

**The Impact of the Blood Pressure-
Associated Genetic Locus at *SLC4A7*
on Gene Expression and
Intracellular pH Regulation**

A thesis submitted to Queen Mary, University of
London for the degree of Doctor of Philosophy

Dr Fu Liang Ng
September 2016

Statement of Originality

In accordance with Queen Mary, University of London regulations, I declare that all the work within this thesis is my own.

Dr Fu Liang Ng

September 2016

Abstract

Genome-wide association studies have revealed an association between variation at the *SLC4A7* locus and blood pressure. *SLC4A7* encodes the electroneutral $\text{Na}^+/\text{HCO}_3^-$ co-transporter NBCn1 which regulates intracellular pH (pH_i) in a range of tissues, including vascular smooth muscle and endothelium. Notably, the *SLC4A7* knockout mouse has been shown to have an altered blood pressure phenotype.

This thesis presents a functional study of variants at this locus in primary cultures of vascular smooth muscle and endothelial cells. There were genotype-dependent differences in DNA-nuclear protein interactions by formaldehyde-assisted isolation of regulatory elements, electrophoretic mobility shift assays and DNA pulldown assays. Subsequently, there were also genotype-dependent differences in *SLC4A7* expression level and NBCn1 availability at the plasma membrane. In turn, *SLC4A7* genotype is associated with $\text{Na}^+/\text{HCO}_3^-$ -dependent steady-state pH_i and recovery from intracellular acidosis. The genotypic effect on pH_i regulation was independent of the calcineurin activity, or the amino acid substitution E326K resulting from a missense polymorphism. However, in the presence of Na^+/H^+ exchange activity, the *SLC4A7* genotypic effect on net base uptake and steady-state pH_i was detected only in vascular smooth muscle cells but not endothelial cells.

The finding of a genotypic influence on *SLC4A7* expression and pH_i regulation in vascular smooth muscle cells provide an insight into the molecular mechanism underlying the association of variation at the *SLC4A7* locus with blood pressure.

Acknowledgements

I would like to thank my supervisors, Professor Shu Ye and Professor Mark Caulfield, for their endless support, intellectual discussions and honesty, even through difficult times.

I would like to thank “the team”, not just for the scientific, emotional and even physical (when I broke my hand!) support. In alphabetical order (as always): Andy, for being such a great buddy and for talks of how football of the highest quality never happens on a Thursday. Kate, for being “the boss” and putting up with me more than anyone should. Meixia, for being the epitome of kindness. Rosie, for getting me out of my shell more. Sue, for doing all the things we don’t see - we know you do it, otherwise the department falls down!

I would also like to thank the rest of the Clinical Pharmacology department, for making my time far more enjoyable, and not to mention the cathartic releases over cups of tea. Thanks need also be extended to the new friends made within the Institute (with special thanks to football buddies and Team-0), who inspire me with the thought that the next generation of scientists will be kind, supportive and enthused about collaborative Science. That, and of course for all the fun times we’ve had.

I am grateful to the team at the Department of Biomedicine in Aarhus University, Denmark. There are again too many to name, but in particular to Christian Aalkaer and Ebbe Boedtkjer for allowing me to go over, and supporting me while I was there. To Viola who was extremely kind and helpful in getting me settled in. To Thomas, who we worked together with ViVian over many hours, and for shifting his experiments to fit my ridiculous schedules. To Renée, who kept me company in the dark room, with interesting discussions, and not just those relating to blank pages and American diners.

I would like to thank other researchers that helped in various ways. Dianne Cooper, Hefin Jones, Patricia Souza and Rachael Wright for their generous efforts in deriving HUVECs. Jay for her help in collecting umbilical cords from the Royal London Hospital maternity unit. To Sarah Headland, for her constant ideas, even if they were way too expensive to be feasible. To Gill Adams and Martin Goss, for training and allowing me to use their equipment.

I would like to thank the anonymous donors of umbilical cords that formed the basis of the work conducted on primary cell lines, and the British Heart Foundation together with the Bart and The London Charity for their generous support. None of the work in this thesis would have been possible without them.

I would like to thank Nilisa Bhagat and Sunil Shah. Although their work during their BMedSci projects did not make it into this thesis, I am grateful for the training they have given me towards being a parent. All those hours of waiting, worrying, cleaning up their mess, surprises (both the good and not-so-good), the occasional disagreements, but always joy. I am proud of the both of you, and I hope I have done as much for you, as you have for me.

I would also like to thank a few of my many inspirations on my path in Science and Medicine. Dr Kay Colston, who took me as a summer student in her lab when I was aged 18. I am grateful not just for her scientific guidance, but her enthusiasm in spreading her love of Science. Without her, I have no doubt my life would look very different right now. I am also very grateful to Dr Helen Swannie and Dr Nicola Walters, my first ever Medical Registrars. If I ever become half as kind, half as thoughtful and half as supportive as them, I know I'll be a good boss.

I am also grateful for my friends outside of the research bubble I have lived in for the past few years. There are too many to mention by name, but you know who you are. Thank you for perking me up when required, keeping me sane throughout, and providing things to look forward to. Additional mention to Sarah for her care and patience when work gets a bit much.

Finally, I would like to express my eternal indebtedness to my family. To my brother, Fu Siong, who has a positive impact on my life in innumerable ways, to an unquantifiable extent, that I hope he truly understands. To my niece, Sophie, who over just a few months inspires me to be better. To my parents – put simply, I owe them everything. Everything.

For my family

Table of contents

Abstract.....	3
Acknowledgements.....	4
Table of contents	7
List of tables	13
List of figures.....	15
Abbreviations	19
1. Background.....	22
1.1. Blood pressure, genetics and GWAS (genome-wide association studies).....	22
1.2. <i>SLC4A7</i> and blood pressure GWAS	24
1.3. Intracellular pH (pH _i) regulation - NBCn1 and other acid-base transporters	28
1.4. Possible mechanisms of pH _i affecting blood pressure – vascular smooth muscle	32
1.5. Possible mechanisms of pH _i affecting blood pressure – endothelial function	33
1.6. Possible mechanisms of pH _i affecting blood pressure – combined effects of vascular smooth muscle and endothelial function.....	33
1.7. Possible mechanisms of pH _i affecting blood pressure – renal salt- and water- handling	34
1.8. Possible mechanisms of pH _i affecting blood pressure – other tissues	35
1.9. NBCn1 splice variants and impact on function	36
1.10. Post-translational modifications and examples from other NBCs.....	39
1.11. Protein trafficking and examples from other pH _i regulators	42
1.12. Inducibility of NBCn1 (and other acid-base co-transporters)	42
1.13. Observational studies of pH _i in hypertension	47
1.14. Limited human <i>in vivo</i> experience with modulation of pH _i	49
1.15. NBCn1 and intracellular sodium (Na ⁺) – an alternative mechanism?	50
1.16. NBCn1 and intracellular lead (Pb ²⁺) – an alternative mechanism?	51
1.17. Pleiotropy of <i>SLC4A7</i> – association with cancer and addiction	53
1.18. Other BP-associated SNPs in the region - Chinese GWAS identifying rs820430 as independent risk SNP	56
1.19. Considerations of NBCn1 as pharmacological target.....	58
1.20. Summary of background.....	59
1.21. Hypothesis and aims	59
2. Methods and materials	61
2.1. Tissue culture and cell lines	61
2.1.1. Principles - Choice of cell lines	61
2.1.2. Principles - Tissue culture	64
2.1.3. Human umbilical cord specimens - Ethical approval and exclusion criteria.....	66
2.1.4. Materials	67
2.1.5. Methods.....	68

2.2.	DNA isolation from cells or tissues	71
2.2.1.	Principles	71
2.2.2.	Methods	71
2.3.	RNA isolation and reverse transcription into cDNA (complementary DNA).....	72
2.3.1.	Principles – RNA isolation	72
2.3.2.	Principles - Reverse transcription	74
2.3.3.	Method.....	74
2.4.	Genotyping by KASP™ (Kompetitive Allele Specific PCR) method.....	78
2.4.1.	Principles – Genotyping overview.....	78
2.4.2.	Principles – KASP™ methodology	79
2.4.3.	Methods	83
2.5.	Polymerase chain reaction (PCR) and subsequent analyses.....	84
2.5.1.	Principles - PCR.....	84
2.5.2.	Principles - Primer design.....	86
2.5.3.	Principles - Reaction optimization and controls	88
2.5.4.	Principles - Agarose gel electrophoresis	89
2.5.5.	Principles - Isolation of PCR products	90
2.5.6.	Methods	90
2.6.	Allelic expression imbalance analysis	92
2.6.1.	Principles.....	92
2.6.2.	Methods	94
2.7.	qRT-PCR (quantitative reverse transcription-polymerase chain reaction).....	97
2.7.1.	Principles - History	97
2.7.2.	Principles - Detection systems	98
2.7.3.	Principles - Analysis.....	101
2.7.4.	Methods.....	102
2.8.	Protein isolation.....	104
2.8.1.	Principles.....	104
2.8.2.	Methods.....	108
2.9.	Bicinchoninic acid assay for sample protein quantification	111
2.9.1.	Principles.....	111
2.9.2.	Methods.....	111
2.10.	SDS-PAGE (Sodium dodecylsulfate polyacrylamide gel electrophoresis) and immunoblotting (Western Blotting)	112
2.10.1.	Principles.....	112
2.10.2.	Methods.....	116
2.11.	DNA pull-down assays.....	119
2.11.1.	Principles.....	119
2.11.2.	Methods	122
2.12.	Co-immunoprecipitation.....	125
2.12.1.	Principles.....	125
2.12.2.	Methods.....	126
2.13.	Isoelectric focusing.....	126
2.13.1.	Principles.....	126
2.13.2.	Methods.....	127

2.14.	Electrophoretic mobility shift assays (EMSA)	128
2.14.1.	Principles	128
2.14.2.	Variations – Western-Electrophoretic mobility shift assay (WEMSA)	132
2.14.3.	Methods	132
2.15.	Chromatin immunoprecipitation (ChIP) and formaldehyde-assisted isolation of regulatory elements (FAIRE) assays	135
2.15.1.	Principles	135
2.15.2.	Method	137
2.16.	Immunocytochemistry	140
2.16.1.	Principles	140
2.16.2.	Materials	143
2.16.3.	Methods	144
2.17.	Collagen gel contractility assay	145
2.17.1.	Principles	145
2.17.2.	Methods	148
2.18.	Cloning, mutagenesis and bacterial transformation	148
2.18.1.	Principles	148
2.18.2.	Methods	154
2.19.	Transfection	167
2.19.1.	Principles	167
2.19.2.	Methods	168
2.20.	Intracellular pH (pH_i), buffering capacity and Na^+/HCO_3^- -dependent pH_i recovery ...	173
2.20.1.	Principles	173
2.20.2.	Materials	181
2.20.3.	Methods	184
2.20.4.	Analysis	185
2.21.	Primers	187
2.21.1.	KASPar genotyping primers	187
2.21.2.	End-point PCR primers	187
2.21.3.	SYBR Green® qRT-PCR primers	187
2.21.4.	EMSA primers	188
2.22.	Antibodies	190
2.23.	Reagents	191
2.24.	Statistical analyses	194
3.	Analysis of single nucleotide polymorphisms (SNP) at the <i>SLC4A7</i> locus using genomic prediction tools	196
3.1.	Genetic location of rs13082711 and other nearby genes – <i>NEK10</i> and <i>EOMES</i>	196
3.2.	Predicted allele frequencies and genotype distributions	200
3.3.	eSNPs (expression single nucleotide polymorphisms) prediction	202
3.4.	TFBS (transcription factor binding site) prediction	208
3.5.	Nucleosome occupancy prediction	209
3.6.	Predicting the functional impact of non-synonymous single nucleotide polymorphisms (nsSNP) and the resultant amino acid change	211

3.7.	Predicting the functional impact of nsSNP on 3-dimensional structure of protein	215
3.8.	Predicting the potential impact of the nsSNP on post-translational modifications	218
3.9.	Predicting the impact of the nsSNP on the isoelectric point (pI).....	224
3.10.	Predicting the impact of the nsSNP on enzymatic protein digestion and potential differences in mass spectrometry signals	225
3.11.	ESE (exonic splicing enhancer) site prediction.....	225
3.12.	miRNA (micro RNA)-mRNA interaction prediction	228
3.13.	Epigenomic predictions - EncODE (Encyclopedia Of DNA Elements), NIH Roadmap and the Multiple Tissue Human Expression Resource (MuTHER)	229
3.14.	Potential inducibility of <i>SLC4A7</i> expression.....	234
3.15.	Other theoretical molecular mechanisms not assessed by prediction tools.....	236
3.16.	Limited supportive bioinformatics evidence at rs820430	236
3.17.	Summary of prediction tool analyses	237
4.	Results	238
4.1.	<i>SLC4A7</i> genotyping	238
4.2.	Gene expression.....	241
4.2.1.	End-point reverse-transcriptase polymerase chain reaction (RT-PCR).....	241
4.2.2.	Allelic imbalance analysis.....	242
4.2.3.	Quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR)	243
4.2.4.	NBCn1 protein immunoblots for total cellular protein.....	248
4.2.5.	Isoelectric focusing.....	249
4.3.	Studies into molecular mechanisms underlying differences in gene expression	250
4.3.1.	Chromatin immunoprecipitation (ChIP).....	250
4.3.2.	Formaldehyde-assisted isolation of regulatory elements (FAIRE) assays.....	251
4.3.3.	Electrophoretic mobility shift assays (EMSA)	252
4.3.4.	DNA pull-down assay	262
4.4.	Impact of non-synonymous SNP rs3755652	264
4.4.1.	Co-immunoprecipitation and calcineurin A (CnA)	264
4.4.2.	Immunocytochemistry	265
4.4.3.	Subcellular fractionation and localization of NBCn1	267
4.4.4.	Total cellular NBCn1 in transfected A10 cells	268
4.5.	Inducibility.....	269
4.6.	<i>In vitro</i> cellular function	271
4.6.1.	Collagen contractility gel.....	271
4.6.2.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis - HUASMCs	274
4.6.3.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis – HUASMCs (by cell morphologies).....	280
4.6.4.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis – HUASMCs and responses to FK506 and BAPTA ...	285
4.6.5.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis - HUVECs	288

4.6.6.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis – comparing HUASMCs and HUVECs	292
4.6.7.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis – overexpression models in A10 cells	296
4.6.8.	Na ⁺ /HCO ₃ ⁻ -dependent, DMA-insensitive pH _i recovery following ammonium chloride prepulse-induced intracellular acidosis – responses to FK506 in A10 overexpression models	299
5.	Discussion.....	301
5.1.	<i>SLC4A7</i> genotyping	301
5.1.1.	rs13082711, rs13096477 and rs3755652 genotype distribution in Hardy-Weinberg equilibrium	301
5.1.2.	rs820430 genotype distribution in Hardy-Weinberg equilibrium	301
5.2.	Gene expression.....	302
5.2.1.	<i>SLC4A7</i> (but not <i>NEK10</i> and <i>EOMES</i>) mRNA detected in a selection of samples ...	302
5.2.2.	Allelic imbalance analysis identifies the risk (minor) allele at rs13096477 as being more prevalent in mRNA.....	303
5.2.3.	qRT-PCR did not demonstrate an allele-dependent association with <i>SLC4A7</i> mRNA/cDNA levels or splicing, but did demonstrate a large inter-sample variability	303
5.2.4.	NBCn1 protein immunoblots identifies the risk (minor) allele as having a higher expression in total cell protein of HUASMCs but not HUVECs.....	305
5.2.5.	<i>SLC4A7</i> E326K variation does not impact on isoelectric focusing	306
5.3.	Molecular mechanisms underlying gene transcription	307
5.3.1.	Chromatin immunoprecipitation (ChIP) suggested an allele-dependent chromatin interaction at rs13096477	307
5.3.2.	Formaldehyde-assisted isolation of regulatory elements (FAIRE) assays identified the protective (major / T) allele at rs13096477 as preferentially chromatin-bound in HUASMCs but not HUVECs.....	308
5.3.3.	Electrophoretic mobility shift assays (EMSA) indicate allele-dependent in vitro nuclear protein binding at rs13096477, rs2371065 and rs13077400.....	308
5.3.4.	DNA pull-down showed in vitro nuclear protein interaction with sequence centered on rs2371065, but unable to reveal identity of protein.....	310
5.4.	Impact of non-synonymous SNP on protein behavior and function	311
5.4.1.	Calcineurin A did not co-immunoprecipitate with NBCn1 in HUASMCs.....	311
5.4.2.	Immunocytochemistry revealed NBCn1 position both on the nuclear membrane, and also intracellularly	312
5.4.3.	Subcellular protein fractionation identified protein variant (326E) associated with the protective (major / T) allele at rs13096477 as preferentially abundant in the membrane fraction	313
5.4.4.	Transfection of overexpression plasmids suggests that the E326K NBCn1 variation is unlikely to impact on NBCn1 degradation	314
5.5.	Impact of <i>SLC4A7</i> genotype on inducibility	314
5.6.	Impact of <i>SLC4A7</i> genotype on cell function	315
5.6.1.	Collagen gel contractility assays – lacks the rapid response required and is unable to withstand the necessary pH range.....	315
5.6.2.	Na ⁺ /HCO ₃ ⁻ -dependent DMA-insensitive pH _i recovery from intracellular acidosis..	316
6.	Conclusions: summary, key limitations, future studies and implications of results	332
6.1.	Key limitations.....	333

6.2. Possible future studies.....	335
6.3. Implications of results.....	337
References	339

List of tables

Table 1: Selected NBC and NHE knockout mice traits	31
Table 2: Summary of studies on the impact of phosphorylation state on NBC and NHE activity	41
Table 3: Summary of studies on the impact of altering pH on expression of acid-, base-, sodium-transporters of channels.....	43
Table 4: Summary of studies on the impact of RAAS activation on expression of acid-, base-, sodium- transporters or channels.....	45
Table 5: Summary of studies on the impact of NBC / NHE knockouts or knockdown on expression of acid-, base-, sodium- transporters or channels	46
Table 6: Thermocycler program for KASP assays.....	83
Table 7: Excitation and emission wavelengths for the utilised fluorophores.....	84
Table 8: Designing RT-PCR primers.....	88
Table 9: Thermocycler program for end-point PCR.....	91
Table 10: Real-time PCR program for qPCR.....	104
Table 11: Advantages and disadvantages of the different EMSA reporters.....	129
Table 12: Commonly used fluorophores with respective absorption and emission wavelengths.	142
Table 13: Antibodies used for immunocytochemistry.....	144
Table 14: Considerations in deciding methodology of choice assessing VSMC contraction	147
Table 15: Thermocycler program for cloning of SLC4A7	155
Table 16: Thermocycler program for in vitro site-directed mutagenesis.....	162
Table 17: Thermocycler program for step 1 of sequence deletion by PCR-driven overlap extension	163
Table 18: Thermocycler program for step 2 of sequence deletion by PCR-driven overlap extension	165
Table 19: Characteristics of the plasmids	166
Table 20: KASPar genotyping primers.....	187
Table 21: End-point PCR primers, annealing temperatures and predicted amplicon sizes.....	187
Table 22: SYBR Green® qRT-PCR primers.....	187
Table 23: EMSA primers.....	188
Table 24: EMSA competitor primers.....	189
Table 25: List of antibodies used.	190
Table 26: Reagents and formulations	194
Table 27: Genotype and allele frequency at rs13082711	200
Table 28: Genotype and allele frequency at rs3755652	201
Table 29: Genotype and allele frequency at rs820430.....	202
Table 30: SCANdb output for gene transcripts that vary with each SNP.....	204
Table 31: Transcription factors with different allele-specific scores for binding sequences	208
Table 32: Predicted functional impact of the E326K polymorphism by online tools.	212
Table 33: Phospho.ELM predictions on possible sites of post-translation phosphorylation.	219
Table 34: DISPHOS v1.3 predictions on possible post-translation phosphorylation.	220
Table 35: PhosphoNet predictions of post-translational phosphorylation sites	221
Table 36: ScanSite3 predictions of post-translational phosphorylation sites.....	222
Table 37: Net-O-Glyc v4.0 predictions of possible sites of post-translational O-glycosylation.....	223
Table 38: Predicted impact of the E326K polymorphism on isoelectric point (pI)	224
Table 39: Predicted splicing changes by Human Splicing Finder	227
Table 40: Summarised met-QTL data from the MuTHER / GENEVar database.....	233
Table 41: Genotyping of HUASMC samples.....	238
Table 42: Genotyping of HUVEC samples	239

Table 43: Haplotypes at rs13082711, rs13096477 and rs3755652, for HUASMCs and HUVECs....	240
Table 44: Genotyping of HUASMC and HUVEC samples at rs820430.....	241
Table 45: Distribution statistics	247
Table 46: Oligonucleotide sequences of EMSA competitor probes	257
Table 47: Distribution of HUASMC morphology by rs13082711 genotype	282
Table 48: Standard variations of HUASMC and HUVEC samples	295

List of figures

Figure 1: Linkage disequilibrium plots at rs13082711 based on data across the CEU, YRI, and a mixed panel of CHB and JPT populations	26
Figure 2: Sodium- and bicarbonate-dependent pH _i recovery in wildtype and NBCn1 ^{LacZ/LacZ} gene trap/knockout mice (Boedtkjer et al., 2008)	28
Figure 3: Cartoon identifying different classes of pH _i regulators	29
Figure 4: Representative diagram of motifs within NBCn1 of first 8 variants discovered.....	37
Figure 5: Differences in amino acid sequences of human and rat NBCn1 Cassette II	37
Figure 6: BLAST protein alignments of NBCn1 and other selected acid-base regulators.....	39
Figure 7: Comparing human and mouse NBCn1 amino acid sequences	48
Figure 8: Linkage disequilibrium plots at rs13082711 and rs497376 SNPs	54
Figure 9: Linkage disequilibrium plots at the rs13082711 and rs3278 SNPs.....	56
Figure 10: Linkage disequilibrium plots at the rs13082711 and rs820430 SNPs.....	57
Figure 11: Schema of hypothesized potential mechanisms via which the blood pressure related SNP (rs13082711) exerts its effect.....	59
Figure 12: Agarose gel electrophoresis of RNA isolated using the acid guanidinium thiocyanate-phenol-chloroform extraction technique	76
Figure 13: Cartoon demonstrating the chemical basis of the KASP™ genotyping system	80
Figure 14: Representative figure of a KASP genotyping assay.....	82
Figure 15: PeakPicker analysis of chromatographs conducted in parallel.....	95
Figure 16: PeakPicker analysis of the “reference sequence”	96
Figure 17: PeakPicker output of normalised peak-heights	97
Figure 18: qPCR amplification plot.....	100
Figure 19: qPCR log-linear relationship.....	101
Figure 20: Schematic of primers used for SYBR® Green qPCR.	102
Figure 21: Endpoint PCR with primers used for SYBR® Green qPCR.	103
Figure 22: Fractionation of cytosolic and nuclear extracts.....	106
Figure 23: Differential fractionation of cellular proteins.....	107
Figure 24: Cartoon of subcellular protein fractionation by differential centrifugation	110
Figure 25: Representative bicinchoninic acid assay standard curve	112
Figure 26: Cartoon outlining the DNA-pull down process	120
Figure 27: Streptavidin-alkaline phosphatase method of detecting labelled oligonucleotides	130
Figure 28: Nucleic acid staining method of detecting EMSA with PCR products	130
Figure 29: Cartoon showing sequence of steps in a modified protocol to FAIRE.....	136
Figure 30: ChIP / FAIRE optimization of probe sonication.....	138
Figure 31: Cartoon overview of in vitro site-directed mutagenesis	150
Figure 32: Cartoon overview of deletion by PCR-driven overlap extension.....	151
Figure 33: Restriction enzyme selection for pcDNA3.1(+)	154
Figure 34: Endpoint PCR of cloning products	156
Figure 35: Results of NEB Double Digest Finder	157
Figure 36: Bacterial growth on petri dishes after ligation of cloning insert and vector	159
Figure 37: Gel electrophoresis of plasmids before and after KpnI/HindIII digestion	161
Figure 38: Design rationale for mutagenic primers	161
Figure 39: End-point PCR of sequence deletion by PCR-driven overlap extension (Step 1).....	164
Figure 40: End-point PCR of splicing by PCR-driven overlap extension (Step 2)	165
Figure 41: Restriction mapping of plasmids.....	166
Figure 42: RT-PCR for transfected of A10 cells	170
Figure 43: Protein immunoblots for transfected HEK293 and A10 cells	171

Figure 44: Protein immunoblots for transfected HEK293 with dose and time variation	171
Figure 45: Merged inverted fluorescent microscopy images for transfected A10 cells	172
Figure 46: pH-dependent fluorescence excitation spectra of BCECF	174
Figure 47: Experimental protocol investigating influence of NHE (by DMA inhibition) and NBC (by removal of CO ₂ and HCO ₃ ⁻) in HUVECs and HUASMCs	177
Figure 48: Experimental protocol investigating influence of calcineurin (by FK506 inhibition) and intracellular Ca ²⁺ (by BAPTA-AM chelation) in HUASMCs	178
Figure 49: Free intracellular calcium in response to ammonium prepulse in HUASMCs with and without BAPTA-AM loading	178
Figure 50: Na ⁺ -independent pH _i recovery	179
Figure 51: Effect of cell culture surface and substrate on pH _i recovery	180
Figure 52: Example of nigericin and pH clamping calibration	181
Figure 53: pH _i calibration curve	186
Figure 54: Linkage disequilibrium plots at the rs13082711 and rs3806624 SNPs	197
Figure 55: Genotype-Tissue Expression project (GTEx) SLC4A7, NEK10 and EOMES expression data on a panel of human tissue samples	199
Figure 56: Genotype-Tissue Expression project (GTEx) SLC4A7 and NEK10 expression data of selected tissues for the rs13082711 SNP	206
Figure 57: Prediction of nucleosome occupancy with the respective alleles at rs3755652	210
Figure 58: BLAST protein query of the human NBCn1 Cassette II	214
Figure 59: Example of prediction tools converting continuous scores to categorical outputs	215
Figure 60: SWISS-MODEL predicted 3-dimensional representation of native conformational changes with the E326K variants	216
Figure 61: PSIPRED prediction of 3-dimensional conformation with the E326K variants	217
Figure 62: I-TASSER prediction of 3-dimensional conformation with the E326K variants	218
Figure 63: Representative output of ESE scores from ESEFinder	226
Figure 64: EncODE outputs for the 93 SNPs of interest	230
Figure 65: Connectivity Score output for NBCn1	235
Figure 66: End-point RT-PCR (SLC4A7, NEK10 and EOMES) for HUASMCs, HUVECs and nephrectomy biopsies	242
Figure 67: Allelic imbalance analyses using heterozygote samples at the exonic SNP rs13096477	243
Figure 68: qRT-PCR of HUASMC and HUVEC cDNA stratified by rs13082711 genotype	244
Figure 69: qRT-PCR of HUASMC and HUVEC cDNA, stratified by rs13082711 genotype or cell type	245
Figure 70: qRT-PCR results demonstrating a skewed distribution with a long tail	246
Figure 71: qRT-PCR of HUASMC and HUVEC cDNA stratified by rs820430 genotype	247
Figure 72: Protein immunoblots for NBCn1 with HUASMC samples stratified by genotype at rs13082711	248
Figure 73: Protein immunoblots for NBCn1 with HUVEC samples stratified by genotype at rs13082711	249
Figure 74: Isoelectric focusing of HUASMCs whole cell lysates	249
Figure 75: Chromatin immunoprecipitation (ChIP) assay with anti-Pol II antibody	251
Figure 76: Formaldehyde-assisted isolation of regulatory elements assays at rs13096477	252
Figure 77: Rationale in selecting 10 SNPs (out of 93) for further investigation with EMSA	253
Figure 78: EMSA with labelled-probes corresponding to both alleles of 10 SNPs	254
Figure 79: EMSA comparing HUVEC and HUASMC nuclear extracts for three SNPs	255
Figure 80: EMSA studies at rs2371065, rs13096477 and rs13077400	256
Figure 81: Unlabelled double-stranded oligonucleotide competitor EMSA studies	258
Figure 82: Supershift assays at rs2371065 targeted against PHOX2A, PLAGL1 and TFAP2C	259

Figure 83: Western-EMSA (WEMSA) at rs2371065	260
Figure 84: End-point RT-PCR for PHOX2A, PLAGL1, TFAP2C and TRMT1	261
Figure 85: DNA-pull down with double-stranded oligonucleotide centering on rs2371065	262
Figure 86: N-terminal amino acid sequencing chromatographs of DNA-pull down products	263
Figure 87: Calcineurin A does not co-immunoprecipitate with NBCn1 in transfected A10 cells ...	264
Figure 88: Confocal microscopy staining for VE-cadherin and NBCn1 in HUVECs.....	266
Figure 89: Inverted fluorescent microscopy images staining for NBCn1 in HUASMCs.....	267
Figure 90: Subcellular fractionation samples from HUASMCs and HUVECs.....	268
Figure 91: NBCn1 expression in transfected A10 cells.....	269
Figure 92: Inducibility of NBCn1 by angiotensin II, and the relevance of genotype.....	269
Figure 93: Digital light microscopy images (10x objective) of serum-starved HUVECs over 48 hours.	270
Figure 94: Time series for collagen contractility gels.....	271
Figure 95: Collagen contractility gels intra-sample replicability and measurement optimisation.	272
Figure 96: Intra-observer variability of collagen gel contractility assays measurements	272
Figure 97: Inter-observer variability of collagen gel contractility assays measurements	273
Figure 98: Time series for collagen contractility gels with various pharmacological stimuli	274
Figure 99: Digital images of collagen gels exposed to pH _o 6.8.....	274
Figure 100: HUASMC pH _i recovery following intracellular acidosis, with and without CO ₂ /HCO ₃ ⁻ or DMA	275
Figure 101: Serum-starved HUASMC pH _i recovery following intracellular acidosis, with CO ₂ /HCO ₃ ⁻ and DMA	276
Figure 102: Initial net base uptake of HUASMCs in sodium-free and sodium-containing PSS in presence of CO ₂ /HCO ₃ ⁻ and DMA.....	276
Figure 103: pH _i at which initial base flux measurements were obtained.....	277
Figure 104: Rates of net base uptake at pH _i intervals for HUASMC following intracellular acidosis, with CO ₂ /HCO ₃ ⁻ and DMA	278
Figure 105: Final plateau pH _i after recovery from intracellular acidosis	279
Figure 106: HUASMC pH _i recovery following intracellular acidosis, without CO ₂ /HCO ₃ ⁻ , and in the presence of DMA	279
Figure 107: Buffering capacity of HUASMCs in the presence and absence of CO ₂ /HCO ₃ ⁻	280
Figure 108: Representative images of two different HUASMC morphologies	281
Figure 109: HUASMC pH _i recovery following intracellular acidosis, with and without CO ₂ /HCO ₃ ⁻ or DMA, or in 48 hour serum-starved cultures	283
Figure 110: HUASMC pH _i recovery following intracellular acidosis, with and without CO ₂ /HCO ₃ ⁻ , with DMA, or in 48 hour serum-starved cultures in confluent subgroup	284
Figure 111: HUASMC pH _i recovery following intracellular acidosis, with CO ₂ /HCO ₃ ⁻ and DMA, ± FK506 or BAPTA-AM	285
Figure 112: Rates of net base uptake at pH _i intervals for HUASMC following intracellular acidosis, with CO ₂ /HCO ₃ ⁻ and DMA, ± FK506 or BAPTA-AM.....	286
Figure 113: Final plateau pH _i after recovery from intracellular acidosis with CO ₂ /HCO ₃ ⁻ and DMA, ± FK506 or BAPTA-AM	287
Figure 114: HUASMC pH _i recovery following intracellular acidosis, with CO ₂ /HCO ₃ ⁻ , DMA and FK506	287
Figure 115: HUVEC pH _i recovery following intracellular acidosis, with and without CO ₂ /HCO ₃ ⁻ or DMA	288
Figure 116: Initial net base uptake of HUVECs in sodium-free and sodium-containing PSS in presence of CO ₂ /HCO ₃ ⁻ and DMA.....	289
Figure 117: pH _i at which initial base flux measurements were obtained.....	289

Figure 118: Net base uptake at pH_i intervals for HUVECs following intracellular acidosis, with CO_2/HCO_3^- and DMA	290
Figure 119: Final plateau pH_i after recovery from intracellular acidosis	291
Figure 120: HUVEC pH_i recovery following intracellular acidosis, without CO_2/HCO_3^- , and in the presence of DMA	291
Figure 121: Buffering capacity of HUASMCs in the presence and absence of CO_2/HCO_3^-	292
Figure 122: HUASMC and HUVEC pH_i recovery following intracellular acidosis	293
Figure 123: HUASMC and HUVEC net base uptake in presence of CO_2/HCO_3^- and DMA.....	294
Figure 124: Final plateau pH_i after recovery from intracellular acidosis	295
Figure 125: SLC4A7-expression plasmid-transfected A10 cells pH_i recovery following intracellular acidosis with CO_2/HCO_3^- and DMA.....	296
Figure 126: Net base uptake of transfected A10 cells in presence of CO_2/HCO_3^- and DMA	297
Figure 127: Final plateau pH_i after recovery from intracellular acidosis	298
Figure 128: NBCn1 immunoblot signals in transfected A10 cells	298
Figure 129: Transfected A10 cells pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- and DMA \pm FK506.....	299
Figure 130: Final plateau pH_i after recovery from intracellular acidosis	300
Figure 131: NBCn1 protein expression in serum-treated and -starved HUASMC samples undergoing pH_i studies.....	317
Figure 132: NBCn1 protein expression in serum-treated and -starved HUASMC samples undergoing pH_i studies.....	318
Figure 133: HUASMC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- , and in the presence of DMA and FK506.....	320
Figure 134: pH_i recovery rates of HUASMCs and HUVECs in different conditions	323
Figure 135: End-point RT-PCR of selected pH_i regulators.....	324
Figure 136: HUASMC pH_i recovery following intracellular acidosis, in the presence of CO_2/HCO_3^-	328
Figure 137: Final plateau pH_i after recovery from intracellular acidosis	329
Figure 138: HUVEC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^-	329
Figure 139: Estimate costings for increasing sample size	334

Abbreviations

[Na ⁺] _i	Intracellular Na concentrations
Ang II	Angiotensin II
ANOVA	Analysis of variance
ANP	Atrial natriuretic peptide
APS	Ammonium persulfate
BAPTA-AM	1',2'-bis(2-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid tetrakis-acetoxymethyl ester
BCA	Bicinchoninic acid
BCECF-AM	2',7'-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein-acetoxymethyl ester
BLAST	Basic Local Alignment Search Tool
BP	Blood pressure
cAMP	3'-5'-cyclic adenosine monophosphate
CCB	Calcium-channel blocker
cDNA	Complementary deoxyribonucleic acid
CEU	Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain registry
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHB	Han Chinese in Beijing, China
ChIP	Chromatin immunoprecipitation
CV	Coefficient of variance
DAPI	4',6-diamidino-2-phenylindole
ddNTP	dideoxynucleoside triphosphates
DIDS	4,4'-Diisothiocyanatostilbene-2,2'-disulfonic acid
DMA	5-(N,N-Dimethyl)amiloride hydrochloride
DMEM	Dulbecco's modified Eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleoside triphosphates
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EIPA	5-(N-Ethyl-N-isopropyl)amiloride
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia Of DNA Elements
eQTL	Expression quantitative trait loci
ESE	Exonic splicing enhancer
eSNP	Expression single nucleotide polymorphism
FAIRE	Formaldehyde-assisted isolation of regulatory elements
FAM	6-Carboxyfluorescein
FBS	Fetal bovine serum
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
GI	Gastrointestinal
GTex	Genotype-Tissue Expression project
GWAS	Genome-wide association study
HBSS	Hanks' Balanced Salt Solution

HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
HUASMC	Human umbilical artery smooth muscle cell
HUVEC	Human umbilical vein endothelial cell
IB	Immunoblot
ICC	Immunocytochemistry
IP	Immunoprecipitation
IRBIT	Inositol 1,4,5-trisphosphate receptor-binding protein
JPT	Japanese in Tokyo, Japan
KASP	Kompetitive Allele Specific PCR
LD	Linkage disequilibrium
L-NAME	L-arginine methyl ester
MCT	Monocarboxylate transporters
miRNA	Micro-ribonucleic acid
M-MLV	Moloney Murine Leukemia Virus
mRNA	Messenger ribonucleic acid
mTAL	Medullary thick ascending loop of Henle
MuTHER	Multiple Tissue Human Expression Resource
NAd	Noradrenaline
NBC	Sodium-bicarbonate co-transporter
NHANES	National Health and Nutrition Examination Survey
NHE	Sodium-hydrogen exchanger
NIH	National Institute of Health
NMDG	<i>N</i> -methyl-D-glucammonium
NO	Nitric oxide
Nonidet P-40	Octyl phenoxy polyethoxy ethanol
NP-40	Nonyl phenoxy polyethoxy ethanol
NTC	No template control
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH _i	Intracellular pH
PI	Protease inhibitor
PMSF	Phenylmethanesulfonyl fluoride
PP1	Protein phosphatase 1
PSS	Physiological saline solution
PVDF	Polyvinylidene fluoride
qRT-PCR	Quantitative reverse transcription-polymerase chain reaction
RAAS	Renin-angiotensin-aldosterone system
RIPA	Radioimmunoprecipitation assay
ROCK	Rho kinase
RT-PCR	Reverse transcription-polymerase chain reaction
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SHR	Spontaneously hypertensive rat
SNAP	SNP Annotation and Proxy Search
SNP	Single nucleotide polymorphism
SPAK	SPS1-related proline/alanine-rich kinase
TBE	Tris-Borate-EDTA

TBS	Tris-buffered saline
TCA	Trichloroacetic acid
TEMED	Tetramethylethylenediamine
TFBS	Transcription factor binding site
T _m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UTR	Untranslated region
VSMC	Vascular smooth muscle cell
W-EMSA	Western-Electrophoretic mobility shift assay
YRI	Yoruba in Ibadan, Nigeria

1. Background

Hypertension is a common (affecting approximately 40% of adults) and major aetiological factor in cardiovascular, cerebrovascular and renovascular disease. It is estimated to cause up to 12.8% of global mortality and 3.7% of morbidity as calculated by disability-adjusted life years (WHO, 2009). Additionally, it has a major economic cost worldwide, estimated at around \$USD 3.6 trillion (3,600,000,000,000) annually when all indirect costs of its complications are taken into account (Gaziano *et al.*, 2009). It poses as a continuous graded risk factor, where a 20 mmHg increase in systolic blood pressure associated with a two-fold increase in risk of cardiovascular and cerebrovascular disease (Lewington *et al.*, 2002). To compound the population health issue, up to 30% of patients with hypertension are inadequately treated (Calhoun *et al.*, 2008), identifying a need for novel therapeutic pathways.

1.1. Blood pressure, genetics and GWAS (genome-wide association studies)

Hypertension is a condition with strong contribution from both genes and their interaction with the environment. Studies on twins reared apart have suggested that the heritability for blood pressure is around the magnitude of 40% (Hong *et al.*, 1994; Kupper *et al.*, 2005), and GWAS has identified a number of candidate genes in the pathogenesis of hypertension (Levy *et al.*, 2009; Newton-Cheh *et al.*, 2009; Ehret *et al.*, 2011; Kato *et al.*, 2011; Wain *et al.*, 2011; Franceschini *et al.*, 2013; Ganesh *et al.*, 2013; Tragante *et al.*, 2013; Kato *et al.*, 2015; as a selection of references fully reviewed as joint author in Cabrera *et al.*, 2015). Our current understanding of the pathophysiology of hypertension has partially been aided by the knowledge gained from monogenetic hypertension, and to a lesser extent candidate gene analyses (Munroe *et al.*, 2013). GWASs have aimed to go further to identify loci that contain DNA variants that are associated with more subtle blood pressure effects. However, these large observational studies can only identify association, and this does not necessarily mean causation, let alone any additional knowledge of an underlying molecular mechanism. Thus far, there have only been a small handful of published studies successfully determining the molecular mechanism underpinning the effects of the SNPs associated with blood pressure from GWAS (Cabrera *et al.*, 2015). This section will summarise some of the few successes thus far.

Within the *UMOD* (uromodulin) locus, the lead GWAS SNP (rs13333226) is a good example of a marker being in high linkage disequilibrium (LD) with the proposed causal SNP (rs4293393). The latter polymorphism is predicted by position weight matrices to be part of a glucocorticoid response element that is lost in the presence of the C (protective) allele. Concordantly, luciferase assays showed that the T (risk) allele resulted in higher promoter activity as compared to the protective allele in three *in vitro* experimental cell lines. This supports the finding of increased *UMOD* transcript levels in human nephrectomy samples of those with the risk haplotype. The physiological relevance of these findings is further emphasised by the authors' *in vivo* human studies identifying a salt-sensitive hypertensive phenotype with the risk genotype. The risk genotype was associated with an elevated diastolic BP in a cohort of 471 never-treated hypertensive patients, and within experimental subgroups, the loop diuretic furosemide produced larger natriuretic and diastolic BP lowering effects (Trudu *et al.*, 2013).

Another example of alterations in regulatory elements is the SNP rs5068 at the *NPPA* locus which was shown to differentially interact with the inhibitory microRNA *miR-425* by co-transfection experiments, where *miR-425* reduced the expression of the *NPPA* plasmid containing the risk (T) allele, but not with the protective (G) allele plasmid. This is concordant with the hypothesis that the risk allele reduces gene expression with deleterious effect on blood pressure regulation (Arora *et al.*, 2012). The role of miRNA also appears to be important in regulating gene expression relating to the rs4846049 locus, positioned at the 3'-end of the *MTHFR* gene (encoding 5,10-methylenetetrahydrofolate). A common coding polymorphism (C677T) has previously been associated with hyperhomocysteinaemia (Frosst *et al.*, 1995), and in turn, blood pressure (Yang *et al.*, 2014). A luciferase assay demonstrated that *miR-149* inhibited the gene expression in the vector carrying the protective (T) allele, but not the risk (G) allele. This result was further supported with the observation that the peripheral blood mononuclear cells of coronary heart disease patients with the T/T genotype has lower 5,10-methylenetetrahydrofolate expression as compared to risk-carrying G/G heterozygotes (Wu *et al.*, 2013).

The potential of one SNP altering transcription factor binding has also been considered. The SNP rs3918226 in the promoter region of the *NOS3* gene was predicted to alter a putative binding site for transcription factors of the ETS (E-twenty six) family. A luciferase reporter study demonstrated that the T (risk) allele was associated with lower transcriptional activity when tested in HeLa and HEK293T cell lines (Salvi *et al.*, 2013). This is concordant with the expectation that reduced

expression of eNOS, the gene product of *NOS3*, would be a risk factor for hypertension. However, there was neither functional confirmation of the predicted interaction with ETS, nor replication in a cell line better reflective of those relevant to blood pressure regulation.

In addition to identifying local effects of SNPs, there is also a drive to determine potential trans-acting regulators; i.e. SNPs that are associated with expression levels of genes that map far away, even on different chromosomes. The SNP rs653178 at the *ATNX2-SH2B3* locus has shown extensive disease pleiotropy including blood pressure levels, but it is still unclear which, if any, of the local genes *ATNX2* (encodes ataxin-2) or *SH2B3* (SH2B adapter protein-3) may be a causative gene in blood pressure regulation. However, rs653178 (located at 12q24) has been shown to be associated with the expression of *CRIP1* (at 14q32.33) and *RAB11FIP1* (at 8p11.22) (Rotival *et al.*, 2011). This gives rise to two more proteins as possible candidate regulatory proteins in blood pressure control, with little known about them in this context.

The limited reports so far of successful follow-ups to GWAS hits in a polygenetic disease with modest individual contributions to an individual blood pressure measurement emphasises the complexities and challenges facing the research in this thesis.

1.2. *SLC4A7* and blood pressure GWAS

A large-scale international GWAS identified a locus encompassing the *SLC4A7* gene, positioned at the 3p24.1 genomic region, as one of several novel loci associated with blood pressure (Ehret *et al.*, 2011). The lead GWAS SNP, rs13082711, showed a significant association with diastolic blood pressure, but not reaching genome-wide significance with systolic blood pressure ($p=3.8 \times 10^{-9}$ and 1.5×10^{-6} respectively). To better describe the locus and the SNPs in strong linkage disequilibrium, the SNP Annotation and Proxy Search (SNAP, Broad Institute) is able to annotate this from HapMap data (Johnson *et al.*, 2008) (Figure 1, Appendix 1 for full dataset). This rs13082711 SNP is located 40kb upstream from the 5' end of the *SLC4A7* gene and is in strong linkage disequilibrium (LD, $r^2 > 0.8$) with 92 other SNPs spanning a genomic region of 136kb in the CEU (Utah residents with Northern and Western European ancestry from the *Centre d'Etude du Polymorphisme Humain* collection) population, with *SLC4A7* being the only gene within this region (Broad Institute, <http://www.broadinstitute.org/mpg/snap/ldsearch.php>). Of these 93 SNPs, only 4 are exonic, of which only one is a non-synonymous polymorphism (E326K).

In comparison, the YRI (Yoruba in Ibadan, Nigeria) population had fewer SNPs in strong LD based on the same pre-specified $r^2 > 0.8$ cut-off. Notably, a third panel data provided includes the combined CHB+JPT panel, comprising of Han Chinese in Beijing, China and Japanese in Tokyo, Japan, despite being two very different populations. With this atypical panel, there is a larger spread of SNPs spanning 250kb, including 18 additional SNPs, of which 8 are 5'-end of the gene, 8 are 3'-end of the gene and 2 are intronic. These SNPs did not span any other genes, and did not reveal any characteristics of interest based on the tools below.

In context, there are currently 441 known common (defined as having a minor allele frequency of >0.01) SNPs in the region of *SLC4A7*. Of these, 2 are within 500 bp upstream of the gene, 2 within the 5'-UTR (untranslated region), 10 within the 3'-UTR, 6 are synonymous exonic, 3 are non-synonymous, and 418 are intronic. Of course, these numbers will likely increase with discoveries from the larger, newer sequencing projects.

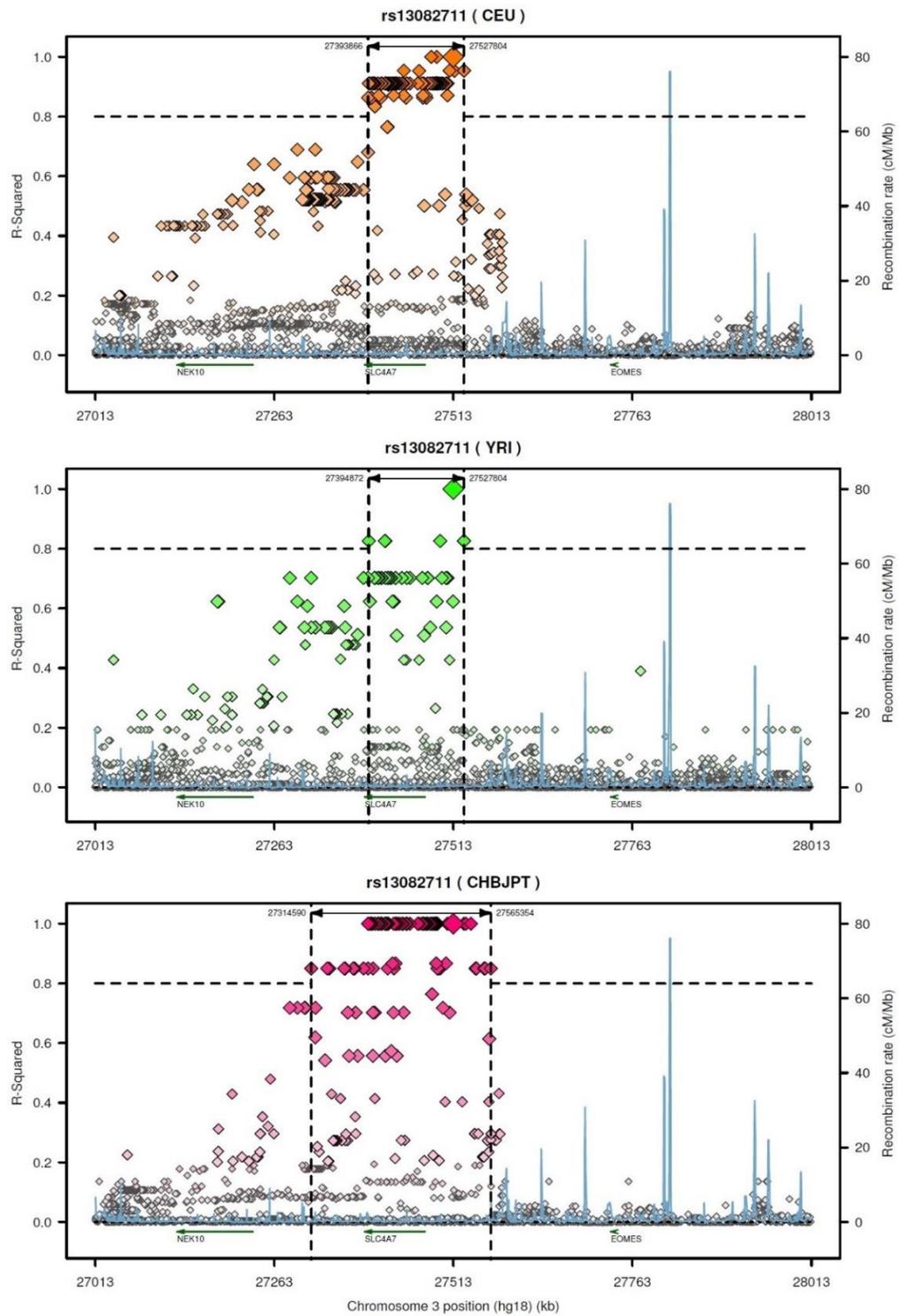


Figure 1: Linkage disequilibrium plots at rs13082711 based on data across the CEU, YRI, and a mixed panel of CHB and JPT populations. Linkage disequilibrium plots at the blood pressure-associated (rs13082711) based on HapMap3 (release 2) data across the CEU, YRI, and a mixed panel of CHB and JPT populations (<http://www.broadinstitute.org/mpg/snap/ldplot.php>). Position along chromosome 3 is matched on the three plots. An LD of $r^2 > 0.8$ (dashed lines) is marked as strong linkage disequilibrium within the loci.

Since GWAS can only identify associations within large populations and cannot infer causality, further studies are required to determine if, and potentially which, polymorphisms can have an effect on gene expression and protein activity, and subsequently on cellular and organ phenotypes.

The *SLC4A7* gene, encodes the electroneutral sodium/bicarbonate co-transporter NBCn1. This symporter functions to co-transport Na^+ and HCO^- ions in a one-to-one ratio (therefore electroneutral) in the intracellular direction. Amongst the cells that express NBCn1 are tissues that are important for blood pressure control – vascular smooth muscle cells (VSMCs) (Damkier *et al.*, 2006; Boedtkjer *et al.*, 2008), endothelial cells (Damkier *et al.*, 2006; Boedtkjer *et al.*, 2008; Boedtkjer *et al.*, 2011) and the renal salt- and water-concentrating cells of the medullary thick ascending loop of Henle cells (Vorum *et al.*, 2000; Damkier *et al.*, 2007; Boedtkjer *et al.*, 2008).

This protein has also been demonstrated to have an important role in controlling intracellular pH (pH_i) (Boedtkjer *et al.*, 2008), where an *in vivo* mouse *SLC4A7* gene trap model reduced NBCn1 expression, decreased resting pH_i and reduced the sodium- and bicarbonate-dependent rate of recovery from ammonium chloride-prepulse induced intracellular acidosis (Figure 2). A recent study showed *SLC4A7*^{-/-} knockout mice had an attenuated response to angiotensin II (Ang II) and nitric oxide synthase inhibition in raising blood pressure, although these mice had higher blood pressure at baseline (Boedtker *et al.*, 2011).

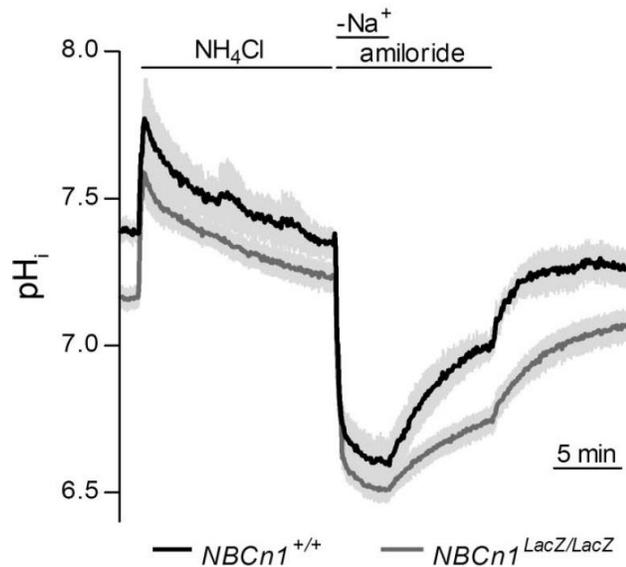


Figure 2: Sodium- and bicarbonate-dependent pH_i recovery in wildtype and $NBCn1^{LacZ/LacZ}$ gene trap/knockout mice (Boedtkjer et al., 2008)

This chapter will initially introduce the potential for pH_i to be an important factor in blood pressure regulation (Sections 1.4 to 1.8), but the potential role of intracellular sodium will also be considered (Section 1.15).

1.3. Intracellular pH (pH_i) regulation - NBCn1 and other acid-base transporters

Intracellular pH is regulated by a range of membrane-bound transporters. The relative importance of the different transporters would depend on the expression profile in the varying tissues. In general, the main transporters responsible for maintaining an pH_i suitable for cellular functioning are NHEs (Na^+/H^+ exchanger) and NBCs (Na^+/HCO_3^- co-transporter). They function to alkalinize (or reduce the acidity of) the cytosol by H^+ extrusion or by importing HCO_3^- respectively. A limited number of cell types utilise NDCBEs (Na^+ -dependent Cl^-/HCO_3^- exchanger), by using the inwardly directed electrochemical Na^+ gradient established by Na^+/K^+ -ATPase pumps. Where anaerobic metabolism occurs, MCTs (monocarboxylate/ H^+ co-transporters) may also play a role. Although may be viewed that intracellular milieu may tend towards acidosis due to the byproducts of cellular metabolism, there are other transporters that acidify the cell. These include AEs (anion exchangers) that imports extracellular Cl^- in exchange for cytosolic HCO_3^- , or less likely PMCA (plasma membrane Ca^{2+} -ATPase), which exchanges cytosolic Ca^{2+} for extracellular H^+ . The latter may therefore decrease pH_i in response to stimuli that elevate intracellular Ca^{2+} . These mechanisms are summarized in

Figure 3.

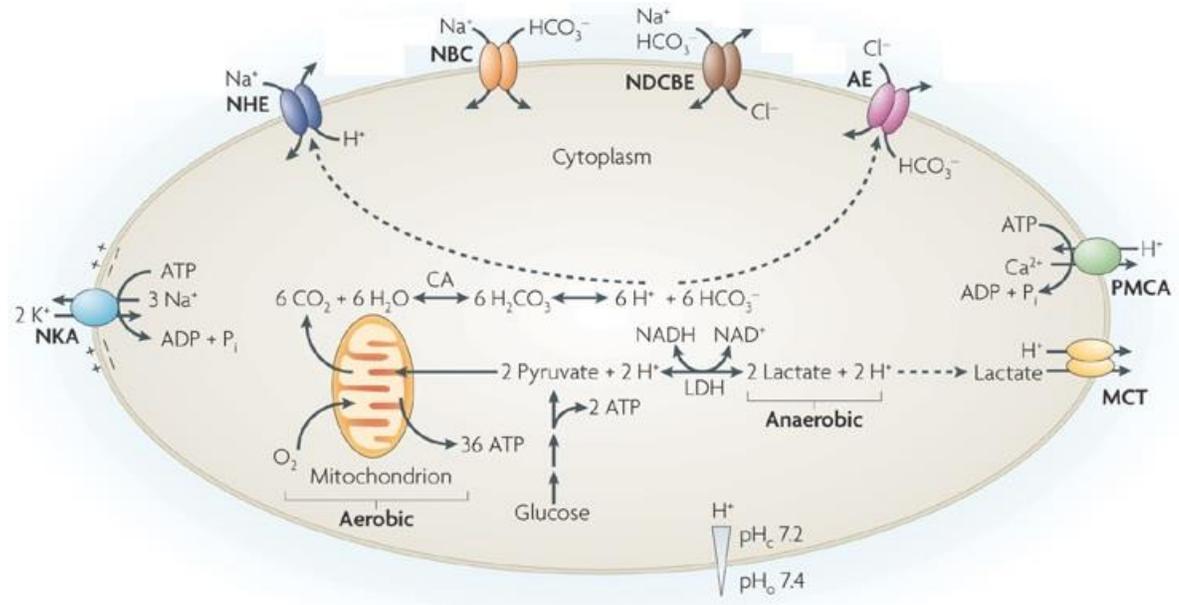


Figure 3: Cartoon identifying different classes of pH_i regulators

There are multiple members within each class of transporter, and the relative importance depends on cell type. NHE, Na^+/H^+ exchanger; NBC, Na^+/HCO_3^- co-transporter; NDCBE, Na^+ -dependent Cl^-/HCO_3^- exchanger; AE, anion exchanger; PMCA, plasma membrane Ca^{2+} -ATPase; MCT, monocarboxylate/ H^+ co-transporters; CA, carbonic anhydrase; NKA, Na^+/K^+ -ATPase pumps; LDH, lactate dehydrogenase; pH_c , cytosolic pH; pH_o , outside/extracellular pH. (Adapted from Casey *et al.*, 2010)

As well-regulated pH_i is crucial for cellular function, it would not be surprising to find compensatory mechanisms in a system with in-built redundancies. However, several pharmacological studies suggest otherwise. Non-selective pharmacological inhibition of all sodium-bicarbonate co-transporters (NBCs) with the small molecule S0859 reduces the pH_i of cat myocytes by 0.2 (De Giusti *et al.*, 2010). This is also reflected by pharmacological agents inhibiting other acid-base transporters in human vascular smooth muscle *in vitro*, with both EIPA (5-(*N*-ethyl-*N*-isopropyl)amiloride) inhibiting NHE and DIDS (4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid) inhibiting the anion exchanger, decreasing pH_i by 0.1 separately (Weissberg *et al.*, 1987; Aalkjaer and Cragoe, 1988). Altogether, this may suggest that despite multiple acid-base transporters having a role in maintaining pH_i , modulation of just one of these proteins may be sufficient to alter pH_i , and potentially cellular phenotype.

It could be argued that the short-term pharmacological *in vitro* studies would not account for potential compensatory changes in expression of other acid-base regulators. However, knockout

models provide better evidence of longer-term reduction in NBCn1 activity. Despite the *SLC4A7*^{-/-} knockout mouse demonstrating an increased NHE1 activity in mesenteric endothelial cells, this does not fully compensate for the absence of NBCn1 function, resulting in a persisting reduction in pH_i (Boedtkjer *et al.*, 2011). This suggests that changes in NBCn1 activity alone may be sufficient to alter intracellular acid-base regulation. Interestingly, the *SLC9A1*^{-/-} (NHE1) knockout mice did not demonstrate a reduction in resting pH_i , but still has the hypotensive phenotype (Boedtkjer *et al.*, 2012). The observation that knockout mice for a variety of acid-base transporters demonstrate persisting phenotypes (Table 1) suggests that these transporters are not redundant.

	<i>SLC4A7</i> ^{-/-} (NBCn1)	<i>SLC4A5</i> ^{-/-} (NBCe2)	<i>SLC9A1</i> ^{-/-} (NHE1)	<i>SLC9A3</i> ^{-/-} (NHE3)
Blood Pressure	↑ BP at rest, but ↓ hypertensive response to AngII or L-NAME (Boedtkjer <i>et al.</i> , 2011)	↑ BP (Groger <i>et al.</i> , 2011)	↓ BP (Boedtkjer <i>et al.</i> , 2012)	↓ BP (Schultheis <i>et al.</i> , 1998).
Intracellular pH	↓ steady state pH _i , ↓ Na ⁺ /HCO ₃ ⁻ -dependent pH _i recovery (Boedtkjer <i>et al.</i> , 2011)		↓ pH _i recovery, but only in absence of CO ₂ /HCO ₃ ⁻ (Boedtkjer <i>et al.</i> , 2012)	
Arterial pH	No change (Chen <i>et al.</i> , 2012)	Metabolic acidosis (Groger <i>et al.</i> , 2011)	No change (Bell <i>et al.</i> , 1999)	Metabolic acidosis (Schultheis <i>et al.</i> , 1998)
Vascular smooth muscle	↓ ROCK-dependent Ca ²⁺ -sensitivity, (Boedtkjer <i>et al.</i> , 2011) ↓ myogenic tone (Thomsen <i>et al.</i> , 2014)		↓ NAd response, ↓ ROCK-dependent Ca ²⁺ -sensitivity in absence of CO ₂ /HCO ₃ ⁻ (Boedtkjer <i>et al.</i> , 2012). ↓ hypoxia-induced pulmonary ↑BP (Yu <i>et al.</i> , 2008).	
Endothelium	↓ NO-mediated relaxation (Boedtkjer <i>et al.</i> , 2011)		↓ NO-mediated relaxation in absence of CO ₂ /HCO ₃ ⁻ (Boedtkjer <i>et al.</i> , 2012)	
Renal		Increased urine volume (Groger <i>et al.</i> , 2011)	↓ transepithelial absorption rate (Good <i>et al.</i> , 2004)	Salt-wasting (Ledoussal <i>et al.</i> , 2001)
GI tract	↓ duodenal HCO ₃ ⁻ secretion (Chen <i>et al.</i> , 2012)	No change in duodenal HCO ₃ ⁻ flux (Chen <i>et al.</i> , 2012)		GI hypertrophy (Schultheis <i>et al.</i> , 1998)
Hormone profile		↓ renin, ↓ aldosterone, ↑ ANP (Groger <i>et al.</i> , 2011)		↑ renin, ↑ aldosterone (Schultheis <i>et al.</i> , 1998)
Others	Non-Mendelian distribution of littermates (Boedtkjer <i>et al.</i> , 2011)		↓ survival (Bell <i>et al.</i> , 1999)	

Table 1: Selected NBC and NHE knockout mice traits

Where multiple publications applicable, earliest is referenced. L-NAME, L-arginine methyl ester; ROCK, rho kinase; NAd, noradrenaline; ANP, atrial natriuretic peptide. Empty cells denote currently unknown phenotype. Yellow cells, likely secondary adaptive changes.

Another key concept is that pH_i is a dynamic variable. pH_i falls in response to a variety of vasoconstrictive stimuli such as angiotensin II, noradrenaline, endothelin-1 and cellular depolarization (Hatori *et al.*, 1987; Aalkjaer and Cragoe, 1988; Touyz and Schiffrin, 1993; Austin

and Wray; 1993b). Changes in NBCn1 expression or activity will impact on the rate of recovery, even if there was an absence of changes in baseline pH_i . This is supported by the observation that NBCn1-mediated Na^+/HCO_3^- co-transport protects vascular smooth muscle cells against intracellular acidification during contractions (Boedtkjer *et al.*, 2006; Danielsen *et al.*, 2013).

1.4. Possible mechanisms of pH_i affecting blood pressure – vascular smooth muscle

Consistent with the hypothesis that pH_i affects vascular function, intracellular acidosis has been previously demonstrated to reduce VSMC contractility, and the converse is also true to intracellular alkalosis (Austin and Wray, 1993a; Horie *et al.*, 1995; Boedtkjer *et al.*, 2011). As intracellular acidification of VSMCs lowers rho-kinase-dependent VSMC Ca^{2+} sensitivity (Boedtkjer *et al.*, 2011; Boedtkjer *et al.*, 2012), the improved ability to eliminate the contraction-induced intracellular acid load may contribute to the higher peripheral resistance of hypertensive patients through increased rho-kinase activity. Consistent with the effect of sustained intracellular acidification on VSMC Ca^{2+} sensitivity, knockout of NBCn1 lowers noradrenaline-induced contractions of mesenteric arteries (Boedtkjer *et al.*, 2011) and myogenic responses of pressurized middle cerebral arteries (Thomsen *et al.*, 2014) after endothelial blockade without affecting VSMC membrane potential or $[Ca^{2+}]_i$.

Notably, studies in the early 1990s using rat mesenteric arteries, a fall of pH_i by 0.1 generated by the weak acid butyrate can reduce *ex vivo* contractility by around 10%, with the converse also true intracellular alkalosis induced by the weak base trimethylamine (Austin and Wray, 1993a). This suggests that modest changes in pH_i , well within potential pharmacological modulation of NBCn1 activity/expression (see Section 1.3), may have an effect on overall peripheral vascular resistance. Additionally, cariporide (NHE1 inhibitor) which reduces pH_i , reversed the hypertensive response to the mineralocorticoid agonist deoxycorticosterone acetate (Young and Funder, 2003). In turn, *SLC4A7*^{-/-} knockouts, mimicking long-term pharmacological inhibition of NBCn1 reduced the hypertensive and vascular reactivity to angiotensin II (Boedtkjer *et al.*, 2011).

There is also evidence that changes in acid-base transport function and/or pH_i impact *in vitro* VSMC proliferation (Wu *et al.*, 2008), migration and viability (Brenninkmeijer *et al.*, 2011), and medial wall thickness (Boedtkjer *et al.*, 2012a). NBCn1 plays a key role for VSMC migration and carotid artery remodeling most likely because it establishes local pH_i gradients and promotes

filopodia, which can explain the decelerated directional migration of VSMCs from NBCn1 knockout mice (Boedtkjer *et al.*, 2016). Altered NBCn1 activity could therefore modify vascular remodeling with long-term impact on peripheral resistance.

1.5. Possible mechanisms of pH_i affecting blood pressure – endothelial function

Intracellular pH also appears to have an impact on endothelial function (Besse *et al.*, 2006; Boedtkjer *et al.*, 2011). In *SLC4A7*^{-/-} knockout mice, endothelial function was impaired as compared to wildtype mice. This was demonstrated by differences in acetylcholine-stimulated NO levels, L-NAME (NO-synthase inhibitor, N-nitro-L-arginine methyl ester)-sensitive L-arginine conversion and *ex vivo* acetylcholine-dependent relaxation. Due to the already diminished NO-tone, these mice were also resistant to the blood pressure raising effect of NO-synthase inhibition (Boedtkjer *et al.*, 2011). Similarly, myogenic tone in cerebral arteries of these mice did not increase as much as wildtypes following treatment with L-NAME. Importantly, there were no differences in membrane potential or $[Ca^{2+}]_i$ between arteries from NBCn1 knockout and wild-type mice, and it was shown that that NO production and α -kinase-dependent Ca^{2+} sensitivity were similarly reduced at low pH_i in pressurized mouse middle cerebral arteries (Thomsen *et al.*, 2014).

Intracellular pH has also been shown to impact on the expression of gene products important in hypertension. In the endothelial cells, this includes the vasoactive peptide endothelin (Cukiernik *et al.*, 2004) and nitric oxide synthase, the enzyme responsible for the generation of another vasoactive substance nitric oxide (Fleming *et al.*, 1994; Thomsen *et al.*, 2014).

1.6. Possible mechanisms of pH_i affecting blood pressure – combined effects of vascular smooth muscle and endothelial function

While these studies provide evidence of how the vascular smooth muscle and endothelium may respond to changes in pH_i independent of each other, this may not be representative of the *in vivo* interaction between the two cell types. There is however only limited studies utilizing both cell types concurrently. The use of whole-animal knockout models (Boedtkjer *et al.*, 2011) is complicated by the impact on other organs, and in particular the kidney.

As expressed in Section 1.5, the potential impact of endothelial pH_i on overall vasculature may be in part via the release of vasoactive substances which may be assayed in isolated endothelial cells. However, when in its native form, there are myoendothelial gap junctions that allow the spread of cellular hyperpolarization and the subsequent vasorelaxation. In *ex vivo* preparations of mouse mesenteric arteries, alkalinisation of the endothelium (from an average pH_i of 7.24 up to greater than 7.5) inhibited acetylcholine-induced vasorelaxation. This effect was NO-synthase independent, but was abolished by gap junction inhibitors or normalisation of pH_i by NHE inhibition (cariporide). Furthermore, dye transfer from endothelial cells to VSMCs was inhibited by the endothelial alkalinisation caused by the removal of $\text{CO}_2/\text{HCO}_3^-$, as was an endothelium-derived hyperpolarization of vascular smooth muscle cells (Boedtkjer *et al.*, 2013).

One study on the pulmonary arteries of heartworm-free dogs demonstrated that ammonium chloride-induced, and thus intracellular alkalosis-related, *ex vivo* vasoconstriction was independent of endothelial function, achieved either by manual denudation or indomethacin. Similarly, whole-organ *ex vivo* preparations of isolated perfused lungs of male Sprague-Dawley rats showed that ammonium chloride induces an increase in pulmonary artery pressures similar to that predicted based on isometric tension studies of artery segments (Krampetz and Rhoades, 1991). While this might suggest that the effect on the vascular smooth muscle is dominant, this is with the caveat of deriving evidence of an *ex vivo* model, and with a relatively underused animal model as well as the pulmonary circulation. This is later supported by the finding that the intracellular acidification of whole endothelium-intact mouse mesenteric small arteries induced by the nominal absence of $\text{CO}_2/\text{HCO}_3^-$ diminished the response to sustained (2 hours) norepinephrine-induced contractions as compared to the presence of $\text{CO}_2/\text{HCO}_3^-$. This was despite the absence of effect of $\text{CO}_2/\text{HCO}_3^-$ -free conditions on intracellular Ca^{2+} (Boedtkjer *et al.*, 2006).

As pH_i may have an impact on the release of a range of vasoactive substances, together with alteration in myoendothelial gap junctions, it makes the interpretation of isolated systems very difficult. Additionally, it is feasible that different vascular beds may have different dominant systems. It is arguable that it is still currently unclear how pH_i influences the *overall* vascular tone.

1.7. Possible mechanisms of pH_i affecting blood pressure – renal salt- and water- handling

Although NBCn1 is known to be expressed in the medullary thick ascending loop of Henle, the overall impact on salt- and water-handling is complicated. As NBCn1 is positioned basolaterally in the medullary thick ascending loop of Henle (mTAL) (Praetorius *et al.*, 2004), at first glance it may seem logical that knockout mice would have reduced Na^+ -flux from the blood to lumen, thus Na^+ -preserving. Alternatively, better parallels might be drawn with *SLC9A1*^{-/-} (encoding NHE1) and *SLC9A3*^{-/-} (encoding NHE3) knockout mice. Just like the *SLC4A7*^{-/-} knockouts, they all lead to reduced intracellular Na^+ and HCO_3^- , but a key difference is that NHE1 is positioned basolaterally and NHE3 is apical. Despite the difference in localization, and thus the overall flux of Na^+ , both these knockout models are salt-resistant (or salt-wasting), suggesting the overall impact of intracellular acidosis might be more important than that of sodium flux. In *SLC9A3*^{-/-} knockouts, one of the key findings is that the NKCC2 co-transporter (which is the target of loop diuretics) is down-regulated (Brooks *et al.*, 2001; Amlal *et al.*, 2003), thus demonstrating a salt- and water-wasting phenotype consistent with loop diuretic treatment. In *SLC9A1*^{-/-} knockout mice, NHE3 expression was down-regulated (Good *et al.*, 2004), although not much is known of its effect on NKCC2.

In the absence of data on salt-loading of *SLC4A7*^{-/-} knockout mice, nor a renal-specific conditional knockout, the overall impact on the salt-and-water handling capabilities remains under speculation. A further caveat is that the murine nephron is not completely analogous to that of humans (Cheval *et al.*, 2012), therefore requiring caution when extrapolating these studies to the human kidney.

1.8. Possible mechanisms of pH_i affecting blood pressure – other tissues

Due to the expression profile of NBCn1, three further tissues should be considered as potential contributors to potential mechanisms affecting blood pressure – the heart, gastrointestinal tract and neural tissue.

Within the heart, NBCn1 has been shown to be expressed in the atria, and mRNA is also present in the ventricles (Boedtkjer *et al.*, 2008). Like vascular smooth muscle, the cardiac muscle appears to be similarly responsive to pH_i , even though it is striated muscle instead. Correspondingly, intracellular alkalosis stimulated an increase in the contractile tension generated by

cardiomyocytes, and intracellular acidosis with a decrease in tension (Bountra and Vaughan-Jones, 1989). Additionally, intracellular acidosis was also associated with a decline in spontaneous Ca^{2+} spark rate (Balnave and Vaughan-Jones, 2000). If an increase in cardiac contractility increases cardiac output (as cardiac output = stroke volume x heart rate), this *theoretically* may result in an increase in blood pressure (as mean arterial pressure = cardiac output x systemic vascular resistance + central venous pressure). However, the effect of changes in cardiac contractility would perhaps be nullified by other homeostatic factors.

NBCn1 is expressed on the mucosal surface in various sections of the gastrointestinal tract, where it is found basolaterally (Praetorius *et al.*, 2001; Boedtkjer *et al.*, 2008). While there is no published data on the impact of *SLC4A7*^{-/-} knockout on gastrointestinal function, it is worth bearing in mind that the *SLC9A3*^{-/-} knockout mice lacking the apical NHE3 results in a diarrhoeal state, volume depletion and subsequently lower blood pressure (Ledoussal *et al.*, 2001).

Within the central nervous system, NBCn1 has been shown to be expressed throughout the cerebrum and cerebellum (Boedtkjer *et al.*, 2008), but very little is known about which specific cell types and anatomical regions they are expressed, let alone how the expression impacts neuronal function. The limited knowledge available shows that pH_i is a factor in the frequency of action potentials of trigeminal cells *in vitro*, where intracellular acidification reduces the activity, and conversely intracellular alkalosis increases the frequency of action potentials (Hwang *et al.*, 2011). Due to the complex nature of neural regulation of blood pressure, in the absence of tissue-specific knockouts, the exact contribution of NBCn1 in each tissue type is difficult to tease out.

1.9. NBCn1 splice variants and impact on function

The *SLC4A7* gene spans over 110,000 bp and contains 28 exons, introducing the possibility of multiple splice variants. Additionally, exons 1 and 2 pose as two potentially different transcription initiation sites (designated MEAD and MERF respectively based on the initial four amino acids). In combination with the two possible start codons, there are four cassettes identified, leading to 32 theoretical combinations, the key isoforms are identified in Figure 4. Thus far, only 21 full length variants have been identified, with a further 6 truncated variants, albeit in mRNA transcripts and not fully explored in protein form. Based on experimental *in vitro* work on *Xenopus laevis* oocytes, the inclusion or exclusion of these optional structural elements have been demonstrated to have

an effect on surface NBCn1 protein abundance and intrinsic bicarbonate transport activity (Yang *et al.*, 2009). Of note, the non-synonymous SNP rs3755652 (see Section 3.6 for more details) is located within Cassette II.

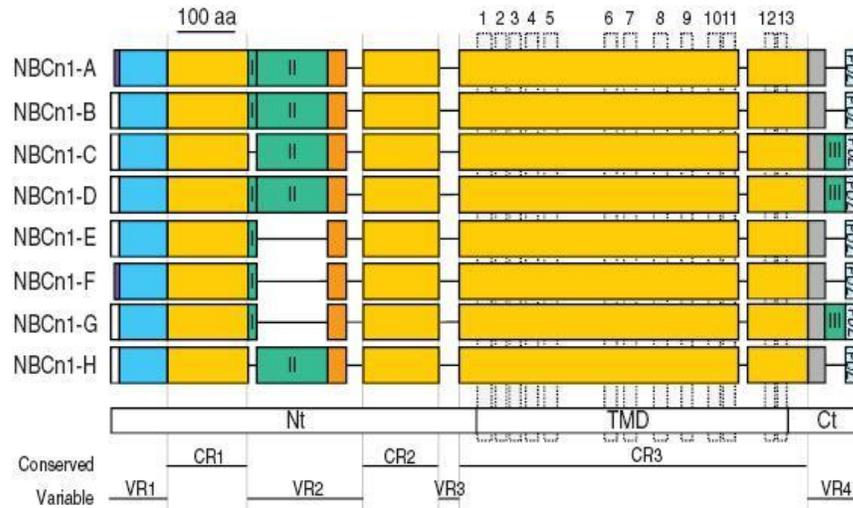


Figure 4: Representative diagram of motifs within NBCn1 of first 8 variants discovered. Representative diagram of the alternative transcription starts points, cassettes, transmembrane domains, conserved/variable regions in relation to NBCn1 isoforms (Adapted from Boron, 2009)

The *in vitro* relevance of splice variants is further emphasized by the observation that the splice Cassette II can bind calcineurin, as demonstrated in A7r5, a fibroblastic cell line derived from *Rattus norvegicus* thoracic aorta smooth muscle cells. In turn, calcineurin inhibitors were able to attenuate the expected activity of NBCn1 in murine mesenteric arteries, after either depolarization or norepinephrine-induced contractions (Danielsen *et al.*, 2013). Any finding with regards to animal models needs to be taken in context of any differences in the amino acid sequences. As indicated in Figure 5, within the Cassette II splice itself, there are 11 amino acid differences between the human and rat sequences or varying predicted significance. Furthermore, the human Cassette II is one amino acid longer than that of rats (and other murine species).

(http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) in

Figure 6, even the closely related members of the SLC4 family (anion exchangers 1, 2, and 3, electroneutral sodium-bicarbonate co-transporter and sodium-driven chloride-bicarbonate exchanger) or the functionally relevant SLC9 family of sodium-hydrogen exchangers did not have any homology with the splice cassette. In fact, there are no known proteins with significant homology with the splice cassette (Appendix 2 for full BLAST results). This limits the ability to extrapolate function from other known proteins.

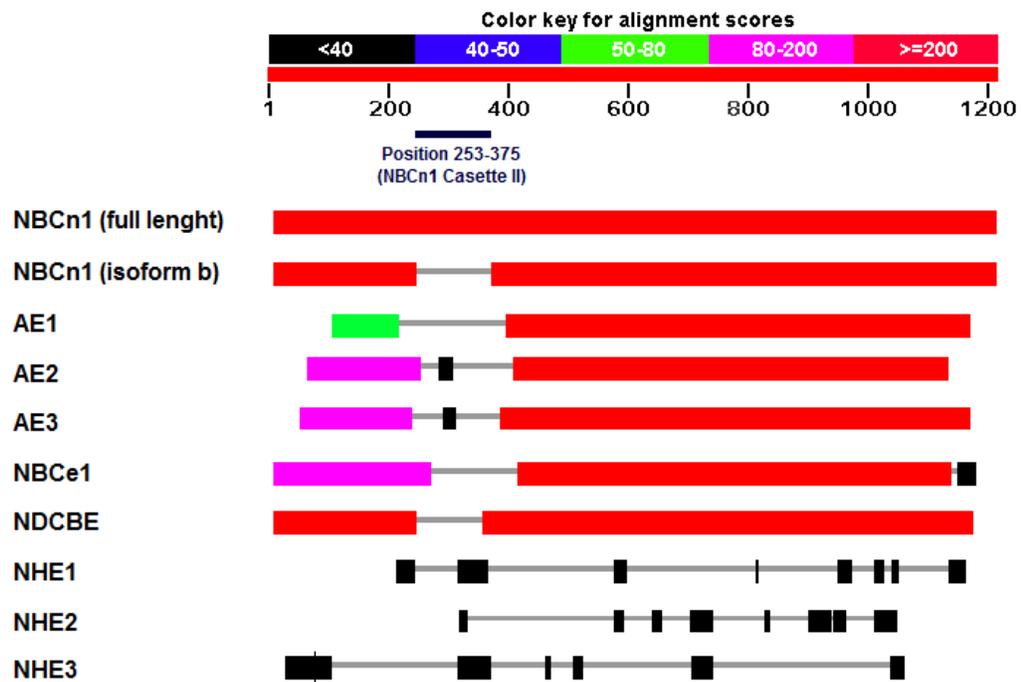


Figure 6: BLAST protein alignments of NBCn1 and other selected acid-base regulators
 BLAST (Basic Local Alignment Search Tool; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>, NCBI) protein alignments of NBCn1-B (full length), NBCn1-E (without Cassette II), other members of the SLC4 family as well as selected acid-base regulators.

Thus, if there is an allele-dependent difference in RNA splicing, it poses a potential mechanism for the polymorphism to have an overall impact on pH_i , either by total protein abundance, intrinsic bicarbonate transport activity, or its response to signaling pathways such as with calcineurin.

1.10. Post-translational modifications and examples from other NBCs

The function or activity of a protein may be affected by various post-translational modifications. Whilst there is a multitude of different potential modifications, there has already been some evidence of the impact of phosphorylation and glycosylation on the family of sodium-bicarbonate co-transporters, mainly on NBCe1 for which NBCn1 shares some homology (Boron et al., 2009, also see

Figure 6). This may be particularly relevant to the blood pressure-associated SNP, as it marks a non-synonymous polymorphism resulting in an amino acid change positioned within a series of serine or threonine residues (both being sites of phosphorylation and glycosylation). As the mechanism of adding a phosphate or glycosyl group to an amino acid is typically enzymatic with the requirement for a surrounding consensus binding sequence (see Section 3.8), *theoretically* the amino acid change can alter any propensity to these modifications.

Phosphorylation is the modification of an amino acid by the addition of a phosphate group (PO_4^{3-}). There are multiple kinases potentially responsible for phosphorylation, which typically occurs on serine, threonine, tyrosine, and less commonly, histidine residues. For phosphorylation to occur, a kinase-specific consensus sequence usually is closely matched to the amino acid sequence at the phosphorylation site (Rust and Thompson, 2011). As indicated earlier in Section 1.9, there is already the potential for calcineurin (a.k.a. protein phosphatase 3) to regulate NBCn1 function.

The phosphorylation status of the NBCe1 Ser¹⁰²⁶ residue appears to have a role in modulating its stoichiometry, shifting between 3:1 and 2:1 ($\text{HCO}_3^- : \text{Na}^+$ ratio). 8-Br-cAMP-stimulated PKA-dependent phosphorylation of that amino acid results in the change to the 2:1 stoichiometry, and this shift was removed in mutant proteins with an (non-phosphorylatable) Ala¹⁰²⁶ residue. Furthermore, Asp¹⁰²⁶ mutants (mimics the negative charge of a phosphorylated serine residue retains a 2:1 function regardless of 8-Br-cAMP stimulation. However it is worth noting that not all phosphorylation sites have an observed effect on co-transporter function, as generating the N-terminus T49A or T49D mutants in NBCe1 did not have an impact (Gross *et al.*, 2003).

A similar dependence on phosphorylation was also noted with NBCn1. Dephosphorylation by exposure to PP1 (protein phosphatase 1) or IRBIT (an inositol 1,4,5-trisphosphate receptor-binding protein which recruits PP1) resulted in an increased activity. Conversely, dephosphorylation by inhibitors of PP1 (I2 and tautomycin), a dominant-negative IRBIT mutation (IRBIT Δ PEST, reducing the effect of IRBIT recruitment of PP1) or introduction of SPAK (SPS1-related proline/alanine-rich

kinase, resulting in phosphorylation of proteins) reduced NBCn1-A activity (Hong *et al.*, 2013). Furthermore, protein kinase A (PKA) also resulted in reduced NBCn1 activity (Loiselle *et al.*, 2004). The same pattern of regulation was also found with these pharmacological challenges with NBCe1 (Shirakabe *et al.*, 2006; Yang *et al.*, 2011; Hong *et al.*, 2013). Overall, it may suggest that NBCs may be more active in the dephosphorylated state.

Further supporting evidence might be inferred from studies on NHE1 and NHE3. With these antiporters, they also appear to be regulated by phosphorylation, but in the opposite direction – where the activity is increased in the phosphorylated state (

Table 2) (Pang *et al.*, 2001; Pang *et al.*, 2002; Pang *et al.*, 2004; Misik *et al.*, 2005; Ammar *et al.*, 2006; Epting *et al.*, 2006; Esteva-Font *et al.*, 2007; Matsushita *et al.*, 2007; Zaun *et al.*, 2008; Di Sole *et al.*; Matsushita *et al.*, 2011; Liu *et al.*, 2013). Of course, care should be taken when interpreting this and extrapolating towards the potential interaction with calcineurin A. Particularly, as the research represented in

Table 2 combines outputs from only 12 groups. Additionally, any summary of results in such a manner should be taken in context of possible publication and confirmation biases.

		PKA	SPAK	I2	Tauto	CHP *	CsA	FK506	IRBIT	PP1	CnA	
Mechanism of action		Kinase		Phosphatase inhibitor		Inhibits calcineurin			Recruits PP1	Phosphatase		
Impact on phosphorylation		Increases						Decreases				
NBCn1	VSMC											
	Experimental cell lines										+ Co-IP	
NBCe1	VSMC											
	Pancreatic duct cells									+ Co-IP		
	Experimental cell lines		+ Co-IP							+ Co-IP		
NHE1	VSMC											
	Experimental cell lines					+ Co-IP					+ Co-IP	
NHE3	VSMC											
	Renal cells											
	Experimental cell lines					+ Co-IP						

*Table 2: Summary of studies on the impact of phosphorylation state on NBC and NHE activity. Studies grouped into overall experimental challenge. Each cell may represent more than one study. Orange cells – increased activity; blue cells - decreased activity; black cells - no change; empty cells – no studies so far. “+ Co-IP” denotes that in addition to change in activity, there is evidence of protein-protein interaction by co-immunoprecipitation. * denotes the presence of multiple studies across these CHP proteins (CHP1, CHP2 and CHP3). Full details and references in Appendix 3. PKA, Protein Kinase A; SPAK, Ste20-related proline/alanine-rich kinase; I2, Protein Phosphatase Inhibitor-2; Tauto, Tautomycin; CHP, calcium-binding protein; CsA, ciclosporin A; FK506, tacrolimus; IRBIT, Inositol 1,4,5-trisphosphate Receptor-binding Protein; PP1, Protein phosphatase-1; CnA, calcineurin A.*

Glycosylation is the modification of an amino acid by the addition of a glycosyl group. There are a range of mechanisms and cellular locations by which glycosylation occurs, but this is beyond the scope of this thesis. For glycosylation to occur, a consensus sequence usually is typically matched to the amino acid sequence at the glycosylation site (Chauhan *et al.*, 2013), although thought not to be as strict as for phosphorylation.

NHE3 is an example of a transporter that is regulated by glycosylation. Treating cultured renal proximal tubule cells LLC-PK₁ with tunicamycin, an inhibitor of *N*-glycosylation, reduced the presence of glycosylated NHE3 at the membrane, and for the non-glycosylated NHE3 to be retained away from the membrane. This in turn resulted in reduced NHE activity, either measured by acid-stimulated ²²Na⁺ uptake, or sodium-dependent pH_i recovery from acid load (Soleimani *et al.*, 1997). NCBE (the sodium-coupled bicarbonate exchanger encoded by *SLC4A10*, and also known as NBCn2) was found to be glycosylated, as the apparent molecular weight decreased following the removal of *N*-glycosylation modifications by *N*-glycosidase F treatment. To assess the physiological impact of these glycosylation modifications, a triple mutant generated by converting the *N*-glycosylation target asparagine to the non-glycosylatable glutamine (N677Q/N687Q/N697Q) residues was generated, resulting in significantly reduced protein expression in *Xenopus laevis* oocytes. Accordingly, the triple mutant did not demonstrate any measurable co-transporter function (Chen *et al.*, 2008). On the other hand, despite finding NBCe1 to be glycosylated, the non-glycosylated triple mutant (N592Q/N597Q/N617Q) did not show any changes in co-transporter activity in *Xenopus laevis* oocytes (Choi *et al.*, 2003). However, none of these protein variants occur in a homologous sequence or domain/regions as the E326K variants in NBCn1.

1.11. Protein trafficking and examples from other pH_i regulators

To fulfill the role as an pH_i regulator by bicarbonate transport, NBCn1 should be, at least classically, positioned on the plasma membrane. However, it has already been observed that NBCn1 is predominantly located on the membrane, although with some intracellular signal from confocal microscopy (Lauritzen *et al.*, 2010), with the function of the intracellular localization being currently yet unknown. Interestingly, changes in trafficking and/or protein activity have been noted with non-synonymous polymorphisms or mutations in *SLC4A1* (Quilty *et al.*, 2001), *SLC4A4* (Toye *et al.*, 2006) and *SLC4A11* (sodium bicarbonate transporter-like protein 11)

(Loganathan and Casey, 2014). However, these variants are typically in the transmembrane domains, and not in the analogous N-terminal region like the NBCn1 E326K variation.

As already explored in Section 1.9, the E326K amino acid change occurs in a region where protein-protein interaction can occur, with potential conformational changes. This may suggest a potential allele-dependent difference in protein trafficking. With changes in cellular localization, it may be a potential mechanism for the polymorphism to have an overall impact on pH_i .

1.12. Inducibility of NBCn1 (and other acid-base co-transporters)

NBCn1, and other regulators of pH_i , is feasibly spatially and temporally dynamic. There is a body of evidence indicating that the expression profiles of NBCn1 and other pH_i regulating co-transporters change in response to various challenges. It appears that in general, intracellular acidosis induces expression of these co-transporters (

Table 3), be it generated by ammonium chloride (Laghmani *et al.*, 1997; Kim *et al.*, 1999; Kwon *et al.*, 2002; Amlal *et al.*, 2004; Praetorius *et al.*, 2004; Jakobson *et al.*, 2004; Nowik *et al.*, 2008; Lee *et al.*, 2010), hydrogen chloride (Karim *et al.*, 2002, Kim *et al.*, 2003; Park *et al.*, 2010) or chronic hypercapnia (Kanaan *et al.*, 2007; Orłowski *et al.*, 2013). Conversely, intracellular alkalosis appears to downregulate expression (Kim *et al.*, 1999; Amlal *et al.*, 2001; Praetorius *et al.*, 2004). As pH_i regulators, this negative feedback is not unexpected.

		NH4Cl- or HCl- induced intracellular acidosis											Chronic hypercapnia-induced intracellular acidosis		Intracellular alkalosis			
		R	R	R*	R	R	R	R	R	M	X*	R	M	X*	R	R	R	
NBCn1	VSMC																	
	Cardiac																	
	Renal cells																	
	Neurones																	
	GI tract																	
NBCe1	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
NHE1	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
NHE3	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AE1	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AE2	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
Na/K-ATPase	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
NKCC2	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
NCC	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
ENaC	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AQP1	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AQP2	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AQP3	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
AQP4	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
H ⁺ -ATPase	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	
Na-PI2	Experimental cell lines																	
	VSMC																	
	Cardiac																	
	Renal cells																	
	GI tract																	

Table 3: Summary of studies on the impact of altering pH on expression of acid-, base-, sodium-transporters of channels

Studies in vascular, endothelial, renal and neuronal cells are listed in chronological order from left-to-right after grouping into overall experimental challenge. Each column represents one experimental condition. Experimental model abbreviation M – mice, R – rat, X – experimental cell line, * - in vitro study. Red cells - acidosis increased expression / alkalosis decreased expression; black cells – no change; empty cells – not assessed in study. There were no studies indicating that acidosis reduced expression or alkalosis increased expression. Full details and references in Appendix 4.

Notes: Summaries such as this table are subject to publication bias.

Additionally, exposure to activators of the renin-angiotensin-aldosterone axis (RAAS) appears to induce the expression or increase the activity of NBCs and NHEs (

Table 4). This includes induction by angiotensin II (Matsui *et al.*, 1995; Ye *et al.*, 1996; Sandmann *et al.*, 2001; Baertz *et al.*, 2002; Kwon *et al.*, 2003; Epting *et al.*, 2006; Li *et al.*, 2008; De Giusti *et al.*, 2009) and aldosterone / corticosteroids (Ebata *et al.*, 1999; Alzamora *et al.*, 2000; Muto *et al.*, 2000; Gekle *et al.*, 2001; Nielsen *et al.*, 2002; Karmanzyn *et al.*, 2003; Michea *et al.*, 2005; Miyata *et*

al., 2005; Musch *et al.*, 2008), as well as reversal of the induction by their respective antagonists (Alzamora *et al.*, 2000; Sandmann *et al.*, 2001; Nielsen *et al.*, 2002; Baetz *et al.*, 2002; Michea *et al.*, 2005; Miyata *et al.*, 2005; Epting *et al.*, 2006; Yamamoto *et al.*, 2007; Li *et al.*, 2008; Musch *et al.*, 2008; De Giusti *et al.*, 2009). The one study that did not follow this pattern was by Muto *et al.* (2000), who demonstrated that short-term dexamethasone exposure (3 hours) with cultured rat aortic VSMCs increased pH_i; recovery rates, but longer-term exposure (24 hours) decreased mRNA levels.

		Renin-Angiotensin-Aldosterone Axis Stimulation																				
		Rb*	R*	R*	H*	R	R	R	R	X*	R	R	R*	R	R*	R*	X*	R	R	X*	C*	
NBCn1	VSMC																					
	Cardiac																					
	Renal cells																					
	Neurones																					
	GI tract																					
	Experimental cell lines																					
NBCe1	VSMC																					
	Cardiac																					
	Renal cells																					
	GI tract																					
	Experimental cell lines																					
	NHE1	VSMC																				
Cardiac																						
Renal cells																						
Neurones																						
GI tract																						
Experimental cell lines																						
NHE3	VSMC																					
	Renal cells																					
	GI tract																					
	Experimental cell lines																					
AE1	Renal cells																					
AE2	Renal cells																					
Na/K-ATPase	Renal cells																					
NKCC2	Renal cells																					
NCC	Renal cells																					
ENaC	Renal cells																					
AQP1	Renal cells																					
AQP2	Renal cells																					
AQP3	Renal cells																					
AQP4	Renal cells																					
H ⁺ -ATPase	Renal cells																					
Na-PI2	Renal cells																					

Table 4: Summary of studies on the impact of RAAS activation on expression of acid-, base-, sodium- transporters or channels

*Studies in vascular, endothelial, renal and neuronal cells are listed in chronological order from left-to-right. Each column represents one experimental condition. Experimental model abbreviation C – cat, H – human, M – mice, R – rat, X – experimental cell line, * - in vitro study. Red cells – increased expression by RAAS activation, blue cells – decreased expression, orange cells – increased transport function; black cells – no change, empty cells – not assessed in study. Full details and references in Appendix 5.*

Notes: (1) Earlier studies (predating cloning of NBCn1) describe any Na⁺- dependent pH_i recovery as NHE. (2) Studies on the acid or base flux cannot differentiate between the NHE isoforms as well as between the NBC isoforms. (3) Summaries such as this table are subject to publication bias.

There is less information and no clear pattern observed when NBCs or NHEs are knocked-out/-down when compared to changes in pH or challenges via the renin-angiotensin-aldosterone axis, (Brooks *et al.*, 2001; Amlal *et al.*, 2003; Lauritzen *et al.*, 2010; Boedtkjer *et al.*, 2011; Groger *et al.*, 2011; Chen *et al.*, 2012) (

Table 5). However, it is notable that in the very few studies on changes in gene expression after disruption of *SLC4A7* expression, no impact was observed for NBCe1, NHE1 and NHE3 expression (Lauritzen *et al.*, 2010; Chen *et al.*, 2012). This should be taken with the usual caveat that with detection by protein immunoblots, these studies may not have sufficient sensitivity to identify subtle but yet physiologically significant changes. Additionally, there may be cell-specific differences in the response to activation of the renin-angiotensin-aldosterone system.

		SLC4A7 (NBCn1)			SLC9A1 (NHE1)	SLC9A3 (NHE3)		SLC4A5 (NBCe2)
NBCn1	VSMC		KO	KO				
	Cardiac		KO	KO				
	Renal cells		KO	KO				
	Neurons		KO	KO				
	GI tract		KO	KO				
	Experimental cell lines	siRNA	KO	KO				
NBCe1	VSMC							
	Cardiac							
	Renal cells							
	GI tract							
	Experimental cell lines							
NHE1	VSMC							
	Cardiac							
	Renal cells							
	Neurons							
	GI tract							
	Experimental cell lines				siRNA			
NHE3	VSMC					KO	KO	
	Renal cells					KO	KO	
	GI tract					KO	KO	
	Experimental cell lines					KO	KO	
AE1	Renal cells							
AE2	Renal cells							
Na/K-ATPase	Renal cells							
NKCC2	Renal cells							
NCC	Renal cells							
ENaC	Renal cells							
AQP1	Renal cells							
AQP2	Renal cells							
AQP3	Renal cells							
AQP4	Renal cells							
H ⁺ -ATPase	Renal cells							
Na-Pi2	Renal cells							

Table 5: Summary of studies on the impact of NBC / NHE knockouts or knockdown on expression of acid-, base-, sodium- transporters or channels
Studies in vascular, endothelial, renal and neuronal cells are listed in chronological order from left-to-right. Each column represents one experimental condition. Red cells – increased expression by the knockout, blue cells – decreased expression, black cells – no change, empty cells – not assessed in study. Full details and references in Appendix 6.

It is conceivable that the impact of functional SNPs at the *SLC4A7* locus acts by altering the inducibility of gene expression over and above the basal level, such as changing the propensity of transcription factor to bind the altered consensus sequence.

A note of caution should be stated when comparing studies across different animal species, particularly when involving a range of different acid-base regulators. This is because the relative contributions of NBC and NHE in regulating cardiomyocyte pH_i have been shown to vary with species assessed (Yamamoto *et al.*, 2005). Additionally, results from HEK293 cell lines should always be interpreted with caution. Whilst they were initially derived from the embryonic human kidney, they have expression profiles similar to neuronal tissue (Shaw *et al.*, 2002). Additionally,

most of the cells have been shown to have a hypotriploid karyotype, with most cells having 64 chromosomes, but some with more (European Collection of Cell Cultures, n.d.). It is unlikely that HEK293 cells have cellular function similar to human cells, let alone that of human kidneys.

1.13. Observational studies of pH_i in hypertension

There is already previous research into pH_i and hypertension. In the experimental murine model of hypertension, the spontaneously hypertensive rat (SHR), resting pH_i in resistance arterioles was more alkaline as compared to Wistar-Kyoto wildtypes (Izzard and Heagerty, 1989a). There is evidence that this difference is amiloride-dependent, thus suggestive of increased NHE activity in SHR. However, the contribution of the DIDS-dependent component, reflecting anion exchanger activity was not different between SHR and WKY wildtypes (Izzard and Heagerty, 1989b). It is important to recognize that in the late 1980s, the function of NBCs would not have been investigated. However, this was further examined to show that SHR did indeed have increased pH_i recovery in the absence of HCO_3^- , better supporting the finding of increased NHE activity in SHR (Foster *et al.*, 1992). On the other hand, this still predates the cloning of NBCs, and thus no direct investigation on whether NBCs contribute to the differences in baseline pH_i between the hypertensive and normotensive animal models are known. With such observational data, it is difficult to establish causality. It is possible that the increased NHE activity predisposes to a hypertensive phenotype, or that hypertension induces secondary changes which includes upregulation of acid extruders. Alternatively, this data also does not exclude a spiral where increased NHE expression predisposes to hypertension which in turn upregulates NHE.

However, the converse seems to apply to circulating blood cells. The lymphocytes of SHR were more acidotic than that of wildtype controls (Batlle *et al.*, 1990). This was consistent with observations that untreated hypertensive patients had more acidotic pH_i erythrocytes as compared to treated hypertensive patients and normotensive subjects (Resnick *et al.*, 1987). Interestingly, there were no observed differences with platelets (Sagnella *et al.*, 1999). The differences between the findings in resistance arterioles and circulating blood cells may however indicate tissue specificity in the distribution and role of pH regulating proteins.

Thus far, the large majority of supportive evidence presented is dependent on animal models, and particularly the knockout mouse model. It is worth emphasising that the mouse NBCn1 protein has up to 80 different amino acids within its sequence, and is also 1 amino acid shorter (Figure 7). These differences are typically conserved across murine species, and it is currently unknown whether any of these differences have any functional impact.

```

1 MERFRLEEKLEGPDEEAVVDLGTKSSSTVNTKFEKEEESHRAVIGVHVFPFSKESRRRHR 60
61 HRGKHHHRRRDKKSDKEDGRESPSYDTPSQRVQFILGTEDDDEEHIPHDLFTMDEL 120
121 YRDGEYEWKETARWLKFEEDVEDGGDRWSKPYVATLSLHSLFELRSCILNGTVMLDMRA 180
181 STLDEIADMVLDNMIASGQLDSIRENVREALKRHHQNEKRFTSRIPLVRSPADIGKK 240
241 HSDPHLLERNGEGLSASRHSRLTGLSASNLSLRGESPLSLLLHLLPSRAGTPAGSRCT 300
301 TPVPTPQNSPPSSPSTISRLSRSQESQQAPELVSPASDDIPVVIHPPEEDTEALK 360
361 GEEQKNEENVDTTPGILASPQSAPGNLDNSKSGEIKNGSGGSRENSTVDFSKVDMNFM 420
421 KIPTGAEASNVLVGEVDFLERPIIAFVRLAPAVLLGLTEVPVPTRFLLFGPAGKAPQ 480
481 YHEIGRSIATLMTDEIFHDVAYKAKDRNDLLSGIDEFLDQVTVLPPGEWDPSIRIEPK 540
541 VPSQEKRKIPVFNHNSPTLGETPKEAAHHAGPELQRTGRLEGGILLDIKRKAPFFLSDF 600
601 KDALSQCLASILFLYCACMSPVITFGGLGEATEGRI SAIESLFGASLTGIAYS LFAGQ 660
661 PLTILGSTGVPVLVFEKILKFCRDYLSYLSLRISGLWTSFLCIVLVATDASSLVCYIT 720
721 RFTEEAFAALICIIFIYEALEKLEFLGELYAFNMHNLDKLTYSVCVTEPPNPSNETLA 780
781 QWKKDINITAHNLSWENLTVSECKRLRGVFLGSACGHGPHYDPDVLFWCVLLEFTTFEFLSS 840
841 FLKQFKTKRYFPTKVRSTISDFAVELTIVIMVVIDYLVGVESPKLHVPEKFEPTPEPRGW 900
901 IISPLGDNPWWTLLIAATPALLCTILIFMQQITAVIINRKEHKLKKGAGYHLDLLMVG 960
961 MLGVCS V MGLPWFVAATVLSISHVNSLKVSECSAPGEQPKFLGIREQRTVGLMIFILMG 1020
1021 LSVFMTSVLKFIPMPVLYGVFLYMGVSSLKGIQ LP DRIKLF FGMPAKHQ PDLI YLRYP PLW 1080
1081 KVH IFTV QLTCLVLLWVIV SAAAVVF PMMVLAL VFVRK LMDLC FTKRE LSWLD DLMPE 1040
1141 S K K K E D D K K K K E E A E R M L Q D D T V H L P F E G S L L Q I P V K L K Y S P K P V S V I S F 1200
1201 DE F K K Y DA E T S L

```

Figure 7: Comparing human and mouse NBCn1 amino acid sequences. Comparison made between the human full-length MERF-start NBCn1 protein (NCBI Reference Sequence: NP_003606.3) and the predicted mouse full length MERF-start NBCn1 protein (NCBI Reference Sequence: XP_011243051). Amino acids in red represent differences between the two species. Amino acid in grey is absent from the mouse NBCn1. Sequences bold and underlined indicate 11 regions predicted by NCBI as transmembrane regions. NOTE: This needs to be taken in context - that the absence of known crystallised structure, NBCn1 is widely thought to contain 10 to 14 transmembrane regions, and that both N- and C-termini are intracellular, indicating an even number of transmembrane regions (Boron et al., 2009). The 326 amino acid in this sequence is represented as 326E.

Alongside this, it is essential to consider the human data, all *in vivo* and *in vitro*, experimental and observational, whilst recognising the limitations of these data. With regards to observational data, resistance arteries from subcutaneous fat biopsies from fourteen hypertensive subjects and fourteen normotensive controls showed no difference in baseline pH_i. However, the biopsies of normotensive controls showed the expected intracellular acidification when treated with noradrenaline, but this is significantly attenuated in hypertensive patients' biopsies. Furthermore, the case-control discrepancy is not present with EIPA-induced intracellular acidosis (by inhibiting

NHEs) and DIDS-induced intracellular acidosis (by inhibiting the $\text{HCO}_3^-/\text{Cl}^-$ exchanger) (Izzard *et al.*, 1991). This leaves room for the discrepancy to be accounted for by a sodium-bicarbonate transporter. However, this study was also conducted prior to in-depth characterisation of NBCs, thus their contribution currently cannot be deduced.

1.14. Limited human *in vivo* experience with modulation of pH_i

There has also been some clinical experience with inhibitors of pH_i regulators, in the guise of the NHE1 inhibitor cariporide as a postulated therapy in ischaemic heart disease and high-risk coronary artery bypass grafting. This included two large randomised controlled trials – GUARDIAN, n=3000 (Boyce *et al.*, 2004) and EXPEDITION, n=6000 (Mentzer *et al.*, 2008). Both the clinical trials reported reduced cardiac events, but the larger EXPEDITION trial reported an increase in mortality, predominantly due to cerebrovascular accidents (strokes). Neither trials report data on blood pressure changes, except for a statement that “changes observed in blood pressure, pulse, ECG parameters, and laboratory evaluations that included hematology, electrolytes, and metabolism were not considered clinically relevant”. While this might suggest that pharmacological modulation of pH_i may not have a strong impact on managing hypertension, it should be recalled that NHEs and NBCs may have different expression profiles, different inducibility, and that the finding of Izzard *et al.* (1991), that altered pH_i regulation in hypertensive patients is not explained by NHEs, leaving a possibility of NBCs being a dominant factor. Furthermore, these studies were conducted in the acute phase of ischaemic events, where the regulation of blood pressure would be managed by the clinical teams, potentially removing any signal that might have been present.

In a smaller, but better controlled context, amiloride was infused into forearms of 28 healthy volunteers undergoing venous occlusion plethysmography (Pickkers *et al.*, 1999). Amiloride, in addition to its well-known effect of inhibiting ENaC (epithelial sodium channel), is also reported to inhibit NHE (Besterman *et al.*, 1985). In this setting, amiloride reduced the angiotensin II-induced vasoconstriction, and was itself vasodilatory in high concentrations (Pickkers *et al.*, 1999), suggesting pharmacological modulation of pH_i impacts on human vascular function *in vivo*. Whilst this effect was initially attributed to the effect on NHE1, it is now complicated by the identification of ENaC in vascular smooth muscle (Drummond *et al.*, 2004) and endothelial cells (Kusche-Vihrog *et al.*, 2008).

1.15. NBCn1 and intracellular sodium (Na^+_i) – an alternative mechanism?

Whilst there is biological plausibility for genetic polymorphism in *SLC4A7* exerting an influence on blood pressure regulation via pH_i , the impact of the concomitant sodium flux should also be considered.

Some observational evidence can be seen in both animal models and humans. Red blood cells and cultured aortic vascular smooth muscle cells from SHR were found to have higher intracellular Na^+ as compared to wildtype (Losse *et al.*, 1984). Similar findings were identified in the red blood cells of untreated hypertensive patients having higher intracellular sodium concentrations as compared to healthy adults (Cooper *et al.*, 1989), but with the caveat that all 217 participants were black Americans, limiting the generalisability. Interestingly, intracellular Na^+ of red blood cells demonstrated a negative correlation with the response to calcium-channel blockers in 17 African Americans (M’Nuyamba-Kabangu *et al.*, 1988).

With this observation that hypertensive patients have higher intracellular sodium levels, it is interesting to note that potent vasoconstrictors do not only result in intracellular acidification as explored in Section 1.3, but an elevated $[\text{Na}^+]_i$ in vascular smooth muscle cells after exposure to angiotensin II (Johnson *et al.*, 1991), vasopressin and endothelin-1 (Neylon *et al.*, 1994). This makes the interpretation of the observational evidence more complicated as it opens up the possibility of the association between intracellular sodium and blood pressure as due to confounding factors, rather than a causative relationship.

Unlike the impact of pH_i , the impact of intracellular sodium on vascular reactivity is more difficult to tease out experimentally. Unlike intracellular acidosis and alkalosis caused by weak acids and alkali which can diffuse through the plasma membrane, generating changes in intracellular sodium concentrations often requires manipulation of a co-transporter, thus a change in concentrations of another ion as well.

Ouabain is a Na^+/K^+ -ATPase inhibitor with secondary effects on reducing $\text{Na}^+/\text{Ca}^{2+}$ exchanger activity. When used to challenge rat mesenteric arteries, it increased intracellular sodium from 11 mM to 65 mM without a change in the tone of rat mesenteric arteries, but did potentiate noradrenaline-induced vasoconstriction. It is however noteworthy there was an associated

depolarization (Mulvany *et al.*, 1984). As the potentiation was not reversed by the calcium channel blockers felodipine or D600, it was proposed that the effect was mediated via depolarisation of the cell membrane, but the contribution via intracellular sodium cannot be fully excluded. There is also some *in vivo* evidence with ouabain. Local infusion induced an elevation of vascular resistance utilising venous occlusion forearm plethysmograph in 15 normotensive subjects. Unlike the study with *ex vivo* rat mesenteric arteries, this effect was completely abolished when combined with nifedipine (Robinson *et al.*, 1983). Once again, it is difficult to isolate the contribution of intracellular sodium, from all the other changes in intracellular ions, pH_i and membrane potential.

Another method of experimentally increasing intracellular sodium is by introducing a sodium ionophore (such as callipeptin A or gramicidin), which allows the influx of Na^+ ions due to the electrochemical gradient. Callipeltin A, while increasing $[\text{Na}^+]_i$, provokes vasoconstriction in guinea pig aortic rings. This effect was inhibited by amiloride, but not the calcium channel blocker, verapamil (Trevesi *et al.*, 2004). A different sodium ionophore gramicidin, was also demonstrated to induce hyperplasia in SHR vascular smooth muscle cells (Osanai *et al.*, 1996). This provides an association with intracellular sodium and a hypertensive phenotype for vascular smooth muscle cells, but like experiments with ouabain, it is still difficult to separate the effect of intracellular sodium and the change in membrane potential.

Whilst pH_i and $[\text{Na}^+]_i$ appears inextricably linked, any increased activity of NBCn1 is likely to result in both intracellular alkalosis and increased intracellular sodium – both of which has evidence towards increased peripheral vascular resistance. Although it may be near impossible to define which (or both) is the primary cause, this does not affect the assessment of *SLC4A7* as a candidate gene in hypertension.

1.16. NBCn1 and intracellular lead (Pb^{2+}_i) – an alternative mechanism?

This thesis will concentrate on the potential impact of the blood pressure associated SNPs on NBCn1, and mainly via pH_i . However, it should be borne in mind that twin studies have identified that the genetic region around *SLC4A7* is associated with intracellular lead concentrations of erythrocytes (Whitfield *et al.*, 2007; Whitfield *et al.*, 2010). A major caveat is that the peak genetic region (within the 1-LOD score confidence interval of this peak) spans 15 Mb, and contains 62 genes or putative genes. Even the authors recognize that although *SLC4A7* is in their opinion the

most biologically plausible candidate, other genes within the region may be responsible through other mechanisms.

The role of Pb^{2+}_i with relation to *SLC4A7* and blood pressure may have some support through a longitudinal cohort study identified that blood Pb^{2+} was a significant predictor of blood pressure independent of confounders. Additionally, over the course of 11 years where the exposure to environmental lead in Copenhagen was markedly reduced, the individual's reduction in blood Pb^{2+} levels were also predictive of a reduction in their blood pressure (Moller and Kristensen, 1992). However, in a cross-sectional study in Birmingham with industrial lead exposure of the early 1980s, this effect was much weaker (Maheswaren *et al.*, 1993). The American NHANES (National Health and Nutrition Examination Survey) study including 16,222 participants demonstrated a positive correlation between blood Pb^{2+} levels and blood pressure, although the finding was more apparent in Black Americans as compared with White and Mexican-Americans (Scinicariello *et al.*, 2011). These observational studies are backed by experimental evidence where Pb^{2+} reduced the endothelium-dependent acetylcholine relaxation of aortic rings *ex vivo* (Zhang *et al.*, 2007) and rat tail arteries *in vivo* (Silveria *et al.*, 2010). Whilst this *could* be of relevance, it is essential to recognize that the genetic association is with *intracellular* lead concentrations, whereas the observational and experimental data associated with blood pressure is with *extracellular* lead levels.

Although feasible, there is thus far no evidence to support suggestions that *intracellular* lead levels impacting on blood pressure regulation or *direct experimental* evidence linking NBCn1 and Pb^{2+}_i . With this, there is still more supportive evidence of *SLC4A7* acting via pH_i or $[Na^+]_i$, as opposed to Pb^{2+}_i , to result in altered risks of hypertension. It is important to note that the cell culture media used for this thesis is in the nominal absence of Pb^{2+} (<http://www.sigmaaldrich.com/life-science/cell-culture/learning-center/media-formulations/dme.html> and <http://www.sigmaaldrich.com/life-science/cell-culture/learning-center/media-formulations/medium-199.html>), suggesting that any observed differences is unlikely to be Pb^{2+} -dependent.

1.17. Pleiotropy of *SLC4A7* – association with cancer and addiction

Interestingly, *SLC4A7* also demonstrates pleiotropy with an association with breast cancer susceptibility (Ahmed *et al.*, 2009; Chen *et al.*, 2012). Notably, the SNP associated with cancer susceptibility (rs4973768) is placed in the 3'-UTR of the gene, and nearly 140kb away from the

blood pressure-associated SNP, and is only in weak LD ($r^2 < 0.2$) (Figure 8). *SLC4A7* demonstrates strong biological plausibility as a candidate gene as intracellular alkalosis has already been associated with cancer cell growth and survival *in vitro* (Pedersen *et al.*, 2006). More recently, rs4973768 has been associated with volume density of breast cancer in African-American women (Keller *et al.*, 2015), although the authors accept the need for further validation, particularly with the apparent ethnicity-specificity of the findings.

This link is strengthened by the finding that S0859, inhibiting NBCs non-selectively, reduces the *in vitro* growth of the HCT116 (human colonic carcinoma) cell line (Hulikova *et al.*, 2011), and that NBCn1 expression is up-regulated in biopsies of primary human breast carcinomas (Lee *et al.*, 2014) and somatic mutations in *SLC4A7* have been identified to up to 5% of breast, pancreatic, lung, colonic and endometrial cancers (Gorbatenko *et al.*, 2014). On the other hand, the non-selective NBC inhibitor S0859 did not have a clear impact on MCF-7 (human breast cancer) cell line *in vitro* survival (Lauritzen *et al.*, 2010). To date, there are no published studies on the potential molecular mechanism of rs4973768 leading to changes in NBCn1 expression or function.

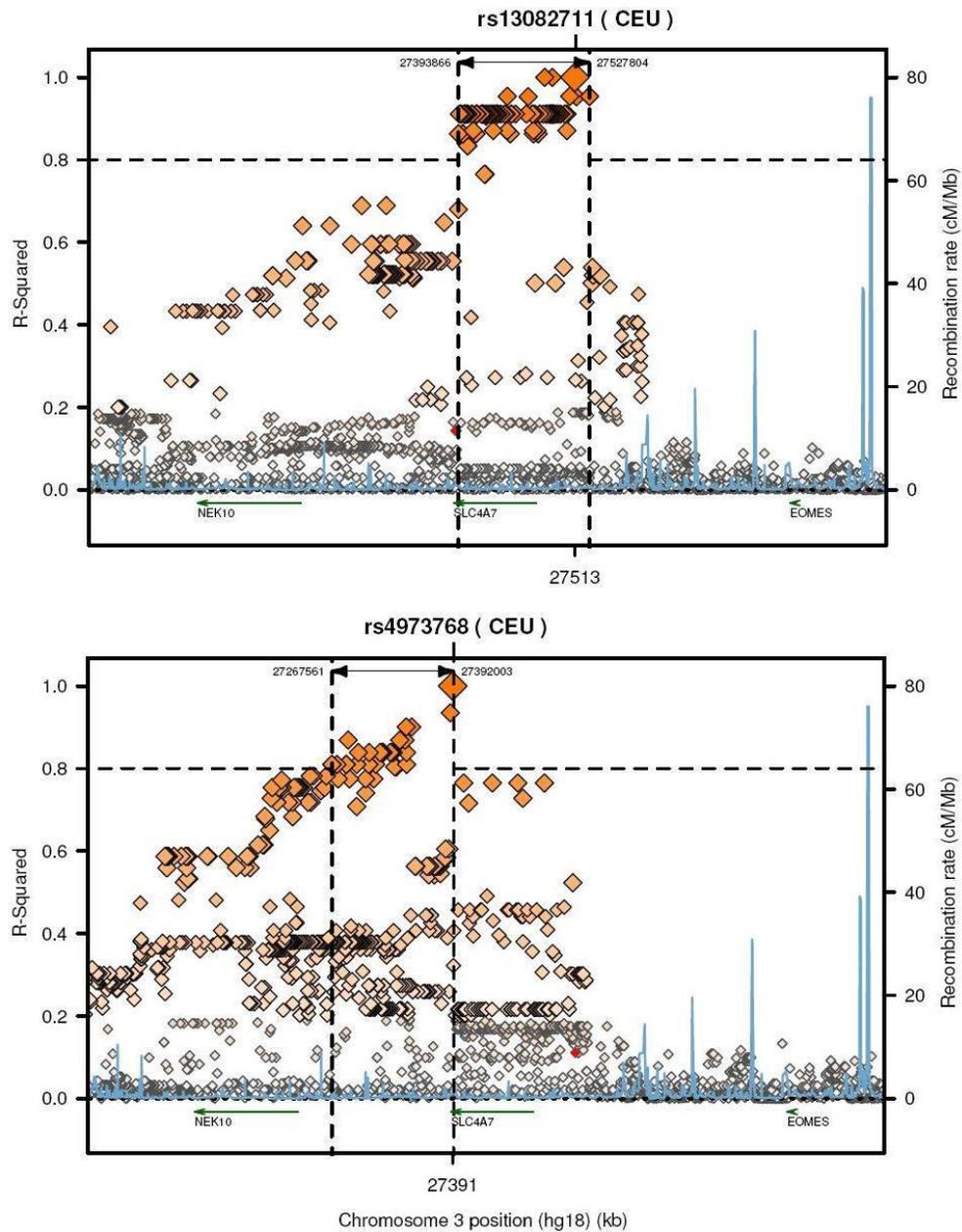


Figure 8: Linkage disequilibrium plots at rs13082711 and rs497376 SNPs
 Linkage disequilibrium plots at the blood pressure-associated (*rs13082711*) and breast cancer susceptibility-associated (*rs4973768*) SNPs based on HapMap3 (release 2) data for the CEU population (<http://www.broadinstitute.org/mpg/snp/ldplot.php>). Position along chromosome 3 is matched on both plots. An LD of $r^2 > 0.8$ (dashed lines) is marked as strong linkage disequilibrium within the loci.

Whilst there is no strong positional link between the two different GWAS loci to suggest that they share a common molecular genetic mechanism, there is still strong biological support that both affect its disease susceptibility via NBCn1 and pH_i regulation. Additionally, it appears that increased NBCn1 expression or activity, and the resulting protection from intracellular acidosis are

common to both diseases. A lesson worth considering from the field of cancer is that a gene-by-environment (GxE) interaction analysis identified the rs4973768 as having a suggestive (non-genome-wide significant) interaction with ever-smoker status (Barrdahl *et al.*, 2014). Whilst the authors linked this with the potential interplay with Pb²⁺, it is important to recognize that it does not reach a genome-wide level of significance, and the caveats expressed above in Section 1.16 still apply. Thus far, the blood pressure-associated rs13082711 is not known to GxE interactions with regards to smoking status (Sung *et al.*, 2015) but larger scale studies are in progress.

With these developments, there may still be much to learn from evidence derived from rs4973768 with regards to hypertension.

Other than an association with malignancies, a case-control study of 1158 individuals from three populations indicated that rs3278, an intronic SNP within *SLC4A7*, was associated with criteria for dependence on at least one substance of abuse (Ishiguro *et al.*, 2007). Although rs3278 and rs13082711 having genomic loci that spans a similar region, none of the SNPs are in LD with the other loci ($r^2 < 0.2$) (Figure 9). Unlike the impact of pH_i in malignancies, little is known of its impact on addictive behaviors.

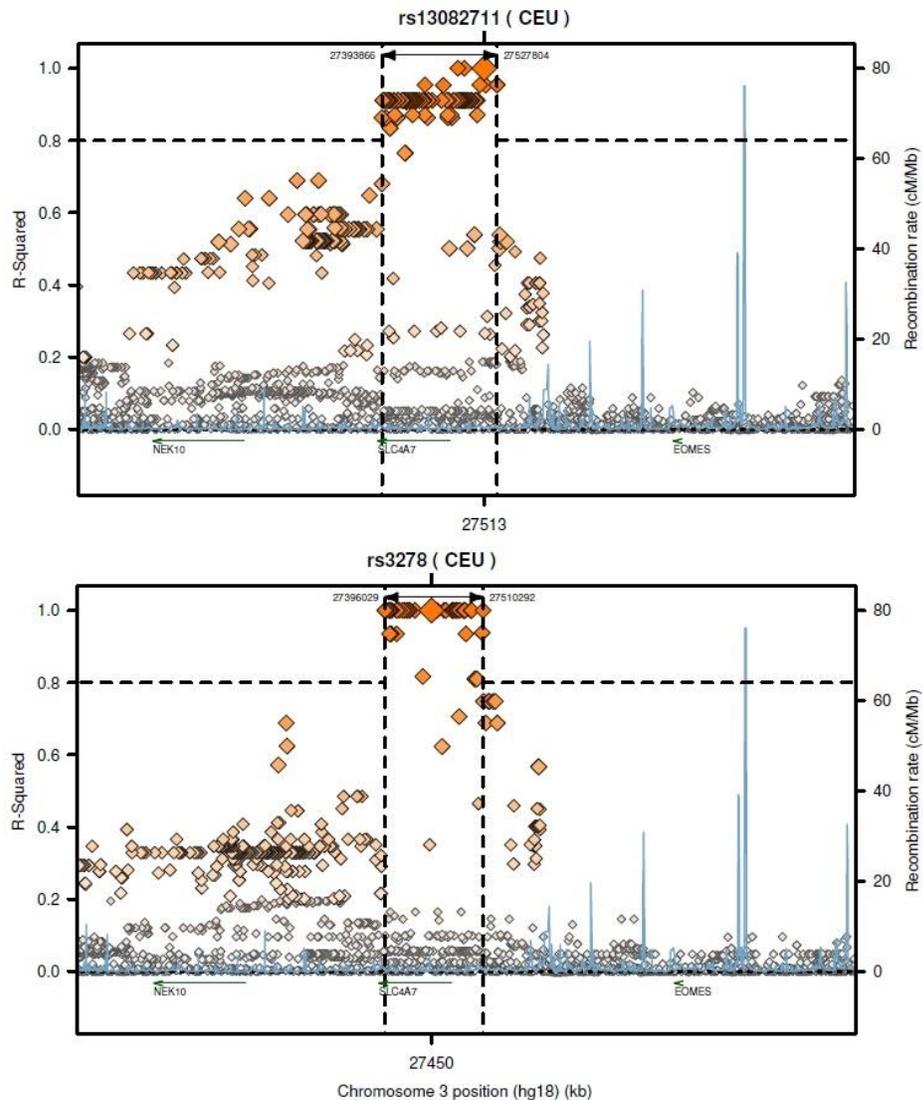


Figure 9: Linkage disequilibrium plots at the rs13082711 and rs3278 SNPs
 Linkage disequilibrium plots at the blood pressure-associated (rs13082711) and addiction susceptibility-associated (rs3278) SNPs based on HapMap3 (release 2) data for the CEU population (<http://www.broadinstitute.org/mpg/snap/ldplot.php>). Position along chromosome 3 is matched on both plots. An LD of $r^2 > 0.8$ (dashed lines) is marked as strong linkage disequilibrium within the loci.

1.18. Other BP-associated SNPs in the region - Chinese GWAS identifying rs820430 as independent risk SNP

In addition to rs13082711, a newer GWAS investigating blood pressure genetics in a Chinese population (11,816 in discovery meta-analysis, 69,146 in replication cohort) identified rs820430 as a new and independent SNP associated with blood pressure, where this time, only systolic blood pressure reached genome-wide significance (Lu *et al.*, 2014). These two SNPs are positioned just

under 11kb from each other, both 5' of *SLC4A7* (Figure 10). It is notable that the prevalence of the rs13082711 risk allele was >20% in the ICBP (Ehret *et al.*, 2011) cohort, but only 2% in the HapMap-CHB, perhaps accounting for the failed detection in the Lu cohort as a risk SNP. Importantly, the two SNPs, rs13082711 and rs820430, are not in LD ($r^2 = 0.06$ in HapMap-JPY+CHB, and 0.19 in HapMap-CEU populations). Although these are two independent signals, whether both SNPs (additively or synergistically), one or actually none impacts on *SLC4A7* expression or NBCn1 function is not possible to be determined based purely on the GWAS. The lack of clues of how rs820430, or the SNPs in high LD, has a molecular genetic effect is explored further in Section 3.15.

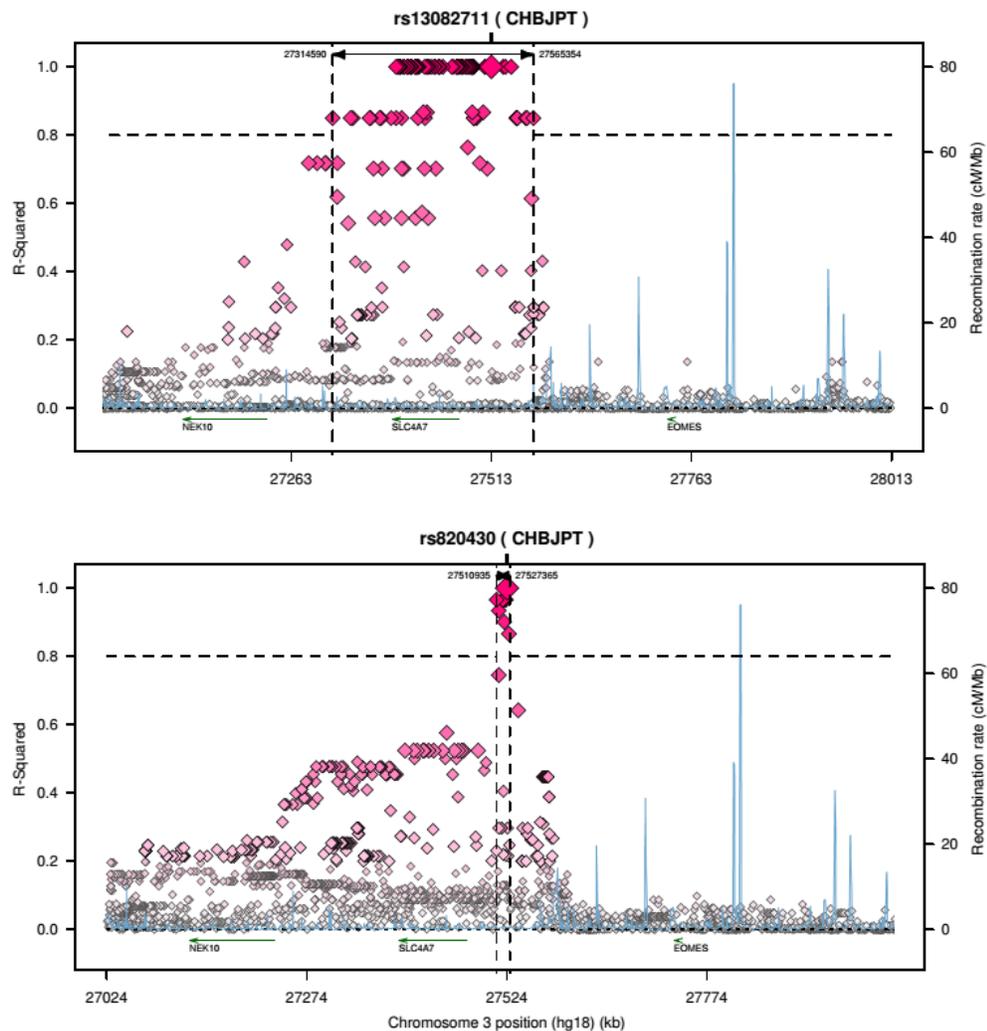


Figure 10: Linkage disequilibrium plots at the rs13082711 and rs820430 SNPs
 Linkage disequilibrium plots at the Ehret *et al.*, 2011 (rs13082711) and Le *et al.*, 2014 (rs820430) blood pressure-associated SNPs based on HapMap3 (release 2) data for the CHBJPT population (<http://www.broadinstitute.org/mpg/snap/ldplot.php>). Position along chromosome 3 is matched on both plots. An LD of $r^2 > 0.8$ (dashed lines) is marked as strong linkage disequilibrium within the loci.

1.19. Considerations of NBCn1 as pharmacological target

The widespread expression profile of NBCn1 (Boedtkjer *et al.*, 2008) would raise the potential for adverse effects based on the impact on other tissues. This is exemplified by developmental abnormalities observed with the *SLC4A7*^{-/-} knockout mouse is similar to Usher's Syndrome, with developmental abnormalities in photoreceptor cells together with reduction of auditory brainstem responses between one and four months, with resultant hearing and sight impairment (Bok *et al.*, 2003).

Additionally, *SLC4A7*^{-/-} knockouts demonstrate non-Mendelian inheritance, with a breeding pair of two *SLC4A7*^{+/-} heterozygotes produces only 18% *SLC4A7*^{-/-} homozygotes as opposed to the expected 25%, suggesting some selection pressures, although maintaining equal gender distributions (Boedtkjer *et al.*, 2011). However, there does not appear to be a significant selection pressure exerted by the blood-pressure associated SNP rs13082711, as it appears that its genotype frequency does not deviate from Hardy-Weinberg equilibrium in a number of populations (see Section 3.2).

On the other hand, knockout mice can be crude models to predict potential pharmacological inhibition of the NBCn1 protein, due to the constitutive absence of the function of the protein of interest. While constitutive knockouts may affect maturing organs, pharmacological agents used in adults should not impact on tissue maturation. Additionally, gene knockout exerts the equivalent of full pharmacological antagonism, whereas pharmacological agents can behave as partial antagonists or even as perhaps partial agonist. Additionally, the current published *SLC4A7*^{-/-} knockout model is not tissue-selective. Pharmacological agents may have the possibility of tissue-specific targeting, or at least limiting the crossing of the blood-brain barrier.

The only NBC inhibitor developed thus far is S0859, but is not selective between the different NBC-subtypes. Additionally, it has been shown to have a very strong affinity to components in plasma, and as a result showed no effect on isolated murine tissues at concentrations up to 50 µM (Larsen *et al.*, 2012). While S0859 may be a useful *in vitro* tool, there is currently no small molecule close to *in vivo* use.

1.20. Summary of background

From the recent literature, *SLC4A7* appears to be a biologically plausible candidate gene for blood pressure regulation. A genetic locus encompassing the gene has been identified as being associated with levels of diastolic blood pressure. Additionally, *SLC4A7*^{-/-} knockout mice demonstrate a blood pressure phenotype that appears to be resistant to hypertensive stressors. This follows the overall understanding of pH_i being important in the function of tissues relevant to hypertension. There is also some evidence of altered pH_i regulation in the hypertensive population.

1.21. Hypothesis and aims

The blood-pressure associated SNP rs13082711 may exert its effect by one or more genetic mechanisms, and via pH_i regulation (Figure 11). This may impact on a variety of tissue types, but this project concentrates in vascular smooth muscle and endothelial cells.

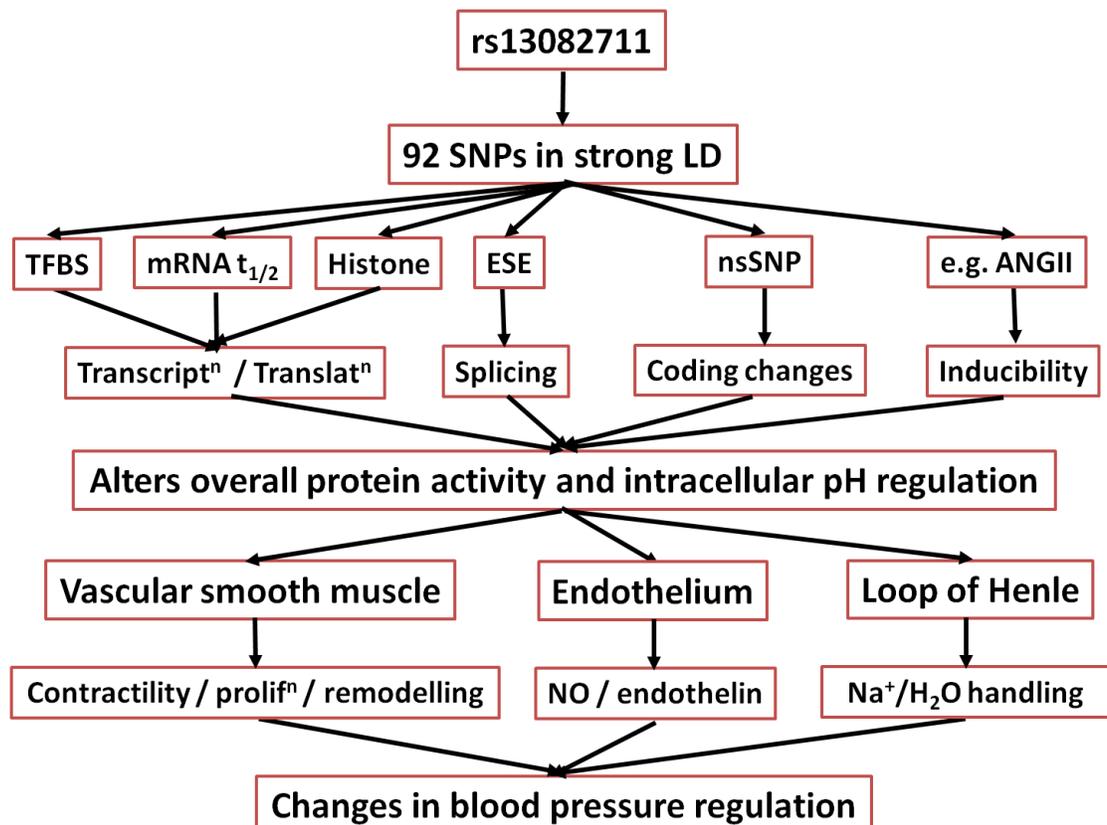


Figure 11: Schema of hypothesized potential mechanisms via which the blood pressure related SNP (rs13082711) exerts its effect

LD, linkage disequilibrium; TFBS, transcription factor binding site; ESE, exonic splicing enhancers; ANGII, angiotensin II.

Not all paths are to be investigated

Using *in vitro* cell lines relating to blood pressure regulation, the work in this thesis aims to investigate the impact of the blood pressure-associated SNP at the *SLC4A7* locus or the SNP(s) in strong LD on:

1. DNA-nuclear protein interactions at the *SLC4A7* locus,
2. NBCn1 expression and localization, and
3. intracellular pH regulation.

2. Methods and materials

2.1. Tissue culture and cell lines

2.1.1. Principles - Choice of cell lines

Human umbilical artery smooth muscle cells (HUASMC)

There are several commonly used vascular smooth muscle cell lines for tissue culture, all grown as adherent cultures. However, the necessity of comparing genotypes / haplotypes in most of the experiments conducted for this thesis limits the potential of experimental cell lines such as the BD (Berlin Drukrey) IX rat *embryonic* aortic vascular smooth muscle A7r5 or A10. Additionally, knowledge gained from animal cell lines can occasionally be difficult to extrapolate to human disease.

On the other hand, there are many commercial suppliers of primary human vascular smooth muscle cell lines, including derived from aorta, coronary artery, pulmonary and umbilical arteries. These are likely to be derived from a very small of donors, and thus not achieving the aim of assessing a population of different genotypes / haplotypes.

The issues highlighted above indicates that to be able to assess the necessary number of samples of different genotypes / haplotypes, an in-house supply of primary cells lines needs to be generated. Whilst human coronary arteries may occasionally be excised in coronary artery bypass grafting, the availability of samples is rare, and even if available would represent a wide range of disease states. Umbilical cords then becomes an ideal source of primary cell lines, as they are widely available, non-diseased and would otherwise be regarded as clinical waste. A further advantage of utilising umbilical cords is the less potential for co-morbidities to affect cellular phenotype as in the case of coronary arteries following bypass grafting – by definition diseased vessels.

There are a couple of counter-arguments against the use of HUASMCs. These primarily center on the nature of umbilical cords being conduit arteries (i.e. not resistance vessels), carriers deoxygenated blood (unlike systemic arteries) and are of embryonic origin (i.e. not adult arteries). Of course, all other primary cell lines derived from aorta and coronary arteries are also not

resistance vessels, and the use of pulmonary arteries would not be applicable for research into systemic hypertension.

Human umbilical vein endothelial cells (HUVEC)

The rationale for utilizing HUVECs as the endothelial cell line of choice is similar to that of HUASMCs above. For the same reasons requiring heterogeneity of samples, the mouse yolk sac endothelial cell line (C166) and the simian virus 40-transfected immortalised human dermal microvascular endothelial cells (CDC/EU.HMEC-1) is not feasible. Although primary endothelial cell lines, especially HUVECs are available commercially, they are also likely to be from a small pool of donors. As above, umbilical cords are also a ready source of primary endothelial cells. However, the same counter-arguments also apply in the case of HUVECs, with the added caveat that they are derived from veins, instead of artery, but notably carry oxygenated blood. Like the HUASMCs, HUVECs are adherent cell cultures.

MCF-7

MCF-7 cells are a commercially-available adherent epithelial cell line initially isolated from the metastatic pleural effusion from human breast adenocarcinoma, and is typically used as a transfection host (ATCC; http://www.lgcstandards-atcc.org/products/all/HTB-22.aspx?geo_country=gb). For this thesis, it is a reasonable positive control for the gene expression of several genes of interest (e.g. *SLC4A7*, *NEK10*). As a cancer cell line with multiple mutations, no further interpretation is made of the profile and relative quantity of gene expression, as it is not a comparable cell line for the work in this thesis.

A10 cells

A10 cells are a commercially-available adherent myoblastic cell line initially isolated from the medial layer of the embryonic *Rattus norvegicus* (DBIX strain) thoracic aorta (Kimes and Brandt, 1976), and is typically used as a transfection host and has characteristics of smooth muscle cells including expression of myokinase and creatine phosphokinase (ATCC; http://www.lgcstandards-atcc.org/products/all/CRL-1476.aspx?geo_country=dk). For this thesis, it was utilised as a transfection host, to provide a smooth muscle intracellular milieu, albeit murine and embryonic.

HEK293 cells

HEK293 cells are a commercially-available adherent epithelial cell line which may act as a positive control for the expression of several genes of interest (e.g. *SLC4A7*, *NEK10* and *TFAP2C*). Results from HEK293 cell lines should however be interpreted with caution. Whilst they were initially derived from embryonic kidney cells, they have expression profiles similar to neuronal tissue (Shaw *et al.*, 2002). Additionally, most of the cells have been shown to have a hypotriploid karyotype, with most cells having 64 chromosomes, but some with more (European Collection of Cell Cultures, n.d.). Thus, no further interpretation is made of the profile and relative quantity of gene expression, as it is not a comparable cell line for the work in this thesis. For this thesis, it was utilised as a transfection host, to provide an easily transfectable cell from a human source, albeit an experimental cell line.

NIH Swiss 3T3 cells

3T3 cells are a commercially-available adherent fibroblast cell line derived from embryonic fibroblasts from NIH Swiss mice. These cells were used purely as an experimental model, and were not intended to be extrapolated to vascular smooth muscle or endothelial cell function. Additionally, most of the cells have been shown to have a hypertriploid karyotype, with most cells having 68 chromosomes (note diploid number of chromosomes is 20), but some with more (Leibiger *et al.*, 2013).

CD8⁺ T-lymphocytes (cDNA kind gift from Dr S Nadkarni)

CD4⁺ T-lymphocytes isolated by flow cytometry from human peripheral blood is used as a positive control for *EOMES* gene expression (McLane *et al.*, 2013).

cDNA from renal nephrectomies (kind gift from Professor P Munroe)

A limited collection of cDNA derived from the non-diseased pole of renal carcinoma nephrectomies were gifted by Professor Munroe. This had a restricted role in the thesis for several reasons:

- Kidneys are highly heterogeneous organs, with a multitude of cell types. As the surgical samples could not be obtained uniformly, they are likely to have significant variations in the proportions of cells within them. This precludes the samples from a meaningful qRT-PCR assessment.

- Although designated as the non-disease pole, this is a macroscopic label, and was not histologically examined to consider whether there are microscopic regions of tumours.
- There are no provisions for a larger sample size, and any attempts at establishing primary cultures were not feasible.

2.1.2. Principles - Tissue culture

The conduct of tissue culture required sterile conditions. This included the use of Class II laminar flow biological safety cabinets, single-use sterile pipettes and flasks, autoclaved liquids, sterile reagents, and where necessary sterile-filtered using a 0.22 µm filter (Milipore, SLGVV255F). Waste from such cultures needed to be treated with broad-spectrum disinfectants (e.g. Virkon®, Du Pont) which have anti-viral, anti-fungal and anti-bacterial properties. For all centrifugation steps, cells were spun down at low speeds (1,000g) to maintain cell viability.

As the work conducted for this thesis utilised adherent cells, they needed to be detached from the tissue culture surface with minimal damage, either harvesting for experiments, or for subculturing. In most instances, unless stated, harvesting cells for experiments was conducted with single-use cell lifters. For subculturing, there are several enzymatic options, one of which is trypsin. Trypsin is a serine protease and is able to hydrolyse proteins. It functions to cleave peptide chains mainly at the carboxyl- side of lysine or arginine residues, unless when followed by proline. As trypsin is able to undergo autolysis, they were stored at cold temperatures when not in use. Conversely, to optimise enzymatic efficiency, it was warmed to 37°C when in use and was prepared at a pH of ~8. Once the cells have lifted off the tissue culture surface, it can be inactivated with serum (either fetal bovine, bovine calf or human).

While in use in a tissue culture system, trypsin digests extracellular proteins, primarily to detach adherent cells from the tissue culture surface. In addition to digesting adhesion proteins, it will also interact with most of the other membrane-bound proteins. This is particularly relevant for the current study, as trypsin would also degrade the membrane-bound protein of interest, NBCn1. Thus, for experiments that require intact NBCn1 proteins, it is essential that cell lifters were used for harvesting the cells for protein assays.

Primary cells used for the experiments in this thesis were younger than 7th passage. This is important as cellular phenotypes of primary cells change with subsequent population doublings,

confounding the ability to draw physiologically relevant conclusions. For example, as demonstrated with baboon femoral artery endothelial cells, most cells become senescent by the 30th population doubling, with around a 50% reduction in eNOS protein levels, but no significant changes in E-selectin, VCAM-1 and ICAM-1 expression (Shi *et al.*, 2004).

Where indicated, serum-free media was used to induce synchronised cell arrest of the cultured cells (Pardee, 1974). Even if there may be some dispute on whether if cells arrest at a specific point within the G₁ phase of a cell cycle or not (Cooper, 2003), this provides a standardized state of the cells to conduct experiments. Additionally, the presence of serum may breakdown pharmacological challenges – for example, angiotensin II has a very short half-life in serum, approximately 16 seconds in the circulatory system of the rat (Al-Merani *et al.*, 1978). The dose of external challenges used are planned to be as close to circulating physiological levels as possible, ranging from 10 to 100 µg/ml for angiotensin II, and 1 to 16 µM for norepinephrine (Bergeron *et al.*, 2001).

A large part of the work conducted for this required the establishment of primary cell lines, both of vascular smooth muscle and endothelial cells. This can be achieved by using adherent explants (such as with vascular smooth muscle cells) where the tissue is dissected and placed opposing a tissue culture surface for cells to grow onto. Alternatively, this can be through enzymatic digestion (such as with endothelial cells) where collagenase type I was used to lift the endothelial cells off the luminal surface, and this was then seeded onto an appropriate cell culture surface.

The work in this thesis also requires preservation of cell lines for future utilization, for which the cells are stored at low temperatures. While ultra-low temperature freezers (-80°C) are appropriate in the short term, storage in liquid nitrogen storage tanks (less than -180°C) virtually stops the biological activity of the cells, allowing safe storage.

The main damage to cells during cryopreservation occurs during the freezing stage. There are several contributory reasons. Firstly, solutes are excluded as ice crystals grow in freezing water, causing them to become very concentrated in the remaining liquid water, potentially damaging cells with osmotic pressures. Secondly, when tissues are cooled slowly, water molecules migrate out of cells. This forms ice outside of the cells, and may cause mechanical damage to the cell membrane. This migration of water also results in cellular dehydration, and associated stresses on

the cell can cause damage directly. Finally, and perhaps most importantly, intracellular ice formation can be fatal to cells.

To counter the potential damage by freezing, dimethyl sulfoxide (DMSO) was used in tissue culture as a cryoprotectant. When added prior to the freezing process, it reduces ice formation and thereby prevents cell death during the freezing process. This protective effect needs to be balanced against the intracellular toxicity of DMSO – and 10% v/v is a typical concentration in cryopreservation media. To further enhance survival through the cryopreservation process, the cooling is slow and controlled, with the optimal rate of being lowered at 1°C per minute (Mazur, 1984).

A further complication is the necessity to slowly dilute out the DMSO from the cells upon re-warming the cells by adding the resuspending media in a dropwise manner (Weed and Jenkins, 1979). Unlike the cooling period, cells survive the cryopreservation process better when warmed up rapidly (Mazur, 1984).

2.1.3. Human umbilical cord specimens - Ethical approval and exclusion criteria

The work conducted for this thesis had ethical approval (Protocol No.: Plaque-WHRI-01; NRES ref: 08/ H0704 / 140 dated 21st January 2009 (Appendix 7a); AM04 Substantial Amendment dated 8th March 2012 (Appendix 7b) and further expanded with AM05 Minor Amendment dated 01 Jan 2013 (Appendix 7c).

Umbilical cords were obtained from The Royal London Hospital, Barts and The London NHS Trust, Whitechapel Road, Whitechapel, London E1 1BB, which is one of the three listed Clinical Source Centers. As per the ethical approval, all tissue samples were fully anonymous and devoid of patient clinical data before distribution to the recipient analysis groups.

In brief, prior to the placenta being discarded by the clinical team, the umbilical cord was cut and placed into a specimen pot filled with Hanks' Balanced Salt Solution (HBSS), and kept at 4°C before transfer to the named Study Centre.

Exclusion criteria for the study includes known infective status of the mother (e.g. viral hepatitis B and C, HIV, Group-B streptococcus), insufficient cord length (< 20cm), monozygotic twins (where

only one cord will be assessed) and stillbirths. Pre-eclampsia was not a pre-stated exclusion criteria based on the ethical approval, but the umbilical cords were also not collected.

2.1.4. Materials

Medium 199 (M199) (Sigma, D5671) supplemented with 15% v/v fetal bovine serum (FBS) (Appleton Woods, FB021), 2.5 µg/ml human beta-endothelial cell growth factor (Sigma, E1388), 2.25 mg/ml endothelial cell growth supplement from bovine neural tissue (Sigma, E2759), 1.25 mg/ml thymidine (Sigma, 89270), 5,000 units/ml heparin from porcine intestinal mucosa (Sigma, H3393), 2 mM glutamine (Sigma, G7513), 100 units/ml penicillin (Sigma, P0781) and 100 µg/ml streptomycin (Sigma, P0781), hereby in this thesis termed “supplemented M199”. Where stated, “unsupplemented M199” denotes Medium 199 without any additives.

Dulbecco’s Modified Eagle Medium (DMEM) (Sigma, M4530) supplemented with 15% v/v FBS, 0.5 ng/ml human epidermal growth factor (ProSpec, CYT-217), 2 ng/ml human fibroblast growth factor (ProSpec, CYT-218), 5 µg/ml human insulin (ProSpec, CYT-270), 2mM glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin, hereby in this thesis termed “supplemented DMEM”. Where stated, “unsupplemented DMEM” denotes Dulbecco’s Modified Eagle Medium without any additives. When culturing HEK293 or A10 cells, DMEM was supplemented with 10% v/v FBS, 2mM glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin, hereby in this thesis termed “DMEM+10%FBS”.

Phosphate buffered saline (PBS) was formulated initially at 10x concentration using 80g NaCl, 2g KCl, 14.4g Na₂HPO₄ and 2.4g KH₂PO₄ dissolved into 800ml distilled-deionised water. This was adjusted to pH 7.4 (with concentrated HCl or NaOH) and the volume adjusted to 1 L using distilled-deionised water. This solution was later diluted to 1x working concentration (1 part 10x PBS, 9 parts distilled-deionised water) and autoclaved for sterility. The final concentration is 137 mM Na⁺, 2.7 mM K⁺, 139.7 mM Cl⁻, 10 mM PO₄³⁻.

Physiological saline solution (PSS) was formulated initially at 10x concentration using 64.28 g NaCl, 3.73 g KCl, 4.07 g MgCl₂, 23.83 g HEPES, 8.4 g NaHCO₃, 0.68 g KH₂PO₄, 1.82 g EDTA dissolved into 1L distilled-deionised water. This solution was later diluted to 1x working concentration and autoclaved separately from glucose and CaCl₂ (which is added to reach the final concentration of 10 mM and 0.15 mM respectively).

Hanks' Balanced Salt Solution (HBSS) had a final formulation of 138 mM NaCl, 5.33 mM KCl, 4 mM NaHCO₃, 1.26 mM CaCl₂, 0.5 mM MgCl₂, 0.44 mM KH₂PO₄, 0.41 mM MgSO₄, 0.3 mM Na₂HPO₄ and 5.6 mM glucose.

Cryopreservation media consisted of 70% v/v FBS, 20% v/v basal media (either unsupplemented DMEM for HUASMCs or unsupplemented M199 for HUVECs) and 10% v/v DMSO, constituted just prior to use.

Specific reagents used in tissue culture systems for various experiments are discussed in their respective methods sections. All are sterile when used in cell culture systems. Unless stated, all materials were warmed to 37°C prior to use.

2.1.5. Methods

Derivation of primary human umbilical artery vascular smooth muscle cell (HUASMC) lines

The following protocol was adapted from Leik *et al.*, 2004, and conducted in a designated Class II laminar flow biological safety cabinets specifically reserved for primary cultures. In summary, the umbilical cord was kept moist in a petri dish filled with PSS. It was then dissected to expose the umbilical arteries (usually two per cord, but occasionally one), which then using forceps, the outer fascia and adventitia was peeled off to reveal the underlying tunica media. These arteries were then trimmed to approximately 6 to 8 cm, and transferred to a new petri dish with fresh PSS. The arteries were dissected longitudinally and any remaining intravascular blood clots were flushed away. The arteries were then cut further into sections of approximately 3mm in length. These small sections were washed twice with PSS, and once with unsupplemented DMEM. The residual wash solutions were removed, and the sections were transferred to a T-25 flask pre-coated with 0.2% w/v gelatin. The tissue culture flask was then placed upright in the tissue culture incubator to allow further draining of wash media, which would otherwise reduce attachment and formation of explants. After two hours, the media that has drained from the artery segments was removed from the tissue culture flask, and replaced with supplemented DMEM while being careful not to displace the attached segments. Primary cells usually appear after seven days. From then on, the culture media was replaced every two to three days. Once the tissue sections detach, they are removed from the cell culture system.

Derivation of primary human umbilical vein endothelial cell (HUVEC) lines

The following protocol was adapted from Jaffe *et al.* (1973), and was conducted in a designated Class II laminar flow biological safety cabinets specifically reserved for primary cultures. In summary, the umbilical vein was identified, and cannulated using a 18G cannula at one end, and clamped in position. Through this cannula, the umbilical vein was washed with 50ml of Hanks' Balanced Salt solution (HBSS), removing the intraluminal clots. Following the wash, 0.2% w/v type I collagenase (Sigma, C9891) diluted in HBSS was introduced via the cannula. When the collagenase started to drip out the distal end, the cord was clamped on both ends. Further type I collagenase was injected until the vein was significantly distended. On average, a 15 cm section of umbilical vein required 15 to 20ml of collagenase. The cord was then incubated with collagenase and clamped *in situ* for ten minutes at 37°C with gentle massaging of the cord at the start.

Following the incubation the distal clamp was released and the fluid collected in a 50ml sterile centrifuge tube. The vein was further flushed with 30 to 35ml of PBS until to a final volume of 50ml. The solution, now containing the HUVECs, was centrifuged slowly, at 1,000g for 10 minutes at room temperature. The supernatant was carefully discarded, and the cell pellet resuspended in supplemented M199. The cells are then plated in T-75 flasks pre-coated with 0.2% w/v gelatin. The culture media was changed the next day to remove non-adherent cells. Subsequently, the culture media was replaced every two to three days until suitable for further subculture. Confluency was typically achieved in 6 to 8 days with "cobblestone appearance" on optical microscopy.

General maintenance

All tissue culture procedures were conducted in sterile conditions, and where required, in a Class II laminar flow biological safety cabinets. The culture media was replaced every two to three days unless the cells required subculturing or were suitable for the start of an experimental protocol. Tissue culture was conducted in a humidified incubator kept at 37°C and 5% CO₂. Cellular morphology and growth was regularly monitored.

Subculturing (passaging) cells

When the cells reached 80-90% confluency, they were suitable for subculturing, if required. The adherent cells were initially washed with PBS, and the wash solution is subsequently fully removed. A volume of trypsin (Sigma, T4049) just enough to cover the cell culture surface (350 µl for a T-25 flask; 1ml for a T-75 flask) was added to the flask, gently swirled and left in the

humidified cell culture incubator – approximately 2 minutes for HUASMCs, 30 seconds for HUVECs. The cell culture surface was then observed under the optical microscope for cell detachment. Once the cell detachment is confirmed, supplemented media (DMEM for HUASMCs, M199 for HUVECs, defined in Section 2.1.4) was added at a sufficient volume for further cell culture flasks. The new tissue culture flasks were pre-coated with 0.04% w/v porcine skin gelatin (Sigma, G1890).

Cryopreservation and subsequent resuscitation of cryopreserved cell lines

When the cells were suitable for cryopreservation, the flask was washed with PBS, and all the residual wash solution aspirated. Then a volume of trypsin just sufficient to cover the tissue culture surface was added. This was allowed to incubate in the cell culture incubator for the cell type-dependent time, until cell detachment is observed on the optical microscope. Once the cells have detached, supplemented media (DMEM for HUASMCs, and M199 for HUVECs) was added to neutralise the trypsin. This cell suspension was collected into a 15ml centrifuge tube, and slow centrifuged at 1,000g for 5 minutes at room temperature. The supernatant is discarded, and the cell pellet was resuspended with cryopreservation media (see Materials above). A confluent T-25 flask contained sufficient number of cells for 1ml of cryopreservation media, and in turn sufficient for resuscitation into a new gelatin pre-coated T-25 flask in the future. Once resuspended, the cells were transferred into cryovial(s) (Thermo Scientific, 5000-0020), which is then placed in a cryopreservation container (Thermo Scientific, 5100-0001) filled with isopropanol. This container is initially transferred to ultra-low temperature (-80°C) freezers, with the container controlling the decline in temperature. Once at -80°C, usually left overnight, it is then suitable to be transferred to liquid nitrogen storage tanks.

To resuscitate cryopreserved cells, the cryovials were transferred from its storage on dry ice to the relevant cell culture areas. It was then rapidly warmed up by a water bath set at 37°C. Once it has reached the desired temperature, usually after 10 minutes, the cell suspension within the cryovial was transferred into a 15ml centrifuge tube. Unsupplemented media (DMEM for HUASMCs, and M199 for HUVECs) was added, initially drop-by-drop, to dilute out the DMSO. The final volume added is 10x that of the cryopreservation media. This cell suspension was then slow centrifuged at 1,000g for 5 minutes at room temperature. The supernatant was discarded, and the cell pellet resuspended with the appropriate supplemented media (DMEM for HUASMCs, and M199 for HUVECs), and seeded onto a T-25 flask pre-coated with 0.04% w/v gelatin.

2.2. DNA isolation from cells or tissues

2.2.1. Principles

DNA isolation with isopropanol is based on the insolubility of DNA in isopropanol (and ethanol). The first step is lysing the cells, usually with lysis buffers containing detergents or surfactants to aid disruption of the lipid membrane. Cell lysis can be aided by physical methods such as sonication or using the vortex mixer.

When extracting DNA from tissue samples with connective tissue, digestion with Proteinase K is required. Proteinase K is a broad-spectrum endopeptidase / serine protease with broad activity over a wide range of pH as well as an optimal temperature of 50°C to 60°C. Proteases and RNases may also be used to remove proteins and RNAs respectively from the sample as potential contaminants. Alternatively, treating with high concentrations of sodium chloride precipitates out the proteins in the sample. Sodium chloride “salts out” the proteins as the sodium ions attract water molecules away from interacting with the hydrophilic / charged sections of proteins. In turn, the hydrophobic portions of the proteins become relatively stronger than hydrophilic interactions. This results in protein molecules coagulating by forming hydrophobic interactions with each other.

DNA is highly soluble in water, primarily due to its polar and negatively charged phosphate backbone. Due to its polarity, it forms a hydration shell when dissolved in water. On the addition of isopropanol, which is much less polar than water, this removes the hydration shell and allows the phosphate groups to form stable ionic bonds with positive ions in the solution, usually with sodium ions in the lysis buffer. This precipitate can be further washed with ethanol. Eventually, when the ethanol is evaporated, the DNA pellet can be dissolved with nuclease-free water.

Alternatively, commercial kits with mini-columns are available for DNA extraction.

2.2.2. Methods

Isopropanol / ethanol method

The samples were initially mixed with 500 µl lysis buffer (10mM Tris pH 8.0, 10mM EDTA pH 8.0, 100mM NaCl, 0.5% w/v sodium dodecylsulfate), with pipette mixing. This is sufficient for up to 10^7 cells from tissue culture. However, pieces of tissue up to 20 mg required additional digestion by 400 µg Proteinase K (Sigma, P2308) added to the lysis buffer and incubated overnight at 55°C with occasional vortex mixing.

At this stage, isolation of DNA from either cell pellets or digested tissue samples required the same steps. Following cell lysis, 250 µl of 5M NaCl is added and mixed on the roller shaker for 5 minutes at room temperature. This precipitates the proteins, and was separated from the soluble DNA by fast centrifugation at 13,000g for 7 minutes at room temperature.

The supernatant with the soluble DNA is transferred to a new microcentrifuge tube, to which 500 µl of isopropanol is added. This was mixed for 2 minutes on a roller shaker at room temperature. At this stage, the DNA is precipitated and collected by fast centrifugation at 13,000g for 5 minutes at room temperature. The supernatant was discarded, and the DNA pellet was washed with 70% v/v ethanol and once again pelleted by fast centrifugation at 13,000g for 2 minutes at room temperature. The supernatant was removed, and the residual ethanol was left to air-dry.

Once dry, the DNA pellet was dissolved with 50 µl of nuclease-free water and incubated at 55°C for 10 minutes. The DNA concentration and purity was then measured using a spectrophotometer (Nanodrop® ND-1000, Thermo Fisher Scientific Inc., USA). For a sample that is free of contaminants, DNA is expected to have an A_{260}/A_{280} ratio (ratio of absorbance of ultraviolet light at wavelengths of 260 nm and 280 nm respectively) of approximately 2.0, where A_{260} typically reflects nucleic acid content. If the ratio is much lower, it may indicate the presence of proteins, phenols or other contaminants that absorb strongly at or near 280 nm. A secondary measure of nucleic acid purity is that of the A_{260}/A_{230} ratio, commonly expected to be around 2.0 to 2.2. In the presence of excess EDTA, carbohydrates, phenols and guanidine HCl, all absorbing ultraviolet light at around 230 nm, the sample will reveal a reduced A_{260}/A_{230} ratio.

During the studies conducted for this thesis, this method produced a larger quantity of DNA and of higher purity as compared to commercial kits, albeit was more time consuming.

2.3. RNA isolation and reverse transcription into cDNA (complementary DNA)

2.3.1. Principles – RNA isolation

There are several methods that are commonly used for extracting total cellular RNA. This includes the phenol-chloroform extraction method, as first described in 1987 (Chomczynski *et al.*, 1987) as well as column-based protocols.

RNA isolation – column-based

Column-based nucleic acid purification is a solid phase extraction method to quickly purify nucleic acids. It also has the benefit compared to phenol-chloroform based methods as it does not use hazardous materials.

This method relies on nucleic acids binding to the solid phase (usually silica) depending on the pH and the salt content of the lysis buffer. There are typically three stages. First, a buffer is added to lower the pH, then the appropriate solute content thus allowing the sample to bind the silica column. The lysis buffer also often contains a denaturing agent such as guanidine hydrochloride, which denatures proteins including RNases.

The column is then washed, usually containing a higher percentage of ethanol, several times. Finally, the column can be eluted with buffer or water.

RNA isolation – liquid-liquid extraction technique

Acid guanidinium thiocyanate-phenol-chloroform extraction is a liquid-liquid extraction technique that is able to separate out RNA from other cellular components. This method typically takes longer than column-based systems, but usually provides higher purity and recovery of RNA. Additionally, standard RNA columns are unsuitable for purification of short (<200 nucleotides) RNA species, such as siRNA, miRNA, gRNA and tRNA.

The lysis buffer for this method, amongst others, includes both guanidinium thiocyanate and phenol. The former denatures proteins, including RNases, and separates rRNA from ribosomal proteins, while the latter is a solvent with poor solubility. Following the addition of chloroform, these solvents separate into two phases with different colours. The clear, upper aqueous phase contains nucleic acids and the lower phase containing the proteins dissolved in phenol and the lipids dissolved in chloroform. With the introduction of acidic conditions, RNAs remains in the aqueous phase, whilst DNA partitions into the organic phase.

The aqueous phase can then be used for further isopropanol precipitation. Like DNA, isopropanol precipitates RNAs by removing its hydration shell and allows the phosphate groups to form stable ionic bonds with positive ions in the solution. This precipitate can be further washed with ethanol.

Eventually, when the ethanol is evaporated, the RNA pellet can be dissolved with nuclease-free water.

2.3.2. Principles - Reverse transcription

First strand synthesis primers

There are two main options for primers in first strand synthesis – oligo(dT) and random primers.

Oligo(dT) primers are based on targeting the poly-adenosine tail present in mRNA. They are usually either oligo(dT)₁₂₋₁₈ (a mixture of 12-mer to 18-mer thymidines), oligo(dT)₂₀ (a homogenous mixture of 20-mer thymidines) or anchored oligo(dT) primers which anneals at the 3'-UTR/polyA junction.

An advantage of oligo(dT) primers is that it targets the poly-adenine tail present only in mRNA – which significantly increases the specificity of the amplification. However, as this also means all reverse transcription is initiated at the 3' end, the difficult secondary structure may interfere with complete cDNA synthesis. Additionally, a commonly used internal control in real-time quantitative PCR, 18S rRNA (ribosomal RNA), does not have a poly-A tail. Thus, if 18S is to be used as an internal control, oligo(dT) primers should not be used.

Random primers are oligonucleotide sequences synthesized entirely randomly. Commonly these primers are six (hexamers) or fifteen nucleotide long (pentadecamers). This gives a numerous range of sequences that have the potential to anneal at many random points on a RNA sequence and act as primers to commence first strand cDNA synthesis.

This potential to anneal and act as a primer at many points of an RNA sequence allows it to be good for synthesizing large pools of DNA. Additionally, as first strand synthesis starts throughout the RNA molecule, it does not have the same issues with non-polyadenylated RNAs.

2.3.3. Method

RNA isolation – Acid guanidinium thiocyanate-phenol-chloroform extraction technique

The liquid-liquid method used for RNA isolation was based on the TRI reagent (acid guanidinium thiocyanate-phenol-chloroform extraction). The protocol was based on the manufacturer's

recommendations (<http://www.sigmaaldrich.com/technical-documents/protocols/biology/tri-reagent.html>) and is summarized as below.

As TRI reagent (Sigma, T9424) is not compatible with plastic tissue culture flasks, cells were initially isolated by cell lifters and centrifugation into a cell pellet. 1 ml of TRI reagent was used to process up to 10^7 cells, therefore suitable for either a confluent T-25 or T-75 flask. The cell lysate was homogenized by pipette and vortex mixing. The cell lysates were then centrifuged at 13,000 g for 10 minutes at 4°C to pellet the insoluble material such as plasma membranes, polysaccharides, and high molecular mass DNA. The supernatant contained RNA and protein, and was transferred into a fresh microcentrifuge tube.

To ensure complete dissociation of nucleoprotein complexes, the samples were allowed to stand for 5 minutes at room temperature. Then, 200 µl of chloroform : isoamyl alcohol (24:1 v/v) (Sigma, C0549) was added and mixed vigorously with a vortex mixer for 15 seconds. This mixture was allowed to stand for 2–15 minutes at room temperature before centrifugation at 13,000g for 15 minutes at 4°C. The centrifugation separates the mixture into 3 phases - a red lower organic phase (containing protein), an interphase (containing DNA), and a colorless upper aqueous phase (containing RNA).

The aqueous phase containing RNA was then transferred into to a fresh microcentrifuge tube. Then, 500 µl of isopropanol is added to the sample, vortex mixed, and allowed to stand for 10 minutes at room temperature to precipitate the RNA. The mixture was then centrifuged at 13,000g for 10 minutes at 4°C to pellet the RNA precipitate. The supernatant was discarded, and the RNA pellet washed with 1 ml of 70% v/v ethanol. The suspension was then centrifuged at 13,000g for 10 minutes at 4°C to pellet the RNA precipitate again.

The wash ethanol was discarded, and the RNA pellet allowed to air dry for 5–10 minutes by air-drying. However, the RNA pellet should not completely dry out, as this greatly decrease its solubility. The pellet was then dissolved in nuclease-free water and pipette-mixed, and if required, incubated at 55–60 °C for 10–15 minutes.

As part of optimization, the RNA sample was also run on an agarose gel demonstrating the 28S, 18S and 5S bands together with an RNA smear (Figure 12).

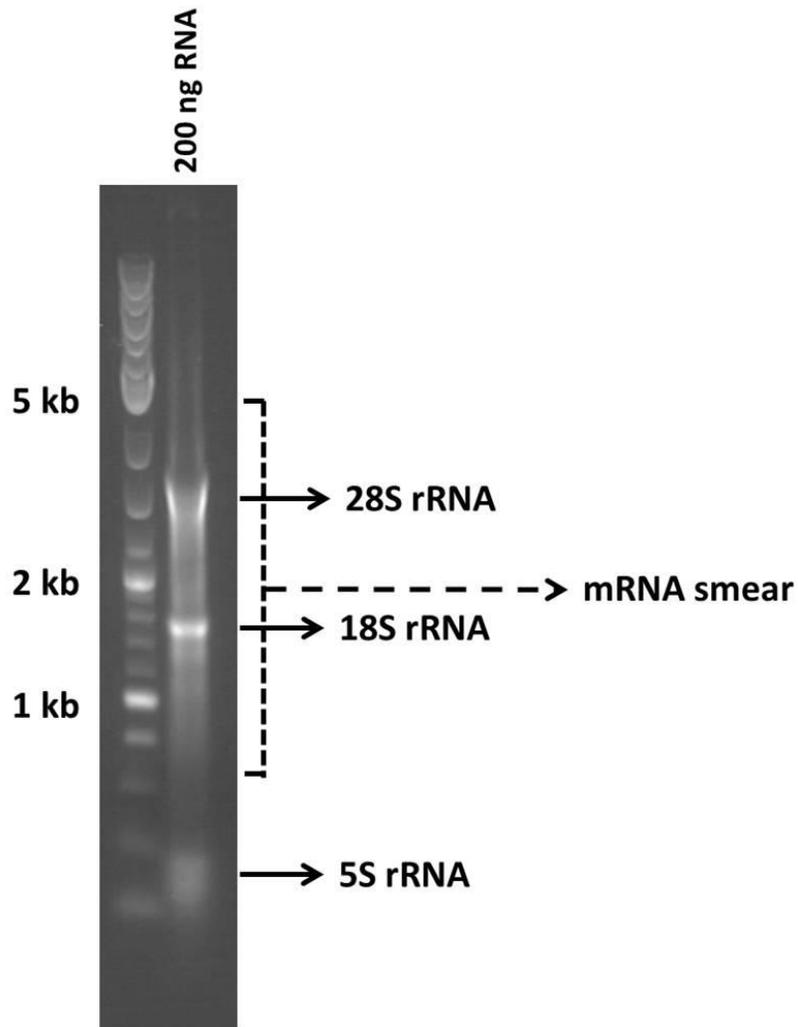


Figure 12: Agarose gel electrophoresis of RNA isolated using the acid guanidinium thiocyanate-phenol-chloroform extraction technique
 Thick bands represent the 28S, 18S and 5S RNA species, whereas the smear is the expected sizes for total cellular RNA. No appreciable genomic DNA signal seen

RNA isolation – column method

The column method used for total cellular RNA isolation for this thesis utilised Nucleospin RNA II columns (Macherey-Nagel, 740955), and the protocol is summarized below. This column is able to process up to 5×10^6 cells, which equated to a 25 cm^2 tissue culture surface area (T-25 flask).

To lyse cell pellets, 350 μl Buffer RA1 and 3.5 μl beta-mercaptoethanol was added to each sample in a microcentrifuge tube and vortex mixed. This cell lysate was then transferred into the NucleoSpin Filter, and centrifuged at 13,000g for 1 minute at 4°C . This step filters the lysate and reduces the viscosity of the sample.

To ensure RNA binding to the spin columns, 350 μ l of 70% v/v ethanol was added to adjust RNA binding conditions, and pipette mixed. The mixture was then transferred into the Nucleospin RNA II Column, and centrifuged at 13,000g for 1 minute at 4°C. This resulted in RNA binding to the column. The flowthrough was discarded. The silica membrane was desalted by adding 350 μ l of Membrane Desalting Solution B, and centrifuged at 13,000g for 1 minute at 4°C with the flowthrough discarded.

To digest any contaminating DNA, a DNase reaction mixture (90 μ l rDNase Reaction Buffer and 10 μ l rDNase) was dropped onto the silica membrane and left for 15 minutes at room temperature. To wash the silica, 200 μ l Buffer RA2 was added to the column, and centrifuged at 13,000g for 30 seconds at 4°C with the flowthrough discarded. After that, 600 μ l Buffer RA3 was added to the column, and centrifuged at 13,000g for 30 seconds at 4°C with the flowthrough discarded. The final wash was conducted with 250 μ l Buffer RA3 added to the column, and centrifuged at 13,000g for 2 minutes at 4°C with the flowthrough discarded. To dry the silica, the column was further centrifuged at 13,000g for 1 minute at 4°C and the flowthrough discarded.

To elute the RNA, 40 μ l of RNase-free water was added to the silica membrane and centrifuged at 13,000g for 2 minutes at 4°C with the flowthrough collected. The eluate was quantified by a spectrophotometer.

RNA yield and purity quantification

The RNA content and purity was quantified using a spectrophotometer. For a sample free of contaminants, RNA is expected to have an A_{260}/A_{280} ratio (ratio of absorbance of ultraviolet light at wavelengths of 260nm and 280nm respectively) of approximately 2.0, where A_{260} typically reflects nucleic acid content. If the ratio is much lower, it may indicate the presence of proteins, phenols or other contaminants that absorb strongly at or near 280nm. A secondary measure of nucleic acid purity is that of the A_{260}/A_{230} ratio, commonly expected to be around 2.0 to 2.2. In the presence of excess EDTA, carbohydrates, phenols and guanidine HCl, all absorbing ultraviolet light at around 230nm, the sample will reveal a reduced A_{260}/A_{230} ratio.

Reverse transcription

The reverse transcription conducted for the work in this thesis utilises the M-MLV (Moloney murine leukemia virus) reverse transcriptase (Promega, M170), and the manufacturer's protocol is summarized below.

The protocol was optimised for the reverse transcription of 1000 ng of RNA. This was added to an RNase-free microcentrifuge tube together with 0.1 ng random primers (Promega, C1181), and made up to a final volume of 10 µl with nuclease free water. This was heated to 70°C in a thermocycler (PTC-0225, Peltier DNA Engine Tetrad) for 5 minutes, and immediately returned to 4°C on an ice bath. To this mixture, reagents were added to achieve a final concentration of 1x M-MLV reaction buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂ and 10 mM DTT), 1 mM dNTPs together with 200 units of M-MLV reverse transcriptase. The reaction was made up with nuclease-free water to a final volume of 25 µl. This reaction mixture was returned to the thermocycler, kept at 25°C for 5 minutes, followed by 42°C for 90 minutes and finally 70°C for 15 minutes.

Subsequent PCRs are conducted as per Section 2.5.

Positive and negative controls

For end-point RT-PCR, the following controls should be utilised:

- No template control – whereby a signal indicates cDNA contamination within the PCR system.
- RT negative control (i.e. RNA sample underwent the same process of reverse transcription, but with the reverse transcriptase enzyme replaced with water) – whereby a signal indicates potential cDNA or genomic DNA contamination of the template.
- Positive control (cDNA of a sample that is known to express the RNA of interest) – whereby the absence of a signal indicates a failed reaction.

2.4. Genotyping by KASP™ (Kompetitive Allele Specific PCR) method

2.4.1. Principles – Genotyping overview

Most estimates place the overall frequency of polymorphisms within the human genome at 1 in 1000. This places the overall number of likely polymorphisms at the order of 50 million autosomal

variations (GRCh37, Ensembl genome browser release 68, July 2012; http://jul2012.archive.ensembl.org/Homo_sapiens/Location/Genome?r=1:1-1000000).

The current definition of a genetic polymorphism is the occurrence in the same population, of two or more alleles at one locus, each with appreciable frequency (typically 1%). These variations are not *de novo* mutations, but are actively and steadily maintained by natural selection, in contrast to transient polymorphisms where a form is progressively replaced by another. By this definition, genetic polymorphisms include single nucleotide polymorphisms (SNPs), insertion-deletion (Indel) and variable tandem number repeats (VTNR). Whilst accepting the importance of the other types of polymorphisms, the scope of this thesis only includes SNPs.

Current methods of genotyping SNPs include restriction fragment length polymorphism identification (RFLPI), random amplified polymorphic detection (RAPD), amplified fragment length polymorphism detection (AFLPD), polymerase chain reaction (PCR), DNA sequencing, allele specific oligonucleotide (ASO) probes, and hybridization to DNA microarrays. Most of these methodologies can be viewed as either enzyme-based, or hybridization-based. Enzyme-based methods utilize a range of enzymes, including nucleases (restriction fragment length polymorphism) and DNA polymerases (PCR-based methods where the primer or product interacts with a reporter/quencher system). Hybridization-based methods often include probes that are complementary to the allele-specific sequence of interest, linked together with a reporter/quencher system.

DNA sequencing, such as Sanger sequencing used in this thesis, also provides genotype data, and is often regarded as the gold standard. Currently, the main limitation is the cost per unit, as compared to the other methods.

2.4.2. Principles – KASP™ methodology

The genotyping method used for this thesis was the KASP™ method (Applied Bioscience). It is an endpoint genotyping technology based on fluorescence and associated quenchers, as summarized in Figure 13. It has the benefit of being able to be conducted on 96-, 384- and 1536-well plate formats, and has a relatively low cost-per-unit.

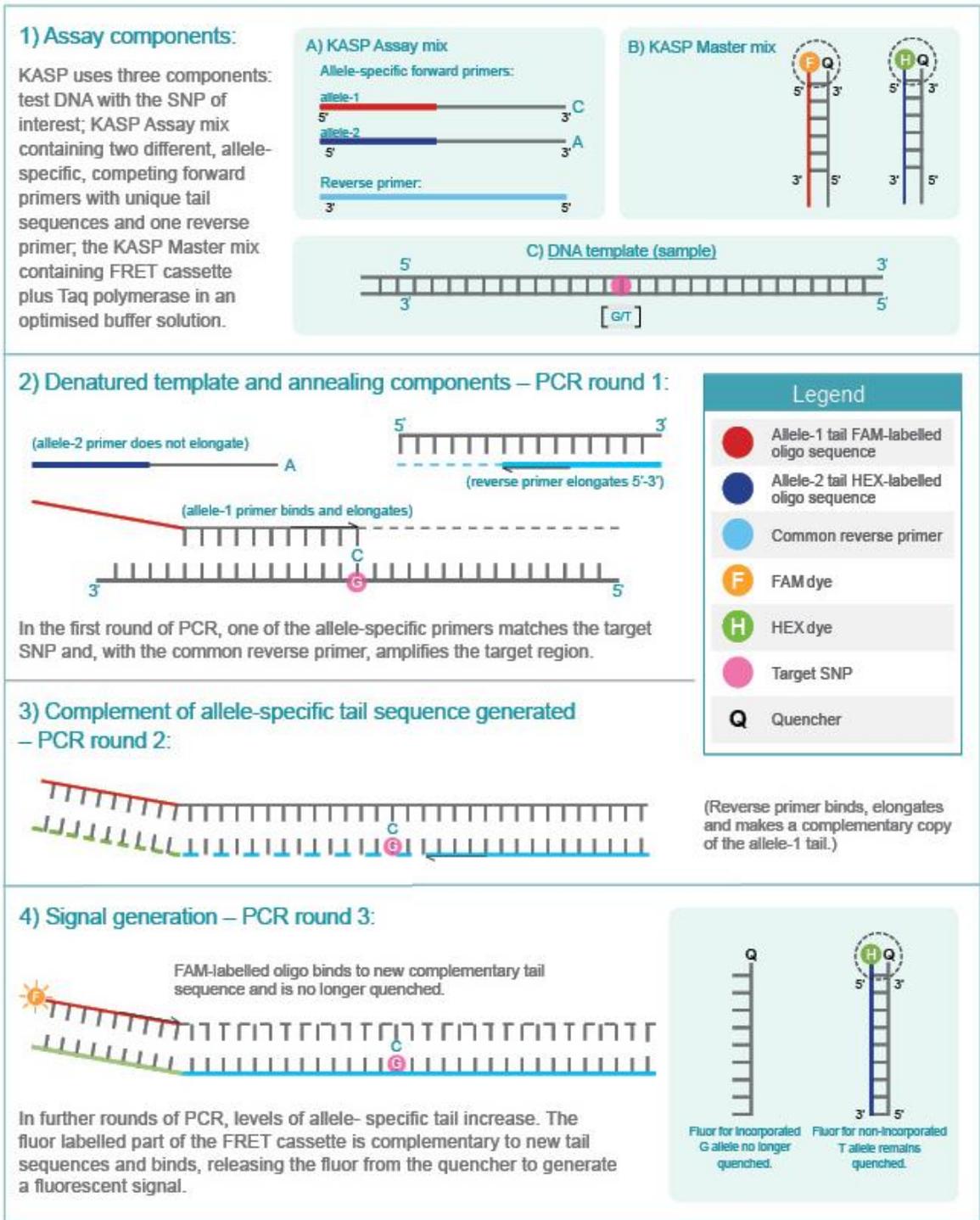


Figure 13: Cartoon demonstrating the chemical basis of the KASP™ genotyping system
 Adapted from <http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/how-does-kasp-work/>.

The assay consists of two key mixtures – the master mix, and the primer mix. The master mix consists of the DNA polymerase and appropriate buffers to catalyze the polymerase chain reaction, and two distinct FRET (Förster resonance energy transfer) cassettes. These cassettes contain a double-stranded oligonucleotide with a fluorescent dye on one 5'-end, and a quencher on the complementary 3'-end. Whilst annealed together, there is no detectable fluorescent signal, but when separated, the signal is detectable. The primer mix includes two allele-specific forward primers and one common reverse primer that are designed by the proprietary primer design program PrimerPicker™ (Applied Bioscience), with an algorithm set, at least in part, for a T_m between 63°C to 66°C. The 5'-end of the primers contain an allele-specific sequence that corresponds to that of the corresponding fluorescence-tagged oligonucleotide, but does not relate to the sequence around the SNP of interest. It is however noteworthy that, the source code for the PrimerPicker™ is proprietary.

When the PCR occurs, the allele-specific primer binds and elongates. This generates a first PCR strand that includes the allele-specific tail corresponding to the designated fluorescent dye. During subsequent PCR rounds, the complementary sequence to the fluorescent tagged oligonucleotide is generated, allowing it to anneal, and be separated from its quencher. Thus, this allows allele-specific fluorescence to be detected.

The fluorescent data is viewed graphically (see Figure 14 as example), and genotypes can then be determined by sample clusters.

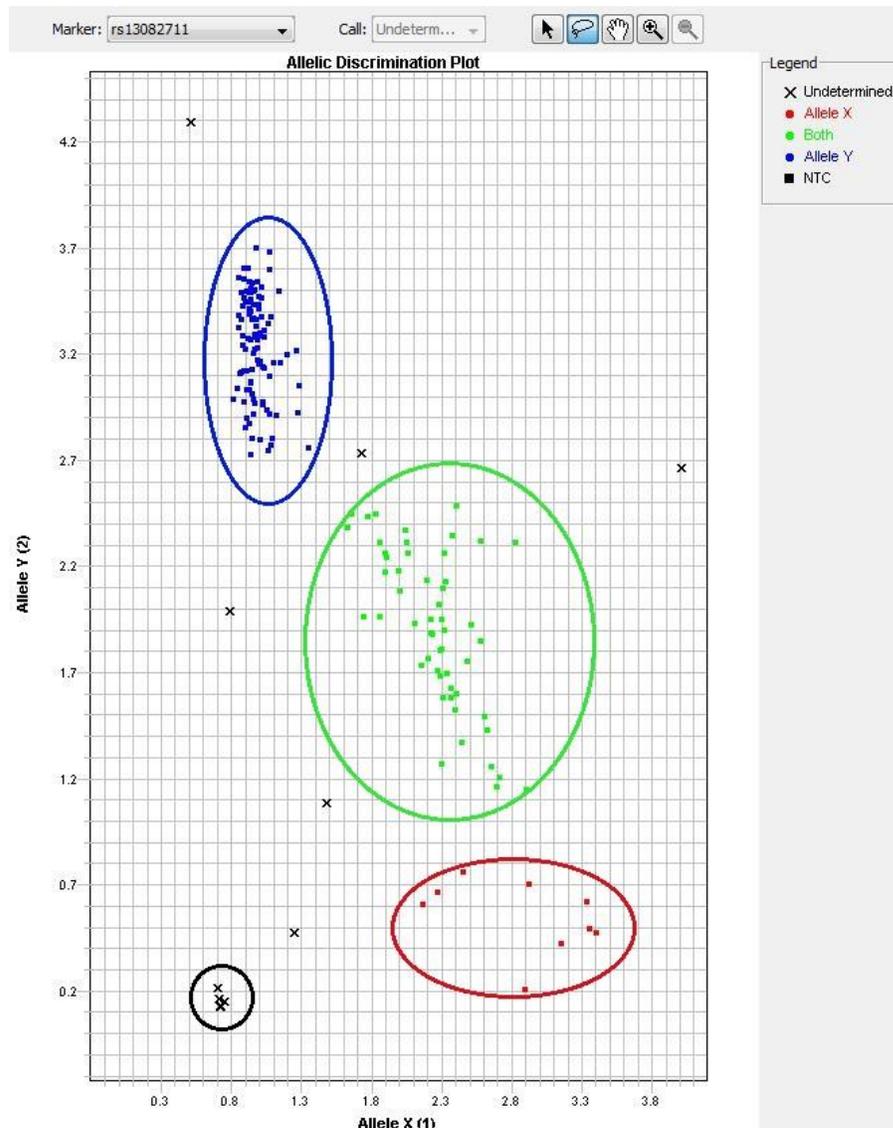


Figure 14: Representative figure of a KASP genotyping assay
 Distinct clusters are identified to indicate homozygotes (red and blue circles) and heterozygotes (green circle). No template controls are located at near the point of origin (black circle). Where present, samples not in clusters are also unclassified.

KASP (Kompetitive Allele Specific PCR) methodology – advantages and limitations

The main advantage of the KASP methodology compared to other genotyping methods is based on a lower cost-per unit, as well as the flexibility to conduct low- to high-throughput studies, particularly compared to restriction enzyme-based, or DNA sequencing. There is a marginal cost-benefit compared to other FRET-dependent technologies such as TaqMan® (Life Technologies).

Limitations are based on the principle that samples are assayed in bulk and results are dependent on the different fluorescence detected between the different samples. Should the templates be

non-uniform, or of insufficient quality, there is a potential for poor clustering on the fluorescence scatterplot. Additionally, there is anecdotal evidence based on the experience of the research team suggesting that some SNPs and their surrounding sequences are not suitable. The gold-standard currently applied to genotyping is that of DNA sequencing. As shown here, the KASP assays used have a high concordance with the gold standard, as well as inter-batch concordance (Results Section 4.1.).

2.4.3. Methods

At first, the DNA template was stamped into a 384-well plate. Each well was plated with either 5ng DNA or left empty as to be a designated NTC (no template control). As the “dry template” method was used, this was left in a 37°C oven for the liquid to evaporate overnight, leaving the DNA left at the bottom of the well.

The Reaction Mix was made by combining in a 1:63 ratio of Assay Mix (12 nM of each allele-specific primer and 30 nM of the common primer) and the Buffer Mix (KASPar master mix with added 0.15 mM MgCl₂ to achieve a final concentration of 1.8mM). 2.5 µl of this Reaction Mix was added to each well. The plate was then sealed with an optically-clear adhesive film to prevent evaporation. The reaction mixture was placed into a thermocycler and the program in Table 6 was used.

		Temperature	Duration
Initialisation step (Hot Start)		94°C	15 minutes
First 10 cycles	Denaturing step	94°C	20 seconds
	Annealing and elongation step	61°C to 55°C (dropping 0.6°C per cycle)	1 minute
Next 26 cycles	Denaturing step	94°C	20 seconds
	Annealing and elongation step	55°C	1 minute

Table 6: Thermocycler program for KASP assays

The plates were cooled to 4°C prior to reading on a FRET-capable plate reader (7900HT Fast Real-Time PCR System, Applied Biosystems). FAM (6-Carboxyfluorescein) and VIC (structure unknown, Applied Bioscience proprietary compound) probes were used for this thesis, and their excitation and emission wavelengths are listed below in

Table 7.

Fluorophore	Excitation (nm)	Emission (nm)
VIC	535	556
FAM	485	520

Table 7: Excitation and emission wavelengths for the utilised fluorophores

If further cycles were required to achieve tight clusters, a further 3 cycles of 94°C denaturing for 20 seconds and annealing/elongation for 1 minute at 57°C and cooled to 4°C prior to reading. This was repeated as necessary.

Analysis of results by SDS (Sequence Detection Systems, Applied Biosystems) software

The results from the plate reader was obtained in a *.sds file. This was analysed by the SDS v2.4.1 software (Applied Biosystems, USA). Manual calling was utilised to identify clusters of samples based on their florescence levels for the respective probes. An example is shown in Figure 14. A successful call rate at 95% and above was regarded as an adequate assay.

2.5. Polymerase chain reaction (PCR) and subsequent analyses

2.5.1. Principles - PCR

The polymerase chain reactions are able to amplify a single copy of a piece of DNA (or PCR amplicon) across several orders of magnitude, generating thousands to millions of copies of the original sequence (Saiki *et al.*, 1985). This method relies on an enzyme, DNA polymerase, which is able to extend nucleotide sequences with an open 3'-end with a complementary strand present. To successfully complete a reaction to copy, there are several steps, and different temperatures that are required. This means a thermal cycler is required to create cycles of repeated heating and cooling. The main components of the reaction are a DNA (or PCR amplicon) template, *Taq* polymerase enzyme, primers, deoxynucleoside triphosphates (dNTPs), divalent cations (generally Mg²⁺ as co-enzymes) and buffer solutions. Their relevance will be explained in each step.

Before the temperature of the reaction starts to cycle, some DNA polymerases require “hot-start” heat activation. In these instances, it undergoes an initialization step, with temperatures (usually 94°C to 96°C) and duration (usually 5 to 10 minutes) dependent on the enzyme.

In the first step (denaturing step), the two strands of the DNA double helix are physically separated at a high temperature in a process called DNA melting by disrupting the hydrogen bonds between the complementary bases, which results in two single-stranded DNA molecules. This is often set at a temperature greater than 90°C.

With the second step (annealing step), the temperature is lowered and the two DNA strands become templates for complementary primers to anneal to. The optimal temperature is dependent on the primer and complementary template. In sequences with higher GC (guanine or cytosine nucleotides) content have the potential to form stronger hydrogen bonds with the template, and this generally require higher annealing temperatures. The annealing temperature ranges from 50°C to 65°C, and is empirically determined (see Section 2.5.3), but is typically about 3–5 °C below the T_m (primer melting temperature) of the primers used. Stable DNA–DNA hydrogen bonds are formed when the primer sequence very closely matches the template sequence, and the DNA polymerase is able to combine with the primer-template interaction. Nowadays, almost all PCR amplifications utilises a heat-stable DNA polymerase (such as *Taq* polymerase, an enzyme that was initially isolated from the *Thermus aquaticus* bacteria, but is now typically bulk produced from pTaq cloned in *E. coli*).

During the third step (elongation step), the DNA polymerase synthesizes a new DNA strand complementary to the template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, by condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. This step is set at a temperature which is optimal for the DNA polymerase, typically between 70°C and 75°C. The buffer solution enhances the stability of the polymerase and provides a suitable chemical environment for optimum enzymatic activity. Magnesium acts as a co-factor for the polymerase enzyme, and increases the enzymatic activity, thus increasing the concentration increases sensitivity, but decreases specificity. The duration of this step is dependent on the predicted size of the PCR amplicon. As *Taq* polymerase extends by approximately 1,000 base pairs per minute, this is also the approximate time used.

The final elongation step is often performed, and is maintained at the same temperature as all previous elongation steps, but for a longer duration (usually between 5 to 15 minutes). This is to ensure that after the last PCR cycle, any remaining single-stranded DNA is fully extended.

As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. Typically, PCR protocols consist of a series of 25-40 repeated temperature cycles. Once again, this is established empirically for each reaction.

2.5.2. Principles - Primer design

A PCR primer is a single-stranded oligonucleotide that is able to hybridize with complementary stands of the DNA template, and thus identify the region to be copied. An appropriate pair of primers is used to amplify each region that it flanks. There are several considerations during primer design:

- Primer length between 18-30 bases
 - Longer primers do not necessarily improve specificity, but risks secondary structures and formation of primer dimers
- (Predicted) primer melting temperatures (T_m) between 55°-70°C
 - A simple prediction is calculated from $T_m = 4(G + C) + 2(A + T)$ °C
 - However, the actual T_m is more complicated. T_m is also influenced by the concentration of electrolytes within the reaction. There is a range of online tools that uses more complex algorithms, for example www.basic.northwestern.edu/biotools/oligocalc.html.
 - In practice, the range of acceptable T_m is broad enough that post-design optimisation is able to determine the optimal annealing temperature.
- (Predicted) T_m difference between forward and reverse primers of less than 5°C
 - Pairs of primers with a large difference in T_m may result in inadvertent preferential amplification of the most efficiently primed product strand, and lead to a reduction in yield or even no amplification.
- GC content of 50-60% and good 3' stability
 - In addition to the absolute number of G or C bases, the position also appears to matter. If present at the 3' end of primers (GC clamp), it helps to promote correct binding at the 3' end due to the stronger hydrogen bonding of G and C bases. However, when present as sequences in the middle of the primer may disrupt stable primer binding due to the higher probability of primer-dimer formation.

- No secondary structures
 - Nucleic acids naturally fold into secondary structures, particularly at lower temperatures. The stability of these template secondary structures depends largely on their free energy and melting temperatures. If these secondary structures in the oligonucleotides are stable even above the annealing temperatures, then the primers are unable to bind to the template DNA, and the yield of PCR product is significantly reduced. There are a range of online tools that are able to predict the secondary structures of primers, for example <http://www.premierbiosoft.com/primerdesign/index.html>.
- No self-complementarity
 - Self-complementarity can result in hairpin formation with just four GC base pairs in the stem and three bases in the loop. If primers form hairpins, they are not available for to anneal to their target regions. There are a range of online tools that is able to predict the secondary structures of primers, for example <http://www.premierbiosoft.com/primerdesign/index.html>.
- No complementarity to other primers in the reaction
 - Complementarity between the two primers, especially at the 3' ends, can lead to the formation of primer-dimers.
- No long runs with the same base of greater than 4 pairs
 - As this would otherwise lead to an increased risk of overall misalignment of primer and template.
- Depending on the polymerase enzyme, some are not able to reliably replicate beyond 2,000 base pairs.

There are several online programs to assist in primer design. This includes Primer3web (<http://primer3.ut.ee/>).

Whilst a 17-mer or longer primer should be complex enough so that the likelihood of annealing to sequences other than its chosen target is low it is important to ensure that portions of the primer do not have sequence or cross-homology with regions other than the desired target. Computer programs such as BLAST (Basic Local Alignment Search Tool, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) can be used to find any other regions of similarity within a genome. If there is significant homology with other sections of the genome, it increases the potential for non-specific amplification.

Primer design for RT-PCR

When designing primer pairs for RT-PCR, to amplify cDNA, at least one of the primers should span an intron-exon border or the pair to be placed in two different exons. This would allow differentiation of amplification of cDNA from that of genomic DNA (Table 8). Of course, this is not possible with intronless genes.

Template	cDNA	Genomic DNA
Both primers within one exon	Amplifies with expected amplicon size	Amplifies with expected amplicon size
One primer spanning intron-exon border	Amplifies with expected amplicon size	May amplify, depending on the 3'-end of the primer interaction with the template
Both primers on differing exons (spanning at least one intron)	Amplifies with expected amplicon size	Amplifies with larger than expected amplicon size (and might not amplify if too large an intron)

Table 8: Designing RT-PCR primers

2.5.3. Principles - Reaction optimization and controls

As laid out above, there are two variables for the polymerase chain reaction that are often determined empirically – annealing temperature and Mg^{2+} concentration. The range of annealing temperatures is often $4^{\circ}C$ either side of the lower T_m out of the two primers. The range for Mg^{2+} concentration is typically between 1mM to 4mM, with starting experiments typically at 2mM. Lowering the annealing temperature or increasing the Mg^{2+} concentration increases the sensitivity of the reaction, but increases the chances of non-specific products. Conversely, to increase the specificity of the reaction (for example when there are non-specific products already present), the annealing temperature can be increased or the Mg^{2+} concentration reduced – but potentially reducing yield of the desired amplicon.

In certain circumstances, nested PCR can be conducted to reduce non-specific products or increase sensitivity in reactions with a low yield. This procedure involves two sets of primers, used in successive polymerase chain reactions. The second pair of primers is designed to amplify a region within the PCR amplicon generated by the first pair.

Polymerase chain reactions should be conducted with:

- positive controls (known samples that produce an amplicon with the reaction) whereby an absence of an amplicon indicates a faulty reaction,
- negative controls (no template in the reaction) whereby any detected amplicon indicates contamination of the reaction, and
- in the case of RT-PCR, an RT-negative control (where the RNA is processed without the reverse transcriptase reaction) whereby any detected amplicon suggests contamination of the RNA sample.

2.5.4. Principles - Agarose gel electrophoresis

Detection of PCR products is typically conducted with agarose gel electrophoresis made in Tris-Acetate-EDTA (TAE) or Tris-Borate-EDTA (TBE) buffers.

The quantity (w/v percentage) of agarose determines the pore size to sieve the electrophoresed PCR products, with the higher percentage resulting in smaller pores, and better separation over smaller amplicon sizes. The 2% w/v agarose gels used for this thesis have a good separation of products between 100 and 1,200 bp. To separate larger products, such as in the cloning / mutagenesis process, 0.7% w/v gels were used.

TAE and TBE buffers are chosen as these solutions are slightly basic, keeping the PCR amplicons deprotonated and soluble in water. Additionally, EDTA is a chelator of divalent cations, particularly Mg^{2+} which is also a co-factor for many enzymes, including contaminant nucleases, thus providing protection against enzymatic degradation.

The samples are loaded near the cathode end of the gel. With the net negative charge of the sugar-phosphate backbone of the PCR amplicons, when an electric field is introduced, it induces the nucleic acids to migrate towards the anode. The separation of these fragments is dependent on the mobility of different sized molecules passing through the gel. Longer molecules experience more resistance within the gel, thus they migrate more slowly across the gel, and the converse is true for smaller molecules. The gel is run until the bromophenol front, which approximates to the mobility of a 100 bp PCR product, is three-quarters down the gel.

There are several nucleic acid stains that aid detection. Previously, ethidium bromide was commonly used as it intercalates between double stranded DNA, and fluoresces under ultraviolet light. However, due to its mechanism of action, it deforms DNA and affects DNA replication, thus potentially mutagenesis. Newer and larger molecules such as GelRed™ and SYBR® Green work similarly, are more sensitive in detecting double stranded DNA, and are marketed as being less toxic by virtue of being less likely to enter cells. To approximate the sizes of PCR amplicons, a DNA ladder of known sizes is electrophoresed in parallel to the samples.

2.5.5. Principles - Isolation of PCR products

If the PCR amplicons are required for downstream processes, such as sequencing, enzymatic processing, cloning and microarray analyses, they need to be purified of unincorporated primers, left-over dNTPs, salts, and enzymes. This can be performed straight from the PCR product, or from excised bands of an agarose gel.

If the sample is being extracted from an agarose gel, it is initially melted before it is bound on a silica membrane within a spin column. The bound DNA is washed with an ethanol-based solution and the cleaned DNA can be eluted in the buffer of choice, such as nuclease-free water, TAE or TBE. Most manufacturers of commercial columns claim that up to 95% of PCR products, between 100 bp and 10 kb, can be reclaimed in this process. Furthermore, they would claim that all dNTPs and salts, more than 99% of the primers and most primer-dimers are removed.

2.5.6. Methods

Template

The PCR template could be genomic DNA, complementary DNA (cDNA) as derived from reverse transcription of RNA, formaldehyde-derived chromatin-associated DNA, or PCR amplicons for nested PCRs.

Polymerase chain reaction

The reaction mix was kept at 0°C using an ice bath to inhibit any polymerase activity. The final concentration of the reagents within the reaction include 400 µM primers, 80 µM dNTPs, variable Mg²⁺ (see Table 21 for reaction specific concentration), 1x reaction buffer and 2.5 units of Taq polymerase (Sigma, D1806). The final volume of the reaction was made up to 25 µl using nuclease-free water. The reaction was placed into a thermocycler, and the following program was used (

Table 9). After the reaction, the samples were held at 4°C until agarose gel electrophoresis.

		Temperature	Duration
Initialisation step (Hot Start)		95°C	5 minutes
35 cycles	Denaturing step	95°C	1 minute
	Annealing step	Variable (see Section 2.21.2 for primer pair-specific conditions)	1 minute
	Elongation step	72°C	30 seconds
Final elongation step		72°C	10 minutes

Table 9: Thermocycler program for end-point PCR

Agarose gel electrophoresis

2% w/v agarose powder was suspended in 1x TBE buffer (89 mM Tris-base, 89 mM boric acid, 2 mM EDTA pH 8.0; where the 10x preparation is adjusted to pH 8.3), and brought to the boil to dissolve the agarose. This was let to cool at room temperature, and before the gel forms, GelRed™ (Biotium, 41003) was added to achieve a final 1x concentration. The resultant mixture was poured into horizontal gel casts with the appropriate comb in place to create the sample wells. The gel is suitable for use once it solidifies and reaches room temperature.

Just before electrophoresis, 6x DNA Loading Buffer (60% v/v glycerol, 60 mM Tris-HCl pH 8.0, 6 mM EDTA pH 8.0, 0.06% w/v bromophenol blue) was added to the PCR products to achieve a final concentration of 1x. Electrophoresis was conducted in a horizontal electrophoresis tank filled with 1x TBE. The samples are loaded in the wells placed at the cathode end of the tank. The quantity loaded is dependent on the size of the wells created. A 100 bp DNA ladder (New England Biolabs, N3231L) was electrophoresed in parallel.

Once the bromophenol blue front reaches approximately three-quarters, the gel was visualised by an ultraviolet transilluminator which also captures a digital image (GelDoc-It Imager, Ultra-Violet Products Ltd., UK). In semi-quantitative analyses, the intensities of the signals were quantified by the public domain, Java-based image processing program ImageJ.

PCR product isolation

Isolation of PCR products was conducted using the Wizard® SV Gel and PCR Clean-Up System (Promega, A9281), and the protocol is summarised as follows.

When isolating PCR amplicons from an excised gel piece, Membrane Binding Solution was added to the gel slice in a microcentrifuge at a ratio of 1 µl for every 1 mg of agarose gel. This was then incubated in a water bath equilibrated to 50°C to 65°C until it dissolved. When isolating PCR amplicons from the reaction mixture, add an equal volume of Membrane Binding Solution to the PCR amplification.

For both sample types, the mixture can now be loaded into the SV mini-column assembly. This was then centrifuged for 1 minute at 13,000g at room temperature. The flow-through was discarded, and 700 µl of Membrane Wash Solution was added. This was then centrifuged for 1 minute at 13,000g at room temperature. The flow-through is again discarded, and a second wash of 500 µl of Membrane Wash Solution was added to the mini-column assembly. This was then centrifuged for 5 minute at 13,000g at room temperature. The flow-through was discarded again, and the mini-column was re-centrifuged to allow evaporation of any residual ethanol.

Once the residual ethanol has evaporated, the PCR products was eluted by adding 50 µl of nuclease-free water to the mini-column, and incubated at room temperature for 1 minute. The eluate was collected by centrifugation of the mini-column for 1 minute at 13,000g at room temperature. The concentration of the eluate was quantified by spectrophotometer before downstream processes.

2.6. Allelic expression imbalance analysis

2.6.1. Principles

The allelic expression imbalance analysis method used in this study utilises sequencing chromatographs of genomic DNA and cDNA of heterozygotes at the SNP of interest to assess any allele-dependent levels of RNA (i.e. cDNA). The relative chromatograph peak heights at the SNP could be a good indicator of relative expression levels of each allele. However, due to the nature of the Sanger sequencing method, the peak heights are not just dependent on other quantity of the allele present, but also the position of the SNP and the surrounding nucleotides. With this, the genomic DNA is used as a control.

This method has an advantage of being internally controlled, where both alleles are exposed to the same intracellular environment, unlike qPCR (Section 2.7). On the other hand, it is not able to easily differentiate between isotypes or comparisons across different cell types.

Sanger sequencing

Sanger sequencing utilises the chain-termination method, and requires a single-stranded DNA (or PCR amplicon) template, a primer, DNA polymerase, deoxynucleoside triphosphates (dNTPs), and modified di-deoxynucleoside triphosphates (ddNTPs). The template, primer, polymerase and dNTPs work similarly to that in a standard PCR with cycling temperatures. However, the ddNTPs is the key variation, as once incorporated, they terminate elongation of the chain. They have chain-terminating properties as these nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond. The ddNTPs may be radioactively or fluorescently labeled for detection in automated sequencing machines.

The processed sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and the DNA polymerase. To each reaction *only one* of the four dideoxynucleosides (ddATP, ddGTP, ddCTP, or ddTTP) is added. Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat denatured and separated by size. The relative positions of the different bands among the four lanes, from bottom to top, are then used to read the DNA sequence. In the past, separation may be conducted on a denaturing polyacrylamide-urea gel, and radiolabelled-ddNTPs detected by autoradiography. With modern DNA sequencers typically utilise capillary electrophoresis for size separation and detection is based on fluorescent dye-labelled ddNTPs. The data output is usually shown as fluorescent peak trace chromatograms.

Allelic imbalance analysis by PeakPicker

As peak-heights along a chromatogram vary depending on sample, base type and position, the absolute values are not fully representative of the relative abundance of any one allele. With this, the PeakPicker software (McGill University and Génome Québec Innovation Centre, California; previously available at <http://genomequebec.mcgill.ca/EST-HapMap>) was used to determine relative allele ratios from paired genomic DNA and cDNA samples, as first published in 2005 (Ge *et al.*, 2005). For each sequencing chromatogram, the peak-heights of the polymorphism of interest

were normalized to the peak-heights of the adjacent non-polymorphic positions. The software utilises the same adjacent positions across the whole dataset, analyzed simultaneously. The input files for PeakPicker are the raw sequence files from ABI (Applied Biosciences, USA) sequencers (*.ab1 file), whilst the output of this software are normalized peak-heights (arbitrary units) for each allele at the polymorphism.

Utilizing validated software such as PeakPicker allows for more uniform quantification and removes potential of observer bias. However, it should be recognised that these normalized peak-heights is at most semi-quantitative as it is not linearly correlated with nucleic acid copy numbers. Additionally, the source code for the PeakPicker is not publically available, leaving the result with reduced transparency.

Complementary methods

Another method of determining allele-associated mRNA expression levels is to compare the qPCR C_t values (number of PCR cycles to reached a defined level of signal) of different genotypes. A proposed benefit of qPCR over allelic imbalance analyses is that it allows comparisons across the three genotypes, potentially determining if there is an additive, or perhaps a dominant, model present. Unlike allelic imbalance, qPCR also allows comparison of haplotypes.

However, this is balanced against the experimental context of allelic imbalance comparisons. In this setting, the cellular environment in a heterozygotic cell line is the same for both alleles. This reduces the significant variability and potentially confounding factors that would otherwise cloud the overall picture of qPCR results. This also provides the logistical advantage of being able to conduct paired statistical analyses, which in general reduces the number of experiments required to achieve a certain level of statistical power.

2.6.2. Methods

Sanger sequencing

Sanger sequencing was conducted by an in-house commercial service at the Genome Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ. Samples (greater than 10 μ l, typically at 10 ng/ μ l) together with primers (2 μ l at 10 mM) were delivered to the service at 4°C. Results were returned as an .ab1 file with the chromatograph raw data.

Allelic imbalance analysis by PickPicker

The PeakPicker software (McGill University and Génome Québec Innovation Centre, California; previously available at <http://genomequebec.mcgill.ca/EST-HapMap>) was used to determine relative allele ratios from heterozygotes of interest. Once the raw sequencing data was available, it is uploaded into the program (Figure 15). This was done after all the samples to be analyzed are available for uniformity of analysis.

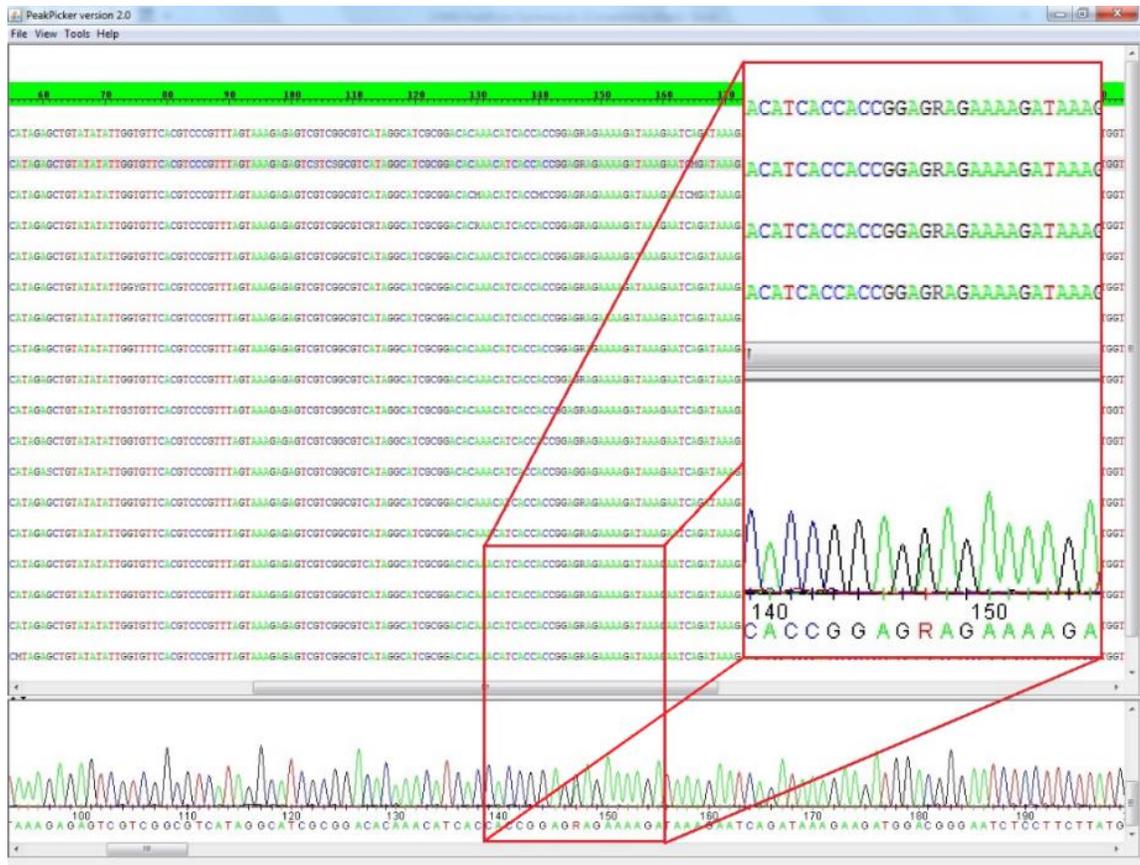


Figure 15: PeakPicker analysis of chromatographs conducted in parallel. Each row of nucleotide sequence represents one uploaded sample. The chromatograph at the bottom is that set as “reference sequence”.

After uploading the sequences, the SNP of interest is selected, and the program was prompted to automatically select surrounding peaks of the same nucleotides to allow for normalisation to the peak-heights of the adjacent non-polymorphic positions (Figure 16).

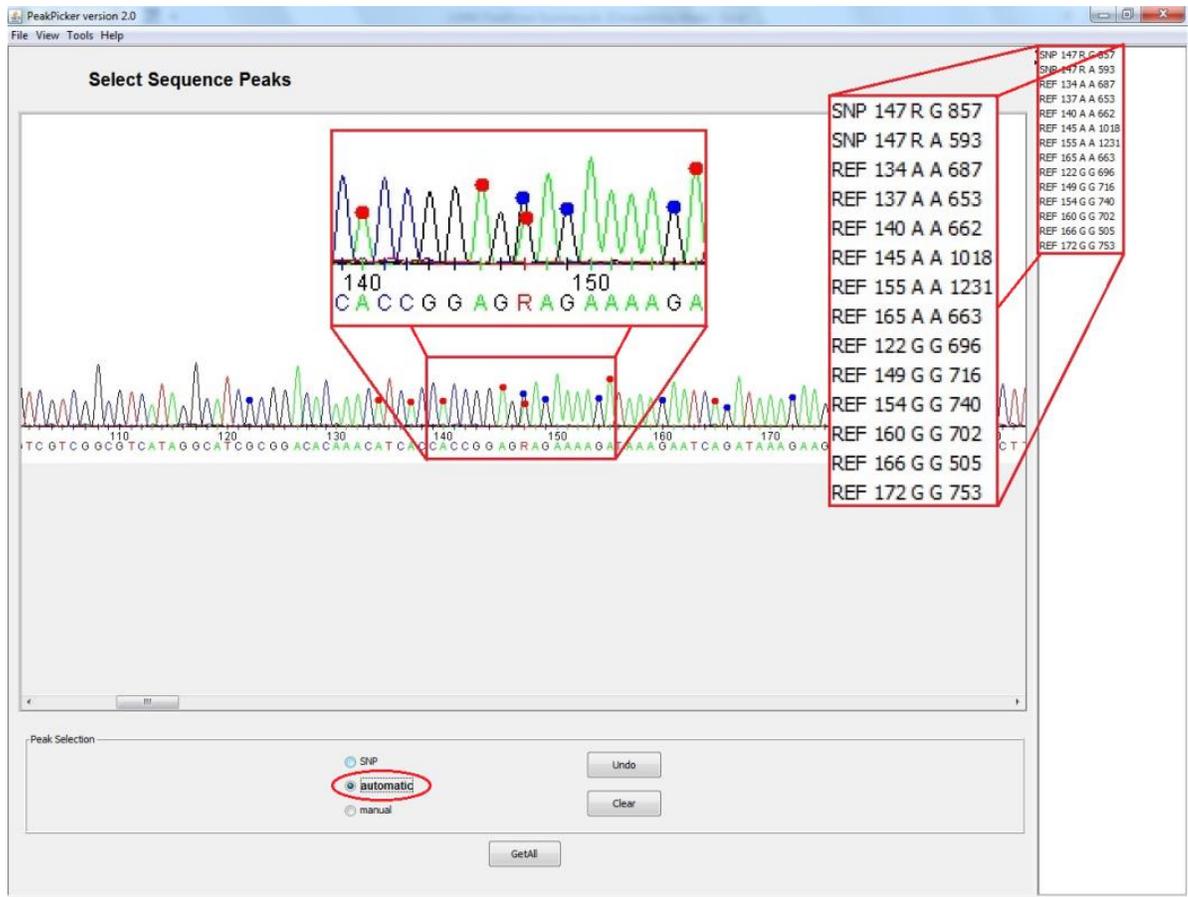


Figure 16: PeakPicker analysis of the “reference sequence”
 Red and blue dots indicate selection of the SNP of interest. The values in the box highlighted on the right represent the measure peak-heights at the specified positions.

After the program conducts the analysis, the normalised peak-heights (arbitrary units) are listed for each sample (Figure 17).

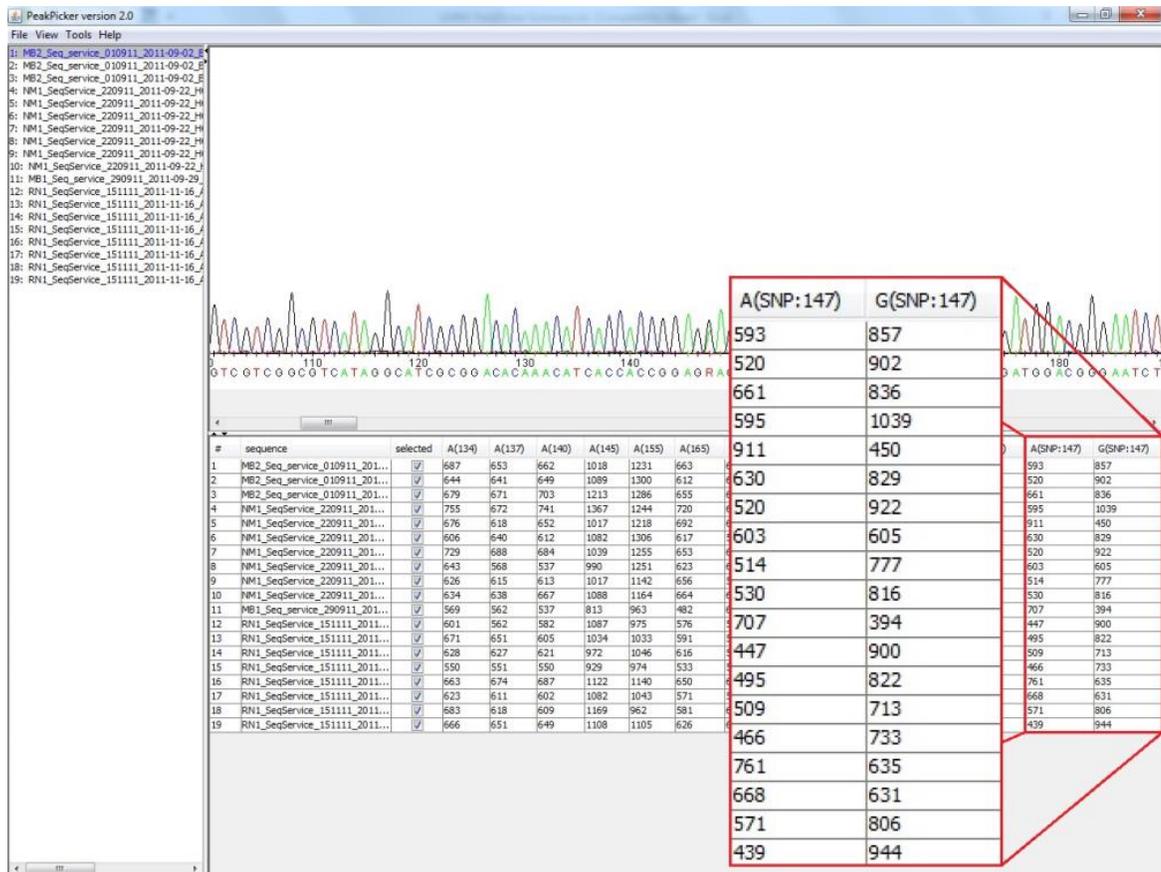


Figure 17: PeakPicker output of normalised peak-heights
 The highlighted box lists the normalised peak-heights (arbitrary units) is listed for each sample.

2.7. qRT-PCR (quantitative reverse transcription-polymerase chain reaction)

2.7.1. Principles - History

Before the age of the automated real-time PCR system, attempts were made to provide quantitation using end-point PCR and ethidium bromide staining, for which one of the first reports is that from Wang and colleagues (Wang *et al.*, 1989). Then, the target mRNA is co-amplified with a known internal standard in one reaction. A standard curve could be generated with varying the number of cycles, or serial-diluting the templates. The slope of the curves, as compared to the known internal standard would provide the relative quantification. The real-time PCR systems that we know now bear significant resemblance to this precursor technique, except for:

- replacing the ethidium bromide staining with either fluorescent-quencher tagged probes or PCR-compatible nuclei acid stains, and

- replacing the standard curve with an automated process with fluorescence readings at each cycle, and detecting a threshold fluorescent signal is reported, instead of a manually generated standard curve.

Each of these factors will be explored below.

2.7.2. Principles - Detection systems

Intercalators

Initially techniques for measuring real-time PCR products utilised intercalator dyes. These dyes (such as SYBR® Green I) behave similarly to ethidium bromide and GelRed™, but instead of binding the major or minor grooves of double-stranded DNA, these dyes intercalate with DNA instead. The key difference between dyes like SYBR® Green I, and other intercalators is that they do not inhibit polymerase chain reactions. These intercalators also have an increased fluorescence when bound to double-stranded DNA. As the PCR progresses, more amplicons are created which allows more SYBR Green I dye to bind double-stranded DNA, and results in an increased fluorescence intensity proportional to the amount of PCR product produced.

An advantage with the intercalator system is that multiple dye molecules bind to a single PCR amplicon, increasing the sensitivity for detecting amplification products. On the other hand, this multiple dye binding means the amount of signal is dependent on the mass of double-stranded DNA produced in the reaction. Thus, if the amplification efficiencies are the same, amplification of a longer product will generate more signal than a shorter one. A further advantage is that it is relatively cheaper than fluorogenic probe-based methods (see below).

The primary disadvantage of the SYBR Green I dye chemistry is that it binds to any double-stranded DNA. With this, it can also bind to nonspecific double-stranded DNA or oligonucleotide sequences, including primer dimers, resulting in either an inflated quantification, or in some cases, a false positive. This can be limited by initially conducting end-point PCR reactions to confirm specificity of the PCR primers, and melting curve analyses.

Fluorogenic-labelled probes

Another common method uses fluorogenic-labeled probes and the 5' nuclease activity of Taq DNA polymerase. An example of this is the TaqMan® probe-based expression assays. It requires two primers, and an oligonucleotide probe containing a reporter fluorescent dye on the 5' end and a

quencher dye on the 3' end. While the probe is intact, the proximity of the quencher dye greatly reduces the fluorescence emitted by the reporter dye by fluorescence resonance energy transfer through space. As the PCR continues, the Taq DNA polymerase cleaves the probe, and separates the reporter dye from the quencher dye, increasing the reporter dye signal. As this also removes the probe from the target strand, primer extension continues to the end of the template strand without inhibition of the overall PCR process. With each further cycle, more probes are cleaved, producing an increased fluorescent signal.

The advantage of the fluorogenic-labelled probes centers around the increased specificity, as the probe only binds a specific sequence. Additionally, the probes can be labelled with different reporter dyes allowing the possibility of duplex studies, allowing an internal control. Finally, unlike that of intercalator dyes, the one probe to one amplicon ratio means that there is no dependence on amplicon size for the overall signal, allowing for more flexibility when designing primers and probes.

The primary disadvantage of the TaqMan[®] probe-based chemistry is that the synthesis of different probes is required for different sequences. The requirement for a short amplicon (80 to 150 bp) to fit three binding sequences (a ~20 bp forward primer, a ~20 bp reverse primer and a ~30 bp probe) makes primer design in a predefined area difficult. Although usually not an issue, the ratio of one fluorescent probe to one amplicon means the fluorescent signal may not be as greatly amplified as intercalators.

Passive reference dyes

Passive reference dyes are commonly used in qPCR to normalize for non-PCR related fluorescence signal variation. As the passive reference dyes do not take part in the reaction, the detectable fluorescence remains constant throughout the numerous PCR cycles. This provides an internal reference to which the reporter dye signal can be normalized. This normalization is typically automatically considered with newer real-time PCR systems, and is taken as a ratio ($R_n = \text{Reporter fluorescence} / \text{Passive Reference fluorescence}$). The use of a passive reference is best for normalizing well-to-well optical variations caused by bubbles. It also may help, to a very limited degree, with pipetting errors and changes in reaction volumes.

Signal detection

The automated real-time PCR systems now allow a high-throughput, sensitive and accurate gene quantitation assay. The assay uses fluorescent signals, either by fluorescent DNA intercalators or fluorophore-tagged probes (described above), and an instrument capable of measuring fluorescence in real-time.

During each extension step (when either the DNA intercalator is able to bind double-stranded amplicons or fluorescent-probes have already been cleaved), the fluorescence value of the PCR is assessed and plotted in an amplification plot as exemplified in Figure 18 (Heid *et al.*, 1996). The reaction initially demonstrates a baseline phase, with minimal change in detectable fluorescence signal. At a certain number of amplicon copies, the signal rapidly increases into the exponential phase, then the linear phase and finally (after a large number of cycles), the plateau phase.

The threshold value is calculated by the analysis software as a set number of standard deviations (usually 10) above the mean of the baseline emissions calculated from the initial cycles (usually 1 to 15). Once the threshold is chosen, the point at which the amplification plot crosses the threshold is defined as C_t (threshold cycle). C_t is reported as the cycle number at this point, although not necessarily as an integer

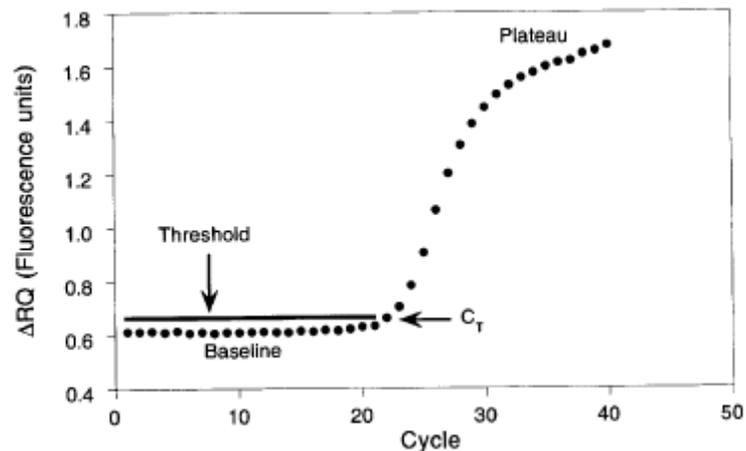


Figure 18: qPCR amplification plot

Example amplification plot (Heid *et al.*, 1996). ΔRQ (also termed ΔRn , or Delta Rn) is the normalization of ratio of the fluorescence emission intensity of the reporter dye to the fluorescence emission intensity of the passive reference dye obtained by subtracting the baseline.

As identified by various reports, the C_t value is predictive of the quantity of input target (Figure 19, Heid *et al.*, 1996). Depending on the PCR efficiency, each decrease in C_t of one cycle equates to an x-fold increase in template quantity. In a theoretical 100% efficient PCR system, this equates to a doubling of the template.

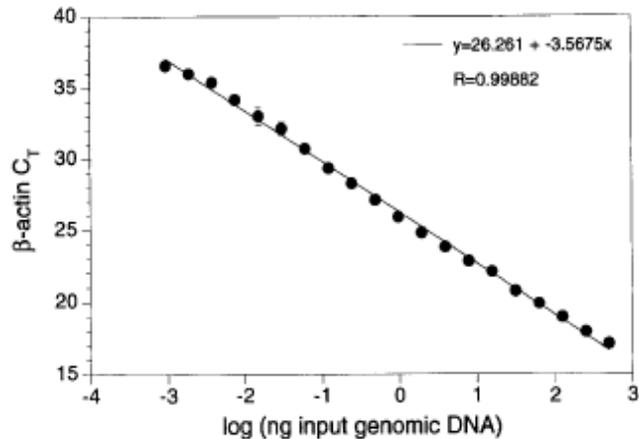


Figure 19: qPCR log-linear relationship

Example of the log-linear relationship between input template (x-axis) and threshold cycle (y-axis) (Heid *et al.*, 1996).

As the work in this thesis concentrates on relative differences in gene expression, the above relative quantification is applicable. *Absolute* quantification adds a further layer of complexity of relating the PCR signal to a standard curve, and is neither used, nor further discussed in this thesis.

2.7.3. Principles - Analysis

One way of analyzing and presenting relative quantification qPCR data is the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001), and refers to the change in expression of a target gene in a sample (or population) relative to another. This method utilises an internal control gene to normalize for the amount of RNA/cDNA added to the reaction. Standard housekeeping genes such as GAPDH, β -actin and 18S rRNA are often used. *Theoretically*, a difference of $\Delta\Delta C_t$ of x equates to a 2^x -fold difference in signal level, of course, assuming 100% PCR efficiency which is not typical.

2.7.4. Methods

Primer design

A key reason for conducting qRT-PCR studies, in addition to supporting the initial findings of the allelic imbalance experiments on allele-dependent mRNA levels, is to assess any potential differences in *SLC4A7* isoforms based on genotype. Original intentions to use a duplex design were excluded as the positioning of probes and primers would lead to potential duplicated and thus non-specific signals. The eventual primer design utilised a common forward primer. The amplicons of the reaction detecting either isoform were similar in size (near 100 bp) to aid comparable PCR efficiency (Figure 20).

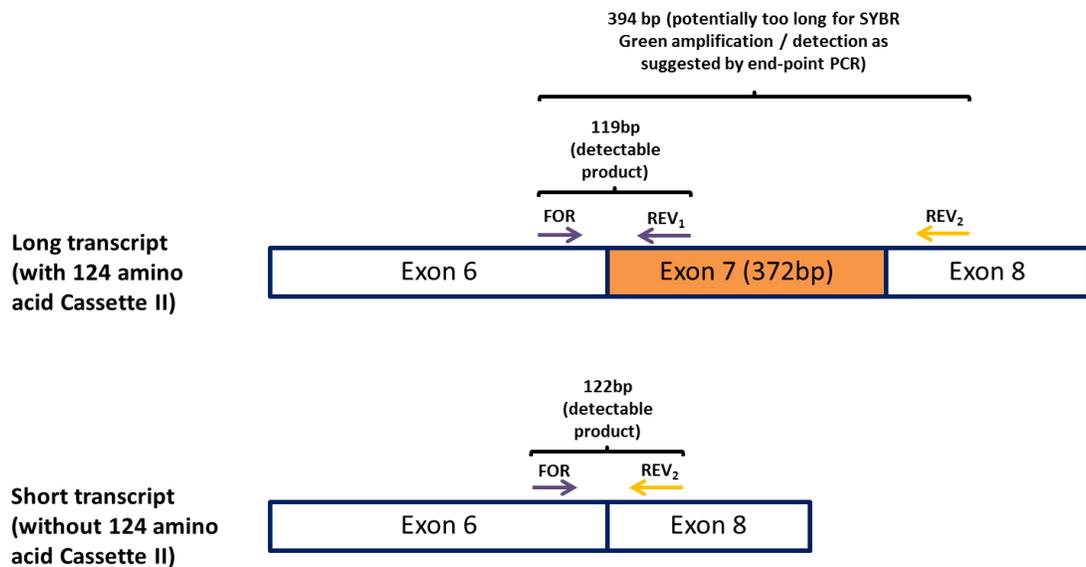


Figure 20: Schematic of primers used for SYBR[®] Green qPCR. To detect the longer transcript (above), only the FOR and REV₁ primers are used, and the position of REV₂ is illustrated only to show the relative sizes of PCR amplicons. For the shorter *SLC4A7* transcript (below), only the FOR and REV₂ primers are used.

End-point PCR using these pairs of primers confirmed that there were no non-specific signals, no primer-dimers, and the absence of the larger (394 bp) product when using FOR and REV₂ primers (Figure 21). This indicates the two different pairing of primers is appropriate to specifically amplify the sequences of interest, and produce only little, if at all, non-specific signals.

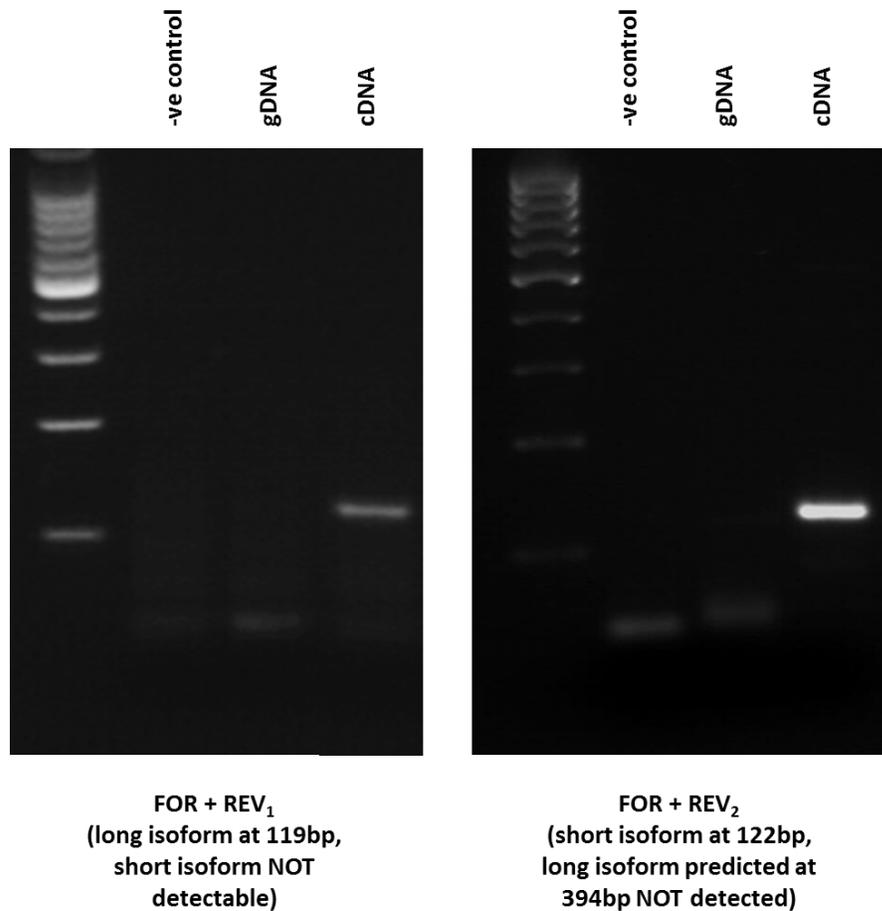


Figure 21: Endpoint PCR with primers used for SYBR[®] Green qPCR. Endpoint PCR with primer pairs detecting the variants with (left) and without (right) the 372 amino acid exon 7 (equating to the 124 amino acid Cassette II). There were neither non-specific products nor primer-dimers present.

If required, magnesium and primer concentrations could be varied to optimize the signal while eliminating non-specific products as well as primer-dimers. However, the starting concentrations of 3 mM Mg²⁺ and 400 nM of each primer produced the appropriate results on end-point PCR, and thus did not require further optimization.

Quantitative polymerase chain reaction (qPCR)

The qPCR was conducted in 384-well plates with the SYBR[®] Green system (Applied Biosystems, 4309155) with 5 µl for each reaction. For each reaction, 2.5 µl of 2x SYBR[®] Green PCR Master Mix, 0.2 µl of each forward and reverse (10 mM) primer to reach a final concentration of 400 nM each and 0.1 µl of nuclease-free water is added to 2 µl of 2.5 ng/µl cDNA template. Note that the 2x

SYBR[®] Green PCR Master Mix already contains 6 mM Mg²⁺, to reach a final concentration of 3 mM in each reaction.

All reactions were conducted in duplicate, together with duplicates of the housekeeping gene (18S ribosomal RNA).

The 384-well plate was placed into a real-time PCR system (Applied Biosystems 7900HT Fast Real-Time PCR System) and the thermocycler program in Table 10 was used.

		Temperature	Duration
Initialisation step (Hot Start)		95°C	10 minutes
40 cycles	Denaturing step	95°C	15 seconds
	Annealing and elongation step	60°C	1 minute

Table 10: Real-time PCR program for qPCR

Control / housekeeping gene expression

18S ribosomal RNA was used as the control housekeeping gene to normalize for RNA/cDNA quantity in the reaction. The pre-designed TaqMan probe and primer set (4333760F) was used according to the manufacturer's protocol.

Analysis of results by SDS (Sequence Detection Systems, Applied Biosystems) software

The results from the plate reader was obtained in a *.sds file. This was analyzed by the SDS v2.4.1 software (Applied Biosystems, USA). C_t values were automatically calculated based on population-dependent threshold levels.

2.8. Protein isolation

2.8.1. Principles

Methods and components of lysis buffers for protein isolation depend ultimately on the end use. Isolation is generally conducted at 4°C and with protease inhibitors to reduce sample degradation.

(a) Total cellular protein isolation for SDS-PAGE

Isolation of total cellular protein is predominantly based on surfactants. Surfactants are amphiphilic, which is to say that they have both hydrophilic and hydrophobic groups. This allows them to disrupt lipid membrane layers, thus freeing up intracellular and membrane-bound proteins. Typical surfactants used are SDS (sodium dodecyl sulphate), Triton X-100, CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), NP-40 (nonyl phenoxyethoxyethanol) or Nonidet P-40 (octyl phenoxyethoxyethanol). Mechanical disruption with sonication, dounce homogenizers, vortex mixers or syringe-and-needle methods could also be further used.

(b) Protein isolation for co-immunoprecipitation

The basis of co-immunoprecipitation requires the retention of protein-protein interactions. Thus, lysis buffers containing milder detergents such as NP-40 or Nonidet P-40 were utilised as they have a lower propensity to interrupt protein-protein interactions. Unlike total cellular protein isolation, mechanical disruption is not utilised as it may promote disruption of protein-protein interactions, particularly sonication. Alternatively, other protocols utilise formaldehyde to form protein-protein crosslinks, which is maintained during, and reversed after immunoprecipitation by heating.

(c) Nuclear extract isolation for electrophoretic mobility shift assays (EMSA)

Isolation of nuclear extracts occurs in two stages with first the cytosolic lysis buffer followed by a nuclear lysis buffer. Initially, a cytosolic extract is derived when the plasma membrane is lysed without lysing the nuclear membrane. The cytosolic lysis buffer is often hypotonic, resulting in cellular swelling and disruption of the plasma membrane but not the nuclear membrane. Milder detergents such as NP-40 or Nonidet P-40 is utilised as they have an effect on the plasma membrane and not nuclear membrane. After the use of the cytosolic lysis buffer, the sample undergoes centrifugation separating out the cytosolic fraction (supernatant) and nuclei (pellet). The resultant pellet is lysed with a non-ionic, non-denaturing nuclear lysis buffer. The end-use in EMSAs require the proteins remain in native conformation, and this the need for non-denaturing conditions.

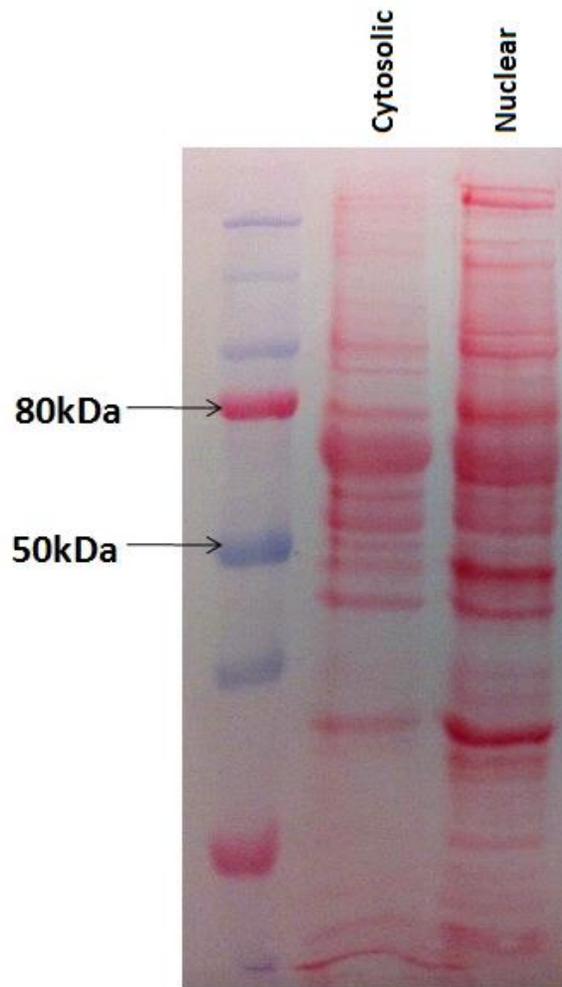


Figure 22: Fractionation of cytosolic and nuclear extracts
Ponceau S staining of cytosolic and nuclear extracts (both 20 μg per lane) showing different banding patterns.

(d) Subcellular protein fractionation

Isolation of membrane proteins takes advantage of the differences in the sedimentation coefficient of various cellular compartments. Typically, the larger and denser the particle, the higher the sedimentation coefficient, and in practical terms, forms pellets with centrifugation at lower speeds.

Initially, the cells are lysed with a non-ionic, hypotonic lysis buffer. The non-ionic nature of the lysis buffer results in maintained lipid bilayers, thus nuclei, large complexes of cytoskeleton, mitochondria, lysosomes, peroxisomes, vesicles, and importantly, large fragments of the plasma membrane, remains intact. The cell homogenate is centrifuged at progressively higher speeds to result in different cell fractions.

A first low speed centrifugation (approximately 1,000g) pellets out whole cells, nuclei and large insoluble proteins complexes. A further centrifugation at medium speeds (approximately 10,000g) pellets out mitochondria, lysosomes and peroxisomes. A final ultracentrifugation at fast speeds (approximately 100,000g) pellets out vesicles and the cellular lipid membrane. The resultant supernatant contains the soluble cytosolic proteins (Figure 23).

A key limitation to this method is that the 10,000g fraction contains the whole mixture of cellular organelles. If there is an interesting signal in this fraction, it is not possible to determine which subcellular location it is derived from. Alternatives, or more precisely complementary methods, would include immunocytochemistry (Section 2.16) and flow cytometry, although they would also lack the specificity of identifying specific organelles, and suffers complications in quantification.

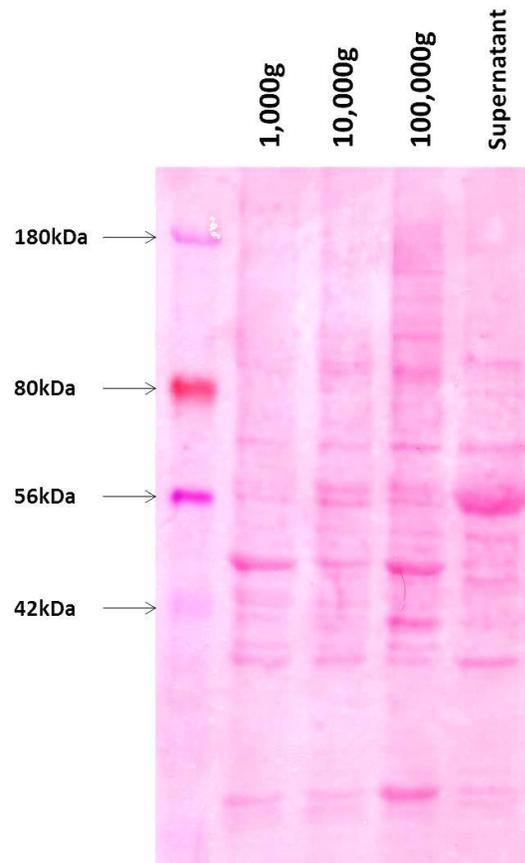


Figure 23: Differential fractionation of cellular proteins
Representative Ponceau S staining of subcellular fractionation by progressively higher speed centrifugation showing different banding of proteins for each fraction. 20 ng was loaded onto each lane.

(e) Protease inhibitors

Proteases are found in cells, and thus also present in cell lysates, independent of lysis method. There are broadly six groups of proteases – serine proteases, threonine proteases, cysteine proteases, aspartate proteases, glutamic acid proteases and metalloproteases – for which the names identify the mechanism to cleave the peptide bond. Metalloproteases are typically inhibited by the EDTA within most lysis buffers, as it chelates the necessary metal ions. With the remaining proteases, there are a variety of protease inhibitors with varying specificities for different protease classes. With this, usually a cocktail of different inhibitors are required.

Particular care should be taken when handling and storing protease inhibitors, as several (such as phenylmethylsulfonyl fluoride) have a short half-life in aqueous solution and are even shorter with temperatures above 4°C. With this, the cocktails should be kept at -20°C before use.

2.8.2. Methods

(a) Total cellular protein isolation for SDS-PAGE

Cells from a confluent T-25 flask (25 cm² surface area) with an approximate cell count of 2.5 million HUVECs or HUASMCs were used for whole-cell protein extraction. The cell culture media is removed from the T-25 flask, and subsequently the adherent cells were washed twice with ice-cold PBS. Following removal of the PBS wash, 200 µl of RIPA buffer (50 mM Tris pH7.4, 150 mM NaCl, 0.5% w/v sodium deoxycholate, 1% v/v NP-40, 0.1% w/v sodium dodecylsulfate) with protease inhibitor mixture was added. The volume of RIPA buffer was scaled-up or –down dependent on the tissue culture surface area. Single-use cell lifters were used to detach the adherent cells from the cell culture surface. The resultant cell suspension was transferred to a 2 ml microcentrifuge tube. This was rested on ice for 30 minutes with cell disruption using a needle and syringe. Following cell lysis, the sample was centrifuged at 13,000g for 10 minutes at 4°C. The supernatant was kept as the cell lysate, and the cell debris pellet discarded.

(b) Protein isolation for co-immunoprecipitation

Cells from a confluent T-75 flask (75 cm² surface area) with an approximate cell count of 7.5 million HUASMCs were used for co-immunoprecipitation. The cell culture media was removed from the T-75 flask, and subsequently the adherent cells were washed twice with ice-cold PBS. Following removal of the PBS wash, 500 µl of IP lysis buffer (25 mM Tris HCl pH 7.2, 150 mM NaCl, 1 mM EDTA pH 8.0, 1% v/v NP-40), with protease inhibitor mixture was added. Single-use cell

lifters were used to detach the adherent cells from the cell culture surface. The resultant cell suspension was transferred to a 2 ml microcentrifuge tube. This was agitated on a roller shaker for 30 minutes in the cold room (4°C). Following cell lysis, the sample was centrifuged at 13,000g for 10 minutes at 4°C.

(c) Nuclear extract isolation for electrophoretic mobility shift assays (EMSA)

Cells from a confluent T-75 flask (75 cm² surface area) with an approximate cell count of 7.5 million HUVECs or HUASMCs were used for nuclear extract isolation. The cell culture media is removed from the T-75 flask, and subsequently the adherent cells are washed twice with ice-cold PBS. Following removal of the PBS wash, 500 µl of hypotonic lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.004% v/v NP-40), with protease inhibitor mixture was added. The volume of lysis buffer was scaled-up or -down dependent on the tissue culture surface area. Single-use cell lifters were used to detach the adherent cells from the cell culture surface.

The resultant cell suspension was transferred to a 2 ml microcentrifuge tube. This was agitated on a roller shaker for 30 minutes in the cold room (4°C). Following cell membrane disruption, the sample was centrifuged at 13,000g for 3 minutes at 4°C. The supernatant was kept as the cytoplasmic fraction, and the cell debris pellet contains unlysed nuclei.

The nuclei pellet is lysed with the nuclear lysis buffer (20 mM HEPES pH 7.9, 400 mM NaCl, 1 mM EDTA, 10% v/v glycerol), with protease inhibitor mixture. The nuclei were disrupted using a needle and syringe, and placed roller shaker for 2 hours in the cold room (4°C). Then, the sample was centrifuged at 13,000g for 5 minutes at 4°C. The supernatant nuclear fraction was retained, and the insoluble debris is discarded.

(d) Subcellular protein fractionation

Cells from a confluent T-75 flask (75 cm² surface area) with an approximate cell count of 7.5 million were used for membrane protein preparations. The cell culture media was removed from the T-75 flask, and subsequently the adherent cells are washed twice with ice-cold PBS. Following removal of the PBS wash, 500 µl of homogenization buffer (10 mM Tris HCl pH 7.2, 1 mM EDTA pH 8.0, 250 mM sucrose), with protease inhibitor mixture was added. Single-use cell lifters were used to detach the adherent cells from the cell culture surface.

The cell homogenates was transferred to a microcentrifuge tube and centrifuged at 900g for 10 minutes at 4°C. The pellet is kept as the nuclear (and cell debris) fraction and dissolved in 50 µl of RIPA buffer (50 mM Tris pH7.4, 150 mM NaCl, 0.5% w/v sodium deoxycholate, 1% v/v NP-40, 0.1% w/v sodium dodecylsulfate) with protease inhibitor mixture.

The subsequent supernatant was transferred into a fresh microcentrifuge tube and centrifuged at 10,000g for 5 minutes at 4°C. The pellet is kept as the mitochondrial (and lysosomes / peroxisomes) fraction and dissolved in 50 µl of RIPA buffer with protease inhibitors.

The subsequent supernatant was again transferred into a fresh microcentrifuge tube and centrifuged at 100,000g for 60 minutes at 4°C. The pellet is kept as the membrane fraction and dissolved in 50 µl of RIPA buffer with protease inhibitors. The supernatant is kept as the cytoplasmic fraction.

If the cytoplasmic fraction was not of a sufficient concentration, it was concentrated using the acetone precipitation method (see Section 2.11.2), and later dissolved in 50 µl of RIPA buffer with protease inhibitors.

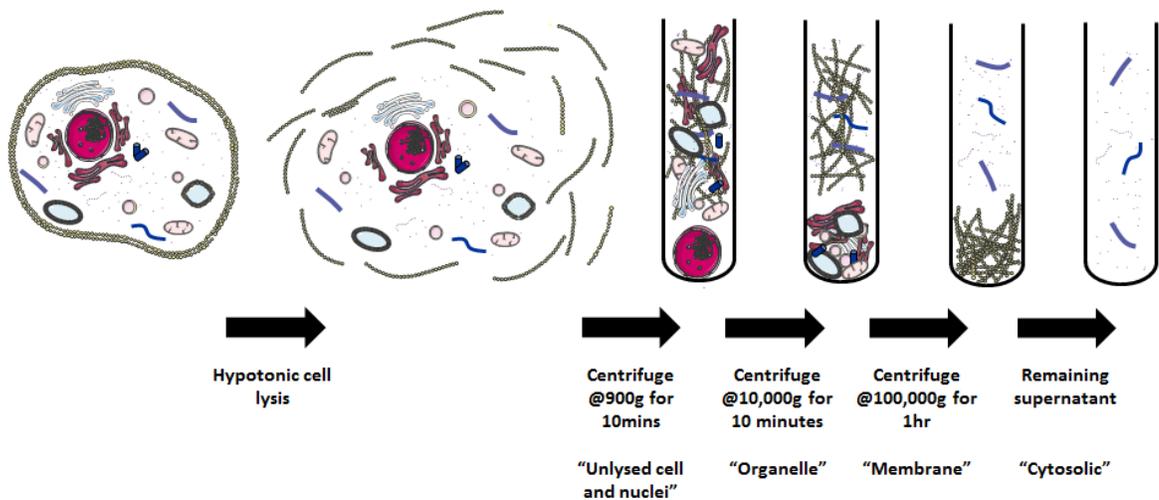


Figure 24: Cartoon of subcellular protein fractionation by differential centrifugation
 By sequentially increasing the centrifugational speed, the proteins are separated by their sedimentation coefficient. The large nuclei are the first to be isolated, followed by organelles and then membrane-bound proteins. Being very small, free cytosolic proteins are not able to be isolated at these centrifugational speeds. This method was utilized to assess the presence of NBCn1 that is localized to the plasma membrane.

(e) Protease inhibitors

Protease inhibitors were made into individual stocks of 100x stock solution of 100mM phenylmethanesulfonyl fluoride (PMSF) (Sigma, P7626), 200 μ M leupeptin hemisulfate (Sigma, L2884), 150 μ M pepstatin A (Sigma, P2465) and 15 μ M aprotinin (Sigma, A1153). When preparing for use, equivolumines of each protease inhibitor was combined for a 25x stock concentration. This was used in a 24:1 (sample : protease inhibitor) ratio to achieve the final concentration of 1x.

2.9. Bicinchoninic acid assay for sample protein quantification

2.9.1. Principles

Bicinchoninic acid (BCA) assays are based on alkaline solutions of copper salts. Under alkaline conditions, cupric ions (Cu^{2+}) chelate with the peptide bonds, and this reduces the cupric ion to cuprous ions (Cu^+). The cuprous ions are then able to chelate with bicinchoninic acid forming a bicinchoninic acid-cuprous ion complex that absorbs light at a wavelength of 562nm. The formation of the cuprous ions and the subsequent complex with BCA is proportional to the amount of protein, and particularly influenced by asparagine, histidine, cysteine, tyrosine, and tryptophan side chains due to their reducing potential (Smith, 1985). These reactions are also temperature-dependent, and are conducted between 37°C and 60°C to increase the sensitivity of the assay. BCA assays are routinely conducted with samples of known protein concentrations, forming standard curves.

When using this assay, caution should be exercised to ensure the components of the lysis buffer itself does not react with the assay. A list of interfering reagents is available with most commercial kits.

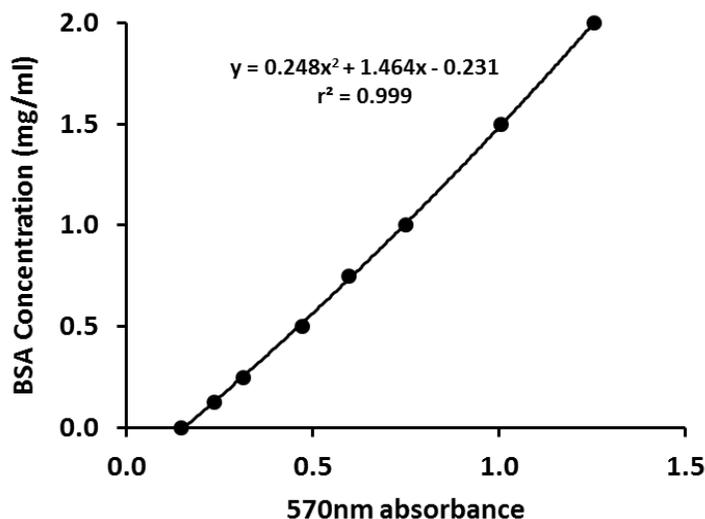
2.9.2. Methods

Protein quantification was conducted using the bicinchoninic acid (BCA) assay. The protocol for the commercial kit (Pierce, 23225) was used and summarised as follows.

For every reaction, 200 μ l of Reagent A was mixed with 4 μ l of Reagent B to form the working reagent. 200 μ l of this mixture was added to 10 μ l of protein sample (or standard / blank) in 96-well plates. The resultant mixture was mixed on a plate shaker for 30 seconds, and then incubated for 30 minutes at 37°C. Following incubation, the plate was cooled to room temperature and the

absorbance read at 570nm on a microplate spectrophotometer (MRX-TC Revelation, Dynex Technologies). Although the optimal absorbance is 562nm, the kit manufacturers noted that a wavelength between 540nm to 590nm is acceptable.

A standard curve is prepared of 0 (blank), 0.125, 0.25, 0.5, 0.75, 1, 1.5 and 2 mg/ml bovine serum albumin (BSA) diluted in the respective lysis buffer was prepared in parallel with the samples. A quadratic best-fit line was plotted with BSA standards with known concentrations to provide a reference for the samples with unknown concentrations (Figure 25).



*Figure 25: Representative bicinchoninic acid assay standard curve
Absorbance plotted on x-axis and protein standard BSA (bovine serum albumin) plotted on the y-axis to allow computerized generation of quadratic best-fit line (as recommended by manufacturers).*

2.10. SDS-PAGE (Sodium dodecylsulfate polyacrylamide gel electrophoresis) and immunoblotting (Western Blotting)

2.10.1. Principles

SDS-PAGE electrophoresis

SDS-PAGE electrophoresis was used for a variety of experiments, including progression to Western protein immunoblotting, as well as separation of proteins in DNA-pull down assays (Section 2.11). This method separates proteins from a mixture based on their molecular weight. There has been many different developments to this methodology, but most of the credit is attributed to the

Burnette, Stark and Towbin groups in the late 1970s and early 1980s (Towbin *et al.*, 1979; Renart *et al.*, 1979; Burnette, 1981).

This method utilises protein samples that are denatured with heat and the surfactant sodium dodecylsulfate (SDS) and a reducing agent such as 2-betamercaptoethanol or dithiothreitol (DTT). The SDS coats the linearised protein, resulting in a negatively charge, overcoming the native charge of the protein. Denatured protein samples were loaded onto polyacrylamide gels, and an electrical current applied. The negatively-charged, SDS-coated proteins (anions) would migrate towards the positive (anode) electrode.

The polymerised acrylamide within the SDS-PAGE gel forms a meshwork for which the proteins would migrate across. The higher the percentage (volume per volume) of acrylamide results in increased cross-linking and denser meshwork – allowing greater resolution between smaller proteins. Conversely, the lower the acrylamide content allows greater resolution between larger proteins. For most SDS-PAGE gels in this thesis, either 6% or 8% separating gels were used depending on sizes of the proteins of interest. Mobility across the gel is dependent on the conformation, size and charge of the molecule. However, due to the denaturing and linearization of the protein together with the SDS coat, the effect of native conformation and charge no longer is a significant factor (See Section 2.14. for comparison with native gels). This allows the separation whereby the smaller proteins, with a larger charge-to-mass-ratio, would move faster than the larger proteins.

The discontinuous Laemmli buffer system employs different buffer ions and pH in the gel and in the electrode reservoirs. Samples are loaded onto a “stacking” non-restrictive large pore gel (lower pH of 6.8 and lower polyacrylamide concentration) which overlays a smaller pore “resolving” gel (higher pH of 8.3 to 9.0 and higher polyacrylamide concentration). This allows stacking of proteins at the interphase before final separation, so that relatively large volumes of dilute protein samples can be applied with greater resolution.

As the protein itself within the samples is not visible on a SDS-PAGE gel, the rate of mobility is monitored by commercial pre-stained protein standards, marking out a range of protein sizes. Additionally, the loading buffer for the protein samples contains bromophenol blue which

maintains its blue colour in the alkaline environment of the gel, and travels in a front that approximates to 5 to 10 kDa.

After electrophoretic separation, the proteins are electrotransferred onto polyvinylidene fluoride (PVDF) membranes, which are hydrophobic and binds protein with high affinity. This membrane was preferred to the alternative, nitrocellulose, due to the superior binding capacity, particularly if SDS is used in the transfer buffer. PVDF membranes also have the additional benefit for being able to be processed in downstream applications such as protein sequencing. The buffers during electrotransfer need to be kept cool in face of the current that is applied through it, to prevent buffer breakdown and the subsequent loss of transfer efficiency.

Following electrotransfer, the membrane can proceed to be used for immunoblotting, or Ponceau S staining for DNA-pulldown studies.

Immunoblotting

Immunoblotting allows the identification of specific (or as specific as the antibody allows – see later) proteins electrotransferred onto membranes. It typically utilises signal amplification via secondary antibodies and a further enzymatic step before signal detection.

Once the proteins are electrotransferred to the PVDF or nitrocellulose membrane, a “blocking” step is often conducted. This typically utilises non-fat milk available commercially from food stores, or purified proteins (such as bovine serum albumin). Application of these proteins occupies the regions of the membrane that has not had sample proteins transferred over from the gel. If this is not performed, the primary antibody used in later steps may bind the membrane itself (independent of the target protein of interest) resulting in increased background signal.

Once “blocked”, the membrane is incubated with a primary antibody that has been raised against the protein of interest. Most antibodies however are raised against an epitope, typically a shorter sequence of amino acids from the protein of interest. This leads to the possibility of non-specific signals where other proteins may mimic the epitope, thus also binding the primary antibody. This incubation step is typically conducted overnight at 4°C, but may also be at room temperature over a shorter duration. Where indicated, the addition of the immunising epitope to the primary

antibody step can compete for binding of the antibody, hence reducing or removing the signal, thus identifying “specific” bands.

To amplify the signal, a secondary antibody is utilised. The secondary antibody is raised to target a portion of the primary antibody, typically the F_c region (fragment crystallisable region, or more colloquially known as the tail of the antibody). This is usually a conserved sequence for any particular species, as opposed to the F_{ab} (antigen-binding fragments) region, which allows consistent and cross-assay use. The signal amplification occurs as for every one primary antibody bound to the membrane, more (often estimated to be in the region of 10 to 20) secondary antibodies can bind the protein-primary complex.

The secondary antibody is usually labelled with an enzyme such as HRP (horseradish peroxidase) or alkaline phosphatase, although fluorescent-labelled secondary antibodies are also occasionally used for protein immunoblots. Labelling with an enzyme further amplifies the signal, as the bound enzymes are able to catalyse large quantities of substrate to the signal-generating products. This thesis utilises enhanced chemiluminescence with the horseradish peroxidase / luminol / hydrogen peroxide / p-coumaric acid combination.

Enhanced chemiluminescence

When the luminol is exposed to an activating oxidant, in this case hydrogen peroxide, it can undergo chemical reactions via several different intermediates to emit a low intensity blue glow. This reaction is catalysed by the horseradish peroxidase attached to the secondary antibodies, allowing specificity of the signal. The p-coumaric acid serves as the enhancer for the reaction, improving the light signal.

This end-product of signal amplification allows for detection of very low protein abundance in samples. However, the light emission has a short half-life (in the order of minutes), and so needs to be exposed to an autoradiography film within the appropriate time window.

Ponceau S staining

Ponceau S is part of a family of azo dyes. It is a reversible colloidal stain that can be used on PVDF, nitrocellulose and cellulose acetate membranes. Ponceau S has an overall negative charge, and binds to both the positive and non-polar amino acids, with a lower detection limit of 250 ng

protein. Thus, protein-bound sections of the membrane are coloured pink/red, and sections without protein unstained. This allows visualisation of efficiency of the electrotransfer, as well as identifying any specific bands of proteins in the sample. Staining can be reversed by mild alkaline or water washes without detriment to subsequent processes such as immunoblotting or mass spectrometry.

Other stains are also applicable, such as colloidal metals (typically gold or silver) or other diazo dyes (e.g. amido black). These methods are typically more sensitive but with slightly more complex experimental protocols or unlike Ponceau S, or is not reversible to a state of use in downstream studies.

Advantages and limitations of protein immunoblotting

Protein immunoblots are well-established in protein studies. It enables identification of proteins within a sample with relative ease. It also allows semi-quantitative assessment of relative amounts of the protein of interest present in different samples.

It is however limited by the sensitivity and specificity of the antibody. Whilst conducting a competitive blockade of the primary antibody using the immunising epitope can somewhat help identify immunoblot bands specific to the protein of interest, it still may identify other proteins containing similar epitopes.

Additionally, like most assays that have an output dependent on signal intensity on an autoradiograph, with a non-linear relationship with the quantity of protein present, the method of quantifying protein immunoblots is only semi-quantitative.

Another drawback is the limited number of samples that can be assessed concurrently. The inherent variability in the assay output signal, dependent on the efficiencies of the electrotransfer, primary and secondary antibodies, and the enhanced chemiluminescence reaction, meaning that comparing across two blots / autoradiographs is technically challenging.

2.10.2. Methods

Whole-cell protein extraction

See Section 2.8.2

Protein quantification

See Section 2.9.2

Preparation of protein samples for SDS-PAGE electrophoresis

Following quantification of the protein samples, 20 µg of protein was used and diluted with the above RIPA / PI mixture to 20 µl, followed by 5 µl of a 5x loading buffer (200 mM Tris pH 6.8, 10% w/v sodium dodecylsulfate, 20% v/v glycerol, 5% v/v beta-mercaptoethanol, 0.05% w/v bromophenol blue). This mixture was heated to 80°C for 2 minutes to denature the proteins for the work in this thesis, although a higher temperature and longer duration may be required for other proteins. These samples were then loaded into the wells of SDS-PAGE gels.

Casting of SDS-PAGE gels

Polyacrylamide gels were made based on the Laemmli discontinuous buffer system consisting of a lower separating gel and an upper stacking gel. The content of the SDS-PAGE gels were dependent on the size of the protein of interest. For the work in this thesis, separating gels of 8% or 10% acrylamide was used. The volume prepared to cast one separating gel of 8.6 cm width x 6.7 cm height x 0.1 cm thickness was 6 ml, using the Mini-PROTEAN Casting System (Biorad, 165-8000).

The formulation for separating gels included the variable w/v percentage of acrylamide: (w/v) bis-Acrylamide (37.5:1 stock) to suit the gel, 0.375 M Tris pH 8.8, 0.1% w/v sodium dodecyl sulfate (SDS), 0.1% w/v ammonium persulfate (APS) and 0.1% w/v tetramethylethylenediamine (TEMED).

Once the separating gel mixture was placed into the casting system, 1 ml of butanol was added atop the separating gel to ensure a level gel. They were set aside for the acrylamide to polymerise for 30 minutes. Following this, the 2 ml per stacking gel (formulation below) was added to the formed separating gel and the plastic well combs were fixed in place. The formulation for separating gels included the variable w/v percentage of acrylamide: (w/v) bis-Acrylamide (37.5:1 stock) to suit the gel, 0.375 M Tris pH 6.8, 0.1% w/v sodium dodecyl sulfate (SDS), 0.1% w/v ammonium persulfate (APS) and 0.1% w/v tetramethylethylenediamine (TEMED). These gels were cast up to 48 hours before use and kept moist with distilled-deionised water at 4°C. Alternatively, pre-cast gels can be purchased commercially.

Separation of proteins with SDS-PAGE electrophoresis

The gel casts were assembled into the Mini-PROTEAN Tetra Cell system (Biorad, #165-8000), and filled with 1x running buffer (0.1% w/v sodium dodecyl sulphate, 25 mM Tris-base, 192 mM glycine). Electrophoresis was conducted initially at 80 V, until the bromophenol blue front is compressed at the stacking-separating gel interface. At that point, electrophoresis was conducted at the higher voltage of 140V. This continued until the appropriate separation of pre-coloured standard proteins is visualised, usually lasted approximately 100 minutes.

Electrotransfer of proteins from SDS-PAGE gel to PVDF membrane

Following completion of the electrophoresis above, the SDS-PAGE gel was carefully lifted off the glass casting system, and opposed against a methanol-activated PVDF membrane. This was sandwiched by a single Whatman® qualitative Grade 1 filter paper on either side, and further sandwiched by manufacturer-provided sponges. This construct was placed in the mini-gel holder cassette, and placed within the Mini Trans-Blot module. Electrophoresis using Towbin transfer buffer (20% v/v methanol, 25 mM Tris-base, 192 mM glycine) was conducted at 60 V for 150 minutes at 4°C. Successful transfer was visualised with the transfer of the pre-coloured standard protein ladder onto the PVDF membrane.

Immunoblotting

The PVDF membrane was placed in a plate shaker incubated with blocking buffer comprising 5% (w/v) non-fat milk (Marvel, Premier Foods) in TBS-Tween (150 mM NaCl, 50 mM Tris adjusted to pH 7.4, 0.5% w/v Tween) for 1 hour at room temperature to block potential non-specific binding of antibodies. In experiments with high background signal, such as immunoblotting for calcineurin A, the alternative blocking buffer comprising 3% w/v BSA (Sigma, A9418) in TBS-Tween was used. Following blocking, the membrane is incubated with the primary antibody (varying dilutions in the blocking buffer, see Section 2.22) with agitation, overnight at 4°C. This was then washed with TBS-Tween thrice for 10 minutes each on a plate shaker for gentle agitation. After washing, the membrane was incubated with the corresponding HRP (horseradish peroxidase)-labelled secondary antibody (varying dilutions in blocking buffer, see Section 2.22) with agitation, for 1 hour at room temperature. This was again washed with TBS-Tween thrice for 10 minutes each on a plate shaker for gentle agitation in preparation for chemiluminescence detection.

Enhanced chemiluminescence

The PVDF membrane with the protein - primary antibody – HRP-labelled secondary antibody is suitable for enhanced chemiluminescence detection. One part of luminol / p-coumaric acid solution (2.5 mM luminol, 0.4 mM p-coumaric acid, 0.1 M Tris pH 8.5) and one part stabilised hydrogen peroxide (0.02% hydrogen peroxide, 0.1 M Tris pH 8.5) is mixed, and 2 ml is applied to sufficiently cover 100 cm² (10 cm x 10 cm membrane). Following 5 minutes incubation at room temperature, the membrane is exposed to an autoradiograph film (CL-XPosure Film, Thermoscientific, 34089) in a dark room, and developed (Konica-Minolta). The developed film is later scanned (Hewlett-Packard) for digital storage and quantification.

Analysis

In semi-quantitative analyses, the intensities of the autoradiograph signals were quantified by the public domain, Java-based image processing program ImageJ.

Ponceau S staining

Ponceau S is a sodium salt diazo dye that stains protein bound sections of the membrane red. It has the added advantage of being reversible with no significant deleterious effect on downstream processes. Once a PVDF (or nitrocellulose) membrane was ready for staining, it was placed within the staining solution (0.1% w/v Ponceau S, 40% v/v isopropanol, 10% v/v acetic acid), and gently agitated for 10 minutes on a plate shaker at room temperature. The staining solution was discarded, and the membrane washed with de-staining solution (45% v/v isopropanol, 10% v/v acetic acid) for 5 minutes per wash. The washes were repeated as necessary until the protein bands are clear. To fully reverse the staining, the membrane was washed in distilled-deionised water, and repeated as necessary. This is mainly an optional step for confirmation of adequate protein transfer, but occasionally for specifically comparing protein bands in different samples.

2.11. DNA pull-down assays

2.11.1. Principles

DNA pull-down assays can be utilised to determine *in vitro* DNA-nuclear protein interactions, as summarized in Figure 26. It relies on a labelled sequence of double-stranded oligonucleotides or polymerase chain reaction amplicons. The label is often a high affinity tag such as biotin that forms a strong interaction when introduced to streptavidin or avidin.

The labelled probe is incubated with nuclear proteins, either in the form of crude whole nuclear extracts or specific (often recombinant) proteins. In the case of biotin-labelled probes, a streptavidin coated bead is used to provide a heavier anchor to allow separation from unbound proteins. The beads are often agarose bead-based which requires slow centrifugation, or magnetic bead-based, which allows separation without centrifugation.

Once the bead - streptavidin - biotin - DNA probe - nuclear protein complex is isolated, it undergoes several washes to reduce non-specific binding. The proteins are then eluted with heating. As the protein of interest may be low in abundance, it may be necessary to concentrate the eluate.

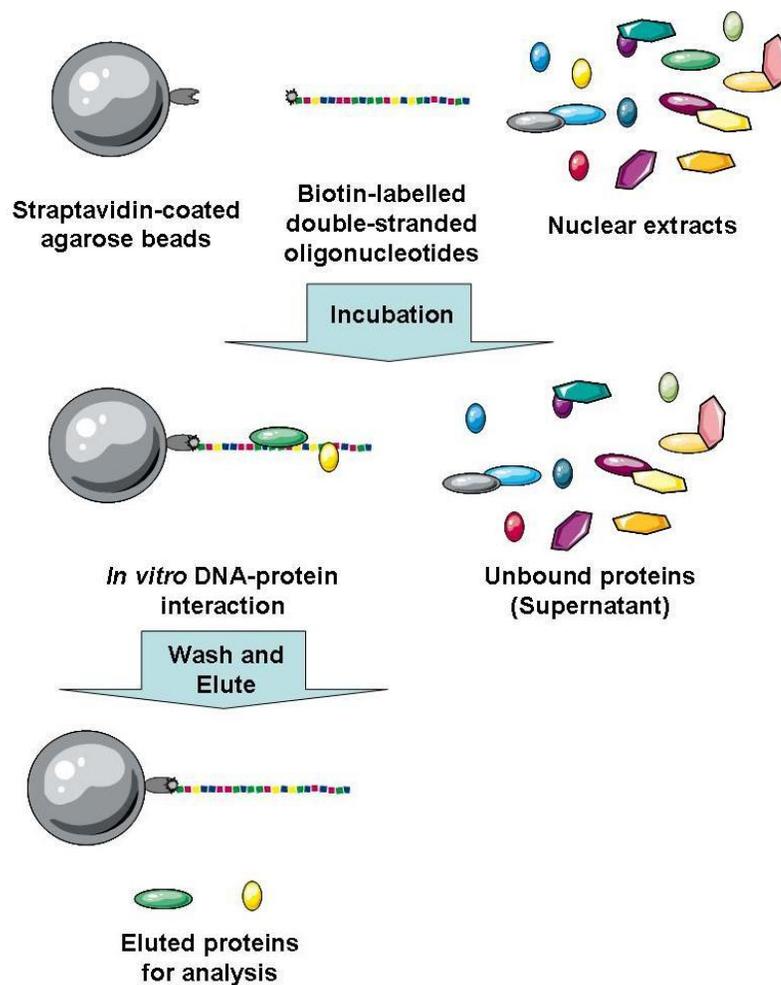


Figure 26: Cartoon outlining the DNA-pull down process
Initially the streptavidin-coated agarose beads, biotin-labelled double-stranded oligonucleotides and nuclear extracts are incubated together to allow for the relevant DNA-nuclear protein to occur. Subsequently, the beads are isolated by centrifugation and the unbound proteins removed. Finally, the agarose-oligonucleotide-protein complex is heated to release the proteins for analysis.

These samples are then suitable for separation using SDS-PAGE or 2D-electrophoresis, and further identification.

Advantages, limitations and complementary methodologies

Like most other studies in determining DNA-nuclear protein interactions, there are significant advantages and limitations to each method. The key advantage of DNA-pulldown assays are based on the ability to enrich low abundance targets and the ability to isolate intact DNA-nuclear protein complexes that are compatible with downstream processes such as immunoblotting, amino acid sequencing and mass-spectrometry.

A major limitation is that the assay is conducted in an artificial *in vitro* environment which may not necessarily mimic that of an intracellular or nuclear milieu, notwithstanding the added complexities of epigenomic control. An additional concern is that DNA probes may be prone to non-specific binding, particularly with longer probes. This may result in false-positives, thus the requirement for complementary methods.

Whilst DNA-pulldown assays can indicate whether a DNA sequence may interact with nuclear proteins *in vitro*, results need to be concordant with other assays. One such method is the chromatin immunoprecipitation (ChIP) assay, which utilises the crosslinking of the DNA-nuclear protein complex in chromatin, and antibodies to proteins of interest to pull-down DNA fragments which the protein is bound to. ChIP has the additional advantage of the DNA-nuclear protein interaction occurring within a cellular system as opposed to DNA-pulldown, occurring purely within an experimental context.

Concentrating proteins by acetone precipitation

Acetone precipitation, like most other methods of protein precipitation utilising organic solvents such as trichloroacetic acid (TCA), is based on hydrophobic aggregation within a solution. Acetone is able to dehydrate the shells of water on a protein in solution. This produces hydrophobic surfaces on proteins, whereby if they are present in sufficient numbers, aggregates can start to form. As this method relies on the proteins “finding each other”, it is also a time-dependent process. These aggregates are protein precipitates that can be separated from solution by centrifugation at a high speed.

Acetone, like TCA, is also a protein denaturant at room temperature. Thus the process is conducted on ice, or kept at -80°C when incubating. Unlike TCA, acetone is not an acid, and thus does not require further steps of acetone washes or pH titration when the protein precipitate is dissolved in the relevant SDS-PAGE buffer.

SDS-PAGE electrophoresis

See Section 2.10.1

N-terminal amino acid sequencing

N-terminal sequencing uses a chemical process based on the technique initially developed by Pehr Edman. During the chemical reaction, the N-terminal amino acid reacts with phenylisothiocyanate (PITC), this derivatization process results in a phenylthiohydantoin (PTH) - amino acid link. This amino acid is then sequentially removed while the rest of the peptide chain remains intact. Each derivatization process is a cycle, and each cycle removes a new amino acid. The amino acids are sequentially analysed by mass spectrometry to give the sequence of the protein or peptide (Jacobs and Niall, 1975). Samples that are suitable for this method of analysis include PVDF-membrane bound gel electrophoresis separated proteins, high-performance liquid chromatograph-separated tryptic digest fragments or purified proteins. A typical successful analysis requires 100 picomoles of purified protein.

However, the N-terminus is reported to be “blocked”, by formyl, acetyl or pyroglutamyl groups in approximately 60 to 70% of eukaryotic proteins. These N-terminal modifications prevents the derivatization process, and thus does not produce a successful sequencing analysis. These blocks can also occur during isolation of the protein.

2.11.2. Methods

Nuclear protein isolation and bicinchoninic acid (BCA) assay quantification

See sections 2.8.2 and 2.9.2

Oligonucleotide annealing

Commercial single-stranded oligonucleotides were suspended in annealing buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) to an initial concentration of 200 mM. Equimolar quantities of forward and reverse single-stranded oligonucleotides were mixed and heated to 95°C for 2

minutes with a thermocycler. This was left to cool at room temperature over an hour, and could be stored at -20°C.

DNA pulldown

The protocol was based on the manufacturer's recommendations (http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/1/s1638pis.pdf), and is summarised below. For every reaction, 50 µl of straptavidin-conjugated agarose beads mixture (Sigma, S1638) was centrifuged at 4,000g for one minute at room temperature. The preservative supernatant was discarded and the beads were washed with 100 µl of Binding/Washing buffer (10 mM Tris pH 7.0, 1 mM EDTA, 2 M NaCl). The wash was repeated twice. 1 nmol of double-stranded oligonucleotides were immobilised onto the beads by incubating with the beads with 100 µl of Binding/Washing buffer for 15 minutes on a roller mixer at room temperature.

Following biotin-straptavidin interaction, the bead-oligonucleotide complex was centrifuged 4,000g for one minute at room temperature, and the supernatant discarded. The beads were further washed once with 100 µl of Binding/Washing buffer, and twice with incubation buffer (50 mM Tris pH 7.0, 1 mM EDTA, 100 mM KCl, 5% v/v glycerol, 0.1% v/v Triton-X100) supplemented with protease inhibitors. All centrifugation steps for the washes were conducted at 4,000g for one minute at room temperature.

The bead-oligonucleotide complex was incubated with 250 µl of nuclear extract (2 µg/µl) and 250 µl of incubation buffer for 30 minutes on a roller mixer at room temperature. Following incubation, the bead-oligonucleotide-protein complex was formed, and the sample was centrifuged 4,000g for one minute at room temperature, and the supernatant kept as a reference sample. The bead-oligonucleotide-protein complex was washed twice with 500 µl of nuclear extract buffer (20 mM HEPES pH 7.9, 400 mM NaCl, 1 mM EDTA, 10% v/v glycerol) supplemented with protease inhibitors. All centrifugation steps for the washes were conducted at 4,000g for one minute at room temperature. Supernatant from the washes were also kept as reference samples.

The washed bead-oligonucleotide-protein complex is then suitable for elution. 100 µl of 2x Laemmli buffer (4% w/v sodium dodecyl sulfate, 20% w/v glycerol, 120 mM Tris pH 6.8) supplemented with protease inhibitors was added to the washed bead-oligonucleotide-protein

complex, and incubated for 10 minutes at 70°C with occasional agitation. The sample was then centrifuged at 4,000g for one minute at room temperature.

Acetone precipitation

In these studies, the final quantity of nuclear extracts used was 2,000 µl, in other words, four of the above 500 µl reactions. The combined protein eluate in Laemlli buffer was added to cold acetone (-20°C), in a 1:4 ratio, to an acetone-compatible tube. This mixture was kept for 24 hours at -80°C with occasional vortex mixing.

The sample was then centrifuged at 13,000g for 10 minutes at 4°C. The supernatant was decanted and disposed as per safety protocols. The residual acetone was left to evaporate from the uncapped tube in a fume hood at room temperature for 30 minutes. The pellets were not over-dried, as this could have hindered dissolution. In this protocol, the pellet was resuspended in 20 µl of RIPA buffer (50 mM Tris pH7.4, 150 mM NaCl, 0.5% w/v sodium deoxycholate, 1% v/v NP-40, 0.1% w/v sodium dodecylsulfate). In general, the buffer of choice is dependent on the downstream process. Recent literature suggest that acetone precipitation is at least, if not more, efficient as triacetic acid precipitation in murine samples (Fic *et al.*, 2010).

Control sample

As the agarose beads have the ability to bind proteins themselves, control samples were run as above, but in place of the oligonucleotides, the same volume of annealing buffer with oligonucleotides was added.

SDS-PAGE electrophoresis, electrotransfer and Ponceau staining

See Section 2.10.2

N-terminal amino acid sequencing

Following identification of specific bands from the pull down (i.e. neither present in the final wash nor control sample), these were excised and assessed with N-terminal amino acid sequencing commercially with Alta Biosciences, UK (<http://www.altabioscience.com/>).

2.12. Co-immunoprecipitation

2.12.1. Principles

The basis of co-immunoprecipitation requires the retention of protein-protein interactions. Thus, lysis buffers containing milder detergents such as NP-40 or Nonidet P-40 were utilised as they have a lower propensity to interrupt protein-protein interactions. Unlike total cellular protein isolation, mechanical disruption is not utilised as it may promote disruption of protein-protein interactions, particularly sonication. Alternatively, other protocols utilise formaldehyde to form protein-protein crosslinks, which is maintained during, and reversed after immunoprecipitation by heating.

After isolation of the proteins in their native and complexed conformations, they are immunoprecipitated using specific antibodies targeted against the protein of interest. The protein-antibody complex is then pulled down by agarose beads, which provides a heavier anchor to allow separation from unbound proteins.

Once the bead - antibody - protein complex is isolated, it undergoes several washes to reduce non-specific binding. The proteins are then eluted with heating in a loading buffer including a reducing agent (such as beta-mercaptoethanol or dithiothreitol) to break the protein-protein interactions.

These samples are then suitable for separation using SDS-PAGE or 2D-electrophoresis, and immunoblotting.

Limitations

The main limitation of co-immunoprecipitation assays stem from the antibodies available. Most commercial antibodies are raised to a linear epitope, which may not be easily recognised when in native conformation. Additionally, the target region for the antibody may also be involved any potential protein-protein interaction, thus shielded from immunoprecipitation.

Additionally there are limitations with regards to quantification of results. As with other analytical systems with a final output dependent on signal intensity, immunoblotting a co-immunoprecipitated sample would at best, allow only semi-quantitative comparisons.

2.12.2. Methods

The protocol was based on a recent study identifying the protein-protein interaction between NBCn1 and calcineurin A (CnA) (Danielsen *et al.*, 2013).

Protein isolation was conducted with IP lysis buffer (25 mM Tris HCl pH 7.2, 150 mM NaCl, 1 mM EDTA pH 8.0, 1% v/v NP-40), with protease inhibitor mixture. After quantification, 20 µl of protein was set aside as “input” samples.

500 µl of protein, topped up to 1 ml with IP lysis buffer, was incubated with 2 ng of rabbit anti-NBCn1 polyclonal IgG (Abcam, ab82335) or 2 ng of the negative control, normal rabbit IgG (Santa Cruz, sc-2027) overnight on a roller mixer at 4°C.

Following overnight incubation, the samples were incubated with 20 µl of Protein A/G PLUS-Agarose beads (Santa Cruz, sc-2003) for 3 hours on a roller mixer at 4°C. The samples were centrifuged at 1,000g for 5 minutes at 4°C, and the supernatant discarded. The beads were washed three times with 1ml of IP buffer (25 mM Tris HCl pH 7.2, 150 mM NaCl, 1 mM EDTA pH 8.0), each time the samples centrifuged at 1,000g for 5 minutes at 4°C, and the supernatant discarded. After discarding the supernatant from the final wash, 50 µl of Co-IP Loading Buffer (62.5 mM Tris-HCl pH 6.8, 10% v/v glycerol, 10% w/v sodium dodecylsulfate, 100 mM dithiothreitol) was added to the beads and incubated at 95°C for 2 minutes. After elution, the samples were centrifuged at 1,000g for 5 minutes at room temperature, and the supernatant saved as the co-immunoprecipitation sample. Bromophenol blue was added to the sample to achieve a final concentration of 0.02% w/v.

The samples were analysed by SDS-PAGE as described in Section 2.10.

2.13. Isoelectric focusing

2.13.1 Principles

Isoelectric focusing utilises a pH gradient, as first described by Hoch and Barr (1955) with paper electrophoresis, and has been developed further to utilize an electrophoretic gel. With this, the sample components migrate towards the anode or the cathode depending on the pH of the gel gradient. The pH gradient is provided by carrier ampholytes, which are mixtures of amphoteric

compounds with a spectrum of isoelectric points, thus forming a gradient under the influence of an electric field. As they have high buffering capacities, the pH at the region does not change with the migration of proteins within the gel. Additionally, due to their hydrophilic nature, they do not bind proteins.

Upon electrophoresis, proteins come to a rest within the gradient where their net charges are zero, also known as their isoelectric points (pI). Should a protein diffuse away from its pI, it would gain a charge and migrate back which provides the focusing effect. As the proteins become highly concentrated at their pI, this results in a high sensitivity for detection, and small charge differences can be differentiated. There is flexibility in the choice of pH gradients within the gels, which allows improvement of resolution by using narrower gradients.

As indicated in Section 3.9, the non-synonymous SNP rs3755652 results in an amino acid change from the negatively charged glutamic acid to the positively charged lysine (E326K). Prediction tools suggest that this amino acid change alters the pI of NBCn1 by approximately 0.1.

2.13.2. Methods

Protein isolation

Whole-cell protein extraction with RIPA buffer, and protein quantification was conducted at the same as with SDS-PAGE studies (see Sections 2.8.2 and 2.9.2.). The desired quantity of proteins was acetone precipitated (see Section 2.11.2), and re-suspended in urea / thiourea / CHAPS buffer (6 M urea, 2 M thiourea and 2% w/v CHAPS). Just before electroseparation, 5x loading buffer (with a final 1x concentration of 10% v/v glycerol, 20 mM lysine, 20 mM arginine, 20 mM DTT, 0.001% w/v bromophenol blue) was added.

Casting of isoelectric focusing gels

An isoelectric focusing gel (pH 3-10) was casted using the Mini-PROTEAN Casting System. The final concentrations of the components are 5% (w/v) Acrylamide : Bis-Acrylamide (37.5 : 1), 2% (w/v) 3-10 ampholytes, 0.1% (w/v) APS, 0.1% (w/v) TEMED and 8 M urea. The plastic well combs were fixed in place, and the gels were set aside for 30 minutes to allow the acrylamide to polymerise.

Separation of proteins by isoelectric focusing

The gel casts were assembled into the Mini-PROTEAN Tetra Cell system. The inner chamber was filled with the cathode buffer (20 mM lysine, 20 mM arginine), and the outer chamber filled with anode buffer (70 mM phosphoric acid). Electrophoresis was conducted initially at 100 V for 60 minutes, then 250 V for 60 minutes and finally at 400 V for 90 minutes. When the proteins have reached their respective isoelectric points on the gel, the current eventually drops to zero, indicating the focusing is complete. The module was kept at 10°C throughout at the experiment.

Electrotransfer of proteins from isoelectric focusing gel to PVDF membrane, immunoblotting and enhanced chemiluminescence

These steps were performed as the same with SDS-PAGE gels (see Section 2.10.2).

2.14. Electrophoretic mobility shift assays (EMSA)

2.14.1. Principles

The regulation of gene expression is in a large part based on DNA-nuclear protein and RNA-protein interactions. DNA-nuclear protein interactions may take the form of transcription factor binding to DNA elements, up- or down-regulating gene expression. Additionally, DNA-histone interactions are a significant epigenetic regulator of gene expression. Histone proteins combine to form nucleosomes, which in turn binds segments of DNA to form closed condensed chromatin. This results in downregulation of the corresponding DNA regions that is spooled around the nucleosome. RNA-protein interactions may occur at various points post-transcription, including control of splicing by spliceosomes, polyadenylation by polynucleotide adenylyltransferase, and effects on mRNA stability, localization and translation. This section will only discuss EMSAs as a tool to investigate DNA-nuclear protein interactions.

EMSAs utilize the differences in charge-to-size ratios of unbound and protein-bound double-stranded oligonucleotides (Fried and Crothers, 1981; Garner and Revzin, 1981). As proteins need to retain its native conformation to be able to interact with oligonucleotides (or DNA sections), the protein extraction (see Section 2.8), protein-oligonucleotide reaction buffers and separating gel are all conducted in non-denaturing conditions.

These oligonucleotides are labelled to be able to detect non-protein bound and protein bound species. This section describe in detail the biotin / streptavidin / horseradish peroxidase method (Li *et al.*, 2004), but variations include biotin / streptavidin / alkaline phosphatase, digoxigenin (Denkin *et al.*, 2004) or phosphorus-32 (³²P), as well as nucleic acid stains of unlabeled oligonucleotides (Förster-Fromme and Jendrossek, 2010; Busmann *et al.*, 2010). The advantages and disadvantages of the various reporters are summarized in Table 11.

	Biotin / streptavidin / horseradish peroxidase	Biotin / streptavidin / alkaline phosphatase	Digoxigenin / alkaline phosphatase labelled anti-digoxigenin	Phosphorus-32	Nucleic acid stains
Sensitivity	High	Low (see Figure 27)	Moderate	High	Low (see Figure 28)
Specificity	High	High	High	High	High (But lower with long templates)
Additional safety factors	None	None	None	Special precautions required for handling radioactive materials	None
Cost	Moderate	Moderate	Moderate	High	Low

Table 11: Advantages and disadvantages of the different EMSA reporters

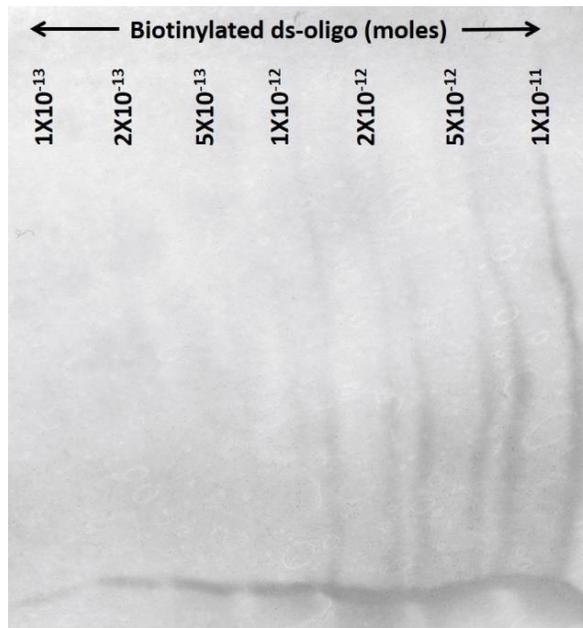


Figure 27: Streptavidin-alkaline phosphatase method of detecting labelled oligonucleotides
 Note relatively poor signal-to-noise ratio as compared to the biotin-streptavidin-horseradish peroxidase using only 10^{-14} moles of biotin-labelled oligonucleotides.

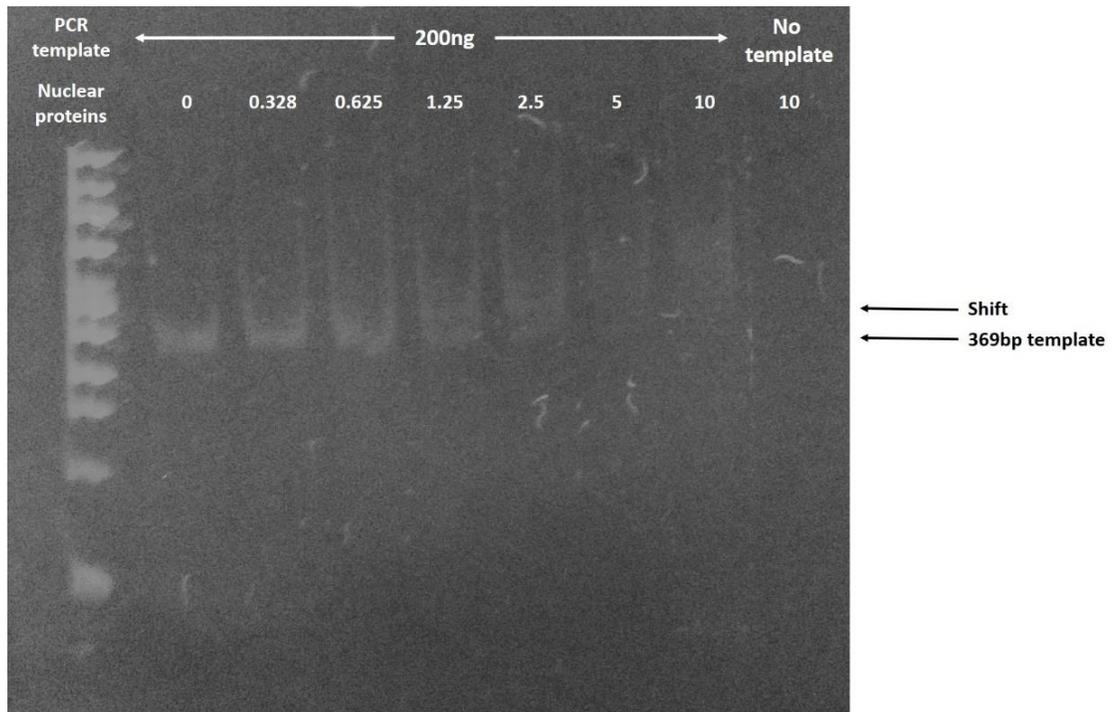


Figure 28: Nucleic acid staining method of detecting EMSA with PCR products
 Note relatively poor signal-to-noise ratio (200 ng 369 bp PCR template approximates to 1.2×10^{-12} moles) as compared to the biotin-streptavidin-horseradish peroxidase using only 10^{-14} moles of biotin-labelled oligonucleotides. Also there is smearing of the shifted products from 5ng nuclear protein and above.

In the absence of protein binding, double-stranded oligonucleotides have a strong negative charge due to their sugar-phosphate backbone, whilst still being small. This results in rapid mobility when exposed to an electrophoretic force attracting it towards the anode. In the presence of protein-binding the oligonucleotide-protein complex will be larger and have a smaller charge as most DNA-binding proteins carry an overall positive charge. This results in a smaller charge-to-size ratio and thus retards its mobility in an electrophoretic field. A difference in mobility can be visualized as two separate bands when the label is appropriately probed.

The reaction mixture is combined with variable quantities of nuclear proteins, labeled oligonucleotide probes and unlabeled competitor probes. Unlabeled competitor probes serve to determine whether any shifted signals are specific. Addition of poly(dI.dC) functions as a non-specific competitor to increase the specificity of the assay. In supershift experiments, the addition of an antibody targeted against specific nuclear proteins may produce a further increase in size with no significant change to charge, thus further reducing its complex's electrophoretic mobility.

This reaction mixture is then electrophoresed in a non-denaturing gel, and electrotransferred onto a nylon membrane, as this membrane is more suitable for retaining nucleic acids. The transferred oligonucleotides are then crosslinked with oven drying and 254nm ultraviolet light. Following crosslinking the biotin-labelled oligonucleotide can be detected by a variety of methods. The work in this thesis uses a commercial streptavidin-linked horseradish peroxidase kit, the LightShift[®] Chemiluminescent EMSA Kit (ThermoScientific).

Limitations

As with other analytical systems with a final output dependent on signal intensity, EMSAs would at best, allow only semi-quantitative comparisons. Another important consideration when interpreting EMSAs is that the DNA-nuclear protein interaction is conducted within a purely artificial experimental context. This reaction environment may not necessarily mimic that of an intracellular or nuclear milieu, notwithstanding the added complexities of epigenomic control.

Complementary methods

EMSAs can indicate an *in vitro* interaction between a nuclear protein and a short segment of double-stranded oligonucleotide, but further complementary methods may be required to strengthen any hypothesis. One such method is the chromatin immunoprecipitation (ChIP) assay.

This technique utilises the crosslinking of the DNA-nuclear protein complex in chromatin, and antibodies to proteins of interest to pull-down DNA fragments which the protein is bound to. ChIP has the additional advantage of the DNA-nuclear protein interaction occurring within a cellular system as opposed to EMSAs, occurring purely within an experimental context.

A further alternative, not used in this thesis, is the DNase footprinting assay which can be used to identify the specific binding site of a protein to DNA *in vitro*. This technique utilises short fragments of labeled DNA (or PCR product), and an agent that cleaves the fragments into even smaller fragments. When a protein that can bind the DNA template is introduced into the reaction, it will protect the DNA it is bound to from the cleavage agent, thus identifying the “footprint” or the binding site of the protein.

2.14.2. Variations – Western-Electrophoretic mobility shift assay (WEMSA)

Some variants of the method combine western blotting together with EMSAs, gaining the name Western-Electrophoretic mobility shift assay (WEMSA) (Deckmann *et al.*, 2012). This utilises the same binding conditions and native gel electrophoresis as EMSAs, but have the added component of electrotransfer to either a PVDF or nitrocellulose membrane for detecting specific proteins with their respective targeted antibodies. The principle of varying electrophoretic mobilities still applies. An unbound protein, being large and carrying relatively low charge, will not be particularly mobile in a native gel. On the other hand, a DNA-oligonucleotide complex will have significantly larger negative charge with only a small gain in size, thus increasing its electrophoretic mobility.

2.14.3. Methods

Nuclear extract isolation and bicinchoninic acid (BCA) assay quantification

See sections 2.8.2 and 2.9.2

Oligonucleotide annealing

Commercial single-stranded oligonucleotides were suspended in annealing buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA) to an initial concentration of 200 mM. Equimolar quantities of forward and reverse single-stranded oligonucleotides are mixed and heated to 95°C for 2 minutes with a thermocycler. This was left to cool at room temperature over an hour, and could be stored at -20°C.

Electrophoretic mobility shift reactions

The reaction mixture (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA, 5% v/v glycerol, 0.5 mg/ml poly(dI.dC), 1 mM phenylmethanesulfonylfluoride, 1 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µg/ml pepstatin) was combined with variable quantities of nuclear proteins, labeled antibody, unlabeled competitor antibodies and supershift antibodies. Unless where specified, quantities of nuclear proteins were 10 µg (2 µg/µl), labelled oligonucleotides were 10 femtomoles (10^{-14} moles), unlabeled competitor oligonucleotides were 100 femtomoles (10^{-13} moles). Accounting for the variable volumes of reactants, the final volume totals 20 µl. The reaction occurs at room temperature for 30 minutes. At the end of the reaction, 2 µl of 10x loading buffer (10 mM Tris pH 7.5, 1 mM EDTA, 50% v/v glycerol, 0.001% w/v bromophenol blue) is added.

When required for supershift assays, 1 or 5 ng of the antibody (rabbit anti-AP2γ IgG antibody, Santa Cruz, sc-8977 X; goat anti-PHOX2A IgG antibody, Santa Cruz, sc-13229 X; or rabbit anti-ZAC1 IgG antibody, Santa Cruz, sc-22811 X) was added into the reactions, while maintaining the overall volume at 20 µl. The rationale for selecting these antibodies is discussed in results (Results Section 4.3.3).

Non-denaturing gel electrophoresis

For EMSAs, electrophoresis was conducted using a vertical 6% polyacrylamide gel (6% w/v 37.5:1 acrylamide:bisacrylamide, 0.5x Tris-Borate-Edetate (TBE) buffer, 0.001% w/v ammonium persulphate (APS) and 0.001% v/v tetramethylethylenediamine (TEMED)). The running buffer was 0.5x TBE. Although APS is important for the polymerization reaction, may interfere with protein-oligonucleotide binding. Thus, it was electrophoresed away from the gel by a “pre-run” before the samples are applied. TBE buffer was used as they are effective keeping DNA or oligonucleotides soluble and protect from enzymatic degradation.

The separating electrophoresis was conducted in 0.5x TBE buffer at 100 V at 4°C, until the bromophenol blue front, approximating the electrophoretic mobility of free unbound oligonucleotides, is three-quarters down the gel.

Electrotransfer to nylon membrane and crosslinking

Following electrophoresis, the oligonucleotides are electrotransferred onto a nylon membrane at 100 V for 30 minutes, in 0.5x TBE buffer. The transfer could be visually confirmed with the transfer

of the bromophenol blue front. The oligonucleotides were then crosslinked on to the membrane with oven drying for 30 minutes, followed by 254nm ultraviolet light at a dose of 0.12 J/cm² (Biolink BLX-254E).

Detection of biotin-labeled oligonucleotides

The studies conducted for this thesis used the LightShift[®] Chemiluminescent EMSA Kit (#89880, ThermoScientific), for which the manufacturer's recommended protocol is summarized below.

For each membrane, 20 mL of Blocking Buffer was added and incubated at room temperature for 15 minutes with gentle shaking on a plate shaker. After blocking, the solution was discarded and replaced with a solution containing 66.7 µl of Stabilized Streptavidin-Horseradish Peroxidase Conjugate and 20 mL Blocking Buffer with a further incubation for 15 minutes with gentle shaking on a plate shaker. The membrane was then washed four times with 20 mL of 1x Wash Solution. Each wash lasted 5 minutes at room temperature with gentle shaking on a plate shaker.

Following washing, the membrane was transferred to a new container and 30 mL of Substrate Equilibration Buffer was added, with a further incubation at room temperature for 5 minutes on a plate shaker. After the equilibration step, the membrane was again transferred to a new container and exposed to a Substrate Working Solution containing 6 mL Luminol / Enhancer Solution and 6 mL Stable Peroxide Solution. This was incubated at room temperature for 5 minutes without agitation.

The membrane was then removed from the Working Solution and an edge of the membrane was blotted on a paper towel for 2-5 seconds to remove excess buffer without allowing the membrane to become dry. The membrane was then exposed to an autoradiograph in the dark room and developed. The exposure time was adjusted to obtain the desired signal.

In semi-quantitative analyses, the intensities of the signals were quantified by the public domain, Java-based image processing program ImageJ.

Western-Electrophoretic mobility shift assay (WEMSA)

WEMSAs were conducted similar to EMSAs until the stage of electrotransfer. Electrotransfer and immunoblotting was conducted as per the Western blotting protocol (see Section 2.10.2).

2.15. Chromatin immunoprecipitation (ChIP) and formaldehyde-assisted isolation of regulatory elements (FAIRE) assays

2.15.1. Principles

ChIP and FAIRE assays allow the identification of DNA sequences that are protein-bound within the cell nucleus. They are both based on the ability of formaldehyde to form reversible DNA-nuclear protein crosslinks, which are resistant to shearing damage or denaturing when exposed to ultrasound pulses. Conversely, sequences of DNA that is not interacting with proteins will not form crosslinks and thus not protected from shearing damage. Formaldehyde initially interacts with the amine structure of nucleotides (e.g. guanine, adenine or cytosine) to form a methylol structure, followed by a Schiff base (protein-N=CH₂). This Schiff base forms the basis of the cross-link with an amine group of an amino acid on the interacting protein.

For ChIP assays, the isolated sheared chromatin undergoes immunoprecipitation with an antibody against a specific (nuclear) protein with beads providing an anchor to precipitate out of solution. The DNA sequence of interest within the immunoprecipitated complex can be released following sodium hydroxide elution, de-crosslinking by heating and enzymatic digestion of residual proteins (Gilmour and Lis, 1984). These DNA fragments can form templates for end-point of quantitative PCR. Alternatively, FAIRE analyses of the chromatin preparation alone provides an overall assessment of which DNA sequences are protein/nucleosome-bound, as exemplified in Figure 67 (Smith *et al.*, 2011).

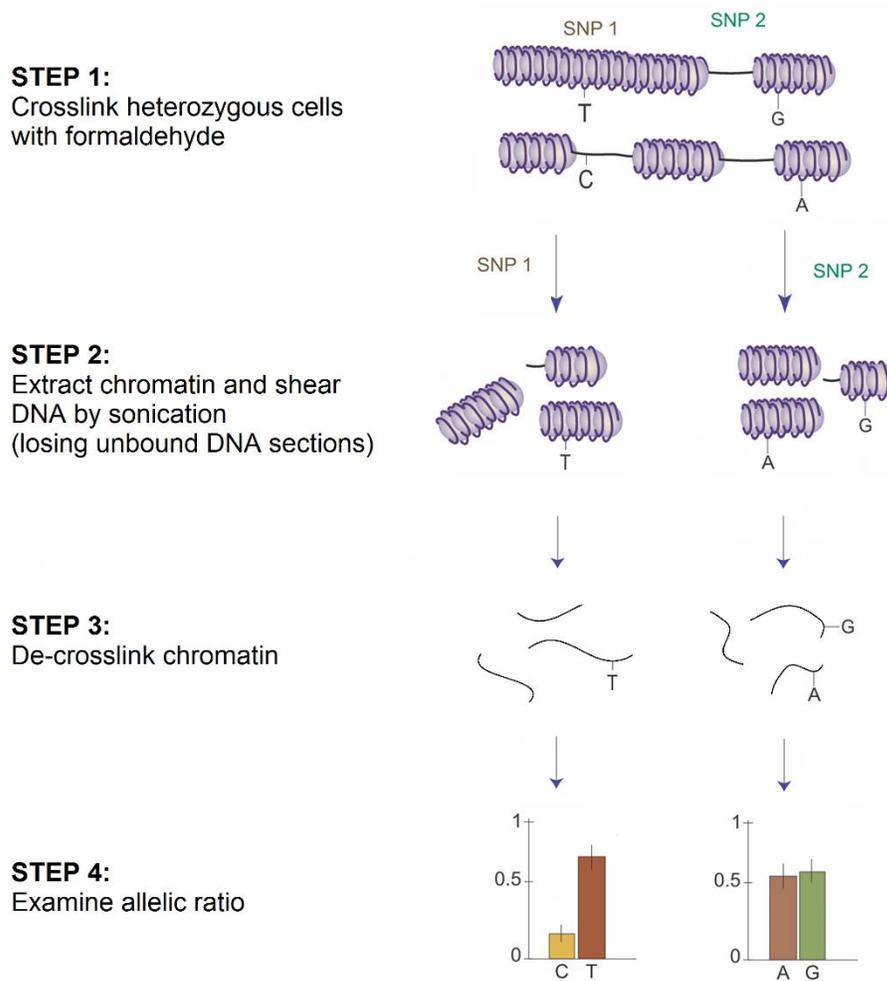


Figure 29: Cartoon showing sequence of steps in a modified protocol to FAIRE. The protocol identifies preferentially protein/nucleosome-bound alleles of SNPs (SNP1) as compared to polymorphisms has no impact (SNP2) (Significantly adapted from Smith et al., 2011).

In the work for this thesis, the DNA fragments are amplified by PCR, and the amplicons analysed with Sanger sequencing to assess allelic imbalances (see Section 2.6.2).

Limitations

Whilst FAIRE and ChIP utilises the same principle of formaldehyde crosslinking of chromatin, their main limitations actually contrast with each other.

FAIRE is non-specific in identifying DNA-nuclear protein interactions. In other words, the assay would be sensitive in identifying *any* DNA-nuclear protein interaction, but identity of the interacting protein would not be identified. Conversely, ChIP would be more specific to the antibody target, but less sensitive. The reduced sensitivity would often require a large number of

cells for chromatin isolation. The additional limitation is the difficulty in predicting the exact identify of the protein interacting with the DNA. As identified in Section 3.4, the process of predicting transcription factor binding has its flaws.

These studies gain strength when conducted in parallel with other studies aiming to determine the regulatory elements of gene transcription, such as electrophoretic mobility shift assays (Section 2.14) or luciferase assays.

2.15.2. Method

Formaldehyde crosslinking and DNA shearing

While the cells were still adherent to the tissue culture flasks, formaldehyde (stock 37% w/v) was added to the culture medium to achieve a final concentration of 1% w/v. For the purpose of the studies in this thesis, an 80-90% confluent T-25 flask (25 cm² surface area) was sufficient. This was left to incubate for 10 minutes in room temperature to allow DNA-nuclear protein cross-linking. Following this, glycine (stock 1.25 M) was added to achieve a final concentration of 125 mM to quench the residual formaldehyde. This was left at room temperature for 5 minutes, and then place on ice.

The culture media was then removed, and the cells washed with ice cold (4°C) PBS containing protease inhibitors twice. A cell lifter was then used detach the adherent cells from the tissue culture surface area and collected into a microcentrifuge tube. The cells were pelleted by slow centrifugation, at 1,000g for 4 minutes at 4°C, and the supernatant discarded.

The cells were then resuspended in ChIP Lysis Buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% w/v sodium dodecylsulfate) with protease inhibitors. To aid cell lysis, the suspension was pipette mixed and incubated on ice for 10 minutes. The DNA in the sample was then exposed to a probe sonicator (Jencons Scientific, 690-024) for three pulses of 10 seconds. This was optimised for HUASMCs and the particular probe sonicator to result in sheared DNA of lengths between 200 and 800 base pairs (Figure 30) – optimal for polymerase chain reactions. At this stage, for ChIP assays, the samples proceed to immunoprecipitation. For FAIRE assays, the samples do not undergo immunoprecipitation, and proceeds straight to de-crosslinking.

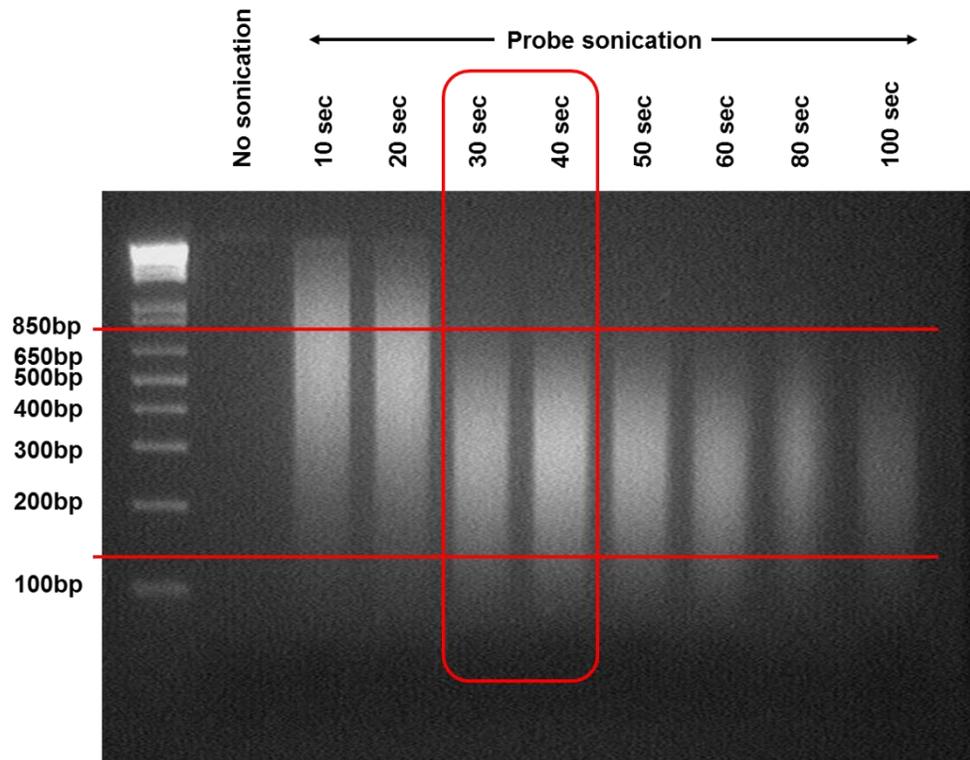


Figure 30: ChIP / FAIRE optimization of probe sonication
DNA smear (200ng per lane) with increasing duration of probe sonication of formaldehyde cross-linked chromatin. Probe sonication for 30 to 40 seconds provided the optimal sheared DNA sizes of approximately 150 to 800 base pairs.

Immunoprecipitation of chromatin

The sonicated lysates were then diluted with 800 μ l of ChIP Dilution Buffer (16.7 mM Tris pH 8.0, 1.2 mM EDTA, 150 mM NaCl, 1.1% v/v Triton X-100, 0.01% w/v sodium dodecylsulfate), together with the immunoprecipitating antibody (in this thesis, 10 μ g Rabbit Anti-Pol II IgG antibody; Santa Cruz, sc-9001) is added. The sample incubated overnight at 4°C on a roller shaker. Following that, 20 μ l Protein A/G PLUS-Agarose (Santa Cruz, sc-2003) was added to the sample and incubated again for 3 hours at 4°C on a roller shaker to collect the DNA-protein-antibody complex. The agarose beads were pelleted by slow centrifugation, at 1,000g for 4 minutes at 4°C, and the supernatant discarded.

This was followed by four separate washes of 1ml:

- Low salt buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% w/v sodium dodecylsulfate, 1% v/v Triton X-100);

- High salt buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl. 0.1% w/v sodium dodecylsulfate, 1% v/v Triton X-100);
- LiCl buffer (250mM lithium chloride, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% v/v NP-40, 1% w/v sodium dodecylcholate);
- TE buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA).

For each of the washes, the agarose beads were pelleted by slow centrifugation, at 1,000g for 4 minutes at 4°C, and the supernatant discarded.

The DNA-nuclear protein complex was eluted from the agarose by adding 200 µl Elution Buffer (1% w/v sodium dodecylsulfate, 100 mM NaHCO₃) and incubated at room temperature for 30 minutes on a roller shaker. The agarose beads were then separated from the eluate by centrifugation at 1,000g for 4 minutes at room temperature. The supernatant was collected as the DNA-nuclear protein complex eluate. The sample then undergoes de-crosslinking.

De-crosslinking

De-crosslinking of either CHIP or FAIRE samples requires adding of 8 µl of 5 M sodium chloride to the 200 µl sample. This was incubated at 65°C for 4 hours to reverse the DNA-nuclear protein crosslinks. To digest the residual proteins, proteinase K was used (addition of 20 µg proteinase K, 10 µl 0.5 M EDTA, 20 µl 1 M Tris pH 6.5), and incubated at 45°C for 1 hour. The DNA fragments were then suitable for isolation by DNA extraction columns.

Isolation of DNA fragments

The DNA fragments were isolated using the Wizard® SV Gel and PCR Clean-Up System (Promega, A9281) (See Section 2.5.6)

Polymerase chain reaction and Sanger sequencing

See Section 2.6.2

Control samples

For negative controls, parallel reactions are conducted, but instead of the primary antibody of interest, the same quantity of isotype control normal IgG from the same animal (in this thesis, 10 µg Normal Rabbit IgG; Santa Cruz, sc-2027) is used.

2.16. Immunocytochemistry

2.16.1. Principles

Immunocytochemistry allows the identification of proteins down to a cellular level, and particularly if used in combination with high-magnification confocal microscopy. Due to the requirements of fluorescent microscopy, cells are often cultured on a coverslip, but occasionally on tissue culture surfaces may be suitable. Adherent cells usually require a coated surface for successful adherence. The process of immunocytochemistry has similar principles to that of immunoblotting (Section 2.10.1). It requires the fixation of proteins (and cells), “blocking” to reduce background signal, and signal amplification with labelled antibodies.

Fixation plays several critical roles; to stabilize cell morphology and tissue architecture, disable proteolytic enzymes, strengthen samples to withstand further processing and staining, as well as protect samples against decomposition. Chemical fixatives work by either crosslinking or precipitating sample proteins. Care should be taken on the duration of exposure to the fixation agent, as prolonged fixation can mask target antigens or prevent antibody accessibility. No single fixative is ideal for all tissues, samples or antigens. Each fixative procedure must therefore be optimized to balance adequate fixation without altering the antigen or disturbing the endogenous location and the cellular detail of the tissue.

The most widely used chemical fixative is formaldehyde, which crosslinks primary amines on proteins and nucleic acids to form partially-reversible methylene bridges (see Section 2.15.1). Cross-linkers are better at preserving the cellular structure, but may reduce the antigenicity of some cell components. Alternatively, precipitating fixatives include organic solvents such as ethanol, methanol and acetone. They precipitate large protein molecules and are good for cytological preservation. These reagents have the added actions of also permeabilizing cells as they strip the lipid membrane. However, these reagents, requiring cold incubation, and typically causes cell shrinkage. Notably, both fixation types may denature protein antigens.

To evaluate intracellular (and if the experiment is not purely to assess membrane-bound proteins), the cells must be permeabilized to ensure free access of the antibody to its antigen. If the cells are permeabilized by the choice of fixatives, this can be conducted using detergents that disrupt the lipid membrane, including Triton X-100, NP-40 and Tween-20. Similar to fixation, the

concentration and duration of applying the permeabilisation buffer should be optimised as excess application of detergents can also remove the protein of interest.

Following permeabilisation, a blocking step is conducted reduce non-specific antibody binding to the cell culture surface, and thus background signal. Blocking agents are typically a 2–10% w/v solution of bovine serum albumin or addition of a 5–10% v/v solution of serum (ideally from the species in which the secondary antibodies were raised).

To achieve a fluorescent signal indicating the presence of a protein of interest, antibodies are used. Typically, immunofluorescence utilises primary-secondary antibody immunolabelling. Like in Western blots (Section 2.10.1) the primary antibody is raised against the protein of interest, and the secondary antibody is raised against the F_c section of the primary antibody. The difference as compared to Western blots, the primary antibody needs to be able to recognise the native conformation of the protein (as opposed to just the linearized amino acid epitope as the minimum sufficient criteria for Western blots) and the secondary protein is labelled with a fluorescent probe. Less commonly, directly fluorescent dye-conjugated primary antibodies can be useful in highly abundant targets, most protein targets require further signal amplification using primary and secondary antibodies.

To be able to identify the position of the protein of interest relative to sub-cellular structures, the main options are double immunofluorescence and/or counterstaining via non-antibody based techniques. To identify the nucleus, nucleic acid stains such as DAPI (4',6-diamidino-2-phenylindole) or Hoescht dyes bind the minor groove of double-stranded DNA (hence localising to nuclei), and when bound, it has a maximum fluorescence emission at 461 nm, in the blue spectra. To identify the plasma membrane, the commonly used counterstain is fluorophore-labelled wheat germ agglutinin. Wheat germ agglutinin is a lectin and has a high affinity for the membrane-associated N-acetylglucosamine. To identify the cytoskeletal structures, the commonly used counterstain is fluorophore-labelled phalloidin. Phalloidin is derived from a fungal toxin and has a strong affinity for F-actin. With both the fluorophore-labelled wheat germ agglutinin and phalloidin, the absorbance/emission wavelengths are dependent on the conjugated fluorophore. Alternatively, antibodies that specifically target certain organelles or cell types might be used as counterstains when used in a double immunofluorescence protocol.

Whenever more than one fluorophore is used, it is important that there is no significant overlap in their emission spectra. This can be aided by various charts or tables (for example, Table 12).

Fluorescent label	Color	Abs (nm)	Em (nm)	MW (daltons)
DyLight™ 405	Violet	400	420	793
Aminomethylcoumarin (AMCA)	Violet Blue	353	442	410
ATTO 425	Blue	436	484	498
Cy2™	Blue Green	489	505	897
DyLight™ 488	Blue Green	493	518	1,011
ATTO 488	Green	501	523	981
Fluorescein (FITC)	Green	495	528	390
ATTO 532	Yellow Green	532	553	1081
Cy3™	Yellow Green	552	565	949
DyLight™ 549	Yellow Green	550	568	982
Rhodamine (TRITC)	Orange	550	570	444
R-Phycoerythrin (RPE)	Orange	488	575	240,000
ATTO 550	Orange	554	576	791
Cy3.5™	Orange Red	581	596	1,286
Texas Red®	Red	596	620	625
ATTO 594	Red	601	627	1389
Allophycocyanin	Far-Red	650	660	100,000
Cy5™	Far-Red	650	667	975
ATTO 647N	Far-Red	644	669	843
DyLight™ 649	Far-Red	646	674	1,008
ATTO 655	Far-Red	663	684	887
IRDye®700DX	Near Infra-Red	689	700	1,954
Cy5.5™	Near Infra-Red	678	703	1,312
DyLight™ 680	Near Infra-Red	682	715	950
DyLight™ 800	Infra-Red	770	794	1,050
IRDye® 800CW	Infra-Red	774	800	1,166
IRDye® 800	Infra-Red	778	806	1,166

Table 12: Commonly used fluorophores with respective absorption and emission wavelengths http://www.rockland-inc.com/fluorescence_microscopy.aspx

Advantages, limitations and complementary methods

Immunocytochemistry has the advantage over protein immunoblots with the ability to localise the protein of interest to a particular region of the cell. Confocal microscopy has the added benefit of increased optical resolution and contrast due to being able to take images from one focal plane, and even reconstruct 3-D images.

Similar to protein immunoblots, the technique is only as good as the primary antibody used. This introduces the possibility of low sensitivity or specificities, especially as not all proteins are able to appropriately detect the native conformation. Additionally, protein-protein interactions within the cellular environment may block the target epitope of the antibody, limiting its ability to detect the protein. A further limitation is that, unlike protein immunoblots, there is no way of identifying different protein sizes and therefore different isoforms (unless the antibody is isoform-specific).

The quantification of immunofluorescence images is also complicated. Firstly, the signal is not directly proportional to protein quantity and therefore it is at best, semi-quantitative. There is also the difficulty of selecting appropriate cells to obtain images from, as this may potentially lead to biased selections.

2.16.2. Materials

Tris buffered saline (TBS – final concentration 20 mM Tris, 150 mM NaCl) was formulated initially at 10x concentration with 24 g Tris-HCl, 5.6 g Tris-base and 8 g NaCl dissolved into 900 ml distilled-deionised water. This was adjusted to pH 7.6 (with concentrated HCl or NaOH) and the volume adjusted to 1 L using distilled-deionised water. This solution is later diluted to 1x working concentration (1 part 10x TBS, 9 parts distilled-deionised water) and further adjusted to pH 7.6.

The fixative, 4% (w/v) formaldehyde was constituted from 37% (w/v) formaldehyde (Sigma, 252549) diluted in serum-free media (unsupplemented DMEM for HUASMCs and unsupplemented M199 for HUVECs). The permeabilisation buffer was 0.1% Triton X-100 (v/v) (Sigma, T8787) diluted in TBS. The wash buffer was constituted with 1% FBS v/v in TBS. The blocking buffer was constituted with 10% FBS v/v in TBS. DAPI was diluted to final concentration of 300 nM in TBS.

All rinses were done thrice.

2.16.3 Methods

Cells (HUASMC or HUVEC) were grown on coverslips in a 6-well plate pre-coated with 0.2% (w/v) porcine skin gelatin. Once the cells were between 60 to 80% confluent, the coverslips were suitable for immunocytochemistry. The coverslips were rinsed in wash buffer, followed by incubation with the fixative 4% w/v formaldehyde in TBS for 10 minutes at room temperature. The coverslips were again rinsed in wash buffer. Where indicated, this was the stage where the coverslip was incubated with permeabilisation buffer for 10 minutes at room temperature followed by rinses with wash buffer. The coverslip was then incubated with blocking buffer for 2 hours at room temperature. Once again, the coverslips were rinsed and then incubated in the primary antibody(ies) of interest, diluted in blocking buffer (see Table 13 below for concentrations) for 1 hour at room temperature. At this stage, parallel control studies were conducted with normal (untargeted) antibodies of the same isotype/host as the primary antibody of interest.

Following incubation with the primary antibody(ies), the coverslips were rinsed in wash buffer, and incubated in the corresponding secondary antibody(ies) diluted in blocking buffer (see Table 13 for concentrations) for 1 hour at room temperature in a humidified chamber. The coverslips were then rinsed in TBS, and then incubated with 300nM DAPI for 10 minutes at room temperature. The coverslips were then given a final rinse in deionised-distilled water. Once the coverslips were air-dried, they were mounted on glass microscope slides (VWR, 631-0110) using fluorescence mounting medium (Dako, S3023). To seal the coverslips onto the microscope slide, nail varnish was applied to the edges of the coverslip.

Antigen	Primary antibody	Control (for primary) antibody	Secondary antibody
NBCn1	Rabbit IgG anti-human NBCn1 5µg/ml (Abcam, ab82335)	Rabbit normal IgG 5µg/µl (Santa Cruz, sc-2027)	Donkey polyclonal Secondary Antibody to Rabbit IgG - H&L Alexa Fluor® 594 2µg/µl (Abcam, ab150076) Ex: 590nm, Em: 617nm
VE-Cadherin	Mouse IgG anti-human VE-Cadherin 2µg/ml (Abcam, ab7047)	Mouse normal IgG 2µg/µl (Santa Cruz, sc-2025)	Goat polyclonal Anti-Mouse IgG - H&L Alexa Fluor® 488 2µg/ml (Abcam, ab150113) Ex: 495nm, Em: 519nm

Table 13: Antibodies used for immunocytochemistry
Ex, excitation wavelength; Em, emission wavelength.

The slides were initially visualised on an inverted fluorescent microscope (EVOS® FL Cell Imaging System, Life Technologies), using the DAPI (Ex: 357 nm, Em: 447 nm), GFP (Ex: 470 nm / Em: 510 nm) and Texas Red® (Ex: 585 nm, Em: 624 nm) light cubes as recommended by the manufacturers for the respective fluorophores. Following this the slides were viewed on the ZEISS confocal microscope with filters for the respective wavelengths. Digital images were obtained from the microscopy for analysis.

Analysis

Digital images were quantified by the public domain, Java-based image processing program ImageJ. Channel-specific signals, matching the fluorophore of interest, were used. A line was created across a single cell, and the signal profile for each fluorophore was determined. The two peaks for membrane-specific proteins (e.g. VE-Cadherin) were used to identify the cell membrane. The mean signal of NBCn1 at the membrane was compared to the mean cytoplasmic signal (See Results Figure 88).

2.17. Collagen gel contractility assay

2.17.1. Principles

Collagen gel contractility assays (CGCs) provide an option in assessing *in vitro* cell contraction. One of the earliest descriptions of this methodology was by Bell and colleagues using human foreskin fibroblasts. They showed that the contraction, as visualized by a reduction in collagen gel size, occurs over the duration of days, and the extent of contraction was time- and cell-number dependent (Bell *et al.*, 1979). Over the years, this method has been adapted for a range of cell types including cardiac fibroblasts, airways smooth muscle, vascular smooth muscle, and particularly relevant to this thesis, vascular smooth muscle cells derived from human umbilical arteries (Sainio *et al.*, 2010). The basis of the assay is that of a collagen matrix. When kept cool in acidic conditions, collagen does not polymerise easily. A collagen gel is formed through polymerization when the pH is brought up 7-8 and allowed to warm up in the humidified environment of a tissue culture incubator. Further added into this mixture are the cells dissociated by trypsin, and the basal media.

The polymerization takes up to 60 minutes, and a further layer of media is overlaid on the gel to aid nutrition and hydration. Over the course of hours, the trypsin-dissociated cells regain their

previous morphology, and in fact form a more contractile phenotype in a three-dimensional matrix. There is however, no clear standardized protocol in the literature, but at least a working range is established.

Once fully formed, collagen gels provide a platform to assess cellular responses to various pharmacological stimuli and comparing cells of different populations or patient groups (Matsumoto *et al.* 2007; Ihalainen *et al.*, 2007).

Advantages, limitations and competing methodology

The advantages and limitations of CGCs should be viewed relative to the two main alternatives - small vessel wire myography, and forearm venous occlusion plethysmography (Table 14).

	Collagen gel contractility assays	Small vessel wire myography	Forearm venous occlusion plethysmography
Availability of samples	(+) Derived from otherwise discarded umbilical cords	(-) Discarded samples from abdominal surgery, which is not always viable for experiments (-) Invasive muscle or buttock biopsies procedures and therefore difficult to recruit volunteers	(-) Invasive, thus difficult to recruit in large numbers
Current ethical approval	Already approved through NRES ref: 08/ H0704/140; Amendment 4	Requires further Research Ethics Committee approvals	Requires further Research Ethics Committee approvals
Relevance to biological equivalent	(±) <i>In vitro</i> response, out of context of continuous, circumferential vessels, and absence of endothelial cells	(±) <i>Ex vivo</i> response, and difficult to standardize across different human samples	(+) <i>In vivo</i> localized response to pharmacological agent
Possible number of experiments conducted	(+) Due to tissue culture techniques and relatively small cell number required, numerous experiments can be conducted	(-) Samples from surgery or biopsies tend to be small, and do not allow replicates. Additionally, limited by number of wire myographs available.	(-) Time intensive procedures, thus limited number of experiments per subject possible
Range of experiments possible	(+) Large range possible, including pharmacological agents not licensed for human use, manipulation of gene expression such as siRNA, shRNA and overexpression plasmids.	(±) Able to use the range of pharmacological agents, but difficult to manipulate gene expression while retaining tissue viability	(-) Only a limited range of pharmacological agents would be suitable, often not targeting novel proteins of interest.
Speed of response	(-) Days	(+) Seconds to minutes	(+) Seconds to minutes
Cost	(+) No fixed costs and low variable cost	(±) High fixed cost, but low variable costs	(-) High fixed and variable costs

Table 14: Considerations in deciding methodology of choice assessing VSMC contraction (+), (±) and (-) denotes author's views on relative advantages and disadvantages.

In context of a study that compares genotypes, especially one with a minor genotype frequency of less than 5%, the need for a large population is paramount. Additionally the number and type of planned experiments also favours the *in vitro* method of assessing vascular smooth muscle contractility. Although sometimes regarded as the “gold standard” in assessing vascular response, the key disadvantage of forearm venous occlusion plethysmography, in the setting of NBCn1, is

the inability to challenge the physiological system of choice, as it would be inappropriate to aim to intravenously infuse agents that changes intracellular or extracellular pH *in vivo*.

2.17.2. Methods

A 250 µl mixture of collagen gel with HUASMCs was dispensed into each well of a 48-well plate. The final formulation of the collagen gel mixture was 1x DMEM (from 10x solution, Sigma), 1.5 mg/ml rat tail collagen type I (Sigma, C3867), 1.25×10^6 cells/ml, 2 mM glutamine, 40 mM sodium bicarbonate and 20 mM sodium hydroxide (optimised to achieve a final solution pH of 8). While preparing this mixture, all components were kept on ice to prevent early collagen polymerisation.

Once dispensed into the wells, the plate was placed in the humidified cell culture incubator for 30 minutes to polymerise. Following polymerisation, the gels were overlaid with 250 µl of unsupplemented DMEM and left to incubate for 24 hours. After 24 hours, the gels were manually and gently detached from the walls of the well. Once detached, the appropriate pharmacological challenge could be used.

At pre-determined time-points, digital images (Ivanov *et al.*, 2005) were obtained with a specially-fabricated camera housing case to ensure a fixed height and light exposure when obtaining images by a 14 megapixel digital camera (Nikon, S3100). The digital images were analysed using the public domain, Java-based image processing program ImageJ to calculate the gel surface area as a percentage of total well surface area.

2.18. Cloning, mutagenesis and bacterial transformation

2.18.1. Principles

Background

Mutagenesis and overexpression plasmids provide an important tool in molecular biology. Although used in a wide variety of settings, in the context of this study, it was utilised to investigate the impact of amino acid changes, and even splice variants, on protein function. There were many essential studies leading up to a functional insert, the earliest report of the process bearing resemblance to current protocols was described in Cohen *et al.* (1973). Cohen and colleagues managed to insert DNA encoding antibiotic resistance into unrelated plasmids, using restriction enzymes that produce overhanging digested ends, followed by ligating the insert and

vector together. Subsequent transformation into otherwise antibiotic-susceptible bacteria with the newly generated plasmid conferred it with newly-acquired antibiotic-resistance. Since then, the understanding of the range of restriction enzymes and the use of standardised vectors have made molecular cloning a key tool in research. Furthermore, improvements in transfection technologies (see Section 2.19), have allowed the use of expression plasmids in a range of cell lines.

The current range of expression plasmids includes those that are untagged or tagged. Untagged inserts will express the protein in its native form, with no added moieties. Tags can take several forms, primary classified by the added function. These include affinity-tags (to aid isolation, such as polyhistidine-tags), epitope-tags (to aid antibody detection, such as HA- and *myc*-tags), chromatography-tags (to aid separation, such as FLAG-tags) and fluorescent-tags (to aid visualisation or quantification, such as GFP-tags). Despite these advantages of a tagged protein, the tag itself can be sizable, and has the theoretical potential to modulate the function of the protein of interest. As the work in this thesis prioritises a functioning protein *in vivo*, an untagged vector is preferred.

***In vitro* site-directed mutagenesis**

There are several methods to achieve mutagenesis within an expression plasmid. The method used for this thesis is based on complementary mutagenic primers (Figure 31). These primers are able to anneal to the plasmid, but introduces the desired mutation, typically in the middle of the primer set. Following PCR with high-fidelity *Taq* polymerase, there is the combination of the new daughter plasmids that contain the mutation, as well as the original parental plasmids with the unwanted original sequence. The system then exploits the methylated and hemi-methylated nature of parental plasmids that has been produced in most bacteria. The methylation-sensitive *DpnI* endonuclease enzyme has restriction site that recognises the target sequence 5'-Gm6A⁺TC-3' (methylated sixth carbon of the adenine) sites, but not unmethylated 5'-GATC-3' sequences. Thus it only digests the parental plasmid, leaving the mutated daughter sequences undigested. The remaining plasmid is then suitable for bacterial transformation. It is worth noting that plasmid DNA from most *E. coli* strains (which are *dam*⁺) is methylated and is a suitable template for mutagenesis by this method. However, plasmid DNA derived from the *dam*⁻ *E. coli* strains, such as JM110 and SCS110, being unmethylated are not suitable.

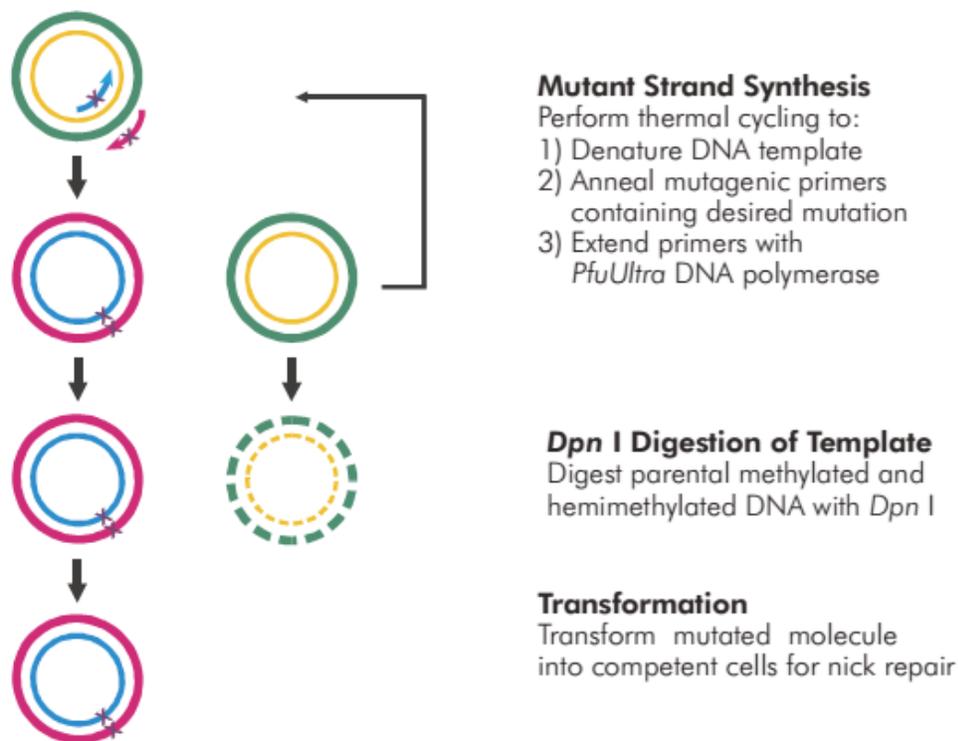


Figure 31: Cartoon overview of *in vitro* site-directed mutagenesis
 Adapted from <http://www.chem.agilent.com/library/usermanuals/Public/200523.pdf>

An alternative to *in vitro* site-directed mutagenesis is PCR-driven overlap extension (Heckman and Pease, 2007), which is described below for sequence deletion.

Deletion by PCR-driven overlap extension

This method has been previously described by Heckman and Pease (2007), utilising a two-step procedure. To generate a deletion from the original sequence, deletion primers are designed to “jump” the deleted sequence by matching 15 base pairs either side of the desired deletion. The reverse primer is then designed to be complementary to the forward primer (Figure 32, A). The first step generates two PCR fragments, which contain segments of overlap with each other (B). The second step aims to hybridise the two fragments (C), exploiting the region of overlap to form a larger product after PCR (D). This method can also be used to generate mutations or insertions, but not described for this thesis.

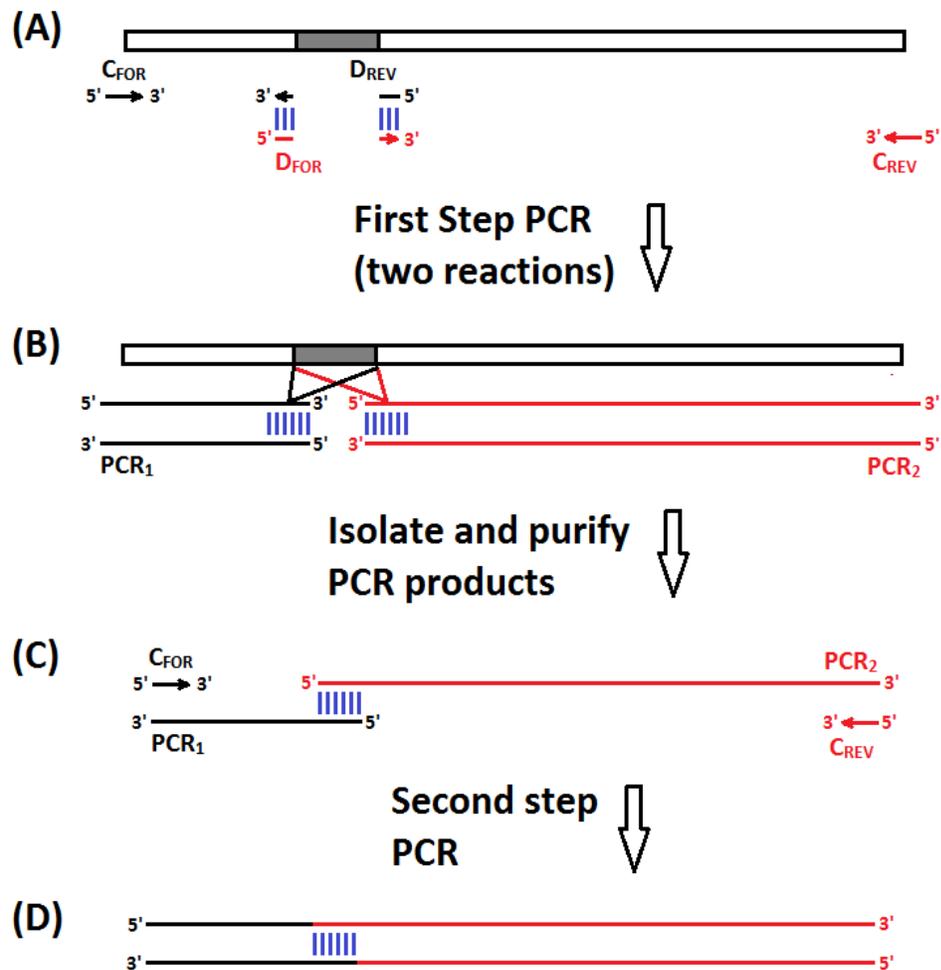


Figure 32: Cartoon overview of deletion by PCR-driven overlap extension

The grey box indicates the region to be deleted. The split primers D_{FOR} and D_{REV} are actually continuous, but represented as split only to demonstrate which sequences they are complementary to. Blue vertical bars identify complementary sequences at the region of overhang extension. Colour only identifies sequences for continuity through steps, and does not indicate any changed sequences.

Bacterial transformation

E. coli, which is typically used for propagating plasmids, is not naturally transformable. Their ability to take up DNA, also known as competency, must be induced and is now commonly based on the early description by Mandel and Higa (1970). This method is typically by heat-shocking, often assisted by divalent and multivalent cations (for example Ca^{2+} , Mg^{2+} , Mn^{2+} , Rb^+ or hexamine cobalt(III)). It is likely that a major reason for the lack of ability to take up DNA in its natural state is its lipid bilayer cell membrane. The cell membrane is negatively charged on the external surface, as is the overall charge of plasmids due to the phosphate backbone. With this, the cations are proposed to be able to “coat” the membrane, as well as the plasmid, serving to neutralise the

repellent forces of the two. This is further assisted by heat-shocking the bacteria. The rapid change in temperature (but not too high as to damage the cell) is proposed to temporarily alter the fluidity of the membrane, allowing the DNA molecule to enter the cell.

In uncommon circumstances that the standard heat-shock does not induce transformation, electroporation is another method of promoting competence (Wirth *et al.*, 1989). The cells are briefly shocked with an electric field, temporarily creating gaps in the cell membrane through which the DNA molecule enters. This permeability is a temporary phenomenon as the cell's membrane-repair mechanisms can seal up the hole.

Once transformed, the bacteria are typically plated on a selective agar plate, whereby an antibiotic (e.g. ampicillin or kanamycin) which the plasmid carries resistance against is added. With this, only bacteria with the antibiotic-resistance sequences can grow. However, this may be growth of empty vectors, or abnormal inserts. The growth is promoted by the presence of typtone (peptides derived from the tryptic digest of casein) and yeast extracts, which are also a source of amino acids and peptides

Further selection can be aided by the β -galactosidase Blue-White screening which requires addition of the chromogenic substrate X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) to the agar plate. This screening takes advantage of the characteristic of *E.coli* producing β -galactosidase in the presence of lactose by activation of the *lacZ* operon. However, host *E. coli* strains are containing *lacZ* Δ M15 mutation produces β -galactosidase which are inactive on its own. When transfected with a plasmid vector, which typically codes for the first 146 amino acids of β -galactosidase (also inactive on its own), it complements the mutated host enzyme to activate it. Thus, transfection with a native vector plasmid results in the β -galactosidase hydrolysing X-gal to form 5-bromo-4-chloro-indoxyl, which spontaneously dimerizes to produce an insoluble blue pigment. However, if the DNA insert is placed within the multiple cloning sites of the vector *lacZ* sequence, this will disrupt the production of an active β -galactosidase, and the colony will remain white in colour.

Once the colony is selected, it can be propagated in liquid culture, which continues to contain the antibiotic to maintain selectivity. The volume of the culture will depend on the eventual quantity of the plasmid desired.

Cryopreservation of bacterial cultures in glycerol stocks

Although plasmids can be re-transformed into more bacteria, there is an advantage of freezing the transformed bacteria, particularly as it takes a shortened time to produce a large quantity of plasmids. Hollander and Nell (1954) aimed to find a solution to the concern of damage to bacteria when frozen. They found that freezing conditions supplemented with 15% v/v glycerol significantly preserved survival of bacterial after multiple freeze-thaw cycles. The glycerol method also maintained its original bacterial characteristics, which in this case was the virulence of *Treponema pallidum* and syphilitic lesions in rabbits.

Like DMSO in cryopreservation of primary cell lines (see Section 2.1), glycerol is proposed to function as a cryopreservant by depressing the freezing point of the cells, serving to encourage greater dehydration of the cells, and minimises the solution effects as discussed previously. Glycerol is less toxic than DMSO to cells, but is also less penetrating. Thus glycerol is better for simpler cells such as bacteria, and DMSO for more complex cells such as in primary cell cultures.

Plasmid isolation

The transformed bacteria is usually pelleted from liquid culture, and re-suspended in an RNase-containing formulation to prevent bacterial RNA contaminating the final product. The Tris in the solution provides a pH buffer, and the EDTA chelates divalent cations to prevent activity of DNases. It is then lysed, typically with sodium dodecylsulfate as the surfactant. The presence of sodium hydroxide not just aids in lysing the cells, but also denatures plasmids and genomic DNA by breaking hydrogen bonds into single stranded DNA (ssDNA). Following a short incubation, the mixture is neutralised, typically with potassium acetate. The small circular plasmid DNA is more easily re-natured, but the huge genomic DNA cannot easily re-anneal, thus staying as single-stranded DNA. This allows for the selection of plasmids by staying in solution, whereas the single-stranded genomic DNA precipitate out with the sodium dodecylsulfate and denatured cellular proteins due to hydrophobic interactions, which in turn can easily be separated by centrifugation.

At this point, there are two main options. The plasmid DNA in solution can be extracted by the phenol/chloroform method, or by affinity chromatography-based methods. In the former, the plasmid DNA will be isolated in the upper aqueous phase, and can be purified in the same manner as TRIzol-based RNA extraction (see Section 2.3.1). The work in this thesis utilises commercial affinity chromatography-based columns. The resin preferentially binds to plasmid DNA under

certain pH or electrolyte conditions (as during the equilibration and wash phases), and releases the plasmid with other conditions (such as during the elution phase). Once eluted, the plasmid can be precipitated by isopropanol and washed with ethanol.

2.18.2. Methods

Cloning from HA-tagged *SLC4A7* plasmid into untagged vector

The expression plasmids used were cloned originally from a HA-tagged human *SLC4A7* plasmid kindly gifted by Dr K Witkowska. The vector which was to be used was pcDNA3.1(+). The known restriction sites of the *SLC4A7* insert was assessed using the online tool NEBCutter (<http://nc2.neb.com/NEBcutter2/>) to search for one which did not coincide with that of the vector (Figure 33).

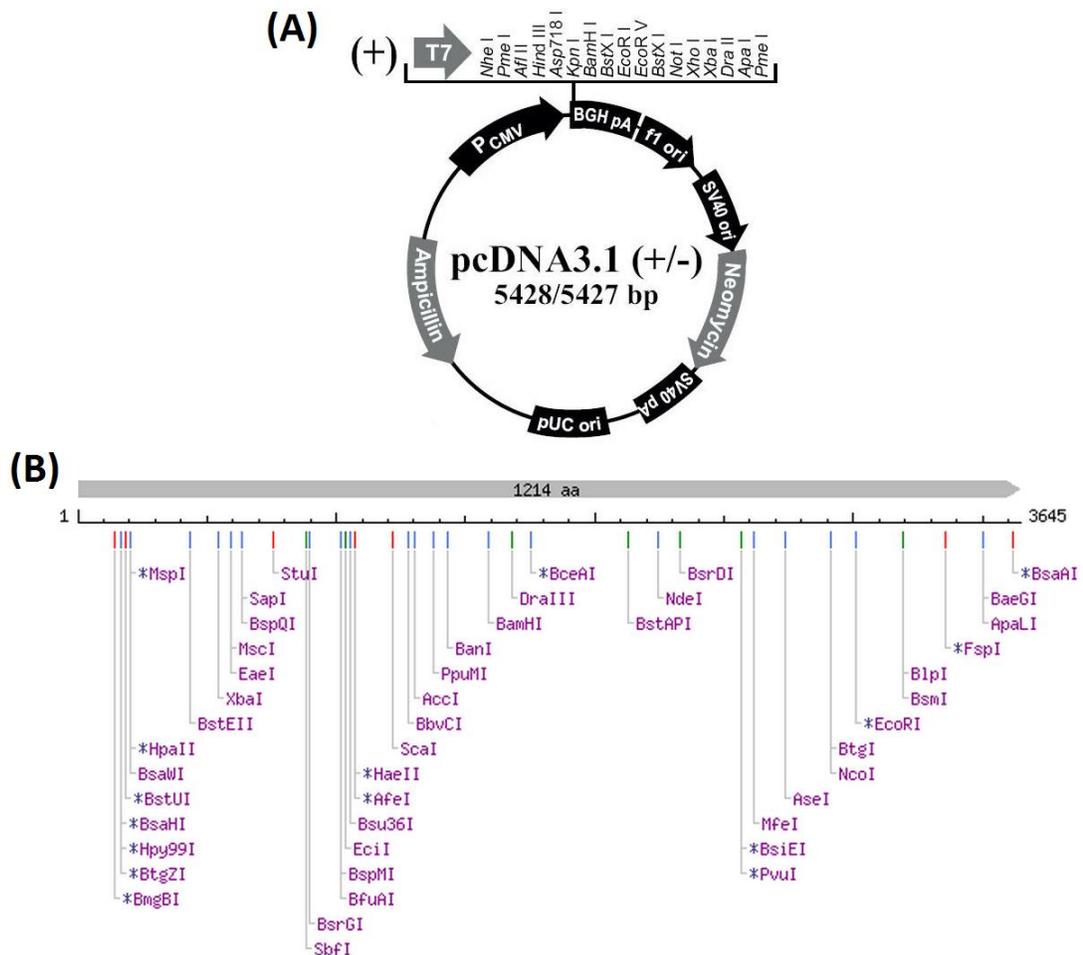


Figure 33: Restriction enzyme selection for pcDNA3.1(+)
 pcDNA3.1(+) restriction enzyme map (A) and NEBCutter output for full length *SLC4A7* insert (B)
 The only restriction enzymes identified as being present on the pcDNA3.1(+) map and the *SLC4A7* insert were *EcoRI* and *XbaI*. Thus any other pair of restriction enzymes would be acceptable.

Primers were designed with the sequence including a 5' extended segment GAAAAA (green below), followed by the desired restriction site (AAGCTT for HindIII, and GGTACC for KpnI; red below) and the subsequent 18 to 20 base pairs. This resulted in the following primer sequences:

- Forward with *HinIII* site (C_{FOR}): **GAAAAA**AAGCTTATGAAAGATTTTCGTCTGG
- Reverse with *KpnI* site (C_{REV}): **GAAAAA**GGTACCCTATATAATGAAGTTTCAGCA

Note that the underlined ATG and CTA sequences represent the start codon and the reverse complement of the stop codon (TAG).

As the primers had an estimated T_m of 65°C, thus the PCR reaction was initially conducted at annealing temperatures of 68°C. However, due to the low product yields, optimisation identified an annealing temperature of 54°C was ideal.

The PCR reaction utilised a high-fidelity Q5® High Fidelity DNA polymerase (New England BioLabs, M0491S). The reaction mix was kept at 0°C using an ice bath to inhibit any early polymerase activity. The final concentration of the reagents within the reaction included 10 ng of template, 100 µM primers, 1x reaction master mix buffer (which included the optimised concentrations of Taq enzyme, dNTPs and Mg²⁺). The final volume of the reaction was made up to 20 µl using nuclease-free water. The reaction was placed into a thermocycler, and the following program was used as based on the manufacturer's recommendations (Table 15).

		Temperature	Duration
Initialisation step (Hot Start)		98°C	5 minutes
35 cycles	Denaturing step	98°C	10 seconds
	Annealing step	54°C	30 seconds
	Elongation step	72°C	5 minutes (>1 minute per 1kb of PCR product)
Final elongation step		72°C	10 minutes

Table 15: Thermocycler program for cloning of SLC4A7

The products were electrophoresed on a 0.7% w/v agarose gel. Once resolved, the desired PCR product (Figure 34, 3666bp) was excised and extracted as described above (Section 2.5.6), and the concentration quantified measured using a spectrophotometer. Notably, attempts to derive full

length PCR products from primary cell line (HUVEC) cDNA was unsuccessful, perhaps due to the relatively low concentrations of target template.

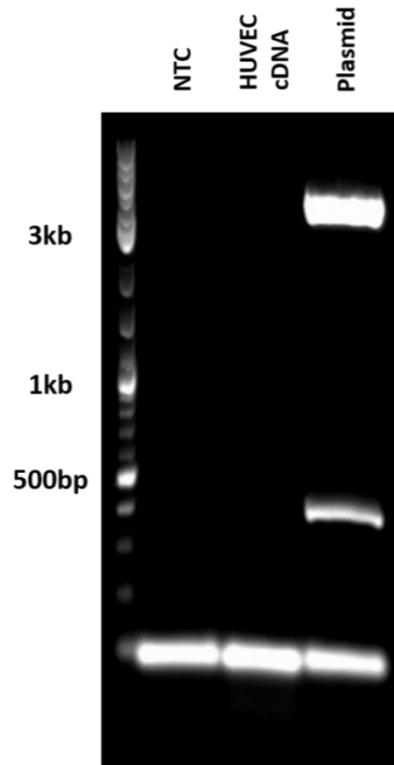


Figure 34: Endpoint PCR of cloning products

Endpoint PCR with primer pairs amplifying the full length SLC4A7 cDNA (3666 bp) from a HA-tagged plasmid (lane 3), with non-specific products around 400bp. No products amplified with primary cell line cDNA (lane 2).

Once purified, the PCR products were digested with *HindIII* and *KpnI* enzymes simultaneously. To achieve near complete digestion, NEBuffer 2 was chosen (

Figure 35). This process was also performed with the pcDNA3.1(+) vector. The final reaction included either 1 μ g *SLC4A7* insert or 10 μ g pcDNA3.1(+), 1x NEBuffer 2, 150 units of *HindIII* enzyme (New England BioLabs, R0104) and 350 units of *KpnI* enzyme (New England BioLabs, R0142), made up to a final volume of 50 μ l with nuclease-free water. The different quantities of enzyme were based on the relative activities in NEBuffer 2, and that this reached the maximal volume of enzyme mixture as 10% of the final volume, due to its glycerol content. The reaction was incubated at 37°C for 3 hours.

Enzyme	Cat#	Temp	Supplied NEBuffer	Supplements		% Activity in NEBuffer			
				BSA	SAM	1	2	3	4
HindIII*	R0104	37°C	NEBuffer 2	No	No	50	100	10	50
KpnI*	R0142	37°C	NEBuffer 1	Yes	No	100	75	0	50

Figure 35: Results of NEB Double Digest Finder

The NEB Double Digest Finder (<https://www.neb.com/tools-and-resources/interactive-tools/double-digest-finder>) identifying NEBuffer 2 as the optimal buffer for concomitant use of HindIII and KpnI.

The digested vector was electrophoresed on a 0.7% w/v agarose gel. Once resolved, the electrophoretic pattern of an undigested plasmid typically shows three bands, representing the different electrophoretic mobilities of the different conformations – relaxed, circular and supercoiled. After restriction enzyme digestion, this is now a single band of a linear plasmid (Figure 37 later). The digested plasmid band was excised and extracted as described above (see Section 2.5.5), and the concentration quantified measured using a spectrophotometer.

The digested vector was subsequently dephosphorylated. The reaction included 1 µg digested vector, 1 unit of calf intestinal phosphatase (New England BioLabs, M0290S), a final concentration of 1x NEBuffer 2, and made up to a final volume of 50 µl with nuclease-free water. The reaction mix was incubated at 37°C for one hour.

The digested insert, together with the digested and dephosphorylated vector were then ligated together. The reaction consisted of 10 ng of pcDNA3.1(+) vector, and equimolar (14.8 ng) quantity of the *SLC4A7* insert, a final concentration of 1x reaction buffer and 10,000 units of T4 DNA Ligase (New England BioLabs, M0202) and made up to a final volume of 20 µl with nuclease-free water. The reaction mix was incubated at 4°C for 72 hours. Alternatively, the reaction can be conducted at room temperature for 24 hours. A negative control plate is conducted concurrently, using only digested vector but without insert, also treated with T4 DNA ligase as above. Note that although the reaction was successful at a 1:1 molar ratio of insert to vector, alternative ratios commonly used are 3:1, 5:1 and 10:1.

Transformation

In preparation for the bacterial transformation process, selective bacterial agarose plate was made using a pre-mixed LB agar powder (Sigma, L2897), achieving a final concentration of 15 g/l agar, 10

g/l tryptone, 5 g/l yeast extract and 5 g/l sodium chloride. Each 100-mm petri dish required approximately 20 ml of the mixture. The mixture was sterilised by autoclaving to a temperature of 121°C for 15 minutes. When cooling after autoclaving, at a temperature less than 40°C, ampicillin was added to achieve a final concentration of 100 µg/ml. The mixture was then poured into individual plates. Once set, they are pre-warmed in the incubator at 37°C prior to plating of bacteria.

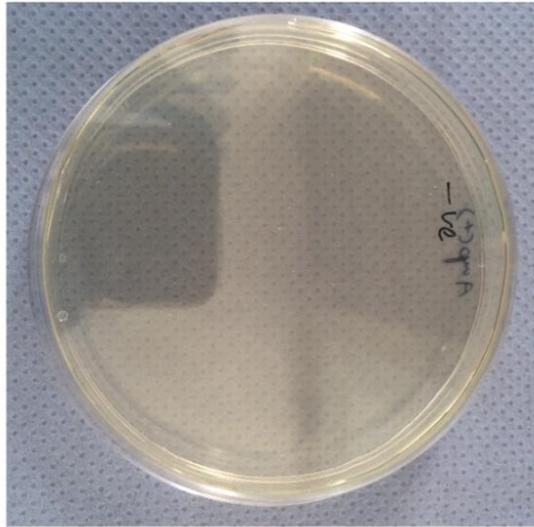
Prior to transformation of chemically competent cells (in the work for this thesis, either XL10 Gold or JM109), SOB medium (Super Optimal Broth) was created with a final concentration of 10 g/l tryptone, 5 g/l yeast extract, 0.584 g/l NaCl, 0.186 g/l KCl. This was autoclaved to 121°C for 15 minutes to achieve sterility. To create the final SOB medium, concentrated sterile-filtered MgCl₂ was added to achieve a final concentration of 10 mM. To create a SOC (Super Optimal Broth with Catabolite repression), concentrated sterile-filtered glucose was added to achieve a final concentration of 20 mM. Magnesium and glucose are not autoclaved with the rest of the solution and at the high temperatures, they could react with tryptic peptides.

To transform the chemically competent cells, they were brought to thaw slowly from -80°C storage on ice. Once thawed, 5 µl of the ligated mixture was added to the 50 µl aliquot of cells in a 14 ml round-bottomed polypropylene culture tube. This was mixed gently, without pipette mixing or vortex mixing. The mixture was incubated on ice for 30 minutes. Following this, the cells were heat shocked at 42°C in a water bath for 30 seconds. The reaction mixture was then immediately returned to incubate on ice for 2 minutes. After the 2 minutes, 250 µl of pre-warmed (37°C) SOC media was added to the reaction and shaken horizontally on a shaking incubator at 37°C for 1 hour, rotating at 225 rpm. Following this incubation, the mixture was then spread on the pre-warmed selective plate. The plates were placed inverted in the 37°C bacterial incubator overnight, for a maximum of 16 hours to prevent overgrowth.

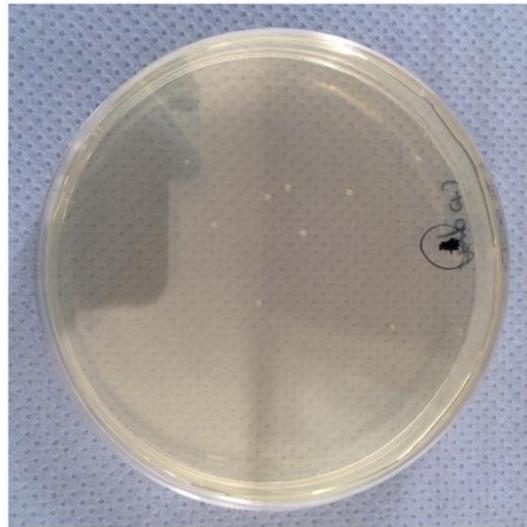
The next day (

Figure 36), for singular colonies were picked and immersed into 2.5 ml of SOB medium in a 14 ml round-bottomed polypropylene culture tube supplemented with ampicillin to reach a final concentration of 100 µg/ml, and incubated in a shaking incubator at 37°C for 24 hours, rotating at 250 rpm. After the 24 hours, if required, a glycerol stock of the bacterial culture was made by adding 200 µl to 500 µl of 20% v/v glycerol, and immediately vortex mixed and flash frozen with

dry ice or liquid nitrogen, to be stored at -80°C . The remaining mixture of bacterial cells was pelleted with centrifugation at 8,000 g for 10 minutes at 4°C .



**Vector only
-ve control
(0:1 ratio)**



**Insert + Vector
(1:1 ratio)**

Figure 36: Bacterial growth on petri dishes after ligation of cloning insert and vector. No growth observed on negative control plate with only digested vector treated with T4 DNA ligase as above (left), and 8 visible colonies with ligated insert and vector in a 1:1 ratio (right).

Plasmid isolation and purification

Plasmid purification was conducted using the column method from commercial kits. The protocol for the Mini-Prep commercial kit (Invitrogen, K2100-03) was used and summarised as follows. When purifying larger quantities, the Midi-prep (Invitrogen, K2100-04) was used, which follows the same principle, but using different volumes (protocol not summarised here).

Before the start of the purification process, the column resin was prepared by allowing 2 ml of equilibration buffer (EQ1; 0.1 M sodium acetate pH 5.5, 0.6 M NaCl, 0.15% v/v Triton X-100) to flow through the column, with the flow-through discarded. The harvested cells within the ultracentrifuge tube were then re-suspended with 400 μl of Resuspension Buffer containing RNase (R3; 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 20 mg/ml RNase A), and then lysed with 400 μl of Lysis Buffer (L7; 0.2 M NaOH, 1% w/v SDS). This was allowed to settle for 5 minutes with occasional inversion of the ultracentrifuge tube. Following lysis, 400 μl of Precipitation Buffer (N3; 3.1 M potassium acetate, pH 5.5) is added, followed by centrifugation at 16,000 g for 10 minutes at room

temperature. The pelleted precipitant was discarded, and the resultant supernatant was allowed to flow through the column. The flow-through from this step is discarded. The resin membrane, containing the plasmids were washed twice with two 2.5 ml of Wash Buffer (W8; 0.1 M sodium acetate pH 5.0, 825 mM NaCl), once again with the flow-through discarded. The column is then transferred to be place upon a collection microcentrifuge, and the plasmids are then eluted with 900 µl of Elution Buffer (E4; 100 mM Tris-HCl pH 8.5, 1.25 M NaCl).

The plasmids within the eluent was then precipitated with 630 µl of ice-cold isopropanol, and pelleted by centrifugation at 16,000 g for 30 minutes at 4°C. The supernatant was discarded, and the pellet washed with 1 ml of ice-cold 70% ethanol. Once again, the washed plasmids were pelleted by centrifugation, this time at 16,000 g for 5 minutes at 4°C. The supernatant was discarded, and the pellet left to air-dry. Once dried, the pellet was resuspended 50 µl TE buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA). The concentration of the purified plasmid was quantified using a spectrophotometer.

Restriction mapping

To conduct restriction mapping of plasmids with an approximate indication of success in transformation with the insert present, a reaction mix of 200 ng plasmid, 1x NEBuffer 2, 150 units of *HindIII* enzyme and 350 units of *KpnI* enzyme, made up to a final volume of 20 µl with nuclease-free water was incubated at 37°C for 16 hours. The digestion products were then electrophoresed on a 0.7% w/v agarose gel (Figure 37).

The mutagenesis consisted of a final concentration of 1x reaction buffer, 10 ng plasmid template, 100 ng of both primers, 2 % v/v proprietary dNTP mix, and 2.5 U of PfuUltra HF DNA polymerase, made up to a final volume of 50 µl with nuclease-free water. The reaction was placed into a thermocycler, and the following thermocycler program was used as based on the manufacturer’s recommendations (Table 16).

		Temperature	Duration
Initialisation step (Hot Start)		95°C	30 seconds
16 cycles	Denaturing step	95°C	30 seconds
	Annealing step	55°C	1 minute
	Elongation step	68°C	10 minutes (>1 minute per 1kb of PCR product)

Table 16: Thermocycler program for in vitro site-directed mutagenesis

Following the amplification of the mutated plasmids, 1 µl of 10 U/µl *DpnI* restriction enzyme was added to each reaction tube. This was pipette mix, and heated at 37°C for 1 hour to digest the parental (methylated, non-mutated) plasmid.

Transformation of XL10-Gold supercompetent cells were conducted as described above, except for the addition of 2 µl of the provided 2-β-mercaptoethanol to the 50 µl aliquots of cells just prior to mixing with the mutated molecule.

Sequence deletion by PCR-driven overlap extension

PCR-driven overlap extension was based on the protocol described in Heckman and Pease (2007). Primers were designed to span (and omit) the spliced Exon 7. In other words it included the 15 final base pairs of Exon 6, and the first 15 base pairs of Exon 8 in consecutive order (Figure 32 above). These primers were used in concert with the aforementioned cloning primers in two steps.

The first of the two steps includes two separate high-fidelity PCRs; one between the forward cloning primer (C_{FOR}), and the reverse deletion primer (D_{REV}) to generate the first fragment (PCR_1). The other reaction is between the reverse cloning primer (C_{REV}) and forward deletion primer (D_{FOR}) to generate the second fragment (PCR_2). These PCR reactions utilised the Q5® High Fidelity DNA

polymerase as before. In this instance, the final concentration of the reagents within the reaction included 100 ng of original full-length *SLC4A7* plasmid, 100 μ M of each primer, 1x reaction master mix buffer (which includes the optimised concentrations of Taq enzyme, dNTPs and Mg^{2+}). The final volume of the reaction was made up to 50 μ l using nuclease-free water. The reaction was placed into a thermocycler using the following program (Table 17).

		Temperature	Duration
Initialisation step (Hot Start)		98°C	5 minutes
35 cycles	Denaturing step	98°C	10 seconds
	Annealing step	54°C	30 seconds
	Elongation step	72°C	3 minutes (>1 minute per 1kb of PCR product)
Final elongation step		72°C	10 minutes

Table 17: Thermocycler program for step 1 of sequence deletion by PCR-driven overlap extension

The products of the two PCR reactions were then run on a 0.7% w/v agarose gel (Figure 39, Lanes 4 and 5). Upon resolution, the desired PCR₁ and PCR₂ products (777 bp and 2548 bp, respectively) were excised, purified (see Section 2.5.5) and quantified measured using a spectrophotometer.

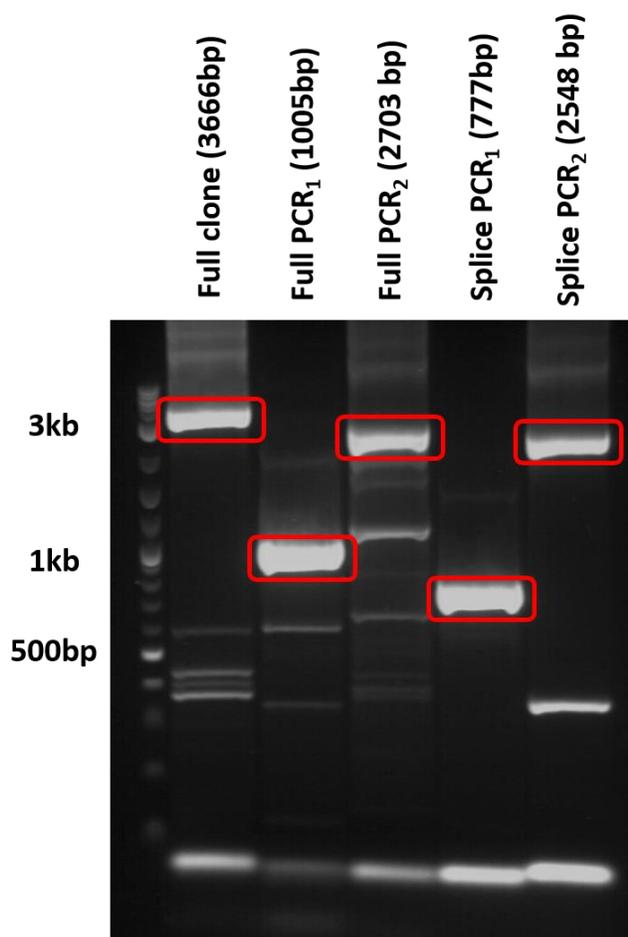


Figure 39: End-point PCR of sequence deletion by PCR-driven overlap extension (Step 1)
 Lane 1 was a positive control using the HA-tagged plasmid as template. Lanes 2 and 3 were parallel attempts at mutagenesis by overlap extension (not described further). Lanes 4 and 5 were PCR1 and PCR2 products respectively. There were multiple non-specific products, and only the products of the correct sizes were extracted. All products yielded >750 ng DNA, and had the sequences confirmed by Sanger sequencing.

The second step combines the two PCR products as templates, together with the flanking cloning primers, utilising the overhang to form deleted products. Once again, using the Q5[®] High Fidelity DNA polymerase kit, the final concentration of the reagents within the reaction include 100 ng of both PCR₁ and PCR₂, 100 μM of each C_{FOR} and C_{REV} primer, 1x reaction master mix buffer (which includes the optimised concentrations of Taq enzyme, dNTPs and Mg²⁺). The final volume of the reaction was made up to 50 μl using nuclease-free water. The reaction was placed into a thermocycler, and the following program (Table 18).

		Temperature	Duration
Initialisation step (Hot Start)		98°C	5 minutes
35 cycles	Denaturing step	98°C	10 seconds
	Annealing step	54°C	30 seconds
	Elongation step	72°C	5 minutes (>1 minute per 1kb of PCR product)
Final elongation step		72°C	10 minutes

Table 18: Thermocycler program for step 2 of sequence deletion by PCR-driven overlap extension

The PCR reaction products were run on a 0.7% w/v agarose gel (Figure 40). The desired smaller product (3270 bp) was then excised from the gel, purified (see Section 2.5.5) and quantified measured using a spectrophotometer. This PCR product now forms an insert to be placed within the pcDNA3.1(+) vector, same as when cloning the full-length insert described above.

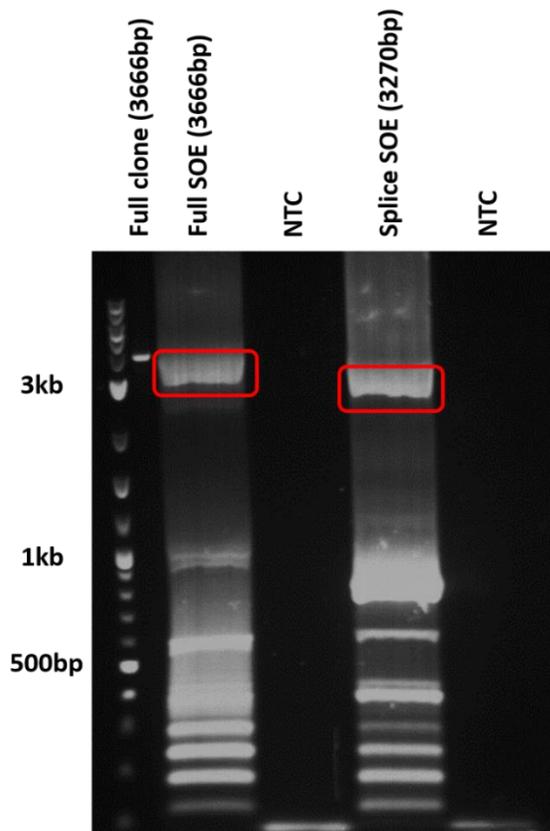


Figure 40: End-point PCR of splicing by PCR-driven overlap extension (Step 2)
 Lane 1 was a positive control using the HA-tagged plasmid as template. Lanes 2 was the parallel attempts at mutagenesis by overlap extension (not described further). Lanes 4 was the product of

sequence deletion by overlap. Lanes 3 and 5 are no template controls. There were multiple non-specific products, and only the products of the correct sizes were extracted. All products yielded >250 ng DNA, and had the sequences confirmed by Sanger sequencing.

As confirmation of the plasmids used for this thesis, plasmids S1G, U1 and Z3A (all named in chronological experimental order), together with vector control, underwent restriction enzyme mapping as previously described (

Figure 41).

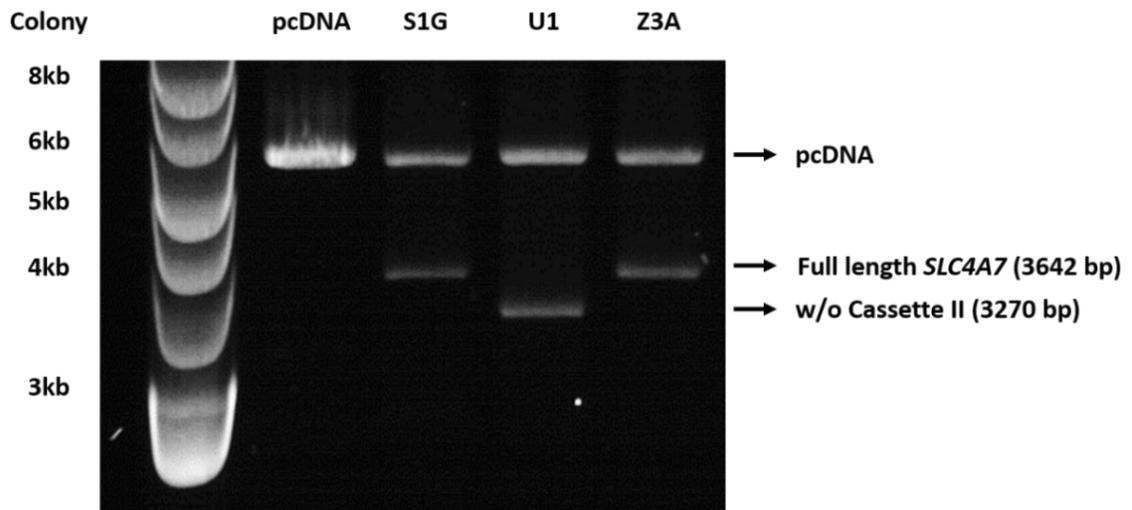


Figure 41: Restriction mapping of plasmids

200 ng of plasmid DNA were digested overnight with *KpnI* and *HindIII*. *SLC4A7* inserts of S1G, U1 and Z3A plasmid sequences were confirmed by Sanger sequencing. Note that the indicated fragments lengths are approximate, as does not account for the restriction enzyme sites overhangs.

Original plasmid name	Characteristics	Functional name
pcDNA3.1(+)	No change	Vec
S1G	Full <i>SLC4A7</i> insert (into pcDNA3.1(+)) via <i>HindIII</i> and <i>KpnI</i> restriction sites) with A allele at position 976 (AAG codon)	326K
U1	<i>SLC4A7</i> insert (into pcDNA3.1(+)) via <i>HindIII</i> and <i>KpnI</i> restriction sites) with deletion of base pairs 751 to 1122	Spl
Z3A	Full <i>SLC4A7</i> insert (into pcDNA3.1(+)) via <i>HindIII</i> and <i>KpnI</i> restriction sites) with G allele at position 976 (GAG codon)	326E

Table 19: Characteristics of the plasmids

Sequencing

All plasmids had their sequences confirmed by Sanger sequencing (see Section 2.6.2).

Reviving transformed bacterial from cryopreservation in glycerol stock

To revive transformed bacterial from frozen glycerol stocks, 50 ml of SOB media with 100 µg/ml ampicillin was initially prepared as above in a conical flask, and left to cool to 37°C. Once cooled, the glycerol stock was scraped with a sterile pipette tip, and immediately left immersed in the SOB/ampicillin. The vial containing the glycerol stock was always kept frozen by being placed on dry ice throughout. The conical flask was then incubated in a shaking incubator overnight, rotating at 250 rpm. After 16 to 24 hours, the bacterial growth was then pelleted and the plasmids extracted as above.

2.19. Transfection

2.19.1. Principles

Background

Transfection describes the process of introducing nucleic acids (DNA, RNA, siRNA, plasmids etc.) into eukaryotic cells, as opposed to transformation which refers to process involving non-eukaryotic cells (e.g. bacteria). Typically, it also excludes virus-driven methods, referred to as viral transduction.

There are now a wide variety of methods of transfection. An early but still commonly used, relatively cheap and simple method is calcium phosphate transfection, as initially described by Graham and Van der Eb (1973). By mixing the various components containing calcium, phosphate and DNA, precipitates form and are allowed to reach the cells. Although the precise mechanism is unclear, the precipitates are presumed to be taken up by the cells via endocytosis. Whilst useful, the calcium phosphate transfection method has several limitations, not least the range of cell lines that it can be successfully transfect. A concern that often occurs with calcium phosphate transfections is the variability in transfection efficiencies from experiment-to-experiment. It is also unclear whether the uptake of the precipitates significantly alters the intracellular milieu in addition to the genetic material.

A common alternative is lipofection. This is a lipid-mediated method of DNA transfection into living cells, which one of the first descriptions was by Fraley *et al.* (1980). Most lipofection methods currently use proprietary molecules, and thus no further information is readily available. In general they utilise cationic lipids, having a positively charged polar head group and a non-polar lipid chain tail. Once again, the exact mechanism is unknown, but it is presumed that it is based on the electrostatic interactions between the positively charged cationic head and the negatively charged phosphate backbone of plasmid DNA. In turn the amphiphilic nature of the cationic lipid then forms a lipid bilayer containing the DNA, in other words a liposome. The outer cationic head of the liposome is then able to interact and merge with the anionic lipids present on the plasma membrane, allowing the DNA contents to be internalised. Lipofection is very efficient with majority of cell lines, but still fails in some. It is also relatively more expensive.

Electroporation utilises an electric field placed across cell in suspension to enhance uptake of DNA macromolecules across the plasma membrane, with an early description by Neumann *et al.*, (1982). Since then, the method has been refined, and standardised electroporators are commercially available with a wide range of settings to suit the cell line of choice. As with the two previous descriptions, the exact chemical / molecular mechanism by which electroporation functions is uncertain, but broadly, the applied electric field temporarily moves ions present both inside *and* outside the cells. This is believed to be the mechanism which increases membrane permeability to exogenous molecules through small “holes”. Once the electrical field is turned off, the cell membrane repair mechanisms will seal up the passages and restore membrane integrity. Although this is a relatively easy method to use, and applicable in the majority of cell lines, there is the potential disadvantage of cell mortality, particularly with adherent cell lines that need to recover immediately post-electroporation to allow cellular function in re-adhering to the cell culture surface.

Other alternatives include, viral vectors, microprojectiles, microinjections, cationic polymers and magnetic nanoparticles, all of which have their advantages and disadvantages, but typically are not first-line (or even second-line) techniques. Additionally, this is not an exhaustive list as there are newer biotechnologies that also achieve cell transfection via different methods.

2.19.2. Methods

HEK293 and A10 culture

Studies with HEK293 and A10 cells were cultured as described in Section 2.1.5, with DMEM+10%FBS as the culture media. When subculturing the cells to proliferate, the cell culture surface did not require any gelatin coating. However, when utilising glass cover slips, the culture surface was pre-coated with 0.2% w/v gelatin (Sigma, G1890) to minimise cells detachment when undergoing multiple washes which is otherwise particularly noticeable with HEK293 cells with the immunocytochemistry (see Section 2.16) protocols.

HEK293 and A10 cells for transfection experiments were grown on a gelatin pre-coated 16-mm coverslip within 6-well plates or 35-mm petri dishes. Once the cells reach the desired confluence (40%), the reaction mix was prepared. At this confluency, the cells are likely to be in their log-growth phase, providing just enough cell-cell contact to help cell cultures grow, but not too much contact that results in contact inhibition, making cells resistant to uptake of DNA. The media was replaced prior to transfection, with 2ml of fresh DMEM+10%FBS media.

All studies were conducted in parallel, where all four plasmids (vector, S1G, U1 and Z3A) were transfected in parallel to minimise potential variability in the baseline characteristics of the cells.

Calcium phosphate transfection

Calcium phosphate transfection was conducted using 2x HEPES buffered saline (50 mM HEPES, 10 mM KCl, 12 mM dextrose, 280 mM NaCl, 1.5 mM Na₂PO₄, pH adjusted to 7.4 using HCl, and sterile filtered) and 2 M CaCl₂ solution. Initially, a calcium chloride-plasmid mixture was made with 1 (up to 2) µg of plasmid and 24.4 µmol (12.2 µl of 2 M) CaCl₂ made up to a final volume of 100 µl with sterile-filtered nuclease-free water. This mixture was then added drop-wise into 100 µl of 2x HEPES buffered saline and left to incubate for 10 minutes at room temperature. Following this incubation, the resultant mixture was added drop-wise into the well, aiming to cover the whole well. The cells were then incubated a humidified incubator kept at 37°C and 5% CO₂. After 24 hours, the media is replaced with fresh DMEM+10%FBS media. After a total of 24 to 72 hours after the start of transfection, the cells are suitable for further studies. This working volume of plasmids and reagents were optimized to a well in a 6-well plate (approx. cell growth area 10 cm²), and subsequent work is scaled up or down dependent on cell culture growth area.

Liposome-based transfection (X-tremeGENE, Roche)

The liposome-based transfection with the proprietary X-tremeGENE reagent (Roche, 06366244001) was conducted as per manufacturer's recommendations (http://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/06989195001_03.13.pdf), as summarised below.

For every well, 1 (up to 2) μg of plasmid is added to 200 μl of basal media, in the context of the work for this thesis, DMEM (Sigma, M4530). Following that, 3 μl of the X-tremeGENE reagent was added to the plasmid/DMEM mixture with pipette mixing, and allowed to incubate for 20 minutes at room temperature. Following incubation, the mixture was added drop-wise into the well, aiming to cover the whole well. The cells were then incubated a humidified incubator kept at 37°C and 5% CO₂ for (24 to) 48 hours without the need to replace the culture media. This working volume of plasmids and reagents were optimized to a well in a 6-well plate (approx. cell growth area 10 cm²), and subsequent work is scaled up or down dependent on cell culture growth area.

There is overexpression of RNA (as detected by RT-PCR) in A10 cells after 24 hours post-transfection (Figure 42).

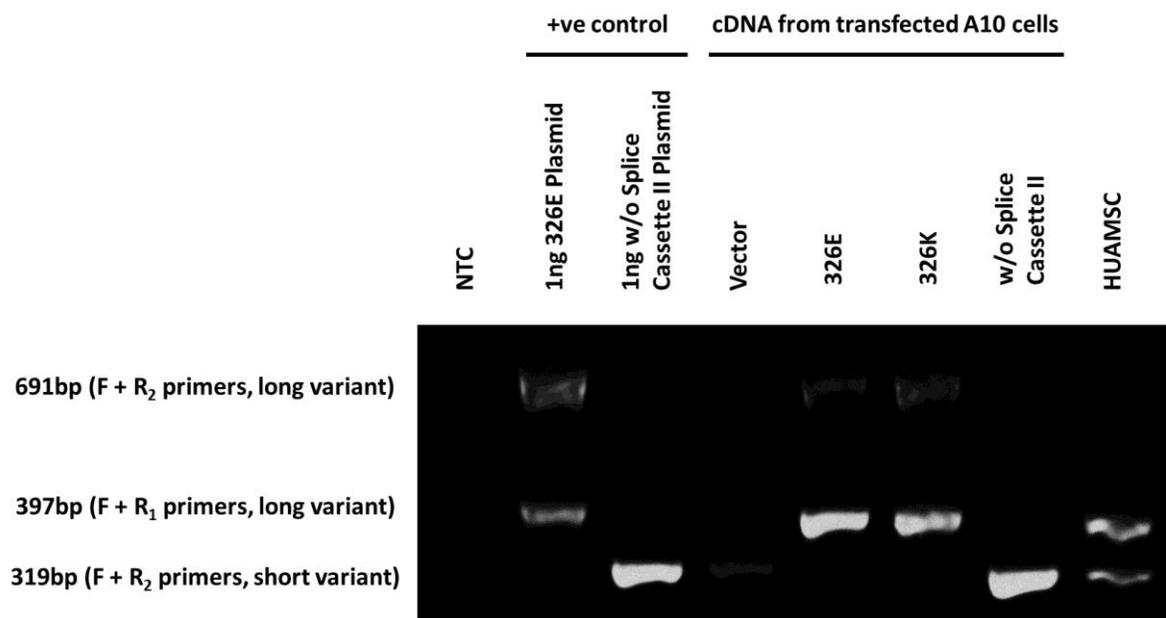


Figure 42: RT-PCR for transfected of A10 cells

End-point RT-PCR using 10 ng of reverse-transcribed RNA from transfected A10 cells and untransfected HUAMSC cells. Three primers (F, R₁ and R₂, see Section 2.21.2) were used concurrently. 1 ng of plasmid served as positive control marking appropriate product sizes. A10 cells were transfected with 1 ng plasmid per 10 cm² cell culture surface (326E, 326K and Spl or with vector control) for 24 hours.

These conditions produced an overexpression of NBCn1 of the appropriate sizes, with either 2 μg for 24 hours (Figure 43), or 1 μg for 48 hours (Figure 44)

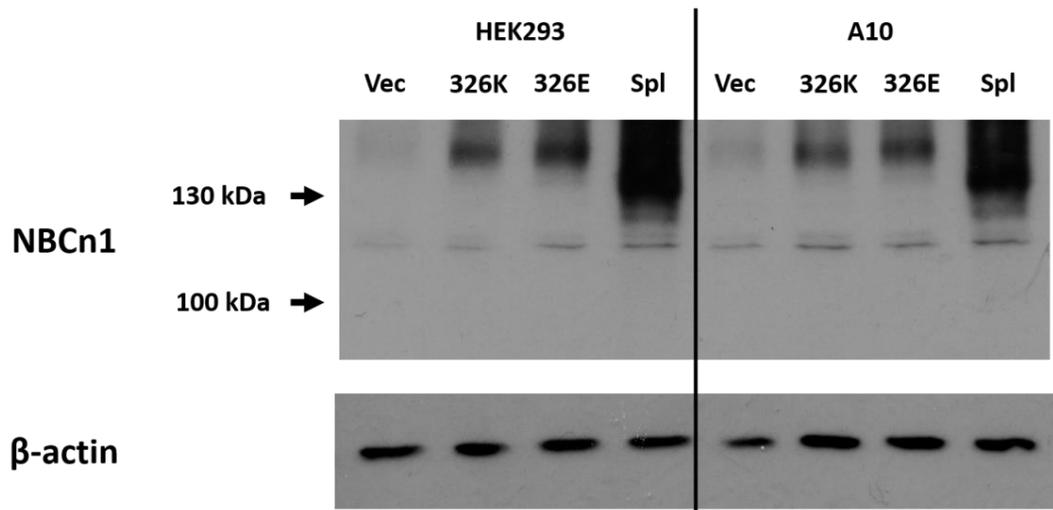


Figure 43: Protein immunoblots for transfected HEK293 and A10 cells
HEK293 and A10 cells were transfected with 2 ng plasmid per 10 cm² cell culture surface (326E, 326K and Spl or with vector control) for 24 hours.

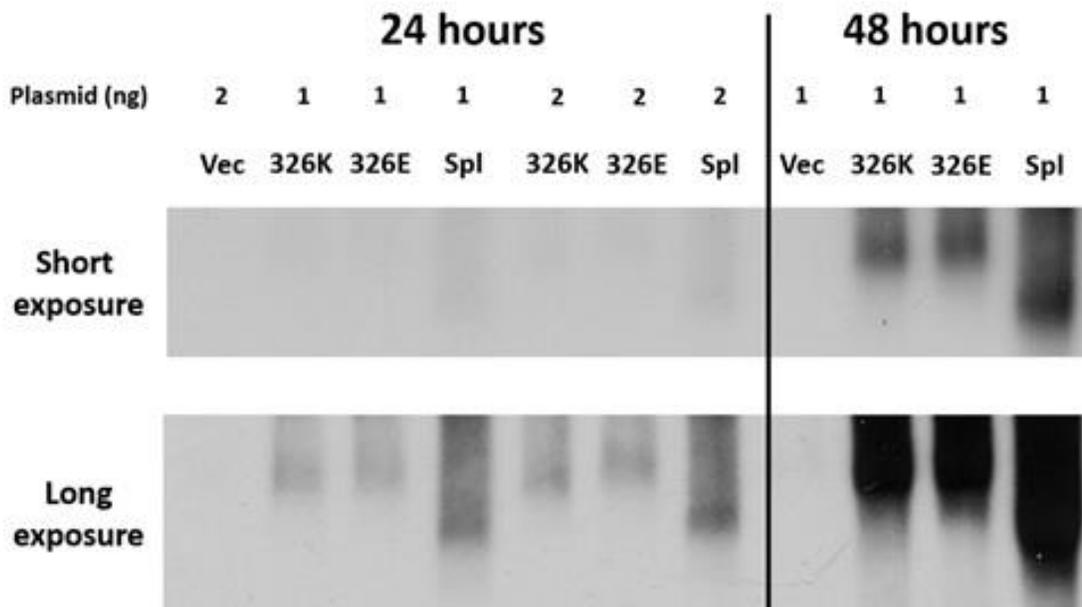


Figure 44: Protein immunoblots for transfected HEK293 with dose and time variation
HEK293 cells in 6-well plates were transfected with 1 or 2 ng plasmid per 10 cm² cell culture surface (326E, 326K and Spl or with vector control) for 24 or 48 hours.
The short (30 seconds) and long (5 minutes) exposure of the autoradiograph film were done consecutively on the same ECL-treatment. The short exposure allows better differentiation of sizes of the products. The long exposure allows better differentiation of the degree of expression between the conditions.

The overexpression was also confirmed by fluorescent immunocytochemistry (Figure 45). The distribution of signal that is not confined to the plasma membrane is further discussed in Chapter 5.4.2).

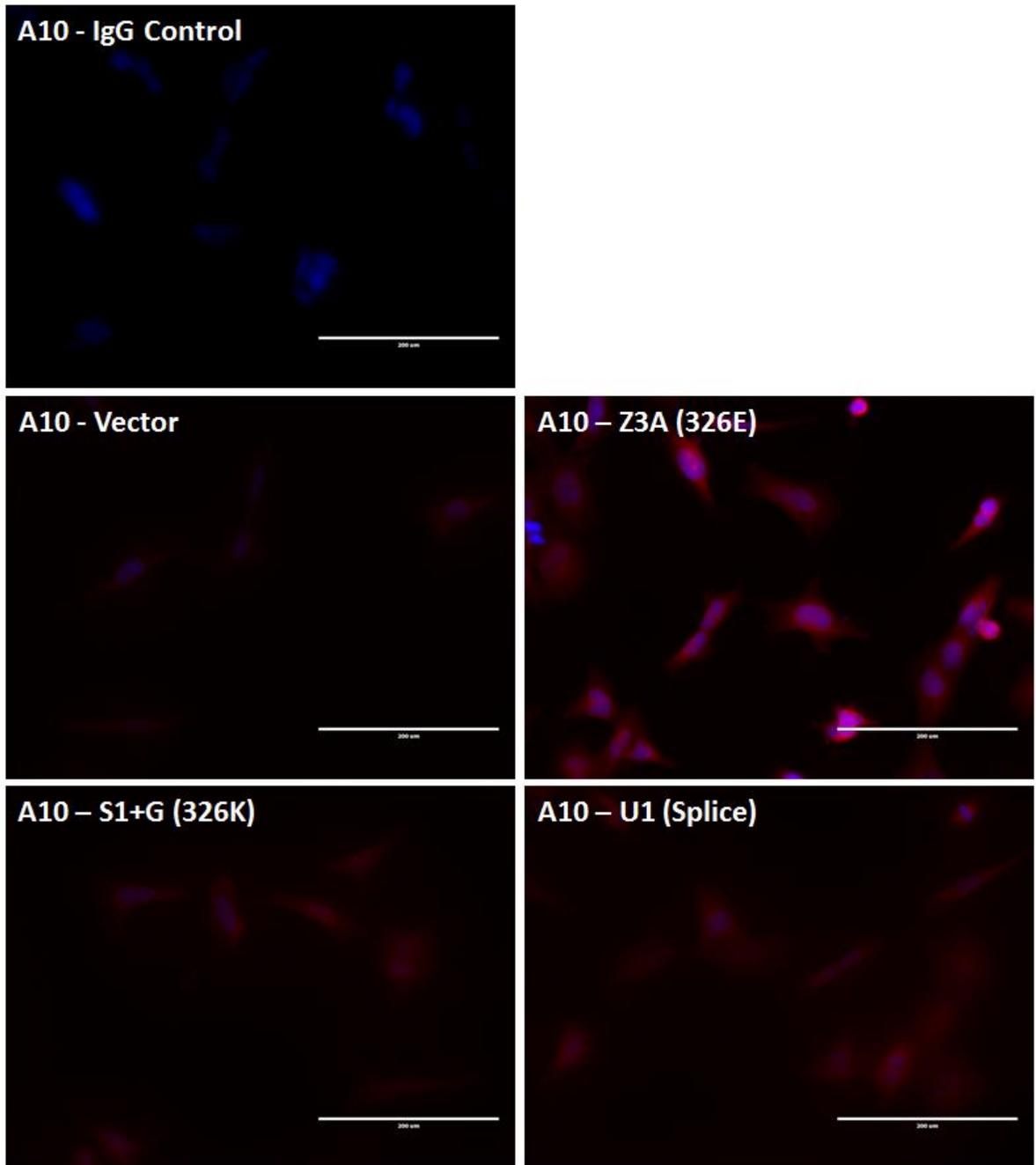


Figure 45: Merged inverted fluorescent microscopy images for transfected A10 cells
Representative images of A10 cells transfected with 2 μg of plasmid per 10 cm² cell culture surface using the X-tremeGENE reagent for 24 hours. Cover slips imaged using the inverted fluorescent

microscope (EVOS® FL Cell Imaging System, Life Technologies). Blue, DAPI; red, NBCn1. Scale bars represent 200 μm .

Expression plasmids and the range of NBCn1 variants

It should be recognised that there are several different variations to NBCn1 apart for Cassette II. The expression plasmids utilised here contain the MERF N-terminal sequence, and also contains Cassette I, III and IV. Although the MERF sequence is classically regarded as “human” (as opposed to the “classically murine” MEAD sequence), it has been recently shown that both MERF and MEAD sequences are expressed in human cells. Additionally, with the other cassettes, there are now 10 known full length human NBCn1 variants, where the absence/presence of these cassettes may alter surface availability and intrinsic transporter function (Liu *et al.*, 2013). It should be recognised that there is a *theoretical* possibility that the E326K amino acid change behaves differently in depending on the whole NBCn1 sequence.

2.20. Intracellular pH (pH_i), buffering capacity and $\text{Na}^+/\text{HCO}_3^-$ -dependent pH_i recovery

2.20.1. Principles

pH_i measurement

Early methods of measuring pH_i initially involved direct use of pH electrodes. Originally, glass electrodes were used (Caldwell, 1954), but they were too large for most cells except for crustacean muscle fibers or cephalopod giant axons. This method progressed to smaller microelectrodes (Thomas, 1974), but were still limited by the size of the cell, technical difficulties and the inability to assess more than one cell at any one time. This is now very much superseded by the use of pH-dependent fluorescent dyes. A commonly used dye is BCECF-AM ((2',7'-bis-(2-carboxyethyl)-5-(and-6)-carboxyfluorescein-acetoxymethyl ester), as first reported in 1982 (Rink *et al.*, 1982), and it continues to be popular almost a quarter of a decade on as it has some key properties important for pH_i assessment.

Although its active moiety is BCECF, it is formulated as an AM ester to allow for easier usage. The ester derivative itself is non-fluorescent, and membrane permeant. This allows non-invasive loading of the cells. Additionally, intracellular non-specific esterases within viable cells are able to free up the BCECF molecule, thus producing a signal only when present within viable cells. When the hydrolysis of the ester bond occurs, the resultant BCECF with a pK_a of approximately 7.0, has a negative charge at physiological pH, which also aids intracellular retention. However, it is also

worth noting that this property, results in BCECF having a buffering capacity of its own, thus it is important not to overload cells and distort the observed intracellular buffering capacity. The key to the utility of BCECF as a pH indicator is that the fluorescence excitation profile is pH-dependent (Figure 46), allowing the implementation of ratiometric measurement techniques (see Section 2.20.4.). Notably, these dyes have a frequency whereby the signal is dependent on, and preferably linearly correlated, to the pH. It also has an isobestic point, specific wavelength at which the absorbance of a sample does not change with the pH of the sample. This absorbance at the isobestic point accounts for dye loading, thus a ratio of the emitted light following excitation at the two wavelengths allows for calibration of the experiment and eliminates artefacts caused by movement, photobleaching or dye loss.

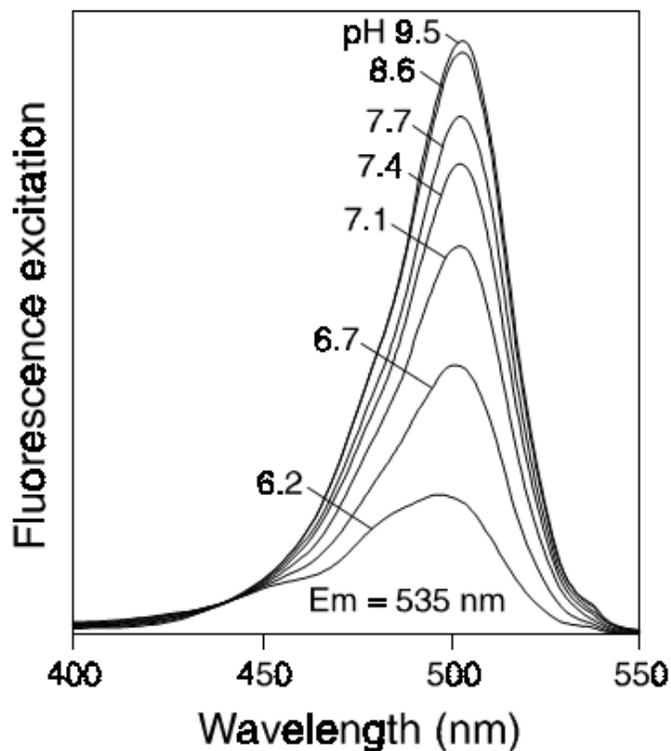


Figure 46: pH-dependent fluorescence excitation spectra of BCECF. Fixed emission wavelength of 535 nm with peak excitations near 490 nm and a pH-independent isobestic point near 440 nm. (Adapted from <https://tools.lifetechnologies.com/content/sfs/manuals/mp01150.pdf>). The studies conducted here utilized excitation wavelengths of 495 and 440 nm.

Carboxy SNARF-1-AM (Spiro[7H-benzo[c]xanthene-7,1'(3H)-isobenzofuran]-ar'-carboxylic acid, 3-(acetyloxy)-10-(dimethylamino)-3'-oxo-, (acetyloxy)methyl ester) is an alternative pH_i indicator,

which like BCECF-AM easily loads into viable cells, and is retained after the non-specific esterases frees up carboxy SNARF-1 molecule. Carboxy SNARF-1 have both dual-emission and dual-excitation properties, making them particularly useful for confocal laser-scanning microscopy, flow cytometry and microplate reader-based measurements. Additionally, the spectra of carboxy SNARF-1 are well resolved from those of multiple Ca^{2+} and Na^+ indicators, permitting simultaneous measurements of pH_i and Ca^{2+} or Na^+ .

Newer fluorimetric microscope imaging systems now allow for assessment of multiple regions within one single microscopy field. With an adequately dye-loaded specimen, it is possible to obtain images with exposures of less than 500 milliseconds, allowing for an *almost* real-time, second-to-second measurement. This frequency of measurements is important for an accurate measurement of pH_i recovery. Using more advanced systems, not used for the work in this thesis, it is even possible to combine with scanning confocal microscopy, although this would be limited by longer duration between full scans.

Ammonium chloride pre-pulse and experimental protocols

To induce intracellular acidification a pre-pulse of ammonium chloride is administered and later washed-out, of which an early version was described in 1974 but using microelectrodes and an external Ringer's solution of pH 8.0 (Thomas, 1974). A protocol closer to current practice was described in 1977, with an external solution of pH 7.7 (Boron and De Weer, 1976), but this time with giant axons of live specimens of the squid *Loligo pealei*, with a relatively high (but not high relative to the squid) extracellular sodium content of 450 mM.

After an initial period of baseline measurements (Figure 47, a), the addition of NH_4Cl to the extracellular solution initially leads to the equilibrium of $\text{NH}_4^+ \leftrightarrow \text{NH}_3 + \text{H}^+$. As the plasma membrane is permeable to NH_3 , it rapidly diffuses intracellularly, where it can combine with H^+ , resulting in abrupt intracellular alkalinisation (b). This is followed by a slower acidification as NH_4^+ enters the cells and base extrusion (e.g. via $\text{Cl}^-/\text{HCO}_3^-$ exchange) is activated. Upon washout of extracellular NH_4Cl , the intracellular NH_3 rapidly escapes the cell, leaving H^+ behind and causing intracellular acidification, and the recovery of pH_i from this acidification is initially observed in the absence of sodium (c). In this step, the sodium content within the PSS (physiological salt solution) is replaced with either equimolar *N*-methyl-D-glucammonium (NMDG) chloride or choline bicarbonate. NMDG is an organic cation and choline is a quaternary ammonium cation, both of

which are *relatively* inert and largely impermeable to the plasma membrane. This combination allows the generation of an isotonic, but sodium-free solution. Following the period of exposure to sodium-free solution, the samples are re-introduced to sodium with the original PSS to determine sodium-dependent base uptake (d). The main variations on the initial experiment are conducted in the presence of an NHE inhibitor; or the absence of $\text{CO}_2/\text{HCO}_3^-$ to identify a bicarbonate-dependent component to pH_i recovery (Figure 47).

There are several options for NHE inhibitors, all with different inhibitory potencies and target specificities. An early option was amiloride, but is not as potent an inhibitor of NHE (IC_{50} of 84 μM) and also has the disadvantage of being non-specific at experimental conditions with a Na^+ -channel (ENaC) IC_{50} of 0.34 μM . EIPA (5-(N-Ethyl-N-isopropyl)amiloride), on the other hand, is a more potent inhibitor of NHE (IC_{50} 0.33 μM) but relatively non-specific with regards to the $\text{Na}^+/\text{Ca}^{2+}$ exchanger (NCX) IC_{50} of 129 μM . A potential compromise is that of dimethyl amiloride (DMA) with IC_{50} of 4.3 μM and 550 μM for NHE and NCX respectively (Lesburg *et al.*, 1990). Utilising DMA, instead of EIPA, as the NHE inhibitor has the additional benefit of being easier to handle logistically, being able to be frozen in stock solutions.

For $\text{CO}_2/\text{HCO}_3^-$ -free studies, the bicarbonate within the solution is replaced with equimolar sodium chloride (or NMDG chloride for sodium- and bicarbonate-free solution), once again maintaining isotonicity. Instead of bubbling CO_2 gas into the solution, air (with the nominal absence of CO_2) is used instead. HEPES is used as one of the pH buffering agents in these studies. It is an organic zwitterion with strong buffering capacity in the physiological range (6.8 to 8.2), and is typically independent of carbon dioxide for its buffering potential. This makes it an appropriate basis for the $\text{CO}_2/\text{HCO}_3^-$ -free buffers. It is notable that HEPES has been found to have inhibitory effects on ion channels, for example Cl^- channels (Yamamoto and Suzuki, 1987) and to control for such potential effects; HEPES is also included in the $\text{CO}_2/\text{HCO}_3^-$ -containing solutions.

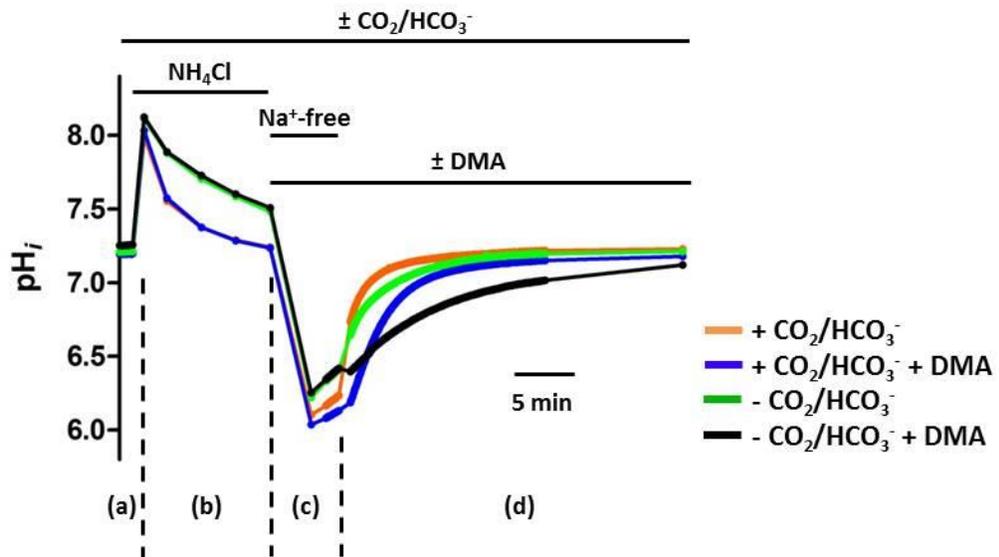


Figure 47: Experimental protocol investigating influence of NHE (by DMA inhibition) and NBC (by removal of CO₂ and HCO₃⁻) in HUVECs and HUASMCs (a) Initial baseline pHi, (b) abrupt intracellular alkalinisation with ammonium chloride with initial rapid influx of NH₃ and later slower acidification with NH₄⁺ entry, (c) washout of ammonium chloride results in intracellular NH₃ rapidly escapes the cell, leaving H⁺ behind and causing intracellular acidification, and the recovery of pHi from this acidification is initially observed in the absence of Na⁺ and (d) where the recovery is observed in the presence of Na⁺. Representative figure from HUASMCs, and reflects the format of representing results for the rest of the thesis.

Separate to the above studies, HUASMCs were also cultured for 48 hours in serum-free media prior to experimentation attempting to limit any differences in cell cycle stages (Griffin, 1976). Notably, serum starvation has previously been shown to reduce sodium-dependent recovery from intracellular acidosis in Caco-2 human epithelial colorectal adenocarcinoma cells (Watson *et al.*, 1992). This study was conducted before the identification of NBCs, and therefore the finding was attributed to a post-translational effect on NHEs. However, studies were performed in the presence of CO₂/HCO₃⁻, therefore could also be related to an effect on NBCs. This was not possible for HUVECs, as culture in serum-free media resulted in cell death, as reflecting previous studies (Karsan *et al.*, 1997).

Other variations for HUASMCs include the addition of 10 μM of the calcineurin inhibitor FK506 (or equivolume 2 μl DMSO vehicle control) at the point of ammonium chloride washout, with (or with equivolume 2 μl DMSO vehicle control) incubation with 30 μM of the intracellular calcium chelator 1',2'-Bis(2-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid tetrakis-acetoxymethyl ester (BAPTA-AM) for 20 minutes before the start of the experiment (Figure 48). The concentration of FK506 and

BAPTA-AM used were reflective of previous studies on VSMCs of rat mesenteric isolated arteries (Danielsen *et al.*, 2013). Additionally, the effective concentration and loading duration was also confirmed to have a sufficient influence on intracellular calcium of HUASMCs by Fura-2 (F-1201, Fisher Scientific) studies (Figure 49).

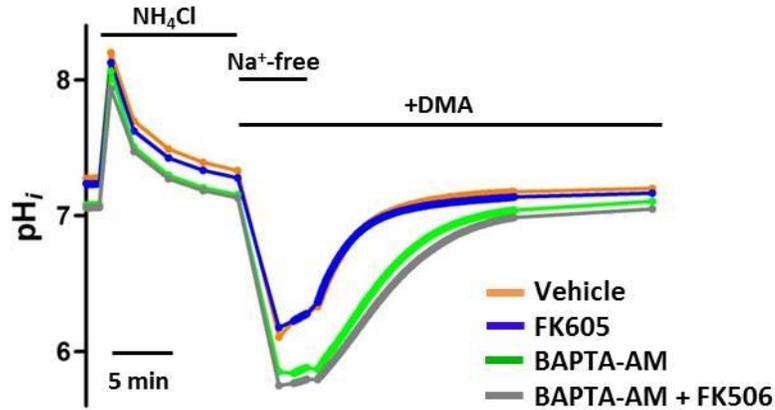


Figure 48: Experimental protocol investigating influence of calcineurin (by FK506 inhibition) and intracellular Ca²⁺ (by BAPTA-AM chelation) in HUASMCs

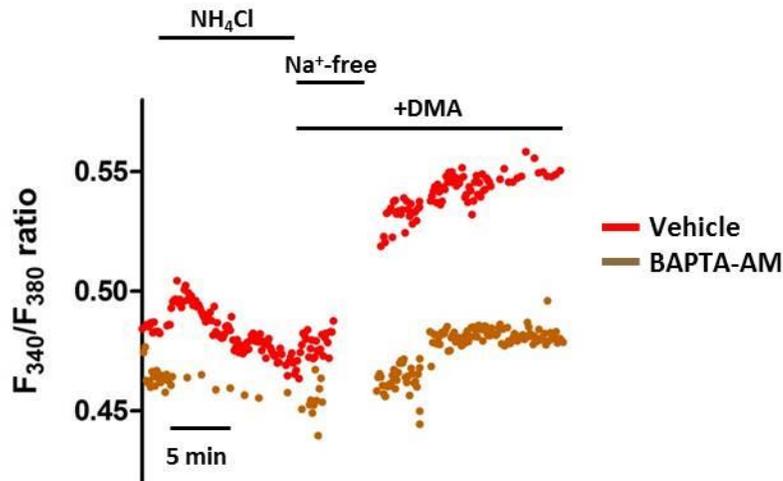


Figure 49: Free intracellular calcium in response to ammonium prepulse in HUASMCs with and without BAPTA-AM loading
30 μM BAPTA-AM loading for 20 minutes is sufficient to attenuate increases in intracellular calcium in response to the alkalinisation, acidification and depolarisation of HUASMCs during ammonium chloride prepulse. The ratios were not calibrated to absolute values of intracellular free Ca²⁺ as this was not necessary to demonstrate an effect of BAPTA-AM.

Of note, the studies demonstrate a sodium-independent component of pH_i recovery (Figure 50), which is also independent of CO₂/HCO₃⁻ (Figure 50, lower panel). The potential mechanisms for this are not further investigated for this thesis, although the functional presence of

monocarboxylate transporters (inhibited by its specific blocker α -cyano-4-hydroxycinnamate) have been demonstrated in an *in vitro* assessment of human radial artery smooth muscle cell explants (Lee *et al.*, 2014).

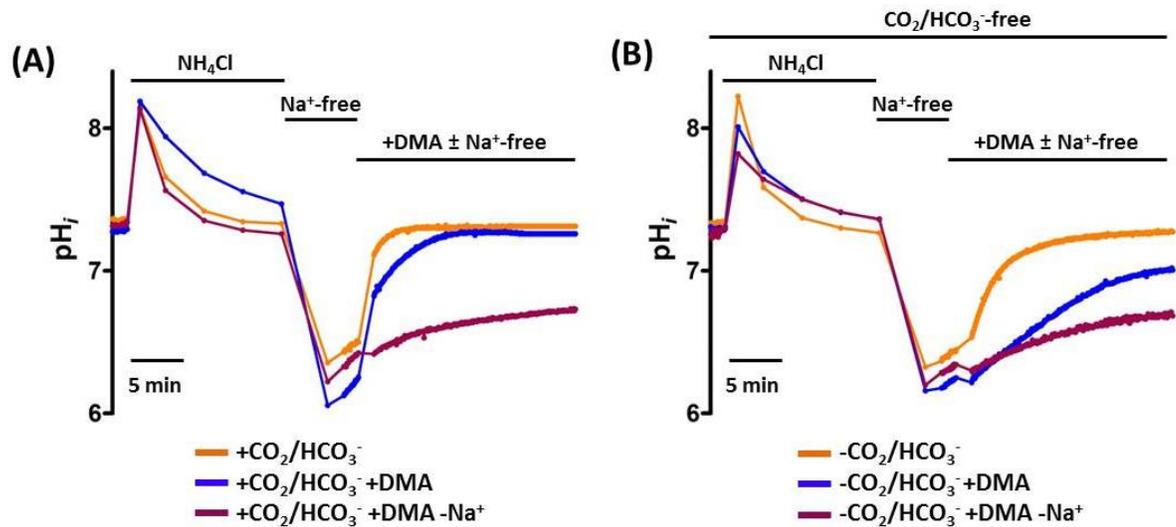


Figure 50: Na^+ -independent pH_i recovery

Paired experiments on HUVECs. Note the purple curves (denoting the absence of Na^+ following ammonium chloride washout) are nearly overlapping in the (A) presence or (B) absence of $\text{CO}_2/\text{HCO}_3^-$.

Cell culture surface, culture substrate and cell type selection

Experiments were initially conducted with cells cultured on 0.04% w/v gelatin-coated glass coverslips. Under these condition, cells were prone to detachment, particularly after ammonium chloride exposure. Furthermore, this is exacerbated with HUASMCs which retain their contractile abilities. Cell detachment is markedly reduced when using the cell culture surface of 0.1% w/v poly-lysine coated-polystyrene flasks. Further to improving cell attachment, NBC activity, as represented by pH_i recovery in PSS (+DMA) buffer is more apparent with the poly-lysine/polystyrene cell culture surface (Figure 51). Hence, all studies in this thesis are conducted on cells cultured lysine/polystyrene cell culture surface.

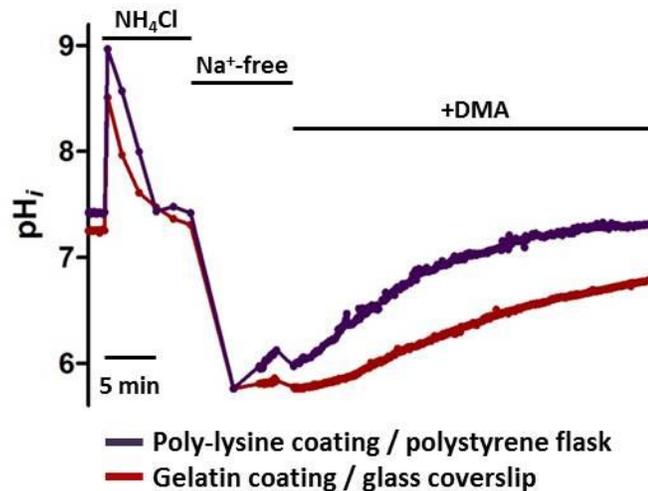


Figure 51: Effect of cell culture surface and substrate on pH_i recovery
 Representative results for 1 HUASMC, 2 HUVECs and 1 A10 cell cultures on different cell culture surfaces. The gelatin-coated glass coverslip culture surface consistently had lower resting pH_i and rate of recovery as compared to poly-lysine-coated polystyrene flasks, across all cell types.

Two experimental cell lines were not followed up for different reasons. HEK293 cells shrank significantly upon addition of the standard 20 mM NH_4Cl prepulse. This was even observed with concentrations of NH_4Cl as low as 5 mM. On top of this problem, HEK293 cells were not tolerant of the multiple washes that was required as part of this protocol. NIH3T3 fibroblasts did not have any less endogenous sodium- and bicarbonate-dependent recovery from intracellular acidosis as compared to A10 cells and thus not further utilised.

Calibration curves for pH_i

Calibration of the initial fluorescence output (F_{495}/F_{440} ratio) to a numerical pH value typically utilises nigericin, which acts as a K^+/H^+ -exchanger, allowing equilibration of these ions across the plasma membrane (Harold and Baarda, 1968). In the presence of nigericin, bathing the sample with solutions of known pH will equilibrate with the intracellular space, and allows the dye-measured ratio to equate to a known pH (Chaillet and Boron, 1985) (Figure 52).

The presence of the high K^+ concentration approximates to the intracellular $[K^+]$, which allows nigericin to mediate K^+-H^+ exchange until extracellular $[H^+]$ is equal to intracellular $[H^+]$, and therefore pH_i equals pH_o . The calibration curves for the work done for this thesis are further explored in Section 2.20.4.

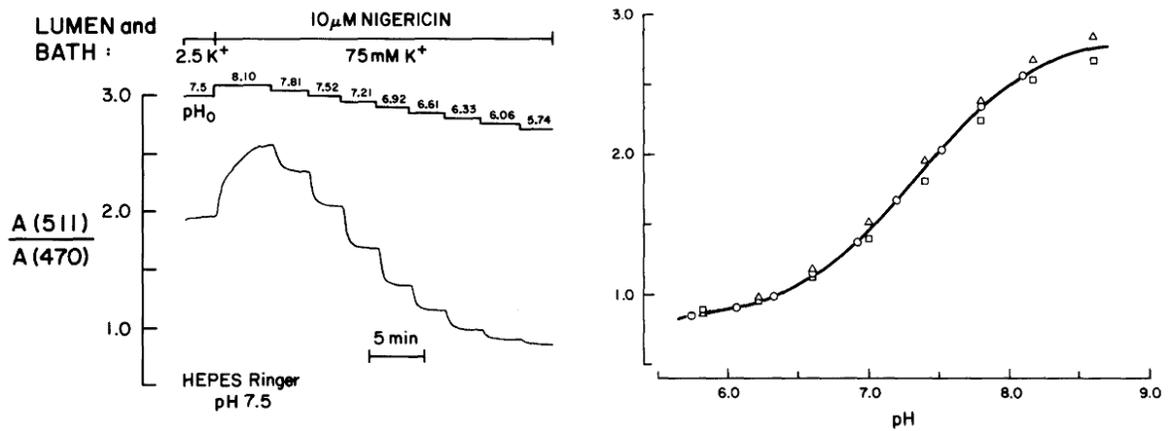


Figure 52: Example of nigericin and pH clamping calibration Adapted from Chaillet and Boron (1985). Note the difference from current protocol of the measured fluorescence ratio (due to use of different dye) and the concentration of extracellular K^+ (due to different cell type), but the principle remains the same. It also refers to “lumen” as the sample assessed was a proximal tubule from a salamander.

Limitations

Perhaps the most important limitation of this methodology is the dependence on the relevance of the samples assayed. With an *in vitro* assay, there are always the concerns of extrapolating to *in vivo* situations (as explored in Section 2.1.1) and the lack of interactions between adjacent cell types (as explored in Section 1.6). The work in this thesis also found a major influence of the cell culture surface and substrate (Figure 51), indicating that different experimental conditions may influence pH_i recovery. With that, consistency in conditions must be ensured, and particularly relevant when comparing results across studies/publications.

2.20.2. Materials

All solutions were warmed to 37°C and bubbled with the corresponding gas for 30 minutes prior to use.

Physiological saline solution (PSS) had a final formulation of 116 mM NaCl, 2.82 mM KCl, 25 mM NaHCO_3 , 1.6 mM CaCl_2 , 1.2 mM MgSO_4 , 1.18 mM KH_2PO_4 , 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with 5% CO_2 in air. Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 141 mM Na^+ , 122 mM Cl^- , 25 mM HCO_3^- , 10 mM HEPES, 5.5 mM glucose, 4 mM K^+ , 1.6 mM Ca^{2+} , 1.2 mM Mg^{2+} , 1.2 mM SO_4^{2-} , 1.18 mM H_2PO_4^- and 0.03 mM EDTA.

Bicarbonate-free physiological saline solution (-HCO₃⁻ PSS) had a final formulation of 141 mM NaCl, 2.82 mM KCl, 25 mM NaHCO₃, 1.6 mM CaCl₂, 1.2 mM MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with air (nominally 0% CO₂). Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 141 mM Na⁺, 147 mM Cl⁻, 10 mM HEPES, 5.5 mM glucose, 4 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

Sodium-free physiological saline solution (-Na⁺ PSS) had a final formulation of 116 mM NMDG (N-methyl-D -glucamine), 2.82 mM KCl, 25 mM choline bicarbonate, 1.6 mM CaCl₂, 1.2 mM MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with 5% CO₂ in air. Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 116 mM NMDG⁺, 25 mM choline⁺, 122 mM Cl⁻ (in theory after titrating pH Cl), 25 mM HCO₃⁻, 10 mM HEPES, 5.5 mM glucose, 4 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

Sodium- and bicarbonate-free physiological saline solution (-Na⁺/-HCO₃⁻ PSS) had a final formulation of 141 mM NMDG, 2.82 mM KCl, 25 mM NaHCO₃, 1.6 mM CaCl₂, 1.2 mM MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with air (nominally 0% CO₂) with HCl. Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 141 mM NMDG⁺, 147 mM Cl⁻ (in theory after titrating pH Cl), 10 mM HEPES, 5.5 mM glucose, 4 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

High-potassium physiological saline solution (High-K⁺ PSS) had a final formulation of 68.8 mM NaCl, 50 mM KCl, 25 mM NaHCO₃, 1.6 mM CaCl₂, 1.2 mM MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with 5% CO₂ in air. Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 93.8 mM Na⁺, 122 mM Cl⁻, 25 mM HCO₃⁻, 10 mM HEPES, 5.5 mM glucose, 51.2 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

Sodium-free, high-potassium physiological saline solution (-Na⁺/High-K⁺ PSS) had a final formulation of 68.8 mM NMDG, 50 mM KCl, 25 mM choline bicarbonate, 1.6 mM CaCl₂, 1.2 mM

MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of 7.45 at 37°C while bubbled with air (nominally 0% CO₂). Calcium chloride was the last to be added after achieving the desired pH. This provides a final concentration of 68.8 mM NMDG⁺, 25 mM choline⁺, 122 mM Cl⁻, 25 mM HCO₃⁻, 10 mM HEPES, 5.5 mM glucose, 51.2 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

In calibrating the experiments, a range of bicarbonate-free, high-potassium PSS calibration (Calibration PSS) were used. The buffers had a final formulation of 11 mM NaCl, 132.8 mM KCl, 1.6 mM CaCl₂, 1.2 mM MgSO₄, 1.18 mM KH₂PO₄, 10 mM HEPES, 0.03 mM EDTA and 5.5 mM glucose. The solution was adjusted to pH of a range between 6.0 and 8.0 at 37°C while bubbled with air (nominally 0% CO₂). Calcium chloride was the last to be added after achieving the desired pH. Just before use, nigericin (N7143, Sigma) was added to achieve a final concentration of 5 mg/l. This provides a final concentration of 11 mM Na⁺, 147 mM Cl⁻, 10 mM HEPES, 5.5 mM glucose, 134 mM K⁺, 1.6 mM Ca²⁺, 1.2 mM Mg²⁺, 1.2 mM SO₄²⁻, 1.18 mM H₂PO₄⁻ and 0.03 mM EDTA.

Poly-L-lysine (P8920, Sigma) was used at a working concentration of 0.1% w/v, covering the cell culture surface at 0.5ml for each 12.5 cm² cell culture flask. This was allowed to rest at room temperature for 5 minutes, before being aspirated. The flask was then washed with sterile distilled deionised water and aspirated to dryness. The flask was then left to dry at 37°C for 2 hours.

In studies with the calcineurin inhibitor FK506 (F4679, Sigma), a final concentration of 10 μM is used. Additionally, to increase the intracellular calcium levels, a 50 mM K⁺ solution is used where the increased K⁺ is provided by KCl in direct equimolar replacement of NaCl (or NMDG). Dimethyl amiloride (A4562, Sigma) was used at a final concentration of 10 μM.

In studies with transfected cells, transfection with the lipofectamine method is as previously described (Section 2.19.).

2.20.3. Methods

Equilibration and set-up

The desired cells were sub-cultured on polystyrene cell culture flasks with 12.5 cm² cell culture surface (#353107, BD-Falcon) pre-coated with 0.1% w/v poly-l-lysine. Once at the desired confluence, the cultured cells were ready to be assayed.

The experiments were conducted in a temperature controlled set-up with the stand and ambient air warmed up to 37°C for 30 minutes before initiation of experiments. Just before starting the experiments, two ports were created on the top of the culture flask, one to allow for aspiration and the other for bubbling of gasses (5% CO₂ / 95% air or 100% air depending on study). The flask was then placed in a custom manufactured holder to ensure stability throughout the study. Once in place the culture flask washed twice with PSS (or -HCO₃⁻ PSS), and the relevant gas bubbled through at a constant rate. After the washes, the flask was filled with 2 ml of PSS (or -HCO₃⁻ PSS) with 2 µl of 2mM stock BCECF-AM (B-1170, ThermoFisher) dissolved in DMSO, to reach a final concentration of 1 µM.

The cells were incubated with the BCECF-AM for 20 minutes. After BCECF-AM loading, the adherent cells were washed twice with PSS (or -HCO₃⁻ PSS), and the recordings were initiated (Visiview Version 3, Visitron Systems).

Experimental protocols

Experimental protocols for HUASMCs and HUVECs were otherwise as described in Section 2.20.1., and can be summarised as 20 mM ammonium chloride prepulse:

- In PSS buffer,
- In PSS buffer followed by 30 µM DMA (NHE inhibition),
- In CO₂/HCO₃⁻-free PSS (NBC inhibition), and
- In CO₂/HCO₃⁻-free PSS (NBC inhibition) followed by 30 µM DMA (NHE inhibition).

For HUASMCs, additional studies can be summarised as 20 mM ammonium chloride prepulse:

- In PSS buffer followed by 30 µM DMA (NHE inhibition) for cells cultured in serum-free media for 48 hours,
- With 2 µl equivolume DMSO vehicle incubation for 20 minutes, in PSS buffer followed by 30 µM DMA (NHE inhibition) and 2 µl equivolume DMSO vehicle,

- With 2 μ l equivolume DMSO vehicle incubation for 20 minutes, in PSS buffer followed by 30 μ M DMA (NHE inhibition) and 10 μ M FK506 (calcineurin inhibition),
- With 30 μ M BAPTA-AM loading for 20 minutes (intracellular calcium chelation), in PSS buffer followed by 30 μ M DMA (NHE inhibition) and 2 μ l equivolume DMSO vehicle, and
- With 30 μ M BAPTA-AM loading for 20 minutes (intracellular calcium chelation), in PSS buffer followed by 30 μ M DMA (NHE inhibition) and 10 μ M FK506 (calcineurin inhibition).

Paired digital images at excitation wavelengths of 495 nm and 440 nm were obtained with exposure times of 800 and 1000 milliseconds respectively. These images and their corresponding fluorescence measurements were obtained every 3 seconds at times of interest, otherwise at 15 second intervals. Paired videos for both 495 nm and 440 nm excitation wavelengths are shown in Appendix 8a and 8b.

Experimental protocols for transfected A10 cells were otherwise as described in Section 2.20.1., and can be summarised as 20 mM ammonium chloride prepulse:

- In PSS buffer followed by 30 μ M DMA (NHE inhibition), and
- In PSS buffer followed by 30 μ M DMA (NHE inhibition) and 2 μ l equivolume DMSO vehicle, and
- In PSS buffer followed by 30 μ M DMA (NHE inhibition) and 10 μ M FK506 (calcineurin inhibition).

2.20.4. Analysis

pH_i calibration curves

The calibration curves for the work done for this thesis follows a highly reproducible curve, with coefficients of variation of <5% (Figure 53).

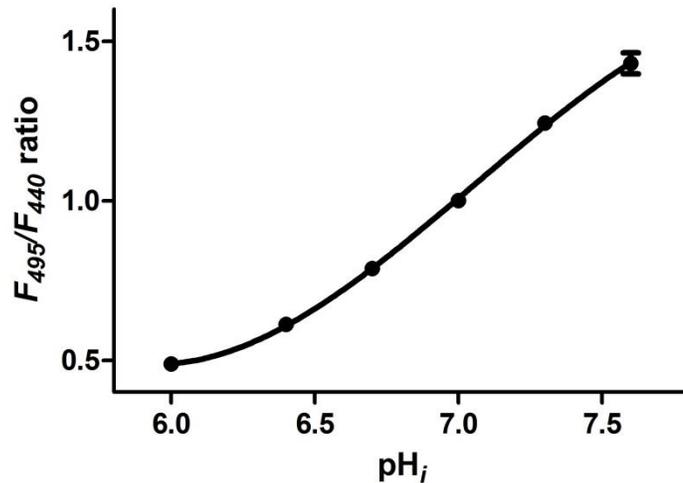


Figure 53: pH_i calibration curve

F₄₉₅/F₄₄₀ ratios (mean ± SEM, n=6 sets) at a range of pH_i normalised to the ratio at pH_i of 7.0. Error bars not easily visualized at datapoints between pH 6.0 and 7.4 due to the very small inter-experiment variation. The curve axes were inverted to derive the cubic polynomial formula (pH_i) = 2.2855x³ - 7.2398x² + 8.7642x + 3.194, where x = normalised F₄₉₅/F₄₄₀ ratio. The relationship is near linear between the pH ranges of 6.5 to 7.5.

Baselines

The initial resting baseline was recorded for 1 minute (20 timepoints at 3 seconds intervals) before the ammonium chloride prepulse. The end resting baseline was noted for studies with inhibitors (DMA or FK506) for 1 minute (20 timepoints at 3 seconds intervals) when the pH_i trace plateaus.

Buffering capacity, pH_i recovery rate and net base uptake

The buffering capacity (mmol/L) was calculated based on the response to wash-out of NH₄Cl using the formula $\beta = \Delta[\text{NH}_4^+]_i / \Delta\text{pH}_i$. The [NH₄⁺]_i was calculated from the Henderson-Hasselbalch equation assuming equilibration of NH₃ across the membrane and using a pK_a of 9.2 (Equation 1 below), and ΔpH_i calculated by the difference in pH_i before and after washout of NH₄Cl (Boetkjer *et al.*, 2006).

$$\text{pH} = \text{pK}_a + \log_{10} \left(\frac{[\text{A}^-]}{[\text{HA}]} \right)$$

Equation 1: Henderson-Hasselbalch equation

The rate of net base uptake (mmol/L/min) during recovery from intracellular acidosis was calculated as the product of the rate of pH_i gain (ΔpH_i per minute calculated by linear regression over 15 seconds) and the buffering capacity (Boetkjer *et al.*, 2006).

2.21. Primers

2.21.1. KASPar genotyping primers

SNP	Primer type (reporter)	Sequence
rs13082711	C allele (VIC)	GAAGGTCGGAGTCAACGGATTAAGATCTGTTTAAAAACAGCTCCCTAC
	T allele (FAM)	GAAGGTGACCAAGTTCATGCTTAAAGATCTGTTTAAAAACAGCTCCCTAT
	Common primer	AGTTTAGCAAAGCTACAGAACAAGGACAT
rs13096477	C allele (VIC)	GAAGGTCGGAGTCAACGGATTACACAAACATCACCACCGGAGG
	T allele (FAM)	GAAGGTGACCAAGTTCATGCTGACACAAACATCACCACCGGAGA
	Common primer	CCCGTCCATCTTCTTTATCTGATCTTTA
rs3755652	C allele (VIC)	GAAGGTCGGAGTCAACGGATTCTGACCTCCAGAAGTTCCCAAG
	T allele (FAM)	GAAGGTGACCAAGTTCATGCTCCTGACCTCCAGAAGTTCCCAAA
	Common primer	TCACTGGCAGGTGAAACCAGTAGTT

Table 20: KASPar genotyping primers

2.21.2. End-point PCR primers

Assay		Sequence	Annealing temp / Mg ²⁺ conc	Amplicon size (bp)
rs13096477 genomic DNA	Forward	CCAATACAATACATTCCCCATAC	57°C / 2mM	459
	Reverse	TGGGCACAATATTCTGTGTAGTG		
rs13096477 complementary DNA	Forward	CCTGGTCTCTGATGAAGAAGC	57°C / 2mM	406
	Reverse	CGCCATCTTCAACATCTCTCT		
SLC4A7 splice (all three primers in a single PCR)	Forward	CTGGCCAATTAGACGAGTCC (Exon 6)	57°C / 2mM	Variable (see Figure 42)
	Rev 1	TCAGCACCCGTAGGAATTTT (Exon 7)		
	Rev 2	CGCTGACTCTTTTGGGAAGT (Exon 8)		
SLC4A4 cDNA	Forward	GGCTTCTTCTCTCCACAGT	57°C / 2mM	129
	Reverse	TTCTTGGTTTTGATGCCGGTG		
SLC9A1 cDNA	Forward	CGGTTCTGGCTGTCTTTGAG	57°C / 2mM	181
	Reverse	ACCACGAAGAAGCTCAGGAA		
β-actin cDNA	Forward	TTCTACAATGAGCTGCGTGTG	57°C / 2mM	122
	Reverse	GGGTGTTGAAGTCTCAAA		
EOMES cDNA	Forward	CCACTGCCACTACAATGTG	55°C / 3mM	480
	Reverse	GACCTCCAGGGACAATCTGA		
NEK10 cDNA	Forward	AATGGCCAATGAAGCTTTTG	57°C / 2mM	377
	Reverse	TACGCTGAGGATGTTTGTCTG		
PHOX2A cDNA	Forward	TTTCGCTGAGACCCACTACC	57°C / 2mM	216
	Reverse	GACTCCTTGAATCGTCGTC		
PLAGL1 cDNA	Forward	GGGACCATTGAAGAATTCCA	57°C / 2mM	2 variants (720 & 416)
	Reverse	GGCTTTGAGGTGGTCCAGTA		
TFAP2C cDNA	Forward	CCACTGAGGTCTTCTGCTC	57°C / 2mM	235
	Reverse	AGAGTCACATGAGCGGCTTT		
TRMT1 cDNA	Forward	AGGAGAGAAGGACACGCAA	57°C / 2mM	290
	Reverse	CTAGGGCAAATCGAATGGAA		

Table 21: End-point PCR primers, annealing temperatures and predicted amplicon sizes

2.21.3. SYBR Green® qRT-PCR primers

Assay		Sequence
SLC4A7 splice *	Common Forward (Exon 6)	CAGTCGGATTCTCTTGTTCGA
	Reverse 1 (Exon 7)	GACCTGTTCGCAAAGAGTGG
	Reverse 2 (Exon 8)	TGTCCAAGTTTCCAGGAGCA
18S	Forward	CCCAGTAAGTCCGGGTCATAA
	Reverse	CCGAGGGCCTCACTAAACC

Table 22: SYBR Green® qRT-PCR primers

* Only one pair (common forward + reverse 1 OR reverse 2) used in any one reaction

2.21.4. EMSA primers

SNP	Allele		Sequence
rs13077400	A	Forward	GAAGATGACCAAGAAGAAGAG
		Reverse	CTCTTCTTCTTGGTCATCTTC
	G	Forward	GAAGATGACCGAGAAGAAGAG
		Reverse	CTCTTCTTCTCGGTCATCTTC
rs13078798	A	Forward	CAACCTCACCAATGAAACTG
		Reverse	CAGTTTCCATTGGTGAGGTTG
	G	Forward	CAACCTCACCGATGAAACTG
		Reverse	CAGTTTCCATCGGTGAGGTTG
rs13082711	C	Forward	CAGCTCCCTACGATGGGGGCC
		Reverse	GGCCCCATCGTAGGGAGCTG
	T	Forward	CAGCTCCCTATGATGGGGGCC
		Reverse	GGCCCCATCATAGGGAGCTG
rs13096477	C	Forward	TATCTTTTCTCCTCCGGTGGT
		Reverse	ACCACCGGAGGAGAAAAGATA
	T	Forward	TATCTTTTCTTCTCCGGTGGT
		Reverse	ACCACCGGAGAAGAAAAGATA
rs1309717	C	Forward	CTGCAACATCCGAAAGAGTC
		Reverse	GACTCTTTCCGGATGTTGCAG
	T	Forward	CTGCAACATCTGAAAGAGTC
		Reverse	GACTCTTTCCAGATGTTGCAG
rs17019809	A	Forward	CCCGTAAACAAATGGAAACC
		Reverse	GGTTTCCATTTGTTTTACGGG
	G	Forward	CCCGTAAACGAATGGAAACC
		Reverse	GGTTTCCATTCGTTTTACGGG
rs2003886	C	Forward	TGCGTCTCTGCGTCTGCAGCT
		Reverse	AGCTGCAGACGCAGAGACGCA
	T	Forward	TGCGTCTCTGTGTCTGCAGCT
		Reverse	AGCTGCAGACACAGAGACGCA
rs2371065	A	Forward	TCCCACCTTCATAATTTTGTT
		Reverse	AACAAAATTATGAAGGTGGGA
	C	Forward	TCCCACCTTCCTAATTTTGTT
		Reverse	AACAAAATTAGGAAGGTGGGA
rs35701251	A	Forward	TTCAGAATACATCCTGAGTAT
		Reverse	ATACTCAGGATGTATTCTGAA
	T	Forward	TTCAGAATACTTCTGAGTAT
		Reverse	ATACTCAGGAAGTATTCTGAA
rs3755652	C	Forward	CGCTGACTCTCTTGGGAACTT
		Reverse	AAGTCCCAAGAGAGTCAGCG
	T	Forward	CGCTGACTCTTTTGGGAACTT
		Reverse	AAGTCCCAAAGAGTCAGCG

Table 23: EMSA primers

Listed in alphabetical order. Where relevant, the forward primer was commercially labelled with biotin.

Competitor primers		Sequence
Non-specific competitor	Forward	AGTACGGGGGTCTCCCCAGA
	Reverse	TCTGGGGAGGACCCCCGTACT
TFAP2C	Forward	TCCCACCTTCCAAATTTTGTT
	Reverse	AACAAAATTTGGAAGGTGGGA
Elk1	Forward	TCCCACCTCCGGCTCTTGTT
	Reverse	AACAAGAGCCGGAAGGTGGGA
c-fos	Forward	TCCCACCTGACTCACTTTGTT
	Reverse	AACAAAGTGAGTCAGGTGGGA
TRMT1	Forward	TCCCACCTTCATTTTTTTGTT
	Reverse	AACAAAAAATGAAGGTGGGA
EBF1 (position 1)	Forward	TCCCACCTTCCCGTTTTTGTT
	Reverse	AACAAAACGGGAAGGTGGGA
EBF1 (position 2)	Forward	TCCCACCTTCCCGTTTTTGTT
	Reverse	AACAAAACGGGAAGGTGGGA

Table 24: EMSA competitor primers

2.22. Antibodies

Type	Name	Mono / poly clonal	Manufacturer	Product No	Working concentration (µg/ml)
1 ^o ry IB / IP / ICC	Rabbit Anti-SLC4A7 IgG antibody	Poly	Abcam	ab82335	1 for IB / IP 5 for ICC
1 ^o ry IB (comparator)	Rabbit Anti-SLC4A7 IgG antibody	Poly	Sigma	HPA035857	1
Primary IB	Goat Anti-PPAR-Abeta (Calcineurin A) IgG antibody	Poly	Santa Cruz	sc-6124	1
Blocking peptide	Immunizing peptide for sc-6124	N/A	Santa Cruz	sc-6214P	40
2 ^o ry IB	Horse anti-mouse IgG, HRP-linked Antibody	N/A	New England Biolabs	7076S	0.1
2 ^o ry IB	Goat anti-rabbit IgG, HRP-linked Antibody	N/A	New England Biolabs	7074S	0.13
2 ^o ry IB	Donkey anti-goat IgG-HRP	N/A	Santa Cruz	sc-2020	0.08
Control antibody IP / ICC	Normal mouse IgG	Poly	Santa Cruz	sc-2025	Variable
Control antibody IP / ICC	Normal rabbit IgG	Poly	Santa Cruz	sc-2027	Variable
1 ^o ry IB	Mouse Anti-GAPDH IgG antibody	Mono	Santa Cruz	sc-47724	0.1
1 ^o ry IB	Mouse Anti-α Tubulin IgM antibody	Mono	Santa Cruz	sc-53646	1
IP	Rabbit Anti-Pol II IgG antibody	Poly	Santa Cruz	sc-9001	10ng per reaction
EMSA Supershift / WEMSA	Rabbit anti-AP-2γ IgG Antibody	Poly	Santa Cruz	sc-8977X	0.05 to 0.25
EMSA Supershift	Rabbit anti-ZAC1 IgG Antibody	Poly	Santa Cruz	sc-22811X	0.05 to 0.25
EMSA Supershift	Goat anti-PHOX2A IgG antibody	Poly	Santa Cruz	sc-13229X	0.05 to 0.25
1 ^o ry IB / ICC	Mouse IgG anti-human VE-Cadherin	Mono	Abcam	ab7047	10 for ICC, 1 for IB
1 ^o ry IB / ICC	Mouse IgG anti-human N-Cadherin	Poly	Abcam	ab18203	1
Fluorescent 2 ^o ry ICC	Donkey anti-rabbit IgG - H&L Alexa Fluor® 594	Poly	Abcam	ab150076	2
Fluorescent 2 ^o ry ICC	Goat anti-mouse IgG - H&L Alexa Fluor® 488	Poly	Abcam	ab150113	2

Table 25: List of antibodies used.

1^ory denotes primary; 2^ory denotes secondary; IB denotes immunoblots; IP denotes immunoprecipitation; ICC denotes immunocytochemistry; EMSA denotes electrophoretic mobility shift assays; WEMSA denotes Western-EMSA.

2.23. Reagents

Thesis section	Reagent / mixture	Formulation
2.1 Tissue culture	Cryopreservation media	70% v/v FBS 20% v/v base media (M199 or DMEM) 10% v/v dimethyl sulfoxide
2.1 Tissue culture	Hanks' Balanced Salt Solution (HBSS)	138 mM NaCl 5.33 mM KCl 4 mM NaHCO ₃ 1.26 mM CaCl ₂ 0.5 mM MgCl ₂ 0.44 mM KH ₂ PO ₄ 0.41 mM MgSO ₄ 0.3 mM Na ₂ HPO ₄ 5.6 mM glucose
2.1 Tissue culture	HUASMC media supplement	Final concentration: 0.5 ng/ml human epidermal growth factor 2 ng/ml human fibroblast growth factor 5 µg/ml human insulin
2.1 Tissue culture	HUVEC media supplement	Final concentration: 2.5 µg/ml human beta-endothelial cell growth factor 2.25 mg/ml endothelial cell growth supplement from bovine neural tissue 1.25 mg/ml thymidine 5,000units/ml heparin from porcine intestinal mucosa
2.2 DNA isolation	Lysis buffer for DNA isolation	10 mM Tris pH 8.0 10 mM EDTA pH 8.0 100 mM NaCl 0.5% w/v sodium dodecylsulfate If digesting tissue, 400 µg Proteinase K was added
2.5 PCR	6x DNA Loading Buffer	60% v/v glycerol 60 mM Tris-HCl pH 8.0 6 mM EDTA pH 8.0 0.06% w/v bromophenol blue
2.8 Protein isolation	Protease inhibitor (PI)	1 mM PMSF 1 µg/ml Leupeptin 1 µg/ml Aprotinin 1 µg/ml Pepstatin
2.8 Protein isolation – Membrane-bound protein	Homogenization Buffer (for membrane prep)	10 mM Tris HCl pH 7.2, 1mM EDTA pH 8.0, 250 mM sucrose
2.8 Protein isolation – nuclear extracts	Hypotonic Lysis Buffer (Buffer A for nuclear extract isolation)	10 mM HEPES pH 7.9 10 mM KCl 0.1 mM EDTA 0.004% v/v NP-40

2.8 Protein isolation – nuclear extracts	Nuclear Lysis Buffer (Buffer B for nuclear extract isolation)	20 mM HEPES pH 7.9 400 mM NaCl 1 mM EDTA 10% v/v glycerol
2.8 Protein isolation – total cellular protein	Radio-Immunoprecipitation Assay (RIPA) lysis buffer	50 mM Tris.HCl pH7.4 150 mM NaCl 0.5% w/v sodium deoxycholate 1% v/v NP-40 0.1% w/v sodium dodecylsulfate
2.10 Immunoblotting	10x Tris-buffered saline (TBS)	24 g Tris-HCl 5.6 g Tris-base 88 g NaCl Diluted into 1 L of ddH ₂ O
2.10 Immunoblotting	Destaining Solution (for Ponceau)	45% v/v isopropanol 10% v/v acetic acid
2.10 Immunoblotting	Enhanced chemiluminescence (ECL)	Solution 1: 1 ml of 250 mM Luminol stock 0.44 ml of 88 mM p-coumaric acid 10 ml of 1 M Tris-base pH8.5 88.56 ml of ddH ₂ O Solution 2: 64 µl of 30% v/v H ₂ O ₂ 10 ml of 1 M Tris-base pH8.5 90 ml of ddH ₂ O Working solution (constitute before use): 1 ml of each solution
2.10 Immunoblotting	Western blot (denaturing) 6x loading buffer	240 mM Tris pH 6.8 12% w/v sodium dodecylsulfate 24% v/v glycerol 6% v/v beta-mercaptoethanol 0.06% w/v bromophenol blue
2.10 Immunoblotting	Western blot 1x Running Buffer	192 mM glycine 25 mM Tris-base 0.1% w/v sodium dodecyl sulphate
2.10 Immunoblotting	Western blot 1x Transfer (Towbin) Buffer	192 mM glycine 25 mM Tris-base 20% v/v methanol
2.10 Immunoblotting, 2.11 DNA pull-down	Ponceau Staining Solution	0.1% w/v Ponceau S 40% v/v isopropanol 10% v/v acetic acid
2.11 DNA pull-down	DNA pull-down Elution Buffer	120 mM Tris pH 6.8 20% w/v glycerol 4% w/v sodium dodecylsulphate After elution, bromophenol blue added to reach final concentration of 0.02% w/v

2.11 DNA pull-down	DNA pull-down Incubation Buffer	50 mM Tris pH 7.0 1 mM EDTA 100 mM KCl 5% v/v glycerol 0.1% v/v Triton-X100
2.11 DNA pull-down	DNA pull-down Wash Buffer	10 mM Tris pH 7.0 1 mM EDTA 2 M NaCl
2.12 Co-Immunoprecipitation	IP Loading Buffer	62.5 mM Tris-HCl pH 6.8 10% v/v glycerol 10% w/v sodium dodecylsulfate 100 mM dithiothreitol After elution, bromophenol blue added to reach final concentration of 0.02% w/v
2.12 Co-Immunoprecipitation	IP Lysis Buffer	25 mM Tris HCl pH 7.2 150 mM NaCl 1 mM EDTA pH 8.0 1% v/v NP-40
2.12 Co-Immunoprecipitation	IP Wash Buffer	25 mM Tris HCl pH 7.2 150 mM NaCl 1 mM EDTA pH 8.0
2.14 EMSA	EMSA (non-denaturing) loading buffer	10 mM Tris pH 7.5 1 mM EDTA 50% v/v glycerol 0.001% w/v bromophenol blue
2.14 EMSA	EMSA oligonucleotide annealing buffer	10 mM Tris pH 8.0 50 mM NaCl 1 mM EDTA
2.16 Immunocytochemistry	Immunocytochemistry Blocking Buffer	10% FBS v/v in TBS
2.16 Immunocytochemistry	Immunocytochemistry Permeabilisation Buffer	0.1% Triton X-100 v/v in TBS
2.16 Immunocytochemistry	Immunocytochemistry Wash Buffer	1% FBS v/v in TBS
2.20 Intracellular pH	Bicarbonate-free physiological saline solution ($-\text{HCO}_3^-$ PSS)	141 mM NaCl 2.82 mM KCl 25 mM NaHCO_3 1.6 mM CaCl_2 1.2 mM MgSO_4 1.18 mM KH_2PO_4 10 mM HEPES 0.03 mM EDTA 5.5 mM glucose
2.20 Intracellular pH	Bicarbonate-free, high-potassium physiological saline solution ($-\text{HCO}_3^-$ /High- K^+ PSS)	11 mM NaCl 132.8 mM KCl 1.6 mM CaCl_2 1.2 mM MgSO_4

		1.18 mM KH_2PO_4 10 mM HEPES 0.03 mM EDTA 5.5 mM glucose
2.20 Intracellular pH	Physiological saline solution (PSS)	116 mM NaCl 2.82 mM KCl 25 mM NaHCO_3 1.6 mM CaCl_2 1.2 mM MgSO_4 1.18 mM KH_2PO_4 10 mM HEPES 0.03 mM EDTA 5.5 mM glucose
2.20 Intracellular pH	Sodium- and bicarbonate-free physiological saline solution ($-\text{Na}^+/-\text{HCO}_3^-$ PSS)	141 mM NMDG chloride 2.82 mM KCl 25 mM NaHCO_3 1.6 mM CaCl_2 1.2 mM MgSO_4 1.18 mM KH_2PO_4 0.03 mM EDTA 5.5 mM glucose
2.20 Intracellular pH	Sodium-free physiological saline solution ($-\text{Na}^+$ PSS)	116 mM NMDG chloride 2.82 mM KCl 25 mM choline bicarbonate 1.6 mM CaCl_2 1.2 mM MgSO_4 1.18 mM KH_2PO_4 10 mM HEPES 0.03 mM EDTA 5.5 mM glucose
Various	10x Phosphate-buffered saline (PBS)	80 g NaCl 2 g KCl 14.4 g Na_2HPO_4 2.4 g KH_2PO_4 Diluted into 1 L of ddH ₂ O

Table 26: Reagents and formulations

2.24. Statistical analyses

Comparisons between two independent groups were conducted using two-sample t-test or the Mann-Whitney U-test for parametric and non-parametric distributions respectively. Comparisons between two paired groups were conducted using one-sample t-test or the Wilcoxon sign-ranked for parametric and non-parametric distributions respectively. Predicted linear relationships were analysed by least-squares linear regression and the derived slopes and y-axis intercepts compared.

Comparisons between multiple paired groups were conducted using repeated-measures analysis of variance (ANOVA). Examination of the influence of two different independent variables (e.g. genotype and dose) on one dependent variable was conducted using two-way ANOVA.

Contingency tables (e.g. comparing observed and predicted genotype frequencies) were assessed by a chi-squared test, or the Fisher's Exact Test, when at least one cell had a value of less than 5.

Graphical presentation and statistical analyses other than power calculations were conducted using Prism v5 (GraphPad Software). Power calculations were conducted using G*Power v3.1.9.2 (Universität Kiel, Germany). Input parameters included an α -value of 0.05, and a sample size to obtain a power of 80%. Two-tailed comparisons were used when comparing two groups.

Values were expressed as mean \pm standard error of the mean, unless stated. When values expressed as a percentage, the total values may occasionally not add up to 100% due to rounding. A p-value of less than 0.05 was considered statistically significant. Multiple testing was further adjusted for by Bonferroni correction.

3. Analysis of single nucleotide polymorphisms (SNP) at the *SLC4A7* locus using genomic prediction tools

The complexity surrounding investigating the numerous SNPs in high LD with the lead SNP (rs13082711) in aforementioned genome-wide association study (Ehret *et al.*, 2011) has already been visited in Section 1.2.

There are many mechanisms whereby a polymorphism can *theoretically* impact on gene expression or protein function. The following sections sets out to utilize a range of prediction tools relating to the 93 polymorphisms of interest - ranging from prediction of eSNPs (expression SNPs), transcription factor binding sites, properties of amino acid changes, post-translational modification, exonic splicing enhancer sites, miRNA (micro RNA) interactions and epigenetic influences. These sections will also aim to describe the potential flaws of the tools and what, *if any*, can be learnt from them.

3.1. Genetic location of rs13082711 and other nearby genes – *NEK10* and *EOMES*

As genome-wide association studies are only able to identify genomic loci, as opposed to specific genes, it is necessary to consider the other genes near the locus, one of which is *NEK10*, encoding the NIMA (never in mitosis gene a)-related kinase 10. It has a low probability as a candidate gene by both positional and functional criteria. This gene is situated 255kb away from the sentinel blood-pressure associated SNP rs13082711 and does not have any SNPs in strong LD with (all $r^2 < 0.2$). Furthermore, the gene product has no clear biological function related to blood pressure. Recently, *NEK10* has been reported to play a role in the cellular response to ultraviolet irradiation (Moniz and Stambolic, 2011), and an nsSNP within *NEK10* was found to be associated with risk of breast cancer (Milne *et al.*, 2014).

The other gene flanking *SLC4A7* is *EOMES*, encoding eomesodermin. It is not a strong positional candidate being 232kb away from rs13082711, and the *EOMES* gene and its flanking regions do not contain any SNPs in LD with rs13082711, even when applying a markedly lowered threshold of $r^2 > 0.2$. *EOMES* has in fact been implicated as a candidate gene in a GWAS for the haematological malignancy, Hodgkin's lymphoma (Frampton *et al.*, 2013). Notably, the rs3806624 locus is more

than 220kb from the *SLC4A7* hypertension loci, with very weak linkage disequilibrium ($r^2 = 0.015$, $d' = 0.232$) (Figure 54).

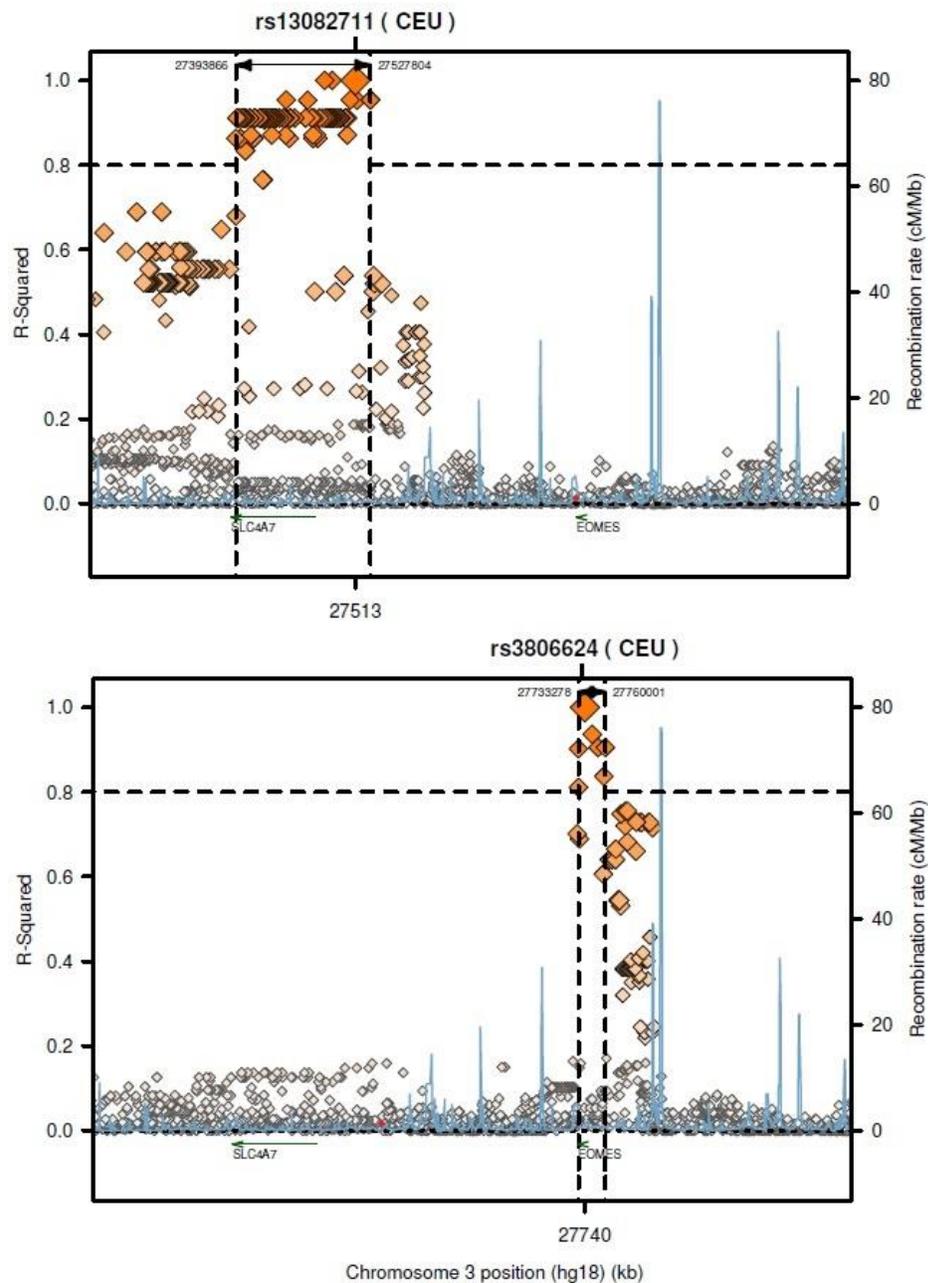


Figure 54: Linkage disequilibrium plots at the *rs13082711* and *rs3806624* SNPs
 Linkage disequilibrium plots at the blood pressure-associated (*rs13082711*) and Hodgkin's lymphoma susceptibility-associated (*rs3806624*) SNPs based on HapMap3 (release 2) data for the CEU population (<http://www.broadinstitute.org/mpg/nap/ldplot.php>). Position along chromosome 3 is matched on both plots. An LD of $r^2 > 0.8$ (dashed lines) is marked as strong linkage disequilibrium within the loci.

Like *NEK10*, the gene product for *EOMES* does not appear to have any clear biological function related to blood pressure. Eomesodermin-deficiency in *EOMES*^{-/-} knockout mice is embryonically lethal, and appears to be essential in trophoblast development and mesoderm formation (Russ *et al.*, 2000). Whilst blood vessels are embryonically derived from the lateral plate of the mesoderm (Risau *et al.*, 1988; Pardanaud *et al.*, 1989), and nephrons from the intermediate mesoderm (Obara-Ishihara *et al.*, 1999), this embryonically lethal knockout does not provide a direct biologically plausible link between *EOMES* and hypertension.

Conditional *EOMES* knockout mice provide further insight into the potential function of its gene product, but currently it does not add much to the understanding of blood pressure regulation. Mice with haematopoietic-specific deficiency of both transcription factors *EOMES* and T-bet failed to develop NK cells (Gordon *et al.*, 2009), and CD8⁺ T cells lacking *EOMES* compete poorly in contributing to the pool of Ag-specific central memory cells (Banerjee *et al.*, 2010). There are currently no published murine models of conditional knockouts relating to the circulatory, renal or neurological systems.

With this, *NEK10* and *EOMES* do not seem to be strong candidate genes in terms of genomic position as well as known protein function. Adding to this hypothesis, the GTEx (Genotype-Tissue Expression project; <http://www.gtexportal.org/home/>) RNA-seq data demonstrates high transcript levels of *SLC4A7*, but low *NEK10* and *EOMES*, in arteries. The results on renal samples are more difficult to interpret as the samples are from the renal cortex, whereas *SLC4A7* is localized to the medulla. Nevertheless, the levels of *NEK10* and *EOMES* in the renal cortex are still lower than *SLC4A7* (Figure 55). Of course, care needs to be taken when using PRKM (reads per kilobase per million reads) as a comparator (Wagner *et al.*, 2012). This is further supported by the finding of this thesis that *NEK10* and *EOMES* transcripts were not found in a panel of tissues key to blood pressure regulation (Results Section, Figure 66).

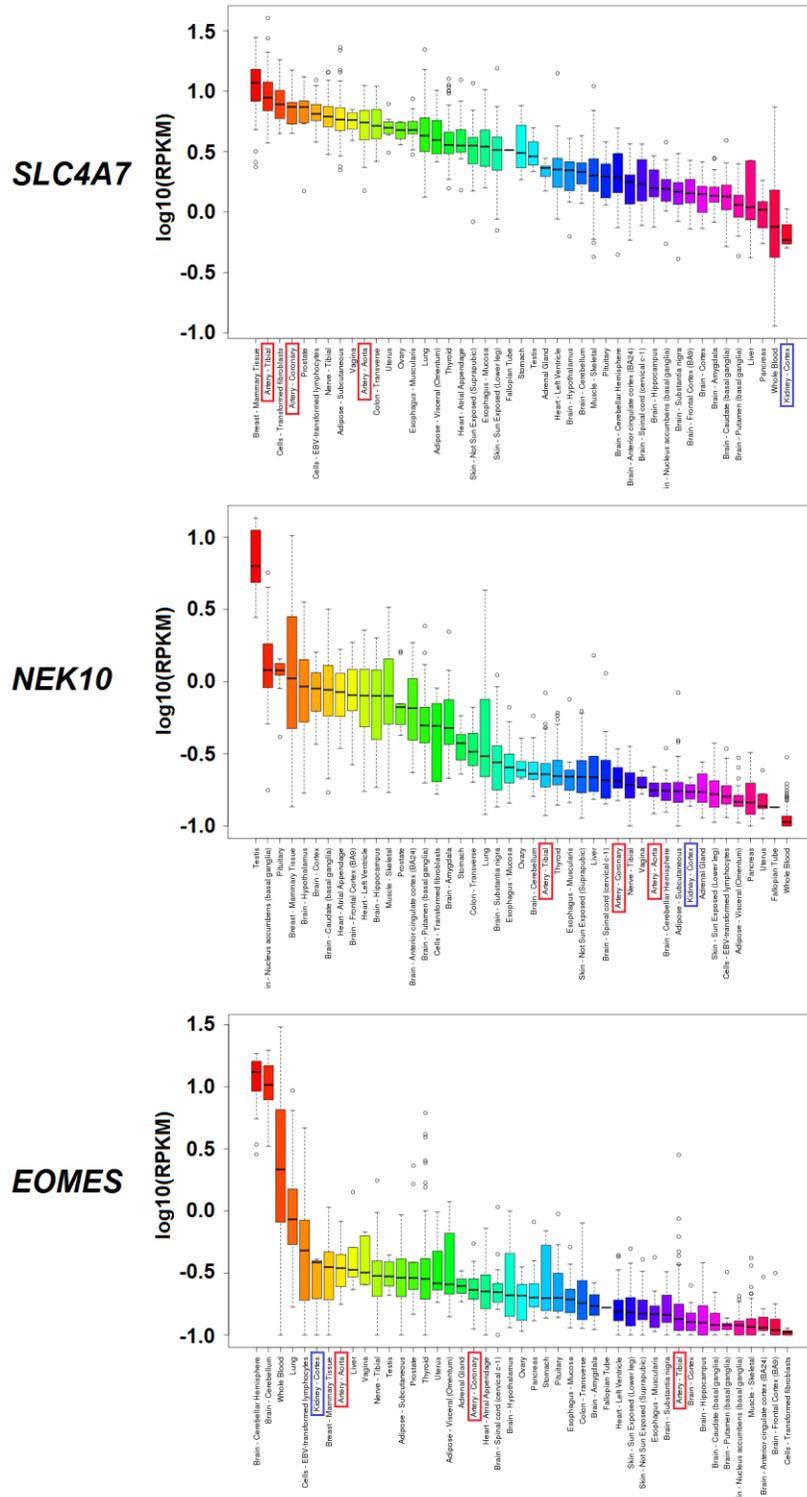


Figure 55: Genotype-Tissue Expression project (GTEx) SLC4A7, NEK10 and EOMES expression data on a panel of human tissue samples.

RPKM denotes reads per kilobase per million reads, as a method of quantifying gene expression by normalizing for total read length and the number of sequencing reads. Coronary artery, tibial artery and aortae indicated in red box; renal cortex indicated in blue box. From <http://www.gtexportal.org/home/>

3.2. Predicted allele frequencies and genotype distributions

In planning future experiments, it is important to consider the genotype distribution of the accessible samples. Based on HapMap data, the minor allele frequency ranges from 2.2% to 20.2%, and a minor homozygote frequency of 0% to 9.6% (Table 27) at rs13082711.

	Sample size	Genotype frequency			Allele frequency	
		C/C	C/T	T/T	C (risk allele)	T
HapMap-CEU	104	0.096	0.212	0.692	0.202	0.798
Study data (Genetically heterogeneous cohort)	404	0.040	0.317	0.643	0.198	0.802
HapMap-CHB	90	0.022	0.000	0.978	0.022	0.978
HapMap-JPT	88	0.000	0.114	0.886	0.057	0.943
HapMap-YRI	116	0.000	0.086	0.914	0.043	0.957

Table 27: Genotype and allele frequency at rs13082711

Genotype and allele frequency of HapMap populations, and from this study's data at rs13082711 in descending risk allele frequency. CEU denotes Utah residents with Northern and Western European ancestry from the CEPH (Centre d'Etude du Polymorphisme Humain) collection; CHB, Han Chinese in Beijing; JPT, Japanese in Tokyo; YRI, Yoruba in Ibadan, Nigeria.

Although in high LD with the GWAS-associated rs13082711, the genotype frequency of the investigated nsSNP rs3755652 should also be considered (Table 28). There is actually more data available for rs3755652 due to the nature of interest as an nsSNP, as compared to the intergenic rs13082711.

	Sample size	Genotype frequency			Allele frequency	
		C/C	C/T	T/T	C (risk allele)	T
HapMap-CEU	226	0.080	0.327	0.593	0.243	0.757
HapMap-GIH	176	0.057	0.341	0.602	0.227	0.773
CS Agilent	1137	0.047	0.350	0.603	0.222	0.778
HapMap-MEX	100	0.020	0.400	0.580	0.220	0.780
Study data (Genetically heterogeneous cohort)	394	0.036	0.362	0.602	0.217	0.783
HapMap-TSI	174	0.011	0.368	0.621	0.195	0.805
ESP Cohort Populations	4552	0.037	0.293	0.669	0.184	0.816
HapMap-MKK	286	0.021	0.189	0.790	0.115	0.885
HapMap-JPT	170	0.000	0.106	0.894	0.053	0.947
HapMap-ASW	96	0.000	0.083	0.917	0.042	0.958
HapMap-LWK	178	0.000	0.079	0.921	0.039	0.961
HapMap-YRI	226	0.000	0.115	0.885	0.037	0.942
HapMap-CHD	170	0.000	0.071	0.929	0.035	0.965
HapMap-CHB	86	0.023	0.000	0.977	0.023	0.977

Table 28: Genotype and allele frequency at rs3755652

Genotype and allele frequency of HapMap populations, and from this study's data at rs3755652 in descending risk allele frequency. CEU denotes Utah residents with Northern and Western European ancestry from the CEPH (Centre d'Etude du Polymorphisme Humain) collection; GIH, Gujarati in Houston, Texas; MEX, Mexican ancestry in Los Angeles; TSI, Tuscans in Italy; MKK, Maasai in Kinyawa, Kenya; JPT, Japanese in Tokyo; ASW African ancestry in Southwest USA; LWK, Luhya in Webuye, Kenya; YRI, Yoruba in Ibadan, Nigeria; CHD, Chinese in Metropolitan Denver; CHB, Han Chinese in Beijing.

CS Agilent denotes cohort of European descent with whole-genome sequencing.

ESP denotes Exome Sequencing Project.

The genotype frequency of the Chinese GWAS identified independent risk SNP rs820430 is also shown in

Table 29.

	Sample size	Genotype frequency			Allele frequency	
		A/A	A/G	G/G	A (risk allele)	G
Study data (Genetically heterogeneous cohort)	383	0.439	0.439	0.123	0.658	0.342
HapMap-CEU	116	0.414	0.448	0.138	0.638	0.361
HapMap-CHB	90	0.178	0.442	0.400	0.389	0.611
HapMap-JPT	90	0.156	0.489	0.356	0.400	0.600
HapMap-YRI	120	1.000	0.000	0.000	0.000	0.000

Table 29: Genotype and allele frequency at rs820430

Genotype and allele frequency of HapMap populations, and from this study's data at rs820430 in descending risk allele frequency. CEU denotes Utah residents with Northern and Western European ancestry from the CEPH (Centre d'Etude du Polymorphisme Humain) collection; CHB, Han Chinese in Beijing; JPT, Japanese in Tokyo; YRI, Yoruba in Ibadan, Nigeria.

Also of note, the overall genotype distribution of our samples does not readily approximate into one particular dataset of HapMap ancestry. This in part reflects the significant ethnicity admixture in the Borough of Tower Hamlets, London, from where the samples are obtained. According to the latest 2011 census (Tower Hamlets Council, 2012) the two largest self-reported ethnic groups are Bangladeshi (32%, not covered by HapMap data), and White British (31%, which in itself is a genetically heterogeneous population). Additionally, 21% of households are multi-ethnic, indicating a large proportion of our samples is derived from mixed ancestry. Importantly, it must also be recognised that each of the CEU, YRI, JPT and CHB datasets represent a very specific population, and commonly misused terminologies such as "Caucasian", "Asian", "East Asian", "Sub-Saharan African", "West African", "Nigerian", or "East Asian" are inaccurate (International HapMap Project, 2005).

3.3. eSNPs (expression single nucleotide polymorphisms) prediction

A potential objective of this thesis is to determine whether the SNPs of interest impacts on *SLC4A7* gene expression. There is already a limited amount of data publically available via the GENEVAR (GENe Expression VARIation) database from the Sanger Institute (<http://www.sanger.ac.uk/resources/software/genevar/>) (Yang *et al.*, 2010), which the public datasets allow interrogation from four different published studies (Grundberg *et al.*, 2012, Stranger *et al.*, 2012, Nica *et al.*, 2011, Dimas *et al.*, 2009), from very specific tissue types.

Out of these datasets, none were found to be an eQTL (expression quantitative trait loci) at the default significance level of $P < 0.001$. However, there were several SNPs that were associated with gene expression at a $P < 0.01$ level (Appendix 9). However, these findings seemingly appear to be present in only certain populations (correlation with *SLC4A7* expression in CEU population, and with *NEK10* in the GIH population). If we are to consider the multiple testing conducted, these associations need to be interpreted with caution.

The limited support from these datasets on our SNPs of interest having an impact on *SLC4A7* gene expression however may not be sufficiently sensitive for several reasons. Most of all, these datasets are derived from relatively small sample size ($n = 60$ to 856). Furthermore, the tissues are of limited relevance in the context of blood pressure regulation – lymphoblastoid cells, skin, adipose tissue, fibroblasts, and mainly from healthy populations or umbilical samples. Additionally, lymphoblastoid cell lines, which forms a large proportion of the samples, are Epstein Barr virus (EBV)-transformed lymphocytes, which in itself alters gene expression profiles (Carter *et al.*, 2002), and even mosaicism (Shirley *et al.*, 2012), limiting its relevance in gene expression studies. Finally, these studies only consider the relationship between SNPs and RNA/cDNA levels; whereas the overall impact on gene expression is far more than gene transcription or RNA stability alone.

A separate source of information on eQTL mapping of SNPs to gene expression comes from SCANDb (SNP and CNV Annotation Database, <http://www.scandb.org>). It utilises data from SNPs and RNA/cDNA from EBV-transformed B-lymphocytes in HapMap CEU and YRI samples (Gamazon *et al.*, 2010). Table 30 shows SCANDb predicting trans-eQTLs (altered gene expression located on different chromosome from SNP of interest) for 9 gene and 2 likely pseudogenes with a pre-set significance level of $P < 0.0001$. Once again, the caveats regarding EBV-transformation on gene expression (Carter *et al.*, 2002; Shirley *et al.*, 2012) still apply.

	CHMP5	CHSY1	DNAJC12	IMPDH2	MIB2	PLEKHA4	PLOD1	SLC9A7	TXNDC13	C10orf59	LOC100130069
rs35523848											
rs67282221											
rs35124794											
rs34234702											
rs13082711											
rs35944308											
rs34348584											
rs13086867											
rs35701251											
rs13097178			■			■				■	
rs2003886											
rs6776318											
rs13081368											
rs73050094											
rs77685358											■
rs7625337			■	■	■						■
rs35512045											
rs57317249											
rs35605663											
rs7644466			■								
rs73050029											
rs7640957											
rs7640929											
rs34422761											
rs73037289											
rs74279676											
rs5792391											
rs67104968											
rs17019809											
rs35262121											
rs6791212											
rs11719386											
rs6773733			■							■	
rs6797389											
rs34611823											
rs35572617											
rs17682751			■							■	
rs13061912											
rs13062327											
rs2371065											
rs2887926		■		■			■			■	■
rs13096477											
rs6777272											
rs66694067											
rs13077400		■	■	■			■			■	■
rs375562			■							■	
rs66743183											
rs34722035											
rs7624790											
rs34307140											
rs34878935											
rs7650165				■			■			■	■
rs923940											
rs923941										■	
rs11717669											
rs11708710											
rs11716531											
rs13100751											
rs73046194											
rs73046191											
rs11718991											
rs34098124											
rs4973772											
rs13089870											
rs34775440											
rs13078798		■	■	■			■			■	■
rs6789539											
rs997680		■	■	■			■			■	■
rs35904501											
rs17682229											
rs3736312											
rs36101969											
rs36116174											
rs34981505											
rs3605621											
rs34209332											
rs3816881			■				■			■	■
rs6775176											
rs2131205			■	■					■	■	■
rs2172297											
rs2840088											
rs2370995		■	■	■			■		■	■	■
rs13075550											
rs4973770		■	■	■			■		■	■	■
rs4973769		■	■	■			■		■	■	■
rs2172298											
rs13019721											
rs77135084											
rs35998108											
rs13082777		■	■	■			■			■	■
rs11709954	■							■		■	
rs7637211											
rs35797288											

Table 30: SCANDb output for gene transcripts that vary with each SNP Summarised SCANDb (<http://www.scandb.org>) output identifying gene transcripts that vary with each SNP (one row each, and rs13082711 identified in grey). Each column represents the gene or pseudogene with altered expression levels. Green cells denotes SNP-dependent expression $P < 0.0001$ (but none $P < 0.00001$), blank cells denote no SNP-expression associated found. The SNP IDs are not intended to be read in this format, as full results in Appendix 10.

Whilst this may pique interest, there are several caveats to remember. All of the proposed trans-eQTLs had a significance level of $P < 0.0001$ (10^{-4}), but > 0.00001 (10^{-5}). It could be argued that for a significant trans-eQTL with an even larger number of multiple testing conducted, the uncorrected significance level should be much lower. There should also be caution exercised with the limited range of populations used. Last of all, the concerns of using lymphoblastoid cell lines as highlighted above also applies.

More recently, the Genotype-Tissue Expression (GTEx) project (<http://www.gtexportal.org/home>) aimed to provide gene expression data (mRNA/cDNA, not protein) across multiple human tissues. The tissue type with most samples at this stage ($n = 60$ to 120) include subcutaneous adipose tissue, tibial artery, blood, cardiac left ventricle, lung, skeletal muscle, tibial nerve, skin (sun exposed) and thyroid tissue. As exemplified in Figure 56 there were no significant effect of the rs13082711 SNP on *SLC4A7* and *NEK10* gene expression after considering the Bonferroni correction for multiple (200,000) hypotheses as set out by the project (GTEx Consortium, 2013). Additionally, the potential direction of effect on *NEK10* is conflicting, being in opposite directions for the tibial artery and left ventricle. Furthermore, all the other tissues studied did not reveal an impact of the polymorphism on neither *SLC4A7* nor *NEK10* expression. An additional caveat to consider is that the magnitude of change is small.

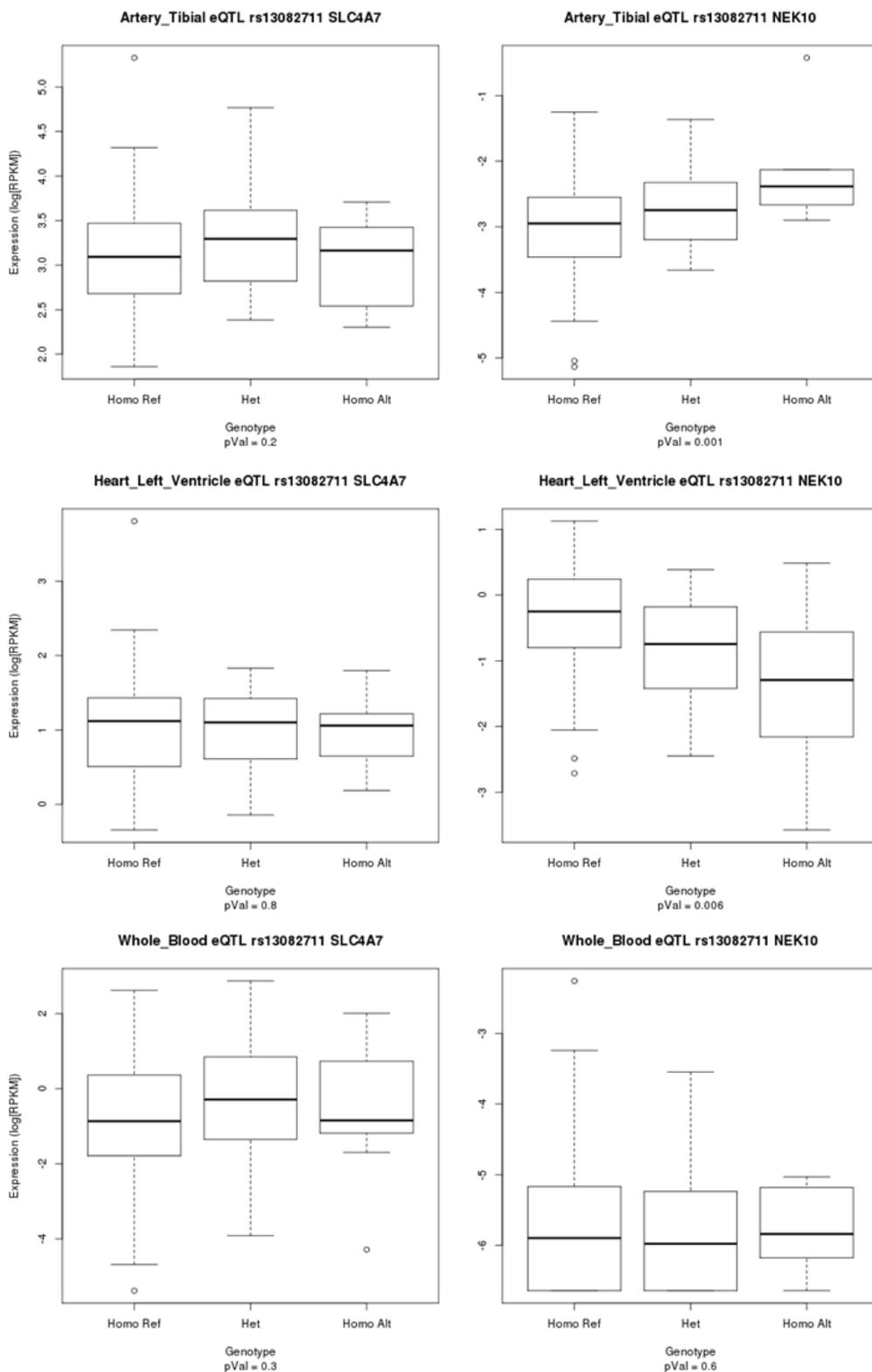


Figure 56: Genotype-Tissue Expression project (GTEx) SLC4A7 and NEK10 expression data of selected tissues for the rs13082711 SNP
 Note the differences in y-axis scales, and smaller magnitude of changes in the NEK10 samples.

Although the attempt to provide a panel of expression data from different tissues is highly laudable, there are significant issues with the interpretation. As an indication of the population that the samples were derived, the pilot phase demographics (GTEx Project, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v1.p1) show they were 65% male, median age 53, >80% white, and with a median BMI 26.7. Importantly, 94.3% of samples were derived following deaths (the remaining from surgical samples), of which 41.4% were from deaths relating to intracranial haemorrhage or ischaemic strokes, and 30% from myocardial infarction. There is also other significant cardiovascular past medical history that includes hypertension (51.4%), type II diabetes (22.7%) and renal failure (10.1%). This means the impact of the environment is significant, and potentially outweighs that of the initial genetic influence on gene expression. These subjects also had their (based on their past medical history, likely to be non-inconsequential) drug history collected, but this was not available publicly. Even if medications data were available, it would be near impossible to tease out the effect of medications vs environment vs genetics based on this sample size (pilot n=214, eventual target n=750).

Further consideration should also be given to the choices of tissue samples. Most of them are not a uniform cell type, for example, the tibial artery is likely to contain vascular smooth muscle cells, endothelial cells, fibroblasts and any other surrounding connective tissue. Also notably, the magnitude of genotype-related *NEK10* mRNA levels is far smaller as compared to the magnitude of detected *SLC4A7* mRNA levels. It is worth recognising that statistically significant results are not necessarily biologically significant. These predictions should be taken in context of a recent meta-analysis of identifying 34 gene differentially expressed in treatment-naïve hypertensives (Huan *et al.*, 2015). *SLC4A7* was not identified as being differentially expressed, but the main limitation of this study is that it centres on the RNA expression in samples of whole blood.

Therefore, a lack of observed relationship between *SLC4A7* genotype and its expression may be related to multiple environmental confounders, or in this context of increased noise, an underpowered sample size.

3.4. TFBS (transcription factor binding site) prediction

If a SNP occurs at the consensus binding site of transcription factors, the nucleotide change may increase or decrease the probability of binding and thus up- or down-regulating RNA transcription (the direction of action depending on the interacting nuclear protein itself). It follows that these particular SNPs may have an overall effect on gene expression. There are several online prediction tools that use position weight matrices to predict possible TFBS, which includes AliBaba2 (Grabe, 2002), JASPAR (Bryne *et al.*, 2008) and P-Match (Chekmenev *et al.*, 2005). Using these tools, the four SNPs within 5000 bp of the start of transcription were assessed for allele-dependent potential transcription factor binding sites – in other words, where the predicted scores change with the polymorphism. The results are shown in

Table 31.

	Reference	rs2003886 A/C	rs13097178 C/T	rs35701251 A/T	rs13086867 A/G
AliBaba2	http://www.gene-regulation.com/pub/programs/alibaba2/index.html	C-EBP α (C)	C-EBP α (T)	N/A	SP-1 (G) NF- κ B (G)
JASPAR	http://jaspar.genereg.net/	NFE2L2 (A) AP1 (A) Arnt/AHR (C)	ELK1 (C) ELK4 (C) SPIB (C) TAL1/TCF3 (T) Hand1-Tcfe2a (T) NFATC2 (T) (+ 18 others)	FEV (T) SPI1 (T) ELF5 (T) SPIB (T) ETS1 (T) GABPA (T) (+ 3 others)	AP1 (A) SPIB (A) RELA (G) MZF1 (G) SP1 (G) MIZF (G) (+ 2 others)
Match	http://www.gene-regulation.com/pub/programs/match/index.html	N/A	N/A	N/A	N/A
Patch	http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi	N/A	NIP (C) PEA (C) E12 (T) NF-ATp (T)	N/A	N/A
P-Match	http://www.gene-regulation.com/cgi-bin/pub/programs/pmatch/bin/p-match.cgi	N/A	N/A	N/A	N/A

Table 31: Transcription factors with different allele-specific scores for binding sequences. The polymorphism with preferential binding scores is listed in parentheses. SNPs are listed together with minor/major allele, and position relative to transcription start. Where more than 6 are

predicted, only the top 6 hits are listed. N/A denotes transcription factor binding was not predicted to be different based on the two alleles.

Whilst these predictions may be helpful as a starting point, there are several limitations to these tools. One striking observation is the inconsistencies between the four prediction tools utilised. This not just limits the utility of using the predictions as a starting point, but also raises concerns of its positive predictive value.

These prediction tools utilize position weight matrices from databases such as TRANSFAC (Matys *et al.*, 2003) and JASPAR (Bryne *et al.*, 2008). Thus, the predictions are only as generalizable as their input databases. The experimental evidence that inform these databases include gel shift assays, footprinting analysis and ChIP-seq studies, and is based on a range of cell types. With all of these experimental methods, there are some limitations to the generalizability from *in vitro* assays to *in vivo* function, as well as across cell types. Whilst the TRANSFAC database provides a “quality value” to describe the confidence of the observed DNA-binding activity, this continuous variable can sometimes be difficult to interpret in terms of an *in vivo* probability.

With these caveats in mind, a recent study has identified a false positive rate of 0.13 and 0.64 for P-Match and JASPAR respectively (Fazius *et al.*, 2011), based on the re-identification of twenty-two hidden true sites in random sequences. This however may not fully reflect the low positive predictive value of these prediction tools due to the low underlying base rate within a whole genome.

3.5. Nucleosome occupancy prediction

A separate mechanism of DNA-nuclear protein interaction controlling gene expression involves the organization of DNA into nucleosomes. Nucleosomes are 147 base pair segments of DNA, wrapped around an octamer of histone proteins, which in turn can be connected together by the linker histone H1. The presence of histone binding and nucleosome structures can influence and reduce the accessibility of DNA to other proteins such as transcription factors. In turn, if a SNP sufficiently changes the DNA sequence to result in differences in affinity for histone octamers; this may result in an overall change in gene expression.

An online nucleosome occupancy prediction tool (http://bio.physics.leidenuniv.nl/~noort/cgi-bin/nup3_st.py) has been designed using an empirical statistical mechanics model based solely on

the periodic occurrence of the dinucleotides TA, TT, AA, and GC found in regions of high nucleosome affinity (van der Heijden *et al.*, 2012). The authors calculated a Pearson's correlation coefficient (r) of 0.66 between their predictions and *in vivo* (yeast) experiments. However, there is no direct evidence of the strength of correlation between these prediction and human nucleosomes *in vivo*. With the proviso that the tool is not designed to predict changes in nucleosome occupancy relating to SNPs, it did highlight some interesting results. Out of the 93 SNPs of interest, up to 18 SNPs were predicted to change the probability of the nucleotide being at a nucleosome dyad (Appendix 11), introducing an interesting possibility of blood pressure associated SNPs exerting epigenetic effects. However, only rs3755652 (Figure 57) produced a magnitude change in P of greater than 0.01.

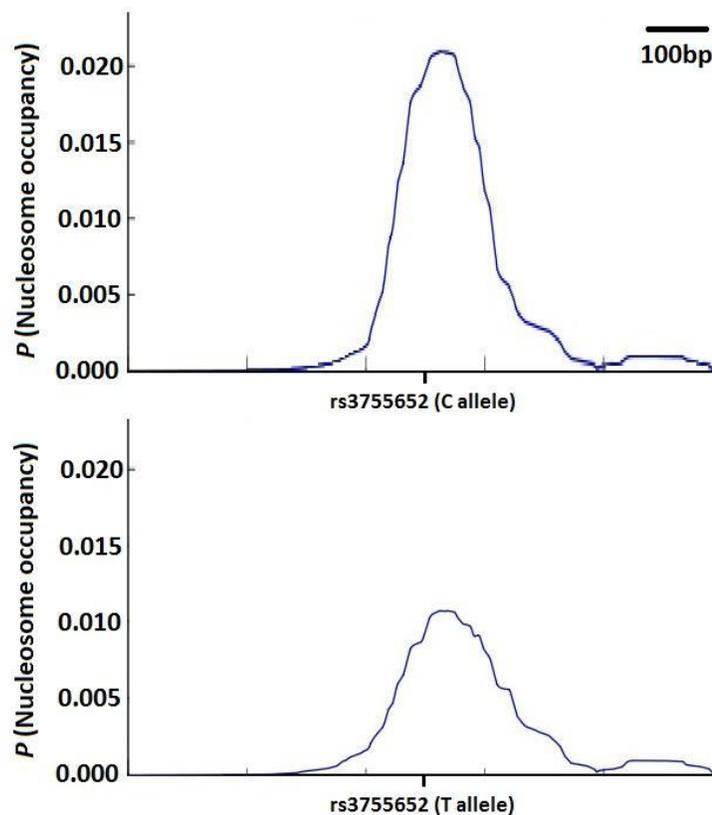


Figure 57: Prediction of nucleosome occupancy with the respective alleles at rs3755652

It is important to recognize that whilst some of the changes in predicted probabilities of nucleosome binding may appear relatively large (up to 100% increase in numerical value), the biological relevance is uncertain. Additionally, the algorithm is unable to consider any impact of DNA methylation or the variations of *in vivo* cellular milieu.

3.6. Predicting the functional impact of non-synonymous single nucleotide polymorphisms (nsSNP) and the resultant amino acid change

Only one of the 93 SNPs of interest (rs3755652) results in a non-synonymous coding substitution. The guanine to adenine (G1047A) polymorphism produces an amino acid change from glutamic acid to lysine at position 326 (E326K), and has some potential for impact on the function of the encoded NBCn1 protein.

Taken in the wider context, the presence of glutamic acid or lysine at that position is not commonly found in the animal kingdom. With a BLAST search of the 30 most similar NBCn1 Cassette II from other species (Appendix 12), only the common chimpanzee (*Pan troglodyte*) is predicted to have a glutamic acid (E) residue, the same as humans. The common marmoset (*Callithrix jacchus*) is predicted to have a histidine (H) residue there. Of the 28 remaining, of which 7 are primates, all are predicted to have a glutamine (Q) residue in that position. None are predicted to have the rarer human risk-allele amino acid lysine (K) residue there. Whilst potentially interesting, this does not provide further evidence on whether (or not) the E326K amino acid substitution has any functional significance.

The main difference between glutamic acid and lysine amino acid residues is that the former has a negative charge; whilst the latter has a positive charge. Apart from their charges, they are otherwise relatively similar - having similar molecular weights (147.13 kDa and 146.19 kDa respectively), both being hydrophilic, and neither are bulky enough to classically result in steric hindrance and changes in secondary structures. This is reflected by the early analysis by Grantham (1974), noting that the difference between the two amino acids based on composition, polarity and molecular volume to be relatively small. Set out in alphabetical order,

Table 32 shows the predictions by eight different online tools on the effect of the amino acid change. Three other prediction tools Panther (<http://www.pantherdb.org/tools/csnpscoreForm.jsp>), SNPs3D (<http://www.snps3d.org/>) and nsSNP analyser (<http://snpanalyzer.uthsc.edu/>) did not return any results as they admit to a lack of homologous structures in their training database.

Prediction tool	Reference	Prediction (Score)	From Bromberg and Rost, 2007; Thusberg et al., 2011; Frousios et al., 2013			Calculated from Sens and Spec	
			Sens	Spec	FN*	LR+	LR-
Grantham	Grantham, 1974	Moderately conservative AA change (56)	N/A	N/A	N/A	N/A	N/A
i-Mutant	http://folding.uib.ee/i-mutant/i-mutant2.0.html	Destabilising mutation (-0.80 kcal/mol at 37C, pH 7.4)	N/A	N/A	N/A	N/A	N/A
Mutation Assessor	http://mutationassessor.org/	Neutral (0.145)	0.76	0.86	0.19	5.43	0.28
MutPred	http://mutpred.mutdb.org/	No hypothesis (0.332). Leading features: <ul style="list-style-type: none"> Gain of MoRF binding (P = 0.0041) Gain of methylation at E326 (P = 0.044) 	0.85	0.78	0.05	3.86	0.19
PHD-SNP	http://gpcr2.biocomp.unibo.it/~emidio/PhD-SNP/PhD-SNP.htm	Neutral (8)	0.79	0.77	0.23	3.43	0.27
PolyPhen2	http://genetics.bwh.harvard.edu/pph2/	Benign	0.70	0.73	0.12	2.50	0.41
SIFT	http://sift.jcvi.org/www/SIFT_enst_submit.html	Tolerated (0.87)	0.63	0.84	0.22	3.94	0.44
SNAP	https://www.rostlab.org/services/SNAP	Effect (2)	0.80	0.76	N/A	3.33	0.26
SNP&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go/	Neutral (6)	0.92	0.71	0.15	3.17	0.11

Table 32: Predicted functional impact of the E326K polymorphism by online tools.

Sens denotes sensitivity; Spec, specificity; FN, false negative rate; LR+, likelihood ratio for positive result; LR-, likelihood ratio for negative result; N/A, not available.

* indicates false negative rates reported using benchmarking set by Frousios et al., 2013 and not general population, see text for discussion.

LR+ calculated as $(Sens)/(1-Spec)$, LR- calculated as $(1-Sens)/(Spec)$

Whilst some tools were unable to provide a prediction based on their respective algorithms and training sets, all but one predicted that the amino acid change would be tolerated. Only SNAP and i-Mutant hinted at a potential change in function relating to the nsSNP. SNAP returned a result of 2, from a range of -100 to 100, where anything below 0 is regarded as neutral, and those above 0 are regarded as having an effect. It however does not provide any guidance on the degree of confidence based on the numerical value. A similar issue is present with the i-Mutant prediction. It suggested the E326K amino acid change results in a “destabilizing” mutation. However, it should be noted that prediction (-0.80 kcal/mol) is difficult to put in physiological context. Furthermore, a dataset comparing the predicted and experimentally obtained protein stability, identified only a modest correlation coefficient ($r = 0.62$), and a relatively large standard error (1.45 kcal/mol). This indicates the signal-to-noise ratio of this prediction is low.

This might initially be interpreted as the E326K polymorphism would not have an impact on the function of the NBCn1 protein. However, caution should be exercised with these predictions, as exemplified by the E6V mutation in the β -globin gene. This common beta-globin mutation results in the abnormal HbS responsible for sickle cell disease and results in a significant change in protein behavior *in vivo* but was predicted to be benign by PolyPhen. The likely reason for this is that most prediction tools are unable to take into account the wide variety of macromolecule-macromolecule interactions that occur *in vivo*. The same study identified that PolyPhen and SIFT only correctly identified approximately 70% of mutations causing the clinical manifestation of G6PD (Glucose-6-phosphate dehydrogenase) deficiency (Tchernitchko *et al.*, 2004). Further studies identified the sensitivity and specificities of these tools mainly ranges from 70% to 85%, as summarised in

Table 32 (Bromberg and Rost, 2007; Thusverg *et al.*, 2011; Frousios *et al.*, 2013). However, as the studies utilize datasets enriched with well-characterised (and presumably mostly high-impact Mendelian) mutations, these parameters may be overestimated. Additionally, sensitivities and specificities may not be the ideal parameter, as the positive and negative predictive values are dependent on the underlying population prevalence. Alternatively, the utility of these data sets may be viewed in terms of their likelihood ratios. The tools have calculated likelihood ratio for negative results ranging from 0.11 to 0.44. This highlights the difficulty of interpreting results from prediction tools on their own, as the pre-test odds for whether an nsSNP alters protein function is at best unknown, or at worst biased and untrustable.

In addition, these tools often utilize overlapping algorithms and training sets of experimental data. This would lead to often comparable predictions, and the consistency across the different tools may be falsely reassuring. Furthermore, the training sets for these algorithms are unlikely to have encountered sequences that resemble the area of interest, particularly as the sequence of the 124 amino acid Cassette II does not contain any putative conserved domains, and bares minimal similarities to other known human proteins (Figure 58). The E (Expect)-value reflects the number of expected hits purely by chance in that size of database, and are 0.7, 1.2 and 5.4 for the three hits respectively. In other words, the low alignment scores of the three “hits” suggest that they are there most probably by chance.

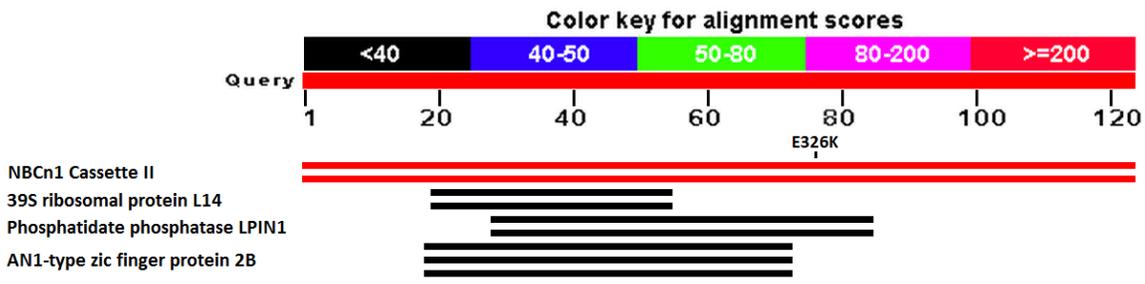


Figure 58: BLAST protein query of the human NBCn1 Cassette II BLAST (Basic Local Alignment Search Tool; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>, NCBI) protein query with the NBCn1 Cassette II using the protein-protein BLAST algorithm limited to Homo sapien proteins only. Black bars indicate low alignment scores, limited to the default maximum E-value of 10.

The lack of sequence alignments within the libraries of prediction tools should be of concern, as most tools provide a prediction output irrespective of confidence of predictive ability. This is further exacerbated by a false dichotomy, where the prediction is often presented as a definitive categorical variable. A good example is provided by MutationAssessor (Figure 59), whereby a cut-off is placed at the score where the prevalence of false negative will equal that of false positives (of up to 22%!). A categorical output could be misleading with results just either side of a decision boundary.

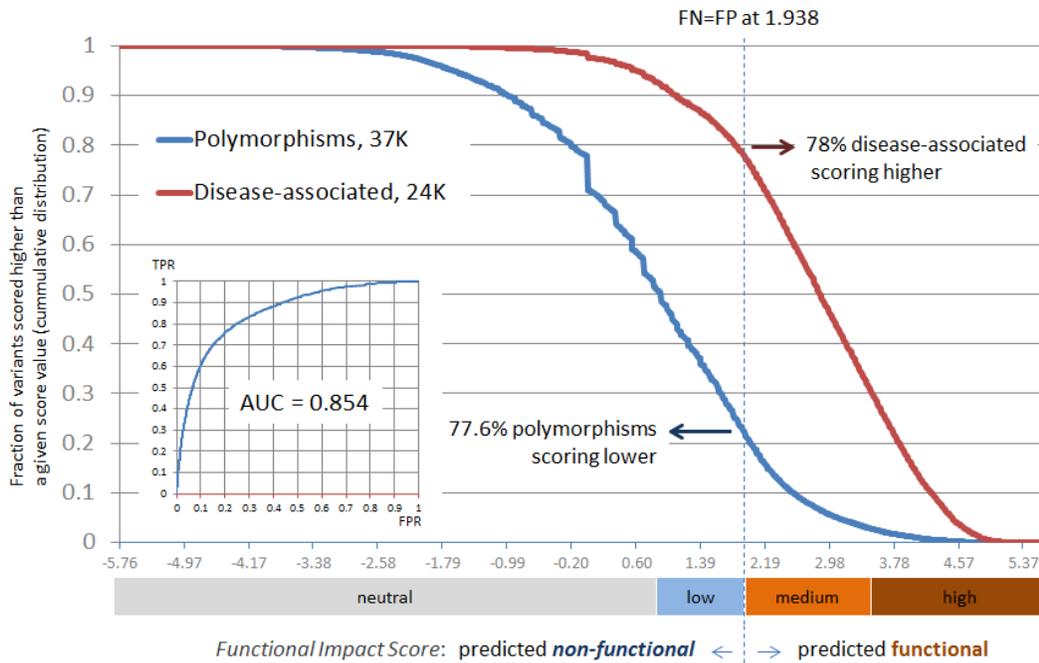


Figure 59: Example of prediction tools converting continuous scores to categorical outputs
Adapted from <http://mutationassessor.org/howitworks.php>

Another consideration of the algorithm training sets is that they are predominantly derived from distinct and unambiguous cases of deleterious mutations, as seen in monogenetic conditions. This may not be powerful enough to determine potentially subtle changes that may be identified from GWAS hits.

Finally, these tools are also usually unable to consider the impact of nearby regulatory elements or post-translational modifications. The potential of nearby regulatory elements may still be relevant to NBCn1, as the E326K polymorphism lie near to *postulated* protein kinase C phosphorylation sites of Ser³²⁴ and Ser³²⁷ (Pushkin *et al.*, 1999). Similarly, the E326K amino acid change lies only 6 amino acids away from a likely calcineurin binding site (Danielsen *et al.*, 2013), which would not be accredited by the prediction tools.

3.7. Predicting the functional impact of nsSNP on 3-dimensional structure of protein

Another way of considering potential impacts of amino acid changes is via 3-dimensional modelling of the protein variants. One such tool is SWISS-MODEL (<http://swissmodel.expasy.org/>), which in short, utilises evolutionarily related structures which combine to generate a structural

model of a protein of interest based on sequence homology. The quality of the computed model is estimated to indicate the expected quality of the obtained model (Biasini *et al.*, 2014). This model predicted structures that on first glance appeared to vary at the intracellular N-terminus, where the 326K variant may have a more open conformation compared to the 326E variant (Figure 60). However, it is essential to recognize that both predictions were provided together with very low GMQE (Global Model Quality Estimate) and QMEAN4 scores, indicating poor prediction quality. This is often due to the queried sequence having insufficient homology with sequences of known structures, as already explored above. With this, the result is extremely unreliable for forming any further hypotheses.

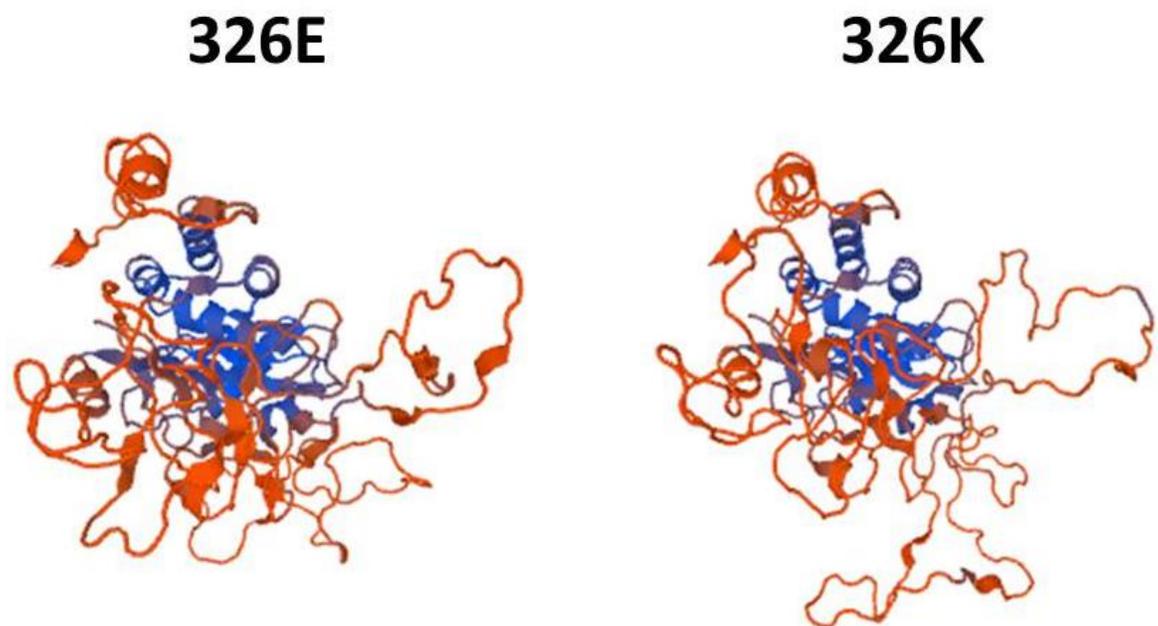


Figure 60: SWISS-MODEL predicted 3-dimensional representation of native conformational changes with the E326K variants

An alternative tool is PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) which concentrates on the secondary structure. This tool reports a Q_3 score (percentage of correct predictions of the three states – helix, sheet and coil) of 73.4%. These predictions are based on a window of 15 amino acids to predict the central residue using an artificial neural network linking the input sequence with known protein structures (McGuffin *et al.*, 2000). The summary data (Figure 61) initially suggests that the E326K changes the secondary structure of the NBCn1 protein. However, the detailed outputs (Appendix 13) indicate that the prediction at these specific

sequences are modest at best. Once again, the result is perhaps unreliable for forming any further hypotheses on its own.

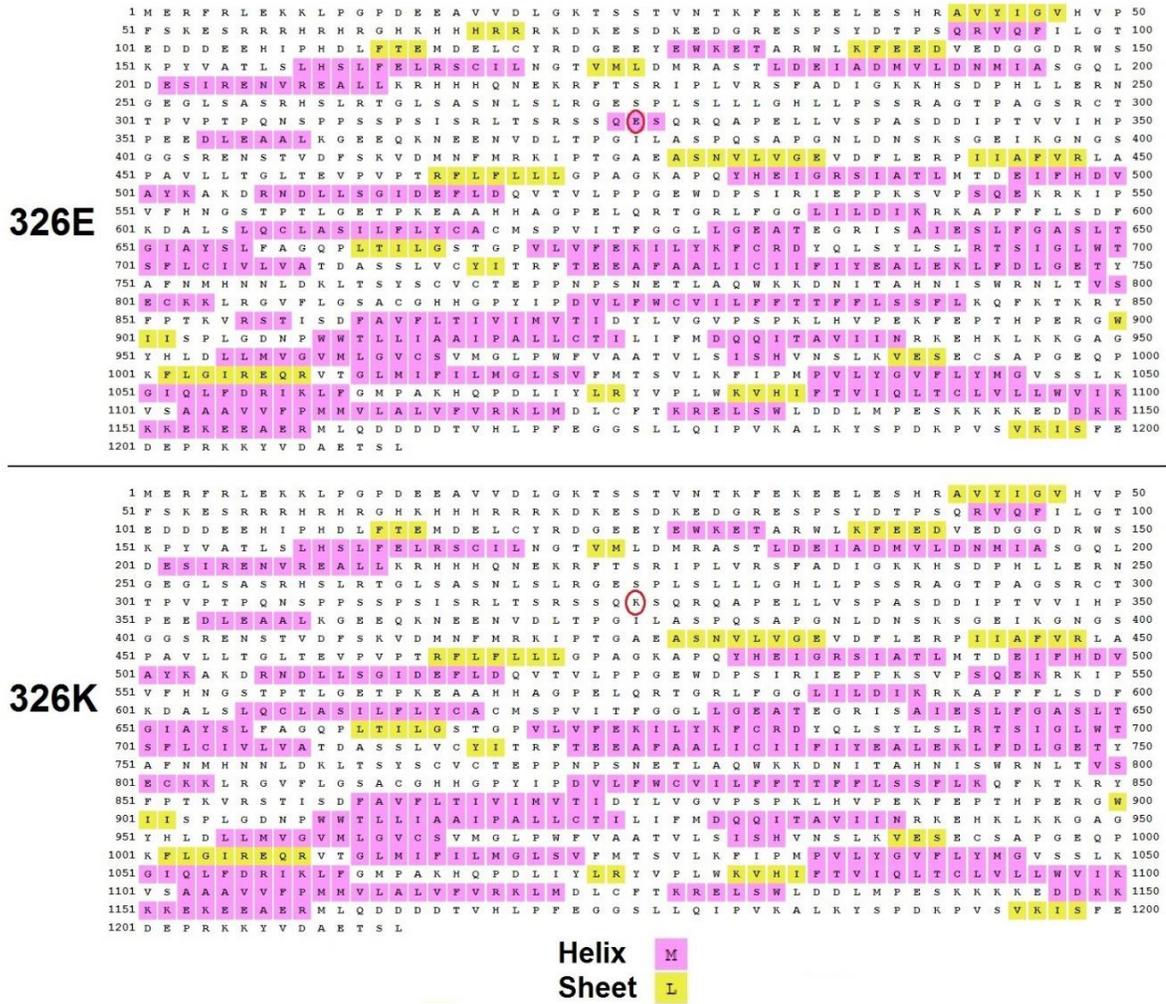


Figure 61: PSIPRED prediction of 3-dimensional conformation with the E326K variants. Highlighted are the predicted helices (purple boxes), predicted sheets (yellow boxes) and the E326K amino acid change (red circle).

Another tool that predicts secondary structures is I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Roy *et al.*, 2010), one of the highest ranking 3D-model prediction servers (Critical Assessment of Techniques for Protein Structure Prediction, 2012). Similar to PSIPRED, I-TASSER predicts that the 326E variant results in a protein that has a helical secondary structure at the 326 residue, but not with the 326K variant (Figure 62). Like the previous tool, I-TASSER also provides a global estimate of quality of predicted models (C-score) which in the case of these predictions were low.

sodium-bicarbonate co-transporters was previously discussed in Section 1.10, and polymorphisms that have an effect on the propensity to these modifications may in turn impact on protein function.

One of the commonly used phosphorylation site prediction tools is PhosphoBase (<http://phospho.elm.eu.org/>), first released in 1999 (Blom *et al.*, 1999) and is regularly updated (Dinkel *et al.*, 2010). It utilises position weight matrices based on experimentally verified phosphorylation sites. The sensitivity of the prediction tool was recorded to range from 69% to 96%. However it was also noted that there were false positives, resulting in positive predictive value between 65% and 89% (Blom *et al.*, 1999). PhosphoBase does not predict any differences in phosphorylation sites when considering the E326K amino acid change (Table 33).

Phosphorylated residue	Position	Sequence	Conservation score	IUPRED score	Phospho3D accessibility score
S	84	KESDKEDGRE S PSYDTPSQRV	0.91	0.76	Medium
S	382	LTPGILASPQ S APGNLDNSKS	0.81	0.71	Medium
S	400	SKSGEIKGNG S GGGRENSTVD	0.04	0.76	Medium
S	403	GEIKGNGSGG S RENS TVDFSK	0.89	0.68	Medium
S	407	GNGSGGRENS T VDFSKVDMN	0.89	0.6	Medium
T	1167	AERMLQDDDD T VHLPFEGGSL	0.31	0.29	Medium
S	1176	DTVHLPFEGG S LLQIPVKALK	0.04	0.12	Medium
S	1213	PRKKYVDAET S L	-	0.32	Medium

Table 33: Phospho.ELM predictions on possible sites of post-translation phosphorylation. There were no differences noted as relevant when considering the E326K amino acid change. Conservation score quantifies the conservation of each phosphor-site with scores from 0 to 1, where 1 indicates highest conservation. IUPRED score is a prediction of disordered amino acids (where disorder better allows modification) with scores from 0 to 1, where scores greater than 0.5 are considered as disordered. Phospho3D score is the prediction of surface accessibility by kinases.

This limited impact of E326K is also reported by DISPHOS v1.3 (DISorder-Enhanced PHOSphorylation Sites Predictor, <http://www.dabi.temple.edu/disphos/>) (

Table 34) a prediction tool trained on over 2000 non-redundant experimentally confirmed protein phosphorylation sites as well as Phospho.ELM (

Table 33) outputs. Both these tools utilize the observation that intrinsically unstructured / disordered sequences of amino acids have no single well-defined tertiary structure in their native

state, which in turn affords the flexibility for post-translational modification (Iakoucheva *et al.*, 2004).

Position	Residue		326E		326K
55	S	0.846	FSKE S RRRH	0.846	FSKE S RRRH
76	S	0.913	KDKE S DKED	0.913	KDKE S DKED
84	S	0.949	DGRE S PSYD	0.949	DGRE S PSYD
86	S	0.606	RESP S YDTP	0.606	RESP S YDTP
87	Y	0.624	ESPS Y DTPS	0.624	ESPS Y DTPS
89	T	0.675	PSYD T PSQR	0.675	PSYD T PSQR
91	S	0.545	YDTP S QRVQ	0.545	YDTP S QRVQ
257	S	0.519	GLSA S RHSL	0.519	GLSA S RHSL
260	S	0.738	ASRH S LRTG	0.738	ASRH S LRTG
271	S	0.629	ASN L SLRGE	0.629	ASN L SLRGE
276	S	0.749	LRGE S P LSL	0.749	LRGE S P LSL
293	T	0.536	SRAG T PAGS	0.535	SRAG T PAGS
297	S	0.598	TPAG S RCTT	0.596	TPAG S RCTT
301	T	0.812	SRCT T PVPT	0.811	SRCT T PVPT
309	S	0.798	TPQN S PPSS	0.795	TPQN S PPSS
312	S	0.779	NSPP S SPSI	0.776	NSPP S SPSI
313	S	0.843	SPPS S PSIS	0.839	SPPS S PSIS
315	S	0.887	PSSP S ISRL	0.884	PSSP S ISRL
317	S	0.902	SPSI S R L TS	0.892	SPSI S R L TS
321	S	0.919	SRLT S RSSQ	0.922	SRLT S RSSQ
323	S	0.834	LTSR S <u>S</u> QES	0.801	LTSR S <u>S</u> QKS
324	S	0.928	TSRS S <u>Q</u> ESQ	0.916	TSRS S <u>Q</u> KSQ
327	S	0.741	SSQ E <u>S</u> QRQA	0.723	SSQ K <u>S</u> QRQA
392	S	0.509	DNSK S GEIK	0.509	DNSK S GEIK
403	S	0.605	GSGG S RENS	0.605	GSGG S RENS
407	S	0.63	SREN S TVDF	0.63	SREN S TVDF
532	S	0.534	EWDP S IRIE	0.534	EWDP S IRIE
540	S	0.561	EPPK S VPSQ	0.561	EPPK S VPSQ
1132	S	0.518	KREL S WLDD	0.518	KREL S WLDD
1141	S	0.854	LMPE S KKKK	0.854	LMPE S KKKK

Table 34: DISPHOS v1.3 predictions on possible post-translation phosphorylation.

Position of the E326K change is underlined in the 323^{Ser}, 324^{Ser} and 327^{Ser} sequences. There were no differences in scores noted as relevant when considering the E326K amino acid change. Phosphorylation scores ranges from 0 to 1, where 1 indicates highest probability and residues with the score >0.5 are considered as likely phosphorylation sites.

On the other hand, these results conflict with a third tool, PhosphoNet (<http://www.phosphonet.ca/>). PhosphoNet also utilises position weight matrices, but predicted a difference in the propensity to phosphorylation at the 323^{Ser} residue; there the 326E amino acid increases the probability, whilst 326K reduces it (

Table 35). This introduces an interesting possibility of the non-synonymous polymorphism at rs3755652 resulting in a change in protein function by altering post-translational phosphorylation.

		326E		326K				326E		326K	
Position	Residue	Sequence	Score	Sequence	Score	Position	Residue	Sequence	Score	Sequence	Score
26	S	GKTSSTVNT	0.678	GKTSSTVNT	0.678	323	S	LTSRSSQES	0.945	LTSRSSQKS	0.243
30	T	STVNTKFEK	0.923	STVNTKFEK	0.923	324	S	TSRSSQESQ	0.997	TSRSSQKSQ	0.998
39	S	EELESHRAV	0.824	EELESHRAV	0.824	327	S	SSQESQROA	0.996	SSQKSQROA	0.997
52	S	HVPFSKESR	0.966	HVPFSKESR	0.966	337	S	ELLVSPASD	0.81	ELLVSPASD	0.81
55	S	FSKESRRRH	0.994	FSKESRRRH	0.994	390	S	NLDNSKSGE	0.861	NLDNSKSGE	0.861
76	S	KDKESDKED	0.998	KDKESDKED	0.998	400	S	KGNGSGGSR	0.819	KGNGSGGSR	0.819
84	S	DGRESPSYD	0.995	DGRESPSYD	0.995	403	S	GSGGSRENS	0.92	GSGGSRENS	0.92
86	S	RESPSYDTP	0.777	RESPSYDTP	0.777	407	S	SRENSTVDF	0.996	SRENSTVDF	0.996
87	Y	ESPSYDTPS	0.953	ESPSYDTPS	0.953	412	S	TVDFSKVDM	0.971	TVDFSKVDM	0.971
89	T	PSYDTPSQR	0.893	PSYDTPSQR	0.893	424	T	RKIPTGAEA	0.563	RKIPTGAEA	0.563
91	S	YDTPSQRVQ	0.953	YDTPSQRVQ	0.953	459	T	LTGLTEVPV	0.522	LTGLTEVPV	0.522
121	Y	DELQYRDGE	0.695	DELQYRDGE	0.695	481	Y	KAPQYHEIG	0.942	KAPQYHEIG	0.942
127	Y	DGEEYEWKE	0.943	DGEEYEWKE	0.943	502	Y	HDVAYKAKD	0.567	HDVAYKAKD	0.567
132	T	EWKETARWL	0.792	EWKETARWL	0.792	512	S	NDLLSGIDE	0.997	NDLLSGIDE	0.997
161	S	LSLHSLFEL	0.988	LSLHSLFEL	0.988	540	S	EPPKSVPSQ	0.988	EPPKSVPSQ	0.988
181	S	DMRASTLDE	0.981	DMRASTLDE	0.981	543	S	KSVPSQEKR	0.988	KSVPSQEKR	0.988
182	T	MRASTLDEI	0.914	MRASTLDEI	0.914	559	T	GSTPTLGET	0.815	GSTPTLGET	0.815
203	S	QLDESIREN	0.996	QLDESIREN	0.996	563	T	TLGETPKEA	0.971	TLGETPKEA	0.971
225	T	EKRFTSRIP	0.947	EKRFTSRIP	0.947	639	S	EGRISAIES	0.997	EGRISAIES	0.997
226	S	KRFTSRIPL	0.947	KRFTSRIPL	0.947	695	S	SLRTSIGLW	0.543	SLRTSIGLW	0.543
233	S	PLVRSFADI	0.703	PLVRSFADI	0.703	723	T	ITRFTEEAF	0.744	ITRFTEEAF	0.744
242	S	GKKHSDPHL	0.59	GKKHSDPHL	0.59	793	S	AHNISWRNL	0.977	AHNISWRNL	0.977
255	S	GEGLSASRH	0.51	GEGLSASRH	0.51	798	T	WRNLTVSEC	0.813	WRNLTVSEC	0.813
260	S	ASRHSLRTG	0.996	ASRHSLRTG	0.996	800	S	NLTVSECKK	0.508	NLTVSECKK	0.508
271	S	ASNLSLRGE	0.932	ASNLSLRGE	0.932	847	T	KQFKTKRYF	0.509	KQFKTKRYF	0.509
276	S	LRGESPLSL	0.973	LRGESPLSL	0.973	857	S	TKVRSTISD	0.938	TKVRSTISD	0.938
288	S	HLLPSSRAG	0.652	HLLPSSRAG	0.652	858	T	KVRSTISDF	0.902	KVRSTISDF	0.902
293	T	SRAGTPAGS	0.974	SRAGTPAGS	0.974	882	S	VGVPSPKLH	0.949	VGVPSPKLH	0.949
297	S	TPAGSRCTT	0.818	TPAGSRCTT	0.818	894	T	KFEPTHPER	0.553	KFEPTHPER	0.553
300	T	GSRCTTPVP	0.509	GSRCTTPVP	0.509	982	S	VLSISHVNS	0.885	VLSISHVNS	0.885
301	T	SRCTTPVPT	0.994	SRCTTPVPT	0.994	986	S	SHVNSLKVE	0.937	SHVNSLKVE	0.937
309	S	TPQNSPPSS	0.956	TPQNSPPSS	0.956	991	S	LKVESECSA	0.931	LKVESECSA	0.931
312	S	NSPPSSPSI	0.551	NSPPSSPSI	0.551	1127	T	DLCFTKREL	0.869	DLCFTKREL	0.869
313	S	SPPSSPSIS	0.98	SPPSSPSIS	0.98	1132	S	KRELSW added	0.997	KRELSW added	0.997
315	S	PSSPSISRL	0.932	PSSPSISRL	0.932	1141	S	LMPESECKK	0.97	LMPESECKK	0.97
317	S	SPSISRLTS	0.547	SPSISRLTS	0.547	1188	S	ALKYSPDKP	0.886	ALKYSPDKP	0.886
320	T	ISRLTSRSS	0.988	ISRLTSRSS	0.988	1194	S	DKPVSVKIS	0.992	DKPVSVKIS	0.992
321	S	SRLTSRSSQ	0.996	SRLTSRSSQ	0.996	1198	S	SVKISFEDE	0.996	SVKISFEDE	0.996
						1207	Y	PRKKYVDAE	0.906	PRKKYVDAE	0.906

Table 35: PhosphoNet predictions of post-translational phosphorylation sites

The predicted change due to the E326K amino acid change is highlighted in green (increased probability), red (decreased probability). Yellow denotes minimal change. Scores range from 0 to 1, where 1 indicates highest probability, and scores >0.5 are considered as likely phosphorylation sites.

Of note, another commonly used phosphorylation prediction tool, ScanSite3 (Massachusetts Institute of Technology, <http://scansite3.mit.edu>) also predicted an impact of the E326K non-synonymous polymorphism on phosphorylation, but in this case, on the 324^{Ser} residue instead (Table 36).

Site	326E		326K	
	Percentile	Sequence	Percentile	Sequence
Y87	0.11%	DGRES ^{Sy} DTPSQRV	0.11%	DGRES ^{Sy} DTPSQRV
Y87	0.15%	DGRES ^{Sy} DTPSQRV	0.15%	DGRES ^{Sy} DTPSQRV
Y87	0.03%	DGRES ^{Sy} DTPSQRV	0.03%	DGRES ^{Sy} DTPSQRV
T305	0.18%	RCTTPVPtPQNSPPS	0.18%	RCTTPVPtPQNSPPS
S309	0.02%	PVPTPQNsPPSSPSI	0.02%	PVPTPQNsPPSSPSI
S309	0.05%	PVPTPQNsPPSSPSI	0.05%	PVPTPQNsPPSSPSI
S313	0.09%	PQNSPPSsPSISRLT	0.09%	PQNSPPSsPSISRLT
S313	0.16%	PQNSPPSsPSISRLT	0.16%	PQNSPPSsPSISRLT
S317	0.16%	PPSSPSIsRLTSRSS	0.16%	PPSSPSIsRLTSRSS
S317	0.11%	PPSSPSIsRLTSRSS	0.11%	PPSSPSIsRLTSRSS
S324	0.18%	SRLTSRSsQESQRQA	0.76%	SRLTSRSsQKSQRQA
T459	0.09%	AVLLTGLtEVPVPTR	0.09%	AVLLTGLtEVPVPTR
T858	0.06%	FPTKVRStISDFAVF	0.06%	FPTKVRStISDFAVF

Table 36: ScanSite3 predictions of post-translational phosphorylation sites

The predicted change due to the E326K amino acid change is highlighted in green (increased probability), red (decreased probability). Scores are expressed as percentiles ranking the best fit compared to all other ScanSite records, and those at 0.2% or lower are considered as likely phosphorylation sites.

Glycosylation is another post-translational modification that may impact on protein function. However, compared to phosphorylation, the amount of protein glycosylation data is limited and its annotation in databases is scarce, thus having relatively lower predictive values.

NetOGlyc 4.0 (<http://www.cbs.dtu.dk/services/NetOGlyc/>) is a predictor of O-glycosylation trained on a mass spectrometry-based dataset of genetically engineered human cell lines (Steentoft *et al.*, 2013). It predicts that the 317^{Ser} and 337^{Ser} residues are more likely to be O-glycosylated with the polymorphism resulting in 326E, as opposed to 326K (Table 37). This introduces the possibility, albeit low in context of limited prior relevant examples, of the non-synonymous polymorphism at rs3755652 resulting in a change in protein function by altering post-translational glycosylation.

Position	326E score		326K score		Position	326E score		326K score	
24	0.918	Positive	0.918	Positive	305	0.803	Positive	0.847	Positive
25	0.836	Positive	0.839	Positive	309	0.718	Positive	0.771	Positive
26	0.918	Positive	0.919	Positive	312	0.738	Positive	0.734	Positive
27	0.919	Positive	0.916	Positive	313	0.898	Positive	0.845	Positive
52	0.763	Positive	0.744	Positive	315	0.622	Positive	0.596	Positive
55	0.938	Positive	0.935	Positive	317	0.576	Positive	0.408	Negative
76	0.868	Positive	0.871	Positive	321	0.773	Positive	0.577	Positive
84	0.915	Positive	0.920	Positive	323	0.632	Positive	0.527	Positive
86	0.904	Positive	0.905	Positive	324	0.641	Positive	0.577	Positive
89	0.573	Positive	0.580	Positive	327	0.747	Positive	0.659	Positive
91	0.796	Positive	0.794	Positive	337	0.538	Positive	0.395	Negative
100	0.528	Positive	0.518	Positive	379	0.897	Positive	0.844	Positive
225	0.735	Positive	0.733	Positive	382	0.852	Positive	0.784	Positive
226	0.847	Positive	0.844	Positive	390	0.821	Positive	0.765	Positive
233	0.820	Positive	0.831	Positive	392	0.885	Positive	0.846	Positive
242	0.853	Positive	0.852	Positive	400	0.803	Positive	0.778	Positive
255	0.923	Positive	0.906	Positive	403	0.651	Positive	0.618	Positive
257	0.916	Positive	0.910	Positive	407	0.774	Positive	0.773	Positive
260	0.863	Positive	0.811	Positive	408	0.514	Positive	0.528	Positive
263	0.827	Positive	0.678	Positive	412	0.520	Positive	0.556	Positive
266	0.848	Positive	0.708	Positive	532	0.747	Positive	0.726	Positive
268	0.724	Positive	0.609	Positive	540	0.856	Positive	0.843	Positive
288	0.743	Positive	0.736	Positive	543	0.789	Positive	0.748	Positive
289	0.876	Positive	0.865	Positive	556	0.958	Positive	0.949	Positive
293	0.841	Positive	0.839	Positive	557	0.966	Positive	0.960	Positive
297	0.791	Positive	0.870	Positive	559	0.902	Positive	0.887	Positive
300	0.892	Positive	0.916	Positive	563	0.929	Positive	0.918	Positive
301	0.930	Positive	0.942	Positive					

Table 37: Net-O-Glyc v4.0 predictions of possible sites of post-translational O-glycosylation. The differences in scores resulting from the E326K amino acid change is highlighted in green (increased probability), red (decreased probability). Scores range from 0 to 1, where 1 indicates highest probability, and scores above 0.5 is regarded as potential glycosylation sites.

NetOGlyc 4.0 has an estimated sensitivity of approximately 80% (Steentoft *et al.*, 2013), but positive and negative predictive values of this tool are likely to be low in context of an amino acid sequence with no *a priori* hypothesis, as applies in the case of NBCn1. Additionally, the use of the (arbitrary) cut-off of 0.5 may provide a false impression of a definitive prediction, where on a continuous scale of probability, the change may not be that definitive.

Another sequence-based prediction program, OGLcNAcScan (<http://cbsb.lombardi.georgetown.edu/hulab/OGAP.html>), predicted only 6 O-glycosylation sites, none of which are affected by the E326K amino acid change. However, this result should be taken in context of an estimated sensitivity of only 30% (Jochmann *et al.*, 2014).

A different form of glycosylation, *N*-glycosylation, is predicted to occur at 10 potential *N*-glycosylation sites by NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>), but similarly, none of them are affected by the non-synonymous polymorphism.

Although these tools may appear to suggest the possibility of the rs3755652 non-synonymous polymorphism altering post-translational modifications, and potentially resulting in a change in protein function, it has to be interpreted with caution in context of limitations of the training sets, conflicting results, modest predictive values, or limited prior evidence to suggest biological plausibility. Additionally, as will all prediction tools, there are complexities that cannot be factored in; particularly that *in silico* analyses may not reflect the *in vivo* milieu, with varying macromolecule interactions, unknown 3D protein conformations, as well as varying levels of enzymatic activity to result in the post-translational modification.

3.9. Predicting the impact of the nsSNP on the isoelectric point (pI)

While there are limited predictions on a functional change in the protein, there is a notable difference in the individual pI (isoelectric points) of the negatively charged glutamic acid and positively charged lysine amino acids (3.08 and 9.47 respectively). This potentially influences the overall pI of the protein. Although there are slightly different predictions for the pI of the glutamic acid and lysine variants, the latter consistently has a pI 0.1 larger than the former (Table 38). Caution should be exercised when extrapolating these predictions to *in vivo* situations as they are unable to account for protein folding and post-translational modifications.

Prediction tool	Reference	pI (326E)	pI (326K)
PepStats / EMBOSS	http://www.bi.up.ac.za/cgi-bin/emboss.pl?_action=input&_app=pepstats	6.70	6.79
ExPSPy, SIB Swiss Institute of Bioinformatics	http://web.expasy.org/compute_pi/	6.26	6.35
Scripps Research Institute	http://www.scripps.edu/~cdputnam/protcalc.html	6.71	6.81
ScanSite, Massachusetts Institute of Technology	http://scansite.mit.edu/calc_mw_pi.html	6.26	6.35

Table 38: Predicted impact of the E326K polymorphism on isoelectric point (pI)

Whilst this difference in pI may not be sufficient to result in a physiological impact on the protein function, it should be sufficient to produce a difference in experiments utilizing isoelectric focusing to separate proteins.

3.10. Predicting the impact of the nsSNP on enzymatic protein digestion and potential differences in mass spectrometry signals

Another potential method of utilizing the non-synonymous polymorphism in assessing the relative expression of the two alleles is by mass spectrometry. As the two amino acid residues, lysine and glutamic acid have similar molecular masses, this would be difficult to differentiate by overall mass alone. However, the change of amino acid may change its behavior under tryptic digestion. This utilises the enzymatic property of trypsin cleaving peptide chains mainly at the carboxyl side of the amino acids lysine or arginine, except when either is followed by proline.

As the E326K polymorphism results in one of the amino acids having a propensity to be cleaved (lysine, K), and the other does not (glutamic acid, E), the online tool MS-Digest (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest>) predicts different sized fragments produced after tryptic digestion with the two different amino acids (Appendix 14). These differences in fragment sizes would allow relative quantification of the respective proteins by mass spectrometry in heterozygotes for the non-synonymous (E326K) rs3755652 polymorphism, thus indicating if the protein representing a particular allele is preferentially expressed.

3.11. ESE (exonic splicing enhancer) site prediction

In view of the recent findings that NBCn1 splice variants demonstrated difference in activity (see Section 1.9 for details, Yang *et al.*, 2009; Danielsen *et al.*, 2013; Liu *et al.*, 2013), any impact of *SLC4A7* genotype on mRNA splicing may contribute to the blood pressure effects. Out of the 93 SNPs of interest, four are exonic. Using the online ESEfinder prediction tool (Cold Harbour Spring Laboratories; <http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese finder.cgi?process=home>), all four exonic SNPs are predicted to be in potential exonic splicing enhancer sites, and the nucleotide change potentially alters the probability of a splicing event (example Figure 63, Appendix 15 for full results).

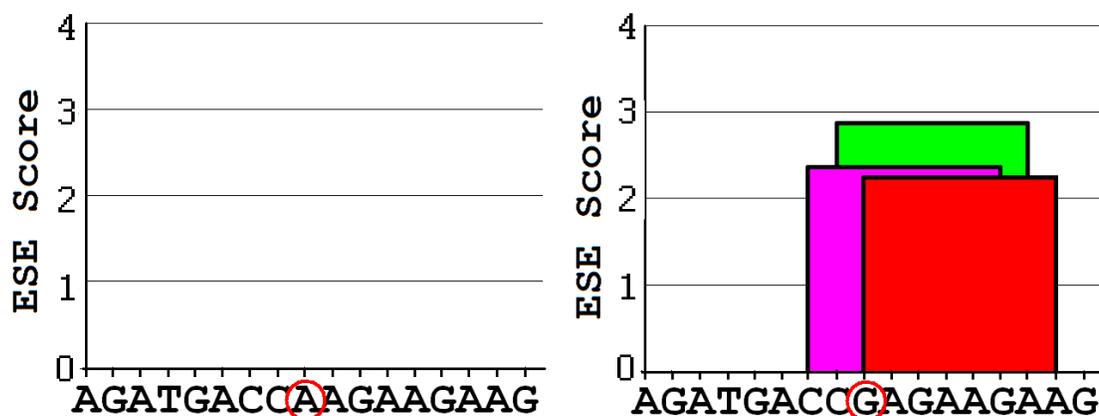


Figure 63: Representative output of ESE scores from ESEFinder
 ESE scores above threshold at rs13077400 for the A (left) and G allele (right), from ESEFinder (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>). Bars represent scores for binding of SF2/ASF (red), SF2/ASF - IgM-BRCA1 (pink), SRp40 (green). No hits for SC35 and SRp55 detected in these sequences. Similar changes at rs13096477 and rs3736312.

The tool utilises position weight matrices to score potential exonic splicing enhancer sequences (Cartegni *et al.*, 2003). As with other prediction tools utilising position weight matrices, these predictions may lack specificity, and have already been clearly identified by the authors as a caveat in its use. This is emphasised by a later study on the positive predictive value of ESEfinder in predicting allele-dependent splicing was 0%, or 0 out of 42 predictions (El Sharway *et al.*, 2009). Again, tools like ESEfinder that use cut-offs may provide a false impression of a definitive prediction, where on a continuous scale of probability, the change may not be that definitive. On the other hand, the study was conducted on a total of 92 samples, based on a combination of 25 whole blood samples, 57 lymphoblastoid cell lines and 10 operative brain tissue biopsies from patients with pharmaco-resistant temporal lobe epilepsy. In view of this heterogeneous sample and limited size, it could be argued that it may have lacked the power to detect allele-dependent splicing. Nevertheless, this is not encouraging.

A newer tool, Human Splicing Finder (<http://www.umd.be/HSF/>) was developed with updated position weight matrices, not just for the spliceosome, but also includes exonic splicing silencing sequences, hnRNP (heterogeneous ribonucleoprotein particle) and branch sites (Desmet *et al.*, 2009). As summarized in

Table 39 below, both rs13077400 and rs3755652 polymorphisms may have an effect on components important for mRNA splicing. As our SNPs of interest around Cassette II are exonic, the analyses for branch sites (which are always intronic) were disregarded. The authors of the

program also evaluated the algorithms using a set of mutations known to alter splicing, and another set of control polymorphisms thought to be unrelated to splicing. This returned a positive predictive value ranging from 22 to 56%, and a negative predictive value of 76 to 95% (Desmet *et al.*, 2009).

SNP	Sequence Position	Linked SR protein	Reference Motif (value 0-100)	Linked SR protein	Mutant Motif (value 0-100)	Variation
rs13077400 (T>C)	912	-	-	SRp40	cttctcg (86.35)	New site
rs3755652 (G>A)	1041	SRp40	tccaag (82.22)	-	-	Site broken
	1042	SF2/ASF	ccaaga (77.85)	-	-	Site broken
	1042	SF2/ASF	ccaaga (75.31)	-	-	Site broken
	1043	-	-	SRp40	ccaaaag (85.03)	New site
	Sequence Position	Enhancer motif reference sequence		Enhancer motif mutant sequence		Variation
rs3755652 (G>A)	1044	-		caaaag		New Site
	1046	-		aaagag		New Site
	Sequence Position	Sironi Motif Reference	Reference silencer (value 0-100)	Sironi Mutant motif	Mutant silencer (value 0-100)	Variation
rs13077400 (T>C)	914	Motif 2 - [T/G]G[T/A]	tcttggtc (63.24)	-	-	Site broken
rs13077400 (T>C)	1040	-	-	Motif 3 - TCTCCCAA	ttccaaa (60.19)	New site
	Sequence Position	Linked hnRNP protein	Reference Motif (value 0-100)	Linked hnRNP protein	Mutant Motif (value 0-100)	Variation
rs13077400 (T>C)	1045	hnRNP A1	aagaga (73.33)	-	-	Site broken

Table 39: Predicted splicing changes by Human Splicing Finder

Predicted changes by Human Splicing Finder (<http://www.umd.be/HSF/>), relating to the two exonic SNPs located within Cassette II (rs13077400 and rs3755652). “-” denotes results below significance thresholds as evaluated by the program algorithm.

An alternative online prediction tool, RESCUE-ESE (Massachusetts Institute of Technology; <http://genes.mit.edu/burgelab/rescue-ese/>), found that only one (rs13077400) of the four exonic SNPs was predicted to potentially affect interactions with exonic splicing enhancers (Fairbrother *et al.*, 2002) (Appendix 16).

There is some consistency of the rs13077400 SNP being predicted to have an increase the probability of functioning as an exonic splicing enhance, as well as reducing probability in being a splicing silencer. This is interesting as the nearest identified splice cassette is known to be an important region of protein-protein interaction (Section 1.9). There is less consistency in the predictions for the other SNPs. Once again, caveats already discussed with position weight matrices also apply here. Additionally, these predictions do not consider the distance from *known* splice sites, which is important for *in vivo* mRNA splicing.

3.12. miRNA (micro RNA)-mRNA interaction prediction

miRNAs are small non-coding mRNAs that typically interact with the 3' UTR of mRNA, usually to silence gene expression, but can rarely stabilize the mRNA and promote gene expression. Of the 93 SNPs examined, none lie within this classical region of action. This is supported by the prediction tools mirBase (<http://www.mirbase.org/>), miRDB (<http://mirdb.org/miRDB/index.html>) and TargetScan (<http://www.targetscan.org/>). The lack of predicted miRNA interaction sequences is particularly important to be considered with the low specificity of these tools. Out of a range of prediction tools, the best positive predictive value is around 10%, and most lie between 0 to 5% (Ben-Moshe *et al.*, 2012). Although the negative predictive values of these tools are not investigated, often in the cases of the prediction tools, it is likely to be higher than the positive predictive value.

3.13. Epigenomic predictions - ENCODE (Encyclopedia Of DNA Elements), NIH Roadmap and the Multiple Tissue Human Expression Resource (MuTHER)

In recent years, there is an increasing drive towards exploring the epigenomic control of gene expression. One major international collaborative effort formed the ENCODE Consortium, which aims to identify functional elements in the human genome. Currently, there is data available from experimental assays exploring DNA methylation, open chromatin status, RNA-binding proteins, RNA profiling, transcription factor binding and histone binding, together with others such as known genetic polymorphisms and nucleosome status.

The data relevant to *SLC4A7* and hypertension is summarized in (Figure 64 and Appendix 17), where 15 of the 93 SNPs of potential interest was noted to be in genomic regions that had signals for DNase hypersensitivity, transcription factor binding (typically by CHIP), CpG islands or nucleosome interactions. Like the rest of this chapter, the summarized data from ENCODE is presented *as is*, and should only be interpreted with caution. The reproducibility of the ENCODE outputs across cell types and experimental methods is not yet known, as is whether it has any predictive value (both positive and negative) when extrapolating to human diseases. These issues and other causes for concerns are discussed further below.

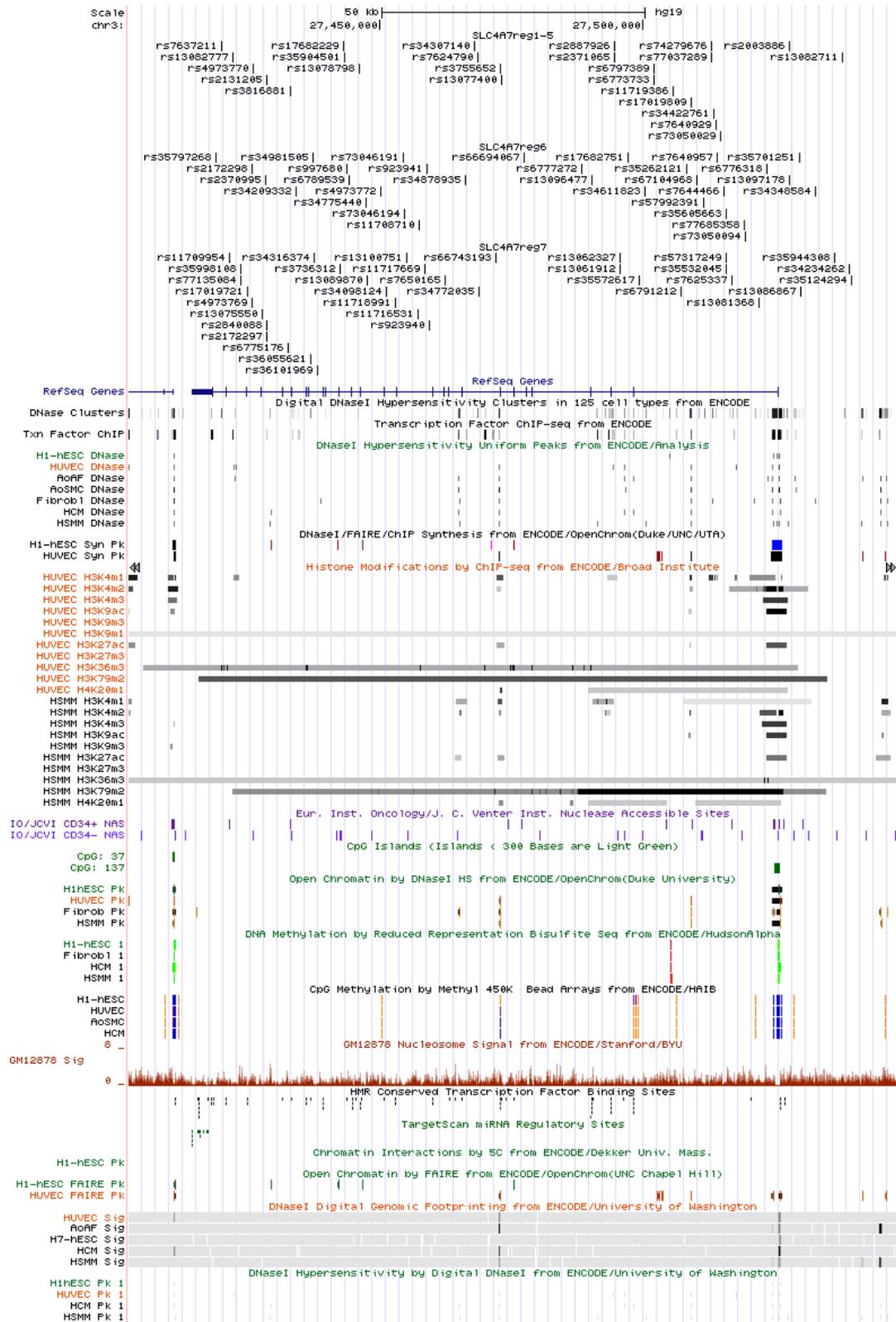


Figure 64: ENCODE outputs for the 93 SNPs of interest
 ENCODE outputs for the 93 SNPs of interest (top), and various results for DNase hypersensitivity, FAIRE, CHIP, histone modifications, CpG modifications and chromatin status in cell types of potential interest.

Although the dataset is continuously expanding, data is concentrated on the “Tier 1” cell lines, GM12878 (lymphoblastoid cell line produced by EBV transformation), K562 (immortalized cell line produced from a patient with chronic myelogenous leukemia) and H1 human embryonic stem cells (obtained commercially from Cellular Dynamics International) (ENCODE Experiment Matrix, 2012), which are experimental cell lines with little relevance to blood pressure regulation.

“Tier 2” cell lines, the next level of cell types with less experimental data available, are typically cancer cell lines, for example HeLa-S3 (immortalized cervical cancer cell line), HepG2 (cell line derived from liver carcinoma) and MCF-7 (epithelial cell line derived from the metastatic pleural effusion of human breast adenocarcinoma), and healthy white blood cells from donors, such as monocytes and B-cells. The only remotely relevant cell type to hypertension in this tier are HUVECs (obtained commercially from Lonza Biosciences) (ENCODE Experiment Matrix, 2012). Although HUVECs do have biological relevance to the study of blood pressure regulation, the use from a commercial source results in an inability to assess the impact of genotype.

“Tier 3” with even less experimental data includes most of the primary cell lines (ENCODE Experiment Matrix, 2012). Of these primary cell lines, not many are directly relevant to blood pressure control. At a stretch, AoSMC (aortic smooth muscle cells), HAoEC (thoracic aortic endothelial cells pooled from two individuals), HBMEC (brain microvascular endothelial cells), HBVSMC (brain vascular smooth muscle cells), HPAEC (pulmonary artery endothelial cells) and HSaVEC (saphenous vein endothelial cells pooled from two individuals) may show some relevance. However, data from these cell lines are sparse. Other tissue of potential relevance (Fetal Adrenal Gland, BC_Adrenal_Gland_H12803N, BC_Brain_H11058N, BC_Kidney_01-11002, Cerebellum_OC, Cerebrum_frontal_OC, Fetal Kidney, Kidney Adult, Kidney_OC) in addition to having limited experimental data, are derived from heterogeneous tissues, and sometimes derived from multiple subjects. Thus these samples cannot attribute any observations to one particular cell type. The current information available does not allow for delineation of which cell type within the tissue is responsible for the accessible data.

Whilst these datasets can hint at the epigenomic control in blood pressure regulation, the lack of information from relevant cells is a severely limiting factor. A further barrier to utilizing ENCODE data for the purpose of this thesis stems from the fact that the data is predominantly derived from one cell line, or in the case of primary cell lines, one individual. This means that the results only

applies to one genotype / haplotype, and thus cannot provide any information on whether a polymorphism may alter the phenotype. There are also skeptics of the interpretation of this large epigenomic dataset. They argue that it may inappropriately attribute a causal role to observational findings, is prone to analytical methods that yield biased errors and inflate estimates of functionality, favours statistical sensitivity over specificity, and emphasizes statistical significance rather than the magnitude of the effect (Graur *et al.*, 2013).

A separate collaboration has resulted in the NIH Roadmap Epigenomics Program also with the aim of developing a community, freely-accessible resource of genome-wide epigenetic maps in a range of human primary cells and tissues (Chadwick, 2012). The two datasets are now possible to view in one combined site, the UCSC Genome Browser website (<http://www.epigenomebrowser.org/>). However, the caveats expressed regarding utilizing ENCODE data would also apply to that of the Roadmap Epigenomics Program.

Another consideration of epigenetic influences would be that of DNA methylation. This is the process of adding a methyl group onto cytosine or adenine DNA nucleotides. DNA methylation occurs mainly in CpG islands (string of cytosine-guanine repeats) at the C5 position, and is carried out by the enzyme DNA methyltransferase. Physiologically, this phenomenon is typically age-dependent (Bell *et al.*, 2012). With methylated nucleotides in promoter regions, this typically reduces gene expression. On the other hand, hypomethylation increases gene expression as seen with some oncogenes and cancers. However, a recent GWAS aiming to detect enrichment of CpG island methylation in blood pressure associated-SNPs found 35 SNPs associated with methylation markers, but none of them relating to *SLC4A7* (Kato *et al.*, 2015).

Recently, the Multiple Tissue Human Expression Resource (MuTHER) was used to generate methylome data from subcutaneous adipose tissue punch biopsies from 648 twins. In turn, a metQTL (methylation-quantitative trait loci) analysis between DNA methylation levels and probabilities of genotypes (minor allele frequency > 5%, info > 0.8) was conducted. This was considered a step forward from previous analyses typically used either whole-blood-derived samples or arrays only targeting promoter regions, which may reveal only a portion of DNA methylation variation. Like that of eQTLs, the dataset can be accessed via the GENEVAR database (<http://www.sanger.ac.uk/resources/software/genevar/>).

Although the results summarised in Table 40 shows multiple SNPs associated with three main CpG islands in different patterns, it should be recalled that this is an observational association and that these SNPs are in strong LD. With this, it is still possible that *any* of the 93 SNPs may be causal, although those positioned closer to the CpG islands would be stronger candidates. This brings up the interesting possibility of the SNP affecting gene expression by the propensity to methylation. On the other hand, similar to other results on the GENEVar database, the pre-set cut off of $P < 0.0001$ might not be stringent enough, noting that the majority of these observed associations were between $P = 10^{-4}$ and 10^{-5} .

	Intergenic (SLC447-EOMES)	5' of SLC447	5' of NEK10
rs1523848			
rs1298223			
rs1212434			
rs4434402			
rs13082713	Green	Black	Black
rs1044384			
rs14148148			
rs13088827			
rs15761213	Green		
rs13097126	Green	Black	
rs1009386			
rs8778138			
rs13085568			
rs17050094			
rs17748518			
rs7626137	Green		
rs15131045	Green	Black	
rs17121240			
rs10505683	Green		
rs1544444			
rs17050029			
rs17480917			
rs14046319			
rs14822703			
rs17317289			
rs16276974			
rs137991391			
rs17170868			
rs17018809			
rs15267121			
rs4974212			
rs13731386	Green		
rs1717133	Green		
rs4972389	Green		
rs14811823	Green		
rs15219817	Green		
rs137487713	Green		
rs13088181	Green		
rs13082827	Green		
rs1371905	Green		
rs14881376	Green	Black	
rs13086477	Green		
rs1717171	Green		
rs16668007	Green		
rs13077460	Green		
rs1745642	Green		
rs8241133	Green		
rs18710111	Green		
rs1624290	Green		
rs14820740	Green		
rs14820815	Green		
rs17650155	Green		
rs121348	Green		
rs1021041	Green		
rs11717669	Green		
rs13708130	Green		
rs13716513	Green		
rs13089751	Green		
rs170441104	Green		
rs170441101	Green		
rs1374991	Green		
rs13088126	Green		
rs1871712	Green		
rs13088870	Green		
rs14775440	Green		
rs13081788	Green		
rs1089539	Green		
rs1095100	Green		
rs10908501	Green		
rs13082229	Green		
rs1778132	Green		
rs130101909	Green		
rs101011514	Green		
rs148881505	Green		
rs10054012	Green		
rs14898181	Green		
rs13810881	Green		
rs10715176	Green		
rs1313205	Green		
rs1272787	Green		
rs1486088	Green		
rs13709915	Green		
rs130935501	Green		
rs4871770	Green		
rs1887738	Green		
rs1272788	Green		
rs13098723	Green		
rs17148484	Green		
rs10948108	Green		
rs13086777	Green		
rs1309644	Green		
rs17617711	Green		
rs15761213	Green		

Table 40: Summarised met-QTL data from the MuTHER / GENEVar database. Summarised met-QTL data from the MuTHER / GENEVar database (<http://www.scandb.org>) output identifying gene transcripts that vary with each SNP (one row each, and rs13082713 identified in grey). Each column indicates region of CpG island. Green cells denotes SNP-dependent DNA methylation $P < 0.0001$, black cells denote no SNP-expression associated found, open cells denote no SNP information available. The SNP IDs are not intended to be read in this format, as full results in Appendix 18.

3.14. Potential inducibility of *SLC4A7* expression

While most of the above predictions are based on the impact of genotype on gene expression in an unstimulated state, it should also be considered if the genetic polymorphism has a functional impact in a *stimulated* state. It has previously been identified that acid-base regulating transporters are subject to endocrine stimulation, and particularly relevant to blood pressure control, the renin-angiotensin-aldosterone axis (Section 1.12).

To consider the impact of a hormonal or pharmacological challenge on NBCn1 expression, some limited clues may be gained from the online dataset, Connectivity Map (Broad Institute; <http://www.broadinstitute.org/cmap/>). This is an online interface allowing access to collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules (Lamb *et al.*, 2006).

Based on the data on NBCn1 expression, there does not appear to be strong support for the hypothesis of angiotensin II or aldosterone inducing NBCn1 expression (Figure 65), with gluco- and mineralocorticoids being as equally spread out along the Connectivity Score spectrum as the renin-angiotensin-aldosterone axis antagonists. Tacrolimus, a calcineurin inhibitor, was also highlighted as the potential interaction between NBCn1 and calcineurin inhibitors have been discussed in Section 1.9. In Connectivity Map data, tacrolimus appears to reduce NBCn1 expression, which the direction of action is at least congruent with the finding that tacrolimus also attenuated NBCn1 activity in murine mesenteric arteries (Danielsen *et al.*, 2013). However, it is the only calcineurin inhibitor assessed in the dataset.

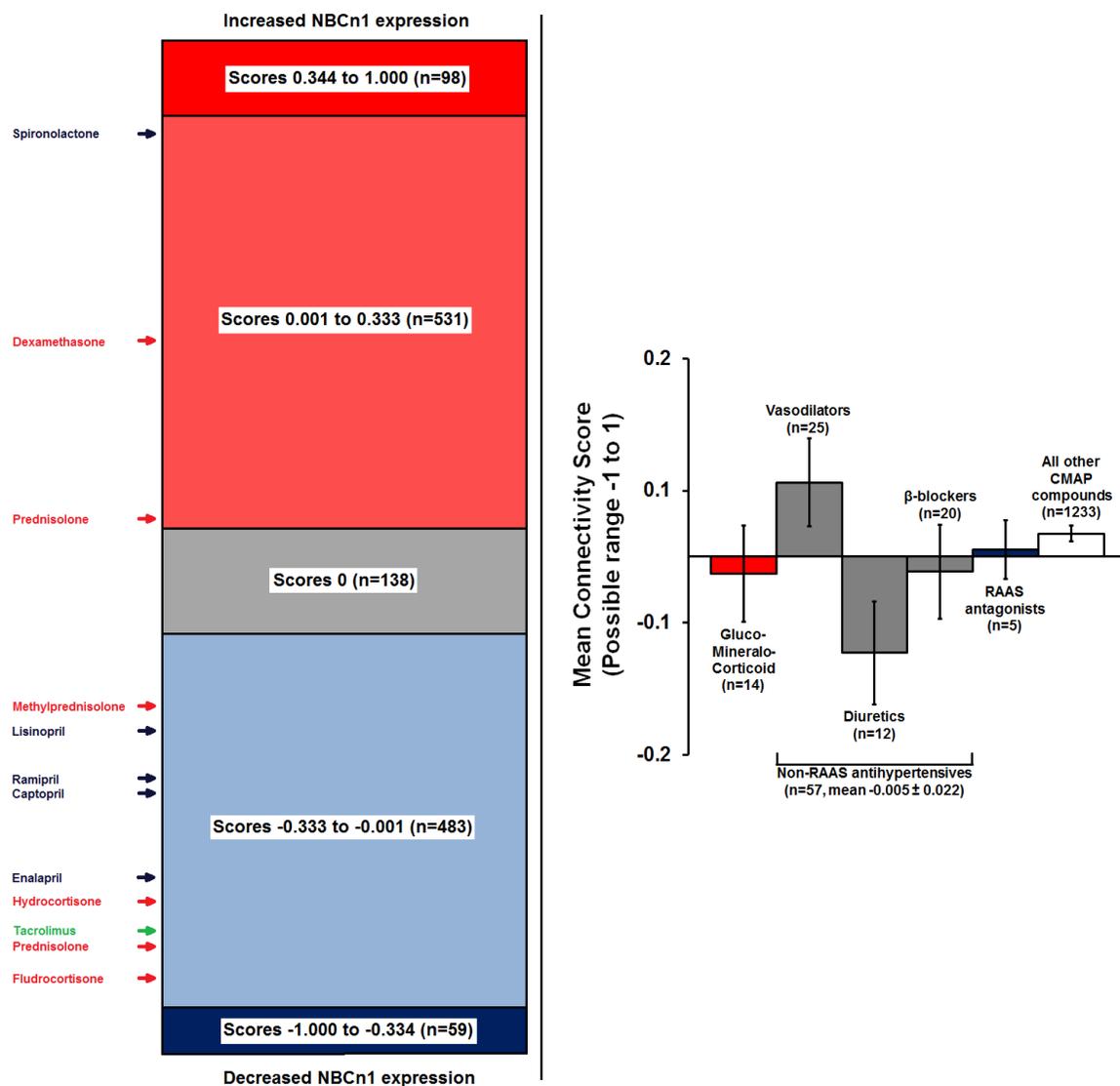


Figure 65: Connectivity Score output for NBCn1

(Left) Ranking of small molecules and their effect on NBCn1 expression, with Connectivity Scores of -1 (decreased expression) to 1 (increased expression). Bars on the left of the scores indicated glucocorticoids (red), renin-angiotensin-aldosterone axis antagonists (blue) and other anti-hypertensive medications (grey). Small molecules of particular interest are listed in the colour of their grouping. (Right) Mean Connectivity Score based on small molecule classification, represented as mean ± SEM. See Appendix 19 for full results.

However, this lack of supportive information should be considered with several caveats. All of the experiments are conducted on cancer cell lines - MCF7 (breast cancer), PC3 (prostate cancer) and HL60 (promyelocytic leukaemia). These experimental models may have only limited biological relevance to blood pressure regulation. Additionally, for approximately 70% of the pharmacological agents tested, the overall result was an amalgamation of 4 or less experiments, and is an amalgam of experiments on different cell lines and varying treatment concentrations.

Perhaps most importantly, most of the pharmacological antagonism is conducted in the absence of their physiological agonist or substrate while in an *in vitro* cell culture system - for example an angiotensin converting enzyme inhibitor (ramipril) in absence of its substrate (angiotensin I).

3.15. Other theoretical molecular mechanisms not assessed by prediction tools

Whilst the above already provides quite a large amount of information to carefully interpret, it should be noted that there are other potential molecular mechanisms where an SNP may lead to changes in gene expression or product function. This may relate to mRNA conformation and stability (leading to changes in gene expression) and protein localization (leading to changes in product function).

3.16. Limited supportive bioinformatics evidence at rs820430

Although this chapter has already explored the myriad of pitfalls that bioinformatics tools may provide, it is worth discussing these prediction tools in context of rs820430. As previously shown in Figure 10, rs820430 is in high LD with 10 other SNPs, all clustered within 16kb of each other, of which none are within 140kb from the start of *SLC4A7*. As none of these SNPs are near *SLC4A7*, there is no scope for impact of protein function (by amino acid substitution), splicing (by exonic splicing enhancers) or miRNA interactions (within the 3'-UTR). This leaves the potential impacts on transcription factor binding, and epigenetic impact of nucleosome interactions.

In short, the only finding of note was that the GeneVar database indicated that rs820430 was weakly associated with *SLC4A7* ($P < 0.05$) only with the JPT, but not the CHB population. This makes the interpretation in context of the Lu *et al.*, Chinese GWAS difficult. However, the caveats as outlined in Section 3.3 remains. Outputs from other eQTL databases (ScanDB), tissue-specific eQTLs (GTEx), methylation and CpG islands (MuTHER) and nucleosome occupancy (Leiden University) were essentially negative.

3.17. Summary of prediction tool analyses

Prediction tools may occasionally be helpful in prioritizing lines of investigations. However, several key limitations should always be remembered, in that:

- the algorithms are often derived from experimental models that are not always relevant to blood pressure regulation (or any other investigated disease),
- the insufficient specificity can readily result in false positives, especially if multiple predictions are undertaken,
- the achievement of statistical significance (or a “cut-off score”) is not necessarily biologically significant, and provides an artificial dichotomy when in reality at best represents a continuum of likelihood,
- the observational nature of the data means any observed correlation is not necessarily a causative relationship,
- there are occasions where a multitude of tools may provide conflicting outputs for the same prediction, or in other cases where these tools provide concurring outputs based on similar training datasets may be falsely reassuring,
- the “black box” nature of the algorithms lacks transparency,
- with the primary data often inaccessible to confirm appropriateness of data and relevance to algorithm, and last but not least,
- it often neglects the underlying pre-test probability as derived from known physiological understanding of the proteins of interest.

With this, outputs from these prediction tools should always be treated with a large degree of caution.

4. Results

4.1. SLC4A7 genotyping

229 HUASMC and 178 HUVEC samples were genotyped by the KASP method for the lead SNP at the SLC4A7 loci (rs13082711; Ehret *et al.*, 2011) and two others in high LD (rs13096477 and rs3755652), and the results are in

Table 41,

Table 42 and

Table 43.

HUASMC		rs13082711 (T>C)	rs13096477 (T>C)	rs3755652 (A>G)	Haplotype
Genotype result	Major homozygote	146 (64.3%)	142 (62.3%)	133 (59.6%)	125 (64.8%)
	Heterozygote	72 (31.7%)	78 (34.2%)	82 (36.8%)	63 (32.6%)
	Minor homozygote	9 (4.0%)	8 (3.5%)	8 (3.6%)	5 (2.6%)
	Unclassified	2	1	2	N/A
	Insufficient sample	0	0	4	N/A
Calculated frequency	Major allele frequency	80.2%	79.4%	78.0%	81.1%
	Minor allele frequency	19.8%	20.6%	22.0%	18.9%
Hardy-Weinberg predicted	Major homozygote	145.9 (64.3%)	143.7 (63.0%)	135.8 (60.9%)	126.9 (65.8%)
	Heterozygote	72.2 (31.8%)	74.6 (32.7%)	76.5 (34.3%)	59.2 (30.7%)
	Minor homozygote	8.9 (3.9%)	9.7 (4.2%)	10.8 (4.8%)	6.9 (3.6%)

Table 41: Genotyping of HUASMC samples

Genotypes, allele frequencies and Hardy-Weinberg predicted distributions at rs13082711, rs13096477 and rs3755652 of 229 HUASMC samples. "Haplotype" for heterozygotes refers to a concordance of heterozygosity across all three SNPs, as it is technically impossible to be certain that they form the same haplotype without sequencing. "Unclassified" denotes an inability to call genotype by KASP assay. Not all columns add to exactly 100% as numbers are rounded to one decimal place.

HUVEC		rs13082711	rs13096477	rs3755652	Haplotype
Genotype result	Major homozygote	114 (64.4%)	110 (62.5%)	104 (60.8%)	100 (62.9%)
	Heterozygote	56 (31.6%)	60 (34.1%)	61 (35.7%)	53 (33.3%)
	Minor homozygote	7 (4.0%)	6 (3.4%)	6 (3.5%)	6 (4.2%)
	Unclassified	1	2	6	N/A
	Insufficient sample	0	0	1	N/A
Calculated frequency	Major allele frequency	80.2%	79.5%	78.7%	79.6%
	Minor allele frequency	19.8%	20.5%	21.3%	20.4%
Hardy-Weinberg predicted	Major homozygote	113.9 (64.4%)	111.4 (63.3%)	105.8 (61.9%)	100.6 (63.3%)
	Heterozygote	56.2 (31.7%)	57.3 (32.5%)	57.4 (33.6%)	51.7 (32.5%)
	Minor homozygote	6.9 (3.9%)	7.4 (4.2%)	7.8 (4.6%)	6.6 (4.2%)

Table 42: Genotyping of HUVEC samples

Genotypes, allele frequencies and Hardy-Weinberg predicted distributions at rs13082711, rs13096477 and rs3755652 of 178 HUVEC samples. “Haplotype” for heterozygotes refers to a concordance of heterozygosity across all three SNPs, as it is technically impossible to be certain that they form the same haplotype without sequencing. “Unclassified” denotes an inability to call genotype by KASP assay. Not all columns add to exactly 100% as numbers are rounded to one decimal place.

The genotype distributions for both HUASMCs and HUVECs were in Hardy-Weinberg equilibrium ($P > 0.05$ for deviation from equilibrium by chi-squared test).

All 6 genotyping runs had a call rate of $>95\%$. As there were multiple genotyping runs, there were samples that were present in more than one genotyping batch. Inter-assay concordance of 97.8% (88/90), 97.9% (94/96) and 96.1% (73/76) for the rs13082711, rs13096477 and rs3755652 respectively. Where there are discordant results from two runs, the samples are “unclassified”. There was no intra-assay variability tests conducted. Of those assessed, 39 KASP assays showed

100% (14/14 HUASMCs, 16/16 HUVECs and 9/9 renal biopsies) concordance with gold standard DNA sequencing. In 14 paired samples with HUASMCs and HUVECs from single donors, all three SNPs had 100% matching genotype by KASP assay. With the SNPs in high LD, haplotypes were also identified. Samples with a full data set are shown below (Table 43).

	Major homozygotes (TT/TT/AA)	Heterozygotes (CT/CT/AG)	Minor homozygotes (CC/CC/GG)	Non-major haplotype (<3% for each individual combination)
HUASMCs	125 (56.8%)	63 (28.6%)	5 (2.3%)	27 (12.3%)
HUVECs	100 (59.5%)	52 (31.0%)	6 (3.6%)	10 (6.0%)

Table 43: Haplotypes at rs13082711, rs13096477 and rs3755652, for HUASMCs and HUVECs
“Haplotype” for heterozygotes refers to a concordance of heterozygosity across all three SNPs, while it is impossible to be certain that they form the same haplotype without sequencing.

The genotype distributions for both HUASMCs and HUVECs were in Hardy-Weinberg equilibrium when considering the three main haplotypes ($P > 0.05$ for deviation from equilibrium chi-squared test), and there were no differences in the distribution between these two sample groups ($P > 0.05$). Notably there was no difference in the proportion of samples with incomplete datasets (HUASMCs 3.9%, 9/229; HUVECs 5.6%, 10/178; $P > 0.05$).

213 HUASMC and 174 HUVEC samples were genotyped by the KASP method for the Chinese GWAS lead SNP at the *SLC4A7* loci (rs820430; Lu *et al.*, 2014).

		HUASMC	HUVEC
Genotype result	Major homozygote	95 (44.8%)	73 (42.7%)
	Heterozygote	91 (42.9%)	77 (45.0%)
	Minor homozygote	26 (12.3%)	21 (12.3%)
	Unclassified	1	3
Calculated frequency	Major allele frequency	66.3%	65.2%
	Minor allele frequency	33.7%	34.8%
Hardy-Weinberg predicted	Major homozygote	93.1 (43.9%)	72.7 (42.5%)
	Heterozygote	94.8 (44.7%)	77.6 (45.4%)
	Minor homozygote	26 (11.4%)	21 (12.1%)

Table 44: Genotyping of HUASMC and HUVEC samples at rs820430
Genotypes, allele frequencies and Hardy-Weinberg predicted distributions at rs820430 of 213 HUASMC and 174 HUVEC samples. "Unclassified" denotes an inability to call genotype by KASP assay. Not all columns add to exactly 100% as numbers are rounded to one decimal place.

4.2. Gene expression

4.2.1. End-point reverse-transcriptase polymerase chain reaction (RT-PCR)

In a panel of HUASMCs, HUVECs and biopsies of non-diseased poles of cancer nephrectomies, *SLC4A7* mRNA was detected by end-point RT-PCR, but not of the surrounding genes *NEK10* and *EOMES* (Figure 66). Primers for *SLC4A7* spanned exons 2 to 5, and therefore detected the majority of *SLC4A7* isoforms.

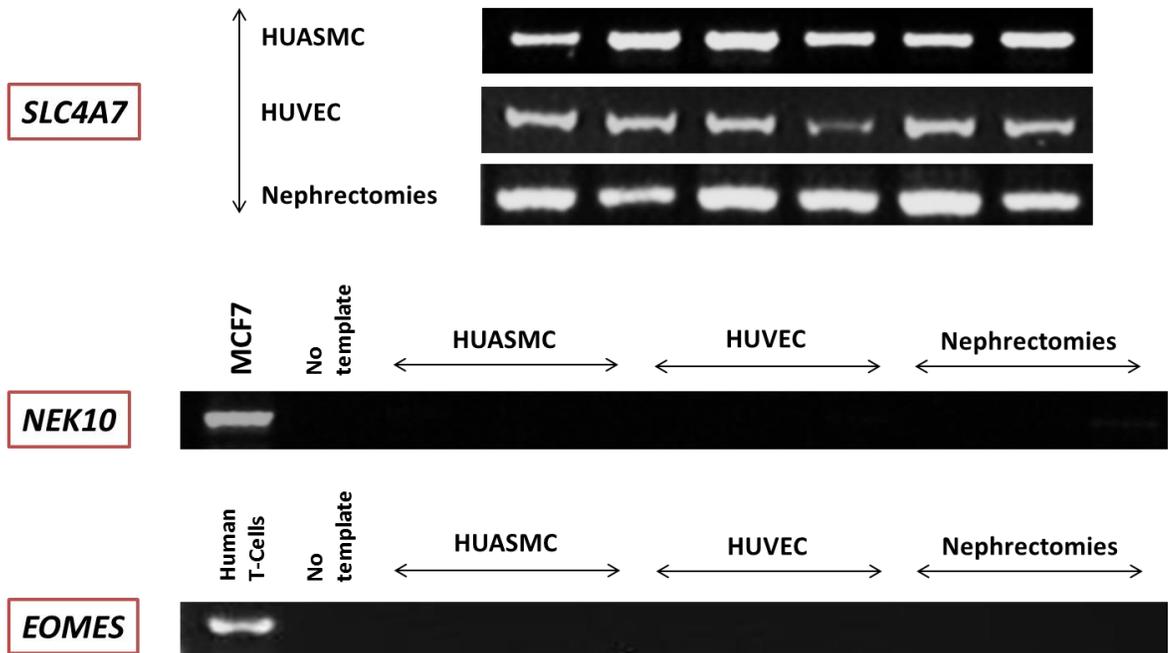


Figure 66: End-point RT-PCR (*SLC4A7*, *NEK10* and *EOMES*) for HUASMCs, HUVECs and nephrectomy biopsies

MCF7 and human T-lymphocytes were positive controls.

Input of 20ng reverse-transcribed RNA for each reaction / lane.

4.2.2. Allelic imbalance analysis

Known heterozygotes at three SNPs rs13082711, rs13096477 and rs3755652 (which are all in high LD) were selected for allelic imbalance analysis. The PCR amplicon surrounding the SNP was sequenced and analysed by PeakPicker. In HUASMCs, HUVECs and nephrectomy biopsies, the risk (minor) allele C at rs13096477 was preferentially detected as compared to the protective T (major) allele (Figure 67).

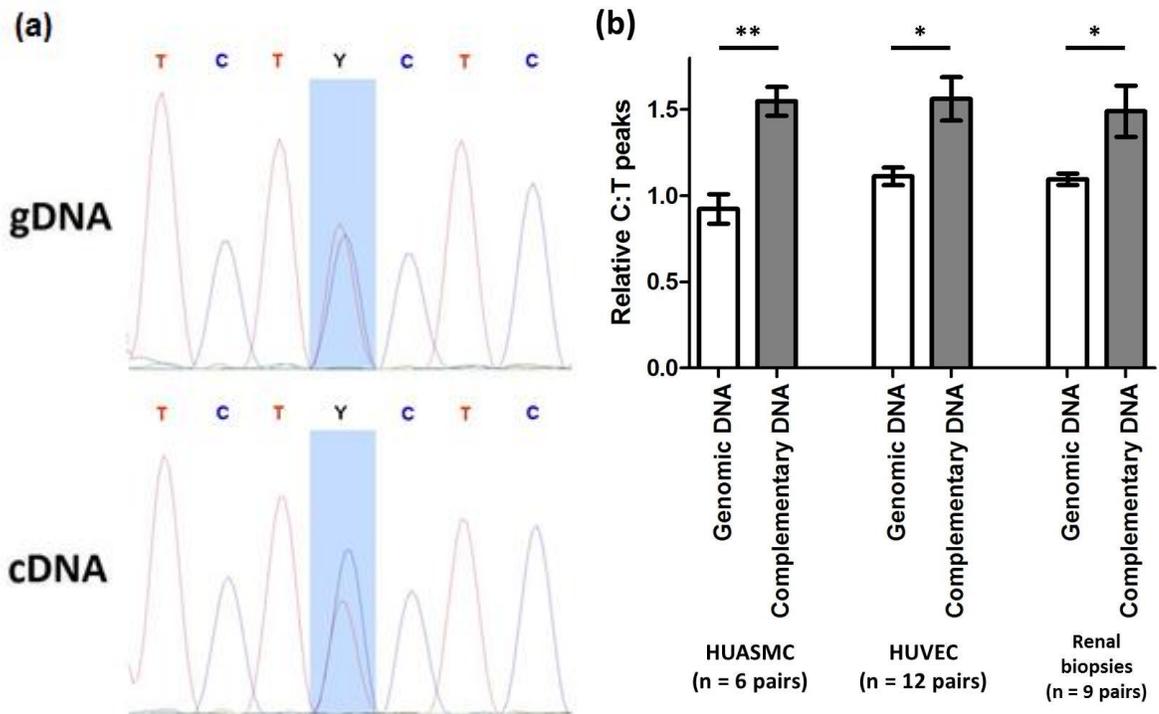


Figure 67: Allelic imbalance analyses using heterozygote samples at the exonic SNP rs13096477. Allelic imbalance analyses using heterozygotes at an exonic SNP (rs13096477), which is in high LD with the GWAS lead SNP (rs13082711). (a) Representative chromatographs of genomic DNA (gDNA) and complementary DNA (cDNA) for the same sample, with the C/T SNP rs13096477 highlighted in blue. (b) Bar chart of normalized C/T peak height ratios calculated using PeakPicker, where the C allele is preferentially observed in cDNA of HUASMC and renal biopsies. * $P < 0.05$, ** $P < 0.01$, paired t-test.

4.2.3. Quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR)

173 HUASMC and 145 HUVEC cDNA samples were successfully assayed in the qRT-PCR experiment. Quality control criteria were set for with C_t standard deviation of replicates >1 for at least one of the three probes used (18S, *SLC4A7* long transcript and *SLC4A7* short transcript). The samples were obtained between 2010 and 2014. All samples shown were concordant for genotypes at rs13082711, rs13096477 and rs3755652. There were no statistically significant differences in ddC_t values between the genotypes in either long or short transcripts in both cell types (Figure 68).

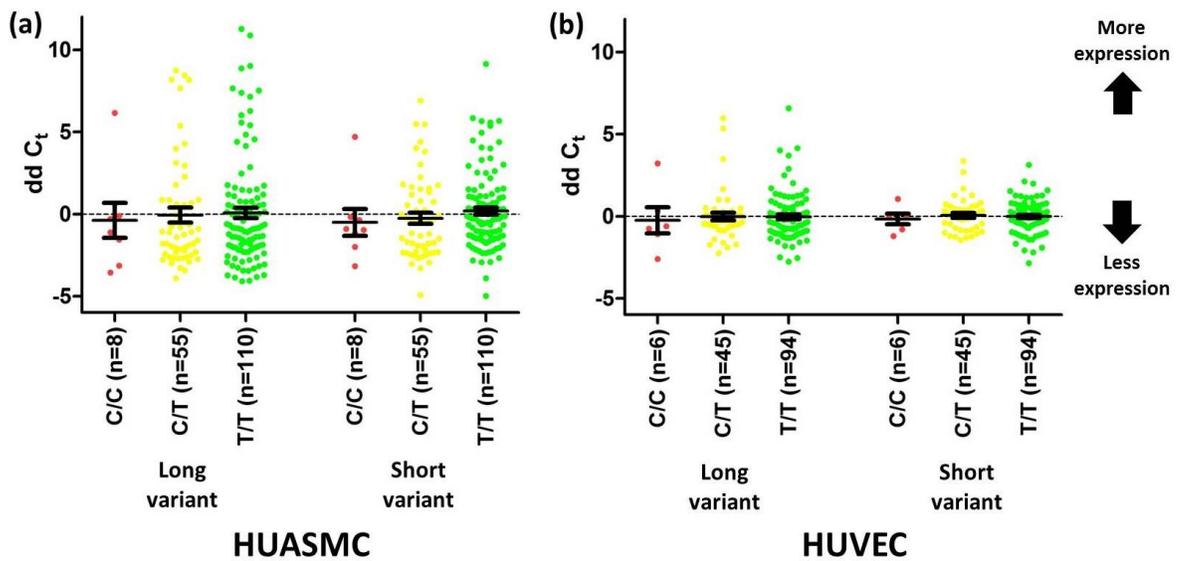


Figure 68: qRT-PCR of HUASMC and HUVEC cDNA stratified by rs13082711 genotype ddC_t was calculated by subtracting the corresponding C_t values of the 18S control probe from C_t values of the text probe (dC_t), and later referenced to the population mean. Positive value denotes increased transcript levels. No difference between genotypes for each variant by one-way ANOVA.

When considering the difference between HUASMC and HUVEC samples, HUVECs have a larger number of transcripts for both the long variant (difference in dC_t of 1.73 ± 1.58) and short variant (difference in dC_t of 3.79 ± 1.33) (14 paired samples, Figure 69a). The difference in dC_t may extrapolate to a fold difference of 4.3x and 14.9x for the long and short variants respectively. There is no genotype impact on the ratio of long to short variants, but there is a cell type difference, where HUASMCs have a larger proportion (despite lower overall expression) levels of long variant transcripts (difference in dC_t of 3.45 ± 0.18 , approximating to 11.9x fold difference) (Figure 69b).

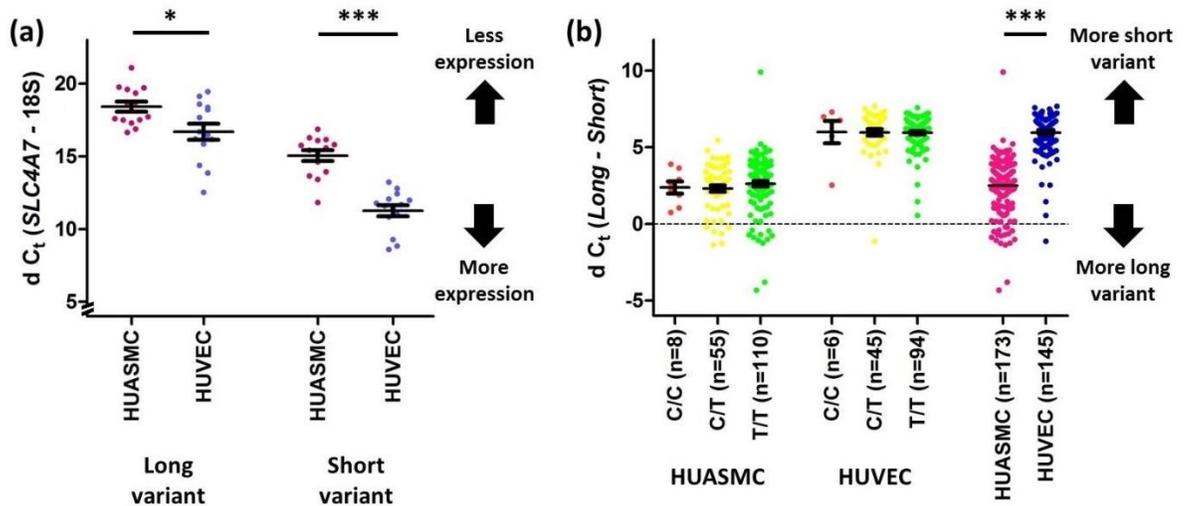


Figure 69: qRT-PCR of HUASMC and HUVEC cDNA, stratified by rs13082711 genotype or cell type (a) 14 pairs of HUASMC and HUVEC samples from the same umbilical cord were assessed for SLC4A7 mRNA expression, expressed as the difference in C_t of SLC4A7 and 18S transcripts. Positive value denotes increased transcript levels. (b) Ratio of long and short transcripts in HUASMCs and HUVECs, expressed as the difference of C_t of the two variants. Positive value denotes greater proportion of short variants.

* $P < 0.05$, *** $P < 0.001$. Groups in (A) compared by paired t-test. Groups in (B) compared by unpaired t-test.

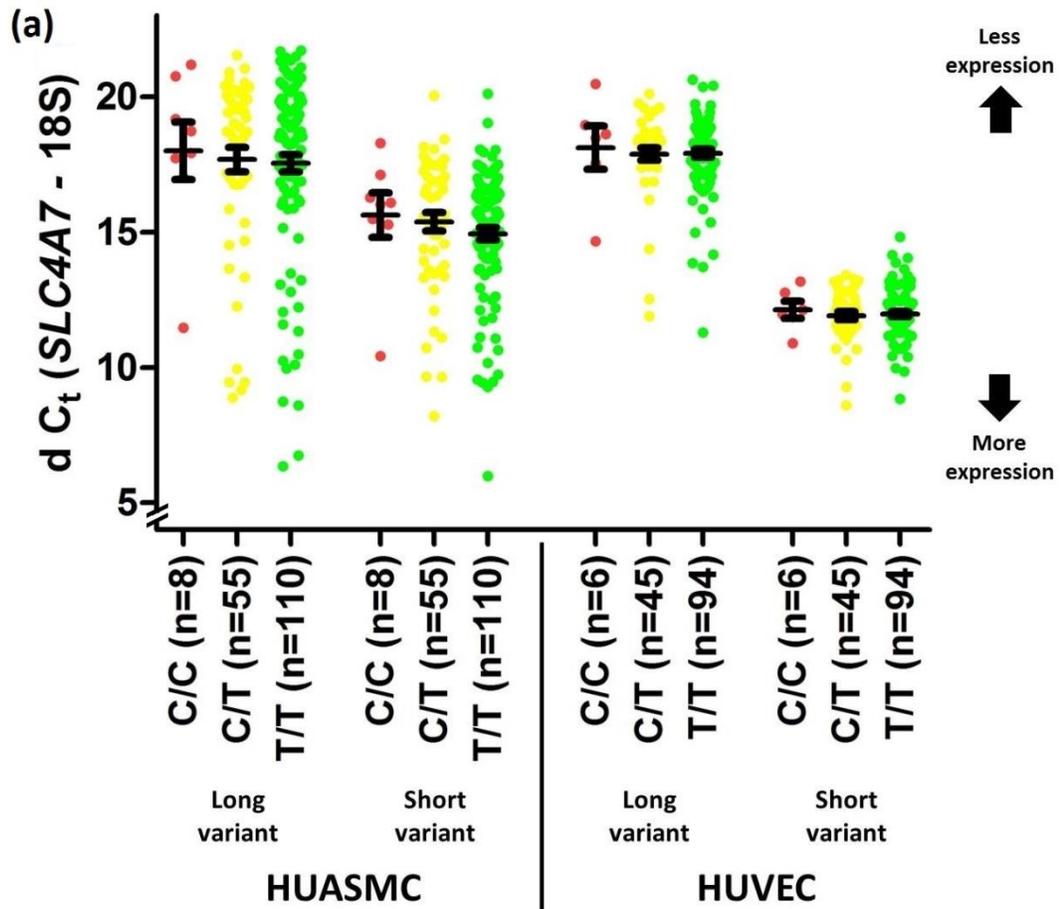
No difference between genotypes for each variant by one-way ANOVA.

It is not possible to directly compare the C_t of the long and short variants as to determine which is more prevalent, as there may be different PCR efficiencies, and it should be recalled that the primers of the short variant could also technically detect that of the long variant, despite the single end-point PCR product as seen in the Methods Section 2.7.4.

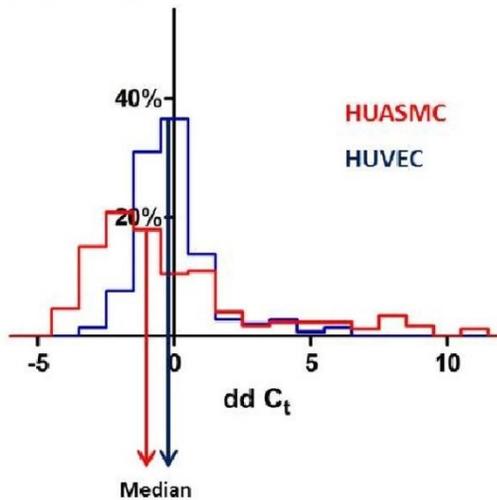
From Figure 68, it is apparent that there is a large distribution of ddC_t values, regardless of genotype for both HUASMCs and HUVECs (although larger in HUASMCs). This is expressed in dC_t values in

Figure 70a. There is an apparent range of C_t of up to 12, even when excluding outliers, approximating to 2^{12} -fold (4096x) difference. Even with more conservative estimates of PCR efficiency (e.g. 80%), this still equates to differences in the magnitude of thousand-fold. A histogram showing the distribution of ddC_t (by definition a mean of 0) shows a positively skewed distribution (

Figure 70b and c) with interquartile ranges that are large (Table 45) relative to expected differences between genotypes.



(b) Long variant



(c) Short variant

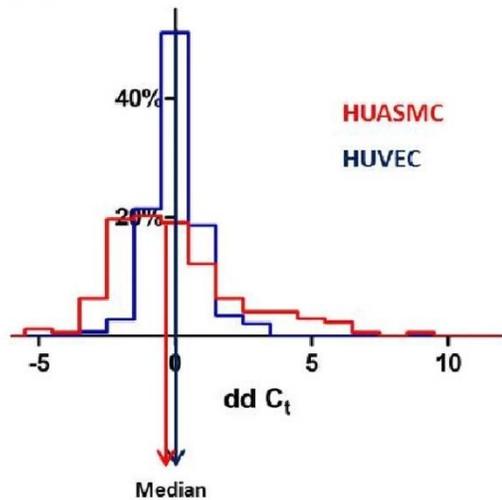


Figure 70: qRT-PCR results demonstrating a skewed distribution with a long tail
 (a) qRT-PCR of HUASMC and HUVEC cDNA expressed as $dd C_t$ by subtracting the corresponding C_t values of the 18S control probe from C_t values of the text probe. Smaller value denotes increased transcript levels. No difference between genotypes for each variant by one-way ANOVA.
 (b) and (c) Histograms of $dd C_t$ showing percentage of sample falling within C_t bins of 1.

	HUASMC		HUVEC	
	Long variant	Short variant	Long variant	Short variant
Standard deviation ddC_t^*	3.313	2.395	1.503	0.949
Lower quartile ddC_t	-2.200	-1.689	-0.852	-0.528
Median ddC_t	-0.938	-0.372	-0.242	0.017
Upper quartile ddC_t	0.862	1.075	0.396	0.457
IQR (ddC_t)	3.062	2.764	1.248	0.985
IQR (fold difference in transcript)**	9.351	7.792	3.375	2.979

Table 45: Distribution statistics

* For illustrative purposes, accepting that standard deviations are not ideal representations of variance in non-normally distributed datasets

** Fold difference calculated as equal to $1+2^{IQR}$. As before this carries the caveat that it assumes 100% PCR efficiency which is improbable.

Similar to the results for rs13082711, there were no statistically significant differences in qRT-PCR C_t values between the rs820430 genotypes (Figure 71). This includes the comparison of C/C vs the combined group of C/T and T/T for HUASMCs.

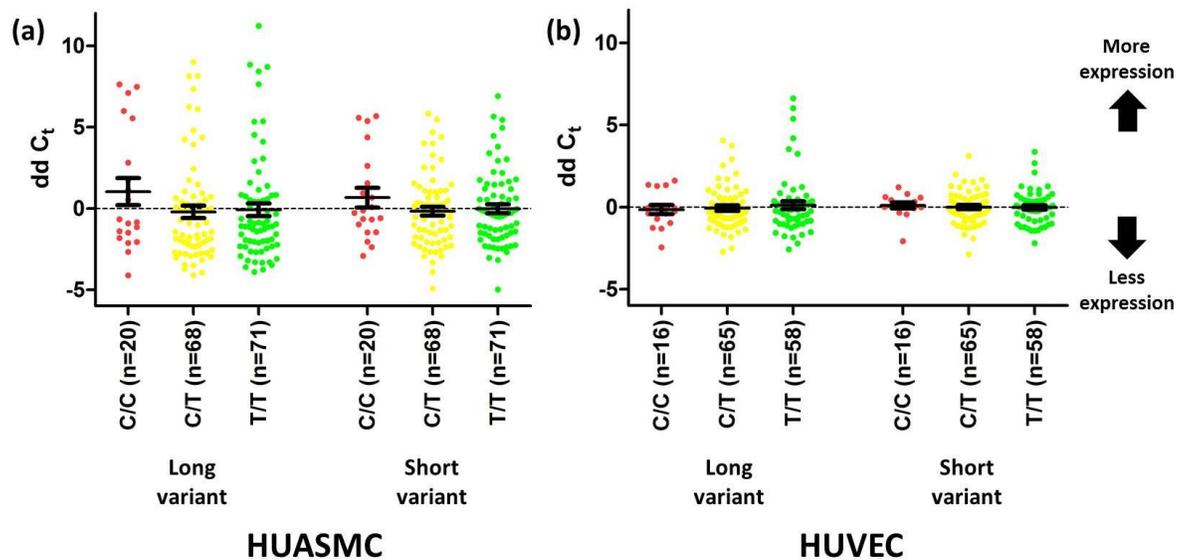


Figure 71: qRT-PCR of HUASMC and HUVEC cDNA stratified by rs820430 genotype ddC_t was calculated by subtracting the corresponding C_t values of the 18S control probe from C_t values of the text probe (dC_t), and later referenced to the population mean. Positive value denotes increased transcript levels.

No difference between genotypes for each variant by one-way ANOVA.

4.2.4. NBCn1 protein immunoblots for total cellular protein

Samples with known genotypes at rs13082711 were assessed for protein expression by protein immunoblots. There is an allele-dose effect, with the greatest signal with minor homozygotes, followed by heterozygotes and lowest signal with major homozygotes (

Figure 72). This is present when analyzing the larger or smaller isoform, as well as the combined total.

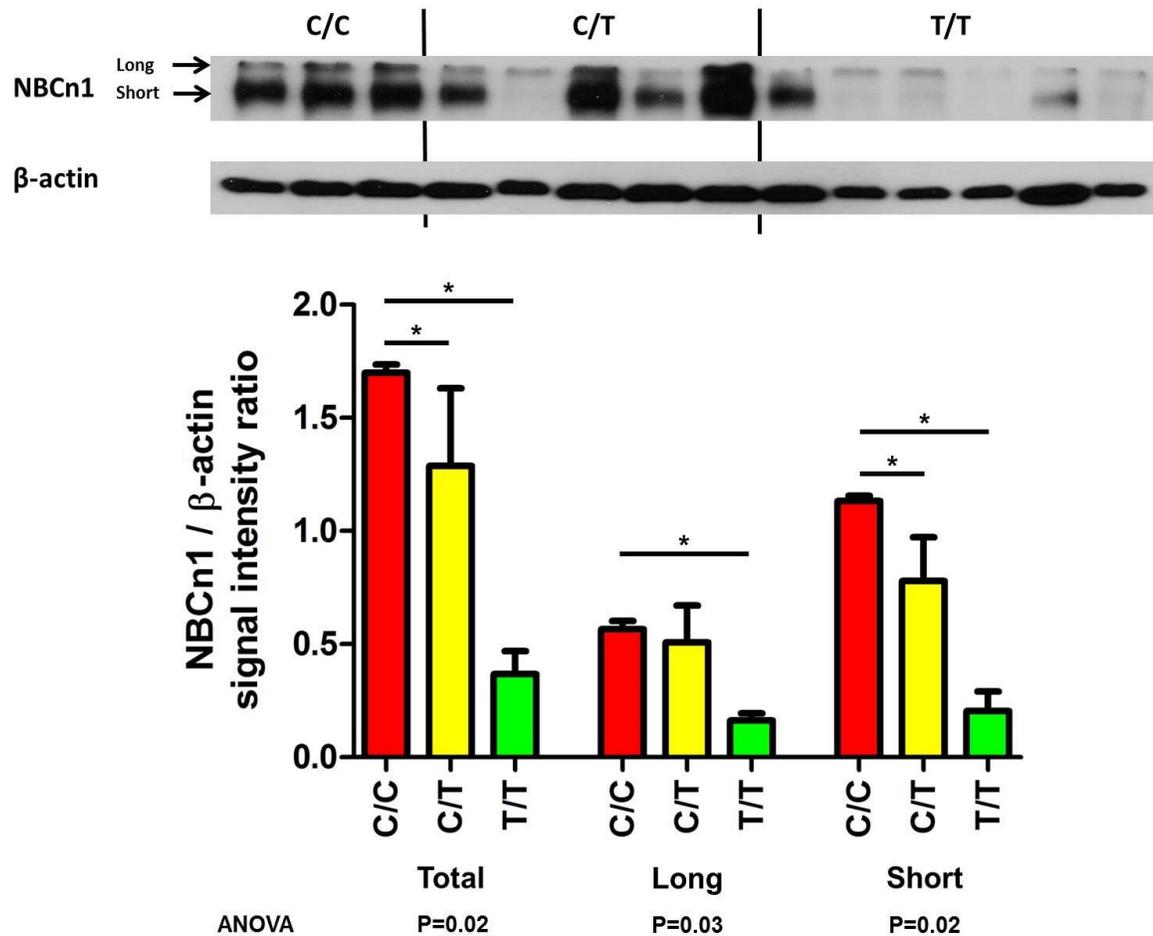


Figure 72: Protein immunoblots for NBCn1 with HUASMC samples stratified by genotype at rs13082711

(a) Immunoblots for samples with genotype at all three SNPs. The two bands for NBCn1 approximate to the 136kDa and 127kDa predicted larger and smaller isoforms (with and without Cassette II). (b) Bar graph for the signal intensities normalized to β -actin signal.

Comparisons across the three genotypes by one-way ANOVA.

*P<0.05 after correcting for multiple comparisons

There were no statistical differences between percentage of overall signal represented by the longer variant (C/C, 33.2 \pm 2.8%; C/T 39.9 \pm 13.7%; T/T 52.8 \pm 22.1; one-way ANOVA P>0.05).

Unlike HUASMCs, HUVECs did not show a genotype-dependent NBCn1 total protein expression (Figure 73, NBCn1:β-actin ratio C/T 1.00 ± 0.13 ; T/T 1.00 ± 0.15).

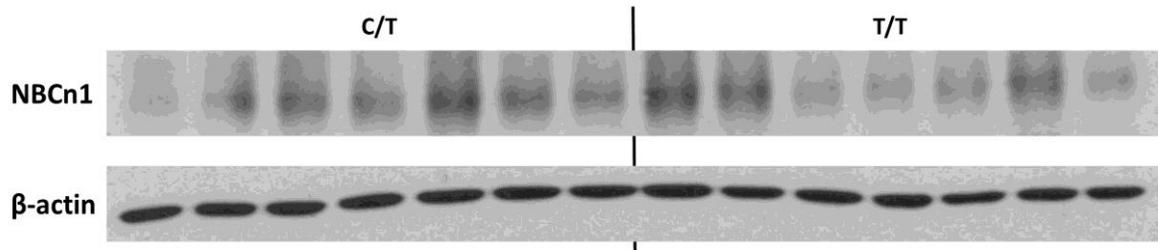


Figure 73: Protein immunoblots for NBCn1 with HUVEC samples stratified by genotype at rs13082711
There are two signals for NBCn1 present.

4.2.5. Isoelectric focusing

HUASMC samples representing the different genotypes at the non-synonymous SNP rs3755652, resulting in the E326K amino acid, was examined by isoelectric focusing. There were no differences in the isoelectric focusing of NBCn1 based on genotype (Figure 74).

rs13082711	C/C	T/T	C/T
rs3755652	G/G	A/A	A/G
Amino acid	326K	326E	Both

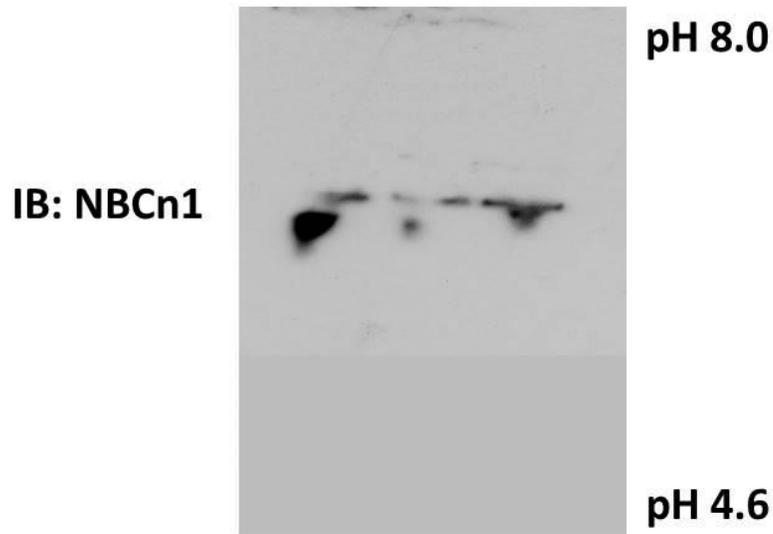


Figure 74: Isoelectric focusing of HUASMCs whole cell lysates

Isoelectric focusing of whole cell lysates of HUASMCs did not reveal the predicted changes (Section 3.9) in mobility for the different E326K variants. The pH gradient is identified on the right. Representative blot of three experiments, all showing no difference in electrophoretic mobility.

4.3. Studies into molecular mechanisms underlying differences in gene expression

Following the identification of a genotype-associated effect on *SLC4A7* gene transcription in both the assessed cell types (Section 4.2.2), and total cellular NBCn1 expression at least in HUASMCs (Section 4.2.4), the next set of experiments aimed to explore the underlying molecular genetic mechanisms.

4.3.1. Chromatin immunoprecipitation (ChIP)

A ChIP assay using HUASMCs was initially conducted attempting to identify which, if any, allele/haplotype is preferentially transcribed. To do this, 2 ng of anti-RNA polymerase II (Pol II) was used to immunoprecipitate sections of DNA that were being actively transcribed by RNA polymerase. Allelic imbalance with heterozygotes was used to improve the sensitivity and reduce external confounders by having a system where both alleles/haplotypes are exposed to the same environment.

There was no significant difference between the input chromatin and the immunoprecipitated sample (Figure 75). However, there was a striking reduction in the C allele : T allele ratio in the chromatin samples as compared to what would be expected if there were no preferential chromatin binding. This finding cannot be interpreted without further experiments as due to the chemistry of Sanger sequencing (see Section 2.6.1), there may be differences in the C and T chromatograph signals purely due to experimental conditions. With this, the study progressed to the formaldehyde-assisted isolation of regulatory elements (FAIRE) assays to determine if the T (major) allele at rs13096477 is preferentially bound.

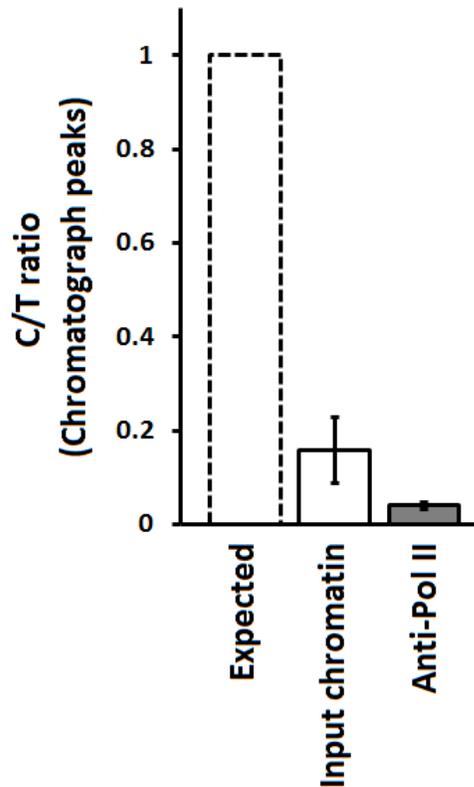


Figure 75: Chromatin immunoprecipitation (ChIP) assay with anti-Pol II antibody Allelic imbalance at rs13096477 of ChIP products. rs13096477 was chosen as a marker for the 93 SNPs in high LD.

No significant difference between input chromatin and anti-Pol II. It was not appropriate to conduct statistical tests to compare the expected C/T ratio of heterozygotes with either experimental group.

n=5 pairs

4.3.2. Formaldehyde-assisted isolation of regulatory elements (FAIRE) assays

To continue from the ChIP studies (Section 4.3.1), FAIRE assays in both HUASMCs and HUVECs were conducted to identify sections of DNA that are protein-bound (chromatin) within a cellular environment. At rs13096477, the major/protective (T) allele is preferentially chromatin-bound in HUASMCs, but not HUVECs (Figure 76).

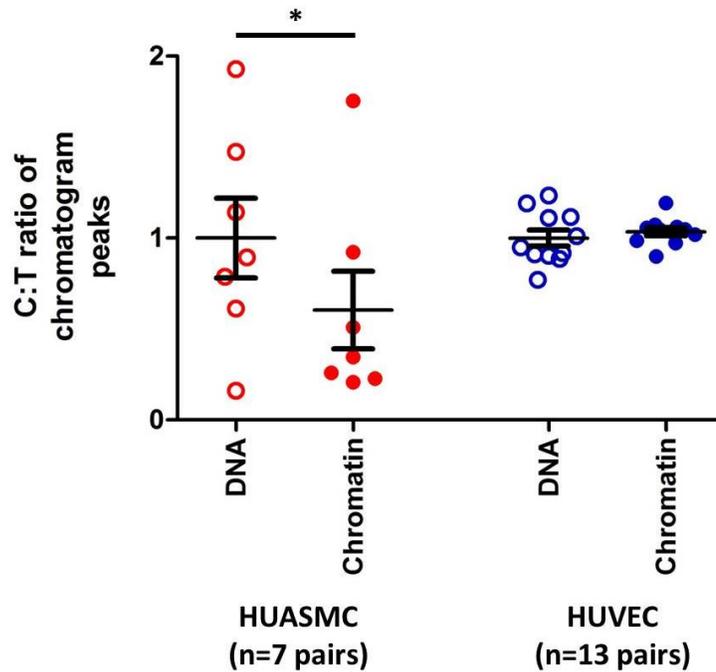


Figure 76: Formaldehyde-assisted isolation of regulatory elements assays at rs13096477 Heterozygotes at rs13096477 were used to obtain chromatin-associated DNA, and paired with the respective genomic DNA.
* $P < 0.05$, Wilcoxon signed-rank test.

4.3.3. Electrophoretic mobility shift assays (EMSA)

In order to utilize electrophoretic mobility shift assays in investigating the potential SNPs of interest, a decision was made to prioritise the initial list of 93 SNPs down to 10 SNPs. The decision-making rationale included the position of the SNPs relative to start of transcription, predictions of transcription factor binding, ENCODE and RegulomeDB predictions (Figure 77), of which the limitations have already been discussed in (Section 3.17).

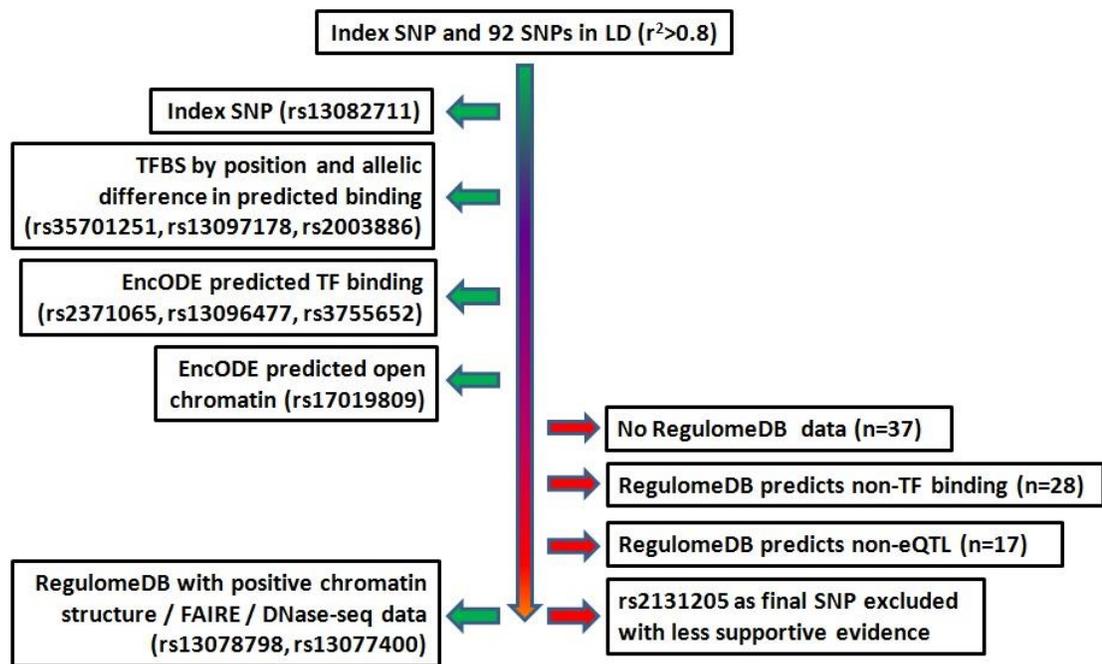


Figure 77: Rationale in selecting 10 SNPs (out of 93) for further investigation with EMSA. Arrows to left indicate SNPs selected, and arrows to the right indicate SNPs with lower priority.

With the 10 SNPs to be investigated by EMSA, the scanning panel shows that the electrophoretic mobility of labelled double-stranded oligonucleotides representing both the alleles of rs13096477, the minor alleles of rs2371065 and rs13077400, were retarded in the presence of 10 ng HUASMC nuclear extracts (Figure 78). The doublet of the signal shift for the rs2371065 probe is consistent throughout all EMSA studies, and even the DNA-pulldown assay shown later (Figure 85). It should be highlighted that the GWAS blood pressure-associated index SNP rs13082711 did not demonstrate DNA-nuclear protein interactions *in vitro*.

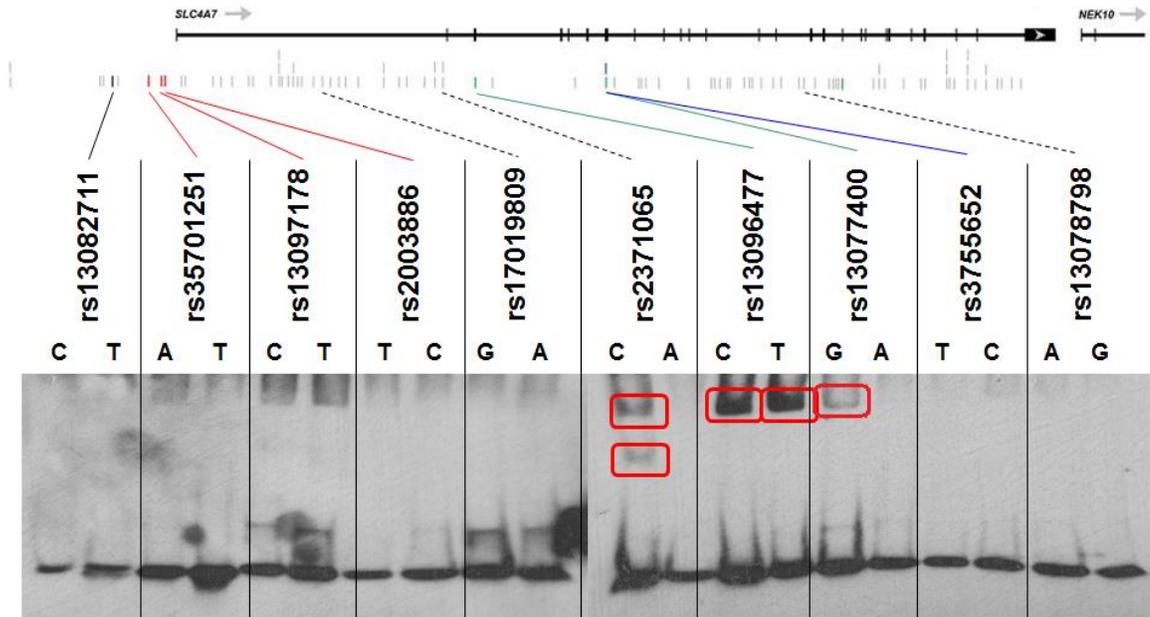


Figure 78: EMSA with labelled-probes corresponding to both alleles of 10 SNPs
 For each SNP, the minor allele is listed first; solid black line indicates GWAS-lead SNP, red lines indicate predicted TFBS, broken black lines indicate intronic, green lines indicate synonymous exonic SNPs, blue lines indicate non-synonymous SNP. All lanes had 10 ng HUASMC nuclear extracts added to 10 pmol of labelled double-stranded oligonucleotide probe. Unbound probes as visualized at bottom of autoradiograph. Red boxes denote observed electrophoretic mobility shifts, indicating protein-oligonucleotide interaction. Representative of n=2 sets.

To test for consistency of the potential cell-type dependent effect of DNA-nuclear protein interaction as seen in the FAIRE studies above (Figure 76), further EMSA studies were conducted comparing HAUSMC and HUVEC nuclear extracts with regards to the rs2371065, rs13096477 and rs13077400 (Figure 79).

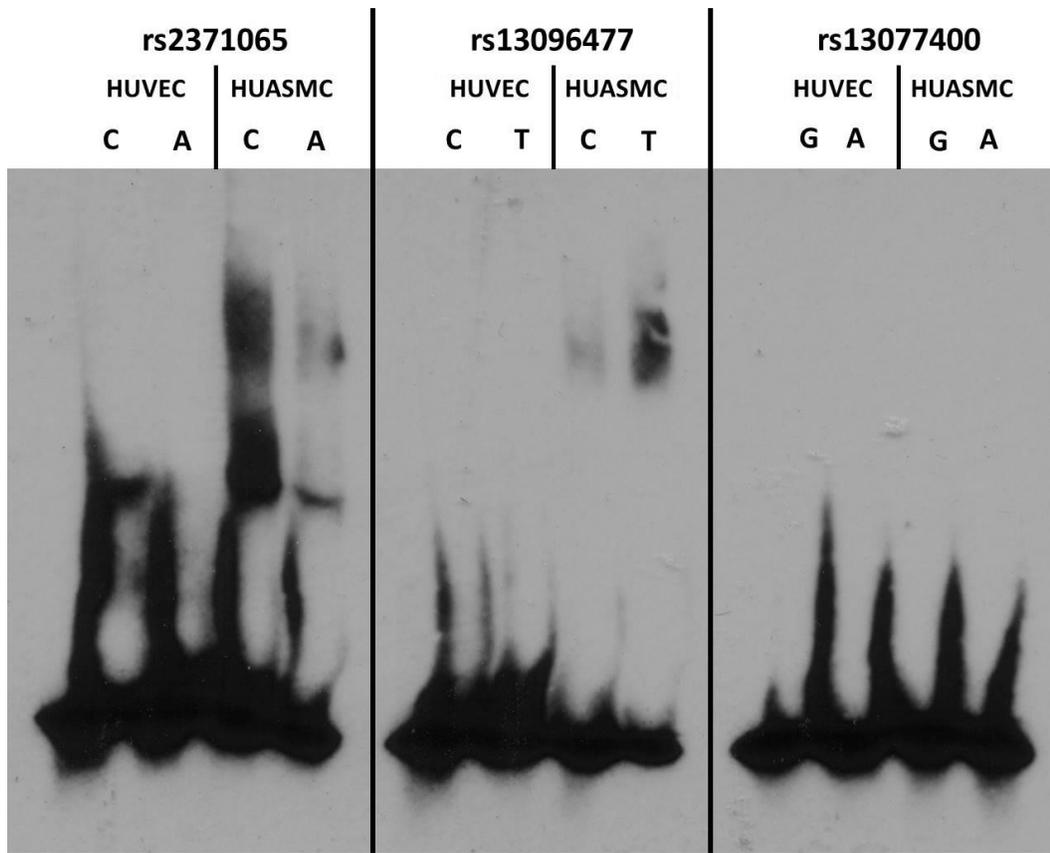


Figure 79: EMSA comparing HUVEC and HUASMC nuclear extracts for three SNPs. For each SNP, the minor (risk) allele is listed first. All lanes had 10 ng HUVEC or HUASMC nuclear extracts added to 10 pmol of labelled double-stranded oligonucleotide probe. Unbound probes as visualized at bottom of autoradiograph.

With all three SNPs, HUVEC nuclear extracts did not retard the mobility of the biotin-labelled oligonucleotide probes. HUASMC nuclear extracts continue to indicate *in vitro* DNA-nuclear protein interaction consistent with previous FAIRE results (Figure 76).

The apparent allele-dependence of the electrophoretic mobility shifts was further assessed by a dose-dependent assay for the sequences surrounding rs2371065, rs13096477 and rs13077400 using HUASMC nuclear extracts (Figure 80). The shifted signals for rs2371065 and rs13096477 continue to be persistently strong, whereas the signals for rs13077400 were often faint. These sets of experiments confirm an allelic-dependent *in vitro* interaction with HUASMC nuclear proteins.

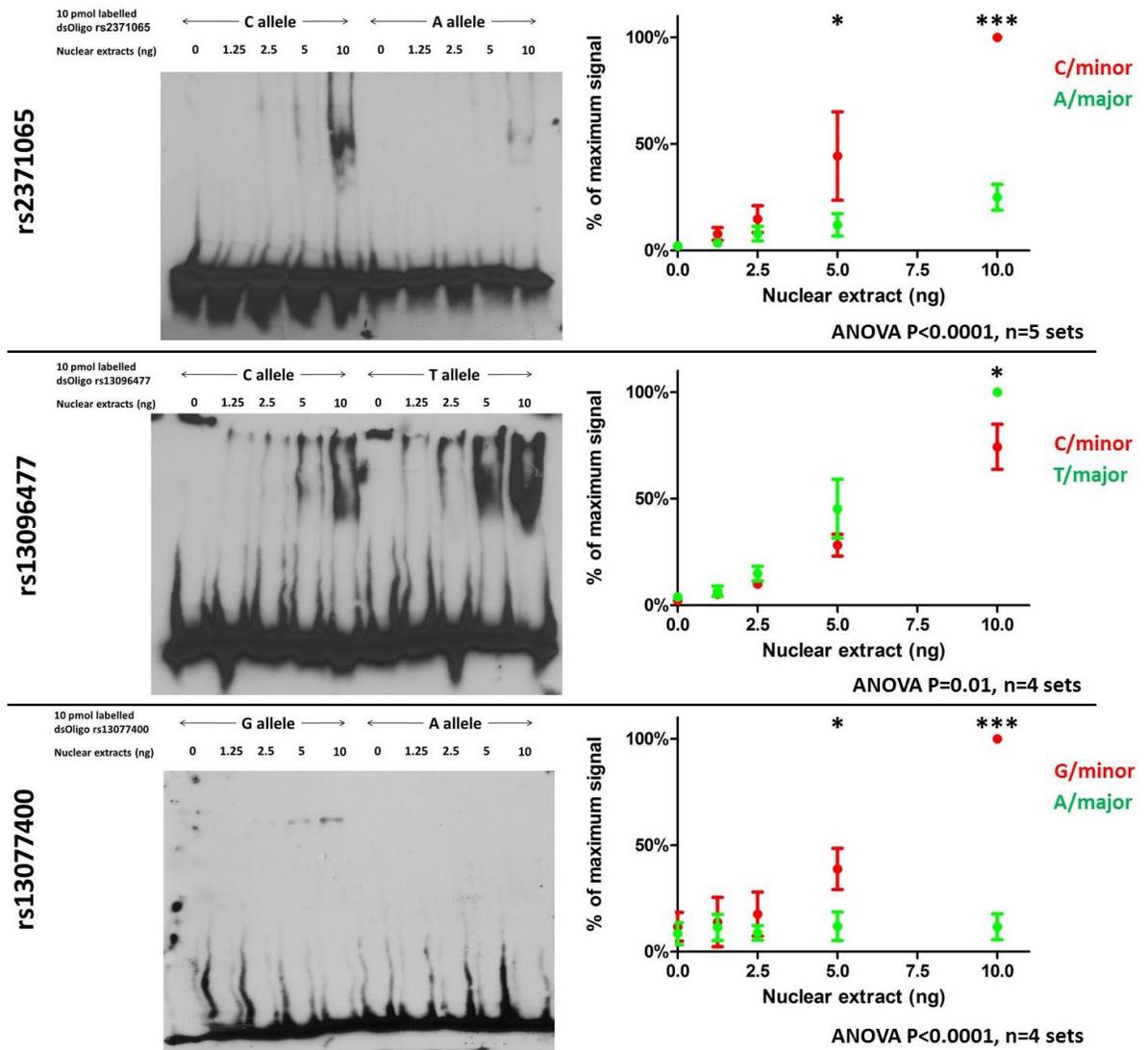


Figure 80: EMSA studies at rs2371065, rs13096477 and rs13077400

Dose-response of the double-stranded oligonucleotide probes for minor (first 5 lanes on gel, red on chart) or major (last 5 lanes on gel, green on graph) alleles at rs2371065, rs13096477 and rs13077400, shifting to increasing quantities of HUASMC nuclear extract.

ANOVA denotes two-ways repeated measures ANOVA for interaction.

* $P < 0.05$, *** $P < 0.001$ for comparison between alleles at each concentration of nuclear extract after Bonferroni correction for multiple comparisons.

Initial attempts at to identify the interacting nuclear protein was using the C allele at rs2371065, due to strong shifted signal, and the large difference when compared against the T allele. To narrow down the list of potential nuclear proteins that may interact with the rs2371065 double-stranded oligonucleotide, databases from Match, Patch and Wilmer, together with data from the ENCODE database was used. Those that are predicted to interact with the sequence that overlaps

with the rs2371065 SNP are listed in Table 46. Unlabelled double-stranded oligonucleotides were generated as competitor probes.

	Prediction Source	Sequence reference	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
rs2371065			T	C	C	C	A	C	C	T	T	C	C/A	T	A	A	T	T	T	T	G	T	T
PHOX2A	Wilmer	Wilmer	T	C	C	C	A	C	C	T	T	C	C	T	A	A	T	T	T	T	G	T	T
PLAGL1	Wilmer	Wilmer	T	C	C	C	A	C	C	T	T	C	C	T	A	A	T	T	T	T	G	T	T
c-Fos	Patch	JASPAR	T	C	C	C	A	C	C	T	G	A	C	T	C	A	C	T	T	T	G	T	T
EBF1 (pos 1)	ENCODE	Wilmer	T	C	C	C	A	C	C	T	T	C	C	C	G	T	T	T	T	T	G	T	T
EBF1 (pos 2)	ENCODE	Wilmer	T	C	C	C	A	C	C	T	T	C	C	C	C	G	T	T	T	T	G	T	T
Elk1	Match	JASPAR	T	C	C	C	A	C	C	T	T	C	C	G	G	C	T	C	T	T	G	T	T
TFAP2C	Wilmer	Wilmer	T	C	C	C	A	C	C	T	T	C	C	A	A	A	T	T	T	T	G	T	T
TRMT1	Wilmer	Wilmer	T	C	C	C	A	C	C	T	T	C	A	T	T	T	T	T	T	T	G	T	T

Table 46: Oligonucleotide sequences of EMSA competitor probes

The sense strand is shown, and a reverse complimentary strand is used to construct double-stranded oligonucleotides. The reference SLC4A7 sequence is shown on the first row, with the rs2371065 SNP at position 11. The competitor probe representing PHOX2A and PLAGL1 is the same as the C allele probe due to the shared sequence. Coloured boxes are proposed consensus sequences; green, no change from refernce sequence; red, different from consensus sequence; large font, proposed key sequences. See Section 3.4 for description of prediction sources and sequence references.

Using competitor assays, the initially observed shift is reduced and almost amilorated by the competitor probes corresponding to PHOX2A, PLAGL1, TFAP2C and TRMT1 (Figure 81). Of note, the competitor corresponding to the SLC4A7 C allele sequence reduced the shifted signal more than that with the A allele.

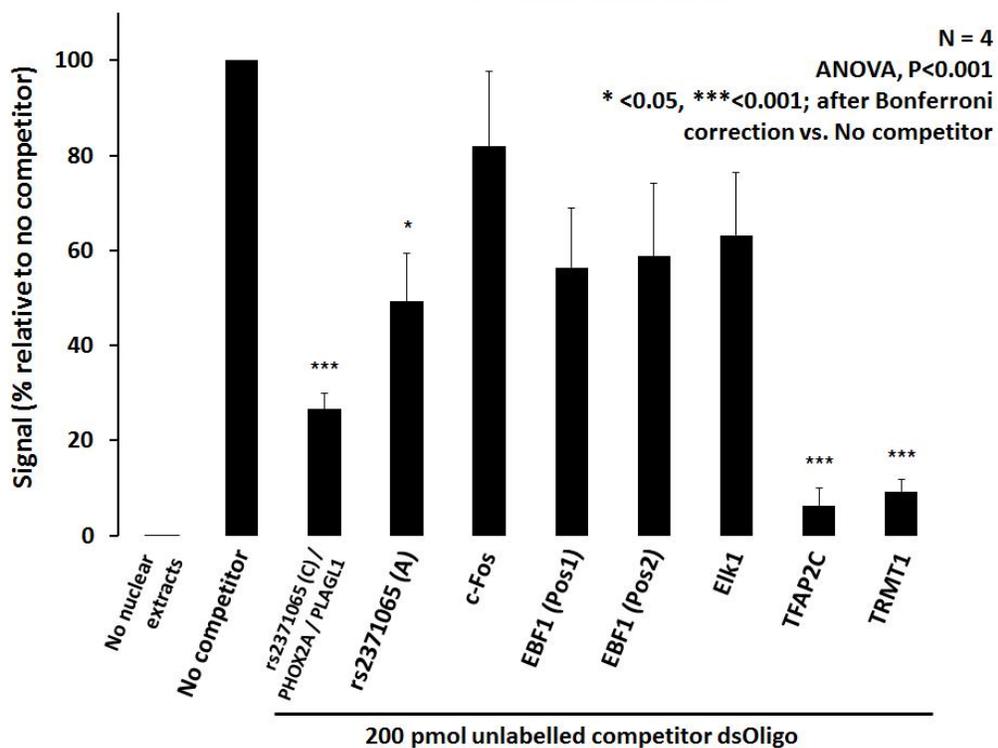
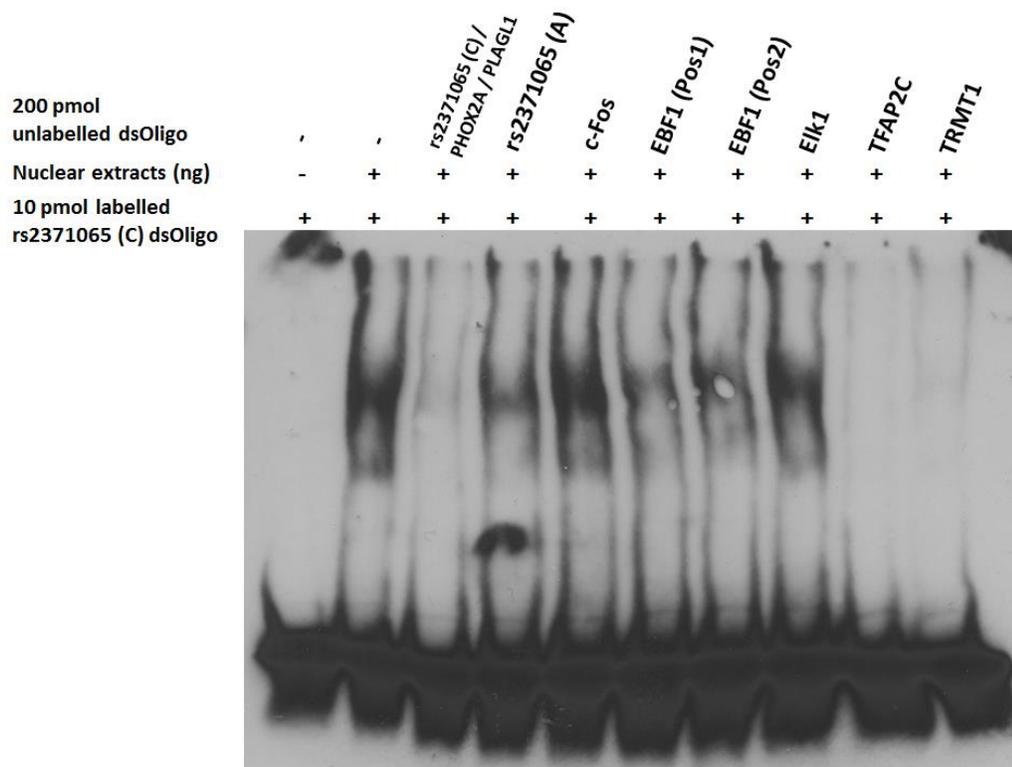


Figure 81: Unlabeled double-stranded oligonucleotide competitor EMSA studies
Effect of 200 pmol unlabelled double-stranded oligonucleotide competitor probes on the previously seen protein-labelled oligonucleotide (rs2371065 (C) dsOligo) shifted signal.
ANOVA denotes one-way repeated measures ANOVA for interaction.
* $P < 0.05$, *** $P < 0.001$ vs. no competitor, after Bonferroni correction for multiple comparisons.

Antibodies targeted against the most promising nuclear proteins (PHOX2A, PLAGL1 and TFAP2C) were used in supershift assays. TRMT1 was not assessed as it functions primarily to bind tRNA (transfer RNA). In two separate studies, there were no supershifts observed with two different concentrations of anti-PHOX2A, anti-PLAGL1 and anti-TFAP2C antibodies (Figure 82).

Antibody (ng)	-	-	1 5		1 5		1 5		1	-
			PHOX2A		PLAGL1		TFAP2C		IgG	
0.125% gelatin, 0.125% NaN ₃	-	-	+	+	+	+	+	+	+	+
Nuclear extracts (10 ng)	-	+	+	+	+	+	+	+	+	+
Labelled rs2371065 dsOligo (10 pmol)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)	(C)

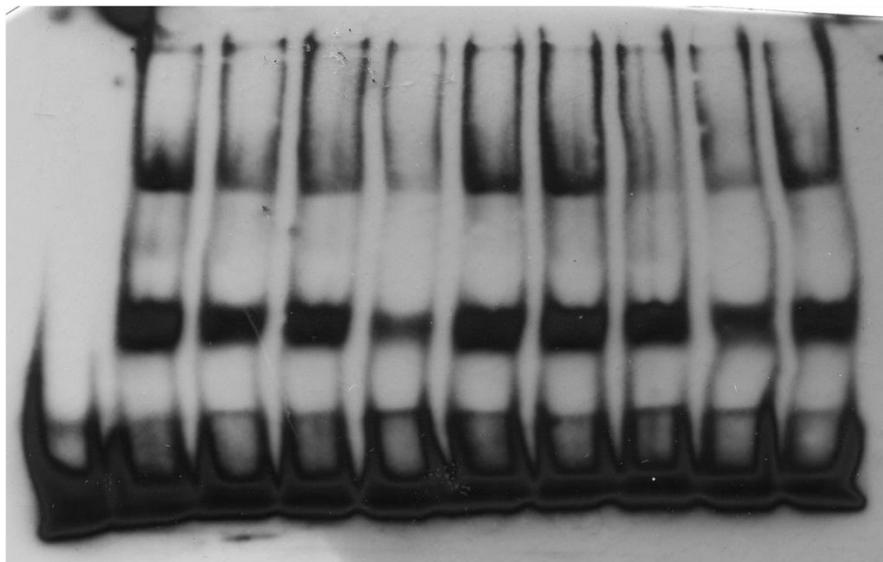
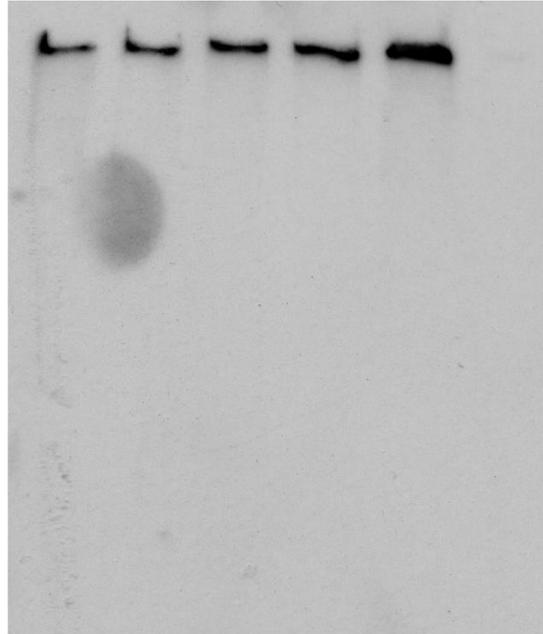


Figure 82: Supershift assays at rs2371065 targeted against PHOX2A, PLAGL1 and TFAP2C. Controls included normal IgG and gelatin / NaN₃ (sodium azide) as they can potentially interfere with the protein-oligonucleotide interaction. Representative of n=2. No consistent pattern observed.

As supershift assays are highly dependent on antibody-protein interactions that may be difficult to achieve *in vitro*, a Western-EMSA (see Section 2.14.2) was conducted. With a native polyacrylamide gel, the nuclear proteins normally migrate slowly. However, if there were protein-oligonucleotide binding, this would increase the charge-to-mass ratio and accelerate the mobility, shifting the detectable protein downwards. This was not observed (Figure 83).

EMSA Buffer	-	+	+	+	+	+
Unlabelled rs2371065 C allele dsOligo (fmol)	-	-	200	500	-	500
Unlabelled non-specific competitor dsOligo (fmol)	-	-	-	-	500	-
HEK293 Nuclear Extract (20ng)	+	+	+	+	+	-



Western-EMSA
IB: Anti-AP2γ

Figure 83: Western-EMSA (WEMSA) at rs2371065

Western-EMSA (WEMSA) with HEK293 nuclear extracts and double-stranded oligonucleotides corresponding to the sequence surrounding rs2371065. No shift of AP2γ (the gene product of TFAP2C) was detected. Representative of n=2. HEK293 nuclear extracts were chosen due to the higher expression levels.

No PHOX2A and PLAGL1 signal was detected when replicating this study with the respective antibodies. In addition, a CHIP assay using 2ng anti-AP2γ antibodies did not identify the sequence around rs2371065 as interacting with AP2γ.

These negative results were followed up with end-point RT-PCR to determine if there was at least mRNA expression of *PHOX2A*, *PLAGL1* or *TFAP2C* in tissues of interest. Out of the three, only *PLAGL1* mRNA was detected by RT-PCR in HUASMCs (Figure 84).

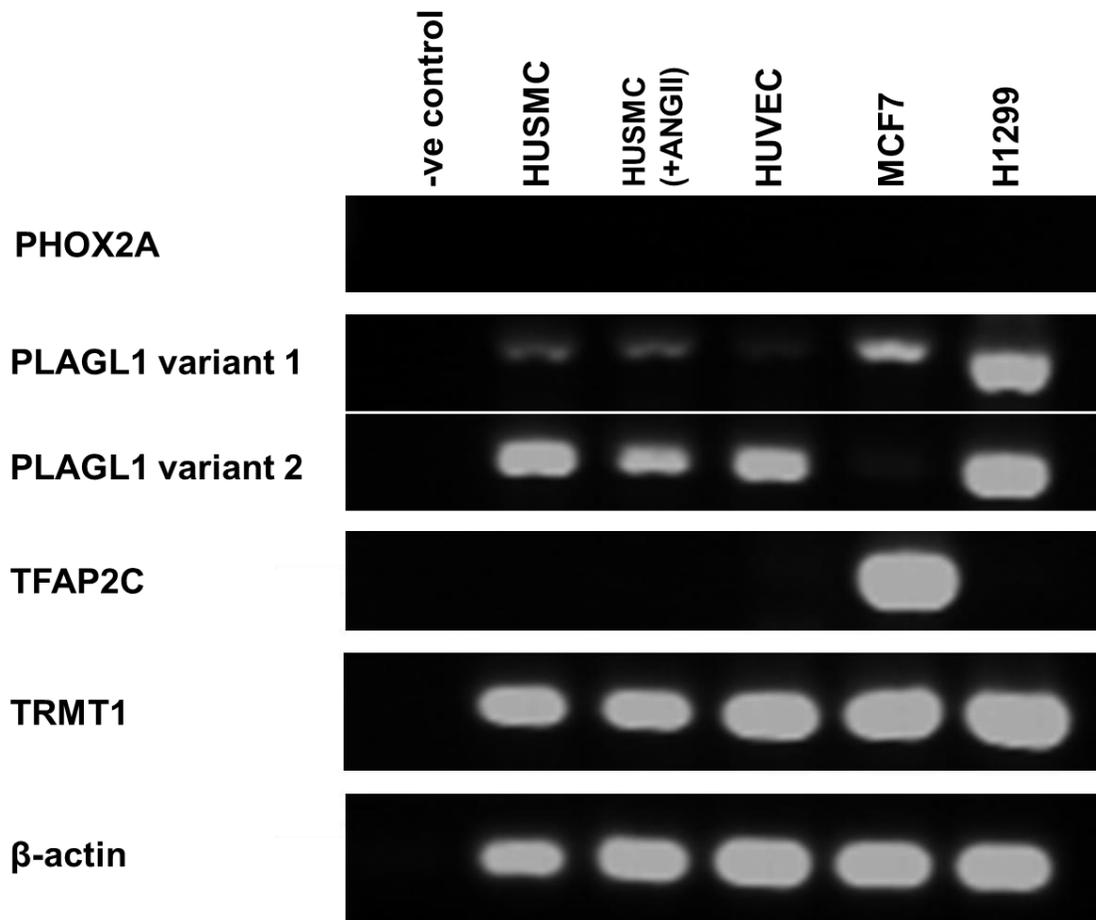


Figure 84: End-point RT-PCR for PHOX2A, PLAGL1, TFAP2C and TRMT1
End-point RT-PCR (10ng input RNA) of a panel of cell lines for PHOX2A, PLAGL1, TFAP2C and TRMT1. β-actin shown as control. MCF7 and H1299 cells were used as possible positive controls.

As the attempts at identifying the protein responsible for the EMSA shift with rs2371065 probes via supershift assays were unproductive, a DNA pull-down assay was conducted (see Section 4.3.4).

Due to the limited success of the supershift (Figure 82) and DNA pull-down assay (Section 4.3.4), this result of an allele-preferential nuclear protein binding with the sequences surrounding rs13096477 and rs13077400 was not investigated further.

4.3.4. DNA pull-down assay

A DNA pull-down assay was conducted with biotin-labelled double stranded oligonucleotides targeted against the 21 bp sequence centered on rs2371065 (with the C allele in position 11; forward strand 5'-TCCCACCTTCCCTAATTTTGT-3'). It identified two specific bands of proteins, neither present in the final (2nd) wash and the negative (no oligo, but with agarose beads) control (Figure 85). The presence of two signals was reassuringly consistent with previous EMSA studies. These bands were excised and sent for N-terminal protein sequencing commercially by Alta Bioscience (Cambridge, UK).

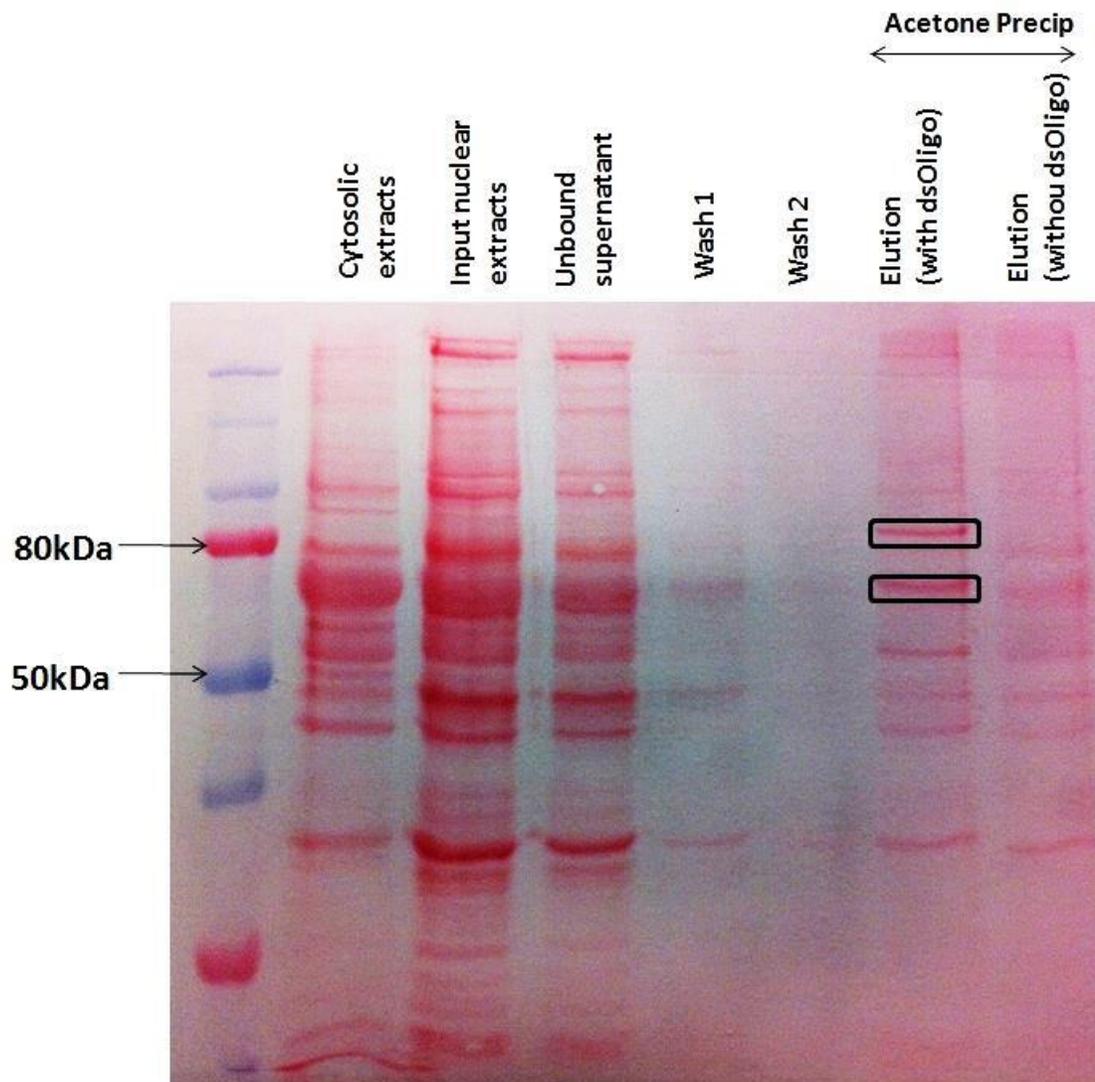
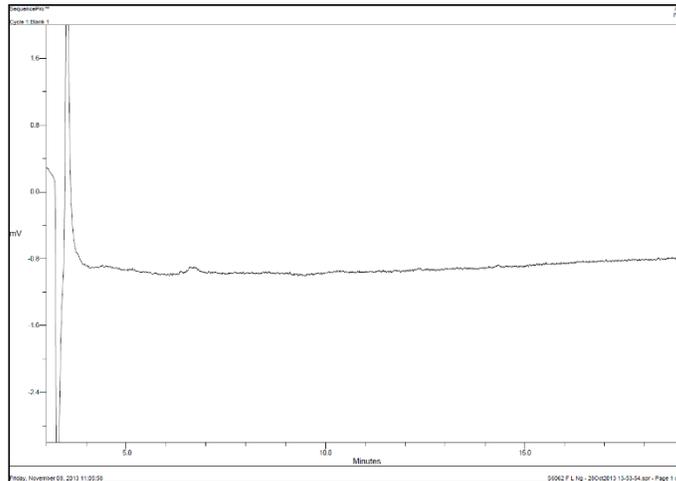


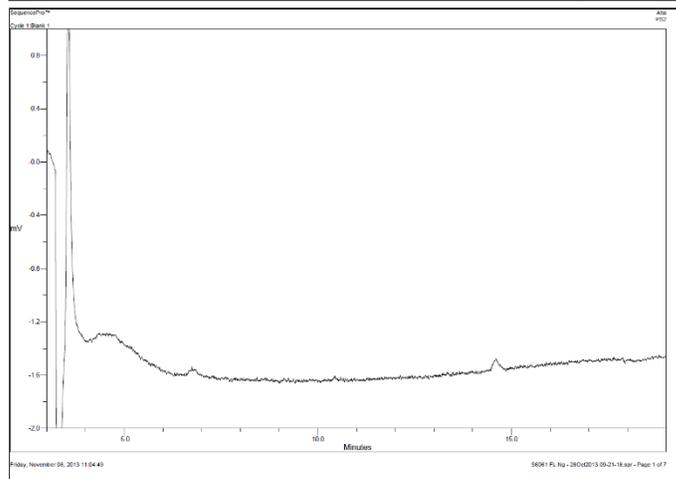
Figure 85: DNA-pull down with double-stranded oligonucleotide centering on rs2371065
Digital image of a Ponceau S-stained PVDF membrane demonstrating the different banding patterns. The boxes indicate bands that is neither present in the final wash nor the negative control. It was given the nominal names 80 kDa and 60 kDa pending sequencing results.

The commercial service was unable to return any sequenced amino acids despite a successful positive control run (Figure 86).

**~80kDa
band**



**~60kDa
band**



**Standard
protein**

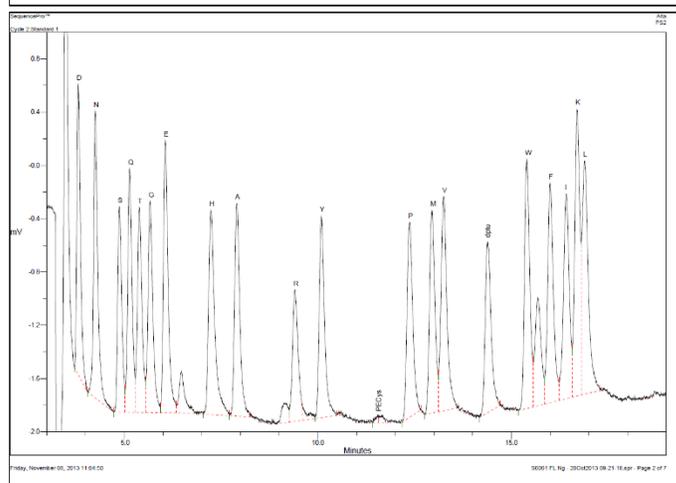


Figure 86: N-terminal amino acid sequencing chromatographs of DNA-pull down products N-terminal amino acid sequencing chromatographs of the two excised bands (top and middle), and the proprietary positive control sequencing standard profile (bottom).

4.4. Impact of non-synonymous SNP rs3755652

4.4.1. Co-immunoprecipitation and calcineurin A (CnA)

In studies using A10 cells transfected with *SLC4A7* variants (326E, 326K or splice variant without Cassette II) or vector control, there was no co-immunoprecipitation of NBCn1 with calcineurin A detected (Figure 87). The presence of calcineurin A signal in lanes 1 to 4 acts as a positive control for the system to detect calcineurin A. The absence of NBCn1 signal in lanes 5 to 8 act as a negative control to confirm immunoprecipitation of NBCn1 in lanes 9 to 12. The absence of calcineurin A signal in lanes 9 to 12 indicates that despite the immunoprecipitation of NBCn1, it does not co-immunoprecipitate with calcineurin A in this experimental set up.

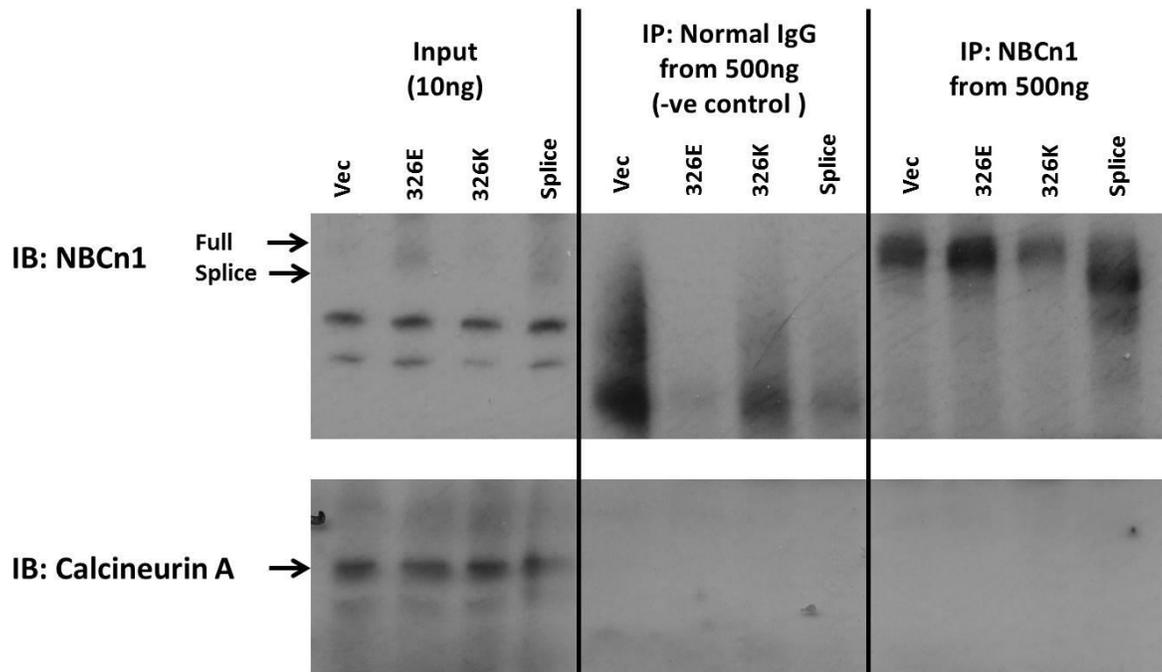


Figure 87: Calcineurin A does not co-immunoprecipitate with NBCn1 in transfected A10 cells
Input total protein from A10 cells transfected with 1 ng plasmid (vector control, 326E, 326K or Spl) per 10 cm² cell culture surface for 24 hours (lanes 1 to 4). Eluates from immunoprecipitating 500 µg of protein using 2 ng of the negative control, normal rabbit IgG (lanes 5 to 8) or using 2 ng of rabbit anti-NBCn1 polyclonal IgG (lanes 9 to 12).

Studies repeated with HUASMC primary cell cultures also did not show calcineurin A to co-immunoprecipitate with NBCn1.

4.4.2. Immunocytochemistry

Immunocytochemistry identified NBCn1 present in both the membrane and cytosol, in HUVECs (Figure 88) and HUASMCs (Figure 89). Similarly, transfection with A10 cells were as initially shown in Section 2.19 (Figure 45). Negative controls IgG (normal mouse IgG, Santa Cruz sc-2025; normal rabbit IgG, Santa Cruz sc-2027) did not produce any detectable signal with the same exposure. It is apparent that in all three cell models, with the different variants, the is NBCn1 signal present in both the plasma membrane, as well as intracellularly. The signal ratio calculated from the mean NBCn1 signals within the membrane (corresponding VE-Cadherin peaks) and cytoplasmic areas it encompasses were similar in the two genotypes.

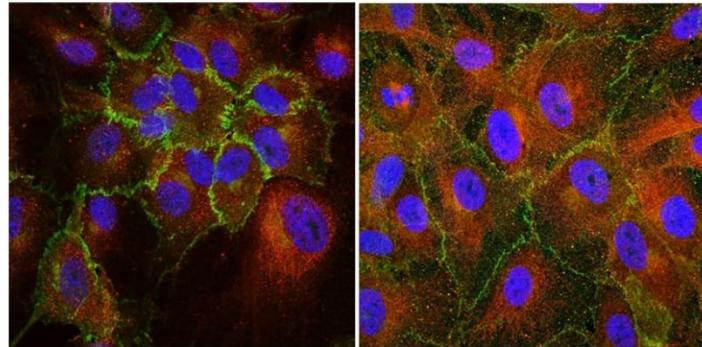
As intracellular organelle markers were not used, it is difficult to localise the intracellular signal, but the perinuclear pattern might suggest presence within the Golgi apparatus. However, the quantification of the relative presence in membrane or cytosol is fraught with difficulties as explored further in Discussion Section 5.4.2. With this, subcellular fractionation (Section 4.4.3) was used.

rs3755652
 Position 326 AA
 Minor/major
 GWAS phenotype

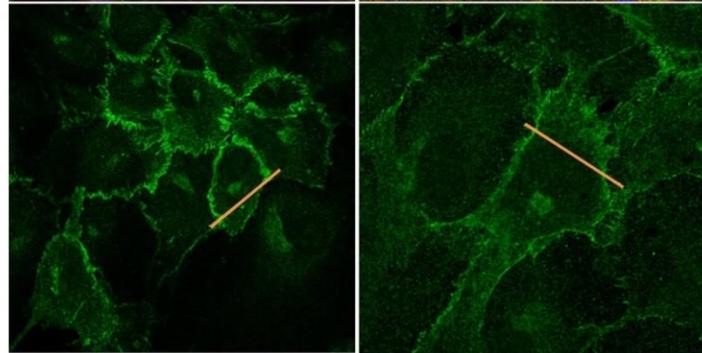
G/G
 Lys/Lys
 Minor
 Risk

A/A
 Glu/Glu
 Major
 Protective

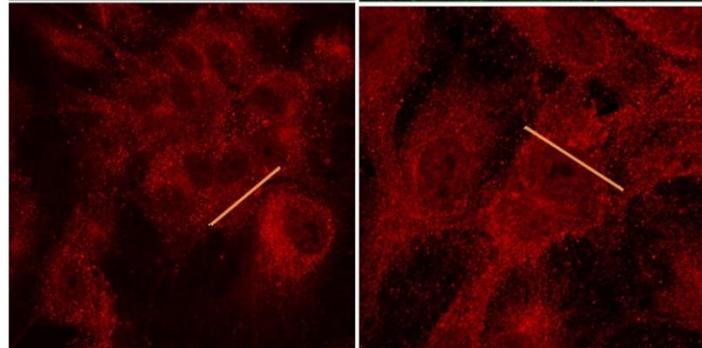
DAPI
 VE-cadherin
 NBCn1



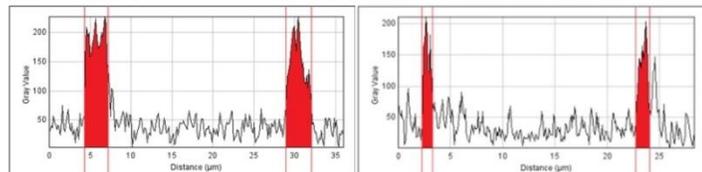
VE-cadherin



NBCn1



VE-cadherin signal



NBCn1 signal

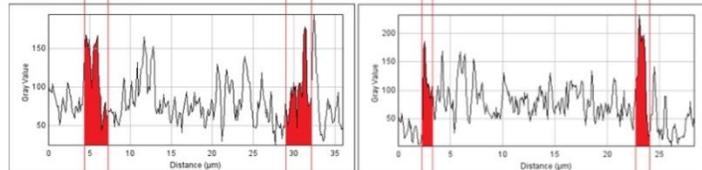


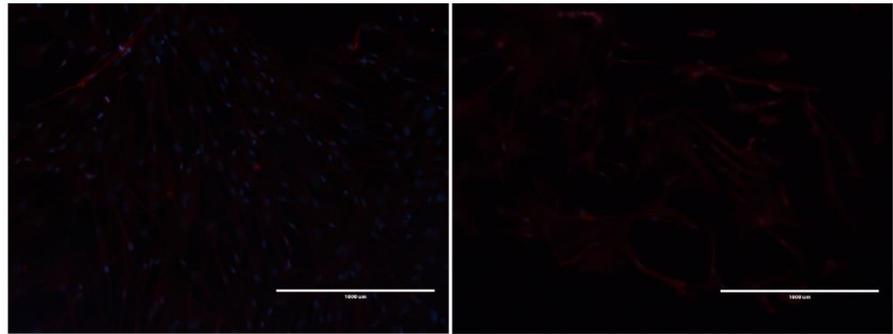
Figure 88: Confocal microscopy staining for VE-cadherin and NBCn1 in HUVECs
 Confocal microscopy (63x objective) of HUVECs of rare and common homozygotes at the nsSNP rs3755652. Histograms show profile plots of the respective signals across the yellow line (ImageJ).

rs3755652
 Position 326 AA
 Minor/major
 GWAS phenotype

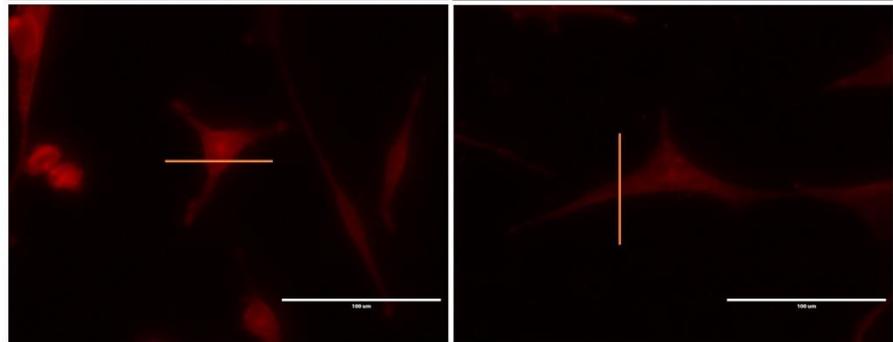
G/G
 Lys/Lys
 Minor
 Risk

A/A
 Glu/Glu
 Major
 Protective

DAPI
 NBCn1
 (4x objective)



NBCn1
 (40x objective)



NBCn1 signal

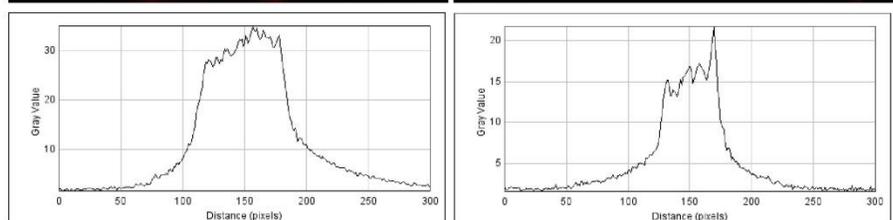


Figure 89: Inverted fluorescent microscopy images staining for NBCn1 in HUASMCs
 Representative images of HUASMCs of rare and common homozygotes at the nsSNP rs3755652. Cells grown to high confluence for 4x objective images, solitary cells for 40x objective images. Scale bars represent 1000 μm (4x objective) and 100 μm (40x objective). Histograms show profile plots of the respective signals across the yellow line of 40x objective images (ImageJ).

4.4.3. Subcellular fractionation and localization of NBCn1

With the differential centrifugation method of subcellular fractionation (Section 2.8.1), one fraction was discarded, and three others were obtained. The first centrifugation pellet (at 1,000g) was discarded as it would have contained a mixture of nuclei as well as whole unlysed cells, making interpretation difficult. The second centrifugation pellet (at 10,000g) would contain a mixture of cellular organelles, the third centrifugation pellet (at 100,000g) would predominantly contain the plasma membrane-bound proteins and the resultant supernatant reflects the content of the free cytosolic compartment.

In both HUASMCs and HUVECs, there was a greater amount of NBCn1 in the “membrane” (100,000 g) fraction in the cells encoding NBCn1 proteins with the risk Lys/Glu variants (G/A at rs3755652) compared to the protective Glu/Glu variant (A/A at rs3755652) (Figure 90). Notably, there were no differences between the genotypes in the “organelle” (10,000 g) fraction. There were no detectable NBCn1, N- or VE-cadherin signal in the cytosolic (supernatant) fraction in all the samples.

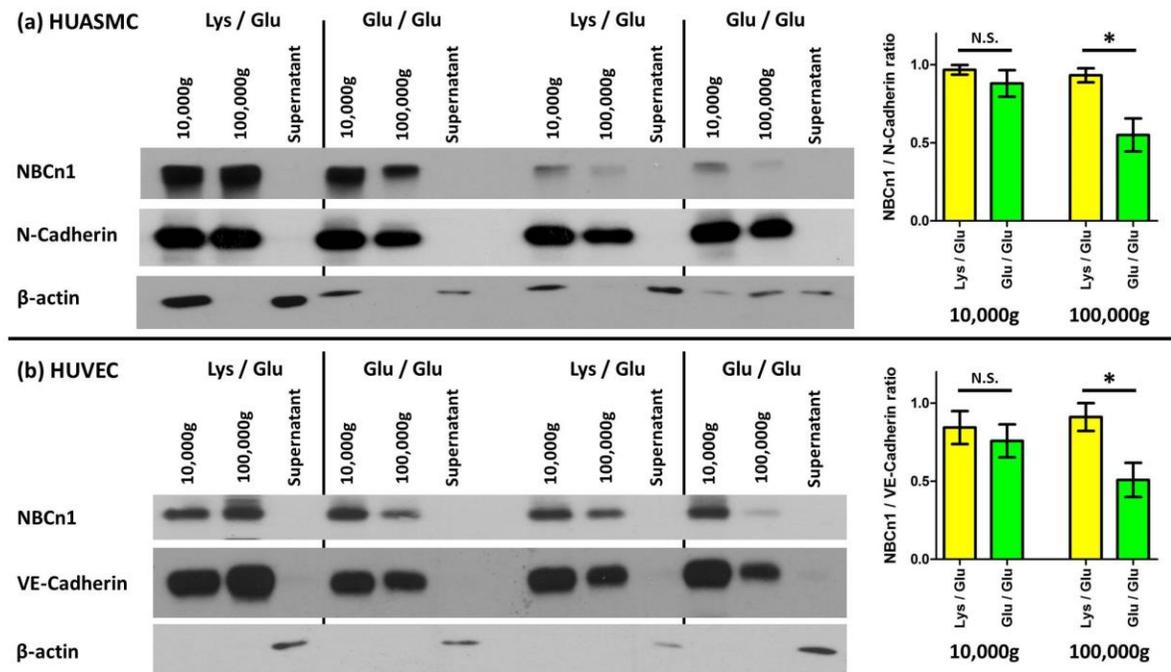


Figure 90: Subcellular fractionation samples from HUASMCs and HUVECs. Samples paired for E326K variants in each immunoblot. 2 ng protein of each cellular fraction, grouped by cell sample, was loaded for each lane. $n = 9$ each group for HUASMCs, 8 each group for HUVECs. * $P < 0.05$ by Mann-Whitney U-test

4.4.4. Total cellular NBCn1 in transfected A10 cells

Transfecting equimolar quantities of the two 326E and 326K *SLC4A7* plasmids for 48 hours resulted in the same total cellular protein NBCn1 expression as detected by protein immunoblots (Figure 91).

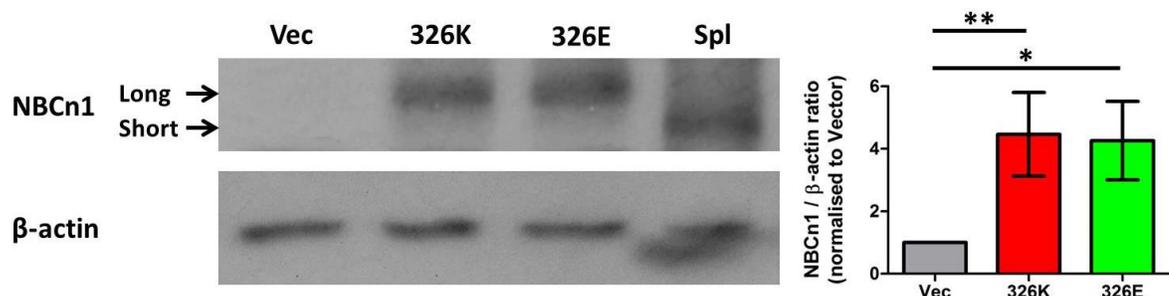


Figure 91: NBCn1 expression in transfected A10 cells
 HEK293 and A10 cells were transfected with 1 ng plasmid per 10 cm² cell culture surface (326E, 326K and Spl or with vector control) for 48 hours.
 n = 8 sets, one-way ANOVA P=0.003. *P<0.05, **P<0.01 after correcting for multiple comparisons

4.5. Inducibility

Treating HUASMCs with 10⁻⁹ M angiotensin II in a serum-starved culture system over 3 days increased total NBCn1 expression as compared to vehicle control (Figure 92a). Treatment with a dose-range of angiotensin II over 12 hours changed the mRNA/cDNA allelic imbalance of the two polymorphic alleles, where there is a return from an imbalance from favouring the minor (C) allele to a balance between the two alleles with increasing angiotensin II exposure (Figure 92b).

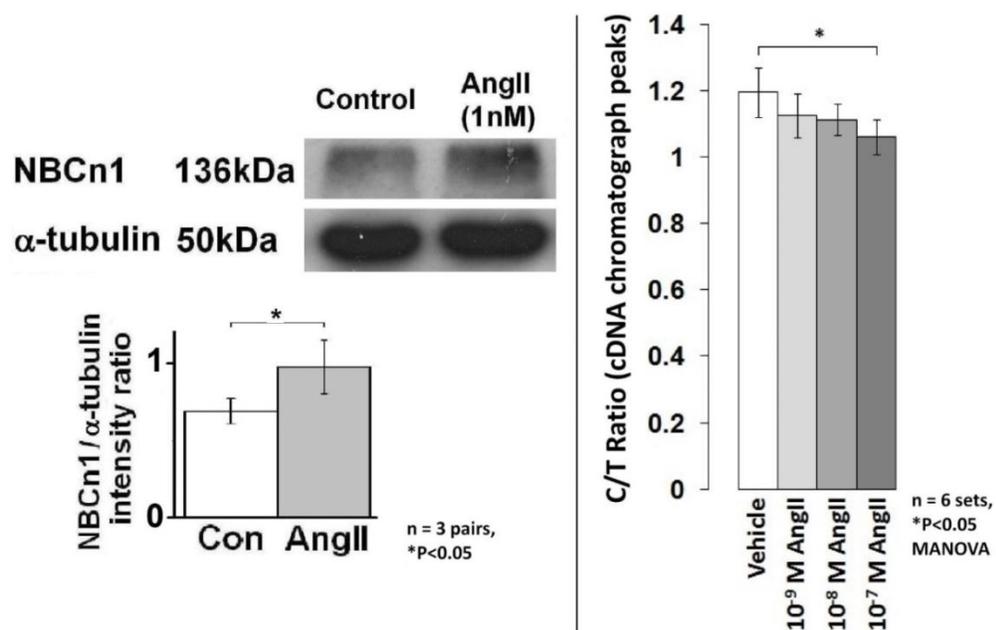


Figure 92: Inducibility of NBCn1 by angiotensin II, and the relevance of genotype
 Samples were exposed to varying concentrations of angiotensin II or vehicle in an unsupplemented culture system. (a) Protein immunoblots indicate an induction of NBCn1 by angiotensin II, (b) allelic imbalance analyses indicate the induction may be genotype-dependent.

These studies could not be replicated in HUVECs particularly as they demonstrated an apoptotic phenotype in unsupplemented media (Figure 93). This need for media supplementation would likely negate the effect of angiotensin II due to its short half-life in serum (Al-Merani *et al.*, 1978). Whilst some reports utilise “low serum” supplementation, it is still likely to impact on the availability of angiotensin II to act on the cells. Therefore, this was not considered further.

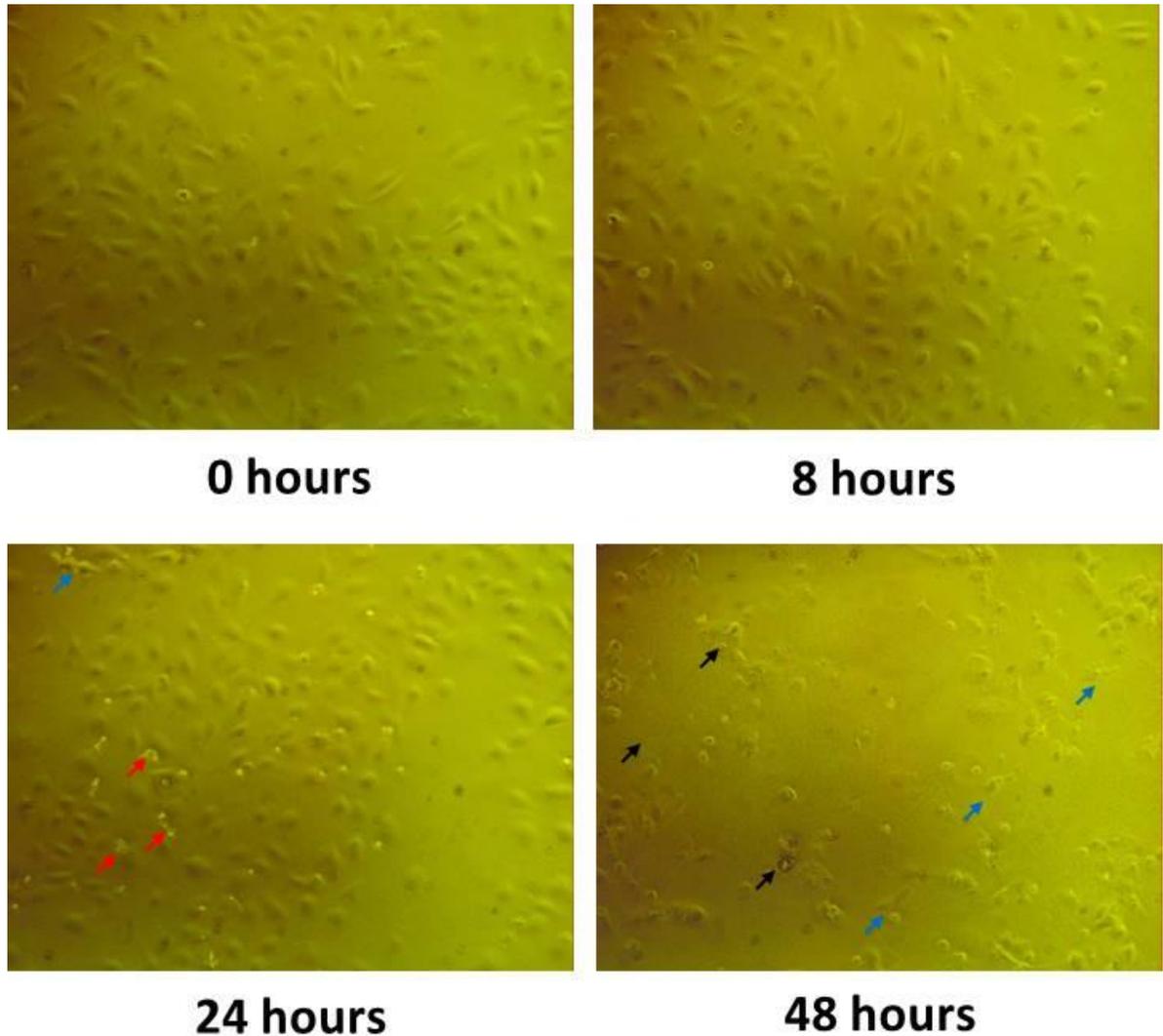


Figure 93: Digital light microscopy images (10x objective) of serum-starved HUVECs over 48 hours. Digital images taken from the same position over the time course. Marked reduction in cell numbers observed, as compared to paired cultures in serum-supplemented conditions. Red arrows denote detached cells; blue arrows denote cell shrinkage; black arrows denote clumps of dead cells. Only selected examples are highlighted, and more are present for each field.

4.6. *In vitro* cellular function

4.6.1. Collagen contractility gel

An initial trial of collagen contractility gel assays revealed the optimal cell density of 1.25×10^6 cells/ml gel. Following this, a time-course was assessed for untreated, serum-supplemented and angiotensin II-stimulated contraction. It showed that maximal contraction is achieved by 72 hours (

Figure 94). On top of the unstimulated contraction, addition of serum-supplementation produces a further contraction that is not added to by 10^{-8} M angiotensin II. This lack of additive effect is not unexpected, as previously mentioned, the half-life of angiotensin II is markedly reduced in the presence of serum (Al-Merani *et al.*, 1978). With this, all further studies are conducted over 72 hours, and without serum-supplementation.

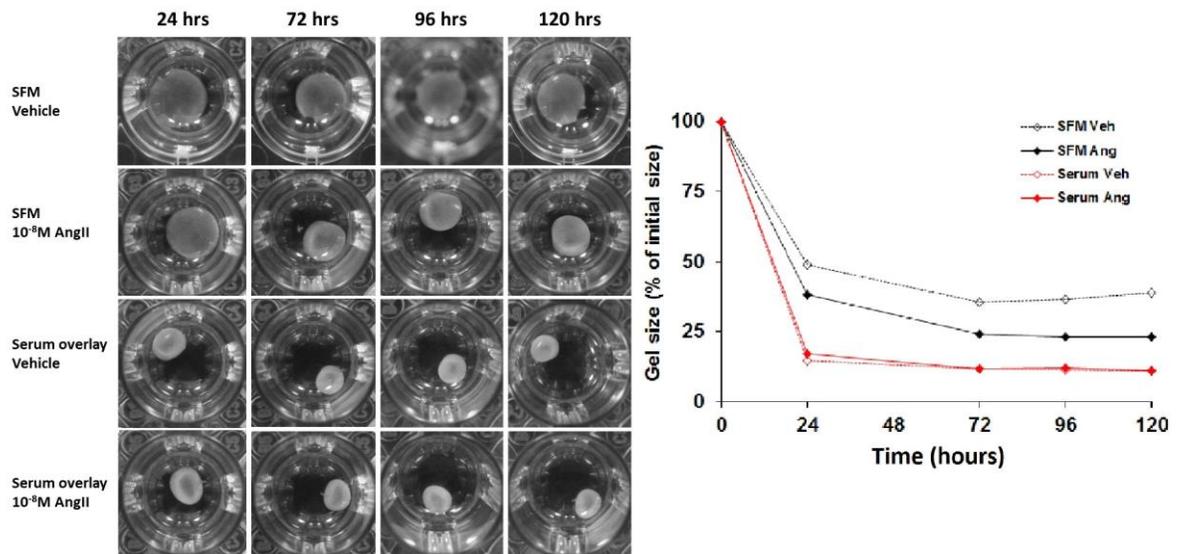


Figure 94: Time series for collagen contractility gels

Time series for collagen contractility gels (1.25×10^6 cells/ml gel) with various conditions.

“SFM” denotes serum-free DMEM media as overlay, “Serum” denotes 15% FBS-supplemented DMEM media as overlay.

Having established a contractile system, the replicability was assessed, showing good correlation between replicates (Figure 95a). For appropriate quantification, an imaging system had to be established. This involved initially assessing the difference between colour and black-and-white digital images. The independent measurements of the two image formats were well correlated (Figure 95b). Anecdotally, the black-and-white images showed clearer borders, as it is not complicated by the colour of the culture media. This opinion was also independently reflected by

Observer 2 (Dr AD Moore) and Observer 3 (Dr K Witkowska). With this, all further images were taken in black-and-white.

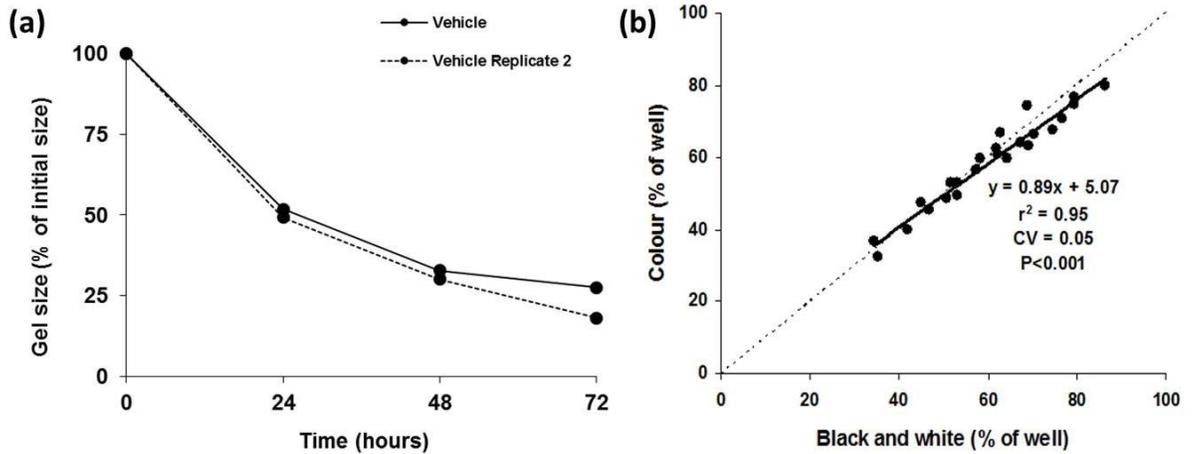


Figure 95: Collagen contractility gels intra-sample replicability and measurement optimisation (a) Replicates of collagen contractility gels demonstrating low variability. (b) Correlation between independent measurements of the same gel surface areas, either in colour or black-and-white digital images (n=24).

The variability of the measurements itself needed to be assessed. Intra-observer variability was good (Figure 96; $P < 0.001$, $r^2 > 0.9$, $CV < 0.05$), and inter-observer variability was reasonable and was again confirmed to be better with black-and-white images (Figure 97; $P < 0.001$, $r^2 > 0.7$, $CV < 0.15$). With this, future analysis would be conducted by three independent blinded observers with black-and-white digital images.

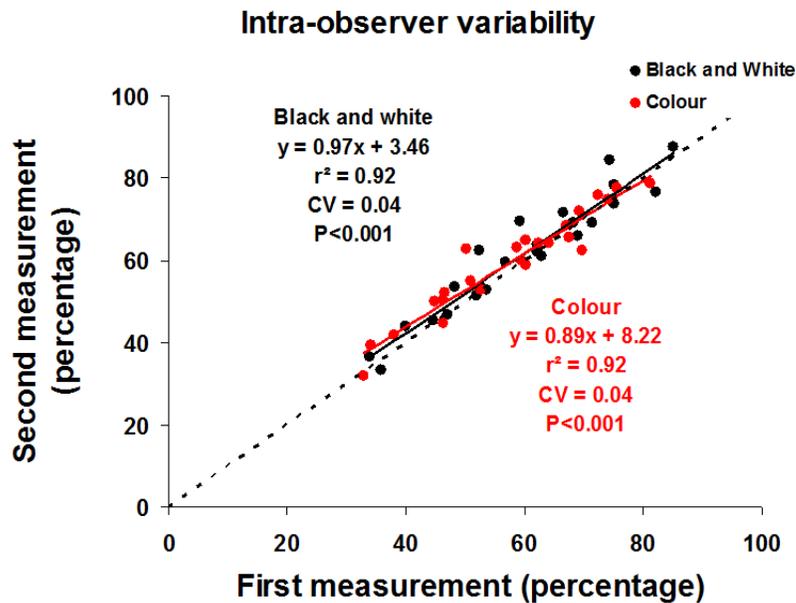


Figure 96: Intra-observer variability of collagen gel contractility assays measurements

Intra-observer variability between a first and second measurement of the same digital image, with three days separating the measurements (n=24).

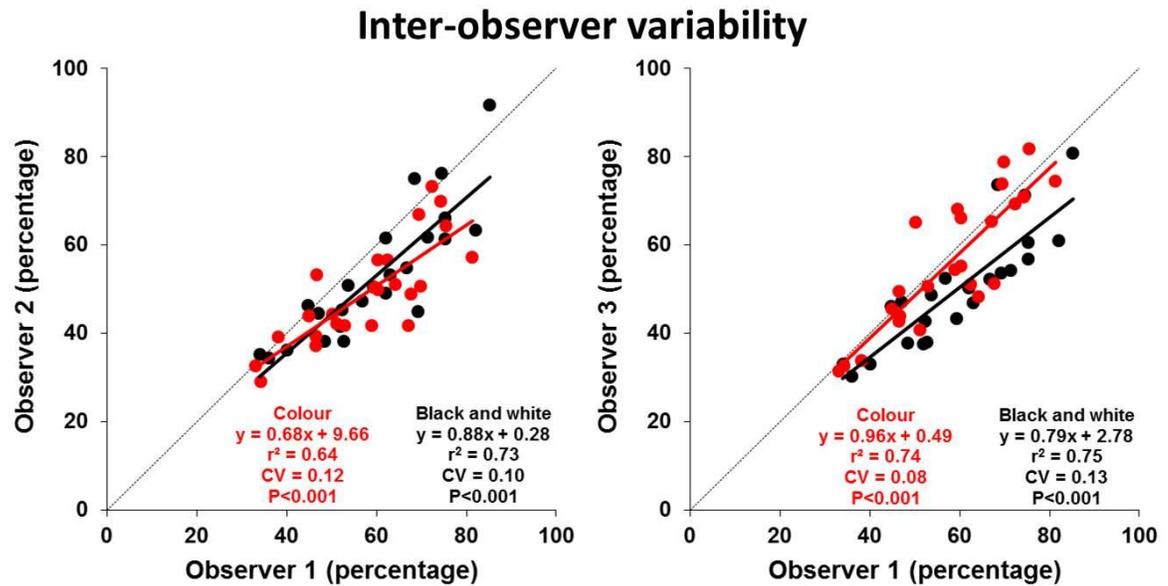


Figure 97: Inter-observer variability of collagen gel contractility assays measurements
Inter-observer variability between a three independent observers of the same digital image, with all observers blinded to the identity of the image (n=24). With gratitude to Dr AD Moore (Observer 2) and Dr K Witkowska (Observer 3).

However, after establishing the methodology, it became apparent that collagen contractility gels were not suitable for analyzing the impact of intracellular acidosis. The ammonium chloride pre-pulse method was not deemed appropriate due to the multiple washes that is necessary and would disturb the positioning of the collagen gel. With this, noradrenaline (as adrenergic receptor agonist), sodium acetate as weak base that results in intracellular acidosis and extracellular acidosis using a monovalent strong acid (HCl) were use, in comparison to vehicle control.

Unsurprisingly, 10^{-5} M noradrenaline resulted in a smaller collagen gel, reflecting vascular smooth muscle contraction, whereas 40 mM sodium acetate did not produce a clear shift in gel size as opposed to vehicle control (Figure 98). However, upon introducing an overlying culture media with an acidotic pH of 6.8, the structural integrity of the collagen gels notably changes with blurring of the edges (Figure 99). This is likely due to the alkaline pH that is necessary for collagen polymerization. This is also apparent in the methodology of extracting collagen from rat tails involving acidification of the tissue (Rajan *et al.*, 2006).

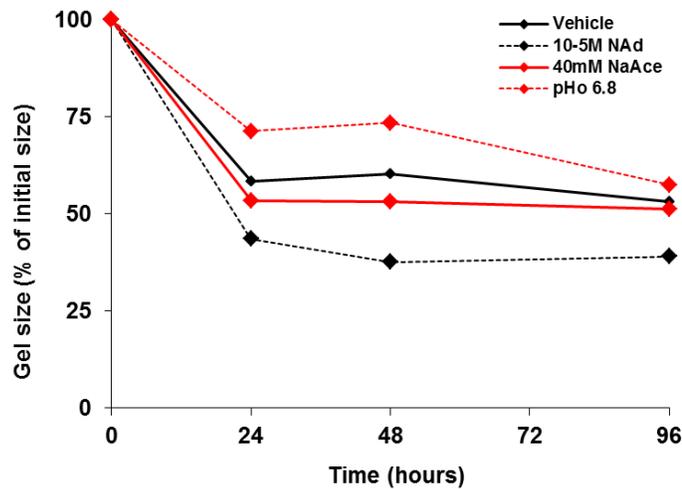


Figure 98: Time series for collagen contractility gels with various pharmacological stimuli
 Time series for collagen contractility gels (1.25×10^6 cells/ml gel) with various conditions. NAd, noradrenaline; NaAce, sodium acetate. Experiments conducted with unsupplemented DMEM as overlay. pH_o 6.8 achieved by adding 8 μ l of 1M HCl per 1 ml of unsupplemented DMEM.

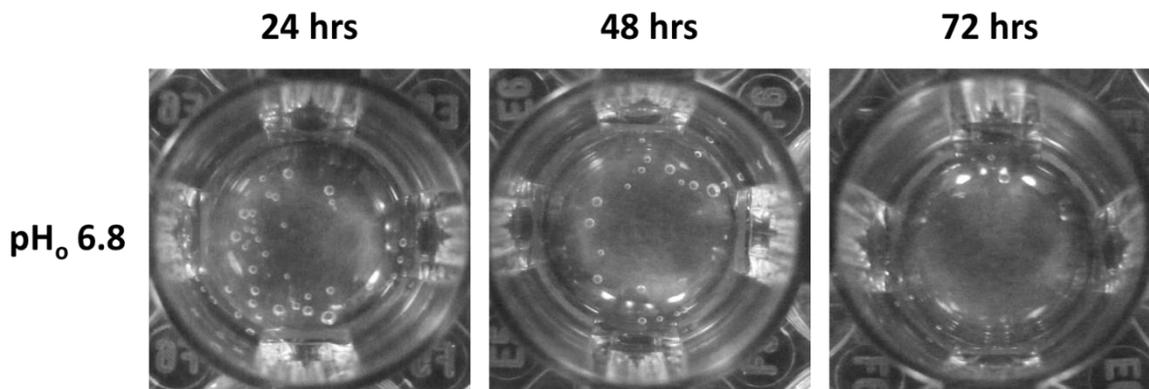


Figure 99: Digital images of collagen gels exposed to pH_o 6.8

With this, attempts to establish an *in vitro* system of investigating the influence of *SLC4A7* genotype on vascular smooth muscle contractility under acidic conditions were unsuccessful.

4.6.2. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis - HUASMCs

With *in vitro* studies of 19 HUASMC samples (9 C/T heterozygotes and 10 T/T major/protective allele homozygotes at rs13082711), any potential difference in pH_i recovery following intracellular acidosis was not detected by repeated measures two-way ANOVA (Figure 100). This was despite the apparent divergence of curves in cells exposed to NHE inhibition (DMA) in the presence of $\text{CO}_2/\text{HCO}_3^-$ (Figure 100b).

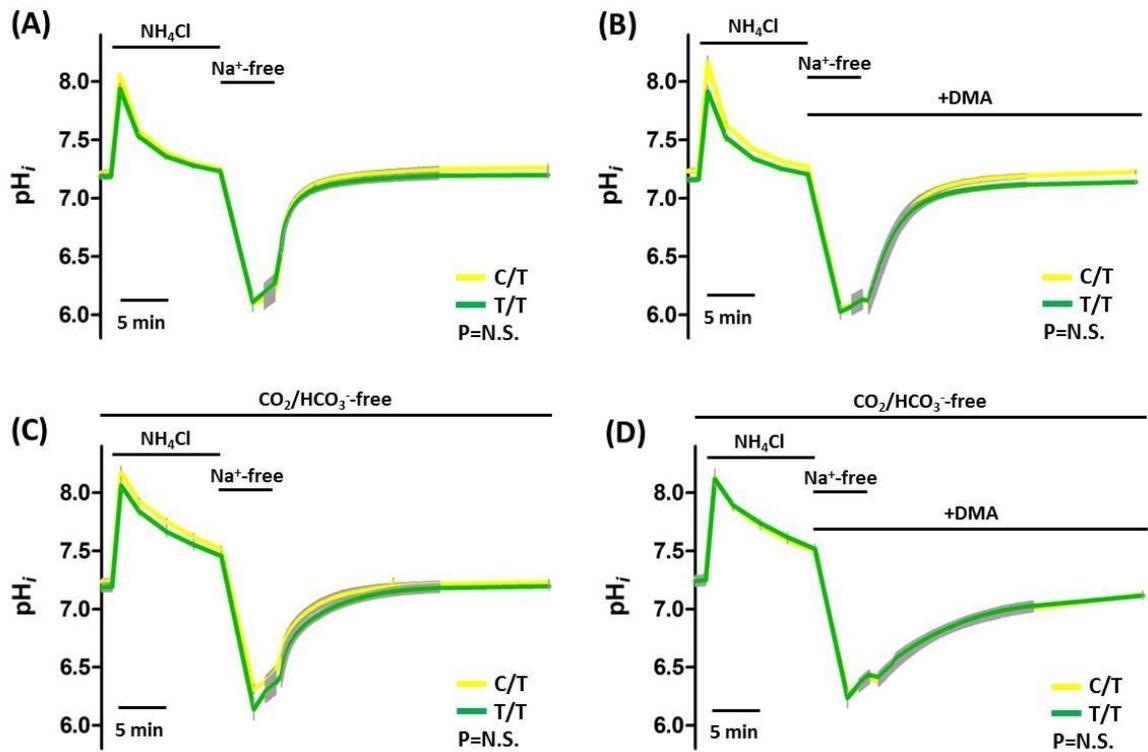


Figure 100: HUASMC pH_i recovery following intracellular acidosis, with and without CO_2/HCO_3^- or DMA

rs13082711 genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$). Grey error bars denotes SEM.

All $P > 0.05$ when comparing the two genotypes by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

To minimize the potential effects of variances in cell cycles by returning the cells into a quiescent state, HUASMCs were serum-starved for 48 hours (Pardee, 1974). In this synchronised state, they were then assessed for pH_i recovery in the presence of CO_2/HCO_3^- and DMA, thus mainly isolating NBC activity. There was a markedly reduced rate of pH_i recovery in serum starved cells (Figure 101a). As with the standard cultures (Figure 100b), any potential differences in genotypes were not detected by repeated measures two-way ANOVA, despite the divergence of recovery curves (Figure 101b).

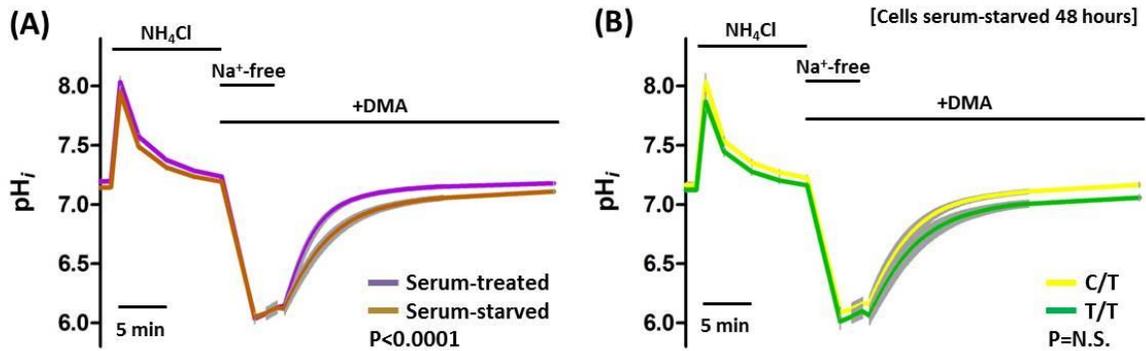


Figure 101: Serum-starved HUASMC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- and DMA

(A) $n=19$ paired HUASMCs. (B) $rs13082711$ genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$). Grey error bars denotes SEM.

Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

When comparing the initial rate of net base uptake (for the first 15 recorded seconds after reintroduction of sodium-containing buffer), there were no differences between the genotypes in either standard serum-containing or serum-starved culture systems (Figure 102). Notably, there was a reduced (but non-statistically significant for the sample size) initial net base uptake when HUASMCs are serum starved ($n=19$ pairs, -0.40 mmol/L/min, 95% CI -0.84 to 0.03 , $P=0.07$).

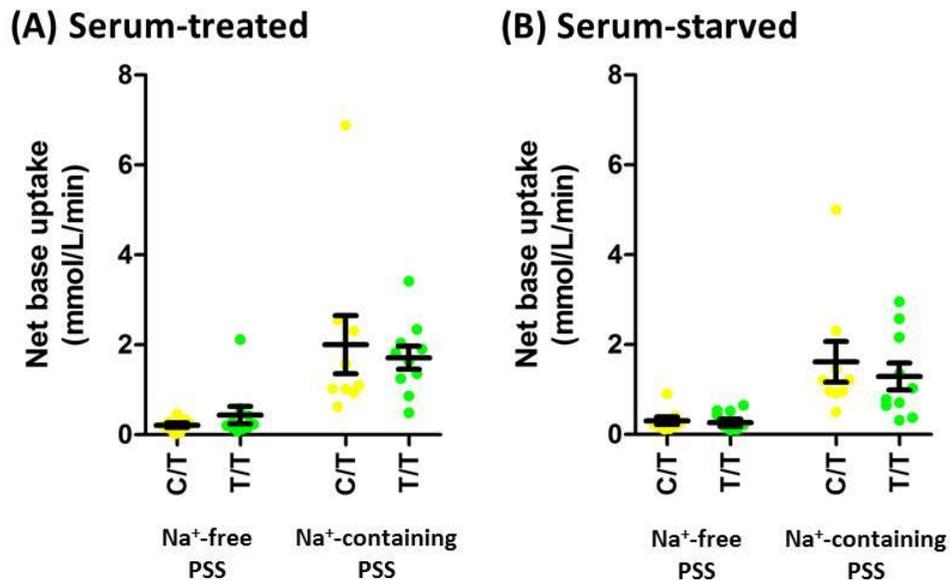


Figure 102: Initial net base uptake of HUASMCs in sodium-free and sodium-containing PSS in presence of CO_2/HCO_3^- and DMA

$rs13082711$ genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$).

$P>0.05$ for comparisons across genotype by Mann-Whitney U-test.

Although there were no differences in the initial rate of net base uptake across the genotypes, it is apparent that there is a large range of initial “starting” pH_i (Figure 103). Although there was no overall difference between the groups, the standard deviation and range were large for the both serum-containing (SD 0.23 and range 0.79) and serum-starved (SD 0.17 and range 0.68) cultures.

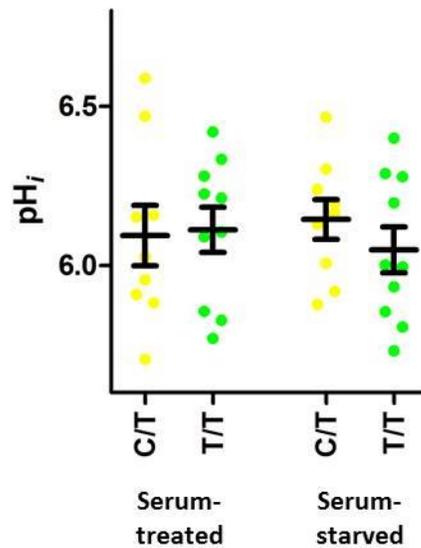
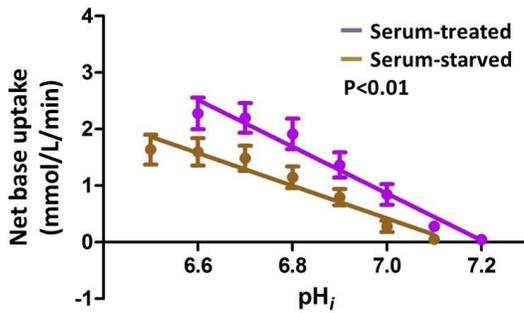


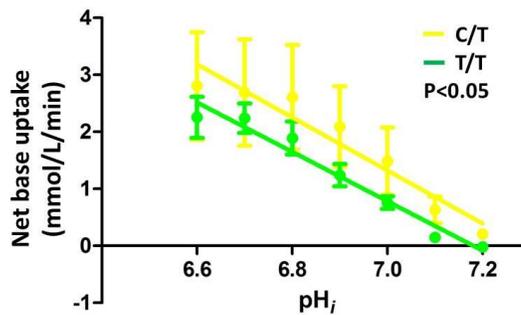
Figure 103: pH_i at which initial base flux measurements were obtained $P > 0.05$ for comparisons across genotype by Mann-Whitney U-test.

As activities of acid-base transporters are regulated by pH_i , these varying pH_i points from which the initial rates of net base uptake were assessed introduces marked variation in the net base uptake values of Figure 102. This could disguise any potential differences in HCO_3^- transport as suggested by the diverging curves in Figure 100b and Figure 101b. With this, the rate of net base uptake was determined at regular pH_i intervals (Figure 104). There is a genotype effect reflective of the aforementioned divergent pH_i recovery curves, where the C/T genotypes have a steeper slope when plotting net base uptake against each pH_i in both serum-treated and serum starved cells, although there were no statistically difference at each pH_i .

(A) Serum-treated vs serum-starved



(B) Serum-treated



(C) Serum-starved

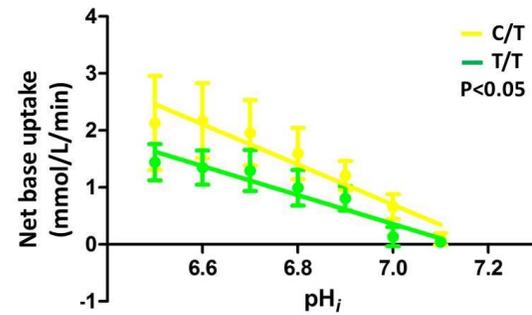


Figure 104: Rates of net base uptake at pH_i intervals for HUASMC following intracellular acidosis, with CO_2/HCO_3^- and DMA

(A) $n=19$ paired HUASMCs. (B and C) $rs13082711$ genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$).

Groups were compared by linear regression.

In addition to differences in net base uptake rates, there is also a genotype effect on final plateau pH_i after recovery from intracellular acidosis that is revealed in the presence of DMA ($\Delta pH_i = 0.087 \pm 0.027$, $P < 0.01$), which is replicated with the serum-starved cells ($\Delta pH_i = 0.108 \pm 0.035$, $P < 0.01$) (Figure 105). Notably, 48-hour serum starvation resulted in a reduction in final plateau pH_i in the presence of CO_2/HCO_3^- and DMA ($\Delta pH_i = 0.070 \pm 0.015$, $n=19$ pairs, $P < 0.001$). On the other hand, there is no significant difference when the cells were assessed under CO_2/HCO_3^- conditions.

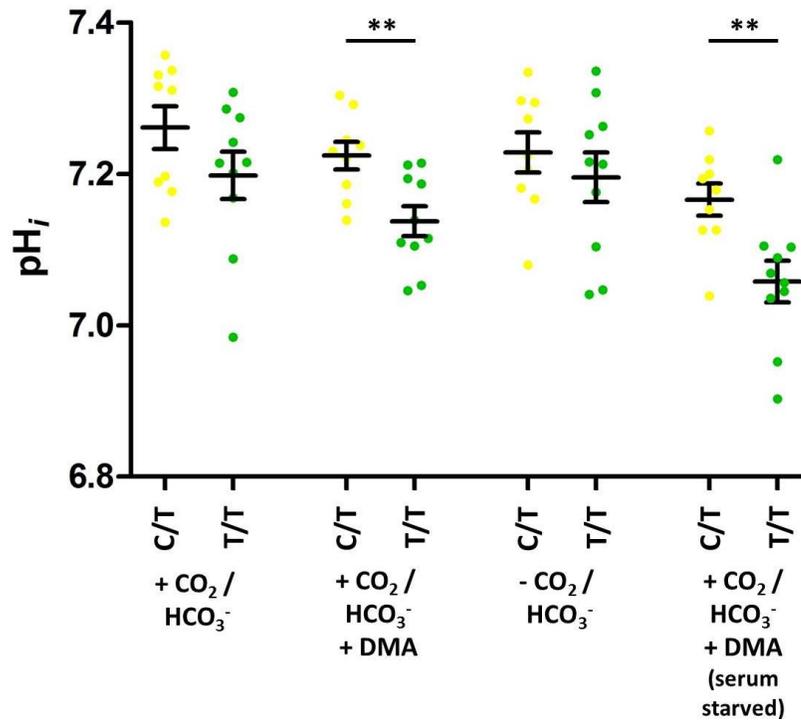


Figure 105: Final plateau pH_i after recovery from intracellular acidosis *rs13082711* genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$). ** $P < 0.01$ by Mann-Whitney U-test.

These observed genetic influence in net base flux (Figure 104) and resting pH_i (Figure 105) was confirmed to be CO_2/HCO_3^- -dependent as the differences disappear in CO_2/HCO_3^- -free conditions (Figure 106).

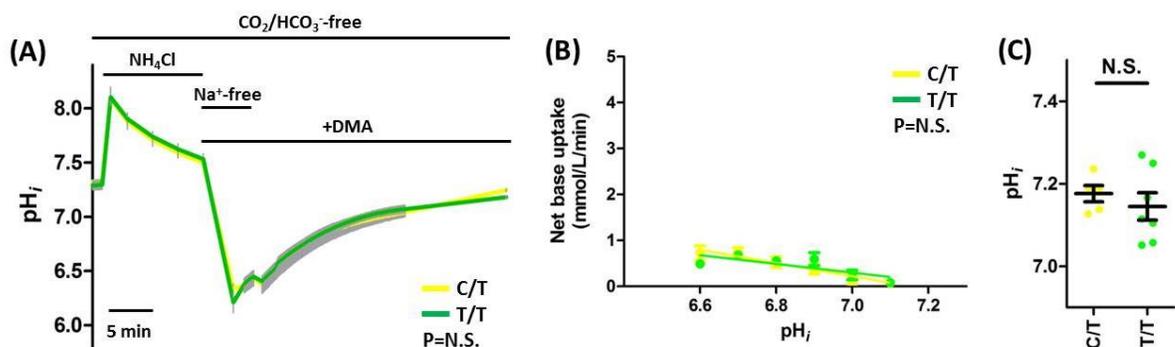


Figure 106: HUASMC pH_i recovery following intracellular acidosis, without CO_2/HCO_3^- , and in the presence of DMA *rs13082711* genotypes: C/T (yellow, $n=9$) and T/T (green, $n=10$). Grey error bars denotes SEM. Groups in (A) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B) compared by linear regression. Groups in (C) compared by Mann-Whitney U-test.

As expected, the buffering capacity of the cells was higher in the presence of $\text{CO}_2/\text{HCO}_3^-$, as compared to its nominal absence. The buffering capacity is similarly low regardless of the presence or absence of $\text{CO}_2/\text{HCO}_3^-$ at the lower ranges pH_i , relating to the lower intracellular HCO_3^- concentrations. The difference was particularly evident at the pH_i ranges closer to physiological levels where lower intracellular HCO_3^- concentrations are high and adds substantially to the buffering power (Figure 107). There were no differences in buffering capacity between cells of the two assessed genotypes in the presence or absence of $\text{CO}_2/\text{HCO}_3^-$, indicating that any genetic influence on NBCn1 function does not occur via changes in buffering capacities. Serum-starvation for 48 hours did not alter buffering capacity in the presence of $\text{CO}_2/\text{HCO}_3^-$ ($n=19$ pairs; with serum, 17.8 ± 0.7 mmol/L; serum-starved, 17.6 ± 0.7 mmol/L), and did not reveal any differences between the genotypes.

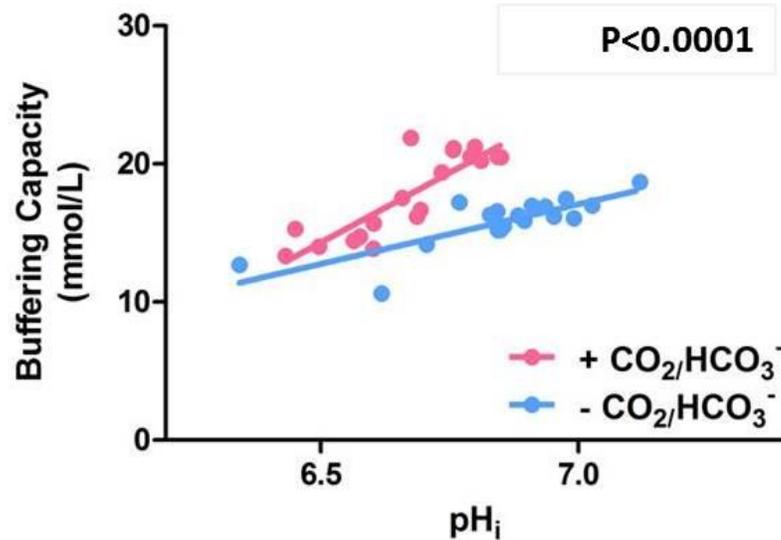


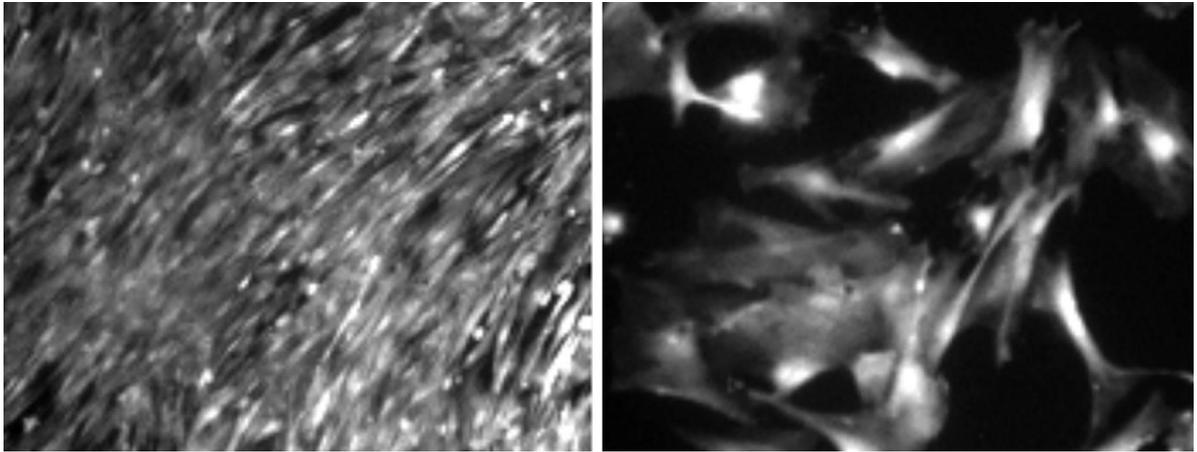
Figure 107: Buffering capacity of HUASMCs in the presence and absence of $\text{CO}_2/\text{HCO}_3^-$. Buffering capacity in the presence / absence of $\text{CO}_2/\text{HCO}_3^-$ as a function of mid-washout pH_i . $n=19$ paired HUASMCs. *** $P < 0.001$ by Wilcoxon sign-ranked test. Difference between slopes assessed by linear regression.

4.6.3. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis – HUASMCs (by cell morphologies)

Of the 19 HUASMCs assessed, it was apparent that it was divisible into two different macroscopic morphologies and cell culture surface confluencies. One group has a higher proliferation rate, achieves cell culture surface confluency and has an elongated spindle-shaped morphology (Figure 108, left). The second group has a lower proliferation rate, does not achieve cell culture surface

confluency, and has a larger surface area per cell. It is important to consider these differences as it poses potential confounding factors such as the cells that achieve confluency:

- will have a larger surface area-to-volume ratio, allowing for greater transporter density,
- may have higher cell-cell interactions,
- may be in different stages of the cell cycle, and
- has a macroscopic appearance closer to a "contractile" phenotype (Chamley *et al.*, 1974).



*Figure 108: Representative images of two different HUASMC morphologies
Fluorescence images of BCECF-loaded cultures. Typical appearance of HAUSMC primary cultures reaching confluence appearing elongated (left); and not those not reaching confluence even after two week's culture appearing to occupy a larger culture surface area (right). Same magnification for both images.*

The digital images were categorized by two independent assessors (Dr AD Moore and Dr FL Ng) blinded to the other characteristics of the samples. There were no statistical differences in the distribution of the two different morphologies when separated by rs13082711 genotype (Table 47), with the caveat of small sample numbers.

	rs13082711 genotype	
	C/T	T/T
Confluent	5	7
Non-confluent	4	3

Table 47: Distribution of HUASMC morphology by rs13082711 genotype
P>0.05 by Fisher's exact test.

The cells from the two different morphological groups had markedly different pH_i recovery profiles (Figure 109), particularly that of studies conducted in CO_2/HCO_3^- -free conditions (Figure 109C), predominantly reflecting the NHE component. Although there was no statistical difference in the pH_i recovery in cells studied in the presence of CO_2/HCO_3^- and DMA (isolating the NBC component, Figure 109B), there is a difference in net base flux at each pH_i level as reflected by the divergent curves.

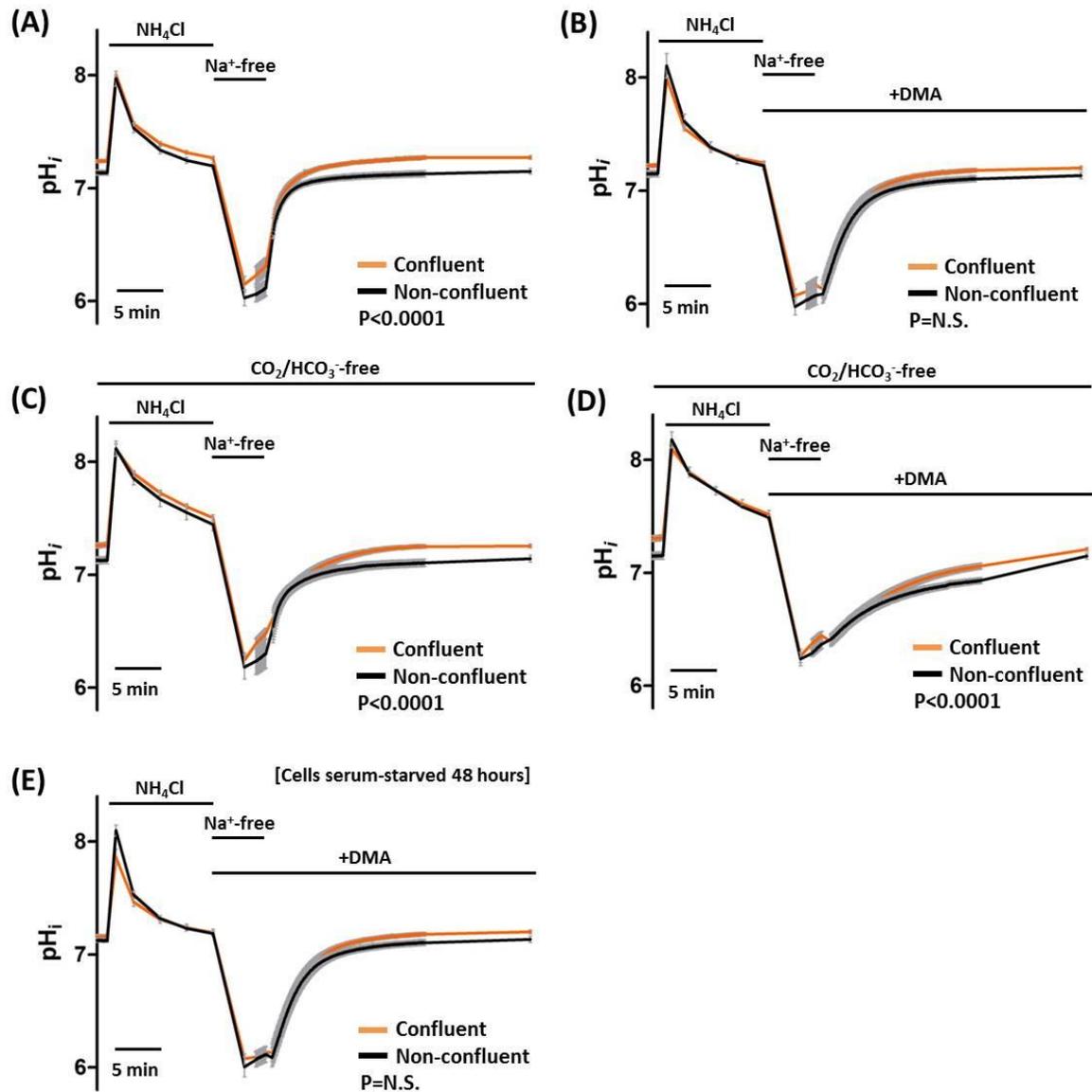


Figure 109: HUASMC pH_i recovery following intracellular acidosis, with and without CO_2/HCO_3^- or DMA, or in 48 hour serum-starved cultures
 Confluent (orange, $n=12$) and non-confluent subgroup (grey, $n=7$). Grey error bars denotes SEM.
 Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

The presence of two morphologically different groups that behave differently with regards to pH_i recovery necessitates a further assessment of the subgroups, as it would have potentially introduced a degree of variability of results in Figure 100. Most of the following will only consider the confluent subgroup, as there is a larger sample size and is more likely to reflect healthier proliferating cells with a more “contractile” morphology.

When considering this subgroup, the differences between the two genotypes are even more apparent than before. The genotypic effect on $\text{Na}^+/\text{HCO}_3^-$ -dependent net base flux and resting pH_i persists (

Figure 110A-C), and is also continues in the growth-arrested serum-starved cells (

Figure 110D-F), although as with the unstarved group, there were no statistically difference at each pH_i . The difference in pH_i plateau after recovery from intracellular acidosis revealed in the

presence of DMA is 0.079 ± 0.024 ($P < 0.01$), which is replicated with the serum-starved exposed to DMA ($\Delta\text{pH}_i = 0.118 \pm 0.043$, $P < 0.05$). These observed genetic influence in net base flux and resting

pH_i was confirmed to be $\text{CO}_2/\text{HCO}_3^-$ -dependent as the differences disappear in $\text{CO}_2/\text{HCO}_3^-$ -free conditions (

Figure 110G-I).

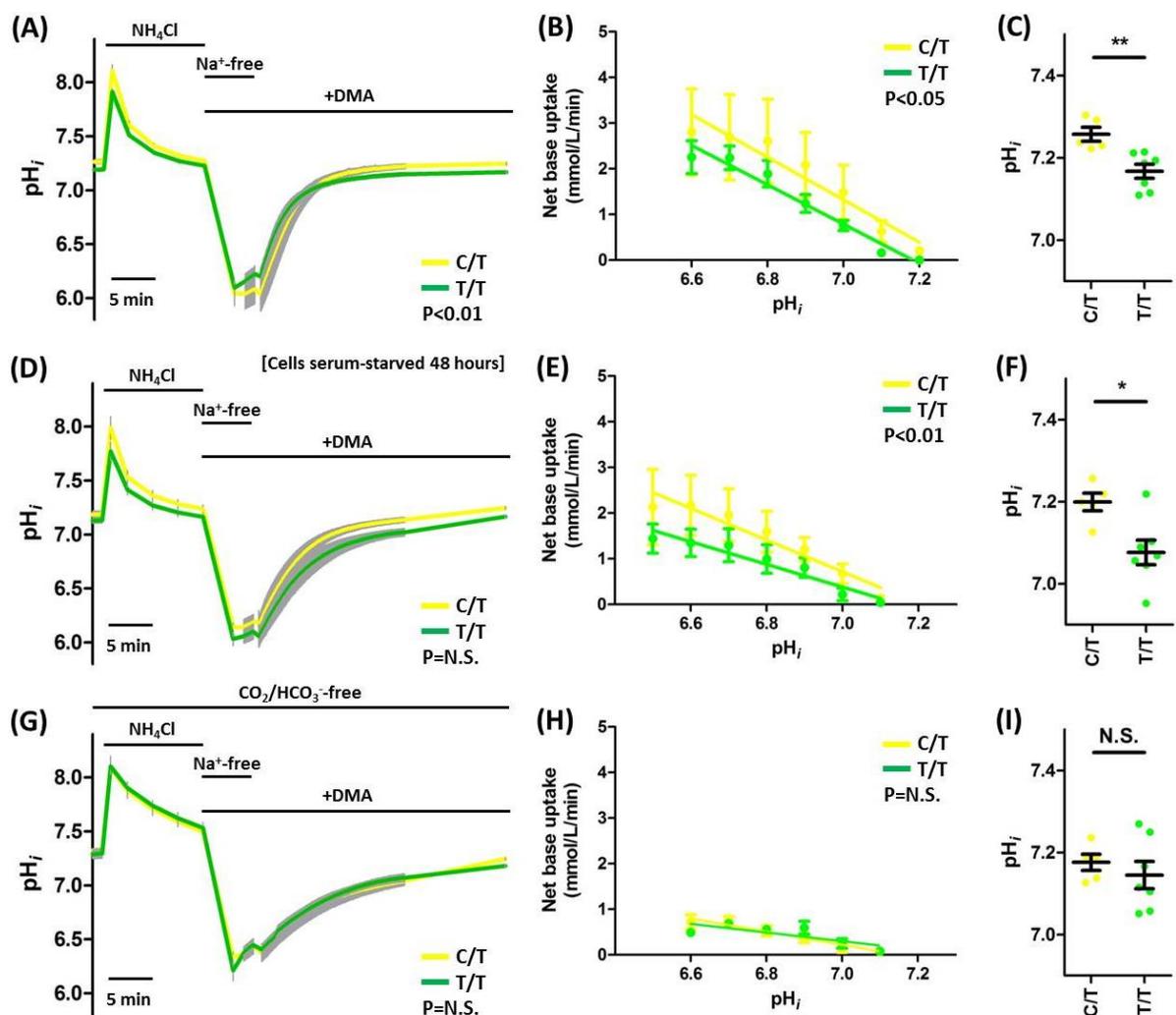


Figure 110: HUASMC pH_i recovery following intracellular acidosis, with and without $\text{CO}_2/\text{HCO}_3^-$, with DMA, or in 48 hour serum-starved cultures in confluent subgroup

rs13082711 genotypes: C/T (yellow, n=5) and T/T (green, n=7). Grey error bars denotes SEM. Groups in (A, D and G) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds. Groups in (B, E and H) compared by linear regression. Groups in (C, F and I) compared by Mann-Whitney U-test. The non-confluent subgroup is not further investigated due to the low numbers after further stratification by genotype (4 vs 3).

4.6.4. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis – HUASMCs and responses to FK506 and BAPTA

In 11 matched experiments (rs13082711 genotype: C/T n=4, T/T n=7), cells from four concurrent cultures of the same cell lineage were assessed for $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis while being exposed to (A) DMSO vehicle, (B) 10 μM FK506, (C) 30 μM BAPTA-AM loading, or (D) both FK506 and BAPTA-AM as previously represented in Figure 48.

There was no effect of 10 μM FK506 on $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery (Figure 111b), but a marked effect with BAPTA-AM incubation (Figure 111d).

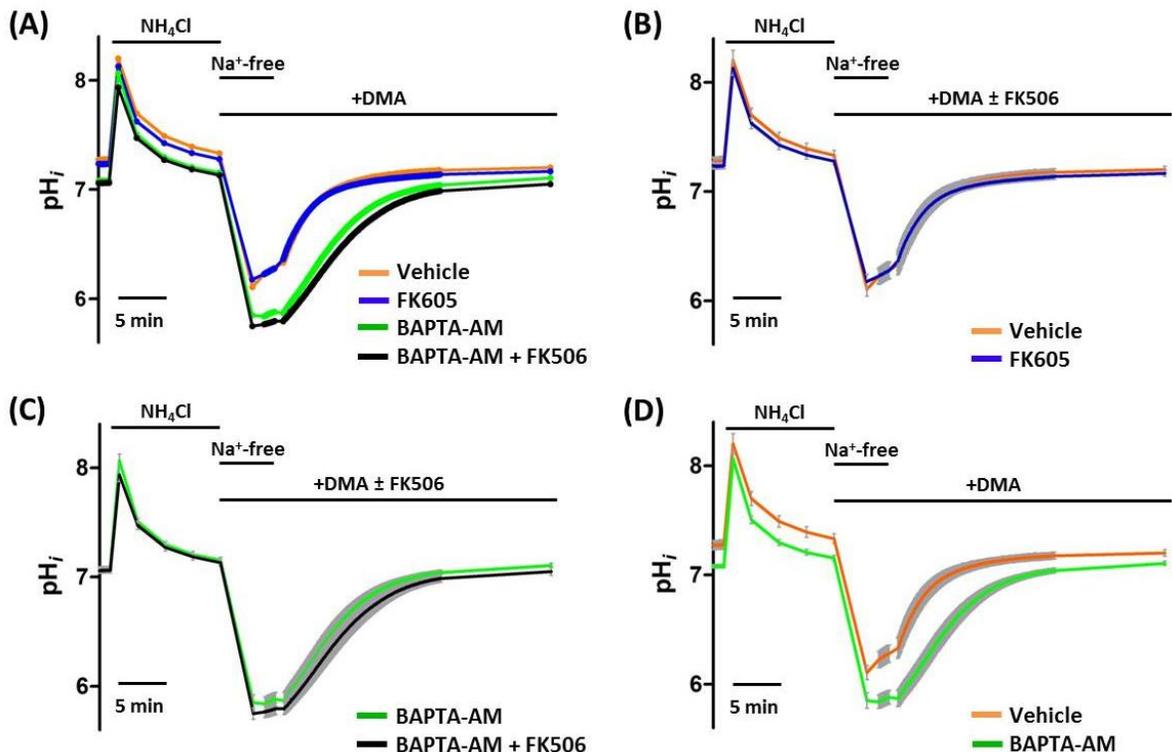


Figure 111: HUASMC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- and DMA, \pm FK506 or BAPTA-AM

$N=11$ sets of matched quadruplets. Grey error bars denotes SEM.

(A) only shows mean of all four matched experiments, whereas (B-D) shows mean \pm SEM of a pair. Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

When considering the rate of net base uptake at each specified pH_i , the absence of an effect of 10 μM FK506 is confirmed, as is the notable impact of pre- incubation with 30 μM BAPTA-AM for 20 minutes (Figure 112).

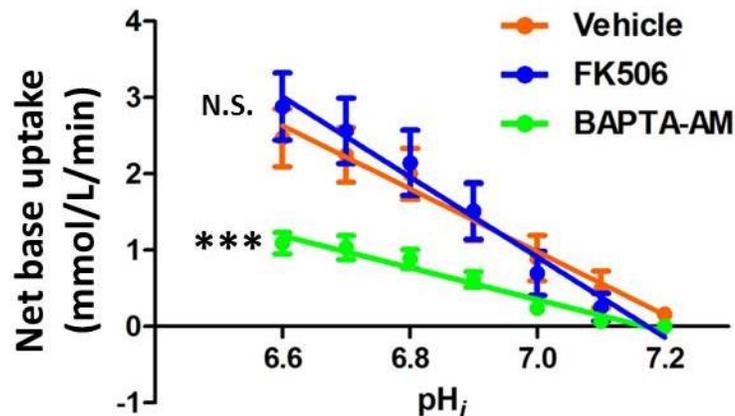


Figure 112: Rates of net base uptake at pH_i intervals for HUASMC following intracellular acidosis, with CO_2/HCO_3^- and DMA, \pm FK506 or BAPTA-AM

$n=12$ sets of matched quadruplets.

*** $P < 0.0001$ Vehicle vs. BAPTA-AM for slope by linear regression.

$P < 0.05$ Vehicle vs BAPTA-AM at pH_i of 6.6 and 6.7, after Bonferroni correction for multiple comparisons.

N.S. no significant difference Vehicle vs. FK506.

Similar to net base flux, FK506 also did not have an effect on plateau pH_i . However, intracellular calcium chelation with 30 μM BAPTA-AM loading significantly reduced plateau pH_i ($\Delta pH_i = 0.094 \pm 0.029$, $P < 0.01$) (Figure 113).

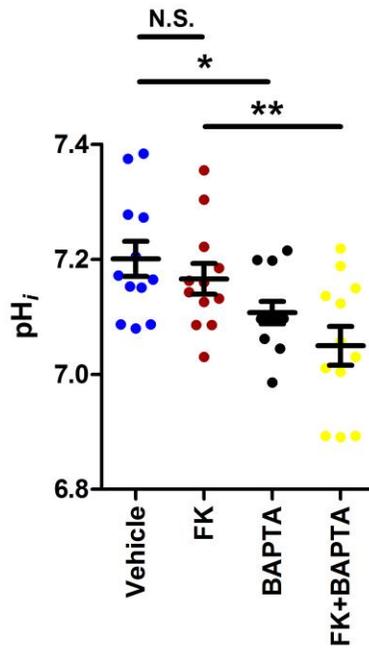


Figure 113: Final plateau pH_i after recovery from intracellular acidosis with CO_2/HCO_3^- and DMA, \pm FK506 or BAPTA-AM $n=12$ sets of matched quadruplets. Bonferroni correction for multiple testing applied. * $P<0.05$, ** $P<0.01$ by Wilcoxon sign-ranked test. N.S. no significant difference Vehicle vs. FK506.

It is unsurprising that as 10 μM FK506 did not have an effect on Na^+/HCO_3^- -dependent, DMA-insensitive pH_i recovery, the *SLC4A7* genotypic effect persists during calcineurin inhibition by FK506 (Figure 114), although there were no statistically difference at each pH_i .

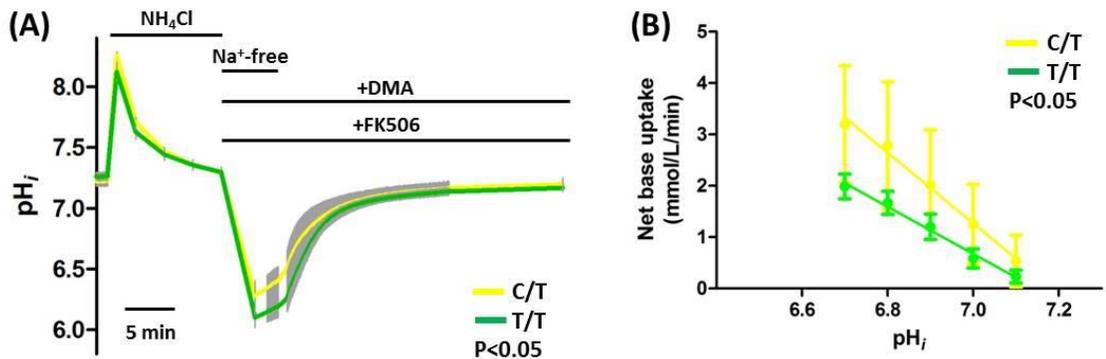


Figure 114: HUASMC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- , DMA and FK506 *rs13082711* genotypes: C/T ($n=4$) and T/T ($n=7$). Grey error bars denotes SEM. Groups in (A) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B) compared by linear regression.

4.6.5. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis - HUVECs

With *in vitro* studies of 20 HUVECs samples (10 C/T heterozygotes and 10 T/T major/protective allele homozygotes at rs13082711), any potential difference in pH_i recovery following intracellular acidosis was not detected by repeated measures two-way ANOVA (Figure 115). This was despite the apparent divergence of curves in cells exposed to NHE inhibition (DMA) in the presence of $\text{CO}_2/\text{HCO}_3^-$ (Figure 115b).

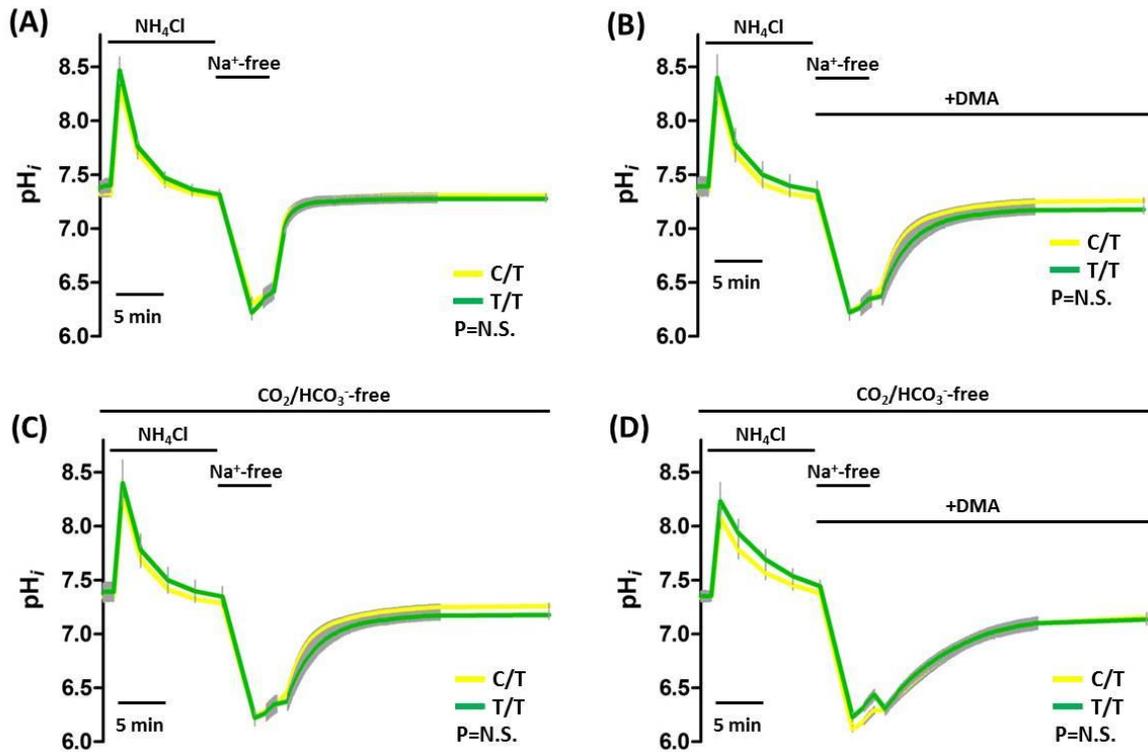


Figure 115: HUVEC pH_i recovery following intracellular acidosis, with and without $\text{CO}_2/\text{HCO}_3^-$ or DMA

rs13082711 genotypes: C/T (yellow, $n=10$) and T/T (green, $n=10$). Grey error bars denotes SEM.

Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

When comparing the initial rates of net base uptake (for the first 15 recorded seconds after reintroduction of sodium-containing buffer), there was a larger rate of net base uptake in *rs13082711* C/T samples compared to T/T samples ($+1.72 \pm 0.77$ mmol/L/min, $P<0.05$, Figure 116).

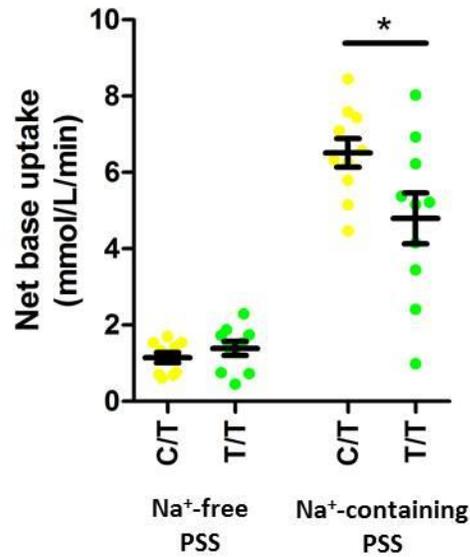


Figure 116: Initial net base uptake of HUVECs in sodium-free and sodium-containing PSS in presence of $\text{CO}_2/\text{HCO}_3^-$ and DMA *rs13082711* genotypes: C/T (yellow, $n=10$) and T/T (green, $n=10$). * $P < 0.05$ by Mann-Whitney U-test.

Similar to the finding in HUASMCs, there is a large range of pH_i where the “initial” base uptake was measured (Figure 117), with a large standard deviation (0.21) and range (0.73).

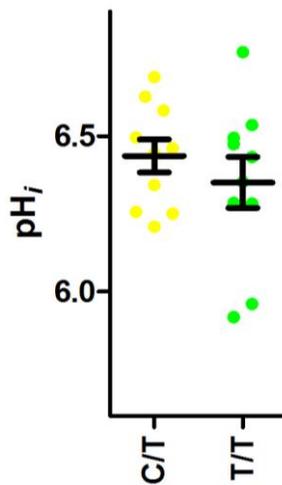


Figure 117: pH_i at which initial base flux measurements were obtained $P > 0.05$ for comparisons across genotype by Mann-Whitney U-test.

Just like with HUASMCs, the varying pH_i points from which the initial rates of net base uptake were obtained would introduce marked variation in the net base uptake values. Once again, the rates of net base uptake were determined at regular pH_i intervals (Figure 118). This supports a genotype

effect reflective of the aforementioned divergent pH_i recovery curves, where the rs13082711 C/T genotypes have a steeper slope when plotting net base uptake against each pH_i , although there were no statistically difference at each pH_i . This is consistent with the HUASMC results (Figure 104).

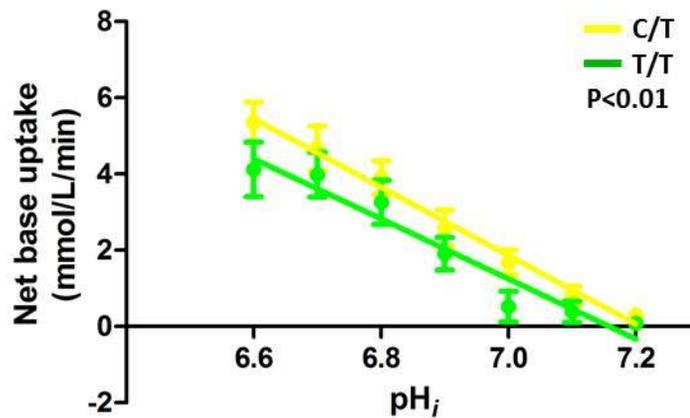


Figure 118: Net base uptake at pH_i intervals for HUVECs following intracellular acidosis, with CO_2/HCO_3^- and DMA
rs13082711 genotypes: C/T (yellow, n=10) and T/T (green, n=10).
Groups compared by linear regression.

In addition to differences in rates of net base uptake, it is worth considering any potential genotype-dependent effect on final plateau pH_i after recovery from intracellular acidosis that is revealed in the presence of DMA. In a similar direction to HUASMCs, the final plateau pH_i in C/T samples were higher (statistically non-significant) as compared to T/T samples ($\Delta pH_i = 0.082$, 95% CI -0.020 to 0.185, $P=0.10$,

Figure 119). This is a similar magnitude of that of HUASMCs, but due to the larger confidence interval, did not reach statistical significance. Of note, there is no difference when the cells were assessed under CO_2/HCO_3^- conditions, despite the different slopes of pH_i recovery (Figure 115).

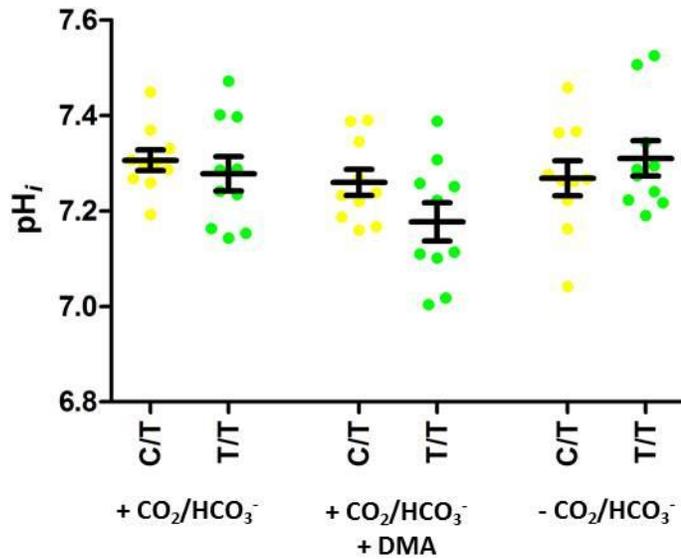


Figure 119: Final plateau pH_i after recovery from intracellular acidosis *rs13082711* genotypes: C/T (yellow, $n=10$) and T/T (green, $n=10$). No statistical differences between genotypes by Mann-Whitney U-test.

Just like that of HUASMCs, this observed genetic influence in net base flux and resting pH_i was confirmed to be CO_2/HCO_3^- -dependent as the differences disappear in CO_2/HCO_3^- -free conditions (Figure 120).

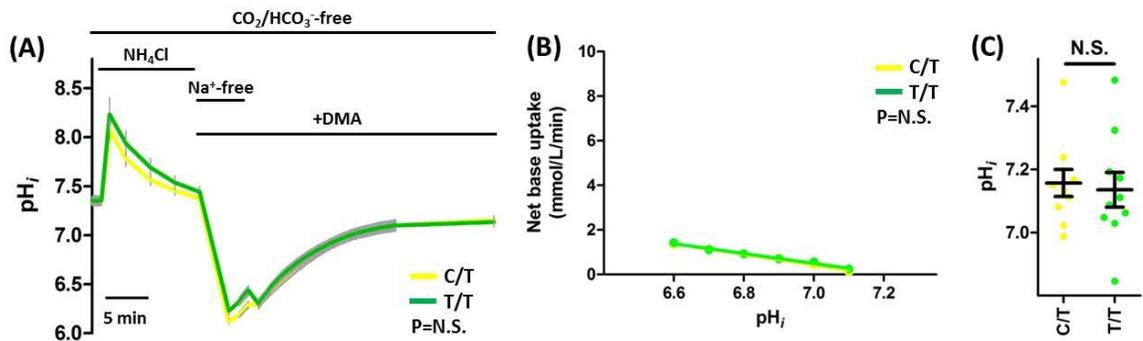


Figure 120: HUVEC pH_i recovery following intracellular acidosis, without CO_2/HCO_3^- , and in the presence of DMA *rs13082711* genotypes: C/T (yellow, $n=10$) and T/T (green, $n=10$). Grey error bars denotes SEM. Groups in (A) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B) compared by linear regression. Groups in (C) compared by Mann-Whitney U-test.

Similar to HUASMCs, the buffering capacity of the cells was higher in the presence of CO_2/HCO_3^- , as compared to its absence. As expected, and like HUASMCs, the buffering capacities are low

regardless of the presence or absence of $\text{CO}_2/\text{HCO}_3^-$ at the lower ranges pH_i . At higher pH_i there is still the divergence of the curves where the presence of $\text{CO}_2/\text{HCO}_3^-$ is associated with higher buffering capacities (Figure 121). There were no differences in buffering capacity for the two assessed genotypes in the presence and absence of $\text{CO}_2/\text{HCO}_3^-$, indicating that any genetic influence on NBCn1 function does not occur via changes in buffering capacities.

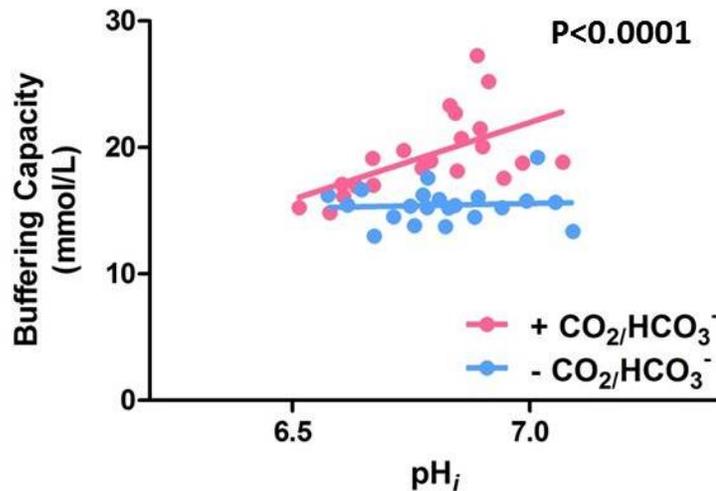


Figure 121: Buffering capacity of HUASMCs in the presence and absence of $\text{CO}_2/\text{HCO}_3^-$. Buffering capacity in the presence / absence of $\text{CO}_2/\text{HCO}_3^-$ as a function of mid-washout pH_i , $n=20$ paired HUVECs. Difference between slopes assessed by linear regression.

4.6.6. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis – comparing HUASMCs and HUVECs

To fully interpret the HUASMCs and HUVECs results, they should also be considered in relation to each other. It should be noted that none of the samples were paired for arterial and venous samples from the same umbilical cord, thus all comparisons were based on unpaired statistical tests.

HUVECs demonstrated faster pH_i recovery as compared to HUASMCs in $\text{CO}_2/\text{HCO}_3^-$ -containing buffer (Figure 122a), $\text{CO}_2/\text{HCO}_3^-$ -containing buffer with NHE inhibition by DMA (Figure 122b) and removal of NBC activity in $\text{CO}_2/\text{HCO}_3^-$ -free buffer (Figure 122c). It is notable that the difference is absent following the inhibition of both NHEs and NBCs, during studies with $\text{CO}_2/\text{HCO}_3^-$ -free buffer with the addition DMA (Figure 122d).

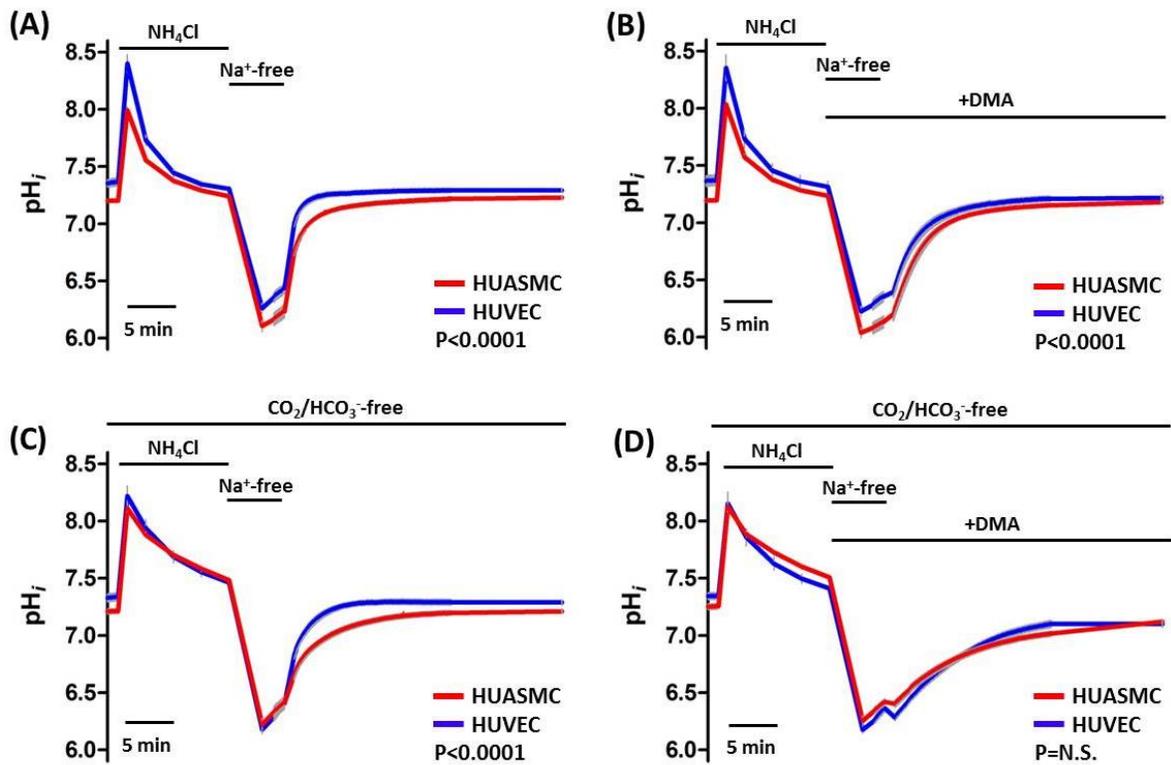


Figure 122: HUASMC and HUVEC pH_i recovery following intracellular acidosis HUASMC (red, $n=19$), HUVEC (blue, $n=20$). Grey error bars denotes SEM. Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer.

There were no differences in buffering capacities of HUASMCs and HUVECs, in both the CO_2/HCO_3^- -containing and CO_2/HCO_3^- -free conditions.

The net base uptake measured in the 15 recorded seconds before and after re-introduction of sodium-containing solutions is markedly different between the two cell types. HUVECs demonstrated higher net base uptake as compared to HUASMCs, in both the Na^+ -independent ($+0.94 \pm 0.15$ mmol/L/min, $P<0.0001$) and Na^+ -containing conditions ($+3.81 \pm 0.54$ mmol/L/min, $P<0.0001$) (Figure 123a). These initial rates of base uptake were obtained at a large range of pH_i values, which were also significantly different between the two cell types (Figure 123b). HUVECs also continue to show an increased net base uptake at regular pH_i intervals in Na^+ -containing buffer for (Figure 123c).

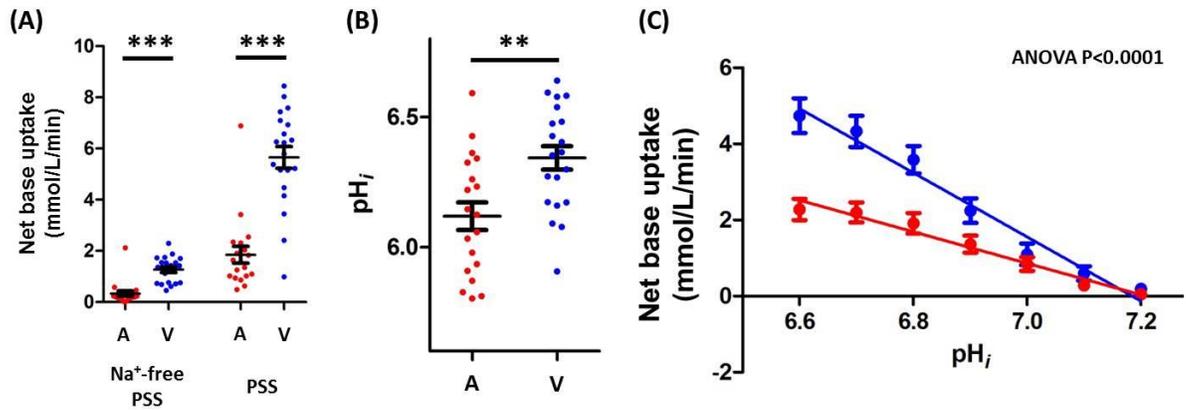


Figure 123: HUASMC and HUVEC net base uptake in presence of $\text{CO}_2/\text{HCO}_3^-$ and DMA
 (A) Net base uptake in the last 15 seconds in Na^+ -free buffer (+DMA) and the first 15 recorded seconds in Na^+ -containing PSS (+DMA), (B) pH_i at point of assessing Na^+ -dependent base uptake, (C) net base uptake at regular pH_i intervals.
 HUASMC (red, $n=19$), HUVEC (blue, $n=20$).
 ** $P < 0.01$. *** $P < 0.001$ by Mann-Whitney U-test.
 $P < 0.001$ HUASMC vs HUVEC at pH_i of 6.6 to 6.8, after Bonferroni correction for multiple comparisons.

In addition to the increased pH_i recovery, HUVECs also have a higher plateau baseline after recovery from intracellular acidosis in both $\text{CO}_2/\text{HCO}_3^-$ -containing ($\Delta\text{pH}_i = 0.064 \pm 0.030$, $P < 0.05$ by Mann-Whitney U-test) and $\text{CO}_2/\text{HCO}_3^-$ -free conditions ($\Delta\text{pH}_i = 0.080 \pm 0.033$, $P < 0.05$ by Mann-Whitney U-test). However, despite the increased $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake in HUVECs, there were no statistically significant difference in the plateau baseline ($\Delta\text{pH}_i = 0.040 \pm 0.031$, $P = 0.20$ by Mann-Whitney U-test) (Figure 124).

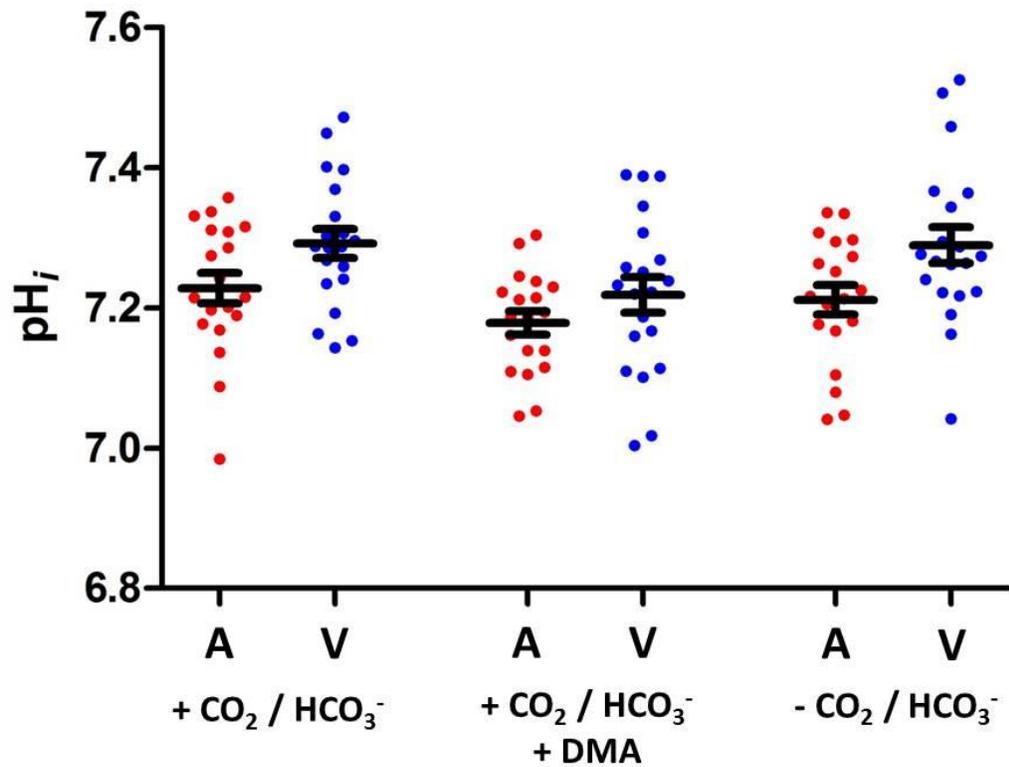


Figure 124: Final plateau pH_i after recovery from intracellular acidosis HUASMC (red, $n=19$), HUVEC (blue, $n=20$).
 * $P < 0.05$ by Mann-Whitney U-test.

HUVECs appear to have a larger population variance as compared to HUASMCs. The standard deviation of the HUVEC samples are larger than that of HUASMCs in the three main Na^+/HCO_3^- -dependent, DMA-insensitive characteristics assessed (Table 48). Of note, and as hypothesized, serum-starved HUASMCs had a generally lower variance as compared to serum-treated HUASMCs.

		Standard deviation		
		HUASMCs (n=20)		HUVECs (n=19)
		Serum-treated	Serum-starved	
Na^+/HCO_3^--dependent, DMA-insensitive	Initial net base uptake (mmol/L/min)	1.14	1.43	1.89
	Net base uptake at pH_i 6.6 (mmol/L/min)	1.05	1.22	2.04
	Plateau pH_i	0.09	0.07	0.11

Table 48: Standard variations of HUASMC and HUVEC samples

HUASMCs have consistently smaller standard deviations than HUVECs with regards to $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake and final plateau pH_i .

4.6.7. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis – overexpression models in A10 cells

A10 cells were transfected with 1 ng plasmid per 10 cm² cell culture surface (326E, 326K, Spl or vector control) for 48 hours. The transfected A10 cells were assessed for their pH_i recovery following ammonium-induced intracellular acidosis in the presence of $\text{CO}_2/\text{HCO}_3^-$ and DMA. The three overexpressing plasmids were different from that of the vector plasmid, but there were no differences between the three overexpression plasmids (Figure 125).

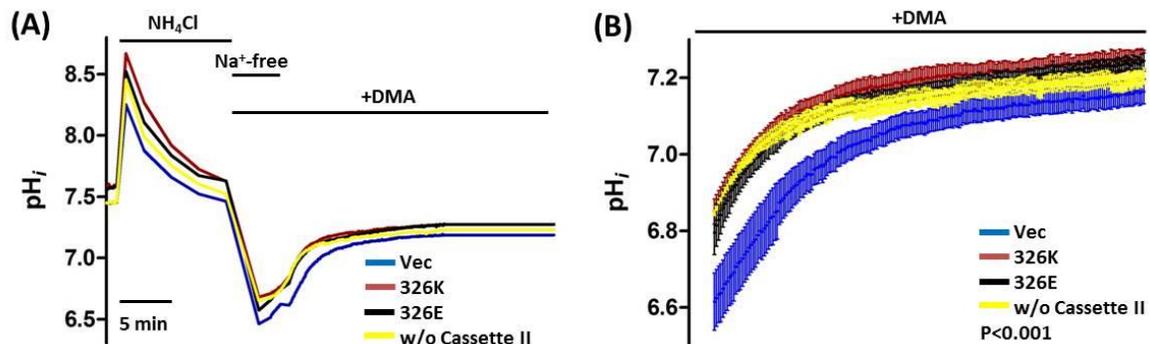


Figure 125: SLC4A7-expression plasmid-transfected A10 cells pH_i recovery following intracellular acidosis with $\text{CO}_2/\text{HCO}_3^-$ and DMA

(Left) only shows the mean at each timepoint. (Right) zoomed view of mean \pm SEM of the first 600 seconds after reintroduction of Na^+ -containing buffer.

Groups compared by repeated measures two-way ANOVA over the 190 time points between 30 seconds to 600 seconds after re-introduction of Na^+ -containing buffer.

Blue, Vector; Brown, 326K; Black, 326E; Yellow, without Cassette II.

n = 6 sets of transfections.

The buffering capacities in the presence of $\text{CO}_2/\text{HCO}_3^-$ in the four different plasmid-transfected A10 cells were not different (Vector 20.4 ± 1.6 mmol/L, 326K 21.5 ± 1.7 mmol/L, 326E 19.4 ± 1.8 mmol/L, without Cassette II 22.9 ± 1.2 mmol/L; one-way ANOVA).

The net Na^+ -dependent base uptake measured based on the 15 recorded seconds before and after re-introduction of sodium-containing solutions is elevated in A10 cells transfected with *SLC4A7*-overexpressing plasmids as compared to vector plasmid (326K vs Vec $+2.4 \pm 1.1$ mmol/L/min; 326E vs Vec 3.2 ± 1.2 mmol/L/min; w/o Cassette II vs Vec 3.0 ± 0.9 mmol/L/min), but with no difference between the three overexpression plasmids (

Figure 126a). Like that of HUASMCs and HUVECs, initial rates of net base uptake were obtained at a large range of pH_i values, which was also significantly different between the overexpression plasmids and vector (Figure 126b). The analysis that accounted for net base uptake at regular pH_i intervals also show a higher rate of net base uptake in Na^+ -containing buffer for the expression plasmids as compared to vector, but no difference between the three expression plasmids (Figure 126c).

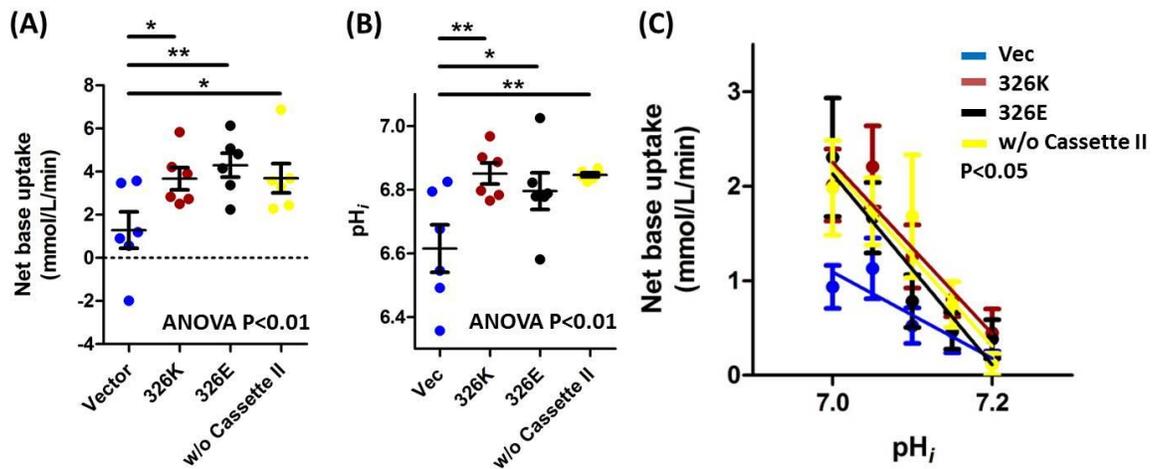


Figure 126: Net base uptake of transfected A10 cells in presence of CO_2/HCO_3^- and DMA (A) Net base uptake in the last 15 seconds in Na^+ -free buffer (+DMA) and the first 15 recorded seconds in Na^+ -containing PSS (+DMA), (B) pH_i at point of assessing Na^+ -dependent net base uptake, (C) net base uptake at regular pH_i intervals, where groups were compared by linear regression.

$n=6$ sets of parallel transfections.

* $P < 0.05$, ** $P < 0.01$ by repeated measures one-way ANOVA with Bonferroni correction.

(C) $P < 0.05$ by two-way ANOVA.

In addition to the increased pH_i recovery, A10 cells transfected with overexpression plasmids also have a higher plateau baseline after recovery from intracellular acidosis as compared to vector (326K vs Vec $\Delta pH_i = 0.085 \pm 0.016$; 326E vs Vec $\Delta pH_i = 0.087 \pm 0.023$; w/o Cassette II vs Vec $\Delta pH_i = 0.041 \pm 0.027$). Again, there were no differences between the three expression plasmids (Figure 127).

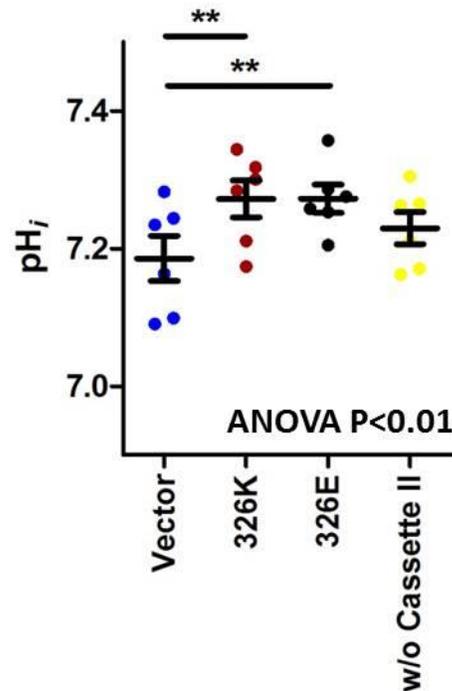


Figure 127: Final plateau pH_i after recovery from intracellular acidosis n=6 sets of parallel transfections. **P<0.01 by repeated measures one-way ANOVA.

In these studies with transfected A10 cells, the overexpression plasmids produced more NBCn1 signal as compared to transfection with vector. There were no differences between the three overexpression plasmids (

Figure 128). Care should be taken with the numerical values of the y-axis, as there can be inter-blot variability in signal (although all sets of transfections were assessed in one blot), and quantification of immunoblots signal would not have a linear relationship with actual protein quantity.

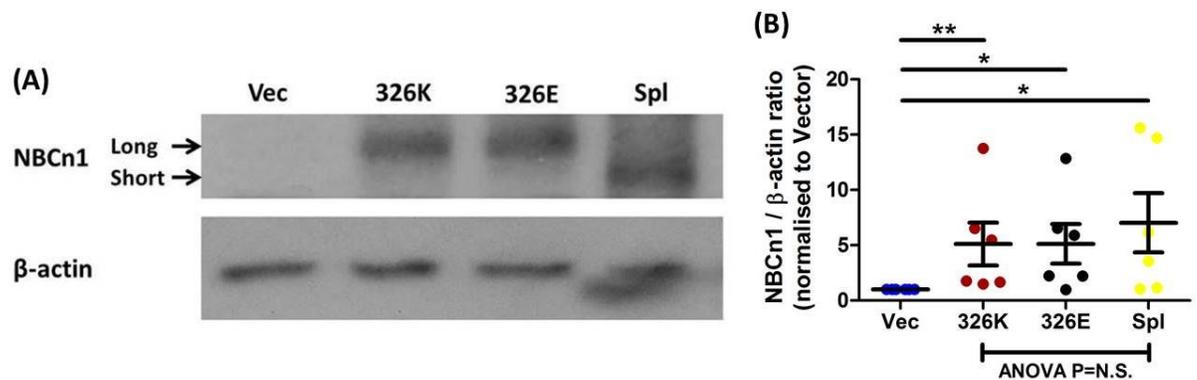


Figure 128: NBCn1 immunoblot signals in transfected A10 cells Representative immunoblot on left. n=9 sets of parallel transfections.

* $P < 0.05$ by repeated measures one-way ANOVA.

4.6.8. $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis – responses to FK506 in A10 overexpression models

A10 cells transfected with the three overexpression plasmids (326K, 326E and splice variant without Cassette II) were assessed for their $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery following ammonia-induced intracellular acidosis. Any potential difference in pH_i recovery was not detected by repeated measures two-way ANOVA (

Figure 129). This was despite the apparent divergence of curves in cells transfected with 326K and 326E, but not the splice variant.

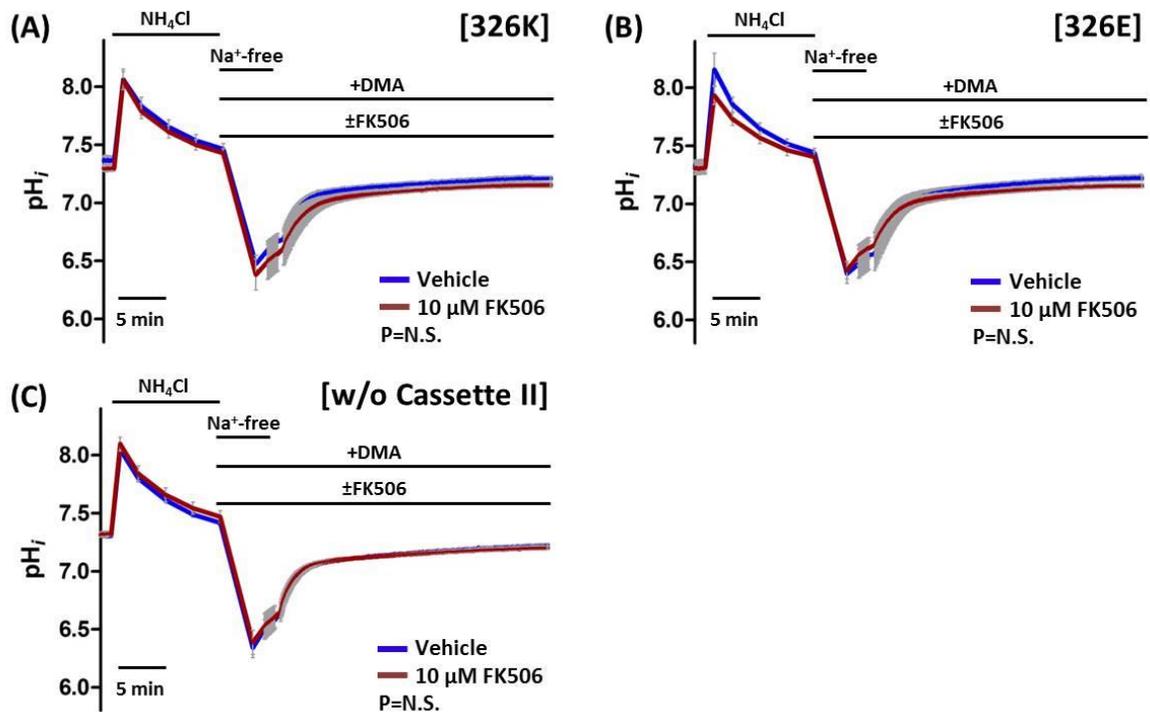


Figure 129: Transfected A10 cells pH_i recovery following intracellular acidosis, with $\text{CO}_2/\text{HCO}_3^-$ and DMA \pm FK506

Blue denotes vehicle; brown denotes FK506. $n = 4$ parallel transfections and paired experiments. Grey error bars denotes SEM.

Groups compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer, all $P > 0.05$.

Similar to HUASMCs, A10 cells treated with 10 μM FK506 did not reveal a statistically significant difference in plateau baseline pH_i as compared to vehicle in 326K-transfected ($\Delta\text{pH}_i = -0.057 \pm$

0.028, P=0.09), 326E-transfected ($\Delta\text{pH}_i = -0.065 \pm 0.028$, P=0.06), splice variant-transfected ($\Delta\text{pH}_i = -0.022 \pm 0.023$, P=0.38) cells (

Figure 130). The lack of a statistically significant result may however be due to the small sample size.

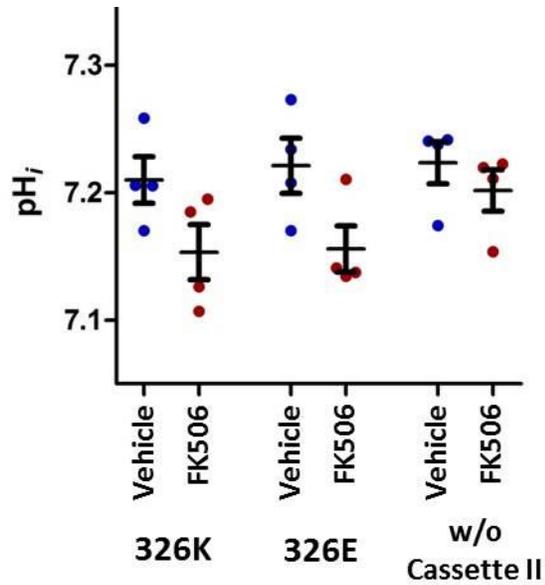


Figure 130: Final plateau pH_i after recovery from intracellular acidosis. Blue denotes vehicle; brown denotes FK506. $n=4$ parallel transfections and paired experiments. $P>0.05$ for differences between Vehicle and FK506 for each pair.

5. Discussion

5.1. SLC4A7 genotyping

5.1.1. rs13082711, rs13096477 and rs3755652 genotype distribution in Hardy-Weinberg equilibrium

The quality control parameters for the genotyping assay were adequate, with call rates and inter-assay concordance of greater than 95%, as is the 100% concordance with the gold standard DNA sequencing (Section 4.1).

The genotype distribution of the three SNPs is in Hardy-Weinberg equilibrium (Table 27 and

Table 28), which is consistent with public databases (Section 3.2). The absence of a deviation from Hardy-Weinberg equilibrium implies (although does not absolutely confirm) that there is no large selection effect of the genetic variation. As previously discussed, the genotype distribution is not expected to closely approximate that of public databases, as regional differences would be important, and particularly the local sample pool is more heterogeneous as compared to those databases.

There were also no differences in the genotype, haplotype distribution and rate of incomplete datasets between HUASMCs and HUVECs (Table 43) suggesting there were no systematic errors in genotyping across samples.

5.1.2. rs820430 genotype distribution in Hardy-Weinberg equilibrium

Similar to the results in Section 5.1.1 above, the rs820430 genotyping assay demonstrated adequate quality control parameters, and the genotype distribution was in Hardy-Weinberg equilibrium (Table 44). Although the genotype distribution is not expected to directly approximate to public databases, it does appear close to that of HapMap-CEU (Table 29).

With this, the preceding interpretation of the genotyping results is also relevant here.

5.2. Gene expression

5.2.1. *SLC4A7 (but not NEK10 and EOMES) mRNA detected in a selection of samples*

An expression panel with end-point RT-PCR identified *SLC4A7* mRNA in all HUASMCs, HUVECs and renal biopsy samples selected (Figure 66). On the other hand, mRNA of the flanking genes (see Section 3.1), *NEK10* and *EOMES* were not detected. This supports the hypothesis that *SLC4A7* is the *most likely* gene, if any, that the genome-wide associated SNP rs13082711 has a functional impact on. This is compatible with publically available data, where the GTEx (Genotype-Tissue Expression project; <http://www.gtexportal.org/home/>) RNA-seq data demonstrates high transcript levels of *SLC4A7*, but low *NEK10* and *EOMES*, in arteries. The results on renal samples are more difficult to interpret as the sample is from the renal cortex, whereas *SLC4A7* is localized to the medulla. Nevertheless, the levels of *NEK10* and *EOMES* in the renal cortex are still lower than *SLC4A7* (Figure 55).

Whilst it is *unlikely* that *NEK10* or *EOMES* have an important role in blood pressure regulation through these tissues, a few caveats should be recalled:

- End-point RT-PCR has lower sensitivity, and may not detect low copy numbers. It could also be argued, that very low copy numbers of cell-cycle regulators (*NEK10*) and transcription factors (*EOMES*) could still be physiologically relevant.
- Blood pressure regulation occurs in concert with many different organs (see Section 1.8), and the absence of expression in vascular smooth muscle cells, endothelial cells and kidneys does not *necessarily* exclude a role in blood pressure regulation.
- That from a generalisability viewpoint, HUASMCs and HUVECs are not derived from resistance vessels (see Section 2.1.1) and that the nephrectomy biopsies are sampled from cancerous kidneys, even if they are from the non-cancer pole. It could be argued that they are not representative samples of tissues regulating blood pressure.

On balance, these are weak counter-arguments, and *SLC4A7* remains as the best candidate at this genomic locus.

A final theoretical point for consideration is that the blood pressure-associated SNP may actually have long-ranging effects, past the local region, or even trans-chromosomes. While theoretically possible, the results set out in this thesis where there is a genotypic effect on *SLC4A7* gene expression and protein function means a trans-effect is also less likely.

5.2.2. Allelic imbalance analysis identifies the risk (minor) allele at rs13096477 as being more prevalent in mRNA

The allelic imbalance in the cDNA chromatographs of rs13096477 heterozygotes (Figure 67) implies that there is more C (minor) allele mRNA as compared to the T (major) allele in HUASMC, HUVEC and nephrectomy biopsies. The method has the advantage that both alleles are exposed to the same cellular environment thus is unlikely to be due to an external factor.

As the exonic SNP rs13096477 is in high LD with 92 other SNPs, it should be emphasized that it is not necessarily the causative polymorphism, and could have been a result of any other the other SNPs in high LD.

In theory, the increased presence of the C allele mRNA could be due to either differences in mRNA transcription (i.e. increased C allele or decreased T allele transcription), or reduced mRNA turnover / degradation (i.e. the converse, decreased C allele or increased T allele degradation). In the absence of a nonsense mutation or an SNP of interest in the 3'-UTR nor predicted to interact with miRNAs, the more likely explanation for the difference between the alleles is that of differential transcription rather than degradation.

The consistency of effect across the three assessed sample types suggests that this is a common mechanism across multiple cell types. Also reassuringly, a combined analysis of the samples of all three cell types, the C allele appears to have 50% more transcript levels compared to T allele (n=27, P<0.0001).

A limitation of the allelic imbalance analysis methodology is the inability to discriminate between isoforms. As the exonic rs13096477 is placed in exon 3, outside of Cassette II, the assay will detect the differences as a total of both long and short isoforms. With this, it is unable to determine whether if (or not) either isoform is dominant in its effect.

5.2.3. qRT-PCR did not demonstrate an allele-dependent association with SLC4A7 mRNA/cDNA levels or splicing, but did demonstrate a large inter-sample variability

Unlike allelic imbalance studies (Section 5.2.2), qRT-PCR did not demonstrate an allele-dependent association with *SLC4A7* transcript levels (Figure 68). Whilst this could be true result, it should be recalled *SLC4A7* is prone to up- and down-regulation (Section 1.12), dependent on pH, other hormonal stimuli and its position in the cell cycle. This is important with the potential variability in

cell states when using cell culture systems with supplemented culture media and not cell cycle-synchronised before mRNA isolation. Additionally, although mRNA isolation was conducted at passage number 1 for each primary culture, the initial number of seeded cells may vary and thus the number of cell divisions to reach confluency and its phenotypic state may also be variable.

However, the qRT-PCR studies indicate a smaller dC_t (*SLC4A7* – *18S*) in HUVECs as compared to HUASMCs in paired samples (

Figure 69a). There are two main interpretations which will be considered in turn. Firstly, this may reflect that HUVECs have a higher *SLC4A7* expression, which would also be consistent with measurements of NBCn1 function (Section 4.6.6). The alternative explanation is that the expression levels of *18S* mRNA per cell is not equal between the two cell types, and therefore the dC_t (*SLC4A7* – *18S*) would not be an appropriate housekeeping control across cell types.

In addition to the potential increased *SLC4A7* mRNA levels in HUVECs, there also appears to be a difference in relative proportions of long and short variants. HUASMCs had a higher proportion of long variants as compared to HUVECs (

Figure 69b). Although the C_t values of the short variant were consistently lower than those of the longer variant, this does not necessarily indicate there is more of the shorter variant mRNA present. As previously discussed, it may be differences in PCR efficiency, and in addition probes for the short variant can also detect that of the longer variant. More relevant to the study of the genetic polymorphism, there was no genetic effect of rs13082711 on splicing.

Another striking result of the qRT-PCR results is the inter-sample variability (

Figure 70). The positively skewed distribution, particularly in HUASMCs, may imply that there is a baseline level of expression, but can be induced. Additionally, the smaller interquartile ranges of HUVECs may be related to the finding above that they have a larger overall transcript levels, perhaps already being in an induced state. It should be noted that this is unlikely to be related to intra-sample variability as an exclusion criteria for a successful assay includes any sample with replicates greater than 1 C_t apart is excluded, even for just one of the three (*SLC4A7_{long}*, *SLC4A7_{short}* and *18S*) probes.

Based on the population standard deviation, an expected difference between C/T and T/T genotypes of 25% transcript (as the C allele is predicted to have 50% more transcript than T allele

by the allelic imbalance method, equating to a ddC_t of 1.322) and a 1:2 ratio of genotype distribution, around 400 HUVECs and 1900 HUASMCs would be required to achieve a power of 80% with an α -value of 0.05. As previously addressed, the caution for this power calculation is that it relies on a standard deviation value in the presence of a non-parametric distribution, together with the assumption of a 100% PCR efficiency.

Nephrectomy biopsy samples were not assessed by qRT-PCR. This is owing to *SLC4A7*/*NBCn1* expression being localized to the mTAL, and non-selective biopsies will obtain highly heterogeneous samples with varying proportions of mTAL. This will mean the main determinant of the quantity of *SLC4A7* mRNA within a sample is the biopsy contents rather than any genotype influence. Once again, this highlights the advantage of allelic imbalance analyses over qRT-PCR.

With all these caveats to qRT-PCR analysis, the allelic imbalance analysis would appear to be a better reflection of allele-dependent mRNA/cDNA levels.

5.2.4. *NBCn1* protein immunoblots identifies the risk (minor) allele as having a higher expression in total cell protein of HUASMCs but not HUVECs

Consistent with the allelic imbalance analyses findings, the minor allele is dose-dependently associated with increased quantity of total cellular *NBCn1* protein detected (Figure 72). The two bands present on the immunoblot may potentially reflect the two main isoforms (predicted sizes 136kDa and 127kDa), with and without Cassette II. In the absence of an antibody raised against Cassette II, or utilising highly-specialised studies such as protein sequencing / mass spectrometry, it is not possible to be certain whether these reflect two different variants or differences in post-translational modification.

Like the notable variability in the HUASMC *SLC4A7* mRNA levels, there was also a corresponding large variability in the β -actin normalised immunoblot signals. Accepting the significant caveat of the lack of linearity at the extremes of signal detection, arithmetically there is up to 70-fold signal difference comparing the maximum and minimum signals. This was despite culturing in serum-free media after reaching confluency. This is consistent with the findings of a large range of mRNA transcript levels as detected by qRT-PCR (Section 5.2.3).

The presence of increased *total* protein, does not imply functionality, and can theoretically potentially reflect sequestration in non-functioning regions of the cell. This will be explored later in Section 5.4.

On the other hand, HUVEC samples did not show a genotype-dependent NBCn1 total protein expression (Figure 73). It should be noted that the probability of a false negative appears low. Another notable difference is that immunoblots for HUVECs show much lower variability in NBCn1 expression levels, just over a 3.5-fold signal difference between the maximum and minimum signal intensities of the 14 sample panel, as compared to the HUASMCs. This is consistent with the lower variability in HUVEC *SLC4A7* mRNA transcript levels.

To explain the difference in the HUASMC and HUVEC results, it can be argued that the higher transcript levels of HUVECs may saturate the translation mechanism, and therefore eliminates any potential differences in total protein expression. The lower expression and increased variability in HUASMCs opens up the possibility of a larger impact of genotype on regulating gene expression and protein function.

5.2.5. *SLC4A7 E326K variation does not impact on isoelectric focusing*

In attempt to reduce the influence of cell culture conditions, isoelectric focusing was attempted under the initial prediction that the 326E and 326K variants based on the rs3755652 non-synonymous polymorphism (Section 3.9). If the two variants produced separate signals, it would be possible to assess the impact of both alleles in the same intracellular environment in heterozygotes, in a manner like allelic imbalance analysis for RNA (Section 5.2.2).

However, isoelectric focusing did not reveal any separation of the 326E or 326K variants (Figure 74). This could be:

- A true result in the E326K amino acid change does not influence the isoelectric point of NBCn1.
- Alternatively, it should be recalled that the difference in predicted isoelectric point only applies to the full length protein that includes Cassette II. The splicing variant without Cassette II (amino acid 253-375, including E326K), would mean that the non-synonymous polymorphism does not actually alter the protein sequence. Therefore, if the

predominant isoform of NBCn1 in the samples assayed are that of the shorter variant, the results would be expected.

As the antibody is raised against the C-terminus, the antibody itself is unable to differentiate between these isoforms.

5.3. Molecular mechanisms underlying gene transcription

Following the identification of a genotype-associated effect on *SLC4A7* gene transcription in both assessed cell types, and the total cellular NBCn1 expression at least in HUASMCs (Section 5.2), the next set of experiments aimed to determine the underlying molecular genetic mechanisms driving these processes.

5.3.1. Chromatin immunoprecipitation (ChIP) suggested an allele-dependent chromatin interaction at rs13096477

There was no observed statistical significance in the allelic balance of chromatin immunoprecipitation outputs using anti-Pol II antibodies (Figure 75).

There are two main potential explanations for the negative finding.

- The main hypothesis is related to the striking difference in the allelic balance in chromatin-associated DNA inputs (C:T ratio 0.15 ± 0.07), whereas it would have been expected to be closer to 1 in heterozygotes. In context of a marked allelic imbalance in the ChIP inputs, it would be difficult to identify any further imbalances related to the immunoprecipitation. However, this observed difference cannot be adequately interpreted without further experiments due to the chemistry of Sanger sequencing (see Section 2.6.1). With this, the study progressed to the formaldehyde-assisted isolation of regulatory elements (FAIRE) assays (Section 5.3.2).
- The second explanation is based on the use of anti-Pol II antibodies. This methodology is dependent on *active* transcription at the ~400 bp sequence around rs13096477. Binding of transcription factors to DNA may not be consistent but may be transient. This transient nature may reduce the probability of capturing the interaction.

It should also be considered that this may be a true negative result where there is no allelic-dependence on binding of the transcription initiation complex. However, this conflicts with subsequent and stronger evidence, and it appears to be the less likely explanation.

5.3.2. Formaldehyde-assisted isolation of regulatory elements (FAIRE) assays identified the protective (major / T) allele at rs13096477 as preferentially chromatin-bound in HUASMCs but not HUVECs

FAIRE assays at rs13096477 identified the T (major) allele as being preferentially chromatin bound in HUASMCs (Figure 76). These results are consistent with those of the allelic imbalance between genomic DNA and cDNA (Section 4.2.2). In the previous study, the C (minor/risk) allele was associated with increased mRNA/cDNA levels. In turn, the T (major/protective) allele is associated with increased chromatin-binding, suggesting that at baseline, the transcription of *SLC4A7* may be limited by the T allele and its interaction with nuclear proteins. However, it is not possible with FAIRE to determine the protein species that has caused the interaction. This could potentially be addressed by either EMSA (Section 5.3.3) or DNA pull-down assays (Section 5.3.4).

Notably, the FAIRE result is also only positive in HUASMCs, but not HUVECs. The sample size for HUVEC studies should be sufficiently powered (98%) to detect a difference. The potential for this tissue-specific effect could be due to:

- The *potentially* higher baseline expression with lower inter-sample variation in HUVECs (Section 5.2.3) suggests an overall higher probability being in an open chromatin conformation. As the allelic imbalance studies only detects a *relative* difference between the alleles, a near absence of rs13096477 in HUVEC closed chromatin would not show any differences in the alleles. This hypothesis is consistent with the EMSA results (Figure 80, middle panels), and further supported by the lower variability in the HUVEC chromatin sample C:T ratios (standard deviation 0.07) as compared to HUASMC chromatin sample C:T ratios (standard deviation 0.57).
- Alternatively, the results for the HUASMC samples could be a false positive. Although this should always be considered, the consistency of positive results in the same allelic direction for HUASMCs, suggests that this is less likely.

5.3.3. Electrophoretic mobility shift assays (EMSA) indicate allele-dependent in vitro nuclear protein binding at rs13096477, rs2371065 and rs13077400

There was an early need to narrow down the number of SNPs to be investigated, as the study of the blood pressure associated SNP, together with all 92 SNPs in high LD is logistically near impossible, particularly this means assessing 186 different alleles. Furthermore, with multiple testing of 93 independent SNPs, the risk of a false positive is markedly increased (or the need to conduct large number of replicates in view of a lower Bonferroni corrected significance threshold).

The prioritization of the 10 SNPs (Figure 77) to investigate was conducted in a stepwise manner. The first priority was the GWAS associated SNP followed by SNPs in the classical positions of TFBS with predicted differences in transcription factor binding (Table 31). This is followed by prediction tools, with ENCODE and RegulomeDB. Of course, as indicated in Section 3.13, there are limitations to these prediction tools particularly with the paucity of input data, notably with 40% (37/93) of the SNPs of interest having no data within RegulomeDB. However, this is still the only information available to aid the decision. With that, the pragmatic view would be to accept these limitations, and accept that the true (if any) causative SNP(s) may lay outside of the 10 investigated SNPs.

Of the 10 SNPs investigated in the panel (Figure 78), 3 SNPs (intronic SNP rs2371065, synonymous exonic SNPs rs13096477 and rs13077400) demonstrated electrophoretic mobility shifts, interacting with HUASMC nuclear extracts. Notably, in a paired study, nuclear extracts from HUVECs did not retard the electrophoretic mobilities of those oligonucleotides (Figure 79). This can be interpreted as consistent with the previous qRT-PCR results (Section 5.2.3). As *SLC4A7* mRNA appears to be more prevalent in HUVECs as compared to HUASMCs, this might suggest their respective nuclear extracts may have different levels of the proteins responsible for the DNA-nuclear protein interaction. This interpretation should be taken with the caveats of being an *in vitro* experiment and does not account for potential epigenetic modifications.

The three SNPs (rs2371065, rs13096477 and rs13077400) were further investigated to confirm allelic-dependent oligonucleotide-protein interactions (Figure 80). The result for rs13096477 emphasises the consistency of results across methodologies where the protective major (T) allele preferentially binds nuclear extracts as compared to the risk (C) allele, just as demonstrated in the FAIRE studies (Section 4.3.2).

Having identified a DNA-nuclear protein interaction, it would be interesting to identify the protein responsible for the observed shift. With this, competitor studies were performed. Unlabeled competitors designed to reflect the consensus binding sequences of transcription factors predicted to bind the sequence on and around rs2371065 (Table 46) suggested that PHOX2A, PLAGL1, TFAP2C and TRMT1 may be good candidates by significantly reducing the shifted signal (Figure 81). To take this further, supershift assays were attempted with antibodies targeted

against PHOX2A, PLAGL1 and TFAP2C. TRMT1 was not assessed as it functions primarily to bind tRNA. However, there were no supershifts noted in these experiments (Figure 82). This was confirmed by an absence of shift in the WEMSA (Figure 83). This failure to identify the responsible protein led to DNA pull-down assays (see Section 5.3.4).

5.3.4. DNA pull-down showed *in vitro* nuclear protein interaction with sequence centered on rs2371065, but unable to reveal identity of protein

A DNA pull-down assay with biotin-labelled double stranded oligonucleotides representing the 21 bp sequence centered on rs2371065 was successful in obtaining two specific protein bands as identified by Ponceau S staining (Figure 85). The specificity of the bands is supported by their absence in the final wash and the negative control. The presence of two bands with the DNA pull-down reflects the findings of two shifted bands in EMSAs (Figure 78). Just like EMSAs, the results have to be taken with the caveat that it is an *in vitro* system that may not always reflect an *in vivo* cellular environment.

With this colloidal stain's lower detection limit of 250 ng, the bands were of a sufficient quantity for commercial protein sequencing (http://www.altabioscience.com/media/repo/Protein_sequencing_Technical_Brochure.pdf). The commercial service (AltaBioscience, UK) was unable to return a protein sequence read, not even position 1. The most likely cause is N-terminal blockage, which is present in ~50% of eukaryotic proteins as a result of glycosylation or pyroglu-cyclisation. There is no method of determining this prior to sequencing. Other potential causes of altered protein structure during processing that impedes protein sequencing chemistry include processing at elevated pH (>9.0), inferior grade reagents and water, exposure to elevated temperatures if glutamine is the N-terminal residue, protease inhibitors interacting with terminal amino groups or the use of formic acid or urea; all of which were known to be present.

Potential remedies include fragmenting the protein into smaller components, but this makes the interpretation of the sequence more difficult or requires HPLC (high-performance liquid chromatography) purification, which in turn increases the initial protein yield required. Acidification of the protein sample may (but rarely) remove the blockage. However, it should be considered that these *in vitro* interactions may actually be "false positives" that did not reflect *in*

vivo biology. As these potential remedies are logistically complicated, expensive and still unlikely to yield useful results, DNA pull-down was not further pursued.

5.4. Impact of non-synonymous SNP on protein behavior and function

Although Section 5.3 has already discussed the impact of *SLC4A7* genotype on gene expression, it should be recalled that the GWAS lead SNP rs13082711 is in high LD with a non-synonymous SNP (rs3755652). This section explores the potential impact of the E326K (glutamic acid to lysine at position 326) amino acid change.

5.4.1. *Calcineurin A did not co-immunoprecipitate with NBCn1 in HUASMCs*

There has been previously reported interactions between NBCn1 and calcineurin A in mice (Danielsen *et al.*, 2013) and experimental *in vitro* models (Gill *et al.*, 2014), but not yet with human NBCn1. There is also an increasing body of evidence, albeit again in animal models, that those co-transporters that regulate pH_i can be modulated by kinases and phosphatases (Table 2). Coupled with the presence of postulated calcineurin binding sites close to the E326K amino acid change (Danielsen *et al.*, 2013), it was important to consider whether there is an interaction between calcineurin A and NBCn1. Co-immunoprecipitation would show only a temporal and spatial physical relationship, whereas a physiological relationship, as assayed by pharmacological inhibition will be examined later (Section 5.6.2).

There was an absence of co-immunoprecipitation of NBCn1 and calcineurin A together in transfected A10 cells as well as HUAMSC primary cultures (Figure 87). This was initially surprising as application of the same protocol for murine NBCn1 as conducted by Danielsen *et al* (2013) yielded positive results. However, there are several potential explanations for this:

- This could be a true negative result.
 - The lack of consistency with other studies could potentially be explained by the difference in human and mouse NBCn1 amino acid sequences (Figure 5). It may well be that the relevance of calcineurin is lower in human NBCn1 than murine NBCn1.
 - Another explanation that accepts the lack of interaction as a true negative relates to the *in vitro* nature of this study. Positive co-immunoprecipitation in murine arteries (Danielsen *et al.*, 2013) was performed in freshly dissected vessels, and under conditions that elevated intracellular calcium concentrations. An *in vitro* primary cell culture may not be conducive to this interaction. This is a hypothesis that is backed by

the finding that different cell culture surfaces significantly impacted on NBCn1 co-transport activity (Figure 51).

- Alternatively, this result could relate to the cell type used. HUASMCs, being from embryonic, conduit arteries that carry deoxygenated blood, do not reflect the adult resistance arteries, as previously expanded when discussing the limitations of the cell type in Section 2.1.1. Additionally, it may be possible that the combination of the transfected human NBCn1 and the endogenous murine (A10 cells) calcineurin does not interact across species.
- In view of the absence of a positive co-immunoprecipitation, this could be a false negative result.
 - It could be that the interaction between calcineurin A and NBCn1 is transient, and therefore it would be difficult to capture the interaction.
 - On a similar note, the detection limit of protein immunoblots may not be sensitive enough to detect low levels of co-immunoprecipitation. Theoretically, a higher sensitivity detection system such as ELISA (enzyme-linked immunosorbent assay), HPLC (high-performance liquid chromatography) or mass spectrometry may be an alternative assay.
 - There is still a possibility that the co-immunoprecipitation conditions that were successful for murine NBCn1 are not optimized for human NBCn1.

This system was interrogated again in a cell physiology setting with tacrolimus, a calcineurin inhibitor, and its impact on NBCn1 $\text{Na}^+/\text{HCO}_3^-$ transport function (Section 5.6.2).

5.4.2. Immunocytochemistry revealed NBCn1 position both on the nuclear membrane, and also intracellularly

Immunocytochemistry studies of HUVECs and HUASMCs were stained for NBCn1. This was conducted together with VE-cadherin as a membrane marker for HUVECs, but as a preliminary study, a membrane marker such as N-cadherin for HUASMCs was not used. The immunocytochemistry on HUVECs demonstrates a significant proportion of NBCn1 signal that does not appear to be membrane-bound (Figure 45, Figure 88 and Figure 89). Based on appearance, this may be representative of Golgi or endoplasmic reticulum, but cannot be determined without additional organelle-specific markers. Caveats that have been expressed in previous sections on

(1) *in vitro* preparations, (2) embryonic conduit vessels and (3) the importance of cell culture surfaces should also be considered in the context of these “negative” results.

The appearance of NBCn1 intracellularly may initially appear counterintuitive, but reflects previous literature where it is detected both along the plasma membrane, as well as intracellularly, even if not clearly localised to any particular organelle (Lauritzen *et al.*, 2010; Namkoong *et al.*, 2015).

Although there are well recognised methods of quantifying membrane vs. non-membrane proteins (such as in Figure 88), there is the possibility of bias. In the absence of a categorical (yes/no) difference between the genotypes, any difference would be on a continuous scale. This means selection of cells to analyze (out of tens of thousands on a slide) and the direction of the slice (infinite possibilities) are subject to selection bias. With the difficulties in quantification, the attempt to determine differences in cellular localization, in particular the proportion of NBCn1 that appears membrane-bound, prompted the use of subcellular fractionation by differential centrifugation method (Section 5.4.3).

5.4.3. Subcellular protein fractionation identified protein variant (326E) associated with the protective (major / T) allele at rs13096477 as preferentially abundant in the membrane fraction

To reduce the impact of inter-experiment variation, immunoblots were run in pairs (Lys/Glu, G/A at rs3755652 and Glu/Glu, A/A at rs3755652), and if possible as two pairs on one blot. Due to the limited number of rare homozygotes (Lys/Lys, G/G at rs3755652), they were not used for this set of studies. The cadherins (N- and VE-) were used as a loading control as their relative expressions should reflect the quantity of membrane proteins loaded. Of note, there was no difference in total cellular expression of VE-cadherin comparing 7 G/A and 7 A/A HUVEC samples (normalized ratio G/A 0.97 ± 0.08 , A/A 1.03 ± 0.05 ; $P > 0.05$), supporting its suitability as a loading control.

Heterozygotes HUASMCs and HUVECs carrying the risk variant (Lys/Gly) had a higher NBCn1 signal in the “membrane” fraction as compared to the homozygotes for the protective variant (Gly/Gly) (Figure 90). This was despite a similar NBCn1 signal in the “organelle” fraction. The statistically significant difference remains even when taking a ratio of “membrane” : “organelle” fractions. These results are consistent with the hypothesis that the risk allele, has a higher level of NBCn1 present at the *expected site of action* – the plasma membrane; not just higher *overall* protein levels.

It is interesting to note that HUVECs have a higher expression of NBCn1 in the “membrane fraction” for this set of studies, but not as previously detected on a panel of total cellular protein (Figure 73). This perhaps relates to the relative presence of NBCn1 in each compartment. Anecdotally, a confluent 25 cm² cell culture flask provides on average 5 ng of “membrane fraction” proteins, 20 ng “organelle fraction” proteins and 200 ng “cytosolic fraction” proteins. As 2 ng is loaded for immunoblots of each fraction, this accounts for 40%, 10% and 1% of the fractions respectively. Considering this complexity, it would be safe to conclude that the cells carrying the risk allele has more abundance of NBCn1 at the cell membrane, but not enough evidence to suggest that it affects trafficking to the cell membrane. If there is a subtle effect on trafficking to be detected, perhaps protein quantification using membrane biotinylation would be the next step to pursue. However, as the probability that NBCn1 trafficking is the mechanism whereby the GWAS-associated polymorphism exerts its effect appears low, it has not been prioritized.

5.4.4. Transfection of overexpression plasmids suggests that the E326K NBCn1 variation is unlikely to impact on NBCn1 degradation

In *SLC4A7* overexpression studies in A10 mouse aortic smooth muscle cells, there was a consistent increase in total cellular NBCn1 expression of either 326E or 326K variants when compared with control vector, but with no difference between the two variants (Figure 91). This suggests that over the 48 hour period, there is no difference in protein production or degradation.

This provides some evidence that the E326K amino acid change may not play a role in total cellular protein levels. However, it should be emphasized that experimental confirmation of equal mRNA translation and protein degradation of the two variants would require further investigation with combination of mRNA translation inhibitors (e.g. cyclohexamide) and proteasome inhibitors (e.g. MG132).

5.5. Impact of *SLC4A7* genotype on inducibility

Consistent with the findings in other acid-base regulators, such as NHE1 and NHE3 (see Section 1.12), angiotensin II induced NBCn1 expression in vascular smooth muscle (Figure 92a). This also reflects the previous findings for NBCn1 induction in cardiac and renal cells, both *in vivo* and *in*

vitro. Of note, the concentration of 10^{-8} M used was at a physiologically relevant level (Bergeron *et al.*, 2001).

The allelic imbalance method was used due to the previously discussed marked inter-sample variability in HUASMCs (Section 5.2.3), which showed an allele-dependent effect of AngII-related upregulation. The allelic imbalance observed at baseline returns to parity upon induction (Figure 92b). This suggests that the C (minor) allele being more transcribed at baseline, indicating constitutive transcription, whereas expression of the T (major) allele may be inhibited at rest, but could be induced upon stimulation. This interpretation can also be consistent with the findings of the FAIRE study (Section 5.3.2), where the major allele is observed to be preferentially chromatin-bound.

5.6. Impact of *SLC4A7* genotype on cell function

5.6.1. Collagen gel contractility assays – lacks the rapid response required and is unable to withstand the necessary pH range

The initial optimization of the collagen gel contractility was successful, with the cell density and time course (Figure 94), replicability (Figure 95), inter-assay (Figure 95), intra-observer (Figure 96) and inter-observer (Figure 97) variability established for primary HUASMC contractility gels.

However, three main issues have limited the utility of this assay as an indicator of vascular smooth muscle cell function in terms of contractility, which would have relevance to blood pressure regulation. The time course to achieve maximal contraction in such as set up is near 72 hours (Figure 94), which is not reflective of the rapid recovery of intracellular acidosis by NCBn1 (in magnitudes of minutes, see Figure 2) and other acid-base regulating transporters and channels. This delayed contraction is likely to mask any effect of genotype on pH_i and vascular smooth muscle contractility.

A separate concern is that the collagen gel is not resistant to lower pH values. In the methodology, the gel polymerises at a pH higher than 8.0 (see Section 2.17.1), and conversely, acidification is the common method of dissolving collagen from rat tails (Rajan *et al.*, 2006). This is reflected by fact that the exposure of the gels to a media of pH 6.8 resulted in blurred gel borders (Figure 99) which

was difficult to quantify, and more importantly unable to guarantee a connective tissue matrix for the vascular smooth muscles to anchor and produce a quantifiable contraction of the gel.

The final caveat, as previously expressed for HUASMCs, is that there is significant inter-sample variability in terms of NBCn1 expression, perhaps reflecting state in cell cycle or cellular phenotype of each studies sample. This would markedly increase the overall numbers that are required to detect potential genotype-associated differences.

With these three concerns, it appears that collagen gel contractility assays may not provide the quality of evidence required to identify any differences in cell function based on *SLC4A7* genotype. In the subsequent studies, cellular function was predominantly assessed with pH_i recovery following ammonium chloride pre-pulse.

5.6.2. Na^+/HCO_3^- -dependent DMA-insensitive pH_i recovery from intracellular acidosis

Thus far, the data presented shows the *SLC4A7* genotype at rs13082711 impacting on gene expression, where the risk allele is associated with increased NBCn1 expression in terms of mRNA and protein, and subsequently availability at the plasma membrane. This section aims to explore whether this translates to differences in protein function in terms of Na^+/HCO_3^- -dependent, DMA-insensitive pH_i recovery following ammonium-induced intracellular acidosis. This would reveal NBC activity, although it should be recognised that it would not differentiate between NBCn1 and NBCE1 or even other sodium-bicarbonate co-transporters. Overall, there is a consistent trend where the risk allele, associated with higher NBCn1 expression, demonstrates increased protein activity as defined by Na^+/HCO_3^- -dependent, DMA-insensitive net base uptake and final plateau pH_i . This appears to be independent of the co-inherited non-synonymous polymorphism rs3755652 and the subsequent E326K amino acid change. The differing strengths of evidence may be dependent on relative protein expression levels of the three investigated cell lines.

HUASMCs - The protective (major / T) allele at rs13096477 is associated with greater Na^+/HCO_3^- -dependent, DMA-insensitive net base uptake and resting pH_i

Based purely on the Na^+/HCO_3^- -dependent, DMA-insensitive pH_i recovery curves (Figure 100) and initial net base uptake (Figure 102), there was no statistically significant differences between the two genotypes. However, there is a divergence in the curves that merited further analysis. It

became apparent that there was significant inter-sample variability, which led to a large range of pH_i start points from which recovery was measured (Figure 103). This is a potential confounder as net base uptake is pH_i -dependent. When taking this into account, it showed that rs13082711 genotype was likely to be associated with NBC activity, where the higher-expressing risk allele is associated with a steeper slope when plotting net base uptake against each pH_i (Figure 104b), although they were not statistically different at each pH_i , and final resting pH_i (Figure 105).

Serum-starvation of 48 hours, intending to reflect the expression levels of quiescent cells resulted in lower protein expression (Figure 131, $P=0.15$, $n=16$ pairs). As hypothesised, this also reduced the variability potentially introduced by having cells in different phases of the cell cycle. With the lower expression levels, this reduced $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery curves (Figure 101a), although once again, no statistically difference at each pH_i , and final resting pH_i (Figure 105). In this context of quiescent HUASMCs, the genotypic effect of rs13082711 persists, where the higher-expressing risk allele continues to be associated a steeper $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake relationship to pH_i (Figure 104c), and final resting pH_i (Figure 105).

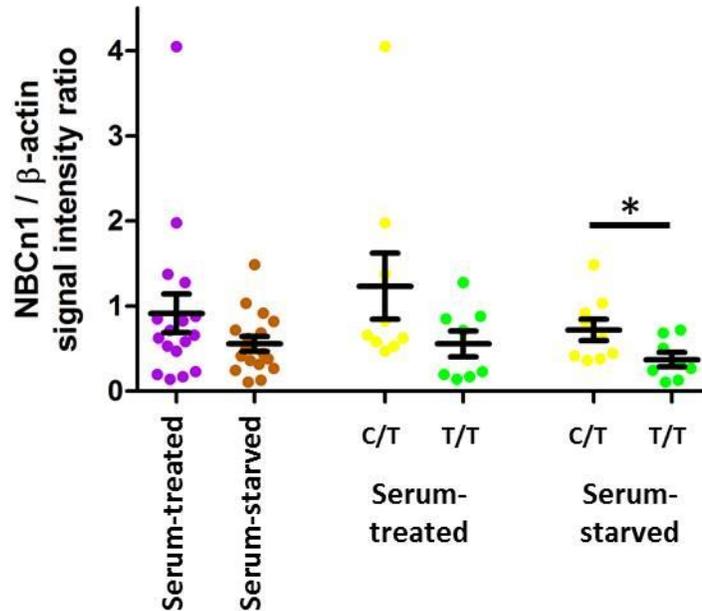


Figure 131: NBCn1 protein expression in serum-treated and -starved HUASMC samples undergoing pH_i studies

Serum-treated and -starved, $n=17$ pairs; rs13082711 genotype C/T, $n=9$; T/T, $n=8$.

* $P<0.05$

Following completion of the studies, it became apparent that the HUASMCs could be classified into two different morphological categories (Figure 108). In a post-hoc analysis, it was seen that the confluent subgroup was associated with higher NHE and NBC activity. As the confluent subgroup is more likely to be reflective of healthier and more proliferative cells, with a morphology that appears more “contractile”, this subgroup was further assessed. The rs13082711 genotype-associated differences in $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake and final resting pH_i (

Figure 110) was even greater in this subgroup. However, it is interesting that despite the differences in $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake, there does not appear to be an influence of the morphological groups on total cellular NBCn1 protein expression by immunoblots (Figure 132). This may be related to the overall surface-to-volume differences in the two subgroups.

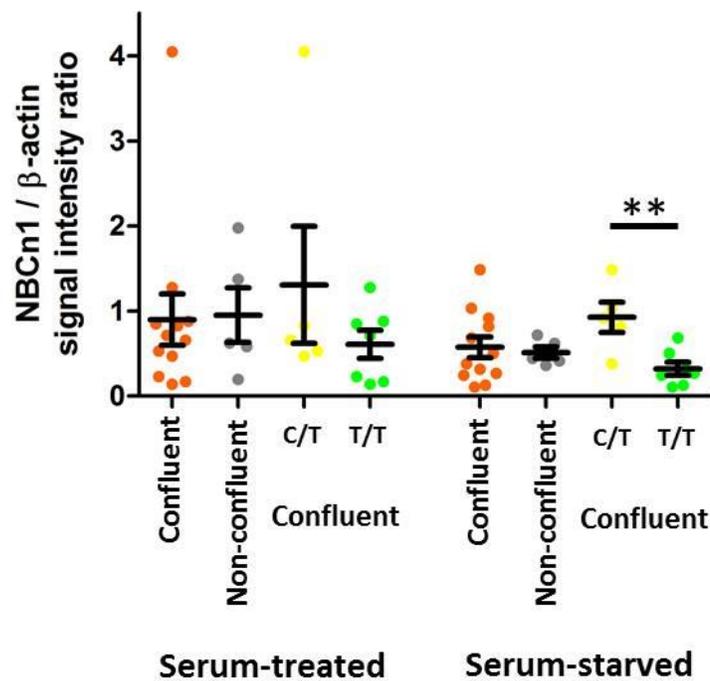


Figure 132: NBCn1 protein expression in serum-treated and -starved HUASMC samples undergoing pH_i studies
 Confluent, $n=12$; non-confluent $n=5$; confluent subgroup rs13082711 genotype C/T, $n=5$; T/T, $n=7$.
 Paired samples for serum-treated and -starved.
 ** $P < 0.05$

It is worth noting that although *SLC4A7* genotype and its associated variation on expression levels may influence $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake and final resting pH_i , it

does not have an impact on the cell's buffering capacities. This is not unexpected, as buffering capacities were not different between *SLC4A7* knockout mice lacking NBCn1 as compared to wildtype controls; both in the presence and absence of CO₂/HCO₃⁻ (Boedtkjer *et al.*, 2011). Thus, the genotypic effect on NBC activity is not related to changes in buffering capacity.

HUASMCs – calcineurin inhibition (by FK506) did not alter net base uptake, whereas intracellular calcium chelation (by BAPTA-AM) decreased net base uptake and resting pH_i, but was not dependent on rs13096477 genotype

As previously discussed in Section 1.9, there has been reported physical interaction between calcineurin A together with Cassette II of NBCn1 (Danielsent *et al.*, 2013; Gill *et al.*, 2014). In turn, calcineurin inhibitors were able to attenuate the expected activity of NBCn1 in rat mesenteric arteries, after either depolarization or norepinephrine-induced contractions (Danielsen *et al.*, 2013). The hypothesis that calcineurin, and intracellular Ca²⁺, may have a role in NBCn1 function of HUASMCs and HUVECs was also tested here. In 12 matched experiments (rs13082711 genotype: C/T n=4, T/T n=8), cells from four concurrent cultures of the same cell lineage were assessed for Na⁺/HCO₃⁻-dependent, DMA-insensitive pH_i recovery following ammonium chloride prepulse-induced intracellular acidosis while being exposed to (A) DMSO vehicle, (B) 10 μM FK506, (C) 30 μM BAPTA-AM loading, or (D) both FK506 and BAPTA-AM.

Na⁺/HCO₃⁻-dependent, DMA-insensitive pH_i recovery and net base uptake was not changed in the presence of 10 μM FK506 (

Figure 111b and Figure 112). Considering how the “vehicle” and “FK” curves practically overlap, the lack of a difference detected is unlikely to be a false negative. This is also true when stratified into *SLC4A7* genotype subgroups (Figure 114). Potential reasons for a lack of an effect could be that:

1. Calcineurin and NBC proteins can interact, but the 10 μM exposure of FK506 is too low. Differing cell preparations (*ex vivo* vessels vs. *in vitro* cell culture) can feasibly have differing IC₅₀s, but this is perhaps a less likely explanation. However, it would be typical to expect higher sensitivities in cell cultures as compared to intact vessels due to difficulties in penetrating the layers of intact vessels, combined with potential local degradation. If there were more opportunities, perhaps a dose-range study could be conducted, but however this will massively expand the workload for unlikely gain.
2. Calcineurin and NBC proteins can interact, but calcineurin has already exerted its phosphatase activity *before* the introduction of FK506. In other words, NBCn1 may already be in a

dephosphorylated state, where inhibition of calcineurin A phosphatase activity does not change the phosphorylation status. It is for instance possible that the culture conditions alter the expression or activity of counteracting kinases and therefore changes the phosphorylation state even under resting conditions.

3. Calcineurin and NBC proteins can interact, but the intracellular conditions (e.g. intracellular Ca^{2+}) may be insufficient to appropriately activate calcineurin A. It is worth recollecting the study by Danielsen *et al.* (2013) assessed the impact of FK506 in a high- K^+ -depolarized state which activates Ca^{2+} influx via voltage-gated Ca^{2+} -channels. If there were more opportunities, the cells could be exposed to an initial high- K^+ environment. However with a preliminary study of n=4 pairs, exposure to a 60 mM K^+ (replacing Na^+) throughout the study did not notably alter the pH_i recovery of HUASMCs after ammonium-prepulse washout.

This study is unable to differentiate between the three hypotheses above, and any of the three may be potential explanations, either on their own or in combination. Nonetheless, our data show that NBCn1 activity is possibly increased in VSMCs carrying the BP risk allele also after inhibition of calcineurin activity (Figure 133).

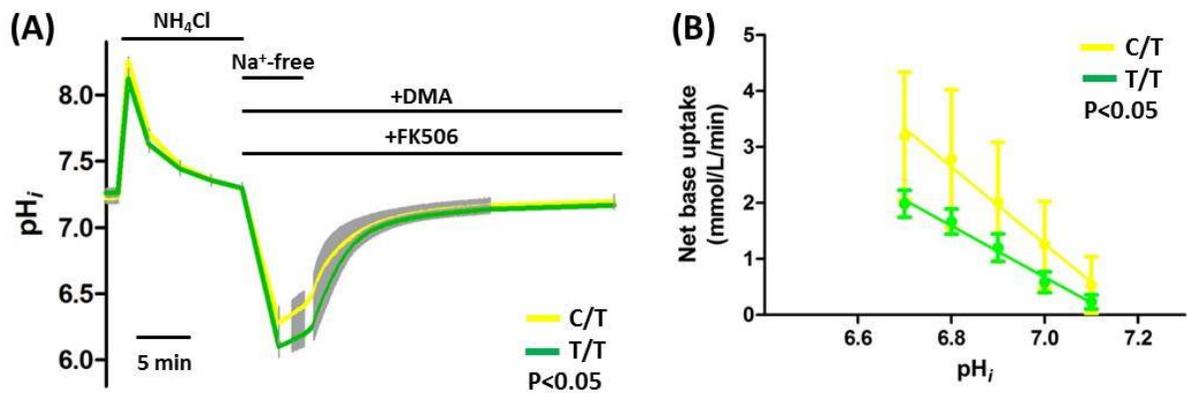


Figure 133: HUASMC pH_i recovery following intracellular acidosis, with $\text{CO}_2/\text{HCO}_3^-$, and in the presence of DMA and FK506

rs13082711 genotypes: C/T (yellow, n=4) and T/T (green, n=7). Grey error bars denotes SEM.

Groups in (A) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B) compared by linear regression.

The impact of intracellular Ca^{2+} chelation was however much clearer. 30 μM BAPTA-AM loading for 20 minutes significantly reduced $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery and net base uptake of HUASMCs (

Figure 111d and Figure 112). The BAPTA-AM-attenuated NBC activity in HUASMCs reflects the findings in rat mesenteric arteries where the depolarisation (50 mM KCl)-stimulated NBC activity was greatly attenuated by BAPTA-AM (Danielsen *et al.*, 2013). It should be noted that the Ca^{2+} chelation was deemed to be successful with this specific concentration and duration of initial loading, based on the marked attenuation of the Ca^{2+} -spike of intracellular acidosis following washout of ammonium (Figure 49). In the context of markedly reduced $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery, it is unsurprising that 30 μM BAPTA-AM also decreased final resting pH_i (Figure 113a). However, this is in the absence of a significant change in buffering capacity (Vehicle 16.7 ± 1.0 mmol/L, BAPTA 15.7 ± 1.0 mmol/L, difference 1.0 ± 1.3 mmol/L, $n = 12$ pairs).

While these results may support the hypothesis that $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery was reduced by directly influencing Ca^{2+} -dependent signaling pathways, it was apparent that the morphology of the HUASMCs changes in a manner that is not noted with the vehicle control. Over the course of the 45 minute recordings, the cells begin to detach from the cell culture surface and appear to shorten, particularly after the washout of ammonium. Therefore, the change in $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery is likely to be relating to a change in overall cellular phenotype.

HUVECs - The protective (major / T) allele at rs13096477 is associated with greater $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake

The findings from the studies on HUVECs and their $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive recovery from intracellular acidosis largely reflects the observations made with HUASMCs, which provides reassuring consistency.

As with HUASMCs, based purely on the $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery curves (Figure 115b), there were no statistically significant differences between the two genotypes. However, there is a divergence in the curves that merited further analysis. Despite the large inter-sample variability in pH_i at the first recording when Na^+ -containing buffer is reintroduced, there is an increased $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake carried by the risk allele (Figure 116). This is reiterated by showing that rs13082711 genotype was associated with NBC activity, where the risk allele is associated with stonger relationship between $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake and pH_i (Figure 118), and final resting pH_i

(

Figure 119, albeit non-statistically significant $P=0.10$). The consistency of direction and magnitude of effect on final resting pH_i opens up the possibility of a false negative. Despite the same magnitude of difference and similar sample numbers, the larger standard deviation from the HUVEC samples as compared to HUASMCs (Figure 124, Table 48), results in the study having only a statistical power of 58%, and the needs $n=30$ for each genotype to reach a power of 80%.

As with HUASMCs, *SLC4A7* genotype did not influence the cell's buffering capacities, both in the presence and absence of CO_2/HCO_3^- , which is unsurprising given the knowledge gained from the aforementioned knockout mouse model (Boedtkjer *et al.*, 2011). Thus, the genotypic effect on NBC activity is not related to changes in buffering capacity.

Unlike HUASMCs, the assessment of HUVECs in the thesis did not include serum-starved cells, stratification by cellular morphology or use of calcineurin inhibition. This is due to HUVECs:

- not tolerating serum-starvation for any significant duration of time (Figure 93),
- not revealing any notable difference in morphology, and
- not regarded as having a large contribution of calcineurin signaling to its cellular function.

HUASMCs vs HUVECs - HUASMCs have a comparable contribution from NBCs and NHEs to pH_i recovery from intracellular acidosis, but HUVECs with a much faster recovery has a dominant NHE activity

As mentioned in Sections 1.3 to 1.8, NBCn1 is widely expressed in a range of tissues. The work for this thesis has predominantly explored vascular smooth muscle cells and endothelial cells. As there appears to be opposing effects of these cell types on blood pressure as demonstrated by the *SLC4A7* knockout mouse (Boedtkjer *et al.*, 2011), this section attempts to highlight the similarities and differences between the two cell types.

From Figure 122, it is apparent that HUVECs have a faster recovery from intracellular acidosis as compared to HUASMCs. This is in addition to a higher final plateau pH_i (Figure 124). The absence of a statistically significant difference in the final plateau pH_i of the two cell types in CO_2/HCO_3^- buffer with DMA treatment (predominantly isolating the NBC component) is likely to be due to an underpowered sample size (power 26%, requiring a $n=89$ each group sample size to reach a power of 80% with an α -value of 5%). The Na^+ -independent net base uptake in HUVECs is almost 4 times

larger than that of HUASMCs. This may reflect the activity of Na^+ -independent pH_i regulators such as monocarboxylate/ H^+ co-transporter or H^+ -ATPase, but was not within the scope for this thesis.

Taken together, this indicates that HUVECs have higher overall ability to recover from intracellular acidosis (Figure 122a), which comprises mainly of higher activity for both NBCs (Figure 122b) and NHEs (Figure 122c). In the absence of $\text{CO}_2/\text{HCO}_3^-$ (inhibiting NBCs) and the presence of DMA (inhibiting NHEs), the pH_i recovery was slow and similar in the two cell types, suggesting that non-NBC and non-NHE components play a very small role in pH_i regulation.

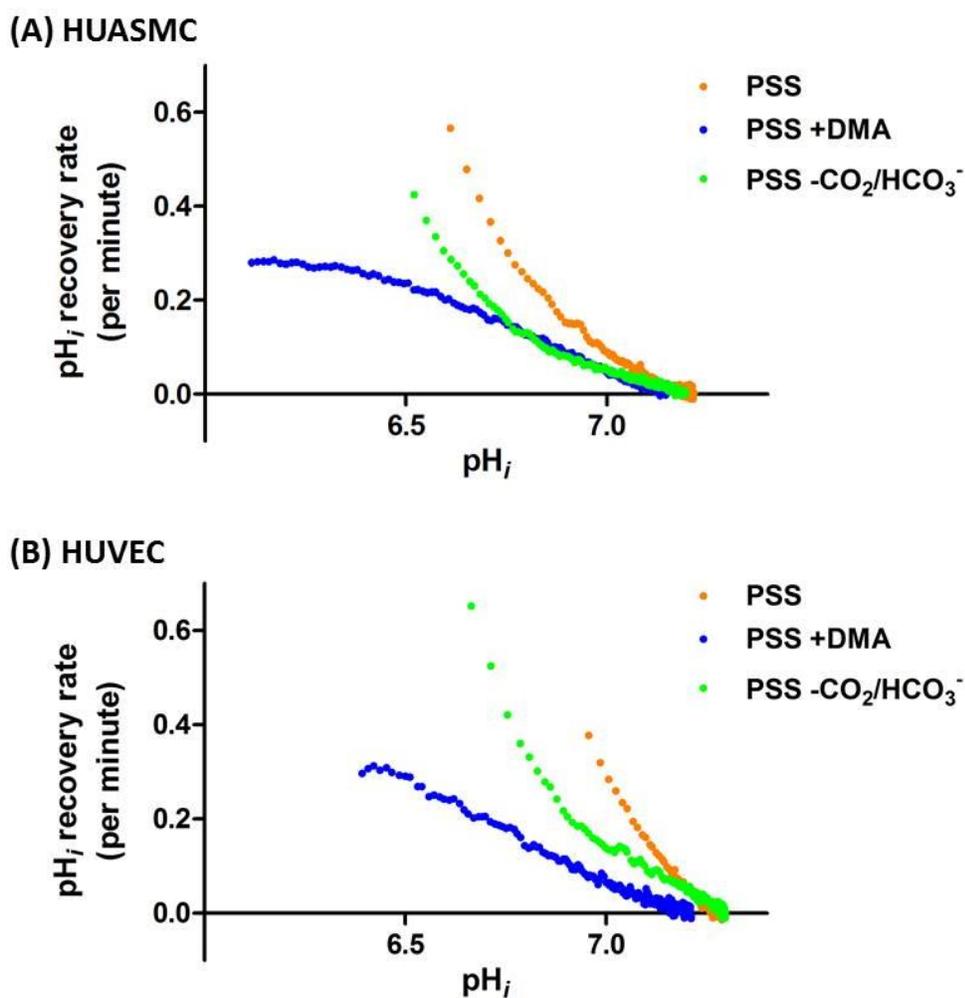


Figure 134: pH_i recovery rates of HUASMCs and HUVECs in different conditions. Each data point reflects one of the 290 timepoints, plotted for the average pH_i and recovery rate. Orange, PSS; blue, PSS + DMA; green, PSS without $\text{CO}_2/\text{HCO}_3^-$.

Although not easily quantifiable, it appears that the relative contributions of NHE and NBC may be closer to each other in HUASMCs as compared to HUVECs, especially in the pH_i ranges from around 6.7 upwards (

Figure 134). Also of note, there is less variability in the HUASMCs as compared to HUVECs in three key Na^+/HCO_3^- -dependent, DMA-insensitive characteristics assessed (Table 48). Taken in combination, the higher overall contribution and lower population variability may mean that any genotypic differences in NBCn1 activity in HUASMCs could play a larger role in the overall pH_i regulation as compared to HUVECs.

These results should be taken in relation to expression of other pH_i regulators (Figure 135). This end-point RT-PCR panel of HUASMC and HUVEC samples paired from the same donor, shows inter-sample variability is present not just for *SLC4A7* short and long isoforms (confirming the previous qRT-PCR and immunoblot results), but also *SLC4A4* (NBCe1) and *SLC9A1* (NHE1). Accepting that this is not qRT-PCR or immunoblot results, it is apparent that in these paired samples the overall signal for *SLC4A7* and *SLC9A1* is stronger in HUVECs compared to HUASMCs. Notably, there is minimal signal for *SLC4A4* in HUVECs.

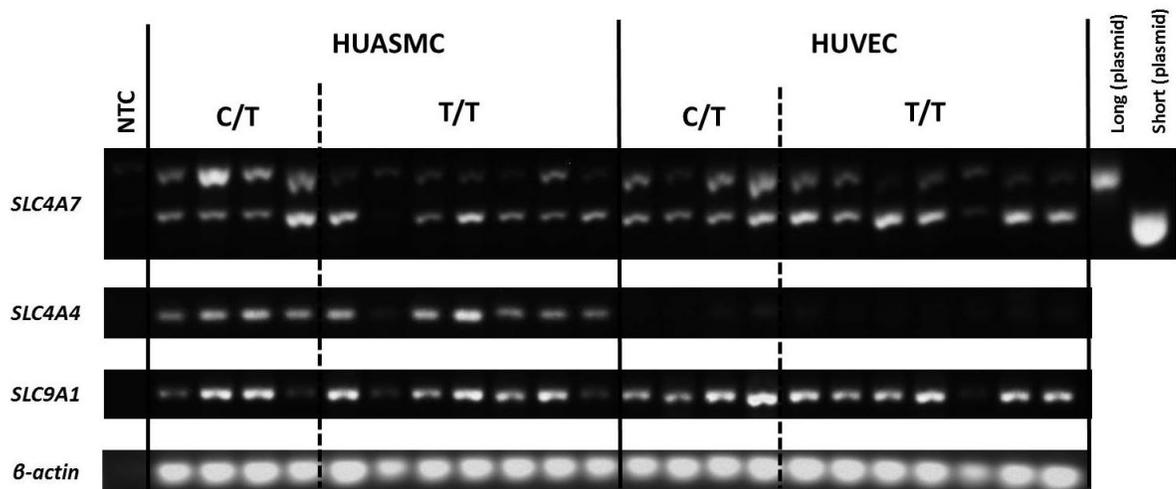


Figure 135: End-point RT-PCR of selected pH_i regulators
 End-point RT-PCR (10 ng of reverse transcribed RNA) for the long and short variants of *SLC4A7*, together with other pH_i regulators *SLC4A4* (NBCe1) and *SLC9A1* (NHE1). β -actin included as positive control. 0.1 ng of plasmids for the full length and variant without Cassette II was also underwent PCR for *SLC4A7* as markers of product size.

It is worth recollecting the $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive pH_i recovery and baseline would be composed of the actions of all NBCs (with likely minor contribution from non-NHE co-transporters), that also includes *SLC4A4/NBCE1*. Notably, NBCn1 KO appears to abolish all $\text{Na}^+/\text{HCO}_3^-$ co-transport. Additionally, despite the observation of mRNA expression of multiple SLC4-family $\text{Na}^+/\text{HCO}_3^-$ co-transporters has also been identified in mouse carotid arteries, NBCn1 has been found to functionally dominate net acid extrusion *ex vivo* (Boedtkjer *et al.*, 2016). However, after culture by explants, there is a remaining 40% NBC-activity in NBCn1 KO cells, suggesting that other NBCs may play a functional role in cell culture conditions (Boedtkjer, unpublished data). This means that the genotype-effect on HUASMC $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake and end pH_i is potentially to be *in addition to* baseline NBCE1 contribution. There is currently no small molecule pharmacological inhibition of the different NBC isoforms available.

A10 mouse aortic smooth muscle cells - *SLC4A7* overexpression increased $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive net base uptake, but is not related to the E326K variation or the presence/absence of Cassette II

Overexpression models were used to test two main research questions – (1) whether increased *SLC4A7/NBCn1* expression affects $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive baseline pH_i and recovery from intracellular acidosis, and (2) whether the non-synonymous polymorphism and the subsequent E326K amino acid change or the splice variant alters NBCn1 function, which is difficult to assess in the primary cell lines as there is a genotype-associated difference in gene expression. In summary, there is an effect of expression on $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive baseline pH_i and recovery, but *independent* of the E326K or splice variations.

From

Figure 125, there is a clear difference between the overexpression plasmids (326E, 326K and w/o Cassette II) and vector, although there is no statistically significant difference between the three overexpression plasmids. This is further supported by assessing the net base uptake, both at the initial phase, or at fixed pH_i intervals (

Figure 126). In concordance with the increased $\text{Na}^+/\text{HCO}_3^-$ -dependent, DMA-insensitive base uptake in the three overexpression plasmids, there is also an associated increased in baseline pH_i (Figure 127). The resting pH_i for the short splice variant is difficult to interpret. It occupies an intermediate level between vector and the two long variants, but not statistically significant from

each other. However, when considering additional *SLC4A7* transfection studies in the next section (total n=10 sets of parallel transfections), there is no difference between the three overexpression plasmids (end pH_i: 326K 7.247 ± 0.020, 326E 7.252 ± 0.016, w/o Cassette II 7.227 ± 0.015).

As with the primary cell line studies, NBCn1 expression levels and E326K variation did not influence buffering capacities.

Throughout these overexpression studies, all three plasmids showed similar levels of overexpression on immunoblot signals (Figure 128). Additionally, there appears to be equal efficiency in uptake of plasmids, or at least as identified by qRT-PCR (Figure 42). This suggests findings that the three variants are similar to each other in Na⁺/HCO₃⁻-dependent, DMA-insensitive net base uptake and resting pH_i are true negatives, and not confounded by differences in expression levels. However, a caveat to consider is the possibility of overexpression studies saturating the system, which disguises any potential subtle differences between the variants.

A10 mouse aortic smooth muscle cells - Calcineurin inhibition (by FK506) did not alter net base uptake in *SLC4A7* overexpressed cells, regardless of E326K variation or the presence/absence of Cassette II

Similar to that of the findings with HUASMC primary cultures, Na⁺/HCO₃⁻-dependent, DMA-insensitive pH_i recovery and net base uptake was not changed in the presence of 10 μM FK506 (Figure 129). Potential reasons for a lack of an effect are similar to that explored earlier in this section, with the additional factor that in this setting it relies on interaction between human NBCn1 and rat calcineurin A. A potential method of overcoming this is to co-transfect another plasmid that overexpresses human calcineurin A, but this is likely to introduce even more noise in to the system, and makes detecting a signal even more difficult.

When assessing the final plateau pH_i, it is interesting to note that there is an overall pattern (by repeated-measures ANOVA) where calcineurin inhibition lowers plateau pH_i in both variants with Cassette II but not with the plasmid that lacks Cassette II. It would be tempting to suggest that this supports the hypothesis of calcineurin A interacting with Cassette II. However, the main caveats to this are that the individual paired comparisons between “Vehicle” and “FK506” do not reach statistical significance, and even more so when correcting for multiple comparisons. Additionally,

as with the HUASMC studies, when considering the reduction of pH_i from the start of the experiment, to the end following introduction of DMA and FK506, the differences between the treatment groups is less apparent. With this, it is likely that there is no true effect of calcineurin inhibition in this experimental model.

Does the use of DMA to inhibit NHEs produce an artificial result?

When considering overall (NBC and NHE) pH_i recovery from acidosis, the protective (major / T) allele at rs13096477 is associated with greater net base uptake and resting pH_i only in HUASMCs, but not HUVECs

These studies exploring NBC activity utilises 10 μ M DMA to inhibit NHE transporters, with the aim to leave NBCs as the major pH_i regulator of the cells assessed. This provides an appropriate experimental protocol to assess the specific effect of *SLC4A7* genotype on NBC function, but however, does not provide a true representation of *in vivo* pH_i regulation. This is particularly as NHE function appears to dominate in HUVECs but not as much in HAUSMCs (Figure 134). This opens up the possibility that variations in NBC activity may have a relevant impact on overall pH_i regulation for HAUSMCs, but not HUVECs. This is particularly relevant as the *SLC4A7* knockout mouse model demonstrates an increased HCO_3^- -independent, amiloride-sensitive (NHE) base uptake, suggesting some degree of compensatory effect between the different pH_i regulators. With this, it is worth revisiting studies conducted without NHE inhibition (HUASMCs, Figure 100a; HUVECs, Figure 115a).

With HUASMCs (Figure 100a), there is a pH_i recovery curve that appears to diverge with rs13082711 heterozygotes (C/T) recovering faster than common homozygotes (T/T), even in the presence of uninhibited NHE. This is better represented when considering the net base uptake at regular pH_i intervals (Figure 136). When grouping by cell morphology, as previously discussed, there is a marked difference in pH_i recovery (Figure 136a). In both subgroups, the genotypic effect on HUASMC NBCn1 activity does not appear to be overcome by the NHE function.

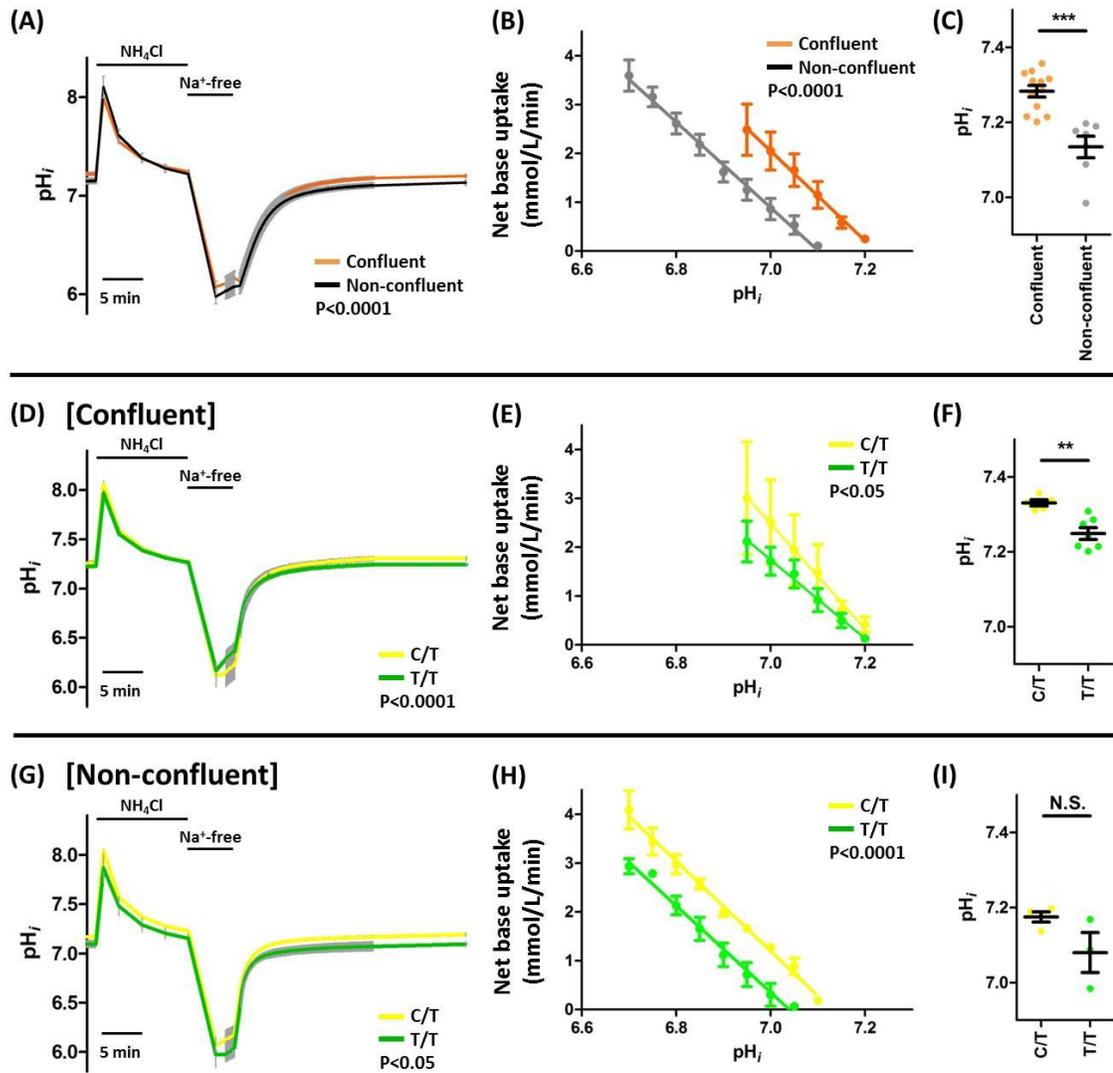


Figure 136: HUASMC pH_i recovery following intracellular acidosis, in the presence of CO_2/HCO_3^- . Confluent subgroup (orange, $n=12$) and non-confluent subgroup (grey, $n=7$) (A-C). *rs13082711* genotypes in confluent subgroup: C/T (yellow, $n=5$) and T/T (green, $n=7$) (D-F). *rs13082711* genotypes in non-confluent subgroup: C/T (yellow, $n=4$) and T/T (green, $n=3$) (G-I). Grey error bars denote SEM.

Groups in (A, D and G) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B, E and H) compared by linear regression. Groups in (C, F and I) compared by Mann-Whitney U-test.

The genotype effect on final plateau pH_i is also observed in the confluent subgroup (Figure 137), although the non-confluent subgroup was underpowered to detect the difference. The difference in pH_i plateau after recovery from intracellular acidosis in the presence of CO_2/HCO_3^- and absence of DMA is 0.081 ± 0.020 ($P < 0.01$), which numerically is close to that of the studies in presence of

DMA (Figure 136), potentially suggesting that NHEs may have only a relatively low, if at all, contribution to the difference.

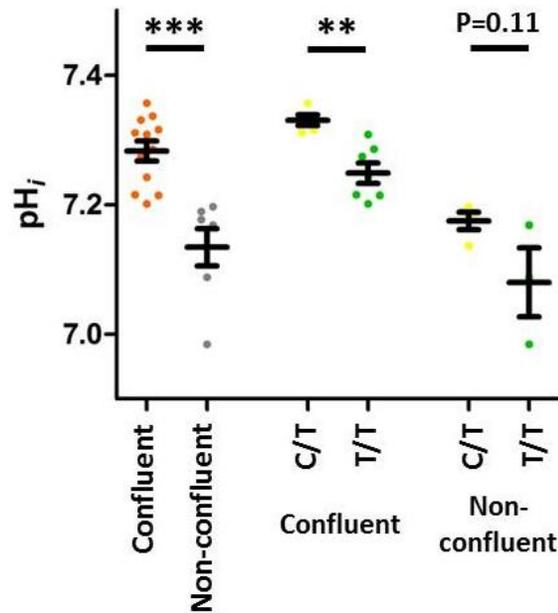


Figure 137: Final plateau pH_i after recovery from intracellular acidosis. Confluent (orange, $n=12$) and non-confluent subgroup (grey, $n=7$). *rs13082711* genotypes in confluent subgroup: C/T (yellow, $n=5$) and T/T (green, $n=7$). ** $P < 0.01$, *** $P < 0.001$, by Mann-Whitney U-test.

On the other hand, the pH_i recovery curves for HUVECs in CO_2/HCO_3^- -containing conditions are practically overlapping (Figure 138). This suggests that genotypic effects on NBCn1 in HUVECs (Figure 118) can be overcome by the proportionally larger contribution of NHEs to HUVEC pH_i regulation. Similarly, genotype did not have an influence on final plateau pH_i in the presence of CO_2/HCO_3^- .

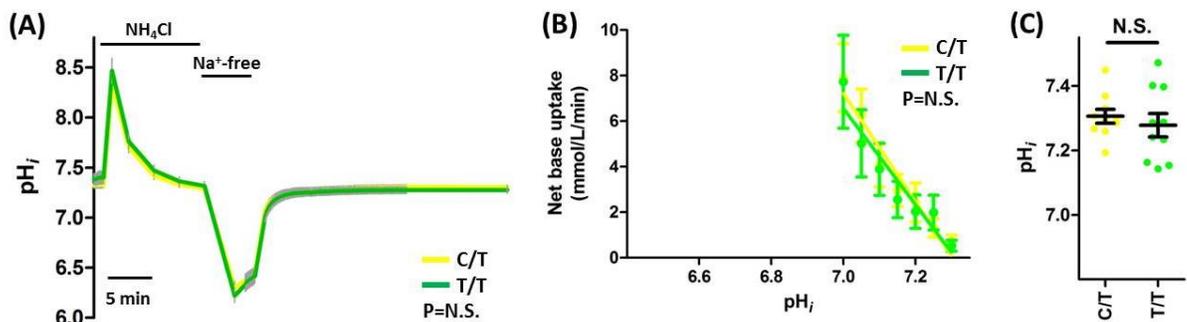


Figure 138: HUVEC pH_i recovery following intracellular acidosis, with CO_2/HCO_3^- *rs13082711* genotypes: C/T (yellow, $n=10$) and T/T (green, $n=10$). Grey error bars denote SEM. Groups in (A) compared by repeated measures two-way ANOVA over the 280 time points between 60 seconds to 900 seconds after re-introduction of Na^+ -containing buffer. Groups in (B) compared by linear regression. Groups in (C) compared by Mann-Whitney U-test. Note differing y-axis scale in (B) compared to

Figure 136.

This difference between HUASMCs and HUVECs should also be considered with the qRT-PCR results (Section 5.2.3) where the inter-sample variability in HUASMC is higher. This suggests that the expression differences in genotype have the potential to be of a larger range, hence more likely to be detected on a cellular function level. This view on pH_i regulation, without inhibition of any acid-base transporters, may reveal that the genotypic effect of rs13082711 on blood pressure may have a larger role in HUASMCs than in HUVECs.

Is the pH_i recovery or resting pH_i that is important? Or neither? Or both? Does it matter?

As already explored in Sections 1.4 to 1.8, steady state pH_i has the potential to alter cellular function in a variety of tissues that may affect blood pressure. In summary, this could be via pH-dependent changes in enzyme activity, channel activity and receptor activity. Similarly, different transport activities may cause Na^+ overload and secondary Ca^{2+} -overload with consequent effects on cellular behaviour. However, the effect may be related not just to resting pH_i , but also to the rate from which pH_i recovers from intracellular acidosis; particularly as pH_i is not constant and pH_i recovery rates are important in response to a variety of *in vivo* vasoconstrictive stimuli such as angiotensin II, noradrenaline, endothelin-1 and cellular depolarization (Hatori *et al.*, 1987; Aalkjaer and Cragoe, 1988; Touyz and Schiffrin, 1993; Austin and Wray, 1993).

It is hitherto unknown whether it is the baseline pH_i , or the ability to return towards its set-point that predominantly influences the overall *in vivo* phenotype. However, these results indicate a genotypic effect on *both* the rate of recovery and resting levels of pH_i . Thus, although the *exact* mechanism is not yet clear, the genetic link is still present.

Is this clinically significant?

This is a much more difficult question to answer.

As discussed in Sections 1.7 and 1.8, the role of NBCn1 in influencing blood pressure is unlikely to be limited to the vasculature, but is not further explored in this thesis. It is very much feasible that genetic variants in *SLC4A7* may influence renal, gastrointestinal, cardiac or neuronal tissue function and subsequently blood pressure. However, regardless of this possibility, the work shown here suggests that the variants influence vascular smooth muscle pH_i regulation.

As discussed in Section 1.13, there has been some observational data from hypertensive arterial segments suggesting pH_i regulation may play a role in blood pressure regulation. Accepting the caveats and complexities of comparing different inbred animal strains, the pH_i in mesenteric artery segments of spontaneously hypertensive rats were only 0.03 higher than Wistar-Kyoto controls, with approximately 70% increased rate of pH_i recovery (Foster *et al.*, 1992). In human studies, the difference between the pH_i of arterial segments after noradrenaline contraction from hypertensive subjects were protected from intracellular acidosis by about 0.04 as compared to those from normotensive subjects (Izzard *et al.*, 1991). Furthermore, non-selective pharmacological inhibition of all NBCs with S0859 reduces the pH_i of cat myocytes by 0.2 (De Giusti *et al.*, 2010), and the blood vessels of *SLC4A7* knockout mice had resting pH_i that were 0.14 lower than their wildtype littermates (Boedtkjer *et al.*, 2011).

In the data presented in this thesis, the difference between the genotypes in resting pH_i (between 0.08 and 0.12) and pH_i recovery rates (between 15 and 55%), with both parameters higher in rs13082711 risk allele carriers as compared to the protective allele homozygotes. This should be regarded in the context the studies explored in Section 1.4 – in rat mesenteric artery studies, a fall of pH_i by 0.1 was associated with a decreased *ex vivo* contractility by around 10%, with the converse also true with the increase in pH_i (Austin and Wray, 1993a), and in *SLC4A7* knockout mice, a fall in pH_i of 0.1 was associated with a decreased *in vitro* cell proliferation by around 10% and *in vivo* medial hypertrophy by around 5%. With this, it could be argued that if the observed *in vitro* genotypic difference continues to persist *in vivo* this may be clinically significant.

6. Conclusions: summary, key limitations, future studies and implications of results

GWAS have identified multiple genetic loci associated with blood pressure, each with a modest contributory effect to overall blood pressure levels, one of these was found to be encompassed by *SLC4A7/NBCn1*. The function of NBCn1 was well described prior to the work conducted for this thesis. It is known to regulate pH_i by the electroneutral symport of sodium and bicarbonate into the cell. Reduction of NBCn1 activity, by knockdown or knockout markedly reduces, but does not abolish recovery from intracellular acidosis. The *SLC4A7*^{-/-} knockout mouse demonstrates a complicated blood pressure phenotype where they were modestly hypertensive at rest, but significantly resistant to hypertensive stressors such as angiotensin II nitric oxide synthase inhibition (Boedtkjer *et al.*, 2011). As NBCn1 is expressed in various tissues, it is unclear which one (or more) contributes to blood pressure regulation.

There is a relative large literature base regarding other sodium-dependent pH_i regulators, such as NHE1 and NHE3. Less is known about NBCn1, and how subtle genetic variations might affect its function, and potentially lead to blood pressure changes. Based on other co-transporters, various hypotheses could be proposed to explain the impact of genetic changes on NBCn1 expression and/or function. These include effects on baseline expression or inducibility, splice variants, protein trafficking, post-translational modifications and intracellular protein-protein interactions. This is complicated by the understanding that the blood pressure-associated SNP rs13082711 is in a linkage disequilibrium block with 92 other SNPs. With this, rs13082711 could merely be a marker, and *theoretically* could be any (or none, or more) of these polymorphisms that is the true causative SNP.

The work presented here has identified that the blood pressure risk allele rs13082711 is associated with:

- Allele-dependent DNA-nuclear protein interactions
- greater transcription levels,
- NBCn1 protein expression and cell membrane availability, particularly in vascular smooth muscle cells, but perhaps also in endothelial cells, and
- increased Na^+/HCO_3^- -dependent, DMA-insensitive pH_i recovery, base uptake and resting pH_i , once again more apparent in HUASMCs.

Notably, due to the large contribution of NHE in HUVEC pH_i regulation, the small genotypic effect is removed when NHE function is present. This is not the case for HUASMCs, where the genotypic effect persists even in the absence of DMA. This effect was independent of the associated E326K amino acid change, suggesting that the genotypic influence on function is predominantly related to overall expression levels.

In terms of negative findings, significant efforts attempting to identify the underlying molecular mechanism for this genotype-associated difference in expression were unsuccessful, despite the initial promise of FAIRE and EMSA results. In assessing potential impact on protein behavior, the studies did not reveal any clear evidence of the polymorphism influencing gene splicing, protein trafficking or interaction with calcineurin A.

6.1. Key limitations

One of the limitations of the current work is the restricted number of rare homozygotes, with a genotype frequency of less than 4% in this population. To feasibly conduct all the studies, while maintaining primary cultures to a low passage number, this would require a minimum of 12 rare homozygotes and even higher for assays with higher variability such as qRT-PCR. This suggests that an overall sample size of 532 provides a 95% chance of having sufficient rare homozygotes (calculated using binomial distributions with event probability of 3.4%, also accounting for unclassified genotype calls, see Table 41). This would cost approximately £36,700 if both HUASMCs and HUVECs are required, or £19,700 and £18,400 for HUASMCs or HUVECs alone, respectively. Furthermore, this calculation merely estimates the costs for cell culture to get to a state of being able to conduct the necessary assays, and not the cost of the studies themselves. This needs to be considered against a consumables budget of £30,000 for the duration of the PhD as provided by the kind support of the British Heart Foundation. In addition to monetary cost, this endeavor would take up to an estimated 1,200 researcher-hours, which approximates to 1/5th of the allocated PhD time. Perhaps this is not a time- and cost-effective expenditure.

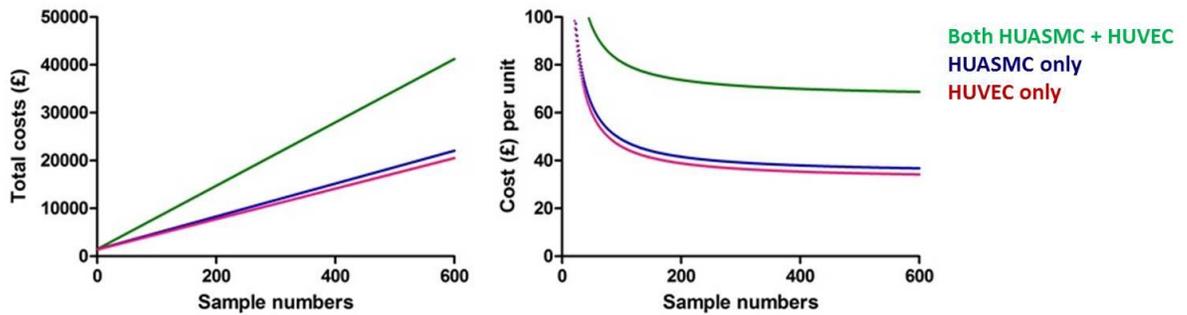


Figure 139: Estimate costings for increasing sample size
 Total costs (left) and cost per unit (right) with increasing sample sizes of deriving primary cell cultures of both HUASMCs and HUVECs (green), HUASMCs only (blue) and HUVECs only (red). Full costings available in Appendix 20.

To circumvent the need for a large repository of cells due to the low prevalence of rare homozygotes, some may suggest the use of the CRISPR (clustered regularly-interspaced short palindromic repeats)/Cas9 technology. It is often argued that this would allow the investigation of an isolated region, with less noise from the variation in other genetic and environmental (*in vitro* cell culture) factors. There are several counter-arguments against this suggestion. The most relevant is that CRISPR/Cas9 technology is not able to replace the whole LD block. Thus it would be impossible to investigate all 93 SNPs in high LD individually. Next, despite the rapidly improving methodology, the off-target indels introduced by the technique is still being quoted at around 0.1% (Kim *et al.*, 2015), which equates to thousands of undesired changes in the genetic code. Last of all, there are already genotype-associated differences in protein expression and function with limited access to rare homozygotes. In the absence of evidence (or a strong hypothesis) that there is a dominant/recessive effect, the pursuit of rare homozygotic samples is unnecessary.

The failure to identify the molecular mechanism underlying the allele-associated changes in gene expression has been frustrating. This is especially as FAIRE and EMSA studies have potentially narrowed down to a few candidate SNPs from the large LD block. An alternative to the attempted protein sequencing would require specific antibodies for immunoblots, and this is very difficult as prediction tools have been seen to be poor in this setting. With this, it is perhaps not a time- and cost-effective pursuit.

Although there are exciting results with regards to an genotypic effect on gene expression and overall protein function, it should always be taken with the caution (as previously indicated) that this is an *in vitro* study of embryonic (with the additional issues of carrying arterial deoxygenated

blood and lower arterial pressures) conduit vessels. This may not represent fully the function in adult resistance vessels. Additionally, the work conducted here, like all *in vitro* cell culture work, is based on isolated cells, and not in an whole organism setting, with complicated interactions, either by cell-cell contact, paracrine, endocrine and neurogenic factors.

Potential further studies (some more feasible than others) to further the understanding on *SLC4A7* and hypertension will be discussed in the next section, also outlining logistical and financial considerations.

6.2. Possible future studies

Possible future studies that have been considered, or sometime kindly suggested elsewhere, can be split into (1) animal studies, (2) *in vitro* studies on other cell types, (3) *ex vivo* studies on human blood vessels or (4) *in vivo* studies with healthy volunteers. The limitations of pharmacological inhibition of NBCn1 as an experimental tool will also be discussed, although this may be an area of interest for the future.

To take a step away from human studies, some have suggested a return to animal models. It is now possible to introduce point mutations in mouse models which in theory would be able to assess a single polymorphism. This has two main (and fatal) drawbacks. The first have been previously discussed with suggestions of CRISPR/Cas on human cells. Due to the large LD block, there are multiple potential causative (if at all) SNPs to explore – too many to realistically do so by single point mutations. The second consideration is that although the final NBCn1 sequence of humans and mouse are similar, there are 80 amino acid differences (11 within Cassette II itself), *including* that of the E326K human variation (where the murine NBCn1 has Glutamine/Q in the 326 position). Therefore, introducing the point mutation changing residue to E or K may not be appropriate. Additionally, work for this study has suggested that genotype-associated variation in expression levels is more likely the mechanism action rather than the amino acid change.

The work here has already sampled human adult kidneys, identifying a genotype-associated level of mRNA expression by allelic imbalance. However, due to the heterogenous nature of renal biopsies, it would be difficult to determine protein expression by immunoblots as there will be no appropriate control for the relative quantity of cells from the thick ascending loop of Henle (where NBCn1 is predominantly expressed in kidneys). Obtaining large samples of kidneys is difficult, as is

samples of other tissue types that may play a role in blood pressure regulation, such as the brain and GI tract. One possible sample is from gluteal biopsies for resistance vessels. This would be a potential source to confirm these results in adult resistance vessels. Of course, the previous caution with sample sizes and costs should be considered, and may be prohibitive. Additionally, the task of recruiting large numbers of healthy volunteers for an invasive tissue biopsy may also be logistically difficult. The final consideration of adult vessels is that environmental factors would also have a large effect to potentially overcome subtle genetic effects (as reflected by the huge cohorts required for GWASs).

Perhaps the most interesting possibility lies with *ex vivo* wire myograph studies of arterial segments. The arteries could either be from umbilical cords or gluteal biopsies, for which the relative advantages and disadvantages have been discussed above. The attraction of this set-up is that it progresses on the current knowledge that *SLC4A7* genotype influences NBCn1 expression and pH_i regulation onto whole arterial function (with both vascular smooth muscle and endothelial contributions). Arterial segments could be studied in a various combination of settings – with or without $\text{CO}_2/\text{HCO}_3^-$, at external pH of 6.8 or 7.4, with or without NHE inhibition, exposure to a range of vasoconstrictors or vasodilators and with or without endothelial contribution (pharmacologically or mechanically). This would be able to determine whether there is a genotype-associated difference in vascular function overall, and in isolated settings.

In vivo healthy volunteer studies would be interesting if there were a clear mechanism of assessing NBCn1. There are two main considerations, either (1) local vascular beds in the short term, or (2) whole person in the medium term. The first proposal would take the form of forearm plethysmography. As there are currently no licensed NBC inhibitors of human use, the closest pharmacological challenge could be in the form of amiloride. As previously discussed, local amiloride infusion reduced the angiotensin II-induced vasoconstriction in healthy volunteers, and was vasodilatory in high concentrations (Pickkers *et al.*, 1999). This could be reproduced to identify any genotype-associated changes in response to amiloride. In theory, this short-lived local infusion would mainly have an effect on the vasculature, and not due to its renal diuretic effects, and can reveal the contribution of NBCs, as NHEs could be inhibited by the amiloride.

Studies on the whole person is even more difficult. The dominant effects of oral, long-term amiloride is probably from the diuretic effect of ENaC inhibition, and therefore difficult to

attribute any (if at all) differences to NBC. However, there is the possibility (although currently untested even in the knockout mouse model) that NBCs can play a role in salt-and-water regulation due to its expression in the ascending loop of Henle. A potential study is to salt-load and -restrict healthy volunteers, with end-points being 24-hour ambulatory blood pressure measurements and 24-hour urinary electrolyte excretion. Although a potentially interesting, there is perhaps not enough supportive evidence yet of NBCn1 contributing to salt-and-water regulation to justify this time- and cost-heavy study.

The lack of NBCn1 inhibitors is currently a limiting factor in exploring its function. However, there is a possibility of advancing the small molecule strategy (currently S0859 as the best candidate, but is not suitable for *in vivo* use due to its high albumin binding, and also non-specificity for NBC isoforms). Alternatively, there is a possibility of inhibitory monoclonal antibodies, although there is none close to fruition as yet.

6.3. Implications of results

At the end of this thesis, it is worth considering what clinical implications may follow from this study. As previously discussed, this is an *in vitro* study, for which there are still several hurdles to identify the genotypic effect on hypertension patients *in vivo*. Three main questions that might be posed are (1) is this work transferable to other conditions, (2) whether this work aid personalized medicine, and (3) does this work bring any potential new therapeutics closer to reality? Personally, the answers are yes, no and maybe (but only by a slight amount).

As indicated in Section 1.17, a separate SNP at the 3'-UTR *SLC4A7* also demonstrates an association with breast cancer susceptibility (Ahmed *et al.*, 2009; Chen *et al.*, 2012). Additionally, intracellular alkalosis, or a resistance to intracellular acidosis, has already been associated with cancer cell growth and survival *in vitro* (Pedersen *et al.*, 2006). The results here could stimulate a similar study on breast cancer cells, but with the rs4973768 SNP instead.

It is the opinion of the author that personalized medicines based on genotype, or even a panel of SNPs is far from reality when applied to complex diseases such as hypertension. Pharmacogenetics studies in hypertension has generally not been able to show *clinically significant* results, let alone be cost-effective for a condition like hypertension, which already has a large

range of therapeutic options. The knowledge gained from this study does not take this further; particularly there were no proven impact of calcineurin inhibition.

Although the responsible mechanism that results in genotype-associated expression levels was not ascertained it may be postulated that this study adds weight of evidence (albeit very small) to suggest that NBCn1 is a potential target. All previous work on NBCn1 in blood pressure regulation was based on animal models and, for the first time, the work presented here indicates a link between genotype to human *SLC4A7* expression and NBCn1 protein function. This may stimulate progress on developing an NBCn1 inhibitor that may be appropriate for *in vivo* use. However, the author is fully aware that the attrition rate from candidate target to actual marketed pharmacological agent is in the order of thousands to one. Additionally, there is a lower desire for pharmaceutical development of drugs for hypertension due to a flooded marketplace with older off-patent drugs. The most likely path for NBCn1 inhibitors to reach the cardiovascular market is for one to be initially developed for the oncology market. Watch this space.

References

- Aalkjaer C, Cragoe EJ Jr. Intracellular pH regulation in resting and contracting segments of rat mesenteric resistance vessels. *J Physiol*. 1988 Aug;402:391-410.
- Ahmed S, Thomas G, Ghossaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos Santos Silva I, *et al*. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet*. 2009 May;41(5):585-90.
- Ali R, Amlal H, Burnham CE, Soleimani M. Glucocorticoids enhance the expression of the basolateral Na⁺:HCO₃⁻ cotransporter in renal proximal tubules. *Kidney Int*. 2000 Mar;57(3):1063-71.
- Al-Merani SA, Brooks DP, Chapman BJ, Munday KA. The half-lives of angiotensin II, angiotensin II-amide, angiotensin III, Sar1-Ala8-angiotensin II and renin in the circulatory system of the rat. *J Physiol*. 1978 May;278:471-90.
- Alzamora R, Michea L, Marusic ET. Role of 11beta-hydroxysteroid dehydrogenase in nongenomic aldosterone effects in human arteries. *Hypertension*. 2000 May;35(5):1099-104.
- Amlal H, Chen Q, Greeley T, Pavelic L, Soleimani M. Coordinated down-regulation of NBC-1 and NHE-3 in sodium and bicarbonate loading. *Kidney Int*. 2001 Nov;60(5):1824-36.
- Amlal H, Ledoussal C, Sheriff S, Shull GE, Soleimani M. Downregulation of renal AQP2 water channel and NKCC2 in mice lacking the apical Na⁺-H⁺ exchanger NHE3. *J Physiol*. 2003 Dec 1;553(Pt 2):511-22.
- Amlal H, Sheriff S, Soleimani M. Upregulation of collecting duct aquaporin-2 by metabolic acidosis: role of vasopressin. *Am J Physiol Cell Physiol*. 2004 May;286(5):C1019-30.
- Ammar YB, Takeda S, Hisamitsu T, Mori H, Wakabayashi S. Crystal structure of CHP2 complexed with NHE1-cytosolic region and an implication for pH regulation. *EMBO J*. 2006 Jun 7;25(11):2315-25.
- Arora P, Wu C, Khan AM, Bloch DB, Davis-Dusenbery BN, Ghorbani A, Spagnolli E, Martinez A, Ryan A, Tainsh LT, Kim S, Rong J, Huan T, Freedman JE, Levy D, *et al*. Atrial natriuretic peptide is negatively regulated by microRNA-425. *J Clin Invest*. 2013 Aug 1;123(8):3378-82.
- Austin C, Wray S (a). Extracellular pH signals affect rat vascular tone by rapid transduction into intracellular pH changes. *J Physiol*. 1993 Jul;466:1-8.
- Austin C, Wray S (b). Changes of intracellular pH in rat mesenteric vascular smooth muscle with high-K⁺ depolarization. *J Physiol*. 1993 Sep;469:1-10.
- Baetz D, Haworth RS, Avkiran M, Feuvray D. The ERK pathway regulates Na⁽⁺⁾-HCO₃⁽⁻⁾ cotransport activity in adult rat cardiomyocytes. *Am J Physiol Heart Circ Physiol*. 2002 Nov;283(5):H2102-9. Erratum in: *Am J Physiol Heart Circ Physiol* 2002 Dec;283(6).
- Balnavae CD, Vaughan-Jones RD. Effect of intracellular pH on spontaneous Ca²⁺ sparks in rat ventricular myocytes. *J Physiol*. 2000 Oct 1;528 Pt 1:25-37.
- Banerjee A, Gordon SM, Intlekofer AM, Paley MA, Mooney EC, Lindsten T, Wherry EJ, Reiner SL. Cutting edge: The transcription factor eomesodermin enables CD8⁺ T cells to compete for the memory cell niche. *J Immunol*. 2010 Nov 1;185(9):4988-92.

- Barrdahl M, Canzian F, Joshi AD, Travis RC, Chang-Claude J, Auer PL, Gapstur SM, Gaudet M, Diver WR, Henderson BE, Haiman CA, Schumacher FR, Le Marchand L, Berg CD, Chanock SJ, *et al.* Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. *Hum Mol Genet.* 2014 Oct 1;23(19):5260-70.
- Batlle DC, Saleh A, Rombola G. Reduced intracellular pH in lymphocytes from the spontaneously hypertensive rat. *Hypertension.* 1990 Jan;15(1):97-103.
- Bell E, Ivarsson B, Merrill C. Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro. *Proc Natl Acad Sci U S A.* 1979 Mar;76(3):1274-8.
- Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 2012;8(4):e1002629.
- Bell SM, Schreiner CM, Schultheis PJ, Miller ML, Evans RL, Vorhees CV, Shull GE, Scott WJ. Targeted disruption of the murine Nhe1 locus induces ataxia, growth retardation, and seizures. *Am J Physiol.* 1999 Apr;276(4 Pt 1):C788-95.
- Bergeron R, Kjaer M, Simonsen L, Bülow J, Skovgaard D, Howlett K, Galbo H. Splanchnic blood flow and hepatic glucose production in exercising humans: role of renin-angiotensin system. *Am J Physiol Regul Integr Comp Physiol.* 2001 Dec;281(6):R1854-61.
- Besse S, Tanguy S, Boucher F, Huraux C, Riou B, Swynghedauw B, de Leiris J. Protection of endothelial-derived vasorelaxation with cariporide, a sodium-proton exchanger inhibitor, after prolonged hypoxia and hypoxia-reoxygenation: effect of age. *Eur J Pharmacol.* 2006 Feb 15;531(1-3):187-93.
- Besterman JM, May WS Jr, LeVine H 3rd, Cragoe EJ Jr, Cuatrecasas P. Amiloride inhibits phorbol ester-stimulated Na⁺/H⁺ exchange and protein kinase C. An amiloride analog selectively inhibits Na⁺/H⁺ exchange. *J Biol Chem.* 1985 Jan 25;260(2):1155-9.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014 Apr 29. 42(Web Server issue):W252-8.
- Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 1999 Dec 17;294(5):1351-62.
- Boedtkjer E, Bentzon JF, Dam VS, Aalkjaer C. Na⁺, HCO₃⁻-cotransporter NBCn1 increases pHi gradients, filopodia, and migration of smooth muscle cells and promotes arterial remodelling. *Cardiovasc Res.* 2016;111:227-39.
- Boedtkjer E, Damkier HH, Aalkjaer C. NHE1 knockout reduces blood pressure and arterial media/lumen ratio with no effect on resting pH(i) in the vascular wall. *J Physiol.* 2012 Apr 15;590(Pt 8):1895-906.
- Boedtkjer E, Kim S, Aalkjaer C. Endothelial alkalisation inhibits gap junction communication and endothelium-derived hyperpolarisations in mouse mesenteric arteries. *J Physiol.* 2013 Mar 15;591(Pt 6):1447-61.
- Boedtkjer E, Praetorius J, Aalkjaer C. NBCn1 (slc4a7) mediates the Na⁺-dependent bicarbonate transport important for regulation of intracellular pH in mouse vascular smooth muscle cells. *Circ Res.* 2006 Mar 3;98(4):515-23.

- Boedtkjer E, Praetorius J, Füchtbauer EM, Aalkjaer C. Antibody-independent localization of the electroneutral Na⁺-HCO₃⁻ cotransporter NBCn1 (slc4a7) in mice. *Am J Physiol Cell Physiol*. 2008 Feb;294(2):C591-603.
- Boedtkjer E, Praetorius J, Matchkov VV, Stankevicius E, Mogensen S, Füchtbauer AC, Simonsen U, Füchtbauer EM, Aalkjaer C. Disruption of Na⁺,HCO₃⁻ cotransporter NBCn1 (slc4a7) inhibits NO-mediated vasorelaxation, smooth muscle Ca²⁺ sensitivity, and hypertension development in mice. *Circulation*. 2011 Oct 25;124(17):1819-29.
- Bok D, Galbraith G, Lopez I, Woodruff M, Nusinowitz S, BeltrandelRio H, Huang W, Zhao S, Geske R, Montgomery C, Van Sligtenhorst I, Friddle C, Platt K, Sparks MJ, Pushkin A, *et al*. Blindness and auditory impairment caused by loss of the sodium bicarbonate cotransporter NBC3. *Nat Genet*. 2003 Jul;34(3):313-9.
- Boron WF, Chen L, Parker MD. Modular structure of sodium-coupled bicarbonate transporters. *J Exp Biol*. 2009 Jun;212(Pt 11):1697-706.
- Boron WF, De Weer P. Intracellular pH transients in squid giant axons caused by CO₂, NH₃, and metabolic inhibitors. *J Gen Physiol*. 1976 Jan;67(1):91-112.
- Bossel Ben-Moshe N, Avraham R, Kedmi M, Zeisel A, Yitzhaky A, Yarden Y, Domany E. Context-specific microRNA analysis: identification of functional microRNAs and their mRNA targets. *Nucleic Acids Res*. 2012 Nov;40(21):10614-27.
- Bountra C, Vaughan-Jones RD. Effect of intracellular and extracellular pH on contraction in isolated, mammalian cardiac muscle. *J Physiol*. 1989 Nov;418:163-87.
- Boyce SW, Bartels C, Bolli R, Chaitman B, Chen JC, Chi E, Jessel A, Kereiakes D, Knight J, Thulin L, Theroux P; GUARD During Ischemia Against Necrosis (GUARDIAN) Study Investigators. Impact of sodium-hydrogen exchange inhibition by cariporide on death or myocardial infarction in high-risk CABG surgery patients: results of the CABG surgery cohort of the GUARDIAN study. *J Thorac Cardiovasc Surg*. 2003 Aug;126(2):420-7.
- Brenninkmeijer L, Kuehl C, Geldart AM, Arons E, Christou H. Heme oxygenase-1 does not mediate the effects of extracellular acidosis on vascular smooth muscle cell proliferation, migration, and susceptibility to apoptosis. *J Vasc Res*. 2011;48(4):285-96.
- Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. 2007;35(11):3823-35.
- Brooks HL, Sorensen AM, Terris J, Schultheis PJ, Lorenz JN, Shull GE, Knepper MA. Profiling of renal tubule Na⁺ transporter abundances in NHE3 and NCC null mice using targeted proteomics. *J Physiol*. 2001 Feb 1;530(Pt 3):359-66.
- Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res*. 2008 Jan;36(Database issue):D102-6.
- Burnette WN. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem*. 1981 Apr;112(2):195-203.
- Bussmann M, Baumgart M, Bott M. RosR (Cg1324), a hydrogen peroxide-sensitive MarR-type transcriptional regulator of *Corynebacterium glutamicum*. *J Biol Chem*. 2010 Sep 17;285(38):29305-18.

- Cabrera CP, Ng FL, Warren HR, Barnes MR, Munroe PB, Caulfield MJ. Exploring hypertension genome-wide association studies findings and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley Interdiscip Rev Syst Biol Med*. 2015 Mar;7(2):73-90.
- Caldwell PC. An investigation of the intracellular pH of crab muscle fibres by means of micro-glass and micro-tungsten electrodes. *J Physiol*. 1954 Oct 28;126(1):169-80.
- Calhoun DA, Jones D, Textor S, Goff DC, Murphy TP, Toto RD, White A, Cushman WC, White W, Sica D, Ferdinand K, Giles TD, Falkner B, Carey RM. Resistant hypertension: diagnosis, evaluation, and treatment. A scientific statement from the American Heart Association Professional Education Committee of the Council for High Blood Pressure Research. *Hypertension*. 2008 Jun;51(6):1403-19.
- Cartegni L., Wang J., Zhu Z., Zhang M. Q., Krainer A. R.; 2003. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acid Research*, 2003, 31(13): 3568-3571.
- Carter KL, Cahir-McFarland E, Kieff E. Epstein-barr virus-induced changes in B-lymphocyte gene expression. *J Virol*. 2002 Oct;76(20):10427-36.
- Casey JR, Grinstein S, Orlowski J. Sensors and regulators of intracellular pH. *Nat Rev Mol Cell Biol*. 2010 Jan;11(1):50-61. doi: 10.1038/nrm2820. Epub 2009 Dec 9. Review.
- Chadwick LH. The NIH Roadmap Epigenomics Program data resource. *Epigenomics*. 2012 June ; 4(3): 317–324. doi:10.2217/epi.12.18.
- Chaillet JR, Boron WF. Intracellular calibration of a pH-sensitive dye in isolated, perfused salamander proximal tubules. *J Gen Physiol*. 1985 Dec;86(6):765-94.
- Chamley JH, Campbell GR, Burnstock G. Dedifferentiation, redifferentiation and bundle formation of smooth muscle cells in tissue culture: the influence of cell number and nerve fibres. *J Embryol Exp Morphol*. 1974 Oct;32(2):297-323.
- Chauhan JS, Rao A, Raghava GP. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One*. 2013 Jun 28;8(6):e67008.
- Chekmenev DS, Haid C, Kel AE. P-Match: transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W432-7.
- Chen LM, Kelly ML, Rojas JD, Parker MD, Gill HS, Davis BA, Boron WF. Use of a new polyclonal antibody to study the distribution and glycosylation of the sodium-coupled bicarbonate transporter NCBE in rodent brain. *Neuroscience*. 2008 Jan 24;151(2):374-85.
- Chen M, Praetorius J, Zheng W, Xiao F, Riederer B, Singh AK, Stieger N, Wang J, Shull GE, Aalkjaer C, Seidler U. The electroneutral Na⁺:HCO₃⁻ cotransporter NBCn1 is a major pHi regulator in murine duodenum. *J Physiol*. 2012 Jul 15;590(Pt 14):3317-33.
- Chen W, Zhong R, Ming J, Zou L, Zhu B, Lu X, Ke J, Zhang Y, Liu L, Miao X, Huang T. The SLC4A7 variant rs4973768 is associated with breast cancer risk: evidence from a case-control study and a meta-analysis. *Breast Cancer Res Treat*. 2012 Dec;136(3):847-57.
- Cheval L, Pierrat F, Rajerison R, Piquemal D, Doucet A. Of mice and men: divergence of gene expression patterns in kidney. *PLoS One*. 2012;7(10):e46876.
- Choi I, Hu L, Rojas JD, Schmitt BM, Boron WF. Role of glycosylation in the renal electrogenic Na⁺-HCO₃⁻ cotransporter (NBCe1). *Am J Physiol Renal Physiol*. 2003 Jun;284(6):F1199-206.

- Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 1987 Apr;162(1):156-9.
- Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A.* 1973 Nov;70(11):3240-4.
- Cooper R, Aina O, Chaco L, Achilihu AG, Shamsi N, Ford E. Red cell sodium and potassium in hypertension among blacks. *J Natl Med Assoc.* 1989 Apr;81(4):365-70.
- Critical Assessment of Techniques for Protein Structure Prediction (CASP10), 2012. 10th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction. <http://predictioncenter.org/casp10/>. Accessed 26/5/13.
- Cukiernik M, Hileeto D, Downey D, Evans T, Khan ZA, Karmazyn M, Chakrabarti S. The role of the sodium hydrogen exchanger-1 in mediating diabetes-induced changes in the retina. *Diabetes Metab Res Rev.* 2004 Jan-Feb;20(1):61-71.
- Damkier HH, Nielsen S, Praetorius J. An anti-NH₂-terminal antibody localizes NBCn1 to heart endothelia and skeletal and vascular smooth muscle cells. *Am J Physiol Heart Circ Physiol.* 2006 Jan;290(1):H172-80
- Damkier HH, Nielsen S, Praetorius J. Molecular expression of SLC4-derived Na⁺-dependent anion transporters in selected human tissues. *Am J Physiol Regul Integr Comp Physiol.* 2007 Nov;293(5):R2136-46.
- Danielsen AA, Parker MD, Lee S, Boron WF, Aalkjaer C, Boedtkjer E. Splice cassette II of Na⁺,HCO₃⁻ cotransporter NBCn1 (slc4a7) interacts with calcineurin A: implications for transporter activity and intracellular pH control during rat artery contractions. *J Biol Chem.* 2013 Mar 22;288(12):8146-55.
- De Giusti VC, Garciarena CD, Aiello EA. Role of reactive oxygen species (ROS) in angiotensin II-induced stimulation of the cardiac Na⁺/HCO₃⁻ cotransport. *J Mol Cell Cardiol.* 2009 Nov;47(5):716-22.
- De Giusti VC, Orłowski A, Aiello EA. Angiotensin II inhibits the electrogenic Na⁺/HCO₃⁻ cotransport of cat cardiac myocytes. *J Mol Cell Cardiol.* 2010 Nov;49(5):812-8.
- Deckmann K, Rörsch F, Geisslinger G, Grösch S. Identification of DNA-protein complexes using an improved, combined western blotting-electrophoretic mobility shift assay (WEMSA) with a fluorescence imaging system. *Mol Biosyst.* 2012 Apr;8(5):1389-95.
- Denkin SM, Sekaric P, Nelson DR. Gel shift analysis of the empA promoter region in *Vibrio anguillarum*. *BMC Microbiol.* 2004 Oct 29;4:42.
- Di Sole F, Babich V, Moe OW. The calcineurin homologous protein-1 increases Na⁽⁺⁾/H⁽⁺⁾ -exchanger 3 trafficking via ezrin phosphorylation. *J Am Soc Nephrol.* 2009 Aug;20(8):1776-86.
- Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science.* 2009 Sep 4;325(5945):1246-50.
- Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F. Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D261-7.
- Drummond HA, Gebremedhin D, Harder DR. Degenerin/epithelial Na⁺ channel proteins: components of a vascular mechanosensor. *Hypertension.* 2004 Nov;44(5):643-8.

Ebata S, Muto S, Okada K, Nemoto J, Amemiya M, Saito T, Asano Y. Aldosterone activates Na⁺/H⁺ exchange in vascular smooth muscle cells by nongenomic and genomic mechanisms. *Kidney Int.* 1999 Oct;56(4):1400-12.

El Sharawy A, Hundrieser B, Brosch M, Wittig M, Huse K, Platzer M, Becker A, Simon M, Rosenstiel P, Schreiber S, Krawczak M, Hampe J. Systematic evaluation of the effect of common SNPs on pre-mRNA splicing. *Hum Mutat.* 2009 Apr;30(4):625-32.

EncODE Experimental Matrix, 2012.
<http://encodeproject.org/ENCODE/dataMatrix/encodeDataMatrixHuman.html>. Accessed 3/4/16.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012 Sep 6;489(7414):57-74.

Epting T, Hartmann K, Sandqvist A, Nitschke R, Gordjani N. Cyclosporin A stimulates apical Na⁺/H⁺ exchange in LLC-PK1/PKE20 proximal tubular cells. *Pediatr Nephrol.* 2006 Jul;21(7):939-46.

Esteva-Font C, Ars E, Guillen-Gomez E, Campistol JM, Sanz L, Jiménez W, Knepper MA, Torres F, Torra R, Ballarín JA, Fernández-Llama P. Cyclosporin-induced hypertension is associated with increased sodium transporter of the loop of Henle (NKCC2). *Nephrol Dial Transplant.* 2007 Oct;22(10):2810-6.

European Collection of Cell Cultures, n.d. General Cell Collections: 293.
http://www.hpacultures.org.uk/products/celllines/generalcell/detail.jsp?refId=85120602&collection=ecacc_gc. Accessed 13/9/13.

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science.* 2002 Aug 9;297(5583):1007-13.

Fazius E, Shelest V, Shelest E. SiTaR: a novel tool for transcription factor binding site prediction. *Bioinformatics.* 2011 Oct 15;27(20):2806-11.

Fic E, Kedracka-Krok S, Jankowska U, Pirog A, Dziedzicka-Wasylewska M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis.* 2010 Oct;31(21):3573-9.

Fleming I, Hecker M, Busse R. Intracellular alkalinization induced by bradykinin sustains activation of the constitutive nitric oxide synthase in endothelial cells. *Circ Res.* 1994 Jun;74(6):1220-6.

Förster-Fromme K, Jendrossek D. AtuR is a repressor of acyclic terpene utilization (Atu) gene cluster expression and specifically binds to two 13 bp inverted repeat sequences of the atuA-atuR intergenic region. *FEMS Microbiol Lett.* 2010 Jul;308(2):166-74.

Foster CD, Honeyman TW, Scheid CR. Alterations in Na⁽⁺⁾-H⁺ exchange in mesenteric arteries from spontaneously hypertensive rats. *Am J Physiol.* 1992 Jun;262(6 Pt 2):H1657-62.

Fraley R, Subramani S, Berg P, Papahadjopoulos D. Introduction of liposome-encapsulated SV40 DNA into cells. *J Biol Chem.* 1980 Nov 10;255(21):10431-5.

Frampton M, da Silva Filho MI, Broderick P, Thomsen H, Försti A, Vijayakrishnan J, Cooke R, Enciso-Mora V, Hoffmann P, Nöthen MM, Lloyd A, Holroyd A, Eisele L, Jöckel KH, Ponader S, *et al.* Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. *Nat Commun.* 2013;4:2549.

Franceschini N, Fox E, Zhang Z, Edwards TL, Nalls MA, Sung YJ, Tayo BO, Sun YV, Gottesman O, Adeyemo A, Johnson AD, Young JH, Rice K, Duan Q, Chen F, *et al.* Genome-wide association analysis of blood-pressure

traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am J Hum Genet.* 2013 Sep 5;93(3):545-54.

Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* 1981 Dec 11;9(23):6505-25.

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet.* 1995 May;10(1):111-3.

Frousios K, Iliopoulos CS, Schlitt T, Simpson MA. Predicting the functional consequences of non-synonymous DNA sequence variants--evaluation of bioinformatics tools and development of a consensus strategy. *Genomics.* 2013 Oct;102(4):223-8.

Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ. SCAN: SNP and copy number annotation. *Bioinformatics.* 2010 Jan 15;26(2):259-62.

Ganesh SK, Tragante V, Guo W, Guo Y, Lanktree MB, Smith EN, Johnson T, Castillo BA, Barnard J, Baumert J, Chang YP, Elbers CC, Farrall M, Fischer ME, Franceschini N, *et al.* Loci influencing blood pressure identified using a cardiovascular gene-centric array. *Hum Mol Genet.* 2013 Apr 15;22(8):1663-78.

Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* 1981 Jul 10;9(13):3047-60.

Gaziano TA, Bitton A, Anand S, Weinstein MC; International Society of Hypertension. The global cost of nonoptimal blood pressure. *J Hypertens.* 2009 Jul;27(7):1472-7.

Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, Hudson TJ, Pastinen T. Survey of allelic expression using EST mining. *Genome Res.* 2005 Nov;15(11):1584-91.

Gekle M, Freudinger R, Mildenerger S, Schenk K, Marschitz I, Schramek H. Rapid activation of Na⁺/H⁺-exchange in MDCK cells by aldosterone involves MAP-kinase ERK1/2. *Pflugers Arch.* 2001 Mar;441(6):781-6.

Gill HS, Roush ED, Dutcher L, Patel S. Direct Evidence for Calcineurin Binding to the Exon-7 Loop of the Sodium-Bicarbonate Cotransporter NBCn1. *Int J Biol Sci.* 2014 Jul 4;10(7):771-6.

Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci U S A.* 1984 Jul;81(14):4275-9.

Good DW, Watts BA 3rd, George T, Meyer JW, Shull GE. Transepithelial HCO₃⁻ absorption is defective in renal thick ascending limbs from Na⁺/H⁺ exchanger NHE1 null mutant mice. *Am J Physiol Renal Physiol.* 2004 Dec;287(6):F1244-9.

Gorbatenko A, Olesen CW, Boedtker E, Pedersen SF. Regulation and roles of bicarbonate transporters in cancer. *Front Physiol.* 2014 Apr 16;5:130. doi: 10.3389/fphys.2014.00130. eCollection 2014.

Gordon SM, Chaix J, Rupp LJ, Wu J, Madera S, Sun JC, Lindsten T, Reiner SL. The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. *Immunity.* 2012 Jan 27;36(1):55-67.

Grabe N. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol.* 2002;2(1):S1-15.

- Graham FL, van der Eb AJ. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*. 1973 Apr;52(2):456-67.
- Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974 Sep 6;185(4154):862-4.
- Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol*. 2013;5(3):578-90
- Griffin MJ. Synchronization of some human cell strains by serum and calcium starvation. *In Vitro*. 1976 May;12(5):393-8.
- Gröger N, Vitzthum H, Fröhlich H, Krüger M, Ehmke H, Braun T, Boettger T. Targeted mutation of SLC4A5 induces arterial hypertension and renal metabolic acidosis. *Hum Mol Genet*. 2012 Mar 1;21(5):1025-36.
- Gross E, Fedotoff O, Pushkin A, Abuladze N, Newman D, Kurtz I. Phosphorylation-induced modulation of pNBC1 function: distinct roles for the amino- and carboxy-termini. *J Physiol*. 2003 Jun 15;549(Pt 3):673-82.
- Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin SY, Glass D, *et al.*; Multiple Tissue Human Expression Resource (MuTHER) Consortium. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012 Oct;44(10):1084-9.
- Guo L, Liu F, Chen S, Yang X, Huang J, He J, Jaquish CE, Zhao Q, Gu CC, Hixson JE, Gu D. Common variants in the Na(+)-coupled bicarbonate transporter genes and salt sensitivity of blood pressure: the GenSalt study. *J Hum Hypertens*. 2015 Nov 19. doi: 10.1038/jhh.2015.113. [Epub ahead of print] PubMed PMID: 26582410.
- Harold FM, Baarda JR. Effects of nigericin and monactin on cation permeability of *Streptococcus faecalis* and metabolic capacities of potassium-depleted cells. *J Bacteriol*. 1968 Mar;95(3):816-23.
- Hatori N, Fine BP, Nakamura A, Cragoe E Jr, Aviv A. Angiotensin II effect on cytosolic pH in cultured rat vascular smooth muscle cells. *J Biol Chem*. 1987 Apr 15;262(11):5073-8.
- Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc*. 2007;2(4):924-32.
- Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res*. 1996 Oct;6(10):986-94.
- Hoch H, Barr GH. Paper electrophoresis with superimposed pH gradient. *Science*. 1955 Aug 5;122(3162):243-4.
- Hollander DH, Nell EE. Improved preservation of *Treponema pallidum* and other bacteria by freezing with glycerol. *Appl Microbiol*. 1954 May;2(3):164-70.
- Hong JH, Yang D, Shcheynikov N, Ohana E, Shin DM, Muallem S. Convergence of IRBIT, phosphatidylinositol (4,5) bisphosphate, and WNK/SPAK kinases in regulation of the Na⁺-HCO₃⁻ cotransporters family. *Proc Natl Acad Sci U S A*. 2013 Mar 5;110(10):4105-10.
- Hong Y, de Faire U, Heller DA, McClearn GE, Pedersen N. Genetic and environmental influences on blood pressure in elderly twins. *Hypertension*. 1994 Dec;24(6):663-70.
- Horie S, Yano S, Watanabe K. Intracellular alkalinization by NH₄Cl increases cytosolic Ca²⁺ level and tension in the rat aortic smooth muscle. *Life Sci*. 1995;56(21):1835-43.

Huan T, Esko T, Peters MJ, Pilling LC, Schramm K, Schurmann C, Chen BH, Liu C, Joehanes R, Johnson AD, Yao C, Ying SX, Courchesne P, Milani L, Raghavachari N, *et al.* A meta-analysis of gene expression signatures of blood pressure and hypertension. *PLoS Genet.* 2015 Mar 18;11(3):e1005035.

Hulikova A, Vaughan-Jones RD, Swietach P. Dual role of CO₂/HCO₃⁻ buffer in the regulation of intracellular pH of three-dimensional tumor growths. *J Biol Chem.* 2011 Apr 22;286(16):13815-26.

Hwang SM, Koo NY, Jin M, Davies AJ, Chun GS, Choi SY, Kim JS, Park K. Intracellular acidification is associated with changes in free cytosolic calcium and inhibition of action potentials in rat trigeminal ganglion. *J Biol Chem.* 2011 Jan 21;286(3):1719-29.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 2004 Feb 11;32(3):1037-49.

Ihalainen S, Soliymani R, Iivanainen E, Mykkänen K, Sainio A, Pöyhönen M, Elenius K, Järveläinen H, Viitanen M, Kalimo H, Baumann M. Proteome analysis of cultivated vascular smooth muscle cells from a CADASIL patient. *Mol Med.* 2007 May-Jun;13(5-6):305-14.

International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011 Sep 11;478(7367):103-9.

International HapMap Project, 2005. Guidelines for Referring to the HapMap Populations in Publications and Presentations. <http://hapmap.ncbi.nlm.nih.gov/citinghapmap.html>. Accessed 7/5/2014.

Ishiguro H, Walther D, Arinami T, Uhl GR. Variation in a bicarbonate co-transporter gene family member SLC4A7 is associated with propensity to addictions: a study using fine-mapping and three samples. *Addiction.* 2007 Aug;102(8):1320-5.

Ivanov V, Roomi MW, Kalinovsky T, Niedzwiecki A, Rath M. Bioflavonoids effectively inhibit smooth muscle cell-mediated contraction of collagen matrix induced by angiotensin II. *J Cardiovasc Pharmacol.* 2005 Nov;46(5):570-6.

Izzard AS, Cragoe EJ Jr, Heagerty AM. Intracellular pH in human resistance arteries in essential hypertension. *Hypertension.* 1991 Jun;17(6 Pt 1):780-6.

Izzard AS, Heagerty AM (a). The measurement of internal pH in resistance arterioles: evidence that intracellular pH is more alkaline in SHR than WKY animals. *J Hypertens.* 1989 Mar;7(3):173-80.

Izzard AS, Heagerty AM (b). Resting intracellular pH in mesenteric resistance arteries from spontaneously hypertensive and Wistar-Kyoto rats: effects of amiloride and 4,4'-diisothiocyanatostilbene-2,2'-disulphonic acid. *J Hypertens Suppl.* 1989 Dec;7(6):S128-9.

Jacobs JW, Niall HD. High sensitivity automated sequence determination of polypeptides. *J Biol Chem.* 1975 May 25;250(10):3629-36.

Jaffe EA, Nachman RL, Becker CG, Minick CR. Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria. *J Clin Invest.* 1973 Nov;52(11):2745-56.

Jakobsen JK, Odgaard E, Wang W, Elkjaer ML, Nielsen S, Aalkjaer C, Leipziger J. Functional up-regulation of basolateral Na⁺-dependent HCO₃⁻ transporter NBCn1 in medullary thick ascending limb of K⁺-depleted rats. *Pflugers Arch.* 2004 Sep;448(6):571-8. Epub 2004 Aug 13.

Jochmann R, Holz P, Sticht H, Stürzl M. Validation of the reliability of computational O-GlcNAc prediction. *Biochem Biophys Acta.* 2014 Feb;1844(2):416-21. doi: 10.1016/j.bbapap.2013.12.002.

- Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008 Dec 15;24(24):2938-9.
- Johnson EM, Theler JM, Capponi AM, Vallotton MB. Characterization of oscillations in cytosolic free Ca²⁺ concentration and measurement of cytosolic Na⁺ concentration changes evoked by angiotensin II and vasopressin in individual rat aortic smooth muscle cells. Use of microfluorometry and digital imaging. *J Biol Chem*. 1991 Jul 5;266(19):12618-26.
- Kanaan A, Douglas RM, Alper SL, Boron WF, Haddad GG. Effect of chronic elevated carbon dioxide on the expression of acid-base transporters in the neonatal and adult mouse. *Am J Physiol Regul Integr Comp Physiol*. 2007 Sep;293(3):R1294-302.
- Karim Z, Attmane-Elakeb A, Sibella V, Bichara M. Acid pH increases the stability of BSC1/NKCC2 mRNA in the medullary thick ascending limb. *J Am Soc Nephrol*. 2003 Sep;14(9):2229-36.
- Karmazyn M, Liu Q, Gan XT, Brix BJ, Fliegel L. Aldosterone increases NHE-1 expression and induces NHE-1-dependent hypertrophy in neonatal rat ventricular myocytes. *Hypertension*. 2003 Dec;42(6):1171-6.
- Karsan A, Yee E, Poirier GG, Zhou P, Craig R, Harlan JM. Fibroblast growth factor-2 inhibits endothelial cell apoptosis by Bcl-2-dependent and independent mechanisms. *Am J Pathol*. 1997 Dec;151(6):1775-84.
- Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, Kelly TN, Saleheen D, Lehne B, Mateo Leach I, Drong AW, Abbott J, Wahl S, Tan ST, Scott WR, *et al*. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet*. 2015 Nov;47(11):1282-93.
- Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X, Tay WT, Chen CH, Zhang Y, Yamamoto K, Katsuya T, Yokota M, Kim YJ, Ong RT, Nabika T, *et al*. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet*. 2011 Jun;43(6):531-8.
- Keller BM, McCarthy AM, Chen J, Armstrong K, Conant EF, Domchek SM, Kontos D. Associations between breast density and a panel of single nucleotide polymorphisms linked to breast cancer risk: a cohort study with digital mammography. *BMC Cancer*. 2015;15(1):1159. doi: 10.1186/s12885-015-1159-3.
- Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*. 2015 Mar;12(3):237-43, 1 p following 243.
- Kim GH, Ecelbarger C, Knepper MA, Packer RK. Regulation of thick ascending limb ion transporter abundance in response to altered acid/base intake. *J Am Soc Nephrol*. 1999 May;10(5):935-42.
- Kim YH, Kwon TH, Christensen BM, Nielsen J, Wall SM, Madsen KM, Frøkiaer J, Nielsen S. Altered expression of renal acid-base transporters in rats with lithium-induced NDI. *Am J Physiol Renal Physiol*. 2003 Dec;285(6):F1244-57. Epub 2003 Aug 26.
- Kimes BW, Brandt BL. Characterization of two putative smooth muscle cell lines from rat thoracic aorta. *Exp Cell Res*. 1976 Mar 15;98(2):349-66. PubMed PMID: 943301.
- Krampetz IK, Rhoades RA. Intracellular pH: effect on pulmonary arterial smooth muscle. *Am J Physiol*. 1991 Jun;260(6 Pt 1):L516-21.
- Kupper N, Willemsen G, Riese H, Posthuma D, Boomsma DI, de Geus EJ. Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertension*. 2005 Jan;45(1):80-5.

Kusche-Vihrog K, Sobczak K, Bangel N, Wilhelmi M, Nechyporuk-Zloy V, Schwab A, Schillers H, Oberleithner H. Aldosterone and amiloride alter ENaC abundance in vascular endothelium. *Pflugers Arch*. 2008 Feb;455(5):849-57.

Kwon TH, Fulton C, Wang W, Kurtz I, Frøkiaer J, Aalkjaer C, Nielsen S. Chronic metabolic acidosis upregulates rat kidney Na-HCO cotransporters NBCn1 and NBC3 but not NBC1. *Am J Physiol Renal Physiol*. 2002 Feb;282(2):F341-51.

Kwon TH, Nielsen J, Kim YH, Knepper MA, Frøkiaer J, Nielsen S. Regulation of sodium transporters in the thick ascending limb of rat kidney: response to angiotensin II. *Am J Physiol Renal Physiol*. 2003 Jul;285(1):F152-65.

Laghmani K, Borensztein P, Ambühl P, Froissart M, Bichara M, Moe OW, Alpern RJ, Paillard M. Chronic metabolic acidosis enhances NHE-3 protein abundance and transport activity in the rat thick ascending limb by increasing NHE-3 mRNA. *J Clin Invest*. 1997 Jan 1;99(1):24-30.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006 Sep 29;313(5795):1929-35.

Larsen AM, Krogsgaard-Larsen N, Lauritzen G, Olesen CW, Honoré Hansen S, Boedtkjer E, Pedersen SF, Bunch L. Gram-scale solution-phase synthesis of selective sodium bicarbonate co-transport inhibitor S0859: in vitro efficacy studies in breast cancer cells. *Chem Med Chem*. 2012 Oct;7(10):1808-14.

Lauritzen G, Jensen MB, Boedtkjer E, Dybboe R, Aalkjaer C, Nylandsted J, Pedersen SF. NBCn1 and NHE1 expression and activity in DeltaNERbB2 receptor-expressing MCF-7 breast cancer cells: contributions to pH regulation and chemotherapy resistance. *Exp Cell Res*. 2010 Sep 10;316(15):2538-53.

Ledoussal C, Lorenz JN, Nieman ML, Soleimani M, Schultheis PJ, Shull GE. Renal salt wasting in mice lacking NHE3 Na⁺/H⁺ exchanger but not in mice lacking NHE2. *Am J Physiol Renal Physiol*. 2001 Oct;281(4):F718-27.

Ledoussal C, Woo AL, Miller ML, Shull GE. Loss of the NHE2 Na⁺/H⁺ exchanger has no apparent effect on diarrheal state of NHE3-deficient mice. *Am J Physiol Gastrointest Liver Physiol*. 2001 Dec;281(6):G1385-96.

Lee CY, Tsai YT, Chang CY, Chang YY, Cheng TH, Tsai CS, Loh SH. Functional characterization of intracellular pH regulators responsible for acid extrusion in human radial artery smooth muscle cells. *Chin J Physiol*. 2014 Oct 31;57(5):238-48.

Lee S, Lee HJ, Yang HS, Thornell IM, Bevenssee MO, Choi I. Sodium-bicarbonate cotransporter NBCn1 in the kidney medullary thick ascending limb cell line is upregulated under acidic conditions and enhances ammonium transport. *Exp Physiol*. 2010 Sep;95(9):926-37.

Lee S, Mele M, Vahl P, Christiansen PM, Jensen VE, Boedtkjer E. Na⁺,HCO₃⁻ cotransport is functionally upregulated during human breast carcinogenesis and required for the inverted pH gradient across the plasma membrane. *Pflugers Arch*. 2014 May 2.

Leibiger C, Kosyakova N, Mkrtchyan H, Gleib M, Trifonov V, Liehr T. First molecular cytogenetic high resolution characterization of the NIH 3T3 cell line by murine multicolor banding. *J Histochem Cytochem*. 2013 Apr;61(4):306-12.

Leik CE, Willey A, Graham MF, Walsh SW. Isolation and culture of arterial smooth muscle cells from human placenta. *Hypertension*. 2004 Apr;43(4):837-40.

Lemeer S, Heck AJ. The phosphoproteomics data explosion. *Curr Opin Chem Biol*. 2009 Oct;13(4):414-20. doi: 10.1016/j.cbpa.2009.06.022. Epub 2009 Jul 19. Review.

Lesburg C, Li S, Cragoe EJ Jr, Deth RC. Influence of amiloride derivatives on alpha-1 adrenergic receptor-induced contractions of the rabbit aorta. *J Pharmacol Exp Ther*. 1990 May;253(2):530-6.

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Köttgen A, Vasan RS, Rivadeneira F, *et al*. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009 Jun;41(6):677-87.

Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002 Dec 14;360(9349):1903-13. Erratum in: *Lancet*. 2003 Mar 22;361(9362):1060.

Li XC, Zhuo JL. Intracellular ANG II directly induces in vitro transcription of TGF-beta1, MCP-1, and NHE-3 mRNAs in isolated rat renal cortical nuclei via activation of nuclear AT1a receptors. *Am J Physiol Cell Physiol*. 2008 Apr;294(4):C1034-45.

Li Y, Jiang Z, Chen H, Ma WJ. A modified quantitative EMSA and its application in the study of RNA--protein interactions. *J Biochem Biophys Methods*. 2004 Aug 31;60(2):85-96.

Liu Y, Qin X, Wang DK, Guo YM, Gill HS, Morris N, Parker MD, Chen LM, Boron WF. Effects of optional structural elements, including two alternative amino termini and a new splicing cassette IV, on the function of NBCn1 (SLC4A7). *J Physiol*. 2013 Aug 19. Oct 15;591(Pt 20):4983-5004.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCt} method. *Methods* 2001; 25: 402-408.

Loganathan SK, Casey JR. Corneal dystrophy-causing SLC4A11 mutants: suitability for folding-correction therapy. *Hum Mutat*. 2014 Sep;35(9):1082-91.

Loiselle FB, Morgan PE, Alvarez BV, Casey JR. Regulation of the human NBC3 Na⁺/HCO₃⁻ cotransporter by carbonic anhydrase II and PKA. *Am J Physiol Cell Physiol*. 2004 Jun;286(6):C1423-33.

Losse H, Zidek W, Vetter H. Intracellular sodium and calcium in vascular smooth muscle of spontaneously hypertensive rats. *J Cardiovasc Pharmacol*. 1984;6 Suppl 1:S32-4.

Lu X, Wang L, Lin X, Huang J, Charles Gu C, He M, Shen H, He J, Zhu J, Li H, Hixson JE, Wu T, Dai J, Lu L, Shen C, *et al*. Genome-wide association study in Chinese identifies novel loci for blood pressure and hypertension. *Hum Mol Genet*. 2015 Feb 1;24(3):865-74.

Maheswaran R, Gill JS, Beevers DG. Blood pressure and industrial lead exposure. *Am J Epidemiol*. 1993 Mar 15;137(6):645-53.

Mandel M, Higa A. Calcium-dependent bacteriophage DNA infection. *J Mol Biol*. 1970 Oct 14;53(1):159-62.

Matsui H, Barry WH, Livsey C, Spitzer KW. Angiotensin II stimulates sodium-hydrogen exchange in adult rabbit ventricular myocytes. *Cardiovasc Res*. 1995 Feb;29(2):215-21..

Matsumoto H, Moir LM, Oliver BG, Burgess JK, Roth M, Black JL, McParland BE. Comparison of gel contraction mediated by airway smooth muscle cells from patients with and without asthma. *Thorax*. 2007 Oct;62(10):848-54.

- Matsushita M, Tanaka H, Mitsui K, Kanazawa H. Dual functional significance of calcineurin homologous protein 1 binding to Na⁺/H⁺ exchanger isoform 1. *Am J Physiol Cell Physiol*. 2011 Aug;301(2):C280-8.
- Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, *et al*. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003 Jan 1;31(1):374-8.
- Mazur P. Freezing of living cells: mechanisms and implications. *Am J Physiol*. 1984 Sep;247(3 Pt 1):C125-42. Review.
- M'Buyamba-Kabangu JR, Lepira B, Lijnen P, Tshiani K, Fagard R, Amery A. Intracellular sodium and the response to nitrendipine or atenolol in African blacks. *Hypertension*. 1988 Jan;11(1):100-5.
- McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000 Apr;16(4):404-5.
- McLane LM, Banerjee PP, Cosma GL, Makedonas G, Wherry EJ, Orange JS, Betts MR. Differential localization of T-bet and Eomes in CD8 T cell memory populations. *J Immunol*. 2013 Apr 1;190(7):3207-15.
- Mentzer RM Jr, Bartels C, Bolli R, Boyce S, Buckberg GD, Chaitman B, Haverich A, Knight J, Menasché P, Myers ML, Nicolau J, Simoons M, Thulin L, Weisel RD; EXPEDITION Study Investigators. Sodium-hydrogen exchange inhibition by cariporide to reduce the risk of ischemic cardiac events in patients undergoing coronary artery bypass grafting: results of the EXPEDITION study. *Ann Thorac Surg*. 2008 Apr;85(4):1261-70.
- Michea L, Delpiano AM, Hitschfeld C, Lobos L, Lavandero S, Marusic ET. Eplerenone blocks nongenomic effects of aldosterone on the Na⁺/H⁺ exchanger, intracellular Ca²⁺ levels, and vasoconstriction in mesenteric resistance vessels. *Endocrinology*. 2005 Mar;146(3):973-80.
- Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menéndez-Rodríguez P, Hardisson D, Mendiola M, González-Neira A, Pita G, Alonso MR, Dennis J, Wang Q, Bolla MK, Swerdlow A, *et al*. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*. 2014 Nov 15;23(22):6096-111.
- Misik AJ, Perreault K, Holmes CF, Fliegel L. Protein phosphatase regulation of Na⁺/H⁺ exchanger isoform I. *Biochemistry*. 2005 Apr 19;44(15):5842-52.
- Miyata Y, Muto S, Kusano E. Mechanisms for nongenomic and genomic effects of aldosterone on Na⁺/H⁺ exchange in vascular smooth muscle cells. *J Hypertens*. 2005 Dec;23(12):2237-50.
- Møller L, Kristensen TS. Blood lead as a cardiovascular risk factor. *Am J Epidemiol*. 1992 Nov 1;136(9):1091-100.
- Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci*. 1995 Sep-Oct;1(5):319-29.
- Moniz LS, Stambolic V. Nek10 mediates G2/M cell cycle arrest and MEK autoactivation in response to UV irradiation. *Mol Cell Biol*. 2011 Jan;31(1):30-42.
- Mulvany MJ, Aalkjaer C, Petersen TT. Intracellular sodium, membrane potential, and contractility of rat mesenteric small arteries. *Circ Res*. 1984 Jun;54(6):740-9. Munroe PB, Barnes MR, Caulfield MJ. Advances in blood pressure genomics. *Circ Res*. 2013 May 10;112(10):1365-79.

- Musch MW, Lucioni A, Chang EB. Aldosterone regulation of intestinal Na absorption involves SGK-mediated changes in NHE3 and Na⁺ pump activity. *Am J Physiol Gastrointest Liver Physiol*. 2008 Nov;295(5):G909-19.
- Muto S, Ebata S, Okada K, Saito T, Asano Y. Glucocorticoid modulates Na⁺/H⁺ exchange activity in vascular smooth muscle cells by nongenomic and genomic mechanisms. *Kidney Int*. 2000 Jun;57(6):2319-33.
- Namkoong E, Shin YH, Bae JS, Choi S, Kim M, Kim N, Hwang SM, Park K. Role of Sodium Bicarbonate Cotransporters in Intracellular pH Regulation and Their Regulatory Mechanisms in Human Submandibular Glands. *PLoS One*. 2015 Sep 16;10(9):e0138368.
- Neumann E, Schaefer-Ridder M, Wang Y, Hofschneider PH. Gene transfer into mouse lymphoma cells by electroporation in high electric fields. *EMBO J*. 1982;1(7):841-5.
- Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, *et al*. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009 Jun;41(6):666-76.
- Neylon CB, Avdonin PV, Dilley RJ, Larsen MA, Tkachuk VA, Bobik A. Different electrical responses to vasoactive agonists in morphologically distinct smooth muscle cell types. *Circ Res*. 1994 Oct;75(4):733-41.
- Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, *et al*.; MuTHER Consortium. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*. 2011 Feb 3;7(2):e1002003.
- Nielsen J, Kwon TH, Masilamani S, Beutler K, Hager H, Nielsen S, Knepper MA. Sodium transporter abundance profiling in kidney: effect of spironolactone. *Am J Physiol Renal Physiol*. 2002 Nov;283(5):F923-33.
- Nowik M, Lecca MR, Velic A, Rehrauer H, Brändli AW, Wagner CA. Genome-wide gene expression profiling reveals renal genes regulated during metabolic acidosis. *Physiol Genomics*. 2008 Feb 19;32(3):322-34
- Obara-Ishihara T, Kuhlman J, Niswander L, Herzlinger D. The surface ectoderm is essential for nephric duct formation in intermediate mesoderm. *Development*. 1999 Mar; 126(6):1103-8.
- Orlowski A, Vargas LA, Aiello EA, Álvarez BV. Elevated carbon dioxide upregulates NBCn1 Na⁺/HCO₃⁻ cotransporter in human embryonic kidney cells. *Am J Physiol Renal Physiol*. 2013 Dec 15;305(12):F1765-74.
- Osanai T, Kanazawa T, Okuguchi T, Kamada T, Metoki H, Oike Y, Onodera K. Sodium ionophore converts growth manner of vascular smooth muscle cells from spontaneously hypertensive rats. *Cardiovasc Res*. 1996 Jan;31(1):124-31.
- Pang T, Hisamitsu T, Mori H, Shigekawa M, Wakabayashi S. Role of calcineurin B homologous protein in pH regulation by the Na⁺/H⁺ exchanger 1: tightly bound Ca²⁺ ions as important structural elements. *Biochemistry*. 2004 Mar 30;43(12):3628-36.
- Pang T, Su X, Wakabayashi S, Shigekawa M. Calcineurin homologous protein as an essential cofactor for Na⁺/H⁺ exchangers. *J Biol Chem*. 2001 May 18;276(20):17367-72.
- Pang T, Wakabayashi S, Shigekawa M. Expression of calcineurin B homologous protein 2 protects serum deprivation-induced cell death by serum-independent activation of Na⁺/H⁺ exchanger. *J Biol Chem*. 2002 Nov 15;277(46):43771-7.

- Pardanaud L, Yassine F, Dieterlen-Lievre F. Relationship between vasculogenesis, angiogenesis and haemopoiesis during avian ontogeny. *Development*. 1989 Mar;105(3):473-85.
- Pardee AB. A restriction point for control of normal animal cell proliferation. *Proc Natl Acad Sci U S A*. 1974 Apr;71(4):1286-90.
- Park HJ, Rajbhandari I, Yang HS, Lee S, Cucoranu D, Cooper DS, Klein JD, Sands JM, Choi I. Neuronal expression of sodium/bicarbonate cotransporter NBCn1 (SLC4A7) and its response to chronic metabolic acidosis. *Am J Physiol Cell Physiol*. 2010 May;298(5):C1018-28.
- Pedersen SF. The Na⁺/H⁺ exchanger NHE1 in stress-induced signal transduction: implications for cell proliferation and cell death. *Pflugers Arch*. 2006 Jun;452(3):249-59.
- Pickkers P, van Beek M, Hughes AD, Russel FG, Thien T, Smits P. Presence and mechanism of direct vascular effects of amiloride in humans. *J Cardiovasc Pharmacol*. 1999 Sep;34(3):388-93.
- Praetorius J, Hager H, Nielsen S, Aalkjaer C, Friis UG, Ainsworth MA, Johansen T. Molecular and functional evidence for electrogenic and electroneutral Na⁺-HCO₃⁻ cotransporters in murine duodenum. *Am J Physiol Gastrointest Liver Physiol*. 2001 Mar;280(3):G332-43.
- Praetorius J, Kim YH, Bouzinova EV, Frische S, Rojek A, Aalkjaer C, Nielsen S. NBCn1 is a basolateral Na⁺-HCO₃⁻ cotransporter in rat kidney inner medullary collecting ducts. *Am J Physiol Renal Physiol*. 2004 May;286(5):F903-12.
- Pushkin A, Abuladze N, Lee I, Newman D, Hwang J, Kurtz I. Cloning, tissue distribution, genomic organization, and functional characterization of NBC3, a new member of the sodium bicarbonate cotransporter family. *J Biol Chem*. 1999 Jun 4;274(23):16569-75.
- Quilty JA, Li J, Reithmeier RA. Impaired trafficking of distal renal tubular acidosis mutants of the human kidney anion exchanger kAE1. *Am J Physiol Renal Physiol*. 2002 May;282(5):F810-20.
- Rajan N, Habermehl J, Coté MF, Doillon CJ, Mantovani D. Preparation of ready-to-use, storable and reconstituted type I collagen from rat tail tendon for tissue engineering applications. *Nat Protoc*. 2006;1(6):2753-8.
- Renart J, Reiser J, Stark GR. Transfer of proteins from gels to diazobenzoyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proc Natl Acad Sci U S A*. 1979 Jul;76(7):3116-20.
- Resnick LM, Gupta RK, Sosa RE, Corbett ML, Laragh JH. Intracellular pH in human and experimental hypertension. *Proc Natl Acad Sci U S A*. 1987 Nov;84(21):7663-7. Erratum in: *Proc Acad Sci U S A* 1988 Apr;85(7):2367.
- Rink TJ, Tsien RY, Pozzan T. Cytoplasmic pH and free Mg²⁺ in lymphocytes. *J Cell Biol*. 1982 Oct;95(1):189-96.
- Risau W, Sariola H, Zerwes HG, Sasse J, Ekblom P, Kemler R, Doetschman T. Vasculogenesis and angiogenesis in embryonic-stem-cell-derived embryoid bodies. *Development*. 1988 Mar;102(3):471-8.
- Robinson BF, Phillips RJ, Wilson PN, Chiodini PL. Effect of local infusion of ouabain on human forearm vascular resistance and on response to potassium, verapamil and sodium nitroprusside. *J Hypertens*. 1983 Aug;1(2):165-9.
- Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, Castagné R, Deiseroth A, Proust C, Brocheton J, Godefroy T, Perret C, Germain M, Eleftheriadis M, Sinning CR, et al. Integrating genome-wide

genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet.* 2011 Dec;7(12):e1002367.

Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010 Apr;5(4):725-38.

Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC, Wilson V, Evans MJ. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature.* 2000 Mar 2;404(6773):95-9.

Rust HL, Thompson PR. Kinase consensus sequences: a breeding ground for crosstalk. *ACS Chem Biol.* 2011 Sep 16;6(9):881-92

Sagnella GA, Miller MA, Khong TK, MacGregor GA. Platelet sodium/hydrogen exchanger activity in normotensives and hypertensives. *Clin Chim Acta.* 1999 May;283(1-2):105-18.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science.* 1985 Dec 20;230(4732):1350-4.

Sainio A, Jokela T, Tammi MI, Järveläinen H. Hyperglycemic conditions modulate connective tissue reorganization by human vascular smooth muscle cells through stimulation of hyaluronan synthesis. *Glycobiology.* 2010 Sep;20(9):1117-26.

Salvi E, Kuznetsova T, Thijs L, Lupoli S, Stolarz-Skrzypek K, D'Avila F, Tikhonoff V, De Astis S, Barcella M, Seidlerová J, Benaglio P, Malyutina S, Frau F, Velayutham D, Benfante R, *et al.* Target sequencing, cell experiments, and a population study establish endothelial nitric oxide synthase (eNOS) gene as hypertension susceptibility gene. *Hypertension.* 2013 Nov;62(5):844-52.

Sandmann S, Yu M, Kaschina E, Blume A, Bouzinova E, Aalkjaer C, Unger T. Differential effects of angiotensin AT1 and AT2 receptors on the expression, translation and function of the Na⁺-H⁺ exchanger and Na⁺-HCO₃⁻ symporter in the rat heart after myocardial infarction. *J Am Coll Cardiol.* 2001 Jun 15;37(8):2154-65.

Schultheis PJ, Clarke LL, Meneton P, Miller ML, Soleimani M, Gawenis LR, Riddle TM, Duffy JJ, Doetschman T, Wang T, Giebisch G, Aronson PS, Lorenz JN, Shull GE. Renal and intestinal absorptive defects in mice lacking the NHE3 Na⁺/H⁺ exchanger. *Nat Genet.* 1998 Jul;19(3):282-5.

Scinicariello F, Abadin HG, Murray HE. Association of low-level blood lead and blood pressure in NHANES 1999-2006. *Environ Res.* 2011 Nov;111(8):1249-57.

Shaw G, Morse S, Ararat M, Graham FL. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* 2002 Jun;16(8):869-71.

Shi Q, Aida K, Vandenberg JL, Wang XL. Passage-dependent changes in baboon endothelial cells--relevance to in vitro aging. *DNA Cell Biol.* 2004 Aug;23(8):502-9.

Shirakabe K, Priori G, Yamada H, Ando H, Horita S, Fujita T, Fujimoto I, Mizutani A, Seki G, Mikoshiba K. IRBIT, an inositol 1,4,5-trisphosphate receptor-binding protein, specifically binds to and activates pancreas-type Na⁺/HCO₃⁻ cotransporter 1 (pNBC1). *Proc Natl Acad Sci U S A.* 2006 Jun 20;103(25):9542-7.

Shirley MD, Baugher JD, Stevens EL, Tang Z, Gerry N, Beiswanger CM, Berlin DS, Pevsner J. Chromosomal variation in lymphoblastoid cell lines. *Hum Mutat.* 2012 Jul;33(7):1075-86.

- Silveira EA, Lizardo JH, Souza LP, Stefanon I, Vassallo DV. Acute lead-induced vasoconstriction in the vascular beds of isolated perfused rat tails is endothelium-dependent. *Braz J Med Biol Res.* 2010 May;43(5):492-9.
- Smith AJ, Howard P, Shah S, Eriksson P, Stender S, Giambartolomei C, Folkersen L, Tybjærg-Hansen A, Kumari M, Palmen J, Hingorani AD, Talmud PJ, Humphries SE. Use of allele-specific FAIRE to determine functional regulatory polymorphism using large-scale genotyping arrays. *PLoS Genet.* 2012;8(8):e1002908.
- Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. Measurement of protein using bicinchoninic acid. *Anal Biochem.* 1985 Oct;150(1):76-85. Erratum in: *Anal Biochem* 1987 May 15;163(1):279.
- Soleimani M, Singh G, Bookstein C, Rao MC, Chang EB, Dominguez JH. Inhibition of glycosylation decreases Na⁺/H⁺ exchange activity, blocks NHE-3 transport to the membrane, and increases NHE-3 mRNA expression in LLC-PK1 cells. *J Lab Clin Med.* 1996 Jun;127(6):565-73.
- Stentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 2013 May 15;32(10):1478-88.
- Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.* 2012;8(4):e1002639.
- Sung YJ, de Las Fuentes L, Schwander KL, Simino J, Rao DC. Gene-smoking interactions identify several novel blood pressure Loci in the framingham heart study. *Am J Hypertens.* 2015 Mar;28(3):343-54.
- Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. *Clin Chem.* 2004 Nov;50(11):1974-8.
- Thomas RC. Intracellular pH of snail neurones measured with a new pH-sensitive glass micro-electrode. *J Physiol.* 1974 Apr;238(1):159-80.
- Thomsen AB, Kim S, Aalbaek F, Aalkjaer C, Boedtkjer E. Intracellular acidification alters myogenic responsiveness and vasomotion of mouse middle cerebral arteries. *J Cereb Blood Flow Metab.* 2014 Jan;34(1):161-8.
- Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011 Apr;32(4):358-68.
- Touyz RM, Schiffrin EL. Effects of angiotensin II and endothelin-1 on platelet aggregation and cytosolic pH and free Ca²⁺ concentrations in essential hypertension. *Hypertension.* 1993 Dec;22(6):853-62.
- Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A.* 1979 Sep;76(9):4350-4.
- Tower Hamlets Council, 2012. 2011 Census: Second Release Headline Analysis. Available online: <http://www.towerhamlets.gov.uk/idoc.ashx?docid=e01ef46e-4740-4fe2-9853-5187d150ea41&version=-1>. Accessed 29/8/2013.
- Toye AM, Parker MD, Daly CM, Lu J, Virkki LV, Pelletier MF, Boron WF. The human NBCe1-A mutant R881C, associated with proximal renal tubular acidosis, retains function but is mistargeted in polarized renal epithelia. *Am J Physiol Cell Physiol.* 2006 Oct;291(4):C788-801.

Tragante V, Barnes MR, Ganesh SK, Lanktree MB, Guo W, Franceschini N, Smith EN, Johnson T, Holmes MV, Padmanabhan S, Karczewski KJ, Almqvister B, Barnard J, Baumert J, Chang YP, *et al.* Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. *Am J Hum Genet.* 2014 Mar 6;94(3):349-60.

Trevisi L, Cargnelli G, Ceolotto G, Papparella I, Semplicini A, Zampella A, D'Auria MV, Luciani S. Callipeltin A: sodium ionophore effect and tension development in vascular smooth muscle. *Biochem Pharmacol.* 2004 Oct 1;68(7):1331-8.

Trudu M, Janas S, Lanzani C, Debaix H, Schaeffer C, Ikehata M, Citterio L, Demarets S, Trevisani F, Ristagno G, Glaudemans B, Laghmani K, Dell'Antonio G; Swiss Kidney Project on Genes in Hypertension (SKIPOGH) team, Loffing J, Rastaldi MP, *et al.* Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med* 2013, 19:1655-1660.

van der Heijden T, van Vugt JJ, Logie C, van Noort J. Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome occupancy. *Proc Natl Acad Sci U S A.* 2012 Sep 18;109(38):E2514-22.

Vorum H, Kwon TH, Fulton C, Simonsen B, Choi I, Boron W, Maunsbach AB, Nielsen S, Aalkjaer C. Immunolocalization of electroneutral Na-HCO₃⁻ cotransporter in rat kidney. *Am J Physiol Renal Physiol.* 2000;279:F901-9.

Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 2012 Dec;131(4):281-5.

Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, Bochud M, Rice KM, Henneman P, Smith AV, Ehret GB, Amin N, Larson MG, Mooser V, Hadley D, *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet.* 2011 Sep 11;43(10):1005-11.

Wang AM, Doyle MV, Mark DF. Quantitation of mRNA by the polymerase chain reaction. *Proc Natl Acad Sci U S A.* 1989 Dec;86(24):9717-21. Erratum in: *Proc Natl Acad Sci U S A* 1990 Apr;87(7):2865.

Watson AJ, Levine S, Donowitz M, Montrose MH. Serum regulates Na⁺/H⁺ exchange in Caco-2 cells by a mechanism which is dependent on F-actin. *J Biol Chem.* 1992 Jan 15;267(2):956-62. Erratum in: *J Biol Chem* 1993 Feb 5;268(4):3016.

Weed RG, Jenkins EC. Slow versus fast postthaw dilution for the recovery of lymphocytes in whole blood frozen in either dimethyl sulfoxide or glycerol. *Exp Cell Biol.* 1979;47(5):351-9.

Weissberg PL, Little PJ, Cragoe EJ Jr, Bobik A. Na-H antiport in cultured rat aortic smooth muscle: its role in cytoplasmic pH regulation. *Am J Physiol.* 1987 Aug;253(2 Pt 1):C193-8.

Whitfield JB, Dy V, McQuilty R, Zhu G, Heath AC, Montgomery GW, Martin NG. Genetic effects on toxic and essential elements in humans: arsenic, cadmium, copper, lead, mercury, selenium, and zinc in erythrocytes. *Environ Health Perspect.* 2010 Jun;118(6):776-82.

Whitfield JB, Dy V, McQuilty R, Zhu G, Montgomery GW, Ferreira MA, Duffy DL, Neale MC, Heijmans BT, Heath AC, Martin NG. Evidence of genetic effects on blood lead concentration. *Environ Health Perspect.* 2007 Aug;115(8):1224-30.

Wirth R, Friesenegger A, Fiedler S. Transformation of various species of gram-negative bacteria belonging to 11 different genera by electroporation. *Mol Gen Genet.* 1989 Mar;216(1):175-7.

World Health Organisation, 2009. Global Health Risks: Mortality and burden of disease attributable to selected major risks.

http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf. Accessed 9/5/2014.

Wu C, Gong Y, Sun A, Zhang Y, Zhang C, Zhang W, Zhao G, Zou Y, Ge J. The human MTHFR rs4846049 polymorphism increases coronary heart disease risk through modifying miRNA binding. *Nutr Metab Cardiovasc Dis*. 2013 Jul;23(7):693-8.

Wu S, Song T, Zhou S, Liu Y, Chen G, Huang N, Liu L. Involvement of Na⁺/H⁺ exchanger 1 in advanced glycation end products-induced proliferation of vascular smooth muscle cell. *Biochem Biophys Res Commun*. 2008 Oct 24;375(3):384-9.

Yamamoto T, Shirayama T, Sakatani T, Takahashi T, Tanaka H, Takamatsu T, Spitzer KW, Matsubara H. Enhanced activity of ventricular Na⁺-HCO₃⁻-cotransport in pressure overload hypertrophy. *Am J Physiol Heart Circ Physiol*. 2007 Aug;293(2):H1254-64.

Yamamoto D, Suzuki N. Blockage of chloride channels by HEPES buffer. *Proc R Soc Lond B Biol Sci*. 1987 Feb 23;230(1258):93-100.

Yamamoto T, Swietach P, Rossini A, Loh SH, Vaughan-Jones RD, Spitzer KW. Functional diversity of electrogenic Na⁺-HCO₃⁻ cotransport in ventricular myocytes from rat, rabbit and guinea pig. *J Physiol*. 2005 Jan 15;562(Pt 2):455-75.

Yang B, Fan S, Zhi X, Li Y, Liu Y, Wang D, He M, Hou Y, Zheng Q, Sun G. Associations of MTHFR gene polymorphisms with hypertension and hypertension in pregnancy: a meta-analysis from 114 studies with 15411 cases and 21970 controls. *PLoS One*. 2014 Feb 5;9(2):e87497.

Yang D, Li Q, So I, Huang CL, Ando H, Mizutani A, Seki G, Mikoshiba K, Thomas PJ, Muallem S. IRBIT governs epithelial secretion in mice by antagonizing the WNK/SPAK kinase pathway. *J Clin Invest*. 2011 Mar;121(3):956-65.

Yang HS, Cooper DS, Rajbhandari I, Park HJ, Lee S, Choi I. Inhibition of rat Na⁺(-)-HCO₃⁻ cotransporter (NBCn1) function and expression by the alternative splice domain. *Exp Physiol*. 2009 Nov;94(11):1114-23.

Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*. 2010 Oct 1;26(19):2474-6.

Ye M, Flores G, Batlle D. Angiotensin II and angiotensin-(1-7) effects on free cytosolic sodium, intracellular pH, and the Na⁺-H⁺ antiporter in vascular smooth muscle. *Hypertension*. 1996 Jan;27(1):72-8.

Young M, Funder J. Mineralocorticoid action and sodium-hydrogen exchange: studies in experimental cardiac fibrosis. *Endocrinology*. 2003 Sep;144(9):3848-51.

Yu L, Quinn DA, Garg HG, Hales CA. Deficiency of the NHE1 gene prevents hypoxia-induced pulmonary hypertension and vascular remodeling. *Am J Respir Crit Care Med*. 2008 Jun 1;177(11):1276-84

Zaun HC, Shrier A, Orlowski J. Calcineurin B homologous protein 3 promotes the biosynthetic maturation, cell surface stability, and optimal transport of the Na⁺/H⁺ exchanger NHE1 isoform. *J Biol Chem*. 2008 May 2;283(18):12456-67.

Zhang LF, Peng SQ, Wang S, Li BL, Han G, Dong YS. Direct effects of lead (Pb²⁺) on the relaxation of in vitro cultured rat aorta to acetylcholine. *Toxicol Lett*. 2007 Apr 25;170(2):104-10.