Characterising novel genetic causes of growth failure

Dr Emily Jane Hamilton Cottrell

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Supervisors: Professor Helen L. Storr
Professor Louise A. Metherell

Centre for Endocrinology
William Harvey Research Institute
Barts and the London School of Medicine and Dentistry
Queen Mary University of London

Statement of originality

I, Emily Jane Hamilton Cottrell, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Date: 24/01/2023

Abstract

To identify novel genetic causes of growth failure, I developed a unique, targeted whole gene panel for rapid and accurate genetic testing of patients with short stature and features of Growth Hormone Insensitivity (GHI) or unexplained short stature. This included 64 genes of interest, including those in the GH-IGF1 pathway and genes linked to conditions with overlapping features. In parallel, I also assessed these patients for copy number variants. Using custom bioinformatic pipelines to filter these data sets and a variety of *in silico* prediction programs, I identified interesting novel genetic defects in both known and candidate growth genes. I then performed functional analysis of these defects to determine if they affected gene structure/function and could explain the patient phenotype.

I identified several novel splicing mutations in the *Growth Hormone Receptor (GHR)* causing a spectrum of GHI. These include a novel mutation deep within intron 6 *GHR* that leads to mis-splicing and pseudoexon inclusion. Pseudoexon inclusion leads to frameshift of the *GHR* and thus causes a non-functional Growth Hormone Receptor and severe GHI. I discovered two novel heterozygous *GHR* mutations in patients with milder GHI phenotypes. These mutations both led to mis-splicing of exon 9 of the *GHR* and act in a dominant negative effect on the GHR, reducing the efficacy of signalling and explaining their milder phenotypes. I identified a rare novel heterozygous *IGF1* variant that I hypothesised would impair IGF-1 cleavage causing functional IGF-1 deficiency. Our patient cohort was enriched for low frequency CNVs, particularly in patients with subtle features of Silver Russell Syndrome. This is the first study to assess CNVs in patients with GHI. From my CNV analysis, I identified *CHD1L* and *HMGA2* as key candidate growth genes and functionally assessed several patient variants identified within our cohort.

Collaborators

Dr Vivian Hwa, Cincinnati Children's Hospital Medical Center

Vivian has provided expert input onto many aspects of my doctorate research. I also spent 4 months under her supervision at the Cincinnati Children's Hospital Medical Center, Cincinnati, USA, functionally assessing the novel *GHR*, *IGF1* and *HMGA2* variants I had identified.

Professors Thomas Klonisch and Sabine Hombach-Klonisch, University of Mannitoba

Thomas and Sabine are experts on the role of *HMGA2* in cancer. They have collaborated on our *HMGA2* work, helping to functionally assess the novel variants identified and generated novel transgenic mice with heterozygous and homozygous c.166A>G p.(Lys56Glu) mutations (*Hmga2*^{K56E}) to further study the effect of this novel variant.

Department of Clinical Genetics, Leiden University Medical Centre

We have collaborated with several members of the Clinical Genetics team at Leiden University Medical Centre, Leiden, The Netherlands, especially Hermine van Duyvenvoorde, Clinical Geneticist and Dr Christiaan de Bruin, Paediatric Endocrinologist. They have identified several patients with novel *HMGA2* mutations and arranged for fibroblast samples from these kindreds to be sent to our centre for functional analysis.

Dr Martin Bidlingmaier, LMU Klinikum

Dr Bidlingmaier and his team at the Endocrine Laboratory, LMU Klinikum (Munich, Germany) performed biochemical assays for the families with novel *GHR* variants and novel c.356G>A *IGF1* variant.

Dr Claudia Cabrera, Bioinformatician, QMUL

Pathway enrichment analysis was performed by Dr Claudia Cabrera using Ingenuity Pathway Analysis (IPA) software (Qiagen, inc). IPA identified biological pathways and functions that were enriched within the CNV regions (CNV gene lists) identified in our GHI and unexplained short stature cohorts.

Dr Sumana Chatterjee, Clinical Research Fellow, QMUL

Phenotypic predictors (height SDS, age, sex, BW SDS or IGF-1 SDS) associated with the identification of subjects with and without CNVs in the cohort were compared by Dr Sumana Chatterjee using 2-tailed t-tests and logistic regression analysis. The frequency of deletions vs duplications and the size of the CNVs between the 2 cohorts (GHI and unexplained short stature groups) were analysed by Fisher's exact t test and Mann-Whitney u test, respectively.

Dr Tasneem Ladha, MsC Student, QMUL

Whilst working with our group on her MsC project, Tasneem performed splicing assays for Patients 6, 7 and 8 with novel *GHR* variants.

Publications arising from this thesis

Emily Cottrell, Avinaash Maharaj, Jack Williams, Sumana Chatterjee, Grazia Cirillo, Emanuele Miraglia del Giudice, Adalgisa Festa, Stefania Palumbo, Donatella Capalbo, Mariacarolina Salerno, Claudio Pignata, Martin O Savage, Katharina Schilbach, Martin Bidlingmaier, Vivian Hwa, Louise A Metherell, Anna Grandone, Helen L Storr. Growth Hormone Receptor *(GHR)* 6Ω Pseudoexon Activation: A Novel Cause Of Severe Growth Hormone Insensitivity (GHI), *The Journal of Clinical Endocrinology & Metabolism*, 2021; dgab550, https://doi.org/10.1210/clinem/dgab550

Cottrell E, Cabrera CP, Ishida M, Chatterjee S, Greening J, Wright N, Bossowski A, Dunkel L, Deeb A, Basiri IA, Rose SJ, Mason A, Bint S, Ahn JW, Hwa V, Metherell LA, Moore GE, Storr HL. Rare CNVs provide novel insights into the molecular basis of GH and IGF-1 insensitivity. *Eur J Endocrinol.* 2020 Dec;183(6):581-595. Doi: 10.1530/EJE-20-0474.

Afiya Andrews, Avinaash Maharaj, **Emily Cottrell**, Sumana Chatterjee, Pratik Shah, Louise Denvir, Katja Dumic, Artur Bossowski, Talat Mushtaq, Rade Vukovic, Mohamed Didi, Nick Shaw, Louise A Metherell, Martin O Savage, Helen L Storr. Genetic characterization of short stature patients with overlapping features of growth hormone insensitivity syndromes, *The Journal of Clinical Endocrinology & Metabolism*, 2021; dgab437, https://doi.org/10.1210/clinem/dgab437

Chatterjee S, **Cottrell E**, Rose SJ, Mushtaq T, Maharaj AV, Williams J, Savage MO, Metherell LA, Storr H. *GHR* gene transcript heterogeneity may explain phenotypic variability in *GHR* pseudoexon (6Ψ) patients. *Endocr Connect*. 2020 Feb 1;9(3):211-22. Doi: 10.1530/EC-20-0026.

Afiya Andrews, **Emily Cottrell**, Avinaash Maharaj, Tasneem Ladha, Jack Williams, Louise A Metherell, Peter J McCormick*, Helen L Storr* 'Characterisation of dominant-negative GH receptor variants reveals a potential therapeutic target for short stature' *Eur J Endocrinol*. April 2023. DOI: 10.1093/ejendo/lvad039 *Joint senior authorship

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to the following people who have all played such important roles in my PhD journey.

I would like to thank my husband and family for their unwavering support and encouragement throughout this PhD project.

I would like to thank my primary supervisor Professor Helen Storr, for the opportunity to be part of such an exciting, cutting-edge project and for always pushing me to achieve my maximum potential. To my secondary supervisor Professor Lou Metherell, for encouraging a fascination of genetics and for helping me to explore the vast genetic landscape. To Dr Vivian Hwa, for her wonderful supervision during my 4-month placement at Cincinnati Children's Hospital Medical Center and for her expert contributions to my project.

I am grateful Dr Jack Williams for his patience, guidance and excellent advice throughout my project. To Dr Afiya Andrews and the members of my lab group for sharing their scientific knowledge and enthusiasm. To all of the researchers at the Centre for Endocrinology for helping to nurture an inquisitive and encouraging research environment.

I would like to thank all of my collaborators for their invaluable expert input, helping to shape my project and planning exciting future work. To the patients and their families, without whom this work would not have been possible. This work was supported by the ESPE 2018 Research Fellowship (sponsored by Novo Nordisk A/S) and a Barts Charity Clinical Research Training Fellowship award (MGU0519). I am extremely grateful for these awards which enabled me to dedicate the time and resources required to complete this exciting project.

Table of Contents

Title 1

Statement of Originality 2

Abstract 2

Collaborators 3

Publications arising from this thesis 5

Acknowledgements 6

Table of Contents 7

List of Figures 14

List of Tables 19

List of abbreviations 21

Chapter 1. Introduction

1.1 The GH/IGF-1 axis

1.1.1 Overview 23

1.1.2 Relationship between GH and IGF-1 24

1.1.3 The Growth Hormone Receptor (GHR)

1.1.3.1 The *Growth Hormone Receptor* (*GHR)* gene 25

1.1.3.2 GHR protein structure 27

1.1.3.3 GHBP 30

1.1.3.4 GH signalling via the GHR 30

1.1.4 Insulin-like growth factor 1 (IGF-1) 34

1.2 Growth Hormone Insensitivity (GHI)

1.2.1 Definition 36

1.2.2 Recognised genetic defects in the *GHR* 36

1.2.2.1 Classic Laron Syndrome 37

1.2.2.2 *GHR* 6Ψ pseudoexon 38

1.2.2.3 *GHR* dominant negative variants 39

1.2.3 Other genetic variants causing GHI 41

1.3 IGF-1 insensitivity

1.3.1 Definition 42

1.3.2 Recognised genetic causes of IGF-1 insensitivity 42

1.4 Overlapping short stature syndromes presenting with GH/IGF-1 insensitivity 43

1.5 Genetic variation affecting gene function 45

1.5.1 Missense variants 46

1.5.2 Nonsense variants 48

1.5.3 Frameshift variants 48

1.5.4 Synonymous variants 48

1.5.5 Abnormal splicing of mRNA

1.5.5.1 The process of RNA splicing 49

1.5.5.2 Naturally occurring alternative splicing 51

1.5.5.3 Abnormal splicing affecting gene function 52

1.5.6 Copy number variation (CNV) 53

1.6 Pathogenicity scores for genetic variants 54

1.7 Our cohort: investigations to date

1.7.1 Cohort 55

1.7.2 Previous genetic investigation in our patient cohort 56

1.8 Thesis rationale 57

Chapter 2. Methods

2.1 Ethical approval 61

2.2 Patient cohort 61

2.2.1 Patients assessed on the short stature gene panel 61

2.2.2 Patients assessed for copy number variation (CNVs) 62

2.3 Genetic analysis

2.3.1 DNA extraction from whole blood 63

2.3.2 Creation of custom short stature gene panel 63

2.3.3 Array Comparative Genomic Hybridisation (aCGH) to identify copy number variants (CNVs)

2.3.3.1 Array details 68

2.3.3.2 Array CGH data analysis 68

2.3.3.3 CNV classification 69

2.4 Bioinformatic analysis

2.4.1 Ingenuity Variant Analysis 69

2.4.2 *In silico* splicing prediction software 71

2.4.3 Assessment of genes within the identified CNV regions 71

2.4.3.1 Pathway enrichment analysis 73

2.4.3.2 *In silico* protein-protein interaction analysis using Candidate gene lists 73

2.4.3.3 Statistical analysis of phenotypic predictors in CNV patients 74

2.5 Verifying identified variants of interest

2.5.1 Primer design 74

2.5.2 Agarose gel 75

2.5.3 DNA loading dye 76

2.5.4 Polymerase Chain Reaction (PCR) 76

2.5.5 Sanger sequencing analysis

2.5.5.1 Sending samples for Sanger sequencing 78

2.5.5.2 Interpreting sequencing results 78

2.6 Cell culture 79

2.6.1 Preparation of Luria-Bertani (LB) broth 81

2.6.2 Preparation of LB agar plates 81

2.6.3 Transformation of chemically competent bacterial cells 81

2.6.4 Glycerol stocks 82

2.6.5 Miniprep 82

2.6.6 Midiprep 83

2.6.7 Transient transfection with Lipofectamine 2000 84
2.6.8 Obtaining whole cell lysate 85
2.6.9 BCA (Bicinchoninic acid) Protein assay 86

2.7 Western blotting
2.7.1 Western blot reagents 86
2.7.2 Western blot procedure 87
2.7.3 Modifications to western blot procedure to aid detection of small proteins (<30KDa) 88

2.8 *In vitro* splicing assay 89

2.9 Creation of custom GHR vectors using Gibson assembly 96
2.9.1 Gibson vector and primer design 96
2.9.2 Assembly of mutant *GHR* vectors 100

2.10 Assessing the effect of mutant GHR constructs on downstream pSTAT5 signalling 101

2.11 Site-directed mutagenesis 102
2.11.1 Primer design 102
2.11.2 PCR amplification 103
2.11.3 *Dpn1* treatment 104
2.11.4 Transformation with 'Pure Gold' ultra-competent *E. coli* cells 104

2.12 Electrophoretic Mobility Shift Assay (EMSA)
2.12.1 EMSA reagents and Polyacrylamide gel 105
2.11.2 EMSA protocol 106

2.13 Immunoprecipitation using anti-FLAG® M2 beads 107

2.14 Co-immunoprecipitation (Co-IP) 108

2.15 Restriction enzyme digest 109

2.16 Custom *IGF1* vectors 110

2.17 Immunocytochemistry 114

2.18 Isolating Streptomycin-tagged constructs using StrepTrap® columns 115

2.19 IGF-1 furin cleavage assay 117

2.20 Biochemical assays for patients with c.618+836T>G, c.876-15T>G and c.902T>G *GHR* variants and c.356G>A *IGF1* variant 119

Chapter 3. Variants of interest identified from custom short stature gene panel

3.1 Variants identified following bioinformatic analysis of the short stature gene panel 121

3.2 Novel splicing mutations in the *Growth Hormone Receptor* causing a spectrum of Growth Hormone Insensitivity

3.2.1 Identification of novel *GHR* pseudoexon variant

3.2.1.1 Patient phenotypes 132

3.2.1.2 Bioinformatic analysis and splicing prediction 139

3.2.1.3 Segregation of novel 6Ω *GHR* variant 141

3.2.1.4 Functional assessment of novel c.618+836T>G *GHR* variant

3.2.1.4.1 *In vitro* splicing assay 143

3.2.1.4.2 RT-PCR fibroblasts from Kindred 2 146

3.2.1.4.3 Predicted effect of novel 6Ω pseudoexon inclusion on *GHR* sequence and protein 148

3.2.1.5 Functional assessment of *GHR* 6Ω pseudoexon 149

3.2.2 Potential *GHR* pseudoexon variant identified in Patient 6

3.2.2.1 Patient phenotype 152

3.2.2.2 Bioinformatic analysis and splicing prediction 152

3.2.2.3 Segregation 153

3.2.2.4 Splicing assay results 153
3.2.2.5 Analysis of Patient 6 fibroblast cDNA 154
3.2.2.6 BAM file identification of pathogenic variant 156

3.2.3 *GHR* exon 9 dominant negative variants causing abnormal splicing
3.2.3.1 Patient phenotypes 158
3.2.3.2 Bioinformatic analysis and splicing prediction 159
3.2.3.3 Segregation 160
3.2.3.4 Splicing assay results 162
3.2.3.5 Effect of novel variants on *GHR* sequence 163
3.2.3.6 Creation of Patient 7 and 8 mutant *GHR* constructs by Gibson assembly 165
3.2.3.7 Functional assessment of mutant *GHR* constructs 166

3.3 Novel mutation in *IGF1* hypothesised to impair IGF-1 cleavage
3.3.1 Patient phenotype 169
3.3.2 Variant details and functional prediction 171
3.3.3 Segregation 174
3.3.4 Biochemical analysis of family members 175
3.3.5 Immunocytochemistry of FLAG and HA tagged IGF-1 constructs 176
3.3.6 Furin assay with Strep II and HA tagged IGF-1 constructs
3.3.6.1 Isolating streptomycin II tagged IGF-1 constructs for the furin assay 178
3.3.6.2 Furin assay results 185

3.4 Discussion 187

Chapter 4. Copy number variants in patients with Growth Hormone Insensitivity (GHI) and unexplained short stature

4.1 Introduction 195

4.2 Results 196
4.2.1 Clinical and biochemical features of the CNV subjects 197
4.2.2 SRS features in the CNV subjects 201

4.2.3 Details of the CNVs identified in the GHI and unexplained short stature subjects 204
4.2.4 CNVs in the GHI subjects 210
4.2.5 CNVs in the unexplained short stature patients 211

4.3 *In silico* analysis
4.3.1 Pathway enrichment analysis 211
4.3.2 Identification of candidate genes 214
4.3.3 *In silico* protein-protein interaction analysis of candidate gene list 215

4.4 Discussion 218

Chapter 5. Novel candidate genes identified following CNV analysis

5.1 Chromodomain-helicase-DNA-binding protein 1-like (*CHD1L*)
5.1.1 Introduction 226
5.1.2 Details of variants in the patients 228
5.1.3 Functional analysis of patient variants 231
5.1.3.1 Site-directed mutagenesis 232
5.1.3.2 Co-immunoprecipitation 232
5.1.3.3 *In vitro* splicing assay 235
5.1.4 Segregation 236
5.1.5 Conclusions and discussion 236

5.2 High-mobility group AT-hook 2 (*HMGA2*)
5.2.1 Introduction 237
5.2.2 Details of patient variants 239
5.2.3 Electrophoretic mobility shift assay 245
5.2.4 Creation and detection of the frameshift *HMGA2* variants 249
5.2.5 Subcellular localisation of wild-type and mutant HMGA2 proteins 250
5.2.6 Limitation of vector design for long mutant HMGA2 proteins 252
5.2.7 HMGA2 protein expression in patient fibroblasts 253
5.2.8 Creation of transgenic mouse model 254
5.2.9 Conclusions and discussion 255

Chapter 6. Overall conclusions and future directions 257

References 265

Appendix 1: Composition of buffers and reagents utilised from commercially available kits 287

Appendix 2. Primers to confirm variants identified in genomic DNA 290

Appendix 3. Antibodies used and applications 291

Appendix 4. Primer sequences for *In vitro* splicing assay, site-directed mutagenesis, Gibson assembly and cDNA synthesis 293

Appendix 5. Oral presentations of this research 296

Appendix 6. Awards arising from this work 297

List of Figures

Figure 1.1 Overview of the GH-IGF-1 axis in humans 23

Figure 1.2 The coding sequence of the full length *GHR* gene *(GHRfl)* 26

Figure 1.3 Translated full length GHR amino acid sequence 26

Figure 1.4 The coding sequence of the *GHR* variant lacking exon 3 (*GHRd3* isoform) 27

Figure 1.5 Translated amino acid sequence of the *GHRd3* isoform 27

Figure 1.6 Protein structure of the GHR 29

Figure 1.7 Activation of the Growth Hormone Receptor by Growth Hormone 31

Figure 1.8 The intracellular signalling pathway activated by GH binding to the GHR 33

Figure 1.9 *IGF1* gene transcripts 34

Figure 1.10 Key steps in IGF-1 processing post-translation 35

Figure 1.11 Schematic showing the predicted effect of the *GHR* 6Ψ pseudoexon inclusion on protein and ternary protein structure 39

Figure 1.12 Amino acid codons 46

Figure 1.13 The effect of a missense mutation causing Sickle Cell Disease 47

Figure 1.14 Splicing of pre-mRNA to form mRNA 50

Figure 1.15 Consensus sequences of 5' and 3' splice sites 51

Figure 1.16 Findings of CGS and WES sequencing in the GH/IGF-1 Insensitivity cohort prior to my period of research 57

Figure 2.1 Genes included in the custom short stature gene panel 64

Figure 2.2 Filtering pipeline to identify pathogenic variants from short stature gene panel 70

Figure 2.3 Flowchart showing the bioinformatic pipeline used to identify key pathways and candidate growth genes within the class 3-5 CNV regions identified in the subjects 72

Figure 2.4 Chromatogram of Sanger sequencing generated from ABI file on Bioedit 79

Figure 2.5 Alignment of wild-type and mutant sequences on Bioedit identifying a homozygous C to T change 79

Figure 2.6 Vector map for MoBiTec-Exontrap Cloning Vector pET01 (MoBiTec GmbH, Germany) 90

Figure 2.7 Vector map generated on Benchling including location of designed primers to generate mutant *GHR* vector for Patient 7 97

Figure 2.8 Primer sequences for deleting 26bp region in *GHR* to mimic Patient 7 transcript 98

Figure 2.9 Schematic of 6Ω pseudoexon *GHR* vector created by Gibson assembly 99

Figure 2.10 Base and amino acid sequence of wild-type *IGF1* vector insert with FLAG and HA tags 111

Figure 2.11 Base and amino acid sequence of mutant c.356G>A *IGF1* vector insert with FLAG and HA tags 112

Figure 2.12 Base and amino acid sequence of wild-type *IGF1* vector insert with Strep II and His tags and no signal peptide 113

Figure 2.13 Base and amino acid sequence of mutant *IGF1* vector insert with Strep II and His tags and no signal peptide 114

Figure 3.1 Clinical images and growth chart of Patient 5 133

Figure 3.2 Growth chart of Patient 18 137

Figure 3.3 Clinical images and growth chart of Patient 19 139

Figure 3.4 Splicing prediction generated by Human Splicing Finder for c.618+836T>G *GHR* variant 140

Figure 3.5 Location of original 6Ψ *GHR* pseudoexon and the novel 6Ω variant identified on the short stature gene panel 140

Figure 3.6 Pedigrees and sequencing chromatograms for Kindreds 1 and 2 142

Figure 3.7 Splicing assay confirms retention of novel 6Ω *GHR* pseudoexon 143

Figure 3.8 Novel 6Ω *GHR* pseudoexon utilises same dormant acceptor site as the original 6Ψ *GHR* pseudoexon 145

Figure 3.9 Expression of wild-type and mutant transcripts in affected family members with the heterozygous c.618+836T>G *GHR* 6Ω variant 147

Figure 3.10 *GHR* genomic DNA sequence with the 151bp 6Ω pseudoexon inclusion between exons 6 and 7 148

Figure 3.11 GH receptor amino acid sequence with 6Ω pseudoexon inclusion 148

Figure 3.12 Predicted effect of the novel *GHR* 6Ω pseudoexon on protein structure 149

Figure 3.13 The novel *GHR* 6Ω pseudoexon diminishes GH-dependent STAT5B activation and accumulates extracellularly 151

Figure 3.14 Splicing prediction generated by Human Splicing Finder for Patient 6 c.70+32137A>G *GHR* variant 153

Figure 3.15 Splicing assay results for Patient 6 c.70+32137A>G *GHR* variant 154

Figure 3.16 Deletion of 63 bases exon 7 *GHR* identified in Patient 6 cDNA 155

Figure 3.17 c.723C>T *GHR* variant visualised on Patient 6 BAM file 156

Figure 3.18 Section of VCF for Patient 6 missing 42711413C>T variant 157

Figure 3.19 Location of c.876-15T>G *GHR* variant of interest in Patient 7 159

Figure 3.20 Splicing prediction generated by Human Splicing Finder for Patient 8 c.902T>G *GHR* variant 160

Figure 3.21 Location of c.902T>G *GHR* variant of interest in Patient 8 160

Figure 3.22 Segregation of c.810-15C>T *GHR* variant identified in Patient 7 161

Figure 3.23 Segregation of c.836T>G *GHR* variant identified in Patient 8 161

Figure 3.24 Gel electrophoresis of PCR cDNA splicing product 162

Figure 3.25 *GHR* coding base sequence with deletion first 26 bases exon 9 (seen in Patient 7) 163

Figure 3.26 Translated GHR amino acid sequence with Patient 7 mutant exon 9 163

Figure 3.27 *GHR* coding base sequence with deletion first 26 bases exon 9 and T>G change at start of mutant exon 9 (seen in Patient 8) 164

Figure 3.28 Translated GHR amino acid sequence with Patient 8 mutant exon 9 164

Figure 3.29 Deletion of 26 bases exon 9 using Gibson assembly to mimic Patient 7 mutant *GHR* 165

Figure 3.30 Deletion of 26 bases exon 9 and introduction of T>G change using Gibson assembly to mimic Patient 8 mutant *GHR* 166

Figure 3.31 *GHR* c.876-15T>G exerts dominant negative effect on WT GHR signalling 167

Figure 3.32 *GHR* c.902T>G exerts dominant negative effect on WT GHR signalling 168

Figure 3.33 Growth chart for Patient 10 170

Figure 3.34 Arginine at this base position is highly conserved across the species 171

Figure 3.35 Summary of novel *IGF1* variant identified in Patient 10 172

Figure 3.36 Segregation for the novel heterozygous c.356G>A *IGF1* variant 174

Figure 3.37 Expert biochemical analysis for Patient 10 kindred 175

Figure 3.38 Immunocytochemistry for wild-type IGF-1 protein 177

Figure 3.39 Immunocytochemistry for mutant c.356G>A IGF-1 protein 178

Figure 3.40 Western blot showing streptomycin signal from transfected whole cell lysates 179

Figure 3.41 Coomassie stain showing protein signal from StrepTrap purification stages 180

Figure 3.42 Western blot showing streptomycin signal from StrepTrap purification stages 181

Figure 3.43 Coomassie stain showing protein signal from StrepTrap purification stages when lysing in 0.5% triton 182

Figure 3.44 Coomassie stain showing protein signal from StrepTrap purification stages when lysing in RIPA buffer 183

Figure 3.45 Western blot of products from furin assay 184

Figure 3.46 Western blot of wild-type and mutant IGF-1 furin assay products 185

Figure 3.47 Furin digestion of wild-type and mutant IGF-1 proteins 186

Figure 3.48 Furin digestion of wild-type and mutant IGF-1 proteins (3 replicates) 187

Figure 4.1 Summary of the CNVs identified in the GHI and unexplained short stature subjects 197

Figure 4.2 Pathway enrichment analysis identified the WNT pathway as enriched within the gene list from CNV regions of the Growth Hormone Insensitivity (GHI) subjects 213

Figure 4.3 Pathway enrichment analysis identified CLOCK as a transcription regulator enriched within the CNV gene list of the unexplained short stature subjects 214

Figure 4.4 Genes in our patients CNVs associated with Height in GWAS Catalog 215

Figure 4.5 Protein-protein interactions of the candidate genes in the CNV regions of the Growth Hormone Insensitivity subjects 216

Figure 4.6 Protein-protein interactions of the candidate genes in the CNV regions of the unexplained short stature subjects 217

Figure 5.1 The role of *CHD1L* in cell growth and prevention of apoptosis 227

Figure 5.2 Schematic showing the location of the novel missense *CHD1L* variants 231

Figure 5.3 *CHD1L* variants created by site-directed mutagenesis 232

Figure 5.4 Input for co-immunoprecipitation 233

Figure 5.5 Western blot of PARP and CHD1L constructs immunoprecipitated using FLAG beads 234

Figure 5.6 cDNA PCR of spliced *CHD1L* product 235

Figure 5.7 Reported *HMGA2* variants and our patient variants 239

Figure 5.8 Principles of the Electrophoretic mobility shift assay (EMSA) 246

Figure 5.9 Electrophoretic mobility shift assay (EMSA) using FLAG IP product 247

Figure 5.10 Western blot of FLAG IP product used for EMSA in Figure 5.9. HMGA2 antibody used to detect presence of HMGA2 protein 248

Figure 5.11 EMSA using FLAG IP input quantified by western blot for HMGA2 249

Figure 5.12 Western blot of HMGA2 wild-type and mutant proteins 250

Figure 5.13 Immunocytochemistry for HMGA2 wild-type and mutants 251

Figure 5.14 Amino acid sequences for *HMGA2* variants c.144delC and c.145delA: translation differences to *HMGA2* vector 253

Figure 5.15 Western blot analysis of HMGA2 expression in patient fibroblasts 254

List of Tables

Table 1.1 Reported *GHR* dominant negative variants 40

Table 1.2 Netchine-Harbison Clinical Scoring System 45

Table 2.1 Details of genomic regions included in custom short stature gene panel 66

Table 3.1 Phenotypic details of patients with genetic variants identified on short stature gene panel 125

Table 3.2 Genetic variants identified by the custom short stature gene panel 128

Table 3.3 Clinical and auxological details for the patients with the novel c.618+836T>G *GHR* 6Ω pseudoexon mutation 134

Table 3.4 Biochemical details of the patients with the novel c.618+836T>G *GHR* 6Ω pseudoexon mutation and their parents 135

Table 3.5 Phenotypic and biochemical details of Patients 7 and 8 158

Table 3.6 Phenotype and biochemistry of reported heterozygous *IGF1* gene mutations compared to Patient 10 173

Table 4.1 Clinical and biochemical features of the patients harbouring copy number variants 199

Table 4.2 SRS features in the patients harbouring CNVs 202

Table 4.3 Details of copy number variants identified in our patients 205

Table 4.4 Complete list of genes contained in each CNV region identified 208

Table 5.1 Phenotypic details of the 7 patients with variants affecting *CHD1L* 228

Table 5.2 Biochemical details of the 7 patients with variants affecting *CHD1L* 229

Table 5.3 Details of the genetic variants affecting *CHD1L* identified in our cohort 230

Table 5.4 Phenotypic details of patients with *HMGA2* variants 240

Table 5.5 Patient scores for Netchine-Harbison SRS Clinical Scoring System 242

Table 5.6 Biochemistry of patients with novel *HMGA2* variants 243

Table 5.7 Genetic details of novel *HMGA2* variants 244

List of abbreviations

A2ML1	*Alpha-2-macroglobulin-like 1*
µl	Microlitre
ACAN	*Aggrecan*
ADAMTS3	*A disintegrin and metalloproteinase with thrombospondin motifs 3*
ALS	Acid-labile subunit
ANKRA2	*Ankyrin repeat family A member 2*
ARAF	*A-Raf proto-oncogene, serine/threonine kinase*
BP	Base pair
BRAF	*B-Raf proto-oncogene, serine/threonine kinase*
BSA	Bovine serum albumin
CBL	*Casitas B-lineage lymphoma*
CCDC8	*Coiled coil domain–containing 8*
CGS	Candidate gene sequencing
CHD1L	*Chromodomain-helicase-DNA-binding protein 1-like*
CNV	Copy number variant
CRH	Corticotrophin releasing hormone
CUL7	*Cullin 7*
CUL9	*Cullin 9*
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
DNTPs	Deoxyribonucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
FBS	Foetal bovine serum
FBXW8	*F-box/WD repeat-containing protein 8*
GH	Growth hormone
GHBP	Growth hormone binding protein
GHI	Growth Hormone Insensitivity
GHR	*Growth Hormone Receptor*
GHRH	Growth hormone releasing hormone
HEK293T	Human embryonic kidney cell line
HMGA2	*High-mobility group AT-hook 2*
HRAS	*HRas proto-oncogene, GTPase*
ICR1	*Imprinting control region 1*
IGF1	*Insulin-like growth factor 1*
IGF2	*Insulin-like growth factor 2*
IGF1R	*Insulin-like growth factor 1 receptor*
IGF2R	*Insulin-like growth factor 2 receptor*
IGFALS	*Insulin-like growth factor binding protein acid labile subunit*
IGFBP	Insulin-like growth factor binding protein
IGFBP1	*Insulin-like growth factor binding protein 1*
IGFBP2	*Insulin-like growth factor binding protein 2*
IGFBP3	*Insulin-like growth factor binding protein 3*
IGFBP4	*Insulin-like growth factor binding protein 4*
IGFBP5	*Insulin-like growth factor binding protein 5*
IGFBP6	*Insulin-like growth factor binding protein 6*
IKBKB	*Inhibitor of nuclear factor kappa B kinase subunit beta*

IL11RA	*Interleukin 11 receptor subunit alpha*
IL2RG	*Interleukin 2 receptor subunit gamma*
JAK2	*Janus kinase 2*
kDa	Kilodaltons
KRAS	*Kirsten rat sarcoma 2 viral oncogene homolog*
LB Broth	Luria-bertani broth
LZTR1	*Leucine-zipper-like transcriptional regulator 1*
MAP2K1	*Mitogen-activated protein kinase 1*
MAP2K2	*Mitogen-activated protein kinase 2*
MEG3	*Maternally expressed 3*
mRNA	Messenger ribonucleic acid
NF1	*Neurofibromin 1*
Ng	Nanograms
NOX4	*NADPH oxidase 4*
NPR2	*Natriuretic peptide receptor B*
NRAS	*Neuroblastoma RAS viral oncogene homolog*
Nur77	Nuclear receptor 4A1
OBSL1	*Obscurin-like 1*
OMIM	Online Mendelian inheritance in man
PAGE	Polyacrylamide gel electrophoresis
PAPPA	*Pregnancy-associated plasma protein A*
PAPPA-2	*Pregnancy-associated plasma protein A2*
PARP1	*Poly (ADP-ribose) polymerase 1*
PCR	Polymerase chain reaction
PLAG1	*PLAG1 zinc finger*
PTPN11	*Tyrosine-protein phosphatase non-receptor type 11*
RAF1	*Raf-1 proto-oncogene, serine/threonine kinase*
RASA2	*RAS P21 protein activator 2*
RIT1	*RIC-like protein without CAAX motif 1*
ROC1	*Regulator of cullins-1*
RRAS	*Related RAS viral oncogene homolog*
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Standard deviation score
SEC16A	*SEC16 homolog A, endoplasmic reticulum export factor*
SHOC2	*SHOC2 leucine rich repeat scaffold protein*
SHOX	*Short stature homeobox*
SKP1	*S-phase kinase-associated protein 1*
SOCS2	*Suppressor of cytokine signaling 2*
SOS1	*Son of sevenless homolog 1*
SOS2	*Son of sevenless homolog 2*
SS	Short stature
STAT3	*Signal transducer and activator of transcription 3*
STAT5B	*Signal transducer and activator of transcription 5B*
STC1	*Stanniocalcin-1*
STC2	*Stanniocalcin-2*
TCTP	Translationally controlled tumour protein
TENM4	*Teneurin transmembraneprotein 4*
WES	Whole exome sequencing

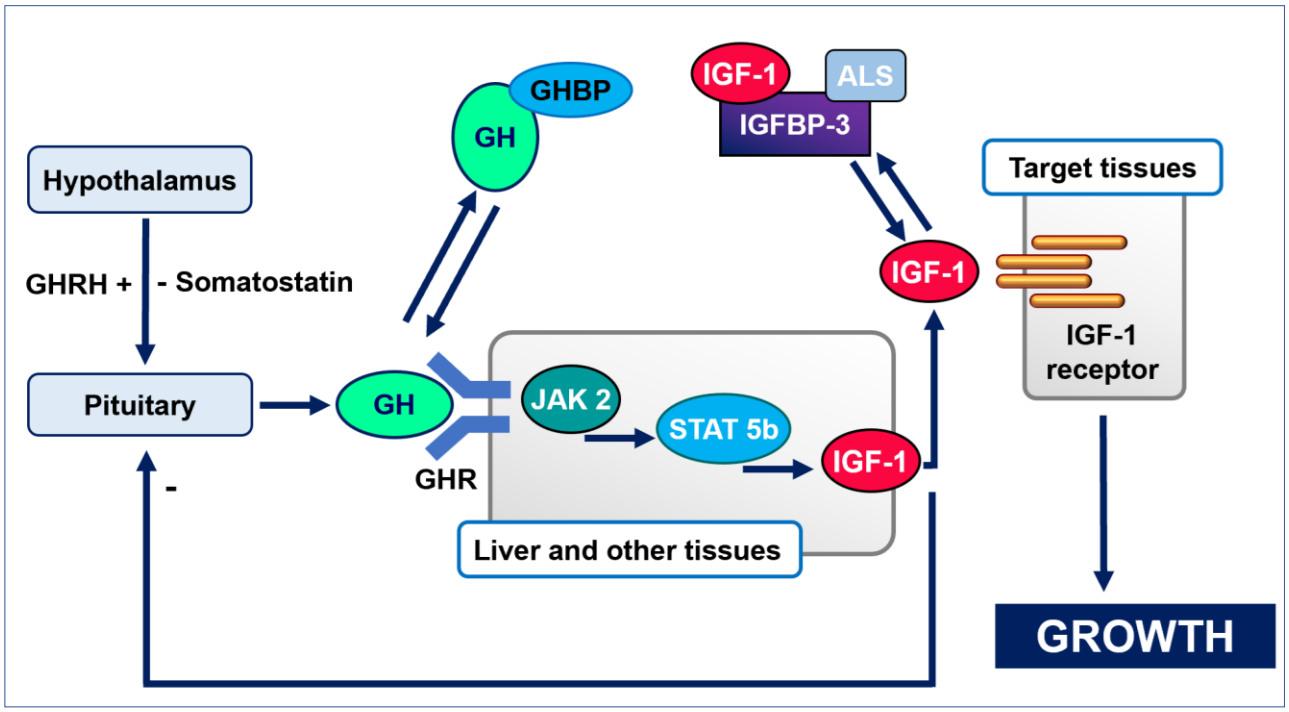
Chapter 1. Introduction

1.1 The GH-IGF-1 axis

1.1.1 Overview

The growth hormone-insulin-like growth factor-1 (GH-IGF-1) axis is essential for normal postnatal human growth (**Figure 1.1**).

Figure 1.1 Overview of the GH-IGF-1 axis in humans



GHRH, Growth Hormone Releasing Hormone; GH, Growth Hormone; GHBP, Growth Hormone Binding Protein; JAK2, Janus kinase 2; STAT5B, Signal Transducer and Activator of Transcription 5B; IGF-1, Insulin-like growth factor 1; IGFBP3, IGF-binding protein 3; ALS, Acid Labile Subunit. GHRH released from the hypothalamus stimulates the release of GH from the anterior pituitary gland. GH circulates bound to GHBP then binds to the GHR. This binding triggers an intracellular signalling cascade which results in the release of IGF-1. IGF-1 circulates bound in a ternary complex with ALS and one of the IGFBPs (most commonly IGFBP3). Being bound in this ternary complex increases the half-life of IGF-1 and its stability in the circulation. Free (unbound) IGF-1 binds to the IGF1R and ultimately results in normal linear growth. Somatostatin has an inhibitory effect on GH release. However, this diagram represents a simplified version of the GH-IGF-1 axis and in fact GH and IGF-1 have multiple actions on human physiology, both direct and indirect. Original figure.

Growth hormone (GH) is secreted in a pulsatile fashion from the pituitary somatotroph cells under the positive and negative influence of hypothalamic hormones GH-releasing hormone (GHRH) and

somatostatin, respectively. The presence of ghrelin stimulates GHRH and thus GH release. Adipocytes inhibit GH secretion, both via leptin-mediated somatostatin secretion and by the direct effect of free fatty acids (FFAs) on the pituitary somatotrophs. GH acts on numerous cell types, tissues and organs, but its main targets for growth are the liver and the epiphyseal plates in the long bones and spine (1). Growth Hormone Binding Protein (GHBP) binds around 50% of the circulating GH and prolongs the half-life of GH, therefore serving as its buffer; however, its physiological role is not completely understood (2). Binding of GH to the Growth Hormone Receptor (GHR) leads to activation of a signalling cascade involving Janus kinase 2 (JAK2) and signal transducer and activator of transcription 5B (STAT5B). STAT5B translocates to the nucleus and leads to transcription of key genes including those encoding *insulin-like growth factor 1* (*IGF1*), *IGF-binding proteins* (*IGFBPs*, of which *IGFBP3* is the most abundant) and *acid labile subunit* (*ALS*), which circulate as a ternary complex. IGF1 is cleaved from this ternary complex by proteases, enabling it to bind the IGF-1 receptor and initiate signal cascades that ultimately promote linear growth. Circulating IGF-1 serves as a negative feedback for GH secretion from the pituitary gland (1).

1.1.2 Relationship between GH and IGF-1

The original "Somatomedin hypothesis" formulated in 1972, proposed that growth was due to the direct effect of GH on hepatic production of IGF-1 (then known as Somatomedin C), with the resulting circulating increase in IGF-1 exerting growth promoting effects on target tissues such as the cartilage and bone (3,4). Whilst the liver is the primary producer of IGF-1, it can also be secreted by non-hepatic tissues such as chondrocytes and exert autocrine/paracrine growth promoting effects (5). Recent genomic and transcriptomic approaches have demonstrated that IGF-1 mRNA is expressed in many human tissues (https://www.ncbi.nlm.nih.gov/gene/3479) (6,7). It is also unclear whether IGF-1 production is mediated by GH in other tissues, however administration of GH injections into hypophysectomised rats resulted in direct GH regulation of IGF-1 in multiple non-

hepatic tissues, including the growth plate and adipose tissues (8–13). Murine models suggest important roles for both peripheral and hepatic IGF-1. Liver-specific *igf1* knockout mice demonstrate normal growth, supporting the role of peripheral IGF-1 (14). However, restoration of only hepatic Igf1 expression in *igf1*$^{-/-}$ mice was sufficient to induce normal growth. These studies demonstrated that in the absence of tissue IGF-1, elevated serum levels of IGF-1 can support normal body size at puberty and beyond, although not normal female reproductive function. This suggests that the majority of autocrine/paracrine actions of IGF-1 that determine organ growth and function can be compensated by elevated levels of circulating IGF-1 (15).

GH and IGF-1 also appear to have additional independent functional roles. GH has been shown to have complex IGF-1-independent anabolic effects. It is recognised that IGF-1 can be regulated independent of GH, for example during *in utero* growth (16). These findings suggest that the growth promoting effects of GH and IGF-1 are not as straightforward as the original 'Somatomedin hypothesis' and our understanding of the complex and diverse roles of GH and IGF-1 is expanding. Several revisions to the original hypothesis have been proposed (17,18). However, it remains clear that both GH and IGF-1 play crucial roles in normal human linear growth.

1.1.3 The Growth Hormone Receptor (GHR)

1.1.3.1 The *Growth Hormone Receptor* (*GHR)* gene

The Growth Hormone Receptor (GHR) is encoded by the *GHR* gene which is located on chromosome 5 and consists of ten exons, nine of which are protein-coding. Multiple *GHR* transcripts are generated from the *GHR* gene, with human full-length *GHR* mRNA widely distributed in human tissues. The two main forms are full-length *GHR* (*GHRfl*) and the common *GHR* variant that lacks

exon 3 (*GHRd3*). The *GHRd3* isoform is the result of a genomic level deletion that causes an in-frame alteration to the extracellular domain resulting in an isoform that may be more sensitive to GH than wild-type GHR (19). The coding sequence and translated amino acid sequence for full length *GHR* are shown below in **Figure 1.2** and **Figure 1.3,** respectively. The coding sequence and translated amino acid sequence for *GHRd3* are shown below in **Figure 1.4** and **Figure 1.5,** respectively.

Figure 1.2 The coding sequence of the full length *GHR* gene *(GHRfl)*

ATGGATCTCTGGCAGCTGCTGTTGACCTTGGCACTGGCAGGATCAAGTGATGCTTTTTCTGGAAGTGAGGCCACAGCAGCTATCCTTAGCAGAGCACCCTGGA
GTCTGCAAAGTGTTAATCCAGGCCTAAAGACAAATTCTTCTAAGGAGCCTAAATTCACCAAGTGCCGTTCACCTGAGCGAGAGACTTTTTCATGCCACTGGACA
GATGAGGTTCATCATGGTACAAAGAACCTAGGACCCATACAGCTGTTCTATACCAGAAGGAACACTCAAGAATGGACTCAAGAATGGAAAGAATGCCCTGATTA
TGTTTCTGCTGGGGAAAACAGCTGTTACTTTAATTCATCGTTTACCTCCATCTGGATACCTTATTGTATCAAGCTAACTAGCAATGGTGGTACAGTGGATGAAAAG
TGTTTCTCTGTTGATGAAATAGTGCAACCAGATCCACCCATTGCCCTCAACTGGACTTTACTGAACGTCAGTTTAACTGGGATTCATGCAGATATCCAAGTGAGAT
GGGAAGCACCACGCAATGCAGATATTCAGAAAGGATGGATGGTTCTGGAGTATGAACTTCAATACAAAGAAGTAAATGAAACTAAATGGAAAATGATGGACCC
TATATTGACAACATCAGTTCCAGTGTACTCATTGAAAGTGGATAAGGAATATGAAGTGCGTGTGAGATCCAAACAACGAAACTCTGGAAATTATGGCGAGTTCA
GTGAGGTGCTCTATGTAACACTTCCTCAGATGAGCCAATTTACATGTGAAGAAGATTTCTACTTTCCATGGCTCTTAATTATTATCTTTGGAATATTTGGGCTAACA
GTGATGCTATTTGTATTCTTATTTTCTAAACAGCAAAGGATTAAAATGCTGATTCTGCCCCCAGTTCCAGTTCCAAAGATTAAAGGAATCGATCCAGATCTCCTCAA
GGAAGGAAAATTAGAGGAGGTGAACACAATCTTAGCCATTCATGATAGCTATAAACCCGAATTCCACAGTGATGACTCTTGGGTTGAATTTATTGAGCTAGATAT
TGATGAGCCAGATGAAAAGACTGAGGAATCAGACACAGACAGACTTCTAAGCAGTGACCATGAGAAATCACATAGTAACCTAGGGGTGAAGGATGGCGACTC
TGGACGTACCAGCTGTTGTGAACCTGACATTCTGGAGACTGATTTCAATGCCAATGACATACATGAGGGTACCTCAGAGGTTGCTCAGCCACAGAGGTTAAAA
GGGGAAGCAGATCTCTTATGCCTTGACCAGAAGAATCAAAATAACTCACCTTATCATGATGCTTGCCCTGCTACTCAGCAGCCCAGTGTTATCCAAGCAGAGAA
AAACAAACCACAACCACTTCCTACTGAAGGAGCTGAGTCAACTCACCAAGCTGCCCATATTCAGCTAAGCAATCCAAGTTCACTGTCAAACATCGACTTTTATGC
CCAGGTGAGCGACATTACACCAGCAGGTAGTGTGGTCCTTTCCCCGGGCCAAAAGAATAAGGCAGGGATGTCCCAATGTGACATGCACCCGGAAATGGTCTC
ACTCTGCCAAGAAAACTTCCTTATGGACAATGCCTACTTCTGTGAGGCAGATGCCAAAAAGTGCATCCCTGTGGCTCCTCACATCAAGGTTGAATCACACATACA
GCCAAGCTTAAACCAAGAGGACATTTACATCACCACAGAAAGCCTTACCACTGCTGCTGGGAGGCCTGGGACAGGAGAACATGTTCCAGGTTCTGAGATGCC
TGTCCCAGACTATACCTCCATTCATATAGTACAGTCCCCACAGGGCCTCATACTCAATGCGACTGCCTTGCCCTTGCCTGACAAAGAGTTTCTCTCATCATGTGGCT
ATGTGAGCACAGACCAACTGAACAAAATCATGCCTTAG

Colouring indicates coding regions of the *GHR*: Exon 2, Exon 3, Exon 4, Exon 5, Exon 6, Exon 7, Exon 8, Exon 9, Exon 10

This base sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html). Each exon of the *GHR* is indicated by a different colour, as explained in the key at the bottom of the image.

Figure 1.3 Translated full length GHR amino acid sequence

MDLWQLLLTLALAGSSDAFSGSEATAAILSRAPWSLQSVNPGLKTNSSKEPKFTKCRSPERETFSCHWTDEVHHGTKNLGPIQLFYTRRNTQEWTQEWKECPDYVS
AGENSCYFNSSFTSIWIPYCIKLTSNGGTVDEKCFSVDEIVQPDPPIALNWTLLNVSLTGIHADIQVRWEAPRNADIQKGWMVLEYELQYKEVNETKWKMMDPILTT
SVPVYSLKVDKEYEVRVRSKQRNSGNYGEFSEVLYVTLPQMSQFTCEEDFYFPWLLIIIFGIFGLTVMLFVFLFSKQQRIKMLILPPVPVPKIKGIDPDLLKEGKLEEVNTI
LAIHDSYKPEFHSDDSWVEFIELDIDEPDEKTEESDTDRLLSSDHEKSHSNLGVKDGDSGRTSCCEPDILETDFNANDIHEGTSEVAQPQRLKGEADLLCLDQKNQNN
SPYHDACPATQQPSVIQAEKNKPQPLPTEGAESTHQAAHIQLSNPSSLSNIDFYAQVSDITPAGSVVLSPGQKNKAGMSQCDMHPEMVSLCQENFLMDNAYFCEA
DAKKCIPVAPHIKVESHIQPSLNQEDIYITTESLTTAAGRPGTGEHVPGSEMPVPDYTSIHIVQSPQGLILNATALPLPDKEFLSSCGYVSTDQLNKIMP*

638 amino acids followed by a stop codon (*)

This sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html) and translated into amino acids using Bioedit Sequence Alignment Editor (20).

Figure 1.4 The coding sequence of the *GHR* variant lacking exon 3 (*GHRd3* isoform)

ATGGATCTCTGGCAGCTGCTGTTGACCTTGGCACTGGCAGGATCAAGTGATGCTTTTTCTGGAAGTGAGGATTCTTCTAAGGAGCCTAAATTCACCAAGTGCCG
TTCACCTGAGCGAGAGACTTTTTCATGCCACTGGACAGATGAGGTTCATCATGGTACAAAGAACCTAGGACCCATACAGCTGTTCTATACCAGAAGGAACACTC
AAGAATGGACTCAAGAATGGAAAGAATGCCCTGATTATGTTTCTGCTGGGGAAAACAGCTGTTACTTTAATTCATCGTTTACCTCCATCTGGATACCTTATTGTAT
CAAGCTAACTAGCAATGGTGGTACAGTGGATGAAAAGTGTTTCTCTGTTGATGAAATAGTGCAACCAGATCCACCCATTGCCCTCAACTGGACTTTACTGAACG
TCAGTTTAACTGGGATTCATGCAGATATCCAAGTGAGATGGGAAGCACCACGCAATGCAGATATTCAGAAAGGATGGATGGTTCTGGAGTATGAACTTCAATAC
AAAGAAGTAAATGAAACTAAATGGAAAATGATGGACCCTATATTGACAACATCAGTTCCAGTGTACTCATTGAAAGTGGATAAGGAATATGAAGTGCGTGTGAG
ATCCAAACAACGAAACTCTGGAAATTATGGCGAGTTCAGTGAGGTGCTCTATGTAACACTTCCTCAGATGAGCCAATTTACATGTGAAGAAGATTTCTACTTTCC
ATGGCTCTTAATTATTATCTTTGGAATATTTGGGCTAACAGTGATGCTATTTGTATTCTTATTTTCTAAACAGCAAAGGATTAAAATGCTGATTCTGCCCCCAGTTCC
AGTTCCAAAGATTAAAGGAATCGATCCAGATCTCCTCAAGGAAGGAAAATTAGAGGAGGTGAACACAATCTTAGCCATTCATGATAGCTATAAACCCGAATTCC
ACAGTGATGACTCTTGGGTTGAATTTATTGAGCTAGATATTGATGAGCCAGATGAAAAGACTGAGGAATCAGACACAGACAGACTTCTAAGCAGTGACCATGA
GAAATCACATAGTAACCTAGGGGTGAAGGATGGCGACTCTGGACGTACCAGCTGTTGTGAACCTGACATTCTGGAGACTGATTTCAATGCCAATGACATACATG
AGGGTACCTCAGAGGTTGCTCAGCCACAGAGGTTAAAAGGGGAAGCAGATCTCTTATGCCTTGACCAGAAGAATCAAAATAACTCACCTTATCATGATGCTTGC
CCTGCTACTCAGCAGCCCAGTGTTATCCAAGCAGAGAAAAACAAACCACAACCACTTCCTACTGAAGGAGCTGAGTCAACTCACCAAGCTGCCCATATTCAGCT
AAGCAATCCAAGTTCACTGTCAAACATCGACTTTTATGCCCAGGTGAGCGACATTACACCAGCAGGTAGTGTGGTCCTTTCCCCGGGCCAAAAGAATAAGGCA
GGGATGTCCCAATGTGACATGCACCCGGAAATGGTCTCACTCTGCCAAGAAAACTTCCTTATGGACAATGCCTACTTCTGTGAGGCAGATGCCAAAAAGTGCAT
CCCTGTGGCTCCTCACATCAAGGTTGAATCACACATACAGCCAAGCTTAAACCAAGAGGACATTTACATCACCACAGAAAGCCTTACCACTGCTGCTGGGAGGC
CTGGGACAGGAGAACATGTTCCAGGTTCTGAGATGCCTGTCCCAGACTATACCTCCATTCATATAGTACAGTCCCCACAGGGCCTCATACTCAATGCGACTGCCT
TGCCCTTGCCTGACAAAGAGTTTCTCTCATCATGTGGCTATGTGAGCACAGACCAACTGAACAAAATCATGCCTTAG

Colouring indicates coding regions of the *GHR*: Exon 2, Exon 4, Exon 5, Exon 6, Exon 7, Exon 8, Exon 9, Exon 10

This base sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html)

Figure 1.5 Translated amino acid sequence of the *GHRd3* isoform

MDLWQLLLTLALAGSSDAFSGSEDSSKEPKFTKCRSPERETFSCHWTDEVHHGTKNLGPIQLFYTRRNTQEWTQEWKECPDYVSAGENSCYFNSS
FTSIWIPYCIKLTSNGGTVDEKCFSVDEIVQPDPPIALNWTLLNVSLTGIHADIQVRWEAPRNADIQKGWMVLEYELQYKEVNETKWKMMDPILTT
SVPVYSLKVDKEYEVRVRSKQRNSGNYGEFSEVLYVTLPQMSQFTCEEDFYFPWLLIIIFGIFGLTVMLFVFLFSKQQRIKMLILPPVPVPKIKGIDPDL
LKEGKLEEVNTILAIHDSYKPEFHSDDSWVEFIELDIDEPDEKTEESDTDRLLSSDHEKSHSNLGVKDGDSGRTSCCEPDILETDFNANDIHEGTSEVA
QPQRLKGEADLLCLDQKNQNNSPYHDACPATQQPSVIQAEKNKPQPLPTEGAESTHQAAHIQLSNPSSLSNIDFYAQVSDITPAGSVVLSPGQKN
KAGMSQCDMHPEMVSLCQENFLMDNAYFCEADAKKCIPVAPHIKVESHIQPSLNQEDIYITTESLTTAAGRPGTGEHVPGSEMPVPDYTSIHIVQ
SPQGLILNATALPLPDKEFLSSCGYVSTDQLNKIMP*

616 amino acids followed by a stop codon (*)

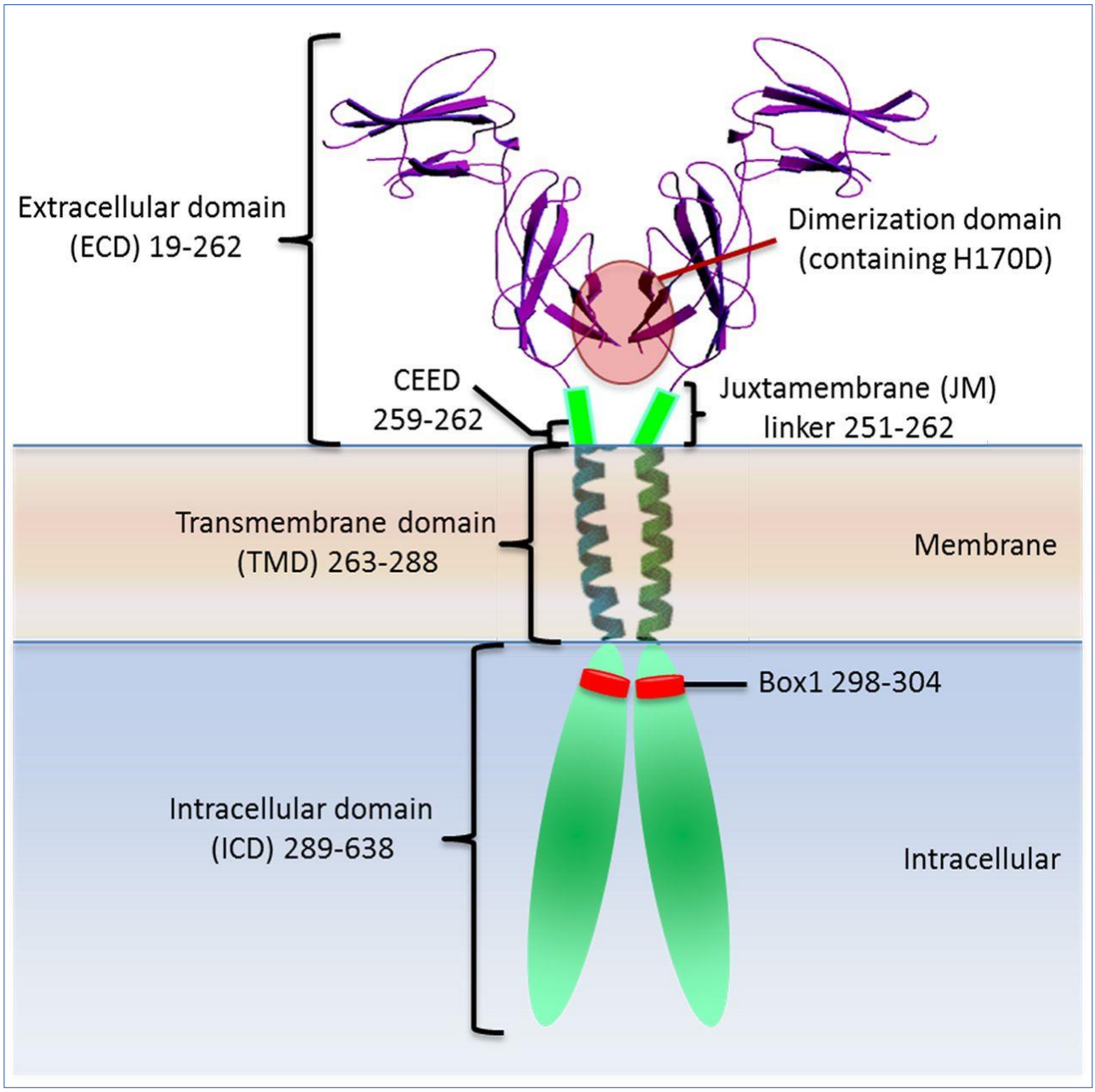
GHR, Growth Hormone Receptor. Exon 3 consists of 66 bases which translates into 22 amino acids, meaning this isoform is 22 amino acids shorter than the full length GHR protein. This sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html) and translated into amino acids using Bioedit Sequence Alignment Editor (20).

1.1.3.2 GHR protein structure

The GHR is a homodimeric, cell-surface, transmembrane protein, and is a member of the class I cytokine receptor family, which includes more than 30 receptors such as the prolactin receptor (PRLR) and the erythropoietin receptor (EPOR) (21). The protein structure of the GHR is shown in **Figure 1.6**. Exons 2 to 10 encode for a prepeptide of 638 amino acid residues. The first 18 amino

acids comprise the signal peptide, which is proteolytically removed upon the insertion of the receptor into the plasma membrane. The mature GHR protein, 620 amino acids in length, is comprised of three domains: an extracellular domain encoded by exons 2 to 7 (246 amino acids), a transmembrane domain encoded by exon 8 (24 amino acids) and the intracellular domain encoded by exons 8, 9 and 10 (350 amino acids) (21).

Figure 1.6 Protein structure of the GHR



GHR, Growth Hormone Receptor. Overview of GHR structure showing the key regions and amino acid regions (22). The signal peptide (amino acids 1-18) is proteolytically removed upon the insertion of the receptor into the plasma membrane so is not included in this visual (21). Amino acids 19-262 of the GHR make the extracellular domain, 263-288 the transmembrane domain and 289-638 the intracellular domain.

The GHR has two main recognised functions – the creation of Growth Hormone Binding Protein

(GHBP) and GH signalling.

1.1.3.3 GHBP

Growth Hormone Binding Protein (GHBP) is cleaved from the extracellular domain of the GHR in a process known as 'receptor ecto-domain shedding' (23). GHBP was first identified in human serum in 1986, but regulation of this shedding is not fully understood (24,25). In humans, GHBP can be cleaved by the metalloproteinase TNF-converting enzyme (TACE) and circulates in the bloodstream (2). Physiological factors including platelet-derived-growth factor and serum also induce GHBP shedding (26). The concentration of circulating GHBP is thought to reflect GHR expression status. GHBP binds ~50% of the circulating GH and prolongs the half-life of GH, therefore serving as a 'buffer'; however, its physiological role is incompletely understood (2). Recently identified dominant negative mutations affecting exon 9 of the *GHR* have been shown to increase GHBP levels and result in GH sequestration and a reduction in GHR signalling. Preventing this GHBP sequestration could therefore be a novel therapeutic target for this subset of patients (27).

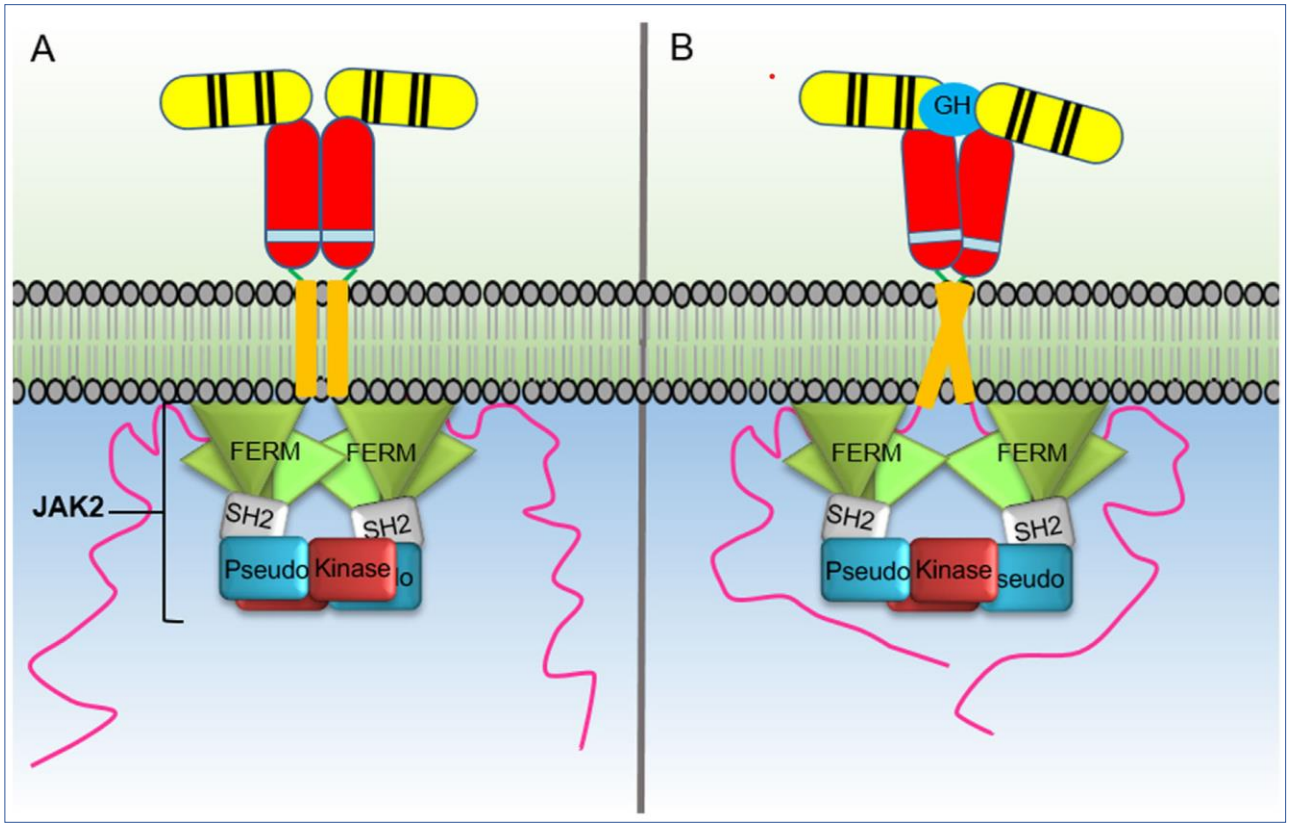
The dimerised GHR has been found to be resistant to proteolysis so GHBP shedding cannot occur (28). However, the presence of GH is required for the interaction between the GHR and TACE to occur (29). The cleavage of GHBP from the GHR may be a mechanism of downregulating GHR function as, without the extracellular domain, the receptor is expected to be non-functional in terms of GH signalling. However, it is unknown if this remnant receptor possesses any regulatory function, as has been reported for other membrane receptors lacking their extracellular domain (30).

1.1.3.4 GH signalling via the GHR

The GHR lacks intrinsic kinase activity and therefore relies on the activation of cytosolic Janus kinase 2 (JAK2) for intracellular signalling. GH binds to a GHR dimer and causes conformational changes that activate JAK2 (**Figure 1.7**). In the inactive state two JAK2 molecules, each bound to a GHR

intracellular domain, interact so that the kinase domain of one JAK2 is inhibited by the pseudokinase domain of the other JAK2 molecule. GH binding to the GHR extracellular domain causes the transmembrane domains to transition from a parallel interaction to a left-handed crossover interaction. This causes a separation of the intracellular domains and the associated JAK2 molecules. The movement of the JAK2 molecules dissociates the pseudokinase–kinase trans-interaction and results in a kinase–kinase trans-interaction and activation (31,32).

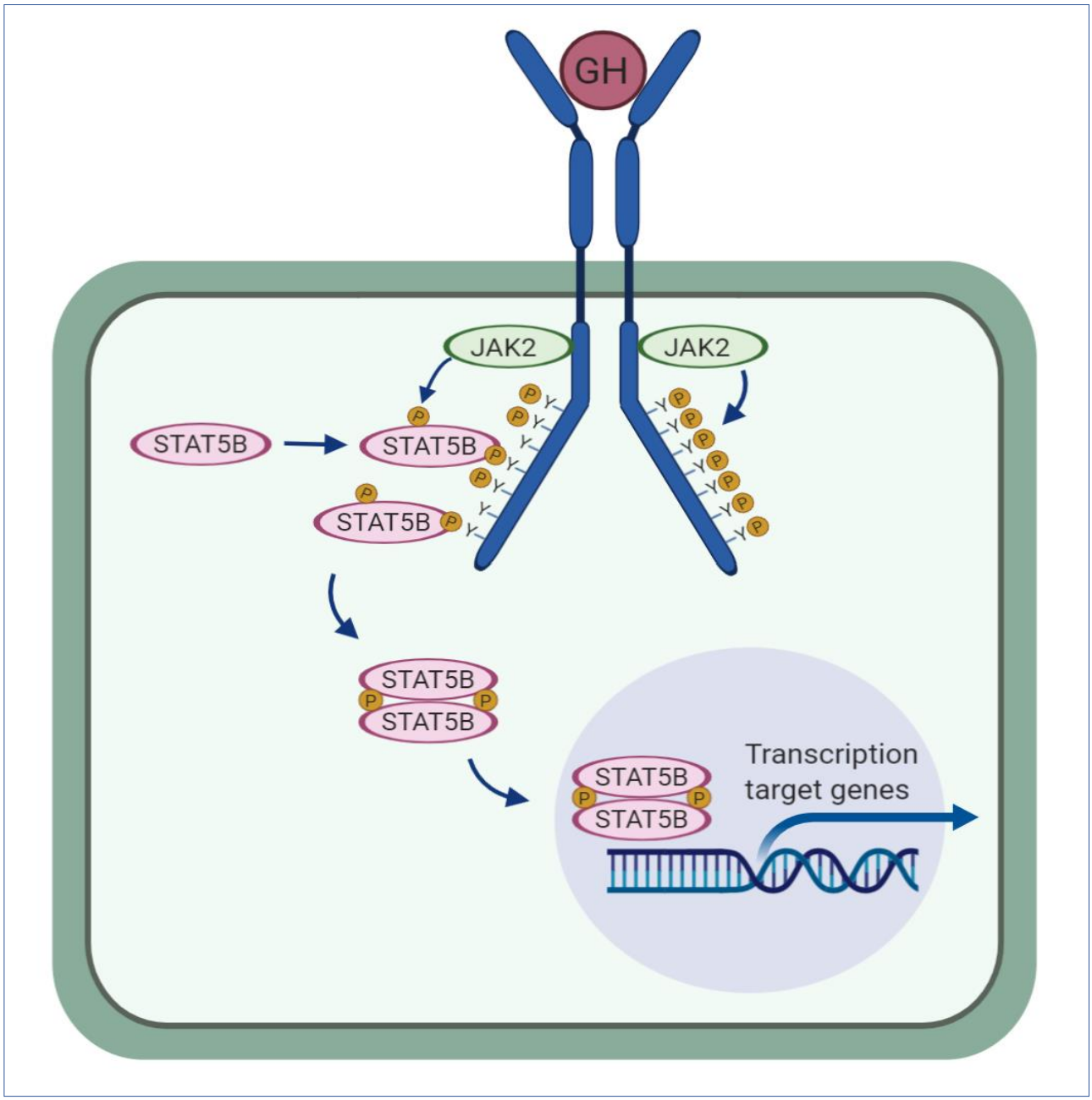
Figure 1.7 Activation of the Growth Hormone Receptor by growth hormone



A) The inactive GHR homodimer. B) The conformational changes that occur upon GH binding (32). GH binding to the GHR extracellular domain causes the transmembrane domains to change from a parallel interaction to a left-handed crossover interaction. This leads to a separation of the intracellular domains and the associated JAK2 molecules. This movement of the JAK2 molecules dissociates the pseudokinase–kinase trans-interaction and results in a kinase–kinase trans-interaction and activation. Once activated, JAK2 phosphorylates tyrosines on the intracellular domain of the GHR. This enables STAT5B to become phosphorylated, which can then translocate to the nucleus.

Once activated, JAK2 phosphorylates tyrosines on the intracellular domain of the GHR. Following recruitment to activated receptors, the latent cytosolic STAT5B protein is phosphorylated on a single tyrosine at position 699 by kinases, including JAK2, which serves to facilitate subsequent homodimerization. Active-STAT5 dimers are then translocated to the nucleus, bind DNA, and act as transcription factors driving transcription of genes including *IGF-1, ALS* and *IGFBPs* (1,21). This pathway is depicted in **Figure 1.8**.

Figure 1.8 The intracellular signalling pathway activated by GH binding to the GHR



Binding of GH to the Growth Hormone Receptor (in blue) is shown. Y represents the intracellular tyrosines of the Growth Hormone Receptor to which phosphates (P) are attached. GH, Growth Hormone; JAK2, Janus kinase 2; STAT5B, Signal Transducer and Activator of Transcription 5B. GH binding to the GHR initiates conformational changes to the GHR and activation of JAK2 as described in **Figure 1.7**. Once activated, JAK2 phosphorylates tyrosines on the intracellular domain of the GHR. The latent cytosolic STAT5B protein is phosphorylated on a single tyrosine at position 699 by kinases, including JAK2, which serves to facilitate subsequent homodimerization. Active-STAT5 dimers are then translocated to the nucleus, bind DNA, and act as transcription factors driving transcription of genes including *IGF-1, ALS* and *IGFBPs* (1,21). Original figure created using Biorender (https://app.biorender.com/).

1.1.4 Insulin-like growth factor 1 (IGF-1)

Insulin-like growth factor 1 (IGF-1) is a cellular and secreted growth factor which is essential for normal growth, development and maintenance and plays key roles in several biological pathways (33–35).

The *IGF1* gene is located on the long arm of chromosome 12 in humans. It is around 80Kb in size and is highly conserved in mammals and primates (36). It contains 6 exons but due to several transcription initiation sites, alternative splicing and different polyadenylation signals, gives rise to multiple *IGF1* transcripts with a variety of biological functions (37,38). The composition of these multiple transcripts is displayed in **Figure 1.9.**

Figure 1.9 *IGF1* gene transcripts

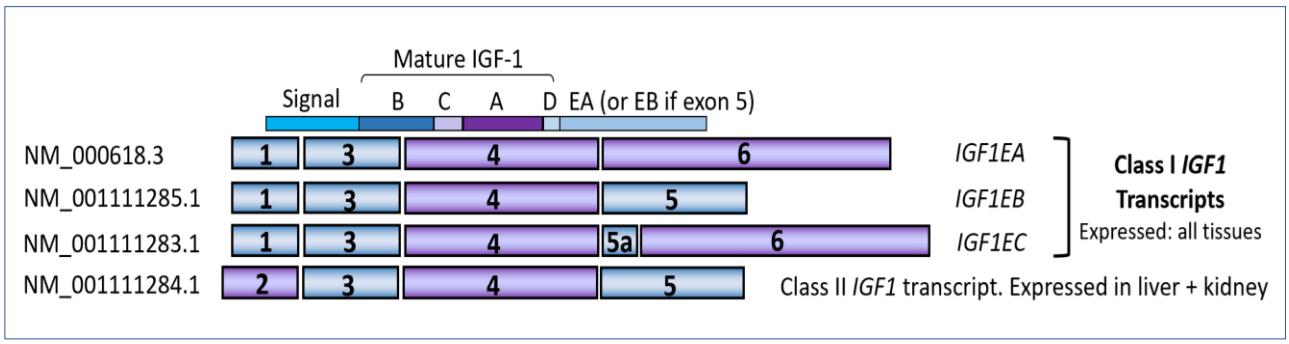


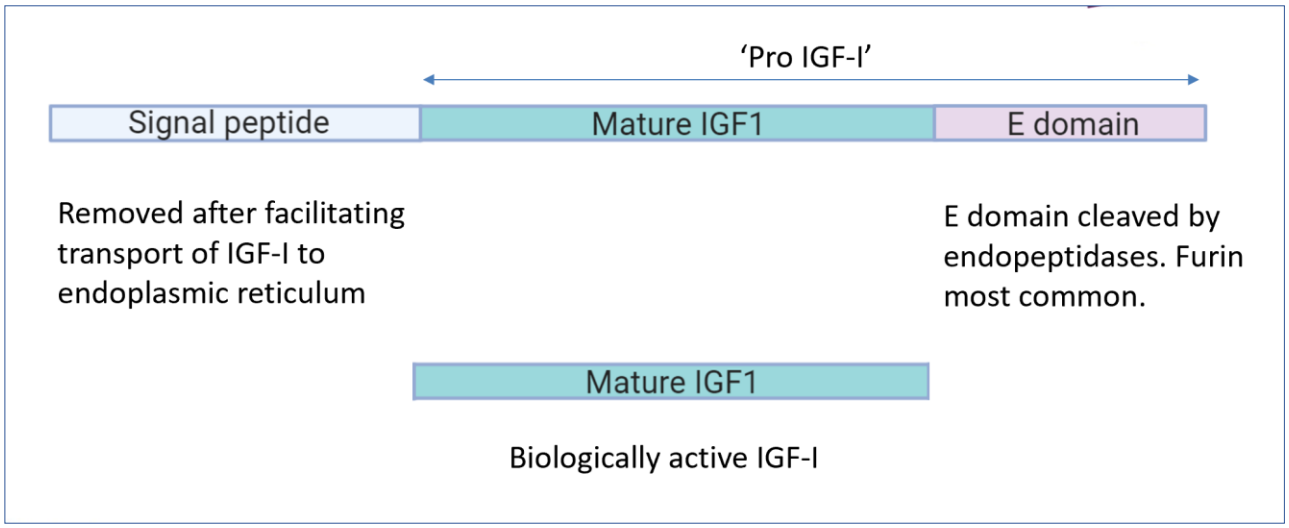
Figure showing the exonic make-up of the multiple *IGF1* transcripts created by differential splicing. Mature *IGF1* is encoded by sections B,C, A and D so is the same in all of the transcripts and composed of exons 3 and 4. The Signal peptide and E domain have different components in each transcript. The 3 class I *IGF1* transcripts all contain exon 1, 3 and 4 but differ in their exon 5 and 6 composition. They are expressed in all tissues. The class II *IGF1* transcript contains exon 2, 3, 4 and 5 and is expressed only in the liver and kidney. Modified version of original figure by Dr Vivian Hwa.

There are 2 different classes of *IGF1* mRNA variants: class I transcripts have their initiation sites on exon 1 (promoter 1), whereas class II transcripts have their initiation sites on exon 2 (promoter 2). Differential splicing creates class one (exon 1 and 3) or class 2 (exon 2 and 3) variants. Differential splicing also results in variants containing different sections of exon 5 and 6. Class I transcripts are

expressed widely in many tissues of the body, whereas class II transcripts are expressed mostly in the liver and kidney and are thought to be GH responsive/dependent (39–42).

The translated IGF-1 protein consists of the signal peptide, the mature IGF-1 sequence and the E domain. After it is translated, IGF-1 undergoes several key steps in the generation of mature IGF-1, the biologically active form of IGF-1 that circulates in the bloodstream. These key stages are shown below in **Figure 1.10**. After facilitating the passage of the polypeptide to the endoplasmic reticulum and the secretory pathway, the signal peptide is thought to be removed leaving the mature IGF-1 and the E domain, also known as 'Pro-IGF1' (36,43). The cleavage of the E domain by endopeptidases such as furin has been reported to occur both intra- and extracellularly (44–46). Removal of the E domain results in the mature, bioactive IGF-1 (47).

Figure 1.10 Key steps in IGF-1 processing post-translation



The signal peptide of IGF-1 is removed after facilitating the transport of IGF-1 to the endoplasmic reticulum and into the secretory pathway. This leaves Pro-IGF-1, which is composed of mature IGF-1 plus the E domain. The E domain is cleaved by endopeptidases, most commonly furin. The cleavage of the E domain creates mature IGF-1, which circulates in the bloodstream and exerts its growth promoting effects. Original diagram.

The biological actions of IGF-1 are modulated by the insulin-like growth factor binding proteins (IGFBPs). These are a family of six homologous proteins with a high binding affinity for IGF-1 and IGF-2 (48). The IGFBPs bind IGF-1 and increase its half-life by protecting it from proteolysis (49).

Hence, they regulate the free IGF-1 concentration and local bioavailability, thought to be specific to the target tissue (48,50). The proportion of free circulating IGF-1 is very low. The majority of circulating IGF-1 forms a ternary complex with IGFBP-3 and the glycoprotein acid-labile subunit (ALS). Given its large size, the IGFBP-3 molecule is unable to pass through the capillary membrane (49). IGFBP3 acts as an inhibitor of proliferation, evoking a pro-apoptotic effect. A number of potential mechanisms have been suggested to explain this, including sequestration of IGF-1 from the receptor and binding competition with IGF1R (51). Some of the IGFBPs also display IGF-1 potentiating effects (52). The interactions between these molecules of the IGF-1 system are complex and are incompletely understood (53).

1.2 Growth Hormone Insensitivity (GHI)

1.2.1 Definition

Perturbations of the GH-IGF-1 axis can lead to Growth Hormone Insensitivity (GHI) (54). GHI is characterised by a triad of short stature (SS), IGF-1 deficiency and normal/high GH levels. The most commonly identified genetic cause of GHI are defects in the *GHR* (55).

1.2.2 Recognised genetic defects in the *GHR*

Since 1966, over 90 different naturally-occurring mutations affecting the *GHR* gene have been reported (56). The majority of these mutations affect the extracellular domain of the GHR, resulting in abnormal GH binding and hence low or undetectable GHBP levels (55). Missense, nonsense, splice mutations and deletions have also been reported. Splice mutations represent approximately 20% of *GHR* defects (54). The majority of mutations causing aberrant *GHR* mRNA splicing disrupt major regulatory elements, such as the canonical donor and acceptor splice sites (57,58). Homozygous and compound heterozygous mutations affecting the exonic and intron-exon boundaries of the *GHR* are

recognized to cause classic Growth Hormone Insensitivity (GHI), or Laron Syndrome. However, non-classical mutations of the *GHR* have also been identified in recent years, including the *GHR* pseudoexon and heterozygous mutations affecting the intracellular region of the *GHR* and exerting a dominant negative effect on receptor function (59–63).

1.2.2.1 Classic Laron Syndrome

Classical GHI (Online Mendelian Inheritance in Man (OMIM) #262500), also known as 'Laron Syndrome,' was first described by Laron *et al.* in 1966 (64). This paper reported three children from a consanguineous Jewish family who showed striking postnatal growth failure despite high serum GH concentrations. The initial hypothesis for this clinical picture was the presence of an abnormal GH molecule, but was subsequently shown to be due to *GHR* gene defects (65). This syndrome is now recognised to be the result of homozygous and compound heterozygous *GHR* mutations (66,67).

Globally, an estimated 350 people are affected by Laron syndrome (68). Given the autosomal recessive nature of the condition, it is seen more frequently in populations with a high incidence of consanguinity (69). The 'Israeli cohort' consists of 69 affected individuals living in Israel (70) and the 'Ecuadorian cohort' is made up of 90 individuals who inhabit the remote villages of the Loja province in southern Ecuador (71–73). Most of the affected members of this Ecuadorian community are homozygous for the *GHR* splice site mutation c.594A>G, p.Val181_Met188del which is thought to have a common founder effect (74).

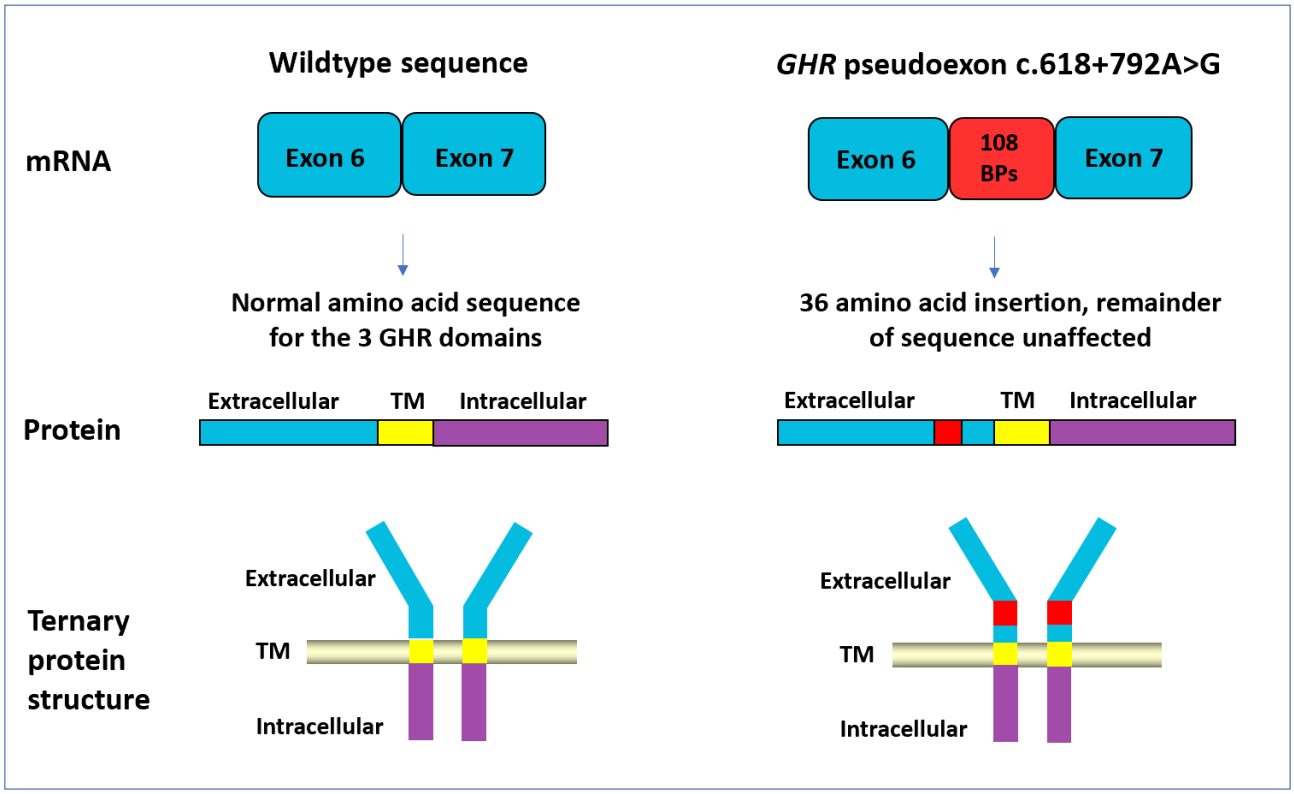
Laron syndrome is associated with classic phenotypic features. Growth failure is the most striking feature, with untreated females attaining an adult height of 108-136 cm (height SDS -9.0 to -4.3)

and males 116-142cm (height SDS -8.8 to -4.9). The body length of newborns ranges between 42-45 cm and their average birthweight is 2500g (69). Patients have small hands and feet and a relative macrocephaly as their head size tends to be relatively normal in striking contrast to their body (75). There is midface hypoplasia and underdevelopment of the facial bones, causing the appearance of a protruding forehead and saddle nose. Sparse, slowly growing hair and nails and crowded, defective teeth are also classic features. Many patients, especially female, have a high-pitched voice due to a narrow oropharynx and larynx. Patients may be overweight and their obesity increases progressively with age (76). Despite this obesity, patients with Laron syndrome have lower incidence of both diabetes and malignancy, which may be attributable to their reduced IGF-1 levels (70–72,77). Interestingly both genders attain full sexual development and reproductive potential, despite their hypogenitalism, hypogonadism and delayed puberty (delayed by 3-7 years on average) (78).

1.2.2.2 *GHR* 6Ψ pseudoexon

Our centre previously described the first *GHR* pseudoexon mutation in 2001 in four siblings with mild GHI from a highly consanguineous Pakistani family (79). This homozygous variant (42700896A>G, c.618+792A>G) alters the intronic sequence from CAGT to CGGT and activates a cryptic donor splice site. Due to the presence of a nearby dormant cryptic acceptor site, this region is recognised as an exon by the spliceosome and this additional exon is retained during *GHR* splicing (a 'pseudoexon'). The inclusion of this pseudoexon is predicted to cause in-frame insertion of 36 amino acid residues between exons 6 and 7 (**Figure 1.11**). This insertion in the dimerization domain of the *GHR* results in defective trafficking rather than impaired signalling, causing a partial loss-of-function. As such, moderate postnatal growth failure is observed (Height SDS -3.3 to -6.0) (80).

Figure 1.11 Schematic showing the predicted effect of the *GHR* 6Ψ pseudoexon inclusion on protein and ternary protein structure



mRNA, messenger RNA; GHR, Growth Hormone Receptor; BPs, base pairs; TM, transmembrane. The c.618+792A>G mutation leads to mis-splicing and the inclusion of a 108-base pair pseudoexon inclusion between exons 6 and 7 of the *GHR*. When translated, this leads to an in-frame insertion of 36 amino acids to the extracellular domain of the Growth Hormone Receptor. Original figure.

This variant is 792 base pairs (bp) into the intron and was identified using homozygosity mapping of several polymorphic markers surrounding the *GHR* (79). It would not be detected by whole exome sequencing as this technique only covers exons and intron-exon boundaries. Candidate gene sequencing also focuses on sequencing exons of the gene of interest and would not identify changes deep within non-coding regions.

1.2.2.3 *GHR* dominant negative variants

Seven *GHR* 'dominant negative' (DN) variants (OMIM #604271) have been reported to date (59–62). The presence of one mutant allele exerts a dominant effect on the function of the GHR, unlike the classic *GHR* homozygous defects. These DN variants affect the transmembrane and/or intracellular

domain of the GHR (**Table 1.1**). Interestingly, the phenotype observed in these patients is not as severe as classic GHI. Height SDS ranges from -3.0 to -4.3 and often no dysmorphic features are seen (59–62). This is likely due to the presence of 25% WT-WT GHR homodimers (62).

Table 1.1 Reported *GHR* dominant negative variants

	Location in *GHR gene*	Variant	Protein change	Effect at protein level	Length of GHR protein (AA)	Ref
1	Intron 8	c.876–1G>C	p.I293Kfs*4	Splicing out of exon 9. Premature stop 3 altered amino acids into exon 10	295	(60)
2	Exon 9	c.899dupC	p.V300SFs*7	Premature stop after 6 altered amino acids	306	(81)
3	Exon 9	c.920_921ins14	p.K307Nfs*18	Premature stop after 17 altered amino acids	323	(62)
4	Intron 9	c.945+1G>A	p.I293Kfs*4	Splicing out of exon 9. Premature stop 3 altered amino acids into exon 10	295	(61)
5	Intron 9	c.945+2C>T	p.I293Kfs*4	Splicing out of exon 9. Premature stop 3 altered amino acids into exon 10	295	(62)
6	Exon 10	c.964dupG	p.V322Gfs*9	Premature stop after 8 altered amino acids	329	(62)
7	Exon 10	c.1011_1012delCA	p.His337Gln*2	Premature stop after 1 altered amino acid	337	(59)

GHR, Growth Hormone Receptor; AA, amino acid; Ref, reference. GHR amino acid lengths were calculated using sequences from Ensembl37 Genome Browser. (http://grch37.ensembl.org/index.html) and translating the modified base sequences to amino acids using Bioedit Sequence Alignment Editor (20).

The interactions of the resulting truncated *GHR* transcripts with the full-length 'normal' *GHR* transcript have previously been investigated. Immunoprecipitation and Western blotting experiments of cells transfected with a mutant *GHR* 279bp in length with a full-length GHR revealed that the mutant and full-length GHR could form heterodimers, and the mutant GHR could only be internalised when complexed with full-length GHR (82). This is likely due to the truncated GHR

proteins lacking the UbE motif at amino acids 340-349, necessary for ubiquitin-dependent endocytosis (83). Several of the *GHR* dominant negative variants have been shown to impair GHR signalling i.e. the phosphorylation of STAT5B is reduced when co-transfected with WT *GHR*, thus acting in a dominant negative manner (81,84). As discussed above, phosphorylation of STAT5B is a key aspect of GHR downstream signalling and the generation of IGF-1.

1.2.3 Other genetic variants causing GHI

Monogenic defects responsible for GHI have been identified in *STAT5B* (85–87), *IGFALS* (88), *PAPPA2* (89) and *IGF1* (90) genes. Many of these mutations are associated with additional clinical and phenotypic features; immunodeficiency (*STAT5B*) (91,92), reduced bone mineral density (*PAPPA2)* (89) and pubertal delay (*STAT5B*, *IGFALS*) (93,94). These additional features facilitate the clinical recognition. However, a significant proportion of patients with clear evidence of GHI remain without a genetic diagnosis despite extensive investigation.

Pathogenic *IGF1* gene mutations causing childhood growth failure are extremely rare. Only 5 autosomal recessive mutations, 1 *IGF1* copy number variant and 2 heterozygous frameshift mutations are reported. Homozygous *IGF1* mutations cause pre and postnatal growth failure, severe learning difficulties, microcephaly and delayed bone age. Heterozygous *IGF1* defects lead to postnatal growth failure and mild learning difficulties, with or without microcephaly and delayed bone age. This observed phenotype is thought to be due to *IGF1* haploinsufficiency (90,95–103). Heterozygous missense *IGF1* mutations have not previously been described.

1.3 IGF-1 insensitivity

1.3.1 Definition

IGF-1 insensitivity (OMIM #270450) causes impaired foetal and postnatal growth and is associated with high or normal GH and IGF-1 levels (104).

1.3.2 Recognised genetic causes of IGF-1 insensitivity

Mutations in *IGF1R* were first reported by Abuzzahab *et al*. following analysis of DNA from children with short stature and unexplained intrauterine growth restriction (104). Haploinsufficiency of the *IGF1 receptor* (*IGF1R*) gene is characterised by IGF-1 insensitivity causing impaired fetal and postnatal growth associated with high/normal IGF-1 levels. Additional features include microcephaly and developmental delay. A complete absence of IGF1R protein in humans may be lethal, such as with a loss-of-function homozygous mutation. This could explain why most identified cases have heterozygous mutations, except for two patients with compound heterozygous mutations and two with homozygous mutations (104–107). The few patients with biallelic mutations appear to retain some degree of IGF1R activity. Functional studies of *IGF1R* patient mutations suggest different mechanisms for the IGF1 resistance and poor growth: receptor haploinsufficiency, decreased biosynthesis, reduction of binding affinity, interference of transmembrane signalling and disruption of the tyrosine kinase activity have all been proposed (108). The impact of *IGF1R* mutations on intrauterine growth is variable, but often more severe when the mutation is of maternal origin. This indicates that maternal IGF-1 resistance during pregnancy is a contributing factor to the severity of the growth retardation, possibly by decreasing placental growth (95). Patients with *IGF1R* mutations have shown poor to moderate clinical responses to recombinant Growth Hormone treatment (108).

1.4 Overlapping short stature syndromes presenting with GH-IGF-1 insensitivity

In recent years, there has been a rapid expansion of the clinical, biochemical, and genetic characteristics of GH/IGF-1 insensitivity and other overlapping short stature syndromes. This is likely due to a combination of increasing awareness of these disorders and the advances in genetic analysis available. Several short stature (SS) syndromes overlap with GHI-IGF-1 axis defects such as Noonan syndrome (OMIM #163950; #605275; #609942; #610733; #611553; #613224), 3M syndrome (OMIM #273750; #612921; #614205) and Silver-Russell Syndrome (SRS, OMIM #180860) (109–111).

Noonan syndrome is an autosomal dominant condition that causes postnatal growth failure and distinctive facial features in addition to a number of additional features including congenital heart defects, mild-to-moderate learning disability and skeletal abnormalities (112,113). Noonan syndrome (NS) is caused by mutations in genes involved in the RAS/MAPK signalling pathway which plays key roles in various cellular processes, including proliferation, survival, differentiation, metabolism and migration (113,114). Around 50% of genetically confirmed cases of NS are identified in *Protein tyrosine phosphatase, nonreceptor type 11* (*PTPN11*), which codes for the cytoplasmic tyrosine phosphatase *SH2 domain–containing phosphatase 2* (SHP-2) (113,115). SHP-2 binds to and dephosphorylates signalling molecules such as STAT5b (116). As discussed previously, STAT5b plays important roles in GH signalling and thus gain-of-function mutations of *PTPN11* may reduce the efficacy of GH signalling and thus cause a degree of GHI. Consistent with this, serum IGF-1 levels were decreased in NS carrying *PTPN11* mutations (117,118). The mechanism of GHI in NS patients without *PTPN11* mutations is less well understood. It is hypothesized that abnormal functioning of these genes hyperactivate the RAS–MAPK pathway, inhibiting the JAK–STAT pathway and impairing the production of IGF-1 (119). Consistent with this, pharmacological inhibition of the RAS–MAPK

pathway improves growth in mouse models carrying mutations of *RAF1* and *SOS1*, both recognised to cause Noonan syndrome (120,121).

3M syndrome is characterised by pre- and post-natal growth failure, prominent heels, triangular-shaped face, normal intelligence, and distinct radiological features (slender long bones and tall vertebral bodies) (122). There is an interesting phenotypic overlap between 3M syndrome and Laron syndrome, with both groups displaying frontal bossing, depressed nasal bridge, bitemporal hair thinning, and a high-pitched voice (122). 3 genetic causes of 3M syndrome are recognised - mutations in *Cullin 7* (*CUL7*) (70% of genetically diagnosed cases), O*bscurin-like 1* (*OBSL1*) (25%), and *Coiled coil domain–containing 8* (*CCDC8*) (5%) (123–129). Many 3M patients show a biochemical profile suggestive of GHI. Patients have normal or high peak GH levels and low or normal IGF-1 levels. The growth response to hGH therapy in these patients is variable but typically poor, also suggesting a degree of GHI (122,129).

Silver-Russell Syndrome (SRS) is a clinical diagnosis based on a combination of characteristic features including being born small for gestational age (SGA), postnatal growth failure, relative macrocephaly at birth/protruding forehead, body asymmetry, feeding difficulty and/or low BMI. Diagnosis requires fulfilment of ≥4/6 Netchine-Harbison Clinical Scoring System (NH-CSS; including both prominent forehead and relative macrocephaly, termed 'Clinical SRS'), or 3/6 NH-CSS in association with a genetic defect recognised to cause SRS (130). These criteria are described in **Table 1.2**.

Table 1.2 Netchine-Harbison Clinical Scoring System

Clinical criteria	Definition
SGA (birth weight and/or birth length)	≤−2 SDS for gestational age
Postnatal growth failure	Height at 24 ± 1 months ≤−2 SDS or height ≤−2 SDS below mid-parental target height
Relative macrocephaly at birth	Head circumference at birth ≥1.5 SDS above birth weight and/or length SDS
Protruding forehead*	Forehead projecting beyond the facial plane on a side view as a toddler (1–3 years)
Body asymmetry	LLD of ≥0.5 cm or arm asymmetry or LLD <0.5 cm with at least t wo other asymmetrical body parts (one non-face)
Feeding difficulties and/or low BMI	BMI ≤−2 SDS at 24 months or current use of a feeding tube or cyproheptadine for appetite stimulation

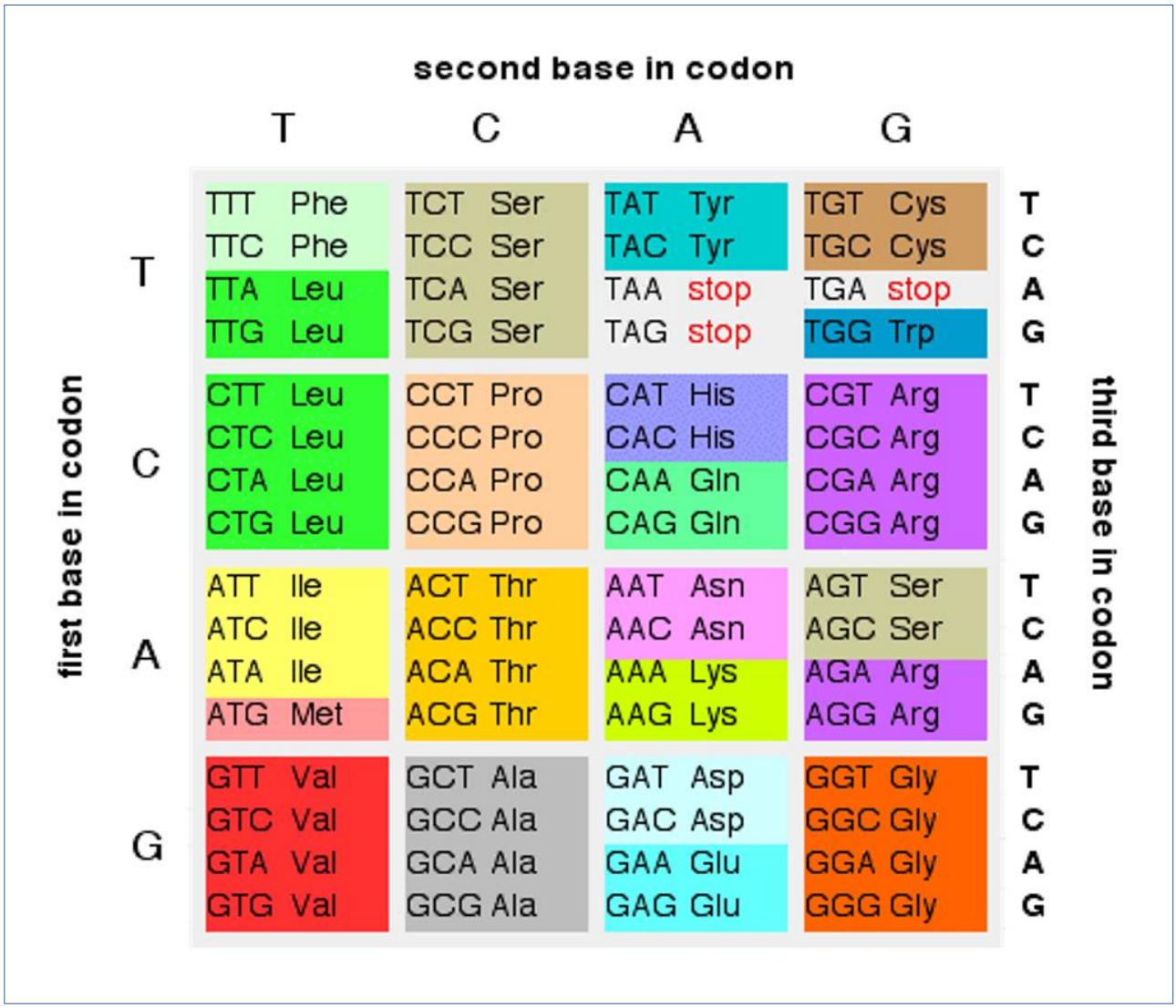
A clinical diagnosis of SRS requires fulfilment of ≥4/6 Netchine-Harbison Clinical Scoring System (NH-CSS; including both prominent forehead and relative macrocephaly, termed 'Clinical SRS'), or 3/6 NH-CSS in association with a genetic defect recognised to cause SRS (130). LLD, leg length discrepancy; SDS, Standard Deviation Score; SGA, small for gestational age; BMI, Body Mass Index (130,131).

Hypomethylation of the imprinted *H19/IGF2* domain of chromosome 11p15 (11p15LOM) and maternal uniparental disomy of chromosome 7 (upd(7)mat) are identified in 50-60% and 10% SRS cases, respectively (130). We previously reported 2 patients with growth failure and GHI who were subsequently diagnosed with 11p15LOM and upd(7)mat (109,132). The genetic aetiology remains unknown in around 40% clinical SRS cases (131). Children with SRS and 11p15LOM have significantly higher IGF-1 and IGFBP-3 levels compared with other children born small for gestational age, suggesting they may have a degree of IGF-1 resistance (133,134).

1.5 Genetic variation affecting gene function

Three nucleotides code for one amino acid as shown in **Figure 1.12**. The effect of changing a single base or inserting or deleting bases on the resulting amino acid and protein structure can vary widely. Some of the common types of changes are discussed below.

Figure 1.12 Amino acid codons



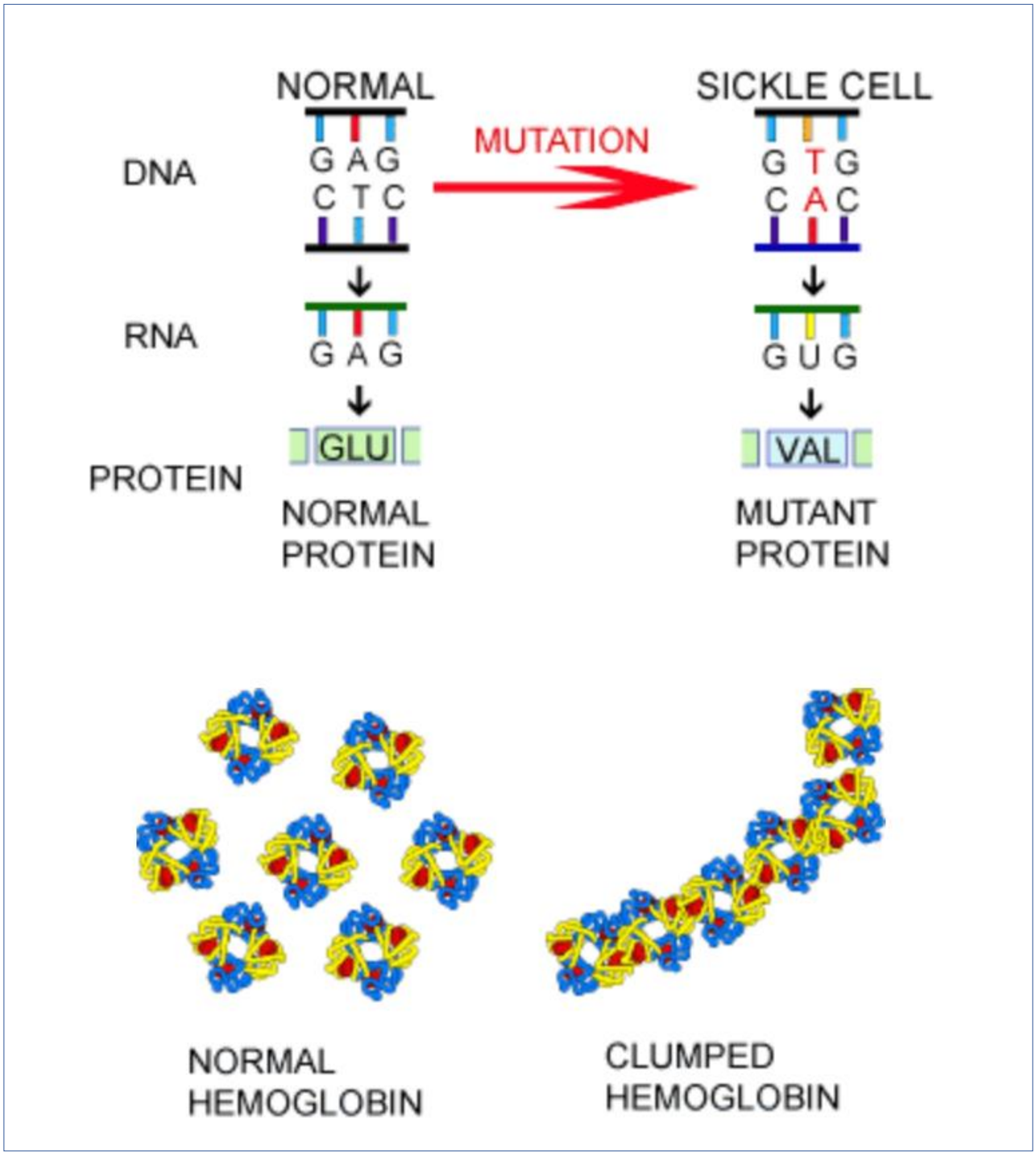
Base sequence of amino acid codons. Image obtained from https://www.chemguide.co.uk/organicprops/aminoacids/dna4.html

1.5.1. Missense variants

A missense mutation is a point mutation in a single nucleotide that leads to one altered amino acid in a protein. It does not affect the rest of the amino acid sequence, so the effect of the missense variant is largely dependent on the role of the amino acid that has been changed and whether the new amino acid has similar properties to the original. Missense mutations can have dramatic affects on protein function. One example of this is the most common cause of Sickle Cell Disease, illustrated below in **Figure 1.13**. A single A→T base change (GAG→GTG) in the *Haemoglobin Subunit Beta*

(*HBB*) gene alters the 6th amino acid from Glutamine to a Valine (Glu6Val) with dramatic effects on the protein function (135).

Figure 1.13 The effect of a missense mutation causing Sickle Cell Disease



This image demonstrates the profound impact the missense Glu6Val mutation has on the formation of haemoglobin in affected individuals. Image obtained from https://evolution.berkeley.edu/evolibrary/article/mutations_06

1.5.2 Nonsense variants

A nonsense mutation is caused by a single base change that results in a premature stop codon. For example, TGC→TGA which would change a Cysteine to a stop codon. These nonsense mutations result in a truncated protein that often lacks the function of the normal, or 'wild-type,' protein. Around 10% of genetic disease is due to nonsense mutations (136).

1.5.3 Frameshift variants

Frameshift variants are caused when any number of nucleotides are inserted or deleted from a sequence that are not divisible by 3. Because 3 bases code for one amino acid, this insertion or deletion alters the reading frame (the grouping of the codons) and thus results in the translation of different amino acids. A frameshift mutation will usually result in the translation of a different length amino acid sequence due to the earlier or later appearance of a stop codon in the sequence. Usually the resulting protein will be very different to the original and often non-functional. This is more likely the earlier in the sequence that the frameshift occurs.

1.5.4 Synonymous variants

Several 3-base combinations code for the same amino acid. For this reason, a single nucleotide variant can have no effect on the resultant amino acid and thus be described as a 'synonymous variant.' However, although these variants can be predicted benign due to their lack of effect on amino acid translation, they can have significant effects on splicing if they create a donor or acceptor splice site. For example, synonymous variants in *CYP11A1* have been shown to cause Primary Adrenal Insufficiency through missplicing (137).

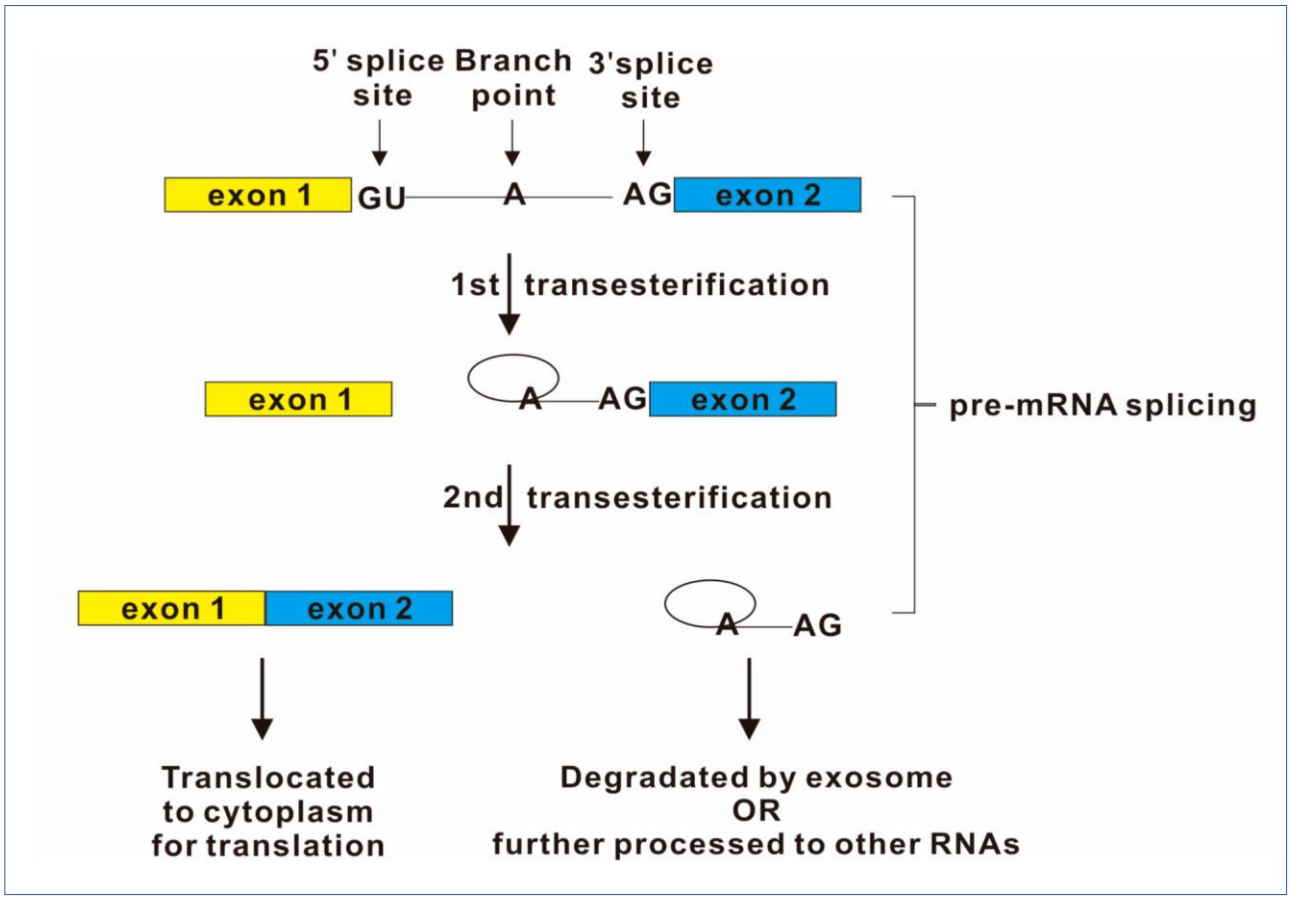
1.5.5 Abnormal splicing of mRNA

1.5.5.1 The process of RNA splicing

RNA splicing is a tightly regulated and highly complex pathway necessary for the removal of non-coding material from precursor messenger mRNA (pre-mRNA) and formation of mature messenger RNA (mRNA). The regions retained in the mature RNA molecule after RNA splicing are termed 'exons,' which refers to these regions in both RNA and DNA sequences. The regions removed in splicing are termed 'introns.' These terms were introduced by Walter Gilbert in 1978 (138). The annotation of exons and introns is transcript specific. A region that is exonic in one transcript may in fact be intronic in another transcript, due to the process of alternative splicing (139).

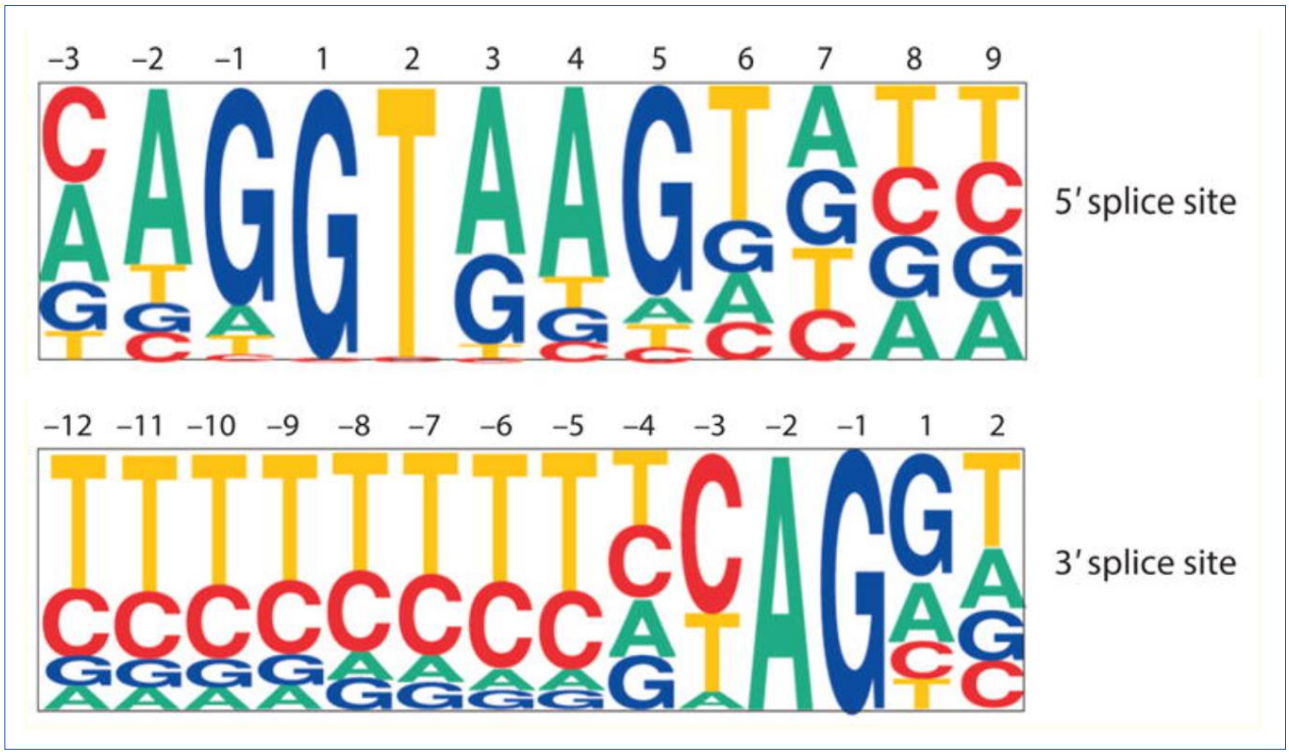
Splicing involves a series of reactions which are catalyzed by the spliceosome, a complex of small nuclear ribonucleo proteins (snRNPs). An overview of the process is shown in **Figure 1.14**. Key regions for splicing to occur are the 5' splice site, also known as the donor splice site, and the 3' splice site, also known as the splice acceptor site. The consensus sequences of these two sites are shown in **Figure 1.15**. Many other components influence splicing and the recognition of the exon-intron boundary by the splicing machinery. These include Exonic Splicing Silencer sites, Exonic Splicing Enhancer sites, Branch points and Polypyrimidine tracts. The Polypyrimidine tract is a region rich in pyrimidine nucleotides just prior to the 3' end of the intron that promotes assembly of the spliceosome (140).

Figure 1.14 Splicing of pre-mRNA to form mRNA



Pre-mRNA splicing includes intron exclusion and exon ligation. In most cases, introns start from the sequence GU as 5' splice sites (the donor splice site) and end with the sequence AG as 3' splice site (the acceptor splice site). A highly conserved nucleotide A at the branch site located 20–50 bases upstream of the 3' splice site (141). Image taken from (141).

Figure 1.15 Consensus sequences of 5’ and 3’ splice sites



Approximately 20,000 5’ and 3’ splice sites from annotated GenBank files were extracted and aligned to show the most common bases occurring at the splice sites (142,143). In these pictograms, the size of a letter corresponds to the frequency with which that base is present at each position. For the 5’ splice-site consensus sequence, the position labelled 1 is the first nucleotide of the intron, and the position –1 is the last nucleotide of the upstream exon. For the 3’ splice-site consensus, the position labelled –1 is the last nucleotide of the intron, and the position 1 is the first nucleotide of the downstream exon (144).

1.5.5.2 Naturally occurring alternative splicing

Alternative RNA splicing enables a single gene to code for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final mRNA sequence (145). When the different mRNA transcripts are translated, the amino acid sequences result in distinct proteins often with diverse biological functions. Whilst the human genome only contains around 20,000 protein-coding genes, alternative splicing enables far more than 20,000 proteins to be created. In humans, around 95% of multi-exonic genes are alternatively spliced (146). There are numerous modes of alternative splicing observed. The most common is exon skipping, where a

particular exon may be included in mRNAs under some conditions or in particular tissues, and omitted from the mRNA in others (145).

1.5.5.3 Abnormal splicing affecting gene function

It has been suggested that one third of all disease-causing mutations are due to abnormal splicing (147). Commonly this is due to mutations occurring directly at the 'canonical' or classic intron-exon boundary splice site resulting in reduced or complete loss of splicing at that site. This can lead to the exon being excluded from the mRNA or the spliceosome incorrectly recognising nearby alternative dormant splice sites and a mutant form of the exon being included in the mRNA. This causes an abnormal length of RNA with a range of possible effects on protein structure, including frameshift if the inclusion causes a change to the reading frame. Mutations occurring in the polypyrimidine tract have also been recognised to cause disease, including Growth Hormone Insensitivity. A homozygous mutation of the polypyrimidine tract just prior to exon 8 of the *GHR* (c.785-6T>A) led to complete skipping of exon 8. This causes frameshift to the *GHR*, a premature stop codon and classic Growth Hormone Insensitivity (148).

Intronic mutations can also cause abnormal splicing due to the inclusion of 'pseudoexons' into the mature mRNA, most often due to the creation of a new donor or acceptor splice site within an intronic sequence. Intronic sequences that are flanked by non-canonical splice sites and are normally not seen in mature mRNA are referred to as 'pseudo-exons' or cryptic exons. In comparison to genuine exons, pseudo-exons tend to have fewer splicing enhancer and more splicing silencer motifs (149–152). The inclusion of these pseudoexons can have significant effects on the resulting protein especially if their inclusion leads to a frameshift. Mutations resulting in aberrant pseudoexon inclusion have been found to be disease-causing in more than 50 genes (153). Disease-associated pseudoexons range in size from 30 to 344bp (154). It is likely that pseudoexons are

currently under-recognised as a cause of disease and with the advent of whole genomic sequencing our exploration and understanding of non-coding variants is rapidly advancing.

1.5.6 Copy number variation (CNV)

Advances in chromosomal microarray technologies have identified chromosomal imbalances across the human genome. These imbalances or copy number variations (CNVs), comprise deletions or duplications which can affect single or multiple genes or sections of a chromosome. The simplest explanation for CNVs exerting an effect on the phenotype of a patient is that they delete or duplicate one or more critical gene(s) responsible for a certain trait. CNVs are part of natural genetic variation and are often classified as 'benign.' However, the impact of CNVs on complex human traits remains understudied. A recent study assessed CNVs in UK Biobank participants, performing GWAS between the copy number of CNV-proxy probes and 57 continuous traits. This study found that total CNV burden negatively impacted 35 traits, leading to decreased intelligence, increased adiposity and liver/kidney damage. 30 traits remained burden associated once correcting for CNV-GWAS signals, suggesting a polygenic CNV architecture (155). Several rare CNVs are recognised to cause complex neurological traits such as autism and schizophrenia and also affect the susceptibility to HIV, Crohn's disease and psoriasis (156–163).

The interpretation of potential disease-causing CNVs is challenging as they exhibit variable penetrance and/or variable expressivity (i.e. not all patients carrying the same CNV have the recognised disease phenotype or disease severity) (164). Nevertheless, identification of CNVs in patients is important as it secures a diagnosis and facilitates the identification of key genomic regions and/or genes important for normal physiological processes. Whilst CNVs are recognised in

many short stature syndromes, they have not previously been investigated in patients with Growth Hormone Insensitivity.

1.6 Pathogenicity scores for genetic variants

Several pathogenicity scoring systems exist for assessing genetic variants.

Polyphen-2 scores can be generated for non-synonymous Single Nucleotide Polymorphisms (SNPs) that alter an amino acid (http://genetics.bwh.harvard.edu/pph2). It is an automatic tool based on computer algorithms that predict the impact of an amino acid substitution on the structure and function of a human protein. The tool compares the properties of the two different amino acids, maps the substitution site to known protein structures, predicts if the structure would be affected by the substitution and if a key domain of the protein would be altered. With this information, SNPs are predicted to be Probably Damaging (most confident prediction), Possibly Damaging (less confident prediction due to higher likelihood of a false positive result based on existing datasets) or Benign. A score of between 0.0 and 1.0 is generated for each variant. 0.0 to 0.15 is predicted to be Benign, 0.15 to 0.85 Possibly Damaging and 0.85 to 1.0 Probably Damaging. For example, an amino acid substitution that had very different properties to the original and occurred in a critical region of the protein is likely to be predicted to be Probably Damaging (165).

Sorting Intolerant From Tolerant (SIFT) scores (http://blocks.fhcrc.org/sift/SIFT.html) also use a computer algorithm to predict whether an amino acid substitution affects protein function. In contrast to Polyphen-2, the SIFT score ranges from 0.0 (deleterious) to 1.0 (tolerated). Variants with scores 0.0 to 0.05 are considered deleterious, and the closer to 0.0 the more confidence the variant is deleterious. Variants with scores 0.05 to 1.0 are predicted to be tolerated (benign). The closer the score is to 1.0 the more confidently the variant is predicted to be tolerated (166).

Combined Annotation Dependent Depletion (CADD) Score (https://cadd.gs.washington.edu/) is a tool for assessing the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome. Unlike SIFT and Polyphen-2, it is not limited to missense changes or coding regions and can generate scores for intronic variants. The score considers allelic diversity, annotations of functionality, pathogenicity, experimentally measured regulatory effects and complex trait associations. A scaled-C score is generated for each variant ranking it relative to all possible substitutions of the human genome (8.6×10^9). A score of greater than or equal to 10 indicates the variant is predicted to be amongst the 10% most deleterious changes that can occur to the human genome. A score of greater than or equal to 20 indicates the 1% most deleterious changes, 30 the 0.1% deleterious and so on (167,168).

For CNVs, a different scoring system is used. CNVs are usually divided into 5 categories based on evidence including population, computational, functional and segregation data in line with accepted best practice guidelines from the American College of Medical Genetics (ACMG) (169). These 5 categories are: class 1 (benign), class 2 (likely benign), class 3 (variant of uncertain significance; VUS), class 4 (likely pathogenic) and class 5 (pathogenic) (169).

1.7 Our cohort: investigations to date

1.7.1 Cohort

The cohort comprised subjects who were referred to our centre for genetic analysis between 2008 and 2020. Professor Helen Storr directs an international referral centre for patients with undiagnosed severe short stature many with suspected defects in the GH-IGF-1 axis. These patients were investigated at their home institutions and no genetic diagnosis had been identified at the time of their referral.

Referring physicians completed a proforma detailing the clinical and biochemical data at the time of sending DNA samples. The referring clinicians excluded growth hormone deficiency (basal or peak GH level of 10 and 7 μg/l, respectively) and causes of secondary GHI/IGF-1 insensitivity (e.g. undernutrition). Birth weight, height and BMI were expressed as Standard Deviation Scores (SDS) according to appropriate national standards. Biochemical investigations included: basal and/or peak GH and basal IGF-1 levels. IGF-1 was expressed as SDS based on the age and sex appropriate range provided by the institution. Where serum IGF-1 was undetectable (less than the lower limit of the assay), we calculated the lowest possible detectable SDS and assigned this value for the statistical analysis. In these patients, the IGF-1 SDS ranged between -2.4 and -3.0 but this is likely to underestimate the degree of IGF-1 insensitivity.

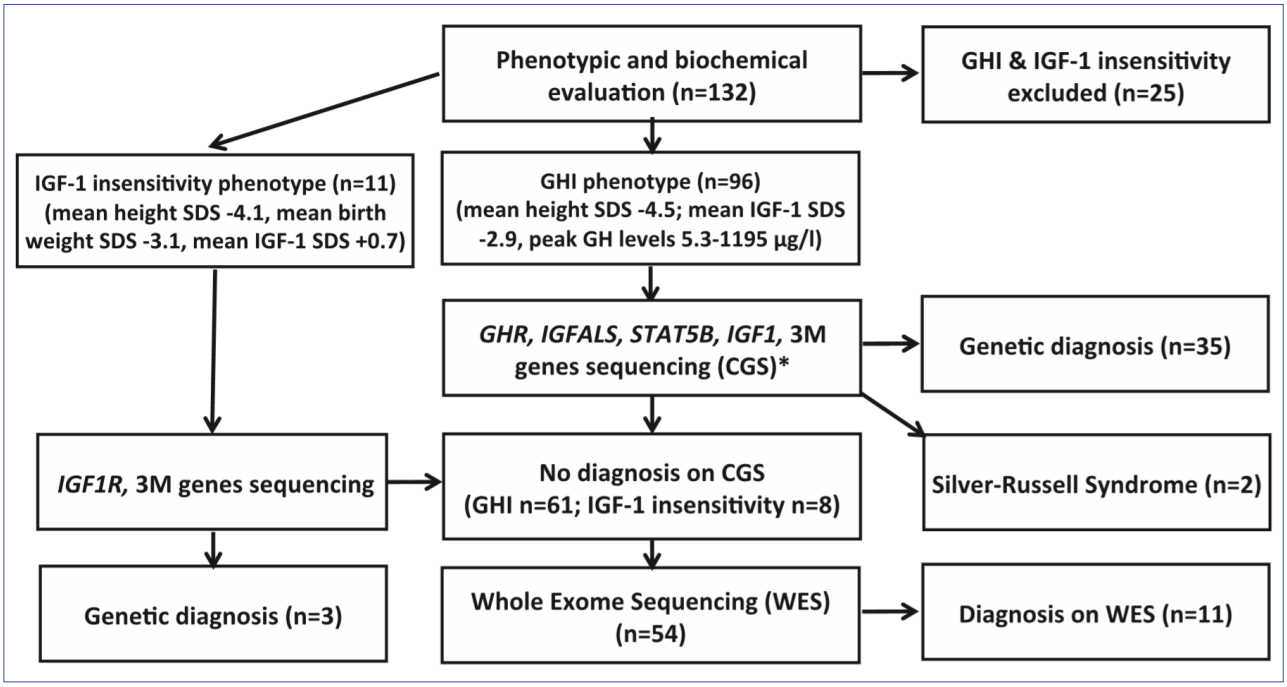
1.7.2 Previous genetic investigation in our patient cohort

Our centre is an international referral centre for patients with suspected defects in the GH-IGF-1 pathway. Prior to my research period, patients had been investigated by two main approaches, candidate gene sequencing (CGS) and whole exome sequencing (WES). CGS considers the biochemical and phenotypic features of each patient and involves sequencing key genes suspected to be responsible. Whole exome sequencing involves sequencing the protein-coding exons of every gene, casting a wider net for diagnosis then filtering the genetic defects identified based on pathogenicity, mean allele frequency, and so forth. Patients that had not been diagnosed following a CGS approach were analysed by WES.

These approaches identified genetic diagnoses in 38/107 (36%) patients by candidate gene sequencing (CGS) and a further 11/107 (10%) by whole exome sequencing (WES). This cohort of 107 short patients included 96 with GHI and 11 with IGF-1 insensitivity. Two patients in this cohort

had hypomethylation in the H19/IGF2 region or mUPD7 consistent with SRS (total with genetic diagnosis 51/107, 48%) (109) (**Figure 1.16**).

Figure 1.16 Findings of CGS and WES sequencing in the GH/IGF-1 Insensitivity cohort prior to my period of research



3M syndrome genes, *CUL7*, *CCDC8* and *OBSL1*; BW, birth weight. *The candidate genes sequenced depended on the clinical and biochemical features. The majority of patients were screened for mutations in the Growth Hormone Receptor gene (*GHR*) +/− *IGFALS*. Other genes were selected depending on the phenotype e.g. *STAT5B* if there was evidence of immune deficiency/ eczema/atopy and *IGF1* and 3M genes if birth weight SDS was ≤2.0 SDS. (109).

Following candidate gene sequencing +/- whole exome sequencing +/- SRS testing, a diagnosis was achieved for 51/107 (48%) of our patients. The remaining 56/107 (52%) of patients had no diagnosis (109).

1.8 Thesis rationale

Following CGS and WES, many patients remained without a diagnosis despite having short stature and classic features of GH/IGF-1 Insensitivity. This was hypothesised to be due to undiscovered growth genes, non-canonical/non-exonic variants or CNVs in recognised growth genes. To try and identify the genetic cause of disease in the patients without a genetic diagnosis, I planned to create

a custom next-generation targeted short stature gene panel and in parallel assess the undiagnosed patients for CNVs by array Comparative Genomic Hybridisation (array CGH).

WES focuses on the protein-coding exons but excludes many exonic regions that don't form part of the final coding sequence, termed untranslated regions or 'UTRs'. It also excludes many non-coding RNAs, which are often subject to splicing but are not translated (170). More than 50,000 loci in the human genome transcribe long non-coding RNAs (lncRNAs) (171,172). These are transcripts over 200 nucleotides in length that have low or no protein-coding potential (173). Although these lncRNAs outnumber protein-coding genes in mammalian genomes, they are generally expressed at lower levels, more poorly conserved and more cell-type specific (171,174). However, the structural motifs of lncRNAs and the evolutionary conservation of lncRNA promoters would indicate that they play key roles in biological regulation (175–177). Of the few hundred lncRNAs that have been well characterised so far, important regulatory roles in transcription, translation and chromatin state have been demonstrated (178–180).

Whilst WES examines the intronic regions just adjacent to the protein-coding exons to identify mutations to the canonical splice sites, it doesn't sequence the vast intronic material that lies between these exons. Creating a custom short stature gene panel which examines the entire genomic sequences for specific growth genes of interest would enable not only the rapid identification of coding variants in recognised genes, it would also allow for detailed exploration of non-coding variants. I hypothesised that by analysing the entire genomic sequence including UTRs, non-coding RNAs and all intronic material, I would be able to identify novel genetic causes of growth failure in these undiagnosed patients. For example, it would enable me to interrogate intronic variants which may exert aberrant effects on splicing, akin to the one previously reported *GHR* pseudoexon. It would also enable exploration of changes in 5'UTR, which may affect ATG start sites

of the gene and other non-coding regions with important regulatory roles, many of which may be currently unrecognised. With this work I aimed to improve the understanding of non-classic mechanisms of genetic disease.

Whilst WES and the custom short stature gene panel would identify single nucleotide polymorphisms and small insertions and deletions, they are not able to reliably detect larger deletions or duplications of genetic material. 'Copy number variants' (CNVs) can range in size and contain part of one gene or several hundred genes. Assessing our undiagnosed patients for CNVs enabled an unbiased identification of altered regions which may reveal novel candidate genes with important roles in growth. CNVs may be identified that have been seen in other patients with similar features or recognised deletion/duplication syndromes associated with short stature, in this way achieving a diagnosis for patients. If a gene or gene(s) in these regions responsible for growth are not established, bioinformatic analysis can help to identify novel potential candidate genes and enriched pathways in our patients with CNVs.

Widening the spectrum of recognised genetic defects affecting the normal function of the GH-IGF-1 axis will improve our understanding of the normal physiology of this complex pathway. Non-classic Growth Hormone Insensitivity is an important clinical entity and may have higher prevalence than classic GHI. It is likely to be under-recognised currently due to the less severe postnatal growth failure and absence of the dysmorphic features classically associated with GHI.

An accurate genetic diagnosis in growth failure is critical for patients, families and clinicians to appreciate disease implications and define prognosis. A definitive diagnosis ends uncertainty, enables genetic counselling and can secure funding for growth promoting therapies. It also helps predict disease progression and informs about co-morbidities that may require additional screening.

The aim of this research was to further our understanding of the genetic causes of growth failure and increase the understanding of the intricate pathways essential for normal linear growth.

Chapter 2. Methods

2.1 Ethical approval

Informed written consent for genetic research and publication of clinical details was obtained from patients and/or their parents. The study was approved by the Health Research Authority, East of England - Cambridge East Research Ethics Committee (REC reference: 17/EE/0178).

2.2 Patient cohort

Subjects were investigated at their home institutions and referring physicians completed a proforma detailing the clinical and biochemical data at the time of sending DNA samples. The referring clinicians excluded growth hormone deficiency (basal or peak GH level of 10 and 7µg/l, respectively) and causes of secondary GHI/IGF-1 insensitivity (such as poor nutrition or chronic disease). Birth weight, height and BMI were expressed as Standard Deviation Scores (SDS) according to appropriate national standards. Biochemical investigations included: basal and/or peak GH and basal IGF-1 levels. IGF-1 was expressed as SDS based on the age and sex appropriate range provided by the institution. Where serum IGF-1 was undetectable (less than the lower limit of the assay) we calculated the lowest possible, detectable SDS and assigned this value for the statistical analysis. In these patients, the IGF-1 SDS ranged between -2.4 and -3.0 but this is likely to underestimate the degree of IGF-1 insensitivity. Subjects were referred to our centre for genetic analysis between 2008 and 2021.

2.2.1 Patients assessed on the short stature gene panel

I have assessed 92 patients with suspected defects in the GH/IGF-1 axis on the short stature gene panel alongside 13 control samples. 37 patients were female and 55 male, with a mean age at referral of 8.32 years (range 0.5 to 59 years), mean height SDS -3.67 (range -1.6 to -7.44). 17/92

(13%) patients were born small for gestational age (SGA), with mean birth weight (BW) SDS -1.08 (range 1.64 to -4.66). All patients had height SDS ≤-2 except for one younger sibling of an index case who was included as she shared her brother's clinical phenotype and had IGF-1 deficiency (height SDS -1.6).

2.2.2 Patients assessed for copy number variation (CNVs)

For genetic analysis of copy number variation, patients were divided into Growth Hormone Insensitivity and unexplained short stature cohorts.

Growth Hormone Insensitivity (GHI) subjects

We assessed 54 patients with GHI (36 males, mean age 7.3 years, median 7.0 years, range 0.5-17.0 years). All had SS (mean height SDS -3.8, median -3.7, range -1.6 to -7.4 SDS) and evidence of GHI (normal/high levels of growth hormone and IGF-1 deficiency; mean IGF-1 SDS -2.5; range -2.0 to -4.1). Mean birth weight (BW) SDS was -1.1 (median -0.95, range 1.1 to -4.8) and 9/54 (17%) GHI subjects were born SGA (BW SDS <-2.0). All patients had height SDS ≤-2 except for one younger sibling of an index case who was included as she shared her brother's clinical phenotype and had IGF-1 deficiency (height SDS -1.6, IGF-1 SDS -2.4).

Unexplained short stature subjects

We assessed 10 patients with unexplained short stature (6 male; mean age 5.8 years, median 4.25 years, range 1.1-16.5). All had short stature (mean height SDS -3.4, median -3.5, range -2.0 to -4.5) and normal/high GH and IGF-1 levels. Mean BW SDS was -2.0 (median -2.12, range -0.3 to -3.8) and 6/10 were born SGA (BW SDS <-2.0).

2.3 Genetic analysis

2.3.1 DNA extraction from whole blood

Blood was collected in EDTA tubes and Deoxyribonucleic acid (DNA) extraction performed using the GE Healthcare Nucleon™ BACC2 Genomic DNA Extraction Kit (catalogue number 10637265, Fisher Scientific, UK). Composition of buffers and reagents from this kit can be found in **Appendix 1.** Using an aseptic procedure, four times the volume of Reagent A was added to each blood sample in a 50ml falcon tube and placed on a rotary mix for 4 minutes at room temperature. The sample was then centrifuged at 1300g for 5 minutes. The supernatant was discarded and 2ml Reagent B added to the pellet. The pellet was resuspended by Vortex and the suspension transferred to a 15 ml Falcon tube. 500µl of sodium perchlorate solution was added to the solution and the tube inverted 7 times. 2mls of chloroform was then added, inverting the tube 7 times. 300µl of Nucleon resin was added to the tube, taking care to avoid remixing the phases. The sample was then centrifuged at 1300g for 3 minutes. Holding the tube vertically without disturbing the nucleon resin layer, the upper phase was transferred into a clean tube (approximately 2.5ml). Twice the volume of cold absolute ethanol was added to the solution and mixed by inversion until precipitate became visible. This precipitate was hooked using a heat-sealed Pasteur glass pipette and suspended in TE in an Eppendorf tube. After rotary mixing the sample at 4°C for a minimum of 2 hours, the DNA sample quality and concentration was assessed using a Nanodrop spectrophotometer.

2.3.2 Creation of the custom short stature gene panel

The gene panel was created to enable detailed exploration of key genes of interest in GHI and overlapping syndromes. The panel included 60 genes of interest, 3 non protein-coding regions and one intergenic region (**Figure 2.1**). Whole genomic sequences were included for each gene, including the intronic regions, 2kb upstream and 500bp downstream of the gene. This enabled detailed

exploration and assessment of non-coding regions in known genes (in addition to the coding/exonic sequences), which could potentially uncover diagnoses for patients with phenotypes classic for a genetic defect but with no exonic pathogenic variants identified. This work would therefore develop our understanding of non-coding variants.

Figure 2.1 Genes included in the custom short stature gene panel

GH-IGF-1 axis	3M	Noonan	Other SS genes of interest
	OBSL1	***PTPN11***	
GHR	***CCDC8***	***SOS1***	***ACAN***
IGF1	***CUL7***	***SOS2***	***NPR2***
IGFBP1	*CUL9*	***RAF1***	***SHOX***
IGFBP2	*SEC16A*	*RASA2*	*TENM4*
IGFBP3	*FBXW8*	***LZTR1***	*ADAMTS3*
IGFBP4	*ANKRA2*	***CBL***	*IL2RG*
IGFBP5	*ROC1*	***SHOC2***	*CHD1L*
IGFBP6	*SKP1*	*ARAF*	*STC1*
PAPPA		***BRAF***	*STC2*
PAPPA-2	**SRS**	***HRAS***	*NOX4*
STAT5B		***KRAS***	*IKBKB*
STAT3	*MEG3*	***NRAS***	*IL11RA*
IGF1R	*H19*	*RRAS*	
IGFALS	*KCNQ1OT1*	***NF1***	
JAK2	*ICR1*	***RIT1***	
SOCS2	***IGF2***	*A2ML1*	
	IGF2R	***MAP2K1***	
	HMGA2	***MAP2K2***	
	PLAG1		

SRS, Silver Russell Syndrome; SS, short stature. Shown in bold are genes in which mutations have been recognised to cause growth failure. Shown in non-bold are genes with important roles in growth pathways in which I hypothesised mutations could lead to growth failure.

Genes were selected for the panel based on their relevance to GHI phenotypes. Recognised genetic causes of overlapping syndromes (SRS, 3M and Noonan) were included, in addition to other short stature genes of interest that may present with similar phenotypes. Several novel genes, which were good candidates, such as genes with key roles in known growth pathways but with no currently recognised human mutations causing growth failure, were also included.

A limited number of genes were included for Silver Russell Syndrome (SRS). As discussed previously, SRS is primarily a disorder of methylation and imprinting. Hypomethylation of the imprinted *H19/IGF2* domain of chromosome 11p15 (11p15LOM) and maternal uniparental disomy of chromosome 7 (upd(7)mat) are identified in 50-60% and 10% SRS cases, respectively (130). However, defects in single genes have also been identified in single genes causing SRS phenotypes and 40% of patients with clinical SRS remain without a genetic diagnosis (131). Since 2015, 5 mutations have been reported in *IGF2* causing growth failure and SRS features (133,181–183). These include a family with multiple affected members (paternal transmission) with a heterozygous nonsense substitution c.191C→A, a *de novo* heterozygous indel c.110_117delinsAGGTAA, a *de novo* heterozygous c.101G>A missense mutation, a *de novo* heterozygous nonsense mutation c.78C>G and a *de novo* heterozygous two-nucleotide duplication c.158_159dup (133,181–183). *IGF2* is a paternal effect gene and silenced on the maternal allele, and all reported cases were either paternal transmission or *de novo*. Since 2015, 5 mutations of *HMGA2* have been reported causing growth failure and SRS features (184–187). All of these mutations were heterozygous and caused a frameshift change to the protein: c.189delA, a microdeletion including exon 1 and 2, a microdeletion of exon 2, c.193C>T and a 7bp deletion just prior to exon 5 (184–187). *HMGA2* is not an imprinted gene and paternally inherited, maternally inherited and *de novo* mutations have been reported. *PLAG1* is a non-imprinted gene thought to be an upstream regulator of *HMGA2*. Two heterozygous

PLAG1 mutations causing frameshift were reported in 2017 in patients with growth failure and SRS phenotypes (188).

Once the genetic regions of interest were decided, I generated a list of co-ordinates of the target regions to be covered by the panel. Otogenetics (Otogenetics Corporation, 4553 Winters Chapel Road, Ste 100 Atlanta, GA CLIA CERTIFIED 11D2066426, GA St Clinical laboratory License 067-071) designed probes to cover this region in as much detail as possible, within the limitations of highly repetitive regions. The total number of probes was 89527, and the average coverage of the panel for the regions of interest was 97.05% (**Table 2.1**). Probe design, preparation of libraries, capture and sequencing was performed by Otogenetics. Sequencing was performed using an Illumina HiSeq 2500 platform (paired ends 100-125, designated average coverage of 100x). Otogenetics performed data mapping, duplicate removing, snv/Indel calling, vcf annotation and generated VCF, BAM and Bam.bai files that I used as the input for my bioinformatic analysis.

Table 2.1 Details of genomic regions included in custom short stature gene panel

Gene	Chromosome	Start point	End point	Size of region (BP)	Coverage (%)
NRAS	1	115245091	115260015	14924	99.34
CHD1L	1	146712292	146767943	55651	99.04
RIT1	1	155865600	155881695	16095	93.32
PAPPA-2	1	176430308	176815235	384927	98.36
OBSL1	2	220413452	220437081	23629	99.78
SOS1	2	39206538	39351986	145448	98.04
IGFBP2	2	217495552	217529659	34107	96.25
IGFBP5	2	217534829	217560748	25919	99.35
RAF1	3	12623101	12706225	83124	92.72
RASA2	3	141203890	141334684	130794	98.10
ADAMTS3	4	73144687	73435016	290329	95.68
GHR	5	42421880	42722479	300599	94.39
ANKRA2	5	72846161	72862011	15850	99.02
SKP1	5	133482634	133513229	30595	96.04
STC2	5	172739717	172757006	17289	100.00

CUL7	6	43003356	43022183	18827	95.59
CUL9	6	43147914	43192825	44911	97.05
IGF2R	6	160388132	160535039	146907	98.23
IGFBP3	7	45925957	45933767	7810	100.00
BRAF	7	140417128	140625064	207936	95.09
IGFBP1	7	45925957	45933767	7810	100.00
IKBKB	8	42126821	42190473	63652	96.31
PLAG1	8	57,071,464	57124383	52919	97.79
STC1	8	23697429	23712820	15391	99.73
PAPPA	9	118914084	119165101	251017	99.46
JAK2	9	4983034	5128683	145649	93.10
SEC16A	9	139332550	139372641	40091	97.83
NPR2	9	35790152	35810229	20077	99.50
IL11RA	9	34648700	34662389	13689	99.77
SHOC2	10	112677302	112773925	96623	98.47
H19	11	2014407	2023200	8793	100.00
KCNQ1OT1	11	2627559	2721724	94165	93.60
ICR1	11	2016813	2025240	8427	100.00
IGF2	11	2148343	2171033	22690	96.06
CBL	11	119074753	119179359	104606	96.97
HRAS	11	530243	537787	7544	100.00
TENM4	11	78361877	79152492	790615	98.27
NOX4	11	89055525	89323279	267754	95.53
IGF1	12	102787646	102874923	87277	94.71
FBXW8	12	117346762	117469453	122691	98.26
PTPN11	12	112854156	112948217	94061	95.83
KRAS	12	25355724	25404370	48646	98.21
A2ML1	12	8973069	9040097	67028	95.09
IGFBP6	12	53489221	53496629	7408	97.75
HMGA2	12	66215912	66360575	144663	98.66
SOCS2	12	93961591	93977763	16172	99.08
SOS2	14	50581848	50698776	116928	97.31
MEG3	14	101243748	101327868	84120	98.49
IGF1R	15	99190201	99508259	318058	99.25
MAP2K1 (MEK1)	15	66677156	66785150	107994	94.02
ACAN	15	89344675	89419085	74410	99.63
IGFALS	16	1838415	1845472	7057	100.00
STAT5B	17	40349187	40429225	80038	94.78
STAT3	17	40463343	40541086	77743	94.39
NF1	17	29419946	29709634	289688	96.41
IGFBP4	17	38597703	38614483	16780	96.39

CCDC8	19	46911630	46917341	5711	99.39
RRAS	19	50136550	50143958	7408	98.89
MAP2K2 (MEK2)	19	4088320	4124626	36306	99.07
ROC1	22	41345352	41369813	24461	94.35
LZTR1	22	21331752	21353827	22075	99.04
ARAF	X	47418517	47431807	13290	99.35
IL2RG	X	70325255	70332458	7203	98.97
SHOX	X	583080	620646	37566	91.74
Total size of 64 genetic regions (BP)				5921965	
Total size gene panel (Mb)				5.92	
Average coverage of target regions					97.05

BP, base pairs; Mb, Megabases. Co-ordinates and coverage of regions included in the GHI custom panel including 2000 bases upstream and 500 downstream of genetic regions of interest. *H19, KCNQ1OT1* and *MEG3* are non-coding and *ICR1* is an intergenic region. GRCh37 (http://grch37.ensembl.org/index.html) was used as the reference genome for generating the co-ordinates of each region.

2.3.3 Array Comparative Genomic Hybridisation (aCGH) to identify copy number variants (CNVs)

2.3.3.1 Array details

DNA samples were analysed by aCGH at an ISO15189 accredited genetics laboratory, using a 60K oligonucleotide array (Agilent design 028469 or 085030) as previously described (189). Briefly, 1µg DNA was labelled using CGH Labelling Kit for Oligo Arrays (Enzo Life Sciences, USA), labelled DNA was purified post-labelling using QIAquick PCR purification Kit (catalogue number 28104, Qiagen, Crawley, UK), DNAs were applied to a 60K oligonucleotide array (Agilent, USA) and hybridisation, washing and scanning was performed following the manufacturers' protocols.

2.3.3.2 Array CGH data analysis

Array CGH data analysis was undertaken as previously described (189). Fluorescence signal intensity analysis was performed using Feature Extraction software (Agilent, USA). CNV detection was

performed using Genomic Workbench software (Agilent, USA) and the ADM-2 algorithm (threshold 6). Secondary analysis was performed using the ADM-1 algorithm (threshold 6) for detection of low-level mosaicism (190). Population polymorphisms were filtered out and each CNV was assessed for pathogenicity in the context of each subject's phenotype following the Association for Clinical Genetic Science (ACGS) best practice guidelines (191). GRCh37 (http://grch37.ensembl.org/index.html) was used as the reference genome. Where possible, parental samples were obtained and segregation studies were performed to ascertain the parent of origin or confirm *de novo* events. The median resolution of our CGH array was 120kb. Five GHI subjects were analysed at a reduced resolution (ranging from 1Mb-10Mb) due to poor DNA quality. No significant CNVs were identified in any patients analysed at reduced resolution.

2.3.3.3 CNV classification

CNVs were classified into 5 categories based on evidence including population, computational, functional and segregation data in line with accepted best practice guidelines for SNVs (169). This included: class 1 (benign), class 2 (likely benign), class 3 (variant of uncertain significance; VUS), class 4 (likely pathogenic) and class 5 (pathogenic) (169). Class 1 or 2 CNVs were discarded and only those considered pathogenic, likely pathogenic or VUS were investigated further.

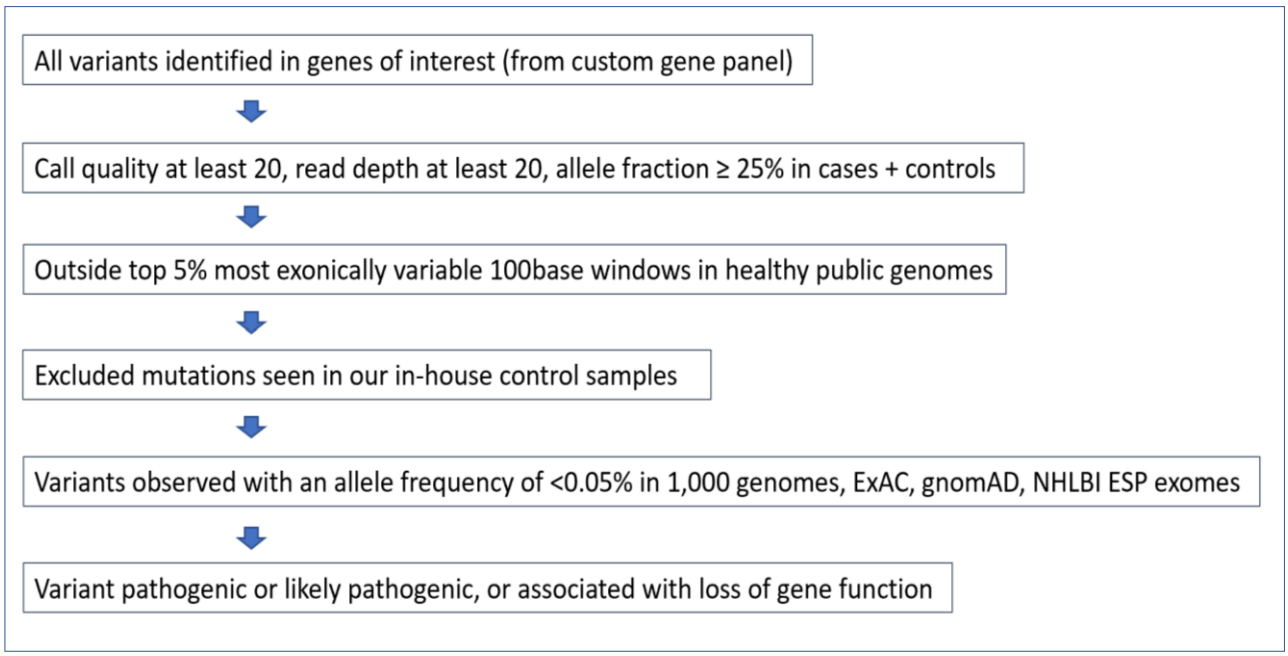
2.4 Bioinformatic analysis

2.4.1 Ingenuity Variant Analysis

Ingenuity Variant Analysis (IVA) (https://www.qiagenbioinformatics.com/ products/ingenuity-variant-analysis from QIAGEN, Inc) is a bioinformatics tool for rapidly filtering genetic variants (192). To use the IVA programme, I uploaded Variant Call Files (VCFs) which annotate any changes seen in the patient and compare it to the reference genome. VCFs can be generated from several next

generation sequencing techniques including whole exome sequencing (WES), whole genome sequencing and targeted gene panels. VCFs contain thousands of genetic variants per patient, many of which are synonymous and IVA allows filtering based on a number of parameters e.g. type of variant or inheritance pattern. For example, it is possible to select parameters to identify rare compound heterozygous mutations in a gene when analysing a trio of patient with unrelated parents. I used IVA to filter the variants in VCFs generated from the short stature custom gene panel. An example of a filtering pipeline is shown in **Figure 2.2**.

Figure 2.2 Filtering pipeline to identify pathogenic variants from short stature gene panel



13 patients were assessed on the short stature gene panel as control samples. 12 of these have known defects in genes featured on the panel and 1 was a healthy patient with a normal phenotype. 1,000 genomes (https://www.internationalgenome.org/); ExAC genome browser (http://exac.broadinstitute.org/gene/ENSG00000086200); gnomAD genome aggregation database (https://gnomad.broadinstitute.org/) NHLBI ESP exomes (https://evs.gs.washington.edu/EVS/)

This example filtering pipeline enabled the rapid identification of clearly pathogenic exonic mutations responsible for a patient's phenotype. However, searching for novel and non-classic variants required some alterations to this pipeline. For example, non-coding variants are often predicted benign based on the algorithms in place, so I filtered for very rare, homozygous variants

in patients that remained undiagnosed but removed the criteria for the variant to be predicted pathogenic or associated with loss of gene function.

A significant advantage of using tools such as IVA is the speed of filtering and volume of variants which can be assessed at one time. This was invaluable when assessing the wealth of data generated from increasingly sophisticated genetic analysis, and IVA provided detailed annotation including pathogenicity scores from three different sources and links to gene summaries. In cases where the mutation has been previously published, it also provides details of the publication. It allows the user to tailor each genetic search to the genetic variants in which they are interested. However, it is limited in its assessment of non-coding variants.

2.4.2 *In silico* splicing prediction software

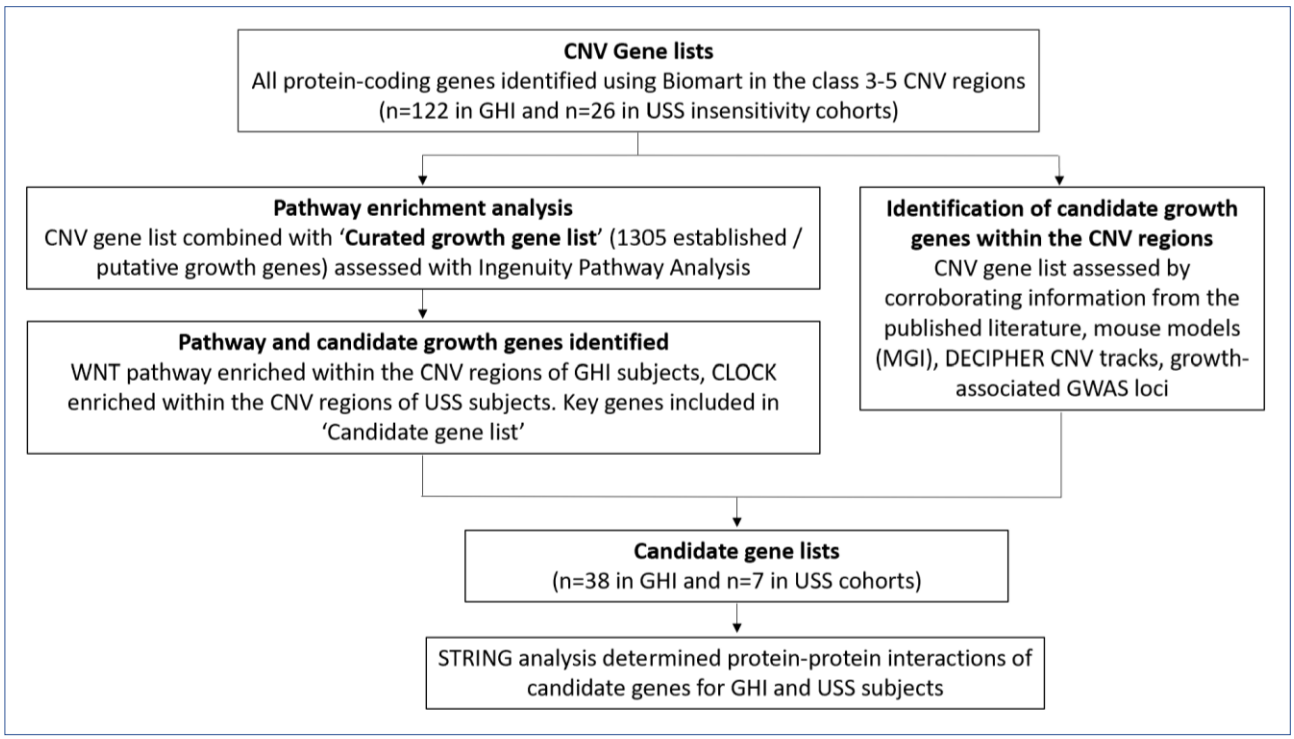
Whilst IVA is able to identify variants that may alter splicing if they affect the canonical splice sites directly, it does not identify variants deep within non-coding regions that may have aberrant effects on splicing. For this reason, I used other splicing prediction tools to assess non-coding variants and their potential for affecting splicing. Human Splicing Finder (http://umd.be/HSF3/) was a valuable tool for splice predictions (193). I identified all intronic homozygous variants with an allele frequency of ≤0.05% in the 1,000 genomes, ExAC, gnomAD and NHLBI ESP exomes. I then assessed the list of variants generated using Human Splicing Finder (http://umd.be/HSF3/) which calculated the consensus values of potential splice sites, splice enhancer and splice silencer sites.

2.4.3 Assessment of genes within the identified CNV regions

Class 3-5 CNV regions identified were explored using bioinformatic analysis techniques to identify key pathways and candidate growth genes (**Figure 2.3**). A 'CNV gene list' was generated. This included all the protein coding genes within the class 3-5 CNV regions identified in the subjects.

These genes were identified using Biomart (http://grch37.ensembl.org/biomart/) (CNV gene list, **table 4.4**). The CNV gene list was evaluated in detail to identify potential candidate growth genes and generate the 'Candidate gene list.'

Figure 2.3 Flowchart showing the bioinformatic pipeline used to identify key pathways and candidate growth genes within the class 3-5 CNV regions identified in the subjects



Given the distinct phenotypes, the CNV regions identified in the GHI and unexplained short stature patients were assessed separately. CNV, copy number variant; GHI, Growth Hormone Insensitivity; USS, unexplained short stature; Biomart http://grch37.ensembl.org/biomart/; MGI, Mouse Genome Informatics http://www.informatics.jax.org/; DECIPHER https://decipher.sanger.ac.uk/; GWAS catalogue, Genome Wide Association Studies catalogue https://www.ebi.ac.uk/gwas/; STRING database https://string-db.org/.

UCSC genome browser enabled visualisation of DECIPHER data (http://genome.ucsc.edu/) (194). DECIPHER genome data tracks aided the identification of overlapping CNVs previously reported in growth failure subjects and therefore allowed the identification of key regions within the CNVs identified in our subjects (195). Mouse Genome Informatics (MGI) identified genes within the CNV gene list associated with pre- or post-natal growth failure phenotypes in mouse models

(http://www.informatics.jax.org/). Literature searches, Online Mendelian Inheritance in Man (OMIM, https://www.omim.org/) and pathway analyses also determined established and putative growth genes within the CNV regions. The NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/) was queried to identify genes from the CNV gene list with loci associated with "height." GWAS studies aim to elucidate genotype-phenotype correlations by assessing differences in the allele frequencies of genetic variants between individuals who share similar ancestry but display different phenotypes. Single-nucleotide polymorphisms are most commonly studied (196). Statistical analysis identifies genomic loci that are significantly associated with the phenotypic variable of interest, in this case 'height.'

2.4.3.1 Pathway enrichment analysis

Pathway enrichment analysis was performed by Dr Claudia Cabrera using Ingenuity Pathway Analysis (IPA) software (Qiagen, inc). IPA identified biological pathways and functions enriched within the CNV regions (CNV gene lists) identified in the GHI and unexplained short stature cohorts. To further investigate the enriched pathways and their role in growth, we created a 'Curated growth gene list', 1305 established and candidate growth genes generated from published data and in-house analysis (109,197–200). Overlaying the curated growth gene list with the pathway results allowed detection of pathways harbouring growth-related genes. Pathways with evidence of enrichment in more than 3 subjects were further investigated.

2.4.3.2 *In silico* protein-protein interaction analysis using Candidate gene lists

The combined bioinformatic analysis described above produced 'Candidate gene lists' for the GHI and unexplained short stature subjects (n=38 GHI, n=7 unexplained short stature). The STRING database (https://string-db.org/) was used to explore protein-protein interactions within the candidate genes. Default settings were applied with the exception of interaction sources, where

text mining and neighbourhood were excluded. Direct interactions between two candidate genes were explored and also those via intermediate proteins.

2.4.3.3 Statistical analysis of phenotypic predictors in CNV patients

Phenotypic predictors (height SDS, age, sex, BW SDS or IGF-1 SDS) associated with the identification of subjects with and without CNVs in the cohort were compared by Dr Sumana Chatterjee using 2-tailed t-tests and logistic regression analysis. The frequency of deletions vs duplications and the size of the CNVs between the 2 cohorts (GHI and unexplained short stature groups) were analysed by Fisher's exact t test and Mann-Whitney u test, respectively.

2.5 Verifying identified genetic variants of interest

2.5.1 Primer design

Primers were designed to verify variants of interest identified by the short stature gene panel and to assess segregation of the variant. The relevant region was identified using Ensembl (http://grch37.ensembl.org/index.html) and primers were designed to the following criteria:

- To amplify a genomic region 400-600bp in size.
- Oligonucleotide length between 17 to 27 nucleotides.
- GC content 40-60%.
- A maximum length of 4 mononucleotide repeats (e.g. AAAA).
- Primer pairs to have similar Tm (melting temperature values). This can be estimated by Tm = 2°C x (number of A and T residues) + 4°C x (number of G and C residues).
- Primers that form weak secondary structures only.
- Primer sequences that preferably end in a GC clamp (last 2 nucleotides in sequence are G or C).

Primers were designed manually and by using ExonPrimer (https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html). They were then entered into the BLAST search tool on Ensembl (http://grch37.ensembl.org/Multi/Tools/Blast) to check if the oligonucleotide sequences occurred elsewhere in the genome (thus increasing the risk of non-specific PCR products). Primers that occurred only once in the genome were preferentially ordered. Primer sequences used for genomic DNA amplification are provided in **Appendix 2.**

GRCh37 (http://grch37.ensembl.org/index.html) was used as the reference genome throughout this work. This was chosen by my PhD supervisors as they felt the most comfortable with this build and felt it would be most compatible with previously attained WES data mapped to GRCh37 and more easily allow combined analysis. However, it should be noted that an updated reference genome GRCh38 (https://www.ensembl.org/Homo_sapiens/Info/Index) was released in 2013 and most recently patched in 2019. This build uses more extensive alternate contigs to represent common complex variation, including HLA loci. It also corrects thousands of small sequencing artefacts present in the GRCh37 build that could cause false SNP and indels to be called (201). Furthermore, in 2022 the Telomere-to-Telomere (T2T) Consortium published a complete 3.055 billion base pair (bp) sequence of a human genome, T2T-CHM13. Made possible by high accuracy long read sequencing, this maps complex regions previously unchartered due to technological barriers and unlocks endless potential for comprehensive studies of genomic variation across the genome (202).

2.5.2 Agarose gel

1g of Agarose was combined with 100ml 1x Tris Acetate EDTA solution and heated until dissolved. Once cooled, 6 microliters of Gel RED™ Nucleic Acid Stain (catalogue number BT41003, Biotium, UK) was added, taking care to avoid contact with skin and inhalation. After arranging a gel cassette

and comb with desired number of wells securely in frame, the solution was added, any bubbles dispersed and the mixture left to solidify.

2.5.3 DNA loading dye

12.5ml distilled water was added to 25mg Orange G. Once dissolved, 12.5ml glycerol was slowly added and the solution mixed well. The loading dye was divided into 1ml aliquots and stored in a cool dark place until use. For use with agarose gels, 4µl DNA was added to 1µl loading dye to give a clear signal.

2.5.4 Polymerase Chain Reaction (PCR)

Ratio of reagents for PCR mix

Taq polymerase Buffer	2.5 µl
dNTP mix (10mM of each dNTP)	1 µl
Forward primer (10mM concentration)	0.5 µl
Reverse primer (10mM concentration)	0.5 µl
Taq polymerase	0.4 µl
Genomic DNA	100-200ng
DEPC treated water, DNA and RNA free	to make up 25 µl total volume

Total volume for each PCR reaction = 25 µl

A master mix can be made using all reagents except the DNA and primers (unless all reactions use the same DNA/primers) which will improve the accuracy by avoiding the use of small volumes. Scale up the volumes listed above depending on number of PCR samples.

Reagents were thawed on ice prior to use, except for *Taq* polymerase which was removed from the -20°C freezer just before use. A master mix was created by combining all reagents except DNA and primers into an Eppendorf tube, adding *Taq* polymerase last. The mixture was mixed well using the vortex. The master mix was pipetted into individual PCR tubes, to which the relevant primers and DNA were added. The mixture was mixed by pipetting up and down, ensuring there were no bubbles in the samples. The samples were loaded into the G-Storm GS1 thermal cycler in the central slots and the appropriate PCR program selected.

Touchdown PCR (in which the initial annealing temperature is higher than the expected primer melting temperature (Tm) with gradual decrease in temperature over subsequent cycling) was used for standard reactions. This reduces the likelihood of non-specific PCR products forming. The most common settings are shown below, but the temperatures used in steps b and c were adjusted depending on the Tm of primers used.

a. Denaturing: 95°C for 5 minutes

b. Touchdown: 10 x (95°C for 30 seconds, 65°C for 30 seconds (-1°C per cycle), 72°C for 30 seconds)

c. Amplification: 25 x (95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds)

d. Elongation: 72°C for 5 minutes

e. Storage: store at 10°C until samples retrieved from machine

The lid was heated to 110°C throughout to reduce the temperature gradient within the PCR tubes and minimise evaporation of samples.

During the PCR reaction, an agarose gel was made as per the steps above and placed into the Electrophoresis chamber. 1 x Tris Acetate EDTA solution was added to the chamber to the fill line.

3µl of 100bp DNA Ladder (catalogue number 15628019, Invitrogen, Paisley, UK) was loaded into the chosen well. 4µl of each PCR product was combined with 1µl DNA loading buffer on a clean section of parafilm, mixed by pipette and loaded into the wells of the agarose gel. Care was taken to use different areas of parafilm for each sample to avoid cross contaminating loaded samples, and to clearly document the loading order of the samples.

Electrophoresis ran at 110V for 90 mins. Bubbles were immediately visible within the chamber, demonstrating flow of current. After the 90-minute period, the gel was removed from the tank if the visible dye had run to within 1cm of the end of agarose gel. If not, additional running time was used. Once complete, the gel was visualised at 600nm wavelength using the Odyssey® Fc Imaging System by LI-COR Biosciences.

2.5.5 Sanger sequencing analysis

2.5.5.1 Sending samples for Sanger sequencing

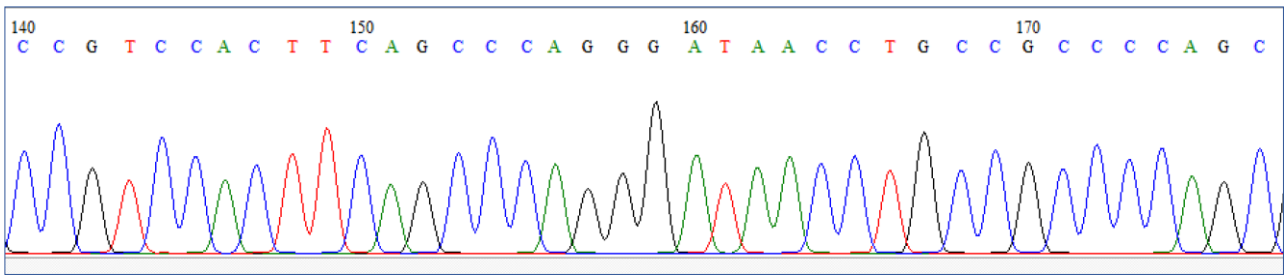
If gel electrophoresis demonstrated a single, clean band of PCR product, samples were sent for Sanger sequencing analysis. Concentration of each PCR product was determined using the Nanodrop spectrophotometer. Sample were diluted using RNA and DNA free water to a concentration of 10-50ng/µl and 20µl of product sent to Eurofins sequencing service. (https://www.eurofinsgenomics.eu/)

2.5.5.2 Interpreting sequencing results

For each specified combination of primer and PCR product, Eurofins generated an ABI file which can be visualised using Bioedit Sequence Alignment Editor (20). BioEdit is a free biological sequence alignment editor designed to make alignment and manipulation of sequences relatively easy on

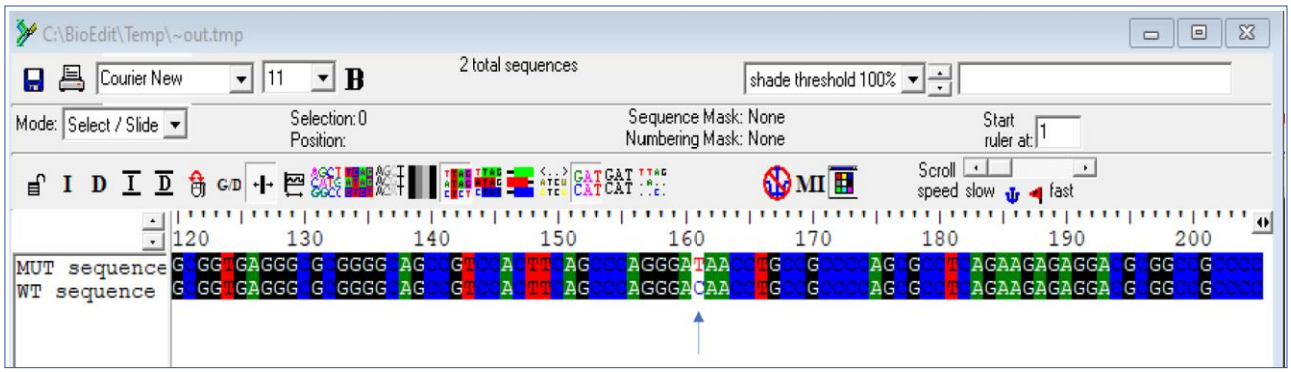
desktop computers. Example images generated from the programme are shown (**Figures 2.4 and 2.5**).

Figure 2.4 Chromatogram of Sanger sequencing generated from ABI file on Bioedit



Sanger sequencing showing a classic four-colour chromatogram. Numbering is relative to the nucleotides of the PCR product.

Figure 2.5 Alignment of wild-type and mutant sequences on Bioedit identifying a homozygous C to T change



MUT, Mutant; WT, wild-type. Using the alignment tool and colour coding the bases of the sequence enables easy identification of base mismatches. A homozygous C to T change is indicated by the blue arrow.

2.6 Cell culture

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium High Glucose (Catalogue number D6429, Sigma Aldrich, Poole, UK) supplemented with 10% fetal bovine serum (FBS; catalogue number 11573397, Fisher Scientific, UK) and 1% penicillin/streptomycin solution (Pen/Strep; catalogue number 15140122, ThermoFisher Scientific, UK). Fibroblast cells were

cultured in Dulbecco's Modified Eagle Medium High Glucose supplemented with 20% fetal bovine serum, 1% penicillin/streptomycin solution and 1% Non-essential Amino Acids (NEAA; catalogue number 11140050, ThermoFisher Scientific, UK). All cells were incubated in a humidified incubator at 37°C and 5% CO_2. Cells were assessed for mycoplasma contamination on monthly basis using the MycoAlert™ Mycoplasma Detection Kit (catalogue number LT07-318, Lonza Bioscience) and were contamination-free.

Cells were passaged and/or frozen once cultured to 90% confluency in a 75cm^2 (T75) flask. They were gently washed with 10ml Phosphate Buffered Saline solution (PBS; catalogue number 806552, Sigma Aldrich, Poole, UK) prior to incubation at 37°C and 5% CO_2 with 3ml 1x Trypsin/Ethylenediaminetetraacetic acid solution (Trypsin EDTA; R001100, ThermoFisher Scientific, UK) until detachment of cells from the flask was visible under the microscope. 7ml media was then added to inactivate the Trypsin EDTA and this solution was centrifuged at 1200rpm (fibroblasts) or 1500rpm (HEK293T) for 5 minutes at room temperature to pellet the cells. Following centrifugation, the supernatant was discarded.

When splitting the cells, this pellet was then resuspended with the relevant cell media, divided between several tissue culture flasks and incubated at 37°C and 5% CO_2. When freezing the cells, the pellet was resuspended in 4ml 90% FBS/10% Dimethylsulfoxide (DMSO, catalogue number C6164, Sigma Aldrich, Poole, UK) solution and frozen in 1ml aliquots in cryovials. These were stored at -80°C for short-term storage or placed in liquid nitrogen storage for long-term storage. Frozen samples were thawed by placing in the 37°C water bath for 1-2 minutes. 5ml cell media was added to each sample and centrifuged at 1000rpm for 5 minutes. The supernatant was discarded and the cell pellet resuspended in cell media and placed into tissue culture flasks.

2.6.1 Preparation of Luria-Bertani (LB) broth

10 g LB powder was dissolved in 400mls MilliQ water and sterilised using an autoclave. The broth was stored at room temperature with a tight seal when not in use. The LB broth was prepared in proximity to a flame to help maintain the sterility of the broth.

2.6.2 Preparation of LB agar plates

10g LB powder and 7.5g agar were dissolved in 400ml MilliQ water and sterilised using an autoclave. The solution was cooled by placing into a 37°C water bath. Ampicillin was added to final concentration 100micrograms/ml, and the agar mixture was poured into 10cm sterile dishes in the presence of an open flame to reduce the risk of contamination. Once set, the plates were sealed using parafilm and stored at 4°C until use.

2.6.3 Transformation of chemically competent bacterial cells

NEB 5-alpha competent *E. coli* (High Efficiency) (New England Biolabs, Hertfordshire, UK) were used for high efficiency transformation. Competent cells were stored at -80°C and thawed on ice for 10 minutes prior to use. 1µl of the plasmid DNA was added to the tube, which was then incubated on ice for 30 minutes. Samples were heat shocked for 30 seconds at 42°C then placed immediately back onto ice for 5 minutes. 950µl LB broth was then added to each vial and placed at 37°C in a shaking incubator set to 220rpm for 1 hour. Each sample was then centrifuged at 13000g for 1 minute at room temperate. 500µl of the supernatant was discarded and the remainder used to resuspend the pellet by pipetting. In the presence of an open flame to prevent contamination, 100µl of this solution was then plated onto a LB agar ampicillin plate and spread with a sterilised glass Pasteur pipette. Plates were incubated at 37°C overnight then colonies selected for culture the following morning.

2.6.4 Glycerol stocks

Prior to lysing cells in miniprep or midiprep methods to obtain the DNA construct, it is always advisable to make glycerol stocks of the living cells. This means the *E. coli* cells containing the DNA construct of interest can be easily amplified without the need for repeat transformation and colony selection. To create the glycerol stock, 200µl glycerol and 800 µl bacterial culture were combined in an autoclaved Eppendorf tube and mixed well. The stock was stored at -80°C until required. To amplify the bacterial culture from the glycerol stock, 200µl of the defrosted stock was added to 200ml autoclaved LB broth containing 100micrograms/ml ampicillin and left in a shaking incubator at 37°C overnight.

2.6.5 Miniprep

Miniprep of DNA constructs was performed using the QIAprep Spin Miniprep Kit (catalogue number 27106, Qiagen, Crawley, UK). Composition of buffers and reagents from this kit can be found in **Appendix 1.** Using a sterile tip, a single colony was picked from a freshly streaked (overnight incubation) LB agar ampicillin selection plate and used to inoculate a starter culture of 10ml LB broth containing 100micrograms/ml ampicillin. This step was performed in the presence of an open flame to prevent contamination. This was incubated overnight at 37°C in a shaking incubator set to 220rpm.

The following morning, glycerol stocks were created and the bacterial cells harvested by centrifuging the culture at 4000g for 10 minutes at room temperature. The supernatant was discarded and the pelleted bacterial cells resuspended in 250µl Buffer P1. The sample was placed into an autoclaved Eppendorf tube and 250µl Buffer P2 was added and mixed by inversion of the tube 4-6 times until the solution turned blue (due to the presence of Lyse Blue). 350µl Buffer N3 was added and mixed by inversion 4-6 times. The solution at this point became colourless and was then centrifuged for 10 minutes at room temperature at 17000g. 800µl of the supernatant was added to a QIAprep 2.0

spin column. The sample was centrifuged for 1 minute at 17000g and the flow through discarded. The QIA prep 2.0 spin column was washed by adding 0.5ml Buffer PB. The sample was centrifuged for 1 minute at 17000g and the flow through discarded. 750 µl Buffer PE was then added to the QIAprep 2.0 spin column. The sample was centrifuged for 1 minute at 17000g and the flow through discarded. The sample was then centrifuged for an additional 1 minute at 17000g to remove residual wash buffer and discard the flow through. The QIA prep 2.0 column was then placed into an autoclaved 1.5ml Eppendorf tube. To elute the DNA, 30 µl Elution Buffer EB (10mM TrisCl, pH8.5) was added to the centre of the QIAprep 2.0 spin column. After 1 minute incubation at room temperature, the sample was then centrifuged at 17000g for 1 minute. The flow through was reapplied to the centre of the QIAprep 2.0 spin column and the incubation and centrifugation repeated. This step is taken to maximise DNA elution from the column and thus concentration of final product. Finally, the quality and concentration of the eluted miniprepped sample was assessed using a Nanodrop spectrophotometer.

2.6.6 Midiprep

Midiprep of DNA was performed using the HiSpeed Plasmid Midi Kit (catalogue number 12643, Qiagen, Crawley, UK). Composition of buffers and reagents from this kit can be found in **Appendix 1.** Using a sterile tip, a single colony was picked from a freshly streaked (overnight incubation) LB agar ampicillin selection plate and used to inoculate a starter culture of 10ml LB broth containing 100micrograms/ml ampicillin. This step was performed in the presence of an open flame to prevent contamination. This was incubated for 8 hours at 37°C in a shaking incubator set to 220rpm. Following this incubation, 200µl of this bacterial culture was added to 200ml LB broth containing 100micrograms/ml ampicillin and this was incubated overnight at 37°C in a shaking incubator set to 220rpm.

The following morning, glycerol stocks were created and the bacterial cells harvested by centrifuging the culture at 4000g for 15 minutes at room temperature. Supernatant was discarded and the bacterial pellet resuspended by adding 6ml of chilled Buffer P1. The sample was vortexed to ensure the bacterial pellet is well dissolved. 6ml buffer P2 was added and the solution was mixed by inversion 4-6 times, turning blue due to the presence of Lyse Blue. The solution was incubated at room temperature for 5 minutes. 6ml chilled Buffer P3 was added and mixed thoroughly by inverting 4-6 times, at which stage the solution became colourless.

This solution was then poured onto a QIAfilter cartridge and incubated at room temperature for 10 minutes. During this time, 4ml buffer QBT was added to a Hispeed tip to calibrate the tip. The solution was then filtered through the QIAfilter cartridge and onto the HiSpeed tip. Flow through from the HiSpeed tip was discarded. 20ml Buffer QC was added to the HiSpeed Tip and flow through discarded. The HiSpeed tip was then transferred to a new collection tube. 5mls Buffer QF was added to elute the DNA. DNA was precipitated by adding 3.5ml isopropanol directly to the eluted DNA. After incubating the sample at room temperature for 5 minutes, the solution entered a QIAprecipitator Module. This module was washed with 2ml 70% ethanol and then air was pushed into the module to dry the membrane and remove the ethanol. DNA captured within the QIAprecipitator Module was then eluted by adding 300µl Buffer TE. The flow through was again passed through the QIAprecipitator Module to optimise DNA elution. The quality and concentration of eluted sample was then assessed using a Nanodrop spectrophotometer.

2.6.7 Transient transfection with Lipofectamine 2000

All steps were performed in a tissue culture hood under aseptic conditions. HEK293T cells were seeded in a 6-well plate to achieve 70% confluency after 24 hours incubation at 37°C and 5% CO_2. Opti-MEM (Catalogue number 31985062, Gibco, UK) and HEK293T cell culture media (DMEM with

10% Fetal Bovine Serum and 1% Pen/Strep solution) were warmed to 37°C using a water bath. 1µg of plasmid DNA was added to 150 µl Opti-MEM in an autoclaved Eppendorf tube. In a separate tube, 6µl Lipofectamine 2000 reagent (Catalogue number 11668027, Invitrogen, Paisley, UK) was added to 150µl Opti-MEM. The tubes were incubated for 5 minutes at room temperature. The lipofectamine containing solution was then added to the DNA solution and mixed by flicking the tube. The tubes were left at room temperature for 20 minutes under the hood. After 20 minutes, the 6-well plate of HEK293T cells was removed from the incubator and the media gently aspirated. The DNA/lipofectamine was gently added in a dropwise manner to the appropriately labelled well. After incubating for 2 minutes at room temperature, 2ml HEK293T cell culture media was gently added to each well and the cells returned to the incubator at 37°C and 5% CO_2. After 18 hours the cells were lysed.

2.6.8 Obtaining whole cell lysate

RIPA lysis buffer system (Santa Cruz Biotechnology; RIPA buffer with added protease inhibitor cocktail, sodium orthovanadate and phenylmethanesulfonyl fluoride) was used for whole cell lysis. Cells were removed from the incubator and placed on ice. Media was removed and 1ml cold PBS applied to each well, taking care not to detach the cells. After 2-3 minutes, this PBS was removed and the PBS wash repeated. All PBS was then removed from the wells. 100µl RIPA buffer was applied per well and left at room temperature for 2-3 minutes. A cell scraper was used to detach the cells and they were aspirated into autoclaved Eppendorf tubes. The tubes were placed on a rotary mixer at 4°C for 30 minutes. They were then sonicated for 5 seconds at 20KHz to break up cell membranes and fragment the DNA. Following sonication, the samples were placed immediately on ice for 5 minutes then the process was repeated once more. They were then placed into a pre-chilled 4°C centrifuge for 15 minutes at 15,000g. Supernatant was aspirated and placed in new Eppendorf tubes

and the pellet discarded. Protein concentration of whole cell lysates was then assessed using BCA assay.

2.6.9 BCA (Bicinchoninic acid) Protein assay

BCA assay was performed using the Pierce™ BCA Protein Assay Kit (catalogue number 23225, ThermoFisher Scientific, UK). Diluted Albumin (BSA) Protein standards A-I were made as per kit instructions (ranging in concentration from 0µg/mL to 2000µg/mL). Whole cell lysate samples were diluted 1:10 by adding 2µl lysate to 18µl RIPA lysis buffer system. 10µL of BCA protein standards A-I and 10µL of the diluted 1:10 whole cell lysate samples were loaded in duplicate to the 96 well plate. 10mls Reagent A and 200µl Reagent B were combined (1:50 dilution) and 200µl of this solution was added to each well. The plate was gently mixed using an Ika-Vibrax-VXR vibrating plate, then covered in foil and placed into an incubator at 37°C for 30 mins. After this time, the plate was removed from the incubator and assessed on a plate reader at 570 nm wavelength. The values for the whole cell lysate samples were interpreted in the context of the values of the protein standards using a protein assay standard curve.

2.7 Western blotting

2.7.1 Western blot reagents

20x MOPS Running Buffer

I dissolved 104.6g MOPS, 60.6g Trisbase, 10g SDS and 3g EDTA in 300mls MilliQ distilled water, then added MilliQ water to make up 500mls total volume. I made sure to cover the glass beaker in foil as the solution is photosensitive. When I required 1xMOPS buffer, I added 50ml of this solution to 950ml to MilliQ water and mixed well.

Transfer Buffer

I dissolved 9g glycine and 2.4g Tris Base in 800ml MilliQ distilled water. I then added 200ml methanol and mixed well.

Phosphate buffered saline (PBS)

I dissolved 10 PBS tablets in 1000ml MilliQ water.

PBS Tween 0.1%

I added 1ml Tween 20 to 999ml PBS and mixed well.

Blocking buffer (5% milk/PBS Tween 0.1%)

I added 2.5g milk powder to 50ml PBS tween 0.1% and mixed by inversion.

2.7.2 Western blot procedure

Desired volumes of 4x Laemmli sample buffer, RIPA buffer and whole cell lysate were combined according to BCA concentrations and desired protein loading per lane. Samples were heated to 95°C for 5 minutes on a heat block to denature the proteins and then placed immediately onto ice. Samples were then centrifuged for 30 seconds at 10,000rpm and pipetted to ensure well mixed. Two 10 well NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.5 mm thickness, were secured in the electrophoresis chamber and MOPS running buffer poured between the gels. After ensuring no leakage of buffer, MOPS was then poured into the main chamber and samples loaded onto the gel. 6µl Novex® Sharp Pre-stained Protein Standard ladder was loaded into the first well and the loading order of the samples was well documented. The electrophoresis was run at 110V for 30 minutes until visible dye had formed a flat straight line. The voltage was then increased to 130V and run until the visible dye had reached the bottom of the gel. Once electrophoresis is complete, the gels were

carefully removed from their plastic casing and any sections of the gel no longer required were removed. The proteins on the gel were then transferred onto a nitrocellulose membrane using a transfer machine set at 15v for 45 minutes (for 2 gels).

The membrane was soaked in transfer buffer for 5 minutes prior to transfer. Quality of protein transfer was assessed using Ponceau red stain, which was subsequently removed with PBS washing. Membranes were then blocked for one hour at room temperature in 10ml blocking buffer. After this, the membranes were incubated at 4°C for 2 days with primary antibody 1 in 1000 dilution in 5ml 5% milk/PBS Tween 0.1% and 1 in 100 sodium azide as a bacteriostatic preservative. A comprehensive list of antibodies used and their applications is listed in **Appendix 3**. The membranes were then washed three times with 10 ml PBS Tween 0.1% for 5 minutes each. Secondary antibodies were added to the membrane at 1 in 10,000 dilution in 5ml blocking buffer and incubated at room temperature for 1 hour protected from light. The membranes were then washed with PBS-Tween 0.1% for 5 minutes and PBS for 2 washes of 5 minutes each at room temperature. Plain PBS was used for these final washes as Tween can interfere with the image generated on the LiCor machine. The membrane was then imaged on Odyssey® Fc Imaging System by LI-COR Biosciences using 700nm and/or 800nm waveforms (depending on the antibodies used).

2.7.3 Modifications to western blot procedure to aid detection of small proteins (<30KDa)

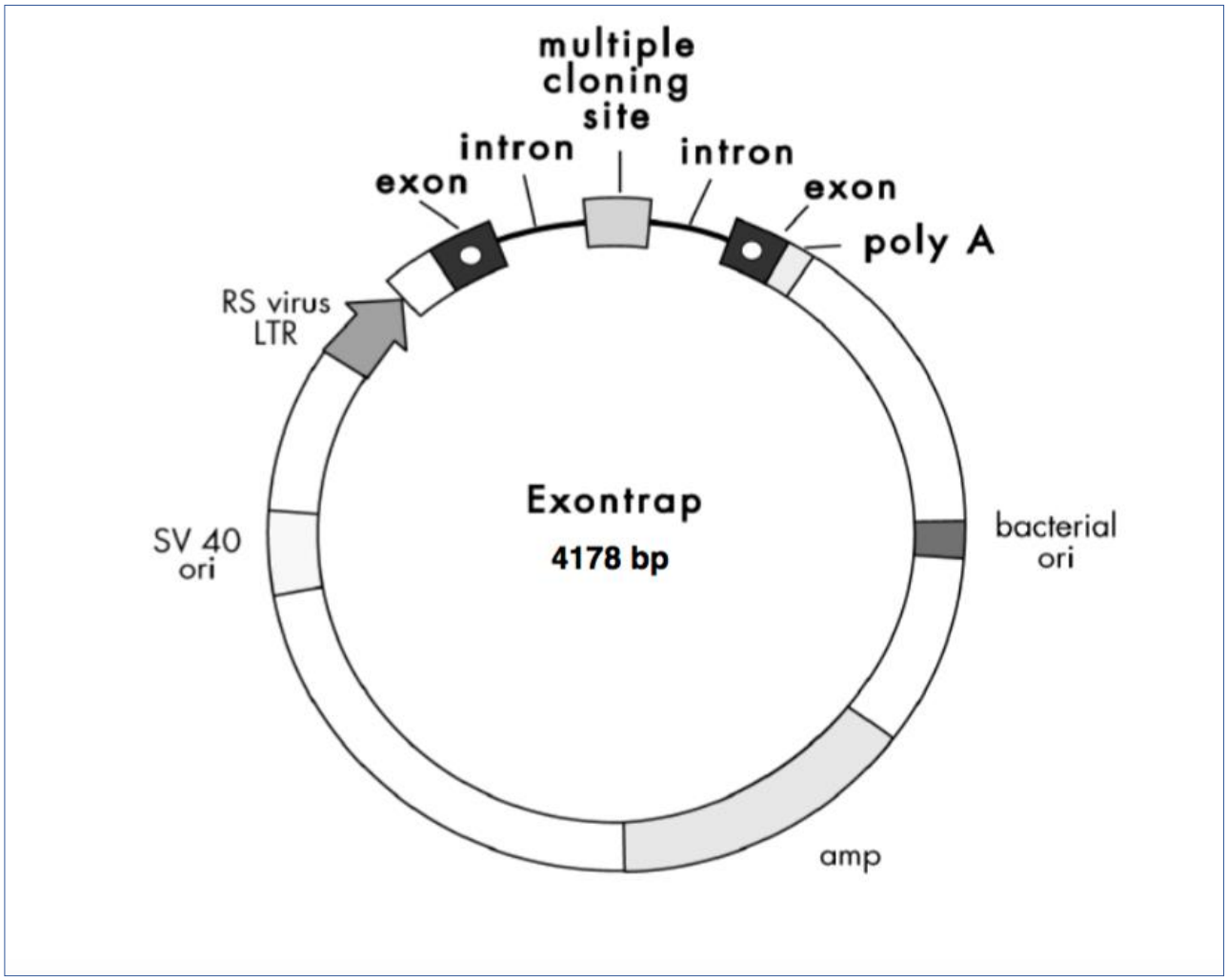
Modifications were made to the above protocol to improve separation and detection of small proteins. The samples were combined with 4 x non-reducing, fluorescent compatible sample buffer (catalogue number LC2570, ThermoFisher Scientific, UK) with added beta mercaptoethanol. Once boiled, the samples were loaded onto 16% Tricine gels (catalogue number 12040116 (10 well) and 12050116 (12 well), ThermoFisher Scientific, UK) with 6µl of Spectra™ Multicolor Low Range Protein

Ladder (catalogue number 26628, ThermoFisher Scientific, UK). The electrophoresis was run at 80V for 3-4 hours total in 1 x Tricine SDS running buffer (catalogue number 12020146 ThermoFisher Scientific, UK). 0.2µm nitrocellulose membranes (catalogue number LC2000, Invitrogen, Paisley, UK) were used for protein transfer on the Trans-Blot Turbo system (Bio-Rad, UK) for 30 minutes at 15V.

2.8 *In vitro* splicing assay

MoBiTec-Exontrap Cloning Vector pET01 (Catalogue reference PET01, MoBiTec, Germany) shown in **Figure 2.6** was utilised for cloning of a DNA fragment of interest into its multiple cloning site using the restriction enzyme *XbaI*.

Figure 2.6 Vector map for MoBiTec-Exontrap Cloning Vector pET01 (MoBiTec GmbH, Germany)



The vector contains an intronic sequence interrupted by a multiple cloning site. The DNA fragment of interest was cloned into this polylinker. If the DNA was incorporated in the correct orientation, the vector was then transfected into HEK293 cells. RNA was generated from these cells and reverse transcription of mRNA was then achieved with subsequent DNA amplification**.**

The DNA fragment of interest was amplified using standard touchdown PCR protocol and specifically designed primers containing the restriction enzyme sequence for XbaI (AGCTATATCTAGA) at the start of both the forward and reverse primer. Primer sequences used for the *in vitro* splicing assays are provided in **Appendix 4.** 5 PCR samples were run for both the patient and a control (25 µl volume per sample). Samples were assessed by electrophoresis to check the PCR had amplified products of the intended length.

The amplified product was column purified using the QIAquick® Gel Extraction Kit (catalogue number 28704, Qiagen, Crawley, UK). Composition of reagents from this kit can be found in **Appendix 1.** 100µl of product for either patient or control PCR reactions was combined with 500µl Buffer QG, placed into a column and centrifuged at 17000g for 1 minute at room temperature. The flow through was discarded. 750µl Buffer PE was added and the sample was centrifuged at 17000g for 1 minute at room temperature. The flow through was discarded and the sample centrifuged again at the same settings. The column was then transferred to an autoclaved Eppendorf tube and 30µl elution buffer added. The sample was incubated at room temperature for 2 minutes, then centrifuged at 17000g for 1 minute. Quality and concentration of the purified samples was assessed using a NanoDrop spectrophotometer.

The products and the vector were then digested with *XBal* for 4 hours at 37°C in the ratios shown below:

For the insert:

Insert (PCR product)	30µl
Bovine Serum Albumin (BSA)	0.5µl
Xbal	2.0µl
XBal buffer	4.0µl
DEPC treated water, DNA and RNA free	3.5µl
Total volume	40µl

For the vector:

Insert (PCR product)	500ng
Bovine Serum Albumin (BSA)	0.3µl
Xbal	1.0µl
Xbal buffer	3.0µl
DEPC treated water, DNA and RNA free	To make up to 30µl total volume

During the last 30 minutes of this 4-hour incubation at 37°C, alkaline phosphatase (ALP) treatment was applied to the vector sample only. This enzyme catalyses dephosphorylation, which prevents re-ligation of linearized plasmid DNA. This was added in the following ratios:

Vector digestion volume	30µl
ALP	3µl
ALP buffer	4µl
DEPC treated water, DNA and RNA free	3µl
Total volume	40µl

Samples then underwent repeat purification as described above. As samples were 40µl volume, 200µl Buffer QG was added (5x volume). The same volumes were used for Buffer PE and elution buffer. Following repeat purification, the insert and vector were incubated together with T4 DNA ligase at 16°C for 1 hour to anneal the constructs together.

Amount insert required = $\frac{\text{50 nanogram of vector}}{\text{Size of vector}} \times \text{size of insert (bp)} \times 2$

Vector	50 nanograms
Insert	as calculated above
T4 DNA ligase	same volume as insert + vector
DEPC treated water, DNA and RNA free	to make up to 15µl total volume

Transformation using NEB 5-alpha competent *E. coli* (High Efficiency) (New England Biolabs, Hertfordshire, UK) was performed as described previously. All of the 15µl reaction mixture was added to one *E. coli* vial rather than the 1µl described in the methods previously.

The following day, colonies were selected from each plate. Part of each colony was added to a PCR mixture to determine if ligation had been successful and if the colony contained the vector plus insert. The remainder of the colony was placed into 10ml of LB broth with 100micrograms/ml ampicillin and placed at 37°C in a shaking incubator at 220rpm overnight.

Specific primers for the pET01 vector were used:

ET Primer 6 (Forward): GCGAAGTGGAGGATCCACAAG

ET Primer 7 (Reverse): ACCCGGATCCAGTTGTGCCA

The following PCR programme was used:

a. Touchdown: 30 x (95°C for 20 seconds, 75°C for 20 seconds (-0.7°C per cycle), 72°C for 30 seconds)

b. Amplification: 10 x (95°C for 20 seconds, 60°C for 20 seconds, 72°C for 30 seconds)

c. Storage: store at 10°C until samples retrieved from machine

The lid was heated to 110°C throughout to reduce the temperature gradient within the PCR tubes and minimise evaporation of samples.

Assessment of the size of PCR product on gel electrophoresis enabled selection of colonies that appeared the correct size for the expected vector plus insert sequence. PCR products with the correct size were then sent for Sanger sequencing to verify the sequence and orientation of the insert.

For colonies that were confirmed to contain the insert and vector in correct orientation, the corresponding bacterial culture was miniprepped and the constructs transfected into HEK293T cells using Lipofectamine 2000 as described previously. 18 hours post transfection, cells were lysed and RNA extraction and cDNA synthesis performed.

RNA extraction was performed using the RNeasy Plus Mini kit (catalogue number 74134, Qiagen, Crawley, UK). Composition of buffers and reagents from this kit can be found in **Appendix 1.** Cell media was removed and 1 ml cold PBS was applied to each well. This was removed and 350µl RLT buffer was added to the cells, and cell scrapers were used to detach cells from the wells. The contents of the wells were transferred into RNAse free Eppendorf tubes. 350 µl of 70% ethanol was added to each sample. The samples were then transferred into RNeasy spin columns and centrifuged at 17000g for 1 minute. Flow through was discarded. 350 µl of Buffer RW1 was added to the columns. The tubes were again centrifuged at 17000g for 1 minute and flow through discarded. 10 µl DNase1 and 70 µl Buffer RDD was added to each sample and the samples incubated at room temperature for 15 minutes. 350 µl of Buffer RW1 was added to the columns. The tubes were centrifuged at 17000g for 1 minute and flow through was discarded. 500µl of Buffer RPE was added to the columns and the tubes were centrifuged at 17000g for 1 minute. Flow through was

discarded. A further 500µl of Buffer RPE was added to the columns and the tubes were centrifuged at 17000g for 1 minute. Flow through was discarded. The samples were then centrifuged at 17000g for 2 minutes. The columns were transferred to autoclaved Eppendorf tubes 30 µl of RNAse free water was added to each column and the samples were left to stand for 3 minutes. The samples were then centrifuged at 17000g for 1 minute. The eluted RNA was quantified using a NanoDrop spectrophotometer and stored at -80°C until cDNA synthesis was performed.

cDNA synthesis was performed using the High-Capacity RNA-to-cDNA^{TM} Kit (ThermoFisher Scientific, UK).

For each RNA sample, a reaction mixture of the following reagents was combined:

RNA	1µg
10x RT Buffer	2.0µl
dNTP	0.8µl
Random primer	2.0 µl
Reverse transcriptase	1.0 µl
DEPC treated water, DNA and RNA free	to make up to 20µl total volume

The reactions were then incubated sequentially at the following temperatures:

25°C for 10 minutes

37°C for 120 minutes

85°C for 5 minutes

PCR was performed using 2µl cDNA and ET Primer 2 (Forward) GAGGGATCCGCTTCCTGGCCC and ET Primer 3 (Reverse) CTCCCGGGCCACCTCCAGTGCC and the programme described for colony PCR.

These primers amplified the region of interest starting from the two exons of the exon trap vector and including any retained splicing product between these two exons. Electrophoresis was performed using a 2% agarose gel (2g agarose in 100ml 1x Tris Acetate EDTA buffer) and the fragment sizes were assessed. PCR products were then sent for Sanger sequencing to confirm the sequences.

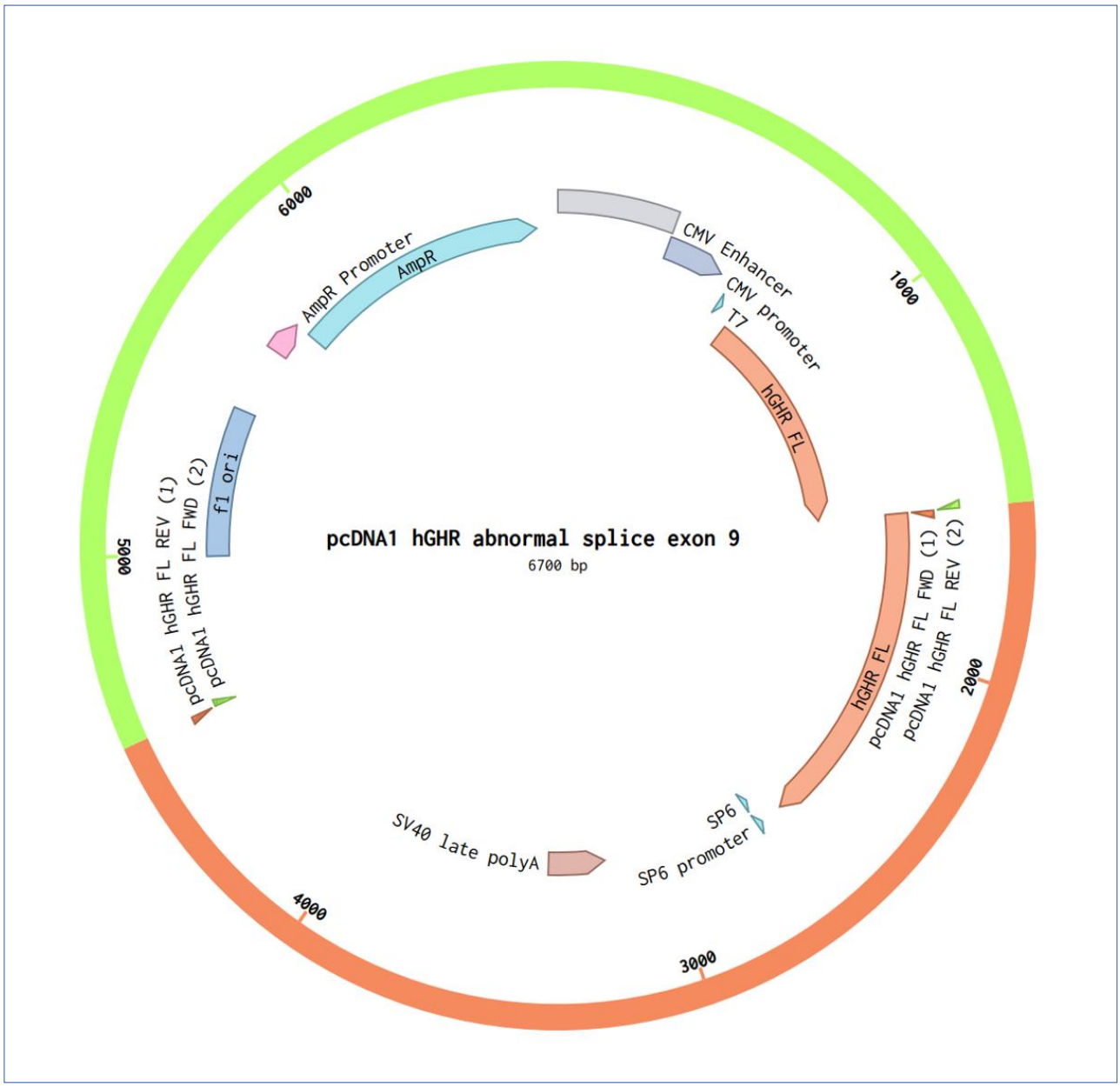
2.8 Creation of custom *GHR* vectors using Gibson assembly

Gibson assembly is an effective technique for assembling multiple overlapping DNA molecules using a 5' exonuclease to generate single stranded-DNA overhangs that specifically anneal (203). Using this technique enabled me to recreate the *GHR* sequence seen in both patients with suspected dominant negative *GHR* variants. It also enabled me to clone in the novel pseudoexon sequence to the *GHR* vector to enable functional assessment of the mutant protein.

2.9.1 Gibson vector and primer design

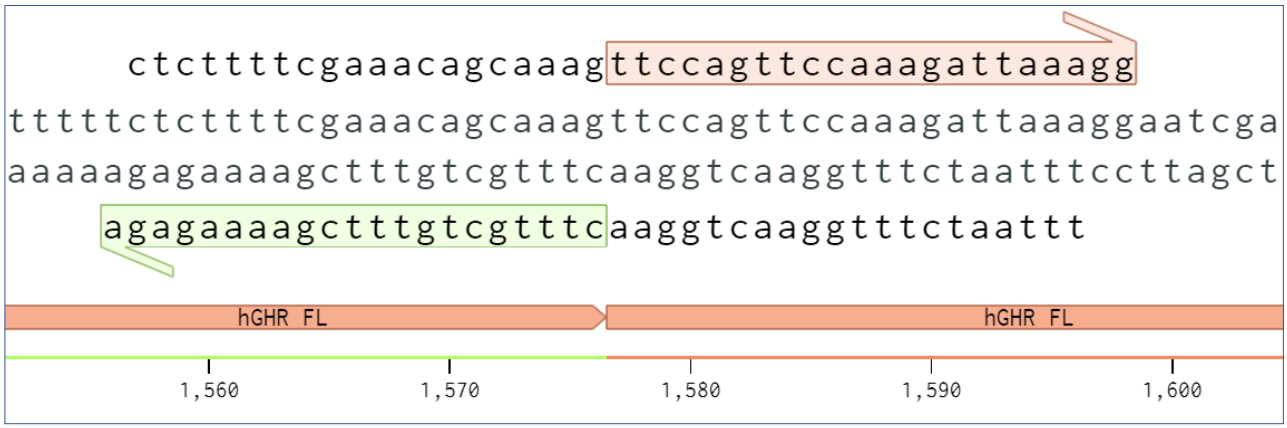
Primers were designed using Benchling assembly wizard for Gibson assembly (Benchling Biology Software 2020, https://benchling.com). The designed vector product and primers for deleting 26bp of the *GHR* to mimic the mutant exon 9 seen in Patient 7 are shown in **Figures 2.7 and 2.8**. The primer sequences used for all Gibson assemblies are provided in **Appendix 4.**

Figure 2.7 Vector map generated on Benchling including location of designed primers to generate mutant *GHR* vector for Patient 7



Custom primers designed to assemble green and orange segments of vector. These two fragments were then annealed using Gibson assembly. Vector map generated using Benchling assembly wizard for Gibson assembly (Benchling Biology Software 2020, https://benchling.com).

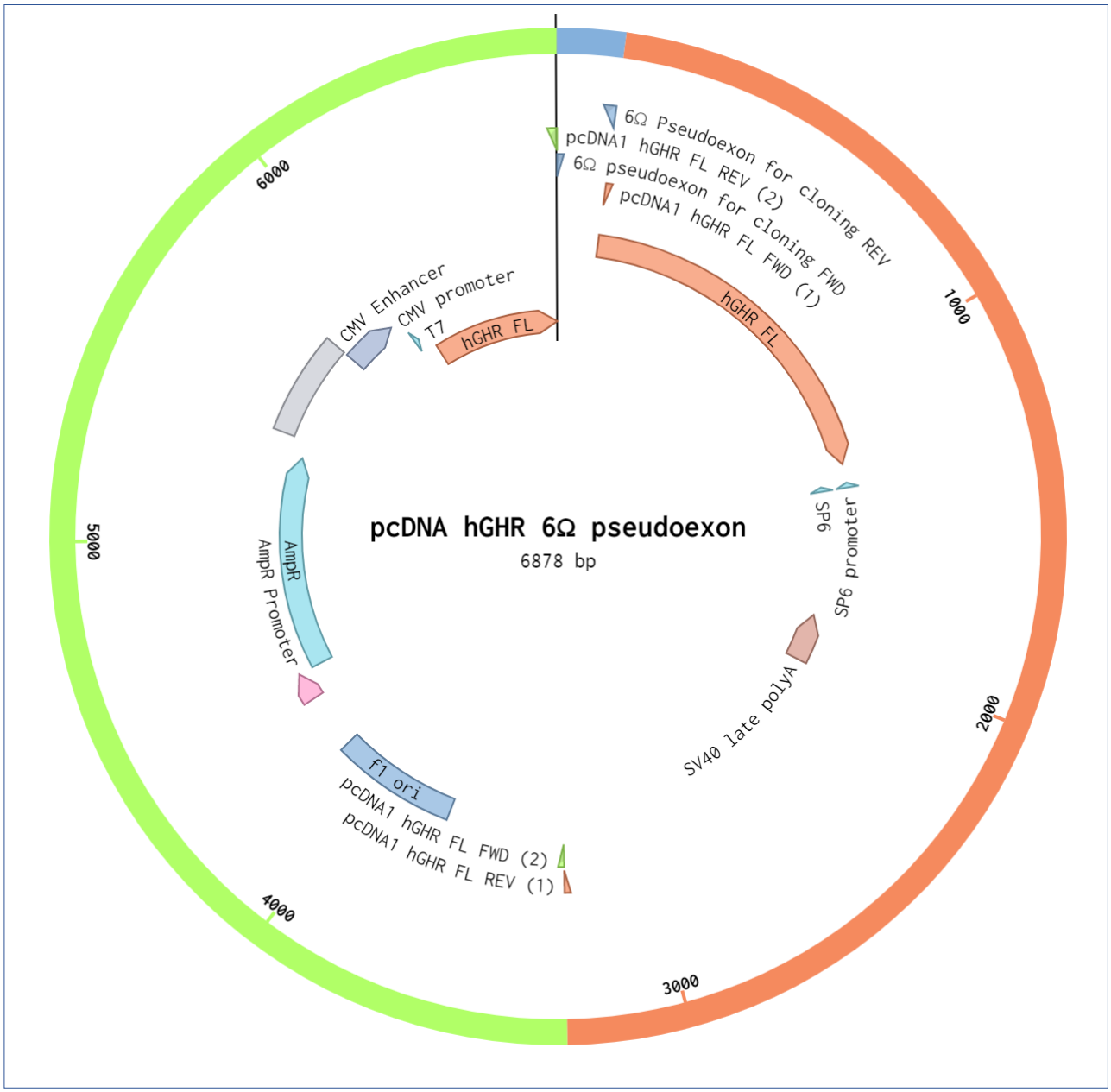
Figure 2.8 Primer sequences for deleting 26bp region in *GHR* to mimic Patient 7 transcript



Shown in green is the base sequence of the forward primer and orange is the sequence of the reverse primer to generate the 26-base deletion in the *GHR* vector.

To assemble a *GHR* vector containing the 6Ω pseudoexon, a 3-part Gibson assembly was required due to the length of the insert. Shown below in **Figure 2.9** is the vector assembly map.

Figure 2.9 Schematic of 6Ω pseudoexon *GHR* vector created by Gibson assembly



The 6Ω pseudoexon *GHR* vector was assembled from 3 amplified fragments, the outer ring (pseudoexon insert in blue) and the two other amplified fragments of the WT GHR vector (orange and green). The sets of primers used to amplify the 3 fragments (pcDNA1 hGHR FL FWD (1); pcDNA1 hGHR FL REV (1); pcDNA1 hGHR FL FWD (2); pcDNA1 hGHR FL REV (2); 6Ω pseudoexon for cloning FWD (3); 6Ω pseudoexon for cloning REV (3)) are indicated by the triangles of the corresponding colour. Image generated by Benchling (Benchling Biology Software 2020, https://benchling.com).

2.9.2 Assembly of mutant *GHR* vectors

Phusion® High-Fidelity PCR Kit (New England Biolabs, Hertfordshire, UK) was used to amplify the target sequences and minimise the risk of incorrect bases being incorporated into the sequence.

The reaction mixture was made up as follows:

5 x phusion Buffer	4µl
10mM dNTP mix	0.4µl
Forward primer (10mM concentration)	0.5µl
Reverse primer (10mM concentration)	0.5µl
Phusion DNA polymerase	0.2µl
DMSO	0.6µl
template DNA	25ng
(volume will depend on concentration of sample)	
DEPC treated water, DNA and RNA free	to 20 µl total volume

Negative control reactions were included, containing all above reagents except primers and all above reagents except DNA template. A positive control reaction was included using a 1.3KB Lambda template.

The samples were run on the following PCR programme:

a. Denaturing: 98°C for 3 minutes

b. Amplification cycles: 35 x (98°C for 30 seconds, 72°C for 2 minutes)

d. Elongation: 72°C for 10 minutes

e. Storage: store at 10°C until samples retrieved from machine

PCR products were visualised on gel electrophoresis to verify product size. They were then *Dpn1* treated to get rid of methylated DNA (i.e the original wild-type vector template). 1µl of *Dpn1* (concentration of 10units/µl) was added to each PCR tube and the sample incubated at 37°C for 3 hours. PCR product clean-up was then performed using the Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit.

Sample quality and concentration of the samples was assessed using a NanoDrop spectrophotometer. NEBiocalculator was used to calculate the volume needed from each product for optimum annealing (https://nebiocalculator.neb.com/#!/ligation). 0.02-0.5pmols of DNA are recommended when two fragments are to be annealed in Gibson assembly.

The two fragments were combined with 2 times the volume of NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, Hertfordshire, UK) and incubated at 50°C for 60 minutes to anneal the two fragments into a circular vector. This construct was then transformed in *E. coli* cells as previously described. Single colonies were selected for miniprep and DNA products verified by Sanger sequencing.

2.10 Assessing the effect of mutant GHR constructs on downstream pSTAT5 signalling

HEK293T cells were transfected with the constructs using Lipofectamine 2000 as described in **Section 2.6.7**. A total of 1µg of plasmid DNA was transfected into each well. Cells were transfected with the construct combinations in duplicate to allow for one growth hormone (GH) unstimulated well for every GH stimulated well. Post transfection, 2ml of serum free media (DMEM only) was applied to each well to serum starve the cells prior to GH stimulation. 24 hours post transfection, GH was added to DMEM to a final GH concentration of 100ng/ml. The media was then changed in

all wells – half of the wells had 2ml DMEM with GH added and the other half 2ml DMEM only. 20 minutes later, all cells were lysed and protein concentrations were determined by BCA assay as described in **Section 2.6.9**.

Western blotting was performed (detailed in **Section 2.7**) with 20µg protein loaded per lane. Protein transfer was performed for a longer duration due to the large size of GHR protein (>100KDa on gel). The proteins on the gel were then transferred onto a nitrocellulose membrane using a transfer machine set at 15V for 90 minutes (for 2 gels). Primary antibodies were used at a concentration of 1 in 1000: Rabbit STAT5B antibody, PA1841 (Booster Bio); Rabbit Polyclonal Phosho-pSTAT5 (Tyr694) antibody #9351 (Cell Signalling); Rabbit Polyclonal Anti-GHR Antibody, #TA334654 (Origene); Mouse Beta Actin Monoclonal Antibody, catalogue number: 66009-1-Ig (Protein Tech).

2.11 Site-directed mutagenesis

Site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit Instruction (catalogue number 200522, Agilent, UK). Composition of buffers and reagents from this kit can be found in **Appendix 1.**

2.11.1 Primer design

Primers were designed using the Quikchange Primer Design tool available at www.agilent.com/store/primerDesignProgram.jsp

The primers generated using this tool were then entered into the BLAST search tool on Ensembl (http://grch37.ensembl.org/Multi/Tools/Blast) to check if the oligonucleotide sequences occurred elsewhere in the genome (thus increasing the risk of non-specific PCR products). Primers that

occurred only once in the genome were preferentially ordered. All primers were diluted to 125ng/µl. Primer sequences are provided in **Appendix 4**.

2.11.2 PCR amplification

The concentration of the target vector was assessed using a NanoDrop spectrophotometer. The vector was then diluted to 10ng/µl.

Ratio of reagents for PCR mix

10 x Buffer	5 µl
dNTP mix (10mM of each dNTP)	1 µl
Forward primer (125ng/µl concentration)	1 µl
Reverse primer (125ng/µl concentration)	1 µl
Quikchange	3 µl
Vector	1 µl
DEPC treated water, DNA and RNA free	to make up 50 µl total volume

Total volume for each PCR reaction = 50 µl

Once the reactions had been mixed and were placed on ice, 1µl PfuUltra High-Fidelity DNA polymerase was added just prior to starting the PCR reaction. Total volume per reaction was then 51µl.

The PCR programme for this mutagenesis is as follows:

a. Denaturing: 95°C for 1 minute

b. Amplification: 12 x (95°C for 50 seconds, 60°C for 50 seconds, 68°C for 10 minutes*)

d. Elongation: 68°C for 7 minutes

e. Storage: store at 10°C until samples retrieved from machine

The lid was heated to 110°C throughout to reduce the temperature gradient within the PCR tubes and minimise evaporation of samples. *variable dependent on length of vector plus insert, one minute per kb extension time, so if the total size was 12Kb this would be increased to 12 minutes.

Upon completion of this PCR, the products were visualised on a 1% agarose gel to assess the success of the experiment at this stage.

2.11.3 *Dpn1* treatment

After confirming products of the correct size, the products were put onto ice to cool. 1µl DPNI was added to each reaction tube and mixed well by pipetting. The reactions were then stored at 37°C for 1 hour. In contrast to the newly synthesised SDM constructs, the original vector is methylated as it was grown in bacteria. DPNI's ability to recognise and cut methylated DNA enables removal of the original vector template.

2.11.4 Transformation with 'Pure Gold' ultra-competent *E. coli* cells

Following this incubation, the reactions were placed on ice. 45µl of 'Pure Gold' ultra-competent *E. coli* cells were added to newly labelled tubes on ice. 2µl of beta-mercaptoethanol mix (provided with the kit) were then added to each tube on ice for ten minutes, flicking gently every few minutes to combine. 2µl of PCR product was then added to each respective tube and the reactions were left for 30 minutes on ice. After this time, the reaction tubes were heat shocked at 42°C for 60 seconds then placed straight back onto ice. After 2 minutes on ice, 500µl of LB broth was added to each tube and the samples were placed into a shaking incubator for 1 hour at 37°C.

Following this incubation, the samples were centrifuged for 2 minutes at 2000rpm and 400µl of supernatant removed. The samples were then resuspended and 100µl of each sample was spread onto pre-warmed agar plates with ampicillin. These plates were incubated overnight at 37°C and the following day several colonies were selected for miniprep to amplify the created product. Samples were sent for Sanger sequencing to verify the success of the SDM and check for any off-target effects.

2.12 Electrophoretic Mobility Shift Assay (EMSA)

2.12.1 EMSA reagents and Polyacrylamide gel

10x TBE buffer

Combine 108g tris base, 55g boric acid and 7.44g EDTA in 800ml deionised water. Make up to 1 litre total volume with additional deionised water as required.

Custom Oligonucleotides

Fluorescent oligos labelled with Cy-5 were used in the EMSA with sense strand 5'Cy5-CGCTATAAGCGCTAATAACGC and antisense strand 5'Cy5-GCGTTATTAGCGCTTATAGCG. The oligo sequences were based on a previous publication by a group using EMSA to assess a non-synonymous *HMGA2* variant (c.83G>A; p.G28E) affecting height in Shetland ponies (204).

DNA Polyacrylamide gels

Having tried a variety of different concentration gels, I found that 5% Polyacrylamide gels were the most effective for my EMSA experiments.

The glass plates and spacers were cleaned thoroughly with deionized water and ethanol, then assembled in the gel caster once dry. For 5% Polyacrylamide gels, the gel solution was assembled in the following quantities:

2ml 30% Acrylamide (29:1)

8.8ml water

1.2ml 10 x TBE electrophoresis buffer

200µl Ammonium persulfate (10% w/v)

10µl TEMED (Tetramethylethylenediamine; Thermo Scientific Pierce)

The TEMED was added last as the acrylamide rapidly polymerizes following this addition. The solution was gently added to the cast and the appropriate comb inserted, taking care not to trap any bubbles in the gel. The gels were left to polymerise for 60 minutes at room temperature.

2.12.2 EMSA protocol

After thawing the custom oligonucleotides slowly on ice, I combined the two oligos in a 1:1 ratio and placed them in a 65°C water bath for 10 minutes. I removed from heat and allowed them to cool to room temperature, then left them at room temperature for 1 hour. I diluted the duplexed oligos 1:1000 using ice cold TE buffer (pH 8.8) and thawed the FLAG IP product and the other EMSA reagents on ice.

I prepared the EMSA master mix in following order:

7µl 5 X EMSA buffer (125mM HEPES, 25mM $MgCl_2$, 37.5% glycerol, 0.5mM EDTA)

1.75µl poly dI/dC (2µg/µl)

1.75µl BSA (10mg/ml)

1.75µl DTT (20mM)

1.75µl Bromophenol blue (0.1%)

7µl KCl (500mM)

2µg FLAG IP product

Double distilled water – to make total volume 35µl

I incubated each reaction on ice for 30 minutes. During the incubation, I ran the pre-cast 5% Acrylamide/TBE gel for 40 minutes at 160V in 1 x TBE running buffer. I then added 50fmol pre-duplexed oligo to each reaction and incubate the samples at room temperature for 20 minutes protected from light. After this incubation, I added 2 µl anti-FLAG antibody to each reactions and incubated the samples for an additional 10 minutes at room temperature protected from light. I then loaded the entire reaction volume per well, and ran the samples at room temperature at 110V, protected from light for 70-80 minutes. I scanned the gel using the LiCor machine (LiCor, Lincoln, NE) set to the 700 channel.

2.13 Immunoprecipitation using anti-FLAG® M2 beads

I used ANTI-FLAG® M2 Magnetic Beads (product number M8823, Sigma Aldrich, Poole, UK) for my FLAG immunoprecipitation. These are 4% agarose beads bound with the Anti-FLAG M2 (mouse monoclonal) antibody.

I firstly prepared the 3X FLAG® elution solution with 3X FLAG® peptide (product number F4799, Sigma Aldrich, Poole, UK). I dissolved the FLAG® peptide in 0.5M Tris-Hcl, pH 7.5, with 1M NaCl at a concentration of 25µg/µl and stored at -20°C until required. I then diluted these samples 5-fold with MilliQ water to prepare a 3x FLAG stock solution at 5µg/µl of 3 X FLAG peptide. For the elution, I added 3µl of this solution to 100µl TBS, giving a final concentration of 150ng/µl.

I thoroughly resuspended the ANTI-FLAG® M2 Magnetic Beads by gentle inversion, then aliquoted 50µl FLAG bead suspension per reaction with 5 times the volume of TBS and mixed thoroughly. I centrifuged at 4°C 10,000 rpm for one minute. I repeated this TBS wash 3 times. I then combined 50µl bead suspension with 100µg cell lysate (quantified by BCA) and left the samples on the rotary mixer at 4°C overnight. The following day, I centrifuged the samples at 8,000g for 30 seconds and gently removed supernatant without disturbing the bead pellet. I then wash the samples 3 times with 500µl TBS.

I then applied 100µl of 3 x FLAG® elution solution to each washed bead pellet. I rotary mixed the samples for 30 mins at 4°C, then centrifuged them at 4°C for 30 seconds at 6500g. I then transferred the supernatants (FLAG IP product) to new Eppendorf tubes and stored them at 2-8°C (short-term) or -20°C (long-term).

2.14 Co-immunoprecipitation (Co-IP)

Co-IP lysis buffer

50mM Tris, pH 8.0

150mM NaCl

1% Triton x 100

10% glycerol

1mM EDTA

2.5mM EGTA

1 Roche protein inhibitor tablet per 50ml lysis buffer

Co-immunoprecipitation was performed to compare the interaction of wild-type and mutant CHD1L and PARP1 proteins. HEK293T cells were seeded to 70% confluency in 10cm dishes. They were transfected with a total of 3µg DNA using Lipofectamine 2000 reagent (Invitrogen, Paisley, UK).

The following day, the cells were placed on ice and washed with 5ml pre-chilled PBS. 0.6ml Co-IP lysis buffer was added to each 10cm dish. The cells were scraped, collected into Eppendorf tubes and placed on ice for 30 minutes. The samples were vortexed every 5-10 minutes during this 30-minute incubation period. The tubes were then centrifuged at 15000g for 15 minutes at 4°C and the supernatant collected. This was quantified by BCA protein assay.

FLAG IP was performed as described previously. This captured FLAG tagged CHD1L constructs which were then released by competitive inhibition using 3 x FLAG peptide. The products were assessed by western blot.

2.15 Restriction enzyme digest

Following incubation of a vector of interest with the relevant restriction enzymes, it should be possible to visualise the digested product on electrophoresis and compare to undigested sample incubated without restriction enzymes. Upon visualisation of the samples, a 'drop out' of correct size band should be visible in the digested sample representing the vector insert. I performed this technique with my *HMGA2* vector which contained the coding *HMGA2* sequence with a pcDNA3.1 vector backbone. *Kpn1* (catalogue number R0142, New England Biolabs, USA) and *Bam*HI (catalogue number R0136S, New England Biolabs, USA) were selected as the restriction enzymes for digest of the *HMGA2* vector following review of the vector map showing possible restriction enzyme sites. These were incubated with NEBuffer r1.1 and NEBuffer r2.1 (catalogue number B7030S, New England Biolabs, USA) to optimise digestion conditions. Composition of these buffers can be found in **Appendix 1.**

Reagents were combined as follows:

Vector DNA	300ng
NEBuffer r1.1	2 µl
Kpn1	1 µl
NEBuffer r2.1	2 µl
*Bam*HI	1 µl
MilliQ water	to make up to 20 µl total volume

Samples were incubated at 37°C for 40 minutes. They were then combined in 4:1 ratio sample: loading dye and loaded onto a 1% agarose gel, run at 100V for 60 minutes then visualised.

2.16 Custom *IGF1* vectors

All custom *IGF1* vectors were created on a pcDNA3.1 vector backbone. I designed them and ordered them from Genscript Biotech (Oxford, UK) who synthesised them. On their arrival, I then midiprepped the samples and sent them for Sanger sequencing to verify the sequence.

For my immunocytochemistry experiments looking at the localisation of wild-type compared to mutant IGF-1 proteins, I designed vectors with a FLAG tag between the signal peptide and the mature IGF-1 sequence, and a HA tag after the E domain. These are both small tags that are well detected by immunofluorescence and unlikely to interfere with the normal function of the IGF-1 protein. I added the FLAG tag after the signal peptide to avoid it being cleaved off when the signal peptide was removed as IGF-1 enters the secretory pathway. The base and amino acid sequences of the wild-type and mutant *IGF1* vector inserts are shown below in **Figure 2.10** and **Figure 2.11** respectively.

Figure 2.10 Base and amino acid sequence of wild-type *IGF1* vector insert with FLAG and HA tags

Base sequence:

ATGGGAAAAATCAGCAGTCTTCCAACCCAATTATTTAAGTGCTGCTTTTGTGATTTCTTGAAG
GTGAAGATGCACACCATGTCCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTT
CACCAGCTCTGCCACGGCTGATTACAAGGATGACGACGATAAGGGACCGGAGACGCTCTG
CGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAGACAGGGGCTTTTATTTCAAC
AAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTCAGACAGGCATCGTGGATGAG
TGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGTATTGCGCACCCCTCAAGCCTG
CCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCGACATGCCCAAGACCCAGAAGG
AAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAAACAAGAACTACAGGATGTACAC
CCATACGATGTTCCAGATTACGCT

Amino acid sequence:

MGKISSLPTQLFKCCFCDFLKVKMHTMSSSHLFYLALCLLTFTSSATADYKDDDDKGPETLCGA
ELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA
RSVRAQRHTDMPKTQKEVHLKNASRGSAGNKNYRMYPYDVPDYA

Colour key:

IGF1 signal peptide, FLAG tag between signal peptide and mature *IGF1*, mature *IGF1*, location of mutation (first base/amino acid of E domain), *IGF1* E domain, HA tag after E domain

The different regions of *IGF1* are colour coded as explained in the colour key. Wild-type base and amino acid sequence is shown, with the location of the change in the mutant highlighted in red.

Figure 2.11 Base and amino acid sequence of mutant c.356G>A *IGF1* vector insert with FLAG and HA tags

Base sequence:

ATGGGAAAAATCAGCAGTCTTCCAACCCAATTATTTAAGTGCTGCTTTTGTGATTTCTTGAAG
GTGAAGATGCACACCATGTCCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTT
CACCAGCTCTGCCACGGCTGATTACAAGGATGACGACGATAAGGGACCGGAGACGCTCTG
CGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAGACAGGGGCTTTTATTTCAAC
AAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTCAGACAGGCATCGTGGATGAG
TGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGTATTGCGCACCCCTCAAGCCTG
CCAAGTCAGCTCACTCTGTCCGTGCCCAGCGCCACACCGACATGCCCAAGACCCAGAAGG
AAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAAACAAGAACTACAGGATGTACAC
CCATACGATGTTCCAGATTACGCT

Amino acid sequence:

MGKISSLPTQLFKCCFCDFLKVKMHTMSSSHLFYLALCLLTFTSSATADYKDDDDKGPETLCGA
ELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA
HSVRAQRHTDMPKTQKEVHLKNASRGSAGNKNYRMYPYDVPDYA

Colour key:

IGF1 signal peptide, FLAG tag between signal peptide and mature *IGF1*, mature *IGF1*, location of mutation (first base/amino acid of E domain), *IGF1* E domain, HA tag after E domain

The different regions of *IGF1* are colour coded as explained in the colour key. Mutant base and amino acid sequence is shown, with the location of the novel *IGF1* variant (102813333C>T, c.356G>A, p.R119H) highlighted in red.

In an attempt to isolate an IGF-1 substrate for my furin assay, I designed a Streptomycin II tagged IGF-1 vector which could hopefully be isolated using StrepTrap columns (Details of the StrepTrap purification process can be found in methods **Section 2.18**). I added the Streptomycin II tag at the start of the mature IGF-1 sequence and a His tag after the E domain. Designing the vector in this way enabled me to identify the cleaved E domain, mature IGF-1 and the pro-IGF-1. It is a high affinity tag that should bid well to the StrepTrap columns. His was chosen as it was a small tag and was not predicted to interfere with the purification process. I also designed the vector without a signal peptide. This should prevent the IGF-1 from being excreted from the cell as it is the signal peptide that facilitates the movement of IGF-1 to the endoplasmic reticulum and into the secretory pathway. This meant that I could harvest large amounts of protein from whole cell lysates as the IGF-1 should

remain inside the cell in the absence of the signal peptide. The base and amino acid sequences of the wild-type and mutant *IGF1* vector inserts are shown below in **Figure 2.12** and **Figure 2.13** respectively.

Figure 2.12 Base and amino acid sequence of wild-type *IGF1* vector insert with Strep II and His tags and no signal peptide

Base sequence:

TGGAGCCACCCACAGTTTGAGAAAGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGAT
GCTCTTCAGTTCGTGTGTGGAGACAGGGGCTTTTATTTCAACAAGCCCACAGGGTATGGCT
CCAGCAGTCGGAGGGCGCCTCAGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTG
ATCTAAGGAGGCTGGAGATGTATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGT
CCGTGCCCAGCGCCACACCGACATGCCCAAGACCCAGAAGGAAGTACATTTGAAGAACGC
AAGTAGAGGGAGTGCAGGAAACAAGAACTACAGGATGCATCACCATCATCATCAC

Amino acid sequence:

KLATMWSHPQFEKGPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFR
SCDLRRLEMYCAPLKPAKSARSVRAQRHTDMPKTQKEVHLKNASRGSAGNKNYRMHHHHHH

Colour key:

Strep II tag before mature *IGF1*, mature *IGF1*, location of mutation (first base/amino acid of E domain), *IGF1* E domain, His tag after E domain

The different regions of *IGF1* are colour coded as explained in the colour key. Wild-type base and amino acid sequence is shown. This insert has no signal peptide, to prevent it being secreted from the cell and enable it to be harvested in large quantities from whole cell lysates.

Figure 2.13 Base and amino acid sequence of mutant *IGF1* vector insert with Strep II and His tags and no signal peptide

Base sequence:

TGGAGCCACCCACAGTTTGAGAAAGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGAT
GCTCTTCAGTTCGTGTGTGGAGACAGGGGCTTTTATTTCAACAAGCCCACAGGGTATGGCT
CCAGCAGTCGGAGGGCGCCTCAGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTG
ATCTAAGGAGGCTGGAGATGTATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCACTCTGT
CCGTGCCCAGCGCCACACCGACATGCCCAAGACCCAGAAGGAAGTACATTTGAAGAACGC
AAGTAGAGGGAGTGCAGGAAACAAGAACTACAGGATGCATCACCATCATCATCAC

Amino acid sequence:

KLATMWSHPQFEKGPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFR
SCDLRRLEMYCAPLKPAKSAHSVRAQRHTDMPKTQKEVHLKNASRGSAGNKNYRMHHHHHH

Colour key:

Strep II tag before mature *IGF1*, mature *IGF1*, location of mutation (first base/amino acid of E domain), *IGF1* E domain, His tag after E domain

The different regions of *IGF1* are colour coded as explained in the colour key. Mutant base and amino acid sequence is shown, with the location of the novel *IGF1* variant (102813333C>T, c.356G>A, p.R119H) highlighted in red.. This insert has no signal peptide, to prevent it being secreted from the cell and enable it to be harvested in large quantities from whole cell lysates.

2.17 Immunocytochemistry

Place one 12mm Deckgläser cover glass (catalogue number 631-0713) into each well of a 12 well plate. Coat with Poly-D-Lysine (Catalogue number: A3890401, Gibco™) to improve cell adherence to coverslips. Dilute Poly-D-Lysine 1:10 with autoclaved MilliQ water and apply to cover glasses in 12 well plate. Incubate at 37°C for 1-4 hours and then wash 4 times with autoclaved water. Leave to dry under hood.

Seed HEK293T cells to 40-50% confluency on these Poly-D-Lysine coated cover glasses. Transfect with constructs of interest. The following day, place cells on ice and gently apply 4% Paraformaldehyde (PFA) directly to the cell media. Leave for 15 minutes under fume hood, then discard PFA/media mix and apply 500µl 4% PFA directly to each well. Leave for 15 minutes under fume hood.

Following this incubation, wash the cells 3 times with pre-chilled PBS at 4°C. To lyse the cells, apply PBS with 0.2% triton and leave at room temperature for 15 minutes. Wash 3 times with PBS. Apply blocking buffer (PBS tween 0.1% with 10% goat serum) for 1 hour. Incubate overnight at 4°C with primary antibodies at 1:100 dilution with blocking buffer. Ensure samples stay protected from light and do not dehydrate by creating a damp, dark chamber for the cells.

The following day, wash the cells 3 times in PBS tween 0.1%. Apply the secondary antibodies at 1:500 dilution with blocking buffer. Incubate for one hour in a damp chamber protected from light at room temperature. Wash the cells 3 times with PBS tween 0.1%, then 3 times with MilliQ water. Aqueous mountant with DAPI (catalogue number 104139, Abcam, Cambridge, UK) was applied to the slides and left to set overnight at room temperature. Slides were stored at 4°C until acquisition of images on the Zeiss LSM 880 confocal microscope**.**

2.18 Isolating Streptomycin-tagged constructs using StrepTrap™ columns

I used 1ml StrepTrap™ High Performance prepacked chromatography columns (Catalogue number GE28-9075-46, Sigma Aldrich, Poole, UK) to isolate my streptomycin-tagged IGF1 constructs. This would enable me to compare cleavage of wild-type IGF1 compared to mutant constructs in my furin assay. These columns are prepacked with StrepTactin Sepharose™ High Performance, a medium for purifying Streptag™ II proteins.

Reagents

Binding buffer

PBS (20 mM sodium phosphate, 280 mM NaCl, 6 mM potassium chloride), pH 7.4.

Lysis buffer

0.1% PBS tween with added protease inhibitor.

Elution buffer

2.5mM desthiobiotin in PBS

10mM stock solution was created by adding 10ml PBS to 21.43g desthiobiotin and stirring well until dissolved. To create the desired 2.5mM concentration desthiobiotin, 2.5ml of this stock solution with combined with 7.5ml PBS.

Regeneration buffer

0.5 M NaOH

Method

HEK293T cells transfected with strep tagged constructs were cultured to 80-90% confluency in T75 dishes. Media was aspirated and the cells were gently washed with PBS. 3ml of 1x trypsin was then added to each T75 and flasks were incubated at 37°C in 5% CO_2 for 2-3 minutes. Following this incubation, 9ml pre-warmed media was added to each flask. This media was Dulbecco's Modified Eagle Medium High Glucose (Catalogue number D6429, Sigma Aldrich, Poole, UK) supplemented with 10% fetal bovine serum (FBS; catalogue number 11573397, Fisher Scientific, UK). The suspended cells were collected into 50ml falcon tubes and centrifuged at 1500rcf for 5 minutes. The cell pellet was then resuspended in 90% FBS/10% Dimethylsulfoxide (DMSO, catalogue number C6164, Sigma Aldrich, Poole, UK) solution and frozen in cryovials. These were stored at -80°C for short-term storage.

On the day of the StrepTrap™ Purification process, these samples were thawed and centrifuged at 4000rpm for 10 minutes at 4°C. The supernatant was removed and the custom lysis buffer described above was added to each cell pellet and resuspended. The samples were then sonicated for 10 seconds at 20KHz to break up cell membranes and fragment the DNA. Following sonication, the samples were placed immediately on ice for 5 minutes then the process was repeated once more.

During the cell lysis, the 1ml StrepTrap™ Purification columns were equilibrated using 10ml PBS over 10-15 minutes. Following equilibration, the soluble fraction from the whole cell lysates was added to the columns and run through at a similar speed to the PBS. The columns were then washed with 10ml PBS. At this stage, the Strep constructs would be expected to be bound within the columns. 1ml of elution buffer was then added to each column. This was repeated 3 times, with the 3 serial fractions collected in Eppendorf tubes. The StrepTrap™ Purification columns were then washed with 10ml PBS. To regenerate the columns, they were washed with 3ml MilliQ distilled water followed by 3ml 0.5 M NaOH and 3ml distilled water. The columns were once again equilibrated using 10ml PBS over 10-15 minutes prior to repeat use.

10µl of sample from each stage was obtained for Coomassie stain to assess the success of the different stages of the process. The samples were combined with 4x non-reducing fluorescent compatible sample buffer (catalogue number LC2570, ThermoFisher Scientific, UK) with added beta mercaptoethanol. They were boiled at 95°C for 5 minutes. 10µl of each sample was loaded onto an SDS-page gel and run at 80V for 150 minutes. InstantBlue® Coomassie Protein Stain (ISB1L) (ab119211, Abcam, Cambridge, UK) was added to the gel and incubated for 45 minutes at room temperature. Images of the gel were obtained using a whitebox imager.

2.19 IGF-1 furin cleavage assay

This assay enabled comparison of furin cleavage of mutant vs wild-type IGF-1 protein. Following a period of optimisation, these were the conditions I found optimal for my HEK293T whole cell lysates furin digestion.

Furin assay buffer

25mM Tris base

1mM $CaCl_2$

0.5% w/v Brij-35

Combine reagents and make up to 50ml with MilliQ water. Adjust pH to 9.0.

Furin resuspension

10µg rhFurin (catalogue number 1503-SE, R&D systems, Minneapolis, USA) was dissolved in 1ml furin assay buffer, giving a concentration of 10ng/µl

Furin cleavage assay

For the assay, the following reagents were combined in Eppendorf tubes:

Furin (10ng/µl)	16µl
Whole cell lysis protein	40µg
RIPA	to adjust for differences in whole cell lysate volumes of wild-type vs mutant
Furin assay buffer	10.25µl

Total volume 40µl. Negative control samples were included with no furin and 26.25µl furin assay buffer.

Samples were incubated at 37°C for 0, 60, 120 and 240 minutes. The negative control sample was incubated for 240 minutes to ensure the apparent digestion was not due to protein degradation. Following incubation, 20µl sample running buffer with beta-mercaptoethanol was added to the samples which were then boiled at 95°C for 5 minutes.

The samples were then placed onto ice for 1-2 minutes, briefly centrifuged once cool and stored at 4°C overnight. The following day, the samples were assessed for IGF-1 digestion by western blot,

probing with Anti-Strep-tag II antibody (ab76949). Intensity of western blot bands was quantified using the Image J application (https://imagej.nih.gov/ij/).

2.20 Biochemical assays for patients with c.618+836T>G, c.876-15T>G and c.902T>G *GHR* variants and c.356G>A *IGF1* variant

Biochemical assays for these families were performed by Dr Martin Bidlingmaier's team at the Endocrine Laboratory, LMU Klinikum (Munich, Germany). For each assay, all samples from the same family were analysed in the same analytical run. Family members who were homozygous, heterozygous and wild-type for the mutations of interest were assessed.

Serum IGF-1, GH and IGFBP 3 were measured using the IDS-iSYS platform (Immunodiagnostic Systems, Boldon, England, UK). The assays were calibrated against recombinant standards (98/574 for GH, 02/254 for IGF-1 and 93/560 for IGFBP 3). Intra- and inter-assay coefficients of variability (CVs) at various concentrations ranged from 4.0–8.7% (IGF-1), 1.3–5.4% (GH) and 5.5-12.4% (IGFBP 3). The limits of quantification are 10.0 ng/mL (IGF-1), 0.04 µg/L (GH) and 80.0 ng/mL, respectively (205–207). Serum ALS levels were measured in duplicate by sandwich immunometric assay using monoclonal antibodies directed against specific N- and C-terminal oligopeptides as previously described (208). A serum pool of healthy male volunteers was used for calibration and assigned 1000 mU/mL. Intra- and interassay CVs are <9%, the limit of quantification is 100 mU/mL, and the linear assay range is 100 to 5000 mU/mL (208). Serum GHBP concentrations were measured by an in-house, time-resolved fluorescence immunoassay (IFMA) based on monoclonal antibodies (209). The assay is standardized against recombinant non-glycosylated GHBP with concentration assigned by amino acid analysis (PRL Rehovot, Israel). Within-assay CVs were 3.4% at 312 pM and 3.4% at 2034 pM. At the same concentrations, between-assay CVs were 16.0% and 11.7%, respectively. The

lower limit of quantification was 80 pM and the linear range covered concentrations between 80-4880 pM.

Chapter 3. Variants of interest identified from custom short stature gene panel

3.1 Variants identified following bioinformatic analysis of the short stature gene panel

I analysed the short stature gene panel data of 92 undiagnosed patients in total alongside 13 control samples. 37 patients were female and 55 male, with a mean age at referral of 8.32 years (range 0.5 to 59 years), mean height SDS -3.67 (range -1.6 to -7.44). 17/92 (13%) patients were born small for gestational age (SGA), with mean birth weight (BW) SDS -1.08 (range 1.64 to -4.66). All patients had height SDS ≤-2 except for one younger sibling of an index case who was included as she shared her brother's clinical phenotype and had IGF-1 deficiency (height SDS -1.6).

I identified key genetic variants of interest in 17 of these patients. The phenotypic and biochemical details of these patients are provided in **Table 3.1**. The details of the genetic variants are described in **Table 3.2**. These variants were identified using the VCFs and BAM files generated from the gene panel and the bioinformatics filtering pipelines and *in silico* tools discussed previously (Methods **Sections 2.3 and 2.4**).

Genetic variants which were in keeping with the patient phenotypes were validated by Sanger sequencing in patient(s) and family members, where possible. Those variants which did not segregate with the phenotype in other family members were excluded.

Analysing the data generated from the custom gene panel, I identified several genetic variants of interest in genes of the GH-IGF-1 axis. I identified recognised pathogenic *GHR* mutations in Patients 1-4, in keeping with their classic GHI phenotypes. Patients 5 and 6 had classic GHI phenotypes but

no diagnostic exonic mutations. However, in both patients I identified very rare, deep intronic mutations with the potential to affect splicing of the *GHR* gene with the potential for pseudoexon inclusion. The functional analysis of these novel intronic *GHR* variants identified in Patients 5 and 6 is shown in results **Section 3.2.1** and **Section 3.2.2** respectively. Patients 7 and 8 showed milder growth failure and did not have classic GHI phenotypic features. In both patients I identified heterozygous *GHR* mutations that I predicted would affect splicing of exon 9 of the *GHR*. I hypothesised these variants may be acting in a dominant negative manner. As discussed in my introduction **Section 1.2.2.3**, patients with dominant negative *GHR* mutations have much milder growth failure and often no dysmorphic features. The functional analysis of the *GHR* variants identified in Patients 7 and 8 is shown in results **Section 3.2.3**. In Patient 9, I identified a novel heterozygous *GHR* variant in exon 10 that I hypothesised may also act in a dominant negative manner. Patient 9 had moderate postnatal growth failure (height SDS -4.0 at 8.0 years of age) which would be in keeping with the other patients reported in the literature with dominant negative *GHR* mutations but this variant requires functional analysis to ascertain its pathogenicity. Patient 10 had no dysmorphic features, a height SDS of -3.4 at 10.1 years of age and a low normal IGF-1 level. In Patient 10, I identified a novel heterozygous missense *IGF1* variant occurring at the key binding site of furin that I hypothesised may impair furin cleavage and thus the generation of the mature IGF-1, leading to postnatal growth failure and functional IGF-1 deficiency. The functional analysis of this *IGF1* variant is shown in results **Section 3.3.**

I also identified several variants in genes associated with 3M syndrome, Silver Russell syndrome and Noonan syndrome. I identified compound heterozygous mutations in *Cullin 7 (CUL7)* in Patient 11 and 12 that had not previously been reported in the literature. In Patient 11 I identified two missense variants in *CUL7 (*c.3349C>T; c.3490C>T) that were both predicted damaging by SIFT and assigned high CADD scores. The growth failure seen in Patient 11 (Height SDS -2.0 at 7.0 years) is at

the milder end of the spectrum for 3M syndrome. In Patient 12 I identified an in-frame CTC deletion and a missense variant in *CUL7* (c.1171_1173delGAG; c.2711G>A). The missense variant was assigned a very high CADD score of 34 suggesting pathogenicity, however the CADD score of the in-frame deletion was lower at 18 and this variant had a mean allele frequency of 0.446%. There is not much literature about the features of 3M syndrome in adulthood. Unfortunately, parental DNA was not available for either patient which would have helped in the assessment of pathogenicity. These results have been communicated back to the referring clinicians and with the research team at The University of Manchester with a specialist interest in 3M syndrome.

Patient 13 had several features of 3M syndrome with postnatal growth failure, dolicephaly, frontal bossing and lumbar lordosis. I identified a recognised pathogenic homozygous mutation in *Obscurin-like 1* (*OBSL1,* c.2134+1G>A) that disrupts a canonical splice site, explaining their clinical features. This was confirmed in the patient and both parents were found to be heterozygous upon Sanger sequencing. Patient 14 also had features of 3M syndrome, with pre and postnatal growth failure, triangular shaped face and prominent sternum. In this patient I identified a recognised pathogenic homozygous mutation in *Coiled Coil Domain–Containing 8 (CCDC8*; c.612dupG), in keeping with a diagnosis of 3M syndrome and explaining their phenotype.

Patient 15 exhibited pre and postnatal growth failure, macrocephaly and puffy hands and feet were noted at birth. I identified a novel heterozygous frameshift variant in *High-mobility group AT-hook 2 (HMGA2*, c.145delA). The functional analysis of this novel variants can be found in results **Section 5.2.**

Patient 16 had poor weight gain, frequent chest infections, mild learning difficulties and pre and postnatal growth failure. I identified a novel heterozygous mutation in *Insulin-like growth factor 2*

(IGF2, c.-7+130G>T). I predicted that this mutation could cause activation of an intronic splice donor site, leading to the inclusion of a pseudoexon with an alternative ATG start codon for the *IGF2* gene. Functional work is required to confirm this.

Patient 17 had a mild Noonan syndrome phenotype, with low set ears, delayed bone age and a degree of hypertelorism. I identified a missense heterozygous mutation in the *Son of Sevenless, Homolog 2* gene (*SOS2*, c.572C>G). Sanger sequencing verified the heterozygous mutation in the patient and his mother, who has a height SDS of -1.67. SIFT Function Prediction of this variant is Damaging, PolyPhen-2 Function Prediction is Benign and the CADD score is 23.4. The frequency of this mutation in Europeans is 0.631% in the gnomAD genome aggregation database. Upon balance, this was felt likely to be responsible for his phenotype and growth failure. He was referred to a clinical geneticist for their expert input and had an echocardiogram due to the recognised cardiac associations of Noonan syndrome. Achieving this diagnosis enabled the patient to commence GH therapy, as Noonan syndrome is a licensed indication for GH treatment.

Patients 5-8 with novel variants in the *GHR*, Patient 10 with a novel *IGF1* variant and Patient 15 with a novel *HMGA2* variant were selected for further functional analysis. In addition, two siblings were identified by collaborators (Patients 18 and 19) that were compound heterozygous for the novel c.618+836T>G intronic *GHR* variant identified in Patient 5 and a recognised nonsense point mutation in exon 4 of the *GHR* (c.181C>T; R43X) (74,210). These patients are all discussed in detail in my following results chapters.

Table 3.1 Phenotypic details of patients with genetic variants identified on short stature gene panel

Pt	Gene	Variant details	Phenotype	Ethnicity	Sex	Age (years)	Height SDS	BW SDS	GH peak (ug/L)	IGF-1 SDS
GH/IGF-1 axis genes										
1	*GHR*	Homozygous c.344A>C	Short, broad features. Sibling of Patient 2 and 3.	Syrian	M	15.3	-4.8	NK	40	-2.8
2	*GHR*	Homozygous c.344A>C	Short, broad features. Sibling of Patient 1 and 3.	Syrian	M	10.8	-3.7	NK	NK	NK
3	*GHR*	Homozygous c.344A>C	Short, broad features. Sibling of Patient 1 and 2.	Syrian	F	6.3	-4.5	NK	NK	NK
4	*GHR*	Homozygous c.508G>C	Laron phenotype, frontal bossing.	Pakistani	F	6.5	-3.7	-1.1	NK	-2.8
5	*GHR*	Homozygous c.618+836T>G	Prominent forehead, depressed nasal bridge, small chin, delayed tooth eruption, small hands and feet, micropenis and empty and hypoplastic scrotum. Delayed psychomotor development, high pitched voice.	Caucasian (Italian)	M	2.9	-7.4	-0.4	Baseline GH 38ug/L	-2.3
6	*GHR*	Homozygous c.70+32137A>G	Depressed nasal bridge, midfacial hypoplasia, frontal bossing, high pitched voice.	South Indian	M	12.3	-5.9	-1.3	>40	IGF-1 <25ng/ml*
7	*GHR*	Heterozygous c.876-15T>G	Relative macrocephaly, borderline mesomelic shortening in the upper limb, slight lumbar hyperlordosis.	Caucasian (British)	M	16.5	-3.2	-2.4	NK	+2.2

8	*GHR*	Heterozygous c.902T>G	No dysmorphic features.	Caucasian (Slovakian)	M	14.6	-2.7	0.2	57.5	-3.0
9	*GHR*	Heterozygous p.R386C	Developmental delay, visually impaired, nystagmus.	Caucasian (British)	F	8.0	-4.0	0	15.2	IGF-1 38.8ug/L*
10	*IGF1*	Heterozygous c.356G>A	No dysmorphic features.	Caucasian (British)	F	10.1	-3.4	-0.2	17.1	-1.3
3M syndrome genes										
11	*CUL7*	Compound heterozygous c.3349C>T; c.3490C>T	Pectus carinatum, high-pitched voice.	Latin-Mexican	M	7.0	-2.0	NK	19	-2.6
12	*CUL7*	Compound heterozygous C.1171_1173del GAG; c.2711G>A	Central obesity, possible cataracts. 3 previous miscarriages, non-specified uterine tumour.	Latin-Mexican	F	59.0	-4.5	-1.1	0.09	-1.72
13	*OBSL1*	Homozygous c.2134+1G>A	Frontal bossing, flat nasal bridge, relatively large head with increased antero-posterior diameter, mid facial hypoplasia, dolichocephaly, bushy eyebrows, mild hirsutism, lumber lordosis, and protuberant abdomen.	Sri Lankan	F	10.0	-4.5	-0.8	Basal GH 4.1ug/L	IGF-1 295ug/L*
14	*CCDC8*	Homozygous c.612dupG	Triangular face, prominent sternum.	Pakistani	M	1.6	-7.4	-3.7	9.9	-2.4
SRS/SRS-like genes										
15	*HMGA2*	Heterozygous c.145delA	Macrocephaly, puffy hands and feet noted at birth.	Latin-Mexican	F	0.5	-3.7	-3.5	NK	-1.91

16	*IGF2*	Heterozygous c.-7+130G>T	Poor weight gain, regular chest infections, mild learning difficulties.	Caucasian (British)	F	9.6	-3.7	-3.0	30	-2.4
Noonan syndrome genes										
17	*SOS2*	Heterozygous c.572C>G	Low set ears, mild hypertelorism, joint hypermobility.	Caucasian (British)	M	9.4	-2.0	1.2	10.3	-1.2

Pt, Patient; NK, not known; SDS, Standard Deviation Score. * unable to calculate IGF-1 SDS as assay normal range not known.

Table 3.2 Genetic variants identified by the custom short stature gene panel

Pt	Gene	Variant details	Segregation	Predicted effect of variant	Base change, transcript + AA change	Reported gnomAD frequency	Pathogenicity scores	Reference
GH-IGF-1 axis genes								
1,2,3	*GHR*	Homozygous	Confirmed in patients (3 siblings)	Destabilizing effect on GHR due to disruption of hydrogen bond; the OD1 of the native AA formed a H bond with N of residue B99; the replacement AA is unable to maintain this hydrogen bond (SAAPdap predictor)	42695096A>C, c.344A>C, pN115T	Not reported in gnomAD	SIFT: Damaging, Polyphen-2: Probably Damaging, CADD score: 23.7	(211)
4	*GHR*	Homozygous	Confirmed in patient	GHR with changed amino acid at this key position in the extracellular domain is unable to dimerize so GH signalling cannot occur	42699994G>C, c508G>C, pD170H	0.001%	SIFT: Damaging, Polyphen-2: Probably Damaging, CADD score: 28.5	(212)
5	*GHR*	Homozygous	Confirmed in patient, both parents heterozygous	Predicted to cause activation of deep intronic cryptic donor site and inclusion of novel pseudoexon into *GHR.* Please see results **Section 3.2.1**	42700940T>G, c.618+836T>G	Novel	CADD score: <10	n/a
6	*GHR*	Homozygous	Confirmed in patient, both parents heterozygous	Predicted to activate an intronic cryptic donor site deep within intron 2 *GHR.* Please see results **Section 3.2.2**	42598173, c.70+32137A>G	0.058%	CADD score 17.5	n/a

7	*GHR*	Heterozygous DN	Confirmed in patient, inherited from short mother	Precited to disrupt polypyrimidine tract prior to exon 9 *GHR* canonical splice site. Alternative splice site within exon 9 may be recognised by spliceosome, causing loss of 26bp from exon 9 causing a frameshift. Please see results **Section 3.2.3**	42718139T>G, c.876-15T>G	0.029%	CADD score <10	n/a
8	*GHR*	Heterozygous DN	Confirmed *de novo* in patient. Both parents and sibling wild-type	Predicted to activate cryptic splice site within exon 9 of *GHR*, which would lead to loss of 26bp from exon 9 causing a frameshift. Please see results **Section 3.2.3**	42718180T>G, c.902T>G	Novel	SIFT: Damaging, Polyphen: Probably damaging, CADD score: 27.7	n/a
9	*GHR*	Heterozygous	Confirmed *de novo* in patient, both parents wild-type	Predicted to exert dominant negative effect on *GHR* function but further work required to confirm this	42718765C>T, c.1156C>T, p.R386C	0.034%	SIFT: Damaging, Polyphen-2: Probably Damaging, CADD score: 34.0	n/a
10	*IGF1*	Heterozygous	Confirmed in patient, inherited from short father	Predicted to impair mature IGF-1 formation. Missense change affecting key binding site of furin. Furin binds and cleaves Pro IGF-1 to mature IGF-1. Please see results **Section 3.3**	102813333C>T, c.356G>A, p.R119H	0.004%	SIFT: Damaging, Polyphen: Probably damaging, CADD score: 32.0	n/a

3M syndrome genes								
11	*CUL7*	Compound heterozygous	Confirmed in patient. Parental DNA not available as adopted	Missense variant with predicted deleterious effect on gene function Missense variant with predicted deleterious effect on gene function	43011192G>A, c.3349C>T, p.R1117W; 43010695G>A, c.3490C>T, p.R1164W	0.007% 0.174%	SIFT: Damaging, PolyPhen-2: Possibly damaging, CADD score 23.1 SIFT: Damaging, PolyPhen-2: Possibly damaging, CADD score 28.1	n/a n/a
12	*CUL7*	Compound heterozygous	Confirmed in patient. No parental DNA available	In-frame CTC deletion; Missense variant with predicted deleterious effect on gene function	43018766, C.1171_1173delGAG; 43013779G>A, c.2711G>A.	0.446% 0.002%	CADD score 18 CADD score 34	n/a n/a
13	*OBSL1*	Homozygous	Confirmed in patient, both parents heterozygous	Decreases splicing efficiency by 96% (MaxEntScan score) Strongest nearby alternative splice site 52bp 5' resulting in deletion causing a frameshift	220431551C>T, c.2134+1G>A	0.001%	CADD score 34.0	(213)
14	*CCDC8*	Homozygous	Confirmed in patient	Insertion of C leading to frameshift of coding sequence	46915455 ins C, c.612dupG, p.K205FS*59	0.001%	CADD score: 33.0	(127,214)
SRS/SRS-like genes								
15	*HMGA2*	Heterozygous	Confirmed heterozygous in patient	Deletion of A leading to frameshift of coding sequence. Please see results **Section 5.2**	66221814delA, c.145delA, p.Arg49Glyfs*117	Novel	Mutation Taster: disease causing	n/a

16	*IGF2*	Heterozygous	Awaiting repeat DNA samples	Activation of intronic donor site. Could cause inclusion of pseudoexon with alternative ATG start codon. Functional work required to confirm this.	2159329C>A, c.-7+130G>T	0.003%	CADD score: 13.7	n/a
Noonan syndrome genes								
17	*SOS2*	Heterozygous	Confirmed in patient, inherited from short mother (height SDS -1.7)	Missense change	50655357G>C, c.572C>G, p.P191R	0.631%	SIFT: Damaging, PolyPhen-2: Benign CADD score: 23.4	n/a

Pt, Patient; AA, amino acid; SRS, Silver Russell Syndrome; DN, Dominant Negative. SAAPdap predictor (www.bioinf.org.uk); gnomAD genome aggregation database (https://gnomad.broadinstitute.org/); MaxEntScan score (http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) (215); Polyphen-2 score, (http://genetics.bwh.harvard.edu/pph2) (165) SIFT score, Sorting Intolerant From Tolerant score (http://blocks.fhcrc.org/sift/SIFT.html) (166); CADD score, Combined Annotation Dependent Depletion Score (https://cadd.gs.washington.edu/) (167)

Chapter 3.2 Novel splicing mutations in the *Growth Hormone Receptor* causing a spectrum of Growth Hormone Insensitivity

3.2.1 Identification of novel *GHR* pseudoexon variant

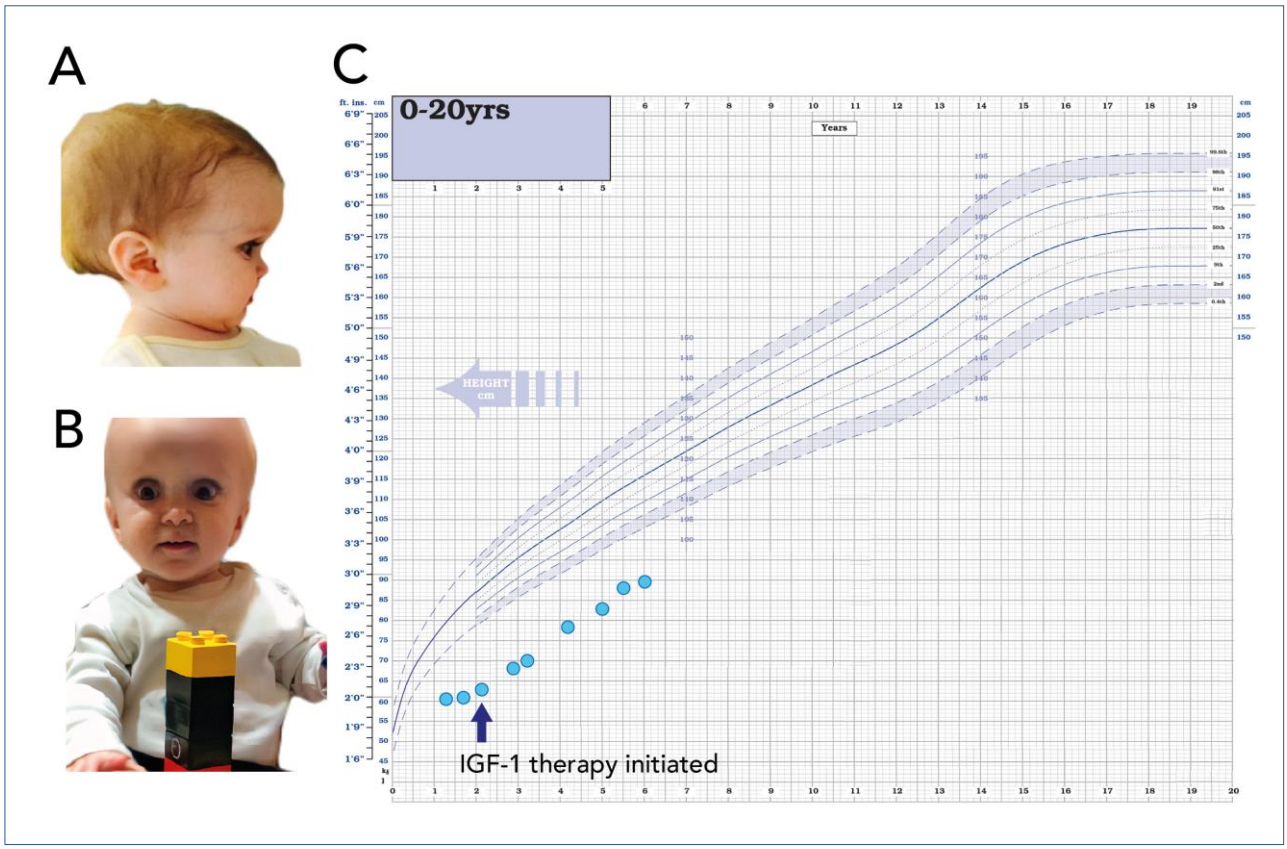
3.2.1.1 Patient phenotypes

Kindred 1

The Index case (Patient 5, **Figures 3.1A and 3.1B**) was referred to our genetic sequencing service at 2.9 years of age with classical biochemical and phenotypic features of GH Insensitivity (GHI; 'Laron syndrome') (**Table 3.3**). He was the second child of unrelated non-dysmorphic Caucasian parents. He was born at 37 weeks' gestation with a normal birth weight (BW SDS -0.4). Severe postnatal growth failure was first noted at 5 months of age and by 1.7 years his height was 61 cm (height SDS -7.4) (**Figure 3.1C**), he had a normal BMI 16.5 kg/m^2 (SDS -0.6) and relative macrocephaly (head circumference SDS -1.2). At presentation he had classical Laron facial features, delayed tooth eruption, small hands and feet, micropenis, bilateral undescended testes and hypoplastic scrotum. Maternal and paternal heights were -2.0 and -1.5 SDS, respectively and there was no family history of growth failure. Random serum GH was extremely elevated (38 µg/L; normal range (NR) 0-20 ng/mL). At diagnosis, he was noted to have severe deficiencies of IGF-1 (<10 ng/mL; NR 13-143) and IGFBP 3 (<80 ng/ml; NR 1612-4525), ALS and GHBP levels were undetectable (<100 mU/ml and <80 pM, respectively). IGF-1 levels during a 5-day IGF-1 generation test (IGFGT; GH 0.033 mg/kg/day, performed according to established protocols) demonstrated an IGF-1 level of <10 ng/mL at baseline and at 4 days following GH administration, indicating severe GH resistance (**Table 3.4**). He was diagnosed with severe GH resistance/primary IGF-1 deficiency and commenced recombinant human IGF-1 therapy (rhIGF-1; 120 µg/kg by subcutaneous injection twice daily) at 2.1 years of age. He had many episodes of hypoglycemia which required continuous glucose monitoring for 6 months. He developed a mild, isolated but persistent elevation of TSH (maximum 7 µU/mL; NR 0.3-

4). Following commencement of rhIGF-1 therapy, his height velocity improved considerably from 2.2cm/year to 8.1cm/year and has remained consistently above baseline (5.0-8.5cm/year) suggesting a good response to rhIGF-1 therapy. At the latest assessment aged 6 years, his height was 89 cm (-5.0 SDS) (**Figure 3.1C**).

Figure 3.1 Clinical images and growth chart of Patient 5



A and B. Clinical images of Patient 5 aged 1.7 years with classical Laron features of mid-facial hypoplasia, depressed nasal bridge and frontal bossing. **C.** Growth chart showing severe postnatal growth failure and response to rhIGF-1 therapy.

Table 3.3 Clinical and auxological details for the patients with the novel c.618+836T>G *GHR* 6Ω pseudoexon mutation

Phenotypic details	Patient 5	Patient 18	Patient 19
Age (years)	1.7	9.6	3.4
Height (cm) [SDS]	61.0 [-7.4]	83.2 [-9.3]	67.0 [-6.9]
Height velocity (cm/year) [SDS]	1.6 [-4.5]	1.5 [-5.2]	2.0 [-4.4]
Weight (kg) [BMI SDS]	6.1 [-0.6]	10.7 [-1.0]	5.5 [-4.4]
Head circumference [SDS]	45.9cm [-1.2]	45.0 cm [-5.7]	43.0 cm [-5.6]
Bone age (years)	NK	4.0	1.5
Birth weight (kg) (gestation) [SDS]	2.8 (37/40) [-0.4]	2.6 (40/40) [-2.3]	2.1 (41/40) [-3.8]
Other phenotypic details	Small hands and feet Undescended testes Hypoplastic scrotum Micropenis Delayed tooth eruption	Small hands and feet Undescended testes Micropenis Mild learning difficulties Bilateral hearing loss Pubertal delay	Small hands and feet Undescended testes Micropenis Mild learning difficulties Recurrent hypoglycemia Mild papilloedema Necrotising enterocolitis

SDS, Standard Deviation Score; NK, not known.

Table 3.4 Biochemical details of the patients with the novel c.618+836T>G *GHR* 6Ω pseudoexon mutation and their parents

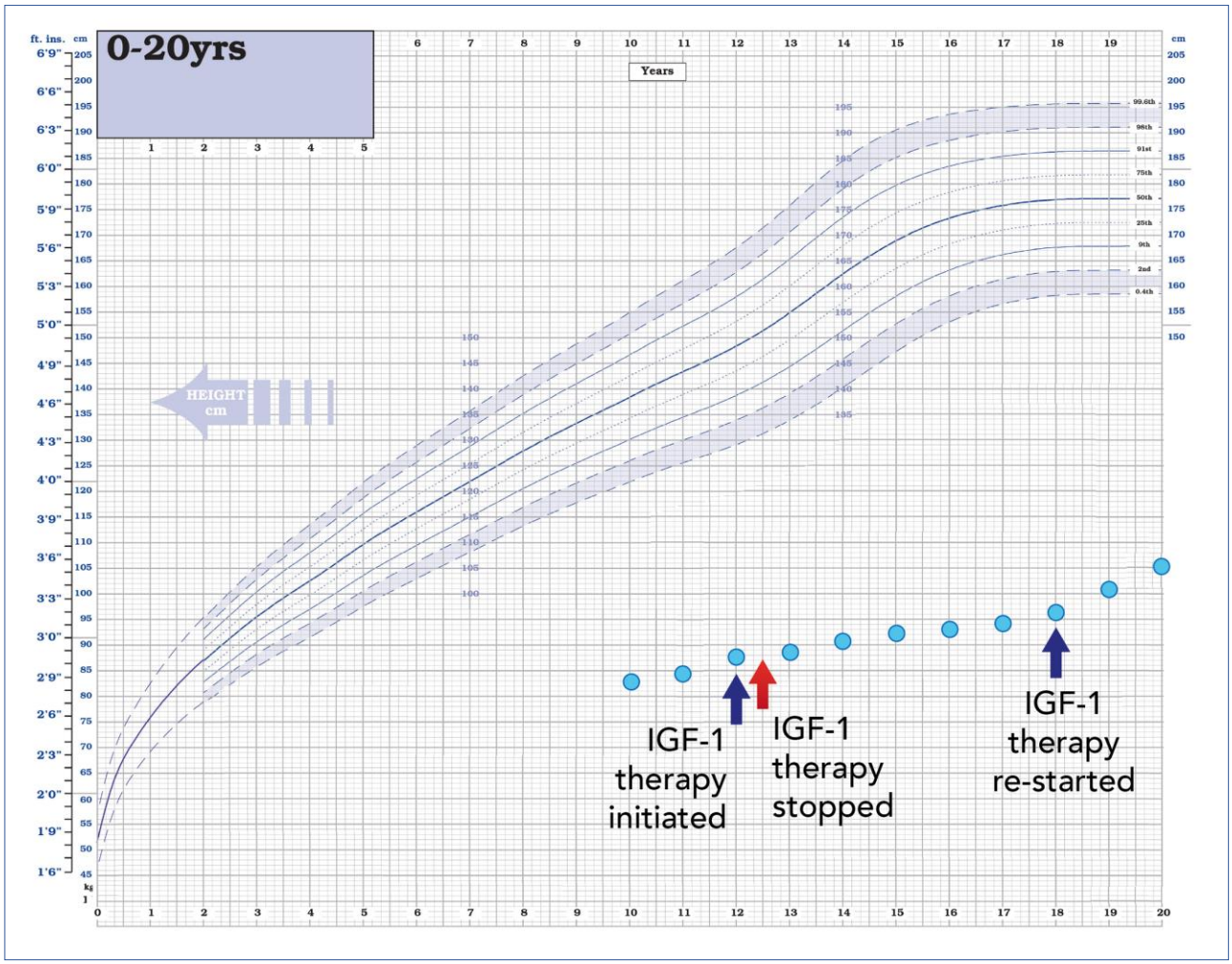
	Kindred 1			Kindred 2			
	Patient 5 (P5)	**P5 Mother**	**P5 Father**	**Patient 18 (P18)**	**Patient 19 (P19)**	**P18/19 Mother**	**P18/19 Father**
Age at presentation (years)	1.7	42.5	42.1	9.6	3.4	42.5	42.9
Height SDS	**-7.4**	**-2.0**	-1.5	**-9.3**	**-6.9**	**-3.2**	-0.7
Basal GH (μg/L)	**38.0**	ND	ND	**52.0**	**110.0**	0.21	5.23
IGF-1 (ng/mL) [SDS]	**<10 [-2.6]***	165 [+0.9]	**244 [+2.7]**	**<10 [-3.4]**	**<10 [-2.7]**	70 [-1.7]	**45 [-2.8]**
IGFGT: Basal; Peak IGF-1 (ng/ml)	**<10; <10**	ND	ND	**<10; <10**	**<10; <10**	ND	ND
IGFBP 3 (ng/mL) [SDS]	**<80 [-4.1]***	3333 [-1.1]	3897 [-0.1]	**<80 [-4.6]**	**274 [-3.8]**	**1603 [-3.0]**	**1700 [-2.9]**
ALS (mU/mL) [SDS]	**<100 [-2.6]***	620 [0.1]	594 [-0.1]	**<100 [-4.3]**	**<100 [-2.6]**	**183 [-2.6]**	**184 [-2.6]**
GHBP (pM) [SDS]	**<80***	1345 [-1.0]	701 [-1.8]	**<80**	**<80**	**247 [-2.4]**	**111 [-2.6]**
TSH (μU/mL) [NR]	7.0 **[0.3-4.0]**	ND	ND	**6.8 [0.3-4.2]**	**13.8 [0.3-4.2]**	ND	ND

ND, not done; NR, normal range; SDS, Standard Deviation Scores calculated based on the normal ranges for age and sex; IGFGT, IGF-1 generation tests (following established protocols using GH 0.033 mg/kg/day for 5 days (Patient 5) and 7 days (Patients 18 and 19)). *Samples obtained at 3.2 years of age. Results highlighted in bold are outside of the normal range, SDS -2 to +2.

Kindred 2

Patient 18 presented with severe growth failure at 9.6 years of age with a height of 83.2 cm (-9.3 SDS) and height velocity of 1.5 cm/year (-5.2 SDS) (**Table 3.3 and Figure 3.2**). Head circumference was 45 cm (-5.7 SDS) and BMI within the normal range (-1.0 SDS). At a chronological age of 9.6 years, bone age was significantly delayed at 4.0 years. At presentation, he had small hands and feet, undescended testes and micropenis. He did not have obvious 'classical' Laron facial features (frontal bossing or mid facial hypoplasia) but had reduced facial height (nasion to menton; -4.9 SDS) compared to head width (maximal biparietal diameter; -1.2 SDS) (216). He also suffered from mild learning difficulties, bilateral hearing loss and pubertal delay was later noted. At 40 weeks' gestation, he was born small for gestational age (SGA) with a birth weight of 2.6 kg (-2.3 SDS). The end stages of the pregnancy were complicated by pre-eclampsia. At diagnosis his basal GH levels were very elevated at 52.0 µg/mL with undetectable IGF-1 and GHBP (<10 ng/mL and <80 pM, respectively) and severe deficiencies of IGFBP 3 and ALS (<80 ng/mL and <100 mU/mL, respectively). IGFGT (0.033 mg/kg/day for 7 days, as above) showed no response to GH, with baseline and peak levels of IGF-1 <10 ng/mL (**Table 3.4**). His parents were non-consanguineous with no dysmorphic features. The father had a normal height (0.7 SDS) and his mother had short stature (-3.2 SDS). TSH levels were slightly elevated (6.8 µU/mL; NR 0.3-4.2) but FT4 1.5 ng/dl (NR 0.9-1.7) and FT3 3.0 pg/ml (NR 3.0-4.7) were consistently normal and levothyroxine therapy was never required. He commenced rhIGF-1 at 12 years (120 µg/kg by subcutaneous injection twice daily) but stopped after 6 months. He recommenced rhIGF-1 therapy at 18 years (height 96.2 cm, -11.8 SDS) and continued until 21 years of age (height 109.1 cm, -9.9 SDS). His height velocity improved considerably during the periods of rhIGF-1 therapy but unfortunately compliance was poor, and the duration of treatment was suboptimal (**Figure 25**). His final adult height at 23 years is 110 cm (-9.7 SDS) (**Figure 25**). He did not give consent for the clinical photographs at diagnosis to be included within this thesis.

Figure 3.2 Growth chart of Patient 18

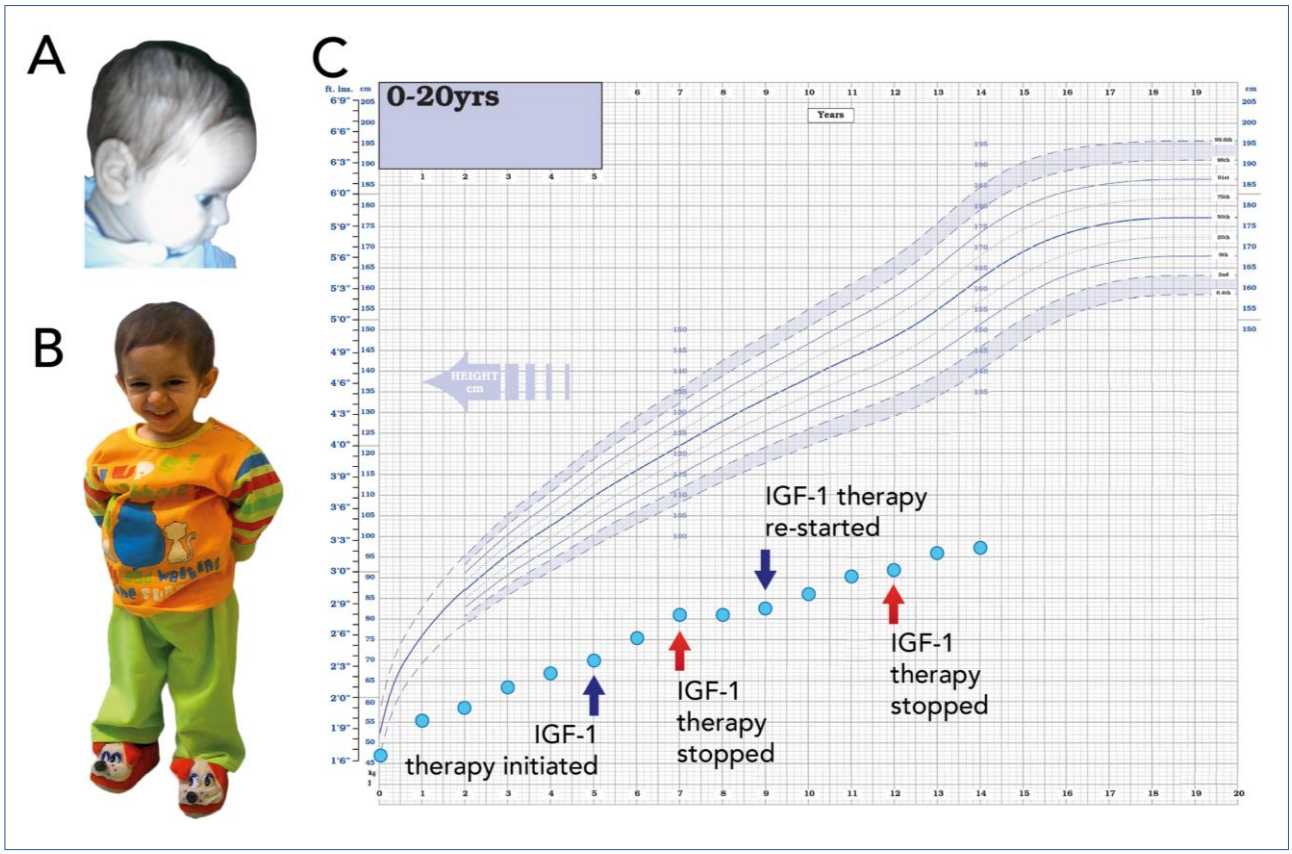


Growth chart of Patient 18 showing severe postnatal growth failure. Periods of rhIGF-1 therapy are indicated by the blue arrows.

Patient 19 (**Figure 26A and B**), the younger sibling of Patient 18, was also born small for gestational age (birth weight 2.1kg; -3.8 SDS) at 41 weeks gestation. At 3.4 years, he presented with a height of 67 cm (-6.9 SDS) and height velocity of 2.0 cm/year (-4.4 SDS) (**Table 3.3**). His head circumference was 43.0 cm (-5.6 SDS) and BMI -4.4 SDS. At a chronological age of 3.4 years, his bone age was significantly delayed at 1.5 years. At presentation he had small hands and feet, undescended testes, micropenis and mild papilledema. Images from early infancy showed frontal bossing **(Figure 26A)**. He also suffered from recurrent hypoglycemia, mild learning difficulties and an episode of necrotising enterocolitis. At diagnosis, baseline GH levels were elevated at 110.1 µg/L and IGFGT (0.033 mg/kg/day for 7 days, as above) showed no response to GH, with baseline and peak IGF-1

levels of <10 ng/mL (**Table 3.4**). He had deficiencies of IGFBP 3 and ALS (274 ng/mL and <100 mU/mL, respectively). He was commenced on levothyroxine at 2 months of age due to hyperthyrotropinaemia with TSH of 13.8 µU/ml (NR 0.3-4.2). FT4 and FT3 have remained within the normal range on treatment (latest FT3 level 3.1 pg/ml; NR 3.0-4.70). He has undergone periods of rhIGF-1 therapy (120 µg/kg subcutaneous injection twice daily) with variable compliance (similar to his sibling) initially commencing treatment at 5 years of age (height 70 cm, -8.1 SDS) until aged 7 years (81 cm, -7.2 SDS), restarting at 9 years (height 83 cm, -8.1 SDS) until 12 years (92 cm, -7.6 SDS). Subsequently he remained off treatment and his height at latest assessment at 14 years of age is 97.0 cm (-7.7 SDS) . His height velocity improved during the initial period of rhIGF-1 therapy, but the treatment response and outcome was likely affected by the significant compliance issues (**Figure 26C**).

Figure 3.3 Clinical images and growth chart of Patient 19

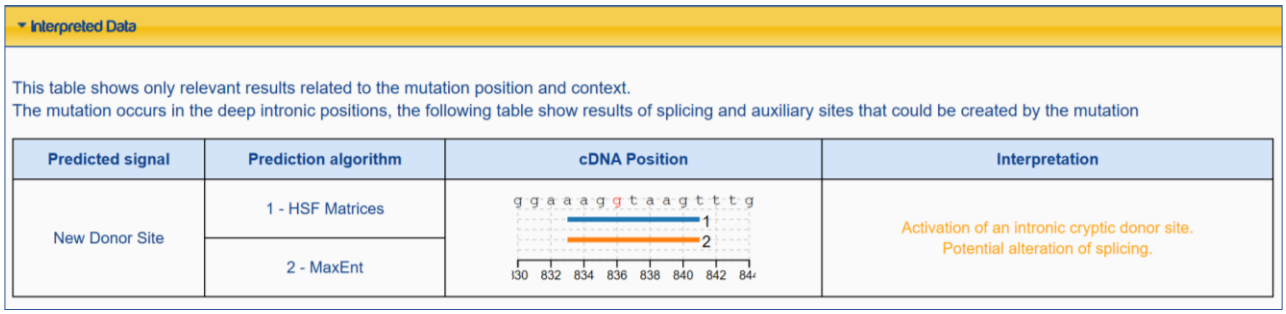


A and B. Clinical images showing Patient 19 (P19; younger sibling of P18) aged 3.5 months (A) and 5.4 years (B) displaying classical Laron features of mid-facial hypoplasia, depressed nasal bridge and frontal bossing. **C**. Growth chart showing severe postnatal growth failure. Periods of IGF-1 therapy are indicated.

3.2.1.2 Bioinformatic analysis and splicing prediction

I identified an exceedingly rare, homozygous variant in intron 6 of the *GHR* gene (42700940T>G, c.618+836T>G). This variant, altering the sequence from AGTT to AGGT, was predicted to activate an intronic cryptic donor splice site deep within intron 6 of the *GHR* (**Figure 3.4**). This is a novel variant not listed in the 1,000 genomes, ExAC, gnomAD and NHLBI ESP exomes. It was assigned a CADD score <10, which is not unusual for a non-coding variant.

Figure 3.4 Splicing prediction generated by Human Splicing Finder for c.618+836T>G *GHR* variant



▾ Interpreted Data

This table shows only relevant results related to the mutation position and context.
The mutation occurs in the deep intronic positions, the following table show results of splicing and auxiliary sites that could be created by the mutation

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
New Donor Site	1 - HSF Matrices	g g a a a g g t a a g t t t g 1 2 830 832 834 836 838 840 842 844	Activation of an intronic cryptic donor site. Potential alteration of splicing.
	2 - MaxEnt		

The novel c.618+836T>G *GHR* variant is predicted to activate an intronic cryptic donor site. Image generated from Human Splicing Finder (http://umd.be/HSF3/index.html)

Interestingly, the novel intronic c.618+836T>G variant is 44bp downstream of the original *GHR* pseudoexon variant (6Ψ; c.618+792A>G) (**Figure 3.5**) (63). This original 6Ψ *GHR* pseudoexon variant is described in detail in section **1.2.2.2.** This homozygous c.618+792A>G deep intronic *GHR* mutation leads to mis-splicing and inclusion of an in-frame insertion to the Growth Hormone Receptor, causing a partial loss of function due to impaired trafficking.

Figure 3.5 Location of original 6Ψ *GHR* pseudoexon and the novel 6Ω variant identified on the short stature gene panel

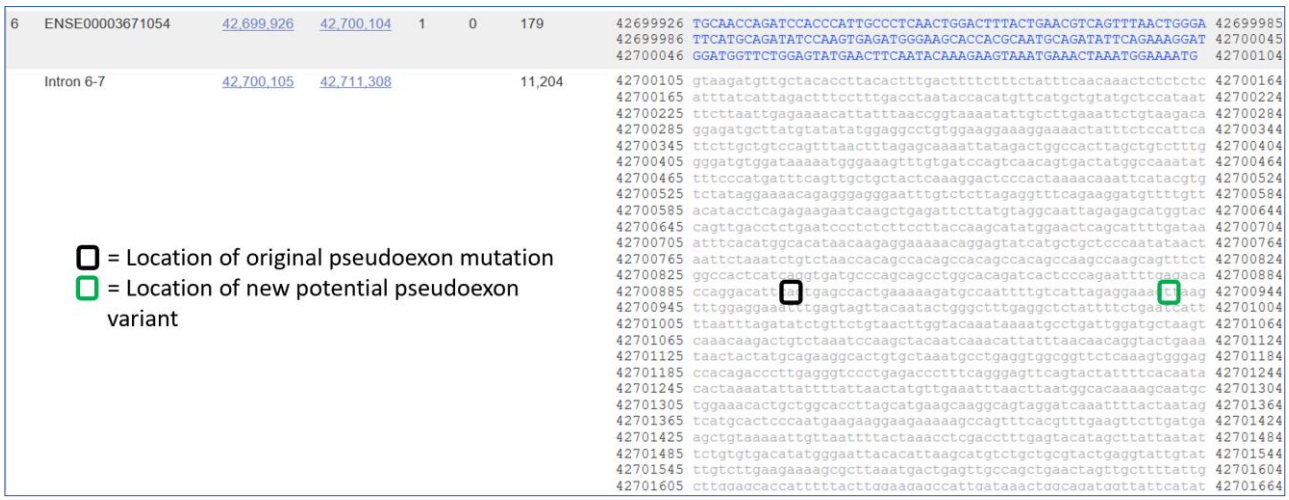
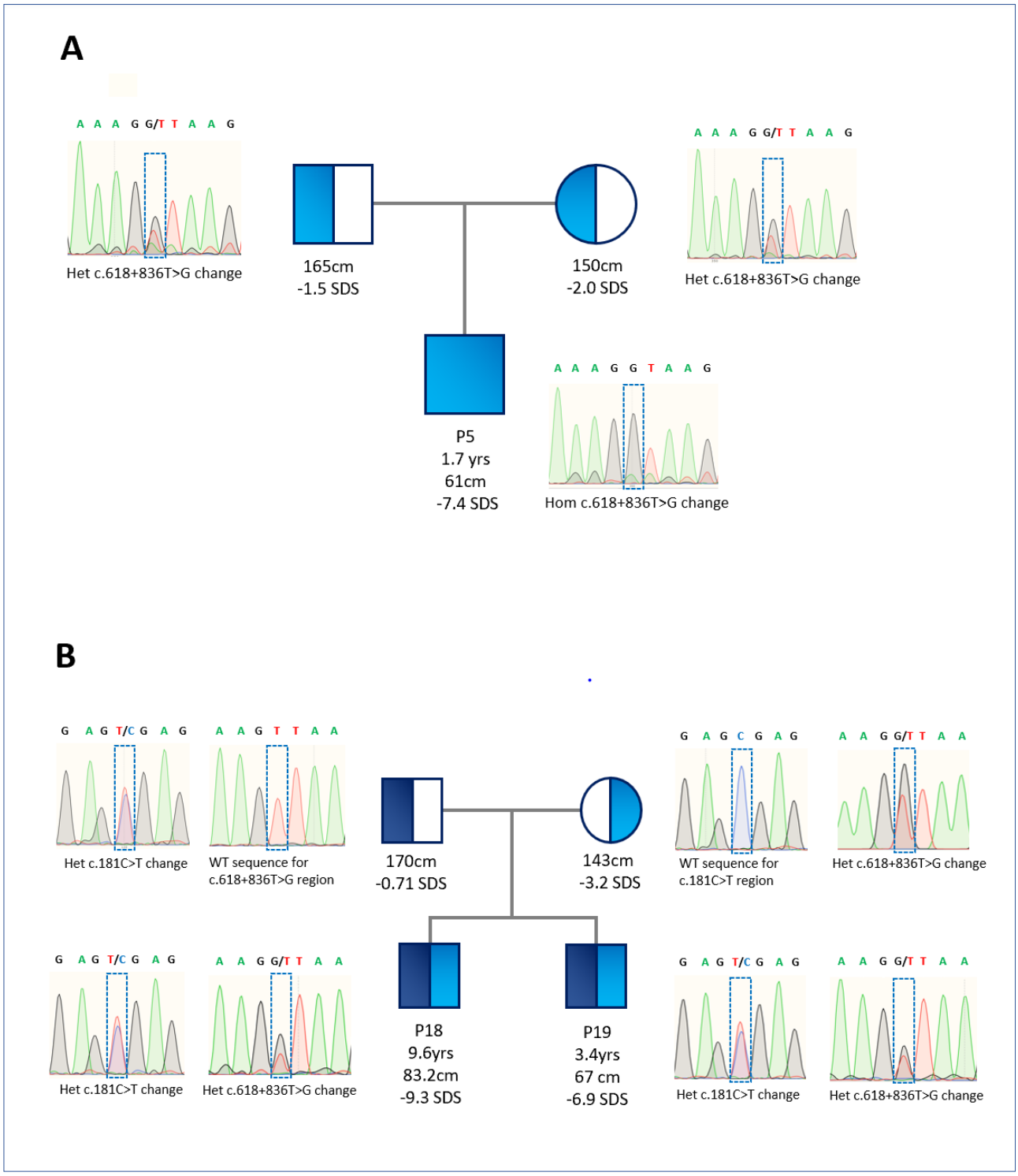


Image created using *GHR* sequence available on Ensembl37 Genome Browser. (http://grch37.ensembl.org/index.html). The T which is altered in the novel c.618+836T>G *GHR* change can be seen in the green box downstream of the original *GHR* pseudoexon variant in black box (c.618+792A>G).

3.2.1.3 Segregation of novel 6Ω *GHR* variant

This novel 42700940T>G, c.618+836T>G *GHR* variant was confirmed by PCR using primers I designed to cover the affected region (GHR intron 6F (Forward) and GHR intron 6R (Reverse); Sequences are provided in **Appendix 1**) in Patient 5 and his parents. This confirmed Patient 5 as homozygous for this novel variant and revealed both parents to be heterozygous (**Figure 3.6A**). Following my identification of this novel intronic change, targeted Sanger sequencing of the intron 6 region of the *GHR* was undertaken in Patients 18 and 19 and their parents at a collaborating centre in Italy. This was performed in these patients as they had a classic GHI phenotype but after candidate gene sequencing including the exons of the *GHR*, only a heterozygous mutation had been identified in exon 4 of the *GHR*, c.181C>T; R43X (**Figure 3.6B**). This previously reported mutation was inherited from their father and is predicted to lead to frameshift and induction of an early stop at residue 43 of the GHR (74,210). However in isolation was not felt to explain their classic GHI phenotype. Sequencing of intron 6 revealed that Patients 18 and 19 were both heterozygous for the novel intronic *GHR* variant (42700940T>G, c.618+836T>G), inherited from their mother.

Figure 3.6 Pedigrees and sequencing chromatograms for Kindreds 1 and 2



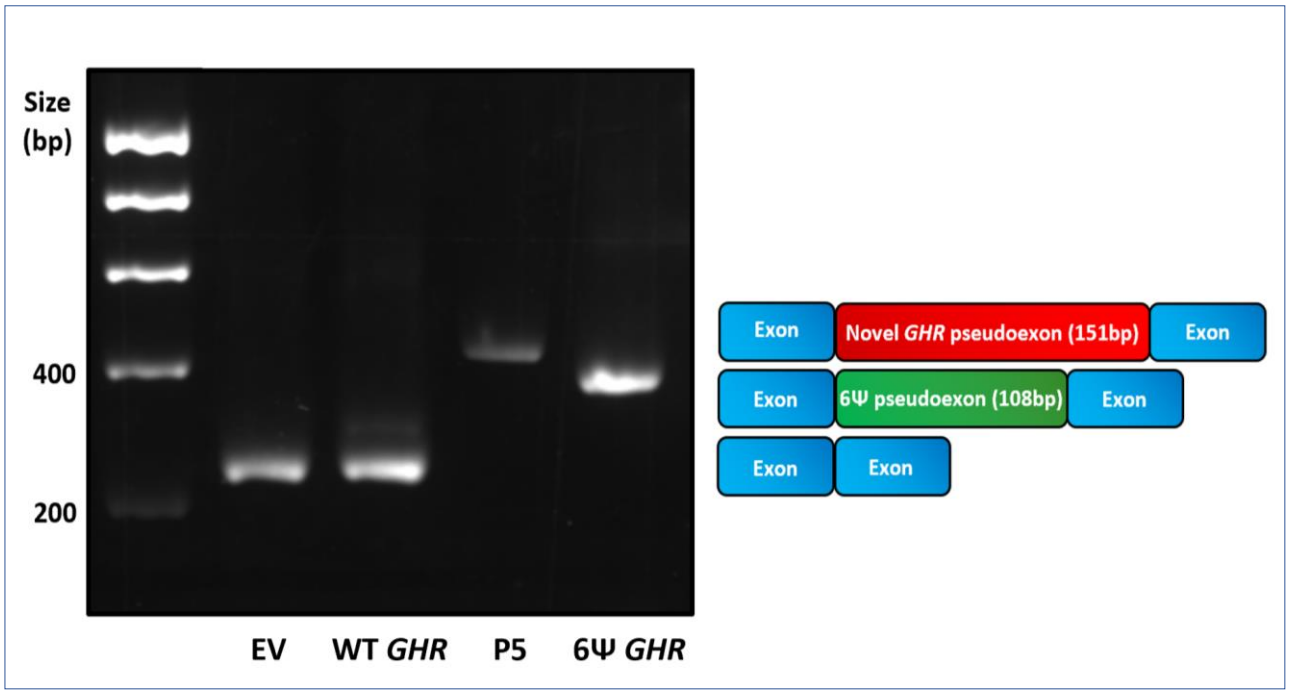
Family pedigrees and chromatograms showing the segregation of the novel c.618+836T>G *GHR* variant in affected families. **A**. Homozygous and heterozygous c.618+836T>G *GHR* variants in Patient 5 (P5) and both parents, respectively. **B.** Patients 18 and 19 (P18 and 19) harboured compound heterozygous c.618+836T>G (maternally inherited) and c.181C>T (R43X) (paternally inherited) *GHR* mutations.

3.2.1.4 Functional assessment of novel c.618+836T>G *GHR* variant

3.2.1.4.1 *In vitro* splicing assay

To assess whether this novel intronic variant affected *GHR* splicing, I performed an *in vitro* splicing assay, as described previously. I designed primers to amplify a 600bp DNA fragment of interest from Patient 5, a control sample and also a patient with the previously described 6Ψ pseudoexon mutation (c.618+792A>G) for comparison. This splicing assay revealed the inclusion of 151bp in addition to the two exons of the exon trap vector confirming 6Ω pseudoexon inclusion (**Figure 3.7**).

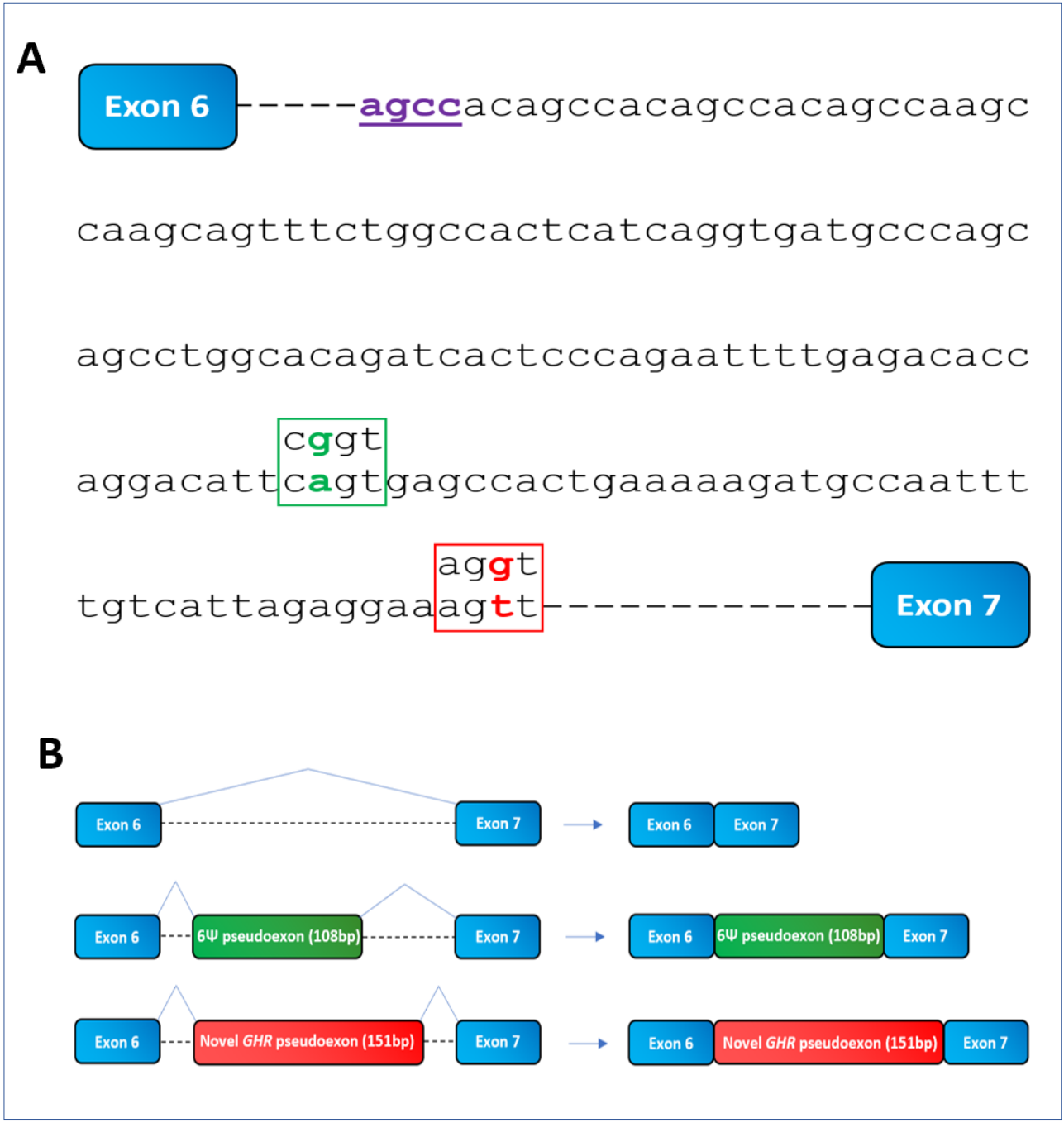
Figure 3.7 Splicing assay confirms retention of novel 6Ω *GHR* pseudoexon



Gel electrophoresis of cDNA splicing products following the splicing assay using an exon trap vector (MoBiTec-Exontrap cloning vector pET01). EV, empty vector, pET01 alone; WT-*GHR*, pET01 with 600bp of wild-type *GHR* intron 6 sequence inserted; Patient 5, pET01 with 600bp of Patient 5 intron 6 sequence inserted (including the c.618+836T>G variant). *GHR*-6Ψ, pET01 with 600bp sequence from a patient with the original *GHR* pseudoexon (6Ψ) c.618+792A>G variant. The spliced products were amplified by PCR and visualized on a 2% agarose gel. Lanes 1 and 2: A 250bp band is seen in empty vector and WT sequence, as expected, representing the two exons of the exon trap vector and confirming normal splicing with WT sequence (lane 2). Lane 3: A 401bp band is seen for Patient 5 and sequencing revealed 151bp insert between the two exons of the exon trap vector (250bp) confirming novel 6Ω pseudoexon inclusion. Lane 4: A 358bp band is seen in the GHR 6Ψ patient sample and sequencing revealed a 108bp insert between the two exons (250bp) of the exon trap vector confirming the original pseudoexon inclusion, as expected. WT, wild-type; bp, base pairs; *GHR*, Growth Hormone Receptor.

Sanger sequencing of the spliced product verified this prediction confirming that the novel variant activates an intronic cryptic donor site deep within intron 6 of the *GHR* and utilises the same dormant acceptor splice site as the originally described 6Ψ pseudoexon. The close proximity of this dormant splice acceptor site leads to misrecognition of this region as an exon by the spliceosome and its retention during the splicing process. Interestingly, the same dormant acceptor site is involved in the mis-splicing and inclusion of the original 6Ψ *GHR* pseudoexon (**Figure 3.8**).

Figure 3.8 Novel 6Ω *GHR* pseudoexon utilises same dormant acceptor site as the original 6Ψ *GHR* pseudoexon



A. The novel 6Ω *GHR* pseudoexon c.618+836T>G variant creates an AGGT splice donor site (red) downstream of the original *GHR* 6Ψ pseudoexon variant (c.618+792A>G) (green) which produces a CGGT splice donor site. The dormant intronic AGCC acceptor splice site involved in mis-splicing and inclusion of both pseudoexons are shown in purple. Dashed lines indicate interrupted intronic sequence. **B.** Schematic of the 6Ψ and novel 6Ω *GHR* pseudoexon inclusion events into the mRNA.

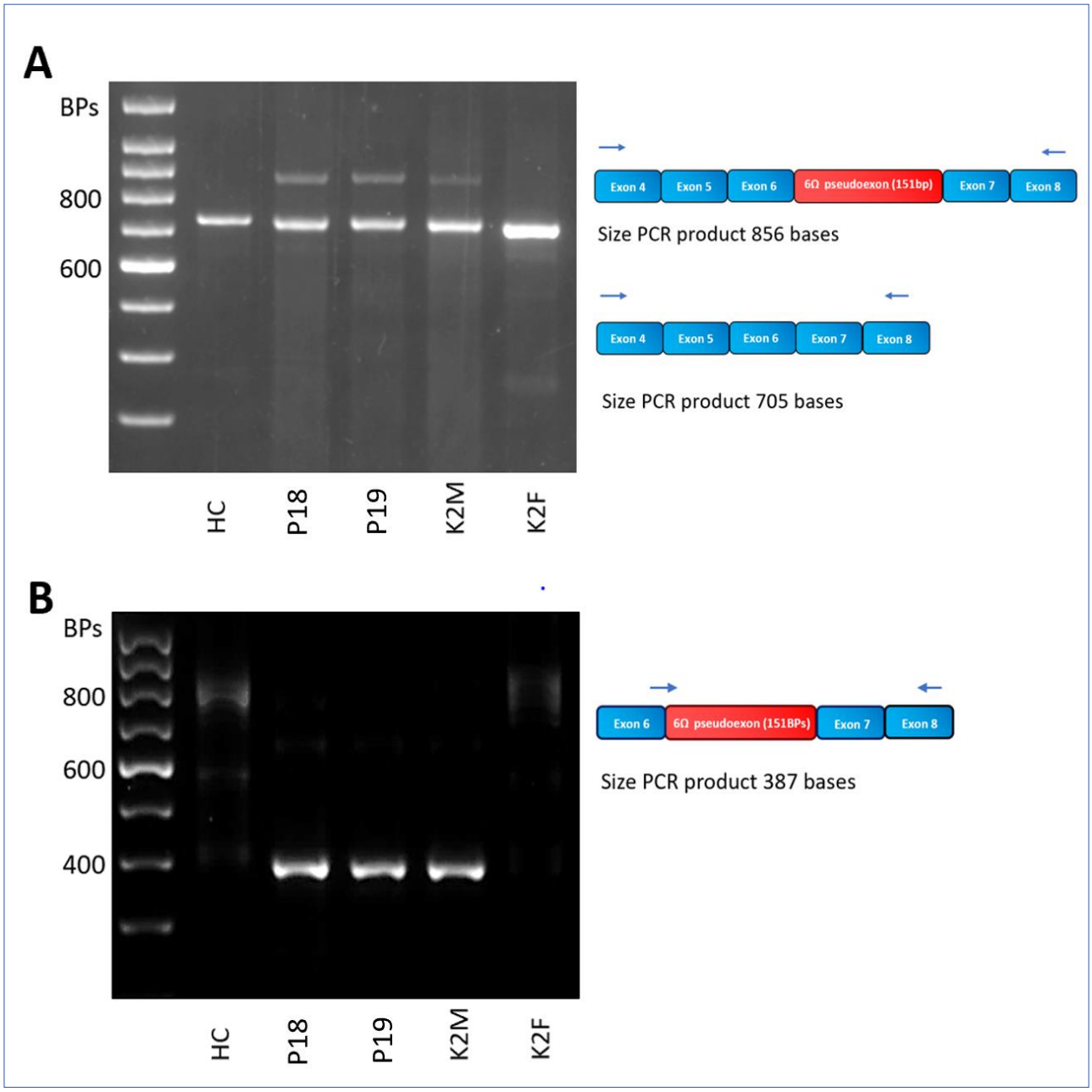
3.2.1.4.2 RT-PCR fibroblasts from Kindred 2

Patients 18 and 19 of the second kindred were compound heterozygous for the *GHR* 6Ω pseudoexon variant (c.618+836T>G) and another previously published nonsense point mutation in exon 4 of the *GHR* (c.181C>T; R43X) (74,210).

RNA samples derived from dermal fibroblasts of healthy control, Patients 18 and 19 and their parents were used to generate cDNA. RT-PCR was performed using primers for *GHR* cDNA Exon 4F (Forward) and *GHR* cDNA Exon 8R (Reverse) to amplify both wild-type *GHR* sequence and the 6Ω pseudoexon insertion. A 'normal' band of expected size (705bp) was seen in all the samples, and a larger (856bp) band was seen in Patients 18, 19 and their mother, who were all heterozygous for the *GHR* 6Ω variant (c.618+836T>G) (**Figure 3.9A**). This larger band corresponds to the retention of the additional 151 bases 6Ω pseudoexon.

RT-PCR was also performed using primers designed to amplify only sequences containing the 6Ω pseudoexon inclusion (GHR cDNA pseudo F1 (Forward) and GHR cDNA Exon 8R (Reverse)). A 387bp band was seen in Patients 18 and 19 and their mother, all of whom are heterozygous for the *GHR* 6Ω variant (c.618+836T>G) (**Figure 3.9B**). Sanger sequencing confirmed the inclusion of the 151 bases *GHR* 6Ω pseudoexon in keeping with the *in vitro* findings of the MoBiTec-Exontrap splicing assay.

Figure 3.9 Expression of wild-type and mutant transcripts in affected family members with the heterozygous c.618+836T>G *GHR* 6Ω variant



cDNA was prepared from dermal fibroblasts derived from a healthy control and Patients 18, 19 and both parents. A. Schematic showing the locations of the primers (GHR cDNA Exon 4F (Forward) and GHR cDNA Exon 8R (Reverse) (blue arrows)) used to amplify the region encompassing wild-type *GHR* the *GHR* 6Ω pseudoexon insertion. A 'normal' 705bp PCR product was seen in all the samples. An additional larger (856bp) PCR product was seen in Patient 18, Patient 19 and their mother, who were heterozygous for the c.618+836T>G *GHR* 6Ω variant, indicating the additional 151bp 6Ω pseudoexon insertion. B. Schematic showing the locations of the primers (GHR cDNA pseudo F1 (Forward) and GHR cDNA Exon 8R (Reverse) (blue arrows)). The forward primer at the junction of the 6Ω pseudoexon insertion means only sequences containing the *GHR* 6Ω pseudoexon insertion are amplified. The expected 387bp PCR product is seen in Patient 18, Patient 19 and their mother, all of whom are heterozygous for c.618+836T>G *GHR* variant. BPs, base pairs; HC, Healthy Control; P18, Patient 18; P19, Patient 19; K2M, Kindred 2 Mother; K2F Kindred 2 Father.

3.2.1.4.3 Predicted effect of novel 6Ω pseudoexon inclusion on *GHR* sequence and protein

The inclusion of this novel 151bp *GHR* 6Ω pseudoexon is predicted to lead to a frameshift and introduction of a premature stop codon after 245 amino acids (**Figures 3.10 and 3.11)**.

Figure 3.10 *GHR* genomic DNA sequence with the 151bp 6Ω pseudoexon inclusion between exons 6 and 7

ATGGATCTCTGGCAGCTGCTGTTGACCTTGGCACTGGCAGGATCAAGTGATGCTTTTTCTGGAAGTGAGGCCACAGCAGCTATCCTTAGCAGAGCACCCTGGAGT
CTGCAAAGTGTTAATCCAGGCCTAAAGACAAATTCTTCTAAGGAGCCTAAATTCACCAAGTGCCGTTCACCTGAGCGAGAGACTTTTTCATGCCACTGGACAGAT
GAGGTTCATCATGGTACAAAGAACCTAGGACCCATACAGCTGTTCTATACCAGAAGGAACACTCAAGAATGGACTCAAGAATGGAAAGAATGCCCTGATTATGTT
TCTGCTGGGGAAAACAGCTGTTACTTTAATTCATCGTTTACCTCCATCTGGATACCTTATTGTATCAAGCTAACTAGCAATGGTGGTACAGTGGATGAAAAGTGTTTC
TCTGTTGATGAAATAGTGCAACCAGATCCACCCATTGCCCTCAACTGGACTTTACTGAACGTCAGTTTAACTGGGATTCATGCAGATATCCAAGTGAGATGGGAAG
CACCACGCAATGCAGATATTCAGAAAGGATGGATGGTTCTGGAGTATGAACTTCAATACAAAGAAGTAAATGAAACTAAATGGAAAATGCCACAGCCACAGCCAC
AGCCAAGCCAAGCAGTTTCTGGCCACTCATCAGGTGATGCCCAGCAGCCTGGCACAGATCACTCCCAGAATTTTGAGACACCAGGACATTCAGTGAGCCACTGA
AAAAGATGCCAATTTTGTCATTAGAGGAAAGATGGACCCTATATTGACAACATCAGTTCCAGTGTACTCATTGAAAGTGGATAAGGAATATGAAGTGCGTGTGAGA
TCCAAACAACGAAACTCTGGAAATTATGGCGAGTTCAGTGAGGTGCTCTATGTAACACTTCCTCAGATGAGCCAATTTACATGTGAAGAAGATTTCTACTTTCCATG
GCTCTTAATTATTATCTTTGGAATATTTGGGCTAACAGTGATGCTATTTGTATTCTTATTTTCTAAACAGCAAAGGATTAAAATGCTGATTCTGCCCCCAGTTCCAGTTC
CAAAGATTAAAGGAATCGATCCAGATCTCCTCAAGGAAGGAAAATTAGAGGAGGTGAACACAATCTTAGCCATTCATGATAGCTATAAACCCGAATTCCACAGTG
ATGACTCTTGGGTTGAATTTATTGAGCTAGATATTGATGAGCCAGATGAAAAGACTGAGGAATCAGACACAGACAGACTTCTAAGCAGTGACCATGAGAAATCAC
ATAGTAACCTAGGGGTGAAGGATGGCGACTCTGGACGTACCAGCTGTTGTGAACCTGACATTCTGGAGACTGATTTCAATGCCAATGACATACATGAGGGTACCT
CAGAGGTTGCTCAGCCACAGAGGTTAAAAGGGGAAGCAGATCTCTTATGCCTTGACCAGAAGAATCAAAATAACTCACCTTATCATGATGCTTGCCCTGCTACTCA
GCAGCCCAGTGTTATCCAAGCAGAGAAAAACAAACCACAACCACTTCCTACTGAAGGAGCTGAGTCAACTCACCAAGCTGCCCATATTCAGCTAAGCAATCCAAG
TTCACTGTCAAACATCGACTTTTATGCCCAGGTGAGCGACATTACACCAGCAGGTAGTGTGGTCCTTTCCCCGGGCCAAAAGAATAAGGCAGGGATGTCCCAATG
TGACATGCACCCGGAAATGGTCTCACTCTGCCAAGAAAACTTCCTTATGGACAATGCCTACTTCTGTGAGGCAGATGCCAAAAAGTGCATCCCTGTGGCTCCTCAC
ATCAAGGTTGAATCACACATACAGCCAAGCTTAAACCAAGAGGACATTTACATCACCACAGAAAGCCTTACCACTGCTGCTGGGAGGCCTGGGACAGGAGAACA
TGTTCCAGGTTCTGAGATGCCTGTCCCAGACTATACCTCCATTCATATAGTACAGTCCCCACAGGGCCTCATACTCAATGCGACTGCCTTGCCCTTGCCTGACAAAG
AGTTTCTCTCATCATGTGGCTATGTGAGCACAGACCAACTGAACAAAATCATGCCTTAG

Colouring indicates coding regions of the *GHR:* Exon 2, Exon 3, Exon 4, Exon 5, Exon 6, Pseudoexon inclusion, Exon 7, Exon 8, Exon 9, Exon 10

Genomic DNA *GHR* sequence (Ensembl37 Genome Browser; http://grch37.ensembl.org/index.html). The 151bp pseudoexon sequence is shown in red.

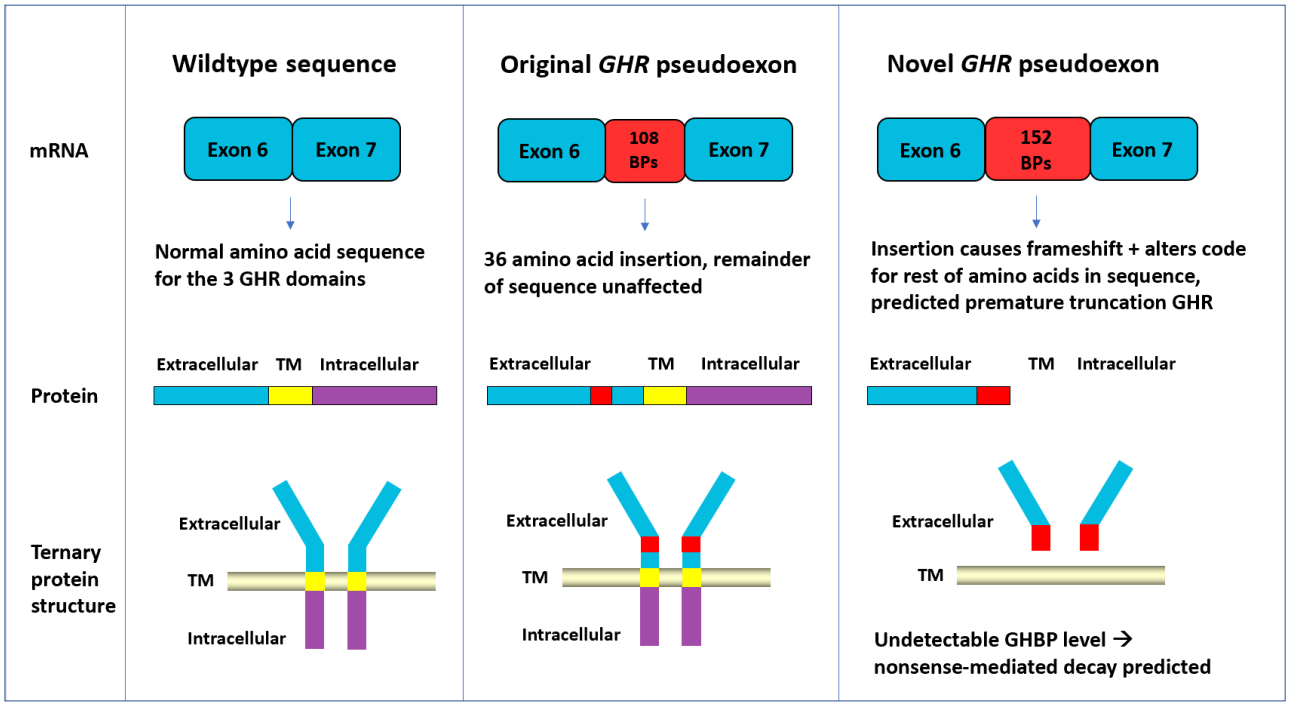
Figure 3.11 GH receptor amino acid sequence with 6Ω pseudoexon inclusion

MDLWQLLLTLALAGSSDAFSGSEATAAILSRAPWSLQSVNPGLKTNSSKEPKFTKCRSPERETFSCHWTD
EVHHGTKNLGPIQLFYTRRNTQEWTQEWKECPDYVSAGENSCYFNSSFTSIWIPYCIKLTSNGGTVDEKC
FSVDEIVQPDPPIALNWTLLNVSLTGIHADIQVRWEAPRNADIQKGWMVLEYELQYKEVNETKWKMPQ
PQPQPSQAVSGHSSGDAQQPGTDHSQNFETPGHSVSH*

Inclusion of the pseudoexon causes a frameshift and inclusion of 39 altered amino acids (red) into the *GHR* sequence followed by a premature stop codon. *, stop codon (Bioedit Sequence Alignment Editor (20)). *GHR* signal peptide is shown in blue. Total size of mutant GHR protein is 245 amino acids.

The resultant truncated protein is expected to be non-functional given it would lack both the transmembrane (encoded by exon 8; residues 265-288) and intracellular (encoded by exons 9 & 10; residues 289-638) domains of the GHR (**Figure 3.12**).

Figure 3.12 Predicted effect of the novel *GHR* 6Ω pseudoexon on protein structure



Schematic of the novel 6Ω *GHR* pseudoexon inclusion event and predicted GHR protein compared to wild-type sequence and 6Ψ pseudoexon. The 6Ω pseudoexon inclusion is predicted to cause a frameshift and result in premature truncation of the GHR lacking both transmembrane (TM) and intracellular domains.

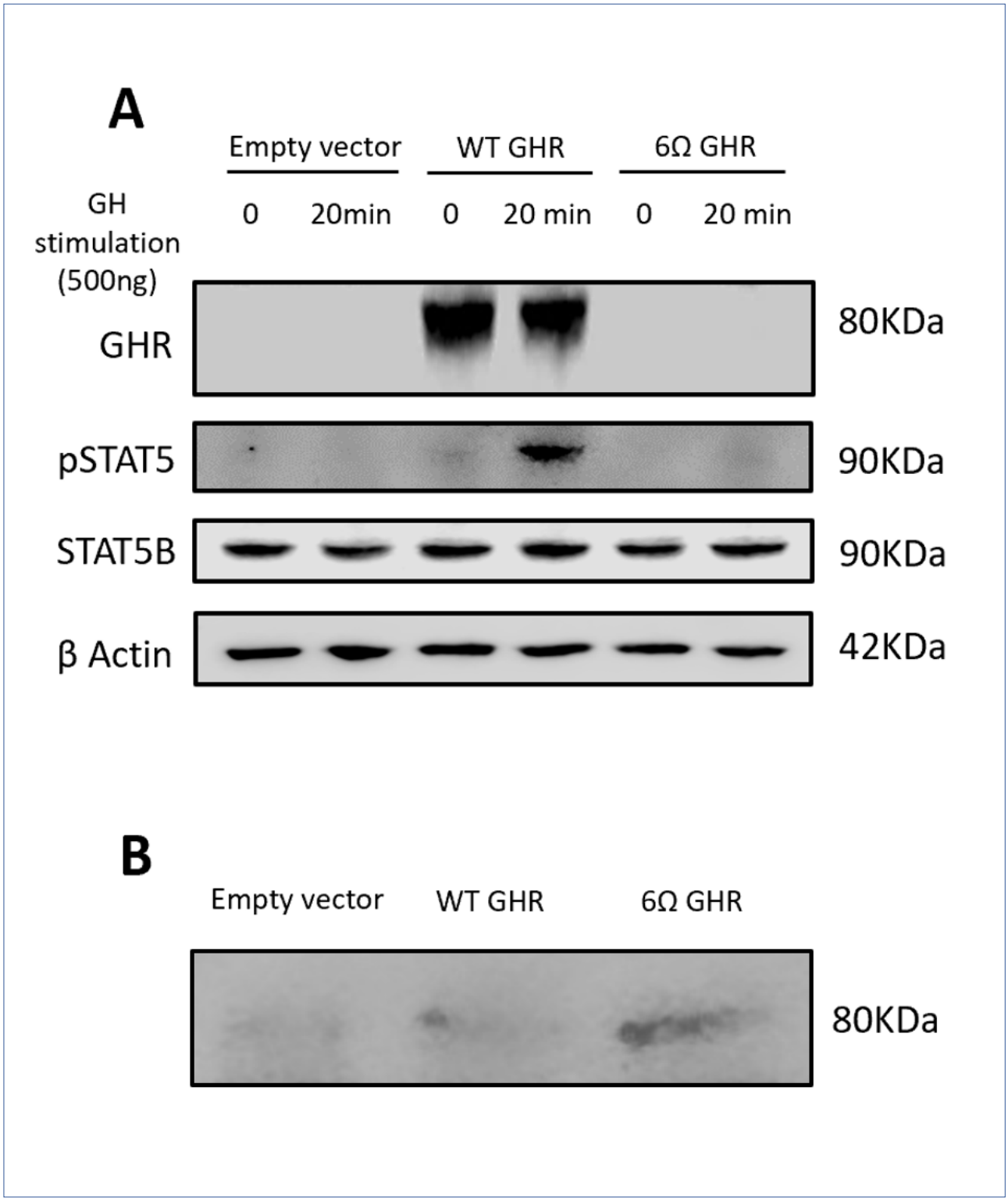
3.2.1.5 Functional assessment of *GHR* 6Ω pseudoexon

I created a *GHR* 6Ω pseudoexon construct using a 3-part Gibson assembly as previously described in methods **Section 2.9.1**. The impact of the *GHR* 6Ω pseudoexon on GHR signaling was assessed following growth hormone stimulation (500 ng) of both wild-type and 6Ω pseudoexon *GHR* constructs expressed in HEK293T cells. Tyrosine phosphorylation of STAT5B was used as a marker of intact GHR signalling. When compared to wild-type GHR, the 6Ω pseudoexon construct exhibited reduced phosphorylated-STAT5B following GH-stimulation (**Figure 3.13A**). As the truncated 6Ω GHR pseudoexon lacks both transmembrane and intracellular domains, it is unlikely to be able to anchor

onto the cell surface or dimerize, significantly impairing the activation of STAT5B and the downstream effects of growth hormone stimulation.

48 hours following transfection of the GHR 6Ω pseudoexon construct into HEK293T cells, the serum free conditioned media was probed using a GHBP antibody. This revealed extracellular accumulation of mutant (truncated) GHR in the GHR 6Ω pseudoexon transfected cells, present at much higher quantity than serum from WT GHR transfected cells (**Figure 3.13B**). The GHR 6Ω pseudoexon protein lacks both transmembrane and intracellular domains and would result in defective anchoring to the plasma membrane. The truncated protein is secreted extracellularly and recognized by the polyclonal GHBP antibody. Interestingly, biochemical assays using serum from all 3 patients revealed undetectable GHBP (**Table 3.4**). The GHBP assay relies on highly specific monoclonal antibodies and the GHR 6Ω pseudoexon protein lacks an epitope crucial for one of these monoclonal antibodies.

Figure 3.13 The novel *GHR* 6Ω pseudoexon diminishes GH-dependent STAT5B activation and accumulates extracellularly



A. Whole cell lysates from untreated or GH-stimulated (20 min) HEK293 cells transfected with pcDNA3.1 empty vector, wild-type (WT) *GHR* or 6Ω *GHR* mutant constructs. Representative immunoblots of three experiments are shown. **B.** Immunoblot analysis of conditioned media with anti-GHBP antibody from HEK293 cells transfected with the 6Ω *GHR* mutant construct showing extracellular accumulation of the truncated mutant 6Ω *GHR* protein. GH, Growth Hormone: WT, wild-type; B Actin, Beta Actin.

3.2.2 Potential *GHR* pseudoexon variant identified in Patient 6

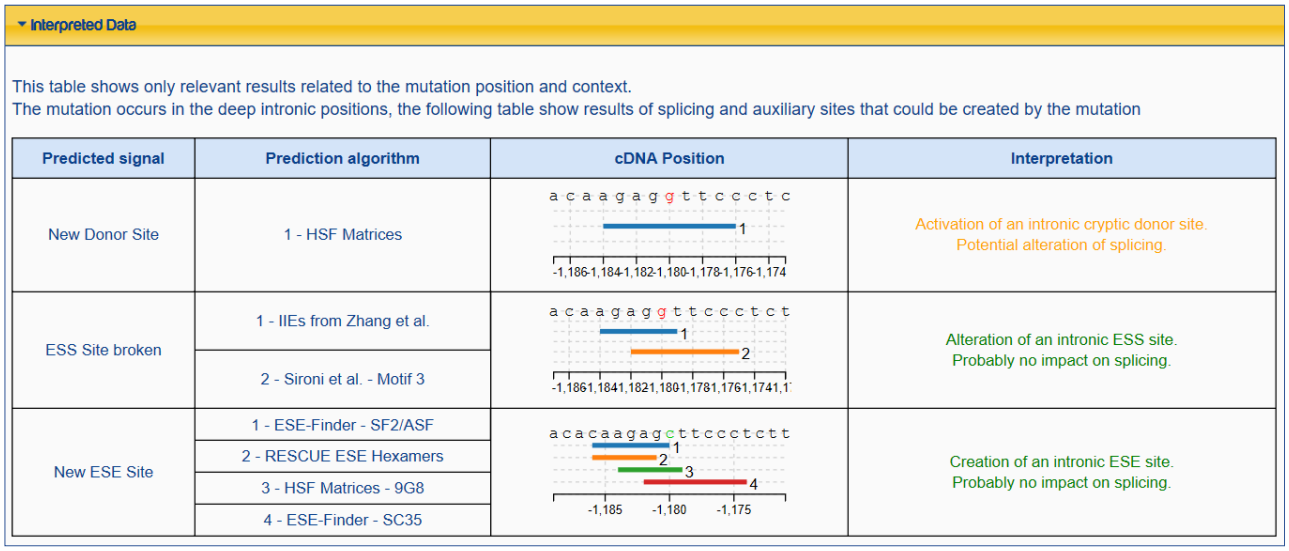
3.2.2.1 Patient phenotype

Patient 6 is a 12-year-old boy with phenotypic and biochemical features of Classic GHI. Born at term weighing 3kg (BW SDS -1.34), growth problems had been noted in early childhood and by 12.3 years his height SDS was -5.87. He had a depressed nasal bridge, mid-facial hypoplasia, frontal bossing and a high-pitched voice. Biochemically he had a high GH peak on stimulation >40ug/L with a baseline IGF-1 <25ng/ml and IGFBP3 <0.5mg/L. There was no family history of short stature and his parents were 3rd degree cousins.

3.2.2.2 Bioinformatic analysis and splicing prediction

Bioinformatic analysis identified no exonic variants in candidate genes that appeared pathogenic. However, upon interrogation of the non-coding material I identified a deep intronic homozygous change in *GHR*, 42598173, c.70+32137A>G. Ingenuity predicted the change to have a CADD score of 17.5, which is very unusual for a non-coding variant so deep within the intron. The gnomAD frequency was 0.058% with a gnomAD homozygous count of 0. This variant was predicted to activate an AGGT intronic cryptic donor site deep within the intron 2 of the GHR (**Figure 3.14**) so I hypothesised that this variant may lead to a novel *GHR* pseudoexon inclusion.

Figure 3.14 Splicing prediction generated by Human Splicing Finder for Patient 6 c.70+32137A>G *GHR* variant



Interpreted Data

This table shows only relevant results related to the mutation position and context.
The mutation occurs in the deep intronic positions, the following table show results of splicing and auxiliary sites that could be created by the mutation

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
New Donor Site	1 - HSF Matrices	a c a a g a g g t t c c c t c 1 -1,186 -1,184 -1,182 -1,180 -1,178 -1,176 -1,174	Activation of an intronic cryptic donor site. Potential alteration of splicing.
ESS Site broken	1 - IIEs from Zhang et al. 2 - Sironi et al. - Motif 3	a c a a g a g g t t c c c t c t 1 2 -1,1861 -1841 -1821 -1801 -1781 -1761 -1741 -1	Alteration of an intronic ESS site. Probably no impact on splicing.
New ESE Site	1 - ESE-Finder - SF2/ASF 2 - RESCUE ESE Hexamers 3 - HSF Matrices - 9G8 4 - ESE-Finder - SC35	a c a c a a g a g c t t c c c t c t t 1 2 3 4 -1,185 -1,180 -1,175	Creation of an intronic ESE site. Probably no impact on splicing.

Image generated from Human Splicing Finder (http://umd.be/HSF3/index.html). The novel c.70+32137A>G *GHR* variant was predicted to activate an AGGT intronic cryptic donor site.

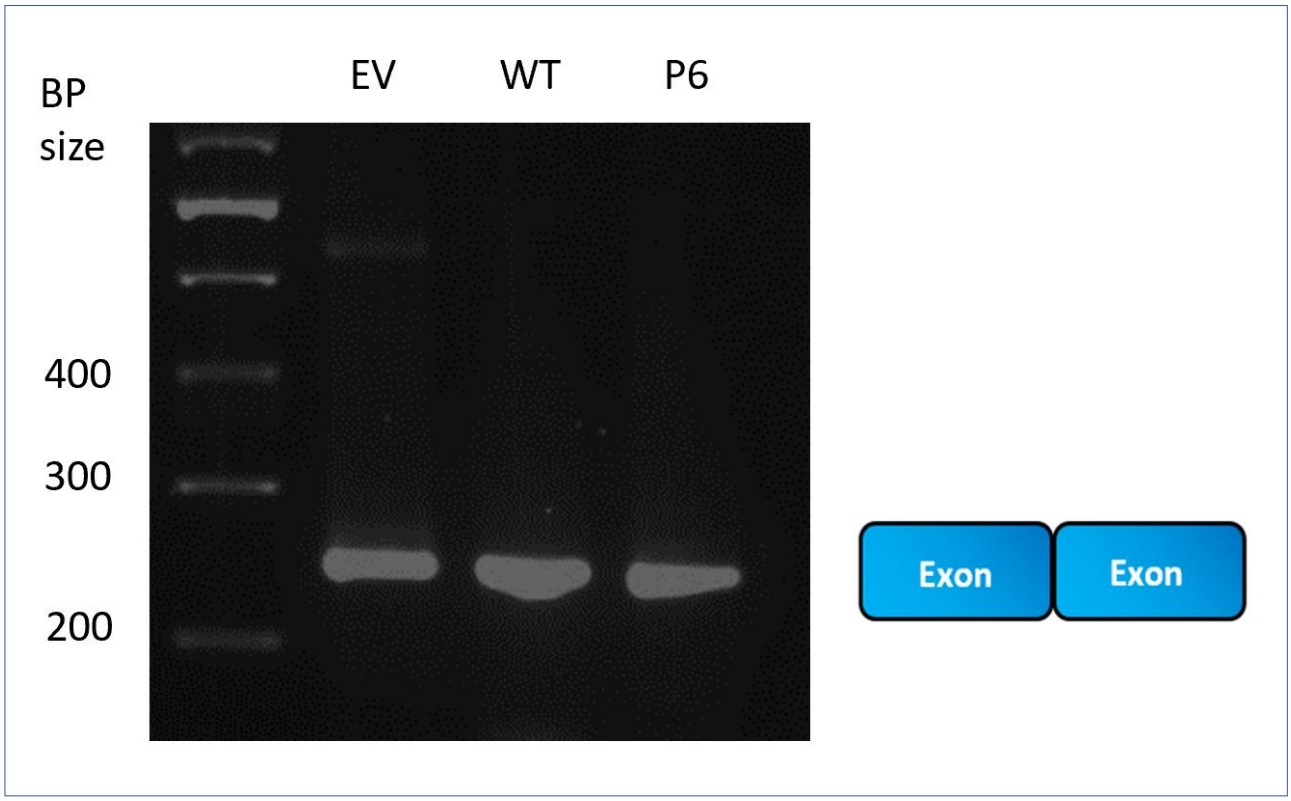
3.2.2.3 Segregation

Sanger sequencing confirmed Patient 6 to be homozygous for this novel c.70+32137A>G *GHR* variant and both parents to be heterozygous.

3.2.2.4 Splicing assay results

I designed primers to amplify an 800bp region including the variant of interest and assess *in vitro* splicing using the MoBiTec-Exontrap cloning vector pET01). Sanger sequencing of the spliced PCR product confirmed the same sequence for Patient 6 as for the wild-type sample (**Figure 3.15**). The assay showed no evidence of novel pseudoexon inclusion due to the c.70+32137A>G *GHR* variant.

Figure 3.15 Splicing assay results for Patient 6 c.70+32137A>G *GHR* variant

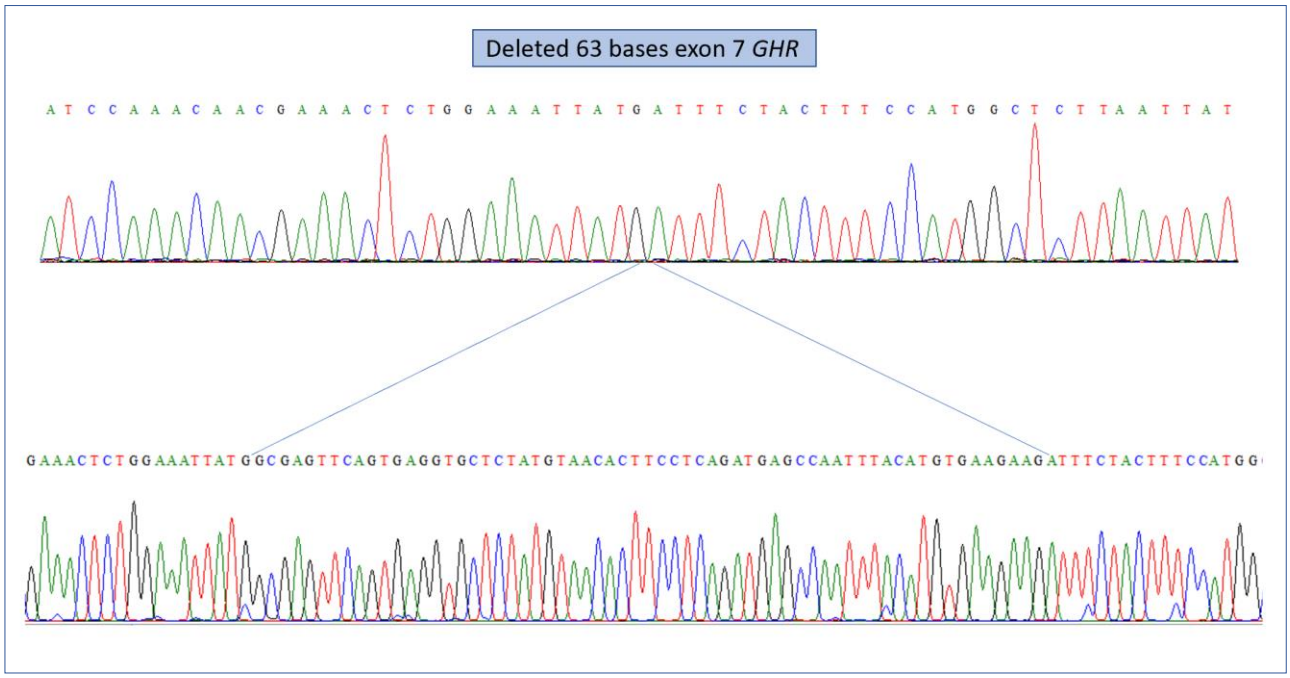


Gel electrophoresis of cDNA splicing products following the splicing assay using an exon trap vector (MoBiTec-Exontrap cloning vector pET01). EV, empty vector, pET01 alone; WT-*GHR*, pET01 with 800bp of wild-type *GHR* intron 2 sequence inserted; Patient 6, pET01 with 800bp of Patient 6 intron 2 sequence inserted (including the c.618+836T>G variant). The spliced products were amplified by PCR and visualized on a 2% agarose gel. In all 3 lanes a 250bp band is seen representing only the two exons of the exon trap vector. No pseudoexon inclusion is seen in the patient. This finding was confirmed by Sanger sequencing of the PCR products. BP, base pairs; EV, empty vector; WT, wild-type; P6, Patient 6; *GHR*, Growth Hormone Receptor. Splicing assay performed by MsC student Tasneem Ladha.

3.2.2.5 Analysis of Patient 6 fibroblast cDNA

Given that the splicing assay had not identified a diagnosis for Patient 6 and given his classic GHI phenotype and biochemistry, I assessed his *GHR* cDNA which I generated from patient fibroblasts. PCR of fibroblast cDNA revealed a 63 base pair deletion from exon 7 (**Figure 3.16**).

Figure 3.16 Deletion of 63 bases exon 7 *GHR* identified in Patient 6 cDNA



Sanger sequencing from Patient 6 cDNA (top) compared to healthy control cDNA (bottom) from cultured dermal fibroblasts. The 63-base deletion in exon 7 is visible in the Patient 6 sequence, demonstrated by the blue lines. Images were generated using ABI files from Sanger sequencing on Bioedit Sequence Alignment Editor (20)

This was a surprising result as I had not identified any variants of interest in exon 7 using my bioinformatics filtering pipelines on Ingenuity Variant Analysis. When I removed all applied filters to this dataset, such as read depth and allele frequency, there was still no listed variants in exon 7 for this patient. To determine whether this was a DNA deletion (copy number variant) or more likely a splicing problem, I consulted the original BAM file for this patient to see if there was a missing segment in the reads at this point. This would suggest they couldn't sequence across this region because it was missing in the genomic DNA.

3.2.2.6 BAM file identification of pathogenic variant

Upon examination of the Binary Alignment Map (BAM) file, it was clear that the reads covered all of exon 7 in good depth. However, I also noted a homozygous C>T change 2 base pairs after the start of deleted 63 base region (**Figure 3.17**).

Figure 3.17 c.723C>T *GHR* variant visualised on Patient 6 BAM file

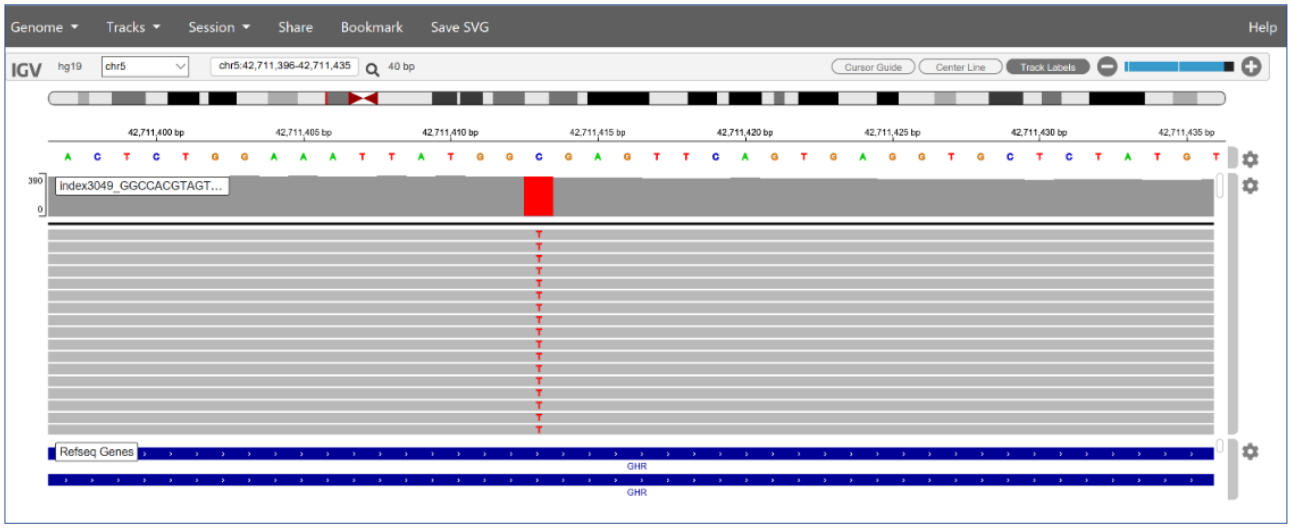


Figure created by taking screenshot whilst viewing the BAM file in Integrative Genomics Viewer (217). The C>T change is visible in all of the reads, demonstrated by the red 'T.'

This 42711413C>T, c.723C>T *GHR* variant was not called in the Variant Call File (VCF), likely in error by the Otogenetics Team (**Figure 3.18**).

Figure 3.18 Section of VCF for Patient 6 missing 42711413C>T variant

#CHROM	POS	ID	REF	ALT	QUAL	FILTER
chr5	42708728	rs376484521	TAAAC	T	5705.73	PASS
chr5	42708739	rs141011355	G	GA	4355.73	PASS
chr5	42709604	rs34382565	T	C	11624.8	PASS
chr5	42710148	rs151225774	TA	T	2971.73	PASS
chr5	42710487	rs5867599	T	TG	12392.7	PASS
chr5	42710504	rs7444761	A	C	12936.8	PASS
chr5	42711100	rs6873340	G	A	10212.8	PASS
chr5	42711737	rs35604733	G	A	5991.77	PASS
chr5	42711890	rs35291899	A	G	7864.77	PASS
chr5	42711940	rs10805665	G	A	8348.77	PASS
chr5	42712153	rs5025611	G	A	5840.77	PASS
chr5	42712324	rs75345822	A	G	8111.77	PASS
chr5	42713472	rs33972388	TG	T	11634.7	PASS
chr5	42714850	rs2972780	T	C	4370.77	PASS
chr5	42715499	rs2910878	A	G	11971.8	PASS
chr5	42716271	rs35197877	CA	C	2579.73	PASS
chr5	42716575	rs80293541	G	A	4232.77	PASS
chr5	42717529	rs6451641	G	A	4951.77	PASS

This figure contains a section from Patient 6's Variant Call File (VCF). The VCF lists all variants identified in this patient in order of chromosomal position prior to any bioinformatic filtering being applied. The 42711413C>T variant clearly visualised on the BAM file is not listed here.

This c.723C>T changes the sequence from CGG**C** to CGG**T**, creating a potential splice site. This base change was first reported in 1997 and predicted to create a cryptic donor splice site that would result in mis-splicing of exon 7 and deletion of 63bp (218). This hypothesis was later confirmed by functional analysis by RT-PCR of patient lymphocyte mRNA (219). My functional analysis is in keeping with these findings. Therefore, the homozygous c.723C>T *GHR* variant identified on the BAM file is likely to be the genetic change responsible for the GHI phenotype observed in this patient. I confirmed Patient 6 to be homozygous for this variant by Sanger sequencing.

3.2.3 *GHR* exon 9 dominant negative variants causing abnormal splicing

3.2.3.1 Patient phenotypes

Patients 7 and 8 (**Table 3.5**) were referred with moderate postnatal growth failure and did not have a phenotype of classic GHI discussed previously. Patient 7 was born small for gestational age with BW SDS -2.42 and by 16.5 years of age his height SDS was -3.2. Patient 8 had a BW SDS +0.18 but by 15 years of age has a height SDS of -2.7. Further details are shown in the table below.

Table 3.5 Phenotypic and biochemical details of Patients 7 and 8

Patient	Patient 7	Patient 8
Sex	Male	Male
Age (years)	16.5	15.0
Phenotype	Relative macrocephaly, disproportionate short stature. Borderline mesomelic shortening in the upper limb, slight lumbar hyperlordosis.	Nil dysmorphic features
Height SDS	-3.24	-2.7
BW SDS	-2.42	0.18
GH peak (ug/L)	NK	57.5
Baseline IGF-1 (ng/ml)	501	<25 (NR 220-972)
IGF-1 SDS	2.15	-3.03
GHBP (IFMA, pM)	1666 (NR 536-3634)	14567 (NR 536-3634)
GHBP (LIA, pM)	467 (NR 154-1073)	3366 (NR 154-1073)

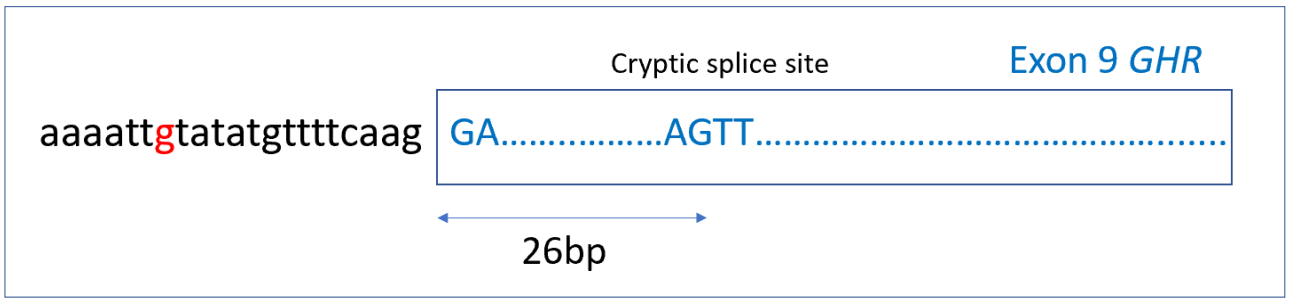
BW, birth weight; SDS, standard deviation score; GHBP, growth hormone binding protein; IFMA, immunofluorometric assay; LIA, ligand immunofunctional assay. GHBP levels were confirmed by serial dilutions.

3.2.3.2 Bioinformatic analysis and splicing prediction

Using my custom IVA filtering pipelines, I identified interesting heterozygous *GHR* variants in both patients predicted to affect splicing. Patient 7 had a variant in intron 8 which disrupted the polypyrimidine tract just prior to exon 9 (42718139T>G, c.876-15T>G). This mutation has a gnomAD frequency of 0.029% and was assigned a CADD score of <10.

As discussed previously, the polypyrimidine tract is a region rich in pyrimidine nucleotides just prior to the 3' end of the intron that promotes assembly of the spliceosome (140). Disruption of this region was predicted to reduce the efficacy of splicing at the canonical splice site by 20% by MaxEntScan score (215). A nearby cryptic splice site within exon 9 of the *GHR* may be recognised by the spliceosome and mis-splicing would result in a mutant exon 9, 26 bases smaller than wild-type. The proximity of these regions is demonstrated in **Figure 3.19.**

Figure 3.19 Location of c.876-15T>G *GHR* variant of interest in Patient 7

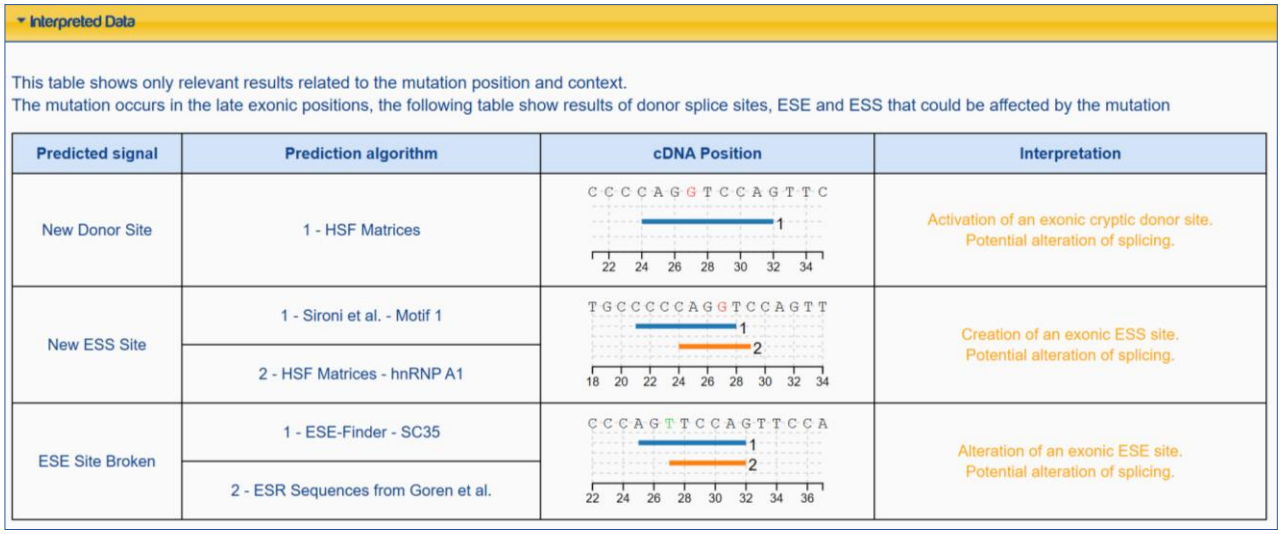


The c.876-15T>G intronic change identified in Patient 7 is highlighted in red. This variant is predicted to disrupt the polypyrimidine tract and reduce the efficacy of splicing at the canonical splice site. This may cause the spliceosome to instead recognise the nearby cryptic splice site highlighted, which would result in a mutant exon 9, 26 bases smaller than the wild-type exon 9.

The variant I identified in Patient 8 was in exon 9 of GHR (42718180T>G, c.902T>G) and predicted to activate an exonic cryptic donor splice site as shown in **Figure 3.20**. Misrecognition by the spliceosome of this donor site would result in a mutant exon 9, 26 bases smaller than the wild-type

exon 9. The proximity of these regions is shown in **Figure 3.21.** This variant was novel, not listed on gnomAD database and was assigned a CADD score of 27.7.

Figure 3.20 Splicing prediction generated by Human Splicing Finder for Patient 8 c.902T>G *GHR* variant



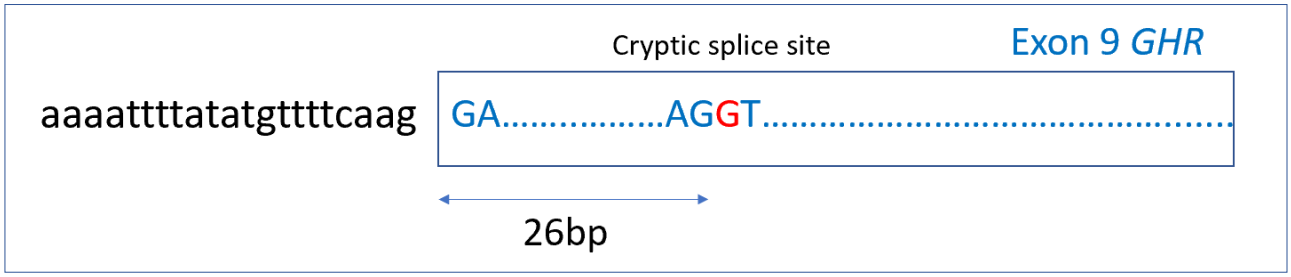
Interpreted Data

This table shows only relevant results related to the mutation position and context.
The mutation occurs in the late exonic positions, the following table show results of donor splice sites, ESE and ESS that could be affected by the mutation

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
New Donor Site	1 - HSF Matrices	C C C C A G G T C C A G T T C 1 22 24 26 28 30 32 34	Activation of an exonic cryptic donor site. Potential alteration of splicing.
New ESS Site	1 - Sironi et al. - Motif 1 2 - HSF Matrices - hnRNP A1	T G C C C C C A G G T C C A G T T 1 2 18 20 22 24 26 28 30 32 34	Creation of an exonic ESS site. Potential alteration of splicing.
ESE Site Broken	1 - ESE-Finder - SC35 2 - ESR Sequences from Goren et al.	C C C A G T T C C A G T T C C A 1 2 22 24 26 28 30 32 34 36	Alteration of an exonic ESE site. Potential alteration of splicing.

The c.902T>G *GHR* variant was predicted to activate an exonic cryptic donor site. Image generated from Human Splicing Finder (http://umd.be/HSF3/index.html)

Figure 3.21 Location of c.902T>G *GHR* variant of interest in Patient 8



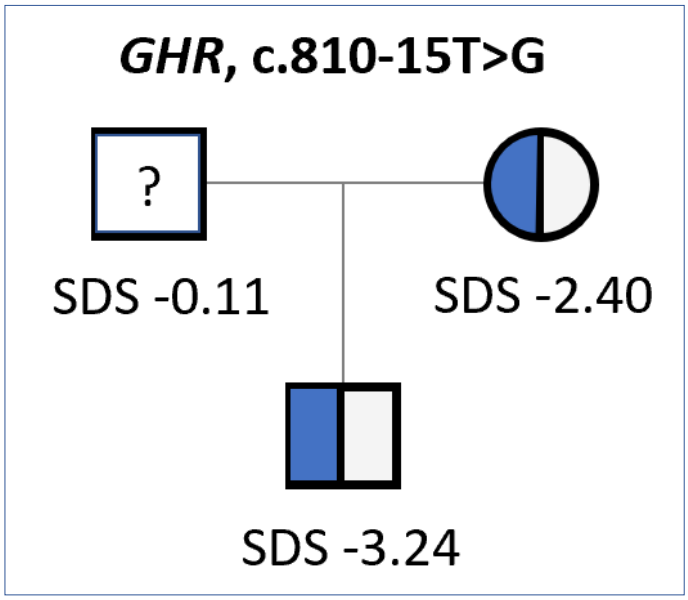
The c.902T>G *GHR* variant identified in Patient 8 is highlighted in red. This variant is predicted to activate the cryptic splice site within exon 9, leading to misrecognition by the spliceosome and creation of a mutant exon 9 26bp smaller than wild-type exon 9.

3.2.3.3 Segregation

Both patient variants were confirmed to be heterozygous by Sanger sequencing. The variant in Patient 7 was inherited from his short mother (height SDS -2.4) on Sanger sequencing (**Figure 3.22**).

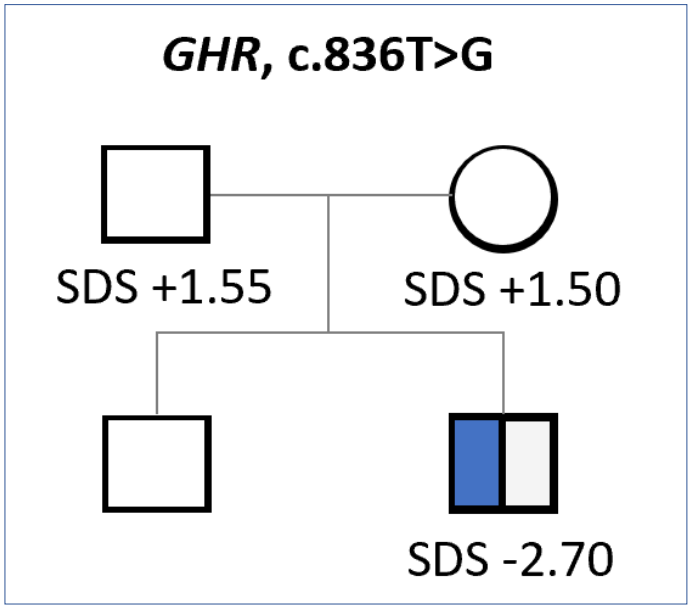
Patient 8 did not have any short relatives and his variant was found to be *de novo* on testing (**Figure 3.23**).

Figure 3.22 Segregation of c.810-15T>G GHR variant identified in Patient 7



Family tree showing both mum and Patient 7 heterozygous for the c.810-15T>G *GHR* variant with height SDS for each family member. DNA from Patient 7's father was not available for genetic testing.

Figure 3.23 Segregation of c.836T>G *GHR* variant identified in Patient 8

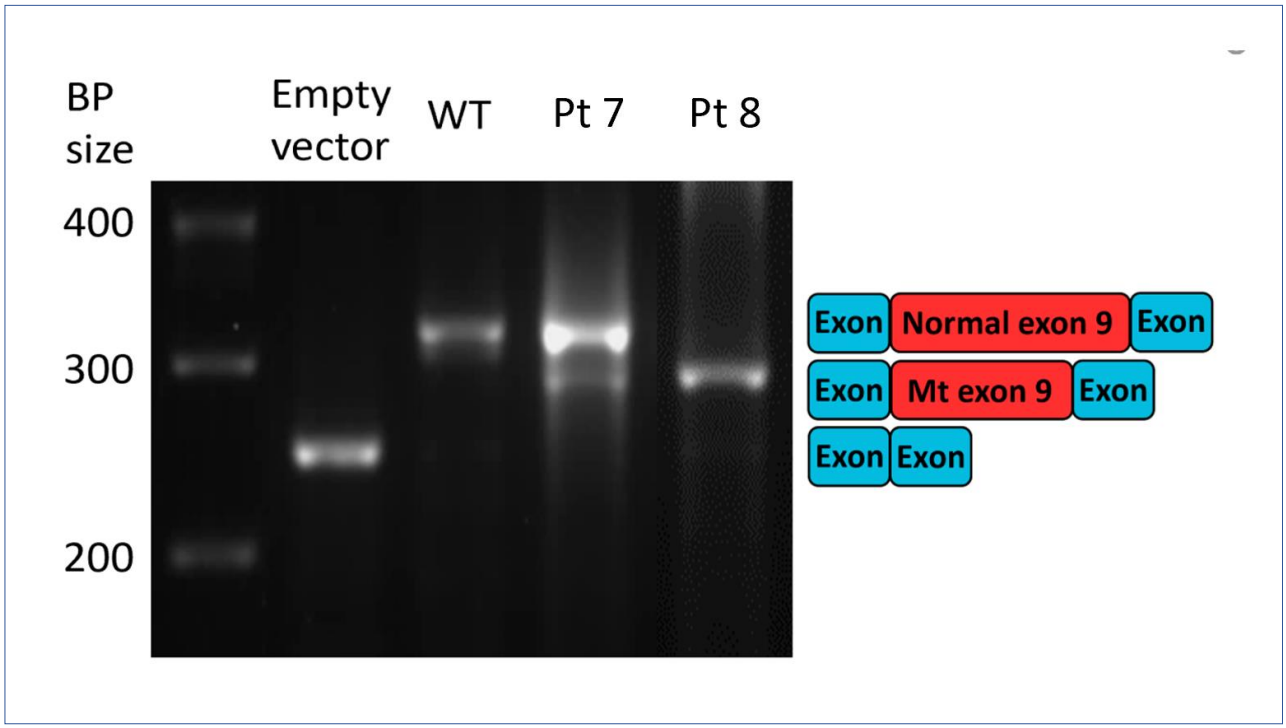


Family tree showing *de novo* c.836T>G *GHR* variant in Patient 8. Height SDS are shown for Patient 8 and his parents. Patient 8's sibling is described as 'normal height' but exact measurements not known.

3.2.3.4 Splicing assay results

To assess whether the variants identified in Patients 7 and 8 affected *GHR* splicing, an *in vitro* splicing assay as performed as described in Methods **Section 2.8**. I designed primers to amplify an 800bp DNA fragment of interest including exon 9 from both patients and a control sample. The PCR electrophoresis gel showed a smaller band in both patients (**Figure 3.24**). Sanger sequencing confirmed this to contain a mutant exon 9, 26bp smaller than wild-type. Wild-type exon 9 sequence was also detected in Patient 7, suggesting disruption to the polypyrimidine tract reduced the efficacy of canonical splicing but did not abolish it.

Figure 3.24 Gel electrophoresis of PCR cDNA splicing product



BP, base pair; WT, wild-type; Pt, Patient; Mt, Mutant. Splicing assay was performed using an exon trap vector (MoBiTec-Exontrap cloning vector pET01). Empty vector, pET01 alone; WT, pET01 with 800bp of WT GHR sequence including exon 9 GHR; Pt 7 and Pt 8, pET01 with 800bp of patient sequence inserted which included exon 9 and the respective missense mutations. The electrophoresis gel shows a 250bp band in the empty vector and a 320bp band in the WT sequence representing the two exons of the exon trap vector and normally spliced exon 9 which is 70bp. A smaller band was detected in the spliced products of both Patient 7 and 8, which sequencing confirmed to contain a mutant exon 9 26bp smaller than WT, leading to frameshift of the GHR. WT exon 9 sequence was also detected in Patient 7, explaining the two visible bands on the gel. 1 of 3 replicates of this assay was performed by MsC student Tasneem Ladha.

3.2.3.5 Effect of novel variants on *GHR* sequence

The variant identified in Patient 7 (c.876-15T>G) leads to abnormal splicing resulting in a deletion of the first 26 bases of exon 9 *GHR* (in addition to presence of wild-type sequence). The base sequence of *GHR* mutant transcript is shown below in **Figure 3.25** and the translated amino acid sequence shown in **Figure 3.26**.

Figure 3.25 *GHR* coding base sequence with deletion first 26 bases exon 9 (seen in Patient 7)

ATGGATCTCTGGCAGCTGCTGTTGACCTTGGCACTGGCAGGATCAAGTGATGCTTTTTCTGGAAGTGAGGCCACAGCAGCTATCCTTAGCAGAGCACCCTGGA
GTCTGCAAAGTGTTAATCCAGGCCTAAAGACAAATTCTTCTAAGGAGCCTAAATTCACCAAGTGCCGTTCACCTGAGCGAGAGACTTTTTCATGCCACTGGACA
GATGAGGTTCATCATGGTACAAAGAACCTAGGACCCATACAGCTGTTCTATACCAGAAGGAACACTCAAGAATGGACTCAAGAATGGAAAGAATGCCCTGATTA
TGTTTCTGCTGGGGAAAACAGCTGTTACTTTAATTCATCGTTTACCTCCATCTGGATACCTTATTGTATCAAGCTAACTAGCAATGGTGGTACAGTGGATGAAAAGT
GTTTCTCTGTTGATGAAATAGTGCAACCAGATCCACCCATTGCCCTCAACTGGACTTTACTGAACGTCAGTTTAACTGGGATTCATGCAGATATCCAAGTGAGATG
GGAAGCACCACGCAATGCAGATATTCAGAAAGGATGGATGGTTCTGGAGTATGAACTTCAATACAAAGAAGTAAATGAAACTAAATGGAAAATGATGGACCCTA
TATTGACAACATCAGTTCCAGTGTACTCATTGAAAGTGGATAAGGAATATGAAGTGCGTGTGAGATCCAAACAACGAAACTCTGGAAATTATGGCGAGTTCAGT
GAGGTGCTCTATGTAACACTTCCTCAGATGAGCCAATTTACATGTGAAGAAGATTTCTACTTTCCATGGCTCTTAATTATTATCTTTGGAATATTTGGGCTAACAGTG
ATGCTATTTGTATTCTTATTTTCTAAACAGCAAAGTTCCAGTTCCAAAGATTAAAGGAATCGATCCAGATCTCCTCAAGGAAGGAAAATTAGAGGAGGTGAACAC
AATCTTAGCCATTCATGATAGCTATAAACCCGAATTCCACAGTGATGACTCTTGGGTTGAATTTATTGAGCTAGATATTGATGAGCCAGATGAAAAGACTGAGGAA
TCAGACACAGACAGACTTCTAAGCAGTGACCATGAGAAATCACATAGTAACCTAGGGGTGAAGGATGGCGACTCTGGACGTACCAGCTGTTGTGAACCTGACA
TTCTGGAGACTGATTTCAATGCCAATGACATACATGAGGGTACCTCAGAGGTTGCTCAGCCACAGAGGTTAAAAGGGGAAGCAGATCTCTTATGCCTTGACCAG
AAGAATCAAAATAACTCACCTTATCATGATGCTTGCCCTGCTACTCAGCAGCCCAGTGTTATCCAAGCAGAGAAAAACAAACCACAACCACTTCCTACTGAAGGA
GCTGAGTCAACTCACCAAGCTGCCCATATTCAGCTAAGCAATCCAAGTTCACTGTCAAACATCGACTTTTATGCCCAGGTGAGCGACATTACACCAGCAGGTAGT
GTGGTCCTTTCCCCGGGCCAAAAGAATAAGGCAGGGATGTCCCAATGTGACATGCACCCGGAAATGGTCTCACTCTGCCAAGAAAACTTCCTTATGGACAATGC
CTACTTCTGTGAGGCAGATGCCAAAAAGTGCATCCCTGTGGCTCCTCACATCAAGGTTGAATCACACATACAGCCAAGCTTAAACCAAGAGGACATTTACATCAC
CACAGAAAGCCTTACCACTGCTGCTGGGAGGCCTGGGACAGGAGAACATGTTCCAGGTTCTGAGATGCCTGTCCCAGACTATACCTCCATTCATATAGTACAGTC
CCCACAGGGCCTCATACTCAATGCGACTGCCTTGCCCTTGCCTGACAAAGAGTTTCTCTCATCATGTGGCTATGTGAGCACAGACCAACTGAACAAAATCATGCC
TTAG

Colouring indicates coding regions of the *GHR*: Exon 2, Exon 3, Exon 4, Exon 5, Exon 6, Exon 7, Exon 8, Exon 9, Exon 10

This sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html) with the mutant exon 9 sequence interpreting using ABI files from Sanger sequencing of spliced products using the Bioedit Sequence Alignment Editor (20).

Figure 3.26 Translated GHR amino acid sequence with Patient 7 mutant exon 9

MDLWQLLLTLALAGSSDAFSGSEATAAILSRAPWSLQSVNPGLKTNSSKEPKFTKCRSPERETFSCHWTDEVHHGTKNLGPIQLFYTRRNTQEWTQ
EWKECPDYVSAGENSCYFNSSFTSIWIPYCIKLTSNGGTVDEKCFSVDEIVQPDPPIALNWTLLNVSLTGIHADIQVRWEAPRNADIQKGWMVLEY
ELQYKEVNETKWKMMDPILTTSVPVYSLKVDKEYEVRVRSKQRNSGNYGEFSEVLYVTLPQMSQFTCEEDFYFPWLLIIIFGIFGLTVMLFVFLFSKQ
QSSSSKD*RNRSRSPQGRKIRGGEHNLSHS**L*TRIPQ**LLG*IY*ARY**AR*KD*GIRHRQTSKQ*P*EIT**PRGEGWRLWTYQLL*T*HSG
D*FQCQ*HT*GYLRGCSATEVKRGSRSLMP*PEESK*LTLS*CLPCYSAAQCYPSREKQTTTTSY*RS*VNSPSCPYSAKQSKFTVKHRLLCPGERHY
TSR*CGPFPGPKE*GRDVPM*HAPGNGLTLPRKLPYGQCLLL*GRCQKVHPCGSSHQG*ITHTAKLKPRGHLHHHRKPYHCCWEAWDRRTCSR
F*DACPRLYLHSYSTVPTGPHTQCDCLALA*QRVSLIMWLCEHRPTEQNHALX

Inclusion of the mutant exon 9 causes 6 abnormal amino acids then a premature stop (*) shown in red

Total size of mutant GHR protein 247 amino acids

*, stop codon; GHR, Growth Hormone Receptor. This sequence was translated into amino acids using Bioedit Sequence Alignment Editor (20).

The variant identified in Patient 8 (c.902T>G) leads to abnormal splicing resulting in a deletion of the first 26 bases of exon 9 GHR. The base sequence of *GHR* mutant transcript is shown below in **Figure 3.27** and the translated amino acid sequence shown in **Figure 3.28**.

Figure 3.27 *GHR* coding base sequence with deletion first 26 bases exon 9 and T>G change at start of mutant exon 9 (seen in Patient 8)

ATGGATCTCTGGCAGCTGCTGTTGACCTTGGCACTGGCAGGATCAAGTGATGCTTTTTCTGGAAGTGAGGCCACAGCAGCTATCCTTAGCAGAGCACCCTGGA
GTCTGCAAAGTGTTAATCCAGGCCTAAAGACAAATTCTTCTAAGGAGCCTAAATTCACCAAGTGCCGTTCACCTGAGCGAGAGACTTTTTCATGCCACTGGACA
GATGAGGTTCATCATGGTACAAAGAACCTAGGACCCATACAGCTGTTCTATACCAGAAGGAACACTCAAGAATGGACTCAAGAATGGAAAGAATGCCCTGATTA
TGTTTCTGCTGGGGAAAACAGCTGTTACTTTAATTCATCGTTTACCTCCATCTGGATACCTTATTGTATCAAGCTAACTAGCAATGGTGGTACAGTGGATGAAAAGT
GTTTCTCTGTTGATGAAATAGTGCAACCAGATCCACCCATTGCCCTCAACTGGACTTTACTGAACGTCAGTTTAACTGGGATTCATGCAGATATCCAAGTGAGATG
GGAAGCACCACGCAATGCAGATATTCAGAAAGGATGGATGGTTCTGGAGTATGAACTTCAATACAAAGAAGTAAATGAAACTAAATGGAAAATGATGGACCCTA
TATTGACAACATCAGTTCCAGTGTACTCATTGAAAGTGGATAAGGAATATGAAGTGCGTGTGAGATCCAAACAACGAAACTCTGGAAATTATGGCGAGTTCAGT
GAGGTGCTCTATGTAACACTTCCTCAGATGAGCCAATTTACATGTGAAGAAGATTTCTACTTTCCATGGCTCTTAATTATTATCTTTGGAATATTTGGGCTAACAGTG
ATGCTATTTGTATTCTTATTTTCTAAACAGCAAAGGTCCAGTTCCAAAGATTAAAGGAATCGATCCAGATCTCCTCAAGGAAGGAAAATTAGAGGAGGTGAACAC
AATCTTAGCCATTCATGATAGCTATAAACCCGAATTCCACAGTGATGACTCTTGGGTTGAATTTATTGAGCTAGATATTGATGAGCCAGATGAAAAGACTGAGGAA
TCAGACACAGACAGACTTCTAAGCAGTGACCATGAGAAATCACATAGTAACCTAGGGGTGAAGGATGGCGACTCTGGACGTACCAGCTGTTGTGAACCTGACA
TTCTGGAGACTGATTTCAATGCCAATGACATACATGAGGGTACCTCAGAGGTTGCTCAGCCACAGAGGTTAAAAGGGGAAGCAGATCTCTTATGCCTTGACCAG
AAGAATCAAAATAACTCACCTTATCATGATGCTTGCCCTGCTACTCAGCAGCCCAGTGTTATCCAAGCAGAGAAAAACAAACCACAACCACTTCCTACTGAAGGA
GCTGAGTCAACTCACCAAGCTGCCCATATTCAGCTAAGCAATCCAAGTTCACTGTCAAACATCGACTTTTATGCCCAGGTGAGCGACATTACACCAGCAGGTAGT
GTGGTCCTTTCCCCGGGCCAAAAGAATAAGGCAGGGATGTCCCAATGTGACATGCACCCGGAAATGGTCTCACTCTGCCAAGAAAACTTCCTTATGGACAATGC
CTACTTCTGTGAGGCAGATGCCAAAAAGTGCATCCCTGTGGCTCCTCACATCAAGGTTGAATCACACATACAGCCAAGCTTAAACCAAGAGGACATTTACATCAC
CACAGAAAGCCTTACCACTGCTGCTGGGAGGCCTGGGACAGGAGAACATGTTCCAGGTTCTGAGATGCCTGTCCCAGACTATACCTCCATTCATATAGTACAGTC
CCCACAGGGCCTCATACTCAATGCGACTGCCTTGCCCTTGCCTGACAAAGAGTTTCTCTCATCATGTGGCTATGTGAGCACAGACCAACTGAACAAAATCATGCC
TTAG

Colouring indicates coding regions of the *GHR*: Exon 2, Exon 3, Exon 4, Exon 5, Exon 6, Exon 7, Exon 8, Exon 9, Exon 10

This sequence was generated using *GHR* sequence available on Ensembl37 Genome Browser (http://grch37.ensembl.org/index.html) with the mutant exon 9 sequence interpreting using ABI files from Sanger sequencing of spliced products using the Bioedit Sequence Alignment Editor (20).

Figure 3.28 Translated GHR amino acid sequence with Patient 8 mutant exon 9

MDLWQLLLTLALAGSSDAFSGSEATAAILSRAPWSLQSVNPGLKTNSSKEPKFTKCRSPERETFSCHWTDEVHHGTKNLGPIQLFYTRRNTQEWT
QEWKECPDYVSAGENSCYFNSSFTSIWIPYCIKLTSNGGTVDEKCFSVDEIVQPDPPIALNWTLLNVSLTGIHADIQVRWEAPRNADIQKGWMV
LEYELQYKEVNETKWKMMDPILTTSVPVYSLKVDKEYEVRVRSKQRNSGNYGEFSEVLYVTLPQMSQFTCEEDFYFPWLLIIIFGIFGLTVMLFVFL
FSKQQRSSSKD*RNRSRSPQGRKIRGGEHNLSHS**L*TRIPQ**LLG*IY*ARY**AR*KD*GIRHRQTSKQ*P*EIT**PRGEGWRLWTYQLL*T
*HSGD*FQCQ*HT*GYLRGCSATEVKRGSRSLMP*PEESK*LTLS*CLPCYSAAQCYPSREKQTTTTSY*RS*VNSPSCPYSAKQSKFTVKHRLLCP
GERHYTSR*CGPFPGPKE*GRDVPM*HAPGNGLTLPRKLPYGQCLLL*GRCQKVHPCGSSHQG*ITHTAKLKPRGHLHHHRKPYHCCWEAWD
RRTCSRF*DACPRLYLHSYSTVPTGPHTQCDCLALA*QRVSLIMWLCEHRPTEQNHALX

Inclusion of the mutant exon 9 causes 5 abnormal amino acids then a premature stop (*) shown in red. One normal amino acid (R) is included in the protein from the mutant exon 9 prior to this.

Total size of mutant GHR protein 247 amino acids

*, stop codon; GHR, Growth Hormone Receptor. This sequence was translated into amino acids using Bioedit Sequence Alignment Editor (20).

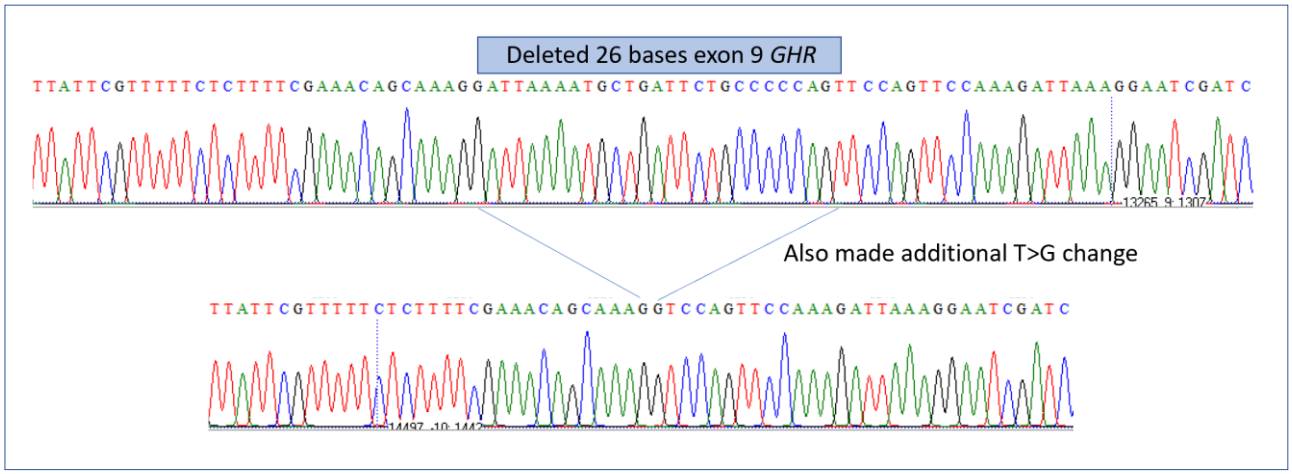
3.2.3.6 Creation of Patient 7 and 8 mutant *GHR* constructs by Gibson assembly

These mutant *GHR* constructs were created using Gibson assembly and custom primers designed to delete 26 bases from the GHR pcDNA1 vector and for Patient 8 to introduce the T>G change seen in the spliced product. The Sanger sequencing of the assembly products is shown in **Figure 3.29** and **Figure 3.30**. The entire sequence of the vector was assessed to ensure the products had reassembled correctly and no other changes had been introduced to the vector.

Figure 3.29 Deletion of 26 bases exon 9 using Gibson assembly to mimic Patient 7 mutant *GHR*

Sanger sequencing of Gibson assembly product demonstrating deletion of 26 bases of exon 9 to reflect mis spliced product seen with Patient 7. Images generated using ABI files from Sanger sequencing on Bioedit Sequence Alignment Editor (20)

Figure 3.30 Deletion of 26 bases exon 9 and introduction of T>G change using Gibson assembly to mimic Patient 8 mutant *GHR*

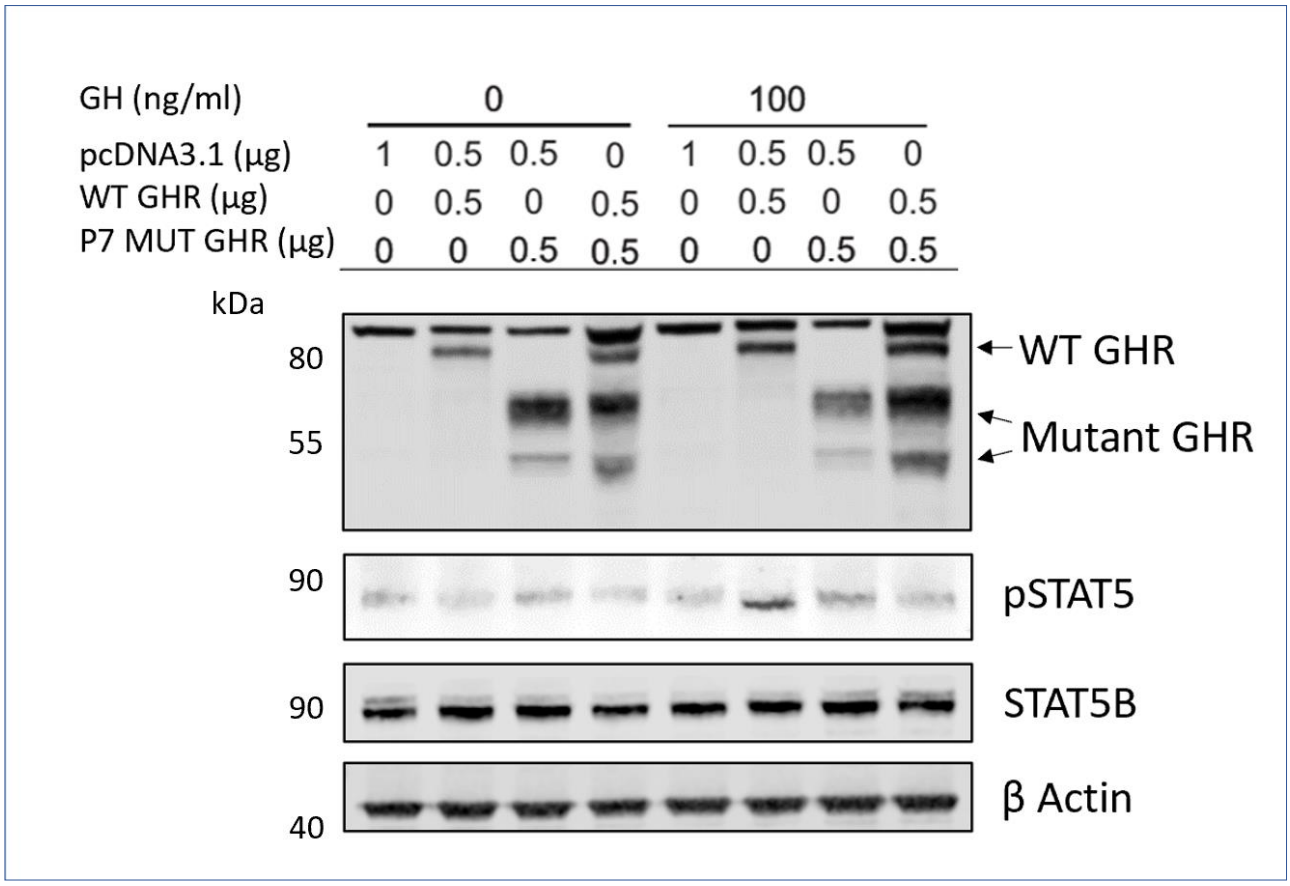


Sanger sequencing of Gibson assembly product demonstrating deletion of 26 bases of exon 9 and creation of T>G change to mis spliced product seen with Patient 8. Images generated using ABI files from Sanger sequencing on Bioedit Sequence Alignment Editor (20)

3.2.3.7 Functional assessment of mutant *GHR* constructs

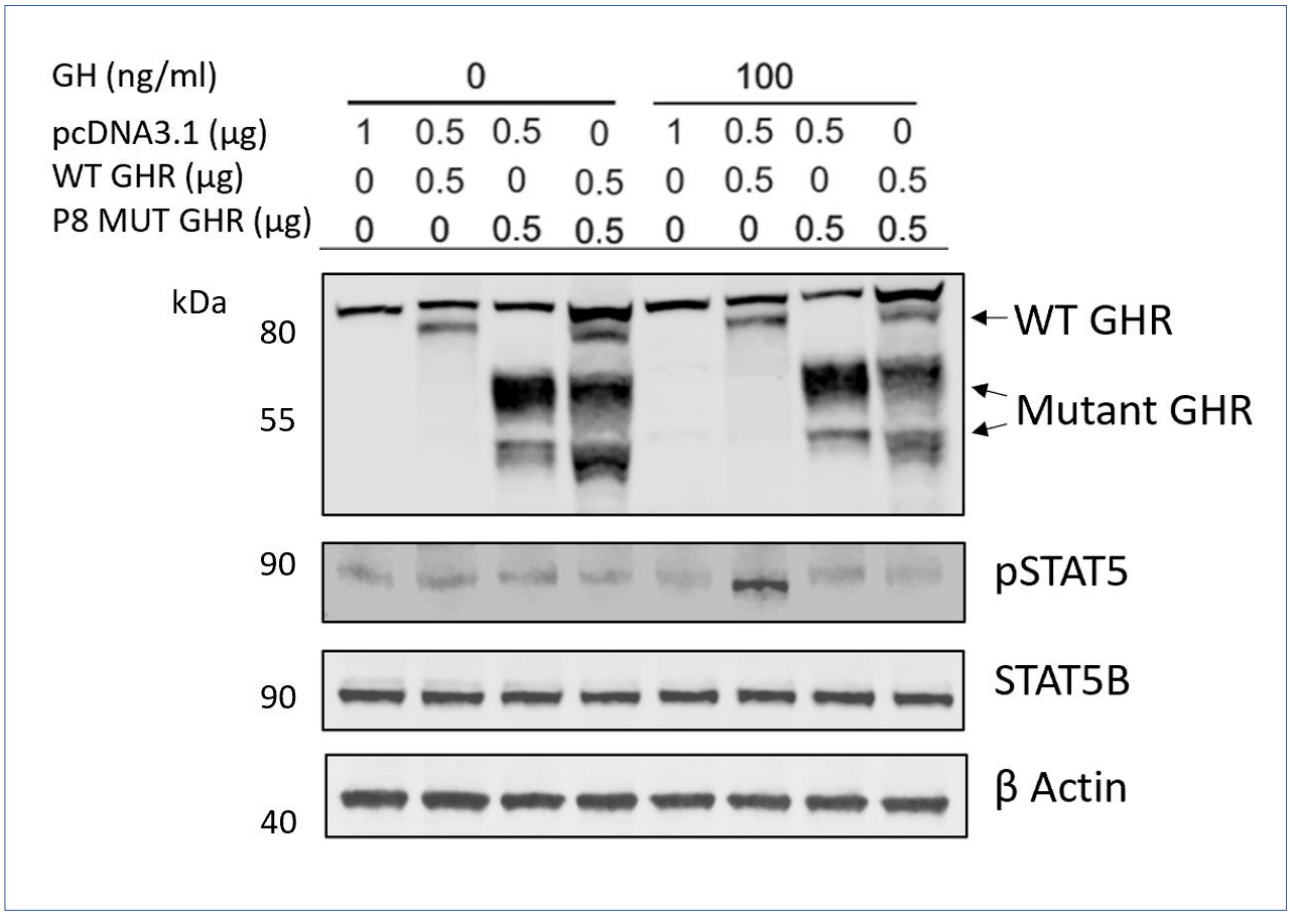
To determine if these mutants exert dominant negative effects on GH signalling, I assessed the effects on downstream signalling by assessing phosphorylation of STAT5, a key component of GH signalling. Western blot panels assessing the mutants are displayed below in **Figures 3.31** and **3.32**.

Figure 3.31 *GHR* c.876-15T>G exerts dominant negative effect on WT GHR signalling



GH, Growth Hormone; WT, wild-type; P7 MUT, Patient 7 mutant created by Gibson assembly. Whole cell lysates from transfected HEK293T cells, 20µg protein loaded per lane. The latter 4 lanes were growth hormone stimulated for 20 minutes at 100ng/ml prior to lysis. Wild-type GHR is seen around 80kDa and the mutant GHR bands are smaller. A non-specific band is seen for the GHR antibody in all lanes around 90kDa. Robust phosphorylation is seen upon GH stimulation of the wild-type GHR when alone but when co-transfected with the Patient 7 mutant, this phosphorylation is not seen. This demonstrates that the Patient 7 mutant exerts a dominant negative effect on signalling. Beta actin is used as a loading control.

Figure 3.32 *GHR* c.902T>G exerts dominant negative effect on WT GHR signalling



GH, Growth Hormone; WT, wild-type; P8 MUT, Patient 8 mutant GHR created by Gibson assembly. Whole cell lysates from transfected HEK293T cells, 20µg protein loaded per lane. The latter 4 lanes were growth hormone stimulated for 20 minutes at 100ng/ml prior to lysis.
Wild-type GHR is seen around 80kDa and the mutant GHR bands are smaller. A non-specific band is seen for the GHR antibody in all lanes around 90kDa. Robust phosphorylation is seen upon GH stimulation of the wild-type GHR when alone but when co-transfected with the Patient 8 mutant, this phosphorylation is not seen. This demonstrates that the Patient 8 mutant exerts a dominant negative effect on signalling. Beta actin is used as a loading control.

In contrast to the robust phosphorylation of STAT5 seen upon GH stimulation of the wild-type GHR construct, no such phosphorylation was seen upon GH stimulation of either mutant construct. Co-transfecting wild-type and mutant constructs impaired this phosphorylation, suggesting both mutants act in a dominant manner and impair downstream signalling of the GHR. This is in keeping with the patient phenotypes. Patients with dominant negative variants tend to have a milder phenotype than those with classic homozygous *GHR* mutations, thought to be due to the presence of 25% WT-WT *GHR* homodimers (62).

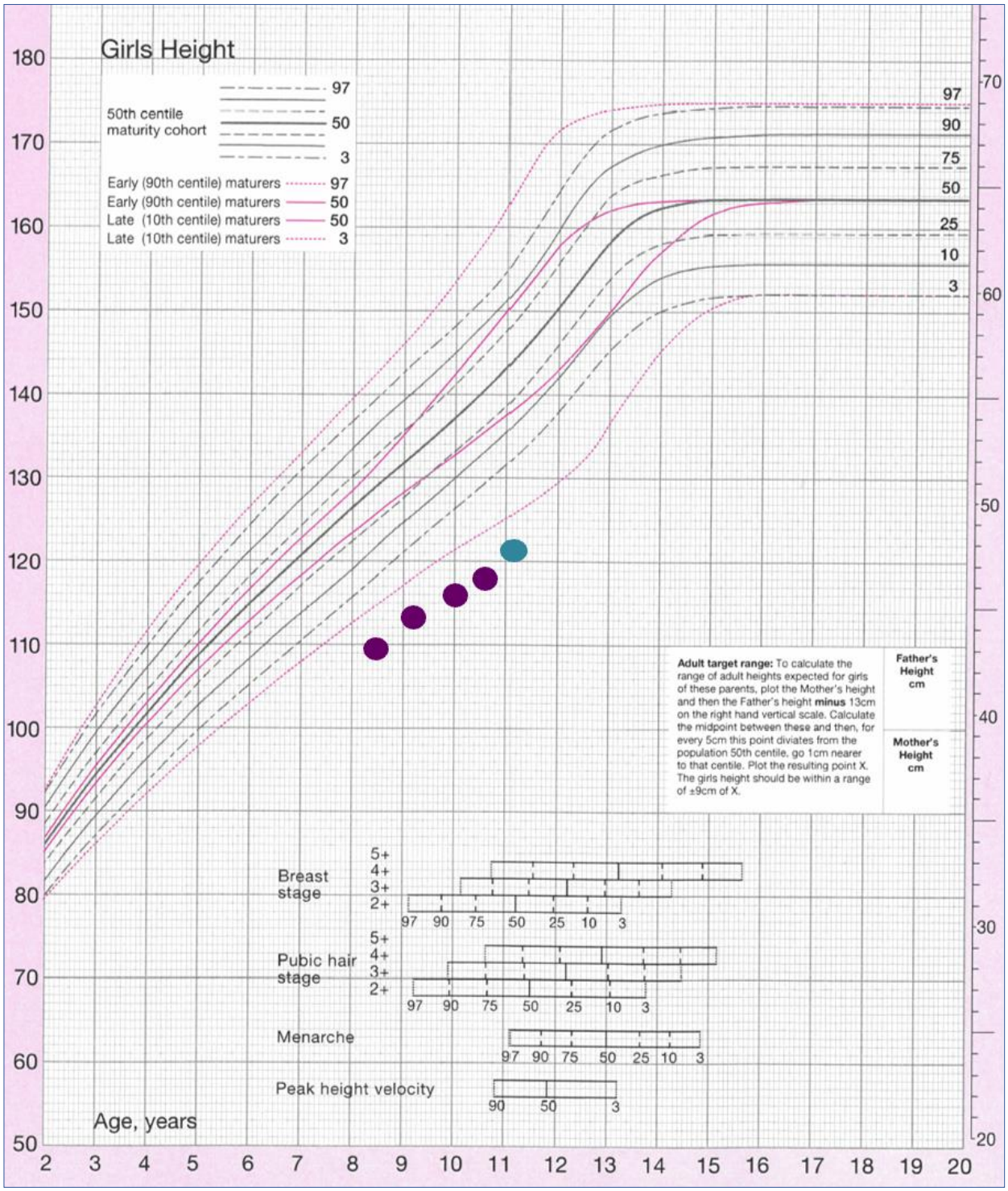
3.3 Novel mutation in *IGF1* hypothesised to impair IGF-1 cleavage

3.3.1 Patient phenotype

Patient 10 had a normal birth weight of 3.38kg (-0.2 SDS) but poor growth was noted from the first year of life. At presentation, she had a height of 117.2cm (-3.4 SDS) and head circumference of 49.5cm (-1.6 SDS) at 10.1 years of age. Her growth chart is below shown in **Figure 3.33**. Her bone age at 10.1 years was delayed by 2.5 years. She had a normal 46XX karyotype and a normal CGH array. Coeliac screen was negative, with thyroid function tests and cortisol within the normal ranges. On examination, she had no dysmorphic features. She had a low normal IQ and was described as a picky eater.

A high peak GH was observed upon glucagon stimulation (17.1mcg/L). Baseline IGF-1 levels were low/normal (144micrograms/L; -1.3 SDS) and responded poorly (increase <15micrograms/L) following IGF-1 generation testing. Given her poor response to IGF-1 generation testing, she was commenced on recombinant IGF-1 therapy. After 6 months therapy, serum IGF-1 levels were very high (1,044micrograms/L; SDS +5.9). Height velocity improved from 3.6cm/year to 5.9cm/year.

Figure 3.33 Growth chart for Patient 10



Growth chart for Patient 10. The blue spot indicates the height measurement taken once recombinant IGF-1 therapy had been commenced.

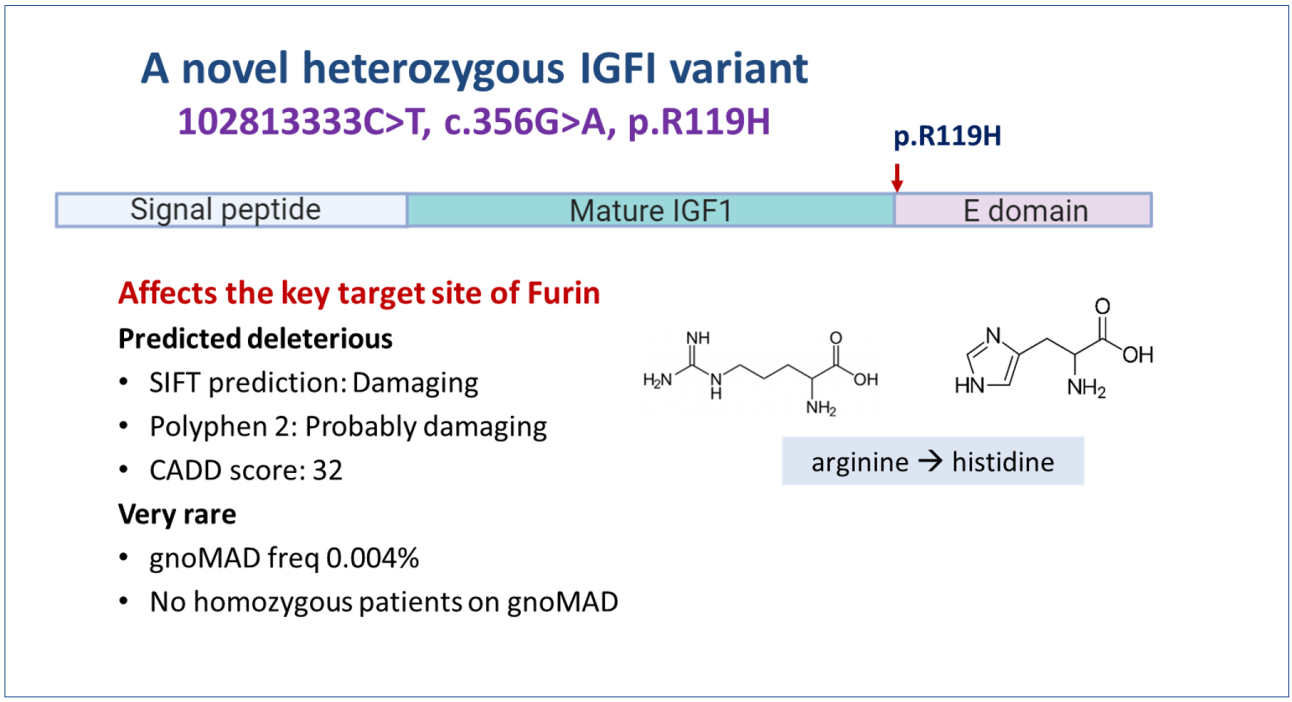
3.3.2 Variant details and functional prediction

Using custom bioinformatic pipelines that I created to analyse the short stature gene panel data, I identified a rare novel heterozygous *IGF1* variant (102813333C>T, c.356G>A, p.R119H; gnomAD frequency 0.004%) which was predicted damaging by SIFT; CADD score 32. This variant alters the first amino acid of IGF-1 E domain, the most critical residue for furin binding, which is highly conserved across species (**Figure 3.34**). This novel mutation changes arginine to histidine at this critical furin binding site. This is considerable change from a positively charged amino acid to a negatively charged amino acid with a very different structure. This novel *IGF1* variant is summarised in **Figure 3.35**.

Figure 3.34 Arginine at this base position is highly conserved across the species

Red box indicates position of highly conserved arginine, R. This amino acid is changed to histidine by the novel *IGF1* variant (102813333C>T, c.356G>A, p.R119H) identified in Patient 10. Figure created using Uniprot alignment, Align (uniprot.org).

Figure 3.35 Summary of novel *IGF1* variant identified in Patient 10



The location of the novel *IGF1* variant (102813333C>T, c.356G>A, p.R119H) identified in Patient 10 is indicated by the red arrow. The structural change in amino acid from arginine to histidine is demonstrated in the figures.

Furin cleaves pro-IGF-1 to mature, biologically active IGF-1 by removing the IGF-1 E domain. I hypothesised that if furin was unable to bind to the pro-IGF-1 due to the p.R119H amino acid change at the key binding site, this may result in impaired IGF-1 cleavage and thus reduced circulating mature IGF-1 levels. As the IGF-1 assay used for serum IGF-1 is an ELISA which recognises an epitope common to both mature and pro-IGF-1, the IGF-1 serum measurement would include both forms. As the pro-IGF-1 would be expected to be less biologically active, this would lead to a functional IGF-1 deficiency and explain the patient's poor postnatal growth despite low-normal IGF-1 levels.

Patients reported in the literature with heterozygous *IGF1* defects show postnatal growth failure and mild learning difficulties, with or without microcephaly and delayed bone age. These reported patients are compared to Patient 10 in **Table 3.6** below.

Table 3.6 Phenotype and biochemistry of reported heterozygous *IGF1* gene mutations compared to Patient 10

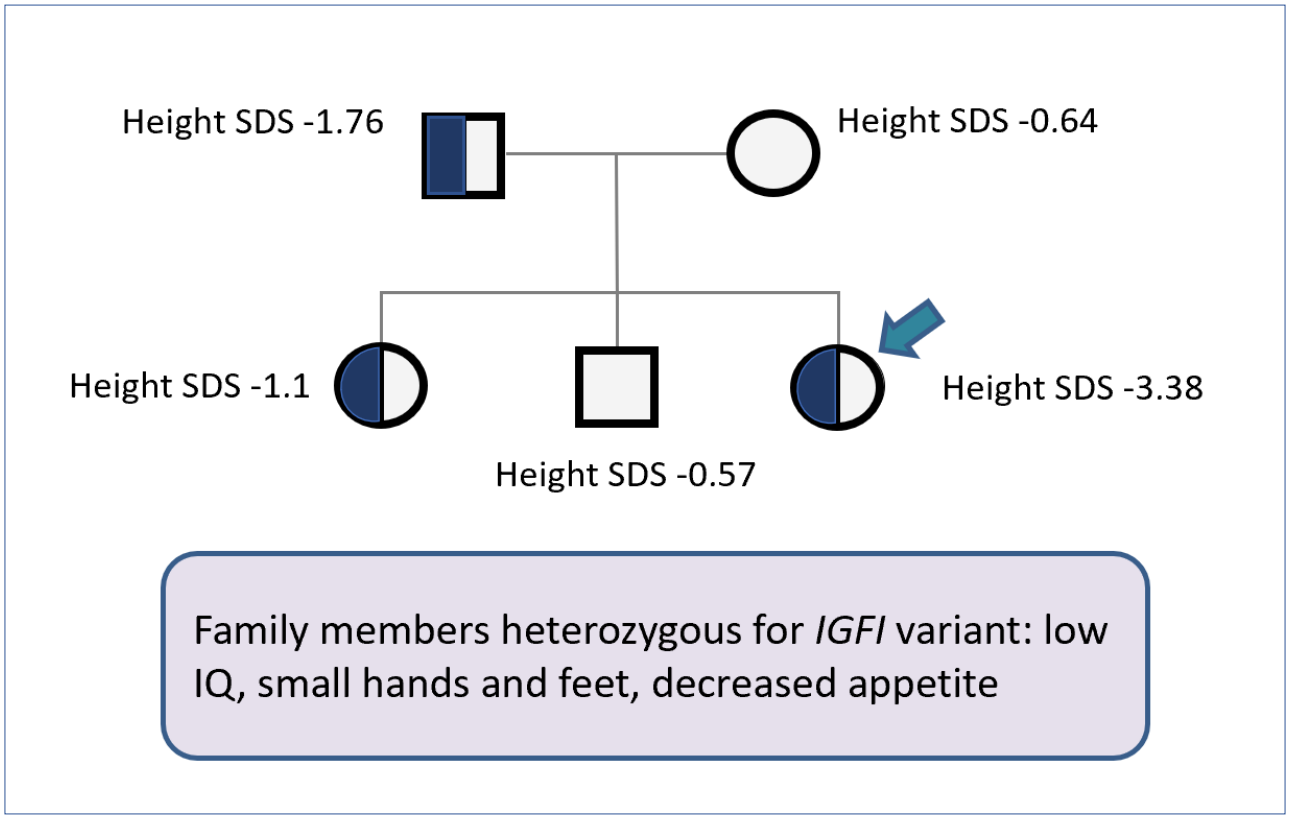
Measures	**Batey, 2014** (102)	**Fuqua, 2012** (103)	**Van Duyvenvoorde, 2010, sibling 1** (95)	**Van Duyvenvoorde, 2010, sibling 2** (95)	**Patient 10**
***IGF1* variant**	Heterozygous *IGF1* gene deletion (262KB deletion including whole IGFI gene)	Heterozygous exon 4 splicing excision → frameshift and early stop, c.402+1G>C, p.N74Rfs*8	Heterozygous 4bp deletion → frameshift and early stop, c.243-246dupCAGC, p.S83Qfs*13	Heterozygous 4bp deletion → frameshift and early stop, c.243-246dupCAGC, p.S83Qfs*13	Heterozygous missense c.356G>A, p.R119H
Birth weight SDS	-1.5	-1.5	-2.9	-1.2	-0.15
Height SDS	-2.7	-4.2	-4.1	-4.6	-3.38
HC SDS	-3.4	No microcephaly	-2.4	-1.6	-1.6
Serum IGF-I	Low-normal	Low-normal	Low	Low	Low-normal
ALS	High-normal	High-normal	Normal	Normal	Low
IGFBP3	Normal	Normal	Normal	Normal	Normal
Bone age	Normal	Normal	Delayed	Delayed	Delayed

SDS, standard deviation score; HC, head circumference.

3.3.3 Segregation

The novel heterozygous *IGF1* variant (102813333C>T, c.356G>A, p.R119H) identified in Patient 10 was confirmed by Sanger sequencing and found to be inherited from her father. Her younger sister also shared the variant. The family tree is shown in **Figure 3.36**. Interestingly, these family members all had low IQ, small hands and feet and described themselves as picky eaters. The mother and son reported higher school achievement and good appetites. The father of the kindred recalled how he had very notably been the shortest in his class at school but towards his late teenage years he had had some degree of catch-up growth which resulted in him being a similar height to his shorter friends. He reported normal pubertal timing.

Figure 3.36 Segregation for the novel heterozygous c.356G>A *IGF1* variant

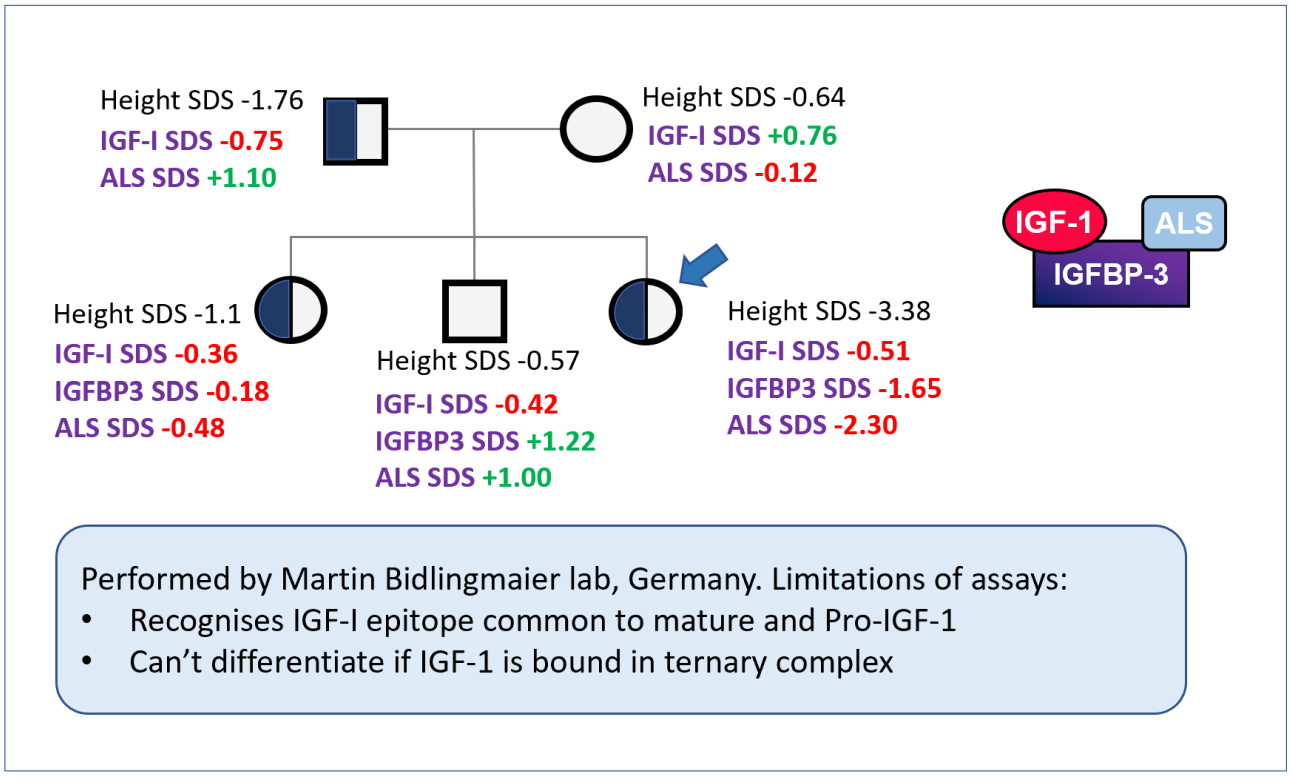


Segregation for the novel heterozygous *IGF1* variant (102813333C>T, c.356G>A, p.R119H). Patient 10 is indicated by a blue arrow. The variant was inherited from her father and her sister was also heterozygous for the variant. Her mother and younger brother were wild-type. Interestingly the family members who were heterozygous for the variant had several similar features, described in the purple box.

3.3.4 Biochemical analysis of family members

Biochemical assays were performed by Martin Bidlingmaier's lab and are described in detail in methods **Section 2.20**. These assessed serum levels of the components of the IGF-1 ternary complex and the results are shown below in **Figure 3.37**. As with the ELISA techniques used in most laboratories that analyse serum IGF-1, this IGF-1 assay recognises an epitope common to both pro-IGF-1 and mature IGF-1. The assays are also unable to differentiate if the IGF-1 is bound in the ternary complex.

Figure 3.37 Expert biochemical analysis for Patient 10 kindred

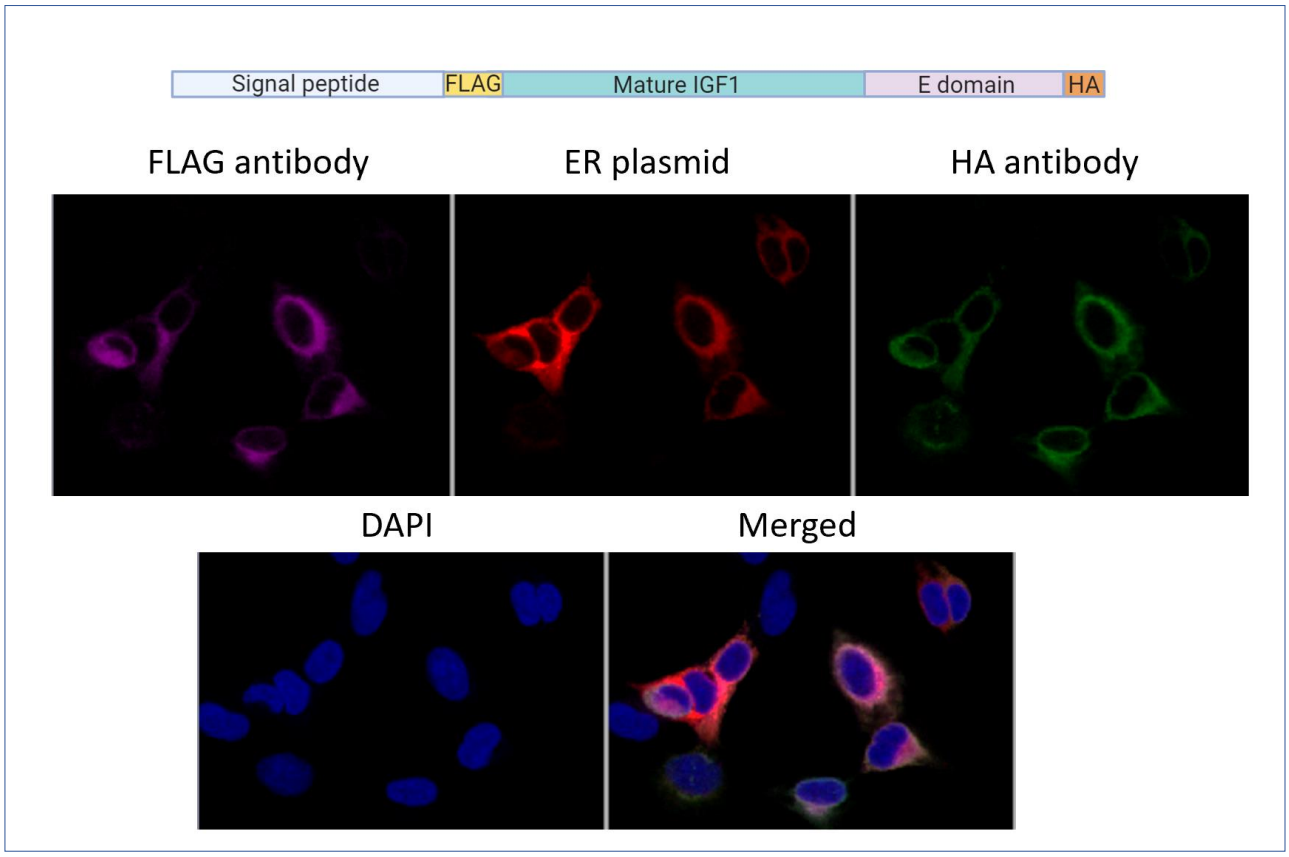


ALS, Acid-labile subunit. There are no accepted normal reference ranges for adult IGFBP3 levels so it was not possible to calculate an SDS value for these. The mother's IGFBP3 level was 4062ng/ml and the father's 5608ng/ml. Limitations of the assays are highlighted in the blue box.

3.3.5 Immunocytochemistry of FLAG and HA tagged IGF-1 constructs

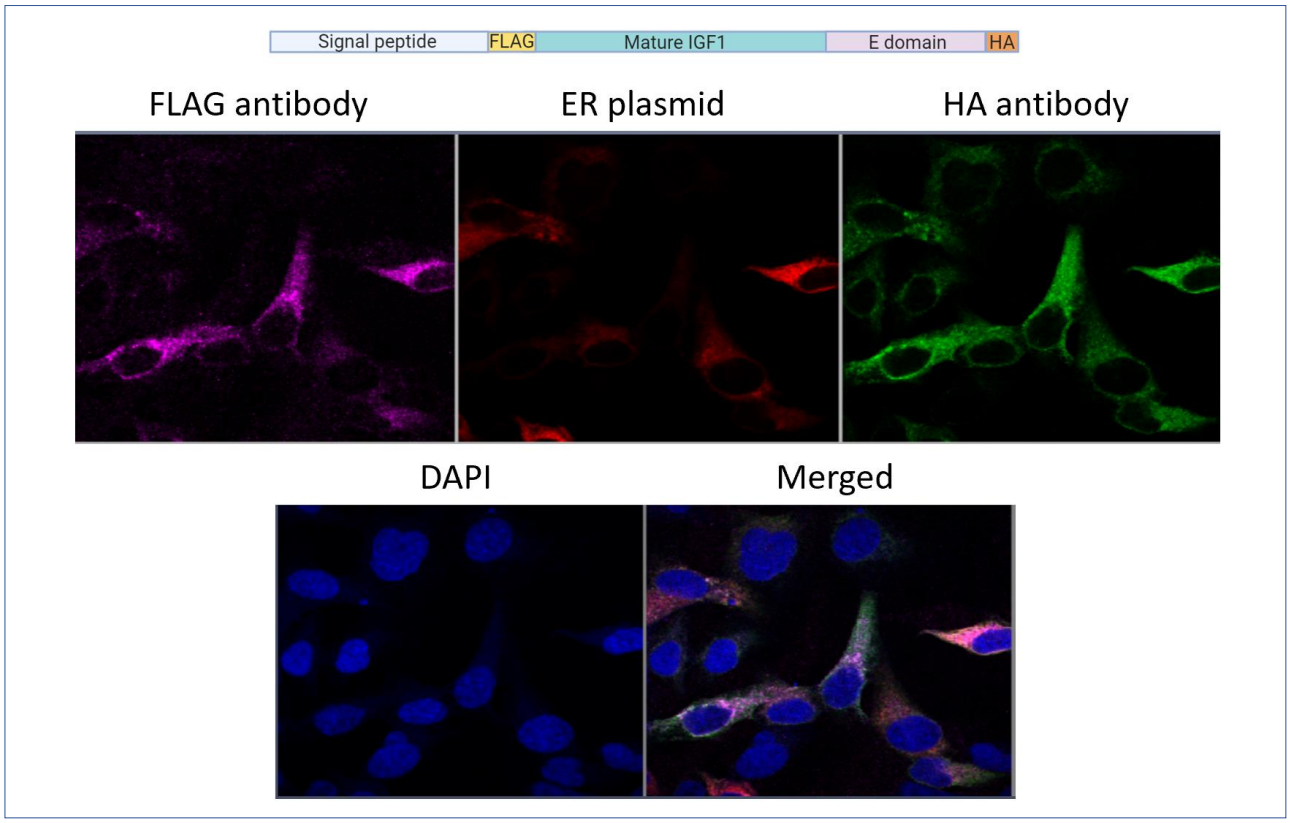
The signal peptide of the IGF-1 facilitates transport of the molecule to the endoplasmic reticulum and into the secretory pathway. I used immunocytochemistry with FLAG and HA antibodies to assess if this novel IGF-1 mutation affected localisation of the IGF-1 protein. The details of these *IGF1* vectors are provided in methods **Section 2.16**. With this experiment I used a plasmid known to localise to the endoplasmic reticulum to aid visualisation of this region. I transfected the wild-type and mutant constructs into HEK293T cells and assessed them using confocal microscopy. Further details of my immunocytochemistry can be found in methods **Section 2.17**. Given that I was using FLAG and HA antibodies, this experiment should detect only transfected constructs rather than endogenous and thus in the HEK293T cells transfected with the mutant should represent only the mutant IGF-1 localisation. My results for the wild-type and mutant IGF-1 constructs are shown in **Figure 3.38** and **Figure 3.39** respectively.

Figure 3.38 Immunocytochemistry for wild-type IGF-1 protein



Representative image of the custom *IGF-1* vector is shown at top of the figure to aid visualisation. DAPI stain highlights the nucleus. ER plasmid localises to the endoplasmic reticulum. These images show that the FLAG and HA tagged wild-type IGF-1 protein localises to the endoplasmic reticulum. This is how IGF-1 enters the secretory pathway.

Figure 3.39 Immunocytochemistry for mutant c.356G>A IGF-1 protein



Representative image of the custom *IGF-1* vector is shown at top of the figure to aid visualisation. DAPI stain highlights the nucleus. ER plasmid localises to the endoplasmic reticulum. These images show that the FLAG and HA tagged mutant IGF-1 protein localises to the endoplasmic reticulum. This is how IGF-1 enters the secretory pathway.

My results showed that both wild-type and mutant IGF-1 constructs were able to translocate to the nucleus. Both FLAG and HA signals were present in the region of the ER suggesting the presence of both the mature IGF1 and E domain, although it is not possible to know if they were joined as Pro-IGF-1 or present as cleaved mature IGF-1 and separate E-domain.

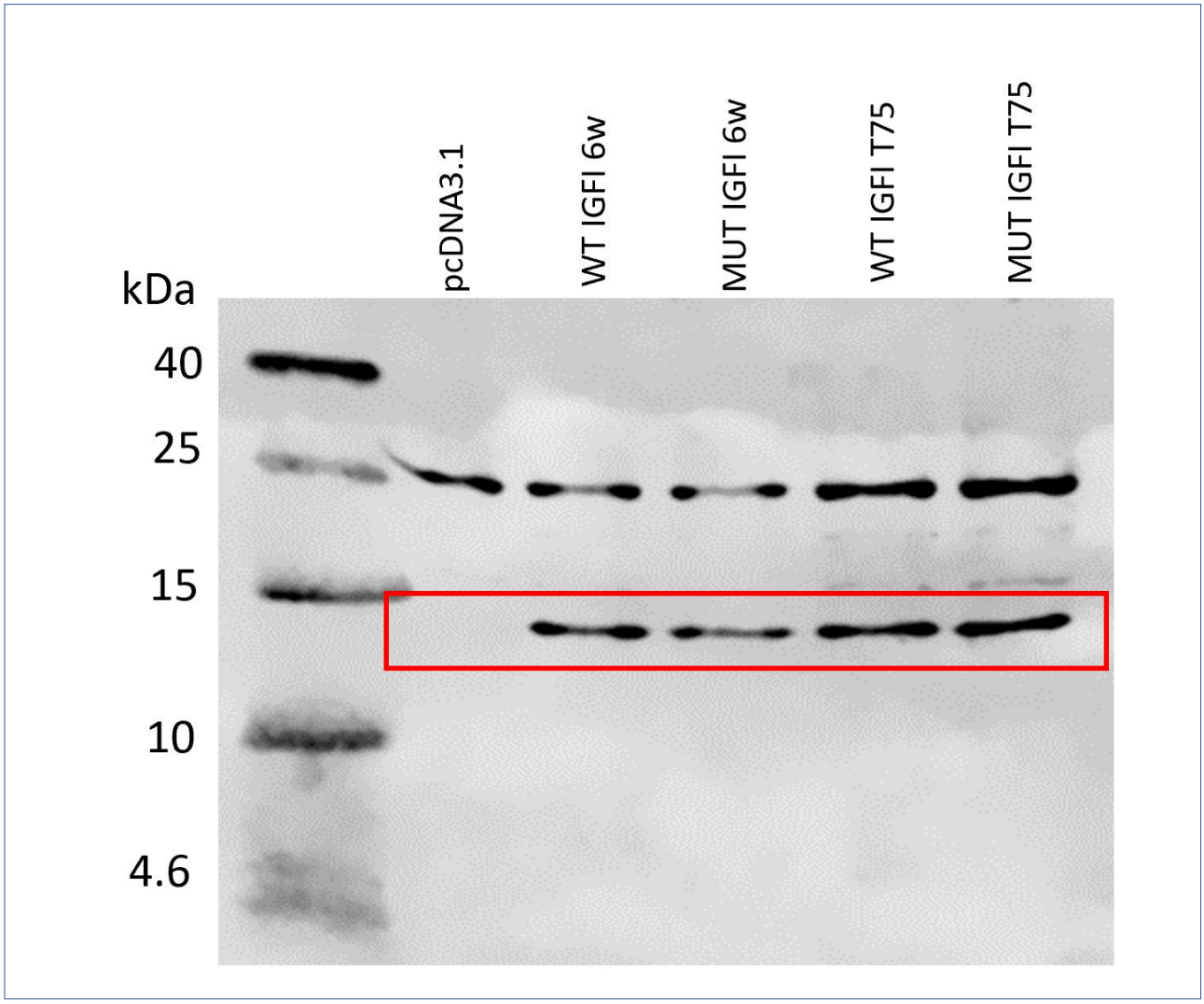
3.3.6 Furin assay with Strep II and HA tagged IGF-1 constructs

3.3.6.1 Isolating streptomycin II tagged IGF-1 constructs for the furin assay

I began by assessing the best method to generate large quantities of whole cell lysates containing the streptomycin tagged vector, as these would be the input for my purification using StrepTrap columns. I transfected HEK293T cells with my wild-type and mutant IGF-1 constructs in both 6 well

and T75 cell culture flasks and loaded equal protein volumes to all lanes. The expected size of the streptomycin tagged IGF-1 protein (proIGF-1 plus streptomycin and his tag) was 14KDa. The western blot is shown below in **Figure 3.40.**

Figure 3.40 Western blot showing streptomycin signal from transfected whole cell lysates



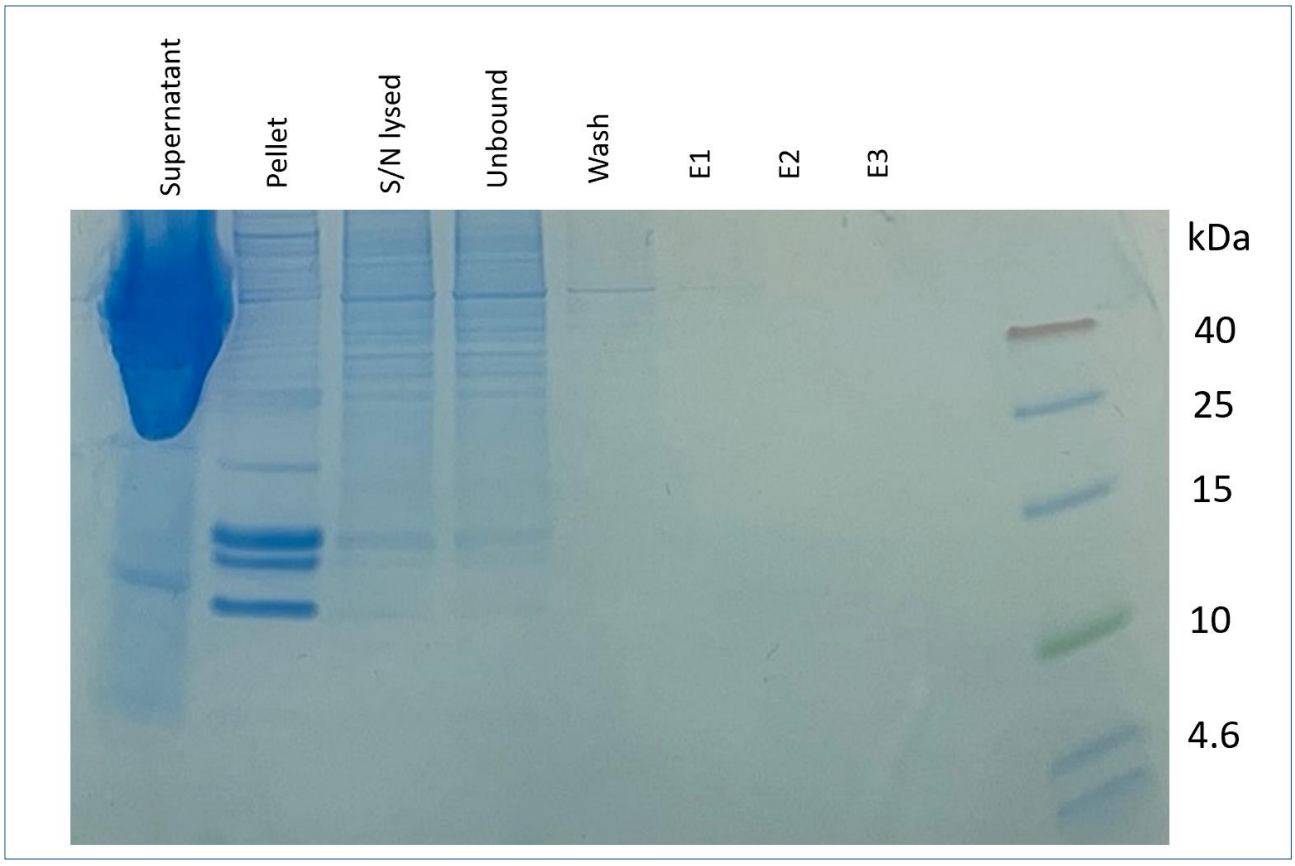
KDa, Kilodaltons; pcDNA3.1, empty vector (negative control); WT, wild-type; MUT, mutant (c.356G>A); 6w, 6 well plate; T75, T75 cell culture dish. Signal generated probing with Anti-Strep-tag II antibody (ab76949). The 25KDa band is presumed nonspecific as it is seen in all sample lanes. The Signal of interest is highlighted by the red box and is presumed to be Strep II and His tagged pro-IGF-1.

A strong streptomycin signal around the expected size of 14KDa was detected in wild-type and mutant samples for both the HEKs transfected in 6 well plates and T75 dishes. It was slightly stronger in the T75 flasks so I proceeded to generate my lysates using this method. No signal was seen in my

HEKs transfected with an empty vector, in keeping with the band being specific for streptomycin containing constructs. In addition, a larger band around 25KDa was detected. As this was present in all samples including my negative control, I felt this was a non-specific band. I also probed for His using the same lysates, but no signal was detected. This could be due to the conformation of the protein or the charge. The samples were boiled in reducing buffer at 95°C for 5 minutes.

Following my first purification using the StrepTrap columns, I collected 20µl from each stage and used them as input for a Coomassie stain, shown in **Figure 3.41**. This is a non-specific protein stain that could show me if any protein had been captured in my eluted samples.

Figure 3.41 Coomassie stain showing protein signal from StrepTrap purification stages

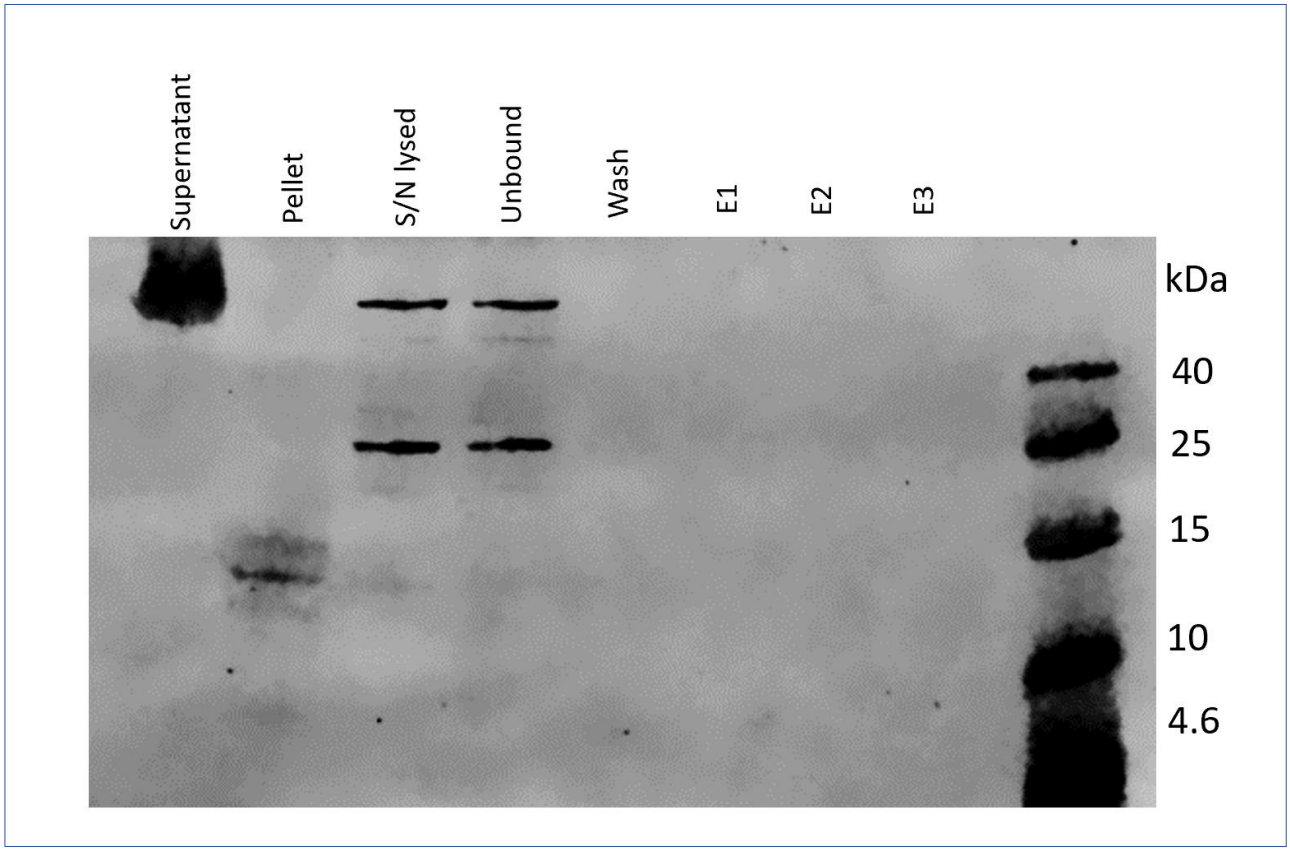


KDa, Kilodaltons; S/N, supernatant; E1, eluate 1; E2, eluate 2; E3, eluate 3. Representative image from WT IGF-1 lysates. The same results were obtained for the mutant c.356G>A IGF-1 lysates.

The strongest band was for the pellet, suggesting the protein was not dissolved into the triton solution. To verify if the streptomycin tagged constructs were still in the insoluble fraction, i.e. the

pellet, I ran a western blot using the same inputs as the Coomassie stain, probing with streptomycin antibody to achieve a more specific signal. This is shown below in **Figure 3.42.**

Figure 3.42 Western blot showing streptomycin signal from StrepTrap purification stages

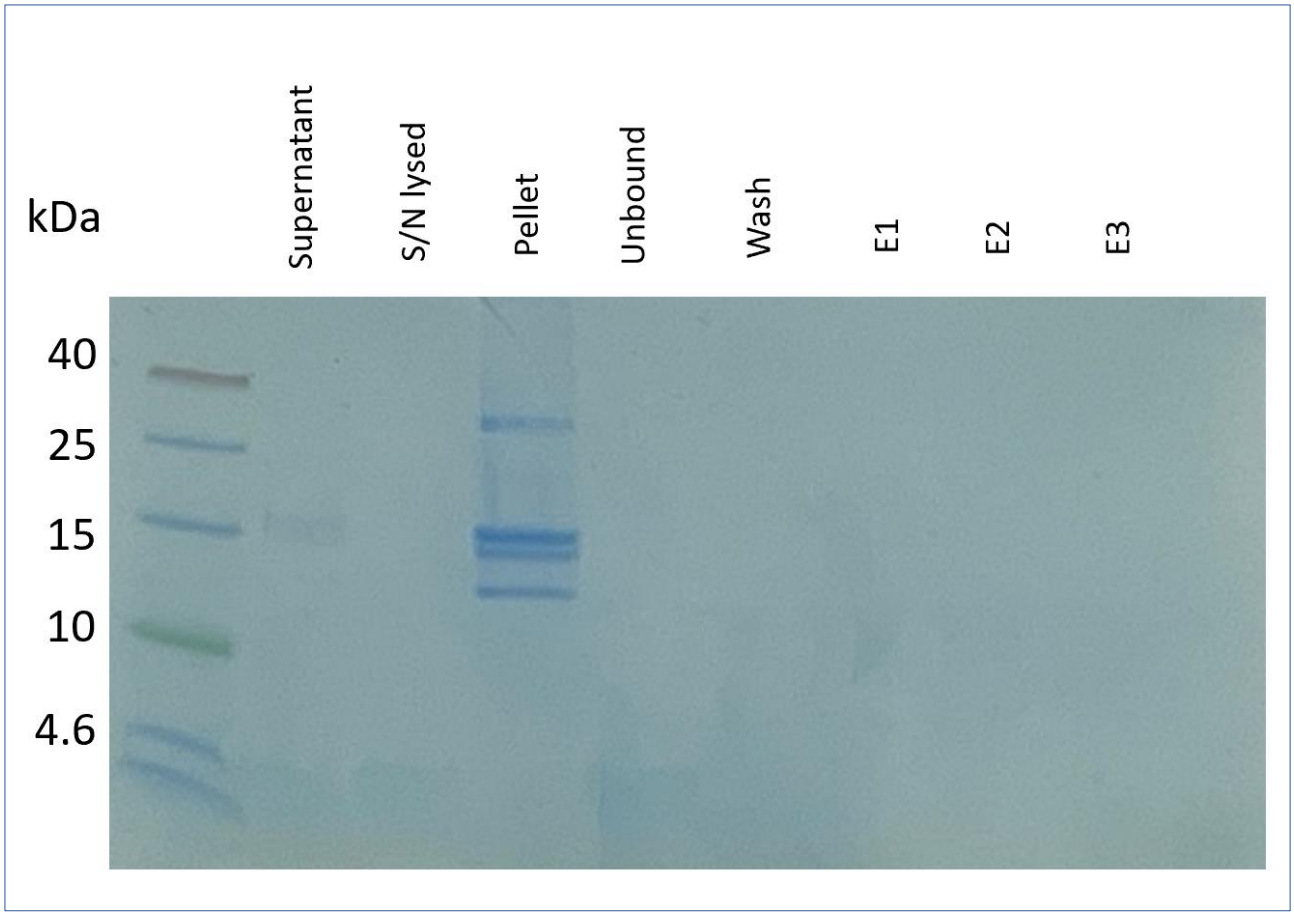


KDa, Kilodaltons; pcDNA3.1, S/N, supernatant; E1, eluate 1; E2, eluate 2; E3, eluate 3. Signal generated probing with Anti-Strep-tag II antibody (ab76949). Representative image from WT IGF-1 lysates. The same results were obtained for the mutant c.356G>A IGF-1 lysates

The protein (expected size 14kDa) was detected most strongly in the insoluble fraction, suggesting it didn't dissolve into the triton suspension. The bands shown in supernatant lysed and unbound were likely to be non-specific. The smaller band around 25kDa was detected on my previous western blot (**Figure 3.40**) and the larger band around 50kDa was likely too large to be detected given the scale of the western blot. Upon reflection, it was felt that as the protein is overexpressed it may have formed insoluble aggregates or the protein could be bound to the membrane, for example at the IGF1R.

To try and overcome this issue, I added a higher concentration of triton (0.5%) to the pellet suspension and incubated at 1 hour at room temperature to optimise solubility. I then repeated the StrepTrap column purification and Coomassie stain. This Coomassie stain is shown below in **Figure 3.43.**

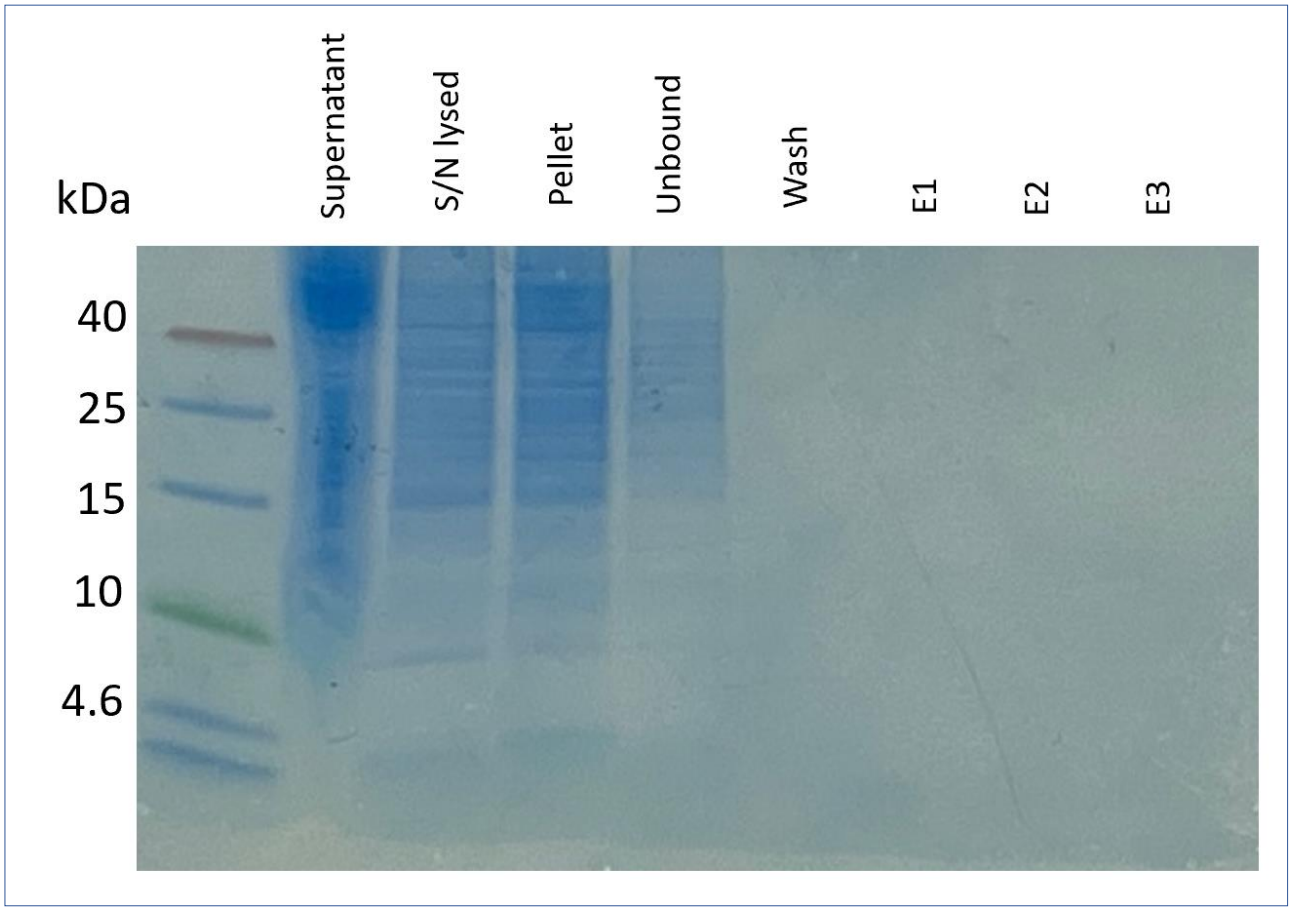
Figure 3.43 Coomassie stain showing protein signal from StrepTrap purification stages when lysing in 0.5% triton



KDa, Kilodaltons; S/N, supernatant; E1, eluate 1; E2, eluate 2; E3, eluate 3. Representative image from WT IGF-1 lysates. Same results obtained for mutant c.356G>A IGF-1 lysates

As there still did not appear to be protein in the eluate, I tried the purification again this time using RIPA buffer. The resulting Coomassie stain is shown below in **Figure 3.44.**

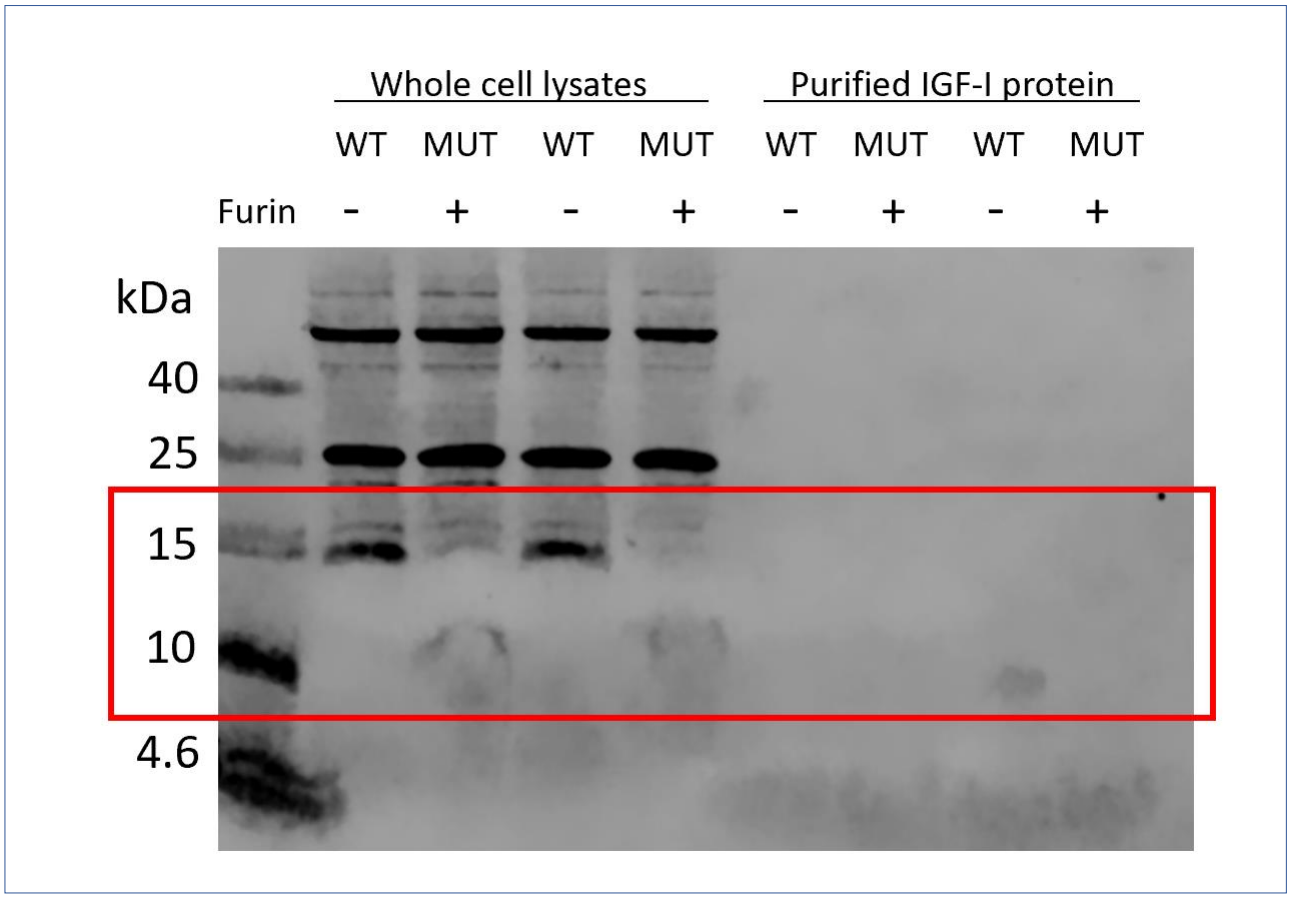
Figure 3.44 Coomassie stain showing protein signal from StrepTrap purification stages when lysing in RIPA buffer



KDa, Kilodaltons; S/N, supernatant; E1, eluate 1; E2, eluate 2; E3, eluate 3. Representative image from WT IGF-1 lysates. Same results obtained for mutant c.356G>A IGF-1 lysates.

Even when using the RIPA buffer, there was not a strong protein signal detected in any of the eluates. Given that a strong, clear streptomycin signal had been generated on my initial transfected whole cell lysate blots (**Figure 3.40**), I decided to trial both the eluant and the whole cell lysates in the furin assay. For this initial furin assay, I incubated samples with 160ng furin or with furin assay buffer only to generate negative controls and incubated all samples overnight at 37°C. The results of this assay are shown below in **Figure 3.45**. For further details of the furin assay please see methods **Section 2.19.**

Figure 3.45 Western blot of products from furin assay



KDa, Kilodaltons; WT, wild-type IGF-1; MUT, mutant c.356G>A IGF-1. Signal generated probing with Anti-Strep-tag II antibody (ab76949). Red box indicates region of interest. Pro-IGF-1 constructs (i.e. undigested IGF-1) would be expected to be around 14KDa and mature IGF-1 (i.e. digested IGF-1) around 10KDa.

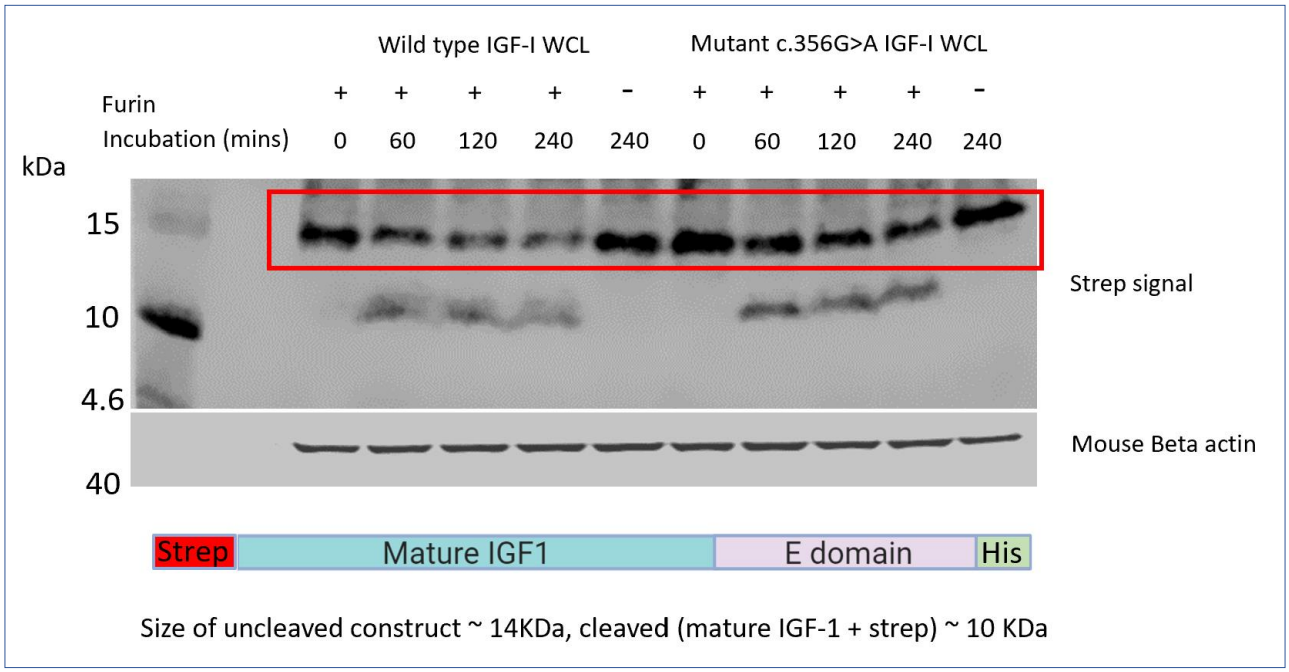
This initial furin assay suggested that furin was able to digest the protein of interest when incubated under the conditions I had created based on the literature. A smaller band around 11kDa is thought to represent the mature IGF-1. The same non-specific bands were seen around 25kDa and 50kDa. Of note, the cleaved E-domain would not be detectable on this blot as it doesn't contain a streptomycin signal. When I probed these products for His by western blot, no signal was observed in any of the lanes.

Given the relatively clear signals obtained using the whole cell lysates and my multiple unsuccessful attempts at isolating the strep tagged IGF-1 constructs, I decided to proceed with the whole cell lysates only.

3.3.6.2 Furin assay results

Using my optimised protocol described in methods **Section 2.19,** I incubated whole cell lysates with furin and then assessed digestion by the intensity of signals detected on western blot imaging. An example of this western blot is shown below in **Figure 3.46.**

Figure 3.46 Western blot of wild-type and mutant IGF-1 furin assay products

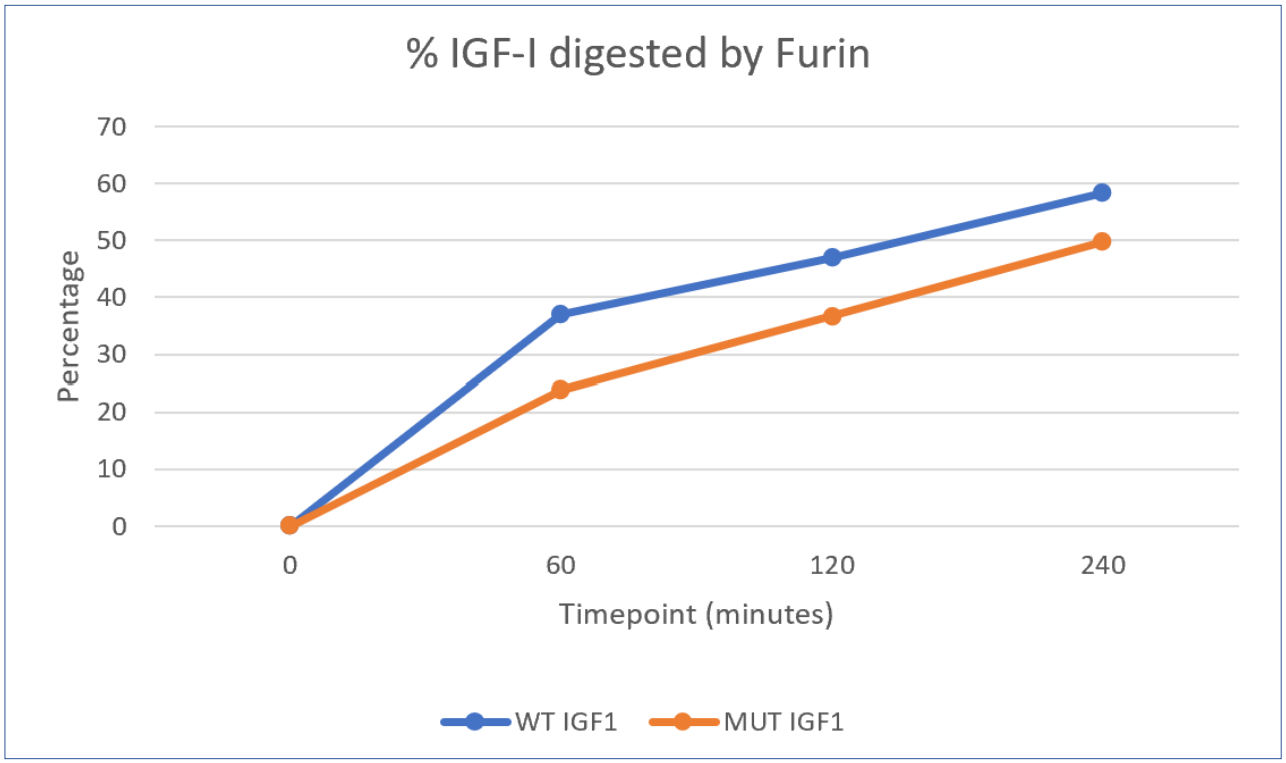


KDa, kilodaltons; WCL, whole cell lysates. + or – indicates presence of 160ng furin in the sample incubated. Samples without furin were incubated with furin assay buffer alone. Mouse Beta actin was used as a loading control. The diagram on the bottom helps to visualise the IGF-1 construct. The band ~ 14KDa is expected to be the Pro-IGF-1 construct (uncleaved product, strep + mature IGF-1 + E domain + His). The band ~ 10 KDa is expected to be the mature IGF-1 construct (cleaved product, Strep + mature IGF-1 only).

Streptomycin signal was detectable for both the cleaved and uncleaved construct, but only reliably quantifiable for the uncleaved construct, highlighted above in the red box. I therefore quantified how much the pro-IGF-1 had reduced from timepoint 0 using Image J software (as described in

methods **Section 2.19**) and deduced this to be the amount digested. Reassuringly, the whole cell lysates incubated for the maximal time of 240 minutes with no furin do not appear to have undergone digestion. My analysis of the pro-IGF-1 construct for the western blot above is shown below in **Figure 3.47**.

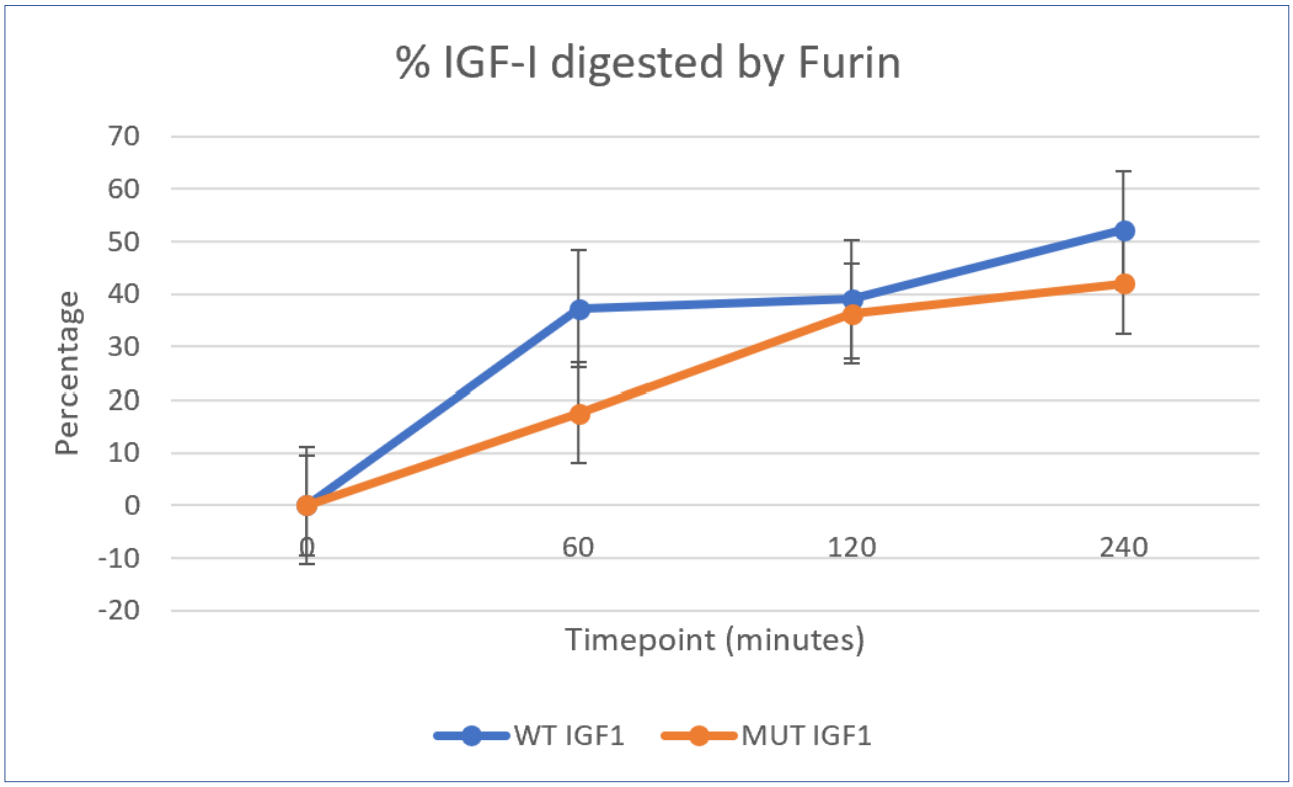
Figure 3.47 Furin digestion of wild-type and mutant IGF-1 proteins



WT, wild-type IGF-1; MUT, mutant c.356G>A IGF-1. Results from quantifying western blot bands using Image J software (https://imagej.nih.gov/ij/).

After performing 3 identical replicate experiments for this furin assay, I pooled the results and these are shown in **Figure 3.48.**

Figure 3.48 Furin digestion of wild-type and mutant IGF-1 proteins (3 replicates)



WT, wild-type IGF-1; MUT, mutant c.356G>A IGF-1. Results from quantifying western blot bands for three replicate furin assays using Image J software (https://imagej.nih.gov/ij/).

Although not significant, there did seem an overall trend for the mutant IGF-1 to be more resistant to digestion by furin than the wild type IGF-1. Another limitation of this analysis is that I was only able to reliably assess the pro-IGF-1 bands (i.e. uncleaved IGF-1) on western blot, rather than the smaller mature IGF-1 (i.e. cleaved).

3.4 Discussion

Creating the custom gene panel enabled detailed analysis of key growth genes of interest. The data enabled rapid diagnosis of recognised genetic causes of growth failure, for example recognised exonic mutations in established genes, but also enabled detailed exploration of non-coding regions in the selected candidate genes. Achieving genetic diagnosis for patients ends uncertainty and helps

guide future management. For example, the *SOS2* mutation identified in Patient 17, in keeping with a diagnosis of Noonan syndrome, meant he was eligible for GH treatment and enabled screening for cardiac comorbidities with an echocardiogram.

The majority of the novel genetic variants identified and selected for further functional analysis involved genes in the GH-IGF-1 pathway. This is not surprising given the many of the patients were referred to our centre with suspected defects in the GH-IGF-1 axis.

Through bioinformatic filtering of non-coding variants called on our custom gene panel, I identified a novel homozygous variant c.618+836T>G in intron 6 of the *GHR*, 44bp downstream of the previously recognized pseudoexon mutation. My *in vitro* splicing assay revealed inclusion of a 151bp pseudoexon due to activation of the same dormant acceptor site involved in the mis-splicing and inclusion of the previously reported 6Ψ *GHR* pseudoexon. In contrast to the 6Ψ *GHR* pseudoexon, inclusion of the 6Ω pseudoexon into the mRNA transcript leads to a frameshift and introduction of a premature termination codon after 245 amino acids. This results in a 45KDa mutant 6Ω GHR protein lacking both transmembrane and intracellular domains, required to anchor the receptor in the cell membrane and induce intracellular signaling, respectively. Using Gibson assembly to recreate this mutant, I demonstrated that the mutant 6Ω GHR leads to diminished STAT5 signaling *in vitro* and accumulation of mutant protein extracellularly, likely due to its inability to anchor to the cell membrane. The predicted deleterious impact of the 6Ω pseudoexon inclusion is in keeping with the severe postnatal growth failure seen in all 3 patients.

Our centre previously described the first *GHR* pseudoexon (6Ψ) mutation in 2001 in four siblings from a highly consanguineous Pakistani family with mild GHI (63). This homozygous point mutation (c.618+792A>G) altered the intronic sequence activating a cryptic donor splice site. Due to the

presence of a nearby dormant cryptic acceptor site, this region is recognized as an exon (a 'pseudoexon') by the spliceosome and is retained during *GHR* splicing. The inclusion of this pseudoexon caused an in-frame insertion of 36 amino acid residues (lacking a stop codon) between exons 6 and 7 in the dimerization domain of the GHR. This resulted in defective trafficking (and concomitant reduced cell surface expression) rather than impaired signalling, causing a partial loss-of-function (220). As such, moderate postnatal growth failure was observed (Height SDS -3.3 to -6.0) and 50% of these patients had no dysmorphic features (221,222). The 6Ψ *GHR* mutation, 792 bases into the intron, was identified using homozygosity mapping of several polymorphic markers surrounding the *GHR* to identify the locus then discovered by cDNA analysis (63). It would not be detected by conventional candidate gene sequencing or whole exome sequencing which only covers exons and intron-exon boundaries. This novel c.618+836T>G pseudoexon variant would not have been identified by whole exome or classic candidate gene sequencing. At 836 bases into the intron 6, it could only be identified by targeted sequencing of this region, custom whole gene panel or whole genome sequencing. It would also be possible to identify by cDNA analysis, but a high index of suspicion would first be required in order to decide to perform this.

Both the 6Ψ and 6Ω *GHR* pseudoexon inclusion events occur in the same intronic region, suggesting that intron 6 may harbour several cryptic splice sites. Alternatively, the dormant cryptic acceptor splice site may be recognised favourably by the spliceosome, predisposing this region to be recognised as an exon. Given the classic GHI phenotype of the patients with the novel 6Ω *GHR* pseudoexon, it would be logical to include this region of intron 6 for candidate gene sequencing of all patients referred with GHI (not just mild cases as previously may have been considered due to knowledge of the 6Ψ pseudoexon).

The heterozygous nonsense *GHR* mutation c.181C>T (R43X) identified in Patients 18 and 19 has been reported in Ecuadorian, Mediterranean and Russian populations (210,223,224). It is thought to have arisen independently in these diverse populations. The c.181C>T mutation occurs at a highly mutable CpG dinucleotide 'hot spot' and has been detected in patients with a variety of *GHR* haplotypes (210,223,224).

Consistent with the severe IGF-1 deficiency, patients 5 and 19 also had 'classical' Laron syndrome facial features. Although Patient 18 did not have the typical frontal bossing and depressed nasal bridge, he did have reduced facial height, suggesting some phenotypic variability despite comparable biochemical abnormalities. The mothers of both kindreds who carried the 6Ω pseudoexon had short stature (-2.0 and -3.2 SDS). It has been recognized that some mothers and sisters who are heterozygous for deleterious *GHR* mutations, have more significant growth failure (height <-2 SDS) compared to male carriers (225). Heterozygosity of the functionally null E180 Ecuadorian mutation was not associated with a reduction in circulating GHBP, IGF-1, IGF-2, IGFBP 2, or IGFBP 3 levels (226,227)(226,228). The parents of Patient 5 and the mother of patients 18 and 19 who all carried the novel GHR 6Ω pseudoexon variant had normal IGF-1 levels. Patient 5's parents (novel GHR 6Ω pseudoexon variant carriers) also had normal IGFBP 3 and ALS levels. In contrast the father of Patients 18 and 19, who carried the known GHR R43X mutation, had low IGF-1 levels which may suggest that this mutation in heterozygosity has a greater impact on IGF-1 secretion than either of the *GHR* pseudoexon variants.

It is notable that all 3 6Ω *GHR* pseudoexon patients had elevated TSH. Most patients with Growth Hormone Insensitivity have thyroid function within the normal range (69). Furthermore, exogenous administration of IGF-1 in individuals with Laron syndrome did not negatively impact thyroid function (229). However, the relationship between the GH-IGF-1 system and the hypothalamic-

pituitary-thyroid axis is complex and incompletely understood. GH therapy in children and adults with GH deficiency can induce a fall in serum T4 (230). This is thought to be due to the GH effect on deiodination of T4 to T3, leading to higher serum T3 levels (231). It could be hypothesized that the supraphysiological levels of GH seen in patients 18 and 19 are responsible for their raised TSH levels due to reduced T4 feedback, but the mechanisms are not fully understood.

Both sets of parents from the 2 kindreds originate from Frattamaggiore, a town in the Campania region of Southern Italy, suggesting they share common ancestry. Interestingly, the majority of reported patients with GHR 6Ψ mutations are of Pakistani origin and previous work by our research group suggests the presence of a common ancestor (80). The E180 *GHR* splice mutation is the most common mutation identified in patients with classic GHI, comprising approximately one third of the known population with GHR deficiency. This mutation is concentrated in a large population of individuals with Laron syndrome in Southern Ecuador and is thought to have also originated from a single common ancestor (232,233).

The GHBP assays performed by Martin Bidlingmaier's lab, described in detail in methods **Section 2.20**, used highly specific monoclonal GHBP capture antibodies. One of these recognizes a critical epitope which is significantly altered by the 6Ω pseudoexon inclusion. This explains the undetectable GHBP levels in the sera of all 3 patients harbouring the novel 6Ω pseudoexon. We know from previous analyses in Laron patients that this particular antibody is also unable to bind to GHBP in patients carrying missense R161C and R211G GHR mutations, the nonsense R217X mutation and the E180 and G223 splicing mutations which also modify this region.

When assessing patient variants from the custom gene panel, I first identified patients with pathogenic or likely pathogenic exonic mutations and excluded them from further analyses. If

present on the variant call file, the c.723C>T *GHR* variant would have been easily identifiable as a pathogenic mutation in Patient 6. This patient would therefore have been filtered out from further analysis pending Sanger sequencing confirmation of this variant in the patient and family members.

As this variant was not called in the Variant Call File and thus not identifiable on Ingenuity Variant Analysis, I explored variants in non-coding regions that could be responsible for his phenotype given that he had no likely causative exonic variants and had a classic GHI phenotype. This led to the identification of the deep intronic c.70+32137A>G *GHR* variant, which did not appear to affect splicing upon *in vitro* analysis with the MoBiTec splicing assay.

Patient fibroblasts played a key role in the diagnosis of this patient, as through assessing this GHR cDNA I was able to identify the genetic abnormality responsible for his Growth Hormone Insensitivity and postnatal growth failure. Patient 6 proved to be a diagnostic challenge but through assessment of patient fibroblasts and visual inspection of BAM BAI files I am pleased we were able to provide a diagnosis for this family. This diagnosis ends uncertainty, helps secure treatment, informs prognosis and guides future management.

The splicing assay for Patient 8 showed complete absence of normal *GHR* transcript and is in keeping with his biochemical GHI and lack of dysmorphic features, similar to other reported patients with *GHR* dominant negative mutations. However, Patient 7 had a significant proportion of wild-type GHR transcript. Considering he had some skeletal disproportion and does not have classic GHI biochemistry, it is possible that he has (an)other genetic defects contributing to his phenotype. He has been assessed for mutations in *SHOX* and *FGFR3* and no other causative mutations were identified on whole exome sequencing. It would be worthwhile assessing him by whole genome sequencing and applying a virtual panel of the long list of potential skeletal dysplasia genes.

The mutant *GHR* created by the abnormal splicing seen in Patients 7 and 8 is of particular interest as it mimics a naturally occurring GHR transcript that arises in the absence of the mutation, that may play an important regulatory role in normal physiology. This different *GHR* transcript, missing the first 26 bases of exon 9 (without the additional T>G change seen in Patient 8), is found in a significant percentage of *GHR* transcripts in healthy people. One study by Ross *et al.* looked at alternative mRNA splicing of intracellular domain of GHR by RT-PCR of human liver and cultured IM-9 cells. The presence of two alternative transcripts was shown. 1-10% of all *GHR* transcripts were homologous to the mutant *GHR* transcript missing the first 26bp of exon 9, and 1% of transcripts skipped all of exon 9 (82). The authors suggest that the mutant *GHR* lacking the first 26bp of exon 9 modulates the function of the full length receptor, inhibits signalling and generates large amounts of GHBP (82). Interestingly, Patient 8 who appeared to have almost 100% of this mutant GHR on the *in vitro* splicing assay had exceedingly high GHBP levels on 2 independent assays.

The regulation of GHR expression and GHBP generation is not well understood and exploring the function and interaction of this mutant GHR protein will help provide insights into these crucially important interactions.

The novel heterozygous *IGF1* c.356G>A variant I identified in Patient 10 did not impair localisation of IGF-1, with both wild-type and mutant constructs able to translocate to the endoplasmic reticulum of HEK293T cells. Unfortunately, I was unable to isolate the Streptomycin tagged IGF-1 constructs for the furin assay using a variety of techniques with the StrepTrap columns, so proceeded with the whole cell lysates. With further optimisation, it may be possible to isolate these and thus create a 'cleaner' and stronger signal for the western blot analysis of cleavage. Using whole cell lysates in the furin assay, the streptomycin signal was detectable for both the cleaved and

uncleaved construct, but only reliably quantifiable for the uncleaved construct. I therefore quantified how much the pro-IGF-1 had reduced from timepoint 0 using Image J software and deduced this to be the amount digested. The results did not show significantly reduced cleavage, rather a mild trend towards reduced cleavage. Repeating the experiments further times may reduce the size of the error bars and improve the significance. I could also try to functionally assess this variant using other techniques. I could transfect the FLAG tagged wild-type and mutant *IGF1* constructs into HEK293T cells, then harvest the serum and apply it to human fibroblasts to assess downstream signalling. Fibroblasts are known to express IGF1R so are IGF-1 responsive. Harvesting these whole cell lysates and western blotting for total IGF1R and phospho1135 IGF1R. Phospho1135 IGF1R is a specific marker of IGF1R and reduced levels of this in the mutant sample would suggest impaired signalling.

Chapter 4. Copy number variants in patients with Growth Hormone Insensitivity (GHI) and unexplained short stature

4.1 Introduction

Children with short stature are recognised to have a greater burden of rare CNVs (234–237) and a greater average CNV length compared to those with normal height (234,235). The incidence of pathogenic and likely pathogenic CNVs is reported as 4–13% in patients with idiopathic short stature (234,238,239) and 16% in patients born small for gestational age with persistent short stature (236,237). Many of these patients display growth retardation in association with malformations and/or neurodevelopmental disorders (236,237). In one study, pathogenic or likely pathogenic CNVs were identified in 14% of patients with syndromic short stature (240). Overall, the findings of these studies support that rare CNVs contribute to short stature and have helped to reveal potential candidate genes and/or loci (234–237,239).

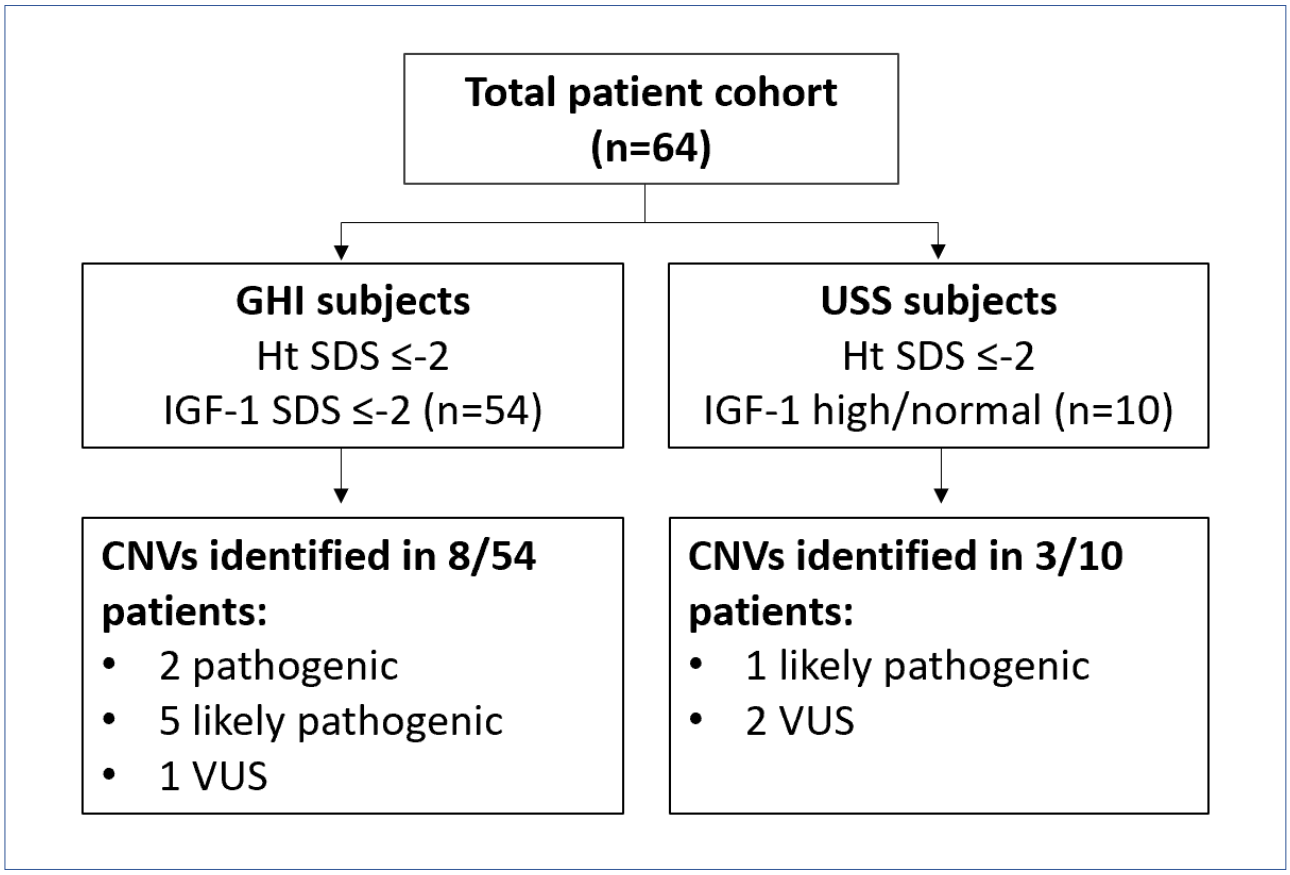
More than 30 pathogenic CNVs have been identified in patients with suspected Silver Russell Syndrome (SRS) (130,236,241,242). Some have SRS-compatible phenotypes but many do not fulfil the strict NH-CSS criteria (detailed in **Table 1.2**) and frequently have more severe developmental delay and/or intellectual disability compared to 'classic' SRS patients (131). In one study, pathogenic CNVs were identified in 6% SRS-compatible patients and 7% SRS-like patients who were negative for classic SRS imprinting defects (243). Some SRS patients, especially those with upd(7)mat, have fewer 'typical' clinical SRS features compared with 11p15 LOM individuals. These patients are part of the SRS-like spectrum and corroborate the wide clinical and genetic heterogeneity of SRS (244–247).

The identification of a pathogenic molecular defect is important for patients, families and clinicians as it can avoid unnecessary investigations and/or treatment, ends uncertainty and allows appropriate genetic counselling. It is also fundamental to identify co-morbidities associated with syndromic SS. I investigated the role of CNVs in the aetiology of GHI and unexplained short stature and interrogated the genomic regions for novel candidate genes and pathways. The clinical and biochemical phenotypes of the CNV patients were analysed in detail to classify this novel subtype of GHI and unexplained short stature patients. The classes of CNVs are described in methods **Section 2.3.3.3.**

4.2 Results

Pathogenic or likely pathogenic CNVs (class 4 or 5) were discovered in 8/64 (13%) patients (seven GHI subjects and one unexplained short stature subject). Additionally, three variants of uncertain significance CNVs (VUS, class 3) were identified in one GHI and two unexplained short stature patients. (**Figure 4.1** and **Table 4.1).**

Figure 4.1 Summary of the CNVs identified in the GHI and unexplained short stature subjects



The most deleterious CNVs detected in each patient are listed. One patient with GHI and two patients with unexplained short stature had more than one CNV, so listed above is the CNV predicted to be most deleterious. GHI, Growth Hormone Insensitivity; USS, unexplained short stature; SDS, Standard Deviation Score, CNV, copy number variant; VUS, variant of uncertain significance.

4.2.1 Clinical and biochemical features of the CNV subjects

GHI subjects

I identified CNVs in classes 3 to 5 in 8/54 (15%) patients (6 males; mean height SDS -3.9, range -1.6 to -5.7). 7/54 (13%) patients (5 males, mean height SDS -3.9) had pathogenic or likely pathogenic CNVs (class 4 or 5). Interestingly, a total of 75% GHI subjects with class 3-5 CNVs were male. However, phenotypic predictors for CNV in the GHI cohort did not reveal any significant differences between the height SDS, age, sex, BW SDS or IGF-1 SDS in the GHI subjects harbouring CNVs compared to those without CNVs.

Unexplained short stature subjects

I identified CNVs classes 3-5 in 3/10 (30%) subjects (2 males, mean height SDS -2.8, range -2.0 to -5.7). Pathogenic or likely pathogenic CNVs were found in 1/10 (10%). There were no significant differences between the mean height, age, sex, BW SDS or IGF-1 SDS in the unexplained short stature patients harbouring CNVs compared to those without CNVs.

Table 4.1 Clinical and biochemical features of the patients harbouring copy number variants

Patient	Age at referral (years)	Sex	BW SDS	Height SDS	IGF-1 SDS	Previous genetic testing	Additional features
Growth Hormone Insensitivity (GHI) subjects							
20	3.8	M	-1.6	-3.6	-2.0	CGS, WES, SS Panel, SRS**	Autistic spectrum
21	1.1	F	-1.7	-1.6	-2.4	CGS, WES, SS Panel	Language delay, dyslexia, recurrent ear infections
22	9.1	M	-0.4	-3.7	-2.3	CGS, WES, SS Panel	-
23	11.3	M	-1.9	-5.1	-2.7*	CGS, WES, SS Panel	-
24	1.9	M	-3.2	-5.7	-2.4	CGS, WES, SS Panel, SRS**	Delayed motor development
25	17.0	M	-0.3	-4.0	-2.1	CGS, WES, SS Panel	Delayed puberty, learning difficulties
26	2.8	M	-0.7	-4.9	-2.8	CGS, SS Panel, SRS**	Persistent abdominal distention, bloating, severe constipation.
27	12.4	F	0.3	-2.5	-2.7	SS panel	Migraine, normal brain MRI

Unexplained short stature subjects							
28	14.4	M	-2.2	-2.7	-0.6	CGS, SS Panel	-
29	2.7	M	-2.1	-2.0	-0.8	CGS, SS Panel	Adrenal insufficiency
30	2.5	F	-1.3	-3.6	1.3	SS Panel	-

CGS, candidate gene sequencing (*GHR*, *GHR 6ψ, IGFALS* and *IGF1* for GHI group and 3M syndrome genes, *CUL7*, *CCDC8, OBSL1* and *IGF1R* for unexplained short stature group*)*; WES, whole exome sequencing; SS Panel, custom gene panel covering entire genomic sequence of 64 genes associated with GH-IGF-1 axis defects causing GHI and IGF-1 insensitivity and overlapping short stature syndromes; F, Female; M, Male; BW, birth weight; NK, not known; *, IGF-1 undetectable on assay. ** negative testing for 11p15 LOM +/- upd(7)mat undertaken at the referring centre. Patients 20-26, 28 and 29 are included in previous publication from our centre (109).

4.2.2 SRS features in the CNV subjects

Five GHI patients with class 3-5 CNVs had clinical features of Silver Russell Syndrome (SRS; NH-CSS ≥2 in addition to SRS-like features, **Table 4.2**). However, only GHI patient 22 fulfilled the NH-CSS criteria with 3/6 clinical criteria in addition to a recognized genetic cause (1q21 deletion (248)). Six patients (5 GHI) had CNVs previously reported in suspected SRS: 1q21 deletion (n=3), 12q14 deletion (n=1), Xp22 duplication (n=1) and Xq26 duplication (n=1). A number of additional features, not part of the NH-CSS criteria but previously associated with SRS, were identified in 7 (64%) patients (**Table 4.2**) (131).

Table 4.2 SRS features in the patients harbouring CNVs

Patient	CNV	NH-CSS criteria	Additional SRS features*	CNV previously associated with short stature / SRS
Growth Hormone Insensitivity (GHI) subjects				
20	1q21 deletion	2 (b,f)	Triangular face, high arched palate, hypoglycaemia, clinodactly	SRS features (248) and 1q21 deletion syndrome (OMIM #612474)
21	1q21 deletion	1 (f)	Speech delay	SRS features (248) and 1q21 deletion syndrome (OMIM #612474)
22	1q21 deletion	3 (b,c,f)	Clinodactly	SRS features (248) and 1q21 deletion syndrome (OMIM #612474)
23	12q14 deletion	2 (b,f)	Triangular face, high pitched voice	SRS features (241) and 12q14 deletion syndrome** (249)
24	7q21 deletion, 7q31 deletion	2 (a,b)	Triangular face, low set ears, delayed motor development	None reported
25	5q12 deletion	1 (b)	Nil	5q12 deletion syndrome, OMIM #615668 (250)
26	15q11 deletion	3 (b,d,f)	Triangular face, hypoglycaemia	Yes - 15q11 deletion, OMIM #615656
27	Xq26 duplication	1 (b)	Brachydactly, downturned mouth	SRS features (248)***

Unexplained short stature subjects				
28	7q21 duplication, Xp22 duplication	2 (a,b)	Nil	SRS features with Xp22 duplication (242). Xp22 duplication also identified in SGA cohort (237).****
29	7q36 duplication	2 (a,b)	Nil	None reported
30	3p22 deletion, 15q13 duplication	1 (b)	Nil	None reported

NH-CSS, Netchine-Harbison SRS Clinical Scoring System: diagnosis of SRS requires fulfilment of 4/6 (including both prominent forehead and relative macrocephaly, termed 'Clinical SRS') or 3/6 in addition to a genetic diagnosis associated with SRS. The criteria are: a. SGA (Birth weight and/or birth length ≤−2 SDS for gestational age) b. Postnatal growth failure (Height at 24 ± 1 months ≤−2 SDS or height ≤−2 SDS below mid-parental target height) c. Relative macrocephaly at birth (Head circumference at birth ≥1.5 SDS above birth weight and/or length SDS) d. Protruding forehead (Forehead projecting beyond the facial plane on a side view at 1–3 years) e. Body asymmetry (Leg length discrepancy (LLD) of ≥0.5 cm or arm asymmetry or LLD <0.5 cm with at least two other asymmetrical body parts, one non-face) f. Feeding difficulties and/or low BMI (BMI ≤−2 SDS at 24 months or use of feeding tube or cyproheptadine as appetite stimulant) (130,131). * additional clinical features recognised in SRS (131); ** no OMIM number assigned to this syndrome currently; *** the duplication described in Spengler *et al*. begins at Xq25 (genomic co-ordinates 129,132,238- 139,650,444) whilst our patient duplication begins at Xq26 (co-ordinates 134,842,275-141,407,613). **** Both these duplications described in the literature are larger than the CNV identified in our patient and encompass the SHOX region and/or the SHOX enhancer region, whilst our CNV does not include either.

4.2.3 Details of the CNVs identified in the GHI and unexplained short stature subjects

A total of 14 CNVs classified as VUS, likely pathogenic or pathogenic were identified in 11 subjects. Of these, 6/14 CNVs (43%) have previously been reported in SRS-like patients (**Table 4.3**). Genomic deletions occurred in 8/11 subjects and ranged in size from 143,487 to 9,111,383bp. Genomic duplications were found in 4/11 patients and ranged from 491,138 to 2,260,322bp. Two of the genomic duplications identified were found in association with another deletion or duplication (Patients 28 and 30, respectively). Two deletions were identified in Patient 24. Interestingly, three subjects harboured a 1q21 deletion, two of whom were siblings (Patients 20 and 21). A total of 122 protein-coding genes were contained in the class 3-5 CNV regions of the GHI cohort and 26 genes in the class 3-5 CNV regions of the unexplained short stature cohort (**Table 4.4**).

Statistical analysis of the CNVs identified was performed by Dr Sumana Chatterjee. Except for one outlier in each group, the CNVs in the GHI group were significantly larger (>500,000bp) than those identified in the unexplained short stature group (<500,000bp) ($p=0.03$). All CNVs identified in GHI subjects were deletions except for one, and similarly all CNVs identified in the unexplained short stature cohort were duplications except for one. However, this was not significant by Fisher's exact t test, likely due to our small sample size.

Table 4.3 Details of copy number variants identified in our patients

Patient	CNV[1]	Class[2]	Inheritance (parental height SDS)	Size (Mb)	Number of affected protein coding genes	DECIPHER[3]	Mouse Genome Database[4]	Main Candidate gene(s) in this region[5]
Growth Hormone Insensitivity patients								
20[6]	1q21.1q21.2(146564742_147735011) x1	4	Maternal (SDS +0.08)	1.17	9	26	3	*BCL9*[b,d,e], *PRKAB2*[c,d], *FMO5*[c,d], *GPR89B*[c], *CHD1L*[d,e]
21[6]	1q21.1q21.2(146641600_147735011) x1	4	Maternal (SDS +0.08)	1.09	9	26	3	*BCL9*[b,d,e], *PRKAB2*[c,d], *FMO5*[c,d], *GPR89B*[c], *CHD1L*[d,e]
22	1q21.1q21.2(145987155_147735011) x1	4	*De novo*	1.74	11	26	3	*BCL9*[b,d,e], *PRKAB2*[c,d], *FMO5*[c,d], *GPR89B*[c], *CHD1L*[d,e]
23	12q14.2q15(64413681_67794677)x1	5	NK	3.38	21	9	4	*HMGA2*[a,c,d], *WIF1*[b], *XPOT*[c], *IRAK3*[c], *GRIP1*[c], *LLPH*[d], *MSRB3*[d], *SRGAP1*[d]

24	7q21.2(91914300_92762100)x1 7q31.1q31.31(111130598_120241981)x1	3 4	*De novo* *De novo*	0.85 9.11	9 29	4 7	3 8	*ANKIB1*[c,d], *PEX1*[c,d], *CDK6*[c,d], *SAMD9*[d,e], *GATAD1*[d] *WNT2*[b,c], *IMMP2L*[b, c], *IFRD1*[c], *GPR85*[c], *FOXP2*[c,d], *CAV1*[c], *MET*[c], *CFTR*[c], *CAV2*[d] *PPP1R3A*[d], *TFEC*[d]
25	5q12.1q12.2(60371468_62950178)x1	3	NK	2.58	10	3	2	*ZSWIM6*[c], *CKS1B*[c,e], *DIMT1*[e]
26	15q11.2(22765627_23085096)x1	4	NK	0.32	4	29	0	-
27	Xq26.3q27.2(134842275_141407613)x3	4	*De novo*	6.57	38	14	6	*SOX3*[a,b], *ZIC3*[c], *BRS3*[c], *RBMX*[c], *CD40LG*[c], *FHL1*[c]
Unexplained short stature patients								
28	7q21.13(89733373_90035738)x3 Xp22.33(1793445_2213992)x3	3 3	Maternal (SDS -1.67) Paternal (SDS -0.41)	0.30 0.42	5 1	1 8	0 0	

29	7q36.1(151748853_152032715)x3	3	Maternal (SDS +0.33)	0.28	2	3	1	*GALNT11*[b], *KMT2C*[c]
30	3p22.1(41611009_41754496)x1 15q13.2q13.3(30653877_32914199)x3	3 4	Maternal (SDS -0.75) Paternal (SDS -3.35)	0.14 2.26	1 17	2 14	1 2	*ULK4*[c] *CHRNA7*[b], *KLF13*[b], *FAN1*[c], *OTUD7A*[c]

[1] Co-ordinates given are relative to version 37 of the Human Genome. [2]CNV class 3, Variant uncertain significance (VUS); 4, likely pathogenic; 5, pathogenic (169). [3] Number of patients in DECIPHER with overlapping deletion/duplication (as per patient CNV) and pre/postnatal growth restriction. [4]Number of genes causing a growth restricted phenotype in the mouse model (Mouse Genome Informatics http://www.informatics.jax.org/). [5] Candidate genes were identified by corroborating information from current literature, MGI, DECIPHER, GWAS database information and our bioinformatic pathway analysis. [6] siblings. Candidate genes in the CNV regions were selected as they were: [a]established in literature to have important role in normal linear growth [b]identified as candidate gene from our bioinformatic pathway analysis; [c]causing a growth restricted phenotype in the mouse model; [d]Containing loci associated with height in GWAS catalog (https://www.ebi.ac.uk/gwas/) or in GWAS literature [e]Putative growth role based on current literature.

Table 4.4 Complete list of genes contained in each CNV region identified

Patient	CNV	Genes in the CNV region(s)
Growth Hormone Insensitivity (GHI) subjects		
20	1q21 deletion	*PRKAB2, FMO5, CHD1L, BCL9, ACP6, GJA5, GJA8, GPR89B, NBPF24*
21	1q21 deletion	*PRKAB2, FMO5, CHD1L, BCL9, ACP6, GJA5, GJA8, GPR89B, NBPF24*
22	1q21 deletion	*NBPF11, NBPF12, PRKAB2, FMO5, CHD1L, BCL9, ACP6, GJA5, GJA8, GPR89B, NBPF24*
23	12q14 deletion	*SRGAP1, C12orf66, C12orf56, XPOT, TBK1, RASSF3, AC025262.1, GNS, TBC1D30, WIF1, LEMD3, MSRB3, HMGA2, RP11-366L20.2, AC090673.2, LLPH, TMBIM4, IRAK3, HELB, GRIP1, CAND1*
24	7q21 deletion 7q31 deletion	*ANKIB1, GATAD1, ERVW-1, PEX1, RBM48, FAM133B, CDK6, SAMD9, SAMD9L* *IMMP2L, DOCK4, ZNF277, IFRD1, LSMEM1, TMEM168, C7orf60, GPR85, LINC00998, TSRM, PPP1R3A, FOXP2,* *MDFIC, TFEC, TES, CAV2, CAV1, MET, CAPZA2, ST7, ST7-OT4, WNT2, ASZ1, CFTR, CTTNBP2, NAA38, ANKRD7, KCND2*
25	5q12 deletion	*NDUFAF2, AC008498.1, SMIM15, ZSWIM6, C5orf64, KIF2A, DIMT1, IPO11, CKS1B, LRRC70*
26	15q11 deletion	*TUBGCP5, CYFIP1, NIPA2, NIPA1*

27	Xq26 duplication	*CT45A1, CT45A2, CT45A3, CT45A4, CT45A5, CT45A6, SAGE1, MMGT1, SLC9A6, FHL1, MAP7D3, GPR112, BRS3, HTATSF1, VGLL1, CD40LG, ARHGEF6, RBMX, GPR101, ZIC3, FGF13, F9, MCF2, ATP11C, CXorf66, SOX3, LINC00632, CDR1, SPANXB2, SPANXB1, LDOC1, SPANXC, SPANXA1, SPANXA2, SPANXD, MAGEC3, MAGEC1, MAGEC2*
Unexplained short stature subjects		
28	7q21 duplication Xp22 duplication	*STEAP1, STEAP2, C7orf63, GTPBP10, CLDN12* *DHRSX*
29	7q36 duplication	*GALNT11, KMT2C*
30	3p22 deletion 15q13 duplication	*ULK4* *CHRFAM7A, GOLGA8R, GOLGA8Q, GOLGA8H, ARHGAP11B, FAN1, MTMR10, TRPM1, KLF13, OTUD7A, CHRNA7, GOLGA8K, AC139426.2, GOLGA8O, AC135983.2, GOLGA8N, ARHGAP11A*

CNV, copy number variant.

4.2.4 CNVs in the GHI subjects

Details of the CNVs identified in the GHI subjects are provided in **Table 4.3**. Three patients had 1q21 deletions including two siblings from a non-consanguineous family, Patients 20 and 21. The older male sibling (Patient 20) had a height SDS -3.6 with a small triangular face and high arched palate and BW SDS -1.6. He had negative SRS testing (11p15 LOM and upd(7)mat) at the referring centre. The younger female sibling (Patient 21) had a height SDS -1.6 and BW SDS -1.7 with no dysmorphic features. Both had feeding problems in infancy and recurrent infections. The male sibling experienced hypoglycaemic episodes and was also diagnosed with autistic spectrum disorder. The younger sibling had dyslexia. The deletion was inherited from their mother, whose height SDS was 0.1. Patient 22 had no dysmorphic features, height -3.8 SDS and BW SDS -0.4. He was the offspring of non-consanguineous, Caucasian parents who had normal phenotypes and the deletion was *de novo*.

The 12q14 deletion was identified in Patient 23, a Caucasian male with BW SDS -1.9, height SDS -5.1, triangular face and high-pitched voice. He required nasogastric feeding for poor weight gain. Interestingly, he had an older sibling with similar facial features and poor growth and two half siblings, who also had growth failure. Patient 24 had two *de novo* deletions of chromosome 7 (7q21 and 7q31). This male Kuwaiti patient was born SGA (BW SDS -3.2) and had a triangular face, low set ears and delayed motor development. He had negative SRS testing (11p15 LOM and upd(7)mat) at the referring centre.

The 5q12 deletion was identified in Patient 25, a Caucasian male with postnatal growth failure (height SDS -4.0), delayed puberty and learning difficulties. Patient 26 had a 319Kb deletion of chromosome 15 at the BP1-BP2 Prader-Willi syndrome/Angelman syndrome locus. This male

Caucasian patient had postnatal growth failure (height SDS -4.9), relative macrocephaly, triangular face, hypoglycaemic episodes and feeding difficulties. He had low BMI (-2.0 SDS) and received supplemental enteral feeds. He suffered with persistent abdominal distention, bloating and severe constipation but intestinal biopsies were unremarkable. He had negative SRS testing (11p15 LOM and upd(7)mat) at the referring centre.

A *de novo* Xq26.3q27.2 duplication was detected in Patient 27, a 12-year-old Caucasian female patient with height SDS -2.5, brachydactyly, a downturned mouth and a broad nasal bridge.

4.2.5 CNVs in the unexplained short stature patients

Details of the CNVs identified in the subjects with unexplained short stature are provided in **Table 4.3**. Two CNVs were detected in Patient 28, a 14-year-old male who was born SGA (BW SDS -2.19) with poor catch-up growth (height SDS -2.7),7q21 and Xp22 duplications inherited from his mother and father with normal stature (height SDS -1.7 and -0.41, respectively).

The 7q36 duplication was identified in a 2-year-old Caucasian male (Patient 29) with adrenal insufficiency who was born SGA (BW -2.05) with no catch-up growth. The 3p22 deletion and 15q13 duplication were identified in Patient 30, a 2-year-old female Caucasian patient with postnatal growth failure (height SDS -3.6).

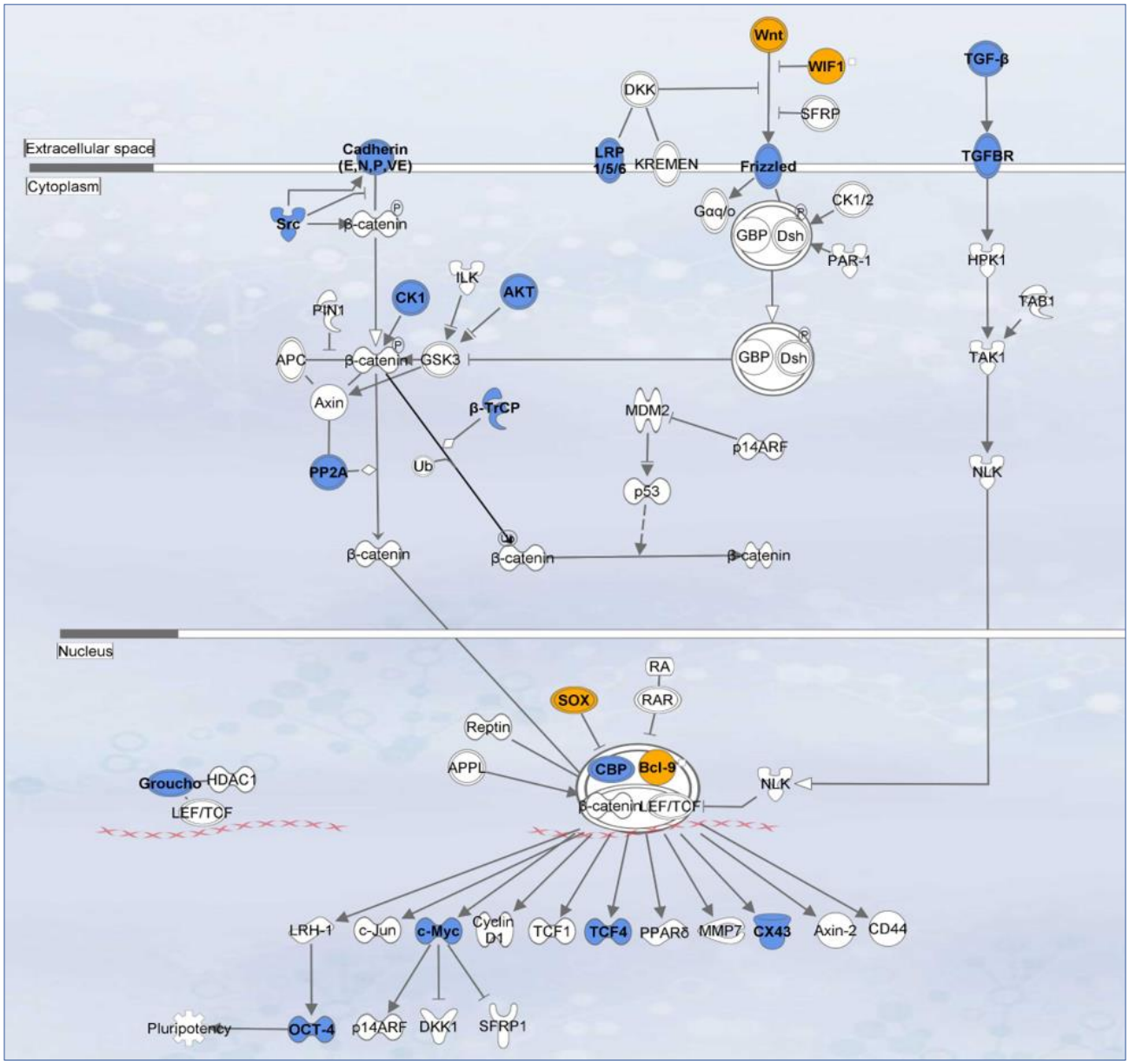
4.3 *In silico* analysis

4.3.1 Pathway enrichment analysis

Ingenuity pathway analysis (IPA) (Qiagen, inc) was performed by bioinformatician Dr Claudia Cabrera. This identified pathways and biological functions enriched within the CNV regions. The analysis was focussed on enriched pathways in individuals of the same group (i.e. GHI or

unexplained short stature). The WNT canonical pathway was identified with evidence of enrichment in three GHI individuals (Patients 20-22; Benjamini-Hochberg adjusted p-value = 0.11). Additionally, WNT pathway genes were observed in 3 other GHI subjects whose enrichment evidence was weak (Patients 23, 24 and 27) (**Figure 4.2**). The WNT pathway genes were *WIF1* in the 12q14 deleted region of Patient 23*, WNT2* in the 7q31 deletion of Patient 24, *SOX3* in the Xq26 duplicated region of Patient 27 and *BCL9* in the 1q21 deleted region of Patients 20, 21 and 22. All of these patients have deletions except for Patient 27.

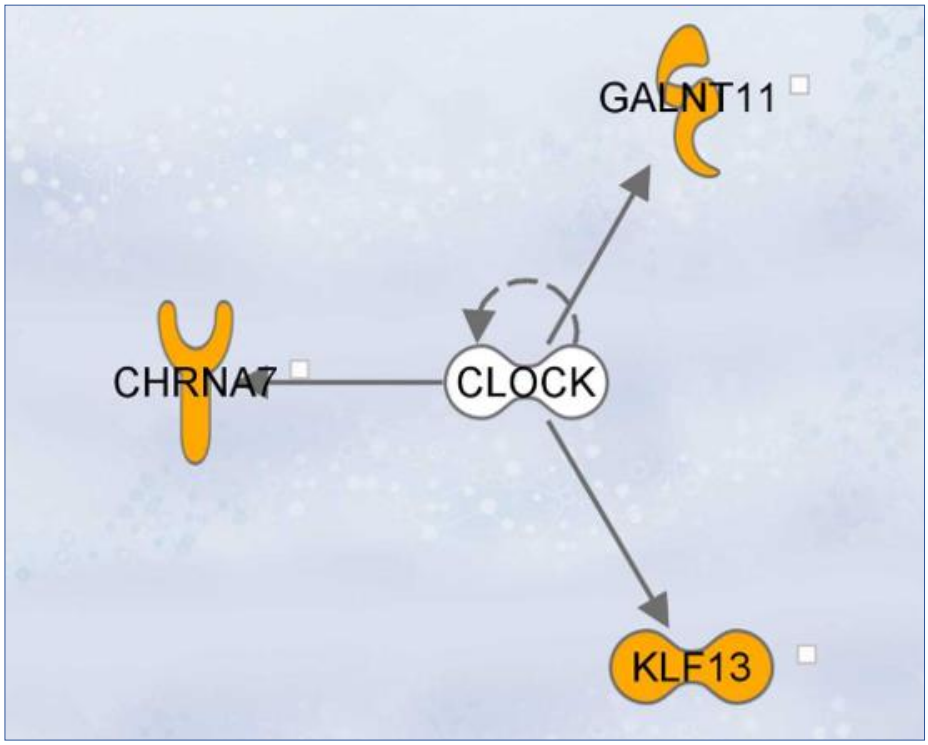
Figure 4.2 Pathway enrichment analysis identified the WNT pathway as enriched within the gene list from CNV regions of the Growth Hormone Insensitivity (GHI) subjects



Genes highlighted in blue are from our curated growth candidate gene list of 1305 genes. Yellow genes are those genes found within in the CNV regions.

IPA upstream analysis of unexplained short stature subjects identified CLOCK as a plausible common upstream regulator in 2 unexplained short stature patients (**Figure 4.3**). CLOCK is a transcription regulator for genes *GALNT11*, *CHRNA7* and *KLF13* within the CNV regions. *GALNT11* lies within the 7q36 region duplicated in Patient 29 and both *KLF13* and *CHRNA7* are found within the 15q13 region duplicated in Patient 30.

Figure 4.3 Pathway enrichment analysis identified CLOCK as a transcription regulator enriched within the CNV gene list of the unexplained short stature subjects

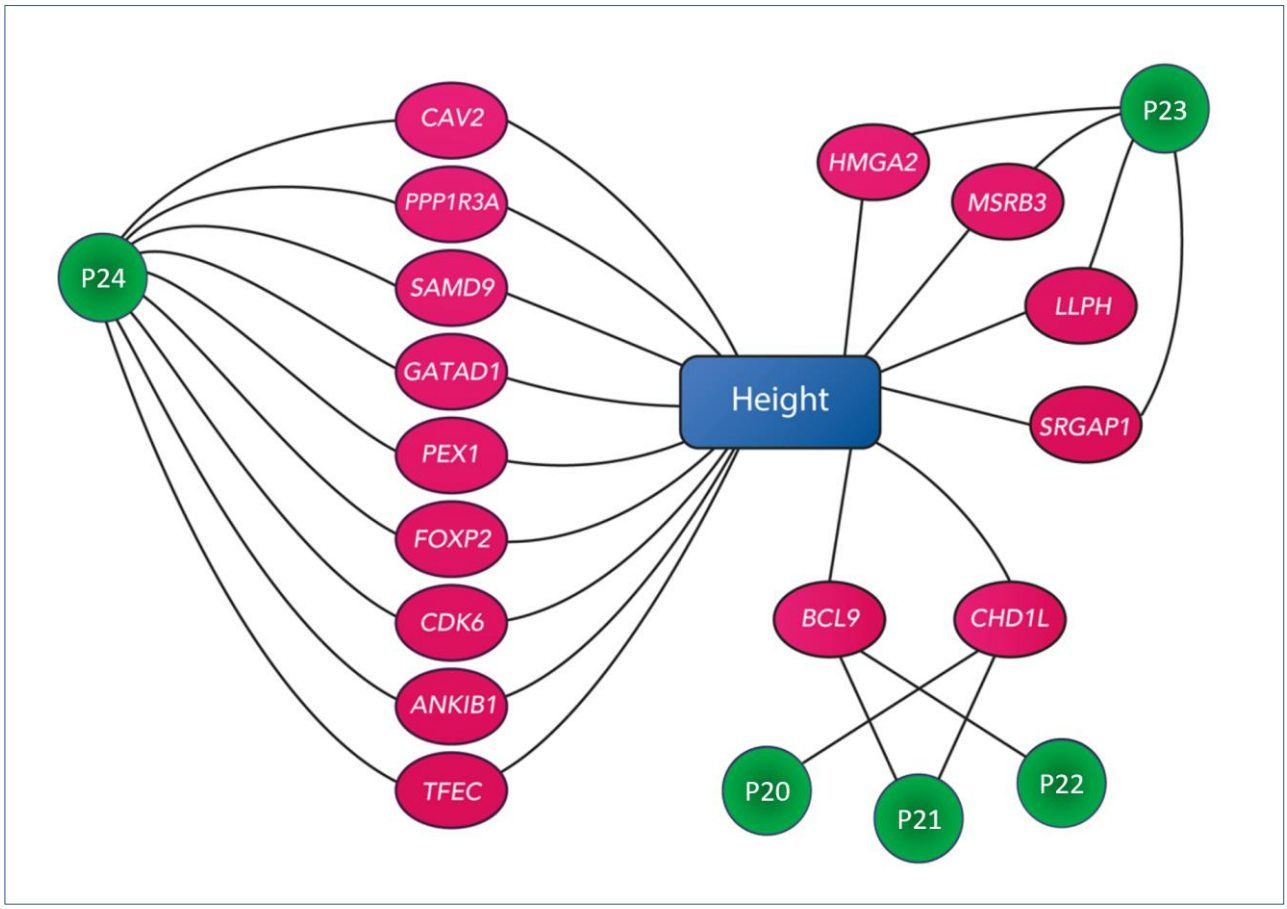


Genes highlighted in yellow are contained within our patient CNV regions.

4.3.2 Identification of candidate genes

Candidate growth genes were identified in 9/11 (82%) CNV subjects (**Table 4.3**). Some of the CNV regions harboured strong candidate growth genes, such as *HMGA2* in the 12q14 deletion (Patient 23). In other regions, candidate genes were proposed based on their role in known growth pathways or the presence of growth restriction phenotypes in mouse models. The limitation of this approach is that this would only identify protein-coding candidate genes and not imprinting control regions which may affect growth for example in chromosome 7 alterations. Genes in our patients CNVs associated with height in the NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/) are shown in **Figure 4.4**.

Figure 4.4 Genes in our patients CNVs associated with Height in GWAS Catalog



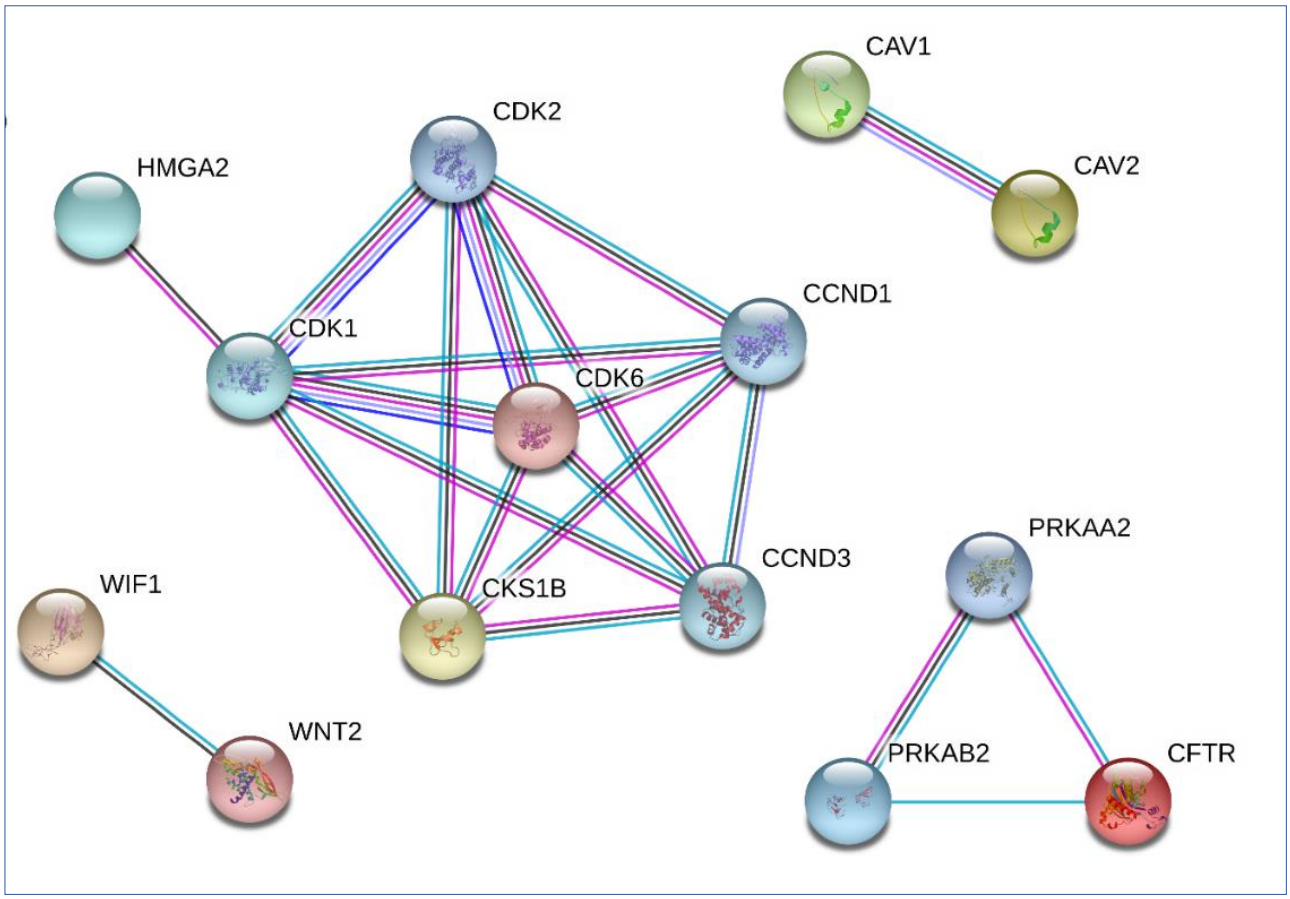
P, patient; Genes are shown in blue and connected to the patient(s) in which the CNV containing that gene was identified. The NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/) was used to examine all protein-coding genes within the CNV regions (class 3-5) in our patients. Of these CNV genes, those that have loci with associations to height were identified.

4.3.3 *In silico* protein-protein interaction analysis of candidate gene list

STRING analysis identified interactions (direct and via intermediate proteins) between several genes in the CNV regions (candidate gene list) of the GHI cohort (38 genes) (**Figure 4.5**). Interactions were identified between several genes involved in regulating cell cycle progression. *Cyclin-dependent kinase 6* (*CDK6*, Patient 24) and *Cyclin-dependent kinases regulatory subunit 1B* (*CKS1B,* Patient 25) interact directly and also with *High Mobility Group AT-Hook 2* (*HMGA2*, Patient 23) via *Cyclin-dependent kinase 1* (*CDK1*). *Cyclin-dependent kinase 2* (*CDK2*), *Cyclin D1* (*CCND1*) and *Cyclin D3* (*CCND3*) also interacted with both *CDK6* (Patient 24) and *CKS1B* (Patient 25). Direct interaction was identified between *Wnt Inhibitory Factor 1* (*WIF1*, Patient 23) and *Wnt Family Member 2* (*WNT2*,

Patient 24). *Protein Kinase AMP-activated Non-catalytic Subunit B2* (*PRKAB2*, Patients 20-22) and *Cystic Fibrosis Transmembrane Conductance Regulator* (*CFTR*, Patient 24) interacted directly and also via *5'-AMP-activated Protein Kinase Catalytic Subunit Alpha-2* (*PRKAA2*).

Figure 4.5 Protein-protein interactions of the candidate genes in the CNV regions of the Growth Hormone Insensitivity subjects

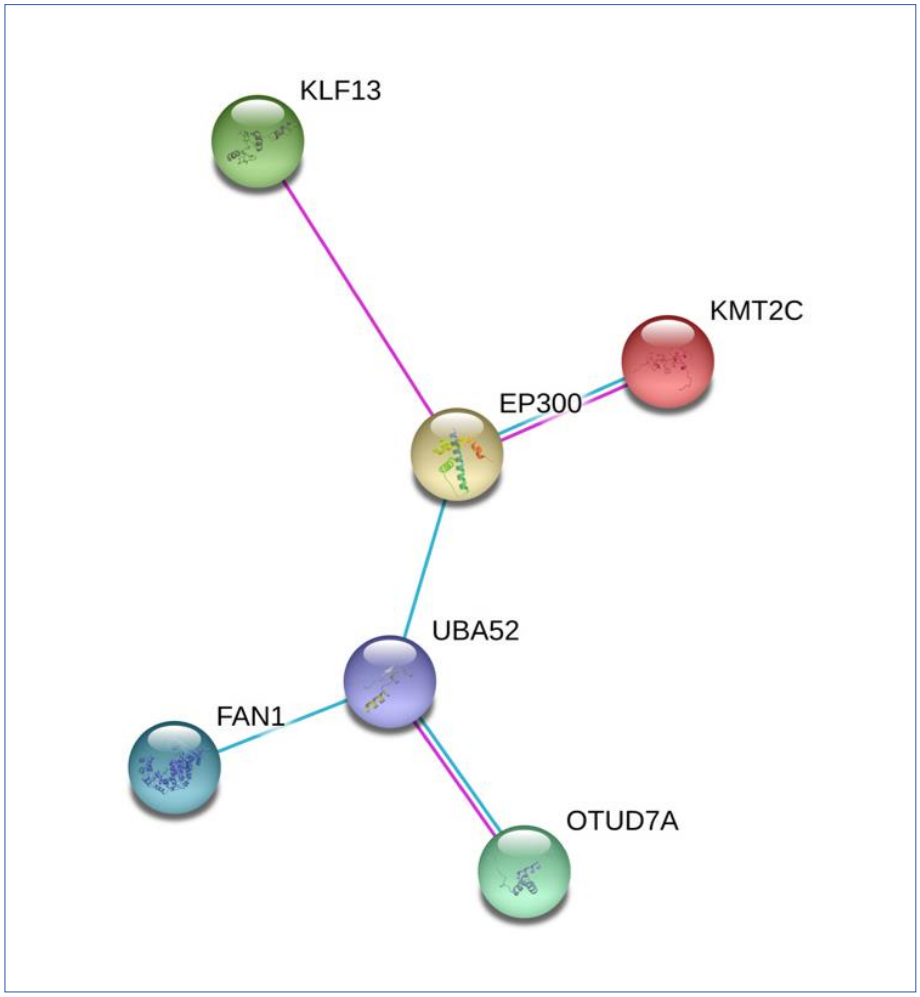


The STRING database (https://string-db.org/) explored protein-protein interactions of the curated list of candidate genes for our Growth Hormone Insensitivity patients with CNVs. Default settings were used with the exception of interaction sources, where text mining and neighbourhood sources were excluded. Direct interactions between two candidate genes were explored and also those via intermediate proteins.

STRING analysis identified interactions (direct and via intermediate proteins) between several genes in the CNV regions (candidate gene list) of the unexplained short stature cohort (7 genes), (**Figure 4.6**). *Kruppel-like Factor 13* (*KLF13*, Patient 30) interacts with *Lysine Methyltransferase 2C* (*KMT2C*, Patient 29) via *E1A-associated protein p300* (*EP300*). Additionally, *OUT Domain-containing Protein*

7A (*OTUD7A,* Patient 30) and *Fanconi-associated Nuclease 1* (*FAN1*, Patient 30) interact with *EP300* via *Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1* (*UBA52*).

Figure 4.6 Protein-protein interactions of the candidate genes in the CNV regions of the unexplained short stature subjects



The STRING database (https://string-db.org/) explored protein-protein interactions of the curated list of candidate genes for our unexplained short stature patients with CNVs. Default settings were used with the exception of interaction sources, where text mining and neighbourhood sources were excluded. Direct interactions between two candidate genes were explored and also those via intermediate proteins.

4.4 Discussion

Growth Hormone Insensitivity has classically been associated with a Laron phenotype causing severe postnatal growth failure, dysmorphic features and metabolic abnormalities. However with advancing genetic techniques, there is increasing recognition of a spectrum of genetic defects responsible for milder GHI phenotypes and less severe short stature (21). These include non-classic *GHR* defects such as dominant-negative heterozygous variants, the 6Ψ *GHR* pseudoexon and other defects in the GH-IGF-1 signalling pathway (*STAT5B, IGFALS*, *PAPPA2*, *IGF1*, *IGF1R*, *IGF2*) (60,79,86,88–90,104,162,183,251,252). Our findings suggest that CNVs contribute to the genetic aetiology of undiagnosed patients with mild to moderate GHI phenotypes. Haploinsufficiency or duplication of gene(s) in the affected region(s) may be sufficient to affect growth but not as severely as total loss of function/homozygous mutations seen in 'classic' GHI.

Whilst 5/11 (45%) CNV patients with GHI had clinical features of SRS, only Patient 22 strictly fulfilled the Netchine-Harbison SRS clinical scoring system criteria (NH-CSS) with a score of 3/6 and a genetic diagnosis previously associated with SRS. The NH-CSS is the only SRS scoring system that was developed using prospective data (Azzi *et al*. 2015). As with many tools based on clinical assessment there is a degree of subjectivity, such as protruding forehead, but scoring criteria for each domain are well-defined. These criteria were highly sensitive (97.9%) for detecting classic 11p15 LOM and matUPD patients, but specificity was low (36%). They also had a high negative predictive value (88.9%), hence those not fulfilling the criteria were unlikely to have either of the two classic molecular defects (130).

For this reason, the NH-CSS may not be sensitive enough to detect patients with milder forms of SRS due to other genetic abnormalities. The latest SRS consensus paper recognises over 30 different

CNVs identified in patients with suspected SRS. These patients often do not fulfil the NH-CSS criteria and are frequently associated with more severe learning difficulties than classic SRS. Current recommendations are to manage the patient according to their CNV rather than SRS (131). Whilst our patients with CNVs did not have a high incidence of learning difficulties, my results are in keeping with the observation that patients with suspicion of SRS and CNVs often have milder phenotypes. This is an interesting finding, since 40% patients with clinical SRS have no genetic diagnosis, and exploring CNV regions in detail may enable the discovery of novel genes and pathways responsible for SRS-like phenotypes (131). My findings potentially expand the SRS-like spectrum and may lead to new SRS subtypes.

A significant challenge in assessing the pathogenicity of CNVs is their variable penetrance in the population. There may be several contributing factors which account for this, including the background genetic variation of an individual and/or epigenetic mechanisms such as imprinting, expression or regulatory variation among genes in the affected CNV region. It is also possible that in some individuals with genetic deletions, recessive variants residing on the single remaining allele are unmasked (253–255). If parents harbouring the same deletion / duplication do not share the same phenotype as the index case, it is possible that the CNV is benign. However, in the case of 1q21 deletions, Mefford *et al*. assessed 4,737 control samples and no 1q21.1 deletions were identified (256,257).

I identified 1q21 deletions in 3 GHI patients, including 2 siblings. The 1q21 deletion syndrome is recognised to cause a variable degree of dysmorphism, learning difficulties, SS in 50% and autism (OMIM #612474). Dysmorphic features include microcephaly, frontal bossing, deep-set eyes, epicanthal folds, large nasal bridge, long philtrum and a high-arched palate. 1q21 deletions can occur *de novo*, but many are inherited in an autosomal dominant manner from a parent (256–258).

1q21 deletion is characterised by variable penetrance, with parents frequently having no or very mild phenotype (256). Potential causes of the recognised phenotypic variability have been sought but not yet delineated. Sequence analysis of *GJA5* and *GJA8* genes within the affected regions of 11 1q21 deletion carriers yielded no data to support the unmasking of recessive variants as a cause of the variable phenotypes observed. Preliminary data from an affected 1q21 deletion patient and her unaffected mother, who also carried the deletion, suggested that differences in the methylation status of the non-deleted 1q21.1 locus also did not contribute to the phenotypic variability. Parent-of-origin studies show both maternal and paternal transmission of the deletion, therefore it is also unlikely that imprinting plays a role (256).

The genes in region 1q21 responsible for the growth failure are not established, but we identified several candidate genes within the commonly deleted region such as Chromodomain-helicase-DNA-binding protein 1-like *(CHD1L). CHD1L* regulates chromatin relaxation/cell cycle progression and 50% knockdown in glioma cells results in significantly reduced proliferation (259). B-cell CLL/lymphoma 9 protein *(BCL9)* is an essential co-activator in the Wnt/β-catenin signalling pathway (260). 5'-AMP-activated protein kinase *(PRKAB2)* and Flavin containing monooxygenase 5 *(FMO5)* are both within genomic loci associated with height in genome-wide association studies (197). In addition, loci in BCL9 and CHD1L were associated with height in the GWAS Catalog (https://www.ebi.ac.uk/gwas/; p-value=7.00E-13 and 7E-20 respectively).

Patient 23 had a 12q14 deletion and there is strong evidence to suggest that haploinsufficiency of High-mobility group AT-hook 2 *(HMGA2)* is responsible for the observed growth failure in patients with deletions in this region (249). Additionally, sub microscopic 12q14 deletions, spanning only part of *HMGA2,* have been reported in three patients with SS and SRS features (185). An association study of 155 idiopathic SS patients and 318 normal stature controls examined three *HMGA2* single-

nucleotide polymorphisms (SNPs) and concluded they contributed to short stature susceptibility (204). Heterozygous *HMGA2* point mutations have been identified in 2 subjects with SS and SRS-like features, causing a frameshift and a premature stop codon, respectively (184). A heterozygous 7bp intronic deletion causing aberrant splicing of *HMGA2* has also been described in a patient with a similar phenotype (261). *HMGA2* is thought to function as an upstream regulator of IGF2 but the mechanism is not well characterised or understood (184). Loci in *HMGA2, LLPH, MSRB3* and *SRGAP1* were reported in the GWAS catalog to be associated with height with minimum p-values of 1.00E-287, 3.00E-21, 4.00E-39, 3.00E-12 respectively (GWAS catalogue, Genome Wide Association Studies catalogue, https://www.ebi.ac.uk/gwas/).

Patients 24 and 28 had CNVs affecting 7q21 and one additional region. SRS patients with Chromosome 7 defects have milder phenotypes compared with *ICR1* hypomethylation patients and less asymmetry (244,262). Patient 24 had 7q21 and 7q31 deletions and fulfilled 2/6 NH-CSS (SGA and postnatal growth failure) in addition to a triangular face, low set ears and delayed motor development. Both deletions are close to the 7q21.3 (*PEG10/SGCE*) and 7q32 (*MEST*) imprinting clusters. Several genes within the CNV regions identified in patient 24 had loci associated with height in the GWAS Catalog: *TFEC* (p-value 5.00E-07), *ANKIB1* (p-value 2.00E-15), *CDK6* (p-value 3.00E-240), *FOXP2* (p-value 2.00E-15), *PEX1* (p-value 1.00E-08), *GATAD1* (p-value 1.00E-08), *SAMD9* (2.00E-25), *PPP1R3A* (p-value 2.00E-08) and *CAV2* (p-value 1.00E-11). Patient 28 had duplications of 7q21 and Xp22 and fulfilled 2/6 NH-CSS (SGA with postnatal growth failure). Xp22 duplications have been associated with SRS and SGA. The duplication identified in the patients with SRS features and SGA included *SHOX* and *SHOX* enhancer region (*PAR3-12*), respectively (237,242). However, the CNV identified in our subject does not encompass either of these regions and both duplications were larger than the CNV identified in our patient suggesting the mechanism may be independent of *SHOX* function.

Patient 25 had 5q12 deletion. The 5q12 deletion syndrome has been reported in several patients with postnatal growth failure and developmental delay (250). This region harbours two interesting candidate genes that could be responsible for the growth failure. *Cyclin-dependent kinases regulatory subunit 1 (CKS1B)* has an essential role in cell cycle regulation. *Cks1*-depleted breast cancer cells exhibit slow G1 cell cycle progression and G2-M arrest due to blocked mitotic entry (263). *Dimethyladenosine transferase 1 homolog* gene *(DIMT1)* is a methyltransferase that is overexpressed in several cancers (264,265). The functional roles of both genes, in cell cycle regulation and ribosomal biology respectively are likely to be critical for normal linear growth and thus haploinsufficiency may lead to postnatal growth failure.

Prader-Willi (PWS) and Angelman syndrome are classically caused by deletions of different parental origin involving the distal breakpoint BP3 and proximally placed breakpoints BP1 or BP2 at 15q11 (266). Patient 26 had a smaller 15q11 deletion encompassing only BP1 and BP2. 15q11.2 deletions (including just BP1 and BP2) have variable phenotypes and expressivity. Short stature and unspecified dysmorphic features are reported in 10% and 39% patients, respectively. Feeding problems are not reported, in contrast to the significant feeding difficulties seen in our patient. Interestingly feeding problems are often a feature of Prader Willi syndrome (266).

Patient 27 was the only GHI subject with a chromosomal duplication. The Xq26.3q27.2 duplicated region harbours *SRY-related HMG-box 3 (SOX3)*, which encodes a transcription factor involved in the regulation of embryonic development. *SOX3* under- and over-expression in males causes multiple pituitary hormone or isolated growth hormone deficiency associated with infundibular hypoplasia, ectopic/undescended posterior pituitary and abnormalities of the corpus callosum with or without intellectual disability (267). Our patient had normal brain MRI and sufficient GH secretion on stimulation testing but a low baseline IGF-1 (SDS -2.7). Interestingly, a similar female patient is

reported with poor growth and low IGF-1. It is suggested her poor growth is due to GH deficiency, but formal GH stimulation testing was not performed (268). Another female patient with a slightly larger duplication is reported in a cohort of patients with SRS features. This patient had a broad nasal bridge similar to Patient 27, with a triangular face and relative macrocephaly. The duplication begins at Xq25 (genomic co-ordinates 129,132,238-139,650,444) whilst duplication identified in our subject begins at Xq26 (co-ordinates 134,842,275-141,407,613) (248). The overlapping regions include *SOX3* in addition to several other genes.

Chromosome 7q36 rearrangements are very rare, with deletions being more prevalent than duplications. 70% 7q36 deletion individuals exhibit growth retardation (269), however, few duplications have been reported. One patient with a larger 7q36 duplication (151,602,419-152,956,632) compared to our subject (Patient 10; 151,748,853-152,032,715) was born SGA with no catch-up growth by 22 months of age. He also had developmental delay and multiple congenital abnormalities including a cardiovascular malformation, sensorineural hearing loss, myopia, astigmatism, cryptorchidism, hypospadias, microphallus and dysmorphic facial features (270). One interesting candidate gene, *GALNT11,* is common to both duplicated regions. *GALNT11* encodes Polypeptide N-acetylgalactosaminyltransferase 11, which O-glycosylates *NOTCH1* and activates the Notch signalling pathway (271). Abnormal *GALNT11* dosage (either too much or too little) may alter Notch signalling and adversely affect growth in these patients. However, Patient 29's duplication was inherited from his mother who is of normal stature. This suggests either variable penetrance or that this CNV is not responsible for the observed phenotype in our patient. Variable penetrance could be effect of the synergistic activity of two separate "mild" defects, for example a deletion or duplication in addition to another variant in a growth genes. This oligogenic model is a feasible mechanism for growth failure.

A 3p22 deletion and 15q13 duplication were identified in Patient 30. These CNVs have not previously been associated with short stature or SRS features. Interestingly, this 15q13 duplication was inherited from her father who has significant short stature (height SDS -3.4). Within this region lies several candidate genes. Both *FAN1* and *OTUD7A* cause a growth restricted phenotype in their respective mouse models, whilst *CHRNA7* and *KLF13* were identified from our Ingenuity Pathway Analysis.

Pathway analysis identified genes with links to established or novel growth pathways in the CNV regions of our subjects. Six GHI individuals had CNVs harbouring genes involved in the canonical WNT pathway. *WIF1* in the 12q14 deleted region of Patient 23*, WNT2* in the 7q31 deletion of Patient 24, *SOX3* in the Xq26 duplicated region of Patient 27 and *BCL9* in the 1q21 deleted region of Patients 20, 21 and 22. The WNT pathway determines cell fate and is a key regulator of cell proliferation (272). Aberrant Wnt activity leads to uncontrolled cell growth and oncogenesis and is a potential novel therapeutic target for cancer (273). Abnormal copy numbers of these genes may impair this pathway and thus normal linear growth. The 12q14 deletion identified in Patient 23 includes *WIF1* and *HMGA2,* which is thought to be the most crucial gene regulating growth in this region. However, it is possible both genes contribute important regulatory effects on growth.

CLOCK regulates the transcription of *GALNT11*, *CHRNA7* and *KLF13* genes residing within the CNV regions identified in the unexplained short stature subjects. This is particularly interesting as IGF-1 regulates clock gene expression and function as a zeitgeber for cellular hypothalamic circadian rhythms (274). In addition to its central functional role in the regulation of circadian rhythm, CLOCK modulates G2-to-M cell cycle transition facilitating cell cycle progression and proliferation (275).

Whilst genome wide association studies have successfully identified numerous loci associated with height, most SNPs have regulatory roles often on genes other than those in closest proximity. This is an important limitation of using GWAS data and SNP-to-gene linking strategies are being developed to help establish the gene(s) regulated by each identified SNP (276). Experimental evidence is required to validate the candidate growth genes, interactions and biological pathway enriched in our cohort.

Following CNV identification and bioinformatic analysis of our cohort, I identified *CHD1L* and *HMGA2* as the strongest candidate growth genes. In addition to patient deletions of regions containing these genes, I identified several other patients in our cohort with point mutations occurring within these candidate genes. Further details of these patients, their genetic changes and my functional analysis of these variants are provided in **Chapter 5**.

Chapter 5. Novel candidate genes identified following CNV analysis

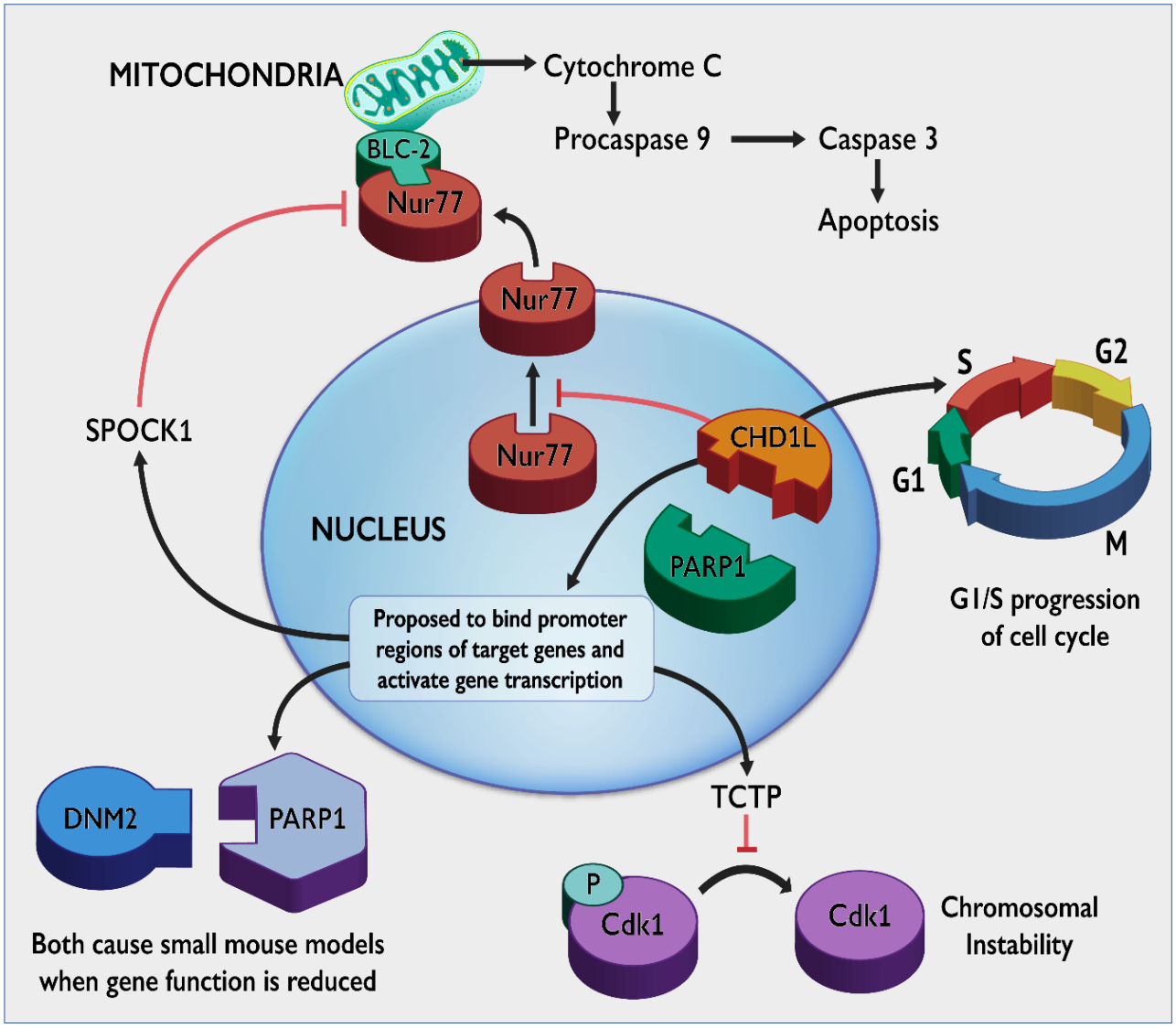
5.1 Chromodomain-helicase-DNA-binding protein 1-like (*CHD1L*)

5.1.1 Introduction

Copy number variant analysis (Chapter 4) identified 3 patients with 1q21 deletion in our undiagnosed cohort, suggesting this region may harbour gene(s) with important roles in growth. Patients with 1q21 syndrome show a variable phenotype, with short stature observed in around 50% of patients (158). Interrogating the genes harboured within the common genomic region deleted in all 3 of these patients, I identified 4 further patients in our cohort with short stature and heterozygous exonic missense variants in *CHD1L, Chromodomain Helicase DNA Binding Protein 1 Like*. *CHD1L* is within the region deleted in 1q21 syndrome, so the hypothesis would be that only one functional copy or haploinsufficiency is sufficient to impair normal growth. Heterozygous mutations of the *CHD1L* gene may render one copy of the gene non-functional and thus explain the similar phenotype observed in all 7 patients (3 individuals with a 1q21 deletion and the 4 patients with heterozygous exonic missense variants in *CHD1L*).

CHD1L was considered a good candidate growth gene due to its recognised ubiquitous role in DNA replication, repair and transcription and expression. It has been shown to have a key role in G1/S cell cycle progression in glioma cells, with 50% knockdown of *CHD1L* significantly inhibited cell growth (259). The role of *CHD1L* in cell growth and prevention of apoptosis is shown in **Figure 5.1**. Overexpression of *CHD1L* in a transgenic mouse model has been shown to promote cell proliferation (277). A *CHD1L* knockout mouse model is not reported. *CHD1L* is named due to its similarity to *CHD1.* Heterozygous *CHD1* mouse models show growth restriction and homozygous mutations are lethal.

Figure 5.1 The role of *CHD1L* in cell growth and prevention of apoptosis



Poly (ADP-ribose) polymerase 1 (PARP1) regulates gene expression and activates *CHD1L*(278,279). Heterozygous missense variants in the macro domain impair interaction of *CHD1L* and *PARP1* (280). The macro domain also interacts with Nuclear receptor 4A1 (Nur77) to inhibit apoptosis (281). Whilst *CHD1L* has been shown to affect cell growth and cell cycle progression, the genes responsible are not well established. Regulation of both Testican-1 (encoded by the *SPOCK1* gene) and Translationally controlled tumour protein (TCTP) by *CHD1L* was identified by studies looking directly for these interaction in cancer cell lines(282,283). Original diagram.

5.1.2 Details of variants in the patients

The phenotypic details and biochemistry for these patients is shown in **Table 5.1** and **Table 5.2**, respectively. The details of the genetic variants are described in **Table 5.3**.

Table 5.1 Phenotypic details of the 7 patients with variants affecting *CHD1L*

Patient	Bp position/base change (Grch37)	BW SDS	Height SDS	BMI SDS	Age (years)	Phenotype
20*	Chr1: (146564742-147735011)x1	-1.59	-3.64	-1.02	3.8	Small triangular face and high arched palate. Slender frame with clinodactly. Feeding difficulties, food refusal, gastrostomy, recurrent infections in early childhood, hypoglycaemic episodes. On autistic spectrum.
21*	Chr1: (146641600-147735011)x1	-1.68	-1.58	NK	1.1	Feeding difficulties, recurrent infections and poor weight gain/growth. Dyslexia and language delay, recurrent ear infections.
22	Chr1: (145987155-147735011)x1	-0.38	-3.74	-1.33	9.1	No notable dysmorphic features.
31	146757987, G>C	-0.76	-2.63	+0.16	6.1	Dental caries. No notable dysmorphic features.
32^	146759326, A>G	+1.38	-2.84	-1.23	13.7	No notable dysmorphic features.
33^	146759326, A>G	+1.07	-1.56	-3.13	11.2	No notable dysmorphic features.
34	146765379, A>G	+0.14	-4.43	+1.00	4.0	Born at 28/40 gestation. Mild delay in gross motor skills but fine motor skills and speech very good for age. Large head circumference, small hands and feet, protuberant abdomen and convergent squint.

Patients indicated by * and ^ are siblings. Bp, base pair; BW, Birth weight; SDS, Standard Deviation Score; BMI, Body Mass Index; HC, Head circumference; NR, Normal range; NK, not known.

Table 5.2 Biochemical details of the 7 patients with variants affecting *CHD1L*

Patient	Bp position/base change (Grch37)	Basal GH (ug/L)	GH peak (ug/L)	Basal IGF-1 (ng/ml)	IGF-1 SDS	Bone age
20*	Chr1: (146564742-147735011)x1	12.7	98	34.3	-1.9	1.5 years (at 2.3 years)
21*	Chr1: (146641600-147735011)x1	NK	18	NK	-2.4	NK
22	Chr1: (145987155-147735011)x1	15.8	20.8	48	-2.3	6.6 years (at 9.1 years)
31	146757987, G>C	44.8	18.8	25	-2.4	5.0 years (at 5.5 years)
32^	146759326, A>G	9.4	21.2	98	-2.9	10.5 years (at 12.8 years)
33^	146759326, A>G	NK	NK	NK	NK	NK
34	146765379, A>G	1.72	9.21	<25	-2.5	2.2 years (at 3.10 years)

Patients indicated by * and ^ are siblings. Bp, base pair; GH, growth hormone; NK, not known.

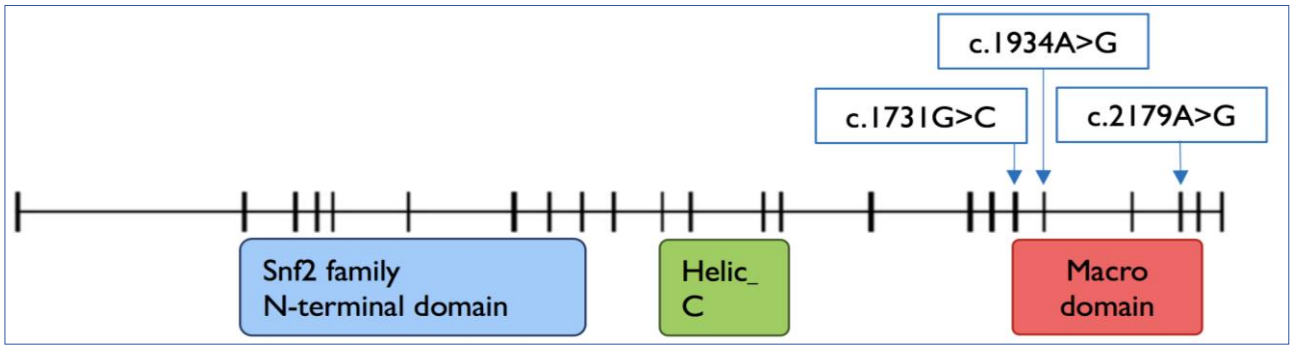
Table 5.3 Details of the genetic variants affecting *CHD1L* identified in our cohort

Patient number	Genetic change	Bp position/base change (GRCh37)	Region	Amino acid change	Protein change	Frequency	CADD score
20*	1q21 deletion, including *CHD1L*	Chr1: (146564742-147735011)x1	-	-	-	~240 cases reported worldwide	N/A
21*		Chr1: (146641600-147735011)x1	-	-	-		
22		Chr1: (145987155-147735011)x1	-	-	-		
31	Heterozygous exonic missense *CHD1L* mutation	146757987, G>C	Exon 18	c.1731G>C	p.W577C	0.00001% allele frequency in gnomAD	32
32^		146759326, A>G	Exon 19	c.1934A>G	p.H645R	0.00004% allele frequency in gnomAD	<10
33^		146759326, A>G	Exon 19	c.1934A>G	p.H645R	0.00004% allele frequency in gnomAD	<10
34		146765379, A>G	Exon 21	c.2179A>G	p.I727V	0.00254% allele frequency in gnomAD	14.34. Predicted by HSF3 to create a new exonic donor site

Patients indicated by * and ^ are siblings. Bp, base pair; CADD score, Combined Annotation Dependent Depletion Score (https://cadd.gs.washington.edu/) (167); HSF3, Human Splicing Finder (http://umd.be/HSF3/)

CHD1L has not previously been reported to cause growth problems in humans. The 3 different missense variants in our patients all occur in the macro domain (**Figure 5.2**). *In vitro* studies show this key domain binds poly(ADP-ribose) (PAR) and interacts with chromatin associated PARP1. PARP1 regulates gene expression and activates CHD1L. An intact macrodomain is therefore essential for *CHD1L* function.

Figure 5.2 Schematic showing the location of the novel missense *CHD1L* variants



Locations of novel heterozygous missense CHD1L variants are shown in the blue boxes, with arrows highlighting their location within the gene. Helic_C, Helicase superfamily C-terminal domain. Vertical lines represent exons in the gene.

5.1.3 Functional analysis of patient variants

Given the location of the variants and the recognised interaction between PARP1 and CHD1L at this region, I hypothesised that the variants seen in our patients would impair their interaction. To test this hypothesis, I recreated mutant CHD1L vector constructs by site-directed mutagenesis (**Figure 5.3**) and performed co-immunoprecipitation assays with PARP.

5.1.3.1 Site-directed mutagenesis

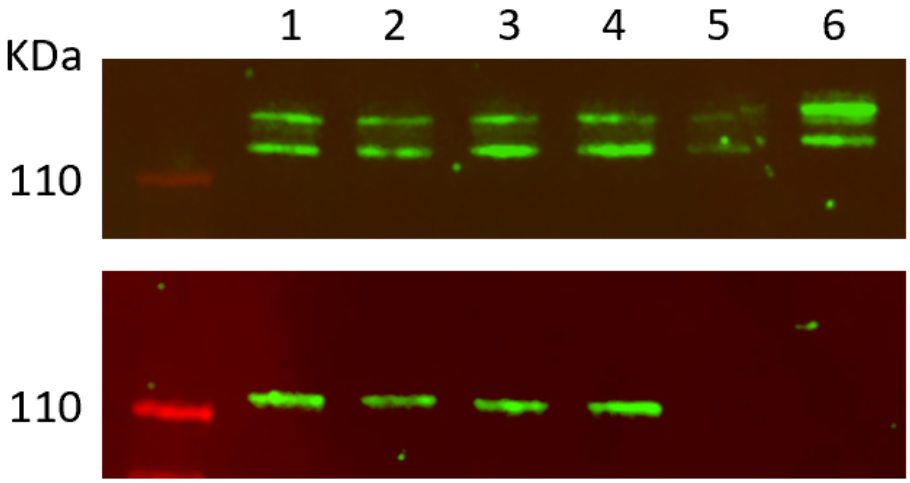
Figure 5.3 *CHD1L* variants created by site-directed mutagenesis

Sanger sequencing across the region of interest demonstrates the single base change created in each vector using site-directed mutagenesis, highlighted by the blue arrows. Mutant sequencing is shown in the top row compared to the wild-type in the bottom row. These mutations mimic the 3 missense *CHD1L* variants identified in patients in our cohort: 146757987G>C exon 18 variant identified in Patient 31; 146759326 A>G exon 19 variant identified in Patients 32 and 33 (siblings); 146765379 A>G exon 21 variant identified in Patient 34.

5.1.3.2 Co-immunoprecipitation

To determine if the interaction between PARP and CHD1L was impaired due to these variants in the macrodomain, I performed co-immunoprecipitation. The CHD1L constructs were FLAG tagged at the N terminal so the key output was the PARP signal of the isolated constructs. The CHD1L-PARP complexes were pulled down using FLAG beads. Western blot analysis of the whole cell lysates used as input for my co-immunoprecpitation are shown in **Figure 5.4** and the IP product are shown in **Figure 5.5**.

Figure 5.4 Input for co-immunoprecipitation

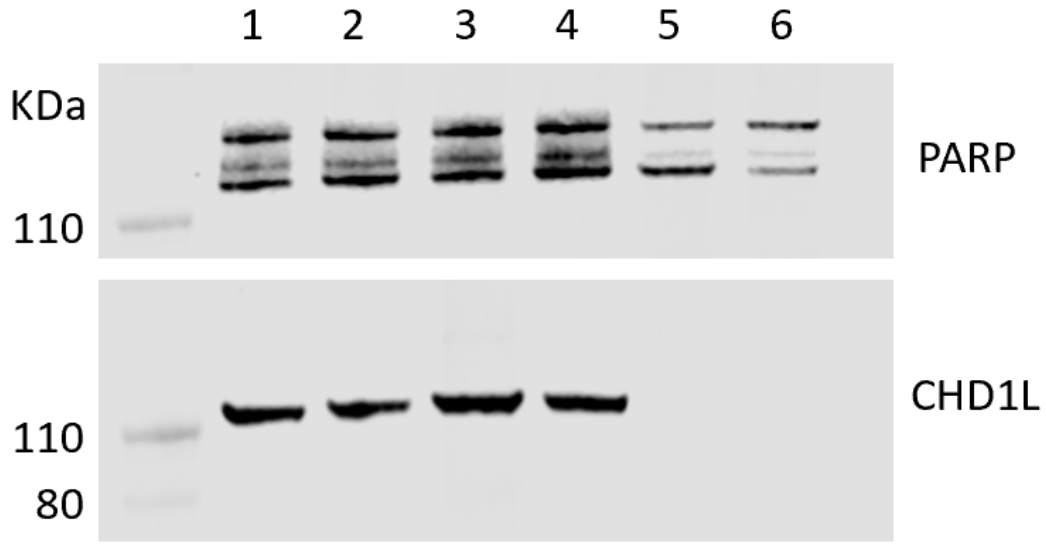


Lanes

1. WT CHD1L with 3 x flag tag + PARP1 with mcherry (positive control)
2. SDM e18 CHD1L with 3 x flag tag + PARP1 with mcherry
3. SDM e19 CHD1L with 3 x flag tag + PARP1 with mcherry
4. SDM e21 CHD1L with 3 x flag tag + PARP1 with mcherry
5. CHD1L missing macrodomain with 1 x flag tag + PARP1 with mcherry (negative control)
6. PARP1 with mcherry only (negative control)

Western blots of whole cell lysates from HEK293T cells transfected with CHD1L and PARP1 constructs as described. No signal is seen in lane 5 as the CHD1L antibody epitope recognises the macrodomain and the macrodomain is missing in this construct.

Figure 5.5 Western blot of PARP and CHD1L constructs immunoprecipitated using FLAG beads



Lanes
1. WT CHD1L with 3 x flag tag + PARP1 with mcherry (positive control)
2. SDM e18 CHD1L with 3 x flag tag + PARP1 with mcherry
3. SDM e19 CHD1L with 3 x flag tag + PARP1 with mcherry
4. SDM e21 CHD1L with 3 x flag tag + PARP1 with mcherry
5. CHD1L missing macrodomain with 1 x flag tag + PARP1 with mcherry (negative control)
6. PARP1 with mcherry only (negative control)

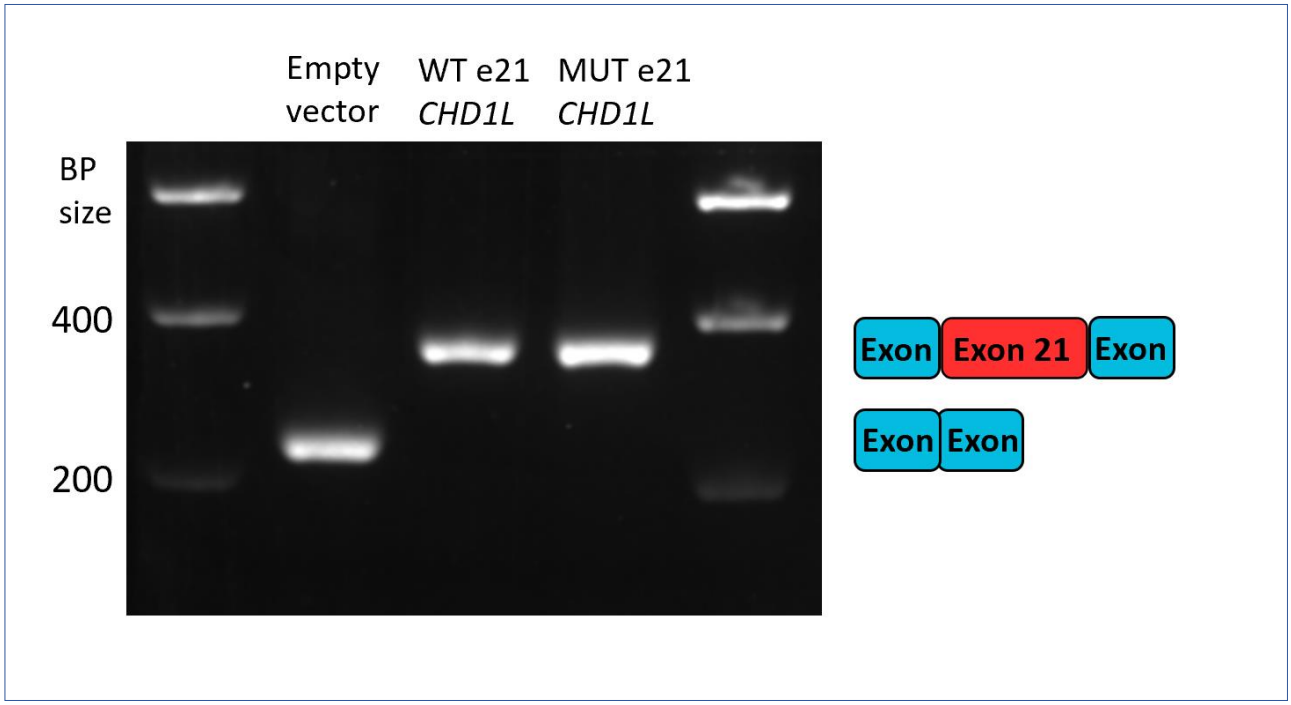
Western blot of constructs immunoprecipitated using FLAG beads. No signal is seen in lane 5 as the CHD1L antibody epitope recognises the macrodomain and the macrodomain is missing in this construct. Similar amounts of PARP protein appear to have been pulled down for the wild-type CHD1L construct and the 3 mutants, suggesting the interaction between PARP and CHD1L was not affected by these mutants.

The experiment successfully isolated constructs containing CHD1L and PARP. Overall, the interaction between PARP1 and the 3 mutant CHD1L constructs did not appear reduced compared to WT. The antibody for CHD1L recognised the macrodomain so did not detect the CHD1L construct missing the macrodomain (lane 5 in **Figure 5.4** and **Figure 5.5**). Unfortunately, the membrane was torn so the PARP transfer may have been suboptimal for the whole cell lysate in lane 5 for **Figure 5.4** but it was decided on balance not to repeat the experiment given the findings of my later experiments.

5.1.3.3 *In vitro* splicing assay

To assess whether the novel exon 21 variant identified in Patient 34 affected *CHD1L* splicing, I performed an *in vitro* splicing assay (as described previously in methods **Section 2.8**). I designed primers to amplify an 800bp DNA fragment of interest from Patient 34 and from a control sample with a wild-type sequence for this region. The cDNA PCR of the spliced products revealed bands of a similar size for both the control and patient sample (**Figure 5.6**). Sanger sequencing of the cDNA products confirmed a normal exon 21 sequence for both the patient and control sample, suggesting that this novel variant does not affect splicing of the *CHD1L* gene.

Figure 5.6 cDNA PCR of spliced *CHD1L* product



Gel electrophoresis of cDNA splicing products following the splicing assay using an exon trap vector (MoBiTec-Exontrap cloning vector pET01). Empty vector represents the sample with pET01 alone; WT e21 *CHD1L*, pET01 with 800bp of wild-type *CHD1L* exon 21 sequence inserted; MUT e21, pET01 with 800bp of Patient 34 sequence inserted (including the c.2179A>G variant in exon 21). The spliced products were amplified by PCR and visualized on a 2% agarose gel. Lane 1: A 250bp band is seen in empty vector sequence, as expected, representing the two exons of the exon trap vector. Lanes 2 and 3: A 370bp band is seen, which was confirmed by Sanger sequencing to contain the wild-type exon 21 sequence and the two flanking exons of the exontrap vector. BP, base pair; WT, wild-type; MUT, mutant.

5.1.4 Segregation

To help assess the likely pathogenicity of these novel *CHD1L* variants, I requested DNA samples from family members to assess if the phenotypes segregated with the variants of interest. It took some time to acquire these samples. After assessing these samples, the missense variants did not segregate with the growth phenotype. For example, Patients 32 and 33 inherited the *CHD1L* variant from their father, who was tall with a height SDS of +3.50.

5.1.5 Conclusions and discussion

Given the preliminary findings of my co-immunoprecipitation, splicing and segregation studies, I did not investigate these novel variants further. However, it should be noted that 1q21 deletion syndrome which affects this same region is recognised to have variable penetrance and often the parent from whom the deletion was inherited is of normal height. The reason for this is not well understood. It is possible that the 1q21 deletion/*CHD1L* mutation may only be one component of an oligogenic growth picture in these patients.

The macro domain of CHD1L also has several other important interactions, such as with Nur77 to inhibit apoptosis (281). It is therefore possible that these variants are impacting interactions not assessed by my co-immunoprecipitation experiments which focused specifically on CHD1L and PARP1 interaction.

5.2 High-mobility group AT-hook 2 (*HMGA2*)

5.2.1 Introduction

From my copy number variant analysis described in **Chapter 4**, a previously undiagnosed patient in our cohort (Patient 23) was diagnosed with 12q14 deletion syndrome which led to interrogation of the genes in this region. The clear candidate growth gene in this region is *High mobility Group AT-hook 2, HMGA2*. This is a gene newly recognised to cause growth failure and a Silver Russell -like phenotype. Silver Russell Syndrome is a genetically heterogenous disorder and 30% of patients with clinical SRS have no genetic diagnosis (244). Previous research at our centre has recognised the clinical and biochemical overlap between SRS and patients with defects in the GH/IGF-1 pathway (109).

HMGA2 deficiency in mouse models causes a pygmy phenotype. *HMGA2* is completely deleted in the transgenic insertional mutant $pg^{T"N4oAcha}$ (A/A), whereas in the spontaneous mutant *pg* the 5' sequences and the first two exons of *HMGA2* are absent. These pygmy mice show growth failure despite sufficient growth hormone levels (284,285). This is especially interesting as our patient cohort show growth failure despite having sufficient growth hormone levels.

A non-synonymous *HMGA2* variant affecting the first AT-hook domain (c.83G>A; P.G28E) has recently been identified in Shetland ponies. This mutant was found to impair DNA binding in an electrophoretic mobility shift assay, suggesting impaired DNA binding *in vivo* due to this variant may be responsible for the reduced growth and small stature in these ponies. Interestingly, height measurements were taken of 110 genotyped Shetland ponies and the height difference between the 3 genotypes (heterozygous, homozygous and wild-type) was statistically significant ($p<2.5 \times 10^{-4}$). The generated correlation plot suggested a mean reduction of height of 9.5cm per copy of the

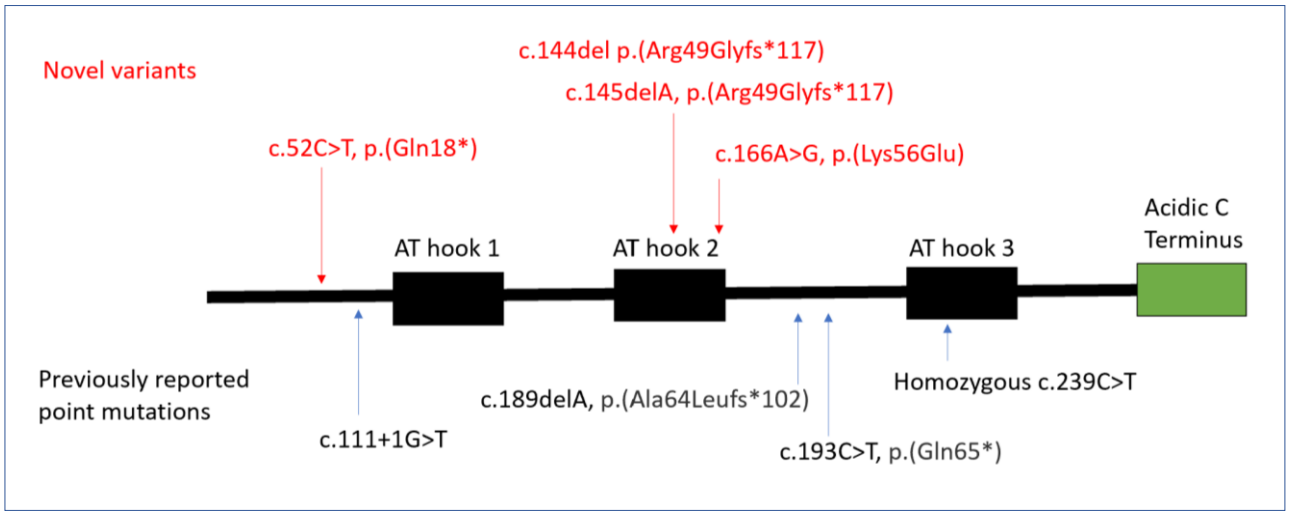
mutant c.83G>A allele. Upon testing several horse and pony breeds, this c.83G>A variant occurred exclusively in small pony breeds supporting its critical role in growth (204).

HMGA2 is an important regulator of cell growth, apoptosis, differentiation and malignant transformation (286). The HMGA2 protein contains 3 AT-hooks which are DNA binding domains that bind to AT rich sequences in the minor groove of the DNA helix (287). By binding to the DNA, they are able to modify the DNA conformation and facilitate the binding of transcription factors. The HMGA proteins are able to alter chromatin structure and interact directly through protein-protein interactions with transcription factors, influencing gene transcription. HMGA2 is highly expressed in embryonic tissue, but expression is not detectable in most adult tissues (286). Overexpression of HMGA2 is found in a variety of human cancers (288,289). Common variants in proximity to the 3'UTR region of *HMGA2* have been strongly associated ($<10^{-10}$) with childhood and final height in several genome wide association studies (290–293). *HMGA2* polymorphisms have also been shown to contribute to idiopathic short stature (294).

Despite the strong evidence of the crucial growth role of *HMGA2* across the species, there is no clear experimental data on the mechanism of action of *HMGA2* in human linear growth.

To date, four point mutations in *HMGA2* have been reported in the literature. The location of these and our patient variants are shown in **Figure 5.7**.

Figure 5.7 Reported *HMGA2* variants and our patient variants



The basic structure of *HMGA2* is shown, with 3 AT hook domains and an acidic C-terminus. Novel variants identified in our cohort and collaborators are shown in red above the *HMGA2* structure. Variants reported in the literature are shown in black below the *HMGA2* structure. c.111+1G>T and Homozygous c.239C>T variants are reported in Meyer *et al.* (295). c.189delA, p.(Ala64Leufs*102) and c.193C>T, p.(Gln65*) variants are reported in Abi Habib *et al.* (188).

5.2.2 Details of patient variants

We identified four novel heterozygous variants in *HMGA*2; two from our own undiagnosed short stature cohort and two referrals which were identified by our collaborators in The Netherlands. I identified the variant in Patient 15 upon assessing our custom gene panel data, and the variant in Patient 35 upon assessing whole exome sequencing data. The variants in Patients 36 and 37 were identified on whole exome sequencing data by our collaborators. The phenotypic details and NH-CSS for these patients and those with previously reported point mutations in *HMGA2* are shown in **Table 5.4** and **Table 5.5,** respectively. The biochemistry and the genetic details of the novel *HMGA2* variants identified are described in **Table 5.6** and **Table 5.7** respectively. All previously reported variants are heterozygous except for the homozygous missense c.239C>T mutation identified in two siblings by Mayer *et al.* (295). These siblings had a much more severe growth phenotype than the patients with heterozygous mutations, with severe pre- and postnatal growth failure, low BMI and microcephaly.

Table 5.4 Phenotypic details of patients with *HMGA2* variants

Patient	Heterozygous *HMGA2* variant	Sex	BW SDS	Height SDS	BMI SDS	HC SDS	Age (years)	Phenotype	Segregation
Novel *HMGA2* variants									
15	c.145delA, p.(Arg49Glyfs*117)	F	-3.48	-3.71	-1.21	NK	0.5	Relative macrocephaly, puffy hands and feet noted at birth.	Not known – no parental samples
35	c.166A>G, p.(Lys56Glu)	F	-3.78	-3.9	-3.02	-4.9	5.8	Triangular face, nasal, high-pitched voice. High arched palate. Very petite looking. Precocious puberty, Feeding problems, mild developmental problems (inattention and poor writing/reading skills).	Inherited from mother who has short stature (height SDS -3.52)
36	c.52C>T, p.(Gln18*)	F	-2.7	-3.4	NK	NK	3.0	Triangular face, relatively big forehead. Ophthalmic issues: hypermetropic and astigmatism	Inherited from affected mother and maternal grandfather. Mother has short stature (height SDS -3.7) and similar facial features.
37	c.144delC p.(Arg49Glyfs*117)	F	NK	NK	NK	NK	NK	IUGR and SRS like facial features.	Adopted
Previously reported *HMGA2* variants									
38	c.111+1G>T	M	-2.6	-3.2	-3.1	-3.3 (at birth)	3.7	IUGR, protruding forehead, microcephaly, feeding difficulties.	*De novo* variant
39	c.189delA, p.(Ala64Leufs*102)	M	-3.1	-3.0	-1.1	-2.3 (at birth)	4.0	Clinical features of Silver Russell Syndrome.	*De novo* variant
40	c.193C>T, p.(Gln65*)	F	-2.5	-3.4	-0.6	-1.2	2.0	IUGR, protruding forehead, feeding difficulties.	*De novo* variant

41*	Hom c.239C>T, p.(Pro80Leu)	F	-4.7	-6.7	-3.5	-7.4	3.0	Severe pre and postnatal growth failure, protruding forehead, feeding difficulties	Inherited one mutant copy from each parent
42*	Hom c.239C>T, p.(Pro80Leu)	M	-5.2	-6.2	-5.3	-6.1	1.1	Severe pre and postnatal growth failure, protruding forehead, feeding difficulties	Inherited one mutant copy from each parent

BW, Birth weight; SDS, Standard Deviation Score; BMI, Body Mass Index; HC, Head circumference; NK, not known. *Patients 41 and 42 are siblings with homozygous *HMGA2* mutations. c.111+1G>T and Homozygous c.239C>T variants are reported in Meyer *et al.* (295). c.189delA, p.(Ala64Leufs*102) and c.193C>T, p.(Gln65*) variants are reported in Abi Habib *et al.* (188).

Table 5.5 Patient scores for Netchine-Harbison SRS Clinical Scoring System

Patient	Heterozygous *HMGA2* variant	Birth weight (SDS)	Gestation	Head circumference at birth (SDS)	Protruding forehead?	Body asymmetry?	Feeding difficulties?	NH-CSS
Novel *HMGA2* variants								
15	c.145delA, p.(Arg49Glyfs*117)	2100g (-3.5)	40/40	'Normal, not relative to BW'	No	No	No	3/6
35	c.166A>G, p.(Lys56Glu)	2000g (-3.8)	40/40	NK	NK	No	Yes	3/6
36	c.52C>T, p.(Gln18*)	(-2.7)	NK	-2.0	Yes	Yes	No	4/6
37	c.144delC p.(Arg49Glyfs*117)	(-1.9)	NK	-1.4	Yes	No	Yes	4/6
Previously reported *HMGA2* variants								
38	c.111+1G>T	2360g (-2.6)	39/40	31cm (-3.31)	Yes	No	Yes	4/6
39	c.189delA, p.(Ala64Leufs*102)	1720g (-3.1)	37/40	31cm (-2.3)	Yes	No	Yes	5/6
40	c.193C>T, p.(Gln65*)	2300g (-2.5)	39/40	32.5cm (-2.0)	Yes	No	Yes	5/6
41*	Hom c.239C>T, p.(Pro80Leu)	1400g (-4.7)	39/40	NK	Yes	NK	Yes	4/6
42*	Hom c.239C>T, p.(Pro80Leu)	1050g (-5.2)	38/40	NK	Yes	NK	Yes	4/6

NH-CSS, Netchine-Harbison SRS (Silver Russell Syndrome) Clinical Scoring System: diagnosis of SRS requires fulfilment of 4/6 (including both prominent forehead and relative macrocephaly, termed 'Clinical SRS') or 3/6 in addition to a genetic diagnosis associated with SRS. The criteria are: a. SGA (Birth weight and/or birth length ≤−2 SDS for gestational age) b. Postnatal growth failure (Height at 24 ± 1 months ≤−2 SDS or height ≤−2 SDS below mid-parental target height) c. Relative macrocephaly at birth (Head circumference at birth ≥1.5 SDS above birth weight and/or length SDS) d. Protruding forehead (Forehead projecting beyond the facial plane on a side view at 1–3 years) e. Body asymmetry (Leg length discrepancy (LLD) of ≥0.5 cm or arm

asymmetry or LLD <0.5 cm with at least two other asymmetrical body parts, one non-face) f. Feeding difficulties and/or low BMI (BMI ≤−2 SDS at 24 months or use of feeding tube or cyproheptadine as appetite stimulant) (130,131). *Patients 41 and 42 are siblings with homozygous *HMGA2* mutations. c.111+1G>T and Homozygous c.239C>T variants are reported in Meyer *et al.* (295). c.189delA, p.(Ala64Leufs*102) and c.193C>T, p.(Gln65*) variants are reported in Abi Habib *et al.* (188).

Table 5.6 Biochemistry of patients with novel *HMGA2* variants

Patient	Heterozygous *HMGA2* variant	Basal GH (ug/L)	GH peak (ug/L)	Basal IGF-1 (ng/ml)	IGF-1 SDS	Bone age
15	c.145delA, p.(Arg49Glyfs*117)	NK	NK	18 (NR 14-192)	-1.91	NK
35	c.166A>G, p.(Lys56Glu)	0.3 mg/l	Not tested	33 (NR 4.4-22.3)	+4.4	Bone age 5 years at chronological age of 6 years
36	c.52C>T, p.(Gln18*)	NK	NK	NK	NK	NK
37	c.144delC p.(Arg49Glyfs*117)	NK	NK	NK	NK	NK

NR, Normal range; SDS, Standard Deviation Score; NK, not known

Table 5.7 Genetic details of novel *HMGA2* variants

Pt	Heterozygous *HMGA2* variant	Pathogenicity scores	GnomAD frequency	Mutant amino acid sequence	Effect of variant on amino acid sequence
15	c.145delA, p.(Arg49Glyfs*117)	Mutation Taster: disease causing	Novel – not listed	MSARGEGAGQPSTSAQGQPAAPAPQKRGRGRPRKQQQEPTGEPSP KRPGEDPKAAKTRVPLKQLKRKQKPLEKNGQEADLGNGHNKLFRRS LLRRKLKRHPHKSLPKRTRGRQRSISTSAAVGSFEGRRHCSDHLFCIA MVFPLSSGVGWGGVGEGGVGWGEIT*	Changes 49th amino acid from Arginine to Glycine and causes frameshift. Mutant protein 164 amino acids, only first 48 same as wild-type sequence.
35	c.166A>G, p.(Lys56Glu)	CADD score: 27.2 Mutation Taster: disease causing	Novel – not listed	MSARGEGAGQPSTSAQGQPAAPAPQKRGRGRPRKQQQEPTGEPSP KRPRGRPKGSENKSPSKAAQKKAEATGEKRPRGRPRKWPQQVVQKK PAQEETEETSSQESAEED*	Missense changing 56th amino acid from Lysine to Glutamic acid
36	c.52C>T, p.(Gln18*)	CADD score: 35 Mutation Taster: disease causing	Novel – not listed	MSARGEGAGQ PSTSAQG*	Frameshift, causes immediate stop codon after 17th amino acid.
37	c.144delC p.(Arg49Glyfs*117)	Mutation Taster: disease causing PyloPhen: pathogenic	Novel – not listed	MSARGEGAGQPSTSAQGQPAAPAPQKRGRGRPRKQQQEPTGEPSP KRPGEDPKAAKTRVPLKQLKRKQKPLEKNGQEADLGNGHNKLFRRS LLRRKLKRHPHKSLPKRTRGRQRSISTSAAVGSFEGRRHCSDHLFCIA MVFPLSSGVGWGGVGEGGVGWGEIT*	Changes 49th amino acid from Arginine to Glycine and causes frameshift. Mutant protein 164 amino acids, only first 48 same as wild-type sequence.

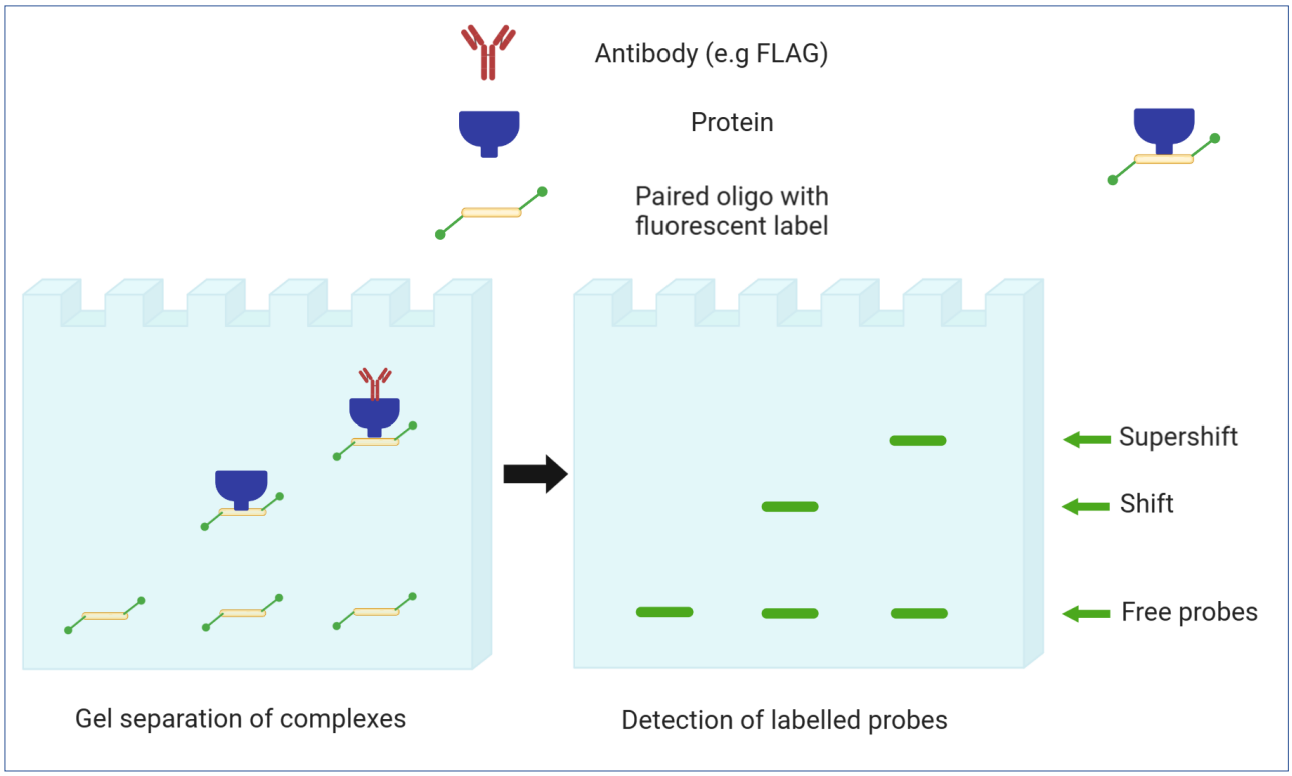
Pt, Patient; CADD score, Combined Annotation Dependent Depletion Score (https://cadd.gs.washington.edu/) (167). *HMGA2* transcript NM_003483.4 (wild-type sequence is 109 amino acids)

Interestingly, Patient 35 harboured the c.166A>G, p.(Lys56Glu), the first missense heterozygous variant identified in a patient with short stature and SRS features. This variant occurs in a critically important region of *HMGA2* next to the 2nd AT hook. The missense variant changes lysine, a positively charged amino acid to glutamic acid, a negatively charged amino acid. The positively charged amino acids surrounding the consensus AT hook sequence are key to effective binding at target sites (296). For this reason, I chose to functionally assess the consequence of this single base change to determine if it impaired DNA binding.

5.2.3 Electrophoretic mobility shift assay

I performed an Electrophoretic mobility shift assay (EMSA) comparing the function of wild-type HMGA2 protein binding to the mutant c.166A>G, p.(Lys56Glu) binding. This is the *HMGA2* variant identified in Patient 35. The principles of the EMSA are demonstrated in **Figure 5.8**.

Figure 5.8 Principles of the Electrophoretic mobility shift assay (EMSA)



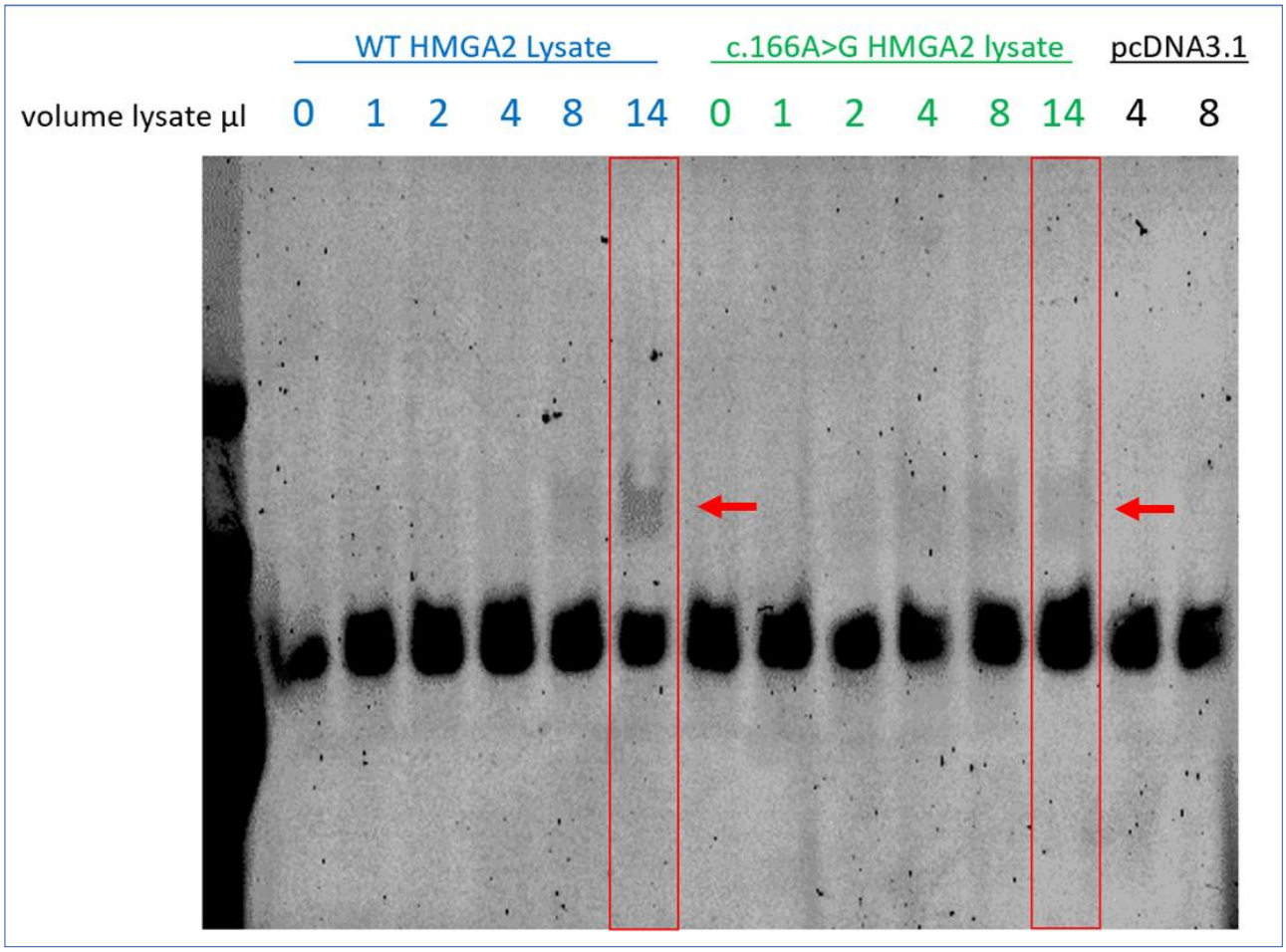
To ensure EMSA results were specific for HMGA2 proteins and not for other proteins present in abundance in the cell lysate, only proteins contained FLAG tags were isolated from whole cell lysates using immunoprecipitation using FLAG beads. An HMGA2 antibody was used to induce supershift. Image created using BioRender.

The EMSA using FLAG IP product is shown in **Figure 5.9** below. This EMSA appeared to show reduced shift for the mutant compared to the wild-type HMGA2 when equal volumes of cell lysate were loaded. However, no supershift was seen.

Despite quantifying the whole cell lysate with a BCA assay and adding the same volume of FLAG IP beads to the cell lysates, the quantification by western blot shown in **Figure 5.10** highlighted that differences in the amount of HMGA2 protein were still present. Interestingly, **Figure 5.10** demonstrated that the mutant protein was almost twice as abundant as the wild-type protein in the same volume of FLAG IP product. This supports the hypothesis of reduced binding, given that the

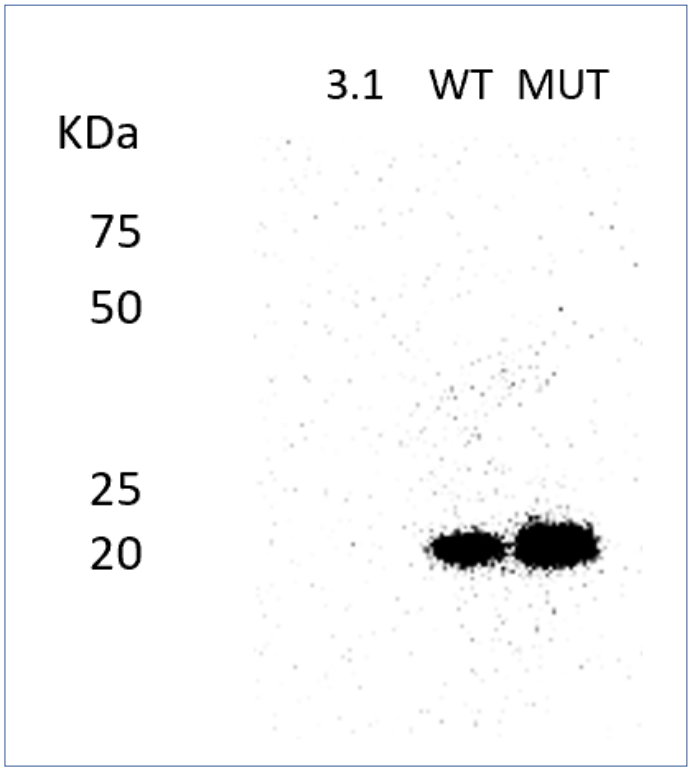
equal volumes loaded would contain more mutant protein than wild-type yet showed more shift for the wild-type protein.

Figure 5.9 Electrophoretic mobility shift assay (EMSA) using FLAG IP product



Red boxes highlight the lanes of interest, comparing the shift seen with 14µl of WT HMGA2 FLAG IP product compared to 14µl of mutant HMGA2 protein FLAG IP product. The observed shift is indicated by red arrows. No supershift is seen.

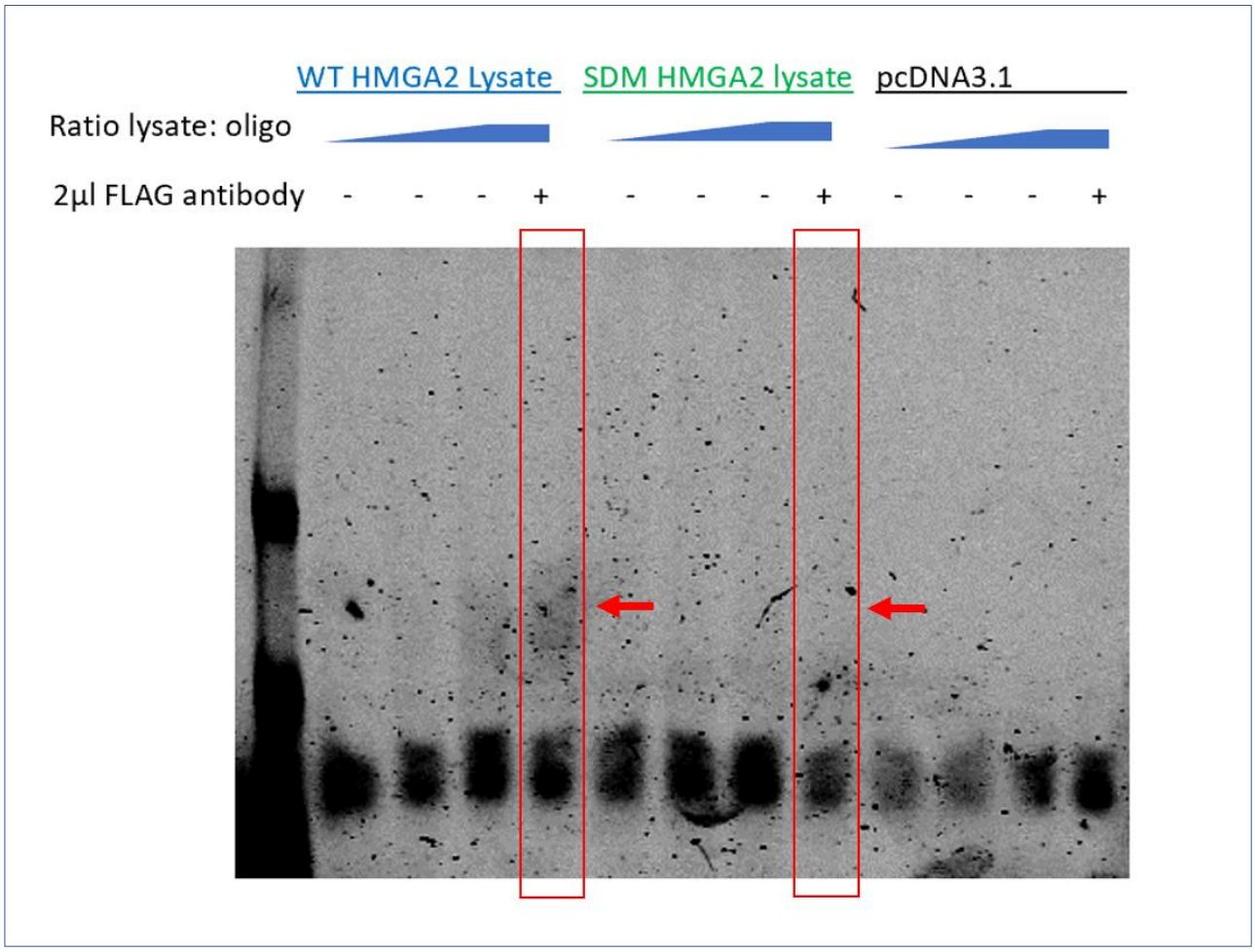
Figure 5.10 Western blot of FLAG IP product used for EMSA in Figure 5.9. HMGA2 antibody used to detect presence of HMGA2 protein



Kda, Kilodaltons; 3.1, pcDNA3.1 empty vector only; WT, wildtype; MUT, c.166A>G, p.(Lys56Glu) mutant. Expected size of FLAG tagged HMGA2 protein 19KDa.

To overcome the variability of the amount of HMGA2 protein loaded, I performed western blot analysis of subsequent FLAG IP lysates prior to EMSA and used the image to quantify my input for EMSA. This additional step improved the likelihood of loading equal HMGA2 protein quantities. The final EMSA using this method can be seen in **Figure 5.11**.

Figure 5.11 EMSA using FLAG IP input quantified by western blot for HMGA2



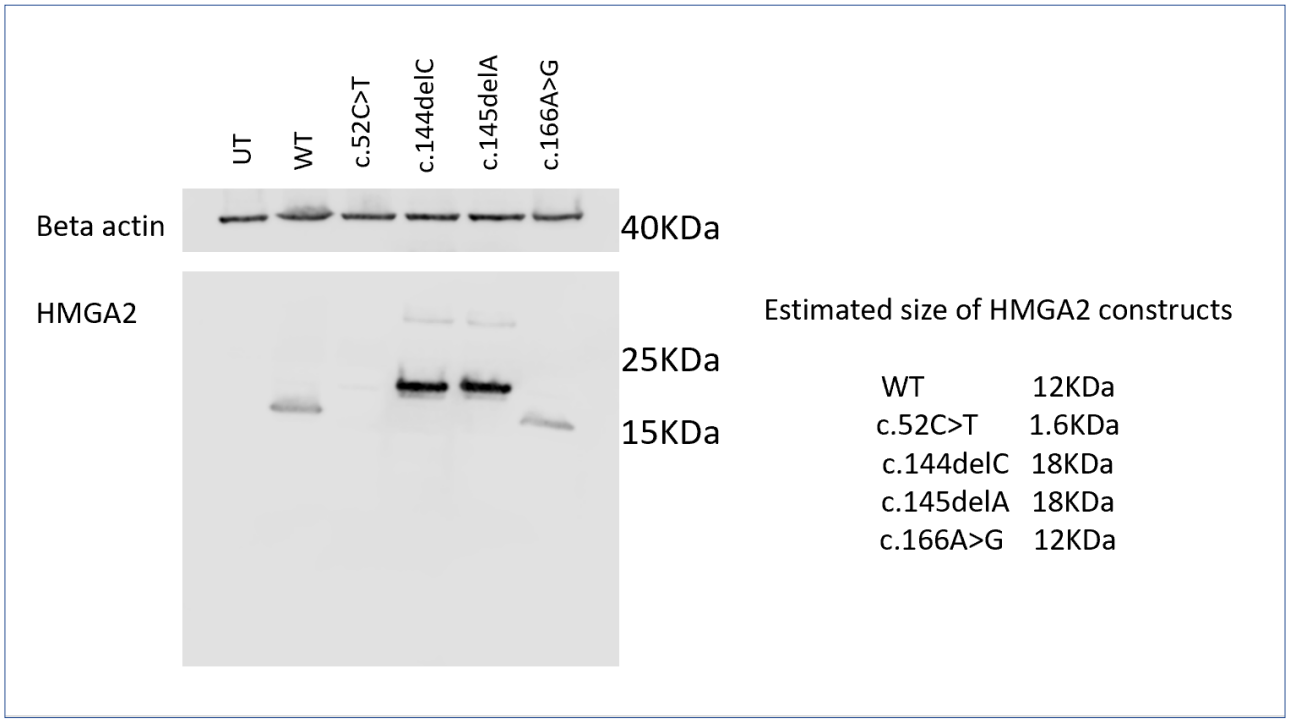
Red boxes highlight the lanes of interest, comparing the shift seen with WT HMGA2 FLAG IP product compared to mutant HMGA2 protein FLAG IP product. The observed shift is indicated by red arrows. Despite the artefact on the image, it is possible to see a reduced shift for the mutant compared to the wild-type. No supershift is seen.

5.2.4 Creation and detection of the frameshift *HMGA2* variants

To assess the expression and localisation of the mutant *HMGA2* variants, I recreated the remaining 3 patient variants in the same *HMGA2* vector using site-directed mutagenesis. Following transfection into HEK293T cells and whole cell lysis 24 hours later, I assessed expression of HMGA2 proteins by western blot (**Figure 5.12**). The c.144delC p.(Arg49Glyfs*117), c.145delA p.(Arg49Glyfs*117) and c.166A>G p.(Lys56Glu) mutant proteins were easily detectable by western blot using an antibody recognising the N terminal of HMGA2 protein. The two larger proteins created by the variants c.144delC p.(Arg49Glyfs*117) and c.145delA p.(Arg49Glyfs*117) appeared

to be expressed at higher levels than the wild-type HMGA2 protein. It was not possible to detect the c.52C>T p.(Gln18*) mutant protein. As it is very small, it is unlikely to be well detected by the HMGA2 antibody epitope (which is N terminal).

Figure 5.12 Western blot of HMGA2 wild-type and mutant proteins

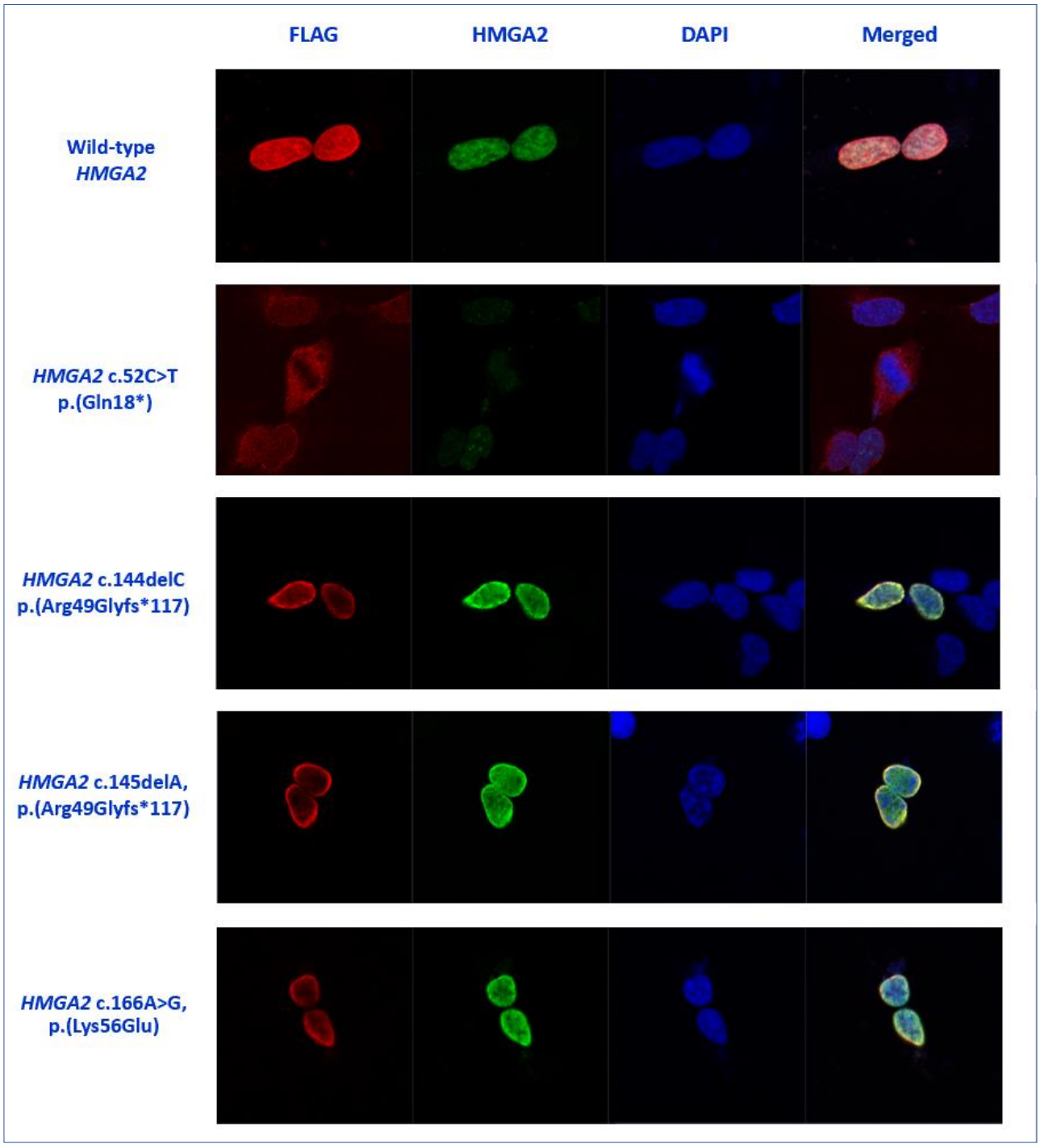


UT, untransfected HEK293T cells; WT, HEK293T cells transfected with the wild-type *HMGA2* vector; KDa, Kilodalton. Beta actin used as loading control.

5.2.5 Subcellular localisation of wild-type and mutant HMGA2 proteins

To assess whether these mutant HMGA2 proteins were able to translocate to the nucleus, I performed immunocytochemistry analysis of HEK293T cells transfected with wild-type and mutant constructs. Representative images are shown in **Figure 5.13**.

Figure 5.13 Immunocytochemistry for HMGA2 wild-type and mutants



Images acquired using the Zeiss LSM 880 confocal microscope of HEK293T cells transfected with wild-type and mutant *HMGA2* constructs. These cells were fixed onto coverslips and incubated with FLAG, HMGA2 and DAPI primary antibodies followed by fluorescent secondary antibodies. For further details please see methods **Section 2.17.** The FLAG tagged wild-type HMGA2 protein is able to translocate to the nucleus, as demonstrated by the correlation with DAPI nuclear staining. These images suggest that the c.52C>T p.(Gln18*) mutant HMGA2 protein is unable to translocate to the nucleus, likely due to its lack of nuclear localisation signals. The c.144delC p.(Arg49Glyfs*117), c.145delA p.(Arg49Glyfs*117) and c.166A>G p.(Lys56Glu) HMGA2 mutants appear able to translocate to the nucleus, but appear to be more densely situated around the nuclear membrane rather than distributed evenly throughout the nucleus.

The location of the FLAG and HMGA2 signal for the wild-type correlated very closely with the DAPI nuclear stain suggesting the FLAG tagged HMGA2 was able to successfully translocate to the nucleus. The c.52C>T p.(Gln18*) mutant protein is very small and would not be well detected by the HMGA2 antibody due to the antibody epitope. This epitope is located at the N terminal of the protein but spans beyond the 52 bases included in this early frameshift protein. However, the FLAG signal at the N terminal end should be relatively well detected by the FLAG antibody. The FLAG signal for this mutant is very interesting and suggests the small c.52C>T p.(Gln18*) mutant protein is unable to localise to the nucleus. This is likely due to the lack of nuclear localisation signals present in this protein. There is only background signal detected for the HMGA2 antibody for the c.52C>T p.(Gln18*) mutant protein. The c.144delC p.(Arg49Glyfs*117), c.145delA p.(Arg49Glyfs*117) and c.166A>G p.(Lys56Glu) HMGA2 mutant proteins appear to be able to translocate to the nucleus, but appear more dense around the nuclear membrane rather than distributed evenly throughout the nucleus. Further functional assessment is ongoing to characterise this observation in more detail.

5.2.6 Limitation of vector design for long mutant HMGA2 proteins

The *HMGA2* vector which contained the entire *HMGA2* coding sequence was used for the above experiments. It was designed for the EMSA experiment and successfully produced WT and missense *HMGA2* variant proteins when expressed in HEK293T cells (**Figure 5.10**). The same vector was used to create the two *HMGA2* variants, c.144delC p.(Arg49Glyfs*117) and c.145delA p.(Arg49Glyfs*117). These are the mutations seen in Patient 15 and Patient 37 respectively, both of which cause a frameshift and the translation occurs beyond the normal stop codon, meaning that the 3'UTR is also translated into the mutant protein. These differences are shown below in **Figure 5.14**. For both variants, there is a significant difference to the mutant sequence when the 3'UTR section is included. This could affect how well the protein is able to translocate to the nucleus or any other function of the mutant protein. For this reason, functional analysis will be repeated for

these variants by performing site-directed mutagenesis to a new *HMGA2* vector which includes a section of the 3'UTR region.

Figure 5.14 Amino acid sequences for *HMGA2* variants c.144delC and c.145delA: translation differences to *HMGA2* vector

Both variants changes 49th amino acid Arginine (AGG) to Glycine (GGG)

Sequence generated from HMGA2 vector
MSARGEGAGQPSTSAQGQPAAPAPQKRGRGRPRKQQQEPTGEPSPKRPGEDPKAAKTRVPLKQLKRKQKPLEKNGQEADLGN
GHNKLFRRSLLRRKLKRHPHKSLPKRTRDPDKNSADIQHSGGRSSLEGPFKPADQPRLCLLVASHLLFAPPPCLP*
157 amino acids

Sequence generated from HMGA2 when 3'UTR region included (true sequence of mutant)
MSARGEGAGQPSTSAQGQPAAPAPQKRGRGRPRKQQQEPTGEPSPKRPGEDPKAAKTRVPLKQLKRKQKPLEKNGQEADLGN
GHNKLFRRSLLRRKLKRHPHKSLPKRTRGRQRSISTSAAVGSFEGRRHCSDHLFCIAMVFPLSSGVGWGGVGEGGVGWGEIT*
164 amino acids

Blue region = similarities within mutant sequence
Red region = differences within mutant sequence

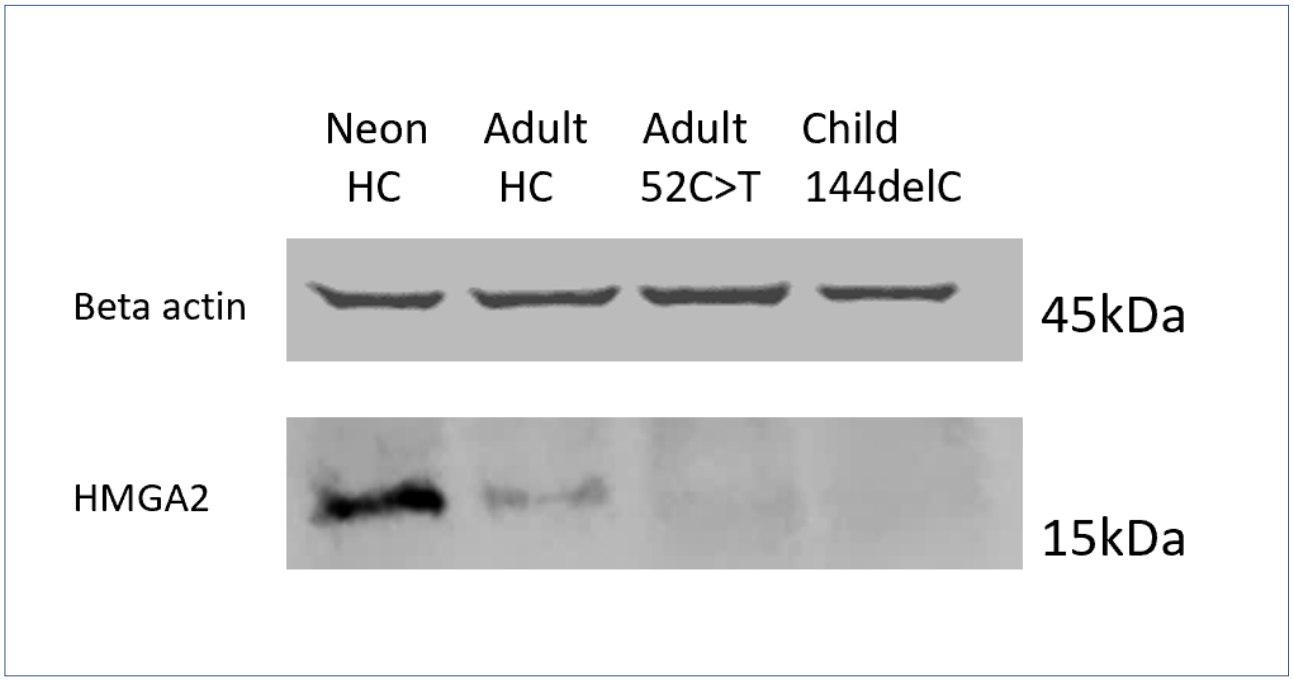
The similarities (highlighted in blue) and differences (highlighted in red) between the sequence generated using my *HMGA2* vector and the true sequence of the variants when the 3'UTR region is included.

5.2.7 HMGA2 protein expression in patient fibroblasts

I obtained dermal fibroblast samples from Patient 37 (c.144delC mutation) and the mother of Patient 36 (c.52C>T mutation), from whom the child inherited the variant. This mother has the heterozygous c.52C>T mutation, has short stature (SDS -3.7) and a triangular shaped face. After culturing the fibroblasts to a similar passage, I assessed levels of HMGA2 protein expression by western blot (shown in **Figure 5.15**). HMGA2 expression was strongest in the neonatal healthy control, which is in keeping with the recognised high expression in utero and in early life. Interestingly, some HMGA2 protein was detectable in the adult healthy control sample. However, HMGA2 was not detectable in either the child or the adult patient with *HMGA2* mutations. It was surprising that so little HMGA2 was detectable in the patient samples, especially as they had

heterozygous mutations so you would expect some wild-type HMGA2 to be present. These are not thought to be dominant negative mutations and are thought to cause the phenotype due to haploinsufficiency for *HMGA2*, in keeping with patients with postnatal growth failure and 12q14 deletion syndrome, who are missing one copy of the *HMGA2* gene. It is possible that this could be an imprinted region, although this is not currently recognised, or that other genes/proteins are playing important regulatory roles in their expression. Further work is required to fully characterise this.

Figure 5.15 Western blot analysis of HMGA2 expression in patient fibroblasts



HC, Healthy control; KDa, Kilodaltons. Beta actin used as a loading control.

5.2.8 Creation of transgenic mouse model

The preliminary findings of the EMSA assessing the c.166A>G p.(Lys56Glu) HMGA2 mutant protein suggests that this amino acid plays a critical role in *HMGA2* function. Our collaborators at The University of Mannitoba, Winnepeg generated novel transgenic mice with heterozygous and homozygous c.166A>G p.(Lys56Glu) mutations (*Hmga2*^{K56E}) to further assess this region. Mice with

homozygous mutations displayed dysmorphic facial features similar to the phenotypes observed in SRS children. 3D reconstructed serial micro-CT images showed an overall shortening of the viscerocranium and asymmetry of the skull bones. *Hmga2*^{K56E} transgenic mice with homozygous mutations were fertile but small for gestational age and showed SRS-like dwarfism compared to heterozygous age- and sex-matched littermates. This team have isolated the mouse embryonic fibroblasts, which stably express the mutant Hmga2 protein. This will enable further functional characterisation of this interesting variant and help develop our understanding of the role of *HMGA2* in growth.

5.2.9 Conclusions and discussion

I identified several novel *HMGA2* mutations in patients with short stature and SRS features. The novel c.166A>G mutation is the first heterozygous missense *HMGA2* mutation identified in a patient with growth failure and SRS features. The mutant c.166A>G HMGA2 protein could translocate to the nucleus but my EMSA findings suggest that binding to DNA target sites is impaired. No supershift was observed in my EMSA, which is usually required to prove the binding seen is specific for the protein of interest. The lack of supershift may be because the binding site of the HMGA2 antibody was similar to the binding site of the protein to the labelled probes. It would be pertinent to repeat this experiment using an HMGA2 antibody with a different binding epitope to see if it possible to induce supershift and therefore confirm that the reducing binding seen in my EMSA is specific for HMGA2. However, as IP FLAG products were used as the input in this experiment and no protein was precipitated from the empty vector whole cell lysates, I am reasonably confident that these are FLAG tagged HMGA2 proteins rather than non-specific proteins. Endogenous proteins in the cell should not be present in the FLAG IP product.

It is an exciting finding that the *Hmga2*^{K56E} transgenic mouse recapitulates the patient phenotype and the team have isolated the mouse embryonic fibroblasts, which stably express the mutant Hmga2 protein. This will enable further functional characterisation of this interesting variant and help develop our understanding of the role of *HMGA2* in growth.

Chapter 6. Overall conclusions and future direction

The custom gene panel included whole genomic sequences for each gene of interest, including the intronic regions, 2kb upstream and 500bp downstream of the gene. This enabled detailed exploration and assessment of non-coding regions in known genes as well as looking for mutations in selected potential candidate growth genes. This generated a wealth of data for the non-coding regions for the genes of interest, given that the classically examined exonic regions of the gene make up only 1-2% of the gene itself. One challenge was interpreting this data, given that a relatively small amount is known about the vast non-coding landscape between exons in a gene and the 3 and 5'UTR regions. After filtering for very rare variants, I relied on *in silico* software to help predict the changes that were most likely to be of interest and then pursued these functionally in my lab experiments. At present, making these novel diagnoses is far more labour intense than identifying a classic exonic mutation. Whilst the exonic mutation would require confirmatory Sanger sequencing, the novel diagnoses would require splicing assays, fibroblast analysis, cell models and so on to fully characterise the effect of the novel change. However, as our knowledge and understanding of these novel mutations develops this may become less labour intense.

I selected the strongest candidates for functional analysis, but it is likely there are still several more diagnoses to be made from this genetic data. I have not assessed any variants that disrupt splicing silencer or enhance splice enhancer elements, affect non-coding regulatory regions or non-coding RNA. As our recognition of disease-causing non-coding variants improves, so will the *in silico* prediction models and tools available. Each new genetic discovery helps our understanding of the mechanisms of disease and the vast genetic landscape that we are still exploring at a rapid pace. Combining genetic sequencing data with studies assessing the mRNA produced in affected patient tissue is key in determining the pathogenicity of deep intronic mutations. This can be done by

conventional RT-PCR analysis and sequencing of cDNA products or by direct RNA-seq analysis (297,298).

When compared to candidate gene sequencing, the custom gene panel enables rapid testing of the 64 genes of interest. This is particularly helpful in patients without strong phenotypic features, such as Patients 7 and 8 with the dominant negative *GHR* variants. As they do not have classic GHI phenotypes, it is possible the *GHR* may not have been selected as a candidate gene to sequence and thus the diagnosis could have been missed. Similarly, heterozygous *SHOX* gene defects are being increasingly recognised in patients who had previously been labelled as having idiopathic short stature (299). As we appreciate an increasing overlap between short stature phenotypes, it is helpful to cast the net wider in our search for genetic diagnoses.

When the human genome was sequenced, far fewer genes were identified than expected. This may be in part explained by the high frequency of alternative splicing, seen for example in both the *GHR* and *IGF1* genes, but it also suggests a complex interplay of other key genetic regulatory mechanisms that may have been underestimated (300). For many years, genomic interrogation has focused heavily on the protein-coding regions of the gene, i.e. the exons, and the non-coding regions have been simply labelled as 'junk DNA.' However, there are several reasons to consider whether non-coding regions have been unfairly overlooked (298).

Firstly, many intronic sequences are remarkably conserved across species, suggesting at least some intronic features are subject to evolutionary constraints (301,302). Well conserved regions include the consensus splice site sequences, the sequences of non-coding RNA genes and the binding sites for regulatory proteins (303). Intronic regions have also been shown to increase transcriptional efficiency. Intron-containing genes have been shown to present higher levels of transcription when

compared to genes lacking introns in yeast, drosophila and mammalian cells (304–307). For transcription to occur in mammalian genes, a complex communication must occur between promoters and enhancers that are often located vast distances from each other in the genome. Recent studies suggest that regulatory sequences located within the introns may work together with promoters to aid transcriptional processes (308). For example, expression of the type II collagen α1 (*Col2a1*) gene is dependent on SOX9, a master transcription factor required for chondrogenesis that binds to key regulatory regions located in *Col2a1* introns 1 and 6 (309).

Furthermore, recent advancement in transcriptome sequencing and analysis has revealed a surprising prevalence of non-canonical splicing mechanisms, ranging from changes in the usual order of splicing to the recognition of atypical splice sites (310). Base sequences resembling the consensus splice sequences can be found in abundance within the introns, and these are referred to as cryptic, non-canonical or pseudo splice sites. The complex process of recognition of a splice site boundary by the spliceosome is not fully understood, especially as more than 9000 sequence variants have been identified in the -3 to +6 region of human 5' splice sites (311). This makes it unlikely that the sequence alone is responsible for splice site boundary recognition. It is likely influenced by other factors such as SR and hnRNP proteins that bind the pre-mRNA and either inhibit recruitment of spliceosomal components or stabilise spliceosome interactions (312,313).

Pseudoexon inclusions are likely an under recognised cause of human disease (314,315). This mis-splicing event is due to intronic mutations, most commonly point mutations but sometimes small intronic deletions, that activate non-canonical splice sites within the intron (316). The commonest mechanism for pseudoexon inclusion involves a mutation that creates a new donor splice site and activates a dormant acceptor splice site (298). This is the mechanism seen in the creation of both *GHR* pseudoexons, which utilise the same dormant acceptor splice site deep within intron 6.

Pseudoexon inclusions can also result from mutations that create a novel acceptor splice site or that disrupt or create splicing silencer or enhancer elements, respectively (298). My findings highlight the importance of studying variation in non-coding regions as a cause of monogenic disorders.

Genomic discovery is advancing at such a pace that new candidate growth genes are frequently being proposed. One limitation of the custom short stature gene panel is that it covers a pre-agreed list of genes. It is possible to amend the gene panel list as new discoveries are made by adding more probes, but for patients that have already been assessed on the original panel this would require more DNA, more expense, etc. Whilst the 'whole gene' panel should enable exploration of UTRs and non-coding RNAs nearby or within the genes of interest, analysis of potential distal genetic regulatory mechanisms is not possible due to the focused nature of the panel. For example, it wouldn't be possible to identify genetic alterations to transcriptional enhancers that control the spatiotemporal regulation of genes. These may play an under-recognised role in Mendelian disease and are reported in several different diseases (317). Gain-of-function point mutations in distal enhancers have been shown to affect protein expression and lead to conditions such as preaxial polydactly (318). Loss-of-function mutations in a single enhancer close to PTF1A have been shown to impair the entire PTF1A enhancer cluster and lead to pancreas agenesis and neonatal diabetes (317).

To overcome these limitations of the gene panel, one option would be to assess the patients using whole genome sequencing and apply a virtual gene panel list in the filtering pipeline of this data. This would mean as new candidate growth genes were discovered, they could be easily added to the virtual gene panel list without the need for repeat testing or expense. Combining whole genome sequencing data with enhancer maps may enable recognition of changes to distal enhancer regions.

However, it is unlikely that whole genome sequencing would cover the regions of interest with such a depth of coverage that is possible with the gene panel. Recent innovations combining gene-engineering technology with computer aided technology such as Oxford Nanopore Technologies (ONT) can provide rapid analysis and offer the possibility of real-time sequencing at a relatively low cost. Long read lengths enable simpler assembly and more comprehensive analysis of repetitive regions. ONT determines base sequences by monitoring voltage changes as the nucleotides pass through biological nanopores. Other similar innovations include NabSys, which uses a solid-state nanopore for sequencing, and Sequenom, which involves simultaneous optical probing by multiple nanopores (319). However, a current limitation of this technology is the read accuracy, with error rates of 5-15% compared to <1% for third generation sequencing technology (320,321). With ongoing optimisation of nanopore structure, machine algorithms of basecalling and library preparation methods, nanopore sequencing technology offers broad application prospects in epigenetics research, gene modifications, cancer genetics and may eventually replace earlier sequencing methods (322).

Identifying a genetic alteration is an important first step, but key to contextualising this change is determining its effect on the complex interactions within the human body. Initiatives such as the Genotype-Tissue Espression (GTEx) project help to characterise how different genetic variants affect gene regulation. GTEx studies the effects of genetic changes on the transcriptome across human tissues, and links these regulatory mechanisms to Mendelian trait and disease associations (323).

Another challenge is interpreting variants where pathogenicity is uncertain, such as the *CUL7* variants identified in Patients 11 and 12 and in the class 3 CNVs identified in Patients 25, 28 and 29. This problem may become increasingly frequent as we explore further into the genome, but reporting of variants and accumulation of pathogenicity evidence and patients with the same

variants and similar phenotypes will help develop the tools to assess these novel regions. In cases where pathogenicity is uncertain, assessing segregation of the phenotype and variant within the wider family pedigree is helpful but still may not provide a conclusive answer. Referral to a clinical geneticist for their expert input may also be beneficial.

The mutant *GHR* missing the first 26 bases of exon 9 seen in Patients 7 and 8 may play an important regulatory role in normal physiology and is found in a significant percentage of *GHR* transcripts in healthy people. 1-10% of all *GHR* transcripts were homologous to the mutant *GHR* transcript missing the first 26bp of exon 9 (82). This mutant *GHR* was proposed to modulate the function of the full length receptor, inhibit signalling and generate large amounts of GHBP (82). The regulation of GHR expression and GHBP generation is not well understood. Leading on from my genomic discoveries, our research group has been exploring the function and interaction of this mutant GHR protein to help provide insights into these crucially important interactions. They are using NanoBiT complementation and flow cytometry to look at the cell surface expression of variant GHR homo/heterodimers compared to wild-type (WT) homodimers. They are also investigating rhGH binding to variant GHR homo/heterodimers and GH binding protein (GHBP) cleaved from the variant GHRs to determine if increased GH sequestration occurs.

Assessing patients for copy number variants could be considered a less biased approach to the targeted gene panel as this is assessing for duplications or deletions of genetic material across the entire genome. Although the pathogenicity algorithm does rely on whether the CNV is associated with disease or phenotypic features in other reported patients. My CNV analysis shown in results **Chapter 4** was the first study investigating CNVs in subjects with growth failure associated with GHI and unexplained short stature. In our cohort, rare CNVs are a relatively common cause of milder non-classical phenotypes. The discovery of CNVs in this cohort expands the known phenotypes of

rare CNVs, potentially identifies new candidate growth genes and expands the spectrum of GHI/IGF-1 insensitivity and the overlapping short stature disorders such as SRS. CNV analysis should be considered in all short patients where no monogenic cause has been identified, particularly those with features of GHI and concomitant subtle features of SRS.

As increasingly sophisticated techniques allow the collection of vast amounts of genetic data, large collaborative studies enable thorough and powered analysis to identify key candidate genes or regions responsible for a phenotype. A recent meta-analysis using CNV data from almost one million individuals enabled construction of a dosage sensitivity map of the human genome for 54 Mendelian disorders. They used an ensemble machine learning model to predict dosage sensitivity probabilities for all autosomal genes. This identified 2,987 haploinsufficient and 1,559 triplosensitive genes and has vast potential applications for future research in Mendelian disease (324).

My initial results from the missense *HMGA2* variant are promising but significant further work is required to fully characterise the role of *HMGA2* in growth. It would be pertinent to repeat the EMSA experiment using an HMGA2 antibody with a different binding epitope to see if it possible to induce supershift and therefore confirm that the reducing binding seen in my EMSA is specific for HMGA2. Immunocytochemistry could be performed on patient fibroblasts using the HMGA2 antibody to compare levels of expression and site of expression between the adult and child patient samples and the adult and neonatal healthy control. In the patient with c.52C>T mutation, detection of mutant HMGA2 protein would be limited due to the antibody epitope and the small size of the protein. However, the results would still provide valuable information about wild-type expression of the protein, given that these patients have heterozygous mutations.

Site-directed mutagenesis to accurately recreate the c.145delA and c.144delC mutants could be performed on a custom *HMGA2* vector which includes a section of the 3' UTR immediately after the *HMGA2* coding sequence. This would enable translation to occur into the 3'UTR and accurately represent the mutant proteins. Functional analysis by western blot and immunocytochemistry could then ascertain if these mutant proteins were detectable in the cell and if their localisation differed to the wild-type HMGA2 protein.

Preliminary data confirm that the homozygous transgenic *Hmga2*^{K56E} mouse model recapitulated the Silver Russell Like phenotype seen in our patient. This is very promising, as it confirms the critical functional importance of this *HMGA2* amino acid residue and validates my preliminary data showing that this region plays a critical role in the function of *HMGA2* and the regulation of growth. Our collaborators have cultured mouse embryonic fibroblasts from these transgenic mice and have shown that the mutant HMGA2 protein is stable and has robust expression. This provides a wealth of possibilities for functional characterisation in the future.

References

1. Ranke MB, Wit JM. Growth hormone-past, present and future. *Nature Reviews Endocrinology*. 2018;14(5): 285–300. https://doi.org/10.1038/nrendo.2018.22.
2. Baumann G. Growth hormone binding protein. The soluble growth hormone receptor. *Minerva endocrinologica*. 2002;27(4): 265–276.
3. Daughaday WH, Hall K, Raben MS, Salmon WDJ, van den Brande JL, van Wyk JJ. Somatomedin: proposed designation for sulphation factor. *Nature*. 1972;235(5333): 107. https://doi.org/10.1038/235107a0.
4. Salmon WD, Daughday WH. A hormonally controlled serum factor which stimulates sulfate incorporation by cartilage in vitro. *The Journal of laboratory and clinical medicine*. 1957;49(6): 825–836.
5. D'Ercole AJ, Applewhite GT, Underwood LE. Evidence that somatomedin is synthesized by multiple tissues in the fetus. *Developmental Biology*. 1980;75(2): 315–328. https://doi.org/https://doi.org/10.1016/0012-1606(80)90166-9.
6. Duff MO, Olson S, Wei X, Garrett SC, Osman A, Bolisetty M, et al. Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature*. 2015;521(7552): 376–379. https://doi.org/10.1038/nature14475.
7. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular and Cellular Proteomics*. 2014;13(2): 397–406. https://doi.org/10.1074/mcp.M113.035600.
8. Nilsson A, Isgaard J, Lindahl A, Dahlstrom A, Skottner A, Isaksson OG. Regulation by growth hormone of number of chondrocytes containing IGF-I in rat growth plate. *Science*. 1986;233(4763): 571 LP – 574. https://doi.org/10.1126/science.3523759.
9. Isaksson OG, Jansson JO, Gause IA. Growth hormone stimulates longitudinal bone growth directly. *Science*. 1982;216(4551): 1237 LP – 1239. https://doi.org/10.1126/science.7079756.
10. Lowe WL, Lasky SR, Leroith D, Roberts CT. Distribution and Regulation of Rat Insulin-Like Growth Factor I Messenger Ribonucleic Acids Encoding Alternative Carboxyterminal E-Peptides: Evidence for Differential Processing and Regulation in Liver. *Molecular Endocrinology*. 1988;2(6): 528–535. https://doi.org/10.1210/mend-2-6-528.
11. Lowe WL, Roberts CT, Lasky SR, LeRoith D. Differential expression of alternative 5' untranslated regions in mRNAs encoding rat insulin-like growth factor I. *Proceedings of the National Academy of Sciences of the United States of America*. 1987;84(24): 8946–8950. https://doi.org/10.1073/pnas.84.24.8946.
12. Isaksson OGP, Lindahl A, Nilsson A, Isgaard J. Mechanism of the Stimulatory Effect of Growth Hormone on Longitudinal Bone Growth*. *Endocrine Reviews*. 1987;8(4): 426–438. https://doi.org/10.1210/edrv-8-4-426.
13. Lindahl A, Isgaard J, Isaksson OGP. Growth Hormone in Vivo Potentiates the Stimulatory Effect of Insulin-Like Growth Factor-1 in Vitro on Colony Formation of Epiphyseal Chondrocytes Isolated from Hypophysectomized Rats*. *Endocrinology*. 1987;121(3): 1070–1075. https://doi.org/10.1210/endo-121-3-1070.
14. Yakar S, Liu JLI, Stannard B, Butler A, Accili D, Sauer B, et al. Normal growth and development in the absence of hepatic insulin-like growth factor I. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(13): 7324–7329. https://doi.org/10.1073/pnas.96.13.7324.
15. Wu Y, Sun H, Yakar S, LeRoith D. Elevated levels of insulin-like growth factor (IGF)-I in serum rescue the severe growth retardation of IGF-I null mice. *Endocrinology*. 2009;150(9): 4395–

4403. https://doi.org/10.1210/en.2009-0272.
16. Saenger P, Czernichow P, Hughes I, Reiter EO. Small for gestational age: Short stature and beyond. *Endocrine Reviews*. 2007;28(2): 219–251. https://doi.org/10.1210/er.2006-0039.
17. Kaplan SA, Cohen P. Review: The somatomedin hypothesis 2007: 50 Years later. *Journal of Clinical Endocrinology and Metabolism*. 2007;92(12): 4529–4535. https://doi.org/10.1210/jc.2007-0526.
18. Le Roith D, Bondy C, Yakar S, Liu JL, Butler A. The somatomedin hypothesis: 2001. *Endocrine Reviews*. 2001;22(1): 53–74. https://doi.org/10.1210/edrv.22.1.0419.
19. Boguszewski CL, Barbosa EJL, Svensson P, Johannsson G, Glad CAM. Clinical and Pharmacogenetic Aspects of the Growth Hormone. *European Journal of Endocrinology*. 2017;(September): 1–42.
20. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser*. 1999;(41): 95–98.
21. Storr HL, Chatterjee S, Metherell LA, Foley C, Rosenfeld RG, Backeljauw PF, et al. *Nonclassical GH insensitivity: Characterization of mild abnormalities of GH action*. Endocrine Reviews. 2019. https://doi.org/10.1210/er.2018-00146.
22. Brooks AJ, Dai W, O'Mara ML, Abankwa D, Chhabra Y, Pelekanos RA, et al. Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science*. 2014;344(6185). https://doi.org/10.1126/science.1249783.
23. Fisker S. Physiology and pathophysiology of growth hormone-binding protein: Methodological and clinical aspects. *Growth Hormone and IGF Research*. 2006;16(1): 1–28. https://doi.org/10.1016/j.ghir.2005.11.001.
24. Herington AC, Ymer S, Stevenson J. Identification and characterization of specific binding proteins for growth hormone in normal human sera. *Journal of Clinical Investigation*. 1986;77(6): 1817–1823. https://doi.org/10.1172/JCI112507.
25. Baumann G, Stolar MW, Amburn K, Barsano CP, DeVries BC. A specific growth hormone-binding protein in human plasma: initial characterization. *The Journal of clinical endocrinology and metabolism*. 1986;62(1): 134–141. https://doi.org/10.1210/jcem-62-1-134.
26. Guan R, Zhang Y, Jiang J, Baumann CA, Black RA, Baumann G, et al. Phorbol ester- and growth factor-induced growth hormone (GH) receptor proteolysis and GH-binding protein shedding: Relationship to GH receptor down-regulation. *Endocrinology*. 2001;142(3): 1137–1147. https://doi.org/10.1210/endo.142.3.8030.
27. Andrews A, Cottrell E, Maharaj A, Ladha T, Williams J, Schilbach K, et al. Characterization of dominant-negative growth hormone receptor variants reveals a potential therapeutic target for short stature. *European journal of endocrinology*. 2023;188(4): 353–365. https://doi.org/10.1093/ejendo/lvad039.
28. Zhang Y, Guan R, Jiang J, Kopchick JJ, Black RA, Baumann G, et al. Growth Hormone (GH)-induced Dimerization Inhibits Phorbol Ester-stimulated GH Receptor Proteolysis. *Journal of Biological Chemistry*. 2001;276(27): 24565–24573. https://doi.org/10.1074/jbc.M101281200.
29. Schantl JA, Roza M, Van Kerkhof P, Strous GJ. The growth hormone receptor interacts with its sheddase, the tumour necrosis factor-α-converting enzyme (TACE). *Biochemical Journal*. 2004;377(2): 379–384. https://doi.org/10.1042/BJ20031321.
30. Mercado M, Baumann G. A growth hormone/prolactin-binding protein in human milk. *The Journal of Clinical Endocrinology & Metabolism*. 1994;79(6): 1637–1641. https://doi.org/10.1210/jcem.79.6.7989468.
31. Waters MJ, Brooks AJ. JAK2 activation by growth hormone and other cytokines. *Biochemical Journal*. 2015;466(February): 1–11. https://doi.org/10.1042/BJ20141293.

32. Dehkhoda F, Lee CMM, Medina J, Brooks AJ. The growth hormone receptor: Mechanism of receptor activation, cell signaling, and physiological aspects. *Frontiers in Endocrinology*. 2018;9(FEB): 1–23. https://doi.org/10.3389/fendo.2018.00035.
33. Liu JP, Baker J, Perkins AS, Robertson EJ, Efstratiadis A. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). *Cell*. 1993;75(1): 59–72.
34. Laviola L, Natalicchio A, Giorgino F. The IGF-I signaling pathway. *Current pharmaceutical design*. 2007;13(7): 663—669. https://doi.org/10.2174/138161207780249146.
35. Barton ER, Park S, James JK, Makarewich CA, Philippou A, Eletto D, et al. Deletion of muscle GRP94 impairs both muscle and body growth by inhibiting local IGF production. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2012;26(9): 3691—3702. https://doi.org/10.1096/fj.11-203026.
36. Wallis M. New insulin-like growth factor (IGF)-precursor sequences from mammalian genomes: the molecular evolution of IGFs and associated peptides in primates. *Growth Hormone & IGF Research*. 2009;19(1): 12–23. https://doi.org/https://doi.org/10.1016/j.ghir.2008.05.001.
37. Adamo ML, Lanau F, Neuenschwander S, Werner H, LeRoith D, Roberts C T J. Distinct promoters in the rat insulin-like growth factor-I (IGF-I) gene are active in CHO cells. *Endocrinology*. 1993;132(2): 935–937. https://doi.org/10.1210/endo.132.2.8425505.
38. Yang H, Adamo ML, Koval AP, McGuinness MC, Ben-Hur H, Yang Y, et al. Alternative leader sequences in insulin-like growth factor I mRNAs modulate translational efficiency and encode multiple signal peptides. *Molecular Endocrinology*. 1995;9(10): 1380–1395. https://doi.org/10.1210/mend.9.10.8544846.
39. Wang X, Yang Y, Adamo ML. Characterization of the Rat Insulin-Like Growth Factor I Gene Promoters and Identification of a Minimal Exon 2 Promoter*. *Endocrinology*. 1997;138(4): 1528–1536. https://doi.org/10.1210/endo.138.4.5061.
40. O'Sullivan DC, Szestak TA, Pell JM. Regulation of IGF-I mRNA by GH: putative functions for class 1 and 2 message. *American journal of physiology. Endocrinology and metabolism*. 2002;283(2): E251—8. https://doi.org/10.1152/ajpendo.00016.2002.
41. Woelfle J, Chia DJ, Rotwein P. Mechanisms of Growth Hormone (GH) Action: IDENTIFICATION OF CONSERVED Stat5 BINDING SITES THAT MEDIATE GH-INDUCED INSULIN-LIKE GROWTH FACTOR-I GENE ACTIVATION *. *Journal of Biological Chemistry*. 2003;278(51): 51261–51266. https://doi.org/10.1074/jbc.M309486200.
42. Chia DJ, Young JJ, Mertens AR, Rotwein P. Distinct alterations in chromatin organization of the two IGF-I promoters precede growth hormone-induced activation of IGF-I gene transcription. *Molecular endocrinology*. 2010;24 4: 779–789.
43. Lingappa VR, Lingappa JR, Blobel G. Signal sequences for early events in protein secretion and membrane assembly. *Annals of the New York Academy of Sciences*. 1980;343: 356–361. https://doi.org/10.1111/j.1749-6632.1980.tb47264.x.
44. Tan NS, Ho B, Ding JL. Engineering a novel secretion signal for cross-host recombinant protein expression. *Protein engineering*. 2002;15(4): 337–345. https://doi.org/10.1093/protein/15.4.337.
45. Horimoto T, Nakayama K, Smeekens SP, Kawaoka Y. Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. *Journal of virology*. 1994;68(9): 6074–6078. https://doi.org/10.1128/JVI.68.9.6074-6078.1994.
46. Thomas G. Furin at the cutting edge: From protein traffic to embryogenesis and disease. *Nature Reviews Molecular Cell Biology*. 2002;3(10): 753–766. https://doi.org/10.1038/nrm934.
47. Philippou A, Maridaki M, Pneumaticos S, Koutsilieris M. The Complexity of the IGF1 Gene

Splicing , Posttranslational Modification and Bioactivity. 2014; 202–214. https://doi.org/10.2119/molmed.2014.00011.
48. Baxter RC. Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities. *American journal of physiology. Endocrinology and metabolism*. 2000;278(6): E967-76. https://doi.org/10.1152/ajpendo.2000.278.6.E967.
49. Werner H. The IGF1 Signaling Pathway: From Basic Concepts to Therapeutic Opportunities. *International journal of molecular sciences*. 2023;24(19). https://doi.org/10.3390/ijms241914882.
50. Duan C. Specifying the cellular responses to IGF signals: roles of IGF-binding proteins. *Journal of Endocrinology*. 2002;175(1): 41–54. https://doi.org/10.1677/joe.0.1750041.
51. Bach LA. What Happened to the IGF Binding Proteins? *Endocrinology*. 2018;159(2): 570–578. https://doi.org/10.1210/en.2017-00908.
52. Baxter RC. IGF binding proteins in cancer: mechanistic and clinical insights. *Nature reviews. Cancer*. 2014;14(5): 329–341. https://doi.org/10.1038/nrc3720.
53. Cohen P. Overview of the IGF-I System. *Hormone Research in Paediatrics*. 2006;65(suppl 1(Suppl. 1): 3–8. https://doi.org/10.1159/000090640.
54. David A, Hwa V, Metherell LA, Netchine I, Camacho-Hübner C, Clark AJL, et al. Evidence for a continuum of genetic, phenotypic, and biochemical abnormalities in children with growth hormone insensitivity. *Endocrine Reviews*. 2011;32(4): 472–497. https://doi.org/10.1210/er.2010-0023.
55. Savage MO, Hwa V, David A, Rosenfeld RG, Metherell LA. Genetic defects in the growth hormone-IGF-I axis causing growth hormone insensitivity and impaired linear growth. *Frontiers in Endocrinology*. 2011;2(DEC). https://doi.org/10.3389/fendo.2011.00095.
56. Lin S, Li C, Li C, Zhang X. Growth Hormone Receptor Mutations Related to Individual Dwarfism. *International journal of molecular sciences*. 2018;19(5). https://doi.org/10.3390/ijms19051433.
57. Domené HM, Fierro-Carrión G. Genetic disorders of GH action pathway. *Growth Hormone and IGF Research*. 2018;38(October): 19–23. https://doi.org/10.1016/j.ghir.2017.12.004.
58. David A, Hwa V, Metherell LA, Camacho-hu C, Clark AJL, Rosenfeld RG, et al. Evidence for a Continuum of Genetic , Phenotypic , and Biochemical Abnormalities in Children with Growth Hormone Insensitivity. 2011;32(August): 472–497. https://doi.org/10.1210/er.2010-0023.
59. Server EV. A novel dominant negative mutation in the intracellular domain of GHR is associated with growth hormone insensitivity. 2016; 669–671.
60. Ayling R, Ross R, Towner P, Von Laue S, Finidori J, Moutoussamy S, et al. A dominant-negative mutation of the growth hormone receptor causes familial short stature. *Nature Genetics*. 1997;15: 57–61.
61. Iida K, Takahashi Y, Kaji H, Nose O, Okimura Y, Abe H, et al. Growth Hormone (GH) Insensitivity Syndrome with High Serum GH-Binding Protein Levels Caused by a Heterozygous Splice Site Mutation of the GH Receptor Gene Producing a Lack of Intracellular Domain 1 . *The Journal of Clinical Endocrinology & Metabolism*. 1998;83(2): 531–537. https://doi.org/10.1210/jcem.83.2.4601.
62. Vairamani K, Merjaneh L, Casano-Sancho P, Sanli ME, David A, Metherell LA, et al. Novel Dominant-Negative GH Receptor Mutations Expands the Spectrum of GHI and IGF-I Deficiency. *Journal of the Endocrine Society*. 2017;1(4): 345–358. https://doi.org/10.1210/js.2016-1119.
63. Metherell L a, Akker S a, Munroe PB, Rose SJ, Caulfield M, Savage MO, et al. Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. *American journal of human genetics*. 2001;69(3): 641–646. https://doi.org/10.1086/323266.

64. Laron Z, Pertzelan A, Mannheimer S. Genetic pituitary dwarfism with high serum concentation of growth hormone--a new inborn error of metabolism ? *Israel journal of medical sciences*. 1966;2(2): 152–155.
65. Eshet R, Laron Z, Pertzelan A, Arnon R, Dintzman M. Defect of human growth hormone receptors in the liver of two patients with Laron-type dwarfism. *Israel journal of medical sciences*. 1984;20(1): 8–11.
66. Amselem S, Duquesnoy P, Attree O, Novelli G, Bousnina S, Postel-Vinay MC, et al. Laron Dwarfism and Mutations of the Growth Hormone–Receptor Gene. *New England Journal of Medicine*. 1989;321(15): 989–995. https://doi.org/10.1056/NEJM198910123211501.
67. Godowski PJ, Leung DW, Meachamt LR, Galganit JP, Hellmiss R, Keret R, et al. Characterization of the human growth hormone receptor gene and demonstration of. 1989;86(October): 8083–8087.
68. Janecka A, Kołodziej-Rzepa M, Biesaga B. Clinical and Molecular Features of Laron Syndrome, A Genetic Disorder Protecting from Cancer. *In vivo (Athens, Greece)*. 2016;30(4): 375–381.
69. Laron Z. Laron Syndrome (Primary Growth Hormone Resistance or Insensitivity): The Personal Experience 1958 – 2003. 2004;89(3): 1031–1044. https://doi.org/10.1210/jc.2003-031033.
70. Laron Z, Kauli R. Fifty seven years of follow-up of the Israeli cohort of Laron Syndrome patients-From discovery to treatment. *Growth Hormone and IGF Research*. 2016;28: 53–56. https://doi.org/10.1016/j.ghir.2015.08.004.
71. Guevara-Aguirre J, Procel P, Guevara C, Guevara-Aguirre M, Rosado V, Teran E. Despite higher body fat content, Ecuadorian subjects with Laron syndrome have less insulin resistance and lower incidence of diabetes than their relatives. *Growth Hormone and IGF Research*. 2016;28: 76–78. https://doi.org/10.1016/j.ghir.2015.08.002.
72. Guevara-Aguirre J, Rosenbloom AL, Balasubramanian P, Teran E, Guevara-Aguirre M, Guevara C, et al. GH receptor deficiency in ecuadorian adults is associated with obesity and enhanced insulin sensitivity. *Journal of Clinical Endocrinology and Metabolism*. 2015;100(7): 2589–2596. https://doi.org/10.1210/jc.2015-1678.
73. Rosenbloom AL, Guevara-aguirre J. The Little Women of Loja — Growth Hormone – Receptor Deficiency in an Inbred Population of Southern Ecuador. 1990;(December). https://doi.org/10.1056/NEJM199011153232002.
74. Berg MA, Guevara-aguirre J, Rosenbloom AL, Rosenfeld RG, Francke U. n the Growth Mutation Creating a New Splice Site i Hormone Receptor Genes of 37 Ecuadorean Patients ith Laron Syndrome. 1992;2434.
75. Laron Z. Diagnosis of Laron Syndrome BT - Laron Syndrome - From Man to Mouse: Lessons from Clinical and Experimental Experience. In: Laron Z, Kopchick J (eds.) Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 27–28. https://doi.org/10.1007/978-3-642-11183-9_4.
76. Laron Z, Klinger B. Body Fat in Laron Syndrome Patients: Effect of Insulin-Like Growth Factor I Treatment. *Hormone Research in Paediatrics*. 1993;40(1–3): 16–22. https://doi.org/10.1159/000183762.
77. Guevara-Aguirre J, Rosenbloom AL. Obesity, diabetes and cancer: insight into the relationship from a cohort with growth hormone receptor deficiency. *Diabetologia*. 2015;58(1): 37–42. https://doi.org/10.1007/s00125-014-3397-3.
78. Backeljauw PF, Dattani MT, Cohen P, Rosenfeld RG. *Disorders of growth hormone/insulin-like growth factor secretion and action*.. FOURTH EDI. Pediatric Endocrinology: Fourth Edition. Elsevier Inc.; 2014. https://doi.org/10.1016/B978-1-4557-4858-7.00019-6.
79. Metherell LA, Akker SA, Munroe PB, Rose SJ, Caulfield M, Savage MO, et al. Pseudoexon Activation as a Novel Mechanism for Disease Resulting in Atypical Growth-Hormone

Insensitivity. 2001; 641–646.
80. David A, Camacho-Hübner C, Bhangoo A, Rose SJ, Miraki-Moud F, Akker SA, et al. An intronic growth hormone receptor mutation causing activation of a pseudoexon is associated with a broad spectrum of growth hormone insensitivity phenotypes. *Journal of Clinical Endocrinology and Metabolism*. 2007;92(2): 655–659. https://doi.org/10.1210/jc.2006-1527.
81. Derr MA, Aisenberg J, Fang P, Tenenbaum-Rakover Y, Rosenfeld RG, Hwa V. The Growth Hormone Receptor (GHR) c.899dupC mutation functions as a dominant negative: Insights into the pathophysiology of intracellular GHR defects. *Journal of Clinical Endocrinology and Metabolism*. 2011;96(11). https://doi.org/10.1210/jc.2011-1597.
82. Ross RJM, Esposito N, Shen XY, Von Laue S, Chew SL, Dobson PRM, et al. A short isoform of the human growth hormone receptor functions as a dominant negative inhibitor of the full-length receptor and generates large amounts of binding protein. *Molecular Endocrinology*. 1997;11(3): 265–273. https://doi.org/10.1210/mend.11.3.9901.
83. Govers R, Ten Broeke T, Van Kerkhof P, Schwartz AL, Strous GJ. Identification of a novel ubiquitin conjugation motif, required for ligand-induced internalization of the growth hormone receptor. *EMBO Journal*. 1999;18(1): 28–36. https://doi.org/10.1093/emboj/18.1.28.
84. David A, Savage MO, Vairamani K, Merjaneh L, Casano-sancho P. Novel Dominant-Negative GH Receptor Mutations Expands the Spectrum of GHI and IGF-I Deficiency Novel Dominant-Negative GH Receptor Mutations Expands the Spectrum of GHI and IGF-I Deficiency. 2017;(October). https://doi.org/10.1210/js.2016-1119.
85. Rosenfeld RG, Belgorosky A, Camacho-hubner C, Savage MO, Wit JM, Hwa V. Defects in growth hormone receptor signaling. 2007;18(4). https://doi.org/10.1016/j.tem.2007.03.004.
86. Kofoed EM, Hwa V, Little B, Woods KA, Buckway CK, Tsubaki J, et al. Growth hormone insensitivity associated with a STAT5b mutation. *New England Journal of Medicine*. 2003;349(12): 1139–1147. https://doi.org/10.1056/NEJMoa022926.
87. Hwa V. STAT5B deficiency: Impacts on human growth and immunity. *Growth Hormone and IGF Research*. 2016;28: 16–20. https://doi.org/10.1016/j.ghir.2015.12.006.
88. Domene H, Bengolea S, Martinez A, Ropelato M, Pennisi P, Scaglia P, et al. Deficiency of the Circulating Insulin-like Growth Factor System Associated with Inactivation of the Acid-Labile Subunit Gene. 2004; 570–577.
89. Dauber A, Muñoz-calvo MT, Barrios V, Domené HM, Kloverpris S, Serra-juhé C, et al. Mutations in pregnancy-associated plasma protein A 2 cause short stature due to low IGF-I availability. 2016;8(4): 363–374.
90. Woods KA, Camacho-Hübner C, Savage MO, Clark AJL. Intrauterine Growth Retardation and Postnatal Growth Failure Associated with Deletion of the Insulin-Like Growth Factor I Gene. *New England Journal of Medicine*. 1996;335(18): 1363–1367. https://doi.org/10.1056/NEJM199610313351805.
91. Cohen AC, Nadeau KC, Tu W, Hwa V, Dionis K, Bezrodnik L, et al. Cutting Edge: Decreased Accumulation and Regulatory Function of CD4 + CD25 high T Cells in Human STAT5b Deficiency . *The Journal of Immunology*. 2006;177(5): 2770–2774. https://doi.org/10.4049/jimmunol.177.5.2770.
92. Nadeau K, Hwa V, Rosenfeld RG. STAT5b deficiency: An unsuspected cause of growth failure, immunodeficiency, and severe pulmonary disease. *Journal of Pediatrics*. 2011;158(5): 701–708. https://doi.org/10.1016/j.jpeds.2010.12.042.
93. Domené HM, Hwa V, Argente J, Wit JM, Camacho-Hübner C, Jasper HG, et al. Human Acid-Labile Subunit Deficiency: Clinical, Endocrine and Metabolic Consequences. *Hormone*

Research in Paediatrics. 2009;72(3): 129–141. https://doi.org/10.1159/000232486.
94. Buckway CK, Tsubaki J, Pratt KL, Bezrodnik L, Jasper H, Tepper A, et al. Growth Hormone Insensitivity Associated with a. 2003; 1139–1147.
95. Van Duyvenvoorde HA, Van Setten PA, Walenkamp MJE, Van Doorn J, Koenig J, Gauguin L, et al. Short stature associated with a novel heterozygous mutation in the insulin-like growth factor 1 gene. *Journal of Clinical Endocrinology and Metabolism*. 2010;95(11): 363–367. https://doi.org/10.1210/jc.2010-0511.
96. Walenkamp MJE, Losekoot M, Wit JM. Molecular IGF-1 and IGF-1 receptor defects: From genetics to clinical management. *Hormone Resistance and Hypersensitivity: From Genetics to Clinical Management*. 2013;24(February): 128–137. https://doi.org/10.1159/000342841.
97. Bonapace G, Concolino D, Formicola S, Strisciuglio P. *A novel mutation in a patient with insulin-like growth factor 1 (IGF1) deficiency.* Journal of medical genetics. 2003. p. 913–917. https://doi.org/10.1136/jmg.40.12.913.
98. Walenkamp MJE, Karperien M, Pereira AM, Hilhorst-Hofstee Y, van Doorn J, Chen JW, et al. Homozygous and Heterozygous Expression of a Novel Insulin-Like Growth Factor-I Mutation. *The Journal of Clinical Endocrinology & Metabolism*. 2005;90(5): 2855–2864. https://doi.org/10.1210/jc.2004-1254.
99. Walenkamp MJE, van der Kamp HJ, Pereira AM, Kant SG, van Duyvenvoorde HA, Kruithof MF, et al. A Variable Degree of Intrauterine and Postnatal Growth Retardation in a Family with a Missense Mutation in the Insulin-Like Growth Factor I Receptor. *The Journal of Clinical Endocrinology & Metabolism*. 2006;91(8): 3062–3070. https://doi.org/10.1210/jc.2005-1597.
100. Irène Netchine, Salah Azzi, Muriel Houang, Danielle Seurin, Laurence Perin, Jean-Marc Ricort, Claudine Daubas, Christine Legay, Jan Mester, Robert Herich, François Godeau and YLB. Partial Primary Deficiency of Insulin-Like Growth Factor (IGF)-I Activity Associated with IGF1 Mutation Demonstrates Its Critical Role in Growth and Brain Development. *Molecular endocrinology (Baltimore, Md.)*. 2009;23(11): 1936. https://doi.org/10.1210/mend.23.11.9994.
101. Solomon-Zemler R, Basel-Vanagaite L, Steier D, Yakar S, Mel E, Phillip M, et al. A novel heterozygous IGF-1 receptor mutation associated with hypoglycemia. *Endocrine connections*. 2017;6(6): 395–403. https://doi.org/10.1530/EC-17-0038.
102. Batey L, Moon JE, Yu Y, Wu B, Hirschhorn JN, Shen Y, et al. A novel deletion of IGF1 in a patient with idiopathic short stature provides insight into IGF1 haploinsufficiency. *Journal of Clinical Endocrinology and Metabolism*. 2014;99(1): 153–159. https://doi.org/10.1210/jc.2013-3106.
103. Fuqua JS, Derr M, Rosenfeld RG, Hwa V. Identification of a novel heterozygous IGF1 splicing mutation in a large kindred with familial short stature. *Hormone research in paediatrics*. 2012;78(1): 59–66. https://doi.org/10.1159/000337249.
104. Abuzzahab J, Schneider A, Goddard A, Grigorescu F, Lautier C, Keller E, et al. IGF-I Receptor Mutations Resulting in Intrauterine and Postnatal Growth Retardation. 2003; 2211–2222.
105. Fang P, Cho YH, Derr MA, Rosenfeld RG, Hwa V, Cowell CT. Severe short stature caused by novel compound heterozygous mutations of the insulin-like growth factor 1 receptor (IGF1R). *Journal of Clinical Endocrinology and Metabolism*. 2012;97(2): 243–247. https://doi.org/10.1210/jc.2011-2142.
106. Gannagé-Yared MH, Klammt J, Chouery E, Corbani S, Meǵarbané H, Ghoch JA, et al. Homozygous mutation of the IGF1 receptor gene in a patient with severe pre- and postnatal growth failure and congenital malformations. *European Journal of Endocrinology*. 2013;168(1): 1–7. https://doi.org/10.1530/EJE-12-0701.
107. Prontera P, Micale L, Verrotti A, Napolioni V, Stangoni G, Merla G. *A New Homozygous*

IGF1R Variant Defines a Clinically Recognizable Incomplete Dominant form of SHORT Syndrome. Human Mutation. 2015. p. 1043–1047. https://doi.org/10.1002/humu.22853.
108. Klammt J, Kiess W, Pfäffle R. IGF1R mutations as cause of SGA. *Best Practice and Research: Clinical Endocrinology and Metabolism*. 2011;25(1): 191–206. https://doi.org/10.1016/j.beem.2010.09.012.
109. Shapiro L, Chatterjee S, G Ramadan D, Davies K, Savage M, Metherell L, et al. *Whole-exome sequencing gives additional benefits compared to candidate gene sequencing in the molecular diagnosis of children with growth hormone or IGF-1 insensitivity*. European Journal of Endocrinology. 2017. https://doi.org/10.1530/EJE-17-0453.
110. Savage M, Dunger DB, Hospital JR. Noonan ' s Syndrome : Abnormalities of the Growth Hormone / IGF-l Axis and the Response to Treatment with Human Biosynthetic Growth Hormone. 1991; 446–450.
111. Binder G, Seidel A kathrin, Martin DD, Schweizer R, Schwarze CP, Wollmann HA, et al. The Endocrine Phenotype in Silver-Russell Syndrome Is Defined by the Underlying Epigenetic Alteration. 2007;93(November): 1402–1407. https://doi.org/10.1210/jc.2007-1897.
112. Roberts AE, Allanson JE, Tartaglia M, Gelb BD, Ematologia D. Noonan syndrome. *The Lancet*. 2013;6736(12): 1–10. https://doi.org/10.1016/S0140-6736(12)61023-X.
113. Tartaglia M, Ph D, Gelb BD, Zenker M, Ph D. Noonan syndrome and clinically related disorders. *Best Practice & Research Clinical Endocrinology & Metabolism*. 2011;25(1): 161–179. https://doi.org/10.1016/j.beem.2010.09.002.
114. Tidyman WE, Rauen KA, Pediatrics D, Davis C. Expansion of the RASopathies. 2017;4(3): 57–64. https://doi.org/10.1007/s40142-016-0100-7.Expansion.
115. Kim S oh, Jiang J, Yi W. Involvement of the Src Homology 2-containing Tyrosine Phosphatase SHP-2 in Growth Hormone Signaling CELL BIOLOGY AND METABOLISM : Involvement of the Src Homology 2-containing Tyrosine Phosphatase SHP-2 in Growth Hormone Signaling. 1998;(September 2014). https://doi.org/10.1074/jbc.273.4.2344.
116. Stofega MR. Mutation of the SHP-2 Binding Site in Growth Hormone (GH) Receptor Prolongs GH-Promoted Tyrosyl Phosphorylation of GH Receptor, JAK2, and STAT5B. *Molecular Endocrinology*. 2000;14(9): 1338–1350. https://doi.org/10.1210/me.14.9.1338.
117. Limal JM, Parfait B, Cabrol S, Bonnet D, Leheup B, Lyonnet S, et al. Noonan syndrome: relationships between genotype, growth, and growth factors. *The Journal of clinical endocrinology and metabolism*. 2006;91(1): 300–306. https://doi.org/10.1210/jc.2005-0983.
118. Yart A, Edouard T. Noonan syndrome: an update on growth and development. *Current opinion in endocrinology, diabetes, and obesity*. 2018;25(1): 67–73. https://doi.org/10.1097/MED.0000000000000380.
119. Padidela R, Camacho-Hübner C, Attie KM, Savage MO. Abnormal growth in Noonan syndrome: Genetic and endocrine features and optimal treatment. *Hormone Research*. 2008;70(3): 129–136. https://doi.org/10.1159/000145016.
120. Lv R, Wu X, Neel BG, Araki T, Wu X, Simpson J, et al. MEK-ERK pathway modulation ameliorates disease phenotypes in a mouse model of Noonan syndrome associated with the Find the latest version : MEK-ERK pathway modulation ameliorates disease phenotypes in a mouse model of Noonan syndrome associated with the R. 2011;121(3): 1009–1025. https://doi.org/10.1172/JCI44929.Delineating.
121. Chen PC, Wakimoto H, Conner D, Araki T, Yuan T, Roberts A, et al. Activation of multiple signaling pathways causes developmental defects in mice with a Noonan syndrome-associated Sos1 mutation. *Journal of Clinical Investigation*. 2010;120(12): 4353–4365. https://doi.org/10.1172/JCI43910.
122. Miller JD, McKusick VA, Malvaux P, Temtamy S, Salinas C. The 3-M syndrome: a heritable

low birthweight dwarfism. *Birth defects original article series*. 1975;11(5): 39–47.
123. Dauber A, Stoler J, Hechter E, Safer J, Hirschhorn J. Whole exome sequencing reveals a novel mutation in CUL7 in a patient with an undiagnosed growth disorder. *Journal of Pediatrics*. 2014;162(1): 202–204. https://doi.org/10.1016/j.jpeds.2012.07.055.Whole.
124. Hanson D, Murray PG. The Genetics of 3-M Syndrome : Unravelling a Potential New Regulatory. *Hormone Research in Paediatrics*. 2011;(December). https://doi.org/10.1159/000334392.
125. Demir K, Altincik A, Bober E. Severe short stature due to 3-M syndrome with a novel OBSL1 gene mutation. *Journal of pediatric endocrinology & metabolism : JPEM*. 2013;26(1–2): 147–150. https://doi.org/10.1515/jpem-2012-0239.
126. Lugli L, Bertucci E, Mazza V, Elmakky A, Ferrari F, Neuhaus C, et al. Pre- and post-natal growth in two sisters with 3-M syndrome. *European Journal of Medical Genetics*. 2016; https://doi.org/10.1016/j.ejmg.2016.01.009.
127. Liao L, Gan HW, Hwa V, Dattani M, Dauber A. Two Siblings with a Mutation in CCDC8 Presenting with Mild Short Stature: A Case of 3-M Syndrome. *Hormone research in paediatrics*. 2017;88(5): 364–370. https://doi.org/10.1159/000477907.
128. Keskin M, Şahin NM, Kurnaz E, Bayramoğlu E, Erdeve ŞS, Aycan Z, et al. A Rare Cause of Short Stature : 3M Syndrome in a Patient with Novel Mutation in OBSL1 Gene. *J Clin Res Pediatr Endocrinol*. 2017;9(1): 91–94.
129. Hanson D, Murray PG, Coulson T, Sud A, Omokanye A, Stratta E, et al. Mutations in CUL7 , OBSL1 and CCDC8 in 3-M syndrome lead to disordered growth factor signalling. *Journal of Molecular Endocrinology*. 2012; 267–275. https://doi.org/10.1530/JME-12-0034.
130. Azzi S, Salem J, Thibaud N, Chantot-Bastaraud S, Lieber E, Netchine I, et al. A prospective study validating a clinical scoring system and demonstrating phenotypical-genotypical correlations in Silver-Russell syndrome. *Journal of Medical Genetics*. 2015;52(7): 446–453. https://doi.org/10.1136/jmedgenet-2014-102979.
131. Wakeling EL, Brioude F, Lokulo-Sodipe O, O'Connell SM, Salem J, Bliek J, et al. Diagnosis and management of Silver-Russell syndrome: First international consensus statement. *Nature Reviews Endocrinology*. 2017;13(2): 105–124. https://doi.org/10.1038/nrendo.2016.138.
132. Storr HL, Dunkel L, Kowalczyk J, Savage MO, Metherell LA. Genetic characterisation of a cohort of children clinically labelled as GH or IGF1 insensitive: Diagnostic value of serum IGF1 and height at presentation. *European Journal of Endocrinology*. 2015;172(2): 151–161. https://doi.org/10.1530/EJE-14-0541.
133. Iliev DI, Kannenberg K, Weber K, Binder G. IGF-I sensitivity in Silver – Russell syndrome with IGF2 / H19 hypomethylation. *YGHIR*. 2014;24(5): 187–191. https://doi.org/10.1016/j.ghir.2014.06.005.
134. Binder G, Seidel AK, Martin DD, Schweizer R, Schwarze CP, Wollmann HA, et al. The endocrine phenotype in Silver-Russell syndrome is defined by the underlying epigenetic alteration. *Journal of Clinical Endocrinology and Metabolism*. 2008;93(4): 1402–1407. https://doi.org/10.1210/jc.2007-1897.
135. Pinheiro LT, Silva RK de S, Barbosa RPT, Moura RRL, Parente DM. The Genetics of Sickle Cell Anemia: A Literature Review. *International Journal of Engineering & Technology IJET-IJENS*. 2015;15(36): 1510403–1511503. http://www.ijens.org/Vol_15_I_03/1410406-1503-9292-IJET-IJENS.pdf
136. Benhabiles H, Jia J, Lejeune F. Chapter 1 - General Aspects Related to Nonsense Mutations. In: Benhabiles H, Jia J, Lejeune FBTNMC in HD (eds.) *Nonsense Mutation Correction in Human Diseases*. Boston: Academic Press; 2016. p. 1–76. https://doi.org/https://doi.org/10.1016/B978-0-12-804468-1.00001-4.
137. Maharaj A, Buonocore F, Meimaridou E, Ruiz-Babot G, Guasti L, Peng HM, et al. Predicted

Benign and Synonymous Variants in CYP11A1 Cause Primary Adrenal Insufficiency Through Missplicing. *Journal of the Endocrine Society*. 2019;3(1): 201–221. https://doi.org/10.1210/js.2018-00130.
138. Gilbert W. Why genes in pieces? *Nature*. 1978;271(5645): 501. https://doi.org/10.1038/271501a0.
139. Black DL. Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology. *Cell*. 2000;103(3): 367–370. https://doi.org/https://doi.org/10.1016/S0092-8674(00)00128-8.
140. EJ W, MA GB. Polypyrimidine tract binding protein antagonizes exon definition. *Molecular and Cellular Biology*. 2001;(August): 3281–3288. https://doi.org/10.1128/MCB.21.10.3281.
141. Pan K, Lee JTH, Huang Z, Wong CM. Coupling and coordination in gene expression processes with pre-mRNA splicing. *International Journal of Molecular Sciences*. 2015;16(3): 5682–5696. https://doi.org/10.3390/ijms16035682.
142. Moore MJ, Query CC, Sharp PA. *Splicing of Precursors to mRNA by the Spliceosome*. The RNA World. 1993. www.cshlpress.com/copyright.
143. Wilson EO. The encyclopedia of life. *Trends in Ecology and Evolution*. 2003;18(2): 77–80. https://doi.org/10.1016/S0169-5347(02)00040-X.
144. Lee Y, Rio DC, Biology S, Biology C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*. 2015;(258): 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316.Mechanisms.
145. Black DL. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*. 2003;72(1): 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720.
146. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 2008;40(12): 1413–1415. https://doi.org/10.1038/ng.259.
147. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(27): 11093–11098. https://doi.org/10.1073/pnas.1101135108.
148. David A, Miraki-Moud F, Shaw NJ, Savage MO, Clark AJL, Metherell LA. Identification and characterisation of a novel GHR defect disrupting the polypyrimidine tract and resulting in GH insensitivity. *European Journal of Endocrinology*. 2010;162(1): 37–42. https://doi.org/10.1530/EJE-09-0583.
149. Corvelo A, Eyras E. Exon creation and establishment in human genes. *Genome Biology*. 2008;9(9). https://doi.org/10.1186/gb-2008-9-9-r141.
150. Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, et al. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Research*. 2004;32(5): 1783–1791. https://doi.org/10.1093/nar/gkh341.
151. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell*. 2004;119(6): 831–845. https://doi.org/10.1016/j.cell.2004.11.010.
152. Zhang XHF, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. *Genes and Development*. 2004;18(11): 1241–1250. https://doi.org/10.1101/gad.1195304.
153. Dhir A, Buratti E. Alternative splicing: Role of pseudoexons in human disease and potential therapeutic strategies: Minireview. *FEBS Journal*. 2010;277(4): 841–855. https://doi.org/10.1111/j.1742-4658.2009.07520.x.
154. Vorechovsky I. Transposable elements in disease-associated cryptic exons. *Human Genetics*.

2010;127(2): 135–154. https://doi.org/10.1007/s00439-009-0752-4.
155. Auwerx C, Lepamets M, Sadler MC, Patxot M, Stojanov M, Baud D, et al. The individual and global impact of copy-number variants on complex human traits. *American Journal of Human Genetics*. 2022;109(4): 647–668. https://doi.org/10.1016/j.ajhg.2022.02.010.
156. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics*. 2009;10: 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217.
157. Singleton A., Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. Alpha-Synuclein locus triplication causes Parkinson's disease. *Science*. 2003;302(5646): 841. https://doi.org/10.1126/science.1090278.
158. Bernier R, Steinman KJ, Reilly B, Wallace AS, Sherr EH, Pojman N, et al. Clinical phenotype of the recurrent 1q21.1 copy-number variant. *Genetics in Medicine*. 2016;18(4): 341–349. https://doi.org/10.1038/gim.2015.78.
159. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. *American Journal of Human Genetics*. 2008;82(2): 477–488. https://doi.org/10.1016/j.ajhg.2007.12.009.
160. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316(5823): 445–449. https://doi.org/10.1126/science.1138659.
161. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005;307(5714): 1434–1440. https://doi.org/10.1126/science.1101160.
162. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A Chromosome 8 Gene-Cluster Polymorphism with Low Human Beta-Defensin 2 Gene Copy Number Predisposes to Crohn Disease of the Colon. *The American Journal of Human Genetics*. 2006;79(3): 439–448. https://doi.org/10.1086/505915.
163. Hollox EJ, Huffmeier U, Zeeuwen PLJM, Palla R, Lascorz J, Rodijk-Olthuis D, et al. Psoriasis is associated with increased β-defensin genomic copy number. *Nature Genetics*. 2008;40(1): 23–25. https://doi.org/10.1038/ng.2007.48.
164. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. 2013;15(6): 478–481. https://doi.org/10.1038/gim.2012.164.
165. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010;7(4): 248–249. https://doi.org/10.1038/nmeth0410-248.
166. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*. 2003;31(13): 3812–3814. https://doi.org/10.1093/nar/gkg509.
167. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*. 2019;47(D1): D886–D894. https://doi.org/10.1093/nar/gky1016.
168. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*. 2014;46(3): 310–315. https://doi.org/10.1038/ng.2892.
169. Richards S, Aziz N, Bale S, Bick D, Das S. ACMG Standards and Guidelines Standards and guidelines for the interpretation of sequence variants : a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. 2015;(March): 1–20. https://doi.org/10.1038/gim.2015.30.
170. Aspden JL, Wallace EWJ, Whiffin N. Not all exons are protein coding: Addressing a common misconception. *Cell genomics*. 2023;3(4): 100296.

https://doi.org/10.1016/j.xgen.2023.100296.
171. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*. 2017;543(7644): 199–204. https://doi.org/10.1038/nature21374.
172. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics*. 2015;47(3): 199–208. https://doi.org/10.1038/ng.3192.
173. Ramilowski JA, Yip CW, Agrawal S, Chang JC, Ciani Y, Kulakovskiy I V, et al. Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome research*. 2020;30(7): 1060–1072. https://doi.org/10.1101/gr.254219.119.
174. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature reviews. Genetics*. 2016;17(10): 601–614. https://doi.org/10.1038/nrg.2016.85.
175. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science (New York, N.Y.)*. 2005;309(5740): 1559–1563. https://doi.org/10.1126/science.1112014.
176. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, et al. Systematic discovery of Xist RNA binding proteins. *Cell*. 2015;161(2): 404–416. https://doi.org/10.1016/j.cell.2015.03.025.
177. Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova I V, Sanbonmatsu KY, et al. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Molecular cell*. 2016;64(1): 37–50. https://doi.org/10.1016/j.molcel.2016.08.010.
178. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291): 1071–1076. https://doi.org/10.1038/nature08975.
179. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012;491(7424): 454–457. https://doi.org/10.1038/nature11508.
180. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature reviews. Molecular cell biology*. 2016;17(12): 756–770. https://doi.org/10.1038/nrm.2016.126.
181. Yamoto K, Hasegawa T, Ogata T, Fujisawa Y. De novo IGF2 mutation on the paternal allele in a patient with Silver – Russell syndrome and ectrodactyly. https://doi.org/10.1002/humu.23253.
182. Liu D, Wang Y, Yang X an, Liu D. De Novo Mutation of Paternal IGF2 Gene Causing Silver – Russell. 2017;8(August): 1–6. https://doi.org/10.3389/fgene.2017.00105.
183. Begemann M, Zirn B, Santen G, Wirthgen E, Ph D, Soellner L, et al. Paternally Inherited IGF2 Mutation and Growth Restriction. 2017; 349–356. https://doi.org/10.1056/NEJMoa1415227.
184. Abi Habib W, Brioude F, Edouard T, Bennett JT, Lienhardt-Roussie A, Tixier F, et al. Genetic disruption of the oncogenic HMGA2–PLAG1–IGF2 pathway causes fetal growth restriction. *GENETICS in MEDICINE*. 2017;00(August). https://doi.org/10.1038/gim.2017.105.
185. Buysse K, Reardon W, Mehta L, Costa T, Fagerstrom C, Kingsbury DJ, et al. The 12q14 microdeletion syndrome: Additional patients and further evidence that HMGA2 is an important genetic determinant for human height. *European Journal of Medical Genetics*. 2009;52(2–3): 101–107. https://doi.org/10.1016/j.ejmg.2009.03.001.
186. Crescenzo A De, Citro V, Freschi A, Sparago A, Palumbo O, Cubellis MV, et al. ORIGINAL ARTICLE A splicing mutation of the HMGA2 gene is associated with Silver – Russell

syndrome phenotype. 2015;60(6): 287–293. https://doi.org/10.1038/jhg.2015.29.
187. Leszinski GS, Warncke K, Hoefele J, Wagner M. A case report and review of the literature indicate that HMGA2 should be added as a disease gene for Silver-Russell syndrome. *Gene*. 2018;663: 110–114. https://doi.org/10.1016/j.gene.2018.04.027.
188. Habib WA, Brioude F, Edouard T, Bennett JT, Lienhardt-roussie A, Tixier F, et al. Genetic disruption of the oncogenic HMGA2 – PLAG1 – IGF2 pathway causes fetal growth restriction. *Nature Publishing Group*. 2017;00(March). https://doi.org/10.1038/gim.2017.105.
189. Ahn JW, Mann K, Walsh S, Shehab M, Hoang S, Docherty Z, et al. Validation and implementation of array comparative genomic hybridisation as a first line test in place of postnatal karyotyping for genome imbalance. *Molecular Cytogenetics*. 2010;3(1): 1–8. https://doi.org/10.1186/1755-8166-3-9.
190. Hoang S, Ahn J, Mann K, Bint S, Mansour S, Homfray T, et al. Detection of mosaicism for genome imbalance in a cohort of 3,042 clinical cases using an oligonucleotide array CGH platform. *European Journal of Medical Genetics*. 2011;54(2): 121–129. https://doi.org/10.1016/j.ejmg.2010.10.010.
191. Ellard S, Baple EL, Owens M, Eccles DM, Turnbull C, Scott R, et al. ACGS Best Practice Guidelines for Variant Classification 2018. 2018; 1–16.
192. Wendelsdorf K, Shah S. Empowered genome community: Leveraging a bioinformatics platform as a citizen-scientist collaboration tool. *Applied and Translational Genomics*. 2015;6: 7–10. https://doi.org/10.1016/j.atg.2015.08.002.
193. Desmet FO, Hamroun D, Lalande M, Collod-Bëroud G, Claustres M, Béroud C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*. 2009;37(9): 1–14. https://doi.org/10.1093/nar/gkp215.
194. James Kent W, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Research*. 2002;12(6): 996–1006. https://doi.org/10.1101/gr.229102. Article published online before print in May 2002.
195. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE Data in the UCSC Genome Browser: Year 5 update. *Nucleic Acids Research*. 2013;41(D1): 56–63. https://doi.org/10.1093/nar/gks1172.
196. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1): 59. https://doi.org/10.1038/s43586-021-00056-9.
197. Wang SR, Carmichael H, Andrew SF, Miller TC, Moon JE, Derr MA, et al. Large-scale pooled next-generation sequencing of 1077 genes to identify genetic causes of short stature. *Journal of Clinical Endocrinology and Metabolism*. 2013;98(8): 1428–1437. https://doi.org/10.1210/jc.2013-1534.
198. García-magariño M, García R, Garza D. Fanconi Anemia and Laron Syndrome. *The American Journal of the Medical Sciences*. 2017;353(5): 425–432. https://doi.org/10.1016/j.amjms.2017.02.001.
199. Gordon M, Crouthamel C, Post EM, Richman RA. Psychosocial aspects of constitutional short stature: social competence, behavior problems, self-esteem, and family functioning. *The Journal of pediatrics*. 1982;101(3): 477–480.
200. Kjaer R, Fine RS. Rare and low-frequency coding variants alter human adult height. 2016; https://doi.org/10.1038/nature21039.
201. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*. 2017;27(5): 849–864. https://doi.org/10.1101/gr.213611.116.

202. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V, Mikheenko A, et al. The complete sequence of a human genome. *Science (New York, N.Y.)*. 2022;376(6588): 44–53. https://doi.org/10.1126/science.abj6987.
203. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009;6(5): 343–345. https://doi.org/10.1038/nmeth.1318.
204. Frischknecht M, Jagannathan V, Plattet P, Neuditschko M. A Non-Synonymous HMGA2 Variant Decreases Height in Shetland Ponies and Other Small Horses. 2015; 1–11. https://doi.org/10.1371/journal.pone.0140749.
205. Manolopoulou J, Alami Y, Petersenn S, Schopohl J, Wu Z, Strasburger CJ, et al. Automated 22-kD growth hormone-specific assay without interference from pegvisomant. *Clinical Chemistry*. 2012;58(10): 1446–1456. https://doi.org/10.1373/clinchem.2012.188128.
206. Bidlingmaier M, Friedrich N, Emeny RT, Spranger J, Wolthers OD, Roswall J, et al. Reference intervals for insulin-like growth factor-1 (IGF-I) from birth to senescence: Results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations. *Journal of Clinical Endocrinology and Metabolism*. 2014;99(5): 1712–1721. https://doi.org/10.1210/jc.2013-3059.
207. Friedrich N, Wolthers OD, Arafat AM, Emeny RT, Spranger J, Roswall J, et al. Age- and sex-specific reference intervals across life span for insulin-like growth factor binding protein 3 (igfbp-3) and the igf-i to igfbp-3 ratio measured by new automated chemiluminescence assays. *Journal of Clinical Endocrinology and Metabolism*. 2014;99(5): 1675–1686. https://doi.org/10.1210/jc.2013-3060.
208. Stadler S, Wu Z, Dressendörfer RA, Morrison KM, Khare A, Lee PDK, et al. Monoclonal anti-acid-labile subunit oligopeptide antibodies and their use in a two-site immunoassay for ALS measurement in humans. *Journal of Immunological Methods*. 2001;252(1–2): 73–82. https://doi.org/10.1016/S0022-1759(01)00335-0.
209. Rowlinson SW, Behncken SN, Rowland JE, Clarkson RW, Strasburger CJ, Wu Z, et al. Activation of chimeric and full-length growth hormone receptors by growth hormone receptor monoclonal antibodies: A specific conformational change may be required for full-length receptor signaling. *Journal of Biological Chemistry*. 1998;273(9): 5307–5314. https://doi.org/10.1074/jbc.273.9.5307.
210. Amselem S, Sobrier ML, Duquesnoy P, Rappaport R, Postel-Vinay MC, Gourmelen M, et al. Recurrent nonsense mutations in the growth hormone receptor from patients with laron dwarfism. *Journal of Clinical Investigation*. 1991;87(3): 1098–1102. https://doi.org/10.1172/JCI115071.
211. Al-Ashwal AA, Al-Sagheir A, Ramzan K, Al-Owain M, Allam R, Qari A, et al. Clinical, Endocrine, and Molecular Genetic Analysis of a Large Cohort of Saudi Arabian Patients with Laron Syndrome. *Hormone Research in Paediatrics*. 2017;88(2): 119–126. https://doi.org/10.1159/000475991.
212. Duquesnoy P, Sobrier M laure, Duriez B, Dastot F, Buchanan CR, Savage MO, et al. A single amino acid substitution in the exoplasmic domain of the human growth hormone (GH) receptor confers familial GH resistance (Laron syndrome) with positive GH-binding activity by abolishing receptor homodimerization. 1994;13(6): 1386–1395.
213. Guo MH, Shen Y, Walvoord EC, Miller TC, Moon JE, Hirschhorn JN, et al. Whole exome sequencing to identify genetic causes of short stature. *Hormone Research in Paediatrics*. 2014;82(1): 44–52. https://doi.org/10.1159/000360857.
214. Hanson D, Murray PG, O'Sullivan J, Urquhart J, Daly S, Bhaskar SS, et al. Exome sequencing identifies CCDC8 mutations in 3-M syndrome, suggesting that CCDC8 contributes in a pathway with CUL7 and OBSL1 to control human growth. *American Journal of Human*

Genetics. 2011;89(1): 148–153. https://doi.org/10.1016/j.ajhg.2011.05.028.
215. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*. 2004;11(2–3): 377–394. https://doi.org/10.1089/1066527041410418.
216. KW G. *Handbook of physical measurements*.. 3rd Editio. Oxford University Press.;
217. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. *Integrative genomics viewer.* Nature biotechnology. 2011. p. 24–26. https://doi.org/10.1038/nbt.1754.
218. Sobrier ML, Dastot F, Duquesnoy P, Kandemir N, Yordam N, Goossens M, et al. Nine Novel Growth Hormone Receptor Gene Mutations in Patients with Laron Syndrome 1 . *The Journal of Clinical Endocrinology & Metabolism*. 1997;82(2): 435–437. https://doi.org/10.1210/jcem.82.2.3725.
219. Baumbach L, Schiavi A, Bartlett E, Perera E, Day J, Brown M, et al. Clinical , Biochemical , and Molecular Investigations of (Laron ' s Syndrome)*. *Journal of Clinical Endocrinology and Metabolism*. 1997;82(2): 444–451.
220. Maamra M, Milward A, Esfahani HZ, Abbott LP, Metherell LA, Savage MO, et al. A 36 residues insertion in the dimerization domain of the growth hormone receptor results in defective trafficking rather than impaired signaling. *Journal of Endocrinology*. 2006;188(2): 251–261. https://doi.org/10.1677/joe.1.06252.
221. Chatterjee S, Shapiro L, Rose SJ, Mushtaq T, Clayton PE, Ten SB, et al. Phenotypic spectrum and responses to recombinant human IGF1 (rhIGF1) therapy in patients with homozygous intronic pseudoexon growth hormone receptor mutation. 2018;
222. Chatterjee S, Cottrell E, Rose SJ, Mushtaq T, Maharaj A V, Williams J, et al. GHR gene transcript heterogeneity may explain phenotypic variability in GHR pseudoexon (6 Ψ) patients. 2020;(8): 211–222.
223. Berg MA, Argente J, Chernausek S, Gracia R, Guevara-Aguirre J, Hopp M, et al. Diverse growth hormone receptor gene mutations in Laron syndrome. *American Journal of Human Genetics*. 1993;52(5): 998–1005.
224. Rosenbloom AL, Francke U, Berg MA, Kasatkina EP, Volkova TN, Skorobogatova VF, et al. Severe Growth Hormone Insensitivity (Laron Syndrome) Due to Nonsense Mutation of the GH Receptor in Brothers from Russia. *Journal of Pediatric Endocrinology and Metabolism*. 1995;8(3): 159–166. https://doi.org/10.1515/JPEM.1995.8.3.159.
225. Laron Z, Klinger B, Erster B, Silbergeld A. Serum GH binding protein activities identifies the heterozygous carriers for Laron type dwarfism. *Acta Endocrinologica*. 1989;121(4): 603–608. https://doi.org/10.1530/acta.0.1210603.
226. Guevara-aguirre J, Rosenbloom AL, Guevara-aguirre M, Yariz K, Saavedra J, Baumbach L, et al. Effects of heterozygosity for the E180 splice mutation causing growth hormone receptor deficiency in Ecuador on IGF-I , IGFBP-3 , and stature. 2007;17: 261–264. https://doi.org/10.1016/j.ghir.2007.01.016.
227. Iida K, Takahashi Y, Kaji H, Takahashi MO, Okimura Y, Nose O, et al. Functional characterization of truncated growth hormone (GH) receptor- (1-277) causing partial GH insensitivity syndrome with high GH-binding protein. *Journal of Clinical Endocrinology and Metabolism*. 1999;84(3): 1011–1016. https://doi.org/10.1210/jc.84.3.1011.
228. Rosenbloom AL, Guevara-Aguirre J, Rosenfeld RG, Fielder PJ. Is there heterozygote expression of growth hormone receptor deficiency? *Acta Paediatrica*. 1994;83(s399): 125–127. https://doi.org/https://doi.org/10.1111/j.1651-2227.1994.tb13304.x.
229. Klinger B, Ionesco A, Anin S, Laron Z. Effect of insulin-like growth factor I on the thyroid axis in patients with Laron-type dwarfism and healthy subjects. *Acta endocrinologica*. 1992;127(6): 515–519. https://doi.org/10.1530/acta.0.1270515.

230. H. Moayeri, A. Hemati KB and HD. EFFECTS OF GROWTH HORMONE REPLACEMENT THERAPY ON THYROID FUNCTION TESTS IN GROWTH HORMONE. (11).
231. Yamauchi I, Sakane Y, Yamashita T, Hirota K, Ueda Y, Kanai Y, et al. Effects of growth hormone on thyroid function are mediated by type 2 iodothyronine deiodinase in humans. *Endocrine*. 2018;59(2): 353–363. https://doi.org/10.1007/s12020-017-1495-y.
232. Berg MA, Guevara-Aguirre J, Rosenbloom AL, Rosenfeld RG, Francke U. Mutation creating a new splice site in the growth hormone receptor genes of 37 Ecuadorean patients with Laron syndrome. *Human Mutation*. 1992;1(1): 24–34. https://doi.org/https://doi.org/10.1002/humu.1380010105.
233. Pinto M, Guevara-aguirre J, Shevah O, Gonc FT, Fridman C, Rosembloom AL, et al. The E180splice Mutation in the GHR Gene Causing Laron Syndrome : Witness of a Sephardic Jewish Exodus from the Iberian Peninsula to the New World ? 2014;(May). https://doi.org/10.1002/ajmg.a.36444.
234. Zahnleiter D, Uebe S, Ekici AB, Hoyer J, Wiesener A, Wieczorek D, et al. Rare Copy Number Variants Are a Common Cause of Short Stature. *PLoS Genetics*. 2013;9(3): 1–11. https://doi.org/10.1371/journal.pgen.1003365.
235. Dauber A, Yu Y, Turchin MC, Chiang CW, Meng YA, Demerath EW, et al. Genome-wide association of copy-number variation reveals an association between short stature and the presence of low-frequency genomic deletions. *American Journal of Human Genetics*. 2011;89(6): 751–759. https://doi.org/10.1016/j.ajhg.2011.10.014.
236. Canton APM, Costa SS, Rodrigues TC, Bertola DR, Malaquias AC, Correa FA, et al. Genome-wide screening of copy number variants in children born small for gestational age reveals several candidate genes involved in growth pathways. *European Journal of Endocrinology*. 2014;171(2): 253–262. https://doi.org/10.1530/EJE-14-0232.
237. Wit JM, Van Duyvenvoorde HA, Van Klinken JB, Caliebe J, Bosch CAJ, Lui JC, et al. Copy number variants in short children born small for gestational age. *Hormone Research in Paediatrics*. 2014;82(5): 310–318. https://doi.org/10.1159/000367712.
238. Hu G, Fan Y, Wang L, Yao RE, Huang X, Shen Y, et al. Copy number variations in 119 Chinese children with idiopathic short stature identified by the custom genome-wide microarray. *Molecular Cytogenetics*. 2016;9(1): 1–7. https://doi.org/10.1186/s13039-016-0225-0.
239. Van Duyvenvoorde HA, Lui JC, Kant SG, Oostdijk W, Gijsbers ACJ, Hoffer MJ V, et al. Copy number variants in patients with short stature. *European Journal of Human Genetics*. 2014;22(5): 602–609. https://doi.org/10.1038/ejhg.2013.203.
240. Homma T, Krepischi A, Katsue Furuya T, Honjo R, Malaquias A, Bertola D, et al. Recurrent Copy Number Variants Associated with Syndromic Short Stature of Unknown Cause. 2018;903: 13–21. https://doi.org/10.1159/000481777.
241. Fokstuen S, Kotzot D. Chromosomal rearrangements in patients with clinical features of Silver–Russell syndrome. *American Journal of Medical Genetics Part A*. 2014;164(6): 1595–1605. https://doi.org/10.1002/ajmg.a.36464.
242. Bruce S, Hannula-Jouppi K, Puoskari M, Fransson I, Simola KOJ, Lipsanen-Nyman M, et al. Submicroscopic genomic alterations in Silver - Russell syndrome and Silver - Russell-like patients. *Journal of Medical Genetics*. 2010;47(12): 816–822. https://doi.org/10.1136/jmg.2009.069427.
243. Inoue T, Nakamura A, Fuke T, Yamazawa K, Sano S, Matsubara K, et al. Genetic heterogeneity of patients with suspected Silver-Russell syndrome: genome-wide copy number analysis in 82 patients without imprinting defects. *Clinical Epigenetics*. 2017;9(1): 1–10. https://doi.org/10.1186/s13148-017-0350-6.
244. Wakeling EL, Abu Amero S, Alders M, Bliek J, Forsythe E, Kumar S, et al. Epigenotype-phenotype correlations in Silver-Russell syndrome. *Journal of Medical Genetics*.

2010;47(11): 760–768. https://doi.org/10.1136/jmg.2010.079111.
245. Bartholdi D, Krajewska-Walasek M, Õunap K, Gaspar H, Chrzanowska KH, Ilyana H, et al. Epigenetic mutations of the imprinted IGF2-H19 domain in Silver-Russell syndrome (SRS): Results from a large cohort of patients with SRS and SRS-like phenotypes. *Journal of Medical Genetics*. 2009;46(3): 192–197. https://doi.org/10.1136/jmg.2008.061820.
246. Bruce S, Hannula-Jouppi K, Peltonen J, Kere J, Lipsanen-Nyman M. Clinically distinct epigenetic subgroups in Silver-Russell syndrome: The degree of H19 hypomethylation associates with phenotype severity and genital and skeletal anomalies. *Journal of Clinical Endocrinology and Metabolism*. 2009;94(2): 579–587. https://doi.org/10.1210/jc.2008-1805.
247. Russo S, Calzari L, Mussa A, Mainini E, Cassina M, Di Candia S, et al. A multi-method approach to the molecular diagnosis of overt and borderline 11p15.5 defects underlying Silver–Russell and Beckwith–Wiedemann syndromes. *Clinical Epigenetics*. 2016;8(1). https://doi.org/10.1186/s13148-016-0183-8.
248. Spengler S, Begemann M, Ortiz Bruechle N, Baudis M, Denecke B, Michael Kroisel P, et al. *Molecular Karyotyping as a Relevant Diagnostic Tool in Children with Growth Retardation with Silver-Russell Features*. The Journal of pediatrics. 2012. https://doi.org/10.1016/j.jpeds.2012.04.045.
249. Lynch SA, Foulds N, Thuresson AC, Collins AL, Annerén G, Hedberg BO, et al. The 12q14 microdeletion syndrome: Six new cases confirming the role of HMGA2 in growth. *European Journal of Human Genetics*. 2011;19(5): 534–539. https://doi.org/10.1038/ejhg.2010.215.
250. Jaillard S, Andrieux J, Plessis G, Krepischi AC V, Lucas J, Brun M Le, et al. Deletion 5q12 : Delineation of a Phenotype Including Mental Retardation and Ocular Defects. 2011; 725–731. https://doi.org/10.1002/ajmg.a.33758.
251. Schoenmakers N, Moran C, Peeters RP, Visser T, Gurnell M, Chatterjee K. Biochimica et Biophysica Acta Resistance to thyroid hormone mediated by defective thyroid hormone receptor alpha ☆. *BBA - General Subjects*. 2013;1830(7): 4004–4008. https://doi.org/10.1016/j.bbagen.2013.03.018.
252. Domené HM, Argente J, Wit M. Human Acid-Labile Subunit Deficiency : Clinical , Endocrine and. 2009;1425: 129–141. https://doi.org/10.1159/000232486.
253. McDonald-McGinn DM, Fahiminiya S, Revil T, Nowakowska BA, Suhl J, Bailey A, et al. Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.12Ds. *Journal of Medical Genetics*. 2013;50(2): 80–90. https://doi.org/10.1136/jmedgenet-2012-101320.
254. Hochstenbach R, Poot M, Nijman IJ, Renkens I, Duran KJ, Van'T Slot R, et al. Discovery of variants unmasked by hemizygous deletions. *European Journal of Human Genetics*. 2012;20(7): 748–753. https://doi.org/10.1038/ejhg.2011.263.
255. Poot M. Beware of hemizygous deletions that may unmask deleterious variants. *Molecular Syndromology*. 2012;3(2): 45–46. https://doi.org/10.1159/000339564.
256. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, et al. Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes. *New England Journal of Medicine*. 2008;359(16): 1685–1699. https://doi.org/10.1056/NEJMoa0805384.
257. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature Genetics*. 2008;40(12): 1466–1471. https://doi.org/10.1038/ng.279.
258. Christiansen J, Dyck JD, Elyas BG, Lilley M, Bamforth JS, Hicks M, et al. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circulation Research*. 2004;94(11): 1429–1435. https://doi.org/10.1161/01.RES.0000130528.72330.5c.

259. Sun J, Zhang L, Zhao H, Qiu X, Chen W, Wang D, et al. CHD1L Regulates Cell Cycle, Apoptosis, and Migration in Glioma. *Cellular and Molecular Neurobiology*. 2016;36(4): 565–576. https://doi.org/10.1007/s10571-015-0237-z.
260. De la Roche M, Worm J, Bienz M. The function of BCL9 in Wnt/β-catenin signaling and colorectal cancer cells. *BMC Cancer*. 2008;8: 1–13. https://doi.org/10.1186/1471-2407-8-199.
261. De Crescenzo A, Citro V, Freschi A, Sparago A, Palumbo O, Cubellis MV, et al. A splicing mutation of the HMGA2 gene is associated with Silver-Russell syndrome phenotype. *Journal of Human Genetics*. 2015;60(6): 287–293. https://doi.org/10.1038/jhg.2015.29.
262. Price SM, Stanhope R, Garrett C, Preece MA, Trembath RC. The spectrum of Silver-Russell syndrome: A clinical and molecular genetic study and new diagnostic criteria. *Journal of Medical Genetics*. 1999;36(11): 837–842.
263. Westbrook L, Manuvakhova M, Kern FG, Estes NR, Ramanathan HN, Thottassery J V. Cks1 regulates cdk1 expression: A novel role during mitotic entry in breast cancer cells. *Cancer Research*. 2007;67(23): 11393–11401. https://doi.org/10.1158/0008-5472.CAN-06-4173.
264. Ikeda S, Kitadate A, Abe F, Saitoh H, Michishita Y, Hatano Y, et al. Hypoxia-inducible microRNA-210 regulates the DIMT1-IRF4 oncogenic axis in multiple myeloma. *Cancer Science*. 2017;108(4): 641–652. https://doi.org/10.1111/cas.13183.
265. Liu G, Peng X, Cai Y, Cheng A, Zha L, Wang Z. DIMT1 overexpression correlates with progression and prognosis in gastric carcinoma. *Human Pathology*. 2017;70: 35–42. https://doi.org/10.1016/j.humpath.2017.02.034.
266. Cox DM, Butler MG. The 15q11 . 2 BP1 – BP2 Microdeletion Syndrome : A Review. 2015; 4068–4082. https://doi.org/10.3390/ijms16024068.
267. Stagi S, Lapi E, Pantaleo M, Traficante G, Giglio S, Seminara S, et al. A SOX3 (Xq26.3-27.3) duplication in a boy with growth hormone deficiency, ocular dyspraxia, and intellectual disability: A long-term follow-up and literature review. *Hormones*. 2014;13(4): 552–560. http://www.hormones.gr/pdf/Hormones_2014-552.pdf%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed12&NEWS=N&AN=2014960862
268. Stankiewicz P, Thiele H, Schlicker M, Cseke-Friedrich A, Bartel-Friedrich S, Yatsenko SA, et al. Duplication of Xq26.2-q27.1, including SOX3, in a mother and daughter with short stature and dyslalia. *American Journal of Medical Genetics*. 2005;138 A(1): 11–17. https://doi.org/10.1002/ajmg.a.30910.
269. Ayub S, Gadji M, Krabchi K, Côté S, Gekas J, Maranda B, et al. Three new cases of terminal deletion of the long arm of chromosome 7 and literature review to correlate genotype and phenotype manifestations. *American Journal of Medical Genetics, Part A*. 2016;170(4): 896–907. https://doi.org/10.1002/ajmg.a.37428.
270. Al Dhaibani MA, Allingham-Hawkins D, El-Hattab AW. De novo chromosome 7q36.1q36.2 triplication in a child with developmental delay, growth failure, distinctive facial features, and multiple congenital anomalies: A case report. *BMC Medical Genetics*. 2017;18(1): 1–5. https://doi.org/10.1186/s12881-017-0482-8.
271. Boskovski MT, Yuan S, Pedersen NB, Goth CK, Makova S, Clausen H, et al. The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature*. 2013;504(7480): 456–459. https://doi.org/10.1038/nature12723.
272. Katoh M, Katoh M. WNT Signaling Pathway and Stem Cell Signaling Network. 2007; 4042–4045. https://doi.org/10.1158/1078-0432.CCR-06-2316.
273. Morgan R, Ankrah R, Loadman PM. *Wnt Signaling as a Therapeutic Target in Cancer and Metastasis*. Introduction to Cancer Metastasis. Elsevier Inc.; 2017. https://doi.org/10.1016/B978-0-12-804003-4.00020-7.

274. Breit A, Miek L, Schredelseker J, Geibel M, Merrow M, Gudermann T. Insulin-like growth factor-1 acts as a zeitgeber on hypothalamic circadian clock gene expression via glycogen synthase kinase-3β signaling. *The Journal of biological chemistry*. 2018;293(44): 17278–17290. https://doi.org/10.1074/jbc.RA118.004429.
275. Karantanos T, Theodoropoulos G, Pektasides D, Gazouli M. Clock genes : Their role in colorectal cancer. 2014;20(8): 1986–1992. https://doi.org/10.3748/wjg.v20.i8.1986.
276. Gazal S, Weissbrod O, Hormozdiari F, Dey KK, Nasser J, Jagadeesh KA, et al. Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nature Genetics*. 2022;54(6): 827–836. https://doi.org/10.1038/s41588-022-01087-y.
277. Chen M, Huang JD, Hu L, Zheng BJ, Chen L, Tsang SL, et al. Transgenic CHD1L expression in mouse induces spontaneous tumors. *PLoS ONE*. 2009;4(8). https://doi.org/10.1371/journal.pone.0006727.
278. Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science*. 2008;319(5864): 819–821. https://doi.org/10.1126/science.1149250.
279. Gottschalk AJ, Timinszky G, Kong SE, Jin J, Cai Y, Swanson SK, et al. Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proceedings of the National Academy of Sciences*. 2009;106(33): 13770–13774. https://doi.org/10.1073/pnas.0906920106.
280. Brockschmidt A, Chung B, Weber S, Fischer DC, Kolatsi-Joannou M, Christ L, et al. CHD1L: A new candidate gene for congenital anomalies of the kidneys and urinary tract (CAKUT). *Nephrology Dialysis Transplantation*. 2012;27(6): 2355–2364. https://doi.org/10.1093/ndt/gfr649.
281. Cheng W, Su Y, Xu F. CHD1L: A novel oncogene. *Molecular Cancer*. 2013;12(1): 1–8. https://doi.org/10.1186/1476-4598-12-170.
282. Li Y, Chen L, Chan THM, Liu M, Kong KL, Qiu JL, et al. SPOCK1 is regulated by CHD1L and blocks apoptosis and promotes HCC cell invasiveness and metastasis in mice. *Gastroenterology*. 2013;144(1): 179–191. https://doi.org/10.1053/j.gastro.2012.09.042.
283. Chan THM, Chen L, Liu M, Hu L, Zheng BJ, Poon VKM, et al. Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. *Hepatology*. 2012;55(2): 491–505. https://doi.org/10.1002/hep.24709.
284. Benson KF, Chada K. Mini-mouse: Phenotypic characterization of a transgenic insertional mutant allelic to pygmy. *Genetical Research*. 1994;64(1): 27–33. https://doi.org/10.1017/S0016672300032511.
285. Zhou X, Benson KF, Ashar HR, Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature*. 1995;376(6543): 771–774. https://doi.org/10.1038/376771a0.
286. Fusco A, Fedele M. Roles of HMGA proteins in cancer. *Nature Reviews Cancer*. 2007;7(12): 899–910. https://doi.org/10.1038/nrc2271.
287. Zhang XY, Ding JX, Tao X, Hua KQ. FSH stimulates expression of the embryonic gene HMGA2 by downregulating let-7 in normal fmbrial epithelial cells of ovarian high-grade serous carcinomas. *Experimental and Therapeutic Medicine*. 2013;5(1): 350–354. https://doi.org/10.3892/etm.2012.794.
288. Fedele M, Pierantoni GM, Visone R, Fusco A. Critical role of the HMGA2 gene in pituitary adenomas. *Cell Cycle*. 2006;5(18): 2045–2048. https://doi.org/10.4161/cc.5.18.3211.
289. Ahlemann M, Zeidler R, Lang S, Mack B, Gires O, Mu M. HMGA2 overexpression in non-small cell lung cancer. *Molecular Carcinogenesis*. 2006;967(May): 957–967. https://doi.org/10.1002/mc.

290. Phase H, Methods S, Methods S, Methods S, Table S, Methods S, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. 2010; https://doi.org/10.1038/nature09410.
291. Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JRB, et al. A common variant of HMGA2 is associated with adult and childhood height in the general population. 2007;39(10): 1245–1250. https://doi.org/10.1038/ng2121.
292. Weedon MN, Lango H, Lindgren CM, Wallace C, David M, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. 2009;40(5): 575–583. https://doi.org/10.1038/ng.121.Genome-wide.
293. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. 2008;40(5): 584–591. https://doi.org/10.1038/ng.125.
294. Fusco I, Babu D, Mellone S, Barizzone N, Prodam F, Fanelli A, et al. Variations in the high-mobility group-A2 gene (HMGA2) are associated with idiopathic short stature. *Pediatric Research*. 2016;79(2): 258–261. https://doi.org/10.1038/pr.2015.225.
295. Meyer R, Begemann M, Hübner CT, Dey D, Kuechler A, Elgizouli M, et al. One test for all: whole exome sequencing significantly improves the diagnostic yield in growth retarded patients referred for molecular testing for Silver–Russell syndrome. *Orphanet Journal of Rare Diseases*. 2021;16(1): 1–10. https://doi.org/10.1186/s13023-021-01683-x.
296. Joynt S, Morillo V, Leng F. Binding the Mammalian High Mobility Group Protein AT-hook 2 to AT-Rich Deoxyoligonucleotides : Enthalpy-Entropy Compensation. *Biophysj*. 2009;96(10): 4144–4152. https://doi.org/10.1016/j.bpj.2009.02.015.
297. Gonorazky H, Liang M, Cummings B, Lek M, Micallef J, Hawkins C, et al. RNAseq analysis for the diagnosis of muscular dystrophy. *Annals of Clinical and Translational Neurology*. 2016;3(1): 55–60. https://doi.org/https://doi.org/10.1002/acn3.267.
298. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Human Genetics*. 2017;136(9): 1093–1111. https://doi.org/10.1007/s00439-017-1809-4.
299. Binder G. Short stature due to SHOX deficiency: genotype, phenotype, and therapy. *Hormone research in paediatrics*. 2011;75(2): 81–89. https://doi.org/10.1159/000324105.
300. Croft L, Croft L, Schandorff S, Clark F, Burrage K, Arctander P, et al. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nature genetics*. 2000;24(4): 340–341. https://doi.org/10.1038/74153.
301. Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Human mutation*. 2007;28(2): 150–158. https://doi.org/10.1002/humu.20400.
302. Hare MP, Palumbi SR. High intron sequence conservation across three mammalian orders suggests functional constraints. *Molecular Biology and Evolution*. 2003;20(6): 969–978. https://doi.org/10.1093/molbev/msg111.
303. Kelly S, Georgomanolis T, Zirkel A, Diermeier S, O'Reilly D, Murphy S, et al. Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Research*. 2015;43(9): 4721–4732. https://doi.org/10.1093/nar/gkv386.
304. Juneau K, Miranda M, Hillenmeyer ME, Nislow C, Davis RW. Introns Regulate RNA and Protein Abundance in Yeast. *Genetics*. 2006;174(1): 511–518. https://doi.org/10.1534/genetics.106.058560.
305. Shabalina SA, Ogurtsov AY, Spiridonov AN, Novichkov PS, Spiridonov NA, Koonin E V. Distinct Patterns of Expression and Evolution of Intronless and Intron-Containing Mammalian Genes. *Molecular Biology and Evolution*. 2010;27(8): 1745–1749. https://doi.org/10.1093/molbev/msq086.

306. Brinster RL, Allen JM, Behringer RR, Gelinas RE, Palmiter RD. Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*. 1988;85(3): 836–840. https://doi.org/10.1073/pnas.85.3.836.
307. McKenzie RW, Brennan MD. The two small introns of the Drosophila affinidisjuncta Adh gene are required for normal transcription. *Nucleic acids research*. 1996;24(18): 3635–3642. https://doi.org/10.1093/nar/24.18.3635.
308. Stadhouders R, van den Heuvel A, Kolovos P, Jorna R, Leslie K, Grosveld F, et al. Transcription regulation by distal enhancers. *Transcription*. 2012;3(4): 181–186. https://doi.org/10.4161/trns.20720.
309. Yasuda H, Oh C do, Chen D, de Crombrugghe B, Kim JH. A Novel Regulatory Mechanism of Type II Collagen Expression via a SOX9-dependent Enhancer in Intron 6 *. *Journal of Biological Chemistry*. 2017;292(2): 528–538. https://doi.org/10.1074/jbc.M116.758425.
310. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. *Nature Reviews Genetics*. 2016;17(7): 407–421. https://doi.org/10.1038/nrg.2016.46.
311. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes and Development*. 2012;26(10): 1098–1109. https://doi.org/10.1101/gad.190173.112.
312. Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & development*. 1998;12(13): 1998–2012. https://doi.org/10.1101/gad.12.13.1998.
313. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular Cell Biology*. 2002;3(3): 195–205. https://doi.org/10.1038/nrm760.
314. Dhir A, Buratti E. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *The FEBS Journal*. 2010;277(4): 841–855. https://doi.org/https://doi.org/10.1111/j.1742-4658.2009.07520.x.
315. Romano M, Buratti E, Baralle D. Role of Pseudoexons and Pseudointrons in Human Cancer. Ghigna C (ed.) *International Journal of Cell Biology*. 2013;2013: 810572. https://doi.org/10.1155/2013/810572.
316. Bovolenta M, Neri M, Fini S, Fabris M, Trabanelli C, Venturoli A, et al. A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. *BMC Genomics*. 2008;9(1): 572. https://doi.org/10.1186/1471-2164-9-572.
317. Miguel-Escalada I, Maestro MÁ, Balboa D, Elek A, Bernal A, Bernardo E, et al. Pancreas agenesis mutations disrupt a lead enhancer controlling a developmental enhancer cluster. *Developmental cell*. 2022;57(16): 1922-1936.e9. https://doi.org/10.1016/j.devcel.2022.07.014.
318. Lettice LA, Williamson I, Wiltshire JH, Peluso S, Devenney PS, Hill AE, et al. Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. *Developmental Cell*. 2012;22(2): 459–467. https://doi.org/10.1016/j.devcel.2011.12.010.
319. Sun X, Song L, Yang W, Zhang L, Liu M, Li X, et al. Nanopore Sequencing and Its Clinical Applications. *Methods in molecular biology (Clifton, N.J.)*. 2020;2204: 13–32. https://doi.org/10.1007/978-1-0716-0904-0_2.
320. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome biology*. 2018;19(1): 90. https://doi.org/10.1186/s13059-018-1462-9.
321. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional

genomics. *Development, growth & differentiation*. 2019;61(5): 316–326. https://doi.org/10.1111/dgd.12608.
322. Lin B, Hui J, Mao H. Nanopore Technology and Its Applications in Gene Sequencing. *Biosensors*. 2021;11(7). https://doi.org/10.3390/bios11070214.
323. Frey-Law LA, Bohr, Nicole L, Sluka KA, Herr K, Clark CR, Noiseux NO, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)*. 2020;369(6509): 1318–1330. https://doi.org/10.1126/science.aaz1776.
324. Collins RL, Glessner JT, Porcu E, Lepamets M, Brandon R, Lauricella C, et al. A cross-disorder dosage sensitivity map of the human genome. *Cell*. 2022;185(16): 3041-3055.e25. https://doi.org/10.1016/j.cell.2022.06.036.

Appendix 1: Composition of buffers and reagents utilised from commercially available kits

GE Healthcare Nucleon™ BACC2 Genomic DNA Extraction Kit (catalogue number 10637265, Fisher Scientific, UK):

Reagent A (lysis agent): 10 mM Tris-HC1; 320 M sucrose; 5 mM MgCl2; 1% (v/v) Triton X-100; pH 8.0 (pH adjusted using 40%(w/v) NaOH).

Reagent B: (precise composition is not disclosed by the company). 5 M Sodium perchlorate.

Nucleon resin: includes potassium hydroxide (precise composition is not disclosed by the company).

QIAprep Spin Miniprep Kit (catalogue number 27106, Qiagen, Crawley, UK):

Buffer P1 (resuspension buffer): 50 mM Tris·Cl; 10 mM EDTA; 100 µg/ml RNase A

Buffer P2 (lysis buffer): 200 mM NaOH; 1% SDS (w/v)

Buffer N3 (neutralisation buffer): 4.2 M Guanidinium hydrochloride, 0.9M potassium acetate, pH 4.8 Buffer PB (wash buffer): 5 M Guanidinium hydrochloride; 30% isopropanol Buffer PE (wash buffer): 10 mM Tris-HCl pH 7.5; 80% ethanol Buffer EB (elution buffer): 10 mM Tris-Cl, pH 8.5

Other components of the kit: QIAprep Spin Columns, Collection Tubes (2 ml)

Qiagen HI SPEED Plasmid Midi Kit (catalogue number 12643, Qiagen, Crawley, UK):

Buffer P1 (resuspension buffer): 50 mM Tris·Cl; 10 mM EDTA; 100 µg/ml RNase A

Buffer P2 (lysis buffer): 200 mM NaOH; 1% SDS (w/v)

Buffer P3 (neutralisation buffer): 3.0 M potassium acetate, pH 5.5

Buffer QBT (equilibration buffer): 750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol (v/v); 0.15% Triton® X-100 (v/v)

Buffer QC (wash buffer): 1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol (v/v) Buffer QF (elution buffer): 1.25 M NaCl; 50 mM Tris·Cl, pH 8.5; 15% isopropanol (v/v)

Other components of the kit: QIAfilter Cartridge; QIAGEN-tip Qiagen

QIAquick® Gel Extraction Kit (catalogue number 28704, Qiagen, Crawley, UK):

Buffer QG (solubilization and binding buffer, with pH indicator): 5.5 M guanidine thiocyanate, 20 M Tris Hcl pH 6.6.

Buffer PE (wash buffer): 10 mM Tris-HCl pH 7.5; 80% ethanol

Other components of the kit: QIAquick spin column; Collection tube (2 ml)

QuikChange II XL Site-Directed Mutagenesis Kit Instruction (catalogue number 200522, Agilent, UK):

PfuUltra High-Fidelity DNA polymerase (2.5 U/ µl)

10× reaction buffer

DpnI restriction enzyme (10 U/µl)

Oligonucleotide control primer 1, 5´ CCA TGA TTA CGC CAA GCG CGC AAT TAA CCC TCA C 3´

Oligonucleotide control primer 2, 5´ GTG AGG GTT AAT TGC GCG CTT GGC GTA ATC ATG G 3´

pWhitescript 4.5-kb control plasmid (5 ng/ µl)

dNTP mix

XL1-Blue supercompetent cells

pUC18 control plasmid (0.1 ng/µl in TE buffer)

NEBuffer r1.1 (catalogue number B7030S, New England Biolabs, USA)

1X Buffer Components

10 mM Bis-Tris-Propane-HCl

10 mM MgCl2

100 µg/ml Recombinant Albumin

pH 7.0@25°C

NEBuffer r2.1 (catalogue number B7030S, New England Biolabs, USA)

1X Buffer Components

50 mM NaCl

10 mM Tris-HCl

10 mM MgCl2

100 µg/ml Recombinant Albumin

pH 7.9@25°C

RNeasy Plus Mini kit (catalogue number 74134, Qiagen, Crawley, UK):

Buffer RLT (lysis buffer): contains a guanidine salt (precise composition is not disclosed by the company).

Buffer RW1 (stringent wash buffer): contains a guanidine salt, as well as ethanol (precise composition is not disclosed by the company).

Buffer RPE (mild wash buffer): Ethanol containing (precise composition is not disclosed by the company.

Appendix 2. Primers to confirm variants identified in genomic DNA

Primers	Sequence
GHR intron 6 (Forward)	5'- GGTTTCAGAAGGATGTTTTG
GHR intron 6 (Reverse)	5'- ATTGCTTTTGTGCCATTAAG
GHR intron 2 (Forward)	5'-AACCAGTGCCCATAGGTG
GHR intron 2 (Reverse)	5'-CTGGAACTAGAGGCAGAAGG
GHR exon 10 (Forward)	5'-AGGTGAACACAATCTTAGCC
GHR exon 10 (Reverse)	5'-TCACCTGGGCATAAAAGTC
CUL7 exon 4 (Forward)	5'-GAGGAGTTGAGTGGGGAG
CUL7 exon 4 (Reverse)	5'-ACACACAGGATGCCTTCTC
CUL7 exon 13 (Forward)	5'-ACCCTACCTGACCGTGAC
CUL7 exon 13 (Reverse)	5'-CAGAGAAAGCTGTCACCTCC
CUL7 exon 17 (Forward)	5'-AGGCGTAAACCACCCTACTCCTTGTATCAT TCCGTCC
CUL7 exon 19 (Forward)	5'-CATCCATGAACATCATTTTGGTGGTGGAGA AGCAGG
CUL7 exon 19 (Reverse)	5'-GATGATCTGCCTGCCTCGAATGGGAATG GAAGGG
CUL7 exon 17 (Reverse)	5'-AAGGAGCTCCAGTCGTGGGCAGGAGCA TGAAGAC
CCDC8 exon 1 (Forward)	5'-ACGTACGACAGCAGCAACTCTCGAGAAAAGCCACG
CCDC8 exon 1 (Reverse)	5'- CTCTGATTACCTGCAGCCCCCAGGTTGTC CTCAAAC
OBSL1 splice (Forward)	5'-GATGTGCAGGTATACGACG
OBSL1 splice (Reverse)	5'-CAGAGAATTGGGATATGATGC
HMGA2 exon 2 (Forward)	5'-TCAGGGTCAATTTCTTTCAG
HMGA2 exon 2 (Reverse)	5'-GGCTCCTGTAGTCAGTCATTG
CHD1L exon 18 (Forward)	5'-GATATTGTTTGTGAACAGCAG
CHD1L exon 18 (Reverse)	5'-ACACTGGGACCACCACC
CHD1L exon 19 (Forward)	5'-TGTGCACTGGACTTCACC
CHD1L exon 19 (Reverse)	5'-GCTGTGTAACCCTCACGATAC
CHD1L exon 21 (Forward)	5'-CCCAAGTGACCCATCAAG
CHD1L exon 21 (Reverse)	5'-GTCGTATCACATGCCTAGAAAC

Appendix 3. Antibodies used and applications

Name of antibody	Manufacturer, catalogue no	Species raised in, monoclonal or polyclonal	Application	Dilution used
Primary antibodies				
Beta actin	Protein Tech, 66009-1-1g	Mouse, monoclonal	Western blot	1 in 5000
p-STAT5 (Y694)	Cell Signalling Technology, D47E7	Rabbit, monoclonal	Western blot	1 in 750
STAT5B	Boster Biological Technology, PA1841	Rabbit, polyclonal	Western blot	1 in 1000
HMGA2	Cell Signalling Technology, D1A7	Rabbit, monoclonal	Western blot, Immunocytochemistry	1 in 1000 for WB, 1 in 100 for ICC
CHD1L	Abcam, ab197019	Rabbit, monoclonal	Western blot	1 in 1000
mCherry	Abcam, ab125096	Mouse, monoclonal	Western blot	1 in 1000
IGF-1	Abcam, ab176523	Mouse, monoclonal	Western blot	1 in 1000
GHBP	BioVision, 6660-100	Rabbit, polyclonal	Western blot	1 in 1000
GHR N terminal	Origene, TA334654	Rabbit, polyclonal	Western blot	1 in 1000
HA	Sigma, H6908	Rabbit, polyclonal	Immunocytochemistry, western blot	1 in 1000 for WB, 1 in 100 for ICC
FLAG	Sigma, F1804	Mouse, monoclonal	Immunocytochemistry, western blot	1 in 1000 for WB, 1 in 100 for ICC

His	Sigma, H1029	Mouse, monoclonal	Western blot	1 in 1000
6 x His	Abcam, ab18184	Mouse, monoclonal	Western blot	1 in 1000
Strep tag II	Abcam, ab76949	Rabbit, polyclonal	Western blot	1 in 1000
Secondary antibodies				
Donkey anti-Rabbit IRDye 800CW	Li-Cor, 926-32213	Donkey	Western blotting	1 in 10000
Donkey anti-Mouse IRDye 680CW	Li-Cor, 926-68072	Donkey	Western blotting	1 in 10000
Goat Anti-Rabbit IgG (Alexa Fluor®488)	Abcam, ab150081	Goat	Immunocytochemistry	1 in 500
Goat anti-Mouse IgG (Alexa Fluor®647)	Invitrogen, A-21235	Goat	Immunocytochemistry	1 in 500

WB, western blotting; ICC, immunocytochemistry

Appendix 4. Primer sequences for *In vitro* splicing assay, site-directed mutagenesis, Gibson assembly and cDNA synthesis

Primers	**Sequence**
Splicing assays	
Intron 6 *GHR XbaI* (Forward)	5'- **AGCTATATCTAGA**TTGTCTCTTAGAGGTTTCAG
Intron 6 *GHR XbaI* (Reverse)	5'- **AGCTATATCTAGA**CAGTGCCTTCTGCATAGTAG
Intron 2 *GHR XbaI* (Forward)	5'- **AGCTATATCTAGA**AACCAGTGCCCATAGGTG
Intron 2 *GHR XbaI* (Reverse)	5'- **AGCTATATCTAGA**CTGGAACTAGAGGCAGAAGG
Exon 9 *GHR XbaI* (Forward)	5'- **AGCTATATCTAGA**CTTTCAAATACTAAATCTGGC
Exon 9 *GHR XbaI* (Reverse)	5'- **AGCTATATCTAGA**GTTCAGAGATTAATTATGGTTACAATAAG
ET PRIM 06 (Forward)	5'- GCGAAGTGGAGGATCCACAAG
ET PRIM 07 (Reverse)	5'- ACCCGGATCCAGTTGTGCCA
ET PRIM 02 (Forward)	5'- GAGGGATCCGCTTCCTGGCCC
ET PRIM 03 (Reverse)	5'- CTCCCGGGCCACCTCCAGTGCC
Gibson assembly	
pcDNA1 *hGHR* FL FWD (1)	5'- ATTTTGTCATTAGAGGAAAGATGGACCCTATATTGACAACA
pcDNA1 *hGHR* FL REV (1)	5'- GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
pcDNA1 *hGHR* FL FWD (2)	5'- CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC
pcDNA1 *hGHR* FL REV (2)	5'- GGCTGTGGCTGTGGCTGTGGCATTTTCCATTTAGTTTCATTTACTTC

6Ω pseudoexon for cloning FWD (3)	5’- ATGAAACTAAATGGAAAATGCCACAGCCACAGCCACAGCC
6Ω pseudoexon for cloning REV (3)	5’- GTTGTCAATATAGGGTCCATCTTTCCTCTAATGACAAAATTGGCATCTTTTTC
pcDNA1 *hGHR* P7 FWD (1)	5’- TCTTTTCGAAACAGCAAAGGTCCAGTTCCAAAGATTAAAGGA
pcDNA1 *hGHR* P7 REV (1)	5’- GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
pcDNA1 *hGHR* P7 FWD (2)	5’- CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC
pcDNA1 *hGHR* P7 REV (2)	5’- TTTAATCTTTGGAACTGGACCTTTGCTGTTTCGAAAAGAGA
pcDNA1 *hGHR* P8 FWD (1)	5’- CTCTTTTCGAAACAGCAAAGTTCCAGTTCCAAAGATTAAAGG
pcDNA1 *hGHR* P8 REV (1)	5’- GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGG
pcDNA1 *hGHR* P8 FWD (2)	5’- CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTC
pcDNA1 *hGHR* P8 REV (2)	5’- TTTAATCTTTGGAACTGGAACTTTGCTGTTTCGAAAAGAGA
RT-PCR	
GHR cDNA pseudo F1 (Forward)	5’- GGAAAATGCCACAGCC
GHR cDNA Exon 4F (Forward)	5’- TCTAAGGAGCCTAAATTCACC
GHR cDNA Exon 8R (Reverse)	5’-TAGCATCACTGTTAGCC
Site-directed mutagenesis	
HMGA2 52C>T SDM (Forward)	5’-TCAGCCCAGGGATAACCTGCCGCCC
HMGA2 52C>T SDM (Reverse)	5’-GGGCGGCAGGTTATCCCTGGGCTGA
HMGA2 144delC SDM (Forward)	5’-CCTCTCCTAAGAGACCAGGGGAAGACCCAAAG
HMGA2 144delC SDM (Reverse)	5’-CTTTGGGTCTTCCCCTGGTCTCTTAGGAGAGG
HMGA2 145delA SDM (Forward)	5’-TCTCCTAAGAGACCCGGGGAAGACCCAAAG

HMGA2 145delA SDM (Reverse)	5’-CTTTGGGTCTTCCCCGGGTCTCTTAGGAGA
CHD1L Exon 18 SDM (Forward)	5’-TGGTAATTGTTGGATTCGCACCAGGCCATCCTTTA
CHD1L Exon 18 SDM (Reverse)	5’-TAAAGGATGGCCTGGTGCGAATCCAACAATTACCA
CHD1L Exon 19 SDM (Forward)	5’-ACCTCTGCCCCAGCGGCCAGAGTCATC
CHD1L Exon 19 SDM (Reverse)	5’-GATCACTCTGGCCGCTGGGGCAGAGGT
CHD1L Exon 21 SDM (Forward)	5’-TCTTTTTTGCTGCTAAAAATACCTTCTTCAGGCCCTCTTCTAG
CHD1L Exon 21 SDM (Reverse)	5’-CTAGAAGAGGGCCTGAAGAAGGTATTTTTAGCAGCAAAAAAGA
GHR vector sequencing	
GHR vector F	5’-GGTTACTCATCATCATATTCGG
GHR vector F1	5’-ATACTCAATGCGACTGCC
GHR vector F2	5’-GCCATACCACATTTGTAGAGG
GHR vector F3	5’-ATTGTAGAGGCTTCCAGAGG
GHR vector F4	5’-GCCACCTCTGACTTGAGC
GHR vector F5	5’-GTGGGTTACATCGAACTGG
GHR vector R1	5’-ACCTCCTCTAATTTTCCTTCC
GHR vector R2	5’-CTTGAGGAGATCTGGATCG

GHR XbaI primers amplified target sequence of interest for the splicing assay and incorporated *XbaI* restriction enzyme target sites (highlighted in bold); ET primers amplified regions of interest in the MoBiTec-Exontrap cloning vector pET01 to confirm splicing products; pcDNA1 and 6Ω pseudoexon primers amplified the 3 target regions for Gibson assembly of 6Ω pseudoexon *GHR* vector; *GHR* cDNA primers amplified regions of interest in fibroblast cDNA from kindred 2 and a healthy control.

Appendix 5. Oral presentations of this research

03/11/2022 **'Characterisation of the first heterozygous missense *HMGA2* variant helps delineate the crucial functional roles of a novel growth gene'** Emily Cottrell, Avinaash V. Maharaj, Barbara Triggs-Raine, Thatchawan Thanasupawat, Jack Williams, Masanobu Fujimoto, Hermine A. van Duyvenvoorde, Christiaan de Bruin, Sjoerd Joustra, Sarina Kant, Danielle van der Kaay, Maria Inmaculada Castilla de Cortázar Larrea, Ahmed Massoud, Louise A Metherell, Vivian Hwa, Sabine Hombach-Klonisch, Thomas Klonisch, Helen L. Storr. Presented at the 2022 British Society for Paediatric Endocrinology and Diabetes Conference in Belfast, UK.

25/11/2021 **'*Growth Hormone Receptor* 6Ω pseudoexon activation: a novel cause of severe Growth Hormone Insensitivity'** Emily Cottrell, Avinaash Maharaj, Jack Williams, Sumana Chatterjee, Grazia Cirillo, Emanuele Miraglia del Giudice, Adalgisa Festa, Stefania Palumbo, Donatella Capalbo, Mariacarolina Salerno, Claudio Pignata, Martin O. Savage, Katharina Schilbach, Martin Bidlingmaier, Vivian Hwa, Louise A. Metherell, Anna Grandone, Helen L. Storr. Presented at the 2021 British Society for Paediatric Endocrinology and Diabetes virtual conference.

22/06/2021 '**A rare heterozygous *IGF1* variant causing postnatal growth failure and offering novel insights into IGF-1 physiology'** Emily Cottrell, Sumana Chatterjee, Vivian Hwa, Helen L. Storr. Presented at 2021 British Endocrine Society National Clinical Cases virtual meeting.

27/11/2019 '**A rare but very important cause of growth failure'** Emily Cottrell, Tasneem Ladha, Hanna Borysewicz-Sańczyk, Beata Sawicka, Artur Bossowski, Helen Storr. Presented at the 2019 British Society for Paediatric Endocrinology and Diabetes Conference in Cardiff, UK (CME day).

28/11/2019 '**Novel genetic defects in a cohort of Silver-Russell Syndrome (SRS) and SRS-like patients'** Emily Cottrell, Miho Ishida, Gudrun Moore, Helen L. Storr. Presented at 2019 British Society for Paediatric Endocrinology and Diabetes Conference in Cardiff, UK

26/10/2019 **'An update on new genetic causes of poor growth'** Emily Cottrell, Helen L. Storr. Presented at the 2019 Child Growth Foundation Annual Convention, Stratford, UK.

28/09/2018 '**Patients with short stature and GH/IGF-1 Insensitivity Harbour Copy number variants causing a Silver-Russell-like phenotype'** Emily Cottrell, Sumana Chatterjee, Gudrun Moore, Miho Ishida, James Greening, Neil Wright, Artur Bossowski, Asma Deeb, Iman Al Basiri, Stephen Rose, Avril Mason, Joowook Ahn, Susan Bint, Martin O. Savage, Louise A. Metherell, Helen L. Storr. Presented at the 2018 European Society for Paediatric Endocrinology 57th annual meeting in Athens, Greece.

Appendix 6. Awards arising from this work

June 2021	**2nd best oral case presentation** National award at the BES National Clinical Cases 2021 meeting for my presentation 'A rare heterozygous *IGF1* variant causing postnatal growth failure and offering novel insights into IGF-1 physiology.'
February 2021	**Barts Charity Clinical Research Training Fellowship award** Awarded for completion of PhD research. Total value awarded £68,629.
November 2018	**Best Clinical Case** National award at the British Endocrine Society annual meeting for novel 6Ω *GHR* pseudoexon case.
August 2018	**The 2018 ESPE Research Fellowship** Internationally competitive funding for 2-year period of research in Paediatric Endocrinology. Total value awarded €140,000.
May 2018	**Sandoz Educational Award, Annual Retreat of Trainees in Endocrinology** National award for presentation on preliminary research findings.