

A Genomic Approach Towards an Understanding of Clonal Evolution and Disease Progression in Multiple Myeloma

Ankit Kumar Dutta

BSc (Mol Biol), BHlthSc (Hons)

Myeloma Research Laboratory
Discipline of Physiology
Adelaide Medical School
Faculty of Health and Medical Sciences
The University of Adelaide

&

Cancer Theme
South Australian Health & Medical Research Institute
(SAHMRI)



Supervisors

Prof. Andrew C. W. Zannettino

Dr. Duncan R. Hewett

A thesis submitted for the degree of Doctor of Philosophy
August 2018

TABLE OF CONTENTS

| | |
|---|-------------|
| TABLE OF CONTENTS | i |
| ABSTRACT | v |
| DECLARATION | vii |
| ACKNOWLEDGEMENTS | viii |
| LIST OF ABBREVIATIONS | ix |
| PUBLICATIONS | xii |
| Scientific Manuscripts | xii |
| Conference Proceedings | xii |
| Awards | xiv |
| | |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1 Biology of Multiple Myeloma | 3 |
| 1.2 Development of Multiple Myeloma | 4 |
| 1.3 A New Paradigm In Multiple Myeloma Development..... | 9 |
| 1.3.1 <i>Linear Model of Tumour Evolution</i> | 10 |
| 1.3.2 <i>Expansionist Model of Tumour Evolution</i> | 13 |
| 1.3.3 <i>Intraclonal Heterogeneity Model of Tumour Evolution</i> | 17 |
| 1.4 Single Cell Analysis of Multiple Myeloma | 21 |
| 1.5 Limitations of Published Studies in Multiple Myeloma..... | 22 |
| 1.6 Epigenetics in Multiple Myeloma Development..... | 23 |
| 1.7 Current Therapies and Impacts in Multiple Myeloma..... | 26 |
| 1.8 Summary and Objectives | 31 |
| 1.9 References..... | 32 |
| | |
| CHAPTER 2: SUBCLONAL EVOLUTION IN DISEASE PROGRESSION FROM MGUS/SMM TO MULTIPLE MYELOMA IS CHARACTERISED BY CLONAL STABILITY | 41 |
| 2.1 Abstract..... | 43 |
| 2.2 Introduction..... | 44 |
| 2.3 Materials & Methods | 46 |
| 2.3.1 Clinical samples | 46 |
| 2.3.2 Cell sorting..... | 46 |
| 2.3.3 DNA isolation, QC and sequencing..... | 46 |
| 2.3.4 Analysis of data | 47 |
| 2.3.5 Data deposition | 47 |
| 2.3.6 Code Availability | 47 |
| 2.4 Results..... | 48 |

| | | |
|-----------|---|----|
| 2.4.1 | A changing spectrum of acquired mutations, not mutational load, is associated with MM progression | 48 |
| 2.4.2 | The subclonal architecture required for MM progression exists at MGUS/SMM diagnosis | 54 |
| 2.4.2.1 | <i>Subclonal tumour evolution in MGUS-MM patients</i> | 55 |
| 2.4.2.2 | <i>Subclonal tumour evolution in SMM-MM patients</i> | 58 |
| 2.5 | Discussion..... | 62 |
| 2.6 | Acknowledgements..... | 65 |
| 2.7 | Supplementary | 66 |
| 2.7.1 | Supplementary Methods | 66 |
| 2.7.1.1 | Whole Exome Sequencing..... | 66 |
| 2.7.1.2 | Analysis of Whole Exome Sequencing Data..... | 66 |
| 2.7.1.2.1 | Sequence alignment | 66 |
| 2.7.1.2.1 | Somatic variant calling | 66 |
| 2.7.1.3 | Tumour heterogeneity and subclonal evolution..... | 67 |
| 2.7.2 | Supplementary Appendix 1 | 68 |
| 2.7.2.1 | <i>Subclonal tumour evolution in MGUS-MM patients</i> | 68 |
| 2.7.2.2 | <i>Subclonal tumour evolution in SMM-MM patients</i> | 69 |
| 2.7.3 | Supplementary Figures | 71 |
| 2.8 | References..... | 79 |

CHAPTER 3: TRANSCRIPTOMIC AND DNA METHYLATION ANALYSES IN MGUS/SMM TO MM PROGRESSION 84

| | | |
|-------|--|-----|
| 3.1 | Abstract..... | 86 |
| 3.2 | Introduction..... | 87 |
| 3.3 | Materials & Methods | 89 |
| 3.3.1 | Clinical samples..... | 89 |
| 3.3.2 | Cell sorting..... | 89 |
| 3.3.3 | Nucleic acids isolation and QC..... | 89 |
| 3.3.4 | RNA sequencing and analysis | 90 |
| 3.3.5 | Whole Genome Bisulphite Sequencing and analysis | 90 |
| 3.4 | Results..... | 91 |
| 3.4.1 | The progression of MGUS/SMM to MM is associated with minimal variation in gene expression pattern..... | 91 |
| 3.4.2 | MGUS/SMM to MM progression is associated with specific expression changes | 96 |
| 3.4.3 | The progression of MGUS/SMM to MM is characterised by the maintenance of hypomethylation acquired at the asymptomatic disease stage..... | 102 |
| 3.4.4 | Gene ontology pathways associated with the transformation to MM | 112 |
| 3.4.5 | Allelic specific expression of genes associated with MGUS/SMM to MM progression..... | 115 |
| 3.5 | Discussion..... | 121 |
| 3.6 | Supplementary | 124 |
| 3.6.1 | Supplementary Figures..... | 124 |

| | | |
|-----|-----------------|-----|
| 3.7 | References..... | 134 |
|-----|-----------------|-----|

CHAPTER 4: *SP140*, A GENE RECURRENTLY MUTATED IN HUMAN MYELOMA, IS THE TARGET OF RNA EDITING IN THE 5TGM1 MURINE MYELOMA CELL LINE 137

| | | |
|---------|---|-----|
| 4.1 | Abstract..... | 139 |
| 4.2 | Introduction..... | 140 |
| 4.3 | Materials & Methods | 143 |
| 4.3.1 | Sequencing and analysis | 143 |
| 4.3.2 | Cell culture..... | 143 |
| 4.3.2.1 | Murine cell lines | 143 |
| 4.3.2.2 | Human MM cell lines | 144 |
| 4.3.3 | Nucleic acids isolation..... | 144 |
| 4.3.3.1 | DNA isolation..... | 144 |
| 4.3.3.2 | RNA isolation | 144 |
| 4.3.4 | Validation | 144 |
| 4.3.4.1 | Confirming SNVs in 5TGM1 cell line | 144 |
| 4.3.5 | Screening | 145 |
| 4.3.5.1 | Murine cell lines | 145 |
| 4.3.5.2 | Human MM cell lines | 145 |
| 4.3.6 | Generation of Apobec1 sgRNA expression vectors | 146 |
| 4.3.6.1 | Transfection of 5TGM1 cells..... | 146 |
| 4.3.6.2 | Mutation screening | 147 |
| 4.3.7 | Generation of Apobec3 sgRNA expression vectors | 147 |
| 4.3.7.1 | Viral packaging and 5TGM1 infection..... | 147 |
| 4.3.7.2 | Mutation screening | 148 |
| 4.3.8 | Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) | 148 |
| 4.4 | Results..... | 149 |
| 4.4.1 | <i>Sp140</i> exhibits high impact mutations in the transcriptome of 5TGM1 cells | 149 |
| 4.4.2 | <i>Sp140</i> is a target of RNA editing in 5TGM1 PCs | 150 |
| 4.4.3 | <i>SP140</i> is not a target of RNA editing in human MM PCs..... | 158 |
| 4.4.4 | Molecular mechanisms inducing RNA editing of <i>SP140</i> | 161 |
| 4.5 | Discussion..... | 174 |
| 4.6 | Supplementary | 177 |
| 4.6.1 | Supplementary Figures | 177 |
| 4.7 | References..... | 199 |

CHAPTER 5: DISCUSSION AND CONCLUSIONS 203

| | | |
|-----|----------------------------|-----|
| 5.1 | General Discussion | 204 |
| 5.2 | Towards A Cure for MM..... | 211 |
| 5.3 | Conclusion | 215 |
| 5.4 | References..... | 216 |

ABSTRACT

Multiple myeloma (MM) is a largely incurable haematological malignancy characterised by the aberrant proliferation of malignant plasma cells (PCs) in the bone marrow (BM). Next generation sequencing (NGS) studies have shown that MM patients display complex mutational landscapes involving intraclonal genetic heterogeneity. While intraclonal heterogeneity is now an established feature of MM, the genomic changes and tumour evolution associated with the transformation from the asymptomatic disease stages of Monoclonal Gammopathy of Undetermined Significance (MGUS) and Smouldering Multiple Myeloma (SMM) to MM remains unknown.

This thesis presents a unique assessment of the genomic architecture and subclonal evolution associated with the natural history of disease transformation, with the analyses of a rare cohort of paired BM samples from patients when first diagnosed with MGUS or SMM, who later went on to develop MM (n = 10). Whole exome sequencing (WES) and bioinformatic analyses identified that clonal heterogeneity was present at the asymptomatic MGUS/SMM stages of disease, with a changing spectrum of acquired mutations associated with transition to MM. Subclonality was observed at MGUS/SMM, with the presence of between 5 to 11 subclones. The progression to MM was characterised by a prevailing model of subclonal evolution defined by clonal stability, where the transformed PC subclones of MM were already present at the MGUS/SMM stage.

RNA sequencing (RNAseq) revealed that the patterns of expressed genes at MGUS/SMM to MM were found to be relatively homogeneous. Moreover, RNAseq revealed that mutant genes identified by WES were generally not expressed, expressed at low levels, with most genes showing wild-type expression. Analysis of the methylome was carried out using whole genome bisulphite sequencing (WGBS). Significant hypomethylation was observed in PCs recovered at all disease stages (MGUS, SMM and MM) compared to normal PCs. Interestingly, the degree of hypomethylation observed at MGUS was maintained with progression to SMM and MM stages.

In addition, the phenomenon of RNA editing in *SPI40*, a recurrently mutated gene in human MM patients, was investigated in the 5TGM1 MM PC line. A high impact C>T (ie. U) RNA editing change was identified in exon 2 of *Sp140*, resulting in an early STOP

codon, which was hypothesised to result in the formation of truncated Sp140 protein that may contribute to MM pathogenesis. In mouse cell lines, *Sp140* RNA editing was not restricted to the 5TGM1 cell line, but editing was not observed in any human MM PC lines. CRISPR-Cas9 mediated mutation of the mouse *Apobec1* and *Apobec3* genes, showed that neither of these cytidine deaminases were responsible for this RNA editing phenomenon.

These studies show that MGUS/SMM patients, that progress in a short time frame, appear to be sufficiently genetically complex to be on the threshold of transformation to MM. Furthermore, the intrinsic genomic complexity of MM is present at the asymptomatic stages of disease (MGUS and SMM), suggesting that extrinsic factors from the tumour microenvironment play an important role in mediating progression. Indeed, these studies suggest that early intervention at MGUS/SMM may be possible to prevent progression and result in durable cure for patients.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, 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 in my name, 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 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 Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed:



Ankit K. Dutta

August 2018

ACKNOWLEDGEMENTS

"You need people around you who want to help you along with your goals, because when you're at it alone, no matter how much passion you have, you can never make it, it's very very hard" – Daniel Brunoli

So here I am 5 seconds out from the end, reflecting back on an incredible 4 year PhD journey. More than just developing my scientific expertise, this has also been a huge personal growth experience. At the end of the day you realise it's not as easy as just getting it done. Success can never be achieved on your own, and I would like to extend my gratitude to everyone that has helped me to make it to this day (which at times I thought would never come haha).

Firstly to my amazing mentors Prof. Andrew Zannettino and Dr. Duncan Hewett for their scientific guidance, support and enthusiasm throughout my PhD. I know this project has been a new venture for all of us, but I can unequivocally say, "We did it". It has been a dream run with the world standard achievements we have reached (ASH and Leukemia journal). Thank you for your critical review in perfecting my thesis. I credit a huge portion of my success during this time to your encouragement and efforts to push me to excellence. Thank you for giving me a shot in 2012 as a young undergraduate, never once did I feel like you didn't believe in me over the years.

I would also like to thank Dr. J. Lynn Fink, Dr. John Grady, Dr. Chung Kok, Manuel Linares and Dr. Benjamin Mayne for sharing their computational biology expertise during the project. Their bioinformatics support and discussion had helped me to develop a solid understanding in this interdisciplinary research field.

Thank you to the funding bodies: Australian Government, Leukaemia Foundation of Australia and The University of Adelaide, who have supported my studentship with research and travel scholarships during my PhD.

A very special thank you to the PhD Eagles, Dr. Chee Man "CMC" Cheong, Dr. Krzysztof "Mr Muscles" Mrozik and Jia Qi "JiJi" Ng for their companionship during the program. Our Peter Rabbit coffee runs, lunches and mischief have made this invigorating.

Thank you to all the students of the Myeloma Research Laboratory and Mesenchymal Stem Cell Laboratory: Mara “Mars Bar” Zeissig, Kimberley “Kimbruleè” Clark, Natasha “Strong” Friend, Pawanrat “Quinny” Tangseefa, Khatora Opperman, Alanah “Pusheen” Bradey, Elizabeth “Libbee” Coulter, Chee “Chee 2” Ho H’ng, Clara “Mono-pole-y” Pribadi, Elyse Bell, Ana Klisuric and Tamara Scheffer. Our dinners, outings, quiz nights (damn 4th place!) and general banter have made this a very enjoyable experience. Also thank you to SAHMRI students Sanuja “SJ” Fernando, Dr. Kartini “Kart” Asari, Dr. Theodosia “Sissy” Charitou and Alex Janssan for their friendship and support.

Thank you also to all the post doctorate researchers and laboratory officers who have also provided their support and expertise during my time: Rosa Harmer, Vicki Wilczek, Sharon Paton, Dr. Stephen Fitter, Dr. Sally Martin. Dr. Kate Vandyke, Dr. Melissa Cantley, Dr. Jacqueline Noll, Dr. Agnes Arthur, Dr. Thao Nguyen, Dr. Bill “The Gift” Panagopoulos Prof. Stan Gronthos and Dr. Jim “Mr Cool” Cakouros.

Travelling to the USA to present at the ASH Annual Meeting and visit laboratories was a big highlight of my candidature, and I’d like to thank Jia, Bill and Mara for being a part of the experience of a lifetime (when science takes you places!).

Thank you to my friends Dr. Reuben Jacob and Dr. Lauren Giorgio for their encouragement during my PhD and also their enthusiasm in supporting my interest in the biotechnology industry.

I would like to thank my good friends Daniel and Leah Brunoli (and Miss P!), for their support and positive energy during my PhD research. You have been inspiring to me because of the hard work you put in building The Edit. Dan, thank you for the great brunch catch ups (go UR Caffe!), piano and watch days that have helped to relax and re-energise me. Maybe now I will have more time to practice the piano... haha

Thank you to the A-Team: Marie Anastasi (Ellul), Avi Saini and Anmol Saini for your motivation and encouragement to achieve my best. Marie, thank you for the coffee and lunch meets, and your words of advice with problem solving.

Thank you to all the brothers: Daniel “DJ Dan Dan” Nagy, Liguò “DJ CalcuZilla” Dai, Philip “DJ Pheel” Uroda, Dr. Matthew “Ducktor Mat” Emes, Raj Datta, Craig Aistrophe, Avi Saini, Anmol Saini, Suliman “Suli” Yagoub, Sam Gili and Adam “Mr. Italia” Fazzini that have always supported me throughout this time. Our catch-ups, coffees, lunches,

dinners, outings, shopping trips, training sessions, POWER marathons and banter have brought a lot of enjoyment in my life. Thank you also to my good friends Helena Trang, Rachel Bala and Anh-Hong Nguyen for supporting and believing in me.

I would also like to thank my piano teacher and mentor, Dr. John Phillips, for his guidance, support and friendship as I've grown up.

“Don't let school get in the way of your education, the best lessons in life don't come from a teacher, they come from experience” – Arvin Lal

Its funny, at the beginning I used to think completing a PhD was the final hurdle, to then get a stable job to excel my career in. Now I realise that this is only the beginning, the bare minimum of what is required for I want to contribute and who I want to become.

Finally, the biggest thank you goes to my parents for their love, support and guidance throughout the years. You have always motivated me to work harder and achieve more. Thank you for building a foundation from which I have been able to jump for success. I hope this year has made you proud, now there are three Dr.'s in the house!

(I made it! Everyone should address me as Dr. Dutta now...)

Thank you.

LIST OF ABBREVIATIONS

| | |
|-------------|--|
| <i>ActB</i> | β -actin gene (mouse) |
| Adar | Adenosine deaminases acting on RNA (mouse) |
| ADAR | Adenosine deaminases acting on RNA (human) |
| AID | Activation-induced cytidine deaminase |
| AMG | Asymptomatic monoclonal gammopathy |
| aPC | Aberrant plasma cell |
| Apobec | Apolipoprotein B mRNA editing enzyme, catalytic polypeptide like (mouse) |
| APOBEC | Apolipoprotein B mRNA editing enzyme, catalytic polypeptide like (human) |
| AS | Active site |
| ASCT | Autologous stem cell transplantation |
| BM | Bone marrow |
| BMPC | Bone marrow plasma cell |
| Bp | Base pair |
| Cas9 | CRISPR associated protein 9 |
| cDNA | Complementary DNA |
| cfDNA | Cell free DNA |
| CNV | Copy number variant |
| CPM | Counts per million of reads |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| CSR | Class switch recombination |
| CTC | Circulating tumour cell |
| DMEM | Dulbecco's Modified Eagle Medium |
| DMR | Differentially methylated region |
| DNMT | DNA methyltransferase |
| EMC | Erasmus University Medical Center |
| EV | Empty vector |
| FACS | Fluorescence-activated cell sorting |
| FCS | Fetal calf serum |
| FISH | Fluorescence in situ hybridisation |
| FMO | Fluorescence minus one |

| | |
|----------|---|
| FPKM | Fragments Per Kilobase of transcript per Million mapped reads |
| GATK | Genome Analysis Toolkit |
| gDNA | Genomic DNA |
| GEP | Gene expression profiling |
| GO | Gene ontology |
| HAT | Histone acetyl transferase |
| HDAC | Histone deacetylase |
| IFC | Integrated fluidic chip |
| Ig | Immunoglobulin |
| IGH | Immunoglobulin heavy locus |
| IMDM | Iscove's Modified Dulbecco's Media |
| KO | Knockout |
| MDE | Myeloma defining event |
| MGUS | Monoclonal Gammopathy of Undetermined Significance |
| MM | Multiple Myeloma |
| MPT | Melphalan-Prednisone-Thalidomide |
| MRD | Minimal residual disease |
| MRI | Magnetic resonance imaging |
| NGS | Next Generation Sequencing |
| NPC | Normal plasma cell |
| NS-SNV | Non-synonymous single nucleotide variant |
| OS | Overall survival |
| PC | Plasma cell |
| PCL | Plasma Cell Leukaemia |
| PFS | Progression free survival |
| PR | Partial response |
| Rd | Lenalidomide-Dexamethasone |
| RNAseq | RNA sequencing |
| RPMI | Roswell Park Memorial Institute |
| RT-qPCR | Reverse transcription-quantitative polymerase chain reaction |
| SACRB | South Australian Cancer Research Biobank |
| SCATTome | Single-cell analysis of targeted transcriptome |
| SHM | Somatic hypermutation |
| SMM | Smoldering Multiple Myeloma |
| SNP | Single nucleotide polymorphism |

| | |
|--------------|---|
| SNV | Single nucleotide variant |
| Sp140/SP140 | Speckled Protein 140 KDa |
| <i>Sp140</i> | Gene (mouse) |
| <i>SP140</i> | Gene (human) |
| Sp140 | Protein (mouse) |
| SP140 | Protein (human) |
| TTP | Time to progression |
| UAMS | University of Arkansas for Medical Sciences |
| VAF | Variant allele frequency |
| VCD | Bortezomib-Cyclophosphamide-Dexamethasone |
| VRD | Bortezomib-Lenalidomide-Dexamethasone |
| VTD | Bortezomib-Thalidomide-Dexamethasone |
| v/v | Volume per volume |
| WES | Whole Exome Sequencing |
| WGS | Whole Genome Sequencing |
| WGBS | Whole Genome Bisulphite Sequencing |

LIST OF PUBLICATIONS

Scientific Manuscripts

1. **Dutta, A.K.**, Hewett, D.R., Kok, C.H., Mrozik, K.M. & Zannettino, A.C.W. (2018) *SPI40*, a gene recurrently mutated in human myeloma, is the target of RNA editing in the 5TGM1 murine myeloma cell line. *Manuscript in preparation*.
2. **Dutta, A.K.**, Fink, J.L., Mayne, B.T., Harliwong, I., Morgan, G., Mullighan, C., To, L. B., Hewett, D.R. & Zannettino, A.C.W. (2018) Integrative analysis of differential DNA methylation and gene expression in longitudinal MGUS/SMM to MM progression. *Manuscript in preparation*.
3. **Dutta, A.K.**, Fink, J.L., Grady, J.P., Morgan, G.J, Mullighan, C.G, To, L. B., Hewett, D.R. & Zannettino, A.C.W. (2018) Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability. *Leukemia*. <https://doi.org/10.1038/s41375-018-0206-x>.
4. **Dutta, A. K.**, Hewett, D. R., Fink, J. L., Grady, J. P. & Zannettino, A. C. W. (2017) Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma. *British Journal of Haematology*, 178, 196-208.

Conference Proceedings

1. Australian Society for Medical Research (ASMR) SA Scientific Meeting. Adelaide, South Australia, Australia. June 2018.

Oral presentation: *Genomic Analysis of Paired Patient Samples Reveals Novel Insights into Clonal Heterogeneity and Disease Progression of MGUS/SMM to Multiple Myeloma*.
2. 59th American Society of Haematology (ASH) Annual Meeting & Exposition. Atlanta, Georgia, United States of America. December 2017.

Oral presentation: *Whole Exome Sequencing of Paired MGUS/SMM to MM Patients Reveals Novel Subclonal Tumour Evolution Models in Progression of MM*.

Dutta, A. K., Grady, J. P., Hewett, D. R., To, L. B., Fink, L., & Zannettino, A. C. (2017). Whole Exome Sequencing of Paired MGUS/SMM to MM Patients Reveals Novel Subclonal Tumour Evolution Models in Disease Progression of Multiple Myeloma. *Blood*, 130(Suppl 1), 391.

3. 11th Annual Florey International Postgraduate Research Conference. Adelaide, South Australia, Australia. September 2017.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v3.*

4. Australian Society for Medical Research (ASMR) SA Scientific Meeting. Adelaide, South Australia, Australia. June 2017.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v3.*

5. South Australian Health and Medical Research Institute (SAHMRI) Annual Scientific Meeting. Adelaide, South Australia, Australia. November 2016.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v2.*

6. 10th Annual Florey International Postgraduate Research Conference. Adelaide, South Australia, Australia. September 2016.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v2.*

7. Australian Society for Medical Research (ASMR) SA Scientific Meeting. Adelaide, South Australia, Australia. June 2016.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v2.*

8. Australian Mathematical Sciences Institute (AMSI) BioInfoSummer Symposium. Sydney, New South Wales, Australia. December 2015.

Poster and Oral presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v1.*

9. 2nd Annual EMBL Australia PhD Symposium. Melbourne, Victoria, Australia. November 2015.

Poster presentation: *What? Cancer is evolving?... A Genomic Approach Towards An Understanding of Clonal Heterogeneity in Myeloma. v1.*

Awards

2017 American Society of Haematology Abstract Achievement Award

2017 University of Adelaide, Adelaide Medical School Research Travel Award

2016 Australian Society for Medical Research SA Meeting Best PhD Poster Award

2015 Australian Mathematical Sciences Institute BioInfoSummer Travel Award

2015 Florey Medical Research Foundation PhD Cancer Research Grant

2015 Leukaemia Foundation of Australia National Research Program PhD Scholarship

2014 Australian Postgraduate Award (APA) Scholarship

“Peak performance is meditation in motion”

- *Greg Louganis*

Chapter 1

Introduction

Statement of Authorship

| | |
|---------------------|--|
| Title of Paper | Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma |
| Publication Status | <input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style |
| Publication Details | Dutta, A. K. , Hewett, D. R., Fink, J. L., Grady, J. P. & Zannettino, A. C. W. (2017) Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma. <i>British Journal of Haematology</i> , 178, 196-208. |

Principal Author

| | | | |
|--------------------------------------|--|------|---------------|
| Name of Principal Author (Candidate) | Ankit K.Dutta | | |
| Contribution to the Paper | Literature interpretation Manuscript development, writing and review | | |
| Overall percentage (%) | 80% | | |
| Certification: | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. | | |
| Signature | | Date | 31 / 7 / 2018 |

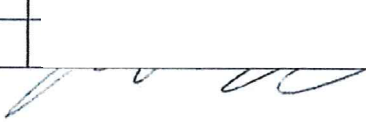
Co-Author Contributions


By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---------------------------|--|------|---------------|
| Name of Co-Author | Duncan R. Hewett | | |
| Contribution to the Paper | Supervision of work Manuscript development and review | | |
| Signature | | Date | 26 / 7 / 2018 |

| | | | |
|---------------------------|-------------------|------|------------|
| Name of Co-Author | J. Lynn Fink | | |
| Contribution to the Paper | Manuscript review | | |
| Signature | | Date | 2018-07-11 |

| | | | |
|---------------------------|---|------|---------|
| Name of Co-Author | John P. Grady | | |
| Contribution to the Paper | Manuscript review | | |
| Signature |  | Date | 11/7/18 |

| | | | |
|---------------------------|---|------|----------|
| Name of Co-Author | Andrew C. W. Zannettino | | |
| Contribution to the Paper | Supervision of work Manuscript development and review Project funding | | |
| Signature |  | Date | 24/7/18. |

Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma

Ankit K. Dutta^{1, 2}, Duncan R. Hewett^{1, 2}, J. Lynn Fink³, J. Grady³, Andrew C. W. Zannettino^{1, 2}

¹Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA 5000 Australia.

²Cancer Theme, South Australian Health & Medical Research Institute, Adelaide, SA 5000 Australia.

³The University of Queensland, Diamantina Institute, Woolloongabba, QLD 4102.

Keywords: Myeloma, Genomics, Tumour evolution, Intraclonal heterogeneity, Clinical impacts

Chapter 1 incorporates the review article published: **Dutta, A. K., Hewett, D. R., Fink, J. L., Grady, J. P. & Zannettino, A. C. W. (2017) Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma. *British Journal of Haematology*, 178, 196-208.**

1.1 Biology of Multiple Myeloma

Multiple Myeloma (MM) is a haematological malignancy resulting from the uncontrolled proliferation of malignant plasma cells (PCs) within the bone marrow (BM). MM is an age-dependent malignancy and the second most common haematological cancer (Singhal and Mehta 2006). Despite advances in therapy, MM remains a largely incurable disease with a median survival of 6 years. Notably, MM accounts for 20% of all deaths from haematological cancers and 2.1% of deaths from all cancers (NCI 2014, Zweegman, *et al* 2014). Strategies such as chemotherapy with combinations of newly developed drugs and autologous stem cell transplantation (ASCT) are used to manage the disease following diagnosis (Palumbo and Anderson 2011). Over the last ten years, an increase in the spectrum of available treatment options has seen a two-fold increase in patient survival (Palumbo and Anderson 2011).

The initiating oncogenic events that lead to the development of MM are thought to originate with the establishment of a founder precursor PC clone within the germinal centre of a peripheral lymphoid organ (Morgan, *et al* 2012). Healthy PCs are derived from B lymphocytes, which undergo rearrangement of their immunoglobulin (Ig) genes within the BM to generate precursor cells which express functional B cell receptors (surface immunoglobulins) (Shapiro-Shelef and Calame 2005). Following this, immature B cells migrate from the BM to the germinal centre of a peripheral lymphoid organ where they undergo affinity maturation in response to antigen exposure, specific for the B cell receptor. Somatic hypermutation (SHM) initiates point mutations in the hypervariable regions of the immunoglobulin heavy chain locus (*IGH*), resulting in the generation of highly specific Ig (Klein, *et al* 1998). Furthermore, class switch recombination (CSR) initiates antibody class switching, through deletional recombination of the Ig locus switch region, producing functional Ig of different isotype (Stavnezer 1996). Both molecular mechanisms of SHM and CSR are initiated by the expression of the enzyme activation-induced cytidine deaminase (AID), which generates double stranded DNA breaks in the Ig loci (Morgan, *et al* 2012). Maturation of B cells in the germinal centre leads to the development of memory B cells and plasmablasts, which are rapidly produced and short-lived effector cells of the early antibody response (Nutt, *et al* 2015). Terminal differentiation of plasmablasts leads to the development of long-lived antibody producing PCs. The PCs subsequently migrate to the bone marrow and/or lymph nodes and are involved in the body's immune response, producing Ig which serve to specifically bind to and destroy foreign antigens present in the body (Shapiro-Shelef and Calame 2005).

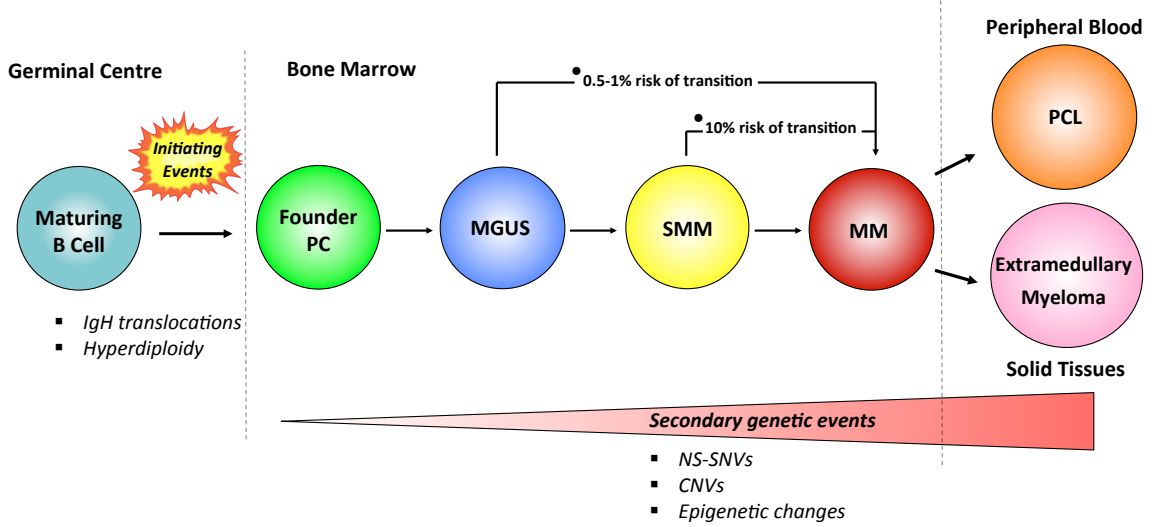
In MM however, the neoplastic PC clone, having sustained primary and potentially secondary mutations, migrates to the BM, where it expands in number and produces abundant, intact clonal Ig, known as Monoclonal (or “M”) protein or paraprotein (Wang and Young 2001). Abnormal PCs can also migrate to, and settle at other sites within the BM, where they expand in number to form myelomatous tumours at multiple sites (Ghobrial 2012). The myelomatous tumours disrupt normal homeostatic bone turnover, blood cell production and organ function leading to the clinical hallmarks of MM including elevated paraprotein in conjunction with suppressed immunoglobulin production and organ or tissue impairment. These clinical determinants of MM are known as the CRAB criteria (hypercalcaemia, renal insufficiency, anaemia and bone lesions) (Kyle, *et al* 2011).

1.2 Development of Multiple Myeloma

The development of MM is thought to involve a multistage transformational process, as a result of the acquisition of multiple genetic mutations that deregulate normal PC activity (*Fig. 1*) (Walker, *et al* 2014). Common initiating events in PCs include *IGH* translocations and hyperdiploidy, which result in the proliferation of monoclonal PCs, leading to development of a pre-cancerous, asymptomatic disease stage known as Monoclonal Gammopathy of Undetermined Clinical Significance (MGUS) (Bergsagel, *et al* 2005). MGUS is a slowly proliferative and relatively stable pre-myeloma stage in which paraprotein levels in the serum are < 30g/L and BM PC numbers account for < 10% of the nucleated cell count within the BM (International Myeloma Working Group 2003, Landgren, *et al* 2009). MGUS affects approximately 3% of the population aged over 50 years and 5% of people aged over 70 years (Kyle, *et al* 2006). Each year, MGUS patients have a 0.5-1% risk of progressing to MM (Rajkumar, *et al* 2014). For this reason, MGUS patients are currently monitored but remain untreated until their condition progresses to a symptomatic disease stage. However the risk of progression is not uniform in all cases, influenced by disease biology factors such as the type and concentration of M protein, serum free light chain ratio, bone marrow plasmacytosis, proportion of clonal plasma cells and presence of immunoparesis (Rajkumar, *et al* 2014). Risk of progression can be stratified by the Mayo Clinic model based on clonal plasma cell burden with M protein values and skewed free light-chain ratios (Rajkumar, *et al* 2005) or the Spanish study group multiparametric flow cytometry model (Perez-Persona, *et al* 2007). For MGUS the Mayo Clinic identifies 3 important risk factors of progression: non-IgG isotype, serum M-protein concentration > 1.5 g/dL and a skewed free light-chain ratio (< 0.26 or > 1.65) (Rajkumar, *et al* 2005). While the Spanish study group assesses the ratio of aberrant PC

(aPC) to normal BM PCs, where MGUS risk factors are a aPC/BMPC ratio > 95% and DNA aneuploidy (Perez-Persona, *et al* 2007). MGUS is followed by an intermediate asymptomatic Smouldering Multiple Myeloma (SMM) stage, where patients do not show evidence of Myeloma Defining Events (MDEs) or amyloidosis (Rajkumar, *et al* 2014). The SMM stage is known to have an annual risk of transition to MM of 10% in the first 5 years following diagnosis, 3% in the next 5 years and 1.5% in the subsequent years thereafter (Rajkumar 2016). For SMM, the Mayo Clinic identifies risk factors for progression being BM PCs > 10%, M-protein concentration > 3 g/dL and a skewed free light-chain ratio (< 0.125 or > 8) (Dispenzieri, *et al* 2008). While the Spanish study group identifies risk factors of a aPC/BMPC ratio > 95% and immunoparesis (Perez-Persona, *et al* 2007). The transition to MM is accompanied by an increased plasmacytosis with the presence of \geq 10% clonal BM PCs or the presence of biopsy-proven bony or extramedullary plasmacytomas, and the presence of at least one of the MDEs (Rajkumar, *et al* 2014). The MDEs include evidence of CRAB features or at least one of the biomarkers of malignancy; including either the presence of clonal BM PCs percentage \geq 60% or a serum free light chain ratio \geq 100 or > 1 focal lesion (from MRI studies) (Rajkumar, *et al* 2014). In the final stage of the transformation process, malignant PC clones may gain independence from the BM, enter the peripheral circulation, leading to Plasma Cell Leukaemia (PCL), or form tumours in soft tissue or organs, leading to Extramedullary Myeloma (Walker, *et al* 2014).

Figure 1. The development of MM is a multistage transformational process. Common initiating events of IgH translocations and hyperdiploidy during B cell development deregulate normal PC behaviour, leading to asymptomatic MGUS. Further mutational load leads to the intermediate stage of SMM before final transformation to symptomatic MM. MGUS can take > 25 years to progress, with patients having an annual 0.5-1% risk of transition to MM. Whereas SMM usually takes < 5 years to progress, with an annual 10% risk of progressing to MM in the first 5 years. However, the risk of progression is not uniform, and is dependant on disease biology factors as discussed above.



Using karyotyping and molecular cytogenetic techniques MM PCs have been classified under three key subtypes; hyperdiploid, non-hyperdiploid or unclassified (Fonseca, *et al* 2009). Hyperdiploid MM cases are characterised by trisomies and account for 50-60% of patients with MM (Sawyer, *et al* 2016). Non-hyperdiploid cases exhibit chromosomal translocations of the IgH locus, which are present in 45% of patients with MM (Mikhael, *et al* 2013). Interestingly IgH translocations arise through the normal process of B cell development. As previously described, during late B cell development AID induced CSR leads to double stranded breaks in the IgH locus, most of which are repaired locally. However, error in reassembly can lead to illegitimate recombination of the IgH locus with double stranded breaks elsewhere in the genome, resulting in aberrant chromosomal translocations of the IgH locus, where partner oncogenes are put into proximity of strong IGH enhancers leading to the hallmark event in MM (Gonzalez, *et al* 2007, Morgan, *et al* 2012). The three most common IgH abnormalities include the t(4;14) (chromosome bands p16q32), t(11;14) (chromosome bands q13q32) and t(14;16) (chromosome bands q32q23) (Fonseca, *et al* 2009) translocations. Non-hyperdiploid cases represent a more aggressive form of MM and are associated with poor prognosis (Bergsagel and Chesi 2013). Other cytogenetic abnormalities including monosomies of chromosome 14, and unaltered chromosome structure, are also present at a lower frequency of MM patients (Mikhael, *et al* 2013, Rajkumar 2016). Further studies have identified that this may represent a novel hyperhaploid subtype of high risk MM disease (Sawyer, *et al* 2016). Risk type of active MM can be stratified based on the specific genetic lesions harboured by a patient into high risk (t(14;16), t(14;20) and del 17p), intermediate risk (t(4;14, gain 1q, hypodiploidy and del 13) and standard risk (t(11;14), t(6;14) and trisomies) (Mikhael, *et al* 2013, Rajkumar 2016). However, the early common primary pathogenic events of hyperdiploidy and IgH translocations are believed to be insufficient to cause active symptomatic MM, as asymptomatic MGUS patients can harbour these abnormalities and show no clinical symptoms (Chapman, *et al* 2011, Fonseca, *et al* 2002). The acquisition of secondary genetic events of non-synonymous single nucleotide variations (NS-SNVs), copy number variations (CNVs) and epigenetic changes are also required to initiate MGUS and PC malignancy leading to MM disease progression (Chesi and Bergsagel 2011, Morgan, *et al* 2012). In addition to the oncogenic mutations occurring in the PC, there has been significant advancement in the understanding of an important role played by the BM microenvironment in supporting MM disease progression through proliferation, survival and drug resistance of MM PCs (Abe 2011, Lawson, *et al* 2015, Manier, *et al* 2012, Mitsiades, *et al* 2006, Noll, *et al* 2012).

1.3 A New Paradigm In Multiple Myeloma Development

New high-resolution Next Generation Sequencing (NGS) techniques represent an important advance in genomics, providing researchers with powerful tools for genetic analysis at single nucleotide resolution, enabling the identification of critical disease mutations and disease vulnerability. To date, no single gene mutation or combination of mutations have been identified as being common to all MMs at presentation (Weston-Bell, *et al* 2013). These findings suggest that multiple diverse genetic aberrations, and molecular pathways, are responsible for the onset of disease. Furthermore, the genetic abnormalities that are characteristic of each transformational stage of MM (MGUS, SMM and MM) have not been fully identified. This has been attributed to the low-resolution cytogenetic techniques previously used, which possess relatively low sensitivity of identification.

Of particular interest is the elucidation of how these key genetic aberrations contribute to tumour evolution in MM. Defining these critical disease mutations will provide insight into the possible mechanisms underlying disease progression, and identify key biomarkers of disease risk and provide novel therapeutic targets. Evaluation of the available cytogenetic and NGS studies reveals evidence for the existence of multiple tumour evolution models within MM. As a result three models of tumour evolution in MM are postulated:

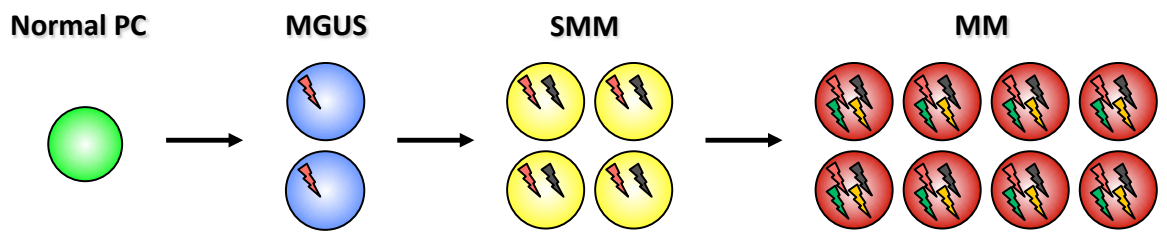
- Linear model
- Expansionist model
- Intraclonal heterogeneity model

1.3.1 Linear Model of Tumour Evolution

Classical cancer biology theory proposes a linear model of evolution, where tumours are derived from a unicellular origin with clonal growth pattern as a result of sequential accumulation of genetic mutations (Bahlis 2012).

As discussed earlier, the development of MM is considered to be a multistage transformational process where patients with MGUS progress through an intermediate SMM transition stage before developing symptomatic MM. Initial genome analyses by Chapman, *et al.*, identified MM tumour-specific mutations by comparing corresponding tumour to normal PCs using Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) techniques. An average of 35 NS-SNVs and 21 chromosomal rearrangements that disrupted protein coding regions were identified in MM (Chapman, *et al* 2011). Following this, Walker, *et al.*, carried out the first WGS study comparing a small number of premalignant MM stages (MGUS n = 4, SMM n = 4) and symptomatic MM stages (MM n = 26, PCL n = 2) to reveal tumour acquired NS-SNVs as a function of disease progression, where genetic complexity increases through the stages of MM. MGUS PCs were found to harbour a median of 13 NS-SNVs which increased to 28, 31 and 59 through SMM, MM and PCL, respectively (Walker, *et al* 2014). These findings are consistent with a linear model of tumour evolution in MM, where the sequential acquisition of NS-SNVs reaches a mutational threshold in the SMM PC, resulting in uncontrolled clonal proliferation leading to MM disease progression (*Fig. 2*).

Figure 2. The linear model of tumour evolution: The sequential acquisition and accumulation of multiple genetic mutations (represented by distinct bolts increasing through MGUS (red), SMM (red + black) and MM (red + black + green + yellow)) in PCs leads to MM disease progression.



1.3.2 Expansionist Model of Tumour Evolution

The expansionist model of tumour evolution infers that all necessary genetic mutations are present in a subset of PCs at the MGUS disease stage, and it is their subsequent expansion that leads to MM disease progression.

Studies using low-resolution molecular cytogenetic techniques have detected the same genetic mutations throughout all stages of MM transformation. Fluorescence *in situ* hybridisation (FISH) analysis has identified IgH translocations, chromosome 13q and 17p deletions and gain of 1q throughout all stages of MGUS, SMM and MM (Chiecchio, *et al* 2009, Lopez-Corral, *et al* 2011). Interestingly, the number of PCs harbouring these specific genetic abnormalities increased with disease progression, suggesting that clonal PC expansion was due to selective advantages (Avet-Loiseau, *et al* 1999, Chiecchio, *et al* 2009, Lopez-Corral, *et al* 2011). Notably also, sequential analysis of 5 patients (2 MGUS-MM, and 3 SMM-MM) revealed a higher proportion of PC harbouring the genetic abnormality that was observed at MGUS/SMM diagnosis upon progression to MM (Lopez-Corral, *et al* 2011). Further high resolution CNV analysis using high density single nucleotide polymorphism (SNP) arrays have indicated an increasing genetic complexity as disease progresses towards MM, with a progressive increase in the median number of CNVs through MGUS, SMM and MM from 5 to 7.5 to 12, respectively (Lopez-Corral, *et al* 2012). Frequent abnormalities observed in MM include gains on chromosome 1q, 3p, 6p, 9p, 11q, 19p, 19q and 21q together with deletions on chromosome 1p, 6q, 8p, 12p, 13q, 14q, 16q, 17p, 17q and 22q (Lopez-Corral, *et al* 2012, Walker, *et al* 2012). Alterations of 11q and 21q gains and, 16q and 22q deletions were previously viewed to be MM specific, however, it has been shown that these alterations are present within minor subclones at the MGUS stage (Lopez-Corral, *et al* 2012). Furthermore, WGS of sequential SMM-MM has demonstrated little difference in the median number of NS-SNVs present at both stages, with 28 to 31 respectively, reported (Walker, *et al* 2014). These findings suggest that the predominant PC clone may already be present at the SMM stage, and it is the outgrowth that leads to the initiation of MM disease progression. Interestingly, a recent WES and genotyping study of paired random bone marrow-focal lesion samples has revealed insights into spatial heterogeneity, where both similarities and differences of site specific SNVs and CNVs were observed to contribute to disease progression (Weinhold, *et al* 2015). In some paired samples up to 90% of variants were shared between both site samples, and a prominent pattern of the outgrowth of subclonal bone marrow sample CNVs as clonal events in focal lesions samples was observed (Weinhold, *et al* 2015).

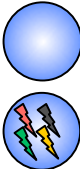
Taken together, these findings support an expansionist model of MM evolution, where a subpopulation of MGUS PCs harbour all the required genetic mutations, and their subsequent expansion results in MM disease progression (*Fig. 3*).

Figure 3. The expansionist model of tumour evolution: All necessary genetic mutations are present in a subpopulation of MGUS PCs (represented by all the distinct bolts (red + black + green + yellow) present in one), and it is their eventual expansion that leads to MM disease progression.

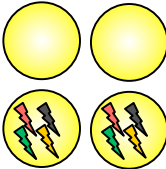
Normal PC



MGUS



SMM



MM



1.3.3 Intraclonal Heterogeneity Model of Tumour Evolution

The intraclonal heterogeneity model of tumour evolution posits a genetically heterogeneous clonal structure at the asymptomatic MGUS disease stage, where “Darwinian” competition occurs between distinct PC subclones in response to selective pressures, leading to the outgrowth of dominant PC subclones and subsequent MM disease progression (Greaves and Maley 2012, Nowell 1976).

Initial cytogenetic studies have indicated a genomic complexity where only a proportion of the clonal PC population carried any specific abnormalities at each stage of MM disease (Chiecchio, *et al* 2009, Lopez-Corral, *et al* 2011, Lopez-Corral, *et al* 2012). The advent of NGS techniques has allowed genetic analysis at single nucleotide resolution resulting in a higher power of detection of clonal architecture. A specific mutation being clonal or subclonal within a tumour sample can be determined by the proportion (of mutated reads) of total tumour cells that harbour the specific mutation, adjusted to any normal (non-tumour) contamination and the copy number of the locus. From this a phylogenetic tree of clonal/subclonal fractions can then be constructed to estimate the intraclonal evolution that may be occurring within the tumour. Initial sequencing studies on MM patients using WGS and WES have led to the discovery of frequent significantly mutated genes (“driver” genes), and the existence of a heterogeneous genetic landscape, where coexisting clones of differing genetic architecture arise during the evolution of MM (Bolli, *et al* 2014, Egan, *et al* 2012, Lohr, *et al* 2014, Walker, *et al* 2012). These “driver” genes are thought to influence clonal fitness (selective advantage and dominance) of malignant PC clones harbouring these mutants, driving MM disease progression in a branching manner.

While MGUS is known to be a benign disease stage that is far less genetically complex than MM, intraclonal heterogeneity has been demonstrated through all stages of MM, suggesting that disease progression is mediated via competition between subclones and outgrowth of the fittest of these subclones from the earliest stage of MM (Bolli, *et al* 2014, Egan, *et al* 2012, Lohr, *et al* 2014, Walker, *et al* 2012). Large cohort sequencing studies in MM have identified recurrently mutated genes associated with disease pathogenesis. The initial study of 38 MM tumours with matched normal genomes identified significant frequent somatic mutations occurring in MM, involving genes *KRAS*, *NRAS*, *TP53*, *BRAF* and *IRF4*, and six newly discovered cancer-associated genes; *CCND1*, *FAM46C*, *DIS3*, *PNRC1*, *ALOX12B*, *HLA-A* and *MAGED1* (Chapman, *et al* 2011). Further

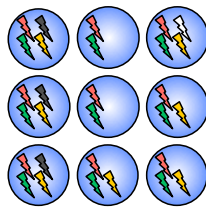
large cohort NGS studies performed by Lohr et. al. (n = 203 patients), Bolli et. al. (n = 67 patients) and Walker et. al. (n = 463 patients) have also identified the presence of mutations in *KRAS*, *NRAS*, *TP53*, *BRAF*, *FAM46C*, *DIS3*, *IRF4*, *TRAF3* and *CYLD* (Bolli, et al 2014, Lohr, et al 2014, Walker, et al 2015), of which *KRAS*, *NRAS*, *TP53*, *BRAF*, *FAM46C* and *DIS3* are now believed to be “driver” genes in MM disease progression. The RAS/MAPK pathway is frequently observed to be deregulated in MM, with recurrent mutations occurring in *KRAS*, *NRAS* and *BRAF* (Bolli, et al 2014, Lohr, et al 2014, Walker, et al 2015). Mutations in *KRAS* and *NRAS* tend to mainly occur with mutual exclusivity, however have been observed to co-exist in 2% of patients (Walker, et al 2015). Interestingly, these “driver” mutations have been identified as being present in clonal fractions in some patients and subclonal fractions in other patients, suggesting “driver” events may also arise during later stages of MM tumour evolution (Bolli, et al 2014, Lohr, et al 2014). Furthermore, although affecting the same pathway, they have been identified to occur subclonally or in a nested fashion, where one mutation clone is identified as a subclone of another (Lohr, et al 2014). While it would be expected that these subclonal populations would exhibit improved survival advantage, owing to the presence of mutations in multiple “driver” genes, they do not appear to have sufficient selective advantage for outgrowth to clonality (Lohr, et al 2014). This advancement in our understanding of the intraclonal heterogeneity in MM illustrates the consideration required towards therapeutic choices, where suboptimal outcome would be observed if a patient is treated for a “driver” mutation that exists only subclonally. These unique findings support an intraclonal heterogeneity model of MM evolution, where subclones harbour differing combinations of mutations, with the genetic landscape of subclones changing as MM disease progresses (*Fig. 4*).

Figure 4. The intraclonal heterogeneity model of tumour evolution: Distinct PC subclones carry different combinations of genetic mutations (represented by MGUS subpopulations harbouring distinct combinations of coloured bolts), with the dominance of subclones changing with MM disease progression. Predominant subclones harbouring “driver” mutations, conferring increased fitness potential, are able to survive the microenvironment pressures and contribute to MM disease progression (represented by the PC subpopulation harbouring distinct bolts red + green + yellow + white).

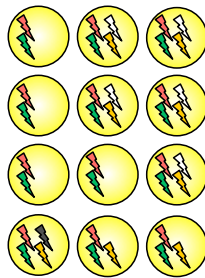
Normal PC



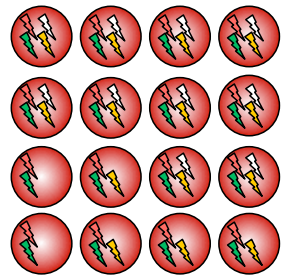
MGUS



SMM



MM



1.4 Single Cell Analysis of Multiple Myeloma

In more recent years, novel single cell analysis techniques have provided a sophisticated method for unravelling tumour genomics and evolution at a more detailed level. By unravelling clonal diversity and frequency of genetic mutations, single cell analysis is able to provide a greater understanding of the genomic complexity and clonal architecture that is present at the individual cell level within a bulk tumour. Current platforms available for single cell capture and subsequent genomic interrogation include Fluorescence Activated Cell Sorting (FACS) based automated single cell deposition and the Fluidigm C1 integrated fluidic chip (IFC) system (Fluidigm Corporation).

The existence of clonal heterogeneity presents a complex situation when analysing tumours. Distinct clonal populations of cells cannot be identified by conventional tissue average data analysis (Gundry, *et al* 2012). Furthermore, random low abundance mutations are currently inaccessible by standard high throughput sequencing approaches because they cannot be distinguished from sequencing errors (Gundry, *et al* 2012). At the genomic level, it is difficult to identify whether mutations are present in the same tumour cells or arise from distinct clonal populations of cells present within the bulk tumour (Yadav and De 2014). Consequently, isolated detection of dominant clonal populations of tumour cells could bias classification, prognosis and treatment of disease (Yadav and De 2014).

Application of single cell analysis techniques in MM supports the existence of multiple independent subclones within tumours. Single cell studies of t(11;14) MM have described the existence of 2 to 6 different clones, which are related by linear and branching phylogenies, highlighting the presence of intraclonal heterogeneity at MM diagnosis (Melchor, *et al* 2014). Melchor and colleagues demonstrated the existence of subclonal heterogeneity with parallel evolution of the RAS/MAPK pathway in distinct single cells, giving rise to divergent subclonal populations of mutant *NRAS* clone lineage and mutant *KRAS* clone lineage (Melchor, *et al* 2014). Furthermore, single cell genotyping has revealed a subclonal substructure in MM with 3 main clones harbouring *ATM* mutations where 2 subclones exhibited other mutually exclusive mutations (Walker, *et al* 2012). Walker and colleagues also identified intraclonal heterogeneity of mutations in “driver” genes *NRAS* (present in 32% and 96% of tumour cells) and *KRAS* (20% and 72% of tumour cells) (Walker, *et al* 2012). Interestingly, activating *KRAS* mutations were present as minor subclones in one case, observed in 20% and 48% of tumour cells (Walker, *et al* 2012). This suggests that subclones are continually at risk of developing “driver” mutations that can confer a growth and survival advantage leading to clonal dominance

over time (Walker, *et al* 2012). This advancement in the understanding of the intratumour heterogeneity in MM illustrates the consideration required towards therapy and its clinical implications. A recent study using *in vitro* modelling of MM cell lines that are bortezomib-sensitive or -resistant has generated a novel gene expression profiling (GEP) signature that can predict drug response to proteasome inhibitors (Stessman, *et al* 2013). The analysis of this GEP at the single cell level in pre-treated MM cell lines and drug naïve patient samples was able to identify pre-existing single cell sub populations that were resistant to proteasome inhibitors, demonstrating the possible requirement of therapies tailored towards subclonal populations within bulk tumours (Mitra, *et al* 2016). This has also been developed into the computational software package SCATTome (single-cell analysis of targeted transcriptome) that can predict probability of drug response of single cells based on the MM expression signature (Mitra, *et al* 2016).

Collectively, these NGS and single cell studies support clonal heterogeneity as a model of MM evolution where “Darwinian” competition between heterogeneous PC subclones initiates non-linear accumulation of mutations and outgrowth of dominant subclones driving MM disease progression.

1.5 Limitations of Published Studies in Multiple Myeloma

To date, a limited number of NGS studies of MM have been performed, with the first “Initial genome sequencing and analysis of multiple myeloma” carried out by Chapman *et al.* in 2011 (Chapman, *et al* 2011). The majority of these studies involved single time point studies of MM PC samples or have relied on the use of unpaired MGUS and MM PC samples. Consequently, these studies are only able to provide a detailed indication of the genetic landscape at the MM disease stage and an indirect indication of genetic mutations associated with MM disease progression.

WES analysis of serial samples ($n = 15$), collected at disease progression or relapse post-treatment with later time point samples collected at relapse/progression after further lines of treatment has revealed major patterns of tumour evolution associated with MM progression: 1. no change in clonal composition, 2. differential clonal response, with proportions of subclones changing over time, 3. linear evolution, with a new subclone emerging over time, and 4. branching evolution, with the emergence of new clones and decline of other clones (Bolli, *et al* 2014). Only one study has investigated paired SMM-MM samples ($n = 4$), finding that intraclonal heterogeneity is a typical feature in MM,

where SMM is a transition state where subclonal structure is evolving (Walker, *et al* 2014). Interestingly, only one unique NS-SNV was identified in MM, demonstrating that most of the required mutations for transition to symptomatic MM are already present. Additionally, comprehensive analysis of paired presentation and relapse/progression samples after combination high dose therapy (n = 33), using WES as well as gene expression and copy number profiles has revealed a majority of patients (n = 22) relapse through a branching tumour evolution pattern, with others showing linear evolution and differential response (Weinhold, *et al* 2016). Furthermore an increase in bi-allelic inactivation of tumour suppressor genes (mainly *TP53* and *FAM46C*) was associated with relapse, with double hit events in *TP53*: del(17p)/*TP53*^{mut} or del(17p)/*TP53*^{del} characterising a subgroup of patients with worse outcome after relapse (Weinhold, *et al* 2016). Further serial sample studies of this nature and/or sampling of different sites from the same patient are required for a greater understanding of genetic heterogeneity in MM disease progression.

A comprehensive approach would be to perform NGS studies of sequential paired MGUS-MM samples from the same patient. At present, no longitudinal progression studies of paired MGUS-MM PC samples have been performed. This is due, in large part, to the difficulty in establishing a cohort of patient samples from individuals when they were first diagnosed with MGUS, and subsequently symptomatic MM. Additionally due to the nature of premalignant disease, MGUS samples contain a low number of tumour PCs compared to normal healthy PCs, which results in high contamination and low yield of tumour PCs on isolation. The thorough genomic analysis of both bulk tumours and single cells on paired MGUS-MM patients represents a unique approach to identify key “driver” genes that are mutated and/or aberrantly expressed during disease progression. This approach would derive gene signatures indicative of pathways that are deregulated during the MGUS to MM transition. Furthermore, genomic data derived from such a study may allow for the identification of biomarkers that can predict which MGUS patient will progress to MM.

1.6 Epigenetics in Multiple Myeloma Development

Extensive studies of MM have been performed using cytogenetic and genomic approaches, however, relatively little is known about the role of epigenetics in driving MM disease progression. The rate of epigenetic change in cancers has been estimated to be orders of magnitude higher than that of genetic change occurring, and could be a major determinant of clonal evolution (Greaves and Maley 2012). The key epigenetic mechanisms known to alter and regulate gene expression are DNA methylation and histone modifications.

Changes in DNA methylation status, such as hypermethylation leading to gene inactivation and hypomethylation inducing genomic instability, have been observed in many cancer types (Kulis and Esteller 2010). Methylome studies comparing MM and its transition stages (MGUS, SMM and PCL) with normal PCs, have shown an increase in the number of differentially methylated gene loci associated with disease progression (Heuck, *et al* 2013, Salhia, *et al* 2010, Walker, *et al* 2011). The presence of genetic hypomethylation has been implicated as an important and early mechanism which drives MM disease progression (Salhia, *et al* 2010). Hypomethylation is associated with genomic instability and often coupled with altered chromatin structure, changes in DNA methyltransferase activity, loss of imprinting and an increased frequency of CNVs (Walker, *et al* 2011). Further studies have identified distinct profiles of epigenetic modifications linked with MM disease transition stages, with global hypomethylation occurring at MGUS-MM transition and hypermethylation occurring at MM-PCL transition (Walker, *et al* 2011). Specifically, DNA methyltransferase DNMT3A was observed to be underexpressed in MM due to the actions of hypermethylation, providing insight into the possible mechanism of hypomethylation observed in premalignant stages of MM (Heuck, *et al* 2013). Gene ontology enrichment analysis has revealed that hypomethylation in MM may favour bone invasion by increasing interactions with the bone marrow extracellular matrix, initiating adhesive interactions and the formation of lytic bone lesions (Salhia, *et al* 2010). Interestingly, highlighting the heterogeneity also observed at the methylation level, gene-specific hypermethylation has also been associated with MGUS-MM transition, with 77 affected genes having roles in developmental, cell cycle and transcriptional regulatory pathways identified (Walker, *et al* 2011). Further gene-specific hypermethylation has been identified at MM-PCL transition, with 1802 affected genes that are associated with cell signalling and cell adhesion pathways (Walker, *et al* 2011). Hypermethylation is proposed to deregulate adhesion of MM PCs to the bone marrow, facilitating independence of malignant PCs from the bone marrow niche, leading to PC egress from the marrow and entry into the peripheral circulation and development of PCL (Walker, *et al* 2011). Furthermore, a recent study has also identified hypermethylation of developmentally regulated B cell enhancers as a new type of epigenetic modification associated with the pathogenesis of MM (Agirre, *et al* 2015).

Initial methylome analysis had revealed that methylation status is not associated with specific genetic alterations (Salhia, *et al* 2010). In contrast, other studies have identified specific MM cytogenetic subgroups which exhibit individual methylation

profiles with a t(4;14) group, two separate t(11;14) groups and two separate hyperdiploid groups described (Walker, *et al* 2011). The t(4;14) cytogenetic subgroup displays frequent hypermethylation, akin to that observed in PCL, signifying that the methylation status may influence the aggressive clinical phenotype usually observed in both cases (Walker, *et al* 2011). However, the mechanisms that cause abnormal DNA methylation patterns in MM are yet to be determined (Dimopoulos, *et al* 2014).

Complex epigenetic mechanisms involving histone modifications are also reported to contribute to the pathogenesis of cancer (Plass, *et al* 2013). While DNA methylation is relatively constant, histone modifications are dynamic in nature. The main regulatory mechanism of the epigenome is histone acetylation, which is maintained by the interplay of two enzymes; histone acetyl transferases (HATs) catalysing the addition and histone deacetylases (HDACs) catalysing the removal of acetyl groups on lysine residues of histones.

In MM, HDAC inhibitors have been reported to have potent anti-myeloma activity both *in vitro* and *in vivo* (Smith, *et al* 2010). However, a clear understanding of which HDACs are expressed by MM PCs is lacking (Smith, *et al* 2010). Alternatively, histone methylation has been implicated to play an important role in MM development. For example, the high risk cytogenetic t(4;14) subgroup exhibits dysregulated expression of Fibroblast Growth Factor Receptor 3 (FGFR3) and Multiple Myeloma Set Domain (MMSET), a histone methyl transferase (Kalff and Spencer 2012). Universal expression of MMSET in t(4;14) MM suggests that MMSET is critical for myeloma pathogenesis and/or progression (Marango, *et al* 2008). Interestingly, further to its histone methyl transferase activity, MMSET has been identified to enhance the activity of HDACs (HDAC1 and HDAC2) and therefore plays a role in altering histone acetylation. MMSET has been shown to be beneficial to the survival of MM PCs as *in vitro* MMSET knockdown affects genes involved in key survival processes such as cell cycle, apoptosis and adhesion (Brito, *et al* 2009). These findings reveal insights into new epigenetic therapeutic strategies in MM. However, further laboratory and clinical studies are required in this emerging area.

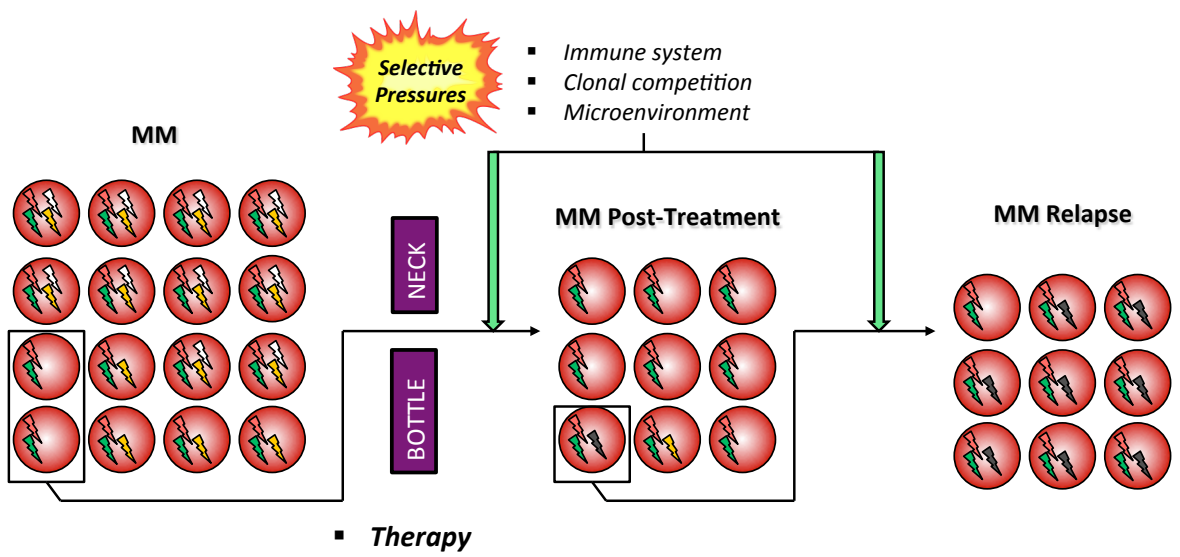
1.7 Current Therapies and Impacts in Multiple Myeloma

Although MM remains a largely incurable disease, advances in clinical research have produced effective treatment methods for disease control. Different strategies are employed for asymptomatic and symptomatic disease stages where MGUS/SMM stages require clinical monitoring while active MM is treated immediately and, in most cases, aggressively to induce disease remission (Palumbo and Anderson 2011). In the past decade, the survival of MM patients has more than doubled due to the introduction of new effective drug classes including immunomodulatory drugs (lenalidomide, thalidomide, pomalidomide, daratumumab and elotuzumab), histone deacetylase inhibitors (panobinostat) and proteasome inhibitors (bortezomib, carfilzomib and ixazomib), and the increased use of ASCT (Gertz 2014, MMRF 2015, Palumbo and Anderson 2011). Patient risk status is evaluated based on the factors of age/performance, renal function and presence or absence of high-risk genetic abnormalities, which all plays an important role in the treatment selection (Gertz 2014). Treatment comprises of three phases: induction, consolidation and maintenance. Current induction treatments for newly diagnosed patients, who are eligible for ASCT, include two-drug combination therapy of dexamethasone with lenalidomide (Rd), thalidomide or bortezomib. Three-drug combination treatments for newly diagnosed patients include the use of bortezomib with: bortezomib-cyclophosphamide-dexamethasone (VCD), bortezomib-thalidomide-dexamethasone (VTD) or bortezomib-lenalidomide-dexamethasone (VRD) (Rajkumar 2016). These combination treatment strategies are used to induce a complete response in patients before ASCT, followed by maintenance treatment with thalidomide or lenalidomide (Palumbo and Anderson 2011). For patients ineligible for transplantation, preferred treatments include melphalan based combinations of melphalan-prednisone-thalidomide (MPT), Rd (for elderly patients) or bortezomib based combinations VRD, VCD or VTD (Rajkumar 2016). In more recent years the treatment of relapsed MM patients (who have undergone previous lines of treatment) has seen promising results with the clinical trials and approval of monoclonal antibodies daratumumab (targeting cell surface protein CD38) and elotuzumab (targeting cell surface protein CD319) in 2015. Daratumumab has shown great efficacy in clinical trials with the ability to induce a deep response as both a monotherapy (Lonial, *et al* 2016, Usmani, *et al* 2016) and in combination therapy with bortezomib-dexamethosone (Palumbo, *et al* 2016) or lenalidomide-dexamethosone (Dimopoulos, *et al* 2016). Similarly, elotuzumab has shown efficacy in its clinical trial as a combination therapy with lenalidomide-dexamethosone, but does not show any single-agent activity

(Lonial, *et al* 2015, Rajkumar 2016). These novel and combination therapies have been observed to provide a durable response and greater progression free survival for MM patients, although there are no current studies performed that investigate how a patient tumour composition may change genetically with the administration of these new treatments.

Heterogeneity is thought to be characteristic of MM, and the administration of therapy acts as a potent source of artificial selection, which changes the dynamics of cancer clones (*Fig. 5*). The therapeutic strategies employed to control tumour growth are genotoxic and result in massive death of aggressive tumour PC clones with “driver” mutations, but also provides a selective pressure for the proliferation of indolent tumour PC clones with “passenger” mutations that may resist treatment (Brioli, *et al* 2014, Greaves and Maley 2012). Thus, therapy can initiate a selective bottleneck facilitating the death or survival of different subclones based on fitness (Brioli, *et al* 2014, Greaves and Maley 2012). Indolent clones surviving treatment may mutate further, acquiring “driver” mutations, thereby conferring improved fitness and malignant potential, which in turn, leads to disease relapse (Brioli, *et al* 2014, Greaves and Maley 2012). Clonal tiding, the rise and fall in dominance of subclones as selective pressures change, has been associated with the occurrence of MM disease relapse (Melchor, *et al* 2014). As a result, distinct clones may dominate at different times during the disease course making MM disease control difficult (Brioli, *et al* 2014, Greaves and Maley 2012).

Figure 5. The impact of therapy on intraclonal heterogeneity in MM: The introduction of treatment presents a new selective pressure on a heterogeneous MM tumour, in conjunction with those already existing due to the bone marrow microenvironment, immune system and competing clonal architecture. While therapy is effective in eliminating dominant PC populations harbouring critical “driver” mutations, it may be ineffective against indolent PC populations which do not have “driver” mutations. These residual PC clones surviving treatment may mutate further, acquiring “driver” mutations, thereby conferring improved fitness and malignant potential, which in turn leads to MM disease relapse.



In view of these new findings, the impact of therapy on clonal evolution and disease progression in MM should be considered at the outset of treatment. To date, a limited number of studies have investigated the changing clonal architecture in MM associated with treatment. Longitudinal WGS study of a single patient tumour through transformation stages - diagnosis, first relapse, second relapse and secondary PCL - has revealed substantial tumour heterogeneity, with clonal tiding in response to selective pressures from treatment, and resulting clonal evolution (Egan, *et al* 2012). These findings have been validated by WES analysis of a patient tumour at diagnosis and first relapse where 81 novel NS-SNVs were identified (33 shared at diagnosis and relapse, while 48 were new) following relapse after first line therapy (Weston-Bell, *et al* 2013). Genomic analysis of paired diagnosis-relapse samples (n = 24) using the Genome-Wide Human SNP array has identified patients exhibiting branching, non linear evolution following therapy driven by the survival of a minor subclone which expanded at relapse (Magrangeas, *et al* 2013). Similarly, targeted genomic mutation panel sequencing of sequential pre- and post-therapy MM samples (n = 25) investigating the most commonly mutated genes in MM has revealed clonal evolution in the majority of patients, including clonal expansion, retraction and/or extinction (Kortum, *et al* 2015). To this end, the complete extinction of subclones (with mutations of *KRAS* and *TP53*) and emergence of new subclones (with mutations of *FAM46C*, *FAT1*, *SPEN* and *TP53*) was identified following therapy (Kortum, *et al* 2015). Conversely, however, WES on paired high-risk SMM-MM post-treatment samples has also identified that therapy is able to reduce clonal complexity of disease (Walker, *et al* 2014). These observations suggest inherent disease complexity at relapse in response to changing selective pressures attributed to the different chemotherapeutic agents and illustrates the need for tumour clones to be monitored for regressing or reappearing subclones which may contribute to disease aggressiveness following specific treatment regimes.

It has been suggested that combinatory treatment regimens should be utilised for a deeper response to reduce both bulk tumour and eliminate clonal and subclonal populations. It has also been suggested that continued therapy versus selective therapy at specific stages of progression for disease control may provide better treatment outcomes. Unfortunately, there is no evidence to support the notion that continuous therapy is more effective than repeated therapy following disease relapse (Gertz 2014). Ultimately, improvements in the outcomes of future treatment will need to take into account the plasticity of MM PCs and altering dominance of genetically distinct subclones that occur as results of previous treatment (Hajek, *et al* 2013, Zhou, *et al* 2009).

1.8 Summary and Objectives

In recent years, rapid advances in genomic technologies, including the application of NGS and single cell analysis techniques, has led to a revolution in our understanding of MM biology and provides direct evidence that MM is a genetically complex disease. Studies also suggest that MM development can be accounted for by a number of tumour development models including the linear model, expansionist model and intraclonal heterogeneity model. In addition, these methods have shown that MM development is associated with significant recurrent probable “driver” mutations in *KRAS*, *NRAS*, *TP53*, *BRAF*, *FAM46C* and *DIS3* which are central to MM disease pathogenesis. Additionally, these new insights will impact current therapeutic strategies used to control MM disease. At this stage, research has mainly been performed on unpaired MGUS/SMM/MM samples, which limits our full understanding of the key “drivers” of MGUS to MM transition. Furthermore, few studies have examined the impact of treatment on intraclonal heterogeneity. With the progress of NGS technologies and the development of more cost-effective methods, thorough analysis of paired MGUS-MM samples and analysis of post-treatment samples should reveal the genetic and molecular mechanisms that play a central role in MM tumour development and disease progression. Ultimately, these insights will heavily influence the future therapeutic strategies used to control MM disease development and relapse.

This project aims to characterise the underlying genomic complexity and tumour evolution associated with disease progression from MGUS/SMM to MM in paired patient samples. These samples were isolated from patients when they were first diagnosed with MGUS or SMM and subsequently when they were diagnosed at MM.

The studies outlined in this thesis address the following aims:

1. To identify the somatic variants associated with the progression of MGUS/SMM to MM using Whole Exome Sequencing (WES)
2. To infer the subclonal tumour evolution model characteristic of disease progression from MGUS/SMM to MM
3. To identify the transcriptomic and methylomic changes associated with the progression of MGUS/SMM to MM using RNA sequencing (RNAseq) and Whole Genome Bisulphite Sequencing (WGBS)

1.9 References

1. Abe, M. (2011) Targeting the interplay between myeloma cells and the bone marrow microenvironment in myeloma. *Int J Hematol*, 94, 334-343.
2. Agirre, X., Castellano, G., Pascual, M., Heath, S., Kulis, M., Segura, V., Bergmann, A., Esteve, A., Merkel, A., Raineri, E., Agueda, L., Blanc, J., Richardson, D., Clarke, L., Datta, A., Russinol, N., Queiros, A.C., Beekman, R., Rodriguez-Madoz, J.R., San Jose-Eneriz, E., Fang, F., Gutierrez, N.C., Garcia-Verdugo, J.M., Robson, M.I., Schirmer, E.C., Guruceaga, E., Martens, J.H., Gut, M., Calasanz, M.J., Flicek, P., Siebert, R., Campo, E., Miguel, J.F., Melnick, A., Stunnenberg, H.G., Gut, I.G., Prosper, F. & Martin-Subero, J.I. (2015) Whole-genome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res*, 25, 478-487.
3. Avet-Loiseau, H., Facon, T., Daviet, A., Godon, C., Rapp, M.J., Harousseau, J.L., Grosbois, B. & Bataille, R. (1999) 14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma. *Intergroupe Francophone du Myelome. Cancer Res*, 59, 4546-4550.
4. Bahlis, N.J. (2012) Darwinian evolution and tiding clones in multiple myeloma. *Blood*, 120, 927-928.
5. Bergsagel, P.L. & Chesi, M. (2013) V. Molecular classification and risk stratification of myeloma. *Hematol Oncol*, 31 Suppl 1, 38-41.
6. Bergsagel, P.L., Kuehl, W.M., Zhan, F., Sawyer, J., Barlogie, B. & Shaughnessy, J., Jr. (2005) Cyclin D dysregulation: an early and unifying pathogenic event in multiple myeloma. *Blood*, 106, 296-303.
7. Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., Hinton, J.W., Li, Y., Tubio, J.M., McLaren, S., S, O.M., Butler, A.P., Teague, J.W., Mudie, L., Anderson, E., Rashid, N., Tai, Y.T., Shamas, M.A., Sperling, A.S., Fulciniti, M., Richardson, P.G., Parmigiani, G., Magrangeas, F., Minvielle, S., Moreau, P., Attal, M., Facon, T., Futreal, P.A., Anderson, K.C., Campbell, P.J. & Munshi, N.C. (2014) Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*, 5, 2997.
8. Brioli, A., Melchor, L., Cavo, M. & Morgan, G.J. (2014) The impact of intra-clonal heterogeneity on the treatment of multiple myeloma. *Br J Haematol*, 165, 441-454.

9. Brito, J.L., Walker, B., Jenner, M., Dickens, N.J., Brown, N.J., Ross, F.M., Avramidou, A., Irving, J.A., Gonzalez, D., Davies, F.E. & Morgan, G.J. (2009) MMSET deregulation affects cell cycle progression and adhesion regulons in t(4;14) myeloma plasma cells. *Haematologica*, 94, 78-86.
10. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M., Anderson, K.C., Ardlie, K.G., Auclair, D., Baker, A., Bergsagel, P.L., Bernstein, B.E., Drier, Y., Fonseca, R., Gabriel, S.B., Hofmeister, C.C., Jagannath, S., Jakubowiak, A.J., Krishnan, A., Levy, J., Liefeld, T., Lonial, S., Mahan, S., Mfuko, B., Monti, S., Perkins, L.M., Onofrio, R., Pugh, T.J., Rajkumar, S.V., Ramos, A.H., Siegel, D.S., Sivachenko, A., Stewart, A.K., Trudel, S., Vij, R., Voet, D., Winckler, W., Zimmerman, T., Carpten, J., Trent, J., Hahn, W.C., Garraway, L.A., Meyerson, M., Lander, E.S., Getz, G. & Golub, T.R. (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, 471, 467-472.
11. Chesi, M. & Bergsagel, P.L. (2011) Many multiple myelomas: making more of the molecular mayhem. *Hematology Am Soc Hematol Educ Program*, 2011, 344-353.
12. Chiecchio, L., Dagrada, G.P., Ibrahim, A.H., Dachs Cabanas, E., Protheroe, R.K., Stockley, D.M., Orchard, K.H., Cross, N.C., Harrison, C.J., Ross, F.M. & Foroni, U.K.M. (2009) Timing of acquisition of deletion 13 in plasma cell dyscrasias is dependent on genetic context. *Haematologica*, 94, 1708-1713.
13. Dimopoulos, K., Gimsing, P. & Gronbaek, K. (2014) The role of epigenetics in the biology of multiple myeloma. *Blood Cancer J*, 4, e207.
14. Dimopoulos, M.A., Oriol, A., Nahi, H., San-Miguel, J., Bahlis, N.J., Usmani, S.Z., Rabin, N., Orlowski, R.Z., Komarnicki, M., Suzuki, K., Plesner, T., Yoon, S.S., Ben Yehuda, D., Richardson, P.G., Goldschmidt, H., Reece, D., Lisby, S., Khokhar, N.Z., O'Rourke, L., Chiu, C., Qin, X., Guckert, M., Ahmadi, T., Moreau, P. & Investigators, P. (2016) Daratumumab, Lenalidomide, and Dexamethasone for Multiple Myeloma. *N Engl J Med*, 375, 1319-1331.
15. Dispenzieri, A., Kyle, R.A., Katzmann, J.A., Therneau, T.M., Larson, D., Benson, J., Clark, R.J., Melton, L.J., 3rd, Gertz, M.A., Kumar, S.K., Fonseca, R., Jelinek, D.F. & Rajkumar, S.V. (2008) Immunoglobulin free light chain ratio is an independent risk factor for progression of smoldering (asymptomatic) multiple myeloma. *Blood*, 111, 785-789.
16. Egan, J.B., Shi, C.X., Tembe, W., Christoforides, A., Kurdoglu, A., Sinari, S., Middha, S., Asmann, Y., Schmidt, J., Braggio, E., Keats, J.J., Fonseca, R., Bergsagel, P.L., Craig, D.W., Carpten, J.D. & Stewart, A.K. (2012) Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood*, 120, 1060-1066.

17. Fluidigm Corporation (2014) Reveal Hidden Variation - C1 Single-Cell DNA Sequencing.
18. Fonseca, R., Bailey, R.J., Ahmann, G.J., Rajkumar, S.V., Hoyer, J.D., Lust, J.A., Kyle, R.A., Gertz, M.A., Greipp, P.R. & Dewald, G.W. (2002) Genomic abnormalities in monoclonal gammopathy of undetermined significance. *Blood*, 100, 1417-1424.
19. Fonseca, R., Bergsagel, P.L., Drach, J., Shaughnessy, J., Gutierrez, N., Stewart, A.K., Morgan, G., Van Ness, B., Chesi, M., Minvielle, S., Neri, A., Barlogie, B., Kuehl, W.M., Liebisch, P., Davies, F., Chen-Kiang, S., Durie, B.G., Carrasco, R., Sezer, O., Reiman, T., Pilarski, L., Avet-Loiseau, H. & International Myeloma Working, G. (2009) International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. *Leukemia*, 23, 2210-2221.
20. Gertz, M.A., Rajkumar, S. V. (2014) Multiple Myeloma Diagnosis and Treatment. (ed. by Springer).
21. Ghobrial, I.M. (2012) Myeloma as a model for the process of metastasis: implications for therapy. *Blood*, 120, 20-30.
22. Gonzalez, D., van der Burg, M., Garcia-Sanz, R., Fenton, J.A., Langerak, A.W., Gonzalez, M., van Dongen, J.J., San Miguel, J.F. & Morgan, G.J. (2007) Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood*, 110, 3112-3121.
23. Greaves, M. & Maley, C.C. (2012) Clonal evolution in cancer. *Nature*, 481, 306-313.
24. Gundry, M., Li, W., Maqbool, S.B. & Vijg, J. (2012) Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic Acids Res*, 40, 2032-2040.
25. Hajek, R., Okubote, S.A. & Svachova, H. (2013) Myeloma stem cell concepts, heterogeneity and plasticity of multiple myeloma. *Br J Haematol*, 163, 551-564.
26. Heuck, C.J., Mehta, J., Bhagat, T., Gundabolu, K., Yu, Y., Khan, S., Chrysofakis, G., Schinke, C., Tariman, J., Vickrey, E., Pulliam, N., Nischal, S., Zhou, L., Bhattacharyya, S., Meagher, R., Hu, C., Maqbool, S., Suzuki, M., Parekh, S., Reu, F., Steidl, U., Grealley, J., Verma, A. & Singhal, S.B. (2013) Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis. *J Immunol*, 190, 2966-2975.
27. International Myeloma Working Group (2003) Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br J Haematol*, 121, 749-757.

28. Kalf, A. & Spencer, A. (2012) The t(4;14) translocation and FGFR3 overexpression in multiple myeloma: prognostic implications and current clinical strategies. *Blood Cancer J*, 2, e89.
29. Klein, U., Goossens, T., Fischer, M., Kanzler, H., Braeuninger, A., Rajewsky, K. & Kuppers, R. (1998) Somatic hypermutation in normal and transformed human B cells. *Immunol Rev*, 162, 261-280.
30. Kortum, K.M., Langer, C., Monge, J., Bruins, L., Zhu, Y.X., Shi, C.X., Jedlowski, P., Egan, J.B., Ojha, J., Bullinger, L., Kull, M., Ahmann, G., Rasche, L., Knop, S., Fonseca, R., Einsele, H., Stewart, A.K. & Braggio, E. (2015) Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann Hematol*, 94, 1205-1211.
31. Kulis, M. & Esteller, M. (2010) DNA methylation and cancer. *Adv Genet*, 70, 27-56.
32. Kyle, R.A., Buadi, F. & Rajkumar, S.V. (2011) Management of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Oncology (Williston Park)*, 25, 578-586.
33. Kyle, R.A., Therneau, T.M., Rajkumar, S.V., Larson, D.R., Plevak, M.F., Offord, J.R., Dispenzieri, A., Katzmann, J.A. & Melton, L.J., 3rd (2006) Prevalence of monoclonal gammopathy of undetermined significance. *N Engl J Med*, 354, 1362-1369.
34. Landgren, O., Kyle, R.A., Pfeiffer, R.M., Katzmann, J.A., Caporaso, N.E., Hayes, R.B., Dispenzieri, A., Kumar, S., Clark, R.J., Baris, D., Hoover, R. & Rajkumar, S.V. (2009) Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood*, 113, 5412-5417.
35. Lawson, M.A., McDonald, M.M., Kovacic, N., Hua Khoo, W., Terry, R.L., Down, J., Kaplan, W., Paton-Hough, J., Fellows, C., Pettitt, J.A., Neil Dear, T., Van Valckenborgh, E., Baldock, P.A., Rogers, M.J., Eaton, C.L., Vanderkerken, K., Pettit, A.R., Quinn, J.M., Zannettino, A.C., Phan, T.G. & Croucher, P.I. (2015) Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. *Nat Commun*, 6, 8983.
36. Lohr, J.G., Stojanov, P., Carter, S.L., Cruz-Gordillo, P., Lawrence, M.S., Auclair, D., Sougnez, C., Knoechel, B., Gould, J., Saksena, G., Cibulskis, K., McKenna, A., Chapman, M.A., Straussman, R., Levy, J., Perkins, L.M., Keats, J.J., Schumacher, S.E., Rosenberg, M., Multiple Myeloma Research, C., Getz, G. & Golub, T.R. (2014) Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*, 25, 91-101.

37. Lonial, S., Dimopoulos, M., Palumbo, A., White, D., Grosicki, S., Spicka, I., Walter-Croneck, A., Moreau, P., Mateos, M.V., Magen, H., Belch, A., Reece, D., Beksac, M., Spencer, A., Oakervee, H., Orłowski, R.Z., Taniwaki, M., Rollig, C., Einsele, H., Wu, K.L., Singhal, A., San-Miguel, J., Matsumoto, M., Katz, J., Bleickardt, E., Poulart, V., Anderson, K.C., Richardson, P. & Investigators, E.-. (2015) Elotuzumab Therapy for Relapsed or Refractory Multiple Myeloma. *N Engl J Med*, 373, 621-631.
38. Lonial, S., Weiss, B.M., Usmani, S.Z., Singhal, S., Chari, A., Bahlis, N.J., Belch, A., Krishnan, A., Vescio, R.A., Mateos, M.V., Mazumder, A., Orłowski, R.Z., Sutherland, H.J., Blade, J., Scott, E.C., Oriol, A., Berdeja, J., Gharibo, M., Stevens, D.A., LeBlanc, R., Sebag, M., Callander, N., Jakubowiak, A., White, D., de la Rubia, J., Richardson, P.G., Lisby, S., Feng, H., Uhlar, C.M., Khan, I., Ahmadi, T. & Voorhees, P.M. (2016) Daratumumab monotherapy in patients with treatment-refractory multiple myeloma (SIRIUS): an open-label, randomised, phase 2 trial. *Lancet*, 387, 1551-1560.
39. Lopez-Corral, L., Gutierrez, N.C., Vidriales, M.B., Mateos, M.V., Rasillo, A., Garcia-Sanz, R., Paiva, B. & San Miguel, J.F. (2011) The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells. *Clin Cancer Res*, 17, 1692-1700.
40. Lopez-Corral, L., Sarasquete, M.E., Bea, S., Garcia-Sanz, R., Mateos, M.V., Corchete, L.A., Sayagues, J.M., Garcia, E.M., Blade, J., Oriol, A., Hernandez-Garcia, M.T., Giraldo, P., Hernandez, J., Gonzalez, M., Hernandez-Rivas, J.M., San Miguel, J.F. & Gutierrez, N.C. (2012) SNP-based mapping arrays reveal high genomic complexity in monoclonal gammopathies, from MGUS to myeloma status. *Leukemia*, 26, 2521-2529.
41. Magrangeas, F., Avet-Loiseau, H., Gouraud, W., Lode, L., Decaux, O., Godmer, P., Garderet, L., Voillat, L., Facon, T., Stoppa, A.M., Marit, G., Hulin, C., Casassus, P., Tiab, M., Voog, E., Randriamalala, E., Anderson, K.C., Moreau, P., Munshi, N.C. & Minvielle, S. (2013) Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*, 27, 473-481.
42. Manier, S., Sacco, A., Leleu, X., Ghobrial, I.M. & Roccaro, A.M. (2012) Bone marrow microenvironment in multiple myeloma progression. *J Biomed Biotechnol*, 2012, 157496.
43. Marango, J., Shimoyama, M., Nishio, H., Meyer, J.A., Min, D.J., Sirulnik, A., Martinez-Martinez, Y., Chesi, M., Bergsagel, P.L., Zhou, M.M., Waxman, S., Leibovitch, B.A., Walsh, M.J. & Licht, J.D. (2008) The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood*, 111, 3145-3154.

44. Melchor, L., Brioli, A., Wardell, C.P., Murison, A., Potter, N.E., Kaiser, M.F., Fryer, R.A., Johnson, D.C., Begum, D.B., Hulkki Wilson, S., Vijayaraghavan, G., Titley, I., Cavo, M., Davies, F.E., Walker, B.A. & Morgan, G.J. (2014) Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia*, 28, 1705-1715.
45. Mikhael, J.R., Dingli, D., Roy, V., Reeder, C.B., Buadi, F.K., Hayman, S.R., Dispenzieri, A., Fonseca, R., Sher, T., Kyle, R.A., Lin, Y., Russell, S.J., Kumar, S., Bergsagel, P.L., Zeldenrust, S.R., Leung, N., Drake, M.T., Kapoor, P., Ansell, S.M., Witzig, T.E., Lust, J.A., Dalton, R.J., Gertz, M.A., Stewart, A.K., Rajkumar, S.V., Chanan-Khan, A., Lacy, M.Q. & Mayo, C. (2013) Management of newly diagnosed symptomatic multiple myeloma: updated Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) consensus guidelines 2013. *Mayo Clin Proc*, 88, 360-376.
46. Mitra, A.K., Mukherjee, U.K., Harding, T., Jang, J.S., Stessman, H., Li, Y., Abyzov, A., Jen, J., Kumar, S., Rajkumar, V. & Van Ness, B. (2016) Single-cell analysis of targeted transcriptome predicts drug sensitivity of single cells within human myeloma tumors. *Leukemia*, 30, 1094-1102.
47. Mitsiades, C.S., Mitsiades, N.S., Munshi, N.C., Richardson, P.G. & Anderson, K.C. (2006) The role of the bone microenvironment in the pathophysiology and therapeutic management of multiple myeloma: interplay of growth factors, their receptors and stromal interactions. *Eur J Cancer*, 42, 1564-1573.
48. MMRF (2015) Multiple Myeloma Drugs Guide.
49. Morgan, G.J., Walker, B.A. & Davies, F.E. (2012) The genetic architecture of multiple myeloma. *Nat Rev Cancer*, 12, 335-348.
50. NCI (2014) SEER Stat Fact Sheets: Myeloma. National Cancer Institute at the National Institutes of Health.
51. Noll, J.E., Williams, S.A., Purton, L.E. & Zannettino, A.C. (2012) Tug of war in the haematopoietic stem cell niche: do myeloma plasma cells compete for the HSC niche? *Blood Cancer J*, 2, e91.
52. Nowell, P.C. (1976) The clonal evolution of tumor cell populations. *Science*, 194, 23-28.
53. Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M. & Corcoran, L.M. (2015) The generation of antibody-secreting plasma cells. *Nat Rev Immunol*, 15, 160-171.
54. Palumbo, A. & Anderson, K. (2011) Multiple myeloma. *N Engl J Med*, 364, 1046-1060.

55. Palumbo, A., Chanan-Khan, A., Weisel, K., Nooka, A.K., Masszi, T., Beksac, M., Spicka, I., Hungria, V., Munder, M., Mateos, M.V., Mark, T.M., Qi, M., Schecter, J., Amin, H., Qin, X., Deraedt, W., Ahmadi, T., Spencer, A., Sonneveld, P. & Investigators, C. (2016) Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma. *N Engl J Med*, 375, 754-766.
56. Perez-Persona, E., Vidriales, M.B., Mateo, G., Garcia-Sanz, R., Mateos, M.V., de Coca, A.G., Galende, J., Martin-Nunez, G., Alonso, J.M., de Las Heras, N., Hernandez, J.M., Martin, A., Lopez-Berges, C., Orfao, A. & San Miguel, J.F. (2007) New criteria to identify risk of progression in monoclonal gammopathy of uncertain significance and smoldering multiple myeloma based on multiparameter flow cytometry analysis of bone marrow plasma cells. *Blood*, 110, 2586-2592.
57. Plass, C., Pfister, S.M., Lindroth, A.M., Bogatyrova, O., Claus, R. & Lichter, P. (2013) Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet*, 14, 765-780.
58. Rajkumar, S.V. (2016) Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management. *Am J Hematol*, 91, 719-734.
59. Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., Sezer, O., Kristinsson, S.Y., Caers, J., Usmani, S.Z., Lahuerta, J.J., Johnsen, H.E., Beksac, M., Cavo, M., Goldschmidt, H., Terpos, E., Kyle, R.A., Anderson, K.C., Durie, B.G. & Miguel, J.F. (2014) International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*, 15, e538-548.
60. Rajkumar, S.V., Kyle, R.A., Therneau, T.M., Melton, L.J., 3rd, Bradwell, A.R., Clark, R.J., Larson, D.R., Plevak, M.F., Dispenzieri, A. & Katzmann, J.A. (2005) Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance. *Blood*, 106, 812-817.
61. Salhia, B., Baker, A., Ahmann, G., Auclair, D., Fonseca, R. & Carpten, J. (2010) DNA methylation analysis determines the high frequency of genic hypomethylation and low frequency of hypermethylation events in plasma cell tumors. *Cancer Res*, 70, 6934-6944.
62. Sawyer, J.R., Tian, E., Shaughnessy Jr, J.D., Epstein, J., Swanson, C.M., Stangeby, C., Hale, C.L., Parr, L., Lynn, M., Sarmartino, G., Lukacs, J.L., Stein, C., Bailey, C., Zangari, M., Davies, F.E., Van Rhee, F., Barlogie, B. & Morgan, G.J. (2016) Hyperhaploidy is a novel high-risk cytogenetic subgroup in multiple myeloma. *Leukemia*.

63. Shapiro-Shelef, M. & Calame, K. (2005) Regulation of plasma-cell development. *Nat Rev Immunol*, 5, 230-242.
64. Singhal, S. & Mehta, J. (2006) Multiple myeloma. *Clin J Am Soc Nephrol*, 1, 1322-1330.
65. Smith, E.M., Boyd, K. & Davies, F.E. (2010) The potential role of epigenetic therapy in multiple myeloma. *Br J Haematol*, 148, 702-713.
66. Stavnezer, J. (1996) Immunoglobulin class switching. *Curr Opin Immunol*, 8, 199-205.
67. Stessman, H.A., Baughn, L.B., Sarver, A., Xia, T., Deshpande, R., Mansoor, A., Walsh, S.A., Sunderland, J.J., Dolloff, N.G., Linden, M.A., Zhan, F., Janz, S., Myers, C.L. & Van Ness, B.G. (2013) Profiling bortezomib resistance identifies secondary therapies in a mouse myeloma model. *Mol Cancer Ther*, 12, 1140-1150.
68. Usmani, S.Z., Weiss, B.M., Plesner, T., Bahlis, N.J., Belch, A., Lonial, S., Lokhorst, H.M., Voorhees, P.M., Richardson, P.G., Chari, A., Sasser, A.K., Axel, A., Feng, H., Uhlar, C.M., Wang, J., Khan, I., Ahmadi, T. & Nahi, H. (2016) Clinical efficacy of daratumumab monotherapy in patients with heavily pretreated relapsed or refractory multiple myeloma. *Blood*, 128, 37-44.
69. Walker, B.A., Boyle, E.M., Wardell, C.P., Murison, A., Begum, D.B., Dahir, N.M., Proszek, P.Z., Johnson, D.C., Kaiser, M.F., Melchor, L., Aronson, L.I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J.R., Brioli, A., Mikulasova, A., Cairns, D.A., Gregory, W.M., Quartilho, A., Drayson, M.T., Russell, N., Cook, G., Jackson, G.H., Leleu, X., Davies, F.E. & Morgan, G.J. (2015) Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol*, 33, 3911-3920.
70. Walker, B.A., Wardell, C.P., Chiecchio, L., Smith, E.M., Boyd, K.D., Neri, A., Davies, F.E., Ross, F.M. & Morgan, G.J. (2011) Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*, 117, 553-562.
71. Walker, B.A., Wardell, C.P., Melchor, L., Brioli, A., Johnson, D.C., Kaiser, M.F., Mirabella, F., Lopez-Corral, L., Humphray, S., Murray, L., Ross, M., Bentley, D., Gutierrez, N.C., Garcia-Sanz, R., San Miguel, J., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2014) Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*, 28, 384-390.
72. Walker, B.A., Wardell, C.P., Melchor, L., Hulkki, S., Potter, N.E., Johnson, D.C., Fenwick, K., Kozarewa, I., Gonzalez, D., Lord, C.J., Ashworth, A., Davies, F.E. & Morgan, G.J. (2012) Intraclonal heterogeneity and distinct molecular mechanisms

- characterize the development of t(4;14) and t(11;14) myeloma. *Blood*, 120, 1077-1086.
73. Wang, L. & Young, D.C. (2001) Suppression of polyclonal immunoglobulin production by M-proteins shows isotype specificity. *Ann Clin Lab Sci*, 31, 274-278.
74. Weinhold, N., Ashby, C., Rasche, L., Chavan, S.S., Stein, C., Stephens, O.W., Tytarenko, R., Bauer, M.A., Meissner, T., Deshpande, S., Patel, P.H., Buzder, T., Molnar, G., Peterson, E.A., van Rhee, F., Zangari, M., Thanendrarajan, S., Schinke, C., Tian, E., Epstein, J., Barlogie, B., Davies, F.E., Heuck, C.J., Walker, B.A. & Morgan, G.J. (2016) Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. *Blood*, 128, 1735-1744.
75. Weinhold, N., Chavan, S.S., Heuck, C., Stephens, O.W., Tytarenko, R., Bauer, M., Peterson, E.A., Ashby, T.C., Meissner, T., Stein, C.K., Johann, D., Johnson, S.K., Yaccoby, S., Epstein, J., van Rhee, F., Zangari, M., Schinke, C., Thanendrarajan, S., Davies, F.E., Barlogie, B. & Morgan, G.J. (2015) High Risk Multiple Myeloma Demonstrates Marked Spatial Genomic Heterogeneity Between Focal Lesions and Random Bone Marrow; Implications for Targeted Therapy and Treatment Resistance. *Blood*.
76. Weston-Bell, N., Gibson, J., John, M., Ennis, S., Pfeifer, S., Cezard, T., Ludwig, H., Collins, A., Zojer, N. & Sahota, S.S. (2013) Exome sequencing in tracking clonal evolution in multiple myeloma following therapy. *Leukemia*, 27, 1188-1191.
77. Yadav, V.K. & De, S. (2014) An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform*.
78. Zhou, Y., Barlogie, B. & Shaughnessy, J.D., Jr. (2009) The molecular characterization and clinical management of multiple myeloma in the post-genome era. *Leukemia*, 23, 1941-1956.
79. Zweegman, S., Palumbo, A., Bringhen, S. & Sonneveld, P. (2014) Age and aging in blood disorders: multiple myeloma. *Haematologica*, 99, 1133-1137.

Chapter 2

**Subclonal evolution in disease progression from
MGUS/SMM to multiple myeloma is characterised by
clonal stability**

Statement of Authorship

| | |
|---------------------|---|
| Title of Paper | Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability |
| Publication Status | <input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style |
| Publication Details | Dutta, A. K. , Fink, J. L., Grady, J. P., Morgan, G. J., Mullighan, C. G., To, L. B., Hewett, D. R. & Zannettino, A. C. W. (2018) Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability. <i>Leukemia</i> . https://doi.org/10.1038/s41375-018-0206-x |

Principal Author

| | | | |
|--------------------------------------|--|------|-----------|
| Name of Principal Author (Candidate) | Ankit K.Dutta | | |
| Contribution to the Paper | Experimental design Generation of data Analysis and interpretation of data Manuscript development, writing and review | | |
| Overall percentage (%) | 65% | | |
| Certification: | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. | | |
| Signature | | Date | 31/7/2018 |

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | |
|---------------------------|---|------|------------|
| Name of Co-Author | J. Lynn Fink | | |
| Contribution to the Paper | Experimental design Analysis and interpretation of data Manuscript development and review | | |
| Signature | | Date | 2018-07-11 |

| | | | |
|---------------------------|--|------|---------|
| Name of Co-Author | John P. Grady | | |
| Contribution to the Paper | Analysis and interpretation of data Manuscript review | | |
| Signature | | Date | 11/7/18 |

| | | | |
|---------------------------|---|------|---------|
| Name of Co-Author | Gareth J. Morgan | | |
| Contribution to the Paper | Experimental design Grant funding application Manuscript review | | |
| Signature | | Date | 7-11-18 |

| | | | |
|---------------------------|---|------|----------------------------|
| Name of Co-Author | Charles G. Mullighan | | |
| Contribution to the Paper | Experimental design Grant funding application Manuscript review | | |
| Signature | | Date | 10 th July 2018 |

| | | | |
|---------------------------|---|------|---------|
| Name of Co-Author | Luen B. To | | |
| Contribution to the Paper | Experimental design Grant funding application Manuscript review | | |
| Signature | | Date | 11.7.18 |

| | | | |
|---------------------------|---|------|-----------|
| Name of Co-Author | Duncan R. Hewett | | |
| Contribution to the Paper | Experimental design Supervision of work Manuscript development and review | | |
| Signature | | Date | 26/7/2018 |

| | | | |
|---------------------------|--|------|-----------|
| Name of Co-Author | Andrew C. W. Zannettino | | |
| Contribution to the Paper | Experimental design Supervision of work Manuscript development and review Project funding | | |
| Signature | | Date | 31/7/2018 |

Chapter 2: Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability

Ankit K. Dutta^{1,2}, J. Lynn Fink^{3*}, John P. Grady³, Gareth J. Morgan⁴, Charles G. Mullighan⁵, Luen B. To^{6,7}, Duncan R. Hewett^{1,2} & Andrew C.W. Zannettino^{1,2*}

¹Myeloma Research Laboratory, Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA, 5005, Australia.

²Cancer Theme, South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, 5000, Australia.

³Genomic Medicine Division, The University of Queensland, Diamantina Institute (UQDI), Brisbane, QLD, 4102, Australia.

⁴The Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, 72205, USA.

⁵Department of Pathology and the Hematological Malignancies Program, St Jude Children's Research Hospital, Memphis, TN, 38105, USA.

⁶SA Pathology, Adelaide, SA, 5000, Australia.

⁷Haematology and Bone Marrow Transplant Unit, Royal Adelaide Hospital, Adelaide, SA, 5000, Australia.

*co-senior authors

Running title: Clonal stability characterises progression to myeloma

Keywords: MGUS, Smoldering MM, Myeloma, Next generation sequencing, Intraclonal heterogeneity, Clonal evolution, Clonal stability

Chapter 2 incorporates the original article published: **Dutta, A.K., Fink, J.L., Grady, J.P., Morgan, G.J, Mullighan, C.G, To, L. B., Hewett, D.R. & Zannettino, A.C.W.** (2018) Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability. *Leukemia*. <https://doi.org/10.1038/s41375-018-0206-x>

2.1 Abstract

Multiple myeloma (MM) is a largely incurable haematological malignancy defined by the clonal proliferation of malignant plasma cells (PCs) within the bone marrow. Clonal heterogeneity has recently been established as a feature in MM, however, the subclonal evolution associated with disease progression has not been described. Here, we performed whole exome sequencing of serial samples from 10 patients, providing new insights into the progression from monoclonal gammopathy of undetermined significance (MGUS) and smouldering MM (SMM), to symptomatic MM. We confirm that intraclonal genetic heterogeneity is a common feature at diagnosis and that the driving events involved in disease progression are more subtle than previously reported. We reveal that MM evolution is mainly characterised by the phenomenon of clonal stability, where the transformed subclonal PC populations identified at MM are already present in the asymptomatic MGUS/SMM stages. Our findings highlight the possibility that PC extrinsic factors may play a role in subclonal evolution and MGUS/SMM to MM progression.

2.2 Introduction

Multiple Myeloma (MM) is a haematological malignancy characterised by the uncontrolled proliferation of neoplastic plasma cells (PCs) within the bone marrow (BM). MM accounts for ~10% of all haematological malignancies¹, with a median survival rate of 5.2 years². Although 10 new therapeutic agents for MM have been approved in the last 20 years, and patient outcomes have improved significantly, individual responses to therapy and overall survival are varied³. MM remains largely incurable, with relapse being a common feature of disease.

The development of MM has been classically viewed as a multistage process involving the acquisition of multiple genetic mutations, with immunoglobulin heavy chain translocations and hyperdiploidy known to be common initiating events that deregulate normal PC behaviour leading to the development of monoclonal gammopathy of undetermined significance (MGUS)⁴⁻⁷. Further mutational load leads to an intermediate stage of smouldering multiple myeloma (SMM)^{4,8}. However, these common initiating events are insufficient to cause MM transformation, as MGUS/SMM patients commonly harbour these abnormalities and show no clinical symptoms of MM^{9,10}. Studies have shown that progression to MM is associated with additional genetic changes including aneuploidy, chromosomal translocations, single nucleotide variants, small insertions and deletions, and copy number variants affecting one or more genes, with mutations present at a frequency of 0.1 to 10 per megabase¹¹.

Recent studies show that MM patients display complex mutational landscapes involving intraclonal genetic heterogeneity at the bulk tumour level, where mutations are acquired in a non-linear branching pattern¹²⁻¹⁷. Intraclonal heterogeneity has been observed at all stages of MM, suggesting that disease progression may be mediated through inter-subclone competition and outgrowth of the fittest of these subclones. Genomic studies on large cohorts of unmatched MGUS-SMM-MM patients have led to the discovery of recurrently mutated genes, of which *KRAS*, *NRAS*, *TP53*, *BRAF*, *FAM46C* and *DIS3* are believed to be drivers of MM transformation^{10,18-20}. While clonal heterogeneity is now an established feature in MM, the subclonal evolution associated with MGUS/SMM to MM progression remains poorly understood.

A comprehensive approach to identifying the key somatic mutations and infer the subclonal evolution associated with MM transformation, involves the longitudinal study of sequential MGUS-MM or SMM-MM samples from the same patient. However, because MGUS is often an incidental finding, it is extremely rare to have diagnostic BM samples

from the same patient at both the MGUS and MM stages. In addition, because there are no cell line or mouse models of MGUS or SMM, there has been limited opportunity to study the specific genetic changes and molecular mechanisms that characterise the progression from MGUS/SMM to MM.

Here, we report a longitudinal genomics investigation of MM, based on paired MGUS-MM (n = 5) or SMM-MM (n = 5) patient samples obtained from the same patient when initially diagnosed at MGUS/SMM, and subsequently when they developed MM. Using whole exome sequencing, we have obtained a detailed description of the genomic complexity and subclonal evolution underlying progression from MGUS/SMM to MM.

2.3 Materials & Methods

2.3.1 Clinical samples.

Bone marrow mononuclear cell aspirates were collected from patients at MGUS/SMM, and subsequently at later diagnosis of MM (MGUS-MM (n = 5) and SMM-MM (n = 5)). The median time to progression of MGUS to MM was 3.2 years (range 1 - 13 years) and SMM to MM was 1.2 years (range 0.48 – 4.1 years). Where available, the cytogenetic status of patients is reported in *Supplementary Table 1*. Samples were collected from patients prior to treatment. All patients provided informed consent in accordance with the Declaration of Helsinki. Samples were cryopreserved by the South Australian Cancer Research Biobank (SACRB) at SA Pathology. The studies were approved by the Royal Adelaide Hospital Human Research Ethics Committee (HREC/13/RAH/569 No: 131133). Samples were collected over a period of 22 years, and we initially began this study with paired-samples from 18 patients. However, due to our strict criteria for sample purity and mutation calling resolution, final analysis was only performed on samples from 10 patients.

2.3.2 Cell sorting.

PCs and non-tumour cells were purified using multicolour flow cytometry as previously described²¹. Briefly, approximately 1×10^5 mononuclear cells were prepared for single stain antibody control (CD138-PE (Beckman Coulter #A54190) and CD38-PE-Cy7 (Biolegend #303515)) and compensation/FMO tubes (1: unstained; 2: hydroxystilbamidine (FluoroGold; Life Technologies) only; 3: CD38-PE-Cy7+FluoroGold; 4: CD138-PE+FluoroGold; and 5: CD38-PE-Cy7+CD138-PE). The sort sample was stained with CD138-PE and CD38-PE-Cy7 antibody at $1 \mu\text{L}/100 \mu\text{L}$ cells. Cells were stained with FluoroGold immediately prior to sorting. Viable PCs ($\text{CD138}^+\text{CD38}^+$ and FluoroGold negative) and non-tumour cells were sorted on the FACS Aria Fusion sorter (BD Biosciences). FACS purity check was carried out on sorted cells, using 100-500 cells from each sample.

2.3.3 DNA isolation, QC and sequencing.

DNA was isolated from purified PC and non-tumour populations using the All Prep DNA/RNA Micro Kit (QIAGEN) as per manufacturers' instructions. Yields and quality was assessed using the NanoDrop 8000 and Qubit 2.0 fluorometer (Thermo Fisher Scientific).

115ng of gDNA were used as input for fragmentation on the Covaris E220, followed by end-repair/A-tailing and ligation of SureSelect Adapter Oligos (Agilent). Pre-Capture PCR amplification of 10 cycles, or 12 cycles for low input samples, were performed. A total of 750ng of each sample was hybridised to SureSelect XT Clinical Research Exome (Agilent) probes overnight. Captured DNA was amplified with 11 cycles of post-capture PCR incorporating index barcodes. Sequencing was performed on the Illumina HiSeq4000 (2x100 bp paired-end reads) and NextSeq 500 (2x150 bp paired-end reads). Samples were sequenced to a minimum depth of ~140X mean coverage. Isolated non-tumour cells were also sequenced to a similar average depth (138x).

2.3.4 Analysis of data.

Sequencing reads were mapped to the human decoy genome (hs37d5) using Novoalign (v3.02.08), followed by post-processing according to GATK best practices²². Somatic single nucleotide and small indel variants were called using MuTect2²³ and multiSNV²⁴. Variants were filtered based on: 10+ reads covering the variant site; 5+ reads covering the variant in the tumour sample. Variants were annotated with SnpEff²⁵.

R 3.3.2 was used throughout for analyses. Somatic copy number variants were called using CNVkit²⁶ v0.7.11 and custom in-house methods developed to support highly aneuploid genomes to perform segmentation and calculate log2 changes.

Clonal evolution was investigated using PhyloWGS²⁷ and visualised using fish plot in R²⁸. PhyloWGS is noted to inflate the number of subclones, thus we recognise that subclone numbers may be overestimated. All phylogenetic trees constructed were based on the assumption that there is a single founder clone.

Additional information on sequencing and somatic mutation analysis is given in the supplementary methods.

2.3.5 Data deposition.

All raw sequencing reads have been deposited in the EGA repository (Accession number: EGAS00001002850).

2.3.6 Code Availability.

Custom script generated for CNV analysis is available on request from corresponding authors.

2.4 Results

2.4.1 A changing spectrum of acquired mutations, not mutational load, is associated with MM progression.

Whole exome sequencing was performed on paired MGUS/SMM to MM patients [detailed in *Supplementary Table 1*] to a minimum average depth of 140x [*Supplementary Table 2*]. A total of 1614 somatic non-synonymous single nucleotide variants (NS-SNVs) were identified across the MGUS/SMM samples (range 30-220) with a median 161 per patient. Interestingly, in the MM samples, we identified a total 1508 somatic NS-SNVs (range 59-226), with a median 152 per patient. There was an average of 27 NS-SNVs that were shared between the MGUS/SMM and MM stage (range 0-53) [*Supplementary Table 3*]. We observed a moderately higher mutation load compared to previous larger cohort studies of MM, which identified median SNV numbers of 31 (range 15-46)⁴, and 52 mutations per patient (range 2-488)¹⁸.

Recent sequencing studies of *unpaired* MM samples have described an increasing median NS-SNV burden from MGUS to SMM to MM, with MGUS patients harbouring approximately half the number of NS-SNVs when compared to unmatched MM patients⁴, with an average of 35 at the MM stage¹⁰. Here, we observed the opposite upon progression from MGUS/SMM to MM, where 7 out of 10 patients showed a decrease in total mutational load [*Figure 1a*]. While the total mutational burden is not considerably different between MGUS/SMM and MM, the presence of intraclonal heterogeneity and changes in the spectrum of mutated genes between disease stages, suggests that there is waxing and waning of subclones over time²⁹.

We next examined the changing mutational landscape associated with MGUS/SMM to MM progression, to identify the genetic aberrations associated with this process, including both previously reported “drivers” of MM and frequently acquired mutations present at MM transition. To this end, we identified 2566 unique genes with acquired variants at MM transition across all patients. The most common genes harbouring mutations at MM include *KRAS*, *KMT2D*, *RYR2*, *DNAH5*, *PCDH8*, *RP4-669P10.16*, *DIS3*, *FAT3*, *PKHD1* and *SPI40* [*Figure 1b*]. Moreover, we identified 15 previously reported recurrently mutated genes: *KRAS*, *FAT3*, *DIS3*, *TRAF3*, *SPI40*, *RBI*, *PTEN*, *ROBO1*, *PRDM1*, *NRAS*, *MYC*, *MAGED1*, *IRF4*, *HLA-A* and *CDK4*^{10,18-20,30} [*Figure 1c*], including mutations in 3 known “drivers” of MM: *KRAS*, *NRAS* and *DIS3* [*Supplementary Table 4*]. In our samples, mutations in *KRAS* and *NRAS*, were mutually exclusive, consistent with

previous observations that report the rare co-occurrence of mutations in these genes (in 2% of patients)²⁰.

The RAS/MAPK pathway was highly mutated with 40% of patients at MGUS/SMM, and 70% at MM, harbouring mutations in *KRAS* and *NRAS*. *DIS3* was mutated in 30% of patients at the MM stage only [Figure 1d]. This highlights that driver mutations can be acquired at both the asymptomatic stages and be maintained during progression to MM, or be acquired later at MM. However, we found that acquisitions of driver mutations are subclonal in nature. Low variant allele frequencies were identified for RAS pathway mutations (*KRAS* range 0.024 - 0.53, *NRAS* 0.03 - 0.28), suggesting that these mutations were present in subclonal PC populations during progression (P01, P02, P05 and P06) [Figure 1d]. Interestingly, acquisition of *DIS3* mutations at the MM stage in patients P07 and P10 was observed to be clonal in nature³¹ [Figure 1d].

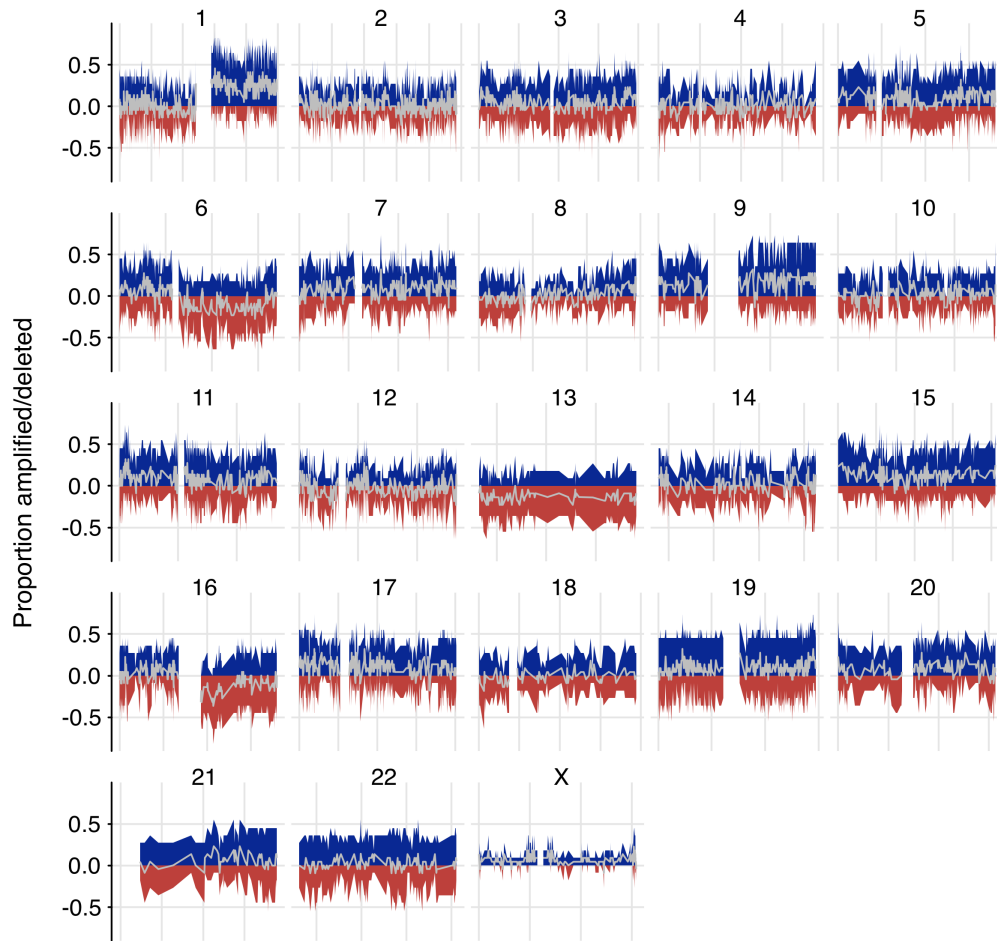
We also characterised the copy number variation (CNV) landscape associated with MGUS/SMM to MM progression, finding copy changes to be widespread [Figure 2a]. This contrasts a recent small longitudinal study of SMM to MM transformation that showed that copy changes are a feature of early stage of MM disease and not associated with progression⁴. We observed that MGUS/SMM patients harbour a similar frequency of chromosomal loci copy gains and losses than MM patients, with a median of 70 at MGUS/SMM (range 19 to 114), and 67.5 at MM (range 43 to 103) [Supplementary Table 5]. Upon progression, we observed known frequent chromosomal copy number abnormalities in MM, including amplifications on chromosome arms 1q, 3p, 6p, 9p, 11q, 19p, 19q and 21q, coupled with losses on chromosome arms 1p, 6q, 8p, 13q, 16q and 22q across patients in the cohort [Figure 2b]. While we performed a gene level copy number analysis, we did not find any genes that were consistently gained or lost in our cohort upon progression. Interestingly, we also observed that many of the cytogenetic abnormalities associated with MM are present at MGUS/SMM stages and that standard cytogenetic methods did not accurately capture these abnormalities [Supplementary Figure 4].

Figure 1. Pattern of genetic mutations in MGUS/SMM to MM progression. (a) Total NS-SNV mutational load associated with progression from MGUS/SMM to MM in individual patients (b) Waterfall diagram indicating the 10 most frequently mutated genes associated with progression from MGUS/SMM to MM (c) We identified mutations in 15 reported recurrently mutated MM genes in our MGUS/SMM to MM samples. However, individual patients harbour a heterogeneous genetic architecture, with a combination of mutations in known driver (*KRAS*, *NRAS* and *DIS3*) and other candidate genes. Mutations in RAS/MAPK pathway genes are most prevalent. (d) Gradient diagram across all patients indicating the variant allele frequencies (VAF) of identified known cancer drivers. RAS mutations are observed to mainly occur subclonally.

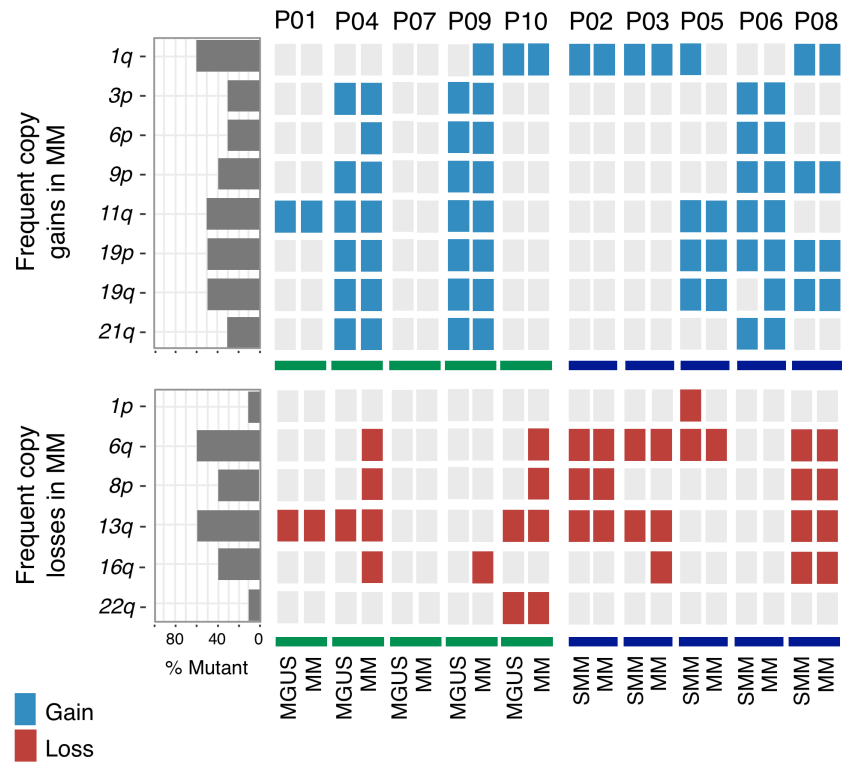


Figure 2. The chromosomal copy number variation landscape associated with MGUS/SMM to MM progression. (a) Chromosomal copy number landscape plot illustrating the proportion of patients with copy amplifications (blue) and deletions (red) of chromosomes across all patients associated with MGUS/SMM to MM progression. Grey traces the average in the cohort as a whole, where below 0 indicates loss and above 0 represents gain. **(b)** Frequent previously reported copy number changes of MM were identified, including gains on 1q, 3p, 6p, 9p, 11q, 19p, 19q and 21q; and losses on 6q, 8p, 13q, 16q and 22q being present at the asymptomatic stages and maintained with progression to MM.

a



b



2.4.2 The subclonal architecture required for MM progression exists at MGUS/SMM diagnosis

While clonal heterogeneity is an established feature in MM, the subclonal evolution associated with disease progression has not been well explored. Due to the nature of our paired longitudinal patient samples, we were able to directly examine the relationship between genetic variegation and clonal structure to construct evolutionary trajectories accompanying progression to MM in 8 individual patients.

Comparisons of unpaired MGUS/SMM and MM samples have shown that MGUS/SMM exhibits mutational similarity with MM, but many mutations were present in a smaller proportion of aberrant PCs^{32,33}. Similarly, small paired-sample studies examining the evolution over time of asymptomatic monoclonal gammopathies (AMGs) to MM (n = 4)³⁴, and high risk SMM to MM patients (n = 4)⁴, have also found that most somatic changes required for MM were present at the asymptomatic stages, with the clinically dominant MM subclone present at the SMM stage. Therefore, the occurrence of clonal evolution in MM represents a change in clonal heterogeneity over time from the asymptomatic stages to MM¹³.

In both MGUS-MM (P01, P04 and P10) and SMM-MM (P02, P03, P05, P06 and P08) progression, we find a prevailing model of evolution defined by clonal stability. This is where the transformed subclonal PC populations identified at MM were already present in the asymptomatic MGUS/SMM stages. Progression to MM involved subtle changes in the existent subclonal structure from MGUS/SMM, coupled with a degree of emergence and/or extinction of child subclonal branches [*Figure 3, 4*]. Of note, we observed that subclonal evolution has already begun prior to MGUS/SMM sampling. While multiple subclonal populations are present at MGUS/SMM diagnosis, each patient harbours unique set of oncogenic mutations driving MM progression.

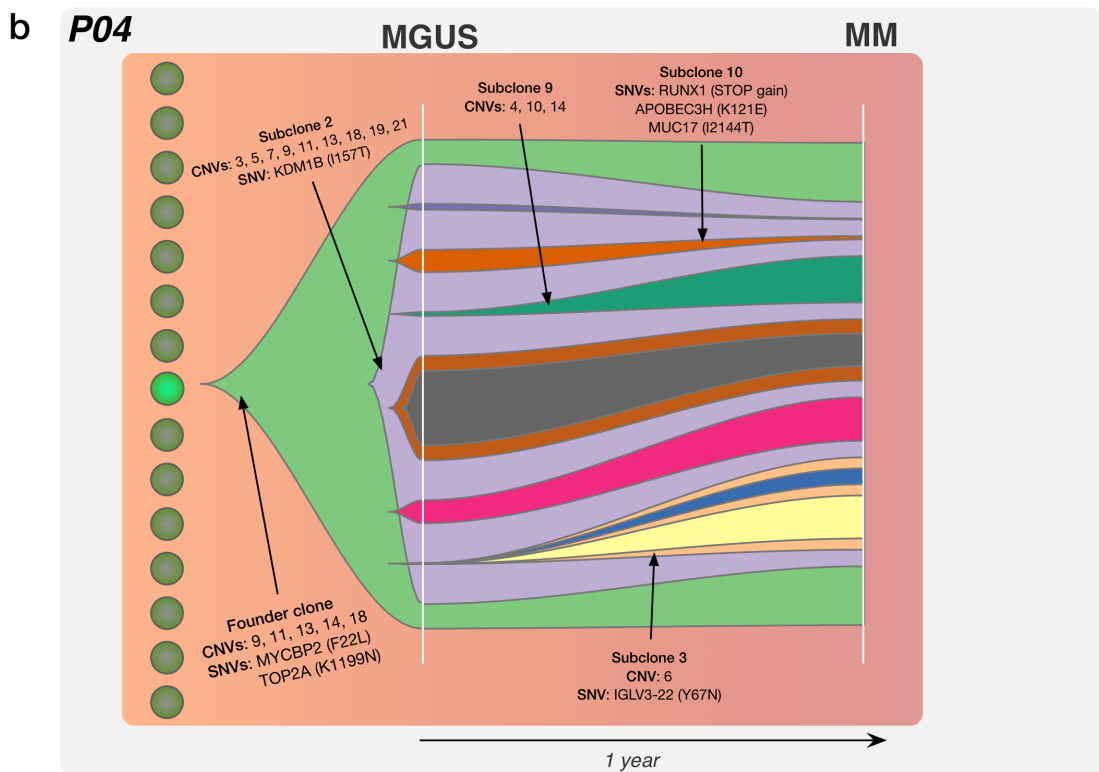
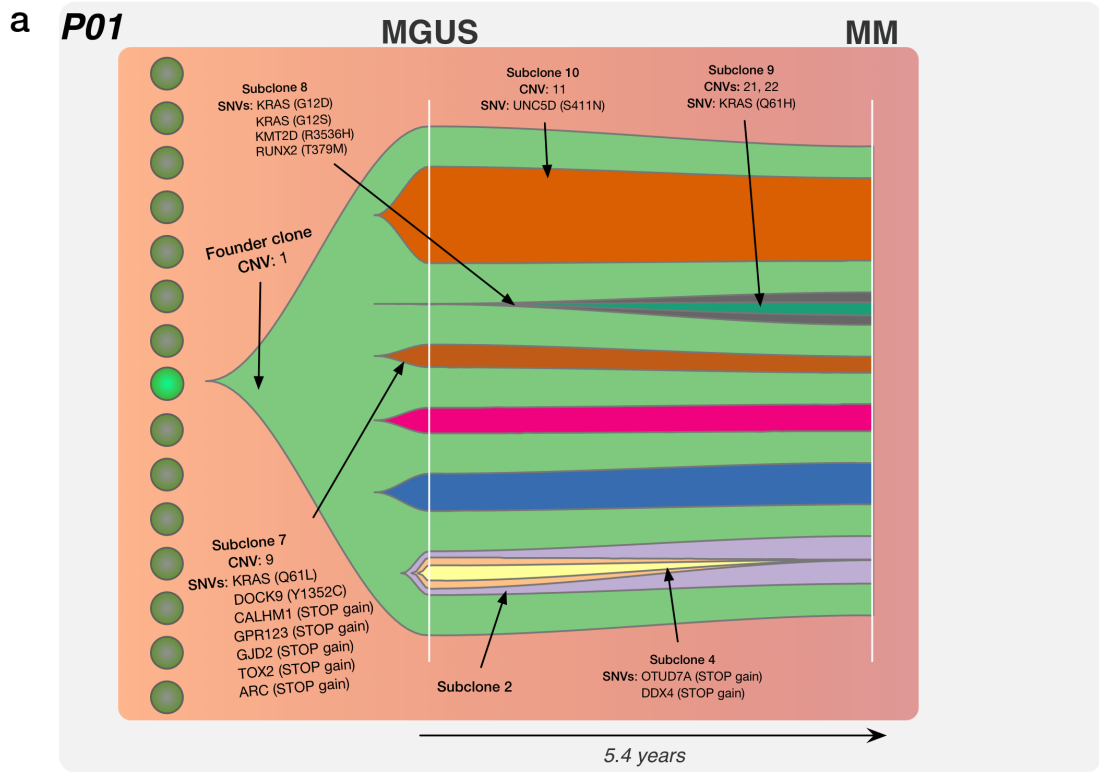
2.4.2.1 Subclonal tumour evolution in MGUS-MM patients

Three patients (P01, P04 and P10) were initially diagnosed with MGUS, and subsequently with MM. Typically, an average of 7 subclones were identified in MGUS sample pairs. We describe two examples, with the full-annotated subclonal architecture for all MGUS-MM patients found in *Supplementary Figure 1/Appendix 1*.

Patient P01 exhibited a modest increase in NS-SNV mutations with progression and was composed of eight subclones at diagnosis. The founder clone had a copy number change on chromosome 1. Interestingly, while P01 mainly exhibited stable progression of subclones from MGUS to MM, we observed *KRAS* mutations to be newly acquired in multiple child subclones. Subclone 7 [brown] harboured a mutation causing an amino acid change at Q61L, with a resultant neutral growth observed. Furthermore, we identified mutations occurring in a nested fashion, with outgrowth of subclone 8 [grey from <1% to ~6%] harbouring mutations at G12D and G12S, with further emergence of child subclone 9 [green] harbouring additional change at Q61H with MM progression. This was coupled with the extinction of child subclonal branches of subclone 2 [purple] [*Figure 3a*].

Patient P04 exhibited an interesting subclonal evolution pattern, where initially one subclone [subclone 2 purple] evolved from the founder clone, which was followed by substantial branching evolution resulting in six child subclones involved in MM progression. The founder clone harboured mutations in *MYCBP2* (F22L) and *TOP2A* (K1199N) and copy number changes on chromosomes 9, 11, 13, 14 and 18. While most of the child subclones exhibit stability, subclone 3 [orange from <1% to ~18%] and subclone 9 [green from <1% to ~9%] appear to have a selective advantage and showed emergence towards MM [*Figure 3b*]. Similarly, P10 was composed of eight subclones at MGUS with the neutral growth of subclonal populations coupled with the emergence of multiple subclones [subclone 4 yellow from ~3% to ~25%, and subclone 5 blue from ~1% to ~10%] and extinction of child subclone 3 [orange from ~5% to <1%] with progression [*Supplementary Figure 1c*].

Figure 3. The subclonal tumour evolution associated with MGUS to MM progression. Fishtail plots illustrate the subclonal architecture in MGUS-MM of two patients (**a**: P01 and **b**: P04), which was defined by the existence of between 5 to 8 PC subclones at MGUS diagnosis. These subclonal populations generally progress to MM in a stable manner, in combination with the coupled emergence and/or extinction of child subclones. Key mutations in the founder clone and subclones are highlighted, with mutations in driver genes identified at both the clonal and subclonal level. The full-annotated subclonal genetic architecture for all patients can be found in Supplementary Figure 1.



2.4.2.2 Subclonal tumour evolution in SMM-MM patients

Five patients (P02, P03, P05, P06 and P08) were diagnosed for SMM, and then subsequently MM at a later time point. Generally, an average of 8 subclones were identified in SMM-MM pairs. We report two examples, with the full-annotated subclonal architecture for all SMM-MM patients found in *Supplementary Figure 2/Appendix 1*.

Patient P02 was composed of eleven subclones at diagnosis and exhibited stable growth during progression with mainly the emergence of child subclone 5 and its branches [blue from ~5% to ~13%] and extinction of subclone 9 [dark green from ~6% to <1%] [*Figure 4a*]. The founder clone showed copy changes on chromosomes 6, 8 and 13, and mutations in *HERPUDI* (STOP gain), *FGFR3* (809S) and *DAPK1* (K435R). Furthermore, we identified a *KRAS* mutation (A146P) in subclone 11, whose population proportion size, interestingly, did not change during MM progression.

Patient P03 displayed an interesting evolution pattern with massive extinction of subclone 2 [purple] from ~47% to ~6%, and almost all of its child subclones, by MM diagnosis. The founder clone harboured mutations in *NOD2* (STOP gain) and CNVs on chromosomes 1, 6, 9, 13 and 16. Furthermore, two individual subclones that contained distinct *DIS3* mutants M566K and R689P were identified at SMM diagnosis in subclone 8 [black] and child subclone 11 [dark purple], respectively [*Figure 4b*]. While recent single cell analysis has demonstrated parallel evolution of the RAS/MAPK pathway in MM through the occurrence of *RAS* mutations in individual clones leading to distinct subclonal populations³⁵, here we uniquely identify parallel evolution of *DIS3*, with the resultant emergence of both subclonal lineages with MM progression. Additionally, subclone 13 and its child subclones exhibited outgrowth with a mutation in *NEK2* (L39H) [light green from <1% to ~7%].

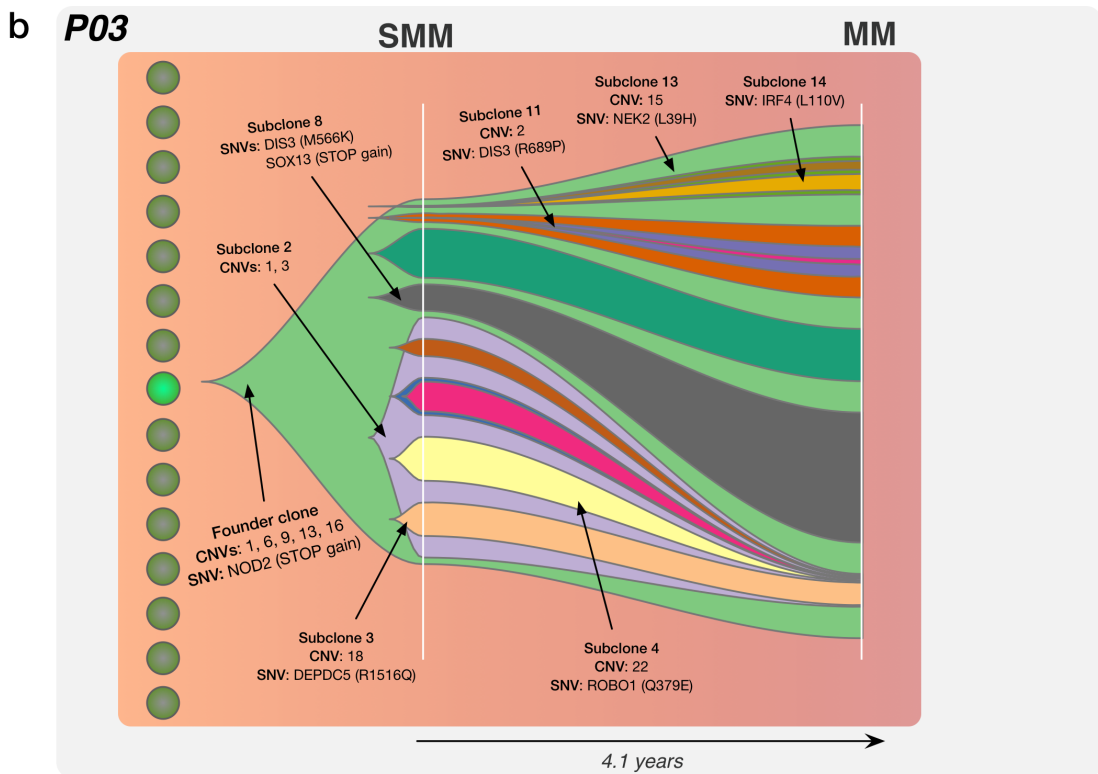
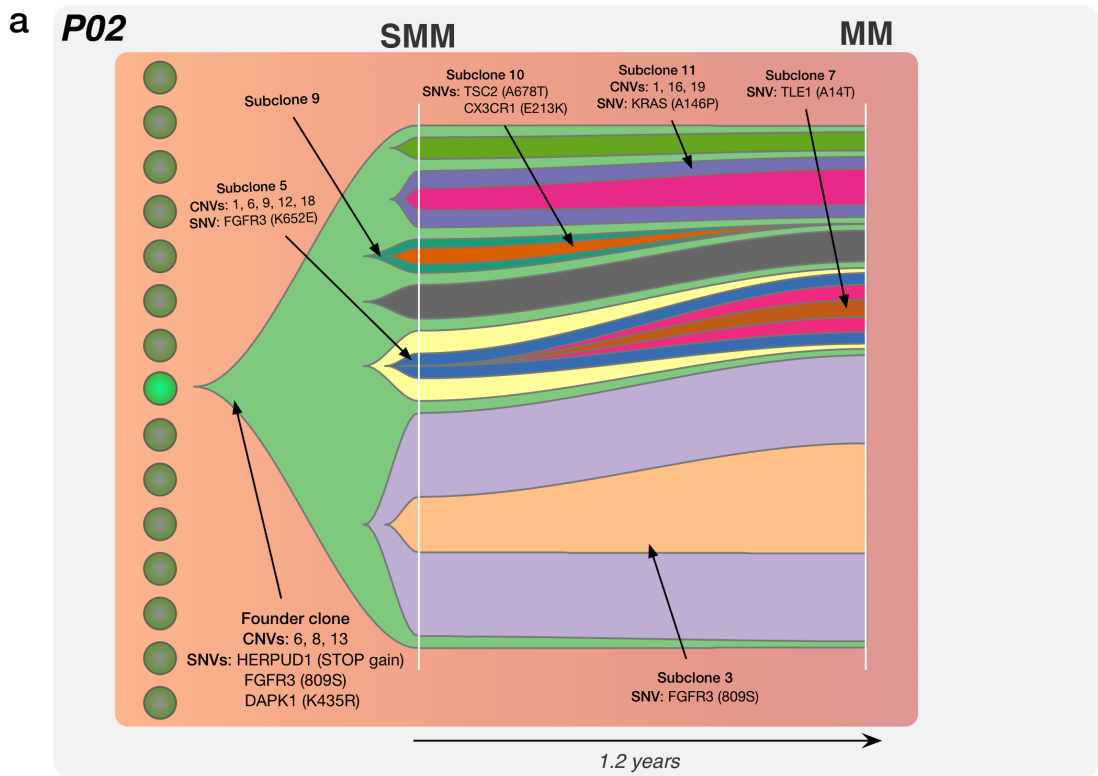
Clonal stability in tumour evolution is also exemplified in other SMM-MM patients [P05, P06, and P08], which were characterised by 7, 5 and 8 subclones at diagnosis, respectively, and exhibited coupled emergence and extinction of child subclones in the progression to MM. In P05, there were initially two subclones that progressed to MM with the emergence and extinction of child clones from subclonal branch 5 [blue], combined with the neutral growth from subclonal branch 2 [purple] [*Supplementary Figure 2c*]. Similarly, in P06, with progression we observed the emergence of subclone 8 [black from <1% to ~23%] and its child subclone 9 [dark green from <1% to ~4%], and subclone 2

[purple from ~8% to ~28%] and its child subclones 6 [pink from <1% to ~10%] and 7 [brown from <1% to ~6%]. The proportions of child subclonal population 3 [orange] remained unchanged between SMM and MM [*Supplementary Figure 2d*]. Patient P08 exhibited neutral growth, which was coupled with the emergence of child subclone 9 [green from <1% to ~5%] and extinction of child subclone 8 [black from ~6% to <1%] with MM progression [*Supplementary Figure 2e*].

Our analysis reveals conclusive evidence of intraclonal heterogeneity and subclonality from the earliest MGUS/SMM stages, where most of the transformed subclonal populations involved in progression to MM are already present at diagnosis. Notably, we do not observe a remarkable difference in the subclonality characteristic at the initial asymptomatic MGUS stage (average 7 subclones) and the intermediate SMM stage (average 8 subclones). This suggests that major subclonal remodelling is also not a phenomenon associated with advancement between the asymptomatic stages.

Figure 4. The subclonal tumour evolution associated with SMM to MM progression.

Fishtail plots illustrating the subclonality in SMM-MM of two patients (**a**: P02 and **b**: P03), with the existence of between 5 to 11 PC subclones at the SMM stage. Notably, in comparison to the subclonal architecture at MGUS diagnosis, we observe a similar number of subclones present at SMM. Similar to the MGUS subclones, these SMM subclonal populations generally progress to MM in a stable manner, in combination with the coupled emergence and/or extinction of child subclones. Key mutations in MM genes in the founder clone and subclones are highlighted, where mutations in driver genes were found to be both clonal and subclonal in nature. The full-annotated subclonal genetic architecture for all patients can be found in Supplementary Figure 2.



2.5 Discussion

The longitudinal investigation of MGUS/SMM to MM samples using NGS has revealed a new understanding of the underlying genetic architecture and subclonal evolution associated with MM progression. Analysis of MGUS-MM, and SMM-MM transition has shown that intraclonal heterogeneity is present at the asymptomatic stages. We find that progression is associated with an altered landscape of acquired mutations, rather than an increased total mutational burden.

Cancer progression models propose either the sequential accumulation of key genetic mutations throughout progressive disease and clonal expansions (“Darwinian” evolution), or punctuated bursts of large-scale chromosomal alterations (“Saltationist” evolution)³⁶. The current understanding of MM transformation involves a sequential nature of evolution from the well-defined asymptomatic stages of MGUS and SMM, characterised by clonal expansion of PCs, and branching “Darwinian” evolution with the presence of 2 to 6 subclones, highlighting clonal heterogeneity at MM presentation^{10,18-20,29,30,35,37-41}. In this model it is recognised that progression from the asymptomatic stages is dependent on the rise and fall in dominance of PC subclones based on their clonal fitness. The acquisition of driver mutations confers a selective advantage and facilitates better survival properties allowing the subclones to survive the selective pressures of the microenvironment/immune system and progress to symptomatic MM.

Notably, our study establishes MM disease progression to be characterised by the phenomenon of clonal stability, where substantial remoulding of the subclonal populations from the asymptomatic stages is not a necessary prerequisite for progression to MM. We found the existence of multiple PC subclones (range 5 to 11) at both MGUS and SMM that were intrinsic in the development and progression of MM. Furthermore, by comparing patients at MGUS and SMM stages we identified no significant difference in the number of PC subclones present at diagnosis (with an average of 7 versus 8, respectively). This is striking, as progression between the asymptomatic stages of MGUS and SMM is currently distinguished by an increased BM PC% and monoclonal protein level. We also found no correlation between the extent of subclonality and BM PC% at the MM stage in patients [Supplementary Figure 3]. Similarly, a recent study of four high risk SMM to MM transformation patients revealed that clonal progression was the key feature of MM onset, where the invasive clinically predominant clone typical of MM, was already present at SMM⁴. Similarly, in their investigation, Walker *et. al.* reported a shifting clonal structure

with the outgrowth and reduction of subclonal populations from SMM to MM⁴. Taken together, we hypothesise that patients who progress within a short time frame, MGUS/SMM to MM transformation does not always required the acquisition of many additional mutations and clonal selection. These MGUS/SMM patients appear to be sufficiently genetically complex to be on the threshold of transformation to MM, which may possibly be driven by extrinsic factors.

Being able to define the crucial oncogenic events in the founder clonal population could facilitate treatment strategies for early intervention to arrest MM progression. However, as the subclones responsible for MM are evident at the asymptomatic stages this poses the question as to why these patients are not symptomatic. A strong possibility is that further to intrinsic genetic factors, extrinsic factors such as the tumour microenvironment may also play an important role in defining both the subclonal architecture and the overall tumour cell burden for progression to clinical malignancy. The complex interactions of the tumour microenvironment with subclones provide signals that may support tumour growth or dormancy, which may influence their transformation⁴²⁻⁴⁹. Of note, a recent study from our group which used the C57BL/KaLwRij mouse model of MM, demonstrated habitual clonal dominance, where only a few establishing MM cells subsequently contributed to tumour burden while most remained dormant. This illustrates the strong selection pressures present within the BM microenvironment which plays a role in defining the clonal architecture⁴⁶.

The current standard of care at the asymptomatic stages involves monitoring patients, with no treatment options until they progress to symptomatic MM. Here our study has demonstrated that there is no significant shift in subclonal structure associated with MM progression. As such, subclonal populations present at MGUS/SMM diagnosis would be just as amenable to treatment, and eradication of these subclonal populations prior to disease transformation could delay progression and may provide the prospect of a durable cure⁵⁰. However, we recognise that intraclonal heterogeneity has been shown to be characteristic of MGUS/SMM/MM, with multiple subclones having differing survival properties, therefore the risk of further mutation and tumour evolution due to drug selection pressures would eventually lead to relapse. Furthermore, intraclonal heterogeneity with clonal selection may not be the only defining evolution associated with progression of MM, with a recent study illustrating the involvement of spatial heterogeneity with regional site seeding and outgrowth resulting in progression⁵¹.

Therefore, a combined longitudinal and spatial study of progression in patients would further elucidate genomic biomarkers of MM tumour evolution⁵², although the ability to sample from multiple sites in asymptomatic patients has significant ethical and practical challenges.

Our findings reveal new insights into the genomic complexity and subclonal tumour evolution that is present from MGUS/SMM through to MM transformation. The existence of subclonality and clonal stability as a model of tumour evolution not only provides a more comprehensive understanding of the underlying biology of MM disease progression, but also new considerations required for patients at diagnosis and future therapeutic approaches to control this disease.

2.6 Acknowledgements

This study was supported by funding from the Australian Government National Health and Medical Research Council (APP1065314). A.K.D was supported by a Leukaemia Foundation of Australia National Research Program PhD Scholarship and Florey Medical Research Foundation Project Grant in Cancer Research for PhD students.

The authors would like to thank the South Australian Cancer Research Biobank (SACRB) for assistance with sourcing patient samples; Dr. Annie Chow for curation of patient samples used in the study; Dr. Stanley Cheung for curation of clinical data of patient samples used in the study; Dr. Randall Grose, SAHMRI Australian Cancer Research Foundation (ACRF) Innovative Imaging Centre for assistance with flow cytometry cell sorting; Lisa Anderson and team from UQDI/QUT for the exome library preparation and sequencing; Jeremy Dumsday, Agilent Technologies and Jenny Peters, Integrated Sciences for the CRE kit; Prof. Maher Gandhi, UQDI, for facilitating access to the MMRF CoMMpass raw sequencing data (dbGaP phs000348, phs000748); and members of the Myeloma Research Laboratory for their support and advice.

2.7 Supplementary

2.7.1 Supplementary Methods

2.7.1.1 Whole Exome Sequencing.

Exome libraries were generated using the Nimblegen KAPA Hyper Library Prep kit (Kapa Biosystems, PN KP-KK8504) followed by the SureSelectXT Clinical Research Exome (CRE) (Agilent, S06588914) capture kit. 115ng of gDNA were used as input for fragmentation on the Covaris E220 followed End-Repair/A-Tailing and ligation of the SureSelect Adapter Oligos, excepting samples which had low input and required additional PCR cycles [*Supplementary Table 6*]. 10 cycles, or 12 cycles in the case low input samples, of Pre-Capture PCR amplification were performed to produce sufficient library for exome capture. Libraries were quantified on the LabChip GX II (LCGXII) using the 5K HT DNA assay (Perkin Elmer, PN 760435 and CLS760675) and 750ng of each sample was input to the Agilent CRE capture workflow, hybridised to the CRE probes overnight. Following the capture washes and 11 cycles of Post-Capture PCR incorporating index barcodes, captured libraries were validated on the LCGXII using the 5K HT DNA assay (Perkin Elmer, PN 760435 and CLS760675). An equimolar pool was prepared from the captured libraries and the pool validated on an Agilent HS DNA Bioanalyzer chip (Agilent, PN 5067-4626), and by qPCR using the KAPA Library Quantification Kits (Kapa Biosystems, PN KK4824), to assess quantity and quality of the samples ready for sequencing. Sequencing was performed on the Illumina HiSeq4000 (2x100 bp paired-end reads) and NextSeq 500 (2x150 bp paired-end reads). Samples were sequenced to a minimum depth of ~140X mean coverage. Isolated non-tumour cells were also sequenced to a similar average depth (138x).

2.7.1.2 Analysis of Whole Exome Sequencing Data.

2.7.1.2.1 Sequence alignment.

Sequencing reads were mapped to the human decoy genome (hs37d5) using Novoalign (v3.02.08) followed by post-processing according to GATK best practices²².

2.7.1.2.1 Somatic variant calling.

Somatic single nucleotide and small indel variants were called using MuTect²³ and multiSNV²⁴. Variants were filtered using the following criteria: 10+ reads covering the variant site; 5+ reads covering the variant in the tumour sample. Variant annotation was performed with SnpEff²⁵.

R 3.3.2 was used throughout for analyses. Somatic copy number variants were called using CNVkit²⁶ v0.7.11 and custom in-house methods developed to support highly aneuploid genomes to perform segmentation and calculate log2 change between matched non-tumour and MGUS/SMM/MM. Log2 changes were corrected for sample purity to calculate ploidy at each stage. To investigate common copy number changes between MGUS/SMM and MM stages, for each patient and each gene the ploidy change was calculated between MGUS/SMM and MM, and a score generated for each gene by calculating the number of patients demonstrating ploidy increase minus the number of patients demonstrating ploidy decrease (ploidy change was threshold at 0.2 copies to reduce noise). Broad copy number changes were compared to cytogenetics data when available to examine concordance. Focal copy number changes were defined as regions < 3Mb in length⁵³

To investigate total copy number change at each stage, the CNVkit segmentation was used. Purity adjusted ploidy changes > 0.2 were summed for each patient. The Y chromosome, Immunoglobulin heavy region on chromosome 14, and T cell receptor A variable regions on chromosomes 7 and 14 were excluded from this analysis as they were frequently hyper-segmented in the CNVkit analysis.

2.7.1.3 Tumour heterogeneity and subclonal evolution.

Clonal evolution was investigated using PhyloWGS²⁷ and visualised using fish plot in R²⁸. It is worth noting that PhyloWGS can inflate the number of subclones so, although we based our analysis on the inferred subclonal architecture, we recognise that the numbers of subclones may be overestimated. There are some mutation discrepancies between the subclonal trees and the SNVs called by MuTect2 due to threshold differences between the two analyses. PhyloWGS requires one time point to demonstrate that the mutation is present at a non trivial level, therefore a lower burden of proof is required to infer its existence between the two time points. However, if both time points demonstrate a low proportion then that mutant is not inferred in the subclonal tree. In our subclonal tumour evolution models, we did not consider polyclonal evolution, where multiple founder PC clones were present at the MGUS/SMM diagnosis stage, due to the computational difficulty in modeling polyclonal evolution. All phylogenetic trees constructed were based on the assumption that there is a single founder PC clone.

2.7.2 Supplementary Appendix 1

2.7.2.1 Subclonal tumour evolution in MGUS-MM patients

Patients P01, P04 and P10 were initially diagnosed with MGUS, and subsequently with MM. We observed that MGUS-MM patients exhibited an average time to progression (TTP) of ~5 years. All three patients showed a decrease in total non-synonymous SNVs associated with progression [Figure 1a]. Patients P04 and P10 were composed of eight and five subclones at MGUS diagnosis, respectively. Patient P04 exhibited an interesting subclonal evolution pattern, where initially one subclone [subclone 2 purple] evolved from the founder clone, which was followed by substantial branching evolution resulting in six child subclones involved in MM progression, and a rapid TTP of 1 year. The founder clone harboured mutations in *MYCBP2* (F22L) and *TOP2A* (K1199N) and copy number changes on chromosomes 9, 11, 13, 14 and 18. While most of the child subclones exhibit stability, subclone 3 [orange] and subclone 9 [green] appear to have a selective advantage and showed emergence towards MM [Supplementary Figure 1b].

In patient P10 we observed a decreased proportion of the founder clone, likely due to some normal PCs contamination [Supplementary Table 2]. We identified neutral growth of subclonal populations coupled with the emergence of multiple subclones [subclone 4 yellow from ~3% to ~25%, and subclone 5 blue from ~1% to ~10%] and extinction of child subclone 3 [orange from ~5% to <1%] with progression. The founder clone showed multiple high impact mutations in *DUSP27* (STOP gain), *SPI40* (F133I) and *FAM110B* (P339L) [Supplementary Figure 1c]. The TTP of P10 was noted to be 13 years, possibly representing an earlier diagnosis and sampling time for this patient.

Patient P01 exhibited a slight decrease in NS-SNV mutations with progression and was composed of eight subclones at diagnosis. The founder clone had a copy number change on chromosome 1. Interestingly, while P01 mainly exhibited stable progression of subclones from MGUS to MM, we observed *KRAS* mutations to be newly acquired in multiple child subclones. Subclone 7 [brown] harboured a mutation causing an amino acid change at position Q61L, with a resultant neutral growth observed. Furthermore, we identified mutations occurring in a nested fashion, with outgrowth of subclone 8 [grey from <1% to ~6%] harbouring mutations at G12D and G12S, with further emergence of child subclone 9 [green] harbouring additional change at Q61H with MM progression. This was coupled with the extinction of child subclonal branches of subclone 2 [purple] [Supplementary Figure 1a].

2.7.2.2. Subclonal tumour evolution in SMM-MM patients

Patients P02, P03, P05, P06 and P08 were diagnosed for SMM, and then subsequently MM at a later time point. SMM patients demonstrated an average TPP of 2 years. Patients were separated on the basis of total non-synonymous SNV burden associated with MM progression, where 3 patients showed a decrease (P02, P05 and P08) and 2 patients showed an increase (P03 and P06) [Figure 1a]. Patients P02 and P06 were composed of eleven and five subclones at diagnosis, respectively. Patient P02 exhibited stable growth during progression, with mainly the emergence of child subclone 5 and its branches [blue from ~5% to ~13%] and extinction of subclone 9 [dark green from ~6% to <1%] [Supplementary Figure 2a]. The founder clone showed copy changes on chromosomes 6, 8 and 13, and mutations in *HERPUD1* (STOP gain), *FGFR3* (809S) and *DAPK1* (K435R). Furthermore, we identified a *KRAS* mutation (A146P) in subclone 11, whose population proportion size, interestingly, did not change during MM progression.

Patient P06 displayed a small founder clone proportion, possibly due to variants such as structural changes unable to be characterised by WES, which harboured copy changes on chromosomes 3, 5, 6, 15, 17, 19 and 21, and point mutation in *NRAS* (Q61R). We mainly observed the emergence of subclone 8 [black from <1% to ~23%] with *KLC3* mutation (R442H) and its child subclone 9 [dark green from <1% to ~4%], and subclone 2 [purple from ~8% to ~28%] and its child subclones 6 [pink from <1% to ~10%] with mutations in *MYCBP2* (E730K), *FGFR3* (A165T) and *PRDMI* (G214R) and 7 [brown from <1% to ~6%] with progression. The proportions of child subclonal population 3 [orange] remained unchanged between SMM and MM [Supplementary Figure 2d].

Patient P08 exhibited neutral growth, which was coupled with the emergence of child subclone 9 [green from <1% to ~5%] and extinction of child subclone 8 [black from ~6% to <1%] with MM progression. The founder clone had widespread mutations with CNVs in chromosome 2, 8, 9, 13, 16, 18, 19, 20 and 22, and SNVs in *RBI* (G449E), *PLEKHA7* (STOP gain), *RBM4B* (STOP gain), *DDX55* (R222Q), *CCDC105* (STOP gain), *HIST1H3J* (STOP gain) and *MLIP* (STOP gain) [Supplementary Figure 2e].

Similar to P06, patient P05 showed a smaller founder clone proportion at diagnosis. There were initially two subclones present at the SMM stage, which progressed to MM with the emergence and extinction of child clones from subclonal branch 5 [blue], combined with the neutral growth from subclonal branch 2 [purple]. Subclone branch 5

[blue] and its child subclonal branches 9 [green] and 10 [light brown] harboured multiple stop mutations in genes *NRG3*, *EZH2*, *KLHL20*, *SNX9*, *C8orf87*, *ACTL6A* and *MTA3*. However, these child subclones became progressively extinct with MM progression, from ~5% to ~2% and ~3% to <1%, respectively. While child subclones 7 [dark brown] and 8 [black] exhibit emergence towards MM, from <1% to ~10% and <1% to ~5%, respectively. The founder clone harboured mutations in *KRAS* (G12V) and *ICAM5* (R85L), and CNVs on chromosome 2 and 19 [*Supplementary Figure 2c*].

Patient P03 displayed an interesting evolution pattern with massive extinction of subclone 2 [purple] from ~47% to ~6%, and almost all of its child subclones, by MM diagnosis. The founder clone harboured mutations in *NOD2* (STOP gain) and CNVs on chromosomes 1, 6, 9, 13 and 16. Furthermore, two individual subclones that contained distinct *DIS3* mutants M566K and R689P were identified at SMM diagnosis in subclone 8 [black] and child subclone 11 [dark purple], respectively [*Supplementary Figure 2b*]. While recent single cell analysis has demonstrated parallel evolution of the RAS/MAPK pathway in MM through the occurrence of RAS mutations in individual clones leading to distinct subclonal populations²⁴, here we uniquely identify parallel evolution of *DIS3*, with the resultant emergence of both subclonal lineages with MM progression. Additionally, subclone 13 and its child subclones exhibited outgrowth with a mutation in *NEK2* (L39H) [light green from <1% to ~7%].

2.7.3 Supplementary Figures

| Patient | Age | Diagnosis | Time To Progression (years) | Hyperdiploid | Translocations | Chromosomal Abberations | CRAB Symptoms |
|---------|-----|-----------|-----------------------------|--------------|----------------|---|--|
| P01 | 72 | MGUS-MM | 5.4 | nil | nil | nil | Bone lesions |
| P02 | 80 | SMM-MM | 1.2 | nil | t(4;14) | +1q,-8p21,-13q14 | Anaemia |
| P03 | 83 | SMM-MM | 4.1 | nil | nil | -13q14 | Pancytopenia |
| P04 | 79 | MGUS-MM | 1 | y | t(8;19) | +3,+5,+7,-8,+9,+11,-13,+15,+15,-18,+der(19) | Anaemia |
| P05 | 61 | SMM-MM | 0.9 | n/a | n/a | n/a | Anaemia |
| P06 | 65 | SMM-MM | 1.8 | nil | nil | nil | Anaemia, Peripheral neuropathy |
| P07 | 59 | MGUS-MM | 3 | nil | nil | +1q | Anaemia, Bone lesions, Renal insufficiency |
| P08 | 95 | SMM-MM | 0.48 | nil | n/a | -13q | Anaemia |
| P09 | 89 | MGUS-MM | 3.2 | nil | nil | nil | Bone lesions, Anaemia |
| P10 | 65 | MGUS-MM | 13 | nil | t(4;14) | +1q | Anaemia |

Supplementary Table 1. Clinical cytogenetic data for MGUS/SMM to MM patients.

Clinically recorded data at MM diagnosis for patients in the study. The median age of patients at MM diagnosis was 75.5 years. Molecular cytogenetics of patients was performed using FISH analysis on interphase spreads of bone marrow smears. Nil represents parameter not being present. N/A represents that data was not available.

| Patient | Purity | | Mean Target Coverage (X) | |
|---------|----------|-------|--------------------------|-----|
| | MGUS/SMM | MM | MGUS/SMM | MM |
| P01 | >0.99 | >0.99 | 179 | 196 |
| P02 | >0.99 | >0.99 | 165 | 154 |
| P03 | >0.99 | >0.99 | 146 | 201 |
| P04 | >0.99 | >0.99 | 140 | 194 |
| P05 | >0.99 | >0.99 | 155 | 149 |
| P06 | >0.99 | >0.99 | 157 | 158 |
| P07 | 0.37 | >0.99 | 1170 | 167 |
| P08 | >0.99 | 0.82 | 159 | 304 |
| P09 | >0.99 | >0.99 | 185 | 205 |
| P10 | 0.66 | >0.99 | 470 | 210 |

Supplementary Table 2. Estimated sample purity and exome sequencing coverage. Purity of FACS sorted patient PCs was assessed by FACS purity check post sort on sorted cells tube, with 100-500 cells through the flow cytometer for each sample. Mean depth of sequencing describing the average number of reads over bases in the targeted exome region of samples sequenced on the HiSeq4000 and NextSeq500.

| <i>Patient</i> | MGUS/SMM | MM | Shared |
|----------------|-----------------|-----------|---------------|
| P01 | 217 | 212 | 42 |
| P02 | 220 | 175 | 111 |
| P03 | 161 | 226 | 48 |
| P04 | 160 | 59 | 44 |
| P05 | 196 | 159 | 83 |
| P06 | 30 | 133 | 10 |
| P07 | 133 | 116 | 4 |
| P08 | 150 | 125 | 91 |
| P09 | 157 | 158 | 50 |
| P10 | 190 | 145 | 20 |

Supplementary Table 3. The shared NS-SNVs between MGUS/SMM and MM, and unique NS-SNVs in MM patients. Analysis of the total mutational load reveals a median of 161 NS-SNVs at MGUS/SMM and 152 NS-SNVs at MM. The table describes the shared and MM unique NS-SNVs identified in each patient.

| | Chr | Position | dbGaP | Ref allele | Alt allele | Impact | cDNA position | Amino acid position |
|--------------------|-----|-----------|-------------|------------|------------|----------|---------------|---------------------|
| <i>DIS3</i> | | | | | | | | |
| P03 MM | 13 | 73337650 | | C | G | Moderate | c.2066 G>C | p.Arg689Pro |
| | 13 | 73345100 | | A | T | Moderate | c.1697 T>A | p.Met566Lys |
| P07 MM | 13 | 73336078 | | A | T | Moderate | c.2325 T>A | p.Phe775Leu |
| P10 MM | 13 | 73337651 | | G | C | Moderate | c.2065 C>G | p.Arg689Gly |
| <i>NRAS</i> | | | | | | | | |
| P06 SMM | 1 | 115256529 | rs11554290 | T | C | Moderate | c.182 A>G | p.Gln61Arg |
| P06 MM | 1 | 115256529 | rs11554290 | T | C | Moderate | c.182 A>G | p.Gln61Arg |
| <i>KRAS</i> | | | | | | | | |
| P01 MGUS | 12 | 25380275 | rs17851045 | T | G | Moderate | c.183 A>C | p.Gln61His |
| P01 MM | 12 | 25380275 | rs17851045 | T | G | Moderate | c.183 A>C | p.Gln61His |
| | 12 | 25380276 | rs121913240 | T | A | Moderate | c.182 A>T | p.Gln61Leu |
| | 12 | 25398284 | rs121913529 | C | T | Moderate | c.35 G>A | p.Gly12Asp |
| | 12 | 25398285 | rs121913530 | C | T | Moderate | c.34 G>A | p.Gly12Ser |
| P02 SMM | 12 | 25378562 | | C | G | Moderate | c.436 G>C | p.Ala146Pro |
| P02 MM | 12 | 25378562 | | C | G | Moderate | c.436 G>C | p.Ala146Pro |
| P05 SMM | 12 | 25378561 | | G | A | Moderate | c.437 C>T | p.Ala146Val |
| P05 MM | 12 | 25378561 | | G | A | Moderate | c.437 C>T | p.Ala146Val |
| | 12 | 25398284 | rs121913529 | C | A | Moderate | c.35 G>T | p.Gly12Val |
| P07 MM | 12 | 25380275 | rs17851045 | T | G | Moderate | c.183 A>C | p. Gln61His |
| P09 MM | 12 | 25380275 | rs17851045 | T | A | Moderate | c.183 A>T | p.Gln61His |
| P10 MM | 12 | 25398284 | | C | A | Moderate | c.35 G>T | p.Gly12Val |

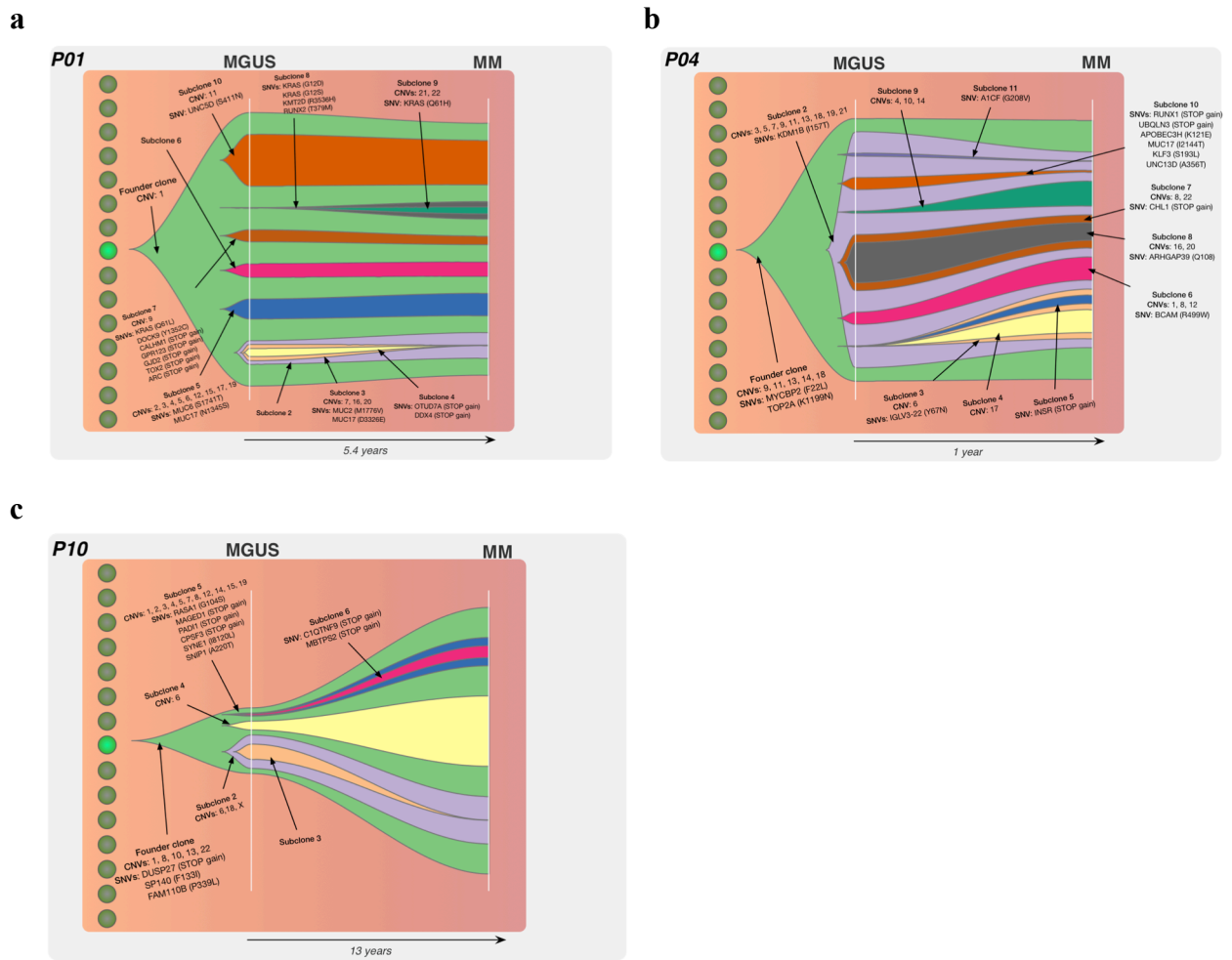
Supplementary Table 4. The full characterisation of driver mutations in MGUS/SMM to MM patients. Single nucleotide variants in previously reported driver genes were identified in *KRAS*, *NRAS* and *DIS3*. The table describes the genomic positions and subsequent impact on cDNA and amino acid changes.

| MGUS/SMM | | MM | |
|----------|------|---------|------|
| Patient | CNVs | Patient | CNVs |
| P01 | 19 | P01 | 65 |
| P02 | 67 | P02 | 65 |
| P03 | 50 | P03 | 70 |
| P04 | 114 | P04 | 73 |
| P05 | 97 | P05 | 55 |
| P06 | 71 | P06 | 103 |
| P07 | 88 | P07 | 43 |
| P08 | 110 | P08 | 78 |
| P09 | 69 | P09 | 78 |
| P10 | 26 | P10 | 49 |

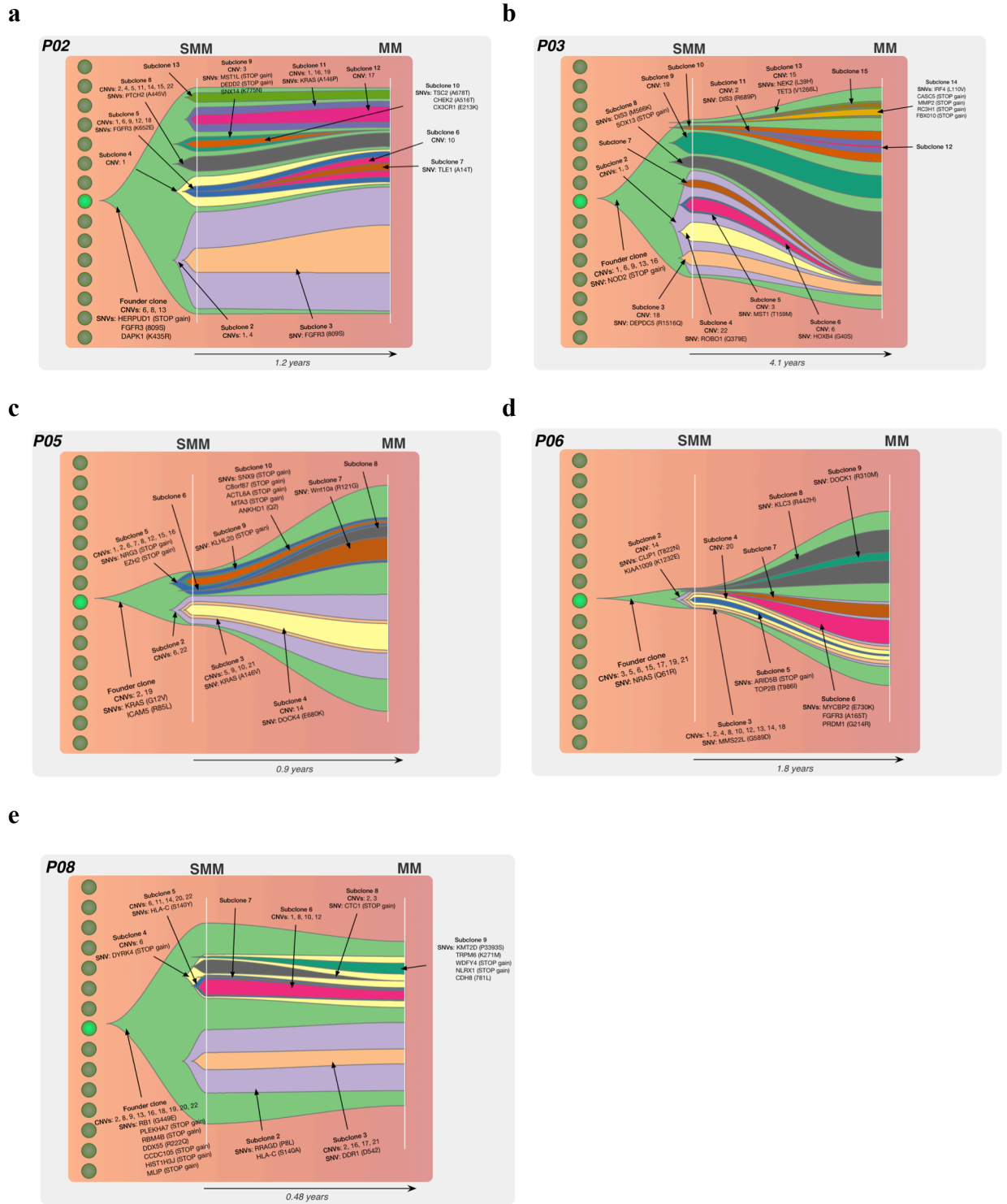
Supplementary Table 5. The copy number landscape of MGUS/SMM to MM patients. We identified numerous CNV changes in each patient at MGUS/SMM and MM. MGUS/SMM patients harboured a higher median number of changes than at MM (70 vs. 67.5, respectively).

| Patient | Input (ng) | PCR Cycles |
|----------|------------|------------|
| P07 NT | 67.77 | 12 |
| P10 MGUS | 30.24 | 12 |

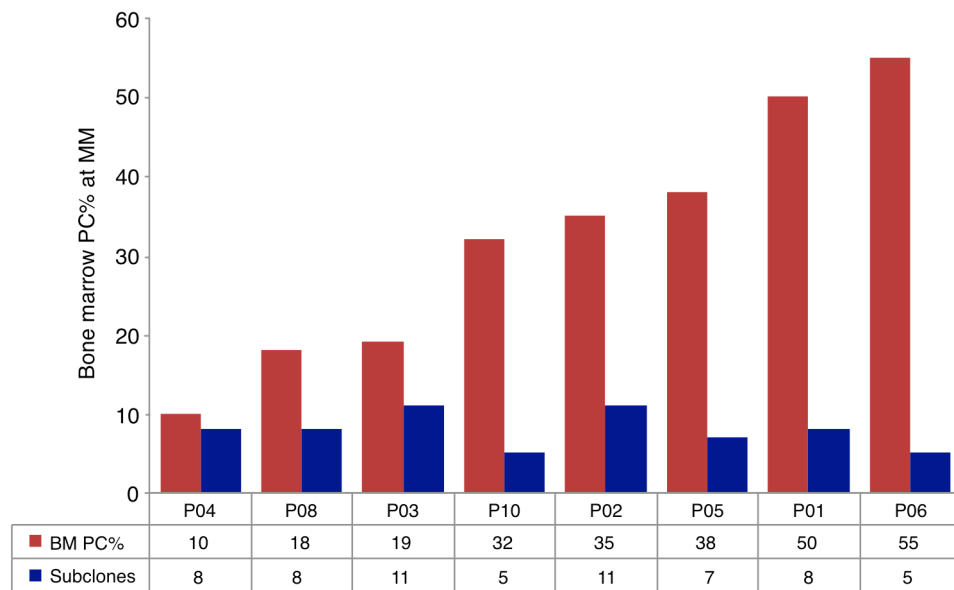
Supplementary Table 6. Exome library preparation of low input samples requiring additional PCR cycles. Two samples had yields lower than the required 115ng gDNA input and required extra PCR amplification during library preparation to generate sufficient library for exome capture.



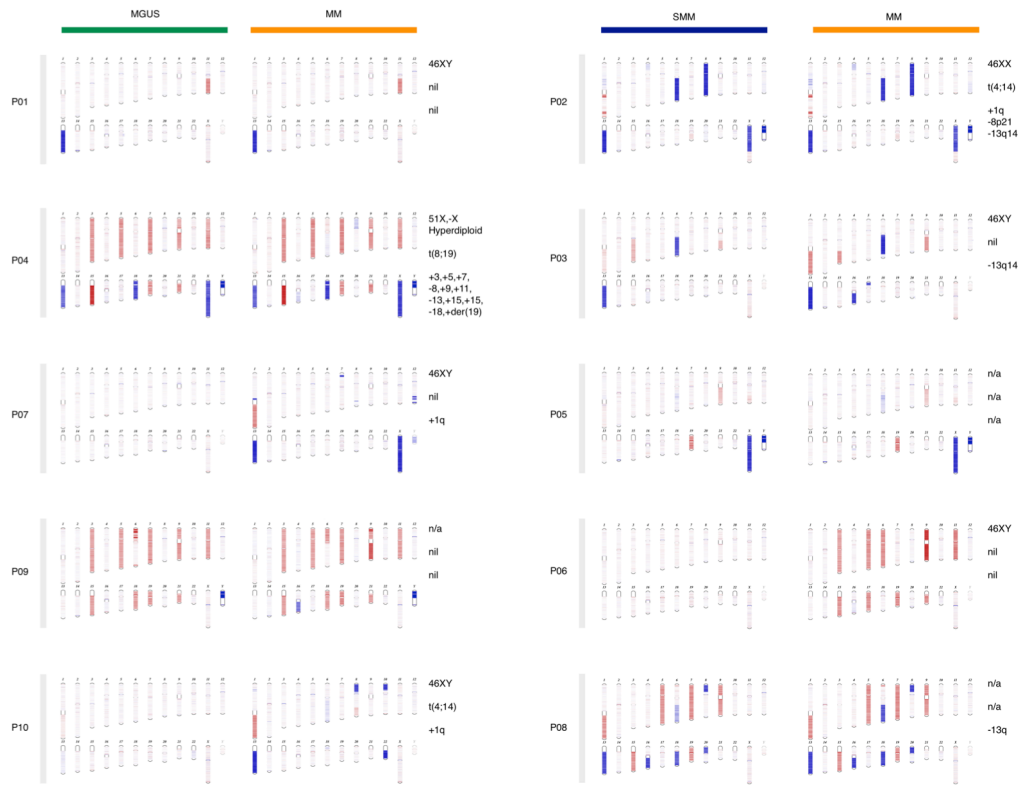
Supplementary Figure 1. The subclonal tumour evolution associated with MGUS to MM progression. Fishtail plots annotated with the complete subclonal genetic architecture in three patients (**a**: P01, **b**: P04 and **c**: P10) from Figure 3 of the main article.



Supplementary Figure 2. The subclonal tumour evolution associated with SMM to MM progression. Fishtail plots annotated with the complete subclonal genetic architecture in five patients (a: P02, b: P03, c: P05, d: P06 and e: P08) from Figure 4 of the main article.



Supplementary Figure 3. The correlation of BM PC% and the subclones identified with progression in MGUS/SMM to MM patients. Progression to MM is characterised by an increase in BM PC% and monoclonal protein levels, however, we find no correlation between the extent of subclonality and BM PC% at MM.



Supplementary Figure 4. A comparison of cytogenetic abnormalities with virtual karyotypes. Virtual karyotypes were generated from genome-wide copy number changes inferred from whole exome sequencing data at both MGUS/SMM and MM stages. Chromosomal copy amplifications are illustrated by red, while copy deletions are shown in blue. Several patients exhibit very similar karyotypes at both MGUS/SMM and MM. In some patients, hyperdiploidy is present at MGUS/SMM and is undetected by standard cytogenetics even at MM (molecular cytogenetics results listed to the right of each patient figure). Nil represents parameter not being present. N/A represents that data was not available.

2.8 References

1. Rajkumar SV. Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management. *Am J Hematol.* 2016;91(7):719-734.
2. Kumar SK, Dispenzieri A, Lacy MQ, et al. Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia.* 2014;28(5):1122-1128.
3. Smith D, Yong K. Advances in understanding prognosis in myeloma. *Br J Haematol.* 2016;175(3):367-380.
4. Walker BA, Wardell CP, Melchor L, et al. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia.* 2014;28(2):384-390.
5. Palumbo A, Anderson K. Multiple myeloma. *N Engl J Med.* 2011;364(11):1046-1060.
6. Landgren O, Kyle RA, Pfeiffer RM, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood.* 2009;113(22):5412-5417.
7. Kyle RA, Therneau TM, Rajkumar SV, et al. A long-term study of prognosis in monoclonal gammopathy of undetermined significance. *N Engl J Med.* 2002;346(8):564-569.
8. Kyle RA, Remstein ED, Therneau TM, et al. Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. *N Engl J Med.* 2007;356(25):2582-2590.
9. Fonseca R, Bailey RJ, Ahmann GJ, et al. Genomic abnormalities in monoclonal gammopathy of undetermined significance. *Blood.* 2002;100(4):1417-1424.
10. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature.* 2011;471(7339):467-472.
11. Pawlyn C, Morgan GJ. Evolutionary biology of high-risk multiple myeloma. *Nat Rev Cancer.* 2017;17(9):543-556.
12. Dutta AK, Hewett DR, Fink JL, Grady JP, Zannettino ACW. Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma. *Br J Haematol.* 2017;178(2):196-208.

13. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol*. 2017;14(2):100-113.
14. Szalat R, Munshi NC. Genomic heterogeneity in multiple myeloma. *Curr Opin Genet Dev*. 2015;30:56-65.
15. Brioli A, Melchor L, Cavo M, Morgan GJ. The impact of intra-clonal heterogeneity on the treatment of multiple myeloma. *Br J Haematol*. 2014;165(4):441-454.
16. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313.
17. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. *Nat Rev Cancer*. 2012;12(5):335-348.
18. Bolli N, Avet-Loiseau H, Wedge DC, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*. 2014;5:2997.
19. Lohr JG, Stojanov P, Carter SL, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. 2014;25(1):91-101.
20. Walker BA, Boyle EM, Wardell CP, et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol*. 2015;33(33):3911-3920.
21. Vandyke K, Zeissig MN, Hewett DR, et al. HIF-2alpha Promotes Dissemination of Plasma Cells in Multiple Myeloma by Regulating CXCL12/CXCR4 and CCR1. *Cancer Res*. 2017;77(20):5452-5463.
22. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 10 11-33.
23. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219.
24. Josephidou M, Lynch AG, Tavare S. multiSNV: a probabilistic approach for improving detection of somatic point mutations from multiple related tumour samples. *Nucleic Acids Res*. 2015;43(9):e61.
25. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.

26. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol.* 2016;12(4):e1004873.
27. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* 2015;16:35.
28. Miller CA, McMichael J, Dang HX, et al. Visualizing tumor evolution with the fishplot package for R. *BMC Genomics.* 2016;17(1):880.
29. Egan JB, Shi CX, Tembe W, et al. Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood.* 2012;120(5):1060-1066.
30. Walker BA, Wardell CP, Melchor L, et al. Intracлонаl heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. *Blood.* 2012;120(5):1077-1086.
31. Weissbach S, Langer C, Puppe B, et al. The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Br J Haematol.* 2015;169(1):57-70.
32. Lopez-Corral L, Sarasquete ME, Bea S, et al. SNP-based mapping arrays reveal high genomic complexity in monoclonal gammopathies, from MGUS to myeloma status. *Leukemia.* 2012;26(12):2521-2529.
33. Mikulasova A, Wardell CP, Murison A, et al. Somatic mutation spectrum in monoclonal gammopathy of undetermined significance indicates a less complex genomic landscape compared to multiple myeloma. *Haematologica.* 2017.
34. Zhao S, Choi M, Heuck C, et al. Serial exome analysis of disease progression in premalignant gammopathies. *Leukemia.* 2014;28(7):1548-1552.
35. Melchor L, Brioli A, Wardell CP, et al. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia.* 2014;28(8):1705-1715.
36. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell.* 2017;168(4):613-628.
37. Lopez-Corral L, Gutierrez NC, Vidriales MB, et al. The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells. *Clin Cancer Res.* 2011;17(7):1692-1700.

38. Keats JJ, Chesi M, Egan JB, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood*. 2012;120(5):1067-1076.
39. Magrangeas F, Avet-Loiseau H, Gouraud W, et al. Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*. 2013;27(2):473-481.
40. Kortum KM, Langer C, Monge J, et al. Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann Hematol*. 2015;94(7):1205-1211.
41. Weston-Bell N, Gibson J, John M, et al. Exome sequencing in tracking clonal evolution in multiple myeloma following therapy. *Leukemia*. 2013;27(5):1188-1191.
42. Dhodapkar MV. MGUS to myeloma: a mysterious gammopathy of underexplored significance. *Blood*. 2016;128(23):2599-2606.
43. Ghobrial IM, Detappe A, Anderson KC, Steensma DP. The bone-marrow niche in MDS and MGUS: implications for AML and MM. *Nat Rev Clin Oncol*. 2018.
44. Noll JE, Williams SA, Purton LE, Zannettino AC. Tug of war in the haematopoietic stem cell niche: do myeloma plasma cells compete for the HSC niche? *Blood Cancer J*. 2012;2:e91.
45. Manier S, Sacco A, Leleu X, Ghobrial IM, Roccaro AM. Bone marrow microenvironment in multiple myeloma progression. *J Biomed Biotechnol*. 2012;2012:157496.
46. Hewett DR, Vandyke K, Lawrence DM, et al. DNA Barcoding Reveals Habitual Clonal Dominance of Myeloma Plasma Cells in the Bone Marrow Microenvironment. *Neoplasia*. 2017;19(12):972-981.
47. Lawson MA, McDonald MM, Kovacic N, et al. Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. *Nat Commun*. 2015;6:8983.
48. Das R, Strowig T, Verma R, et al. Microenvironment-dependent growth of preneoplastic and malignant plasma cells in humanized mice. *Nat Med*. 2016;22(11):1351-1357.
49. Mitsiades CS, Mitsiades NS, Munshi NC, Richardson PG, Anderson KC. The role of the bone microenvironment in the pathophysiology and therapeutic management of multiple myeloma: interplay of growth factors, their receptors and stromal interactions. *Eur J Cancer*. 2006;42(11):1564-1573.

50. Barlogie B, Mitchell A, van Rhee F, Epstein J, Morgan GJ, Crowley J. Curing myeloma at last: defining criteria and providing the evidence. *Blood*. 2014;124(20):3043-3051.
51. Rasche L, Chavan SS, Stephens OW, et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat Commun*. 2017;8(1):268.
52. Bustoros M, Mouhieddine TH, Detappe A, Ghobrial IM. Established and Novel Prognostic Biomarkers in Multiple Myeloma. *Am Soc Clin Oncol Educ Book*. 2017;37:548-560.
53. Krijgsman O, Carvalho B, Meijer GA, Steenbergen RD, Ylstra B. Focal chromosomal copy number aberrations in cancer-Needles in a genome haystack. *Biochim Biophys Acta*. 2014;1843(11):2698-2704.

Chapter 3

Transcriptomic and DNA methylation analyses in MGUS/SMM to MM progression

Statement of Authorship

| | |
|---------------------|---|
| Title of Paper | Transcriptomic and DNA methylation analyses in MGUS/SMM to MM progression |
| Publication Status | <input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style |
| Publication Details | Dutta, A. K. , Mayne, B.T., Harliwong, I., Morgan, G. J., To, L. B., Hewett, D. R., Fink, J. L. & Zannettino, A. C. W. (2018) Transcriptomic and DNA methylation analyses in MGUS/SMM to MM progression. <i>Manuscript in preparation.</i> |

Principal Author

| | |
|--------------------------------------|---|
| Name of Principal Author (Candidate) | Ankit K.Dutta |
| Contribution to the Paper | Experimental design Generation of data Analysis and interpretation of data Manuscript development, writing and review |
| Overall percentage (%) | 65% |
| Certification: | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. |
| Signature | <div style="display: flex; justify-content: space-between;"> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 15%; text-align: center;">Date</div> <div style="border-bottom: 1px solid black; width: 5%; text-align: center;">31 / 7 / 2018</div> </div> |

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | |
|---------------------------|---|
| Name of Co-Author | Benjamin T. Mayne |
| Contribution to the Paper | Analysis of data Manuscript review |
| Signature | <div style="display: flex; justify-content: space-between;"> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 15%; text-align: center;">Date</div> <div style="border-bottom: 1px solid black; width: 5%; text-align: center;">31/7/2018</div> </div> |

| | |
|---------------------------|--|
| Name of Co-Author | Ivon Harliwong |
| Contribution to the Paper | Experimental design Sequencing library preparation |
| Signature | <div style="display: flex; justify-content: space-between;"> <div style="border-bottom: 1px solid black; width: 80%;"></div> <div style="border-bottom: 1px solid black; width: 15%; text-align: center;">Date</div> <div style="border-bottom: 1px solid black; width: 5%; text-align: center;">15/07/18</div> </div> |

| | | | |
|---------------------------|--|------|---------|
| Name of Co-Author | Gareth J. Morgan | | |
| Contribution to the Paper | Grant funding application Manuscript review | | |
| Signature | | Date | 7.11.18 |

| | | | |
|---------------------------|--|------|---------|
| Name of Co-Author | Luen B. To | | |
| Contribution to the Paper | Grant funding application Manuscript review | | |
| Signature | | Date | 11.7.18 |

| | | | |
|---------------------------|---|------|-----------|
| Name of Co-Author | Duncan R. Hewett | | |
| Contribution to the Paper | Experimental design Supervision of work Manuscript development and review | | |
| Signature | | Date | 26-7-2018 |

| | | | |
|---------------------------|--|------|------------|
| Name of Co-Author | J. Lynn Fink | | |
| Contribution to the Paper | Experimental design Analysis of data Manuscript review | | |
| Signature | | Date | 2018-07-11 |

| | | | |
|---------------------------|--|------|------------|
| Name of Co-Author | Andrew C. W. Zannettino | | |
| Contribution to the Paper | Experimental design Supervision of work Manuscript development and review Project funding | | |
| Signature | | Date | 31/07/2018 |

Chapter 3: Transcriptomic and DNA methylation analyses in MGUS/SMM to MM progression

Ankit K. Dutta^{1,2}, Benjamin T. Mayne², Ivon Harliwong³, Gareth J. Morgan⁴, Luen B. To^{5,6}, Duncan R. Hewett^{1,2}, J. Lynn Fink^{3*} and Andrew C.W. Zannettino^{1,2*}

¹Myeloma Research Laboratory, Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA, 5005, Australia.

²Cancer Theme, South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, 5000, Australia.

³Genomic Medicine Division, The University of Queensland, Diamantina Institute (UQDI), Brisbane, QLD, 4102, Australia.

⁴The Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, 72205, USA

⁵SA Pathology, Adelaide, SA, 5000, Australia.

⁶Haematology and Bone Marrow Transplant Unit, Royal Adelaide Hospital, Adelaide, SA, 5000, Australia.

*co-senior authors

Running title: Transcriptomic and methylomic landscape in MGUS/SMM to MM progression

Keywords: MGUS, Smouldering MM, Myeloma, Next generation sequencing, RNAseq, Gene expression, Whole genome bisulphite sequencing, DNA Methylation

3.1 Abstract

Multiple myeloma (MM) is a largely incurable haematological malignancy characterised by the uncontrolled proliferation of neoplastic plasma cells (PCs) within the bone marrow. Recent studies have focused on the investigation of the genetic landscape in MM and its asymptomatic stages of monoclonal gammopathy of undetermined significance (MGUS) and smouldering multiple myeloma (SMM). While genetic analyses of patient samples have identified recurrently mutated genes and clonal heterogeneity which are characteristic of MM, our understanding of the transcriptomic and methylomic changes associated with progression from MGUS/SMM to MM remains poor. Here, we have performed RNA sequencing (n = 6) and whole genome bisulphite sequencing (n = 4) on serial samples from patients who progressed from MGUS or SMM to MM, to analyse the gene expression and methylation changes associated with disease transformation. Our findings suggest that progression from MGUS/SMM to MM was accompanied by relatively few changes in gene expression. Of the 250 differentially expressed genes that approached statistical significance, the majority of genes showed down regulation upon transition to MM. The top 10 differentially expressed genes included, *THEMIS2*, *BTBD19*, *HBB*, *ATP8A2*, *CELSR1*, *CD69*, *TWF2*, *SLC20A1*, *ALGIL* and *SLC23A3*. Mutated genes, previously identified in whole exome sequencing analyses of the same patients, were found to be expressed at low levels or not at all. In most cases, only the wild type allele of a gene harbouring heterozygous mutation was expressed. Analysis of the methylome revealed significant DNA hypomethylation in MGUS, SMM and MM PCs compared to normal PCs. Extreme DNA hypomethylation was acquired at the initiation of MGUS, and maintained with progression to SMM and MM. Our study suggests that most of the genomic changes of MM occur during the oncogenic transition from a normal PC to a MGUS/SMM PC, with minimal changes in the gene expression and DNA methylation landscape accompanying with the progression to MM.

3.2 Introduction

Multiple myeloma (MM) is a haematological malignancy characterised by the clonal expansion of neoplastic plasma cells (PCs) within the bone marrow. MM is a genetically complex disease, characterised by heterogeneity that influences the disparate treatment and survival outcomes of patients¹⁻⁶. Despite recent advances in therapeutic strategies, MM remains a largely incurable disease, with relapse being a common occurrence⁷.

The initiation of MM involves a multistep transformational process, evolving from the asymptomatic stages of monoclonal gammopathy of undetermined significance (MGUS) and smouldering multiple myeloma (SMM)⁶. Genomic studies of large patient cohorts using next generation sequencing (NGS) techniques have compared the DNA mutational landscape between the asymptomatic stages of MGUS, SMM, and symptomatic MM, identifying recurrently mutated genes (“drivers”) and establishing intraclonal genetic heterogeneity and clonal evolution patterns^{1-3,5,6}. While many studies have performed molecular cytogenetic and NGS analysis to characterise the genetic architecture of MM, the role of transcriptome and methylome changes associated with disease progression are poorly understood. Identification of mutations in genes which alter gene expression, influencing molecular pathways and oncogenic signalling, could inform the treatment strategies used for MM patients^{8,9}.

To date, analysis of the MM transcriptome has relied on array-based technologies which provide a global snapshot of the gene expression profile (GEP) in individual tumours for risk stratification and prognosis of patients¹⁰⁻¹². GEPs have been used to define molecular heterogeneity and classify patients based on common expression signatures into 7 distinct subgroups of MM: MF [*MAF/MAFB*], MS [*MMSET*], CD-1 and CD-2 [*CCND1/CCND3*], HY [hyperdiploid], LB [low bone disease] and PR [proliferation]¹³. Studies have also investigated the association of expression profiles with disease progression, resulting in robust GEP signatures for stratification of high-risk MM patients, such as the UAMS-70¹² and EMC-92¹¹. However, due to the marked heterogeneity that characterises MM, GEPs only provide a broad-brush insight into the disease biology and major clones associated with MM progression. Thus, with the rapid advances in genomic technologies, GEP may not represent the best methodology to identify clinically relevant expression changes in patients. A recent study of MM patient samples using RNA sequencing (RNAseq) has demonstrated allele-specific expression of mutated genes in

MM¹⁴. This illustrates the need for the application of more current technologies such as RNAseq for transcriptomic interrogation of MM samples.

Epigenetic mechanisms are known to play an integral role in regulating gene expression¹⁵. However, the key epigenetic mechanisms of DNA methylation and histone modification contribute to transformation in MM is not well understood. The rate of epigenetic change in cancers has been estimated to be orders of magnitude higher than that of genetic change, and could be a major determinant of clonal evolution¹⁶. Studies investigating the methylome in unmatched MGUS, SMM and MM patient samples in comparison to normal PCs, have illustrated that initiation of disease is characterised by global hypomethylation, which either increases or reduces upon disease progression¹⁷⁻²⁰. Moreover, MM PCs were observed to exhibit extreme heterogeneity in DNA methylation patterns compared to MGUS PCs¹⁷.

As current sequencing studies have focused on the genetic changes and intraclonal heterogeneity in MM, what remains unknown is the expression changes of recurrently mutated genes and their implications for MM transformation. Here, we report an integrative analysis of differential DNA methylation and gene expression in paired MGUS/SMM to MM patient samples. These patients were also the subjects of our previous whole exome sequencing analysis study, which identified clonal stability as a model of tumour evolution in MM²¹.

Longitudinal analysis using RNAseq was carried out on paired MGUS-MM (n = 2), or SMM-MM (n = 4) PCs to assess the expression changes associated with MM transformation. Additionally, investigation of the methylome was carried out using whole genome bisulphite sequencing (WGBS) on normal PCs (NPCs: n = 3), and paired MGUS-MM PCs (n = 1) or SMM-MM PCs (n = 3) for a greater understanding of the underlying methylomic landscape, and regulation of transcriptome associated with the progression to MM.

3.3 Materials & Methods

3.3.1 Clinical samples.

Bone marrow mononuclear cell aspirates were collected from patients at MGUS/SMM, and subsequently at later diagnosis of MM (MGUS-MM (n = 2) and SMM-MM (n = 4)). Samples were collected from patients prior to treatment. All patients provided informed consent in accordance with the Declaration of Helsinki. Samples were cryopreserved by the South Australian Cancer Research Biobank (SACRB) at SA Pathology. The studies were approved by the Royal Adelaide Hospital Human Research Ethics Committee (HREC/13/RAH/569 No: 131133).

3.3.2 Cell sorting.

PCs were purified using multicolour flow cytometry as previously described²². Briefly, approximately 1×10^5 mononuclear cells were prepared for single stain antibody control (CD138-PE (Beckman Coulter #A54190) and CD38-PE-Cy7 (Biolegend #303515)) and compensation/FMO tubes (1: unstained; 2: hydroxystilbamidine (FluoroGold; Life Technologies) only; 3: CD38-PE-Cy7+FluoroGold; 4: CD138-PE+FluoroGold; and 5: CD38-PE-Cy7+CD138-PE). The sort sample was stained with CD138-PE and CD38-PE-Cy7 antibodies at $1 \mu\text{L}/100 \mu\text{L}$ cells. Cells were stained with FluoroGold immediately prior to sorting. Viable PCs (CD138⁺CD38⁺ and FluoroGold negative) were sorted on the FACS Aria Fusion sorter (BD Biosciences). FACS purity check was carried out on sorted cells, using 100-500 cells from each sample.

3.3.3 Nucleic acids isolation and QC.

DNA was isolated from purified PC populations using the All Prep DNA/RNA Micro Kit (QIAGEN) as per manufacturers' instructions. Yields and quality was assessed using the NanoDrop 8000 and Qubit 2.0 fluorometer (Thermo Fisher Scientific).

RNA was isolated from purified PC using the All Prep DNA/RNA Micro Kit (QIAGEN) and RNAAqueous-Micro Kit (Thermo Fisher Scientific) as per manufacturers' instructions. Yields and quality was assessed using the Qubit 2.0 fluorometer (Thermo Fisher Scientific) and BioAnalyzer 2100 (Agilent). The median RIN scores of extracted RNA samples was 6.6.

3.3.4 RNA sequencing and Analysis.

Between 33 to 380ng of RNA was used to generate RNA libraries using the NEXTflex Rapid RNA-Seq Kit (BIOO Scientific) according to manufacturers' instructions. Sequencing was carried out on the Illumina NextSeq500 (2x75 bp paired-end reads) with approximately 90 million reads per sample.

RNA sequencing reads were aligned to the human reference genome (hg19) using the STAR 2-pass method as previously described²³. Variant calling was performed using the GATK best practices²⁴ and were annotated using ANNOVAR²⁵. For gene expression analyses, read counts were quantified using Rsubread²⁶. The TMM method within edgeR²⁷ was used to normalise the expression data, and FactoMineR²⁸ was used for principle component analysis. The Benjamini Hochberg method was used to assess potential false discovery from multiple comparison testing²⁹. The PANTHER (Protein Analysis Through Evolutionary Relationships, <http://www.pantherdb.org>) classification system was used to interpret pathway level analysis³⁰⁻³².

3.3.5 Whole Genome Bisulphite Sequencing and Analysis

Approximately 50ng of DNA was used for bisulphite library preparation using the Ovation Methyl-Seq System (NuGen) according to manufacturers' instructions. Sequencing was performed on the Illumina NextSeq500 (2x150bp paired-end reads). Samples were sequenced to an average coverage of ~30x.

Prior to alignment, reads were trimmed to a minimum of 15bp to improve mapping efficiency and to remove any contaminating adaptor sequences using AdapterRemoval v2.1.7. The human reference genome (GRCh37) was obtained from the Illumina igenomes link and was converted into the bisulphite sequence using Bismark v0.18.1. Trimmed reads were aligned to the bisulphite converted reference genome using Bismark v0.18.1 and Bowtie2 v2.2.5 v, which was also used for cytosine methylation calling. Differential methylation analyses were performed using BiSeq. Differentially methylated regions (DMRs) were considered to be significant if they were within clusters with a false discovery rate (FDR) < 0.05. Principle component analysis was performed using the smoothed methylation values of CpG sites within the top 100 DMRs using FactoMineR.

3.4 Results

3.4.1 The progression of MGUS/SMM to MM is associated with minimal variation in gene expression.

To characterise the transcriptomic landscape associated with progression to MM, we performed RNAseq on 6 paired MGUS/SMM to MM patient sample sets, with an average of approximately 90 million reads per sample. The median time to progression of MGUS to MM was 3.2 years (range 1 – 5.4 years), and SMM to MM was ~1 year (range 0.5 – 4.1 years). The median age of MM diagnosis in the cohort was 79.5 years, which was higher than the established median of 65 years³³ [*Supplementary Table 1*].

To date, our understanding of the transcriptomic landscape of MM is derived from the GEP analyses of unmatched MGUS/SMM/MM samples isolated from different patients³⁴⁻³⁶. As such, current studies are affected by the significant intrinsic natural variation in gene expression patterns that exist between unrelated individuals. To overcome this limitation of inter-patient transcriptional “noise” and identify changes in gene expression associated with the natural history of disease transformation, we analysed matched MGUS/SMM to MM samples. Indeed, principal component analysis revealed that while each patient exhibited great variability in their overall expression patterns, paired samples from each individual patient clustered closely together [*Figure 1*]. Pairwise XY scatterplot analysis of gene expression in individual MGUS to MM, or SMM to MM patients demonstrated little variation between each stage of disease [*Figure 2*]. The median correlation coefficient across all MGUS/SMM to MM cases was 0.915, indicating relative homogeneity in gene expression pattern in each paired sample case. Overall, this shows that there is limited change in the expression of genes associated with the progression of MGUS/SMM to MM. Furthermore, this highlights the limitation of previous studies using unmatched patient sample comparisons, where the majority of gene expression differences identified would be occurring in grouped samples from MGUS/SMM/MM from different patients. Similarly, a previous expression study of normal PCs (NPCs) in comparison to PCs from MGUS and MM patient samples using microarray analysis has revealed that most gene expression changes occur during the initiation of MGUS (i.e. NPCs vs. MGUS). These studies also showed that the expression differences between MGUS and MM are much smaller than that between NPCs and MGUS or MM PCs, with only 74 differentially expressed genes distinguishing MGUS from MM³⁶. Taken together, these data suggest that most of the transcriptomic changes of MM occur during the aberrant transition of NPCs to MGUS/SMM.

Figure 1. The transcriptomic landscape associated with MGUS/SMM to MM transformation. Principal component analysis illustrates intra-patient clustering of expression profiles in the progression of MGUS/SMM to MM in individual patients. MGUS/SMM stage is represented by circles, and MM stage is represented by triangles.

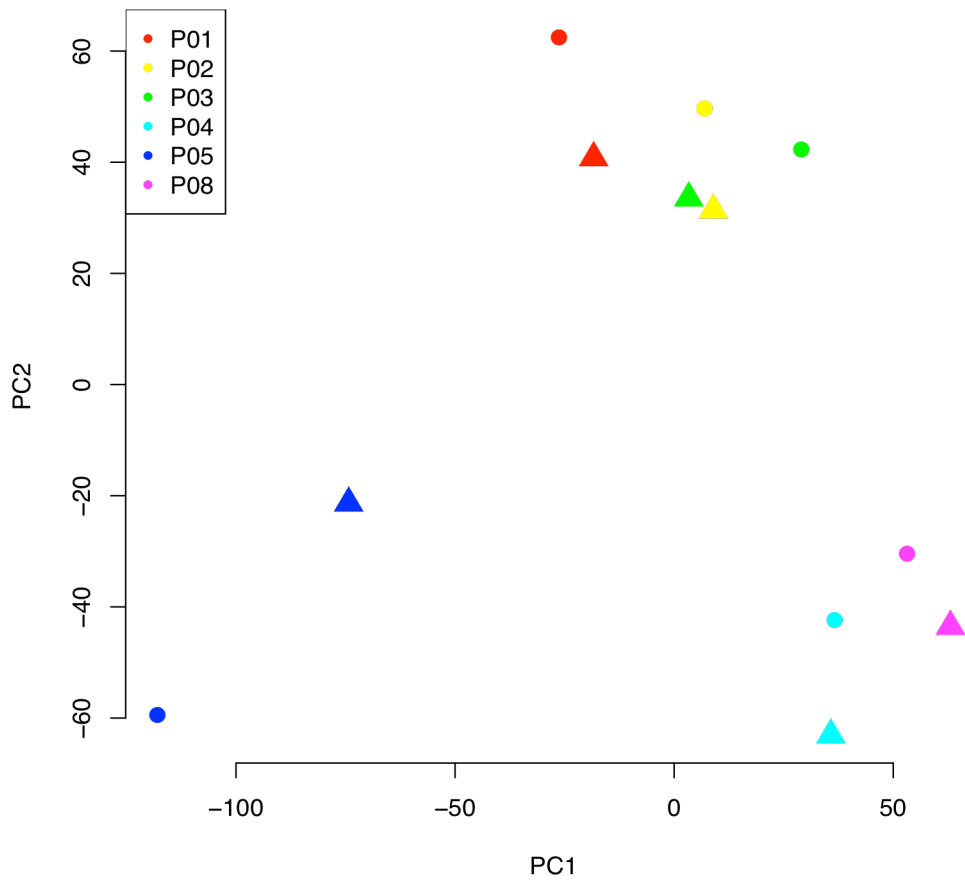
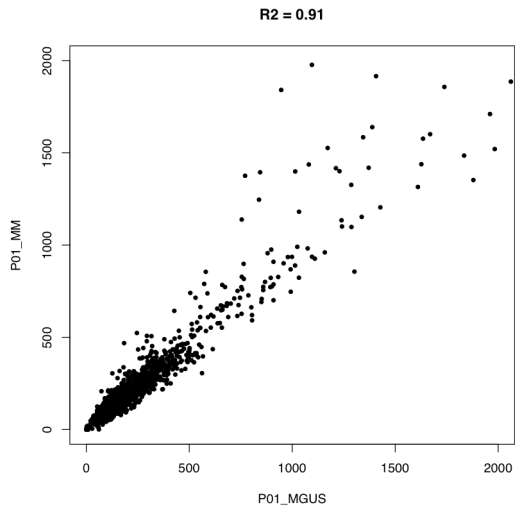
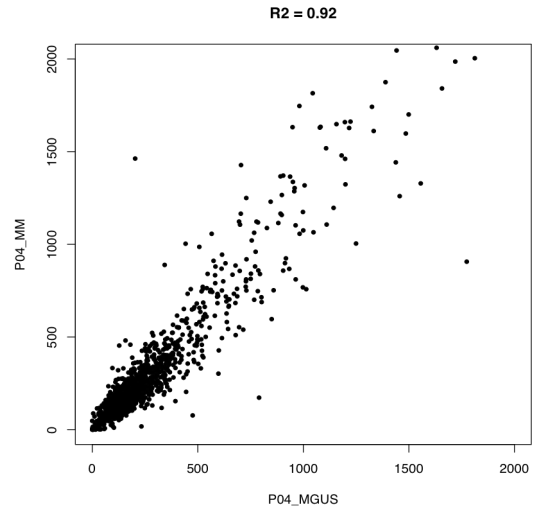
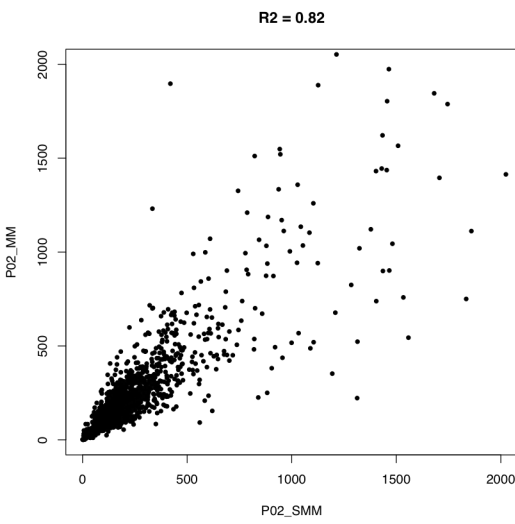
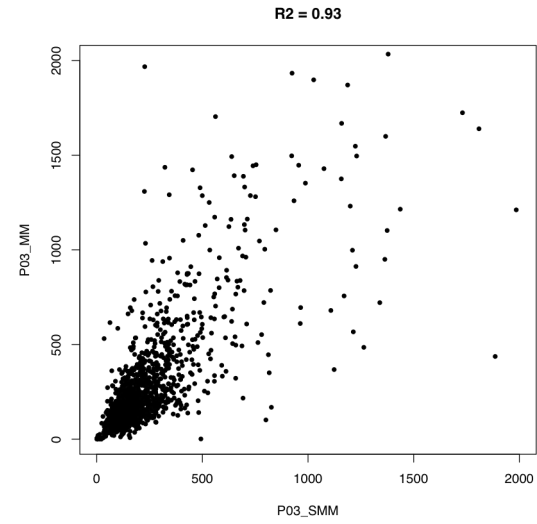
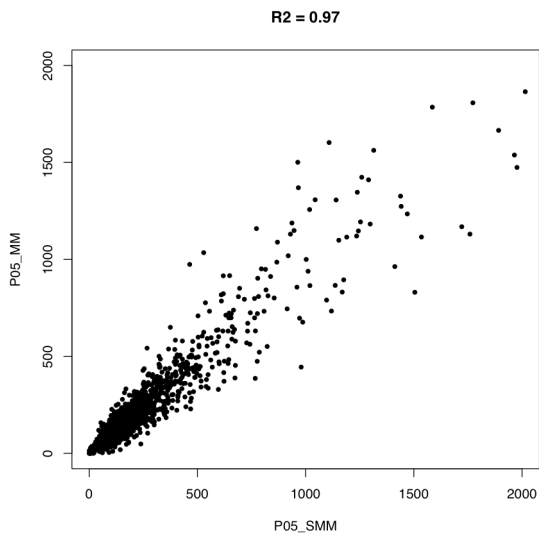
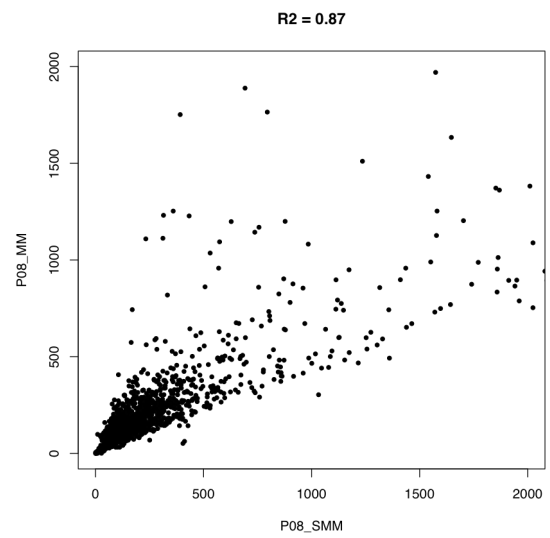


Figure 2. The relationship of gene expression changes associated with MGUS/SMM to MM progression. XY scatterplots illustrate the similarity in gene expression between the asymptomatic stages of MGUS/SMM and MM in paired MGUS-MM (**a**: P01, **b**: P04) and SMM-MM patients (**c**: P02, **d**: P03, **e**: P05, **f**: P08). The median R^2 value was 0.915, demonstrating homogeneity in expression between disease stages. R^2 correlation values are given for each patient above individual plots.

a**b****c****d****e****f**

3.4.2 MGUS/SMM to MM progression is associated with specific expression changes.

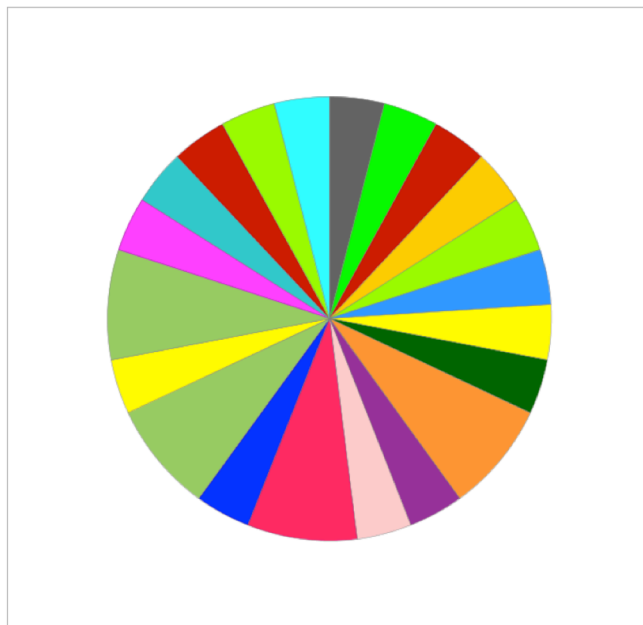
The clustering of matched samples allowed gene expression changes associated with transition that would otherwise be masked by inter-patient variability. Differential gene expression testing between MGUS/SMM and MM groups, identified 250 genes reaching statistical significance (i.e. raw p value < 0.05, before controlling the false discovery rate using the Benjamini Hochberg method²⁹) [*Supplementary Table 3*]. Overall, the majority of differential expression changes involved a reduction in gene expression upon MM transformation, with only 109 genes that were upregulated [*Supplementary Table 3a*], and 141 genes that were downregulated [*Supplementary Table 3b*]. The mean log fold change of upregulated genes was +1.65, while the mean log fold change of downregulated genes was -2.35. These findings highlight that of the group of genes that do change at transition, there is predominantly a downregulation in gene expression, which is consistent with previous GEP studies of MM³⁶. Pathway analysis revealed gene expression signatures associated with the transition of MGUS/SMM to MM, with 196 differentially expressed genes affecting 69 biological pathways [*Supplementary Figure 1*]. Overall, we did not observe pronounced pathway enrichment due to the similar transcriptional profiles identified between the MGUS/SMM and MM stages. Of genes that showed highly increased expression (> +2 fold change) upon progression to MM, there were 5 affected pathways including apoptosis signalling, angiogenesis, Wnt signalling, cadherin signalling and Alzheimer disease-presenilin pathways [*Figure 3a*]. Genes that were highly down regulated (> -2 fold change) upon MM transition, associated with 21 affected molecular pathways, including Metabotropic glutamate receptor group I and III, inflammation mediated by chemokine and cytokine signalling and heterotrimeric G-protein signalling-Gi alpha and Gs alpha mediated pathways [*Figure 3b*].

(a)



- Alzheimer disease-presenilin pathway (P00004)
- Angiogenesis (P00005)
- Apoptosis signaling pathway (P00006)
- Cadherin signaling pathway (P00012)
- Wnt signaling pathway (P00057)

(b)

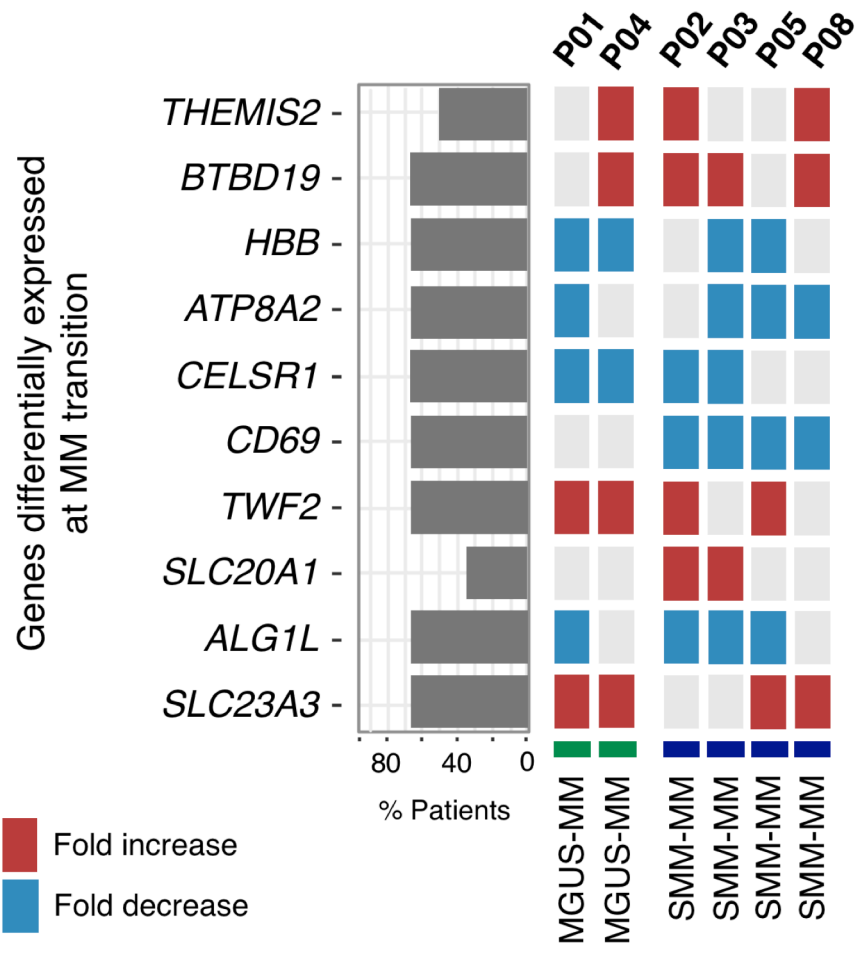


- 5-Hydroxytryptamine degradation (P04372)
- Alzheimer disease-amyloid secretase pathway (P00003)
- B cell activation (P00010)
- Cadherin signaling pathway (P00012)
- Endothelin signaling pathway (P00019)
- GABA-B receptor II signaling (P05731)
- Gonadotropin-releasing hormone receptor pathway (P06664)
- Hedgehog signaling pathway (P00025)
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)
- Huntington disease (P00029)
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031)
- Ionotropic glutamate receptor pathway (P00037)
- Metabotropic glutamate receptor group I pathway (P00041)
- Metabotropic glutamate receptor group II pathway (P00040)
- Metabotropic glutamate receptor group III pathway (P00039)
- Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042)
- Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043)
- Synaptic vesicle trafficking (P05734)
- Transcription regulation by bZIP transcription factor (P00055)
- Wnt signaling pathway (P00057)

Figure 3. Deregulation of molecular pathways associated with MGUS/SMM to MM progression. (a) The upregulation of genes with differential expression of > 2 fold, was associated with 5 biological pathways including apoptosis signalling [red], angiogenesis [yellow], Wnt signalling [blue], cadherin signalling [orange] and Alzheimer disease-presenilin [green]. (b) Genes that were highly downregulated, with differential expression of > -2 fold upon MM transition, were most associated with biological pathways including Metabotropic glutamate receptor group I and III [dark greens], inflammation mediated by chemokine and cytokine signalling [red] and heterotrimeric G-protein signalling-Gi alpha and Gs alpha mediated pathways [orange].

Investigating MGUS-MM and SMM-MM transition on an individual patient-by-patient basis, we identified dynamic fold changes in gene expression associated with progression. We found genes whose expression levels demonstrated fold increases (2x), or fold decreases (0.5x) associated with MM progression. The top 10 genes differentially expressed, based on fold change and frequency, between MGUS/SMM and MM included *THEMIS2*, *BTBD19*, *HBB*, *ATP8A2*, *CELSR1*, *CD69*, *TWF2*, *SLC20A1*, *ALGIL* and *SLC23A3* [Figure 4]. At MGUS to MM progression, we observed fold increases in expression of *THEMIS2*, *BTBD19*, *TWF2* and *SLC23A3*. Conversely, fold decreases in the expression levels of *ATP8A2*, *ALGIL*, *HBB* and *CELSR1* were identified. In the progression of SMM to MM, we identified fold increases in the expression of *THEMIS2*, *BTBD19*, *TWF2*, *SLC20A1* and *SLC23A3*. Fold decreases were found in the expression of *HBB*, *ATP8A2*, *CELSR1* and *ALGIL*. Interestingly, all SMM-MM patients exclusively showed fold decreases in *CD69* expression. Due to the nature of our small and rare sample size, differentially expressed genes between MGUS/SMM and MM samples did not reach statistical significance. These findings reveal a number of genes and pathways deregulated upon MM progression, and highlight the need to perform a comprehensive larger cohort study with validated sample sets.

Figure 4. Waterfall diagram illustrating fold changes in gene expression associated with MGUS/SMM to MM progression. The top 10 differentially expressed genes, based on fold change and frequency, upon MM transition showing fold increase of at least 2x (*THEMIS2*, *BTBD19*, *TWF2*, *SLC20A1* and *SLC23A3*) or fold decrease of at least 0.5x (*HBB*, *ATP8A2*, *CELSR1*, *CD69* and *ALG1L*).

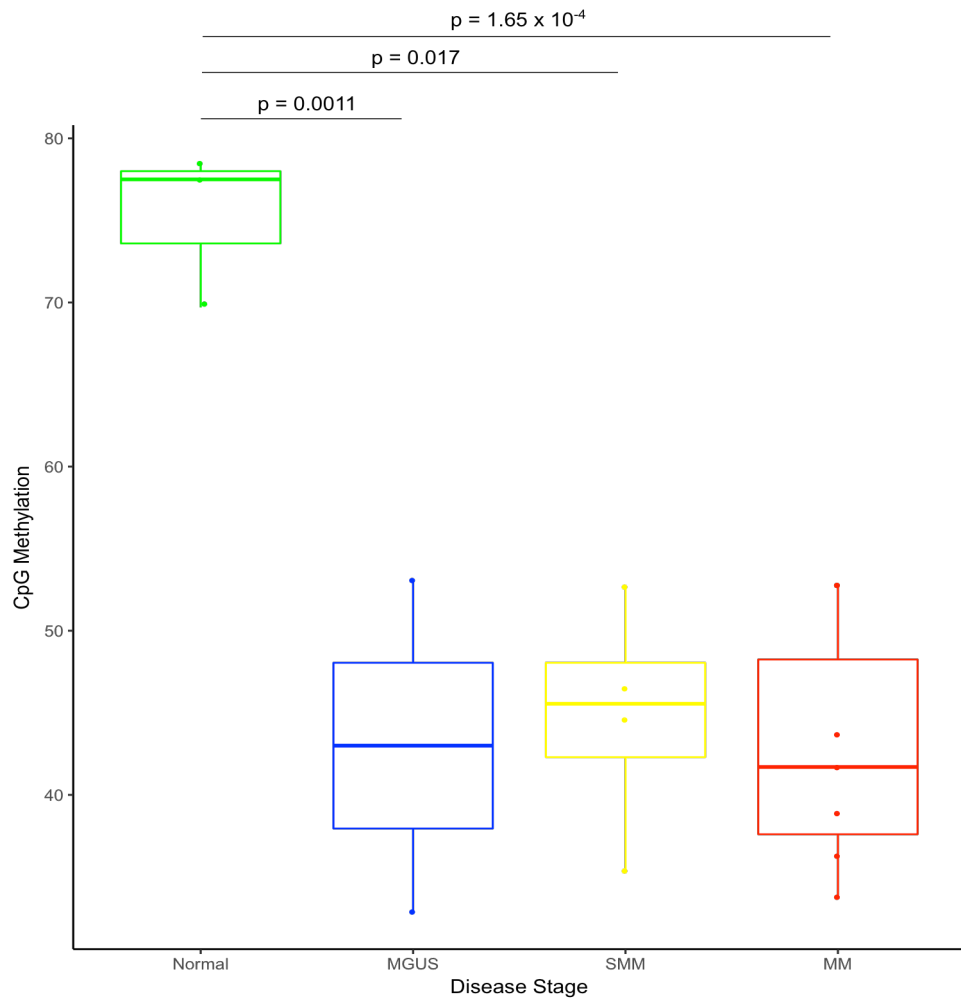


3.4.3 The progression of MGUS/SMM to MM is characterised by the maintenance of hypomethylation acquired at the asymptomatic disease stage.

We next examined the DNA methylation landscape associated with MM transformation in paired MGUS/SMM to MM patient samples (n = 4) and NPCs (n = 3) [*Supplementary Table 2*]. WGBS was performed to a depth of approximately 30x to assess the methylation profiles of MGUS, SMM, MM and NPC samples. For disease stage specific comparison, additional single samples with sufficient DNA from our WES cohort were included: 1 MGUS sample (P11), 1 SMM sample (P05), and 3 MM samples (P04, P06 and P10) [*Supplementary Table 2*].

Analysis of the methylation status of MGUS/SMM and MM PCs in comparison to NPCs revealed that both the asymptomatic and symptomatic stages of disease were characterised by extreme DNA hypomethylation. The average methylation of CpG sites across NPC samples was 77.5%, which is consistent with levels previously reported of ~80% CpG methylation in mammalian cells³⁷⁻³⁹. However, a significant decrease in CpG methylation was observed at each stage of disease, with MGUS, SMM and MM exhibiting an average methylation of 43% (NPC vs. MGUS $p = 0.0011$), 45.6% (NPC vs. SMM $p = 0.017$) and 41.7% (NPC vs. MM $p = 0.000165$), respectively [*Figure 5*]. Interestingly, we observed that in the progression from MGUS to SMM and MM, a similar range of hypomethylation was maintained as that initially acquired at MGUS. These findings are in stark contrast to previous studies of the MM methylome in unmatched samples, which have identified global hypomethylation as a key feature in the transition of MGUS to MM¹⁹, with disease transformation characterised by a progressive increase in hypomethylation¹⁷⁻¹⁹. A previous study of unmatched MGUS (n = 16) and MM (n = 104) PCs compared to NPCs (n = 3) demonstrated that 98.3% of CpG sites in MM PCs were hypomethylated, while 73.2% were hypomethylated in MGUS PCs¹⁷.

Figure 5. The initiation of MGUS is characterised by significant hypomethylation, which is maintained through progression to SMM and MM. Analysis of the overall methylome in asymptomatic MGUS/SMM and symptomatic MM in comparison to non-disease, demonstrates disease onset and progression is characterised by significant hypomethylation of CpG sites. While an average of 77.5% of CpG sites were methylated in NPCs, an average of 43%, 45.6% and 41.7% were methylated at MGUS, SMM and MM, respectively (NPC vs. MGUS $p = 0.0011$, NPC vs. SMM $p = 0.017$, NPC vs. MM $p = 0.000165$). Interestingly, the range of methylation remains constant between MGUS/SMM and MM patients (MGUS vs. MM $p = 0.09911$, SMM vs. MM $p = 0.6843$).



CpG sites were smoothed into clusters that were then used to identify differentially methylated regions (DMRs) between sample groups. Differential methylation testing comparing NPCs with MGUS/SMM, and MM PCs, identified a total of 190,401 DMRs between MGUS/SMM and NPCs, and 177,535 DMRs between MM and NPCs. Thus, there was only a small decrease in the overall number of DMRs observed between NPCs and MGUS/SMM, and NPCs and MM PCs. Overall, all statistically significant DMRs identified between NPCs and MGUS/SMM/MM were located on chromosomes 14, 16, 17, 18, 20, 21, 22, 3, 4, 5, 8, Y and X. Interestingly, however, upon progression from MGUS/SMM to MM, we found no statistically significant DMRs (MGUS vs. MM $p = 0.9911$, and SMM vs. MM $p=0.6843$). Analysis of the change in methylation near promoters within 2000bp of the transcription start site of all genes found within DMRs, revealed that the average level of methylation decreased in MGUS/SMM/MM near the transcription start site of genes as it does for NPCs, whereas NPCs maintained a higher level of methylation [Figure 6].

Investigating the methylome profile of patients on an individual basis, principal component analysis revealed intra-patient clustering [Figure 7]. CpG sites within the top 100 DMRs were plotted as a representative analysis of the relationship between the asymptomatic stages of MGUS/SMM and MM in each patient. Pairwise XY scatterplot comparisons of paired MGUS/SMM to MM samples illustrated minimal variation in the methylation profiles associated with disease progression [Figure 8]. A median correlation coefficient of 0.90 was calculated, demonstrating a strong positive linear relationship and homogeneity in methylation patterns. Conversely, stage specific comparison of NPCs with MGUS, SMM or MM demonstrated considerable heterogeneity of methylation, with a median correlation coefficient of 0.36, highlighting the greater difference in CpG methylation patterns between the transition of NPCs to MGUS/SMM/MM [Supplementary Figure 2].

Figure 6. Promoter methylation analysis of all genes identified within DMRs between NPCs and MGUS/SMM and MM PCs. Comparison of the average level of methylation within 2000bp of the transcription start site of genes within DMRs, between disease stages and NPCs, demonstrates a similar drop in the average level of methylation towards the promoter of genes in MGUS, SMM and MM. BS is the transcription binding site.

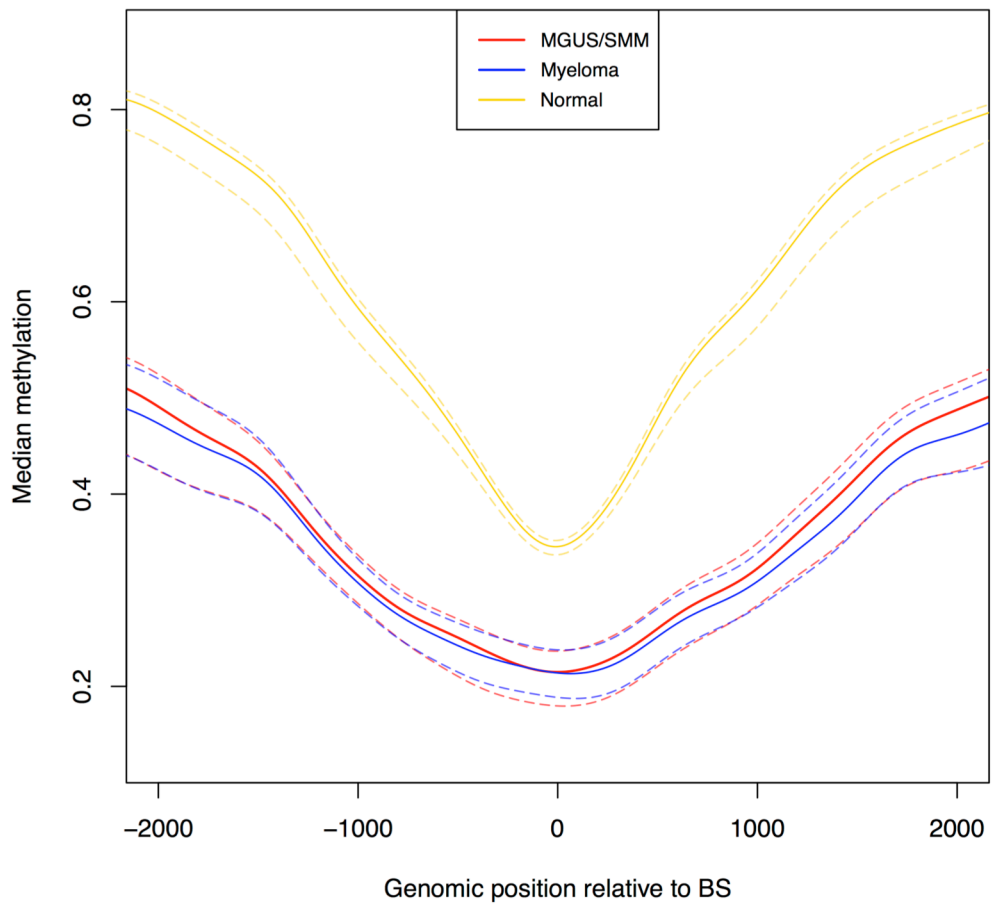


Figure 7. The methylomic landscape associated with MGUS/SMM to MM transformation. Principal component analysis of the methylome profiles of MGUS/SMM to MM patients demonstrates clustering of paired samples from individual patients, with no significant differential methylation associated with progression to MM. Normal samples are represented by squares; MGUS/SMM group is represented by circles; MM group is represented by underlined circles.

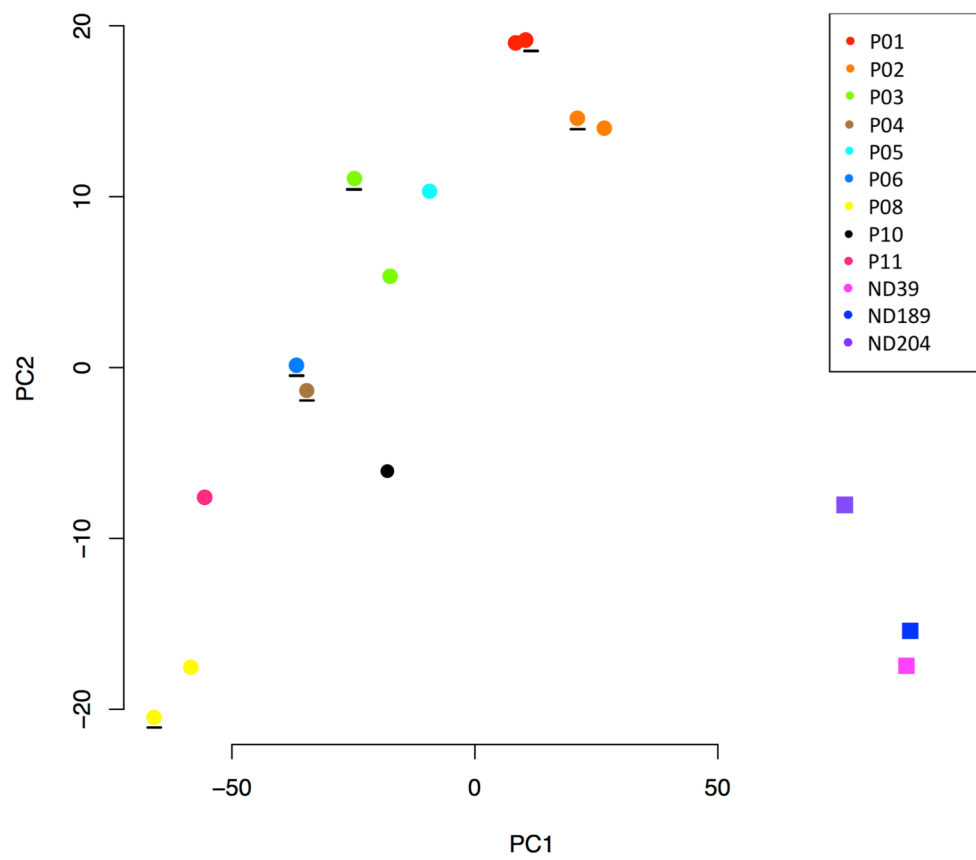
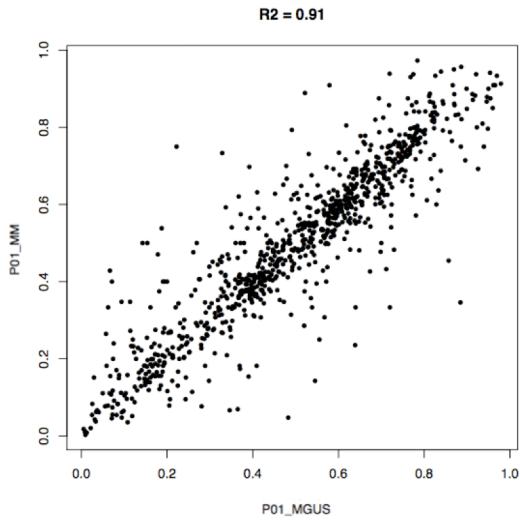
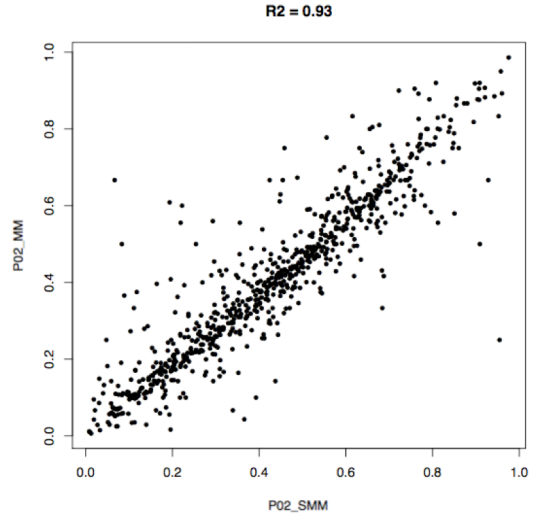


Figure 8. The relationship in methylation landscape associated with MGUS/SMM to MM progression. XY scatterplots illustrate minimal variation in methylation patterns between the asymptomatic stages of MGUS/SMM and MM in paired MGUS-MM (**a**: P01) and SMM-MM patients (**b**: P02, **c**: P03, **d**: P08). The median R^2 value was 0.90, demonstrating relative homogeneity in CpG methylation between disease stages. R^2 correlation values are given for each patient above individual plots.

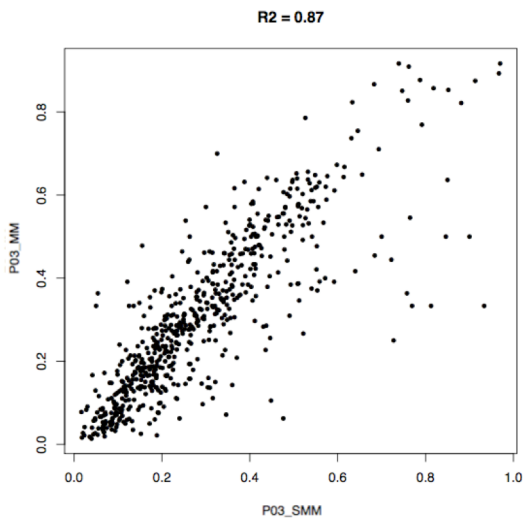
a



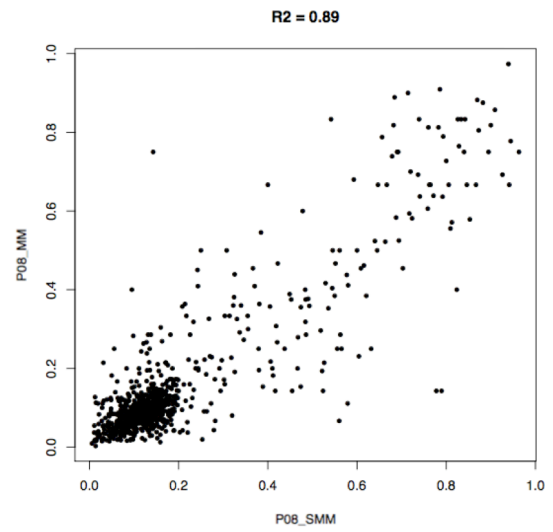
b



c



d

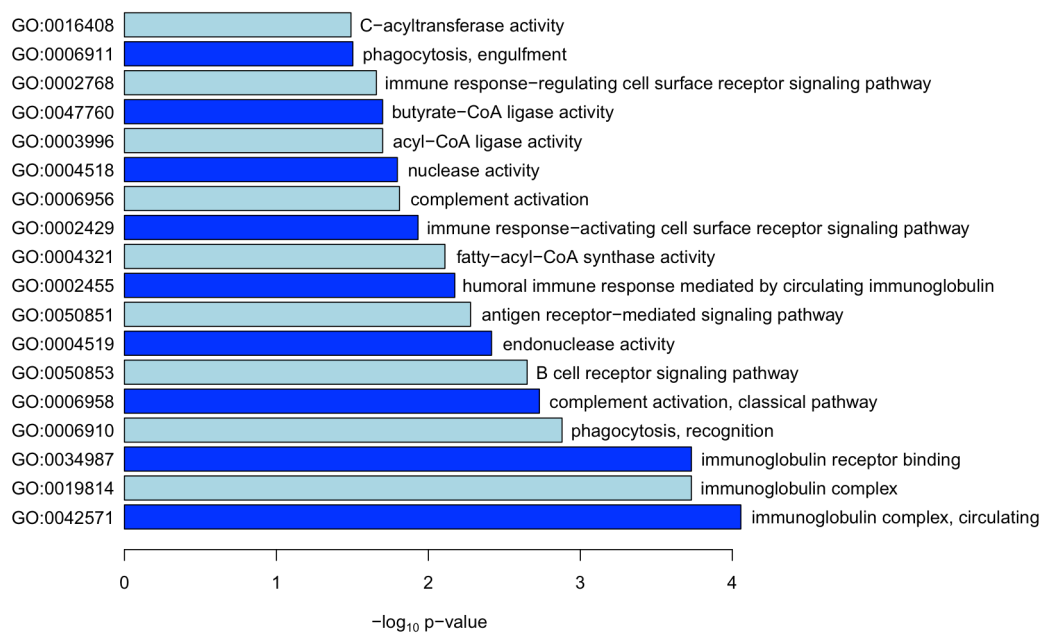


3.4.4 Gene ontology pathways associated with the transformation to MM.

Gene ontology (GO) analysis was performed to identify the effects of differentially methylated regions found between NPCs and MGUS/SMM and MM PCs. Epigenetic deregulation significantly affected 321 genes, with roles in cellular processes such as immunoglobulin complex assembly, immunoglobulin receptor binding, phagocytosis, complement activation and B cell receptor signalling [Figure 9].

Previous study by Salhia *et al.*, examining the pathways associated with differentially methylated genes in MGUS, SMM and MM, identified that extracellular matrix adhesion and remodelling were the most significantly affected pathways¹⁸. Associated dysregulation of cellular processes such as proteolysis and adhesion/extracellular matrix modifications suggested that hypomethylation, during myelomagenesis, may favour bone invasion by increasing interactions with the bone marrow extracellular matrix, initiating the required adhesive forces underlying bone invasion and the formation of lytic lesions¹⁸. Interestingly, a further study by Walker *et al.*, also revealed gene-specific hypermethylation associated with the transition of MGUS to MM, with 77 genes that have biological roles in development, cell cycle and transcription regulation pathways¹⁹.

Figure 9. Gene Ontology analysis of pathways associated with DMRs identified between NPCs and MGUS/SMM and MM. Gene Ontology reveals the top ranked cellular pathways/processes affected by differential methylation in MGUS/SMM and MM, compared to NPCs. There were 321 genes significantly affected by differential methylation. A larger $-\log(p)$ value indicates a more significant effect.



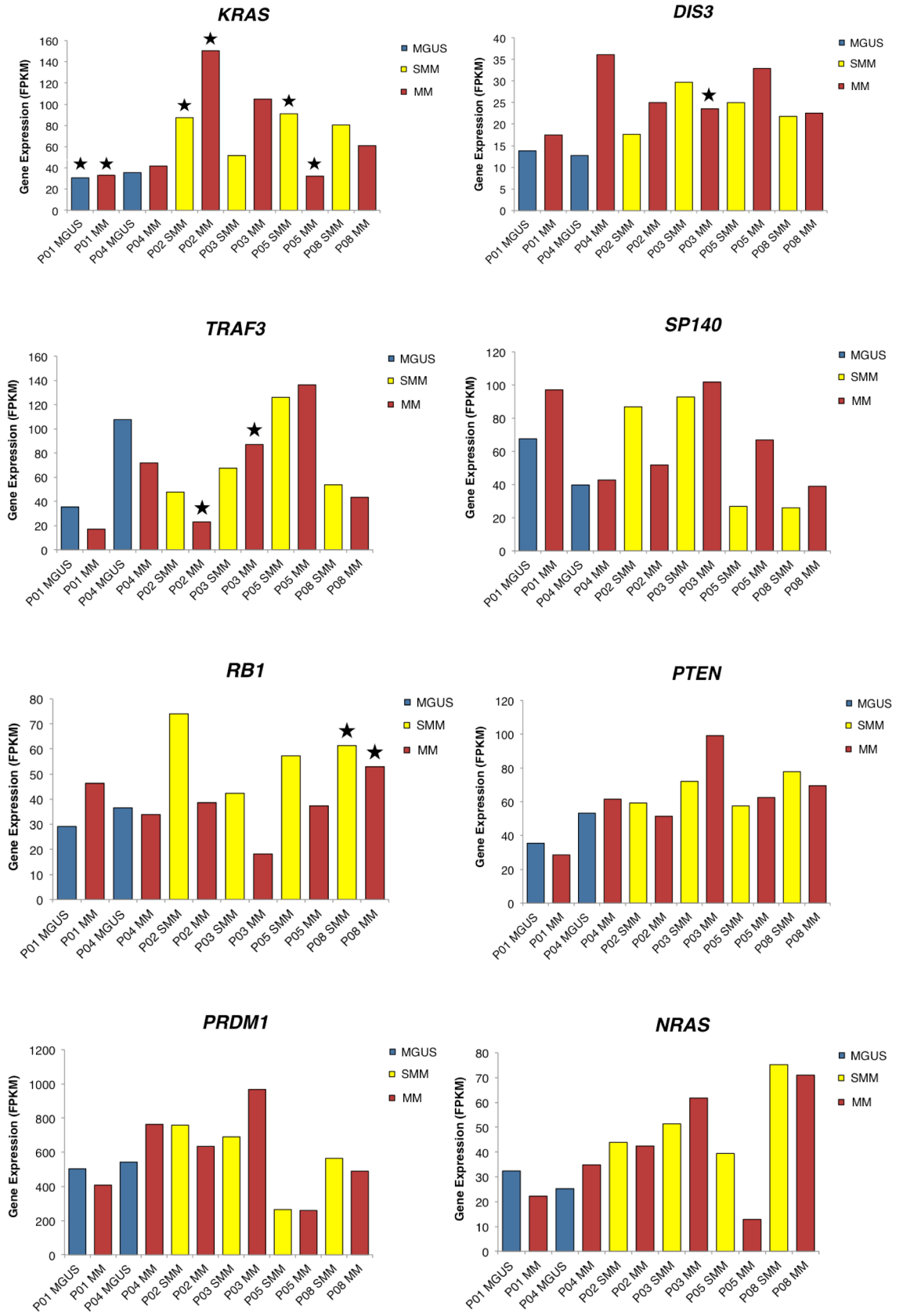
3.4.5 Allelic specific expression of genes associated with MGUS/SMM to MM progression.

NGS studies have identified recurrently mutated genes in MM^{1-3,5}, however ultimately, it is expression of these mutated genes in MM PCs that determines the phenotype of the tumour cells. Mutations in genes that are not expressed in PCs are more likely to be passenger mutations rather than driver mutations of disease. Our previous WES analysis of these paired samples identified 15 previously reported mutated genes. We assessed normalized RNAseq read counts (FPKM: Fragments Per Kilobase of transcript per Million mapped reads) as a measure of expression levels, finding most genes were generally not expressed, or expressed at low levels across all patients at both MGUS/SMM and MM. This included *KRAS*, *DIS3*, *TRAF3*, *SP140*, *RBI*, *PTEN*, *PRDM1*, *NRAS*, *MYC*, *MAGED1*, *IRF4*, *HLA-A* and *CDK4*. Patient-specific expression showed that, on average, most genes exhibited < 100 transcript counts [Figure 10a]. *PRDM1*, *MAGED1*, *IRF4* and *HLA-A* were observed to have high expression levels, consistently showing > 100 counts. Conversely, *FAT3* and *ROBO1*, were identified to not be expressed at all. There were no DMRs identified in the analysis between MGUS/SMM and MM, indicating transcriptional changes are unlikely to be caused directly at those genes.

We then assessed allelic exclusivity of expression of mutated genes in patients in our RNAseq data. WES analysis revealed patients that harboured mutations at the genetic level included *KRAS* (P01 MGUS-MM, P02 SMM-MM and P05 SMM-MM), *DIS3* (P03 MM), *TRAF3* (P02 MM, P03 MM), *RBI* (P08 SMM-MM), *MYC* (P01 MM), *IRF4* (P03 MM) and *CDK4* (P01 MM) [Figure 10a stars]. While these genes were identified to have DNA mutations, it was generally observed these mutants were expressed at similar levels to patients that didn't harbour the mutant. Assessing mutant allele specific expression, we mostly observed wild-type allele transcript expression, with only *KRAS* and *DIS3* having mutant allele expression in 2 patients (*KRAS*: P05 MM; and *DIS3*: P03 MM) [Figure 11]. Intriguingly, one of the *DIS3* transcript variants in P03 was unique to that of the two that were identified by WES. This suggests that this mutation was not detected in exome data due to the mutation residing in a minor subclone and had a low variant allele frequency (VAF). Furthermore, of genes with frequently acquired mutations at MM progression in our WES study, *KMT2D* was also expressed. *KMT2D* harboured genetic mutations in 4/6 patients (P01, P03, P05 and P08), and was also observed to be highly expressed at both MGUS/SMM and MM stages [Figure 10b]. However, mutant *KMT2D* was only expressed at low levels in one patient at MM (P03).

Figure 10. Expression of previously reported recurrently mutated genes in MM, identified in our MGUS/SMM to MM cohort. (a) RNAseq transcript counts show 13 recurrently mutated genes were expressed at both MGUS/SMM and MM. Average expression levels demonstrate most are expressed at low levels, or not at all (*FAT3* and *ROBO1*). *PRDMI*, *MAGED1*, *IRF4* and *HLA-A* were observed to have high expression levels (consistently > 100 counts). **(b)** Of the 10 most common genes harbouring mutations at MM in our WES data, only 4 genes exhibited expression including *KRAS*, *DIS3*, *SP140* and *KMT2D*. Stars represent mutation of a gene at specific disease stages in patients.

(a)



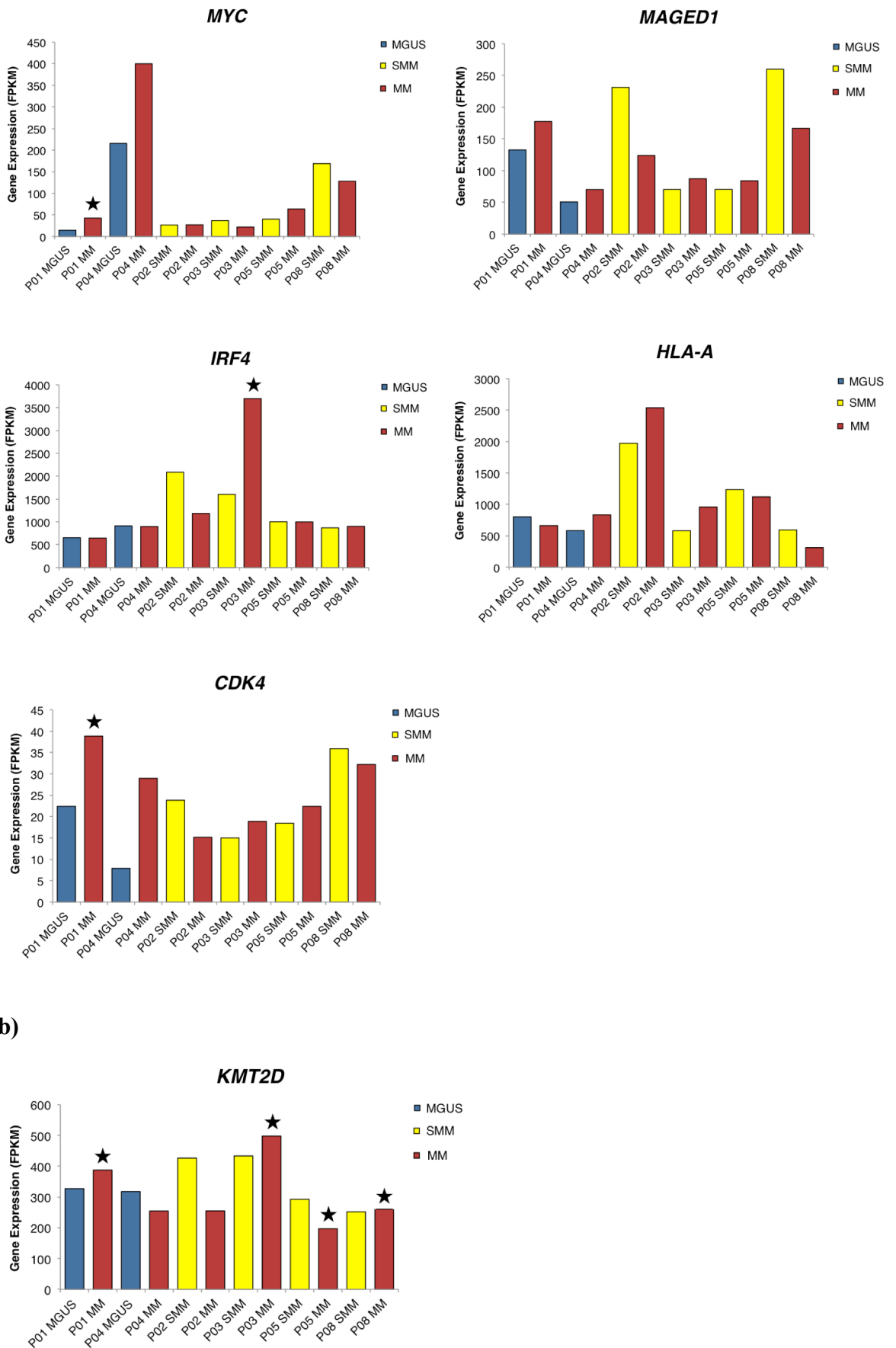


Figure 11. Correlation of WES data with RNAseq expression data. DNA mutations in recurrently mutated MM genes were identified in specific patients, however, these mutants were expressed at similar levels to patients that didn't harbor the mutant. Alignment of whole exome sequencing data with transcript expression data reveals most identified genetic mutations were lowly expressed, with only 3 genes: *KRAS*, *DIS3* and *KMT2D* showing patient specific mutant allele expression.

| Gene | WES | | | | RNAseq | | | | |
|--------------|-----|-----------|------------|------------|----------|------------|------------|--------------------|-----------|
| | Chr | Position | Ref allele | Alt allele | Position | Ref allele | Alt allele | Total Count (FPKM) | Alt reads |
| KRAS | | | | | | | | | |
| P01 MGUS | 12 | 25380275 | T | G | | | | 31 | |
| P01 MM | 12 | 25380275 | T | G | | | | | |
| | 12 | 25380276 | T | A | | | | | |
| | 12 | 25398284 | C | T | | | | | |
| | 12 | 25398285 | C | T | | | | 33 | |
| P02 SMM | 12 | 25378562 | C | G | | | | 87 | |
| P02 MM | 12 | 25378562 | C | G | | | | 150 | |
| P05 SMM | 12 | 25378561 | G | A | | | | 91 | |
| P05 MM | 12 | 25378561 | G | A | | | | | |
| | 12 | 25398284 | C | A | 25398284 | C | A | 32 | 2 |
| DIS3 | | | | | | | | | |
| P03 MM | 13 | 73337650 | C | G | | | | | |
| | 13 | 73345100 | A | T | 73345100 | A | T | | 6 |
| | 13 | | | | 73349359 | G | C | 24 | 5 |
| TRAF3 | | | | | | | | | |
| P02 MM | 14 | 103352559 | G | A | | | | 23 | |
| P03 MM | 14 | 103342759 | A | G | | | | 87 | |
| RB1 | | | | | | | | | |
| P08 SMM | 13 | 48953743 | G | A | | | | 61 | |
| P08 MM | 13 | 48953743 | G | A | | | | 53 | |
| MYC | | | | | | | | | |
| P01 MM | 8 | 128750965 | C | A | | | | 43 | |
| IRF4 | | | | | | | | | |
| P03 MM | 6 | 394932 | C | G | | | | 3698 | |
| CDK4 | | | | | | | | | |
| P01 MM | 12 | 58144762 | G | T | | | | 39 | |
| KMT2D | | | | | | | | | |
| P01 MM | 12 | 49427983 | C | T | | | | 388 | |
| P03 MM | 12 | 49416084 | G | A | | | | | |
| | | | | | 49445088 | T | C | 498 | 7 |
| P05 MM | 12 | 49415528 | A | G | | | | 197 | |
| P08 MM | 12 | 49430962 | G | A | | | | 260 | |

3.5 Discussion

The application of NGS technologies in genetic analyses of large cohorts of MM patient samples has established intraclonal genetic heterogeneity as a feature of disease. We have previously performed WES of longitudinal MGUS/SMM to MM patients, identifying clonal heterogeneity throughout all stages of disease, with progression being characterised by a model of clonal stability. Here, genomic analysis of a subset of these paired patient samples has revealed new insights into the transcriptomic and methylomic landscape associated with disease progression to MM. By using matched samples, we were able to negate any intrinsic inter-patient differences to determine the gene expression patterns between MGUS/SMM and MM, and methylation changes that accompany the transition of NPCs through to MGUS/SMM and MM.

Overall, we found minimal variation in the expression of genes associated with the progression of MGUS/SMM to MM. PCA of the transcriptome of each individual patient clustered very closely between the MGUS/SMM and MM stages. This highlights the limitation of previous studies using unmatched patient sample comparisons to identify common gene expression changes associated with MGUS/SMM to MM transition. We found 250 genes that were reaching significance in differential expression testing (raw $p < 0.05$), with a majority being downregulated upon MM transition. Due to the rare nature and small sample size of our matched cohort, gene expression changes did not reach statistical significance. Similarly, previous gene expression profiles of unmatched MGUS/MM samples have noted that MGUS can be clearly defined from NPCs, however, MGUS and MM samples appear to be identical at the gene expression level^{20,34,36,40}. Pathway analysis identified changes in gene expression associated with molecular processes such as angiogenesis^{41,42}, Wnt signalling^{43,44} and cadherin signalling^{45,46}, which are characteristically deregulated during MM disease progression. These pilot studies into the transcriptomic landscape associated with MGUS/SMM to MM progression have revealed an indication of the genes and pathways that are deregulated in the longitudinal progression to MM, and highlights the need for further, larger paired cohort studies to validate key changes associated with MM progression.

Patient-by-patient analysis revealed that a number of the top 10 differentially expressed genes consistently exhibited at least a fold change in expression in more than half of the patients. *THEMIS2* showed fold increase in one MGUS to MM, and two SMM to MM patients, and a recent study characterising its function has found that its overexpression results in enhanced downstream activation of MAPK kinases suggesting an

increase in the RAS signalling pathway⁴⁷. *HBB* exhibited fold decrease in two MGUS to MM, and two SMM to MM patients and its level of expression in breast and prostate cancer cells has previously been correlated with adverse clinical outcomes⁴⁸. *CD69* was identified to exclusively show at least a one-fold decrease in all SMM to MM patients. Previous *in vitro* and patient data has shown that the absence, or decrease in *CD69* expression, is linked with poor prognosis and resistance to the proteasome inhibitor bortezomib, and thus may guide clinical therapeutic choices for these specific SMM patients⁴⁹. Indeed, the median time to progression of SMM patients in this cohort was ~1 year, indicating their high-risk nature and poor prognosis.

We then sought to characterise the DNA methylation profile associated with the progression of MGUS/SMM to MM. Our data is partially concordant with the current understanding of the MM methylome during progression. To this end, previous unpaired studies comparing MGUS, SMM, MM or PCL samples with NPCs showed that extensive hypomethylation was an important early event in disease progression. The transformation from MGUS to MM was characterised by global hypomethylation¹⁸⁻²⁰, and this was associated with altered chromatin structure, changes in DNA methyltransferase activity, loss of imprinting and increased frequencies of CNVs¹⁹. The resulting aberrant transcription and chromosomal instability within PC clones was postulated to contribute to disease progression, and is one of the critical differences distinguishing MGUS from MM¹⁹.

Notably however, while these previous studies showed progressive hypomethylation upon MM transition, we revealed minimal variation in the methylation profiles of patients at the asymptomatic stages of MGUS or SMM, compared to MM. Moreover, the initiation of MGUS was characterized by extreme hypomethylation, a phenotype that was maintained with progression to SMM and MM. By contrast, one previous study used methylation data to clearly distinguish disease stages, where MGUS was defined by predominant hypomethylation and the later MM stage by acquired hypermethylation²⁰. It is conceivable that global hypomethylation identified at MGUS, maybe relates to the need for proliferating oncogenic PCs to have open chromatin and active gene transcription which could facilitate the MM transformation process^{18,19}. The initial investigation of the MM methylome showed increasing global hypomethylation with disease stages, and therefore proposed that the overall degree of methylation may have prognostic value¹⁸. However, our analysis shows that this may not be beneficial for patients that progress in a short time frame, where the levels of hypomethylation acquired

at MGUS is maintained during disease progression. The differences between the present study and previously reported findings may relate to sample dilution from non-tumour cells, as these previous studies have relied on CD138⁺ magnetic-activated cell sorting (MACS) sorting only. In contrast, we have used FACS to purify the CD138⁺ and 38⁺⁺ tumour PC populations from MGUS/SMM/MM BM samples, to minimise any sample heterogeneity.

It is possible that only subtypes of MM exhibit hypermethylation, and indeed heterogeneity in methylation has been shown for different cytogenetic groups of MM^{17,19}. Studies of other cancer types have shown the association of hypermethylation with treatment resistance, via the inactivation of various cell cycles and genes involved in chemo-sensitivity. As such, in these patients, DNA demethylating agents could be proposed to be effective in reducing the methylation levels in tumour cells to levels comparable to normal non-tumour cells. *In vitro* studies in human MM cell lines have shown that DNA methylase inhibitor decitabine (5-aza-2-deoxycytidine) can inhibit proliferation; with a hypomethylating effect in hypermethylated MM without any adverse effects²⁰. However, as our study, along with others, show that hypomethylation is maintained, or increased, with disease progression, the use of DNA methylase inhibitors would only be effective in hypermethylated subgroups of MM¹⁷⁻²⁰.

We assessed recurrently mutated genes in MM, finding that most mutated genes of MM have low or no expression in MM PCs, and in the instance where genes were expressed, this was generally associated with the occurrence of differential allelic expression. Differential and limited expression of mutant alleles has also been previously shown in a larger set of MM patient samples (n = 10 patients, 14 samples) analysed by RNAseq¹⁴. Notably, this would have significant implications when considering the use of targeted treatment strategies, which are solely based on driver mutations status revealed by genetic sequencing data only.

In conclusion, our studies reveal new insights into the gene expression and methylation patterns associated with the progression of MGUS/SMM to MM. Interestingly, PCs from patients at the asymptomatic stages of disease, appear to be as genetically complex as PCs recovered from the MM stage. Further, exhibiting minimal variation in gene expression, with wild type allele specific expression of mutants, and significant hypomethylation that is acquired at MGUS/SMM. These studies highlight the importance of transcriptomic and epigenetic interrogation of patient PCs to inform prognosis and treatment stratification.

3.6 Supplementary

3.6.1 Supplementary Figures

| Patient | Age | Diagnosis | Time To Progression (years) | Hyperdiploid | Translocations | Chromosomal Abberations | CRAB Symptoms |
|---------|-----|-----------|-----------------------------|--------------|----------------|---|---------------|
| P01 | 72 | MGUS-MM | 5.4 | nil | nil | nil | Bone lesions |
| P02 | 80 | SMM-MM | 1.2 | nil | t(4;14) | +1q,-8p21,-13q14 | Anaemia |
| P03 | 83 | SMM-MM | 4.1 | nil | nil | -13q14 | Pancytopenia |
| P04 | 79 | MGUS-MM | 1 | y | t(8;19) | +3,+5,+7,-8,+9,+11,-13,+15,+15,-18,+der(19) | Anaemia |
| P05 | 61 | SMM-MM | 0.9 | n/a | n/a | n/a | Anaemia |
| P08 | 95 | SMM-MM | 0.48 | nil | n/a | -13q | Anaemia |

Supplementary Table 1. Clinical cytogenetic data for MGUS/SMM to MM patients with RNAseq. Clinically recorded data at MM diagnosis for patients in the study. The median age of patients at MM diagnosis was 79.5 years old. Molecular cytogenetics of patients was performed using FISH analysis on interphase spreads of bone marrow smears. Nil represents parameter not being present. N/A represents that data was not available.

| Patient | Age | Diagnosis | Time To Progression (years) | Hyperdiploid | Translocations | Chromosomal Abberations | CRAB Symptoms |
|---------|-----|-----------|-----------------------------|--------------|----------------|---|---|
| P01 | 72 | MGUS-MM | 5.4 | nil | nil | nil | Bone lesions |
| P02 | 80 | SMM-MM | 1.2 | nil | t(4;14) | +1q,-8p21,-13q14 | Anaemia |
| P03 | 83 | SMM-MM | 4.1 | nil | nil | -13q14 | Pancytopenia |
| P04 | 79 | MGUS-MM | 1 | y | t(8;19) | +3,+5,+7,-8,+9,+11,-13,+15,+15,-18,+der(19) | Anaemia |
| P05 | 61 | SMM-MM | 0.9 | n/a | n/a | n/a | Anaemia |
| P06 | 65 | SMM-MM | 1.8 | nil | nil | nil | Anaemia, Peripheral neuropathy |
| P08 | 95 | SMM-MM | 0.48 | nil | n/a | -13q | Anaemia |
| P10 | 65 | MGUS-MM | 13 | nil | t(4;14) | +1q | Anaemia |
| P11 | 68 | MGUS | n/a | n/a | n/a | n/a | Anaemia, Bone lesions, Renal impairment |
| NPC1 | 24 | Normal | | | | | |
| NPC2 | 27 | Normal | | | | | |
| NPC3 | 33 | Normal | | | | | |

Supplementary Table 2. Clinical cytogenetic data for MGUS/SMM to MM patients with WGBS. Clinically recorded data at MM diagnosis for patients in the study. Molecular cytogenetics of patients was performed using FISH analysis on interphase spreads of bone marrow smears. Nil represents parameter not being present. N/A represents that data was not available.

(a)

| Genes | Symbol | logFC | logCPM | LR | p Value | FDR |
|-----------------|-----------------|--------------|---------------|-------------|----------------|------------|
| ENSG00000130812 | ANGPTL6 | 5.59773003 | 0.362576511 | 4.608683591 | 0.031810444 | 1 |
| ENSG00000213657 | RPL31P44 | 5.094653224 | 0.264262573 | 5.020705009 | 0.025045969 | 1 |
| ENSG00000213901 | SLC23A3 | 4.565223007 | 1.092082408 | 8.327880903 | 0.003904117 | 1 |
| ENSG00000260396 | NA | 4.521783271 | 0.628406087 | 4.687573004 | 0.030381531 | 1 |
| ENSG00000129173 | E2F8 | 4.446364925 | 1.697791267 | 7.381120584 | 0.006591209 | 1 |
| ENSG00000143228 | NUF2 | 4.081469523 | 1.265613637 | 5.107143182 | 0.023827524 | 1 |
| ENSG00000265218 | ENSG00000265218 | 3.873487561 | 0.535305572 | 7.097647153 | 0.007718521 | 1 |
| ENSG00000106070 | GRB10 | 3.760016403 | 1.625397912 | 4.071128404 | 0.043622125 | 1 |
| ENSG00000263033 | ENSG00000263033 | 3.739649174 | 0.596568017 | 4.078834776 | 0.043423594 | 1 |
| ENSG00000105374 | NKG7 | 3.716842278 | 5.504708007 | 6.441167048 | 0.011150537 | 1 |
| ENSG00000258891 | ENSG00000258891 | 3.073077177 | 1.121849217 | 5.350413476 | 0.020717357 | 1 |
| ENSG00000235560 | ENSG00000235560 | 3.070651553 | 1.06236116 | 4.209655004 | 0.040194511 | 1 |
| ENSG00000252061 | RNU6-415P | 3.056835205 | 1.836723997 | 4.39496998 | 0.036045094 | 1 |
| ENSG00000155254 | MARVELD1 | 3.009516417 | 3.53762875 | 3.985138022 | 0.04590334 | 1 |
| ENSG00000267554 | ENSG00000267554 | 2.981053864 | 1.546698827 | 5.247137875 | 0.021982901 | 1 |
| ENSG00000237765 | FAM200B | 2.740824089 | 1.478501314 | 6.894846179 | 0.00864446 | 1 |
| ENSG00000206650 | SNORA70G | 2.682573707 | 1.007744817 | 4.615122446 | 0.031691233 | 1 |
| ENSG00000170190 | SLC16A5 | 2.659007198 | 2.659935649 | 6.23907012 | 0.012496207 | 1 |
| ENSG00000100453 | GZMB | 2.636159743 | 4.329505166 | 5.549395363 | 0.018487072 | 1 |
| ENSG00000222009 | BTBD19 | 2.515115933 | 3.8061088 | 12.30553519 | 0.000451617 | 1 |
| ENSG00000260417 | ENSG00000260417 | 2.480131459 | 1.246217355 | 5.240923471 | 0.02206156 | 1 |
| ENSG00000154920 | EME1 | 2.440482556 | 1.915432918 | 4.140957832 | 0.041857479 | 1 |
| ENSG00000099985 | OSM | 2.417570983 | 1.494434748 | 4.537262615 | 0.033164587 | 1 |
| ENSG00000260274 | ENSG00000260274 | 2.404916132 | 1.056828284 | 4.604203654 | 0.031893663 | 1 |
| ENSG00000221886 | ZBED8 | 2.373231786 | 0.776523997 | 4.049509159 | 0.04418419 | 1 |
| ENSG00000163251 | FZD5 | 2.248344449 | 1.852410979 | 4.883267943 | 0.02711823 | 1 |
| ENSG00000259781 | HMGB1P6 | 2.192382482 | 1.871208518 | 4.073790972 | 0.043553424 | 1 |
| ENSG00000260006 | NA | 2.182465038 | 2.846382402 | 7.682994347 | 0.00557436 | 1 |
| ENSG00000241634 | ENSG00000241634 | 2.155798765 | 4.074640484 | 5.202191131 | 0.022558435 | 1 |
| ENSG00000186642 | PDE2A | 2.155159979 | 1.468172633 | 4.501825758 | 0.033858684 | 1 |
| ENSG00000170160 | CCDC144A | 2.150131996 | 5.618496323 | 4.0111624 | 0.045199978 | 1 |
| ENSG00000140379 | BCL2A1 | 2.004935153 | 3.53176948 | 5.880745842 | 0.015307336 | 1 |
| ENSG00000162073 | PAQR4 | 1.967387502 | 2.606875637 | 5.657893393 | 0.017376982 | 1 |
| ENSG00000135378 | PRRG4 | 1.909978623 | 2.521457866 | 5.227194951 | 0.022236365 | 1 |
| ENSG00000272016 | NA | 1.862981329 | 3.074052598 | 5.34304043 | 0.020805156 | 1 |
| ENSG00000156831 | NSMCE2 | 1.831393241 | 3.154126284 | 4.576160522 | 0.032419779 | 1 |
| ENSG00000204519 | ZNF551 | 1.760167606 | 3.45577717 | 5.967170679 | 0.014574647 | 1 |
| ENSG00000163521 | GLB1L | 1.755546949 | 2.910539378 | 4.0918879 | 0.043089487 | 1 |
| ENSG00000007384 | RHBDF1 | 1.746786161 | 2.44065237 | 4.325874782 | 0.037537115 | 1 |
| ENSG00000247596 | TWF2 | 1.673407196 | 3.560156351 | 8.723244961 | 0.003141784 | 1 |
| ENSG00000166592 | RRAD | 1.527594213 | 3.200898839 | 4.216526071 | 0.040032041 | 1 |
| ENSG00000185158 | LRRC37B | 1.526029938 | 3.309693689 | 3.865427699 | 0.049290629 | 1 |
| ENSG00000106991 | ENG | 1.525475514 | 3.827272591 | 4.378246705 | 0.036400422 | 1 |
| ENSG00000261474 | ENSG00000261474 | 1.455249643 | 2.056027838 | 4.918104082 | 0.026576667 | 1 |
| ENSG00000141441 | GAREM1 | 1.393411092 | 3.697500714 | 4.92204574 | 0.026516101 | 1 |

| | | | | | | |
|-----------------|-----------------|-------------|-------------|-------------|-------------|---|
| ENSG00000260065 | NA | 1.386707083 | 3.276824785 | 5.688229197 | 0.017079081 | 1 |
| ENSG00000168528 | SERINC2 | 1.385178092 | 3.111496391 | 6.492203687 | 0.010834858 | 1 |
| ENSG00000059915 | PSD | 1.348355238 | 3.204416065 | 6.328145739 | 0.01188365 | 1 |
| ENSG00000159674 | SPON2 | 1.285530958 | 5.196369142 | 5.338024476 | 0.020865106 | 1 |
| ENSG00000151725 | CENPU | 1.282862563 | 3.111251171 | 5.498284358 | 0.01903514 | 1 |
| ENSG00000167984 | NLRC3 | 1.264543285 | 4.027507446 | 4.254748556 | 0.039140742 | 1 |
| ENSG00000166886 | NAB2 | 1.258994642 | 4.239620787 | 8.042922978 | 0.004568177 | 1 |
| ENSG00000234925 | ATP5HP4 | 1.246794796 | 3.153536453 | 4.93436922 | 0.02632767 | 1 |
| ENSG00000257242 | LINC01619 | 1.231701382 | 4.052566143 | 5.499565064 | 0.019021204 | 1 |
| ENSG00000116514 | RNF19B | 1.19590434 | 4.648284417 | 4.561813684 | 0.032692435 | 1 |
| ENSG00000090339 | ICAM1 | 1.166667393 | 6.699785553 | 8.281076687 | 0.004006037 | 1 |
| ENSG00000119403 | PHF19 | 1.166257851 | 4.526416958 | 5.184712214 | 0.022786446 | 1 |
| ENSG00000184207 | PGP | 1.156251989 | 3.901283167 | 6.00182375 | 0.014291098 | 1 |
| ENSG00000128039 | SRD5A3 | 1.117616883 | 3.706187836 | 4.946065234 | 0.02615012 | 1 |
| ENSG00000146757 | ZNF92 | 1.078053681 | 3.869228731 | 3.978004628 | 0.046098141 | 1 |
| ENSG00000168826 | ZBTB49 | 1.075403941 | 3.391477533 | 4.684985801 | 0.030427315 | 1 |
| ENSG00000162702 | ZNF281 | 1.073593106 | 4.122926474 | 4.060181582 | 0.043905778 | 1 |
| ENSG00000130775 | THEMIS2 | 1.031872191 | 8.13617753 | 13.82926278 | 0.000200194 | 1 |
| ENSG00000186594 | MIR22HG | 1.01283964 | 5.569450039 | 5.422431079 | 0.019879661 | 1 |
| ENSG00000197872 | FAM49A | 0.989629942 | 5.548124839 | 6.445291446 | 0.011124679 | 1 |
| ENSG00000162971 | TYW5 | 0.987417058 | 3.909111358 | 4.529033559 | 0.033324432 | 1 |
| ENSG00000101445 | PPP1R16B | 0.966767518 | 6.03598865 | 5.766007181 | 0.016339117 | 1 |
| ENSG00000243317 | C7orf73 | 0.95172681 | 3.841854925 | 3.879552073 | 0.048877596 | 1 |
| ENSG00000167553 | TUBA1C | 0.936073097 | 5.003187147 | 5.050792863 | 0.024614672 | 1 |
| ENSG00000181467 | RAP2B | 0.921544818 | 5.541995314 | 5.001715988 | 0.025322201 | 1 |
| ENSG00000182541 | LIMK2 | 0.920040786 | 4.832437155 | 7.964049997 | 0.004771553 | 1 |
| ENSG00000120063 | GNA13 | 0.916721676 | 6.574926601 | 6.542520044 | 0.010532605 | 1 |
| ENSG00000173846 | PLK3 | 0.907355807 | 6.309176292 | 4.112954779 | 0.042555951 | 1 |
| ENSG00000158019 | BABAM2 | 0.903782247 | 4.189787631 | 3.982641709 | 0.045971411 | 1 |
| ENSG00000263244 | ENSG00000263244 | 0.900103969 | 6.521734646 | 6.898919526 | 0.008624786 | 1 |
| ENSG00000183696 | UPP1 | 0.889747433 | 5.174400937 | 4.082448513 | 0.043330825 | 1 |
| ENSG00000146112 | PPP1R18 | 0.888695895 | 6.8322549 | 6.785610064 | 0.009189562 | 1 |
| ENSG00000168389 | MFSD2A | 0.879138461 | 4.60020596 | 4.429007589 | 0.035333046 | 1 |
| ENSG00000189159 | JPT1 | 0.868240646 | 3.708887728 | 3.947055711 | 0.046953452 | 1 |
| ENSG00000144136 | SLC20A1 | 0.866610591 | 5.816218597 | 8.720482028 | 0.003146548 | 1 |
| ENSG00000020633 | RUNX3 | 0.861212016 | 6.69306518 | 4.587644364 | 0.032203244 | 1 |
| ENSG00000176890 | TYMS | 0.859750802 | 4.180663644 | 4.183890505 | 0.040809917 | 1 |
| ENSG00000164615 | CAMLG | 0.850187442 | 4.15655429 | 4.45490277 | 0.034801197 | 1 |
| ENSG00000160856 | FCRL3 | 0.846897549 | 4.962550147 | 4.286441355 | 0.038417496 | 1 |
| ENSG00000101365 | IDH3B | 0.844103263 | 5.61575824 | 8.28705444 | 0.00399287 | 1 |
| ENSG00000117632 | STMN1 | 0.84105229 | 5.068401108 | 4.853227979 | 0.027594445 | 1 |
| ENSG00000117614 | SYF2 | 0.814464041 | 5.01080814 | 5.441791499 | 0.019660494 | 1 |
| ENSG00000054967 | RELT | 0.802782033 | 5.92571288 | 5.781929724 | 0.016191755 | 1 |
| ENSG00000127554 | GFER | 0.76586362 | 4.115137122 | 4.23856659 | 0.039515516 | 1 |
| ENSG00000059728 | MXD1 | 0.765267415 | 5.689713318 | 5.31367617 | 0.021158664 | 1 |
| ENSG00000034152 | MAP2K3 | 0.763509686 | 6.072777207 | 4.411075166 | 0.03570633 | 1 |
| ENSG00000137193 | PIM1 | 0.761691184 | 7.842099377 | 5.882592484 | 0.015291291 | 1 |
| ENSG00000196950 | SLC39A10 | 0.755886339 | 4.83105715 | 5.424983591 | 0.019850622 | 1 |

| | | | | | | |
|-----------------|--------|-------------|-------------|-------------|-------------|---|
| ENSG00000070495 | JMJD6 | 0.739648438 | 5.476770449 | 5.88673918 | 0.015255323 | 1 |
| ENSG00000167470 | MIDN | 0.734117142 | 7.010780925 | 4.042867302 | 0.044358394 | 1 |
| ENSG00000141543 | EIF4A3 | 0.726748769 | 5.724494984 | 7.651092656 | 0.005673795 | 1 |
| ENSG00000101558 | VAPA | 0.719601535 | 6.034219757 | 4.875957239 | 0.027233332 | 1 |
| ENSG00000080371 | RAB21 | 0.717771907 | 5.256629701 | 4.79632292 | 0.028520546 | 1 |
| ENSG00000154358 | OBSCN | 0.703502616 | 6.306876813 | 4.189430232 | 0.040676766 | 1 |
| ENSG00000115685 | PPP1R7 | 0.694582031 | 4.98493528 | 4.145528404 | 0.041744625 | 1 |
| ENSG00000117868 | ESYT2 | 0.691788109 | 6.316277694 | 4.830224389 | 0.027964991 | 1 |
| ENSG00000175376 | EIF1AD | 0.690428953 | 5.80840658 | 4.67764417 | 0.030557627 | 1 |
| ENSG00000164081 | TEX264 | 0.680190517 | 5.386480396 | 4.023427237 | 0.04487244 | 1 |
| ENSG00000165609 | NUDT5 | 0.668390669 | 4.713849585 | 4.253822407 | 0.039162091 | 1 |
| ENSG00000171222 | SCAND1 | 0.634605978 | 5.725477751 | 3.934479473 | 0.047305778 | 1 |
| ENSG00000162413 | KLHL21 | 0.616725409 | 6.014719633 | 4.190839969 | 0.040642955 | 1 |
| ENSG00000135801 | TAF5L | 0.597008775 | 4.824529436 | 4.392673775 | 0.036093667 | 1 |
| ENSG00000100664 | EIF5 | 0.553810713 | 8.644360299 | 3.870303916 | 0.049147621 | 1 |
| ENSG00000023734 | STRAP | 0.548869638 | 5.434474156 | 4.029838467 | 0.044702221 | 1 |

(b)

| Genes | Symbol | logFC | logCPM | LR | p Value | FDR |
|-----------------|-----------------|--------------|---------------|-------------|----------------|-------------|
| ENSG00000272140 | ENSG00000272140 | -5.952498835 | 1.885835518 | 5.287043455 | 0.021484665 | 1 |
| ENSG00000225796 | MTND4P23 | -5.908020984 | 0.736226569 | 6.320774208 | 0.011933152 | 1 |
| ENSG00000270878 | ENSG00000270878 | -5.581537232 | 0.2483212 | 7.761508681 | 0.005337136 | 1 |
| ENSG00000197385 | ZNF860 | -5.46259309 | 2.017194426 | 5.232892022 | 0.022163651 | 1 |
| ENSG00000228171 | ENSG00000228171 | -5.210151138 | 1.091689302 | 3.928802477 | 0.047465733 | 1 |
| ENSG00000223648 | IGHV3-64 | -5.092070201 | 3.39540472 | 5.641912445 | 0.017536069 | 1 |
| ENSG00000189366 | ALG1L | -5.022309004 | 0.473919404 | 8.507505761 | 0.003536845 | 1 |
| ENSG00000198785 | GRIN3A | -4.984021898 | 0.633707833 | 3.894154205 | 0.048454431 | 1 |
| ENSG00000128606 | LRRC17 | -4.919132696 | 0.909072344 | 4.364904446 | 0.036686541 | 1 |
| ENSG00000211950 | IGHV1-24 | -4.917608925 | 2.272410422 | 17.50352622 | 2.87E-05 | 0.422420029 |
| ENSG00000269305 | NA | -4.914212362 | 0.62749366 | 7.351700971 | 0.006699934 | 1 |
| ENSG00000260423 | LINC02367 | -4.776349445 | 0.732204719 | 5.297240222 | 0.02135924 | 1 |
| ENSG00000207770 | NA | -4.751101012 | 0.814754393 | 5.426290494 | 0.019835771 | 1 |
| ENSG00000250327 | RPSAP70 | -4.729113515 | 0.713329432 | 4.116640882 | 0.042463314 | 1 |
| ENSG00000163554 | SPTA1 | -4.615834282 | 1.751636027 | 3.991562758 | 0.045728633 | 1 |
| ENSG00000234840 | LINC01239 | -4.581231634 | 2.56883438 | 4.638841538 | 0.031256095 | 1 |
| ENSG00000211938 | IGHV3-7 | -4.571456261 | 2.984870593 | 5.843946071 | 0.015630733 | 1 |
| ENSG00000234536 | NA | -4.438132327 | 0.63957825 | 4.302819958 | 0.038049233 | 1 |
| ENSG00000171084 | FAM86JP | -4.433154584 | 1.011104527 | 4.404055325 | 0.035853577 | 1 |
| ENSG00000211659 | IGLV3-25 | -4.358518469 | 2.36349412 | 5.785031476 | 0.016163208 | 1 |
| ENSG00000238061 | ENSG00000238061 | -4.335698786 | 0.691866138 | 5.675683214 | 0.01720164 | 1 |
| ENSG00000164695 | CHMP4C | -4.188786663 | 1.444276807 | 6.964395709 | 0.008314751 | 1 |
| ENSG00000269578 | ENSG00000269578 | -4.094618848 | 0.444702767 | 4.712863376 | 0.029937747 | 1 |
| ENSG00000234709 | UPF3AP3 | -3.73491562 | 0.334353817 | 4.018241914 | 0.04501061 | 1 |
| ENSG00000225151 | GOLGA2P7 | -3.724560156 | 0.832866737 | 4.660441203 | 0.030865259 | 1 |
| ENSG00000118402 | ELOVL4 | -3.687809517 | 1.523571922 | 5.477522507 | 0.019262541 | 1 |
| ENSG00000151304 | SRFBP1 | -3.649866163 | 2.075693602 | 3.916218546 | 0.047822335 | 1 |
| ENSG00000198711 | SSBP3-AS1 | -3.586417747 | 0.793069981 | 6.21147422 | 0.012692489 | 1 |
| ENSG00000223599 | ENSG00000223599 | -3.516669336 | 0.987193931 | 4.541262318 | 0.033087184 | 1 |
| ENSG00000080224 | EPHA6 | -3.449710341 | 1.725517255 | 5.625815801 | 0.017697827 | 1 |
| ENSG00000184984 | CHRM5 | -3.440548036 | 0.948070667 | 4.524186708 | 0.033418956 | 1 |
| ENSG00000219470 | ENSG00000219470 | -3.433551561 | 0.033822024 | 6.071087439 | 0.013741234 | 1 |
| ENSG00000198513 | ATL1 | -3.417146908 | 1.46065116 | 4.805097119 | 0.028375667 | 1 |
| ENSG00000113966 | ARL6 | -3.369941472 | 1.054492685 | 5.224853904 | 0.022266317 | 1 |
| ENSG00000232233 | LINC02043 | -3.2854874 | 1.582357566 | 4.926557736 | 0.026446949 | 1 |
| ENSG00000143858 | SYT2 | -3.097979752 | 0.884021841 | 4.237430372 | 0.039541972 | 1 |
| ENSG00000226002 | GTF2IP14 | -2.959822123 | 0.960959013 | 3.963487913 | 0.046497265 | 1 |
| ENSG00000211956 | IGHV4-34 | -2.934554149 | 2.352190274 | 5.336977647 | 0.02087764 | 1 |
| ENSG00000134297 | PLEKHA8P1 | -2.88388571 | 0.519689112 | 5.574769694 | 0.018221061 | 1 |
| ENSG00000153558 | FBXL2 | -2.86762063 | 1.256166107 | 5.104906608 | 0.023858263 | 1 |
| ENSG00000125869 | LAMP5 | -2.866917217 | 6.334825109 | 5.148179659 | 0.023270756 | 1 |
| ENSG00000244734 | HBB | -2.82839705 | 3.951500023 | 11.63621708 | 0.0006468 | 1 |
| ENSG00000211949 | IGHV3-23 | -2.80124484 | 4.287558188 | 6.601742862 | 0.0101879 | 1 |
| ENSG00000160791 | CCR5 | -2.750873225 | 2.119348898 | 7.094637499 | 0.007731493 | 1 |
| ENSG00000139160 | ETFBKMT | -2.717965905 | 2.168615593 | 4.131744811 | 0.042085937 | 1 |

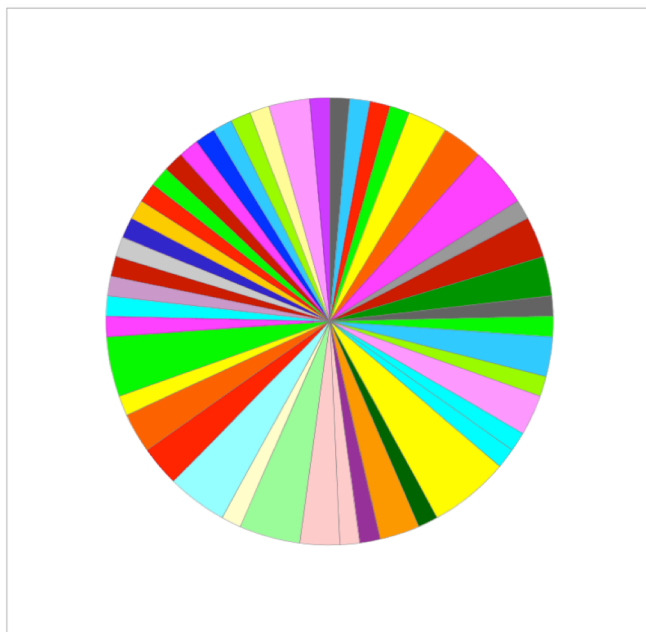
| | | | | | | |
|-----------------|-----------------|--------------|-------------|-------------|-------------|---|
| ENSG00000132932 | ATP8A2 | -2.697925 | 4.537678402 | 9.836896781 | 0.001710458 | 1 |
| ENSG00000112782 | CLIC5 | -2.685038654 | 2.891668225 | 4.655566046 | 0.030953025 | 1 |
| ENSG00000246273 | SBF2-AS1 | -2.618815409 | 1.189189091 | 3.903858138 | 0.048175355 | 1 |
| ENSG00000117322 | CR2 | -2.583033599 | 1.16451107 | 4.834473741 | 0.027896154 | 1 |
| ENSG00000226549 | SCDP1 | -2.580823945 | 3.337616185 | 4.273558498 | 0.038709784 | 1 |
| ENSG00000169403 | PTAFR | -2.579930264 | 5.550437296 | 3.868643752 | 0.049196261 | 1 |
| ENSG00000211898 | IGHD | -2.556609483 | 4.070314768 | 4.19429458 | 0.040560225 | 1 |
| ENSG00000110848 | CD69 | -2.529781658 | 4.358278124 | 8.829270831 | 0.002964368 | 1 |
| ENSG00000064989 | CALCRL | -2.525758131 | 3.393464976 | 6.963530368 | 0.008318773 | 1 |
| ENSG00000160539 | PLPP7 | -2.495545958 | 2.259401839 | 4.94660077 | 0.02614202 | 1 |
| ENSG00000176125 | UFSP1 | -2.474866362 | 1.347517488 | 5.89234079 | 0.015206874 | 1 |
| ENSG00000241527 | CA15P1 | -2.467551177 | 1.863774615 | 5.41572513 | 0.019956163 | 1 |
| ENSG00000167693 | NXN | -2.43453088 | 2.050326983 | 3.898570389 | 0.048327215 | 1 |
| ENSG00000233695 | GAS6-AS1 | -2.39540206 | 1.987592662 | 4.835110074 | 0.027885861 | 1 |
| ENSG00000211651 | IGLV1-44 | -2.377708208 | 3.584987505 | 6.526161559 | 0.010629911 | 1 |
| ENSG00000188191 | PRKAR1B | -2.343549651 | 3.283215504 | 4.073781691 | 0.043553663 | 1 |
| ENSG00000163794 | UCN | -2.329110748 | 1.570406646 | 4.165018154 | 0.041266971 | 1 |
| ENSG00000243466 | IGKV1-5 | -2.31385205 | 4.439275156 | 5.002730249 | 0.025307367 | 1 |
| ENSG00000233806 | LINC01237 | -2.266151637 | 2.800753973 | 7.932544317 | 0.004855348 | 1 |
| ENSG00000211959 | IGHV4-39 | -2.260361407 | 6.90392344 | 4.009818725 | 0.045236014 | 1 |
| ENSG00000075275 | CELSR1 | -2.25479695 | 4.71732946 | 9.620450668 | 0.001924225 | 1 |
| ENSG00000153140 | CETN3 | -2.245813645 | 2.306627964 | 3.865332661 | 0.049293421 | 1 |
| ENSG00000250654 | ENSG00000250654 | -2.236569125 | 1.392633527 | 4.098230781 | 0.042928112 | 1 |
| ENSG00000198477 | NA | -2.232627109 | 2.130492298 | 4.054868732 | 0.044044143 | 1 |
| ENSG00000144908 | ALDH1L1 | -2.1930315 | 3.714049151 | 5.570093796 | 0.018269782 | 1 |
| ENSG00000012124 | CD22 | -2.169659144 | 3.864959183 | 3.963126129 | 0.046507258 | 1 |
| ENSG00000268041 | ENSG00000268041 | -2.125264856 | 2.659605612 | 4.797994676 | 0.028492883 | 1 |
| ENSG00000163520 | FBLN2 | -2.097246847 | 4.926336827 | 7.543944366 | 0.006021209 | 1 |
| ENSG00000168916 | ZNF608 | -2.091355427 | 3.522837759 | 4.660715459 | 0.03086033 | 1 |
| ENSG00000187870 | RNFT1P3 | -2.040096722 | 4.557979814 | 4.659198899 | 0.0308876 | 1 |
| ENSG00000241127 | YAE1D1 | -1.928511982 | 1.832404908 | 4.510552852 | 0.033686351 | 1 |
| ENSG00000110811 | P3H3 | -1.89152959 | 4.39257172 | 5.69200071 | 0.017042414 | 1 |
| ENSG00000204248 | COL11A2 | -1.879514501 | 3.007209998 | 4.477083642 | 0.034352292 | 1 |
| ENSG00000199568 | RNU5A-1 | -1.818473095 | 4.106512986 | 4.811932394 | 0.028263335 | 1 |
| ENSG00000163126 | ANKRD23 | -1.786419112 | 2.824102778 | 4.650998986 | 0.031035481 | 1 |
| ENSG00000259712 | ENSG00000259712 | -1.786291136 | 1.458067007 | 3.994183505 | 0.045657569 | 1 |
| ENSG00000154553 | PDLIM3 | -1.769447687 | 3.872653354 | 5.793752614 | 0.016083222 | 1 |
| ENSG00000214456 | PLIN5 | -1.760757379 | 2.274869889 | 5.125223195 | 0.023580538 | 1 |
| ENSG00000129282 | NA | -1.746849353 | 2.377494774 | 3.943638971 | 0.047048898 | 1 |
| ENSG00000248996 | ENSG00000248996 | -1.741823284 | 1.637791611 | 4.009452404 | 0.045245844 | 1 |
| ENSG00000148926 | ADM | -1.71153944 | 5.917847554 | 4.422795963 | 0.035461884 | 1 |
| ENSG00000137338 | PGBD1 | -1.697145114 | 2.579139013 | 4.132971321 | 0.042055447 | 1 |
| ENSG00000112379 | ARFGEF3 | -1.676225349 | 4.610654505 | 6.034009692 | 0.014032821 | 1 |
| ENSG00000122778 | KIAA1549 | -1.667278508 | 2.968993317 | 4.058820985 | 0.043941169 | 1 |
| ENSG00000211662 | IGLV3-21 | -1.645088037 | 12.25352334 | 6.628454926 | 0.010036228 | 1 |
| ENSG00000211666 | IGLV2-14 | -1.639183161 | 5.650023863 | 4.122668158 | 0.042312295 | 1 |
| ENSG00000240041 | IGHJ4 | -1.576767407 | 3.325330077 | 4.498448639 | 0.033925619 | 1 |
| ENSG00000203814 | HIST2H2BF | -1.546617402 | 3.514633843 | 6.560335342 | 0.010427671 | 1 |

| | | | | | | |
|-----------------|--------------|--------------|-------------|-------------|-------------|---|
| ENSG00000149403 | GRIK4 | -1.507817066 | 5.201427158 | 6.612655837 | 0.010125654 | 1 |
| ENSG00000117407 | ARTN | -1.467584203 | 3.847478148 | 5.411262018 | 0.020007247 | 1 |
| ENSG00000262943 | ALOX12P2 | -1.353814106 | 3.184684189 | 4.006832188 | 0.045316219 | 1 |
| ENSG00000211892 | IGHG4 | -1.350690295 | 8.517813894 | 5.960853562 | 0.014626958 | 1 |
| ENSG00000161277 | THAP8 | -1.340221363 | 2.349664214 | 4.321048402 | 0.037643723 | 1 |
| ENSG00000237638 | LINC02245 | -1.328971929 | 3.629445924 | 3.85837259 | 0.049498318 | 1 |
| ENSG00000179909 | ZNF154 | -1.318921941 | 4.240688648 | 4.534494647 | 0.033218263 | 1 |
| ENSG00000155189 | AGPAT5 | -1.303248593 | 3.200331411 | 4.745309845 | 0.029378263 | 1 |
| ENSG00000215424 | MCM3AP-AS1 | -1.260328798 | 3.25277696 | 4.680153558 | 0.030513021 | 1 |
| ENSG00000203772 | SPRN | -1.258429576 | 3.103016704 | 4.191139008 | 0.040635787 | 1 |
| ENSG00000237940 | LINC01238 | -1.251744801 | 3.328418042 | 5.394611304 | 0.020199024 | 1 |
| ENSG00000211677 | IGLC2 | -1.234687676 | 7.751354291 | 5.513697877 | 0.018868114 | 1 |
| ENSG00000187824 | TMEM220 | -1.215489622 | 2.624393088 | 3.930592401 | 0.047415239 | 1 |
| ENSG00000048162 | NOP16 | -1.193773706 | 3.834441305 | 5.83191734 | 0.015737962 | 1 |
| ENSG00000133740 | E2F5 | -1.139307269 | 4.515135523 | 7.838967665 | 0.00511318 | 1 |
| ENSG00000139737 | SLAIN1 | -1.133077728 | 3.420055415 | 3.944186446 | 0.047033591 | 1 |
| ENSG00000092096 | SLC22A17 | -1.115029514 | 6.028606206 | 8.132447627 | 0.004348001 | 1 |
| ENSG00000163534 | FCRL1 | -1.107910974 | 5.283049201 | 4.34092371 | 0.037206732 | 1 |
| ENSG00000213398 | LCAT | -1.099954761 | 4.585586113 | 4.12000928 | 0.042378846 | 1 |
| ENSG00000171444 | MCC | -1.069469943 | 7.31792472 | 4.210730196 | 0.040169042 | 1 |
| ENSG00000100154 | TTC28 | -1.049628994 | 4.863663331 | 4.20191939 | 0.040378252 | 1 |
| ENSG00000171044 | XKR6 | -0.995645929 | 3.073036969 | 4.03918769 | 0.044455215 | 1 |
| ENSG00000196562 | SULF2 | -0.947931233 | 8.515725284 | 7.248259759 | 0.007096976 | 1 |
| ENSG00000162600 | OMA1 | -0.930762683 | 4.266898095 | 4.041425761 | 0.044396299 | 1 |
| ENSG00000238164 | TNFRSF14-AS1 | -0.927582561 | 4.494708225 | 3.847726469 | 0.049813471 | 1 |
| ENSG00000253797 | UTP14C | -0.898538735 | 3.287708227 | 4.121031254 | 0.042353253 | 1 |
| ENSG00000134809 | TIMM10 | -0.895119223 | 3.479561631 | 4.429104376 | 0.035331043 | 1 |
| ENSG00000138074 | SLC5A6 | -0.887615885 | 5.432692276 | 6.140326571 | 0.013213291 | 1 |
| ENSG00000137936 | BCAR3 | -0.858841419 | 5.156230483 | 4.593684495 | 0.03208996 | 1 |
| ENSG00000197798 | FAM118B | -0.846777058 | 3.497312751 | 4.041443968 | 0.04439582 | 1 |
| ENSG00000103995 | CEP152 | -0.845395614 | 4.477584202 | 4.677824302 | 0.030554423 | 1 |
| ENSG00000144827 | ABHD10 | -0.833234171 | 3.807547552 | 4.119629316 | 0.042388365 | 1 |
| ENSG00000145088 | EAF2 | -0.819394884 | 6.300140577 | 3.908097283 | 0.048053974 | 1 |
| ENSG00000079950 | STX7 | -0.770904878 | 4.936922778 | 4.870972328 | 0.027312107 | 1 |
| ENSG00000133874 | RNF122 | -0.747238275 | 5.776211338 | 4.516240757 | 0.033574526 | 1 |
| ENSG00000178425 | NT5DC1 | -0.747028468 | 4.223160216 | 4.715964134 | 0.029883803 | 1 |
| ENSG00000117226 | GBP3 | -0.724020218 | 4.689972501 | 4.44398236 | 0.035024458 | 1 |
| ENSG00000130396 | AFDN | -0.718105402 | 5.424790542 | 5.339377681 | 0.020848915 | 1 |
| ENSG00000068654 | POLR1A | -0.684453578 | 5.947250436 | 5.051632511 | 0.024602748 | 1 |
| ENSG00000140563 | MCTP2 | -0.67325732 | 6.05901058 | 5.013957439 | 0.025143766 | 1 |
| ENSG00000066933 | MYO9A | -0.663896209 | 5.142943611 | 4.156665878 | 0.041470961 | 1 |
| ENSG00000168421 | RHOH | -0.663137789 | 6.988708035 | 3.875009239 | 0.04901004 | 1 |
| ENSG00000001631 | KRIT1 | -0.661329675 | 5.512385589 | 5.446314854 | 0.019609649 | 1 |
| ENSG00000196914 | ARHGEF12 | -0.648104834 | 6.891782884 | 4.749061265 | 0.029314282 | 1 |
| ENSG00000178764 | ZHX2 | -0.624249954 | 6.05272826 | 3.871665269 | 0.049107774 | 1 |
| ENSG00000125746 | EML2 | -0.608988324 | 6.011431403 | 4.617960034 | 0.031638845 | 1 |
| ENSG00000117899 | MESD | -0.551249111 | 5.772087115 | 4.043527302 | 0.044341051 | 1 |
| ENSG00000197943 | PLCG2 | -0.528799517 | 8.013474191 | 5.117448066 | 0.023686425 | 1 |

Supplementary Table 3. List of differentially expressed genes associated with MM progression. Differential gene expression testing between MGUS/SMM and MM groups, identified 250 genes approaching statistical significance (i.e. raw $p < 0.05$, before controlling the false discovery rate using the Benjamini Hochberg method²⁹). There were **(a)** 109 genes upregulated, and **(b)** 141 genes downregulated upon progression to MM. NA represents genes IDs that did not map to any known gene symbols.

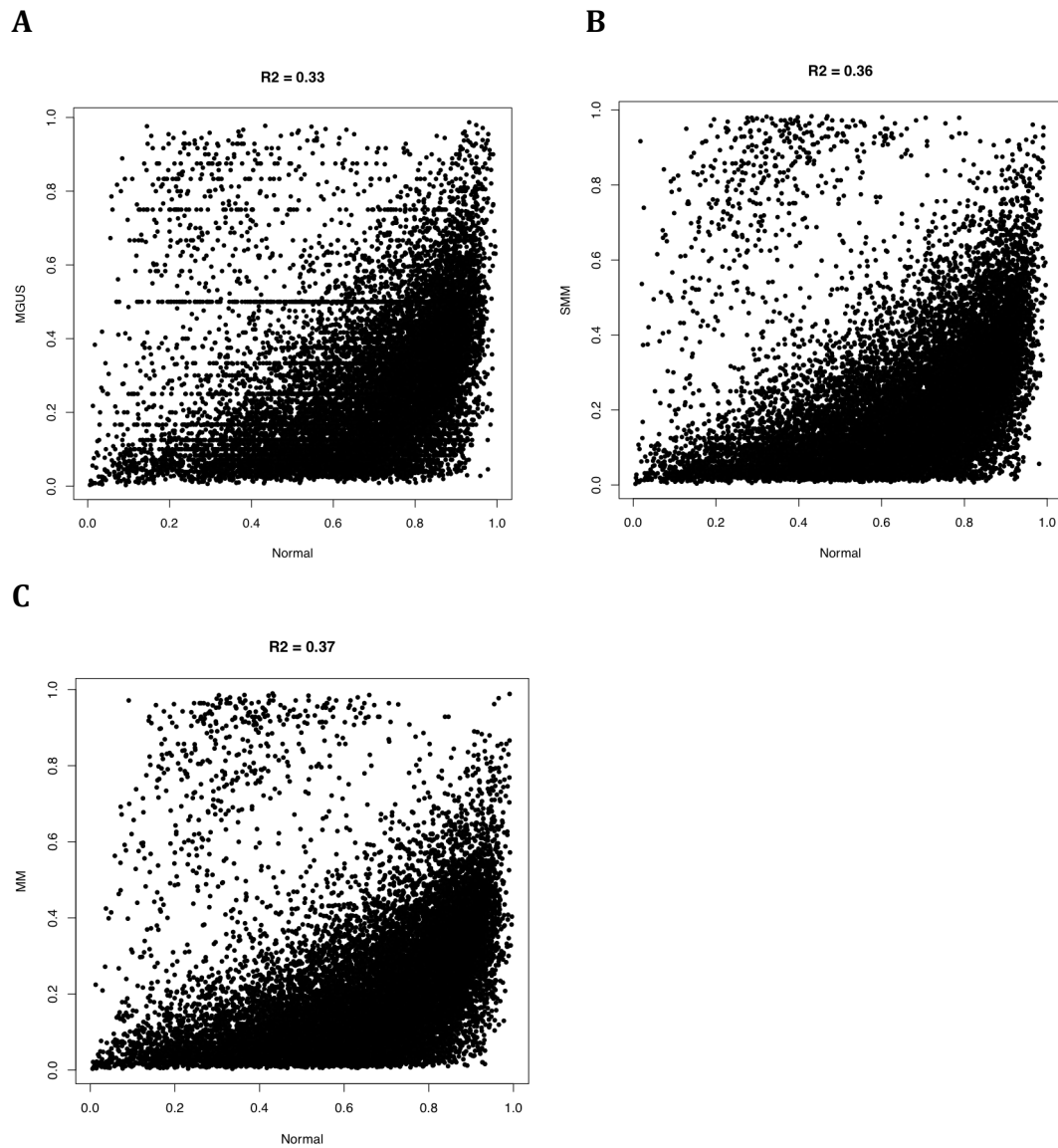
| Gene | P01 MGUS | P01 MM | P04 MGUS | P04 MM | P02 SMM | P02 MM | P03 SMM | P03 MM | P05 SMM | P05 MM | P08 SMM | P08 MM |
|---------|----------|----------|----------|----------|---------|----------|----------|----------|----------|----------|----------|----------|
| THEMIS2 | 97.7038 | 140.5189 | 138.4640 | 275.0199 | 57.8187 | 171.7534 | 465.0760 | 833.3867 | 247.6385 | 324.3043 | 125.0783 | 249.4285 |
| BTBD19 | 4.0761 | 5.0869 | 0.0547 | 2.3837 | 0.8070 | 6.7209 | 1.3894 | 11.0820 | 53.5205 | 68.9619 | 0.9629 | 2.8187 |
| HBB | 2.8409 | 1.4467 | 53.9627 | 0.6385 | 9.7102 | 5.5392 | 28.1909 | 3.4365 | 4.6724 | 0.0573 | 55.2919 | 38.7727 |
| ATP8A2 | 2.9233 | 0.0000 | 124.2892 | 76.8319 | 11.8709 | 6.9425 | 0.2742 | 0.0000 | 0.0531 | 0.0000 | 54.7344 | 27.4077 |
| CELSR1 | 4.2408 | 0.0000 | 2.1344 | 0.9790 | 1.3277 | 0.0369 | 9.0130 | 2.8238 | 152.6502 | 180.6527 | 5.5241 | 1.6493 |
| CD69 | 3.6232 | 2.7534 | 3.5026 | 4.3417 | 2.0306 | 0.0369 | 107.0594 | 5.9139 | 3.1592 | 0.0000 | 10.0346 | 2.9387 |
| TWF2 | 3.9526 | 19.2741 | 1.8608 | 7.4916 | 5.5710 | 13.8481 | 2.6326 | 3.4898 | 9.1059 | 30.4143 | 4.7132 | 5.0977 |
| SLC20A1 | 33.9267 | 50.7753 | 33.1657 | 61.0824 | 21.2427 | 67.2095 | 59.1971 | 140.4962 | 29.5743 | 29.1542 | 35.7801 | 48.5483 |
| ALG1L | 1.3587 | 0.0000 | 2.3533 | 1.7878 | 1.3277 | 0.0000 | 0.9324 | 0.0799 | 4.7255 | 0.0000 | 3.1928 | 1.9491 |
| SLC23A3 | 0.0000 | 4.7135 | 0.0000 | 0.2554 | 2.0566 | 1.4033 | 1.5723 | 2.8238 | 0.1593 | 11.5127 | 0.2027 | 3.2985 |

Supplementary Table 4. Normalised gene expression (FPKM) of genes differentially expressed at MM transition. There were 10 top differentially expressed genes, based on fold change and frequency, upon MM transition showing fold increase (*THEMIS2*, *BTBD19*, *TWF2*, *SLC20A1* and *SLC23A3*) or fold decrease (*HBB*, *ATP8A2*, *CELSR1*, *CD69* and *ALG1L*) based on their normalised gene expression counts.



- S-Hydroxytryptamine degradation (P04372) ⓘ
- SH2 type receptor mediated signaling pathway (P04374) ⓘ
- Adrenaline and noradrenaline biosynthesis (P00001) ⓘ
- Alzheimer disease-amyloid secretase pathway (P00003) ⓘ
- Alzheimer disease-prenelin pathway (P00004) ⓘ
- Angiogenesis (P00005) ⓘ
- Apoptosis signaling pathway (P00006) ⓘ
- Axon guidance mediated by netrin (P00009) ⓘ
- B cell activation (P00010) ⓘ
- Cadherin signaling pathway (P00012) ⓘ
- Cytoskeletal regulation by Rho GTPase (P00016) ⓘ
- De novo pyrimidine deoxyribonucleotide biosynthesis (P02739) ⓘ
- EGF receptor signaling pathway (P00018) ⓘ
- Endothelin signaling pathway (P00019) ⓘ
- FGF signaling pathway (P00021) ⓘ
- Formyltetrahydroformate biosynthesis (P02743) ⓘ
- GABA-B receptor II signaling (P05731) ⓘ
- Gonadotropin-releasing hormone receptor pathway (P06664) ⓘ
- Hedgehog signaling pathway (P00025) ⓘ
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026) ⓘ
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027) ⓘ
- Histamine H1 receptor mediated signaling pathway (P04385) ⓘ
- Huntington disease (P00029) ⓘ
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031) ⓘ
- Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032) ⓘ
- Integrin signalling pathway (P00034) ⓘ
- Ionotropic glutamate receptor pathway (P00037) ⓘ
- Metabotropic glutamate receptor group I pathway (P00041) ⓘ
- Metabotropic glutamate receptor group II pathway (P00040) ⓘ
- Metabotropic glutamate receptor group III pathway (P00039) ⓘ
- Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042) ⓘ
- Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043) ⓘ
- Nicotinic acetylcholine receptor signaling pathway (P00044) ⓘ
- Oxidative stress response (P00046) ⓘ
- Oxytocin receptor mediated signaling pathway (P04391) ⓘ
- PDGF signaling pathway (P00047) ⓘ
- Parkinson disease (P00049) ⓘ
- Ras Pathway (P04393) ⓘ
- Salvage pyrimidine ribonucleotides (P02775) ⓘ
- Synaptic vesicle trafficking (P05734) ⓘ
- Tetrahydrofolate biosynthesis (P02742) ⓘ
- Thyrotropin-releasing hormone receptor signaling pathway (P04394) ⓘ
- Toll receptor signaling pathway (P00054) ⓘ
- Transcription regulation by bZIP transcription factor (P00055) ⓘ
- VEGF signaling pathway (P00056) ⓘ
- Wnt signaling pathway (P00057) ⓘ
- p38 MAPK pathway (P05918) ⓘ

Supplementary Figure 1. Dysregulation of biological pathways associated with differentially expressed genes identified between MGUS/SMM to MM. Pathway analysis revealed gene expression signatures associated with the transition of MGUS/SMM to MM. There were 196 differentially expressed genes mapped, with their deregulation affecting 69 biological pathways. Pathways associated with genes that were highly up regulated, or highly down regulated at MM transition are illustrated in *Figure 3*.



Supplementary Figure 2. The relationship in methylation landscape in normal PCs versus MGUS, SMM and MM stages. XY scatterplots illustrate great variation in the methylation pattern between normal and disease stages of (A) MGUS, (B) SMM and (C) MM, with a skewing towards greater methylation observed in NPC. The median R^2 value was 0.36, demonstrating heterogeneity in CpG methylation between stage specific comparisons. R^2 correlation values are given for group above individual plots.

3.7 References

1. Bolli N, Avet-Loiseau H, Wedge DC, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*. 2014;5:2997.
2. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471(7339):467-472.
3. Lohr JG, Stojanov P, Carter SL, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. 2014;25(1):91-101.
4. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol*. 2017;14(2):100-113.
5. Walker BA, Boyle EM, Wardell CP, et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol*. 2015;33(33):3911-3920.
6. Walker BA, Wardell CP, Melchor L, et al. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*. 2014;28(2):384-390.
7. Kumar SK, Rajkumar SV, Kyle RA, et al. Multiple Myeloma. *Nature Reviews Disease Primers*. 2017;3(17046):1-20.
8. Szalat R, Avet-Loiseau H, Munshi NC. Gene Expression Profiles in Myeloma: Ready for the Real World? *Clin Cancer Res*. 2016;22(22):5434-5442.
9. Giancotti FG. Deregulation of cell signalling in cancer. *FEBS Lett*. 2014;588:2558-2570.
10. Decaux O, Lode L, Magrangeas F, et al. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myelome. *J Clin Oncol*. 2008;26(29):4798-4805.
11. Kuiper R, Broyl A, de Knecht Y, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia*. 2012;26(11):2406-2413.
12. Shaughnessy JD, Jr., Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;109(6):2276-2284.
13. Zhan F, Huang Y, Colla S, et al. The molecular classification of multiple myeloma. *Blood*. 2006;108(6):2020-2028.
14. Rashid NU, Sperling AS, Bolli N, et al. Differential and limited expression of mutant alleles in multiple myeloma. *Blood*. 2014;124(20):3110-3117.
15. Dimopoulos K, Gimsing P, Gronbaek K. The role of epigenetics in the biology of multiple myeloma. *Blood Cancer J*. 2014;4:e207.
16. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313.
17. Agirre X, Castellano G, Pascual M, et al. Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res*. 2015;25(4):478-487.
18. Salhia B, Baker A, Ahmann G, Auclair D, Fonseca R, Carpten J. DNA methylation analysis determines the high frequency of genic hypomethylation and low frequency of hypermethylation events in plasma cell tumors. *Cancer Res*. 2010;70(17):6934-6944.

19. Walker BA, Wardell CP, Chiecchio L, et al. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*. 2011;117(2):553-562.
20. Heuck CJ, Mehta J, Bhagat T, et al. Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis. *J Immunol*. 2013;190(6):2966-2975.
21. Dutta AK, Fink JL, Grady JP, et al. Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability. *Leukemia*. 2018:doi 10.1038/s41375-41018-40206-x.
22. Vandyke K, Zeissig MN, Hewett DR, et al. HIF-2alpha Promotes Dissemination of Plasma Cells in Multiple Myeloma by Regulating CXCL12/CXCR4 and CCR1. *Cancer Res*. 2017;77(20):5452-5463.
23. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics*. 2015;51:11 14 11-19.
24. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 10 11-33.
25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
26. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41(10):e108.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140.
28. Le SJ, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*. 2008;25(1).
29. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57:289-300.
30. Mi H, Huang X, Muruganujan A, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45(D1):D183-D189.
31. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013;8(8):1551-1566.
32. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol*. 2009;563:123-140.
33. Rajkumar SV. Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2016;91(7):719-734.
34. Anguiano A, Tuchman SA, Acharya C, et al. Gene expression profiles of tumor biology provide a novel approach to prognosis and may guide the selection of therapeutic targets in multiple myeloma. *J Clin Oncol*. 2009;27(25):4197-4203.
35. Zhan F, Hardin J, Kordsmeier B, et al. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood*. 2002;99(5):1745-1757.

36. Davies FE, Dring AM, Li C, et al. Insights into the multistep transformation of MGUS to myeloma using microarray expression analysis. *Blood*. 2003;102(13):4504 - 4511.
37. Cedar H, Bergman Y. Programming of DNA methylation patterns. *Annu Rev Biochem*. 2012;81:97-117.
38. Jung S, Kim S, Gale M, et al. DNA methylation in multiple myeloma is weakly associated with gene transcription. *PLoS One*. 2012;7(12):e52626.
39. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*. 1985;40(1):91-99.
40. Hardin J, Waddell M, Page CD, et al. Evaluation of multiple models to distinguish closely related forms of disease using DNA microarray data: an application to multiple myeloma. *Stat Appl Genet Mol Biol*. 2004;3:Article10.
41. Giuliani N, Storti P, Bolzoni M, Palma BD, Bonomini S. Angiogenesis and multiple myeloma. *Cancer Microenviron*. 2011;4(3):325-337.
42. Ria R, Reale A, De Luisi A, Ferrucci A, Moschetta M, Vacca A. Bone marrow angiogenesis and progression in multiple myeloma. *Am J Blood Res*. 2011;1(1):76-89.
43. Spaan I, Raymakers RA, van de Stolpe A, Peperzak V. Wnt signaling in multiple myeloma: a central player in disease with therapeutic potential. *J Hematol Oncol*. 2018;11(1):67.
44. Derksen PW, Tjin E, Meijer HP, et al. Illegitimate WNT signaling promotes proliferation of multiple myeloma cells. *Proc Natl Acad Sci U S A*. 2004;101(16):6122-6127.
45. Mrozik KM, Cheong CM, Hewett D, et al. Therapeutic targeting of N-cadherin is an effective treatment for multiple myeloma. *Br J Haematol*. 2015;171(3):387-399.
46. Groen RW, de Rooij MF, Kocemba KA, et al. N-cadherin-mediated interaction with multiple myeloma cells inhibits osteoblast differentiation. *Haematologica*. 2011;96(11):1653-1661.
47. Peirce MJ, Brook M, Morrice N, et al. Themis2/ICB1 is a signaling scaffold that selectively regulates macrophage Toll-like receptor signaling and cytokine production. *PLoS One*. 2010;5(7):e11465.
48. Zheng Y, Miyamoto DT, Wittner BS, et al. Expression of beta-globin by cancer cells promotes cell survival during blood-borne dissemination. *Nat Commun*. 2017;8:14344.
49. Buda G, Carulli G, Orciuolo E, et al. CD69 Expression Predicts Favorable Outcome in Multiple Myeloma Patients Treated with VTD. *Blood*. 2015;126(23):1768.

Chapter 4

***SP140*, a gene recurrently mutated in human myeloma,
is the target of RNA editing in the
5TGM1 murine myeloma cell line**

Statement of Authorship

| | |
|---------------------|--|
| Title of Paper | SP140, a gene recurrently mutated in human myeloma, is the target of RNA editing in the 5TGM1 murine myeloma cell line |
| Publication Status | <input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style |
| Publication Details | Dutta, A. K., Hewett, D. R., Kok, C. H., Mrozik, K. M. & Zannettino, A. C. W. (2018) SP140, a gene recurrently mutated in human myeloma, is the target of RNA editing in the 5TGM1 murine myeloma cell line. <i>Manuscript in preparation.</i> |

Principal Author

| | | | | |
|--------------------------------------|--|-----------|------|-----------|
| Name of Principal Author (Candidate) | Ankit K. Dutta | | | |
| Contribution to the Paper | Experimental design Generation of data Analysis and interpretation of data Manuscript development, writing and review | | | |
| Overall percentage (%) | 70% | | | |
| Certification: | This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper. | | | |
| Signature | <table border="1" style="width: 100%;"> <tr> <td style="width: 60%;"></td> <td style="width: 20%;">Date</td> <td style="width: 20%;">31/7/2018</td> </tr> </table> | | Date | 31/7/2018 |
| | Date | 31/7/2018 | | |

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

| | | | | |
|---------------------------|--|-----------|------|-----------|
| Name of Co-Author | Duncan R. Hewett | | | |
| Contribution to the Paper | Experimental design and assisted with experiments (Apobec KO cell line data) Interpretation of data Supervision of work Manuscript development and review | | | |
| Signature | <table border="1" style="width: 100%;"> <tr> <td style="width: 60%;"></td> <td style="width: 20%;">Date</td> <td style="width: 20%;">26/7/2018</td> </tr> </table> | | Date | 26/7/2018 |
| | Date | 26/7/2018 | | |

| | | | | |
|---------------------------|--|---------|------|---------|
| Name of Co-Author | Chung H. Kok | | | |
| Contribution to the Paper | Analysis of 5TGM1 RNAseq data | | | |
| Signature | <table border="1" style="width: 100%;"> <tr> <td style="width: 60%;"></td> <td style="width: 20%;">Date</td> <td style="width: 20%;">24/7/18</td> </tr> </table> | | Date | 24/7/18 |
| | Date | 24/7/18 | | |

| | | | |
|---------------------------|---------------------------------|------|---------|
| Name of Co-Author | Krzysztof M. Mrozik | | |
| Contribution to the Paper | Generation of 5TGM1 RNAseq data | | |
| Signature | | Date | 19/7/18 |

| | | | |
|---------------------------|--|------|-----------|
| Name of Co-Author | Andrew C. W. Zannettino | | |
| Contribution to the Paper | Experimental design Supervision of work Manuscript development and review Project funding | | |
| Signature | | Date | 24/7/2018 |

Chapter 4: *SP140*, a gene recurrently mutated in human myeloma, is the target of RNA editing in the 5TGM1 murine myeloma cell line

Ankit K. Dutta^{1,2}, Duncan R. Hewett^{1,2}, Chung H. Kok², Krzysztof M. Mrozik^{1,2}, Andrew C. W. Zannettino^{1,2}

¹Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA 5000 Australia.

²Cancer Theme, South Australian Health & Medical Research Institute, Adelaide, SA 5000 Australia.

Running title: *Sp140* is a target of RNA editing in 5TGM1 myeloma cells

Keywords: Sp140, RNA editing, 5TGM1, Myeloma, Mouse model

4.1 Abstract

Multiple myeloma (MM) is a largely incurable haematological malignancy characterised by the clonal proliferation of antibody secreting plasma cells (PCs) within the bone marrow. Increasingly, research has focused on the identification of key genetic mutations that underlie disease transformation from MGUS/SMM to MM. However to date, no single gene mutation, or combination of mutations, have been identified as being common to all MM patients at presentation¹. Multiple large cohort genomic studies of MM patients have been carried out finding *SP140* to be recurrently mutated in ~3-12% of patients. However, the precise role of mutated SP140 in MM development is unknown. Here, we demonstrate evidence of RNA editing of *Sp140* in the murine MM PC line, 5TGM1. RNA-specific nucleotide variants were identified in exon 2, including c.166 C>T (i.e. U) non-synonymous variant; and c.180 G>A synonymous variant. Screening of multiple non-PC murine cell lines determined that these editing changes were not a PC-specific phenomenon. Despite *Sp140* exon 2 sequence being conserved at c.166 in the human genome, screening of human MM cell line cDNA revealed no evidence of RNA editing in *SP140*. In relation to 5TGM1 cells, the non-synonymous C>T (i.e. U) change at c.166 predicted a STOP gain, prompting us to investigate whether the classical C>U deamination enzymes Apobec1 and Apobec3, were responsible for the RNA editing of *Sp140*. CRISPR-Cas9 gene editing was used to generate a 5TGM1 cell line in which *Apobec1* or *Apobec3* were knocked out. While *Apobec1* and *Apobec3* were efficiently knocked down in these cells, no consequent change in RNA editing of *Sp140* phenotype was observed, suggesting other mechanisms are responsible for the RNA editing of *Sp140*. The discovery that RNA editing of *Sp140* occurs in the 5TGM1 MM PC line warrants further exploration into the mechanisms that underlie this phenomenon, and suggests that RNA editing may influence the transcriptional and translational landscape of other MM associated genes.

4.2 Introduction

Genomic studies of MM patient samples have demonstrated intraclonal heterogeneity as a feature of disease through MGUS, SMM and MM, suggesting that progression is characterised by the acquisition of key genetic mutations which confer a selective advantage to tumour cells²⁻⁵. Significantly mutated genes that were identified in these studies include *KRAS*, *NRAS*, *BRAF*, *TP53*, *DIS3* and *FAM46C*, which are believed to be drivers of MM due to their recurrent nature²⁻⁵. Additionally, 28 novel candidate genes harbouring recurrent mutations associated with MM disease were also identified, including *IRF4*, *CDK4*, *ROBO1*, *FAT3*, *EGRI*, *PEG3*, *LTB*, *TGDS*, *SNX7*, *RASA2*, *USP29*, *TRAF3*, *CYLD*, *RBI*, *CCND1*, *PNRC1*, *ALOX12B*, *HLA-A*, *MAGED1*, *PRDMI*, *ACTG1*, *MAPK*, *NF1*, *NFKBIA*, *CDKN2C*, *PTEN*, *NFKBI* and *SPI40*^{2,3,5}. While many mutated genes have been identified to be associated with MM, few have been biologically interrogated for their causative roles in MM disease development.

SPI40 is a nuclear body protein involved in the antigen response of mature B cells^{2,6}. Further investigation has revealed its function as a bromodomain and plant homeodomain containing epigenetic reader⁷, with single nucleotide variants (SNVs) in *SPI40* significantly associating with immune diseases such as Crohn's disease^{8,9} and chronic lymphocytic leukaemia¹⁰. SNVs identified in *SPI40* resulted in aberrant mRNA transcription and reduced protein levels. Genomic studies of MM patients have found that *SPI40* is mutated in ~3-12% of patients, whose plasma cells (PCs) harbour a range of *SPI40* gene alterations including missense, nonsense, frameshift and splice site changes^{2,3,11,12}. Two of these studies have suggested *SPI40* as a novel candidate tumour suppressor gene in MM, due to the significant frequency of inactivating mutations^{2,3}. Moreover, Bolli *et. al.* showed that mutations in *SPI40* correlated with shorter relapse-free survival of MM patients².

Murine models of human disease play an important role in the screening and characterisation of mutant candidate genes in both *in vitro* and *in vivo* preclinical settings. In MM, the most commonly used mouse MM PC line is 5TGM1¹³. This well characterised cell line is amenable to genetic modification and can be used for both the *in vitro* and *in vivo* characterisation of the function of candidate genes of interest. The 5T MM models were originally identified in aging C57BL/KaLwRij mice that spontaneously developed benign B cell monoclonal proliferative disorders, which resembled human disease, including a monoclonal expansion of PCs within the bone marrow¹⁴. The 5TGM1 cell line

is a subclonal cell line of the 5T33 MM model¹³, a model which was developed by serial transplantation of bone marrow cells from aging C57BL/KaLwRij mice which displayed evidence of a serum monoclonal protein spike, into young recipients¹⁵. The 5TGM1-C57BL/KaLwRij *in vivo* model is the main preclinical model of MM used in the laboratory as it replicates many of the features of human MM disease¹⁶⁻¹⁸. Genetically modified 5TGM1 PCs are inoculated into 6-8 week old C57BL/KaLwRij mice, with subsequent effects on disease development monitored over a 4 week period. While MM is typically known as a single disease type in humans, recent genomic studies have demonstrated a marked genetic heterogeneity between patients, suggesting that MM is rather a collection of monoclonal gammopathies with different genetic subtypes which all share a common clinical phenotype¹⁹. As the 5TGM1 and C57BL/KaLwRij models are frequently used for *in vitro* and *in vivo* studies of MM, respectively, it would be advantageous to know which specific genetic subtype is being modelled in these studies for translatable outcomes in that patient group.

RNA sequencing (RNAseq) has been used widely to examine gene expression, however, there are limited studies investigating variant detection. Mining transcriptome sequencing data of 5TGM1 MM PCs, we investigated the mutation status of mRNAs of genes that are recurrently mutated in MM patients. To this end, we found that *Sp140* mRNA harboured a high impact SNV mutation that induced a STOP gain in exon 2 (c.166 C>T). Unexpectedly, validation studies of this SNV within 5TGM1 cells revealed the phenomenon of RNA editing, where the point mutation was identified in the screening of RNA but was not found in genomic DNA of 5TGM1 cells.

Recent studies have shown that modifications at the transcriptomic level are linked with cancer²⁰⁻²². Of particular interest, is the molecular mechanism of RNA editing, which is catalysed by two known editing enzyme families: apolipoprotein B mRNA editing enzyme, catalytic polypeptide like (APOBECs) and adenosine deaminases acting on RNA (ADARs). To date, the association of cancer and editing changes have been reported in an A>I context by ADARs. ADAR editing enzymes have been implicated in hepatocellular carcinoma²³, glioblastoma²⁴, prostate cancer²⁵, colorectal cancer²⁶, non-small-cell lung cancer²⁷, chronic myeloid leukaemia²⁰ and MM²². Examples in liver and lung cancers have shown that overexpression of ADAR1 lead to recoding changes in *AZIN1* that activates both the development and progression of disease, thus demonstrating the oncogenic potential of RNA editing enzymes^{23,27}. Similarly, APOBECs are a class of cytidine

deaminases known to catalyse C>U RNA editing changes, where currently APOBEC1, APOBEC3A and APOBEC3G are known to cause recoding changes²⁸. In this study, we identified a high impact C>T (ie. U) RNA editing change of *Sp140*, and hypothesized that Apobec family enzymes are the likely enzymatic candidates inducing this phenotype.

As outlined in our Whole Exome Sequencing (WES) study of serial MGUS/SMM to MM patients in Chapter 2, we also found that *SP140* mutations were newly acquired at MM in 30% of patients. In Chapter 3, which details our transcriptomic and methylomic analysis of serial MGUS/SMM to MM patients, we found that *SP140* was expressed in tumour PCs but did not harbour any point mutations. Here, we investigated the extent of RNA editing of *Sp140/SP140* in 6 murine (5TGM1, NIH-3T3, NS1, BA/F3, FDCP1 and RAW264.7) and 9 human MM cell lines (KMS-18, MM.1R, MM.1S, NCI-H929, RPMI 8226, U266, JJN3, MOLP-8 and EJM) to determine the tissue specificity of this RNA editing event and its prevalence in human MM. Moreover, CRISPR-Cas9 gene editing was used to generate 5TGM1 cell lines in which *Apobec1* or *Apobec3* were knocked out, to assess any subsequent changes in RNA editing of *Sp140* due to loss of these deaminase enzymes.

4.3 Materials & Methods

4.3.1 Sequencing and analysis

RNA sequencing of 5TGM1 cells was previously performed by Dr. K. Mrozik. Briefly, total cellular RNA was extracted from cells using 1mL TRIzol as per manufacturer's instructions (Thermo Fisher Scientific). All RNA samples were dissolved in nuclease-free water (30µL). Yields and quality were assessed using the Qubit 2.0 fluorometer (Thermo Fisher Scientific) and 2200 TapeStation System (Agilent). The range of RIN scores of extracted RNA samples was 8.0 – 8.3. RNA libraries were generated using the TruSeq Stranded mRNA Sample Preparation Kit, using the LS (low sample) protocol according to manufacturers' instructions (Illumina). Sequencing was carried out on the Illumina NextSeq500 (2x75 bp paired-end reads) with approximately 100 million reads per sample.

Gene analysis was performed by Dr. C. Kok. Quality of raw RNA sequencing reads was assessed using the FastQC package (Babraham Bioinformatics). Sequencing reads were aligned to the *mus musculus* mm10 genome assembly using the Subread aligner package. Uniquely mapped reads were retained and the number of reads that mapped were counted using featureCounts. Transcripts were filtered from downstream analysis if they did not meet the threshold of at least 1 count per million mapped reads, in at least 2 samples. Reads counts per gene were converted to log₂ counts per million (CPM) with the voom function of the limma package. Single nucleotide variants were called using the Genome Analysis Tool Kit (GATK) best practices pipeline (Broad Institute), and variant annotation was performed using ANNOVAR²⁹.

4.3.2 Cell culture

4.3.2.1 Murine cell lines

All cell culture media was supplemented with additives consisting of 2mM L-glutamine, 1mM sodium pyruvate, 15mM HEPES, 50 U/mL penicillin and 50µg/mL streptomycin (Sigma-Aldrich). The 5TGM1 MM cell line was cultured in Iscove's Modified Dulbecco's Media (IMDM) supplemented with 20% fetal calf serum (FCS) and additives. NIH-3T3, NS1 and RAW264.7 cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FCS and additives. BA/F3 cells were cultured in Roswell Park Memorial Institute media (RPMI-1640) supplemented with 10% FCS, additives and 10% (v/v) conditioned media from the WEHI-3B cell line, as a source of IL-3 (a gift from K. Asari of the SAHMRI Cancer Theme group). FDCP1 cells were cultured in DMEM supplemented with 10% FCS, additives and 10% (v/v) conditioned media of the

WEHI-3B cell line for IL-3. Cell cultures were maintained in humidified incubation chambers at 37°C with 5% CO₂.

4.3.2.2 Human MM cell lines

All human cell lines were cultured in RPMI-1640 supplemented with 10% FCS, with the exception of the EJM cell line, which was cultured in IMDM supplemented with 10% FCS. Cell cultures were maintained in humidified incubator at 37°C with 5% CO₂.

4.3.3 Nucleic acids isolation

4.3.3.1 DNA isolation

Genomic DNA was extracted from cell lines using the DNeasy Blood & Tissue kit, according to the manufacturer's instructions (QIAGEN). Quantity and quality of isolated DNA was determined using the Nanodrop 8000 Spectrophotometer (Thermo Fisher Scientific) and all samples were stored at -80°C until required.

4.3.3.2 RNA isolation

Total RNA was extracted from cell lines using TRIzol (Life Technologies). RNA yields and quality were assessed using the Nanodrop 8000 Spectrophotometer (Thermo Fisher Scientific) and all samples were stored at -80°C until required.

4.3.4 Validation

4.3.4.1 Confirming SNVs in 5TGM1 cell line

cDNA (20µL) was generated from 1.5µg of total RNA extracted from 5TGM1 cells using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). *Sp140* exon 2 was then PCR-amplified using Sp140RT-F: 5'- CCAGAGGACCAGAATGAAGAGG -3' and Sp140RT-R: 5'- TCCCGGCTAAACTTCTTCTGT -3' primers and 2µL cDNA with Phusion High-Fidelity DNA Polymerase (New England BioLabs), as per manufacturers' instructions for 50µL total reaction volume, on a Veriti 96-Well Thermocycler (Applied Biosystems) for 35 cycles with 98°C denaturation, 60°C annealing and 72°C extension. The PCR product was resolved on a 2% agarose gel and subsequently excised, and purified using the Ultra Clean PCR Purification Kit (MoBio Laboratories) before direct Sanger sequencing using PCR primers (as above), at the Australian Genome Research Facility (AGRF Adelaide Node).

DNA was screened by direct purified PCR product screening, as described above with gDNA PCR primers designed to amplify *Sp140* exon 2 (F: 5'-CCAGCCATTGGTAGCATCTTG -3'; R: 5'- GCTTGCTGTCAGGACTGAGT -3').

Same molecule analysis of tandem editing of RNA was validated by PCR product cloning and colony sequencing. *Sp140* exon 2 was PCR amplified using cDNA primers, before adenylation using AmpliTaq Gold (Thermo Fisher Scientific). Amplicons were then ligated into the pGEM-T vector (Promega Corporation). JM109 competent cells were transformed and plated onto Luria broth with ampicilin agar plates for colony growth overnight at 37°C. Bacterial colonies (n = 10) were randomly selected, cultured overnight and then plasmid minipreps were carried out using alkaline lysis miniprep (QIAGEN). Double digest, using restriction enzymes NdeI and SacII, was used to confirm successful ligation of the exon 2 PCR product into the vector. Extracted DNA was quantitated on the Nanodrop 8000 (Thermo Fisher Scientific), prior to direct Sanger sequencing using T7 forward and SP6 reverse vector primers at the Australian Genome Research Facility (AGRF Adelaide Node).

DNA and RNA of *Sp140* exon 7 was screened by direct purified PCR product screening, as described above in 4.3.4.1 *Validation: Confirming SNVs in 5TGM1 cell line*, with PCR primers designed to exon 7 gDNA (F: 5'- CAGGATGCCTCCCTTTCTCC -3'; R: 5'- GAAAGACCCACAGACGCTGT -3') and cDNA (F: 5'- GGCCACAACCTGGTCAAACC -3'; R: 5'- GGTTCTTTTTTCATCACTCCCTTCA -3').

4.3.5 Screening

4.3.5.1 Murine cell lines

Screening of *Sp140* exon 2 changes were carried out on 6 mouse cell lines: 5TGM1, NIH-3T3, NS1, BA/F3, FDCP1 and RAW264.7. Nucleic acids were isolated and quantitated as described above. RNA and DNA were screened by direct purified PCR product sequencing, using gDNA and cDNA PCR primers as described above in 4.3.4.1 *Validation: Confirming SNVs in 5TGM1 cell line*.

4.3.5.2 Human MM cell lines

Screening of *SP140* exon 2 changes was carried out on 9 human MM cell lines: KMS-18, MM.1R, MM.1S, NCI-H929, RPMI 8226, U266, JLN3, MOLP-8 and EJM. Nucleic acids were isolated and quantitated as described above. RNA was screened by direct purified

PCR product screening, as described using above in 4.3.4.1 *Validation: Confirming SNVs in 5TGM1 cell line*, with cDNA PCR primers (F: 5'- GGCAAGTGGAGACAGCAATC - 3'; R: 5'- CTTGTCACTGGGACCAGGTT -3').

4.3.6 Generation of Apobec1 sgRNA expression vectors

Plasmid vector pSpCas9(BB)-2A-GFP (also known as px458) was a gift from Feng Zhang and was obtained from Addgene (plasmid #48138)³⁰. The Cerulean2 reporter was amplified from the LeGO-iCer2 lentiviral vector (Addgene plasmid #27346, gift of Boris Fehse)³¹ and a T2A self-cleaving peptide sequence was also added using two rounds of hemi-nested PCR. An initial 30 cycles of PCR using T2A-Cer2-Forward (5'- GTCGAGGAGAATCCTGGCCCAGTGAGCAAGGGCGAGGAGCTG-3') and EcoRI-TdT-Cer2-Reverse (5'-GCCGGAATTCTTACTTGTACAGCTCGCTCAT-3') primers was followed by another 30 cycles of PCR using EcoRI-T2A-Forward (5'- AAAGGAATTCGGCAGTGGAGAGGGCAGAGGAAGT-CTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA -3') and EcoRI-TdT-Cer2-Reverse primers. Both rounds of PCR used Phusion High-Fidelity DNA Polymerase (New England Biolabs) and a T_m of 58°C. The PCR product was ligated into px458 using the EcoRI sites, replacing the EcoRI-site-flanked T2A-GFP cassette, to generate px458-Cer2. sgRNAs to exon 6 of mouse Apobec1 (ensembl.org transcript ENSMUST00000112586.7) were designed using the MIT CRISPR design software (<http://crispr.mit.edu/>). The following oligonucleotide pairs, specific for each sgRNA, were ligated into px458-Cer2 using the tandem BbsI sites downstream of the U6 promoter as per Ran, *et al.*³⁰. sgRNA **emboldened**.

Apobec1_sgRNA#1_top: 5'-CACCGA**AGAAGACTTCAA**ACTCGTG-3'

Apobec1_sgRNA#1_bottom: 5'-AAACCACGAGTTTGAAGTCTTCTTC-3'

Apobec1_sgRNA#2_top: 5'-CACCGTCTCTTT**CCGAAGCTCCC**GG-3'

Apobec1_sgRNA#2_bottom: 5'-AAACCCGGGAGCTTCGGAAAGAGAC-3'

4.3.6.1 Transfection of 5TGM1 cells

4mL of 5TGM1 BMx1 cells at 2×10^5 per mL in IMDM (20% FCS) were transfected with either 4µg px458-sgRNA plasmid, or 4µg px458-empty vector (no sgRNA) plasmid and 20µL Polyfect (QIAGEN) according to manufacturer's recommendations. 48 hours after transfection, GFP⁺Cerulean⁺ cells were isolated using flow cytometry. Ten days later, single transfected cells (clones) were deposited into 96 well plates by flow cytometry and clones were expanded prior to mutation screening.

4.3.6.2 Mutation screening

DNA was isolated from each clone using DNeasy Blood and Tissue kit (QIAGEN) according to manufacturer's instructions. Exon 6 of mouse Apobec1 (ensembl.org transcript ENSMUST00000112586.7) was amplified using mApobec1.het.F (5'-GTCATCTCAGCCTGGAATATG -3') and mApobec1.het.R (5'-GGCTCAGAACTCTGTAATGG -3') primers using Phusion High-Fidelity DNA Polymerase (New England Biolabs) at a T_m of 70°C. PCR products were purified using an UltraClean™ PCR clean-up kit (MoBio Laboratories) prior to Sanger sequencing at the Australian Genome Research Facility (AGRF Adelaide Node).

4.3.7 Generation of Apobec3 sgRNA expression vectors

The GFP reporter gene was excised from lentiviral vector FgH1tUTG (a gift from Marco Herold, Addgene plasmid # 70183)³² and replaced with the Plum reporter gene using Gibson assembly following the Miller laboratory recommendations (<http://miller-lab.net/MillerLab/protocols/>) by Yu Chinn Joshua Chen at University of Adelaide. Two different guide RNAs were designed using the MIT CRISPR design software (<http://crispr.mit.edu/>) to sequences just upstream of each of those encoding the two active sites (AS) of mouse Apobec3. The following oligonucleotide pairs were used to clone the sgRNAs targeting AS1 and AS2 encoded by exons 3 and 7, respectively, of mouse Apobec3 (ensembl.org transcript ID ENSMUST00000109620.9). The 4 bp CACC- and AAAC- overhangs enabled ligation into BsmBI digested FgH1UTPlum.

AS1-antisense-F 5' CACCGACATTCGAAACAGGGGCTCC 3'

AS1-antisense-R 5' AAACGGAGCCCCTGTTTCGAATGTC 3'

AS1-sense-F 5' CACCGAGATCACCTGGTATATGTCC 3'

AS1-sense-R 5' AAACGGACATATACCAGGTGATCTC 3'

AS2-antisense-F 5' CACCGACAGTTTGGGCAGGGGCTCC 3'

AS2-antisense-R 5' AAACGGAGCCCCTGCCCAAACACTGTC 3'

AS2-sense-F 5' CACCGCAATCACCTGCTACCTCACC 3'

AS2-sense-R 5' AAACGGTGAGGTAGCAGGTGATTGC 3'

4.3.7.1 Viral packaging and 5TGM1 infection

5TGM1 cells were transduced with FUCas9Cherry (a gift from Marco Herold: Addgene plasmid # 70182) to constitutively express Cas9. Viral particles containing FuCas9cherry were generated following Lipofectamine 2000 (Invitrogen) transfection of HEK293T cells concomitantly with psPAX2 (gift from Didier Trono, Addgene plasmid # 12260) and

pECO (Clontech) packaging constructs. 5TGM1 cells were infected with lentiviral particles by centrifugation at 1000g for 1 hour with 8µg/ml polybrene. Forty hours later, cherry-positive 5TGM1 cells that had been successfully transduced with Cas9 library were isolated by flow cytometry. 5TGM1-FuCas9cherry were then transduced with similarly packaged FgH1tUTPlum-Apobec3-sgRNA or FgH1tUTPlum-empty vector lentiviral vectors and Cherry⁺Plum⁺ cells were isolated by flow cytometry. sgRNA expression was induced by the addition of doxycycline hyclate (Sigma-Aldrich) to 1µg/ml for 48 hours, and individual mCherry⁺Plum⁺ cells were deposited into a 96 well plate by flow cytometry. These 5TGM1 clones were expanded prior to mutation screening.

4.3.7.2 Mutation screening

DNA was isolated from each clone using DNeasy Blood and Tissue kit (QIAGEN) according to manufacturer's instructions. Exon 3 or exon 7 of mouse Apobec3 (ensembl.org transcript ID ENSMUST00000109620.9) was amplified using either Exon3-F (5' AACAGGGCTCAGAGTGCTAG 3') and Exon3-R (5' ACACACCCTTCACCATATGG 3') or Exon7-F (5' TGGGAATGTGGAGTTAGTGG 3') and Exon7-R (5' GGCTTGTCATATTGAGGCTG 3') primers and Phusion High-Fidelity DNA Polymerase (New England Biolabs) at T_ms of 70°C or 67°C, respectively. PCR products were purified using an UltraClean PCR clean-up kit (MoBio Laboratories) prior to Sanger sequencing at the Australian Genome Research Facility (AGRF Adelaide Node).

4.3.8 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

cDNA was generated from RNA isolated from murine cell lines using Superscript III First-Strand Synthesis System (Life Technologies). qPCR was performed using RT2 SYBR Green qPCR Mastermix (QIAGEN) on the Bio-Rad CFX Connect machine (Bio-Rad) using primers targeting *Apobec1* (F: 5'- CTGTAGCTGTTGATCCCAC -3'; R: 5'- CTAAGAAGTTGACTTCAACG -3'), *Apobec3* (F: 5'- CAGCAGAATCTTTGCAGG -3'; R: 5'- CAGAATCTCCTGAAGCTTAG -3') and housekeeping control β -actin, *ActB* (F: 5'-TTGCTGACAGGATGCAGAAG-3'; R: 5'-AAGGGTGTAACGCGAGCTC-3'). Gene expression was calculated using the comparative C_T method relative to *ActB* expression³³.

4.4 Results

4.4.1 *Sp140* exhibits high impact mutations in the transcriptome of 5TGM1 cells

Transcriptomic sequencing data of murine 5TGM1 MM PCs was examined for any nucleotide variants in genes that have previously been reported to be recurrently mutated in MM patients²⁻⁵ [*Supplementary Table 1*]. Genome Analysis Toolkit (GATK) annotated RNAseq data was filtered to reveal that 11 out of the 34 reported significantly mutated genes in human MM patients also harboured nucleotide variants in 5TGM1 MM PCs. This included *KRAS*, *NRAS*, *TP53*, *BRAF*, *DIS3*, *SP140*, *RASA2*, *PNRC1*, *HLA-A*, *NF1* and *PTEN*, which harboured SNVs. These variants of interest were then analysed using the Ensembl Variant Effect Predictor tool (<https://asia.ensembl.org/Tools/VEP>) to determine if any of the identified SNVs resulted in changes to the protein synthesis sequence. To this end, variants in 5 genes: *TP53*, *BRAF*, *SP140*, *HLA-A* (H2-Q7) and *NF1* had potential consequences on protein sequence [*Supplementary Table 2*]. Of these, *Sp140* was found to harbour high impact base changes inducing a STOP gain [c.166C>T resulting in p.R56*] and STOP loss [c.478T>A resulting in *160K] in exons 2 and 7, respectively [*Figure 1*]. Of note, as this is RNA data, nucleotide Thymidine (T) should be interpreted as nucleotide Uracil (U). RNAseq sequencing reads were explored in the Integrative Genomics Viewer (IGV) to confirm the identified point mutations of interest, which showed an approximate 50:50 ratio for each nucleotide change [*Supplementary Fig. 1*]. We hypothesised that the STOP gain mutation in exon 2 would result in the formation of a truncated Sp140 protein, while the STOP loss mutation in exon 7 would generate a larger than expected protein. In both cases, the abnormal proteins may contribute to MM pathogenesis. Due to the nature of our discovery in the RNAseq data, we went on to validate our findings using Sanger sequencing of the *Sp140* exonic regions of interest.

a

| Gene | Position | Consequence | Impact | Exon | Codon (Ref) | Codon (Alt) | Amino Acid (Ref) | Amino Acid (Alt) |
|--------------|----------|-------------|--------|------|---------------------|---------------------|------------------|------------------|
| <i>Sp140</i> | 85609827 | Stop gain | HIGH | 2\17 | <u>C</u> GA | <u>T</u> GA | R [ARG] | * [STOP] |
| | 85609841 | Synonymous | LOW | 2\17 | <u>T</u> C <u>G</u> | <u>T</u> C <u>A</u> | S [SER] | S [SER] |
| | 85628219 | Stop loss | HIGH | 7\8 | <u>T</u> AA | <u>A</u> AA | * [STOP] | K [LYS] |

b

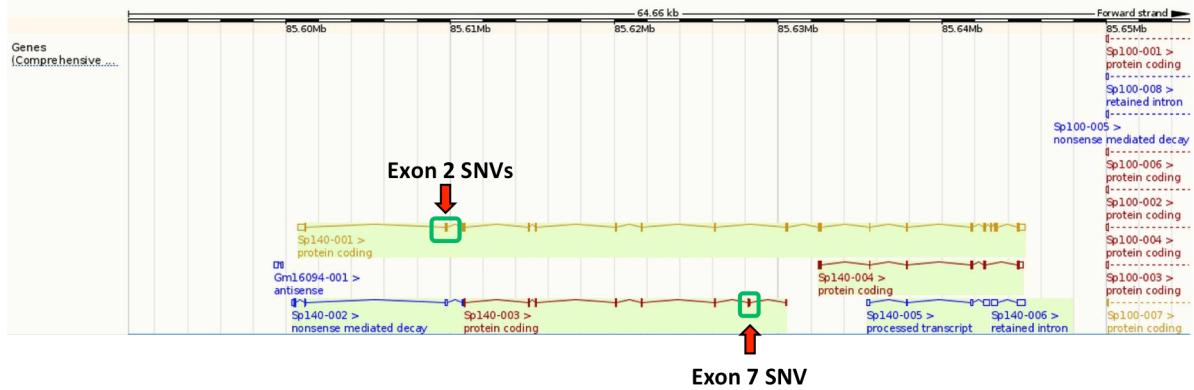
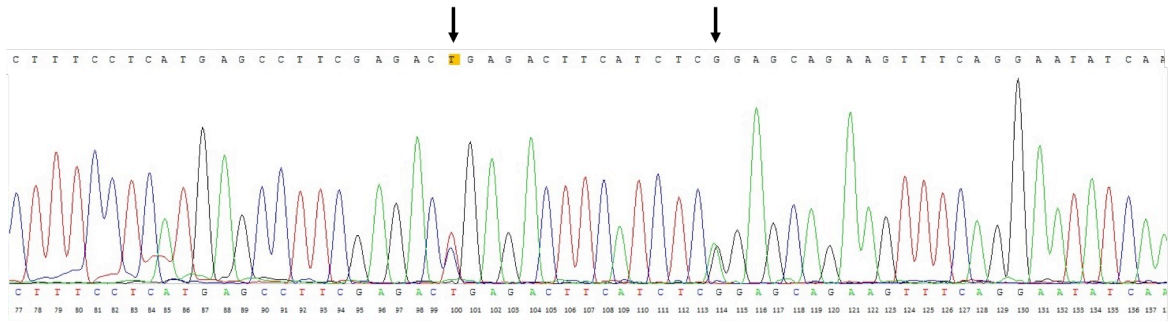


Figure 1. Identification of *Sp140* SNVs in the transcriptome of 5TGM1 MM PCs. (a) Variant Effect Predictor characterises 3 SNVs found in 5TGM1 MM PC mRNA transcripts. Notably, 2 variants are proposed to induce high impact consequences in *Sp140* protein formation through STOP gain in exon 2 and STOP loss in exon 7. **(b)** SNV locations in *Sp140* gene transcripts.

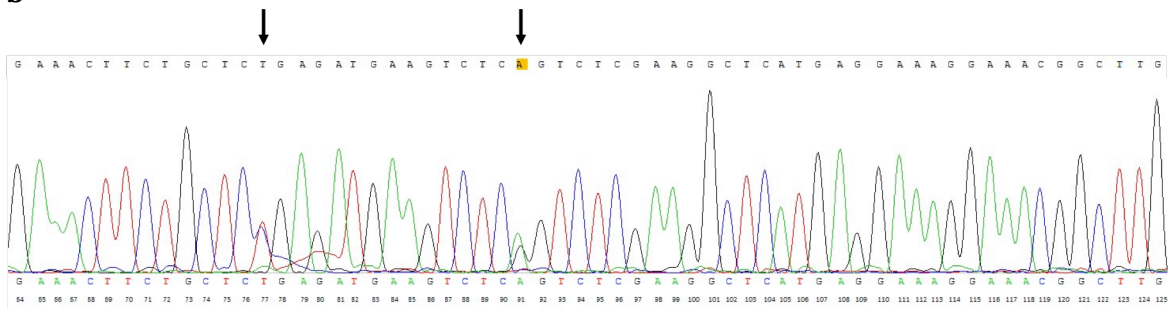
4.4.2 *Sp140* is a target of RNA editing in 5TGM1 PCs

Sanger sequencing validation of the RNA SNVs was carried out on the same samples that were prepared for RNAseq. cDNA was generated for PCR amplification of both exon 2 and exon 7 regions, before sequencing was performed to confirm the identified variants affecting *Sp140*. Sequence traces of the exon 2 amplicon showed heterozygous alterations from the reference genome at both identified SNV sites c.166C>T (i.e. U) (nonsense) and c.180G>A (synonymous) [Figure 2a, 2b]. Exon 7 amplicon trace showed the identified alterations to be homozygous at c.478T>A (STOP loss) [Figure 2c, 2d].

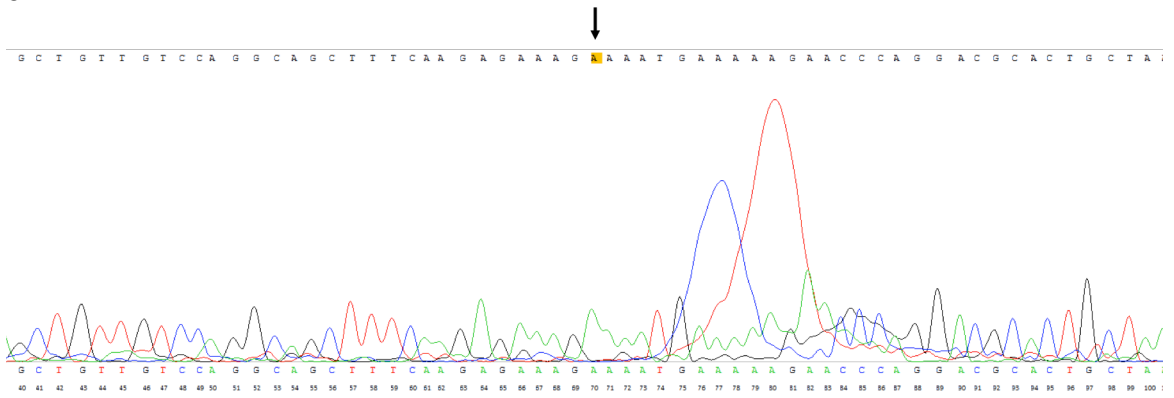
a



b



c



d

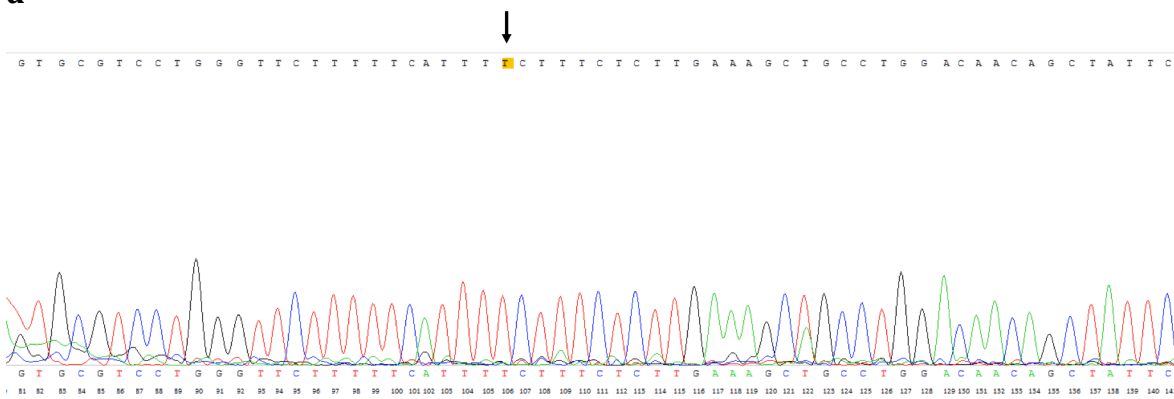
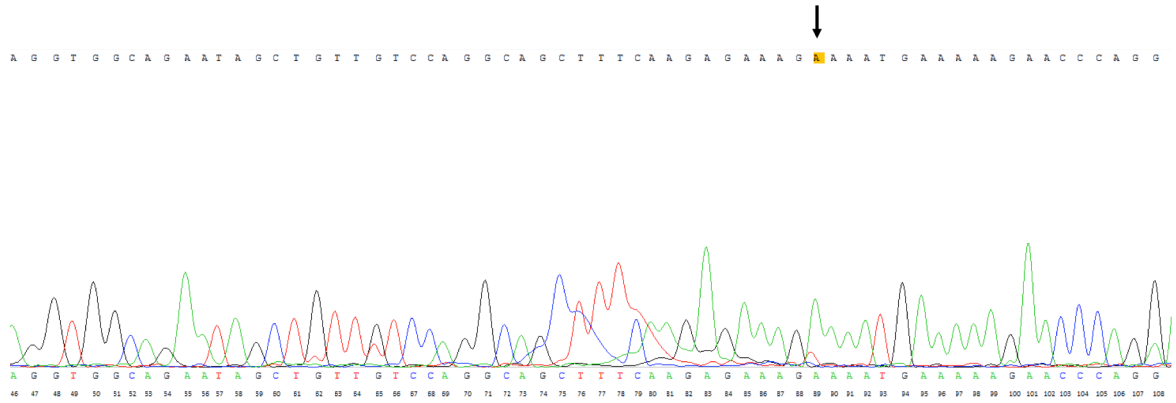


Figure 2. Validation of RNA modification in *Sp140*. Sanger sequencing confirms c.166C>T (i.e. U) and c.180G>A variants in **(a)** forward and **(b)** reverse sequencing traces of *Sp140* exon 2. Both variants appear to be heterozygous. Similarly, exon 7 variant c.478T>A was confirmed in **(c)** forward and **(d)** reverse sequencing traces. Altered sites are indicated by arrows, with missense c.166C>T (i.e. U) and c.478T>A highlighted in yellow.

In order to assess whether the nucleotide changes were due to mutations present at the DNA level, genomic DNA (gDNA) sequencing was performed on PCR amplified gDNA from exon 2 and exon 7. While the initial VEP analysis of RNAseq data had detected a point mutation in exon 7 with potential consequences, validation of gDNA revealed it to be a homozygous SNV at the genomic level [Figure 3a]. Ensembl genome browser also revealed that this site is a catalogued STOP loss T/A variant in mouse (dbSNP rs1131936885). Subsequently, we focused our efforts on the identified STOP gain variant in exon 2. Notably, gDNA sequencing did not verify the identified SNVs in exon 2, demonstrating a clean sequencing trace supporting the reference allele at both positions [Figure 3b]. Moreover, we confirmed that this result was not restricted to a particular primer pair set, with the use of a second set of primers designed to amplify a larger region of exon 2 (termed Exon 2L). Sanger sequencing of these amplicons again supported the reference allele at both positions [Supplementary Figure 3]. Comparison of the Sanger sequencing data from both RNA and DNA suggests these mRNA alterations may be caused by RNA editing. Furthermore, clonal analysis of *Sp140* exon 2 RT-PCR products revealed that both editing changes always occurred in *cis* on the same strand [Figure 4]. RNA editing is a process involving the post transcriptional modification of RNA transcript bases, which is enzymatically catalysed by deaminase enzymes of the APOBEC or ADAR families. As RNA editing is known to increase the complexity of expressed transcripts in cancer cells and may provide a selective advantage for growth²⁸, this phenomena was investigated further.

a



b

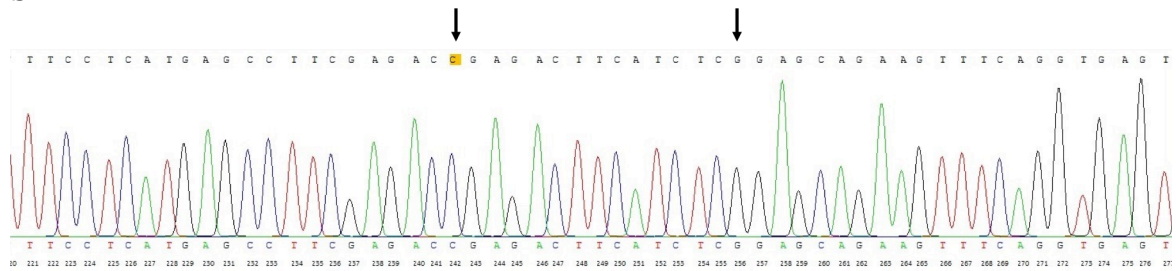
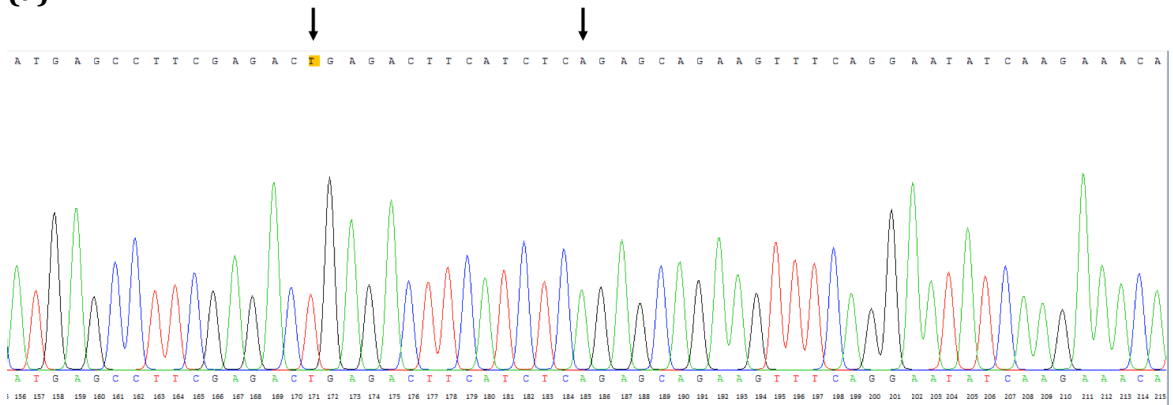
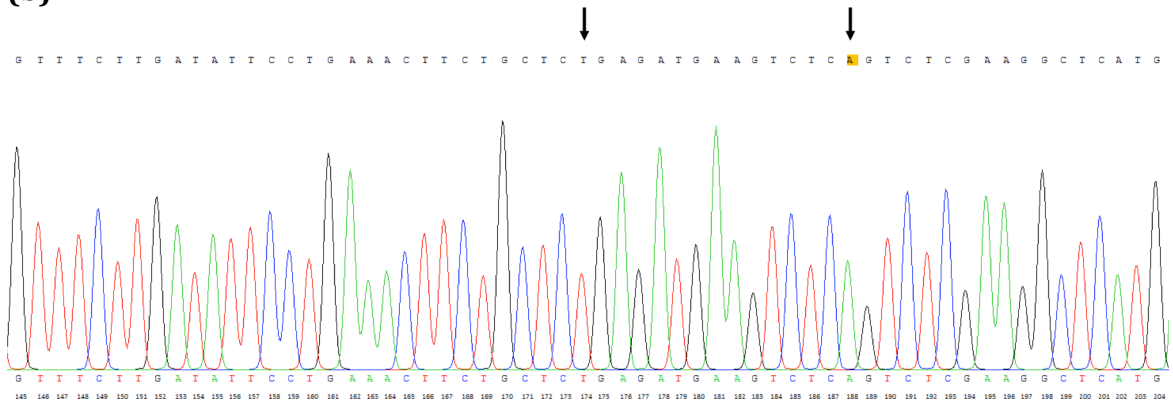


Figure 3. Validation of identified RNA modifications in *Sp140* at the gDNA level. (a) Exon 7 is mutated at the genomic DNA level in 5TGM1 PCs. Sanger sequencing trace supports a prominent missense change to alternate nucleotide A (reference nucleotide is T) at position c.478. **(b)** Exon 2 of *Sp140* is unmutated at the genomic level in 5TGM1 PCs. Sanger sequencing reveals a clean sequencing trace that supports the reference nucleotides of C and G at positions c.166 and c.180, respectively. Sites of interest are indicated by arrows, with position c.166 and c.478 highlighted in yellow.

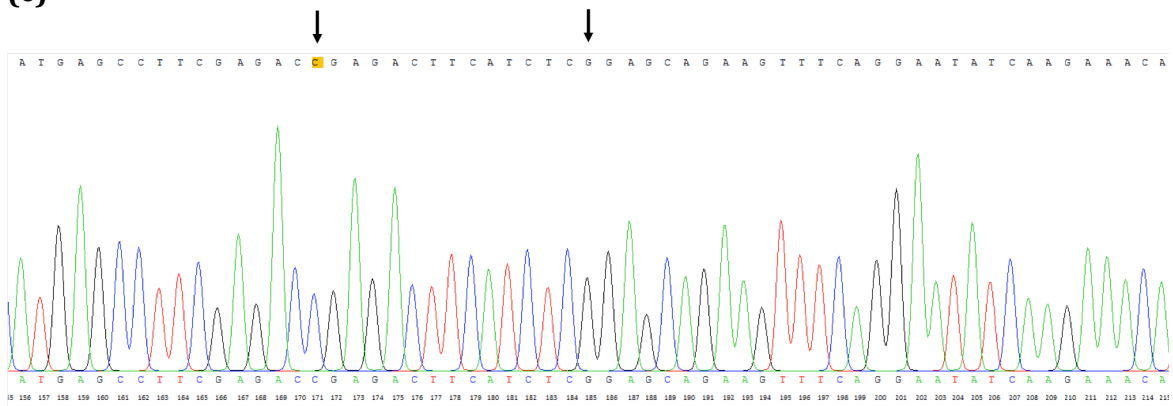
(a)

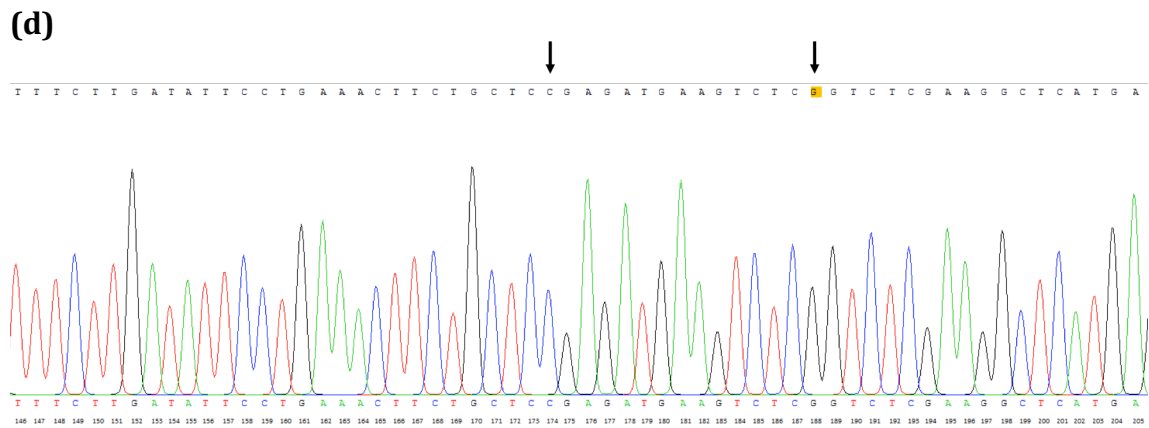


(b)



(c)





(e)

| Colony | Forward (T7) or Reverse (SP6) | Site 1 c.166 | Site 2 c.180 | Status |
|--------|-------------------------------|--------------|--------------|---------------|
| 1 | Forward | C | G | Unedited |
| | Reverse | G | C | |
| 2 | Forward | C | G | Unedited |
| | Reverse | G | C | |
| 3 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |
| 4 | Forward | C | G | Unedited |
| | Reverse | G | C | |
| 5 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |
| 6 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |
| 7 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |
| 8 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |
| 9 | Forward | C | G | Unedited |
| | Reverse | G | C | |
| 10 | Forward | T | A | <i>Edited</i> |
| | Reverse | A | T | |

Figure 4. Clonal analysis demonstrates association of the two *Sp140* exon 2 RNA editing changes. Comparative analysis of representative sequencing traces of cloned RT-PCR products from an edited *Sp140* RNA molecule (clone 3) ((a) forward and (b) reverse sequence) versus an unedited *Sp140* RNA molecule (clone 2) ((c) forward and (d) reverse sequence) illustrates that RNA editing of both c.166 and c.180 sites occur in *cis* on the same RNA molecule. Sites of interest are indicated by arrows, with position c.166 highlighted in yellow. (e) A total of 10 clones were sequenced, with data supporting *cis* RNA editing changes in *Sp140* exon 2 with either tandem editing (6/10 clones) or tandem non-editing (4/10 clones).

In view of these findings, we sought to explore whether RNA editing activity was a plasma cell specific phenomenon. In addition to 5TGM1, five other murine cell lines including NIH-3T3 (embryonic fibroblast), NS1 (B lymphocyte), BA/F3 (pro B cell), FDCP1 (myeloid progenitor) and RAW264.7 (macrophage) were investigated. Specifically, cDNA was generated from RNA isolated from each cell line and the *Sp140* exon 2 region was amplified by RT-PCR. Sanger sequencing analysis of PCR products revealed that both c.166 and c.180 sites were subject to varying levels of RNA editing across the different cell lines [Figure 5a]. With 5TGM1 cells used as an editing reference control, FDCP1 and NS1 cells exhibited prominent non-synonymous C>T (i.e. U) changes with ~57% and ~30% of c.166 edited, respectively. NIH-3T3, BA/F3 and RAW264.7 cells also showed editing changes, however, the occurrence of editing was at much lower frequency with ~12%, ~9% and ~9% of c.166 edited, respectively [Figure 5b] [Supplementary Table 3]. These findings suggest that site specific RNA editing alterations are not exclusive to cells of PC origin and are found in other murine cell lines and their respective different mouse strains including NIH/Swiss (NIH-3T3), BALB/c (NS1 and RAW264.7), C3H (BA/F3) and DBA/2 (FDCP1). While RNA editing of *Sp140* was not unique to 5TGM1 PCs, the rates were highest with ~59% of c.166 edited. Moreover, *SP140* has been identified to be recurrently mutated in patient derived MM PC studies and has been suggested to be a tumour suppressor due to recurrent inactivating nonsense mutations^{2,3,11}, prompting us to examine this phenomenon further.

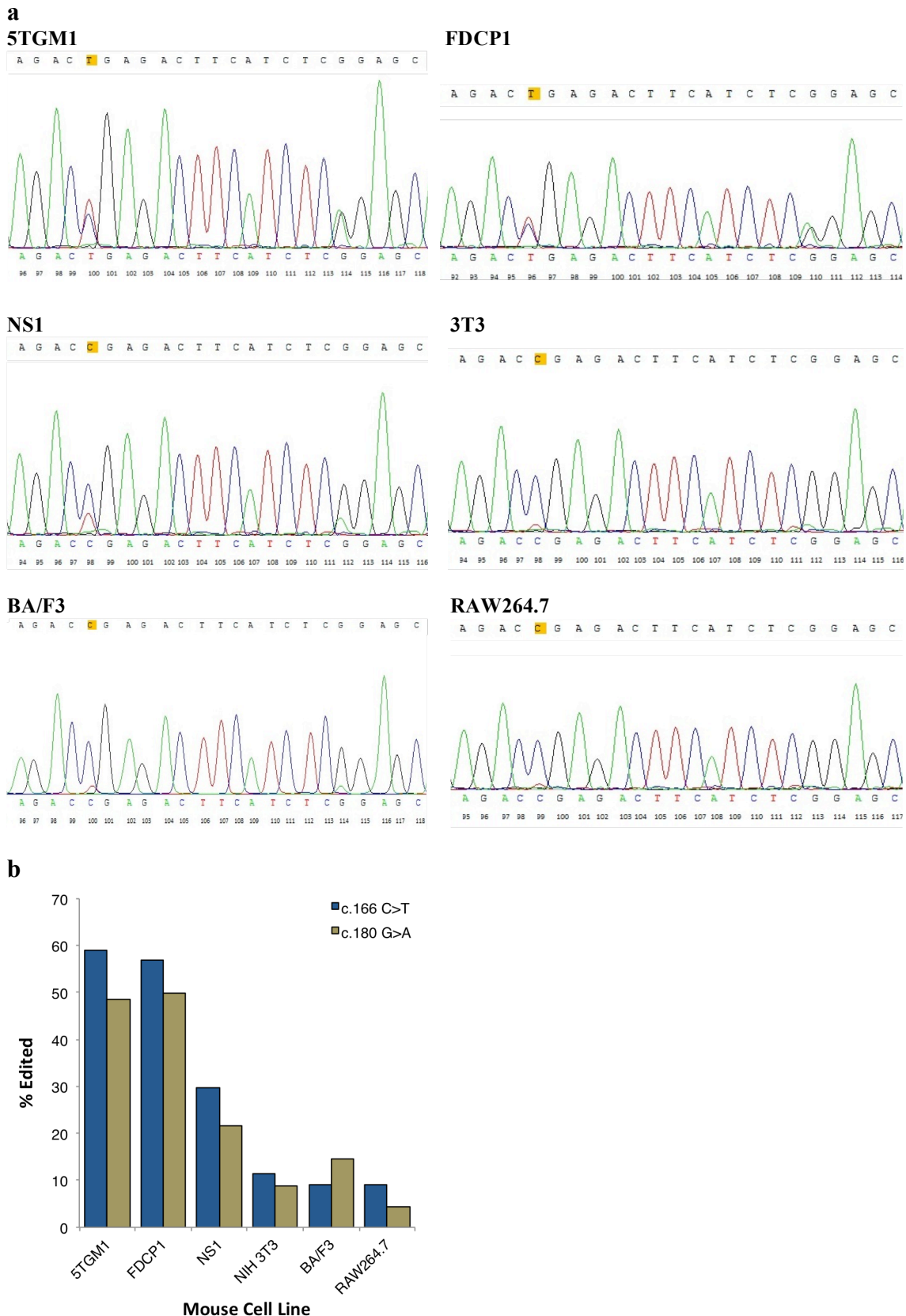


Figure 5. RNA editing of *Sp140* occurs in multiple mouse cell lines (a) Sanger sequencing of the exon 2 region of interest revealed *Sp140* was edited at varying levels in other non-PC cell lines. Position c.166 is highlighted in yellow **(b)** Calculation of the percentage of altered nucleotide at RNA editing sites c.166 and c.180. Sequencing counts revealed that FDCP1 and NS1 cells also exhibited prominent c.166C>T (i.e. U) change.

4.4.3 *SP140* is not a target of RNA editing in human MM PCs

Prior to investigating the occurrence of *SP140* RNA editing in human MM cell lines, the conservation of *Sp140/SP140* gene sequence between mouse and human was examined. Comparison of the *Sp140/SP140* exon 2 region for sequence homology between mouse and human demonstrated that site c.166 from mouse was cross-species conserved, whereas, c.180 was not [Figure 6]. Gene sequences were derived from Ensembl genome browser (<https://asia.ensembl.org>), which also revealed that site c.166 is a catalogued deleterious missense C/T variant in humans (dbSNP rs755024352). As RNA editing at c.166 was identified to lead to a high impact STOP gain change in 5TGM1 PCs, we similarly screened a range of human MM cell lines for this alteration.

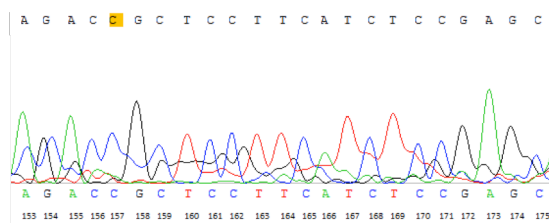


Figure 6. Comparative analysis of mouse *Sp140* and human *SP140* gene sequences. Comparison of the *Sp140/SP140* exon 2 region of interest between mouse and human genome revealed position c.166 to be conserved, while c.180 was not.

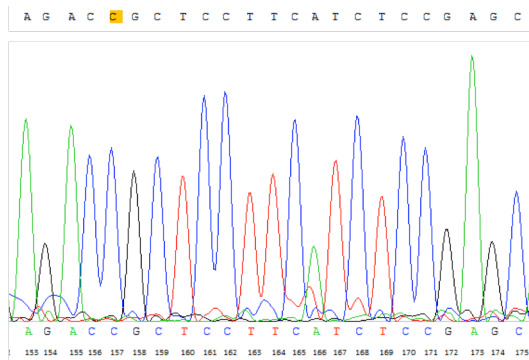
Utilising the human MM PC line RNA sequencing data generated by Dr Jonathan J. Keats laboratory (The Translational Genomics Research Institute, Phoenix, Arizona; <http://www.keatslab.org/data-repository>), the transcript counts of *SP140* was assessed in 14 human MM plasma cell lines available in our laboratory including: JIM-1, KMS-11, KMS-18, LP-1, MM.1R, MM.1S, NCI-H929, OPM2, RPMI 8226, U266, JJN3, KMM-1, MOLP-8 and EJM. We found *SP140* to be expressed at variable levels across all the MM cell lines [Supplementary Table 4], with 9 PC lines exhibiting moderate or high *SP140* expression. These were chosen for screening on the basis of the abundance of transcript available for screening, and protein that would be affected by RNA editing changes. These MM PC lines were curated into 2 groups, namely the “high expressing” group which included KMS-18, MM.1R, MM.1S, U266 and JJN3; and the “moderate expressing” group which included RPMI-8226, NCI-H929, MOLP-8 and EJM.

For each human MM cell line, RNA was extracted for cDNA synthesis followed by PCR amplification of *SPI40* exon 2. PCR products were purified and sequenced to screen for the identified variant affecting *SPI40*. Sanger sequencing trace analysis revealed that the c.166 site was not edited in any of the human MM PC lines examined, with prominent C nucleotide calls in both “high expressing” and “moderate expressing” *SPI40* PC lines [Figure 7a, 7b]. Taken together, our Sanger sequencing data revealed that RNA editing is unlikely to be a common event in exon 2 of *SPI40* in human MM PCs.

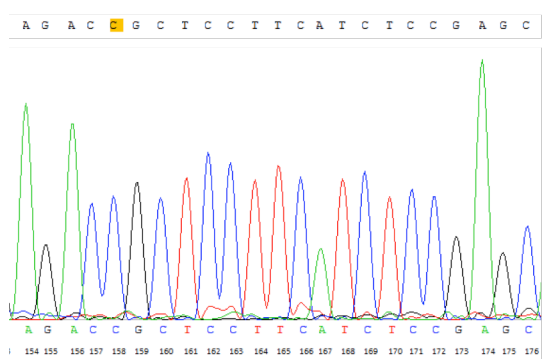
a
KMS-18



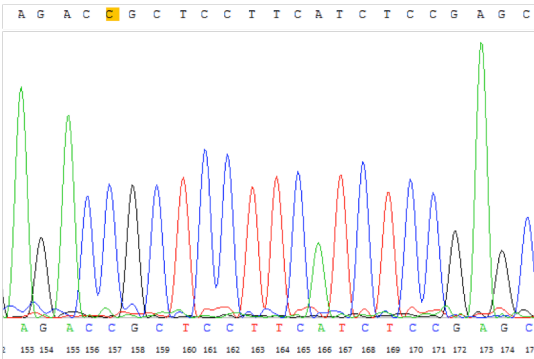
MM.1R



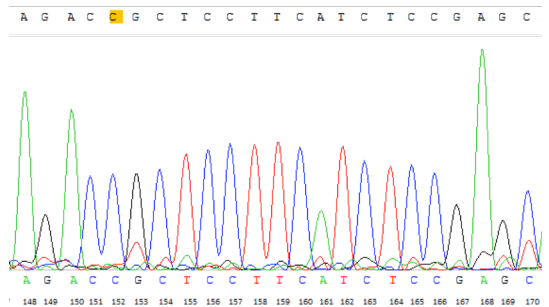
MM.1S



U266



JJN3



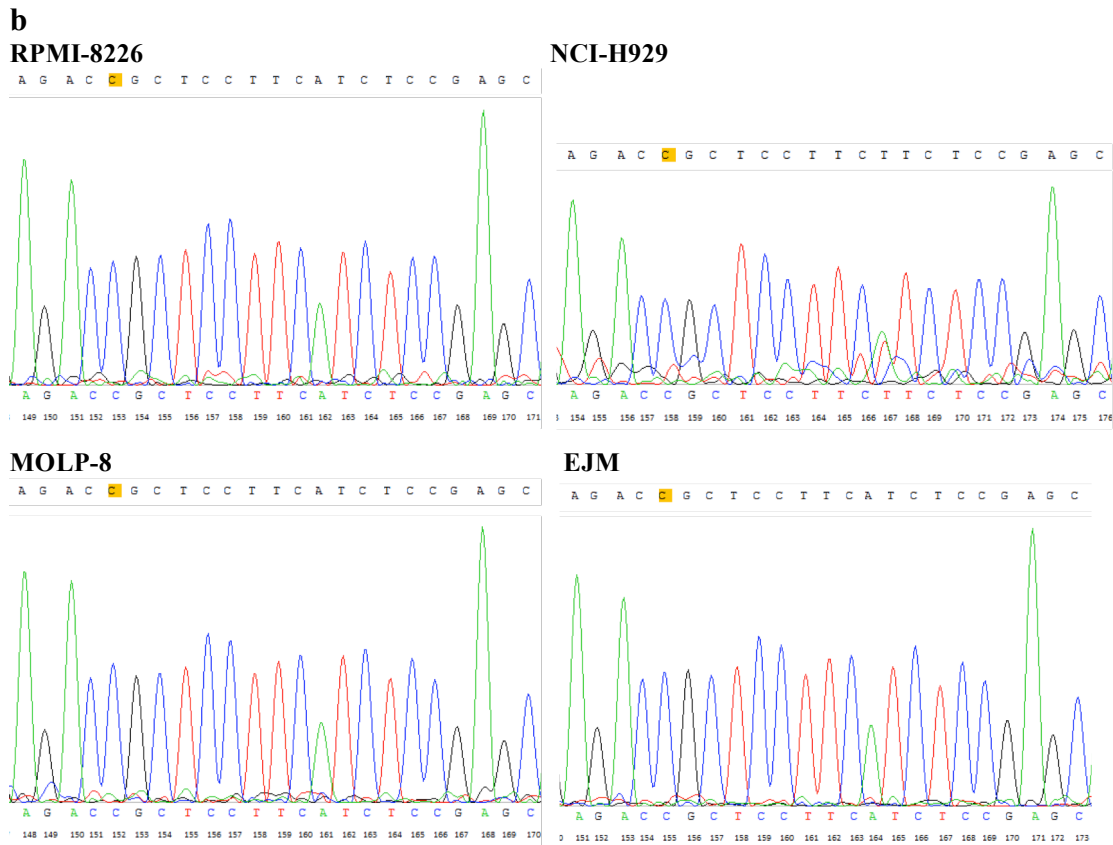


Figure 7. RNA/cDNA sequence screening of *SP140* in human MM cell lines. Sanger sequencing of exon 2 region revealed that *SP140* is not a target of RNA editing, with traces supporting the reference C nucleotide in both *SP140* (a) “high expressing”: KMS-18, MM.1R, MM.1S, U266, JJN3; and (b) “moderate expressing”: RPMI-8226, NCI-H929, MOLP-8 and EJM, human MM cell lines. Position c.322 is highlighted in yellow.

4.4.4 Molecular mechanisms inducing RNA editing of *SP140*

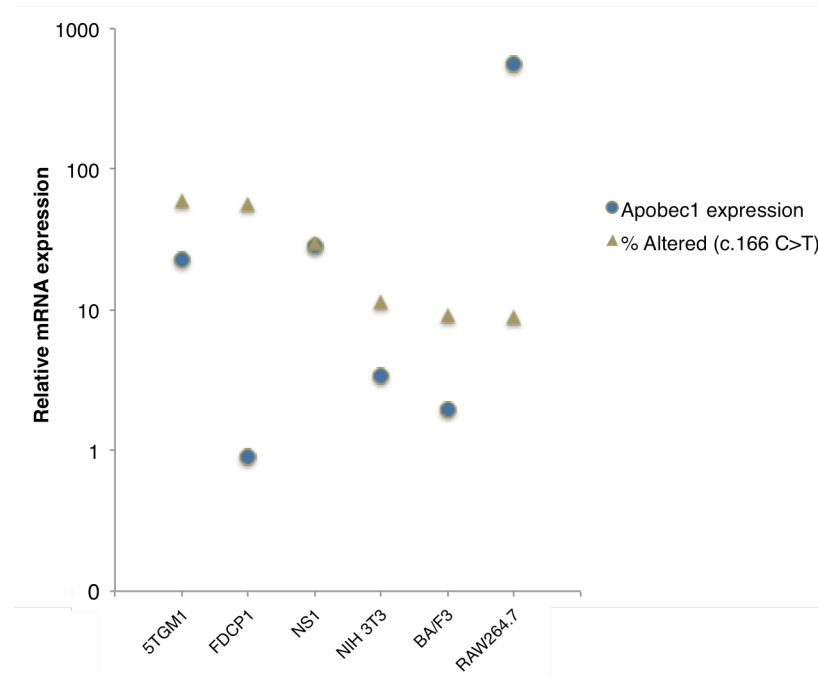
While RNA editing of *SP140* was not a feature in any human MM PC lines, identification of *Sp140* as a target of editing in the mouse 5TGM1 cell line suggests a possible role of RNA editing as a mechanism of PC transformation in this preclinical model of MM. As we identified a C>T (i.e. U) change at c.166, we postulated that this was due to the action of a member of the Apobec family of enzymes. APOBECs are a class of cytidine deaminases known to catalyse C>U RNA editing changes, where currently APOBEC1, APOBEC3A and APOBEC3G are known to cause recoding changes²⁸.

RNAseq data showed that only Apobec1 and Apobec3 were expressed in the 5TGM1 PC line. RT-qPCR analysis revealed that both Apobec1 and Apobec3 were expressed in all of the mouse cell lines that had previously been screened for *Sp140* RNA editing levels. Notably, there was no correlation between the levels of gene expression of either Apobec and the frequency of RNA editing in these cell lines ($R^2 = 0.16$ for Apobec1; and $R^2 = 0.11$ for Apobec3) [Figure 8a, 8b].

The Apobec family members contain a domain structure characteristic of cytidine deaminases³⁴. Mouse Apobec1 contains one putative cytidine deamination domain (CDD) with a conserved active site (AS) that includes a H-X-E motif followed 28 amino acids later by a P-C-X₂₋₄-C motif. The His(H)-Cys(C)-Cys(C) residues of these motifs coordinate zinc binding and the Glu(E) residue acts as a proton shuttle during catalysis³⁵. Point mutation of any of the conserved H, E, C and C amino acids of Apobec1 has been shown to abolish *in vitro* RNA editing activity^{36,37} and greatly reduced cytidine deaminase activity³⁷. C-terminal truncated rabbit Apobec1 (66 amino acids long) also has no detectable RNA editing activity³⁶. Smaller C-terminal truncated Apobec1 (181 amino acids long) also displayed no *in vitro* RNA editing activity, and point mutations in key lysine residues in this region (L185, L189) also greatly reduces editing activity³⁸. Although no RNA cytidine deamination activity has so far been ascribed to mouse Apobec3, critical residues/motifs important for its ssDNA deamination activity have been characterised. Mouse Apobec3 contains two putative CDDs each with a conserved AS that includes the H-X-E motif followed 23-28 amino acids later by a P-C-X₂₋₄-C motif. According to Nair *et al.*, the N-terminal domain appears to be the sole locus of deamination activity in mouse Apobec3³⁹. Replacing any of the conserved E, C and C amino acids from the N-terminal domain completely abolished any *in vitro* deamination activity³⁹, an effect that was not observed following mutation of any of these amino acids in the C-terminal domain. By

contrast, the C-terminus of Apobec3 appeared to be necessary for the encapsidation into retrovirus particles³⁹. Hakata and Landau had previously shown that mutation of the conserved Glu(E) of AS1, but not that in AS2, was sufficient to abolish *in vitro* deaminase activity⁴⁰. Somewhat contradictory to the results of Nair *et al.*, Hakata and Landau stated that mouse Apobec3 with mutations of both of the cysteines at either AS1 or AS2 also lacked deaminase activity, but no data was shown⁴⁰. Similarly they stated that single mutation of either of the two cysteines of AS2 prevented deamination activity despite the presence of an intact AS1 catalytic domain, however, again no data was shown⁴⁰. Together, these data suggest that mutation of the conserved P-C-X₂₋₄-C motif within each AS of either Apobec1 or Apobec3 should be sufficient to inactivate each putative cytidine deamination domain.

(a)



(b)

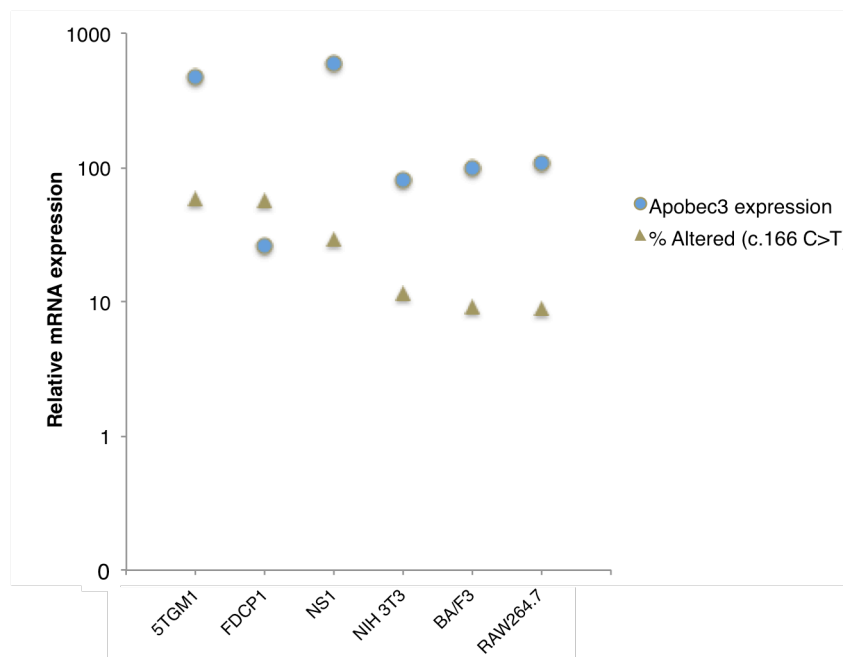
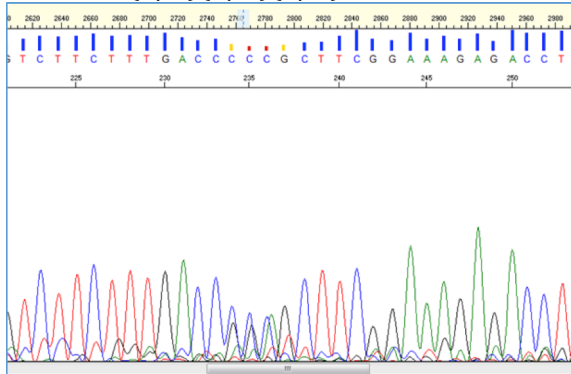


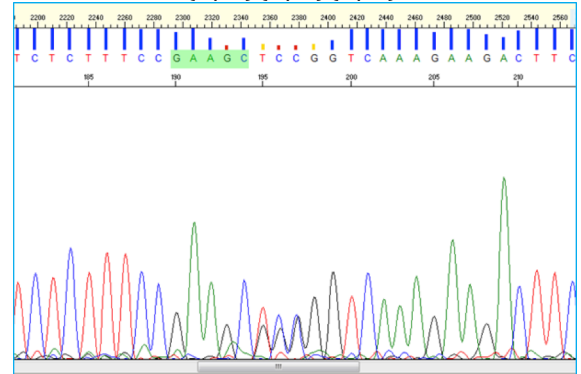
Figure 8. Correlation of Apobec enzyme expression with RNA editing of *Sp140* exon 2 in mouse cell lines. RT-qPCR revealed varying levels of (a) Apobec1 and (b) Apobec3 expression in multiple mouse cell lines: 5TGM1, FDCP1, NS1, HIH-3T3, BA/F3 and RAW264.7. There was no correlation between the levels of expression of either Apobec with the percentage of altered nucleotide at RNA editing site c.166. Correlation coefficients $R^2 = 0.16$ and $R^2 = 0.11$, were calculated for Apobec 1 and Apobec3, respectively, illustrating a weak linear relationship.

Plasmid-based CRISPR-Cas9 gene editing was used to mutate Apobec1 in 5TGM1 cells³⁰. Two different single guide RNAs (sgRNAs) were used to target exon 6 (the largest coding exon) of Apobec1 with the intention of generating insertion/deletion mutations that would generate inactivating frameshift mutations. A total of 36 clonal 5TGM1 cell lines from each Apobec1-sgRNA transfection were screened for mutations. Only one Apobec1-sgRNA2 5TGM1 clonal cell line was found to harbour two frameshift inducing mutations [Figure 9] [Supplementary Figure 5]. Both of the mutated alleles in Apobec1-sgRNA2-clone#1 (Apobec1 KO cell line) were predicted to encode for severely C-terminal truncated proteins, lacking any of the conserved H-X-E and P-C-X₂₋₄-C motifs that are required for cytidine deamination activity [Figure 10a, 10b] [Supplementary Figure 6]. RT-qPCR also showed a ~56% reduction of Apobec1 mRNA levels in the Apobec1 KO cell line compared to an empty vector clonal 5TGM1 cell line (EV11) [Figure 10c]. To assess the effect of Apobec1 mutation on resulting RNA editing of *Sp140*, cDNA was PCR amplified before purification and Sanger sequencing. Analysis of sequencing traces illustrated that the EV sample showed c.166C>U editing activity [Figure 11a]. The Apobec1 KO cell line also demonstrated unchanged levels of RNA recoding activity, suggesting that Apobec1 is not the enzyme responsible for the RNA editing of *Sp140* in mouse [Figure 11b].

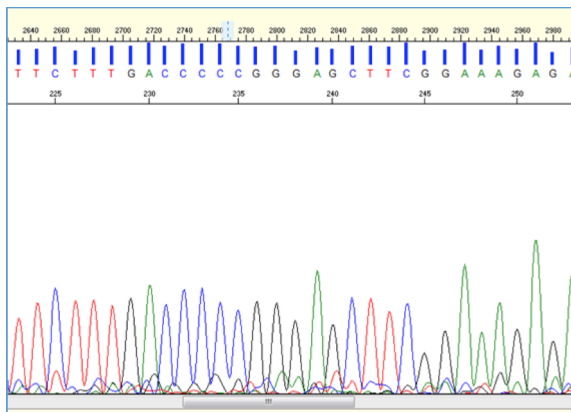
a
Apobec1-sgRNA2-clone#1
 TTTGACC(C/G)(C/G)(C/A)GCTTCGGAAA



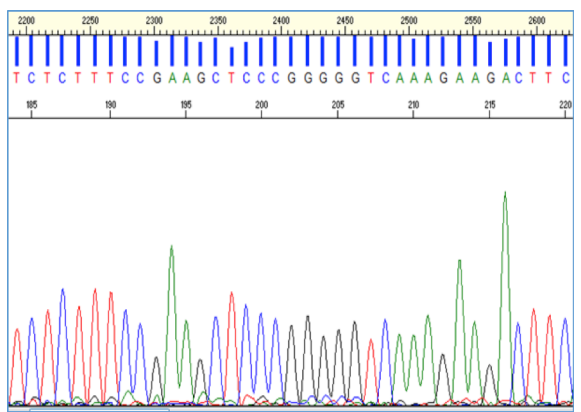
b
 TTTCCGAAGC(G/T)(G/C)(G/C)GGTCAAA



c
Empty vector (EV11)



d



e
Mutation deconvolution

WT sequence: CCCCCGGGAGCTTC

MUT sequence: CCSSM----GCTTC

Where S = G/C and M = C/A

This sequencing result can only be explained by the following two 4 bp deletions:

| | |
|--------------|----------------|
| wt allele | CCCCCGGGAGCTTC |
| mut allele#1 | CC----GGAGCTTC |
| mut allele#2 | CCCC----GCTTC |

Figure 9. Generation of Apobec1 frameshift mutant clone in 5TGM1. CRISPR-Cas9 mediated gene editing of Apobec1 exon 6 reveals two frameshift mutations in Apobec1-sgRNA2-clone#1 **(a)** forward and **(b)** reverse compared to empty vector (EV11) **(c)** forward and **(d)** reverse sequencing traces. **(e)** Mutation deconvolution illustrates the resulting 4 base pair frameshift deletions in the Apobec1-sgRNA2 5TGM1 clonal PC line. Full nucleotide sequences can be found in *Supplementary Figure 5*.

In addition to this, a transgene consisting of *Sp140* exon 2 and its surrounding exon 1 and 3 elements (termed the *Sp140* Minigene) was constructed in the pLeGO Cer2 vector, to assess whether exogenous *Sp140* RNA could be a template for RNA editing in 5TGM1 cells. Sequencing analysis revealed no evidence of editing, with traces supporting reference C and G nucleotides at position c.166 and c.180, respectively. This demonstrates that further elements of the *Sp140* gene beyond the immediate exon 2 region, such as the intronic sequences are required by the RNA editing enzyme responsible for the C>T editing phenomenon we observed [*Supplementary Figure 4*].

a

Predicted Translation of Mutant Allele 1 (65 aa):

MSSETGPVAVDPTLRRRIEPHEFEVFFD**RS**FGKRPVCCMRSTGVEGTVSGDTRAKTPATTLKSTS

b

Predicted Translation of Mutant Allele 2 (65 aa):

MSSETGPVAVDPTLRRRIEPHEFEVFFD**PR**FGKRPVCCMRSTGVEGTVSGDTRAKTPATTLKSTS

c

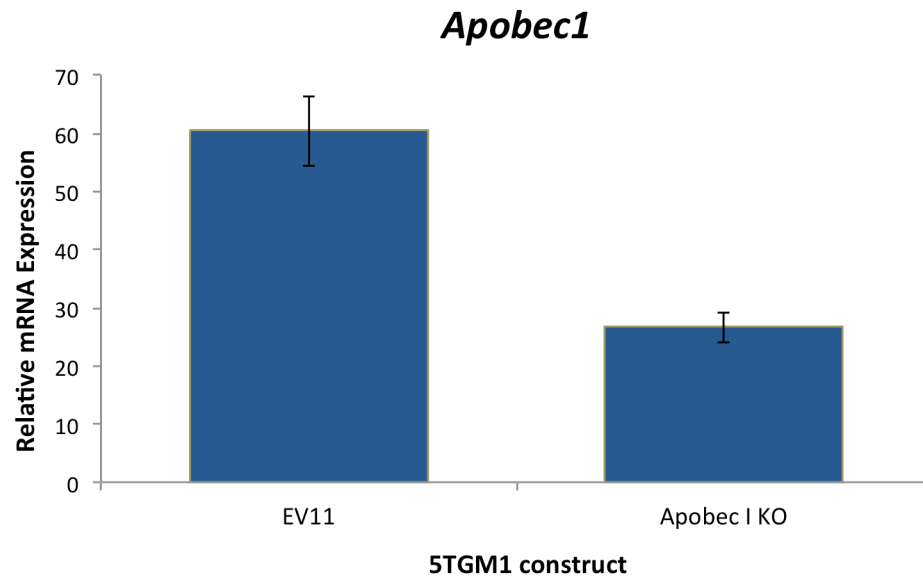


Figure 10. Predicted effects of frameshift mutations on Apobec1 protein formation, and its quantitation in knockout construct Apobec1-sgRNA2-clone#1.

CRISPR-Cas9 mediated (a) mutant allele 1 and (b) mutant allele 2 in the Apobec1 KO cell line are predicted to result in severely truncated Apobec1 proteins. While translation of unmutated *Apobec1* results in a 229 amino acid protein, both mutant alleles are predicted to encode 65 amino acid proteins with significant truncations of the C-terminus, lacking motifs that are required for its deaminase activity. Green highlighted text illustrates the wildtype protein sequence, while yellow highlighted text illustrates the mutated C-terminal protein sequence. Cyan highlighted text shows the altered protein sequence exclusive to Mutant Allele 1. Full protein sequence can be found in *Supplementary Figure 6* (c) Reduction of Apobec1 mRNA levels in Apobec1 KO cells. *Apobec1* expression was measured in cell lines using quantitative real-time PCR, confirming a ~56% knockdown in the Apobec1 KO construct (Apobec1-sgRNA2-clone#1) compared to the empty vector control (EV11). Data were normalised to *ActB* (mean \pm standard deviation of triplicates).

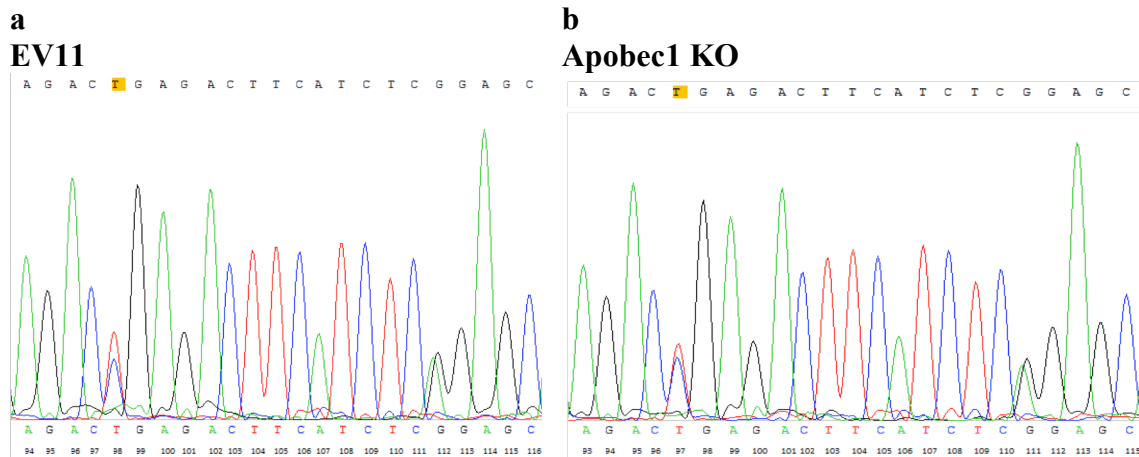


Figure 11. Apobec1 is not responsible for catalysing RNA editing changes in *Sp140*. Sanger sequencing of *Sp140* region of interest reveals RNA editing in exon 2 is not due to the cytidine deaminase action of Apobec1. **(a)** Sequencing traces of empty vector (EV11) cells show both C and T nucleotides at c.166. **(b)** Sequencing traces from Apobec1 KO cells, also showed c.166 C>U RNA editing. Position c.166 is highlighted in yellow.

Given the low targeting efficiency of this plasmid-based approach, an alternative lentiviral transduction methodology (with inducible sgRNA expression)³² was used to mutate Apobec3. Two different sgRNAs were designed to target exonic sequences just upstream of regions encoding the conserved P-C-X₂₋₄-C motifs of each of the two active sites (AS) of mouse Apobec3 (one sense orientation sgRNA and one antisense orientation sgRNA at each AS). A total of 10 clonal cell lines were generated from each of the four Apobec3 sgRNAs and screened for putative inactivating insertion/deletion mutations. 13 different clonal 5TGM1 cell lines were found to harbour homozygous or compound heterozygous frameshift mutations in exon 3 of Apobec3 [Figure 12a, 12b] [Supplementary Figure 7, 8]. All exon 3 mutations were predicted to encode for severely C-terminal truncated proteins that lack the second active site (AS2) and lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS1 [Supplementary Figure 11, 12]. Similarly, homozygous or compound heterozygous frameshift mutations in exon 7 of Apobec3 were observed in a separate 13 independent clonal 5TGM1 cell lines [Figure 12c, 12d] [Supplementary Figure 9, 10]. These exon 7 mutations were all predicted to encode C-terminally truncated Apobec3 proteins that lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS2 [Supplementary Figure 13, 14]. Subsequently, we assessed whether the loss of Apobec3 function resulted in changes to the RNA editing of *Sp140*. cDNA of each clone was PCR amplified before purification and Sanger sequencing.

Analysis of sequencing traces illustrated that both Apobec3 AS1 and AS2 mutant clones showed unchanged levels of RNA editing activity at c.166C>U [Figure 13]. This suggests that neither Apobec1, nor Apobec3, are the enzymes responsible for the RNA editing of *Sp140* observed in mouse. As such, the RNA editing of *Sp140* is likely to be caused by another, as yet unknown, enzyme.

a

Apobec3 AS1 sense sgRNA clones with two mutant (frameshift) alleles

| | | |
|------------------|--|---------------|
| | AGATCACCTGGTATATGTCCTGGAGCCCCTGTTTCGAATGTGCAG | unmutated |
| Clone#1 allele 1 | AGATCACCTGGTATATG--CTGGAGCCCCTGTTTCGAATGTGCAG | -2bp |
| Clone#1 allele 2 | AGATCACCTGGT-----TCCTGGAGCCCCTGTTTCGAATGTGCAG | -5bp |
| Clone#2 allele 1 | AGATCACCTGGTATATG-----CCCCTGTTTCGAATGTGCAG | -8bp |
| Clone#2 allele 2 | AGATCACCTGGT-----TCCTGGAG-----TTTCGAATGTGCAG | -11bp |
| Clone#3 allele 1 | AGATCACCTGGTATATG--CTGGAGCCCCTGTTTCGAATGTGCAG | -2bp |
| Clone#3 allele 2 | AGATCACCTGGT-----TCCTGGAGCCCCTGTTTCGAATGTGCAG | -5bp |
| Clone#4 allele 1 | AGATCACCTGGTATATG-- G TGGAGCCCCTGTTTCGAATGTGCAG | -2bp and mut. |
| Clone#4 allele 2 | AGATCACCTGGT-----TCCT G CAGCCCCTGTTTCGAATGTGCAG | -5bp |
| Clone#5 allele 1 | AGATCACCTGGTATAT--CTGGAGCCCCTGTTTCGAATGTGCAG | -2bp |
| Clone#5 allele 2 | AGATCACCTGGTATATG-- G TGGAGCCCCTGTTTCGAATGTGCAG | -1bp and mut. |
| Clone#6 allele 1 | AGATCACCTGGTATAT-----GGAGCCCCTGTTTCGAATGTGCAG | -5bp |
| Clone#6 allele 2 | AGATCACCTGGT-----TCCT GCT GGAGCCCCTGTTTCGAATGTG | -5bp and +3bp |
| Clone#7 allele 1 | AGATCACCTGGTATATG--CTGGAGCCCCTGTTTCGAATGTGCAG | -2bp |
| Clone#7 allele 2 | AGATCACCTGGT-----TCCTGGAGCCCCTGTTTCGAATGTGCAG | -5bp |
| Clone#9 allele 1 | AGATCACCTGGTATATG--CTGGAGCCCCTGTTTCGAATGTGCAG | -2bp |
| Clone#9 allele 2 | AGATCACCTGGT-----TCCTGGAGCCCCTGTTTCGAATGTGCAG | -5bp |

b

Apobec3 AS1 antisense sgRNA clones with two mutant (frameshift) alleles

| | | |
|------------------|---|---------------|
| | TGGTATATGTCCTGGAGCCCCTGTTTCGAATGTGCAGAGCAGATA | unmutated |
| Clone#1 homo. | TGGTATATGTCCTGGAG G CCCCTGTTTCGAATGTGCAGAGCAGATA | +1bp |
| Clone#2 allele 1 | TGGTATATGTCCTGGA--CCCCTGTTTCGAATGTGCAGAGCAGATA | -1bp |
| Clone#2 allele 2 | TGGTATATGTCCTGG--G--CCCTGTTTCGAATGTGCAGAGCAGATA | -2bp |
| Clone#3 homo. | TGGTAT AC -----GCCCTGTTTCGAATGTGCAGAGCAGATA | -8bp and mut. |
| Clone#6 allele 1 | TGGTATATGTCCTGG--GCCCTGTTTCGAATGTGCAGAGCAGATA | -1bp |
| Clone#6 allele 2 | TGGTATATGTCCTGGAG--CCCTGTTTCGAATGTGCAGAGCAGATA | -1bp |
| Clone#9 allele 1 | TGGTATATGTCCTGG--GCCCTGTTTCGAATGTGCAGAGCAGATA | -1bp |
| Clone#9 allele 2 | TGGTATATGTCCTGG--CCCCTGTTTCGAATGTGCAGAGCAGATA | -2bp |

c**Apobec3 AS2 sense sgRNA clone with two mutant (frameshift) alleles**

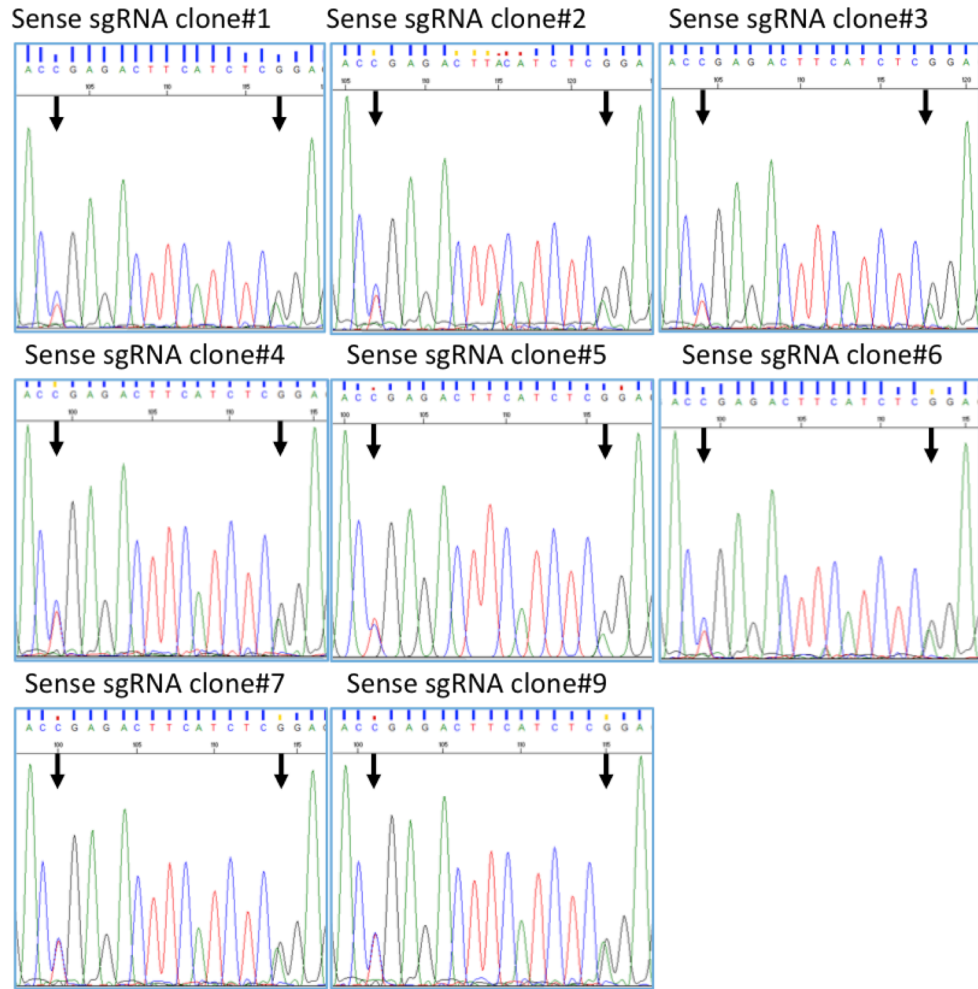
| | | |
|-------------------|--|---------------|
| | CAATCACCTGCTACCTCACCTGGAGCCCTGCCCAAACCTGTGCCTGG | unmutated |
| Clone#1 allele 1 | CAATCACCTGCTACC--ACCTGGAGCCCTGCCCAAACCTGTGCCTGG | -2bp |
| Clone#1 allele 2 | CAATCACCTGCTACCTC-----GAGCCCTGCCCAAACCTGTGCCTGG | -5bp |
| Clone#2 allele 1 | CAATCACCTGCTACC--ACCTGGAGCCCTGCCCAAACCTGTGCCTGG | -2bp |
| Clone#2 allele 2 | CAATCACCTGCTACCTC---- TA AGCCCTGCCCAAACCTGTGCCTGG | -4bp and mut. |
| Clone#4 allele 1 | CAATCACCTGCTACC-----TGGAGCCCTGCCCAAACCTGTGCCTGG | -5bp |
| Clone#4 allele 2 | CAATCACCTGCTACCTC G -CTGGAGCCCTGCCCAAACCTGTGCCTGG | -1bp and mut. |
| Clone#5 homo. | CAATCACCTGCTACCT-ACCTGGAGCCCTGCCCAAACCTGTGCCTGG | -1bp |
| Clone#7 allele 1 | CAATCACCTGCTACC-----TGGAGCCCTGCCCAAACCTGTGCCTGG | -5bp |
| Clone#7 allele 2 | CAATCACCTGCTACCTC- G CTGGAGCCCTGCCCAAACCTGTGCCTGG | -1bp and mut. |
| Clone#8 allele 1 | CAATCACCTGCTACCTC-----GAGCCCTGCCCAAACCTGTGCCTGG | -5bp |
| Clone#8 allele 2 | CAATCACCTGCTACC--ACCTGGAGCCCTGCCCAAACCTGTGCCTGG | -2bp |
| Clone#9 homo. | CAATCACCTGCTACCT-ACCTGGAGCCCTGCCCAAACCTGTGCCTGG | -1bp |
| Clone#10 allele 1 | CAATCACCTGCTACCTC- G CTGGAGCCCTGCCCAAACCTGTGCCTGG | -1bp and mut. |
| Clone#10 allele 2 | CAATCACCTGCTACCT-----GGAGCCCTGCCCAAACCTGTGCCTGG | -5bp |

d**Apobec3 AS2 antisense sgRNA clone with two mutant (frameshift) alleles**

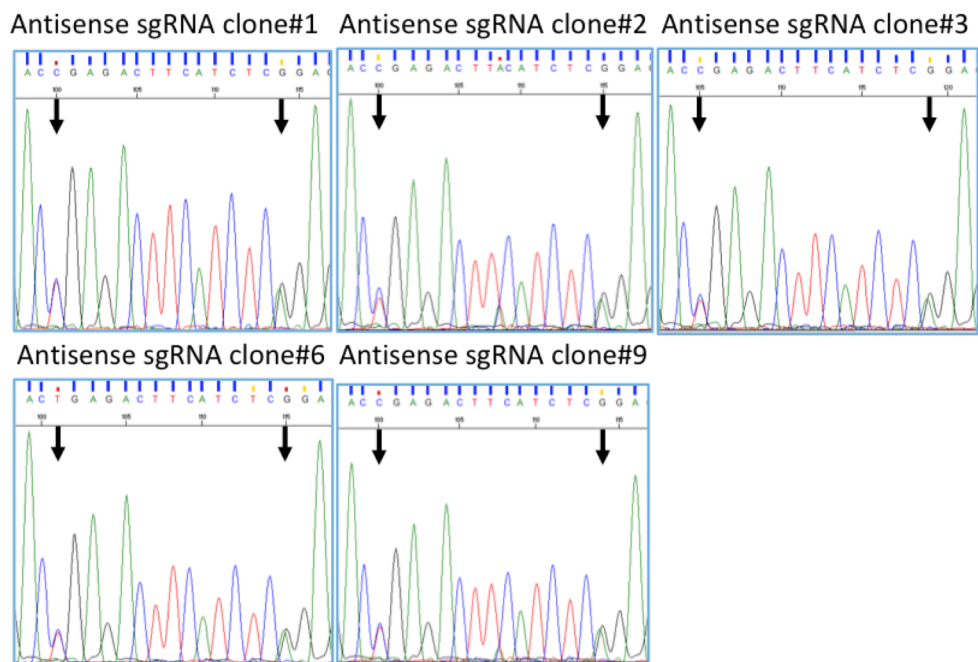
| | | |
|------------------|---|-----------|
| | CAATCACCTGCTACCTCACCTGGAGCCCTGCCCAAACCTGTGCCTG | unmutated |
| Clone#1 allele 1 | CAATCACCTGCTACCTCACCTGG-GCCCTGCCCAAACCTGTGCCTG | -1bp |
| Clone#1 allele 2 | CAATCACCTGCTACCTCACCTGG A TGCCCTGCCCAAACCTGTGCCT | +1bp |
| Clone#5 allele 1 | CAATCACCTGCTACCTCACCTGG-GCCCTGCCCAAACCTGTGCCTG | -1bp |
| Clone#5 allele 2 | CAATCACCTGCTACCTCACCTGG A TGCCCTGCCCAAACCTGTGCCT | +1bp |
| Clone#6 allele 1 | CAATCACCTGCTACCTCACCTGG A AGCCCTGCCCAAACCTGTGCCT | +1bp |
| Clone#6 allele 2 | AAATCCTCTTCCTTG (A) ----- del 85 -----AACTGTGCCT | -85bp |
| Clone#7 allele 1 | CAATCACCTGCTACCTCACCTGG-----TGCCCAAACCTGTGCCTG | -5bp |
| Clone#7 allele 2 | CAATCACCTGCTACCTCACCTGG G GCCCTGCCCAAACCTGTGCCT | +1bp |
| Clone#9 allele 1 | CAATCACCTGCTACCTCACCTGG-CCCCTGCCCAAACCTGTGCCTG | -1bp |
| Clone#9 allele 2 | CAATCACCTGCTACCTCACCTGG--CCCTGCCCAAACCTGTGCCTG | -2bp |

Figure 12. Generation of Apobec3 frameshift mutant 5TGM1 clones. An inducible CRISPR-Cas9 gene editing system was used to mutate two different active sites (AS) of Apobec3 (one sense orientation sgRNA and one antisense orientation sgRNA at each AS). 13 different clonal 5TGM1 cell lines were found to harbour homozygous or compound heterozygous frameshift mutations in exon 3 of Apobec3 using either (a) sense or (b) antisense sgRNAs. Sequencing traces illustrating mutations can be found in *Supplementary Figure 7, 8*. Similarly, homozygous or compound heterozygous frameshift mutations in exon 7 of Apobec3 were observed in a separate 13 independent clonal 5TGM1 cell lines using either (c) sense or (d) antisense sgRNAs. Sequencing traces illustrating mutations can be found in *Supplementary Figure 9, 10*.

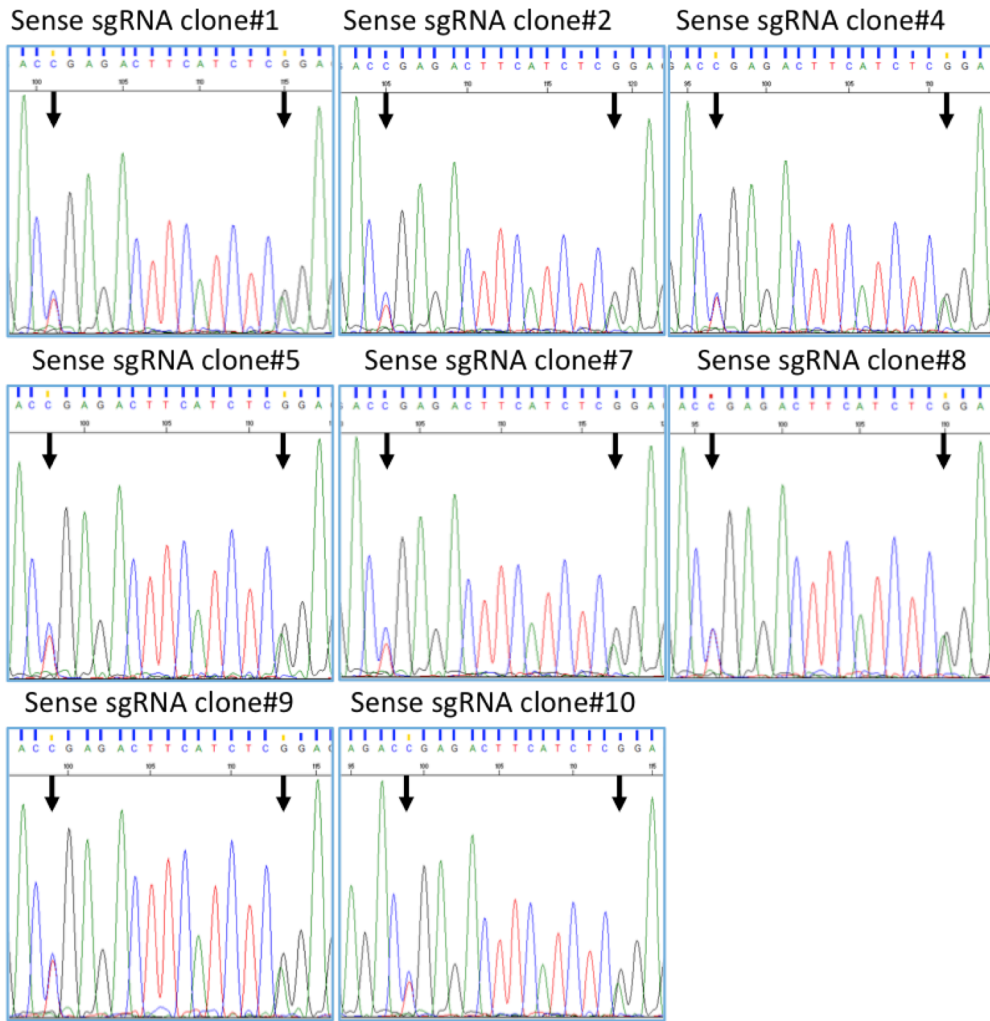
a



b



c



d

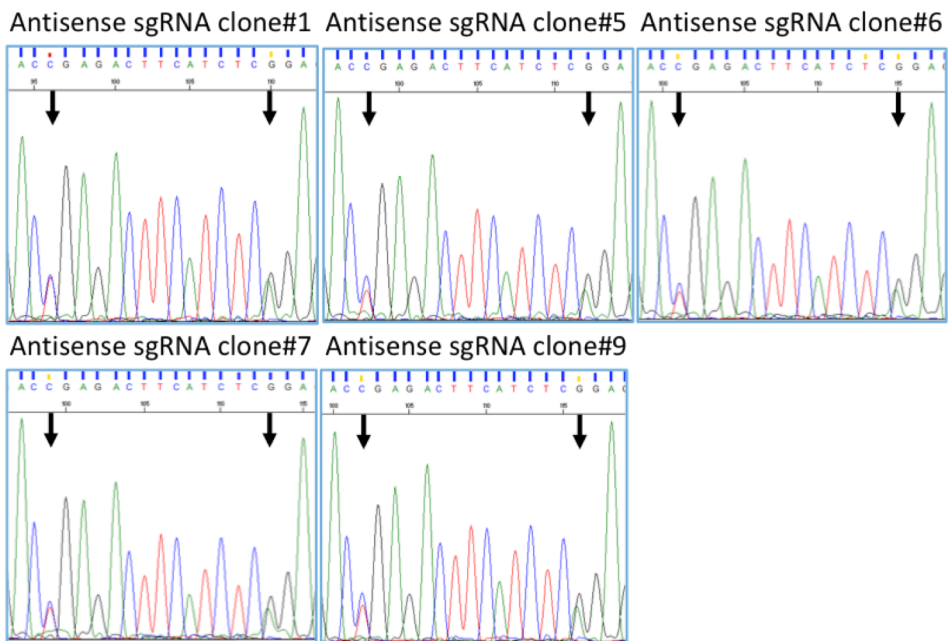


Figure 13. Apobec3 is not responsible for catalysing RNA editing changes in *Sp140*. Sanger sequencing of the *Sp140* region of interest in mutant Apobec3 clones revealed unchanged RNA editing of exon 2, suggesting that this is not due to the cytidine deaminase action of Apobec3. Sequencing traces of *Sp140* in AS1 **(a)** sense sgRNA and **(b)** antisense sgRNA-induced double frameshift clones shows RNA editing with both C and T nucleotides at c.166. Similarly, sequencing traces of *Sp140* in AS2 **(a)** sense sgRNA and **(b)** antisense sgRNA-induced double frameshift clones also demonstrate c.166 C>U RNA editing. Sites of interest c.166 and c.180 are indicated by arrows.

4.5 Discussion

Cancer, in its many forms, is due to the deregulation of biological processes following mutations that affect the normal program of DNA transcription to mRNA, and mRNA translation to protein. RNA editing is an unconventional event that adds to the diversity of expressed transcripts and protein repertoire of tumour cells. Recoding changes have been implicated in many malignancies such as hepatocellular carcinoma²³, glioblastoma²⁴, prostate cancer²⁵, colorectal cancer²⁶ and chronic myeloid leukaemia²⁰. In a MM context, there has been one study that has investigated gene-specific RNA editing, finding that ADAR1 recoding of *GLII* promotes malignant regeneration²². Here, we explored the novel identification of RNA editing in mouse *Sp140*, inducing STOP gain (c.166) and synonymous change (c.180) edits, in the 5TGM1 MM PC line.

Notably, recent NGS analyses of MM patient samples have shown that *SP140* is recurrently mutated at the genetic level in ~3-12% of patients, with a suggested role as a tumour suppressor^{2,3,12}. In chapter 2, our genetic analysis of paired MGUS/SMM to MM using whole exome sequencing found that *SP140* was mutated in 30% of our patients. Therefore, alteration at position c.166 was notable as it induces a nonsense STOP gain codon change in the transcript. Notably, RNA editing in coding regions is a rare occurrence, with editing normally occurring within introns or 3' UTRs²⁸. An early truncation of *Sp140* would result in inactivation and possible tumorigenic consequences. Although this is not a classical complete loss of function of a tumour suppressor gene, it has previously been shown that in a haploinsufficient phenotype (where one mutant and one wild type allele remains), a suboptimal level of gene product can result in lower level of function and oncogenesis^{41,42}.

It is curious that edited *Sp140* transcripts are not subject to nonsense mediated mRNA decay in our samples⁴³. While we initially found RNA editing changes in 5TGM1 cells, they were not PC-specific, with varying levels of *Sp140* editing also identified in other mouse cell lines. Studies show that RNA editing also occurs to some degree in normal cells, and thus recoding is not a cancer specific event, rather cancer associated editing may be more dependent on quantity of edited transcript, rather than site specific changes²⁸. In addition, we also screened a range of human MM PC lines finding *SP140* exon 2 site c.166 to be conserved, however, it was not subject to any RNA editing.

To investigate the underlying mechanism responsible for the RNA editing phenotype in 5TGM1 cells, we explored known RNA editing enzyme Apobec1 that catalyses C>U changes. Verified C>U RNA changes catalysed by APOBECs are far less common than A>I changes by ADARs²⁸. APOBEC family enzymes are principally known for their DNA editing activity, with APOBEC1 originally the only known RNA editor. However, knockout of *Apobec1* did not result in a decrease in RNA editing phenotype at c.166 in 5TGM1 cells. While Apobec1 was not found to be responsible for the observed RNA editing, recent studies have demonstrated that APOBEC3A and APOBEC3G can also induce C>U changes under physiological conditions^{28,44,45}. APOBEC3A shows highly site-specific activity, targeting coding region of genes and inducing missense or nonsense codon changes²⁸. Notable, while there are 7 paralogous *APOBEC3* genes (A3A-H) in humans, the mouse genome encodes one *Apobec3* gene^{34,46,47}. Generation of *Apobec3* double frameshift mutant clones did not result in a change of *Sp140* RNA editing phenotype at c.166 in 5TGM1 cells. This indicates there may be other unknown enzymes capable of catalysing site specific C>U recoding changes. Interestingly, a recent study of APOBEC3A in Wilms Tumour (WT1) has identified tandem *cis* RNA editing changes with C>T (i.e. U) and G>A changes⁴⁸. Similarly, we identified paired RNA editing changes of C>T (i.e. U) and G>A within 14 nucleotides in exon 2 of *Sp140*. As G>A is a non-classical editing change, it has been proposed that this type of dual editing change may be a tandem event, where the lost amine group from the first position (i.e. c.166) through C>U deamination, is shuttled to a linked nucleotide altering a G>A (G to 2,6-diaminopurine that mimics A) at a second position (c.180)⁴⁸.

Intratumour heterogeneity is a common feature in many cancers, and has been shown to be characteristic of MM^{2,3,5,49-53}. Therefore, it remains intriguing as to why tumour cells would require further diversity through RNA editing. The answer may be the dynamic nature of post transcriptional modification, which may confer survival advantages to the tumour cell. It may be that mutations which are advantageous to a cancer cell in most microenvironments are hardwired in the DNA, but those mutations which provide a competitive advantage, either in changing environments or certain disease stages, are regulated by RNA editing²⁸.

The existence of RNA editing in MM represents a novel mechanism inducing post transcriptional changes to mRNA, conferring increased variability in proteins and cell survival. This suggests that RNAseq expression studies of patient samples would also need to consider RNA recoding. Further, they should also consider the impact of the tumour

microenvironment and unnatural selective pressures, such as drug exposure, on cancer cells and their dynamic survival adaptability. The 5TGM1 cell line is frequently used as a preclinical model for *in vitro* testing, however, as we identified no corresponding RNA editing changes in human MM cell lines and patient samples, it is worth noting that this cell line may only model those patients showing *SPI40* post transcriptional modifications. As 5TGM1 cells harboured post transcriptional changes in 50% of their transcripts, these cells may mimic a subset of MM patients that are heterozygous at the DNA levels for *SPI40* post transcriptional changes.

4.6 Supplementary

4.6.1 Supplementary Figures

| Known Drivers | Novel Candidates | | |
|---------------|------------------|----------------|---------------|
| <i>KRAS</i> | <i>SP140</i> | <i>RASA2</i> | <i>PRDM1</i> |
| <i>NRAS</i> | <i>IRF4</i> | <i>USP29</i> | <i>ACTG1</i> |
| <i>TP53</i> | <i>ROBO1</i> | <i>TRAF3</i> | <i>MAPK</i> |
| <i>BRAF</i> | <i>FAT3</i> | <i>CYLD</i> | <i>NF1</i> |
| <i>DIS3</i> | <i>EGR1</i> | <i>RB1</i> | <i>NFKBIA</i> |
| <i>FAM46C</i> | <i>PEG3</i> | <i>CCND1</i> | <i>CDKN2C</i> |
| | <i>LTB</i> | <i>PNRC1</i> | <i>PTEN</i> |
| | <i>TGDS</i> | <i>ALOX12B</i> | <i>NFKBI</i> |
| | <i>SNX7</i> | <i>HLA-A</i> | <i>CDK4</i> |
| | | | <i>MAGED1</i> |

Supplementary Table 1. Significantly mutated genes in MM patient studies. Large cohort patient sample studies (Chapman *et al.*⁵ n = 38, Lohr *et al.*³ n = 203, Bolli *et al.*² n = 67, Walker *et al.*⁴ n = 463) have identified 34 recurrent significantly mutated genes in MM, of which *KRAS*, *NRAS*, *TP53*, *BRAF*, *DIS3* and *FAM46C* are believed to be drivers due to their recurrent mutation across the studies.

| Gene | Chromosome | Position | Consequence | Impact | Exon | CDS Position | Protein Position | Codon (ref) | Codon (new) | Amino Acid (ref) | Amino Acid (new) |
|-------|--------------|----------|----------------|----------|-------|--------------|------------------|-----------------------|-----------------|-----------------------|-----------------------|
| Trp53 | chr3 | 69587616 | Missense var | MODERATE | 4\11 | 341 | 114 | AAG | ATG | K [Lysine LYS] | M [Methionine MET] |
| | | 69587616 | Missense var | MODERATE | 4\11 | 341 | 114 | AAG | ATG | K [Lysine LYS] | M [Methionine MET] |
| | | 69587616 | Missense var | MODERATE | 4\11 | 350 | 117 | AAG | ATG | K [Lysine LYS] | M [Methionine MET] |
| | | 69587616 | Missense var | MODERATE | 4\11 | 350 | 117 | AAG | ATG | K [Lysine LYS] | M [Methionine MET] |
| BRAF | chr6 | 39646740 | Missense var | MODERATE | 12\22 | 1267 | 423 | GGC | CGC | G [Glycine GLY] | R [Arginine ARG] |
| SP140 | chr1 | 85609827 | Stop gained | HIGH | 2\17 | 166 | 56 | CGA | TGA | R [Arginine ARG] | * [STOP] |
| | | 85609841 | Synonymous var | LOW | 2\17 | 180 | 60 | TCG | TCA | S [Serine SER] | S [Serine SER] |
| | | 85628219 | Stop lost | HIGH | 7\8 | 478 | 160 | TAA | AAA | * [STOP] | K [Lysine LYS] |
| H2-Q7 | chr17 | 35440154 | Missense var | MODERATE | 3\6 | 580 | 194 | CAG | GAG | Q [Glutamine GLN] | E [Glutamic Acid GLU] |
| NF1 | chr11 | 79475798 | Missense var | MODERATE | 35\57 | 4693 | 1565 | TTC | GTC | F [Phenylalanine PHE] | V [Valine VAL] |
| | | | Missense var | MODERATE | 36\58 | 4756 | 1586 | TTC | GTC | F [Phenylalanine PHE] | V [Valine VAL] |
| | | 79547114 | Synonymous var | LOW | 41\57 | 6442 | 2078 | GCG | GCA | A [Alanine ALA] | A [Alanine ALA] |
| | | | Synonymous var | LOW | 42\58 | 6472 | 2099 | GCG | GCA | A [Alanine ALA] | A [Alanine ALA] |
| | | | Synonymous var | LOW | 2\18 | 249 | 83 | GCG | GCA | A [Alanine ALA] | A [Alanine ALA] |
| | | 79547135 | Missense var | MODERATE | 41\57 | 6463 | 2085 | TTC | TTG | F [Phenylalanine PHE] | L [Leucine LEU] |
| | | | Missense var | MODERATE | 42\58 | 6493 | 2106 | TTC | TTG | F [Phenylalanine PHE] | L [Leucine LEU] |
| | Missense var | MODERATE | 2\18 | 270 | 90 | TTC | TTG | F [Phenylalanine PHE] | L [Leucine LEU] | | |

Supplementary Table 2. Genes that are significantly mutated in MM patients that show nucleotide variants in 5TGM1 RNAseq data and the predicted consequences on protein formation. Analysis of human MM genes with variants found in 5TGM1 transcripts using the Ensembl Variant Effect Predictor tool revealed variants in 5 genes had potential impacts on their protein formation. *Sp140* was identified to harbour high impact changes inducing a STOP gain and a STOP loss.

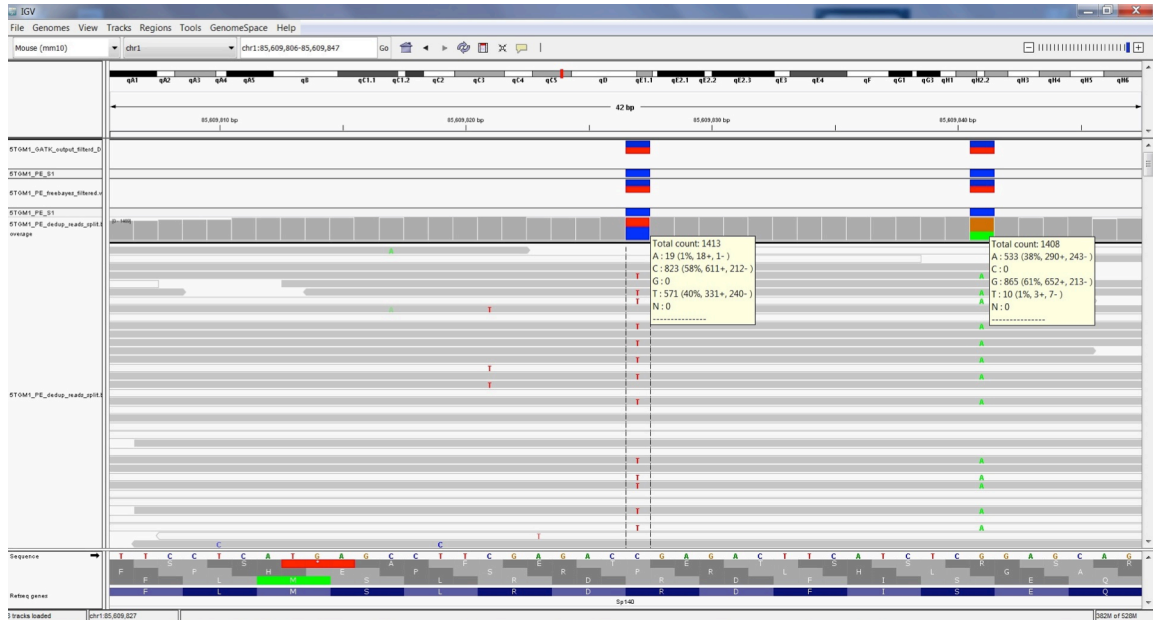
| mm10 cell line | c.166 C | c.166 C>T | c.180 G | c.180 G>A | (%) c.166 C>T | (%) c.180 G>A |
|----------------|---------|-----------|---------|-----------|---------------|---------------|
| 5TGM1 | 368 | 532 | 392 | 371 | 59.11 | 48.62 |
| FDCP1 | 387 | 509 | 341 | 340 | 56.81 | 49.93 |
| NS1 | 612 | 258 | 606 | 168 | 29.66 | 21.71 |
| NIH 3T3 | 773 | 101 | 829 | 80 | 11.56 | 8.80 |
| BA/F3 | 975 | 98 | 830 | 143 | 9.13 | 14.70 |
| RAW264.7 | 833 | 82 | 836 | 39 | 8.96 | 4.46 |

Supplementary Table 3. Differential site specific RNA editing occurring in mouse cell lines. Calculation of the percentage of altered nucleotide at RNA editing sites c.166 and c.180 in *Sp140* exon 2 revealed FDCP1 and NS1 cell lines show much higher editing changes than other mouse cell lines analysed.

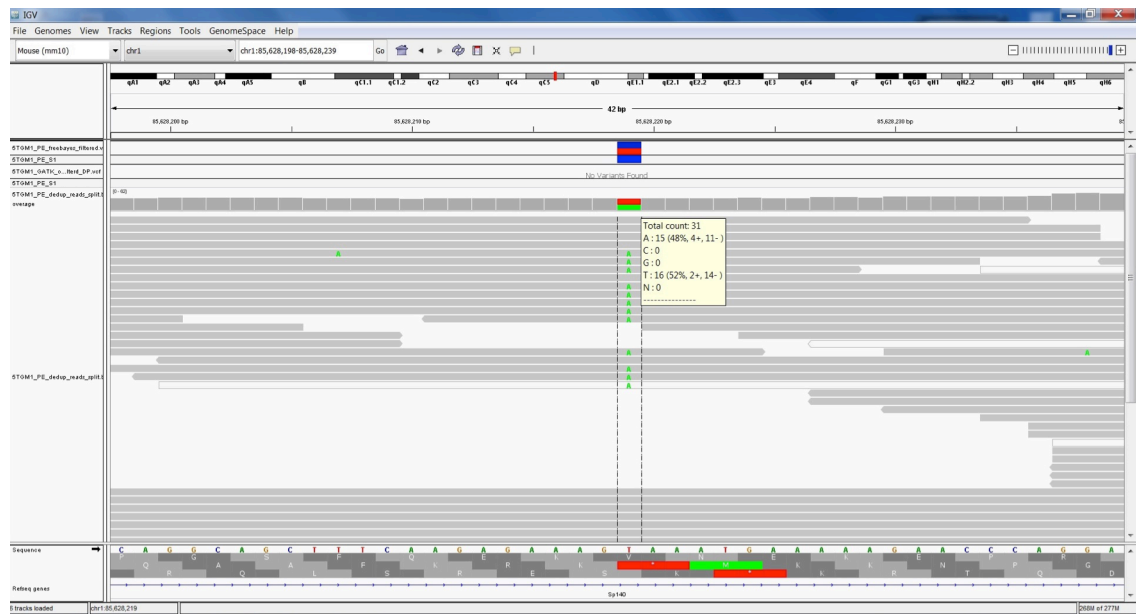
| MM cell line | RNAseq counts | |
|--------------|---------------|-----------------------|
| JIM-1 | 16 | |
| KMS-11 | 9\27 | (Adherent\Suspension) |
| KMS-18 | 1530 | |
| LP-1 | 40 | |
| MM.1R | 1235 | |
| MM.1S | 1418 | |
| NCI-H929 | 153 | |
| OPM2 | 7 | |
| RPMI 8226 | 485 | |
| U266 | 1265 | |
| JJN3 | 735 | |
| KMM-1 | 19 | |
| MOLP-8 | 63 | |
| EJM | 151 | |

High
 Mid

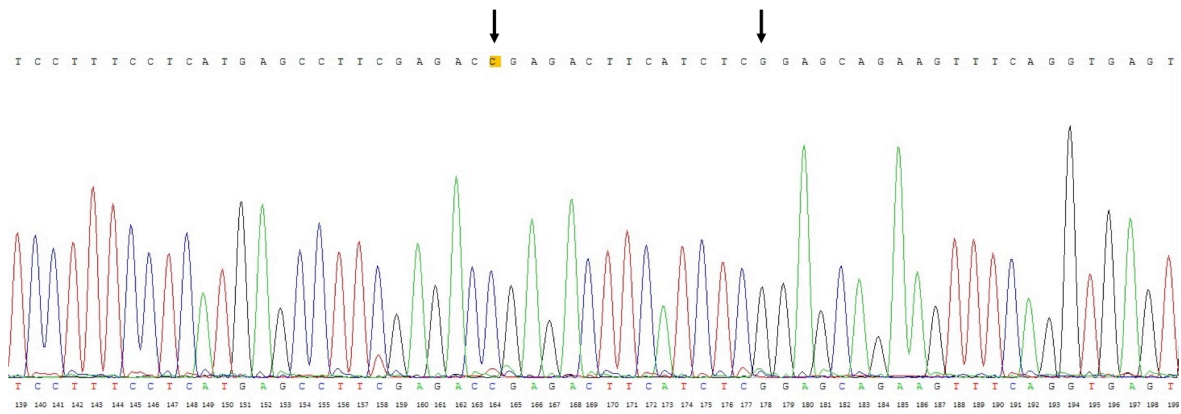
Supplementary Table 4. SP140 expression in human MM cell lines. Examination of RNAseq data from Keats laboratory for *SP140* expression demonstrates varied levels of expression. Two groups were portioned based on “high” expressing (orange) and “moderate” expressing (yellow) cell lines.



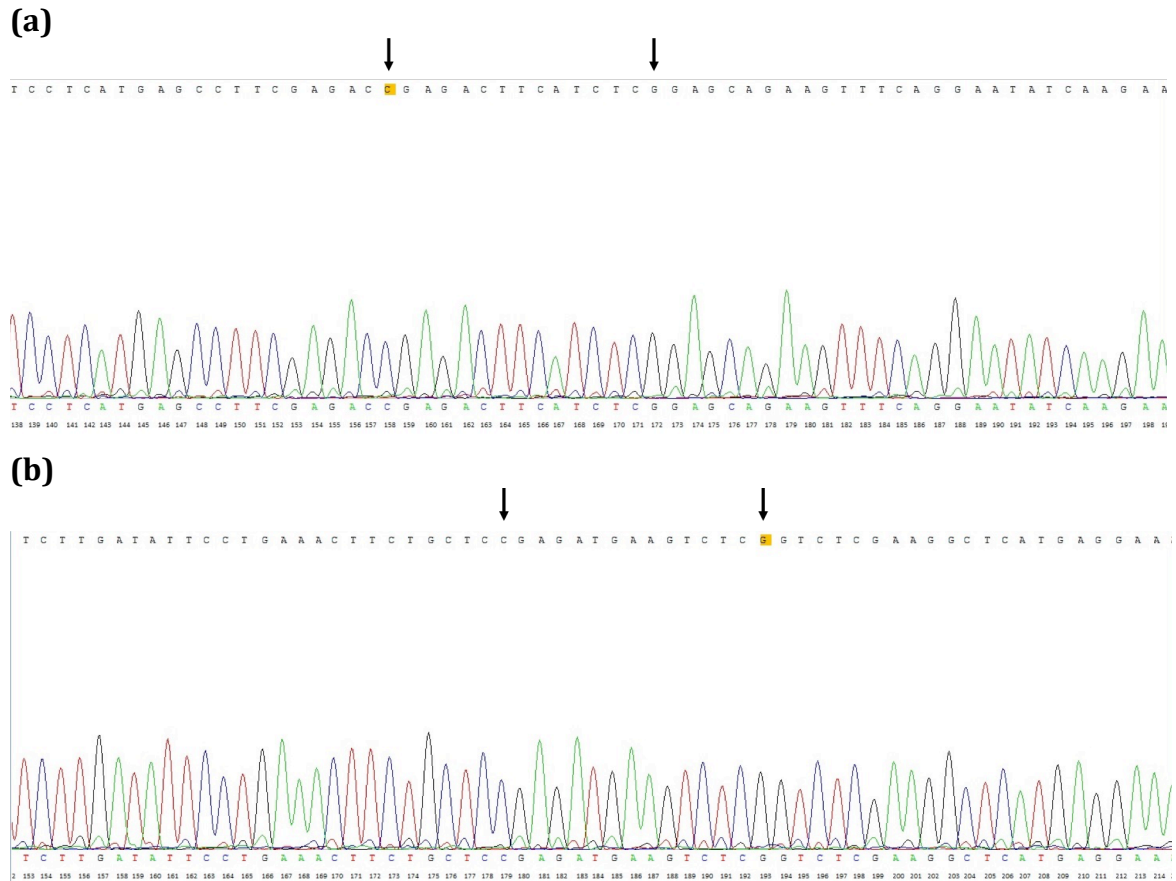
Supplementary Figure 1. IGV sequence trace of *Sp140* Exon 2. Exploring RNAseq reads of *Sp140* exon 2 in the 5TGM1 cell line confirms ~heterozygous C>T changes at c.166 and G>A changes at c.180.



Supplementary Figure 2. IGV sequence trace of *Sp140* Exon 7. Exploring RNAseq reads of *Sp140* exon 7 in the 5TGM1 cell line confirms ~heterozygous T>A changes at c.478.



Supplementary Figure 3. Sequencing of identified RNA modifications in *Sp140* at the gDNA level using Exon2L primer pair set. A secondary genomic primer pair (Exon 2L) was designed to amplify a larger region of *Sp140* exon 2 in 5TGM1 PCs. We again found exon 2 of *Sp140* to unmutated at the genomic level, demonstrating that this observation is not due to a specific primer set. Sanger sequencing revealed a clean sequencing trace that supports the reference nucleotides of C and G at positions c.166 and c.180, respectively. Sites of interest are indicated by arrows, with position c.166 highlighted in yellow.



Supplementary Figure 4. Exogenous *Sp140* RNA construct (Minigene) is not edited in 5TGM1 cells. The Minigene consisting of *Sp140* exon 2 and its surrounding elements of exon 1 and 3 was not sufficient as a template for RNA editing in 5TGM1 PCs. pLeGo Cer2 vector specific primers were used to amplify the Minigene region, and Sanger sequencing revealed a clean sequencing trace that supports the reference nucleotides of C and G at positions c.166 and c.180, respectively. This illustrates further *Sp140* gene elements, such as intronic sequences, are needed for Apobec enzyme activity. Sites of interest are indicated by arrows, with position c.166 highlighted in yellow.

a

Unmutated Nucleotide Sequence (690 nt) of the coding region: alt exons in black-blue

```
ATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCACCTCTGAGGAGAAGAATTGAGCCCCACGAGTTTG
AAGTCTTCTTTGACCCCCGGGAGCTTCGGAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGTGGGAAG
GCACAGTGTCTGGCGACACACGAGCCAAAACACCAGCAACCACGTTGAAGTCAACTTCTTAGAAAAATTT
ACTACAGAAAGATACTTTTCGTCCGAACACCAGATGCTCCATTACCTGGTTCCTGTCTGGAGTCCCTGCG
GGGAGTGCTCCAGGGCCATTACAGAGTTTCTGAGCCGACACCCCTATGTAACCTCTGTTTATTTACATAGC
ACGGCTTTATCACCACACGGATCAGCGAAACCGCCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACT
ATCCAGATCATGACAGAGCAAGAGTATTGTTACTGCTGGAGGAATTTTCGTCAACTACCCCCCTTCAAACG
AAGCTTATTGGCCAAGGTACCCCCATCTGTGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTT
AGGACTTCCACCCTGTTTAAAAATTTTAAGAAGAAAGCAACCTCAACTCACGTTTTTTCACAATTACTCTT
CAAACCTGCCATTACCAAAGGATACCACCCCATCTCCTTTGGGCTACAGGGTTGAAA TGA
```

b

Mutant Allele 1 – Apobec1-sgRNA2-Clone#1 : 4 bp deletion:

```
ATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCACCTCTGAGGAGAAGAATTGAGCCCCACGAGTTTG
AAGTCTTCTTTGACC----GGAGCTTCGGAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGTGGGAAG
GCACAGTGTCTGGCGACACACGAGCCAAAACACCAGCAACCACGTTGAAGTCAACTTCT TAGAAAAATTT
ACTACAGAAAGATACTTTTCGTCCGAACACCAGATGCTCCATTACCTGGTTCCTGTCTGGAGTCCCTGCG
GGGAGTGCTCCAGGGCCATTACAGAGTTTCTGAGCCGACACCCCTATGTAACCTCTGTTTATTTACATAGC
ACGGCTTTATCACCACACGGATCAGCGAAACCGCCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACT
ATCCAGATCATGACAGAGCAAGAGTATTGTTACTGCTGGAGGAATTTTCGTCAACTACCCCCCTTCAAACG
AAGCTTATTGGCCAAGGTACCCCCATCTGTGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTT
AGGACTTCCACCCTGTTTAAAAATTTTAAGAAGAAAGCAACCTCAACTCACGTTTTTTCACAATTACTCTT
CAAACCTGCCATTACCAAAGGATACCACCCCATCTCCTTTGGGCTACAGGGTTGAAATGA
```

c

Mutant Allele 2 – Apobec1-sgRNA2-Clone#1 : 4 bp deletion:

```
ATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCACCTCTGAGGAGAAGAATTGAGCCCCACGAGTTTG
AAGTCTTCTTTGACCCCC----GCTTCGGAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGTGGGAAG
GCACAGTGTCTGGCGACACACGAGCCAAAACACCAGCAACCACGTTGAAGTCAACTTCT TAGAAAAATTT
ACTACAGAAAGATACTTTTCGTCCGAACACCAGATGCTCCATTACCTGGTTCCTGTCTGGAGTCCCTGCG
GGGAGTGCTCCAGGGCCATTACAGAGTTTCTGAGCCGACACCCCTATGTAACCTCTGTTTATTTACATAGC
ACGGCTTTATCACCACACGGATCAGCGAAACCGCCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACT
ATCCAGATCATGACAGAGCAAGAGTATTGTTACTGCTGGAGGAATTTTCGTCAACTACCCCCCTTCAAACG
AAGCTTATTGGCCAAGGTACCCCCATCTGTGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTT
AGGACTTCCACCCTGTTTAAAAATTTTAAGAAGAAAGCAACCTCAACTCACGTTTTTTCACAATTACTCTT
CAAACCTGCCATTACCAAAGGATACCACCCCATCTCCTTTGGGCTACAGGGTTGAAATGA
```

Supplementary Figure 5. Apobec1 frameshift mutant clone in 5TGM1. (a) Full unmutated nucleotide sequence of Apobec1. Mutation deconvolution illustrates the frameshift-inducing 4bp deletions in the Apobec1-sgRNA2 5TGM1 clonal PC line (Apobec1-sgRNA2-clone#1) **(b)** mutant allele 1 and **(c)** mutant allele 2. Green highlighted text illustrates the start codon sites, while red highlighted text show the stop codon sites. Both mutant Apobec1 alleles are observed to harbour early stop codon sites compared to unmutated Apobec1.

(a)

Translation of Unmutated Gene (229aa):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPR^{ELRKETCLLYEINWGGRRHSVWRHTSQNTSNHVEVNFLEKF}
TTERYFRPNTRCSITWFLSWS**PCGEC**SRRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVT
IQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTTIL
QTCHYQRIPPHLLWATGLK

(b)

Predicted Translation of Mutant Allele 1 (65 aa):

MSSETGPVAVDPTLRRRIEPHEFEVFFD**RS**FGKRPVCCMRSTGVEGTVSGDTRAKTPATTLKSTS

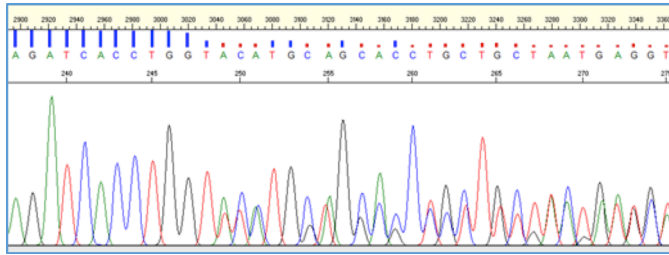
(c)

Predicted Translation of Mutant Allele 2 (65 aa):

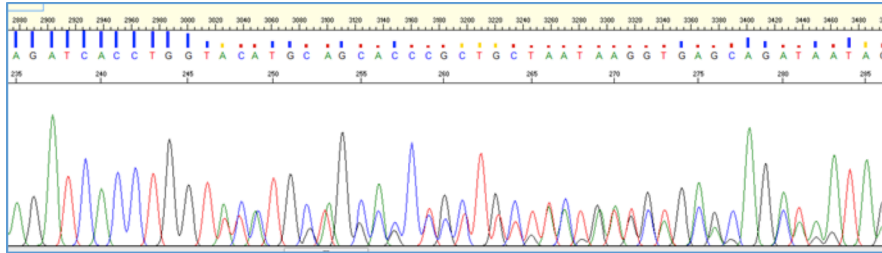
MSSETGPVAVDPTLRRRIEPHEFEVFFD**RS**FGKRPVCCMRSTGVEGTVSGDTRAKTPATTLKSTS

Supplementary Figure 6. Predicted effects of frameshift mutations on Apobec1 protein formation. (a) Translation of unmutated *Apobec1* results in a 229 amino acid protein. CRISPR-Cas9 mediated mutations in *Apobec1* with (b) mutant allele 1 and (c) mutant allele 2 predicted to encode severely truncated protein. Mutated *Apobec1* results in 65 amino acid proteins with significant truncation of C-terminus, missing the motifs that are required for its cytidine deaminase activity. Green highlighted text illustrates the wildtype protein sequence, while yellow highlighted text illustrates the mutated C-terminal protein sequence. Cyan highlighted text shows the altered protein sequence exclusive to Mutant Allele 1. Emboldened text illustrates the conserved active site motifs of *Apobec1*.

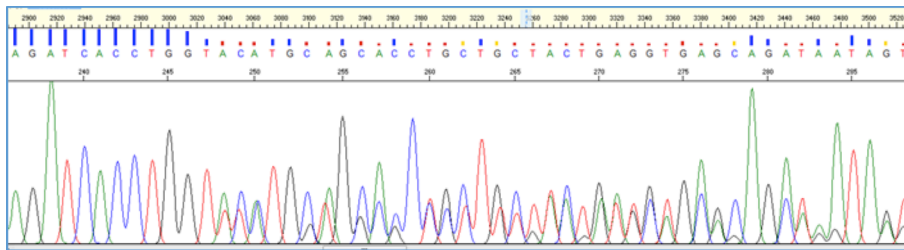
Clone#1



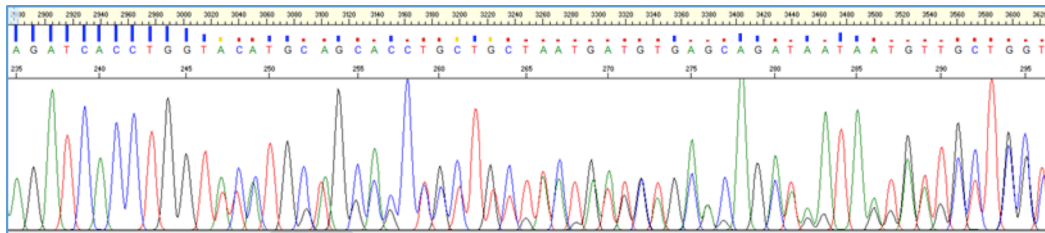
Clone#2



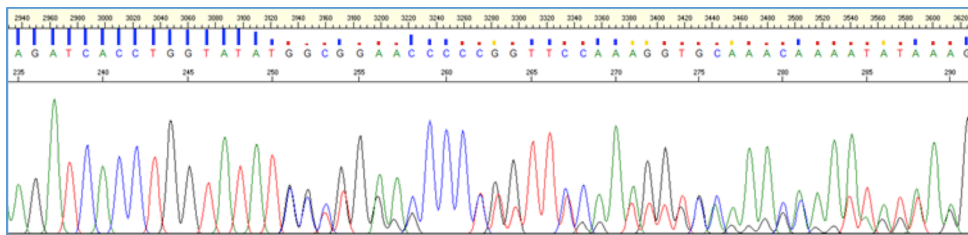
Clone#3



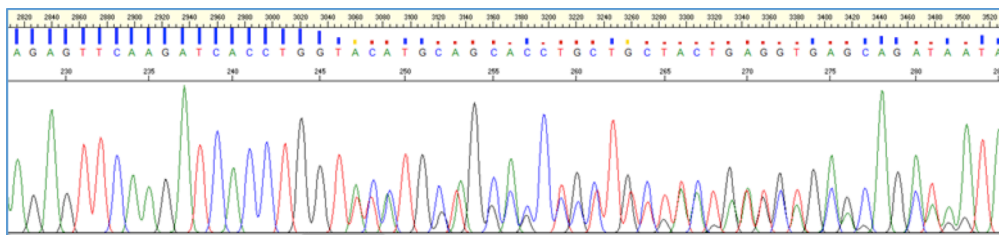
Clone#4



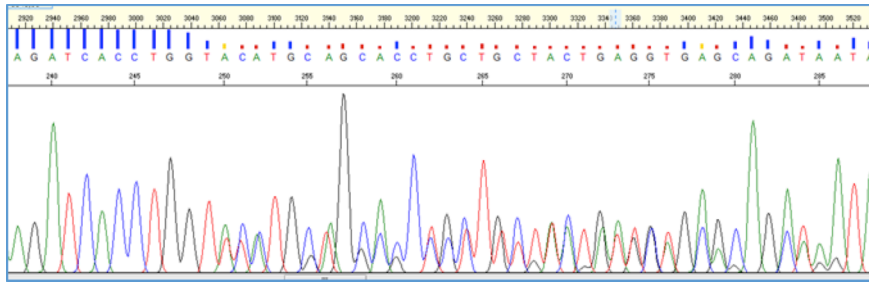
Clone#5



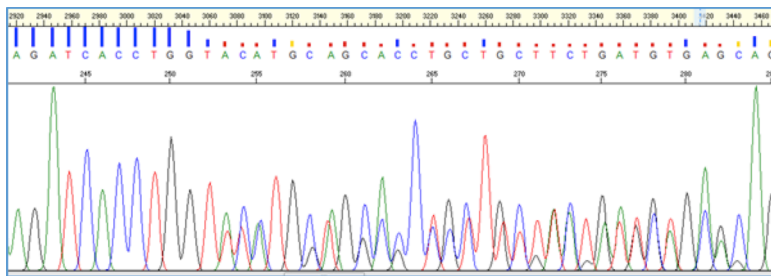
Clone#6



Clone#7

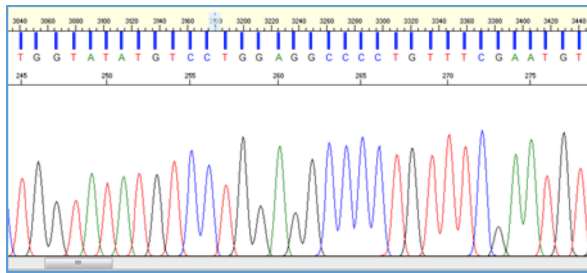


Clone#9

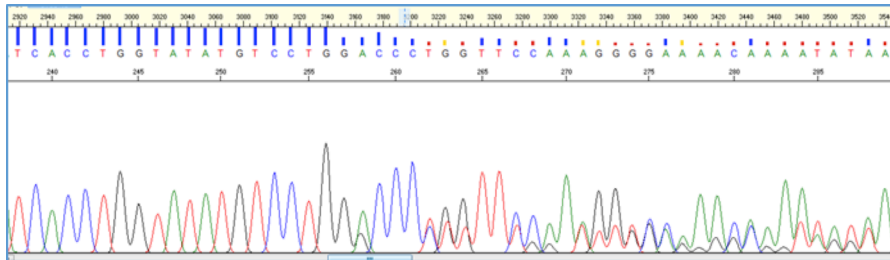


Supplementary Figure 7. Generation of Apobec3 AS1 sense sgRNA clones with two mutant alleles. Sequencing traces show induced CRISPR-Cas9 gene editing using sense sgRNA targeting AS1 of Apobec3 resulting in 8 different clonal 5TGM1 cell lines harbouring homozygous or compound heterozygous frameshift mutations in exon 3 of Apobec3.

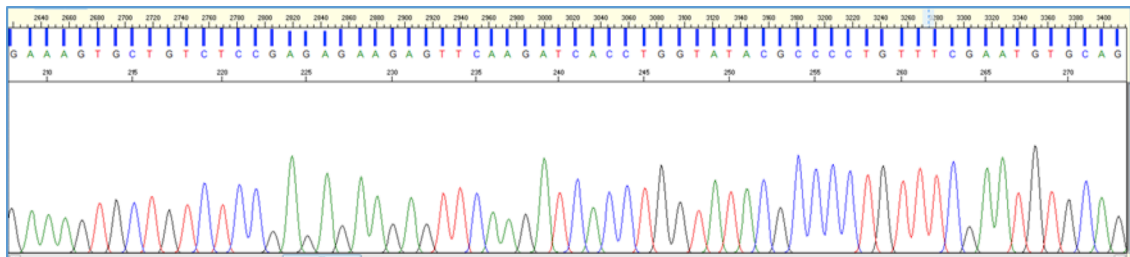
Clone#1



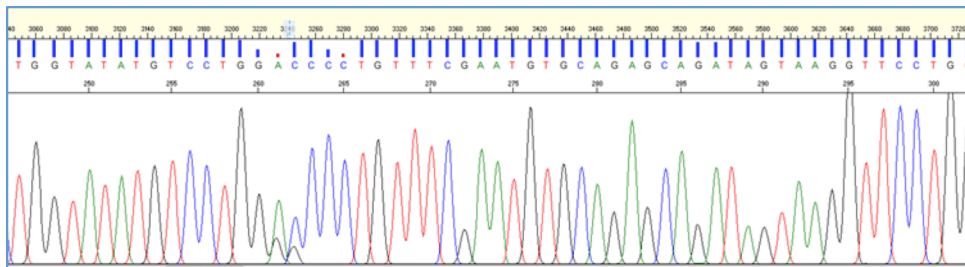
Clone#2



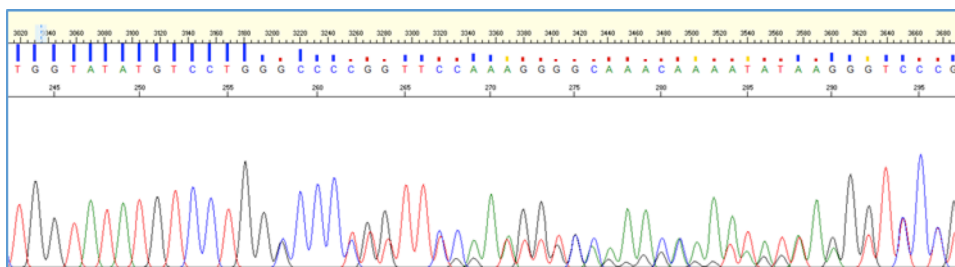
Clone#3



Clone#6

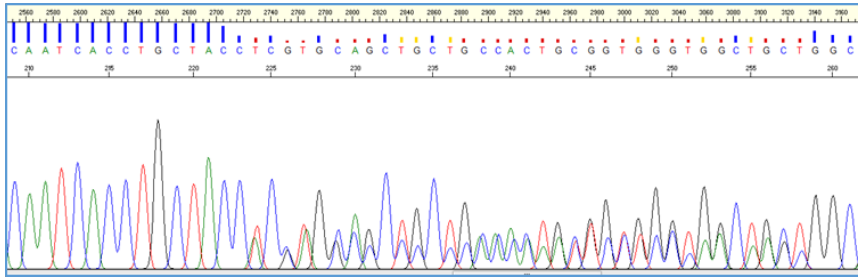


Clone#9

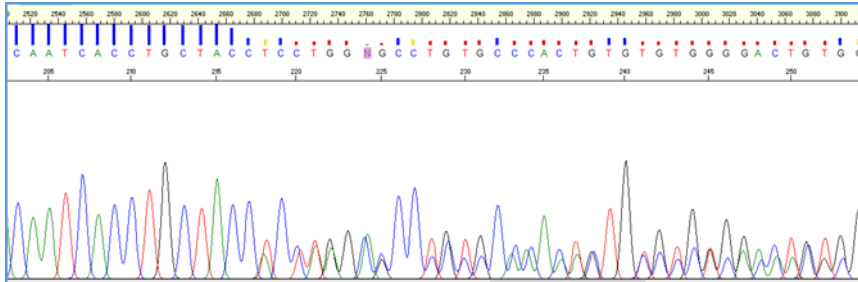


Supplementary Figure 8. Generation of Apobec3 AS1 antisense sgRNA clones with two mutant alleles. Sequencing traces show induced CRISPR-Cas9 gene editing using antisense sgRNA targeting AS1 of Apobec3 resulting in 5 different clonal 5TGM1 cell lines harbouring homozygous or compound heterozygous frameshift mutations in exon 3 of Apobec3.

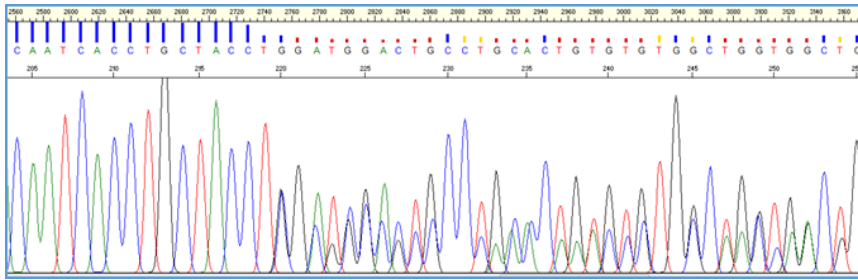
Clone#1



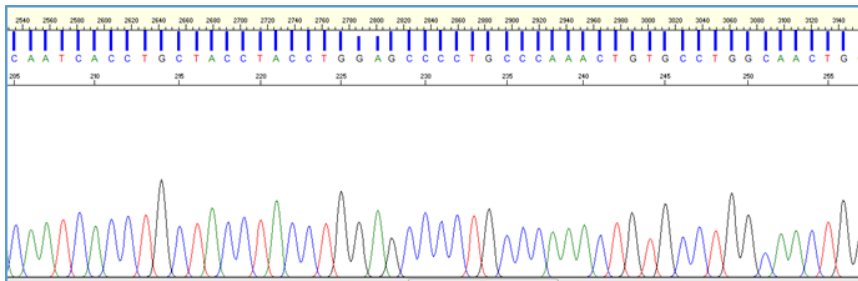
Clone#2



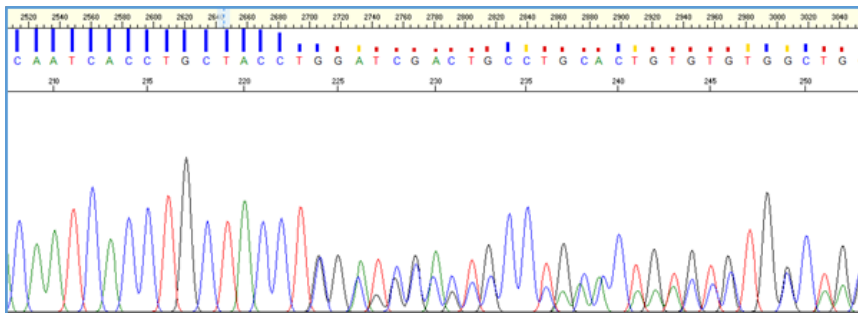
Clone#4



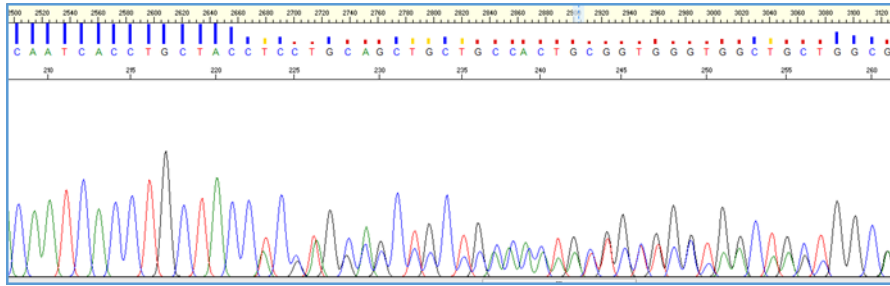
Clone#5



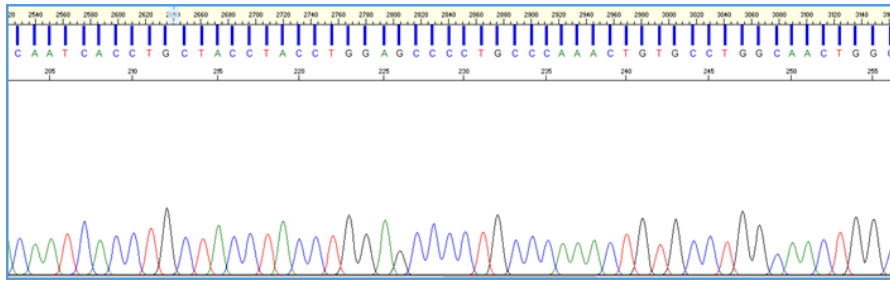
Clone#7



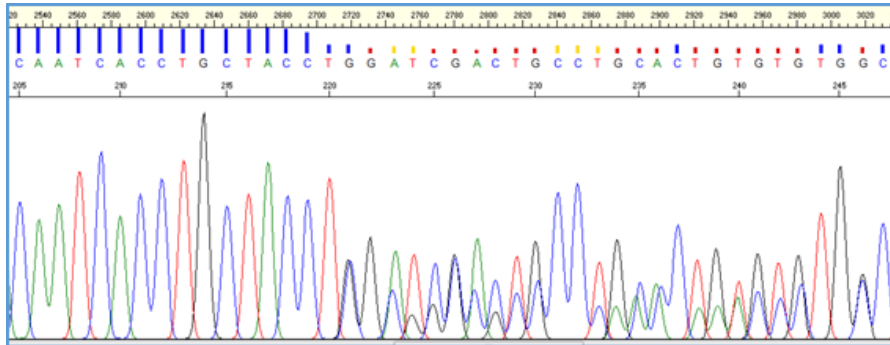
Clone#8



Clone#9

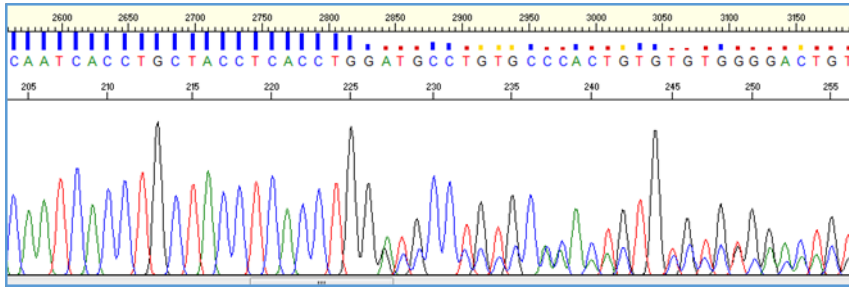


Clone#10

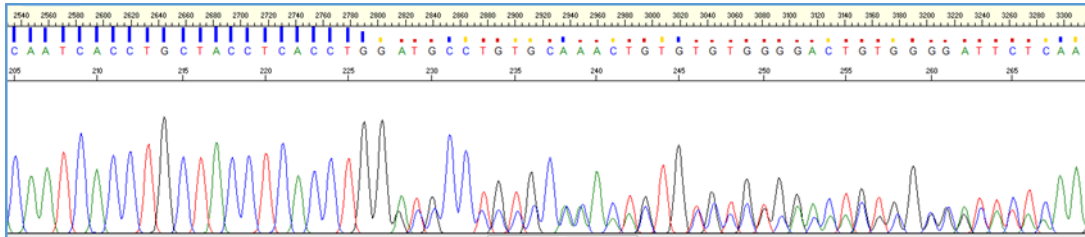


Supplementary Figure 9. Generation of Apobec3 AS2 sense sgRNA clones with two mutant alleles. Sequencing traces show induced CRISPR-Cas9 gene editing using sense sgRNA targeting AS2 of Apobec3 resulting in 8 different clonal 5TGM1 cell lines harbouring homozygous or compound heterozygous frameshift mutations in exon 7 of Apobec3.

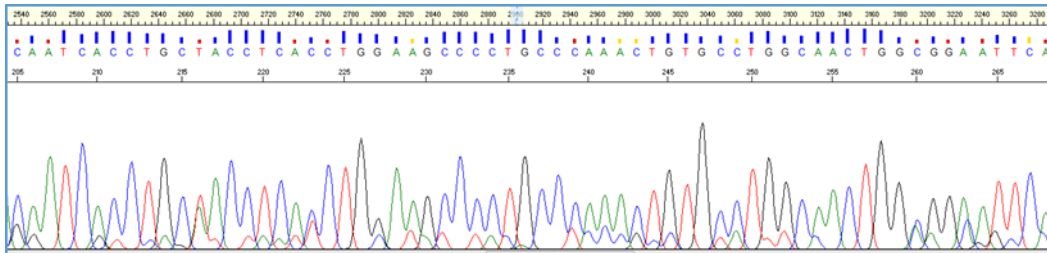
Clone#1



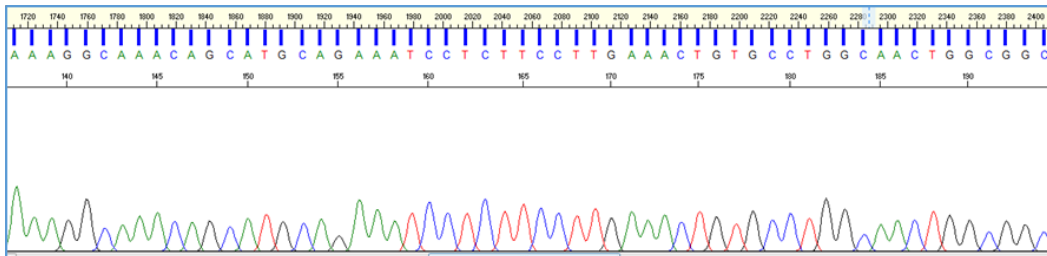
Clone#5



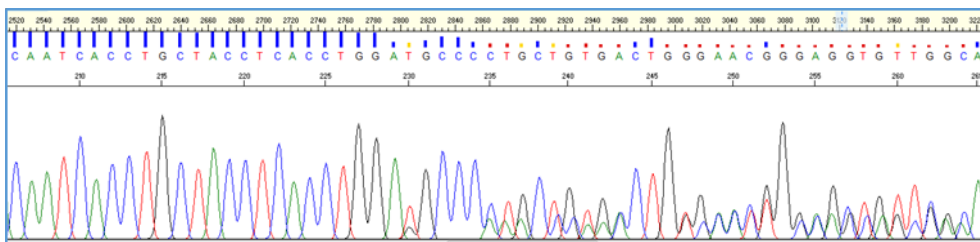
Clone#6 allele 1 (larger PCR product)
(contaminated slightly with smaller PCR product)



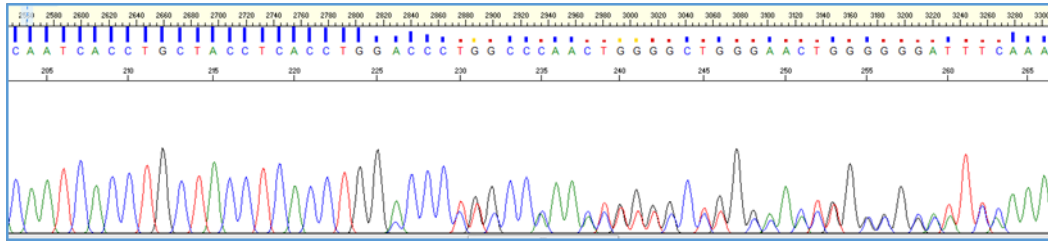
Clone#6 allele 2 (smaller PCR product)



Clone#7



Clone#9



Supplementary Figure 10. Generation of Apobec3 AS2 antisense sgRNA clones with two mutant alleles. Sequencing traces show induced CRISPR-Cas9 gene editing using antisense sgRNA targeting AS2 of Apobec3 resulting in 5 different clonal 5TGM1 cell lines harbouring homozygous or compound heterozygous frameshift mutations in exon 7 of Apobec3.

Clone#1-allele 1 - 160 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMLEPLFRMCRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#1-allele 2 - 160 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQP
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#2 – allele 1 - 158 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMP_LFRMCRADSKVPGYTPQPE
PGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#2 - allele 2 - 157 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEFRMCRADSKVPGYTPQPEP
GHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#3 – allele 1 – 160 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMLEPLFRMCRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#3 – allele 2 – 159 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQP
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#4 – allele 1 - 160 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMVEPLFRMCRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#4 – allele 2 - 159 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQP
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#5 – allele 1 – 160 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYILEPLFRMCRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPRI

Clone#5 – allele 2 – 111 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMAGAPVSNVQSR

Clone#6 – allele 1 – 159 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMEPLFRMCRADSKVPGYTPQ
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Clone#6 – allele 2 – 160 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQ
PEPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Clone#7 – allele 1 – 160 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMLEPLFRMCRADSKVPGYTPQ
PEPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Clone#7 – allele 2 – 159 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQ
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Clone#9 – allele 1 – 160 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMLEPLFRMCRADSKVPGYTPQ
PEPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Clone#9 – allele 2 – 159 amino acids

MGPFCLGCSRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWFLEPLFRMCRADSKVPGYTPQ
EPGHLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGPIRI

Supplementary Figure 11. Predicted effects of Apobec3 AS1 sense sgRNA double frameshift clones. All Apobec3 exon 3 mutant clones are predicted to encode for severely C-terminal truncated proteins that lack AS2 and lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS1. Green highlighted text illustrates the wildtype protein sequence.

Clone#1 – homozygous? – 161 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^RPLFRMC^RRADSKVPGYTP
QPEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGP^IRI

Clone#2 – allele 1 – 11 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^TPVSNVQSR

Clone#2 – allele 2 – 160 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^ALFRMC^RRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGP^IRI

Clone#3 – homozygous? – 158 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWY^TPLFRMC^RRADSKVPGYTPQPE
PGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGP^IRI

Clone#6 – allele 1 – 111 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^APVSNVQSR

Clone#6 – allele 2 – 111 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^SPVSNVQSR

Clone#9 – allele 1 – 111 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^APVSNVQSR

Clone#9 – allele 2 – 160 amino acids

MGPFC LGCSHRKCYSP^{IR}NLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW^PLFRMC^RRADSKVPGYTPQ
PEPGLQLPPLQRTGPRNPAESLQAGSGRSPGGCHGP^IRI

Supplementary Figure 12. Predicted effects of Apobec3 AS1 antisense sgRNA double frameshift clones. All Apobec3 exon 3 mutant clones are predicted to encode for severely C-terminal truncated proteins that lack AS2 and lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS1. Green highlighted text illustrates the wildtype protein sequence.

Clone#1 – allele 1 – 372 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYHLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFSVAIR
DPGGRHGPPPTVY

Clone#1 – allele 2 – 371 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYHLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFSVAIRD
PGGRHGPPPTVY

Clone#2 – allele 1 – 372 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYHLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFSVAIR
DPGGRHGPPPTVY

Clone#2 – allele 2 – 311 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYL

Clone#4 – allele 1 – 371 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFSVAIRD
PGGRHGPPPTVY

Clone#4 – allele 2 – 332 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQD SKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLAGAPAQTVPGNWRHSGKIVQI

Clone#5– homozygous? – 332 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLPGAPAQTVPGNWRHSGKIVQI

Clone#7 – allele 1 – 371 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSVAIRD
PGGRHGPPTVY

Clone#7 – allele 2 – 332 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLAGAPAQTVPGNWRHSGKIVQI

Clone#8 – allele 1 – 371 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSVAIRD
PGGRHGPPTVY

Clone#8 – allele 2 – 372 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYHLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSVAIR
DPGGRHGPPTVY

Clone#9 – homozygous? – 332 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEIILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLDKIRSM
ELSQVTITCYLPGAPAQTVPGNWRHSGKIVQI

Clone#10 – allele 1 – 332 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQONLCRLVQEGAQVAAMDLYEFKCCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGQHAEILFLDKIRSM
ELSQVTITCYLAGAPAQTVPGNWRHSGIVQI

Clone#10 – allele 2 – 371 amino acids

MGPFCLGCSHRKCYSPINRLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHH
NLSLDIFSSRLYNVQDPETQONLCRLVQEGAQVAAMDLYEFKCCWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGQHAEILFLDKIRSM
ELSQVTITCYLEPLPKLCLATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSVAIRD
PGGRHGPPTVY

Supplementary Figure 13. Predicted effects of Apobec3 AS2 sense sgRNA double frameshift clones. All Apobec3 exon 7 mutant clones are predicted to encode for C-terminally truncated proteins that lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS2. Green highlighted text illustrates the wildtype protein sequence.

Clone#1 – allele 1 – 332 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LD^KI^RS^M
E^LS^QV^TI^TC^YL^TW^AP^AQ^TV^PG^NW^RH^SK^GI^VQ^I

Clone#1 – allele 2 – 373 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LD^KI^RS^M
E^LS^QV^TI^TC^YL^TW^MP^LP^KL^CL^AT^GG^IQ^KG^SS^RS^NS^AY^LH^LP^PV^FP^LE^EA^LP^EG^AV^FS^VA^I
R^DP^GG^RH^GP^PT^VY

Clone#5 – allele 1 – 332 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LD^KI^RS^M
E^LS^QV^TI^TC^YL^TW^AP^AQ^TV^PG^NW^RH^SK^GI^VQ^I

Clone#5 – allele 2 – 373 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LD^KI^RS^M
E^LS^QV^TI^TC^YL^TW^MP^LP^KL^CL^AT^GG^IQ^KG^SS^RS^NS^AY^LH^LP^PV^FP^LE^EA^LP^EG^AV^FS^VA^I
R^DP^GG^RH^GP^PT^VY

Clone#6 – allele 1 – 373 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LD^KI^RS^M
E^LS^QV^TI^TC^YL^TW^KP^LP^KL^CL^AT^GG^IQ^KG^SS^RS^NS^AY^LH^LP^PV^FP^LE^EA^LP^EG^AV^FS^VA^I
R^DP^GG^RH^GP^PT^VY

Clone#6 – allele 2 – 310 amino acids

MGPFCLGCSHRKCYSP^IRN^LI^SQ^ET^FK^FH^FK^NL^GY^AK^RK^DT^FL^CY^EV^TR^KD^CD^SP^VS^LH
H^GV^FK^NK^DN^IH^AE^IC^FL^YW^FH^DK^VL^KV^LS^PR^EE^FK^IT^WY^MS^WS^PC^FE^CA^EQ^IV^RF^LA^TH^H
N^LS^LD^IF^SS^RL^YN^VQ^DP^ET^QQ^NL^CR^LV^QE^GA^QV^AA^MD^LY^EF^KK^CW^KK^FV^DN^GG^RR^RF^RP^WK
R^LL^TN^FR^YQ^DS^KL^QE^IL^RP^CY^IS^VP^SS^SS^ST^LS^NI^CL^TK^GL^PE^TR^FW^VE^GR^RM^DP^LS^EE^E
F^YS^QF^YN^QR^VK^HL^CY^YH^RM^KP^YL^CY^QL^EQ^FN^GQ^AP^LK^GC^LL^SE^KG^KQ^HA^EI^LF^LE^TV^PG^N
W^RH^SK^GI^VQ^I

Clone#7 – allele 1 – 371 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HG VFKNKDNI **HAEICFLYWFHDKVLKVLSPREEFKITWYMSWS** **PCFECAEQIVRFLATHH**
NLSDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKCKWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGK**QHAEILFLDKIRSM**
ELSQVTITCYLTM**PKL****CL**ATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSAIRD
PGGRHGPPTVY

Clone#7 – allele 2 – 373 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HG VFKNKDNI **HAEICFLYWFHDKVLKVLSPREEFKITWYMSWS** **PCFECAEQIVRFLATHH**
NLSDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKCKWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGK**QHAEILFLDKIRSM**
ELSQVTITCYLTM**PLPKL****CL**ATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSAI
RDPGGRHGPPTVY

Clone#9 – allele 1 – 332 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HG VFKNKDNI **HAEICFLYWFHDKVLKVLSPREEFKITWYMSWS** **PCFECAEQIVRFLATHH**
NLSDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKCKWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGK**QHAEILFLDKIRSM**
ELSQVTITCYLTM**PAQTV****PGNWRH**SKGIVQI

Clone#9 – allele 2 – 372 amino acids

MGPFC LGCSHRKCYSP I RNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
HG VFKNKDNI **HAEICFLYWFHDKVLKVLSPREEFKITWYMSWS** **PCFECAEQIVRFLATHH**
NLSDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKCKWKKFVDNGGRRFRPWK
RLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEE
FYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGGK**QHAEILFLDKIRSM**
ELSQVTITCYLTM**PKL****CL**ATGGIQKGSSRSNSAYLHLPVFPLEEALPEGAVFVSAIR
DPGGRHGPPTVY

Supplementary Figure 14. Predicted effects of Apobec3 AS2 antisense sgRNA double frameshift clones. All Apobec3 exon 7 mutant clones are predicted to encode for C-terminally truncated proteins that lack one or more of the cysteines within the P-C-X₂₋₄-C motif of AS2. Green highlighted text illustrates the wildtype protein sequence.

4.7 References

1. Weston-Bell N, *et al.* Exome sequencing in tracking clonal evolution in multiple myeloma following therapy. *Leukemia* **27**, 1188-1191 (2013).
2. Bolli N, *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun* **5**, 2997 (2014).
3. Lohr JG, *et al.* Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* **25**, 91-101 (2014).
4. Walker BA, *et al.* Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol* **33**, 3911-3920 (2015).
5. Chapman MA, *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472 (2011).
6. Madani N, *et al.* Implication of the lymphocyte-specific nuclear body protein Sp140 in an innate response to human immunodeficiency virus type 1. *J Virol* **76**, 11133-11138 (2002).
7. Mehta S, *et al.* Maintenance of macrophage transcriptional programs and intestinal homeostasis by epigenetic reader SP140. *Sci Immunol* **2**, (2017).
8. Franke A, *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-1125 (2010).
9. Jostins L, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
10. Sille FC, Thomas R, Smith MT, Conde L, Skibola CF. Post-GWAS functional characterization of susceptibility variants for chronic lymphocytic leukemia. *PLoS One* **7**, e29632 (2012).
11. Kortum KM, *et al.* Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann Hematol* **94**, 1205-1211 (2015).
12. Lionetti M, Neri A. Utilizing next-generation sequencing in the management of multiple myeloma. *Expert Rev Mol Diagn* **17**, 653-663 (2017).
13. Michigami T, *et al.* Cell-cell contact between marrow stromal cells and myeloma cells via VCAM-1 and α 4B1-integrin enhances production of osteoclast-stimulating activity. *Blood* **96**, 1953-1960 (2000).
14. Radl J, Croese JW, Zurcher C, Van den Enden-Vieveen MH, de Leeuw AM. Animal model of human disease. Multiple myeloma. *Am J Pathol* **132**, 593-597 (1988).

15. Radl J, De Glopper ED, Schuit HR, Zurcher C. Idiopathic paraproteinemia. II. Transplantation of the paraprotein-producing clone from old to young C57BL/KaLwRij mice. *J Immunol* **122**, 609-613 (1979).
16. Hewett DR, *et al.* DNA Barcoding Reveals Habitual Clonal Dominance of Myeloma Plasma Cells in the Bone Marrow Microenvironment. *Neoplasia* **19**, 972-981 (2017).
17. Mrozik KM, *et al.* Therapeutic targeting of N-cadherin is an effective treatment for multiple myeloma. *Br J Haematol* **171**, 387-399 (2015).
18. Noll JE, *et al.* SAMS1 is a tumor suppressor gene in multiple myeloma. *Neoplasia* **16**, 572-585 (2014).
19. Kumar SK, *et al.* Multiple Myeloma. *Nature Reviews Disease Primers* **3**, 1-20 (2017).
20. Jiang Q, *et al.* ADAR1 promotes malignant progenitor reprogramming in chronic myeloid leukemia. *Proc Natl Acad Sci U S A* **110**, 1041-1046 (2013).
21. Jiang Q, Crews LA, Holm F, Jamieson CHM. RNA editing-dependent epitranscriptome diversity in cancer stem cells. *Nat Rev Cancer* **17**, 381-392 (2017).
22. Lazzari E, *et al.* Alu-dependent RNA editing of GLI1 promotes malignant regeneration in multiple myeloma. *Nat Commun* **8**, 1922 (2017).
23. Chen L, *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat Med* **19**, 209-216 (2013).
24. Galeano F, *et al.* ADAR2-editing activity inhibits glioblastoma growth through the modulation of the CDC14B/Skp2/p21/p27 axis. *Oncogene* **32**, 998-1009 (2013).
25. Martinez HD, *et al.* RNA editing of androgen receptor gene transcripts in prostate cancer cells. *J Biol Chem* **283**, 29938-29949 (2008).
26. Han SW, *et al.* RNA editing in RHOQ promotes invasion potential in colorectal cancer. *J Exp Med* **211**, 613-621 (2014).
27. Hu X, *et al.* RNA editing of AZIN1 induces the malignant progression of non-small-cell lung cancers. *Tumour Biol* **39**, 1010428317700001 (2017).
28. Baysal BE, Sharma S, Hashemikhabir S, Janga SC. RNA Editing in Pathogenesis of Cancer. *Cancer Res* **77**, 3733-3739 (2017).
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
30. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).

31. Weber K, Bartsch U, Stocking C, Fehse B. A multicolor panel of novel lentiviral "gene ontology" (LeGO) vectors for functional gene analysis. *Mol Ther* **16**, 698-706 (2008).
32. Aubrey BJ, *et al.* An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Rep* **10**, 1422-1432 (2015).
33. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108 (2008).
34. Jarmuz A, *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285-296 (2002).
35. Betts L, Xiang S, Short SA, Wolfenden R, Carter CW, Jr. Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J Mol Biol* **235**, 635-656 (1994).
36. Yamanaka S, Poksay KS, Balestra ME, Zeng GQ, Innerarity TL. Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc motif is essential for catalytic activity, and noncatalytic auxiliary factor(s) of the editing complex are widely distributed. *J Biol Chem* **269**, 21725-21734 (1994).
37. MacGinnitie AJ, Anant S, Davidson NO. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J Biol Chem* **270**, 14768-14775 (1995).
38. Teng BB, Ochsner S, Zhang Q, Soman KV, Lau PP, Chan L. Mutational analysis of apolipoprotein B mRNA editing enzyme (APOBEC1). structure-function relationships of RNA editing and dimerization. *J Lipid Res* **40**, 623-635 (1999).
39. Nair S, Sanchez-Martinez S, Ji X, Rein A. Biochemical and biological studies of mouse APOBEC3. *J Virol* **88**, 3850-3860 (2014).
40. Hakata Y, Landau NR. Reversed functional organization of mouse and human APOBEC3 cytidine deaminase domains. *J Biol Chem* **281**, 36624-36631 (2006).
41. Levine AJ, Hu W, Feng Z. Tumour Suppressor Genes. In: *The Molecular Basis of Cancer* (ed[^](eds Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson CB). Elsevier.
42. Harris SL, *et al.* Detection of functional single-nucleotide polymorphisms that affect apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16297-16302 (2005).
43. Peltz SW, Brown AH, Jacobson A. mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev* **7**, 1737-1754 (1993).
44. Sharma S, *et al.* APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun* **6**, 6881 (2015).

45. Sharma S, Patnaik SK, Taggart RT, Baysal BE. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Sci Rep* **6**, 39100 (2016).
46. Bhattacharya C, Aggarwal S, Kumar M, Ali A, Matin A. Mouse apolipoprotein B editing complex 3 (APOBEC3) is expressed in germ cells and interacts with dead-end (DND1). *PLoS One* **3**, e2315 (2008).
47. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229 (2008).
48. Niavarani A, *et al.* APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts. *PLoS One* **10**, e0120089 (2015).
49. Brioli A, Melchor L, Cavo M, Morgan GJ. The impact of intra-clonal heterogeneity on the treatment of multiple myeloma. *Br J Haematol* **165**, 441-454 (2014).
50. Dutta AK, Hewett DR, Fink JL, Grady JP, Zannettino ACW. Cutting edge genomics reveal new insights into tumour development, disease progression and therapeutic impacts in multiple myeloma. *Br J Haematol* **178**, 196-208 (2017).
51. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* **14**, 100-113 (2017).
52. Szalat R, Munshi NC. Genomic heterogeneity in multiple myeloma. *Curr Opin Genet Dev* **30**, 56-65 (2015).
53. Walker BA, *et al.* Intracolon heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia* **28**, 384-390 (2014).

Chapter 5

Discussion and Conclusions

5.1 General Discussion

Multiple myeloma (MM) is a blood cancer characterised by the aberrant proliferation of malignant plasma cells (PCs) in the bone marrow (BM). MM is a rare disease, accounting for 1.8% of all new cancer cases, and 2.1% of all cancer deaths per year¹. Each year there are approximately 1800 and 31,000 newly diagnosed patient cases in Australia² and the United States of America³, respectively. Despite the development of effective new therapies, which have led to improved outcomes, patients inevitably relapse and require further treatment. The mechanisms underlying MM initiation, therapeutic resistance and disease recurrence are complex and associated with intraclonal genetic heterogeneity⁴⁻⁷.

The advent of Next Generation Sequencing (NGS) technologies has revolutionised our understanding of the genetic heterogeneity and key driver mutations in genes associated with MM disease development. The ‘Initial genome sequencing and analysis of multiple myeloma’ in 38 patients was carried out by Chapman and colleagues in 2011, identifying the first set of significantly mutated genes believed to be drivers of MM⁵. In addition to previously reported mutant genes, such as *KRAS*, *NRAS* and *TP53*, NGS identified new unexpected candidates such as *DIS3* and *FAM46C*. Following this seminal study, 3 further large cohort studies of patients were carried out using patient samples of MM and its pre malignant stages of monoclonal gammopathy of undetermined significance (MGUS) or smouldering multiple myeloma (SMM)^{4,6,7}. These studies demonstrated intraclonal heterogeneity as a hallmark of MM, where distinct PC populations carry differing mutations, with the rise and fall in dominance of clonal populations as disease progresses. The most recurrently mutated genes identified in these studies include *KRAS*, *NRAS*, *BRAF*, *TP53*, *DIS3* and *FAM46C*, which are believed to be drivers of MM due to their recurrent nature^{4,6,7}. Moreover, intraclonal heterogeneity has been observed at both the earliest stages of disease and also at MM, where the dynamics of progression are thought to be characterised by “Darwinian” evolution. In this model, the acquisition of driver mutations is suggested to confer an improved clonal fitness (i.e. selective advantage and dominance) allowing clones to survive and progress to MM. However, as these studies relied on the comparison of genetic changes between MGUS, SMM and MM samples isolated from different individuals, it remains unknown whether the genetic heterogeneity that was present at MM was also present at the asymptomatic MGUS and SMM stages.

Intraclonal genetic heterogeneity is viewed as one of the main reasons for the disparate outcomes for MM patients. As different clonal populations within a tumour carry

differing combinations of mutations they exhibit diversified survival properties, therefore, treatment impacts the death or survival of clonal populations based on their clonal fitness. While standard treatment strategies kill off aggressive dominant clones with driver mutations, they leave indolent clones that resist the effects of treatment. Moreover, as treatment represents an unnatural selective pressure on indolent clones may cause them to mutate further and acquire driver mutations, facilitating an improved clonal fitness and subsequent dominance, leading to MM relapse. To date, few studies have investigated the changing clonal composition of MM in response to treatment using array and NGS techniques, finding substantial tumour heterogeneity with the acquisition of new mutations and resulting clonal tiding at relapse in response to therapy⁸⁻¹³.

Our current knowledge of the genomic complexity in MM has been established on the examinations of *unmatched* samples isolated from different patients at MGUS, SMM, MM and PCL, and how genetic heterogeneity is observed to be related between stages^{4-7,14-20}. However, two very small studies have investigated the tumour evolution associated with the natural history of disease transformation from SMM to MM, using paired SMM-MM samples from the same patient (n = 4 in both studies)^{17,20}. These studies found the majority of genetic changes required for MM were present from the asymptomatic stage of SMM. These initial studies have suggested that progression to MM does not involve many new mutations, where clonal progression is a key feature of transformation¹⁷. While these studies characterised the clonal heterogeneity present from the SMM stage, the subclonal evolution associated with disease progression was poorly understood.

Due to the rare nature and difficulty in collecting paired samples from treatment naïve patients when first diagnosed with MGUS/SMM and then subsequently at MM, the studies presented in this thesis constitute the first genetic analysis of the changes associated with MM transformation. Based on the matched nature of our samples, we hypothesised that analyses of MM patients in a paired setting would reveal commonly mutated genes that represent key drivers of disease progression. To investigate this hypothesis, in Chapter 2 we performed whole exome sequencing (WES) analysis of matched samples from 10 patients, who progressed from MGUS to MM (n = 5) or SMM to MM (n = 5). Similar to previous NGS studies on large cohorts of *unmatched* MGUS-SMM-MM patient samples, we confirmed clonal heterogeneity is a common feature at diagnosis. However, in contrast, we identified that the driving events involved with disease progression are more subtle than previously reported. To this end, we found a changing spectrum of acquired mutations, not

total mutational load, to be associated with MGUS/SMM to MM progression. Interestingly, slightly higher somatic single nucleotide variant (SNV) and copy number variant (CNV) frequency were identified at the asymptomatic stages of MGUS/SMM compared to MM (SNVs: median 161 versus 152 per patient, respectively; CNVs: median 70 versus 67.5 per patient, respectively). We found the RAS/MAPK pathway to be highly mutated in our study, with 40% of patients at MGUS/SMM and 70% of patients at MM harbouring mutations in *KRAS* and *NRAS*. These were identified to be mutually exclusive, consistent with previous observations that report the rare co-occurrence of mutations in these genes. Notably, we revealed that MGUS/SMM to MM progression is characterised by a prevailing model of tumour evolution defined by clonal stability, where the transformed PC subclones of MM were already present at the MGUS/SMM stage. Subclonality was evident at the earliest stages of disease, with the presence of between 5 to 11 subclones at MGUS/SMM which progress to MM with subtle changes in the degree of emergence and/or extinction of child subclonal branches. These findings suggest that patients who progress in a short time frame are already sufficiently genetically complex at MGUS/SMM to be on the threshold of transformation to MM; a process that may be driven by PC-extrinsic selective pressures imparted by the tumour microenvironment.

While Chapter 2 focused on the characterisation of the genetic architecture and subclonal evolution in PCs as a function of disease progression, Chapter 3 examined the changes in the transcriptomic and methylomic landscape associated with disease evolution. To date, little is known about the role of the transcriptomic changes associated with the progression to MM in a matched longitudinal setting. Current understanding of the MM transcriptome is derived from gene expression profiling (GEP) studies using array-based technologies, which provide a broad ranging insight into the transcriptomic landscape of MM at a single point. These insights have led to the generation of robust risk stratification models for the prognosis of patients, such as the UAMS-70²¹ and EMC-92²² gene signatures. However, these tools are not yet routinely utilised in the clinic, as further supporting evidence is required before they become commercially available for use in a clinical setting. The clinical precision of GEPs in MM prognostication is hampered by the presence of intraclonal heterogeneity throughout all stages of disease. As GEP provides a snapshot of expression patterns, it is only able to account for the dominant clones of a tumour, thus missing important information on subclonal changes such as possible driver genes and associated deregulated pathways in minor subclones. Interestingly, it has been shown that minor subclones that resist treatment are able to further mutate, gaining

selective fitness and malignant potential, leading to MM relapse¹⁰. With the advent of NGS technologies, such as RNA sequencing (RNAseq) and its progressive affordability, GEP may not represent the best methodology for detailed interrogation of the MM transcriptome. RNAseq is a highly sensitive technology, which is able to provide a deeper insight into the transcriptomic complexity of a cell as it assesses RNA transcripts with single base resolution in a high throughput manner²³. Therefore, as RNAseq provides a quantitative measure of each RNA fragment from a cell, it can accurately detect both highly and lowly expressed genes, as would be found in a heterogeneous tumour.

Gene expression is known to be regulated by epigenetic mechanisms such as DNA methylation. The rate of epigenetic change in cancers has been estimated to be orders of magnitude higher than that of genetic change, and could be a major determinant of clonal evolution²⁴. Studies investigating the methylomic changes in MM have demonstrated that disease initiation is characterised by global hypomethylation²⁵⁻²⁸. Interestingly, however, analysis of disease progression has yielded conflicting results, with some studies identifying increased hypomethylation^{25,27,28}, while another study showed decreasing hypomethylation due to progressive hypermethylation²⁶. In Chapter 3, we investigated the transcriptomic and DNA methylation changes associated with the progression of MGUS/SMM to MM using RNAseq and whole genome bisulphite sequencing (WGBS), respectively. Interestingly, there was minimal variation in gene expression between the MGUS/SMM and MM stages. There were 250 genes approaching statistical significance in differential expression testing, with the top 10 genes including: *THEMIS2*, *BTBD19*, *HBB*, *ATP8A2*, *CELSR1*, *CD69*, *TWF2*, *SLC20A1*, *ALGIL* and *SLC23A3*. Mutated genes of MM, found in WES analyses of the same patients described in Chapter 2, were expressed at low levels or not at all. In most cases, only the wild type allele of a gene harbouring heterozygous mutation was expressed. Analysis of the methylome using WGBS revealed the initiation and progression of MGUS, to SMM and MM, was associated with extreme hypomethylation. While an average of 77.5% of CpG sites were methylated in NPCs, an average of 43%, 45.6%, 41.7% of CpG sites methylated were methylated in MGUS, SMM and MM, respectively. Notably, compared to NPCs, there were 190,401 differentially methylated regions (DMRs) at MGUS/SMM, and 177,535 DMRs at MM. However, we found no DMRs between MGUS/SMM and MM. These findings indicate that PCs from the asymptomatic MGUS/SMM stages appear to be as genetically complex as the MM stage, with the majority of transcriptomic and methylomic changes occurring during the aberrant transition of NPCs to MGUS/SMM.

The initiating oncogenic events of MM are known to occur in a maturing B cell, when present within the germinal centre. The resultant PCs re-enter the BM, leading to the asymptomatic condition of MGUS. Common initiating events include *IgH* translocations and hyperdiploidy. Generation of highly specific Ig, through the somatic hypermutation process in the hypervariable region of the *IgH* locus is induced by the expression of activation-induced cytidine deaminase (AID). AID is a member of the APOBEC family of DNA/RNA editing enzymes. Genomic studies of patients have identified the enrichment of a mutational signature characterised by C>T transitions at CpG sites, which is hypothesised to be due to the aberrant activity of APOBEC enzymes^{4,29}. Subsequently, a study has shown the association of the APOBEC mutational signature with the t(14;16) and t(14;20) translocation subgroups in MM that exhibit poor prognosis³⁰.

In Chapter 4, we investigated the phenomenon of RNA editing in *SP140*, a gene recurrently mutated in human MM patient studies^{4,6,9}, in the 5TGM1 murine MM PC line. In addition to DNA, APOBECs also catalyse RNA editing, inducing C>U RNA changes, where currently APOBEC1, APOBEC3A and APOBEC3G are known to cause recoding changes³¹. Here, we identified a high impact C>T (ie. U) RNA editing change in exon 2 (c.166) of *Sp140*, resulting in an early STOP codon, and hypothesize that Apobec enzymes are the likely candidates inducing this phenotype. *In vitro* studies revealed site specific RNA editing of *Sp140* was present in the 5TGM1 cell line, with ~59% of exon 2 (c.166) edited. In addition, *Sp140* RNA editing was not unique to murine PCs, but present in other mouse cell lines to a varying degree, with exon 2 (c.166) edited at a frequency of ~57%, ~30%, ~12%, ~9% and ~9% in FDCP1, NS1, NIH-3T3, BA/F3 and RAW264.7 cells, respectively. While the exon 2 region of interest was conserved between the mouse and human genome, Sanger sequencing revealed this site was not edited in a range of SP140 expressing human MM PC lines. Moreover, in studies investigating the underlying mechanism of C>U RNA editing in 5TGM1 cells, we showed that Apobec1 and Apobec3 were not responsible in this process. This suggests there may be other novel uncharacterised enzyme candidates present in 5TGM1 PCs, which can cause dynamic RNA editing changes.

The 5TGM1 cell line and the 5TGM1-C57BL/KaLwRij model are commonly used *in vitro* and *in vivo* preclinical models of MM as they reproduce many of the features of human MM disease³²⁻³⁴. However, as recent genomic studies have demonstrated a marked genetic heterogeneity between patients, it would be valuable to ascertain which specific

clinical subtype(s) of MM are being represented by this mouse model, to better translate findings to the appropriate genetic subgroup(s) of MM patients. As we identified no corresponding RNA editing changes in human MM cell lines and patient samples, it is worth noting this *in vitro* model may replicate those MM patients with post transcriptional modifications resulting in STOP codons and truncated protein formation. Moreover, further large cohort studies of MM patients concurrently using WES and RNAseq will enable the potential to study RNA editing changes throughout MM, which may be widespread in disease.

Intratumour heterogeneity is a common feature of MM^{4-7,17}, therefore it remains intriguing why tumour cells would require further diversity through RNA editing. The answer may be the dynamic nature of post transcriptional modification, which may confer survival advantages to the tumour cell. It may be that mutations advantageous for a cancer cell in most microenvironments are hardwired in the DNA, but those advantageous only in changing environments or certain disease stages are regulated by RNA editing³¹. As such, future transcriptomic studies of patient samples should also consider RNA recoding and the impact of the tumour microenvironment and unnatural selective pressures, such as drug exposure, on cells and their dynamic survival adaptability.

The studies presented in this thesis represent a unique examination the genomic complexity and tumour evolution associated with the progression of MGUS/SMM to MM in a longitudinal nature. Our data supports a clonal stability model of tumour progression in MM. While initial *unmatched* studies of patient samples illustrated “Darwinian evolution”, with the rise and fall in dominance of clonal populations based on selective advantages, our analysis of matched samples generally demonstrates clonal cooperation, not competition, of transformed clones from MGUS/SMM to MM. Similarly, we observed minimal variation in the transcriptomic and methylomic landscape associated with progression to MM. Taken together, these studies highlight that the genomic architecture of MGUS/SMM patients could be prognostic of transformation to MM. However, it remains unknown why patients harbouring transformed PC populations at MGUS/SMM are not symptomatic. It has been noted that that both intrinsic (genetic architecture) and extrinsic (immune cells and bone marrow microenvironment) factors may regulate tumour subclones and their subsequent symptomatic evolution³⁵. The data in this thesis points to a more pronounced role of the tumour microenvironment in the development and progression of MM disease. Notably, it has previously been shown in a haematological

setting of myelodysplasia that changes in the microenvironment can create a “promalignant” niche which precedes the acquisition of any tumorigenic genetic aberrations, highlighting the concept of niche induced oncogenesis^{36,37}. The complex interactions of the tumour microenvironment with subclones can provide signals that may support tumour growth or restriction, which influences their transformation. Indeed in MM, a recent *in vivo* study has demonstrated microenvironment dependent disease progression; with xenograft of MGUS patient derived CD138⁺ bone marrow mononuclear cells into a genetically humanized MIS^(KI)TRG6 mice model exhibiting progressive growth over an 8 - 12 week period³⁸. Humanisation was achieved through a genetic knock-in of 6 genes important for innate immune cell and MM cell development, including *IL-6*, *CSF1*, *IL-3*, *CSF2*, *SIRPA* and *TPO*³⁸. This preclinical model further supports the notion that strong selective pressures are existent within the tumour microenvironment, which mediate clonal stability at asymptomatic disease, and subsequent progression to MM. Conversely, tumour subclones may modify the host-mediated growth control to influence their progression³⁵. Furthermore, interaction of MM PCs with the microenvironment has been shown to influence dormancy of tumour cells, where dormant MM PCs are resistant to standard therapies and contribute to minimal residual disease (MRD), that can be reactivated at a later time point leading to disease progression³⁹. Supporting the importance of the microenvironment, is the observation that failure to control disease long-term in the MRD setting could reflect a damaged microenvironment⁴⁰. The current standard of care at the asymptomatic stages involves monitoring patients, with no treatment options until they display evidence of progress towards symptomatic MM. To fully realise personalised medicine strategies that target “at risk patients” who would benefit from earlier therapeutic intervention, further large cohort longitudinal and spatial sequencing studies of both treatment-naïve and post-treatment patients are required to characterise the response of both clonal drivers and subclonal architecture to current MM treatments.

5.2 Towards A Cure for MM

The marked intraclonal genetic/cytogenetic heterogeneity that characterises MM is thought to be the main obstacle to finding a unifying cure for patients. As MM presents as a heterogeneous disease with a heterogeneous clinical course, it can otherwise be considered not as a single disease entity, rather a collection of monoclonal gammopathies which all share similar clinical symptoms⁴¹. Current treatments are given based on mechanistic purposes such as proteasome inhibition, immune modulation (anti-inflammatory and antineoplastic) and induced cell death⁴², with the objective of managing MM as a chronic disease. NGS has significant potential to be implemented in the clinical assessment of risk stratification and mutation specific treatment strategies for patients; however, currently it is primarily used for research purposes only. As illustrated in Chapter 2, the main hurdle surrounding the use of NGS in the clinical setting for targeted treatment is not only the presence of intraclonal genetic heterogeneity, but also the occurrence of driver mutations within subclonal populations. As such, administration of selected drug would only eliminate a subpopulation of tumour cells, resulting in only a partial response. Alternatively, NGS could be used to screen and select only patients who would respond to established targeted therapies (in a combination therapy setting) against known drivers, such as vemurafenib for activating *BRAF* V600E mutations^{42,43}. However, it should be noted that treatment targeted to mutated genes would only be effective if the mutant is expressed. A recent study by Rashid *et al.* has demonstrated that most mutated genes in MM are lowly expressed or not expressed at all⁴⁴. This finding was confirmed in the studies described in Chapter 3, where our transcriptomic analysis revealed that genetic mutations that were acquired at MM in our cohort were only expressed at low levels or not at all.

Currently, there are no approved target based treatment strategies for MM patients, however, recent clinical trials of the small molecule BCL-2 inhibitor, Venetoclax, as a monotherapy⁴⁵ or in combination⁴⁶, in the t(11;14) subgroup of MM has shown good efficacy and safety. The t(11;14) cytogenetic alteration is the most prevalent subgroup in MM (~15-20% patients)⁴⁷, and PCs from these patients are known to overexpress the anti-apoptotic factor BCL-2 which promotes tumour cell survival. Thus, drug administration facilitates activation of pro-apoptotic pathways to induce cell death of malignant PCs. In light of these findings, Venetoclax may be the first targeted treatment for a specific subgroup of genetically defined MM, thus beginning the personalised medicine revolution in MM⁴⁸.

Notably however, as demonstrated in our longitudinal investigation of progressive patients, and in studies from Zhao *et al*²⁰ and Walker *et al*¹⁷, most of the genetic complexity and subclonal architecture of MM is present at the asymptomatic stages of disease. This proposes the question as to whether MGUS should be considered the real “disease”, as it is the earliest stage of disease where subclonality and clonal heterogeneity originates, evolving under selective pressures to give rise to clonal evolution associated with development and progression to MM. *As such, should patients be treated at the asymptomatic stages to eradicate subclones for the prevention of progression and possible cure in MM?*⁴⁹ Currently, this is not routine practice as MGUS and SMM are asymptomatic conditions therefore patients remain untreated. The standard of care involves a careful “watch and wait” strategy in which patients are monitored for signs of progression to MM. The risk-benefit ratio of early treatment needs to be considered, as therapeutic strategies are highly toxic to patients, such as current induction therapies using the proteasome inhibitor bortezomib that is known to cause peripheral neuropathy in ~55-67% of patients⁵⁰⁻⁵³. Moreover, in the case of MGUS, only 20% of patients actually go on to develop MM, therefore it may be considered more beneficial to leave patients to only be monitored at this stage⁵⁴.

In the case of SMM however, there has been an urge for early treatment of patients before they progress to MM due to the armamentarium of new classes of effective drugs. SMM is an intriguing disease stage where two subsets of patients have been identified, one group exhibits indolent disease akin to MGUS, while another group displays a more aggressive disease course described as “early myeloma”⁴⁹. In 2014, the International Myeloma Working Group reclassified these “early myeloma” SMM patients as being overt MM, through the addition of criteria assessing Myeloma Defining Events (MDEs) into the clinical classification system, which is designed to look beyond just end organ damage (from CRAB features) as a marker of MM onset⁵⁵. Very recent data from NGS studies of high risk SMM presented at The American Society of Hematology Annual Meeting and Exposition 2017, have suggested that these patients are characterized by a higher mutational load (average 1.44 mutations/Mb compared to 0.73 mutations/Mb for low risk SMM patients)⁵⁶, mutations in the MAPK and NFkB pathways⁵⁶, and *MYC-IgH* structural variation, which predicts a rapid progression to MM in < 2 years⁵⁷. Notably, this high mutational load is comparable to that of the median somatic mutation rate of MM, which is observed to be 1.6 mutations/Mb⁴⁷. Indeed the first clinical trial (QuiRedex Phase III trial) investigating the early treatment of high risk SMM patients, using induction therapies

lenalidomide and dexamethasone, has shown significant survival advantage of patients treated early (vs. standard of care monitoring) with an increased median time to progression (TTP) (median not reached vs. 21 months) and overall survival (OS) (94% vs. 80%)⁵⁸. Notably, toxic effects were mainly restricted to moderate/mild side effects (grade 2 or lower). Long term follow up of patients after 6 years, demonstrated continued significant benefit to high risk SMM patients treated early with a prolonged effect on TTP and OS⁵⁹. Another recent clinical trial in intermediate and high risk SMM (CENTAURUS Phase II trial), investigating single agent activity of the new FDA approved treatment Daratumumab (Darzalex), a human anti-CD38 monoclonal antibody, has shown significant activity with approximately 50% of patients in both intermediate and long dosing schedules displaying a partial response (PR) or better and increased 12-month progression free survival (PFS) rates⁶⁰.

Although, it should be again noted that clonal heterogeneity adds an extra layer of complexity and not all patients fall into defined groups. Currently the diagnosis of MGUS/SMM is an incidental process when an individual visits the clinic for a routine blood test. Identification of asymptomatic disease is largely based on increased clonal immunoglobulin in the blood (i.e. paraproteinemia), using relatively insensitive assays such as serum protein electrophoresis³⁵. As it has been shown mutational load increases towards MM, it may be suggested that future diagnostic methods could be based on mutational burden and specific cytogenetic/mutational events, allowing patients to be stratified to a course of clinical care that is specific for that subgroup. In this sense, it may also be recommended that NGS can be implemented to assess patients at regular intervals (along current guidelines of 3 – 6 months) to track mutational load, gene expression signatures and acquisition of known driver mutations⁶¹. However, serial bone marrow biopsies are an intrusive and painful procedure (with possible secondary complications such as bleeding or infection), and thus regular NGS assessment may not be a feasible option for all patients. This is especially true in MM, which is a disease affecting the elderly, with a median age of diagnosis of 65 years old⁶². Recent studies have investigated the merits of non-invasive measures such as simple blood tests that can measure cell free DNA (cfDNA) or circulating tumour cells (CTCs) as a marker of MM disease development. Initial studies performing NGS analysis of matched peripheral blood and BM biopsies, have demonstrated the ability of blood samples to successfully capture the clonal genetic mutations and heterogeneity of MM, similar to that detected in standard BM biopsy sample analysis alone⁶³⁻⁶⁷. A high concordance in the presence of clonal somatic

NS-SNVs (in MM driver genes) and CNV mutations was identified between liquid and BM biopsy samples. These findings suggests that quantitative disease monitoring of patients in the clinic at regular intervals may be a real possibility in the near future. As MM is a blood cancer, this has significant potential for clinical applicability compared to solid tumours due to: 1) the ability of tumour PCs to be repeatedly sampled, as the original tumour is not resected; and 2) the yield of enriched tumour PCs that can be captured for analysis (solid tumours yield ~1-10 cells per/mL of blood, whereas in MM, samples can yield 10-100x this)⁶⁵. Moreover, blood biopsy has the ability to capture the spatial heterogeneity of MM, which otherwise would be impossible due to the practical challenges of sampling patients at multiple sites. Indeed recent NGS analysis of BM samples from newly diagnosed MM patients has illustrated that intraclonal heterogeneity with clonal selection may not be the only defining evolutionary feature associated with progression to MM, with the involvement of spatial heterogeneity and regional site seeding and outgrowth of advanced clones leading to progression⁶⁸. Taken together, liquid biopsies may provide improved detection of the temporal and spatial nature of acquired mutations in MM patients, while also reducing the effects of bias due to single BM site sampling, which may not always reveal the full spectrum of genetic heterogeneity.

As health technologies continue to mature, it is viewed that larger throughput sequencing studies of patients in a longitudinal, spatial and treatment response setting will be carried out in order to derive a comprehensive understanding of MM as a disease and its heterogeneity, ultimately resulting in therapeutic strategies for early intervention and cure for patients.

5.3 Conclusion

Using NGS techniques of WES, RNAseq and WGBS to analyse a rare collection of paired patient samples, the data presented in this thesis reveals a new understanding of the underlying genomic complexity and tumour evolution model involved in disease progression of MM. Here, it has been demonstrated that MGUS/SMM patients that progress, appear to be sufficiently genetically complex to be on the threshold of transformation to MM; a process that may be driven by PC extrinsic factors such as the tumour microenvironment. The existence of subclonality and clonal stability as a model of tumour evolution provides new considerations required for patients at diagnosis and the subsequent treatments that are employed. This work should underpin further longitudinal patient sample studies and the consideration of genomic and subclonal architecture with risk stratification. Ultimately, comprehensive knowledge of the underlying biology involved in MM gained from NGS studies will significantly influence our understanding of disease development, relapse and the required clinical strategies for a cure.

5.4 References

1. National Cancer Institute. Cancer Stat Facts: Myeloma. In: <https://seer.cancer.gov/statfacts/html/mulmy.html> ed.
2. Cancer Australia. Myeloma Statistics Myeloma in Australia. In: <https://myeloma-cancer.canceraustralia.gov.au/statistics> ed.
3. American Cancer Society. Key Statistics About Multiple Myeloma. In: <https://www.cancer.org/cancer/multiple-myeloma/about/key-statistics.html> ed.
4. Bolli N, Avet-Loiseau H, Wedge DC, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun.* 2014;5:2997.
5. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature.* 2011;471(7339):467-472.
6. Lohr JG, Stojanov P, Carter SL, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell.* 2014;25(1):91-101.
7. Walker BA, Boyle EM, Wardell CP, et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol.* 2015;33(33):3911-3920.
8. Egan JB, Shi CX, Tembe W, et al. Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood.* 2012;120(5):1060-1066.
9. Kortum KM, Langer C, Monge J, et al. Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann Hematol.* 2015;94(7):1205-1211.
10. Magrangeas F, Avet-Loiseau H, Gouraud W, et al. Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia.* 2013;27(2):473-481.
11. Weston-Bell N, Gibson J, John M, et al. Exome sequencing in tracking clonal evolution in multiple myeloma following therapy. *Leukemia.* 2013;27(5):1188-1191.
12. Weinhold N, Ashby C, Rasche L, et al. Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. *Blood.* 2016;128(13):1735-1744.
13. Keats JJ, Chesi M, Egan JB, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood.* 2012;120(5):1067-1076.

14. Chiecchio L, Dagrada GP, Ibrahim AH, et al. Timing of acquisition of deletion 13 in plasma cell dyscrasias is dependent on genetic context. *Haematologica*. 2009;94(12):1708-1713.
15. Lopez-Corral L, Gutierrez NC, Vidriales MB, et al. The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells. *Clin Cancer Res*. 2011;17(7):1692-1700.
16. Lopez-Corral L, Sarasquete ME, Bea S, et al. SNP-based mapping arrays reveal high genomic complexity in monoclonal gammopathies, from MGUS to myeloma status. *Leukemia*. 2012;26(12):2521-2529.
17. Walker BA, Wardell CP, Melchor L, et al. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*. 2014;28(2):384-390.
18. Walker BA, Wardell CP, Melchor L, et al. Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. *Blood*. 2012;120(5):1077-1086.
19. Weinhold N, Chavan SS, Heuck C, et al. High Risk Multiple Myeloma Demonstrates Marked Spatial Genomic Heterogeneity Between Focal Lesions and Random Bone Marrow; Implications for Targeted Therapy and Treatment Resistance. *Blood*. 2015.
20. Zhao S, Choi M, Heuck C, et al. Serial exome analysis of disease progression in premalignant gammopathies. *Leukemia*. 2014;28(7):1548-1552.
21. Shaughnessy JD, Jr., Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;109(6):2276-2284.
22. Kuiper R, Broyl A, de Knegt Y, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia*. 2012;26(11):2406-2413.
23. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
24. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313.
25. Agirre X, Castellano G, Pascual M, et al. Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res*. 2015;25(4):478-487.

26. Heuck CJ, Mehta J, Bhagat T, et al. Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis. *J Immunol.* 2013;190(6):2966-2975.
27. Salhia B, Baker A, Ahmann G, Auclair D, Fonseca R, Carpten J. DNA methylation analysis determines the high frequency of genic hypomethylation and low frequency of hypermethylation events in plasma cell tumors. *Cancer Res.* 2010;70(17):6934-6944.
28. Walker BA, Wardell CP, Chiecchio L, et al. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood.* 2011;117(2):553-562.
29. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-421.
30. Walker BA, Wardell CP, Murison A, et al. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat Commun.* 2015;6:6997.
31. Baysal BE, Sharma S, Hashemikhabir S, Janga SC. RNA Editing in Pathogenesis of Cancer. *Cancer Res.* 2017;77(14):3733-3739.
32. Hewett DR, Vandyke K, Lawrence DM, et al. DNA Barcoding Reveals Habitual Clonal Dominance of Myeloma Plasma Cells in the Bone Marrow Microenvironment. *Neoplasia.* 2017;19(12):972-981.
33. Mrozik KM, Cheong CM, Hewett D, et al. Therapeutic targeting of N-cadherin is an effective treatment for multiple myeloma. *Br J Haematol.* 2015;171(3):387-399.
34. Noll JE, Hewett DR, Williams SA, et al. SAMS1 is a tumor suppressor gene in multiple myeloma. *Neoplasia.* 2014;16(7):572-585.
35. Dhodapkar MV. MGUS to myeloma: a mysterious gammopathy of underexplored significance. *Blood.* 2016;128(23):2599-2606.
36. Ghobrial IM, Detappe A, Anderson KC, Steensma DP. The bone-marrow niche in MDS and MGUS: implications for AML and MM. *Nat Rev Clin Oncol.* 2018.
37. Raaijmakers MH, Mukherjee S, Guo S, et al. Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature.* 2010;464(7290):852-857.
38. Das R, Strowig T, Verma R, et al. Microenvironment-dependent growth of preneoplastic and malignant plasma cells in humanized mice. *Nat Med.* 2016;22(11):1351-1357.

39. Lawson MA, McDonald MM, Kovacic N, et al. Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. *Nat Commun.* 2015;6:8983.
40. Pawlyn C, Morgan GJ. Evolutionary biology of high-risk multiple myeloma. *Nat Rev Cancer.* 2017;17(9):543-556.
41. Kumar SK, Rajkumar SV, Kyle RA, et al. Multiple Myeloma. *Nature Reviews Disease Primers.* 2017;3(17046):1-20.
42. Andrulis M, Lehnert N, Capper D, et al. Targeting the BRAF V600E mutation in multiple myeloma. *Cancer Discov.* 2013;3(8):862-869.
43. Raab MS, Lehnert N, Xu J, et al. Spatially divergent clonal evolution in multiple myeloma: overcoming resistance to BRAF inhibition. *Blood.* 2016;127(17):2155-2157.
44. Rashid NU, Sperling AS, Bolli N, et al. Differential and limited expression of mutant alleles in multiple myeloma. *Blood.* 2014;124(20):3110-3117.
45. Kumar S, Kaufman JL, Gasparetto C, et al. Efficacy of venetoclax as targeted therapy for relapsed/refractory t(11;14) multiple myeloma. *Blood.* 2017;130(22):2401-2409.
46. Moreau P, Chanan-Khan A, Roberts AW, et al. Promising efficacy and acceptable safety of venetoclax plus bortezomib and dexamethasone in relapsed/refractory MM. *Blood.* 2017;130(22):2392-2400.
47. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol.* 2017;14(2):100-113.
48. Kortum KM, Einsele H. First targeted therapy in multiple myeloma. *Blood.* 2017;130(22):2359-2360.
49. Ghobrial IM, Landgren O. How I treat smoldering multiple myeloma. *Blood.* 2014;124(23):3380-3388.
50. Argyriou AA, Iconomou G, Kalofonos HP. Bortezomib-induced peripheral neuropathy in multiple myeloma: a comprehensive review of the literature. *Blood.* 2008;112(5):1593-1599.
51. Richardson PG, Weller E, Lonial S, et al. Lenalidomide, bortezomib, and dexamethasone combination therapy in patients with newly diagnosed multiple myeloma. *Blood.* 2010;116(5):679-686.

52. Roussel M, Avet-Loiseau, H., Moreau, P., Huynh, A., Benboubker, L., Hulin, C., Marit, G., Leleu, X., Pegourie, B., Fruchart, C., Caillot, D., Stoppa, A., Facon, T., Harousseau, J., & Attal, M. Frontline Therapy with Bortezomib, Lenalidomide, and Dexamethasone (VRD) Induction Followed by Autologous Stem Cell Transplantation, VRD Consolidation and Lenalidomide Maintenance In Newly Diagnosed Multiple Myeloma Patients: Primary Results of the IFM 2008 Phase II Study. *Blood*. 2010;116(21):624.
53. Voorhees PM, Laubach J, Anderson KC, Richardson PG. Peripheral neuropathy in multiple myeloma patients receiving lenalidomide, bortezomib, and dexamethasone (RVD) therapy. *Blood*. 2013;121(5):858.
54. International Myeloma Foundation. Understanding MGUS and Smoldering Multiple Myeloma. 2017.
55. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15(12):e538-548.
56. Bustoros M, Park, J., Salem, K. Z., Liu, C., Capelletti, M., Huynh, D., Tai, Y., Mouhieddine, T. H., Freeman, S., Ha, G., Rhoades, J., Reed, S., Gydush, G., Rotem, D., Perilla Glen, A., Henrick, P., Caola, A., Rivotto, B., Savell, A., Noonan, K., Reyes, K., Sacco, A., Boehner, C. J., Leblebian, H., Liu, D., Morgan, E. A., Roccaro, A. M., Paba-Prada, C., Van Allen, E., Love, C., Adalsteinsson, V., Laubach, J., Schlossman, R., Munshi, N., Richardson, P. G., Anderson, K. C., Kumar, S. K., Getz, G., Ghobrial, I. M., & Manier, S. Next Generation Sequencing Identifies Smoldering Multiple Myeloma Patients with a High Risk of Disease Progression. *Blood*. 2017;130(Suppl 1):392.
57. Keane N, Stein, C. K., Angelov, D., Tian, S., Viswanatha, D., Kumar, S. K., Dispenzieri, A., Gonzalez De La Calle, V., Misund, K., Kyle, R. A., O'Dwyer, M. E., Fonseca, R., Stewart, A. K., Braggio, E., Asmann, Y., Rajkumar, S. V., & Bergsagel, P. L. MYC Translocations Identified By Sequencing Panel in Smoldering Multiple Myeloma Strongly Predict for Rapid Progression to Multiple Myeloma. *Blood*. 2017;130(Suppl 1):393.
58. Mateos MV, Hernandez MT, Giraldo P, et al. Lenalidomide plus dexamethasone for high-risk smoldering multiple myeloma. *N Engl J Med*. 2013;369(5):438-447.
59. Mateos MV, Hernandez MT, Giraldo P, et al. Lenalidomide plus dexamethasone versus observation in patients with high-risk smoldering multiple myeloma (QuiRedex): long-term follow-up of a randomised, controlled, phase 3 trial. *Lancet Oncol*. 2016;17(8):1127-1136.

60. Hofmeister CC, Chari, A., Cohen, Y., Spencer, A., Voorhees, P. M., Estell, J., Venner, C. P., Sandhu, I., Jenner, M. W., Williams, C., Cavo, M., Van De Donk, N. W., Beksac, M., Kuppens, S., Bandekar, R., Neff, T., Heuck, C., Qi, M., Goldschmidt, H., & Landgren, O. Daratumumab Monotherapy for Patients with Intermediate or High-Risk Smoldering Multiple Myeloma (SMM): Centaurus, a Randomized, Open-Label, Multicenter Phase 2 Study. *Blood*. 2017;130(Suppl 1):510.
61. Dhodapkar MV, Sexton R, Waheed S, et al. Clinical, genomic, and imaging predictors of myeloma progression from asymptomatic monoclonal gammopathies (SWOG S0120). *Blood*. 2014;123(1):78-85.
62. Rajkumar SV. Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2016;91(7):719-734.
63. Kis O, Kaedbey R, Chow S, et al. Circulating tumour DNA sequence analysis as an alternative to multiple myeloma bone marrow aspirates. *Nat Commun*. 2017;8:15086.
64. Lohr JG, Kim S, Gould J, et al. Genetic interrogation of circulating multiple myeloma cells at single-cell resolution. *Sci Transl Med*. 2016;8(363):363ra147.
65. Manier S, Park J, Capelletti M, et al. Whole-exome sequencing of cell-free DNA and circulating tumor cells in multiple myeloma. *Nat Commun*. 2018;9(1):1691.
66. Guo G, Raje NS, Seifer C, et al. Genomic discovery and clonal tracking in multiple myeloma by cell-free DNA sequencing. *Leukemia*. 2018.
67. Mishima Y, Paiva B, Shi J, et al. The Mutational Landscape of Circulating Tumor Cells in Multiple Myeloma. *Cell Rep*. 2017;19(1):218-224.
68. Rasche L, Chavan SS, Stephens OW, et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat Commun*. 2017;8(1):268.