A role for histone H3, histone H4 and histone associating proteins DNMT3A and PHF6 in *JAK2V617F* positive myeloproliferative neoplasms

---

Nisha Rao

Bachelor of Science (Hons) Molecular Biology

November 2014

Thesis submitted for the degree of Doctor of Philosophy in the discipline of Genetics at the University of Adelaide

Table of Contents

Abstract

The Philadelphia chromosome negative myeloproliferative neoplasms (MPN); polycythemia vera, essential thrombocythemia and primary myelofibrosis, are clonal disorders harbouring the specific Janus kinase 2 (*JAK2*) lesion (*JAK2V617F*) at a high frequency. Accumulating evidence from pedigrees of MPN together with the identification of a plethora of heterogeneous lesions identified in sporadic MPN patients suggest that *JAK2V617F*, and other acquired changes in *JAK2*, cooperate with mutations in other genes to generate clonal disease. The nature of the other mutations dictates the disease phenotype and contributes to the potential for transformation to acute leukaemia. Emerging research is focussed on understanding the contribution of these other changes to MPN pathogenesis. As many of the other recurrent mutations reported in MPN affect genes involved in epigenetic regulation, studies have focused on identifying the role of epigenetic changes. Many epigenetic regulators mediate their effects via interaction with post-translationally modified histone H3 and H4 and, given the findings that pathogenic mutations are present in histone H3 in other tumours, the focus here was on the role of histone H3 and H4 variants in MPN pathogenesis.

Thus, a key aim of this PhD project was to identify pathogenic coding variants in the *histone H3* and *histone H4* genes in MPN. In the first study, MPN peripheral blood mononuclear cells or granulocyte patient samples were screened using Sanger sequencing for *histone H4* coding region variants. The screen identified previously unidentified sequence variants in several of the 15 *histone H4* genes. A coding variant of *HIST1H4C*, resulting in the substitution of cysteine for arginine [R, (*HIST1H4C:p.R4C*)], affects a known key residue involved in epigenetic regulation

(R3 residue on the mature protein). This gene was also shown to make a major contribution to the *histone H4* mRNA pool in several haemopoietic cell types, further indicating a potential for this variant to confer functional consequences. This was tested using enforced expression of the variant in two cell lines, HEK293 and a myeloid cell line (FDM cells). Further, it was demonstrated by RNA microarray and QPCR analyses of FDM cells expressing *HIST1H4C*:*p.R4C* that this variant conferred selective differential expression of five genes.

We extended this analysis of *Histone H4* genes to screen for disease-associated variants in *histone H3* genes (*n*=17) and the histone-H3 interacting protein *PHF6* (consisting of 9 coding exons). For this, we used an amplicon-based next generation sequencing (NGS) approach and the Roche 454 sequencing platform. This identified a coding region variant in *HIST1H3E* (*HIST1H3E:p.A96V*), the presence of which was confirmed by Sanger sequencing. A number of other changes identified by the NGS approach were not confirmed by Sanger re-sequencing, however the possibility that these are present in the original patient sample at a level below the detection limit for Sanger sequencing cannot be excluded. Sanger sequencing of the *PHF6* terminal exons 9 and 10 identified a somatic mutation (*PHF6R335fs*) in a PV patient.

Finally a Sanger-based sequencing screen of the terminal exon of gene encoding DNA (cytosine-5-)-methyltransferase 3 alpha (*DNMT3A*) identified somatic R882C and M880V substitutions in two PV patients. Clonal analysis for these mutations in *DNMT3A* indicated that their acquisition can either precede or follow the acquisition of *JAK2V617F*.

.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution. To the best of my knowledge and belief it contains no material previously published or written by another person, except where due reference has been made in the text.

In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to a copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time. (*Please see errata sheet on the following page*)

Nisha Rao

November 2014

Errata Sheet

Due to lack of copyright and authorship certification, the following figures and sections from the thesis are not available publicly in the electronic version of this thesis:

Chapter 1

Figure 1.1, *Figure.1 2*, *Figure 1. 3*, *Figure 1. 4*, *Figure 1.5*, *Figure 1. 6*, *Figure 1. 7*, *Figure 1.8*, *Figure1. 9* and *Figure 10*

Chapter 2

Figure 2.5, *Section 2.3.5* and *Section 2.3.6*

Chapter 4

Figure 4.1

Chapter 5

Figure 5.1

Acknowledgements

I would like to thank the Queen Elizabeth Hospital and the University of Adelaide for their continual financial support during this PhD candidature. I would also like to thank the staff and students of the Queen Elizabeth Hospital Foundation, Royal Adelaide Hospital, and SA Pathology. I am grateful for the scientific communications that made my PhD candidature possible.

I would like to thank my supervisor; Prof. Richard D'Andrea, who continually supported and guided me throughout this PhD. I am truly grateful for all the support and encouragement that you have provided. I would also like to thank my co-supervisor; Ms Carolyn Butcher, whose many scientific discussions made this PhD candidature possible.

I am grateful for all of the friendships with several members of the Haematology department at the Queen Elizabeth Hospital, the Acute Leukaemia Laboratory at SA Pathology, and within the Basil Hetzel Institute for Translational Research. Thank you for all the conversations and after-hours socialising, which made this PhD candidature enjoyable.

I also wish to thank members at work especially Prof. Stephen Nicholls and my wonderful colleagues for being patient and understanding towards my study commitments. I know it would have been very difficult otherwise.

I wish to thank my parents and sister for motivating me continuously. I certainly would not have come this far without your constant encouragement and love.

To friends (Alpana, Jordan, Teresa, and Donna to name only a few), thank you for your emotional support and motivation throughout this PhD candidature. I know I could not have been encouraged more if it had not been for your constant words of support.

And lastly, but certainly not least, I wish to thank Shannon. You have inspired and motivated me to move forward, even in times of a complete lack of positivity or motivation. Thank you for all your love and support.

Abbreviations

2-HG	2-hydroxyglutarate
3´ UTR	3 prime untranslated region
4-OHT	4-Hydroxy tamoxifen
α-KG	α-ketoglutarate
Ac	Acetylated
AID	Activated Induced cytidine Deaminase
AML	Acute Myeloid Leukaemia
Allo-SCT	Allogeneic-Stem Cell Therapy
ALL	Acute lymphoblastic leukaemia
AKT	Protein kinase B
BCR-ABL1	Break point Cluster Region-Abelson murine leukaemia viral oncogenes, homolog 1
BER	Base Excision Repair
BFUE	Blast Forming Colonies Erythroid
BM	Bone Marrow
BMMNC	Bone marrow mononuclear cell
BTG2	BTG family, member 2
$CXCL_{12}$	Chemokine (C-X-C motif) ligand 12
CAR	$CXCL_{12}$-abundant reticular
CALR	Calreticulin
CEL	Chronic Eosinophilic Leukaemia
CMP	Committed Myeloid Progenitor
CLP	Committed Lymphoid Progenitor
Chr	Chromosome

CML	Chronic Myeloid Leukaemia
CMML	Chronic Myelomonocytic Leukaemia
CNL	Chronic Neutrophilic Leukaemia
C/EBPα	CCAAT/enhancer binding protein, alpha
CD	Cluster of Differentiation
CFU	Colony Forming Unit
COSMIC	Catalogue of Somatic Mutations in Cancer database
DNMT3A	*DNA (cytosine-5-)-methyltransferase 3 alpha*
Dock10	*Dedicator of cytokinesis 10*
e-BFUE	endogenous Blast Forming Colonies Erythroid (EPO independent colonies)
EEC	Endogenous erythroid colony
EPO	Erythropoietin
EED	Embryonic Ectoderm Development protein
EGR1	Early growth response protein-1
ERK	Extracellular signal-regulated kinase
ET	Essential thrombocythemia
EP	Erythroid Progenitor
ER-LBD	Estrogen Receptor-Ligand Binding Domain
EVS	Exome Variant Server
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum
FBN-III	Fibronectin type III
FDM	Foetal Derived Murine
FERM	4.1, ezrin, radixin, moesin homology domain

GAPDH	*Glyceraldehyde 3-phosphate dehydrogenase*
GATA1	GATA binding protein 1
GATA2	GATA binding protein 2
*g*DNA	Genomic DNA
GEV16	Gal4-Estrogen receptor-VP16 transactivation domain fusion protein
GAL4-DBD	GAL4-DNA binding domain
GFI-1	Growth factor independence-1
GFP	Green Fluorescent Protein
G-CSF	Granulocyte- Colony Stimulating factor
GM-CSF	Granulocyte Macrophage-Colony Stimulating factor
GMP	Granulocyte Monocyte Progenitor
H3K27me^3	Histone H3 lysine 27 trimethylation
H3Y41ph	Histone H3 tyrosine 41 phosphorylation
H4R3me^{2}a	Histone H4 arginine 3 asymmetric di-methylation
H4R3me^{2}s	Histone H4 arginine 3 symmetric di-methylation
HDE	Histone Downstream Element
HEK293	Human Embryonic Kidney 293
HEL	Human Erythro-Leukaemia 92.1.7
HGF	Haematopoietic growth factor
HIF	Hypoxia-inducible factor
HIST1	*Histone cluster 1*
HIST2	*Histone cluster 2*

HIST3	*Histone cluster 3*
HiNF-P	Histone nuclear factor–P
HIRA proteins	Histone cell cycle regulation defective homolog A (*Saccharomyces cerevisiae*)
HMR	Hidden Mat Right locus
HP-1α	Heterochromatin protein-1alpha
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem/Progenitor cell
IDH	Isocitrate dehydrogenase
IGV2.1.2	Integrative Genomics Viewer version 2.1.2 software
IL	Interleukin
IL-3	Interleukin-3
IL-5	Interleukin-3
Il4ra	*Interleukin-4 receptor alpha*
INF-α	Interferon alpha
JAK2	Janus Kinase 2
JAK2V617F	Janus Kinase 2 Valine 617 Phenylalanine
JAK2V617F^{+}	*JAK2V617F*-positive
JH domain	JAK Homology domain
K-562	Human erythromyeloblastoid leukaemia cell
KIT	v-kit Hardy-Zucherman 4 feline sarcoma viral oncogene homolog
LIMMA	Linear Modelling for Microarray Analysis
LDH	Lactate dehydrogenase

LTHSC	Long-term repopulating haematopoietic stem cell
LMO2	*LIM (Lin11, Isl-1 & Mec-3) domain only 2 (rhombotin-like 1)*
LNK	SH2B adaptor protein 3
LIF-1	Leukaemia Inhibitory Factor-1
LTR	Long Terminating Repeats
µ	Micro (10^{-6})
µM	micro Molar
µg	microgram
µL	microlitre
M	Methionine
MACS	Magnetic Activated Cell Sorting
MAPK	Mitogen-activated protein kinase
MDS	Myelodysplastic syndrome
MDP	Macrophage/dendritic progenitor cell
MEP	Megakaryocyte Erythroid Progenitor
Me	Methylated
MEP50	Methylosome Protein 50
MkP	Megakaryocyte Progenitor
Mk	Megakaryocyte
MF	Myelofibrosis
MPL	Myeloproliferative leukaemia virus oncogene, also known as thrombopoietin receptor
MPN	Myeloproliferative neoplasm
MID	Multiplex Identification adaptors

MIG	MSCV-IRES-GFP plasmid/vector
miR	micro-RNA
Mir340	*MicroRNA 340*
mTOR	Mammalian target of rapamycin
NUSE Plot	Normalised Unscaled Standard Error plot
$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	reduced $NADP^+$
NGS	Next generation sequencing
Ndrg1	*n-Myc downstream regulated gene, 1*
PAX5	Paired box 5
PBMNC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
Ph	Philadelphia chromosome (*BCR-ABL* translocation)
ph	phosphorylated/phosphorylation
PI3K	Phosphoinositide-3-kinase
PHF6	Plant Homeodomain Finger, 6
PIAS	Protein Inhibitors of Activated STATs
Plekho2	*Pleckstrin homology domain containing family O member 2*
PMN	Polymorphonuclear
PMF	Primary myelofibrosis
PolyPhen-2	polymorphism phenotyping v2
PRC2	Polycomb Repressive Complex 2
PRMT	Peptidyl arginine methyltransferase
PRMT5	Peptidyl arginine methyltransferase, 5

PSG	Penicillin Streptomycin Glutamine
PTM	Post translational modification
PU.1	Spleen focus-forming virus (SFFV) proviral integration oncogene
PV	Polycythemia vera
QPCR	Quantitative PCR
p.R4C	Arginine 4 Cysteine, *HIST1H4C*:*c.10C>T, p.R4C*
p.R56Q	Arginine 56 Glutamine, *HIST1H4C*:*c.167C>T, p.R56Q*
RARA	Retinoic Acid Receptor, Alpha
RBAP48	Retinoblastoma binding protein 4 or RBBP4
RUNX1	Runt-related transcription factor 1
RBC	Red Blood Cell
S-phase	Synthesis-phase in DNA replication
S1	Serine 1
SAM	*S*-adenosylmethionine
SAH	*S*-adenosylhomocysteine
SCF	Stem Cell Factor
SCL	Stem Cell Leukaemia
SFFV	Spleen focus-forming virus
SNO	Spindle-shaped N-cadherin^{+}CD45^{-}Osteoblastic cell
SNP	Single nucleotide polymorphism
SLBP	Stem Loop Binding Protein
SIFT	Sorting-Intolerant-From-Tolerant

STHSC	Short-term repopulating haematopoietic stem cell
STAT	Signal Transducer and Activator of Transcription
SUZ12	Suppressor of zeste 12 homologue
T-ALL	T-cell acute lymphoblastic leukaemia
TDG	Thymine DNA glycosylase
TET2	Ten Eleven Translocation, 2
TF	Transcription factor
TPO	Thrombopoietin
TYK2	Tyrosine Kinase 2
U7snRNA	U7 small nuclear RNA
Ub	Ubiquitinated
UniProt	Universal Protein Resource database
U-MPN	MPN unclassifiable
UCSC genome browser	University of California Santa Cruz Genome Bioinformatics, Human (*Homo sapiens*) Genome Browser Gateway
5xUAS	5x *U*pstream *A*ctivation *S*equences
VP16-TD	VP16- Transcriptional activation Domain
WHO	World Health Organisation
WT	wildtype
W	Tryptophan (amino acid)
WSxWS	Tryptophan Serine, non-conserved residue, Tryptophan Serine