Next-Generation Sequencing for Clinically Viable Multiple Myeloma Genome Interrogation

By

Samuel Douglas Cutler

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
June 2021

Table of Contents

LIST OF TABLES v

LIST OF FIGURES vi

ABSTRACT vii

LIST OF ABBREVIATIONS USED viii

ACKNOWLEDGMENTS xi

CHAPTER 1 INTRODUCTION 1

1.0 PREAMBLE 1
1.1 NORMAL PLASMA CELL DEVELOPMENT 2
1.2 PLASMA CELL DISORDERS AND MULTIPLE MYELOMA SPECTRUM DIAGNOSTIC CRITERIA 3
1.3 EPIDEMIOLOGICAL PERSPECTIVE 4
1.4 OVERVIEW ON THE PATHOPHYSIOLOGY OF MULTIPLE MYELOMA 5
1.5 INITIAL DIAGNOSIS AND CLINICAL ASSESSMENT OF MULTIPLE MYELOMA 6
1.5.1 Prognostication and Response to Therapy 7
1.5.2 Radiological and Laboratory Assessments 8
1.5.3 Routine Blood Tests 9
1.5.4 Serum Protein Electrophoresis, Immunofixation and Serum Free Light Chain Testing 9
1.5.5 Bone Marrow Core Biopsy and Aspirate: Immunofixation, Flow Cytometry, and FISH Assessment 10
1.6 GENOMIC LANDSCAPE OF MULTIPLE MYELOMA 10
1.6.1 Overview 10
1.6.2 Evolving Approaches to Determine the Genomic Landscape and Relevance to Prognosis in Multiple Myeloma 11
1.6.3 Primary Cytogenetic Lesions in Multiple Myeloma 12
1.6.3.1 Copy Number Abnormalities and Hyperdiploid in Multiple Myeloma 13
1.6.3.2 Non-Hyperdiploid Multiple Myeloma 13
1.6.3.2.1 Translocations t(4;14) and t(6;14) 13
1.6.3.2.2 Translocation t(11;14) 14
1.6.3.2.3 Translocations t(14;16) and t(14;20) 15
1.6.4 Secondary Cytogenetic Lesions in MM and Their Prognostic Relevance 15
1.6.4.1 Secondary Lesions to 17p13.1: TP53 16
1.6.4.2 Secondary Lesions to Chromosome 1 Arms p and q 16
1.6.4.3 Secondary Lesions to Chromosome 13 17
1.6.4.4 Secondary Lesions to MYC 17
1.6.5 Molecular Classifications, SNPs and Indels, and Prognostic Relevance 18
1.6.5.1 High-Throughput Sequencing of Multiple Myeloma 19
1.6.5.1.1 Aberrations in MAPK Pathway Genes 20
1.6.5.1.2 Aberrations in NF-KB Pathway Genes 22
1.6.5.1.3 Aberration in DNA Repair Genes 23
1.6.5.1.4 Aberrations in Genes Controlling Cell Cycle 24
1.6.5.1.5 Aberrations in RNA Processing Genes 25
1.6.5.1.6 Other Commonly Mutated Genes in MM 25
1.7 MM THERAPEUTIC APPROACHES 26
1.7.1 Historical Therapies 26
1.7.2 Standard of Care Therapies 27
1.7.2.1 Induction Therapies 27
1.7.3 Treating High-Risk Multiple Myeloma 29
1.7.4 Precision and Modern Therapies 29

1.8 NEXT GENERATION SEQUENCING AND ANALYSIS OVERVIEW 30
1.8.1 Background *30*
1.8.1.1 Sequencing Technology Overview and Terminology 31
1.8.1.2 Considerations for NGS Application in Clinic 32
1.8.1.2.1 Single Nucleotide Variation and Indel Detection 32
1.8.1.2.2 Copy Number Variation Detection 33
1.8.1.2.3 Non-CNV Structural Variant Detection 33
1.8.1.2.4 NGS for Clinical MM Genomic Profiling 34
1.8.2 NGS Technologies for Clinical Oncology *35*
1.8.3 Somatic SNP and Indel Calling Algorithms *36*
1.8.4 CNV Calling Algorithms *38*
1.8.5 Other SV Calling Algorithms *39*
1.8.6 Variant Interpretation *41*
1.9 RATIONAL AND AIMS OF THIS STUDY 41

CHAPTER 2 METHODS 42

2.1 TUMOUR BANK SAMPLE DATA COLLECTION: 42
2.2 BONE MARROW PROCESSING 42
2.3 DNA AND RNA EXTRACTION 43
2.4 DMG26 LIBRARY PREPARATION 44
2.5 BIOINFORMATICS VERSIONS 45

CHAPTER 3 DMG26: A TARGETED SEQUENCING PANEL FOR MUTATION PROFILING TO ADDRESS GAPS IN THE PROGNOSTICATION OF MULTIPLE MYELOMA 47

3.1 ABSTRACT 47
3.2 INTRODUCTION 47
3.3 METHODS 49
3.3.1 Sample Acquisition *49*
3.3.2 DMG26 Panel Design *49*
3.3.3 Sequencing and Variant Calling *51*
3.3.4 Clinical data *51*
3.3.5 Statistical Analysis *51*
3.3.6 Availability of Data *52*
3.4 RESULTS 52
3.4.1 Cohort Description *52*
3.4.1 COMPARISON OF TRUSIGHT AND AMPLISEQ PANELS 52
3.4.2 Analysis of Variant Data *56*
3.4.3 Clinical and Prognostic Value of Mutations *56*
3.4.4 Copy Number Calling *68*
3.4.5 Correlation to Clinical Metrics and FISH Data *68*
3.5 DISCUSSION: 74
3.6 AUTHOR CONTRIBUTIONS: 78
3.7 ACKNOWLEDGEMENTS: 78

CHAPTER 4 CLINICALLY VIABLE WGS TO OUTPERFORM FISH IN SV DETECTION 79

4.1 ABSTRACT 79
4.2 INTRODUCTION 79
4.3 METHODS 81
4.3.1 Patient Sample Acquisition *81*
4.3.2 DNA Extraction and Sequencing *82*
4.3.3 Bioinformatics *82*
4.3.4 Statistical Analysis *83*
4.4 RESULTS 83

4.4.1 Cohort Selection: *83*
4.4.2 Copy Number Variant Calling and Comparison to FISH: *83*
4.4.2.1 Comparing to FISH 101
4.4.4 Comparison of Structural Variant Calling Algorithms.................... *104*
4.4.5 Structural Variant Calling at 12X and Comparison to FISH *104*
4.4.6 Structural Variant Calling with Decreasing Sequence Depth *117*
4.4.6 Combining WGS with Targeted Sequencing Better Captures Prognostic Profile *117*
4.5 DISCUSSION: 121

CHAPTER 5 DISCUSSION 125

5.1 DMG26: PANEL SEQUENCING FOR MYELOMA GENOME INTERROGATION: 125
5.2 DMG26 COMPARED TO OTHER PROPOSED CLINICAL ONCOLOGY PANELS 127
5.3 STRUCTURAL VARIANT CALLING 129
5.4 COMBINING TARGETED AND WHOLE-GENOME SEQUENCING CAPTURES HIGH-RISK DOUBLE-HIT GROUP 132
5.5 SUPERIORITY OF SEQUENCING TO FISH FOR GENOME INTERROGATION 133
5.6 LIMITATIONS 134
5.7 FUTURE DIRECTIONS 136
5.8 CONCLUSIONS 137

REFERENCES 139

APPENDICES 176

List of Tables

Table 2.1: Bioinformatic software used 46

Table 3.1: Summary of patient demographics, laboratory and molecular data, therapies, and follow-up 54

Table 3.2: Cell line data matches COSMIC reports 58

Table 3.3: Focal copy number variations 69

Table 4.1: FISH called translocations and corresponding WGS translocation calls 85

Table 4.2: FISH called CNVs and corresponding WGS CNV calls 86

Table 4.3: WGS full depth coverage 98

Table 4.4: Performance of WGS CNV calling compared to FISH 103

Table 4.5: Comparison of translocation calling algorithms at 12X-2X coverage for FISH probed translocations 106

Table 4.6: Performance of WGS translocation calling compared to FISH 114

Table 4.7: Performance of WGS translocation calling compared to FISH excluding likely FISH errors 115

Table 4.8: Performance of WGS translocation calling compared to FISH on R-ISS high-risk translocations 116

Table 4.9: Performance of subsampled WGS translocation calling compared to full-depth WGS translocation calling 118

Appendix Table 1: Sample library preparation 177

Appendix Table 2: Panel identified variants 178

Appendix Table 3: Caller for each interchromosomal translocation 187

Appendix Table 4: Filter passing intrachromosomal translocation calls at 12X-1X coverage 194

List of Figures

Figure 3.1: DMG26 genes minimally overlap with standard FISH assessed loci......50

Figure 3.2: Coverage by amplicon and exon for TruSeq and AmpliSeq......55

Figure 3.3: Oncoprint of DMG26 identified mutations against patient diagnosis, FISH risk, and focal CNVs......57

Figure 3.4: Mutation distribution per gene and per sample in our cohort......59

Figure 3.5: Hazard ratio of gene mutation status......60

Figure 3.6: Concordance maximized at 20% variant allele frequency......61

Figure 3.7: Hazard ratio by clonal gene mutational status......62

Figure 3.8: High-severity mutations in panel-targeted genes significantly impact PFS......63

Figure 3.9: Figure 4: FATHMM-XF predicted pathogenic variants in panel-targeted genes significantly impacts PFS......65

Figure 3.10: Mutational hazard by prediction algorithms, ClinVar, and COSMIC......66

Figure 3.11: High-risk mutations significantly correlate with reduced PFS in both total cohort and diagnostic subgroups......67

Figure 3.12: Panel called focal-CNVs enhance DMG26 risk stratification......70

Figure 3.13: DMG26 risk markers are independent and significant prognostic markers......71

Figure 3.14: FISH 1q risk......73

Figure 4.1: WGS captures more CNVs than FISH......97

Figure 4.2: Number of CNVs per chromosome per patient......99

Figure 4.3: WGS captures a wide range of CNV sizes and abundances across samples......100

Figure 4.4: WGS CNV calls are consistent with FISH CNV......102

Figure 4.5: GRIDSS calls many more translocations than other callers......105

Figure 4.6: Translocation callers capture a diverse set of lesions......107

Figure 4.7: WGS identifies more translocations than FISH......110

Figure 4.8: WGS identifies a more complex translocation landscape than FISH......111

Figure 4.9: WGS is sensitive and specific compered to FISH......119

Figure 4.10: High-Risk double-hit group captured with combination of WGS and targeted sequencing......120

Appendix Figure 1: Cohort mutational, laboratory, and clinical landscape......176

Abstract

Multiple Myeloma (MM) is a severe and common malignancy which presents clinically with a highly aberrated genome, with mutations ranging in size from whole chromosome events to single nucleotide variants. The large chromosomal events, including translocations at the IgH locus and copy number variants, have well defined clinical correlates, whereas the clinical relevance of small-scale lesions is only recently coming to light. Accordingly, the current clinical standard for myeloma genome interrogation, fluorescent *in situ* hybridization (FISH), only reports on the large-scale lesions, leaving gaps in the prognostication of MM. Herein, we explore the use of next-generation sequencing (NGS) technologies for clinical assessment of the MM genome. We developed a targeted sequencing panel that, for the first time, robustly captures significant clinical associations with gene mutational status; it is an independent biomarker and outperforms FISH based prognostication. We also developed a whole genome sequencing (WGS) pipeline which outperforms FISH in the capture of structural variants down to 10X coverage. Taken together, we described a next generation sequencing approach for MM patients which is clinically viable and superior to FISH.

List of Abbreviations Used

MM	Multiple Myeloma
NGS	Next Generation Sequencing
WGS	Whole genome sequencing
CNV	Copy number variant
SNV	Single nucleotide variant
MGUS	Monoclonal gammopathy of Undetermined Significance
SMM	Smoldering multiple myeloma
nsMM	Non secretory multiple myeloma
PCL	Plasma cell leukemia
POEMS	Polyneuropathy, Organomegaly, Endocrinopathy, Monoclonal protein, Skin changes
CT	Computed tomography
PET	Positron emitted tomography
IMWG	International myeloma working group
MDE	Myeloma defining event
MRI	Magneti Resonance Imaging
MDSC	Myeloid derived suppressor cells
BMSC	Bone marrow stromal cells
ICAM-1	Intercellular adhesion moleculare-1
VCAM-1	Vascular cell adhesions molecular-1
IL-6	Interlukin-6
VEGF	Vascular endothelial growth factor
DKK1	Dickkopf-1
Ig	Immunoglobulin
M-protein	Myeloma-protein
ISS	International staging system
B2M	Beta-2 microglobulin

R-ISS	Revised-international staging system
LDH	Lactate dehydrogenase
FISH	Fluorescent *in situ* hybridization
sCR	Stringent complete response
CR	Complete response
VGPR	Very good partial response
PR	Partial response
MR	Minimal Response
CBC	Complete blood count
ALP	Alkaline phosphatase
SPEP	Serum protein electrophoresis
MRD	Minimal residual disease
WES	Whole exome sequencing
GEP	Gene expression profiles
SV	Structural variant
CSR	Class switch recombination
VDJ	Variant, diversity, and joining regions
Rb1	Retinoblastoma 1
mAbs	Monoclonal antibodies
CN-LOH	Copy neutral-loss of heterozygosity
NF-κB	nuclear factor kappa-light-chain-enhancer
MAPK	Mitogen activated protein kinase
PR	Proliferation group
LB	Low bone disease group
MS	MMSET group
HY	Hyperdiploidy group
CD-1	Cyclin D1 dysregulated group
CD-2	Cyclin D2 dysregulated group
MF	MAF or MAFB dysregulated group
ER	Endoplasmic reticulum

FDA	Food and Drug administration
CyBorD	Cyclophosphamide, bortezomib, dexamethasone
VRd	Bortezomib, lenalidomide, dexamethasone
VD	Bortezomib, dexamethasone
PFS	Progression-free survival
OS	Overall survival
MMRF	Multiple Myeloma Research Foundation
TSO500	TruSight Oncology 500
PPV	Positive predictive value
NPV	Negative predictive value
BAM	Binary alignment map
EDTA	Ethylenediaminetetraacetic acid
ACK	Ammonium chloride lysis buffer
MRN	Medical record number
LDH	Lactate dehydrogenase
AST	Aspartate aminotransferase
BM	Bone Marrow
VAF	Variant allele frequency
NPV	Negative Predictive Value
DMG26	Dalhousie Myeloma Genomics 26

Acknowledgments

Firstly, I would like to thank my supervisors, Dr. Daniel Gaston and Dr. Manal Elnenaei. Thank you for your enormous efforts in helping this project along, and mentoring me. Dan, thank you for meeting me twice a week – keeping me on track, exploring new directions, and making the experience of research calm, fun, and social. Manal, thank you for continual and enthusiastic support from my summer studentships onward, willingness and excitement to expand the project and loop in collaborators, and keen guidance and support in science and life.

I would additionally like to thank my committee, Dr. Wenda Greer, Dr. Shashi Gujar, Dr. Nicholas Forward, and Dr. Mohamed Abou El Hassan. Thank you for your guidance, support, and encouragement along the way. Additionally, I would like to acknowledge Marissa Goudie and Dr. Julie Wagner for their help in sample preparation and sequencing. Prominently, I thank Dr. Andrea Thoni for her large help in patient data acquisition, organization and interpretation.

Finally, I thank my friends and family for their food deliveries, support, listening to my rambles, and for the fun distractions you offered up along the way.

Chapter 1 Introduction

1.0 Preamble

Multiple myeloma (MM) is a malignancy of post-germinal centre plasma cells that is associated with significant mortality and morbidity.[1] It is initiated by invasion of the bone marrow by a transformed plasma cell clone which disrupts the homeostatic processes of bone maintenance and hematopoiesis occurring therein. As the disease progresses, symptoms become systemic, with renal failure, liver damage, bone fragility, and anemia all being common presentations.[1,2] Importantly, the cancer is the second commonest hematological malignancy, and incurable, thereby posing prominent need for better interventions.[3–5]

A remarkable feature of this malignancy is the heterogeneity of genomic lesions observed within MM, at both the intra- and inter-tumour level.[6–16] These genetic lesions range in size and type, including copy number variants (CNVs) afflicting whole chromosomes to focal regions, translocations, insertions and deletions (indels), and single nucleotide variants (SNVs).[9,13,16] The large scale CNVs and translocations at the IgH and MYC locus have long been observed in this cancer, and consequently have well described clinical implications.[13,17–19] Smaller lesions, including SNVs and Indels have only more recently been described, and hence have less well defined, but nonetheless important, clinical correlates.[9,13,16,20–23] Accordingly, part of the clinical work-up of MM is genomic assessment for a particular set of CNVs and translocations to designate an individual's risk.[19,24,25] Assessment for these lesions has historically been performed via fluorescent *in situ* hybridization (FISH), which is limited in resolution to detect small lesions, only assesses a small portion of the genome for which the assayer has validated probes, and is liable to err.[26,27] This has important ramifications for the application of precision medicine, as therapies may be indicated for by small lesions bellow FISH's resolving power such as SNVs and indels, or by infrequent lesions otherwise capturable by FISH but for which probes were not used.

This study seeks to address gaps in the clinical genomic assessment of Multiple Myeloma through a two-pronged approach: 1) using targeted sequencing to capture SNVs and indels in genes frequently mutated in MM and determine their clinical and prognostic significance, 2) using a clinically and economically viable whole-genome sequencing (WGS) pipeline to describe the profile of CNVs and translocations throughout the genome in a more comprehensive and accurate manner than FISH. This will allow for better prognostication of patients and improve therapeutic assignment in light of precision therapies.

1.1 Normal Plasma Cell Development

Plasma cells, the antibody producing cells of the adaptive immune system, are terminally differentiated B cells. B cells begin first in the bone marrow as early pro-B cells which arise from hematopoietic stem cells and are committed to the B cell lineage by the activity of several transcription factors, including P.U1 and PAX5.[28,29] During the pro stage, the Immunoglobulin Heavy Chain (IgH) locus undergoes rearrangement of the diversity (D) and joining (J) loci, and subsequent rearrangement of the variable region with the priorly rearranged DJ locus.[30,31] If these rearrangements are productive, the heavy chain is expressed on the cell surface with an invariant surrogate light-chain in a complex termed the pre-B receptor.[31,32] Cell surface expression of the pre-B receptor defines the pre-B cell developmental stage and initiates an intracellular signaling cascade that 1) prevents apoptosis, 2) drives proliferation, and 3) stops rearrangement at the IgH locus.[31,32]

During the pre-B cell stage, rearrangement of the V and J segments at the IgK or IgL locus occurs.[33] If successful, an IgM immunoglobulin is expressed on the cell surface, marking the transition to an immature B-cell.[33] Immature B-cells then undergo negative selection within the bone marrow as they are tested for autoreactivity.[34] If the immature B cell's immunoglobulin does not bind to self-antigen, or so weakly binds self-antigen that receptor crosslinking does not occur, it exits the bone marrow and migrates to the spleen.[34] If the immature B cell's immunoglobulin is self-reactive, it will undergo receptor editing of the light-chain until the immunoglobulin is no longer self-reactive, or undergo apoptosis.[34]

Once in the spleen, the chemokine CXCL13, produced by follicular dendritic cells, binds to CXCR5 on B cells and provides a positive chemotaxis signal that directs B cells through the splenic T cell zone into the follicular zone.[32] Within the follicle, additional rounds of negative selection and the final stages of antigen independent maturation occur, which primarily give rise to follicular B cells and some marginal-zone B cells.[34]

Plasma cells subsequently develop from both follicular and marginal-zone B cells following antigen exposure.[32] Transcription factors IRF4, BLIMP1, and XBP1 drive a transcriptional profile that differentiates plasma cells from B cells and facilitates expression of immunoglobulin.[32] Typically, plasma cells reside in the bone marrow, and represent 0.25% of the cellularity therein.[35]. Dysregulation of MAPK, JAK-STAT, and NF-κB is necessary for the transformation of a plasma cell into a myeloma cell through downregulation of B cell specific transcription factors BCL-6 and PAX5 that may block proliferation.[35,36]

1.2 Plasma Cell Disorders and Multiple Myeloma Spectrum Diagnostic Criteria

Plasma cell disorders are a broad range of conditions arising from the abnormal and excessive proliferation of a plasma cell into a large clonal population.[37,38] Several distinct disease entities exist within this group, including non-IgM monoclonal gammopathy of undetermined significance (MGUS), IgM MGUS, light-chain MGUS, smoldering multiple myeloma (SMM), multiple myeloma (MM), non-secretory multiple myeloma (nsMM), plasma cell leukemia (PCL), solitary bone and extramedullary plasmacytoma, POEMS, amyloid light-chain amyloidosis, and Waldenstrum's macroglobulinemia.[37] Excluding extramedullary plasmacytoma, in which the plasma cell clone infiltrates soft tissue, all other plasma cell disorders are characterized by invasion of the bone marrow by this clone to varying depths and extent of spread.[39] Also common to all plasma cell disorders (apart from nsMM), is the presence of a serum monoclonal immunoglobulin or light-chain subunit, produced by the clonal plasma cell population.[40,41] Invasion of the bone marrow by plasma cells disrupts the normal homeostatic hematopoietic processes otherwise occurring. As the clonal population expands, this disruption becomes more profound, and gives rise to classic myeloma symptoms captured within the pneumonic 'CRAB': Hypercalcemia (>2.75 mmol/L) secondary to plasma cell driven osteoclast activation, and subsequent bone resorption; renal failure (creatinine >177mol/L) secondary to precipitation of light chains in the tubules leading to cast nephropathy; anemia (hemoglobin <100g/L) secondary to disruption of hematopoiesis; osteolytic lesions secondary to degradation of boney tissue, demonstrated on skeletal radiography, computed tomography (CT), or positron emission tomography-computed tomography (PET-CT).[37] The International Myeloma Working Group (IMWG) provides diagnostic criteria for MGUS, SMM, and MM which considers these CRAB criteria (that are attributable to the plasma cell disorder) as well as myeloma defining events (MDEs).[37] These MDEs comprise the following: 60% or greater clonal plasma cells in bone marrow, serum involved over uninvolved free light chain ratio greater than 100 and an involved free light chain quantitation of at least 100 mg/L, and more than one focal lesion of 5mm or greater on magnetic resonance imaging (MRI).[37] At least one CRAB criteria or MDE is required for a diagnosis of MM to be made, while PCL also requires the presence of more than 2x10^9 plasma cell/L in the blood. MGUS and SMM are believed to be requisite pre-clinical stages to MM, and PCL is a rare and severe clinical sequela of MM.[37,42,43] IgM MGUS almost exclusively progresses to Waldenstrum's macroglobulinemia, and is hence a distinct entity from the myeloma continuum.[44]

1.3 Epidemiological Perspective

MM is currently the second most common hematological malignancy in North America accounting for about 1% of all cancer diagnoses and 10% of all blood cancer diagnoses.[4,5] The age standardized incidence in high-income North America is 5.2 per 100,000, and the global incidence has increased 2.26 fold between 1990 and 2016.[4] In Canada, the incidence of MM increased by 0.92 per million people per year between 1992 and 2010, and the average incidence over this period was 54.29 per million Canadians.[45] Importantly, the incidence of MM within Canada has continued to increase, and between 2011 and 2015 the average incidence in Canada was 72.9 per million per year.[45] MM is nearly twice as prevalent in African-Americans than Caucasians and is about 1.5 times more prevalent in males than females.[46,47] MGUS is more common than MM with an overall incidence being reported between 0.05 and 6.1 per hundred people; though this increases to 9% in those above the age of 80.[48,49] SMM is not well characterized epidemiologically consequent to a paucity of data, though in the U.S.A, incidence is estimated at 0.9 per 100,000 individuals.[50] MGUS and SMM have distinct risk profiles for progression to MM.[51,52] MGUS has relatively low risk of progressing to overt MM at about 1% per year.[52] SMM within the first 5 years post diagnosis has a 10% per year chance of progression; in the subsequent 5 years it has a 3% chance per year, and each year thereafter it matches MGUS risk at 1% per year.[51] Importantly, the current 5-year survival rate for MM rarely exceeds 50%, thereby highlighting the urgent need for further research to improve clinical outcomes.[45]

1.4 Overview on the Pathophysiology of Multiple Myeloma

Whether or not MGUS or SMM are clinically identified, they are believed to precede every case of MM, wherein, a clonal population of transformed post-germinal-centre plasma cells invade the bone marrow compartment thereby initiating the dyscrasia.[1,43] Within the bone marrow niche, plasma cells form tightly and complexly interacting cellular and chemical networks that are central to the development of MM, and its persistence against therapy.[35,53–56] The rare cases that progress to PCL represent a striking shift in disease biology, wherein plasma cells depart from the bone marrow niche and present in the peripheral blood; having developed autonomy from the characteristics of the niche on which the malignancy was previously reliant.[57]

Within the bone marrow, an immunosuppressed phenotype emerges through cross talk between MM cells and bone marrow resident cells.[58] Briefly, myeloid derived suppressor cells (MDSCs), bone marrow stromal cells (BMSCs), osteoclasts, and osteoblasts are prominent aspects in this process, and central to progression from MGUS to MM.[59,60] MM cells drive proliferation of MDSCs, which in turn impede T-cell responses to the invading clone; the abundance of MDSCs in the bone marrow of MM patients is associated with prognosis of the disease.[61–64] Adhesion of MM to BMSCs is crucial for MM cell survival, proliferation, and chemo-resistance.[60] Through intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1), beta-1 integrin, and beta-2 integrin mediated interactions, MM cells induce interlukin-6 (IL-6) production by BMSCs, which subsequently drive MM cell release of vascular endothelial growth factor (VEGF), thereby driving MM cell growth through NF-κB mediated signaling.[55] Degradation of boney tissue is a hallmark of myeloma and is driven by an imbalance in osteoclast and osteoblast activity consequent to crosstalk with MM cells.[64] MM cells impede osteoblastic activity through Dickkopf-1(DKK1) mediated activation of Wnt-signaling in osteoblasts; additionally, osteoblasts can produce IL-6 and hence may augment clonal development similar to BMSCs.[64,65] MM cells drive osteoclasts through IL-6 and RANK-L mediated signaling, which both augment osteoclastic action and induce anti-apoptotic signaling in osteoclasts.

Modulation of the bone marrow niche by a transformed clone of plasma cells is crucial for the advancement of disease burden and acquisition of malignant characteristics, and underpins the progression of MGUS to MM.

1.5 Initial Diagnosis and Clinical Assessment of Multiple Myeloma

Immunoglobulins (Ig) produced by myeloma cells are usually made up of one type of heavy chain (G, A, M, D or E) and one type of light chain (kappa or lambda). The most common type of MM is IgG kappa (34%) followed by IgG lambda (18%) then IgA (kappa or lambda) (21%); IgM (0.5%), IgD (2%) and IgE (0.01%) are rarer MM subtypes.[66–68] MM cells producing light chain M-protein only (light-chain MM) occur in about 16% of patients, while no secretion of any myeloma(M)-protein (ns-MM) comprise only 7% of patients.[67] MGUS and SMM are most frequently diagnosed after a M-protein band is incidentally identified on serum protein electrophoresis in the routine diagnostic laboratory.[69] The patient is then followed up for progression to MM by regular quantitation of the M-protein at set intervals, and when indicated, by laboratory and radiological assessment for CRAB features and MDEs.[38] As the originating plasma cell clone expands, the quantity of monoclonal protein (M-protein or paraprotein) produced imperfectly parallels the clonal size, hence is useful as a proximal measure of disease burden.

After monoclonal protein, anemia it the most common presenting feature of MM, occuring in about 73% of patients, while bone pain is the third most commonly satisfied diagnostic criteria of MM, at 58%.[67] Renal failure presents in about 48% of patients, and the least commonly satisfied CRAB criterion is hypercalcemia at 28% of patients.[67]

Disruption of the hematopoietic process increases proportionally with the increase in plasma cell burden, however, anemia in MM can also be secondary to renal failure, thereby obfuscating the relation between anemic levels and disease burden.[70]

1.5.1 Prognostication and Response to Therapy

The International Myeloma Working Group provides guidelines by which MM may be prognosticated.[24,71] In the first iteration of their risk stratification algorithm, the International Staging System (ISS), the malignancy was trichotomized into risk groups dependent on serum albumin and beta-2 microglobulin (B2M) levels.[71] Stage I, the lowest risk category with a median survival of 62 months, required albumin to be 35 g/L and B2M to be less than 3.5 mg/L (297 nmol/L).[71] Stage III, the highest risk category with a median survival of 29 months, required B2M to be greater than 5.5 mg/L (467 nmol/L) irrespective of albumin levels.[71] Stage II has a median survival of 42 months, and comprises patients who do not meet stage I nor III criteria.[71] This staging scheme, however, performed sub-optimally with a relatively low concordance of 0.67, subsequently, in 2015, the revised-ISS (R-ISS) was published, which has a significantly higher concordance at 0.77.[24,72] This iteration is the present standard and incorporates serum lactate dehydrogenase (LDH) and fluorescent *in situ* hybridization (FISH) assessed cytogenetic lesions into the algorithm.[24] High-risk FISH markers include del(17p), translocations t(4;14), and t(14;16).[24] R-ISS stage I, the lowest risk category with 5-year survival of 82%, requires ISS stage I, absence of high-risk cytogenetic lesions by FISH, and LDH less than the upper limit of normal.[24] R-ISS stage III, the highest risk designation with 5-year survival of 42%, requires ISS stage III and either high-risk cytogenetic lesions by FISH or LDH above the upper limit of normal.[24] R-ISS stage II, with a 5-year survival rate of 62%, is assigned to patients not meeting R-ISS stage I or III criteria.[24] Besides the R-ISS, other prognostic schemes predicated on FISH have been put forth, including the mSMART panel from the Mayo clinic, and recently, the Prognostic Index[19,25]. Similar to the R-ISS, the mSMART panel considers lesions del(17p), t(4;14), t(14;16), t(14;20), or gain 1q to be high risk lesions in a binary fashion.[25] The prognostic index considers the sum of risk contributed by the presence of t(4;14), del(17p), trisomy 5, trisomy 21, 1q gain, and del(1p32) in a fashion weighted by their Cox hazard ratio as assessed on a large cohort in a multivariate assessment against each other.[19] Notably, within the prognostic index, trisomy 5 has a negative weighting, and is thus a positive prognostic indicator; this is the only known positive prognostic marker for this malignancy.[19]

Categories of response to therapy have also been defined and standardized by the IMWG.[73] Response categories include stringent complete response (sCR), complete response (CR), very good partial response (VGPR), partial response (PR), minimal response (MR), stable disease, progressive disease, relapse, and relapse from CR.[73]
PR requires greater than 50% reduction in serum M-protein, greater than 50% decrease in difference between involved and uninvolved serum free light chain quantity, or greater than 50% reduction in bone marrow plasma cell infiltration if the infiltrate was greater than 30% at diagnosis; if a plasmacytoma was present at diagnosis, it must have reduced in size by at least 50%.[73] VGPR requires greater than 90% reduction in serum M-protein, or that M-protein be detectable on immunofixation but not electrophoresis.[73] CR requires negative immunofixation, absence of plasmacytoma, and bone marrow infiltrate be less than 5%.[73] sCR response require CR in addition to a normal free light chain ratio, and no clonal plasma cell infiltrate in the bone marrow on immunohistochemistry or immunofluorescence.[73] Progressive disease requires a greater than 25% increase in serum-M protein or plasma cell infiltration, increase in the size or abundance of lytic lesions, or new development of hypercalcemia.[73] Relapse requires a direct indication of increasing CRAB features. Relapse from CR requires the reappearance of serum-M protein on electrophoresis or immunofixation, plasma cell infiltration in the bone marrow > 5%, or appearance of any sign of progression after achieving complete response.[73]

1.5.2 Radiological and Laboratory Assessments

To facilitate IMWG adherent diagnostics, prognostics and response to therapy tracking, MM patients at diagnosis and set intervals, must have blood drawn for laboratory tests, a bone marrow (BM) core biopsy and aspirate taken, and radiological skeletal assessment performed.[24,37,73–75] Examination of the BM core biopsy is performed to determine the plasma cell burden and the BM aspirate is subjected to flow-cytometric assessment of the clonal kappa and lambda light chains as well as cytogenetic signature via FISH, at least at the point of diagnosis.

1.5.3 Routine Blood Tests

Complete blood count (CBC) informs on the number and quality of the cellular components of blood, as well as the hemoglobin level and red cell indices that are essential to determine the presence and type of anemia. Though other CBC data is not a parameter considered by the IMWG, both MM and MM therapies can cause significant damage to the bone marrow compartment and can thereby give rise to clinically significant cytopenias (low cell counts), that require tracking.[24,37,73,76–78] Hence, CBC is an integral part of a patient's diagnostic work up, and as a follow-up assessment.[74,75]

Apart from the CBC (that includes hemoglobin), other routine blood tests include assessments for serum calcium, albumin, LDH, creatinine, B2M, and alkaline phosphatase (ALP). Results of these provide information on the systemic state of a patient, as well as allude to specific organ –liver, kidney, or bone marrow– deficiencies or damage in a patient. Calcium, hemoglobin, and creatinine are diagnostic parameters relating to hypercalcemia, anemia, and renal failure, respectively; thereby constituting three of four CRAB criteria.[37] LDH, ALP, and albumin can be assessed as indicators of liver damage and function.[73]

1.5.4 Serum Protein Electrophoresis, Immunofixation and Serum Free Light Chain Testing

Serum protein electrophoresis (SPEP) separates serum proteins in a size and charge dependent manner on an agarose gel. The subsequent fractionation and pattern of SPEP is informative for the diagnosing of several conditions, the most notable being plasma cell disorders.[79,80] Serum proteins broadly migrate in five regions: from the positive electrode to negative are albumin, alpha 1, alpha 2, beta, and gamma.[79–81] In a SPEP on a healthy individual, albumin will constitute the highest peak.[81] In MM, a large (usually single) spike is present, most frequently in the gamma region, constituted by the large quantity of clonal immunoglobulin produced by the clonal plasma cell population.[81] Quantitation of this spike is informative for assessments of disease burden, response to therapy, and minimal residual disease status.[24,73,81] Immunofixation electrophoresis is performed concurrent to SPEP where a lane is subsequently stained for each of the heavy-chain constant regions (routinely for IgG, IgA, IgM; then for IgD and IgE if no band for these are visualized), and light-chain type (kappa, lambda) to determine the clonal isotype.[81] If none of the heavy chains are present- staining for free kappa and lambda light chains is performed to confirm that it is a light chain MM.[81] Measurement of serum free light chain quantities and free light chain ratio (kappa/lambda) is also an essential aspect of the diagnosis and monitoring of MM, particularly in cases of light chain MM.[24,73,81]

1.5.5 Bone Marrow Core Biopsy and Aspirate: Immunofixation, Flow Cytometry, and FISH Assessment

A bone marrow aspirate and biopsy are taken from the anterior superior iliac crest at time of diagnosis as part of the standard clinical workup, as well as at clinical checkpoints for therapeutic response, relapse, or progression.[24,37,73–75] An aspirate captures the liquid portion of the bone marrow, while a core biopsy captures solid contents of the spongy area of the bone marrow. To assess plasma cell infiltration of the bone marrow, independent measurements are made by visual microscopic inspection of the core biopsy and aspirate, and via flow cytometry. These assessments may sometimes be discordant, with no measure independently correlating with therapeutic response.[82–84] Considering the highest estimation by any method does, however, significantly correlate with therapeutic response.[84] Importantly, highly sensitive next-generation flow cytometry facilitates assessment of minimal residual disease (MRD) down to 1 cancer cell in 10^6 normal cells; flow-MRD negativity following therapy is a strikingly positive prognostic factor with a hazard ratio of 0.18 for progression free survival, and 0.12 for overall survival.[85] FISH is also performed on fixed cell pellets from the bone marrow for cytogenetic assessment utilizing probes targeted to chromosomal locations of current interest.

1.6 Genomic Landscape of Multiple Myeloma

1.6.1 Overview

MM is remarkable for the abundant and heterogenous landscape of genetic alterations found within it.[9,12,13,15,86–90] Classically, the alterations comprising this landscape have been described as primary and secondary events.[6,91] The former are large, chromosomal scale alterations which accumulate prior to, or during MGUS, and were previously believed to persist throughout the course of disease in a stable and clonal nature.[6] The static nature of clonal cytogenetic lesions has since been disputed by several serial sequencing studies.[23,92,93] Primary lesions broadly dichotomize patients into hyperdiploid and non-hyperdiploid cases, each being mutually exclusive and representing about 50% of cases.[8] Within hyperdiploid cases, trisomies of chromosomes 3, 5, 7, 9, 11, 13, 15, 19, and 21 accumulate.[8] Within non-hyperdilpoid cases, translocation at the IgH locus of chromosome 14 juxtapose the immunoglobulin promoter, which is highly active in plasma cells, next to oncogenes on partner chromosomes 4 (*FGFR3*), 6 (*CCND3*), 11 (*CCND1*), 16 (*MAF*), or 20 (*MAFB*), thereby driving their expression.[8] Primary lesions have been well described in MM, with reports of their prevalence and pattern being published as early as 1966.[17] Secondary lesions, which accumulate throughout disease development, are comprised of variants ranging in size from cytogenetic aberrations – del(13), del(13q14.1), del(17p), del(17p13.1), gain(1q), del(1p), and *MYC* sep – to SNVs and indels.[8,9,13,16]

1.6.2 Evolving Approaches to Determine the Genomic Landscape and Relevance to Prognosis in Multiple Myeloma

Studies that interrogate the landscape of SNVs and insertions and deletions (indels) in MM are relatively novel and evolving, as large-scale whole-exome sequencing (WES) and WGS, on which they are heavily reliant, have only recently become feasible.[9,13,15,16,92,94] These studies have highlighted a mutational profile far more abundant and heterogeneous, both intra- and inter-tumoral, than previously understood by standard cytogenetic assessments.[9,13,15,16,92,94] Subsequently, quite distinct MM subgroups with differing clinical courses have been suggested, redefining MM genetic categorization beyond the classic hyperdiploid and non-hyperdiploid.[9,13,16,22,92,95,96]

However, these findings have yet to be acknowledged in the clinical arena as WGS and WES are too costly for standard clinical laboratories, and FISH, which remains the genomic gold-standard for R-ISS staging and prognostication of myeloma, has a resolution too low to resolve the SNVs and indels which define modern disease categories. Hence, current risk stratifications exclusively consider FISH targeted cytogenetic lesions, with the exception of the Mayo's mSMART scheme, which may consider gene expression profiles (GEPs).[19,24,25] GEPs were first proposed in 2007 as alternatives to FISH for risk stratification, which preceded the WGS and WES studies that described this malignancy's mutational landscape.[97] Gene expression microarray data established significant survival difference between patients according to their gene expression, and GEPs were designed based on genes whose expression were most determinant of cluster assignment. Various profiles have been published, prominent among them being the GEP70 by Shaugnessy et al. and the Proliferation Index by the Intergroupe Francophone du Myelome (IFM).[97,98] Strikingly, although many of these schemes were developed using the same microarrays, there was minimal overlap between each scheme, suggesting that co-regulatory expression networks confound the analysis and obfuscate the risk associated with an individual gene's expression.[99] Consistently however, many genes whose low- or high-expression were determined to be risk-markers in these GEPs mapped to chromosome arm 1p or 1q, whose deletion or amplification, respectively, have well known associations with high-risk disease.[19,97] Despite the known impact of these cytogenetic lesions as a whole, the paucity of overlap between these GEPs indicates that the true 'target' genes are still unknown. Nonetheless, risk stratification by GEPs is highly performant, achieving greater concordance than FISH and the R-ISS, and informing on therapeutic appropriateness.[100–103] GEPs have not become standard for clinical assessments due to cost and technical challenges which impede fidelity and utility when assessing one or a few samples.

There hence remains a need for a clinically viable approach to profile the genome of MM with resolution higher than FISH.

1.6.3 Primary Cytogenetic Lesions in Multiple Myeloma

Cytogenetic abnormalities are present in nearly all myeloma patients.[104] Classical primary lesions have been mainly characterized by FISH and are: translocations t(4;14), t(6;14); t(11;14), t(14;16), t(14;20); trisomies +3, +5, +7, +9, +11, +15, +19, +21; and monosomies -13, -14, -16, -22. Although these variants have long been known, their contribution to risk well defined, and their relation to each other well described, the genes targeted by many of these alterations remain obscure.[105–107] FISH probes to assess all of the above lesions are available within the Mayo Clinic's plasma cell proliferative disorder panel, either as initial assessments (del(17p), del(1q), or IgH sep), or reflexively included following the initial findings.

1.6.3.1 Copy Number Abnormalities and Hyperdiploid in Multiple Myeloma

In MM, aneuploidy arises from primary events and is categorized into four groups: hypodiploid with less than 44 chromosomes (4%), pseudodiploid with 45-48 chromosomes (36%), hyperdiploid with at least 3 trisomies (53%), and near-tetraploid with 75 or more chromosomes (7%).[108–110] Acquisition of chromosome copy number abnormalities occurs sequentially; this was recently elucidated through WGS and opposes the historical theory of a single initial metaphase catastrophe giving rise to the copy number profile of the clonal population.[23,92,93] The abundance of CNVs in MM implicates genomic instability as a malignant feature, which is corroborated by WGS studies identifying chromothripsis in MM genomes.[111,112] Generally, hyperdiploidy and near-tetraploidy are prognostically favourable, both in terms of progression-free and overall survival compared to both hypodiploidy and non-hyperdiploidy.[109] Trisomies 3, 5, 13, and 21 are the only specific full-chromosome additions with significant clinical impact; notably, trisomy 3 and 5 are the only known favorable prognostic markers while trisomy 13 and 21 are prognostically unfavourable.[19,113–115]

1.6.3.2 Non-Hyperdiploid Multiple Myeloma

Non-hyperdiploid MM (NHMM), comprised of hypodiploid, pseudodiploid, and near-tetraploid MM, is characterized in 85% of patients as harbouring a canonical IgH translocation as a primary lesion, and is in general more afflicted by structural variants (SV).[116] Five mechanisms are known to contribute to the formation of breakpoints at the IgH locus: class switch recombination (CSR); homologous recombination; somatic hypermutation; aberrant recombination of the variant, diversity, and joining regions (VDJ) at the immunoglobulin locus; and receptor revision.[117]

1.6.3.2.1 Translocations t(4;14) and t(6;14)

Translocations t(4;14) and t(6;14) occur in 15% and 4% of patients, respectively; both commonly result from recombination error, with the chromosome 14 breakpoint occurring within the switch region.[117,118] On chromosome 4, breakpoints are proximal to both *FGFR3* and *MMSET*; hence, the target gene of this translocation is disputed. However, as approximately 30% of t(4;14) lose *FGFR3* consequent to imbalanced translocation and the risk phenotype associated with t(4;14) remains for these patients, *MMSET* is increasingly accepted as the target.[119,120] The mechanism by which *MMSET* overexpression drives MM pathogenesis in these patients is not yet elucidated. On chromosome 6, breakpoints occur ~1Mb centromeric to *CCND3*.[117] Through dysregulation of Cyclin D3, t(6;14) inactivates retinoblastoma 1 (*Rb1*) signaling that would otherwise impede cell-cycle progression.[118,121] Due to the fairly low incidence of t(6;14), its prognostic impact is not as robustly described as other canonical translocations; though it is considered to be a standard-risk marker, patients harbouring it have a higher propensity for bone disease.[118,122] Though t(4;14) is a high-risk marker, modern therapy regimes predicated on the proteasome inhibitor, bortezomib, have mitigated the adverse clinical outlook.[118,123–125] Furthermore, currently available therapies as well as those in clinical trial may be particularly potent with the t(4;14) MM subgroup: carfilzomib – a second generation proteasome inhibitor, as well as dovitinib – an FGFR3 small molecule antagonist, and FGFR3 monoclonal antibodies (mAbs).[126,127]

1.6.3.2.2 Translocation t(11;14)

Dysregulation of Cyclin D and the subsequent cell cycle progression due to inhibited *Rb1* is also produced by t(11;14) which results in overexpression of *CCND1* and is observed in 20% of MM patients.[118,121] Breakpoints for this translocation are reported within the VDJ segments and the gamma enhancer region, and are believed to arise through failure of somatic hypermutation or VDJ recombination.[117] Though patients with the t(11;14) lesion have classically been considered as standard-risk, an intermediate-risk designation may currently be more accurate as these patients appear to have moderately reduced progression-free and overall survival (PFS, and OS) on modern therapeutic regimens.[128–130] Venetoclax, a BCL-2 inhibitor, is a promising therapeutic option for this group of patients as pre-clinical and clinical studies on relapsed or refractory MM patients found this drug to have a potent anti-MM effect, high overall response rates (ORR), and extended progression free survival in t(11;14) subsets.[131–133] The addition of carfilzomib to the regimen is anticipated to boost efficacy as proteasome inhibitors mitigate Venetoclax resistance mediated by Mcl-1.[134,135]

1.6.3.2.3 Translocations t(14;16) and t(14;20)

Translocations t(14;16) and t(14;20) arise from failures of either homologous recombination or VDJ recombination and drive expression of MAF and MAFB, respectively, both of which are part of the MAF family of transcription factors.[117] Upregulation of these transcription factors deregulates *Cyclin D1*, *ARK5*, and *ITGB7*, which drive cellular invasion, cell cycle progression and anti-apoptotic characteristics, and augments cellular adhesion in the bone marrow niche, respectively.[121] These are rare translocations, with t(14;16) comprising about 5% of IgH translocations and t(14;20) comprising 2%.[121] Though both translocations are associated with significantly poor OS and PFS, only t(14;16) is included in the R-ISS as a high-risk cytogenetic lesion.[24,136] Unlike t(4;14), whose risk profile has been tempered by modern therapies, outcomes for patients with t(14;16) and t(14;20) have not improved with the introduction of proteasome inhibitors; and resistance to bortezomib is driven directly by heightened expression of MAF transcription factors which is also observed in t(11;14) MM.[136–138] Expression of MAF transcription factors is regulated by the MEK-ERK pathway which controls binding of *FOS* to *MAF* promoters. MEK inhibitors, which ultimately impede the binding of FOS to MAF promoters that is necessary for *MAF/MAFB* expression, are showing promising pre-clinical results.[139] Additionally, bortezomib impedes degradation of *MAF/MAFB* by *GSK3*□, thereby elevating levels of *MAF*, which is sufficient for resistance to bortezomib.[140] Consequently, *GSK* inhibitors are under investigation for rescuing the anti-MM effect of proteasome inhibitors in *MAF* expressing myeloma.[140]

1.6.4 Secondary Cytogenetic Lesions in MM and Their Prognostic Relevance

Secondary lesions contribute significantly to the cytogenetic profile of MM, especially in relapsing, progressing, or refractory patients.[87] These lesions are comprised of both translocations and CNVs and are significant prognostic factors as well as key parameters in therapeutic assignment.[19,24,25] Common lesions include amplification of 1q and deletions del(1p), del(17p), del(17p13.1), del(13), del(13q14), as well as translocations involving the *MYC* locus.[87] Importantly, most MM patients are assessed via FISH only at time of diagnosis, hence, the emergence of clinically relevant lesions at later disease stages may preclude informed therapeutic selection in late-stage MM. Nonetheless, secondary lesions amp(1q), del(1p), del(13), and del(17) have known hazard at time of diagnosis.[19,24,25]

1.6.4.1 Secondary Lesions to 17p13.1: TP53

The strongest prognostic indicator in MM among the cytogenetic lesions are any secondary deletions which reduce the gene dosage of *TP53* and its proximal genomic locus.[19,95] Such lesions are usually assessed by FISH, using probes to determine either whole or partial chromosomal loss, namely del(17), del(17p), and del(17p13.1).[19,24,25] At diagnosis, positivity for any of these FISH probes has an independent hazard of 2.79 and is thus considered high-risk by the R-ISS and Mayo's mSMART algorithm and is independently sufficient to designate a patient as high-risk by the prognostic index.[19,24,25] Though *TP53* is an attractive target of such lesions, there remains debate as to the actual target of such deletions as the remaining *TP53* allele is wild type in ~30% of patients harboring such deletions.[141] Nonetheless, double hit at *TP53* is a remarkably poor prognostic indicator whose negative impact has not been mitigated by modern therapeutic regimens.[19,95] There are currently no promising therapeutics beyond standard of care within this group of patients.[118] Unfortunately, clinical investigations of therapeutic regimens applied to this group have limited extensibility due to varied study requirements for percent of cells affected by such deletions to be considered positive for *TP53* loss, which ranges from 20 to 60%.[118] For disease staging, either of del(17), del(17p), and del(17p13.1) must be present in at least 60% of myeloma cells to be considered a high-risk indication.[142]

1.6.4.2 Secondary Lesions to Chromosome 1 Arms p and q

Alterations to the p and q arms of chromosome 1 are also significant negative modulators of clinical outcome.[19,97] The specific targets of these CNVs are complex and investigation of these has been undertaken through cytogenetic, array comparative genomic hybridization, and microarray assessment.[143,144] These studies identified that the minimal regions of deletion for 1p are 1p12, 1p18, 1p21, 1p22.1 and 1p32.3 in 10%, 18%, 18%, 33%, and 20% of MM patients at diagnosis, respectively.[143,144] Genes located within these minimal regions of deletion are *HSP90B3P*, *TGFER3*, *BRDT*, *EPHAX4*, *BTBD8, FAF1, CDKN2C, MAN1A2, FAM46C, GDAP2, CDC14A,* and *MTF2*.[143,144] The majority

(~70%) of MM patients harboring amp(1q) have full 1q arm gains, though a minimal region of gain has been proposed which includes 1q21, 1q22 and 1q23; *CKS1B, ANP32E* have been implicated as targets of these alterations.[10,145] Other suggested targets of amp(1q) are *MUC1, MCL1, BCL9, PSMD4, and PDZK1*.[10] Notably though, amp(1q) is an independent and significant risk marker, and is commonly concurrent with del(13q14.1) and del(17p13.1); both of which are also significant independent risk markers.[19,95,146] The mapping of genes who's over- or under-expression in high-risk MM per published GEPs, corroborates a disperse and complex contribution of alterations across 1q and 1p arms to clinical outcomes. Currently, precision therapies for this group of MM patients are limited.[97]

1.6.4.3 Secondary Lesions to Chromosome 13

Rb1 is thought to be the classic target of del (13), del(13p), and del(13q14); alterations of this gene are predominantly bi-allelic and increase in prevalence with later disease stages.[118] *DIS3* and the micro-RNA cluster *Mir15a/16-1* have also been implicated as possible targets.[147,148] *DIS3* is one of the most mutated genes in MM and double hit to it associated with poor outcome, while loss of the *Mir15a/16-1* cluster is salient in initiating plasma cell dyscrasias in murine models.[147,148] Though any of del(13), del(13p), and del(13q14) are classically considered to be poor prognostic markers, modern therapeutics have convoluted this interpretation, since in patients receiving immunomodulatory drugs and proteasome inhibitors del(13q), unlike del(13), is associated with good outcomes and extended PFS.[114,118] Furthermore, it remains unclear if these cytogenetic lesions are independently significant as they are tightly concurrent with other high-risk lesions such as t(4;14).[118] Interpretation of *Rb1* loss may be convoluted by relying on FISH for detection which commonly reports del(13q) incorrectly as monosomy 13; and copy neutral-loss of heterozygosity (CN-LOH), which is not detectable by FISH, is persistent on 13q.[87,149]

1.6.4.4 Secondary Lesions to MYC

MYC is a transcription factor of the basic helix loop helix leucine zipper class which integrates signaling from the MAPK and PI3K pathways to promote expression of genes involved in cellular proliferation.[150–152] Its expression is involved in rapid proliferation of

B-cells during germinal centre formation during normal antigen response.[153] Similarly, the abundance of *MYC* transcript is positively associated with tumour burden, and is prognostically unfavorable.[154–156] *MYC* translocations are complex structural variants which bring *MYC* in close proximity to a super enhancer, thereby driving its expression.[121] In newly diagnosed MM patients, *MYC* translocations present in approximately 33% at diagnosis and increase in prevalence in relapsed and refractory patients.[157] Dysregulation of MYC is a key event in the pathogenesis from preclinical plasma cell dyscrasias to MM and/or PCL.[157–159] Corroborating this, *MYC* translocations are seen at sub-clonal levels in MM, even when the IgH promoter is involved.[159] Three complexities of *MYC* translocations are: 1) while there is a set of standard translocations partners, these translocations are highly varied, 2) at the breakpoints, deletions ranging in size are present, and 3) the majority of *MYC* translocated patients harbour 2-5 distinct MYC translocations.[121,159–161] Hence, as many studies rely on FISH to capture MYC translocations, which uses a separation probe that cannot report the translocation partner, technical limitations preclude resolution of distinct *MYC* translocations groups which themselves may have distinct risk profiles, and probe binding is possibly obstructed by deletions or multiple translocations at the *MYC* locus.

1.6.5 Molecular Classifications, SNPs and Indels, and Prognostic Relevance

Classifications predicated upon molecular data have been investigated within MM, the earliest of these being microarray studies assessing mRNA expression patterns.[162,163] In 2005 and 2006, Bergsagel et al. pioneered microarray mediated classification in two studies which identified 7 clusters of patients: proliferation group (PR) which is largely constituted by over expression of *TOP2A, BIRC5, CCNB2, NEK2, ANAPC7, STK6, BUB1, CDC2, C10orf3, ASPM,* and *CDCA*; low bone disease (LB) which is characterized by low prevalence of lytic lesions and high expression of *EDN1* and low expression of *DKK1*; MMSET (MS) which is constituted by high expression of either *FGFR3* or *MMSET* and is tightly associated with t(4;14); hyperdiploid (HY) which is associated with over expression of *GNG11, TNFSF11, FRZB,* and *DKK1*; Cyclin dysregulation (CD-1 and CD-2) which is comprised of either CCND1 or CCND3 overexpression and is tightly associated with t(6;14) and t(11;14); and MAF/MAFB (MF) which is associated with either MAF or

MAFB over expression and is tightly linked with t(14;16) or t(14;20), respectively.[162–164] These classifications have been largely confirmed by other groups such as HOVON (Haemato Oncology for Adults in the Netherlands), though the less genetically distinct group, LB, was not.[165] The tight connection between underlying primary lesion and expression cluster in conjunction with the LB group's disputability may indicate that expression studies failed to enhance the classification of MM, with perhaps the exception of the PR signature. Nonetheless, a number of groups have built upon this work to utilize microarrays towards in-clinic patient evaluation, publishing microarrays and evaluation schemes that further stratify patients between high- and standard-risk groups.[97,166,167] Indeed, a GEP which prognosticates according to expression of genes explicitly involved in cell cycle/cellular proliferation has been published.[98]

1.6.5.1 High-Throughput Sequencing of Multiple Myeloma

MM has been genomically described during the last decade via high throughout sequencing studies in three seminal papers by Lohr *et al.*, Bolli *et al.*, and Walker *et al* published during 2014 and 2015.[9,13,16] Collectively, these studies profiled ~750 patients and provided insight on mutational patterns, mutational processes, clonal development, and clinical associations.[9,13,16] Chiefly, these elucidated a strikingly diverse mutational spectrum that is remarkably heterogeneous, both between patients and within a single tumour, and lacks monolithic genetic features common to other plasma cell dyscrasia such as Waldenstrum's macroglobulinemia.[9,13,16,168] On average, the genome of MM harbours 60 coding mutations and thousands of non-coding mutations, which places it as a highly mutated hematological malignancy, though it is still considerably less aberrant than carcinogen induced solid tumours.[6,15] Through these studies it was determined that mutations in MM accumulate via a few evolutionary patterns, including linear and branching, which occur independent of treatment regime.[9,23,169] Genes that were found to be most frequently mutated in MM are largely implicated in several canonical pathways that include NF-κB, mitogen activated protein kinase (MAPK), Jak-Stat signaling as well as DNA damage response.[9,13,14,16,170,171] Through accumulation of mutations in these ontological sets, the malignancy may lose dependence on the bone marrow microenvironment for growth and survival signals, and the marrow's chemoprotection.

Across these studies, cytogenetic lesions and translocations were identified to predominate patient classification and prognosis.[9,13,16] This has been recently challenged, with contemporary assessments identifying novel, significant genomic prognostic markers including *BIRC5*, *TP53*, and *PRMD1* contributing to mutational signatures that delineate particularly high-risk individuals.[21,94,95]

1.6.5.1.1 Aberrations in MAPK Pathway Genes

The MAPK pathway is the predominantly afflicted ontological set in MM, being aberrant in about 50% of patients.[9,13,16,172] These aberrations are almost entirely comprised of mutations in *KRAS*, *NRAS*, and *BRAF*, which are altered in about 20%, 20%, and 10% of patients at diagnosis, respectively.[9,13,16,172] Mutations in *EGR1*, *FGFR3*, *EGFR*, and *MAX*, which are upstream and downstream of RAS and RAF, though less frequently aberrant, are still common mutational targets, being altered in 5%, 5%, 1%, and 3% of MM patients at diagnosis, respectively.[9,13,16,94] In relapsed or refractory patients, mutations in the MAPK pathway genes are more prevalent, being observed in about 70% of patients, suggesting a role for dysregulation of MAPK signaling in progression.[173] Indeed, *NRAS* and *KRAS* mutations have been implicated in the progression of MGUS to MM.[174–176] Canonically, MAPK signaling transduces growth or stress signals from respective cytokines and factors to the nucleus.[150] A receptor tyrosine kinase, such as EGFR and FGFR3, phosphorylates RAS proteins upon ligand binding, thereby activating them.[150] These subsequently phosphorylate RAF proteins, thereby activating them to phosphorylate MEK.[150] Phosphorylated MEK subsequently phosphorylates ERK1/2, a transcription factor, thereby activating it.[150] ERK1/2 drives the expression of a number of genes, including other transcription factors such as the MYC MAX complexes, which work in concert producing a gene expression profile that drives proliferation.[150] In MM, mutations in MAPK are implicated in AKT mediated promotion of survival, cytokine independence, and chemoresistance.[177] In *KRAS* and *NRAS*, mutations cluster in codons 12, 13, 60, and 61; in *BRAF*, mutations cluster at codons 600. Mutations at these codons are commonly activating and are the most commonly observed mutations in oncology.[178] Departing from other malignancies though, in which *NRAS*, *KRAS*, and *BRAF* mutations are mutually exclusive, concurrent *NRAS*, *KRAS*, or *BRAF* mutations are observed in about 15% MAPK aberrated MM patients at a clonal level.[172,173,179] Despite the prevalence of these mutations, their prognostic relevance remains unclear, with reports suggesting conflicting contributions to risk.[180] Furthermore, despite the following: (i) the first protein kinase inhibitor for targeted oncology therapy, Imatinib, was approved in 2004, (ii) 48 protein kinase inhibitors exist on the market, (iii) MAPK mutations in MM are largely clonal, and (iv) the preponderance of MAPK mutations in MM has been described for three decades; there are, however, no protein kinase inhibitors currently approved for MM.[172,173,179,181,182] Lacking targeted therapies in this subset of patients is unideal as these mutations are common and hamper sensitivity of MM cells to proteasome inhibitors, a prominent therapy in standard of care, by increasing proteasome efficiency and reducing the endoplasmic reticulum (ER) stress response.[9,13,16,172,183] Genetic lesions to *EGFR* are common in malignant contexts, and serve to upregulate *EGFR* expression, or drive downstream signaling along the MAPK and PI3K pathways.[184,185] Typically, mutations cluster in the kinase domain, with common target codons being 790 and 858, or are truncating the extracellular domain from exons 2-7, or 19. Mutations in the kinase domain, which include codon 790 and 858 as well as exon 19, are classical activators, increasing catalytic activity by more than 50-fold.[186–188] Deletions of the extracellular domain remove a negative-regulatory element which circumvents signal inhibition normally mediated by receptor endocytosis.[189] *EGR1* is expressed upon MAPK signaling and is implicated in promoting apoptosis through either JUN or MYC mediated processes.[190,191] Low expression of *EGR1* defines a subset of patients with poor responsiveness to bortezomib and poor prognostic outlook; contrarily, mutations in this gene have been associated with a marginally favourable prognostic outlook.[13,192] *FGFR3* mutations are common in colon and bladder cancer, and in MM, mutations in *FGFR3* almost exclusively occur in t(4;14) patients, and have patterns of co-occurrence with mutations in *PRKD2* and *DIS3*, amp(1q), del(13q), and are negatively associated with hyperdiploidy.[13,193] Within the t(4;14) subset of MM patients, activating mutations are believed to augment the oncogenic potential of the subsequently abundant *FGFR3* proteins.[194] Indeed, *FGFR3* mutation is associated with poor clinical outcome.[21] Mutations to *MAX* occur in approximately 3% of MM patients at diagnosis, and cluster in the conserved DNA binding domain.[9,13,16] MAX homodimerizes or heterodimerizes with MYC and binds to and supresses expression of genes regulated by

E-box promoters, from which, MYC would otherwise promote expression.[195] Mutations which abrogate MAX DNA binding are common to many cancers, and mitigate this inhibition.[196] In MM, *MAX* mutations are negatively associated with *MYC* expression and the presence of *MAX* mutations has a hazard of 0.35 in a univariate assessment.[159,196] This should be probed in a multivariate model, as the negative association with *MYC* expression, the overexpression of which is a high-risk marker, may be confounding interpretation of the prognostic impact of these mutations.[197]

1.6.5.1.2 Aberrations in NF-κB Pathway Genes

NF-κB signaling is another major contributor to cell survival and proliferation in MM; it transduces many critical signals from the MM bone marrow niche and is frequently aberrant.[9,13,16,198] Two pathways have been described for NF-κB signaling, the canonical and alternative pathways.[199] The latter is involved in B-cell development and is the predominantly afflicted in MM, being altered in 10-15% of patients.[199–201] Within this pathway, upon ligand (CD40L, LTαβ, BAFF, RANKL and TWEAK) binding to TNFR of LTBR, NF-κB inducing kinase (NIK) activates IKKα dimers and the IKKα-IKKβ-IKKγ complex.[199] The activated IKKα dimers phosphorylate NF-κB2, prompting degradation of its inhibitory subunit by the proteasome thereby liberating the active p52 subunit.[199] P52 then complexes with RelB and localizes to the nucleus.[199] LTβ complexes with LTα, thereby forming LTαβ, which is the primary ligand of LTBR.[199] In MM, *LTB* is mutated in about 3.5% of newly diagnosed MM patients, commonly with truncating mutations in exon 2 with unknown functional consequences.[9,202] TRAF2 and TRAF3, which are prominently mutated in MM, act at the TNFR NIK interface to negatively regulate NIK levels/activity.[199] The IKKα-IKKβ-IKKγ complex activates the classical pathway downstream of ligand binding, leading to the localization of NF-κB hemodimers and heterodimers in the nucleus upon proteasome mediated degradation of IkBα, IkBβ, and IkBε.[199] *CYLD* is mutated in approximately 4% of newly diagnosed MM patients and acts on the IKKα-IKKβ-IKKγ complex to negatively regulate canonical NF-κB signaling.[9,13,16,199] Indeed, mutations observed in *TRAF2*, *TRAF3*, and *CLYD* are typically loss of function, thereby un-inhibiting NIK and the IKKα-IKKβ-IKKγ complex.[203] Hence, mutations in the NF-κB pathway contribute to chemoresistance in MM through hampering

NF-κB signaling inhibition, augmenting pro-survival gene expression driven by NF-κB transcription factors, thereby suppressing apoptosis.[201] *TRAF3* is the most frequently mutated gene in this pathway, being the altered component in 50% of NF-κB aberrant MM patients, and up to 7.5% of MM patients in general.[9,13,16,198] Similarly to MAPK, the prognostic implications of dysregulation in this pathway are not well characterized, though it has been suggested to be a prognostically neutral event.[165] Despite this, NF-κB aberration has been suggested as a necessary event for liberation from the bone marrow niche in sPCL, and indeed, *NF-κB* mutation is more common in late-stage MM patients.[169] Interestingly, patients with gene signature indicative of NF-κB aberration/over-activity seem to respond well to proteasome inhibitors.[201] Recently, a coordinated effect of IL-6, IL-1β, and TNFRSF21 signaling has been identified in the augmentation of NF-κB signaling across many cancer types through appositive feedback loop that also involves STAT3 and AP-1 transcription factors.[204] Indeed, MM is crucially dependent upon both IL-6 and IL-1β in the bone marrow niche, and *TNFRSF21* and *STAT3* mutations are frequent in MM.[9,13,16,205]

1.6.5.1.3 Aberration in DNA Repair Genes

Mutations in *TP53* are among the most prognostically significant events in MM, occur in 3-8% of newly diagnosed MM patients and 25% of sPCL patients, and are present in the malignancy at high cancer clonal fractions.[94,95,206,207] Similar to other cancers, mutation of *TP53* is considered to be an oncogenic event, however, the prognostic significance of these events is only partially described.[94] While double hit events to this gene identify a small subset of patients with a remarkably poor prognosis, monoallelic events have an as yet undefined prognostic contribution.[21,95,208] *TP53* mutations in MM, as well as most other cancers, cluster in exons 2-9, with the vast majority being within the DNA binding domain, between codons 110 and 285.[9,13,16,21,209] TP53 is a transcription factor that binds the genome in a homo-tetramer complex and drives transcription of genes which stop the cell cycle at the G1 checkpoint, namely p21 which is cyclin-dependent kinase inhibitor.[210,211] TP53 is homeostatically under negative regulation by MDM2, which ubiquitinates it, thereby marking TP53 for degradation by the proteasome.[212] Upon DNA damage, DNA-damage-activated kinases such as ATM phosphorylate TP53, thereby preventing ubiquitination and subsequent degradation, leading to accumulation of TP53 in the nucleus.[210] Within oncogenic contexts, ARF may inhibit MDM2 as well through TP53 phosphorylation.[212] Through interaction with Rb1 or BCL-2, TP53 may also mediate senescence or apoptosis, respectively, in response to DNA damage or oncogenic signals.[213] Consequently, aberrations to *TP53* which mitigate its plethora of anti-tumor functions are a prominent oncogenic step. Indeed, in malignant contexts, *TP53* mutations abrogate WT functioning and are increasingly recognized as neomorphic, bestowing the construct with functions that aid in malignant invasion, metastasis, chemoresistance, and epigenomic alteration.[214] Similarly, mutations to *ATM*, observed within 4% of newly diagnosed MM patients, cluster in highly conserved domains and impede DNA damage sensing or abrogate kinase activity, both of which impede *TP53* activation.[9,13,16,215]

1.6.5.1.4 Aberrations in Genes Controlling Cell Cycle

Aberration of the RB1 pathway, which is central in the control of cellular proliferation, has a much less clear contribution to the pathogenesis of MM.[216,217] Rb1 is a classic regulator of cell cycle progression that binds to and inactivates the transcription factor, E2F.[218,219] This stalls cell cycle progression until cyclin dependent kinases, classically being CDK4 and CDK6, hyperphosphorylate Rb1 upon cyclin D1 (CCND1) encounter, thereby inducing release of E2F from Rb1, E2F subsequently drives transcription of genes necessary for cell cycle progression.[218,219] Mutation of *Rb1* is observed in approximately 5% of MM patients at diagnosis and are present at a range of cancer clonal fractions.[9,13,16,94] Dysregulation of *CCND1* in MM is a classical feature of the disease consequent to t(11;14), trisomy 11, or otherwise driven overexpression.[162,163,220] Mutations in *CCND1* are also common, at about 4% of newly diagnosed patients, and occur in the amino terminal domain.[9,13,16] CDKN1B and CDKN2C are both cyclin-dependent-kinase inhibitors and are hence negative regulators of Rb1.[218,219] Mutations in these genes are each observed in approximately 3% of patients and are present at a range of cancer clonal fractions from 0.1 to 1 with a mean at ~0.6.[9,13,16,94] Mutations to *CDKN1B* and *CDKN2C* are associated with lower gene expression, impair kinase function, or cause incorrect cellular localization; all of which culminate in Rb1 hyperphosphorylation and cell cycle progression.[221–223] Germline mutations in *CDKN1B* are implicated in pediatric Cushing's disease, and have been observed in MM.[9,13,16,222] Mutations in each of *RB1*, *CDKN1B*, and *CDKN2A* have been identified as driving events, though in a univariate analysis, these genes' mutational status had no prognostic impact.[9,13,16,94]

1.6.5.1.5 Aberrations in RNA Processing Genes

Genes encoding RNA processing proteins, including *FAM46C* and *DIS3* are mutated in 12% and 8-11% of newly diagnosed MM patients, and present at 50-60% cancer clonal fraction.[9,13,16,94,224] FAM46C is a non-canonical poly(A) polymerase which acts as a tumour suppressor in MM.[225–227] Notably, this role seems unique to MM, as no other cancer is statistically enriched for *FAM46C* mutations.[228] Within MM, 70 mutations are known to occur in this gene, many of which are frameshift inducing, or are stop-gains.[227,229] Indeed, *FAM46C* is either lowly expressed or mutationally rendered non-functional; when FAM46C is reintroduced in MM cells, cell death ensues. Though deletion of the *FAM46C's* cytogenetic locus, 1p12, is a known poor risk marker, mutations within this gene have an undefined contribution to prognosis.[87,94,143] Mutation of *DIS3* may be important in the progression of MM to sPCL, as mutation in this gene increases to an incidence of 21% in these late stage patients.[230] DIS3 is a catalytic subunit of the exosome with 3' to 5' catalytic activity which diversely participates in RNA processing and play a central role in processes that drive MM pathogenesis, including Ig class switch recombination and somatic hypermutation.[224,231] Mutations modestly cluster within the RNB domain, though are fairly well distributed across the PIN, CDS2, and S1 domains, all of which are highly conserved.[231] There is suggestion that these mutations severely imped the enzymatic activity of DIS3, and consequently the exosome with impacts on RNA metabolism at large.[230] Mutation in *DIS3* has a prognostically unfavourable association, and mutations in *DIS3* at sub clonal levels are linked to significantly unfavourable prognosis; though the extent of clinical impact of *DIS3* mutations remains under debate.[21,94,232]

1.6.5.1.6 Other Commonly Mutated Genes in MM

The remaining frequently mutated genes in MM, *IRF4* (2.5%), *STAT3* (4.5%), *SP140* (6%), *ACTG1* (4%), *PRDM1* (6%), are an eclectic assortment across ontological categories.[9,13,16] IRF4 is a transcription factor necessary for the maturation of lymphocytes that is important in B and T cell receptor signaling.[233] Interestingly, within MM cells, IRF4 and MYC are direct targets of each other, and MM cells display IRF4 addiction.[234] Mutations of *IRF4* cluster in MM and other cancers at codon 123 (K123R) and are thought to be activating.[235,236] Consistently, overexpression or mutation of *IRF4* is a poor prognostic factor.[237] STAT3 is a transcription factor that is activated upon IL-6 signaling, and is critical to MM survival, proliferation, and persistence.[238–240] Activation of STAT3 via phosphorylation is a poor prognostic marker observed in about 10% of patients and mutations in *STAT3* are associated with a poor prognosis and predict poor response to lenalidomide.[21,241,242] SP140 is a nuclear body protein that is typically afflicted by truncating mutations in MM, its functional involvement and prognostic relevance in the malignancy remains unclear.[9,243] *ACTG1* encodes a cytoplasmic actin which is commonly mutated in MM and is implicated as an oncogenic driver in this malignancy.[94] In MM, the R39I mutation is common in *ACTG1*, and may have implication for actin polymerization.[244] PRDM1 is a transcription factor that functions as a master regulator of B-cell development.[245] PRDM1 is transcribed in two isoforms, the α and β form.[246] The latter of which is shorter and exhibits significantly weaker repression of genes whose expression is involved in oncogenesis.[246] Imbalance between the α and β form can be achieved by mutation, and indeed MM cells express both the α and β form, while normal plasma cells express only the α form.[246]

1.7 MM Therapeutic Approaches

1.7.1 Historical Therapies

The first description of MM were reported in the 1840s, with presentation of easily fractured bones, linen stiffening urine, and red bone marrow upon examination.[248] Initially, this malignancy was treated with bloodletting and leeches, to minimal effect.[249] Subsequent investigations assessed quinine, urethane, and melphalan for their capacity to reduce serum immunoglobulin, and anemia as well as improve patient outcomes.[250] Melphalan was the first compound identified to improve patient outcome in 1958, and its therapeutic potential was substantially evidenced by 1967.[249] However, modulation of aberrant biochemical states remained of prominent concern.[249] Prednisone, a corticosteroid, was assessed as a single agent MM therapy in 1962 and achieved a significant reduction in serum immunoglobulin and increase in hematocrit, though showed no benefit for extension of survival times.[251] In 1969, melphalan and prednisone combination therapy was assessed in a large clinical trial, which demonstrated synergistic survival enhancement, and favorable modulation of biochemical parameters.[252] This combination was termed MP and was a monolith of MM treatment for the subsequent decades.

Melphalan remains one the most frequent therapeutic components administered to MM and lymphoma patients.[253] As an alkylating agent, melphalan crosslinks DNA at GC base pairs, thereby impeding DNA and RNA synthesis and inducing myeloablation.[254] Though common, melphalan may pose considerable risk (6.1%) for therapeutically induced secondary malignancy.[255]

1.7.2 Standard of Care Therapies

Autologous stem cell transplantation is the most effective MM therapy and is the therapeutic target of treatment regimens for eligible patients (usually younger than 70 years of age, with variation between treatment centers).[253] Patients first undergo induction, during which, combinations of chemotherapeutics are administered to de-bulk the tumour (see section 1.7.2.1).[253] Subsequently, hematopoietic stem cells are mobilized using cyclophosphamide and granulocyte-colony stimulating factor (G-CSF), and a Hickman line is used to capture circulating hematopoietic stem cells, aiming at $5x10^6$ $CD34^+$ cells per kg. Next, myeloablation is standardly achieved through high-dose melphalan at $200mg/m^2$, though a combination of busulfan and melphalan is under assessment within a stage III clinical trial and may offer increased PFS.[256,257] Finally, $CD34^+$ cells are readministered, and patients may undergo consolidation and/or maintenance therapy.[253,256] The former is the short-term administration of chemotherapeutics with modest toxicity profiles to push towards CR of sCR status. The latter is the long-term administration of chemotherapeutics with very-low associated toxicities to delay onset of relapse or progression, which is inevitable even in patients that achieved CR or sCR.[258,259]

1.7.2.1 Induction Therapies

For multiple myeloma, Induction therapies are typically combinations of several chemotherapeutics, with triplets being most common. Current Food and Drug

Administration (FDA) approved chemotherapeutics are alkylating agents: cyclophosphamide, Bendamustine, Doxorubicin; proteasome inhibitors: Bortezomib, Ixazomib, carfilzomib; Immunomodulators: thalidomide, lenalidomide, pomalidomide; monoclonal antibodies: Daratumumab, elotuzumab; a glucocorticoid: Dexamethasone; a histone deacetylase: Panobinostat; and the anti-mitotic agent Vincristine.

Mainstay combinations in myeloma care are CyBorD, VRd, and Vd. CyBorD is comprised of Cyclophosphamide, Bortezomib, and Dexamethasone. In 2015, CyBorD was demonstrated to have superiority over the triplet, PAD (doxorubicin, dexamethasone, bortezomib), more frequently achieving VGPR or greater.[260]. In a recent meta analyses on newly diagnosed multiple myeloma patients that did not receive autologous stem cell transplantation, the median survival of those receiving CyBorD was 92.9 months, which is superior to other dexamethasone containing regimens, RD and Vd, which had median survivals of 79.1 and 56.3 months, respectively.[261] VRd, comprised of Bortezomib, lenalidomide, and dexamethasone, is more effective than CyBorD, achieving median survival of 112.6 months, however is associated with greater toxicity.[261,262] While Vd, a combination of Bortezomib and Dexamethasone, is modestly less effective than CyBorD, it is a highly tolerable therapy, and is thus suited to low-risk and frail patients.[261,263]

The above therapies employ first generation proteasome inhibitors, and second-generation immunomodulatory drugs. The former class is now on second generation with Ixazomib and carfilzomib, and the latter is on its third generation with pomalidomide.[264] It remains unclear how these next-generation therapeutics will shape patient care, and whether they offer additional utility in newly diagnosed cases.[265,266] Nonetheless, in relapsed cases, these novel agents outperform standard regimens in the achievement of complete response. Carfilzomib, Lenalidomide, and dexamethasone outperforms VRd in relapsed setting, with 78% of treated individual reaching near-complete response or higher and having a predicted 24-month survival of 92%.[268,269] Another study swapped pomalidomide for lenalidomide in the aforementioned regimen and achieved a higher response rate of 87% with a complete response rate of 31%, which is high for an otherwise difficult to treat population.[270]

1.7.3 Treating High-Risk Multiple Myeloma

The treatment of high-risk MM has been an area of increasing focus. Of specific concern, has been the treatment of groups defined as high-risk by t(4;16), del(17p13.1) and amp(1p).[271] This group is heterogenous in response to therapies, and requires careful consideration of the landscape of genomic lesions to effectively treat an individual.[271]

Firstly, thalidomide, one of the first novel agents for MM, does not improve high-risk patient outcome over previous therapies.[272] Lenalidomide however, a second generation derivative of thalidomide, does improve outcome of del(17p) patients specifically when included in maintenance therapy (PFS extended to 29 months, compared to 24).[273] Patients with t(4;14), should be treated on bortezomib including regimens if they are transplant eligible, and should be considered for tandem over single transplantation.[274]

In general, MM patients defined as high-risk are recommended to receive triple induction therapies which include a proteasome inhibitor, immunomodulatory drug, and a corticosteroid.[271] All transplant eligible patients should receive at least one autologous bone marrow transplant when not contraindicated, and maintenance therapy should reflect a dose reduced triplet as well.[271] There are, as yet, no combination therapies showing significantly improved outcome outcomes for transplant ineligible patients.[271]

1.7.4 Precision and Modern Therapies

Multiple myeloma, having a highly altered genome with many lesions being clonal, may be a good candidate for precision-therapy approaches.[9,13,16,275] While cytogenetic lesions offer modest utility to guide therapeutic decisions, such as indicating bortezomib based regimens in t(4;14), t(14;16) or del(17p) patients, higher-resolutions and broader assessments in clinical settings are warranted to realize precision medicine for MM.[123,276] Gene expression profiles and sequencing offer superior genomic assessments, particularly considering precision medicine; their clinical has however been limited.[275]

Gene expression profiles can detect alterations that predict sensitivity to numerous targeted therapies. High *DKK1* expression, which is associated with lytic lesions in MM may be targeted by an anti-DKK1 monoclonal antibody[277,278]. Mimetics of BH3 may be indicated for in patients with a high *Bcl-2*/*Mcl-1* ratio.[279] While clinical exploration of the aforementioned has been limitedly explored clinically, a more broad assessment of transcriptional profiles identified an 80-gene signature predictive of response to bortezomib based regimens in a cohort of 128 patietns.[280] Nonetheless, due to lack of consensus and technical demands and limitations, gene expression profiling is neither part of standard clinical work-up for therapeutic assignment, nor has its inclusion been recommended.[99]

Mutational profiling in MM remains a viable avenue to achieve the highly resolved genomic information requisite in precision medicine. Mutations in *BRAF*, *KRAS*, *NRAS*, *BRAF*, *IRF4*, *ATM*, *FGFR3* are all targets of precision therapies currently under investigation.[183,281,282] Of these, vemurafenib, an inhibitor of *BRAF* V600E has already demonstrated durable responses in relapsed and refractory patients.[282] Acknowledging the potential of mutational profiling for precision medicine in MM care, The Multiple Myeloma Research Foundation (MMRF) is conducting a trial, MyDRUG, for targeted therapies which assigns individuals with mutations in MAPK pathway genes, cyclin dependent kinases, *FGFR3*, and *IDH* mutations to appropriate experimental arms.[283] Importantly, a number of conditions precede widespread clinically viable precision medicine. Firstly, appropriate therapies and their associated performance with catalogued genetic indications must be widely available. Secondly, appropriate and clinically viable sequencing approaches must be accessible.

1.8 Next Generation Sequencing and Analysis Overview

1.8.1 Background

Sequencing-based methods to assess more detailed genetic aspects of cancer biology in both the research and clinical settings are becoming increasingly appreciated and have driven improvements in diagnostics, subtyping, prognostics, and therapeutic choice.[284] These technologies have thrown greater light on the pathology and natural history of various cancers, allowing progress within the paradigm of ‘precision medicine’. Precision medicine is the use of highly resolved patient molecular characteristics to inform on diagnosis, prognosis, and therapeutic choice.[285] Accordingly, sequencing technologies are now commonplace in clinical molecular labs, and facilitate targeted or genome-scale interrogations of cancers.

1.8.1.1 Sequencing Technology Overview and Terminology

Next generation sequencing technologies can be categorized by the number of nucleotides sequenced within each fragment as short-read, which sequences between 50 and 700 nucleotides of each fragment, and long-read, which sequences kilobases of each fragment.[286] Short-read sequencing is the most efficiently produced, and hence accounts for the bulk of sequencing efforts currently undertaken.[287]

While short-read sequencing data is efficiently produced, it lacks 'long-range' information which allows accurate placement of sequenced reads within a larger genomic context.[286,287] Consequently, short-read data is primarily useful when a reference genome is available, against which the reads may be aligned.[286,287] Importantly, short-read data is minimally informative in extended regions of low-diversity as mapping is impaired, and when extensive structural variation diverges the sequenced genome from the presumed reference.[286,287] While long-read sequencing methods address these issues, they are more expensive than standard short-read approaches, have significantly higher error rates, and may be more technically challenging.[286,287] Nonetheless, they provide information of unparalleled power to resolve complex genomic loci, or construct a *de novo* genome.[286,287]

Bridging the gap between short-read and long-read data is paired-end sequencing, wherein, a DNA fragment which may be longer than twice the read length is sequenced from both ends.[288] Read pairs generated by this approach contain more information as both reads originated from a contiguous DNA segment, hence, if reads map discordantly a mutational event can be inferred to have occurred within the DNA fragment even if the alteration occurred within a non-sequenced portion. Sequencing both ends of a fragment gives rise to a number of read configurations including soft-clipped, one-end anchored, split, and discordant. Soft-clipped reads are aligned to the reference excluding the 5' and/or 3' terminal. One-end anchored read pairs have on read mapping to the reference without the partner mapping. Split read pairs have one end of a read mapping to one region, and the other end mapping to unexpected region or with an unexpected orientation. Discordant reads occur when each read of the read pair map to genomic loci unexpected given the insert size, or with unexpected orientations. All of these may be used as evidence for a structural variant.

1.8.1.2 Considerations for NGS Application in Clinic

Consequent to the dramatic increase in sequencing efficiency afforded by next-generation technologies, genomic interrogation has become commonplace in clinical settings.[284,289] Though NGS has made WGS markedly more feasible, cost and turn-around time considerations remain for many clinical settings.[289] WGS not only requires more sequencing resources, but also computational and data storage infrastructure, and dramatically increases the chances for incidental genetic findings that pose complex problems for interpretation, counselling, and disclosure of results to patients.[289,290] Furthermore, disease assessment often requires identification of low frequency mutations among many DNA copies with wild-type alleles, thereby necessitating higher sequencing depths; this is problematic as cost of sequencing imperfectly scales with depth and scope. Targeted panels address this well, achieving high coverages of 500x-7000x over specific regions in a cost-effective manner.[291] Such depths are infeasible for WES or WGS in a clinical setting due to cost and time constraints, though, whole-genome or -exome assessments are performed clinically when indicated, and for many clinical trials.[292–294] High sequencing depth is often a clinical priority, as identifying 1 malignant cell in $1x10^6$ normal cells for minimal residual disease detection or studying key points of clinical interest, require sequencing depth that can extend many orders of magnitude beyond 1000x.[295] Consequently, restricting the scope of sequencing to regions for which high-depth interrogation is warranted is standard practice; however, the identification of structural variants can be severely impeded by restricting the scope of assessment. Hence, the goals of the test must be clearly defined.

1.8.1.2.1 Single Nucleotide Variation and Indel Detection

SNVs and indels occur throughout the genome; however, clinically relevant alterations – those with prognostic or precision medicine implications – are generally clustered within the coding regions of a limited set of genes for a given cancer type or tissue site.[183,281,282,296] Based on this, a number of sequencing panels have been produced for specific cancer types, broad groups of cancers, or for cancer in general.[20,21,297,298] These facilitate cost effective, high-depth coverage of regions of interest, allowing confident detection of SNVs and indels even if only present in a small subset of DNA molecules assessed in a given patient sample. Depending on the panel design, however, important lesions within the assayed regions may go undetected, such as copy number variation, loss of heterozygosity, and other structural variants.

1.8.1.2.2 Copy Number Variation Detection

Copy number variations occur throughout the genome, ranging in size from whole chromosomes to smaller focal alterations. Cancers often have associated patterns of CNVs, wherein, specific chromosomes, chromosome arms, cytobands, or genes are amplified or deleted.[299,300] For smaller regions, namely cytobands and genes, the CNVs that afflict them may be highly varied in size and placement, and these cytobands or genes are generally the minimal regions afflicted by CNVs when comparing across many patients.[301] This is true for deletion of *TP53* in MM, which can be consequent to del(17p13.1), del(17p), and/or -17.[113,142,302] CNV detection by panels is challenged by 1) The range in size and placement of CNVs and the consequently variable position of CNV break ends does not lend itself to detection by a targeted approach; and 2) Many CNV calling algorithms assess for significance in deviations of read depth from the norm within a genomic region, this norm may be poorly defined and significant deviations may by obfuscated by noise in a targeted assessment.[303,304] Hence, CNV detection is more successful when utilizing larger panels, often designed with special consideration for CNV detection, or with WES or WGS.[305] We have previously demonstrated that ultra-low-depth WGS, where the average coverage was less than 1x, can accurately detect CNVs across the genome when compared to FISH.[26] Ultra-low-depth WGS is uninformative for SNV and translocation detection, however, it is of low-cost, rapid analysis, and is well powered for CNV detection even within genomically complex cases such as clinical myeloma samples.[26]

1.8.1.2.3 Non-CNV Structural Variant Detection

Structural variant detection remains a challenge for algorithms using short-read sequencing data.[286,287] This is because the strongest evidence for identifying such variants is the presence of reads which span the breakpoint, thereby necessitating high sequencing depths for accurate and confident determination, similar to SNV detection.[286,287] Paired-end sequencing has notable utility within structural variant detection as either read in the read pair, which originate from a modestly larger DNA fragment, can span the breakpoint instead of just a single sequencing read.[306] This could result in reads being split across different genomic loci or they may map discordantly, implying a larger or smaller insert size than expected, or with an unexpected orientation relative to one another. With paired-read technology, most structural variant detection protocols recommend depths around 30X, with higher coverage better addressing the complexity of somatic variation in cancer.[307–310] Structural variations in MM were first described at the cytoband level via karyotyping and subsequently by FISH, and many diseases subtypes have well described profiles of structural variant loci using these.[311] However, the loci of the break ends which define these structural alterations may vary by megabases, and non-canonical structural variants with important implications may also occur[88,94,160]. As structural variants within both coding and non-coding regions of the genome may have clinical implications, WGS, though possibly impractical clinically, is best suited for their detection; WES and panel sequencing experiments are impeded by their limited scope.

1.8.1.2.4 NGS for Clinical MM Genomic Profiling

The core challenges of comprehensive genomic profiling of MM by NGS arise because of competing technical demands for structural variant detection compared to SNV and indel detection.[26,307–310,312] Synchronous capture of all MM relevant classes of genomic abnormalities for clinical purposes has not yet been possible in a cost-effective manner. SNVs and indels demand sequencing depth only feasible using a panel approach, but this is ill-suited for detecting CNVs and other structural variants. CNVs are amenable to detection by ultra-low-depth WGS, which is feasible in a clinical laboratory, however this approach is inadequate for detection of other types of structural variation, namely translocations.[313,314] Hence, an economically viable approach for molecular profiling of the MM genome may need to be two-pronged; targeted sequencing to suitable depths of coverage (~1000x) for SNV and indel detection in specific regions of interest, and WGS at the lowest depth at which both copy number, translocations, and other relevant structural variants can be confidently detected.

1.8.2 NGS Technologies for Clinical Oncology

Most commonly, clinical genomic mutational testing for oncological diagnostics or prognostics is performed on panels which are amplicon (suitable for up to ~50-60 genes), or hybrid capture (for panels greater than 100 genes) based.[315] These gene panels range in specificity from being designed for assessment of a few genes or one malignancy, to assessment of over 500 genes or general oncological categories, such as solid and hematological malignancies, to pan-cancer assessments.[21,291,315,316] Additionally, WES and WGS have been effectively employed in a few clinical settings and trails.[315] Importantly, even in difficult to treat cancers, where no standard actionable mutation were identified, inclusion of a multigene sequencing panel, the Ion AmpliSeq Cancer Panel (CP1) (ThermoFisher Scientific) (190 amplicons across 40 cancer associated genes), into the clinical workups demonstrated increased progression-free survival by 30% and overall response rate by 10%.[317]

Panels with broad application include the FoundationOne CDX (FDA approved), MSK IMPACT (FDA approved), TruSight Oncology 500 (TSO500), and Trusight Myeloid.[297,318] The former three are large and can interrogate tumour DNA for SNPs, CNVs, tumour microsatellite instability, tumour mutation burden, and a limited set of structural variants, while the latter is small and can report on SNVs and indels.[318] Subsequent analysis of the data also allows determination of homologous recombination deficiency, and mismatch repair deficiency. Collectively, these cover the genetic indications for all currently FDA approved oncological precision therapies.[319]

For malignancies with lesions of clinical significance that are unique to them, general panels may be unideal, leaving prognostically important information uncaptured.[320] Accordingly, very few MM-specific panels have been developed, none of which have been widely adopted or are commercially available. The two most prominent panels are the myTYPE and M(3)P.[20,21] Both capture SNVs across an assortment of genes, while the myTYPE panel also attempts to describe structural variants at the IgH locus and copy number changes throughout the genome. Hence, the myTYPE panel fits well into the current R-ISS scheme and captures a range of lesions which may soon be indications for precision therapy.[20] However, the M(3)P panel has demonstrated novel patient stratification on small lesions within the target genes, albeit with only marginal difference for a small percentage of patients.[21]

1.8.3 Somatic SNP and Indel Calling Algorithms

Development of accurate methods for the detection of SNPs and Indels with short read sequencing data has been an active area of bioinformatics focus over the last decade, and continues to be an ongoing, albeit more limited, field of research. Accordingly, a plethora of algorithms using a number of related statistical models to assess support for putative variants have been published. Stemming from this variation, algorithms have different performance in different contexts; and therefore, employing multiple algorithms in an ensemble approach has emerged as a 'best practice'. In a recent performance assessment, MuTect was found to have higher positive predictive values (0.77-0.97) than Vardict and Freebayes (0.33-0.73, and 0.35-0.55, respectively), and an ensemble approach including MuTect, FreeBayes, Vardict, Muse, and MuTect2 was most performant, having a positive predictive value (PPV) of (0.94-0.98).[321] Importantly, FreeBayes and Vardict identified a maximum of 2.3% and 7.2% more true positive SNVs than Mutect across this assessment.[321] In recent assessments of indel callers using simulated and real data, Scalpel and Pindel increase in sensitivity and precision with sequencing depth, and at 50x coverage, reached ~90% sensitivity and ~100% precision for indels from -200 to 50 bb in size.[322,323] In another assessment, Platypus, though less precise than Pindel (0.22 compared to 0.42), identified 2.1 times more true positives than Scalpel, and an ensemble approach was found to be most performant.[324] In this work, the variant callers used are Platypus, MuTect, Pindel, VarDict, FreeBayes, and Scalpel as implemented in the standard clinical bioinformatics pipeline within Nova Scotia Health.[325]

Pindel is geared towards detection of indels of varied sizes.[326] It is underpinned by algorithms looking for read pairs in which only one read of the pair does not map; the mapped read provides an anchor point and orientation which the algorithm accounts for while splitting the unmapped reads into two mappable chunks that potentially span a breakpoint.[326] Notably, while Pindel is specific and sensitive for indels -200 to 50 bp in size, its sensitivity quickly decreases for indels less than -300 bp or greater than 100 bp, reaching 0% recall even at 50x coverage.[322,324,327] Additionally, performance is hampered in genomic contexts where the anchor may map to a repetitive sequence element in the genome, or if SNPs or sequencing error introduces mismatches in either segment of the split read.[326]

Scalpel is a modern algorithm which focuses on indel detection.[328] It constructs De Bruijn graphs from reads in a BAM file across the whole genome in segments of a specified size independent from the reference genome.[328] The constructed sequences (branches of the de Bruijn graphs) are then compared to the reference to identify indels using an implementation of the Smith-Waterman alignment algorithm.[328]

VarDict identifies both indels and SNVs.[329] For indel calling, reads that were soft clipped during alignment or have mismatches undergo unsupervised and supervised local realignment which incorporates more variant supporting reads in the analysis than Pindel's method, offering a more accurate estimation of the variant's frequency.[329] VarDict incorporates indel information into its SNV calling as well, so reads which support an SNV but have poor mapping quality due to a nearby indel are not dismissed after local re-alignment.[329] This has important performance implications: when applied to The Cancer Genome Atlas lung adenocarcinoma dataset, VarDict identified driver mutations in *KRAS*, *NRAS*, *BRAF*, *PIK3CA*, and *MET* in 16% more patients than previously reported when analyzed using MutSig2CV.[228,329]

MuTect is a sensitive and specific SNV caller, which does not call indels.[330] Often, it is combined with Scalpel as an aggregate 'single caller'. MuTect evaluates variants under two models, one in which the variant is assumed to be a sequencing error, and in the other the variant is assumed to be present at an allele fraction proportional to the fraction of reads in which the allele occurs.[330] These are then assessed within a Bayesian framework, and if the latter model has log odds likelihood meeting predetermined threshold, the variant is accepted. All accepted variants are then filtered for proximity to sequencing gaps, strand bias, poor mapping quality of supporting reads, multi-allelic evidence, and clustering.[330]

Platypus can detect SNVs, multiple adjacent SNVs (MNVs), and indels up to several kb in length.[331] This algorithm employs three steps: alignment of reads to a reference genome and subsequent variant detection, creating a De Bruijn graph of variant-implicated reads, and identifying and scoring haplotypes from this graph by ascertaining unique paths in the graph.[331] The scoring algorithm aligns each haplotype-supporting read to the haplotype sequence and assesses the quality of this alignment in a hidden markov model using the Viterbi algorithm.[331] The Frequency of these haplotypes are then estimated, and variants are called in accordance with their haplotype quality and frequency.[331]

FreeBayes, similarly to Pindel, constructs a haplotype from sequence data.[332] However, FreeBayes employs a Bayesian approach, rather than hidden markov models, to determine the maximum likelihood of variants comprising a haplotype being real versus a sequencing or alignment artifact.[332] The likelihood of an observed variant being erroneous is estimated by the per-base quality score of reads constituting the/supporting haplotype scaled by the likelihood of sampling variant supporting reads from a normal genotype.[332]

1.8.4 CNV Calling Algorithms

CNV calling may be performed through either *de novo* genome construction, looking for split or discordantly mapped reads that define the junctions of CNVs, or by identifying variation in depth of coverage.[333] The latter benefits from its ability to determine exact copy numbers, large variants, and copy changes which implicate complex structural variation, or implicate low complexity regions.[333] Additionally, some algorithms benefit from a low required depth of coverage, as compared to the other methods which require upwards of 40x coverage for accuracy.[333] Though, algorithms that assess for split and discordant reads may offer much higher resolution of CNV boundaries, and identify tandem duplications.[333] We have previously demonstrated the utility of QDNAseq in low depth interrogation of MM genomes for CNV detection, hence, we used QDNAseq herein.[26,334] QDNAseq implements the CGHcall algorithm to assess for significant deviations in depth of coverage to call CNVs in our work.[26,334] Briefly, reads are aligned to the reference genome, then the genome is segmented into bins of 1,5, 10, 15, 30, 50, 100, 500, or 1000 kb and the coverage for each bin is calculated with normalizations for mapability and GC content on a bin-by-bin basis.[334] The log odds for each bin's variation in coverage from the median is calculated, and CNVs are reported accordingly.[334]

1.8.5 Other SV Calling Algorithms

Structural variant calling is a computationally difficult task, and is an increasingly active area of research, especially as the significant role that structural variants play in cancer is increasingly acknowledged.[327] Broadly, algorithms assess aligned BAM files for differences in coverage, or groups of poorly mapping, split-, and/or discordant reads, with some algorithms subsequently incorporating this data for the construction of a haplotype.[327] A recent performance assessment recommends an ensemble calling approach, and suggests inclusion of GRIDSS and MANTA due to their high performance in simulated and cancer datasets compared to other calling algorithms, having PPVs of 0.81% and 0.59% and sensitivities of 0.85% and 0.88%, respectively.[327] In this same study, LUMPY also had a high PPV of 0.71, though a relatively low sensitivity of 0.33.[327]

In another study on tumour and simulated data, LUMPY was 80% sensitive at 10x depth for translocation identification of heterozygous variants.[310] SVABA, a modern structural variant calling algorithm published in 2018 was not included in these assessments. SVABA can detect complex translocations, which are abundant in MM and which may be cryptic to MANTA, GRIDSS, and LUMPY.[121,308,327] In this work, the structural variant calling algorithms used are SVABA, GRIDSS, LUMPY, and MANTA.[308–310,335]

SVABA first segments the genome into 25kb bins with 2 kb overlap.[308] Then, in each bin the discordant-, split-, or poor mapping quality reads are assembled into a consensus sequence/haplotype using s string graph assembler.[308] These consensus sequences are then joined together in an organization consistent with the discordant reads that map between bins.[308] This joining facilitates detection of interchromosomal variants and alterations larger than 25 kb in size.[308] These haplotypes are then mapped back to the reference using BWA-mem, and variants are called and scored. Scoring considers the alignment score of the contig to the reference, as well as the number of reads which align better with the contig than the reference; reads mapping to the contig support the variant and increase the variant score.[308]

GRIDSS first extracts soft-clipped, split-, one-end-anchored, discordant-, and low map quality reads across the whole genome.[309] Then, using a novel implementation of De Bruijn graphs, these reads are assembled into contigs, in a process termed whole-genome contig-assembly; presumptively, these contigs span breakends.[309]. These contigs are then mapped back to the reference and split and discordant reads are used for variant calling and scoring.[309] A variant's score is the phred scaled likelihood of the supporting reads originating from the implicated gnomic loci in the absence of a structural variants.[309] For discordant reads, this score reflects the probability of the variant supporting read pair having the observed insert size given insert size distribution of the library; the insert size for chimeric reads is considered to be 10 standard deviations from the mean.[309] For split reads, score reflects the probability of the observed soft-clipping given the distribution of soft-clipping within the alignment.[309]

MANTA constructs a graph, wherein nodes represent breakends, and edges represent putative SVs giving rise to the breakends.[335] Firstly, MANTA extracts discordant-, soft clipped-, and poor mapping quality reads from an aligned BAM file.[335] Then, each read pair is used to construct a small single edged graph, as the algorithm assesses more read pairs these graphs are merged.[335] Through merging, the nodes no longer represent a single break end, but a cluster of break ends which may be implicated in more than one structural variant.[335] Contigs are constructed around nodes (break ends) and aligned to the reference using the Swiss-Waterman algorithm.[335] Variants with coverage higher than 3 times the average coverage for the sample, quality score less than 20, or lacking paired end support are filtered out.[335] The quality score represents the likelihood of observing a given variant supporting reads in a diploid model.[335]

LUMPY uses soft-clipped, discordant, and split-reads to identify putative break point regions.[310] These regions are probabilistic ranges around clusters of the aforementioned reads, where each nucleotide position in the interval is scored for its likelihood of being a break point.[310] Scoring is performed using two models, one for discordant-read support, and another for split read support.[310] In the former, positions are scored depending on the abundance of discordant reads which putatively span them, where the probability of a read spanning the point is proportional to the probability of observing an insert size requisite for the read to span the position given the insert size distribution within the library.[310] For split reads, the positional score is highest at the split in the read, and these probabilities are summed across overlapping split reads.[310]

1.8.6 Variant Interpretation

Concomitant with the bulk of data produced by NGS based assessment is the abundant identification of variants with unknown pathological relevance. Addressing this, large databases which catalogue mutations and the disease context in which the variants were found have been developed, Catalogue of Somatic Mutations in Cancer (COSMIC) and ClinVar being prominent examples. These databases permit variant annotation with information relating to the pathological states in which the variants have been previously reported. Additionally, a number of algorithms, often predicated on machine learning models, provide computed scores for pathogenicity (rfPred, SIFT, MutationTaster, Polyphen2, FATHMM-XF) or predict splicing alterations consequent to mutation (SPLICEAI).

1.9 Rational and Aims of this Study

Multiple myeloma presents with numerous genomic lesions that range in size from single nucleotide variants to whole chromosome alterations. The current assessment, FISH, cannot capture the diversity and complexity of this genomic landscape. Next generation sequencing assays are highly adaptable to capture these lesions, however, a clinically viable genomic interrogation approach has yet to be clinically described. In this work, we aim to

1) Develop and describe a multiple myeloma-specific targeted sequencing panel that captures prognostically relevant, and therapeutically informative small-scale lesions in genes frequently mutated in multiple myeloma.
2) Develop a relatively low-depth whole genome sequencing approach to capture the underlying structural variation existing within MM patient genomes, and demonstrate its performance against FISH.

Chapter 2 Methods

The methods used for each study herein are almost completely described in their respective chapters, **3.3** and **4.3**. Absent from these is the methodology for bone marrow processing and CD138 positive selection process which were performed per Nova Scotia Health Authority (NSHA) standard operating procedures as described below, as well as DNA/RNA extraction, and library preparation for the DMG26

2.1 Tumour Bank Sample Data Collection:

Patients included on the DMG26 and in the WGS study consented to the myeloma tumour bank during bone marrow acquisition at time of diagnosis or relapse, at which time their medical registration number (MRN), as well as reason for and date of bone marrow acquisition are recorded into the myeloma tumour bank sample tracking database, and to each sample a unique myeloma tumour bank identifier is assigned for patient de-identification. Using the MRN, each patient's laboratory data at time of, or within one month of bone marrow sampling was collected using NSHA's laboratory information system. The parameters were FISH results, isotype by immunofixation, immunoglobulin quantitation, bone marrow plasma cell infiltration (highest of aspirate, biopsy, or FLOW), B2M levels, hemoglobin, leukocyte counts, neutrophil count, lymphocyte count, eosinophile count, white blood cell count, monocyte count, reticulocyte count, platelet count, LDH, ALT, AST, albumin, bilirubin (total and direct), creatinine, eGFR, calcium, and ALP. Additionally, clinical data, was collected using each patient's MRN as a search parameter within NSHA's clinical information system. For each patient, data was collected from the clinic letters, starting at time of diagnosis. We catalogued time to each event and event type (relapse, progression, death, response to therapy), therapeutic regimen (therapies, number of cycles), and stage at diagnosis for each patient.

2.2 Bone Marrow Processing

Patient bone marrow samples in Ethylenediaminetetraacetic acid (EDTA), which arrive at the molecular lab, are spun at 500g for 10 minutes, following which, the plasma is aliquoted and stored at -80 C. The remaining bone marrow is then split, with 5mL aliquots placed into separate 50mL conical tubes, into each of which, 40mL of ammonium chloride lysis buffer (ACK) is added. Following a two-minute incubation, the samples were centrifuged at 400g for five minutes, after which, the supernatant was discarded, and the pellet was resuspended in EasySep (StemCell, Canada) buffer to the 50 mL mark. The samples were spun again at 400g for five minutes, and the supernatant was again discarded. Following this, the cells were resuspended in 2.5mL of EasySep buffer, and 10µL aliquote was made which is diluted to 1:100 in 990µL of EasySep buffer and was used for cell counting in the QEII (Halifax, N.S.) core lab.

The bone marrow samples were then adjusted to a concentration of $1x10^8$ cells/mL and filtered through a 70µM Filcon sterile filter into a 14mL round bottom tube. The CD138 positive selection cocktail (EasySep) was then added to the bone marrow samples at a volume of 50µL per mL of sample, gently mixed, and incubated for 15 minutes. During this incubation, magnetic positive selection particles (EasySep) were homogenized, and after the incubation, were added to the bone marrow samples at a volume of 50µL per mL of sample and mixed gently via pipetting. After a 10 minute incubation, the volumes were then adjusted to 5 or 10 mL using EasySep buffer, depending on if the initial bone marrow was less, or greater than 1 mL in volume, respectively. Samples were then placed into the magnetic stand, and incubated for 5 minutes, after which, the supernatant was poured off into a 50mL conical tube. The previous two steps were then repeated 2 more times. The supernatant solution, which contains the CD138 negative portion,was then centrifuged at 400g for five minutes, supernatant discarded, and the cell pellet resuspended in 1 mL of EasySep buffer. The CD138 positive cells were washed from the sides of the round bottom tube with 1 mL of EasySep buffer. 1:30 dilutions were then prepared from 10 µL of each the negative and positive fractions, and assessed by the core lab for cell number. Samples were then cryopreserved at a concentration of $0.5x10^7$ cells/mL in either 0.6 mL of RLT buffer and stored at -80 for subsequent nucleic acids assessments, or in freezing media and stored in liquid nitrogen for subsequent functional studies.

2.3 DNA and RNA Extraction

DNA and RNA were extracted using a Qiagen AllPrep kit (Qiagen, Hilden, Germany) per manufacturer instructions. Briefly, 700 µl of sample was loaded onto spin columns, cells were lysed and samples homogenized using the QIAShredder column

(Qiagen, Hilden, Germany), and beta-mercapto ethanol supplemented RLT buffer (Qiagen, Hilden, Germany) and samples were spun into a collection vial. For DNA purification, the homogenized lysate was placed into an ALLPrep DNA spin column (Qiagen, Hilden, Germany) and spun down. The eluted volume was subsequently used for RNA extraction. The DNA spin column was then washed sequentially with AW1 and AW2 buffer (Qiagen, Hilden, Germany), and DNA was eluted using 100 µl EB buffer (Qiagen, Hilden, Germany). For RNA purification, equal parts 70% ethanol and sample eluted during the initial DNA spin-down were added to the RNeasy spin column (Qiagen, Hilden, Germany), and spun down. Then 700 µl of RWI buffer (Qiagen, Hilden, Germany) was added, followed by two washes with 500 µl of EB1 buffer (Qiagen, Hilden, Germany. The RNA was then collected in 50 µl of RNase-free water.

2.4 DMG26 Library Preparation

Libraries were prepared using AmpliSeq for Illumina On-Demand, Custom, and Community Panels Reference Guide (Illumina, California), and the TruSeq Custom Amplicon Low Input Kit Reference Guide. Library quantitation was performed via a Bioanalyzer, and samples were normalized and pooled, and sequenced on a V3 flow-cell to 1000x coverage.

The TruSeq protocol was performed on 48 samples using 40ng of DNA from each sample at a concentration of 10ng/µl. The design included 1090 amplicons, and hence, per Illumina guidelines, the amplification was performed in 25 polymerase chain reaction (PCR) cycles. The AmpliSeq protocol was performed on 33 samples, in a two-pool manner using 40ng of DNA from each sample at a concentration of 10ng/µl. The panel design included 640 amplicons, and hence, per Illumina guidelines, the amplification was performed in 28 PCR cycles. Libraries were quantitated via a Qubit, and samples were normalized and assessed for quality on a nano flow cell. Pool normalization was adjusted in accordance with read-depth per sample on the nano flow cell, and then sequenced on a V3 flow-cell at 1000x.

2.5 Bioinformatics Versions

The bioinformatic software used herein is well described in the related methods sections, chapters 3.3.3, 3.3.5, and 4.3.3. The version number or date accessed for each software is indicated in ***Table 2.1***. In addition, all data were visualized in R (version 4.0.2) using ggplot, cowplot, pCOR, ComplexHeatmap, and Circlize.[336–339]

Table 2.1: Bioinformatic software used

Software	**Version/Data accessed**
BWA-mem [340]	0.7.13-R1126
Vardict [329]	1.4
FreeBayes [332]	1.0.2
Pindel [326]	0.2.5B8
Mutect [330]	3.1-0-G72492BB
Platypus [331]	0.8.1
Scalpel [328]	0.5.3
SNPeff [341]	4.2
Picard (Picard Toolkit, Broad)	1.141
SAMtools [342]	1.3
VCFanno [343]	0.0.11
rfPred [344]	1.28.0
GATK3 [345]	3.4-46-gbc02625
SIFT [346]	(accessed using rfPred)
MutationTaster [347]	(accessed using rfPred)
Polyphen2 [348]	(accessed using rfPred)
FATHMM-XF [349]	Accessed August 13th, 2020
spliceAI [350]	1.3.1
CovCopCan [351]	1.3.3
Survminer (Kassambara, A., 2020)	0.4.9
Survival [352]	3.2-9
QDNAseq [334]	1.24.0
svABA [308]	1.0.1
LUMPY [310]	0.2.13
GRIDSS [309]	2.10.1
MANTA [335]	1.6.0
GATK4 [345]	4.1.9.0
Circlize [337]	0.4.12.1004
ComplexHeatmap [336]	2.5.6
pROC [338]	1.17.0.1
ggplot2 [339]	3.3.3
Cowplot (Wilke, O., 2020)	1.1.1

Chapter 3 DMG26: A Targeted Sequencing Panel for Mutation Profiling to Address Gaps in the Prognostication of Multiple Myeloma

3.1 Abstract

Multiple Myeloma presents with numerous primary genomic lesions that broadly dichotomize cases into hyperdiploidy or IgH translocated. Clinically, these large alterations are assessed by FISH for risk stratification at diagnosis. Secondary focal events, including indels and SNPs, are also reported, however, their clinical correlates are poorly described, and FISH has insufficient resolution to assess many of them. In this study, we examined the exonic sequences of 26 genes reported to be mutated in >1% of myeloma patients using a custom panel. We sequenced these exons to approximately ~1000x in a cohort of 76 patients from Atlantic Canada with detailed clinical correlates and in four multiple myeloma cell lines. Across the 76 patients, 255 mutations and 33 focal-copy number variations were identified. High-severity mutations and mutations predicted by FATHMM-XF to be pathogenic identified patients with significantly reduced progression free survival. These mutations were mutually exclusive from the Revised-International Stating System (R-ISS) high-risk FISH markers, and were independent of the International Staging System stage and all biochemical parameters of the R-ISS. Applying our panel to patients classified by FISH to be standard-risk successfully reclassified patients into high- and standard-risk groups. Furthermore, three patients in our cohort each had two high-risk markers; two of these three went on to develop plasma cell leukemia, a rare and severe clinical sequela of multiple myeloma.

3.2 Introduction

Multiple myeloma (MM) is the second most common hematological cancer worldwide and despite recent advances in therapies, overall survival (OS) of patients remains poor.[1] The cancer progresses from the preclinical stages of monoclonal gammopathy of undetermined significance (MGUS) and smoldering MM (SMM) to overt MM; rarely secondary plasma cell leukemia (PCL) ensues, which has a remarkably poor prognosis.[353] Clinical courses vary dramatically between patients; this variation is attributed to the remarkable heterogeneity of genetic alterations which underpin MM, thus highlighting the clinical importance of MM's genomic landscape.[87,354]

In 2015, the International Staging System (ISS) for MM was revised (R-ISS), to incorporate genetic abnormalities via fluorescent *in situ* hybridization (FISH) interrogations.[24] High-risk R-ISS FISH findings include 17p deletions and translocation t(4;14) or t(14;16). However, genomic alterations that are beyond the scope and resolution of FISH may have greater predictive value on patient outcomes.

Recent whole-genome and whole-exome sequencing (WGS, WES) studies have highlighted the scale, prevalence and associations of such genetic alterations in MM.[6,9,13,86,94] Subsequently, MM genetic categorization is now being redefined beyond the classical hyperdiploid and IgH translocation subgroups, and the impact of smaller genetic alterations on clinical outcomes are emerging.[6,9,10,13,86,94,95,355] However, these findings have not been effectively translated into the clinical arena as WGS and WES are cost-prohibitive and bench to bedside turnaround times are impractical, highlighting the need for high resolution, clinically viable genomic interrogations

We previously demonstrated the superiority of ultra-low-depth WGS over FISH to detect copy number variations (CNVs).[26] This approach does not, however, resolve indels or single nucleotide variants (SNVs). A few myeloma specific panels have been investigated, including M3P, M3Pv2.0, and myTYPE.[12,20,21,355,356] The M3Pv2.0 targets 77 genes in commonly affected pathways or that are drug targetable. It has reported mutations in 11 of these genes to significantly impact progression-free and overall survival (PFS and OS), *STAT3* being chief among them.[21] Despite this, widespread clinical adoption has not occurred, possibly due to unclear clinical utility.

Here, we present a MM-specific targeted-sequencing approach using a custom-designed 26 gene panel, the DMG26, that is applicable in a standard clinical molecular laboratory. We demonstrate that the DMG26 captures prognostically relevant genomic abnormalities which are currently not assessed for in clinic.

3.3 Methods

3.3.1 Sample Acquisition

Patient bone marrow samples were processed as described previously.[26] Briefly, bone marrow was collected from patients with plasma cell dyscrasias at the Victoria General Hospital (Halifax, Nova Scotia) and underwent red cell lysis with ammonium chloride, followed by CD138+ magnetic cell selection (StemCell, Vancouver, Canada) to achieve plasma cell purity of >90% by cytospin. This work was conducted under ethical approval by the Nova Scotia Health Authority (NSHA) Research Ethics Board (#1021520 and #1021397) and patients provided written informed consent for research.

Four MM cell lines were also included in this study: MM1S, KMS-12BM, RPMI-8226, NCI-H929 (ATCC). Cell lines were maintained in suspension in RPMI supplemented with penicillin/streptomycin and 10% FBS and confirmed to be free of mycoplasma contamination.

3.3.2 DMG26 Panel Design

Mutation data from published WES and WGS studies in MM were reviewed and compared.[9,13,16]Across these, genes that were found to be mutated in greater than 1% of the myeloma patient population studied were included (25 genes), as well as *MYC*. These genes have previously been implicated in MM pathogenesis, or reported as driver genes.[9,13,16,94] The chromosomal loci of the selected genes as well as the standard FISH probe loci are shown in **Figure *3.1***. Using Illumina DesignStudio, we designed a custom panel to target exons of these 26 genes. Panel designs are described in Supplementary Data files (AmpliSeq_Manifest.txt, TruSight_Manifest.txt, exons.bed).

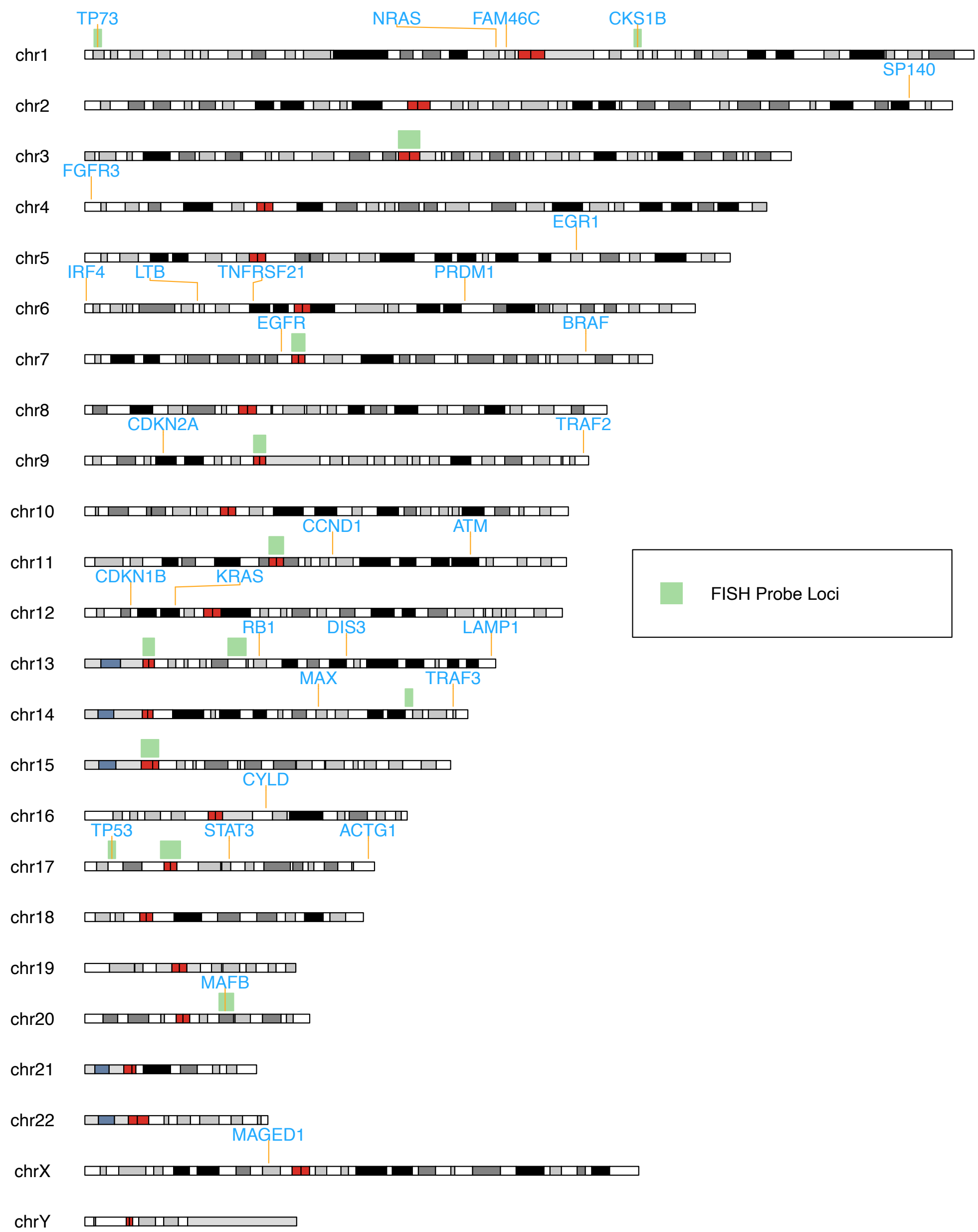


Figure 3.1: DMG26 genes minimally overlap with standard FISH assessed loci
Ideogram showing FISH probe loci (Green) and DMG26 gene loci (names in Blue, position pointer in Yellow)

3.3.3 Sequencing and Variant Calling

DNA library preparations for the DMG26 panel were performed per Illumina TruSight© and AmpliSeq© for Illumina custom panel reference guides. Supplementary Table 1 shows each of the library preparations. Libraries were sequenced at 2x150bp in two runs on an Illumina MiSeq at an average depth of ~1000x. FastQ files were analyzed using an in house bioinformatic pipeline described previously.[325] In brief, reads were aligned by BWA-mem to GRCh37, and variants were called in an ensemble approach using Pindel, Mutect, Vardict, Freebayes, Platypus, and Scalpel and annotated against ClinVar, COSMIC, and other databases using SNPeff and VCFanno.[326,328–332,340,341,343,357,358] Variants were then filtered to include those with at least 20 supporting reads, a variant allele frequency higher than or equal to 10%, two or more supporting callers, and a depth greater than 250x. Variants were also excluded if they were common to >20% of samples in a run and at a VAF less than 1 standard deviation above the mean VAF for the given variant in the run. Filter passing variants were then manually reviewed. Variants were scored for pathogenicity using rfPred, SIFT, MutationTaster, Polyphen2, FATHMM-XF, and SPLICEAI.[344,346–350] Focal copy number variations were called using CovCopCan.[351] Sequencing depth was assessed using SAMtools.[342]

3.3.4 Clinical data

Patient laboratory data, including albumin, beta-2 microglobulin, LDH, bone marrow plasma cell burden, serum M-protein quantity, Ig heavy and light chain type and quantity, serum free light chain ratio, and FISH data, coinciding with the time of bone marrow acquisition, were collected from NSHA laboratory information system. Patient clinical data, including age, sex, diagnosis at time of bone marrow acquisition, therapies received, follow-up period, stage, and time of events were also collected from NSHA hospital information system. Events were defined as relapse, progression, or death and were collected from time of bone marrow acquisition to study end points.

3.3.5 Statistical Analysis

Summary statistics were used to describe the distribution of mutations, clinical, and molecular features across our cohort. Using univariate Cox proportional hazard ratios, Kaplan-Meier and Log-Rank tests from R packages Survival and Survminer, we assessed the association of mutational status with progression-free survival.[359,360] Independence between mutational signatures and standard biochemical prognostic parameters was assessed using the chi-squared test and Wilcoxon-P. Comparison of stratification schemes was performed using Henderson's C. All statistical analyses were performed using R and Python.

3.3.6 Availability of Data

The datasets generated and analysed during this study are available from the Dalhousie Pathology Biobank, BioBank@nshealth.ca. All programing scripts used in this work are available at https://gitlab.com/gaston-lab-genomics/myeloma-amplicon-risk.

3.4 Results

3.4.1 Cohort Description

Seventy-seven patient samples (from 76 patients) and 4 cell lines were included in our study. Our patient samples comprised 20 MGUS, 3 SMM, 52 MM, and 1PCL and two thirds of the cohort were males **(Table 3.1**). Sample MM13 was taken at diagnosis while MM40 was taken at relapse for the same patient. The majority of patients received a combination of Cyclophosphamide, Bortezomib, and Dexamethasone (CyBorD) as first line therapy (44 of 57 treated individuals). The median follow-up time was 19 months (range: 0.4-42), and 30 patients had an event within the follow-up period. Molecular and demographic features for our cohort are summarized in **Table 3.1** and **Appendix *Figure 1***.

3.4.1 Comparison of TruSight and AmpliSeq Panels

The TruSight design targeted a total region of 98859 bp, while the AmpliSeq targeted a total region of 116793 bp. The overlap of the TruSight design with the AmpliSeq design was 89.4%, while the overlap of the AmpliSeq design with the TruSight design was 75.7%. The proportion of coding nucleotides within the target genes captured by our panels was 98.7%, and 99.3% for the TruSight and AmpliSeq designs, respectively. The average coverage of coding nucleotides across samples was 1081x and 1026x on the TruSight and

AmpliSeq panels, respectively (***Figure 3.2***) Bioinformatic variant calling resulted in 210 variants across 45 patient samples and 3 cell lines using TruSight amplicon data and 84 variants across 32 patient samples and one cell line using AmpliSeq amplicon data (**Appendix *Table* 1**; **Appendix *Table* 2**).

Table 3.1: Summary of patient demographics, laboratory and molecular data, therapies, and follow-up

N = 76		
ISS Stage	I	3 (4%)
	II	25 (33%)
	III	18 (24%)
	Unknown	30 (39%)
Diagnosis at time of BM	MM	50 (65%)
	MM relapse	13 (17%)
	MGUS	10 (13%)
	SMM	3 (4%)
	PCL	1 (1%)
Therapies	ASCT	23 (30%)
	Proteasome Inhibitor	47 (62%)
	Monoclonal Ab	9 (12%)
	Immunomodulatory	35 (46%)
Paraprotein type	IgG Kappa	30 (39%)
	IgG Lambda	15 (20%)
	IgA Kappa	8 (11%)
	IgA Lambda	11 (14%)
	IgD Lambda	1 (1%)
	IgM Kappa	1 (1%)
	Non-secretory Kappa	7 (9%)
	Non-secretory Lambda	3 (4%)
Sex	Male	51 (67%)
	Female	25 (33%)
FISH	High-risk	12 (16%)
	Standard-risk	36 (47%)
	Unknown	30 (39%)
Age	Median, Range (Years)	70, 33-86.5
PFS	Median, Range (Months)	12, 1-32
LDH	Median, Range (IU/L)	164, 45-608
Follow-up	Median, Range (Months)	19, 0.4-42
BM PC infiltration	Median, Range (%)	45, 0-95
Albumin	Median, Range (g/L)	33, 16-46
Serum M-protein	Median, Range (g/L)	25.1, 0-105
B2 Microglobulin	Median, Range (nmol/L)	365, 135-2318.6
SFLC I/U	Median, Range	80, 0-4431

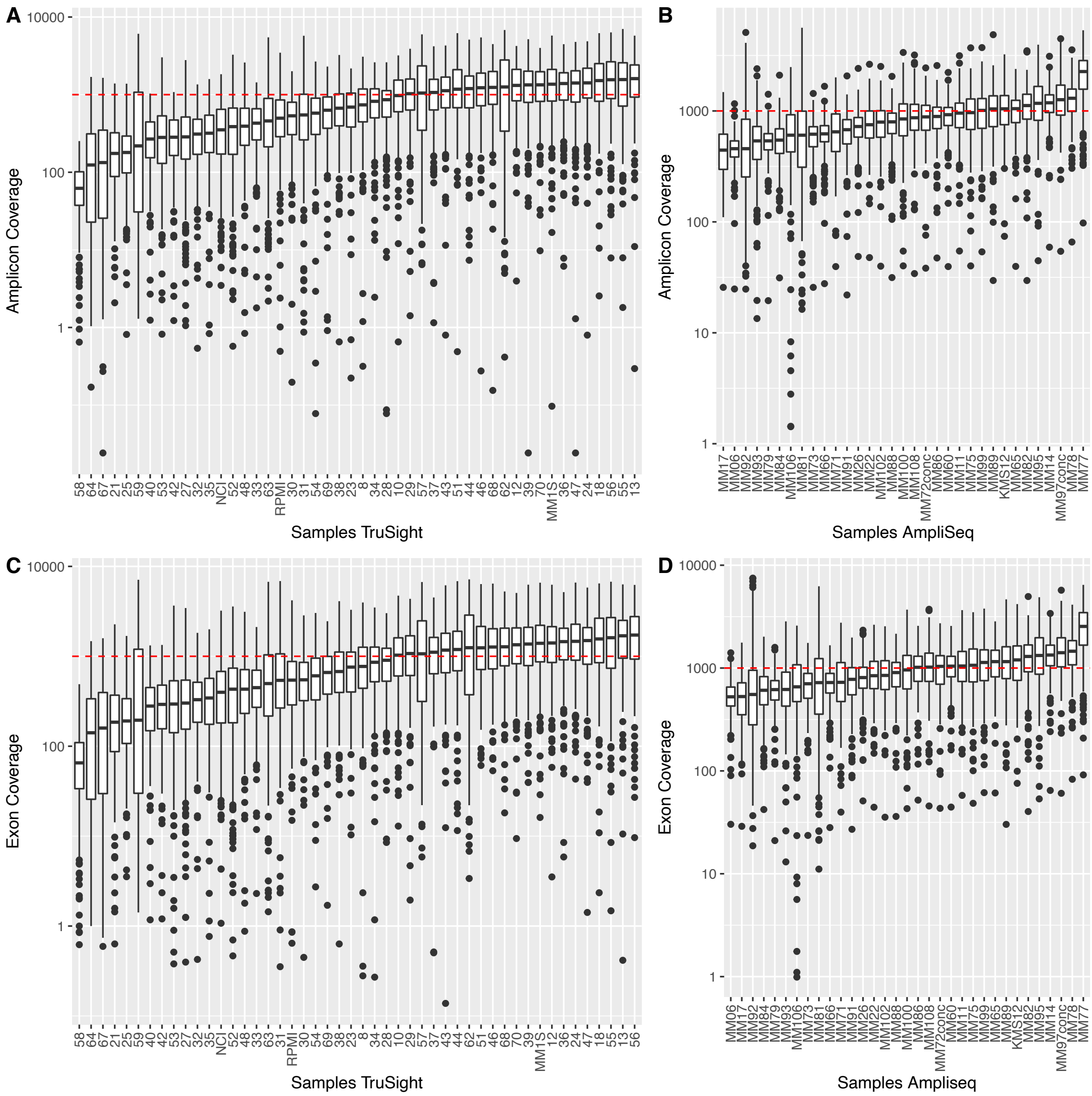


Figure 3.2: Coverage by amplicon and exon for TruSeq and AmpliSeq
Boxplots indicating sequencing depth. Y-axis is log-scaled, and the red line indicates 1000x. **A**) Boxplots of amplicon sequencing depth by patient for samples prepared using TruSight. **B**) Boxplots of amplicon sequencing depth by patient for samples prepared using AmpliSeq. **C**) Boxplots of exon sequencing depth by patient for samples prepared using TruSight. **D**) Boxplots of exon sequencing depth by patient for samples prepared using AmpliSeq.

3.4.2 Analysis of Variant Data

Across our cohort, 294 variants were identified, 39 of which were in the four cell lines (**Figure *3.3***; **Table *3.2***; **Appendix *Figure 1***). All of our cell lines had mutation data catalogued within COSMIC's Cell Line Project which reported 14 verified variants within regions targeted by our panel, 13 (93%) of which we successfully identified (**Table *3.2***). Within our patient samples, 65 patients collectively harboured 255 variants; *ATM* being the most mutated in our cohort (37 variants in 22/77 samples) and *KRAS* the most mutated per kilobase (0.30 mut/kb in 17/77 samples) (**Figure *3.4* A,B**). Low variant allele frequency (VAF) mutations contributed the bulk of variability in mutational burden between patients (**Figure *3.4* C**).

3.4.3 Clinical and Prognostic Value of Mutations

Next, we investigated the clinical associations of the identified variants. In a univariate Cox proportional hazard model, *CDKN1B* was the only gene whose mutational status had a significant correlation with PFS (n = 3, HR 17.21; 95% CI: 3.21-92.14; p = 0.001) (**Figure *3.5***). We then reclassified the mutational status of a gene to require the presence of at least one mutation at a VAF ≥ 20% (**Figure *3.6***). On reassessing the PFS association of mutational status on a per-gene basis, we again identified *CDKN1B* (n = 2, HR =19.87; 95% CI: 3.79-104.11; p < 0.001), as well as *PRDM1* (n = 3, HR = 6.27; 95% CI: 1.79-21.99; p = 0.004) to negatively correlate with PFS (**Figure *3.7***). These genes were only mutated in 2, and 3 patients, respectively, and thus did not capture the majority of risk across our cohort. We then assessed the association of mutation type and severity by sequence ontology[47], across all 26 genes in the panel, with PFS. Most of the mutation types which had a significant correlation with PFS are considered high-severity mutations (**Figure *3.8* A**). Consistently, harbouring at least one high-severity mutation with a VAF ≥ 20% (**Figure *3.6***) was significantly associated with reduced PFS (n =15, HR = 3.0; 95% CI: 1.33-6.72; p = 0.008), and captured risk in a greater proportion of our cohort (**Figure *3.8* B,C**).

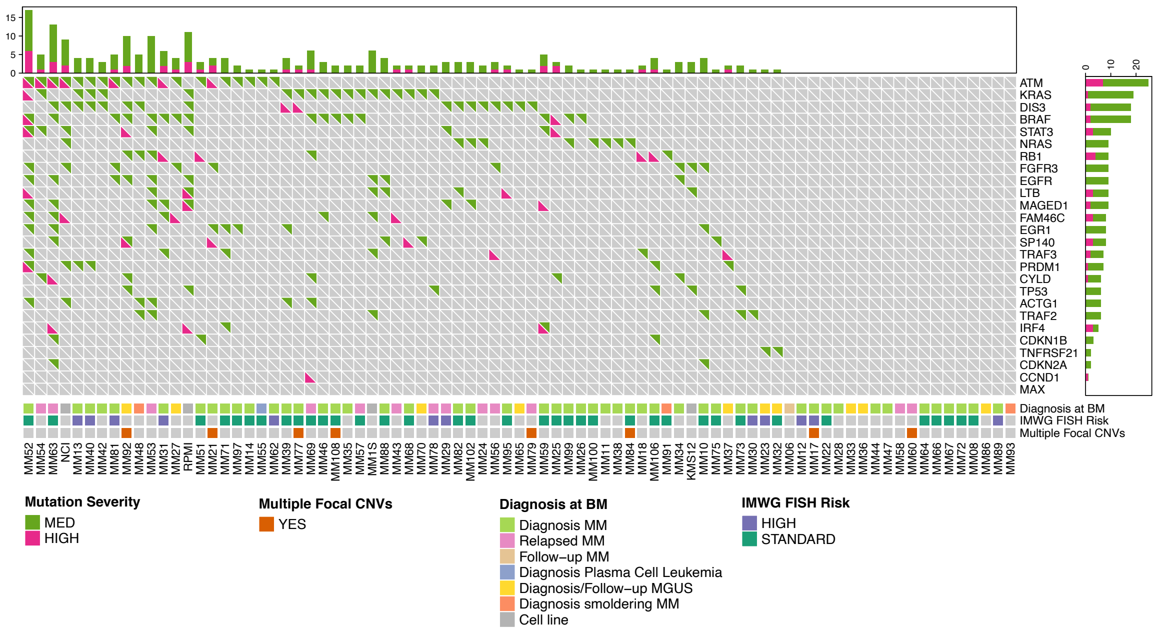


Figure 3.3: Oncoprint of DMG26 identified mutations against patient diagnosis, FISH risk, and focal CNVs

Mutations which are medium or high severity are shown in Green or Pink, respectively. High severity mutations include stop-loss, splice acceptor, splice donor, stop-gain, and frameshift variants. Medium severity mutations include in-frame deletions, in-frame insertions, splice region variants, and missense variants. Low severity mutations, which include stop retained variants, intron variants, and synonymous variants are not shown. Rows are genes, columns are samples. The top bar-plot indicates distribution of mutation severity by patient. The bar-plot on the right indicates distribution of mutation severity by gene. Below the Oncoprint, IMWG FISH risk is indicated in Green (standard-risk) and Purple (high-risk); multiple CovCopCan focal copy number variants are indicated in Orange; and patient diagnosis at time of bone marrow acquisition is indicated in pale Green (MM diagnosis), pale Pink (MM relapse), Turquoise (MM progression), Beige (MM follow-up), pale Purple (MGUS diagnosis), Yellow (MGUS follow-up), Orange (SMM diagnosis), Grey (Cell Line).

Table 3.2: Cell line data matches COSMIC reports

List of all verified mutations reported by COSMIC within cell lines RPMI-8226, NCI-H929, MM1S, and KMS-12-BM. Corresponding mutational status reported by our panel is indicated.

Gene	Codon Alteration	Cell_Line	Panel Called
LTB	c.244G>C	KMS-12-BM	Yes
LTB	c.218A>G	KMS-12-BM	Yes
TP53	c.1010G>T	KMS-12-BM	Yes
EGFR	c.2749G>C	MM1S	Yes
KRAS	c.35G>C	MM1S	Yes
TRAF3	c.1607_1633del	MM1S	Yes
ATM	c.1039G>A	NCI-H929	No
NRAS	c.38G>A	NCI-H929	Yes
TENT5C	c.278_279insC	NCI-H929	Yes
EGFR	c.2252C>T	RPMI-8226	Yes
KRAS	c.35G>C	RPMI-8226	Yes
LTB	c.208+1G>A	RPMI-8226	Yes
LTB	c.208G>A	RPMI-8226	Yes
TP53	c.853G>A	RPMI-8226	Yes

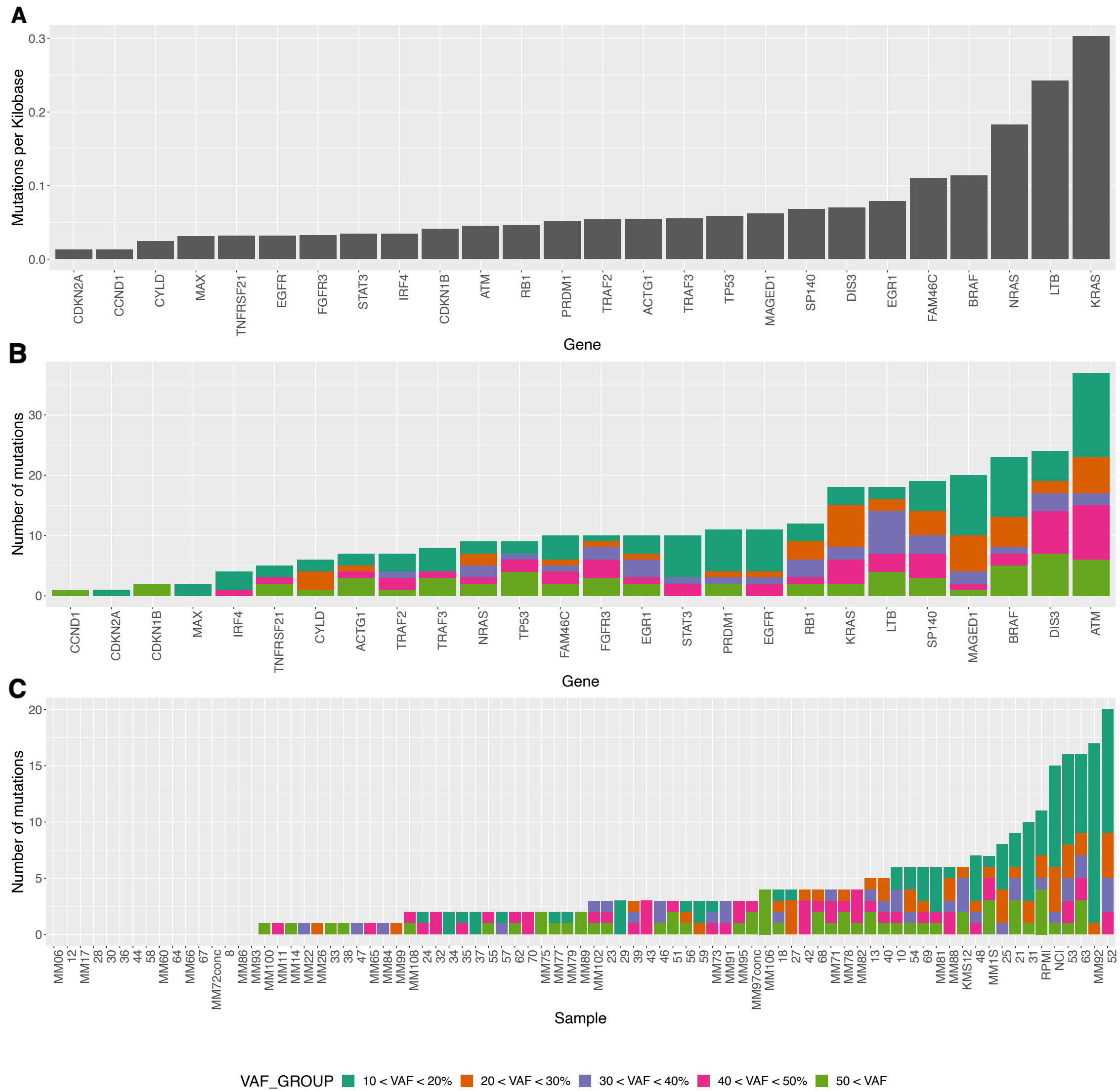


Figure 3.4: Mutation distribution per gene and per sample in our cohort

A) Mutations per sequenced kilobase by gene. B) Number of mutations by gene with variant allele frequency breakdown. C) Number of mutations by sample with variant allele frequency breakdown. For all indications of variant allele frequency, the highest reported value across all somatic variant calling algorithms was used.

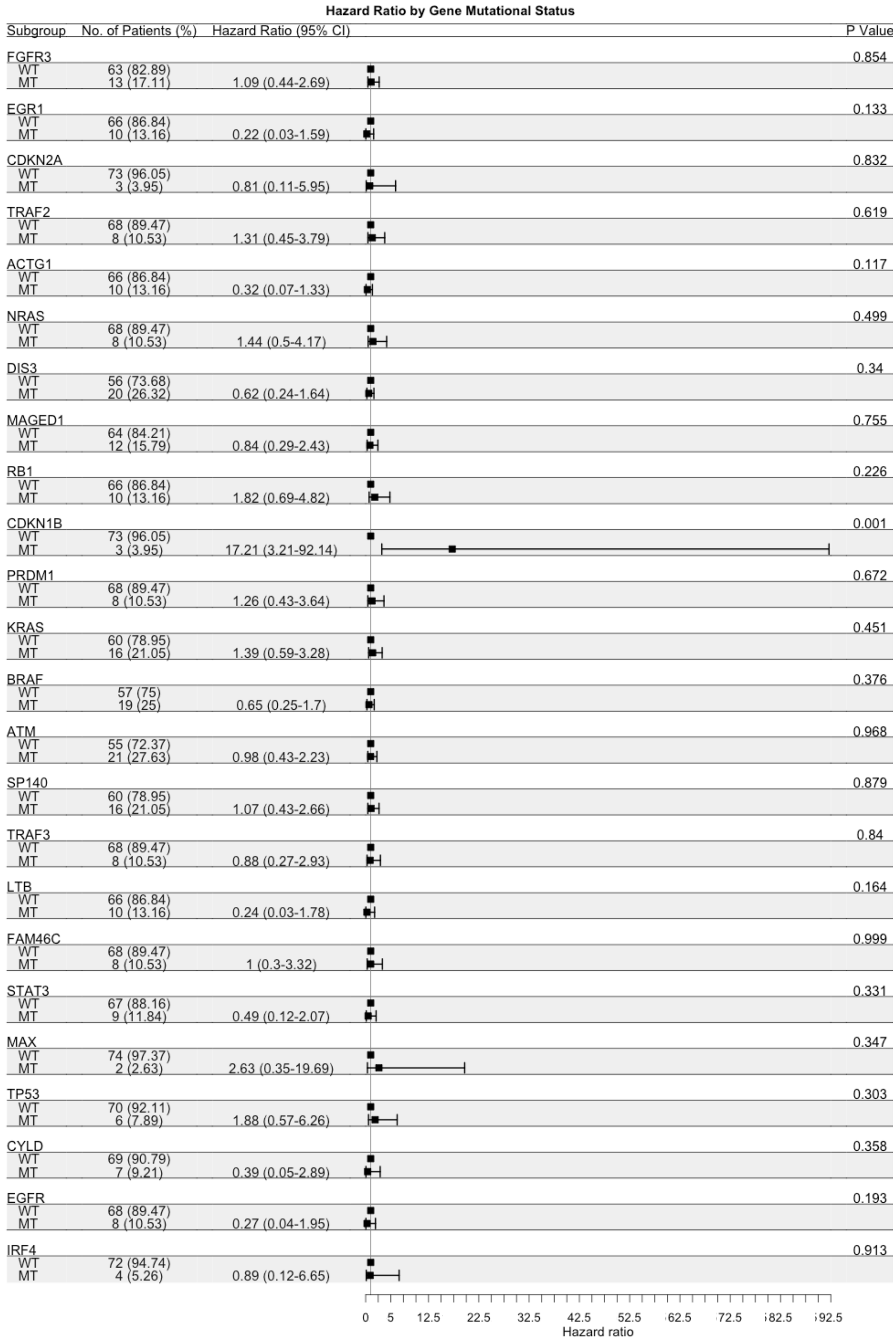


Figure 3.5: Hazard ratio of gene mutation status
Forrest plot showing Cox-proportional hazard of harbouring at least one mutation in the indicated gene.

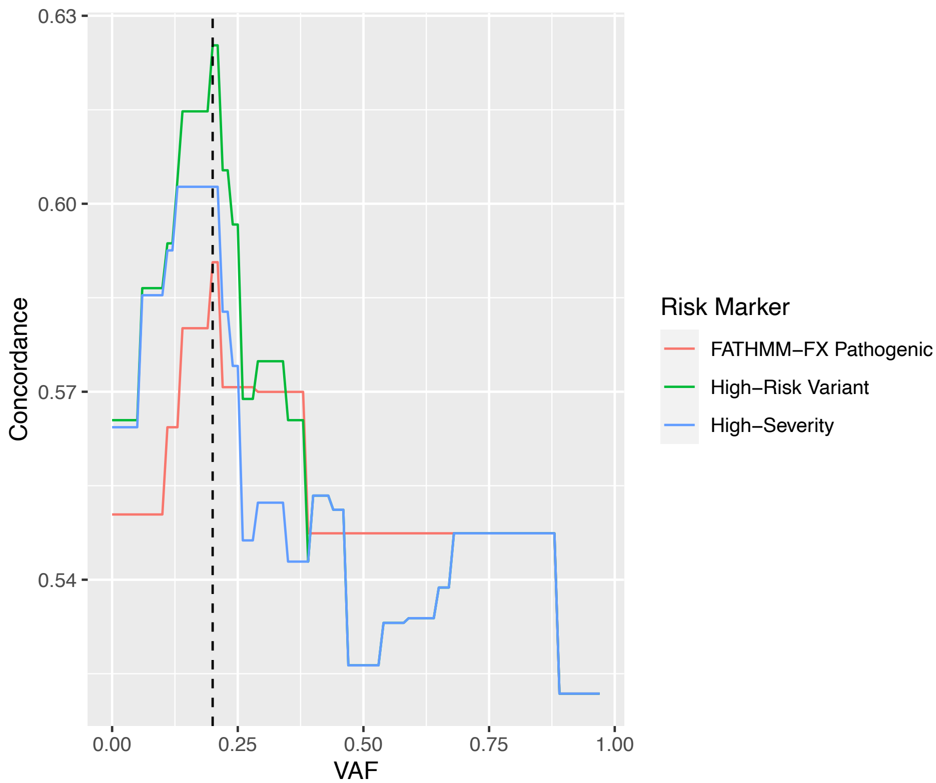


Figure 3.6: Concordance maximized at 20% variant allele frequency
Concordance for hazard assigned by high-severity mutations (Blue), FATHMM-FX predicted pathogenic mutations (red), or both ('high-risk' mutations, Green). Concordance was assessed for each at VAF cut-offs between 0 and 1 at 0.01 increments.

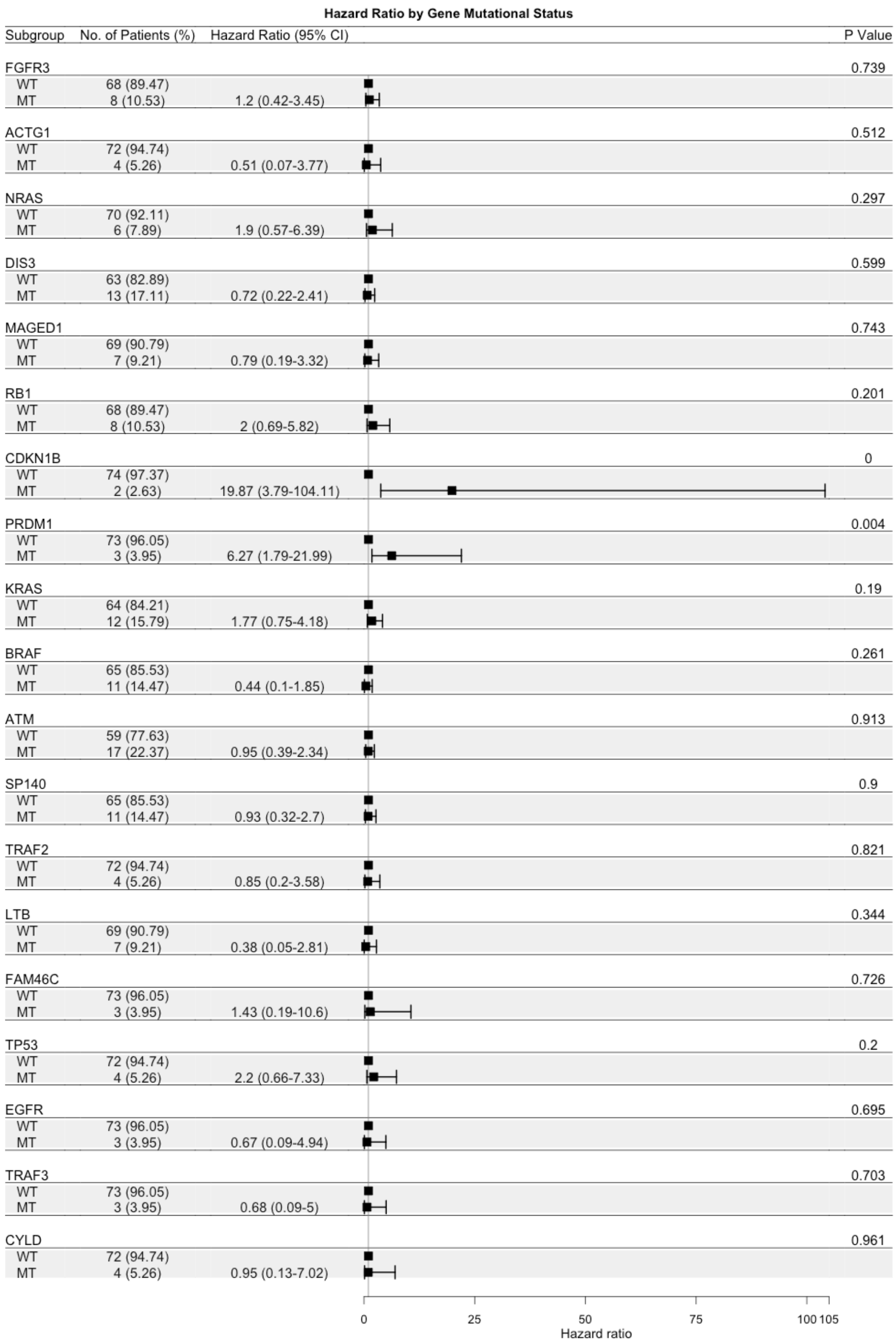


Figure 3.7: Hazard ratio by clonal gene mutational status

Forrest plot showing Cox-proportional hazard of harbouring at least one mutation in the indicated gene with a VAF above 20%.

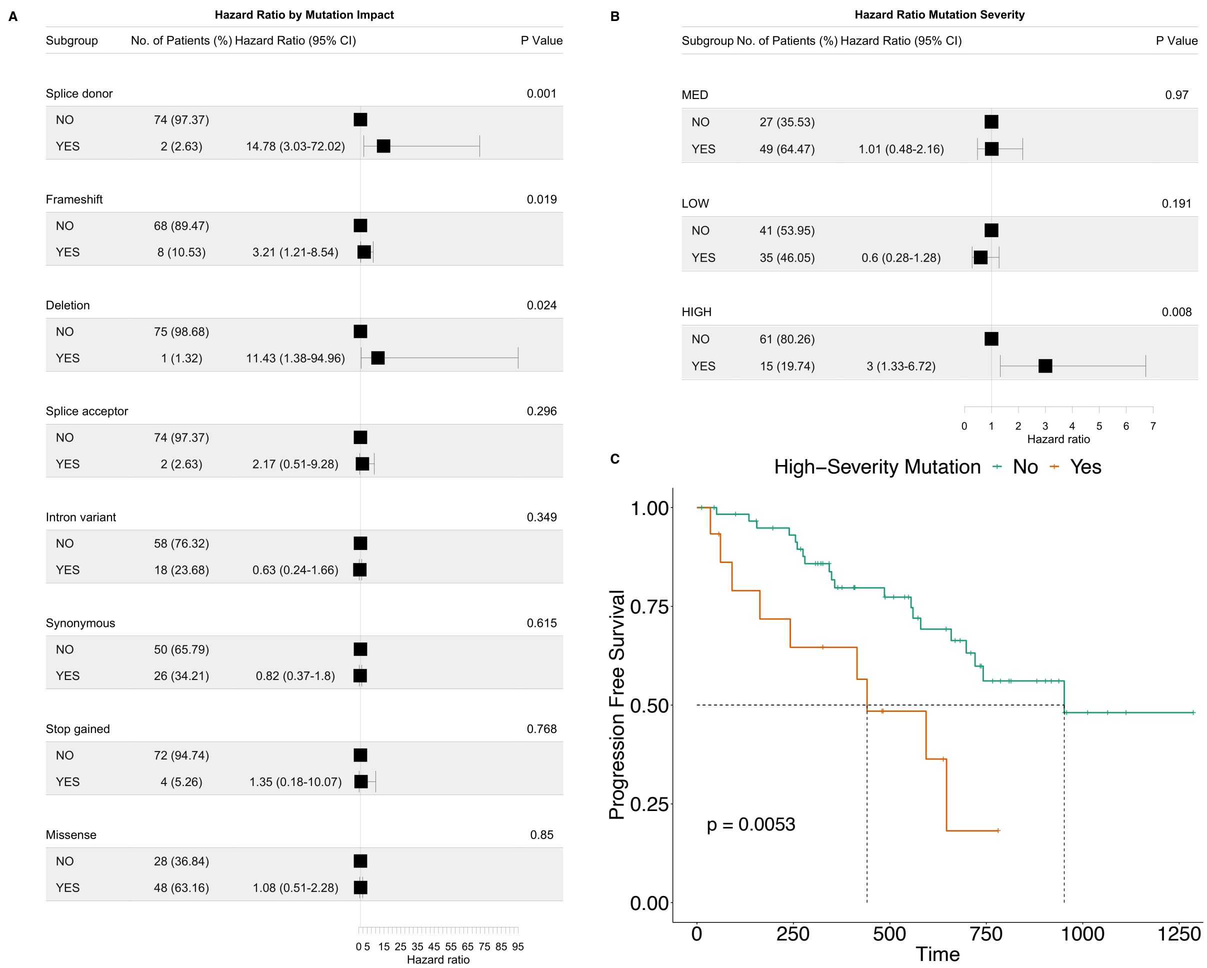


Figure 3.8: High-severity mutations in panel-targeted genes significantly impact PFS A) Forrest plot showing Cox-proportional hazard of harbouring at least one mutation of the indicated impact with a VAF above 20%. B.) Forrest plot showing Cox proportional hazard of harbouring at least one mutation of the indicated severity above a 20% VAF. C.) Kaplan-Meier plot of the different clinical courses between patients harbouring at least one mutation of high severity above a 20% VAF (Orange, n = 15), and those that do not harbour any high severity mutations above a 20% VAF (Green, n = 61). Time is in days. The log-rank p value is indicated.

Examining only high-severity mutations left a large portion of our mutation data uninformative for risk stratification, namely missense mutations (medium severity) which accounted for 155 of our observed variants (**Figure *3.9* A**). We therefore sought to identify which of these mutations may confer additional prognostic value. For this purpose, we considered COSMIC and ClinVar annotations, and scored each variant for pathogenicity with rfPred, SIFT, MutationTaster, Polyphen2, and FATHMM-XF, as well as SPLICEAI which predicts the splicing impacts of mutations. Mutations in COSMIC or those flagged as pathogenic in ClinVar did not significantly correlate with PFS (**Figure 3.10 A,B**). For each algorithm, we assessed the subsequent association of PFS with at least one mutation above the recommended cut-off for predicted pathogenicity. Of these, FATHMM-XF using the upper cut-off of 0.97 performed the best and was hence used for the work herein (**Figure *3.9* B-D**, **Figure 3.10 C-G**). Harbouring at least one FATHMM-XF predicted pathogenic mutation with a VAF ≥ 20% (**Figure *3.6***) had a significant negative correlation with PFS (n = 5; HR = 7.37; 95% CI: 2.7-20.16; p < 0.001) (**Figure *3.9* C,D**).

We then combined these two indicators to define high-risk patients such that a patient is considered high risk if they have one or more mutations that are either high-severity or predicted by FATHMM-XF to be pathogenic and at a VAF ≥ 20%. With this approach, 23/376 mutations were considered high-risk markers, 19 of which were in patient samples. This effectively classified 16 of 76 patients as high risk with significantly reduced PFS (n = 16; log-rank p = 0.0011; HR = 3.46; 95% CI: 1.6-7.6; p = 0.002) (**Figure *3.3***, **Figure *3.11* A,E**). We applied this stratification scheme to our at-diagnosis, pre-treatment, and non-MGUS sub-cohorts and again found significantly reduced PFS in all groups (**Figure *3.11* B-E**). High-risk mutations were found in 12 genes, with *ATM*, *RB1*, *TP53*, *DIS3*, *FAM46C*, *LTB*, and *MAGED1* each harbouring more than one high-risk mutation (**Figure *3.11* G**). No high-risk mutations were found in our MGUS or SMM patients (**Figure *3.11* H**). Notably, three patients: MM43, MM63, and MM106, harboured 2 high risk mutations each and had a striking decrease in PFS (HR = 15.6; 95% CI: 2.9-83; p = 0.0013) (**Figure *3.11* F**). Of these, MM63 had a follow-up time of only 51 days and did not experience an event in this time. However, MM106 and MM43 both relapsed rapidly at 163 and 91 days, respectively, and both progressed to PCL.

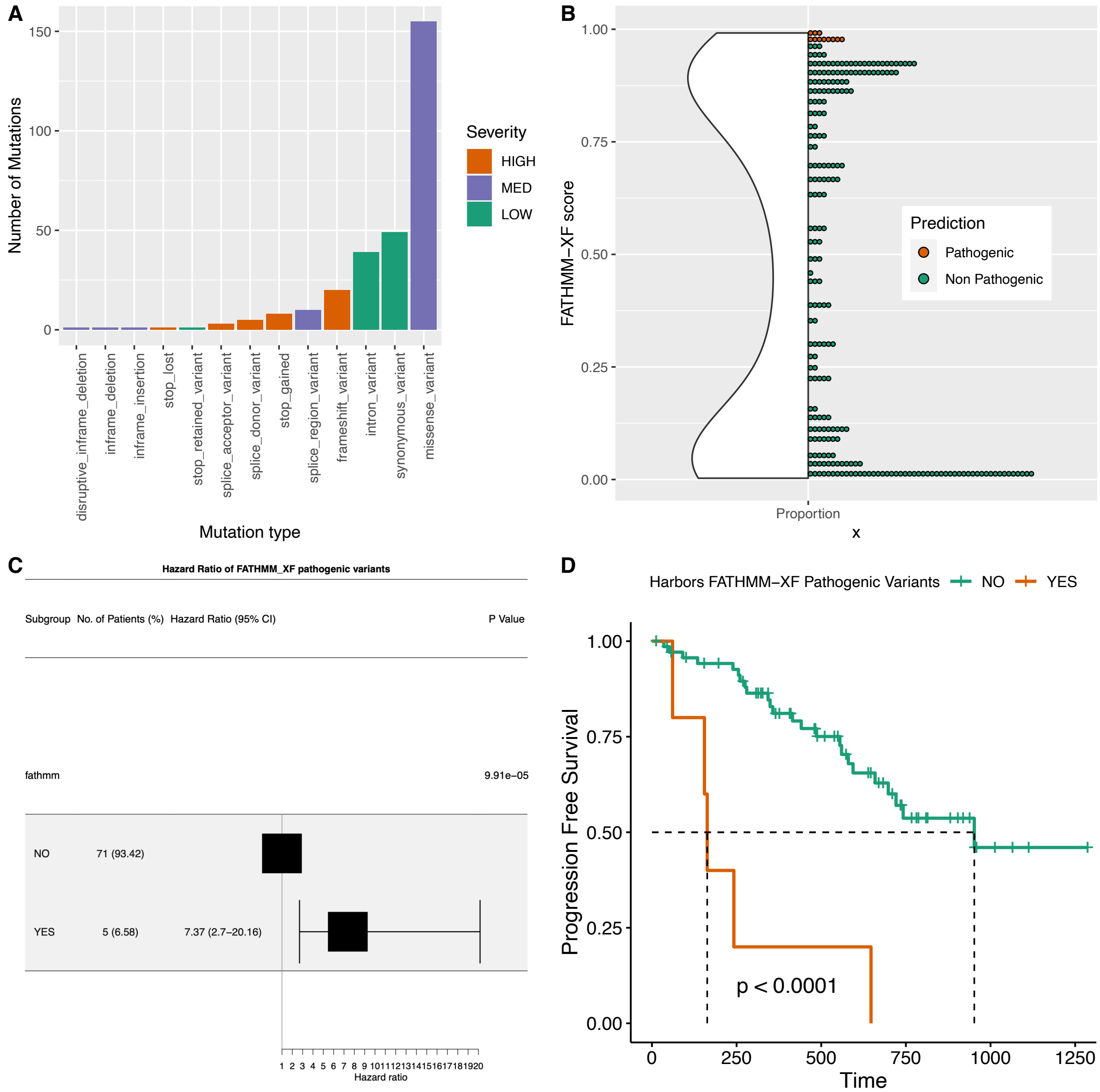


Figure 3.9: Figure 4: FATHMM-XF predicted pathogenic variants in panel-targeted genes significantly impacts PFS

A) Barplot of mutation abundance by impact type. Color indicates impact severity: high, medium (MED), and low are Orange, Purple, and Green respectively. B) Violin plot of FATHMM-XF score of variants. Orange variants have a score above or equal to 0.97, the stringent pathogenic cut-off of FATHMM-XF. C) Forrest plot of Cox proportional hazard of harboring at least one variant above or equal to a FATHMM-XF score of 0.97 and above a 20% VAF. D) Kaplan-Meier plot of the different clinical courses of patients harboring at least on mutation with a FATHMM-XF score above or equal to 0.97 and above a 20% VAF (Orange, n = 5), and those patients who do not harbor a variant scored by FATHMM-XF above or equal to 0.97 and above a 20% VAF (Green, n = 71). Time is in days. The log-rank p value is indicated.

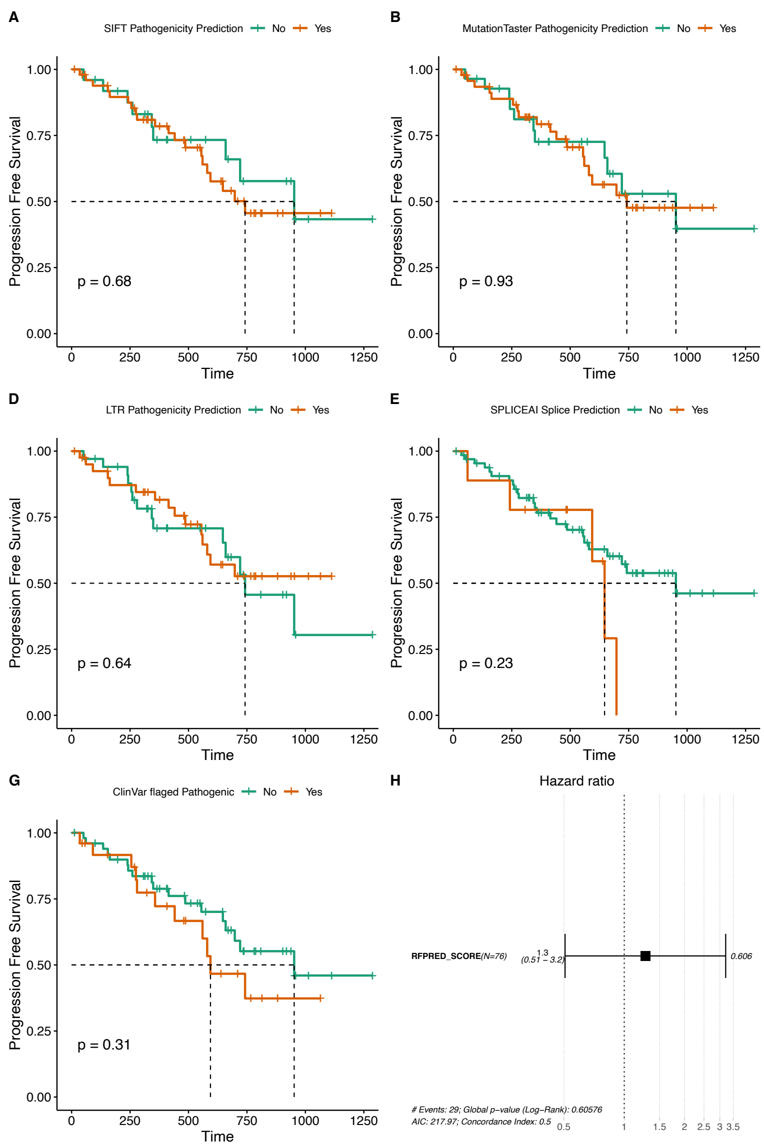


Figure 3.10: Mutational hazard by prediction algorithms, ClinVar, and COSMIC
A) Kaplan-Meier plot of patients harbouring mutations predicted by SIFT to be pathogenic (Orange), and those who do not (Green). **B**) Kaplan-Meier plot of patients harbouring mutations predicted by MutationTaster to be pathogenic (Orange), and those who do not (Green). **C**) Kaplan-Meier plot of patients harbouring mutations predicted by Polyphen2 to be pathogenic (Orange), and those who do not (Green). **D**) Kaplan-Meier plot of patients harbouring mutations predicted by LTR to be pathogenic (Orange), and those who do not (Green). **E**) Kaplan-Meier plot of patients harbouring mutations predicted by SPLICEAE to be splice altering (Orange), and those who do not (Green). **F**) Kaplan-Meier plot of patients harbouring mutations reported in COSMIC (Orange), and those who do not (Green). **G**) Kaplan-Meier plot of patients harbouring mutations flagged in ClinVar as pathogenic (Orange), and those who do not (Green). **H**) Forrest plot showing Cox-proportional hazard for maximum rfPRED score per patient.

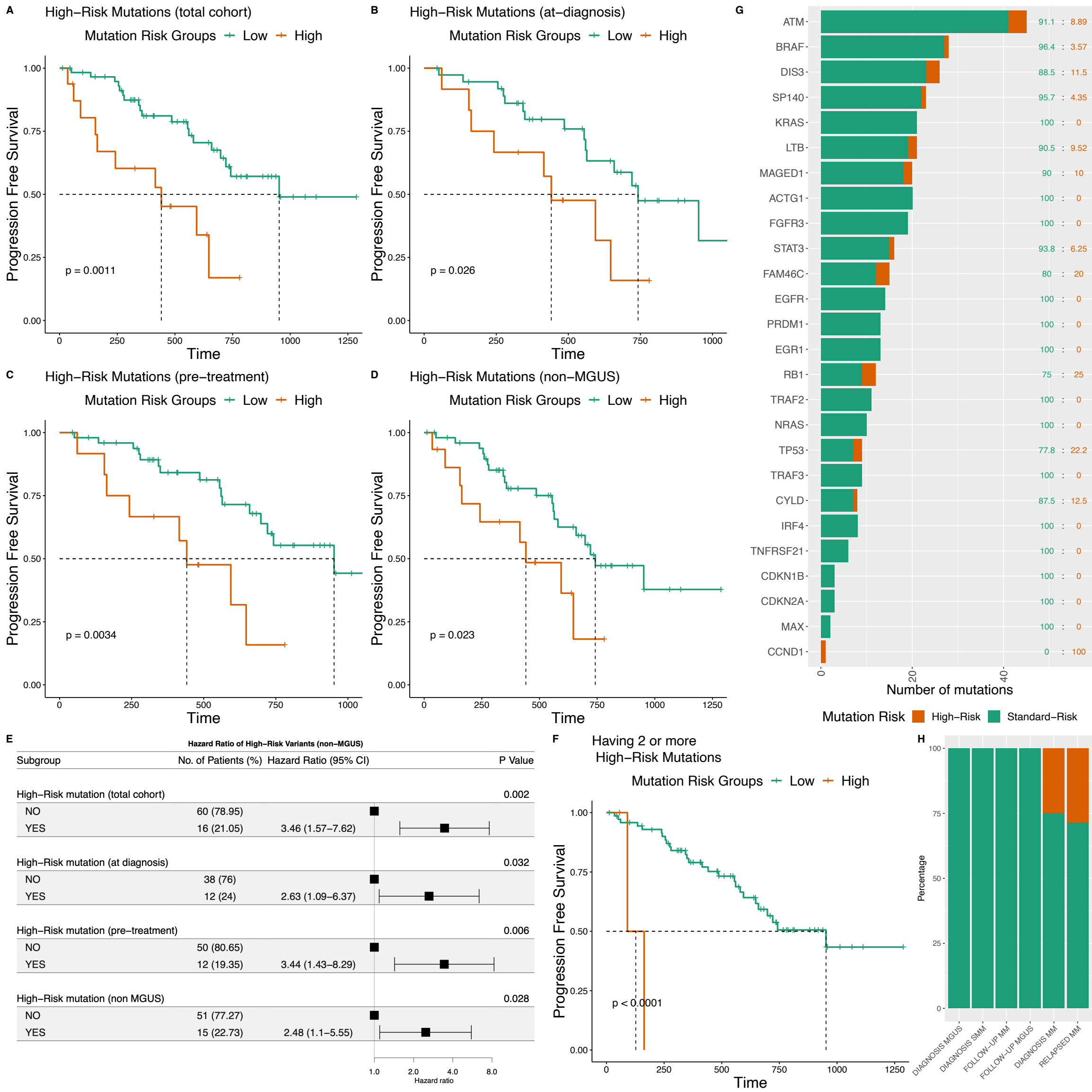


Figure 3.11: High-risk mutations significantly correlate with reduced PFS in both total cohort and diagnostic subgroups

A, **B**, **C**, **D**) Kaplan-Meier Survival plots showing different clinical courses between patients that harbour one or more mutations that are high-severity or predicted by FATHMM-XF to be pathogenic, and are above 20% VAF (Orange), and those who do not (Green) in different disease-stage groups. The log-rank p value is indicated. **A** is total cohort (n = 76), B is at diagnosis of MM (n = 50), C is pre-treatment (MM diagnosis, MGUS, and SMM diagnosis, n = 62), D is non-MGUS patients (n = 66). **E**) Forrest plots of Cox proportional hazard for patients harbouring one or more mutations that are high-severity or predicted by FATHMM-XF to be pathogenic, and are above 20% VAF in different disease-stage groups. **F**) Kaplan-Meier Survival plots showing different clinical courses between patients that harbour multiple mutations that are high-severity or predicted by FATHMM-XF to be pathogenic, and are above 20% VAF (Orange, n = 3), and those who do not (Green, n = 73). The log-rank p value is indicated. G) Proportion of mutations that are high-risk by gene. **H**) Proportion of patients harbouring high-risk variants by reason for bone marrow sample.

3.4.4 Copy Number Calling

Although the DMG26 is designed for SNV and indel calling, using CovCopCan we also analyzed our panel data for focal copy number variations (CNVs); 33 variations were identified across 22 patient samples and 3 CNVs across 3 cell lines (**Table 3.1**). Harbouring 2 or more CNVs significantly reduced PFS by the log rank test (**Figure *3.12* A**), and in a Cox proportional hazard model (n = 8, HR = 3.07, 95% CI: 1.01-9.35, p = 0.048) (**Figure *3.12* B**). Combining focal-CNV data with high-risk mutations enhanced risk stratification, classified 22 patients as high-risk with significantly reduced PFS (n = 22; HR = 4.42; 95% CI: 2.03-9.6; p < 0.001) (**Figure *3.12* C,D**).

3.4.5 Correlation to Clinical Metrics and FISH Data

We assessed the independence of our panel-based risk stratification from other prognostic factors and the ISS. In our cohort, 48 patients had FISH data, 12 of whom harboured R-ISS high-risk markers: t(4;14), t(14;16), del(17), del(17p), and/or del(17p13.1) (**Table 3.1**; **Figure *3.3***; **Appendix *Figure 1***). When re-evaluated based on our DMG26 panel, 12 patients had high-risk mutations and 4 had multiple focal CNVs (**Figure *3.13* A**). Strikingly, the presence of high-risk FISH and high-risk mutations were mutually exclusive, and high-risk findings by our panel outperformed FISH (R-ISS and 1q) in risk classifications, and re-classified FISH standard-risk patients into high- and low-risk groups with significantly different PFS (HR = 3.6; 95% CI: 1.13-11.9; p = 0.031) (**Figure *3.13* A-D; Figure *3.14***). One patient, MM17, was high-risk by FISH (t(14;16)) and harboured 2 panel identified focal-CNVs (**Figure *3.3***; **Figure *3.13* A**). Additionally, 46 patients had ISS staging available at time of bone marrow acquisition: 3 stage 1, 25 stage 2, and 18 stage 3 (**Table 3.1**). This is skewed towards higher stages as incomplete laboratory data can define ISS stage 3, but not stages 1 or 2. Our panel outperformed ISS staging in risk stratification and the occurrence of high-risk mutations was independent of ISS staging (**Figure *3.13* E,F**). Similarly, high-risk mutations were found to be independent of individual laboratory inputs of the ISS and R-ISS algorithms, including beta-2 microglobulin, LDH, and Albumin (**Figure *3.13* G-I**).

Table 3.3: Focal copy number variations

Reported by CovCopCan within our study cohort, using the sequence data from our mutation panel.

Sample	Chromosome	Start	End	Variant	Length (bp)
MM25	chr6	106553047	<DUP>	DUP	106553899
MM24	chr13	48951015	<DUP>	DUP	48954588
MM21	chr6	106553047	<DUP>	DUP	106553899
MM21	chr13	49050792	<DUP>	DUP	73334135
MM52	chr6	106553047	<DUP>	DUP	106553899
NCI	chr6	106553047	<DUP>	DUP	106553899
MM47	chr14	103336486	<DEL>	DEL	103342795
MM44	chr13	48878014	<DEL>	DEL	48923289
MM34	chr6	106553047	<DUP>	DUP	106553899
MM1S	chr7	140477750	<DUP>	DUP	140494515
MM32	chrX	51637270	<DUP>	DUP	51638288
MM84	chr9	21970959	<DUP>	DUP	21974922
MM84	chr9	139802371	<DUP>	DUP	139804595
MM22	chr14	103352474	<DEL>	DEL	103363868
MM82	chr12	25368305	<DUP>	DUP	25380394
MM108	chrX	51637714	<DUP>	DUP	51641751
MM108	chrX	51644447	<DUP>	DUP	51645157
MM99	chr14	103355797	<DEL>	DEL	103372301
MM06	chrX	51637321	<DUP>	DUP	51645157
KMS12	chr6	47220948	<DUP>	DUP	47277277
MM92	chrX	51638323	<DUP>	DUP	51641066
MM92	chrX	51641477	<DUP>	DUP	51645157
MM60	chr13	49047367	<DUP>	DUP	49051659
MM60	chrX	51637321	<DUP>	DUP	51645157
MM17	chrX	51637321	<DUP>	DUP	51638928
MM17	chrX	51639512	<DUP>	DUP	51645157
MM79	chrX	51637321	<DUP>	DUP	51638928
MM79	chrX	51639621	<DUP>	DUP	51640387
MM79	chrX	51640997	<DUP>	DUP	51641792
MM77	chr2	231134181	<DEL>	DEL	231135431
MM77	chr2	231150301	<DEL>	DEL	231155352
MM77	chr2	231175543	<DEL>	DEL	231177480
MM77	chr16	50783469	<DEL>	DEL	50818521
MM73	chrX	51637321	<DUP>	DUP	51645157
MM88	chr13	48954122	<DUP>	DUP	48955736

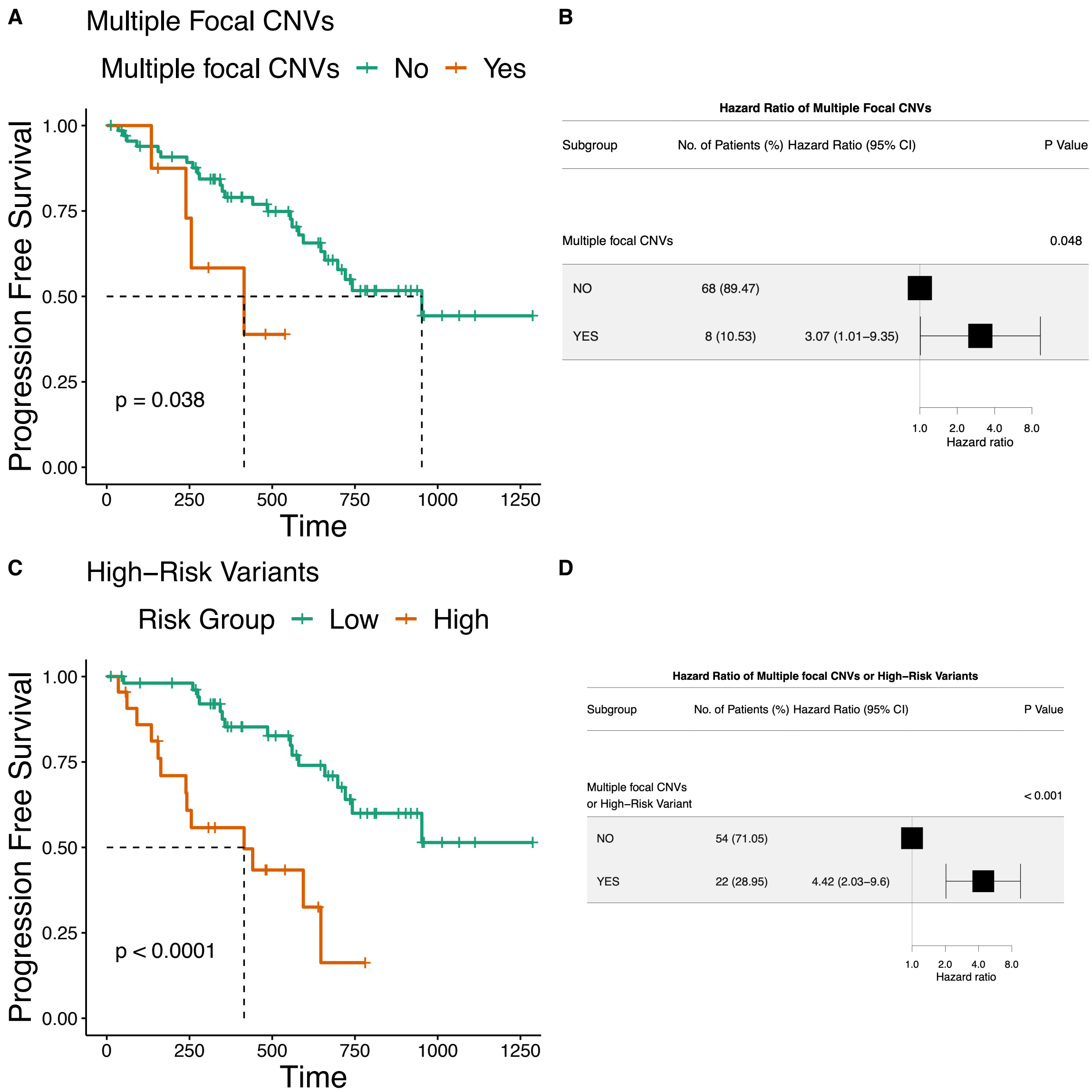


Figure 3.12: Panel called focal-CNVs enhance DMG26 risk stratification
A) Kaplan-Meier survival plot showing different clinical courses for patients harboring multiple focal-CNVs (Orange, n = 8), and those who don't (Green, n = 68). Time is in days. The log-rank p value is indicated. **B**) Forrest plot showing Cox proportional hazard of harboring two or more focal CNVs. **C**) Kaplan-Meier survival plot showing different clinical courses for patients harboring multiple focal-CNVs or high-risk mutations (Orange, n = 22), and those who don't (Green, n = 54). Time is in days. The log-rank p value is indicated. **D**) Forrest plot showing Cox proportional hazard of harboring multiple focal CNVs or high-risk mutations.

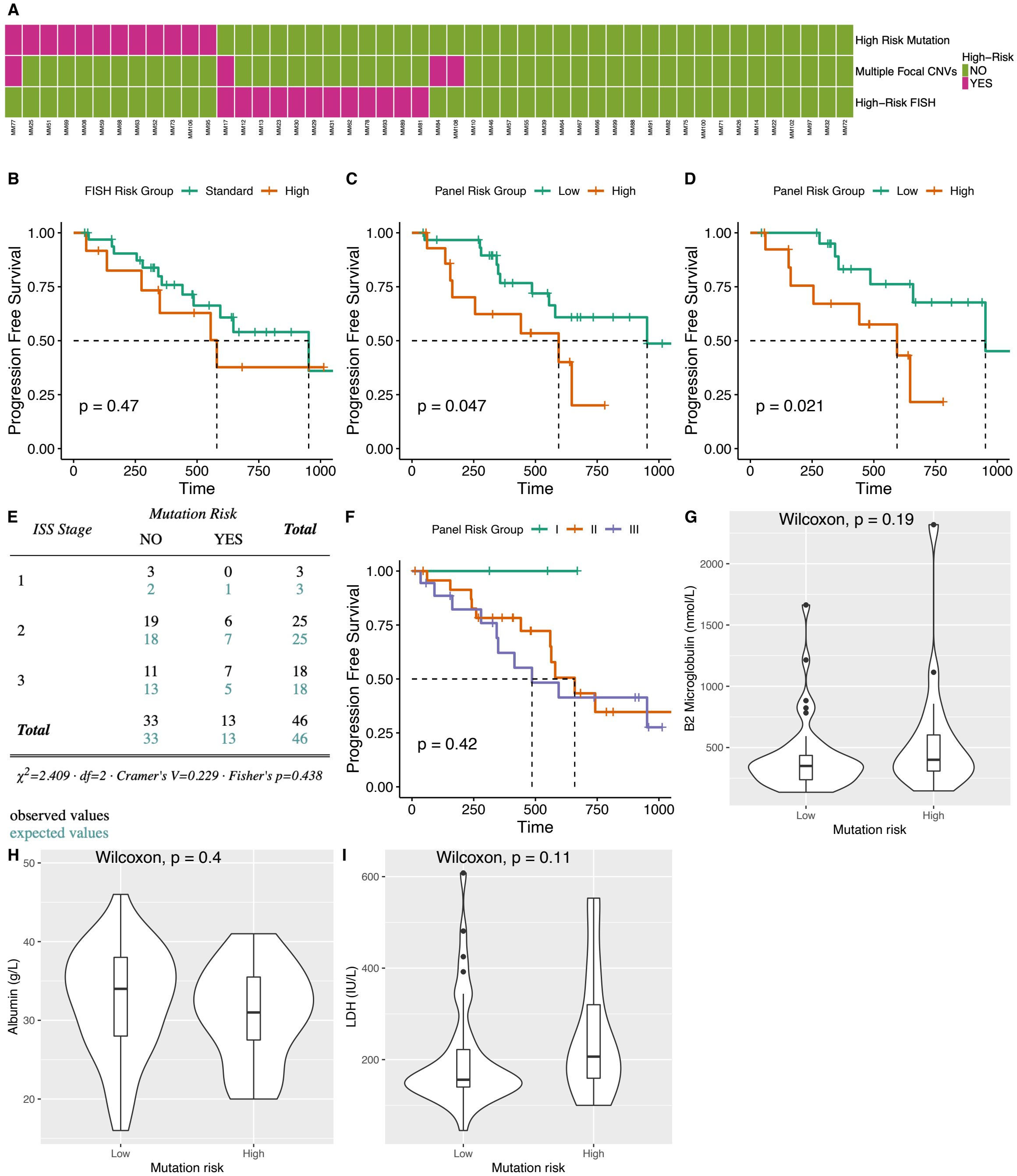


Figure 3.13: DMG26 risk markers are independent and significant prognostic markers

A) plot indicating risk marker status in the FISH assessed portion of our cohort. Columns are patients, rows are risk markers. Pink indicates the patient is positive for that risk marker, Green indicates the patient is negative for that risk marker. **B**) Kaplan-Meier plot of patients classified by R-ISS FISH markers to be high- (Orange, n = 12) and standard-risk (Green, n = 36). The log-rank p value is indicated. **C)** Kaplan-Meier plot of FISH assessed patients with high-risk panel markers (Orange, n = 15), and without (Green, n = 33). The log-rank p value is indicated. **D**) Kaplan-Meier plot showing patients defined by FISH to be standard-risk, with different clinical courses between those with high-risk panel markers

(Orange, n = 12) and those without (Green, n = 24). The log-rank p value is indicated. **E**) Chi-square table showing that the distribution of panel risk is independent of ISS staging. **F**) Kaplan-Meier plot of patients at diagnosis staged by the ISS: I (Green, n = 3), II (Orange, n = 25), III (Purple, n = 18). The log-rank p value is indicated. **G**) Violin plot of beta-2 microglobulin between patients with and without high-risk panel markers. **H**) Violin plot of albumin between patients with and without high-risk panel markers. **I**) Violin plot of LDH between patients with and without high-risk panel markers.

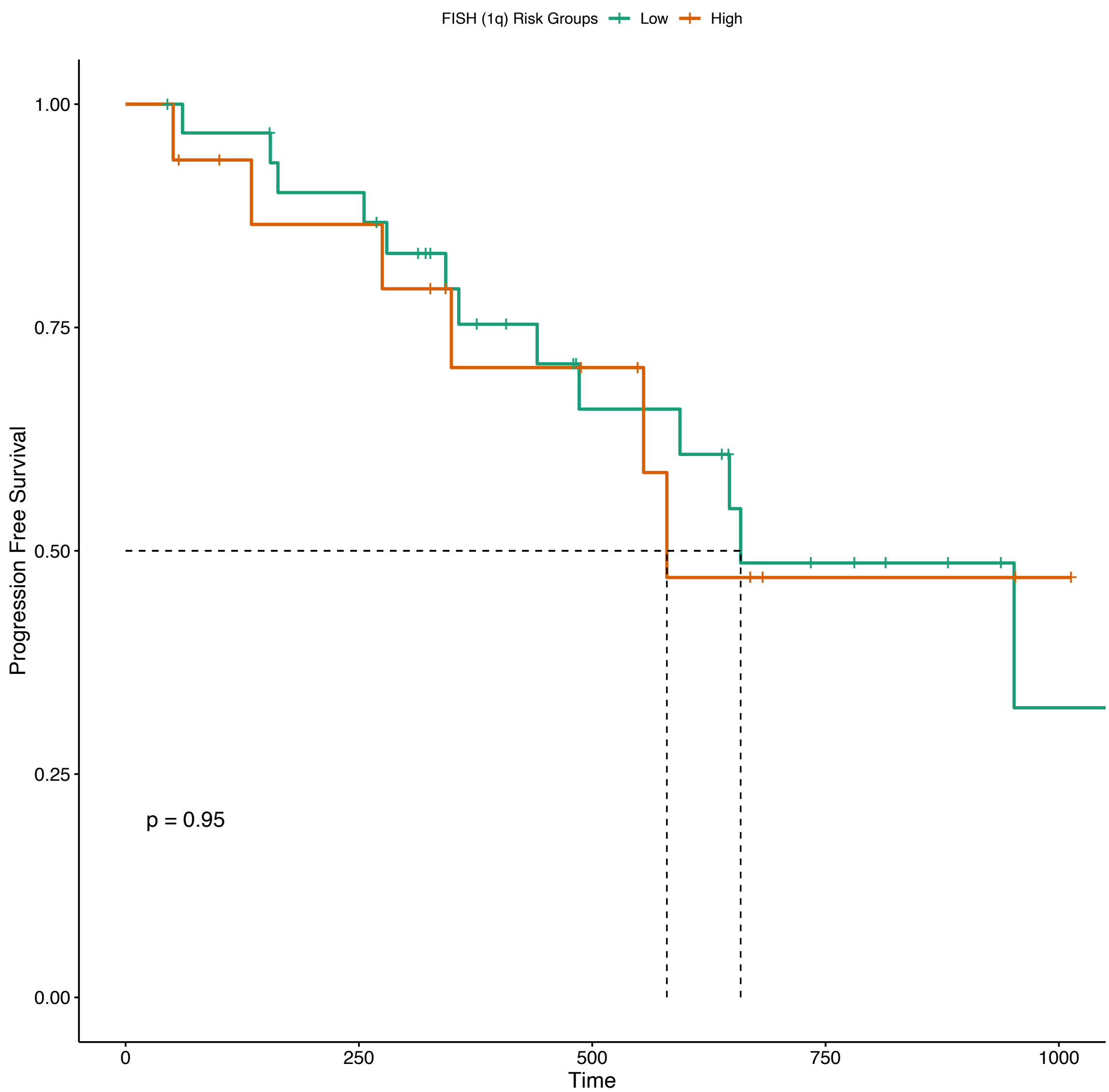


Figure 3.14: FISH 1q risk
Kaplan-Meier plot of the impact of 1q FISH on our patient cohort

3.5 Discussion:

With an increasing number of reports identifying risk contributions by genomic lesions beyond the interrogation scope of FISH, NGS informed prognostication is an attractive option to address current shortcomings. We demonstrate here a clinically relevant targeted sequencing approach using our DMG26 mutation panel. Applying our panel to clinical samples, and cell lines, we identified variants at a frequency and distribution consistent with other panel assessment, though an increase in variant frequency of some genes, namely *ATM*, *FGFR3* and *ACTG1* was observed.[12,20,21,356,360] Compared to the MMRF cohort which reported 1.1 mutations per patients across the panel assessed genes, we identified more lesions per patient at 3.4. In both cases, our higher mutation rate is likely attributable to the higher sequencing depths achieved by our panel, which facilitated more sensitive identification of lower VAF mutations contributing to the bulk of variability in mutational burden between patients. Our panel outperformed both ISS prognostication, and R-ISS FISH based risk assignment, was independent of other prognostic markers, and was mutually exclusive from R-ISS high risk genomic markers in our cohort.

Prognostication is a critical step in the clinical workup of MM patients, and can aid in therapeutic decisions. The current R-ISS for MM is an inadequate means of risk stratification as subsets of individuals stratified as 'low-risk' are later shown to have remarkably progressive disease, while others labelled as 'high-risk' may remain quite stable through the course of their disease.[72,361] The R-ISS is heavily reliant on FISH for the determination of the high-risk genomic abnormalities defined by the International Myeloma Working Group (IMWG), namely del(17p), t(4;14), and t(14;16). There are two major caveats with this approach; firstly, FISH has limited resolution and scope which impedes inclusion of small variants as risk parameters. Secondly, the independent prognostic value of each of these genomic abnormalities is debatable, and in the case of del(17p) it is based on the fraction of clonal involvement.[142,362–366] Therefore, the need to enhance genomic interrogation to redefine risk and evaluate prognosis in MM patients is increasingly recognized within the scientific community, particularly in the light of emerging new therapies. In recent years, there have been a few reports on the value of next-generation sequencing (NGS) technologies in MM to further probe the underlying genetic and transcriptomic landscape. These have spawned clinically focused assessments that employ gene expression profiles and targeted sequencing panels to identify numerous disease groups, each with specific clinical courses.[20,97,98,166,167,356] Yet clinical adoption of these NGS assessments has not occurred in practice, and FISH remains the gold-standard.

It is noteworthy that patient risk stratification is an evolving process which must reflect contemporary treatment modalities. One major benefit to NGS-based genomic interrogations compared to FISH is increased resolution, which allows appropriate use of targeted therapies that are prescribed based on the presence of SNVs. Numerous such drugs are under various stages of development for mutations in *NRAS*, *KRAS*, *BRAF*, *ATM*, and *FGFR3*.[281,282] Mutations within these genes accounted for 97 of the 294 mutations identified within our study, and 5 of these were high-risk (*ATM* c.8530delA, *ATM* c.8338delC, *ATM c*.3349C>T, *ATM* c.4307delA, *BRAF* c.2156delG). Hence, information captured by our panel is relevant to the precision medicine paradigm for targeted therapy.

Beyond targetable variants, our panel also captured mutations with significant associations with PFS. Assessing our panel data in a univariate Cox hazard model, we found mutations in *CDKN1B* and *PRMD1* at a VAF of ≥ 20% to significantly reduce PFS. Our panel targets genes common to the M3P (15 shared genes), and M3Pv2.0 (24 shared genes) panels.[12,21] A recent investigation using the M(3)Pv2.0 panel identified *PRDM1* variants to significantly correlate with both overall and progression-free survival, conferring a hazard similar to that identified by our analysis; though, *CDKN1B* was not reported to confer risk.[21] Both the M(3)Pv2.0 and our assessment identified few patients with mutations in *CDKN1B*; thus, both may be underpowered to probe impacts of this gene on clinical outlook. Additionally, our study contained a mixed cohort of plasma cell dyscrasias at diagnosis and relapse, while the M(3)Pv2.0 assessment was exclusively on newly diagnosed MM patients.[21] Nonetheless, *CDKN1B* has been identified as a driver gene, and the abundance of driver mutational events is associated with poor OS and PFS.[94]

Assessing mutations for their impact on protein sequence and pathogenicity score from FATHMM-XF revealed further panel-captured prognostic information. Both high-severity mutations and FATHMM-XF predicted pathogenic mutations defined patients with significantly reduced PFS, hence these were collectively termed ‘high-risk’ mutations. Genes impacted by high-risk mutations within our cohort are *ATM*, *BRAF*, *CCND1*, *CYLD*, *DIS3*, *FAM46C*, *LTB*, *MAGED1*, *RB1*, *SP140*, *TP53* and *STAT3*, many of which have been described as driver genes previously.[9,14,90,94] Of the 16 patients in our cohort harboring high-risk mutations, three (MM63, MM43, MM106) had two high-risk mutations. Notably, these three patients had remarkably reduced PFS and two, MM43 and MM106, progressed to PCL. The other, MM63, had a brief follow-up period of only 51 days during which no event occurred. MM43 had two high-severity mutations in *FAM46C* (c.138_139dupAA, c.678_679delGCinsTT), while MM106 had a high-severity mutation in *RB1* (c.772_776delAACAG) and a FATHMM-XF predicted pathogenic mutation in *TP53* (c.404G>T). Both high-risk mutations in MM106 are reported in COSMIC (COSM2744945, COSM923). Neither of the *FAM46C* mutations in MM43 have been reported previously; however, both severely alter the protein composition which is consistent with *FAM46C* acting as a tumour suppressor within MM.[226,227] Due to our panel design we cannot determine if these mutations are in *cis* or *trans*, and thus if this patient has a wild-type allele of *FAM46C*.

We did not capture a high-risk variant in any of our MGUS or SMM patients. This is consistent with other reports describing the genomic landscape within these preclinical stages as less heterogenous and characteristically lower risk.[11,367,368] Identification of patients imminently transitioning to MM, especially from SMM, is a key area of research as early identification and subsequent early intervention may be clinically advantageous.[369] Our cohort of SMM and MGUS patients was relatively small and no individual progressed to overt MM during the follow-up period. Hence, further investigations are necessary to probe the prognostic relevance of our panel in preclinical plasma cell dyscrasias.

No high-risk mutations were identified within *NRAS* or *KRAS*, the most commonly mutated genes in MM which are collectively present in about ~40% of cases.[8,23,90,369] The bulk of mutations within *NRAS* and *KRAS* are activating at amino acids 12, 13, 60, and 61, and drive the MAPK pathway.[173,355] The prognostic relevance of mutations within these genes has been previously investigated and neither *NRAS* nor *KRAS* is strongly associated with poor prognosis.[13,370] In fact, a recent report suggest RAS mutations to be a prognostically favourable indication in some treatment groups.[371] Our panel identified 25 patients with a total of 27 mutations in *KRAS* or *NRAS*, 23 of which were at the 12, 13, 60, or 61 amino acid hotspots. Two of the four variants that do not involve these hotspots have not been reported previously; one of which was synonymous. The other variant was *KRAS* c.240delT in MM52. Interestingly, this individual had the highest mutation burden in our patient cohort, but did not experience an event within the 728 days of subsequent follow-up.

Assessing our panel data by CovCopCan identified numerous focal-CNVs. None of the deletions detected by this approach overlapped with mutations within the same patient, indicating that we did not capture double-hit or loss of heterozygosity (LOH) events. However, similar approaches in larger cohorts have successfully identified double-hits to be of particular prognostic importance, especially within the context of *TP53*.[95,372] Additionally, though our panel was not designed for CNV calling and hence did not probe genomic regions that have known clinically significant CNVs we found that harbouring two or more focal CNVs has a significant negative correlation with PFS. Inclusion of focal CNVs with high-risk mutations further enhanced patient risk stratification; though, we could not assess these in a multivariate model as our study cohort was insufficiently powered for multivariate assessments.

Notably, applying our combined risk scheme (high-risk variant, focal CNVs) to patients identified by FISH to be standard-risk, successfully reclassified these patients into high- and low-risk groups. Furthermore, in our cohort, our panel outperformed both R-ISS FISH and ISS based risk stratification. Though we could not assess panel-risk in a multivariate risk model against other risk schemes due to cohort size limitations, we assessed the independence of panel risk markers to standard clinical and prognostic factors. We found no significant association between high-risk variants and ISS stage or biochemical biomarkers.

Compared to other panels proposed for MM, our panel and analysis approach provide strong prognostic information that robustly risk categorizes patients into groups with significantly different outlooks for progression-free survival. Furthermore, the limited panel size makes this approach feasible for clinical laboratories with even modest sequencing capacity and is far more cost-effective than FISH.

Study limitations include a relatively small cohort size, short median follow-up period of 19 months, and FISH data that did not include the whole patient cohort. This underpowered our study for multivariate hazard assessment. Additionally, our cohort size limited comparison of our panel between differently treated groups. Notwithstanding this, the lack of overlap between high-risk FISH and high-risk mutations was a notable finding in our study that deserves further exploration to assess our risk scheme within the context of R-ISS based high-risk FISH. Furthermore, assessment of larger pre-clinical cohorts may provide insight into whether our panel can determine risk of progression to overt MM within MGUS and SMM groups.

3.6 Author Contributions:

Conception and design: M.O.E., D.G., S.D.C., C.JV.C, P.K., M.A.H.
Development of methodology: D.G., S.D.C., M.O.E., P.K.
Acquisition of data: S.D.C.,M.O.E., D.G., A.T., P.K., J.W., M.G., B.E., J.E.B., S.G.
Analysis and interpretation of data: S.D.C., D.G., M.O.E., A.T., M.A.H., N.F.
Writing, review, and/or revision of the manuscript: S.D.C., M.O.E., D.G., C.JV.C, A.T., M.A.H., J.W., J.E.B., B.E.K., S.G., D.W.
Administrative, technical, or material support (i.e. reporting, or organizing data): M.O.E., D.G., S.D.C.
Study supervision: M.O.E., D.G.
Other (funding): M.O.E.

3.7 Acknowledgements:

We thank the molecular diagnostics laboratory lead by Dr. Wenda Greer, with special thanks to Makoto Matsuoka, Judy Park and Kendra MacDonald for excellent technical assistance. This study was supported by a Nova Scotia Health Authority Research grant (#1021520, M.O.E., C.JV.C.), the Beatrice Hunter Cancer Research Institute through the Cancer Research Training Program (S.D.C., B.E.K.), and we thank Illumina (BC, Canada) for providing discounted reagents for library preparation and flow cell for sequencing.

Chapter 4 Clinically viable WGS to outperform FISH in SV detection

4.1 Abstract

Multiple myeloma, the second most common hematological malignancy, is severe and incurable. It presents with a broad range of genomic abnormalities, with translocations at the IgH locus of chromosome 14 and copy number variations (CNVs) of whole-chromosomes to focal regions being recurrent. These events carry significant prognostic information and are hence assessed for in clinical settings using fluorescent *in situ* hybridization (FISH). The use of this technology is suboptimal as it is limited in scope by its targeted nature, susceptibility to miss complex or unbalanced structural variants, high cost, low throughput, and long turn-around-time. We have previously demonstrated that ultra-low-depth whole genome sequencing (WGS) efficiently captures copy number variants in MM genomes, but that it was insufficiently deep to capture other structural variations. Here, we employ high-coverage WGS, and demonstrate that coverages down to 10X facilitate highly concordant translocation calling compared to FISH.

4.2 Introduction

Multiple myeloma (MM) is a malignancy of post germinal centre plasma cells which invade the bone marrow.[1] Currently it is the second most common hematological malignancy in North America accounting for 1-2% of all cancer diagnoses and 10% of hematological cancer diagnoses. Moreover, the incidence of MM doubled between 1990 and 2016.[4,5] Despite dramatic improvements in therapeutic options that have led to increased survival times over the previous decade, the disease is incurable, five-year survival remains low, and therapeutic response is highly varied.[3] The most important determinants of patient outcome and response to therapy are the genetic lesions which underpin this malignancy.[19] Accordingly, genomic assessment is a mainstay of MM patient risk-stratification and is currently performed via fluorescent *in situ* hybridization (FISH).[24] However, FISH has a number of technical short comings, including its targeted nature, low resolution, limited throughput, slow turnaround time, and high cost, and is therefore a sup-optimal technology for this purpose. Moreover, the heterogeneity of the myeloma genome is increasingly appreciated and it may present with numerous subclones, making it a poor candidate for both targeted and low-throughput assessments. Additionally, myeloma cells often harbour small and sometimes cryptic lesions below the resolution limit of FISH; collectively, such considerations highlight the pressing need for better genomic interrogations for MM.[16,88,94]

The classic large scale chromosomal alterations that partition MM into hyperdiploid and non-hyperdiploid subsets are termed primary lesions.[6,8,24,373,374] In hyperdiploidy, aneuploidy of odd numbered chromosomes occurs, and this generally has a favourable prognosis.[375,376] In non-hyperdiplody, translocation at the IgH locus on chromosome 14 position the highly active IgH promoter next to oncogenes on chromosomes 4 (*FGFR3*/*MMSET*), 6 (*CCND3*) 11 (*CCND1*), 16 (*MAF*), and 20 (*MAFB*), thereby driving their expression. Smaller-scale copy number variants (CNVs), translocations at the MYC locus, single nucleotide variants (SNVs), and indels; all termed secondary lesions; are common and occur within a large range of caner clonal fractions.[6,8,9,13,90] Secondary lesions drive the remarkable heterogeneity observed within and between patients, significantly modulate patients' risk, and have important therapeutic implications.[16,21,94,95]

Currently, MM is prognosticated in accordance with the Revised-International Staging System (R-ISS), within which the FISH-detected genomic events t(4;14), t(14;16), and del(17p) are considered to be high-risk markers.[24] Accordingly, the minimum recommended FISH panel probes only for these lesions, which in addition to other technical limitations makes reliance on FISH suboptimal.[27] This limits the application of precision medicine, as unexpected lesions which may guide prognostic and therapeutic decisions may not be captured.

Next-generation sequencing (NGS) techniques are a compelling alternative to FISH, as these technologies can achieve much higher resolution, can be cost-effective, and can be designed to assess the genome either comprehensively or in a targeted manner. A few MM-specific NGS panels have been previously reported.[20,21] These, and other assessments have confirmed that a few large scale alterations, t(4;14), t(14;16), del(17p), del(1p), and gain(1q), are the dominant genetic risk factors.[19,24] Currently, only one MM-specific panel, the myTYPE, reports on such lesions.[20] However, the known translocation breakpoints at the *IgH* locus can vary by megabases, hence a targeted approach may not capture all such translocations. Translocation identification may be better served by whole-genome sequencing (WGS) based approaches due to its untargeted nature.

Whole-genome sequencing of tumours, while decreasing in cost, can still be prohibitively expensive for routine clinical use.[377] The cost of WGS scales primarily with sequencing depth, hence, clinically viable WGS-based assessments require optimization of the sequencing depth for utility versus cost. We have previously demonstrated that ultra-low depth WGS, at coverages less than 0.1x, is superior to FISH for profiling CNVs in MM samples.[26] This approach was, however, unable to reliably detect translocations.[17] Capturing translocations is essential for risk stratification and therapy choice.[19,24] Herein, we investigate WGS at coverages of 1X to 12x as an option for CNV and translocation detection in MM patients.

4.3 METHODS

4.3.1 Patient Sample Acquisition

Patient bone marrow samples were taken from the Nova Scotia Health's MM tumour bank. The samples were processed as described previously to obtain CD138+ selected cells.[26] Nine MM patient sample were taken, one patient sample, MM29, had FISH performed at diagnosis while WGS was performed on bone marrow taken at time of relapse following complete remission, hence was not considered in head-to-head-comparisons between WGS and FISH. MM29 was included in comparison of subsampled WGS translocations call against the full depth call set. Patients were selected to maximize the number and variety of translocations as detected previously by FISH. Patient FISH data was collected from the NS Health laboratory information system. One MM cell line (MM1S) was included as an external control. This work was conducted under ethical approval by the Nova Scotia Health Research Ethics Board (#1021520 and #1021397), and patients provided written informed consent.

4.3.2 DNA Extraction and Sequencing

DNA and RNA were extracted as described previously using a Qiagen AllPrep kit.[26] DNA was sent to Genome Quebec (Canada) for 250 bp insert, 2x150 WGS sequencing at 12X coverage on one lane of a V4 flow cell on a an Illumina NovaSeq.

4.3.3 Bioinformatics

Reads in FastQ format were aligned to the hg19 human reference genome using bwa-mem[340], and processed based on GATK4 best practices[378]. Aligned BAMs were than processed for CNV calling with QDNAseq at all standard window settings (1, 5, 10, 15, 30, 50, 100, 500, and 1000 kb), and break ends defining interchromosomal translocations were called using SVABA, LUMPY, MANTA, and GRIDSS on default settings.[308–310,334,335,379] Analysis with GRIDSS included the ENCODE blacklist. GRIDSS break end calls were compared against FISH and other caller at score cut-offs of 0, 100, and 1000.

Ensemble Variant translocation calling

Translocations, from the union of all callers, were filtered to select only those that are:

1) Interchromosomal
2) Classic MM translocation (at the IgH locus, or MYC locus) or meeting any of the following:
    a.) called by GRIDSS with a score greater than 1000
    b.) called by GRIDSS with a score greater than 500 and are called by at least two other variant callers
    c.) called by SVABA with a score greater than 15
    d.) called by SVABA with a score greater 9 and are called by at least two other variant caller.
3) Passed manual review in IGV and of individual break end calls in VCFs

All filter passing break ends were compared against MM1S super enhancers from the dbSUPER[380] database, and those within 50 kb of one were labeled as proximal to a super enhancer. Aligned BAMs were subsequently *in silico* subsampled using samtools with a random seed of 0 to simulate varying depths of coverage (12X-1X coverage at 1X increments), and translocation calling was repeated as above.[342] The performance of translocation calling was assessed at each subsampled depth through comparison to FISH and to the original WGS variant call set at 12X.

4.3.4 Statistical Analysis

Performance metrics, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated by comparing calls against FISH data, and subsampled WGS data to the 12X call set. For translocations, comparing to FISH, WGS calls were considered to be matching if within 5,000,000 bp of the FISH probed genes/targeted regions. For CNVs, comparing to FISH, WGS-based calls needed to overlap with the FISH probe region. Comparing subsampled WGS based translocation calls to the full-depth call set, breakpoints were required to be within 10,000 bp to be considered matching. ROC curves were generated iterating over sequencing depth. Youden's index, an indication of the optimum cut-off for a binary classifier, was calculated using the R package, pROC.[338]

4.4 RESULTS

4.4.1 Cohort Selection:

Ten samples (9 MM patients and 1 MM cell line) underwent WGS to ~12x (11.9-17X). The clinical features of the patient samples are summarized in Table 1. Briefly, within our patient cohort, FISH reported four *MYC* separations and eight *IgH* separations, two of which had an unidentified partner (***Table 4.1***). Additionally, 17 copy-number gains and 9 copy-number losses were also reported by FISH (***Table 4.2***, ***Figure 4.1***). MM1S has previously been described to have t(14;16) and t(3;8). The full depth of WGS for each patient is described in ***Table 4.3***.

4.4.2 Copy Number Variant Calling and Comparison to FISH:

WGS based CNV calling identified a total of 2394 CNVs, comprising 1471 deletions and 923 amplifications across the cohort (***Figure 4.1*** **A-J**, ***Figure 4.2***). The mean CNV size was 1.60 Mb, with the mean size of deletions and duplications being 2.11 and

1.23 Mb, respectively (**Figure *4.3* A**). MM29 had the fewest number of bases over which CNVs were called, with a total of 11.7 Mb, of which, 6.1 Mb were amplified and 5.6 Mb had reduced copy number (**Figure *4.3* B**). MM97 had the highest number of bases over which CNVs were called, with a total of 882.8 Mb being copy-number altered, of which, 460.5 Mb were amplified, while 699.7 Mb had reduced copy-number (**Figure *4.3* B**). MM68 had the fewest CNVs, with 90 being identified, while MM75 had the most CNVs identified at 513 (**Figure *4.3* C**). The q arm of chromosome 1 in sample MM12 was the most amplified in our cohort, having a max log_2 fold-change of 1.3 and an average across the arm of 1.12. A 200 kb section on chromosome 8p11.21 had the most negative log_2 fold-change of -4.76, and -4.52 in MM12, and MM08, respectively.

Table 4.1: FISH called translocations and corresponding WGS translocation calls

SAMPLE	FISH	WGS CALL	WGS TRANSLOCATION CALL CORRESPONDING TO FISH DATA											
			FULL DEPTH	12X	11X	10X	9X	8X	7X	6X	5X	4X	3X	2X
MM08	IgH sep (+)	t(14;8) (q32.33;q24.21)	TP	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN
MM08	t(4;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM08	t(6;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM08	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM08	t(14;16) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM08	t(14;20) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM08	MYC sep (+)	t(14;8) (q32.33;q24.21)	TP	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN
MM12	IgH sep (+)	t(14;4) (q32.33;p16.3)	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN
MM12	t(4;14) (+)	t(14;4) (q32.33;p16.3)	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN
MM12	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM12	MYC sep (-)	t(3;8) (q13.13;q24.21) t(15;8) (q13.3;q24.21)	FP	FP	FP	FP	FP	FP	FP	FP	TN	TN	TN	TN
MM29	IgH sep (+)	t(14;16) (q32.2;q23.3)	TP	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN
MM29	t(4;14) (+)	Not Called	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN
MM29	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM29	MYC sep (-)	t(3;8) (q13.13;q24.21)	FP	FP	FP	FP	FP	FP	FP	FN	FN	FN	FN	FN
MM30	IgH sep (+)	t(14;16) (q32.33;q23.1)	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	FN
MM30	t(14;16) (+)	t(14;16) (q32.33;q23.1)	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	FN
MM30	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM30	MYC sep (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM40	t(14;16) (+)	t(14;16) (q32.33;q23.1)	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN	FN
MM40	MYC sep (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM46	IgH sep (+)	t(11;14) (q13.3;q32.33)	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN
MM46	t(11;14) (+)	t(11;14) (q13.3;q32.33)	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN
MM46	MYC sep (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM68	IgH sep (+)	t(11;14) (q13.3;q32.33)	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP
MM68	t(11;14) (+)	t(11;14) (q13.3;q32.33)	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP
MM68	MYC sep (+)	t(3;8) (q13.13;q24.21)	TP	TP	TP	TP	TP	TP	TP	TP	TP	FN	FN	FN
MM75	IgH sep (-)	t(14;8) (q32.33;q24.21)	FP	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM75	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM75	MYC sep (+)	(3;8) (q13.13;q24.21) t(14;8) (q32.33;q24.21)	TP	TP	TP	TP	TP	TP	TP	TP	FN	FN	FN	FN
MM97	IgH sep (+)	t(14;20) (q32.33;q11.23)	TP	TP	TP	TP	TP	FN	FN	FN	FN	FN	FN	FN
MM97	t(4;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM97	t(6;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM97	t(11;14) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM97	t(14;16) (-)	Not Called	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
MM97	t(14;20) (-)	t(14;20) (q32.33;q11.23)	FP	FP	FP	FP	FP	TN	TN	TN	TN	TN	TN	TN
MM97	MYC sep (+)	t(3;8) (q13.13;q24.21)	TP	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN	FN

Table 4.2: FISH called CNVs and corresponding WGS CNV calls

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM08	+7CEN(D7Z1x3)	chr7	0	159138663	+	50000:159100000	dup	0.55	0.54	0.54	0.54	0.54	0.54	0.54	0.55	0.54	TP	TP
MM08	+9CEN(D9Z1x3)	chr9	0	141213431	+	200000:141000000	dup	-0.04	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	TP	TP
MM08	+15CEN(D15Z4x3)	chr15	0	102531392	+	22550000:102400000	dup	0.57	0.56	0.56	0.56	0.55	0.55	0.55	0.55	0.54	TP	TP
MM08	+20q12(MAFBx3)	chr20	37600000	41700000	+	29850000:62900000	dup	0.49	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	TP	TP
MM08	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM08	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN
MM08	+1q21(TP73x2,CKS1Bx3)	chr1	154935118	154963725	N		N										TN	TN
MM08	+1q21(TP73x2,CKS1Bx3)	chr1	3557129	3664765	N		N										TN	TN
MM08	+1(TP73,CKS1B)x3	chr1	154935118	154963725	N		N										TN	TN
MM08	+1(TP73,CKS1B)x3	chr1	3557129	3664765	N		N										TN	TN
MM08	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM08	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM08	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM08	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM08	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N		N										TN	TN
MM08	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N		N										TN	TN
MM08	-13(RB1,LAMP1)x1	chr13	48865883	49068026	N		N										TN	TN
MM08	-13(RB1,LAMP1)x1	chr13	113939469	113989741	N		N										TN	TN

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM12	+1q22(TP73x2,1q22x5)	chr1	154935118	154963725	+	149850000:161400000	dup	1.10	1.28	1.28	1.28	1.28	1.28	1.27	1.27	1.28	TP	TP
MM12	+1q22(TP73x2,1q22x5)	chr1	3557129	3664765	+		N											
MM12	+1(TP73,1q22)x3	chr1	154935118	154963725	N	149850000:161400000	dup	1.10	1.28	1.28	1.28	1.28	1.28	1.27	1.27	1.28		
MM12	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-1.20	-0.97	-0.96	-0.96	-0.96	-0.95	-0.96	-0.96	-0.95	TP	TP
MM12	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.98	-0.98	-0.98	-0.98	-0.99	-0.98	-0.97	NA	NA		
MM12	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-1.20	-0.97	-0.96	-0.96	-0.96	-0.95	-0.96	-0.96	-0.95		
MM12	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.98	-0.98	-0.98	-0.98	-0.99	-0.98	-0.97				
MM12	+15CEN(D15Z4x3)	chr15	0	102531392	+	22550000:102400000	dup	0.54	0.55	0.55	0.55	0.54	0.54	0.54	0.54	0.54	TP	TP
MM12	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM12	+9CEN(D9Z1x3)	chr9	0	141213431	N		N										TN	TN
MM12	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM12	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN
MM12	+1(TP73,1q22)x3	chr1	3557129	3664765	N		N										TN	TN
MM12	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM12	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM12	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM12	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM29	+1q21(TP73x2,CKS1Bx3)	chr1	154935118	154963725	+		N										FN	

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM29	+1q21(TP73x2,CKS1Bx3)	chr1	3557129	3664765	+		N											
MM29	+3CEN(D3Z1x3)	chr3	0	198022430	+		N										FN	
MM29	+9CEN(D9Z1x3)	chr9	0	141213431	+		N										FN	
MM29	+11(CCND1-XTx3)	chr11	51600000	55700000	+		N										FN	
MM29	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-		N										TN	
MM29	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-		N										TN	
MM29	+17(TP53,D17Z1)x3	chr17	7559720	7602868	+		N										FN	
MM29	+17(TP53,D17Z1)x3	chr17	22190000	25810000	+		N											
MM29	+15CEN(D15Z4x3)	chr15	0	102531392	N		N										TN	
MM29	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	
MM29	+1(TP73,CKS1B)x3	chr1	154935118	154963725	N		N										TN	
MM29	+1(TP73,CKS1B)x3	chr1	3557129	3664765	N		N										TN	
MM29	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	
MM29	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	
MM29	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	
MM29	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	
MM29	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N		N										TN	
MM29	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N		N										TN	
MM30	+1q22(TP73x2,1q22x3)	chr1	154935118	154963725	+	150600000:161400000	dup	0.45	0.38	0.37	0.36	0.36	0.36	0.36	0.36	0.37	TP	TP

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM30	+1q22(TP73x2,1q22x3)	chr1	3557129	3664765	+		N											
MM30	+1(TP73,1q22)x3	chr1	154935118	154963725	N	150600000:161400000	dup	0.45	0.38	0.37	0.36	0.36	0.36	0.36	0.36	0.37		
MM30	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-0.69	-0.64	-0.64	-0.64	-0.64	-0.65	-0.64	-0.65	-0.65	TP	TP
MM30	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.65	-0.61	-0.61	-0.62	-0.62	-0.62	-0.60				
MM30	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-0.69	-0.64	-0.64	-0.64	-0.64	-0.65	-0.64	-0.65	-0.65		
MM30	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.65	-0.61	-0.61	-0.62	-0.62	-0.62	-0.60				
MM30	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM30	+9CEN(D9Z1x3)	chr9	0	141213431	N		N										TN	TN
MM30	+15CEN(D15Z4x3)	chr15	0	102531392	N		N										TN	TN
MM30	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM30	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN
MM30	+1(TP73,1q22)x3	chr1	3557129	3664765	N		N										TN	TN
MM30	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM30	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM30	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM30	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM40	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-1.33	-0.49	-0.48	-0.48	-0.48	-0.48	-0.48	-0.49	-0.49	TP	TP
MM40	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.55	-0.49	-0.50	-0.50	-0.50	-0.51	-0.51				

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM40	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-1.33	-0.49	-0.48	-0.48	-0.48	-0.48	-0.48	-0.49	-0.49		
MM40	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.55	-0.49	-0.50	-0.50	-0.50	-0.51	-0.51				
MM40	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM40	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM40	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM40	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM46	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-1.14	-0.94	-0.94	-0.93	-0.94	-0.93	-0.94	-0.94	-0.94	TP	TP
MM46	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.90	-0.89	-0.90	-0.90	-0.90	-0.89	-0.87				
MM46	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-1.14	-0.94	-0.94	-0.93	-0.94	-0.93	-0.94	-0.94	-0.94		
MM46	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.90	-0.89	-0.90	-0.90	-0.90	-0.89	-0.87				
MM46	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM46	+9CEN(D9Z1x3)	chr9	0	141213431	N		N										TN	TN
MM46	+15CEN(D15Z4x3)	chr15	0	102531392	N		N										TN	TN
MM46	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM46	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN
MM46	+1q22(TP73x2,1q22x3)	chr1	154935118	154963725	N		N										TN	TN
MM46	+1q22(TP73x2,1q22x3)	chr1	3557129	3664765	N		N										TN	TN
MM46	+1(TP73,1q22)x3	chr1	154935118	154963725	N		N										TN	TN
MM46	+1(TP73,1q22)x3	chr1	3557129	3664765	N		N										TN	TN

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_ CHROM	FISH_PROBE_BINDING_ START	FISH_PROBE_BINDING_ END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM46	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM46	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM46	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM46	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM68	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-1.25	-0.98	-0.97	-0.97	-0.97	-0.98	-0.98	-0.98	-0.99	TP	TP
MM68	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96				
MM68	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-1.25	-0.98	-0.97	-0.97	-0.97	-0.98	-0.98	-0.98	-0.99		
MM68	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96	-0.96				
MM68	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM68	+9CEN(D9Z1x3)	chr9	0	141213431	N		N										TN	TN
MM68	+15CEN(D15Z4x3)	chr15	0	102531392	N		N										TN	TN
MM68	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM68	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN
MM68	+1q22(TP73x2,1q22x3)	chr1	154935118	154963725	N		N										TN	TN
MM68	+1q22(TP73x2,1q22x3)	chr1	3557129	3664765	N		N										TN	TN
MM68	+1(TP73,1q22)x3	chr1	154935118	154963725	N		N										TN	TN
MM68	+1(TP73,1q22)x3	chr1	3557129	3664765	N		N										TN	TN
MM68	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM68	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM68	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM68	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM75	-1p36.3/+1q22(TP73x1,1q22x3)	chr1	3557129	3664765	-	2700000:3850000	del	-0.67	-0.69	-0.69	-0.69	-0.70	-0.70	-0.71			TP	TP
MM75	-1p36.3/+1q22(TP73x1,1q22x3)	chr1	154935118	154963725	+	149850000:161400000	dup	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.47		
MM75	+1(TP73,1q22)x3	chr1	3557129	3664765	N	2700000:3850000	del	-0.67	-0.69	-0.69	-0.69	-0.70	-0.70	-0.71				
MM75	+1(TP73,1q22)x3	chr1	154935118	154963725	N	149850000:161400000	dup	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.47		
MM75	+9CEN(D9Z1x3)	chr9	0	141213431	+		N										FN	FN
MM75	-13(RB1,LAMP1)x1	chr13	48865883	49068026	-	47050000:52050000	del	-0.94	-0.73	-0.72	-0.72	-0.72	-0.73	-0.72	-0.73	-0.72	TP	TP
MM75	-13(RB1,LAMP1)x1	chr13	113939469	113989741	-	113000000:114350000	del	-0.72	-0.71	-0.72	-0.72	-0.73	-0.73	-0.73				
MM75	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	N	47050000:52050000	del	-0.94	-0.73	-0.72	-0.72	-0.72	-0.73	-0.72	-0.73	-0.72		
MM75	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	N	113000000:114350000	del	-0.72	-0.71	-0.72	-0.72	-0.73	-0.73	-0.73				
MM75	-14q32(IGHx1)	chr14	104000001	107349540	-	65650000:107300000	del	-0.69	-0.74	-0.73	-0.73	-0.72	-0.73	-0.71	-0.71		TP	TP
MM75	+11(CCND1-XTx3)	chr11	51600000	55700000	N		N										TN	TN
MM75	+15CEN(D15Z4x3)	chr15	0	102531392	N		N										TN	TN
MM75	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM75	+3CEN(D3Z1x3)	chr3	0	198022430	N		N										TN	TN

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM75	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN
MM75	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM75	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM75	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM97	+3CEN(D3Z1x3)	chr3	0	198022430	+	100000:197800000	dup	0.23	0.30	0.29	0.29	0.30	0.29	0.30	0.30	0.31	TP	TP
MM97	+9CEN(D9Z1x3)	chr9	0	141213431	+	200000:141000000	dup	-0.11	0.41	0.40	0.40	0.40	0.39	0.40	0.39	0.39	TP	TP
MM97	+11(CCND1-XTx4)	chr11	51600000	55700000	+	55050000:58800000	dup	0.76	0.76	0.77	0.77	0.77	0.78	0.78	0.77	NA	TP	TP
MM97	-13q14(RB1x1,LAMP1x2)	chr13	48865883	49068026	-	47050000:52050000	del	-0.70	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.63	TP	TP
MM97	-13q14(RB1x1,LAMP1x2)	chr13	113939469	113989741	-		N											
MM97	-13(RB1,LAMP1)x1	chr13	48865883	49068026	N	47050000:52050000	del	-0.70	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.62	-0.63		
MM97	-13(RB1,LAMP1)x1	chr13	113939469	113989741	N		N											
MM97	+15CEN(D15Z4x3)	chr15	0	102531392	+	22550000:102400000	dup	-0.56	0.68	0.81	0.81	0.81	0.80	0.80	0.80	0.81	TP	TP
MM97	+7CEN(D7Z1x3)	chr7	0	159138663	N		N										TN	TN
MM97	+1q22(TP73x2,1q22x3)	chr1	154935118	154963725	N		N										TN	TN
MM97	+1q22(TP73x2,1q22x3)	chr1	3557129	3664765	N		N										TN	TN
MM97	+1(TP73,1q22)x3	chr1	154935118	154963725	N		N										TN	TN
MM97	+1(TP73,1q22)x3	chr1	3557129	3664765	N		N										TN	TN
MM97	-17p13.1(TP53x1,D17Z1x2)	chr17	7559720	7602868	N		N										TN	TN

SAMPLE	FISH_PROBE_NAME	FISH_PROBE_BINDING_CHROM	FISH_PROBE_BINDING_START	FISH_PROBE_BINDING_END	FISH_PROBE_STATUS	WGS_COORDINATES	CALL_AT_50_KB	LOG_AT_1	LOG_AT_5	LOG_AT_10	LOG_AT_15	LOG_AT_30	LOG_AT_50	LOG_AT_100	LOG_AT_500	LOG_AT_1000	CALL WITH MM29	CALL WITHOUT MM29
MM97	-17p13.1(TP53x1,D17Z1x2)	chr17	22190000	25810000	N		N										TN	TN
MM97	-17(TP53,D17Z1)x1	chr17	7559720	7602868	N		N										TN	TN
MM97	-17(TP53,D17Z1)x1	chr17	22190000	25810000	N		N										TN	TN
MM97	-13(RB1,LAMP1)x1	chr13	113939469	113989741	N		N										TN	TN


A
B
C
D
TP73
NRAS
FAM46C
CKS1B
SP140
FGFR3
EGR1
IRF4
LTB
TNFRSF21
PRDM1
EGFR
BRAF
CDKN2A
TRAF2
CCND1
ATM
CDKN1B
KRAS
RB1
DIS3
LAMP1
MAX
TRAF3
CYLD
TP53
STAT3
ACTG1
MAFB
MAGED1
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
WGS_GAIN
WGS_DEL
FISH_DEL
FISH_GAIN
FISH_NORMAL


E
F
G
H
TP73
NRAS
FAM46C
CKS1B
SP140
FGFR3
EGR1
IRF4
LTB
TNFRSF21
PRDM1
EGFR
BRAF
CDKN2A
TRAF2
CCND1
ATM
CDKN1B
KRAS
RB1
DIS3
LAMP1
MAX
TRAF3
CYLD
TP53
STAT3
ACTG1
MAFB
MAGED1
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
WGS_GAIN
WGS_DEL
FISH_DEL
FISH_GAIN
FISH_NORMAL

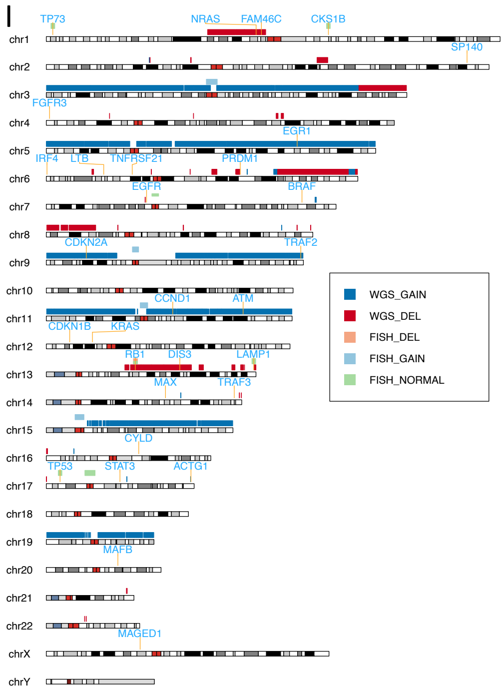


Figure 4.1: WGS captures more CNVs than FISH
Ideograms for each patient sample overlaying FISH and WGS CNV calls. A) MM08. B) MM12. C) MM29. D) MM30. E) MM40. F) MM46. G) MM68. H) MM75. I) MM97. Dark Blue indicates WGS amplifications. Dark Red indicates WGS called deletions. Light Blue indicates FISH called amplifications. Light Red indicates FISH called deletions. Light Green indicates FISH called normal copy number.

Table 4.3: WGS full depth coverage

Sample	Coverage
MM08	12.73
MM12	17.17
MM29	17.44
MM30	14.75
MM40	13.57
MM46	11.94
MM68	14.87
MM75	15.04
MM97	12.04
MM1S	14.87

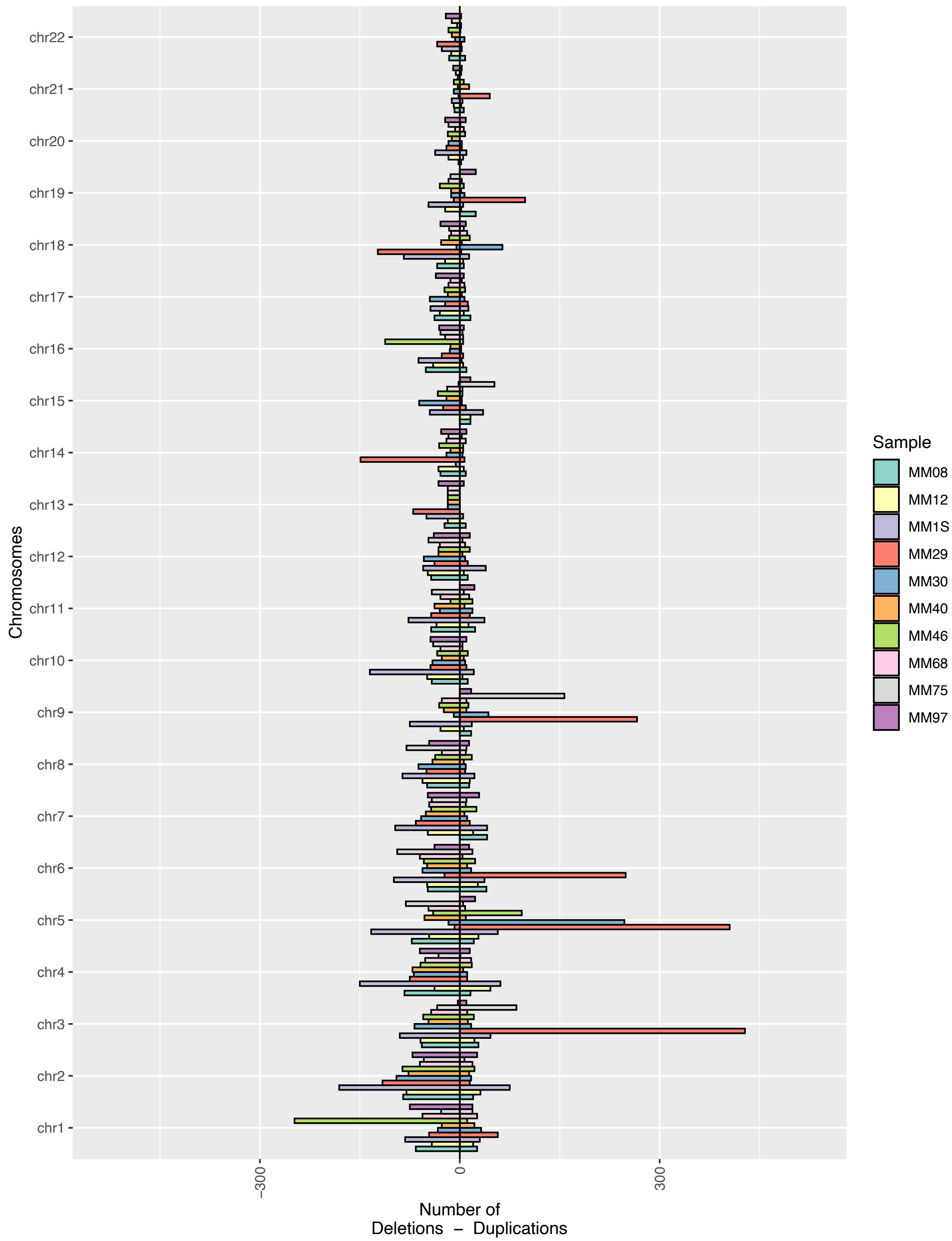


Figure 4.2: Number of CNVs per chromosome per patient
Hurricane plot indicating the number of deletions per chromosome per patients.

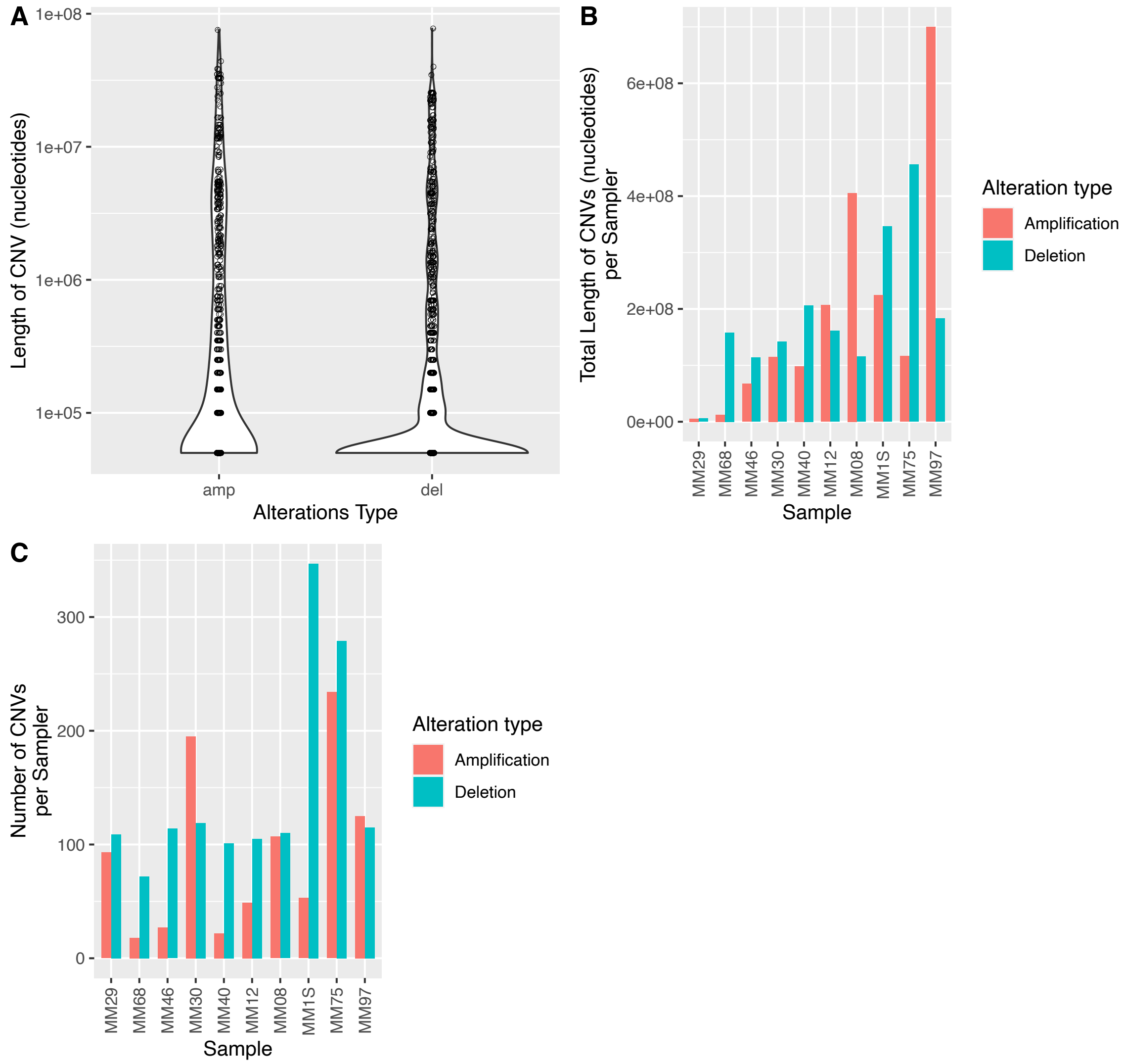


Figure 4.3: WGS captures a wide range of CNV sizes and abundances across samples

A) Violin plot indicating the distribution of CNV sizes (both deletions and amplifications) across all samples. B) Bar plot indicating the total size of deletions (Blue) and amplifications (Red) per samples. C) Bar plot indicating the total number of deletions (Blue) and amplifications (Red) per sample.

4.4.2.1 Comparing to FISH

Of the 22 FISH-identified CNVs, 21 had matching calls in the WGS data (Table 3 and ***Figure 4.1*****A-J**, ***Figure 4.4***). The one false negative was in MM75; a chromosome 9 centromeric FISH probe (D9Z1) had labelled this as trisomy of chromosome 9 (***Table 4.2***, ***Figure 4.1*** **H**, ***Figure 4.4***), while only a small amplification was called by QDNAseq on the q arm of chromosome 9 (***Table 4.2***, ***Figure 4.1*** **H**). This trisomy 9 was observed by FISH in 11 of 50 assessed cells in MM75. No false positive and 81 true negative calls were made by WGS (***Table 4.2***,***Table 4.4***). Hence, comparing to FISH, WGS-based CNV detection had a sensitivity, specificity, positive predictive value, and negative predictive value of 95.2% 1, 1, and 98.8%, respectively (***Table 4.4***).

Four of the eight compared samples (MM08, MM12, MM97, and MM75) were found to have extensive deletions on the p arm of chromosome 1 by WGS (***Figure 4.1*** **A,B,H,I**). Of these, only the deletions present in MM75 overlapped with the *TP73* FISH probe used to assess for such alterations; hence, del(1p) had been reported by FISH only in MM75 (***Table 4.2***; ***Figure 4.1*** **A,B,H,I**; ***Figure 4.4***). Similarly, MM68 had a modestly sized gain on the q arm of chromosome 1, which did not overlap with the *CKS1B* FISH probe used to assess for gain(1q), hence was unreported by FISH (***Figure 4.1*** **G;** ***Figure 4.4***).

Numerous lesions beyond FISH targeted regions were called by QDNASeq. In MM46, WGS identified extensive amplification across the q arm of chromosome 11 which included *CCND1* (***Figure 4.1*** **F**). In MM30 and MM40, extensive deletions across the q arm of chromosome 16, which included *CYLD* in both, were observed by WGS but not probed for by FISH (***Figure 4.1*** **D,E**; ***Figure 4.4***). In MM97, trisomy 5 was observed by WGS, and again unprobed for by FISH (***Figure 4.1*** **I**; ***Figure 4.4***).

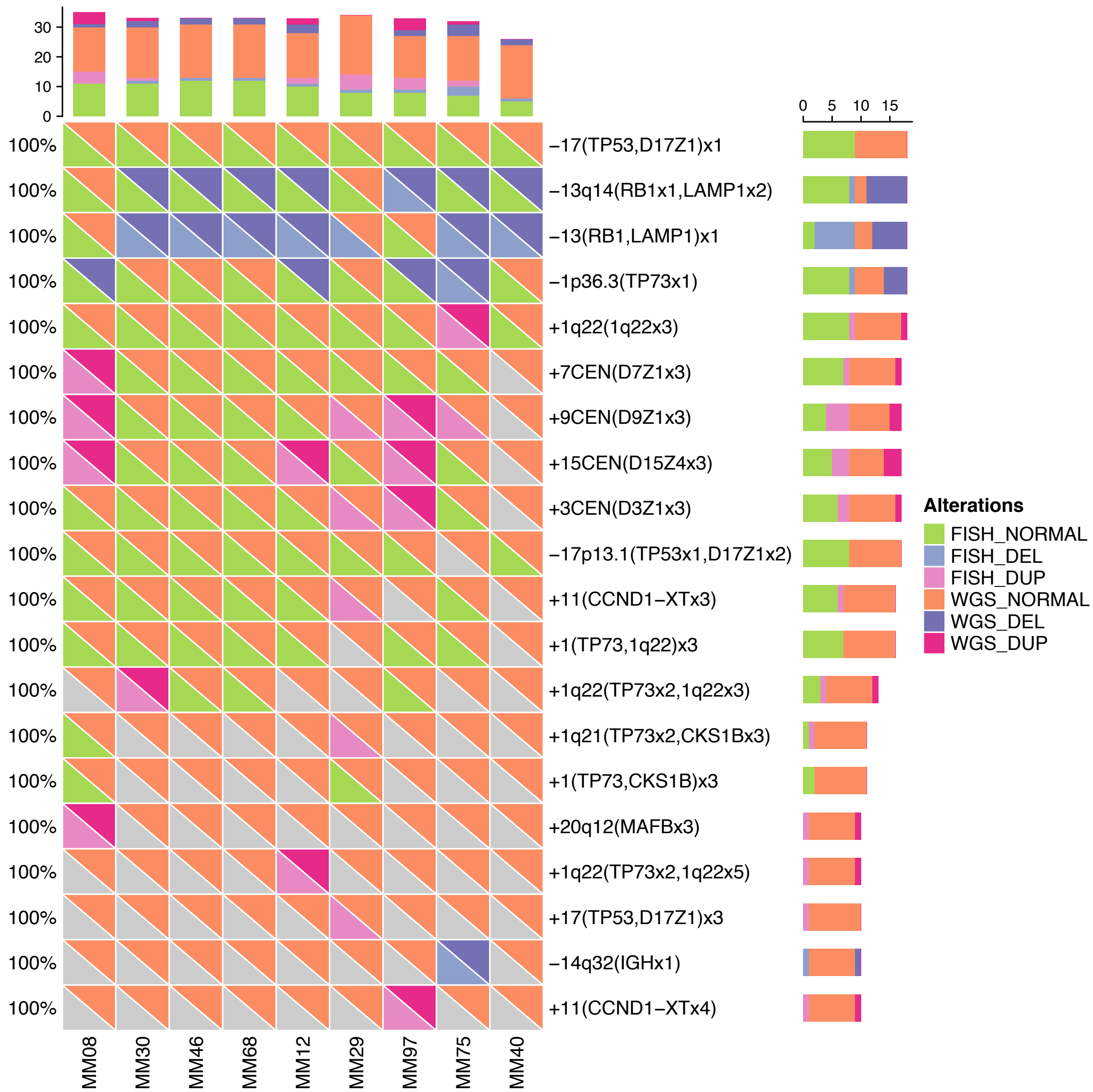


Figure 4.4: WGS CNV calls are consistent with FISH CNV
Oncoprint comparing WGS calls to FISH calls. Green indicates normal copy number call by FISH. Light Purple indicates deletion call by FISH. Light Pink indicates amplification call by FISH. Orange indicates normal copy number call by FISH. Dark Purple indicates deletion call by WGS. Dark Pink indicates amplification call by WGS. Rows are FISH probes, columns are patient samples.

Table 4.4: Performance of WGS CNV calling compared to FISH

	WGS compared to FISH
TP	20
TN	81
FP	0
FN	1
Sensitivity	0.95238095
Specificity	1
PPV	1
NPV	0.98780488

4.4.4 Comparison of Structural Variant Calling Algorithms

Pre-filtering, GRIDSS called the most interchromosomal break ends, at 92497, and identified all IgH translocations reported by FISH. (***Figure 4.5* ATable *4.5***) However, at the two recommended score cut-offs (500 and 1000), GRIDSS called 328, and 20 interchromosomal translocations respectively, and even while using the lower cut-off of 500, it missed the FISH reported IgH translocations in MM1S, MM12, MM40, MM46, MM60, and MM97. (***Figure 4.5* B,C**) Hence, to be sensitive we could not employ stringent score cut-offs, which made the test non-specific. To balance this, we incorporated a number of additional callers in an ensemble approach to corroborate low scoring GRIDSS variants (see ensemble variant calling in methods).

In the ensemble, GRIDSS called the most filter passing breakends, at 245, across all subsampled depths, while LUMPY called the least, at 78 (***Appendix Table* 3**). All filter passing variants called by other algorithms were also called by GRIDSS (***Appendix Table* 3**). Considering only canonical myeloma translocations, GRIDSS and LUMPY were again the most and least frequent breakend callers respectively, which was consistent at all subsampled depths (**Table *4.5***). Cohen's kappa, an assessment for similarity between binary callers, was slightly concordant (0-0.2) for all callers, except MANTA and SVABA, which had a Cohen's kappa of 0.25, placing them as fairly concordant (***Figure 4.6***).[381] Hence, the ensemble combination is well suited to capture and corroborate variants that would be difficult to identify in a specific manner using any one caller.

4.4.5 Structural Variant Calling at 12X and Comparison to FISH

After filtering, WGS-based structural variant calling identified 60 interchromosomal translocations at 12X across our cohort, with a median incidence per sample of 4 (range 2-18) (***Appendix Table* 4**, ***Figure 4.7*A-J**). MM12 had the most translocations, at 18, while MM08 and MM97 had the least, with two interchromosomal translocations each (***Appendix Table* 4**, ***Figure 4.7*A-J**). Chromosome 8 was involved in the most translocations across the cohort, with 19 identified in 9 samples. Chromosome 3 was involved in the second most translocations with 13 in 9 samples (***Appendix Table* 4**, ***Figure 4.8* A,B**).

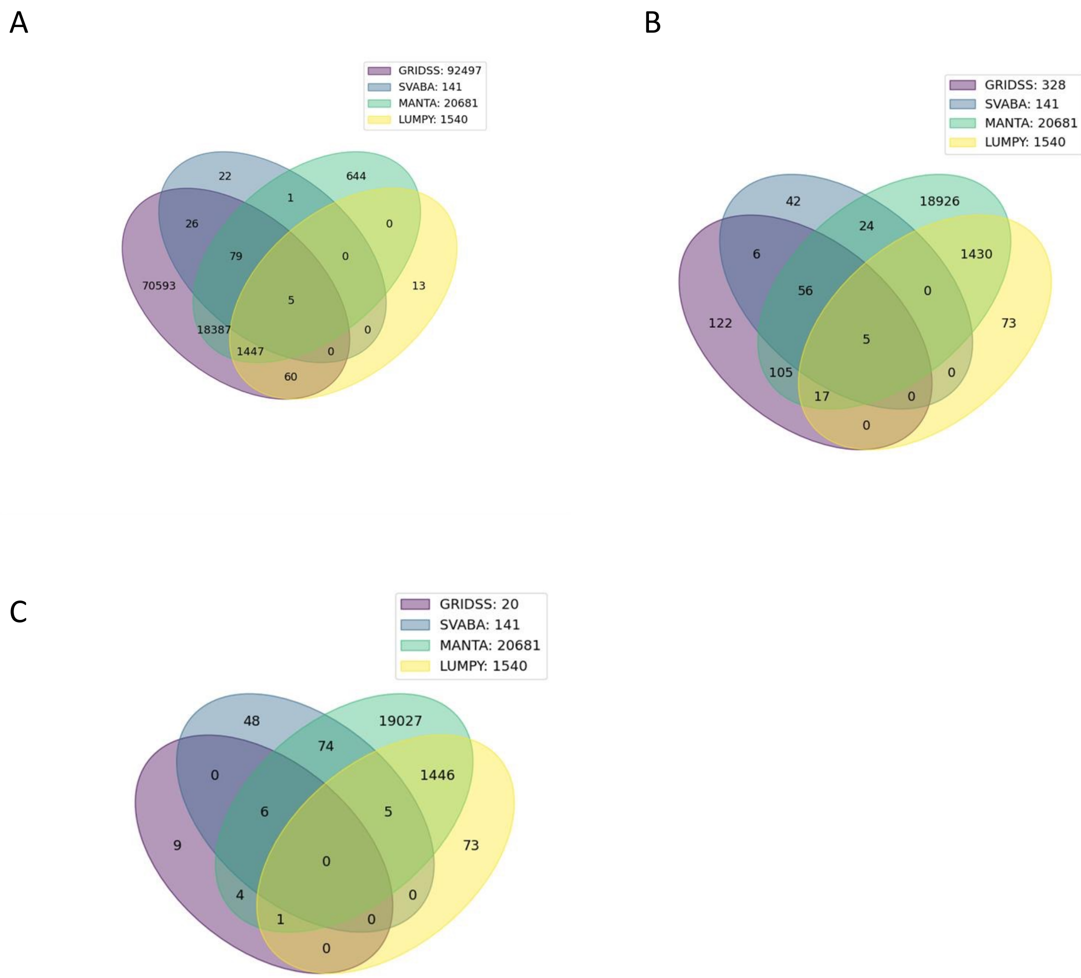


Figure 4.5: GRIDSS calls many more translocations than other callers
Venn-diagrams of interchromosomal translocations calls with no GRIDSS score cut-off (A), a GRIDSS score cut-off of 500 (B), and a GRIDSS score cut-off of 1000 (C). GRIDSS is Purple, SVABA is Blue, MANTA is Green, and LUMPY is Yellow.

Table 4.5: Comparison of translocation calling algorithms at 12X-2X coverage for FISH probed translocations

SAMPLE	FISH	12X				1X				10X				9X			
		GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY
MM08	t(8;14)	Y	Y	Y	N	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	N
MM12	t(4;14)	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N	N	N
	t(3;8)	Y	Y	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
	t(8;15)	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N
MM1S	t(14;16)	Y	N	Y	N	Y	N	Y	N	Y	N	N	N	Y	N	N	N
MM29	t(3;8)	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
MM30	t(14;16)	Y	N	N	N	Y	N	N	N	Y	N	N	N	Y	N	N	N
MM40	t(14;16)	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	N	N	N	N
MM46	t(11:14)	Y	Y	N	N	Y	N	N	N	Y	N	N	N	Y	N	N	N
MM68	t(11;14)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	t(3;8)	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y	N	Y	N	Y	N
MM75	t(3;8)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
MM97	t(4;20)	Y	N	Y	N	Y	N	Y	N	Y	N	N	N	Y	N	N	N

SAMPLE	FISH	8				7X				6X				5X			
		GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY
MM08	t(8;14)	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
MM12	t(4;14)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	t(3;8)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	t(8;15)	Y	Y	Y	N	Y	Y	Y	N	Y	Y	N	N	N	N	N	N
MM1S	t(14;16)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
MM29	t(3;8)	Y	N	Y	N	Y	N	Y	N	N	N	N	N	N	N	N	N
MM30	t(14;16)	Y	N	N	N	Y	N	N	N	Y	N	N	N	Y	N	N	N
MM40	t(14;16)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
MM46	t(11:14)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
MM68	t(11;14)	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	N	N	N
	t(3;8)	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
MM75	t(3;8)	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	N	N	N
MM97	t(4;20)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

SAMPLE	FISH	4X				3X				2X			
		GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY	GRIDSS	svABA	MANTA	LUMPY
MM08	t(8;14)	N	N	N	N	N	N	N	N	N	N	N	N
MM12	t(4;14)	N	N	N	N	N	N	N	N	N	N	N	N
	t(3;8)	N	N	N	N	N	N	N	N	N	N	N	N
	t(8;15)	N	N	N	N	N	N	N	N	N	N	N	N
MM1S	t(14;16)	N	N	N	N	N	N	N	N	N	N	N	N
MM29	t(3;8)	N	N	N	N	N	N	N	N	N	N	N	N
MM30	t(14;16)	Y	N	N	N	Y	N	N	N	N	N	N	N
MM40	t(14;16)	N	N	N	N	N	N	N	N	N	N	N	N
MM46	t(11:14)	N	N	N	N	N	N	N	N	N	N	N	N
MM68	t(11;14)	Y	N	N	N	Y	N	N	N	Y	N	N	N
	t(3;8)	N	N	N	N	N	N	N	N	N	N	N	N
MM75	t(3;8)	N	N	N	N	N	N	N	N	N	N	N	N
MM97	t(4;20)	N	N	N	N	N	N	N	N	N	N	N	N

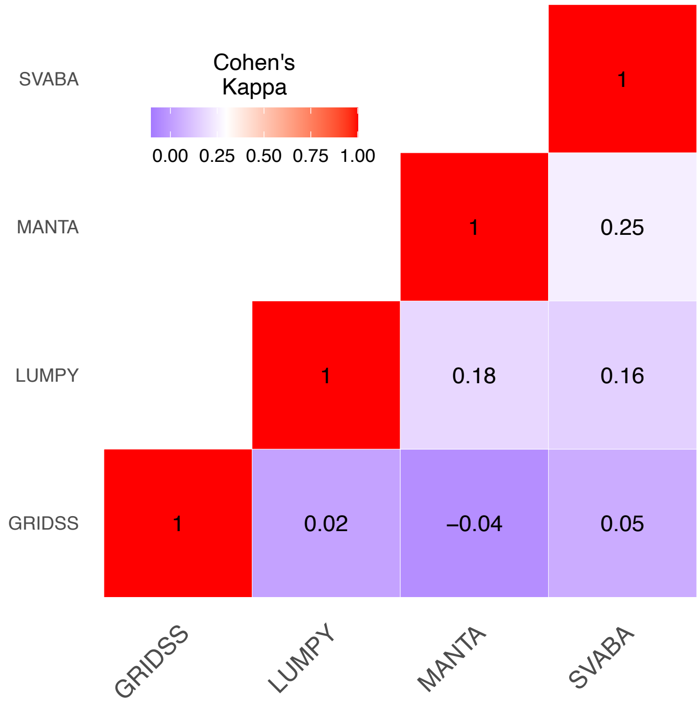
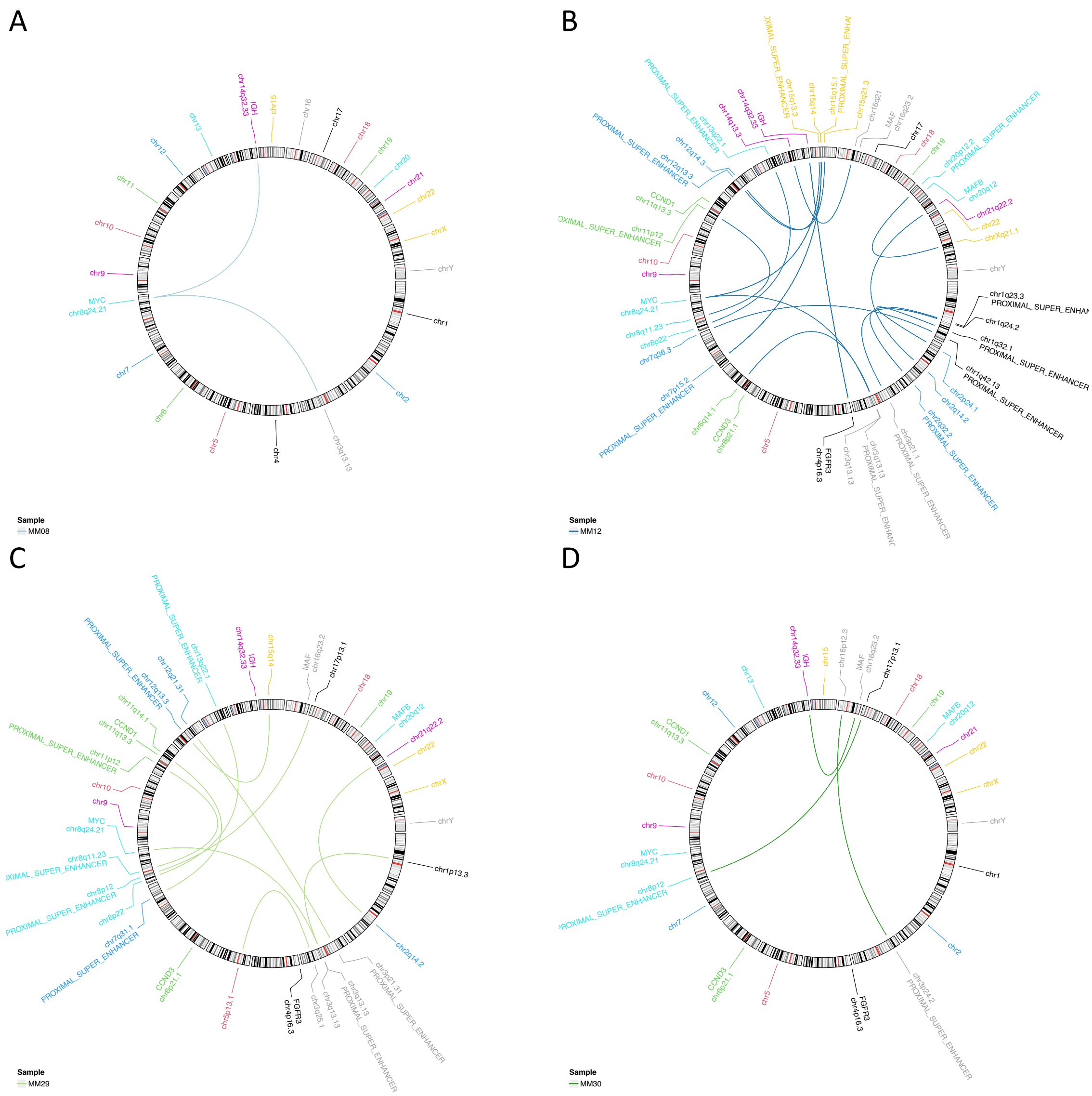
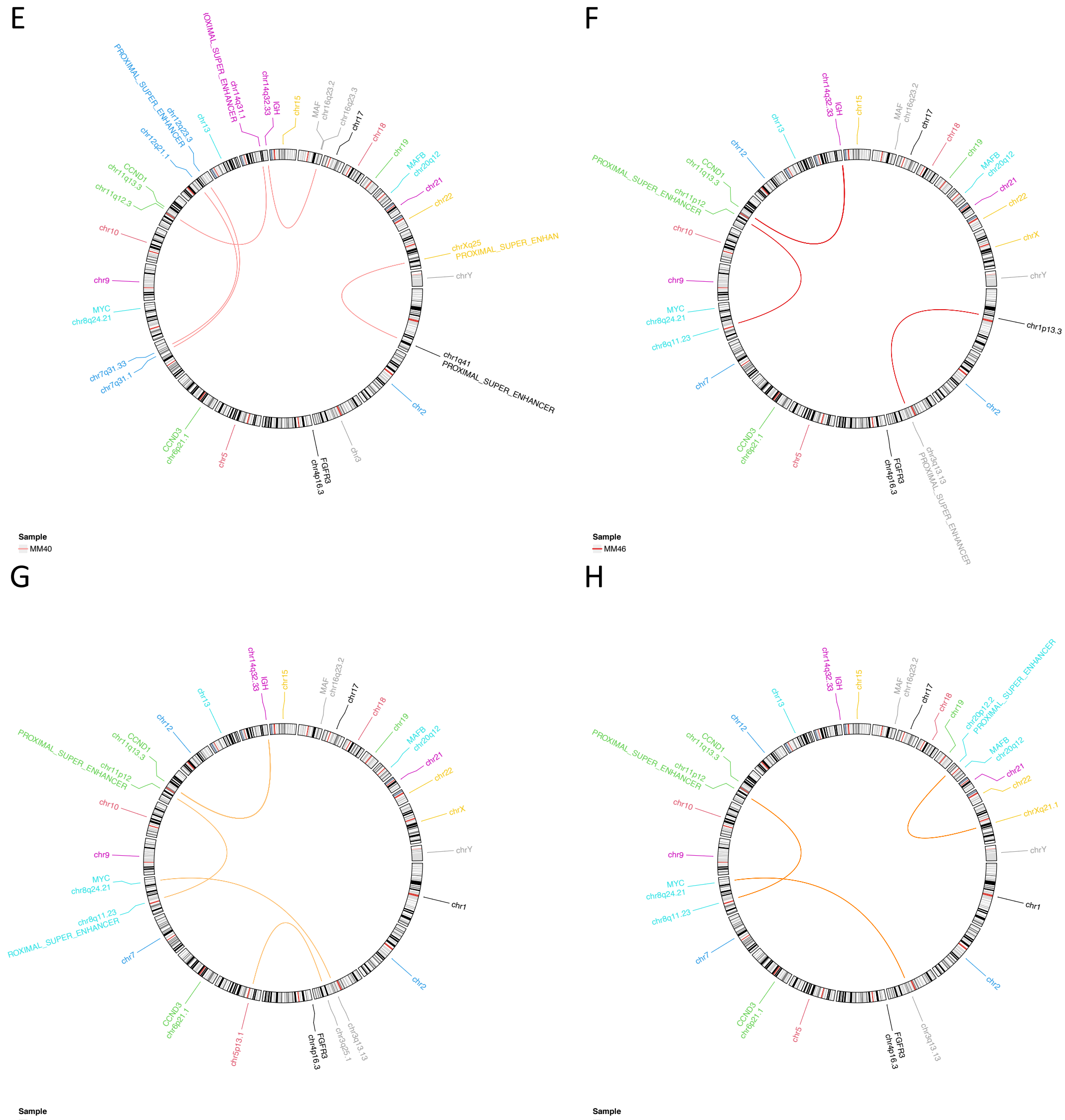


Figure 4.6: Translocation callers capture a diverse set of lesions
Heatmap showing the Cohen's kappa (concordance between two binary classifiers) for all translocations in the filtered call set.


A
B
C
D
Sample
MM08
Sample
MM12
Sample
MM29
Sample
MM30


E
F
G
H
Sample
MM40
Sample
MM46
Sample
MM68
Sample
MM75

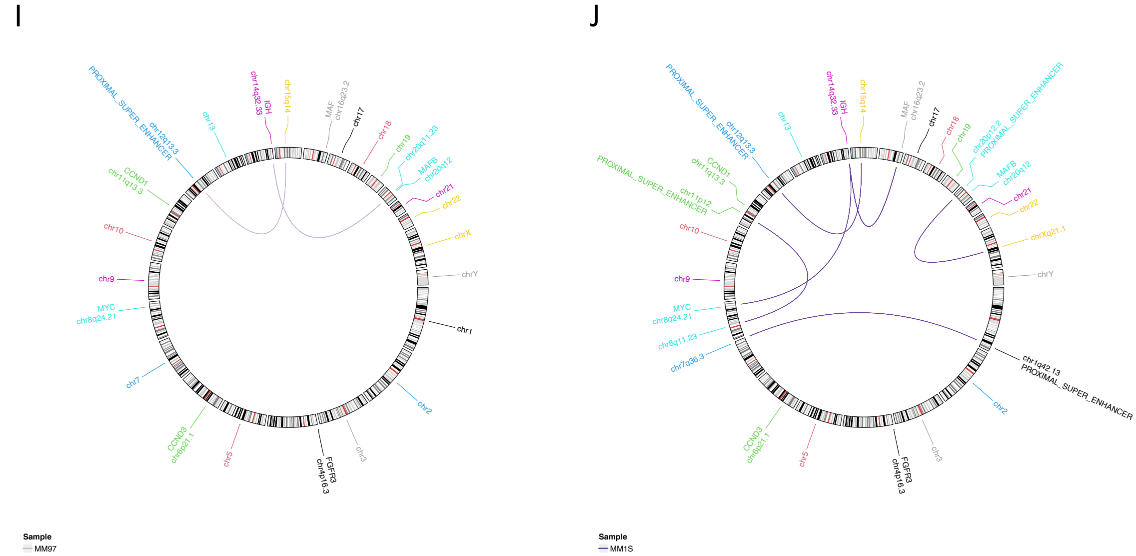


Figure 4.7: WGS identifies more translocations than FISH
Circos plots of interchromosomal translocations identifies by WGS. A) MM08. B) MM12. C) MM29. D) MM30. E) MM40. F) MM46. G) MM68. H) MM75. I) MM97. J) MM1S.

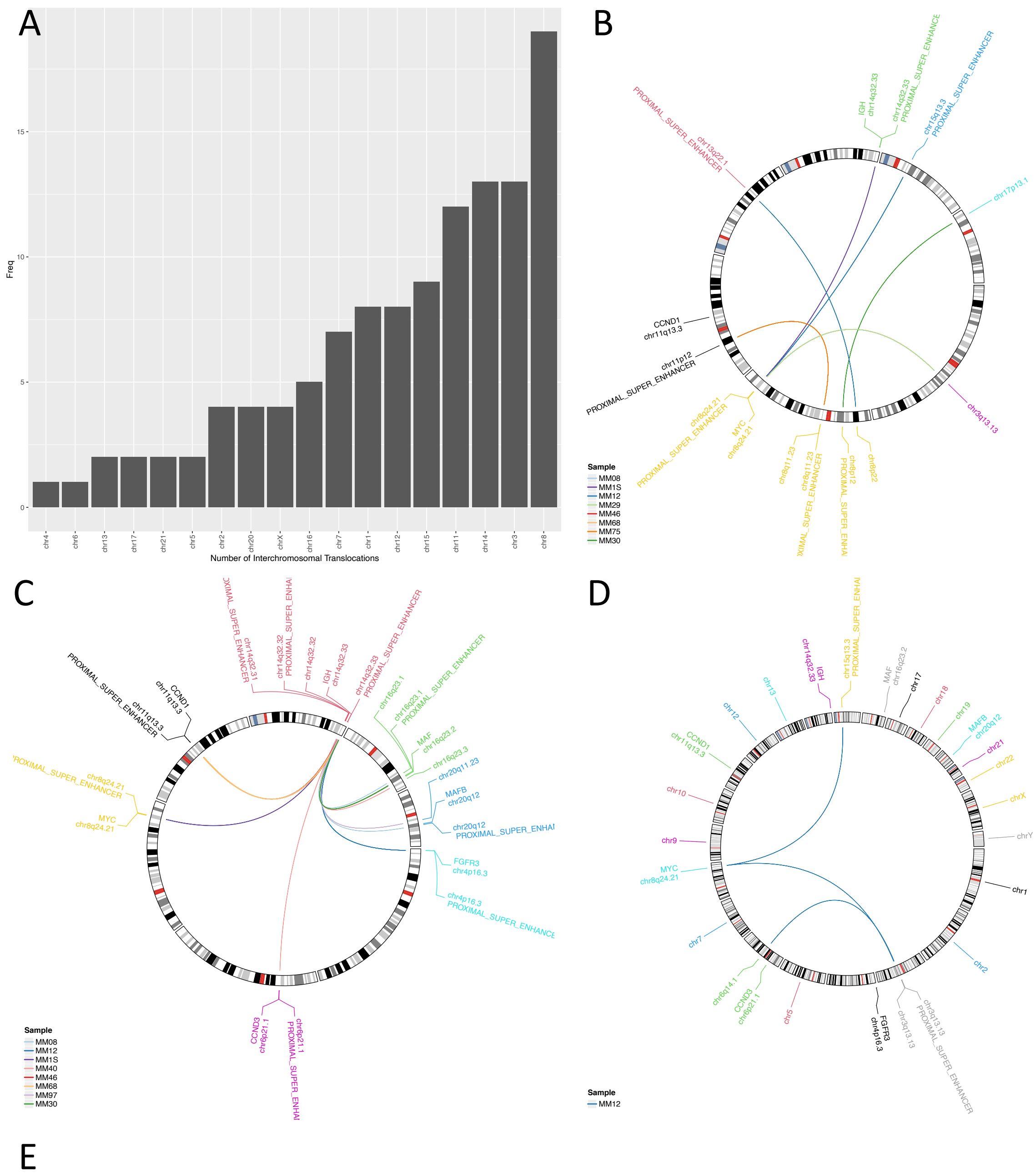


Figure 4.8: WGS identifies a more complex translocation landscape than FISH
A) Number of break points mapping to each chromosome. B) Circos plot showing all interchromosomal translocations identified by WGS with a break end on chromosome 8. C) All canonical MM primary translocations identified by WGS. D) Complex *MYC* translocations in MM08. E) Complex *MYC* translocation in MM75

At the *IgH* locus, WGS-based structural variant calling reported 6 classical myeloma primary translocations (t(4;14), t(6;14), t(11;14), t(16;14), and t(14;20)) and 5 classical myeloma secondary translocations (translocations at the *MYC* locus) (***Table 4.1***; ***Appendix Table* 4**; ***Figure 4.8* B,C**). The primary translocations included one t(4;14) (MM12), one t(14;20) (MM97), two t(11;14) (MM46 and MM68), and three t(14;16) (MM30, MM40, MM1S) (***Table 4.1***; ***Appendix Table* 4**; ***Figure 4.7*A-J**, ***Figure 4.8* C**). Secondary translocations involving the IgH locus were found in MM08, and MM1S, each having at least t(8;14) (***Table 4.1***; ***Appendix Table* 4**; ***Figure 4.7* A,H,J**, ***Figure 4.8* B,C**).

Secondary translocations at the *MYC* locus were identified in a total of 6 patients, 5 of which involved *CD96* on chromosome 3. Three samples, MM1S, MM08, and MM12 harboured complex *MYC* rearrangements (***Figure 4.7* A,B,J**). MM12 was the most complex, having 3 linked translocations involving chromosomes 3, 6, 8, and 15, and the breakpoint loci on chromosomes 3 and 15 were less than 50 kb from a super enhancer (***Figure 4.7* B**; ***Figure 4.8* D**, ***Appendix Table* 4**). MM1S involved two chromosomes, 14 (*IgH* locus), and 16 with the breakpoints on 16 and 14 each being less than 50 kb from a super enhancer (***Figure 4.7* J**, ***Appendix Table* 4**). MM08 also involved two chromosomes, 14 (IgH loci), and 3 (***Figure 4.7* A**, ***Appendix Table* 4**). Many *MYC* translocations had break ends proximal to a super enhancer (***Figure 4.7***).

FISH identified translocations in our cohort are described in Table 2. All primary *IgH* translocations called by FISH had corresponding WGS calls (***Table 4.1***). MM29 had t(4;14) with a normal *MYC* locus reported by FISH on the sample at diagnosis, while t(14;16) and t(3;8) was reported by WGS on the post-relapse sample (***Table 4.1***). WGS captured additional information in MM97; a separation at the *IgH* locus with an unknown partner chromosome was called by FISH; WGS-based structural variant calling identified the partner to be chromosome 20 at the common MAFB locus, though outside of the probe binding region (***Table 4.1***). When comparing against FISH as the 'gold standard', WGS-based structural variant detection had a sensitivity, specificity, PPV, and NPV 93%, 89%, 88%, and 94%, respectively (Table 9). The calls discordant between WGS and FISH were a false negative WGS call for a *MYC* translocations and a false positive WGS call for t(14;20) in MM97, and a false positive WGS call for a *MYC* translocation in MM12. The MM97 t(14;20) break end was outside of the fish probe binding region from this translocation, and the MM12 *MYC* translocation was complex, which may have made it cryptic to FISH. Excluding these from the comparison, WGS-based structural variant detection had a sensitivity, specificity, PPV, and NPV 93%, 100%, 100%, and 94%, respectively (***Table 4.7***). Considering only R-ISS high-risk translocations, t(4;14) and t(14;16), WGS-based SV calling had a sensitivity, specificity, PPV, and NPV that were all 100% (***Table 4.8***)

Table 4.6: Performance of WGS translocation calling compared to FISH

	FULL DEPTH	12x	11x	10x	9x	8x	7x	6x	5x	4x	3x	2x
TP	15	14	14	14	13	8	6	6	5	4	4	2
TN	15	16	16	16	16	17	17	17	18	18	18	18
FP	3	2	2	2	2	1	1	1	0	0	0	0
FN	0	1	1	1	2	7	9	9	10	11	11	13
Sensitivity	1.00	0.93	0.93	0.93	0.87	0.53	0.40	0.40	0.33	0.27	0.27	0.13
Specificity	0.83	0.89	0.89	0.89	0.89	0.94	0.94	0.94	1.00	1.00	1.00	1.00
PPV	0.83	0.88	0.88	0.88	0.87	0.89	0.86	0.86	1.00	1.00	1.00	1.00
NPV	1.00	0.94	0.94	0.94	0.89	0.71	0.65	0.65	0.64	0.62	0.62	0.58

Table 4.7: Performance of WGS translocation calling compared to FISH excluding likely FISH errors

	FULL DEPTH	12x	11x	10x	9x	8x	7x	6x	5x	4x	3x	2x
TP	15	14	14	14	13	8	6	6	5	4	4	2
TN	15	16	16	16	16	16	16	16	16	16	16	16
FP	1	0	0	0	0	0	0	0	0	0	0	0
FN	0	1	1	1	2	7	9	9	10	11	11	13
Sensitivity	1.00	0.93	0.93	0.93	0.87	0.53	0.40	0.40	0.33	0.27	0.27	0.13
Specificity	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
PPV	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NPV	1.00	0.94	0.94	0.94	0.89	0.70	0.64	0.64	0.62	0.59	0.59	0.55

Table 4.8: Performance of WGS translocation calling compared to FISH on R-ISS high-risk translocations

	FULL DEPTH	12x	11x	10x	9x	8x	7x	6x	5x	4x	3x	2x
TP	3	3	3	3	2	1	1	1	1	1	1	0
TN	4	4	4	4	4	4	4	4	4	4	4	4
FP	0	0	0	0	0	0	0	0	0	0	0	0
FN	0	0	0	0	1	2	2	2	2	2	2	3
Sensitivity	1.00	1.00	1.00	1.00	0.67	0.33	0.33	0.33	0.33	0.33	0.33	0.00
Specificity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
PPV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	N/A
NPV	1.00	1.00	1.00	1.00	0.80	0.67	0.67	0.67	0.67	0.67	0.67	0.57

4.4.6 Structural Variant Calling with Decreasing Sequence Depth

The sensitivity, specificity, PPV, and NPV of WGS-based structural variant calling decreased with lower depth when comparing to FISH as the benchmark. WGS performance was stable down to 10x coverage, and after 9X, sensitivity decreased quickly, though specificity and PPV remained high (***Table 4.6***; ***Table 4.7***). The translocation t(14;16), in MM40, a high-risk lesion, was only captured from 12X-10X, and was the single FISH assessable lesion not captured at 9X (***Table 4.1***). Hence, considering only R-ISS high-risk translocations, sensitivity, specificity, PPV, and NPV were each 100% down to and including 10X (***Table 4.8***).

Comparing to the filtered 12X WGS SV call set, the sensitivity of WGS-based SV calling quickly decreased with subsampling, while the PPV increased and stabilized (***Table 4.9***). At depths of 11X and 10X, specificity, and PPV were both 100%, while sensitivity and NPV were 78% and 73% at 11X, and 64% and 59% at 10X(***Table 4.9***). Iterating over subsampled depth, and comparing to full-depth WGS calls, WGS-based SV calling had an area under the curve of 0.89 (95% CI: 0.86-0.92), and Youden's Index suggests an optimum depth of 10x (***Figure 4.9***).

4.4.6 Combining WGS with Targeted Sequencing Better Captures Prognostic Profile

All samples included in the present study were previously assessed on the DMG26 mutation profiling panel, and 25 mutations had been identified in these (***Figure 4.10***).One patient, MM12, carried no mutations identified by the DMG26; *KRAS* was the most frequently mutated (4 samples), and each patient had an average of 2.2 mutations (range 0-6). MM68 and MM08 were designated as high-risk by the DMG26 (chapter 3).

MM08, MM12, MM97, and MM75 all had del(1p), and MM40 had del(13q) identified by WGS. Importantly, MM40 carried a *DIS3* mutation, and hence had a biallelic event at *DIS3*, while. MM08 had both a panel identified high-risk mutation in *DIS3*, ad a high risk del(1p). Both of these patients would hence be within a double-hit high-risk group; and consistently with this, these patients had a short survival of 600, and 647 days, respectively.

Table 4.9: Performance of subsampled WGS translocation calling compared to full-depth WGS translocation calling

	12x	11x	10x	9x	8x	7x	6x	5x	4x	3x	2x	1x
TP	52	41	30	19	14	9	7	3	2	2	1	0
FP	14	11	9	6	2	2	0	0	0	0	0	0
TN	0	3	5	8	12	12	14	14	14	14	14	14
FN	36	47	58	69	74	79	81	85	86	86	87	88
Sensitivity	0.59	0.47	0.34	0.22	0.16	0.10	0.08	0.03	0.02	0.02	0.01	0.00
Specificity	0.00	0.21	0.36	0.57	0.86	0.86	1.00	1.00	1.00	1.00	1.00	1.00
PPV	0.79	0.79	0.77	0.76	0.88	0.82	1.00	1.00	1.00	1.00	1.00	N/A
NP	0.00	0.06	0.09	0.12	0.16	0.15	0.17	0.16	0.16	0.16	0.16	0.16

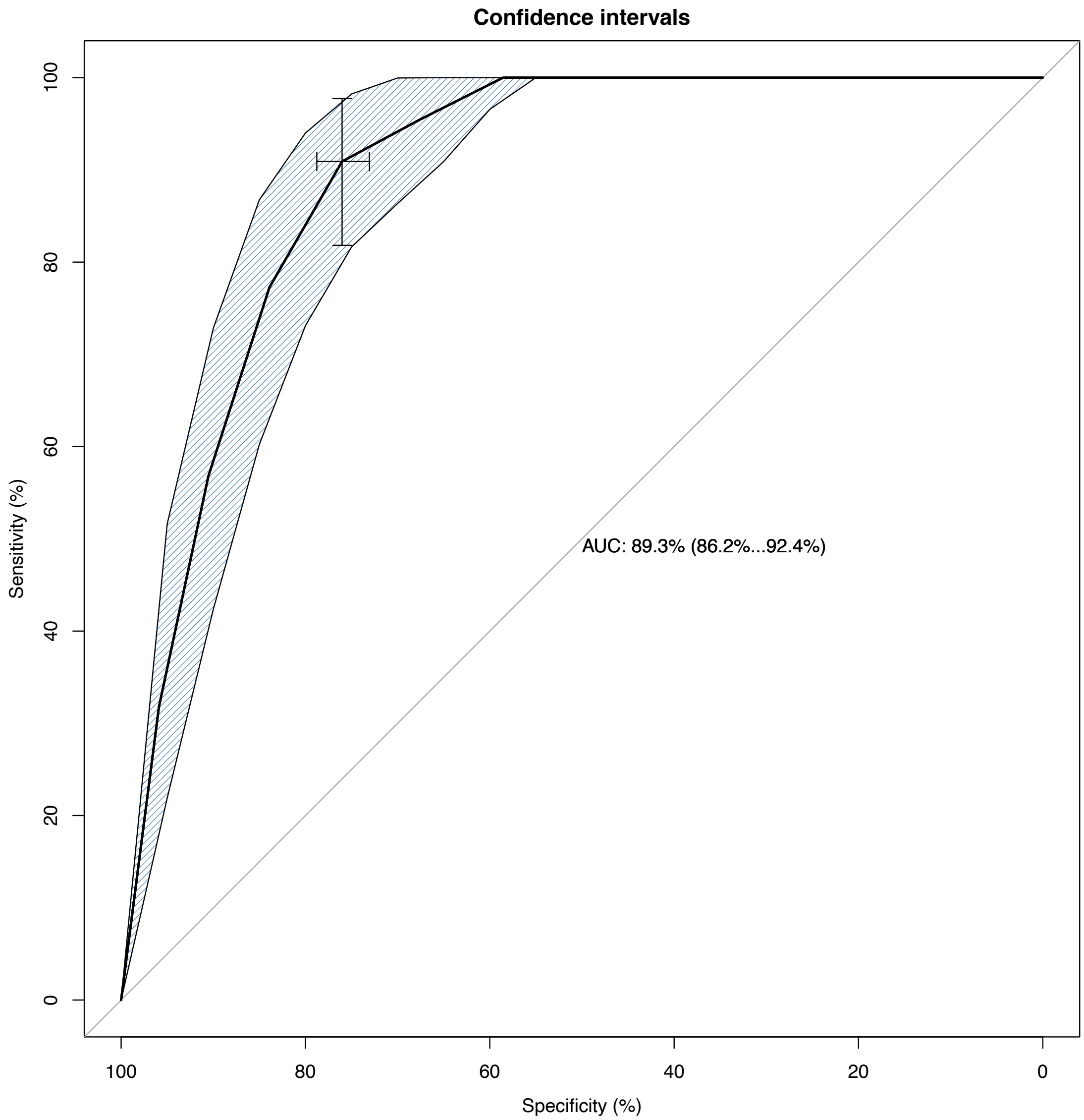


Figure 4.9: WGS is sensitive and specific compered to FISH
ROC curve comparing WGS-based translocation calls to FISH translocation calls, iterating over sequencing depth.

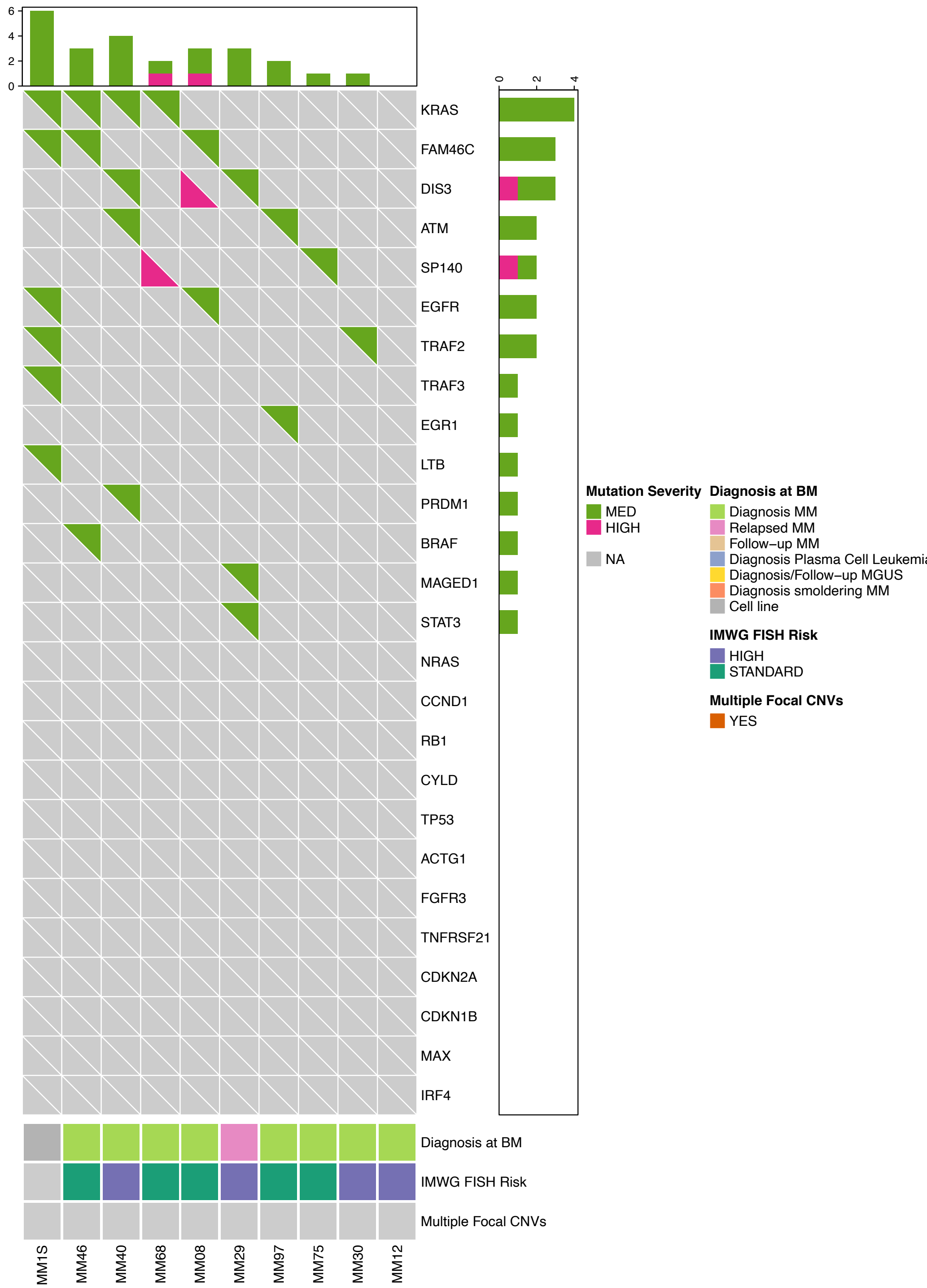


Figure 4.10: High-Risk double-hit group captured with combination of WGS and targeted sequencing
Oncoprint of mutations identified in WGS assessed cohort

4.5 Discussion:

Resolving the diverse set of genomic abnormalities underpinning multiple myeloma is crucial for current clinical practices, and to advance precision medicine in this disease.[19,382] Currently, clinical FISH assessment of MM genomes only characterizes a select set of the translocations at the *IgH* and *MYC* loci, along with common translocation partners, and CNVs at specific sites across the genome. This information is subsequently incorporated into staging methods and has shown a modest utility in guiding precision treatment decisions.[19,118] A growing body of evidence demonstrates that clinically important genomic lesions which are either cryptic, or untargeted by FISH, can be resolved by NGS technologies.[9,20,21,26,161] We demonstrate that canonical translocations at the *IgH* locus are resolved by WGS-based assessments at clinically viable coverage depths of 12X to 9X. Important high-risk translocations and CNVs missed by FISH were resolved by WGS, such as del(1p) in MM08, MM12, and MM97 and t(14;20) in MM97 as well as complex translocations in MM08 and MM12.

In our cohort, WGS-based CNV and SV calling were not only highly concordant with FISH, but also identified important, prognostically relevant, and precision medicine informing translocations where FISH failed. For example, MM97, FISH called an IgH translocation but did not identify the partner chromosome despite probing for all classic IgH partner chromosomes. Through WGS, we identified the partner chromosome to be 20, and the breakpoint to be proximal to the canonical *MAFB* locus. Importantly, this translocation is a marker for poor prognosis, and strongly indicates the use of novel therapies.[15,161,383] The single unmatched CNV call by FISH was trisomy 9 in MM75. This trisomy was identified in 11 (22%) of 50 assessed cells. Our CNV calling threshold was set to identify trisomies in 10% of cells or greater. Nonetheless, modest positive fold changes are observed across this chromosome.

Though FISH and WGS-based variant calls were highly concordant at probe binding sites, the interpretation of some FISH probes were inconsistent with WGS calls. Most notably, del(1q) was called in only one patient, MM75, while WGS called del(1q) in four patients, MM08, MM12, MM97, and MM75. For MM08, MM12, and MM97 the WGS calls were centromeric to the FISH binding area on 1q around *TP73*. Genes whose lower expression is a significant indicator of poor outcome map throughout the q arm of chromosome 1 and deletions along 1q are adverse indicators.[97] Though not included in the R-ISS, a recent large-scale analysis identified del(1q) to be one of four variants carrying significant hazard in a multivariate assessment.[19,24] Hence, missing this lesion represents a significant downfall of FISH, and emphasizes the benefit of the non-targeted nature of WGS-based assessments.

In addition to chromosome 1 abnormalities, WGS identified numerous other CNVs with unmatched FISH data that, while of lesser risk than 1p and 1q copy number changes, are of considerable clinical impact. Prominently, the prognostically unfavourable deletion of the MM tumour suppresser, *CYLD*, was detected and in three patients.[384] Beyond this lesion's prognostic impact, it may also be informative for guiding therapeutic decisions.[277,385] Trisomy of chromosome 5 in MM97 is also notable as this is currently the only known favorable prognostic marker formalized in a risk algorithm, namely, the prognostic index.[19]

Deletions affecting the *IgH* locus on chromosome 14 occur in 22% of newly diagnosed patients, have prognostic relevance, and are indications pomalidomide based therapeutic regimens.[386] We identified IgH deletions in MM1S, MM97, MM75, MM46, MM40, MM30, MM29, and MM08, while an amplification was observed in MM68; only in MM75 was a CNV at the *IgH* locus probed for and identified by FISH (Table 4, Figure 1A,C,D,E,F,H,I). This is positive indication for the sensitivity of WGS based detection, as translocations in the context of CNVs can be cryptic to FISH. Additionally, this further highlights the benefit of untargeted assessments as compared to FISH.

WGS also resolved a far more complex landscape of translocations than did FISH in all samples. This complexity more accurately reflects the underlying genetic aberrations, and is most highlighted in translocations involving the *MYC* locus, which often involved more than two chromosomes.[88,161,387] Consistent with other reports, we identified *MYC* translocations by WGS in more patients than did FISH.[161] Furthermore, we found that break ends involved in complex or simple *MYC* translocations often brought it to within 50 kb of a super enhancer, which is consistent with *MYC* translocations commonly driving *MYC* overexpression.[388] *MYC* overexpression is a significant indicator of poor prognosis and may be a precision medicine indicator.[157]

The sequencing depth for our study ranged between 1X and 12X, and from 9-12X our analysis for SVs and CNVs was performant. Using Youden's index, which is an indication of the optimal cut-off point in a binary classification, indicates that 10X depth is ideal. Direct sequencing costs at a Canadian genome sequencing centre was around $300 (CAD) per patient, which is well below the cost per patient for referred FISH at the MAYO clinic of approximately $1500 (CAD). We did not assess for CNVs on subsampled data as this analysis was previously successfully performed at much lower depths; well below 1X for all major CNVs of interest.[26] We initially set all samples to ~12X coverage, and translocation calls t(14;16) in MM29 and t(8;14) in MM75 which were captured at the original depths of 17.44X, and 15.04X, respectively, respectively, were not identified at 12X (**Table *4.3***). Importantly, t(14;16) is considered to be a high-risk marker, although this has not been confirmed by the Intergroupe Francophone du Myelome (IFM) group, and its designation as high-risk has been disputed by other groups, especially in the context of modern therapies.[24,389] Loss of the *MYC* translocation in MM75 at 12X may be attributable to the complexity of the alteration, which may require higher sequencing depth. Subsampling below 9X yielded progressively less sensitive calling, hence, clinical WGS-based SV calling should be performed at least at 10x depth, with preference for higher depths where feasible.

In this work we limited our analysis to interchromosomal translocations. While all GRIDSS, svABA, LUMPY, and MANTA report intrachromosomal rearrangements, a high-quality dataset of these lesions in MM is not extant, posing significant challenges to filtering for high-quality variants from the total call set, especially given our relatively low-depth experiment. Furthermore, intrachromosomal rearrangements are not currently considered to modulate a patient's risk profile.[88] Importantly, the CNVs and breakends identified by WGS based assessment may be used as evidence for chromothripsis or chromoplexy, which have emerging prognostic implications.[94] Hence, further work on the viable depths at which these more complex rearrangement profiles may be described is an important future direction which may further highlight the value of WGS as compared to FISH.

While we demonstrate high performance of standard short-read WGS for SV calling, superior performance may be achieved using novel long-read (pacbio, oxford nanopore), or synthetic long-read (10X genomics) sequencing approaches. Long-read sequencing is indisputably a better approach for SV detection and allele reconstruction.[49] However, it requires higher-quality DNA, more technical proficiency, higher costs, and often additional equipment.[49,50] Hence, maximizing short read WGS performance is an attractive option for clinical use, with capable sequencers being found in many molecular diagnostic laboratories. A additional short-coming of our analysis is the lack of a true 'gold-standard' assessment, against which to compare the WGS variant calls. While FISH data was available for all patients, FISH has its own sources of error, and thus, even as a current benchmark for MM genomic interrogation, it has shortcomings. Known limitations of FISH include poor performance for unbalanced translocations, and cryptic copy number variants. Furthermore, not all classic lesions were represented in our cohort. Nonetheless, standard WGS reliably detected prognostically important genomic lesions reported by FISH, and identified such lesions where FISH was not targeted or presumptively misreported. Thus, WGS-based assessment for primary genomic assessment of MM is an attractive and more robust alternative to FISH.

Chapter 5 Discussion

5.1 DMG26: Panel Sequencing for Myeloma Genome Interrogation:

The landscape of small mutational events existent within MM has only recently begun to be elucidated, and use of this information in clinical settings remains low. We have developed a targeted sequencing approach, the DMG26, which assesses the mutational landscape of the most frequently altered genomic loci in MM. Importantly, we designated 'high-risk' mutations within these loci, and demonstrated that they were significantly associated with shorter progression free survival. An additional and striking feature of these 'high-risk' markers is that they were mutually exclusive with classical 'high-risk' FISH markers, and were independent of the established R-ISS biochemical parameters. Thus, mutational signatures, as measured by the DMG26, represents a potentially powerful and novel prognostic approach, capturing high-risk patients that may otherwise be incorrectly classified by classic stratification schemes such as the R-ISS. Furthermore, beyond current prognostic utility, 33% of DMG26 identified variants in genes that are currently, or soon-to-be targetable by precision medicines, and therefore may offer therapeutic guidance in future.

Another importantly aspect of our panel is that it identified prognostic relevance of single nucleotide variants where most other studies have not[9,13,16]; although this finding still needs to be replicated in a larger cohort. A potential benefit to our study is that, in addition to assessing on a gene-by-gene basis – as done in the M(3)Pv2.0[21] study and the large-scale WES MM assessments[9,13,16] – we also classified mutations into high- and low-risk groups using contemporary computational methods and assessed across all genes. Evaluating for mutational risk contributions in this manner may be why we have identified significant mutational risk contributions where other studies have failed. Similar strategies have been successfully applied to diverse clinical scenarios including: (i) other cancers such as breast, pancreatic, colorectal, biliary, ovarian, and nasopharyngeal cancer[390]; (ii) including identifying genomic variation implicated in Autism Spectrum Disorder[391]; (iii) inherited cardiomyopathies[392]; (iv) RASopathies and epilepsy[393]. While in our study, designating mutations as either pathogenic or not through computation approaches did successfully identify high-risk patients, it is imperative to acknowledge that mutational contributions to risk are not binary nor isolated, and a model which accounts for this will be more performant.

Beyond capturing prognostic relevance where many other studies have not, our analysis was, strikingly, far more performant than both R-ISS FISH marker and ISS based stratifications. One other panel assessment, the M(3)Pv2.0, has reported a few genes to be of prognostic significance for both progression-free and overall survival.[21] Our assessment corroborated their finding of *PRDM1* mutations conferring hazard; however, we also found mutation of *CDKN1B* to confer risk, while the M(3)Pv2.0 did not. Consequently, the authors claimed modest prognostic relevance of mutations identified on their panel.[21] In contrast, our assessment with the DMG26 captures important risk lesions that stratifies risk in MM patients more robustly.

Multiple myeloma, distinct from other cancers, carries concurrent *NRAS* and *KRAS* mutations in approximately 15% of RAS mutated samples.[172] While we did not observe this in our cohort, we did observe cases of other concurrent mutations in the MAPK pathway. *NRAS* and *BRAF* were co-mutated in MM99. In MM88, MM1S, and RPMI-8226, *BRAF* and *KRAS* were co-mutated. Identification of concurrent RAS-RAF mutations are key in ensuring appropriate therapeutic choice, as vemerafinib, which is indicated for by *BRAF* mutations on their own, is ineffective in contexts with RAS mutation.[394] Interestingly, in MM52, *KRAS* (c.240delT, p.Cys80fs, high-severity but low-risk due to VAF cut-off) and *FGFR3* (c.1117A>G, p.Leu684His, low-risk) were co-mutated. *FGFR3* mutations are infrequently observed with RAS mutations, which has been attributed to the stronger and more robust activation of MAPK signaling consequent to *FGFR3* mutation as compared to RAS mutations.[172] Neither of these MM52 mutations are reported in COSMIC nor ClinVar, nor are indicated in the literature to modulate MAPK activation. Nonetheless, these may pose important considerations, as addressing only *FGFR3* or *KRAS* independently may have reduced efficacy in a similar manner to concurrent *BRAF*/RAS mutations.

Myeloma tumour suppressors on our panel, including *Rb1*, *TP53*, *FAM46C*, *CYLD*, and *DIS3* all carried high rates of high-risk mutations at 25%, 22.2%, 20%, 12.5%, and 11.5%, respectively. On average, 17.1% of mutations in these genes were considered to be high-risk, as compared to only 3.9% of mutations in the other DMG26 genes. The abundance of high-risk lesions within these genes is consistent with findings in late-stage relapse or refractory myeloma, where mutation, LOH, and deletion of tumour suppressors is common.[16] Likely, mutation in these genes precedes relapse, and early capture of mutational events may be an important biomarker of progression if pathogenic mutations can be effectively delineated from mundane variation. Our study offers compelling evidence that the computational methods used herein can be effectively employed to this end within clinical settings.

An important ancillary finding of this assessment is that FISH did not significantly stratify high-risk from standard- or moderate-risk patients. Our analysis represents the first modestly sized assessment of the Atlantic Canadian multiple myeloma population. These findings may warrant further investigation of peculiarities in the relationship between cytogenetic loci and clinical outcome in this patient population.

5.2 DMG26 Compared to other Proposed Clinical Oncology Panels

The DMG26's design was informed by large-scale whole-exome and -genome sequencing studies of myeloma.[9,13,16] While some genes included in the panel are uniquely implicated in MM pathogenesis (e.g. *FAM46C*), there are numerous genes which are nearly ubiquitously mutated across cancer types (e.g. *NRAS*, *KRAS*, *BRAF*, *TP53*, etc.).[227,228] Identifying which genes on the panel capture significant risk may be of economic interest, as currently available commercial panels – such as Illumina's TruSight Myeloid or TruSight Oncology 500 (TSO500) panels – may completely or largely overlap with this subset of genes. However, it is important to counterbalance current economic interest with the potential clinical value of genomic assessment tailored to MM. For example, mutations of *FAM46C*, which are specific to MM and are consequently not assessed for by the myeloid panel, are indications for burgeoning targeted therapies, such as CFI-400945 which is in a phase I/II clinical trial.[225]

To this end, a few MM specific panels have been developed, such as the M(3)pv2.0 and myTYPE.[20,21] In a clinical assessment of the M3Pv2.0 on newly diagnosed MM patients with 47.2% receiving a Bortezomib based regimen, only modest prognostic relevance was reported.[21] Despite the DMG26 performing markedly better on our cohort than did the M3Pv2.0 on their similarly treated cohort, the M3Pv2.0 is considerably larger than the DMG26, and the panels overlap greatly, with *LTB*, and *MAGED1* being the only genes specific to the DMG26. Similarly, the myTYPE panel, which is much larger than both the DMG26 and M3Pv2.0, and assesses all genes included in our panel, did not report mutations that significantly associate with prognosis.[20] As a large hybrid capture panel though, the myTYPE, unlike the DMG26 and M3Pv2.0 panels, is able to capture the broad range of genomic lesions present within an MM genome, including IgH translocations and CNVs.[20] This is an important advantage, as the concurrent capture of all MM variant types in a single clinical test is ideal. The myTYPE panel is, however, limited in its capacity to detect t(8;14) and LOH, and may be poorly suited to detect patterns of chromothripsis that have only been elucidated following the design of this panel.[20,88] Hence, an altered design with consideration for these clinically relevant lesions that employs are analysis approach may be an optimal MM genomic interrogation tool.

Comparing the DMG26 to Illumina's myeloid panel, *BRAF*, *CDKN2A*, *TP53*, *NRAS,* and *KRAS* are the only overlapping genes. Of these shared genes, only *BRAF* and *TP53* harboured high-risk mutations in our cohort, hence, the scope of the myeloid panel is ill-suited for myeloma prognostication. Furthermore, the DMG26 identified lesions in *FAM46C*, *ATM*, *FGFR3* which may have implications for precision medicine. *FGFR3* mutations are targetable in muscle-invasive bladder cancer by erdafitinib, AZD4547, rogaratinib and infigratinib, and may have functional consequences on the affinity of the anti FGFR3 mAb currently under assessment for MM.[395,396] ATM mutations have been associated with sensitivity of mantle cell lymphoma cells, and solid tumours to the ATR inhibitor, AZD6738.[397,398]. Mutation of *FAM46C*, as discussed above, may soon be targetable by CFI-400945.[225] Hence, beyond the prognostic relevance of genes in the DMG26 but not on the myeloid panel, the mutational state of at least 3 genes are relevant for promising targeted therapies currently approved for other cancers, or which are under varying stages of clinical assessment. Presumably, as the repertoire of targeted therapies expands, assessment of MM specific genes will likely carry increasing value for targeted treatment of MM patients.

The TruSight Oncology 500 assesses 523 genes; the large size of this panel facilitates estimation of tumour mutational burden, evaluation of microsatellite instability, CNVs, and loss of heterozygosity in addition to standard mutational profiling. These are significant genomic events with clinical significance that are not assessable by the DMG26. However, it does not cover the genes *TNFRSF21*, *ACTG1*, *TRAF3*, *MAGED1*, *SP140*, nor *LTB*. In our cohort, the DMG26 identified high-risk mutations in the latter three. Though the value of genomic interrogation of *TNFRSF21*, *ACTG1*, and *TRAF3* remains unclear, such assessment of *MAGED1*, *SP140*, and *LTB* appear to potentially carry significant prognostic value.

Expansion of the TSO 500 to include the missed DMG26 genes is an appealing future direction. First, the large panel size facilitates determination and estimation of complex and comprehensive genomic profiles, which are increasingly implicated in myeloma pathogenesis and clinical outcome.[7,88,399] Second, while myeloma is a relatively common blood cancer, small clinical centres may see too few cases to adequately fill a MM-specific sequencing run with specialized panels; larger pan-cancer panels allow for broad pooling of patients to fill sequencing runs. Of course, larger sequencing panels also often require larger sequencers with significant capital investment costs and require more depth of expertise on both the wet-lab and bioinformatics portions of library construction, analysis, and data management.

5.3 Structural Variant Calling

Structural variants in MM are the classic, and currently dominant biomarkers of this malignancy. Hence, any clinically viable genomic assessment must capture these lesions with high confidence. We demonstrated that the ensemble of GRIDSS, svABA, MANTA, and LUMPY successfully identified all FISH tested translocations at minimum depths of 10X sequencing coverage within our cohort, and identified t(14;20) in MM97 which was otherwise unreported by FISH due to the breakpoint on chromosome 20 being distal from the MAFB probe binding region. Beyond this, we also captured the complex structural variation at play at the *MYC* locus. We also reiterated our previous findings, with QDNAseq outperforming FISH in the detection of CNVs.[26] Hence, WGS is a clinically viable approach for MM genome interrogation at 10X coverage.

Previous reports have described clinically validated WGS pipelines for germline and somatic structural variant detection using single algorithms, with most recommending depths at or greater than 30X, and up to 90X to resolve highly complex oncologic genomes.[400–402] WGS at this depth is cost prohibitive for most centres, and would commonly require referral to an external laboratory with a NovaSeq (Illumina), slowing turn-around-time. Numerous performance analyses have found contemporary short-read, paired-end structural variant calling algorithms to be performant down to 10X in genomically complex contexts. However, these have not included comparisons to current clinical assessments for translocations, nor has their performance been previously examines specifically in MM.[310,327,403,404] Additionally, these studies have also iterated the performance gain of ensemble calling over single algorithm calling. Hence, our analysis, which employs an ensemble of contemporary algorithms and compares against the current clinical assessments, is an important extension of these results into MM clinical contexts, with 10X coverage facilitating variant calling concordant with FISH.

Consistent with our previous assessment using ultra-low-depth WGS for CNV detection, we again found that QDNAseq was superior to FISH in profiling the landscape of CNVs in MM.[26] QDNAseq called all CNVs reported by FISH, except for trisomy 9 in MM75, where it identified a small amplification on the q arm distal to the centromeric binding region. Overall, compared to FISH, QDNAseq was highly performant, having a sensitivity of 95% and an NPV of 100%. Comparing FISH to QDNAseq, key variants were missed, including del(1p) in patients MM08, MM12, and MM97, as well as trisomy 5 in MM97. While not included within the R-ISS, the recently suggested Prognostic Index considers del(1p) and trisomy 5 to have a prognostic weighting of 0.8 and -0.3, respectively.[19,24] Hence, missing these CNVs is a significant error by FISH, which, in combination with QDNAseq's otherwise high concordance with FISH, and the identification of additional CNVs outside the scope of FISH, is strong support for the use of WGS-based CNV profiling in MM.

In our assessment, we employed the ensemble of GRIDSS, svABA, MANTA and LUMPY for interchromosomal translocation calling. Employing an ensemble approach has become a best-practice to balance and utilize performance characteristics of different variant calling approaches and has been used in MM previously.[88,327] Our combination is broader and includes more performant algorithms than those used in the only two large-scale assessment of myeloma structural variations, which were DELLY and LUMPY[88], or MANTA[94]. We found that the callers in our ensemble were minimally concordant (max Cohen's kappa at 0.25 for svABA and MANTA), hence, the ensemble was well suited to identify and corroborate variants unidentifiable by callers individually. We also found GRIDSS to be the most sensitive caller (100%), and therefore should be included in clinical assessments, especially at lower sequencing depths.

In MM, *MYC* translocations tend to be involved in complex rearrangements which can implicate more than 8 chromosomes.[387,405] This complexity often impedes detection by FISH, which is reflected in the higher rates of *MYC* translocation reported in MM by WGS studies as compared to FISH studies.[387,405] Consistently, while *MYC* separation was called by FISH in MM08, MM29, and MM1S, only WGS based translocation calling detected the *MYC* translocation in MM12, which was complex. In addition to the improved detection of *MYC* translocation, WGS also resolves the translocation partner, which FISH does not. Unknow translocation partners of a *MYC* separation circumvent careful investigation of the clinical association of *MYC* translocations as different partner chromosomes may drive different phenotypes. Accordingly, *MYC* translocations have been inconsistently associated with differing clinical courses.[154,159,160,197,406,407] None the less, MYC overexpression is associated with poor clinical outcome.[408] Considering this, WGS clearly offers superior information compared to FISH, as translocation breakpoints can be assessed for their proximity to super enhancers, the functional impact of which may be subsequently considered.

The most important finding of this assessment is the high performance of our interchromosomal calling approach at the IgH locus down to 10X, at 100% sensitivity, as these are currently the only translocations currently associated with clinical outcome. In fact, a recent assessment of structural variation on the MMRF COMPASS cohort found that other, less recurrent translocations did not associate clinical outcome.[88] These IgH events occur during preclinical stages and are consequently clonally present. This is advantageous for detection by sequence-based approaches, as variant alleles represent nearly 50% of all alleles sequences, which reduced the coverage required to confidently identify them. The opposite is true for other structural events, which may be present at varying levels of clonality. Indeed, the sensitivity of our approach at 10X compared to the full-depth call set was only 39% for these.

5.4 Combining Targeted and Whole-Genome Sequencing Captures High-Risk Double-Hit Group

Though we did not identify patients with bi-allelic double-hit events in *TP53*, the most prognostically impactful genomic lesion in MM known, by FISH, WGS, or panel sequencing, a bi-allelic double-hit was observed in MM40, and may have been in MM43. Two mutations in *FAM46C* were reported in MM43 by panel sequencing; due to technical limitations we cannot phase the variants, hence it is unclear if the SNVs occurred in *cis* or *trans*. *FAM46C* is the second most commonly afflicted gene by bi-allelic events, and a double hit to this gene is significantly associated with poor outcome.[409] Mutation in *DIS3* and a *DIS3* spanning deletion of chromosome 13 were observed in MM40. Double hits to *DIS3* are also associated with adverse outcomes in MM patients.[410] Consistently, short progression-free survival was observed in both MM40 and MM43, at 564 and 91 days, respectively, and overall survival of 564 days for MM40.

Beyond specific genes with potential bi-allelic mutations, triple- and double-hit groups have recently been defined as those harbouring two or three high-risk lesions throughout the genome. The accumulation of high-risk lesions has progressively worse prognosis.[19,94] Within our cohort, we identified MM08, MM12, MM97, and MM75 to all have extensive deletion on 1p, a high-risk lesion. FISH identified this lesion in only one patient, MM75. Importantly, FISH and WGS identified t(4;14) in MM12, and WGS identified t(14;20) in MM97, both of which are high-risk lesions.[136] Hence, MM12 and MM97 are both in a high-risk double hit group which FISH did not assign them to, thereby highlighting the prognostic value of WGS for structural variant detection. Furthermore, MM08 was identified to have a high-risk mutation in *DIS3*, and therefore is also part of the high-risk double-hit group, thereby highlighting the prognostic value of targeted mutational profiling and comprehensive structural variant calling as a dual-test approach.

Within our cohort, the most strikingly adverse prognostic outlook was observed in patients with multiple high-risk mutations: MM43, MM63, and MM106. These carried mutations in *FAM46C*, *DIS3*, *TP53*, and/or *Rb1*, all of which are tumour suppressors in MM. Bi-allelic events in tumour suppressors are associated with poor prognosis and are more common in relapsed and refractory patients than newly diagnosed MM.[409] Notably, these patients were negative for all high-risk R-ISS FISH lesions; thus, targeted-sequencing identified quickly progressing patients previously classified as standard risk. Promisingly, mutation of these genes may be future indications for targeted therapies currently under development, or have clinical associations with current treatment strategies. High dose chemotherapy, which is administered to all patients undergoing autologous stem cell transplantation, underperforms in patients carrying *DIS3* mutated subclones; achieving an event-free survival of 27 months compared to 70 months.[232] Mutation of *FAM46C* may soon indicate for CFI-400945 therapy.[225] In pre-clinical studies, the adenovirus VCN-01 is showing promising activity against patient derived *Rb1* mutated retinoblastoma samples. Hence, in addition to capturing high-risk patients otherwise misclassified by FISH, NGS based assessments capture genomic lesions which may be future therapeutic targets that mediate the current risk designation.

5.5 Superiority of Sequencing to FISH for Genome Interrogation

Our sequencing-based assessments of myeloma genomes has highlighted two main shortcomings of FISH in the clinic. Firstly, in our targeted sequencing assessed cohort, R-ISS FISH lesions did not stratify patients into groups of significantly different survival, while a mutation-based approach did. Secondly, FISH misses prognostically important lesions. This is attributable to 1) FISH's targeted nature, which circumvents detection of any lesion outside of probe binding regions. Compounding this, FISH panels are built algorithmically with additional probes only included reflexively based on co-occurrence patterns of cytogenetic lesions with those already observed, impeding detection of cytogenetic lesions which occur infrequently together. And 2) certain structural variants, such as complex unbalanced translocations, which are common at the *MYC* locus in MM, are cryptic to FISH.

In our mutation profiling cohort, the underperformance of FISH may be related to the parameters of the scheme. Recently, t(14;16) has come under scrutiny as a modern high-risk lesion, as patients with this lesion respond well to proteasome inhibitor based regimens, such as CyBorD, which has been a mainstay of myeloma treatment over the previous decade.[136–138,411] Furthermore, this lesion is highly concomitant with other high-risk cytogenetic lesions, and as such, has recently been inspected for its independent risk contribution, if any at all.[412,413] Consequently, the poor performance of the FISH-based risk stratification on our cohort may reflect outdatedness of the input parameters, as our cohort was nearly uniformly treated with CyBorD, or the inappropriateness of the risk parameters. Modern FISH based schemes, such as the Prognostic Index, may address this through appropriate weighting of additional cytogenetic lesions.[19] However, it is also clear that high-risk CNVs and translocations go unreported by FISH, which will continue to challenge risk classification regardless of the FISH-panel design, and may be what drives the underperformance of FISH-based stratification in our cohort.

Deletion on 1p and amplification of 1q are well well-known high-risk event which is clear from large FISH- and WGS-based myeloma assessments, and is corroborated by gene expression profiling experiments.[19,97] Importantly, these gene expression studies identified that overexpression of genes mapping across the 1p arm, and under expression of genes mapping across the 1q arm indicate high-risk.[97] This is not acknowledged in FISH-based assessments, which probe only for *TP73* on the terminal end of the q arm, and *CKS1B* on the p arm. Even within our small cohort, we found this limited assessment missed significant deletions on the q arm in three patients, and one amplification on the p arm.

5.6 Limitations

In both our targeted and whole genome sequencing studies, our cohorts were relatively small, at 76 and 9 patients, respectively. Consequently, while these are promising studies which offer strong support for clinical assessment of myeloma by NGS, additional investigations are warranted.

Our panel study, in addition to cohort size limitations, had a limited follow-up period, lacked an external testing set, and had somewhat heterogeneously treated patients. While significant differences in progression-free survival were observed in our study despite the sample size limitation, and the presence of high-risk mutations were independent of all R-ISS high-risk markers, we were unable to assess the mutation profile in a multivariate model. This is a significant shortcoming which will need to be addressed in larger cohorts to validate the clinical relevance of this panel. Additionally, the small cohort-size in conjunction with the mixed treatment profile does not allow interrogation of the performance of the panel assessment in uniformly treated subgroups. It remains unclear if the panel is more or less relevant in different treatment contexts. Furthermore, the limited sample size was insufficient to divide our data into a ‘training’ and ‘testing’ cohort, hence, external validation is required. In addition to sample-size related downfalls, the short follow-up period of our investigation limited our study to correlations with progression-free survival. The extensibility of our findings into overall survival should be assessed on a larger cohort with longer follow-up data.

Within our WGS assessment, consequent to our small sample size, not all prognostically relevant lesions were present, such as del(17p) and t(6;14). Assessment on larger cohorts comprehensive for all prognostically relevant structural variants should be performed. Beyond this limitation, our study was confounded by comparison against FISH as the truth set. While comparison against FISH is sensical for clinical purposes, FISH is not an ideal truth set for performance comparisons due to its limited scope and potential to err, particularly in cases of complex and unbalanced structural variation.[414] Addressing this, higher-depth sequencing, long-insert, long-read, or sanger sequencing could be performed to corroborate or dismiss structural variant call made at lower depths. Additionally, our subsampling was performed once. As subsample is a random process which extracts a portion of reads from the total dataset, variation in the performance of variant calling may be expected across multiple iterations. Hence, it is prudent to assess performance over multiple iterations of subsampling, especially at lower depths.

We employed short-read paired-end sequencing at 2x100bp. While long-read sequencing does outperform short-read sequencing for SV detection (especially in complex contexts), there are a number of benefits of short-read sequencing within the clinical realm compared to long-read approaches. Firstly, the cost for short-read sequencing is lower than that of long-read sequencing. Secondly, greater technical proficiency may be required for long-read sequencing as long DNA molecules are sensitive to shredding and alternative sequencing chemistries and technologies are often required. Thirdly, most clinical molecular labs already have at least modest capacity for short-read sequencing. Taken together, short-read based structural variant calling remains desirable in the clinical setting, at least until long-read sequencing technology further matures.

5.7 Future Directions

A benefit of our assessment being on a mixed population of plasma cell dyscrasias is that 10 MGUS and 3 SMM patients were assessed on the DMG26 panel. Identifying pre-clinical plasma cell dyscrasia patients at risk of progression to overt disease is a prominent area of research; as early intervention in these patients may offer better outcomes.[415–417] With the DMG26, none of the pre-clinical stage patients harboured high-risk lesions, and none progressed to overt disease during the follow-up period. It is a positive indication that the DMG26 did not categorize these pre-clinical samples as high-risk given that they did not progress; however, it remains unclear if DMG26 high-risk designation within pre-clinical contexts would carry clinical significance. Further investigation on larger MGUS and SMM cohorts is warranted. Building on this, assessment of the DMG26 on a larger cohort in general is prudent to validate the clinical significance of its high-risk designation.

For structural variant detection, synthetic approaches to long-read sequencing have been proposed, such as 10X genomics, which couple long-range information to short-read data. Briefly, this technology isolates long DNA molecules into individual oil droplets which are isolated reaction chambers. Each reaction chamber processes the DNA into short-read ready libraries and appends a barcode to the sequence unique to the oil droplet. Reads with the same barcode are then known to originate from the same large DNA molecule. While an attractive option due to its compatibility with Illumina sequencing platforms, it is a technically challenging and costly process that often requires referral to an external laboratory.

Low-coverage long-insert paired-end sequencing is another approach for structural variant detection, and has recently been applied to MM genome interrogation.[88] The authors of this work combined DELLY and LUMPY and identified numerous structural variations (median: 16 per patient) at 4X-8X.[88] They subsequently assessed for chromoplexy and chromothripsis, the latter of which was significantly associated with poor progression-free and overall survival.[88] Importantly, the authors did not compare against FISH, which is necessary to assess for added clinical value of next-generation based assessments as compared to the clinical standard.

5.8 Conclusions

Multiple myeloma is a severe and increasingly common malignancy that remains incurable.[1,4,45] Though there has been dramatic improvement in patient outcome over the previous decade with the introduction of so-called 'novel agents', namely immunomodulatory drugs and proteasome inhibitors, five-year survival remains low at 50% and response to treatment is highly variable.[1] While myeloma presents with a highly varied genome that has long been known to have clinical importance, both the capturing of genomic information and its subsequent use in the diagnosing, prognosing, and treatment of an individual have been performed sub-optimally. This largely relates to technical limitations of the current clinical standard for myeloma genome assessment, FISH, namely its limited throughput, low resolution, high cost, and targeted nature.

Previously, in a head-to-head comparison of ultra-low-depth WGS and FISH, we demonstrated the superiority of WGS in the identification of copy-number lesions.[26]
The work performed herein expands upon this early finding, with prognostication by mutation profiling outperforming FISH based risk stratification, and WGS-based structural variant calling identifying numerous prognostically relevant CNVs and translocations not capturable or cryptic to FISH. Importantly, while both R-ISS high-risk FISH markers and the ISS failed to identify a high-risk group with significantly poor outcome within our cohort, mutation profiling did. Hence, the classic view where large-scale alterations dominate a patient's risk profile and small-scale lesions only modestly decorate this profile may be incorrect. This is in line with recent investigations on *TP53* identifying a double hit profile, which may arise from SNVs and indels, defining the highest-risk myeloma group known.[95] Furthermore, other groups have found significant, and high hazards for mutation in genes that we included in our DMG26 panel including *TP53*, *PRDM1*, *DIS3*, and *CYLD*.[95,409] Using a two-pronged approach, we demonstrate that next-generation sequencing approaches are superior to FISH in resolution, allowing capture of small SNVs and indels which are indications for precision medicine and have prognostic impact, and in scope and sensitivity, allowing capture of CNVs and translocations throughout the genome, many of which are ere either untargeted by or cryptic to FISH and have important prognostic implications. Our findings underscore the value of clinical MM genome

assessment by NGS technology and highlight the need to establish thee as the new benchmark in place of FISH within international myeloma clinical stratification systems.

References

1. Kumar SK, Rajkumar V, Kyle RA, et al. Multiple myeloma. *Nat Rev Dis Primers*. 2017;3(1):17046.

2. Hsu DC, Wilkenfeld P, Joshua DE. Multiple myeloma. *Bmj*. 2012;344(jan05 1):d7953.

3. Kazandjian D. Multiple myeloma epidemiology and survival: A unique malignancy. *Semin Oncol*. 2016;43(6):676–681.

4. Cowan AJ, Allen C, Barac A, et al. Global Burden of Multiple Myeloma: A Systematic Analysis for the Global Burden of Disease Study 2016. *Jama Oncol*. 2018;4(9):1221.

5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *Ca Cancer J Clin*. 2016;66(1):7–30.

6. Manier S, Salem KZ, Park J, et al. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol*. 2017;14(2):100–113.

7. Maura F, Bolli N, Angelopoulos N, et al. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Biorxiv*. 2018;388611.

8. Pont SR du, Cleynen A, Fontan C, et al. Genomics of Multiple Myeloma. *J Clin Oncol*. 2017;35(9):JCO.2016.70.670.

9. Bolli N, Avet-Loiseau H, Wedge DC, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*. 2014;5(1):2997.

10. Carrasco DR, Tonon G, Huang Y, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer Cell*. 2006;9(4):313–325.

11. Weaver CJ, Tariman JD. Multiple Myeloma Genomics: A Systematic Review. *Seminars Oncol Nurs*. 2017;33(3):237–253.

12. Ruiz-Heredia Y, Sànchez-Vega B, Onecha E, et al. Mutational screening of newly diagnosed multiple myeloma patients by deep targeted sequencing. *Haematologica*. 2018;103(11):haematol.2018.188839.

13. 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.

14. Keats JJ, Fonseca R, Chesi M, et al. Promiscuous Mutations Activate the Noncanonical NF-κB Pathway in Multiple Myeloma. *Cancer Cell*. 2007;12(2):131–144.
15. Hoang PH, Dobbins SE, Cornish AJ, et al. Whole-genome sequencing of multiple myeloma reveals oncogenic pathways are targeted somatically through multiple mechanisms. *Leukemia*. 2018;32(11):2459–2470.
16. 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.
17. Dubrova SE, Dygin VP, Ushakova EA. [Cytogenetical and clinico-hematological changes in myeloma disease]. *Tsitologiia*. 1966;8(2):241–9.
18. Yuregir OO, Sahin FI, Yilmaz Z, et al. Fluorescent in situ hybridization studies in multiple myeloma. *Hematology*. 2013;14(2):90–94.
19. Perrot A, Lauwers-Cances V, Tournay E, et al. Development and Validation of a Cytogenetic Prognostic Index Predicting Survival in Multiple Myeloma. *J Clin Oncol*. 2019;37(19):1657–1665.
20. Rustad EH, Hultcrantz M, Yellapantula VD, et al. Baseline identification of clonal V(D)J sequences for DNA-based minimal residual disease detection in multiple myeloma. *Plos One*. 2019;14(3):e0211600.
21. Kortuem KM, Braggio E, Bruins L, et al. Panel sequencing for clinically oriented variant screening and copy number detection in 142 untreated multiple myeloma patients. *Blood Cancer J*. 2016;6(2):e397–e397.
22. Bolli N, Biancon G, Moarii M, et al. Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia*. 2018;32(12):2604–2616.
23. Bolli N, Maura F, Minvielle S, et al. Genomic patterns of progression in smoldering multiple myeloma. *Nat Commun*. 2018;9(1):3363.
24. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *J Clin Oncol*. 2015;33(26):2863–2869.
25. Dispenzieri A, Rajkumar SV, Gertz MA, et al. Treatment of Newly Diagnosed Multiple Myeloma Based on Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART): Consensus Statement. *Mayo Clin Proc*. 2007;82(3):323–341.

26. Elnenaei MO, Knopf P, Cutler SD, et al. Low-depth sequencing for copy number abnormalities in multiple myeloma supersedes fluorescent in situ hybridization in scope and resolution. *Clin Genet*. 2019;96(2):163–168.
27. Huber D, Voithenberg LV von, Kaigala GV. Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Nano Eng*. 2018;1:15–24.
28. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood*. 2008;112(5):1570–1580.
29. Pieper K, Grimbacher B, Eibel H. B-cell biology and development. *J Allergy Clin Immun*. 2013;131(4):959–971.
30. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. MECHANISM AND CONTROL OF V(D)J RECOMBINATION AT THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS. *Annu Rev Immunol*. 2006;24(1):541–570.
31. Lutz J, Heideman MR, Roth E, et al. Pro-B cells sense productive immunoglobulin heavy chain rearrangement irrespective of polypeptide production. *Proc National Acad Sci*. 2011;108(26):10644–10649.
32. Shapiro-Shelef M, Calame K. Regulation of plasma-cell development. *Nat Rev Immunol*. 2005;5(3):230–242.
33. Winkler TH, Mårtensson I-L. The Role of the Pre-B Cell Receptor in B Cell Development, Repertoire Selection, and Tolerance. *Front Immunol*. 2018;9:2423.
34. Nemazee D. Mechanisms of central tolerance for B cells. *Nat Rev Immunol*. 2017;17(5):281–294.
35. Klein B, Tarte K, Jourdan M, et al. Survival and Proliferation Factors of Normal and Malignant Plasma Cells. *Int J Hematol*. 2003;78(2):106–113.
36. Recaldin T, Fear DJ. Transcription factors regulating B cell fate in the germinal centre. *Clin Exp Immunol*. 2016;183(1):65–75.
37. 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–e548.

38. Group TIMW. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Brit J Haematol*. 2003;121(5):749–757.
39. Micco PD, Micco BD. Up-date on solitary plasmacytoma and its main differences with multiple myeloma. *Exp Oncol*. 2005;27(1):7–12.
40. Cook L, Macdonald DHC. Management of paraproteinaemia. *Postgrad Med J*. 2007;83(978):217.
41. Dupuis MM, Tuchman SA. Non-secretory multiple myeloma: from biology to clinical management. *Oncotargets Ther*. 2016;Volume 9:7583–7590.
42. Gundesen MT, Lund T, Moeller HEH, Abildgaard N. Plasma Cell Leukemia: Definition, Presentation, and Treatment. *Curr Oncol Rep*. 2019;21(1):8.
43. Ho M, Patel A, Goh CY, et al. Changing paradigms in diagnosis and treatment of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Leukemia*. 2020;34(12):3111–3125.
44. KYLE RA, GARTON JP. The Spectrum of IgM Monoclonal Gammopathy in 430 Cases. *Mayo Clin Proc*. 1987;62(8):719–731.
45. Tsang M, Le M, Ghazawi FM, et al. Multiple myeloma epidemiology and patient geographic distribution in Canada: A population study. *Cancer*. 2019;125(14):2435–2444.
46. Boyd KD, Ross FM, Chiecchio L, et al. Gender Disparities in the Tumor Genetics and Clinical Outcome of Multiple Myeloma. *Cancer Epidemiology Prev Biomarkers*. 2011;20(8):1703–1707.
47. Waxman AJ, Mink PJ, Devesa SS, et al. Racial disparities in incidence and outcome in multiple myeloma: a population-based study. *Blood*. 2010;116(25):5501–5506.
48. Wadhera RK, Rajkumar SV. Prevalence of Monoclonal Gammopathy of Undetermined Significance: A Systematic Review. *Mayo Clin Proc*. 2010;85(10):933–942.
49. Cohen HJ, Crawford J, Rao MK, Pieper CF, Currie MS. Racial Differences in the Prevalence of Monoclonal Gammopathy in a Community-based Sample of the Elderly. *Am J Medicine*. 1998;104(5):439–444.

50. Ravindran A, Bartley AC, Holton SJ, et al. Prevalence, incidence and survival of smoldering multiple myeloma in the United States. *Blood Cancer J*. 2016;6(10):e486–e486.
51. Kyle RA, Remstein ED, Therneau TM, et al. Clinical Course and Prognosis of Smoldering (Asymptomatic) Multiple Myeloma. *New Engl J Medicine*. 2007;356(25):2582–2590.
52. Kyle RA, Therneau TM, Rajkumar SV, et al. Long-term follow-up of IgM monoclonal gammopathy of undetermined significance. *Blood*. 2003;102(10):3759–3764.
53. Basak G, Srivastava A, Malhotra R, Carrier E. Multiple Myeloma Bone Marrow Niche. *Curr Pharm Biotechno*. 2009;10(3):335–346.
54. Noll JE, Williams SA, Purton LE, Zannettino ACW. Tug of war in the haematopoietic stem cell niche: do myeloma plasma cells compete for the HSC niche? *Blood Cancer J*. 2012;2(9):e91–e91.
55. Hideshima T, Mitsiades C, Tonon G, Richardson PG, Anderson KC. Understanding multiple myeloma pathogenesis in the bone marrow to identify new therapeutic targets. *Nat Rev Cancer*. 2007;7(8):585–598.
56. Alonso S, Matsui W, Jones RJ, Ghiaur G. Bone Marrow Niche - Multiple Myeloma Cross-Talk Generates Bortezomib Resistance. *Blood*. 2015;126(23):914–914.
57. An G, Qin X, Acharya C, et al. Multiple myeloma patients with low proportion of circulating plasma cells had similar survival with primary plasma cell leukemia patients. *Ann Hematol*. 2015;94(2):257–264.
58. Gunn WG, Conley A, Deininger L, et al. A Crosstalk Between Myeloma Cells and Marrow Stromal Cells Stimulates Production of DKK1 and Interleukin-6: A Potential Role in the Development of Lytic Bone Disease and Tumor Progression in Multiple Myeloma. *Stem Cells*. 2006;24(4):986–991.
59. Gámez B, Edwards CM. Contributions of the Bone Microenvironment to Monoclonal Gammopathy of Undetermined Significance Pathogenesis. *Curr Osteoporos Rep*. 2018;16(6):635–641.
60. Bianchi G, Munshi NC. Pathogenesis beyond the cancer clone(s) in multiple myeloma. *Blood*. 2015;125(20):3049–3058.

61. Malek E, Lima M de, Letterio JJ, et al. Myeloid-derived suppressor cells: The green light for myeloma immune escape. *Blood Rev*. 2016;30(5):341–348.
62. Valckenborgh EV, Schouppe E, Movahedi K, et al. Multiple myeloma induces the immunosuppressive capacity of distinct myeloid-derived suppressor cell subpopulations in the bone marrow. *Leukemia*. 2012;26(11):2424–2428.
63. Görgün GT, Whitehill G, Anderson JL, et al. Tumor-promoting immune-suppressive myeloid-derived suppressor cells in the multiple myeloma microenvironment in humans. *Blood*. 2013;121(15):2975–2987.
64. Toscani D, Bolzoni M, Accardi F, Aversa F, Giuliani N. The osteoblastic niche in the context of multiple myeloma. *Ann Ny Acad Sci*. 2015;1335(1):45–62.
65. Karadag, Oyajobi, Apperley, Russell, Croucher. Human myeloma cells promote the production of interleukin 6 by primary human osteoblasts. *Brit J Haematol*. 2000;108(2):383–390.
66. Nair B, Waheed S, Szymonifka J, et al. Immunoglobulin isotypes in multiple myeloma: laboratory correlates and prognostic implications in total therapy protocols. *Brit J Haematol*. 2009;145(1):134–137.
67. Kyle RA, Gertz MA, Witzig TE, et al. Review of 1027 Patients With Newly Diagnosed Multiple Myeloma. *Mayo Clin Proc*. 2003;78(1):21–33.
68. Jákó JM, Gesztesi T, Kaszás I. IgE Lambda Monoclonal Gammopathy and Amyloidosis. *Int Arch Allergy Imm*. 1997;112(4):415–421.
69. Rajkumar SV, Dispenzieri A, Kyle RA. Monoclonal Gammopathy of Undetermined Significance, Waldenström Macroglobulinemia, AL Amyloidosis, and Related Plasma Cell Disorders: Diagnosis and Treatment. *Mayo Clin Proc*. 2006;81(5):693–703.
70. Banaszkiewicz M, Małyszko J, Vesole DH, et al. New Biomarkers of Ferric Management in Multiple Myeloma and Kidney Disease-Associated Anemia. *J Clin Medicine*. 2019;8(11):1828.
71. Greipp PR, Miguel JS, Durie BGM, et al. International Staging System for Multiple Myeloma. *J Clin Oncol*. 2005;23(15):3412–3420.
72. Cho H, Yoon DH, Lee JB, et al. Comprehensive evaluation of the revised international staging system in multiple myeloma patients treated with novel agents as a primary therapy. *Am J Hematol*. 2017;92(12):1280–1286.

73. Rajkumar SV, Harousseau J-L, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. *Blood*. 2011;117(18):4691–4695.
74. Bird JM, Owen RG, D'Sa S, et al. Guidelines for the diagnosis and management of multiple myeloma 2011. *Brit J Haematol*. 2011;154(1):32–75.
75. Kumar SK, Callander NS, Alsina M, et al. Multiple Myeloma, Version 3.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Ne*. 2017;15(2):230–269.
76. Hattori Y, Kakimoto T, Okamoto S, Sato N, Ikeda Y. Thalidomide-induced severe neutropenia during treatment of multiple myeloma. *Int J Hematol*. 2004;79(3):283–288.
77. Qiu L, Zhu F, Wei G, et al. Idiopathic thrombocytopenic purpura treatment in a relapsed/refractory multiple myeloma patient after chimeric antigen receptor T cell therapy. *Regen Ther*. 2020;14:271–274.
78. Sridevi HB, Rai S, Suresh PK, Somesh MS, Minal J. Pancytopenia in Multiple Myeloma- An Enigma: Our Experience from Tertiary Care Hospital. *J Clin Diagnostic Res*. 2015;9(11):EC04-6.
79. O'Connell TX, Horita TJ, Kasravi B. Understanding and interpreting serum protein electrophoresis. *Am Fam Physician*. 2005;71(1):105–12.
80. Lee AY, Cassar PM, Johnston AM, Adelstein S. Clinical use and interpretation of serum protein electrophoresis and adjunct assays. *Brit J Hosp Med*. 2017;78(2):C18–C20.
81. Tripathy S. The Role of Serum Protein Electrophoresis in the Detection of Multiple Myeloma: An Experience of a Corporate Hospital. *J Clin Diagnostic Res*. 2012;6(9):1458–61.
82. Lee N, Moon SY, Lee J, et al. Discrepancies between the percentage of plasma cells in bone marrow aspiration and BM biopsy: Impact on the revised IMWG diagnostic criteria of multiple myeloma. *Blood Cancer J*. 2017;7(2):e530–e530.
83. Štifter S, Babarović E, Valković T, et al. Combined evaluation of bone marrow aspirate and biopsy is superior in the prognosis of multiple myeloma. *Diagn Pathol*. 2010;5(1):30.

84. Rajkumar SV, Fonseca R, Dispenzieri A, et al. Methods for estimation of bone marrow plasma cell involvement in myeloma: Predictive value for response and survival in patients undergoing autologous stem cell transplantation. *Am J Hematol*. 2001;68(4):269–275.
85. Paiva B, Puig N, Cedena M-T, et al. Measurable Residual Disease by Next-Generation Flow Cytometry in Multiple Myeloma. *J Clin Oncol*. 2020;38(8):784–792.
86. Egan JB, Shi C-X, 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.
87. Prideaux SM, O'Brien EC, Chevassut TJ. The Genetic Architecture of Multiple Myeloma. *Adv Hematology*. 2014;2014:1–16.
88. Rustad EH, Yellapantula VD, Glodzik D, et al. Revealing the impact of recurrent and rare structural variants in multiple myeloma. *Biorxiv*. 2019;2019.12.18.881086.
89. Melchor L, Walker BA, Wardell CP, et al. Intra-Clonal Heterogeneity Is a Critical Early Event in the Preclinical Stages of Multiple Myeloma and Is Subject to Darwinian Fluctuation throughout the Disease. *Blood*. 2012;120(21):3941–3941.
90. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471(7339):467–472.
91. Laï JL, Michaux L, Dastugue N, et al. Cytogenetics in Multiple Myeloma A Multicenter Study of 24 Patients with t(11;14)(q13;q32) or Its Variant. *Cancer Genet Cytogen*. 1998;104(2):133–138.
92. Maura F, Bolli N, Angelopoulos N, et al. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nat Commun*. 2019;10(1):3835.
93. Croft J, Ellis S, Sherborne AL, et al. Copy number evolution and its relationship with patient outcome—an analysis of 178 matched presentation-relapse tumor pairs from the Myeloma XI trial. *Leukemia*. 2020;1–11.
94. Walker BA, Mavrommatis K, Wardell CP, et al. Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*. 2018;132(6):587–597.

95. Walker BA, Mavrommatis K, Wardell CP, et al. A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*. 2019;33(1):159–170.
96. Samur MK, Samur AA, Fulciniti M, et al. Genome-Wide Somatic Alterations in Multiple Myeloma Reveal a Superior Outcome Group. *J Clin Oncol*. 2020;38(27):3107–3118.
97. Shaughnessy JD, 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.
98. Hose D, Rème T, Hielscher T, et al. Proliferation is a central independent prognostic factor and target for personalized and risk-adapted treatment in multiple myeloma. *Haematologica*. 2011;96(1):87–95.
99. 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.
100. Kuiper R, Zweegman S, Duin M van, et al. Prognostic and predictive performance of R-ISS with SKY92 in older patients with multiple myeloma: the HOVON-87/NMSG-18 trial. *Blood Adv*. 2020;4(24):6298–6309.
101. Beers EH van, Vliet MH van, Kuiper R, et al. Prognostic Validation of SKY92 and Its Combination With ISS in an Independent Cohort of Patients With Multiple Myeloma. *Clin Lymphoma Myeloma Leukemia*. 2017;17(9):555–562.
102. Laganà A, Beno I, Melnekoff D, et al. Precision Medicine for Relapsed Multiple Myeloma on the Basis of an Integrative Multiomics Approach. *Jco Precis Oncol*. 2018;2018(2):1–17.
103. Harding T, Baughn L, Kumar S, Ness BV. The future of myeloma precision medicine: integrating the compendium of known drug resistance mechanisms with emerging tumor profiling technologies. *Leukemia*. 2019;33(4):863–883.
104. Zandecki M, Laï JL, Facon T. MULTIPLE MYELOMA: ALMOST ALL PATIENTS ARE CYTOGENETICALLY ABNORMAL. *Brit J Haematol*. 1996;94(2):217–227.

105. Liebisch P, Wendl C, Wellmann A, et al. High incidence of trisomies 1q, 9q, and 11q in multiple myeloma: results from a comprehensive molecular cytogenetic analysis. *Leukemia*. 2003;17(12):2535–2537.
106. Kalff A, Spencer A. The t(4;14) translocation and FGFR3 overexpression in multiple myeloma: prognostic implications and current clinical strategies. *Blood Cancer J*. 2012;2(9):e89–e89.
107. Tian E, Sawyer JR, Heuck CJ, et al. In multiple myeloma, 14q32 translocations are nonrandom chromosomal fusions driving high expression levels of the respective partner genes. *Genes Chromosomes Cancer*. 2014;53(7):549–557.
108. Wier SV, Braggio E, Baker A, et al. Hypodiploid multiple myeloma is characterized by more aggressive molecular markers than non-hyperdiploid multiple myeloma. *Haematologica*. 2013;98(10):1586–1592.
109. Fonseca R, Barlogie B, Bataille R, et al. Genetics and Cytogenetics of Multiple Myeloma A Workshop Report. *Cancer Res*. 2004;64(4):1546–1558.
110. Sidana S, Jevremovic D, Ketterling RP, et al. Tetraploidy is associated with poor prognosis at diagnosis in multiple myeloma. *Am J Hematol*. 2019;94(5):E117–E120.
111. Magrangeas F, Avet-Loiseau H, Munshi NC, Minvielle S. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. *Blood*. 2011;118(3):675–678.
112. Neuse CJ, Lomas OC, Schliemann C, et al. Genome instability in multiple myeloma. *Leukemia*. 2020;34(11):2887–2897.
113. Rajan AM, Rajkumar SV. Interpretation of cytogenetic results in multiple myeloma for clinical practice. *Blood Cancer J*. 2015;5(10):e365–e365.
114. Binder M, Rajkumar SV, Ketterling RP, et al. Prognostic implications of abnormalities of chromosome 13 and the presence of multiple cytogenetic high-risk abnormalities in newly diagnosed multiple myeloma. *Blood Cancer J*. 2017;7(9):e600–e600.
115. Chretien M-L, Corre J, Lauwers-Cances V, et al. Understanding the role of hyperdiploidy in myeloma prognosis: which trisomies really matter? *Blood*. 2015;126(25):2713–2719.

116. Fonseca R, Debes-Marun CS, Picken EB, et al. The recurrent IgH translocations are highly associated with nonhyperdiploid variant multiple myeloma. *Blood*. 2003;102(7):2562–2567.
117. Walker BA, Wardell CP, Johnson DC, et al. Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood*. 2013;121(17):3413–3419.
118. Cardona-Benavides IJ, Ramón C de, Gutiérrez NC. Genetic Abnormalities in Multiple Myeloma: Prognostic and Therapeutic Implications. *Cells*. 2021;10(2):336.
119. Santra M, Zhan F, Tian E, Barlogie B, Shaughnessy J. A subset of multiple myeloma harboring the t(4;14)(p16;q32) translocation lacks FGFR3 expression but maintains anIGH/MMSET fusion transcript. *Blood*. 2003;101(6):2374–2376.
120. Mirabella F, Wu P, Wardell CP, et al. MMSET is the key molecular target in t(4;14) myeloma. *Blood Cancer J*. 2013;3(5):e114–e114.
121. Bergsagel PL, Kuehl WM, Zhan F, et al. Cyclin D dysregulation: an early and unifying pathogenic event in multiple myeloma. *Blood*. 2005;106(1):296–303.
122. Greenberg AJ, Rajkumar SV, Therneau TM, et al. Relationship between initial clinical presentation and the molecular cytogenetic classification of myeloma. *Leukemia*. 2014;28(2):398–403.
123. Avet-Loiseau H, Leleu X, Roussel M, et al. Bortezomib Plus Dexamethasone Induction Improves Outcome of Patients With t(4;14) Myeloma but Not Outcome of Patients With del(17p). *J Clin Oncol*. 2010;28(30):4630–4634.
124. Cavo M, Tacchetti P, Patriarca F, et al. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. *Lancet*. 2010;376(9758):2075–2085.
125. Tacchetti P, Pantani L, Patriarca F, et al. Bortezomib, thalidomide, and dexamethasone followed by double autologous haematopoietic stem-cell transplantation for newly diagnosed multiple myeloma (GIMEMA-MMY-3006): long-term follow-up analysis of a randomised phase 3, open-label study. *Lancet Haematol*. 2020;7(12):e861–e873.

126. Avet-Loiseau H, Fonseca R, Siegel D, et al. Carfilzomib significantly improves the progression-free survival of high-risk patients in multiple myeloma. *Blood*. 2016;128(9):1174–1180.
127. Scheid C, Reece D, Beksac M, et al. Phase 2 study of dovitinib in patients with relapsed or refractory multiple myeloma with or without t(4;14) translocation. *Eur J Haematol*. 2015;95(4):316–324.
128. Kaufman GP, Gertz MA, Dispenzieri A, et al. Impact of cytogenetic classification on outcomes following early high-dose therapy in multiple myeloma. *Leukemia*. 2016;30(3):633–639.
129. Paner A, Patel P, Dhakal B. The evolving role of translocation t(11;14) in the biology, prognosis, and management of multiple myeloma. *Blood Rev*. 2019;41:100643.
130. Gran C, Uttervall K, Bruchfeld JB, et al. Translocation (11;14) in newly diagnosed multiple myeloma, time to reclassify this standard risk chromosomal aberration? *Eur J Haematol*. 2019;103(6):588–596.
131. Ghobrial I. BELLINI: a renaissance for an era of precision therapy in multiple myeloma. *Lancet Oncol*. 2020;21(12):1547–1549.
132. Kumar SK, Harrison SJ, Cavo M, et al. Venetoclax or placebo in combination with bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma (BELLINI): a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol*. 2020;21(12):1630–1642.
133. 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.
134. Siu KT, Huang C, Panaroni C, et al. BCL2 blockade overcomes MCL1 resistance in multiple myeloma. *Leukemia*. 2019;33(8):2098–2102.
135. Tron AE, Belmonte MA, Adam A, et al. Discovery of Mcl-1-specific inhibitor AZD5991 and preclinical activity in multiple myeloma and acute myeloid leukemia. *Nat Commun*. 2018;9(1):5341.
136. Abdallah N, Rajkumar SV, Greipp P, et al. Cytogenetic abnormalities in multiple myeloma: association with disease characteristics and treatment response. *Blood Cancer J*. 2020;10(8):82.

137. Qiang Y-W, Ye S, Huang Y, et al. MAFb protein confers intrinsic resistance to proteasome inhibitors in multiple myeloma. *Bmc Cancer*. 2018;18(1):724.
138. Qiang Y-W, Ye S, Chen Y, et al. MAF protein mediates innate resistance to proteasome inhibition therapy in multiple myeloma. *Blood*. 2016;128(25):2919–2930.
139. Annunziata CM, Hernandez L, Davis RE, et al. A mechanistic rationale for MEK inhibitor therapy in myeloma based on blockade of MAF oncogene expression. *Blood*. 2011;117(8):2396–2404.
140. Herath NI, Rocques N, Garancher A, Eychène A, Pouponnot C. GSK3-mediated MAF phosphorylation in multiple myeloma as a potential therapeutic target. *Blood Cancer J*. 2014;4(1):e175–e175.
141. Lodé L, Eveillard M, Trichet V, et al. Mutations in TP53 are exclusively associated with del(17p) in multiple myeloma. *Haematologica*. 2010;95(11):1973–1976.
142. Lakshman A, Painuly U, Rajkumar SV, et al. Natural history of multiple myeloma with de novo del(17p). *Blood Cancer J*. 2019;9(3):32.
143. Boyd KD, Ross FM, Walker BA, et al. Mapping of Chromosome 1p Deletions in Myeloma Identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as Being Genes in Regions Associated with Adverse Survival. *Clin Cancer Res*. 2011;17(24):7776–7784.
144. Smetana J, Frohlich J, Zaoralova R, et al. Genome-Wide Screening of Cytogenetic Abnormalities in Multiple Myeloma Patients Using Array-CGH Technique: A Czech Multicenter Experience. *Biomed Res Int*. 2014;2014:1–9.
145. Hanamura I. Gain/Amplification of Chromosome Arm 1q21 in Multiple Myeloma. *Cancers*. 2021;13(2):256.
146. Schmidt TM, Barwick BG, Joseph N, et al. Gain of Chromosome 1q is associated with early progression in multiple myeloma patients treated with lenalidomide, bortezomib, and dexamethasone. *Blood Cancer J*. 2019;9(12):94.
147. Li F, Xu Y, Deng S, et al. MicroRNA-15a/16-1 cluster located at chromosome 13q14 is down-regulated but displays different expression pattern and prognostic significance in multiple myeloma. *Oncotarget*. 2015;6(35):38270–38282.
148. Roccaro AM, Sacco A, Thompson B, et al. MicroRNAs 15a and 16 regulate tumor proliferation in multiple myeloma. *Blood*. 2009;113(26):6669–6680.

149. Saxe D, Seo E, Bergeron MB, Han J. Recent advances in cytogenetic characterization of multiple myeloma. *Int J Lab Hematol*. 2019;41(1):5–14.
150. ZHANG W, LIU HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res*. 2002;12(1):9–18.
151. Zhu J, Blenis J, Yuan J. Activation of PI3K/Akt and MAPK pathways regulates Myc-mediated transcription by phosphorylating and promoting the degradation of Mad1. *Proc National Acad Sci*. 2008;105(18):6584–6589.
152. Macek P, Cliff MJ, Embrey KJ, et al. Myc phosphorylation in its basic helix–loop–helix region destabilizes transient α-helical structures, disrupting Max and DNA binding. *J Biol Chem*. 2018;293(24):9301–9310.
153. Calado DP, Sasaki Y, Godinho SA, et al. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat Immunol*. 2012;13(11):1092–1100.
154. Chng W-J, Huang GF, Chung TH, et al. Clinical and biological implications of MYC activation: a common difference between MGUS and newly diagnosed multiple myeloma. *Leukemia*. 2011;25(6):1026–1035.
155. Szabo AG, Gang AO, Pedersen MØ, et al. Overexpression of c-myc is associated with adverse clinical features and worse overall survival in multiple myeloma. *Leukemia Lymphoma*. 2016;57(11):1–9.
156. Abdallah N, Baughn LB, Rajkumar SV, et al. Implications of MYC Rearrangements in Newly Diagnosed Multiple Myeloma. *Clin Cancer Res*. 2020;26(24):6581–6588.
157. Jovanović KK, Roche-Lestienne C, Ghobrial IM, et al. Targeting MYC in multiple myeloma. *Leukemia*. 2018;32(6):1295–1306.
158. Kuehl WM, Bergsagel PL. MYC addiction: a potential therapeutic target in MM. *Blood*. 2012;120(12):2351–2352.
159. Network MC, Misund K, Keane N, et al. MYC dysregulation in the progression of multiple myeloma. *Leukemia*. 2020;34(1):322–326.
160. Walker BA, Wardell CP, Brioli A, et al. Translocations at 8q24 juxtapose MYC with genes that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients. *Blood Cancer J*. 2014;4(3):e191–e191.

161. Seliger S, Geirhos V, Haferlach T, et al. Comprehensive Analysis of MYC Translocations in Multiple Myeloma By Whole Genome Sequencing and Whole Transcriptome Sequencing. *Blood*. 2019;134(Supplement_1):1774–1774.
162. Zhan F, Huang Y, Colla S, et al. The molecular classification of multiple myeloma. *Blood*. 2006;108(6):2020–2028.
163. Bergsagel PL, Kuehl WM. Molecular Pathogenesis and a Consequent Classification of Multiple Myeloma. *J Clin Oncol*. 2005;23(26):6333–6338.
164. Zhou Y, Barlogie B, Shaughnessy JD. The molecular characterization and clinical management of multiple myeloma in the post-genome era. *Leukemia*. 2009;23(11):1941–1956.
165. Broyl A, Hose D, Lokhorst H, et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood*. 2010;116(14):2543–2553.
166. Decaux O, Lodé 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 Myélome. *J Clin Oncol*. 2008;26(29):4798–4805.
167. Kuiper R, Broyl A, Knegt Y de, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia*. 2012;26(11):2406–2413.
168. Treon SP, Hunter ZR. A new era for Waldenstrom macroglobulinemia: MYD88 L265P. *Blood*. 2013;121(22):4434–4436.
169. Furukawa Y, Kikuchi J. Molecular basis of clonal evolution in multiple myeloma. *Int J Hematol*. 2020;111(4):496–511.
170. Barwick BG, Neri P, Bahlis NJ, et al. Multiple myeloma immunoglobulin lambda translocations portend poor prognosis. *Nat Commun*. 2019;10(1):1911.
171. Annunziata CM, Davis RE, Demchenko Y, et al. Frequent Engagement of the Classical and Alternative NF-κB Pathways by Diverse Genetic Abnormalities in Multiple Myeloma. *Cancer Cell*. 2007;12(2):115–130.
172. Barwick BG, Gupta VA, Vertino PM, Boise LH. Cell of Origin and Genetic Alterations in the Pathogenesis of Multiple Myeloma. *Front Immunol*. 2019;10:1121.

173. Xu J, Pfarr N, Endris V, et al. Molecular signaling in multiple myeloma: association of RAS/RAF mutations and MEK/ERK pathway activation. *Oncogenesis*. 2017;6(5):e337–e337.
174. Liu P, Leong T, Quam L, et al. Activating mutations of N- and K-ras in multiple myeloma show different clinical associations: analysis of the Eastern Cooperative Oncology Group Phase III Trial. *Blood*. 1996;88(7):2699–2706.
175. Linden MA, Kirchhof N, Carlson CS, Ness BGV. Targeted overexpression of an activated N-ras gene results in B-cell and plasma cell lymphoproliferation and cooperates with c-myc to induce fatal B-cell neoplasia. *Exp Hematol*. 2012;40(3):216–227.
176. Rasmussen T, Kuehl M, Lodahl M, Johnsen HE, Dahl IMS. Possible roles for activating RAS mutations in the MGUS to MM transition and in the intramedullary to extramedullary transition in some plasma cell tumors. *Blood*. 2005;105(1):317–323.
177. Lentzsch S, Chatterjee M, Gries M, et al. PI3-K/AKT/FKHR and MAPK signaling cascades are redundantly stimulated by a variety of cytokines and contribute independently to proliferation and survival of multiple myeloma cells. *Leukemia*. 2004;18(11):1883–1890.
178. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. *J Cell Sci*. 2016;129(7):1287–1292.
179. Lin Y-HT, Way GP, Barwick BG, et al. Integrated phosphoproteomics and transcriptional classifiers reveal hidden RAS signaling dynamics in multiple myeloma. *Blood Adv*. 2019;3(21):3214–3227.
180. Chng WJ, Gonzalez-Paz N, Price-Troska T, et al. Clinical and biological significance of RAS mutations in multiple myeloma. *Leukemia*. 2008;22(12):2280–2284.
181. Neri A, Murphy JP, Cro L, et al. Ras oncogene mutation in multiple myeloma. *J Exp Medicine*. 1989;170(5):1715–1725.
182. Lind J, Czernilofsky F, Vallet S, Podar K. Emerging protein kinase inhibitors for the treatment of multiple myeloma. *Expert Opin Emerg Dr*. 2019;24(3):1–20.
183. Shirazi F, Jones RJ, Singh RK, et al. Activating KRAS, NRAS, and BRAF mutants enhance proteasome capacity and reduce endoplasmic reticulum stress in multiple myeloma. *Proc National Acad Sci*. 2020;117(33):20004–20014.

184. Vitiello PP, Cardone C, Martini G, et al. Receptor tyrosine kinase-dependent PI3K activation is an escape mechanism to vertical suppression of the EGFR/RAS/MAPK pathway in KRAS-mutated human colorectal cancer cell lines. *J Exp Clin Canc Res*. 2019;38(1):41.
185. Leich E, Weißbach S, Klein H-U, et al. Multiple myeloma is affected by multiple and heterogeneous somatic mutations in adhesion- and receptor tyrosine kinase signaling molecules. *Blood Cancer J*. 2013;3(2):e102–e102.
186. Brewer MR, Yun C-H, Lai D, et al. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proc National Acad Sci*. 2013;110(38):E3595–E3604.
187. Yun C-H, Boggon TJ, Li Y, et al. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell*. 2007;11(3):217–227.
188. Ray P, Tan YS, Somnay V, et al. Differential protein stability of EGFR mutants determines responsiveness to tyrosine kinase inhibitors. *Oncotarget*. 2016;7(42):68597–68613.
189. Tomas A, Futter CE, Eden ER. EGF receptor trafficking: consequences for signaling and cancer. *Trends Cell Biol*. 2014;24(1):26–34.
190. Wirth M, Stojanovic N, Christian J, et al. MYC and EGR1 synergize to trigger tumor cell death by controlling NOXA and BIM transcription upon treatment with the proteasome inhibitor bortezomib. *Nucleic Acids Res*. 2014;42(16):10433–10447.
191. Shan J, Balasubramanian MN, Donelan W, et al. A Mitogen-activated Protein Kinase/Extracellular Signal-regulated Kinase Kinase (MEK)-dependent Transcriptional Program Controls Activation of the Early Growth Response 1 (EGR1) Gene during Amino Acid Limitation*. *J Biol Chem*. 2014;289(35):24665–24679.
192. Chen L, Wang S, Zhou Y, et al. Identification of early growth response protein 1 (EGR-1) as a novel target for JUN-induced apoptosis in multiple myeloma. *Blood*. 2010;115(1):61–70.
193. Consortium TAPG. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov*. 2017;7(8):818–831.

194. Benard B, Christofferson A, Legendre C, et al. FGFR3 Mutations Are an Adverse Prognostic Factor in Patients with t(4;14)(p16;q32) Multiple Myeloma: An Mmrf Commpass Analysis. *Blood*. 2017;130(Supplement 1):3027–3027.
195. Grandori C, Cowley SM, James LP, Eisenman RN. THE MYC/MAX/MAD NETWORK AND THE TRANSCRIPTIONAL CONTROL OF CELL BEHAVIOR. *Annu Rev Cell Dev Bi*. 2000;16(1):653–699.
196. Wang D, Hashimoto H, Zhang X, et al. MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res*. 2017;45(5):2396–2407.
197. Glitza IC, Lu G, Shah R, et al. Chromosome 8q24.1/c-MYC abnormality: a marker for high-risk myeloma. *Leukemia Lymphoma*. 2014;56(3):602–607.
198. Keats JJ, Fonseca R, Chesi M, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. *Cancer cell*. 2007;12(2):131–144.
199. Sun S-C. The non-canonical NF-κB pathway in immunity and inflammation. *Nat Rev Immunol*. 2017;17(9):545–558.
200. Roy P, Mukherjee T, Chatterjee B, et al. Non-canonical NFκB mutations reinforce pro-survival TNF response in multiple myeloma through an autoregulatory RelB:p50 NFκB pathway. *Oncogene*. 2017;36(10):1417–1429.
201. Roy P, Sarkar UA, Basak S. The NF-κB Activating Pathways in Multiple Myeloma. *Biomed*. 2018;6(2):59.
202. Rossi A, Voigtlaender M, Janjetovic S, et al. Mutational landscape reflects the biological continuum of plasma cell dyscrasias. *Blood Cancer J*. 2017;7(2):e537–e537.
203. Zhu S, Jin J, Gokhale S, et al. Genetic Alterations of TRAF Proteins in Human Cancers. *Front Immunol*. 2018;9:2111.
204. Ji Z, He L, Regev A, Struhl K. Regulatory network controlling tumor-promoting inflammation in human cancers. *Biorxiv*. 2018;352062.
205. Harmer D, Falank C, Reagan MR. Interleukin-6 Interweaves the Bone Marrow Microenvironment, Bone Loss, and Multiple Myeloma. *Front Endocrinol*. 2019;9:788.
206. Tiedemann RE, Gonzalez-Paz N, Kyle RA, et al. Genetic aberrations and survival in plasma cell leukemia. *Leukemia*. 2008;22(5):1044–1052.

207. Cifola I, Lionetti M, Pinatel E, et al. Whole-exome sequencing of primary plasma cell leukemia discloses heterogeneous mutational patterns. *Oncotarget*. 2015;6(19):17543–17558.
208. Flynt E, Bisht K, Sridharan V, et al. Prognosis, Biology, and Targeting of TP53 Dysregulation in Multiple Myeloma. *Cells*. 2020;9(2):287.
209. Lionetti M, Barbieri M, Manzoni M, et al. Molecular spectrum of TP53 mutations in plasma cell dyscrasias by next generation sequencing: an Italian cohort study and overview of the literature. *Oncotarget*. 2016;7(16):21353–21361.
210. Hafner A, Bulyk ML, Jambhekar A, Lahav G. The multiple mechanisms that regulate p53 activity and cell fate. *Nat Rev Mol Cell Bio*. 2019;20(4):199–210.
211. Benson EK, Mungamuri SK, Attie O, et al. p53-dependent gene repression through p21 is mediated by recruitment of E2F4 repression complexes. *Oncogene*. 2014;33(30):3959–3969.
212. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. *Nature*. 1997;387(6630):296–299.
213. Childs BG, Baker DJ, Kirkland JL, Campisi J, Deursen JM. Senescence and apoptosis: dueling or complementary cell fates? *Embo Rep*. 2014;15(11):1139–1153.
214. Bargonetti J, Prives C. Gain-of-function mutant p53: history and speculation. *J Mol Cell Biol*. 2019;11(7):605–609.
215. Austen B, Barone G, Reiman A, et al. Pathogenic ATM mutations occur rarely in a subset of multiple myeloma patients. *Brit J Haematol*. 2008;142(6):925–933.
216. Zandecki M, Facon T, Preudhomme C, et al. The Retinoblastoma Gene (RB-1) Status in Multiple Myeloma: A Report on 35 Cases. *Leukemia Lymphoma*. 2009;18(5–6):497–503.
217. Chavan SS, He J, Tytarenko R, et al. Bi-allelic inactivation is more prevalent at relapse in multiple myeloma, identifying RB1 as an independent prognostic marker. *Blood Cancer J*. 2017;7(2):e535–e535.
218. Giacinti C, Giordano A. RB and cell cycle progression. *Oncogene*. 2006;25(38):5220–5227.
219. Chinnam M, Goodrich DW. Chapter 5 RB1, Development, and Cancer. *Curr Top Dev Biol*. 2011;94:129–169.

220. Yu CY, Xiang S, Huang Z, et al. Gene Co-expression Network and Copy Number Variation Analyses Identify Transcription Factors Associated With Multiple Myeloma Progression. *Frontiers Genetics*. 2019;10:468.
221. Occhi G, Regazzo D, Trivellin G, et al. A Novel Mutation in the Upstream Open Reading Frame of the CDKN1B Gene Causes a MEN4 Phenotype. *Plos Genet*. 2013;9(3):e1003350.
222. Chasseloup F, Pankratz N, Lane J, et al. Germline CDKN1B loss-of-function variants cause pediatric Cushing's disease with or without an MEN4 phenotype. *J Clin Endocrinol Metabolism*. 2020;105(6):1983–2005.
223. Muscarella P, Bloomston M, Brewer AR, et al. Expression of the p16INK4A/Cdkn2a gene is prevalently downregulated in human pheochromocytoma tumor specimens. *Gene expression*. 2008;14(4):207–216.
224. Tomecki R, Drazkowska K, Kucinski I, et al. Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res*. 2014;42(2):1270–1290.
225. Kazazian K, Haffani Y, Ng D, et al. FAM46C/TENT5C functions as a tumor suppressor through inhibition of Plk4 activity. *Commun Biology*. 2020;3(1):448.
226. Manfrini N, Mancino M, Miluzio A, et al. FAM46C and FNDC3A Are Multiple Myeloma Tumor Suppressors That Act in Concert to Impair Clearing of Protein Aggregates and Autophagy. *Cancer Res*. 2020;80(21):4693–4706.
227. Mroczek S, Chlebowska J, Kuliński TM, et al. The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat Commun*. 2017;8(1):619.
228. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495–501.
229. Barbieri M, Manzoni M, Fabris S, et al. Compendium of FAM46C gene mutations in plasma cell dyscrasias. *Brit J Haematol*. 2016;174(4):642–645.
230. Lionetti M, Barbieri M, Todoerti K, et al. A compendium of DIS3 mutations and associated transcriptional signatures in plasma cell dyscrasias. *Oncotarget*. 2015;6(28):26129–26141.

231. Pertesi M, Vallée M, Wei X, et al. Exome sequencing identifies germline variants in DIS3 in familial multiple myeloma. *Leukemia*. 2019;33(9):2324–2330.
232. Weißbach S, Langer C, Puppe B, et al. The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Brit J Haematol*. 2015;169(1):57–70.
233. Hagman J. Critical Functions of IRF4 in B and T Lymphocytes. *J Immunol*. 2017;199(11):3715–3716.
234. Shaffer AL, Emre NCT, Lamy L, et al. IRF4 addiction in multiple myeloma. *Nature*. 2008;454(7201):226–231.
235. Cherian MA, Olson S, Sundaramoorthi H, et al. An activating mutation of interferon regulatory factor 4 (IRF4) in adult T-cell leukemia. *J Biol Chem*. 2018;293(18):6844–6858.
236. 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.
237. Bai H, Wu S, Wang R, Xu J, Chen L. Bone marrow IRF4 level in multiple myeloma: an indicator of peripheral blood Th17 and disease. *Oncotarget*. 2014;5(0):85392–85400.
238. Arora L, Kumar AP, Arfuso F, Chng WJ, Sethi G. The Role of Signal Transducer and Activator of Transcription 3 (STAT3) and Its Targeted Inhibition in Hematological Malignancies. *Cancers*. 2018;10(9):327.
239. Johnson DE, O'Keefe RA, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat Rev Clin Oncol*. 2018;15(4):234–248.
240. Chong PSY, Chng W-J, Mel S de. STAT3: A Promising Therapeutic Target in Multiple Myeloma. *Cancers*. 2019;11(5):731.
241. Jung S-H, Ahn S-Y, Choi H-W, et al. STAT3 expression is associated with poor survival in non-elderly adult patients with newly diagnosed multiple myeloma. *Blood Res*. 2017;52(4):293–299.
242. Zhu YX, Shi C-X, Bruins LA, et al. Identification of lenalidomide resistance pathways in myeloma and targeted resensitization using cereblon replacement, inhibition of STAT3 or targeting of IRF4. *Blood Cancer J*. 2019;9(2):19.

243. Bloch DB, Monte SM de la, Guigaouri P, Filippov A, Bloch KD. Identification and Characterization of a Leukocyte-specific Component of the Nuclear Body*. *J Biol Chem*. 1996;271(46):29198–29204.
244. Witjes L, Troys MV, Verhasselt B, Ampe C. Prevalence of Cytoplasmic Actin Mutations in Diffuse Large B-Cell Lymphoma and Multiple Myeloma: A Functional Assessment Based on Actin Three-Dimensional Structures. *Int J Mol Sci*. 2020;21(9):3093.
245. Radtke D, Bannard O. Expression of the Plasma Cell Transcriptional Regulator Blimp-1 by Dark Zone Germinal Center B Cells During Periods of Proliferation. *Front Immunol*. 2019;9:3106.
246. Romero-García R, Gómez-Jaramillo L, Mateos RM, et al. Differential epigenetic regulation between the alternative promoters, PRDM1α and PRDM1β, of the tumour suppressor gene PRDM1 in human multiple myeloma cells. *Sci Rep-uk*. 2020;10(1):15899.
247. Vrzalikova K, Leonard S, Fan Y, et al. Hypomethylation and Over-Expression of the Beta Isoform of BLIMP1 is Induced by Epstein-Barr Virus Infection of B Cells; Potential Implications for the Pathogenesis of EBV-Associated Lymphomas. *Pathogens*. 2012;1(2):83–101.
248. KYLE RA. Multiple Myeloma: How Did It Begin? *Mayo Clin Proc*. 1994;69(7):680–683.
249. Kyle RA, Rajkumar SV. Multiple myeloma. *Blood*. 2008;111(6):2962–2972.
250. AUSTIN C, BERGSAGEL DE, SPRAGUE CC. Evaluation of new chemotherapeutic agents in the treatment of multiple myeloma. VI. Metasarcolysin. *Cancer Chemoth Rep*. 1962;21:107–12.
251. MASS RE. A comparison of the effect of prednisone and a placebo in the treatment of multiple myeloma. *Cancer Chemoth Rep*. 1962;16:257–9.
252. Alexanian R, Haut A, Khan AU, et al. Treatment for Multiple Myeloma: Combination Chemotherapy With Different Melphalan Dose Regimens. *Jama*. 1969;208(9):1680–1685.

253. Hari P, Reece DE, Randhawa J, et al. Final outcomes of escalated melphalan 280 mg/m2 with amifostine cytoprotection followed autologous hematopoietic stem cell transplantation for multiple myeloma: high CR and VGPR rates do not translate into improved survival. *Bone Marrow Transpl*. 2019;54(2):293–299.
254. Mcelwain TJ, Powles RL. HIGH-DOSE INTRAVENOUS MELPHALAN FOR PLASMA-CELL LEUKAEMIA AND MYELOMA. *Lancet*. 1983;322(8354):822–824.
255. Yang J, Terebelo HR, Zonder JA. Secondary Primary Malignancies in Multiple Myeloma: An Old Nemesis Revisited. *Adv Hematology*. 2012;2012:801495.
256. Auner HW, Iacobelli S, Sbianchi G, et al. Melphalan 140mg/m2 or 200mg/m2 for autologous transplantation in myeloma: results from the Collaboration to Collect Autologous Transplant Outcomes in Lymphoma and Myeloma (CALM) study. A report by the EBMT Chronic Malignancies Working Party. *Haematologica*. 2017;103(3):haematol.2017.181339.
257. Qazilbash MH, Bashir Q, Thall P, et al. A Randomized Phase III Trial Of Busulfan + Melphalan (Bu-Mel) Vs Melphalan Alone For Multiple Myeloma: Longer PFS In The Bu-Mel Arm. *Clin Lymphoma Myeloma Leukemia*. 2015;15:e72–e73.
258. Palumbo A, Mina R, Cerrato C, Cavallo F. Role of Consolidation/Maintenance Therapy in Multiple Myeloma. *Clin Lymphoma Myeloma Leukemia*. 2013;13:S349–S354.
259. Nathwani N, Larsen JT, Kapoor P. Consolidation and Maintenance Therapies for Newly Diagnosed Multiple Myeloma in the Era of Novel Agents. *Curr Hematol Malig R*. 2016;11(2):127–136.
260. Mai EK, Bertsch U, Dürig J, et al. Phase III trial of bortezomib, cyclophosphamide and dexamethasone (VCD) versus bortezomib, doxorubicin and dexamethasone (PAd) in newly diagnosed myeloma. *Leukemia*. 2015;29(8):1721–1729.
261. He J, Schmerold L, Rampelbergh RV, et al. Treatment Pattern and Outcomes in Newly Diagnosed Multiple Myeloma Patients Who Did Not Receive Autologous Stem Cell Transplantation: A Real-World Observational Study. *Adv Ther*. 2021;38(1):640–659.
262. Areethamsirikul N, Masih-Khan E, Chu C-M, et al. CyBorD induction therapy in clinical practice. *Bone Marrow Transpl*. 2015;50(3):375–379.

263. Niesvizky R, Flinn IW, Rifkin R, et al. Community-Based Phase IIIB Trial of Three UPFRONT Bortezomib-Based Myeloma Regimens. *J Clin Oncol*. 2015;33(33):3921–3929.
264. Rajkumar SV, Kumar S. Multiple Myeloma: Diagnosis and Treatment. *Mayo Clin Proc*. 2016;91(1):101–119.
265. Facon T, Lee JH, Moreau P, et al. Carfilzomib or bortezomib with melphalan-prednisone for transplant-ineligible patients with newly diagnosed multiple myeloma. *Blood*. 2019;133(18):1953–1963.
266. Kumar SK, Jacobus SJ, Cohen AD, et al. Carfilzomib or bortezomib in combination with lenalidomide and dexamethasone for patients with newly diagnosed multiple myeloma without intention for immediate autologous stem-cell transplantation (ENDURANCE): a multicentre, open-label, phase 3, randomised, controlled trial. *Lancet Oncol*. 2020;21(10):1317–1330.
267. Chanan-Khan AA, Swaika A, Paulus A, et al. Pomalidomide: the new immunomodulatory agent for the treatment of multiple myeloma. *Blood Cancer J*. 2013;3(9):e143–e143.
268. Dimopoulos MA, Moreau P, Palumbo A, et al. Carfilzomib and dexamethasone (Kd) vs bortezomib and dexamethasone (Vd) in patients (pts) with relapsed multiple myeloma (RMM): Results from the phase III study ENDEAVOR. *J Clin Oncol*. 2015;33(15_suppl):8509–8509.
269. Jakubowiak AJ, Dytfeld D, Griffith KA, et al. A phase 1/2 study of carfilzomib in combination with lenalidomide and low-dose dexamethasone as a frontline treatment for multiple myeloma. *Blood*. 2012;120(9):1801–1809.
270. Sonneveld P, Zweegman S, Cavo M, et al. Carfilzomib, Pomalidomide and Dexamethasone (KPd) in Patients with Multiple Myeloma Refractory to Bortezomib and Lenalidomide. the EMN011 Trial. *Blood*. 2018;132(Supplement 1):801–801.
271. Usmani SZ, Rodriguez-Otero P, Bhutani M, Mateos M-V, Miguel JS. Defining and treating high-risk multiple myeloma. *Leukemia*. 2015;29(11):2119–2125.

272. Morgan GJ, Davies FE, Gregory WM, et al. Cyclophosphamide, thalidomide, and dexamethasone as induction therapy for newly diagnosed multiple myeloma patients destined for autologous stem-cell transplantation: MRC Myeloma IX randomized trial results. *Haematologica*. 2012;97(3):442–450.
273. Attal M, Lauwers-Cances V, Marit G, et al. Lenalidomide Maintenance after Stem-Cell Transplantation for Multiple Myeloma. *New Engl J Medicine*. 2012;366(19):1782–1791.
274. Cavo M, Salwender H, Rosiñol L, et al. Double Vs Single Autologous Stem Cell Transplantation After Bortezomib-Based Induction Regimens For Multiple Myeloma: An Integrated Analysis Of Patient-Level Data From Phase European III Studies. *Blood*. 2013;122(21):767–767.
275. Russell SJ, Rajkumar SV. Multiple myeloma and the road to personalised medicine. *Lancet Oncol*. 2011;12(7):617–619.
276. Mikhael JR, Dingli D, Roy V, et al. Management of Newly Diagnosed Symptomatic Multiple Myeloma: Updated Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) Consensus Guidelines 2013. *Mayo Clin Proc*. 2013;88(4):360–376.
277. Andel H van, Kocemba KA, Spaargaren M, Pals ST. Aberrant Wnt signaling in multiple myeloma: molecular mechanisms and targeting options. *Leukemia*. 2019;33(5):1063–1075.
278. Fulciniti M, Tassone P, Hideshima T, et al. Anti-DKK1 mAb (BHQ880) as a potential therapeutic agent for multiple myeloma. *Blood*. 2009;114(2):371–379.
279. Touzeau C, Dousset C, Gouill SL, et al. The Bcl-2 specific BH3 mimetic ABT-199: a promising targeted therapy for t(11;14) multiple myeloma. *Leukemia*. 2014;28(1):210–212.
280. Shaughnessy JD, Qu P, Usmani S, et al. Pharmacogenomics of bortezomib test-dosing identifies hyperexpression of proteasome genes, especially PSMD4, as novel high-risk feature in myeloma treated with Total Therapy 3. *Blood*. 2011;118(13):3512–3524.
281. Pawlyn C, Davies FE. Toward personalized treatment in multiple myeloma based on molecular characteristics. *Blood*. 2019;133(7):660–675.

282. Andrulis M, Lehners N, Capper D, et al. Targeting the BRAF V600E Mutation in Multiple Myeloma. *Cancer Discov*. 2013;3(8):862–869.
283. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *New Engl J Medicine*. 2017;377(1):62–70.
284. Berger MF, Mardis ER. The emerging clinical relevance of genomics in cancer medicine. *Nat Rev Clin Oncol*. 2018;15(6):353–365.
285. Yates LR, Seoane J, Tourneau CL, et al. The European Society for Medical Oncology (ESMO) Precision Medicine Glossary. *Ann Oncol*. 2018;29(1):30–35.
286. Mahmoud M, Gobet N, Cruz-Dávalos DI, et al. Structural variant calling: the long and the short of it. *Genome Biol*. 2019;20(1):246.
287. Amarasinghe SL, Su S, Dong X, et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):30.
288. Bashir A, Volik S, Collins C, Bafna V, Raphael BJ. Evaluation of Paired-End Sequencing Strategies for Detection of Genome Rearrangements in Cancer. *Plos Comput Biol*. 2008;4(4):e1000051.
289. Phillips KA, Douglas MP, Marshall DA. Expanding Use of Clinical Genome Sequencing and the Need for More Data on Implementation. *Jama*. 2020;324(20):2029–2030.
290. Ulahannan D, Kovac MB, Mulholland PJ, Cazier J-B, Tomlinson I. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *Brit J Cancer*. 2013;109(4):827–835.
291. Aguilera-Diaz A, Vazquez I, Ariceta B, et al. Assessment of the clinical utility of four NGS panels in myeloid malignancies. Suggestions for NGS panel choice or design. *Plos One*. 2020;15(1):e0227986.
292. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? *Hum Genet*. 2016;135(3):359–362.
293. Initiative MG, Marshall CR, Chowdhury S, et al. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *Npj Genom Medicine*. 2020;5(1):47.

294. LeBlanc VG, Marra MA. Next-Generation Sequencing Approaches in Cancer: Where Have They Brought Us and Where Will They Take Us? *Cancers*. 2015;7(3):1925–1958.
295. Bolli N, Genuardi E, Ziccheddu B, et al. Next-Generation Sequencing for Clinical Management of Multiple Myeloma: Ready for Prime Time? *Frontiers Oncol*. 2020;10:189.
296. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet*. 2013;14(10):703–718.
297. Karlovich CA, Williams PM. Clinical Applications of Next-Generation Sequencing in Precision Oncology. *Cancer J*. 2019;25(4):264–271.
298. Sahajpal NS, Mondal AK, Ananth S, et al. Clinical performance and utility of a comprehensive next-generation sequencing DNA panel for the simultaneous analysis of variants, TMB and MSI for myeloid neoplasms. *Plos One*. 2020;15(10):e0240976.
299. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018;173(2):291-304.e6.
300. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134–1140.
301. Shaughnessy J, Tian E, Sawyer J, et al. High incidence of chromosome 13 deletion in multiple myeloma detected by multiprobe interphase FISH. *Blood*. 2000;96(4):1505–1511.
302. Boyd KD, Ross FM, Tapper WJ, et al. The clinical impact and molecular biology of del(17p) in multiple myeloma treated with conventional or thalidomide-based therapy. *Genes Chromosomes Cancer*. 2011;50(10):765–774.
303. Moreno-Cabrera JM, Valle J del, Castellanos E, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur J Hum Genet*. 2020;28(12):1645–1655.
304. Yao R, Yu T, Qing Y, Wang J, Shen Y. Evaluation of copy number variant detection from panel-based next-generation sequencing data. *Mol Genetics Genom Medicine*. 2019;7(1):e00513.

305. Kerkhof J, Schenkel LC, Reilly J, et al. Clinical Validation of Copy Number Variant Detection from Targeted Next-Generation Sequencing Panels. *J Mol Diagnostics*. 2017;19(6):905–920.
306. Arthur JG, Chen X, Zhou B, Urban AE, Wong WH. Detection of complex structural variation from paired-end sequencing data. *Biorxiv*. 2018;200170.
307. Belzen IAEM van, Schönhuth A, Kemmeren P, Hehir-Kwa JY. Structural variant detection in cancer genomes: computational challenges and perspectives for precision oncology. *Npj Precis Oncol*. 2021;5(1):15.
308. Wala J, Bandopadhayay P, Greenwald N, et al. SvABA: Genome-wide detection of structural variants and indels by local assembly. *Biorxiv*. 2017;105080.
309. Cameron DL, Schröder J, Penington JS, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res*. 2017;27(12):2050–2060.
310. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15(6):R84.
311. Zhan F, Sawyer J, Tricot G. The role of cytogenetics in myeloma. *Leukemia*. 2006;20(9):1484–1486.
312. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med*. 2020;12(1):91.
313. Bailey MH, Meyerson WU, Dursi LJ, et al. Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nat Commun*. 2020;11(1):4748.
314. Zahir FR, Mwenifumbo JC, Chun H-JE, et al. Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. *Bmc Genomics*. 2017;18(1):403.
315. Bewicke-Copley F, Kumar EA, Palladino G, Korfi K, Wang J. Applications and analysis of targeted genomic sequencing in cancer studies. *Comput Struct Biotechnology J*. 2019;17:1348–1359.
316. Zhao C, Jiang T, Ju JH, et al. TruSight Oncology 500: Enabling Comprehensive Genomic Profiling and Biomarker Reporting with Targeted Sequencing. *Biorxiv*. 2020;2020.10.21.349100.

317. Massard C, Michiels S, Ferté C, et al. High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. *Cancer Discov*. 2017;7(6):586–595.
318. Merino DM, McShane LM, Fabrizio D, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J Immunother Cancer*. 2020;8(1):e000147.
319. Malone ER, Oliva M, Sabatini PJB, Stockley TL, Siu LL. Molecular profiling for precision cancer therapies. *Genome Med*. 2020;12(1):8.
320. Horak P, Fröhling S, Glimm H. Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. *Esmo Open*. 2016;1(5):e000094.
321. Bian X, Zhu B, Wang M, et al. Comparing the performance of selected variant callers using synthetic data and genome segmentation. *Bmc Bioinformatics*. 2018;19(1):429.
322. Yang R, Etten JLV, Dehm SM. Indel detection from DNA and RNA sequencing data with transIndel. *Bmc Genomics*. 2018;19(1):270.
323. Yang R, Nelson AC, Henzler C, Thyagarajan B, Silverstein KAT. ScanIndel: a hybrid framework for indel detection via gapped alignment, split reads and de novo assembly. *Genome Med*. 2015;7(1):127.
324. Chen J, Guo J. Comparative assessments of indel annotations in healthy and cancer genomes with next-generation sequencing data. *Bmc Med Genomics*. 2020;13(1):170.
325. Carter MD, Gaston D, Huang W-Y, et al. Genetic profiles of different subsets of Merkel cell carcinoma show links between combined and pure MCPyV-negative tumors. *Hum Pathol*. 2018;71:117–125.
326. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865–2871.
327. Cameron DL, Stefano LD, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun*. 2019;10(1):3240.

328. Narzisi G, O'Rawe JA, Iossifov I, et al. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat Methods*. 2014;11(10):1033–1036.
329. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108–e108.
330. 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.
331. Consortium W, Rimmer A, Phan H, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet*. 2014;46(8):912–918.
332. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *Arxiv*. 2012;
333. Zhang L, Bai W, Yuan N, Du Z. Comprehensively benchmarking applications for detecting copy number variation. *Plos Comput Biol*. 2019;15(5):e1007069.
334. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res*. 2014;24(12):2022–2032.
335. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.
336. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–2849.
337. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811–2812.
338. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *Bmc Bioinformatics*. 2011;12(1):77.
339. Wickham H. ggplot2, Elegant Graphics for Data Analysis. *R*. 2016;

340. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Arxiv*. 2013;

341. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*. 2012;6(2):80–92.

342. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.

343. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. *Biorxiv*. 2016;041863.

344. Jabot-Hanin F, Varet H, Tores F, Alcaïs A, Jaïs J-P. Rfpred: A Random Forest Approach for Prediction of Missense Variants in Human Exome. *Biorxiv*. 2016;037127.

345. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *Biorxiv*. 2018;201178.

346. Sim N-L, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(W1):W452–W457.

347. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575–576.

348. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genetics*. 2013;76(1):7.20.1-7.20.41.

349. Rogers MF, Shihab HA, Mort M, et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*. 2017;34(3):511–513.

350. Jaganathan K, Panagiotopoulou SK, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24.

351. Derouault P, Chauzeix J, Rizzo D, et al. CovCopCan: An efficient tool to detect Copy Number Variation from amplicon sequencing data in inherited diseases and cancer. *Plos Comput Biol*. 2020;16(2):e1007503.

352. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. *Statistics Biology Heal*. 2000;

353. Kyle RA, Durie BGM, Rajkumar SV, et al. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma: IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management. *Leukemia*. 2010;24(6):1121–1127.
354. Rajkumar SV. Multiple myeloma: 2018 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2018;93(8):1091–1110.
355. Kortüm KM, Mai EK, Hanafiah NH, et al. Targeted sequencing of refractory myeloma reveals a high incidence of mutations in CRBN and Ras pathway genes. *Blood*. 2016;128(9):1226–1233.
356. Barrio S, DáVia M, Bruins L, et al. Multiple Myeloma, Methods and Protocols. *Methods Mol Biology Clifton N J*. 2018;1792:117–128.
357. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2018;47(D1):gky1015-.
358. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2017;46(D1):gkx1153-.
359. Kosinski A, Kosinski M, Biecek P. Drawing survival curves using “ggplot2” R package. 2020.
360. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. *Stat Med*. 2001;20(13):2053–2054.
361. Chng WJ, Chung T-H, Kumar S, et al. Gene signature combinations improve prognostic stratification of multiple myeloma patients. *Leukemia*. 2016;30(5):1071–1078.
362. Merz M, Hielscher T, Seckinger A, et al. Baseline characteristics, chromosomal alterations, and treatment affecting prognosis of deletion 17p in newly diagnosed myeloma. *Am J Hematol*. 2016;91(11):E473–E477.
363. Hebraud B, Magrangeas F, Cleynen A, et al. Role of additional chromosomal changes in the prognostic value of t(4;14) and del(17p) in multiple myeloma: the IFM experience. *Blood*. 2015;125(13):2095–2100.
364. Thanendrarajan S, Tian E, Qu P, et al. The level of deletion 17p and bi-allelic inactivation of TP53 has a significant impact on clinical outcome in multiple myeloma. *Haematologica*. 2017;102(9):e364–e367.

365. Neben K, Lokhorst HM, Jauch A, et al. Administration of bortezomib before and after autologous stem cell transplantation improves outcome in multiple myeloma patients with deletion 17p. *Blood*. 2012;119(4):940–948.
366. An G, Li Z, Tai Y-T, et al. The Impact of Clone Size on the Prognostic Value of Chromosome Aberrations by Fluorescence In Situ Hybridization in Multiple Myeloma. *Clin Cancer Res*. 2015;21(9):2148–2156.
367. Nieuwenhuijzen N van, Spaan I, Raymakers R, Peperzak V. From MGUS to Multiple Myeloma, a Paradigm for Clonal Evolution of Premalignant Cells. *Cancer Res*. 2018;78(10):2449–2456.
368. Dutta AK, Hewett DR, Fink JL, Zannettino ACW. Using genomics to better define high-risk MGUS/SMM patients. *Oncotarget*. 2018;9(93):36549–36550.
369. Mouhieddine TH, Weeks LD, Ghobrial IM. Monoclonal gammopathy of undetermined significance. *Blood*. 2019;133(23):2484–2494.
370. Smith D, Armenteros E, Percy L, et al. RAS mutation status and bortezomib therapy for relapsed multiple myeloma. *Brit J Haematol*. 2015;169(6):905–908.
371. Gebauer N, Biersack H, Czerwinska A-C, et al. Favorable prognostic impact of RAS mutation status in multiple myeloma treated with high-dose melphalan and autologous stem cell support in the era of novel agents: a single center perspective. *Leukemia Lymphoma*. 2015;57(1):226–229.
372. Jovanović KK, Escure G, Demonchy J, et al. Deregulation and Targeting of TP53 Pathway in Multiple Myeloma. *Frontiers Oncol*. 2019;8:665.
373. Latreille J, Barlogie B, Johnston D, Drewinko B, Alexanian R. Ploidy and proliferative characteristics in monoclonal gammopathies. *Blood*. 1982;59(1):43–51.
374. Bergsagel PL, Nardini E, Brents L, Chesi M, Kuehl WM. C-Myc in B-Cell Neoplasia, 14th Workshop on Mechanisms in B-Cell Neoplasia. *Curr Top Microbiol*. 1997;224:283–287.
375. Smadja NV, Bastard C, Brigaudeau C, et al. Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood*. 2001;98(7):2229–2238.
376. Avet-Loiseau H, Attal M, Moreau P, et al. Genetic abnormalities and survival in multiple myeloma: the experience of the Intergroupe Francophone du Myélome. *Blood*. 2007;109(8):3489–3495.

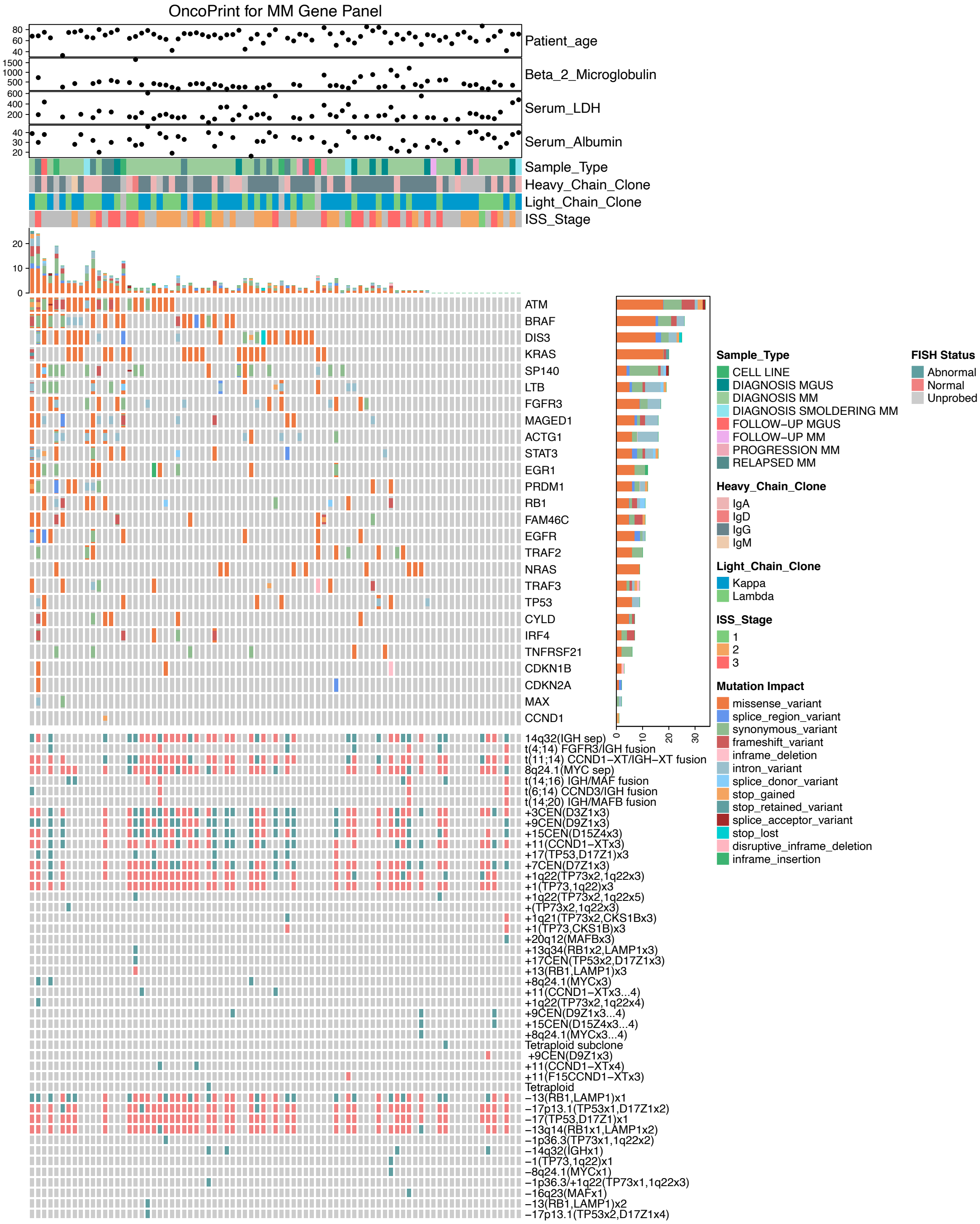
377. Khoury MJ, Coates RJ, Fennell ML, et al. Multilevel Research and the Challenges of Implementing Genomic Medicine. *Jnci Monogr*. 2012;2012(44):112–120.
378. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
379. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep-uk*. 2019;9(1):9354.
380. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res*. 2016;44(D1):D164–D171.
381. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012;22(3):276–282.
382. Lomas O, Gooding S, Ramasamy K, et al. Clinical-Grade Whole Genome Sequencing Reproduces FISH Cytogenetics and Provides Actionable Data in Newly Diagnosed Myeloma - a Pilot Study from the UK 100,000 Genomes Project. *Blood*. 2019;134(Supplement_1):3062–3062.
383. Höllein A, Twardziok SO, Walter W, et al. The combination of WGS and RNA-Seq is superior to conventional diagnostic tests in multiple myeloma: Ready for prime time? *Cancer Genet-ny*. 2020;242:15–24.
384. Jenner MW, Leone PE, Walker BA, et al. Abnormalities of 16q in Multiple Myeloma Are Associated with Poor Prognosis: 500K Gene Mapping and Expression Correlations Identify Two Potential Tumor Suppressor Genes, WWOX and CYLD. *Blood*. 2006;108(11):110–110.
385. Andel H van, Kocemba KA, Haan-Kramer A de, et al. Loss of CYLD expression unleashes Wnt signaling in multiple myeloma and is associated with aggressive disease. *Oncogene*. 2017;36(15):2105–2115.
386. He H, Fu W, Jiang H, et al. The clinical characteristics and prognosis of IGH deletion in multiple myeloma. *Leukemia Res*. 2015;39(5):515–519.
387. Mikulasova A, Ashby C, Tytarenko RG, et al. Microhomology-mediated end joining drives complex rearrangements and over-expression of MYC and PVT1 in multiple myeloma. *Biorxiv*. 2019;515106.

388. Walker BA, Wardell CP, Begum D, et al. MYC Translocations In Multiple Myeloma Involve Recruitment Of Enhancer Elements Resulting In Over-Expression and Decreased Overall Survival. *Blood*. 2013;122(21):274–274.
389. Avet-Loiseau H, Malard F, Campion L, et al. Translocation t(14;16) and multiple myeloma: is it really an independent prognostic factor? *Blood*. 2011;117(6):2009–2011.
390. Fountzilas E, Kotoula V, Koliou G-A, et al. Pathogenic mutations and overall survival in 3,084 patients with cancer: the Hellenic Cooperative Oncology Group Precision Medicine Initiative. *Oncotarget*. 2020;11(1):1–14.
391. Li J, Zhao T, Zhang Y, et al. Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res*. 2018;46(15):gky678-.
392. Zhang X, Walsh R, Whiffin N, et al. Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *Genet Med*. 2021;23(1):69–79.
393. Evans P, Wu C, Lindy A, et al. Genetic variant pathogenicity prediction trained using disease-specific clinical sequencing data sets. *Genome Res*. 2019;29(7):1144–1151.
394. Raab MS, Lehners N, Xu J, et al. Spatially divergent clonal evolution in multiple myeloma: overcoming resistance to BRAF inhibition. *Blood*. 2016;127(17):2155–2157.
395. Kamath AV, Lu D, Gupta P, et al. Preclinical pharmacokinetics of MFGR1877A, a human monoclonal antibody to FGFR3, and prediction of its efficacious clinical dose for the treatment of t(4;14)-positive multiple myeloma. *Cancer Chemoth Pharm*. 2012;69(4):1071–1078.
396. Scholtes M, Akbarzadeh M, Zwarthoff E, et al. Targeted Therapy in Metastatic Bladder Cancer: Present Status and Future Directions. *Appl Sci*. 2020;10(20):7102.
397. Menezes DL, Holt J, Tang Y, et al. A Synthetic Lethal Screen Reveals Enhanced Sensitivity to ATR Inhibitor Treatment in Mantle Cell Lymphoma with ATM Loss-of-Function. *Mol Cancer Res*. 2015;13(1):120–129.
398. Jones CD, Blades K, Foote KM, et al. Abstract 2348: Discovery of AZD6738, a potent and selective inhibitor with the potential to test the clinical efficacy of ATR kinase inhibition in cancer patients. *Cancer Chem*. 2013;2348–2348.

399. Garcia-Heras J. Delineating the Complex Genomic Landscape of Multiple Myeloma Using Next-Generation Sequencing (NGS): Progress and Potential to Supersede Traditional Genetic Testing. *J Assoc Genetic Technologists*. 2020;46(3):131–134.
400. Minoche AE, Lundie B, Peters GB, et al. ClinSV: clinical grade structural and copy number variant detection from whole genome sequencing data. *Genome Med*. 2021;13(1):32.
401. Neerman N, Faust G, Meeks N, et al. A clinically validated whole genome pipeline for structural variant detection and analysis. *Bmc Genomics*. 2019;20(Suppl 8):545.
402. Roepman P, Bruijn E de, Lieshout S van, et al. Clinical validation of Whole Genome Sequencing for cancer diagnostics. *Medrxiv*. 2021;2020.10.29.20222091.
403. Gong T, Hayes VM, Chan EKF. Detection of somatic structural variants from short-read next-generation sequencing data. *Brief Bioinform*. 2020;bbaa056-.
404. Kosugi S, Momozawa Y, Liu X, et al. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol*. 2019;20(1):117.
405. Mikulasova A, Ashby CC, Tytarenko RG, et al. MYC Rearrangements in Multiple Myeloma Are Complex, Can Involve More Than Five Different Chromosomes, and Correlate with Increased Expression of MYC and a Distinct Downstream Gene Expression Pattern. *Blood*. 2017;130(Suppl_1):65–65.
406. Gabrea A, Martelli ML, Qi Y, et al. Secondary genomic rearrangements involving immunoglobulin or MYC loci show similar prevalences in hyperdiploid and nonhyperdiploid myeloma tumors. *Genes Chromosomes Cancer*. 2008;47(7):573–590.
407. Affer M, Chesi M, Chen WD, et al. Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia*. 2014;28(8):1725–1735.
408. Chesi M, Bergsagel PL. Molecular pathogenesis of multiple myeloma: basic and clinical updates. *Int J Hematol*. 2013;97(3):313–323.
409. 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.

410. Boyle EM, Ashby C, Tytarenko RG, et al. BRAF and DIS3 Mutations Associate with Adverse Outcome in a Long-term Follow-up of Patients with Multiple Myeloma. *Clin Cancer Res*. 2020;26(10):2422–2432.
411. Yamga E, Leblanc R, Boudreault J-S. Canadian Retrospective Multicentric Study Comparing the Combination of Bortezomib, Cyclophosphamide and Dexamethasone (Cybord) Versus Bortezomib, Thalidomide and Dexamethasone (VTD) Versus Bortezomib and Dexamethasone (Vd) in Patients with Newly Diagnosed Multiple Myeloma Eligible for Autologous Hematopoietic Stem Cell Transplantation (HCT). *Blood*. 2018;132(Supplement 1):2010–2010.
412. Goldman-Mazur S, Jurczyszyn A, Castillo JJ, et al. A multicenter retrospective study of 223 patients with t(14;16) in multiple myeloma. *Am J Hematol*. 2020;95(5):503–509.
413. Mina R, Joseph NS, Gay F, et al. Clinical features and survival of multiple myeloma patients harboring t(14;16) in the era of novel agents. *Blood Cancer J*. 2020;10(4):40.
414. Atli EI, Gurkan H, Kirkizlar HO, et al. Pros and cons for fluorescent in situ hybridization, karyotyping and next generation sequencing for diagnosis and follow-up of multiple myeloma. *Balk J Med Genet*. 2021;23(2):59–64.
415. Lakshman A, Rajkumar SV, Buadi FK, et al. Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. *Blood Cancer J*. 2018;8(6):59.
416. Mateos M-V, Hernández M-T, Giraldo P, et al. Lenalidomide plus dexamethasone versus observation in patients with high-risk smouldering multiple myeloma (QuiRedex): long-term follow-up of a randomised, controlled, phase 3 trial. *Lancet Oncol*. 2016;17(8):1127–1136.
417. Miguel JS, Mateos M-V, Gonzalez V, et al. Updated risk stratification model for smoldering multiple myeloma (SMM) incorporating the revised IMWG diagnostic criteria. *J Clin Oncol*. 2019;37(15_suppl):8000–8000.

Appendices



Appendix Figure 1: Cohort mutational, laboratory, and clinical landscape
Oncoprint comprising demographic, clinical, laboratory, molecular, and genomic data for our cohort. Data on cell lines have been included. Columns are patients/cell lines.

Appendix Table 1: Sample library preparation

Sample	Library
MM8	TruSight
MM10	TruSight
MM12	TruSight
MM13	TruSight
MM18	TruSight
MM21	TruSight
MM23	TruSight
MM24	TruSight
MM25	TruSight
MM27	TruSight
MM28	TruSight
MM29	TruSight
MM30	TruSight
MM31	TruSight
MM32	TruSight
MM33	TruSight
MM34	TruSight
MM35	TruSight
MM36	TruSight
MM37	TruSight
MM38	TruSight
MM39	TruSight
MM40	TruSight
MM42	TruSight
MM43	TruSight
MM44	TruSight
MM46	TruSight
MM47	TruSight
MM48	TruSight
MM51	TruSight
MM52	TruSight
MM53	TruSight
MM54	TruSight
MM55	TruSight
MM56	TruSight
MM57	TruSight
MM58	TruSight
MM59	TruSight
MM62	TruSight
MM63	TruSight
MM64	TruSight
MM67	TruSight
MM68	TruSight
MM69	TruSight
MM70	TruSight
MM1S	TruSight
NCI	TruSight
RPMI	TruSight
KMS12	AmpliSeq
MM06	AmpliSeq
MM11	AmpliSeq
MM14	AmpliSeq
MM17	AmpliSeq
MM22	AmpliSeq
MM26	AmpliSeq
MM60	AmpliSeq
MM65	AmpliSeq
MM66	AmpliSeq
MM71	AmpliSeq
MM72	AmpliSeq
MM73	AmpliSeq
MM75	AmpliSeq
MM77	AmpliSeq
MM78	AmpliSeq
MM79	AmpliSeq
MM81	AmpliSeq
MM82	AmpliSeq
MM84	AmpliSeq
MM86	AmpliSeq
MM88	AmpliSeq
MM89	AmpliSeq
MM91	AmpliSeq
MM92	AmpliSeq
MM93	AmpliSeq
MM95	AmpliSeq
MM97	AmpliSeq
MM99	AmpliSeq
MM100	AmpliSeq
MM102	AmpliSeq
MM106	AmpliSeq
MM108	AmpliSeq

Appendix Table 2: Panel identified variants

All panel identified variants passing filtering and manual review in patient samples and cell lines.

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
10	FGFR3	A	G	c.1948A>G	p.Lys650Glu	0.47584188	freebayes,mutect,vardict	4	1807888	1807889
10	EGR1	T	A	c.29T>A	p.Leu10Gln	0.32876712	mutect,vardict	5	137801478	137801479
10	CDKN2A	G	A	c.457+8C>T	NA	0.112	freebayes,mutect,vardict ,platypus	9	21970892	21970893
10	TRAF2	G	T	c.984G>T	p.Glu328Asp	0.12666076	freebayes,mutect,vardict ,platypus	9	139814834	139814835
10	ACTG1	G	A	c.684C>T	p.Ala228Ala	0.6231454	freebayes,mutect,vardict	17	79478331	79478332
10	EGR1	T	A	c.99T>A	p.Pro33Pro	0.3721374	freebayes,mutect,vardict ,platypus	5	137801548	137801549
MM100	NRAS	T	C	c.182A>G	p.Gln61Arg	0.59427446	freebayes,mutect,vardict	1	115256528	115256529
MM102	NRAS	T	C	c.182A>G	p.Gln61Arg	0.32473624	freebayes,mutect,vardict ,platypus	1	115256528	115256529
MM102	DIS3	T	C	c.541A>G	p.Lys181Glu	0.4609375	freebayes,mutect,vardict ,platypus	13	73352363	73352364
MM102	MAGED1	C	T	c.1157C>T	p.Pro386Leu	1	freebayes,mutect,vardict	X	51639739	51639740
MM106	RB1	TCAGAA	T	c.772_776delAACAG	p.Asn258fs	1	freebayes,vardict,scalpel ,platypus,pindel	13	48936999	48937005
MM106	CDKN1B	AAGTGGAA TTTCGATTT TC	A	c.178_195delTGGAA TTTCGATTTTCAG	p.Trp60_Gln65del	1	freebayes,vardict,scalpel ,platypus,pindel	12	12870947	12870966
MM106	PRDM1	G	A	c.1279G>A	p.Ala427Thr	0.99553573	freebayes,mutect,vardict	6	106553313	106553314
MM108	KRAS	C	T	c.38G>A	p.Gly13Asp	0.41382667	freebayes,mutect,vardict ,platypus	12	25398280	25398281
MM108	BRAF	AAAAAAAA AAG	A	c.2128-16_2128-7delCTTTTTTTTT	NA	0.90869564	vardict,scalpel,platypus	7	140434575	140434586
13	KRAS	C	G	c.34G>C	p.Gly12Arg	0.27991885	freebayes,mutect,platyp us	12	25398284	25398285
13	ATM	C	T	c.6176C>T	p.Thr2059Ile	0.41678256	freebayes,mutect,vardict	11	108186817	108186818
13	PRDM1	G	C	c.957G>C	p.Glu319Asp	0.3041825	freebayes,mutect,vardict	6	106552991	106552992
13	BRAF	G	A	c.1518-10C>T	NA	0.52337515	freebayes,mutect,vardict ,platypus	7	140476897	140476898
MM14	ATM	A	T	c.5558A>T	p.Asp1853Val	0.5035311	freebayes,mutect,vardict	11	108175462	108175463
18	SP140	G	A	c.2391G>A	p.Glu797Glu	0.56705171	freebayes,mutect,vardict	2	231176195	231176196
18	RB1	G	T	c.264+1G>T	NA	0.21980676	freebayes,mutect,platyp us	13	48881542	48881543
18	TRAF3	C	T	c.35C>T	p.Ala12Val	0.16535123	freebayes,mutect,vardict ,platypus	14	103336572	103336573
18	FGFR3	CAGTGAG	C	c.931-765_931-760delAGAGTG	NA	0.30915618	freebayes,vardict,scalpel ,pindel	4	1804648	1804655
21	ATM	TA	T	c.8530delA	p.Ile2844fs	0.34782609	freebayes,vardict,pindel	11	108216579	108216581
21	SP140	C	CA	c.1404dupA	p.Glu469fs	0.13622291	freebayes,platypus,pind el	2	231134626	231134627
21	FGFR3	A	C	c.1580A>C	p.Glu527Ala	0.51249999	freebayes,mutect	4	1807330	1807331
21	EGR1	A	T	c.487A>T	p.Ser163Cys	0.19387755	freebayes,mutect	5	137802624	137802625
21	ATM	A	T	c.1410A>T	p.Ser470Ser	0.22772278	freebayes,mutect	11	108121601	108121602
21	DIS3	G	A	c.2511+21C>T	NA	0.11310345	freebayes,mutect	13	73335762	73335763
21	PRDM1	G	A	c.2406G>A	p.Leu802Leu	0.95454544	freebayes,mutect	6	106555288	106555289
21	MAGED1	A	G	c.46-82A>G	NA	0.35714287	freebayes,mutect	X	51637640	51637641
MM22	TRAF2	G	A	c.948G>A	p.Ala316Ala	0.38879159	freebayes,mutect,vardict	9	139814798	139814799
23	LTB	T	G	c.208+86A>C	NA	0.33167967	freebayes,mutect,vardict	6	31549504	31549505

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
23	LTB	T	A	c.208+9A>T	NA	0.44487429	freebayes,mutect,vardict,platypus	6	31549581	31549582
24	NRAS	C	T	c.179G>A	p.Gly60Glu	0.15369262	freebayes,mutect,vardict,platypus	1	115256531	115256532
24	DIS3	T	C	c.1463A>G	p.Asp488Gly	0.4067432	freebayes,mutect,vardict	13	73346336	73346337
25	BRAF	GC	G	c.2156delG	p.Arg719fs	0.28	freebayes,platypus,pindel	7	140434540	140434542
25	EGR1	C	T	c.766C>T	p.Leu256Leu	0.15322581	freebayes,mutect,vardict	5	137802903	137802904
25	MAGED1	A	T	c.1950-98A>T	NA	0.30612245	freebayes,mutect	X	51641578	51641579
25	MAGED1	C	T	c.1827-73C>T	NA	0.25396827	freebayes,mutect,vardict	X	51641136	51641137
MM26	BRAF	A	T	c.1799T>A	p.Val600Glu	0.20714286	freebayes,mutect,vardict,platypus	7	140453135	140453136
27	FAM46C	ACT	A	c.400_401delCT	p.Leu134fs	0.12403101	freebayes,platypus,pindel	1	118165886	118165889
27	BRAF	G	C	c.1466C>G	p.Ala489Gly	0.25210086	freebayes,mutect,vardict,platypus	7	140477841	140477842
27	FAM46C	T	G	c.1128T>G	p.Pro376Pro	0.2060606	freebayes,mutect,platypus	1	118166617	118166618
29	STAT3	T	C	c.550+3A>G	NA	0.1328125	freebayes,mutect,vardict,platypus	17	40490745	40490746
29	MAGED1	A	G	c.1832A>G	p.Lys611Arg	0.18779343	freebayes,mutect,platypus	X	51641214	51641215
29	MAGED1	T	G	c.2292T>G	p.Asp764Glu	0.11016949	freebayes,mutect	X	51644812	51644813
31	TRAF3	G	A	c.778G>A	p.Val260Ile	0.14242424	freebayes,mutect,vardict,platypus	14	103357712	103357713
31	SP140	G	A	c.2391G>A	p.Glu797Glu	0.28376845	freebayes,mutect,vardict	2	231176195	231176196
31	FAM46C	T	C	c.313T>C	p.Phe105Leu	0.1245283	freebayes,mutect,vardict	1	118165802	118165803
31	ATM	TA	T	c.8094delA	p.Leu2698fs	0.13110182	freebayes,vardict,scalpel,platypus,pindel	11	108205777	108205779
31	RB1	TG	T	c.253delG	p.Asp85fs	0.11209439	freebayes,platypus,pindel	13	48881528	48881530
31	BRAF	G	T	c.805C>A	p.His269Asn	0.25454545	freebayes,mutect,vardict,platypus	7	140501266	140501267
31	MAGED1	C	T	c.1827-7C>T	NA	0.13414635	freebayes,mutect	X	51641202	51641203
31	MAGED1	A	C	c.1835A>C	p.Asp612Ala	0.11076923	freebayes,mutect	X	51641217	51641218
31	MAX	G	T	c.339C>A	p.Thr113Thr	0.11016949	freebayes,mutect,vardict,platypus	14	65543337	65543338
31	BRAF	T	C	c.1227A>G	p.Ser409Ser	0.72020727	freebayes,mutect,vardict,platypus	7	140482907	140482908
32	TNFRSF21	C	T	c.184G>A	p.Gly62Ser	0.48251748	freebayes,mutect,vardict	6	47254243	47254244
33	TP53	GT	G	c.75-22delA	NA	0.65882355	freebayes,vardict,platypus	17	7579741	7579743
34	FGFR3	T	A	c.2051T>A	p.Leu684His	0.10769231	freebayes,mutect,vardict	4	1808292	1808293
34	EGFR	C	T	c.3629C>T	p.Ala1210Val	0.19300362	freebayes,mutect	7	55273305	55273306
37	MAGED1	C	T	c.45+52C>T	NA	0.12790698	freebayes,mutect	X	51637496	51637497
38	NRAS	T	C	c.182A>G	p.Gln61Arg	0.89393938	freebayes,mutect,vardict,platypus	1	115256528	115256529
39	KRAS	C	G	c.34G>C	p.Gly12Arg	0.27221438	freebayes,mutect,platypus	12	25398284	25398285
39	SP140	G	A	c.2391G>A	p.Glu797Glu	0.32369941	freebayes,mutect,vardict	2	231176195	231176196
40	KRAS	C	G	c.34G>C	p.Gly12Arg	0.25063938	freebayes,mutect,platypus	12	25398284	25398285
40	DIS3	G	A	c.2458C>T	p.Arg820Trp	0.44392523	freebayes,mutect,vardict,platypus	13	73335836	73335837
40	ATM	C	T	c.6176C>T	p.Thr2059Ile	0.35922331	freebayes,mutect,vardict	11	108186817	108186818
40	PRDM1	G	C	c.957G>C	p.Glu319Asp	0.27155173	freebayes,mutect,vardict,platypus	6	106552991	106552992

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
40	BRAF	G	A	c.1518-10C>T	NA	0.51186442	freebayes,mutect,vardict ,platypus	7	140476897	140476898
42	KRAS	C	G	c.34G>C	p.Gly12Arg	0.26666668	freebayes,mutect,platyp us	12	25398284	25398285
42	DIS3	G	A	c.2458C>T	p.Arg820Trp	0.43965518	freebayes,mutect,vardict ,platypus	13	73335836	73335837
42	ATM	C	T	c.6176C>T	p.Thr2059Ile	0.43612334	freebayes,mutect,vardict ,platypus	11	108186817	108186818
42	BRAF	G	A	c.1518-10C>T	NA	0.45222929	freebayes,mutect,vardict ,platypus	7	140476897	140476898
43	FAM46C	T	TAA	c.138_139dupAA	p.Thr47fs	0.46676737	freebayes,vardict,scalpel ,platypus,pindel	1	118165626	118165627
46	BRAF	T	C	c.1801A>G	p.Lys601Glu	0.32785234	freebayes,mutect,vardict	7	140453133	140453134
46	FAM46C	A	G	c.841A>G	p.Ile281Val	0.34750733	freebayes,mutect,vardict	1	118166330	118166331
47	SP140	G	A	c.2391G>A	p.Glu797Glu	0.34026623	freebayes,mutect,vardict	2	231176195	231176196
48	ATM	A	G	c.6914A>G	p.Gln2305Arg	0.10416666	freebayes,mutect	11	108196890	108196891
48	RB1	T	C	c.1312T>C	p.Cys438Arg	0.34558824	freebayes,mutect,platyp us	13	48951149	48951150
48	DIS3	C	T	c.802G>A	p.Asp268Asn	0.15283842	freebayes,mutect,vardict ,platypus	13	73350082	73350083
48	ACTG1	A	T	c.191T>A	p.Ile64Asn	0.29234973	freebayes,mutect,vardict ,platypus	17	79479100	79479101
48	TRAF2	G	T	c.1500G>T	p.Arg500Ser	0.41111112	freebayes,mutect,vardict ,platypus	9	139820190	139820191
48	PRDM1	G	T	c.412-158G>T	NA	0.13707165	freebayes,mutect,vardict ,platypus	6	106547016	106547017
51	RB1	T	A	c.861+2T>A	NA	0.88121545	freebayes,mutect,vardict ,platypus	13	48937094	48937095
51	ATM	C	T	c.4768C>T	p.Leu1590Phe	0.44621515	freebayes,mutect,vardict ,platypus	11	108164195	108164196
51	CDKN1B	G	T	c.151G>T	p.Asp51Tyr	0.99532712	freebayes,mutect,vardict ,platypus	12	12870923	12870924
52	ATM	C	A	c.5697C>A	p.Cys1899*	0.10729023	freebayes,mutect,vardict ,platypus	11	108178645	108178646
52	FAM46C	A	G	c.373A>G	p.Asn125Asp	0.15642458	freebayes,mutect,platyp us	1	118165862	118165863
52	ATM	T	C	c.2408T>C	p.Phe803Ser	0.25943395	freebayes,mutect,vardict ,platypus	11	108129743	108129744
52	ATM	G	A	c.4390G>A	p.Val1464Ile	0.15494978	freebayes,mutect,platyp us	11	108160481	108160482
52	KRAS	CA	C	c.240delT	p.Cys80fs	0.19910179	freebayes,platypus,pind el	12	25380216	25380218
52	TRAF3	G	C	c.1297G>C	p.Val433Leu	0.13875598	freebayes,mutect,vardict ,platypus	14	103371710	103371711
52	STAT3	CT	C	c.1845delA	p.Glu616fs	0.39402986	freebayes,vardict,scalpel ,platypus,pindel	17	40475063	40475065
52	FGFR3	T	A	c.2051T>A	p.Leu684His	0.32768363	freebayes,mutect,platyp us	4	1808292	1808293
52	EGR1	G	A	c.239G>A	p.Ser80Asn	0.14285715	freebayes,mutect,vardict	5	137801688	137801689
52	PRDM1	C	T	c.718C>T	p.Pro240Ser	0.11258278	freebayes,mutect,vardict ,platypus	6	106552752	106552753
52	PRDM1	A	T	c.2317A>T	p.Lys773*	0.18032786	freebayes,mutect,vardict ,platypus	6	106555199	106555200
52	BRAF	T	C	c.1117A>G	p.Thr373Ala	0.48404256	freebayes,mutect,vardict ,platypus	7	140494130	140494131
52	BRAF	GC	G	c.343delG	p.Ala115fs	0.10447761	freebayes,platypus,pind el	7	140534568	140534570
52	MAGED1	A	C	c.292A>C	p.Thr98Pro	0.29834256	freebayes,mutect,platyp us	X	51638226	51638227
52	ATM	T	C	c.546T>C	p.Val182Val	0.21031746	freebayes,mutect,vardict ,platypus	11	108114728	108114729
52	LTB	G	A	c.209-85C>T	NA	0.10954616	freebayes,mutect,vardict ,platypus	6	31549491	31549492
52	EGFR	C	A	c.1881-636C>A	NA	0.26156941	freebayes,mutect,vardict ,platypus	7	55238231	55238232
52	BRAF	G	T	c.2196C>A	p.Ser732Ser	0.10112359	freebayes,mutect,vardict ,platypus	7	140434501	140434502

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
52	KRAS	T	C	c.569A>G	p.Ter190Ter	0.30612245	freebayes,mutect,vardict,platypus	12	25368375	25368376
53	RB1	G	T	c.1573G>T	p.Ala525Ser	0.38675958	freebayes,mutect,vardict,platypus	13	48955456	48955457
53	STAT3	T	C	c.1579A>G	p.Thr527Ala	0.17316018	freebayes,mutect,vardict	17	40476749	40476750
53	EGR1	C	G	c.71C>G	p.Pro24Arg	0.27130976	freebayes,mutect,vardict,platypus	5	137801520	137801521
53	LTB	G	A	c.274C>T	p.Leu92Phe	0.37261146	freebayes,mutect,vardict	6	31549341	31549342
53	EGFR	G	T	c.2061G>T	p.Glu687Asp	0.10557185	freebayes,mutect,vardict	7	55240816	55240817
53	EGFR	G	T	c.2816G>T	p.Cys939Phe	0.19095477	freebayes,mutect,vardict,platypus	7	55266523	55266524
53	BRAF	G	T	c.454C>A	p.Pro152Thr	0.12271541	freebayes,mutect,vardict	7	140534458	140534459
53	MAGED1	T	C	c.2297T>C	p.Ile766Thr	0.41346154	freebayes,mutect	X	51644817	51644818
53	TRAF2	C	T	c.1553C>T	p.Ala518Val	0.53333336	freebayes,mutect,vardict,platypus	9	139820243	139820244
53	STAT3	G	A	c.1452C>T	p.Thr484Thr	0.11267605	freebayes,mutect,vardict	17	40476992	40476993
53	TNFRSF21	C	T	c.1182G>A	p.Leu394Leu	0.10505836	freebayes,mutect,vardict	6	47251734	47251735
53	PRDM1	T	C	c.412-126T>C	NA	0.14869888	freebayes,mutect,vardict,platypus	6	106547048	106547049
53	EGFR	G	T	c.1986G>T	p.Leu662Leu	0.10495627	freebayes,mutect,vardict,platypus	7	55240741	55240742
53	TRAF3	T	A	c.820-10T>A	NA	0.45454547	freebayes,mutect,vardict,platypus	14	103363587	103363588
54	KRAS	C	T	c.38G>A	p.Gly13Asp	0.57884616	freebayes,mutect,vardict,platypus	12	25398280	25398281
54	ATM	TC	T	c.8338delC	p.Val2781fs	0.25125629	freebayes,vardict,scalpel,platypus,pindel	11	108214016	108214018
54	STAT3	G	A	c.1499C>T	p.Thr500Ile	0.1147541	freebayes,mutect,vardict	17	40476829	40476830
54	CYLD	G	A	c.1426G>A	p.Ala476Thr	0.27777779	freebayes,mutect,vardict,platypus	16	50813862	50813863
54	LTB	C	T	c.198G>A	p.Gln66Gln	0.31318682	freebayes,mutect,vardict,platypus	6	31549600	31549601
54	MAGED1	G	C	c.1949+85G>C	NA	0.12398922	freebayes,mutect,vardict	X	51641507	51641508
55	SP140	C	A	c.1890C>A	p.Ile630Ile	0.45895523	freebayes,mutect,vardict,platypus	2	231157424	231157425
56	TRAF3	C	T	c.880C>T	p.Gln294*	0.19235511	freebayes,mutect,platypus	14	103363657	103363658
56	SP140	G	A	c.2391G>A	p.Glu797Glu	0.29245949	freebayes,mutect,vardict	2	231176195	231176196
56	DIS3	G	A	c.1124C>T	p.Pro375Leu	0.8121314	freebayes,mutect,vardict	13	73347936	73347937
57	KRAS	T	A	c.183A>T	p.Gln61His	0.36828241	freebayes,mutect,vardict	12	25380274	25380275
57	BRAF	A	T	c.1799T>A	p.Val600Glu	0.10046457	freebayes,mutect,vardict	7	140453135	140453136
59	BRAF	A	T	c.1799T>A	p.Val600Glu	0.11917808	freebayes,mutect,vardict,platypus	7	140453135	140453136
59	IRF4	TG	T	c.162delG	p.Trp54fs	0.11216566	vardict,platypus,pindel	6	393311	393313
59	MAGED1	GT	G	c.72delT	p.His25fs	0.23563218	freebayes,vardict,platypus,pindel	X	51637747	51637749
62	SP140	G	A	c.2391G>A	p.Glu797Glu	0.61483145	freebayes,mutect,vardict	2	231176195	231176196
62	ATM	T	C	c.4709T>C	p.Val1570Ala	0.41450778	freebayes,mutect,vardict	11	108164136	108164137
63	ATM	C	T	c.3349C>T	p.Gln1117*	0.4710145	freebayes,mutect,vardict,platypus	11	108150281	108150282
63	ATM	TC	T	c.1798delC	p.His600fs	0.10803689	freebayes,platypus,pindel	11	108122752	108122754
63	ATM	T	A	c.5468T>A	p.Ile1823Asn	0.13543308	freebayes,mutect	11	108173727	108173728
63	ATM	T	A	c.8507T>A	p.Met2836Lys	0.431694	freebayes,mutect,platypus	11	108216557	108216558

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
63	DIS3	T	A	c.2600A>T	p.Tyr867Phe	0.32110092	freebayes,mutect,platypus	13	73335570	73335571
63	DIS3	T	C	c.2343-6A>G	NA	0.37142858	freebayes,mutect,platypus	13	73335957	73335958
63	CYLD	GA	G	c.623delA	p.Asp208fs	0.58273381	freebayes,vardict,platypus,pindel	16	50785631	50785633
63	SP140	C	T	c.1031C>T	p.Ser344Phe	0.17368421	freebayes,mutect,vardict,platypus	2	231115749	231115750
63	SP140	G	C	c.1934G>C	p.Arg645Pro	0.11780105	freebayes,mutect,vardict	2	231157468	231157469
63	BRAF	A	G	c.503T>C	p.Val168Ala	0.22887324	freebayes,mutect,platypus	7	140534409	140534410
63	MAGED1	C	A	c.1660C>A	p.Leu554Met	0.25223213	freebayes,mutect,vardict	X	51640647	51640648
63	DIS3	G	C	c.228+153C>G	NA	0.57627118	freebayes,mutect	13	73355589	73355590
63	MAX	G	T	c.172-6221C>A	NA	0.12048193	freebayes,mutect,vardict	14	65550974	65550975
63	TP53	C	A	c.96+11G>T	NA	0.14705883	freebayes,mutect,vardict,platypus	17	7579688	7579689
63	PRDM1	C	T	c.1167C>T	p.Tyr389Tyr	0.16336633	freebayes,mutect,platypus	6	106553201	106553202
63	DIS3	T	C	c.2616A>G	p.Thr872Thr	0.50364965	freebayes,mutect	13	73335554	73335555
MM65	DIS3	C	T	c.586G>A	p.Glu196Lys	0.40305009	freebayes,mutect,vardict,platypus	13	73351625	73351626
68	KRAS	C	G	c.37G>C	p.Gly13Arg	0.23932253	freebayes,mutect,vardict	12	25398281	25398282
68	SP140	G	C	c.2059-1G>C	NA	0.43305278	freebayes,mutect,vardict	2	231174637	231174638
68	ATM	A	G	c.3354A>G	p.Thr1118Thr	0.56905371	freebayes,mutect,platypus	11	108150286	108150287
68	ATM	C	T	c.4473C>T	p.Phe1491Phe	0.52880186	freebayes,mutect,vardict	11	108163381	108163382
69	KRAS	G	T	c.64C>A	p.Gln22Lys	0.21243523	freebayes,mutect,vardict,platypus	12	25398254	25398255
69	BRAF	A	T	c.1799T>A	p.Val600Glu	0.17955439	freebayes,mutect,vardict,platypus	7	140453135	140453136
69	CCND1	C	T	c.742C>T	p.Gln248*	0.67796612	freebayes,mutect,vardict,platypus	11	69465903	69465904
69	RB1	G	C	c.1421G>C	p.Ser474Thr	0.10576923	freebayes,mutect,vardict	13	48954219	48954220
69	CYLD	T	C	c.2471T>C	p.Val824Ala	0.13029316	freebayes,mutect,vardict,platypus	16	50828123	50828124
69	ATM	T	A	c.3336T>A	p.Pro1112Pro	0.49473685	freebayes,mutect,vardict,platypus	11	108150268	108150269
70	SP140	T	C	c.1445-8T>C	NA	0.47103825	freebayes,mutect,vardict,platypus	2	231135292	231135293
70	SP140	G	A	c.1512G>A	p.Gly504Gly	0.49855492	freebayes,mutect,vardict	2	231149073	231149074
MM71	ATM	T	C	c.2258T>C	p.Met753Thr	0.47162426	freebayes,mutect,platypus	11	108128214	108128215
MM71	TRAF3	A	C	c.1483A>C	p.Thr495Pro	0.94613582	freebayes,mutect,vardict	14	103371896	103371897
MM71	IRF4	G	T	c.316G>T	p.Asp106Tyr	0.47333333	freebayes,mutect,vardict	6	394919	394920
MM73	TP53	A	T	c.403T>A	p.Cys135Ser	0.38193202	freebayes,mutect,vardict	17	7578526	7578527
MM73	TRAF2	T	C	c.541T>C	p.Cys181Arg	0.11561866	freebayes,mutect	9	139802539	139802540
MM75	SP140	A	T	c.1678A>T	p.Thr560Ser	0.51241136	freebayes,mutect,vardict,platypus	2	231152638	231152639
MM75	BRAF	T	C	c.1227A>G	p.Ser409Ser	0.55441481	freebayes,mutect,vardict,platypus	7	140482907	140482908
MM77	KRAS	C	G	c.34G>C	p.Gly12Arg	0.16467872	freebayes,mutect,platypus	12	25398284	25398285
MM77	DIS3	A	G	c.2875T>C	p.Ter959Glnext*?	0.53914326	freebayes,mutect,vardict	13	73333934	73333935
MM78	KRAS	T	G	c.183A>C	p.Gln61His	0.28352061	freebayes,mutect,vardict	12	25380274	25380275
MM78	TP53	T	A	c.776A>T	p.Asp259Val	0.43485916	freebayes,mutect,vardict,platypus	17	7577504	7577505

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
MM78	FGFR3	C	T	c.2169-50C>T	NA	0.51284248	freebayes,mutect,vardict	4	1808505	1808506
MM79	DIS3	G	A	c.2458C>T	p.Arg820Trp	0.10727273	freebayes,mutect,vardict,platypus	13	73335836	73335837
MM79	ACTG1	CC	AT	c.363+58_363+59delGGinsAT	NA	1	freebayes,vardict,platypus	17	79478869	79478871
8	DIS3	T	C	c.229-2A>G	NA	0.28448275	freebayes,mutect,platypus	13	73355142	73355143
8	EGFR	G	T	c.1557G>T	p.Glu519Asp	0.42857143	freebayes,mutect,vardict,platypus	7	55229249	55229250
MM81	ATM	A	T	c.5558A>T	p.Asp1853Val	0.66861027	freebayes,mutect,vardict,platypus	11	108175462	108175463
MM81	ATM	T	A	c.7471T>A	p.Trp2491Arg	0.18947369	freebayes,mutect	11	108201103	108201104
MM81	FGFR3	A	T	c.1111A>T	p.Ser371Cys	0.46167883	freebayes,mutect,vardict,platypus	4	1806091	1806092
MM81	ATM	A	T	c.6007-2A>T	NA	0.10039762	freebayes,mutect	11	108186547	108186548
MM81	EGFR	A	G	c.1150A>G	p.Thr384Ala	0.12307692	freebayes,mutect	7	55224467	55224468
MM81	BRAF	T	C	c.249A>G	p.Glu83Glu	0.10132159	freebayes,mutect	7	140534663	140534664
MM82	NRAS	T	C	c.182A>G	p.Gln61Arg	0.43304619	freebayes,mutect,vardict,platypus	1	115256528	115256529
MM82	DIS3	C	T	c.997G>A	p.Ala333Thr	0.54400003	freebayes,mutect,vardict,platypus	13	73348187	73348188
MM84	NRAS	T	C	c.182A>G	p.Gln61Arg	0.3275862	freebayes,mutect,vardict	1	115256528	115256529
MM88	KRAS	G	A	c.437C>T	p.Ala146Val	0.14074074	freebayes,mutect,vardict,platypus	12	25378560	25378561
MM88	EGFR	G	A	c.608G>A	p.Gly203Glu	0.3034682	freebayes,mutect,vardict,platypus	7	55219034	55219035
MM88	SP140	G	C	c.307G>C	p.Glu103Gln	0.29550034	freebayes,mutect,vardict	2	231102996	231102997
MM88	LTB	T	C	c.191A>G	p.Gln64Arg	0.28702012	freebayes,mutect,vardict	6	31549607	31549608
MM88	DIS3	G	A	c.1485C>T	p.Leu495Leu	0.497545	freebayes,mutect,vardict	13	73346314	73346315
MM88	ACTG1	G	A	c.364-87C>T	NA	0.48376259	freebayes,mutect,vardict	17	79478738	79478739
MM89	ATM	T	C	c.1986T>C	p.Phe662Phe	0.50617874	freebayes,mutect,vardict	11	108124627	108124628
MM91	RB1	G	A	c.2360G>A	p.Arg787Gln	0.33776093	freebayes,mutect,vardict,platypus	13	49039374	49039375
MM91	TRAF2	G	A	c.1407G>A	p.Pro469Pro	0.46200609	freebayes,mutect,vardict,platypus	9	139818415	139818416
MM91	LTB	G	A	c.209-48C>T	NA	0.33238637	freebayes,mutect,vardict,platypus	6	31549454	31549455
MM92	ATM	A	T	c.736A>T	p.Asn246Tyr	0.13447432	freebayes,mutect	11	108115587	108115588
MM92	RB1	A	C	c.1148A>C	p.Gln383Pro	0.29807693	freebayes,mutect	13	48947560	48947561
MM92	RB1	G	A	c.1742G>A	p.Gly581Glu	0.1773309	freebayes,mutect	13	49027174	49027175
MM92	DIS3	T	G	c.1862A>C	p.Lys621Thr	0.1150685	freebayes,mutect	13	73342943	73342944
MM92	CYLD	G	T	c.2597G>T	p.Cys866Phe	0.13691275	freebayes,mutect	16	50828249	50828250
MM92	TP53	C	A	c.273G>T	p.Trp91Cys	0.11898017	freebayes,mutect	17	7579413	7579414
MM92	STAT3	T	A	c.1135A>T	p.Arg379*	0.10294118	freebayes,mutect	17	40481768	40481769
MM92	SP140	GT	G	c.892+2delT	NA	0.13043478	freebayes,vardict,platypus,pindel	2	231112780	231112782
MM92	EGFR	A	C	c.748-4A>C	NA	0.13494809	freebayes,mutect	7	55221699	55221700
MM92	BRAF	C	G	c.1169G>C	p.Gly390Ala	0.16710183	freebayes,mutect,vardict	7	140487355	140487356
MM92	BRAF	C	G	c.871G>C	p.Val291Leu	0.1712963	freebayes,mutect	7	140500270	140500271
MM92	ATM	G	A	c.6798G>A	p.Lys2266Lys	0.11214953	freebayes,mutect	11	108196261	108196262

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
MM92	MAGED1	C	T	c.214-127C>T	NA	0.15294118	freebayes,mutect,vardict,platypus	X	51638021	51638022
MM92	MAGED1	T	C	c.1827-46T>C	NA	0.12830189	freebayes,mutect	X	51641163	51641164
MM95	DIS3	C	T	c.2339G>A	p.Arg780Lys	0.47779369	freebayes,mutect,vardict	13	73336063	73336064
MM95	LTB	G	A	c.214C>T	p.Gln72*	0.64026403	freebayes,mutect,vardict,platypus	6	31549401	31549402
MM95	RB1	A	G	c.2463A>G	p.Thr821Thr	0.46830267	freebayes,mutect,vardict	13	49039477	49039478
MM97CONC	ATM	G	A	c.902G>A	p.Gly301Asp	0.47715735	freebayes,mutect,vardict,platypus	11	108117690	108117691
MM97CONC	EGR1	G	A	c.133G>A	p.Ala45Thr	0.51207727	freebayes,mutect,vardict	5	137801582	137801583
MM97CONC	EGR1	A	G	c.184A>G	p.Ser62Gly	0.66612381	freebayes,mutect,vardict	5	137801633	137801634
MM99	NRAS	C	T	c.34G>A	p.Gly12Ser	0.22478992	freebayes,mutect,vardict,platypus	1	115258747	115258748
KMS12	TP53	C	A	c.1010G>T	p.Arg337Leu	1	freebayes,mutect,vardict,platypus	17	7574016	7574017
KMS12	FGFR3	C	T	c.472C>T	p.Arg158Trp	0.2173913	freebayes,mutect,vardict,platypus	4	1803119	1803120
KMS12	LTB	C	G	c.244G>C	p.Asp82His	0.31392404	freebayes,mutect,vardict	6	31549371	31549372
KMS12	LTB	T	C	c.218A>G	p.Lys73Arg	0.3159664	freebayes,mutect,vardict	6	31549397	31549398
KMS12	ACTG1	CC	AT	c.363+58_363+59delGGinsAT	NA	1	freebayes,vardict,platypus	17	79478869	79478871
KMS12	LTB	C	G	c.208+28G>C	NA	0.33190271	freebayes,mutect,vardict	6	31549562	31549563
MM1S	FAM46C	A	G	c.808A>G	p.Met270Val	0.99590498	freebayes,mutect,platypus	1	118166297	118166298
MM1S	KRAS	C	G	c.35G>C	p.Gly12Ala	0.49286199	freebayes,mutect,vardict	12	25398283	25398284
MM1S	EGFR	G	C	c.2749G>C	p.Gly917Arg	0.40333092	freebayes,mutect,vardict	7	55266456	55266457
MM1S	TRAF3	GTCTTTGTGGCCCAAACTGTTCTAGAAA	G	c.1607_1633delTCTTTGTGGCCCAAACTGTTCTAGAAA	p.Val536_Asn545delinsAsp	0.9982332	freebayes,vardict,scalpel,pindel	14	103372019	103372047
MM1S	LTB	T	C	c.239A>G	p.Glu80Gly	0.52457958	freebayes,mutect,vardict	6	31549376	31549377
MM1S	TRAF2	C	T	c.311C>T	p.Pro104Leu	0.19272369	freebayes,mutect,vardict	9	139794916	139794917
MM1S	SP140	A	G	c.2361+27A>G	NA	0.25172412	freebayes,mutect,vardict	2	231175972	231175973
NCI	NRAS	C	T	c.38G>A	p.Gly13Asp	0.29967427	freebayes,mutect,vardict,platypus	1	115258743	115258744
NCI	ACTG1	G	A	c.94C>T	p.Pro32Ser	0.11787073	freebayes,mutect	17	79479286	79479287
NCI	NRAS	C	A	c.521G>T	p.Ser174Ile	0.17344174	freebayes,mutect,vardict,platypus	1	115251204	115251205
NCI	ATM	G	T	c.2893G>T	p.Asp965Tyr	0.10700637	freebayes,mutect,vardict,platypus	11	108141844	108141845
NCI	STAT3	C	A	c.951G>T	p.Met317Ile	0.17848411	freebayes,mutect,vardict,platypus	17	40485913	40485914
NCI	PRDM1	C	G	c.1748C>G	p.Thr583Ser	0.13084112	freebayes,mutect,vardict	6	106553782	106553783
NCI	PRDM1	TG	T	c.1903-4delG	NA	0.12032086	freebayes,vardict,scalpel,platypus,pindel	6	106554780	106554782
NCI	FAM46C	T	TC	c.278_279insC	p.Ile94fs	0.96825397	freebayes,vardict,platypus,pindel	1	118165767	118165768
NCI	ATM	G	T	c.4518G>T	p.Val1506Val	0.13600001	freebayes,mutect,vardict,platypus	11	108163426	108163427
NCI	ACTG1	G	T	c.111C>A	p.Arg37Arg	0.10266159	freebayes,mutect	17	79479269	79479270
NCI	RB1	AT	A	c.2326-35delT	NA	0.2857143	freebayes,vardict,scalpel,platypus,pindel	13	49039303	49039305
RPMI	KRAS	C	G	c.35G>C	p.Gly12Ala	0.79512894	freebayes,mutect,vardict,platypus	12	25398283	25398284
RPMI	TP53	C	T	c.853G>A	p.Glu285Lys	0.99095023	freebayes,mutect,vardict,platypus	17	7577084	7577085
RPMI	EGFR	C	T	c.2252C>T	p.Thr751Ile	0.15830721	freebayes,mutect,vardict,platypus	7	55242481	55242482

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
RPMI	DIS3	T	A	c.229-3A>T	NA	0.3018868	freebayes,mutect,platypus	13	73355143	73355144
RPMI	IRF4	C	CT	c.661dupT	p.Tyr221fs	0.12026726	freebayes,vardict,scalpel,platypus,pindel	6	398846	398847
RPMI	MAGED1	GC	G	c.1363delC	p.Gln455fs	0.24870466	freebayes,vardict,scalpel,pindel	X	51639944	51639946
RPMI	MAGED1	A	C	c.1831A>C	p.Lys611Gln	0.11764706	freebayes,mutect	X	51641213	51641214
RPMI	MAGED1	C	T	c.1377C>T	p.Asp459Asp	0.22522523	freebayes,mutect,vardict	X	51639959	51639960
RPMI	MAGED1	A	G	c.1833A>G	p.Lys611Lys	0.14074074	freebayes,mutect,platypus	X	51641215	51641216
MM106	TP53	C	A	c.404G>T	p.Cys135Phe	0.97000003	freebayes,mutect,vardict	17	7578525	7578526
MM11	FGFR3	C	T	c.1146C>T	p.Gly382Gly	0.46245059	freebayes,mutect,vardict	4	1806126	1806127
13	DIS3	G	A	c.2458C>T	p.Arg820Trp	0.73864484	freebayes,mutect,vardict	13	73335836	73335837
21	TRAF3	C	T	c.690C>T	p.Ser230Ser	0.57317072	freebayes,mutect,vardict	14	103355934	103355935
23	TNFRSF21	C	T	c.184G>A	p.Gly62Ser	0.50046337	freebayes,mutect,vardict	6	47254243	47254244
25	CYLD	C	T	c.1592C>T	p.Ala531Val	0.2820513	freebayes,mutect,vardict	16	50815229	50815230
25	STAT3	C	T	c.645+1G>A	NA	0.10084034	freebayes,mutect,vardict	17	40489779	40489780
25	DIS3	A	G	c.2343-15T>C	NA	0.17391305	freebayes,mutect,vardict	13	73335966	73335967
25	IRF4	G	A	c.291G>A	p.Leu97Leu	0.13592233	freebayes,mutect,vardict	6	394894	394895
27	ATM	G	C	c.6286G>C	p.Glu2096Gln	0.20111732	freebayes,mutect,vardict,platypus	11	108188186	108188187
32	STAT3	C	T	c.1233+31G>A	NA	0.41090909	freebayes,mutect,vardict	17	40481540	40481541
35	KRAS	C	G	c.37G>C	p.Gly13Arg	0.48571429	freebayes,mutect,vardict,platypus	12	25398281	25398282
35	BRAF	T	A	c.443A>T	p.Asn148Ile	0.14	freebayes,mutect,vardict	7	140534469	140534470
37	FAM46C	T	G	c.1128T>G	p.Pro376Pro	0.1147541	freebayes,mutect,platypus	1	118166617	118166618
39	EGR1	A	T	c.184A>T	p.Ser62Cys	0.45805368	freebayes,mutect,vardict	5	137801633	137801634
43	KRAS	T	C	c.182A>G	p.Gln61Arg	0.46074075	freebayes,mutect,vardict	12	25380275	25380276
46	TNFRSF21	C	A	c.609G>T	p.Gly203Gly	0.51346272	freebayes,mutect,vardict,platypus	6	47253818	47253819
48	ATM	A	T	c.901+10A>T	NA	0.10080645	freebayes,mutect,vardict	11	108115762	108115763
52	STAT3	C	T	c.1233+31G>A	NA	0.42703232	freebayes,mutect,vardict	17	40481540	40481541
53	CYLD	A	T	c.1638A>T	p.Ala546Ala	0.22608696	freebayes,mutect,vardict,platypus	16	50815275	50815276
53	LTB	T	A	c.208+74A>T	NA	0.26477271	freebayes,mutect,platypus	6	31549516	31549517
55	ATM	A	T	c.5558A>T	p.Asp1853Val	0.61395693	freebayes,mutect,vardict,platypus	11	108175462	108175463
MM71	EGR1	A	ACAG	c.199_201dupAGC	p.Ser67dup	0.35930735	freebayes,vardict,scalpel,platypus,pindel	5	137801631	137801632
MM73	TP53	C	T	c.993+141G>A	NA	0.49607071	freebayes,mutect,vardict	17	7576711	7576712
MM78	DIS3	T	C	c.120A>G	p.Gly40Gly	0.71769387	freebayes,mutect,vardict	13	73355850	73355851
MM82	LTB	A	T	c.209-6T>A	NA	0.41329479	freebayes,mutect,vardict	6	31549412	31549413
MM82	LTB	G	C	c.209-17C>G	NA	0.41244572	freebayes,mutect,vardict	6	31549423	31549424
MM89	FGFR3	T	TCACCCCG	c.2275-20_2275-14dupACCCCGC	NA	0.51182199	freebayes,vardict,scalpel,pindel	4	1808820	1808821
MM92	ATM	G	A	c.7742G>A	p.Ser2581Asn	0.10824742	freebayes,mutect,vardict	11	108202717	108202718
MM92	SP140	C	A	c.888C>A	p.Asp296Glu	0.1388889	freebayes,mutect,vardict	2	231112775	231112776

SAMPLE	GENE	REF	ALT	CODON	AA	MAX CALLER SOMATIC VAF	CALLERS	CHROM	START	END
MM92	LTB	C	T	c.729G>A	p.Val243Val	0.16571428	freebayes,mutect,vardict	6	31548491	31548492
NCI	ATM	CA	C	c.4307delA	p.His1436fs	0.26839826	freebayes,vardict,scalpel,platypus,pindel	11	108160397	108160399
NCI	DIS3	T	C	c.1606A>G	p.Arg536Gly	0.2364341	freebayes,mutect,vardict,platypus	13	73345282	73345283
NCI	SP140	T	C	c.480T>C	p.Tyr160Tyr	0.30357143	freebayes,mutect,platypus	2	231106191	231106192
NCI	TNFRSF21	G	A	c.1632C>T	p.Asp544Asp	0.10695187	freebayes,mutect,vardict	6	47202511	47202512
RPMI	LTB	CC	TT	c.208_208+1delGGinsAA	p.Gly70Arg	0.64440078	freebayes,vardict,platypus	6	31549589	31549591
RPMI	LTB	CC	TT	c.208_208+1delGGinsAA	p.Gly70Arg	0.64440078	freebayes,vardict,platypus	6	31549589	31549591
43	FAM46C	GC	TT	c.678_679delGCinsTT	p.LeuGln226*	0.44146901	freebayes,vardict	1	118166167	118166169

Appendix Table 3: Caller for each interchromosomal translocation

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM08	12	14:103067474-103067479;20:40318453-40318458	Y	N	N	Y
MM08	12	14:106164864-106164867;8:129238551-129238554	Y	Y	N	Y
MM08	12	14:106164866-106164869;8:129238551-129238554	Y	Y	N	Y
MM08	12	14:103757101-103757153;16:75490080-75490132	Y	N	N	Y
MM08	11	14:103067474-103067479;20:40318453-40318458	Y	N	N	Y
MM08	11	14:106164864-106164867;8:129238551-129238554	Y	N	N	Y
MM08	10	14:106164864-106164867;8:129238551-129238554	Y	Y	N	Y
MM08	10	14:106164866-106164869;8:129238551-129238554	Y	Y	N	Y
MM08	9	14:106164864-106164867;8:129238551-129238554	Y	Y	N	Y
MM08	9	14:106164866-106164869;8:129238551-129238554	Y	Y	N	Y
MM08	8	14:106164866-106164869;8:129238551-129238554	Y	Y	N	N
MM08	12	3:111274085-111274086;8:128533829-128533830	Y	N	N	N
MM12	12	14:96127041-96127123;4:2080373-2080454	Y	N	N	Y
MM12	12	14:106326387-106326515;4:1893350-1893446	Y	N	Y	Y
MM12	12	14:106326390-106326483;4:1893318-1893442	Y	N	Y	Y
MM12	12	14:106326397-106326631;4:1893360-1893559	Y	N	Y	Y
MM12	11	14:96127041-96127123;4:2080373-2080454	Y	N	N	Y
MM12	11	14:106326387-106326515;4:1893350-1893446	Y	N	Y	Y
MM12	11	14:106326390-106326483;4:1893318-1893442	Y	N	Y	Y
MM12	11	14:106326397-106326631;4:1893360-1893559	Y	N	Y	Y
MM12	10	14:96127040-96127123;4:2080373-2080455	Y	N	N	Y
MM12	10	14:106326387-106326515;4:1893350-1893446	Y	N	Y	Y
MM12	10	14:106326390-106326483;4:1893318-1893443	Y	N	Y	Y
MM12	10	14:106326397-106326631;4:1893360-1893559	Y	N	Y	Y
MM12	9	14:96127041-96127123;4:2080373-2080454	Y	N	N	Y
MM12	12	1:164240435-164240436;2:190999444-190999445	Y	Y	N	Y
MM12	12	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	12	1:168186488-168186489;3:53175884-53175885	Y	Y	N	Y
MM12	12	3:111274085-111274086;8:128533829-128533830	Y	Y	N	Y
MM12	12	11:38812657-38812662;8:52731477-52731482	Y	Y	Y	Y
MM12	12	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y
MM12	12	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	12	20:11281583-11281584;X:81651234-81651235	Y	Y	N	Y
MM12	11	1:164240435-164240436;2:190999444-190999445	Y	Y	N	Y

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM12	11	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	11	1:168186488-168186489;3:53175884-53175885	Y	Y	N	Y
MM12	11	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM12	11	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y
MM12	11	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	11	20:11281583-11281584;X:81651234-81651235	Y	Y	N	Y
MM12	10	1:164240435-164240436;2:190999444-190999445	Y	Y	N	Y
MM12	10	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	10	1:168186488-168186489;3:53175884-53175885	Y	Y	N	Y
MM12	10	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM12	10	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y
MM12	9	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	9	1:168186488-168186489;3:53175884-53175885	Y	Y	N	Y
MM12	9	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM12	8	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	8	1:168186488-168186489;3:53175884-53175885	Y	Y	N	Y
MM12	8	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM12	7	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	6	1:200067551-200067555;2:23698252-23698256	Y	Y	N	Y
MM12	12	1:230006665-230006666;7:157274846-157274847	Y	N	Y	Y
MM12	12	2:120451576-120451577;21:40285759-40285760	Y	N	N	Y
MM12	12	3:109751756-109751757;6:78649420-78649421	Y	N	Y	Y
MM12	12	12:66451371-66451387;15:55218263-55218279	Y	N	N	Y
MM12	12	13:74314055-74314062;8:15289357-15289364	Y	N	N	Y
MM12	12	14:37769601-37769602;16:61079284-61079285	Y	N	N	Y
MM12	12	15:40854179-40854180;7:26241364-26241365	Y	N	N	Y
MM12	12	15:40854189-40854190;7:26252970-26252971	Y	N	Y	Y
MM12	11	2:120451576-120451577;21:40285759-40285760	Y	N	N	Y
MM12	11	3:109751756-109751757;6:78649420-78649421	Y	N	Y	Y
MM12	11	12:66451370-66451371;15:55218279-55218280	Y	N	N	Y
MM12	11	14:37769601-37769602;16:61079284-61079285	Y	N	N	Y
MM12	11	15:40854179-40854180;7:26241364-26241365	Y	N	N	Y
MM12	11	15:40854189-40854190;7:26252970-26252971	Y	N	Y	Y
MM12	10	12:66451370-66451371;15:55218279-55218280	Y	N	N	Y
MM12	12	15:31575815-31575820;8:128747170-128747175	Y	Y	N	Y

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM12	11	15:31575819-31575824;8:128747170-128747175	Y	Y	N	N
MM12	10	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	10	15:31575819-31575824;8:128747170-128747175	Y	Y	N	N
MM12	9	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	9	15:31575819-31575824;8:128747170-128747175	Y	Y	N	N
MM12	8	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	7	15:31510529-31510534;8:128803125-128803130	Y	Y	N	Y
MM12	6	15:31510529-31510534;8:128803125-128803130	Y	Y	N	N
MM12	11	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM12	10	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM12	9	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM1S	12	14:106324759-106324761;16:78802777-78802779	Y	N	N	Y
MM1S	12	14:105953824-105953828;8:129711115-129711119	Y	N	N	N
MM1S	11	14:106324759-106324761;16:78802777-78802779	Y	N	N	Y
MM1S	10	14:106324759-106324761;16:78802777-78802779	Y	N	N	N
MM1S	9	14:106324759-106324761;16:78802777-78802779	Y	N	N	N
MM1S	12	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM1S	12	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y
MM1S	12	20:11281583-11281584;X:81651234-81651235	Y	Y	N	Y
MM1S	12	1:230006665-230006666;7:157274846-157274847	Y	N	Y	Y
MM1S	11	1:230006665-230006666;7:157274846-157274847	Y	N	Y	Y
MM1S	11	11:38812657-38812662;8:52731477-52731482	Y	N	N	Y
MM1S	10	1:230006665-230006666;7:157274846-157274847	Y	N	Y	Y
MM29	12	1:109494857-109494858;3:110413393-110413394	Y	Y	N	Y
MM29	12	3:151148543-151148544;5:39787750-39787751	Y	Y	N	Y
MM29	12	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	12	11:38812668-38812669;8:52730142-52730143	Y	Y	Y	Y
MM29	12	12:86120304-86120305;3:49441527-49441528	Y	Y	Y	Y
MM29	12	12:56990095-56990096;15:39994624-39994625	Y	Y	N	Y
MM29	12	13:74313857-74313862;8:15289366-15289371	Y	Y	Y	Y
MM29	12	17:7167889-7167890;8:30145622-30145623	Y	Y	Y	Y
MM29	11	1:109494857-109494858;3:110413393-110413394	Y	Y	N	Y
MM29	11	3:151148543-151148544;5:39787750-39787751	Y	Y	N	Y
MM29	11	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	11	12:86120304-86120305;3:49441527-49441528	Y	Y	N	Y

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM29	11	12:56990095-56990096;15:39994624-39994625	Y	Y	N	Y
MM29	11	13:74313857-74313862;8:15289366-15289371	Y	Y	Y	Y
MM29	11	17:7167889-7167890;8:30145622-30145623	Y	Y	Y	Y
MM29	10	1:109494857-109494858;3:110413393-110413394	Y	Y	N	Y
MM29	10	3:151148543-151148544;5:39787750-39787751	Y	Y	N	Y
MM29	10	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	10	13:74313857-74313862;8:15289366-15289371	Y	Y	Y	Y
MM29	10	17:7167889-7167890;8:30145622-30145623	Y	Y	Y	Y
MM29	9	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	9	13:74313857-74313862;8:15289366-15289371	Y	Y	N	Y
MM29	9	17:7167889-7167890;8:30145622-30145623	Y	Y	Y	Y
MM29	8	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	8	13:74313857-74313862;8:15289366-15289371	Y	Y	N	Y
MM29	8	17:7167889-7167890;8:30145622-30145623	Y	Y	Y	Y
MM29	7	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	6	11:85211619-85211624;7:107829248-107829253	Y	Y	Y	Y
MM29	12	2:120451576-120451577;21:40285759-40285760	Y	N	N	Y
MM29	12	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM29	11	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM29	10	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM29	9	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM29	8	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM29	7	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM30	12	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	12	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	12	14:100231275-100231276;6:41971278-41971279	Y	N	N	N
MM30	11	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	11	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	11	14:100231275-100231276;6:41971278-41971279	Y	N	N	N
MM30	10	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	10	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	10	14:100231275-100231276;6:41971278-41971279	Y	N	N	N
MM30	9	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	9	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	9	14:100231275-100231276;6:41971278-41971279	Y	N	N	N

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM30	8	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	8	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	7	14:106325600-106325601;16:78762503-78762504	Y	N	N	N
MM30	7	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	6	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	5	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	4	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	3	14:106325724-106325725;16:78463187-78463188	Y	N	N	N
MM30	12	17:7167958-7167959;8:30145401-30145402	Y	Y	Y	Y
MM30	11	17:7167958-7167959;8:30145401-30145402	Y	Y	Y	Y
MM30	10	17:7167958-7167959;8:30145401-30145402	Y	Y	Y	Y
MM30	12	14:37769601-37769602;16:61079284-61079285	Y	N	N	Y
MM30	12	16:19332991-19332992;3:25939093-25939094	Y	N	Y	Y
MM30	11	14:37769601-37769602;16:61079284-61079285	Y	N	N	Y
MM30	11	16:19332991-19332992;3:25939093-25939094	Y	N	Y	Y
MM30	10	14:37769601-37769602;16:61079284-61079285	Y	N	N	Y
MM40	12	14:105231547-105231629;16:82329692-82329774	Y	N	N	Y
MM40	12	14:106238517-106238519;16:78884676-78884678	Y	N	N	N
MM40	11	14:105231546-105231629;16:82329691-82329774	Y	N	N	Y
MM40	11	14:106238517-106238519;16:78884676-78884678	Y	N	N	N
MM40	10	14:105231546-105231629;16:82329691-82329774	Y	N	N	Y
MM40	10	14:106238517-106238519;16:78884676-78884678	Y	N	N	N
MM40	12	1:215060384-215060388;X:123611800-123611804	Y	Y	N	Y
MM40	12	11:61841812-61841813;14:81786773-81786774	Y	Y	N	Y
MM40	12	12:74014366-74014367;7:125264219-125264220	Y	Y	N	Y
MM40	12	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM40	11	11:61841812-61841813;14:81786773-81786774	Y	Y	N	Y
MM40	11	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM40	10	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM40	9	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM40	8	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM40	7	12:108203258-108203261;7:111053751-111053754	Y	Y	N	Y
MM46	12	11:69446045-69446047;14:106327421-106327423	Y	Y	N	N
MM46	12	11:69446047-69446048;14:106327422-106327423	Y	Y	N	N
MM46	12	11:69480712-69480717;14:104613880-104613885	Y	N	N	N

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM46	11	11:69446045-69446047;14:106327421-106327423	Y	N	N	N
MM46	11	11:69480712-69480717;14:104613880-104613885	Y	N	N	N
MM46	10	11:69446045-69446047;14:106327421-106327423	Y	N	N	N
MM46	10	11:69480712-69480717;14:104613880-104613885	Y	N	N	N
MM46	9	11:69480712-69480717;14:104613880-104613885	Y	N	N	N
MM46	12	1:109495132-109495133;3:110413394-110413395	Y	Y	N	Y
MM46	12	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM46	11	1:109495132-109495133;3:110413394-110413395	Y	Y	N	Y
MM46	11	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM46	10	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM68	12	11:69260088-69260374;14:106113303-106113625	Y	Y	Y	Y
MM68	12	11:69260289-69260429;14:106113295-106113390	Y	Y	Y	Y
MM68	12	11:69260364-69260365;14:106113314-106113315	Y	Y	Y	Y
MM68	12	11:70188510-70188513;14:105158593-105158596	Y	N	N	N
MM68	11	11:69260088-69260374;14:106113303-106113621	Y	Y	Y	Y
MM68	11	11:69260289-69260429;14:106113295-106113390	Y	Y	Y	Y
MM68	11	11:69260364-69260365;14:106113314-106113315	Y	Y	Y	Y
MM68	10	11:69260088-69260374;14:106113303-106113626	Y	Y	Y	Y
MM68	10	11:69260289-69260429;14:106113295-106113390	Y	Y	Y	Y
MM68	10	11:69260364-69260365;14:106113314-106113315	Y	Y	Y	Y
MM68	9	11:69260088-69260374;14:106113303-106113626	Y	Y	Y	Y
MM68	9	11:69260290-69260429;14:106113295-106113389	Y	Y	Y	Y
MM68	9	11:69260364-69260365;14:106113314-106113315	Y	Y	Y	Y
MM68	8	11:69260089-69260374;14:106113303-106113619	Y	N	Y	Y
MM68	8	11:69260290-69260429;14:106113295-106113389	Y	N	Y	Y
MM68	8	11:69260364-69260365;14:106113314-106113315	Y	N	Y	Y
MM68	7	11:69260089-69260374;14:106113303-106113674	Y	N	Y	Y
MM68	7	11:69260290-69260429;14:106113295-106113389	Y	N	Y	Y
MM68	7	11:69260364-69260365;14:106113314-106113315	Y	N	Y	Y
MM68	6	11:69260089-69260374;14:106113303-106113677	Y	N	Y	Y
MM68	6	11:69260290-69260429;14:106113295-106113389	Y	N	Y	Y
MM68	6	11:69260364-69260365;14:106113314-106113315	Y	N	Y	Y
MM68	5	11:69260364-69260365;14:106113314-106113315	Y	N	N	N
MM68	4	11:69260364-69260365;14:106113314-106113315	Y	N	N	N
MM68	3	11:69260364-69260365;14:106113314-106113315	Y	N	N	N

SAMPLES	SUBSAMPLES	CALLS	GRIDSS	SVABA	LUMPY	MANTA
MM68	2	11:69260364-69260365;14:106113314-106113315	Y	N	N	N
MM68	12	3:111274085-111274086;8:128533829-128533830	Y	Y	N	Y
MM68	12	3:151148543-151148544;5:39787750-39787751	Y	Y	Y	Y
MM68	12	11:38812668-38812669;8:52730142-52730143	Y	Y	Y	Y
MM68	11	3:111274085-111274086;8:128533829-128533830	Y	Y	N	Y
MM68	11	11:38812668-38812669;8:52730142-52730143	Y	Y	Y	Y
MM68	10	11:38812668-38812669;8:52730142-52730143	Y	Y	N	Y
MM68	9	11:38812668-38812669;8:52730142-52730143	Y	Y	N	Y
MM68	9	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM68	10	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM68	8	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM68	7	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM68	6	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM68	5	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM75	12	3:111274085-111274086;8:128533829-128533830	Y	Y	Y	Y
MM75	12	11:38812657-38812662;8:52731477-52731482	Y	Y	N	Y
MM75	11	3:111274085-111274086;8:128533829-128533830	Y	Y	Y	Y
MM75	10	3:111274085-111274086;8:128533829-128533830	Y	Y	Y	Y
MM75	12	20:11281583-11281584;X:81651234-81651235	Y	N	N	Y
MM75	11	20:11281583-11281584;X:81651234-81651235	Y	N	N	Y
MM75	10	20:11281583-11281584;X:81651234-81651235	Y	N	N	Y
MM75	9	3:111274085-111274086;8:128533829-128533830	Y	N	Y	Y
MM75	8	3:111274085-111274086;8:128533829-128533830	Y	N	Y	Y
MM75	7	3:111274085-111274086;8:128533829-128533830	Y	N	Y	Y
MM75	6	3:111274085-111274086;8:128533829-128533830	Y	N	N	Y
MM97	12	14:104938387-104938470;20:35025056-35025139	Y	N	N	Y
MM97	12	14:104938470-104938471;20:35025042-35025043	Y	N	N	Y
MM97	11	14:104938387-104938470;20:35025056-35025139	Y	N	N	Y
MM97	11	14:104938470-104938471;20:35025042-35025043	Y	N	N	Y
MM97	10	14:104938470-104938471;20:35025042-35025043	Y	N	N	N
MM97	9	14:104938470-104938471;20:35025042-35025043	Y	N	N	N
MM97	12	12:56989715-56989716;15:39994636-39994637	Y	Y	Y	Y
MM97	11	12:56989715-56989716;15:39994636-39994637	Y	Y	Y	Y
MM97	10	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y
MM97	9	12:56990095-56990096;15:39994624-39994625	Y	Y	Y	Y

Appendix Table 4: Filter passing intrachromosomal translocation calls at 12X-1X coverage

SAMPLE	CALL: CHROM1	CALL: START1	CALL: END1	CALL: CHROM2	CALL: START2	CALL: END2	CALL: CYTOBAND	FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
MM08	1	202593712	202593719	15	30465294	30465301	t(1;15) (q32.1;q13.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM08	5	37709719	37709720	7	8663313	8663314	t(5;7) (p13.2;p21.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	11	61841812	61841813	14	81786773	81786774	t(11;14) (q12.3;q31.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	11	85211619	85211624	7	107829248	107829253	t(11;7) (q14.1;q31.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	12	56989715	56989716	15	39994636	39994637	t(12;15) (q13.3;q14)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	12	108203263	108203264	7	111053152	111053153	t(12;7) (q23.3;q31.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM08	14	106164864	106164867	8	129238551	129238554	t(14;8) (q32.33;q24.21)	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N
MM12	1	109494857	109494858	3	110413393	110413394	t(1;3) (p13.3;q13.13)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM12	1	109619760	109619763	16	4829176	4829179	t(1;16) (p13.3;p13.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM12	1	164240435	164240436	2	190999444	190999445	t(1;2) (q23.3;q32.2)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM12	1	168186488	168186489	3	53175884	53175885	t(1;3) (q24.2;p21.1)	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N
MM12	1	200067551	200067555	2	23698252	23698256	t(1;2) (q32.1;p24.1)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
MM12	1	230006665	230006666	7	157274846	157274847	t(1;7) (q42.13;q36.3)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM12	2	120451576	120451577	21	40285759	40285760	t(2;21) (q14.2;q22.2)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM12	3	109751756	109751757	6	78649420	78649421	t(3;6) (q13.13;q14.1)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM12	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM12	3	187728893	187728896	6	93097205	93097208	t(3;6) (q27.3;q15)	Y	N	N	N	N	N	N	N	N	N	N	N	N

SAMPLE	CALL CHROM1	START1	END1	CHROM2	START2	END2	CYTOBAND	FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
MM12	11	38812657	38812662	8	52731477	52731482	t(11;8) (p12;q11.23)	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N
MM12	12	56990095	56990096	15	39994624	39994625	t(12;15) (q13.3;q14)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM12	12	66451371	66451387	15	55218263	55218279	t(12;15) (q14.3;q21.3)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM12	13	74314055	74314062	8	15289357	15289364	t(13;8) (q22.1;p22)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM12	14	37769601	37769602	16	61079284	61079285	t(14;16) (q13.3;q21)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM12	14	65447290	65447291	2	194545043	194545044	t(14;2) (q23.3;q32.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM12	14	106326505	106326506	4	1893403	1893404	t(14;4) (q32.33;p16.3)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM12	15	31510529	31510534	8	128803125	128803130	t(15;8) (q13.3;q24.21)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
MM12	15	31575815	31575820	8	128747170	128747175	t(15;8) (q13.3;q24.21)	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM12	15	40854179	40854180	7	26241364	26241365	t(15;7) (q15.1;p15.2)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM12	15	40854189	40854190	7	26252970	26252971	t(15;7) (q15.1;p15.2)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM12	15	63770026	63770027	9	38243843	38243844	t(15;9) (q22.31;p13.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM12	20	11281583	11281584	X	81651234	81651235	t(20;X) (p12.2;q21.1)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM1S	1	38241382	38241383	19	11422015	11422016	t(1;19) (p34.3;p13.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM1S	1	230006665	230006666	7	157274846	157274847	t(1;7) (q42.13;q36.3)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM1S	3	8694997	8694998	8	116902459	116902460	t(3;8) (p26.1;q23.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM1S	11	38812657	38812662	8	52731477	52731482	t(11;8) (p12;q11.23)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM1S	12	56990095	56990096	15	39994624	39994625	t(12;15) (q13.3;q14)	Y	Y	N	N	N	N	N	N	N	N	N	N	N

SAMPLE	CALL							FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
	CHROM1	START1	END1	CHROM2	START2	END2	CYTOBAND													
MM1S	12	74014366	74014367	7	125264219	125264220	t(12;7) (q21.1;q31.33)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM1S	14	105953824	105953828	8	129711115	129711119	t(14;8) (q32.33;q24.21)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM1S	14	106324759	106324761	16	78802777	78802779	t(14;16) (q32.33;q23.1)	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM1S	20	11281583	11281584	X	81651234	81651235	t(20;X) (p12.2;q21.1)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM29	1	109494857	109494858	3	110413393	110413394	t(1;3) (p13.3;q13.13)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM29	2	120451576	120451577	21	40285759	40285760	t(2;21) (q14.2;q22.2)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM29	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
MM29	3	151148543	151148544	5	39787750	39787751	t(3;5) (q25.1;p13.1)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM29	6	107307833	107307835	8	105231727	105231729	t(6;8) (q21;q22.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM29	11	38812657	38812662	8	52731477	52731482	t(11;8) (p12;q11.23)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM29	11	38812668	38812669	8	52730142	52730143	t(11;8) (p12;q11.23)	N	Y	N	N	N	N	N	N	N	N	N	N	N
MM29	11	85211619	85211624	7	107829248	107829253	t(11;7) (q14.1;q31.1)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
MM29	12	56990095	56990096	15	39994624	39994625	t(12;15) (q13.3;q14)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM29	12	86120304	86120305	3	49441527	49441528	t(12;3) (q21.31;p21.31)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM29	12	130027735	130027736	13	113627112	113627113	t(12;13) (q24.33;q34)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM29	13	74313857	74313862	8	15289366	15289371	t(13;8) (q22.1;p22)	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N
MM29	14	98250172	98250173	16	83662276	83662277	t(14;16) (q32.2;q23.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM29	15	40854179	40854180	7	26241364	26241365	t(15;7) (q15.1;p15.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM29	17	7167889	7167890	8	30145622	30145623	t(17;8) (p13.1;p12)	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N

SAMPLE	CALL CHROM1	START1	END1	CHROM2	START2	END2	CYTOBAND	FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
MM30	14	106325724	106325725	16	78463187	78463188	t(14;16) (q32.33;q23.1)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
MM30	16	19332991	19332992	3	25939093	25939094	t(16;3) (p12.3;p24.2)	N	Y	Y	N	N	N	N	N	N	N	N	N	N
MM30	17	7167958	7167959	8	30145401	30145402	t(17;8) (p13.1;p12)	N	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM40	1	164240435	164240436	2	190999444	190999445	t(1;2) (q23.3;q32.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM40	1	215060384	215060388	X	123611800	123611804	t(1;X) (q41;q25)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM40	3	5155530	5155532	8	138095714	138095716	t(3;8) (p26.1;q24.23)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM40	5	89450805	89450808	6	64244711	64244714	t(5;6) (q14.3;q12)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM40	11	20031529	20031533	2	48429373	48429377	t(11;2) (p15.1;p16.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM40	11	61841812	61841813	14	81786773	81786774	t(11;14) (q12.3;q31.1)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM40	12	56989715	56989716	15	39994636	39994637	t(12;15) (q13.3;q14)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM40	12	74014366	74014367	7	125264219	125264220	t(12;7) (q21.1;q31.33)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM40	12	108203258	108203261	7	111053751	111053754	t(12;7) (q23.3;q31.1)	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
MM40	14	106238517	106238519	16	78884676	78884678	t(14;16) (q32.33;q23.1)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM46	1	109495132	109495133	3	110413394	110413395	t(1;3) (p13.3;q13.13)	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
MM46	11	38812657	38812662	8	52731477	52731482	t(11;8) (p12;q11.23)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM46	11	69446045	69446047	14	106327421	106327423	t(11;14) (q13.3;q32.33)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N

SAMPLE	CALL: CHROM1	START1	END1	CHROM2	START2	END2	CYTOBAND	FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
MM46	11	69480712	69480717	14	104613880	104613885	t(11;14) (q13.3;q32.33)	N	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM68	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
MM68	3	151148543	151148544	5	39787750	39787751	t(3;5) (q25.1;p13.1)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM68	11	23038136	23038137	3	81055273	81055274	t(11;3) (p14.3;p12.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM68	11	38812668	38812669	8	52730142	52730143	t(11;8) (p12;q11.23)	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM68	11	69260364	69260365	14	106113314	106113315	t(11;14) (q13.3;q32.33)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
MM68	11	70188510	70188513	14	105158593	105158596	t(11;14) (q13.3;q32.33)	N	Y	N	N	N	N	N	N	N	N	N	N	N
MM68	12	74014366	74014367	7	125264219	125264220	t(12;7) (q21.1;q31.33)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	1	9121442	9121445	14	93713176	93713179	t(1;14) (p36.23;q32.12)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	1	109494857	109494858	3	110413393	110413394	t(1;3) (p13.3;q13.13)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	1	230006665	230006666	7	157274846	157274847	t(1;7) (q42.13;q36.3)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
MM75	10	128580643	128580644	7	26679501	26679502	t(10;7) (q26.2;p15.2)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	11	37429210	37429211	17	10138716	10138717	t(11;17) (p12;p13.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	11	38812657	38812662	8	52731477	52731482	t(11;8) (p12;q11.23)	Y	Y	N	N	N	N	N	N	N	N	N	N	N
MM75	12	56989715	56989716	15	39994636	39994637	t(12;15) (q13.3;q14)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	12	108203263	108203264	7	111053152	111053153	t(12;7) (q23.3;q31.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	14	106094315	106094317	8	128785265	128785267	t(14;8) (q32.33;q24.21)	Y	N	N	N	N	N	N	N	N	N	N	N	N

SAMPLE	CALL							FULL DEPTH CALL	12x CALL	11x CALL	10x CALL	9x CALL	8x CALL	7x CALL	6x CALL	5x CALL	4x CALL	3x CALL	2x CALL	1x CALL
	CHROM1	START1	END1	CHROM2	START2	END2	CYTOBAND													
MM75	16	89942703	89942704	9	107029986	107029987	t(16;9) (q24.3;q31.1)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM75	20	11281583	11281584	X	81651234	81651235	t(20;X) (p12.2;q21.1)	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
MM97	3	111274085	111274086	8	128533829	128533830	t(3;8) (q13.13;q24.21)	Y	N	N	N	N	N	N	N	N	N	N	N	N
MM97	12	56989715	56989716	15	39994636	39994637	t(12;15) (q13.3;q14)	N	Y	Y	Y	Y	N	N	N	N	N	N	N	N
MM97	14	104938470	104938471	20	35025042	35025043	t(14;20) (q32.33;q11.23)	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N