Aberrant downstream mechanisms following depletion of KMT2C and KMT2D in Pancreatic Ductal Adenocarcinoma

Joshua Benjamin Newton Dawkins

Thesis submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy (PhD)

at Barts and the London School of Medicine and Dentistry,

Queen Mary University of London

Barts Cancer Institute,

Queen Mary University of London,

Charterhouse Square,

London

EC1M 6BQ UK

2016

Statement of Originality

I, Joshua Benjamin Newton Dawkins, 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:

Details of collaboration and publications:

- RNA sequencing and was carried out by Source Bioscience, William James House, Cowley Road, Cambridge UK.
- Bioinformatic analysis of the clinically annotated patient gene expression datasets, and the RNA sequencing data were carried out by Dr Jun Wang at the Barts Cancer Institute, Queen Mary University of London, London UK.
- Generation and maintenance of the compound *KC/KPC;Kmt2d* genetically engineered mice was carried out alongside Dr Eleni Maniati at the Barts Cancer Institute, Queen Mary University of London, London UK.
- Some of the KPC mouse data included was kindly provided by Dr Juliana Candido at the Barts Cancer Institute, Queen Mary University of London, London UK.
- Much of the work described has been published in *Cancer Research* (Appendix I).

Abstract

Genomic sequencing of pancreatic ductal adenocarcinoma (PDAC) tumours has highlighted the existence of wide genetic diversity alongside frequent mutations in *KRAS*, *TP53* and *SMAD4*. Within this heterogeneity many components of the epigenetic machinery are mutated, including the histone H3 lysine 4 methyltransferases KMT2C and KMT2D, which are frequently subject to mutation and can identify patients with a more favorable prognosis. In this thesis low expression of *KMT2C* and *KMT2D* were shown to also define better outcome groups, with median survivals of 15.9 vs 9.2 months (p = 0.029), and 19.9 vs 11.8 months (p = 0.001) respectively. Experiments across eight human pancreatic cell lines following their depletion suggest that this improved outcome may be due to attenuated cell proliferation, with decreased progression of cells from G_0/G_1 observed upon KMT2D loss. Whole transcriptome analysis of PDAC cell lines following *KMT2C* or *KMT2D* knockdown identified 31 and 124 differentially expressed genes respectively, with 19 common to both. Gene set enrichment analysis revealed a significant downregulation of genes relating to cell-cycle pathways, confirmed by interrogation of the International Cancer Genome Consortium and The Cancer Genome Atlas PDAC data series. Furthermore, these experiments highlighted a potential role for NCAPD3, a subunit of the condensin II complex, as a PDAC outcome predictor across four patient gene expression series. Alongside this, Kmt2d depletion in cells derived from murine models of pancreatic cancer led to an increase in their response to the antimetabolites 5-fluorouracil and gemcitabine. Taken together, the studies herein suggest that lower levels of this methyltransferase may mediate the sensitivity of PDAC patients to particular treatments. Altogether, these data suggest a potential therapeutic benefit in targeting these methyltransferases within PDAC, especially in those patients that demonstrate higher *KTM2C/D* expression.

Acknowledgements

During my time working towards this thesis there have been a great number of people that without whom, I would never have had such a successful and fulfilling experience. First and foremost, I would like to thank both of my supervisors, Professor Jude Fitzgibbon and Dr Richard Grose, who have both been excellent mentors and without their guidance and generosity I have no doubt that I would have been lost after two years. In addition, I would like to acknowledge Professor Frances Balkwill for the role that she played in a period of transition.

Another exceptional scientist, mentor and friend crucial for helping to develop both this thesis, and myself as a scientist, is Dr Eleni Maniati, where from the very first day until the last she has given limitless advice and support, and for which I will be always thankful.

I am also thankful for the collaborations with many other excellent scientists who have enriched both the work and my experience. I have thoroughly enjoyed working and playing football alongside Dr Jun Wang, whose knowledge and skills in bioinformatics have been key in helping to direct my research. I am also greatly thankful for the input of Dr James Heward, where after each discussion I have always learnt something new.

In addition to those mentioned here, countless other brilliant people have made my time truly enjoyable. Specifically, I would like to give a special thanks to Mark Brown, Dr Lily Keane and Dr Laura King for helping to create the many fond memories both in and out of the lab. I am also grateful to Jonas Mackerodt who has helped me to let off steam by joining me in range of sporting activities. Many thanks are also due to Laura Lecker, who through her enthusiasm, and our shared interests and laughter, has always helped to keep me in high spirits.

I am also very grateful to Becky Parslow for always being someone who was there to offer unconditional support, understanding and kindness; especially at the times when I needed a reminder to come home for food and rest.

In final, I would like to express a great appreciation for my funding body Cancer Research UK who without, this work would have never been possible or indeed reached its full potential.

Table of Contents

Index of Tables

Index of Figures

Abbreviations

°C	Degrees Celsius
% v/v	Percent volume/volume
% w/v	Percent weight/volume
5-FU	5-fluorouracil
A	Adenine
aa	Amino acids
ABCB1	ATP-binding cassette sub-family B member 1
adj.	Adjusted
ALL	Acute lymphoblastic leukaemia
ALR	ALL-1 related
AML	Acute myeloid leukaemia
ASH2	Absent, small or homeotic disc 2
ATCC	American Type Culture Collection
AWS	Associated with SET
BAH	Bromo-adjacent homology
BCA	Bicinchoninic Acid
BCI	Barts Cancer Institute
bp	Base pair
BRD	Bromodomain
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
C	Cytosine
CDK	Cyclin-dependent kinase
CDKI	CDK inhibitor
cDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	CF transmembrane conductance regulator
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP combined with DNA sequencing
Chr	Chromosome
cM	Centimorgan
CO2	Carbon dioxide
COSMIC	Catalogue of Somatic Mutations in Cancer
CpG	Cytosine-phosphate-Guanine
CPM	Count per million
cqn	Conditional quantile normalisation
DAG	Diacylglycerol

DE	Differentially expressed
DLBCL	Diffuse large B-cell lymphoma
Dm	Drosophila melanogaster
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco's Phosphate Buffered Saline
DPY30	Dumpy-30
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FAD	Flavin adenine dinucleotide
FBS	Foetal bovine serum
FDR	False discovery rate
FL	Follicular lymphomas
g	Gravity
G	Guanine
GDP	Guanosine diphosphate
GEMM	Genetically engineered mouse model
GEO	Gene Expression Omnibus
GEP	Gene expression profile
GLM	Generalised linear model
GO	Gene ontology
GOF	Gain-of-function
GPCR	G protein-coupled receptor
GSEA	Gene Set Enrichment Analysis
GTP	Guanosine triphosphate
H3K27ac	Acetylated H3K27
H3K27me3	Trimethylated H3K27
H3K4me1	Monomethylated H3K4
H3K4me2	Dimethylated H3K4
H3K4me3	Trimethylated H3K4
H4K20me1	Monomethylated H4K20
H3K9	Lysine 9 of Histone 3
H4K20	Lysine 20 of Histone 4
HCC	Hepatocellular carcinoma
HCF1	Host cell factor 1
HGNC	HUGO Gene Nomenclature Committee
HUGE	Human Unidentified Gene-Encoded

HUGO	Human Genome Organisation
ICGC	International Cancer Genome Consortium
IMDM	Iscove's Modified Dulbecco's Medium
IP3	Inositol 1,4,5-triphosphate
IPMN	Intraductal papillary mucinous neoplasia
KC	LSL-KrasG12D/+; Pdx-1-Cre
kDa	Kilodalton
KDM	Lysine demethylase
KM	Kaplan-Meier
KMT	Lysine-specific methyltransferase
KP	LSL-KrasG12D/+; LSL-Trp53R172H/+
KPC	LSL-KrasG12D/+; LSL-Trp53R172H/+; Pdx-1-Cre
LOF	Loss-of-function
LSL	LoxP-STOP-LoxP
M	Molar
M-phase	Mitotic phase
MCN	Mucinous cystic neoplasia
MGI	Mouse Genome Informatics
MGNC	Mouse Genomic Nomenclature Committee
MLL	Mixed-lineage leukaemia
MMP	Matrix metalloproteinase
MOPS	3-(N-morpholino)propanesulfonic acid
NES	Normalised enrichment score
NHL	Non-Hodgkin lymphoma
OD	Optical density
OS	Overall survival
PanIN	Pancreatic intraepithelial neoplasia
PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PDX	Patient derived tumour xenograft
PHD	Plant homeodomain
PI	Propidium Iodide
PI(4,5)P2	Phosphatidylinositol 4,5-bisphosphate
PI(5)P	Phosphatidylinositol-5-phosphate
PLC	Phospholipase C
PVDF	Polyvinylidene fluoride
qPT-PCR	Quantitative real-time PCRs
RBBP5	Retinoblastoma binding protein 5
RIN	RNA integrity number

RIPA	Radioimmunoprecipitation Assay
RNA pol II	RNA polymerase II
RNA-seq	RNA sequencing
RNAi	RNA interference
RPKM	Reads per kilobase of transcript per million mapped reads
RPMI	Roswell Park Memorial Institute
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RSEM	RNA-seq by Expectation-Maximization
RT	Room temperature
S-phase	Synthesis phase
SAM	S-adenosyl-L-methionine
SCID	Severe combined immunodeficiency
SET	Su(var)3-9, E(z) (enhancer of zeste) and trithorax
shRNA	Short/small hairpin RNA
siRNA	Small interfering RNA
SMART	Simple Modular Architecture Research Tool
SNP	Single nucleotide polymorphism
T	Thymine
TA	Tris-Acetate
TBS	Tris-Buffered Saline
TBST	TBS-Tween
TCGA	The Cancer Genome Atlas
Trr	Trithorax-related
Trx	Trithorax
TSS	Transcription start site
U	Uracil
WDR5	WD repeat protein 5
WT	Wild-type
YAP	Yes-associated protein
ZF-CxxC	CXXC zinc finger

Chapter 1 Introduction

1.1 Pancreatic ductal adenocarcinoma (PDAC)

1.1.1 Prevalence and epidemiology

Pancreatic ductal adenocarcinoma (PDAC) makes up the majority of all pancreatic malignancies and is notoriously associated with a particularly poor overall survival (1). Typically patients present with invasion and metastases at diagnosis, both of which limit patient suitability for curative surgical resection, thereby leaving chemotherapy as the next best option (2).

In the UK between 2000 and 2010, the incidence of PDAC increased by 3% to 14.5 cases per 100000 person-years (3). In this way, pancreatic cancer has also been predicted to become the second largest leading cause of cancer-related deaths in the USA by 2030 (4). Studies looking at patient demographics have shown PDAC to be a disease predominantly associated with age, where 10% of patients (5), or fewer (3.7%) (3), develop the disease before 50, giving rise to a median age of 68 at diagnosis (2).

The incidence of PDAC is three to four times higher in Northern Europe and America than in tropical countries (3), thus it is likely that genetic, sociodemographic and environmental risk factors are key contributors to the disease aetiology. Unlike with other cancers, only a small number of genetic features are known to predispose to PDAC, where familial clustering identifies only 10% of patients as having a first-degree relative also afflicted (6). The genetic features associated with familial PDAC include germline mutations in *CDKN2A* (7,8), *STK11* (9,10), *PRSS1* (11,12), *BRCA1* (13,14), *BRCA2* (15-17) and *CFTR* (18-21). The penetrance of these mutations is often low and, as suggested by Hezel *et al.*, might be a feature of the complex disease aetiology and development, with a more influential role in the progression of precursor lesions, rather than disease initiation (22). Also supporting a potential underlying genetic component to PDAC development are the increased incidence rates in several different races (*e.g.* Ashkenazi Jews, Maoris, and Cajuns) (5).

In addition to this genetic component, several environmental factors associate with PDAC incidence, with smoking as the most prominent (5). As with many cancers, data for numerous other environmental and lifestyle factors exist, however these are often contradictory. Those reported to confer an increased risk of PDAC include diets high in meat and fat, low in folate, and high in sugar, with diabetes, obesity and chronic pancreatitis therefore strongly associated with the disease (5,22-25). Interestingly, although many of these lifestyle factors are associated with increased social deprivation,

previous studies have not found the disease more prevalent within lower socioeconomic groups (3).

1.1.2 Disease development

The pancreas has both endocrine and exocrine functions key for digestion and blood glucose homeostasis. The endocrine portion contains numerous cell types that form the islets of Langerhans required for hormone secretion, making up 4.5% of the total pancreas volume (26). The remainder of the pancreas contains the acinar and ductal cells required for the production and transport of digestive enzymes into the intestinal tract. Histologically, PDAC, and its early precursor lesions, resemble these ductal cells and therefore the disease is thought to arise within this region (27).

The three neoplastic precursor lesions proposed to be involved in the development of PDAC include pancreatic intraepithelial neoplasia (PanIN), intraductal papillary mucinous neoplasia (IPMN), and mucinous cystic neoplasia (MCN). Of these, PanINs are the most common and, although highly prevalent in PDAC and pancreatitis, can sometimes be found in otherwise unaffected pancreata (28). PanINs are graded into three histological stages where progression from stage I through to III is characterised by increased disorganisation of the columnar mucinous epithelium with abnormal cellular appearance and mitosis (28-30) (Figure 1.1).

In early stage PanIN I lesions, cells are columnar with round nuclei located adjacent to the basement membrane (30). In the intermediate PanIN II lesions, cell atypia and nuclear changes are detected, whilst PanIN III lesions show widespread tissue dysplasia with small clusters of epithelial cells being found in the lumen (30). In the final transition of PanIN III into PDAC, cells begin to invade through the basement membrane (30). Alongside histological evidence, this progression model is supported at the genetic level, where an accumulation of genetic alterations associates with increased PanIN stage (31-37)).

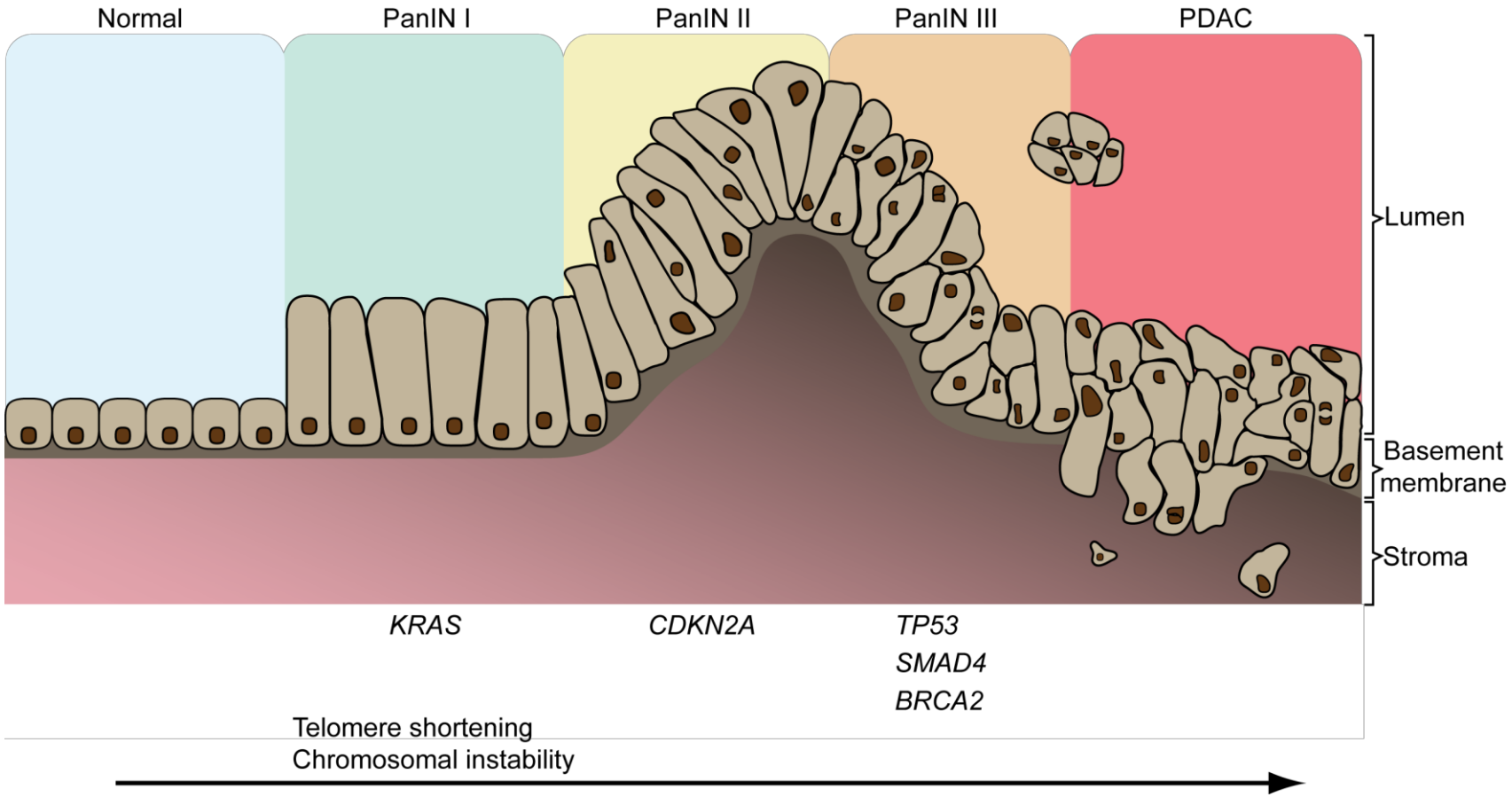

Figure 1.1 – A progression model depicting the genetic and histological changes for the transformation of normal ductal cells into invasive PDAC through increasing grades of PanIN lesions. A schematic for the PDAC progression inspired by the model proposed by Hruban *et al.* (29,30), which shows that normal ductal epithelium progress into infiltrating carcinoma (*left* to *right*) through a series of histologically defined precursor lesions. This progression is associated with an accumulation of genetic aberrations (33,37), where mutations in *KRAS*, telomere shortening (38) and increased chromosomal instability (39) occur as early events, before other mutations arise in the later precursor and disease stages.

1.1.3 Mutational landscape of PDAC

Next-generation sequencing approaches have accelerated our understanding of the recurring mutations present in PDAC (40-45). There appears to be a founder population of cells that have accumulated activating mutations in *KRAS* (present in > 90% of patients) (40,43), alongside frequent loss-of-function (LOF) mutations in *TP53* (50-75%) (43,46-49), and *SMAD4* (~55%) (49,50).

As a proto-oncogene, *KRAS* encodes a GTP(guanosine triphosphate)ase that is constitutively activated by point mutations that cause amino acid substitution. These predominantly involve changing the glycine residue from codon 12 (G12) into an aspartic acid (D), or a valine (V) residue. These nonsynonymous mutations cause GTPase-activating proteins to no longer bind KRAS and hydrolyse GTP into guanosine diphosphate (GDP) (51). This increased GTP causes continuous KRAS activation and persistent downstream signalling (51).

The next most commonly mutated gene in PDAC is *TP53*. Unlike *KRAS*, where one amino acid residue is predominantly substituted, several TP53 residues are commonly altered in PDAC (data from the IARC TP53 Database [http://www-p53.iarc.fr/] (52)), where the most frequent are the arginine (R) residues from codons 175, 248 and 273. Each of these are located in the DNA-binding domain of p53 where, as a transcription factor, these nonsynonymous mutations alter its ability to regulate gene expression (53). wild-type (WT) p53 target genes are implicated in cell-cycle arrest, senescence, apoptosis and DNA repair pathways, thus it commonly acts as a tumour suppressor (53). Alongside a LOF, these mutations can also cause a gain-of-function (GOF), whereby mutant p53 can increase oncogenic gene expression (54). One GOF mechanism proposed, and of particular relevance to this thesis, is that mutant p53 increases the expression of epigenetic enzymes involved in gene activation, such as the histone lysine methyltransferases KMT2A and KMT2D (55).

In addition to these frequently mutated genes, sequencing studies have also highlighted great genetic heterogeneity in the disease, with a long list of less commonly mutated genes (> 5% of cases) found in PDAC (40-45). Therein, a remarkable number of recurring copy number changes and mutations are found in genes that encode components of the epigenetic machinery, thus some form of addiction to epigenetic dysregulation may be a feature of these tumours. Among these genes are those that encode the histone lysine methyltransferases KMT2C (mixed-lineage leukaemia 3

(MLL3)) and KMT2D (MLL2) (42,44,45), where mutations have been reported in up to 16.7% and 6% of PDAC cases, respectively (40-45) (Table 1.1).

In the disease progression model proposed by Hruban *et al.* (29,30), PanINs progress into PDAC via a series of lesions associated with an accumulation of mutations (33,37), telomere shortening (38) and increased chromosomal instability (39) (Figure 1.1). *KRAS* mutation occurs early on in the initial stages, where sequencing identified a codon 12 mutation in 92% of microdissected PanIN I lesions (56)). In the intermediate PanIN II lesions, *CDKN2A* inactivation mutations begin to arise and can be found in 20% of PanIN II, and 36% of PanIN III, lesions (56). In the later PanIN stages, other inactivating mutations in genes such as *SMAD4*, *TP53* and *BRCA2* then begin to arise (37,49,57).

Despite this greater understanding of the mutational and genetic landscape in disease development, there has yet to be an influx of new therapies offering increased clinical benefit. Thus, we need to understand the underlying biology and contribution of these mutations in an effort to develop new therapies with novel mechanisms of action to improve treatment and outcome.

Study	Ref.	Total n =	*KMT2C*		*KMT2D*	
			n =	%	n =	%
Bailey *et al.* 2016	(40)	249[a]	18	7.2	14	5.6
Sausen *et al.* 2015	(41)	101	7	6.9	5	5.0
Waddell *et al.* 2015	(42)	100	8	8.0	6	6.0
Witkiewicz *et al.* 2015[b]	(43)	109	7	6.4	3	2.75
Biankin *et al.* 2012	(44)	99	6	6.1	0	0.0
Jones *et al.* 2008	(45)	24	4	16.7	0	0.0

Table 1.1 – Percentage of PDAC patients with nonsynonomous mutations in *KMT2C* and *KMT2D* reported by various next-generation sequencing studies. Displayed are data for nonsynonomous mutations taken from the supplemental data tables provided by six studies. a This study initially also contained data from Waddell *et al.* (42) but has been excluded and is shown independently, b Data was obtained from microdissected tumours to enrich for tumour cellularity (43).

1.1.4 Preclinical models of PDAC

PDAC tumours have a characteristically complex microenvironment due to a dense stromal desmoplasia comprised of extracellular matrix proteins, fibroblasts and immune cells (reviewed in (58) and (59)). To successfully function as a good preclinical model of PDAC, an *in vivo* system must recapitulate this complex tumour biology. Over recent years there has been a shift from simple cell line based syngeneic and xenograft ectopic mouse models towards more complex orthotopic and patient derived tumour xenograft (PDX) models. These PDX preclinical models are particularly useful because the patient tumour explant also includes patient stroma, thereby more faithfully reflecting primary tumour biology (60). Perhaps because of this retained donor stroma, these complex preclinical models have been effective in predicting clinical response for PDAC (61-63).

PDX models do however have their limitations that prevent their more widespread use (reviewed in (64) and (65)). Briefly, these models rely upon a highly skilled workforce, are costly to run, lengthy in duration, and often limited by supply of clinical tumour samples. As PDX models require a cross-species xenograft, an immunodeficient murine host is also used so that the graft is not rejected. For this reason, although donor stroma and tumour is included, these models lack a complete microenvironment with innate immune system, thereby preventing the study of immunotherapeutics, and the roles of the immune system in disease development, invasion and metastasis. Finally, studies also show that the human tumour stroma transplanted will overtime be replaced by murine host stroma (66,67). For these reasons, PDX models often unfaithfully recapitulate the complex features of disease progression, invasion and metastasis. Since these processes are important features of PDAC, efforts have been invested into developing genetically engineered mouse models (GEMMs) that incorporate them.

1.1.4.1 Genetically engineered mouse models of PDAC

Over the years several PDAC GEMMs have been developed, each with their own specific biological characteristics. Using prior mutational knowledge of $KRAS^{G12D}$ in PDAC development, the first significant pancreatic GEMM was developed by Hingorani *et al.* (68). In this model the Cre-Lox recombination system is employed to initiate $Kras^{G12D}$ expression in early pancreatic progenitor cells during embryogenesis (Figure 1.2). To achieve this, a genetic "STOP" sequence flanked by two LoxP sites is inserted directly before mutant $Kras^{G12D}$ and thereby preventing its transcription. This *LoxP-STOP-LoxP* (LSL) region generates a mouse functionally heterozygous for WT

Kras (*Kras*$^{+/-}$), with a non-functional allele harbouring *Lox-STOP-Lox-Kras*^{G12D} (*LSL-Kras*^{G12D}). By incorporating a transgene with the bacteriophage Cre recombinase under the control of the pancreatic-specific promoter *Pdx-1*, recombination of this LSL site is restricted to the pancreas, and elicits tissue specific expression of *Kras*^{G12D} (Figure 1.2).

These *LSL-Kras*$^{G12D/+}$*; Pdx-1-Cre* (KC) mice develop all three grades of PanINs, akin to those found in humans (68). Alongside these lesions, Hingorani *et al.* also showed that mice could develop PDAC, with this mouse showing both local invasion and metastasis after a long latency period (6.25 months) (68). Following this KC model, several other groups developed models that incorporate other mutations found in human PDAC (including *Cdkn2a*, *Trp53*, *Smad4*, and *Stk11*, reviewed in (69-71)). Whilst many develop PDAC with metastasis, the *LSL-Kras*$^{G12D/+}$*; LSL-Trp53*$^{R172H/+}$*; Pdx-1-Cre* (KPC) is the most commonly used model (72). In the KPC model, an additional LSL region is inserted before *Trp53*^{R172H}, to create a mouse functionally heterozygous for both *Kras* and *Tp53* (*Kras*$^{-/+}$*;Trp53*$^{-/+}$). Crossing these *LSL-Kras*$^{G12D/+}$*; LSL-Trp53*$^{R172H/+}$ (KP) mice with *Pdx-1-Cre* mice creates KPC offspring with pancreatic-specific expression of *Kras*^{G12D} and *Trp53*^{R172H} (72) (Figure 1.2). These KPC mice also develop PanINs, however now all mice develop PDAC with a median survival of five months (72).

Alongside these models, KC mice have been combined with the *Sleeping Beauty* transposon system (73). Instead of designing GEMMs that specifically study the cooperation of designated mutations with *Kras*^{G12D} to drive PDAC, this system uses mobile genetic elements to randomly integrate and interfere with genes throughout the genome. By determining the regions of transposon insertion in the subsequent tumours through sequencing, two groups identified candidate genes that promote progression of PanIN into PDAC (74,75). Interestingly, these screens highlighted significant roles for chromatin binding and chromatin modification pathways in PDAC tumourigenesis (74,75).

The main limitations of these GEMMs are that each arises via a disease aetiology that is unrepresentative of human PDAC. As previously discussed, PDAC occurs predominantly in aged humans through a series of many genetic events in ductal cells. It is worth keeping in mind that *Pdx-1* is expressed early in embryogenesis during pancreatic development (E8.5 onwards) (76), therefore mutant *Kras*^{G12D} is expressed in both exocrine and endocrine cell types (68). These features raise several issues over their effectiveness in modelling the human disease. Firstly, *KRAS* mutations are not

acquired during human embryogenesis, and instead are a later in life event in PanINs (77). Next, the human disease originates from ductal or acinar cells (27) that acquire a heterogeneous range of mutations (41,42,44,45), and therefore it is unrepresentative to only induce two mutations in all pancreatic cells.

Since the KC and KPC models were first generated, techniques in molecular biology have advanced, whereby some limitations could be addressed in a new generation of complex PDAC GEMMs. For example, Guerra *et al.* have developed inducible GEMMs that use a Tet-Off Cre-LoxP system, under the *Elastase* promoter, to induce acinar cell-specific expression of $Kras^{G12V}$, both alone (78) and in combination with loss of *Cdkn2a* or *Trp53* (79). New GEMMs may begin to combine further novel cell specific promoters, with inducible control mechanisms, and multiple DNA recombinases.

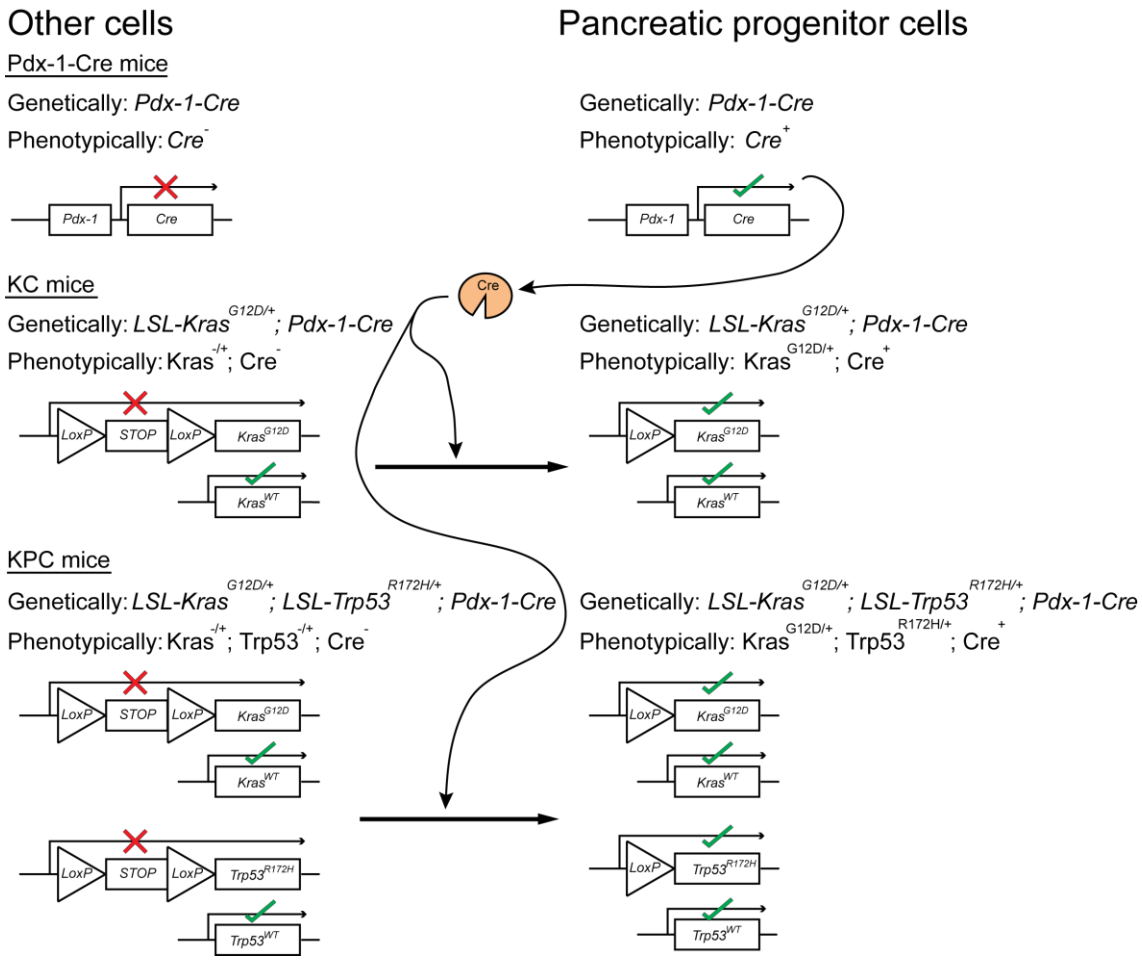

Figure 1.2 – The tissue-specific Cre-Lox recombinase system used in the genetically engineered KC and KPC mouse models of PDAC. The schematic diagram depicts how tissue-specific Cre expression can initiate pancreas-sepcific expression of $Kras^{G12D}$ alone, or in combination with Trp^{R172H}. In this way to generate the KC (68) and KPC (72) models, mice genetically engineered to express *Cre* under the *Pdx-1* pancreatic progenitor cell promoter were crossed with mice genetically engineered to be heterozygous for the mutated genes transcriptionally repressed by a LoxP flanked STOP cassette.

1.1.5 Current treatments, resistance, and future therapies

At diagnosis, PDAC patients typically present with advanced disease characterised by local tissue invasion and metastases. These features severely limit the opportunities for surgical resection but, despite common postoperative complications, it remains the current best option available (80). As a further treatment hurdle, the disease is often highly refractive to chemotherapy and thus the disease has an extremely poor prognosis (5-year survival of 8% (81), and median survival of 7 months).

To date, current chemotherapies used in the adjuvant and neoadjuvant settings include gemcitabine, FOLFIRINOX (combined folinic acid, 5-fluorouracil (5-FU), irinotecan and oxaliplatin) and abraxane (nanoparticle, albumin-bound paclitaxel), whilst newer stromal- and tumour-targeted therapies have been proposed (reviewed in (82)). Despite these efforts resistance still arises frequently and rapidly causing patients to succumb to the disease. This rapid resistance likely involves both innate and acquired mechanisms that involve the complex microenvironment, heterogeneous somatic mutations, and the post-translational regulation of gene expression.

PDAC tumours have a characteristically large and dense stromal compartment consisting of pancreatic stellate cells, fibroblasts, immune cells, and an abundance of extracellular matrix proteins that surrounds the cancer cells. This extensive tissue desmoplasia is often described as a ‘fortress’, whereby it actively provides a tumour supportive niche to promote immune system evasion, and shield the cancer cells from chemotherapy (83). As another hallmark, PDAC tumours develop with poor vascularisation, creating a hypoxic environment (83) that both prevents effective delivery of chemotherapy into the tumour, and promotes tumour aggression (84) and resistance (85). Together this stromal architecture limits the effectiveness of current chemotherapeutics, therefore there is much interest in stromal-targeted therapies to break down the fortress, ‘re-educate’ the immune system, and stabilise tumour vasculature.

Although the role of the stromal compartment is undoubtedly important, resistance also arises through changes within the cancer cell. For example, a cell that once responded to chemotherapy may undergo genetic and epigenetic changes to alter its sensitivity to these chemotherapeutics. This thesis is therefore focused on the understudied component of epigenetics within PDAC, to help identify new therapies with novel mechanisms of action.

1.2 Epigenetics

Historically the term 'epigenetics' has held several subtly different definitions. It has evolved from the original broad definition of 'non-genetic heritable changes in phenotype' to now describe 'the regulatory mechanisms that govern the expression of genetic information' (86). These mechanisms often control temporal and spatial expression of genes via reversible post-translational modifications, without altering DNA sequence (87). Although other epigenetic mechanisms exist, two of the most commonly described are the post-translational modifications on DNA at CpG (Cytosine-phosphate-Guanine) sites, and those on the histone tails that impact on DNA packaging (88). Broadly speaking, these mechanisms work by altering the interactions between DNA and various proteins to affect chromatin reorganisation and gene expression.

1.2.1 The structure of chromatin

In eukaryotic cells, DNA is packaged in nucleosomes through twice being wound around octamers containing pairs of the histones H2A, H2B, H3 and H4 (89) (Figure 1.3). Both the linker histones (H1) and small segments of linker DNA structurally separate these nucleosomes. In regions of inactive gene transcription nucleosomes are further tightly packaged into heterochromatin, whilst near regions of active gene transcription, the appearance is more akin to "beads-on-a-string" that is associated with loosely packaged euchromatin (Figure 1.3). Both forms of chromatin are further coiled into chromatin fibres, which again are further coiled and packaged into chromosomes.

Each histone within the nucleosome octamer has an N-terminal amino acid tail that protrudes out and may be post-translationally modified (90) (Figure 1.3). These modifications alter their interaction with DNA, and thereby facilitate the conversion of chromatin between the two states. Tightly packed heterochromatin renders the DNA inaccessible to transcription machinery, whereas the opposite is true for loosely packaged euchromatin. In this manner, histone modifications act to exert control over the transcription of regions of the genome.

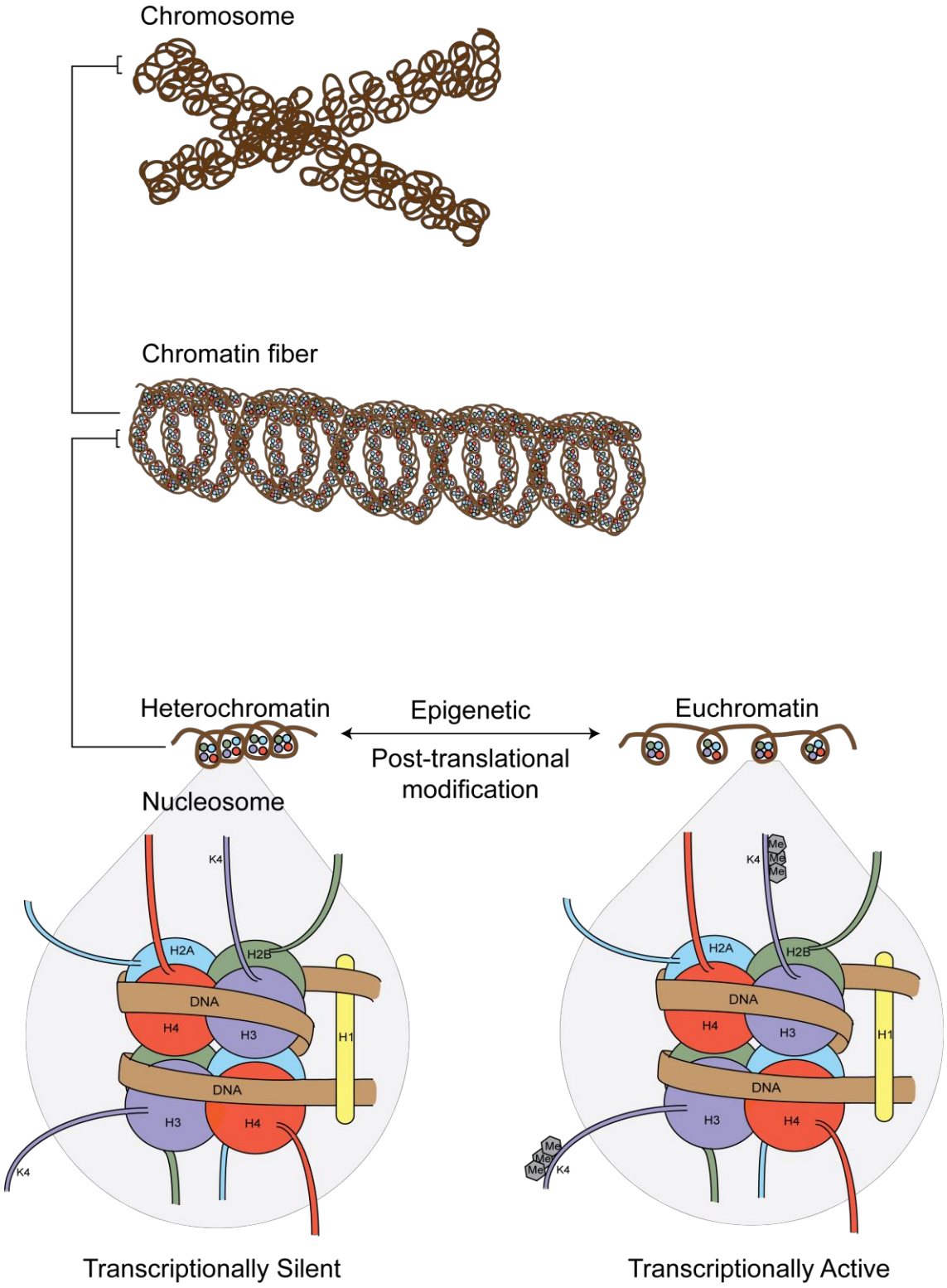

Figure 1.3 – Structure and organisation of DNA packaging from chromatin to the nucleosome and changes associated with transcriptionally active and inactive states. A schematic diagram for different levels of DNA organisation in mammalian cells from the nucleosome to the chromosome (*bottom* to *top*), and for the changes associated with transcriptional activation (*bottom*, *left* to *right*). Chromosomes are made up of condensed 30nm chromatin fibres that, in turn, are comprised of coiled sections of tightly packaged nucleosomes and DNA known as heterochromatin. When this region becomes transcriptionally active, for example through reversible H3K4 methylation, heterochromatin is converted into loosely packaged euchromatin.

1.2.2 'Readers' 'Writers' and 'Erasers'

The proteins involved in the epigenetic control mechanisms can be broadly classed as 'writers', 'erasers' and 'readers' (91). The enzymatic 'writers' and 'erasers' contain domains to catalyse the addition, or removal, of covalent post-translational modifications, while 'readers' contain binding domains to recognise the modifications. This enables further protein recruitment, forming scaffolds and complexes that facilitate a diverse range of biological functions such as gene transcription, DNA repair, and even further epigenetic modification.

In truth this broad classification is overly simplistic, where 'writers' and 'erasers' often also have binding domains for specific epigenetic marks. For example, members of the KMT2 family (see Section 1.3) contain a catalytic Su(var)3-9, E(z) (enhancer of zeste) and trithorax (SET) domain, and plant homeodomain (PHD) zinc fingers that recognise methylated H3 (92). Moreover, the epigenetic machinery often involves large protein complexes with many components considered to be 'writers', 'erasers', and 'readers'. For example, the methyltransferases KMT2C and KMT2D form complexes that also contain the demethylase KDM6A (UTX) (93). This is perhaps unsurprising considering both the relatively large size of the nucleosome substrate, and that epigenetic modifiers work in tandem to alter the balance of modifications.

1.2.3 Histone modification

Epigenetic modifications of the DNA itself include methylation and hydroxymethylation, whereas a much wider range of post-translational marks exist for histone modification. Thus far histones have been described to be enzymatically modified by methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, ADP-ribosylation, and arginine deamination (reviewed in (90)). Depending of the chemistry for each modification, certain ones will only occur on specific amino acid residues with, for example, methylation only found on lysine and arginine residues (90). Epigenetic modifiers are further limited in their specificity, whereby they target select residues (94). Following the discovery of the first epigenetically active enzyme in humans (95), a large number have since been identified across a broad range of families, many of which are likely to have unique role in influencing gene expression and biological processes (reviewed in (96) and (97)).

Together this range of histone modifications cooperates as part of a complex 'histone code', which allows cells to regulate chromatin structure and orchestrate gene expression programs (90). Although dissecting the information contained within this

complex 'histone code' is highly challenging, recent methods that combine chromatin immunoprecipitation (ChIP) with DNA sequencing (ChIP-seq) have helped greatly. ChIP-seq is widely used to identify DNA binding sites for transcription factors and epigenetic enzymes (*e.g.* (55)), and elucidate genome wide patterns of histone modification (*e.g.* (98)) (Figure 1.4). In this way, associations have been made between epigenetic marks and gene regions, such as promoters and enhancers, and provided insights into the functional roles of individual histone modifications.

1.2.3.1 Histone methylation

Epigenetic enzymes differentially methylate a large number of the lysine residues found within histone tails, where the majority of the methylation events known occur upon histone H3. Commonly methylated residues include K4, K9, K26, K27 and K36 amongst others, where all can be un-, mono-, di- and tri-methylated (99). As part of the 'histone code' this methylation of individual residues, to different levels, are associated with a range of functions (100).

Lysine methylation is mediated by lysine-specific methyltransferases (KMTs) and requires the co-factor S-adenosyl-L-methionine (SAM) (101,102). These KMTs can be divided into two groups based on the presence, or absence, of a SET domain, however to date only KMT4 (DOT1L) falls into the latter (96). Across the many SET domain-containing KMTs, there is a large variance in substrate residue specificity (97), which can also depend on methylation status (96), and can be mediated by various structural features and interacting partners (96).

Until the discovery of the first lysine demethylase (KDM), KDM1A (LSD1) by Shi *et al.* (103,104), it was thought that methylated lysine residues were irreversible events that act like scars on the genome. Since then a large number of KDMs have been described that harbour a similar range in substrate specificity as the KMTs (97). Removal of these methyl groups by KDMs occurs through two mechanisms, with KDMs grouped accordingly (96). Briefly, these mechanisms either involve a flavin adenine dinucleotide (FAD)-dependent amine oxidase domain (*e.g.* KDM1A), or a Jumonji C (JmjC) domain, which requires α-ketoglutarate, molecular oxygen, and Fe(II) as cofactors (*e.g.* KDM2A) (96).

1.2.3.1.1 H3K4 methylation

Methylation of H3K4 is associated with transcriptionally poised and active genes (98,105,106). Studies by Barski *et al.* and Mikkelsen *et al*. in the late 2000s used ChIP-

seq to identify the genome-wide profile of H3K4 methylation in both humans (98) and mice (105). Barski *et al.* showed all three H3K4 methylation states as being enriched at transcription start sites (TSSs), and as positively correlating with both RNA polymerase II binding and gene expression (98) (Figure 1.4). In line with this, Mikkelsen *et al*., and another study by Guenther *et al.* using ChIP-chip in human embryonic stem cells, showed that H3K4me3 is commonly associated with promoters (105,106). Promoters marked by H3K4me3 are described as poised while marked with repressive H3K27me3, and active when H3K27ac is present (107,108) (Figure 1.4).

The amount and level of H3K4 methylation at promoters has been shown to increase with proximity to the TSSs (98,109) (Figure 1.4). While H3K4me3 is thought to recruit epigenetic 'readers' with PHD fingers (92), it is unclear as to whether this methylation gradient is transitional feature, or indeed has functions of its own. One role proposed is that this gradient is required for cell specific transcription factor recruitment, where H3K4me2 is identified at 90% of transcription factor binding regions (110) with differences depending on the cell type (111).

Alongside its presence at promoters, H3K4 methylation has been identified at other genomic regions. The study by Barski *et al.*, and another by Heintzman *et al*. using ChIP-chip in HeLa cells (109), also showed that H3K4me1 associates with enhancers (98), which again can be marked as poised or active by H3K27me3 and H3K27ac respectively (107,108) (Figure 1.4). Studies on H3K4me3 levels at enhancers are however somewhat contradictory, with both weak (108,109) and that similar to H3K4me1 (98) having been shown. Perhaps associated with these discrepancies, another group has since shown that the H3K4me3 mark might be indicative of enhancer activation (112).

Besides marking enhancers, H3K4me1 is also present at insulators, where it may recruit proteins that indicate boundaries between differentially expressed genes (98,113). In addition, H3K4me1 can also define a group of promoters, without H3K4me3, that are transcriptionally repressed (113) (Figure 1.4). Therefore despite H3K4 methylation being traditionally associated with active gene transcription, other complex roles for each mark are likely to exist. To this end, rather than studying the marks themselves, it may be more useful to examine the biological functions of the epigenetic enzymes that maintain these marks.

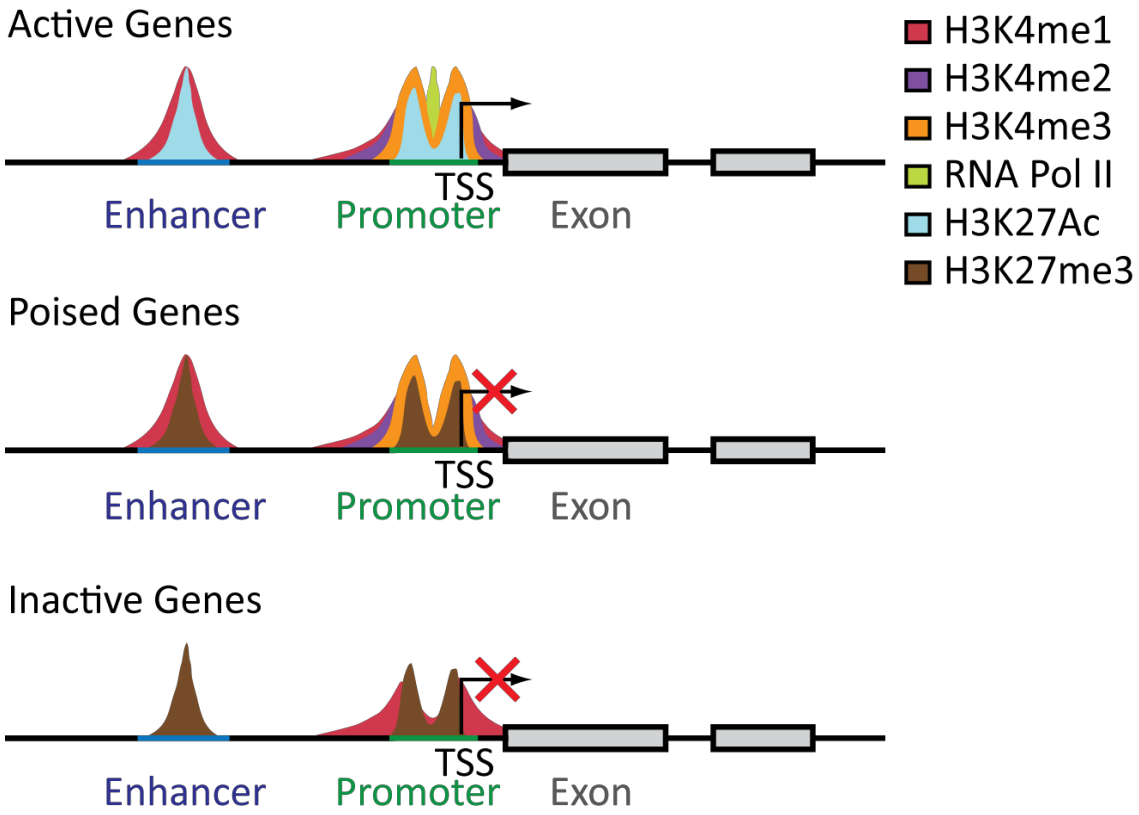

Figure 1.4 – Profiles of chromatin marks found to associate with specific regions of transcriptionally active, poised and inactive genes. A schematic depiction of the associations between selected epigenetic marks and specific regions in active, poised and inactive genes. H3K27me3 is enriched at enhancer and promoter regions of inactive and poised genes to prevent active transcription, indicated by enriched RNA polymerase II (RNA Pol II) binding, until replaced by H3K27ac. While mainly associated with active and poised enhancers, H3K4me1 can also be found at inactive promoters, and on the periphery of active and poised promoters. These inactive promoters exhibit the full range of H3K4 methylation statuses, where the density of H3K4me3 is greatest proximal to the transcription start site (TSS).

1.3 The KMT2 family of lysine methyltransferases

1.3.1 Members of the KMT2 family

Methyltransferases and demethylases are required to actively balance H3K4 methylation. In yeast only one H3K4 targeted SET domain containing protein is known (Set1), whereas four have been described in *Drosophila melanogaster* (dSet1, Trx (Trithorax), Trr (Trithorax-related), Ash1 (Absent, small or homeotic disc 1)) (114), and eight in humans (Figure 1.5). This structurally related group of proteins is called the KMT2 family and consists of KMT2A (MLL1/ALL1), KMT2B (MLL2/MLL4), KMT2C (MLL3/HALR), KMT2D (MLL2/ALR/MLL4), KMT2E (MLL5), KMT2F (SET1A), KMT2G (SET1B) and KMT2H (ASH1L) (97,115). In humans, the relative expression levels for each member has been shown to differ between tissue types (116).

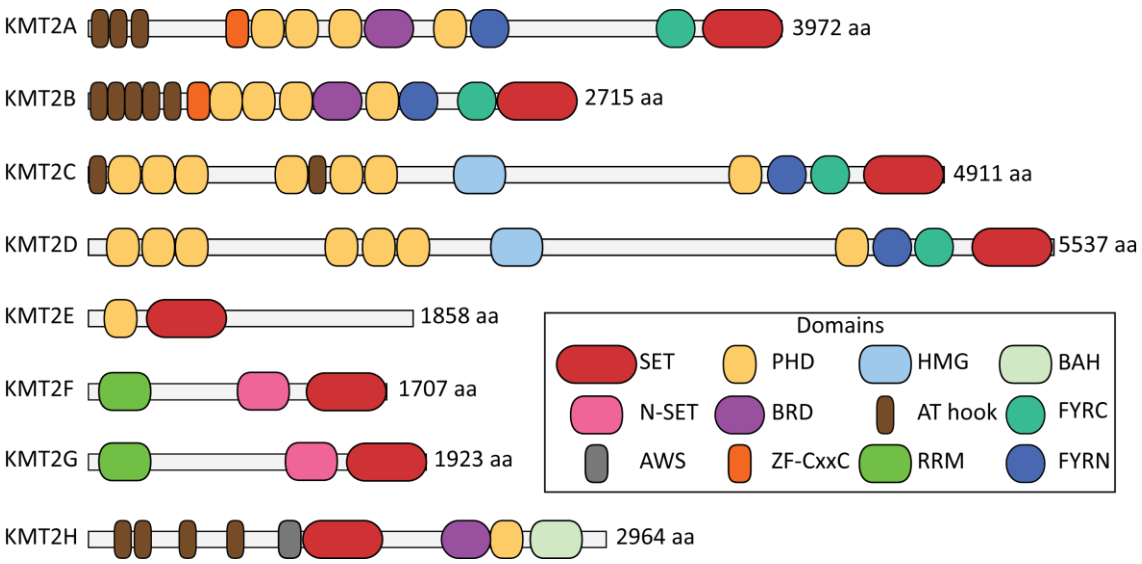

Figure 1.5 – The domain architecture for the KMT2 family of proteins in mammals. A schematic representation of the various domains found in each member of the KMT2 family of lysine specific methyltransferases using data from Simple Modular Architecture Research Tool (SMART) (117), inspired by figures in Zhang *et al.* (118), Rao and Dou (119) and (120). In all but two family members (KMT2E/H) the enzymatic SET domain is located at the C terminus. For KMT2F/G this domain is in close proximity to an N-SET domain and alongside the RNA recognition motif (RRM) is essential for complex formation (121) to promote trimethylation (122). Alongside the SET domain, a PHD finger and two phenylalanine/tyrosine-rich (FYR) regions (FYRC and FYRN) are found towards the C terminus of KMT2A-D, where these are juxtaposed in KMT2C/D. Multiple PHD fingers are also found at the N termini of KMT2A-D, where together they facilitate localisation to methylated histones (92). Unlike the other family members, KMT2C/D contain an HMG-box domain to bind DNA (123). The N termini of KMT2A/B differ from KMT2C/D and instead include multiple AT hooks and a CXXC zinc finger (ZF-CxxC) for DNA binding, and a bromodomain (BRD) to recognise acetylated lysine residues (124). Although its architecture is more distinct, KMT2H also contains many of these domains alongside bromo-adjacent homology (BAH) and associated with SET (AWS) domains.

1.3.2 A note on the literature: *MLL2* and *MLL4* nomenclature inconsistencies

The literature surrounding the KMT2 family of methyltransferases is somewhat convoluted (reviewed in (115)). Over the years the gene located at 12q13.12 with the human accession number NM_003482, has had many aliases. In 1998 the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) initially approved *MLL2* for this gene (115). The following year however, this was confounded when FitzGerald and Diaz referred to another gene at 19q13.12, accession number NM_014727, as *MLL2* (125), before being further propagated by Huntsman *et al.* (126). As a result, there is often uncertainty towards which gene is being addressed when gene loci, accession numbers, and specific knowledge of the complexes are not reported.

In addition to *MLL2* being used for both genes (*e.g.* (116,127-129)), both have also been inconsistently referred to as *MLL4* (*e.g.* (116,130,131)). In this way, the murine orthologue of the human gene on chromosome 12, located on the murine chromosome 15, is often referred to as *Mll4* (*e.g.* (130,132,133)), despite the Mouse Genomic Nomenclature Committee (MGNC) approving the designation *Mll2* (115). To add further confusion, the murine orthologue of the human gene on chromosome 19, located on murine chromosome 7, was initially designated *Wbp7* (115), but it too has been referred to as both *Mll2* (*e.g.* (132,134,135)) and *Mll4* (*e.g.* (129)). For these reasons a great deal of care and attention is required when both interpreting and drawing conclusions from many of the studies using the *MLL2* and *MLL4* nomenclature, especially where gene IDs or chromosomal locations are not stated (*e.g.* (136)).

In light of the ambiguity that these inconsistencies introduced to the literature, other nomenclature systems have been used to provide some clarity. Originally, a little success was found by referring to the human gene on chromosome 12 as *ALL-1* related gene (*ALR*) (137,138), however this was not widely used in the literature. The more recent nomenclature system proposed by Allis *et al.* included not just the KMT2 family, but other epigenetic enzymes that have had similar issues with naming conventions (97), and has since been propagated by Bogershausen *et al.* (115). This nomenclature system is gaining acceptance for many of these genes by both the HGNC and MGNC, however only time will tell if it helps to remove the ambiguity. Despite this however, some studies, and a review, have already begun to use this nomenclature inappropriately (*e.g.* (139-141)).

Dm	Old nomenclature		New nomenclature		Human		Mouse	
	Human	**Mouse**	**Human**	**Mouse**	**HGNC ID**	**Location**	**MGI ID**	**Location**
Trx	*MLL*	*Mll1*	*KMT2A*	*Kmt2a*	7123	Chr11q23.3	96995	Chr9 24.84 cM
	MLL4	*Wbp7*	*KMT2B*	*Kmt2b*	158040	Chr19q13.12	109565	Chr7 18.63 cM
Trr	*MLL3*	*Mll3*	*KMT2C*	*Kmt2c*	13726	Chr7q36.1	2444959	Chr5 12.35 cM
	MLL2	*Mll2*	*KMT2D*	*Kmt2d*	7133	Chr12q13.12	2682319	Chr15 54.8 cM
-	*MLL5*	*Mll5*	*KMT2E*	*Kmt2e*	18541	Chr7q22.1	1924825	Chr5 10.33 cM
Set1	*SETD1A*	*Setd1a*	*KMT2F**	*Kmt2f**	29010	Chr16p11.2	2446244	Chr7 69.73 cM
	SETD1B	*Setd1b*	*KMT2G**	*Kmt2g**	29187	Chr12q24.31	2652820	Chr5 62.8 cM
Ash1	*ASH1L*	*Ash1l*	*KMT2H**	*Kmt2h**	19088	Chr1q22	2183158	Chr3 39.01 cM

Table 1.2 – Nomenclature and gene identification details for the KMT2 family members. Displayed are both the old and new nomenclature (as proposed by (97)) used for the methyltransferases in human and mouse, relative to their homologues in *Drosophila melanogaster* (Dm). For each gene, designated IDs are given for the Human Genome Organisation Gene Nomenclature Committee (HGNC) and the Mouse Genome Informatics (MGI). cM = centimorgan, * = Alias not yet approved.

1.3.3 KMT2 complexes and their roles

Although the KMT2 family is structurally and functionally related, there are significant differences between the members. The most evident difference is that, despite the presence of a SET domain (Figure 1.5), KMT2E appears to lack functional methyltransferase activity (142), unless the SET domain is glycosylated (143). With the exception of KMT2E and KMT2H (142), each of the other KMT2 proteins act as the catalytic subunits within multimeric complexes to catalyse the addition of methyl groups to H3K4 residues (115,144). In the majority of the KMT2 family members the SET domain is located at the C terminus, however in KMT2H this domain is located towards the centre (145). This structural difference is likely a factor that causes KMT2H to form distinct complexes, where it can methylate H3K9 and H4K20 in addition to H3K4 (145).

Each of these KMT2-containing complexes also exhibits differences in substrate specificities. In mammals, three types of KMT2 complexes exist; KMT2A/B, KMT2C/D, and KMT2F/G, where each are related to the *Drosophila* complexes Trx, Trr and dSet1, respectively (**Error! Reference source not found.**). Each of these complexes contains a KMT2 paralogue that likely arose from the *Drosophila* orthologues during mammalian evolution (reviewed in (119)). Each KMT2 complex also commonly includes the four subunits of WD repeat protein 5 (WDR5), retinoblastoma binding protein 5 (RBBP5), absent, small or homeotic disc 2 (ASH2) and Dumpy-30 (DPY30) (93,119). In addition to this core complex, other complex specific subunits can be found, where the expression of these non-core subunits, and therefore the complexes themselves, is greatly heterogeneous (93) and perhaps differs between cells. In this way, label-free quantitative mass spectrometry performed by van Nuland *et al.* showed that in HeLa cells, KMT2A/B complexes make up majority of KMT2 complexes (up to 50%), while the KMT2F/G and KMT2C/D complexes comprise the remaining 32% and 18%, respectively (93). Further to this, it was also estimated that within these cells and of the KMT2C/D complexes, KMT2D made up the majority (approximately 15% of all complexes, compared to 3%) (93).

KMT2A and KMT2B complexes contain the unique subunits of Menin and host cell factor 1 (HCF1) (146), and act to predominantly catalyse H3K4me1/me2 formation (147,148). The KMT2C and KMT2D complexes contain the unique subunits of PTIP, PA1, NCOA6 and the lysine demethylase KDM6A (UTX), and mainly act as monomethyltransferases (94,113,118,120,149-153). The KMT2F and KMT2G

complexes contain the unique subunits of CXXC1, WDR82 and HCF1, and can catalyse all three H3K4 methylation states (H3K4me1/me2/me3) (154).

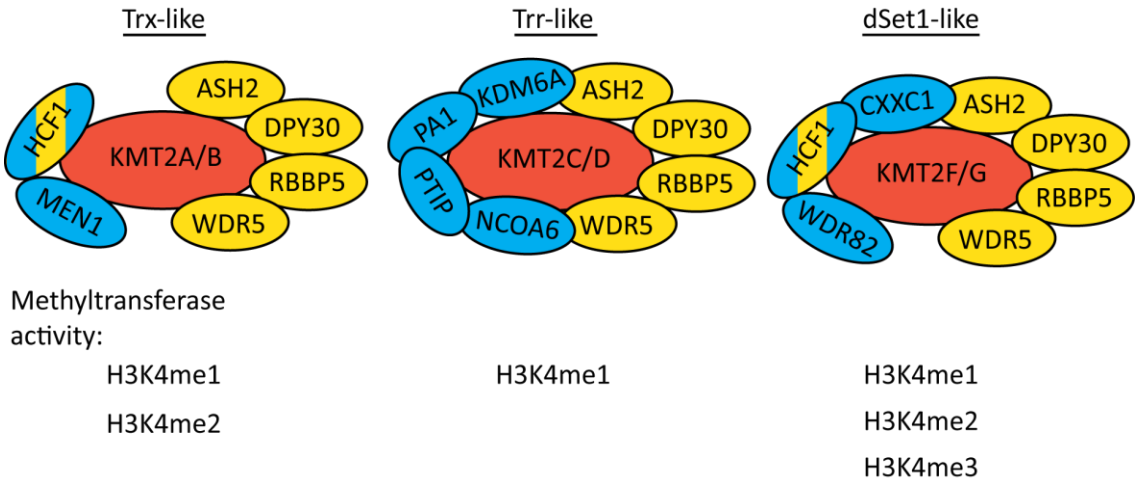

Figure 1.6 – Subunit composition of the similar KMT2 complexes and their H3K4 methyltransferase activity. In mammals the six KMT2 complexes can be sub-grouped into three groups ('Trx-like' for KMT2A/B, 'Trr-like' for KMT2C/D, and 'dSET1-like' for KMT2F/G) using their subunit composition and their similarity to the drosophila melanogaster complexes. For each complex the KMT2 methyltransferase subunit is in red, the four core subunits commonly found in all complexes are in yellow, and those in blue are unique subunits for each complex sub-type. Below each complex is the H3K4 methylation mark for which they have been shown to be active.

The subunit diversity of these KMT2 complexes is likely to play a role in targeting the complexes to different genomic regions. In addition, because each has subtly different substrate and catalytic capabilities, they are likely to have differing functions at their targeted loci. For example, because H3K4me1 is found predominantly at enhancers (98,109), and the KMT2C/D complexes are monomethyltransferases (120,151), it is perhaps unsurprising that many studies have shown these complexes to be enriched, and enzymatically active, at enhancers marked with H3K4me1 (120,149-152). As the KMT2 family members and their complex subunits are heterogeneously expressed across tissue types (116), it is highly likely that the roles of these complexes in regulating transcription are context dependent.

Besides their roles in histone posttranslational modification, it is worth noting that methyltransferases have been reported to have roles in signal transduction through the methylation of non-histone proteins (155,156). Unlike with protein phosphorylation, research into the area of non-histone protein methylation is comparatively understudied with much still remaining to be elucidated. Although little evidence exists, the KMT2

family of lysine methyltransferases may also act in this way having enzymatic activity on proteins besides H3K4. Perhaps supporting this potential is the observation by Zhang *et al.* where the yeast Kmt2 homologue, Set1, can methylate Dam1, which acts as the essential member within the DASH complex for chromosome segregation (157).

1.3.3.1 KMT2 proteins implicated in medical conditions

In normal cells the epigenetic 'readers', 'writers' and 'erasers' work together to maintain a balance of post-translational modifications to enhance or repress gene expression and maintain the chromatin structure. Since a large number of target genes can become mis-regulated if the epigenetic machinery goes awry, it is unsurprising that changes in it have been functionally implicated in number of medical conditions (reviewed in (158,159)). In this way, germline mutations in the *KMT2* genes have been implicated in several neuropsychiatric disorders.

In 2010, a sequencing study identified germline mutations in *KMT2D* as being a major cause of Kabuki syndrome (128). In this study, and a more recent systematic review of 12 studies since, these mutations are described as heterozygous LOF that result in *KMT2D* haploinsufficiency (128,160). This syndrome is phenotypically characterised by distinct facial characteristics, cardiac, skeletal and immunological defects, and intellectual disabilities (128,161), and has been somewhat recapitulated in a mouse model with heterozygous loss of the *Kmt2d* SET domain (162).

KMT2 family germline mutations have also been reported in other neuropsychiatric disorders. Whole-exome sequencing by Jones *et al.,* identified *de novo KMT2A* mutations, with reduced gene expression, in five out of six individuals with Weidermann-Steiner syndrome (163). Furthermore, germline *KMT2C* mutations have been identified in both Autism spectrum disorder (164) and Kleefstra spectrum syndrome (165), while *de novo KMT2F* LOF mutations can be found in patients with schizophrenia (166,167).

Although as yet unconfirmed, observations made in a review by Rao and Dou suggest that there may be a link between *KMT2* germline mutations and cancer (119). The authors noted that schizophrenia patients have a significantly higher risk of developing colon cancer (168), where their own analysis of the Catalogue of Somatic Mutations in Cancer (COSMIC) database (169) showed a link for this cancer and *KMT2F* mutation (119). Alongside this, eight Kabuki syndrome individuals were reported to have developed cancer, however whether this is a *bona fide* link remains unknown (161). For

KMT2A and *KMT2C* germline mutation disorders, associations with cancer incidence have not been described (119). Despite this uncertainty linking *KMT2* germline mutations with cancer, a greater wealth of knowledge exists for their somatic mutation.

1.4 The KMT2 family in Cancer

The KMT2 family of methyltransferases actively catalyse the formation of epigenetic marks associated with gene expression and therefore it is perhaps unsurprising that several members have been implicated in a range of solid and haematological malignancies (reviewed in (119)). Next-generation exome sequencing of 3,281 tumours across 12 tumour types from The Cancer Genome Atlas (TCGA) identified *KMT2* genes as among the most commonly mutated in many cancers (170,171), where these mutations are often nonsense and frameshifts that result in protein truncation (119).

1.4.1 Leukaemia and Lymphoma

As the previous MLL (mixed-lineage leukaemia) nomenclature suggests, mutations in these genes are implicated in haematological malignancies (reviewed in (172)). *KMT2A* is often rearranged by chromosomal translocation, giving rise to oncogenic fusion proteins implicated in both childhood and adult acute leukaemias (172). Although over 70 different translocation fusion partners are known (119), five (*AF4*, *AF6*, *AF9*, *AF10* and *ENL*) represent approximately 80% of those found in these leukaemias (119,172). While the many fusion partners are likely to have different functions (172), KMT2A fusion proteins are hypothesised broadly to interact with transcription factors and add permissive epigenetic marks for continuous transcription of leukaemic genes (119,172). These *KMT2A* translocations correlate with poor outcome for patients with acute lymphoblastic leukaemia (ALL) (173,174), and acute myeloid leukaemia (AML) (174-176).

KMT2D-inactivating mutations are commonly described in the non-Hodgkin lymphomas (NHLs) (177-180), whereby it has been shown that over 80% of follicular lymphomas (FLs) (179,180), and over 30% of diffuse large B-cell lymphomas (DLBCLs) (179) harbour *KMT2D* inactivating mutations. These *KMT2D* LOF mutations are thought to arise as driver mutations during the early events of tumourigenesis (180). In support of this, Kantidakis *et al.* recently used human colorectal carcinoma and mouse embryonic fibroblast cell lines to show that loss of *KMT2D* increased genomic instability, especially at regions of increased transcriptional stress (152). In line with this role of KMT2D as a tumour suppressor in NHLs, its mutation has been shown to increase B-cell proliferation, potentially through reduced levels of H3K4 methylation (181). In another concomitant study, this H3K4 methylation loss was shown at known tumour suppressor genes, and others involved in B-cell differentiation, B-cell receptor signalling and activation (182). These studies in

NHLs, combined with the observation by Kantidakis *et al.* that differential methylation may contribute to genomic instability (152), provide a potential mechanism for *KMT2D* LOF in driving tumour formation.

1.4.2 Solid cancers

KMT2 aberrations have also been reported in a range of solid cancers (119,170,171,183). Upon examining the supplemental data for the 11 solid tumours described in Kandoth *et al.,* 925 of the 1079 *KMT2* mutations are non-silent (170). In line with this, COSMIC database analysis performed by Rao and Dou identified that these mutations are predominantly heterozygous and result in LOF (119).

KMT2A mutations have been found in several cancers, including lung adenocarcinoma (184), transitional cell bladder cancer (185), endometrial carcinoma (186), and follicular thyroid cancer (187). In comparison, mutations in *KMT2B*, *KMT2F* and *KMT2G* appear to occur far less frequently in patient tumours (observations made from the COSMIC database in reviews by Neff and Armstrong (188), and Rao and Dou (119)).

Despite its frequency of mutation in cancer, RNAi-mediated depletion of *KMT2A* reduced viability across a range of cell lines, with this effect stronger in malignant than non-malignant lines (189). *KMT2A* depletion negatively impacted on cell-cycle progression, where cells accumulated in the G_2/M-phase, and induced apoptosis *in vitro* (189). This cell-cycle regulatory role is also supported observations that the KMT2A complex positively regulates expression of cyclin-dependent kinase inhibitors (190). Furthermore the *KMT2A* silencing also reduced tumour growth, with a potential role in angiogenesis, where endothelial cells in regions of hypoxia had high *KMT2A* expression (189).

1.4.2.1 KMT2C and KMT2D

A substantial number of mutations are found in *KMT2C* and *KMT2D* (119,188). As with *KMT2A*, these mutations are found in a wide range of solid malignancies including PDAC (40,41,44,45,191), among others (183,185,192-196). Thus far very few studies have examined these methyltransferases in PDAC, therefore their specific roles and mutation frequencies in this disease remain largely unknown and debatable, and therefore can only be inferred using observations from other solid tumours.

Murine studies have suggested that mutations in *Kmt2c*, and perhaps therefore also *Kmt2d,* may be an important feature in tumour formation and development, whereby they act as tumour suppressors (74,75,133). Two groups, using the murine *Sleeping*

Beauty transposon system, identified *Kmt2c* mutations as cooperating with $Kras^{G12D}$ to accelerate the progression of PanIN to PDAC (74,75). Homozygous *Kmt2c* mutation can also combine with heterozygote loss of *Trp53* to form ureter epithelial tumours (133). Despite *KMT2* mutations being mostly heterozygous in human PDAC (119), these studies suggest that *KMT2C* mutations may be implicated in the development of this disease, where both *KRAS* and *TP53* mutations are common events (37,56). Supporting this role in disease progression, a *KMT2D* mutation was identified as an event in the clonal evolution of one patient's PDAC (197). Furthermore, given that Kantidakis *et al.* have associated KMT2D loss with genomic instability (152), and Waddell *et al.* showed *KMT2D* mutations as being primarily found in a locally rearranged subtype of PDAC (42), their loss may give rise to other somatic mutations that drive cancer progression through increased genomic instability. In this way it is perhaps unsurprising that *KMT2C* and *KMT2D* aberrations have been found during the clonal evolution and progression of many other cancer types (198).

Alongside their apparent tumour suppressor roles, several studies have shown that these methyltransferases are also implicated in regulating cell proliferation across a variety of cancer types (55,133,135,136,138,151,181,199-203). From these studies, the roles of KMT2C and KMT2D in proliferation may be somewhat cell-type dependent. Loss of KMT2D expression and its activity have been shown to exert both negative (55,135,138,151,202) and positive (181,200) effects on cell proliferation, where the former occurs in solid cancers, and the latter in B cell malignancies. For KMT2C, the picture is perhaps clearer, where its activity has been shown to negatively correlate with (133,199,201,203), or at least have no effect on (136), the proliferation of cells derived from both solid (133,136,203) and haematological malignancies (201). This suppressive role of KMT2C in proliferation might indicate why *KMT2C* mutations have been associated with worse prognosis for patients with medulloblastoma (204).

In addition to their mutation, KMT2 family members are differentially expressed in tumours and cancer cell lines compared to normal tissues (116). As KMT2C and KMT2D function as enzymes, it is likely that their reduced expression might functionally resemble LOF mutation. In this way, as with its mutation (204), reduced expression of *KMT2C* has been associated with poor prognosis in both breast (194) and gastric cancers (205). In contrast, the opposite has been found for the expression of *KMT2D* (194,202), and *KMT2A* (194) in breast cancer and their mutation in PDAC (41).

1.5 Statement of problem and thesis aims

Within the treatment of PDAC, there is an unmet need for novel therapeutic approaches. For this reason, research into epigenetic proteins that influence many cell processes, such as cell proliferation, differentiation, metabolism, and response to therapy (206), may be of therapeutic benefit. Despite next-generation sequencing studies identifying *KMT2C* and *KMT2D* mutations in PDAC, little is known about their functional roles and the consequences of their change in disease. The work presented herein aimed to address this and understand how KMT2C and KMT2D may be involved in PDAC. In this way, this study set out to better understand the roles of these methyltransferases in PDAC, and explore whether they present novel opportunities for consideration as epigenetic therapies in this disease. To this end, data presented here have also been published in *Cancer Research* (207) (Appendix I).

The specific aims of this thesis are:

- To examine whether *KMT2C* and *KMT2D* expression levels, like their LOF mutation, can identify PDAC patients with improved prognosis.
- To investigate the proliferative effects associated with their depletion *in vitro*, with *post-hoc* cell-cycle analysis performed to further resolve mechanisms.
- To identify and validate genes differentially expressed in upon the *in vitro* depletion of these methyltransferases in PDAC cells. Further bioinformatic analysis will be performed to identify pathways implicated upon depletion.
- To relate data generated back to the clinical setting by comparing it with the patient gene expression profiles associated with *KMT2C* and *KMT2D* expression levels.
- To explore whether their expression alters chemotherapy response *in vitro*, and to develop PDAC GEMMs to study their role *in vivo*.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Cell lines and culture media

Each cell line was routinely checked for mycoplasma contamination on a monthly basis using the MycoAlert PLUS kit (Lonza, Cat#: LT07-710). The authenticity of each human cell line was previously confirmed either by our colleagues, or ourselves, using small tandem repeat profiling conducted by LGC standards and American Type Culture Collection (ATCC).

2.1.1.1 Human cell lines

Cell line	Cell origin	Growth medium	Reference
PANC-1	Pancreas tumour	DMEM (Sigma-Aldrich, Cat#: D5796), heat inactivated HyClone® FBS (10% v/v) (Thermo Fisher Scientific, Cat#: 12360273), penicillin (100units/ml) and streptomycin (100µg/ml) (Sigma Aldrich, Cat#: P4333)	(208)
Capan-2	Pancreas tumour		(209)
HPDE	Pancreatic ductal epithelium		(210)
BxPC-3	Pancreas tumour	RPMI-1640 (Sigma Aldrich, Cat#: R8758), heat inactivated FBS (10% v/v), penicillin (100units/ml) and streptomycin (100µg/ml)	(211)
SUIT-2	Liver metastasis		(212)
RWP-1	Liver metastasis		(213)
COLO 357	Lymph node metastasis		(214)
CFPAC1	Liver metastasis	IMDM with HEPES (25mM) (Lonza, Cat#: BE12-726F), heat inactivated FBS (10% v/v), L-glutamine (2mM) (Sigma Aldrich. Cat#: G7513), penicillin (100units/ml) and streptomycin (100µg/ml)	(21)

Table 2.1 – List of human cell lines used. DMEM = Dulbecco's Modified Eagle's Medium. FBS = Foetal Bovine Serum. RPMI = Roswell Park Memorial Institute. IMDM = Iscove's Modified Dulbecco's Medium. PANC-1, Capan-2, BxPC-3, Suit-2, RWP-1 and CFPAC1 cell lines were a kind gift from Dr Tatjana Crnogorac-Jurcevic, and the HPDE cell line was a kind gift from Professor Yaohe Wang, both Barts Cancer Institute. The COLO 357 cell line was a kind gift from Dr Caroline Hill, Francis Crick Institute.

2.1.1.2 Murine cell lines

Cell line	Cell origin	Mouse genotype	Growth medium
DT6606	Pancreas tumour	*LSL-Kras*$^{G12D/+}$; *Pdx-1-Cre* (KC) (68)	DMEM, heat inactivated FBS (10% v/v), penicillin (100units/ml) and streptomycin (100µg/ml)
DT6585	Pancreas tumour		
TB32043	Pancreas tumour	*LSL-Kras*$^{G12D/+}$; *LSL-Trp53*$^{R172H/+}$; *Pdx-1-Cre* (KPC) (72)	

Table 2.2 – List of murine cell lines used. Each cell line was a kind gift from Prof David Tuveson, Cold Spring Harbor Laboratory.

For the initial three passages of murine cells generated from the tumours of *LSL-Kras$^{G12D/+}$; LSL-Trp53$^{R172H/+}$; Pdx-1-Cre; Kmt2d$^{Flox/+}$* and *LSL-Kras$^{G12D/+}$; LSL-Trp53$^{R172H/+}$; Pdx-1-Cre; Kmt2d$^{Flox/Flox}$* mice (see Section 2.1.7) were cultured using 'PDAC media' prepared and sterilised using a 0.22µm vacuum filtration system (Millipore, Cat#: SCGPU05RE) before use (Table 2.3). For passages following these initial three DMEM supplemented with FBS (10% v/v), penicillin (100units/ml) and streptomycin (100µg/ml) was used.

Concentration	Regent name	Supplier	Cat #
N/A	DMEM/F12	Thermo Fisher Scientific	11320033
5 mg/ml	D-(+)-glucose	Sigma-Aldrich	G7021
0.1 mg/ml	Trypsin inhibitor I from soybean	Sigma-Aldrich	T9003
0.5 % (v/v)	Insulin-transferrin-selenium	BD Biosciences	354352
25 µg/ml	Bovine pituitary extract	Thermo Fisher Scientific	13028014
20 ng/ml	Epidermal growth factor	BD Biosciences	354001
5 nM	3,3',5 Triiodo-L-thyronine	Sigma-Aldrich	T6397
1 µM	Dexamethasone	Sigma-Aldrich	D4902
100 ng/ml	Cholera toxin B subunit	Sigma-Aldrich	C9903
1.22 mg/ml	Nicotinamide	Sigma-Aldrich	N0636
5% (v/v)	Nu-serum IV culture supplement	BD Biosciences	355104
100units/ml / 100µg/ml	Penicillin/Streptomycin	Sigma Aldrich	P4333

Table 2.3 – Constituents of 'PDAC media' used for the initial passages of cell lines generated from mice tumours.

2.1.2 Buffers, solutions and reagents

Tris-Buffered Saline (TBS) (10x)

1.5M NaCl (Sigma-Aldrich, Cat#: 71380) and 200mM Trizma base (Sigma-Aldrich, Cat#: T1503) in deionised H_2O; pH = 7.6

TBS-Tween (1x)

1:10 dilution of TBS (10x) in deionised H_2O with 0.1% v/v Tween20 (Sigma-Aldrich, Cat#: P7949)

Propidium Iodide (PI) Staining Solution

50µg/ml PI (Sigma, Cat#: P4170) with 100µg/ml RNase A (Qiagen, Cat#: 19101) in Ca^{2+} and Mg^{2+} free Dulbecco's Phosphate Buffered Saline (DPBS) (Sigma Aldrich, Cat#: D8537)

Complete Radioimmunoprecipitation Assay (RIPA) Lysis Buffer

RIPA buffer (Sigma-Aldrich, Cat#: R0278) supplemented with one tablet of protease inhibitor cocktail I (Roche, Cat#: 11836153001) per 10ml, and 1:100 phosphatase inhibitor cocktail II (Sigma-Aldrich, Cat#: P5726)

Isotonic Lysis Buffer

20mM Trizma base pH = 7.5 (1:50 dilution from 1M stock), 100nM NaCl (1:50 dilution from 5M stock), 5mM $MgCl_2$ (1:200 from 1M stock) (Sigma-Aldrich, Cat#: M8266), 10% (v/v) glycerol (1:7.8 from 87% (v/v) stock) (Sigma-Aldrich, Cat#: 10795711), 0.2% (v/v) Nonidet™ P 40 substitute (1:100 from 20% (v/v) stock) (Sigma-Aldrich, Cat#: 74385), 0.5mM DTT (1:1000 from 1M stock) (Roche, Cat#: 10197777001), supplemented with one tablet of protease inhibitor cocktail I per 10ml and 1:100 phosphatase inhibitor cocktail II in deionised H_2O

High-Salt Lysis Buffer

50mM Trizma base pH = 7.5 (1:20 dilution from 1M stock), 600nM NaCl (1:8.3 dilution from 5M stock), 10% (v/v) glycerol (1:7.8 from 87% (v/v) stock), 0.2% (v/v) Nonidet™ P 40 substitute (1:100 from 20% (v/v) stock), 0.5mM DTT (1:1000 from 1M stock), supplemented with one tablet of protease inhibitor cocktail I per 10ml and 1:100 phosphatase inhibitor cocktail II in deionised H_2O

Bicinchoninic Acid (BCA) Assay Reagent

4% (w/v) copper (II) sulphate (Sigma-Aldrich, Cat#: C2284) diluted 1:50 in BCA (Sigma-Aldrich, Cat#: B9643)

Tris-Acetate (TA) Running Buffer (1x)

25ml NuPAGE® TA SDS running buffer (20x) (Thermo Fisher Scientific, Cat#: LA0041) in 475ml of deionised H_2O

3-(N-morpholino)propanesulfonic acid (MOPS) Running Buffer

25ml NuPAGE® MOPS SDS running buffer (20x) (Thermo Fisher Scientific, Cat#: NP0001) in 475ml of deionised H_2O

Equilibration Buffer

25ml NuPAGE® transfer buffer (20x) (Thermo Fisher Scientific, Cat#: NP0006), 500µl NuPAGE® antioxidant (Thermo Fisher Scientific, Cat#: NP0005), and 50ml methanol (Thermo Fisher Scientific, Cat#: 11976961) in 425ml of deionised H_2O

Blocking buffer

5% (w/v) BSA, or 5% (w/v) non-fat dry milk (Marvel), in 1x TBST

DNA extraction buffer

100mM NaCl, 50mM Trizma base, 25mM Ethylenediaminetetraacetic acid (EDTA) and 1% (w/v) SDS in deionised H_2O; pH = 8.0

50x TAE buffer

2M Trizma base, 1M acetic acid and 50mM EDTA pH = 8.0

2.1.3 Antibodies

Protein	Antibody	Company, Cat#, Clone
KMT2D	Rabbit anti-KMT2D	Sigma-Aldrich, HPA035977, polyclonal
β-Actin	Mouse anti-β-Actin	Sigma-Aldrich, A1978, clone: AC-15
PTPN14	Rabbit anti-PTPN14	Cell Signalling Technologies, 13808, clone: D5T6Y
Claudin 1	Rabbit anti-Claudin1	Abcam, ab180158, clone: EPR9306
Calumenin	Rabbit anti-Calumenin	Abcam, ab137019, clone: EPR9075
pro-MET & c-MET	Rabbit anti-MET	Thermo Fisher Scientific, 18-2257, clone: CVD13
NCAPD3	Rabbit anti-CAP-D3	Bethyl Laboratories Inc., A300-604A, polyclonal
ABCB1	Rabbit anti-ABCB1	Abcam, ab170904, clone: EPR10364-57
H3K4me3	Rabbit anti-H3K4me3	Cell Signalling Technologies, 9751, clone: C42D8
H3K4me2	Rabbit anti-H3K4me2	Cell Signalling Technologies, 9725, clone: C64G9
H3K4me1	Rabbit anti-H3K4me1	Cell Signalling Technologies, 5326, clone: D1A9
H3	Rabbit anti-H3	Cell Signalling Technologies, 9715, polyclonal
Rabbit IgG	Donkey anti-rabbit IgG	GE Healthcare, NA9340
Mouse IgG	Sheep anti-mouse IgG	GE Healthcare, NXA931

Table 2.4 – List of antibodies used for western blots. For each protein and targeted antibody the species, company, catalogue number (Cat #), and clone are described.

2.1.4 Small interfering RNA (siRNA) oligonucleotides

Before use, all lyophilised siRNAs were first centrifuged and reconstituted with nuclease-free H_2O (Thermo Fisher Scientific, Cat#: AM9937) to obtain stocks of 50µM (50pmole/µl) siRNA.

Target Gene	**siRNA ID (Cat #)**	**Alias**	**Exons targeted**	**Sequence (5'->3') (Sense); (Antisense)**	**%CG (Sense); (Antisense)**	**Length (bp)**	**Locus (Human hg19)**
N/A	4390844	Ctrl siRNA	N/A	Unknown	Unknown	Unknown	N/A
KMT2D	s15606	siRNA1	48	GCAAAUCGCUAGCAUCAUUtt; AAUGAUGCUAGCGAUUUGCtt	38%; 38%	21	chr12: 49,420,242-49,420,260
KMT2D	s15605	siRNA2	39	GAGUCGAACUUUACUGUCUtt; AGACAGUAAAGUUGCACUCtg	38%; 43%	21	chr12: 49,426,530-49,426,548
KMT2C	s33890	siRNA1	38	CCAAUGAGGUAAAAACGGAtt; UCCGUUUUUACCUCAUUGGta	38%; 38%	21	chr7: 151,874,064-151,874,082
KMT2C	s33889	siRNA2	37	CUCCUGAUGUUGCUAGUAUtt; AUACUAGCAACAUCAGGAGga	38%; 43%	21	chr7: 151,876,943-151,876,961
KMT2C	s33888	siRNA3	38	GCUCAUUCCAGGUCAAUCAtt; UGAUUGACCUGGAAUGAGCcc	43%; 52%	21	chr7: 151,873,544-151,873,562
KMT2C	s55099	siRNA4	18, 19	UCAGUAUUAAGAUCACUAAtt; UUAGUGAUCUUAAUACUGAca	14%; 52%	21	chr7: 151,921,694-151,921,701, 151,927,000-151,927,018

Table 2.5 – List of siRNAs used for human cell transfection. For each siRNA the targeted gene name, catalogue number (Cat #), in-house alias, targeted exons, sequences, percentage content of cytosine and guanine (%CG), base pair length (bp) and target locus (on hg19) are included. All siRNAs were purchased from Thermo Fisher Scientific.

Target Gene	**siRNA ID (Cat #)**	**Alias**	**Exons targeted**	**Sequence (5'->3') (Sense); (Antisense)**	**%CG (Sense); (Antisense)**	**Length (bp)**	**Locus (Mouse mm10)**
N/A	4390844	Ctrl siRNA	N/A	Unknown	Unknown	Unknown	N/A
Kmt2d	s117455	siRNA1	54, 55	GACCUAUGACUAUCAGUUUtt; AAACUGAUAGUCAUAGGUCag	33%; 38%	21	chr15: 98,834,572-98,834,590
Kmt2d	s117456	siRNA2	40	GGGUACUUCCAGCAACAAAtt; UUUGUUGCUGGAAGUACCCtg	43%; 48%	21	chr15: 98,843,598-98,843,616
Kmt2d	s117457	siRNA3	46, 47	CCUGAACGGUGUCAUGGUAtt; UACCAUGACACCGUUCAGGtt	48%; 48%	21	chr15: 98,840,874-98,840,892
Kmt2c	s233052	siRNA1	38	GCCUCCUACUUUAAUAGUAtt; UACUAUUAAAGUAGGAGGCaa	33%; 33%	21	chr5: 25,309,979-25,309,997
Kmt2c	s233053	siRNA2	52	GCCUGAUGAGCUUAAGGUAtt; UACCUUAAGCUCAUCAGGCtt	43%; 43%	21	chr5: 25,284,980-25,284,998

Table 2.6 – List of siRNAs used for murine cell transfection. For each siRNA the targeted gene name, catalogue number (Cat #), in-house alias, targeted exons, sequences, percentage content of cytosine and guanine (%CG), base pair length (bp) and target locus (on mm10) are included. All siRNAs were purchased from Thermo Fisher Scientific.

2.1.5 Genotyping polymerase chain reaction (PCR) primers

To genotype mice and cell lines primers for PCR amplification of targeted DNA fragments were purchased from Sigma-Aldrich. For *Cre*, two primer pairs were used in one reaction with one pair targeted to an endogenous control gene and the other to *Cre*. For *Kras*, two primer pairs were used in two reactions with one pair targeted to the WT gene and the other to *LSL-Kras*^{G12D}. For *Trp53*, three primers were used in one reaction with a pair targeted to the WT gene and a single forward primer targeted to *LSL-Trp53*^{R172H}. For *Kmt2d*, two primers were used in one reaction, with the pair generating different size products for the presence of LoxP sites. Details of the primers can be found in Table 2.7.

Target gene	Primers		Agarose gel % (w/v)
	Name	Sequence (5'->3')	
Cre	Cre F	GCGGTCTGGCAGTAAAAACTATC	3%
	Cre R	GTGAAACAGCATTGCTGTCACTT	
	Internal Ctrl F	CTAGGCCACAGAATTGAAAGATCT	
	Internal Ctrl R	GTAGGTGGAAATTCTAGCATCATCC	
Kras Mutant	LSL Kras F	CCATGGCTTGAGTAAGTCTGC	3%
	WT Kras R	CGCAGACTGTAGAGCAGCG	
Kras WT	WT Kras F	GTCGACAAGCTCATGCGGG	
	WT Kras R	CGCAGACTGTAGAGCAGCG	
Trp53	WT Trp53 F	TTACACATCCAGCCTCTGTGG	3%
	WT Trp53 R	CTTGGAGACATAGCCACACTG	
	LSL Trp53 F	AGCTAGCCACCATGGCTTGAGTAAGTCTGCA	
Kmt2d	Kmt2d F	ACAGCCAGAAGCCGCCTG	1.5%
	Kmt2d R	AGGTTAGCCACTAGCCCTTCC	

Table 2.7 - Primers for Genotyping PCR. For each primer the targeted gene name, primer name, sequences, and agarose gel percentage for electrophoresis are included. F – Forward, R – Reverse. All primers were purchased from Sigma-Aldrich.

2.1.6 RT-qPCR probes

For duplex quantitative real-time PCRs (qRT-PCR), TaqMan® gene expression assays were purchased from Thermo Fisher Sciences. In each experiment the duplex reactions contained a FAM™/MGB probe primer for the gene of interest and an 18S ribosomal RNA (rRNA) endogenous control VIC®/TAMRA™ probe primer. Details of the probe primers, their targets and the product can be found in Table 2.8.

Target Gene	Assay ID (Cat #)	Exons targeted	Amplicon length	Fluorophore-Quencher
KMT2C	Hs01005521_m1	39-40	58	FAM-MGB
KMT2D	Hs00912416_m1	42-43	65	FAM-MGB
18S	4310893E	N/A	187	VIC-TAMRA
Kmt2c	Mm01156942_m1	39-40	53	FAM-MGB

Table 2.8 – List of TaqMan® gene expression assay probes used. For each probe the target gene name, assay ID (Cat #), fluorophore and quencher are included. Details for the exons and transcripts targeted and amplicon length are also included. All probes were obtained from Thermo Fisher Sciences.

2.1.7 Mice

All mice were bred and maintained in a pathogen-free environment at the Biological Services Unit, Barts Cancer Institute, Queen Mary University of London. Mice were housed according to institutional welfare guidelines under the authority of the UK Home Office Project Licence (PPL 70/7411) and the Animals (Scientific Procedures) Act 1986. Experimental protocols and procedures were performed under the personal licence PIL 30/8960. *LSL-Kras*$^{G12D/+}$; *LSL-Trp53*$^{R172H/+}$ (KP) and homozygous *Pdx-1-Cre* mice strains were both a kind gift from Professor Dave Tuveson (Cold Spring Harbor Laboratory, USA). *Kmt2d*$^{Flox/Flox}$ mice were a kind gift from Professor Francis Stewart (Bitechnology Center, Technische Universitat Dresden, Germany). These three strains were interbred to generate compound mutant mice on a 129/SvJae/C57Bl/6 mixed background.

For survival, mice were maintained for up to 8 months and culled when welfare endpoints were met according to the institutional guidelines where all were sacrificed due to PDAC disease burden. For each mouse the primary and any metastatic tumours were dissected (with 4mm^3 portions of tumour taken aside where required to generate cell lines) and snap frozen in liquid nitrogen alongside other organ tissue samples.

2.1.7.1 Generation of cell lines from murine tumours

Roughly 4mm^3 portions of tissue taken from the tumours of mice were placed in a petri dish, washed with 10ml of cold DPBS, and finely diced using a scalpel. Tumour pieces were collected into a 15ml tube using 5ml of 1mg/ml collagenase V (Sigma-Aldrich, Cat#: C9263) diluted in DMEM/F12 and incubated on a shaker at 37°C for 45min. Sample was centrifuged at 214 x *g* for 5min before supernatant was removed and pellet was resuspended in 3ml of 0.05% (v/v) trypsin at 37°C for 5min. To each sample 7ml of DMEM supplemented with 10% (v/v) heat inactivated FBS and 96µM $CaCl_2$ was added before being washed three times by centrifugation at 214 x *g* for 5min and the addition of 7ml of DMEM/F12. After the final centrifugation at 214 x *g* for 5min cells and remaining tumour homogenate were resuspended in 2ml 'PDAC media' (Table 2.3) and plated in 60mm rat collagen I BioCoat™ culture dishes (VWR, Cat#: 734-0275). Cells were then cultured and maintained in Primaria™ 75cm² rectangular straight neck cell culture flasks (BD Biosciences, Cat#: 353810) and 'PDAC media' for three passages as described in Section 2.2 before being cultured and maintained in DMEM as described in Section 2.2.

2.2 Cell culture

2.2.1 Cell line maintenance

All cell lines were routinely cultured in a humidified incubator at 37°C with 5% (v/v) CO_2. Cells were normally cultured in 175cm^2 tissue culture flasks with 35ml of the appropriate growth media used for each cell line described in Table 2.1 and Table 2.1.

2.2.2 Subculture of cell lines

All of the human and murine cell lines were adherent in nature and were therefore subcultured once they reached 80% confluence. Growth medium was removed from the cell monolayer using a vacuum aspirator and the cells were washed twice with DPBS. Cells were then detached from the culture flask with 3ml of 0.1% (v/v) Ca^{2+} and Mg^{2+} free gamma irradiated porcine trypsin with 0.04% (v/v) EDTA (trypsin) (Sigma Aldrich, Cat#: 59418C) at 37°C. Once detached, trypsin was neutralised with fresh growth medium containing heat inactivated FBS to achieve an appropriate dilution (ranged from 1:10 to 1:20 depending on each cell line's growth rate) before adding to 1ml to a new flask with fresh medium.

2.2.3 Storage of cell lines in liquid nitrogen

For storage, cell line stocks were frozen and stored in liquid nitrogen. For each cell line, cells were centrifuged at 180 x *g* for 5mins and resuspended in 10% (v/v) Dimethyl sulfoxide (DMSO) in heat inactivated FBS to get an approximate concentration of 5x10^6 cells/ml and 1ml added to cryovials. These cryovials were then placed in a freezing container (Mr. Frosty, Nalgene, Sigma-Aldrich) and stored at -80°C overnight before being transferred to liquid nitrogen tanks.

2.2.4 Recovery of cell lines from liquid nitrogen

To recover frozen cell lines, the cryovials were rapidly defrosted by placing in a water bath pre-warmed to 37°C until the icy cell pellet was loose enough within the cryovial to be transferred to 30ml of pre-warmed complete culture medium. After the cells were totally thawed they were pelleted by centrifugation at 180 x *g* for 5min. Supernatant containing the freezing medium was aspirated before the pellet was resuspended in fresh complete growth medium and transferred to a 175cm^2 tissue culture flask for culture as described in 2.2.1.

2.3 RNA interference

Prior to the day of transfection, each cell line was harvested using trypsin and plated overnight in 6-well plates with 2.5ml of appropriate growth medium, free from penicillin and streptomycin. Cells were then "forward transfected" by adding 500µl of RNA-lipid complexes to each well to give a final volume of 3ml. RNA-lipid transfection complexes were formed by combing Silencer® Select siRNAs (Thermo Fisher Scientific) with Lipofectamine RNAiMAX (Thermo Fisher Scientific, Cat#: 13778) in Opti-MEM I medium (Thermo Fisher Scientific, Cat#: 11058-021) for 15min at room temperature (RT). Cells were incubated for 48 hours before beginning the experiments. In all experiments, protein was harvested and loss of KMT2D or Kmt2d was confirmed by western blot, as described in 2.5.1.

2.3.1 Human cell transfection

To take into account the differing growth rates of the human cell lines, the cells were initially seeded in 6-well plates with 2.5ml of appropriate growth medium, free from penicillin and streptomycin, at either $5x10^4$ cells/well (SUIT-2) and $8x10^4$ cells/well (RWP1, CFPAC1, HPDE, BxPC-3, COLO 357, PANC-1 and Capan-2) for growth and cell cycle experiments, or $1x10^5$ cells/well (SUIT-2) and $1.6x10^5$ cells/well for RNA sequencing (RNA-seq) and western blot experiments. For the transfection, the final RNA-lipid complex concentration of 8.33nM per well was achieved by combining 25pmol of each siRNA with 5µl Lipofectamine RNAiMAX in 500µl Opti-MEM I medium. These RNA-lipid complexes were added to each well containing cells and incubated for 48 hours.

2.3.1.1 Dual siRNA transfection

PANC-1 cells were seeded at $8x10^4$ cells/well overnight in 2.5ml of appropriate growth medium, free from penicillin and streptomycin and transfection performed as described in 2.3.1. To form the RNA-lipid transfection complexes used to simultaneously transfect cells with two siRNAs, 25pmol of each siRNA was combined with Lipofectamine RNAiMAX, in Opti-MEM I medium, for 15min at RT. These RNA-lipid complexes were added to each well containing cells and incubated for 48 hours.

2.3.2 Murine cell transfection

To transfect the DT6606, DT6585 and TB32043 cell lines, cells were initially seeded at $4x10^4$ cells/well in 2.5ml of DMEM, free from penicillin and streptomycin. The final RNA-lipid complex concentration of 50nM per well for transfection was achieved by combining 150pmol of each siRNA with 5µl Lipofectamine RNAiMAX in 500µl Opti-

MEM I medium. These RNA-lipid complexes were added to each well containing cells and incubated for 48 hours.

2.4 *In vitro* assays

2.4.1 Cell proliferation assay

Following the 48-hour transfections described in 2.3.1, cells were washed with DPBS and 3ml of fresh penicillin and streptomycin free media were added. Cells were then cultured for a further 72 hours (120 hours total) before being detached with trypsin and counted using a Vi-Cell XR automated cell viability analyser (Beckman Coulter). For the dual KMT2C and KMT2D experiment, this cell counting was also performed every 24 hours.

2.4.2 Chemotherapeutic concentration-response assay

Following the 48-hour transfections described in 2.3.2, DT6606, DT6885 and TB32043 cells were washed with DPBS and detached with trypsin. Cells were then re-plated in 96-well plates at $1x10^4$ cells/well (DT6585 and TB32043) or $5x10^3$ cells/well (DT6606), to take into account their differing growth rates. Following adherence overnight, medium in each well was replaced with 100µl DMEM with different concentrations of 5-FU (Accord Healthcare) or Gemcitabine (Hospira). After 72 hours incubation with the drugs, 10µl WST-1 reagent (Roche, Cat#: 11644807001) was added to each well and the optical density (OD) measured at 440nm (reference 630nm) after three hours. To generate log-concentration response curves, the percentage of cell viability was calculated using the maximal OD as representing 100% viability.

2.4.3 Flow-cytometric analysis of cell-cycle

For cell cycle analysis, cells were first seeded in triplicate wells and transfected as described in 2.3.1. At each experimental time point, cells were washed with DPBS, detached with trypsin and pelleted by centrifugation at 180 x *g* and 4°C for 5min before being permeabilised in 1ml ice-cold 70% (v/v) ethanol (Thermo Fisher Scientific, Cat#: 10542382). Permeabilised cells were stored at -80°C until all samples in an experiment were collected. Permeabilised cells were pelleted by centrifugation at 16000 x *g* and 4°C for 5min, washed with ice-cold DPBS and stained with propidium iodide (PI) staining solution for 15min at 37°C. The YG610/20 filter on the Fortessa II flow cytometer was used to examine PI staining of DNA in 30,000 cells.

2.4.3.1 Nocodazole cell-cycle block and release assay

Cell cycle profiles for the eight cell lines were examined by flow-cytometric analysis as described in 2.4.3 at three experimental time points. For each siRNA, the first time point was at 48 hours post-transfection. The next time point, performed on cells in parallel wells, was following a 16-hour incubation with 400ng/ml nocodazole (Sigma, M1404). The final time point, also performed on cells in parallel wells, was at 24 hours after the nocodazole containing medium was replaced with fresh complete culture medium.

2.5 Molecular biology

2.5.1 Western blot

2.5.1.1 Cell lysis

2.5.1.1.1 Whole cell

To prepare whole cell protein lysates, culture medium was aspirated and cells were washed twice with ice cold DPBS. To each well, 80µl of complete RIPA cell lysis buffer was added for 30min on ice. Lysates were harvested with cell scrapers and cellular debris removed by centrifugation at 16000 x *g* and 4°C for 30min. Samples were stored at -20°C.

2.5.1.1.2 Nuclear fraction

To prepare nuclear fraction lysates, medium was aspirated and cells were washed twice with ice cold DPBS. Cells were then detached from the wells by incubating with 500µl of trypsin at 37°C, before being neutralised with 1ml fresh growth medium containing 10% (v/v) heat inactivated FBS. Cells were pelleted by centrifugation at 180 x *g* for 5min, resuspended in 1ml DPBS, and again centrifuged at 180 x *g* for 5min. DPBS was aspirated from the cell pellet and 200µl of isotonic lysis buffer was added. To facilitate lysis, samples were incubated for 10min on ice and homogenised ten times with a P1000 pipette and tip for three cycles. Efficiency of nuclei extraction was confirmed by trypan blue staining. Nuclei were pelleted by centrifugation at 16000 x *g* for 20min and the supernatant containing the cytosolic fraction was retained and stored at -20°C.

To the extracted nuclei, 80µl of high-salt lysis buffer was added for 30min on ice. Samples were sonicated using a Bioruptor® Plus sonication device (Diagenode, Cat#: B01020002) on the low power setting for six cycles of 15sec on and 60sec off. Samples

were centrifuged at 16000 x *g* for 5min to remove cellular debris, and the nuclear fraction supernatant was stored at -20°C.

2.5.1.2 Protein quantification

Protein quantification was determined by BCA assay. In a 96-well plate, 200µl of BCA reagent was added to 10µl of samples, or bovine serum albumin (BSA) (Sigma-Aldrich, Cat#: A4503) standards, diluted in DPBS. Plates were incubated at 37°C for 30min and the OD measured at 595nm. Sample protein concentrations were calculated by interpolation from the standard curve generated.

2.5.1.3 Sample preparation, gel electrophoresis and protein transfer

Samples were prepared for loading by adding 5µl NuPAGE LDS sample buffer (4x) (Thermo Fisher Scientific, Cat#: NP0008), 2µl NuPAGE sample reducing agent (10x) (Thermo Fisher Scientific, Cat#: NP0009), and 13µl protein sample diluted in DPBS to achieve equal amounts of protein. In general, a minimum of 20µg of whole cell lysates, and 10µg of nuclear lysates, were prepared. Prepared samples were then heated at 70°C for 10min before loading on the gel.

Gel electrophoresis was performed using an XCell SureLock mini-cell (Thermo Fisher Scientific, Cat#: EI0001) system. Depending on the protein under investigation, and thus the gel used, either TA (for NuPAGE® Novex® 3-8% TA gels (Thermo Fisher Scientific, Cat#: EA03785BOX)), or MOPS (for NuPAGE® Novex® 4-12% Bis-Tris gels (Thermo Fisher Scientific, Cat#: NP0336BOX)) running buffers were prepared. For the cell's inner chamber, 500µl NuPAGE® antioxidant (Thermo Fisher Scientific, Cat#: NP0005) was added to 200ml of the appropriate 1x running buffer, and the remaining 300ml of running buffer was used for the cell's outer chamber. Gels were run at 150V for 1 hour with a HiMark™ pre-stained protein standard (Thermo Fisher Scientific, Cat#: LC5699) used to follow the course of the gel electrophoresis, and later for analysis.

For large molecular weight proteins, 3-8% TA gels were left for 15min in equilibration buffer before protein transfer. Samples were transferred onto either pre-activated polyvinylidene fluoride (PVDF), or nitrocellulose membranes by dry transfer using mini iBlot® transfer stacks (Thermo Fisher Scientific, Cat#: IB401002 and IB301002, respectively) and the iBlot® gel transfer device (Thermo Fisher Scientific, Cat#: IB1001UK) on program three for 10min.

2.5.1.4 Blocking and probing membrane

Membranes were placed in the appropriate blocking buffer (Table 2.9) for an hour at RT and probed with primary antibodies (diluted in blocking buffer according to Table 2.9) overnight at 4°C. The next day, membranes were washed with 1x TBST (3 x 10min) and incubated with anti-rabbit, or anti-mouse, horseradish peroxidase-conjugated secondary antibodies, diluted in the appropriate blocking buffer (Table 2.9) for an hour at RT.

Protein	Primary staining solution			Secondary staining solution		
	Antibody	Dilution	Blocking buffer	Antibody	Dilution	Blocking buffer
KMT2D	Rabbit anti-KMT2D	1:1000	5% w/v non-fat milk; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v non-fat milk; 1x TBST
β-actin	Mouse anti-β-Actin	1:5000	5% w/v non-fat milk; 1x TBST	Sheep anti-mouse IgG	1:5000	5% w/v non-fat milk; 1x TBST
PTPN14	Rabbit anti-PTPN14	1:1000	5% w/v non-fat milk; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v non-fat milk; 1x TBST
Claudin 1	Rabbit anti-Claudin1	1:2000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
Calumenin	Rabbit anti-Calumenin	1:1000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
pro-MET & c-MET	Rabbit anti-MET	1:1000	3% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
NCAPD3	Rabbit anti-CAP-D3	1:2000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
ABCB1	Rabbit anti-ABCB1	1:1000	5% w/v non-fat milk; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v non-fat milk; 1x TBST
H3K4me3	Rabbit anti-H3K4me3	1:1000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
H3K4me2	Rabbit anti-H3K4me2	1:1000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
H3K4me1	Rabbit anti-H3K4me1	1:1000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST
H3	Rabbit anti-H3	1:1000	5% w/v BSA; 1x TBST	Donkey anti-rabbit IgG	1:2000	5% w/v BSA; 1x TBST

Table 2.9 – List of staining conditions used for western blots. For detection of each protein the details of both the primary and secondary staining solutions are included. For each staining solution details of the antibody (see **Table 2.4**), dilution, and the blocking buffer used are included.

2.5.1.5 Detection and visualisation

Before detection, membranes were first washed with 1x TBST (3 x 10min). Staining was then visualised by incubation of the membrane with Amersham enhanced chemiluminescence (ECL), or ECL prime, western blot detection reagents (GE Healthcare, Cat#: RPN2106 and RPN2232 respectively), and developed using Super Rx X-ray films (Fujifilm) and a Konica Minolta SRX-101A medical film processor.

For each membrane, and where required, the equivalence of protein loading (β-Actin), and total protein (*e.g.* Total H3), was confirmed by stripping and reprobing the membranes.

For stripping the membranes, ReBlot Plus Strong Antibody stripping solution (10x) (Millipore, Cat#: 2504) was diluted 1:10 in deionised H_2O. Membranes were then placed, with the protein coated side facing inwards, in a 50ml tube containing 10ml of 1x ReBlot solution. Membranes were then incubated on a roller at RT for five, ten, or 15min, depending on whether it is the first, second, or third time being stripped. Membranes were then removed from the 1x ReBlot solution and first placed in the appropriate blocking buffer (Table 2.9) for 5min. Membranes were then removed and re-blocked and probed as described in 2.5.1.4.

2.5.2 RNA extraction

To extract RNA from cells, the culture medium was first aspirated and cells washed twice with ice cold DPBS. After complete removal of DPBS, 350µl of the guanidine-thiocyanate–containing buffer RLT (Qiagen, Cat#: 74104) with 1:100 β-mercaptoethanol (Sigma-Aldrich, Cat#: M6250) was added to the cells in each well, whilst on ice. Cells and lysates were immediately removed from wells using a cell scraper and pipette to transfer into a microcentrifuge tube. Each sample lysate was homogenised by pulse vortex mixing and snap frozen with dry ice before being stored at -80°C.

Total RNA species of greater than 200 nucleotides were isolated from cells using the RNeasy Mini Kit (Qiagen, Cat#: 74104) with an on-column DNase digestion (Qiagen, Cat#: 79254) according to the manufacturer's instructions. Briefly, lysates were homogenised by centrifugation at 16000 x *g* through QIAshredder (Qiagen, Cat#: 79654) columns before an equal volume of 70% (v/v) ethanol (analytical grade ethanol (Sigma-Aldrich, Cat#: E7023)) was added. Homogenised lysates were then placed in RNeasy columns and centrifuged at 16000 x *g* for 5sec so the membrane bound total

RNA, whilst other cell debris passed through. To effectively remove other contaminants, such as genomic DNA, the RNeasy columns were washed with 350µl of buffer RW1 and centrifuged at 16000 x *g* for 15sec, before being treated with 80µl (27.3 Kunitz units) of DNase I (10µl DNase I diluted with 70µl of buffer RDD) for 15min at RT. Following this, the columns were again washed with 350µl of buffer RW1 and centrifuged at 16000 x *g* for 15sec. Columns were then washed a further two times using 500µl of buffer RPE (made up with analytical grade ethanol) and centrifugation at 16000 x *g* first for 15sec, and then for 2min, before centrifugation at 16000 x *g* for 1min in an empty collection tube to eliminate carryover of buffer RPE. RNA was finally eluted into 30µl of nuclease-free H_2O by centrifugation at 16000 x *g* for 1min and quantified using a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific).

2.5.3 cDNA synthesis

Single-stranded cDNA was synthesised from total RNA using the High Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific, Cat#: 4368814) according to the manufacturer's instructions. Briefly, for each PCR the RNA samples were diluted with RNA/RNase-free H_2O to get 14.2µl containing 1µg RNA. To get a final reaction volume of 20µl for each sample, 5.8µl of 2x reverse transcription master mix was prepared and added to PCR tubes on ice. The 2x reverse transcription master mix was prepared with 2µl of 10x RT buffer, 0.8µl of 25x dNTP mix (100mM), 2µl of 10x reverse transcription random primers and 1µl of MultiScribe™ reverse transcriptase.

The PCRs were performed using a BioRad thermal cycler and the conditions optimised by the kit's manufacturer (25°C for 10min, 37°C for 120min, 85°C for 10min). Following the completion of the PCR, samples were diluted 1:50 with RNA/RNase-free H_2O to get a final concentration of 8ng of starting RNA/µl.

2.5.4 Quantitative Real-time PCR (qRT-PCR) analysis

Duplex qRT-PCRs were performed on a StepOnePlus™ Real Time PCR system (Thermo Fisher Scientific, 4376600). For each experiment, iTaq™ Universal Probes Supermix (Bio-Rad, Cat#: 1725134) was used according to manufacturer's instructions, with TaqMan® gene expression assays (Thermo Fisher Scientific) for the genes of interest, and an 18S ribosomal RNA (rRNA) endogenous control. Briefly, 8µl of the cDNA samples diluted in H_2O, 8µl of the diluent H_2O, and 8µl of the diluent H_2O following a cDNA synthesis, were separately added in triplicates to 96-well qRT-PCR plates. To each well, 12µl of 2x qRT-PCR master mix was added to get a final reaction volume of 20µl. The 2x qRT-PCR master mix was comprised of 10µl 2x iTaq™

Universal Probes Supermix, 1µl 18S rRNA VIC®/TAMRA™ probe primer, and 1µl target gene FAM™/MGB probe primers (described in 2.1.6 and Table 2.8) per reaction. Plates were sealed and centrifuged briefly before being run in 40 cycles of 95°C for 15sec and 60°C for 60sec.

Data were analysed using the $2^{-\Delta\Delta Ct}$ method (215) to examine relative expression of each gene, where mRNA expression levels were normalised to 18s and then expressed relative to the normalised mRNA expression levels of the control treatment.

2.5.5 Genotyping PCRs

2.5.5.1 DNA extraction

To extract DNA, 360µl of DNA extraction buffer with 15µl Proteinase K (10mg/ml) (Sigma-Aldrich, Cat#: P2308) was added to mouse ear snips, or cell pellets, in a microcentrifuge tube and heated at 55°C for greater than 2 hours. Samples were vortex mixed for 10sec, centrifuged at 16000 x *g* and RT for 30sec and 125µl 5M NaCl. Samples were again vortex mixed for 10sec before being centrifuged at 16000 x *g* and RT for 5min. To a new microcentrifuge tube 375µl of the middle phase (taking care not to disturb the pellet, or collect the top phase) was added to 205µl isopropanol. Samples were again vortex mixed for 10sec before being centrifuged at 16000 x *g* and RT for 10min. Supernatant was discarded and pellet was washed with 500µl of 70% (v/v) ethanol before being centrifuged at 16000 x *g* and RT for 5min. Supernatant was again removed and pellet was left to air dry for 20min. Pellets were resuspended in 100µl nuclease-free H_2O and stored at -20°C.

2.5.5.2 PCRs

PCR amplification of targeted DNA fragments for were set up for each reaction in 200µl PCR tubes using 2µl of extracted DNA, 0.4µl dNTPs (10mM of each) (Qiagen, Cat#: 201912), 0.1µl HotStarTaq Plus polymerase (Qiagen, Cat#: 203601), 2µl 10x CoralLoad PCR buffer (Qiagen, Cat#: 203601), PCR primers for the gene(s) targeted (See Table 2.7) and nuclease-free H_2O up to 20µl final reaction volume. Each PCR reaction was performed using a BioRad thermal cycler and the conditions specified for each reaction (Table 2.10).

2.5.5.3 Agarose gel electrophoresis

PCR products were run on agarose gels of different concentrations depending on the product (Table 2.7). Agarose gels were prepared by dissolving different amounts of agarose (Bioline, Cat#: BIO-41025) in 100µl 1x TAE solution by heating in a

microwave. Agarose gel solution was allowed to cool until hand-hot before 10µl GelRed™ DNA dye (VWR, Cat#: 41003) was added, and the gel was swirled and poured into a gel cassette with comb. Gels were allowed to set at RT before adding to gel tank, removing the comb and 1xTAE buffer added. PCR products and 5µl Hyperladder™ I (Bioline, Cat#: BIO-33026) size marker were loaded into each well. Gels were run at 90V for 30min before being visualised under ultraviolet light.

Cycling programme	Cre	Kras	Trp53	Kmt2d
Initial activation step	5min, 95°C	5min, 95°C	5min, 95°C	5min, 95°C
Denaturation	30sec, 94°C	30sec, 94°C	30sec, 94°C	30sec, 94°C
Annealing	1min, 51.7°C	30sec, 62°C	30sec, 61°C	30sec, 64°C
Extension	1min, 72°C	1min, 72°C	1min, 72°C	2min, 72°C
Number of cycles	35	39	35	30
Final extension	2min, 72°C	7min, 72°C	2min, 72°C	10min, 72°C

Table 2.10 – Reaction conditions used for each gene

2.6 RNA-seq and bioinformatic analysis

2.6.1 Sample preparation, library preparation and RNA-seq

Material for RNA-seq was generated as described in 2.5.2 following a transfection performed as described in 2.3.1. The total RNA isolated was quantified and its purity checked using a nanodrop spectrophotometer before being sent to Source BioScience for RNA-seq.

At Source BioScience, the RNA was again quantified and its integrity measured, this time using an RNA 6000 nano assay kit (Agilent Technologies, Cat#: 5067-1511) on a 2100 BioAnalyser (Agilent Technologies, Cat#: G2939AA). All but one of the samples used in the analysis had an RNA integrity number (RIN) > 9, with this one having a RIN = 7.3 (PANC1 KMT2C siRNA2). A TruSeq stranded total RNA library preparation kit with Ribo-Zero Gold (Illumina, Cat#: MRZG12324) was used to deplete cytoplasmic and mitochondrial rRNA from the RNA samples and prepare cDNA libraries. The libraries prepared for sequencing were next validated on the 2100 BioAnalyser with a DNA 1000 kit (Agilent Technologies, Cat#: 5067-1504) and again after randomizing the samples into seven pools with a high sensitivity DNA analysis kit (Agilent Technologies, Cat#: 5067-4626). Paired end sequences (reads) of 100 bp in length were generated using seven lanes of a HiSeq 2000 sequencing system (Illumina).

RNA-seq data were deposited in Gene Expression Omnibus (GEO) under the accession number GSE75327.

2.6.2 Alignment of reads and bioinformatic analysis

Bioinformatic analysis of RNA-seq data was performed at Barts Cancer Institute, Queen Mary University of London by Dr Jun Wang. After FASTQ data quality check using FastQC, raw reads were aligned to the reference genome hg19 using Tophat2 (216). An average of 27.2M aligned paired-end reads (range 19.7-36.4M), corresponding to an average of 75.0% (range 59.2-82.2%) concordant pair alignment rate, were reported.

The number of reads uniquely aligned (mapping quality score $q > 10$) to the exonic region of each gene were counted using HTSeq (217), based on the Ensembl annotation (version 74). *KMT2C* and *KMT2D* siRNA datasets were first analysed independently. Only genes that achieved at least one count per million (CPM) mapped reads in at least three samples were included, leading to 15,912 and 15,818 filtered genes in total for the respective *KMT2C* and *KMT2D* siRNA datasets. These genes were classified into 15 RNA species, with protein-coding transcripts representing 81.8% and 82.3%, respectively. Read counts were further normalised using the conditional quantile normalisation (cqn) method (218), accounting for gene length and GC content. Differential expression analysis was then performed using the edgeR package (219), employing the generalised linear model (GLM) approach, for each siRNA versus its control pairwise comparison, adjusting for baseline differences between the cell lines, with an additive model design as "model.matrix(~cellline+siRNA_treatment)". For each pairwise comparison, the significantly differentially expressed (DE) genes were selected using a double threshold of false discovery rate (FDR) < 0.05 and an absolute fold change of at least two.

Next, common DE genes were identified between different siRNA versus control comparisons within the *KMT2C* and *KMT2D* siRNA datasets. On the basis of log2 fold changes of siRNA treated over control for all filtered genes, Gene Set Enrichment Analysis (GSEA) was performed for each comparison using the GSEA tool to identify the canonical pathways gene sets from the Molecular Signatures Database (MSigDB-C2 v5.0) (220). The gene ontology (GO) biological process enrichment analysis was also performed for DE genes using the PANTHER classification system (221).

2.7 Statistics

2.7.1 Statistical analysis of clinical and gene expression data

Bioinformatic analysis of two large PDAC datasets, the International Cancer Genome Consortium (ICGC) (222) and The Cancer Genome Atlas (TCGA) (44), with both gene expression and clinical follow-up data available, was used by Dr Jun Wang to complete survival analysis. Data on 87 patients from the ICGC dataset were previously compiled and processed (223). Level 3 gene expression data for TCGA dataset was downloaded via TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). Only annotated and confirmed PDAC patients were selected (108 in total) (Table 2.11). RNA-seq by Expectation-Maximization (RSEM) normalised expression data for 20,501 genes were obtained. For each gene, low and high expression groups were determined using the method described by Mihaly *et al.* (224). Unlike arbitrarily using the median cut-off point this approach instead interrogates each percentile of expression between lower and upper quartiles in the Cox regression analysis to determine the best performing percentile threshold.

Survival modelling and Kaplan-Meier (KM) analysis was undertaken using R statistical environment ("survival" package). Overall survival (OS) was defined as time from diagnosis to death, or to the last follow-up date for survivors. Log-rank test was used to calculate the KM p-values. The Cox proportional hazards model was fitted to every gene independently.

Two additional PDAC gene expression profile (GEP) and clinical follow-up datasets, namely "Stratford" and "BCI_Zhang_merged", compiled and processed previously (223), were also included for validation studies.

2.7.2 Statistical analysis of *in vitro* and *in vivo* data

All statistical analysis of *in vitro* and *in vivo* data was performed using the statistical tools provided in GraphPad Prism 5 for Mac. Specifics of the statistical tests used are detailed in the relevant figure legends.

Demographics and characteristics	ICGC	TCGA
Sex		
Male	53 (60.9%)	59 (54.6%)
Female	34 (39.1%)	49 (45.4%)
Age		
Median (IQR)	68 (60-75)	65 (57-73.25)
>=Population median	44	59
<Population median	43	48
Race		
White		93 (86.1%)
Black/African American		5 (4.6%)
Asian		6 (5.6%)
Hispanic/Latino		1 (0.9%)
Unknown		3 (2.8%)
Tumour grade		
1	2 (2.3%)	14 (13.0%)
2	53 (60.9%)	63 (58.3%)
3	30 (34.5%)	30 (27.7%)
4	2 (2.3%)	1 (0.9%)
Tumour Stage		
1		8 (7.4%)
2		93 (86.1%)
3		3 2.8%)
Discrepancy		1 (0.9%)
Maximum tumour dimension		
Median (IQR)		3.5 (2.85-4.5)
NA		9
>=Population median		53
<Population median		45
Response		
Stable disease		5 (4.6%)
Progressive disease		29 (26.9%)
Partial remission		7 (6.5%)
Complete remission		24 (22.2%)
Unknown		27 (25%)
NA		14 (12.9%)
Discrepancy		2 (1.9%)

Table 2.11 - A table describing clinical characteristics of ICGC and TCGA. Included are the details for the patient demographics and tumour characteristics for the samples from the ICGC and the TCGA datasets that were included in the analysis.

Chapter 3 Functional effect of KMT2C and KMT2D depletion in human PDAC

3.1 Background

The KMT2 family of methyltransferases catalyse the addition of methyl groups to a lysine residue on the tail of histone H3 (H3K4) (reviewed in (144)), where these marks are thought to support a local chromatin environment that facilitates gene expression (see 1.2.3.1.1). Inactivation of *KMT2C* and *KMT2D* arises through a combination of gene deletion and/or mutation in PDAC (40,41,44,45,191), and is a commonly recurring feature in several cancer types (reviewed in (119)).

Previously, the increased expression of oncogenes, and reduced expression of tumour suppressor genes, have both been used to predict cancer prognosis (225,226). We therefore hypothesised that inter-patient fluctuations in *KMT2C* and *KMT2D* expression may also impart significant differences in outcome for PDAC patients. In this way, another group previously showed that low expression of *KMT2D*, and *KDM6A* (another KMT2C/D complex member), identifies breast cancer patients with improved survival (194,202). Whilst still to be examined in PDAC, *KMT2C* and *KMT2D* (and *KMT2A*) mutations have recently been shown to identify patients with better outcome (41). In this chapter, to determine whether a similar expression based phenomenon exists in PDAC, we used existing gene expression profile (GEP) data to assess whether *KMT2C* and *KMT2D* expression identifies patients with improved outcome.

Moreover, the biological roles of KMT2C and KMT2D, and how they impact on PDAC, have thus far remained largely unknown. The roles of these methyltransferases appear to be cell type dependent, with both negative and positive effects on cell proliferation having been shown across a variety of cancer types (55,133,135,136,138,151,181,199-203). As *KMT2C* and *KMT2D* aberrations commonly result in LOF, an RNAi approach was therefore used to examine the effects of depleting KMT2C and KMT2D upon proliferation in a range of human pancreatic cell lines.

3.2 Results

3.2.1 Decreased *KMT2C* and *KMT2D* expression correlates with favourable outcome in PDAC

To assess whether expression levels of these methyltransferases are linked with patient outcome, GEP data from the ICGC data series were used to compare clinical features of patients with tumours expressing different levels of *KMT2C* and *KMT2D*. Low levels of *KMT2C* and *KMT2D* expression were independently associated with better overall survival (OS). Patients whose tumours had low expression of *KMT2C* had a median

survival of 15.9 months compared to 9.2 months for patients with high expression (log-rank p = 0.029) (Figure 3.1A). Perhaps more strikingly, differential expression of *KMT2D* was also effective at identifying patients with improved outcome, where patients with low expression had a median survival of 19.9 months compared to 11.8 months for those with high expression (log-rank p = 0.001) (Figure 3.1B).

Due to the similarities between KMT2C and KMT2D, we also examined whether their combined expression correlated with survival. Although less significant than either gene individually, combined low-level expression of *KMT2C* and *KMT2D* also correlated with improved outcome, with a median survival of 15.9 compared to 9.2 months (log-rank p = 0.044) (Figure 3.1C). These studies provided some confidence that changes in expression of these methyltransferases impacted on PDAC biology and patient outcome.

In support of observations from the ICGC data, similar trends were observed for *KMT2C* and combined *KMT2C/D* expression in TCGA dataset, however these results did not reach statistical significance (log-rank p = 0.07 and 0.24 respectively) (Figure 3.2). In addition, *KMT2D* expression also failed to significantly highlight differences in survival within this dataset (log-rank p = 0.072) (Figure 3.2).

Alongside the survival analysis for these high and low expression PDAC patient groups, the clinically annotated details for each group were identified. For the ICGC data only age and tumour grade were available, whereas for the TCGA data, both of these parameters plus tumour stage and primary therapy outcome success were available. Across the ICGC high/low groups the ages and tumour grades were roughly similar, and although not statistically tested there is a suggestion that the *KMT2D* low expressing group is perhaps slightly older than the high expressing group (Table 3.1). Across the TCGA high/low groups the age, tumour grade, stage and maximum dimension were similar, again with the suggestion that the age is slightly increased in the lower expressing groups (also not statistically tested) (Table 3.2). In light of these data multivariate and univariate analysis were performed to examine whether these clinical factors influence the survival observed. For the ICGC dataset *KMT2C* (p = 0.03566, HR = 1.921 (1.045 - 3.531)), *KMT2D* (p = 0.004160, HR = 5.952 (1.7573 - 20.159)) and their combined expression (p = 0.05750, HR = 1.870 (0.9803 - 3.567)) performed very well as prognostic markers independently of both age and tumour grade (Table 3.3). For the TCGA dataset, the trend of low expression associating with better survival still remained for *KMT2C* (p = 0.259437, HR = 1.439 (0.7644 - 2.710)), and combined *KMT2C* and *KMT2D* expression (p = 0.144319, HR = 1.719 (0.8306 - 3.560)),

independently of age, tumour stage, tumour grade and primary therapy outcome success, however neither were significant using a $p<0.05$ cutoff (Table 3.4).

Overall, these observations are broadly consistent with the recent genetic data that showed 12 out of 101 patients with *KMT2C* or *KMT2D* LOF mutations had superior outcome when compared to patients with WT configurations (41).

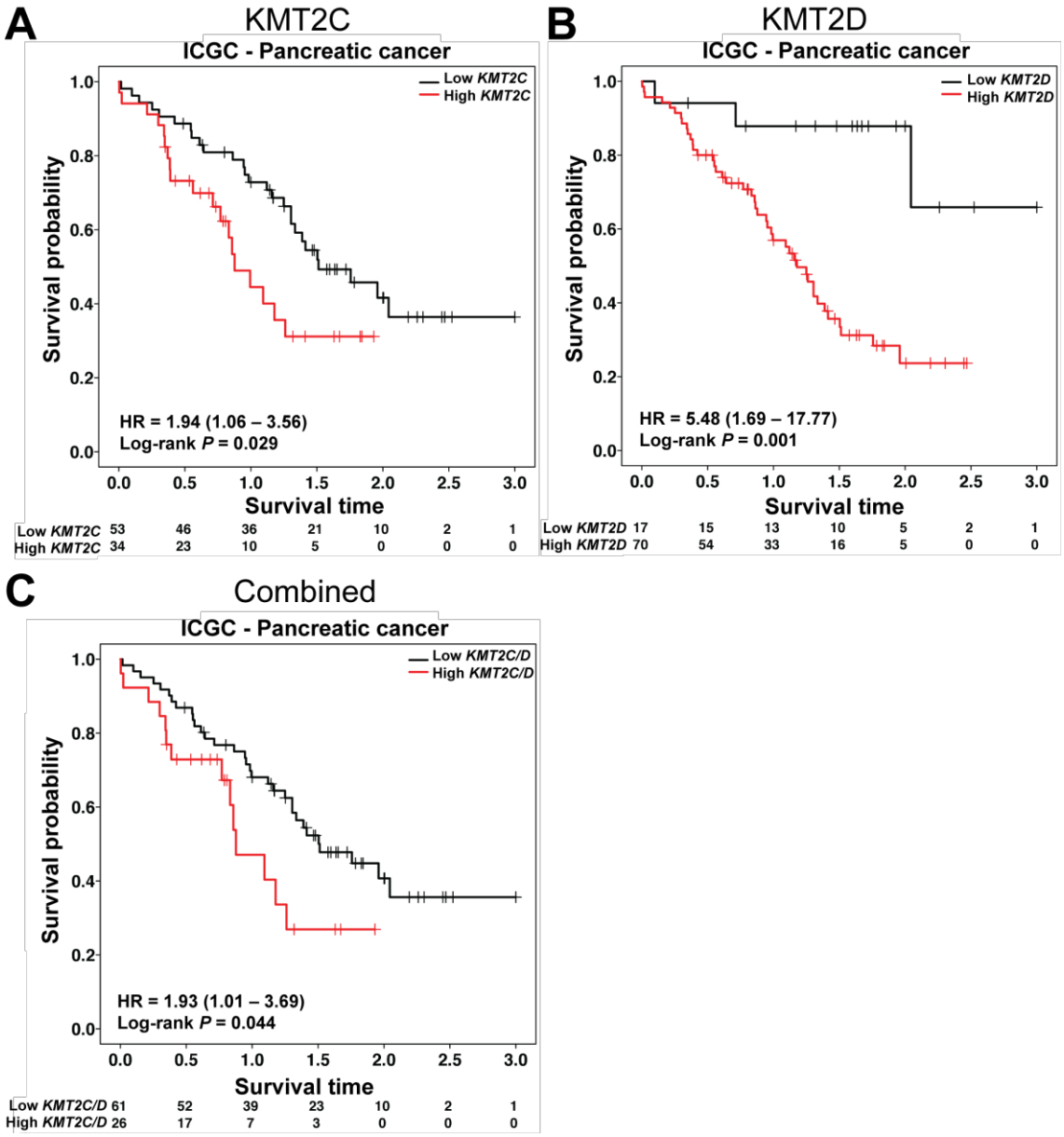

Figure 3.1 – Kaplan-Meier (KM) survival analysis for KMT2C, KMT2D and combined expression in the ICGC dataset. Analysis of PDAC tumours from the ICGC dataset shows that lower KMT2C, KMT2D, and their combined expression correlates with improved patient survival. KM survival analysis graphs for assessing the prognostic value of high (red) and low (black) expression for KMT2C (**A**), KMT2D (**B**), and KMT2C/D combined (**C**) within the ICGC dataset. Numbers on the x-axis represent years.

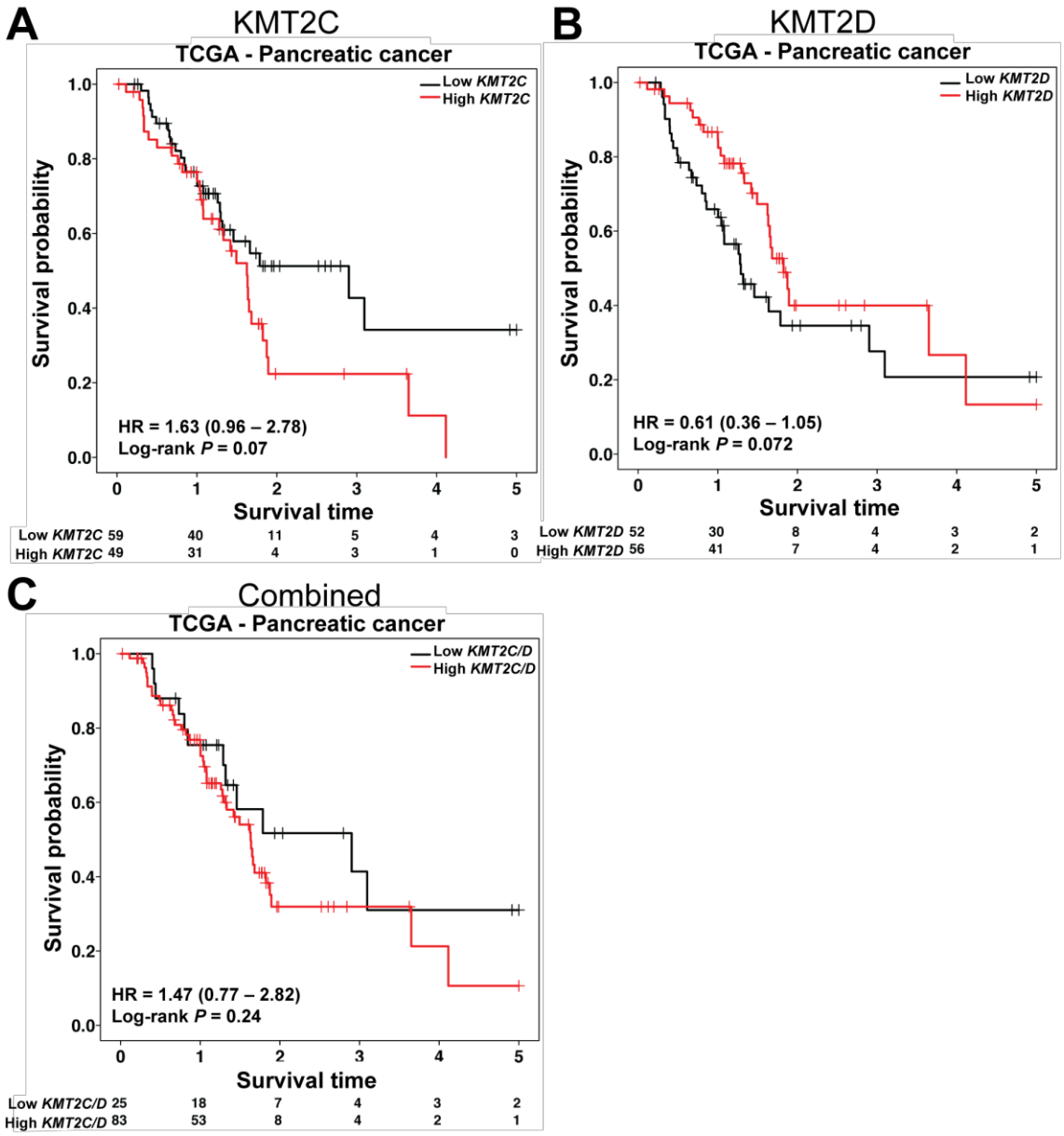

Figure 3.2 - Kaplan-Meier (KM) survival analysis for KMT2C, KMT2D and combined expression in the TCGA dataset. Analysis of PDAC tumours from TCGA dataset suggests that lower KMT2C and combined KMT2C and KMT2D expression may correlate with improved patient survival, whereas the opposite may be true for KMT2D. KM survival analysis graphs for assessing the prognostic value of high (red) and low (black) expression for KMT2C (**A**), KMT2D (**B**), and KMT2C/D combined (**C**) within the TCGA dataset. Numbers on the x-axis represent years.

	KMT2C		KMT2D		KMT2C/D	
Demographics and characteristics	**High**	**Low**	**High**	**Low**	**High**	**Low**
Number	34	53	70	17	26	61
Sex						
Male	23 (67.6%)	30 (56.6%)	42 (60.0%)	11 (64.7%)	18 (69.2%)	35 (57.4%)
Female	11 (32.4%)	23 (43.4%)	28 (40.0%)	6 (35.3%)	8 (30.8%)	26 (42.6%)
Age						
Median (IQR)	68.5 (59-78)	66 (61-73)	65.5 (60-74)	70 (68-77)	67.5 (59-75.75)	68 (61-74)
>=68	17	26	31	13	13	31
<68	16	27	39	4	13	30
Tumour grade						
1	0 (0%)	2 (3.8%)	0 (0%)	2 (11.8%)	0 (0%)	2 (3.3%)
2	21 (61.8%)	32 (60.4%)	41 (58.6%)	12 (70.6%)	16 (61.5%)	37 (60.7%)
3	12 (35.3%)	18 (34.0%)	27 (38.6%)	3 (17.6%)	9 (34.6%)	21 (34.4%)
4	1 (2.9%)	1 (0.2%)	2 (2.9%)	0 (0%)	1 (3.8%)	1 (1.6%)

Table 3.1 – Patients details of ICGC patient groups split using expression of KMT2C and/or KMT2C. Included are details of the patient demographics and tumour characteristics for the ICGC samples when patients are split into the high and low expression groups for *KMT2C*, *KMT2D* and both combined.

	KMT2C		KMT2D		KMT2C/D	
Demographics and characteristics	**High**	**Low**	**High**	**Low**	**High**	**Low**
Number	49	59	56	52	83	25
Sex						
Male	25 (51.0%)	34 (57.6%)	30 (53.6%)	29 (55.8%)	44 (53.0%)	15 (60.0%)
Female	24 (49.0%)	25 (42.4%)	26 (46.4%)	23 (44.2%)	39 (47.0%)	10 (40.0%)
Age						
Median (IQR)	64 (52-72)	67 (61.5-75)	63.5 (54.75-68)	69.5 (62.75-75)	64 (55.5-72)	70 (65-75)
>=65	23	36	24	35	40	19
<65	26	23	32	17	43	6
Tumour grade						
1	8 (16.3%)	6 (10.2%)	9 (16.1%)	5 (9.6%)	9 (10.8%)	5 (20.0%)
2	27 (55.1%)	36 (61.0%)	33 (58.9%)	30 (57.7%)	50 (60.2%)	13 (52.0%)
3	14 (28.6%)	16 (27.1%)	14 (2.0%)	16 (30.8%)	24 (28.9%)	6 (24%)
4	0 (0%)	1 (1.7%)	0 (0%)	1 (1.9%)	0 (0%)	1 (4%)
Tumour Stage						
1	3 (6.1%)	5 (8.5%)	3 (5.4%)	5 (9.6%)	6 (7.2%)	2 (8%)
2	43 (87.8%)	50 (84.7%)	49 (87.5%)	44 (84.6%)	72 (86.7%)	21 (84%)
3	1 (2%)	2 (3.4%)	1 (1.8%)	2 (38%)	2 (2.4%)	1 (4%)
4	2 (4.1%)	1 (1.7%)	3 (5.4%)	0 (0%)	3 (3.6%)	0 (0%)
Discrepancy	0 (0%)	1 (1.7%)	0 (0%)	1 (1.9%)	0 (0%)	1 (4%)
Response						
Stable disease	3 (6.1%)	2 (3.4%)	1 (1.8%)	4 (7.7%)	3 (3.6%)	2 (8%)
Progressive disease	15 (30.6%)	14 (23.7%)	15 (26.8%)	14 (26.9%)	22 (26.5%)	7 (28%)
Partial remission	5 (10.2%)	2 (3.4%)	4 (7.1%)	3 (5.8%)	6 (7.2%)	1 (4%)
Complete remission	10 (20.4%)	14 (23.7%)	14 (25%)	10 (19.2%)	21 (25.3%)	3 (12%)
Unknown	6 (12.2%)	21 (35.6%)	11 (19.6%)	16 (30.8%)	17 (20.5%)	10 (40%)
NA	10 (20.4%)	4 (6.8%)	10 (17.9%)	4 (7.7%)	13 (15.7%)	1 (4%)
Discrepancy	0 (0%)	2 (3.4%)	1 (1.8%)	1 (1.9%)	1 (1.2%)	1 (4%)
Race						
White	40 (81.6%)	53 (89.8%)	49 (87.5%)	44 (84.6%)	72 (86.7%)	21 (84%)
Black/African American	4 (8.2%)	1 (1.7%)	3 (5.4%)	2 (3.8%)	4 (4.8%)	1 (4%)
Asian	2 (4.1%)	4 (6.8%)	2 (3.6%)	4 (7.7%)	3 (3.6%)	3 (12%)
Hispanic/Latino	0 (0%)	1 (1.7%)	1 (1.8%)	0 (0%)	1 (1.2%)	0 (0%)
Unknown	3 (6.1%)	0 (0%)	1 (1.8%)	2 (3.8%)	3 (3.6%)	0 (0%)
Maximum tumour dimension						
Median (IQR)	3.6 (2.925-4.5)	3.5 (2.8-4.5)	3.4 (2.8-4.5)	3.5 (3-4.5)	3.5 (2.8-4.5)	4 (3-4.575)
NA	3	6	4	5	6	3
>=3.5mm	26	27	26	27	40	13
<3.5mm	20	26	26	20	37	9

Table 3.2 - Patients details of TCGA patient groups split using expression of KMT2C and/or KMT2D. Included are details of the patient demographics and tumour characteristics for the TCGA samples when patients are split into the high and low expression groups for *KMT2C*, *KMT2D* and both combined.

	KMT2C		KMT2D		Combined KMT2C and KMT2D	
	***p*-value**	**Hazard ratio**	***p*-value**	**Hazard ratio**	***p*-value**	**Hazard ratio**
Expression risk group (high/low)						
Multivariate	0.03566	1.921 (1.045 - 3.531)	0.004160	5.952 (1.7573 - 20.159	0.05750	1.870 (0.9803 - 3.567)
Univariate	0.02	1.94 (1.06 - 3.56)	0.001	5.48 (1.69 - 17.77)	0.044	1.93 (1.01 - 3.69)
Age	0.00175	1.048 (1.018 - 1.079)	0.000318	1.056 (1.0251 - 1.088)	0.00174	1.048 (1.0176 - 1.079)
Tumour grade	0.00493	2.161 (1.263 - 3.699)	0.144477	1.519 (0.8665 - 2.662)	0.00534	2.127 (1.2507 - 3.618)

Table 3.3 – Statistical analysis for clinical factors on survival in ICGC patient groups. Table displays data for multivariate (and univariate where stated) analysis for the clinical data provided by the ICGC dataset when patents are split using expression level of *KMT2C* and *KMT2D*. *p*-values and Hazard ratios are derived from Cox proportional hazards regression model.

	KMT2C		KMT2D		Combined KMT2C and KMT2D	
	***p*-value**	**Hazard ratio**	***p*-value**	**Hazard ratio**	***p*-value**	**Hazard ratio**
Expression risk group (high/low)						
Multivariate	0.259437	1.439 (0.7644 - 2.710)	0.106925	0.605 (0.3284 - 1.115)	0.144319	1.719 (0.8306 - 3.560)
Univariate	0.07	1.63 (0.96 - 2.78)	0.072	0.61 (0.36 - 1.05)	0.24	1.47 (0.77 - 2.82)
Age	0.077062	1.026 (0.9972 - 1.056)	0.483110	1.010 (0.9823 - 1.038)	0.053058	1.030 (0.9996 - 1.061)
Tumour stage	0.474146	1.220 (0.7077 - 2.103)	0.460816	1.242 (0.6986 - 2.207)	0.373286	1.273 (0.7485 - 2.164)
Tumour grade	0.461113	1.163 (0.7787 - 1.736)	0.534646	1.132 (0.7651 - 1.675)	0.509278	1.149 (0.7610 - 1.734)
Treatment partial remission/response	0.000372	9.026 (2.6875 - 30.316)	0.000134	9.969 (3.0637 - 32.438)	7.91e-05	10.548 (3.2742 - 33.981)

Table 3.4 – Statistical analysis for clinical factors on survival in TCGA patient groups. Table displays data for multivariate (and univariate where stated) analysis for the clinical data provided by the TCGA dataset when patents are split using expression level of *KMT2C* and *KMT2D*. *p*-values and Hazard ratios are derived from Cox proportional hazards regression model.

3.2.2 Validation of human cell lines for use as models of PDAC

A number of cell lines have been established from both primary and metastatic tumours of PDAC patients. To examine the roles of KMT2C and KMT2D in PDAC, a panel of eight human pancreatic cell lines were used. Three cell lines were derived from primary tumours (PANC-1, BxPC-3, Capan-2), four were from metastatic tumour sites (SUIT-2, RWP-1, CFPAC-1, COLO 357), and one was an immortalised cell line derived from normal human pancreatic ductal epithelium (HPDE). As expected, these cancer cell lines harbour a range of mutations in four of the genes most commonly mutated in PDAC (Table 3.5).

Although the mutation statuses of *KMT2C* and *KMT2D* have not been extensively described for these cancer cell lines, the Catalogue Of Somatic Mutations In Cancer (COSMIC) (169) database reports that SUIT-2, CFPAC-1, Capan-2, RWP-1, COLO 357 and PANC-1 as having no mutations in either gene, and for the BxPC-3 only two silent point mutations (c.8091C>T and c.8184T>C) are described for *KMT2D*. For all cell lines, KMT2D expression could be detected by western blot (Figure 3.3), confirming that expression is not lost. Encouragingly, p53 was only detected for the lines reported to harbour *TP53* mutations (SUIT-2, CFPAC-1, RWP-1, BxPC-3 and PANC-1, see Table 3.5 and Figure 3.3), where this expression was less pronounced in the CFPAC-1 cell line with a C242R mutation. These findings also indicate that our Capan-2 cells have a WT *TP53* configuration (60,227), and not a 200bp insertion (228), or a R273H mutation (229), which have both also been reported in the literature.

The expression of *KMT2C* and *KMT2D* were also assessed using qRT-PCR (KMT2C and KMT2D, Figure 3.4), where western blot for KMT2C cannot be performed due to the poor quality of commercially available antibodies. *KMT2C* mRNA was detected in all cell lines, where, assuming equal primer amplification efficiency, expression was typically greater than *KMT2D* mRNA (Figure 3.4). For seven of the eight cell lines, *KMT2D* mRNA expression showed a trend similar to that observed at the protein level with the exception of HPDE where mRNA expression was the greatest of all the cell lines examined.

Overall, these cell lines represented a spectrum of backgrounds, reflecting the complexity of PDAC and therefore were subsequently used to examine the *in vitro* roles of these methyltransferases using RNAi.

Cell line	Origin	*KRAS*	*TP53*	*CDKN2A*	*SMAD4*	*KMT2D*[a]	*KMT2C*[a]
SUIT-2	Liver met (212)	G12D (230)	R273H (230)	E69stop (230)	WT (230)	WT	WT
CFPAC-1	Liver met (21)	G12V (60,230)	C242R (60,227,230)	WT[b] (60,227,231)	HD (50,60,230)	WT	WT
RWP-1	Liver met	G12D (230)	R175H (230)	Arg58stop (230)	WT (230)	WT	WT
HPDE	Pancreas[c]	WT (210)	WT[d] (210)	WT (210)	WT[e] (210)	Unknown	Unknown
COLO 357	Celiac axis lymph node met (214)	G12D (229,232)	WT (229)	WT (229,231)	HD (233)	WT	WT
BxPC-3	Primary tum	WT (60,228,229,234,235)	Y220C (60,227-229,234)	HD (227,229,231,234)	HD (50,60,234,235)	Silent SNPs	WT
PANC-1	Primary tum (208)	G12D (60,228,230,232,234,236)	R273H (60,227,230,234,236) R273C (228)	HD (60,230,231,234)	WT (60,230,233,234)	WT	WT
Capan-2	Primary tum	G12V (60,228,229,235)	WT (60,227) Δ200bp splice site (228,237) R273H (229)	WT (60,229), 7bp ins (231)	WT (60,233)	WT	WT

Table 3.5 – List of mutations that have been reported in the human cell lines. Displayed are reports from the literature for the range and heterogeneity of mutations in the top four significantly mutated PDAC genes within the eight cell lines. Met – Metastasis, Tum – Tumour, HD – Homozygous deletion, WT – Wild-type, Δ – Deletion, ins – Insertion, SNP – single nucleotide polymorphism, a Data from (COSMIC) (169) database, b Promoter methylation, c From normal pancreatic duct, d Immortalised by loss of a functional p53 pathway, e Detected by western blot.

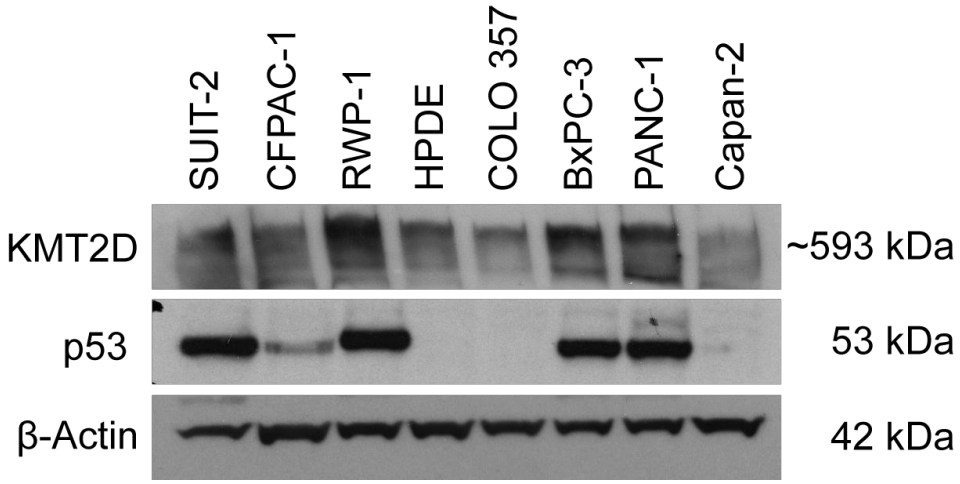

Figure 3.3 – Expression of KMT2D and p53 protein across the eight cell lines. Western blot analysis detected KMT2D protein in each of the cell lines derived from primary (PANC-1, BxPC-3, Capan-2), or metastatic tumours (SUIT-2, CFPAC-1, RWP-1, COLO 357), and in one immortalized normal human pancreatic ductal epithelium (HPDE). Western blot analysis also detected p53 overexpression in the five cell lines previously reported to harbour *TP53* mutations (see **Table 3.5**). β-Actin was used as a loading control.

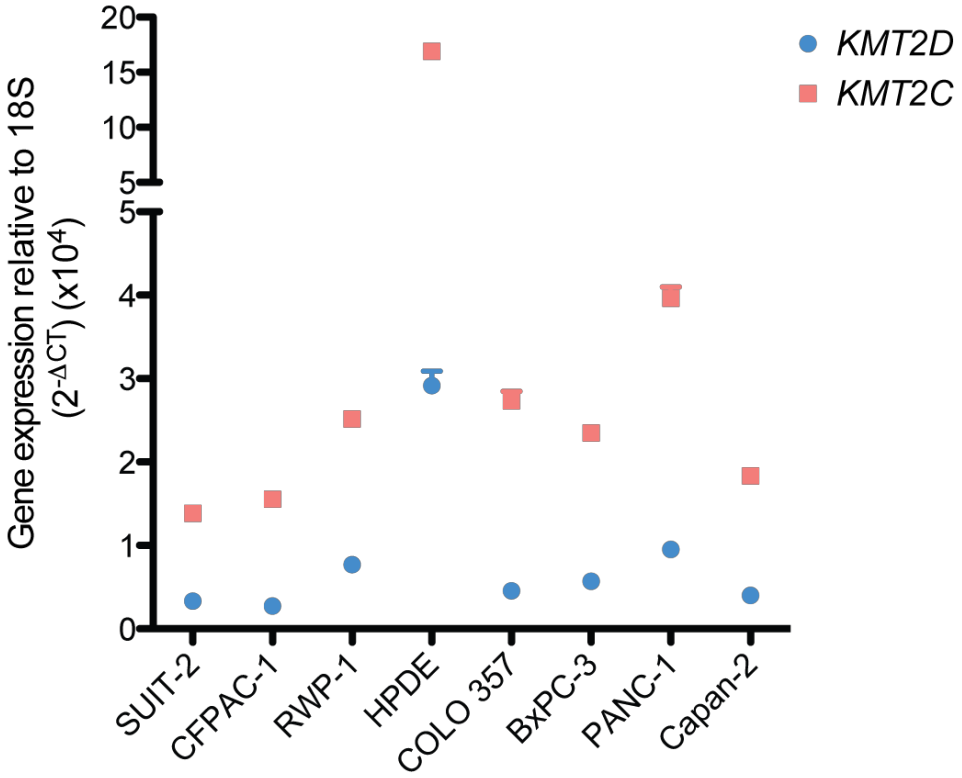

Figure 3.4 – Expression of KMT2C and KMT2D mRNA across the eight cell lines. Expression of *KMT2D* (blue circles) and *KMT2C* (red squares) mRNA relative to the expression of 18S were detected by qRT-PCR in each of the eight cell lines. Data shown are mean values from technical triplicates.

3.2.3 KMT2C and KMT2D are depleted by targeted siRNAs

To deplete *KMT2C*, four unique siRNAs targeting *KMT2C* mRNA were used across the eight cell lines. All four siRNAs were used in the PANC-1, SUIT-2 and COLO 357 cell lines, whereas only two (siRNA1 and siRNA4) were used for silencing in the other five cell lines (Capan-2, BxPC-3, RWP-1, CFPAC-1 and HPDE). The four *KMT2C* siRNAs used were targeted to several exons of the *KMT2C* mRNA transcript (18-19 (siRNA4), 37 (siRNA2), and 38 (siRNA1 and siRNA3)), where all of which were able to deplete *KMT2C* mRNA in all cell lines tested, relative to control siRNA (Figure 3.5). For each cell line examined, *KMT2C* siRNA1 was the most effective (80±10%) at depleting *KMT2C* mRNA. Although knockdown of *KMT2C* by each siRNA was roughly similar, COLO 357 cells appeared to show the smallest relative reduction within the cell lines tested (maximum of 76% for *KMT2C* siRNA1 and minimum of 35% for *KMT2C* siRNA4).

Following western blot analysis, clear reduction in KMT2D was achieved across the eight cell lines for the two siRNAs targeting exons 39 and 48 (Figure 3.6). A third siRNA targeting *KMT2D* also reduced KMT2D expression, however as it was later noted to only elicit minimal effects on cell proliferation, it was not included in later analyses. For seven of the eight cell lines tested, both *KMT2D* siRNAs resulted in a substantial depletion of KMT2D, with more conservative reductions observed in RWP-1 cells. For the BxPC-3 cell line it is worth noting that because the cytostatic effect of *KMT2D* siRNA2 was so strong, minimal protein could be harvested; equal loading of protein was therefore challenging for this cell line.

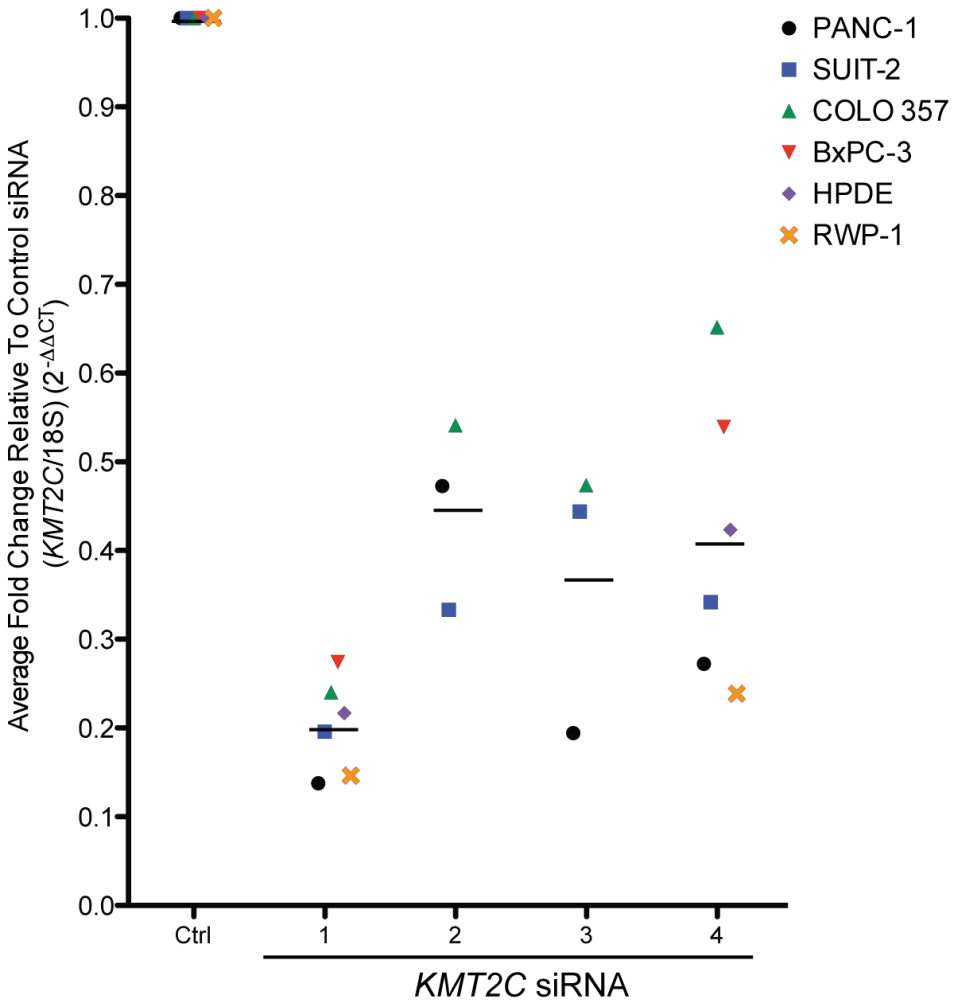

Figure 3.5 – Depletion of *KMT2C* mRNA expression by targeted siRNAs across six pancreatic cell lines. Fold change in *KMT2C* mRNA expression, relative to levels of 18S, for each of the *KMT2C* targeted siRNAs, relative to treatment with untargeted control siRNA, as detected by qRT-PCR across each of the cell lines tested. Data shown are mean values of in triplicates from up to five experimental replicates. For the PANC-1 cells data are from four experiments for control siRNA and *KMT2C* siRNA1, three experiments for *KMT2C* siRNA2 and 3, and one for *KMT2C* siRNA4. For SUIT-2 and COLO 357 cells data are from five experiments for control siRNA and *KMT2C* siRNA1, four experiments for *KMT2C* siRNA2 and 3, and one for *KMT2C* siRNA4. For BxPC-3, HPDE and RWP-1, cells data are from one experiment for control siRNA, *KMT2C* siRNA1, and *KMT2C* siRNA4.

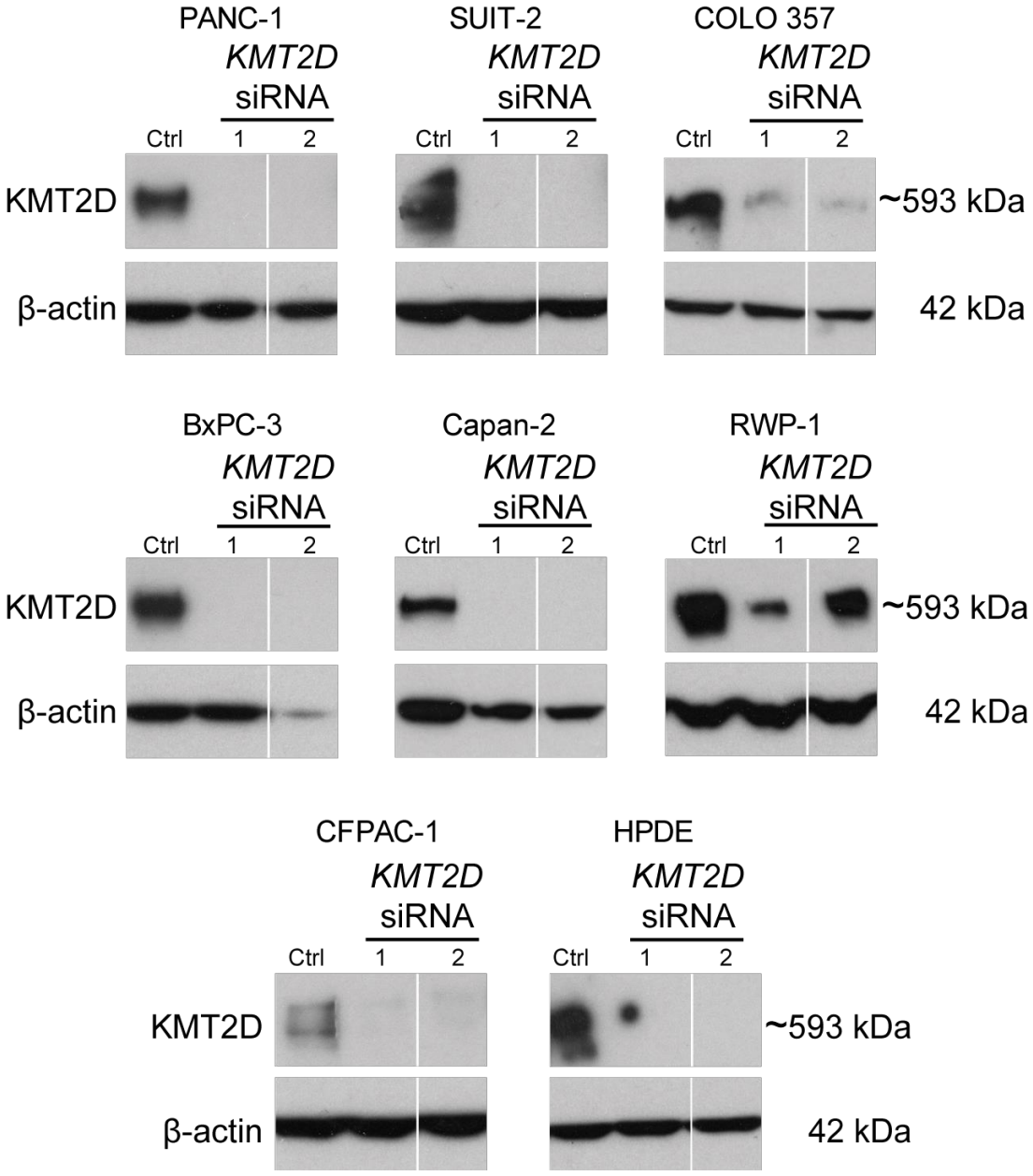

Figure 3.6 – Depletion of KMT2D protein expression by targeted siRNAs in each of the eight cell lines. Western blot analysis to confirm the depletion of KMT2D protein expression by two anti-*KMT2D* siRNAs, compared to an untargeted siRNA, in each of the eight cell lines. A third siRNA was cropped from the images between *KMT2D* siRNA1 and 2. β-Actin was used as a loading control.

3.2.4 Proliferation of cell lines following depletion of *KMT2C* and KMT2D

3.2.4.1 KMT2C

To examine its effect on proliferation, *KMT2C* was first depleted by two siRNAs (*KMT2C* siRNA1 and *KMT2C* siRNA4) across eight human pancreatic cell lines. Across these cell lines *KMT2C* depletion resulted in a variety of changes in their proliferative capacity (Figure 3.7).

KMT2C siRNA1 did not alter proliferation in the RWP-1 and HPDE cell lines, whereas in the remaining six lines proliferation was reduced, although this was not statistically significant for SUIT-2. In five of the cell lines, *KMT2C* siRNA4 also elicited small (PANC-1, Capan-2 and RWP-1), or non-significant (COLO 357 and BxPC-3) decreases in proliferation. Although again non-significant, when the SUIT-2, CFPAC- and HPDE cells were treated with *KMT2C* siRNA4, small increases in cell number were observed.

In light of these discrepancies, two additional siRNAs (*KMT2C* siRNA2 and *KMT2C* siRNA3) were examined in the PANC-1, SUIT-2 and COLO 357 cell lines. As before, these siRNAs elicited differing effects across the cell lines tested. As with *KMT2C* siRNA1, both these siRNAs significantly reduced proliferation in PANC-1 and SUIT-2 cells, however this effect was less for *KMT2C* siRNA3. *KMT2C* siRNA2 did not significantly alter COLO 357 cell proliferation, whereas *KMT2C* siRNA3 did.

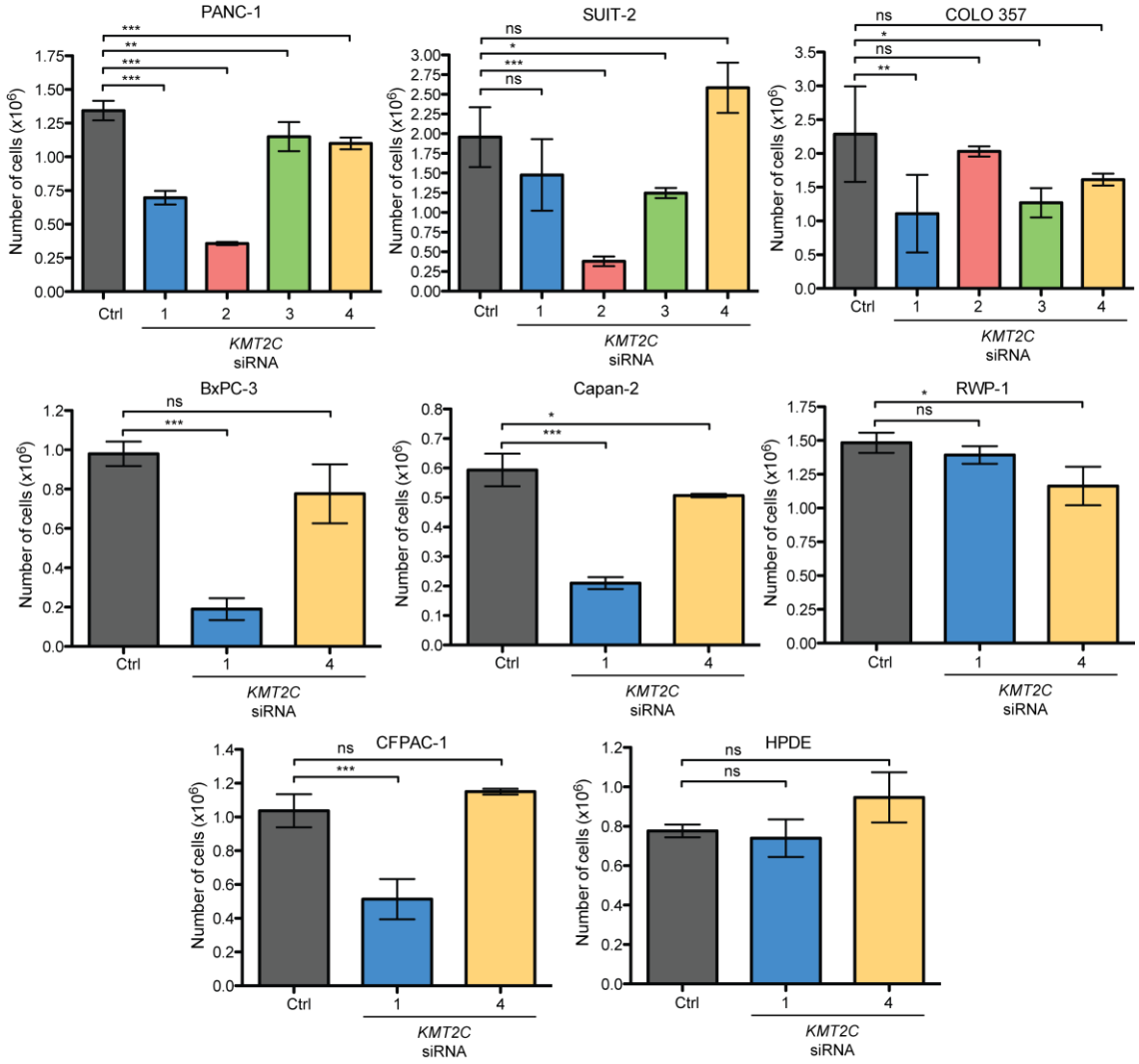

Figure 3.7 – Proliferation of the eight cell lines following *KMT2C* depletion. *KMT2C* mRNA depletion by targeted siRNAs impacts upon cell proliferation, where *KMT2C* depletion variably altered the proliferation of cells. Statistical significance was determined using a One-way ANOVA with Dunnett's post-hoc analysis (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). For PANC-1, SUIT-2 and COLO 357 cells the data shown are mean values ± SD for three replicate wells performed on two separate days (one with control siRNA (grey), *KMT2C* siRNA1 (blue), *KMT2C* siRNA2 (red), *KMT2C* siRNA3 (green); and the other with control siRNA, *KMT2C* siRNA1 and *KMT2C* siRNA4 (yellow)). For the other five cell lines, data shown are mean values ± SD for three replicate wells for control siRNA, *KMT2C* siRNA1, and *KMT2C* siRNA4.

3.2.4.2 KMT2D

Independent silencing by two siRNAs targeted to *KMT2D* resulted in a significant reduction in the proliferation for all of the cell lines tested (Figure 3.8), highlighting an essential role for proliferation. Unlike with *KMT2C* (Figure 3.7), KMT2D depletion results were consistent, where both siRNAs significantly reduced cell proliferation in each cell line tested.

The PANC-1, SUIT-2, COLO 357, BxPC-3 and HPDE cell lines were shown to be the most sensitive to KMT2D depletion, whereas the Capan-2, RWP-1 and CFPAC-1 cell lines were more moderately affected. These differences were not found to correlate with levels of *KMT2C* or *KMT2D* mRNA expression (Figure 3.4), KMT2D protein expression (Figure 3.3), reported mutational status of *KRAS*/*TRP53*/*CDKN2A*/*SMAD4* (Table 3.5), site of tumour (Table 3.5), or response to *KMT2C* siRNA1 (Figure 3.7). For the RWP-1 cell line, this diminished effect on proliferation most likely reflects the weaker knockdown observed by western blot (Figure 3.6).

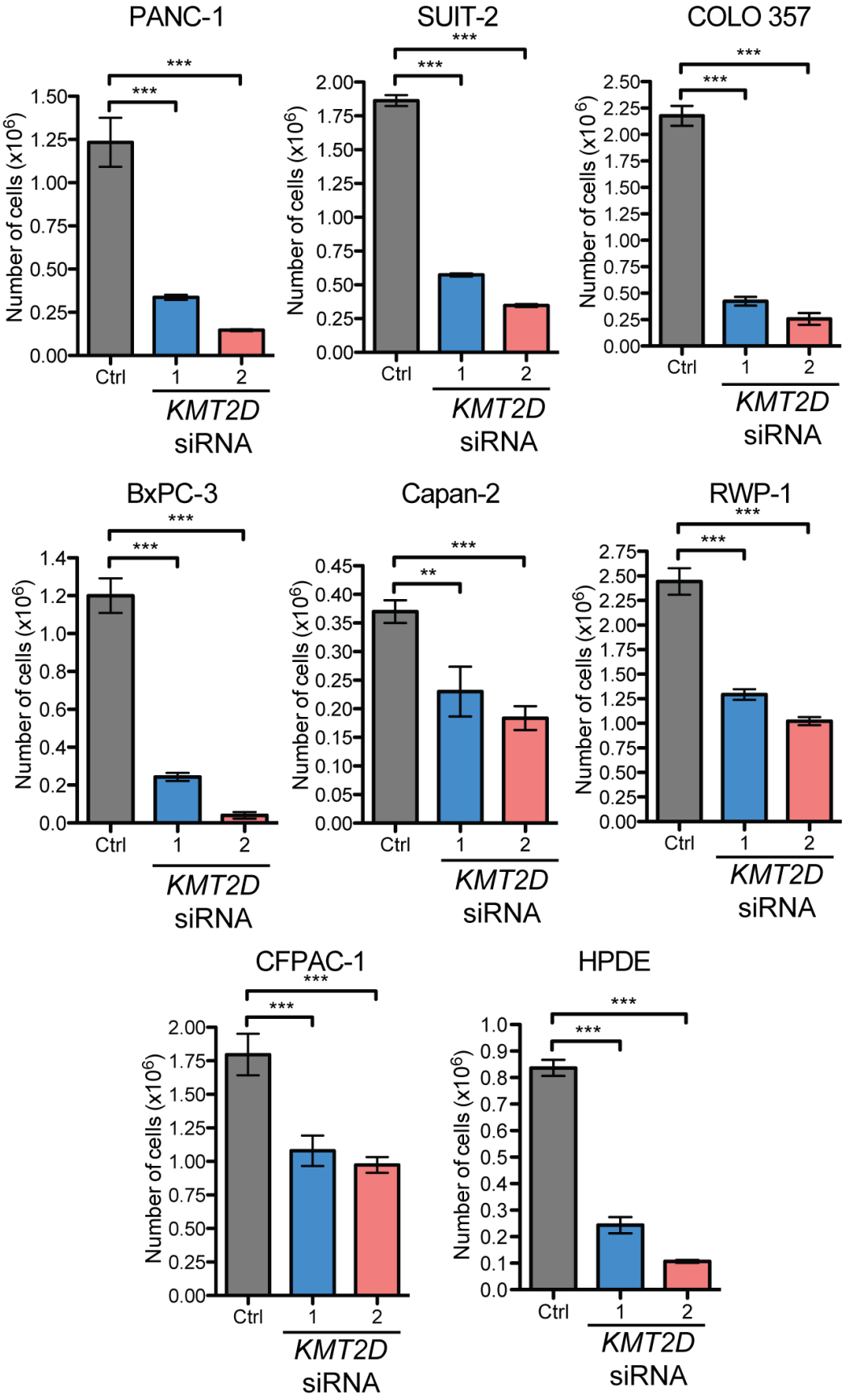

Figure 3.8 – Proliferation of the eight cell lines following KMT2D depletion. Depletion of KMT2D by two anti-*KMT2D* siRNAs significantly inhibits cell proliferation. Statistical significance was determined using a One-way ANOVA with Dunnett's post-hoc analysis (** $p < 0.01$, *** $p < 0.001$). Data shown for each cell line are mean values for three replicate wells ± SD for control siRNA (grey), *KMT2D* siRNA1 (blue), and *KMT2D* siRNA2 (red).

3.2.4.3 Combined KMT2C and KMT2D

KMT2C and *KMT2D* act within similar complexes and thus some functional compensation may exist between the two methyltransferases. As the preceding experiments only considered the proliferative effects of their individual depletion at one time point, the effect of depleting the methyltransferases simultaneously and over time, was examined in PANC1 cells (Figure 3.9). For each methyltransferase, the siRNAs that produced the greatest effect in PANC-1 cells (*KMT2C* siRNA2 (Figure 3.7) and *KMT2D* siRNA2 (Figure 3.8)) were used both separately and in combination.

To account for any potential effect of adding double the amount of oligonucleotide on proliferation, we also included a control with an increased amount of control siRNA. This 'double control' showed no difference to the effect of a single dose of control siRNA (Figure 3.9). As before in our single experiments, *KMT2D* siRNA2 had a strong and significant effect in reducing proliferation at 72 hours, whereas *KMT2C* siRNA2 demonstrated only a moderate effect on cell proliferation, at later time points. Similar to *KMT2C* siRNA2 alone, up to 72 hours a combination of the two siRNAs only produced a weak, albeit significant ($p < 0.05$), effect on proliferation relative to the controls. After this 72-hour time point however, cell proliferation was affected similarly to cells transfected with *KMT2D* siRNA2 alone.

Altogether these three series of experiments suggest that KMT2D is essential for pancreatic cell proliferation, whereas the requirement of KMT2C for proliferation is not as critical and may depend upon other factors.

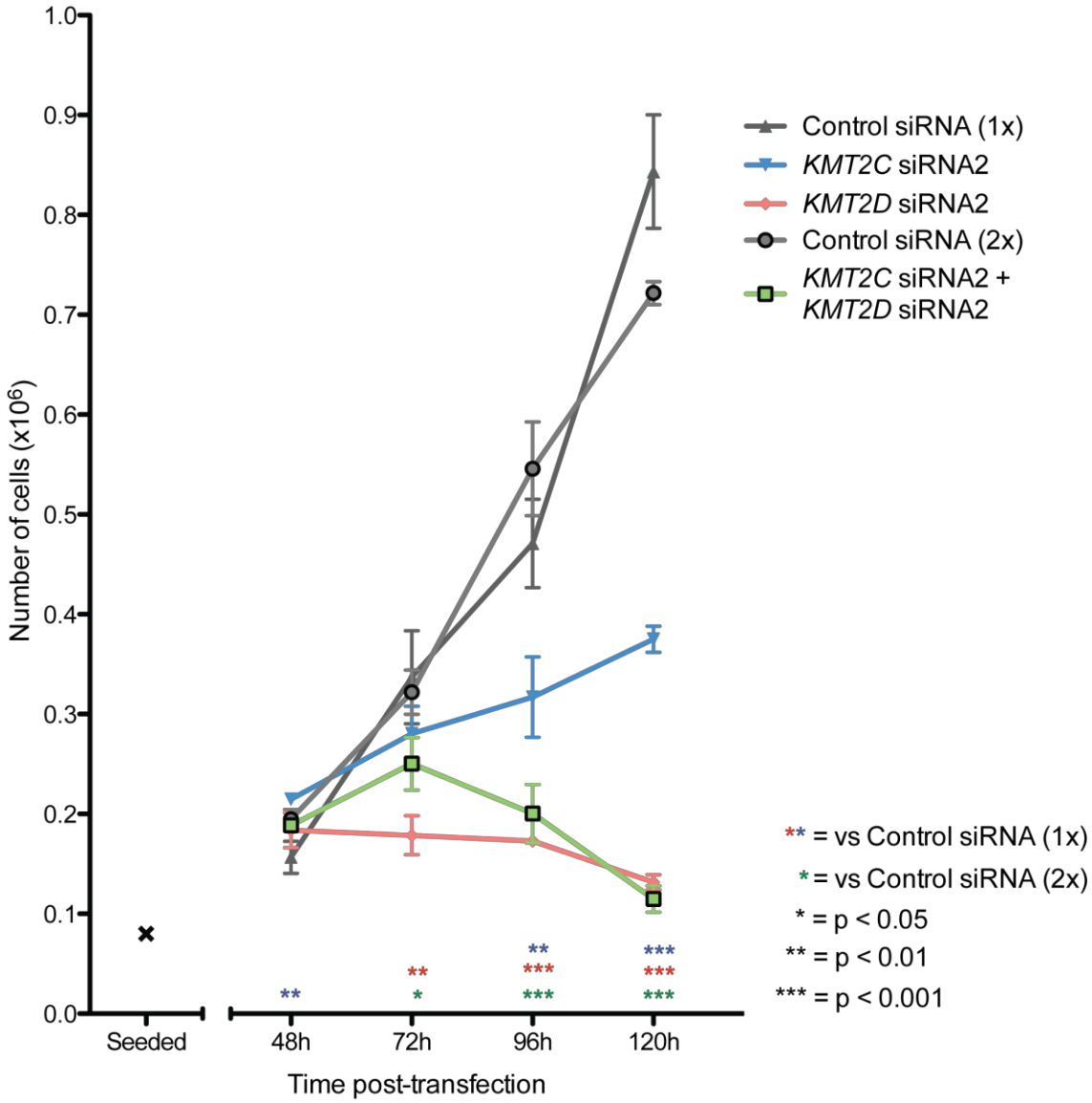

Figure 3.9 - Proliferation of PANC-1 following singular, and combined, depletion of KMT2C and KMT2D. Depletion of KMT2C and KMT2D in parallel imparts a similarly significant inhibition of proliferation as when KMT2D is depleted alone. Statistical significance was determined using a One-way ANOVA with Dunnett's post-hoc analysis (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Data shown for each cell line are mean values ± SD for three replicate wells for control siRNA (1x) (grey), control siRNA (2x) (grey with black outline), *KMT2C* siRNA2 (blue), *KMT2D* siRNA2 (red), and dual depletion of both KMT2C and KMT2D by these siRNAs (green with black outline).

3.2.5 Effects of KMT2D depletion on cell-cycle

Following this observation on cell proliferation, PI staining was used to assess whether the cell-cycle was altered upon KMT2D depletion. Under normal conditions, staining cells with PI produced a histogram with two peaks. The larger of the two peaks, at a lower PI intensity, represents the diploid (2N) cells before DNA synthesis whilst the smaller peak, at double the PI intensity, represents the tetraploid (4N) cells having doubled their DNA content before cell division. Using these peaks, cells within each phase of the cell-cycle can be identified. The cells within the 2N peak are in either of the first two growth phases of the cell-cycle (G_0/G_1), cells within the 4N peak are in either the second growth phase or the mitotic phase (G_2/M), whereas cells found in the region between the two peaks are undergoing DNA synthesis in S-phase. Cells with a PI intensity of less than the 2N peak (<2N) are those undergoing apoptosis, whilst the cells with a PI intensity of greater than the 4N peak (>4N) are considered to be polyploid.

3.2.5.1 Nocodazole block and release assay

Mitosis is a critical stage of the cell-cycle in which the cell divides into two daughter cells. Within the process of mitosis, during prometaphase, cells polymerise microtubules to form the mitotic spindles required for separating chromatids for division. The small molecule nocodazole inhibits microtubule polymerisation, and therefore results in an M-phase cell-cycle brake at the spindle checkpoint. Upon removal of nocodazole, the cells overcome the M-phase brake and recover to resume cell division. To examine how cells without KMT2D progress through the cell cycle, and see whether they reach and then release from this checkpoint, cells transfected with control, or *KMT2D* targeted siRNAs, were treated with nocodazole, and the changes in cell-cycle profiles examined (Figure 3.10).

Under normal "untreated" conditions (U) PI staining gave "normal" cell cycle profiles for each of the six cell lines examined, both when KMT2D was expressed and depleted (Figure 3.10 and Figure 3.11). In the cell lines treated with *KMT2D* siRNAs however, there appeared to be a small yet distinctive increase in proportion of cells in G_0/G_1 (an average of 4% increase for *KMT2D* siRNA1, and 12% for *KMT2D* siRNA2 across the cell lines compared to control siRNA) (Figure 3.10 and Figure 3.11).

For the six cell lines treated with control siRNA, nocodazole treatment (N) produced a large accumulation of cells in G_2/M. When KMT2D was depleted in each of these cell lines, this nocodazole induced G_2/M increase was either completely absent, or diminished (Figure 3.10 and Figure 3.11). This reduced response to nocodazole upon

KMT2D depletion was greatest in the SUIT-2, PANC-1 and COLO 357 cell lines, and weakest in the RWP-1, Capan-2 and BxPC-3 cell lines (Figure 3.11). Upon removal of this nocodazole G_2/M block (R), the majority cells divided and progressed into the G_0/G_1, whilst a small number instead either become polyploid, or undergo apoptosis (Figure 3.11). From these experiments we were able to demonstrate that KMT2D depletion results in a retention of cells in the G_0/G_1-phase, where, unlike cells treated with control siRNA, they do not reach the nocodazole G_2/M block.

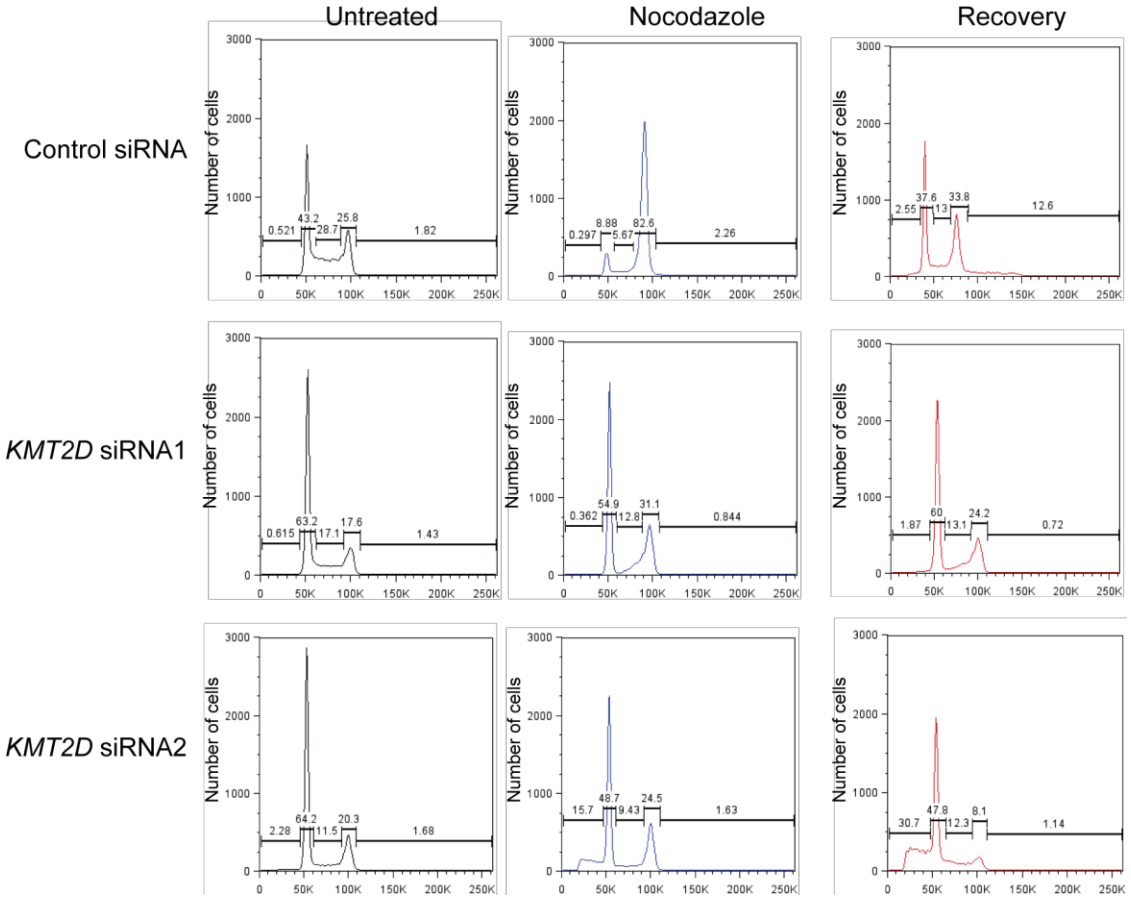

Figure 3.10 – KMT2D is required for cells to reach the nocodazole G_2/M-phase block. Depletion of KMT2D by two anti-*KMT2D* siRNAs prevents G_2/M nocodazole block in six cell lines (PANC-1, SUIT-2, COLO 357, BxPC-3, Capan-2, and RWP-1). Cell DNA content stained using propidium iodide (PI) at three experimental time points. The first was immediately after transfection (Untreated, black traces), next was following a 16-hour incubation with 400ng/ml nocodazole (Nocodazole, blue traces), and finally at 24 hours after the nocodazole block was removed (Recovery, red traces). Data shown are representative histograms from the SUIT-2 cell line for the number of cells and their intensity of PI staining. The gates shown were used to identify the proportion of cells in the apoptotic, G_0/G_1, S, G_2/M, and polyploid phases (from *left* to *right*).

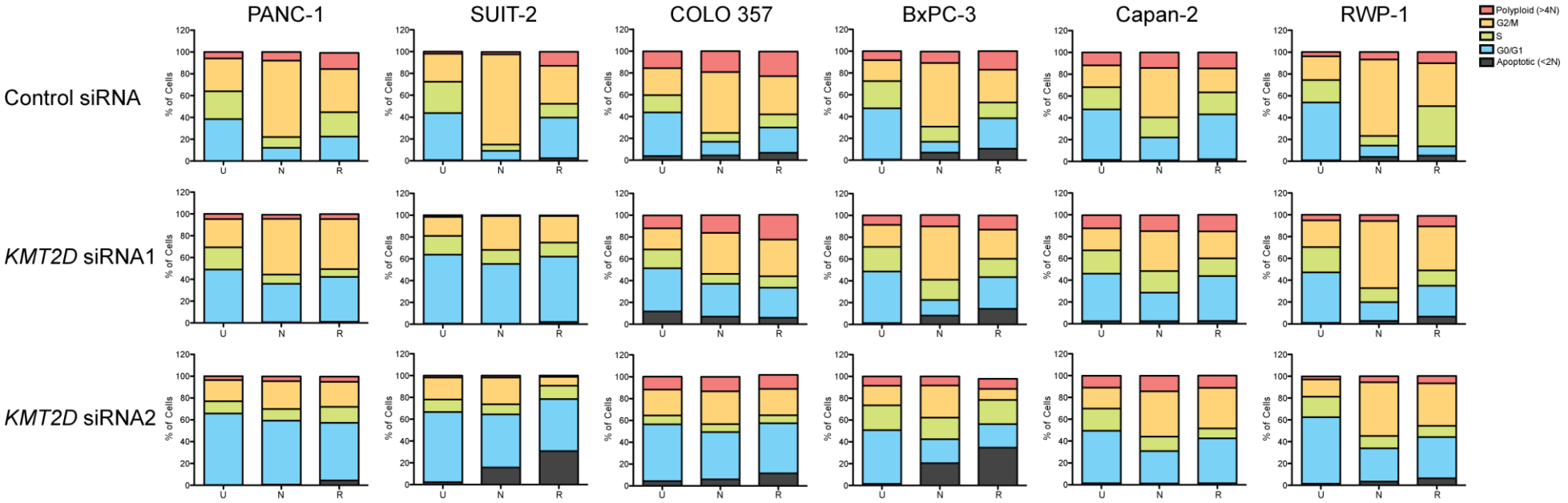

Figure 3.11 – KMT2D is required for cells to accumulate in G_2/M. Depletion of KMT2D by two anti-*KMT2D* siRNAs prevents G2/M nocodazole block in six cell lines (PANC-1, SUIT-2, COLO 357, BxPC-3, Capan-2, and RWP-1). Cell DNA content was stained using propidium iodide (PI) at three experimental time points. The first was immediately after transfection (Untreated, U), next was following a 16-hour incubation with 400ng/ml nocodazole (Nocodazole, N), and finally at 24 hours after the nocodazole block was removed (Recovery, R). Data shown are the proportions of cells in the polyploid (red), G_2/M (orange), S (green), G_0/G_1 (blue), and apoptotic (black) phases determined using the gates shown in **Figure 3.10**.

3.2.5.2 Changes in cell-cycle profile over time

Although the KMT2D depletion appears to perturb cell proliferation by rendering the cells "trapped" in G_0/G_1, it is unknown what happens to these cells over a longer timeframe. It is possible that these cells might remain in G_0/G_1, eventually undergo apoptosis, or pass slowly through G_0/G_1 and into the other cell-cycle phases. To investigate these longer-term effects of KMT2D depletion on cell-cycle, PANC-1 cells were again transfected with either control siRNA or *KMT2D* siRNAs, and the cell-cycle profiles examined at 24-hour intervals over a 120-hour period (Figure 3.12).

When the PANC-1 cells treated with control siRNA were evaluated over several days, the number and proportion of cells in G_0/G_1 slowly increased, while those in the S and G_2/M-phases reduced (Figure 3.12). This is perhaps to be expected where an increase in confluence can affect further growth and proliferation due to increased contact inhibition. In contrast to this, KMT2D depletion led to a decrease in both the number and proportion of PANC-1 cells in G_0/G_1 (Figure 3.12). This decrease in G_0/G_1 was accompanied by a concomitant increase in the apoptotic fraction, whilst the proportions of polyploid and G_2/M-phase cells remained similar (Figure 3.12). From this experiment, it is likely that KMT2D is required for cells to progress out of G_0/G_1, and, if KMT2D is depleted, cells are retained in G_0/G_1 before eventually undergoing apoptosis.

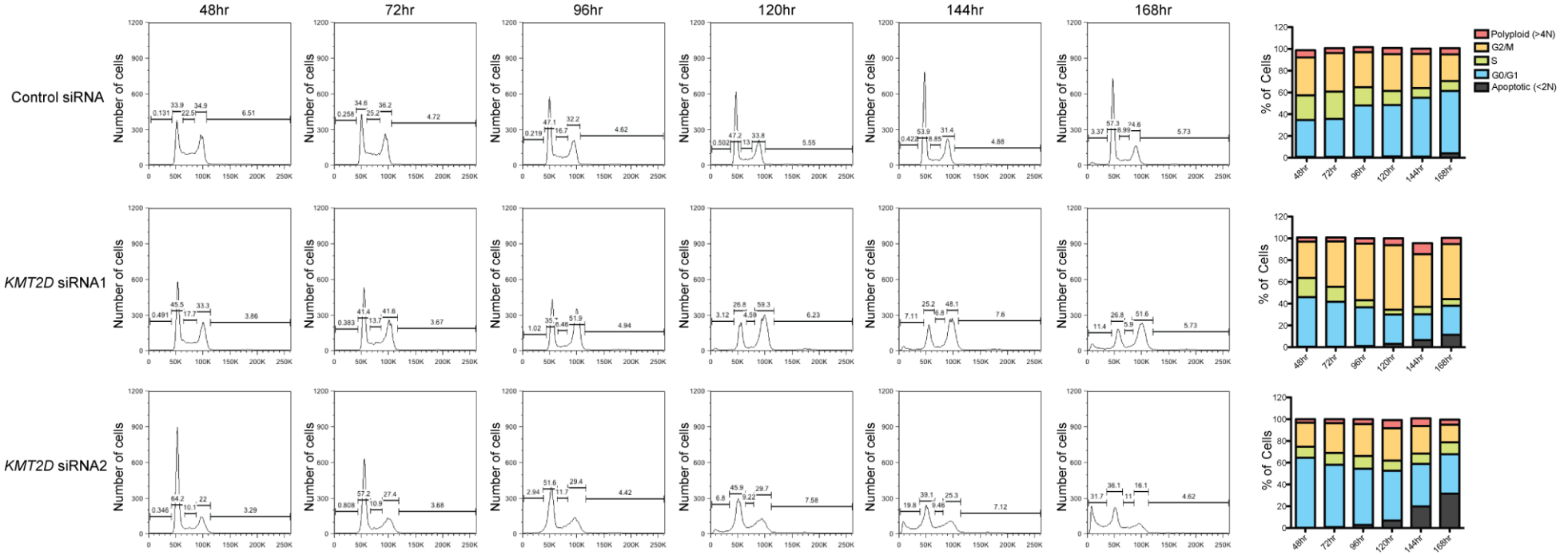

Figure 3.12 – PANC-1 cells blocked in G_0/G_1 due to loss of KMT2D eventually begin to undergo apoptosis. Following the depletion of KMT2D by two anti-*KMT2D* siRNAs, PANC-1 cells blocked in G_0/G_1 eventually begin to undergo apoptosis. PANC-1 cell DNA content stained with propidium iodide (PI) at six 24-hour time points up to 168 hours. Data shown are the histograms for the number of cells and their PI staining intensity, and proportions of cells in the polyploid (red), G_2/M (orange), S (green), G_0/G_1 (blue), and apoptotic (black) phases determined as before using the gates shown.

3.2.6 Effects of KMT2C and KMT2D depletion on the levels of H3K4 methylation and total H3

To examine whether altered levels of H3K4 methylation accompany *KMT2C* and KMT2D depletion, western blots were performed to detect changes in the global levels of mono- (H3K4me1), di- (H3K4me2) and tri- (H3K4me3) methylation (Figure 3.13). Changes in these marks were examined in PANC-1 cells at both 48 and 120 hours following transfection with two *KMT2D* siRNAs, three *KMT2C* siRNAs, and a control siRNA.

For both methyltransferases, at 48 hours post-transfection, no striking changes in any of the methyl marks could be determined (Figure 3.13). At 120 hours, both *KMT2D* siRNAs appeared to reduce the level of all three H3K4 methylation marks compared to control siRNA, where these reductions were perhaps greater for H3K4me1 and H3K4me3, than H3K4me2 (Figure 3.13). Total H3 was also decreased in cells transfected with *KMT2D* siRNAs (Figure 3.13). *KMT2C* depletion had minimal effects on H3K4 methylation at 120 hours, where of the three marks, H3K4me2 particularly unaffected (Figure 3.13).

Taken together the results in this chapter suggest that the methyltransferase KMT2D is required for normal cell-cycle progression, where its loss prevents cell proliferation due to G_0/G_1 cell-cycle arrest and subsequent apoptosis, and thus may confer an improved prognosis.

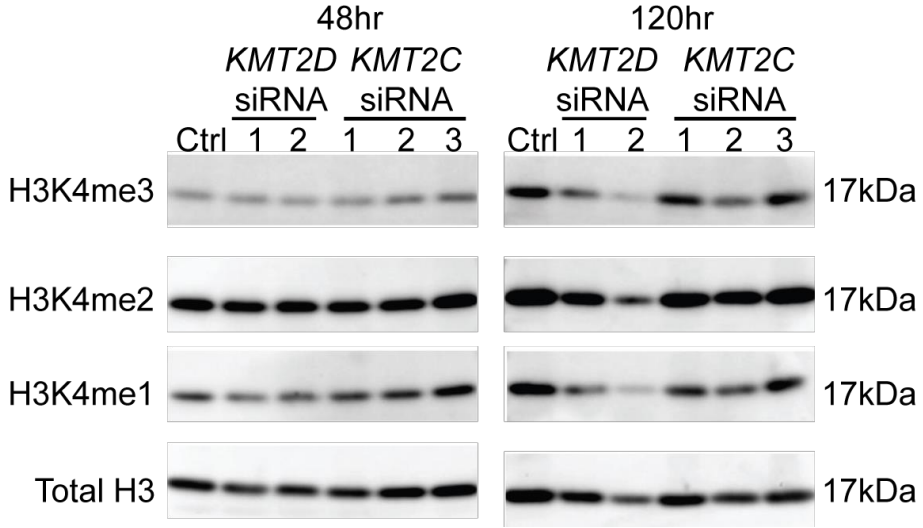

Figure 3.13 – Changes in levels of H3K4 methylation due to depletion of KMT2D. Western blot analysis shows that in PANC-1 cells the levels of H3K4 mono-methylation (H3K4me1), di-methylation (H3K4me2), tri-methylation (H3K4me3), and total H3 are affected by KMT2D depletion at 72 hours post-transfection (120hr), but not immediately following the 48-hour transfection when compared to an untargeted siRNA. Total H3 was used as loading control.

3.3 Discussion

Mutations in *KMT2C* and *KMT2D* (and *KMT2A*) have recently been shown to impart significantly improved overall, and progression free, survival in PDAC (41). In addition, within the setting of breast cancer, *KMT2D* (and *KDM6A*, another member of the KMT2C/D containing complex) expression has been reported also correlate with improved outcome (202). Until now, whether inter-patient fluctuations in the expression of these methyltransferases also confer a benefit to PDAC patients has not been established. By comparing matched GEP and patient outcome in two publically available data series, we could group patients depending on the level of expression for these methyltransferases before comparing differences in survival.

For both genes, independent and combinatory low and high expression patient groups were determined using the commonly used auto-cut-off method (see (238-242) for examples) described by Mihaly *et al.* (224). As opposed to arbitrarily selecting the median as a cut-off point, this approach instead tested each percentile of expression between the lower and upper quartiles in a Cox regression analysis to determine the optimal percentile threshold. For the ICGC PDAC data series, this approach uncovered a strong favourable link with the low expression of these histone modifiers (Figure 3.1), where this could also be resolved using the median cut-off point, albeit this was less significant (data not shown). Although a significant observation was not replicated in the TCGA data series, there was the same trend of low expression correlating with improved survival for either *KMT2C* alone, or in combination with *KMT2D* (Figure 3.2). Unlike within the ICGC dataset, a correlation between improved survival and low *KMT2D* expression alone was not replicated in the TCGA series (Figure 3.2B).

The exact reason for this discordance requires further investigation, where we hypothesise that it may arise from factors that differ between the patient populations (Table 2.11, Table 3.1 and Table 3.2), which are affected by sample inclusion criteria/requirements, or the methods and techniques used to generate the TCGA and ICGC data. Despite this however, the multivariate analysis for the ICGC dataset showed that KMT2C and/or KMT2D expression performed well as prognostic markers independently of age and tumour grade (Table 3.3). Although a similar trend was observed in the TCGA dataset for KMT2C and combined KMT2C and KMT2D expression, independent of age, tumour stage, tumour grade, and primary therapy outcome success, this was not significant ($p > 0.05$) (Table 3.4). Thus, the prognostic potential of KMT2C and KMT2D expression still requires further testing in additional

pancreatic cancer GEP datasets, however this is challenging since such datasets with large patient cohorts and detailed clinical parameters outside of the ICGC and TCGA are still lacking.

Although much research has been done in oncology to identify promising disease biomarkers, relatively few have actually shown clinical effectiveness (243). These shortcomings are often attributed to experimental or analytical problems and biases, and are likely to only be resolved by rigorous reporting and better guidelines (244). To this end, in 2005 McShane *et al.* produced the 'REporting recommendations for tumour MARKer prognostic studies' (REMARK) criteria so that researchers reading studies that follow these criteria can better understand the usefulness and context of any findings (244). To achieve this the REMARK criteria (clearly explained and elaborated on in a checklist by Altman *et al.* (245)) provide information on how best to report and design studies. The information covered by the REMARK checklist (245) can be found in the present study. In this way our identification of *KMT2C* and *KMT2D* expression as a PDAC outcome biomarker is likely to be robust, as highlighted by the similarities to that observed by Sausen *et al.* for LOF mutation (41).

The molecular biological mechanisms by which these methyltransferases may influence patient outcome will be undoubtedly complex, where their role in maintaining H3K4me to facilitate the expression of many genes is likely to be implicated. To begin to understand the roles of these methyltransferases, a series of *in vitro* experiments were performed in this study with RNAi to silence their expression in a panel of human pancreatic cell lines. Throughout the literature there was, in general, a good consensus for the mutational status of the *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* genes in these cell lines (Table 3.5). Reports on the *TP53* and *CDKN2A* mutational status of the Capan-2 cell line are ambiguous with various groups showing *TP53* as WT (60,227), having a 200bp deletion (228,237), or a R273H mutation (229), and *CDKN2A* as being WT (60,229), or having a 7bp insertion (231). These discrepancies may be due to incomplete sequencing, or mismanagement of cell lines arising through misidentification (246), cross contamination, or genomic drift.

The CFPAC-1 cell line was derived from a patient with cystic fibrosis (CF) (21). Mutations in the CF transmembrane conductance regulator (*CFTR*) are reported to be associated with chronic idiopathic pancreatitis (19,20), and therefore acts as risk factor for young onset PDAC (18). Despite the genetic characteristics of CFPAC-1 being well described, the presence of a *CFTR* mutation perhaps renders this cell line better suited

for studies on PDAC development and pathogenesis in CF patients (as has been suggested by Deer *et al*. (247)), where its origins might potentially be confounding and mask observations, such as the true roles of KMT2C and KMT2D in PDAC. For these reasons both the Capan-2 and CFPAC-1 cell lines were only used to examine the effects of KMT2C and KMT2D depletion on cell proliferation and not taken forward in other experiments.

Expression of p53 was examined across the panel of pancreatic cell lines, both as a proxy for the presence of a *TP53* mutation, and also because interactions between KMT2C/KMT2D and p53 have been reported (133,248) (Figure 3.3). Although subtle, there is a suggestion in our data that *TP53* mutant cell lines, which overexpress p53, have slightly increased KMT2D expression when compared to those with a WT configuration (Figure 3.3 and Table 3.5). This supports recent findings by Zhu *et al*. which show that various p53 GOF mutants, including p53(R275H), bind to *KMT2D* whereas their depletion reduces *KMT2D* expression (55). In addition, because Kantidakis *et al.* recently showed that *KMT2D* LOF contribute to genome instability and further DNA damage (152), increased KMT2D expression in *TP53* mutants may act a compensatory mechanism to minimise further DNA damage.

The silencing experiments, which resulted in near complete loss of these methyltransferases (Figure 3.5 and Figure 3.6), showed that depleted KMT2D led to a marked reduction in proliferation in all eight pancreatic cell lines (Figure 3.8). This was consistent with previous studies in a range of human (55,138,151,202) and murine (135) cells, supporting an overlapping role of KMT2D across a variety of solid cancers. While KMT2C was variably implicated in the proliferation of PDAC cells (Figure 3.7), this effect was much less pronounced than that of KMT2D (Figure 3.8 and Figure 3.9). Unlike KMT2D, loss of KMT2C has been shown to either have no effect on (136), or increase (133,199,201,203) the proliferation of cells derived from both solid (133,136,203) and haematological malignancies (201).

Our experiments also go a step further than these studies, whereby we also examined the possibility of complementary roles for KMT2C and KMT2D. Of those mentioned above, only Matkar *et al.* compares data between both, however discrepancies in their use of nomenclature (see 1.3.2) renders their findings somewhat ambiguous, where there is either no effect, or a decrease, on proliferation (136). Since our data showed effects for loss of either methyltransferase on proliferation, we postulate that some, but by no means a complete, functional redundancy exists between the KMT2C and

KMT2D complexes. This may well also explain the observation by Sausen *et al.* that no tumour contains both a *KMT2D* and *KMT2C* mutation (41).

In the *TP53* GOF mutation study by Zhu *et al.,* the authors showed that *TP53* mutations were required for KMT2D depletion to negatively affect proliferation (55), however whether expression of WT p53 could rescue from this effect was not examined. In our experiments, we did not observe this requirement for GOF *TP53* where KMT2D depletion affected both WT and *TP53* mutant cells alike. In addition, it is worth pointing out that despite the varied effects of KMT2C depletion across the cell lines, these effects did not correlate with *TP53* mutation status.

Following KMT2D knockdown by *KMT2D* siRNA2, cell number remained the same, with only a marginal decrease then eventually observed at 120 hours (Figure 3.9). This suggests that the initial effect of KMT2D depletion is cytostatic rather than inducing cell death. To further consider this, we performed cell-cycle analysis following KMT2D depletion. These experiments showed that cells devoid of KMT2D arrest in the G_0/G_1-phases, where nocodazole could not elicit a G_2/M-phase block (Figure 3.10 and Figure 3.11). Previously, Issaeva *et al.* reported that despite KMT2D depletion substantially reducing HeLa cell growth, only a small increase in apoptotic "sub-G_1" cells with no other prominent cell-cycle profile differences were observed (138). Contrary to this and supporting what we observe in "untreated" cell-cycle profiles (Figure 3.10 and Figure 3.11), two studies in murine cells have shown an increase in G_1-phase cells upon Kmt2d depletion (135,249). Using nocodazole we have taken this a step further whereby we showed that these cells accumulated in G_1 without KMT2D do not pass through S-phase to reach the nocodazole pro-metaphase block (Figure 3.10). Following on from this, to determine the fate of these KMT2D depleted and G_1 arrested cells, the cell-cycle profiles of transfected PANC-1 cells were examined up to 120 hours. Over time, without KMT2D the number of cells in the G_0/G_1-phase decreased, while the apoptotic fraction increased (Figure 3.12).

The normal progression of cells through the cell-cycle is largely dependent upon a variety of proteins including cyclins, cyclin-dependent kinases (CDKs), and CDK inhibitors (CDKIs) (reviewed in (250)), the expression of which is tightly controlled. Although we did not explore this possibility further here, it may be that KMT2D is required for the correct temporal and spatial expression of cell cycle regulatory proteins, and thus promotes progression through the cell cycle. Interestingly, from experiments

performed in mouse embryonic cells, Wan *et al.* showed that Kmt2d loss was associated with increased levels of the CDKIs $p16^{ink4a}$ and $p27^{Kip1}$ (135).

The methyltransferases KMT2C and KMT2D are thought to predominantly act by catalysing the monomethylation of H3K4 (94,113,118,120,150,151), where both enzymes contain catalytic SET domains (118). For this reason, their role in facilitating progression of cells through the cell-cycle is likely dependent upon their histone directed methyltransferase activity, therefore as a preliminary investigation we performed western blots to determine changes in the global levels of these marks (Figure 3.13). At the early time point, no discernable change in the three H3K4 methylation states, or total H3, were detected upon KMT2D or KMT2C depletion. At 120 hours however, all three methylation marks were reduced upon KMT2D loss, although since the levels of total H3 also decreased in a similar manner, whether this is a result of decreased methyltransferase activity or a feature of the increased apoptosis cannot be discerned. As with the proliferation experiments, another feature to bear in mind is that for KMT2C only mRNA, and not protein, depletion could be confirmed. Although unlikely, in this way the methyltransferase activity of KMT2C may remain, leaving H3K4 methylation largely unaffected. It may also be that where no changes in methylation were seen, compensatory mechanisms involving the remaining KMT methyltransferases may exist to maintain the marks, where only the depletion of each was studied in isolation. Other limiting features of this technique are that it only examines global levels of these marks on histones, where any subtler and gene loci specific methylation effects, and those also on non-histone proteins, are overlooked.

Previously, several others have also examined the methyltransferase activity of KMT2C and KMT2D where, until more recently, the results have been varied (94,138,150,251,252). Using human cell lines, Goo *et al.* reported that these complexes only have weak H3K4 methyltransferase activity (252), however only KMT2C data was provided. In contrast, Mo *et al.* commented that no methyltransferase activity could be demonstrated for KMT2D (251), whereas Issaeva *et al.* reported clear H3K4 methyltransferase activity (138). In more recent studies however, both complexes are shown to specifically implement H3K4 monomethylation (94,118,153), especially at enhancers (120,150-152) and promoters (113), with perhaps also a weak dimethylation effect (120,151-153), where Zhang *et al.* showed this to increase with time (118). This proposed subtle, yet gene loci-targeted, role might provide one explanation for our observed lack of a change in global H3K4 methylation at 48 hours. In this way,

Kantidakis *et al.* did not observe differences in genomic H3K4me1 upon Kmt2d loss in murine embryonic fibroblasts, and perhaps surprisingly instead noted a marginal increase in H3K4me1-3 at genes damaged when Kmt2d is lost (152). Together these studies suggest that these methyltransferases alter gene expression through specific gene targeted H3K4 methylation changes.

In summary, within this chapter, and in line with the REMARK criteria (245), we have identified a role for KMT2D and KMT2C as prognostic predictors of PDAC outcome, supporting a case for the incorporation of their mutation and expression into existing patient stratification strategies. We speculate that this improved outcome may be a result of the reduced cell proliferation caused by KMT2D depletion, where it appears to hold a key role in cell progression out of G_0/G_1. The functional effects of these methyltransferases are likely due to their roles in orchestrating gene expression through maintaining H3K4 methylation. To better understand specific effects and gene targets of KMT2D and KMT2C, which are likely be both subtle and great in number, in the next chapter we use a large-scale gene expression approach with follow up analysis.

Chapter 4 Effects of KMT2C and KMT2D depletion on the human PDAC transcriptome

4.1 Background

KMT2C and KMT2D function as catalytic subunits within complexes to catalyse H3K4 methylation (144). H3K4 trimethylation is generally associated with active gene expression, particularly at promoters, whereas monomethylated H3K4 (H3K4me1) is found at actively transcribed (94,98,109,112) and poised (107) enhancers. KMT2C and KMT2D preferentially act as monomethyltransferases at these active genomic regions, therefore it is likely our functional observations are regulated through changes in the expression of a range of genes. In support of this, Kantidakis *et al.* have previously shown that *KMT2D* mutation results in a loss of transcription-associated H3K4 methylation (152).

Since complexes containing KMT2C and KMT2D broadly associate with a large range of genomic loci (120,149-152), a large-scale gene expression analysis is required to capture the true extent of the changes to the transcriptome. This approach was used in the present chapter to determine which genes are affected upon the depletion of these epigenetic modifiers. *KMT2C* and *KMT2D* specific RNAi was used to silence gene expression in three PDAC cell lines and isolated total RNA for RNA sequencing (RNA-seq). These RNA samples were sent to Source Bioscience for quality control and RNA-seq whilst Dr Jun Wang at Barts Cancer Institute (BCI) performed all of the bioinformatic analysis.

After initially identifying the genes differentially expressed (DE) upon loss of KMT2D, or KMT2C, gene set enrichment analysis (GSEA) was performed to identify significantly altered pathways. Following this, these profiles were superimposed onto the ICGC and TCGA datasets to assess the biological process enrichment of genes that strongly correlate with *KMT2C* and *KMT2D* expression. Comparing these pathways against the RNA-seq GSEA data helped identify genes commonly associated with histone methyltransferase expression that might also confer an improved clinical outcome.

4.2 Results

4.2.1 Generation of KMT2C and KMT2D depleted RNA samples for RNA-seq

Total RNA samples were generated for RNA-seq from three PDAC cell lines (PANC-1, SUIT-2 and COLO 357) at 48 hours after siRNA transfection. *KMT2C* was depleted in the three cell lines using three targeted siRNAs (*KMT2C* siRNA1, *KMT2C* siRNA2, and *KMT2C* siRNA3), while KMT2D was depleted using two targeted siRNAs (*KMT2D*

siRNA1 and *KMT2C* siRNA2). Before samples were sent to an external service provider Source BioScience (www.sourcebioscience.com) for sequencing, the depletion of both *KMT2C* (mRNA) and *KMT2D* (mRNA and protein) was first confirmed. The concentration and RNA integrity were confirmed to be of sufficient quality in-house before samples were sent to Source BioScience for library preparation, and sequencing.

4.2.1.1 Knockdown of KMT2C and KMT2D for RNA-seq studies

Each targeted siRNA resulted in the knockdown of *KMT2C*, or *KMT2D*, at the mRNA (Figure 4.1A), and/or protein levels (Figure 4.1B) when compared to the scrambled control. A decrease in the expression of *KMT2D* mRNA in the PANC-1 cell line by both *KMT2C* siRNA1 and *KMT2C* siRNA3 was observed (Figure 4.1A, *left*). This effect for *KMT2C* siRNA3 was also detected at the protein level in both the PANC-1 and COLO 357 cell lines (Figure 4.1B). In comparison, *KMT2C* mRNA expression was only slightly decreased by one *KMT2D* siRNA (*KMT2D* siRNA1) in PANC-1 cells, otherwise an increase in *KMT2C* was observed. This stronger increase in *KMT2C* mRNA by *KMT2D* siRNAs, than KMT2D by *KMT2C* siRNAs, may perhaps be indicative of a greater compensatory response to overcome the greater anti-proliferative effects induced by KMT2D depletion.

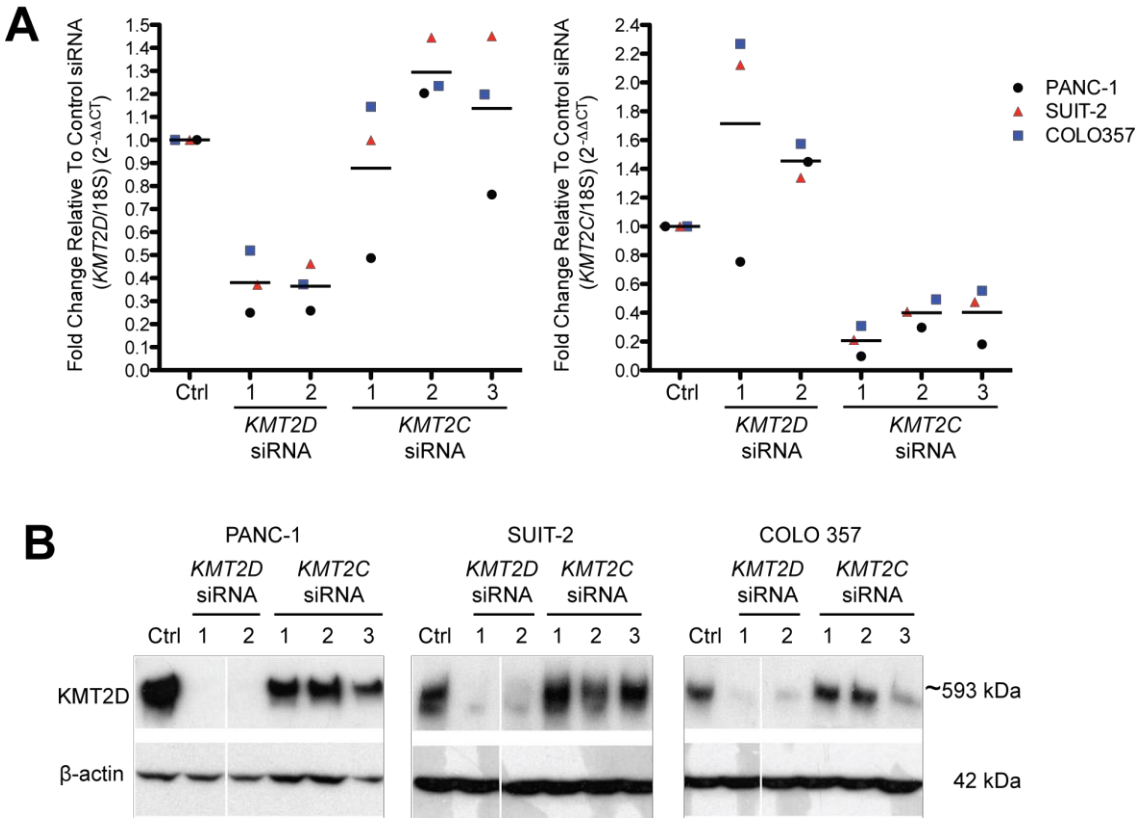

Figure 4.1 – *KMT2D* and *KMT2C* targeted siRNAs reduce expression of *KMT2D* mRNA and protein, and *KMT2C* mRNA respectively. A, RT-qPCR analysis showing changes in *KMT2D* and *KMT2C* mRNA expression in the three cell lines (PANC-1 (black circles), SUIT-2 (red triangles) and COLO 357 (blue squares)) following transfection with the two *KMT2D* siRNAs, the three *KMT2C* siRNAs, or the scramble negative control siRNA. Data shown are normalized mean values from technical triplicates. **B,** Western blot analysis showing that the two *KMT2D* siRNAs reducing KMT2D levels in three cell lines (PANC-1, SUIT-2 and COLO 357), whereas for the three *KMT2C* and control siRNAs it remains relatively unaffected. A third siRNA was cropped from the images between KMT2D siRNA1 and 2.

4.2.1.2 Quality control of samples

Before RNA sequencing was performed, two quality checks were made. Firstly an initial basic measurement was made using a Nanodrop spectrophotometer, and a second more stringent evaluation using a BioAnalyser at Source BioScience (Table 4.1). From the Nanodrop readings, all samples contained at least 167ng/µl of RNA. The 280nm/260nm absorbance ratios suggested that the RNA samples had good levels of purity (> 2). Although the 260nm/230nm ratios suggested that some samples had some guanidine thiocyanate contamination (< 1), it has been reported that 260nm/230nm ratios of < 1 do not affect the library preparation (253). From the Bioanalyser analysis the RNA samples were deemed to be of a quality suitable for sequencing where RNA integrity was close to 10 for all but one sample, where this was > 7.

QMUL (BCI)		**Nanodrop**			**Source Bioscience**	**Bio analyser**	
Sample ID		**Conc (ng/µl)**	**260/280 (nm)**	**260/230 (nm)**	**ID**	**Conc (ng/µl)**	**RIN**
PANC-1	Control siRNA	224.5	2.1	1.83	SOL6268	261	9.8
	KMT2D siRNA1	266.5	2.11	1.83	SOL6269	294	9.6
	KMT2D siRNA2	200.9	2.11	1.93	SOL6271	219	9
	KMT2C siRNA1	372.3	2.09	2.19	SOL6272	342	10
	KMT2C siRNA2	167	2.15	0.67	SOL6273	228	7.3
	KMT2C siRNA3	312.7	2.11	1.46	SOL6274	390	9.2
SUIT-2	Control siRNA	553.8	2.2	1.84	SOL6567	657	10
	KMT2D siRNA1	457.1	2.1	1.97	SOL6568	558	10
	KMT2D siRNA2	312.6	2.12	2.11	SOL6570	405	10
	KMT2C siRNA1	410	2.13	0.72	SOL6571	466	10
	KMT2C siRNA2	314.4	2.13	1.97	SOL6572	245	10
	KMT2C siRNA3	501.3	2.22	1.93	SOL6573	378	10
COLO 357	Control siRNA	588.3	2.15	0.84	SOL6560	564	10
	KMT2D siRNA1	389.4	2.13	1.86	SOL6561	395	10
	KMT2D siRNA2	505.4	2.13	2.18	SOL6563	519	10
	KMT2C siRNA1	312	2.13	2.12	SOL6564	323	10
	KMT2C siRNA2	516.2	2.14	2.14	SOL6565	500	10
	KMT2C siRNA3	252.2	2.14	2.06	SOL6566	276	9.6

Table 4.1 – Nanodrop and BioAnalyser quality control measurements of samples. Table contains data for the RNA concentration, purity (260nm/280nm and 260nm/230nm), and RNA integrity number (RIN) for each sample as quantified using the Nanodrop and BioAnalyser.

4.2.2 Genes differentially expressed following depletion of KMT2C or KMT2D

Transcriptomes for the PANC-1, SUIT-2 and COLO 357 cell lines, where *KMT2C* or *KMT2D* had been depleted using targeted siRNAs, or control siRNA, were assessed using 100 bp paired-end deep RNA sequencing generating approximately 36 million paired-end reads per sample. Bioinformatic analysis was focused on DE protein-coding genes shared between the individual siRNAs for the three PDAC cell lines. There were good correlations in the log2 fold changes for all quantified genes within the *KMT2C* and *KMT2D* siRNA datasets between the different siRNAs when compared to control siRNA (correlation coefficients ranging from 0.39 to 0.60, $p < 0.05$) (Figure 4.2).

KMT2C silencing resulted in differential expression of 790 genes in total, with 31 of these common across all three siRNAs tested (Figure 4.3 and Figure 4.4). For these 31 common DE genes, the fold changes ranged from 0.18 (downregulated, *AKR1B10*) to 6.51 (upregulated, *ANO3*), with 27 genes demonstrating an increase, and four demonstrating a decrease, in expression. *KMT2D* silencing resulted in a total of 1202 DE genes, with 124 of these common to both siRNAs across the three lines (Figure 4.3 and Figure 4.5). Fold changes for the KMT2D DE genes ranged from 0.14 (downregulated, *C2orf54*) to 24.43 (upregulated, *KRT6B*), with 40 genes exhibiting a decrease, and 84 an increase, in gene expression.

Interestingly, there was an overlap of 19 commonly DE genes in the *KMT2C* and *KMT2D* siRNA datasets (Figure 4.3). Each of these genes demonstrated a consistent directional fold change between the two methyltransferase (Table 4.2 and Table 4.3).

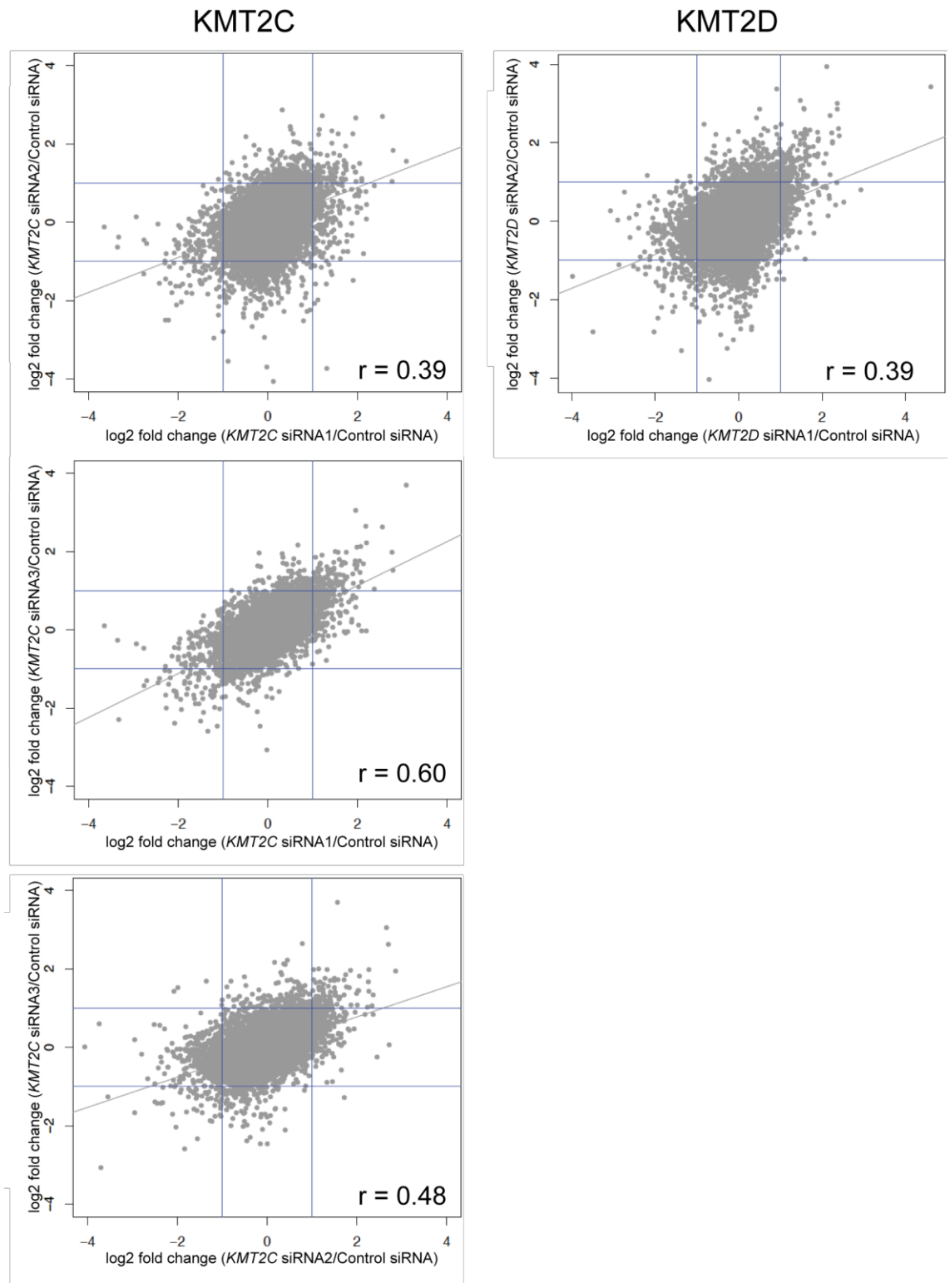

Figure 4.2 – Correlation in differential gene expression between the siRNAs, for both KMT2C and KMT2D within the RNA-seq data. All sufficiently quantified differentially expressed genes were used for the overall correlation analysis between different targeted siRNAs for both genes. Horizontal and vertical lines indicate log2 fold change = 1 and -1, r = Pearson's correlation coefficient.

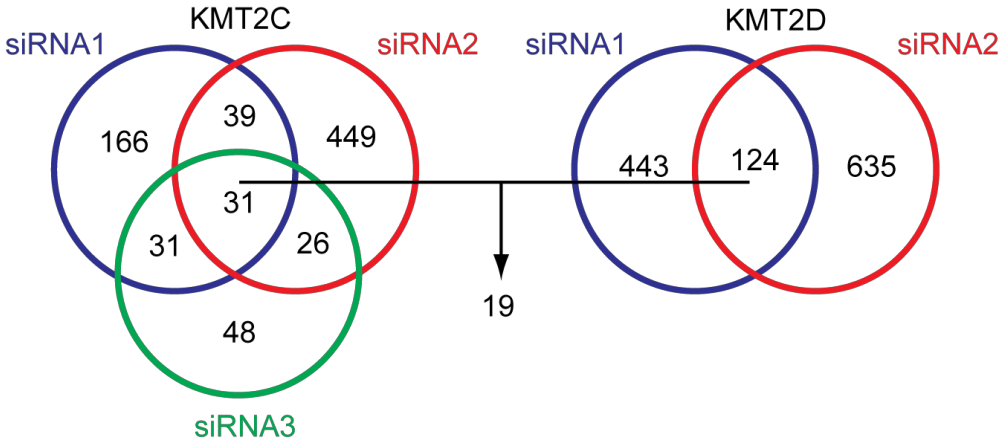

Figure 4.3 - Venn diagram depicting the commonality between the DE genes for each targeted siRNA. Bioinformatic analysis of RNA-seq data identified DE genes common across the three cell lines (PANC-1, SUIT-2 and COLO 357) for the two *KMT2D* siRNAs, and three *KMT2C* siRNAs, compared to the negative control siRNA. Numbers within each set of the Venn diagram displays the number of DE genes for each siRNA across the three cell lines, where 19 genes were found to be common across all siRNAs and cell lines.

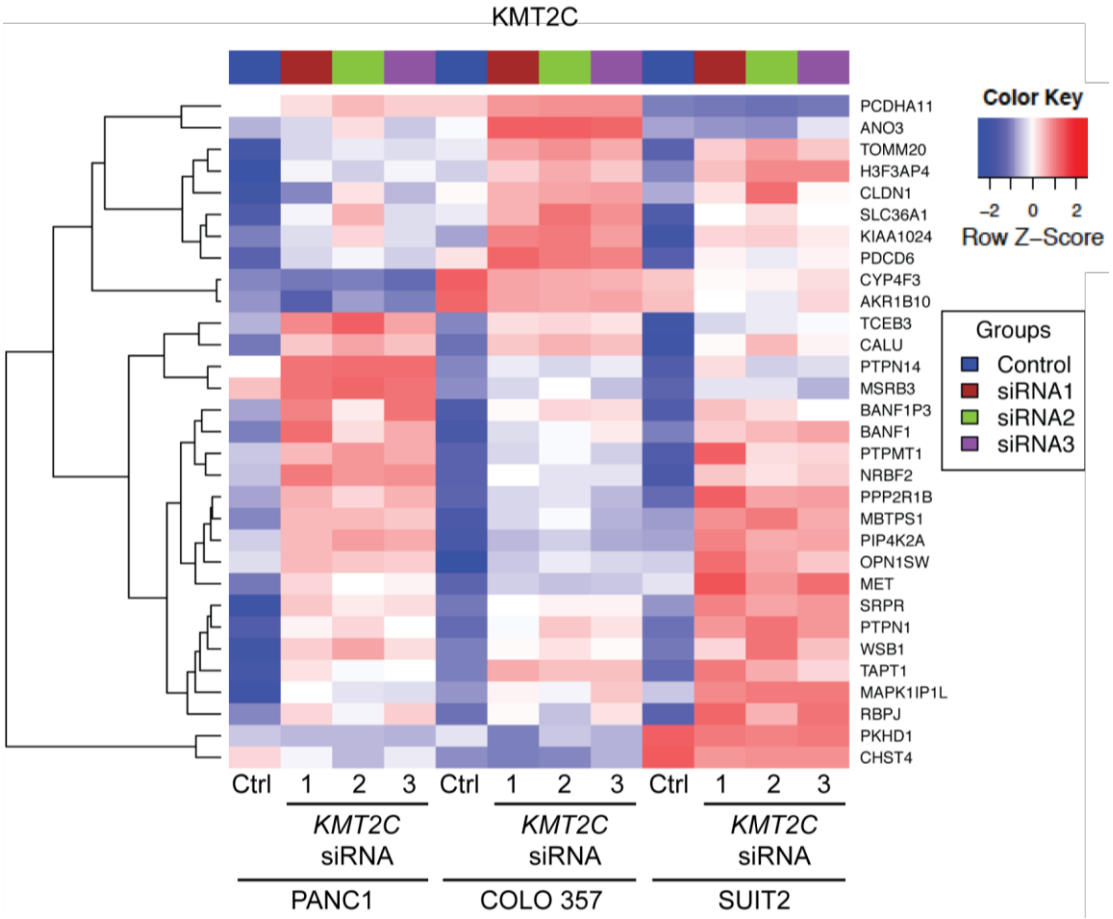

Figure 4.4 – Heatmap depicting the normalised expression of the DE genes for all three KMT2C targeted siRNAs. Bioinformatic analysis of RNA-seq data identified 31 DE genes common across the three cell lines (PANC-1, SUIT-2 and COLO 357) for the three KMT2C siRNAs compared to the negative control siRNA. Heatmaps show the levels of expression for each of the 31 commonly DE genes, where gene downregulation and upregulation are represented by the colours blue and red, respectively. DE genes are listed in Appendix II

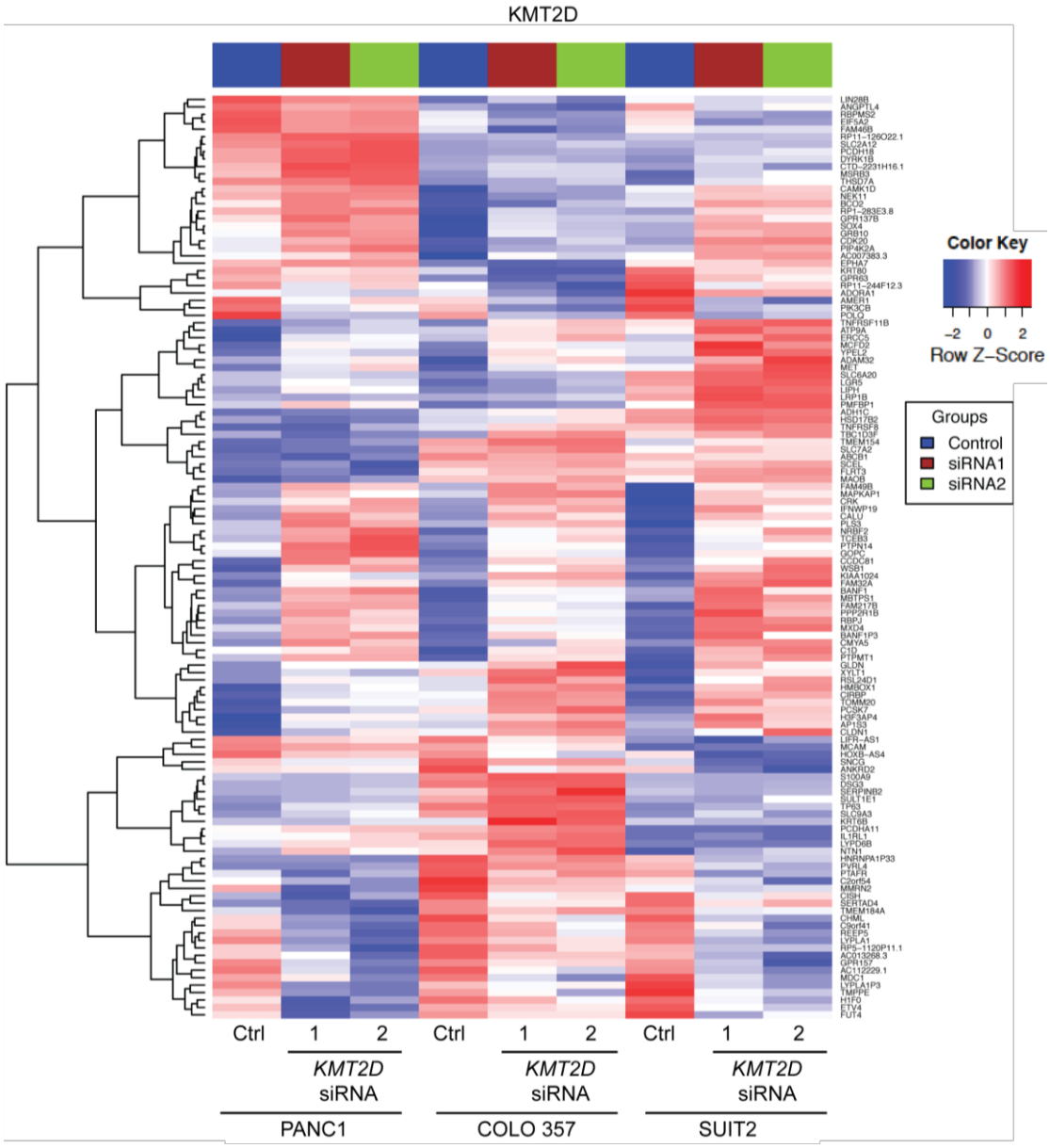

Figure 4.5 - Heatmap depicting the normalised expression of the DE genes for both KMT2D targeted siRNAs. Bioinformatic analysis of RNA-seq data identified 124 DE genes common across the three cell lines (PANC-1, SUIT-2 and COLO 357) for the two KMT2D siRNAs compared to the negative control siRNA. Heatmaps show the levels of expression for each of the 124 commonly DE genes, where gene downregulation and upregulation are represented by the colours blue and red, respectively. DE genes are listed in Appendix III

	KMT2C siRNA1 vs. control				KMT2C siRNA2 vs. control				KMT2C siRNA3 vs. control			
	logFC (treated vs control)	**logCPM**	**PValue**	**FDR**	**logFC (treated vs control)**	**logCPM**	**PValue**	**FDR**	**logFC (treated vs control)**	**logCPM**	**PValue**	**FDR**
TCEB3	1.068	6.709	$6.17x10^{-13}$	$4.09x10^{-10}$	1.217	6.709	$2.92x10^{-16}$	$9.31x10^{-14}$	1.065	6.709	$7.30x10^{-13}$	$7.10x10^{-10}$
MET	2.043	8.702	$1.51x10^{-29}$	$2.41x10^{-25}$	1.495	8.702	$3.71x10^{-17}$	$1.23x10^{-14}$	1.762	8.702	$7.52x10^{-23}$	$5.98x10^{-19}$
WSB1	1.325	6.325	$3.42x10^{-12}$	$1.95x10^{-9}$	1.713	6.325	$5.43x10^{-19}$	$2.47x10^{-16}$	1.328	6.325	$3.10x10^{-12}$	$2.15x10^{-9}$
PTPMT1	1.739	1.821	$5.47x10^{-7}$	$7.13x10^{-5}$	1.508	1.821	$1.21x10^{-5}$	$3.63x10^{-4}$	1.337	1.821	$1.14x10^{-4}$	$7.26x10^{-3}$
CALU	1.470	8.627	$3.39x10^{-26}$	$2.70x10^{-22}$	1.766	8.627	$2.04x10^{-36}$	$1.08x10^{-32}$	1.537	8.627	$2.22x10^{-28}$	$3.53x10^{-24}$
PPP2R1B	1.400	7.008	$7.44x10^{-12}$	$4.08x10^{-9}$	1.167	7.008	$8.27x10^{-9}$	$6.48x10^{-7}$	1.166	7.008	$9.16x10^{-9}$	$2.75x10^{-6}$
MBTPS1	1.146	7.175	$1.02x10^{-15}$	$1.48x10^{-12}$	1.270	7.175	$6.81x10^{-19}$	$3.01x10^{-16}$	1.009	7.175	$1.48x10^{-12}$	$1.12x10^{-9}$
NRBF2	1.285	4.807	$7.40x10^{-15}$	$7.85x10^{-12}$	1.103	4.807	$2.43x10^{-11}$	$3.14x10^{-9}$	1.177	4.807	$1.14x10^{-12}$	$9.05x10^{-10}$
PIP4K2A	1.365	5.889	$2.97x10^{-15}$	$3.63x10^{-12}$	1.393	5.889	$6.30x10^{-16}$	$1.89x10^{-13}$	1.256	5.889	$3.22x10^{-13}$	$3.94x10^{-10}$
PTPN14	1.391	8.218	$1.07x10^{-13}$	$8.49x10^{-11}$	1.070	8.218	$7.56x10^{-9}$	$6.01x10^{-7}$	1.158	8.218	$4.17x10^{-10}$	$1.79x10^{-7}$
CLDN1	1.146	6.203	$6.68x10^{-5}$	$3.30x10^{-3}$	2.362	6.203	$4.84x10^{-15}$	$1.31x10^{-12}$	1.340	6.203	$3.57x10^{-6}$	$5.26x10^{-4}$
RBPJ	1.641	6.001	$1.57x10^{-12}$	$9.23x10^{-10}$	1.080	6.001	$2.51x10^{-6}$	$1.00x10^{-4}$	1.657	6.001	$8.46x10^{-13}$	$7.38x10^{-10}$
KIAA1024	1.244	1.563	$8.19x10^{-6}$	$6.21x10^{-4}$	1.419	1.563	$2.44x10^{-7}$	$1.33x10^{-5}$	1.138	1.563	$5.34x10^{-5}$	$4.13x10^{-3}$
TOMM20	1.151	7.845	$9.14x10^{-12}$	$4.85x10^{-9}$	1.345	7.845	$2.30x10^{-15}$	$6.43x10^{-13}$	1.158	7.845	$7.10x10^{-12}$	$4.52x10^{-9}$
MSRB3	1.896	5.357	$1.14x10^{-14}$	$1.07x10^{-11}$	2.266	5.357	$2.14x10^{-20}$	$1.18x10^{-17}$	1.430	5.357	$4.44x10^{-9}$	$1.57x10^{-6}$
BANF1	1.411	5.184	$1.76x10^{-13}$	$1.34x10^{-10}$	1.251	5.184	$5.74x10^{-11}$	$6.92x10^{-9}$	1.482	5.184	$9.71x10^{-15}$	$2.58x10^{-11}$
H3F3AP4	1.337	6.941	$6.83x10^{-11}$	$3.02x10^{-8}$	1.497	6.941	$3.17x10^{-13}$	$6.07x10^{-11}$	1.526	6.941	$1.37x10^{-13}$	$2.10x10^{-10}$
BANF1P3	1.729	4.141	$2.39x10^{-14}$	$2.11x10^{-11}$	1.497	4.141	$4.84x10^{-11}$	$5.97x10^{-9}$	1.695	4.141	$8.05x10^{-14}$	$1.42x10^{-10}$
PCDHA11	1.506	7.094	$6.03x10^{-17}$	$1.37x10^{-13}$	1.403	7.094	$5.51x10^{-15}$	$1.46x10^{-12}$	1.401	7.094	$6.87x10^{-15}$	$2.57x10^{-11}$

Table 4.2 – Table detailing the 19 commonly DE genes for all three KMT2C siRNA vs control siRNA datasets. Table contains data for the logFC (fold change), logCPM (counts per million), PValue, and false discovery rate (FDR) for each of the 19 commonly DE genes for KMT2C siRNAs.

	KMT2D siRNA1 vs control				KMT2D siRNA2 vs control			
	logFC (treated vs control)	**logCPM**	**PValue**	**FDR**	**logFC (treated vs control)**	**logCPM**	**PValue**	**FDR**
TCEB3	1.036	6.818	$9.66x10^{-11}$	$1.86x10^{-8}$	1.518	6.818	$5.93x10^{-21}$	$3.24x10^{-18}$
MET	1.476	8.694	$4.55x10^{-18}$	$4.23x10^{-15}$	2.068	8.694	$6.73x10^{-33}$	$2.13x10^{-29}$
WSB1	1.555	6.467	$5.48x10^{-15}$	$2.71x10^{-12}$	2.184	6.467	$3.58x10^{-27}$	$5.66x10^{-24}$
PTPMT1	1.618	1.914	$1.64x10^{-8}$	$1.63x10^{-6}$	1.759	1.914	$9.24x10^{-10}$	$6.99x10^{-8}$
CALU	1.551	8.574	$2.00x10^{-26}$	$3.17x10^{-22}$	1.186	8.574	$2.59x10^{-16}$	$6.51x10^{-14}$
PPP2R1B	1.612	7.076	$9.21x10^{-15}$	$4.16x10^{-12}$	1.119	7.076	$4.47x10^{-8}$	$2.28x10^{-6}$
MBTPS1	1.320	7.269	$3.52x10^{-19}$	$4.28x10^{-16}$	1.284	7.269	$2.65x10^{-18}$	$1.05x10^{-15}$
NRBF2	1.136	4.880	$1.22x10^{-10}$	$2.20x10^{-8}$	1.611	4.880	$1.27x10^{-19}$	$5.90x10^{-17}$
PIP4K2A	1.073	5.685	$7.21x10^{-10}$	$1.05x10^{-7}$	1.140	5.685	$5.84x10^{-11}$	$6.03x10^{-9}$
PTPN14	1.355	8.347	$4.65x10^{-17}$	$3.20x10^{-14}$	1.544	8.347	$1.80x10^{-21}$	$1.29x10^{-18}$
CLDN1	1.563	6.727	$8.20x10^{-9}$	$8.82x10^{-7}$	2.871	6.727	$8.56x10^{-24}$	$8.46x10^{-21}$
RBPJ	1.688	6.007	$6.07x10^{-15}$	$2.91x10^{-12}$	1.433	6.007	$2.89x10^{-11}$	$3.17x10^{-9}$
KIAA1024	1.592	1.792	$1.68x10^{-6}$	$8.48x10^{-5}$	1.625	1.792	$1.10x10^{-6}$	$3.62x10^{-5}$
TOMM20	1.551	7.984	$4.50x10^{-18}$	$4.23x10^{-15}$	1.255	7.984	$1.44x10^{-12}$	$1.96x10^{-10}$
MSRB3	1.894	5.529	$1.93x10^{-12}$	$5.25x10^{-10}$	2.259	5.529	$1.33x10^{-16}$	$3.63x10^{-14}$
BANF1	1.233	4.890	$2.94x10^{-10}$	$4.79x10^{-8}$	1.042	4.890	$9.91x10^{-8}$	$4.52x10^{-6}$
H3F3AP4	1.375	6.881	$5.56x10^{-10}$	$8.38x10^{-8}$	1.213	6.881	$3.67x10^{-8}$	$1.95x10^{-6}$
BANF1P3	2.045	4.135	$8.90x10^{-14}$	$3.27x10^{-11}$	1.478	4.135	$4.00x10^{-8}$	$2.09x10^{-6}$
PCDHA11	1.465	0.854	$7.03x10^{-4}$	$1.14x10^{-2}$	1.957	0.854	$3.95x10^{-6}$	$1.03x10^{-4}$

Table 4.3 – Table detailing the 19 commonly DE genes for both KMT2D siRNA vs control siRNA datasets Table contains data for the logFC (fold change), log CPM (counts per million), PValue, and false discovery rate (FDR) for each of the 19 commonly DE genes for KMT2D siRNAs.

4.2.2.1 Increased expression of MET, Calumenin, Claudin-1, PTPN14 and PIP4K2A

Five of the 19 genes common to all siRNAs for both methyltransferases were selected for validation. Four of these were previously associated with pancreatic cancer (*MET* (42,254), *PTPN14* (255), *Calumenin* (255) and *Claudin-1* (256,257)) whilst a fifth (*PIP4K2A*) encodes a lipid kinase involved in cell signal transduction (258). The RNA-seq values for the number of reads per kilobase of transcript per million mapped reads (RPKM) showed that methyltransferase depletion increased the expression for each of these genes (Figure 4.6A). These changes were confirmed at the protein level by western blot (Figure 4.6B). Data generated for each gene are described in the sections below.

4.2.2.1.1 *MET* (MET Proto-Oncogene, Receptor Tyrosine Kinase)

MET is a proto-oncogene that encodes for a precursor protein (pro-MET) for the receptor tyrosine kinase, MET (also known as c-MET, or hepatocyte growth factor receptor (HGFR)). In malignant cells, aberrations in MET activation occur not only through increased amounts of ligand, but also through ligand-dependent and ligand-independent mechanisms such as gene over expression, amplification, or mutation (reviewed in (259)). Previously, large meta-analyses have shown that high MET expression is associated with poor prognosis and survival in both gastric (260) and breast (261) cancers. Moreover, in PDAC patients with stage I and II disease, high MET expression levels have also been shown to identify those with a high risk of relapse and poor survival (254).

Under control conditions, each of the three cell lines expressed the cleaved MET protein, with no uncleaved pro-MET detected (Figure 4.6B). As expected from the RNA-seq data, depletion of KMT2C or KMT2D resulted in increased MET expression. This was most striking in the SUIT-2 cell line, and least so in the PANC-1 cell line, mirroring previous observations for untreated cell lines by Di Renzo *et al.* (262). Interestingly, and in contrast with this study, only the uncleaved 170kDa precursor protein was detected when SUIT-2 cells were devoid of the methyltransferases.

4.2.2.1.2 *PTPN14* (Protein Tyrosine Phosphatase, Non-Receptor Type 14)

The *PTPN14* gene encodes for a Class I, non-receptor type, protein-tyrosine phosphatase called PTPN14 (also called Pez, PTPD2, or PTP36), and has been proposed as a tumour suppressor (263). PTPN14 is involved in the dephosphorylation

of tyrosine residues on range of substrates (264), with roles for PTPN14 reported in cell proliferation (265,266), cell adhesion (265,266), migration (263,264,267), protein trafficking (264), and response to chemotherapy (268).

The expression of PTPN14 before methyltransferase depletion was higher in primary PANC-1 cells than in the metastatic SUIT-2 and COLO 357 cell lines (Figure 4.6). When the PANC-1 cell line was treated with *KMT2C* and *KMT2D* siRNAs, a strong increase in PTPN14 was noted, which was more pronounced following KMT2D depletion (Figure 4.6B). SUIT-2 and COLO 357 cells exhibited smaller increases in PTPN14 expression (Figure 4.6, where for COLO 357 this weak increase was only observed following *KMT2C* depletion, and not KMT2D (Figure 4.6B). Of the *KMT2C* siRNAs, across the three cell lines, *KMT2C* siRNA2 had a much weaker effect on PTPN14 expression, however this was not observed in the RNA-seq RPKM data (Figure 4.6).

4.2.2.1.3 *PIP4K2A* (Phosphatidylinositol-5-Phosphate 4-Kinase Type 2 Alpha)

PIP4K2A encodes a lipid kinase that, like PTPN14, is involved in the transduction of intracellular signals. PIP4K2A functions as a kinase to facilitate the phosphorylation of phosphatidylinositol-5-phosphate (PI(5)P), converting it into phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$) (258). $PI(4,5)P_2$ is a substrate for the membrane bound G protein-coupled receptor (GPCR) activated enzyme, phospholipase C (PLC), where $PI(4,5)P_2$ is converted into the secondary messengers diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3), whereby IP_3 facilitates intracellular Ca^{2+} release (269). Although more subtle than with the other proteins, especially when KMT2C is depleted, PIP4K2A expression was increased when the cells did not express the H3K4 methyltransferases (Figure 4.6B).

4.2.2.1.4 *CALU* (Calumenin)

CALU encodes for a low-affinity Ca^{2+} binding, endoplasmic reticulum (ER) located, (CREC) family protein called Calumenin (270,271). The storage of cellular Ca^{2+} within the ER has two distinct roles in facilitating cellular processes. Firstly, the resulting low cytoplasmic Ca^{2+} levels mean that Ca^{2+} release into the cytoplasm functions as an intracellular signal. Secondly, the high Ca^{2+} levels contained within the ER helps chaperone proteins to facilitate protein folding and secretion (272). Calumenin has previously been implicated in secretory pathways (271,273) where its low Ca^{2+} binding capacity has been proposed to impact Ca^{2+} homeostasis (274) and sensing, rather than in buffering (275). In addition to this ER function, Calumenin itself is secreted into the

extracellular space (255,273,276) where it inhibits cell migration by preventing the degradation of extracellular matrix (ECM) by matrix metalloproteinase (MMP) (255). As with the other highlighted genes, Calumenin expression was increased in the three cell lines upon KMT2C or KMT2D depletion (Figure 4.6B).

4.2.2.1.5 *CLDN1* (Claudin-1)

CLDN1 encodes for the transmembrane protein Claudin-1, a member of the Claudin family of tight junction proteins. Tight junction protein complexes containing Claudins are found between the membranes of neighbouring epithelial cells are essential for the formation of a cell barrier (277). The increase in CLDN1 expression due to the depletion of *KMT2C*, or KMT2D, was found to be greatest in the SUIT-2 and COLO 357 metastatic cell lines, whereas PANC-1 cells only exhibited a minor increase upon KMT2D depletion (Figure 4.6).

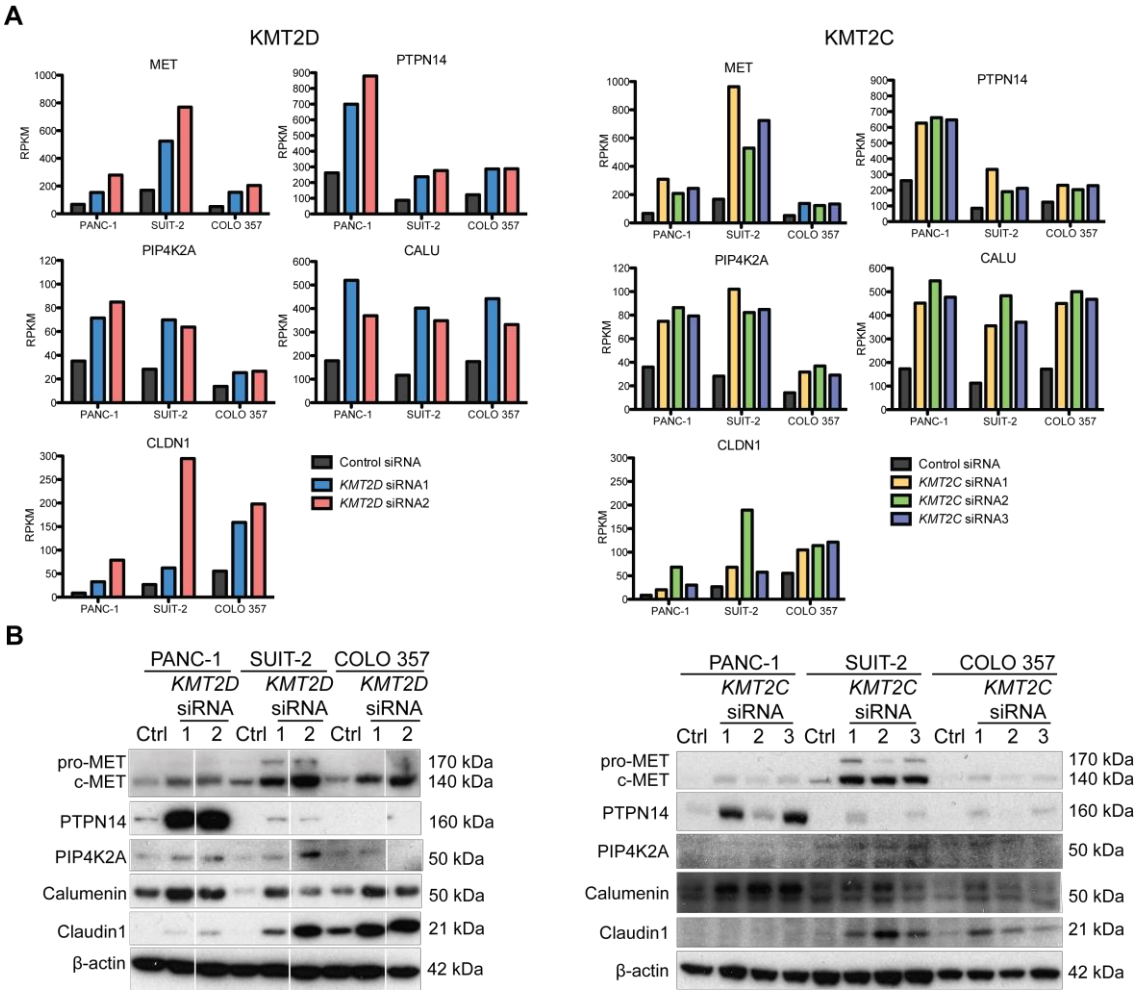

Figure 4.6 – Increased mRNA and protein expression for five genes selected from the 19 commonly DE genes A, Changes in the RPKM values for five of the 19 commonly DE genes (*MET*, *PTPN14*, *PIP4K2A*, *CALU* and *CLDN1*), each increasing upon treatment with *KMT2D* or *KMT2C* siRNAs. **B,** Western blot analysis to validate the expressional changes of the five DE genes following transfection with the two KMT2D siRNAs, the three KMT2C siRNAs, or the scramble negative control siRNA. A third siRNA was cropped from the images between KMT2D siRNA1 and 2. Pro-MET = Uncleaved-MET, c-MET = Cleaved-MET.

4.2.2.2 Decreased expression of ABCB1

In addition to the five common DE genes described above, *ABCB1* was also selected for validation. This gene was selected because, unlike the other genes, its expression was decreased by the loss of *KMT2D*, and thus is more likely to be a direct targeted of H3K4 methylation. *ABCB1* encodes an ATP-dependent efflux pump called ATP-binding cassette sub-family B member 1 (ABCB1, also known as P-glycoprotein 1 (Pgp), multidrug resistance protein 1 (MDR1), or CD243), and has been linked with increased drug resistance in a range of cancers (278), including pancreatic cancer (279). The expression of *ABCB1* was specific to the two cell lines of metastatic origin (SUIT-2 and COLO 357), where its decreased expression upon *KMT2D* silencing was particularly evident in the SUIT-2 cell line (Figure 4.7).

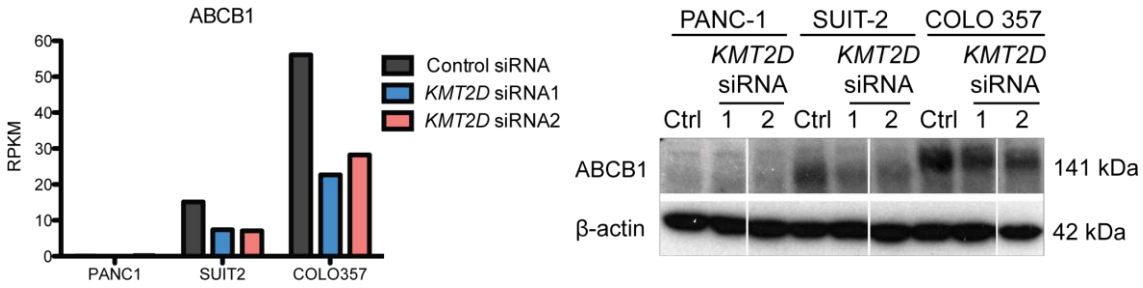

Figure 4.7 – Decreased mRNA and protein expression of ABCB1 upon KMT2D depletion. RNA-seq RPKM data and western blot analysis for the decrease in *ABCB1* expression across the cell lines following treatment with *KMT2D* siRNAs. A third siRNA was cropped from the images.

4.2.3 GSEA of differentially expressed genes

In light of the overlapping data between the transcriptomes for the KMT2C and KMT2D depleted cell lines, GSEA was employed to explore the significant differences between the targeted and control siRNAs in all curated canonical pathways (Table 4.8). In support of the proliferation and cell-cycle studies (Chapter 3), the following six pathways were significantly downregulated (FDR < 0.05) compared to controls in all siRNA experiments; cell-cycle, cell-cycle mitotic, DNA replication, DNA repair, mitotic M-M/G1-phases and Fanconi pathways, all of which relate to cell-cycle and DNA maintenance (Table 4.8). In addition to these, several others relating to cell-cycle checkpoints and cell apoptosis were identified as being affected, however significance was not reached in all five of the pair-wise comparisons.

Moreover, the telomere maintenance, meiotic recombination, and chromosome maintenance pathways, were each significantly downregulated in the *KMT2D* siRNA datasets, whilst also demonstrating an overall less pronounced trend towards downregulation in the *KMT2C* siRNA dataset (Figure 4.8). Interestingly, the translation pathway was differentially altered by each methyltransferase, where its upregulation was associated with KMT2D depletion, and conversely downregulated upon *KMT2C* depletion. In addition, data suggests that other pathways, such as the packaging of telomere ends, processing of capped intron containing pre-mRNA, and mRNA processing pathways may also be differentially altered by the methyltransferases (Figure 4.8). Taken together, these differences may highlight different, non-redundant roles for the two H3K4 methyltransferases.

Figure 4.8 – Heatmap to depict the normalized enrichment score (NES) for the significantly enriched pathways upon KMT2C or KMT2D depletion as determined by GSEA. Bioinformatic analysis of the DE genes produced NES for the most enriched pathways following the depletion of KMT2D or *KMT2C* by siRNAs. A negative NES indicates a downregulation in siRNA treated samples in relation to control (blue), and the opposite for a positive NES (red). Pathways significantly downregulated by all examined siRNAs are noted in red text.

4.2.4 Pathway analysis of PDAC patient GEP data confirms changes in cell-cycle

To assess whether these pathways could also be found in the clinical data biological process enrichment was performed on the genes that correlated strongly (Pearson's correlation $p < 0.001$) with *KMT2C/D* expression levels from the ICGC and TCGA human pancreatic cancer datasets (see 3.2.1). In this way, for the genes that correlated positively with combined *KMT2C/D* expression the PANTHER classification system identified a significant enrichment for the cell-cycle (adjusted (adj.) $p = 1.45\text{x}10^{-2}$ for ICGC dataset (Table 4.4); adj. $p = 1.02\text{x}10^{-6}$ for TCGA (Table 4.5)), mitosis (adj. $p = 1.62\text{x}10^{-3}$ for TCGA (Table 4.5)), and DNA repair (adj. $p = 1.61\text{x}10^{-2}$ for TCGA (Table 4.5)) pathways.

In addition to these commonly downregulated pathways, translation was the most overrepresented biological pathway for these genes, where expression showed significant negative correlation with *KMT2C/D* (adj. $p = 2.21\text{x}10^{-4}$ for ICGC (Table 4.6) and adj. $p = 5.76\text{x}10^{-30}$ for TCGA (Table 4.7)).

Overall, the pathways identified from the correlations with *KMT2C/D* expression in clinical datasets, were consistent with the most significantly downregulated, and upregulated, pathways from in the *in vitro* RNA-seq experiments (Figure 4.8).

PANTHER GO-Slim Biological Process	Homo sapiens - REFLIST (20814)	Client Text Box Input (1170)	Client Text Box Input (expected)	Client Text Box Input (over/under)	Client Text Box Input (fold Enrichment)	Client Text Box Input (p-value)
primary metabolic process	6825	516	383.65	+	1.34	$1.14x10^{-13}$
nucleobase-containing compound metabolic process	3467	299	194.89	+	1.53	$1.89x10^{-12}$
metabolic process	8247	591	463.58	+	1.27	$6.95x10^{-12}$
RNA metabolic process	2360	216	132.66	+	1.63	$1.40x10^{-10}$
transcription from RNA polymerase II promoter	1723	155	96.85	+	1.6	$1.45x10^{-6}$
transcription, DNA-dependent	1941	169	109.11	+	1.55	$2.57x10^{-6}$
regulation of biological process	2846	223	159.98	+	1.39	$4.17x10^{-5}$
organelle organization	571	61	32.1	+	1.9	$5.53x10^{-4}$
biological regulation	3918	283	220.24	+	1.28	$7.01x10^{-4}$
regulation of nucleobase-containing compound metabolic process	1700	138	95.56	+	1.44	$2.49x10^{-3}$
regulation of transcription from RNA polymerase II promoter	1319	109	74.14	+	1.47	$1.11x10^{-2}$
protein metabolic process	2692	198	151.32	+	1.31	$1.16x10^{-2}$
cell cycle	1107	94	62.23	+	1.51	$1.45x10^{-2}$
mRNA splicing, via spliceosome	183	25	10.29	+	2.43	$1.47x10^{-2}$
cellular protein modification process	1317	108	74.03	+	1.46	$1.62x10^{-2}$
cellular component organization or biogenesis	1316	107	73.98	+	1.45	$2.41x10^{-2}$
phosphate-containing compound metabolic process	910	78	51.15	+	1.52	$4.57x10^{-2}$

Table 4.4 – Overrepresented biological pathways for genes with expression strongly positively correlated ($p < 0.001$) with that of KMT2C and KMT2D in the ICGC dataset. Table contains data for the number of genes, the fold enrichment, and the probability for each of the overrepresented biological pathways for the list of genes that positively correlated ($p < 0.001$) with *KMT2C/D* expression in the ICGC dataset. p-values were adjusted using Bonferroni correction.

PANTHER GO-Slim Biological Process	Homo sapiens - REFLIST (20814)	Client Text Box Input (3337)	Client Text Box Input (expected)	Client Text Box Input (over/under)	Client Text Box Input (fold Enrichment)	Client Text Box Input (p-value)
transcription, DNA-dependent	1941	600	311.19	+	1.93	1.10×10^{-51}
nucleobase-containing compound metabolic process	3467	895	555.85	+	1.61	4.50×10^{-47}
RNA metabolic process	2360	669	378.37	+	1.77	1.29×10^{-45}
regulation of nucleobase-containing compound metabolic process	1700	512	272.55	+	1.88	3.47×10^{-40}
transcription from RNA polymerase II promoter	1723	499	276.24	+	1.81	8.43×10^{-35}
metabolic process	8247	1676	1322.2	+	1.27	4.41×10^{-33}
primary metabolic process	6825	1430	1094.22	+	1.31	1.13×10^{-31}
biological regulation	3918	914	628.15	+	1.46	2.64×10^{-31}
cellular process	6708	1371	1075.46	+	1.27	9.93×10^{-25}
regulation of transcription from RNA polymerase II promoter	1319	372	211.47	+	1.76	4.62×10^{-23}
regulation of biological process	2846	663	456.28	+	1.45	8.04×10^{-21}
nitrogen compound metabolic process	1099	302	176.2	+	1.71	8.74×10^{-17}
biosynthetic process	764	210	122.49	+	1.71	2.97×10^{-11}
cellular protein modification process	1317	318	211.15	+	1.51	1.60×10^{-10}
protein phosphorylation	603	164	96.68	+	1.7	3.20×10^{-8}
regulation of catalytic activity	1073	256	172.03	+	1.49	9.88×10^{-8}
regulation of molecular function	1096	260	175.72	+	1.48	1.21×10^{-7}
cell communication	3006	604	481.94	+	1.25	7.17×10^{-7}
cell cycle	1107	257	177.48	+	1.45	1.02×10^{-6}
catabolic process	407	115	65.25	+	1.76	2.54×10^{-6}
phosphate-containing compound metabolic process	910	201	145.9	+	1.38	1.22×10^{-3}
mitosis	367	95	58.84	+	1.61	1.62×10^{-3}
protein transport	1082	227	173.47	+	1.31	7.97×10^{-3}
intracellular protein transport	1052	220	168.66	+	1.3	1.26×10^{-2}
vesicle-mediated transport	895	191	143.49	+	1.33	1.36×10^{-2}
DNA repair	172	50	27.58	+	1.81	1.61×10^{-2}
organelle organization	571	128	91.55	+	1.4	3.30×10^{-2}

Table 4.5 - Overrepresented biological pathways for genes with expression strongly positively correlated ($p < 0.001$) with that of KMT2C and KMT2D in the TCGA dataset Table contains data for the number of genes, the fold enrichment, and the probability for each of the overrepresented biological pathways for the list of genes that positively correlated ($p < 0.001$) with KMT2C/D expression in TCGA dataset. p-values were adjusted using Bonferroni correction.

PANTHER GO-Slim Biological Process	Homo sapiens - REFLIST (20814)	Client Text Box Input (664)	Client Text Box Input (expected)	Client Text Box Input (over/under)	Client Text Box Input (fold Enrichment)	Client Text Box Input (p-value)
translation	435	35	13.88	+	2.52	2.21×10^{-4}
immune system process	1391	72	44.38	+	1.62	9.99×10^{-3}

Table 4.6 - Overrepresented biological pathways for genes with expression strongly negatively correlated ($p < 0.001$) with that of KMT2C and KMT2D in the ICGC dataset. Table contains data for the number of genes, the fold enrichment, and the probability for each of the overrepresented biological pathways for the list of genes that negatively correlated ($p < 0.001$) with KMT2C/D expression in the ICGC dataset. p-values were adjusted using Bonferroni correction.

PANTHER GO-Slim Biological Process	Homo sapiens - REFLIST (20814)	Client Text Box Input (2305)	Client Text Box Input (expected)	Client Text Box Input (over/under)	Client Text Box Input (fold Enrichment)	Client Text Box Input (p-value)
translation	435	149	48.17	+	3.09	5.76×10^{-30}
protein metabolic process	2692	439	298.12	+	1.47	1.98×10^{-14}
metabolic process	8247	1106	913.3	+	1.21	5.64×10^{-14}
oxidative phosphorylation	56	31	6.2	+	5	2.19×10^{-10}
rRNA metabolic process	115	40	12.74	+	3.14	1.56×10^{-7}
primary metabolic process	6825	885	755.82	+	1.17	1.90×10^{-6}
cellular component biogenesis	310	68	34.33	+	1.98	4.48×10^{-5}
regulation of translation	148	40	16.39	+	2.44	1.16×10^{-4}
generation of precursor metabolites and energy	274	61	30.34	+	2.01	1.16×10^{-4}
respiratory electron transport chain	217	51	24.03	+	2.12	2.15×10^{-4}
protein folding	155	39	17.17	+	2.27	8.41×10^{-4}
mitochondrion organization	40	15	4.43	+	3.39	1.34×10^{-2}
mitochondrial transport	25	11	2.77	+	3.97	3.24×10^{-2}

Table 4.7 - Overrepresented biological pathways for genes with expression strongly negatively correlated ($p < 0.001$) with that of KMT2C and KMT2D in the TCGA dataset. Table contains data for the number of genes, the fold enrichment, and the probability for each of the overrepresented biological pathways for the list of genes that negatively correlated ($p < 0.001$) with KMT2C/D expression in the TCGA dataset. p-values were adjusted using Bonferroni correction.

4.2.4.1 Identification of potentially implicated genes

Among the 94 and 257 cell-cycle genes that showed significant positive correlation with combined *KMT2C/D* expression from the respective ICGC (Table 4.4) and TCGA (Table 4.5) datasets, three (*NCAPD3*, *CDKL1* and *EIF2AK4*) were significantly downregulated in at least one *KMT2C/D* siRNA experiment. Each of the three genes showed significant positive correlations with the expression of *KMT2C/D* in both the ICGC and TCGA datasets (*NCAPD3*: $r = 0.36$, $p = 5.56\text{x}10^{-4}$ and $r = 0.38$, $p = 5.19\text{x}10^{-5}$; *CDKL1*: $r = 0.33$, $p = 1.86\text{x}10^{-3}$ and $r = 0.57$, $p = 1.36\text{x}10^{-10}$; *EIF2AK4*: $r = 0.27$, $p = 1.04\text{x}10^{-2}$ and $r = 0.36$, $p = 1.13\text{x}10^{-4}$, respectively (Table 4.8 and Figure 4.9). Using cBioPortal (280,281) the individual gene expression correlation between KMT2C, KMT2D and these three genes could also be investigated in the TCGA dataset. For KMT2C a positive correlation could be found with KMT2D ($r = 0.69$), NCAPD3 ($r = 0.38$) and CDKL1 ($r = 0.46$), where only the correlation between KMT2D and CDKL1 $r = 0.39$) was shown (Figure 4.10). Data for the other gene comparisons could not be distinguished as cBioPortal only shows data for genes with a Pearson's correlation of greater than 0.3 or less than -0.3.

We next tested the ICGC and TCGA datasets to examine whether the expression of each of these genes correlated significantly with patient outcome. For each of the three genes, the patients with low-level expression had significantly better OS rates, however for *EIF2AK4* this only reached the $p < 0.1$ threshold, rather than $p < 0.05$, in the ICGC dataset (Figure 4.11 and Figure 4.12).

Given the striking correlations in these datasets, we went on to test these associations in two additional human pancreatic cancer cohorts, the Stratford and BCI_Zhang_merged datasets (223). For *NCAPD3*, both datasets showed significant correlations (log-rank $p = 0.042$ for Stratford dataset and $p = 2.12\text{x}10^{-4}$ for BCI_Zhang_merged dataset) (Figure 4.12). The candidacy of both *CDKL1* and *EIF2AK4* was not supported in these additional cohorts, where low *EIF2AK4* expression showed only a marginally significant correlation with improved survival in the BCI_Zhang_merged dataset (Figure 4.12). In fact, the opposite was observed for *CDKL1*, where lower expression correlated significantly with a worsened survival in the BCI_Zhang_merged dataset (Figure 4.12).

Interestingly, for *NCAPD3* a significant downregulation of expression was observed in our *KMT2D* siRNA datasets, but not the *KMT2C* siRNA datasets, when compared to

control (*KMT2D* siRNA1, FDR = 0.0299; *KMT2D* siRNA2, FDR = $1.37x10^{-18}$) (Table 4.8 and Figure 4.13A). To examine whether these changes were translated to the protein level, western blots were performed to confirm that reduced NCAPD3 expression was an effect specific to KMT2D, and not also KMT2C (Figure 4.13B). Whilst this reduction in expression was true for both *KMT2D* siRNAs, it is worth pointing out that the effect of *KMT2D* siRNA2 was far greater than that of *KMT2D* siRNA1 across the three cell lines.

Collectively, the results from these experiments suggest that KMT2C and KMT2D have roles in regulating the expression of a wide range of genes, but especially those relating to the cell-cycle and DNA maintenance. This in turn impacts on clinical outcome in patients. Of the DE genes, *NCAPD3* emerges as being a good candidate for further investigation, since its expression is preferentially decreased upon KMT2D depletion. Furthermore, it holds an important role within the cell-cycle, and is a remarkable predictor for improved PDAC outcome.

Name	cor. TCGA	corp. TCGA	cor. ICGC	corp. ICGC	KMT2D siRNA1 vs control		KMT2D siRNA2 vs control		KMT2C siRNA1 vs control		KMT2C siRNA2 vs control		KMT2C siRNA3 vs control	
					logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR	logFC	FDR
NCAPD3	0.3792	$5.19x10^{-5}$	0.3627	$5.56x10^{-4}$	-0.5321	0.0299	-1.7336	$1.37x10^{-18}$	-0.0314	0.9544	-0.1932	0.3899	-0.1831	0.5751
CDKL1	0.5687	$1.36x10^{-10}$	0.3290	$1.86x10^{-3}$	-1.0020	0.0033	-0.5702	0.1139	-0.2745	0.6380	-0.2783	0.4768	-0.2660	0.6495
EIF2AK4	0.3630	$1.13x10^{-4}$	0.2736	0.0104	-0.0146	0.9728	-1.1007	$4.61x10^{-9}$	0.1442	0.7135	0.1113	0.6900	-0.0264	0.9643

Table 4.8 – Genes that are significantly positively correlated with the combined KMT2C/D expression, which are also significantly downregulated in at least one KMT2C/D siRNA experiment. Table contains data for three genes that showed significant positive correlation with combined *KMT2C/D* expression, and show a significant downregulation in expression for at least one *KMT2C/D* siRNA experiment. Data shown for each gene include the expression correlation (cor.) and probability (corp.) within each of the ICGC and TCGA datasets, and the logFC (fold change) and false discovery rate (FDR) for each of the three genes in the siRNA experiments.

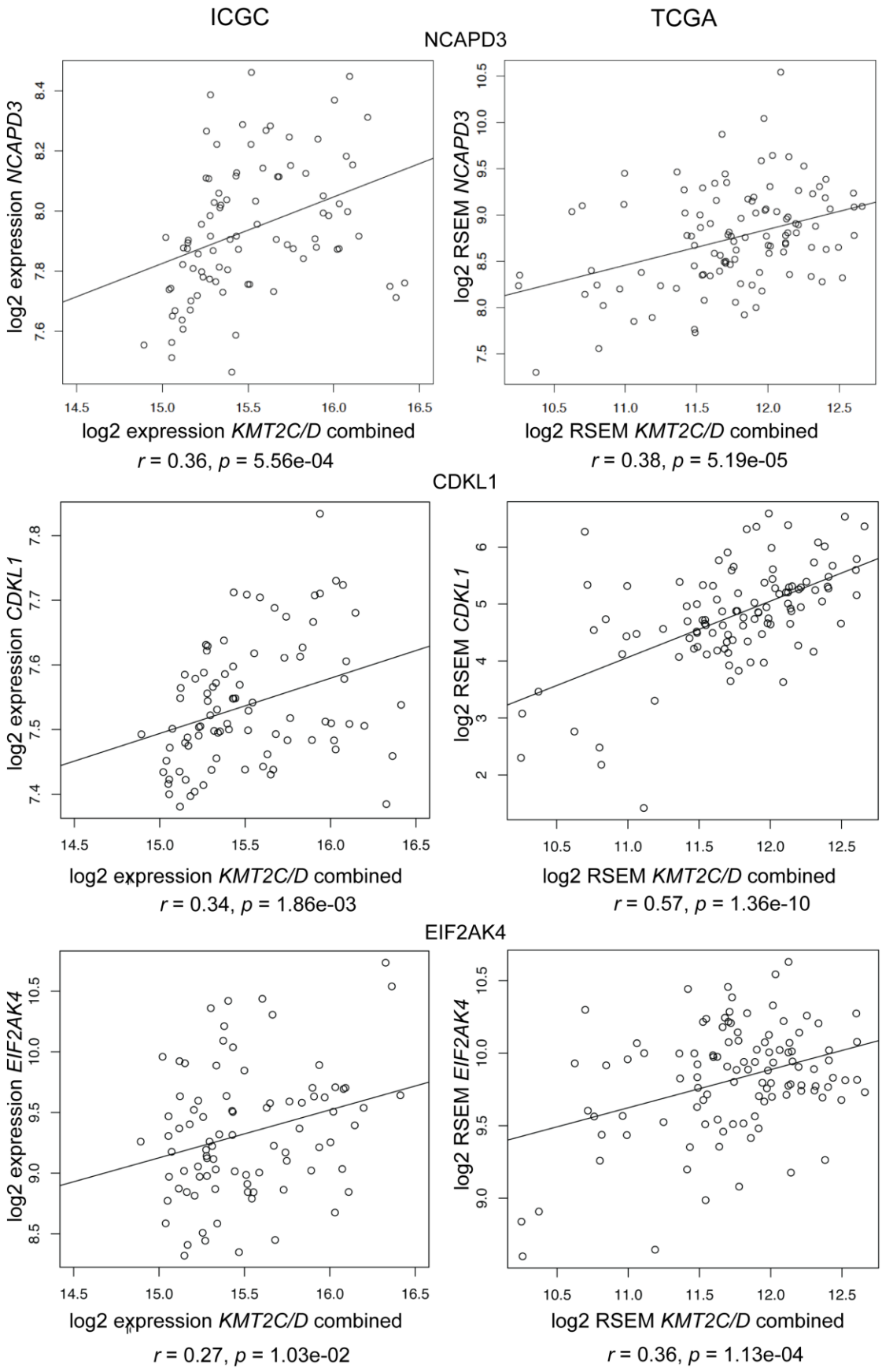

Figure 4.9 – Correlation of gene expression for *NCAPD3*, *CDKL1*, and *EIF2AK4* with *KMT2C/D* Significant positive correlations between the three genes (*NCAPD3, CDKL1* and EIF2AK4) and *KMT2C/D* combined expression levels for both the ICGC and TCGA datasets. The Pearson's correlation coefficient and associated p-values are shown for each set.

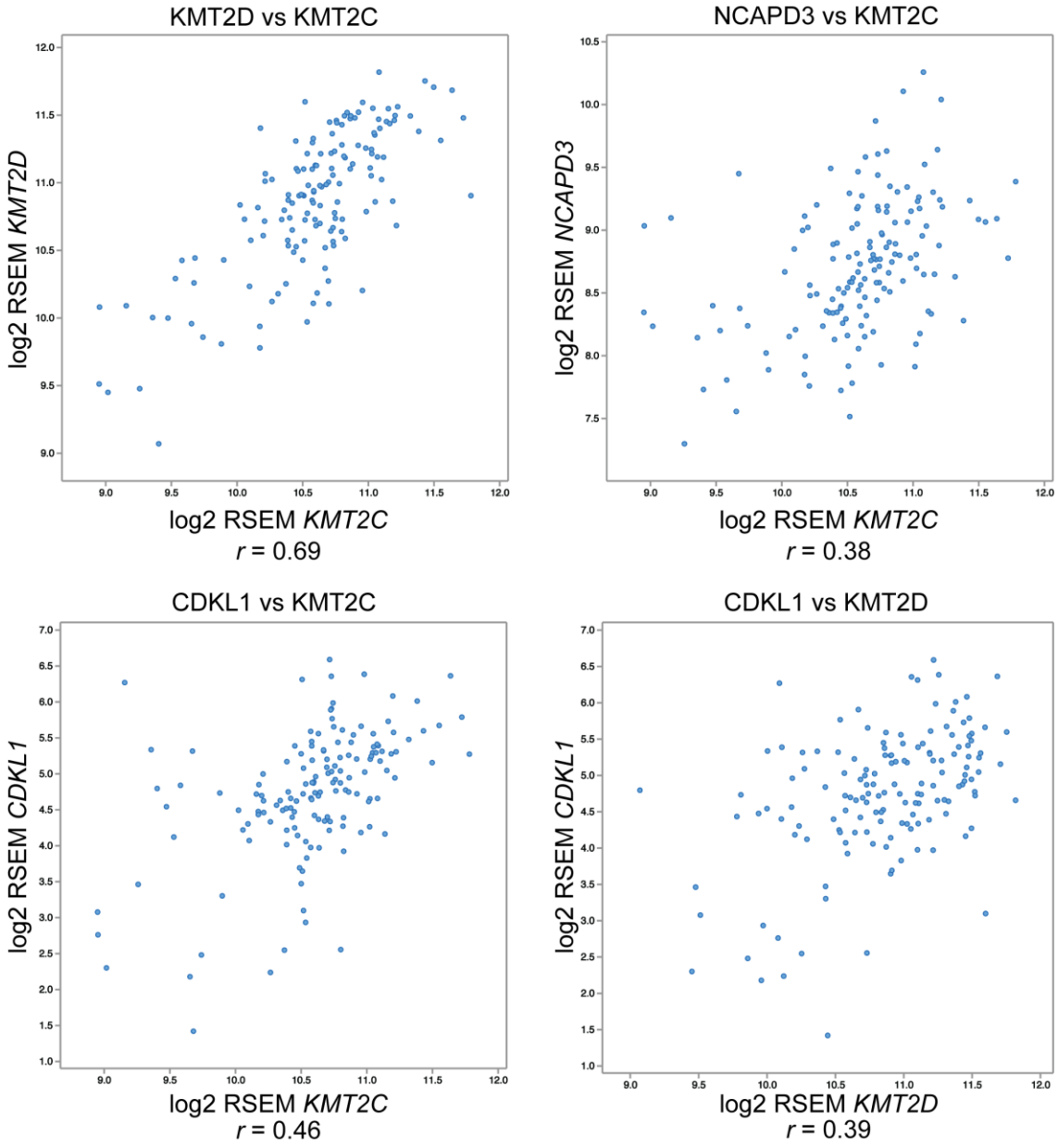

Figure 4.10 – Correlation of individual gene expression for *NCAPD3* or *KMT2D* with *KMT2C*, and *CDKL1* with *KMT2C* or *KMT2D*. Using cBioPortal ((280,281)) for the TCGA provisional dataset highlighted significant positive correlations between these individual genes. The Pearson's correlation coefficient is shown for each. Due to availability only data for the TCGA could be interrogated (and not the ICGC), where only the genes with a Pearson's correlation of greater than 0.3 or less than -0.3 could be obtained.

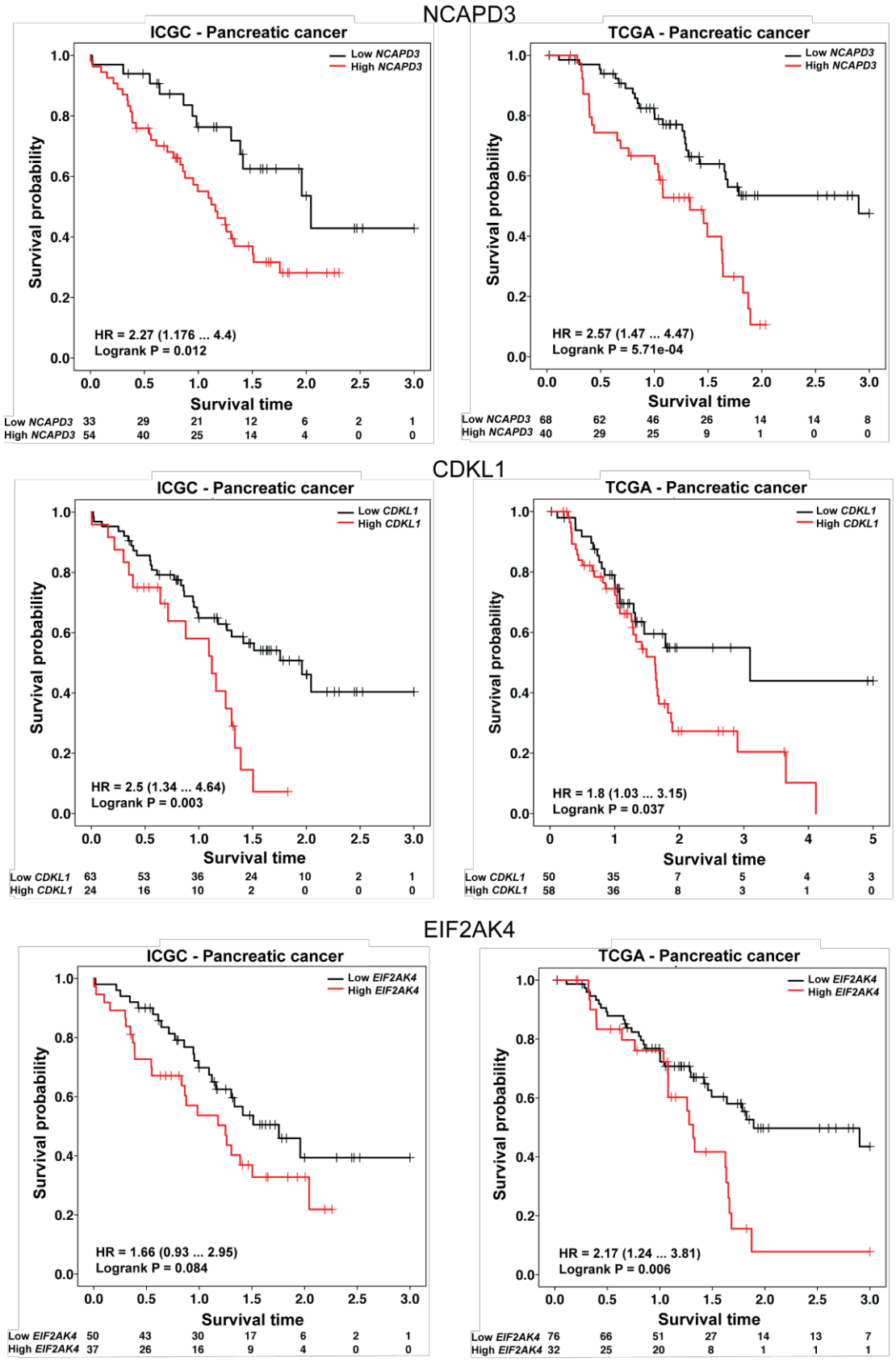

Figure 4.11 – Correlation of *NCAPD3*, *CDKL1*, and *EIF2AK4* expression with overall survival in the ICGC and TCGA datasets. KM survival analysis of PDAC tumours from the ICGC (*left*) and TCGA (*right*) datasets show significant negative correlations of patient survival with high (red) and low (black) expression of *NCAPD3* (top), *CDKL1* (middle) and *EIF2AK4* (bottom). Numbers on the x-axis represent years. HR = Hazard ratio.

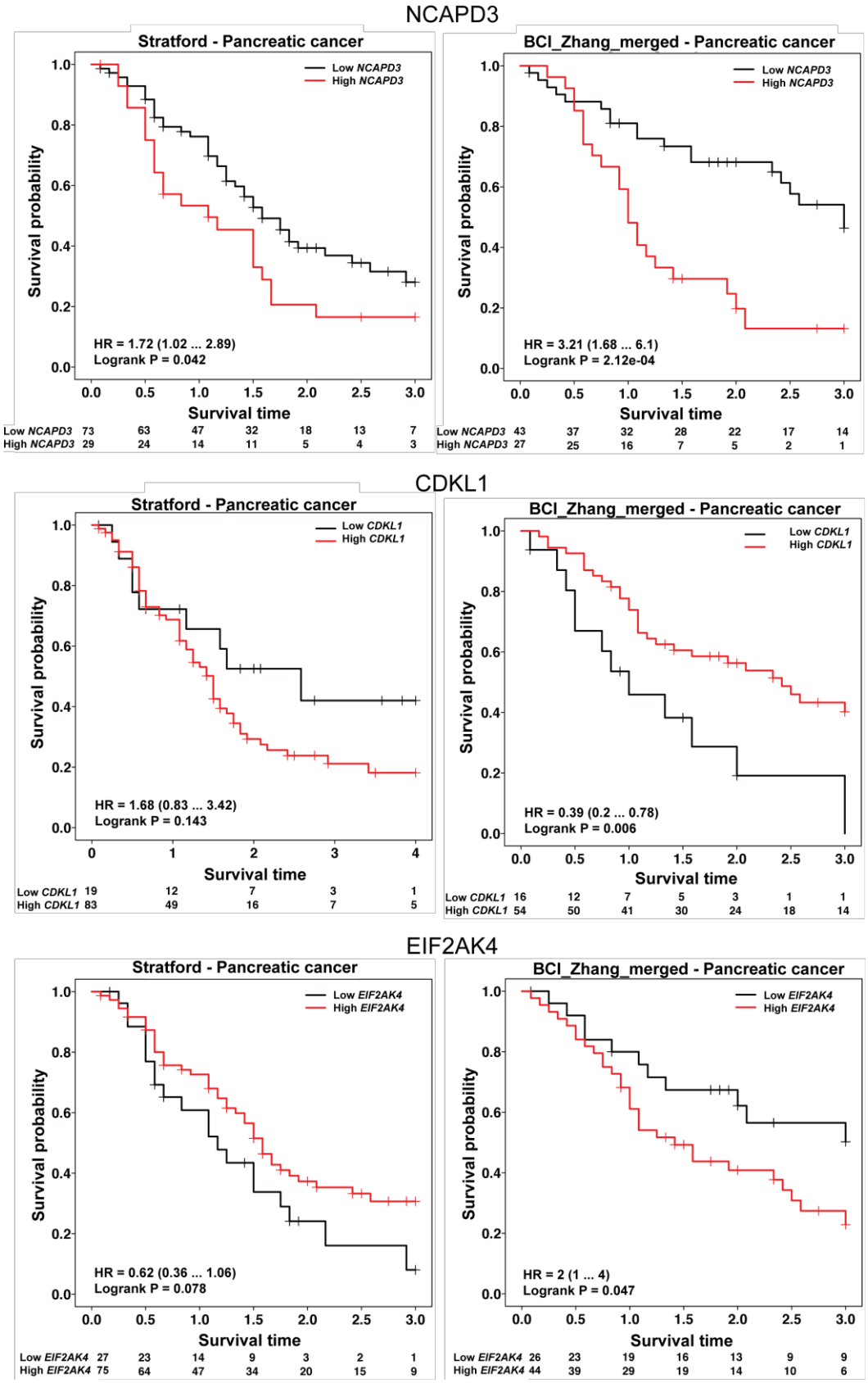

Figure 4.12 – Correlation of *NCAPD3*, *CDKL1* and *EIF2AK4* expression with overall survival in the Stratford and BCI_Zhang_merged datasets. KM survival analysis of PDAC tumours from the Stratford (*left*) and BCI_Zhang_merged (*right*) datasets show negative correlations of patient survival with high (red) and low (black) expression of *NCAPD3* (top), *CDKL1* (middle) and *EIF2AK4* (bottom). Numbers on the x-axis represent years. HR = Hazard ratio.

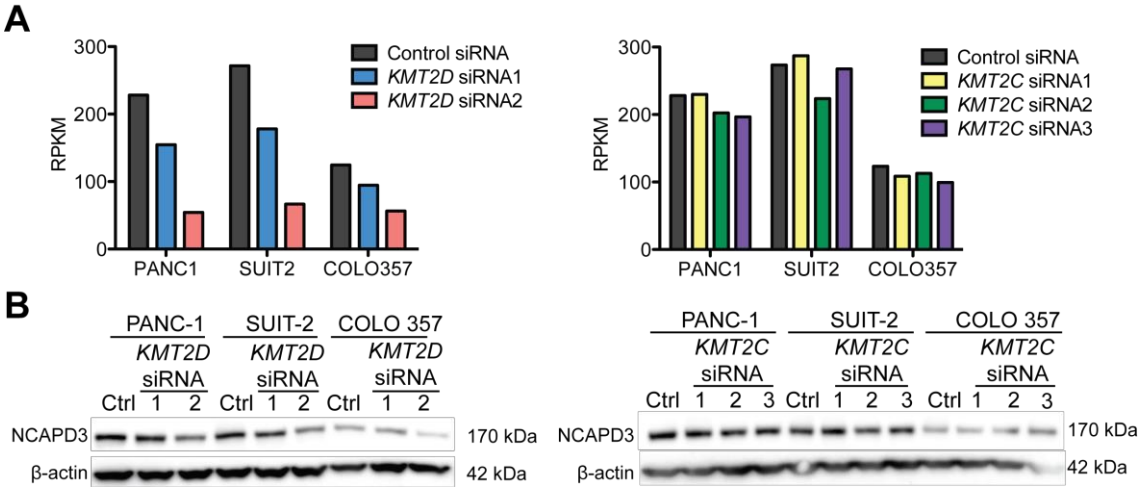

Figure 4.13 – Decreased mRNA and protein expression of NCAPD3 following depletion of KMT2C or KMT2D. A, Changes in the RNA-seq RPKM data for NCAPD3 showing a decrease in expression following the depletion of KMT2D, but not KMT2C, by targeted siRNAs. **B,** Western blot analysis to validate the expression change for NCAPD3 at the protein level upon the depletion of KMT2D, but not KMT2C, by targeted siRNAs.

4.3 Discussion

To explore the potential downstream effectors of KMT2C and KMT2D, and the pathways affected by their loss in pancreatic cancer, the transcriptome changes associated with their depletion were examined. Although not in pancreatic cancer, an earlier study Issaeva *et al.* had performed microarray analysis on HeLa cells that had undergone retroviral transduction with *KMT2D* shRNA vectors (138). In a more recent direct approach to examine genes targeted by the KMT2D complex, Guo *et al.* combined the use of both chromatin immunoprecipitation DNA sequencing (ChIP-seq) and gene expression microarray analysis techniques (248). These experiments were performed in colorectal carcinoma cells (HCT116, which harbours a homozygous *KMT2C* inactivating frameshift mutation (282)) somatically targeted with a recombinant adeno-associated virus (rAAV) vector to either express *KMT2D-Flag* (for ChIP-seq), or to create a *KMT2D*-null cell line (248). Both of these microarray based studies showed that genes regulated by KMT2D were associated with a diverse range of cell functions and pathways. Importantly, and perhaps most interestingly, Guo *et al.* note that the KMT2D complex's binding loci, and its specific functions, are likely to be "exogenous signalling- and/or cellular context-dependent" (248) and so in this chapter the roles of *KMT2C* and *KMT2D* were analysed by RNA-seq in PDAC cell lines.

For the RNA-seq analysis, RNA samples were generated from the PANC-1 SUIT-2 and COLO 357 cell lines at 48 hours post-transfection with siRNA. To reflect the highly metastatic and mutationally diverse nature of PDAC, the three cell lines were selected so as to include lines derived from both primary and metastatic sites, with a range of gene mutations (Table 3.5). PANC-1 was chosen as it originates from a primary tumour (208), whereas SUIT-2 and COLO 357 originate from the metastatic sites (liver (212) and celiac axis lymph node (214), respectively) of PDAC patients. Of those available only cell lines that harboured an oncogenic G12D *KRAS* mutation were selected since this mutation is found in almost all PDAC patient tumours (40). Capan-2 was excluded for consideration as a result of the ambiguous reports on its *TP53* and *CDKN2A* statuses, as discussed in the previous chapter. CFPAC-1 was also excluded due to its potentially confounding CF origins (21). A time point of 48 hours post-transfection was selected for generating the RNA samples in an attempt to minimise the capture of secondary transcriptional effects resulting from reduced proliferation/increased apoptosis described in the previous chapter. Upon confirming the depletion of each methyltransferase for the siRNAs, potential compensatory and off-target effects were also noted, whereby mRNA expression was increased for the non-targeted KMT2 methyltransferase, especially *KMT2C* upon *KMT2D* depletion (Figure 4.1). This is perhaps not surprising given the homology between the two genes and the fact that they form complexes that are identical to one another, where they are proposed to have similar roles in H3K4 methylation.

In order to identify robust gene signatures, we focused primarily on the changes consistent across the three cell lines and siRNA experiments. These RNA-seq experiments highlighted changes in 1202 genes for *KMT2D* siRNAs (124 common to both siRNAs, Figure 4.5) and 790 genes for *KMT2C* siRNAs (31 common to the three siRNAs, Figure 4.4) (Figure 4.3). The more striking phenotypic effect of reduced KMT2D (Chapter 3) was concordant with a higher number of DE genes upon depletion of this methyltransferase. Interestingly, for both methyltransferases the majority of the common DE genes were increased upon their depletion (KMT2D, 84 up and 40 down; KMT2C, 27 up and four down). This is unexpected considering the link between H3K4 methylation and active gene expression and contrasts with that found by Issaeva *et al.* and Guo *et al.* (Issaeva *et al.* HeLa, 74 downregulated with only two genes upregulated (138); Guo *et al.* HCT116 cells, 301 downregulated with 51 genes upregulated (248)). As was also suggested by Guo *et al.,* these upregulated genes may be indirect, or secondary, effects of losing these important methyltransferases. The greater number of

these genes observed in our data relative to these studies may reflect the nature of the models and techniques used. For example, the RNA-seq technique used here is an unbiased approach to examine gene expression, whereas their use of a DNA microarray for gene expression is limited by number of gene probes examined and its reduced accuracy for lowly expressed genes. Both studies also used stably generated cell lines which survive in the absence of KMT2D; it is therefore possible that they do not exhibit such drastic compensatory changes as an acute depletion of these methyltransferases might trigger, particularly where the PDAC cell lines used here are more dependant on their expression. In addition this may be a feature of our stringent analysis, where we focused in on changes common to three cell lines for all siRNAs, which may have filtered out more subtle decreases in gene expression.

Of the 19 DE genes commonly upregulated by KMT2C and KMTD siRNAs, perhaps the most notorious is the proto-oncogene *MET* (Figure 4.6). In a study by Di Renzo *et al.* it was shown that c-MET is upregulated in PDAC compared to the adjacent "normal" tissue (262). In more recent studies it has been shown that MET expression is associated with poor prognosis and survival in multiple cancer types (254,260,261). In the study by Di Renzo *et al.,* MET was highly expressed in PDAC cell lines, especially those that show a high degree of differentiation, where stimulation of these cells with HGF resulted in increased cell proliferation, motility and invasion (262). From this we currently speculate that the increased expression of *MET* may be an indirect result to counteract and compensate for the negative effects of KMT2D depletion on proliferation.

As with MET, *PIP4K2A* has also been proposed to have pro-tumorigenic roles, where its depletion is shown to abrogate the clonogenic potential of AML cells (283). Jude *et al.* showed that this was associated with an accumulation of the cyclin-dependent kinase inhibitors CDKN1A ($p21^{CIP1/WAF1}$) and CDKN1B ($p27^{KIP1}$), which cause G_1 arrest and apoptosis (283). This requirement of PIP4K2A agrees with our *in vitro* findings. In this way, as described with MET above, this inverse relationship in expression may be indicative of compensatory mechanisms.

Unlike MET and PIP4K2A, the protein-tyrosine phosphatase PTPN14 has been reported to act as a tumour suppressor (263). In a previous study that performed microarray analysis of orthotopic xenograft tumours generated by implanting MiaPaca-2 cells into the pancreata of SCID (severe combined immunodeficiency) mice, the expression of PTPN14 was found to be significantly decreased at the tumour invasion

front and in liver metastasis, compared to the primary tumour (284). In addition, PTPN14 was found to be localised to the nucleus in low-density proliferating cells, but cytoplasmic in confluent conditions, where this nuclear translocation was also observed at the cell migration front following scratch wounding (267). This therefore suggests that PTPN14 may have a role in transcriptional regulation at sites of metastasis and invasion. The underlying mechanism may be similar to that proposed by Huang *et al.,* for the transcription factor YAP (Yes-associated protein) in modifying the sensitivity of ovarian cancer cells to various anti-cancer agents (268). In this study, the authors used immunoprecipitation mass spectrometry to show that PTPN14 complexes with YAP through its WW domains, and the PPXY domains of PTPN14, to negatively regulate YAP-associated transcription and sensitise cells to a number of therapeutic agents (268). From these experiments, and the negative associations for PTPN14 in metastasis (268,284), invasion (268,284), proliferation, and drug resistance (268), one possibility is that increasing PTPN14 expression, as shown for KMT2 methyltransferase depletion, might dampen some of these more aggressive features in PDAC.

Another protein that may have a role in invasion and metastasis is the tight junction protein Claudin-1 (285). Tight junction proteins are essential for maintaining intact membrane barriers between neighbouring epithelial cells. During cell migration, invasion, and metastasis, tumour cells dissociate from the bulk primary tumour mass by overcoming these tight junctions (286). Previously, in a study looking at the overexpression and knockdown of Claudin-1 in lung adenocarcinoma cell lines, Chao *et al.* showed that Claudin-1 expression suppressed migration and invasion *in vitro*, and metastasis *in vivo* using an intravenous SCID mouse model (285). In addition, there was a strong positive correlation of improved survival in lung adenocarcinoma patients with increased *CLDN1* expression at the mRNA and protein levels (285). A recent study in 258 patients was unable to replicate this however (287). A final point of interest from Chao *et al.* was that, within their microarray data, *CLDN1* overexpression resulted in a decrease in the expression of *KMT2D* (285), supporting an argument for their inverse regulation. In this regard, and with relation to our observations, increased Claudin-1 due to reduced KMT2C and KMT2D may be indicative of a diminished invasive, migratory, and metastatic cell phenotype, and therefore could be a factor in the associated increase in PDAC patient survival.

Within the list of 19 commonly DE genes upon KMT2C/D depletion we also identified an increase in Calumenin expression (Figure 4.6). Calumenin, a Ca^{2+} binding protein,

has been described to not only have a traditional intracellular role within the secretory pathway (271,273), but to also act extracellularly (255,273,276) to prevent ECM degradation by MMPs (255). Again, as with the overcoming of tight junction cell-cell contacts described above, this degradation of ECM is required for tumour invasion and metastasis. This potential role of Calumenin is further supported by Ding *et al.* who used proteome analysis to show that Calumenin expression was higher in hepatocellular carcinoma (HCC) cells with lower metastatic potential (288). In line with this, Calumenin was also found to be less in a metastatic cell line than a primary cell line generated from the same head and neck squamous cell carcinoma patient (289). Wang *et al.* also demonstrated using immunohistochemistry that Calumenin expression was less in HCC and pancreatic carcinoma tissues than normal tissue, where for HCC this was also much weaker in tumours of higher grades than lower grades (255). These studies suggest that increased Calumenin may also be protective against tumour invasion and metastasis.

Alongside these DE genes common to both methyltransferases, we also chose to investigate *ABCB1*. The observation that ABCB1 expression was decreased upon KMT2D (Figure 4.7), but not KMT2C, depletion was intriguing for several reasons. Firstly, this could be a mechanism by which reduced KMT2D resulted in stronger anti-proliferative (Figure 3.8 and Figure 3.9) and expression-associated survival predictions (Figure 3.1B), when compared to KMT2C (Figure 3.7 and Figure 3.1A, respectively). Secondly, given that the expression of *ABCB1* was decreased following methyltransferase depletion, it is more likely to be a direct target of the KMT2D complex. It is well established that *ABCB1,* and other members of the ATP-binding cassette (ABC) superfamily of transporters are implicated in the development of multidrug resistance, where its expression associates with resistance in many cancer types (278,279). Somewhat supporting this role for KMT2D in regulating ABC drug resistance genes, the study performed by Guo *et al.* showed that the ABC family members *ABCC2* and *ABCC3* were both downregulated upon loss of *KMT2D* (248). Interestingly, expression of ABCB1 was only detected in the metastatic cell lines, and not the primary cell lines. In agreement with our results, Walsh *et al.* found that a significantly higher proportion of malignant melanoma metastases expressed ABCB1, compared to the primary tumours (290). While requiring further experimentation, this may suggest that reduced expression of ABCB1 is indicative of either reduced metastatic potential, or of a less aggressive disease.

Despite this evidence that potentially implicates many of the 19 commonly DE genes in PDAC, together this gene signature did not predict prognosis in any of the four PDAC GEP datasets tested (ICGC, TCGA, Stratford, or BCI_Zhang_merged) based on consensus clustering and OS (data not shown). Moreover, patient clustering based on this 19-gene signature alone did not show a significant enrichment in high or low *KMT2C/D* expression (data not shown). This lack of an association may reflect the complex nature of examining many genes, where although some may be implicated in the reduced KMT2D survival benefit, others could have arisen as compensatory mechanisms. Taken together, observations for PTPN14, Claudin-1, Calumenin and ABCB1, suggests that reduced KMT2C and KMT2D may elicit a favourable phenotype, which perturbs the aggressive disease features of invasion, metastasis, and resistance. As with the 19-gene signature, the expression of these genes was also not associated with improved survival (data not shown). These results may reflect the inherent complexity when making direct comparisons between primary and cell line data, and might also represent secondary, or compensatory, mechanisms caused by complete loss of protein, a feature unlikely to be reflected physiologically.

In addition to testing the contribution of individual genes, GSEA was employed to determine which pathways were altered upon depletion of either KMT2C or KMT2D. This analysis highlighted significant changes (FDR < 0.05) in six pathways, with three directly relating to the cell-cycle, highlighting changes in the mitotic M- and G_1-phases (Figure 4.8). The remaining pathways related to DNA maintenance with DNA replication, DNA repair, and the Fanconi anaemia pathways also demonstrated significant decreases (Figure 4.8). These top downregulated pathways support the *in vitro* work (Chapter 3), which showed cell proliferation as being negatively impacted, due to a reduction in cell cycle, when KMT2D was depleted. In Chapter 3 the *in vitro* cell proliferation assay showed KMT2C depletion to cause a small, but variable and non-statistically significant, decrease in cell proliferation (Figure 3.7). Here, at the transcriptional level, GSEA indicated that genes relating to the cell-cycle were decreased in expression upon KMT2C depletion, in line with that for KMT2D depletion (Figure 4.8). In this way, although a functional role in cell cycle regulation cannot be excluded for KMT2C, it is possible some role still exists, albeit a functionally weaker one than KMT2D.

Other groups have also previously performed microarray analysis to determine differences in gene expression between cells with and without *KMT2D* (138,248). The

affected pathways identified through pathway analysis for both of these studies were not concordant with each other, or our GSEA findings. Issaeva *et al.,* stated that their top significantly downregulated genes were mainly implicated in cell adhesion, cytoskeleton reorganisation and transcriptional regulation (138). The pathways downregulated in the study by Guo *et al.* were mainly comprised of those involved in cAMP signalling, retinoic acid signalling, and B cell signalling (248). This discrepancy between the genes expressed and pathways altered may be indicative of the different cell types and systems used, especially considering our data were generated in transiently transfected cell lines that normally express KMT2C and KMT2D. However, even accounting for this difference, when we focused attention on genes that associated with *KMT2C/D* expression from the ICGC and TCGA PDAC datasets, a significant overrepresentation of cell-cycle genes was found to positively correlate with *KMT2C/D* gene expression (Table 4.4 and Table 4.5). In addition to the cell-cycle pathway, data from the TCGA also showed that both the mitosis and DNA repair pathways showed significant positive correlation with *KMT2C/D* expression providing additional confidence for the robustness of these experiments. From the GSEA for the RNA-seq data, genes associated with the translation pathway were increased upon KMT2D depletion, but intriguingly however, the converse was true for KMT2C depletion (Figure 4.8). In line with the KMT2D data, pathway analysis of the ICGC and TCGA data identified that the translation pathway also showed significant negative correlation with the expression of *KMT2C/D* (Table 4.6 and Table 4.7). As discussed above with the increased differential expression of the individual genes, we are working on the premise that this increased translation may either be a secondary effect of their reduction, or a compensatory mechanism for the proliferative stress caused by their depletion.

Next, to test if there was any other gene specific overlaps between the clinically annotated ICGC and TCGA data and our RNA-seq data, we compared the cell-cycle genes identified from the pathway analysis against our RNA-seq expression profiles. This comparison identified *NCAPD3*, *CDKL1* and *EIF2AK4*, where each showed concurrent direction and magnitude of fold change (Table 4.8). For each of these three genes, the PDAC clinical outcome associated with high and low expression for each gene was examined. Encouragingly, high and low expression of *NCAPD3* proved to be a good predictor of outcome in all four GEP series (ICGC, TCGA, Stratford, or BCI_Zhang_merged) where lower expression was associated with improved outcome (Figure 4.11 and Figure 4.12). For *CDKL1* and *EIF2AK4* the two additional datasets did

not support the associations of their decreased expression with improved outcome (Figure 4.12).

NCAPD3 encodes for the D3 subunit of the condensin II complex known as NCAPD3 (also known as CAP-D3, and previously designated KIAA0056 by the Human Unidentified Gene-Encoded (HUGE) project (291)). NCAPD3 is known to be involved in chromosome condensation and reorganisation (292) and its depletion has been demonstrated by an RNAi screen in HeLa cells to result in mitotic arrest, binucleated cells and abnormal spindle formation, but not cell death (293). This effect may be mediated by the phosphorylation of NCAPD3 at threonine 1415 and serine 1419 by Cdk1 and Plk1 respectively during the prophase, where this is thought to be required for chromatin condensation (294). The condensing II complex is recruited through the interaction of NCAPD3 with H4K20me1, which is regulated by the lysine demethylase PHF8 that in turn is recruited by H3K4 methylation (295). Thus there may be a distinct interaction between loss of H3K4 methylation upon methyltransferase depletion and this mechanism that impedes progression of cells through the cell-cycle.

Within our RNA-seq data, and in the protein validation that followed, a reduction of NCAPD3 expression was found across the three tested cell lines upon *KMT2D* depletion, but not KMT2C depletion (Figure 4.13). Due to the role of NCAPD3 and condensin II in chromosome organisation (292), it also is interesting to note that our GSEA pathway analysis highlighted a predominant role for KMT2D and not KMT2C in chromosome maintenance. Therefore a protein, like NCAPD3, that is involved in chromosome organisation, cell-cycle progression and is only decreased by KMT2D, is likely to be a good candidate for the more notable effects of KMT2D on cell proliferation and patient survival. For this reason we can postulate that the improved survival noted for lower *KMT2D* expression is, at least in part, indicative of its effect on reducing NCAPD3 expression. Despite these striking associations of decreased *NCAPD3* expression with improved PDAC outcome in all four datasets, the converse has previously been found for prostate cancer, where expression instead associated with improved outcome (296). Whilst in contrast with our observations in PDAC, this positive association in prostate cancer is surprising considering the critical role of this protein in cell-cycle progression (293), something also noted by the authors (296).

In summary, in this chapter we combined RNAi and RNA-seq to determine the transcriptome changes associated with KMT2D or KMT2C depletion in three PDAC cell lines. Our data show that while expression of many genes is altered upon the

depletion of these methyltransferases, a significant proportion are decreased in pathways relating to cell-cycle and DNA maintenance. Similar pathways also correlated with *KMT2C/D* expression in human PDAC patients, where comparisons to our *in vitro* data highlighted a potential role for *NCAPD3*. As NCAPD3 expression was diminished upon reduced KMT2D expression, and its expression negatively correlated with patient survival, this gene is a good candidate for the functional effects of reduced KMT2D.

Chapter 5 Kmt2d depletion in the KC and KPC mouse model systems

5.1 Background

In the previous chapters, functional and transcriptional data demonstrated that KMT2D, and to a lesser extent KMT2C, have roles in maintaining the cell-cycle progression and proliferation of human pancreatic cell lines. While this coincides with observations across a range of human solid cancer cell lines (55,138,151,202), few studies have been performed in murine models to examine the roles of *Kmt2d* and *Kmt2c* in solid cancer. GEMMs for *Kmt2d* loss (120,152,162) do exist although these have not been used thus far to investigate its *in vivo* role in solid cancers, with studies instead relying on human cell xenografts (151,202,297). Alongside these, GEMMs with global functionally null *Kmt2c* have also been created (120,132), whereby these mice develop ureter epithelial tumours, especially when combined with *Trp53* haploinsufficiency (133).

In this chapter, we set out to use mouse cell lines generated from the KC (68) and KPC (72) mouse models in RNAi experiments to elucidate the roles of Kmt2c and Kmt2d in *in vitro* models. Alongside this, KC and KPC mice were generated with pancreas specific loss of *Kmt2d* to begin preliminary *in vivo* studies into the role of this methyltransferase in PDAC.

5.2 Results

5.2.1 Depletion of Kmt2c and Kmt2d in mouse cell lines

In order to deplete *Kmt2c* expression, two unique siRNAs targeting exons 38 and 52 of the *Kmt2c* mRNA transcript (*Kmt2c* siRNA1 and *Kmt2c* siRNA2, respectively) were tested across three murine cell lines. Two of these pancreatic cell lines (DT6606 and DT6585) were derived from KC tumours (68), with the third (TB32043) from a KPC animal (72). Due to poor quality of Kmt2c antibodies for western blot, depletion of *Kmt2c* mRNA was examined by qRT-PCR. Both *Kmt2c* siRNAs resulted in a depletion of *Kmt2c* mRNA for the three cell lines, when compared to transfection with control siRNA (Figure 5.1A). Although *Kmt2c* knockdown by each siRNA was similar across the cell lines, *Kmt2c* siRNA2 appeared to elicit a far greater relative reduction in DT6585 cells than *Kmt2c* siRNA1 (88% compared to 43%, Figure 5.1A).

For Kmt2d knockdown, western blot analysis showed a clear reduction in protein was achieved in the three cell lines for two of the siRNAs targeted to *Kmt2d* mRNA (*Kmt2d* siRNA1 targeting exons 54 and 55, and *Kmt2d* siRNA2 targeting exon 40) (Figure 5.1B). A third siRNA targeting *Kmt2d* mRNA was also used (*Kmt2d* siRNA3 targeting

exons 46 and 47), although this siRNA was less efficient than siRNA1 and siRNA2, particularly in the TB32043 cell line (Figure 5.1B).

At 120 hours following siRNA transfection, WST-1 viability assays showed no changes upon *Kmt2c*, or Kmt2d, depletion (Figure 5.2). Although the cell number was not directly measured as previously in human cell lines (see 3.2.4), the assay used here does provide an indirect measure of proliferation over 72 hours since cells were harvested and the same number re-plated following the 48 hours of siRNA treatment. The absence of an effect on proliferation in these murine lines was also confirmed using BrdU incorporation experiments and cell-cycle analysis (Eleni Maniati, data not shown).

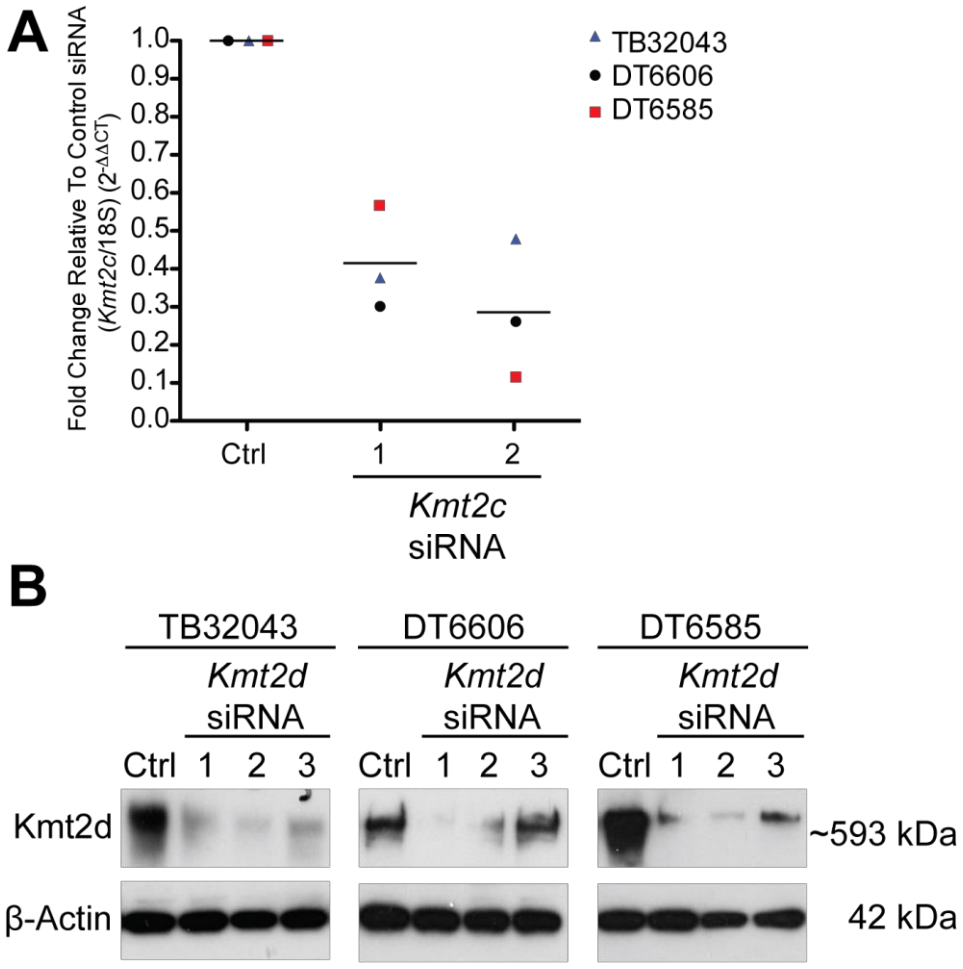

Figure 5.1 – *Kmt2c* and *Kmt2d* targeted siRNAs reduce expression of *Kmt2c* mRNA, and Kmt2d protein respectively. A, RT-qPCR analysis showing fold change in *Kmt2c* mRNA expression, relative to 18S, for the three cell lines (TB32043 (blue triangles), DT6606 (black circles) and DT6585 (red squares)) following transfection with two *Kmt2c* siRNAs, relative to treatment with untargeted control siRNA. Data shown are normalised mean values from technical triplicates. **B,** Western blot analysis showing that two *Kmt2d* siRNAs (*Kmt2d* siRNA1 and *Kmt2d* siRNA2) deplete Kmt2d levels in three cell lines (TB32043, DT6606 and DT6585), relative to treatment with untargeted control siRNA, whereas a third (*Kmt2d* siRNA) is less effective. β-Actin was used as a loading control.

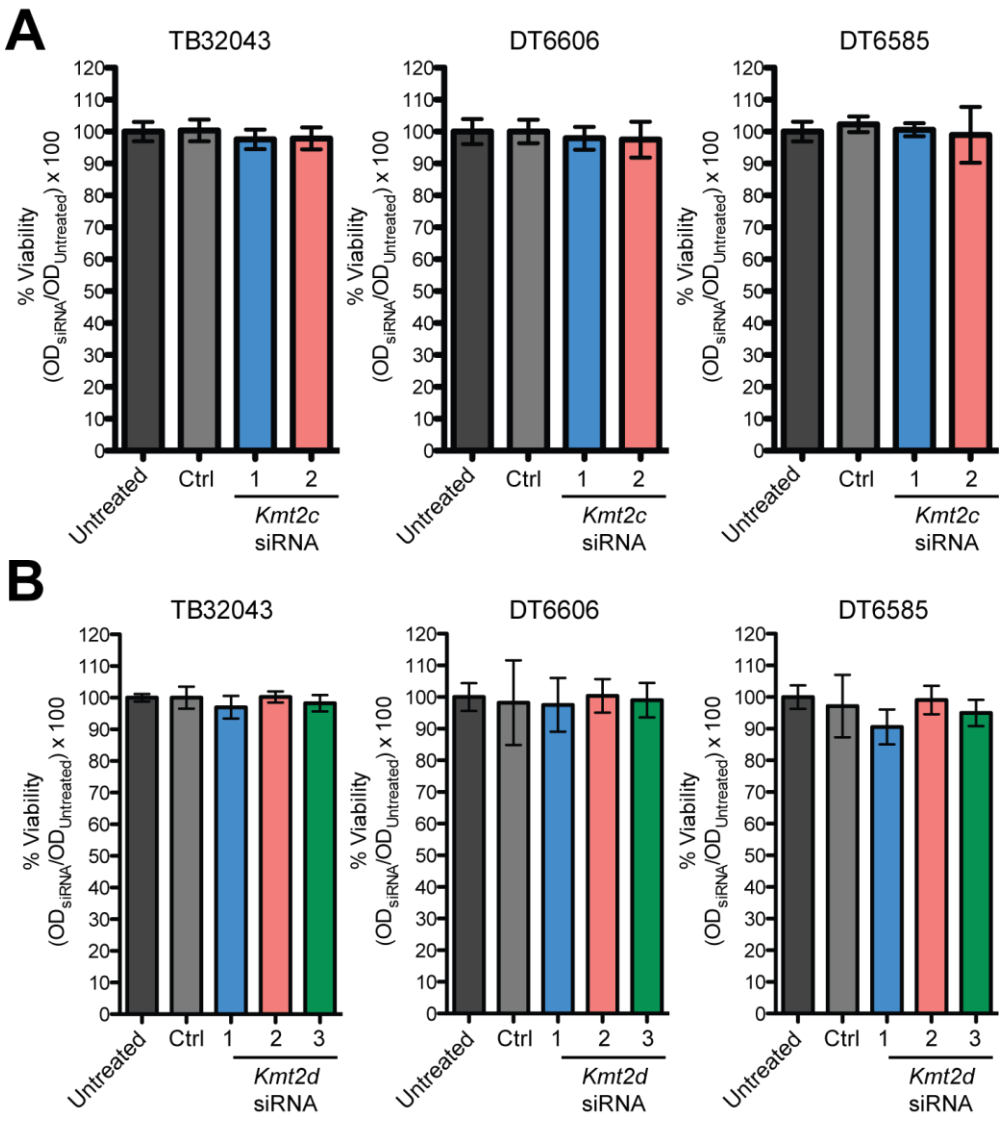

Figure 5.2 – Depleting Kmt2c or Kmt2d does not affect viability of the murine pancreatic cancer cell lines. A, Kmt2c cell viability was examined using WST-1 reagent for cells transfected for 48 hours before being re-plated and incubated for a further 72 hours. Data shown are mean OD values for triplicate wells from five experimental replicates ± SD for untransfected (black), control siRNA (grey), *Kmt2c* siRNA1 (blue), and *Kmt2c* siRNA2 (red). **B,** Kmt2d cell viability was examined using WST-1 reagent for cells transfected for 48 hours before being re-plated and incubated for a further 72 hours. Data shown are from mean OD values for triplicate wells from five experimental replicates ± SD for untransfected (black), control siRNA (grey), *Kmt2d* siRNA1 (blue), and *Kmt2d* siRNA2 (red).

5.2.2 Response of murine cell lines to chemotherapy

Following the absence of a cellular phenotype upon methyltransferase depletion, we next assessed whether changes to the murine cells may be subtler and require an additional stimulus. Since changes in human cell line proliferation were accompanied by cell-cycle abnormalities, it was next examined as to whether these H3K4 methyltransferases have a role in mediating cell response to chemotherapeutics that inhibit the cell cycle. To this end the nucleoside analogue antimetabolites gemcitabine and 5-FU were used, both of which are frequently used in PDAC treatment (see 1.1.5). To determine the response of the cell lines to each drug, a simple experimental design was followed, where an equal number of transfected cells were initially seeded and WST-1 assays performed at 72 hours post-drug treatment, for each sample replicate, at each concentration used.

5.2.2.1 Optimisation of concentration range for Gemcitabine and 5-FU

In order to determine the appropriate range of concentrations to use in the RNAi experiments, response curves were compiled over a range of concentrations for each drug in untransfected cells. For each cell line, gemcitabine treatment resulted in curves with steep Hill slopes (> -1) around $1x10^{-7.25}$ M (Figure 5.3A). Although the three cell lines produced similar growth IC_{50} (gIC_{50}) values (the concentration of drug required to elicit a 50% reduction in cell growth) for gemcitabine, they each differed in their sensitivity to the higher drug concentrations. At these higher gemcitabine concentrations, TB32043 appeared to be the most resistant, with only a reduction in viability to 30% possible. The DT6606 and DT6585 cell lines showed approximately equal sensitivity at these higher concentrations, where a reduction to around 20% viability was achieved.

Across the three cell lines 5-FU treatment produced distinct concentration-response curves, where gIC_{50} values ranged from $1x10^{-6.12}$ M (DT6606) to $1x10^{-3.63}$ M (TB32043) (Figure 5.3A). Mirroring the gemcitabine response data, the TB32043 cell line appeared to be the least sensitive and DT6606 the most sensitive of the three cell lines. Interestingly, TB32043 expresses higher basal levels of *Kmt2d* mRNA than the other two cell lines (Figure 5.3B).

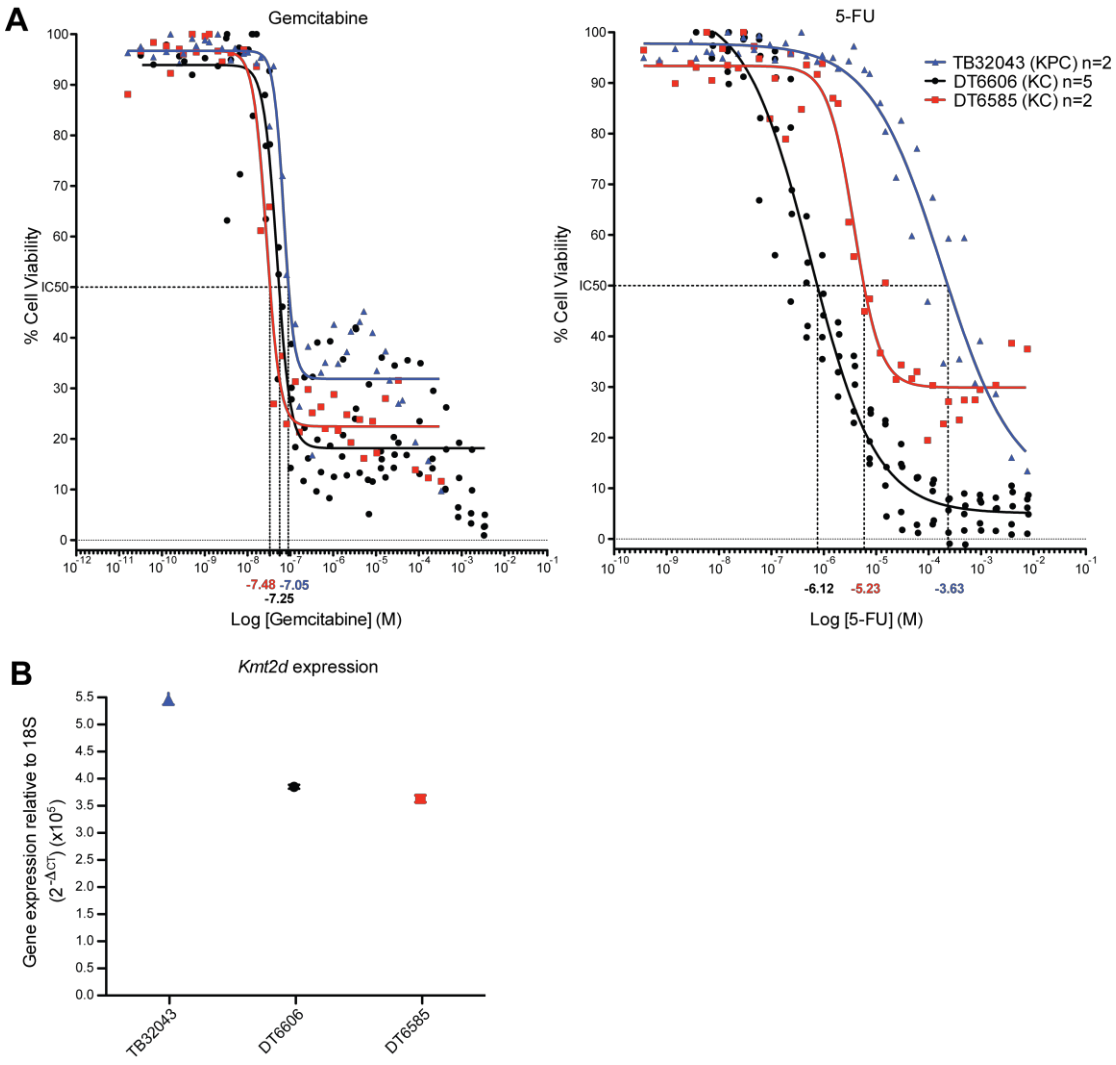

Figure 5.3 – Sensitivity of murine cell lines from the KC and KPC models to Gemcitabine and 5-FU. A, Sensitivity of the three untreated murine cell lines (TB32043 (blue triangles), DT6606 (black circles) and DT6585 (red squares)) to gemcitabine and 5-FU was examined using WST-1 to determine cell viability after 72 hours of drug treatment. Data shown are mean OD values from technical triplicate wells normalised to maximal OD for each biological replicate (n = 2 for TB32043 and DT66585; n = 5 for DT6606). **B,** Expression of *Kmt2d* mRNA relative to the expression of 18S was detected by qRT-PCR for the three cell lines. Data shown are mean values from technical triplicates.

5.2.2.2 Changes in sensitivity following Kmt2c or Kmt2d depletion

Kmt2c depletion had no impact across the three cell lines on their sensitivity to gemcitabine or 5-FU (Figure 5.4). In contrast, striking changes in the concentration-response curves were observed following Kmt2d depletion for both drugs (Figure 5.5). *Kmt2d* siRNA1 and *Kmt2d* siRNA2 notably altered the concentration-response curves while the 'least efficient' third siRNA (*Kmt2d* siRNA3) and the control siRNA closely resembled the untransfected cells (Figure 5.5). *Kmt2d* silencing led to an increase in gemcitabine sensitivity at the higher concentrations, where as little as 10% viability could be achieved, more than two fold less than that of control cells (Figure 5.5). For the 5-FU concentration-response curves, *Kmt2d* silencing elicited a horizontal shift left with approximately 10-fold less 5-FU required to elicit the same reduction in viability (Figure 5.5).

Previously decreased expression of the multidrug resistance gene *ABCB1* was observed in the human PDAC cell lines following depletion of KMT2D (Figure 4.7), but not KMT2C. To examine whether Abcb1 might be implicated in the drug sensitivity observed in these murine cell lines, its expression upon Kmt2d depletion was examined in TB32043 cells. No decrease in Abcb1 expression was observed for this cell line (Figure 5.6) and therefore while a strong link exists between Kmt2d expression and drug response, the precise mechanism of action is still open to speculation.

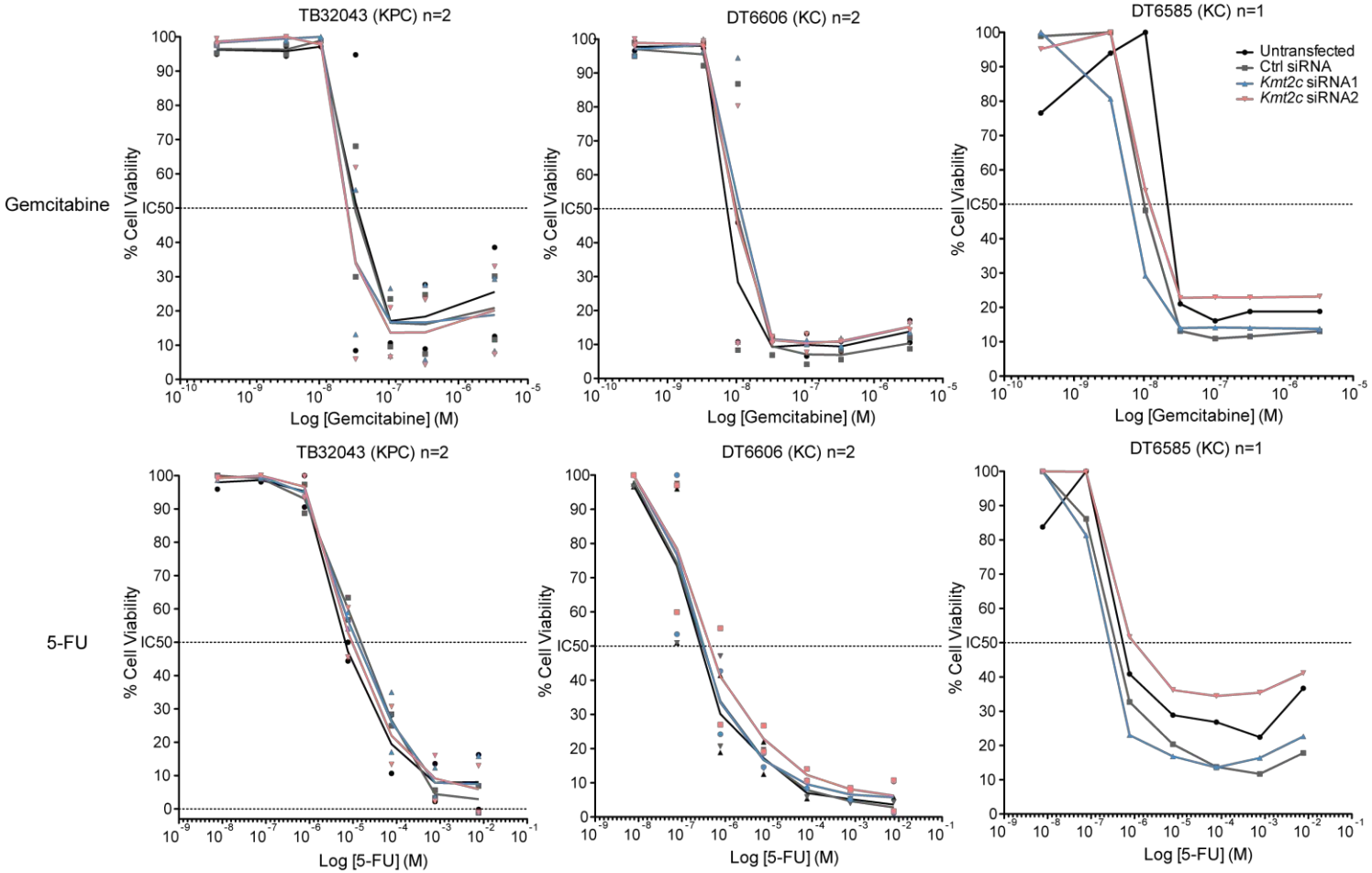

Figure 5.4 – Depletion of Kmt2c does not affect murine cell line sensitivity to gemcitabine or 5-FU. Reduced *Kmt2c* expression does not alter the gemcitabine or 5-FU concentration-response curves for any of the three cell lines examined. Cell viability was examined using WST-1 after 72 hours of treatment with either gemcitabine, or 5-FU. Data shown are mean OD values from technical triplicate wells normalised to maximal OD for each biological replicate (n = 2 for TB32043 and DT6606; n = 1 for DT6585) for untransfected (black), control siRNA (grey), *Kmt2c* siRNA1 (blue), and *Kmt2c* siRNA2 (red).

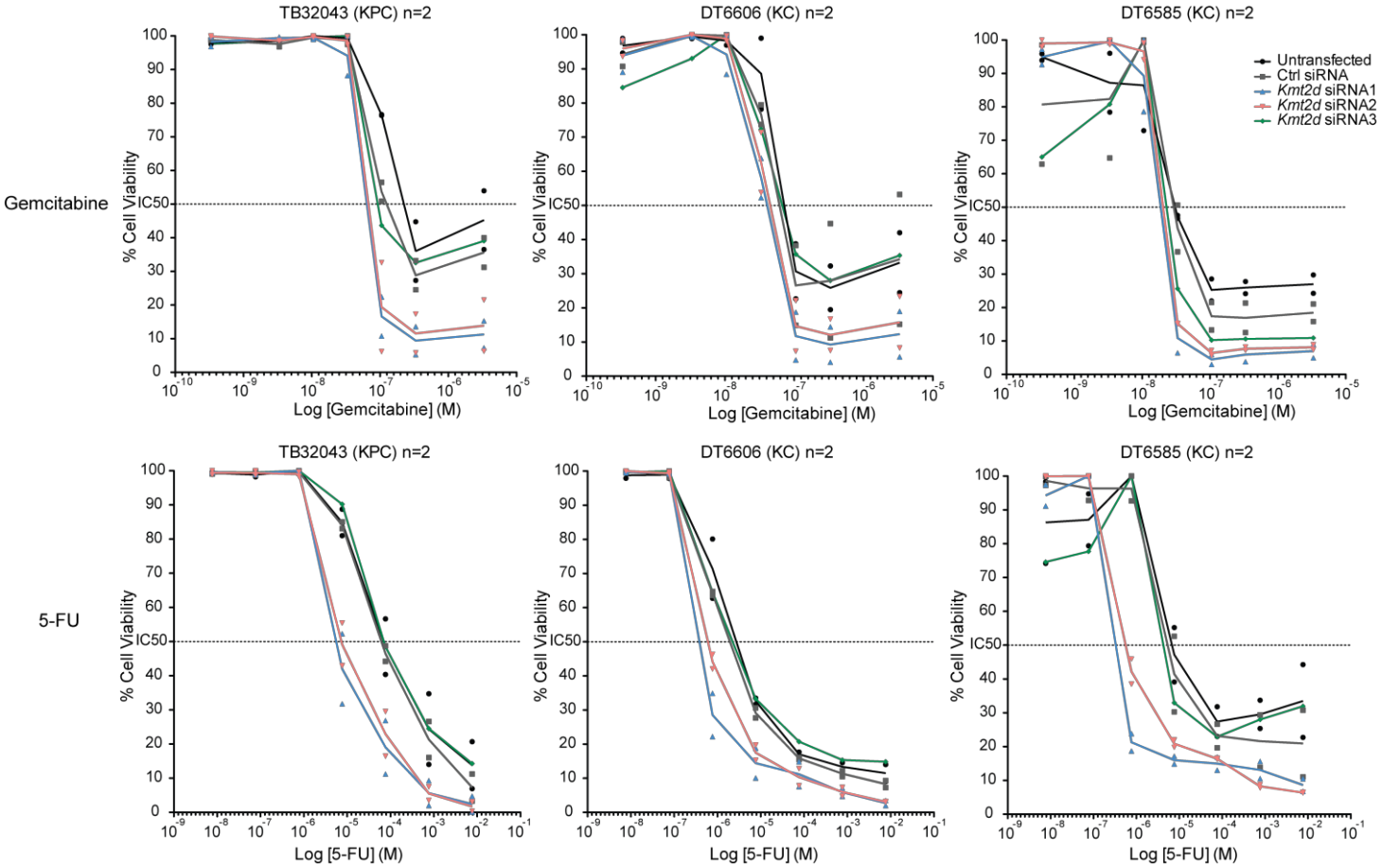

Figure 5.5 – Depleting Kmt2d affects murine cell line sensitivity to Gemcitabine and 5-FU. Depletion of Kmt2d increases murine cell line sensitivity to gemcitabine and 5-FU. Reduced Kmt2d expression increased the effects of gemcitabine at higher concentrations, and resulted in a leftward shift in the 5-FU concentration-response curves for each of the three cell lines. Cell viability was examined using WST-1 after 72 hours of treatment with either gemcitabine, or 5-FU. Data shown are mean OD values from technical triplicate wells normalised to maximal OD for each of two biological replicates for untransfected (black), control siRNA (grey), *Kmt2d* siRNA1 (blue), and *Kmt2d* siRNA2 (red).

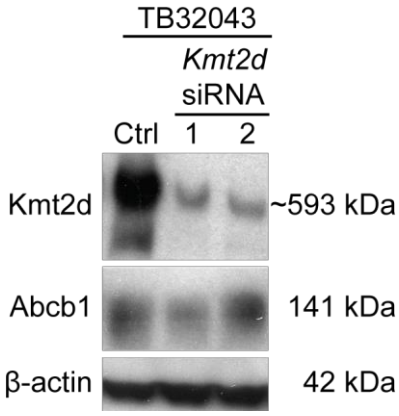

Figure 5.6 – Expression for Abcb1 remains unaffected by Kmt2d depletion. Western blot analysis shows decreased expression of Kmt2d, but not Abcb1, in TB32043 cells following treatment with two *Kmt2d* siRNAs compared to control siRNA.

5.2.3 Generation of *Kmt2d* knockout KC and KPC mice

In light of evidence implicating *Kmt2c* aberration in the formation of uroepithelial (132,133) and PDAC (74,75) tumours, an *in vivo* PDAC model with pancreas specific deletion of *Kmt2d* was developed. To this end, a breeding scheme was designed using *LSL-Kras*$^{G12D/+}$; *LSL-Trp53*$^{R172H/+}$ (KP), homozygous *Pdx-1-Cre*, and *Kmt2d*$^{Flox/Flox}$ mice to generate offspring with WT, heterozygous, or homozygous deletion of *Kmt2d* in the pancreata of KC and KPC mice (Figure 5.7). KP and homozygous *Pdx-1-Cre* mice strains were both obtained from Professor Dave Tuveson (Cold Spring Harbor Laboratory, USA), while *Kmt2d*$^{Flox/Flox}$ mice were obtained from Professor Francis Stewart (Bitechnology Center, Technische Universitat Dresden, Germany). The *Kmt2d* flanked LoxP construct has been previously described (152), where LoxP sites flank exons 3-5 of *Kmt2d* so that upon *Cre* recombination the resulting frame shift results in a premature TGA stop codon.

Kmt2d$^{Flox/Flox}$ mice were first bred with KP mice to generate KP; *Kmt2d*$^{Flox/+}$ offspring at a Mendelian ratio of 1/4. In parallel, *Kmt2d*$^{Flox/Flox}$ mice were also bred with homozygous *Pdx-1-Cre* mice to generate hemizygous *Pdx-1-Cre/-* (C); *Kmt2d*$^{Flox/+}$ offspring. Finally to generate mice with the desired genotypes, KP; *Kmt2d*$^{Flox/+}$ and C; *Kmt2d*$^{Flox/+}$ mice were crossed to generate KC and KPC mice that are heterozygote, homozygote, or WT for *Kmt2d*^{Flox} (Mendelian ratios of 1/16, 1/32 and 1/32, respectively) (Figure 5.7). To ensure the correct pairing of mice, and identify offspring with the desired genotype, PCRs were performed on DNA extracted from mouse ear snips for *Cre*, *Kras*, *Trp53, Kmt2d*, and as required an endogenous positive control gene (Figure 5.8).

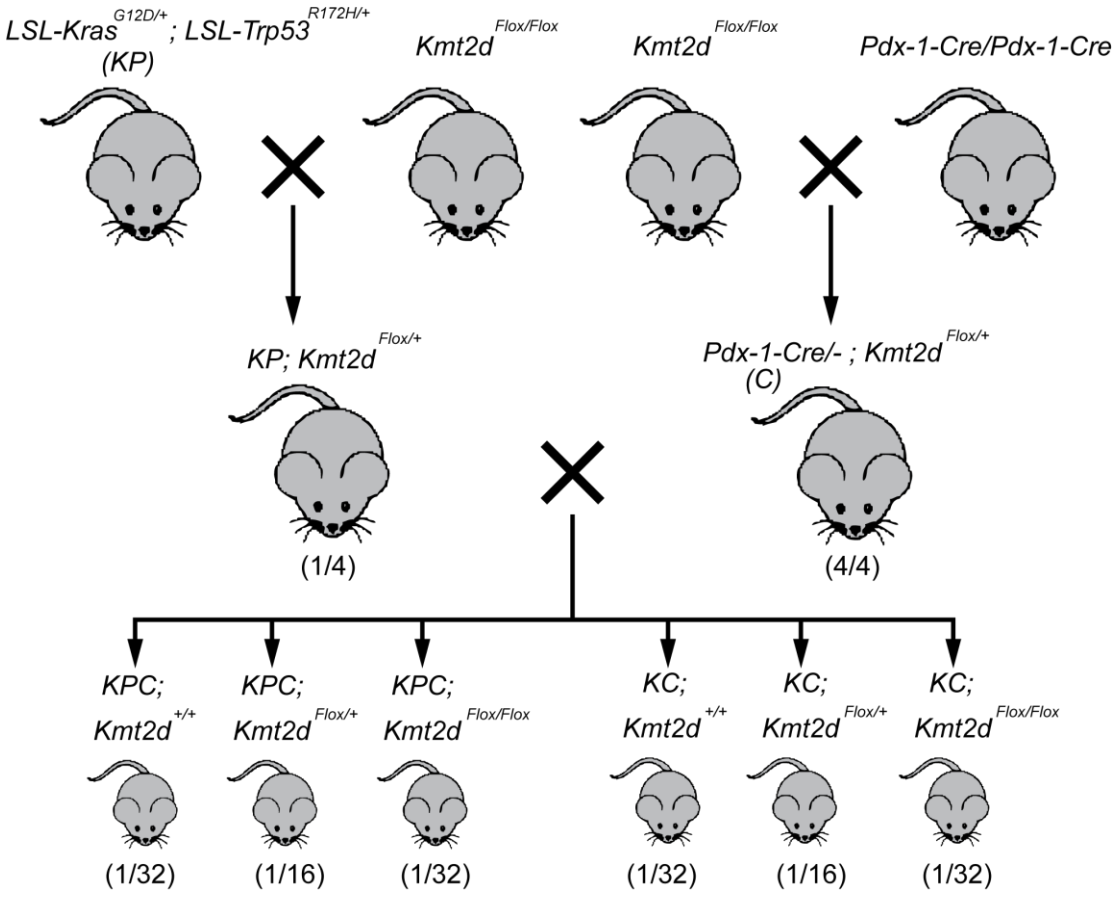

Figure 5.7 – Breeding scheme for generating KPC mice with loss of *Kmt2d.* To generate both KC and KPC mice with different *Kmt2d*^{Flox} zygosities, three breeding pairs were set up over two generations. First *Kmt2d*$^{Flox/Flox}$ mice were separately bred with *LSL-Kras*$^{G12D/+}$*; LSL-Trp53*$^{R172H/+}$ (KP) mice, or *Pdx-1-Cre/Pdx-1-Cre* mice. Resulting offspring were genotyped to set up breeding pairs of *KP; Kmt2d*$^{Flox/+}$ and *C; Kmt2d*$^{Flox/+}$ mice for generating mice with the range of genotypes required. Given in brackets are the Mendelian ratios for the expected numbers of offspring for each genotype.

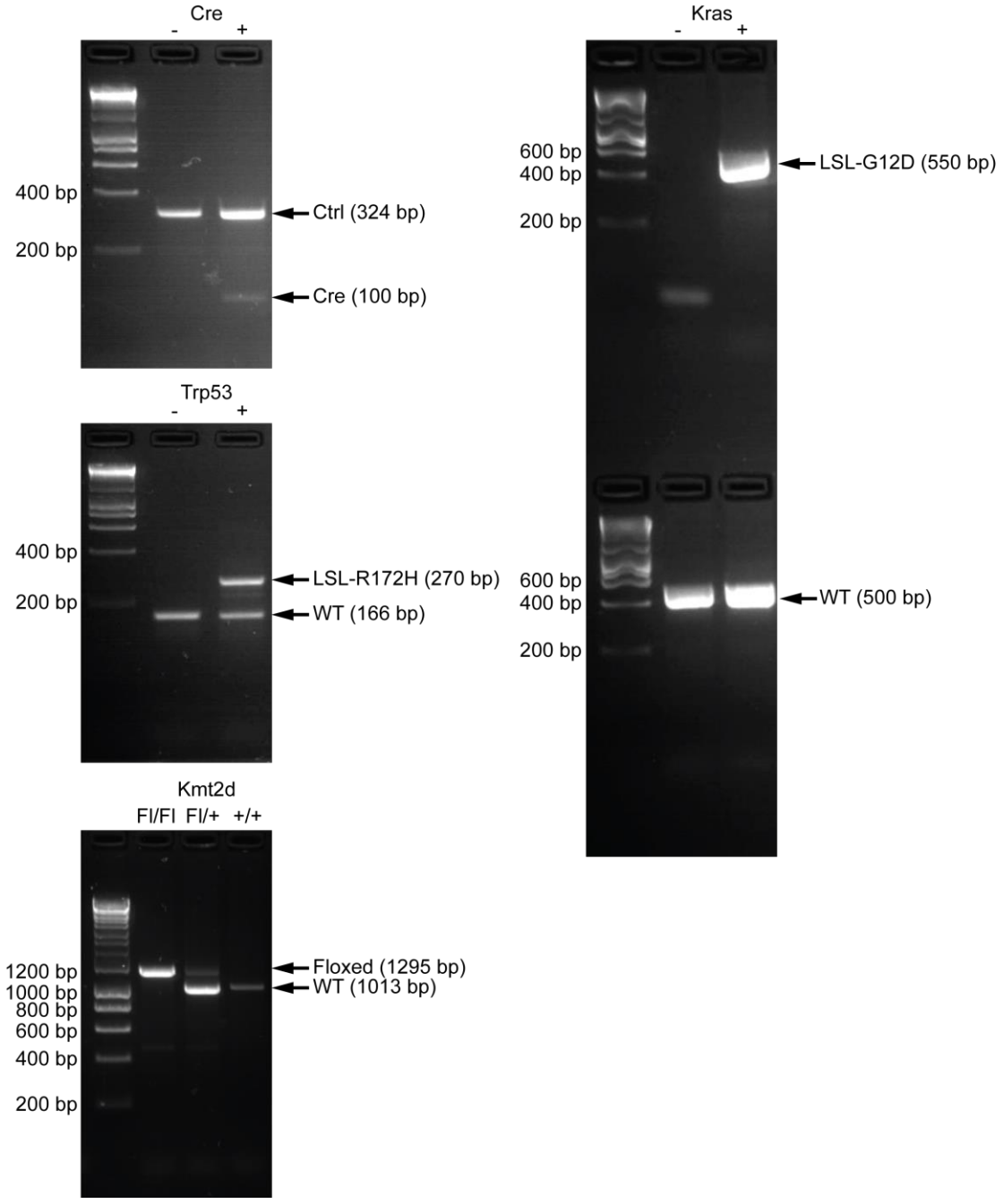

Figure 5.8 – PCR assays used to determine offspring genotype. Agarose gel electrophoresis for PCR products from DNA extracted from mouse ear snips to determine the presence of transgenes, mutant alleles, and WT alleles for *Cre*, *Kras*, *Trp53* and *Kmt2d*. Given in brackets are the sizes of the products generated by the PCRs.

5.2.3.1 Phenotype of Kmt2d loss-of-function in KPC and KC mice

The PCR used for genotyping *Kmt2d* can also produce a 227 bp product for the remaining single LoxP site upon recombination of the flanked LoxP cassette ($Kmt2d^{\Delta}$). In this way, cell lines generated from tumours arising in the KPC; $Kmt2d^{Flox/+}$ and KPC; $Kmt2d^{Flox/Flox}$ mice could be genotyped to confirm that genetic recombination of the LoxP site had occurred in transformed epithelial cells (Figure 5.9). When tested, cell lines generated from KPC; $Kmt2d^{Flox/Flox}$ (E1.1 and E15.1P) and KPC; $Kmt2d^{Flox/+}$(E2.1P) tumours displayed the expected products for homozygote or heterozygote floxed allele recombination respectively. Another cell line was also generated from a KPC; $Kmt2d^{Flox/Flox}$ tumour (E8.5L), however poor-quality DNA resulted in a smeared PCR product. Nevertheless the mouse-of-origin genotype had already been successfully assessed, and thus this cell line was also included for further assessment of Kmt2d protein expression.

To confirm *Kmt2d* knockout in tumour cells at the protein level, western blots were performed using lysates generated from the compound mutant cell lines (E1.1, E15.1P, E2.1P and E8.5L) and control *Kmt2d*-WT KC and KPC cell lines (Figure 5.10). Kmt2d protein knockout was clearly confirmed in KPC; $Kmt2d^{\Delta/\Delta}$ cells (E1.1, E15.1P, E8.5L). KPC; $Kmt2d^{\Delta/+}$ cells (E2.1P) however expressed Kmt2d at similar level to those from control *Kmt2d*-WT KC and KPC mice, indicating that loss of one allele does not result in any detectable reduction of Kmt2d.

The survival for each mouse genotype was assessed to determine whether *Kmt2d* loss alters the disease in the KC and KPC models. Survival for each animal was based on the date at which animals had to be culled for welfare issues associated with PDAC tumour burden. No significant difference was found in the survival of KPC mice harbouring heterozygous, or homozygous deletion of *Kmt2d* ($p = 0.2099$ by log-rank, Mantel-Cox) (Figure 5.11), with no overt differences in the tumour histology noted (Eleni Maniati, data not shown). In addition, there was no earlier onset of disease for either model, where KC mice did not develop disease irrespective of *Kmt2d* status (up to 8 months) (Figure 5.11). As low numbers of WT KPC mice were obtained from our breeding scheme, comparative survival data for 45 untreated KPC mice from other studies ongoing in our laboratory (Dr Juliana Candido) are also included in Figure 5.11.

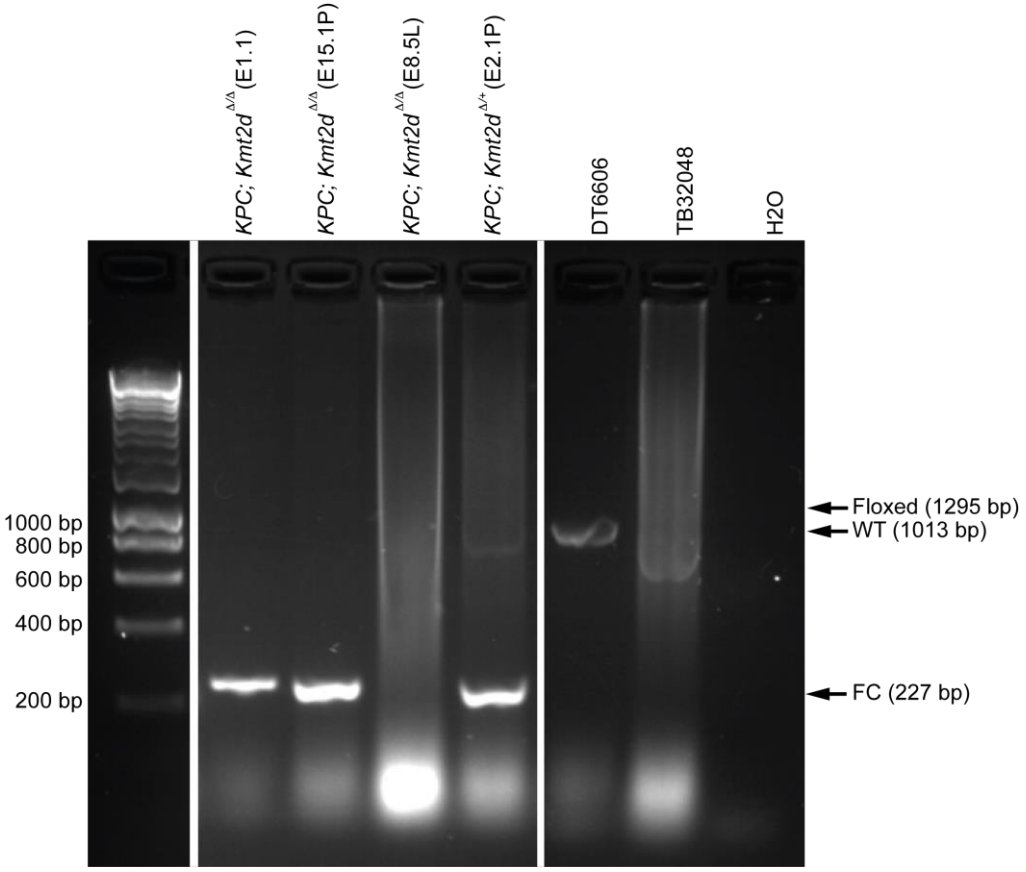

Figure 5.9 – Genotype of Kmt2d cell lines. Agarose gel electrophoresis for the products of PCR primer pairs used to determine the presence of LoxP sites in the *Kmt2d* gene in DNA extracted from cell lines generated from the mice. Given in brackets are the sizes of products generated by the PCRs.

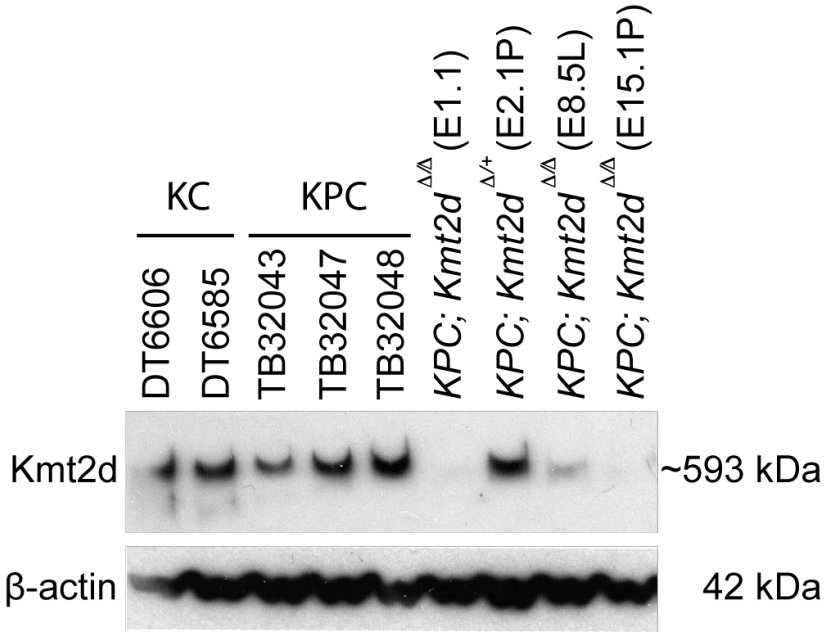

Figure 5.10 – Expression of Kmt2d across a range of murine cell lines generated from the KC, KPC, and KPC; Kmt2d models. Western blot analysis detected Kmt2d protein in each of the cell lines derived from the KC (DT6606 and DT6585), KPC (TB32043, TB 32047 and TB32048), and *KPC; Kmt2d* heterozygous (E2.1P) mouse models. Either weak or no Kmt2d was detected in the cell lines derived from the tumours of *KPC; Kmt2d* homozygous (E1.1, E8.5L and E15.1P) mice. β-Actin was used as a loading control.

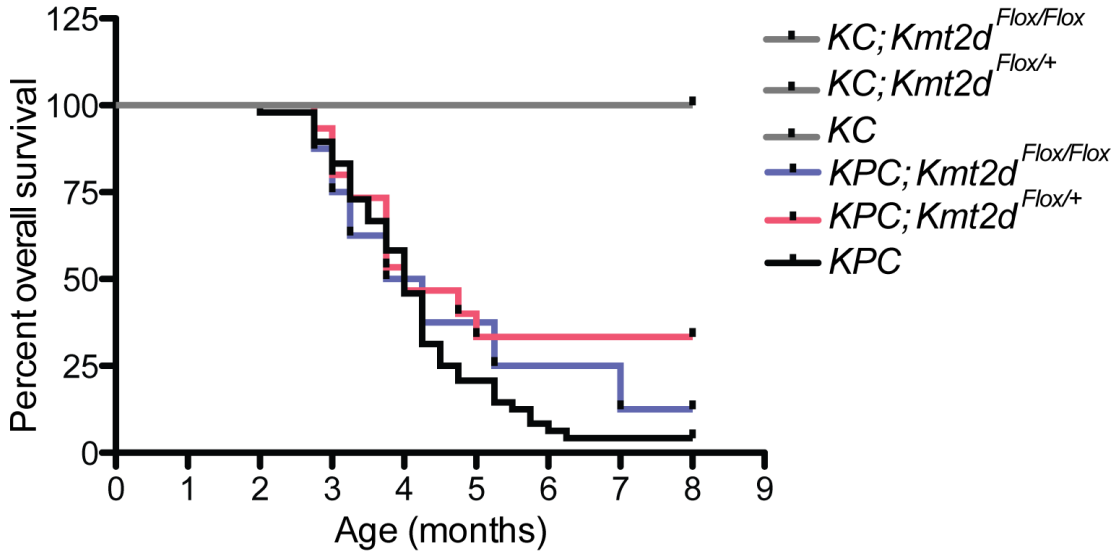

Figure 5.11 – Mice heterozygous and homozygous for loss of *Kmt2d* do not exhibit significant difference in overall survival compared to WT KC and KPC counterparts. KM survival analysis showed that KC mice do not succumb to PDAC regardless of *Kmt2d* status (up to 8 months) (Grey lines, *KC; Kmt2d*$^{Flox/Flox}$ n = 5, *KC; Kmt2d*$^{Flox/+}$ n = 7, *KC* n = 3). For KPC mice KM analysis showed that *Kmt2d* status does not significantly alter overall survival (P = 0.2099 by log-rank, Mantel-Cox) (Blue line: *KPC; Kmt2d*$^{Flox/Flox}$ n = 8, Red line: *KPC; Kmt2d*$^{Flox/+}$ n = 15, Black line: *KPC* n = 48 [45 from laboratory's historic data, including 3 littermates from the present study]).

5.3 Discussion

Considering the vast genetic heterogeneity now known to exist in PDAC (40-45), we sought to further assess the functional impact of *Kmt2d* depletion in the genetically defined KPC and KC murine model systems. Broadly speaking, the majority of functional studies for Kmt2d in murine systems have to date predominantly focused on its effects in the early stages of cell differentiation and embryogenesis (113,120,135,249). Interestingly in these Kmt2d models, Wan *et al.* demonstrated that its depletion reduced embryonic stem cell proliferation *in vitro* (135), while Ang *et al.* observed *in vivo* myocardium hypoplasia when depletion was restricted to cardiomyocytes (249). In contrast to these two studies however, Lee *et al.* found no change in proliferation when Kmt2d was depleted in *Kmt2c* null pre-adipocytes. In line with an effect on proliferation, RNA-seq studies in cardiac tissue and myoblasts have shown a decrease in expression of cell-cycle genes upon *Kmt2d* depletion (113,249) (including Lee *et al.* for myocytes but not for adipocytes (120)), supported by *in vitro* observations for depletion increasing the fraction of G_1 cells (135,249). In our pancreatic KC and KPC cell lines Kmt2d knockdown did not have any detectable effect on cell viability, growth or cell-cycle progression (Figure 5.2 and Dr Eleni Maniati data not shown). Taken together these data indicate *Kmt2d* likely has cell-type and context dependent roles, something also proposed by Lee *et al*.

Despite *Kmt2d* siRNAs not altering KC and KPC cell line proliferation, Kmt2d depletion did render the cells more sensitive to the exposure of gemcitabine or 5-FU (Figure 5.5). Interestingly, increased sensitivity was specific to cells devoid of Kmt2d, where this difference in effects between the methyltransferases is encouragingly in line with their impact on patient outcome (see 3.2.1), and proliferation of human cells *in vitro* (see 3.2.4). This link between Kmt2d expression and chemotherapy sensitivity demonstrated by RNAi might also be inherent to cells based on their basal levels of expression for this gene. In particular, the KPC-derived TB32043 cell line retains both the highest native *Kmt2d* mRNA expression and was found to be the least sensitive to drug treatment among those tested (Figure 5.3). This higher *Kmt2d* mRNA expression in TB32043 (*Trp53* mutant) compared to the KC-derived DT6606 and DT6585 cell lines (*Trp53* WT), is reminiscent of our human cell line data where it may also support a link between *TP53* mutation and an increase in KMT2D expression (Figure 3.3). Nevertheless further experiments into these observations are required since many other

intrinsic factors (*e.g.* acquired mutations or differences in proliferation rates) are likely to be at play.

Since loss of Kmt2d and not Kmt2c increased sensitivity to 5-FU and gemcitabine it will be important to understand the downstream mediators of these methyltransferases. As mentioned in Chapter 4, *ABCB1* (a well-established multidrug resistance gene (278,279,298)) was decreased in human cells upon KMT2D but not KMT2C depletion (Figure 4.7). In the KPC cell line tested, Kmt2d depletion did not result in a decrease of this resistance-associated protein (Figure 5.6), suggesting that Abcb1 does not mediate the increased drug sensitivity.

Gemcitabine and 5-FU are nucleoside analogue antimetabolites that inhibit cell-cycle progression and induce DNA damage. Within the literature Kmt2d loss has been shown to promote both of these features where, alongside the cell-cycle data (discussed above and in the preceding chapters), Kantidakis *et al.* have shown *KMT2D* LOF as increasing genome instability and DNA damage (152). Therefore, our current working hypothesis is that the increased sensitivity to these drugs upon Kmt2d depletion is likely through the enhancement of one, if not both, of these actions. Intriguingly, Kmt2d depletion had different effects on the 5-FU and gemcitabine concentration-response curves (Figure 5.5), which perhaps reflects differences in their pharmacological actions. It has been well reported that 5-FU treatment causes G_1-S arrest across a wide range of cell types (299-305). Considering the proposed role for Kmt2d in G_1 progression, it is perhaps not surprising that murine cells show increased sensitivity at each concentration of 5-FU following Kmt2d depletion, where ten-fold less 5-FU was required to elicit the same response, thus resulting in a leftwards shift in the whole response curve (Figure 5.5). In contrast to 5-FU, Kmt2d depletion enhanced response to gemcitabine only at higher concentrations (Figure 5.5). One plausible explanation is that Kmt2d depletion may work in a more general way to promote gemcitabine cytotoxicity, in comparison to being a direct modulator of the mechanism of action for 5-FU. Gemcitabine predominantly induces S-phase arrest (306-308), rather than a G_1 arrest (309), and therefore the G_1-phase arrest caused by Kmt2d loss may combine synergistically with gemcitabine to cause cytotoxicity.

Within the literature there is very little evidence for the *in vivo* effects of depleting these methyltransferases in solid malignancies. *Kmt2c* has been separately reported to cooperate with $Trp53^{+/-}$ (133) and $Kras^{G12D}$ (74,75) to promote the formation of uroepithelial (132,133) and PDAC tumours (74,75), respectively. To date there are no

in vivo reports into *Kmt2d* mutations and the formation of solid malignancies. To investigate whether *Kmt2d* deletion has a tumorigenic role in PDAC, we combined pancreas specific loss of *Kmt2d* with the well established KC (68) and KPC (72) models (Figure 5.7).

Germline $Kmt2d^{-/-}$ or $Kmt2c^{-/-}$ mice show several abnormal phenotypes (120,162). Lee *et al.* showed whole body homozygous loss of *Kmt2d* to result in embryonic lethality around E9.5 (120) whereas for *Kmt2c* this was lethal at a later perinatal stage (120), thus alluding to their different roles in development. This essential role for *Kmt2d* is likely due to its role in heart development and function where, using a conditional cardiac tissue knockout model, Wan *et al.* have shown that although heterozygous mice are viable, homozygous loss causes lethality between E9 and E14 (249). Although mice with unspecific global heterozygous *Kmt2d* or *Kmt2c* loss are also viable (120), Bjornsson *et al.* have shown functionally heterozygous $Kmt2d^{+/\beta Geo}$ mice to exhibit facial features and hippocampal memory defects (162) in line with human individuals affected by Kabuki syndrome (128,161).

The KC model has been extensively used to examine whether genetic alterations cooperate with mutant $Kras^{G12D}$ in PanIN progression and PDAC formation (reviewed in (69-71)). KC mice, carrying only the $Kras^{G12D}$ mutation, show complete penetrance for all PanIN stages with slow and very infrequent PDAC development (only two mice from a cohort of 29 develop tumours at 6.25 and 8.25 months) (68). Studies introducing other mutations into KC mice, showed that other events are required for full progression into invasive and metastatic PDAC, with the introduction of $Trp53^{R172H}$ causing 100% mortality within 12 months for all 28 mice generated (72). Unlike $Trp53^{R172H}$, *Kmt2d* loss does not appear to substantially cooperate with $Kras^{G12D}$ to accelerate PDAC formation and in accordance with the original KC studies (68), our KC mice regardless of *Kmt2d* status did not form any tumours by eight months (Figure 5.11). This observation might be supported by experiments combining KC mice with the *Sleeping Beauty* transposon system, where no *Kmt2d* insertions were identified in the PDAC tumours formed (74,75). Since *TP53* has been shown to enhance tumourigenesis in *Kmt2c* mutant mice (133), we also examined *Kmt2d* loss in KPC mice. We did not note any significant difference in survival or metastatic disease compared to WT KPC mice (Figure 5.11), with survival again similar to that previously reported by Hingorani *et al.* (72).

The breeding scheme for generating the compound mutant mice resulted in a 129/SvJae/C57Bl/6 mixed genetic background. A mixed strain background has been described as having the potential for introducing variability (310). In addition, the mixed background restricts the use of cell lines generated from the tumours of these mice in transplantable models due to potential histocompatibility mismatches.

Considering the similarities between the two methyltransferases, and the cooperation noted between *Kmt2c* aberration and $Kras^{G12D}$ mutation (74,75) and *Trp53* haploinsufficiency (133), one might have hypothesised similar disease promoting effects for *Kmt2d* loss. Furthermore, KMT2D loss has been shown to increase genomic instability in both mouse and human cells (152), and thus a contribution towards disease progression might have been expected in mice. To examine whether this instability also occurs in the tumours or PanIN lesions of our models it would be interesting to sequence tissue DNA to assess whether Kmt2d loss in this setting increases somatic mutation. We cannot rule out, and indeed it may seem reasonable to conclude, that the absence of pro-tumorigenic effect in the KPC model upon *Kmt2d* loss could be due to the $Kras^{G12D}$ and $Trp53^{R172H}$ mutations already driving a very aggressive disease. In this way, the comparatively weaker effects of *Kmt2d* depletion could have been diluted. This lack of a tumour promotion by Kmt2d loss, compared to that reported for Kmt2c loss in ureter epithelial (133) and PDAC tumours (74,75), perhaps also suggests differences between the roles of their mutation in disease progression.

In addition to examining survival, we also collected tissue samples from these mice for further *ex vivo* analysis and assessment of whether *Kmt2d* loss alters the grade of the PanIN lesions found, tumour architecture, or tumour microenvironment. From these tumours we also generated cell lines to confirm *Kmt2d* recombination and protein loss in the tumour (Figure 5.9 and Figure 5.10), where they may have potential for use in future *in vitro* and *in vivo* studies.

In summary, herein we have used RNAi to determine the *in vitro* effects of Kmt2c and Kmt2d depletion on three murine cell lines derived from the KC and KPC models. Alongside this work, we also developed compound GEMMs to study the *in vivo* role of Kmt2d loss in PDAC. Our data show that unlike in human cell lines, depletion of the methyltransferases alone does not alter the proliferation of our murine KC and KPC cell lines. Interestingly however, the sensitivity of these cell lines to 5-FU and gemcitabine was increased upon the depletion of Kmt2d, but not Kmt2c. We are in the

early stages of exploring whether these findings translate into the clinic, and therefore while our *in vivo* models did not highlight differences in disease progression, they represent important new resources for further investigations.

Chapter 6 Discussion

6.1 Overview of findings, future work and therapeutic implications

New sequencing technologies have greatly improved accessibility to the cancer genome and have highlighted that alongside frequent mutations in *KRAS*, *TP53* and *SMAD4*, a wide genetic diversity exists in PDAC tumours (40-45). Of particular note has been the detection of many mutations in epigenetic regulators that have become an established feature of this disease, and indeed most cancers (159). Members of the KMT2 family are among the most commonly mutated genes in cancer (170,171), with *KMT2C* and *KMT2D* mutations in up to 16.7% (45) and 6% (42) of PDAC cases, respectively. Despite this however, little is known about their roles and the consequences of their aberration. The present work offers an insight into the roles of KMT2C and KMT2D in PDAC and gives some indication of their potential as novel therapeutic opportunities for the treatment of this cancer.

The genetic alterations of *KMT2C* and *KMT2D* in PDAC are almost uniformly LOF mutations (41); thus our initial rationale reasoned that inter-patient fluctuations in *KMT2C* and *KMT2D* expression might also convey favourable outcome to PDAC patients. By comparing matched gene expression profiles and outcome in publically available PDAC datasets, a strong favourable signal linked with low expression of these methyltransferases was uncovered. This directed my subsequent studies into the effects of their depletion and led me to hypothesise that a combination of both their expression and mutation in patient tumours could be clinically used to predict prognosis. While not examined in this thesis, there is potential to test whether these patients with better outcome can also be captured by immunohistochemical analysis. Such an approach may have a much wider application, where it is technically simpler and cheaper than targeted DNA sequencing. Since this would also detect other mechanisms that disrupt these enzymes, this may also help to identify a greater number of patients who might benefit from stratification and targeted therapy, rather than just the minority that carry mutations (40-45). Going forward, a deeper appreciation of the functional differences between mutations is also required, where lesions can be expected to render the methyltransferase enzymatically inactive, unaffected, or even alter substrate specificity (94).

Our *in vitro* KMT2D silencing experiments led to a marked reduction in the proliferation of all eight human pancreatic cell lines tested. This is consistent with previous studies in a variety of human (55,138,151,202) and murine (135) cell types, and thus supports a fundamental and overlapping role for KMT2D across solid cancers.

In contrast to KMT2D, the impact of KMT2C depletion is not as well defined, both within the literature and the present study, where loss of this methyltransferase has been associated with no effect (136), an increase (133,199,201,203), or a decrease (Figure 3.7 and Figure 3.9) in cell proliferation. This suggests its roles may be cell type and context dependent, or indeed depend on the expression of other epigenetic regulators. This difference between KMT2C and KMT2D, is also perhaps supported by patient survival data, where in other cancers reduced *KMT2C* expression, and mutation (204), has been associated with poor prognosis in both breast (194) and gastric cancers (205); while in breast cancer, reduced *KMT2D* expression is associated with improved prognosis (194,202).

To investigate the requirement of KMT2D for proliferation, I determined the cell-cycle changes associated with KMT2D depletion in a series of human PDAC cell lines. My analysis identified that the underlying mechanism of reduced proliferation involves cells arresting in the G_1-phase, implying that KMT2D is permissive for progression into the S-phase. The only prior study in human cells with cell-cycle data was carried out by Issaeva *et al.* in HeLa cells following KMT2D depletion, where no changes to the G_0/G_1 population were noted (138). Nevertheless, they did not incorporate agents to induce a cell-cycle block in either the S or G_2/M-phases, which may thereby explain why a G_1 arrest was not resolved. My results are however consistent with studies using murine embryonic stem cells, and those generated from hearts at E13.5, where substantial increases in G_0/G_1 cells were observed upon Kmt2d loss (135,249).

Alongside these experiments, we sought to explore the broad downstream effects of KMT2C and KMT2D depletion in PDAC through RNA-seq analysis. In order to identify the more robust gene signatures, rigorous criteria whereby the focus was primarily on the changes consistent across the three cell lines and siRNAs used. Applying less stringent criteria may reveal additional potentially interesting genes and pathways, where expression data are stored and open to further interrogation under the accession number GSE75327. In line with a requirement of KMT2D, and perhaps KMT2C, in cell-cycle progression, GSEA identified a significant reduction of many genes associated with the cell-cycle and DNA-maintenance pathways in both KMT2C and KMT2D depleted cells. These results are in accordance with studies from RNA-seq studies in mice cardiac tissue and myoblasts upon *Kmt2d* depletion (113,120,249), but differ to microarray studies on stably depleted human cells, where neither noted changes to the cell-cycle pathway (138,248). The identification of this role in the cell-cycle is

supported further by our pathway analysis of patient expression-survival data (Table 4.4 and Table 4.5) and that of Sausen *et al.*, who commented that patients with *KMT2A*, *KMT2C*, and *KMT2D* mutations also exhibit expressional differences in genes associated with cell-cycle progression (41).

The impact of methyltransferase activity on cell-cycle progression has potentially significant clinical implications. To further refine the cell line RNA-seq data, comparisons were made against publically available gene expression datasets obtained from primary PDAC samples. For the genes identified with a concurrent direction of fold change, we then tested whether their expression also correlated with clinical outcome. This approach highlighted *NCAPD3* as potentially having a novel role in PDAC, establishing its reduced expression as a strong predictor of outcome in all four datasets tested (ICGC, TCGA, Stratford and BCI_Zhang), and demonstrates the power of this experimental approach in identifying potentially drugable pathways. NCAPD3 acts as the critical subunit of the condensin II complex required for chromosome condensation and reorganisation (292). Considering these roles of the complex, it is perhaps unsurprising that *NCAPD3* RNAi results in M-phase arrest (293), rather than the G_1 arrest observed upon KMT2D depletion, where we currently hypothesise that *NCAPD3* is unlikely to mediate the effects of KMT2D in G_1. Indeed, we are far from understanding the role that *NCAPD3* holds in PDAC, with many fundamental questions that remain to be addressed. Whether the gene is a direct or indirect target of KMT2D, what exactly its role may be, and whether it is accountable for the decreased chromosome reorganisation and maintenance observed in our RNA-seq data (Figure 4.8) remain to be uncovered.

The observed effects of KMT2D silencing indicate that it directly or indirectly regulates cell cycle proteins that promote G_1 progression, and thus its depletion results in G_1 arrest. The cyclin-dependent kinase inhibitors $p16^{ink4a}$ and $p27^{Kip1}$ have previously been shown to increase upon Kmt2d depletion (135). Since both block G_1 to S-phase transition (311), a negative regulator of their expression may prove to be a good candidate gene. In a similar way to determine target genes, ChIP based techniques have previously been employed to determine genomic regions directly targeted by KMT2D (150,151,248). As noted by Guo *et al*., the lack of quality antibodies has stifled this approach to identify target genes for KMT2D (248) and KMT2C. To overcome this for KMT2D, Guo *et al.* developed a human colorectal carcinoma cell line with a Flag-tag motif knocked into the endogenous gene (248) so that Flag-tag-ChIP-seq could be

performed. Upon comparison of this ChIP-seq data to known enhancer regions (312), the enhancer elements targeted by KMT2D could be uncovered (150,151). While we appreciate cell specific differences will undoubtedly exist, it is likely that direct targets of this methyltransferase in PDAC could be attained using enhancer mark (*i.e.* H3K4me1, H3K27Ac and H3K27me3) ChIP-seq following RNAi, and comparing this, alongside our RNA-seq data, to those data from Guo *et al*. and Hu *et al.* (150,151). This approach however would be challenging since the current field linking enhancers to their regulated genes remains relatively underdeveloped (313).

While ChIP techniques examine direct chromatin modifications and protein binding, other next-generation sequencing coupled technologies have been developed to interrogate chromatin biology (314) and therefore could be used to examine how KMT2D and KMT2C differentially regulate chromatin structure, and thus genomic expression. Chromosome confirmation capture technologies can be used to investigate the spatial organisation and interactions of chromatin, and thereby may help to better elucidate links between enhancers and their regulated genes (315). DNase-seq, formaldehyde-assisted isolation of regulatory elements (FAIRE-seq), and 'Assay for Transposase Accessible Chromatin' (ATAC-seq) techniques each use a different experimental approach to identify open regions of 'active' chromatin, whereas the micrococci nuclease digestion (MNase-seq) technique instead highlights the regions of high nucleosome occupancy (316) often found at regulatory sequences (317). While these constantly developing approaches each have their own advantages and uses, they also come with biases and limitations that must be carefully considered (314). Predominantly, although promising, there remains significant hurdles in understanding and interpreting these new data types (314,315).

Since differences in clinical benefit associated with these methyltransferases likely involve a number of complex interactions that can only be modelled *in vivo*, we generated compound KC and KPC mouse models with pancreas specific loss of *Kmt2d*. Unlike the improved survival for patients with LOF mutations (41) and low expression (Figure 3.1), our preliminary investigations showed that *Kmt2d* loss did not appear to alter the disease phenotype in mice (Figure 5.11), perhaps reflecting the lesser impact observed in murine cell lines following Kmt2d depletion. Despite murine cell proliferation not being affected, our studies did show that depletion of Kmt2d led to an increase in 5-FU and gemcitabine sensitivity. Whether this phenomenon would occur *in vivo* remains untested. However, 5-FU and gemcitabine are both commonly used in

PDAC treatment (82), where often patients subsequently acquire resistance (318), so this observation is potentially a clinically important finding. Since homozygous LOF mutations and total loss of protein are not found in patients (119,128,160), it would be of interest to evaluate whether heterozygous *Kmt2d*$^{Flox/+}$ PDAC mice also show improved response to these treatments. Although this gives some insight into the potential benefit of KMT2D depletion as a therapeutic avenue, key differences will exist between total loss of protein and the more physiologically relevant scenarios of mutation, reduced expression, and pharmacological intervention. While mice harbouring LOF mutations could be generated using new gene editing techniques, it would be more clinically relevant to study pharmacological inhibition. While many inhibitors have been designed to lysine methyltransferases (319,320), those that target the catalytic site of KMT2C and KMT2D have yet to be developed (321). Since data suggest differing roles for Kmt2c and Kmt2d, where Kmt2d loss is sensitising, further investigations are required into whether a pan-, or homologue specific inhibitor would be required. This is especially true regarding the inhibition of KMT2C, where its mutation and reduced expression has been associated with poor prognosis in other cancers (194,204,205), thus causing concern that targeting this homologue may in some instances be detrimental.

Considering the clinical link between these methyltransferases and genes in the cell-cycle pathway, patients with low *KMT2D* expression may also show enhanced sensitivity to chemotherapy. In this way, the clinically annotated gene expression and mutation datasets could be further investigated to evaluate whether patient KMT2C and KMT2D expression and/or mutation correlate with treatment-based outcome. Notably, the tumour samples used in these studies were obtained at surgery, leading to a potential bias in favour of patients suitable for surgery (2), and therefore while challenging, value may also be found in examining these histone methyltransferases using post-mortem samples. If it is possible to establish a link with treatment response in PDAC, we postulate that the expression and mutation of these methyltransferases could both be used to inform future patient treatment regimes.

Alongside this stratification, KMT2D pharmacological inhibition may be effective in combination therapy, especially for those with high expression and no mutation. Epigenetic therapy promises great benefits over existing therapies currently used in the clinic, where inhibitors can help redress the dysregulated transcriptional balance in cancer cells by altering the expression of several proteins and pathways simultaneously

(322). While this may prevent or modulate resistance (323), the multifaceted mechanism of action may also mean that less drug is required for response, thereby reducing off target side effects (324). Despite these advantages, there are still limitations to consider, including off target toxicity and, since the process of epigenetic regulation takes time through an accumulation of marks, patients may require far longer-term treatment before any benefits may be realised. Also, and particularly within the setting of PDAC, unless these interventions also impact on the tumour stroma they, like current therapies, are likely to be felled by the 'stromal fortress'.

By referring back to our human cell line transcriptome data, we formed two hypotheses for how Kmt2d depletion may increase sensitivity. Since ABCB1 functions as a drug-efflux pump with a broad specificity (298) it is expected that its expression is positively associated with multidrug resistance across many cancer types (278,279,325). While further validation is required, this is unlikely to be a major mechanism of increased sensitivity here, since Abcb1 expression remained stable following Kmt2d depletion in the mouse cell line tested. In addition, ABCB1 overexpression has before been reported to correlate with increased gemcitabine sensitivity in PDAC (326), therefore, we feel other mechanisms, such as those behind the G_1 arrest upon KMT2D depletion, are more credible.

Although not noted in our murine cells, previous studies have showed G_1 arrest to occur (135,249). In this way, because 5-FU induces a G_1-S arrest (299-305), and gemcitabine a S-phase (306-308) and not G_1 (309) arrest, Kmt2d depletion may directly increase 5-FU sensitivity (*i.e.* enhances its effect at all concentrations), while indirectly promoting gemcitabine cytotoxicity. In addition to the cell-cycle effects, KMT2D loss has also been shown to increase genomic instability (152), where its mutation associates with a locally rearranged subtype of PDAC (42). Although not explored within this thesis, it will be worth testing whether depletion may promote increased cytotoxicity in this way, since in PDAC increased genomic instability correlates positively with response to DNA damaging agents (42).

As proposed by Kantidakis *et al.*, this increased genomic instability upon KMT2D loss might also have a role in promoting tumourigenesis (152), especially in PDAC with regards to the disease development model proposed by Hruban *et al.* (29,30), where aberration may facilitate the acquisition of driver mutations. This logic, taken together with studies showing Kmt2c aberration as promoting tumourigenesis (74,75,132,133), make it surprising that disease onset for KC and KPC mice was not altered upon *Kmt2d*

loss. While this may be due to the already aggressive nature of KPC tumours, it is plausible, and worth investigating, that *Kmt2d* loss manifests as promoting increased genetic heterogeneity and PanIN progression.

Overall, this thesis shows that reduced KMT2D expression imparts a better survival in PDAC, likely due a critical role in maintaining cell progression through the G_1-phase of the cell-cycle, which in turn may enhance sensitivity to chemotherapy. Since low *KMT2D* expression also associates with increased breast cancer survival (194,202), these findings for KMT2D might also be applicable in the study of other solid cancers. Although similarities exist between KMT2C and KMT2D, much work still remains to evaluate the specific contribution of these methyltransferases in PDAC.

Chapter 7 References

1. Hariharan D, Saied A, Kocher HM. Analysis of mortality rates for pancreatic cancer across the world. HPB : the official journal of the International Hepato Pancreato Biliary Association 2008;10:58-62.
2. Insulander J, Sanjeevi S, Haghighi M, Ivanics T, Analatos A, Lundell L, et al. Prognosis following surgical bypass compared with laparotomy alone in unresectable pancreatic adenocarcinoma. Br J Surg 2016.
3. Keane MG, Horsfall LJ, Rait G, Pereira SP. Sociodemographic trends in the incidence of pancreatic and biliary tract cancer in UK primary care. PloS one 2014;9:e108498.
4. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research 2014;74:2913-21.
5. Lowenfels AB, Maisonneuve P. Epidemiology and risk factors for pancreatic cancer. Best Pract Res Clin Gastroenterol 2006;20:197-209.
6. Schenk M, Schwartz AG, O'Neal E, Kinnard M, Greenson JK, Fryzek JP, et al. Familial risk of pancreatic cancer. Journal of the National Cancer Institute 2001;93:640-4.
7. Lynch HT, Brand RE, Hogg D, Deters CA, Fusaro RM, Lynch JF, et al. Phenotypic variation in eight extended CDKN2A germline mutation familial atypical multiple mole melanoma-pancreatic carcinoma-prone families: the familial atypical mole melanoma-pancreatic carcinoma syndrome. Cancer 2002;94:84-96.
8. Bartsch DK, Sina-Frey M, Lang S, Wild A, Gerdes B, Barth P, et al. CDKN2A germline mutations in familial pancreatic cancer. Ann Surg 2002;236:730-7.
9. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 2000;119:1447-53.
10. Su GH, Hruban RH, Bansal RK, Bova GS, Tang DJ, Shekher MC, et al. Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. The American journal of pathology 1999;154:1835-40.
11. Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nature genetics 1996;14:141-5.
12. Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK, Jr., Perrault J, et al. Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. Journal of the National Cancer Institute 1997;89:442-6.
13. Thompson D, Easton DF, Breast Cancer Linkage C. Cancer Incidence in BRCA1 mutation carriers. Journal of the National Cancer Institute 2002;94:1358-65.
14. Lynch HT, Deters CA, Snyder CL, Lynch JF, Villeneuve P, Silberstein J, et al. BRCA1 and pancreatic cancer: pedigree findings and their causal relationships. Cancer Genet Cytogenet 2005;158:119-25.
15. Hahn SA, Greenhalf B, Ellis I, Sina-Frey M, Rieder H, Korte B, et al. BRCA2 germline mutations in familial pancreatic carcinoma. Journal of the National Cancer Institute 2003;95:214-21.
16. Ozcelik H, Schmocker B, Di Nicola N, Shi XH, Langer B, Moore M, et al. Germline BRCA2 6174delT mutations in Ashkenazi Jewish pancreatic cancer patients. Nature genetics 1997;16:17-8.

17. Goggins M, Schutte M, Lu J, Moskaluk CA, Weinstein CL, Petersen GM, et al. Germline BRCA2 gene mutations in patients with apparently sporadic pancreatic carcinomas. Cancer research 1996;56:5360-4.
18. McWilliams R, Highsmith WE, Rabe KG, de Andrade M, Tordsen LA, Holtegaard LM, et al. Cystic fibrosis transmembrane regulator gene carrier status is a risk factor for young onset pancreatic adenocarcinoma. Gut 2005;54:1661-2.
19. Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, et al. Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. The New England journal of medicine 1998;339:645-52.
20. Cohn JA, Friedman KJ, Noone PG, Knowles MR, Silverman LM, Jowell PS. Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. The New England journal of medicine 1998;339:653-8.
21. Schoumacher RA, Ram J, Iannuzzi MC, Bradbury NA, Wallace RW, Hon CT, et al. A cystic fibrosis pancreatic adenocarcinoma cell line. Proceedings of the National Academy of Sciences of the United States of America 1990;87:4012-6.
22. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev 2006;20:1218-49.
23. Maitra A, Hruban RH. Pancreatic cancer. Annu Rev Pathol 2008;3:157-88.
24. Kasenda B, Bass A, Koeberle D, Pestalozzi B, Borner M, Herrmann R, et al. Survival in overweight patients with advanced pancreatic carcinoma: a multicentre cohort study. BMC Cancer 2014;14:728.
25. Michaud DS, Giovannucci E, Willett WC, Colditz GA, Stampfer MJ, Fuchs CS. Physical activity, obesity, height, and the risk of pancreatic cancer. JAMA 2001;286:921-9.
26. Ionescu-Tirgoviste C, Gagniuc PA, Gubceac E, Mardare L, Popescu I, Dima S, et al. A 3D map of the islet routes throughout the healthy human pancreas. Sci Rep 2015;5:14634.
27. Esposito I, Konukiewitz B, Schlitter AM, Kloppel G. Pathology of pancreatic ductal adenocarcinoma: facts, challenges and future developments. World J Gastroenterol 2014;20:13833-41.
28. Andea A, Sarkar F, Adsay VN. Clinicopathological correlates of pancreatic intraepithelial neoplasia: a comparative analysis of 82 cases with and 152 cases without pancreatic ductal adenocarcinoma. Mod Pathol 2003;16:996-1006.
29. Hruban RH, Adsay NV, Albores-Saavedra J, Compton C, Garrett ES, Goodman SN, et al. Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. Am J Surg Pathol 2001;25:579-86.
30. Hruban RH, Maitra A, Goggins M. Update on pancreatic intraepithelial neoplasia. International journal of clinical and experimental pathology 2008;1:306-16.
31. Maitra A, Adsay NV, Argani P, Iacobuzio-Donahue C, De Marzo A, Cameron JL, et al. Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. Mod Pathol 2003;16:902-12.
32. Moskaluk CA, Hruban RH, Kern SE. p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. Cancer research 1997;57:2140-3.
33. Yamano M, Fujii H, Takagaki T, Kadowaki N, Watanabe H, Shirai T. Genetic progression and divergence in pancreatic carcinoma. The American journal of pathology 2000;156:2123-33.

34. Heinmoller E, Dietmaier W, Zirngibl H, Heinmoller P, Scaringe W, Jauch KW, et al. Molecular analysis of microdissected tumors and preneoplastic intraductal lesions in pancreatic carcinoma. The American journal of pathology 2000;157:83-92.
35. Hruban RH, Wilentz RE, Kern SE. Genetic progression in the pancreatic ducts. The American journal of pathology 2000;156:1821-5.
36. Hruban RH, Goggins M, Parsons J, Kern SE. Progression model for pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2000;6:2969-72.
37. Luttges J, Galehdari H, Brocker V, Schwarte-Waldhoff I, Henne-Bruns D, Kloppel G, et al. Allelic loss is often the first hit in the biallelic inactivation of the p53 and DPC4 genes during pancreatic carcinogenesis. The American journal of pathology 2001;158:1677-83.
38. van Heek NT, Meeker AK, Kern SE, Yeo CJ, Lillemoe KD, Cameron JL, et al. Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia. The American journal of pathology 2002;161:1541-7.
39. Matsuda Y, Ishiwata T, Izumiyama-Shimomura N, Hamayasu H, Fujiwara M, Tomita K, et al. Gradual telomere shortening and increasing chromosomal instability among PanIN grades and normal ductal epithelia with and without cancer in the pancreas. PloS one 2015;10:e0117575.
40. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47-52.
41. Sausen M, Phallen J, Adleff V, Jones S, Leary RJ, Barrett MT, et al. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nature communications 2015;6:7686.
42. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518:495-501.
43. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nature communications 2015;6:6744.
44. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491:399-405.
45. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801-6.
46. Scarpa A, Capelli P, Mukai K, Zamboni G, Oda T, Iacono C, et al. Pancreatic adenocarcinomas frequently show p53 gene mutations. The American journal of pathology 1993;142:1534-43.
47. Redston MS, Caldas C, Seymour AB, Hruban RH, da Costa L, Yeo CJ, et al. p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. Cancer research 1994;54:3025-33.
48. Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, Kloppel G, et al. Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. British journal of cancer 1991;64:1076-82.
49. Wilentz RE, Iacobuzio-Donahue CA, Argani P, McCarthy DM, Parsons JL, Yeo CJ, et al. Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. Cancer research 2000;60:2002-6.

50. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 1996;271:350-3.
51. Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, et al. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science 1997;277:333-8.
52. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat 2007;28:622-9.
53. Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. Genes Dev 2012;26:1268-86.
54. Muller PA, Vousden KH. p53 mutations in cancer. Nat Cell Biol 2013;15:2-8.
55. Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature 2015;525:206-11.
56. Kanda M, Matthaei H, Wu J, Hong SM, Yu J, Borges M, et al. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology 2012;142:730-33 e9.
57. Goggins M, Hruban RH, Kern SE. BRCA2 is inactivated late in the development of pancreatic intraepithelial neoplasia: evidence and implications. The American journal of pathology 2000;156:1767-71.
58. Neesse A, Algul H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: a changing paradigm. Gut 2015;64:1476-84.
59. Xie D, Xie K. Pancreatic cancer stromal biology and therapy. Genes Dis 2015;2:133-43.
60. Loukopoulos P, Kanetaka K, Takamura M, Shibata T, Sakamoto M, Hirohashi S. Orthotopic transplantation models of pancreatic adenocarcinoma derived from cell lines and primary tumors and displaying varying metastatic activity. Pancreas 2004;29:193-203.
61. Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, et al. Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Molecular cancer therapeutics 2011;10:3-8.
62. Garrido-Laguna I, Uson M, Rajeshkumar NV, Tan AC, de Oliveira E, Karikari C, et al. Tumor engraftment in nude mice and enrichment in stroma- related gene pathways predict poor survival and resistance to gemcitabine in patients with pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2011;17:5793-800.
63. Rubio-Viqueira B, Jimeno A, Cusatis G, Zhang X, Iacobuzio-Donahue C, Karikari C, et al. An in vivo platform for translational drug development in pancreatic cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2006;12:4652-61.
64. Siolas D, Hannon GJ. Patient-derived tumor xenografts: transforming clinical samples into mouse models. Cancer research 2013;73:5315-9.
65. Hidalgo M, Amant F, Biankin AV, Budinska E, Byrne AT, Caldas C, et al. Patient-derived xenograft models: an emerging platform for translational cancer research. Cancer Discov 2014;4:998-1013.
66. Hoffman RM, Bouvet M. Imaging the microenvironment of pancreatic cancer patient-derived orthotopic xenografts (PDOX) growing in transgenic nude mice expressing GFP, RFP, or CFP. Cancer Lett 2015.

67. Delitto D, Pham K, Vlada AC, Sarosi GA, Thomas RM, Behrns KE, et al. Patient-derived xenograft models for pancreatic adenocarcinoma demonstrate retention of tumor morphology through incorporation of murine stromal elements. The American journal of pathology 2015;185:1297-303.
68. Hingorani SR, Petricoin EF, Maitra A, Rajapakse V, King C, Jacobetz MA, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer cell 2003;4:437-50.
69. Mazur PK, Siveke JT. Genetically engineered mouse models of pancreatic cancer: unravelling tumour biology and progressing translational oncology. Gut 2012;61:1488-500.
70. Guerra C, Barbacid M. Genetically engineered mouse models of pancreatic adenocarcinoma. Mol Oncol 2013;7:232-47.
71. Gopinathan A, Morton JP, Jodrell DI, Sansom OJ. GEMMs as preclinical models for testing pancreatic cancer therapies. Dis Model Mech 2015;8:1185-200.
72. Hingorani SR, Wang L, Multani AS, Combs C, Deramaudt TB, Hruban RH, et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer cell 2005;7:469-83.
73. McGrail M, Hatler JM, Kuang X, Liao HK, Nannapaneni K, Watt KE, et al. Somatic mutagenesis with a Sleeping Beauty transposon system leads to solid tumor formation in zebrafish. PloS one 2011;6:e18826.
74. Perez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, et al. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. Nature 2012;486:266-70.
75. Mann KM, Ward JM, Yew CC, Kovochich A, Dawson DW, Black MA, et al. Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. Proceedings of the National Academy of Sciences of the United States of America 2012;109:5934-41.
76. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development 1996;122:983-95.
77. Lohr M, Kloppel G, Maisonneuve P, Lowenfels AB, Luttges J. Frequency of K-ras mutations in pancreatic intraductal neoplasias associated with pancreatic ductal adenocarcinoma and chronic pancreatitis: a meta-analysis. Neoplasia 2005;7:17-23.
78. Guerra C, Schuhmacher AJ, Canamero M, Grippo PJ, Verdaguer L, Perez-Gallego L, et al. Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. Cancer cell 2007;11:291-302.
79. Guerra C, Collado M, Navas C, Schuhmacher AJ, Hernandez-Porras I, Canamero M, et al. Pancreatitis-induced inflammation contributes to pancreatic cancer by inhibiting oncogene-induced senescence. Cancer cell 2011;19:728-39.
80. Sukharamwala P, Thoens J, Szuchmacher M, Smith J, DeVito P. Advanced age is a risk factor for post-operative complications and mortality after a pancreaticoduodenectomy: a meta-analysis and systematic review. HPB : the official journal of the International Hepato Pancreato Biliary Association 2012;14:649-57.
81. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: a cancer journal for clinicians 2016;66:7-30.

82. Kleger A, Perkhofer L, Seufferlein T. Smarter drugs emerging in pancreatic cancer therapy. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2014;25:1260-70.
83. Neesse A, Michl P, Frese KK, Feig C, Cook N, Jacobetz MA, et al. Stromal biology and therapy in pancreatic cancer. Gut 2011;60:861-8.
84. Zhao X, Gao S, Ren H, Sun W, Zhang H, Sun J, et al. Hypoxia-inducible factor-1 promotes pancreatic ductal adenocarcinoma invasion and metastasis by activating transcription of the actin-bundling protein fascin. Cancer research 2014;74:2455-64.
85. Cheng ZX, Wang DW, Liu T, Liu WX, Xia WB, Xu J, et al. Effects of the HIF-1alpha and NF-kappaB loop on epithelial-mesenchymal transition and chemoresistance induced by hypoxia in pancreatic cancer cells. Oncology reports 2014;31:1891-8.
86. Felsenfeld G. A brief history of epigenetics. Cold Spring Harb Perspect Biol 2014;6.
87. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature genetics 2003;33 Suppl:245-54.
88. Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nature medicine 2011;17:330-9.
89. Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403:41-5.
90. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell research 2011;21:381-95.
91. Chi P, Allis CD, Wang GG. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. Nature reviews Cancer 2010;10:457-69.
92. Ali M, Hom RA, Blakeslee W, Ikenouye L, Kutateladze TG. Diverse functions of PHD fingers of the MLL/KMT2 subfamily. Biochimica et biophysica acta 2014;1843:366-71.
93. van Nuland R, Smits AH, Pallaki P, Jansen PW, Vermeulen M, Timmers HT. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. Molecular and cellular biology 2013;33:2067-77.
94. Weirich S, Kudithipudi S, Kycia I, Jeltsch A. Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. Clin Epigenetics 2015;7:36.
95. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 2000;406:593-9.
96. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Molecular cell 2012;48:491-507.
97. Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, et al. New nomenclature for chromatin-modifying enzymes. Cell 2007;131:633-6.
98. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007;129:823-37.
99. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol 2012;19:1218-27.
100. Martin C, Zhang Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 2005;6:838-49.

101. Min J, Feng Q, Li Z, Zhang Y, Xu RM. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell 2003;112:711-23.
102. Schapira M. Structural Chemistry of Human SET Domain Protein Methyltransferases. Curr Chem Genomics 2011;5:85-94.
103. Shi Y, Sawada J, Sui G, Affar el B, Whetstine JR, Lan F, et al. Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature 2003;422:735-8.
104. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 2004;119:941-53.
105. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553-60.
106. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007;130:77-88.
107. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nature genetics 2012;44:148-56.
108. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 2011;21:1273-83.
109. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics 2007;39:311-8.
110. Wang Y, Li X, Hu H. H3K4me2 reliably defines transcription factor binding regions in different cells. Genomics 2014;103:222-8.
111. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 2005;120:169-81.
112. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, et al. H3K4 tri-methylation provides an epigenetic signature of active enhancers. The EMBO journal 2011;30:4198-210.
113. Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Molecular cell 2014;53:979-92.
114. Hallson G, Hollebakken RE, Li T, Syrzycka M, Kim I, Cotsworth S, et al. dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. Genetics 2012;190:91-100.
115. Bogershausen N, Bruford E, Wollnik B. Skirting the pitfalls: a clear-cut nomenclature for H3K4 methyltransferases. Clin Genet 2013;83:212-4.
116. Rabello Ddo A, de Moura CA, de Andrade RV, Motoyama AB, Silva FP. Altered expression of MLL methyltransferase family genes in breast cancer. Int J Oncol 2013;43:653-60.
117. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res 2015;43:D257-60.
118. Zhang Y, Mittal A, Reid J, Reich S, Gamblin SJ, Wilson JR. Evolving Catalytic Properties of the MLL Family SET Domain. Structure 2015;23:1921-33.
119. Rao RC, Dou Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nature reviews Cancer 2015;15:334-46.

120. Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife 2013;2:e01503.
121. Dehe PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, et al. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. The Journal of biological chemistry 2006;281:35404-12.
122. Schlichter A, Cairns BR. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. The EMBO journal 2005;24:1222-31.
123. Malarkey CS, Churchill ME. The high mobility group box: the ultimate utility player of a cell. Trends in biochemical sciences 2012;37:553-62.
124. Mujtaba S, Zeng L, Zhou MM. Structure and acetyl-lysine recognition of the bromodomain. Oncogene 2007;26:5521-7.
125. FitzGerald KT, Diaz MO. MLL2: A new mammalian member of the trx/MLL family of genes. Genomics 1999;59:187-92.
126. Huntsman DG, Chin SF, Muleris M, Batley SJ, Collins VP, Wiedemann LM, et al. MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. Oncogene 1999;18:7975-84.
127. Natarajan TG, Kallakury BV, Sheehan CE, Bartlett MB, Ganesan N, Preet A, et al. Epigenetic regulator MLL2 shows altered expression in cancer cell lines and tumors from human breast and colon. Cancer Cell Int 2010;10:13.
128. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics 2010;42:790-3.
129. Kerimoglu C, Agis-Balboa RC, Kranz A, Stilling R, Bahari-Javan S, Benito-Garagorri E, et al. Histone-methyltransferase MLL2 (KMT2B) is required for memory formation in mice. J Neurosci 2013;33:3452-64.
130. Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. The Journal of biological chemistry 2007;282:20395-406.
131. Cho YW, Hong S, Ge K. Affinity purification of MLL3/MLL4 histone H3K4 methyltransferase complex. Methods Mol Biol 2012;809:465-72.
132. Lee S, Lee DK, Dou Y, Lee J, Lee B, Kwak E, et al. Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proceedings of the National Academy of Sciences of the United States of America 2006;103:15392-7.
133. Lee J, Kim DH, Lee S, Yang QH, Lee DK, Lee SK, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proceedings of the National Academy of Sciences of the United States of America 2009;106:8513-8.
134. Glaser S, Lubitz S, Loveland KL, Ohbo K, Robb L, Schwenk F, et al. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. Epigenetics Chromatin 2009;2:5.
135. Wan X, Liu L, Ding X, Zhou P, Yuan X, Zhou Z, et al. Mll2 controls cardiac lineage differentiation of mouse embryonic stem cells by promoting H3K4me3 deposition at cardiac-specific genes. Stem Cell Rev 2014;10:643-52.
136. Matkar S, Sharma P, Gao S, Gurung B, Katona BW, Liao J, et al. An Epigenetic Pathway Regulates Sensitivity of Breast Cancer Cells to HER2 Inhibition via FOXO/c-Myc Axis. Cancer cell 2015;28:472-85.
137. Prasad R, Zhadanov AB, Sedkov Y, Bullrich F, Druck T, Rallapalli R, et al. Structure and expression pattern of human ALR, a novel gene with strong

homology to ALL-1 involved in acute leukemia and to Drosophila trithorax. Oncogene 1997;15:549-60.
138. Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, et al. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Molecular and cellular biology 2007;27:1889-903.
139. Agha Z, Iqbal Z, Azam M, Ayub H, Vissers LE, Gilissen C, et al. Exome sequencing identifies three novel candidate genes implicated in intellectual disability. PloS one 2014;9:e112687.
140. Hopkin AS, Gordon W, Klein RH, Espitia F, Daily K, Zeller M, et al. GRHL3/GET1 and trithorax group members collaborate to activate the epidermal progenitor differentiation program. PLoS Genet 2012;8:e1002829.
141. Zhang X, Wen H, Shi X. Lysine methylation: beyond histones. Acta Biochim Biophys Sin (Shanghai) 2012;44:14-27.
142. Sebastian S, Sreenivas P, Sambasivan R, Cheedipudi S, Kandalla P, Pavlath GK, et al. MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. Proceedings of the National Academy of Sciences of the United States of America 2009;106:4719-24.
143. Fujiki R, Chikanishi T, Hashiba W, Ito H, Takada I, Roeder RG, et al. GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. Nature 2009;459:455-9.
144. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual review of biochemistry 2012;81:65-95.
145. An S, Yeo KJ, Jeon YH, Song JJ. Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. The Journal of biological chemistry 2011;286:8369-74.
146. Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, et al. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Molecular and cellular biology 2004;24:5639-49.
147. Wu L, Lee SY, Zhou B, Nguyen UT, Muir TW, Tan S, et al. ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. Molecular cell 2013;49:1108-20.
148. Patel A, Vought VE, Dharmarajan V, Cosgrove MS. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. The Journal of biological chemistry 2008;283:32162-75.
149. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Molecular cell 2013;51:310-25.
150. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology 2013;33:4745-54.
151. Guo C, Chen LH, Huang Y, Chang CC, Wang P, Pirozzi CJ, et al. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. Oncotarget 2013;4:2144-53.
152. Kantidakis T, Saponaro M, Mitter R, Horswell S, Kranz A, Boeing S, et al. Mutation of cancer driver MLL2 results in transcription stress and genome instability. Genes Dev 2016;30:408-20.

153. Tang Z, Chen WY, Shimada M, Nguyen UT, Kim J, Sun XJ, et al. SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53. Cell 2013;154:297-310.
154. Shilatifard A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annual review of biochemistry 2006;75:243-69.
155. Liu H, Galka M, Mori E, Liu X, Lin YF, Wei R, et al. A method for systematic mapping of protein lysine methylation identifies functions for HP1beta in DNA damage response. Molecular cell 2013;50:723-35.
156. Biggar KK, Li SS. Non-histone protein methylation as a regulator of cellular signalling and function. Nat Rev Mol Cell Biol 2015;16:5-17.
157. Zhang K, Lin W, Latham JA, Riefler GM, Schumacher JM, Chan C, et al. The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. Cell 2005;122:723-34.
158. Berdasco M, Esteller M. Genetic syndromes caused by mutations in epigenetic genes. Hum Genet 2013;132:359-83.
159. Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nature reviews Genetics 2013;14:765-80.
160. Liu S, Hong X, Shen C, Shi Q, Wang J, Xiong F, et al. Kabuki syndrome: a Chinese case series and systematic review of the spectrum of mutations. BMC Med Genet 2015;16:26.
161. Cheon CK, Ko JM. Kabuki syndrome: clinical and molecular characteristics. Korean J Pediatr 2015;58:317-24.
162. Bjornsson HT, Benjamin JS, Zhang L, Weissman J, Gerber EE, Chen YC, et al. Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. Science translational medicine 2014;6:256ra135.
163. Jones WD, Dafou D, McEntagart M, Woollard WJ, Elmslie FV, Holder-Espinasse M, et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. Am J Hum Genet 2012;91:358-64.
164. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014;515:209-15.
165. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LE, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet 2012;91:73-82.
166. Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. Neuron 2014;82:773-80.
167. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci 2016;19:571-7.
168. Hippisley-Cox J, Vinogradova Y, Coupland C, Parker C. Risk of malignancy in patients with schizophrenia or bipolar disorder: nested case-control study. Arch Gen Psychiatry 2007;64:1368-76.
169. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805-11.
170. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature 2013;502:333-9.

171. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495-501.
172. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nature reviews Cancer 2007;7:823-33.
173. Chen CS, Sorensen PH, Domer PH, Reaman GH, Korsmeyer SJ, Heerema NA, et al. Molecular rearrangements on chromosome 11q23 predominate in infant acute lymphoblastic leukemia and are associated with specific biologic variables and poor outcome. Blood 1993;81:2386-93.
174. Cerveira N, Lisboa S, Correia C, Bizarro S, Santos J, Torres L, et al. Genetic and clinical characterization of 45 acute leukemia patients with MLL gene rearrangements from a single institution. Mol Oncol 2012;6:553-64.
175. Chen Y, Kantarjian H, Pierce S, Faderl S, O'Brien S, Qiao W, et al. Prognostic significance of 11q23 aberrations in adult acute myeloid leukemia and the role of allogeneic stem cell transplantation. Leukemia 2013;27:836-42.
176. Tamai H, Yamaguchi H, Hamaguchi H, Yagasaki F, Bessho M, Kobayashi T, et al. Clinical features of adult acute leukemia with 11q23 abnormalities in Japan: a co-operative multicenter study. Int J Hematol 2008;87:195-202.
177. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proceedings of the National Academy of Sciences of the United States of America 2012;109:3879-84.
178. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nature genetics 2011;43:830-7.
179. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298-303.
180. Okosun J, Bodor C, Wang J, Araf S, Yang CY, Pan C, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nature genetics 2014;46:176-81.
181. Zhang J, Dominguez-Sola D, Hussein S, Lee JE, Holmes AB, Bansal M, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. Nature medicine 2015;21:1190-8.
182. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. Nature medicine 2015;21:1199-208.
183. Ford DJ, Dingwall AK. The cancer COMPASS: navigating the functions of MLL complexes in cancer. Cancer Genet 2015;208:178-91.
184. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008;455:1069-75.
185. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nature genetics 2011;43:875-8.
186. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature 2013;497:67-73.

187. Swierniak M, Pfeifer A, Stokowy T, Rusinek D, Chekan M, Lange D, et al. Somatic mutation profiling of follicular thyroid cancer by next generation sequencing. Mol Cell Endocrinol 2016;433:130-7.
188. Neff T, Armstrong SA. Recent progress toward epigenetic therapies: the example of mixed lineage leukemia. Blood 2013;121:4847-53.
189. Ansari KI, Kasiri S, Mandal SS. Histone methylase MLL1 has critical roles in tumor growth and angiogenesis and its knockdown suppresses tumor growth in vivo. Oncogene 2013;32:3359-70.
190. Milne TA, Hughes CM, Lloyd R, Yang Z, Rozenblatt-Rosen O, Dou Y, et al. Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. Proceedings of the National Academy of Sciences of the United States of America 2005;102:749-54.
191. Balakrishnan A, Bleeker FE, Lamba S, Rodolfo M, Daniotti M, Scarpa A, et al. Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. Cancer research 2007;67:3545-50.
192. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108-13.
193. Wang XX, Fu L, Li X, Wu X, Zhu Z, Fu L, et al. Somatic mutations of the mixed-lineage leukemia 3 (MLL3) gene in primary breast cancers. Pathol Oncol Res 2011;17:429-33.
194. Liu L, Kimball S, Liu H, Holowatyj A, Yang ZQ. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. Oncotarget 2015;6:2466-82.
195. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268-74.
196. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nature genetics 2012;44:760-4.
197. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010;467:1114-7.
198. Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, Schlessinger J, et al. Early and multiple origins of metastatic lineages within primary tumors. Proceedings of the National Academy of Sciences of the United States of America 2016;113:2140-5.
199. Kanda H, Nguyen A, Chen L, Okano H, Hariharan IK. The Drosophila ortholog of MLL3 and MLL4, trithorax related, functions as a negative regulator of tissue growth. Molecular and cellular biology 2013;33:1702-10.
200. Santos MA, Faryabi RB, Ergen AV, Day AM, Malhowski A, Canela A, et al. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. Nature 2014;514:107-11.
201. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. Cancer cell 2014;25:652-65.
202. Kim JH, Sharma A, Dhar SS, Lee SH, Gu B, Chan CH, et al. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. Cancer research 2014;74:1705-17.

203. Xia M, Xu L, Leng Y, Gao F, Xia H, Zhang D, et al. Downregulation of MLL3 in esophageal squamous cell carcinoma is required for the growth and metastasis of cancer cells. Tumour Biol 2015;36:605-13.
204. Bien-Willner GA, Mitra RD. Mutation and expression analysis in medulloblastoma yields prognostic variants and a putative mechanism of disease for i17q tumors. Acta Neuropathol Commun 2014;2:74.
205. Li B, Liu HY, Guo SH, Sun P, Gong FM, Jia BQ. Association of MLL3 expression with prognosis in gastric cancer. Genet Mol Res 2014;13:7513-8.
206. Ellis L, Atadja PW, Johnstone RW. Epigenetics in cancer: targeting chromatin modifications. Molecular cancer therapeutics 2009;8:1409-20.
207. Dawkins JB, Wang J, Maniati E, Heward JA, Koniali L, Kocher HM, et al. Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. Cancer research 2016;76:4861-71.
208. Lieber M, Mazzetta J, Nelson-Rees W, Kaplan M, Todaro G. Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. International journal of cancer Journal international du cancer 1975;15:741-7.
209. Kyriazis AA, Kyriazis AP, Sternberg CN, Sloane NH, Loveless JD. Morphological, biological, biochemical, and karyotypic characteristics of human pancreatic ductal adenocarcinoma Capan-2 in tissue culture and the nude mouse. Cancer research 1986;46:5810-5.
210. Ouyang H, Mou L, Luk C, Liu N, Karaskova J, Squire J, et al. Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. The American journal of pathology 2000;157:1623-31.
211. Tan MH, Nowak NJ, Loor R, Ochi H, Sandberg AA, Lopez C, et al. Characterization of a new primary human pancreatic tumor line. Cancer Invest 1986;4:15-23.
212. Iwamura T, Katsuki T, Ide K. Establishment and characterization of a human pancreatic cancer cell line (SUIT-2) producing carcinoembryonic antigen and carbohydrate antigen 19-9. Jpn J Cancer Res 1987;78:54-62.
213. Dexter DL, Matook GM, Meitner PA, Bogaars HA, Jolly GA, Turner MD, et al. Establishment and characterization of two human pancreatic cancer cell lines tumorigenic in athymic mice. Cancer research 1982;42:2705-14.
214. Morgan RT, Woods LK, Moore GE, Quinn LA, McGavran L, Gordon SG. Human cell line (COLO 357) of metastatic pancreatic adenocarcinoma. International journal of cancer Journal international du cancer 1980;25:591-8.
215. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-8.
216. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 2013;14:R36.
217. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-9.
218. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 2012;13:204-16.
219. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
220. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting

genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 2005;102:15545-50.
221. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 2010;38:D204-10.
222. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database (Oxford) 2011;2011:bar026.
223. Haider S, Wang J, Nagano A, Desai A, Arumugam P, Dumartin L, et al. A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. Genome Med 2014;6:105.
224. Mihaly Z, Kormos M, Lanczky A, Dank M, Budczies J, Szasz MA, et al. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. Breast cancer research and treatment 2013;140:219-32.
225. Lee HS, Lee HK, Kim HS, Yang HK, Kim WH. Tumour suppressor gene expression correlates with gastric cancer prognosis. The Journal of pathology 2003;200:39-46.
226. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507-16.
227. Caldas C, Hahn SA, da Costa LT, Redston MS, Schutte M, Seymour AB, et al. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nature genetics 1994;8:27-32.
228. Berrozpe G, Schaeffer J, Peinado MA, Real FX, Perucho M. Comparative analysis of mutations in the p53 and K-ras genes in pancreatic cancer. International journal of cancer Journal international du cancer 1994;58:185-91.
229. Naumann M, Savitskaia N, Eilert C, Schramm A, Kalthoff H, Schmiegel W. Frequent codeletion of p16/MTS1 and p15/MTS2 and genetic alterations in p16/MTS1 in pancreatic tumors. Gastroenterology 1996;110:1215-24.
230. Moore PS, Sipos B, Orlandini S, Sorio C, Real FX, Lemoine NR, et al. Genetic profile of 22 pancreatic carcinoma cell lines. Analysis of K-ras, p53, p16 and DPC4/Smad4. Virchows Archiv : an international journal of pathology 2001;439:798-802.
231. Huang L, Goodrow TL, Zhang SY, Klein-Szanto AJ, Chang H, Ruggeri BA. Deletion and mutation analyses of the P16/MTS-1 tumor suppressor gene in human ductal pancreatic cancer reveals a higher frequency of abnormalities in tumor-derived cell lines than in primary ductal adenocarcinomas. Cancer research 1996;56:1137-41.
232. Hsiang D, Friess H, Buchler MW, Ebert M, Butler J, Korc M. Absence of K-ras mutations in the pancreatic parenchyma of patients with chronic pancreatitis. American journal of surgery 1997;174:242-6.
233. Schutte M, Hruban RH, Hedrick L, Cho KR, Nadasdy GM, Weinstein CL, et al. DPC4 gene in various tumor types. Cancer research 1996;56:2527-30.
234. Sun C, Yamato T, Furukawa T, Ohnishi Y, Kijima H, Horii A. Characterization of the mutations of the K-ras, p53, p16, and SMAD4 genes in 15 human pancreatic cancer cell lines. Oncology reports 2001;8:89-92.
235. Aoki Y, Hosaka S, Tachibana N, Karasawa Y, Kawa S, Kiyosawa K. Reassessment of K-ras mutations at codon 12 by direct PCR and sequencing from tissue microdissection in human pancreatic adenocarcinomas. Pancreas 2000;21:152-7.

236. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nature reviews Genetics 2007;8:286-98.
237. Saiki AY, Caenepeel S, Cosgrove E, Su C, Boedigheimer M, Oliner JD. Identifying the determinants of response to MDM2 inhibition. Oncotarget 2015;6:7701-12.
238. O'Mara TA, Zhao M, Spurdle AB. Meta-analysis of gene expression studies in endometrial cancer identifies gene expression profiles associated with aggressive disease and patient outcome. Sci Rep 2016;6:36677.
239. Szasz AM, Lanczky A, Nagy A, Forster S, Hark K, Green JE, et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget 2016;7:49322-33.
240. Ouderkirk-Pecone JL, Goreczny GJ, Chase SE, Tatum AH, Turner CE, Krendel M. Myosin 1e promotes breast cancer malignancy by enhancing tumor cell proliferation and stimulating tumor cell de-differentiation. Oncotarget 2016;7:46419-32.
241. Bai Y, Li LD, Li J, Lu X. Targeting of topoisomerases for prognosis and drug resistance in ovarian cancer. J Ovarian Res 2016;9:35.
242. Rai R, Zhang F, Colavita K, Leu NA, Kurosaka S, Kumar A, et al. Arginyltransferase suppresses cell tumorigenic potential and inversely correlates with metastases in human cancers. Oncogene 2016;35:4058-68.
243. Schilsky RL, Taube SE. Tumor markers as clinical cancer tests--are we there yet? Semin Oncol 2002;29:211-2.
244. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). British journal of cancer 2005;93:387-91.
245. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med 2012;9:e1001216.
246. Leroy B, Girard L, Hollestelle A, Minna JD, Gazdar AF, Soussi T. Analysis of TP53 mutation status in human cancer cell lines: a reassessment. Hum Mutat 2014;35:756-65.
247. Deer EL, Gonzalez-Hernandez J, Coursen JD, Shea JE, Ngatia J, Scaife CL, et al. Phenotype and genotype of pancreatic cancer cell lines. Pancreas 2010;39:425-35.
248. Guo C, Chang CC, Wortham M, Chen LH, Kernagis DN, Qin X, et al. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. Proceedings of the National Academy of Sciences of the United States of America 2012;109:17603-8.
249. Ang SY, Uebersohn A, Spencer CI, Huang Y, Lee JE, Ge K, et al. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development 2016;143:810-21.
250. Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Prolif 2003;36:131-49.
251. Mo R, Rao SM, Zhu YJ. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. The Journal of biological chemistry 2006;281:15714-20.
252. Goo YH, Sohn YC, Kim DH, Kim SW, Kang MJ, Jung DJ, et al. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Molecular and cellular biology 2003;23:140-9.

253. Shin H, Shannon CP, Fishbane N, Ruan J, Zhou M, Balshaw R, et al. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. PloS one 2014;9:e91041.
254. Neuzillet C, Couvelard A, Tijeras-Raballand A, de Mestier L, de Gramont A, Bedossa P, et al. High c-Met expression in stage I-II pancreatic adenocarcinoma: proposal for an immunostaining scoring method and correlation with poor prognosis. Histopathology 2015;67:664-76.
255. Wang Q, Shen B, Chen L, Zheng P, Feng H, Hao Q, et al. Extracellular calumenin suppresses ERK1/2 signaling and cell migration by protecting fibulin-1 from MMP-13-mediated proteolysis. Oncogene 2015;34:1006-18.
256. Kondo J, Sato F, Kusumi T, Liu Y, Motonari O, Sato T, et al. Claudin-1 expression is induced by tumor necrosis factor-alpha in human pancreatic cancer cells. International journal of molecular medicine 2008;22:645-9.
257. Borka K, Kaliszky P, Szabo E, Lotz G, Kupcsulik P, Schaff Z, et al. Claudin expression in pancreatic endocrine tumors as compared with ductal adenocarcinomas. Virchows Archiv : an international journal of pathology 2007;450:549-57.
258. Wilcox A, Hinchliffe KA. Regulation of extranuclear PtdIns5P production by phosphatidylinositol phosphate 4-kinase 2alpha. FEBS Lett 2008;582:1391-4.
259. Ho-Yen CM, Jones JL, Kermorgant S. The clinical and functional significance of c-Met in breast cancer: a review. Breast Cancer Res 2015;17:52.
260. Peng Z, Zhu Y, Wang Q, Gao J, Li Y, Li Y, et al. Prognostic significance of MET amplification and expression in gastric cancer: a systematic review with meta-analysis. PloS one 2014;9:e84502.
261. Yan S, Jiao X, Zou H, Li K. Prognostic significance of c-Met in breast cancer: a meta-analysis of 6010 cases. Diagn Pathol 2015;10:62.
262. Di Renzo MF, Poulsom R, Olivero M, Comoglio PM, Lemoine NR. Expression of the Met/hepatocyte growth factor receptor in human pancreatic cancer. Cancer research 1995;55:1129-38.
263. Liu X, Yang N, Figel SA, Wilson KE, Morrison CD, Gelman IH, et al. PTPN14 interacts with and negatively regulates the oncogenic function of YAP. Oncogene 2013;32:1266-73.
264. Belle L, Ali N, Lonic A, Li X, Paltridge JL, Roslan S, et al. The tyrosine phosphatase PTPN14 (Pez) inhibits metastasis by altering protein trafficking. Sci Signal 2015;8:ra18.
265. Ogata M, Takada T, Mori Y, Oh-hora M, Uchida Y, Kosugi A, et al. Effects of overexpression of PTP36, a putative protein tyrosine phosphatase, on cell adhesion, cell growth, and cytoskeletons in HeLa cells. The Journal of biological chemistry 1999;274:12905-9.
266. Wadham C, Gamble JR, Vadas MA, Khew-Goodall Y. The protein tyrosine phosphatase Pez is a major phosphatase of adherens junctions and dephosphorylates beta-catenin. Mol Biol Cell 2003;14:2520-9.
267. Wadham C, Gamble JR, Vadas MA, Khew-Goodall Y. Translocation of protein tyrosine phosphatase Pez/PTPD2/PTP36 to the nucleus is associated with induction of cell proliferation. J Cell Sci 2000;113 (Pt 17):3117-23.
268. Huang JM, Nagatomo I, Suzuki E, Mizuno T, Kumagai T, Berezov A, et al. YAP modifies cancer cell sensitivity to EGFR and survivin inhibitors and is negatively regulated by the non-receptor type protein tyrosine phosphatase 14. Oncogene 2013;32:2220-9.
269. Mikoshiba K. IP3 receptor/Ca2+ channel: from discovery to new signaling concepts. J Neurochem 2007;102:1426-46.

270. Yabe D, Nakamura T, Kanazawa N, Tashiro K, Honjo T. Calumenin, a Ca2+-binding protein retained in the endoplasmic reticulum with a novel carboxyl-terminal sequence, HDEF. The Journal of biological chemistry 1997;272:18232-9.
271. Honore B, Vorum H. The CREC family, a novel family of multiple EF-hand, low-affinity Ca(2+)-binding proteins localised to the secretory pathway of mammalian cells. FEBS Lett 2000;466:11-8.
272. Michalak M, Robert Parker JM, Opas M. Ca2+ signaling and calcium binding chaperones of the endoplasmic reticulum. Cell Calcium 2002;32:269-78.
273. Vorum H, Hager H, Christensen BM, Nielsen S, Honore B. Human calumenin localizes to the secretory pathway and is secreted to the medium. Exp Cell Res 1999;248:473-81.
274. Jung DH, Mo SH, Kim DH. Calumenin, a multiple EF-hands Ca2+-binding protein, interacts with ryanodine receptor-1 in rabbit skeletal sarcoplasmic reticulum. Biochemical and biophysical research communications 2006;343:34-42.
275. Mazzorana M, Hussain R, Sorensen T. Ca-Dependent Folding of Human Calumenin. PloS one 2016;11:e0151547.
276. Wang Q, Feng H, Zheng P, Shen B, Chen L, Liu L, et al. The intracellular transport and secretion of calumenin-1/2 in living cells. PloS one 2012;7:e35344.
277. Sonoda N, Furuse M, Sasaki H, Yonemura S, Katahira J, Horiguchi Y, et al. Clostridium perfringens enterotoxin fragment removes specific claudins from tight junction strands: Evidence for direct involvement of claudins in tight junction barrier. J Cell Biol 1999;147:195-204.
278. Szakacs G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, Bussey KJ, et al. Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. Cancer cell 2004;6:129-37.
279. Pang L, Word B, Xu J, Wang H, Hammons G, Huang SM, et al. ATP-Binding Cassette Genes Genotype and Expression: A Potential Association with Pancreatic Cancer Development and Chemoresistance? Gastroenterology research and practice 2014;2014:414931.
280. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.
281. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
282. Watanabe Y, Castoro RJ, Kim HS, North B, Oikawa R, Hiraishi T, et al. Frequent alteration of MLL3 frameshift mutations in microsatellite deficient colorectal cancer. PloS one 2011;6:e23320.
283. Jude JG, Spencer GJ, Huang X, Somerville TD, Jones DR, Divecha N, et al. A targeted knockdown screen of genes coding for phosphoinositide modulators identifies PIP4K2A as required for acute myeloid leukemia cell proliferation and survival. Oncogene 2015;34:1253-62.
284. Niedergethmann M, Alves F, Neff JK, Heidrich B, Aramin N, Li L, et al. Gene expression profiling of liver metastases and tumour invasion in pancreatic cancer using an orthotopic SCID mouse model. British journal of cancer 2007;97:1432-40.
285. Chao YC, Pan SH, Yang SC, Yu SL, Che TF, Lin CW, et al. Claudin-1 is a metastasis suppressor and correlates with clinical outcome in lung adenocarcinoma. Am J Respir Crit Care Med 2009;179:123-33.

286. Martin TA, Jiang WG. Loss of tight junction barrier function and its role in cancer metastasis. Biochimica et biophysica acta 2009;1788:872-91.
287. Sun B-s, Yao Y-q, Pei B-x, Zhang Z-f, Wang C-l. Claudin-1 correlates with poor prognosis in lung adenocarcinoma. Thoracic Cancer 2016:n/a-n/a.
288. Ding SJ, Li Y, Shao XX, Zhou H, Zeng R, Tang ZY, et al. Proteome analysis of hepatocellular carcinoma cell strains, MHCC97-H and MHCC97-L, with different metastasis potentials. Proteomics 2004;4:982-94.
289. Wu W, Tang X, Hu W, Lotan R, Hong WK, Mao L. Identification and validation of metastasis-associated proteins in head and neck cancer cell lines by two-dimensional electrophoresis and mass spectrometry. Clin Exp Metastasis 2002;19:319-26.
290. Walsh N, Kennedy S, Larkin AM, Tryfonopoulos D, Eustace AJ, Mahgoub T, et al. Membrane transport proteins in human melanoma: associations with tumour aggressiveness and metastasis. British journal of cancer 2010;102:1157-62.
291. Kikuno R, Nagase T, Nakayama M, Koga H, Okazaki N, Nakajima D, et al. HUGE: a database for human KIAA proteins, a 2004 update integrating HUGEppi and ROUGE. Nucleic Acids Res 2004;32:D502-4.
292. Ono T, Losada A, Hirano M, Myers MP, Neuwald AF, Hirano T. Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. Cell 2003;115:109-21.
293. Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, Drechsel D, et al. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature 2004;432:1036-40.
294. Abe S, Nagasaka K, Hirayama Y, Kozuka-Hata H, Oyama M, Aoyagi Y, et al. The initial phase of chromosome condensation requires Cdk1-mediated phosphorylation of the CAP-D3 subunit of condensin II. Genes Dev 2011;25:863-74.
295. Liu W, Tanasa B, Tyurina OV, Zhou TY, Gassmann R, Liu WT, et al. PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. Nature 2010;466:508-12.
296. Lapointe J, Malhotra S, Higgins JP, Bair E, Thompson M, Salari K, et al. hCAP-D3 expression marks a prostate cancer subtype with favorable clinical behavior and androgen signaling signature. Am J Surg Pathol 2008;32:205-9.
297. Bossi D, Cicalese A, Dellino GI, Luzi L, Riva L, D'Alesio C, et al. In Vivo Genetic Screens of Patient-Derived Tumors Revealed Unexpected Frailty of the Transformed Phenotype. Cancer Discov 2016.
298. Higgins CF. Multiple molecular mechanisms for multidrug resistance transporters. Nature 2007;446:749-57.
299. Martino-Echarri E, Henderson BR, Brocardo MG. Targeting the DNA replication checkpoint by pharmacologic inhibition of Chk1 kinase: a strategy to sensitize APC mutant colon cancer cells to 5-fluorouracil chemotherapy. Oncotarget 2014;5:9889-900.
300. Filgueiras Mde C, Morrot A, Soares PM, Costa ML, Mermelstein C. Effects of 5-fluorouracil in nuclear and cellular morphology, proliferation, cell cycle, apoptosis, cytoskeletal and caveolar distribution in primary cultures of smooth muscle cells. PloS one 2013;8:e63177.
301. Guo X, Goessl E, Jin G, Collie-Duguid ES, Cassidy J, Wang W, et al. Cell cycle perturbation and acquired 5-fluorouracil chemoresistance. Anticancer Res 2008;28:9-14.
302. Li MH, Ito D, Sanada M, Odani T, Hatori M, Iwase M, et al. Effect of 5-fluorouracil on G1 phase cell cycle regulation in oral cancer cell lines. Oral Oncol 2004;40:63-70.

303. Mirjolet JF, Didelot C, Barberi-Heyob M, Merlin JL. G(1)/S but not G(0)/G(1)cell fraction is related to 5-fluorouracil cytotoxicity. Cytometry 2002;48:6-13.
304. Yoshikawa R, Kusunoki M, Yanagi H, Noda M, Furuyama JI, Yamamura T, et al. Dual antitumor effects of 5-fluorouracil on the cell cycle in colorectal carcinoma cells: a novel target mechanism concept for pharmacokinetic modulating chemotherapy. Cancer research 2001;61:1029-37.
305. Tokunaga E, Oda S, Fukushima M, Maehara Y, Sugimachi K. Differential growth inhibition by 5-fluorouracil in human colorectal carcinoma cell lines. Eur J Cancer 2000;36:1998-2006.
306. Montano R, Thompson R, Chung I, Hou H, Khan N, Eastman A. Sensitization of human cancer cells to gemcitabine by the Chk1 inhibitor MK-8776: cell cycle perturbation and impact of administration schedule in vitro and in vivo. BMC Cancer 2013;13:604.
307. Pauwels B, Korst AE, Pattyn GG, Lambrechts HA, Van Bockstaele DR, Vermeulen K, et al. Cell cycle effect of gemcitabine and its role in the radiosensitizing mechanism in vitro. Int J Radiat Oncol Biol Phys 2003;57:1075-83.
308. Carpinelli G, Bucci B, D'Agnano I, Canese R, Caroli F, Raus L, et al. Gemcitabine treatment of experimental C6 glioma: the effects on cell cycle and apoptotic rate. Anticancer Res 2006;26:3017-24.
309. Hamed SS, Straubinger RM, Jusko WJ. Pharmacodynamic modeling of cell cycle and apoptotic effects of gemcitabine on pancreatic adenocarcinoma cells. Cancer Chemother Pharmacol 2013;72:553-63.
310. Doetschman T. Influence of genetic background on genetically engineered mouse phenotypes. Methods Mol Biol 2009;530:423-33.
311. Kang YK, Kim WH, Jang JJ. Expression of G1-S modulators (p53, p16, p27, cyclin D1, Rb) and Smad4/Dpc4 in intrahepatic cholangiocarcinoma. Hum Pathol 2002;33:877-83.
312. Consortium EP. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS biology 2011;9:e1001046.
313. Yao L, Berman BP, Farnham PJ. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. Crit Rev Biochem Mol Biol 2015;50:550-73.
314. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nature reviews Genetics 2014;15:709-21.
315. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics 2013;14:390-403.
316. Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nature communications 2016;7:11485.
317. Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, et al. High nucleosome occupancy is encoded at human regulatory sequences. PloS one 2010;5:e9129.
318. Schober M, Jesenofsky R, Faissner R, Weidenauer C, Hagmann W, Michl P, et al. Desmoplasia and chemoresistance in pancreatic cancer. Cancers (Basel) 2014;6:2137-54.
319. Wagner T, Jung M. New lysine methyltransferase drug targets in cancer. Nat Biotechnol 2012;30:622-3.

320. Hui C, Ye T. Synthesis of lysine methyltransferase inhibitors. Front Chem 2015;3:44.
321. Morera L, Lubbert M, Jung M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. Clin Epigenetics 2016;8:57.
322. Azad N, Zahnow CA, Rudin CM, Baylin SB. The future of epigenetic therapy in solid tumours--lessons from the past. Nat Rev Clin Oncol 2013;10:256-66.
323. Brown R, Curry E, Magnani L, Wilhelm-Benartzi CS, Borley J. Poised epigenetic states and acquired drug resistance in cancer. Nature reviews Cancer 2014;14:747-53.
324. Nervi C, De Marinis E, Codacci-Pisanelli G. Epigenetic treatment of solid tumours: a review of clinical trials. Clin Epigenetics 2015;7:127.
325. Shen DW, Pastan I, Gottesman MM. In situ hybridization analysis of acquisition and loss of the human multidrug-resistance gene. Cancer research 1988;48:4334-9.
326. Bergman AM, Pinedo HM, Talianidis I, Veerman G, Loves WJ, van der Wilt CL, et al. Increased sensitivity to gemcitabine of P-glycoprotein and multidrug resistance-associated protein-overexpressing human cancer cell lines. British journal of cancer 2003;88:1963-70.

Appendix I Publication in *Cancer Research*

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

Therapeutics, Targets, and Chemical Biology

Cancer Research

Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma

Joshua B.N. Dawkins, Jun Wang, Eleni Maniati, James A. Heward, Lola Koniali, Hemant M. Kocher, Sarah A. Martin, Claude Chelala, Frances R. Balkwill, Jude Fitzgibbon, and Richard P. Grose

Abstract

Genes encoding the histone H3 lysine 4 methyltransferases KMT2C and KMT2D are subject to deletion and mutation in pancreatic ductal adenocarcinoma (PDAC), where these lesions identify a group of patients with a more favorable prognosis. In this study, we demonstrate that low *KMT2C* and *KMT2D* expression in biopsies also defines better outcome groups, with median survivals of 15.9 versus 9.2 months ($P = 0.029$) and 19.9 versus 11.8 months ($P = 0.001$), respectively. Experiments with eight human pancreatic cell lines showed attenuated cell proliferation when these methyltransferases were depleted, suggesting that this improved outcome may reflect a cell-cycle block with diminished progression from G_0–G_1. RNA-seq analysis of PDAC cell lines following *KMT2C* or *KMT2D* knockdown identified 31 and 124 differentially expressed genes, respectively, with 19 genes in common. Gene-set enrichment analysis revealed significant downregulation of genes related to cell-cycle and growth. These data were corroborated independently by examining *KMT2C/D* signatures extracted from the International Cancer Genome Consortium and The Cancer Genome Atlas datasets. Furthermore, these experiments highlighted a potential role for NCAPD3, a condensin II complex subunit, as an outcome predictor in PDAC using existing gene expression series. Kmt2d depletion in KC/KPC cell lines also led to an increased response to the nucleoside analogue 5-fluorouracil, suggesting that lower levels of this methyltransferase may mediate the sensitivity of PDAC to particular treatments. Therefore, it may also be therapeutically beneficial to target these methyltransferases in PDAC, especially in those patients demonstrating higher KTM2C/D expression. *Cancer Res; 76(16); 4861–71. ©2016 AACR.*

Introduction

Pancreatic ductal adenocarcinomas (PDAC) make up the majority (>90%) of all pancreatic malignancies and are associated with particularly poor overall survival (1). Patients typically present with invasion and metastases at diagnosis, limiting the opportunities for curative surgical resection. The introduction of next-generation sequencing approaches has accelerated our understanding of the recurring coding mutations present in PDAC (2–6). There appears to be a founder population of cells that have accumulated activating mutations in *KRAS* (>90%; ref. 6), alongside loss-of-function mutations in *TP53* (50%–75%; refs. 7–10) and *SMAD4* (~55%; refs. 10, 11). In addition, a significant number of other recurring copy number changes and mutations targeting components of the epigenome have been identified, including the histone lysine (K) methyltransferases *KMT2C* (*MLL3*) and *KMT2D* (*MLL2*; refs. 2–6). Intriguingly, these *KMT2C* and *KMT2D* mutations appear to identify a group of patients with better outcome relative to those with wild-type configuration (5), suggesting that depletion of these methyltransferases may either define less aggressive forms of PDAC, or serendipitously improve the efficacy of existing therapies, where the mechanisms underlying this effect are not known.

The KMT2 family of histone lysine methyltransferases consists of KMT2A (MLL1/ALL1), KMT2B (MLL2/MLL4), KMT2C (MLL3/HALR), KMT2D (MLL2/ALR/MLL4), KMT2E (MLL5), KMT2F (SET1A), KMT2G (SET1B), and KMT2H (ASH1L; ref. 12). These family members, with the exception of KMT2E and KMT2H, act as catalytic subunits within mammalian COMPASS-like complexes to catalyze the addition of methyl groups to a lysine residue on the amino tail of histone H3 (H3K4; ref. 13). H3K4 exists in unmethylated, monomethylated (H3K4me1), dimethylated (H3K4me2), and trimethylated (H3K4me3) states, where H3K4me1 is typically associated with enhancers and H3K4me3 with promoters (14). These KMT2 complexes appear to have different substrate specificities

Barts Cancer Institute, Queen Mary University of London, London, United Kingdom.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

J.B.N. Dawkins and J. Wang share joint first authorship.

J. Fitzgibbon and R.P. Grose share joint senior authorship.

Corresponding Author: Richard Grose, Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, United Kingdom. Phone: 4420-7882-3574; Fax: 4420-7882-3884; E-mail: r.p.grose@qmul.ac.uk

doi: 10.1158/0008-5472.CAN-16-0481

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

to catalyze the formation of H3K4me1 (KMT2C and KMT2D; refs. 15, 16), H3K4me1/me2 (KMT2A and KMT2B; refs. 17, 18), and H3K4me1/me2/me3 (KMT2F and KMT2G; ref. 19).

Our focus here is restricted to two of these methyltransferases identified as potential key players in PDAC. Loss of KMT2C and KMT2D in cancer is expected to impact upon gene expression; however, such changes appear to be cell type–dependent, with both negative and positive effects on cell proliferation reported (15, 20–28). We set out to understand how these methyltransferases impact upon PDAC biology, and whether they may present novel opportunities for patient stratification, personalized therapies, or even therapeutic targets.

Materials and Methods

Cell lines

Human tumor cell lines PANC-1 and Capan-2, and the immortalized human pancreatic ductal epithelial cell line, HPDE, were cultured in DMEM (Sigma Aldrich); BxPC-3, SUIT-2, RWP-1 and COLO 357 in RPMI1640 medium (Sigma Aldrich); and CFPAC-1 cells in Iscove's modified Dulbecco's medium with 25 mmol/L HEPES (Lonza) and 2 mmol/L L-glutamine (Sigma Aldrich). PANC-1, Capan-2, HPDE, BxPC-3, SUIT-2, RWP-1, and CFPAC-1 were obtained from ATCC. All human cell lines were obtained between 2008 and 2012, and authenticated between 2011 and 2016 using small tandem repeat profiling conducted by LGC standards and ATCC. DT6606, DT6586, and TB32034 cell lines, derived from the *LSL-Kras*$^{G12D/+}$;*Pdx-1-Cre* (KC; ref. 29; DT6585 and DT6606) and *LSL-Kras*$^{G12D/+}$;*LSL-Trp53*$^{R172H/+}$;*Pdx-1-Cre* (KPC; ref. 30; TB32043) mice models of PDAC, were cultured in DMEM (Sigma Aldrich). DT6606, DT6585, and TB320343 were kindly provided by David Tuveson (Cold Spring Harbor Laboratory). For each cell line, medium was supplemented with 10% heat-inactivated FBS (Life Sciences Solutions), 100 U/mL penicillin, and 100 μg/mL streptomycin (Sigma Aldrich). All cell lines were resuscitated from authentic stocks, cultured for less than two months at 37°C and 5% CO_2, and routinely screened for mycoplasma.

RNAi transfection

Cells were plated in 6-well plates with 2.5 mL growth medium without penicillin and streptomycin for forward lipid transfection with Silencer Select siRNAs (Life Sciences Solutions) and Lipofectamine RNAiMAX (Life Sciences Solutions; Supplementary Table S1). For human cell transfection, a final concentration of 8.3 nmol/L per well was achieved by combining 25 pmol siRNA with 5 μL Lipofectamine in 500-μL Opti-MEM I medium (Life Sciences Solutions) for 15 minutes, before adding to wells for 48 hours. The protocol was the same for murine cell transfection, with 150 pmol siRNA used instead to achieve a final concentration of 50 nmol/L. In all experiments, loss of *KMT2D*/*Kmt2d* or *KMT2C*/*Kmt2c* expression was confirmed by Western blot analysis or real-time PCR.

Western blot analysis

To prepare protein lysates, culture medium was removed, cells were washed twice with Dulbecco's PBS (DPBS), and RIPA buffer [Sigma-Aldrich; supplemented with protease inhibitor cocktail I (Roche) and 1:100 phosphatase inhibitor cocktail II (Sigma-Aldrich)] was added for 30 minutes on ice. Lysates were harvested with cell scrapers and cellular debris removed by 30-minute centrifugation at 16,000 rcf and 4°C. Protein quantification was determined in a bicinchoninic acid (BCA) assay [4% w/v copper (II) sulphate (Sigma-Aldrich) diluted 1:50 in bicinchoninic acid (Sigma-Aldrich)]. Samples were run on NuPAGE Novex 3%–8% Tris-Acetate or 4%–12% Bis-Tris gels (Life Sciences Solutions) and transferred to preactivated polyvinylidene difluoride membranes by iBlot dry transfer (Life Sciences Solutions). Membranes were blocked for an hour with 5% w/v BSA (Sigma-Aldrich), or 5% w/v nonfat dry milk (Marvel), in Tris-buffered saline with Tween 20 (TBST) and probed with primary antibodies diluted in blocking buffer overnight at 4°C. Membranes were washed with TBST, incubated with rabbit or mouse horseradish peroxidase–conjugated secondary antibodies diluted in blocking buffer, for an hour at room temperature, and washed with TBST (Supplementary Table S2 for primary and secondary antibodies). Staining was visualized by incubation with Amersham Enhanced Chemiluminescence (ECL), or ECL prime, Western blotting detection reagents (GE Healthcare) and developed using Super Rx X-ray films (Fujifilm) and a Konica Minolta SRX-101A medical film processor. Equivalence of protein loaded for each sample was confirmed by β-actin staining.

Cell proliferation assay

The eight cell lines were seeded in 6-well plates at either 5×10^4 cells/well (SUIT-2) or 8×10^4 cells/well (all other cell lines). After a 48-hour transfection, cells were washed with DPBS and 3 mL of fresh penicillin- and streptomycin-free media were added. Cells were cultured for 72 hours before detachment with trypsin (Sigma Aldrich) and counted using a Vi-Cell XR automated cell viability analyzer (Beckman Coulter).

Flow-cytometric analysis of cell cycle

Cells seeded in triplicate wells of 6-well plates at either 1.5×10^5 cells/well (SUIT-2), or 2×10^5 cells/well (all other cell lines), were transfected for 48 hours. Three wells per treatment were harvested and analyzed across three time points (48 hours after transfection, 16-hour treatment with 400 ng/mL nocodazole, and 24 hours after medium was replaced). Cells were pelleted by 5-minute centrifugation at 16,000 rcf and 4°C and permeabilized with 1 mL ice-cold 70% ethanol. Cells were stored at −80°C before washing with ice-cold DPBS and staining with propidium iodide [PI; 50 μg/mL PI (Sigma-Aldrich) with 100 μg/mL RNase A (Qiagen) in DPBS] for 15 minutes at 37°C. The YG610/20 filter on the Fortessa II flow cytometer was used to examine DNA staining in 30,000 cells.

Real-time PCR analysis

To validate selective knockdown of *KMT2C* or *KMT2D* mRNA by siRNAs, total RNA was isolated from cells, after 48-hour transfection, using the RNeasy Mini Kit with an on-column DNase digestion (Qiagen). Extracted RNA was quantified using a Nanodrop spectrophotometer and reverse transcribed into cDNA using a High-Capacity cDNA Reverse Transcription Kit (Life Sciences Solutions). Duplex real-time PCRs were performed using iTaq Universal Probes Supermix (Bio-Rad) and TaqMan gene expression assays (Life Sciences Solutions) to examine expression of the gene of interest and the 18S ribosomal RNA internal housekeeping standard (Supplementary Table S3). Data were analyzed using a $2^{-\Delta\Delta C_t}$ method (31) to examine relative expression of each gene, where mRNA expression levels were normalized to levels of 18S rRNA for the targeted siRNAs, and then expressed relative to

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

normalized mRNA expression levels of the control siRNA treatment.

RNA sequencing and analysis

Material for RNA sequencing (RNA-seq) was generated from cells seeded at either 1×10^5 cells/well (SUIT-2) or 1.6×10^5 cells/well (PANC-1 and COLO 357) in 6-well plates and transfected for 48 hours. Total RNA was isolated (as above), quantified, and integrity measured using a RNA 6000 nano assay kit on a 2100 BioAnalyzer (Agilent). A Ribo-Zero Gold kit (Illumina) was used to deplete rRNA from RNA samples and cDNA libraries were prepared using the TruSeq stranded total RNA library preparation kit (Illumina). Libraries prepared for sequencing were validated on the 2100 BioAnalyzer with a DNA 1000 kit (Agilent) and again after randomizing the samples into seven pools with a high sensitivity DNA analysis kit (Agilent). Paired-end sequences (reads) of 100 bp in length were generated using seven lanes of a HiSeq 2000 (Illumina). RNA-seq data have been deposited in Gene Expression Omnibus (GEO) under the accession number GSE75327 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75327).

After FASTQ data quality check using FastQC, raw reads were aligned to the reference genome hg19 using Tophat2 (32). An average of 27.2 M aligned paired-end reads (range 19.7–36.4 M), corresponding to an average of 75.0% (range 59.2%–82.2%) concordant pair alignment rate, were reported (Supplementary Table S4). The number of reads uniquely aligned (mapping quality score $q > 10$) to the exonic region of each gene were counted using HTSeq (33), based on the Ensembl annotation (version 74). *KMT2C* and *KMT2D* siRNA datasets were first analyzed independently. Only genes that achieved at least one count per million (CPM) mapped reads in at least three samples were included, leading to 15,912 and 15,818 filtered genes in total for the respective *KMT2C* and *KMT2D* siRNA datasets. These genes were classified into 15 RNA species, with protein-coding transcripts representing 81.8% and 82.3%, respectively. Read counts were further normalized using the conditional quantile normalization (cqn) method (34), accounting for gene length and GC content. Differential expression analysis was then performed using the edgeR package (35), employing the generalized linear model (GLM) approach, for each siRNA versus its control pairwise comparison, adjusting for baseline differences between the cell lines, with an additive model design as "model.matrix (~cellline+siRNA_treatment). " For each pairwise comparison, the significantly differentially expressed (DE) genes were selected using a double threshold of false discovery rate (FDR) < 0.05 and an absolute fold change of at least two. Common DE genes were then identified between different siRNA versus control comparisons within the *KMT2C* and *KMT2D* siRNA datasets. On the basis of log2 fold changes of siRNA treated over control for all filtered genes, gene-set enrichment analysis (GSEA) was performed for each comparison using the GSEA tool to identify the canonical pathways gene sets from the Molecular Signatures Database (MSigDB-C2 v5.0; ref. 36). The gene ontology (GO) biological process (BP) enrichment analysis was also performed for DE genes using the PANTHER classification system (37).

Statistical analysis of clinical and expression data

Two large PDAC datasets, the International Cancer Genome Consortium (ICGC; ref. 38) and The Cancer Genome Atlas (TCGA; ref. 3), with both gene expression and clinical follow-up data available, were used to complete survival analysis. Data on 87 patients from the ICGC dataset were previously compiled and processed (39). Level 3 gene expression data for TCGA dataset were downloaded via TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). Only annotated and confirmed PDAC patients were selected (108 in total). RNA-seq by Expectation-Maximization (RSEM) normalized expression data for 20,501 genes were obtained. For each gene, low and high expression groups were determined using the method described previously (40). Briefly, each percentile of expression between lower and upper quartiles was used in the Cox regression analysis and the best performing threshold of percentile was determined. As an example, the selection of high and low expression groups for *KMT2C* and *KMT2D* in the ICGC and TCGA datasets are shown (Supplementary Fig. S1A).

Survival modeling and Kaplan–Meier (KM) analysis was undertaken using R statistical environment ("survival" package). Overall survival (OS) was defined as time from diagnosis to death, or to the last follow-up date for survivors. Log-rank test was used to calculate the KM *P* values. The Cox proportional hazards model was fitted to every gene independently.

Two additional PDAC gene expression profiles (GEP) and clinical follow-up datasets, namely "Stratford" and "BCI_Zhang_-merged," compiled and processed previously (39), were also included for validation studies.

Cell chemotherapy response assay

The three murine cell lines (DT6606, DT6585, and TB32043) were seeded in 6-well plates at a density of 4×10^4 cells/well. After 48-hour transfection, cells were washed with DPBS, detached with trypsin, and replated in 96-well plates at 1×10^4 cells/well (DT6585 and TB32043), or 5×10^3 cells/well (DT6606). After adherence, medium was replaced with DMEM containing different concentrations of 5-fluorouracil (5-FU; Accord Healthcare). After 72-hour incubation, WST-1 reagent (Roche) was added to each well and the optical density (OD) measured at 440 nm (reference 630 nm) after 3 hours. To generate log–dose response curves, percentage of cell viability was calculated using the maximal OD as 100% viability.

Results

Decreased *KMT2C/D* expression correlates with favorable outcome in PDAC patients

Inactivation of *KMT2C* and *KMT2D* arises through a combination of gene deletion and/or mutation in PDAC (2–6). To assess whether expression levels of these methyltransferases are also linked to patient outcomes, we used gene expression profile (GEP) data from the ICGC and TCGA patient series to compare clinical features of patients with tumors expressing different levels of *KMT2C* and *KMT2D*. In the ICGC dataset, we observed that low levels of *KMT2C* and *KMT2D* expression were independently associated with better OS (*KMT2D*, median 19.9 vs. 11.8 months, log-rank $P = 0.001$, Fig. 1A; *KMT2C*, median 15.9 vs. 9.2 months, log-rank $P = 0.029$, Fig. 1B). Combined low-level expression of *KMT2C* and *KMT2D* also correlated with longer survival (median 15.9 vs. 9.2 months, log-rank $P = 0.044$, Supplementary Fig. S1B). A similar trend was also observed in TCGA dataset, for *KMT2C* and combined *KMT2C/D* expression; however, these results did not reach statistical significance (Supplementary Fig. S1C). Overall, these observations are consistent with recent genetic data

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

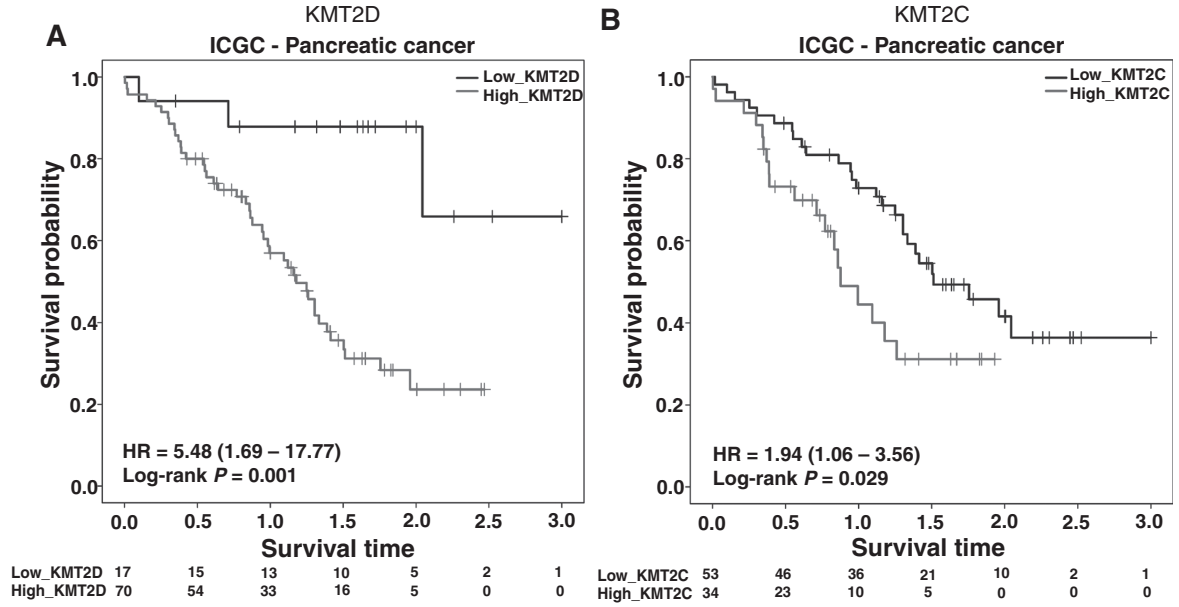

Figure 1.

KM survival analysis to assess prognostic value of *KMT2D* (**A**) and *KMT2C* (**B**) expression in the ICGC dataset. Numbers on the *x*-axis are in the unit of years. Lower expression of *KMT2C* and *KMT2D* in PDAC tumors (black) correlates with improved patient survival compared with higher expression (gray).

demonstrating that 12 of 101 patients with *KMT2C* or *KMT2D* loss-of-function mutations have a superior outcome compared with patients with a wild-type configuration (5).

Silencing of *KMT2C/D* expression reduces proliferation

To determine the cellular effects of depleting *KMT2C* and *KMT2D*, three siRNAs were used to silence both methyltransferases in a panel of eight human pancreatic cell lines; three derived from primary tumors (PANC-1, BxPC-3, Capan-2), four from metastatic sites (SUIT-2, RWP-1, CFPAC-1, COLO 357), and one immortalized from the human pancreatic ductal epithelium (HPDE; Supplementary Fig. S2A and S2B). *KMT2D* silencing by two siRNAs located to different regions of the transcript (exons 39 and 48) resulted in a reduction of KMT2D protein (Fig. 2A) and, consistent with previous studies in medulloblastoma and colorectal cancer cells (15), reduced cell proliferation significantly in all cell lines tested (Fig. 2B). A third siRNA targeting *KMT2D* produced a less pronounced effect on growth inhibition and was not included in further analyses. *KMT2C* silencing resulted in significant, albeit less pronounced, reductions in cell proliferation for the three cell lines tested (Supplementary Fig. S2C and S2D).

To examine whether these changes in cell proliferation were accompanied by cell-cycle anomalies, analysis of *KMT2D*-silenced cells was performed using propidium iodide and a G_2–M blocking agent, nocodazole. In these experiments, cells treated with control siRNA accumulated within the G_2–M phases, contrasting with an observed G_0–G_1 block in *KMT2D*-depleted cells (Fig. 3A). When the cell-cycle profile of PANC-1 cells was evaluated over several days at 24-hour intervals, *KMT2D* silencing led to a decrease in the number and proportion of cells in G_0–G_1 with a concomitant increase in the apoptotic fraction (Fig. 3B and Supplementary Fig. S2E). Taken together, these experiments suggest that KMT2D contributes towards normal cell-cycle progression, with loss of this methyltransferase leading to cell-cycle arrest and subsequent apoptosis.

KMT2C and *KMT2D* share overlapping gene signatures

To better understand the mechanism(s) responsible for the cellular effects of KMT2C/D, we assessed the transcriptional changes in response to *KMT2C* and *KMT2D* depletion in three PDAC cell lines (PANC-1, SUIT-2, and COLO 357). These cell lines were selected to include a range of background mutations and both primary and metastatic tumors. *KMT2C* and *KMT2D* were targeted individually across the three cell lines using siRNAs, with each targeted siRNA resulting in a measurable and specific reduction of *KMT2C* or *KMT2D* expression, compared with the scrambled control (Fig. 4A and B). Total RNA was isolated 48 hours after siRNA treatment and transcriptomes were assessed by RNA-seq using 100-bp paired-end deep sequencing (Supplementary Table S4). Resultant RNA-seq profiles confirmed knockdown of both *KMT2C* and *KMT2D* methyltransferases in their corresponding experiments (Supplementary Fig. S3A and S3B).

We focussed our attention on differentially expressed protein-coding genes shared between the individual siRNAs and all three PDAC cell lines, noting correlations in log2 fold changes for all quantified genes within the *KMT2C* and *KMT2D* siRNA datasets between different siRNAs compared with control (correlation coefficients ranging from 0.39 to 0.60, Supplementary Fig. S3C). As our data support a greater role for KMT2D than KMT2C, in our analysis, we report data in this order. Taking each methyltransferase separately to start, we observed a total of 567 and 759 DE genes for *KMT2D* siRNAs 1 and 2, respectively, with 124 genes common to both siRNAs and the three lines (Fig. 4C and D; Supplementary Table S5). Fold changes for the DE genes ranged

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

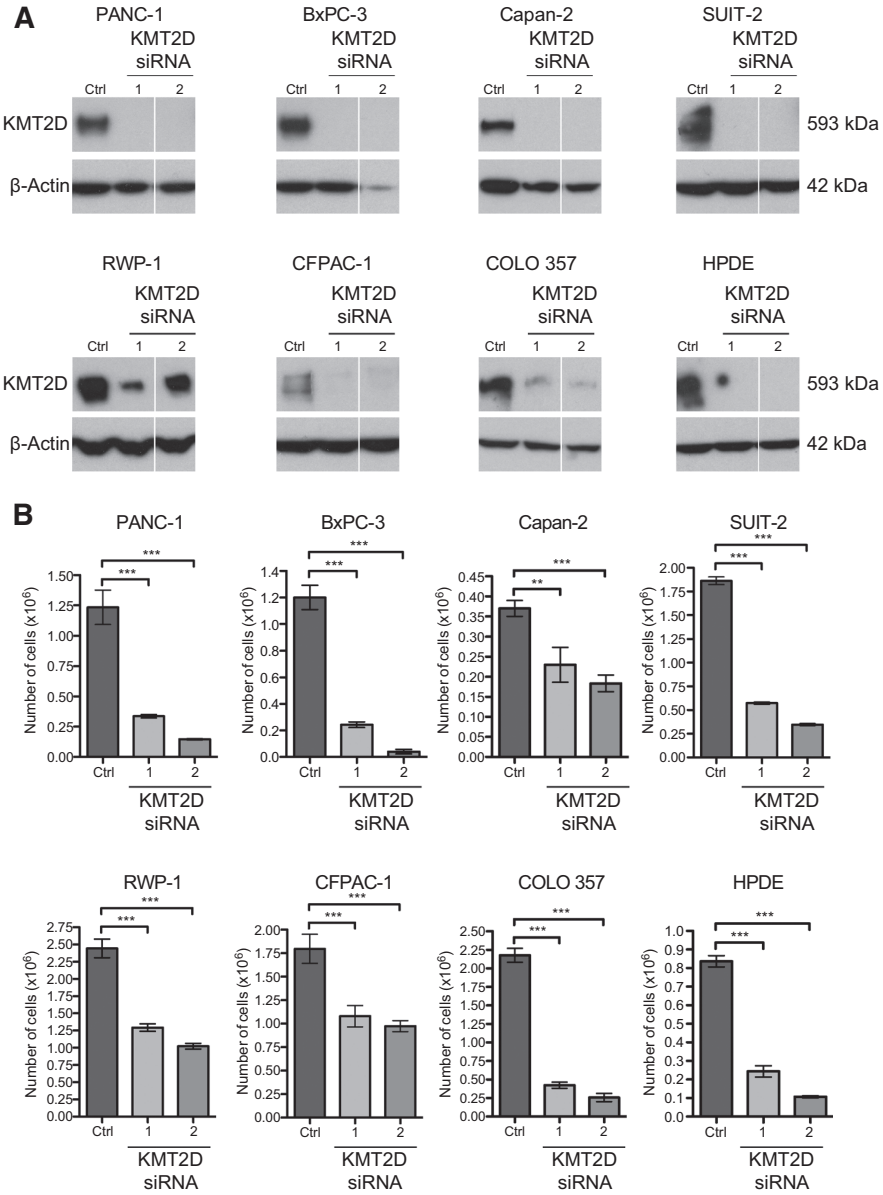

Figure 2.
Depletion of KMT2D reduces cell proliferation. **A,** Western blot analysis confirmed reduced expression of KMT2D at 72 hours after transfection with *KMT2D*-targeted siRNA duplexes in eight cell lines. A third siRNA was cropped from the images. **B,** depletion of KMT2D significantly inhibits cell proliferation (one-way ANOVA with Dunnett *post hoc* analysis, **, $P < 0.01$; ***, $P < 0.001$), where proliferation of the eight cell lines was examined by performing a cell count at 72 hours after *KMT2D* siRNA transfection. Data shown are mean values for three replicate wells performed on the same day ± SD for control siRNA, *KMT2D* siRNA1, and *KMT2D* siRNA2.

from 0.14 (downregulated, *C2orf54*) to 24.43 (upregulated, *KRT6B*), with 40 genes exhibiting a decrease, and 84 an increase, in gene expression. *KMT2C* silencing resulted in differential expression across 790 genes, with 31 common across all three siRNAs tested (Fig. 4C and D; Supplementary Table S6). For these 31 common DE genes, the fold changes ranged from 0.18 (downregulated, *AKR1B10*) to 6.51 (upregulated, *ANO3*), with 27 genes demonstrating an increase, and four a decrease, in expression. There was a striking overlap between DE genes in the *KMT2C* and *KMT2D* siRNA datasets, with 19 genes common to both methyltransferases, each demonstrating consistent directional and fold changes between the two methyltransferases (Supplementary Table S6).

Four of these genes were selected for validation based on their previous associations with pancreatic cancer [*c-MET* (2, 41), *Calumenin* (42), *Claudin-1* (43, 44), and *PTPN14* (45)], with Western blot analysis confirming expression changes of each at the protein level (Fig. 4E and Supplementary Fig. S4A). *ABCB1*, which encodes an ATP-dependent efflux pump and was decreased by *KMT2D* siRNA treatment, was also included in the validation panel, as its expression has been linked with an increase in drug resistance in pancreatic cancer (46). Decreased expression of *ABCB1*, identified by RNA-seq to be specific to *KMT2D* silencing in the two metastatic cell lines, was confirmed by Western blot analysis (Supplementary Fig. S4B). Global levels of H3K4me1/me2/3 remained largely unchanged at 48 hours after

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

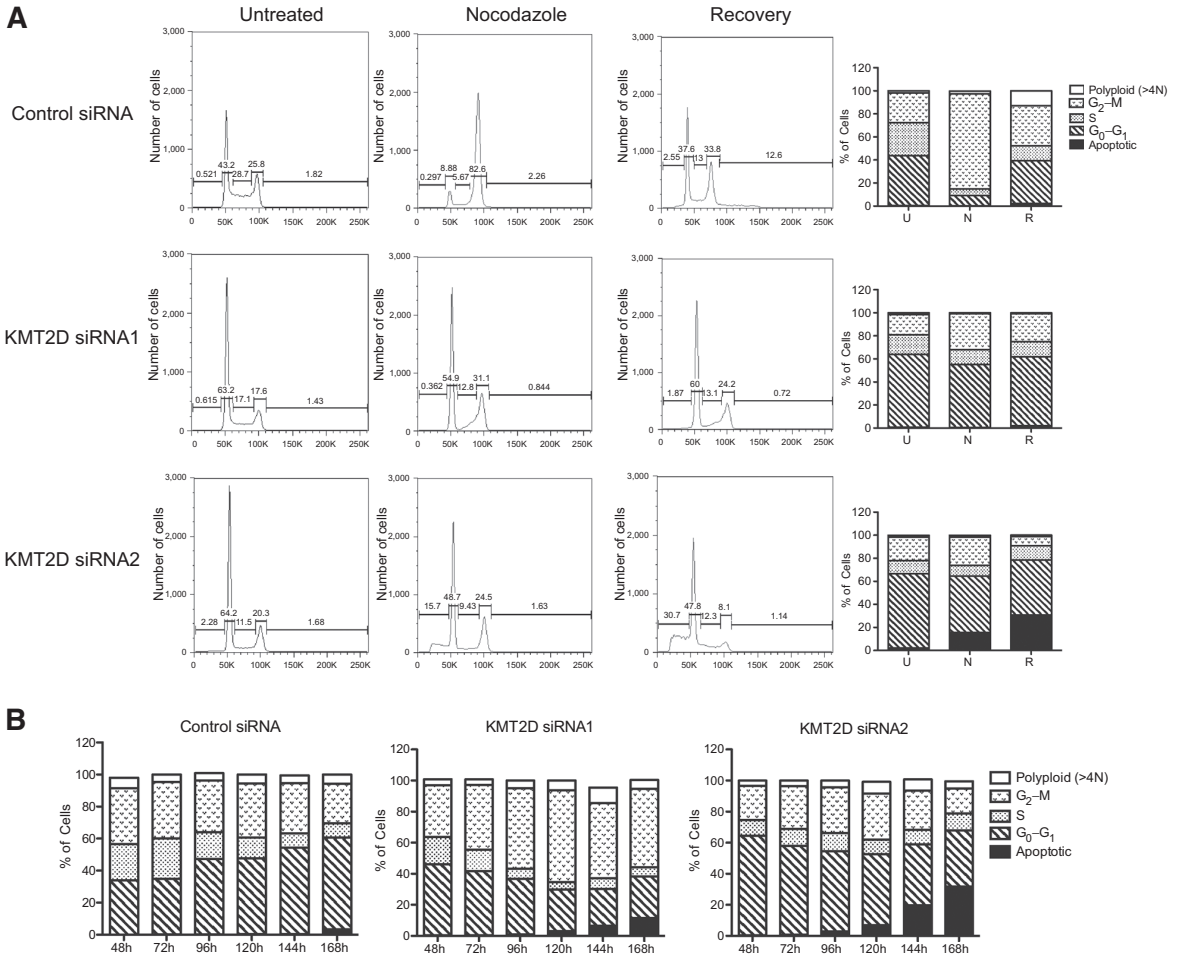

Figure 3.

Loss of KMT2D expression blocks cells in G_0–G_1 before cells undergo apoptosis. **A,** cell-cycle analysis of the SUIT-2 cell line, which is representative of the eight cell lines tested. Profiles and graphs show that cells with reduced KMT2D do not accumulate in G_2–M after 16 hours of 400 ng/mL nocodazole treatment due to their retention in the G_0–G_1 fraction. (U, untreated; N, nocodazole; R, recovery). **B,** graphs depicting the cell-cycle profiles of PANC-1 cells over 168 hours at 24-hour intervals showing that cells blocked in G_0–G_1, due to KMT2D loss, begin to undergo apoptosis.

KMT2D siRNA transfection in PANC-1 cells; however, after 120 hours, KMT2D depletion was accompanied by a global reduction in H3K4me1/3 and, to a lesser extent, H3K4me2 and total H3 (Supplementary Fig. S4C).

Analysis of pathways associated with silenced *KMT2C* or *KMT2D*

In light of the overlapping data between the transcriptomes of KMT2C- and KMT2D-depleted lines, GSEA was employed to explore the significant differences between targeted and control siRNAs in all curated canonical pathways. In support of our proliferation and cell-cycle studies, six pathways (cell-cycle, cell-cycle mitotic, DNA replication, DNA repair, mitotic M-M/G_1 phases, and Fanconi pathways) were significantly downregulated (FDR < 0.05) in all siRNA experiments compared with controls (Fig. 5A and B). Other pathways relating to cell-cycle checkpoints and apoptosis were also affected, although significance was not reached for all five pairwise comparisons. Three pathways of note (telomere maintenance, meiotic recombination and chromosome maintenance) were highly downregulated in the *KMT2D* siRNA datasets, but to a lesser extent in the *KMT2C* siRNA experiments (Fig. 5A and B).

Clinical correlation of expression of targeted genes

Using the ICGC and TCGA human pancreatic cancer datasets, we next assessed the biological process enrichment of genes that strongly correlated (Pearson correlation $P < 0.001$) with *KMT2C/D* expression levels (Supplementary Table S7). To this end, the PANTHER classification system identified a significant enrichment of the cell cycle [adjusted (adj.) $P = 1.45e{-}02$ for ICGC dataset; adj. $P = 1.02e{-}06$ for TCGA], mitosis (adj. $P = 1.62e{-}03$ for TCGA), and DNA repair (adj. $P = 1.61e{-}02$ for TCGA) pathways for genes that positively correlated with combined *KMT2C/D* expression. In addition, translation was found to

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

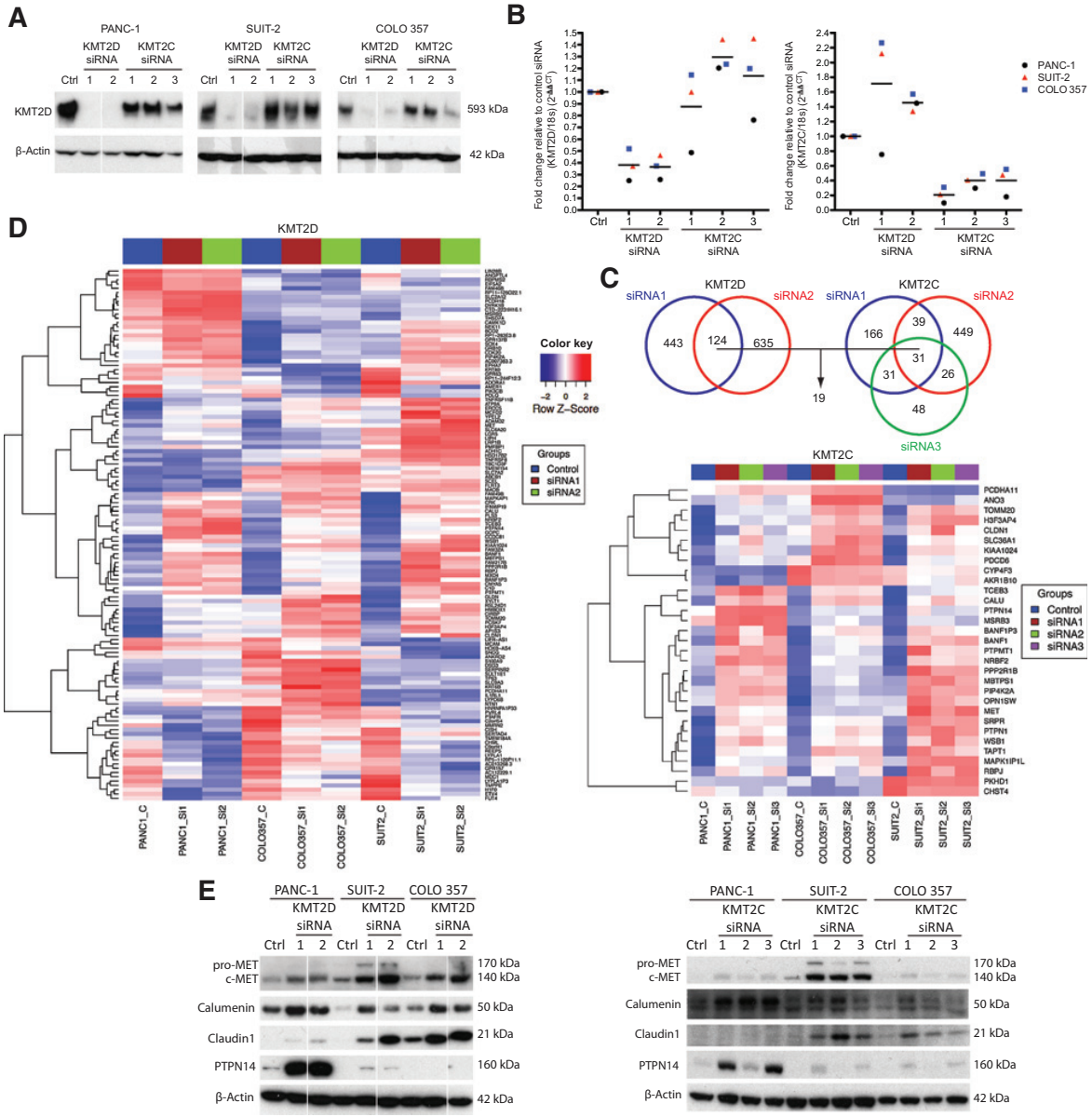

Figure 4.

RNA-seq identifies changes in gene expression following loss of *KMT2D* or *KMT2C*. **A** and **B,** *KMT2D*- and *KMT2C*-targeted siRNAs reduce expression of *KMT2D* mRNA and protein, and *KMT2C* mRNA, respectively. **A,** Western blot analysis showing the two *KMT2D* siRNAs reducing KMT2D levels in three cell lines (PANC-1, SUIT-2, and COLO 357), whereas for the three *KMT2C* and control siRNAs it remains largely unaffected. **B,** RT-qPCR analysis showing changes in *KMT2D* and *KMT2C* mRNA expression in the three cell lines [PANC-1 (black circles), SUIT-2 (red triangles), and COLO 357 (blue squares)] after transfection with two *KMT2D* siRNAs, three *KMT2C* siRNAs, or the scramble negative control siRNA. Data shown are normalized mean values from technical triplicates. **C** and **D,** bioinformatic analysis of RNA-seq data identifies DE genes common across the three cell lines for the two *KMT2D* siRNAs and three *KMT2C* siRNAs, compared with the negative control siRNA. **C,** Venn diagram depicting the commonality for the DE genes of each siRNA across the three cell lines, where 19 were found to be common across all siRNAs and cell lines. **D,** heatmaps for KMT2D and KMT2C datasets showing their 124 and 31 commonly DE genes, respectively. **E,** Western blot analysis validating expression changes of four DE genes at the protein level for the two *KMT2D* siRNAs and three *KMT2C* siRNAs. A third siRNA was cropped from images **A** and **E**.

be the most overrepresented biological pathway for genes, where expression showed significant negative correlation with *KMT2C/D* (adj. $P = 2.21e{-}04$ for ICGC and adj. $P = 5.76e{-}30$ for TCGA). The pathways observed to correlate with *KMT2C/D* expression in human pancreatic cancer datasets were therefore consistent with the most significantly downregulated and upregulated pathways following *KMT2C/D* knockdown in our PDAC cell lines (Fig. 5A).

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

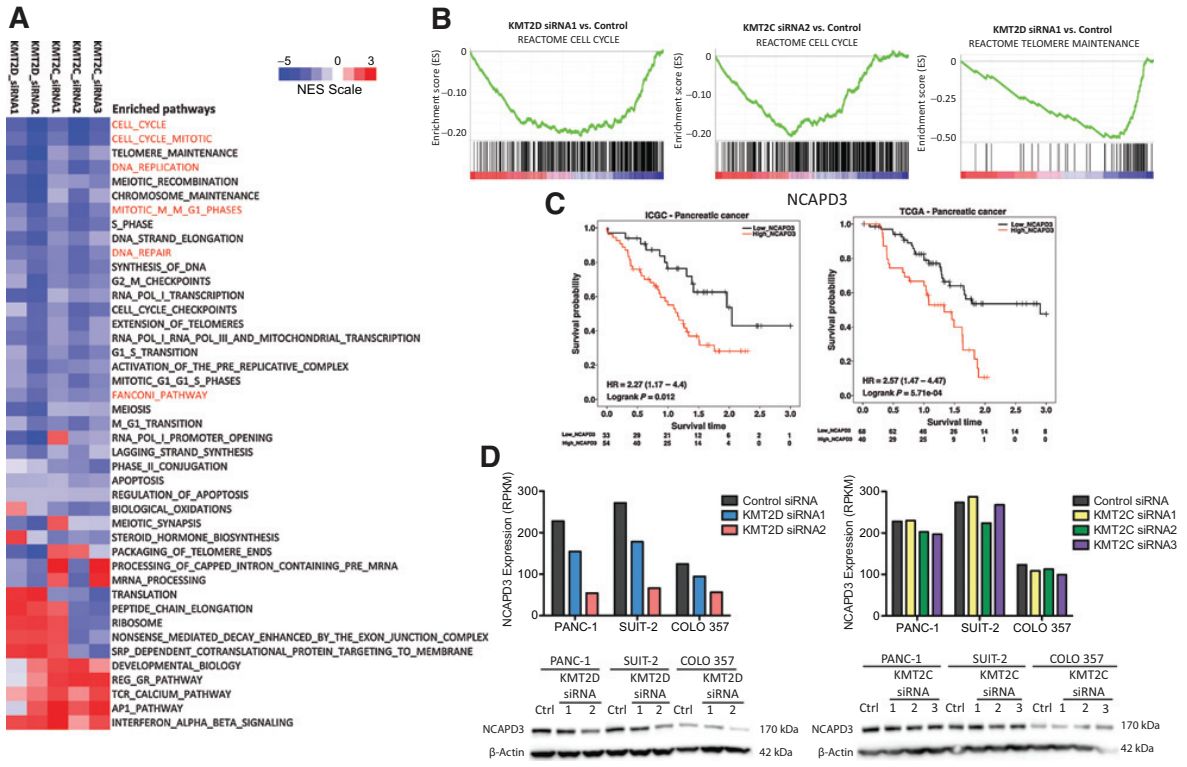

Figure 5.

GSEA identifies significant enrichment of genes relating to the cell-cycle, cell growth, and DNA repair pathways upon loss of KMT2D, potentially contributing to an improved patient outcome. **A,** heatmap showing changes in the normalized enrichment score (NES) for the upmost enriched pathways following loss of KMT2D or *KMT2C*. A negative NES indicated the downregulation in siRNA-treated samples in relation to control. Pathways significantly downregulated by all examined siRNAs are noted in red. **B,** selected enrichment score plots for genes in the REACTOME cell-cycle and telomere maintenance pathways. **C,** KM survival analysis of the ICGC and TCGA datasets to show significant negative correlations of patient survival with high (red) and low (black) expression of *NCPAD3*. Numbers on the *x*-axis represent years. **D,** expression analysis data showing significantly reduced *NCAPD3* expression, both by RNA-seq (top) and Western blotting (bottom), for the two *KMT2D* siRNAs. This reduction was minimal for the three *KMT2C* siRNAs.

Among the 94 and 257 cell-cycle genes (significantly positively correlated with combined *KMT2C/D* expression) from the ICGC and TCGA datasets, respectively (Supplementary Table S8), we identified three (*NCAPD3*, *CDKL1*, and *EIF2AK4*) as being significantly downregulated in at least one *KMT2C/D* siRNA experiment (Supplementary Table S9). Furthermore, patients with low-level expression of these three genes were found to have better OS rates in both the ICGC and TCGA datasets (Fig. 5C and Supplementary Fig. S5A). We focused primarily on NCAPD3, a subunit of the condensin II protein complex involved in chromosome condensation, due to its significant positive correlations with *KMT2C/D* expression in the ICGC and TCGA datasets (r = 0.36, P = 5.56e−04 and r = 0.38, P = 5.19e−05, respectively, Supplementary Fig. S5B), and significant associations with clinical outcome in two additional human pancreatic cancer cohorts (log-rank P = 0.042 for Stratford dataset and P = 2.12e−04 for BCI_Zhang_merged dataset, Supplementary Fig. S5C). The candidacy of *CDKL1* and *EIF2AK4* was not supported in these two additional cohorts (Supplementary Fig. S5C). Interestingly, *NCAPD3* showed significant downregulation in our *KMT2D* siRNA datasets compared with control (FDR < 0.05), whereas its expression remained almost unchanged in the *KMT2C* siRNA datasets, findings that were confirmed at the protein level (Fig. 5D). These data are consistent with GSEA pathway analysis (Fig. 5B), which highlighted a predominant role for *KMT2D*, but not *KMT2C*, in chromosome maintenance.

Kmt2d depletion increases 5-FU sensitivity

The dependence of PDAC cells on KMT2D for survival and the cell-cycle inhibition renders it an appealing therapeutic target. We postulated that lower expression, and/or mutation, could increase sensitivity to commonly used chemotherapies that target cell-cycle processes, providing one potential mechanism for favourable outcome seen in patients with lower *KMT2D* expression. This hypothesis was tested using murine *Kmt2d* siRNAs in three cell lines derived from the KC (DT6585 and DT6606; ref. 29) and KPC (TB32043; ref. 30) genetically engineered mouse models of pancreatic cancer (Fig. 6A). We were unable to perform these experiments in human cell lines as loss of KMT2D affected their cell proliferation (Fig. 2), something that was not seen in the three murine cell lines (Supplementary Fig. S6A). In both KC and KPC lines we observed an increase in their sensitivity to the nucleoside analogue 5-FU, with exposure to 10-fold less 5-FU capable of eliciting the same reduction in cell viability when used in

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

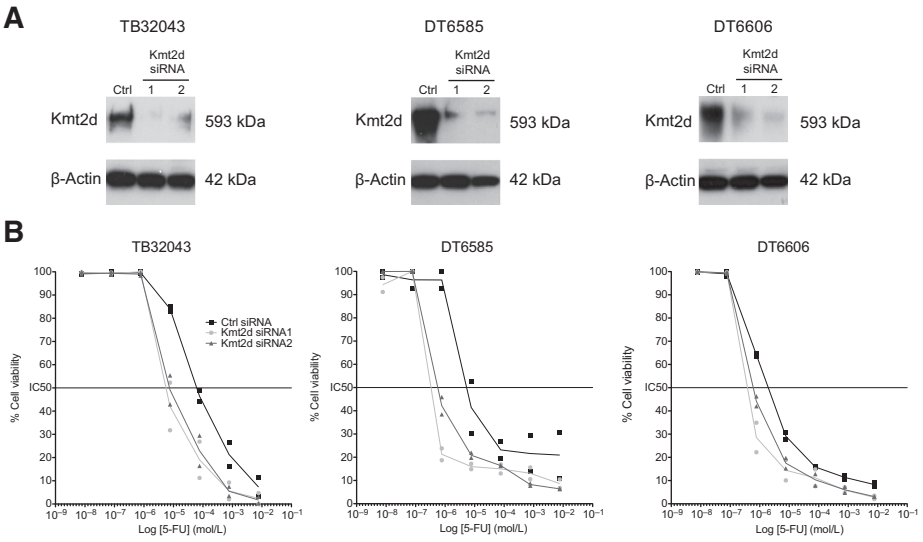

Figure 6.

Depletion of Kmt2d increases sensitivity of murine pancreatic cancer cells to 5FU. **A,** Western blot analysis showing the reduction in Kmt2d expression by two *Kmt2d* targeted siRNAs in the three murine KC (DT6606 and DT6585) and KPC (TB32043) cell lines. **B,** reduced Kmt2d expression results in a leftward shift in 5-FU dose–cell viability response curves, showing that Kmt2d depletion renders each of the cell lines more sensitive to 5-FU treatment. Cell viability was examined using WST-1 after 72 hours of 5-FU treatment. Data shown are mean OD values from technical triplicate wells normalized to maximal OD for each of two biological replicates for control siRNA (black squares), *Kmt2d* siRNA1 (light gray circles), and *Kmt2d* siRNA2 (dark gray triangles).

combination with Kmt2d depletion (Fig. 6B). Interestingly, this increased sensitivity to 5-FU was specific to Kmt2d loss (Supplementary Fig. S6B–S6D), in accordance with our human data where, relative to KMT2D, *KMT2C* loss had weaker effects on cell proliferation (Supplementary Fig. S2D) and less of an impact on patient survival (ICGC data; Fig. 1).

Discussion

The identification of mutations in several key methyl and acetyltransferases are now an established feature of many different cancers (47). Mutations in these enzymes are deemed insufficient for cancer initiation alone, with many of these lesions associated with various developmental disorders (48), where mutations in *KMT2D* are responsible for Kabuki syndrome (49). *KMT2C* and *KMT2D* represent a second tier of mutations in PDAC occurring at a lower frequency compared with mutations in *KRAS*, *TP53*, and *SMAD4*. The actual frequencies of *KMT2C* and *KMT2D* mutations in PDAC are still debatable (2–6). This is further complicated by different mechanisms of inactivation that include chromosomal deletion and variations in the nature and location of mutations, resulting in missense or truncation, and therefore loss of the functional enzymatic methyltransferase SET domain from the carboxyl-terminal.

In our study, we reasoned that inter-patient fluctuations in expression of these methyltransferases might impart significant changes in outcome. We were intrigued by our initial studies comparing matched GEP and outcome in publically available PDAC series that uncovered a strong favorable signal linked with low expression of these histone modifiers; something that has also been previously reported for *KMT2D* in breast cancer (24). It is reassuring that expression of these methyltransferases is consistent with recent studies that link mutations in these genes, and another family member (*KMT2A*), with improved overall and progression-free survival in PDAC (5).

The molecular mechanisms by which these methyltransferases contribute to PDAC development and/or influence patient outcome are likely to be complex, with the H3K4me1-3 chromatin marks potentially altering expression of many target genes. Our silencing experiments, which resulted in near complete loss of these methyltransferases and led to a marked reduction in proliferation effects of *KMT2D* in eight pancreatic cell lines, were consistent with previous studies using gene editing in other tumor settings (15), supporting some overlapping function across tumor types. Indeed, our studies go one step further by also implicating KMT2C in cell proliferation, suggesting that these proteins have somewhat complementary roles in PDAC, and may well explain the observation that *KMT2D* and *KMT2C* mutations arise independently of each other (5, 50).

We therefore went on to explore the potential downstream effectors of *KMT2C* and *KMT2D* loss in pancreatic cancer, focusing on changes common to both methyltransferases. Our RNA-seq experiments and subsequent GSEA analysis highlighted changes in DNA replication, repair, and cell cycle, echoing the results of our initial *in vitro* observations. To identify the more robust gene signatures, we focused primarily on changes consistent across the three cell lines, and siRNA experiments. This highlighted several proteins previously implicated in PDAC biology including c-MET, Calumenin, Claudin-1, and PTPN14, all of which were upregulated in our experiments and are known to impact on a range of distinct pathways and processes (41, 42, 44, 51). Indeed, all four were included in the 19-gene signature that was common to *KMT2D* and *KMT2C* in our siRNAs profiles. This gene signature, however, did not predict prognosis in any of the four PDAC GEP

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

datasets tested (ICGC, TCGA, Stratford, or BCI_Zhang_merged) based on consensus clustering and OS (data not shown), leading us to speculate that these genes may contribute instead to PDAC development. Moreover, patient clustering based on this 19-gene signature alone did not show a significant enrichment in high or low *KMT2C/D* expression (data not shown). While unexpected, this may reflect the inherent complexity in direct comparisons between primary and cell line data, and may also be a measure of the complete loss of protein induced by siRNA, a feature unlikely to be reflected physiologically, where other compensatory mechanisms might be induced. When we instead focused on genes significantly correlated with *KMT2C/D* expression from the ICGC and TCGA datasets, it was encouraging to find that significant overrepresentation of the cell-cycle signature positively correlated with *KMT2C/D* gene expression. We could refine this cell-cycle compartment by comparing against our RNA-seq expression profile, identifying several common genes with concurrent direction of fold change and magnitude. An examination of the contribution of each gene in turn, and their clinical associations, highlighted a potentially novel role for the NCAPD3 subunit of the condensin II complex in PDAC biology. This chromosome condensation protein demonstrated a 1.3–4.1 fold reduction in expression across the three tested cell lines for *KMT2D*, and encouragingly was a good predictor of outcome in all four GEP series. Having confirmed these findings at the protein level, NCAPD3 is now the focus of ongoing experiments.

Further work remains to determine the mechanisms by which loss or reduced activity of these methyltransferases associates with improved patient outcome. In some pilot experiments, we have examined changes in sensitivity to chemotherapy when Kmt2d, or Kmt2c, were depleted. For these experiments, we used cell lines derived from murine models of PDAC whose proliferation, unlike human cell lines, remained unaffected by methyltransferase depletion. Here we noted that Kmt2d silencing increased sensitivity to the antimetabolite 5-FU, suggesting that favorable outcome linked with *KMT2D* low expression might be attributable to an improved response to chemotherapy. This effect was not associated with a change in the levels of Abcb1 (data not shown), which has previously been shown to be a mediator of 5-FU response (52). This overall effect was specific to Kmt2d, as murine cells were not sensitized to 5-FU upon Kmt2c depletion, perhaps reflecting the weaker impact of low *KMT2C* expression on patient outcome.

In summary, we have identified roles for KMT2D, KMT2C, and a new role for NCAPD3 expression as prognostic predictors in PDAC. The data support the incorporation of combined *KMT2C/D* mutation and gene expression into existing risk stratification models. In addition, we report that loss of KMT2D, and to a lesser extent KMT2C, impacts on the cell-cycle and DNA replication pathways, leading to a reduction in cell proliferation. Overall, our studies point to therapeutic benefits of targeting these methyltransferases in PDAC, especially in those patients that demonstrate higher KTM2C/D expression.

Disclosure of Potential Conflicts of Interest

H.M. Kocher reports receiving a commercial research grant from Celgene and is a consultant/advisory board member for Baxalta. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: J.B.N Dawkins, E. Maniati, H.M. Kocher, J. Fitzgibbon, R.P.Grose

Development of methodology: J.B.N Dawkins, J. Wang, E. Maniati

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.B.N Dawkins, J. Wang, E. Maniati, J.A. Heward, C. Chelala

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J.B.N Dawkins, J. Wang, E. Maniati, C. Chelala, J. Fitzgibbon

Writing, review, and/or revision of the manuscript: J.B.N Dawkins, J. Wang, E. Maniati, J.A. Heward, H.M. Kocher, S.A. Martin, C. Chelala, F.R. Balkwill, J. Fitzgibbon, R.P.Grose

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.B.N Dawkins, J. Wang, E. Maniati, L. Koniali, J. Fitzgibbon, R.P.Grose

Study supervision: F.R. Balkwill, J. Fitzgibbon, R.P.Grose

Acknowledgments

The authors thank all members of the Fitzgibbon and Balkwill laboratories for advice and guidance of this project. The authors also thank Dave Tuveson and his laboratory for providing the KC and KPC cell lines and to our other colleagues and collaborators that provided access to the human pancreatic cell lines.

Grant Support

This work was supported by grants from Cancer Research UK (C587/A12888, C587/A16354), the European Research Council (ERC322566), and Barts and The London Charity (467/1307).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 18, 2016; revised May 10, 2016; accepted May 30, 2016; published OnlineFirst June 8, 2016.

References

1. Hariharan D, Saied A, Kocher HM. Analysis of mortality rates for pancreatic cancer across the world. HPB 2008;10:58–62.
2. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518:495–501.
3. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491:399–405.
4. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801–6.
5. Sausen M, Phallen J, Adleff V, Jones S, Leary RJ, Barrett MT, et al. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat Commun 2015;6:7686.
6. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47–52.
7. Scarpa A, Capelli P, Mukai K, Zamboni G, Oda T, Iacono C, et al. Pancreatic adenocarcinomas frequently show p53 gene mutations. Am J Pathol 1993;142:1534–43.
8. Redston MS, Caldas C, Seymour AB, Hruban RH, da Costa L, Yeo CJ, et al. p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. Cancer Res 1994;54:3025–33.
9. Barton CM, Staddon SL, Hughes CM, Hall PA, O'Sullivan C, Kloppel G, et al. Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. Br J Cancer 1991;64:1076–82.
10. Wilentz RE, Iacobuzio-Donahue CA, Argani P, McCarthy DM, Parsons JL, Yeo CJ, et al. Loss of expression of Dpc4 in pancreatic intraepithelial

Published OnlineFirst June 8, 2016; DOI: 10.1158/0008-5472.CAN-16-0481

neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. Cancer Res 2000;60:2002–6.
11. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 1996;271:350–3.
12. Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, et al. New nomenclature for chromatin-modifying enzymes. Cell 2007;131: 633–6.
13. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem 2012;81:65–95.
14. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, et al. H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J 2011;30:4198–210.
15. Guo C, Chen LH, Huang Y, Chang CC, Wang P, Pirozzi CJ, et al. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. Oncotarget 2013;4:2144–53.
16. Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, et al. H3K4 mono- and dimethyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife 2013;2:e01503.
17. Wu L, Lee SY, Zhou B, Nguyen UT, Muir TW, Tan S, et al. ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. Mol Cell 2013;49:1108–20.
18. Patel A, Vought VE, Dharmarajan V, Cosgrove MS. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. J Biol Chem 2008;283: 32162–75.
19. Shilatifard A. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem 2006;75:243–69.
20. Lee J, Kim DH, Lee S, Yang QH, Lee DK, Lee SK, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proc Natl Acad Sci U S A 2009;106:8513–8.
21. Kanda H, Nguyen A, Chen L, Okano H, Hariharan IK. The Drosophila ortholog of MLL3 and MLL4, trithorax related, functions as a negative regulator of tissue growth. Mol Cell Biol 2013;33:1702–10.
22. Santos MA, Faryabi RB, Ergen AV, Day AM, Malhowski A, Canela A, et al. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. Nature 2014;514:107–11.
23. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. Cancer Cell 2014;25:652–65.
24. Kim JH, Sharma A, Dhar SS, Lee SH, Gu B, Chan CH, et al. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. Cancer Res 2014;74: 1705–17.
25. Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, et al. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol 2007;27:1889–903.
26. Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature 2015;525:206–11.
27. Zhang J, Dominguez-Sola D, Hussein S, Lee JE, Holmes AB, Bansal M, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. Nat Med 2015;21:1190–8.
28. Matkar S, Sharma P, Gao S, Gurung B, Katona BW, Liao J, et al. An epigenetic pathway regulates sensitivity of breast cancer cells to HER2 inhibition via FOXO/c-Myc axis. Cancer Cell 2015;28:472–85.
29. Hingorani SR, Petricoin EF, Maitra A, Rajapakse V, King C, Jacobetz MA, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell 2003;4:437–50.
30. Hingorani SR, Wang L, Multani AS, Combs C, Deramaudt TB, Hruban RH, et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell 2005;7:469–83.
31. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402–8.
32. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013;14:R36.
33. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–9.
34. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 2012;13: 204–16.
35. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.
36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
37. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 2010;38:D204–10.
38. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal–a one-stop shop for cancer genomics data. Database (Oxford) 2011;2011:bar026.
39. Haider S, Wang J, Nagano A, Desai A, Arumugam P, Dumartin L, et al. A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. Genome Med 2014;6:105.
40. Mihaly Z, Kormos M, Lanczky A, Dank M, Budczies J, Szasz MA, et al. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. Breast Cancer Res Treat 2013;140:219–32.
41. Neuzillet C, Couvelard A, Tijeras-Raballand A, de Mestier L, de Gramont A, Bedossa P, et al. High c-Met expression in stage I-II pancreatic adenocarcinoma: proposal for an immunostaining scoring method and correlation with poor prognosis. Histopathology 2015;67:664–76.
42. Wang Q, Shen B, Chen L, Zheng P, Feng H, Hao Q, et al. Extracellular calumenin suppresses ERK1/2 signaling and cell migration by protecting fibulin-1 from MMP-13-mediated proteolysis. Oncogene 2015;34: 1006–18.
43. Kondo J, Sato F, Kusumi T, Liu Y, Motonari O, Sato T, et al. Claudin-1 expression is induced by tumor necrosis factor-alpha in human pancreatic cancer cells. Int J Mol Med 2008;22:645–9.
44. Borka K, Kaliszky P, Szabo E, Lotz G, Kupcsulik P, Schaff Z, et al. Claudin expression in pancreatic endocrine tumors as compared with ductal adenocarcinomas. Virchows Arch 2007;450:549–57.
45. Niedergethmann M, Alves F, Neff JK, Heidrich B, Aramin N, Li L, et al. Gene expression profiling of liver metastases and tumour invasion in pancreatic cancer using an orthotopic SCID mouse model. Br J Cancer 2007;97: 1432–40.
46. Pang L, Word B, Xu J, Wang H, Hammons G, Huang SM, et al. ATP-binding cassette genes genotype and expression: a potential association with pancreatic cancer development and chemoresistance? Gastroenterol Res Pract 2014;2014:414931.
47. Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nat Rev Genet 2013;14:765–80.
48. Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. Annu Rev Genomics Hum Genet 2014;15:269–93.
49. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010;42:790–3.
50. Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun 2015;6:6744.
51. Huang JM, Nagatomo I, Suzuki E, Mizuno T, Kumagai T, Berezov A, et al. YAP modifies cancer cell sensitivity to EGFR and survivin inhibitors and is negatively regulated by the non-receptor type protein tyrosine phosphatase 14. Oncogene 2013;32:2220–9.
52. Lu F, Hou YQ, Song Y, Yuan ZJ. TFPI-2 downregulates multidrug resistance protein in 5-FU-resistant human hepatocellular carcinoma BEL-7402/5-FU cells. Anat Rec (Hoboken) 2013;296:56–63.

Appendix II List of genes commonly differentially expressed upon loss of KMT2C

Downregulated

Name	Description
AKR1B10	aldo-keto reductase family 1, member B10 (aldose reductase) [Source: HGNC Symbol; Acc:382]
CHST4	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4 [Source: HGNC Symbol; Acc:1972]
CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3 [Source: HGNC Symbol; Acc:2646]
PKHD1	polycystic kidney and hepatic disease 1 (autosomal recessive) [Source: HGNC Symbol; Acc:9016]

Upregulated

Name	Description
ANO3	anoctamin 3 [Source: HGNC Symbol; Acc:14004]
BANF1	barrier to autointegration factor 1 [Source: HGNC Symbol; Acc:17397]
BANF1P3	barrier to autointegration factor 1 pseudogene 3 [Source: HGNC Symbol; Acc:43883]
CALU	calumenin [Source: HGNC Symbol; Acc:1458]
CLDN1	claudin 1 [Source: HGNC Symbol; Acc:2032]
H3F3AP4	H3 histone, family 3A, pseudogene 4 [Source: HGNC Symbol; Acc:42980]
KIAA1024	KIAA1024 [Source: HGNC Symbol; Acc:29172]
MAPK1IP1L	mitogen-activated protein kinase 1 interacting protein 1-like [Source: HGNC Symbol; Acc:19840]
MBTPS1	membrane-bound transcription factor peptidase, site 1 [Source: HGNC Symbol; Acc:15456]
MET	met proto-oncogene [Source: HGNC Symbol; Acc:7029]
MSRB3	methionine sulfoxide reductase B3 [Source: HGNC Symbol; Acc:27375]
NRBF2	nuclear receptor binding factor 2 [Source: HGNC Symbol; Acc:19692]
OPN1SW	opsin 1 (cone pigments), short-wave-sensitive [Source: HGNC Symbol; Acc:1012]
PCDHA11	protocadherin alpha 11 [Source: HGNC Symbol; Acc:8665]
PDCD6	programmed cell death 6 [Source: HGNC Symbol; Acc:8765]
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha [Source: HGNC Symbol; Acc:8997]
PPP2R1B	protein phosphatase 2, regulatory subunit A, beta [Source: HGNC Symbol; Acc:9303]
PTPMT1	protein tyrosine phosphatase, mitochondrial 1 [Source: HGNC Symbol; Acc:26965]
PTPN1	protein tyrosine phosphatase, non-receptor type 1 [Source: HGNC Symbol; Acc:9642]
PTPN14	protein tyrosine phosphatase, non-receptor type 14 [Source: HGNC Symbol; Acc:9647]
RBPJ	recombination signal binding protein for immunoglobulin kappa J region [Source: HGNC Symbol; Acc:5724]
SLC36A1	solute carrier family 36 (proton/amino acid symporter), member 1 [Source: HGNC Symbol; Acc:18761]
SRPR	signal recognition particle receptor (docking protein) [Source: HGNC Symbol; Acc:11307]
TAPT1	transmembrane anterior posterior transformation 1 [Source: HGNC Symbol; Acc:26887]
TCEB3	transcription elongation factor B (SIII), polypeptide 3 (110kDa, elongin A) [Source: HGNC Symbol; Acc:11620]
TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast) [Source: HGNC Symbol; Acc:20947]
WSB1	WD repeat and SOCS box containing 1 [Source: HGNC Symbol; Acc:19221]

Appendix III List of genes commonly differentially expressed upon loss of KMT2D

Downregulated

Name	Description
PTAFR	platelet-activating factor receptor [Source: HGNC Symbol; Acc:9582]
C2orf54	chromosome 2 open reading frame 54 [Source: HGNC Symbol; Acc:26216]
SERTAD4	SERTA domain containing 4 [Source: HGNC Symbol; Acc:25236]
LIN28B	lin-28 homolog B (C. elegans) [Source: HGNC Symbol; Acc:32207]
CISH	cytokine inducible SH2-containing protein [Source: HGNC Symbol; Acc:1984]
PVRL4	poliovirus receptor-related 4 [Source: HGNC Symbol; Acc:19688]
ANKRD2	ankyrin repeat domain 2 (stretch responsive muscle) [Source: HGNC Symbol; Acc:495]
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta [Source: HGNC Symbol; Acc:8976]
KRT80	keratin 80 [Source: HGNC Symbol; Acc:27056]
LIFR-AS1	LIFR antisense RNA 1 [Source: HGNC Symbol; Acc:43600]
HNRNPA1P33	heterogeneous nuclear ribonucleoprotein A1 pseudogene 33 [Source: HGNC Symbol; Acc:44990]
SNCG	synuclein, gamma (breast cancer-specific protein 1) [Source: HGNC Symbol; Acc:11141]
RBPMS2	RNA binding protein with multiple splicing 2 [Source: HGNC Symbol; Acc:19098]
RP5-1120P11.1	-
ANGPTL4	angiopoietin-like 4 [Source: HGNC Symbol; Acc:16039]
AC112229.1	-
GPR63	G protein-coupled receptor 63 [Source: HGNC Symbol; Acc:13302]
FAM46B	family with sequence similarity 46, member B [Source: HGNC Symbol; Acc:28273]
LYPLA1P3	lysophospholipase I pseudogene 3 [Source: HGNC Symbol; Acc:44007]
MCAM	melanoma cell adhesion molecule [Source: HGNC Symbol; Acc:6934]
MMRN2	multimerin 2 [Source: HGNC Symbol; Acc:19888]
FUT4	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific) [Source: HGNC Symbol; Acc:4015]
AMER1	APC membrane recruitment protein 1 [Source: HGNC Symbol; Acc:26837]
AC013268.3	-
ADORA1	adenosine A1 receptor [Source: HGNC Symbol; Acc:262]
HOXB-AS4	HOXB cluster antisense RNA 4 [Source: HGNC Symbol; Acc:40285]
RP11-244F12.3	-
POLQ	polymerase (DNA directed), theta [Source: HGNC Symbol; Acc:9186]
TMPPE	transmembrane protein with metallophosphoesterase domain [Source: HGNC Symbol; Acc:33865]
CHML	choroideremia-like (Rab escort protein 2) [Source: HGNC Symbol; Acc:1941]
TMEM184A	transmembrane protein 184A [Source: HGNC Symbol; Acc:28797]
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1 [Source: HGNC Symbol; Acc:40]
GPR157	G protein-coupled receptor 157 [Source: HGNC Symbol; Acc:23687]
EIF5A2	eukaryotic translation initiation factor 5A2 [Source: HGNC Symbol; Acc:3301]
C9orf41	chromosome 9 open reading frame 41 [Source: HGNC Symbol; Acc:23435]
LYPLA1	lysophospholipase I [Source: HGNC Symbol; Acc:6737]
ETV4	ets variant 4 [Source: HGNC Symbol; Acc:3493]
MDC1	mediator of DNA-damage checkpoint 1 [Source: HGNC Symbol; Acc:21163]
REEP5	receptor accessory protein 5 [Source: HGNC Symbol; Acc:30077]
H1F0	H1 histone family, member 0 [Source: HGNC Symbol; Acc:4714]

Upregulated

Name	Description
SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12 [Source: HGNC Symbol; Acc:18067]
CRK	v-crk avian sarcoma virus CT10 oncogene homolog [Source: HGNC Symbol; Acc:2362]
CDK20	cyclin-dependent kinase 20 [Source: HGNC Symbol; Acc:21420]
SCEL	sciellin [Source: HGNC Symbol; Acc:10573]
TCEB3	transcription elongation factor B (SIII), polypeptide 3 (110kDa, elongin A) [Source: HGNC Symbol; Acc:11620]
THSD7A	thrombospondin, type I, domain containing 7A [Source: HGNC Symbol; Acc:22207]
ERCC5	excision repair cross-complementing rodent repair deficiency, complementation group 5 [Source: HGNC Symbol; Acc:3437]
RSL24D1	ribosomal L24 domain containing 1 [Source: HGNC Symbol; Acc:18479]
PCSK7	proprotein convertase subtilisin/kexin type 7 [Source: HGNC Symbol; Acc:8748]
C1D	C1D nuclear receptor corepressor [Source: HGNC Symbol; Acc:29911]
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha [Source: HGNC Symbol; Acc:8997]
DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B [Source: HGNC Symbol; Acc:3092]
HMBOX1	homeobox containing 1 [Source: HGNC Symbol; Acc:26137]
IL1RL1	interleukin 1 receptor-like 1 [Source: HGNC Symbol; Acc:5998]
CAMK1D	calcium/calmodulin-dependent protein kinase ID [Source: HGNC Symbol; Acc:19341]
RP1-283E3.8	-
DSG3	desmoglein 3 [Source: HGNC Symbol; Acc:3050]
CIRBP	cold inducible RNA binding protein [Source: HGNC Symbol; Acc:1982]
NRBF2	nuclear receptor binding factor 2 [Source: HGNC Symbol; Acc:19692]
PLS3	plastin 3 [Source: HGNC Symbol; Acc:9091]
EPHA7	EPH receptor A7 [Source: HGNC Symbol; Acc:3390]
TP63	tumor protein p63 [Source: HGNC Symbol; Acc:15979]
FAM32A	family with sequence similarity 32, member A [Source: HGNC Symbol; Acc:24563]
LIPH	lipase, member H [Source: HGNC Symbol; Acc:18483]
SLC9A3	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 [Source: HGNC Symbol; Acc:11073]
NEK11	NIMA-related kinase 11 [Source: HGNC Symbol; Acc:18593]
GRB10	growth factor receptor-bound protein 10 [Source: HGNC Symbol; Acc:4564]
MAPKAP1	mitogen-activated protein kinase associated protein 1 [Source: HGNC Symbol; Acc:18752]
BANF1	barrier to autointegration factor 1 [Source: HGNC Symbol; Acc:17397]
PCDH18	protocadherin 18 [Source: HGNC Symbol; Acc:14268]
ATP9A	ATPase, class II, type 9A [Source: HGNC Symbol; Acc:13540]
NTN1	netrin 1 [Source: HGNC Symbol; Acc:8029]
FLRT3	fibronectin leucine rich transmembrane protein 3 [Source: HGNC Symbol; Acc:3762]
LRP1B	low density lipoprotein receptor-related protein 1B [Source: HGNC Symbol; Acc:6693]
AP1S3	adaptor-related protein complex 1, sigma 3 subunit [Source: HGNC Symbol; Acc:18971]

SOX4	SRY (sex determining region Y)-box 4 [Source: HGNC Symbol; Acc:11200]
BCO2	beta-carotene oxygenase 2 [Source: HGNC Symbol; Acc:18503]
HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2 [Source: HGNC Symbol; Acc:5211]
AC007383.3	-
MBTPS1	membrane-bound transcription factor peptidase, site 1 [Source: HGNC Symbol; Acc:15456]
SLC7A2	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2 [Source: HGNC Symbol; Acc:11060]
GOPC	golgi-associated PDZ and coiled-coil motif containing [Source: HGNC Symbol; Acc:17643]
PTPN14	protein tyrosine phosphatase, non-receptor type 14 [Source: HGNC Symbol; Acc:9647]
H3F3AP4	H3 histone, family 3A, pseudogene 4 [Source: HGNC Symbol; Acc:42980]
MAOB	monoamine oxidase B [Source: HGNC Symbol; Acc:6834]
TMEM154	transmembrane protein 154 [Source: HGNC Symbol; Acc:26489]
FAM217B	family with sequence similarity 217, member B [Source: HGNC Symbol; Acc:16170]
TNFRSF8	tumor necrosis factor receptor superfamily, member 8 [Source: HGNC Symbol; Acc:11923]
PCDHA11	protocadherin alpha 11 [Source: HGNC Symbol; Acc:8665]
MET	met proto-oncogene [Source: HGNC Symbol; Acc:7029]
FAM49B	family with sequence similarity 49, member B [Source: HGNC Symbol; Acc:25216]
MCFD2	multiple coagulation factor deficiency 2 [Source: HGNC Symbol; Acc:18451]
TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast) [Source: HGNC Symbol; Acc:20947]
CALU	calumenin [Source: HGNC Symbol; Acc:1458]
WSB1	WD repeat and SOCS box containing 1 [Source: HGNC Symbol; Acc:19221]
ADAM32	ADAM metallopeptidase domain 32 [Source: HGNC Symbol; Acc:15479]
CLDN1	claudin 1 [Source: HGNC Symbol; Acc:2032]
SLC6A20	solute carrier family 6 (proline IMINO transporter), member 20 [Source: HGNC Symbol; Acc:30927]
KIAA1024	KIAA1024 [Source: HGNC Symbol; Acc:29172]
CCDC81	coiled-coil domain containing 81 [Source: HGNC Symbol; Acc:26281]
PPP2R1B	protein phosphatase 2, regulatory subunit A, beta [Source: HGNC Symbol; Acc:9303]
PTPMT1	protein tyrosine phosphatase, mitochondrial 1 [Source: HGNC Symbol; Acc:26965]
RP11-126O22.1	-
GPR137B	G protein-coupled receptor 137B [Source: HGNC Symbol; Acc:11862]
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide [Source: HGNC Symbol; Acc:251]
LGR5	leucine-rich repeat containing G protein-coupled receptor 5 [Source: HGNC Symbol; Acc:4504]
RBPJ	recombination signal binding protein for immunoglobulin kappa J region [Source: HGNC Symbol; Acc:5724]
XYLT1	xylosyltransferase I [Source: HGNC Symbol; Acc:15516]
S100A9	S100 calcium binding protein A9 [Source: HGNC Symbol; Acc:10499]
MSRB3	methionine sulfoxide reductase B3 [Source: HGNC Symbol; Acc:27375]
GLDN	gliomedin [Source: HGNC Symbol; Acc:29514]
MXD4	MAX dimerization protein 4 [Source: HGNC Symbol; Acc:13906]
CMYA5	cardiomyopathy associated 5 [Source: HGNC Symbol; Acc:14305]
BANF1P3	barrier to autointegration factor 1 pseudogene 3 [Source: HGNC Symbol; Acc:43883]

YPEL2	yippee-like 2 (Drosophila) [Source: HGNC Symbol; Acc:18326]
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2 [Source: HGNC Symbol; Acc:8584]
SULT1E1	sulfotransferase family 1E, estrogen-preferring, member 1 [Source: HGNC Symbol; Acc:11377]
PMFBP1	polyamine modulated factor 1 binding protein 1 [Source: HGNC Symbol; Acc:17728]
IFNWP19	interferon, omega 1 pseudogene 19 [Source: HGNC Symbol; Acc:5451]
TNFRSF11B	tumor necrosis factor receptor superfamily, member 11b [Source: HGNC Symbol; Acc:11909]
TBC1D3F	TBC1 domain family, member 3F [Source: HGNC Symbol; Acc:18257]
LYPD6B	LY6/PLAUR domain containing 6B [Source: HGNC Symbol; Acc:27018]
CTD-2231H16.1	-
KRT6B	keratin 6B [Source: HGNC Symbol; Acc:6444]