

**Identification and Characterisation of Genetic Lesions
that Predispose to and Gene Expression Patterns that
Contribute to Myeloid Malignancies**

PARVATHY VENUGOPAL

Bachelor of Science (Zoology with Medical Microbiology)
Master of Science (Regenerative Medicine)

A thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences
Discipline of Genetics and Evolution

Faculty of Sciences
The University of Adelaide, South Australia

March 2016

Table of Contents

Abstract.....	1
STATEMENT.....	5
Acknowledgements.....	7
List of Figures.....	9
List of Tables.....	11
Abbreviations.....	12
Chapter 1. Review of Literature.....	17
1.1 Haematopoiesis.....	17
1.1.1 Primitive Haematopoiesis.....	18
1.1.2 Regulators of Haematopoiesis.....	18
1.2 Myeloid Malignancies.....	19
1.2.1 Acute Myeloid Leukaemia.....	20
1.2.1.1 FAB Classification Scheme for AML.....	20
1.2.1.2 World Health Organisation (WHO) Classification Scheme for AML.....	21
1.2.2 Myelodysplastic Syndromes.....	23
1.3. Inherited Predisposition to Haematopoietic Malignancies.....	24
1.3.1. Familial Leukaemia.....	24
1.3.2. Syndromic Predisposition to Haematopoietic Malignancies.....	25
1.3.2.1 Inherited Bone Marrow Failure Syndromes.....	25
1.3.2.3 Tumour Suppressor Gene Syndromes.....	25
1.4 GATA2 is an MDS/AML Predisposition Gene.....	26
1.4.1 Structure and Properties of GATA2.....	26
1.4.2 GATA2 in Haematopoiesis.....	27
1.4.2.1 GATA2 in Erythroid Differentiation.....	27
1.4.2.2 GATA2 in Mast Cell Differentiation.....	27
1.4.3 Aberrant GATA2 in MDS/AML.....	28
1.4.4 Other Manifestations of GATA2 Mutations/Deletions.....	28

1.4.5 Somatic <i>GATA2</i> mutations in AML	30
1.5 Gene Expression and Mutations in MDS and AML	31
1.6 Ribosomopathies	33
1.6.1 Ribosome Dysfunction and Bone Marrow Failure	34
1.6.1.1 Diamond Blackfan Anaemia	34
1.6.1.2 Shwachman-Diamond Syndrome	35
1.7 Aim of the study:	37
Chapter 2. Materials and Methods	38
2.1 Introduction	38
2.2 General methods.....	39
2.2.1 DNA Isolation.....	39
2.2.2 Total RNA Isolation.....	39
2.2.3 Routine PCR	39
2.2.4 Tissue Culture	39
2.2.5 Lipofectamine-Mediated Transfection	40
2.2.6 Whole Cell Lysate Preparation	40
2.3 Chapter 3 Methods	40
2.3.1 Animal Handling and LSK Extraction.....	40
2.3.2 Generation of Retrovirus	41
2.3.3 Retrovirus Transduction and Colony Forming Unit Assay	41
2.3.4 Bone marrow transplantation experiment.....	41
2.3.5 <i>In vitro</i> mast cell differentiation	42
2.3.6 Detection of IL-6 and TNF secretion.....	42
2.3.7 β -hexosaminidase release assay.....	42
2.4 Chapter 4 Methods	42
2.4.1 AML Patient Cohort	42
2.4.2 RNA Isolation, cDNA Preparation and Specific Target Amplification	43
2.4.3 96.96 Dynamic Arrays	43

2.4.4 Data Output and Normalisation.....	44
2.4.5 Cohort Characteristics and Correlation Analyses.....	44
2.4.6 Kaplan Meier and Random forest analysis.....	44
2.5 Chapter 5 Methods	45
2.5.1 Animal Handling and LSK Extraction	45
2.5.2 Generation of Retrovirus	46
2.5.1 Retrovirus Transduction and Colony Forming Unit Assay	46
2.5.2 Transplantation:	46
2.6 Chapter 6 Methods	47
2.7 Media and Solutions.....	47
Chapter 3. Clinically important driver mutations in GATA2 zinc finger 2 display functional diversity and tendency for specification of myeloid malignancy subtypes, immunodeficiency disorders and lymphedema.....	50
3.1 Introduction	50
3.2 Statement of Authorship	51
3.3 Manuscript.....	54
3.4 Animal model used in the manuscript.....	88
3.5 Effect of <i>Gata2</i> haploinsufficiency on mast cell differentiation.....	88
Chapter 4: Gene Expression Patterns Improve Existing Prognostic Classification Strategies in Acute Myeloid Leukaemia	93
4.1. Introduction	93
4.2. Results:.....	95
4.2.1 Cohort Characteristics.....	95
4.2.1.1 Mutational Profile	96
4.2.1.2 Quality Control and Data Normalisation	98
4.2.1.3 Gene Expression patterns associated with <i>NPM1</i> mutations.....	99
4.2.2 Correlation of Gene Expression Patterns	102
4.2.2.1 Epigenetic modifier genes are co-regulated across AML.....	102

4.2.2.2	<i>ERG</i> and <i>GATA2</i> expression is positively correlated in AMLs	103
4.2.2.3	<i>EVI2A</i> is consistently upregulated in <i>RAS</i> mutant AML	104
4.2.3.	Random Forest Analysis for Survival Prediction	105
4.2.4.	Kaplan Meier analyses on the LBI Classification Tree	108
4.2.5.	Comparison with ELN Stratification Scheme	109
4.2.6.	Development and Validation of a combined ELN/LBI Classification Scheme	110
4.3.	Discussion	111
Chapter 5. Investigation of potential synthetic lethal interactions between oncogenic RAS and GATA2 in haematopoietic malignancy.....		
5.1	Introduction	116
5.2	Results	118
5.2.1	Selective association of <i>RAS</i> mutations with <i>GATA2</i> low AML	118
5.2.2	<i>NRAS</i> G12D mice develop haematological disease.....	119
5.2.3	<i>NRAS</i> G12D Transduced Lineage Depleted Bone Marrow Cells have <i>in vitro</i> Impaired Differentiation	127
5.2.4	Do <i>NRAS</i> G12D and <i>GATA2</i> cooperate in leukaemogenesis?	128
5.3	Discussion	130
5.4	Recommendations for future work.....	132
Chapter 6. Self-reverting Mutations Autocorrect the Clinical Phenotype in a Diamond Blackfan Anaemia Patient		
6.1	Introduction:	134
6.2	Statement of Authorship.....	135
6.3	Manuscript.....	138
Chapter 7. Final Discussion.....		
7.1	Functional relevance of <i>GATA2</i> ZF2 mutations.....	150
7.2	<i>GATA2</i> haploinsufficiency does not affect <i>in vitro</i> mast cell differentiation.....	151
7.3	Gene expression patterns used in conjunction with current risk stratification strategies can improve risk prediction at the time of AML diagnosis.....	151

7.4 UPD for correction of a gene mutation is a mechanism for spontaneous remission in Diamond Blackfan Anaemia	151
7.5 Conclusion.....	152
Appendix I. Splice factor mutations and alternative splicing as drivers of haematopoietic malignancy.....	153
Appendix II. Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukaemia with concurrent thrombocytopenia	176
Appendix III: Summary of GATA2 mutations and associated disease phenotypes.	180
Appendix IV. Supplementary information for Chapter 3.	188
Appendix V. Antibodies Used for Animal Work	219
Appendix VI. Description of genes whose expression was studied in AML samples in Chapter 4.	220
Appendix VII: Sequences of primers used for q-RTPCR using Fluidigm system in Chapter 4	227
Appendix VIII. Ranking of genes based on random forest analysis for different survival outcomes.....	233
Appendix IX. Endpoints and definitions of survival values.....	236
Appendix X. Supplementary Methods and Figures for Chapter 6.	237
References.....	247

Abstract

Acute Myeloid Leukaemia (AML) is a heterogeneous disease caused by multiple genetic lesions. Our laboratory focuses on understanding the genetics of both inherited and acquired haematopoietic malignancies. In this thesis, I have investigated both inherited and acquired genetic changes that contribute to myeloid malignancies.

One of the key factors regulating haematopoiesis is GATA2, a zinc finger transcription factor. Germline mutations in *GATA2* have been associated with several clinical phenotypes such as myelodysplastic syndrome (MDS)/AML, immunodeficiency disorders (MonoMAC syndrome, DCML deficiency, congenital neutropenia, NK cell deficiency, aplastic anaemia) and Emberger syndrome. Moreover, several somatic mutations in *GATA2* have been reported in MDS/AML. Intriguingly, missense somatic and germline mutations reported to date are mutually exclusive, and several clinical phenotypes are associated with specific mutations. We generated a zinc finger 2 (ZF2) mutant allelic series representing a range of clinical phenotypes to investigate how each mutation effects transactivation, DNA binding, protein structure, protein partner interactions and *in vitro* differentiation. Specific GATA2 mutations perturb the interactions and functions in distinct ways that are beginning to explain differences in observed clinical phenotypes.

We performed gene expression analysis of 91 selected MDS/AML genes, including *GATA2*, on 166 well annotated primary AML samples, bone marrow mononuclear cells (BMMNC) and CD34 controls. Correlation analyses of *GATA2* expression levels with expression of other genes and other mutational and clinical data, was performed to help identify genetic aberrations that cooperate with abnormal levels of *GATA2* in AML.

Statistical correlations of expression levels of various other genes with outcome and mutation status were also identified.

One such correlation was reduced *GATA2* expression with oncogenic RAS mutations. A pilot study was carried out to evaluate and optimise an *NRAS* G12D-induced leukaemia model. All mice transplanted with mutant *NRAS* G12D rapidly developed haematopoietic disease post-transplantation whereas the control group did not. Based on these pilot studies, we have initiated transplantation experiments in a conditional *GATA2* knockout model to investigate the requirement of *GATA2* in *NRAS* G12D induced myeloid disease. Recipient mice continue to be monitored, but are yet to develop disease.

We also identified gene expression patterns of prognostic significance in AML and narrowed down a combination of three genes that are highly predictive of outcome. We devised a strategy integrating these genes into currently used risk stratification strategies and significantly improved risk stratification of AML patients at diagnosis.

Among syndromes that predispose to MDS/AML, is Diamond Blackfan Anaemia (DBA), a congenital disorder characterised by red blood cell deficiency. The underlying genetic cause of DBA in a child was identified using whole genome sequencing (WGS), targeted massively parallel sequencing (MPS) and high density SNP array. A complex scenario of germline and somatic aberrations were identified in two genetic loci that helped to explain the clinical features seen in the patient and the progression of this disease. These have led to the discovery of a mechanism by which spontaneous remissions occur in DBA patients.

Together, these studies have given us valuable insights into malignant myeloid disease biology and offer potential applications in improving therapeutic approaches in AML patients.

STATEMENT

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 and, to the best of my knowledge and belief, 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 this 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. The author acknowledges 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.

Signed,

Parvathy Venugopal

31 March 2016

Acknowledgements

I wish to thank my supervisors Hamish Scott, Chris Hahn and Manuela Klingler-Hoffmann for providing guidance and support throughout my PhD. Thank you for taking time to give me valuable feedback on my thesis in spite of being busy with deadlines and a million other things. The opportunity to work in the Department of Genetics and Molecular Pathology has been invaluable and I have learnt so much in my time here.

I would like to express my sincere gratitude to Hamish for giving me this opportunity and always encouraging me to step out of my comfort zone. I think that is some of the best advice I will ever get. I admire his ability to see the big picture and his brutal honesty which gets the best out of you. He is an incredible individual who is passionate about making a difference in people's lives through his work.

I'm extremely grateful to Chris for his invaluable assistance and encouragement over the last three years. Chris personifies patience, always making time and explaining things as many times and in as many ways as it would take for one to understand. I cannot emphasise enough on the integral role he has played in my PhD. Thank you Chris!

My sincere thanks to Manuela for helping me get started in the lab and giving me that first push I needed to ease into a work environment/culture that was completely new to me. Her role in helping me settle in my early days was immense and I am grateful for that.

I would like to thank all the Scott lab members- Milly, Pete (special thanks for all his help with looking after the mice), Anna, Chan-Eng, Young, Bergithe, Bradley, Jesse and Alicia for their friendship, company and support in all things scientific and non-scientific. Thank you Milly, Alicia and Jesse, for ensuring I didn't go hungry by buying me lunch or chocolate in both solid and liquid form.

Special thanks to Richard D'Andrea, Anna Brown, Mahmoud Bassal, Justine Marum, Jim Manavis, John Finnie, Andreas Schreiber, Joel Geoghegan, David Lawrence, Jinghua (Frank) Feng, Michele Grimbaldston, Dave Yip and all our collaborators for their inputs and contributions to various aspects of my work. Thanks to the patients and their families for their willingness to participate in research without which my thesis wouldn't have been complete. Special thanks to Justine Marum and Ranjini Vaidyanathan for helping

me make sense of random forests and other complex statistical methods used in Chapter 4. All members of the Genetics and Molecular Pathology department, the ACRF genome facility, the animal facility and the Detmold facility deserve special mention for their help and their kind and encouraging words on some of my not-so-great days.

I am grateful to my parents for letting me chase my dreams. I would like to thank my father for being supportive of unconventional career (and partner) choices, my mother for being my super-cool pillar of support and my sister for backing me in all my decisions. I thank my family and friends for being proud of my achievements and I will be forever indebted to my grandfather for encouraging me to ask questions and introducing me to the fascinating world of science.

Most importantly, I would like to express my everlasting gratitude to my partner, Vishal, for his unconditional love and support over the last decade and for patiently putting up with me during these last few months. I fall short of words to express how valuable his contribution to this entire process has been. Thank you for keeping me sane!

List of Figures

Figure 1.1 Haematopoiesis.....	17
Figure 1.2 Complex regulatory network illustrating interactions between 18 core haematopoietic transcription factors.....	19
Figure 1.3 Schematic representation of GATA2 gene and protein and known interacting proteins.....	26
Figure 1.4 Germline missense mutations in <i>GATA2</i> cluster in ZF2 whereas somatic mutations cluster in ZF1.....	30
Figure 1.5 Recurrently mutated genes in AML can be separated into several functional categories.....	33
Figure 1.6 Key ribosomal components mutated in DBA and SDS.....	35
Figure 3.1 Genotypes of mice were confirmed in bone marrow by PCR across the floxed cassette.....	89
Figure 3.2 Flow cytometry analysis of cell surface expression of c-KIT and FcεR.....	89
Figure 3.3 Measurement of IL-6 secretion on stimulation with dinitro-phenyl-albumin (DNP) following IgE sensitisation.....	91
Figure 3.4 Measurement of TNF secretion on stimulation with DNP following IgE sensitisation.....	91
Figure 3.5 Measurement of β-hexosaminidase release.....	92
Figure 4.1 Oncoprint showing mutations identified in genes recurrently mutated in AML.....	97
Figure 4.2 Stability of housekeeping genes was assessed using GeNorm analysis.....	99
Figure 4.3 Distinct gene expression patterns associated with NPM1 mutations.....	101
Figure 4.4 Patients with NPM1 mutations have poorer disease free survival than those with wildtype NPM1.....	102
Figure 4.5 Pearson's correlation coefficient of 80 genes vs 80 genes.....	104
Figure 4.6 Preliminary classification tree for Overall survival in AML.....	107
Figure 4.7 Pruned rpart classification tree segregates patients into three risk groups....	108
Figure 4.8 LAPT4B, BSPRY and IDH1 are predictors of OS and DFS.....	109
Figure 4.9 Integrated Risk Classification Strategy.....	110
Figure 4.10 OS and DFS for LBI risk groups, ELN risk groups and the integrated ELN/LBI model.....	111

Figure 5.1 Ras mediated pathways implicated in oncogenesis.....	116
Figure 5.2 AML samples stratified on GATA2 expression levels.....	118
Figure 5.3 Successful engraftment in transplanted mice.....	120
Figure 5.4 NRAS G12D recipient mice displayed significantly lower RBC counts and haematocrit than EV control mice.....	122
Figure 5.5 Abdominal/thoracic cavity of EV control mouse and NRAS G12D diseased mouse.....	122
Figure 5.6 Diseased NRAS G12D mice exhibited marked splenomegaly.....	123
Figure 5.7 Spleen of NRAS G12D diseased mouse displays disrupted architecture and abnormal morphology.....	124
Figure 5.8 Diseased NRAS G12D marrow contained cells of abnormal morphology...	124
Figure 5.9 Liver of NRAS G12D diseased mice shows periportal infiltration.....	125
Figure 5.10 Extensive tumour infiltration into the renal cortex in NRAS G12D diseased mice.....	125
Figure 5.11 Tumours originate from transplanted GFP positive cells.....	126
Figure 5.12 NRAS G12D recipient mice have significantly shorter survival times than EV recipients.....	126
Figure 5.13 NRAS G12D impacts haematopoietic differentiation.....	127
Figure 5.14 Timeline for NRAS G12D - GATA2 cooperation experiment.....	128
Figure 5.15 Percentage of GFP positive cells in peripheral blood decrease with time...	129
Figure 5.16 Haematological parameters rise to normal ranges in weeks following transplantation.....	129
Figure 5.17 NRAS G12D alters lineage commitment in haematopoietic differentiation.....	131

List of Tables

Table 1.1 French-American-British Classification Scheme for AML.....	20
Table 1.2 WHO classification scheme for MDS and AML.....	21
Table 1.3 FAB Classification for MDS.....	23
Table 1.4 Ribosomopathies with a predisposition to MDS/AML.....	36
Table 3.1 Complete blood pictures of donor mice prior to bone marrow harvest.....	90
Table 3.2 Cells recovery following <i>in vitro</i> mast cell differentiation of wildtype and <i>Gata2</i> heterozygous bone marrow cells.....	90
Table 4.1 Standardised reporting for correlation of cytogenetic and molecular genetic data in AML with clinical data proposed by European LeukemiaNet.....	94
Table 4.2 Number of cases in each of the FAB subtypes.....	95
Table 4.3. Glossary.....	105
Table 4.4 Top ten genes for each survival outcome.....	106
Table 5.1 Weights of spleens and livers from transplant recipient mice from pilot experiment.....	123

Abbreviations

ADBA	Australian Diamond Blackfan Anaemia
AGM	Aorta-gonad mesonephros
ALT	Alanine transaminase
AML	Acute Myeloid Leukaemia
APS	Ammonium persulfate
BFU-E	Burst forming unit-erythroid
BMCMCs	Bone marrow-cultured mast cells
BMMNC	Bone marrow mononuclear cell
BMT	Bone marrow transplant
BSA	Bovine serum albumin
CD	Circular dichroism
CFU-G	Colony forming unit-granulocyte
CFU-GEMM	Colony forming unit-granulocyte, erythroid, macrophage, megakaryocyte
CFU-GM	Colony forming unit-granulocyte, monocyte
CFU-M	Colony forming unit-monocyte
CML-BC	Chronic myeloid leukemia in blast crisis
CMML	Chronic myelomonocytic leukaemia
CMV	Cytomegalovirus
CN	Cytogenetically normal
Co-IP	Co-immunoprecipitation
DBA	Diamond Blackfan Anaemia
DCML	Dendritic cell, monocyte, B and NK lymphoid
DEL	Deletion
DEPC	Diethylpyrocarbonate
DFS	Disease Free Survival
DNP	Dinitro-phenyl-albumin
DSS	Disease Specific Survival
EDTA	Ethylenediaminetetraacetic acid
EFS	Event Free Survival
ELN	European LeukaemiaNet
EMSA	Electromobility shift assay

EV	Empty vector
FAB	French-American-British
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum
G-CSF	Granulocyte-colony stimulating factor
GFP	Green Fluorescent Protein
GMP	Granulocyte Monocyte Progenitor
GOF	Gain-of-function
H&E	Haematoxylin and Eosin
HEK	Human Embryonic Kidney
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
Het	Heterozygous
HSC	Haematopoietic Stem Cell
HSPC	Hematopoietic stem and progenitor cells
ID	Immunodeficiency disorders
IFC	Integrated Fluidic Circuit
IL-6	Interleukin 6
ITC	Isothermal titration calorimetry
JMML	Juvenile myelomonocytic leukaemia
LB	Luria Broth
LBI	<i>LAPTM4B/BSPRY/IDH1</i>
LCL	Lymphoblastic cell line
LOF	Loss-of-function
LOH	Loss of heterozygosity
LSK	$\text{Lin}^- \text{Sca1}^+ \text{c-Kit}^+$
MDS	Myelodysplastic Syndrome
MLL	Mixed-lineage leukemia
MNC	Mononuclear cell
MOI	Multiplicity of infection
MonoMAC	Monocytopenia with <i>Mycobacterium avium</i> complex
MPD	Myeloproliferative disease
MPS	Massively Parallel Sequencing
MRI	Magnetic resonance imaging
MSCV	Murine stem cell virus

Mut	Mutant
NGS	Next Generation Sequencing
NK	Natural Killer
NLS	Nuclear Localisation Signal
NMR	Nuclear magnetic resonance
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
PAGE	Polyacrylamide gel electrophoresis
PBMNC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PBS-T	PBS-Tween 20
p-NAG	p-Nitrophenyl-N-Acetyl- β -D-Glucosaminide
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RA	Refractory anaemia
RAEB	Refractory anaemia with excess blasts
RAEB-T	Refractory anaemia with excess blasts in transformation
RARS	Refractory anaemia with ring sideroblasts
rfsrc	Random forests for survival, regression and classification
RIPA	Radio-Immunoprecipitation Assay
RP	Ribosomal proteins
SACRB	South Australian Cancer Research Biobank
SCF	Stem cell factor
SDS	Shwachman Diamond Syndrome
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
STA	Specific target amplification
TALL	T cell lymphoblastic leukaemia
t-AML	Therapy related AML
TGE	Tris-glycine-EDTA
TNF	Tumour Necrosis Factor
TPO	Thrombopoietin
UPD	Uniparental disomy
VIMP	Variable importance
WB	Western blot

WEMSA	Western blotting-electromobility shift assay
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WT	Wild type
ZF1	Zinc finger 1
ZF2	