GENERATION AND CHARACTERIZATION OF A LYSINE (K)-SPECIFIC METHYLTRANSFERASE 2D KNOCKOUT HUMAN CELL LINE

by

Ryan Donald Huff

B.MSc., The University of Western Ontario, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2015

© Ryan Donald Huff, 2015
Abstract
Lysine (K)-specific methyltransferase 2D (*KMT2D*) encodes a histone-lysine N-methyltransferase that catalyzes the methylation of histone 3 lysine 4 (H3K4me), which is an epigenetic modification involved in transcriptional regulation. *KMT2D* is a recurrent target of somatic mutation in at least 27 types of cancer, with the majority of *KMT2D* mutations (54%) predicted to result in the loss of protein function. In particular, *KMT2D* is mutated in ~85% of patients with follicular lymphoma, with ~50% of cases harboring multiple mutations in *KMT2D*. Disruption of KMT2D function has been linked to a rare pediatric disorder named Kabuki syndrome where ~75% of patients harbour heterozygous loss of function (LOF) mutations. To investigate the impact of LOF *KMT2D* mutations on H3K4 methylation and transcription I inactivated *KMT2D* using zinc finger nuclease (ZFN) technology in the human cell line HEK293A. Consistent with previous studies, HEK293 *KMT2D* LOF cell lines demonstrated loss of KMT2D was sufficient to reduce bulk mono- and dimethylation of H3K4 in the cell. Previous studies have demonstrated that KMT2D’s epigenetic function is involved in nuclear hormone transactivation, and that disruption of nuclear hormone signaling via the retinoic acid receptor (RAR) leads to lymphomagenesis in mouse models. To study the link between RAR signaling and KMT2D, I investigated RAR signaling in HEK293 *KMT2D* LOF cell lines. I observed *KMT2D* was necessary for robust induction of RAR response genes *RARA2*, *RARB2*, and *RARG* in the presence of 9-cis-retinoic acid. These results are compatible with the notion that LOF *KMT2D* mutations may aid cancer cells in escaping RA induced differentiation by impairing RA dependent transcription of differentiation promoting genes.
Preface

Investigation of $KMT2D$ loss of function mutations using a zinc finger nuclease somatic gene knockout strategy was conceived by Dr. Maria Mendez-Lago and Dr. Marco Marra. The TOPO cloning and sequencing validation of the clonal $KMT2D^{\text{-}}$ cell lines generated in this thesis was performed by Diane Trinh. All other research and experimental work in this thesis was conducted by Ryan Huff. Dr. Alessia Gagliardi, Diane Trinh, Dr. Maria Mendez-Lago, and Dr. Marco Marra all provided advice for the optimization and implementation of the experimental work in this thesis.
**Table of Contents**

Abstract ................................................................................................................................. ii

Preface ................................................................................................................................... iii

Table of Contents .................................................................................................................... iv

List of Tables ......................................................................................................................... vii

List of Figures ....................................................................................................................... viii

List of Abbreviations ........................................................................................................... x

List of Genes ........................................................................................................................ xii

Acknowledgements .............................................................................................................. xv

1 Introduction ....................................................................................................................... 1

1.1 Overview ......................................................................................................................... 1

1.2 **KMT2D** ....................................................................................................................... 1

1.2.1 **KMT2D** Nomenclature .......................................................................................... 2

1.2.2 **KMT2D** Complex .................................................................................................. 2

1.2.3 **KMT2D** and Epigenetics ..................................................................................... 3

1.2.4 **KMT2D** and Nuclear Hormone Receptor Transactivation ..................................... 6

1.2.5 **KMT2D** and Constitutive Gene Expression ............................................................. 6

1.2.6 **KMT2D** Cellular Phenotypes .............................................................................. 7

1.2.7 **KMT2D** and Development .................................................................................. 7

1.3 **KMT2D** and Kabuki Syndrome ................................................................................ 9

1.4 **KMT2D** and Cancer .................................................................................................... 10

1.4.1 **KMT2D** Mutations in Cancer ................................................................................ 10

1.4.2 **KMT2D** Mutations in Lymphoma ........................................................................ 11

1.4.3 **KMT2D** and Tumour Development .................................................................... 13

1.5 Genome Engineering with Zinc Finger Nucleases ...................................................... 13

1.5.1 Systems to Enrich for Zinc Finger Nuclease Modified Cells ................................ 14

1.5.2 Zinc Finger Nuclease Off-Target Effects .................................................................. 15

1.6 Hypothesis and Objectives ............................................................................................ 17
1.6.1 Hypothesis ........................................................................................................................................ 17
1.6.2 Objective 1: To Create a Human KMT2D Loss-of-Function Cell Line Model Using a Targeted Zinc Finger Nuclease Somatic Gene Knockout Strategy ................................. 17
1.6.3 Objective 2: To Measure Bulk H3K4 Methylation Profiles in Human Isogenic Wild Type and Mutant KMT2D Cell Lines Created Using ZFNs ......................................................... 17
1.6.4 Objective 3: Evaluate the mRNA Expression of Select KMT2D Target Genes in Human Isogenic Wild Type and Mutant KMT2D Cell Lines Created Using ZFNs ............................ 17

2 Materials and Methods .......................................................................................................................... 22
  2.1 HEK293 Cell Line and Culture ........................................................................................................ 22
  2.2 KMT2D ZFNs .................................................................................................................................... 22
  2.3 GFP and ZFN Reporter Plasmid ...................................................................................................... 22
  2.4 Transfection Methods ..................................................................................................................... 23
  2.5 Flow Cytometry ............................................................................................................................... 23
  2.6 Immunoblotting ............................................................................................................................... 23
  2.7 Acid Histone Extraction .................................................................................................................. 24
  2.8 Surveyor CEL-I Assay .................................................................................................................... 25
  2.9 RT-qPCR .......................................................................................................................................... 26
  2.10 DNA Sequencing ............................................................................................................................ 26
  2.11 Retinoic Acid Time Course ............................................................................................................ 27
  2.12 Alamar Blue Assays ....................................................................................................................... 27

3 Results .................................................................................................................................................. 30
  3.1 Generation of a KMT2D Knockout Cell Line Model ....................................................................... 30
    3.1.1 Screening of HEK293 Cells Transfected with KMT2D Zinc Finger Nucleases ......................... 30
    3.1.2 Characterization of Zinc Finger Nuclease Induced Mutations .................................................. 32
    3.1.3 Transfection of HEK293 with Zinc Finger Nuclease and Reporter Vectors, and Screening for KMT2D/− Clonal Cell Lines ..................................................................................... 32
    3.1.4 Characterization of Zinc Finger Nuclease Induced Mutations .................................................. 35
  3.2 Characterization of KMT2D Knockout Cell Lines ......................................................................... 37
    3.2.1 Analysis of KMT2D/− Cell Lines Rate of Growth .................................................................... 37
    3.2.2 Analysis of KMT2D Dependent Genes .................................................................................... 37
    3.2.3 Expression of KMT2D’s Paralogue KMT2C is Unaffected by KMT2D Genotype................. 38
3.2.4 Bulk H3K4me1 Levels are Reduced in KMT2D−/− Cell Lines.......................... 38
3.2.5 Induction of Retinoic Acid Response Genes is Reduced in KMT2D−/− Cell Lines... 39
3.2.6 Expression of Transcriptional Co-activator CREBBP Correlates with KMT2D...... 41

4 Discussion .................................................................................................................. 65
4.1 Generation of KMT2D Knockout Cell Line Model................................................. 65
4.2 Characterization of KMT2D Knockout Cell Lines ................................................. 69
4.3 Future Directions .................................................................................................... 74

References ..................................................................................................................... 75
List of Tables

Table 2.1: Sequencing, PCR, and RT-qPCR primers. ................................................................. 29
Table 3.1: ZFN induced mutations in selected cell lines detected using western blotting and CEL-I assays, and verified using DNA sequencing. .................................................. 52
List of Figures

Figure 1.1: Histone post-translational modifications mark functional DNA elements in the genome. ................................................................. 18
Figure 1.2: Model of KMT2D’s role in transcriptional activation. ....................... 19
Figure 1.3: Overview of KMT2D mutations reported in human cancers and Kabuki syndrome ........................................................................... 21
Figure 3.1: Diagram of ZFN induced mutagenesis and workflow for ZFN induced mutation screening. .......................................................... 42
Figure 3.2: Evaluation of transfection methods and the CEL-I assay. .................. 43
Figure 3.3: Assessment of ZFN induced KMT2D mutations in HEK293 cells using the CEL-I assay. ................................................................. 45
Figure 3.4: Characterization of KMT2D mutations in three clonal populations isolated from KMT2D ZFN mRNA transfected HEK293 cells ..................................... 46
Figure 3.5: Workflow for creating KMT2D−/− cell lines using KMT2D ZFNs and a fluorescent reporter vector. ...................................................... 47
Figure 3.6: Generation of control clonal cell lines using flow cytometry. .......... 48
Figure 3.7: Generation of KMT2D null cell lines using KMT2D ZFNs, reporter vector, and flow cytometry. ................................................................. 49
Figure 3.8: Screening for additional KMT2D null cell lines. .............................. 50
Figure 3.9: Confirmation of KMT2D’s mutational status in 19 cell lines selected for further characterization. ......................................................... 51
Figure 3.10: Summary of cell line origin, KMT2D mutational status, and KMT2D protein status in 19 cell lines selected for further characterization. .................. 53
Figure 3.11: Characterization of KMT2D mRNA and protein expression in 19 cell lines used in further characterization experiments. .......................... 55
Figure 3.12: Analysis of the top two predicted KMT2D ZFN off-target sites in the cell lines exposed to KMT2D ZFNs. ......................................................... 56
Figure 3.13: Assessment of cell proliferation rates in KMT2D+/+, KMT2D+/−, and KMT2D−/− cell lines. ................................................................. 57
Figure 3.14: Expression of genes previously reported as KMT2D dependent. ....... 58
Figure 3.15: Expression of KMT2D homologues in KMT2D+/+, KMT2D+-, and KMT2D-- cell lines.

Figure 3.16: H3K4 methylation profile in KMT2D+/+, KMT2D+-, and KMT2D-- cell lines.

Figure 3.17: Induction of RARA2, RARB2, and RARG mRNA expression in response to 9-cis-retinoic acid.

Figure 3.18: Induction of the RARA dependent gene CASP9 in response to 9-cis-retinoic acid.

Figure 3.19: H3K4 methylation profiles in KMT2D+/+ and KMT2D-- cell lines treated with 9-cis-retinoic acid.

Figure 3.20: Expression of transcriptional co-activator CREBBP (CBP) correlates with KMT2D expression.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Activated B-cell</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ASCOM</td>
<td>Activating signal cointegrator-2-containing complex</td>
</tr>
<tr>
<td>cAMP</td>
<td>Cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Chromatin immunoprecipitation-sequencing</td>
</tr>
<tr>
<td>CD20</td>
<td>Cluster of differentiation 20</td>
</tr>
<tr>
<td>C/EBPs</td>
<td>CCAAT/enhancer binding proteins</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>COMPASS</td>
<td>Complex proteins associated with Set1</td>
</tr>
<tr>
<td>CRISPR/CAS9</td>
<td>Clustered regulatory interspaced short palindromic repeat/Cas9 RNA</td>
</tr>
<tr>
<td>D CEL-I</td>
<td>CEL-I digested</td>
</tr>
<tr>
<td>DLBCL</td>
<td>Diffuse large B-cell lymphoma</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DSB</td>
<td>Double stranded break</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen receptor</td>
</tr>
<tr>
<td>FL</td>
<td>Follicular Lymphoma</td>
</tr>
<tr>
<td>FYRC</td>
<td>FY-rich domain C-terminal</td>
</tr>
<tr>
<td>FYRN</td>
<td>FY-rich domain N-terminal</td>
</tr>
<tr>
<td>GCB</td>
<td>Germinal center B-cell</td>
</tr>
<tr>
<td>H3K27</td>
<td>Histone 3 Lysine 27</td>
</tr>
<tr>
<td>H3K4</td>
<td>Histone 3 Lysine 4</td>
</tr>
<tr>
<td>H3K9</td>
<td>Histone 3 Lysine 9</td>
</tr>
<tr>
<td>H4R3</td>
<td>Histone 4 Arginine 3</td>
</tr>
<tr>
<td>HEK293</td>
<td>Human embryonic kidney 293</td>
</tr>
<tr>
<td>HeLa</td>
<td>Henrietta Lacks</td>
</tr>
<tr>
<td>HGNC</td>
<td>HUGO Gene Nomenclature Committee</td>
</tr>
<tr>
<td>INDEL</td>
<td>Insertion and or deletion</td>
</tr>
<tr>
<td>KD</td>
<td>Knockdown</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LOF</td>
<td>Loss of function</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblasts</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>NHL</td>
<td>Non-Hodgkin lymphoma</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PHD</td>
<td>Plant homeo domain</td>
</tr>
<tr>
<td>Pol II</td>
<td>Polymerase II</td>
</tr>
<tr>
<td>PPAR</td>
<td>Peroxisome proliferator-activated receptor</td>
</tr>
<tr>
<td>pUC</td>
<td>Plasmid University of California</td>
</tr>
<tr>
<td>pZFN</td>
<td>Plasmid zinc finger nuclease</td>
</tr>
<tr>
<td>RA</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td>RAR</td>
<td>Retinoic acid receptor</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>Ribonucleic acid-sequencing</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse transcription-quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard Error of the Mean</td>
</tr>
<tr>
<td>SET</td>
<td>Su(var)3-9, enhancer-of-zeste, trithorax</td>
</tr>
<tr>
<td>shRNA</td>
<td>short hairpin ribonucleic acid</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering ribonucleic acid</td>
</tr>
<tr>
<td>SNV</td>
<td>Single nucleotide variants</td>
</tr>
<tr>
<td>TFIID</td>
<td>Transcription factor II D</td>
</tr>
<tr>
<td>TALEN</td>
<td>Transcription activator-like effector nucleases</td>
</tr>
<tr>
<td>UD</td>
<td>Undigested</td>
</tr>
<tr>
<td>ZFN</td>
<td>Zinc finger nuclease</td>
</tr>
<tr>
<td>ΔΔCT</td>
<td>Delta delta cycle threshold</td>
</tr>
</tbody>
</table>
### List of Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALR</td>
<td>Acute leukemia locus-1 related</td>
</tr>
<tr>
<td>ANF</td>
<td>Atrial natriuretic factor</td>
</tr>
<tr>
<td>ASC-2</td>
<td>Activating signal cointegrator 2</td>
</tr>
<tr>
<td>ASH2L</td>
<td>Set1/Ash2 histone methyltransferase complex subunit ASH2</td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma 2</td>
</tr>
<tr>
<td>BTBD10</td>
<td>BTB (POZ) domain containing 10</td>
</tr>
<tr>
<td>BTG1</td>
<td>B-Cell Translocation Gene 1</td>
</tr>
<tr>
<td>CASP9</td>
<td>Caspase-9</td>
</tr>
<tr>
<td>CCR2</td>
<td>C-C chemokine receptor type 2</td>
</tr>
<tr>
<td>CCR5</td>
<td>C-C chemokine receptor type 5</td>
</tr>
<tr>
<td>CD20</td>
<td>Cluster of differentiation 20</td>
</tr>
<tr>
<td>Cebpa</td>
<td>CCAAT/enhancer binding protein (C/EBP), alpha</td>
</tr>
<tr>
<td>CREBBP</td>
<td>cAMP-response element-binding binding protein</td>
</tr>
<tr>
<td>CTSD</td>
<td>Cathepsin D</td>
</tr>
<tr>
<td>CYP26A1</td>
<td>Cytochrome P450 26A1</td>
</tr>
<tr>
<td>CYP26B1</td>
<td>Cytochrome P450 26B1</td>
</tr>
<tr>
<td>CYP26C1</td>
<td>Cytochrome P450 26C1</td>
</tr>
<tr>
<td>DPY30</td>
<td>Dpy-30 homolog</td>
</tr>
<tr>
<td>EBAG9</td>
<td>Estrogen receptor binding site associated, antigen, 9</td>
</tr>
<tr>
<td>EGFP</td>
<td>Enhanced green fluorescent protein</td>
</tr>
<tr>
<td>EP300</td>
<td>E1A binding protein p300</td>
</tr>
<tr>
<td>ESR1</td>
<td>Estrogen receptor 1</td>
</tr>
<tr>
<td>FAS</td>
<td>Fas cell surface death receptor</td>
</tr>
<tr>
<td>Gata4</td>
<td>GATA binding protein 4</td>
</tr>
<tr>
<td>GPR56</td>
<td>G protein-coupled receptor 56</td>
</tr>
<tr>
<td>GREB1</td>
<td>Growth regulation by estrogen in breast cancer 1</td>
</tr>
<tr>
<td>HALR</td>
<td>Homologous to ALL-1 related</td>
</tr>
<tr>
<td>HOXA1-3</td>
<td>Homeobox A1-3</td>
</tr>
<tr>
<td>IGH</td>
<td>Immunoglobulin heavy locus</td>
</tr>
<tr>
<td>JMJD2B</td>
<td>Jumonji domain containing 2B</td>
</tr>
</tbody>
</table>
KDM4B  Lysine (K)-specific demethylase 4B
KDM6A  Lysine (K)-specific demethylase 6A
KMT2A-E Lysine (K)-specific methyltransferase 2A-E
LAMB3  Laminin, beta 3
MAL    mal, T-cell differentiation protein
MLL1-5 Myeloid/lymphoid or mixed-lineage leukemia 1-5
MMP-9  Matrix metallopeptidase 9
MMP-11 Matrix metallopeptidase 11
Myf5-cre Myogenic factor 5-causes recombination
MyoD   Myogenic differentiation
NCOA6  Nuclear receptor coactivator 6
NESTIN Nestin
Nkx2.5 NK2 homeobox 5
PA1    PTIP-associated 1 protein
PGK1   Phosphoglycerate kinase 1
PML    Promyelocytic leukemia
PPARG Peroxisome proliferator-activated receptor gamma
PTIP   Pax2 transactivation domain-interacting protein
RARA2 Retinoic acid receptor, alpha isoform 2
RARB2 Retinoic acid receptor, beta isoform 2
RARG  Retinoic acid receptor, gamma
RBBP5 Retinoblastoma binding protein 5
RFP    Red fluorescent protein
RNA18S RNA, 18S ribosomal 5
RSK4   p90 ribosomal S6 kinase 4
SHARPIN SHANK-associated RH domain interactor
SIX1   SIX homeobox 1
Srf    Serum response factor
TAF3   TATA Box Binding Protein Associated Factor 3
Tbx5   T-box 5
TFF1   Trefoil factor 1
TP53  Tumor protein p53
Trr    Trithorax-related
Trx2   Trithorax homolog 2
UTX    Ubiquitously transcribed tetraticopeptide repeat, X chromosome
WBP7   WW domain-binding protein 7
WDR5   WD repeat-containing protein 5
Acknowledgements

I would like to thank my supervisor Dr. Marco Marra for his patience, support, and mentorship throughout the ups and downs of my graduate studies. Thank you to my committee, Dr. Robert Kay and Dr. Keith Humphries for your guidance. Thank you to the Marra lab for all your support and advice. I would especially like to thank Dr. Alessia Gagliardi, Diane Trinh, Dr. Maria Mendez-Lago, and Dr. Ryan Morin for your intellectual and technical mentorship, which contributed immensely to my research and this thesis.

I would also like to thank the Medical Genetics Department, the Canadian Institutes of Health Research, and John Auston for their financial support, which made the research in this thesis possible.

I would like to thank my family, especially my parents, for their unconditional support and encouragement of my scientific curiosity. Lastly I would like to thank Ashley Hearn for her immense patience and unwavering support throughout the completion of this thesis.
To my parents,
and Ashley
1 Introduction

1.1 Overview
Recently, somatic loss-of-function (LOF) mutations in KMT2D (Lysine (K)-specific methyltransferase 2D) have been identified in many different forms of cancer1-16, as well as in a rare pediatric disorder called Kabuki syndrome17-20. Recent work has identified KMT2D as a methylatransferase that monomethylates the 4th lysine on the tail of histone 3 (H3K4) at active cis-enhancer elements that promote gene transcription. However, the mechanism by which KMT2D LOF mutations promote cancer progression is unknown21-23. KMT2D and KMT2D’s paralogue KMT2C have been reported to be functionally redundant in some contexts21,24-27 and independent in others22,28-31. Therefore, in this thesis I was interested in considering two open questions in the field of KMT2D biology: 1) are KMT2D and KMT2C functionally redundant in maintaining bulk H3K4me1, and 2) how do LOF mutations in KMT2D, similar to those identified in cancer, impact H3K4 methylation and transcription in cells with wild type KMT2C.

To investigate the impact of LOF KMT2D mutations on H3K4 methylation and transcription, model KMT2D−/− cell lines were generated by knocking out KMT2D using zinc finger nuclease (ZFN) technology in the human cell line HEK293A (kidney).

1.2 KMT2D
KMT2D (previously known as ALR, MLL2, and MLL4) encodes a histone-lysine N-methyltransferase that catalyzes the methylation of histone 3 lysine 4 (H3K4me), which is an epigenetic modification involved in transcriptional regulation29,32,33. KMT2D belongs to the H3K4 methyltransferase family KMT2A-E (previously MLL1-5) that has been implicated in regulating a range of genes depending on the cellular and developmental context. These genes include homeobox genes, and nuclear hormone responsive-genes34. KMT2A (MLL1) was originally found through its involvement with chromosome translocations in approximately 10% of acute myeloid leukemias, and along with KMT2B (WBP7, KMT2B, TRX2, and previously MLL2) are human homologs of the Drosophila trx gene. In contrast, KMT2D and KMT2C (also MLL3, HALR) are human homologs of the Drosophila trr gene32,34. KMT2D is a very large protein composed of 5537 amino acids. It is encoded by 54 exons spread over 36kb on chromosome 12q13.12, and is ubiquitously expressed throughout the body35.
1.2.1 \textit{KMT2D Nomenclature}

There is considerable confusion in the literature as to the correct name of \textit{KMT2D}, which lies on chr12q13.12. When \textit{KMT2D} was discovered it was originally named \textit{ALR} (ALL-1 related gene)\textsuperscript{32}. \textit{ALL}-1 was soon after named \textit{MLL} (myeloid/lymphoid or mixed-lineage leukemia), therefore the HUGO Gene Nomenclature Committee (HGNC) officially renamed \textit{ALR} as \textit{MLL2} in 1998, as it was the second homologue to \textit{MLL} discovered. However, confusion in the literature started to arise after a paper published in 1999 by FitzGerald and Diaz described the discovery of another \textit{MLL} homologue on chr19q13.12, which they incorrectly named \textit{MLL2}\textsuperscript{36}. This was quickly followed up by another paper that also incorrectly used the name \textit{MLL2} for the gene on chr19\textsuperscript{37}. For a number of years studies used \textit{MLL2} interchangeably to describe the genes on either chr12 or chr19 until 2002 when the name \textit{MLL4} was first used to describe the gene on chr19\textsuperscript{38}. However, by 2006, the name \textit{MLL4} also began to be used to describe the gene on chr12\textsuperscript{24}, and since 2006, \textit{MLL2} and \textit{MLL4} have been used interchangeably for both genes. This has even led to a paper discussing clinical results and interpreting these results in the context of the incorrect gene\textsuperscript{39}. A new system of nomenclature that focused more on the functionality of the proteins encoded by genes was created in 2007\textsuperscript{40}. This removed the MLL designation and classified the MLL homologues as KMT2A-E (lysine (K)-specific methyltransferase 2A-E). This system of naming was adopted in 2013 by the HGNC, designating the \textit{ALR} gene discovered on chr12 as \textit{KMT2D}. The gene on chr19 is officially named \textit{KMT2B}. However, since 2013 several groups have ignored the change in convention and continued to publish interchangeably under the names \textit{MLL2} and \textit{MLL4} despite the confusion this creates in correctly interpreting new results about the function of \textit{KMT2D}.

1.2.2 \textit{KMT2D Complex}

Functionally, KMT2D is known to act as a member of a huge (approximately 2MDa) chromatin-remodeling complex known as the KMT2D/KMT2C, ASCOM (activating signal cointegrator 2 or ASC-2 complex), or COMPASS (complex proteins associated with Set1) complex\textsuperscript{33,34}. In this thesis, the KMT2D complex will be referred to as the KMT2D/KMT2C complex. Along with KMT2D or KMT2C, this complex has been reliably shown to consist of three core components; WD repeat-containing protein 5 (WDR5, ID:P61964), retinoblastoma-binding protein 5 (RBBP5, ID: Q15291), and set1/ash2 histone methyltransferase complex.
subunit ASH2 (ASH2L, ID: Q9UBL3), as well as the secondary components protein dpy-30 homolog (DPY30, ID: Q9C005), lysine-specific demethylase 6A (KDM6A, also UTX, ID: O15550), PAX-interacting protein 1 (PTIP, ID: Q6ZW49), PTIP-associated protein 1(PA1, ID: Q9BTK6), nuclear receptor coactivator 6 (NCOA6, or ASC-2, ID: Q14686), and alpha/beta tubulins

1.2.3 \textit{KMT2D} and Epigenetics

According to the National Institutes of Health Roadmap Epigenomics Project, epigenetics refers to “the study of changes in the regulation of gene activity and expression that are not dependent on gene sequence”\textsuperscript{43}. A major mechanism of epigenetic regulation is the organization and packaging of DNA into chromatin. The packaging of DNA into chromatin is mediated by nucleosomes, which consist of a core histone octomer wrapped around approximately 147 bp of DNA\textsuperscript{23}. The histone octomer consists of two copies each of histones H2A, H2B, H3, and H4. Each histone consists of a structured core with unstructured tails that are subject to post-translational modifications by enzymes known as epigenetic writers. Combinations of these post-translational modifications form a code that mark lightly packed or active euchromatin, tightly packed or silent heterochromatin, as well as functional DNA elements in the genome such as gene enhancers or promoters (Figure 1.1). Specifically, active \textit{cis}-regulatory enhancer elements that regulate transcription are associated with mono- and dimethylated lysine 4 on H3 (H3K4me1/2), and acetylated lysine 27 on H3 (H3K27ac). Active gene promoters on the other hand are associated with di- and trimethylated lysine 4 on H3 (H3K4me2/3)\textsuperscript{23,44}. Functionally, these marks also act as binding sites for epigenetic readers which open up chromatin for transcriptional machinery by remodeling nucleosomes, as well as for transcriptional machinery components such as TAF3, which is a subunit of the transcription complex protein TFIID\textsuperscript{45}.

\textit{KMT2D} was first demonstrated to have histone methyltransferase activity \textit{in vitro} on N-terminal H3 peptides by Goo and collaborators\textsuperscript{33}. In a subsequent study, the \textit{KMT2D} complex, immunoprecipitated from K562 cell lysates, was observed to methylate unmodified and dimethylated K4 on histone 3\textsuperscript{29}. Furthermore, Issaeva and collaborators demonstrated a reduction in \textit{KMT2D} expression via siRNAs correlated with a reduction in expression of several genes such as \textit{CSPG4} and \textit{ENO3}. Correspondingly, reduced \textit{KMT2D} binding and H3K4me3 was observed at the promoters of these genes. These results were compatible with the notion
that KMT2D was functioning as an epigenetic transcriptional activator through the deposition of H3K4me3 at gene promoters. A similar study, examining KMT2D’s role in the transcriptional activation of genes involved in neuronal development, found an association between KMT2D and H3K4me3 at the promoters of several HOX genes including HOX1-346. This study demonstrated that the targeting of KMT2D to H3 was through the interaction of KMT2D’s second of three PHD clusters. Specifically, pull-down assays using H3 and H4 peptides indicated that KMT2D interacts with H4 with unmodified arginine 3 (H4R3) or asymmetrically dimethylated arginine 3 (H4R3me2a). KMT2D’s second PHD cluster was also demonstrated to be necessary for KMT2D’s methyltransferase activity on H3K4.

Interestingly, KMT2D and KMT2D complex member KDM6A play a cooperative role at promoters of genes that are actively transcribed31. KDM6A (also UTX) is a histone 3 lysine 27 demethylase responsible for removing the repressive H3K27me3 chromatin mark31,42. Kim and collaborators demonstrated that KDM6A cooperates with KMT2D in the trimethylation of H3K4 and removal of H3K27me3 at the promoters of several genes; such as MMP-9, MMP-11, and SIX131. Knockdown of either KMT2D or KDM6A led to a decrease in KMT2D and H3K4me3, as well as a decrease in KDM6A and an increase in H3K27me3 at the promoter of MMP-9, resulting in reduced MMP-9 mRNA expression. Taken together, these results indicate KMT2D and KDM6A play a cooperative role at promoters of genes that are actively transcribed.

Recently Hu and collaborators investigated bulk H3K4 methylation in KMT2D null HCT116 (colorectal carcinoma) cells with a KMT2C null background and observed a reduction not in H3K4me3 as expected from previous studies, but in bulk H3K4me121. Moreover, bulk H3K4me1 was restored when KMT2D expression constructs were reintroduced into the cells. Additionally, KMT2D knockdown in mouse embryonic fibroblasts wild type or mutant for KMT2C resulted in changes in the levels of bulk H3K4me1 only in the KMT2C mutant line. From these results the authors concluded KMT2D and KMT2C have redundant roles in the deposition of H3K4me1. However, a loss of bulk H3K4me1 has also been observed in a KMT2D knockout human medulloblastoma cell line wild type for KMT2C28. Therefore, the redundancy of KMT2D and KMT2C in bulk H3K4me1 remains an open question.

Further evidence of KMT2D’s role in the monomethylation of H3K4 was established with ChIP-sequencing experiments, which found a correlation between genome-wide KMT2D bound loci, and H3K4me1 and H3K4me2 deposition21. Interestingly, the majority of these loci
were not near transcription start sites in the genome, but instead were located at enhancer elements. These loci were also observed to be p300 binding loci and enriched for H3K27ac, which also are both known to mark active enhancers\(^47\). In a subsequent study, similar reductions of bulk H3K4me1 were observed upon knocking out KMT2D in mouse preadipocytes with a KMT2C null background. Through ChIP-seq experiments, Lee and collaborators identified a correlation between active expression of genes whose enhancers are marked by Pol II, H3K27ac, H3K4me1, and H3K4me2 and bound by KMT2D\(^22\). Furthermore, the Drosophila homologue of KMT2C and KMT2D, \(trr\), has also been demonstrated to be responsible for bulk H3K4me1 and associated with active enhancers\(^48,49\). Altogether, these studies demonstrate KMT2D predominantly functions as a transcriptional coactivator by monomethylating H3K4 at active enhancer loci (Figure 1.2).

The finding that KMT2D is a major H3K4 monomethylator at enhancer regions may also explain the association of H3K4me3 and KMT2D at gene promoters. A transcriptional activation complex including KMT2D at distal enhancers could be looped into proximity of the promoter, which is then marked by H3K4me3 in a KMT2D independent manner\(^50\). However, due to the interdependent nature of active transcription and H3K4me3 at gene promoters, and the observation of H3K4 trimethylation activity \textit{in vitro}\(^29\), it has not been conclusively ruled out that KMT2D is capable of trimethylating H3K4 once brought into proximity of the promoter. Although no genome wide reduction in H3K4me3 at transcription start sites was observed in KMT2D null HCT116 cells when profiled by ChIP-seq; SET1A, SET1B, KMT2A, and KMT2B have all been previously shown to have H3K4 methylase activity and may compensate for the loss of KMT2C and KMT2D with regards to H3K4me3\(^34\). Interestingly, KMT2D and KMT2C promoter binding has also been observed to be associated with promoter H3K4 monomethylation and transcriptional repression for a small subset of genes (19) involved in myogenesis\(^51\). Therefore these data indicate in some circumstances KMT2D may act as a transcriptional repressor. Although, in addition to the 19 genes KMT2D was implicated in repressing, KMT2D was found to be involved in maintaining the expression of 260 other genes. There are no other examples of KMT2D acting as a transcriptional repressor in the literature.
1.2.4 \textit{KMT2D} and Nuclear Hormone Receptor Transactivation

There have been several biochemical studies that have identified a number of gene networks under KMT2D control. The KMT2D/KMT2C complex has been shown to be involved in nuclear hormone receptor mediated transcriptional activation via NCOA6\textsuperscript{33}. Specifically, the KMT2D/KMT2C complex has been shown to bind the retinoic acid receptor (RAR)\textsuperscript{24,52}, liver X receptor (LXL)\textsuperscript{26}, peroxisome proliferator-activated receptor gamma (PPAR\textgamma)\textsuperscript{22,53}, and estrogen receptor (ER)\textsuperscript{30,54,55}. These studies illustrate the diversity and cell context dependent nature of KMT2D’s biological functions within nuclear hormone receptor mediated transcriptional activation. However, during my thesis work I became interested in the role of KMT2D in RAR transactivation specifically because RA plays a role in regulating the proliferation and differentiation of germinal centre B-cells\textsuperscript{56,57}, which are the cells of origin for follicular lymphoma\textsuperscript{58}. Interestingly, RA has also been demonstrated to induce growth inhibition and differentiation of cancer cells, in particular in the cases of myeloid and lymphoid cancer cells\textsuperscript{57,59-61}. Furthermore, mice have been shown to develop lymphoma when \textit{Rara} is knocked down or knocked out\textsuperscript{62,63}. These results show the importance of RA signaling in B-cell development and differentiation, as well as in lymphoma tumour suppression.

KMT2D and KMT2C were found to be involved in RAR transactivation in experiments that used siRNAs against KMT2D and/or KMT2C and measured the induction of \textit{RAR-\beta2} expression as well as examining the abundance of H3K4me3 at the \textit{RAR-\beta2} promoter\textsuperscript{24}. This study revealed that induction of \textit{RAR-\beta2} in response to RA was impaired only when KMT2D and KMT2C were knocked down together, suggesting redundant roles in mediating RAR transactivation. Another study examined RAR transactivation in a human cell line with homozygous inactivating mutations in KMT2C that was mutant or wildtype for KMT2D. RA induction of \textit{ASB2} was impaired in the mutant KMT2D cell line compared to the wild type KMT2D cell line\textsuperscript{52}. However, the role of KMT2D in RAR transactivation has never been examined in isogenic cell lines mutant or wild type for KMT2D in addition to being wild type for KMT2C.

1.2.5 \textit{KMT2D} and Constitutive Gene Expression

Two studies have used genomic high-throughput approaches to comprehensively identify KMT2D dependent genes in human cells. One of these studies used microarrays to assess the
differential expression of genes in HeLa (cervical carcinoma) control and HeLa KMT2D shRNA KD cells. Issaeva and collaborators found KMT2D regulated the expression of genes involved in cytoskeleton organization, cell adhesion and transcriptional regulation. Over concerns of functional redundancy between KMT2D and KMT2C, KMT2D dependent genes were also examined in KMT2D KO HCT116 cells with a homozygous null KMT2C background. It was found that KMT2D regulated genes were enriched in a number of pathways that include: cAMP signaling (p = 0.0023), nuclear hormone receptor signaling (p = 0.0024 for RXR, p = 0.0062 for FXR/RAR), and interestingly B-cell development (p = 0.01). Notably, the significant enrichment in genes involved in nuclear hormone receptor signaling differentially expressed in the HCT116 cell line without nuclear hormone stimulating conditions further supports KMT2D’s role in nuclear hormone transactivation. However, there is little overlap between the KMT2D dependent genes identified in these studies (1.56%; 4/256), perhaps due to differences in the biology of the cell lines.

1.2.6 KMT2D Cellular Phenotypes

At the cellular level HEK293, MCF-7 (breast adenocarcinoma), and HeLa cell lines were shown to have reduced growth rates when KMT2D was knocked down (KD) with siRNAs. This is thought to be due to increased levels of apoptosis. This slow growth phenotype has been further confirmed in knock out HCT116 and DLD (medulloblastoma) cell lines. In addition, knock down of KMT2D in HeLa cells also resulted in reduced migration, and correspondingly a distinctive actin cytoskeleton organization, with a more mesh-like network of cortical actin rather than definitive focal adhesion stress fibers.

1.2.7 KMT2D and Development

In mice, KMT2D is not expressed during the initial stages of embryogenesis and begins to be expressed in the exterior of the ectoderm and in the first brachial arch of the central mesenchyme around E9.0 during the differentiation and formation of more complex structures. In agreement with the timing of KMT2D expression during development, a knockout KMT2D mouse model resulted in embryonic lethality around E9.5. Molecularly, KMT2D was also demonstrated to play a role in neuronal differentiation. Loss of KMT2D via shRNAs in the pluripotent human embryonal carcinoma cell line NT2/D1 impaired retinoic acid mediated
neuronal differentiation of the cells. Specifically, KMT2D knockdown reduced the expression of HOXA1, HOXA2, HOXA3, and NESTIN, which are regulatory genes involved in neuronal differentiation\(^46\).

KMT2D has also been linked to defects in adipogenesis and myogenesis. In one study, a conditional KMT2D knockout mouse was generated using a cre-loxP system\(^22\). The deletion of KMT2D in somatic precursor cells was induced through the expression of Myf5-Cre. This caused a considerable decrease in brown adipose tissue and skeletal back muscle at E18.5, which led to rib cage muscle defects, breathing difficulties, and consequently death of the mice shortly after birth. This defect in adipogenesis and myogenesis was not observed in KMT2C knockout mice. However, when examining the \textit{in vitro} differentiation of preadipocytes, reduced induction of the master regulatory gene \textit{Cebpa} caused by the loss of KMT2D was exacerbated by the loss of KMT2C. The mechanism of the differentiation block caused by the loss of KMT2D was identified as a loss of H3K4me1 at enhancers, which was associated with reduced transcriptional co-activation of genes involved in adipogenesis and myogenesis\(^22\). KMT2D was observed by ChIP-seq experiments to bind C/EBPs and PPAR\(\gamma\) bound active adipogenic enhancer elements marked by H3K4me1 during the differentiation of brown preadipocytes to adipocytes. KMT2D was then shown to be required for the induction of genes such as \textit{Pparg} and 588 other genes during adipogenesis. KMT2D was also observed to bind active enhancers marked by H3K4me1 and bound by MyoD during myogenesis. KMT2D was then shown to be required for the induction of genes involved in coordinating myogenesis such as \textit{Myog}, as well as 836 other genes.

Conversely, KMT2D and KMT2C promoter binding has also been observed to be associated with promoter H3K4 monomethylation and transcriptional repression for a subset of genes (19) involved in myogenesis. Depletion of KMT2D and KMT2C by siRNAs accelerated the differentiation of the mouse C2C12 myoblast cell line, and correspondingly the myogenesis coordinator \textit{Myog} and other muscle related genes were upregulated\(^51\). However, the transcriptional repression of \textit{Myog} by KMT2D is in conflict with the KMT2D conditional knockout mouse myogenesis model described above\(^22\). Interestingly, KDM6A has been shown to be involved in the differentiation of cardiac lineages in developing mice. Specifically, KDM6A was observed to be associated with the active \textit{ANF} enhancer which regulates four key transcription factors, \textit{Nkx2.5}, \textit{Tbx5}, \textit{Gata4}, and \textit{Srf}, which are involved in the differentiation of
cardiac lineages\textsuperscript{64}. These studies suggest KDM6A, KMT2D, and KMT2C play an important role in marking active enhancers and driving myogenesis, and a questionable role for KMT2D and KMT2C in transcriptional silencing.

1.3 \textit{KMT2D} and Kabuki Syndrome

Heterozygous loss of function \textit{KMT2D} mutations have also been found to be the predominant cause of a rare pediatric disorder named Kabuki syndrome (Figure 1.3), with the majority of mutations arising \textit{de novo} and resulting in haploinsufficiency\textsuperscript{17-20,39}. Kabuki syndrome is characterized by mild to moderate intellectual disability, as well as cardiac and skeletal abnormalities\textsuperscript{20}. A study that examined \textit{KMT2D} mutation status in 110 cases of Kabuki syndrome found that 74\% of cases had a \textit{KMT2D} mutation\textsuperscript{39}. In total, 81 mutations were found amongst these cases, of which 3 were splice site, 37 were nonsense, 22 were frameshift, 3 were in-frame indels, and 16 were missense\textsuperscript{39}. Furthermore, the missense mutations tended to disrupt the enzymatic SET domain, or the histone binding PHD (plant homeo domain). A follow up study investigated the functional impact of \textit{KMT2D} mutations by creating transformed human lymphoblastoid and fibroblast cell lines from individuals with Kabuki syndrome, and examining the mRNA expression of several known \textit{KMT2D} target genes. The mRNA expression of \textit{HOXC6} was reduced in all cell lines established from Kabuki syndrome patients\textsuperscript{65}. Interestingly, 5\% of Kabuki syndrome cases are caused by mutations in the gene \textit{KDM6A}\textsuperscript{66}, which is a histone 3 lysine 27 demethylase known to complex and act cooperatively with \textit{KMT2D}\textsuperscript{31}. These results indicate Kabuki is caused by inactivation of the histone modification activity of the \textit{KMT2D} complex, resulting in the silencing of developmentally important genes such as \textit{HOXC6}\textsuperscript{65}.

An interesting question is if \textit{KMT2D} mutations predispose an individual with Kabuki syndrome to cancer. However, there is only limited evidence in the form of case reports. Cancer has been reported in seven individuals with Kabuki syndrome: two cases of neuroblastoma, one case of hepatoblastoma, one case of fibromyxoid sarcoma, one case of acute lymphocytic leukemia, one case of Burkitt’s lymphoma, and one case of synovial sarcoma\textsuperscript{19}. The mutational status of \textit{KMT2D} is unknown in these patients, although 68\% (215/351) of individuals with Kabuki syndrome have been found to have \textit{KMT2D} mutations\textsuperscript{18,20,39,67,68}. Only one of the seven patients, with neuroblastoma, died from cancer. Therefore it is still questionable an individual
with Kabuki syndrome is predisposed to cancer, and consequently if heterozygous \( KMT2D \) mutations predispose an individual to developing cancer.

### 1.4 \( KMT2D \) and Cancer

#### 1.4.1 \( KMT2D \) Mutations in Cancer

\( KMT2D \) has attracted a lot of attention recently as an increasing number of studies characterizing tumours through whole genome, exome, or transcriptome sequencing discovered \( KMT2D \) as a recurrent target of somatic mutation (Figure 1.3). Analysis of all the \( KMT2D \) mutations reported in the literature revealed the majority of \( KMT2D \) mutations are predicted to result in the loss of protein function (LOF): 54% of the mutations are nonsense, frameshift insertion/deletions, or splice site mutations. In particular, \( KMT2D \) was discovered to be the most frequent target of somatic mutation in patients with follicular lymphoma (FL) (84%; 113/135 cases), with approximately 50% of cases harboring multiple mutations in \( KMT2D \)\(^1\). In addition, \( KMT2D \) mutations and the t(14:18)(q32;q21) (85% of FL patients) translocation are the most frequent genomic alterations of FL. \( KMT2D \) was also found to be recurrently mutated in patients with diffuse large B-cell lymphoma (DLBCL) (41%; 22/54 of cases)\(^1\). These mutations are discussed in the next section. In 27% (35/130) of patients with bladder urothelial carcinoma, 39 \( KMT2D \) mutations were identified, 23 of which were LOF mutations\(^2\). In 20% (59/289) patients with stomach adenocarcinoma, 83 mutations were discovered, of which 42 were LOF mutations\(^3\). In 20% (36/178) of squamous cell lung cancers, of which 19 were LOF mutations\(^4\). A study that examined head and neck squamous cell carcinoma discovered 8 \( KMT2D \) mutations among 11% (8/74) of patients, 5 of which were LOF mutations\(^5\). Furthermore, in 18% (49/279) of an expanded data set of patients with head and neck squamous cell carcinoma, 54 \( KMT2D \) mutations were discovered, of which 37 were LOF mutations\(^6\). In uterine corpus endometrioid carcinoma, 13% (33/248) patients were found to have 57 \( KMT2D \) mutations, which were predominately missense\(^7\). In medulloblastoma, 9% (8/92) of cases were found to have mutations in \( KMT2D \), with 6 out of 8 being LOF mutations\(^8\). \( KMT2D \) mutations have also been identified in a small proportion of patients with small cell lung cancer (7%; 3/42)\(^9\), colorectal adenocarcinoma (5%; 12/224)\(^10\), nasopharyngeal carcinoma (5%; 3/52)\(^11\), esophageal squamous cell carcinoma (5%; 4/88)\(^12\), kidney renal clear cell carcinoma (3%; 13/424)\(^13\), esophageal adenocarcinoma (2%; 3/146)\(^14\), multiple myeloma (2%; 4/205)\(^15\),
glioblastoma (1%; 3/291)\textsuperscript{14}, ovarian serous cystadenocarcinoma (1%; 2/316)\textsuperscript{15}, and acute myeloid leukemia (1%; 1/200)\textsuperscript{76}. Additionally, querying The Cancer Genome Atlas research network (TCGA) database also revealed \textit{KMT2D} mutations in cervical squamous cell carcinoma (15%; 6/39), melanoma (14%; 40/278), kidney renal papillary cell carcinoma (11%; 18/168), lung adenocarcinoma (8%; 19/230), liver hepatocellular carcinoma (5%; 12/231), adrenocortical carcinoma (4%; 4/91), prostate adenocarcinoma (4%; 9/236), kidney chromophobe (3% 2/66), lower grade glioma (2%; 5/289), and breast invasive carcinoma (2%; 23/976) (Figure 1.3).

Comparing the mutational profile of \textit{KMT2D} between these cancers revealed two patterns of mutation characterized by: (1) missense mutations scattered across \textit{KMT2D} as seen in uterine corpus endometrioid carcinoma and melanoma, or (2) frame-shift or nonsense mutations as seen in FL, DLBCL, bladder urothelial carcinoma, stomach adenocarcinoma, and head and neck squamous cell carcinoma (Figure 1.3). Although the second mutational pattern is highly suggestive that \textit{KMT2D} is under selection for acquiring LOF mutations, it is not readily apparent how the first pattern of missense mutations are affecting KMT2D’s function as the missense mutations do not converge on any particular functional domain. Interestingly, \textit{KMT2D} mutations are predominately found in carcinomas 68% (19/28 cancers with reported \textit{KMT2D} mutations), which are epithelial in origin. Altogether, the mutational profile indicates \textit{KMT2D} is a tumour suppressor in at least five different types of cancer, and has an important yet unclear role in epithelial cancers.

1.4.2 \textit{KMT2D} Mutations in Lymphoma

Lymphomas are cancers of the blood, and the most common lymphomas are classified as non-Hodgkin lymphomas (NHL). NHLs are a genetically heterogeneous group of cancers originating from B, T, or natural killer lymphocytes. The two most common forms of NHL are follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL), which account for 25% and 32% of newly diagnosed lymphomas every year, respectively\textsuperscript{58}. FL is the malignant proliferation of germinal center B-cells. It is an indolent disease with an overall survival of 73% over 10 years\textsuperscript{58}. Even though the disease responds well to current chemotherapies, it is incurable and over time patients develop treatment resistant disease. Additionally, approximately 45% of cases transform into a more aggressive form of disease resembling DLBCL\textsuperscript{77}. FL is defined by the t(14;18)(q32;21) translocation that places the anti-apoptotic \textit{BCL2} gene under control of the
IGH locus in approximately 80-90% of cases, resulting in BCL2 constitutive expression. However, the t(14:18) translocation is not necessary or sufficient to cause FL as it has been found in circulating B-cells of 50-70% of healthy individuals\textsuperscript{78,79}. DLBCL is a more aggressive form of lymphoma, and based on gene expression is classified into the prognostic groups; germinal centre B-cell (GCB) like with a 74% 3 year progression free survival, and activated B-cell (ABC) like DLBCL with a 40% progression free survival\textsuperscript{80,81}. DLBCLs have an average 5 year progression free survival of approximately 50\%\textsuperscript{58}.

Recent studies have characterized the mutational landscape of FL and DLBCL. A study by Morin and collaborators reported the sequence analysis of the genomes of 13 DLBCL tumours and integrated these data with whole transcriptome shotgun sequencing (RNA-seq) data from an additional 113 NHLs\textsuperscript{1}. Analysis of these data revealed somatic mutations effecting 109 genes in two or more cases, with KMT2D identified as the most statistically significant gene under selective pressure for acquiring missense or nonsense mutations\textsuperscript{1}. KMT2D was found to be the most frequently mutated gene in FL patients (89% 31/35), and one of the most frequently mutated genes in DLBCL patients (32% 12/37) and DLBCL cell lines (59% 10/17).

Interestingly, the majority (91%) of the mutations discovered in NHL patients were predicted to result in protein LOF, specifically 37% (29/78) nonsense mutations, 46% (36/78) frame shift insertion or deletions, and 8% (6/78) splice site mutations\textsuperscript{1} (Figure 1.3). Additionally, 16 out of the 35 FL patients had two KMT2D mutations, of which 8 FL cases were assessed to determine if the mutations were on separate alleles (\textit{trans}). All 8 cases were found to have KMT2D mutations \textit{in trans}, one mutation per allele. Given the large number of mutated genes in FL and DLBCL, Morin and collaborators identified genes mutated in the same patient. KMT2D mutations were most often found along with somatic mutations in BCL2, TP53, CREBBP, EP300, or BTG1\textsuperscript{1}.

In agreement with Morin and collaborators, a follow up study by Okosun and collaborators examining a larger cohort of FL patients found 82% (82/100) of cases had at least one mutation in KMT2D, predominantly predicted to result in protein LOF\textsuperscript{4}. In addition, 47% (47/100) FL cases had at least two independent KMT2D mutations\textsuperscript{4}. Okosun and collaborators found KMT2D was most frequently co-mutated with CREBBP in FL cases (70%; 57/82). Altogether, the mutational profiles of KMT2D uncovered in these studies are highly suggestive that KMT2D is an important tumour suppressor in NHL.
1.4.3 *KMT2D* and Tumour Development

Following the discovery of frequent inactivating *KMT2D* mutations in FL, several studies sought to determine if *KMT2D* mutations were early or late drivers of lymphomagenesis. Green and collaborators examined intratumoural heterogeneity of FL tumours using exome sequencing\(^\text{82}\). They postulated that somatic mutations found at high allele frequencies likely arose early in tumour development. Exome sequencing revealed 5 *KMT2D* mutations in 4 out of 8 patients in the study. Analysis of the *KMT2D* mutation frequencies suggested *KMT2D* mutations occurred later in tumour development and were likely driver mutations\(^\text{82}\). However, this study did not investigate frameshift insertion-deletions, which according to Morin and collaborators make up 46% of *KMT2D* mutations in NHL\(^1,\text{82}\). Therefore, it is likely not all *KMT2D* mutations present in the tumour samples were identified. A study by Okosun and collaborators also examined the evolution of FL in 10 patients by using the frequencies of nonsynonomous mutations to construct a phylogenetic tree depicting the clonal evolution of each patient’s cancer, across multiple diagnostic and relapse biopsies\(^4\). Taking into account frameshift mutations, Okosun et al. observed *KMT2D* mutations occurred early in FL, and were retained following transformation of FL into more aggressive DLBCL. Interestingly *CREBBP* mutations, which frequently occur along with *KMT2D* mutations in FL patients, were also found to occur early in lymphomagenesis\(^4,\text{82}\). The results of this study indicate that *KMT2D* is a key early driver of lymphomagenesis. However, how *KMT2D* mutations drive lymphomagenesis is unknown.

1.5 Genome Engineering with Zinc Finger Nucleases

In this thesis, zinc finger nucleases were used to generate somatic *KMT2D* cell lines. Over the last two decades ZFNs have emerged as a versatile tool to create precise modifications in the genome\(^\text{83}\). A ZFN is a hybrid enzyme that combines a DNA-binding domain consisting of an array of Cys\(_2\)His\(_2\) zinc fingers and the nonspecific c-terminal DNA cleavage domain of the type II\(_\text{s}\) restriction endonuclease *FokI*\(^\text{84,85}\). Individual zinc finger domains recognize 3-4 contiguous base pairs, binding in the major groove of DNA. The assembly of three to six individual zinc finger domains, which can vary in sequence specificity, results in a ZFN monomer capable of recognizing 9-18bp of DNA. Following the recognition of the two ZFN monomer’s respective target DNA sequences, the ZFNs form an active endonuclease by
heterodimerizing, resulting in the creation of a double stranded break in the DNA at the targeted locus.

The double stranded breaks created by the ZFNs can be used for genome editing in different ways such as addition, alteration, or disruption of genes. The addition or alteration of DNA sequences is achieved through the use of a donor repair plasmid that contains the gene or sequence that is to be introduced flanked by homology arms consisting of DNA sequences identical to the 5’ and 3’ DNA sequences of the targeted locus. The new gene or sequence is then introduced by homologous repair following the double stranded cleavage of the DNA by the ZFNs. Gene disruption is achieved when short insertion and deletions are generated at the ZFN cleavage site through incorrect repair of the double stranded DNA break via Non-Homologous End Joining (NHEJ).

1.5.1 Systems to Enrich for Zinc Finger Nuclease Modified Cells

ZFN efficiency is variable depending on a variety of factors that include cell line, delivery method (i.e. plasmid, mRNA, or protein), expression levels of the ZFNs in the cell, and the accessibility of the chromatin surrounding the target locus. Several different methods have been demonstrated to increase the efficiency of ZFN mediated gene disruption. The first technique that was shown to increase ZFN mediated gene disruption was a simple incubation of the cells in slightly hypothermic conditions following the delivery of ZFNs. This was shown to increase ZFN activity by two to tenfold, and thought to be due to increased steady-state levels of ZFN protein.

One system that has been employed uses a fluorescent reporter plasmid that enables efficient identification and assessment of ZFN activity within a cell. The reporter vector contains a copy of the red fluorescent protein (RFP) gene that is constitutively expressed, along with the ZFN target sequence followed by an out of frame enhanced green fluorescent protein (EGFP) gene that is not expressed. If ZFNs are active within a cell that received a reporter vector they will cleave the vector at the ZFN target sequence, which is susceptible to frameshift mutations introduced through incorrect NHEJ, leading to a 1 in 3 chance of putting EGFP back in frame. Thus, the reporter vector in conjunction with flow cytometry can be used to select and isolate cells with active ZFNs, leading to enrichments in cells with targeted mutations. This
technique was demonstrated to enrich cells with targeted mutations by 38-fold using ZFNs against the human gene CCR5 in HEK293 cells\cite{89}.

A recent system that also employs the use of fluorescent markers in conjunction with flow cytometry has also been developed, where the fluorescent proteins are physically tethered to the ZFN monomers\cite{90}. Enrichments of cells harboring targeted mutations in the gene RSK4 were observed using this technique, depending on the fluorescence intensity in the cells selected for analysis by flow cytometry. A total of 11.8\% of cells with low levels of the fluorescently labeled ZFNs were identified to harbor at least one targeted mutation in RSK4 in K562 cells, compared to 100\% of cells expressing high levels of fluorescently labeled ZFNs\cite{90}.

### 1.5.2 Zinc Finger Nuclease Off-Target Effects

ZFN cytotoxicity is the result of off-target cleavage of the DNA\cite{83}, which can result in the creation of small insertion and or deletions at undesired locations in the genome. Depending on where these mutations occur and their impact on regulatory regions or the coding genome, they can potentially confound the interpretation of a given set of results. Newer generations of ZFN designs, such as the obligate heterodimer design, have aided in reducing off-target cleavage by increasing the length of sequence specificity to ensure the ZFNs target a unique site in the genome, as well as preventing homodimerization of the individual ZFN monomers\cite{91}. Specifically, homodimerization is prevented through an engineered electrostatic interaction between the FokI obligate heterodimer. The introduction of positive or negative amino acids into the FokI interface domain of the ZFN monomers act to repulse like ZFN monomers and attract the oppositely charged different ZFN monomers\cite{92}.

Although new ZFN designs have resulted in a decrease of these side effects, off target cleavage has been documented using a pair of ZFNs that targeted CCR5 with and without the newer obligate heterodimer design\cite{91}. This was examined using integrase-defective lentiviral vectors (IDLV) with homology to the CCR5 ZFN binding locus. Off-target DSBs created by the ZFNs captured and integrated the linearized IDLV, effectively marking the off-target effects of the CCR5 ZFNs. In the case of the obligate heterodimer ZFNs targeting CCR5, four off-target loci were identified, none of which were homodimeric ZFN sites, and all of which had greater than 66.7\% sequence homology to the ZFN target site. These sites were found to have mutated allele frequencies of 0.6-5.8\% when measured using deep sequencing, compared to a mutated
allele frequency of 40% for \textit{CCR5}. The most frequent off target site was in \textit{CCR5}'s homologue \textit{CCR2} (5.8% mutation frequency), with 91.7% sequence identity to the \textit{CCR5} target. Compared to \textit{CCR5}, \textit{CCR2} was targeted tenfold less then \textit{CCR5} (40% mutation frequency). This was confirmed by another study that screened the off-target loci of \textit{CCR5} ZFNs using an \textit{in vitro} synthetic library of potential off-target sequences. The resulting sequences were then used to predict off target genomic locations, which were validated using targeted high-throughput sequencing of K562 cells transfected with \textit{CCR5} obligate heterodimer ZFNs. On target \textit{CCR5} was found to be the most frequently modified (43.5% mutated allele frequency), and similarly the close homologue of \textit{CCR5}, \textit{CCR2}, was most frequently modified off target (10% mutated allele frequency). In this study the only other off-target with the potential to impact a gene was in the promoter of \textit{BTBD10} (0.1% mutation frequency), with 87% binding site homology\textsuperscript{93}. However, the off-targets identified in this study were created using ZFNs with a combined 24bp DNA recognition domain, and combined 36bp DNA recognition domains are now becoming more common. It is not known if this increase in sequence specificity has an effect in limiting off-target cleavage. Altogether, these studies indicate most off-target induced mutations occur at a mutation allele frequency less then 1% in a cell population treated with ZFNs, unless there is greater then 70% homology between the off-target and target locus.
1.6 Hypothesis and Objectives

1.6.1 Hypothesis
Loss of function KMT2D mutations result in reduced bulk levels of histone 3 lysine 4 monomethylation independent of KMT2C genotype, and altered expression of target genes.

1.6.2 Objective 1: To Create a Human KMT2D Loss-of-Function Cell Line Model Using a Targeted Zinc Finger Nuclease Somatic Gene Knockout Strategy
In this thesis, I used a surrogate reporter based strategy in order to create a process that efficiently enriches and screens for ZFN mutated HEK293 clonal cell lines.

1.6.3 Objective 2: To Measure Bulk H3K4 Methylation Profiles in Human Isogenic Wild Type and Mutant KMT2D Cell Lines Created Using ZFNs
Given that almost all of the understanding of KMT2D’s role in the methylation of H3K4 has come from a cell line model lacking KMT2C, I examined the bulk H3K4 mono-, di-, and trimethylation profile in wild type and mutant KMT2D cell lines wild type for KMT2C.

1.6.4 Objective 3: Evaluate the mRNA Expression of Select KMT2D Target Genes in Human Isogenic Wild Type and Mutant KMT2D Cell Lines Created Using ZFNs
I also investigated KMT2D’s role in nuclear hormone transactivation, which has never been examined in wild type and mutant KMT2D cell lines wild type for KMT2C. Specifically, I examined retinoic acid induced expression of several known retinoic acid response genes in wild type and mutant KMT2D cell lines.
Figure 1.1: Histone post-translational modifications mark functional DNA elements in the genome.
A schematic of histone marks associated with transcriptionally active or silent enhancers as well as promoters and gene bodies. Active enhancers are commonly marked by H3K4me1/2 and H3K27ac. Active promoters are enriched for H3K4me2/3 and H3K9ac, and transcribed regions are associated with H3K36me3 and H3K79me2. Inactive enhancers are enriched for H3K9me2/3 and H3K27me3. Transcriptionally silent genes are marked by H3K9me2/3 and H3K27me3 at promoters and H3K27me3 across the gene body. The coloured balls on histone tails represent modifications corresponding to the labels on the diagram. (RNAPII: RNA polymerase II). Adapted with permission from Macmillian Publishers Ltd: Nature Reviews Genetics (Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12, 7–18 (2011)).
Figure 1.2: Model of KMT2D’s role in transcriptional activation.
(A) In response to environmental or developmental signals, transcription factors and epigenetic modifiers are recruited to active enhancer regions. (CBP: CREB-BP, TF1/2: Transcription factor 1/2, ELL3: Elongation Factor RNA Polymerase II-Like 3, Pol II: RNA polymerase II) (B) The active enhancer recruits transcriptional co-activators to the promoter in order to initiate transcription of the regulated gene. Adapted from Molecular Cell 53, Herz, H.M., Hu, D. & Shilatifard, A. Enhancer malfunction in cancer. 859–866 (2014), with permission from Elsevier.
Figure 1.3: Overview of KMT2D mutations reported in human cancers and Kabuki syndrome.  
Each row characterizes the mutational profile of KMT2D in a specific malignancy, with the fraction of patients harboring any KMT2D mutation within each malignancy indicated on the right. The genetic alterations are mapped onto the genomic locus of KMT2D with the exonic structure and corresponding functional protein domains highlighted by colored boxes indicated above. Cancers were sorted in descending order based on the frequency of patients with KMT2D mutations.
2 Materials and Methods

2.1 HEK293 Cell Line and Culture

The human embryonic kidney cell line HEK293A (provided by Dr. Gregg Morin, authenticated by Genetica Laboratories March 16, 2014) was used for ZFN mediated gene disruption. HEK293 cells were cultured in Dulbecco's modified Eagle's medium (Gibco) containing 10% (v/v) fetal bovine serum (Gibco), penicillin (100 units/ml), and streptomycin (100 µg/ml) at 37 degrees Celsius in 5% CO₂ and a humidified atmosphere.

2.2 KMT2D ZFNs

CompoZr™ Knockout Zinc Finger Nuclease (Cat #: CKOZFND14397-1KT, Sigma-Aldrich) targeting KMT2D exon 39 (Chr12:49427414-49427455, NCBI build hg19) were used to knock out KMT2D. Each individual ZFN recognizes an 18 bp DNA sequence on either side of a 6bp spacer sequence where the heterodimerized FokI domains cleave the DNA. In total the ZFN recognition and cleavage site (underlined) consists of 42 bases GGGATCCAGCCCCACCAGAATGTTGCTGTTGCTGTTGGG. Individual ZFNs were encoded in separate plasmids of 4.12Kb. Each pZFN consists of a CMV promoter driving the expression of the ZFN gene, a kanamycin resistance gene, and a pUC origin of replication. Isolated KMT2D ZFN mRNA (2µg each per aliquot) was also included in the ZFN kit.

2.3 GFP and ZFN Reporter Plasmid

A pCMV-GFP plasmid (provided by Dr. Gregg Morin) that robustly expresses GFP was used as a control to test transfection efficacy. A Hygro RS_MLL2 (ToolGen) ZFN reporter plasmid was used to assess ZFN efficiency. The reporter plasmid (6.4Kb) consists of a CMV promoter driving the expression of mRFP followed by the recognition and cleavage sites of the KMT2D ZFNs and an out of frame HYGRO-EGFP gene, a kanamycin resistance gene, and a pUC origin of replication. Vector cleavage by KMT2D ZFNs and subsequent introduction of small insertions or deletions during repair provide a one in three chance of putting HYGRO-EGFP in frame. The expression of EGFP was used as a marker for active KMT2D ZFNs in a cell.
2.4 Transfection Methods

HEK293 cells were seeded at a density of \( \approx 80,000 \) cells/cm\(^2\) (800,000 cells/well) in a 6 well vacuum-gas plasma tissue-culture treated polystyrene culture plate (Falcon) twenty-four hours prior to transfection. Transfections were carried out using a Trans-IT mRNA transfection kit (Mirus) following the manufacturer’s instructions. The Trans-IT mRNA reagents were supplemented with either: GFP control plasmid (2.5\( \mu \)g), KMT2D ZFN 1 and 2 mRNA (2\( \mu \)g each), KMT2D reporter vector (2.5\( \mu \)g), KMT2D ZFN plasmid 1 and 2 (2.5\( \mu \)g each) and KMT2D reporter vector (2.5\( \mu \)g), or Trans-IT mRNA reagents alone. Each treatment was applied to two separate tissue culture wells. The cells were incubated with the treatments overnight at 37 degrees Celsius in 5% CO\(_2\). Twenty-four hours post transfection the culture media was replaced with fresh Dulbecco's modified Eagle's medium (Gibco) containing 10% fetal bovine serum (Gibco), penicillin (100 units/ml), and streptomycin (100 \( \mu \)g/ml), and the cells were incubated in mild hypothermic conditions (30 degrees Celsius) for 48 hours. Qualitatively, the effectiveness of the transfection measured by RFP\(^+\) cells was then assessed by fluorescence microscopy (Axiovert 200, Zeiss).

2.5 Flow Cytometry

HEK293 cells transfected with ZFNs and reporter vectors were trypsinized (0.05%, Gibco), washed in phosphate-buffered saline (PBS) pH 7.4 (Gibco), and resuspended in PBS supplemented with 2% FBS (Gibco). Single-cell suspensions were analyzed and sorted using a FACSARiaII (BS Biosciences) cell sorter. Cells with strong GFP and mCherry signals were isolated and collected as single cells and plated in a 96 well plate or as a pooled sample. The data were analyzed using FlowJo: Flow Cytometry Analysis Software.

2.6 Immunoblotting

Cells were trypsinized (0.05%, Gibco) and washed with PBS pH 7.4 (Gibco). The cells were lysed in RIPA buffer (1x TBS, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 0.004% sodium azide, Santa Cruz Biotechnologies), Complete protease inhibitor cocktail (Roche), and phosphatase inhibitor cocktails A and B (Santa Cruz Biotechnologies) for 1 hour at 4°C. Following centrifugation at 13,000 rpm for 10 mins, the supernatant was collected and total protein was quantified using the Pierce bicinchoninic acid (BCA) protein assay (Thermo
In total, 25µg of protein (1.5µg of protein for acid extracted lysates) were treated with lithium dodecyl sulfate (LDS) sample buffer (Life Technologies), and reducing agent (500mM dithiothreitol, Life Technologies), boiled at 85°C for 15mins, and resolved using gel electrophoresis. When probing for large molecular weight proteins, (KMT2D, KMT2C, EP300, and CREBBP) samples were size fractionated in precast 3-8% Tris-Acetate protein gels (Life Technologies) using Tris-Acetate SDS running buffer (Life Technologies). When probing for smaller molecular weight proteins (H3 and H3K4me1-3), samples were size fractionated in precast 4-12% Bis-Tris protein gels (Life Technologies) using MES SDS running buffer (Life Technologies). The resolved samples were transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore) in transfer buffer (Life Technologies), and blocked with 5% (w/v) skim milk in PBS pH 7.4 (Gibco) with 0.1% Tween 20 (Sigma). Membranes were probed with an antibody against either KMT2D (1:1,000, HPA035977, Sigma-Aldrich), KMT2C (1:500, SAB1300082, Sigma-Aldrich), EP300 (1:1,000, PA1-848, Affinity BioReagents), CREBBP (1:500, sc-369, Santa Cruz Biotechnologies), β-Tubulin (1:1,000, sc-9104, Santa Cruz Biotechnologies), H3 (1:5,000, ab1791, Abcam), H3K4me1 (1:5,000, pAB-037-050, Diagenode), H3K4me2 (1:5,000, pAB-035-050, Diagenode), or H3K4me3 (1:1,500, C42D8, Cell Signaling Technology) in 5% (w/v) skim milk in PBST overnight. The membrane was then probed with a goat anti-rabbit IgG-HRP antibody (1:5,000 or 1:15,000 for H3 and H3K4me1-2, sc-2004, Santa Cruz Biotechnologies) in 5% (w/v) skim milk in PBST for one hour at room temperature. Chemiluminescence was detected using a ChemiDoc MP system (Bio-Rad) following application of either Clarity western ECL substrate (Bio-Rad) or SuperSignal West Femto Substrate (Thermo Scientific). All blots were replicated using successive passages of cell lysates at least twice. Protein densitometry was performed using ImageJ. The intensity of the signal was calculated as the signal of target protein divided by the signal of the corresponding loading control. Statistical analysis was performed using GraphPad Prism6 software.

2.7 Acid Histone Extraction

Cells were collected using a cell scraper, washed twice with PBS pH 7.4 (Gibco), and resuspended in Triton extraction buffer (containing 0.5% Triton X-100 (v/v) (Sigma-Aldrich), protease inhibitor cocktail (Roche), and phosphatase inhibitor cocktails A and B (Santa Cruz Biotechnologies). The samples were then incubated at 4°C for 10mins on a sample rotator
before centrifuging at 2000rpm for 15mins at 4°C. The supernatant was discarded and the samples were washed in 50% of the volume of Triton extraction buffer at 4°C for 10mins on a sample rotator. The samples were then centrifuged at 2000rpm for 15mins at 4°C and the supernatant was discarded. The pellet was resuspended in 0.2N HCL (Fisher Scientific) at a cell density of 4x10^7 cells per mL, and then incubated at 4°C overnight on a sample rotator. Cell debris was then pelleted by centrifuging at 2000rpm for 10mins at 4°C, and the supernatant was collected. Total protein was quantified using Pierce bicinchoninic acid (BCA) protein assay (Thermo Scientific).

2.8 Surveyor CEL-I Assay

The surveyor mutation detection assay (Transgenomic), which utilizes the endonuclease CEL-I isolated from celery, was used to screen clonal populations of ZFN transfected HEK293 cells for KMT2D mutations. Transfected samples were washed with PBS (Gibco) and DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer’s instructions. The KMT2D ZFN target locus was amplified using polymerase chain reaction (PCR) (forward primer: CTCACCAAGCTCCCTGGTCA, reverse primer: GGATTGCCACCTGTCTAGA). For each sample, 200ng of genomic DNA template was used in a 50µL reaction run on a PTC-225 Peltier Thermal Cycler (MJ Research) for 30 cycles with an annealing temperature of 60°C. The PCR products (10µL) were denatured and allowed to randomly re-anneal (95°C 10min, 95°C to 85°C -2°C/s, 85°C to 25°C -0.1°C/s) using a PTC-225 Peltier Thermal Cycler (MJ Research), forming a population of homo- and heteroduplexes. CEL-I, which recognizes and cleaves heteroduplex DNA, was added to each sample and incubated at 42°C for 50 mins. The original PCR product, which was used as a control, and the digested sample were then resolved side by side on 2% agarose (Lonza) gels. The gel was incubated in a SYBR Green Nucleic Acid Gel Stain in Tris-acetate-EDTA (TAE) buffer for 30mins at room temperature and visualized using a FLA-7000 (Fujifilm). The presence of multiple bands in the resolved digested sample indicated a mutation (homoduplex DNA and the products of cleaved heteroduplex DNA)\textsuperscript{94,95}. The amount of cleaved heteroduplexes was determined using ImageJ. The intensity of the signal in the digested lanes was measured, and the fraction of cut heteroduplexes was calculated as the area of the cleaved heteroduplex products.
over the total area of the uncleaved and cleaved bands. This fraction was then used to estimate the percentage of mutated alleles in the population\(^{96}\).

\[
\text{ZFN activity} = 100 \times (1 - (1 - \text{fraction cut})^{0.5})
\]

Primers for analyzing potential \(KMT2D\) off-target effects were derived from the bioinformatic tool PROGNOS, which predicts potential off-target sites similar to the on-target sequence supplied\(^{97}\). The top two loci affecting coding regions in the genome were selected for analysis using the CEL-I assay.

### 2.9 RT-qPCR

The gene expression of \(KMT2A, KMT2B, KMT2C, KMT2D, CSPG4, GPR56, ENO3, LAMB3, RARA2, RARB2, RARG,\) and \(CASP9\) was assessed using real time reverse-transcriptase polymerase chain reaction (RT-PCR). \(RNA18S\) or \(PGK1\) were used as endogenous controls as they did not differ between \(KMT2D^{+/+}, KMT2D^{+/−}\), and \(KMT2D^{−/−}\) cell lines in treated or standard conditions. Primers were designed to bridge exon junctions (Table 2.1). For each sample, total RNA was extracted by column purification using an RNeasy mini kit (Qiagen) following the manufacturer’s instructions. A NanoDrop spectrophotometer (Thermo Scientific) was used to assess the quality and quantity of each sample. All samples were observed to have a 260/280 ratio between 1.95-2.05, and working dilutions of 50ng/\(\mu\)L were prepared. For each real time RT-PCR reaction, 100ng of RNA was used in a reaction volume of 10\(\mu\)L with power SYBR green RNA-CT chemistries (Life Technologies) as per manufacturers instructions. Each sample was run in technical triplicate using an ABI7900 Real time thermal cycler (Applied Biosystems). The \(\Delta\Delta CT\) method was used to calculate gene expression relative to controls\(^{98}\). Statistical analysis was performed using GraphPad Prism6 software.

### 2.10 DNA Sequencing

The ZFN target site was PCR amplified using primers in Table 2.1 (Sequencing Primers) and JumpStart Taq polymerase (Sigma-Aldrich). The PCR was run on a PTC-225 Peltier Thermal Cycler (MJ Research) for 30 cycles with an annealing temperature of 60°C. The PCR products were purified using the Agencourt AMPureXP PCR Purification System (Beckman Coulter) per the manufacturer’s instructions. The chain termination sequencing reactions were performed using BigDye terminator chemistries (Life Technologies) in 384 well reaction plates.
(Applied Biosystems) using a DNA Engine Tetrad 2 Peltier thermal cycler (Bio-Rad) for 50 cycles of 96°C for 10 seconds, 48°C for 5 seconds and 60°C for 4 minutes. The sequencing was performed on an AB3730XL DNA or ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Briefly, for TOPO cloning of the amplification products, the same PCR parameters and primers were used to amplify the ZFN target site using Jumpstart Taq polymerase (Sigma-Aldrich). The PCR products were then blunt-ended using a quick blunting kit (New England Biolabs) as per the manufacturer’s instructions. The processed PCR products were then cloned into E. coli using a Zero Blunt TOPO PCR Cloning Kit (Invitrogen) as per the manufacturer’s instructions. At least six colonies were selected and amplified per sample. Ligated vectors were then isolated using a DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer’s instructions. Purified vectors containing the captured PCR products were then PCR amplified and sequenced.

2.11 Retinoic Acid Time Course

Cells were seeded at a density of ≈50,000 cells/cm² (500,000 cells/well) in a 6 well vacuum-gas plasma tissue-culture treated polystyrene culture plate (Falcon) twenty-four hours prior to treatment. Each cell line was washed with PBS pH 7.4 (Gibco) and treated with either 0.1µM 9-cis-RA (Sigma-Aldrich) or an equal volume of DMSO (Sigma-Aldrich) (vehicle) for 2, 4, 6, 8, 12, or 24 hours. At each time point, cells were washed with PBS pH 7.4 (Gibco) and collected using a cell scraper. RNA was then immediately extracted using an RNeasy mini kit (Qiagen) following the manufacturer’s instructions. RNA expression was measured using RT-qPCR and the ΔΔCT method was used to calculate gene expression relative to the endogenous control PGK1, as well as each clone’s corresponding DMSO treated control sample. Statistical analysis was performed using GraphPad Prism6 software. The data were analyzed using repeated measures two-way ANOVA and a Bonferroni post hoc test with a significance threshold of P=0.05.

2.12 Alamar Blue Assays

Alamar blue (Life Technologies) was used to measure the growth rates of the cell lines. Alamar blue measures the metabolic rate of cells by examining the amount of blue dye (resazurin or 7-Hydroxy-3H-phenoxazin-3-one 10-oxide) that is irreversibly reduced in the presence of
NADPH (nicotinamide adenine dinucleotide phosphate) or NADH (nicotinamide adenine dinucleotide) to a pink dye (resorufin or 7-Hydroxy-3H-phenoxazin-3-one). Cells were seeded at a density of 750 cells/well in 96 well plates. For each time point the optical density (OD) was measured at 585nm using a Safire² (Tecan) plate reader following a 2 hour alamar blue (1:10 v/v) incubation for each time point. Growth rates were calculated relative to a media only control for multi-plate comparisons, the first time point post seeding, and the parental control cell line. The experiment was repeated three times using cells from successive passages. Statistical analysis was performed using GraphPad Prism6 software. The data were analyzed using repeated measures two-way ANOVA using a Dunnett post hoc test with a significance threshold of P=0.05.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Ensembl ID</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KMT2A</strong></td>
<td>ENSG00000118058</td>
<td>Ex26-27 Fwd:</td>
<td>GACCTCCTCATTCAACAACCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex26-27 Rev:</td>
<td>CTTCGAAGTCCAGAGGAGATG</td>
</tr>
<tr>
<td><strong>KMT2B</strong></td>
<td>ENSG00000272333</td>
<td>Ex6-8 Fwd:</td>
<td>GTCTCTGTCGACTGGCTCAGTAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex6-8 Rev:</td>
<td>GGCAACAGCGCTCTCAGTAT</td>
</tr>
<tr>
<td><strong>KMT2C</strong></td>
<td>ENSG00000055609</td>
<td>Ex35-36 Fwd:</td>
<td>CAGAACATGACAGAGAATGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex35-36 Rev:</td>
<td>CTGAGGTGTCAAGGAGACTCTG</td>
</tr>
<tr>
<td><strong>KMT2D</strong></td>
<td>ENSG00000167548</td>
<td>ZFN target site Ex38-39 Fwd:</td>
<td>CACCAGAAATGGTGTGTTGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZFN target site Ex38-39 Rev:</td>
<td>CAGAGCAAGATCCAGAACAACACT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex44-45 Fwd:</td>
<td>TCACTCCCTGAGAGCATCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex44-45 Rev:</td>
<td>AGTTGTGACAGGACGCTTTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex50-51 Rev:</td>
<td>CACTCATGATCAACCCCTACT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex50-51 Rev:</td>
<td>GATGACTTGGAGTGCACAAAC</td>
</tr>
<tr>
<td></td>
<td>Sequencing Primer Fwd:</td>
<td></td>
<td>CTCACCAAGCTCCCTGTTGCA</td>
</tr>
<tr>
<td></td>
<td>Sequencing Primer Rev:</td>
<td></td>
<td>GGATGGCCACCTGCTCTAGA</td>
</tr>
<tr>
<td><strong>GPR56</strong></td>
<td>ENSG00000205336</td>
<td>Ex9-10 Fwd:</td>
<td>CTACCTGAGCCCTCCCTCCTCTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex9-10 Rev:</td>
<td>GTGCACCTTGAGTGTGTGTC</td>
</tr>
<tr>
<td><strong>LAMB3</strong></td>
<td>ENSG00000196878</td>
<td>Ex10-11 Fwd:</td>
<td>AGCTTCCATTCAAGGACCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex10-11 Rev:</td>
<td>GCGTAGGTGACTCTGCTAGT</td>
</tr>
<tr>
<td><strong>RARA2</strong></td>
<td>ENSG00000131759</td>
<td>Ex 1-2 Fwd:</td>
<td>CTGTTCCTCCAGAGAAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex 1-2 Rev:</td>
<td>AGGGCTGAGCACCATATCTTT</td>
</tr>
<tr>
<td><strong>RARB2</strong></td>
<td>ENSG00000077092</td>
<td>Ex 5-6 Fwd:</td>
<td>GGTTCACCTGGCTGGAGCAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex 5-6 Rev:</td>
<td>AAGGCGCTGCTGAGAAGTA</td>
</tr>
<tr>
<td><strong>RARG</strong></td>
<td>ENSG00000172819</td>
<td>Ex 2-3 Fwd:</td>
<td>TGCAATGACAGCTCTCTGAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex 2-3 Rev:</td>
<td>CGATTCCCTGACCTGCTGTT</td>
</tr>
<tr>
<td><strong>CASP9</strong></td>
<td>ENSG00000132906</td>
<td>Ex 1-2 Fwd:</td>
<td>GAGACGGGCTCTCATCACCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex 1-2 Rev:</td>
<td>ATTCGACGCCGAGCTGAGT</td>
</tr>
<tr>
<td><strong>RNA18S5</strong></td>
<td>ENSG00000225840</td>
<td>Fwd:</td>
<td>CGGCTACCACATCCAAGGAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rev:</td>
<td>GCTGGAATTACCGCGCT</td>
</tr>
<tr>
<td><strong>PGK1</strong></td>
<td>ENSG00000102144</td>
<td>Ex4-5 Fwd:</td>
<td>GGGAAAGATGCTTCTGGGAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ex4-5 Rev:</td>
<td>TTGGAAAGATGAGCTCGGAAA</td>
</tr>
<tr>
<td><strong>MAL</strong></td>
<td>ENSG00000172005</td>
<td>PCR Primer Fwd:</td>
<td>AAAGAACCTGGGGGTGGGGGTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCR Primer Rev:</td>
<td>GCACCTCTCGGGTACCCATT</td>
</tr>
<tr>
<td><strong>SHARPIN</strong></td>
<td>ENSG00000179526</td>
<td>PCR Primer Fwd:</td>
<td>GGGGCCAAGGCTATTCCAAATTGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCR Primer Rev:</td>
<td>GCACCTCTCGGGTACCCATT</td>
</tr>
</tbody>
</table>

Table 2.1: Sequencing, PCR, and RT-qPCR primers.
3 Results

3.1 Generation of a KMT2D Knockout Cell Line Model

KMT2D functions in an evolutionarily conserved complex that consists of one KMT2 protein with four core components; Ash2L, WDR5, RBBP5, and DPY30\textsuperscript{33,34,99,100}. In addition to the four core complex components; KDM6A (UTX), PAXIP1 (PTIP), PAGR1 (PA1), and NCOA6 are specific to the KMT2C/KMT2D complex\textsuperscript{29,33,34,101}. Along with the identification of LOF mutations in several types of renal cancer, a previous study has shown the human embryonic kidney (HEK) 293 cell line expresses a fully functional KMT2C/KMT2D complex\textsuperscript{102}. Therefore I selected the HEK293 cell line to create a KMT2D LOF cell line model in which I could examine the effect of KMT2D loss on Histone 3 Lysine 4 methylation and transcriptional co-activation.

3.1.1 Screening of HEK293 Cells Transfected with KMT2D Zinc Finger Nucleases

The KMT2D knockout strategy utilized zinc finger nuclease (ZFN) technology. A pair of ZFNs targeting exon 39 of KMT2D (Chr12:49427414-49427455, NCBI build hg19), specifically targeting the codon for Q3684 of the translated KMT2D protein (5537 amino acids), were used to generate loss-of-function mutations in KMT2D. ZFNs are targeted to the DNA by their respective zinc finger DNA binding domains. Each individual zinc finger domain recognizes and binds an 18bp target sequence, giving the ZFNs a 36bp sequence specificity. This allows the FokI domains on each ZFN to come into contact and heterodimerize, forming an active endonuclease that cleaves the DNA. Incorrect repair of this specific double stranded break (DSB) by the error prone non-homologous end-joining (NHEJ) introduces a permanent small insertion or deletion (Figure 3.1A). If the DSB is repaired correctly, the ZFNs are able to re-bind and cleave the DNA repeatedly until either the ZFNs are degraded or an insertion or deletion is introduced into KMT2D which destroys the 36bp sequence recognized\textsuperscript{83,86,95,96,103}. Frameshift mutations at the KMT2D target locus result in changes to downstream amino acid codons and eventually the creation of a premature stop codon, preventing the translation of KMT2D’s C-terminal enzymatic SET domain, which catalyzes the methylation of histone 3 lysine 4 (H3K4)\textsuperscript{29}. The frameshift mutations that are predicted to be created by the KMT2D ZFNs are representative of the frameshift insertions and deletions, as well as nonsense mutations, in patient tumour
samples that are scattered across the ≈16 Kb of coding DNA prior to the enzymatic SET domain (Figure 1.3). The surveyor mutation detection assay (CEL-I assay) allows the identification of the ZFN induced mutations (Figure 3.1B).

ZFN mRNAs or plasmids were transfected into HEK293 cells using TransIT mRNA reagents. The efficiency of the transfection method was tested using an EGFP control vector prior to using the ZFNs, and was estimated to be greater than 50% as judged by fluorescence microscopy (Figure 3.2A). The sensitivity of the CEL-I assay was also tested prior to using the ZFNs. Control plasmids provided with the CEL-I assay kit, which differed only by one nucleotide that allow heteroduplex formation, were PCR amplified together in different ratios, digested using CEL-I, and run on a 2% agarose gel (Figure 3.2B). The assay was found to be sensitive to mutant allele frequencies from 50% down to 10%, which covered the expected frequency of $KMT2D$ mutant alleles (≈16 to 25% indicated by the ZFN manufacturer) induced by the ZFNs in a heterogeneous population of cells.

To determine whether ZFN delivery in the form of either mRNA or plasmid led to more frequent $KMT2D$ disruption, HEK293 cells were transiently transfected with ZFN mRNA, or pZFN1 and pZFN2, and then the frequency of $KMT2D$ disruption was assessed using the CEL-I assay (Figure 3.3A). Gene disruption signals were observed in HEK293 cells transfected with either ZFN mRNA or pZFNs, however $KMT2D$ was disrupted more frequently (10.9%) in the ZFN mRNA transfected HEK293 cells then the pZFN transfected HEK293 cells (3.4%). The heterogeneous population of ZFN mRNA HEK293 cells was serially diluted and individual monoclonal cell lines were tested for $KMT2D$ disruption using the CEL-I assay, as the mRNA ZFNs were observed to have a stronger $KMT2D$ disruption signal (Figure 3.3B). In total 32 clonal populations were tested, ten of which were positive for ZFN induced disruption of $KMT2D$ (Figure 3.3C). The efficiency of ZFN induced disruption of $KMT2D$ was higher than expected, however this may be due to continued disruption of $KMT2D$ by ZFNs as the cells continued to grow past the initial test of efficiency (Figure 3.3A). One limitation of the CEL-I assay is the inability to identify monoclonal cell lines with homozygous mutations, as no heteroduplexes would be formed during the reannealing step for CEL-I to cleave. Mixing PCR product from the parental wild type cell line with PCR product from the cell lines undergoing the screening can negate this limitation. The initial screen was not performed using mixed PCR
products; therefore its possible monoclonal cell lines with homozygous KMT2D were overlooked.

3.1.2 Characterization of Zinc Finger Nuclease Induced Mutations

The first objective of this thesis was to find clonal lines with homozygous or compound heterozygous frameshift mutations in KMT2D, and correspondingly no KMT2D protein expression. Therefore five of the positive clones (4, 14, 15, 16, 37, 40) were selected for further characterization (Figure 3.4). Two of the clones (14 and 15) failed to thrive in culture, and DNA sequencing revealed clone 4 was wild type for KMT2D (not shown), possibly as a result of the clonal line being founded by more than one cell with the wild type cells over taking the mutated cells in the population. Western blotting was performed to assess KMT2D expression at the protein level. Comparison between the parental untransfected cell line and three of the clones revealed all three clones (16, 37, 40) had reduced levels of KMT2D compared to the untreated parental HEK293 cell line (Figure 3.4B). The reduced levels of KMT2D protein corresponded to the presence of mixed signals in the sequencing traces following the targeted ZFN cut site (Figure 3.4C), which is indicative that insertions or deletions are present at the cut site. Although these results confirmed the activity of the KMT2D ZFNs at the targeted locus, the presence of KMT2D at the protein level, albeit reduced, suggested that each clonal line harbored an inactivating frameshift mutation in only one allele. Therefore, the screening process was improved to increase the likelihood of obtaining KMT2D−/− clonal cell lines.

3.1.3 Transfection of HEK293 with Zinc Finger Nuclease and Reporter Vectors, and Screening for KMT2D−/− Clonal Cell Lines

Although transfection of HEK293 cells with the zinc finger nucleases followed by serial dilution of cells resulted in the identification of 3 KMT2D+/+ monoclonal cell lines, no KMT2D−/− clones were identified. Therefore a new strategy was employed to improve the selection and isolation of HEK293 clones with biallelic LOF KMT2D mutations. Specifically, a reporter vector designed to assess the level of ZFN activity in a cell was co-transfected with the ZFN plasmids into HEK293 cells (Figure 3.5). The reporter vector consists of a CMV promoter followed by a red fluorescent protein (RFP) gene, a copy of the KMT2D ZFN target sequence, and an out of frame EGFP gene. KMT2D ZFNs recognize and cleave target sequences in the
genome as well as in the reporter vector, which then become susceptible to frameshift mutations introduced by error prone NHEJ. On the reporter vector, the presence of a frameshift mutation leads to a 1 in 3 chance of putting \textit{EGFP} in frame resulting in a functional GFP protein and green fluorescence. This fluorescence thus provides a fluorescent selectable marker of ZFN activity. RFP$^+$ and GFP$^+$ cells can then be sorted using flow cytometry and isolated as monoclonal lines. The CEL-I assay does not distinguish between monoclonal cell lines with monoallelic or biallelic \textit{KMT2D} mutations. Instead, screening the resulting clones for loss of \textit{KMT2D} using western blotting was selected as a method for identifying candidate monoclonal cell lines with LOF biallelic \textit{KMT2D} frameshift mutations. In summary, this strategy sought to increase the chances of finding \textit{KMT2D$^\sim$} monoclonal cell populations through flow cytometry enrichment while overcoming the limitations of the CEL-I assay to distinguish between lines with monoallelic or biallelic \textit{KMT2D} mutations.

The reporter vector as well as pZFNs were delivered to HEK293 cells using Trans-IT mRNA reagents and the same transfection protocol previously optimized (Figure 3.2A). The pZFNs were used to optimize the reporter flow cytometry screening because limited aliquots of the ZFN mRNAs were available. During optimization of the transfection procedure the pZFNs were found to induce a robust reporter driven GFP signal, therefore they were used for the remainder of the experiments. HEK293 cells were transfected with reagents only, reporter vector alone, or the reporter and two ZFN vectors (Figure 3.6 & 3.7). The cells were cultured for twenty-four hours at 37°C and then for forty-eight hours at 30°C, as previous studies have shown that mild hypothermic conditions increase the efficiency of ZFN induced gene disruption$^{88}$. Seventy-two hours post transfection the cells were assessed using fluorescence microscopy and sorted using flow cytometry (Figure 3.6 & 3.7). To determine the impact of the screening strategy on HEK293 cells and in follow up characterization experiments, control clonal cell lines were generated from HEK293 cells transfected with the reporter alone (Figure 3.6). Fluorescence microscopy revealed a population of RFP$^+$ GFP$^-$ cells that was determined by flow cytometry to be 17.9% of the transfected population (Figure 3.6A and B). However, a more stringent gating strategy targeting cells expressing higher levels of RFP (10.7% of the population) was used to isolate 192 individual RFP$^+$ cells in two 96 well tissue culture plates, as well as a heterogeneous pooled population of RFP$^+$ cells. After two weeks of growth, approximately 50% of the colonies survived and ten were selected to derive the clonal control
lines (denoted A1-A10). In addition to the clonal lines, the heterogeneous RFP\(^+\) cell population (denoted A parental) was used as a control line in follow up characterization experiments. KMT2D expression was observed using western blotting in all the derived control cell lines (Figure 3.6C).

HEK293 cells transfected with the reporter in addition to pZFN1 and pZFN2 were examined using fluorescence microscopy. The population was found to have approximately 10-20% RFP\(^+\) cells of which a proportion were also GFP\(^+\) (Figure 3.7A). Quantification of RFP\(^+\) and GFP\(^+\) cells using FACS analysis revealed the population of double positive cells to be 6.4% of the total population (Figure 3.7B). A more specific gate (3.10% of the transfected population) that selected for cells with high RFP and GFP expression was used to isolate and plate 480 individual cells into five 96 well plates. In total, 27 out of 480 monoclonal cell lines were derived using this method, of which thirteen clones survived expansion in culture (2.7%; 13/480) and were screened for biallelic inactivation of KMT2D using western blotting. Compared to the parental control line, six of these lines (C2, C8, C17, C18, C22, and C27) had reduced levels of KMT2D, and two lines (C19 and C20) were found to be negative for KMT2D (15.4%; 2/13 lines screened)(Figure 3.7C). In addition to the isolated RFP\(^+\) GFP\(^+\) clones, a heterogeneous pooled population of RFP\(^+\) GFP\(^+\) cells was collected and expanded for 7 days in culture before being stored at -80°C. This pooled population was then used in a second round of screening in order to identify more KMT2D negative clonal cell lines. After thawing, the pooled population was grown for an additional 4 days to increase the number of cells analyzed using FACS. As this population had previously been enriched with cells containing active ZFNs, flow cytometry was used to isolate and plate 1152 cells into 96 well plates using the side and forward scatter gate established in the previous experiment. In total, 384 clonal lines survived expansion in culture (33.3%; 384/1152), of which 131 cell lines were screened using western blotting for KMT2D. The cell lines were prioritized for KMT2D western blot screening based on the approximate growth rate of the cells through clonal expansion, as a reduction in KMT2D had previously been reported to negatively effect cell proliferation\(^29\). Of the 131 lines screened, 23 lines were found to be negative for KMT2D protein expression in a preliminary screen (Figure 3.8A). As no loading control was used in the primary screen the potential KMT2D negative lines were validated using western blotting and six were found to be true KMT2D negative lines (5.3%;
7/131 of lines screened) (Figure 3.8B). In total 9 KMT2D null lines were identified from the 144 cell lines isolated using flow cytometry and screened using western blotting.

3.1.4 Characterization of Zinc Finger Nuclease Induced Mutations

In addition to the 11 control cell lines isolated using flow cytometry, one cell line with wild type KMT2D expression (C3), two cell lines with reduced KMT2D expression (C17 and C22), four cell lines lacking KMT2D expression (C19, C20, D152, D243), and the untreated parental HEK293 cell lines (denoted C parental) were chosen for further characterization (Figure 3.9). To confirm the mutational status of KMT2D at the ZFN target locus, a CEL-I assay was performed on the selected clones. Confirming the western blot results, a KMT2D disruption signal indicated by the presence of three DNA bands was observed in the two cell lines with reduced KMT2D expression (C17 and C22), and the four cell lines lacking KMT2D expression (C19, C20, D152, D243). KMT2D disruption signals were not observed in the control cell lines (A pooled, A1-A10, and C parental), and the cell line with robust KMT2D expression (C3). Subsequently, the cell lines with a gene disruption signal (C17, C22, C19, C20, D152, D243), as well as the C3 and the C parental lines, were selected for DNA sequencing of the ZFN target locus in order to identify the specific mutation (Figure 3.9B and Table 3.1). Sequencing revealed the C parental and C3 line to be wild type, C17 to have only one allele mutated carrying a 4bp insertion, C22 to have a 9bp deletion in one allele, and C19 to harbor a homozygous 5bp insertion. However, the sequencing traces for lines C20, D152, and D243 were not clear enough to determine the exact mutations. Individual alleles from C20, D152, and D243 were then isolated using a TOPO cloning strategy and sequenced. This technique involved amplifying the ZFN target locus using PCR and capturing single alleles from the PCR products in vectors using blunt end ligation. Ligation products were introduced into E. coli, amplified and then isolated. Purified vectors were then sequenced. Using this method, C20, D152, and D243 were found to be compound heterozygotes; with C20 harboring a 4bp and 13bp deletion, D152 a 2bp and 8bp deletion, and D243 a 17bp deletion and 1bp insertion (Table 3.1).

The genotype of the cell lines chosen for further characterization (Figure 3.10) confirmed the western blot results (Figure 3.11A), as the four lines negative for KMT2D expression (C19, C20, D152, D243) contained biallelic frameshifting mutations resulting in the creation of a premature stop codon between 18 and 65 codons downstream of the codon Q3684 targeted by
the ZFNs (Table 3.1, Figure 3.11A). The heterozygous cell line C17, which contained a frameshifting mutation, resulting in a premature stop codon (p.S3686AfsX27), correspondingly showed a reduction in KMT2D expression. Interestingly, the line C22, which harbored a 9bp in frame deletion (p.Q3682_Q3684del), also showed a reduction in KMT2D expression, suggesting the deletion had a negative impact on the protein’s stability.

To determine the effect of the ZFN induced mutations on KMT2D mRNA transcripts, qPCR was performed on the set of cell lines selected for further characterization (Figure 3.10 and 3.11) using three different sets of primers targeting different exon junctions along the mRNA transcript (Figure 3.11B). The expression of KMT2D was found to remain constant using primer sets targeting Ex44-45 and Ex50-51 across individual KMT2D+/+, KMT2D+/-, and KMT2D−/− cell lines, with the exception of C19 and C20. Unexpectedly, the clones C19 and C20 were found to have a significant increase in KMT2D mRNA expression over the C Parental control (Ex44-45 P=0.0043 and P<0.0001 respectively, Ex50-51 P=0.0002 and P<0.0001 respectively). These results suggest that the mutations generated in C19 and C20 are impacting the stability of the mRNA transcript, which may be slowing the rate of the transcript’s degradation.

One set of primers (KMT2D ZFN target site Ex38-39) was designed specifically to target the ZFN cleavage site with the aim of distinguishing between wild type and mutant KMT2D mRNA transcripts (Figure 3.11B, C and D). Although the primers were not able to statistically differentiate between individual wild type and mutant KMT2D lines, grouping the expression data points from each cell lines according to genotype showed overall the KMT2D+/+, and KMT2D−/− cell lines were significantly different from KMT2D+/− cell lines (P=0.0013 and P<0.0001 respectively), but not from one another (P>0.9999). Overall, these results demonstrate limited utility of this RT-qPCR assay in differentiating between individual wild type and mutant KMT2D lines.

A limitation of using ZFNs is the potential for off-target cleavage and introduction of mutations in the genome, which could possibly obscure results in the cell lines created. Therefore I assessed two potential KMT2D ZFN off-target loci in the genome predicted by the bioinformatic tool PROGNOS97. This tool was designed to predict off-target sites for ZFNs by assessing target site homology as well as co-operative ZFN monomer binding based on in vitro data of the off targets of the CCR5 ZFNs97. The top two predicted off-target sites of the KMT2D ZFNs within the coding genome were SHARPIN (77% sequence homology) and MAL (72%
sequence homology). Potential SHARPIN and MAL mutations were assessed by CEL-I assay in the cell lines exposed to ZFNs (C3, C17, C22, C19, C20, D152, and D243) (Figure 3.12). The issue of detecting homozygous mutations in the clones was addressed by individually mixing PCR product from the control line (C parental) with the PCR products of the cell lines exposed to ZFNs. No mutations were identified in the two potential KMT2D ZFN off-target loci in the cell lines exposed to ZFNs.

3.2 Characterization of KMT2D Knockout Cell Lines

3.2.1 Analysis of KMT2D+/- Cell Lines Rate of Growth

To begin understanding the functional impact of the KMT2D mutations in the isolated cell lines, I examined proliferation rates because reduced cell growth is a known phenotype of KMT2D reduced or null cells\textsuperscript{28,29}. The growth rates of KMT2D+/+, KMT2D+/-, and KMT2D-/- cell lines were examined over a period of four days using alamar blue (Figure 3.13). Although individual variation between clones of each genotype was observed, overall KMT2D-/- cell lines were observed to have significantly reduced growth rates compared to KMT2D+/+ at 24 hours (P=0.0360), 48 hours (P<0.0001), and at 72 hours (P<0.0001) (Figure 3.13B). KMT2D+/- cell lines were observed to have reduced growth rates at 48 hours (P=0.0388) compared to KMT2D+/+ cell lines. This phenotype was consistent with other studies that have observed slight reductions in growth rates of KMT2D knock down or knock out cells using different methods of preventing KMT2D expression such as siRNA knockdown, or ZFN- and rAAV- (recombinant adeno-associated virus) induced gene knockout\textsuperscript{28,29}.

3.2.2 Analysis of KMT2D Dependent Genes

In addition to examining the growth phenotype of the KMT2D null cells to assess the functional impact of the ZFN induced KMT2D mutations, I analyzed the mRNA expression of several genes previously reported as KMT2D regulated genes. Given KMT2D’s role in gene expression regulation through H3K4 methylation, two studies assessed the target genes of KMT2D by comparing the gene expression profiles of wild type to knockdown HeLa or knockout KMT2D HCT116 (colorectal) cells\textsuperscript{41,52}. Four genes (CSPG4, ENO3, LAMB3, and GPR56) were demonstrated to be bound by KMT2D in the promoter region and strongly downregulated (-17.2 to -21.3 fold change) in the absence of KMT2D by Issaeva and
collaborators\textsuperscript{29}. Guo and collaborators identified two of these genes (\textit{LAMB3}, and \textit{GPR56}) to be downregulated (-0.63 and -1.48 log2 fold change respectively) in the knockout KMT2D HCT116 cells. Therefore, these genes were chosen for mRNA expression analysis by RT-qPCR in the \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, and \textit{KMT2D}\textsuperscript{−/−} cell lines (Figure 3.14). \textit{GPR56} was observed to be significantly downregulated in \textit{KMT2D}\textsuperscript{−/−} cell lines (P<0.0001) compared to \textit{KMT2D}\textsuperscript{+/+} cell lines (Figure 3.14A). The average expression of \textit{GPR56} trended down in \textit{KMT2D}\textsuperscript{+/−} compared to \textit{KMT2D}\textsuperscript{+/+} cell lines, however the difference was not significant. \textit{LAMB3} average expression also trended down in \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines compared to \textit{KMT2D}\textsuperscript{+/+} cell lines although the trend was not significant, as \textit{LAMB3} expression was variable across the \textit{KMT2D}\textsuperscript{+/+} cell lines (Figure 3.14B). The finding of reduced \textit{GPR56} expression in \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines, consistent with the previous studies, provided support for the notion that the \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines were valid LOF mutation models.

3.2.3 Expression of \textit{KMT2D}'s Paralogue \textit{KMT2C} is Unaffected by \textit{KMT2D} Genotype

Previous studies have reported that \textit{KMT2D} and \textit{KMT2C} can functionally compensate for one another\textsuperscript{21,24,27}. Additionally, in a mouse adipogenesis model, loss of \textit{KMT2D} resulted in increased \textit{KMT2C} mRNA expression (approximately 1.5-2 fold increase) during the differentiation of brown preadipocytes to adipocytes\textsuperscript{22}. Therefore I assessed if the mRNA expression of \textit{KMT2C} and that of two other genes with closely related functions, \textit{KMT2A} and \textit{KMT2B}, were altered in the \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines compared to \textit{KMT2D}\textsuperscript{+/+} cell lines. RT-qPCR was used to determine the effect of \textit{KMT2D} genotype on the mRNA expression of \textit{KMT2C}, \textit{KMT2A}, and \textit{KMT2B} (Figure 3.15A). However, no significant differences of \textit{KMT2A}, \textit{KMT2B}, or \textit{KMT2C} mRNA transcripts were observed between \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, and \textit{KMT2D}\textsuperscript{−/−} cell lines. To validate the mRNA data, \textit{KMT2C} protein levels were analyzed using western blotting (Figure 3.15B). Correlating with the results of the mRNA expression analysis, no differences in KMT2C protein levels were observed between \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, and \textit{KMT2D}\textsuperscript{−/−} cell lines.

3.2.4 Bulk H3K4me1 Levels are Reduced in \textit{KMT2D}\textsuperscript{−/−} Cell Lines

As KMT2D is known to function as a histone lysine methyltransferase targeting enhancer H3K4\textsuperscript{21,28,29,33}, it was of interest to investigate the bulk H3K4 methylation profile in \textit{KMT2D}\textsuperscript{+/+},
Control lines A6-A10 (Figure 3.10) were not used as previous results (Figure 3.13-15) indicated control lines A Pooled and A1-5 were sufficient to evaluate clonal heterogeneity. Notably, western blot analysis of H3K4me1, H3K4me2, and H3K4me3 using antibodies that recognize the different methylation states revealed a significant reduction in H3K4me1 in KMT2D-/- cell lines compared to KMT2D+/+ cell lines (P=0.037) and a further significant reduction compared to KMT2D+/- cell lines (P<0.0001) (Figure 3.16). Bulk H3K4me1 was also significantly reduced in KMT2D+/- cell lines compared to KMT2D+/- cell lines (P=0.0006). Additionally, a reduction in bulk H3K4me2 was also observed in KMT2D+/- (P=0.025) and KMT2D-/- (P=0.0009) cell lines compared to KMT2D+/- cell lines. However there was no difference in H3K4me2 between KMT2D+/- and KMT2D-/- cell lines. These results demonstrate loss of KMT2D is sufficient to significantly affect levels of bulk mono- and di-methylation of H3K4, independent of KMT2C. This result is in agreement with the results of a previous study that examined the H3K4 methylation profile in human medulloblastoma KMT2D knockout cell lines, further validating the KMT2D+/- and KMT2D-/- cell lines I isolated as KMT2D LOF mutation models.

3.2.5 Induction of Retinoic Acid Response Genes is Reduced in KMT2D-/- Cell Lines

Results from the H3K4 methylation profile demonstrated a KMT2C independent role for KMT2D in maintaining H3K4 monomethylation in the isogenic cell lines I isolated. Previous studies have demonstrated that KMT2D, in cells with a null KMT2C background, is associated with transcriptional coactivation by monomethylating H3K4 at active enhancer loci\(^{21,22}\). Therefore I wanted to determine if KMT2D also had an independent role in transcriptional co-activation in the isogenic cell lines I isolated. The KMT2D/KMT2C complex has been shown to associate with nuclear hormone receptors through an interaction with NCOA6\(^{33}\). In particular, when individually knocked down by siRNAs, KMT2D and its paralogue KMT2C were found to have redundant roles in retinoic acid receptor (RAR) transactivation of RARB2 mRNA expression in response to RA\(^{24}\). Therefore RAR transactivation of gene expression in response to 0.1µM 9-cis-RA or DMSO (vehicle) was assessed in KMT2D+/- and KMT2D-/- cell lines. The concentration of 0.1µM 9-cis-RA was chosen based on a previous study that examined RAR transactivation in HEK293 cells\(^{24}\). RAR transactivation was assessed by measuring mRNA expression of three RA inducible genes (RARA2, RARB2, and RARG) at select time points over...
24 hours using RT-qPCR (Figure 3.17). Interestingly, although there was some variation between individual clonal lines, RAR transactivation was significantly reduced in $KMT2D^{+/+}$ cell lines compared to $KMT2D^{+/+}$ cell lines for all three genes between 4 and 12 hours, with the greatest difference observed at the activation peak (6 hours). The most robust difference in RAR transactivation was observed in the RA induced expression of $RARA2$, which by 6 hours showed an 8.6 fold change difference between $KMT2D^{+/+}$ and $KMT2D^{-/-}$ cell lines (Figure 3.17A). Moreover, the induction of a RARA specific RA response gene, $CASP9$, in response to 9-cis-RA was also assessed in $KMT2D^{+/+}$ and $KMT2D^{-/-}$ cell lines using RT-qPCR. The expression of $CASP9$ mRNA was assessed after 6 and 8 hours of 0.1µM 9-cis-RA treatment (Figure3.18A). Although the effect was small, the expression of $CASP9$ mRNA was significantly reduced in $KMT2D^{-/-}$ cell lines compared to $KMT2D^{+/+}$ cell lines (P<0.0001 for 6 and 8 hours) (Figure3.18B). Interestingly, two of the wild type clonal lines (A3 and A5) that did not show RA induced induction of $CASP9$ as robustly as the other wild type clones also expressed KMT2D at lower levels that the other wild type clones (Figure 3.11). These results are compatible with the notion that KMT2D plays a role in mediating efficient RAR transactivation.

Given KMT2D is associated with transcriptional co-activation through the monomethylation of H3K4 at enhancer loci$^{21,22}$ and the large number of genes associated with RARs$^{104}$, I investigated if RAR transactivation exacerbated the differences in bulk H3K4 methylation between $KMT2D^{+/+}$ and $KMT2D^{-/-}$ cell lines. The cell lines were treated with either DMSO (vehicle) or 9-cis-RA for 6 hours and H3K4 mono-, di-, and tri-methylation profiles were assessed using western blotting (Figure 3.19). Consistent with previous results (Figure 3.16), a ≈50% reduction of bulk H3K4me1 was observed in $KMT2D^{-/-}$ cell lines. The reduction in H3K4me1 was significant (P<0.0001) and consistent between 9-cis-RA treated and DMSO treated, however the difference in H3K4me1 levels was not exacerbated using RA. A significant (P=0.0055) reduction of H3K4me2 in 9-cis-RA $KMT2D^{-/-}$ cell lines compared to $KMT2D^{+/+}$ cell lines was also observed, a difference that was not observed between DMSO treated $KMT2D^{-/-}$ and $KMT2D^{+/+}$ cell lines. However, a significant decrease in bulk H3K4me2 (P=0.0009) was also detected in the H3K4 methylation profile of $KMT2D^{-/-}$ cell lines in untreated cells (Figure 3.16), therefore the difference may not be induced by RA. Therefore, it remains questionable that RA treatment exacerbates differences at the level of bulk H3K4 methylation between $KMT2D^{+/+}$ and $KMT2D^{-/-}$ cell lines.
3.2.6 Expression of Transcriptional Co-activator CREBBP Correlates with KMT2D

Besides H3K4me1\textsuperscript{21}, active enhancers are characterized by the presence of H3K27ac, which is deposited by the acetyltransferases p300 (KAT3B) and its parologue CREBBP (KAT3A)\textsuperscript{21}. KMT2D has also been found to colocalize with the acetyltransferase p300 (KAT3B), and H3K27ac genome wide at enhancer elements. In order for H3K27 to be acetylated, any H3K27 methylation marks have to be removed, an activity catalyzed by the KMT2D complex member KDM6A\textsuperscript{105,106}. Therefore, as an initial step in investigating the KMT2D complex and the other enzymes associated with active enhancers, the expression of acetyltransferases p300 and CREBBP was examined in $\text{KMT2D}^{+/+}$, $\text{KMT2D}^{+/-}$, and $\text{KMT2D}^{-/-}$ cell lines using western blotting (Figure 3.20). Surprisingly, the expression of CREBBP but not p300 correlated with KMT2D expression across $\text{KMT2D}^{+/+}$, $\text{KMT2D}^{+/-}$, and $\text{KMT2D}^{-/-}$ cell lines. These results suggest KMT2D may play a role in regulating the expression of, or maintaining the protein stability of H3K27 acetyltransferase CREBBP.
Figure 3.1: Diagram of ZFN induced mutagenesis and workflow for ZFN induced mutation screening.

(A) ZFNs target sequence and corresponding location in exon 39 of KMT2D (Chr12:49427414-49427455, NCBI build hg19). ZFNs cleave the DNA following the recognition and binding of the zinc finger domains to their respective target sequences and heterodimerization of the FokI domains. Correctly repaired double stranded breaks in KMT2D via the NHEJ repair pathway can be targeted again by the ZFNs as the binding sequence remains intact. Incorrect repair of the double stranded DNA break by NHEJ creates short insertions and or deletions in the coding sequence of KMT2D, which can lead to a frameshift that results in the downstream creation of a premature stop codon. (B) Workflow used to create and screen for KMT2D<sup>+</sup> HEK293 cells. Mutations induced by ZFNs were detected using the surveyor mutation detection assay (CEL-I assay). A red x represents a mutation and the scissors represent the CEL-I enzyme. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a gel.
Figure 3.2: Evaluation of transfection methods and the CEL-I assay.
(A) Transfection efficiency of HEK293 cells using Trans-IT mRNA reagents and a control EGFP plasmid. Brightfield and fluorescence photomicrographs were taken three days post transfection. Scale bars represent 100µm. (B) Assessment of CEL-I assay on PCR products generated from control vectors (Ctrl. G and Ctrl. C) that differ by one nucleotide. Three different ratios of the vectors were used to assess the sensitivity of the assay. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a 2% agarose gel. The red arrows indicate bands generated from cleaved heteroduplex DNA.
Figure 3.3: Assessment of ZFN induced \textit{KMT2D} mutations in HEK293 cells using the CEL-I assay.

(A) ZFN mRNAs induce a higher frequency of \textit{KMT2D} mutations. ZFNs were delivered to HEK293 cells encoded in either plasmids (2.5µg of each pZFN) or mRNA (2µg of each ZFN mRNA). Heteroduplex formation was assessed 72 hours post transfection using the CEL-I assay. The red arrows indicate bands generated from cleaved heteroduplex DNA. The expected size of the PCR amplicon was 416 bp, and the heteroduplex bands were 230 bp and 184 bp. (B) Clonal serial dilutions of the pooled population of \textit{KMT2D} ZFN mRNA transfected cells were performed and individual clonal colonies were screened for \textit{KMT2D} mutations using the CEL-I assay. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a 2% agarose gel. Clones positive for \textit{KMT2D} disruption are highlighted in red. The red arrows indicate bands generated from cleaved heteroduplex DNA.
Figure 3.4: Characterization of KMT2D mutations in three clonal populations isolated from KMT2D ZFN mRNA transfected HEK293 cells.

(A) Three clonal lines were found to harbour KMT2D mutations using the CEL-I assay. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a 2% agarose gel. The red arrows indicate bands generated from cleaved heteroduplex DNA. The expected size of the PCR amplicon was 416bp, and the heteroduplex bands were 230 bp and 184 bp. (B) The level of KMT2D protein abundance was assessed using western blot in the untransfected HEK293 parental population and the three clonal cell lines. β-Tubulin was used as a loading control. (C) Electropherograms of ZFN induced mutations detected using the CEL-I assay verified by Sanger sequencing. The top panel shows the parental wild type sequence with the ZFN cut site highlighted with the black bar. The bottom three panels show the same locus in the three monoclonal cell lines. Mixed traces following the ZFN cut site indicate the presence of a ZFN induced insertion/deletion.
Figure 3.5: Workflow for creating $KMT2D^{+\cdot}$ cell lines using $KMT2D$ ZFNs and a fluorescent reporter vector.

The ZFN reporter vector contained a copy of the red fluorescent protein ($RFP$) gene under the control of a CMV promoter, along with a copy of the $KMT2D$ ZFN target sequence and an out of frame enhanced green fluorescent protein ($EGFP$) gene. Active $KMT2D$ ZFNs within cells that received a reporter vector will cleave the vector, which is susceptible to frameshift mutations introduced through incorrect NHEJ, leading to a 1 in 3 chance of putting $EGFP$ in frame and subsequently expression of a green fluorescent selectable marker. Flow cytometry was then used to select and isolate cells with active ZFNs by sorting out and plating individual RFP$^+$ GFP$^+$ cells into 96 well plates. Isolated cell lines were then screened using western blotting and an antibody against KMT2D.
Figure 3.6: Generation of control clonal cell lines using flow cytometry.

(A) HEK293 cells were transfected with the KMT2D ZFN reporter or transfection reagent only (Untransfected). Brightfield and fluorescence photomicrographs were taken 3 days post transfection prior to flow cytometry. Scale bars represent 100µm. (B) Untransfected and KMT2D ZFN reporter transfected HEK293 cells were sorted using flow cytometry. RFP⁺ cells were isolated from the highlighted population (10.7% 1,430/13,390 events) using flow cytometry, and collected as a pool, or plated one cell per well. (C) The status of KMT2D in ten clonal lines and the pooled sample were assessed using western blotting. β-Tubulin was used as a loading control.
Figure 3.7: Generation of KMT2D null cell lines using KMT2D ZFNs, reporter vector, and flow cytometry.

(A) HEK293 cells were transfected with plasmids encoding ZFNs targeting KMT2D (pZFNs) and a KMT2D ZFN reporter vector. RFP+ cells indicate the transfection efficiency and GFP+ cells indicate active ZFNs within the cell. Brightfield and fluorescence photomicrographs were taken 3 days post transfection and prior to flow cytometry. Scale bars represent 100µm. (B) The left plot indicates negative, RFP+, GFP+, and RFP+ GFP+ cells. The GFP+ only population represent false positives as indicated by Figure 3.7B. RFP+ GFP+ cells were isolated from the highlighted red population (3.1%) of cells indicated in the right plot, and plated one cell per well in 96 well tissue culture plates. (C) KMT2D−/− cell lines were screened by western blot. The untransfected HEK293 parental line (C Parental) and clone 40, a KMT2D+/− clone previously identified, were used as controls. β-Tubulin was used as a loading control.
Figure 3.8: Screening for additional KMT2D null cell lines.

(A) Cell lines isolated from a pooled population of enriched RFP$^+$ GFP$^+$ HEK293 ZFN and reporter transfected cells were screened for KMT2D expression using western blotting. $KMT2D^-$ clones C19 and C20 previously identified were used as a control for KMT2D loss, and the untransfected HEK293 parental line was used as control for the presence of KMT2D. Clones highlighted in red indicate potential KMT2D protein null cell lines. (B) As the western blot screen in A did not account for loading controls, potential $KMT2D^-$ clones identified in the primary KMT2D western blot screen were confirmed using western blotting and protein lysates from a successive passage of the clones. KMT2D null clones C19 and C20 previously identified were used as positive controls for KMT2D loss, and the untransfected HEK293 C Parental line was used as a positive control for KMT2D expression. β-Tubulin was used as a loading control. Clones highlighted in red indicate KMT2D protein null cell lines.
Figure 3.9: Confirmation of KMT2D’s mutational status in 19 cell lines selected for further characterization.

The 19 cell lines were selected for further characterization based on the KMT2D western blots used to screen for KMT2D-/- cell lines. (A) KMT2D mutations were confirmed using the CEL-I assay. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a 2% agarose gel. (B) DNA sequencing analysis of ZFN induced mutations detected using the CEL-I assay. Each line of sequence represents one allele. Nucleotide deletions are shown by (-) and insertions are shown in blue boxes. ZFN cutsites are underlined.
Table 3.1: ZFN induced mutations in selected cell lines detected using western blotting and CEL-I assays, and verified using DNA sequencing.

<table>
<thead>
<tr>
<th>Clone</th>
<th>KMT2D Genotype</th>
<th>Nucleotide Change</th>
<th>Chromosome Position (hg19)</th>
<th>Amino-Acid Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>Homozygous</td>
<td>WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C17</td>
<td>Heterozygous</td>
<td>WT</td>
<td>c.11055_11056insGCAA</td>
<td>chr12:49,427,433 p.S3686AfsX27</td>
</tr>
<tr>
<td>C22</td>
<td>Heterozygous</td>
<td>WT</td>
<td>c.11046_11054delACAGCAACA</td>
<td>chr12:49,427,442 p.Q3682_Q3684del</td>
</tr>
<tr>
<td>C19</td>
<td>Homozygous</td>
<td>c.11055_11056insAACAT</td>
<td></td>
<td>chr12:49,427,433 p.S3686NfsX65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c.11055_11056insAACAT</td>
<td></td>
<td>chr12:49,427,433 p.S3686NfsX65</td>
</tr>
<tr>
<td>C20</td>
<td>Compound</td>
<td>c.11052_11055delACAT</td>
<td></td>
<td>chr12:49,427,437 p.Q3684HfsX64</td>
</tr>
<tr>
<td></td>
<td>Heterozygous</td>
<td>c.11050_11062delCAACATTCTGGTG</td>
<td></td>
<td>chr12:49,427,438 p.Q3684GfsX60</td>
</tr>
<tr>
<td>D152</td>
<td>Compound</td>
<td>c.11055_11056delTT</td>
<td></td>
<td>chr12:49,427,433 p.S3686WfsX26</td>
</tr>
<tr>
<td></td>
<td>Heterozygous</td>
<td>c.11043_11051delCAACAGCA</td>
<td></td>
<td>chr12:49,427,444 p.S3682TfsX26</td>
</tr>
<tr>
<td>D243</td>
<td>Compound</td>
<td>c.11061_11077delTGGGGCTGGATCCCTGG</td>
<td></td>
<td>chr12:49,427,427 p.G3688WX18</td>
</tr>
<tr>
<td></td>
<td>Heterozygous</td>
<td>c.11055_11056insA</td>
<td></td>
<td>chr12:49,427,433 p.S3686FfsX26</td>
</tr>
</tbody>
</table>
Figure 3.10: Summary of cell line origin, *KMT2D* mutational status, and KMT2D protein status in 19 cell lines selected for further characterization.

Cell lines selected for characterization of *KMT2D*<sup>+/+</sup>, *KMT2D*<sup>+/−</sup>, and *KMT2D*<sup>−/−</sup> phenotypes. The letters appended to the clone number represent cell lines that were isolated from the same experiment. All cell lines were derived from the C parental HEK293 line. Cell lines with an ‘A’ originated from C parental cells that were transfected with the *KMT2D* ZFN reporter and then isolated using FACS to select RFP<sup>+</sup> cells. Cell lines with ‘C’ originated from C parental cells transfected with *KMT2D* ZFNs and ZFN reporter, and isolated using FACS to select RFP<sup>+</sup> and GFP<sup>+</sup> cells. Cell lines with ‘D’ were isolated using FACS from a pooled population of RFP<sup>+</sup> and GFP<sup>+</sup> cells collected alongside the ‘C’ lines.
Figure 3.11: Characterization of *KMT2D* mRNA and protein expression in 19 cell lines used in further characterization experiments.

(A) *KMT2D* expression was assessed using western blotting on a selection of *KMT2D*^{+/+}, *KMT2D*^{+/-}, and *KMT2D*^{-/-} cell lines chosen for further characterization experiments. β-Tubulin was used as a loading control. (B) *KMT2D* mRNA expression was assessed using qPCR on a selection of *KMT2D*^{+/+}, *KMT2D*^{+/-}, and *KMT2D*^{-/-} cell lines chosen for further characterization experiments. mRNA expression was normalized to the parental HEK293 cell line (C parental) and 18S rRNA. Log10 fold change (S.D.) is shown (n=3). Three different primers targeting different positions along the mRNA transcript were used. The primer pair *KMT2D ZFN* target site Ex38-39 was designed so that the forward primer annealed to the endonuclease cleavage site. For each set of primers, differences between individual cell lines compared to the C parental line were analyzed using a one-way ANOVA with Dunnett’s post hoc test (adjusted P values; **P<0.01, *** P<0.001, **** P<0.0001). (C) Analysis of *KMT2D* mRNA expression using *KMT2D ZFN* target site Ex38-39 primers. Log10 fold change (S.D.) is shown for the individual cell lines (n=3). (D) The three *KMT2D ZFN* target site Ex38-39 expression data points for each of the individual cell lines was grouped according to genotype. Log10 fold change (S.E.M.) and adjusted P values from a one-way ANOVA with Bonferroni’s post hoc test are shown for the grouped analysis.
Figure 3.12: Analysis of the top two predicted KMT2D ZFN off-target sites in the cell lines exposed to KMT2D ZFNs.

The parental cell line (C parental) and cell lines exposed to KMT2D ZFNs were assessed for off-target mutations at two predicted homologous sites within the coding genome (MAL and SHARPIN). (A) MAL and (B) SHARPIN mutations were assessed using the CEL-I assay. Undigested (UD) and CEL-I digested (D) PCR fragments were separated on a 2% agarose gel. The presence of bands resulting from cleaved heteroduplex DNA indicates the presence of a mutation at the locus under examination. The size of the predicted cleavage products are indicated.
Figure 3.13: Assessment of cell proliferation rates in $KMT2D^{+/+}$, $KMT2D^{+/-}$, and $KMT2D^{-/-}$ cell lines. 

(A) The proliferation of cell lines was assessed using alamar blue. Measurements were normalized to readings taken 24 hours after initial seeding (denoted 0Hrs) and the parental HEK293 control cell line (C Parental). The optical density (OD) was measured at 585nm following a 2 hour alamar blue incubation for each time point. The Log10 mean (S.D.) is displayed (n=3). 

(B) Grouped genotype analysis of proliferation rates. The Log10 mean (S.E.M.) is displayed. Grouped data were analyzed by repeated measures two-way ANOVA using a Dunnett’s post hoc test (significant values are relative to $KMT2D^{+/+}$ lines, *P<0.05, ****P<0.0001). * The $KMT2D^{+/+}$ cell line C3 was considered an outlier and not included in the grouped genotype analysis.
Figure 3.14: Expression of genes previously reported as KMT2D dependent.
(A and B) GPR56 and LAMB3 mRNA expression were assessed using RT-qPCR. mRNA expression was normalized to the parental HEK293 cell line (C parental) and 18S rRNA. Log10 fold change (S.D.) is shown for plots displaying individual clonal lines (n=3). Log10 fold change (S.E.M.) is shown for grouped genotype analysis. Grouped data were analyzed by one-way ANOVA using a Bonferroni post hoc test (**** P<0.0001; n.s. not significant).
Figure 3.15: Expression of KMT2D homologues in KMT2D<sup>+/+</sup>, KMT2D<sup>−/−</sup>, and KMT2D<sup>−/−</sup> cell lines.

(A) KMT2A, KMT2B, and KMT2C mRNA expression was assessed using RT-qPCR. mRNA expression was normalized to the parental HEK293 cell line (C parental) and 18S rRNA. Log10 fold change (S.D.) is shown (n=3). Data were analyzed using a one-way ANOVA with a Dunnett’s post hoc test and the results were not significant. (B) Western blot of KMT2C, and KMT2D with β-Tubulin used as a loading control.
Figure 3.16: \( \text{H3K4 methylation profile in } KMT2D^{+/+}, KMT2D^{+/-}, \text{ and } KMT2D^{-/-} \text{ cell lines.} \) (A) Western blot of total H3, H3K4me1, K3K4me2, and H3K4me3. (B) Analysis of H3K4 methylation profile using protein densitometry. Relative protein abundance normalized to total H3 is displayed as mean (S.D.) of \( KMT2D^{+/+}, KMT2D^{+/-}, \) or \( KMT2D^{-/-} \) cell lines. Data were analyzed using one-way ANOVA using a Bonferroni post hoc test (n.s. not significant).
Figure 3.17: Induction of RARA2, RARB2, and RARG mRNA expression in response to 9-cis-retinoic acid.

(A-C) RT-qPCR was used to assess mRNA expression of selected genes in RNA samples isolated from cells treated with either DMSO (vehicle) or 9-cis-RA (0.1µM) at selected time points over 24 hours. mRNA expression was normalized to each cell line’s corresponding DMSO treated (vehicle) for each time point and PGK1 (Phosphoglycerate kinase 1). Mean Log10 fold change (S.D.) is shown for plots displaying individual clonal lines (n=3). Solid black lines indicate KMT2D+/+ cell lines and dotted lines indicate KMT2D−/− cell lines. Mean Log10 fold change (S.E.M.) is shown for grouped genotype analysis of RA induced genes. Grouped data were analyzed using repeated measures two-way ANOVA using a Bonferroni post hoc test (**P<0.01, ***P<0.001, ****P<0.0001).
Figure 3.18: Induction of the RARA dependent gene CASP9 in response to 9-cis-retinoic acid.

(A) RT-qPCR was used to assess CASP9 mRNA expression in RNA samples isolated from cells treated with either DMSO (vehicle) or 9-cis-RA (0.1µM) at 6 or 8 hours. mRNA expression was normalized to each cell line’s corresponding DMSO treated (vehicle) for each time point and PGK1. Mean Log10 fold change (S.D.) is shown (n=3). (B) Grouped genotype analysis of CASP9 mRNA expression. Log10 fold change (S.E.M.) is shown. Grouped data were analyzed using repeated measures two-way ANOVA using a Bonferroni post hoc test (****P<0.0001).
Figure 3.19: H3K4 methylation profiles in \textit{KMT2D}^{+/+} and \textit{KMT2D}^{-/-} cell lines treated with 9-cis-retinoic acid.

(A) Western blot of total H3, H3K4me1, K3K4me2, and H3K4me3. \textit{KMT2D}^{+/+}, and \textit{KMT2D}^{-/-} cell lines were treated with either DMSO (vehicle) or 9-cis-RA (0.1µM) for 6 hours before total protein was acid extracted. (B) Analysis of H3K4 methylation profiles using protein densitometry. Relative protein abundance normalized to total H3 is displayed as mean (S.D.) of \textit{KMT2D}^{+/+} or \textit{KMT2D}^{-/-} cell lines from two independent experiments. Data were analyzed using two-way ANOVA and a Bonferroni post hoc test (n.s. not significant).
Figure 3.20: Expression of transcriptional co-activator CREBBP (CBP) correlates with KMT2D expression.
Western blot of p300, CREBBP, and KMT2D with β-Tubulin used as a loading control.
4 Discussion

4.1 Generation of \textit{KMT2D} Knockout Cell Line Model

In this thesis I implemented several techniques to identify monoclonal cell lines with ZFN induced mutations. The generation of gene knockout models that mimic disease-associated LOF mutations in human cells has generally been challenging in the past due to the low efficiency of the methods\textsuperscript{83}. At the time the work in this thesis was carried out, the use of targeted ZFNs to disrupt a gene locus was an effective method of creating gene knockout models in human cells. A somatic gene knockout approach was chosen over a knockdown approach with short interfering RNAs to ensure that low levels of KMT2D protein expression were not obscuring phenotypes in the generated cell lines.

A pair of ZFNs targeting Exon39 or Q3684 of the translated KMT2D protein (5537 amino acids) was used to permanently introduce targeted deleterious insertions or deletions into the \textit{KMT2D} gene (Figure 3.1A). The location of the targeted frame shift mutations would prevent the translation of the enzymatic SET domain, which is known to catalyze the methylation of H3K4\textsuperscript{29,101}. As almost all deleterious mutations in human cancers prevent the translation of the SET domain (Figure 1.3), the insertion/deletion mutations induced by the ZFNs are consistent with the deleterious insertion/deletion mutations that have been identified in cancers. In addition to the set domain, ZFN induced mutations also prevent the translation of FYRN and FYRC, four LXXLL motifs, and a zinc finger motif. Little is known about the function of the FYRN and FYRC domains except that they are present in a large number of chromatin associated proteins\textsuperscript{107}. Two of the LXXLL motifs have been shown to be involved in associating KMT2D with ESR1\textsuperscript{30}.

Although the targeted mutations prevent the translation of the SET domain, there is a possibility that a truncated protein consisting of first \approx3650 amino acids of KMT2D would be expressed. The functional domains that are known to exist in the first \approx3650 amino acids of KMT2D are two clusters of three PHD domains, and an HMG box domain. In general, PHD domains are known to mediate a protein’s interaction with histones, and the second cluster of PHD domains in KMT2D (4-6) have been shown to mediate an interaction between KMT2D and H4R3me0 or H4R3me2a (asymmetrically dimethylated), an interaction which was required for KMT2D’s methyltransferase activity on H3K4\textsuperscript{46,108}. HMG box domains are known to interact
and bind DNA\textsuperscript{109}. Therefore it cannot be ruled out that a truncated version of KMT2D may be able to competitively inhibit other proteins from binding histones. However, more than half of all the nonsense and frameshift mutations in patient tumours are found to disrupt the translation of KMT2D after these domains, suggesting the artificial mutations induced by the ZFNs are representative and valid models of LOF KMT2D mutations in cancer, which may also generate truncated versions of KMT2D.

In an initial test of the ZFN mediated \textit{KMT2D} disruption, I observed the ZFNs were more efficient at disrupting KMT2D in a heterogeneous population of HEK293 cells when delivered in the form of mRNAs, 10.9\% compared to 3.4\%. This was lower than the expected frequencies of 25.8\% and 16.7\%, which were calculated during the validation of the ZFNs by the manufacturer. However, this might be due to differences in the cell lines, as the validation was done on K562 cells. In addition the ZFN efficiencies I calculated reflected those found in other studies (6-18\%)\textsuperscript{86,103}. Given the better efficiency of the ZFN mRNAs, serial dilutions of the ZFN mRNA treated populations were performed to isolate clonal cell lines. In total, 31\% (10/32) of the clones isolated showed a KMT2D disruption signal using the CEL-I assay, which detects heteroduplex formation. This was higher than expected based on the previous results on the population of ZFN transfected cells; however when the clonal lines were further analyzed, no biallelically mutated \textit{KMT2D} clones were identified. Although CEL-I screening worked relatively well to process and screen a large number of clonal lines, these results highlighted the major drawback to CEL-I screening; the lack of differentiation between clones with mutations in one or both alleles of \textit{KMT2D}. Therefore I moved to a more detailed screening strategy that took into account ZFN activity, as well as used western blotting to distinguish between clones with \textit{KMT2D} mutations that induce partial or complete loss of KMT2D.

Screening for clones using a reporter vector that fluorescently marks cells with active ZFNs allowed me to quantify the efficiency of the ZFNs in the population and then select cells expressing active ZFNs for follow up KMT2D western blot screening in one step. In total, 2.77\% of the population of HEK293 cells co-transfected with the reporter vector and pZFNs were found to express active ZFNs, which was close to the percentage (5\%) obtained in another study using the same reporter and \textit{TP53} ZFNs in HEK293 cells\textsuperscript{89}. The reporter vector contains a hygromycin resistance gene (\textit{HYGRO}), which is expressed in conjunction with \textit{EGFP} after ZFN induced mutations shift the \textit{EGFP} back into frame. This selectable marker was not utilized in
my process, as there were concerns that during the selection for drug resistance, the wild type KMT2D cells containing a vector expressing HYGRO-EGFP would overtake the heterogeneous population and bias against the recovery of KMT2D<sup>−</sup> clones. This is because the loss of KMT2D has been linked to reduced growth rates in the literature<sup>28,29</sup>.

From the population of ZFN positive cells, 480 were isolated using flow cytometry for clonal expansion. A difficulty encountered during this phase of the screening was the low survival rate of the isolated cells in culture (2.7% 13/480). A recovery of only 2.7% of ZFN positive cells compared to the recovery of approximately 40% of isolated cells from the untransfected and reporter alone transfected HEK293 cells suggests that the ZFNs may be to some degree cytotoxic. This is consistent with previous work that demonstrated ZFN cytotoxicity resulted in relative cell death rates of ≈40% in HEK293 cells<sup>110</sup>. This may also explain why a higher recovery rate (33.3%) of individually sorted cells was observed for the second round of screening. The second round of screening involved collecting a heterogeneous population of ZFN positive cells during the first round of screening and selecting individual cells (D clones) by flow cytometry after two weeks of growth in culture. Therefore, by the time cells were selected using flow cytometry during the second round of screening, the cells may have already titrated out the ZFN vectors. ZFN cytotoxicity also suggests off-target effects may be occurring, since ZFN cytotoxicity is the result of off-target cleavage of the DNA<sup>83</sup>. However, in the case of the obligate heterodimer ZFNs targeting CCR5, all off-target loci that were identified had greater then 66.7% sequence homology to the ZFN target site. Furthermore, the locus with the most off target effects was CCR5’s paralogue CCR2 (91% sequence homology). The top two closest gene coding genomic loci resembling the KMT2D ZFN were in SHARPIN and MAL with 77% and 72% sequence homology respectively. However, no off-target mutations were observed. There is the potential other off-target mutations were induced by the KMT2D ZFNs, although they are expected to be limited due to limited homology of other loci to the KMT2D ZFN target site. The presence of multiple copies of the KMT2D ZFN target site in the reporter vectors also may act to titrate out excess KMT2D ZFNs that could have potentially targeted other locations in the genome. Additionally, the use of several clonal KMT2D<sup>−</sup> cell lines is expected to negate any specific differences that may be due to off-target effects.

In total, 9 KMT2D null clonal lines were identified using the reporter vector screening approach, representing 0.55% (9/1632) of all clones isolated by flow cytometry. Of the clonal
lines screened by western blot the 9 clonal lines represent 6.25.5% (9/144), which means one clonal line with suspected biallelic mutations in KMT2D was identified for every 16 clonal lines screened. Only one other study has used these ZFNs (Cat #: CKOZFND14397-1KT, Sigma-Aldrich), however they do not report how many lines or types of lines they screened, and were only able to identify KMT2D+/− cell lines by CEL-I screening in medulloblastoma cell lines28.

The impact of ZFN induced mutations on the KMT2D mRNA transcript was also evaluated using RT-qPCR. Overall the mutations had little effect on the levels of KTM2D mRNA transcript (Figure 3.10). This is not altogether unexpected as the presence of a mutation may or may not affect the stability of the mRNA, depending on whether nonsense codons downstream of the frameshift mutations activate nonsense mediated mRNA decay (NMD) pathways. Relative decreases in KMT2D transcripts may not have been observed because the target site for the ZFN induced frameshifts and downstream nonsense codons are found in a large open reading frame, which has been linked to a reduction in NMD111. KMT2D−/− cell lines C19 and C20 showed significant increases in KMT2D transcripts relative to the C Parental line (Figure 3.11). It is not known why KMT2D mRNA transcript levels were elevated in C19 and C20. A possible explanation for the increased mRNA transcript levels may be that the mutations stabilized the secondary structure of the mRNA transcripts, impacting the degradation rate of the transcripts. However, there are no similar examples reported in the literature.

One RT-qPCR assay was tested that was designed to distinguish between wild type and mutant KMT2D transcripts by using a primer that binds to the ZFN target locus. Significant differences were observed between the average levels of KMT2D transcripts for each genotype. However, this assay was not able to identify any significant differences between individual KMT2D+/+, KMT2D+/−, or KMT2D−/− cell lines. The inability to reliably identify a KMT2D−/− cell line largely comes from the specific differences in mutations induced by the ZFNs. For example, the KMT2D−/− cell line D243 had no observable difference in KMT2D transcript levels compared to the C parental. This is explained by the fact one allele harbors a deletion outside the primer target site, and the other allele has a 1bp insertion which may not change the primer binding region sufficiently to prevent the primer from binding. All together, these results demonstrate limited use for screening for KMT2D knockout clones by RT-qPCR.

Altogether, several methods were integrated in order to accomplish the first objective of this thesis; to create a human KMT2D LOF cell line model using a targeted ZFN somatic gene
knockout strategy. These results emphasize the importance in screening methods that select for ZFN activity within cells and the mutation status of each allele, as the rate of identifying KMT2D clonal cell lines with biallelic mutations was low. These methods are also transferable to newer genome editing tools such as transcription activator-like effector nucleases (TALENs) and clustered regulatory interspaced short palindromic repeat/Cas RNA-guided DNA endonucleases (CRISPR/CAS9)\textsuperscript{112}. These methods were not available when this project was initiated but also cleave DNA at targeted loci in order to induce mutations by error prone NHEJ.

4.2 Characterization of KMT2D Knockout Cell Lines

I determined \textit{KMT2D}\textsuperscript{+/+} and \textit{KMT2D}\textsuperscript{−/−} cell lines had reduced growth rates compared to the \textit{KMT2D}\textsuperscript{+/+} cell lines, confirming previous findings\textsuperscript{28,29}. However, the wild type cell line C3 was observed to have a considerable reduction in growth rate, outside the observed clonal variability of the other isolated wild type cell lines not exposed to ZFNs. This gave rise to the possibility that ZFN off target effects were responsible for an overall reduction in growth rate of the clonal lines exposed to the ZFNs. Conversely, C3 may have originated from a cell with a growth rate outside the variability observed in the wild type lines. None the less, the reduction in growth rate observed in the \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines is consistent with reductions in growth rates observed in other studies using different methods of preventing KMT2D expression such as siRNAs, and rAAV-induced gene knockout\textsuperscript{28,29}. Therefore, C3 was considered an outlier and not included in the grouped analysis of \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, and \textit{KMT2D}\textsuperscript{−/−} cell line growth rates.

To establish the cell lines I generated as valid \textit{KMT2D} mutant models, the mRNA expression of two genes that had been previously reported as \textit{KMT2D} regulated genes was assessed\textsuperscript{29,52}. One of the two genes (GPR56) was found to have significantly reduced expression in the \textit{KMT2D}\textsuperscript{+/−} and \textit{KMT2D}\textsuperscript{−/−} cell lines, consistent with the previous studies. This finding provided support for the notion that the \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, and \textit{KMT2D}\textsuperscript{−/−} cell lines I isolated were valid LOF mutation models.

In pursuit of my second objective, I examined the H3K4 methylation profile of \textit{KMT2D}\textsuperscript{+/+}, \textit{KMT2D}\textsuperscript{+/−}, or \textit{KMT2D}\textsuperscript{−/−} cell lines. The results of the H3K4 methylation profile demonstrate loss of KMT2D is sufficient to reduce bulk methylation of H3K4me1 and H3K4me2 in the cell. These results are consistent with a previous study that generated a similar
isogenic KMT2D knockout cell line model\textsuperscript{28}, validating the KMT2D\textsuperscript{+/-}, KMT2D\textsuperscript{+/+}, and KMT2D\textsuperscript{-/-} cell lines I isolated as KMT2D knockout models. Currently, the independent nature of KMT2D’s role in maintaining bulk H3K4me1 is under some debate in the literature. Recently, Hu and collaborators investigated bulk H3K4 methylation in KMT2D null HCT116 cells with a KMT2C null background\textsuperscript{21}. They observed a reduction in bulk H3K4me1 in KMT2C null human HCT116 cells. However, in mouse embryonic fibroblasts wild type or mutant for KMT2C, KMT2D knockdown resulted in reduced bulk H3K4me1 only in the KMT2C mutant line, suggesting redundant roles for KMT2D and KMT2C. A possible explanation for the consistent bulk H3K4me1 levels observed in KMT2C wild type cells with and without KMT2D knock down may be that low levels of KMT2D expression (20-40\%) were enough to maintain bulk H3K4me1 levels. Supporting this notion, reduction in bulk H3K4me1 was also observed in a KMT2D knockout human medulloblastoma cell line wild type for KMT2C\textsuperscript{28}. Only one study has examined the H3K4 methylation profile in isogenic KMT2C wild type and mutant cells, and slight reductions in H3K4me1 were observed\textsuperscript{22}. However, this difference may be due to loading differences between the KMT2D wild type and mutant samples. Therefore since my results and the results of the other study using a KMT2D isogenic medulloblastoma cell line model indicate knockout of KMT2D is sufficient to observe a large reduction in bulk H3K4me1, it is unlikely KMT2C plays a role in maintaining bulk H3K4me1 in these cell lines.

The results of the H3K4me profile also suggest KMT2D may play a role in maintaining bulk H3K4 dimethylation. This result has also been previously observed\textsuperscript{21,22,28,46}, and has been linked to active enhancer regions. Correspondingly, KMT2D has been shown to associate with H3K4me1 and H3K4me2 deposition genome wide, as well as with p300 and H3K27ac and Pol II at active enhancers in ChIP-sequencing experiments\textsuperscript{21,22}. Altogether, these results suggest a model where KMT2D plays a role in the activation of enhancers by catalyzing the monomethylation of H3K4 at enhancer loci, which in turn promotes the expression of enhancer regulated genes. Therefore, KMT2D mutations in cancer might function to silence the expression of a gene or set of genes that block cancer progression.

In this thesis I also demonstrated a unique role for KMT2D in nuclear hormone transactivation in pursuit of my third objective. Specifically, I observed reduced induction of retinoic acid receptor (RAR) response genes in the presence of 9-cis-retinoic acid in KMT2D\textsuperscript{-/-} cell lines. This is a novel finding, as previous results demonstrated only knockdown of both
*KMT2C* and *KMT2D* resulted in reduced RAR transactivation, as well as LXR transactivation\(^{24,26}\). Utilization of a knockdown rather than a somatic gene knockout approach, which as stated above with regard to KMT2D’s role in bulk H3K4me1, may not have reduced the levels of KMT2D expression enough to observe reduced RAR transactivation in a wildtype KMT2C background. However, this does not mean that KMT2C has no role in RAR transactivation, as the strongest reduction in RAR transactivation in the literature was observed in KMT2C and KMT2D knockdown HEK293 cells. Induction of KMT2C expression may explain why the RAR response, as assessed by measuring the expression of RAR response genes, equalized between *KMT2D\(^{+/+}\)* and *KMT2D\(^{-/-}\)* cell lines over time. However, the equalization in RAR response may also be due to a feed back regulatory loop in which the induction of RARA expression, which aids in strengthening the cells RA response, also in turn induces the expression of a set of catabolic genes belonging to the cytochrome P450 family; *CYP26A1*, *CYP26B1*, and *CYP26C1*. The expression of these genes have previously been shown to be induced by RA, and the proteins they encode function to catabolize the derivatives of RA\(^{21,24-27,113}\). Therefore, it remains an open question if KMT2C can fully compensate for the loss of KMT2D in RAR transactivation over time.

In spite of the equalization of RAR transactivation between *KMT2D\(^{+/+}\)* and *KMT2D\(^{-/-}\)* cell lines by 24 hours, the reduction in RAR transactivation at earlier time points may have a functional impact on the cell. It has been reported that exposure to RA for as little as 8 hours has a functional effect on differentiating mouse embryonic stem cells through activation of erk1/2 signalling, an effect that correspondingly was blocked when RAR agonists were used\(^{50,114}\). Additionally, in a different study, genome wide changes in gene expression have been observed in mouse embryonic stem cells treated with RA for 8 hours\(^{22,28,104}\). Mahony and collaborators found RAR bound at 1145 genes, enrichment in Pol II initiation and elongation at 214 of these genes, and the differential expression of 81 genes\(^{104}\). These results may explain why no changes in bulk H3K4 methylation were observed between *KMT2D\(^{+/+}\)* and *KMT2D\(^{-/-}\)* cell lines, assuming KMT2D’s methyltransferase activity is responsible for promoting RA induced transcription, as short term RA induction is only observed for a limited number of genes. Therefore examining enhancer elements of specific genes such as *RARA* using ChIP-qPCR with antibodies against KMT2D, and H3K4me1/2/3 would determine if KMT2D’s methyltransferase activity is mediating RA induction.
KMT2D and RA signaling might explain the mechanism of some of the KMT2D mutations in cancer since RAR signaling has been linked to growth inhibition and differentiation of cancer cells, and in particular myeloid and lymphoid cancer cells. Furthermore, knockdown or knockout of RARA in mice leads to the development of lymphoma, and a common translocation in acute promyelocytic leukemia fusing PML to RARA creates an oncoprotein that prevents transcription of RAR target genes in response to physiological concentrations of RA. Interestingly, increasing the expression of the protein KDM6A, which is also found in the KMT2D complex and has H3K27 demethylase activity, is known to help overcome PML-RARalpha RA insensitivity by inducing RAR target genes and differentiation of NB4 leukemia cells. Therefore, my results suggest a model where inactivating KMT2D mutations might aid cancer cells in evading RA induced differentiation.

The finding that CREBBP protein expression correlates with KMT2D protein expression also presents an interesting link to enhancer-associated proteins, which has not been observed before. CREBBP and its parologue p300 are overall structurally similar (63% sequence identity) with strongly homologous histone acetyltransferase domains and some, but not complete, functional redundancies in gene regulation. CREBBP and p300 are known to act as transcriptional co-activators through their acetyltransferase activity on a number of residues on the tails of H2B, H3 (including H3K27), and H4 tails, as well as activity on a large number of transcription factors that play key roles in many cellular processes. In Drosophila, CREBBP has been shown to interact with KDM6A and the KMT2C/KMT2D homologue TRR, as well as co-localize with KDM6A, H3K27ac, and H3K4me1. Additionally, p300 and correspondingly H3K27ac have been found to co-localize genome wide in human cells. Interestingly, CREBBP has also been linked to RA induced RAR and RXR mediated transcriptional activity, directly interacting with RAR through its LXXLL motifs as well as through the KMT2D complex member NCOA6. Therefore, reduction in CREBBP in KMT2D−/− cell lines may further reduce enhancer activity for certain genes by reducing enhancer H3K27ac. Furthermore, this correlation in expression also may help explain why bulk H3K27ac is also reduced in KMT2D null cells. Alternatively, reduced H3K27ac may also be the result of decreased availability of unmodified H3K27 to other acetyltransferase proteins such as p300, due to reduced targeting of the H3K27 demethylase KDM6A to genomic loci in KMT2D null cell lines. The mechanism by which CREBBP protein expression correlates with KMT2D is
unknown, but may have to do with protein stability since the mRNA expression of CREBBP is not altered in TCGA patient samples with KMT2D mutations\textsuperscript{70}. Further work examining CREBBP mRNA expression and protein stability via proteasome inhibition is needed to answer this question, as well as assessment of bulk H3K27ac in the cell lines I generated. Additionally, rescue of CREBBP protein levels by reintroducing KMT2D back into the cell lines would further establish KMT2D’s role in maintaining CREBBP expression.

Interestingly, CREBBP is the second most frequently mutated gene in follicular lymphoma (64% of patients) behind KMT2D (82-89% of patients)\textsuperscript{1,4}. The majority of patients carry monoallelic mutations consisting of a deletion, frameshift, nonsense, or missense mutations in the histone acetyltransferase domain\textsuperscript{4}. Besides a role in histone acetylation, CREBBP has also been shown to be involved in modulation of BCL6 and p53 activity through acetylation of the proteins. CREBBP mutations in follicular lymphoma impair the acetylation of BCL6, which is required to inactivate BCL6’s transcriptional repressor activity on DNA damage, cell cycle arrest, and differentiation genes\textsuperscript{122}. CREBBP mutations prevent the acetylation of p53 which is required for its transcriptional activity and cancer surveillance activities\textsuperscript{122}. Therefore, the correlation between KMT2D and CREBBP protein expression could possibly provide a mechanism by which KMT2D mutations lead to the indirect disruption of BCL6 and p53 tumour suppressor activity.

Summarizing the findings in this thesis, I created human LOF $KMT2D$ cell line models that aided in the investigation of KMT2D biology. Although functional compensation between KMT2D and its paralogue KMT2C have been reported\textsuperscript{21,24-27}, the results in this thesis suggest KMT2D plays a role in (1) maintaining bulk H3K4me1 marks in the genome consistent with previous studies, and (2) the robust transactivation of nuclear hormone receptors. I also determined that the loss of KMT2D does not impact the mRNA or protein expression of $KMT2C$, or the mRNA expression of $KMT2A$ and $KMT2B$; a finding that is consistent with the notion that the loss of KMT2D is not compensated for by an increase in expression of similar functioning genes.

Based on my results and the literature I propose a model, similar to the model proposed by Herz et al.\textsuperscript{50}, where KMT2D plays a unique role in the deposition of bulk H3K4me1 marks in the genome, which negatively impacts activation of gene transcription in response to different stimuli, such as nuclear hormones. Additionally, I add CREBBP protein levels are associated
with KMT2D protein expression, which may explain why a reduction in bulk H3K27ac has been observed in cells without KMT2D\textsuperscript{22,28}. Furthermore, inactivating mutations in KMT2D might aid cancer cells in escaping RA induced differentiation or growth arrest, by interrupting the activation of enhancer elements of genes that induce differentiation, as seen in neural RA differentiation models in the literature.

4.3 Future Directions

Although it is known H3K4me1 is enriched at and necessary in the activation of enhancer elements, the diversity of proteins modifying the targeting of KMT2D’s enzymatic activity to specific loci in the genome is unknown. In addition, given the number of enhancers in the cell is estimated to be greater than 55,000, and the epigenetic state of enhancers is more variable then promoters across cell lineages\textsuperscript{23}, it is unknown if enhancer elements dysregulated by KMT2D LOF mutations in cancer are common across, or specific to different cell lineages. Despite these unknowns, the retinoic acid signaling axis is known to play a role in the differentiation of different cell lineages, including neuronal and hematopoietic lineages, and may represent a common signaling axis that is suppressed by KMT2D LOF mutations. To further investigate this hypothesis, I would determine if B-cell differentiation by retinoic acid signaling is impaired in the context of KMT2D LOF mutations. In order to more broadly understand the enhancers dysregulated across cancers that harbor KMT2D LOF mutations, I would examine the enhancer elements and genes dysregulated by loss of KMT2D in B-cells genome wide. Integrating these data with the recently published human epigenome roadmap\textsuperscript{123} could reveal KMT2D regulated enhancers shared between cell lineages that develop KMT2D mutated cancers. Although cell lineage specific KMT2D enhancers are likely involved in KMT2D mutated cancers, understanding the common aspects of KMT2D biology in cancer may reveal a vulnerability that could be exploited therapeutically.
References


42. Lee, S., Roeder, R. G. & Lee, J. W. Roles of histone H3-lysine 4 methyltransferase

43. *Overview of the Roadmap Epigenomics Project* at <http://www.roadmapepigenomics.org/overview>


66. Banka, S. *et al.* Novel KDM6A (UTX) mutations and a clinical and molecular review of


120. Chakravarti, D. et al. Role of CBP/P300 in nuclear receptor signalling. Nature 383, 99–
103 (1996).

