The genetic architecture of plasma cell disorders

Eileen Mary Boyle

N°d'ordre NNT : 2020LYSE1344

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 340
Biologie Moléculaire Intégrative Et Cellulaire

Spécialité de doctorat : Science de la vie
Discipline : Biologie moléculaire

Soutenue publiquement le 15/12/2020, par :

Eileen M. Boyle

Aspects moléculaires des dyscrasies plasmocytaires: intérêts diagnostiques, pronostiques et théranostiques

Devant le jury composé de :

Caers, Jo	Professeur, Université de Liège, Belgique	Examinateur
Corre, Jill	Maitre de Conférence, Université de Toulouse	Examinatrice
Dumontet, Charles	Professeur, Université de Lyon	Co-directeur de thèse
Ghesquières, Herve	Professeur, Université de Lyon	Examinateur
Gutiérrez, Norma C.	Professeur, University of Salamanca, Espagne	Examinatrice
Mohty, Mohamad	Professeur première classe, Université de Paris	Rapporteur
Thieblemont, Catherine	Professeur première classe, Université de Paris	Rapportrice
Walker, Brian A.	Professeur, Indiana University	Co-directeur de thèse

Acknowledgments

This work was carried out from May 2018 to October 2020 in the Myeloma Center at the University of Arkansas for Medical Science and at NYU-Langone's Perlmutter Cancer Centre.

I owe my deepest gratitude to my supervisor Professor Brian A. Walker, without his help, and support this work would hardly have been completed.

I am extremely grateful to Professor Gareth Morgan, and Professor Faith Davies, for their help and trust, and for making it possible to carry out this work in their departments.

I express my deepest thanks to Professor Charles Dumontet for accepting to supervise this work.

I am indebted to my many colleagues especially those involved in the sample processing (Shayu Deshpande and Ruslana Tytarenko), data management (Philip Farmer, Michael Rutherford, and Clyde Bailey) and bioinformatics (Christopher Wardell, Cody Ashby, Patrick Blaney, Yan Wang, and Michael Bauer). This work would not exist without you.

I would like to thank all of those who contributed to this work through rich discussions over the years: Sarah Johnson, Niels Weinhold, Leo Rasche, Katie Ryan, Carolina Schinke, Louis Williams, Marc Braunstein, James Stoeckle, Jessica Caro, Beatrice Razzo, Yubao Wang, Kylee Maclachlan, Even Rustad, Francesco Maura, Paula Proszek, Martin Kaiser, Aneta Mikulasova, David Cairns, Charlotte Pawlyn, Anna-Maria Brioli, Lauren Aronson, Daniel Itzak, Poppy Begum, Alex Murison, Nazrin Dahir, John Jones, Claire Bories, Rachel Litke, Jean-Baptiste Gibier, Morgane Nudel, Zoe Van de Wyngaert, Charles Herbaux, Mark Bustoros, Marie-Pierre Noel, Marie-Odile Petillon, Thierry Facon, and Bruno Quesnel.

I would also like to thank my friends, old and new, for their constant support and unwavering friendship: Anne-Charlotte Gagey, Nadège Gallet, Anahita Rouzé, Delphine Colling, Perrine Bortolotti, Marie Kauv, Claire Bories, Zoe Van De Wyngaert, Morgane Nudel, Sarah Barbieux, Eleonora Piccinini, and last but not least, Robert Daniello.

Finally, I would like to thank my family for their continuous and unparalleled love, help and support. I am grateful to my sisters and wonderful nieces and nephews for always being there for me. I am forever thankful to my parents for all the emphasis they put on education and family.

To my Father,

Funding

This work was funded by the Fédération Française de Recherche sur le myélome et les gammapathies.

The Walkerlab and Morganlab acknowledge continued support from Multiple Myeloma Research Foundation and the Perelman Family Foundation.

The Walkerlab and Morganlab received grant support through a Translational Research Program award from the Leukemia & Lymphoma Society (6020-20).

Publication

- **Part of this work was accepted for publication**

Boyle, E.M., Ashby, C., Tytarenko, R.G., Deshpande, S., Wang, H., Wang, Y., Rosenthal, A., Sawyer, J., Tian, E., Flynt, E., Hoering, A., Johnson, S.K., Rutherford, M.W., Wardell, C.P., Bauer, M.A., Dumontet, C., Facon, T., Thanendrarajan, S., Schinke, C.D., Zangari, M., et al (2020) *BRAF* and *DIS3* Mutations Associate with Adverse Outcome in a Long-term Follow-up of Patients with Multiple Myeloma. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 26, 2422–2432.

Boyle, E.M., Ashby, C., Wardell, C.P., Rowczenio, D., Sachchithanantham, S., Wang, Y., Johnson, S.K., Bauer, M.A., Weinhold, N., Kaiser, M.F., Johnson, D.C., Jones, J.R., Pawlyn, C., Proszek, P., Schinke, C., Facon, T., Dumontet, C., Davies, F.E., Morgan, G.J., Walker, B.A., et al (2018) The genomic landscape of plasma cells in systemic light chain amyloidosis. Blood, 132, 2775–2777.

Boyle, E.M., Deshpande, S., Ashby, C., Tytarenko, R.G., Wang, Y., Johnson, S.K., Wardell, C.P., Bauer, M.A., Dumontet, C., Facon, T., Schinke, C.D., Zangari, M., et al The Molecular Make Up of Smoldering Myeloma Highlights the Evolutionary Pathways Leading to Multiple Myeloma (Nat Comm, *in press*)

- **Part of this work was presented at the American Society of Hematology**

Boyle, E.M., Deshpande, S., Ashby, C., Tytarenko, R.G., Wang, Y., Johnson, S.K., Wardell, C.P., Bauer, M.A., Dumontet, C., Facon, T., Schinke, C.D., Zangari, M., et al The Molecular Make Up of Smoldering Myeloma Highlights the Evolutionary Pathways Leading to Multiple Myeloma, ASH 2018 San Diego (poster)

Boyle, E.M. Williams L, Blaney P, Ashby TC, Bauer M, Walker BA, Choi J, Wang Y, Caro J, Stoeckle J, Arbini A, Kaminetsky D, Braunstein M, Grossbard M, Razzo B, Maclachlan, Maura F, Landgren F, Litke R, Fegan C, Davies F, Morgan G. Telomere Maintenance and Aging Processes Impact the Outcome of Myeloma Patients, ASH 2020 (poster).

Résumé en Français : Aspects moléculaires des dyscrasies plasmocytaires: intérêts diagnostiques, pronostiques et théranostiques.

Les dyscrasies plasmocytaires regroupent un ensemble d'hémopathies lymphoïdes matures caractérisées par la production d'un composant monoclonal. Il s'agit d'un ensemble d'hémopathies hétérogènes regroupant des entités asymptomatiques telles que le myélome indolent (SMM) ou la gammapathie monoclonale de signification indéterminée (MGUS), et des entités aux symptômes divers allant du myélome multiple (MM) à l'amylose à chaine légère (AL). L'objectif de ce travail est de caractériser ces entités d'un point de vue moléculaire, en vue d'établir des similarités et des différences qui pourraient être utiles en pratique clinique.

A ces fins, un séquençage exomique complet de 24 AL au diagnostic a permis de mettre en évidence l'absence de spécificité, d'un point de vue mutationnel ou cytogénétique, de cette entité, qui s'inscrit donc dans le spectre complet des hémopathies plasmocytaires allant de la MGUS au MM symptomatique.

En vue de caractériser les différences entre le myélome symptomatique et indolent, un séquençage ciblé de 82 patients atteints de SMM et de 223 patients atteints de MM au diagnostic a permis de montrer que certaines anomalies, telles que la t(4 ;14), les réarrangements de *MYC*, les gain(1q) et les mutations de *NRAS* et *FAM46C,* sont moins fréquentes dans le myélome indolent que dans le MM symptomatique, suggérant qu'il s'agit soit d'anomalies associées à une courte phase asymptomatique, soit de marqueurs de transformation. La contribution d'APOBEC à la mutagénèse était également moindre dans les échantillons de SMM. L'ensemble de ces données est donc en faveur d'un modèle en deux temps, où l'acquisition d'évènements secondaires déstabilise la cellule tumorale qui peut alors proliférer, ce qui conduit au développement de symptômes cliniques. Il s'agit donc d'un phénomène dynamique, et l'analyse séquentielle de 53 échantillons médullaires, issus de 9 patients à différents moments du suivi, a permis de déterminer l'évolution de l'architecture clonale, ainsi que les facteurs favorisant la compétition clonale au cours du temps. L'ensemble des données collectées

a permis d'identifier des populations à risque de développer une forme symptomatique de la maladie, et pourrait être utilisé à des fins de diagnostic personnalisé précoce en clinique, s'appuyant sur des critères clinico- biologiques.

Enfin concernant le MM proprement dit, un séquençage ciblé de 223 patients jeunes au diagnostic a permis d'identifier des facteurs mutationnels pronostiques, tels que *BRAF* et *DIS3*. Chez le sujet âgé, l'analyse d'un groupe de 980 patients, séquencés au préalable, a mis en évidence un excès de chromothrypsis. Par ailleurs, ces données suggèrent l'existence d'une hématopoïèse clonale chez 10% des patients, qui augmente avec l'âge, mais ne conditionne pas le pronostic. Enfin, un lien potentiel entre longueur des télomères tumoraux et instabilité génomique, sous-tendant une survie défavorable chez le sujet âgé a pu être mise en évidence, confirmant ainsi la place de la génétique dans l'appréciation du pronostic de ces patients.

Au total, l'ensemble de ces données est en faveur de l'intérêt croissant pour la génomique dans les dyscrasies plasmocytaires. Ces données, en affinant le diagnostic et le pronostic des formes asymptomatiques, permettront d'optimiser la prise en charge des patients à des stades précoces, et constituent une base pour des études futures tant cliniques que fondamentales.

Mots clefs :

Myélome-Génétique- Séquençage- Myélome indolent-Amylose AL-Biologie Moléculaires

Résumé en Anglais: The genetic architecture of plasma cell disorders.

The clinical presentation of plasma cell disorders is heterogeneous ranging from asymptomatic conditions such as Monoclonal Gammopathy of Undetermined Significance (MGUS) or Smouldering Multiple Myeloma (SMM) to symptomatic form of diseases such as Multiple Myeloma (MM) or light chain Amyloidosis (AL). This work aims at understanding the genetic basis of some of these clinical differences.

To achieve this, 24 CD138+AL samples were sequenced. These data identified a genetic make-up of AL, not only in terms of copy number abnormalities and translocations but also mutations, that is similar to other plasma cell disorders suggesting the phenotype differences are not related to the genetic architecture of the underlying plasma cell disorder. To identify the molecular keys of progression SMM) we studied 82 patients with targeted sequencing. We compared these results to MM patients. We found fewer *NRAS* and *FAM46C* mutations together with fewer adverse translocations, del(1p), del(14q), del(16q), and del(17p) in SMM consistent with their role as drivers of the transition to MM. SMM samples also had fewer APOBEC mutations. In a unique analysis of change in clonal structure over time, we studied 53 samples from nine patients at multiple time points. Branching evolutionary patterns, novel mutations, biallelic hits in crucial tumour suppressor genes, and segmental copy number changes are key mechanisms underlying the transition to MM, which can precede progression and may be used to guide early intervention strategies. When considering MM, we show that Double-Hit, *BRAF* and *DIS3* mutations had an impact on outcome alongside classical risk factors in the context of an intensive treatment approach. In the elderly population, using previously published cases, we show an excess of chromothrypsis. Finally, we show a correlation between DNA instability, tumour telomere length and outcome suggesting a role for genetic markers in the elderly.

Overall, these data, highlight the importance of genomic in our understanding of plasma cell disorders and set the basis for future research strategies.

Mots clefs: Myeloma- Genetic- Sequencing-Smouldering Myeloma -Amyloidosis

Address

INSERM 1052/CNRS 5286
Cancer Research Center of Lyon
Faculté Rockefeller, 4th floor, stairs B
8 avenue Rockefeller
69008 Lyon
France

University of Arkansas for Medical Sciences
Myeloma Institute
4301 W Markham St,
Little Rock, AR 72205
United States of America

Table of Contents

List of abbreviations

Ab: antibody

ABCC2: ATP-binding cassette, subfamily C, member 2

ACD: adrenocortical dysplasia homolog

ACS: antibody secreting cells

ADP: Adenosine 5′-diphosphate

AIC: Akaike information criterion

AID: Activation-induced cytidine deaminase

AKT: Protein kinase B

AL: light chain amyloidosis

ALL: acute lymphoblastic leukemia

ALMS1: Alstrom syndrome 1

AML: acute myeloid leukemia

APCS: Amyloid P Component, serum

APOBEC: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

ASCT: autologous stem cell transplant

ASXL1: Additional Sex Combs Like 1

ATM: Ataxia-Telangiesctasia mutated

ATR: Ataxia telangiectasia and Rad3 related

ATRX: ATP-dependent helicase, X-linked

BAM: Binary Alignment Map

BCL10: B-cell lymphoma 10

BCR: B-cell receptor

BER: base excision repair

BF: Bayes factor

BIRC2: Baculoviral IAP repeat-containing protein 2

BIRC3: Baculoviral IAP repeat-containing protein 3

BLM: Bloom syndrome gene

BM: bone marrow

BMPC: bone marrow plasmacells

BRAF: B-Raf proto-oncogene

BRCA1: Breast Cancer Type 1 Susceptibility Protein

BRCA2: Breast Cancer Type 2 Susceptibility Protein

BS: single base substitution

BWA: Burrows-Wheeler Aligner

CAMK1D: Calcium/calmodulin-dependent protein kinase type 1D

CARD11: Caspase Recruitment Domain Family Member 11

CAR-T: Chimeric antigen receptor T cells

CASC5: Cancer susceptibility candidate gene 5 protein

CBX3: Chromobox 3

CCDN3: Cyclin D3

CCF: Cancer Clonal Fraction.

CCND1: Cyclin D1

CD: cluster of differenciation

CDKN2C: Cyclin Dependent Kinase Inhibitor 2C

CH: clonal hematopoiesis

CHD2: Chromodomain Helicase DNA Binding Protein 2

CHEK2: Checkpoint Kinase 2

CHGB: Chromogranin B

CI: confidence intervals.

CLL : chronic lymphocytic leukemia

CNA: copy number aberration

CoMMpass: Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile

CP: chromoplexy

CRAB: calcium renal anemia bone

CRAF: Raf-1 proto-oncogene, serine/threonine

CRISPR: clustered regularly interspaced short palindromic repeats

CSR: Class switch recombination

CT: chromothipsis

CTC1: CST Telomere Replication Complex Component 1

CYLD: CYLD Lysine 63 Deubiquitinase

dBGAP: database of Genotypes and Phenotype

ddPCR: Droplet Digital Polymerase chain reaction

DDR: DNA damage response

DH : Double hit

DIS3: DIS3 Homolog, Exosome Endoribonuclease And 3'-5' Exoribonuclease

DKC1: Dyskerin Pseudouridine Synthase 1

DNA: desoxyribonucleic acid

DNAH17: Dynein Axonemal Heavy Chain 17

DNAH5: Dynein Axonemal Heavy Chain 5

DNMT3A: DNA methyltransferase 3A

DSB: double strand break

DUSP2: Dual Specificity Phosphatase 2

ECM: extracellular matrix

ECOG: eastern cooperative oncology group

EFS: event free survival

EGA: European Genome Archive

EGFR: Epidermal growth factor receptor

EGR1: Early Growth Response 1

EM: early myeloma

EP300: Histone acetyltransferase p300

ERCC6: Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 6

ERK: Extracellular regulated kinases

F: female

FAF1: Fas Associated Factor 1

FAM46C: family with sequence similarity 46, member C

FASTA: FAST-All file

FASTQ: FAST-Quality file

FFPE: Formalin-Fixed Paraffin-Embedded

FISH: fluorescence in-situ hybridization

FLT3: fms-like tyrosine kinase 3

FOXO: Forkhead box O

FU: follow-up

G1: Gap 1 phase

G2: Gap 2 phase

GALNT2: polypeptide N-Acetylgalactosaminyltransferase 2

GAP: GTPase-activating proteins

GAR1: GAR1 ribonucleoprotein

GEF: Guanine nucleotide exchange factor

GEP: gene expression profiling

GEP4: four gene expression profiling

GEP70: seventy gene expression profiling

GTP: guanosine triphosphate

GWAS: genome-wide association study

Hi-C: high throughput Chromosome Conformation Capture

HiR: high risk

HR: hazard ratio,

HRD: hyperdiploid

HRMM: High risk multiple myeloma

ICL: interstrand cross-links

IDH2: Isocitrate dehydrogenase 2

IF: Intermediate filament

IG: Immunoglobulin

IgA: immunoglobulin A

IgD : immunoglobulin D

IgE: immunoglobulin E

IGF1: Insulin-Like Growth Factor 1

IGFN1: Immunoglobulin Like And Fibronectin Type III Domain Containing 1

IgG : immunoglobulin G

IGH: immunoglobulin heavy chain

IGHA1: immunoglobulin A1 gene

IGHG1: immunoglobulin G1 gene

IGHG4 : immunoglobulin G4 gene

IGL : immunoglobulin lambda

IGM : immunoglobulin M gene

IL2RG : interleukin 2 receptor, gamma chain

IL7R : Interleukin 7 Recepto

IMID: immunomodulatory drug

IMWG: International Myeloma Working Group

IQR: interquartile range

IR: intermediate risk,

IRB Institutional Review Boars

IRF4: Interferon Regulatory Factor 4

ISS: International staging system

ITT: Intention to treat

JAK2: Janus kinase 2

Kde: Kappa-Deleting-Element

*KDM5A: L*ysine Demethylase 5A

KLHL6: Kelch Like Family Member 6

KMT2B: Lysine Methyltransferase 2B

KMT2B: Lysine Methyltransferase 2B

KRAS: Kirsten Rat Sarcoma virus

L: liter

LCDD: Light chain deposition disease:

LIWGS: Long-insert whole-genome sequencing

LPD: lymphoproliferative disorders

LR: low risk

LRP1B: low-density lipoprotein receptor-related protein 1B

LRRK2: Leucine-rich repeat kinase 2

LTBP4: Latent Transforming Growth Factor Beta Binding Protein 4

LTL: leucocyte thelomere length

LYN: LYN proto-oncogene

M: male

MAF: v-maf musculoaponeurotic fibrosarcoma oncogene homolog

MAFB: v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog B

MAGI1: Membrane-associated guanylate kinase 1

MAP3K14: Mitogen-Activated Protein Kinase Kinase Kinase 14

MAPK: Mitogen-activated protein kinase

MAX: MYC Associated Factor X

MCL: Mantle cell lymphoma

MDE: myeloma defining event

MDGA2: MAM Domain Containing Glycosylphosphatidylinositol Anchor 2

MEK: Mitogen-activated protein kinase kinase

MF: MAF subgroup

mg: milligram

MGP: myeloma genome project

MGUS: Monoclonal gammopathy of undetermined significance

MIR1208: micro RNA 1208

miRNA: microRNA

MM: multiple myeloma

MMR: mismatch repair

MMSET: Multiple Myeloma SET Domain-Containing Protein

MRI: magnetic resonance imaging

MS: MMSET subgroup

MS: M-spike.

mTOR: mammalian target of rapamycin

MTRR: Methionine synthase reductase

MUC16: Mucin 16

MYC: v-myc avian myelocytomatosis viral oncogene homolog

N: number

NDMM: newly diagnosed multiple myeloma

NDMM= newly diagnosed myeloma,

NER: nucleotide excision repair

NF1: Neurofibromin 1

NFKBIE: NFKB Inhibitor Epsilon

NF-κB: nuclear factor-kappa B

NGS: next generation sequencing,

NHEJ: non homologous nuclear endjoining

NHP2: Nucleolar Protein Family A, Member 2

nHRD: non-hyperdiploid,

nNMF: non Negative matrix factorisation

NOP10: Nucleolar Protein Family A, Member 3

NRAS: Neuroblastoma RAS viral oncogene homolog

NRK: Nik-related protein kinase

NSCLC: non small cell lung cancer

NSD2: Nuclear SET Domain-Containing Protein 2

N*SD3:* Nuclear SET Domain-Containing Protein 3

OR11G2: Olfactory receptor 11G

OS: overall survival

P: p-values

PANTHER: Protein ANalysis THrough Evolutionary Relationships

PCL: plasmacell leukemia

PCR: polymerase chain reaction

PFS: progression free survival

PI: proteasome inhibitor

PIM1: Pim-1 proto-oncogene

PKHD1L1: Polycystic Kidney and Hepatic Disease 1-Like 1

PMID: PubMed Identifier

POEMS: polyneuropathy, organomegaly, endocrinopathy, M protein, and skin changes

PON3: Paraoxonase 3

POT1: Protection Of Telomeres 1

PPM1D: Protein Phosphatase, Mg2+/Mn2+ Dependent 1D

PR: proliferation

PRKD2: Serine/threonine-protein kinase D2

PS: performance status

PSMA2: Proteasome 20S Subunit Alpha 2

PSMA8: Proteasome 20S Subunit Alpha 8

PTPN11 : Tyrosine-protein phosphatase non-receptor type 11

PTPN13: Tyrosine-protein phosphatase non-receptor type 13

r: adjusted coefficient of determination.

RAF1: Raf-1 proto-oncogene, serine/threonine

RASA1: RAS P21 Protein Activator

RB1: retinoblastoma 1

RCC1: Regulator of chromosome condensation 1

RCCD: RCC1 Domain Containing 1

R-ISS: revised international staging system

RMI1: RecQ-mediated genome instability protein 1

RNA: Ribonucleic acid

RP1L1: Retinitis pigmentosa 1-like 1

RTEL1: Regulator Of Telomere Elongation Helicase 1

RUVBL1: Pontin

RUVBL2 : Reptin

SALL4: Sal-like protein 4

SAMHD1: SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1

SAP: serum amyloid P component protein.

SEER: Surveillance, Epidemiology, and End Results

SET2D: SET domain containing 2

SF3B1: Splicing factor 3B subunit 1

sFLC: serum free light chain

SHH: somatic hypermutation

SHP2: Src homology region 2 (SH2)-containing protein tyrosine phosphatase 2

SIGLEC12: Sialic acid-binding Ig-like lectin 12

SMM: smoldering myeloma

SNP: Single Nucleotide Polymorphisms

SNVL Sincgle Nucleotide Variant

SRSF11: Serine And Arginine Rich Splicing Factor 11

STELA: Single TElomere Length Analysis

SUSD5: Sushi Domain Containing 5

SVIL: Supervillin

SWOG: Southwest Oncology Group

SYNM: Synemin

TA: telomeras activity

TdT: Terminal deoxynucleotidyl transferase

TENT5C: Terminal Nucleotidyltransferase 5C

TERC: Telomerase RNA Component

TERF2: Telomeric repeat-binding factor 2.

TERF2IP: Telomeric repeat-binding factor 2 interacting protein

TERT: Telomerase reverse transcriptase

TET2: Tet Methylcytosine Dioxygenase 2

TFH: follicular helper T-cells

TGFB: Transforming Growth Factor Beta 1

TI: templated insertion

TIAM2: T-lymphoma invasion and metastasis-inducing protein 2

TINF2: Telomeric Repeat Binding Factor 1 Interacting Nuclear Factor 2

TL: telomere length

TNFAIP3: Tumor necrosis factor, alpha-induced protein 3

*TNIP1:*Tumor necrosis factor, alpha-induced protein 3 Interacting Protein 1

TP53: tumor protein 53

TRAF2: Tumor necrotic factor Receptor Associated Factor 2

TRAF3: Tumor necrotic factor Receptor Associated Factor 3

TRPS1: Tricho-rhino-phalangeal syndrome Type 1

TT: total therapy

TTL: tumor telomere length

Tx: translocation,

TXNDC5: Thioredoxin Domain Containing 5

UAMS: University of Arkansas for Medical Science

UK: United-Kingdom

ULP-WGS

US: United States

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

WRN: Werner's syndrome RecQ Like Helicase

XBP1: X-Box Binding Protein 1

ZBTB21: Zinc Finger And BTB Domain Containing 21

ZFHX4: zinc Finger Homeobox 4

ZFP36L1: zinc Ring Finger Protein Like 1

ZNF292: zinc finger protein 292

K: kappa

Λ: lambda

Chapter 1: Introduction

1.1. Definitions

Plasma cell disorders derive from a proliferation of a single mature B-cell clone. They have in common the presence of a monoclonal, complete or partial, immunoglobulin (Ig) in the serum, or urine. They are believed to arise from a plasma cell that expands and accumulates in the bone marrow. The clinical presentation of these distinct entities is heterogeneous ranging from asymptomatic conditions such as Monoclonal Gammopathy of Undetermined Significance (MGUS) or Smouldering Multiple Myeloma (SMM) to symptomatic form of diseases such as Multiple Myeloma (MM) or light chain Amyloidosis (AL). In 2003, the International Myeloma Working Group (IMWG) first defined a set of criteria to add precision to the definition of these disorders (International Myeloma Working Group, 2003) that were subsequently revised in 2014 (Rajkumar *et al*, 2014) based on a better understanding of the clinical phenotype of these disorders, **Table 1.1**.

Multiple myeloma (MM) is characterized by both the presence of more than 10% clonal plasma cells in the bone marrow and the presence of end-organ damage. Historically, end-organ damage consisted of the CRAB criteria that include hypercalcemia, anaemia, bone disease, and renal failure. More recently, three additional criteria have been added: the presence of more than one focal lesion on Magnetic Resonance Imaging (MRI), the presence of more than 60% plasma cells in the marrow, and the presence of a light chain ratio (the ratio of involved over uninvolved light chain) greater than or equal to a hundred. MM is preceded by two premalignant conditions known as MGUS and SMM, which may be clinically silent and remain undetected.

MGUS is divided into three groups depending on the isotype. Non-IgM MGUS, is defined by the presence of a small tumour burden with an IgA, IgG, IgD, or IgE paraprotein less or equal to 30g/L, less than 10% bone marrow plasma cells (BMPC), and the

absence end-organ damage. Similarly, light-chain MGUS has been defined by a low tumour burden with a low-level circulating light chain (increased κSFLC in patients with ratio ≥1.65 and increased λSFLC in patients with ratio ≤0.26). Both subtypes of MGUS have been associated with a 1% *per annum* risk of progression to symptomatic MM (Kyle *et al*, 1992). The last subtype of MGUS, IgM MGUS, is characterised by an IgM monoclonal protein less or equal to 30g/L without evidence of anaemia, constitutional B-symptoms, hyper viscosity, lymphadenopathy, hepatosplenomegaly, or other end-organ damage that can be attributed to the underlying proliferative disorder. Unlike the other subtypes, IgM MGUS rarely progress to MM, but more often progress to other B-cell malignancies such as Waldenström Macroglobulinemia or Marginal Zone Lymphoma (Kyle *et al*, 2009).

SMM, on the other hand is distinguished from MGUS by a higher risk of progression to symptomatic myeloma. Unlike MGUS, it has a higher tumour burden (paraportein> 30g/L or ≥10% BMPC) but no evidence of end-organ damage (Rajkumar *et al*, 2011). This definition is somewhat blurry as there seems to be a *continuum* from MGUS to MM, and efforts are currently being made to remove this category and elaborate criteria to reclassify low-risk SMM as MGUS and high-risk cases as MM requiring treatment intervention.

At the other end of the spectrum, three other entities are characterised by a high tumour burden. Plasma cell leukaemia (PCL) is clinically and genetically distinct from MM. It is defined by the observation in blood of more than 20% clonal plasma cells on leucocyte differential count or by counting more than 2×10^9 per litre circulating clonal plasma cells. Generally regarded as dire, it can either occur at presentation or in the relapsed and refractory disease setting (van de Donk *et al*, 2012). Another form of extramedullary disease is referred to as plasmacytoma, a tumour consisting of abnormal plasma cells that grow within the soft tissue or bony skeleton. A plasmacytoma can be present as a solitary tumour composed abnormal plasma cells, in which case it is termed a “solitary”

plasmacytoma or it can be associated with a low marrow involvement (Solitary plasmacytoma with minimal marrow involvement) (Caers *et al*, 2018).

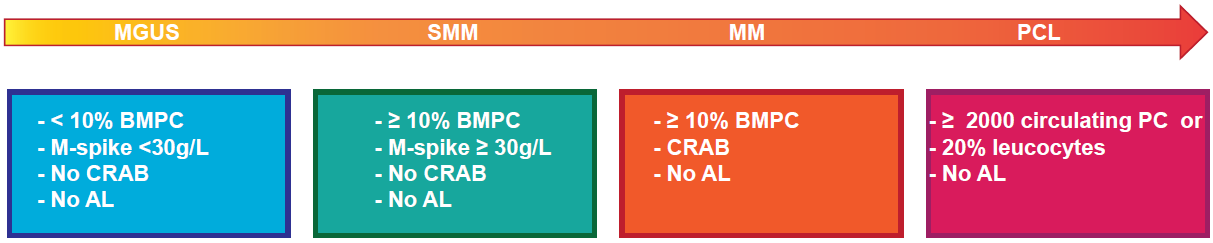


Figure 1.1. Summary of plasmacell disorders

Finally, two last symptomatic entities are not characterised by their tumour burden but their multiple symptoms. Light chain amyloidosis (AL) bears a separate type of end organ damage related to the deposition of immunoglobulin light chain that progressively accumulate throughout the body and form toxic protein aggregates leading to organ dysfunction and eventually organ failure. These deposits explain a vast array of clinical symptoms such as congestive heart failure, kidney failure and neuropathy (Merlini & Bellotti, 2003). The second, POEMS syndrome, also known as Crow-Fukase Syndrome, is a paraneoplastic disorder associated with an underlying plasma cell dyscrasia. The acronym POEMS (polyneuropathy, organomegaly, endocrinopathy, M-protein, and skin changes) captures several dominant features of the syndrome, but excludes other important clinical findings (papilledema, peripheral oedema, ascites, effusions, fatigue, clubbing), biological findings (elevated levels of vascular endothelial growth factor (VEGF), thrombocytosis, polycythaemia), pathological findings (Castleman disease), or even imaging findings (sclerotic bone lesions). The major clinical feature of the syndrome is a chronic progressive polyneuropathy with a predominantly motor disability. Patients usually have low tumour burdens with *circa.* 5% BMPC, almost always λ restricted, and rarely exhibit conventional CRAB criteria (Dispenzieri, 2019).

Introduction

	Definition
Myeloma	Clonal bone marrow plasma cells >10% or biopsy-proven bony or extramedullary plasmacytoma and any one or more of the following CRAB features and myeloma-defining events: • Hypercalcemia: serum calcium 0.25 mmol/L (1mg/dL) higher than the upper limit of normal or ≥2.75 mmol/L (11mg/dL) • Renal insufficiency: creatinine clearance ≤40 mL per minute or serum creatinine ≥177mol/L (≥2mg/dL) • Anaemia: haemoglobin value of 20g/L below the lowest limit of normal, or a haemoglobin value ≤100g/L • Bone lesions: one or more osteolytic lesion on skeletal radiography, CT, or PET/CT. If bone marrow has ≥10% clonal plasma cells, more than one bone lesion is required to distinguish from solitary plasmacytoma with minimal marrow involvement Any one or more of the following biomarkers of malignancy (MDEs): • 60% or greater clonal plasma cells on bone marrow examination • Serum involved / uninvolved free light chain ratio of 100 or greater, provided the absolute level of the involved light chain is at least 100mg/L (a patient's involved free light chain either kappa or lambda is the one that is above the normal reference range; the uninvolved free light chain is the one that is typically in, or below, the normal range) • More than one focal lesion on MRI that is at least 5mm or greater in size.
Smouldering Multiple Myeloma	Both criteria must be met: • Serum monoclonal protein (IgG or IgA) >30g/L or urinary monoclonal protein >500mg per 24h and/or clonal bone marrow plasma cells 10-60% • Absence of myeloma-defining events or amyloidosis
Non-IgM MGUS	• Serum monoclonal protein ≤ 30g/L • Clonal bone marrow plasma cells ≤10% • Absence of end-organ damage such as hypercalcemia, renal insufficiency, anaemia, and bone lesions (CRAB) or amyloidosis that can be attributed to the plasma cell proliferative disorder
IgM MGUS	• Serum IgM monoclonal protein ≤30g/L • No evidence of anaemia, constitutional symptoms, hyperviscosity, lymphadenopathy, hepatosplenomegaly, or other end-organ damage that can be attributed to the plasma cell proliferative disorder
Light chain MGUS	• Abnormal FLC ratio (≤0.26 or ≥1.65) • Increased level of the appropriate free light chain (increased κFLC in patients with ratio ≥1.65 and increased λFLC in patients with ratio ≤0.26) • No immunoglobulin heavy chain expression on immunofixation • Absence of end-organ damage • Clonal bone marrow plasma cells ≤10%

	• Urinary monoclonal protein ≤500mg/24h
Plasma cell Leukaemia	Plasma cell leukaemia is defined by the observation in blood • of more than 20% clonal plasma cells by differential count of the leucocytes • or by counting more than 2×10^9 per litre circulating clonal plasma cells.
Solitary plasmacytoma	• Biopsy-proven solitary lesion of bone or soft tissue with evidence of clonal plasma cells • Normal bone marrow with no evidence of clonal plasma cells • Normal skeletal survey and MRI (or CT) of spine and pelvis (except for the primary solitary lesion) • Absence of end-organ damage
Solitary plasmacytoma with minimal marrow involvement	• Biopsy-proven solitary lesion of bone or soft tissue with evidence of clonal plasma cells • Clonal bone marrow plasma cells ≤10% • Normal skeletal survey and MRI (or CT) of spine and pelvis (except for the primary solitary lesion) • Absence of end-organ damage such as hypercalcemia, renal insufficiency, anaemia, and bone lesions (CRAB) or amyloidosis that can be attributed to the plasma cell proliferative disorder

POEMS syndrome	• Polyneuropathy
	• Monoclonal plasma cell proliferative disorder
	• Any one of the 3 other major criteria: sclerotic bone lesions, Castleman's disease, elevated levels of VEGFA
	• Any one of the following 6 minor criteria:
	  ○ Organomegaly (splenomegaly, hepatomegaly, or lymphadenopathy)
	  ○ Extravascular volume overload (oedema, pleural effusion, or ascites)
	  ○ Endocrinopathy (adrenal, thyroid, pituitary, gonadal, parathyroid, pancreatic)
	  ○ Skin changes (hyperpigmentation, hypertrichosis, glomeruloid hemangiomata, plethora, acrocyanosis, flushing, white nails)
	  ○ Papilledema
	  ○ Thrombocytosis/polycythaemia

Systemic AL amyloidosis	• Presence of an amyloid-related systemic syndrome (e.g., renal, liver, heart, gastrointestinal tract, or peripheral nerve involvement) • Positive amyloid staining by Congo red in any tissue (e.g., fat aspirate, bone marrow, or organ biopsy) • Evidence that amyloid is light-chain-related established by direct examination of the amyloid using mass spectrometry-based proteomic analysis or immune-eletronmicroscopy • Evidence of a monoclonal plasma cell proliferative disorder (serum monoclonal protein, abnormal free light chain ratio, or clonal plasma cells in the bone marrow)

Table 1.1. IMWG criteria of plasma cell disorders (Rajkumar *et al*, 2014)

Such diversity in the clinical presentations is likely to arise from a significant heterogeneity at the plasma cell level. Understanding the plasma cell biology that underlies these entities could shed some light on these disorders and refine patient management.

1.2. Plasma cell biology

Our understanding of B-cell malignancies relies on a system whereby a normal B-cell precursor acquires a genetic event such as a translocation or a copy number change leading to the development of a malignancy and progress through the various phases of disease by acquiring secondary events, **Figure 1.2.1.1.**

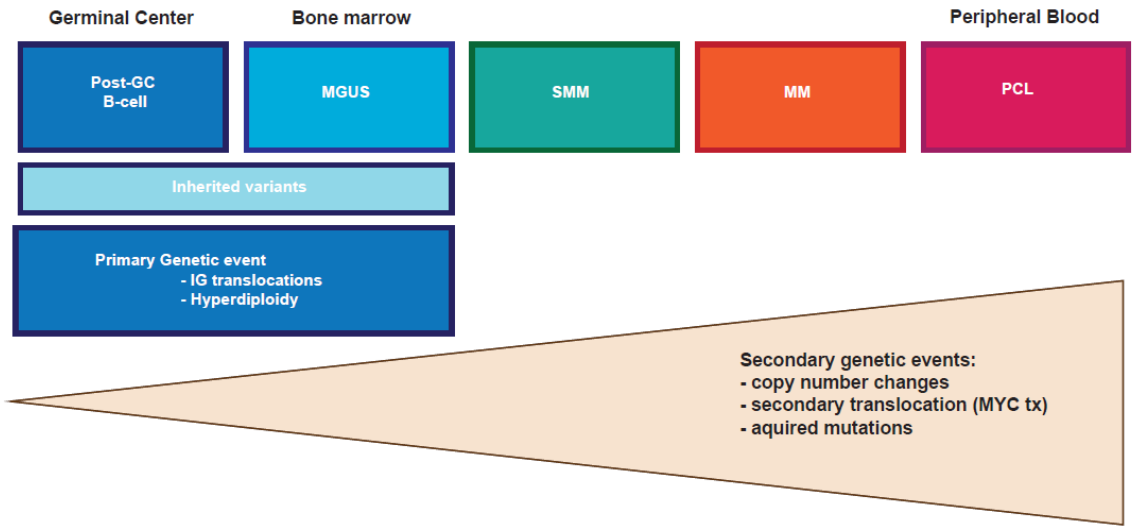


Figure 1.2.1.1. Schematic representation of our current understanding of the genetics of plasma cell disorders where a predisposed post-germinal centre B-cell acquires an initiating event (hyperdiploidy or an IG translocation) and progresses through the various phases of disease by acquiring secondary events (adapted from (Morgan *et al*, 2012)).

The timing and nature of the genetic events will lead to different disease types. Although acquiring hyperdiploidy (HRD) in the pre-B-cell stage leads to acute lymphoblastic leukaemia (ALL), if acquired later, the same event, in a poorly understood process, will lead to MM. This working hypothesis has shed light onto the initiation of MM and expanding this concept through to normal plasma cell biology has helped decipher

the mechanism that will lead a normal plasma cell to MGUS through SMM, NDMM and relapsed and refractory disease.

The aim of this section is to dissect plasma cell development to gain insight into the mechanism of plasma cell disorders at the molecular level.

1.2.1. B-cell terminal differentiation

1.2.1.1. Generation of normal mature B-cell

The B-cell lineage is derived from lymphoid progenitor cells that differentiate from haematopoietic stem cells. They undergo a complex, highly controlled maturation process leading to the expression of a functional Immunoglobulin (Ig) formed of two identical heavy chains [coded by the Immunoglobulin Heavy Chain genes (*IGH*)] and two identical light chains [coded by either the Immunoglobulin Light chain Kappa (*IGK*) or Lambda (*IGL*) genes] (Delves & Roitt, 2000).

The *IGH* genes are located on chromosome 14, and may be divided into four groups depending on the segments of the heavy chain they code for: the variable (V), the Diversity (D), the Joining (J) and the Constant (C) segments. At the pro-B-cell stage, cells undergo consecutive DNA recombinations of single D, J and V segments, under the control of the recombination signal sequences (RSS) that flank each individual gene segments. This site-specific DNA recombination process is catalysed by the proteins encoded by the recombination activating genes 1 and 2 (*RAG1* and *RAG2*) (Nishana & Raghavan, 2012). DNA repair processes such as mismatch repair, base excision repair and non-homologous end joining pathways (Delves & Roitt, 2000; Vuong *et al*, 2013) re-join them. Junction diversity by random addition of nucleotides by Terminal deoxynucleotidyl transferase (TdT), adds further diversity (Victor & Capra, 1994). Once a functional VDJ rearrangement has been achieved the other allele is excluded from recombination attempts in a process called ‘allelic exclusion’ (Jung *et al*, 2006). Finally, at the pre-B cell stage, the VDJ segment formed as a consequence of these rearrangement events is attached to a constant region (M or D) by splicing at the RNA level (Xu *et al*, 2012), **Figure 1.2.1.1.1**.

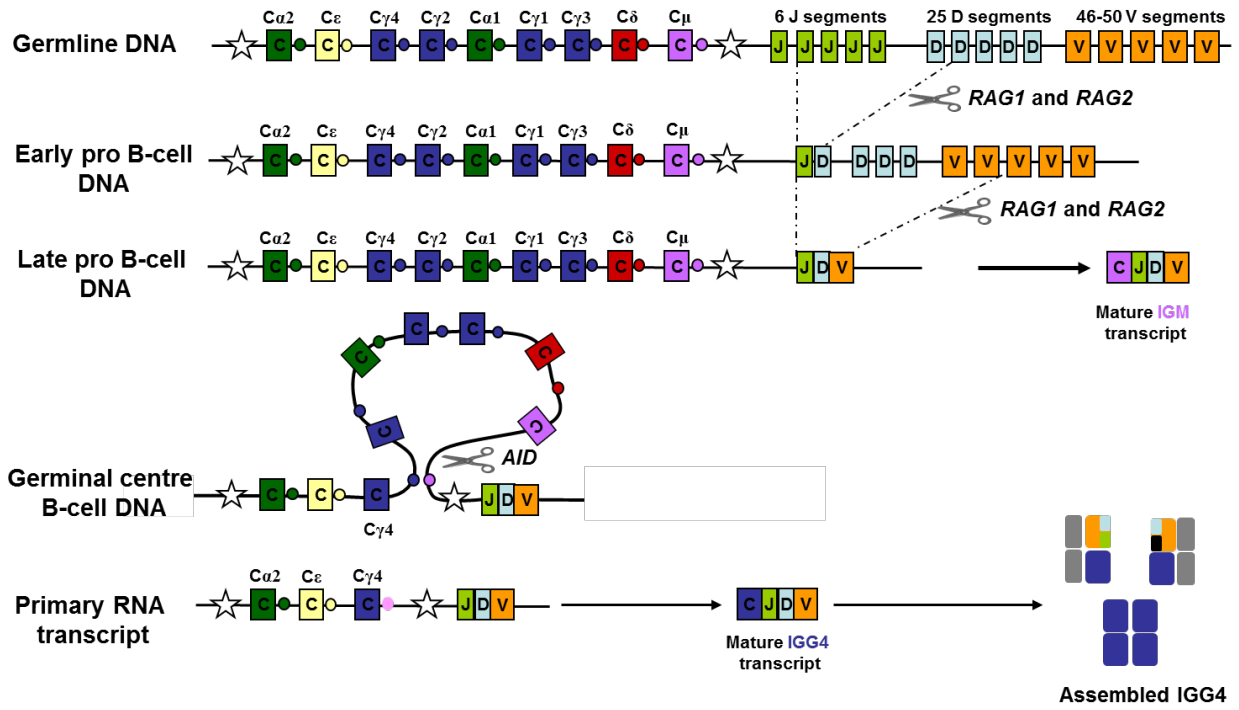


Figure 1.2.1.1.1. Summary of the VDJ rearrangement (Boyle *et al*, 2014)

Following the *IGH* rearrangement, the light chain loci proceed to rearrange. The *IGK* and *IGL* genes are located at loci 2p12 and 22q11.2 respectively. Like the *IGH* locus, the light chain loci include V, J and C segments but lack D segments. First, rearrangements start at the *IGK* locus where J and V segments rearrange. There is limited junctional diversity (Victor *et al*, 1994). Non-functional rearrangements may occur when a downstream element, the Kappa-Deleting-Element (Kde), rearranges with the V segment or an intron (Langerak & van Dongen, 2006) thus terminating the involvement of that locus. This will lead the recombination to switch from the first to the second κ locus (i.e., on the alternate chromosome), or from the second κ locus to the *IGL*-bearing chromosomes. If the κ rearrangement is not functional, the *IGL* locus proceeds to rearrange and the *IGK* locus is deleted.

Finally, assembly of the heavy and light chain on the cell surface, the so called B-cell receptor (BCR), at the immature B-cell stage enables the cell to escape apoptosis and proceed to maturation outside the bone marrow (Delves & Roitt, 2000).

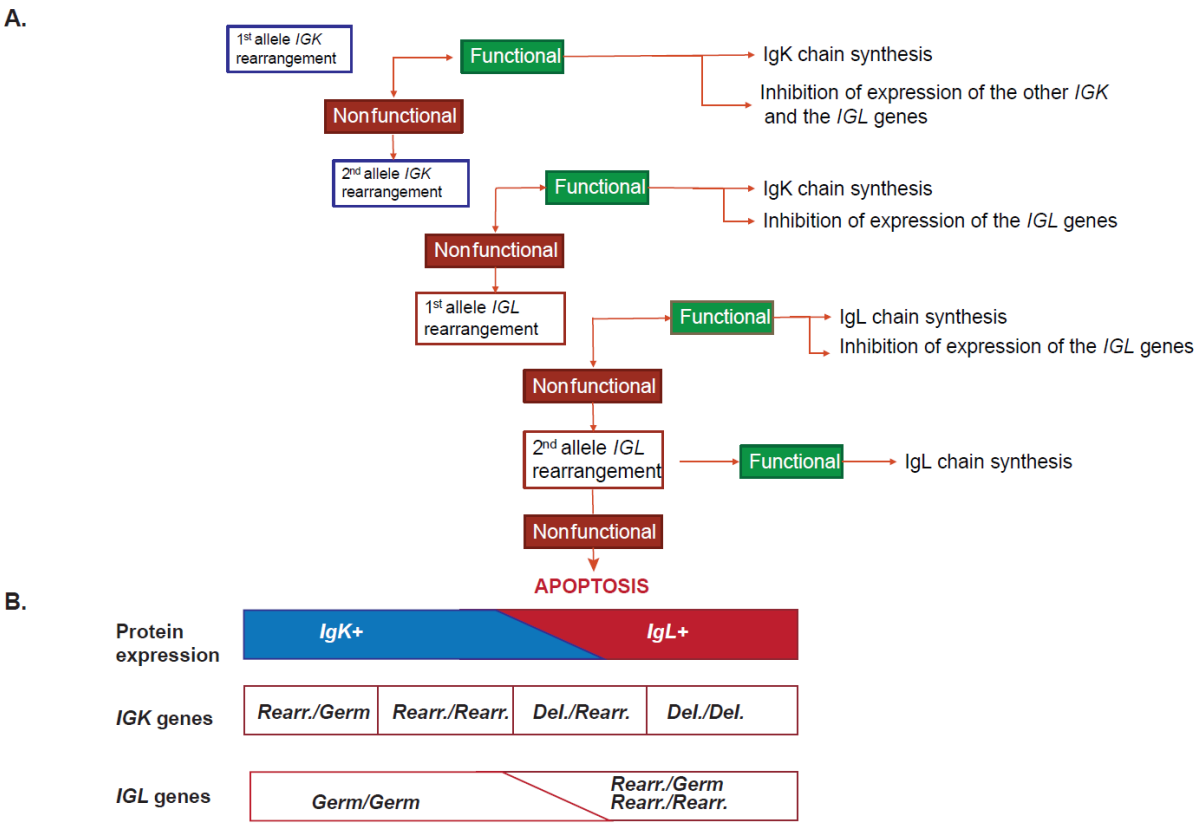


Figure 1.2.1.1.2. Light chain rearrangement. A. Sequencing of light chain rearrangement. B. Correspondence between the DNA and protein level.

1.2.1.2. Production of Antibody secreting cells (ASC)

Mature B-cells have three subsets: follicular B-cells, marginal zone B-cells and B1 cells. Follicular B-cells may be found in the lymphoid follicles of lymph nodes and spleen whereas marginal zone B-cells in the marginal sinus and spleen making them good detectors of blood borne pathogens and partial antigens. B1 cells reside mainly in the peritoneum, pleural cavities and mucosal sites making them ideal actors to recognise environmental pathogens (Rothstein *et al*, 2013).

The production of ASC is a 2-step process. First, an extrafollicular response happens upon receiving an antigen receptor dependant signal, leading to lymphoblast division, class switch recombination (CSR) and differentiation into plasmablasts. Plasmablasts are short-lived, cycling, ASC that are present in extrafollicular foci. Little somatic hypermutation (SHH) occurs and the receptor's affinity is moderate. This step will lead to an early protective immunity. After this initial process creating the primary Ig

repertoire, these immature yet immuno-competent B-cells exit the bone marrow and enter the lymphoid organs through the mantle zone into the germinal centre (GC) where they proceed to proliferate and undergo affinity maturation. This maturation or germinal centre reaction requires close interaction between B-cells, antigen presenting dendritic cells and T-cells. B-cells that have an antigen-specific BCR are selected to survive and proliferate. These positive selected cells termed centroblasts, rapidly divide and expand in the dark zone of the GC. From there, some enter the light zone, cease to divide and become centrocytes. Centrocytes go on to terminally differentiate into either memory B-cells or antibody-secreting cells (plasmablasts or plasma cells). In a poorly understood process, some centrocytes can return to the germinal centre and undergo further clonal selection (Nutt *et al*, 2011). During this GC reaction two type of DNA modifications occur: class switch recombination and somatic hypermutation.

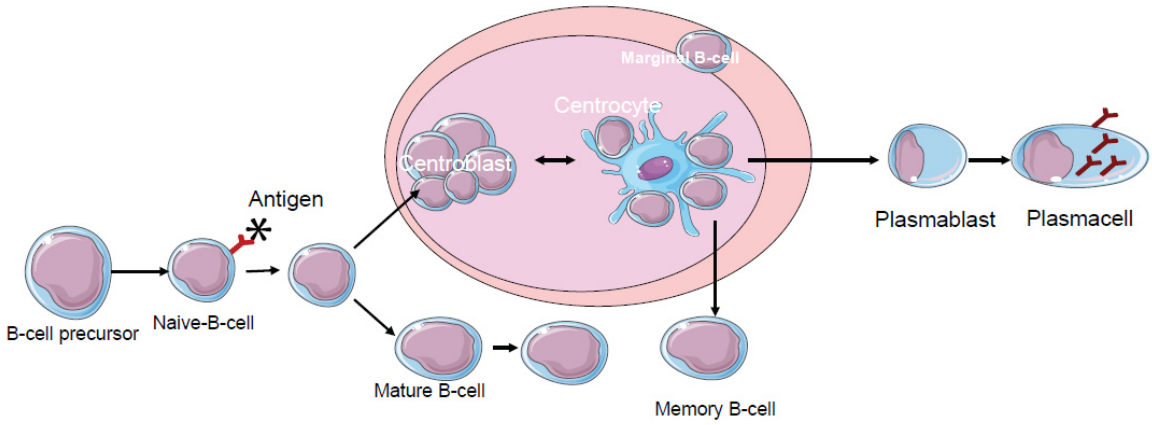


Figure 1.2.1.2.1. Schematic representation of the B-cell maturation process leading to the development of mature plasma cells

Secondly, activated B-cells re-enter the follicle under the influence of T-follicular helper cells (TFH), proliferate, and an antigen specific receptor is selected by SHH. Somatic hypermutation adds point mutations into the variable regions of the IG genes (Fraenkel *et al*, 2007) increasing antibody affinity. As the process of SHH is repeated several times, populations of B-cells bearing receptors with increasing affinity toward decreasing levels of antigen are selected. This process is referred to as “affinity maturation”. Class switch recombination (CSR) occurs when the VDJ segment is brought

into proximity to another constant region (*IGHG1-4, IGHA1-2* or *IGHE*) by DNA deletion. CSR provides an important means of altering the effector function of the antibodies produced by the B-cells and is central to the maturation of the antibody response. Although they differ mechanistically, both these processes are mediated by DNA double strand breaks induced Activation-induced cytidine deaminase (AID) activity. The AID protein (Muramatsu *et al*, 2000) belongs to the APOBEC family of cytidine deaminases and is capable of deaminating cytidine to uracil *in vitro* on both single strand DNA substrates and single strand DNA-RNA hybrids (Chaudhuri *et al*, 2003; Ramiro *et al*, 2003). Although highly regulated (Keim *et al*, 2013) it can induce DNA mutations at a relatively high rate (up to 10^4 to 10^3 per base per division) and is a known oncogene (Muramatsu *et al*, 2000).

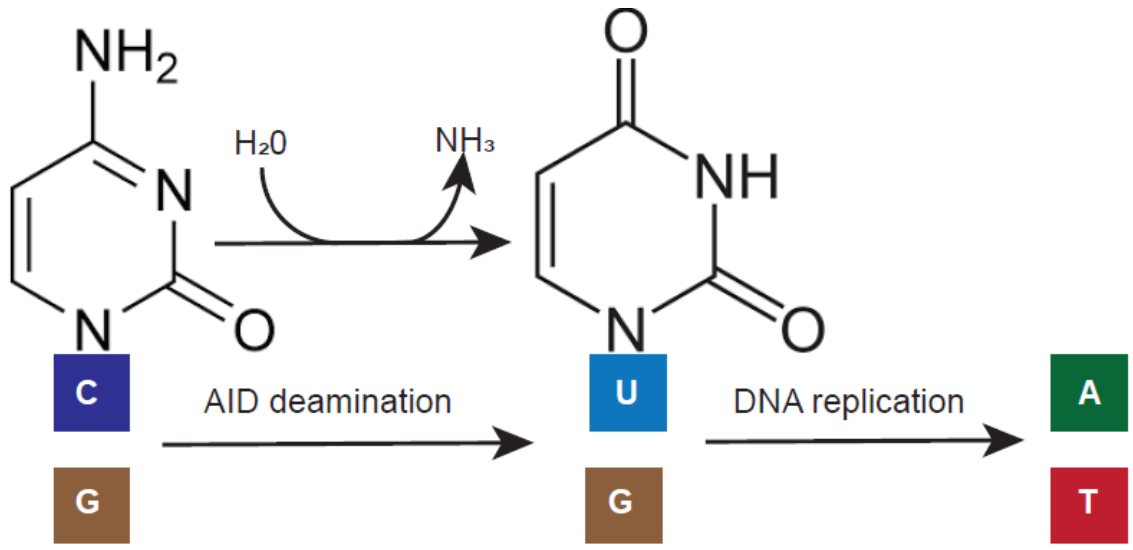


Figure 1.2.1.2.2. AID-cytidine deamination activity

The definitive aim of all these processes is the development of a wide B-cell repertoire that contains self-tolerant, high-affinity antibodies that recognize and bind foreign antigens and display suitable effector properties. Although, tightly regulated, these processes are key in the genesis and progression of plasma cell disorders.

1.2.1.3. Implication of B-cell terminal differentiation in plasma cell disorders

Evidence has arisen, from the study of paraproteins and the Ig loci, that these processes may be implicated in the ontogeny of plasma cell disorders

1.2.1.3.1. Paraprotein

In a series of 10,000 newly diagnosed multiple myeloma patients Greipp et al (Greipp *et al*, 2005a) estimated that 60% of cases were IgG, 24% IgA, 11% Light chain only (LCO), and 3% IgD. Further analysis of the IgG patients suggests that 68% were IgG1, 17% IgG2, 11% IgG3, and 4% IgG4. Similarly for IgA patients, IgA1 represent 93% of cases and IgA2 7% (Papadea *et al*, 1989; Fasullo *et al*, 1989). Overall, these suggest that the proportion of each heavy chain isotype is in keeping with the normal daily immunoglobulin production (Schroeder & Cavacini, 2010). In a poorly understood process, there is an isotype bias between cytogenetic subgroups. For instance, there is an over-representation of the t(11;14) among the IgM, IgE, and LCO dyscrasias (Avet-Loiseau *et al*, 2003b), although overall the frequency is respected.

Similarly, the proportion of light chains κ and λ cases are similar in MM than in the normal serum with 63% and 37% of κ and λ cases respectively (Greipp *et al*, 2005a). There is nonetheless a few biased subsets such as IgD myeloma (Modi *et al*, 2015), AL (Sanchorawala, 2006), and POEMS (Stankowski-Drengler *et al*, 2010) where λ is more prevalent. The reasons for this remain unknown.

1.2.1.3.2. IG rearrangement

1.2.1.3.2.1. VDJ rearrangement

For many years, the study of the functional VDJ rearrangement in plasma cell disorders was hindered by the presence of somatic hypermutation that prevented the hybridization of PCR primers. It has gained recent interest in AL as some rearrangements have been associated with a higher probability of developing AL (see Chapter 1.3.1.).

1.2.1.3.2.2. Translocations

Evidence suggest that translocations into the *IG* loci are generated secondary to double strand breaks occurring through V(D)J rearrangements, CSR or receptor revision. By mapping the breakpoints to the switch regions, next generation sequencing (NGS) studies suggest that most translocations are generated via CSR with the exception of a

small subset of t(11;14) and t(14;16) that are generated via VDJ rearrangement (Walker *et al*, 2010a). Interestingly, recent reports suggest that in mantle cell lymphoma, t(11;14) can also be generated via VDJ rearrangement or CSR (Nadeu *et al*, 2020), thus partly refuting the cell of origin hypothesis to explain differences between these two entities. Furthermore, as these events occur through dysregulation of a in normal process, this explains why *IGL* translocations are not seen in κ myeloma patients and the distribution of VDJ segments used in light chain rearrangements are similar to those in normal B-cell.

1.2.1.3.2.3. Somatic hypermutation

Somatic hypermutation is a highly controlled process that targets selected regions such as the IG loci. Aberrant targets can be seen in lymphoid malignancies and has been well described in lymphoma (Khodabakhshi *et al*, 2012). It has also been described to some extent in MM (Gaidano *et al*, 2003) on loci such as *BCL2* or *IGLL5*. Some authors suggest that the malignant clone has the possibility of re-entering the germinal centre thus explaining various levels of SHH as measured by the AID signature (Francesco Maura, personal communication). This is still highly debated and evidence remains scarce.

Most importantly, deamination processes such as the APOBEC hypermutational processes, are present in MM and are responsible for *circa*. 60% of mutations in the t(14;16) and t(14;20) subgroups (Walker *et al*, 2015b). The mechanism is unclear as is why it only affects this specific subgroup of MM. This process is associated with a high mutational load and frequent dysregulation of DNA repair pathway. To date, there is little evidence supporting the role of APOBEC in premalignant disease phases such as MGUS and SMM. Whether it is a feature associated with disease progression or whether its occurrence early on is responsible for an aggressive disease phenotype remains to be determined.

1.2.2. Risk factors associated with the development of plasma cell disorders

To date, there is no clear study that has revealed undisputed risk factors for the development or progression of plasma cell disorders. There is some evidence supporting lifestyle and environmental factors, oncogenic viruses, inherited conditions, and aging itself but unfortunately the different studies have failed to reach a consensus.

1.2.2.1. Lifestyle or environmental exposures

1.2.2.1.1. Weight and physical activity

Excess body weight is believed to constitute a risk factors for MM, as it has both been associated with increased MM incidence and mortality (Wallin & Larsson, 2011; Teras *et al*, 2014). Furthermore, the molecular epidemiology studies in MM and type II diabetes have uncovered the same single nucleotide polymorphisms (SNP) (Ríos *et al*, 2015). On the other hand, observational studies have, so far, provided conflicting results, likely attributed to confounding factors and reverse causation. Recently, a mendelian randomisation approach, provided no evidence of body mass index (BMI) or other adiposity traits, influencing MM risk or survival (Went *et al*, 2017). In another study, the inverse association between physical activity and MM risk did not reach statistical significance (Marinac *et al*, 2018). Therefore, this remains an active field of ongoing research.

1.2.2.1.2. Smoking

Cigarette smoking is one of the most important risk factors for cancer, but the association with MM or its precursor states has never been ascertained when considering either ever or current smoking (Andreotti *et al*, 2015; Psaltopoulou *et al*, 2013; Castillo *et al*, 2012). As such, it is not considered a risk factor for plasmacell disorders.

1.2.2.1.3. Professional exposure

Many studies have suggested that professional exposures are related to plasma cell disorders.

In the Swedish workers cohort, an excess risk was detected among various workers including agricultural, horticultural and forestry entrepreneurs, bakers and pastry chefs, dental technicians, stone cutters and carvers, and penitentiary officials (V *et al*, 2008). Occasional but intense exposure to pesticides was also associated with MM in that study (V *et al*, 2008) but not others (Alavanja *et al*, 2004). Occupation as a firefighter also emerged as a risk factor for MM, conferring an approximately 50% increased risk; similarly, hairdressers had a 40% increased risk for MM. Comparable risks were found to be associated with MGUS (Landgren *et al*, 2009, 2018).

Other professional exposures such as DDT (dichloro-diphenyl-trichloroethane) exposure, exposure to phenoxy-acetics and chlorophenols, and contact with animals (sheep, horses, dairy cattle) have also been linked to an excess risk of MM. Occupational exposure to methylene chloride was also linked with increased MM risk (Liu *et al*, 2013). Regarding exposure to benzene, a meta-analysis addressing cohort studies pointed to a slight but not significant increase risk (Yoon *et al*, 2018); similarly no association was seen in a meta-analysis of case control studies (Infante, 2006). The data regarding exposure to radiation are also conflicting with positive association seen in some studies (Schubauer-Berigan *et al*, 2015) but not others (Leuraud *et al*, 2015).

An increased risk of MGUS including light-chain MGUS has been shown among individuals exposed to known and suspected carcinogens including PCB (Polychlorinated biphenyl), PAHs (Polycyclic aromatic hydrocarbon), asbestos, (Pukkala *et al*, 2014; Landgren *et al*, 2009; Ruder *et al*, 2014) and Agent Orange (which contains the human carcinogen 2,3,7,8-tetrachlorodibenzo-p-dioxin) (Landgren *et al*, 2009).

1.2.3.2. Infection

The role of viruses in plasmacell disorders has long been hypothesised. First, when considering HIV patients, prior to the development of highly active antiretroviral therapies (HAART), the incidence of MGUS was believed to range from 4-26% in patients with HIV infection (Briault *et al*, 1988; Coker *et al*, 2013). The standardized incidence ratios (SIRs) for MM from 1980-1996 and 1996-2002 were 2.60 (95% CI 1.92-3.44) and 2.20 (95% CI 1.10-3.94), respectively (Grulich *et al*, 2002). Combined, these data support an increased risk of plasmacell disorders in HIV infected patients.

In the 1990's, Rettig et al (Rettig *et al*, 1997) suggested a role for HHV8 in the malignant transformation from MGUS to MM, as it was found in the BM dendritic cells of MM patients but neither in the malignant plasma cells nor BM dendritic cells derived from normal individuals or patients with other malignancies. Nonetheless, these findings were never confirmed in larger series.

Finally, there is no evidence supporting the role of other viruses such as the hepatitis C virus (Duberg *et al*, 2005) in population based studies.

1.2.3.3. Inherited condition

Although no lifestyle or environmental exposure has been consistently linked to an increased risk of MM, the two to four-fold increased risk observed in relatives of MM patients supports the notion of an inherited genetic predisposition (Frank *et al*, 2016; Kristinsson *et al*, 2009).

The largest study addressing this is a meta-analysis of previous Genome Wide Association Studies (GWAS) associated with a replication series, that included over 9,974 MM cases and 247,556 controls of European ancestry. These data, identified, 23 MM risk loci. Cross-referencing these loci with gene expression, epigenetic and in situ High-throughput Chromatine (Hi-C) data suggested they disrupted developmental transcriptional regulators, compatible with altered B-cell differentiation, autophagy/apoptosis, and cell cycle signalling. Nonetheless, these 23 loci, only explained an estimated 16% of the SNP heritability for MM in European populations (Went *et al*, 2018). Recently these 23 SNP have been validated as a risk factor for MM in an independent MM case-control study, and were found to be associated with the risk of developing MGUS (Clay-Gilmour *et al*, 2020).

Furthermore, inherited differences in the risk of developing MM are well recognised. Indeed, there is a greater prevalence and younger onset of MM in African Americans as compared with those with European ancestry. Thus far, there has only been limited evaluation of the possibility that different SNPs were responsible for these differences in phenotype. Rand et al, identified three loci in the African American and European population 7p15.3, 17p11.2, and 22q13.1, and a variant on 3p22.1 that was associated in European ancestry only (Rand *et al*, 2016).

Overall, neither lifestyle nor inherited SNP, to date, explain the entirety of the risk associated with plasma cell disorders, suggesting such a risk is likely multifactorial.

1.2.3.4 Aging

Plasma cell disorders are diseases of the elderly with a median age of 74 years. Paediatric cases are exceptional, (Pilbeam & Lund, 2017) and tend to share more similarities with plasmablastic lymphoma (HIV-associated, bulky extramedullary disease

etc.) than MM. There seems therefore to be a close relationship between age and plasma cell disorders.

Ageing is a term used to describe a correlated set of declines in function with advancing chronological age. It can be seen as the primary cause of many chronic diseases of later life, including Alzheimer disease, chronic kidney disease, coronary artery disease, type II diabetes and even cancer. Malignant plasma cells and aged cells are nonetheless fundamentally opposed as malignant plasma cells can be thought of as hyperactive, rapidly dividing cells with advantageous mutations, and increased energy consumption, while aged cells are hypoactive, unable to divide, with slow metabolism, and an accrual of disadvantageous mutations. Nonetheless the ageing molecular processes and plasma cell disorders are interconnected both in time and space. There are however very few studies looking at the relationship between ageing molecular processes and the genesis of plasma cell disorders.

1.2.3.2.1. DNA instability

DNA damage is defined as a deviation from the normal chemical structure of DNA, which is different from mutations (i.e., alterations of the information content). Unrepaired DNA damage can give rise to genomic instability and induce signalling cascades leading to senescence or cell death, both of which are part of the aging phenotype (Rodier & Campisi, 2011). However, if unrepaired, these chemical modifications may ultimately undermine genetic integrity, and eventually be hijacked for survival by malignant cells.

These lesions are believed to appear early through life. Natural selection processes protect against these events. Highly toxic lesions are unlikely to accumulate, unless they present some survival advantage. Overall, this leads to genetic mosaics with variable mutation rates between cells and tissues based mainly on cell division rates. Once amplified these mutations may cause age-related disorders such as MM. Not all the genome has the same instability. Telomers, microsatellite, retro-transposon, and micro-nuclei are hotspots for these events. Mitochondrial DNA, which is highly sensitive to Reactive Oxygen Species (ROS), also undergoes a high frequency of mutations but the many copies of them compensate for the frequent loss of function mutations (Yuan *et al*, 2020).

In malignant plasma cells, aberrant DNA repair pathways are involved in disease onset (primary translocations), and MM progression (secondary translocations and mutations) (Morgan *et al*, 2012). It is not clear whether the rate of these defects are increased depending on the stage of malignancy. They can be measured on whole genome sequencing approaches using mutational signatures (Alexandrov *et al*, 2015), **Annex 3**. If this phenomenon affects malignant plasma cells, it also affects the other cells in the hematopoietic compartment. Indeed, Haematopoietic Stem Cell (HSC) are particularly sensitive to the cumulative effects of DNA damage. When this happens in the HSC compartment it is called “clonal haematopoiesis”(CH). Mutations in genes involved in epigenetic regulation (*DNMT3A*, *TET2*, *ASXL1*) account for the majority of CH. These mutations are rare in the young but highly prevalent in the elderly, with between 10 and 20% of those older than age 70 harbouring a significant clone (> 4% of nucleated blood cells). Its impact on myeloma remains unknown although several recent studies have highlighted the important crosstalk between aging plasma cells and the myeloid compartment (Pioli *et al*, 2019).

1.2.3.2.2. Telomere biology

The telomere -derived from the Greek nouns τέλος "end" and μέρος "part"- is a region of repetitive hexameric nucleotide sequences at each end of a chromosome, which protects chromosomes from deterioration and fusion with other chromosomes.

The telomerase complex (TC) adds hexameric nucleotide repeats (TTAGGG) to the 3′-hydroxyl end of the telomeric leading strand, using a specific sequence in the RNA component as the template known as *TERC*. It is composed of the enzyme telomerase reverse transcriptase (*TERT*), its RNA component (*TERC*), the protein dyskerin (encoded by *DKC1*) and other associated proteins (*NHP2, NOP10*, and *GAR1,* Pontin and Reptin (encoded by *RUVBL1* and *RUVBL2*)). They are associated with chromatin-remodelling complexes and ensure the regulation of the transcription repertoire, DNA damage repair, and telomerase activity. The Pontin/Reptin complex are essential components for the assembly of the complex (Huber *et al*, 2008). The loading of the complex to the TC is mediated by *SRSF11* through an interaction with TERC and TRF2, and provides a

potential target for modulating telomerase activity in cancer (Lee *et al*, 2015), **Figure 1.2.3.2.2.1**.

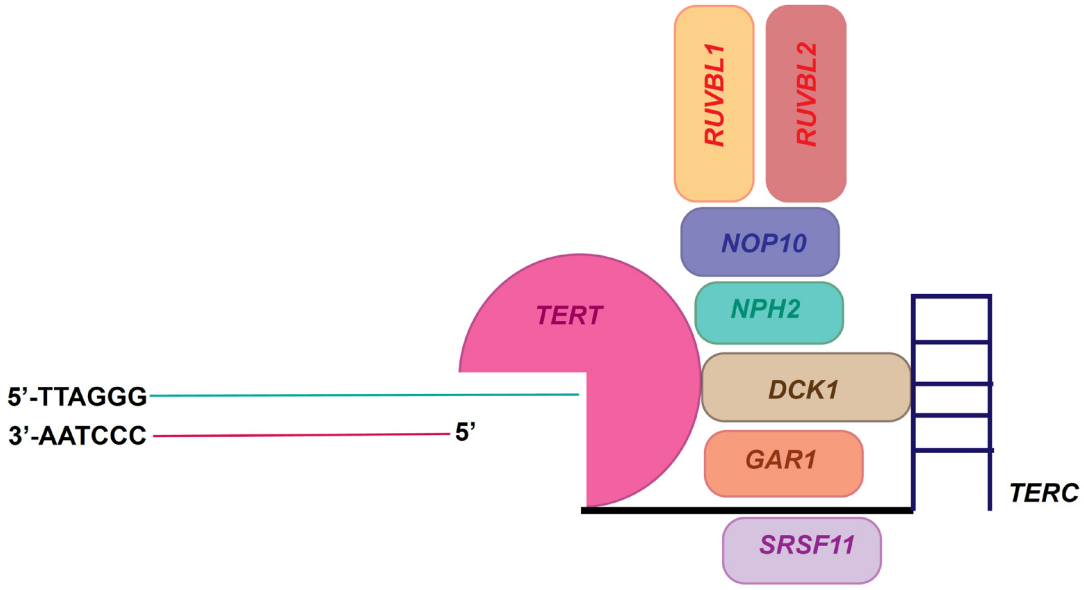


Figure 1.2.3.2.2.1. Telomerase complex. The enzyme telomerase reverse transcriptase (TERT), its RNA component (TERC), the protein DCK1, and other associated proteins (NHP2, NOP10, and GAR1) are shown. They are recruited by the Pontin/Reptin complex and the loading of the complex is mediated via *SRSF11*

The regulation of the TC complex is ensured by multiple protein complexes, **Figure 1.2.3.2.2.2.** Shelterin (also known as telosome) protects telomeres from aberrant DNA repair mechanisms, and regulate telomerase activity. This protective complex is formed proteins associated with telomere DNA: TRF1 (*TERF1*), TRF2 (*TERF2*), RAP1 (*RAP1*), POT1 (*POT1)*, TPP1 (*ACD*), and TIN2 (*TERF2IP*)**.** By connecting the double-stranded DNA TRF1 and TRF2 to the single-stranded DNA-binding unit TPP1, TIN2 (*TINF2*) plays a vital role in both the assembly and function of the shelterin complex. There are many other regulatory complexes that act in close relationship with other DNA repair pathways, cell cycle regulators, and apoptosis, **Figure 1.2.3.2.2.2.**

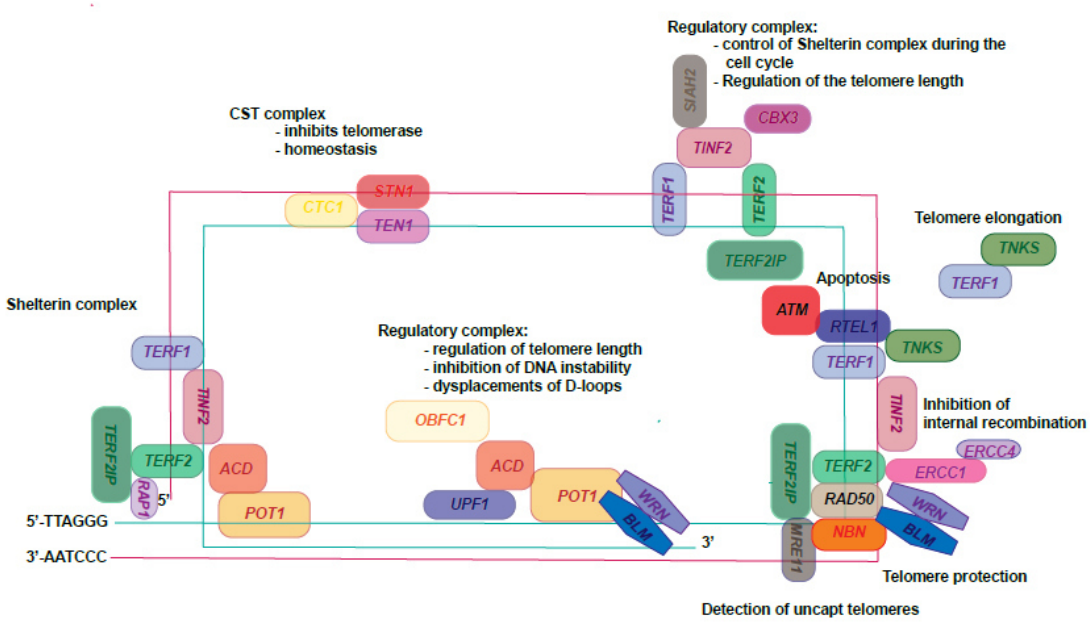


Figure 1.2.3.2.2.2. Regulation of the telomere complex (adapted from (Dreesen *et al*, 2007))

In MM, there is a certain heterogeneity in telomerase expression ranging from levels comparable to normal plasma cells (30% of cases) to levels greater than those of paediatric cancer cell lines (13%of cases). Telomerase activity correlates with poor prognostic factors and to a lesser extent short tumour telomeres. High telomerase activity and short tumour telomeres can be used to a subgroup of patients with poor prognosis (Wu *et al*, 2003). There is also evidence supporting the idea that tumour telomere length are longer in MGUS than in MM. (Panero *et al*, 2010; Yu *et al*, 2019). Relationship with ethnicity and other risk factor remains unexplored.

1.2.3.2.3. Immunosenescence and immune evasion

Cellular senescence is the process in which a cell irreversibly stops dividing and undergoes phenotypic alterations including chromatin changes and alterations in the secretome. It can be triggered by various stress mechanisms such as telomere shortening, oncogene activation, DNA damage response (DDR) pathway and activation

of p16INK4a. The critical actors of cellular senescence seem to be Retinoblastoma 1 (Rb1) and p53, both of which act as tumour suppressor (McHugh & Gil, 2018).

Senescence works as an anticancer mechanism by inhibiting the proliferation of cells with telomere attrition and cells overexpressing oncogenes (Bernadotte *et al*, 2016). This suppressive effect is supported by the many senescent cells found in haematopoietic tissue in MGUS samples (Wang *et al*, 2018). Senescence can also have tumour promoting effects since senescent cells are also active secretors of pro inflammatory cytokines and growth factors. These secretions have been implicated in both aging processes and oncogenesis. Senescent cells are usually eliminated by immune surveillance and phagocytosis. Yet, they accumulate gradually with age, most likely due to an increased production and as decreased elimination secondary to immunosenescent mechanism (Hellmich *et al*, 2020).

The quality and quantity of the immune response, especially when considering T and B-cells, are significantly reduced in quantity and quantity in the aging population(Frasca & Blomberg, 2011). This leads to an insufficient immune response against newly encountered antigens such as neoantigens and foreign pathogens. This is of particular importance when considering IMIDs, monoclonal antibodies or CAR-T cell therapy in the elderly.

Overall, plasma cell disorders are not a “one cause” type disease and arise following multiple environmental, infectious, or physiological aggressions. They also have an increasingly variable clinical presentation. For the sake of time, we will focus on AL, MM, and SMM from this point forward.

1.3. Clinical consequences

1.3.1. Light-Chain amyloidosis

1.3.1.1. Definition

Amyloid refers to abnormal fibrous, extracellular, proteinaceous deposits. The deposits are insoluble and structurally dominated by a β-sheet structure. They are present in a wide variety of disorders called amyloidosis (Rambaran & Serpell, 2008).

The disease nomenclature is based on the nature of the amyloid fibrils for which at least 31 subtypes have been identified (Sipe *et al*, 2014). The abbreviation for amyloidosis is constructed as follows: the first letter “A” stands for amyloid; the suffix designates the precursor protein e.g.; AL is amyloid derived form a light-chain. Other systemic forms include reactive (AA with serum A) and senile amyloidosis. AA is believed to be a secondary condition accompanying chronic inflammation. AL, on the other hand, relates to the presence of a light chain and thought to be one of one of the most common form of systemic amyloidosis (Hemminki *et al*, 2012).

The key event in the pathogenesis of AL is the unstable secondary or tertiary structure of a monoclonal immunoglobulin light-chain, that assembles into monomers, that stack together, form fibrils and precipitate in the extracellular compartments alongside the serum amyloid protein (SAP)(Merlini & Bellotti, 2003). They progressively accumulate throughout the body and form toxic protein aggregates leading to organ dysfunction and eventually organ failure explaining a vast array of clinical symptoms such as congestive heart failure, kidney failure and neuropathy (Merlini & Bellotti, 2003).

1.3.1.2. Epidemiology

1.3.1.2.1. Incidence and Survival

Epidemiological studies of AL are scarce and a correct assessment of incidence is challenging as there was no ICD code until the most recent revision of the ICD-10-CM in 2018 and the creation of the E85.81 light chain AL code, **Table 1.3.1.1** (WHO).

E85.0 Non-neuropathic heredofamilial amyloidosis
E85.1 Neuropathic heredofamilial amyloidosis
E85.2 Heredofamilial amyloidosis, unspecified
E85.3 Secondary systemic amyloidosis
E85.4 Organ-limited amyloidosis
E85.8 Other amyloidosis
E85.81 Light chain (AL) amyloidosis
E85.82 Wild-type transthyretin-related (ATTR) amyloidosis
E85.89 Other amyloidosis
E85.9 Amyloidosis, unspecified

Table 1.3.1.1. Current ICD-10-CM classification (2018 revision)

There are nonetheless four studies that have assessed incidence. The first, published in 1992, was based on the analysis of incidence rates and long term trends in Olmsted county (Kyle *et al*, 1992) and estimated the incidence of AL between 0.61 to 1.05/100,000 person-year. The second was published by the National Amyloidosis Center (NAC), UK. They analysed data from deaths certificates and referrals to the British NAC and estimated a minimal incidence rate of 0.3 /100,000-person year. The Swedish registry studies cross matched the incidence of secondary systemic amyloidosis and that of MM and MGUS and suggested an incidence of 0,32/100.000 person-year (Hemminki *et al*, 2012). Finally, a study of the American 'claims' data estimated the prevalence of AL being 1.55/100,000 person in 2007 and 4.05/100,000 person in 2015. The incidence ranged from 0.97/100,000 in 2007 to 1.4/100,000 in 2015. These data are consistent with a stable incidence and an increase in disease prevalence over this recent eight year period (Quock *et al*, 2018).

Additionally, studies have shown that men have a slight but consistent increased rate of AL compared to women (Hemminki *et al*, 2012; Kyle *et al*, 1992; Quock *et al*, 2018). Also, the majority of patients are over the age of 65 years, with a median age at diagnosis of from 73 years (Hemminki *et al*, 2012). Less than 10% of cases occur before the age of fifty (Quock *et al*, 2018). Some evidence suggests, that cardiac AL occurs in slightly younger patients with a median age of 59 years (Dubrey *et al*, 1998).

Based on population based studies and cancer registries, the median overall survival (OS) is believed to be around thirty-six months (Hemminki *et al*, 2012; Pinney *et al*, 2011). In those, diagnosed after 75, OS is estimated at 20 months (Sachchithanantham *et al*, 2015). An excess of early deaths occur in AL for patients with higher dFLC (differential Free Light Chain: involved Free Light Chain-uninvolved Free Light Chain)(Dittrich *et al*, 2017; Kumar *et al*, 2010). The main prognostic factor is the presence of heart involvement that can be divided into three categories depending on blood-pressure (normal or low) and two serum biomarkers [cardiac troponine (cTnT) and N-terminal pro-B-type natriuretic peptide (NT-ProBNP)](Dispenzieri *et al*, 2004). Based on these Kumar et al, defined a revised staging system where patients were assigned a score of 1 for each of dFLC ≥ 180 mg/L, cTnT ≥ 0.025 ng/mL, and NT-ProBNP ≥ 1,800

pg/mL, creating stages I to IV with scores of 0 to 3 points, respectively. The proportions of patients with stages I, II, III and IV disease were 25%, 27%, 25%, and 23%, and their median OS from diagnosis were 94.1, 40.3, 14, and 5.8 months, respectively (Kumar *et al*, 2012). Other putative adverse prognostic factors include: high von Willebrand factor (VWF) as a surrogate for endothelial dysfunction (Kastritis *et al*, 2016), high GDF-15 levels (Growth Differentiation Factor 15) as a surrogate marker for heart and kidney damage (Kastritis *et al*, 2018), osteopontine as a surrogate marker for heart disease (Kristen *et al*, 2014), percentage of plasma cells in the BM (Kourelis *et al*, 2013; Paiva *et al*, 2011; Perfetti *et al*, 1999), coexistant MM (Kourelis *et al*, 2013), gain(1q21) (Bochtler *et al*, 2014), and malnutrition (Caccialanza *et al*, 2014). The impact of t(11;14) is debated (Bryce *et al*, 2009; Hammons *et al*, 2018; Warsame *et al*, 2015) but seems unfavourable for most authors. On the contrary, 14q gain could be associated with a favourable outcome in one study (Granzow *et al*, 2017).

In comparison to MGUS, the standardized incidence rate is 3.76/100,000 inhabitants, increases regularly with age, and the median OS is similar to an aged-matched MGUS-free population. The mean age at MGUS diagnosis is 72 years (Cabrera *et al*, 2014). In the UK, a study by the Haematological Malignancy Research Network (HMRN) showed an annual incidence rate of 6.2 per 100,000 (2004-2008)(Smith *et al*, 2010). The Mayo clinic found similar rates (Kyle *et al*, 1992; Varettoni *et al*, 2010). Regarding MM, the age-standardized incidence has been reported to be approximately 3-5 cases per 100,000 (Phekoo *et al*, 2004; Sant *et al*, 2010). The median age at diagnosis is 66-70 years with a 5-year OS rate estimated at 50.7% (SEER data, 2018).

In summary, the incidence of AL is roughly between 0.3-1.4/100,000 per year and is believed stable. This incidence is 5-10 times less than that of MM and MGUS. The median age at diagnosis is similar to MGUS and MM. There is a small male predominance in AL. Finally, the outcome of patients with AL is far worse than both MGUS and MM combined.

1.3.1.2.2. Susceptibility and Risk Factors

Although it is estimated that ten percent of patients with MM may develop concurrent AL (Rajkumar *et al*), risk factors for AL are vastly unknown and to some extent

speculative. To gain insight into susceptibility to AL, a gene wide association study was performed on 1229 AL patients from Germany, the United-Kingdom and Italy, and 7526 healthy local controls. Single nucleotide polymorphisms (SNPs) at ten loci showed evidence of an association consistently across the datasets. Some of the SNPs were previously documented to influence MM risk, such as a SNP at the *IRF4* binding site. In AL, rs9344 at the splice site of *CCND1*, promoting t(11;14), reached the highest significance, $p=7{,}80\times10^{-11}$; despite being marginally significant in MM. This may be explained by the high prevalence of t(11;14) in AL. SNP rs79419269 close to the *SMARCD3* gene which is involved in chromatin remodelling was also significant (da Silva Filho *et al*, 2017).

1.3.1.3. Pathology and proteomics

1.3.1.3.1. Pathology findings

The diagnosis of amyloid is based on the light microscopy finding of extracellular, acellular, amorphous, silver-negative, pale pink on hematoxylin and eosin stain, cotton candy-like material (Bain, Barbara *et al*, 2011). Diagnosis is confirmed by Congo-red staining, with characteristic dichroism and apple-green birefringence under polarized light. Thioflavin T, which binds to β-sheet rich structures, is another stain which can also be used to visualise the material, although it is sensitive, it is less specific than Congo red (Khurana *et al*, 2005). In most cases diagnosis is made on minor salivary gland or abdominal fat biopsy. When these tissue biopsies fail to identify any deposits, a biopsy of the clinically affected organ is considered. Electron microscopy may help identify the presence of rigid, non-branching fibrils 7.5 to 10 nm in diameter (Merlini & Bellotti, 2003). Despite having heterogeneous structures depending on aetiology, these differences are morphologically indistinguishable.

The persistence of Congo-red positivity after treatment with potassium permanganate is evocative, but not specific of AL (Noel *et al*, 1987). Furthermore, the presence of a serum or urine paraprotein is not sufficient to establish the diagnosis of AL, because of the frequency of monoclonal gammopathies in patients aged over 50 (Cabrera *et al*, 2014).

Amyloid typing is currently achieved by a multidisciplinary diagnostic approach and requires immunohistochemistry (IHC) and immunofluorescence (IF) studies, clinical evaluation, biochemical tests, and sometimes functional imaging. When applied to FFPE (Formalin-fixed paraffin embedded) samples, IHC lacks specificity and sensitivity (Murphy *et al*, 2001). IF microscopy on frozen sections increases the chance for an accurate typing by lowering background noise but still fails to type 14% of cases on renal biopsies (Picken, 2007). Immuno-gold electron microscopy is a technique which combines immunohistochemistry with electron microscopy and ensures that antibodies react with the amyloid and not the surrounding tissue. It is mainly used for research and shares all the limitation with IHC (Herrera *et al*, 1986).

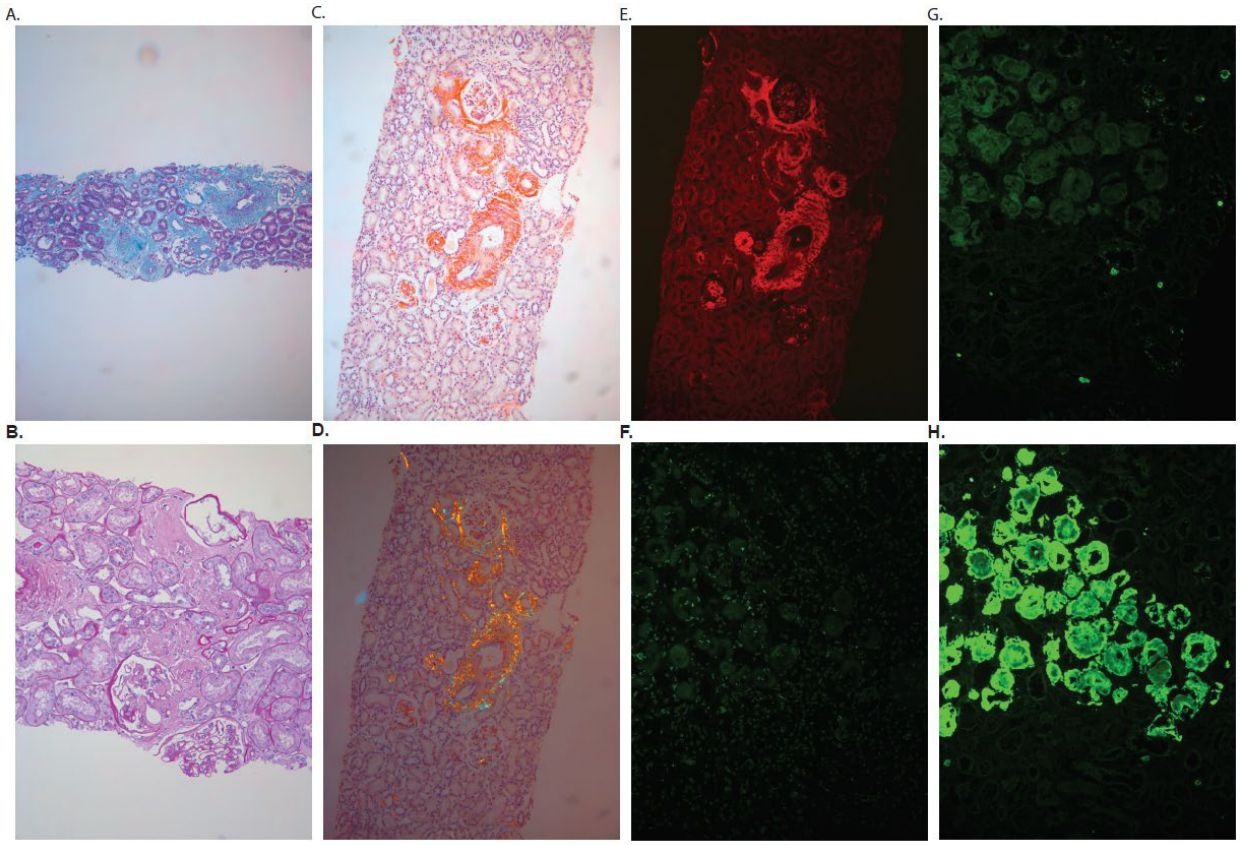


Figure 1.3.1.3.1. Morphological and immunofluorescence study of an AL sample (courtesy of Dr JB Gibier, Lille, University Hospital). A. Masson's trichrome. B. Periodic acid–Schiff staining. C. Congo Red. D. Congo Red and Polarised light. E. CR stain F. Fluorescein G Kappa IF H.Lambda IF.

Unlike previous methods, proteomics focuses on the entire proteome eliminating the need to test for the individual protein sequentially. It is a non-biased, comprehensive, tissue-sparing, approach that is readily applicable in a routine setting, and allowing rapid typing of most cases. Its utility has been enhanced by micro-dissecting the deposits therefore enhancing the signal (Lavatelli *et al*, 2011).

1.3.1.3.2. Proteomic findings

Isolation of the protein components were key in the understanding of the biology of amyloidosis (Glenner *et al*, 1971).

All amyloid deposits contain SAP*,* a glycoprotein member of the pentraxin family, which binds amyloid independently of the protein of origin (Baltz *et al*, 1986). SAP is a stable, protease resistant, circulating protein, that has a specific binding motif that recognises the common conformation of amyloid fibrils (Pepys *et al*, 1994). Proteoglycans are also common in amyloid and constitute the carbohydrate composition of amyloid deposits (Kisilevsky, 1990). Heparan sulfate proteoglycans in particular have a similar kinetics of deposition in tissues to that of fibrillary proteins (Snow *et al*, 1987), and localize with elements of the extracellular matrix such as lamilin, and type IV collagen. These constitute the scaffold that facilitates the initial phase of fibril nucleation (McLaurin *et al*, 2000).

One of the most important features lies in the property of these deposits to aquire several conformation (Perutz, 1997). The actual mechanism remains to be understood. However, there is a general agreement that the protein monomers are bound to each other by hydrogen bonds to form stable beta-sheets. These monomers are oriented perpendicularly to the fibril axis and twisted around each other to form amyloid fibrils. This structure is shared by all amyloid regardless of their nature (Glenner, 1980). The conversion of the structure of the native protein into a predominantly antiparallel β-sheet secondary structure is a pathological process related to the physiological folding process. At a similar energy, some polypeptides can acquire an alternative misfolded and relatively stable state prone to aggregation (Schultz, 2000). Once secreted, they continue the final part of the folding process, forming either a normal or misfolded protein. (Lavatelli *et al*, 2011).

In AL, only a small proportion of light-chains are believed to be amyloidogenic. This is believed to be related to the variable (V) chain. Lavatelli and colleagues also showed a role for both the joining and constant region (Lavatelli *et al*, 2011). Other studies demonstrated that the deposits contained a significant part of the constant region in virtually every case, suggesting that the constant region was also necessary for the generation of the fibrils (Theis *et al*, 2013). When or why truncation occurs is unclear. Lysosomes have also been implicated in amyloid formation. After endocytosis from the mesangial cells, amyloidogenic light chains are directed toward the lysosomal compartment. This acidic environment, participates in the fibril formation before they are excreted in the extracellular matrix (Teng *et al*, 2004).

Finally, the study of urinary exosomes among patients identified high molecular weight immunoreactive proteins corresponding to decamer were found in AL absent from other paraprotein related glomerular disease (Ramirez-Alvarado *et al*, 2012, 2012). These oligomeric light chain species may represent the initial steps of amyloidogenesis but further studies are required to better understand their potential.

1.3.1.3.3. Mechanism of tissue damage

Pronounced deposits may replace normal tissue, subvert architecture, and influence the mechanical organ function. They may also affect the exchange of nutrients and other molecules between cells and the blood stream. However, there are many studies indicating that this does not account for the toxic effect of the deposits. For instance, it has been observed that similar amounts of cardiac amyloid deposits, evaluated by echocardiography, have significantly worse functional and prognostic impact in patients with light chain amyloidosis than in patients with transthyretin amyloidosis (Dubrey *et al*, 1997). Evidence from mouse models suggest a direct toxic effect of the amyloidogenic light-chain through an increase in cellular oxidative-stress leading to endothelial dysfunction (Brenner *et al*, 2004; Migrino *et al*, 2010, 2011; Sikkink & Ramirez-Alvarado, 2010). This would suggest that infiltration itself is not solely responsible for organ dysfunction.

1.3.1.4. Underlying disease

The underlying disease is a plasma cell disorder likely a monoclonal gammopathy of undetermined significance (MGUS), smouldering myeloma (SMM) or symptomatic myeloma (MM). Cases have been described with other lymphoproliferative disorders (LPD) but are less common(Sanchorawala *et al*, 2006).

Limited data are available on the biology of the underlying plasma cell clone. Most studies have concentrated on chromosomal abnormalities; however, recent studies have focused on other aspects of the disease biology such as mutations and gene expression.

1.3.1.4.1. The genetic makeup suggests high predominance of t(11;14) and low incidence of hyperdiploidy

AL, like light chain only MM (Avet-Loiseau *et al*, 2003a), has previously been associated with t(11;14) translocations present in 37-63% of cases depending on the series (Bochtler *et al*, 2011, 2015; Bryce *et al*, 2009; Granzow *et al*, 2017; Harrison *et al*, 2002; Warsame *et al*, 2015) and are believed to be early events (Hayman *et al*, 2001). Other canonical *IGH* translocations have been described but are less frequent such as t(4;14) (0-4%) and t(14;16) (0-5%) and are underrepresented in comparison to MM cohorts (Avet-Loiseau *et al*, 2012; Walker *et al*, 2013a). Some series also report a high incidence of *IGH* translocations with no partner ranging from to 9-20% of patients in some iFISH cohorts (Bochtler *et al*, 2011; Warsame *et al*, 2015). This could account for either technical limitations, non-canonical translocations or/and break-points. Non-canonical translocations have been described in case reports such as t(1;20) (q21;q11) (Otokida *et al*, 1990) with no further information regarding suspected partner genes or fusion transcripts.

The incidence of HRD is lower in amyloidosis than in myeloma ranging from 11-19% (Wuilleme *et al*, 2005). Copy number arrays estimate the frequency of HRD at 19% (Granzow *et al*, 2017). Del(17p) is believed to be less frequent in AL than in MM ranging from 0-4% by iFISH and 5% on copy number arrays. The frequency of gain(1q), when reported is similar to the MM cohorts (24-36%). Additional attention has been paid to del(14q) seen in 19% of patients (Granzow *et al*, 2017) which has a lower incidence in AL than in MM (Walker *et al*, 2010a) and may have prognostic significance.

When considering risk calssifications such as the mSMART classification (Dispenzieri *et al*, 2007), 91% of AL-MM patients would be standard risk, 2% intermediate and 6.5% high-risk. For AL-SMM patients, the risk groups would break down as follows with 19%, 27%, 8,5% and 4,5% being considered low, standard, intermediate, high risk respectively (Warsame *et al*, 2015).

In summary to date studies show that there is no disease defining copy number change or translocation in AL. The pattern of copy number changes and translocation are similar to those described in MM. The frequency, however, of changes vary in comparison to MM and appear consistent across the two main datasets (German and Mayo) with more cases harbouring a t(11;14), and fewer cases having either hyperdiploidy, del(17p), or del(14q). A summary of this information may be found in **Table 1.3.1.4.1.**

1.3.1.4.2. The full mutational landscape remains to be determined

To date only five cases have been sequenced (Paiva *et al*, 2016) and no unifying mutations were identified. Ninety-four single nucleotide variants were identified in 93 genes with a median of 15 variants per patient. Sixty-seven percent of them had previously been described in MM. The only recurrently mutated gene in this series was *XKR8,* a Kell blood-group complex subunit related family member that promotes phosphatidylserine exposure on apoptotic cells by mediating phospholipid scramble. One case of *KRAS* (G12V) mutation was seen.

Targeted panel approaches, based on myeloma targeted panels, have also been published (Kim *et al*, 2016; Rossi *et al*, 2017). In the largest series comparing AL to MGUS and MM, the frequencies of the most mutated genes were as follows: *BRAF* (16.7%), *TP53* (12.5%), *NRAS* (12.5%), *KRAS* (4.2%), and *CCND1* (4.2%). The incidence of these were not statistically different from those expected in MM and MGUS although the number of mutations per sample was significantly smaller in AL cases than MM suggesting a less complex disease process. In a study led by Kim et al, twelve AL samples were tested for mitogen-activated protein kinase (MAPK) pathway mutations (*NRAS*, *KRAS* and *BRAF*). These results suggested that MAPK pathway mutations would only occur in AL patient that met criteria for symptomatic MM (Kim *et al*, 2016).

In summary there does not seem to be any disease defining mutation in AL. The majority of genes mutated in AL are shared with MM, although the datasets are small and results should be interpreted with caution.

1.3.1.4.3. V(D)J rearrangements and somatic mutations may be key determinants of AL

1.3.1.4.3.1. Light chain repertoire

Cloning and sequencing studies of the V(D)J light chain rearrangement in AL identified that the selection of Vλ family gene was biased toward certain gene segments (Abraham *et al*, 2007; Comenzo *et al*, 2001; Ozaki *et al*, 1994; Perfetti *et al*, 2012; Solomon *et al*, 1982). There was a significant overexpression of two germline genes *IGVL6-57* and *IGVL3-1* with a frequency ranging from 20-30% in AL in comparison to 5% in controls (Perfetti *et al*, 2012). These sequence biases are not found in MM (Kiyoi *et al*, 1998; Kosmas *et al*, 1996; Sahota *et al*, 1996), suggesting this bias probably relates to the amyloidogenic potential of the light chain rather than to a general feature of plasma cell dyscrasias.

The specific organ-tropism has to some extent been attributed to the light chain isotype although there is some discrepancies between authors (Comenzo *et al*, 2001; Perfetti *et al*, 2002, 2012). Comenzo et al, in a series of 60 AL cases suggested that the *IGVL6-57* light chains were exclusively associated with kidney disease. This was supported by the amyloid formation from these light-chains cultures with mesangial cell lines. This observation was confirmed by Perfetti et al. Targeting of bone and soft tissue was shown to be more frequent when light chains belonged to the *IGKVI* family of genes (Prokaeva *et al*). Regarding *IGVL3-1* light chains, associations are debated: for some authors they are not associated to any distinctive phenotype (Perfetti *et al*, 2002) for others, they are overrepresented among patients with bone/joint disease (Abraham *et al*, 2003) or major cardiac /multiorgan disease (Comenzo *et al*, 2001). Regarding heart involvement, *IGVL1-44*, was associated with a five-fold increase odds of heart involvement and *IGVL6-57* was underrepresented emphasizing the idea that the variable region are responsible for organ specificity (Perfetti *et al*, 2012).

1.3.1.4.3.2. Somatic mutation

Ig somatic mutations may be found in all AL patients but the overall frequency is smaller than the polyclonal counterparts. This may be explained by the higher frequency of *IGVL6-57* that was more homologous to germline than *IGVL3-1* (Perfetti *et al*, 2012).

Additional data supports the idea that amino acid substitutions in several key sites within the variable regions of the light chain can account for the presence or absence of the amyloidogenic properties (Stevens *et al*, 2000). Taking in consideration that somatic hypermutation adds to the complexity of AL, each patient appears to have an unique protein sequence (Ramirez-Alvarado, 2012). Finally, somatic mutations have a global destabilizing effect on AL proteins: these proteins require less energy to unfold or lose their interaction with the heavy chain (Hurle *et al*, 1994; Stevens *et al*, 1995; Wetzel, 1997). It is therefore probable that a combination of destabilizing and compensatory mutations leads to the formation of fibrils in AL rather than a single mutation (Ramirez-Alvarado, 2012).

1.3.1.4.4. RNA expression

1.3.1.4.4.1. Gene expression studies

Small GEP datasets have been performed and compared to MM (Abraham *et al*, 2005). They identified genes that were differentially expressed such as *TNFRSF7*, *CXCL12* and *PSMA2,* most of which have been associated with tumour progression from normal plasma cells through to MGUS and MM (Abraham *et al*, 2005; D'Apuzzo *et al*; Guikema *et al*). This suggests that the underlining disease is likely to be MGUS or SMM. Based on gene expression and network analysis, attempts were made to explain the greater number of lambda light chain isotype (Gertz, 2013; Kyle *et al*, 1992) and the restricted use of select light chain germ-line genes (Abraham *et al*, 2003; Comenzo *et al*, 2001; Perfetti *et al*, 2012) by a potential role in the regulation of V(D)J rearrangements through the *CCND1-CDK4-Rb* axis (Abraham *et al*, 2005).

1.3.1.4.4.2. miRNA

miR-16 and several others miRNAs have been shown to be deregulated in AL (Weng *et al*, 2011). In chronic lymphocytic leukaemia, miR-16 has been shown to be a

negative regulator of BCL2 with a loss of activity through deletion of 13q14 (Pekarsky & Croce, 2015), a feature shared with plasma cell dyscrasias (Avet-Loiseau *et al*, 1999)

Author	Method	n	% abnormal	Translocation	13-/ 13q14-	HRD	17p-	1q+	14q-
Harrison, 2002	iFISH	24	-	t(11;14) 37%	33%	-	-	-	-
Bryce, 2009	iFISH	54	70%	t(11;14) 39% t(4;14) 0% t(14;16) 2%	30%	-	0%	-	-
Bochtler, 2011	iFISH	246	95%	t(11;14) 53% t(4;14) 4% t(14;16) 1%	35%	11%	2%	24%	-
Warsame, 2015	iFISH	401	81%	t(11;14) 43% t(4;14) 2% t(14:16) 3%	30%	12%	2%	NR	NR
Bochtler, 2015	iFISH	133	-	t(11;14) 63% t(4;14) 4% t(14:16) 5%	32%	19%	4%	27%	-
Granzow, 2017	iFISH and high-density copy number arrays	118	-	t(11;14) 62% t(4;14) 3% t(14:16) 2%	29% and 12%	19%	5%	36%	19%

Table 1.3.1.4.1. Summary of the main cytogenetics published data in AL

1.3.2. Multiple Myeloma

1.3.2.1. Definition

Multiple myeloma is the second most common haematological malignancy after non-Hodgkin lymphoma (Myeloma - Cancer Stat Facts). In most patients, MM is characterized by the secretion of a monoclonal Ig, however, in 15-20% of patients, the MM cells secrete only monoclonal SFLC, and, in less than 3% of patients, these cells do not secrete any monoclonal protein (Bain, Barbara *et al*, 2011). The clinical symptoms of MM are driven by the paraprotein, the malignant cells themselves or cytokines secreted by the malignant plasmacells, and include signs of end-organ damage.

1.3.2.2. Incidence and mortality

Myeloma constitutes 1.5% of all malignant diseases, with an incidence of 6.9/100,000 person per year accounting for 4,500 new cases each year in the UK (representing roughly 2,600 deaths per year)(Home - Office for National Statistics) and 32,270 cases (12,830 deaths) in the USA (Myeloma - Cancer Stat Facts). The median age at diagnosis is 69 years but 34.8% of patients are diagnosed over the age 75, and 9.6% above 85 years (Home - Office for National Statistics), **Figure 1.3.2.2.1**. This prevalence of elderly cases is likely to increase in the near future given the progressive aging of western populations (Ageing Europe — 2019 edition; Population Projections).

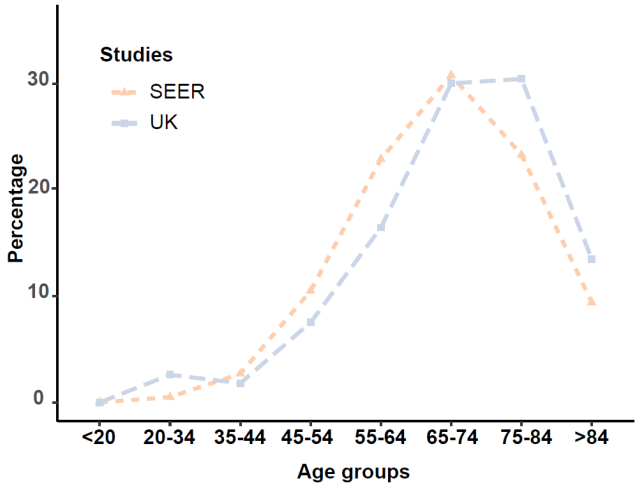


Figure 1.3.2.2.1. Age distribution of MM cases according to the American and British data

Following recent improvements in treatment, the 5-year overall survival (OS) rate for MM patients is approximatively 54%. In older age-groups, the impact of treatment has been less than would have been expected given the improved outcomes seen in younger populations with US population-based data showing median survival figures for the over 65 of only 37-39% at 5-years compared with 57-59% for the 50-64-year-old group. Such age related differences in outcomes are not restricted to the USA with similar UK based population-based data showing a 5-year OS of 14% for the 80-99 age group compared to 26% for the 70-79 age group (Home - Office for National Statistics).

1.3.2.3. Pathology

Malignant plasmacells in MM display many aberrant features that help distinguish them from their normal or reactive counterparts.

Nuclear-cytoplasmic asynchrony is the most distinctive and reliable cytologic characteristic of these cells. Nuclear chromatin is often finely dispersed resembling that of a blast, and the nucleus may be more centrally located, with large, sharply demarcated basophilic nucleoli. Occasionally, inclusion bodies may be seen in the nucleoli. The more immature the chromatin looks, the larger the nucleoli are, the more likely the plasmacell is to be malignant. The nucleus itself can also show some variations; it can be notched, cleaved, multi-lobated, convoluted or even cerebriform. The nucleus can also vary in size, number and some carry multiple nucleoli (Bain, Barbara *et al*, 2011).

Many cytoplasmic variations have been described: some related to inclusions some relating to staining heterogeneity. The most common inclusion bodies are Russell bodies and Dutcher bodies. Russell bodies are eosinophilic, large, homogeneous, immunoglobulin-containing inclusions. They are formed of distended endoplastic-reticulum, where excess proteins are stored, and gradually build-up in the cell. Dutcher bodies are intracytoplasmic spherical Ig inclusions that can invaginate into the nucleus or even overlie it. They are believed to result from the accumulation of Ig in the perinuclear cisternae (Bain, Barbara *et al*, 2011). Other morphological specificities include grape cells which carry many vacuoles that extends to the very edge of the cytoplasm (Stich *et al*, 1955) or flaming cells that have vermillion-staining glycogen-rich overstuffed fibrils (Ranjbaran & Golafshan, 2018).

From a phenotypic perspective, like all plasmacells, they co-expression of CD138 and CD38. Malignant plasmacells can either gain some markers (CD28, CD33, CD56, or CD117) or lose some (CD27). Finally, they are clonotypic, expressing only either kappa or lambda light chains homogeneously (Rawstron *et al*, 2008).

Beyond morphological changes, there is evidence for abnormal topographic arrangements between the tumour cells, the microenvironment and the vasculature. Tumour cells infiltrate fat and haematopoietic tissue. Denser aggregates of malignant cells are seen along endosteal surfaces and around ectatic sinusoids or arteries. With increasing infiltration, MM cells eventually replace the haematopoietic compartment forming dense sheets or nodules.

Moving forward from morphology and flow cytometry, progress in molecular genetics have helped classify MM further and refined our understanding of the disease.

1.3.2.4. Genetics of myeloma

A better understanding of the genetics of MM has occurred over the last 20 years where progress in technology has enabled the study of the genetics of MM at several echelons such as the structural, mutational, and epigenetic level. These data have enhanced the classification of the disease, refined our understanding of the biology, and led to substantial progress in patient management.

1.3.2.4.1. Structural variants

1.3.2.4.1.1. Translocations

A translocation is a type of chromosomal abnormality that involves a chromosome break that reattaches to a different part of a chromosome. Classically, we distinguish inter and intra-chromosomal translocations depending on the partner (same or different chromosomes). Translocations can disrupt topologically associated domains (TAD). TADs represent functional domains as a given TAD encompasses the regulatory elements for the genes inside the same domain. TAD rearrangements rewire the regulatory landscape of genes, which can lead to gene dysregulation, **Figure 1.3.2.4.1.1.** This is a common mechanism in MM, where TAD-TAD rearrangements result in the hijacking of a strong enhancer (usually an Ig or *MYC* enhancer) by an oncogene.

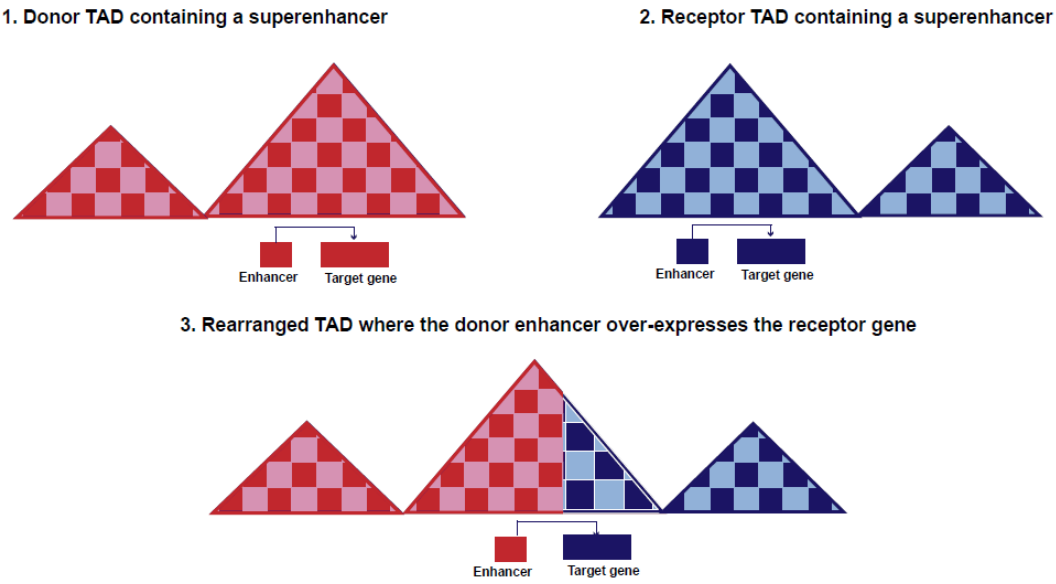


Figure 1.3.2.4.1.1. Schematic representation of a TAD-TAD rearrangement where enhancer from one TAD are hijacked by another gene following a TAD-TAD rearrangement

Over the years, a set of canonical inter-chromosomal translocations have been described in MM, **Table 1.3.2.4.1.1.** They are considered initiating events. They all have in common the hijacking of an Ig super enhancer by an oncogene through aberrant CSR or VDJ rearrangement. The main partners are cyclin D1 (*CCND1*), cyclin D3 (*CCND3)*, fibroblast growth factor receptor 3 (*FGFR3*), multiple myeloma SET domain (*MMSET*; also known as *NSD2*), musculoaponeurotic fibrosarcoma oncogene homolog (*MAF*) and *MAFB* (Walker *et al*, 2013b).

Some of these key molecular events lead to the deregulation of the G1/S transition. The consistent deregulation of the D group cyclin was first noted as a consequence of studying t(11;14) and t(6;14) translocations, which directly deregulate cyclin D1 and cyclin D3, respectively. This concept has been adapted in the TC classification of MM based on gene expression profiles (Zhan *et al*, 2006).

1. Introduction

The t(4;14)(p16.3;q32.3) is found in 15% of newly-diagnosed cases and is classically associated with a worse outcome although the outcome of these patients has improved in the past decade with novel treatment (Neben *et al*, 2010). As a consequence of the translocation, two genes are aberrantly expressed: *FGFR3* and *NSD2*, both of which have potential oncogenic activity (Mirabella *et al*, 2014). Importantly, *FGFR3* shows only weak transforming activity and is eventually lost in 30% of patients, suggesting that it is not the main oncogenic factor (Kalff & Spencer, 2012). In contrast, *NSD2* gene overexpression is universal. *NSD2* is a histone methyl transferase that is responsible for deposition of the H3K36 mono-methyl and di-methyl mark. In a wild-type setting, H3K36me2 accumulates on active gene bodies and acts as a signature of transcriptional activity. However, when *NSD2* is overexpressed as a result of the t(4;14), H3K36me2 spreads outside of active gene bodies into intergenic regions, resulting in the reduction of H3K27me3 domains, thus altering the cellular gene expression programs (Popovic *et al*, 2014).

Finally, the *MAF* and *MAFB* translocations constitute approximatively 5-10 % of MM and associate with intrinsic chemoresistance and, as a result, a dire outcome. A high proportion of patients have high-risk disease, extramedullary disease (Usmani *et al*, 2012) or primary plasma cell leukaemia (Schinke *et al*, 2020). Eighty percent of cases carry a dominant APOBEC profile (Walker *et al*, 2015b) associated with an increased mutational burden and frequent *ATM* and *ATR* mutations. The outcome of these cases has not improved since the introduction of bortezomib which is in sharp contrast to t(4;14) patients (Qiang *et al*, 2016).

Translocation	Frequency	Specificity
t(4;14): *FGFR3 NSD2/IGH*	12-15%	Adverse prognosis in the pre novel agent era
t(11;14): *CCND1/IGH*	15-20%	Neutral prognosis in MM More frequent in plasmacell leukaemia
t(6;14): *CCND3/IGH*	1%	Unknown
t(8;14): *MAFA/IGH*	1%	Unknown
t(14;16): *MAF/IGH*	2-3%	Adverse prognosis Associated with a dominant APOBEC mutational signature
t(14;20) *MAFB/IGH*	2-3%	Adverse prognosis Associated with a dominant APOBEC mutational signature

Table 1.3.2.4.1.1. Summary of the canonical translocations in MM

Other translocations are seen in MM and tend to occur later in the disease process. *MYC* is the most common dysregulated gene and leads to a more aggressive disease phenotype. Nonetheless, it is still seen in 20% of MM at presentation (Walker *et al*, 2014a; Mikulasova *et al*, 2020)

1.3.2.4.1.2. Copy number abnormalities

The other major set of recurrent genetic abnormalities seen in MM is hyperdiploidy (HRD), associated with the gain of the odd numbered chromosomes including 3, 5, 7, 9, 11, 15, 19 and 21. The mechanism underlying HRD is much less obvious but has been may involve telomere attrition and in-born mitosis errors (Paulsson *et al*, 2005). There is some evidence supporting the role for chromosome segregation defects occurring during a single catastrophic mitosis (Molina *et al*, 2020). In terms of specific trisomies affecting prognosis, trisomies 3 and/or 5 significantly improve overall survival, whereas trisomy 21 is associated with an adverse outcome (Chretien *et al*, 2015; Perrot *et al*, 2019).

Gain(1q21) is seen in 35-40% of patients and is believed to be a secondary event. It influences progression and survival through dosage effect of oncogenes such as *CKS1B*, *MCL1*, or *SLAMF7*. Amplification of gain(1q) defined as the presence of four or

more copies is an independent prognostic factor (Walker *et al*, 2019). Del(1p) affects 30% of patients and frequently associates with gain(1q). It includes loss of genes such as *CDKN2C*, *FAF1*, and *FAM46C*. Del(1p) is also generally considered an adverse genetic factor (Hebraud *et al*, 2014).

Del(13q) is the most frequent CNA and is seen in 50% of cases, 80% of which are monosomies 13 and 20% segmental CNA with focal loss of *DIS3* and *RB1*. Although it was classically considered high-risk, evidence suggests that it is the closely associated t(4;14) that drives most of its outcome (Walker *et al*, 2010b).

Finally, del(17p) is seen in 10% of cases and is responsible for the loss of the tumour suppressor gene *TP53*. Most of the prognostic significance of this is driven by high clonal fraction or bi-allelic events (Walker *et al*, 2019; Thakurta *et al*, 2019).

1.3.2.4.1.3. Complex structural rearrangement

In other cancers, SVs have been noted to be complex and to involve more than two sites. These complex events have also been noted in MM but their clinical significance is uncertain (Mikulasova *et al*, 2019; Magrangeas *et al*, 2011). Molecular characterisation of complex structural events are known to involve focal copy number changes at the breakpoints. These have been termed chromoplexy when there is associated copy number loss and templated insertions when there is copy number gain (Shen, 2013). A further complex event, chromothripsis (“chromosome shattering”) is defined by localized clustered rearrangements associated with loss of heterozygosity and copy number oscillation (Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing | bioRxiv), **Figure 1.3.4.1.3.**

A recent report, mapping SVs in a whole genome dataset of MM has shown that the incidence of chromothripsis and templated insertion is approximatively 24% and 19% in MM (Rustad *et al*, 2020). These MM related events that frequently use a templated insertion mechanism, can involve hotspots where key driver genes are impacted and can both upregulate and downregulate driver genes. However, much remains to be understood about the full contribution of structural events to the patterns of gene deregulation and progression in MM.

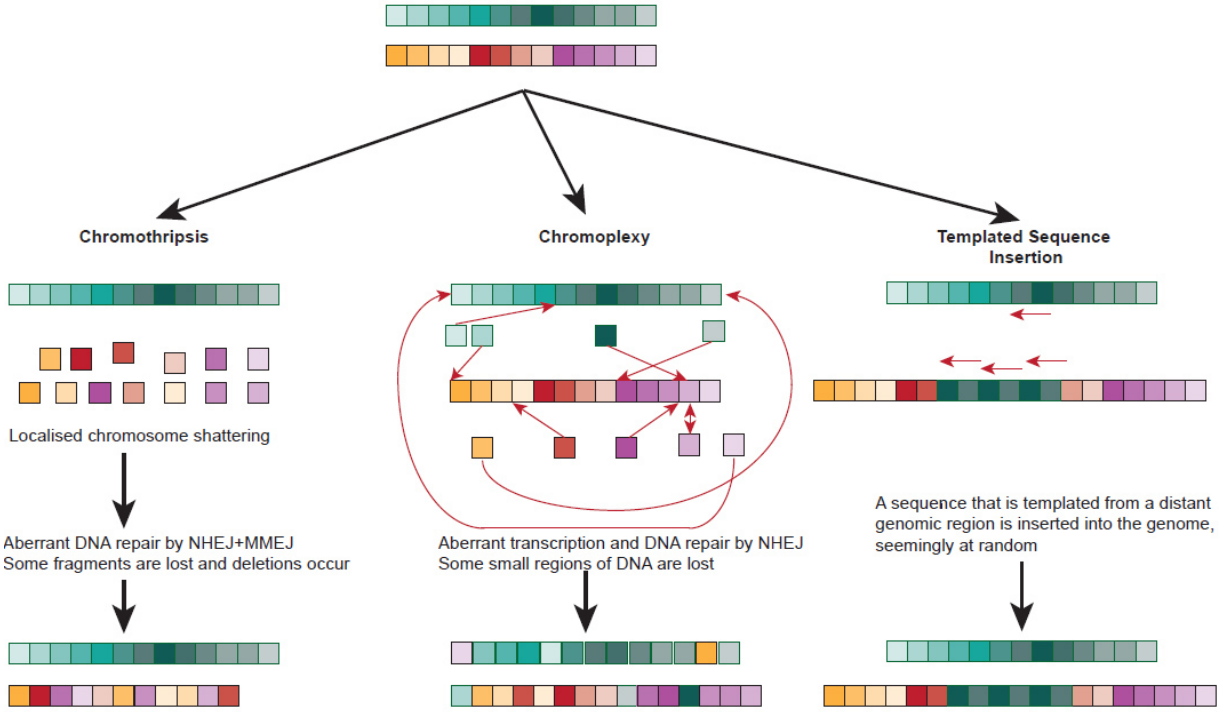


Figure 1.3.2.4.1.3. Complex chromosomal rearrangement (Boyle et al, in press).

1.3.2.4.2. Mutations

1.3.2.4.2.1. Mutational drivers and frequently mutated genes

Several studies have looked at the mutational landscape in MM (Walker *et al*, 2015a, 2018; Chapman *et al*, 2011). The average myeloma sample carries approximatively 50 non-synonymous exonic mutations with some samples carrying more than a thousand. There is no unifying mutation in MM, a feature shared with most haematological malignancies with the exception of Waldenstrom's disease (Treon *et al*, 2012) and Hairy cell leukaemia (Tiacci *et al*, 2011). However, there are recurrently mutated pathway including the Mitogen-activated protein kinase (MAPK) pathway, nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) pathway, and DNA repair pathway.

The most frequently mutated genes are *NRAS* and *KRAS* which are seen in 25 and 20% of patients respectively (Walker *et al*, 2018, 2015a). These mutations are in the

typical hotspots shared with other cancers. Additionally, ten percent of patients also have *BRAF* mutations suggesting the MAPK pathway is crucial in MM. NFκB pathway mutations in eleven genes involved in regulation of the NFκB pathway have been identified in 17-20% of MM tumours. DNA-repair pathway alterations in *TP53, ATM*, and *ATR* are seen in approximatively 10% of patients. (Walker *et al*, 2019).

The largest study to date, identified 63 significantly mutated genes that act as oncogenic drivers, **Figure 1.3.2.4.2.1.** Many are shared with other cancers but some appear MM specific such as *MAX* or *XBP1.*

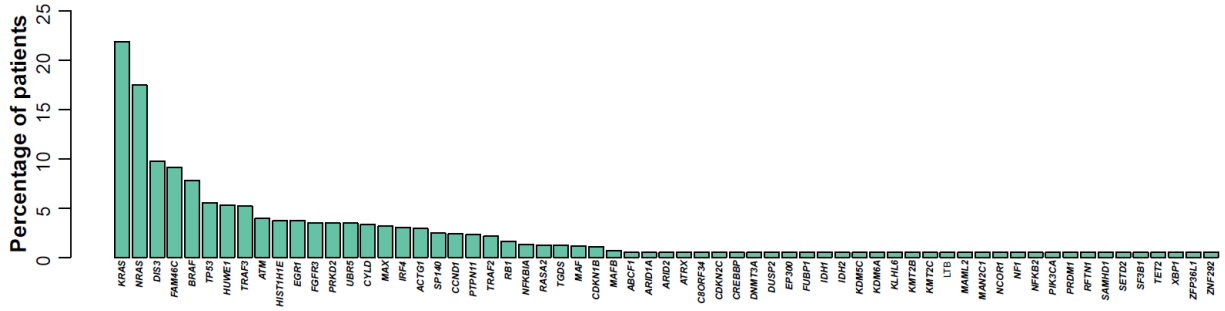


Figure 1.3.2.4.2.1. Frequency of driver genes in the MGP dataset (adapted from (Walker *et al*, 2018))

1.3.2.4.2.2. Co-segregation

Mutations are not evenly distributed across cytogenetic subgroups and frequently co-segregate with other variants. *NRAS* and *KRAS* are believed to be mutually exclusive. HRD is associated with *NRAS* mutations and *MYC* translocations. t(11;14) are associated with *CCND1* and *IRF4* mutations and negatively associated with gain(1q). t(14;16) are associated with gain(1q) del(13q) and *DIS3* mutations. t(4;14) are associated with *DIS3, PRKD2*, and *FGFR3* mutations, and del(12p), del(13q), and gain(1q). t(4;14) are also negatively correlated with *MYC* translocation (Walker *et al*, 2015a, 2018). Combined these data suggest distinct mechanism of disease progression and differences in disease biology that extend beyond the initiating event.

1.3.2.4.3. Epigenetics changes

Genetic modifications have been considered essential in MM, but another type of alterations, the epigenetic events, also contribute to the oncogenic transformation of normal plasma cells by altering gene expression. The methylation of DNA and histones is one of the main physiological processes to down-regulate the gene expression and has been implicated in cancer progression.

The first epigenetic change identified in MM was the global hypomethylation patterns associated with relapsed and refractory disease. The most pronounced DNA methylation changes are seen in the t(4;14). In a wild-type setting, H3K36me2 accrues on active gene bodies and acts as a signature of transcriptional activity. However, when *NSD2* is overexpressed as a result of the t(4;14), H3K36me2 spreads outside of active gene bodies into intergenic regions, resulting in smaler of H3K27me3 domains, altering gene expression programs in the absence of driver mutations. Other histone modifiers deregulated in myeloma include *KDM6A, KDM6B, KMT2A,* and *HOXA9* (Popovic *et al*, 2014; Walker *et al*, 2011; Pawlyn *et al*, 2016).

1.3.2.5. Risk factors

Over the past 20 years, a number of risk-models have been developed in MM from simple models such as International staging system (ISS) to more complex models involving cytogenetics and gene expression profiling.

1.3.2.5.1. International staging system

The International staging system (ISS) for myeloma (Greipp *et al*, 2005b) classifies patients into three groups based on β-2 microglobulin (B2M) and serum albumin (ALB) as follows **(Annex 1)**:

- Stage 1: β2M < 3.5 mg/dL and ALB ≥ 35
- Stage 2: ALB < 35 mg/dL and β2M < 3.5; ALB < 35; or β2M 3.5–<5.5 mg/dL
- Stage 3: β2M ≥ 5.5 mg/dL

The major criticism of ISS was the lack of the use of known other prognostic markers in MM including cytogenetics abnormalities (CA).

1.3.2.5.2. Revised International staging system (R-ISS)

In 2015, Palumbo et al. (Palumbo *et al*, 2015) published the revised international staging system (R-ISS) which combined ISS with CA and LDH as follows (**Annex 1**):

- Stage I: ISS I, standard risk by FISH and normal LDH
- Stage II: not R-ISS I or III
- Stage III: ISS III, either high risk by FISH or high LDH

The high-risk CA considered were the presence by FISH of del(17p), t(4;14) , or t(14;16). R- ISS is a powerful prognostic staging system, but currently its use a daily clinical setting is limited. The major criticism of R-ISS was the lack of the use of known other prognostic markers in such as gain(1q), its reliance on iFISH, and the lack of specificity of LDH levels.

1.3.2.5.3. Plasma cell Labelling Index

Proliferative features of MM such as the plasma cell labelling index (PCLI) developed by the Mayo clinic have been associated with outcome.

PCLI is used as a proliferation index and measures the number of plasma cells in the marrow undergoing the S phases (Gertz *et al*, 1989). Steensma et al., demonstrated that even in clinically stable patients, a high PCLI was associated with a short PFS and OS (Steensma *et al*, 2001). The PCLI was only used in select institutions in the US and has been widely adopted given the more practical prognostic markers available.

1.3.2.5.4. Cytogenetics

In the past decades, chromosomal aberrations (CA) such as t(4;14), t(14;16), t(14;20), amp(1q21), del(1p) and del(17p) have been shown to be associated with poor prognosis mainly in young newly diagnosed patients.

Regarding the translocations, the impact of t(4;14) has been lessened since the introduction of bortezomib (Neben *et al*, 2010) but remains part of many prognostic models (Perrot *et al*, 2019; Palumbo *et al*, 2015). Debate remains regarding the t(14;16) which has been associated with an adverse outcome (Walker *et al*, 2015b) although a retrospective analysis of the 32/1003 patients did not show any survival difference (Avet-Loiseau *et al*, 2011).

Regarding changes on chromosome 1, the role of amp(1q) as a key prognosis factors has been shown in many studies (Walker *et al*, 2019, 2015a; Boyle *et al*, 2015). Regarding 1p, most of the interest has focused on 1p32 (Hebraud *et al*, 2014) that has been associated with an inferior outcome, especially in the young.

TP53 deletion, a well-known tumour suppressor associated with an adverse outcome in many cancers, has also been associated with a negative impact on outcome. Its incidence is believed to increase at relapse (Jones *et al*, 2019). Lodé and colleagues showed that *TP53* mutations were exclusively associated with del(17p) (Lodé *et al*, 2010) and more than the actual event it was the presence of a biallelic inactivation that was responsible for the adverse prognosis (Walker *et al*, 2019; Thakurta *et al*, 2019).

1.3.2.5.5. Gene expression profile

In the early 2000, experts from the University of Arkansas for Medical Science (UAMS) first identified molecular subgroups of MM using gene expression profiling. They defined a seven clusters classification, CD-1, CD-2, MS, MF, HY, PR, and LB that loosly matched the cytogenetic subgroups (Zhan *et al*, 2006). Three novel subsets were identified later on, by the HOVON-65/GMMG-HD4 group which were NFκB, CTA, and PRL3 clusters (Broyl *et al*, 2010).

The IFM model used a combination of 15 most stable genes associated with survival in the IFM 99 trials to stratify patients into prognostic groups (Decaux *et al*, 2008). Similarly the EMC-92-gene signature was generated from the HOVON65/GMMG-HD4 trial (Kuiper *et al*, 2012) and the GEP70 risk score developed at UAMS (Shaughnessy *et al*, 2007) with the aim of risk stratifying patients.

1.3.2.5.6. Risk stratification models

Several risk stratifications models have been developed for prognostication of MM patients. The most widely used are shown in, **Table 1.3.2.5.6.1**. Unfortunately, most of them remain confidential within clinical trials or specific tertiary centres.

	Definition	Outcome
Double-Hit (Walker *et al*, 2019)	Bi-allelic TP53 or amp(1q) and ISS III	6.1% of the population PFS = 15.4 months; OS = 20.7 months
CPI (Perrot *et al*, 2019)	Weighted del(17p), t(4;14), del(1p32), and 1q21 gain and trisomies 3, 5, and 21	Intermediate: HR 1.4 for PFS and 2-5 for OS High risk: 1.76 HR for PFS and 6.4-15 for OS
MSMART (Mikhael *et al*, 2013)	Cytogenetics, R-ISS, PCLI, GEP, *TP53* mutations	Treatment adapted model

Table 1.3.2.5.6.1. Most commonly used risk stratification models in MM.

1.3.3. Smouldering myeloma (SMM)

1.3.3.1. Definition

Smouldering myeloma (SMM) is an asymptomatic clonal plasma cell disorder. Historically, SMM differed from MM by the absence of end organ damage and from MGUS by its risk and rate of progression to overt MM. The natural history of SMM is variable with heterogeneous disease behaviours ranging from MM-like to MGUS-like disease courses.

The first cases of smouldering myeloma were described by Norgaard et al in 1971 in the British Journal of Haematology. A few years later, Kyle and Greipp described the case of six patients with more than 30g/L of paraprotein and more than 10% BMPC who remained stable for 5 to 16 years without treatment and first coined them as Smouldering MM in a 1980 key paper (Kyle & Greipp, 1980). The same year, Alexanian defined indolent myeloma based on the description of 20 cases with ≥ 15% BMPC, < 3 lytic lesions, minimal monoclonal protein level depending on the type of immunoglobulin of 25 g/L for IgG and 10 g/L for IgA and a time to progression > 2 years (Alexanian, 1980) and delayed chemotherapy for up to 8 years

This definition was replaced by an easier definition of smouldering or asymptomatic myeloma by the International Myeloma Working Group (IMWG) consensus in 2003 to include patients with more than 10% BMPC and a serum paraprotein greater or equal to 30g/L without evidence of CRAB criteria.

These criteria were subsequently updated in 2014 by the IMWG (Rajkumar *et al*, 2014). SMM is currently defined by the presence of a paraprotein ≥30 g/L and/or 10% to

60% clonal BMPCs with no evidence of end-organ damage or other myeloma defining event. This revised definition, removed a particularly high risk of progression group of patients.

1.3.3.2. Epidemiology

The lack of population-based registries until recently and the changes in definition have made epidemiological data difficult to acquire. In a study based on the National Cancer Database in the United States, among the 86,327 MM patients included in the study, 13.7% were SMM with a median age at diagnosis of 67 years. The estimated incidence was thus 0.9 cases per 100 000 persons (Ravindran *et al*, 2016). These findings are similar to another Swedish population-based study, where 14.4% of MMs were smouldering at diagnosis suggesting an incidence of 0.4 cases per 100 000 persons (Kristinsson *et al*, 2013). The incidence of SMM was higher among women, African Americans, less educated patients and those who lived closer to the Northeast of the United-States. There also was a negative correlation between SMM and the number of co-morbidities (Ravindran *et al*, 2016). The median OS for SMM and MM patients diagnosed in 2003-2007 were 54.8 and 28.6 months, respectively, whereas the median OS for those diagnosed in 2008-2011 were 67.1 and 40.2 months, respectively suggesting than unlike MGUS, these patients do not have the same outcome as the general population (Ravindran *et al*, 2016).

1.3.3.3. Histopathology

From a quantitative perspective, the current definition of SMM states a BM infiltrate ranging from 10% to 60%. Any infiltrate beyond 60% is now considered as MM; any infiltrate below 10% is considered MGUS (unless the paraprotein is >30g/L).

From a qualitative perspective, there is no single feature that clearly distinguishes the two entities in the same manner that there is none that distinguishes between normal and neoplastic plasma cells. Although it has been suggested that in SMM, the tumour infiltration is smaller, follows an interstitial growth pattern, and is composed of predominately mature plasma cells with low proliferation indices (Bain, Barbara *et al*, 2011).

1.3.3.4. Genetics

Given the scarcity of SMM, there are little data available regarding the incidence of SVs in MM. It has been suggested that there are fewer t(4;14) (9% vs 15% in MM) and more t(11;14) (27% vs 18%). Similarly, there are fewer del(13q) (20% vs 50%). Reported incidences of HRD, del(17p) and gain(1q) were similar than in MM (Neben *et al*, 2013; Bustoros *et al*, 2020).

Most series of SMM have suggested that the genetic make-up of SMM was similar to MM (Walker *et al*, 2014b; Bolli *et al*, 2018; Bustoros *et al*, 2020) and that intraclonal heterogeneity was already present at the SMM stage. Recently, Bustoros et al (Bustoros *et al*, 2020) have suggested that forty-six percent of patients had alterations in the mitogen-activated protein kinase (MAPK) pathway. DNA repair pathway alterations were seen in 10% of cases similar to what was seen in MM. Mutations in genes pertaining to the NFκB factors, protein processing, and cell cycle pathways were found in 22%, 21%, and 6.7% of cases, respectively, consistent with data seen in MM.

Biallelic inactivation events in genes such as *TP53*, *RB1*, *CDKN2C*, *ZNF292*, DIS3, or *FAM46C* were seen in only 6% of patients which was less than what has been described in MM (Walker *et al*, 2018).

1.3.3.5. Risk factors for progression

1.3.1.5.1. The tumour burden and the 20/2/20

Risk stratification may be achieved using the Mayo 2018 (Lakshman *et al*, 2018) risk stratification system (also called the 20/2/20 criteria), which takes into consideration the following three risk factors for progression (**Annex 2**):

- Bone marrow plasma cells greater than 20 percent
- M protein greater or equal 20 g/L
- Involved/uninvolved free light chain (FLC) ratio greater or equal than 20

These factors were able to stratify patients into three risk groups with distinctly different patterns of progression. Patients in the High-risk group (two or three factors present) had an estimated time to progression (TTP) of 29 months with an estimated risk of progression of 24 percent per year during the first two years, 11 percent per year for

the next three years, and 3 percent per year for the next five years. Patient in the intermediate risk (one factor present) have an estimated median TTP of 68 months; estimated rate of progression of 15 percent per year during the first two years, 7 percent per year for the next three years, and 4 percent per year for the next five years. Finally, the low risk group where no factors present had an estimated median TTP of 110 months and an estimated rate of progression of 5 percent per year during the first 10 years.

1.3.1.5.2. Genetic features

Beyond these clinical features, genetic features affecting TTP have been described. In one study, TTP and overall survival (OS) were shorter in patients with t(4;14) when compared with those with t(11;14) (Neben *et al*, 2013). A trend toward shorter TTP has also been seen in patients with del(17p), gain(1q21), and HRD (Neben *et al*, 2013).

A high-risk SMM signature based on only four genes that, alone, predicted a 2-year probability of therapy requirement of 85.7% (Khan *et al*, 2015). When combined with standard clinical tests (serum M-protein <30 g/L, albumin ≥35 g/L), a powerful high-fidelity risk stratification model was generated, in which absence of risk factors successfully selected for an indolent MGUS-like population of 59% of patients with an estimated 2-year progression rate of 5%.

1.3.1.5.3. Dynamics

Progressive increases in paraprotein over short period of time have been described in SMM and associated with a short PFS. This evolving phenotype was defined by an increase in paraprotein levels by ≥25% on two successive evaluations within a 6-month period and was associated with a HR for progression of 5 (Fernández de Larrea *et al*, 2018)

The combination of an evolving phenotype, an SFLC ratio ≥ 8, and BMPC ≥ 20% emerged as significant risk factors for progression where the presence of none or one of these risk factors distinguished a population with a significantly lower risk of progression than patients with two or more risk factors. The 2-year progression rate for subjects with three risk factors was 25%, and for those with evolving changes in both evolving Haemoglobin and evolving paraproteins was 18.5% (Atrash *et al*, 2018). These rates are comparatively much lower than described by Ravi et al (Ravi *et al*, 2016)

1.3.1.5.4. Other factors

Other risk factors include the presence of a focal lesion by PET-CT without underlying osteolytic bone destruction (Siontis *et al*, 2015). Immunoparesis defined by the reduction of two uninvolved immunoglobulin isotypes and the presence of abnormal circulating plasma cell immunophenotype in the setting of immunoparesis has also been associated with a short PFS (Pérez-Persona *et al*, 2010).

1.4. Objectives and unsolved questions

The clinical presentation of plasma cell disorders is heterogeneous.This work aims at understanding the genetic basis of some of these clinical differences.

The role of genetics still remains to be ascertained in AL. Indeed, although it has been associated with t(11;14), little is known about the genetic makeup of the underlying tumours. Genetic comparison of AL to MGUS and MM could help distinguish that entity and offer either diagnostic and therapeutic tools. To this intent we sequenced 24 newly-diagnosed AL patients using whole exome sequencing (WES) with an additional capture of the IG and *MYC* loci to identify the complete spectrum of exonic mutations, copynumber changes, and translocations. By comparing these results to previously obtained results in the laboratory (Mikulasova *et al*, 2017; Walker *et al*, 2015a) we will highlight the differences between AL and other plasma cell disorders.

We also suggest that genetics may help distinguish SMM and MM and refine this group in terms of outcome in order to better identify patients with SMM requiring treatment. Given the risk for normal plasma cell contamination in SMM samples, we will perform targeted sequencing on 83 samples and compared them to MM processed in the same way. This analysis will aim at identifying mutational drivers, CNA, translocations, mutational signatures and compare them to MM with the aim of highlighting differences that could be used in the clinic. We will also evaluate the changes in SMM over time to highlight the genetic *substratum* of these changes and understand the clonal architecture of this disease entity over time. Using 53 samples derived from nine patients we will attempt to reconstruct the changes in the clonal architecture and highlight potential markers suitable for an inception strategy.

1. Introduction

Although there is extensive data regarding the role for cytogenetic and copy number abnormalities in MM, little data regarding the role of mutations in NDMM. With a series of 223 samples treated homogeneously, and a long-term follow-up, we will confirm the frequency of mutations and study the impact of mutations on outcome to better risk stratify patients in the context of the total therapy trials. Finally, in order to gain insight into the pathogeny of MM, and the role of aging processes, we will analyse the Multiple Myeloma Research Fondation CoMMpass cohort to identify hallmarks of aging and use these data to gain insight into the aetiology of MM by analysing features such as telomere length and clonal haematopoiesis and identify prognostic markers in the elderly.

Combined these data could help unravels the complexity of these disease and could add precision to the diagnosis, and management of AL, SMM, and MM.

1.5. References

Abraham, R.S., Ballman, K.V., Dispenzieri, A., Grill, D.E., Manske, M.K., Price-Troska, T.L., Paz, N.G., Gertz, M.A. & Fonseca, R. (2005) Functional gene expression analysis of clonal plasma cells identifies a unique molecular profile for light chain amyloidosis. *Blood*, **105,** 794–803.

Abraham, R.S., Geyer, S.M., Price-Troska, T.L., Allmer, C., Kyle, R.A., Gertz, M.A. & Fonseca, R. (2003) Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain–associated amyloidosis (AL). *Blood*, **101,** 3801–3807.

Abraham, R.S., Manske, M.K., Zuckerman, N.S., Sohni, A., Edelman, H., Shahaf, G., Timm, M.M., Dispenzieri, A., Gertz, M.A. & Mehr, R. (2007) Novel analysis of clonal diversification in blood B cell and bone marrow plasma cell clones in immunoglobulin light chain amyloidosis. *Journal of Clinical Immunology*, **27,** 69–87.

Ageing Europe — 2019 edition Available at: https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-02-19-681 [Accessed August 7, 2020].

Alavanja, M.C.R., Hoppin, J.A. & Kamel, F. (2004) Health effects of chronic pesticide exposure: cancer and neurotoxicity. *Annual Review of Public Health*, **25,** 155–197.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S. & Stratton, M.R. (2015) Clock-like mutational processes in human somatic cells. *Nature Genetics*, **47,** 1402–1407.

Alexanian, R. (1980) Localized and indolent myeloma. *Blood*, **56,** 521–525.

Andreotti, G., Birmann, B.M., Cozen, W., Roos, A.J.D., Chiu, B.C.H., Costas, L., Sanjosé, S. de, Moysich, K., Camp, N.J., Spinelli, J.J., Pahwa, P., Dosman, J.A., McLaughlin, J.R., Boffetta, P., Staines, A., Weisenburger, D., Benhaim-Luzon, V., Brennan, P., Costantini, A.S., Miligi, L., et al (2015) A Pooled Analysis of Cigarette Smoking and Risk of Multiple Myeloma from the International Multiple Myeloma Consortium. *Cancer Epidemiology and Prevention Biomarkers*, **24,** 631–634.

Atrash, S., Robinson, M., Slaughter, D., Aneralla, A., Brown, T., Robinson, J., Ndiaye, A., Sprouse, C., Zhang, Q., Symanowski, J.T., Friend, R., Voorhees, P.M., Usmani, S.Z. & Bhutani, M. (2018) Evolving changes in M-protein and hemoglobin as predictors for progression of smoldering multiple myeloma. *Blood Cancer Journal*, **8,** 1–5.

Avet-Loiseau, H., Attal, M., Campion, L., Caillot, D., Hulin, C., Marit, G., Stoppa, A.-M., Voillat, L., Wetterwald, M., Pegourie, B., Voog, E., Tiab, M., Banos, A., Jaubert, J., Bouscary, D., Macro, M., Kolb, B., Traulle, C., Mathiot, C., Magrangeas, F., et al (2012) Long-term analysis of the IFM 99 trials for myeloma: cytogenetic abnormalities [t(4;14), del(17p), 1q gains] play a major role in defining long-term survival. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **30,** 1949–1952.

Avet-Loiseau, H., Facon, T., Daviet, A., Godon, C., Rapp, M.J., Harousseau, J.L., Grosbois, B. & Bataille, R. (1999) 14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma. Intergroupe Francophone du Myélome. *Cancer Research*, **59,** 4546–4550.

Avet-Loiseau, H., Garand, R., Lodé, L., Harousseau, J.-L. & Bataille, R. (2003a) Translocation t(11;14)(q13;q32) is the hallmark of IgM, IgE, and nonsecretory multiple myeloma variants. *Blood*, **101,** 1570–1571.

Avet-Loiseau, H., Garand, R., Lodé, L., Harousseau, J.-L., Bataille, R. & Intergroupe Francophone du Myélome (2003b) Translocation t(11;14)(q13;q32) is the hallmark of IgM, IgE, and nonsecretory multiple myeloma variants. *Blood*, **101,** 1570–1571.

Avet-Loiseau, H., Malard, F., Campion, L., Magrangeas, F., Sebban, C., Lioure, B., Decaux, O., Lamy, T., Legros, L., Fuzibet, J.-G., Michallet, M., Corront, B., Lenain, P., Hulin, C., Mathiot, C., Attal, M., Facon, T., Harousseau, J.-L., Minvielle, S., Moreau, P., et al (2011) Translocation t(14;16) and multiple myeloma: is it really an independent prognostic factor? *Blood*, **117,** 2009–2011.

Bain, Barbara, Clark, David M. & Wilkins, Bridget (2011)Bone Marrow Pathology. Wiley-Blackwell.

Baltz, T., Giroud, C., Baltz, D., Roth, C., Raibaud, A. & Eisen, H. (1986) Stable expression of two variable surface glycoproteins by cloned Trypanosoma equiperdum. *Nature*, **319,** 602–604.

Bernadotte, A., Mikhelson, V.M. & Spivak, I.M. (2016) Markers of cellular senescence. Telomere shortening as a marker of cellular senescence. *Aging (Albany NY)*, **8,** 3–11.

Bochtler, T., Hegenbart, U., Heiss, C., Benner, A., Moos, M., Seckinger, A., Pschowski-Zuck, S., Kirn, D., Neben, K., Bartram, C.R., Ho, A.D., Goldschmidt, H., Hose, D., Jauch, A. & Schonland, S.O. (2011) Hyperdiploidy is less frequent in AL amyloidosis compared with monoclonal gammopathy of undetermined significance and inversely associated with translocation t(11;14). *Blood*, **117,** 3809–3815.

Bochtler, T., Hegenbart, U., Kunz, C., Benner, A., Seckinger, A., Dietrich, S., Granzow, M., Neben, K., Goldschmidt, H., Ho, A.D., Hose, D., Jauch, A. & Schönland, S.O. (2014) Gain of chromosome 1q21 is an independent adverse prognostic factor in light chain amyloidosis patients treated with melphalan/dexamethasone. *Amyloid: The International Journal of Experimental and Clinical Investigation: The Official Journal of the International Society of Amyloidosis*, **21,** 9–17.

Bochtler, T., Hegenbart, U., Kunz, C., Granzow, M., Benner, A., Seckinger, A., Kimmich, C., Goldschmidt, H., Ho, A.D., Hose, D., Jauch, A. & Schönland, S.O. (2015) Translocation t(11;14) is associated with adverse outcome in patients with newly diagnosed AL amyloidosis when treated with bortezomib-based regimens. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **33,** 1371–1378.

Bolli, N., Maura, F., Minvielle, S., Gloznik, D., Szalat, R., Fullam, A., Martincorena, I., Dawson, K.J., Samur, M.K., Zamora, J., Tarpey, P., Davies, H., Fulciniti, M., Shammas, M.A., Tai, Y.T., Magrangeas, F., Moreau, P., Corradini, P., Anderson, K., Alexandrov, L., et al (2018) Genomic patterns of progression in smoldering multiple myeloma. *Nature Communications*, **9,** 3363.

Boyle, E.M., Proszek, P.Z., Kaiser, M.F., Begum, D., Dahir, N., Savola, S., Wardell, C.P., Leleu, X., Ross, F.M., Chiecchio, L., Cook, G., Drayson, M.T., Owen, R.G., Ashcroft, J.M., Jackson, G.H., Anthony Child, J., Davies, F.E., Walker, B.A. & Morgan, G.J. (2015) A molecular diagnostic approach able to detect the recurrent genetic prognostic factors typical of presenting myeloma. *Genes, Chromosomes & Cancer*, **54,** 91–98.

Boyle, E.M., Walker, B.A., Wardell, C.P., Leleu, X., Davies, F.E. & Morgan, G.J. (2014) B-cell malignancies: capture-sequencing strategies for identification of gene rearrangements and translocations into immunoglobulin gene loci. *Blood and Lymphatic Cancer: Targets and Therapy*, **4,** 107–119 Available at: https://www.dovepress.com/b-cell-malignancies-capture-sequencing-strategies-for-identification-o-peer-reviewed-fulltext-article-BLCTT [Accessed October 20, 2020].

Brenner, D.A., Jain, M., Pimentel, D.R., Wang, B., Connors, L.H., Skinner, M., Apstein, C.S. & Liao, R. (2004) Human amyloidogenic light chains directly impair cardiomyocyte function through an increase in cellular oxidant stress. *Circulation Research*, **94,** 1008–1010.

Briault, S., Courtois-Capella, M., Duarte, F., Aucouturier, P. & Preud'Homme, J.L. (1988) Isotypy of serum monoclonal immunoglobulins in human immunodeficiency virus-infected adults. *Clinical and Experimental Immunology*, **74,** 182–184.

1. Introduction

Broyl, A., Hose, D., Lokhorst, H., de Knegt, Y., Peeters, J., Jauch, A., Bertsch, U., Buijs, A., Stevens-Kroef, M., Beverloo, H.B., Vellenga, E., Zweegman, S., Kersten, M.-J., van der Holt, B., el Jarari, L., Mulligan, G., Goldschmidt, H., van Duin, M. & Sonneveld, P. (2010) Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood*, **116,** 2543–2553.

Bryce, A.H., Ketterling, R.P., Gertz, M.A., Lacy, M., Knudson, R.A., Zeldenrust, S., Kumar, S., Hayman, S., Buadi, F., Kyle, R.A., Greipp, P.R., Lust, J.A., Russell, S., Rajkumar, S.V., Fonseca, R. & Dispenzieri, A. (2009) Translocation t(11;14) and survival of patients with light chain (AL) amyloidosis. *Haematologica*, **94,** 380–386.

Bustoros, M., Sklavenitis-Pistofidis, R., Park, J., Redd, R., Zhitomirsky, B., Dunford, A.J., Salem, K., Tai, Y.-T., Anand, S., Mouhieddine, T.H., Chavda, S.J., Boehner, C., Elagina, L., Neuse, C.J., Cha, J., Rahmat, M., Taylor-Weiner, A., Van Allen, E., Kumar, S., Kastritis, E., et al (2020) Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *Journal of Clinical Oncology*, **38,** 2380–2389.

Cabrera, Q., Macro, M., Hebert, B., Cornet, E., Collignon, A. & Troussard, X. (2014) Epidemiology of Monoclonal Gammopathy of Undetermined Significance (MGUS): The experience from the specialized registry of hematologic malignancies of Basse-Normandie (France). *Cancer Epidemiology*, **38,** 354–356.

Caccialanza, R., Palladini, G., Klersy, C., Cereda, E., Bonardi, C., Cameletti, B., Quarleri, L., Montagna, E., Foli, A., Milani, P., Lavatelli, F., Marena, C. & Merlini, G. (2014) Malnutrition at diagnosis predicts mortality in patients with systemic immunoglobulin light-chain amyloidosis independently of cardiac stage and response to treatment. *JPEN. Journal of parenteral and enteral nutrition*, **38,** 891–894.

Caers, J., Paiva, B., Zamagni, E., Leleu, X., Bladé, J., Kristinsson, S.Y., Touzeau, C., Abildgaard, N., Terpos, E., Heusschen, R., Ocio, E., Delforge, M., Sezer, O., Beksac, M., Ludwig, H., Merlini, G., Moreau, P., Zweegman, S., Engelhardt, M. & Rosiñol, L. (2018) Diagnosis, treatment, and response assessment in solitary plasmacytoma: updated recommendations from a European Expert Panel. *Journal of Hematology & Oncology*, **11,** 10 Available at: https://doi.org/10.1186/s13045-017-0549-1 [Accessed April 4, 2020].

Castillo, J.J., Dhami, P.K., Curry, S. & Brennan, K. (2012) No association between cigarette smoking and incidence of plasma cell myeloma: a meta-analysis of 17 observational studies. *American Journal of Hematology*, **87,** 729–731.

Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.-P., Ahmann, G.J., Adli, M., Anderson, K.C., Ardlie, K.G., Auclair, D., Baker, A., Bergsagel, P.L., Bernstein, B.E., Drier, Y., Fonseca, R., Gabriel, S.B., Hofmeister, C.C., et al (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471,** 467–472.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E. & Alt, F.W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*, **422,** 726–730.

Chretien, M.-L., Corre, J., Lauwers-Cances, V., Magrangeas, F., Cleynen, A., Yon, E., Hulin, C., Leleu, X., Orsini-Piocelle, F., Blade, J.-S., Sohn, C., Karlin, L., Delbrel, X., Hebraud, B., Roussel, M., Marit, G., Garderet, L., Mohty, M., Rodon, P., Voillat, L., et al (2015) Understanding the role of hyperdiploidy in myeloma prognosis: which trisomies really matter? *Blood*, **126,** 2713–2719.

Clay-Gilmour, A.I., Hildebrandt, M.A.T., Brown, E.E., Hofmann, J.N., Spinelli, J.J., Giles, G.G., Cozen, W., Bhatti, P., Wu, X., Waller, R.G., Belachew, A.A., Robinson, D.P., Norman, A.D., Sinnwell, J.P., Berndt, S.I., Rajkumar, S.V., Kumar, S.K., Chanock, S.J., Machiela, M.J., Milne, R.L., et al (2020) Coinherited genetics of multiple myeloma and its precursor, monoclonal gammopathy of undetermined significance. *Blood Advances*, **4,** 2789–2797.

Coker, W.J., Jeter, A., Schade, H. & Kang, Y. (2013) Plasma cell disorders in HIV-infected patients: epidemiology and molecular mechanisms. *Biomarker Research*, **1,** 8.

Comenzo, R.L., Zhang, Y., Martinez, C., Osman, K. & Herrera, G.A. (2001) The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V(L) germ line gene use and clonal plasma cell burden. *Blood*, **98,** 714–720.

Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing | bioRxiv Available at: https://www.biorxiv.org/content/10.1101/333617v1 [Accessed January 8, 2020].

D'Apuzzo, M., Rolink, A., Loetscher, M., Hoxie, J.A., Clark-Lewis, I., Melchers, F., Baggiolini, M. & Moser, B. The chemokine SDF-1, stromal cell-derived factor 1, attracts early stage B cell precursors via the chemokine receptor CXCR4. *European Journal of Immunology*, **27,** 1788–1793.

Decaux, O., Lodé, L., Magrangeas, F., Charbonnel, C., Gouraud, W., Jézéquel, P., Attal, M., Harousseau, J.-L., Moreau, P., Bataille, R., Campion, L., Avet-Loiseau, H. & Minvielle, S. (2008) 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. *Journal of Clinical Oncology*, **26,** 4798–4805.

Delves, P.J. & Roitt, I.M. (2000) The immune system. First of two parts. *The New England journal of medicine*, **343,** 37–49.

Dispenzieri, A. (2019) POEMS Syndrome: 2019 Update on diagnosis, risk-stratification, and management. *American Journal of Hematology*, **94,** 812–827.

Dispenzieri, A., Gertz, M.A., Kyle, R.A., Lacy, M.Q., Burritt, M.F., Therneau, T.M., Greipp, P.R., Witzig, T.E., Lust, J.A., Rajkumar, S.V., Fonseca, R., Zeldenrust, S.R., McGregor, C.G.A. & Jaffe, A.S. (2004) Serum cardiac troponins and N-terminal pro-brain natriuretic peptide: a staging system for primary systemic amyloidosis. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **22,** 3751–3757.

Dispenzieri, A., Rajkumar, S.V., Gertz, M.A., Lacy, M.Q., Kyle, R.A., Greipp, P.R., Witzig, T.E., Lust, J.A., Russell, S.J., Hayman, S.R., Kumar, S., Zeldenrust, S.R., Fonseca, R., Bergsagel, P.L., Reeder, C.B., Stewart, A.K., Roy, V. & Dalton, R.J. (2007) Treatment of Newly Diagnosed Multiple Myeloma Based on Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART): Consensus Statement. *Mayo Clinic Proceedings*, **82,** 323–341.

Dittrich, T., Bochtler, T., Kimmich, C., Becker, N., Jauch, A., Goldschmidt, H., Ho, A.D., Hegenbart, U. & Schönland, S.O. (2017) AL amyloidosis patients with low amyloidogenic free light chain levels at first diagnosis have an excellent prognosis. *Blood*, **130,** 632–642.

van de Donk, N.W.C.J., Lokhorst, H.M., Anderson, K.C. & Richardson, P.G. (2012) How I treat plasma cell leukemia. *Blood*, **120,** 2376–2389 Available at: https://ashpublications.org/blood/article/120/12/2376/30567/How-I-treat-plasma-cell-leukemia [Accessed April 4, 2020].

Dreesen, O., Li, B. & Cross, G. (2007) Telomere structure and function in trypanosomes: a proposal. *Nature Reviews Microbiology*.

Duberg, A.-S., Nordström, M., Törner, A., Reichard, O., Strauss, R., Janzon, R., Bäck, E. & Ekdahl, K. (2005) Non-Hodgkin's lymphoma and other nonhepatic malignancies in Swedish patients with hepatitis C virus infection. *Hepatology (Baltimore, Md.)*, **41,** 652–659.

Dubrey, S.W., Cha, K., Anderson, J., Chamarthi, B., Reisinger, J., Skinner, M. & Falk, R.H. (1998) The clinical features of immunoglobulin light-chain (AL) amyloidosis with heart involvement. *QJM: monthly journal of the Association of Physicians*, **91,** 141–157.

Dubrey, S.W., Cha, K., Skinner, M., LaValley, M. & Falk, R.H. (1997) Familial and primary (AL) cardiac amyloidosis: echocardiographically similar diseases with distinctly different clinical outcomes. *Heart (British Cardiac Society)*, **78,** 74–82.

Fasullo, F.J., Fritsche, H.A., Liu, F.J. & Hamilton, R.G. (1989) IgG heavy-chain subclass typing of myeloma paraproteins by isoelectric focusing immunoblot analysis. *Clinical Chemistry*, **35,** 364–368.

Fernández de Larrea, C., Isola, I., Pereira, A., Cibeira, M.T., Magnano, L., Tovar, N., Rodríguez-Lobato, L.-G., Calvo, X., Aróstegui, J.I., Díaz, T., Lozano, E., Rozman, M., Yagüe, J., Bladé, J. & Rosiñol, L. (2018) Evolving M-protein pattern in patients with smoldering multiple myeloma: impact on early progression. *Leukemia*, **32,** 1427–1434.

Fraenkel, S., Mostoslavsky, R., Novobrantseva, T.I., Pelanda, R., Chaudhuri, J., Esposito, G., Jung, S., Alt, F.W., Rajewsky, K., Cedar, H. & Bergman, Y. (2007) Allelic 'choice' governs somatic hypermutation in vivo at the immunoglobulin kappa-chain locus. *Nature immunology*, **8,** 715–722.

Frank, C., Fallah, M., Chen, T., Mai, E.K., Sundquist, J., Försti, A. & Hemminki, K. (2016) Search for familial clustering of multiple myeloma with any cancer. *Leukemia*, **30,** 627–632.

Frasca, D. & Blomberg, B.B. (2011) Aging Affects Human B Cell Responses. *Journal of clinical immunology*, **31,** 430–435.

Gaidano, G., Pasqualucci, L., Capello, D., Berra, E., Deambrogi, C., Rossi, D., Larocca, L.M., Gloghini, A., Carbone, A. & Dalla-Favera, R. (2003) Aberrant somatic hypermutation in multiple subtypes of AIDS-associated non-Hodgkin lymphoma. *Blood*, **102,** 1833–1841.

Gertz, M.A. (2013) Immunoglobulin light chain amyloidosis: 2013 update on diagnosis, prognosis, and treatment. *American journal of hematology*, **88,** 416–425.

Gertz, M.A., Kyle, R.A. & Greipp, P.R. (1989) The plasma cell labeling index: a valuable tool in primary systemic amyloidosis. *Blood*, **74,** 1108–1111.

Glenner, G.G. (1980) Amyloid Deposits and Amyloidosis. *New England Journal of Medicine*, **302,** 1283–1292.

Glenner, G.G., Terry, W., Harada, M., Isersky, C. & Page, D. (1971) Amyloid Fibril Proteins: Proof of Homology with Immunoglobulin Light Chains by Sequence Analyses. *Science*, **172,** 1150–1151.

Granzow, M., Hegenbart, U., Hinderhofer, K., Hose, D., Seckinger, A., Bochtler, T., Hemminki, K., Goldschmidt, H., Schönland, S.O. & Jauch, A. (2017) Novel recurrent chromosomal aberrations detected in clonal plasma cells of light chain amyloidosis patients show potential adverse prognostic effect: first results from a genome-wide copy number array analysis. *Haematologica*, **102,** 1281–1290.

Greipp, P.R., San Miguel, J., Durie, B.G.M., Crowley, J.J., Barlogie, B., Bladé, J., Boccadoro, M., Child, J.A., Avet-Loiseau, H., Harousseau, J.-L., Kyle, R.A., Lahuerta, J.J., Ludwig, H., Morgan, G., Powles, R., Shimizu, K., Shustik, C., Sonneveld, P., Tosi, P., Turesson, I., et al (2005a) International staging system for multiple myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **23,** 3412–3420.

Greipp, P.R., San Miguel, J., Durie, B.G.M., Crowley, J.J., Barlogie, B., Bladé, J., Boccadoro, M., Child, J.A., Avet-Loiseau, H., Harousseau, J.-L., Kyle, R.A., Lahuerta, J.J., Ludwig, H., Morgan, G., Powles, R., Shimizu, K., Shustik, C., Sonneveld, P., Tosi, P., Turesson, I., et al (2005b) International staging system for multiple myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **23,** 3412–3420.

Grulich, A.E., Li, Y., McDonald, A., Correll, P.K.L., Law, M.G. & Kaldor, J.M. (2002) Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis. *AIDS*, **16,** 1155–1161.

Guikema, J.E.J., Hovenga, S., Vellenga, E., Conradie, J.J., Abdulahad, W.H., Bekkema, R., Smit, J.W., Zhan, F., Shaughnessy, J. & Bos, N.A. CD27 is heterogeneously expressed in multiple myeloma: low CD27 expression in patients with high-risk disease. *British Journal of Haematology*, **121,** 36–43.

Hammons, L., Brazauskas, R., Pasquini, M., Hamadani, M., Hari, P. & D'Souza, A. (2018) Presence of fluorescent in situ hybridization abnormalities is associated with plasma cell burden in light chain amyloidosis. *Hematology/Oncology and Stem Cell Therapy*, **11,** 105–111.

Harrison, C.J., Mazzullo, H., Ross, F.M., Cheung, K.L., Gerrard, G., Harewood, L., Mehta, A., Lachmann, H.J., Hawkins, P.N. & Orchard, K.H. (2002) Translocations of 14q32 and deletions of 13q14 are common chromosomal abnormalities in systemic amyloidosis. *British Journal of Haematology*, **117,** 427–435.

Hayman, S.R., Bailey, R.J., Jalal, S.M., Ahmann, G.J., Dispenzieri, A., Gertz, M.A., Greipp, P.R., Kyle, R.A., Lacy, M.Q., Rajkumar, S.V., Witzig, T.E., Lust, J.A. & Fonseca, R. (2001) Translocations involving the immunoglobulin heavy-chain locus are possible early genetic events in patients with primary systemic amyloidosis. *Blood*, **98,** 2266–2268.

Hebraud, B., Leleu, X., Lauwers-Cances, V., Roussel, M., Caillot, D., Marit, G., Karlin, L., Hulin, C., Gentil, C., Guilhot, F., Garderet, L., Lamy, T., Brechignac, S., Pegourie, B., Jaubert, J., Dib, M., Stoppa, A.-M., Sebban, C., Fohrer, C., Fontan, J., et al (2014) Deletion of the 1p32 region is a major independent prognostic factor in young patients with myeloma: the IFM experience on 1195 patients. *Leukemia*, **28,** 675–679.

Hellmich, C., Moore, J.A., Bowles, K.M. & Rushworth, S.A. (2020) Bone Marrow Senescence and the Microenvironment of Hematological Malignancies. *Frontiers in Oncology*, **10,** Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7052485/ [Accessed October 18, 2020].

Hemminki, K., Li, X., Försti, A., Sundquist, J. & Sundquist, K. (2012) Incidence and survival in non-hereditary amyloidosis in Sweden. *BMC public health*, **12,** 974.

Herrera, G.A., Paul, R., Turbat-Herrera, E.A., Work, J., Viale, G., dell'Orto, P., Lott, R.L., Colombi, R., Old, C.W., Lyerly, R.H. & Coggi, G. (1986) Ultrastructural Immunolabeling in the Diagnosis of Light-Chain-Related Renal Disease. *Pathology and Immunopathology Research*, **5,** 170–187.

Home - Office for National Statistics Available at: https://www.ons.gov.uk/ [Accessed August 7, 2020].

Huber, O., Ménard, L., Haurie, V., Nicou, A., Taras, D. & Rosenbaum, J. (2008) Pontin and Reptin, Two Related ATPases with Multiple Roles in Cancer. *Cancer Research*, **68,** 6873–6876.

Hurle, M.R., Helms, L.R., Li, L., Chan, W. & Wetzel, R. (1994) A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proceedings of the National Academy of Sciences of the United States of America*, **91,** 5446–5450.

Infante, P.F. (2006) Benzene exposure and multiple myeloma: a detailed meta-analysis of benzene cohort studies. *Annals of the New York Academy of Sciences*, **1076,** 90–109.

International Myeloma Working Group (2003) Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *British Journal of Haematology*, **121,** 749–757.

Jones, J.R., Weinhold, N., Ashby, C., Walker, B.A., Wardell, C., Pawlyn, C., Rasche, L., Melchor, L., Cairns, D.A., Gregory, W.M., Johnson, D., Begum, D.B., Ellis, S., Sherborne, A.L., Cook, G., Kaiser, M.F., Drayson, M.T., Owen, R.G., Jackson, G.H., Davies, F.E., et al (2019) Clonal evolution in myeloma: the impact of maintenance lenalidomide and depth of response on the genetics and sub-clonal structure of relapsed disease in uniformly treated newly diagnosed patients. *Haematologica*, **104,** 1440–1450.

Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F.W. (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual review of immunology*, **24,** 541–570.

Kalff, A. & Spencer, A. (2012) The t(4;14) translocation and FGFR3 overexpression in multiple myeloma: prognostic implications and current clinical strategies. *Blood Cancer Journal*, **2,** e89.

Kastritis, E., Papassotiriou, I., Merlini, G., Milani, P., Terpos, E., Basset, M., Akalestos, A., Russo, F., Psimenou, E., Apostolakou, F., Roussou, M., Gavriatopoulou, M., Eleutherakis-Papaiakovou, E., Fotiou, D., Ziogas, D.C., Papadopoulou, E., Pamboucas, C., Dimopoulos, M.A. & Palladini, G. (2018) Growth differentiation factor-15 is a new biomarker for survival and renal outcomes in light chain amyloidosis. *Blood*, **131,** 1568–1575.

Kastritis, E., Papassotiriou, I., Terpos, E., Roussou, M., Gavriatopoulou, M., Komitopoulou, A., Skevaki, C., Eleutherakis-Papaiakovou, E., Pamboucas, C., Psimenou, E., Manios, E., Giannouli, S., Politou, M., Gakiopoulou, H., Papadopoulou, E., Stamatelopoulos, K., Tasidou, A. & Dimopoulos, M.A. (2016) Clinical and prognostic significance of serum levels of von Willebrand factor and ADAMTS-13 antigens in AL amyloidosis. *Blood*, **128,** 405–409.

Keim, C., Kazadi, D., Rothschild, G. & Basu, U. (2013) Regulation of AID, the B-cell genome mutator. *Genes & development*, **27,** 1–17.

Khan, R., Dhodapkar, M., Rosenthal, A., Heuck, C., Papanikolaou, X., Qu, P., van Rhee, F., Zangari, M., Jethava, Y., Epstein, J., Yaccoby, S., Hoering, A., Crowley, J., Petty, N., Bailey, C., Morgan, G. & Barlogie, B. (2015) Four genes predict high risk of progression from smoldering to symptomatic multiple myeloma (SWOG S0120). *Haematologica*, **100,** 1214–1221.

Khodabakhshi, A.H., Morin, R.D., Fejes, A.P., Mungall, A.J., Mungall, K.L., Bolger-Munro, M., Johnson, N.A., Connors, J.M., Gascoyne, R.D., Marra, M.A., Birol, I. & Jones, S.J.M. (2012) Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget*, **3,** 1308–1319.

Khurana, R., Coleman, C., Ionescu-Zanetti, C., Carter, S.A., Krishna, V., Grover, R.K., Roy, R. & Singh, S. (2005) Mechanism of thioflavin T binding to amyloid fibrils. *Journal of Structural Biology*, **151,** 229–238.

Kim, S.J., Shin, H.-T., Lee, H.-O., Kim, N.K.D., Yun, J.W., Hwang, J.H., Kim, K. & Park, W.-Y. (2016) Recurrent mutations of MAPK pathway genes in multiple myeloma but not in amyloid light-chain amyloidosis. *Oncotarget*, **7,** 68350–68359.

Kisilevsky, R. (1990) Heparan sulfate proteoglycans in amyloidogenesis: an epiphenomenon, a unique factor, or the tip of a more fundamental process? *Laboratory Investigation; a Journal of Technical Methods and Pathology*, **63,** 589–591.

Kiyoi, H., Naito, K., Ohno, R., Saito, H. & Naoe, T. (1998) Characterization of the immunoglobulin light chain variable region gene expressed in multiple myeloma. *Leukemia*, **12,** 601–609.

Kosmas, C., Viniou, N.A., Stamatopoulos, K., Courtenay-Luck, N.S., Papadaki, T., Kollia, P., Paterakis, G., Anagnostou, D., Yataganas, X. & Loukopoulos, D. (1996) Analysis of the kappa light chain variable region in multiple myeloma. *British Journal of Haematology*, **94,** 306–317.

Kourelis, T.V., Kumar, S.K., Gertz, M.A., Lacy, M.Q., Buadi, F.K., Hayman, S.R., Zeldenrust, S., Leung, N., Kyle, R.A., Russell, S., Dingli, D., Lust, J.A., Lin, Y., Kapoor, P., Rajkumar, S.V., McCurdy, A. & Dispenzieri, A. (2013) Coexistent multiple myeloma or increased bone marrow plasma cells define equally high-risk populations in patients with immunoglobulin light chain amyloidosis. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **31,** 4319–4324.

Kristen, A.V., Rosenberg, M., Lindenmaier, D., Merkle, C., Steen, H., Andre, F., Schönland, S.O., Schnabel, P.A., Schuster, T., Röcken, C., Giannitsis, E., Katus, H.A. & Frey, N. (2014) Osteopontin: a novel predictor of survival in patients with systemic light-chain amyloidosis. *Amyloid: The International Journal of Experimental and Clinical Investigation: The Official Journal of the International Society of Amyloidosis*, **21,** 202–210.

Kristinsson, S.Y., Björkholm, M., Goldin, L.R., Blimark, C., Mellqvist, U.-H., Wahlin, A., Turesson, I. & Landgren, O. (2009) Patterns of hematologic malignancies and solid tumors among 37,838 first-degree relatives of 13,896 multiple myeloma patients in Sweden. *International journal of cancer. Journal international du cancer*, **125,** 2147–2150.

Kristinsson, S.Y., Holmberg, E. & Blimark, C. (2013) Treatment for high-risk smoldering myeloma. *The New England Journal of Medicine*, **369,** 1762–1763.

Kuiper, R., Broyl, A., de Knegt, Y., van Vliet, M.H., van Beers, E.H., van der Holt, B., el Jarari, L., Mulligan, G., Gregory, W., Morgan, G., Goldschmidt, H., Lokhorst, H.M., van Duin, M. & Sonneveld, P. (2012) A gene expression signature for high-risk multiple myeloma. *Leukemia*, **26,** 2406–2413.

Kumar, S., Dispenzieri, A., Katzmann, J.A., Larson, D.R., Colby, C.L., Lacy, M.Q., Hayman, S.R., Buadi, F.K., Leung, N., Zeldenrust, S.R., Ramirez-Alvarado, M., Clark, R.J., Kyle, R.A., Rajkumar, S.V. & Gertz, M.A. (2010) Serum immunoglobulin free light-chain measurement in primary amyloidosis: prognostic value and correlations with clinical features. *Blood*, **116,** 5126–5129.

Kumar, S., Dispenzieri, A., Lacy, M.Q., Hayman, S.R., Buadi, F.K., Colby, C., Laumann, K., Zeldenrust, S.R., Leung, N., Dingli, D., Greipp, P.R., Lust, J.A., Russell, S.J., Kyle, R.A., Rajkumar, S.V. & Gertz, M.A. (2012) Revised prognostic staging system for light chain amyloidosis incorporating cardiac biomarkers and serum free light chain measurements. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **30,** 989–995.

Kyle, R.A., Benson, J., Larson, D., Therneau, T., Dispenzieri, A., Melton, L.J. & Rajkumar, S.V. (2009) IgM Monoclonal Gammopathy of Undetermined Significance and Smoldering Waldenström's Macroglobulinemia. *Clinical lymphoma & myeloma*, **9,** 17–18 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3773469/ [Accessed April 4, 2020].

Kyle, R.A. & Greipp, P.R. (1980) Smoldering multiple myeloma. *The New England Journal of Medicine*, **302,** 1347–1349.

Kyle, R.A., Linos, A., Beard, C.M., Linke, R.P., Gertz, M.A., O'Fallon, W.M. & Kurland, L.T. (1992) Incidence and natural history of primary systemic amyloidosis in Olmsted County, Minnesota, 1950 through 1989. *Blood*, **79,** 1817–1822.

Lakshman, A., Rajkumar, S.V., Buadi, F.K., Binder, M., Gertz, M.A., Lacy, M.Q., Dispenzieri, A., Dingli, D., Fonder, A.L., Hayman, S.R., Hobbs, M.A., Gonsalves, W.I., Hwa, Y.L., Kapoor, P., Leung, N., Go, R.S., Lin, Y., Kourelis, T.V., Warsame, R., Lust, J.A., et al (2018) Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. *Blood Cancer Journal*, **8,** 59.

Landgren, O., Kyle, R.A., Hoppin, J.A., Beane Freeman, L.E., Cerhan, J.R., Katzmann, J.A., Rajkumar, S.V. & Alavanja, M.C. (2009) Pesticide exposure and risk of monoclonal gammopathy of undetermined significance in the Agricultural Health Study. *Blood*, **113,** 6386–6391.

Landgren, O., Zeig-Owens, R., Giricz, O., Goldfarb, D., Murata, K., Thoren, K., Ramanathan, L., Hultcrantz, M., Dogan, A., Nwankwo, G., Steidl, U., Pradhan, K., Hall, C.B., Cohen, H.W., Jaber, N., Schwartz, T., Crowley, L., Crane, M., Irby, S., Webber, M.P., et al (2018) Multiple Myeloma and Its Precursor Disease Among Firefighters Exposed to the World Trade Center Disaster. *JAMA Oncology*, **4,** 821–827.

Langerak, A.W. & van Dongen, J.J.M. (2006) Recombination in the human IGK locus. *Critical reviews in immunology*, **26,** 23–42.

Lavatelli, F., Brambilla, F., Valentini, V., Rognoni, P., Casarini, S., Di Silvestre, D., Perfetti, V., Palladini, G., Sarais, G., Mauri, P. & Merlini, G. (2011) A novel approach for the purification and proteomic analysis of pathogenic immunoglobulin free light chains from serum. *Biochimica Et Biophysica Acta*, **1814,** 409–419.

Lee, J.H., Jeong, S.A., Khadka, P., Hong, J. & Chung, I.K. (2015) Involvement of SRSF11 in cell cycle-specific recruitment of telomerase to telomeres at nuclear speckles. *Nucleic Acids Research*, **43,** 8435–8451.

Leuraud, K., Richardson, D.B., Cardis, E., Daniels, R.D., Gillies, M., O'Hagan, J.A., Hamra, G.B., Haylock, R., Laurier, D., Moissonnier, M., Schubauer-Berigan, M.K., Thierry-Chef, I. & Kesminiene, A. (2015) Ionising radiation and risk of death from leukaemia and lymphoma in radiation-monitored workers (INWORKS): an international cohort study. *The Lancet. Haematology*, **2,** e276–e281.

Liu, T., Xu, Q., Zhang, C. & Zhang, P. (2013) Occupational exposure to methylene chloride and risk of cancer: a meta-analysis. *Cancer causes & control: CCC*, **24,** 2037–2049.

Lodé, L., Eveillard, M., Trichet, V., Soussi, T., Wuillème, S., Richebourg, S., Magrangeas, F., Ifrah, N., Campion, L., Traullé, C., Guilhot, F., Caillot, D., Marit, G., Mathiot, C., Facon, T., Attal, M., Harousseau, J.-L., Moreau, P., Minvielle, S. & Avet-Loiseau, H. (2010) Mutations in TP53 are exclusively associated with del(17p) in multiple myeloma. *Haematologica*, **95,** 1973–1976.

Magrangeas, F., Avet-Loiseau, H., Munshi, N.C. & Minvielle, S. (2011) Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. *Blood*, **118,** 675–678.

Marinac, C.R., Birmann, B.M., Lee, I.-M., Rosner, B.A., Townsend, M.K., Giovannucci, E., Rebbeck, T.R., Buring, J.E. & Colditz, G.A. (2018) Body mass index throughout adulthood, physical activity, and risk of multiple myeloma: a prospective analysis in three large cohorts. *British Journal of Cancer*, **118,** 1013–1019.

McHugh, D. & Gil, J. (2018) Senescence and aging: Causes, consequences, and therapeutic avenues. *The Journal of Cell Biology*, **217,** 65–77.

McLaurin, J., Yang, D., Yip, C.M. & Fraser, P.E. (2000) Review: modulating factors in amyloid-beta fibril formation. *Journal of Structural Biology*, **130,** 259–270.

Merlini, G. & Bellotti, V. (2003) Molecular mechanisms of amyloidosis. *The New England Journal of Medicine*, **349,** 583–596.

Migrino, R.Q., Hari, P., Gutterman, D.D., Bright, M., Truran, S., Schlundt, B. & Phillips, S.A. (2010) Systemic and microvascular oxidative stress induced by light chain amyloidosis. *International Journal of Cardiology*, **145,** 67–68.

Migrino, R.Q., Truran, S., Gutterman, D.D., Franco, D.A., Bright, M., Schlundt, B., Timmons, M., Motta, A., Phillips, S.A. & Hari, P. (2011) Human microvascular dysfunction and apoptotic injury induced by AL amyloidosis light chain proteins. *American Journal of Physiology - Heart and Circulatory Physiology*, **301,** H2305–H2312.

Mikhael, J.R., Dingli, D., Roy, V., Reeder, C.B., Buadi, F.K., Hayman, S.R., Dispenzieri, A., Fonseca, R., Sher, T., Kyle, R.A., Lin, Y., Russell, S.J., Kumar, S., Bergsagel, P.L., Zeldenrust, S.R., Leung, N., Drake, M.T., Kapoor, P., Ansell, S.M., Witzig, T.E., et al (2013) Management of Newly Diagnosed Symptomatic Multiple Myeloma: Updated Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) Consensus Guidelines 2013. *Mayo Clinic Proceedings*, **88,** 360–376.

Mikulasova, A., Ashby, C., Tytarenko, R.G., Qu, P., Rosenthal, A., Dent, J.A., Ryan, K.R., Bauer, M.A., Wardell, C.P., Hoering, A., Mavrommatis, K., Trotter, M., Deshpande, S., Yaccoby, S., Tian, E., Keats, J., Auclair, D., Jackson, G.H., Davies, F.E., Thakurta, A., et al (2019) Microhomology-mediated end joining drives complex rearrangements and over expression of MYC and PVT1 in multiple myeloma. *Haematologica*.

Mikulasova, A., Ashby, C., Tytarenko, R.G., Qu, P., Rosenthal, A., Dent, J.A., Ryan, K.R., Bauer, M.A., Wardell, C.P., Hoering, A., Mavrommatis, K., Trotter, M., Deshpande, S., Yaccoby, S., Tian, E., Keats, J., Auclair, D., Jackson, G.H., Davies, F.E., Thakurta, A., et al (2020) Microhomology-mediated end joining drives complex rearrangements and overexpression of MYC and PVT1 in multiple myeloma. *Haematologica*, **105,** 1055–1066.

Mikulasova, A., Wardell, C.P., Murison, A., Boyle, E.M., Jackson, G.H., Smetana, J., Kufova, Z., Pour, L., Sandecka, V., Almasi, M., Vsianska, P., Gregora, E., Kuglik, P., Hajek, R., Davies, F.E., Morgan, G.J. & Walker, B.A. (2017) The spectrum of somatic mutations in monoclonal gammopathy of undetermined significance indicates a less complex genomic landscape than that in multiple myeloma. *Haematologica*, **102,** 1617–1625.

Mirabella, F., Murison, A., Aronson, L.I., Wardell, C.P., Thompson, A.J., Hanrahan, S.J., Fok, J.H.L., Pawlyn, C., Kaiser, M.F., Walker, B.A., Davies, F.E. & Morgan, G.J. (2014) A Novel Functional Role for MMSET in RNA Processing Based on the Link Between the REIIBP Isoform and Its Interaction with the SMN Complex. *PLoS ONE*, **9,** Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4055699/ [Accessed October 8, 2020].

Modi, J., Kamal, J., Eter, A., El-Sayegh, S. & El-Charabaty, E. (2015) Immunoglobulin D Multiple Myeloma With Rapidly Progressing Renal Failure. *Journal of Clinical Medicine Research*, **7,** 653–655 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4471757/ [Accessed April 10, 2020].

Molina, O., Abad, M.A., Solé, F. & Menéndez, P. (2020) Aneuploidy in Cancer: Lessons from Acute Lymphoblastic Leukemia. *Trends in Cancer*, **0,** Available at: https://www.cell.com/trends/cancer/abstract/S2405-8033(20)30240-5 [Accessed October 8, 2020].

Morgan, G.J., Walker, B.A. & Davies, F.E. (2012) The genetic architecture of multiple myeloma. *Nature Reviews. Cancer*, **12,** 335–348.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102,** 553–563.

Murphy, C.L., Eulitz, M., Hrncic, R., Sletten, K., Westermark, P., Williams, T., Macy, S.D., Wooliver, C., Wall, J., Weiss, D.T. & Solomon, A. (2001) Chemical typing of amyloid protein contained in formalin-fixed paraffin-embedded biopsy specimens. *American Journal of Clinical Pathology*, **116,** 135–142.

Myeloma - Cancer Stat Facts *SEER* Available at: https://seer.cancer.gov/statfacts/html/mulmy.html [Accessed August 7, 2020].

Nadeu, F., Martin-Garcia, D., Clot, G., Díaz-Navarro, A., Duran-Ferrer, M., Navarro, A., Vilarrasa-Blasi, R., Kulis, M., Royo, R., Gutiérrez-Abril, J., Valdés-Mas, R., López, C., Chapaprieta, V., Puiggros, M., Castellano, G., Costa, D., Aymerich, M., Jares, P., Espinet, B., Muntañola, A., et al (2020) Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood*, **136,** 1419–1432.

Neben, K., Jauch, A., Bertsch, U., Heiss, C., Hielscher, T., Seckinger, A., Mors, T., Müller, N.Z., Hillengass, J., Raab, M.S., Ho, A.D., Hose, D. & Goldschmidt, H. (2010) Combining information regarding chromosomal aberrations t(4;14) and del(17p13) with the International Staging System classification allows stratification of myeloma patients undergoing autologous stem cell transplantation. *Haematologica*, **95,** 1150–1157.

Neben, K., Jauch, A., Hielscher, T., Hillengass, J., Lehners, N., Seckinger, A., Granzow, M., Raab, M.S., Ho, A.D., Goldschmidt, H. & Hose, D. (2013) Progression in smoldering myeloma is independently determined by the chromosomal abnormalities del(17p), t(4;14), gain 1q, hyperdiploidy, and tumor load. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **31,** 4325–4332.

Nishana, M. & Raghavan, S.C. (2012) Role of recombination activating genes in the generation of antigen receptor diversity and beyond. *Immunology*, **137,** 271–281.

Noel, L.-H., Droz, D. & Ganeval, D. (1987) Immunohistochemical Characterization of Renal Amyloidosis. *American Journal of Clinical Pathology*, **87,** 756–761.

Nutt, S.L., Taubenheim, N., Hasbold, J., Corcoran, L.M. & Hodgkin, P.D. (2011) The genetic network controlling plasma cell differentiation. *Seminars in immunology*, **23,** 341–349.

Otokida, K., Yoshida, H., Mizunuma, Y., Yamada, M. & Hiramori, K. (1990) Multiple myeloma associated with amyloidosis and t(1;20)(q21;q11) translocation. *The Tohoku Journal of Experimental Medicine*, **162,** 363–365.

Ozaki, S., Abe, M., Wolfenbarger, D., Weiss, D.T. & Solomon, A. (1994) Preferential expression of human lambda-light-chain variable-region subgroups in multiple myeloma, AL amyloidosis, and Waldenström's macroglobulinemia. *Clinical Immunology and Immunopathology*, **71,** 183–189.

Paiva, B., Martinez-Lopez, J., Corchete, L.A., Sanchez-Vega, B., Rapado, I., Puig, N., Barrio, S., Sanchez, M.-L., Alignani, D., Lasa, M., Coca, A.G. de, Pardal, E., Oriol, A., Garcia, M.-E.G., Escalante, F., González-López, T.J., Palomera, L., Alonso, J., Prosper, F., Orfao, A., et al (2016) Phenotypic, transcriptomic, and genomic features of clonal plasma cells in light-chain amyloidosis. *Blood*, **127,** 3035–3039.

Paiva, B., Vídriales, M.-B., Pérez, J.J., López-Berges, M.-C., García-Sanz, R., Ocio, E.M., de Las Heras, N., Cuello, R., García de Coca, A., Pardal, E., Alonso, J., Sierra, M., Bárez, A., Hernández, J., Suárez, L., Galende, J., Mateos, M.-V. & San Miguel, J.F. (2011) The clinical utility and prognostic

value of multiparameter flow cytometry immunophenotyping in light-chain amyloidosis. *Blood*, **117,** 3613–3616.

Palumbo, A., Avet-Loiseau, H., Oliva, S., Lokhorst, H.M., Goldschmidt, H., Rosinol, L., Richardson, P., Caltagirone, S., Lahuerta, J.J., Facon, T., Bringhen, S., Gay, F., Attal, M., Passera, R., Spencer, A., Offidani, M., Kumar, S., Musto, P., Lonial, S., Petrucci, M.T., et al (2015) Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **33,** 2863–2869.

Panero, J., Arbelbide, J., Fantl, D.B., Rivello, H.G., Kohan, D. & Slavutsky, I. (2010) Altered mRNA Expression of Telomere-Associated Genes in Monoclonal Gammopathy of Undetermined Significance and Multiple Myeloma. *Molecular Medicine*, **16,** 471–478.

Papadea, C., Reimer, C.B. & Check, I.J. (1989) IgG subclass distribution in patients with multiple myeloma or with monoclonal gammopathy of undetermined significance. *Annals of Clinical and Laboratory Science*, **19,** 27–37.

Paulsson, K., Mörse, H., Fioretos, T., Behrendtz, M., Strömbeck, B. & Johansson, B. (2005) Evidence for a single-step mechanism in the origin of hyperdiploid childhood acute lymphoblastic leukemia. *Genes, Chromosomes & Cancer*, **44,** 113–122.

Pawlyn, C., Kaiser, M.F., Heuck, C., Melchor, L., Wardell, C.P., Murison, A., Chavan, S.S., Johnson, D.C., Begum, D.B., Dahir, N.M., Proszek, P.Z., Cairns, D.A., Boyle, E.M., Jones, J.R., Cook, G., Drayson, M.T., Owen, R.G., Gregory, W.M., Jackson, G.H., Barlogie, B., et al (2016) The Spectrum and Clinical Impact of Epigenetic Modifier Mutations in Myeloma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, **22,** 5783–5794.

Pekarsky, Y. & Croce, C.M. (2015) Role of miR-15/16 in CLL. *Cell Death and Differentiation*, **22,** 6–11.

Pepys, M.B., Rademacher, T.W., Amatayakul-Chantler, S., Williams, P., Noble, G.E., Hutchinson, W.L., Hawkins, P.N., Nelson, S.R., Gallimore, J.R. & Herbert, J. (1994) Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proceedings of the National Academy of Sciences of the United States of America*, **91,** 5602–5606.

Pérez-Persona, E., Mateo, G., García-Sanz, R., Mateos, M.-V., de Las Heras, N., de Coca, A.G., Hernández, J.M., Galende, J., Martín-Nuñez, G., Bárez, A., Alonso, J.M., Martín, A., López-Berges, C., Orfao, A., San Miguel, J.F. & Vidriales, M.-B. (2010) Risk of progression in smouldering myeloma and monoclonal gammopathies of unknown significance: comparative analysis of the evolution of monoclonal component and multiparameter flow cytometry of bone marrow plasma cells. *British Journal of Haematology*, **148,** 110–114.

Perfetti, V., Casarini, S., Palladini, G., Vignarelli, M.C., Klersy, C., Diegoli, M., Ascari, E. & Merlini, G. (2002) Analysis of Vλ-Jλ expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r(λIII) as a new amyloid-associated germline gene segment. *Blood*, **100,** 948–953.

Perfetti, V., Colli Vignarelli, M., Anesi, E., Garini, P., Quaglini, S., Ascari, E. & Merlini, G. (1999) The degrees of plasma cell clonality and marrow infiltration adversely influence the prognosis of AL amyloidosis patients. *Haematologica*, **84,** 218–221.

Perfetti, V., Palladini, G., Casarini, S., Navazza, V., Rognoni, P., Obici, L., Invernizzi, R., Perlini, S., Klersy, C. & Merlini, G. (2012) The repertoire of λ light chains causing predominant amyloid heart

involvement and identification of a preferentially involved germline gene, IGLV1-44. *Blood*, **119,** 144–150.

Perrot, A., Lauwers-Cances, V., Tournay, E., Hulin, C., Chretien, M.-L., Royer, B., Dib, M., Decaux, O., Jaccard, A., Belhadj, K., Brechignac, S., Fontan, J., Voillat, L., Demarquette, H., Collet, P., Rodon, P., Sohn, C., Lifermann, F., Orsini-Piocelle, F., Richez, V., et al (2019) Development and Validation of a Cytogenetic Prognostic Index Predicting Survival in Multiple Myeloma. *Journal of Clinical Oncology*, **37,** 1657–1665.

Perutz, M.F. (1997) Mutations make enzyme polymerize. *Nature*, **385,** 773–775.

Phekoo, K.J., Schey, S.A., Richards, M.A., Bevan, D.H., Bell, S., Gillett, D., Møller, H. & Consultant Haematologists, South Thames Haematology Specialist Committee (2004) A population study to define the incidence and survival of multiple myeloma in a National Health Service Region in UK. *British Journal of Haematology*, **127,** 299–304.

Picken, M.M. (2007) Immunoglobulin light and heavy chain amyloidosis AL/AH: renal pathology and differential diagnosis. *Contributions to nephrology*, **153,** 135–155.

Pilbeam, K.L. & Lund, T.C. (2017) Pediatric multiple myeloma. *Blood*, **129,** 395–395.

Pinney, J.H., Lachmann, H.J., Bansi, L., Wechalekar, A.D., Gilbertson, J.A., Rowczenio, D., Sattianayagam, P.T., Gibbs, S.D.J., Orlandi, E., Wassef, N.L., Bradwell, A.R., Hawkins, P.N. & Gillmore, J.D. (2011) Outcome in renal Al amyloidosis after chemotherapy. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **29,** 674–681.

Pioli, P.D., Casero, D., Montecino-Rodriguez, E., Morrison, S.L. & Dorshkind, K. (2019) Plasma Cells Are Obligate Effectors of Enhanced Myelopoiesis in Aging Bone Marrow. *Immunity*, **51,** 351-366.e6.

Popovic, R., Martinez-Garcia, E., Giannopoulou, E.G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M.Y., Zheng, Y., Will, C.M., Small, E.C., Hua, Y., Bulic, M., Jiang, Y., Carrara, M., Calogero, R.A., Kath, W.L., Kelleher, N.L., Wang, J.-P., Elemento, O. & Licht, J.D. (2014) Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. *PLoS genetics*, **10,** e1004566.

Population Projections Available at: https://www.census.gov/programs-surveys/popproj.html [Accessed August 7, 2020].

Prokaeva, T., Spencer, B., Kaut, M., Ozonoff, A., Doros, G., Connors, L.H., Skinner, M. & Seldin, D.C. Soft tissue, joint, and bone manifestations of AL amyloidosis: Clinical presentation, molecular features, and survival. *Arthritis & Rheumatism*, **56,** 3858–3868.

Psaltopoulou, T., Sergentanis, T.N., Kanellias, N., Kanavidis, P., Terpos, E. & Dimopoulos, M.A. (2013) Tobacco smoking and risk of multiple myeloma: a meta-analysis of 40 observational studies. *International Journal of Cancer*, **132,** 2413–2431.

Pukkala, E., Martinsen, J.I., Weiderpass, E., Kjaerheim, K., Lynge, E., Tryggvadottir, L., Sparén, P. & Demers, P.A. (2014) Cancer incidence among firefighters: 45 years of follow-up in five Nordic countries. *Occupational and Environmental Medicine*, **71,** 398–404.

Qiang, Y.-W., Ye, S., Chen, Y., Buros, A.F., Edmonson, R., van Rhee, F., Barlogie, B., Epstein, J., Morgan, G.J. & Davies, F.E. (2016) MAF protein mediates innate resistance to proteasome inhibition therapy in multiple myeloma. *Blood*, **128,** 2919–2930.

Quock, T.P., Yan, T., Chang, E., Guthrie, S. & Broder, M.S. (2018) Epidemiology of AL amyloidosis: a real-world study using US claims data. *Blood Advances*, **2,** 1046–1053.

Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., et al (2014) International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet. Oncology*, **15,** e538-548.

Rajkumar, S.V., Gertz, M.A. & Kyle, R.A. Primary systemic amyloidosis with delayed progression to multiple myeloma. *Cancer*, **82,** 1501–1505.

Rajkumar, S.V., Larson, D. & Kyle, R.A. (2011) Diagnosis of Smoldering Multiple Myeloma. *New England Journal of Medicine*, **365,** 474–475 Available at: https://doi.org/10.1056/NEJMc1106428 [Accessed March 27, 2020].

Rambaran, R.N. & Serpell, L.C. (2008) Amyloid fibrils. *Prion*, **2,** 112–117.

Ramirez-Alvarado, M. (2012) Amyloid formation in light chain amyloidosis. *Current Topics in Medicinal Chemistry*, **12,** 2523–2533.

Ramirez-Alvarado, M., Ward, C.J., Huang, B.Q., Gong, X., Hogan, M.C., Madden, B.J., Charlesworth, M.C. & Leung, N. (2012) Differences in immunoglobulin light chain species found in urinary exosomes in light chain amyloidosis (AI). *PloS One*, **7,** e38061.

Ramiro, A.R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M.C. (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature immunology*, **4,** 452–456.

Rand, K.A., Song, C., Dean, E., Serie, D.J., Curtin, K., Sheng, X., Hu, D., Huff, C.A., Bernal-Mizrachi, L., Tomasson, M.H., Ailawadhi, S., Singhal, S., Pawlish, K., Peters, E.S., Bock, C.H., Stram, A., Berg, D.J.V.D., Edlund, C.K., Conti, D.V., Zimmerman, T., et al (2016) A Meta-analysis of Multiple Myeloma Risk Regions in African and European Ancestry Populations Identifies Putatively Functional Loci. *Cancer Epidemiology and Prevention Biomarkers*, **25,** 1609–1618.

Ranjbaran, R. & Golafshan, H. (2018) Flaming Plasma Cell Leukemia. *Turkish Journal of Hematology*, **35,** 134.

Ravi, P., Kumar, S., Larsen, J.T., Gonsalves, W., Buadi, F., Lacy, M.Q., Go, R., Dispenzieri, A., Kapoor, P., Lust, J.A., Dingli, D., Lin, Y., Russell, S.J., Leung, N., Gertz, M.A., Kyle, R.A., Bergsagel, P.L. & Rajkumar, S.V. (2016) Evolving changes in disease biomarkers and risk of early progression in smoldering multiple myeloma. *Blood Cancer Journal*, **6,** e454.

Ravindran, A., Bartley, A.C., Holton, S.J., Gonsalves, W.I., Kapoor, P., Siddiqui, M.A., Hashmi, S.K., Marshall, A.L., Ashrani, A.A., Dispenzieri, A., Kyle, R.A., Rajkumar, S.V. & Go, R.S. (2016) Prevalence, incidence and survival of smoldering multiple myeloma in the United States. *Blood Cancer Journal*, **6,** e486.

Rawstron, A.C., Orfao, A., Beksac, M., Bezdickova, L., Brooimans, R.A., Bumbea, H., Dalva, K., Fuhler, G., Gratama, J., Hose, D., Kovarova, L., Lioznov, M., Mateo, G., Morilla, R., Mylin, A.K., Omedé, P., Pellat-Deceunynck, C., Andres, M.P., Petrucci, M., Ruggeri, M., et al (2008) Report of the European Myeloma Network on multiparametric flow cytometry in multiple myeloma and related disorders. *Haematologica*, **93,** 431–438.

Rettig, M.B., Ma, H.J., Vescio, R.A., Põld, M., Schiller, G., Belson, D., Savage, A., Nishikubo, C., Wu, C., Fraser, J., Said, J.W. & Berenson, J.R. (1997) Kaposi's Sarcoma-Associated Herpesvirus Infection of Bone Marrow Dendritic Cells from Multiple Myeloma Patients. *Science*, **276,** 1851–1854.

Ríos, R., Lupiañez, C.B., Campa, D., Martino, A., Martínez-López, J., Martínez-Bueno, M., Varkonyi, J., García-Sanz, R., Jamroziak, K., Dumontet, C., Cayuela, A.J., Wętek, M., Landi, S., Rossi, A.M., Lesueur, F., Reis, R.M., Moreno, V., Marques, H., Jurczyszyn, A., Andersen, V., et al (2015) Type 2 diabetes-related variants influence the risk of developing multiple myeloma: results from the IMMEnSE consortium. *Endocrine-Related Cancer*, **22,** 545–559.

Rodier, F. & Campisi, J. (2011) Four faces of cellular senescence. *The Journal of Cell Biology*, **192,** 547–556.

Rossi, A., Voigtlaender, M., Janjetovic, S., Thiele, B., Alawi, M., März, M., Brandt, A., Hansen, T., Radloff, J., Schön, G., Hegenbart, U., Schönland, S., Langer, C., Bokemeyer, C. & Binder, M. (2017) Mutational landscape reflects the biological continuum of plasma cell dyscrasias. *Blood Cancer Journal*, **7,** e537.

Rothstein, T.L., Griffin, D.O., Holodick, N.E., Quach, T.D. & Kaku, H. (2013) Human B-1 cells take the stage. *Annals of the New York Academy of Sciences*, **1285,** 97–114.

Ruder, A.M., Hein, M.J., Hopf, N.B. & Waters, M.A. (2014) Mortality among 24,865 workers exposed to polychlorinated biphenyls (PCBs) in three electrical capacitor manufacturing plants: a ten-year update. *International Journal of Hygiene and Environmental Health*, **217,** 176–187.

Rustad, E.H., Yellapantula, V.D., Glodzik, D., Maclachlan, K.H., Diamond, B., Boyle, E.M., Ashby, C., Blaney, P., Gundem, G., Hultcrantz, M., Leongamornlert, D., Angelopoulos, N., Agnelli, L., Auclair, D., Zhang, Y., Dogan, A., Bolli, N., Papaemmanuil, E., Anderson, K.C., Moreau, P., et al (2020) Revealing the impact of structural variants in multiple myeloma. *Blood Cancer Discovery* Available at: https://bloodcancerdiscov.aacrjournals.org/content/early/2020/09/15/2643-3230.BCD-20-0132 [Accessed September 21, 2020].

Sachchithanantham, S., Offer, M., Venner, C., Mahmood, S.A., Foard, D., Rannigan, L., Lane, T., Gillmore, J.D., Lachmann, H.J., Hawkins, P.N. & Wechalekar, A.D. (2015) Clinical profile and treatment outcome of older (>75 years) patients with systemic AL amyloidosis. *Haematologica*, **100,** 1469–1476.

Sahota, S.S., Leo, R., Hamblin, T.J. & Stevenson, F.K. (1996) Ig VH gene mutational patterns indicate different tumor cell status in human myeloma and monoclonal gammopathy of undetermined significance. *Blood*, **87,** 746–755.

Sanchorawala, V. (2006) Light-Chain (AL) Amyloidosis: Diagnosis and Treatment. *Clinical Journal of the American Society of Nephrology*, **1,** 1331–1341 Available at: https://cjasn.asnjournals.org/content/1/6/1331 [Accessed April 10, 2020].

Sanchorawala, V., Blanchard, E., Seldin, D.C., O'Hara, C., Skinner, M. & Wright, D.G. (2006) AL amyloidosis associated with B-cell lymphoproliferative disorders: frequency and treatment outcomes. *American Journal of Hematology*, **81,** 692–695.

Sant, M., Allemani, C., Tereanu, C., De Angelis, R., Capocaccia, R., Visser, O., Marcos-Gragera, R., Maynadié, M., Simonetti, A., Lutz, J.-M., Berrino, F. & HAEMACARE Working Group (2010) Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood*, **116,** 3724–3734.

Schinke, C., Boyle, E.M., Ashby, C., Wang, Y., Lyzogubov, V., Wardell, C., Qu, P., Hoering, A., Deshpande, S., Ryan, K., Thanendrarajan, S., Mohan, M., Yarlagadda, N., Khan, M., Choudhury, S.R., Zangari, M., van Rhee, F., Davies, F., Barlogie, B., Morgan, G., et al (2020) Genomic analysis of primary plasma cell leukemia reveals complex structural alterations and high-risk mutational patterns. *Blood Cancer Journal*, **10,** Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7303180/ [Accessed October 19, 2020].

Schroeder, H.W. & Cavacini, L. (2010) Structure and Function of Immunoglobulins. *The Journal of allergy and clinical immunology*, **125,** S41–S52 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3670108/ [Accessed April 10, 2020].

Schubauer-Berigan, M.K., Daniels, R.D., Bertke, S.J., Tseng, C.-Y. & Richardson, D.B. (2015) Cancer Mortality through 2005 among a Pooled Cohort of U.S. Nuclear Workers Exposed to External Ionizing Radiation. *Radiation Research*, **183,** 620–631.

Schultz, C.P. (2000) Illuminating folding intermediates. *Nature Structural & Molecular Biology*, **7,** 7–10.

SEER data, US Population data 1969-2012 - SEER Datasets. Available at: http://seer.cancer.gov/data/citation.html [Accessed January 4, 2015].

Shaughnessy, J.D., Zhan, F., Burington, B.E., Huang, Y., Colla, S., Hanamura, I., Stewart, J.P., Kordsmeier, B., Randolph, C., Williams, D.R., Xiao, Y., Xu, H., Epstein, J., Anaissie, E., Krishna, S.G., Cottler-Fox, M., Hollmig, K., Mohiuddin, A., Pineda-Roman, M., Tricot, G., et al (2007) A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*, **109,** 2276–2284.

Shen, M.M. (2013) Chromoplexy: a new category of complex rearrangements in the cancer genome. *Cancer cell*, **23,** 567–569.

Sikkink, L.A. & Ramirez-Alvarado, M. (2010) Cytotoxicity of amyloidogenic immunoglobulin light chains in cell culture. *Cell Death & Disease*, **1,** e98.

da Silva Filho, M.I., Försti, A., Weinhold, N., Meziane, I., Campo, C., Huhn, S., Nickel, J., Hoffmann, P., Nöthen, M.M., Jöckel, K.-H., Landi, S., Mitchell, J.S., Johnson, D., Morgan, G.J., Houlston, R., Goldschmidt, H., Jauch, A., Milani, P., Merlini, G., Rowcieno, D., et al (2017) Genome-wide association study of immunoglobulin light chain amyloidosis in three patient cohorts: comparison with myeloma. *Leukemia*, **31,** 1735–1742.

Siontis, B., Kumar, S., Dispenzieri, A., Drake, M.T., Lacy, M.Q., Buadi, F., Dingli, D., Kapoor, P., Gonsalves, W., Gertz, M.A. & Rajkumar, S.V. (2015) Positron emission tomography-computed tomography in the diagnostic evaluation of smoldering multiple myeloma: identification of patients needing therapy. *Blood Cancer Journal*, **5,** e364.

Sipe, J.D., Benson, M.D., Buxbaum, J.N., Ikeda, S., Merlini, G., Saraiva, M.J.M. & Westermark, P. (2014) Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis. *Amyloid*, **21,** 221–224.

Smith, A., Roman, E., Howell, D., Jones, R., Patmore, R., Jack, A. & Haematological Malignancy Research Network (2010) The Haematological Malignancy Research Network (HMRN): a new information strategy for population based epidemiology and health service research. *British Journal of Haematology*, **148,** 739–753.

Snow, A.D., Willmer, J. & Kisilevsky, R. (1987) A close ultrastructural relationship between sulfated proteoglycans and AA amyloid fibrils. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, **57,** 687–698.

Solomon, A., Frangione, B. & Franklin, E.C. (1982) Bence Jones proteins and light chains of immunoglobulins. Preferential association of the V lambda VI subgroup of human light chains with amyloidosis AL (lambda). *The Journal of Clinical Investigation*, **70,** 453–460.

Stankowski-Drengler, T., Gertz, M.A., Katzmann, J.A., Lacy, M.Q., Kumar, S., Leung, N., Hayman, S.R., Buadi, F., Kyle, R.A., Rajkumar, S.V. & Dispenzieri, A. (2010) Serum immunoglobulin free light chain measurements and heavy chain isotype usage provide insight into disease biology in patients with POEMS syndrome. *American journal of hematology*, **85,** 431–434 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2902776/ [Accessed April 10, 2020].

Steensma, D.P., Gertz, M.A., Greipp, P.R., Kyle, R.A., Lacy, M.Q., Lust, J.A., Offord, J.R., Plevak, M.F., Therneau, T.M. & Witzig, T.E. (2001) A high bone marrow plasma cell labeling index in stable plateau-phase multiple myeloma is a marker for early disease progression and death. *Blood*, **97,** 2522–2523.

Stevens, F.J., Pokkuluri, P.R. & Schiffer, M. (2000) Protein conformation and disease: pathological consequences of analogous mutations in homologous proteins. *Biochemistry*, **39,** 15291–15296.

Stevens, P.W., Raffen, R., Hanson, D.K., Deng, Y.L., Berrios-Hammond, M., Westholm, F.A., Murphy, C., Eulitz, M., Wetzel, R. & Solomon, A. (1995) Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for in vitro characterization of light-chain amyloid proteins. *Protein Science: A Publication of the Protein Society*, **4,** 421–432.

Stich, M.H., Irving Swiller, A. & Morrison, M. (1955) The “Grape Cell” of Multiple Myeloma. *American Journal of Clinical Pathology*, **25,** 601–602.

Teng, J., Russell, W.J., Gu, X., Cardelli, J., Jones, M.L. & Herrera, G.A. (2004) Different types of glomerulopathic light chains interact with mesangial cells using a common receptor but exhibit different intracellular trafficking patterns. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, **84,** 440–451.

Teras, L.R., Kitahara, C.M., Birmann, B.M., Hartge, P.A., Wang, S.S., Robien, K., Patel, A.V., Adami, H.-O., Weiderpass, E., Giles, G.G., Singh, P.N., Alavanja, M., Beane Freeman, L.E., Bernstein, L., Buring, J.E., Colditz, G.A., Fraser, G.E., Gapstur, S.M., Gaziano, J.M., Giovannucci, E., et al (2014) Body size and multiple myeloma mortality: a pooled analysis of 20 prospective studies. *British Journal of Haematology*, **166,** 667–676.

Thakurta, A., Ortiz, M., Blecua, P., Towfic, F., Corre, J., Serbina, N.V., Flynt, E., Yu, Z., Yang, Z., Palumbo, A., Dimopoulos, M.A., Gutierrez, N.C., Goldschmidt, H., Sonneveld, P. & Avet-Loiseau, H. (2019) High subclonal fraction of 17p deletion is associated with poor prognosis in multiple myeloma. *Blood*, **133,** 1217–1221.

Theis, J.D., Dasari, S., Vrana, J.A., Kurtin, P.J. & Dogan, A. (2013) Shotgun-proteomics-based clinical testing for diagnosis and classification of amyloidosis. *Journal of mass spectrometry: JMS*, **48,** 1067–1077.

Tiacci, E., Trifonov, V., Schiavoni, G., Holmes, A., Kern, W., Martelli, M.P., Pucciarini, A., Bigerna, B., Pacini, R., Wells, V.A., Sportoletti, P., Pettirossi, V., Mannucci, R., Elliott, O., Liso, A., Ambrosetti, A., Pulsoni, A., Forconi, F., Trentin, L., Semenzato, G., et al (2011) BRAF mutations in hairy-cell leukemia. *The New England Journal of Medicine*, **364,** 2305–2315.

Treon, S.P., Xu, L., Yang, G., Zhou, Y., Liu, X., Cao, Y., Sheehy, P., Manning, R.J., Patterson, C.J., Tripsas, C., Arcaini, L., Pinkus, G.S., Rodig, S.J., Sohani, A.R., Harris, N.L., Laramie, J.M., Skifter, D.A., Lincoln, S.E. & Hunter, Z.R. (2012) MYD88 L265P somatic mutation in Waldenström’s macroglobulinemia. *The New England Journal of Medicine*, **367,** 826–833.

Usmani, S.Z., Heuck, C., Mitchell, A., Szymonifka, J., Nair, B., Hoering, A., Alsayed, Y., Waheed, S., Haider, S., Restrepo, A., Van Rhee, F., Crowley, J. & Barlogie, B. (2012) Extramedullary disease portends poor prognosis in multiple myeloma and is over-represented in high-risk disease even in the era of novel agents. *Haematologica*, **97,** 1761–1767.

V, L., B, P.-G., N, A., G, L.-A., P, G., N, P., Jp, Z. & M, P. (2008) Occupation, exposure to chemicals, sensitizing agents, and risk of multiple myeloma in Sweden. *Cancer Epidemiology, Biomarkers & Prevention : a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, **17,** 3123–3127.

Varettoni, M., Corso, A., Cocito, F., Mangiacavalli, S., Pascutto, C., Zappasodi, P., Pica, G. & Lazzarino, M. (2010) Changing pattern of presentation in monoclonal gammopathy of undetermined significance: a single-center experience with 1400 patients. *Medicine*, **89,** 211–216.

Victor, K.D. & Capra, J.D. (1994) An apparently common mechanism of generating antibody diversity: length variation of the VL-JL junction. *Molecular immunology*, **31,** 39–46.

Victor, K.D., Vu, K. & Feeney, A.J. (1994) Limited junctional diversity in kappa light chains. Junctional sequences from CD43+B220+ early B cell progenitors resemble those from peripheral B cells. *Journal of immunology (Baltimore, Md.: 1950)*, **152,** 3467–3475.

Vuong, B.Q., Herrick-Reynolds, K., Vaidyanathan, B., Pucella, J.N., Ucher, A.J., Donghia, N.M., Gu, X., Nicolas, L., Nowak, U., Rahman, N., Strout, M.P., Mills, K.D., Stavnezer, J. & Chaudhuri, J. (2013) A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. *Nature immunology*, **14,** 1183–1189.

Walker, B.A., Boyle, E.M., Wardell, C.P., Murison, A., Begum, D.B., Dahir, N.M., Proszek, P.Z., Johnson, D.C., Kaiser, M.F., Melchor, L., Aronson, L.I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J.R., Brioli, A., Mikulasova, A., Cairns, D.A., Gregory, W.M., Quartilho, A., et al (2015a) Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **33,** 3911–3920.

Walker, B.A., Leone, P.E., Chiecchio, L., Dickens, N.J., Jenner, M.W., Boyd, K.D., Johnson, D.C., Gonzalez, D., Dagrada, G.P., Protheroe, R.K.M., Konn, Z.J., Stockley, D.M., Gregory, W.M., Davies, F.E., Ross, F.M. & Morgan, G.J. (2010a) A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*, **116,** e56–e65.

Walker, B.A., Leone, P.E., Chiecchio, L., Dickens, N.J., Jenner, M.W., Boyd, K.D., Johnson, D.C., Gonzalez, D., Dagrada, G.P., Protheroe, R.K.M., Konn, Z.J., Stockley, D.M., Gregory, W.M., Davies, F.E., Ross, F.M. & Morgan, G.J. (2010b) A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*, **116,** e56-65.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2019) A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*, **33,** 159–170.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018) Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*, **132,** 587–597.

Walker, B.A., Wardell, C.P., Brioli, A., Boyle, E., Kaiser, M.F., Begum, D.B., Dahir, N.B., Johnson, D.C., Ross, F.M., Davies, F.E. & Morgan, G.J. (2014a) Translocations at 8q24 juxtapose MYC with genes

that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients. *Blood Cancer Journal*, **4,** e191.

Walker, B.A., Wardell, C.P., Chiecchio, L., Smith, E.M., Boyd, K.D., Neri, A., Davies, F.E., Ross, F.M. & Morgan, G.J. (2011) Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*, **117,** 553–562.

Walker, B.A., Wardell, C.P., Johnson, D.C., Kaiser, M.F., Begum, D.B., Dahir, N.B., Ross, F.M., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2013a) Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood*, **121,** 3413–3419.

Walker, B.A., Wardell, C.P., Johnson, D.C., Kaiser, M.F., Begum, D.B., Dahir, N.B., Ross, F.M., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2013b) Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood*, **121,** 3413–3419.

Walker, B.A., Wardell, C.P., Melchor, L., Brioli, A., Johnson, D.C., Kaiser, M.F., Mirabella, F., Lopez-Corral, L., Humphray, S., Murray, L., Ross, M., Bentley, D., Gutiérrez, N.C., Garcia-Sanz, R., San Miguel, J., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2014b) Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*, **28,** 384–390.

Walker, B.A., Wardell, C.P., Murison, A., Boyle, E.M., Begum, D.B., Dahir, N.M., Proszek, P.Z., Melchor, L., Pawlyn, C., Kaiser, M.F., Johnson, D.C., Qiang, Y.-W., Jones, J.R., Cairns, D.A., Gregory, W.M., Owen, R.G., Cook, G., Drayson, M.T., Jackson, G.H., Davies, F.E., et al (2015b) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nature Communications*, **6,** 6997.

Wallin, A. & Larsson, S.C. (2011) Body mass index and risk of multiple myeloma: a meta-analysis of prospective studies. *European Journal of Cancer (Oxford, England: 1990)*, **47,** 1606–1615.

Wang, J., Cao, X., Zhao, A., Cai, H., Wang, X. & Li, J. (2018) Increased activated regulatory T cell subsets and aging Treg-like cells in multiple myeloma and monoclonal gammopathy of undetermined significance: a case control study. *Cancer Cell International*, **18,** 187.

Warsame, R., Kumar, S.K., Gertz, M.A., Lacy, M.Q., Buadi, F.K., Hayman, S.R., Leung, N., Dingli, D., Lust, J.A., Ketterling, R.P., Lin, Y., Russell, S., Hwa, L., Kapoor, P., Go, R.S., Zeldenrust, S.R., Kyle, R.A., Rajkumar, S.V. & Dispenzieri, A. (2015) Abnormal FISH in patients with immunoglobulin light chain amyloidosis is a risk factor for cardiac involvement and for death. *Blood Cancer Journal*, **5,** e310.

Weng, L., Spencer, B.H., SoohHoo, P.T., Connors, L.H., O'Hara, C.J. & Seldin, D.C. (2011) Dysregulation of miRNAs in AL amyloidosis. *Amyloid: The International Journal of Experimental and Clinical Investigation: The Official Journal of the International Society of Amyloidosis*, **18,** 128–135.

Went, M., Sud, A., Försti, A., Halvarsson, B.-M., Weinhold, N., Kimber, S., van Duin, M., Thorleifsson, G., Holroyd, A., Johnson, D.C., Li, N., Orlando, G., Law, P.J., Ali, M., Chen, B., Mitchell, J.S., Gudbjartsson, D.F., Kuiper, R., Stephens, O.W., Bertsch, U., et al (2018) Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. *Nature Communications*, **9,** 3707.

Went, M., Sud, A., Law, P.J., Johnson, D.C., Weinhold, N., Försti, A., van Duin, M., Mitchell, J.S., Chen, B., Kuiper, R., Stephens, O.W., Bertsch, U., Campo, C., Einsele, H., Gregory, W.M., Henrion, M., Hillengass, J., Hoffmann, P., Jackson, G.H., Lenive, O., et al (2017) Assessing the effect of obesity-

related traits on multiple myeloma using a Mendelian randomisation approach. *Blood Cancer Journal*, **7,** e573–e573.

Wetzel, R. (1997) Domain stability in immunoglobulin light chain deposition disorders. *Advances in Protein Chemistry*, **50,** 183–242.

WHO 2018 ICD-10-CM Codes. Available at: https://www.icd10data.com/ICD10CM/Codes/E00-E89/E70-E88/E85- [Accessed June 6, 2018].

Wu, K.-D., Orme, L.M., Shaughnessy, J., Jacobson, J., Barlogie, B. & Moore, M.A.S. (2003) Telomerase and telomere length in multiple myeloma: correlations with disease heterogeneity, cytogenetic status, and overall survival. *Blood*, **101,** 4982–4989.

Wuilleme, S., Robillard, N., Lodé, L., Magrangeas, F., Beris, H., Harousseau, J.-L., Proffitt, J., Minvielle, S. & Avet-Loiseau, H. (2005) Ploidy, as detected by fluorescence *in situ* hybridization, defines different subgroups in multiple myeloma. *Leukemia*, **19,** 275–278.

Xu, Z., Zan, H., Pone, E.J., Mai, T. & Casali, P. (2012) Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nature reviews. Immunology*, **12,** 517–531.

Yoon, J.-H., Kwak, W.S. & Ahn, Y.-S. (2018) A brief review of relationship between occupational benzene exposure and hematopoietic cancer. *Annals of Occupational and Environmental Medicine*, **30,** 33.

Yu, P.L.I., Wang, R.R., Johnston, G., Wang, Y., Tammur, P., Tamm, A., Punab, M., Rangel-Pozzo, A. & Mai, S. (2019) Distinct Nuclear Organization of Telomeres and Centromeres in Monoclonal Gammopathy of Undetermined Significance and Multiple Myeloma. *Cells*, **8,** Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6678424/ [Accessed October 7, 2020].

Yuan, Y., Ju, Y.S., Kim, Y., Li, J., Wang, Y., Yoon, C.J., Yang, Y., Martincorena, I., Creighton, C.J., Weinstein, J.N., Xu, Y., Han, L., Kim, H.-L., Nakagawa, H., Park, K., Campbell, P.J. & Liang, H. (2020) Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nature Genetics*, **52,** 342–352.

Zhan, F., Huang, Y., Colla, S., Stewart, J.P., Hanamura, I., Gupta, S., Epstein, J., Yaccoby, S., Sawyer, J., Burington, B., Anaissie, E., Hollmig, K., Pineda-Roman, M., Tricot, G., van Rhee, F., Walker, R., Zangari, M., Crowley, J., Barlogie, B. & Shaughnessy, J.D. (2006) The molecular classification of multiple myeloma. *Blood*, **108,** 2020–2028.

Chapter 2: Methods and Validation

2.1. Patients

2.1.1. Newly diagnosed light-chain amyloidosis patients

Newly diagnosed light-chain amyloidosis (AL) patients were included prospectively, at the National Amyloid Centre, University College London Medical School, United-Kingdom, from 2013-2014, prior to any treatment, after informed consent (Prof A. Wechalekar). All cases were histologically proven by an expert pathologist and typing confirmed using either immunohistochemistry or mass-spectrometry.

2.1.2. UAMS Newly diagnosed myeloma patients

Newly diagnosed myeloma patients (NDMM) were recruited onto one of the Total Therapy Trials (TTT) 3a through 6 at the University of Arkansas for Medical Science's (UAMS) myeloma centre, Little Rock, Arkansas, United-States of America (USA), between 2004 and 2017.

In brief, these are a series of phase II and III clinical trials for fit newly-diagnosed patients where all patients received both immunomodulatory drugs (IMIDs) and proteasome inhibitors (PI) in alkylator heavy regimens. All patients were planned to receive two autologous stem cell transplants (ASCT) with Melphalan based conditioning regimens, **Figure 2.1.2.1**. The median follow-up of these trials was 8.35 years (95% CI 8.00-8.63). As part of their staging and follow-up, bone marrow aspirates were performed, and samples stored after informed consent.

Among these patients, we retrospectively selected 223 patients with paired tumour and whole blood for germline control based on sample quality and availability. The breakdown of these patients among the trials may be found in **Table 2.1.2.1.**

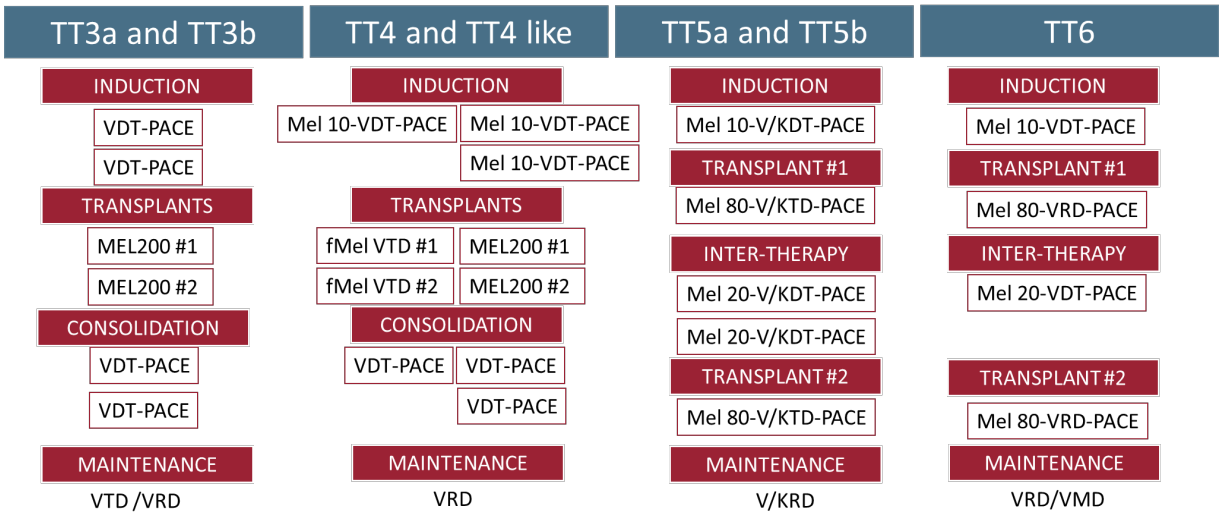


Figure 2.1.2.1. Summary of the Total Therapy Trial Designs. V= bortezomib, D=dexamethasone, T=thalidomide, P=cisplatin, A=doxorubicine, C=cyclophosphamide, E=etoposide, R=lenalidomide, M=melphalan, MEL= high dose melphalan, fMel=fractionated melphalan.

.

	223-baseline study	**Combined TT3a-3b-4-4like-5a-5b-6**
TT3a	18.67% (n=42)	26.56% (n=276)
TT3b	18.67%(n=42)	16.17 % (n=168)
TT4	42.22%(n=95)	34.94 % (n=363)
TT4-Like	0.90%(n=2)	1.06 % (n=11)
TT5a	5.33% (n=12)	7.12 % (n=74)
TT5b	5.28% (n=12)	1.54 % (n=16)
TT6	8.00% (n=18)	12.61 % (n=131)

Table 2.1.2.1. Breakdown of the 223 NDMM samples selected for sequencing according their trial of origin.

Overall, these patients were representative of the complete cohort in terms of sex, ethnicity, risk scores (ISS, and GEP70 risk group), follow-up, and outcome, **Table 2.1.2.2**.

	223-baseline study	**Combined TT3a-3b-4-4like-5a-5b-6**
Number of patients	223	1039
Inclusion dates	02/2004 to 08/2017	02/2004 to 08/2017
Median Follow Up	8.14 years (95% CI 7.39-9.02)	8.35 years (95% CI 8.00-8.63)
Median EFS	6.16 years (95% CI 5.18-7.75)	4.8 years (95% CI 52%-58%)
Median OS	8-year 61% (95% CI 54-69%)	8-year 42% (95% CI 39%-45%)
Median age (years)	59 (range: 30-75)	61 (range: 30-76)
Sex ratio M:F	1.8:1	1.6:1
Ethnicity % - **White** - **African-American** - **Other**	 88% (n=197) 10% (n=22) 2% (n=4)	 86.8% (n=902) 9.7% (n=101) 3.5% (n=36)
ISS % - **I** - **II** - **III**	 26.5% (n=59) 43.5% (n=97) 30.0% (n=67)	 34.0% (n=352) 40.2% (n=416) 25.8% (n=267)
R-ISS % - **I** - **II** - **III**	 17.0% (n=38) 67.7% (n=151) 15.2% (n=34)	
GEP70 high risk %	16.1% (n=36)	15.9% (n=165)

Table 2.1.2.2. Comparison between the 225-baseline study and the combined TT3a-3b-4-4like-5a-5b-6 trial.

2.1.3. UAMS Smouldering myeloma patients

Smouldering myeloma patients (SMM) patients were prospectively recruited on the SO-120 or MO-120 registration trials at the UAMS Myeloma centre, Little-Rock, Arkansas, USA. The SO-120 was an observational multicentric study sponsored by the South West Oncology Group (SWOG) that prospectively enrolled patients with myeloma precursor states (MGUS and SMM), solitary plasmocytomas and other plasmacell disorders

between 2006 and 2016. The aim of the study was to identify biomarkers of progression. When enrolment finished, the MO-120 was designed to recruit these patients locally at UAMS.

As part of the trial, demographic, clinical, radiological, and biological features were collected. Bone marrow samples were taken after informed consent, upon trial enrolment, and Gene expression profiling (GEP) performed at study entry and during follow-up. Unused samples were stored and patients were consent updated and obtained for genetic analysis.

A total of 225 patients were identified with SMM and had GEP available. Eighty-three of these patients had DNA or cells available for further genetic studies. Additionally, SMM sequential samples from nine patients were taken, part of the routine follow-up, and leftover DNA analysed.

This study was approved by the Institutional Review Board (IRB) of the university of Arkansas for Medical Science (#261281). All research was conducted in accordance with the Declaration of Helsinki.

2.1.4. MMRF CoMMpass cohort

The Multiple Myeloma Research Foundation (MMRF) CoMMpass trial has enrolled 1,154 participants (mean age of 63 years, range 27 to 93) across four countries (United States, Canada, Italy, Spain) beginning on July 2011 with the data updated through March 20, 2019. All participants were newly diagnosed at the start of the trial and were sampled at baseline then followed up with every six months for eight years. All trial participants gave informed consent and the data were de-identified before use in this analysis. We used the IA14 and IA15 survival update for chapter 4 and 6 respectively. The median follows up was 3.2 (95% CI 3.1-3.3) and 3.84 years (95% CI 3.71-3.93).

2.2. Sample processing

2.2.1. Amyloidosis samples

CD138+ cells were isolated from bone marrow cells using MACSorting (Miltenyi Biotech, Bisley, UK). Cells were lysed in RLT+ buffer and DNA/RNA extracted using the AllPrep kit (Qiagen, Manchester, UK). Peripheral blood was isolated from patients' white blood cells purified by Ficoll-Pacque and DNA extracted using the QIAamp DNA mini kit (Qiagen). Purity was determined by flow cytometry and only samples with >85% purity were used in this study. All samples were processed at the National Amyloid Centre, University College London Medical School, United-Kingdom, by Dr Dorota Rowczenio.

2.2.2. UAMS samples

CD138+ plasma cells were isolated from bone marrow aspirates by magnetic-activated cell sorting using the AutoMACS Pro (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) or RoboSep (STEMCELL Technologies, Vancouver, Canada). Plasma cell purity was determined by flow cytometry and only samples with >85% purity were used in targeted panel study. A lower threshold (65%) was permitted for the sequential SMM samples. DNA from peripheral blood was used as a matched non-tumour control sample for each patient to exclude germline variants. Nucleic acids were isolated using the AllPrep DNA/RNAor Puregene kits (Qiagen, Hilden, Germany). All samples were processed by the UAMS Myeloma Accessioning Laboratory team under the supervision of Dr Shayu Deshpande.

2.2.3. MMRF CoMMpass samples

Bone marrow aspirates from each patient were subjected to immunomagnetic bead separation using the Miltenyi MACS Cell Separation System (Miltenyi, San Diego, CA) to enrich for CD138-positive malignant MM plasma cells. Only clinically eligible samples with greater than 250,000 cells recovered after CD138 enrichment, which are greater than 80% monoclonal light chain restricted plasma cells move forward for nucleic acid extraction. Genomic DNA was extracted from purified CD138-positive plasma cells (tumour) and matched peripheral blood samples (constitutional) using QIAamp DNA Mini Kit (Qiagen). Total RNA was extracted from CD138-positive plasma cells the using QiaAmp RNeasy Mini Kit (Qiagen). Nucleic acids were quality assessed using the Qubit

2.0 (Thermo Fisher) and Agilent Tape Station to determine quantity and integrity. Control samples were obtained from peripheral blood. If plasma cells were detected in the blood, a CD3 selection was performed to positively select for T-cell. All samples were processed at the TGen myeloma research unit under the supervision of Dr Jonathan Keats.

2.3. Next Generation Sequencing

2.3.1. Principles

2.3.1.1. Principle of sequencing by synthesis

Next-generation sequencing (NGS) systems have been introduced in the past decade and allow for massively parallel DNA sequencing reactions that can analyse 10^6 to 10^9 reactions simultaneously. Several methods exist, the main ones being, pyrosequencing, sequencing by ligation or synthesis. This work focused on sequencing by synthesis (SBS) as developed by Illumina technologies.

In short, input genomic DNA is fragmented, and adaptors ligated to both ends of the fragments. Single stranded fragments are then randomly bound to the inside surface of a flow cell channel. Unlabelled nucleotides and enzymes are added to initiate solid phase bridge amplification, thus building double-stranded bridges. A denaturation step then leaves single stranded templates anchored. This process is repeated until dense clusters of double stranded DNA are generated in each channel on the flow cell.

The sequencing step then formally begins with the incorporation of fluorescent and terminated nucleotide. These nucleotides have been modified in two ways:

- Each is reversibly attached to a single fluorescent molecule with a unique emission wavelength.
- Each nucleotide is also reversibly terminator-bound to ensure a single nucleotide will be incorporated per cycle.

All unincorporated nucleotides are washed. Fluorescent signals are read at each cluster and recorded. Both the fluorescent and terminator group are then cleaved and washed. Modified nucleotides are then added again and the process is repeated until the

sequencing of bases in the fragment is complete. Base calls are made directly from signals intensity measured during each cycle which reduces the error rates. The base-by-sequencing eliminate sequence context specific errors, especially those seen in repetitive sequences regions or derived from homopolymers, **Figure 2.3.1.1**. Raw data are then processed and analysed.

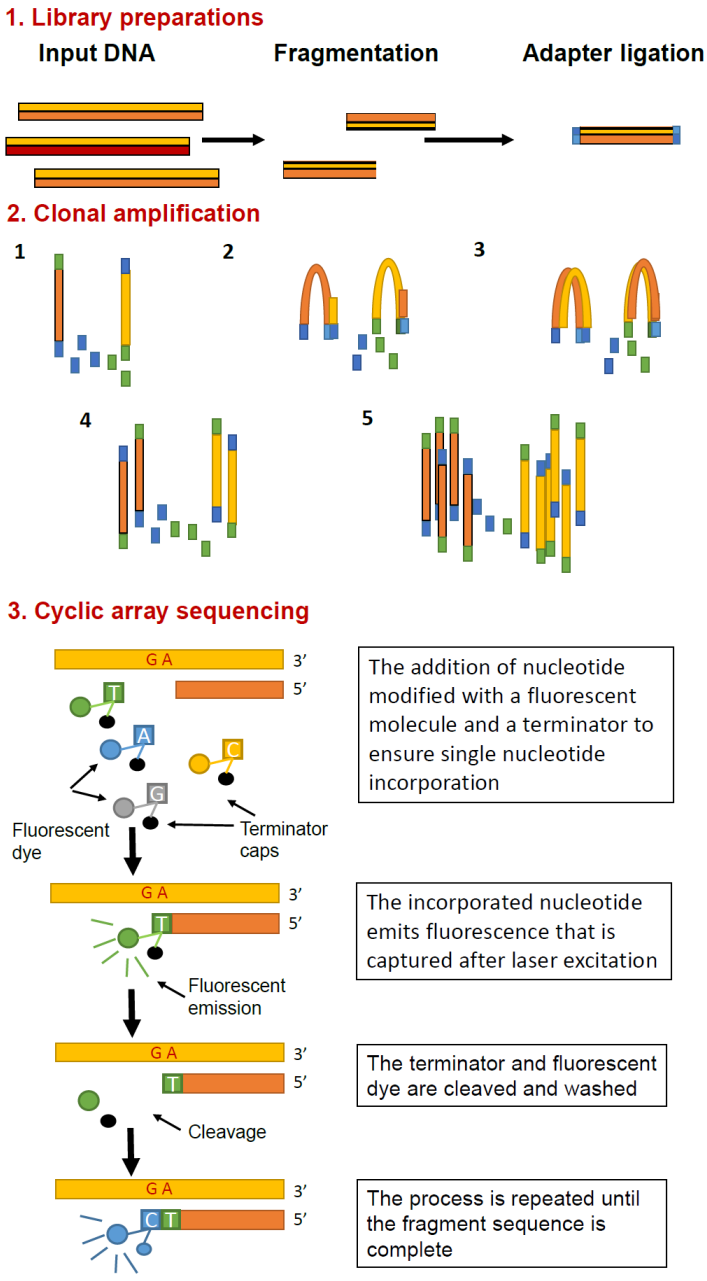


Figure 2.3.1.1. Illustration of the principles of SBS sequencing technology (Modified from (www.illumina.com))

2.3.1.2. Principles of bioinformatic analysis

NGS technologies generate a significant amount of data that require a thorough analytical process. A typical data analysis pre-processing pipeline includes, raw reads quality control, demultiplexing, alignment or mapping, and post-alignment processing. This step ensures the quality of the data that can be then analysed to determine single nucleotide variants (SNV) and Indel (Insertion Deletions), structural variants (SV), V(D)J rearrangements and Telomere Length.

2.3.1.2.1. Pre-processing

2.3.1.2.1.1. Pre-Alignment

The aim of this step is to remove the least reliable clusters from the image analysis results given by the sequencer, separate each sample individually, and generate FASTQ files. In this pre-alignment step, Illumina sequencers perform an internal quality filtering procedure termed “chastity filter”, defined as the ratio of the brightest base intensity divided by the sum of the brightest and second brightest base intensities. Clusters of reads pass the filter provided no more than one base call has a chastity value less than 0.6 in the first 25 cycles.

Following their removal, samples are demultiplexed which mean that each sample that was combined in one sequencing experiment lane are separated out and processed individually. BAMs are then split by read group and converted to the FASTQ format.

2.3.1.2.1.2. Alignment

The aim of this step is to map the reads back to the genome and remove PCR and optical duplicates that could negatively affect the quality of the variant calling.

Read groups are aligned to the reference genome using a Burrows-Wheeler aligner (BWA) algorithm. Current practice employs one of the two most recent human reference genome versions: Hg19 or Hg38. The conversion between one or the other are possible but the conversion is much higher from Hg19 to Hg38 than from Hg38 to Hg19 suggesting that, when possible, the newer version should be used at the time of SNV analysis. In contrast, this conversion has minimal impact on structural variant calling (Pan *et al*, 2019). BWA rearranges a character string into runs of similar characters. This

process is reversible without requiring further data except the position of the first character. One of the most commonly used tools for mapping is BWA-MEM (Li, 2013). This approach automatically chooses between local and end-to-end alignments, supports paired-end reads and performs chimeric alignments. This algorithm is robust to sequencing errors and applicable to a wide range of sequence lengths range from 70bp up to several mega bases. For shorter reads, algorithm such as BWA-Aln, may be used.

Every read is aligned to the reference genome separately and all read group alignments that belong to a single group are merged using Picard (Picard Tools - By Broad Institute). This tool also locates and tags duplicate reads, meaning duplicate reads that originate from a single fragment of DNA. Duplicates can arise during sample preparation following PCR based library construction. They can also result from a single amplification cluster, incorrectly detected as multiple clusters by the optical sensor of the sequencing instrument (termed “optical duplicates”). The MarkDuplicates tool works by comparing sequences in the 5’ positions of both reads and read-pairs. After duplicate reads are collected, the tool differentiates the primary and duplicate reads using an algorithm that ranks reads by the sums of their base-quality scores. Duplicated reads, are flagged, and then removed to prevent downstream variant calling errors.

2.3.1.2.1.3. Post-alignment

This step aims at reducing the amount of misalignment generated by indels, improve base quality metrics, and determine the overall quality metrics of the sequencing process.

The quality of the alignment can be improved by Indel local realignment. When not flagged properly, misalignments resulting from Indels, can incorrectly be scored as substitions and reduce the accuracy of the variant calling steps. Variant callers with reassembly steps such as Mutect2 (Mutect2) do not require an indel realignment but this step may still be useful to improve base quality score recalibration.

The next important post-alignment step is to determine basic sample metrics such as coverage, and GC bias. Coverage represents the number of times a read maps to a specific genomic target: the deeper the coverage of a target region, the more reliable and sensitive the sequencing assay. Achieving robust sequencing results requires that a

certain percentage of the targeted regions reach a certain coverage depth. While 0.2-0.5x is sufficient to determine CNA in a ULP-WGS, it is not sufficient for SNV. A good estimate for the required depth is to divide 20x by the expected allele frequency: e.g. to determine a mutation with a 10% variant allele frequency, requires a 200x coverage.

The AT/GC dropout metrics indicate the degree of inadequate coverage of a particular region based on its AT or GC content. Regions of high and low G/C content have been shown to interfere with mapping/aligning, ultimately leading to fragmented genome assemblies and poor coverage in a phenomenon known as “GC bias” (Chen *et al*, 2013). The CalculateHsMetrics tool in Picard (Picard Tools - By Broad Institute) can be used at this step to compute all these quality markers.

Aligned and cleaned BAM files are then processes as paired tumour and normal (tumour analysis) or normal only (germline analysis).

2.3.1.2.2. Processing

2.3.1.2.2.1. Variant calling and annotation

The aim of this step is to correctly identify SNVs and Indels. The majority of current somatic variant callers are designed to analyse matched tumour and normal samples. The basic idea is to identify potential variants using the tumour and use the normal to distinguish them from somatic variants from germline and loss of heterozygosity (LOH). Aneuploidy is a concern with most variant callers that model joint genotypes which has led to the adoption of approaches that model joint allele frequency such as Mutect2, LoFreq, or Strelka2. Systematic comparison of each caller are available and suggest that Strekla2 performs better for higher mutation frequency (≥20%), while Mutect2 and LoFreq performs better when the mutation frequency is lower than 10% (Chen *et al*, 2020). Several callers can be used and a consensus list of variants reported and merged in a unique VCF.

Additional filters can be positioned at this step:

- A strand bias filter detects artifacts that are seen on a single strand and not the other. They are based on Fisher’s test to identify imbalanced distributions across both strands

- A filter that focuses on repetitive regions which lead to false calls due to alignment errors.

For germline data, the underlying assumption is that the expected VAF is expected to be either 50% or 100%. Variant callers determine, which of the three genotypes (AA, AB, BB) best fits the data. Most artefact are present at low frequencies, like some real variants such as clonal haematopoiesis. This requires sensitive models with specific error correction technologies such as LoFreq.

Once called, variants are annotated which aims to assign functional information relative to the variants. Variants in the VCF files are also matched from external databases such as COSMIC using tools such as Variant Effect Predictor (VEP).

2.3.1.2.2.2. Structural variant calling

Structural variants (SV) are large genomic alterations, which usually encompasses at least 50 bp. These genomic variants are typically classified as translcations, deletions, duplications, insertions, and inversions. They can be detected using short read sequencing by mapping distance and orientation of reads. To add precision and increase resolutions split reads can be used by packages such as DELLY, that still lack resolution for the detection of larger event. To overcome this, Manta, has included coverage information. This makes it one of the most reliable inter or intra-chromosomal rearrangement packages.

Copy number variations (CNVs) are a particular subtype of SVs mainly represented by gains, deletions and duplications. Many tools for CNVs detection from NGS data have been developed. Most of them can reliably call large CNVs (in the order of Mb). In addition, most of these tools were designed to work with whole-genome or whole-exome data and struggle with the sparser data from target panel which still require visual validations.

2.3.1.2.2.3. Telomere length assessment

A number of approaches for estimating telomere length from whole-genome sequencing data have been proposed. Among them Telomerecat is designed to be

agnostic to the number of telomeres present, by quantifying the length of number of telomeric motifs and the subtelomeric regions (Farmery *et al*, 2018).

2.3.1.2.2.4. Immunoglobulin rearrangement assessment

The immunological diversity of lymphocytes mainly comes from the V(D)J rearrangements. Usual SV algorithms are poorly suited for this given the mapping difficulties in this highly repetitive region and highly rearranged region. Many algorithms are available the easiest to implement being ViDjil (Giraud *et al*, 2014) which also offers an estimation of the tumour fraction.

Figure 2.3.1.2. Overview of the bioinformatic pipeline

2.3.2.2. Targeted sequencing

The targeted panel was divided into a translocation panel and a mutation/copy number panel to provide high depth coverage for mutation analysis (0.6 Mb), whilst providing lower depth sequencing of translocation regions (4.2 Mb).

Each patient had their tumour DNA from bone marrow and control DNA from peripheral blood sequenced, to identify somatic mutations, copy number changes and translocations. 50 ng of DNA was used to prepare libraries using the HyperPlus kit (Kapa Biosystems) and split for hybridizing to both mutation and translocation captures (SeqCap EZ target enrichment; Nimblegen), after which mutation and translocation captures were combined. The HiSeq 2500 or NextSeq500 (Illumina, San Diego, CA, USA) were used for sequencing with 75 bp paired-end reads.

2.3.2.3. Pipeline

bcl2fastq was used for demultiplexing and BWA mem (v. 0.7.12) for alignment to Ensembl (GRCh37/hg19) human reference genome. Strelka (v.1.0.14) was used for variant calling and single nucleotide variants (SNVs) were filtered using fpfilter (https://github.com/ckandoth/variant-filter). Indels were filtered using a 10% variant allele frequency (VAF) cut-off. Variants were annotated using Variant Effect Predictor (v.85). To determine copy number, a normalized depth comparison between tumour and control samples was used and segments of SNP variance were utilized to identify regions of chromosomal deletion and gain. Copy number was manually normalized based on the ratio and SNP allele calls using the best fitting chromosomes with the least variance (usually chromosome 2 or 10). Data were visualized using a custom built R Shiny application showing the mutations, translocations, copy number, QC metrics and cross-sample contamination estimations. Intra- and inter-chromosomal rearrangements were called using Manta (v0.29.6) with default settings and the exome flag specified. QC metrics estimated the cross-sample contamination of samples using homozygous SNPs in the germline with 95% or higher VAF examined in the tumour sample. A VAF density plot on those SNPs was generated, as well as reporting the minimum, maximum and median of their values in the germline and tumour, **Figure 2.3.2.3.1.**

2. Methods

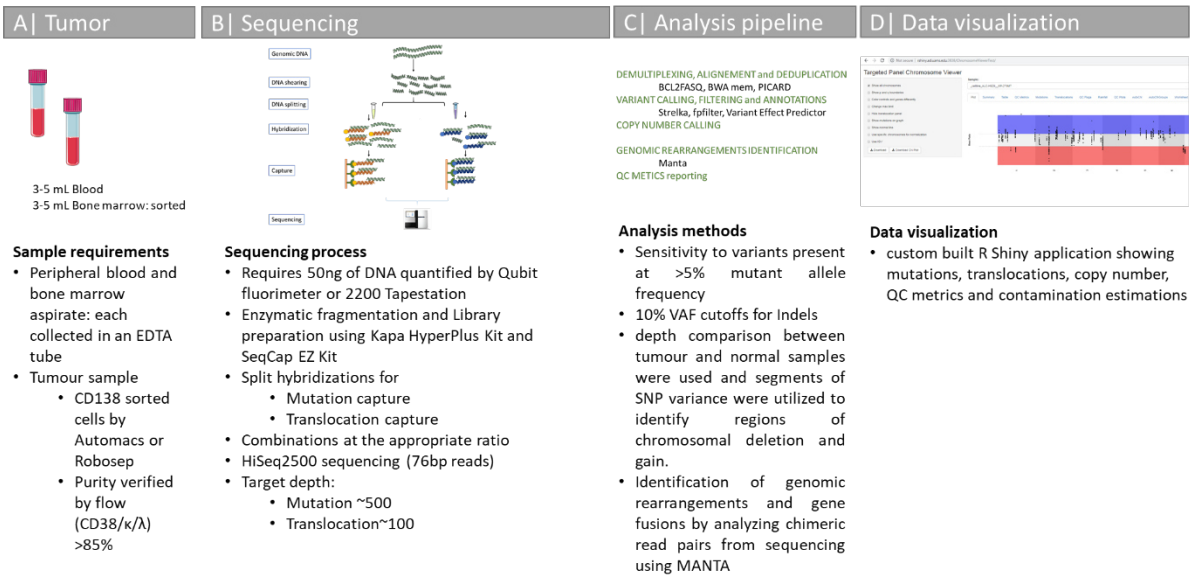


Figure 2.3.2.3.1. Summary of the workflow

2.3.2.4. Validation

SNVs: SNVs were compared and validated using seven samples (Horizon Diagnostics) with known SNVs and VAF. The VAF of mutations found in the validation samples matched those found on the panel with r^2=0.93, **Figure 2.3.2.4.1**.

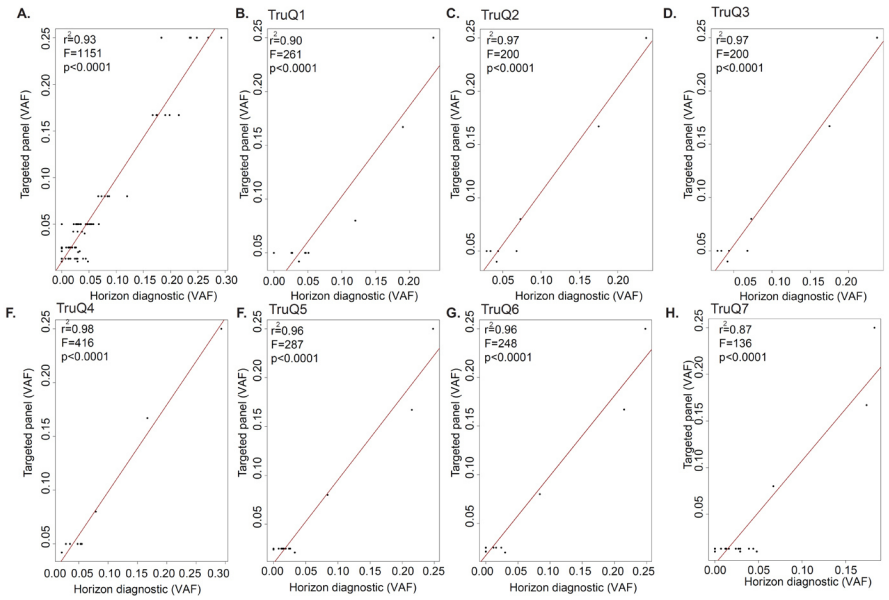


Figure 2.3.2.4.1. Validation of Mutations. All mutations with VAF ≥5% were found on the targeted panel with a good overall correlation (r2=0.93) A. All samples combined, B-H. Individual reference standard

FISH: Copy number data generated from the sequencing panel were validated against existing FISH data for del(1p) (1p13 FISH vs. 1p12 (*FAM46C*) seq.), gain/amp(1q21), del(13q) (*D13S31* vs. *RB1*), and del(17p) (*TP53*). Plots of comparisons between FISH and sequencing data are shown with specificities and sensitivities of each region at the 20%, 25%, 40%, and 50% FISH cut-off, **Figure 2.3.2.4.2 and Table 2.3.2.4.1**. Copy number was determined using smaller segments using two consecutive segments to define a copy number changes thus detecting interstitial deletions in genes such as *TP53* with a greater accuracy.

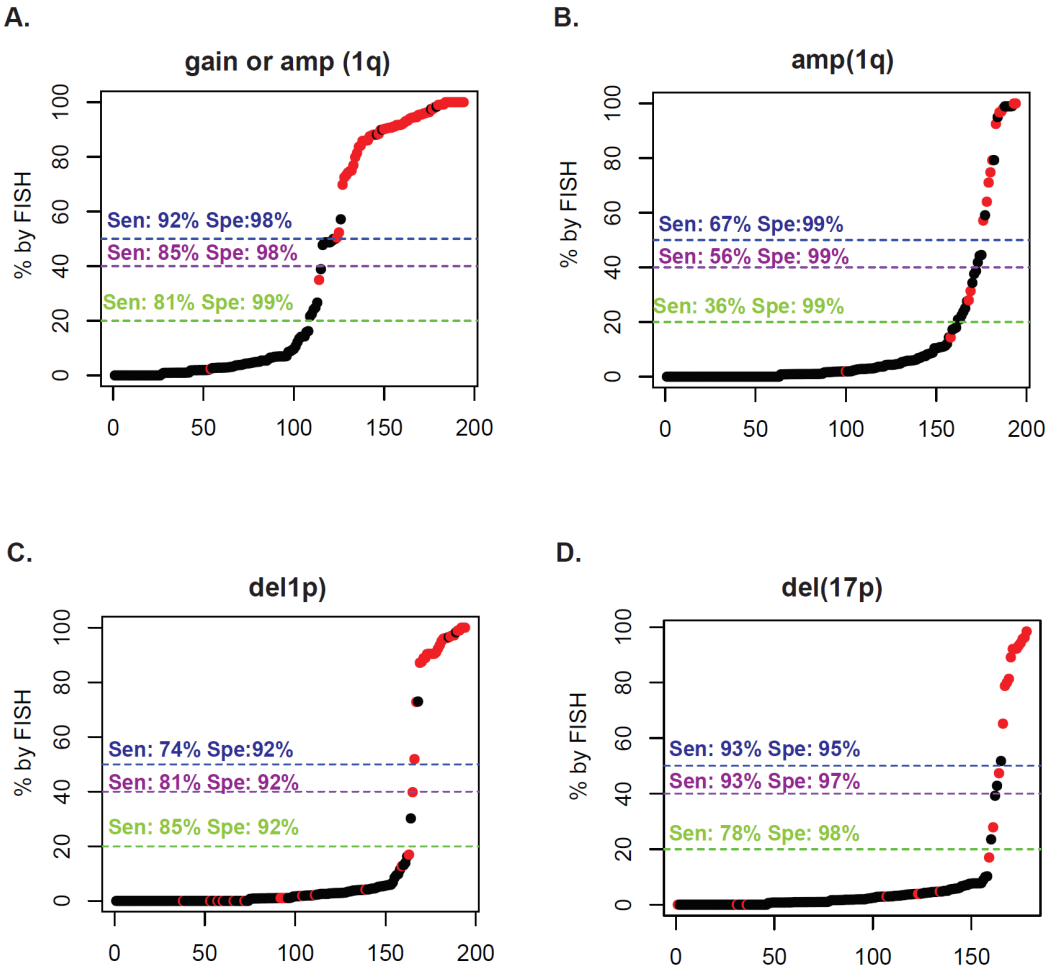


Figure 2.3.2.4.2. Validation of Copy Number Metrics Sequencing vs. FISH copy number calling for A. del(17p) (*TP53*), B. del(1p) (1p13 FISH probe compared to FAM46C (1p12) deletion) C. gain/amp (1q)(*CKS1B*), D. amp(1q)(*CKS1B*).

Sequencing	FISH (20% cut-off)		
del1p (n=166)	del1p13*	normal	
Del1p12 (*FAM46C*)	22	11	sensitivity = 84.62% (95% CI 65.13-95.64)
normal	4	129	specificity = 92.14% (95% CI 86.38-96.01)
gain1q (n=166)	gain/amp CKS1B	normal	
gain/amp1q (1q21.3)	58	1	sensitivity = 80.56% (95% CI 69.53-88.94)
normal	14	93	specificity = 98.94% (95% CI 94.21-99.97)
amp1q (n=166)	amp CKS1B	not amp	
amp1q (1q21.3)	9	1	sensitivity = 36.00% (95% CI 17.97-57.48)
not amp	16	140	specificity = 99.29% (95% CI 96.11-99.98)
del13q (n=66)	del13q (D13S31)	normal	
del13q (*RB1*)	32	1	sensitivity = 94.12% (95% CI 80.32-99.28)
normal	2	31	specificity = 96.88% (95% CI 83.78-99.92)
del17p (n=158)	del17p (TP53)	normal	
del*TP53*	14	3	sensitivity = 77.78% (95% CI 52.36-93.59)
normal	4	137	specificity = 97.86% (95% CI 93.87-99.56)

Table 2.3.2.4.2. Comparison of Sequencing and FISH calls for deletions of 1p12, 13q, and 17p13.1 and gain/amplification of 1q21.

An additional comparison for *TP53* was made using the prognostic cut-off 55% (Thakurta *et al*, 2019). All deletions identified by FISH were identified using the targeted panel. Five additional deletions were called using the panel, 3/5 of them having a del(17p) in at least 20% of cells by FISH, **Figure 2.3.2.4.3.**

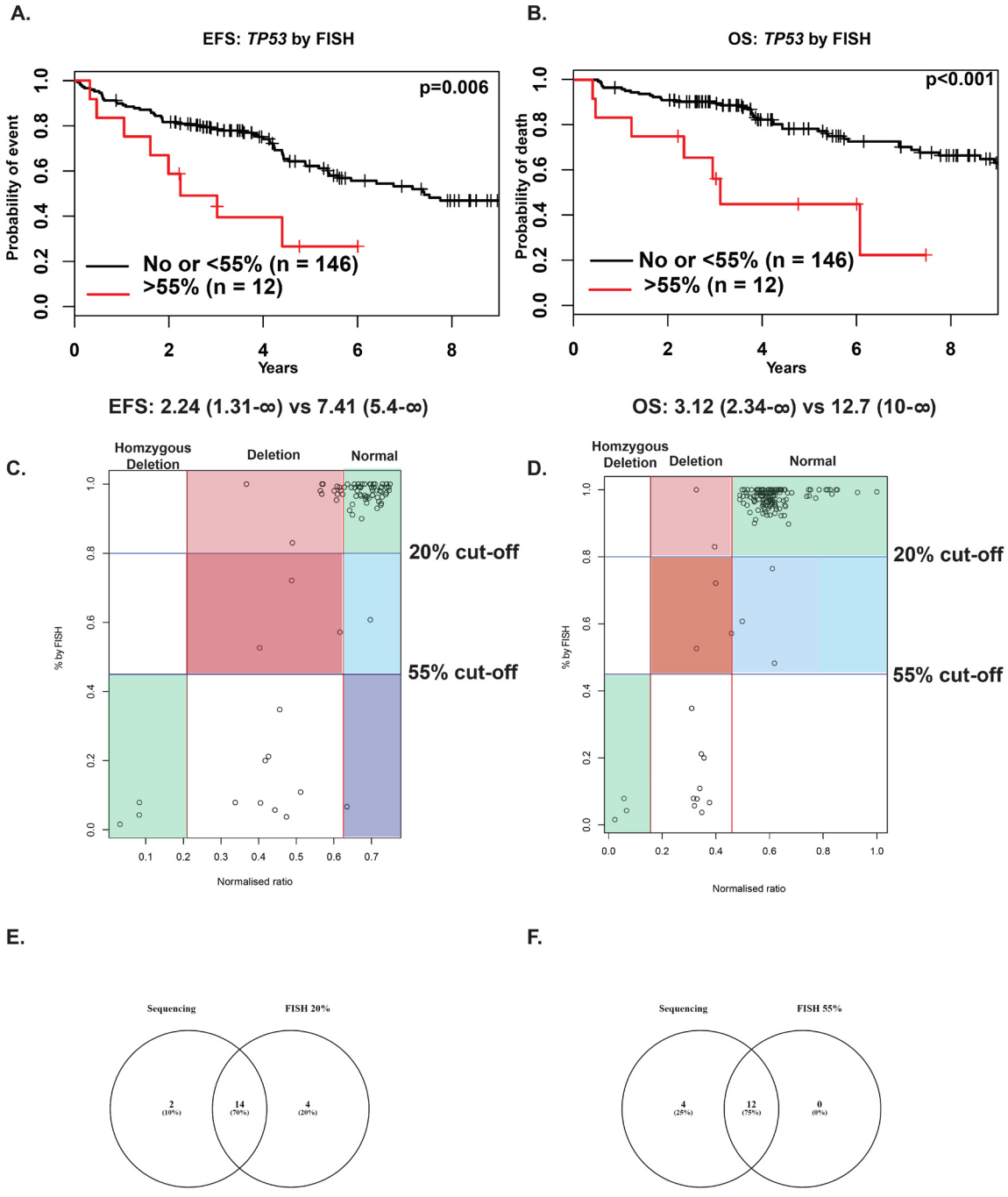


Figure 2.3.2.4.3. Validation of Copy number changes (*TP53*). Using a 55% cut-off for TP53 loss by FISH is prognostic on EFS (Panel A) and OS (Panel B). By using the non-normalized ratio (Panel C) there are many false positives (pink: 20% cut-off, red: 55% cut-off) and false negative (light blue 20% cut-off and navy blue 55% cut-off). By choosing the chromosome with the least variance as reference, normalized ratio are corrected and FP/FN rates decrease (Panel D). Overlap between diagnostic methods and cut-offs are presented on Panel E and F.

2.3.2.5. Metrics

A total of 400 samples were sequenced. Cross sample contamination was assessed visually and sample displaying an unfavourable SNP density profile excluded. To develop a more objective user-independent approach to quality control, we performed a nNMF to quantify the Mismatch Repair Signature (MMR). As MM do not have an MMR signature (**Annex 3**) we performed an nNMF on all samples and patients that displayed evidence of an MMR signature were excluded, including seven SMM samples and eleven NDMM for cross sample contamination, **Figure 2.3.2.5.1**.

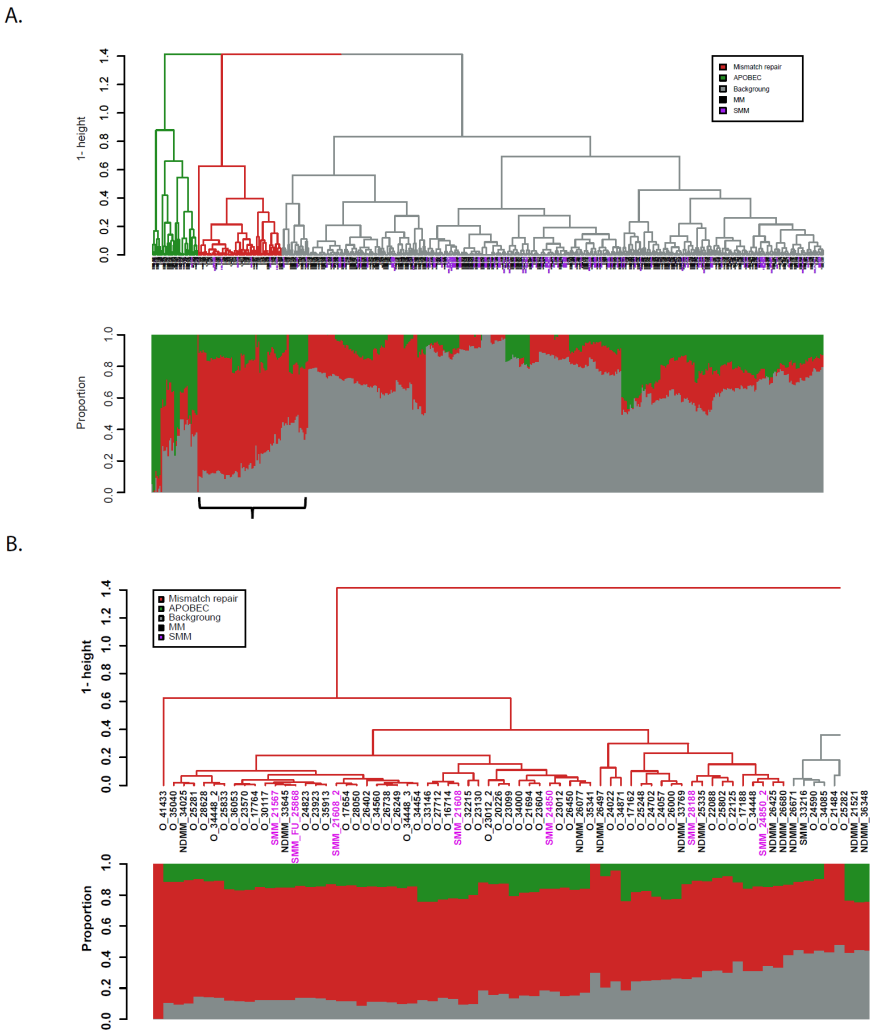


Figure 2.3.2.5.1. nNMF analysis identified a significant mismatched repair signature samples in 68 samples that cluster together A. Overall nNMF B. Zoom on the MMR cluster

None of the samples were excluded for depth of coverage issues. The summary metrics of the target panel are presented in **Table 2.3.2.5.1.**

	Translocation panel	Mutation panel
MM (n=223)	135 (44-197)	452 (200-729)
SMM (n=82)	363 (332-394)	786 (692-867)
EM (n=10)	357 (321-406)	735 (731-891)
MGUS (n=17)	413 (340-433)	974 (834-1036)

Table 2.3.2.5.1. Summary of sequencing metrics

2.3.3. Whole exome sequencing

2.3.3.1. SMM timeline samples

79 samples (9 controls and 70 samples) underwent custom-enriched exome sequencing. 200 ng of DNA was used to prepare libraries using the HyperPlus kit (Kapa Biosystems) and hybridized (SeqCap EZ MedExome; Roche Nimblegen). Eight samples per run were sequenced using a HiSeq 2000 (Illumina, San Diego, CA, USA). Cross sample contamination was assessed by SNP mismatch analysis. Fifteen samples were excluded for purity concerns and two for low coverage (<50X), **Table 2.3.3.1.1**. The median coverage of the 62 remaining samples was 93 (IQR 68-128) and 100 (IQR 95-103) for tumours and controls, respectively. All samples had a depth >50X.

bcl2fastq was used for demultiplexing and BWA mem (v. 0.7.12) for alignment to Ensembl hg38 (exomes) human reference genome. Strelka (v.1.0.14) was used for variant calling and single nucleotide variants (SNVs) were filtered using fpfilter (https://github.com/ckandoth/variant-filter). Indels were filtered using a 10% variant allele frequency (VAF) cut-off. Variants were annotated using Variant Effect Predictor (v.85). Somatic copy-number aberration detection and tumour purity and ploidy estimation were performed using Sequenza v3.0.0 (Favero *et al*, 2015). Intra- and inter-chromosomal rearrangements were called using Manta (v0.29.6) with default settings and the exome flag specified.

Name	Median depth	Purity by flow	PYCLONE
A2	99	74	Yes
A3	88	70	Yes
A4	89	90.4	Yes
A5	86	83.4	Yes
A6	103	91.26	Yes
B2	85	94	Yes
B3	110	98	Yes
B6	116	67	Yes
B7	103	87.1	Yes
B8	106	64	Yes
C1	90	97.9	Yes
C2	93	94.7	Yes
C4	105	89	Yes
C5	89	85	Yes
C6	92	80	Yes
C7	90	60	Yes
C8	100	80.2	Yes
D3	126	86	Yes
D4	154	89	Yes
D5	128	96	Yes
D6	100	100	Yes
D7	98	97	Yes
E10	80	99.6	Yes
E2	151	85.3	Yes
E3	98	84	Yes
E4	108	91	Yes
E5	123	87.9	Yes
E6	115	94.3	Yes
E7	103	87	Yes
E8	91	87.6	Yes
E9	90	98.7	Yes
F2	112	85.9	Yes
F3	90	87.4	Yes
F4	77	90	Yes
F5	96	89	Yes
F6	91	68	Yes
F7	104	100	Yes
F8	91	94.2	Yes
G1	81	91	Yes
G2	99	92	Yes
H1	99	60	No
H10	68	98	No
H2	70	80	No
H5	81	90	No
H6	81	97.5	No
H8	76	99.1	No
H9	68	98.5	No
I1	78	96	Yes
I4	106	87.6	Yes
I5	85	92	No
I6	89	95	No
I8	108	99.5	Yes
I9	84	99.4	Yes

Figure 2.3.3.1.1. Summary of sequencing metrics

2.3.3.2. CoMMpass MMRF WES germline samples

The WES paired-end read data from 980 participants were assessed for quality using FastQC (v0.11.7). Adapter sequences and low-quality bases detected within the reads were removed using Trimmomatic (v0.36). All surviving reads were then mapped to the GRCh38 build of the human reference genome using BWA MEM algorithm (v0.7.17). Sambamba (v0.6.8) was used to convert the resulting SAM files to BAM format, remove low-mapping-quality reads (less than 10), sort the BAM files, and remove duplicate sequences from the BAM files. The BAM files were then realigned and recalibrated then mean coverage was calculated using GATK (v3.8.1). Single nucleotide variants (SNVs) and insertions/deletions (indels) were called using LoFreq (v2.1.3.1) to give best chance of detecting the lowest frequency somatic mutations. Variant calls were targeted to all canonical exons plus a 10 bp padding using BEDTools (v2.27.1). The resulting VCF file was modified for downstream annotation using SAMtools (v1.9), BCFtools (v1.9), and custom scripts. Functional annotation of the called variants was done using ANNOVAR (v2017Jul16).

The downstream filtering process of all called variants was performed using custom scripts written in R (v3.6.1). Before filtering the VCF files by variant, the samples were evaluated based on distribution of coverage to determine outliers. Four samples were removed as a result (mean coverage less than 20X). The variants were broadly filtered to leave only hits in “exonic”, “splicing”, or “exonic;splicing” regions as annotated by the RefGene database. Then the variants were filtered leave only those that resulted in an amino acid sequence change: frameshift insertions, frameshift deletions, stop gains, stop losses, and nonsynonymous SNVs. Next, any variant with a QUAL score less than 90 was removed. Due to the ultra-sensitive variant calling algorithm used by LoFreq and coverage of the samples, we expected the number of false-positive variants with a variant allele frequency (VAF) below 1% to be higher than acceptable. Therefore, the lower limit of variant detection was set at 1% VAF. Additionally, we expect that the truly rare somatic variants would have a VAF below 50%, thus this was set as the upper limit of variant detection.

2. Methods

In an effort to capture variants with very low allele frequencies but maintain a consistent minimal false positive rate among all samples, the variants were filtered based on their loci. If the variant fell within the region set by filters 1-4, it was filtered out. Conversely, if they variant fell within the region set by filter 5, it was kept.

1. Low complexity region (LCR) filter: regions that have shown a significantly higher false positive rate for called indels and SNVs due to either misalignments (https://github.com/lh3/varcmp/blob/master/scripts/LCR-hs38.bed.gz)
2. Hardy Weinberg (HW) filter: regions with sites that have failed the Hardy Weinberg equilibrium test in phase 1 of the 1000 Genomes Project (https://github.com/lh3/varcmp/blob/master/scripts/1000g.hwe-bad.bed) (Graffelman *et al*, 2017)
3. Excess coverage (EC) filter: regions that were shown to have excess read coverage in phase 1 of the 1000 Genomes Project. We converted the coordinates in the files from the GRCh37 build to GRCh38 using Ensembl Assembly Converter (v99).
4. Segmental Duplication (SD) filter – regions of the human genome where there are multiple matches to other parts of the genome (http://humanparalogy.gs.washington.edu/build38/)
5. Strict Mask (SM) filter: regions that require total coverage be within 50% of the average, no more than 0.1% of reads have a mapping quality of zero, and that the average mapping quality for the position be 56 or greater (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/accessible_genome_masks/20120824_strict_mask.bed)
6. After completing the filtering process, the remaining variants were searched for suspected candidate driver mutations. Considering that all participants of this series have multiple myeloma, we expected to see a strong signature of mutations in genes reported to be associated with myeloid development as well as hematologic and lymphoid malignancies. Thus, we define a variant as a candidate driver mutation if it has a VAF below 30% and satisfies one of the following criteria:
    1. Nonsynonymous SNV, stop gain, stop loss, frameshift insertion, or frameshift deletion mutations in any exon between 7 to 23 of *DNMT3A*

2. Stop gain, stop loss, frameshift insertion, or frameshift deletion mutations in *ASXL1*
3. Stop gain, stop loss, frameshift insertion, or frameshift deletion mutations in *TET2*
4. Stop gain, stop loss, frameshift insertion, or frameshift deletion mutations in *PPM1D*
5. Nonsynonymous SNV mutation p.V617F in *JAK2*
6. Any other mutation that has been reported to occur more than 6 times in hematopoietic and lymphoid tissue within the Catalogue of Somatic Mutations in Cancer (COSMIC) database17
7. Finally, all candidate driver mutations which were identified based on occurrence within COSMIC were manually checked against the most current release (v91) to determine if the mutation has been flagged as a single nucleotide polymorphism or if the amino acid change of the mutation matches what is reported in the database.

2.3.4. Ultra-low pass whole genome

Eight-one SMM and 123 MM samples underwent ultra-low-pass WGS. Libraries for tumour DNA and control DNA were prepared as described above using the HyperPlus kit (Kapa Biosystems). Before hybridization to the panels, library DNA was removed and sequenced directly using paired tumour and control libraries which were pooled for sequencing on the on NextSeq500 using 75-bp single end reads. Sequence reads were aligned to Ensembl GRCh37/hg19 and copy number was determined using Control-FREEC (v 3.0.0). Samples with low coverage were excluded (SMM=13 and MM=7). The average coverage was 0.29 (IQR 0.16-0.51) and 0.21 (IQR 0.18-0.3) for SMM and MM, respectively.

2.3.5. MMRF Compass Whole genome sequencing

For WGS,100ng-1000 ng of genomic DNA was fragmented to an average size of that is fragmented to a target size of 900bp. Long-insert whole genome libraries were constructed using the Kapa Hyper Prep Kit (Kapa Biosystems). Data were aligned to hg38 using BWA (v. 0.7.17)(Li & Durbin, 2009) and deduplicated using samblaster (v.

0.1.24)(Faust & Hall, 2014). Copy number was called with Control-FREEC (v. 11.4)(Boeva *et al*, 2012) and structural events with Manta (v. 1.4.0)(Ben-Bassat & Chor, 2014). After primary analysis was complete, data was aligned to hg19 to identify cases with complex chained rearrangements using ChainFinder (v. 1.0.1)(Baca *et al*, 2013) and complex clustered events using ShatterSeek(Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing | bioRxiv).Telomere length was determined using Telomerecat (Farmery *et al*, 2018).

2.4. Expression analysis

2.4.1. Gene expression profiling

Total RNA from plasma cells was used for gene expression profiling (GEP) using U133 Plus 2.0 microarrays (Affymetrix). CEL files were normalized using GCRMA (Wu J & Gentry J, 2020) for application of updated TC algorithm. MAS5 normalization was also performed when necessary, e.g. for calculation of GEP4 and NF-κB scores. All expression data was normalized using R Bioconductor and transformed to the UAMS TT2 and TT3 NDMM standard according to a variant of M-ComBat (Stein *et al*, 2014). All samples were processed as part of the routine clinical management in the accessioning laboratory and analysed using a pipeline developed by Clyde Bailey and Phil Farmer.

Gene Set Enrichment analysis was performed after rma normalisation using the “affy” (affy package | R Documentation) and “fgea” packages (Sergushichev, 2016).

2.4.2. RNA-seq

For RNA-sequencing, either 150ng or 500ng of total RNA was used to enrich for poly-adenylated RNA molecules, which were subsequently fragmented to a target size of 180bp by heat fragmentation. Fragmented molecules were then converted to cDNA using random primers with Superscript II (Invitrogen). After second strand synthesis, the resulting molecules were used for library prep using the Illumina TruSeqRNA library kit. These samples were processed at Tgen by Johnathan Keats.

A subset of 643 samples from the CoMMpass study had RNA sequencing available (Manojlovic *et al*, 2017). Both datasets where aligned to hg38 using STAR (2.5.1b)(Dobin *et al*, 2013) and quality controlled using QoRTS (v1.2.42) (Hartley & Mullikin, 2015) with

alignment and quantification of gene read count with Salmon (v0.7.2) (Patro *et al*, 2017). Normalization of counts and differential gene expression analysis was performed using DESeq2 (v1.14.1) (Torres *et al*, 2019). This step was performed by Michael Bauer.

2.5. Droplet Digital PCR (ddPCR)

Primers were used to amplify the *IGHG3-MYC* translocation and the non-translocated *IGH* locus. Primers and probes may be found in **Table 2.5.1.** 25 µL reaction mixtures were prepared containing primers (40X, Thermo Fisher Scientific), 10 ng template and ddPCR™ Supermix (2X,Bio-Rad). Droplet generation and transfer of emulsified samples to PCR plates was performed according to manufacturer's instructions (Instruction Manual, QX200™ Droplet Generator – Bio-Rad). The cycling protocol started with 95 °C enzyme activation step for 10 minutes followed by 40 cycles of a two-step cycling protocol (94 °C for 30 seconds and 60 °C for 1 minute). The ramp rate between these steps was slowed to 2 °C/second. The sample were read using a QX200 droplet reader (Bio-Rad). The absolute number of positive droplets was calculated using QuantaSoft (v.1.7.4).

	IGHG3-MYC (5'-3')	**non-translocated *IGH* locus (5'-3')**
Forward primer	CAGTATTTTAGTAGCTCAAAGACACCTCTT	AGCTGCCACCTGCTTGT
Reverse primer	GCTTAGGTCAGTTTTGCCCATCT	CTGGGCTGGGCTGAGTT
Probes	FAM-TCCATTTCTGAAGACTTA-MGBNFQ VIC-AGTCCATTTCTGATGACTTA-MGBNFQ	FAM- TCCATTTCTGAAGACTTA-MGBNFQ VIC- AGTCCATTTCTGATGACTTA-MGBNFQ

Table 2.5.1. ddPCR primers

2.6. Statistical analysis

2.6.1. Time to event analysis

Time-to-event analysis was performed in R.

2.3.2.1. MM

All genetic events with n>15 for MM were analysed. The Kaplan–Meier estimator was used to calculate time-to-event distributions. Stepwise Cox regression (R: The R Stats Package) in both directions, based on Akaike information criterion (AIC), using variables with p<0.1 on univariate, estimated the effects of significant covariates for time-to-event outcomes. When multiple features relating to the same event were present, such as del(1p) (*FAF1*) and del(1p) (*CDKN2C*), the one explaining the greatest variance was selected for the analysis. All variables with the exceptions of chromosome X copy number changes were included. The final Cox model consisted only of statistically significant factors at a level of p<0.05. An additional bootstrap was performed using the rms package(Jr, 2018) (B=100) and corrected indices (Dxy and r^2) computed. As bivariable selection methods can induce biaises we repeated the analysis using the well-defined previously published consensus risk factors (ISS, t(4;14), t(14;20), del(1p), gain(1q)) and mutations. Furthermore, we attempted to show the impact of mutations on other risk models such as the IFM copy number model and GEP70.

2.3.2.2. SMM

Time-to-event analysis was performed in R with all genetic events with n>7. The Kaplan–Meier estimator was used to calculate time-to-event distributions. Stepwise Cox regression (R: The R Stats Package) using previously published risk factors (GEP4, and IMWG 2018 criteria (Lakshman *et al*, 2018)) and potential novel factors (del(6q), del(13q), and *KRAS* mutation) was performed.

2.3.2.3. MMRF data

Time-to-event was performed in R. The Kaplan–Meier estimator was used to calculate time-to-event distributions. Stepwise Cox regression (R: The R Stats Package) in both directions, based on Akaike information criterion (AIC), estimated the effects of significant covariates for time-to-event outcomes. We included Age (as a continuous variable), ISS, ECOG performance status, and a composite genetic factor defined by the

presence of either a t(4;14), t(14;16), del(1p), gain(1q), del(17p)) The final Cox model consisted only of statistically significant factors at a level of $p<0.05$. An additional bootstrap was performed using the rms package(Jr, 2018) (B=100) and corrected indices (Dxy and r^2) computed.

2.6.2. Proportional testing

Kruskal-Wallis or Fisher's exact tests were used to compare the median of a continuous variable or the distribution of discrete variables across groups, when appropriate.

2.6.3. Correlation analysis

Correlation between mutated genes and cytogenetic abnormalities was performed using the R package "stats". The covariance was computed using the Pearson method. The test statistic is based on Pearson's product moment correlation coefficient and follows a t distribution with length(x)-2 degrees of freedom assuming independent normal distributions. Correction for multiple testing was performed using the Bonferroni method. The covariance matrix was plotted using corrplot (Wei *et al*, 2017). Alternatively, correlations between mutated genes and cytogenetic abnormalities using Bayesian inference was determined using the program JAGS and the *R*-interface Bayesmed. The probability of the observed data under the null hypothesis versus the alternative hypothesis or Bayes factor (BF) was computed. A BF greater than 1 was considered significant. BFs of 1 to 3, 3 to 20, 20 to 150, and greater than 150 were considered weak, positive, strong, and very strong associations, respectively. Correlation coefficients were plotted using corrplot. The covariance matrix was plotted using corrplot (Wei *et al*, 2017).

2.6.4. Signature analysis

2.6.4.1. Nonnegative matrix factorization

Mutational signatures were called using non-negative matrix factorization (NMF) with counts per sample calculated for the six possible SNV types and the 16 possible 3-base sequence contexts, creating a table with 96 columns. The R package "NMF" was used for all calculations (CRAN - Package NMF). The number of signatures was determined by running 50 iterations of the algorithm for 2-7 signatures. A number of signatures was chosen that maximized the cophenetic distance and dispersion values.

One thousand (1,000) iterations of the algorithm were run for that number of signatures. Cosine similarity was used to determine the Sanger signatures that were closest to the detected signatures.

2.6.4.2. Fitting Signature analysis

Fitting Signature analysis was performed using the fitting algorithm mmSig (Maura *et al*, 2019), which fits the entire mutational catalogue of each patient with the mutational signatures involved in MM pathogenesis was used to determine the signature admixture in each individual sample and among samples more than two years away from progression, samples that were within two years of progression, and samples of patients that had not progressed. The contribution of each mutational signature was then corrected based on the cosine similarity between the original 96-mutational profile and the reconstructed profile generated without that signature.

2.6.5. Clonal architecture analysis

We attempted to reconstruct the clonal population structure of WES samples that had adequate CNA (n=44) using Pyclone (Roth *et al*, 2014) using default settings. PyClone is a Bayesian clustering method for grouping sets of deeply sequenced somatic mutations into putative clonal clusters while estimating their cellular prevalences and accounting for allelic imbalances introduced by segmental copy-number changes and normal-cell contamination. Data were visualized using the Fishplot package (Miller *et al*, 2016).

2.6.6. Predicting the *BRAF* function

The predicted functions of *BRAF* mutations were determined using the Clinical Knowledgebase (CKB) database (JAX CKB) using the mutations present in the MGP dataset (n=103)(Walker *et al*, 2018) and this dataset (n=26).

2.6.7. Diversity analysis

The Shannon diversity index (H) is an index that is commonly used to characterize species diversity in a community. Shannon's index accounts for both abundance and evenness of the species present. The proportion of species i relative to the total number of species (p_i) is calculated, and then multiplied by the natural logarithm of this proportion ($\ln(p_i)$). The resulting product is summed across species, and multiplied by -1.

$$H = -\sum_{i=1}^{R} p_i \ln(p_i)$$

2.7. Data Availability

Sequencing data have been deposited on the European Genome-phenome Archive (EGA) under the following assertions.

- Amyloidosis project: EGAS00001003164
- MM project: EGAD00001004373
- SMM project: EGAD00001005285, and EGAD00001005056

2.8. References

affy package | R Documentation Available at: https://www.rdocumentation.org/packages/affy/versions/1.50.0 [Accessed September 28, 2020].

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., Van Allen, E., Kryukov, G.V., Sboner, A., Theurillat, J.-P., Soong, T.D., Nickerson, E., Auclair, D., Tewari, A., Beltran, H., Onofrio, R.C., et al (2013) Punctuated evolution of prostate cancer genomes. Cell, 153, 666–677.

Ben-Bassat, I. & Chor, B. (2014) String graph construction using incremental hashing. Bioinformatics, 30, 3515–3523.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O. & Barillot, E. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics, 28, 423–425.

Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y. & Hwang, C.-C. (2013) Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. PLOS ONE, 8, e62856.

Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing | bioRxiv Available at: https://www.biorxiv.org/content/10.1101/333617v1 [Accessed January 8, 2020].

CRAN - Package NMF Available at: https://cran.r-project.org/web/packages/NMF/index.html [Accessed January 18, 2019].

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15–21.

Farmery, J.H.R., Smith, M.L. & Lynch, A.G. (2018) Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. Scientific Reports, 8, 1300.

Faust, G.G. & Hall, I.M. (2014) SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics, 30, 2503–2505.

Favero, F., Joshi, T., Marquard, A.M., Birkbak, N.J., Krzystanek, M., Li, Q., Szallasi, Z. & Eklund, A.C. (2015) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Annals of Oncology, 26, 64–70 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4269342/ [Accessed April 16, 2020].

Giraud, M., Salson, M., Duez, M., Villenet, C., Quief, S., Caillault, A., Grardel, N., Roumier, C., Preudhomme, C. & Figeac, M. (2014) Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. BMC genomics, 15, 409.

Graffelman, J., Jain, D. & Weir, B. (2017) A genome-wide study of Hardy–Weinberg equilibrium with next generation sequence data. Human Genetics, 136, 727–741.

Hartley, S.W. & Mullikin, J.C. (2015) QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics, 16, 224.

JAX CKB Available at: https://ckb.jax.org/ [Accessed January 18, 2019].

Jr, F.E.H. (2018)rms: Regression Modeling Strategies Available at: https://CRAN.R-project.org/package=rms [Accessed January 25, 2019].

Lakshman, A., Rajkumar, S.V., Buadi, F.K., Binder, M., Gertz, M.A., Lacy, M.Q., Dispenzieri, A., Dingli, D., Fonder, A.L., Hayman, S.R., Hobbs, M.A., Gonsalves, W.I., Hwa, Y.L., Kapoor, P., Leung, N.,

Go, R.S., Lin, Y., Kourelis, T.V., Warsame, R., Lust, J.A., et al (2018) Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. Blood Cancer Journal, 8, 59.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25, 1754–1760.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio] Available at: http://arxiv.org/abs/1303.3997 [Accessed September 29, 2020].

Manojlovic, Z., Christofferson, A., Liang, W.S., Aldrich, J., Washington, M., Wong, S., Rohrer, D., Jewell, S., Kittles, R.A., Derome, M., Auclair, D., Craig, D.W., Keats, J. & Carpten, J.D. (2017) Comprehensive molecular profiling of 718 Multiple Myelomas reveals significant differences in mutation frequencies between African and European descent cases. PLoS Genetics, 13, Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5699827/ [Accessed January 8, 2020].

Maura, F., Bolli, N., Angelopoulos, N., Dawson, K.J., Leongamornlert, D., Martincorena, I., Mitchell, T.J., Fullam, A., Gonzalez, S., Szalat, R., Abascal, F., Rodriguez-Martin, B., Samur, M.K., Glodzik, D., Roncador, M., Fulciniti, M., Tai, Y.T., Minvielle, S., Magrangeas, F., Moreau, P., et al (2019) Genomic landscape and chronological reconstruction of driver events in multiple myeloma. Nature Communications, 10, 1–12 Available at: https://www.nature.com/articles/s41467-019-11680-1 [Accessed March 27, 2020].

Miller, C.A., McMichael, J., Dang, H.X., Maher, C.A., Ding, L., Ley, T.J., Mardis, E.R. & Wilson, R.K. (2016) Visualizing tumor evolution with the fishplot package for R. BMC Genomics, 17, 880 Available at: https://doi.org/10.1186/s12864-016-3195-z [Accessed April 16, 2020].

Mutect2 GATK Available at: http://gatk.broadinstitute.org/hc/en-us/articles/360037593851 [Accessed September 29, 2020].

Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., Sakkiah, S., Guo, W., Gong, P., Zhang, C., Ge, W., Shi, L., Tong, W. & Hong, H. (2019) Similarities and differences between variants called with human reference genome HG19 or HG38. BMC Bioinformatics, 20, 101.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature methods, 14, 417–419.

Picard Tools - By Broad Institute Available at: https://broadinstitute.github.io/picard/ [Accessed September 29, 2020].

R: The R Stats Package Available at: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html [Accessed January 25, 2019].

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A. & Shah, S.P. (2014) PyClone: statistical inference of clonal population structure in cancer. Nature Methods, 11, 396–398 Available at: https://www.nature.com/articles/nmeth.2883 [Accessed March 27, 2020].

Sergushichev, A.A. (2016) An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv, 060012.

Stein, C.K., Qu, P., Epstein, J., Rosenthal, A., Hunter-Merrill, R., Williams, R., Crowley, J. & Barlogie, B. (2014) Modified Combat Removes Batch Effects from Myeloma Cell GEP–derived Risk Scores and Molecular Subgroup Assignment. Blood, 124, 3355–3355.

Thakurta, A., Ortiz, M., Blecua, P., Towfic, F., Corre, J., Serbina, N.V., Flynt, E., Yu, Z., Yang, Z., Palumbo, A., Dimopoulos, M.A., Guttierez, N., Goldschmidt, H., Sonneveld, P. & Avet-Loiseau, H. (2019) High sub-clonal fraction of 17p deletion is associated with poor prognosis in Multiple Myeloma. Blood.

Torres, A.G., Reina, O., Attolini, C.S.-O. & Pouplana, L.R. de (2019) Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. Proceedings of the National Academy of Sciences, 116, 8451–8456.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018) Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. Blood, 132, 587–597.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. & Zemla, J. (2017)corrplot: Visualization of a Correlation Matrix Available at: https://CRAN.R-project.org/package=corrplot [Accessed January 28, 2019].

Wu J & Gentry J (2020) gcrma: Background Adjustment Using Sequence Information. R package version 2.60.0.

Chapter 3: The genomic landscape of plasma cells in systemic light chain amyloidosis

3.1. Summary

Systemic light chain amyloidosis (AL) is characterized by the deposition of immunoglobulin light chains as amyloid fibrils in different organs, where they form toxic protein aggregates. The underlying disease is likely a plasma-cell disorder, but limited sequencing data are available. We performed an exome sequencing study in AL and compared this to monoclonal gammopathy of undefined significance (MGUS) and myeloma (MM). Twenty-four samples from unselected newly diagnosed AL patients were analysed. Thirty percent of patients yielded a mutation in a MM driver gene. There was evidence of mitogen activated protein kinase (MAPK) activation with *NRAS* mutations, nuclear factor-kappa B (NF-κB) activation and DNA repair pathway alterations. Exome data were used to determine the cytogenetics of AL samples and identified hyperdiploidy in 30% of cases and t(11;14) in 30% of cases. A novel translocation, was identified involving *RCC1*. Finally, further analysis of the *APCS* gene encoding the SAP protein did not show any biased genotype in AL patients in comparison to the general population. These data therefore identified a genetic make-up of AL, not only in terms of copy number abnormalities and translocations but also nonsynonymous mutations, that is similar to other plasma cell disorders such as MM and MGUS.

3.2. Introduction

Systemic light chain amyloidosis (AL) is a rare disease characterized by the presence of extracellular depositions of immunoglobulin light chains as amyloid fibrils in different organs, where they form protein aggregates leading to a vast array of symptoms such as congestive heart failure, kidney failure and neuropathy (Merlini & Bellotti, 2003). The key event in the pathogenesis of AL is an unstable secondary or

tertiary structure of a monoclonal immunoglobulin light chain, that assemble into monomers, stack together, form fibrils, and precipitate in the extracellular compartments with the serum amyloid P component (SAP) protein (Merlini & Bellotti, 2003).

The underlying disease is a plasma cell disorder, likely a monoclonal gammopathy of undetermined significance (MGUS), smouldering myeloma (SMM) or symptomatic myeloma (MM). Cases have been described with other lymphoproliferative disorders (LPD) but are less common (Sanchorawala *et al*, 2006). The entire spectrum of plasma cell disorders, ranging from MGUS to MM, arise from an initial event such as a translocation into the *IGH* locus or hyperdiploidy, and progress throughout the various disease stages following the successive acquisition of mutations in key pathways such as the nuclear factor-kappa B (NF-κB) or mitogen activated protein kinase (MAPK) as well as additional copy number changes. As our understanding of the complex determinants of disease progression from MGUS and MM has improved,(Morgan *et al*, 2012) data regarding AL and where it lies in the spectrum remain to be determined.

Although limited data are available on the biology of the plasma cell clone underlying AL, existing studies have associated AL, like light chain only myeloma, with t(11;14) that make-up 47% of cases (Bochtler *et al*, 2015, 2016, 2008). The specific organ-tropism has to some extent been attributed to specific light chain rearrangements (Merlini & Bellotti, 2003; Abraham *et al*, 2007; Comenzo *et al*, 2001; Perfetti *et al*, 2012). To date five cases have been sequenced (Paiva *et al*, 2016) and did not identify any unifying mutation.

As little is currently known about the genetic make-up of AL in terms of mutations and other translocations, we performed whole exome sequencing on 24 AL patient'sCD138+ marrow cells and compared these results to those obtained in previously published MM, MGUS and other LPD datasets.

3.3. Results

3.3.1. Patient demographics

To assess how representative of a general AL population, we determined the baseline demographic of our group of patients. The median age at diagnosis was 66 (range: 40-82) years old. All cases were histologically proven, newly diagnosed, systemic, light chain amyloidosis (AL). As expected, 75% were lambda light chain restricted and 25% kappa light chain restricted with a respective median involved serum free light chain (sFLC) measured at 180mg/L (range: 59-986 mg/L) and 1580 mg/L (range: 619-3190 mg/L), respectively. All patients had a light chain differential (dFLC) greater than fifty. The median plasmacytosis on bone marrow trephine was 12.5% (range: 2-90%). Among them, 75% had evidence of heart involvement, 71% had renal involvement and 33% had liver involvement, **Table 3.3.1**. Five patients had symptomatic myeloma: two according to standard CRAB criteria (Kyle & Rajkumar, 2009) and three with other myeloma defining events (MDE; SFLC ratio for two and plasmacytosis greater than 60% for one) (Rajkumar *et al*, 2014).

	Value (range)
Median age (range)	66 (40-82)
Gender (M/F)	15/9
Isotype (κ/λ)	6/18
Median SFLC values in κ/λ patients • **Kappa patients** • **Lambda patients**	 1580 mg/L (609-3190 mg/L) 180 mg/L (59-986 mg/L)
Median marrow involvement	12.5% (2-90%)
Organ damage • **Heart including Mayo III** • **Renal** • **Liver**	 • 75% including 56% • 71% • 33%
CRAB criteria • **CRAB criteria** • **IMWG 2014**	 2/24 5/24

Table 3.3.1. Summary of the patient's characteristics

3.3.2. The mutational burden in AL is greater than MGUS and similar to MM.

One of the metrics of disease complexity, that distinguishes MGUS and MM, is tumour mutational burden. To ensure comparability between the datasets, we counted the number of exonic, non-Ig, non-synonymous mutations per sample that had a VAF>5%, that passed quality control filtering. The median number of mutations per sample was 39 (interquartile range (IQR): 5-185) which is more than was seen in MGUS (Mikulasova *et al*, 2017) (median 20 (IQR: 1-41), Fisher's test p=0.002) but not statistically different from MM (Walker *et al*, 2015) (median 35 (IQR: 3-74), Fisher's test, p=0.4)**, Figure 3.3.2.1**.

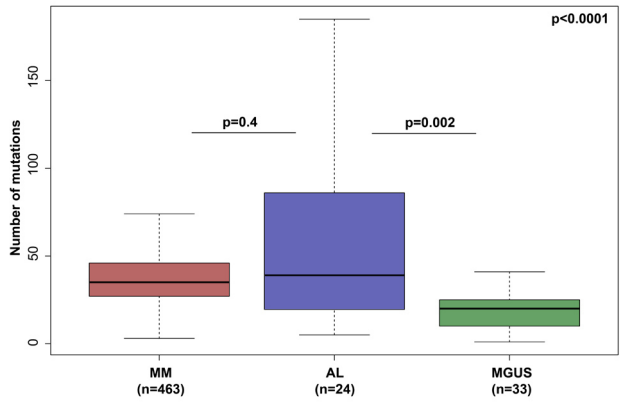


Figure 3.3.2.1. Boxplot suggesting AL has more mutations than MGUS but not a statistically different number than MM

As it were suggested that MM had more mutations than MGUS samples, we assessed whether there was a difference in the number of mutations per sample between AL patients that met MM criteria (n=5) and those who did not (n=19;). The median number of mutations was 63 (IQR: 5-88) median 38 (IQR: 0-176), which was not statistically different (Fisher's test, p=0.5)

3.3.3. There are no unifying mutations in AL

Moving forward to the analysis of mutations, we identified 1491 mutated genes in our dataset with 236 genes mutated more than once. As expected, there was no unifying mutation in AL. The dataset was too small to identify any significantly mutated genes.(Lawrence *et al*, 2013) Of interest to us, *IL7R* mutations were among the most common: four mutations were seen in three patients, **Figure 3.3.3.1 Panel F.** The median tumour fraction for these mutations was 18% suggesting they were secondary events. They did not co-occur with other MAPK mutations. Interestingly, three mutations were seen in *FLT3*, including one involving the tyrosine kinase domain (M837I) in close proximity to the D835 hotspot reported in AML.(Abu-Duhier *et al*, 2001) A summary of the recurrent variants of interest is shown in **Table 3.3.3.1**

3. Amyloidosis

We performed an nNMF to determine whether there were specific signatures present in AL that were not seen in MM or MGUS. Unsurprisingly, there was no specific signatures and all the mutations they were generated via an age-related process (SBS2 and SBS5).

Gene Name	Number of mutations	Gene size	Protein function
MUC16	15	132,499	Mucin barrier, protecting epithelial cells from thogens
LTBP4	6	36,938	Binds transforming growth factor beta GFB) as it is secreted and targeted to the tracellular matrix
RP1L1	5	105,839	Bind microtubules and regulate microtubule lymerization
SYNM	5	37,379	Intermediate filament (IF) family member, toskeletal proteins that confer resistance to echanical stress
ALMS1	4	225,036	Microtubule organization
CASC5	4	70,323	Component of the multiprotein assembly t is required for creation of kinetochore-crotubule attachments and chromosome gregation
CHGB	4	14,034	Tyrosine-sulfated secretory protein undant in peptidergic endocrine cells and urons
IGFN1	4	38,147	Fibronectin Type III Domain Containing
IL7R	4	26,909	The function of this receptor requires the erleukin 2 receptor, gamma chain (IL2RG). is protein has been shown to play a critical e in V(D)J recombination during lymphocyte velopment.
OR11G2	4	10,182	Olfactory receptors interact with odorant lecules in the nose
PKHD1L1	4	170,210	May encode for a large receptor (fibrocystin-
SIGLEC12	4	10,563	Sialic acid-binding immunoglobulin-like tins
SUSD5	4	69,174	Hyaluronic acid binding ECM protein
ZBTB21	4	23,557	Methyl-CpG binding protein.

Table 3.3.3.1. Recurrent variants

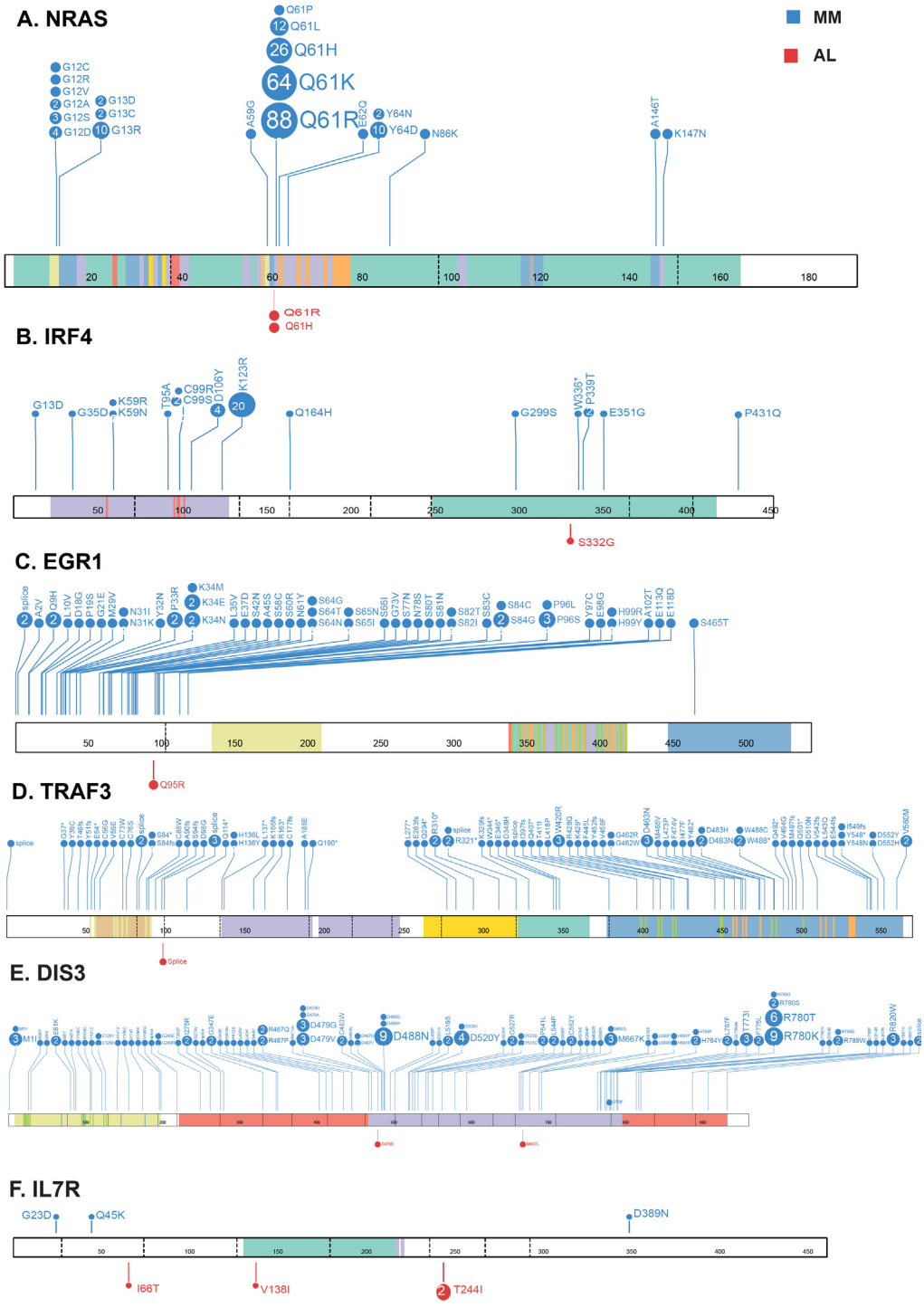


Figure 3.3.3.1. Protein change in genes of interest in AL (red) and MM (blue). A: NRAS, B: IRF4, C: EGR1, D: TRAF3, E: DIS3, F: IL7R

3.3.4. MM driver genes are mutated in AL at lower frequencies and are consistent with MAPK activation, NF-κB activation, and DNA repair pathway alterations

Mutations that provide a selective growth advantage, and thus promote cancer development, are termed driver mutations. Sixty-three driver genes have been previously described MM (Walker *et al*, 2018). Thirty-seven percent of samples (n=9) had a mutation in one of 63 MM driver genes (Walker *et al*, 2018). This is less than MM (84.1%)(Walker *et al*, 2018) and similar to MGUS (36%)(Mikulasova *et al*, 2017). The number of mutated driver genes per sample ranged from zero to five. Among the 63 driver genes, 13 were mutated, **Table 3.3.4.1.** We identified cases with hotspot mutations in *NRAS* (Q61R and Q61H) but not in *KRAS.* These mutations have previously been seen in MM. Interestingly there were mutations in some of the other driver genes such as *EGR1* (Q95R), *DIS3* (D479E, M667L), *IRF4* (S332G), and *TRAF3* (K99_sp). Only the *TRAF3* (K99_sp) and *DIS3* (D479E, M667L) mutations were seen in the MM dataset (Walker *et al*, 2018). In MM, *IRF4* mutations are associated with t(11;14), and in this AL dataset the *IRF4* mutation (S332G) occurred in a non t(11;14) patient and was neither at the K123R hotspot seen in MM (Walker *et al*, 2015) nor the L116R mutation seen in CLL (Havelange *et al*, 2011).

Gene	Number of mutations	Number of patients	Variants
DIS3	2	1	D479E, M667L
DUSP2	1	1	Splice donor variant
EGR1	1	1	Q95R
EP300	1	1	I997V
IRF4	1	1	S332G
KLHL6	1	1	V337M
KMT2B	1	1	P204H
NRAS	2	2	Q61R; Q61H
SAMHD1	1	1	F363L
TET2	1	1	P363L
TRAF3	1	1	Splice donor variant
ZFP36L1	1	1	L82S
ZNF292	1	1	I1740V

Table 3.3.4.1. Driver mutations found in AL

We went on to compare the frequency of MM drivers in MM, MGUS, and AL. Even though a trend for fewer *KRAS* mutations in AL was observed, there was no significant difference in the incidence of driver mutations between MM, MGUS, and AL, **Figure 3.3.4.1**. Most of them were sub-clonal with estimated tumour fractions ranging from 6% to 80%.

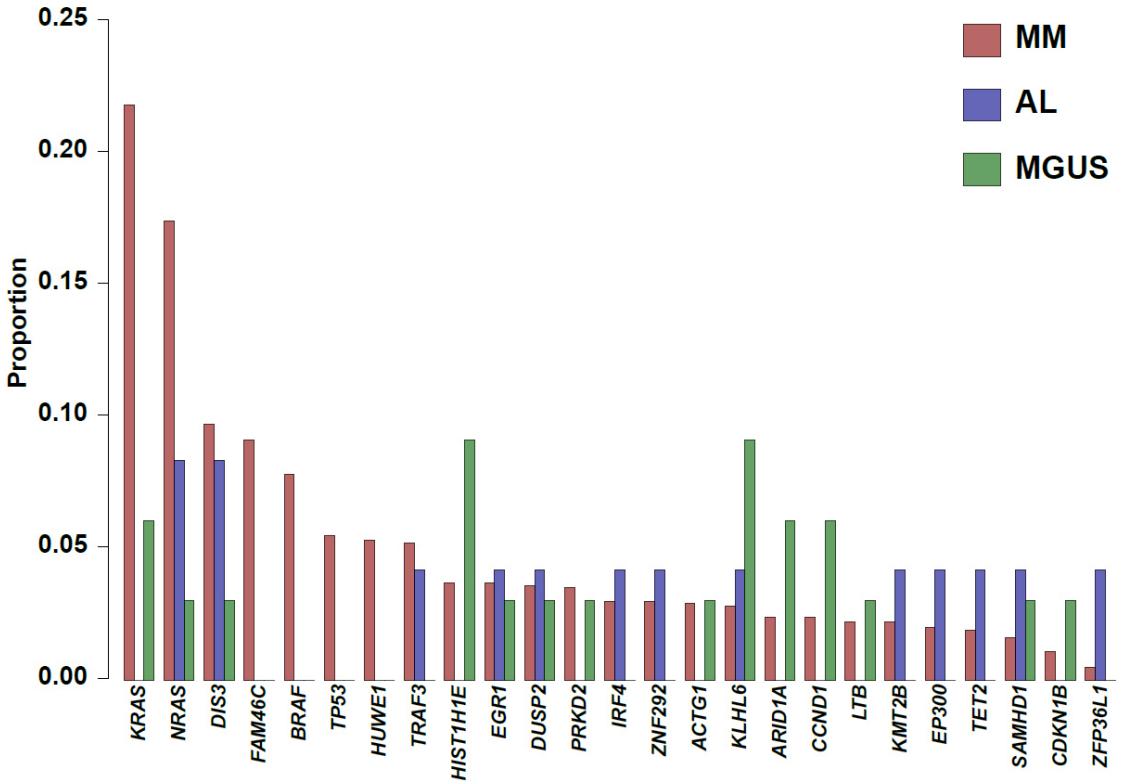


Figure 3.3.4.1. Frequencies of select driver mutations in AL, MGUS and MM.

We looked for driver genes involved in LPDs. Interestingly, one patient had a *CARD11* (R1077W) mutation, which is a landmark mutation of more immature B-cell malignancies, that has been reported in MM with lower frequencies (<1% of cases) (Kim *et al*, 2011).

Finally, we looked at mutations by key myeloma pathways such as DNA repair or NF-κB pathways. There were no mutations genes such as *ATM*, *ATR*, *ZFHX4* and *TP53,* nevertheless mutations in *BRCA2* (P1088T, N372H) and the driver gene *EP300* (I997V) were seen suggesting DNA repair pathway involvement in AL. Copy-number analysis did not reveal any copy number changes at these loci suggesting the absence

of bi-allelic inactivation. Finally, evidence would suggest, the presence of NF-κB pathway activation with not only mutations in the driver genes (such as *TRAF3*, *IRF4*) but also mutations in kinases (such as *LYN* (I165T)), downstream transducing molecules (*CARD11* (R1077V)) and inhibitors (*NFKBIE* (splice))**, Figure 3.3.4.2.**

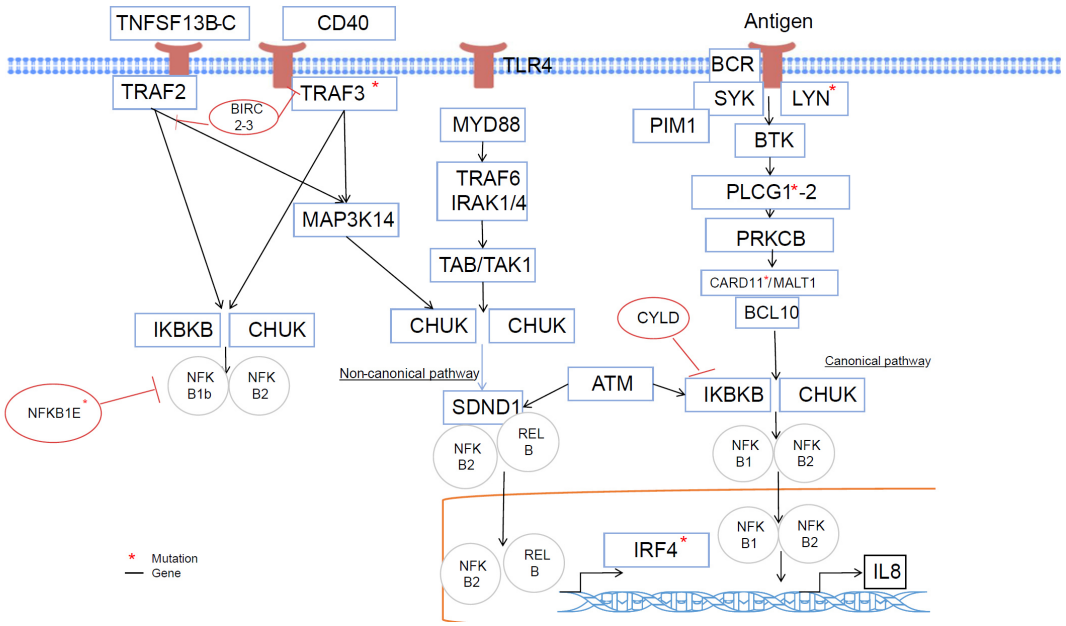


Figure 3.3.4.2. Mutations in the NF-κB pathway: Mutations seen mainly in the canonical pathway and affected kinases such as (*LYN* and *PIM1*), downstream signalling proteins *CARD11*) and inhibitors (*NFKBIE*). As expected from what has been described in MM, mutations were also seen in the non-canonical pathway (*TRAF3*) and in transcription factors (*IRF4*)

3.3.5. The mutational landscape of AL is similar to MM and MGUS suggesting these three disease entities are closely related.

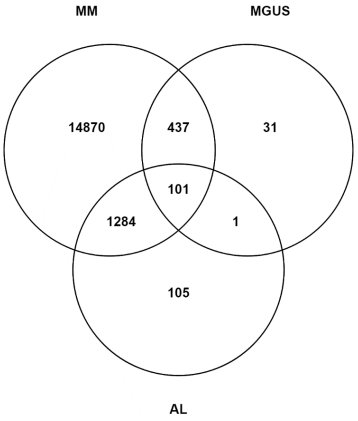


Figure 3.3.5.1. Venn diagram representation of the overlap between MM, MGUS, and AL

We compared the mutational landscape of AL to previously sequenced MM (n=1273)(Walker *et al*, 2018) and MGUS (n=33)(Mikulasova *et al*, 2017) samples. Overall, there were 101 genes in common between AL, MGUS, and MM. Ninety-three percent of the AL-mutated genes (n=1386) were shared with either MM or MGUS. Only 7% of these genes (n=105) had not previously been reported in MM or MGUS, **Figure 3.3.5.1**. None of these mutations was recurrent. A Gene enrichment analysis of the 105 genes that were mutated in AL only was performed using PANTHER(Thomas *et al*, 2003) and did not reveal any specific pathway enrichment, suggesting they are random. There were no mutations in the ribosome subunit genes (Kryukov *et al*, 2016).

Moving forward, as previously studies have reported 68 differentially expressed genes between AL and MM,(Paiva *et al*, 2016) we attempted to identify mutations in these genes. This analysis only showed a mutation in *TIAM2* previously reported in MM (Paiva *et al*, 2016). Of note, we did identify a mutation in the eighth exon of *PSMA2* (P223S) that was among the differentially expressed genes found by Abraham and colleagues (Abraham *et al*, 2005).

To determine which disorder AL resembled most, we performed Ward hierarchical clustering for driver mutation frequencies, based on previously published data,(Walker *et al*, 2015; Mikulasova *et al*, 2017; Lohr *et al*, 2012; Landau *et al*, 2017)

Interestingly, AL clustered with MGUS and was closely related to MM suggesting a common mutational architecture between them, **Figure 3.3.5.2.**

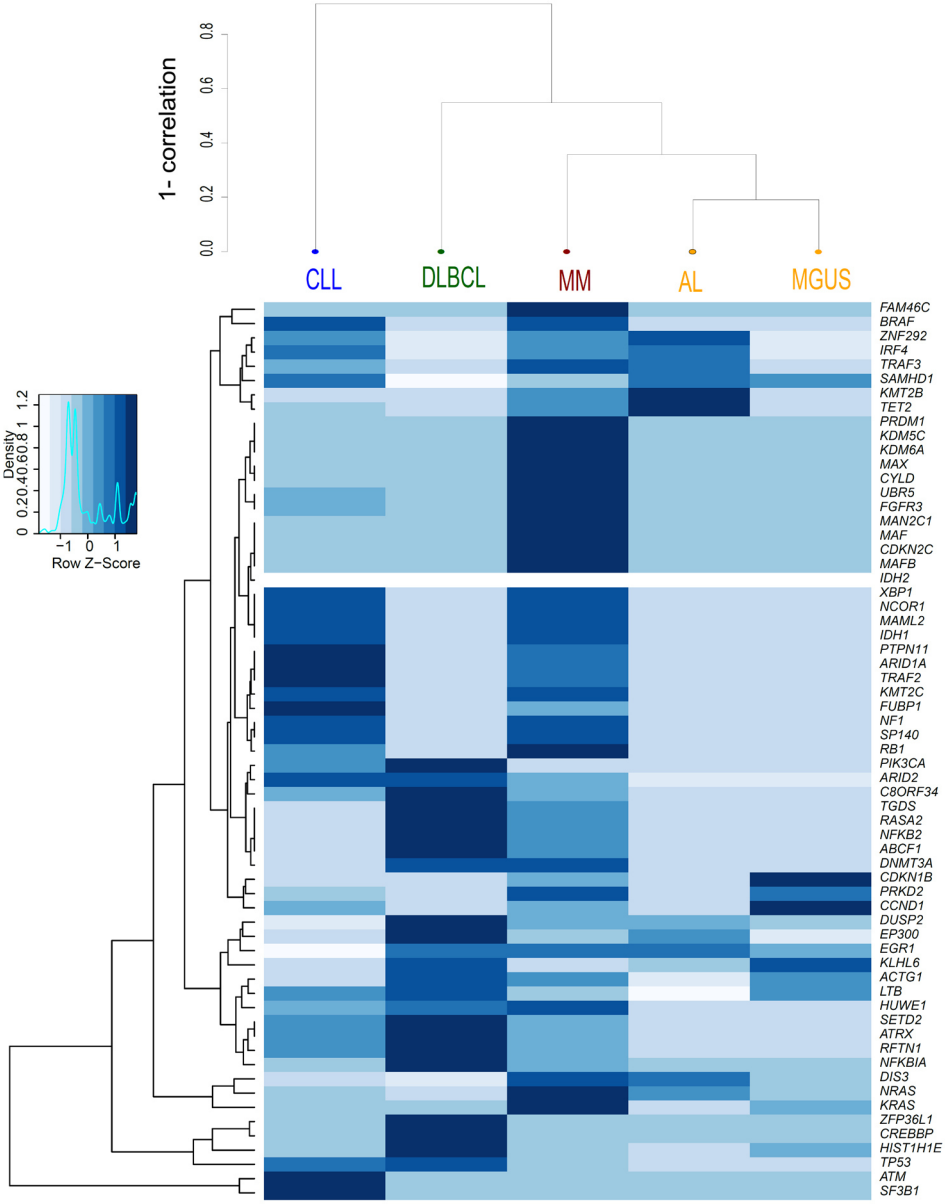


Figure 3.3.5.2. Clustering of driver-mutation frequencies

3.2.6. Molecular karyotyping identifies overlapping translocations and copy number abnormalities with MM

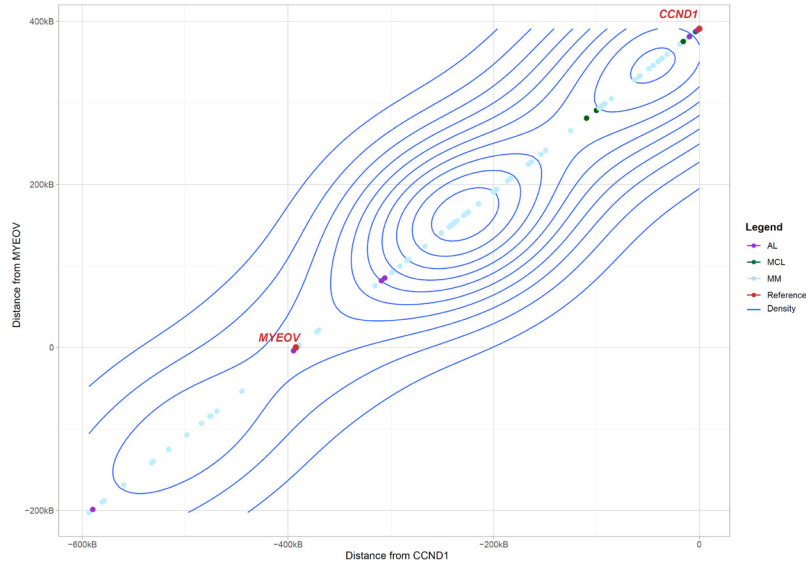


Figure 3.3.6.1. Analysis of the t(11;14) breakpoints suggest they are similar to those observed in myeloma (MM) and Mantle cell lymphoma (MCL).

We hypothesised that the overall architecture of AL was less complex at the chromosom level with fewer translocations and copynumber changes. To confrim this, first we assessed inter-chromosomal rearrangement to the *IGH* or *MYC* regions. We were particularly interested in the t(11;14) translocations that may help distnguish AL from MM or Mantle cell lymphoma (MCL). We identified seven t(11;14) translocations making up for 30% of cases. The breakpoints were located 2-600 kb upstream of *CCND1*. These breakpoints were consistent with those seen in other lymphoid malignancies bearing a t(11;14) such as MM and Mantle cell lymphoma (MCL), **Figure 3.3.6.1**. They were all generated via class switch recombination with breakpoints occurring in the *IGHA1* (2/7) and *IGM* (5/7) switch regions. As expected we did not identify any additional canonical translocations [such as t(4;14), t(14;16), t(14;20)] that occur with lower frequencies in AL. There was no evidence of translocations involving *MYC*: one patient had an 8q24 gain, 5' to *MIR1208*, suggesting *MYC* rearrangements may also occur in AL. One non-canonical translocation included a t(11;14) with a concurrent t(1;14)(q35.3; q32) involving the *RRC1* gene and the *IGHG1* switch region, **Figure 3.3.6.2**..

Regarding other cytogenetic abnormalities, the incidence of del(1p) (n=1, 4%, p=0.008), and del(14q) (n=0, p<0.0001) was lower than expected in MM but similar to MGUS. There was no difference in the incidence of del(17p) (p=0.3), hyperdiploidy (n=7, 30%, p=0.5), gain 1q (n=3, 12.5%, including one case of amplification, p=0.09), t(11;14) (n=7, 30%, p=0.2) or t(4;14) (p=0.09) in comparison to MM. **Table 3.3.6.1**.

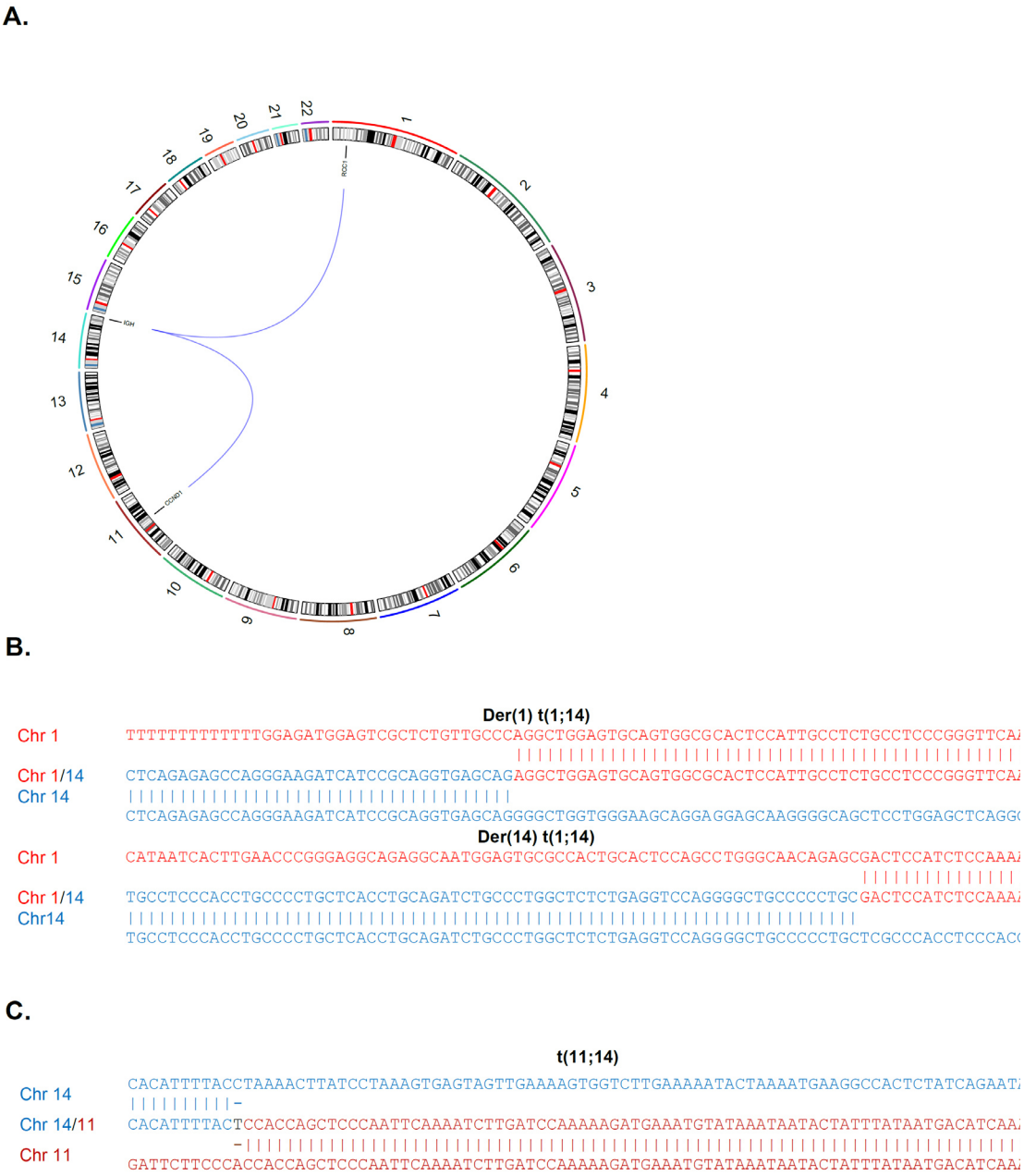


Figure 3.3.6.2. Identification of a t(1;14)(p35;q32). A: Circos plot. B: Complex breakpoint on chromosomes 1 and 14. C: Breakpoint on chromosomes 11 and 14

	AL	MM	MGUS
t(11;14)	30% (n=7/24)	19% (n=86/463)	12% (n=4/33)
t(4;14)	0% (n=0/24)	13% (n=59/463)	9.1% (n=3/33)
HRD	30% (n=7/24)	46% (n=52/114)	38.9% (n=35/90)
gain(1q)	12% (n=3/24) one amp(1q)	36% (n=32/114)	23% (n=21/90)
del(1p)	4% (n=1/24)	30% (n=34/114)	5.6% (n=5/90)
del(17p)	0%	7% (n=8/114)	1.1% (n=1/90)
del(14q)	0%*	38% (n=53/114)	14.4% (n=13/90)

Table 3.3.6.1. Frequencies of main cytogenetic alterations in plasma cell diseases

3.3.7. AL is not determined by *APCS* polymorphisms

To explain the occurrence of AL in these patients, we attempted to analyse the role of either germline or tumour variants in the *APCS* gene that codes for the serum amyloid P (SAP) component protein. There were no mutations in the *APCS* gene. Twenty SNPs were identified but only two had alternative alleles in our dataset: rs2808661 and rs28383572. Frequencies were in keeping with reported data from European populations (ss1293064888) (Consortium, 2015) and aggregated populations (ss1685860962),(Lek *et al*, 2016) **Table 3.3.7.1.** There is therefore no evidence in this dataset suggesting the role of SAP polymorphism in the occurrence of AL.

SNP	Genotype	AL	ss1293064888	ss1685860962
rs2808661	A/G	11.1%	16%	15%
	G/G	89.3%	83%	84.6%
rs28383572	A/G	3.6%	2%	2.7%
	G/G	96.4%	97%	97.3%

Table 3.3.7.1. Reported frequencies of the APCS gene SNP compared to a previously reported European (ss1293064888) and worldwide (ss1685860962) population datasets

3.4. Discussion

This is the largest dataset of whole exome sequencing of AL to date. Combined, these data suggest the underlying disease in AL resembles other plasma-cell disorders such as MGUS and MM.

It is now established that MGUS cases have fewer mutations than MM patients and that the number of mutations per sample reflect disease complexity (Mikulasova *et al*, 2017; Walker *et al*, 2014). We had hypothesised that AL would have a similar number of mutations to MGUS. Nevertheless, the number of mutations in AL was comparable to MM suggesting AL is a more complex disease than MGUS. This may be explained by the number of patients that met MM criteria in our dataset and the trend suggesting these patients had more mutations than the others.

To date five AL cases have undergone whole exome sequencing (Paiva *et al*, 2016) and these studies failed to identify any unifying mutation. In our 24 cases, we also did not identify any disease defining mutation. Recurrent mutations in *IL7R*, a gene that encodes for a cell surface receptor known to regulate V(D)J recombination by altering the accessibility of DNA substrates to the recombinase in pro-B-cells (Corcoran *et al*, 1998) were noted. Attenuating or escaping the IL-7Rα proliferative signals is required to initiate light chain rearrangement, and mutations in this gene have previously been reported in MM cases (Walker *et al*, 2018) and both T and B-ALL cases (Cramer *et al*, 2016). These mutations are believed to enable homodimerization of the IL-7 receptor leading to constitutive signalling and enabling increased cell survival and proliferation (Zenatti *et al*, 2011). Previous studies suggest they require a secondary event such as a *NRAS* mutation to drive ALL (Cramer *et al*, 2018).The *IL7R* mutations did not co-occur with *NRAS* mutations in this dataset.

We were able to identify MM-defined driver mutations. They occurred at similar frequencies to MGUS and were less common than in MM. They were predominantly sub-clonal suggesting they occurred late during disease progression. These mutations translated into alterations of the NF-κB, MAPK and DNA-repair pathways. Two samples had an *NRAS* mutation and occurred at codon 61, one of the typical hotspots of *NRAS* mutations (Hall *et al*, 1983; Walker *et al*, 2015). None of the *NRAS* mutated patients had classical CRAB criteria but one had another myeloma defining event (SFLC ratio>100 with an involved chain greater than 1000 mg/L). In previous data

published by Kim et al.,(Kim *et al*, 2016) only patients with CRAB criteria had *NRAS* mutations. These data are consistent with observations from MGUS and SMM where *NRAS* mutations may be found (Mikulasova *et al*, 2017; Bezieau *et al*, 2001; Rossi *et al*, 2017) with lower frequencies than in MM.

There was a 93% overlap in mutated genes indicating a common spectrum of mutations between AL and other plasma-cell disorders. Attempts to cluster patients based on MM driver genes from a previously published sequencing study placed AL close to MM and MGUS and not with other lymphoid malignancies. This is, to some extent, expected, especially as we only sequenced AL amyloidosis plasma cells, but emphasises the proximity of AL to other plasma-cell disorders.

The only recurrent translocation was t(11;14) with no evidence of other classical non-canonical *IGH* translocations such as t(4;14), t(14;16) or t(14;20). There was nevertheless a t(1;14)(q35.3; q32) involving the *RRC1* gene between exon 1 and 2 and the *IGHG4* switch region. The *RCC1* gene has previously been implicated in chronic lymphocytic leukaemia (CLL) in a t(1;6) translocation that involved *IRF4* instead of the *IGH* locus (Put *et al*, 2012). *RCC1* is a cell cycle regulatory protein that acts as a Ran guanine exchange factor. This gene is thought to play a key role in nucleo-cytoplasmic transport, mitosis and nuclear envelope assembly under Myc regulation (Tsuneoka *et al*, 1997). The effect of this translocation remain unclear but the breakpoint (intron 2) occurs upstream from both the main translation initiation site (exon 5) and the Myc binding domains. It can therefore be hypothesised that it places the *IGH* enhancer within 25kb from the main translation initiation site leading to an overexpression of the normal RCC1 isoforms. Some AL iFISH cohorts (Bochtler *et al*, 2011; Warsame *et al*, 2015) report a high incidence of *IGH* translocations with unknown partners (ranging from 9-20%): this could be explained by either technical limitations or other non-canonical translocations or/and break-points. So far, non-canonical translocations, such as t(1;20)(q21;q11),(Otokida *et al*, 1990) have been described in case reports with no further information regarding suspected partner genes or fusion transcripts.

Copy number changes were similar to myeloma and MGUS with hyperdiploidy being a common feature in a third of our samples, unlike previous reports (Bochtler *et al*, 2011). This may explain the relatively low frequency of t(11;14) translocations in

our dataset (Bochtler *et al*, 2015) and the high sensitivity to copy number change of WES approaches. Unlike data published by Granzow et al (Granzow *et al*, 2017) we did not find any evidence of del(14q) in our dataset. Incidence of copy number changes were lower than previously published in myeloma, but in keeping with MGUS datasets (Mikulasova *et al*, 2017). There were no del(17p) in AL patients, but this was not statistically different to MM as it only affects 8-10% of MM patients (Walker *et al*, 2015; Avet-Loiseau *et al*, 2012).

Beyond the genetic makeup of the tumour samples, we were also interested in the presence of SAP (Serum Amyloid P component) polymorphisms. SAP is a highly conserved, pentameric extracellular protein that makes up 14% of the dry mass of amyloid deposits (Botto *et al*, 1997). The absence of *APCS* SNP imbalance in our cohort suggests that protein polymorphisms do not play a role in the occurrence of AL amyloidosis. The relative frequency of *APCS* genotypes was similar to that expected in a general European population (Lek *et al*, 2016), in keeping with previous GWAS studies (da Silva Filho *et al*, 2017).

Overall, these data support the idea that AL is a heterogeneous entity that can arise from the complete spectrum of plasma cell diseases ranging from MGUS (that bear a lower tumour burden, fewer mutations, fewer driver mutations) to MM (that yield a larger tumour burden, more mutations, including more driver mutations). This is supported by similar copy number abnormalities and translocations and non-synonymous mutations. The genetic changes identified are unlikely to explain the clinical presentation of AL amyloidosis; neither does the SAP genotype. Any causal link between genetic changes in plasma cells and light chain instability remains unknown. We suggest future studies should examine more closely the interaction between SAP, light-chains, and target peripheral organs. Our data, however, support the ongoing use of myeloma-based therapy for this disease.

3.5. References

Abraham, R.S., Ballman, K.V., Dispenzieri, A., Grill, D.E., Manske, M.K., Price-Troska, T.L., Paz, N.G., Gertz, M.A. & Fonseca, R. (2005) Functional gene expression analysis of clonal plasma cells identifies a unique molecular profile for light chain amyloidosis. *Blood*, **105,** 794–803.

Abraham, R.S., Manske, M.K., Zuckerman, N.S., Sohni, A., Edelman, H., Shahaf, G., Timm, M.M., Dispenzieri, A., Gertz, M.A. & Mehr, R. (2007) Novel analysis of clonal diversification in blood B cell and bone marrow plasma cell clones in immunoglobulin light chain amyloidosis. *Journal of Clinical Immunology*, **27,** 69–87.

Abu-Duhier, F.M., Goodeve, A.C., Wilson, G.A., Care, R.S., Peake, I.R. & Reilly, J.T. (2001) Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *British Journal of Haematology*, **113,** 983–988.

Avet-Loiseau, H., Attal, M., Campion, L., Caillot, D., Hulin, C., Marit, G., Stoppa, A.-M., Voillat, L., Wetterwald, M., Pegourie, B., Voog, E., Tiab, M., Banos, A., Jaubert, J., Bouscary, D., Macro, M., Kolb, B., Traulle, C., Mathiot, C., Magrangeas, F., et al (2012) Long-term analysis of the IFM 99 trials for myeloma: cytogenetic abnormalities [t(4;14), del(17p), 1q gains] play a major role in defining long-term survival. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **30,** 1949–1952.

Bezieau, S., Devilder, M.C., Avet-Loiseau, H., Mellerin, M.P., Puthier, D., Pennarun, E., Rapp, M.J., Harousseau, J.L., Moisan, J.P. & Bataille, R. (2001) High incidence of N and K-Ras activating mutations in multiple myeloma and primary plasma cell leukemia at diagnosis. *Human Mutation*, **18,** 212–224.

Bochtler, T., Hegenbart, U., Cremer, F.W., Heiss, C., Benner, A., Hose, D., Moos, M., Bila, J., Bartram, C.R., Ho, A.D., Goldschmidt, H., Jauch, A. & Schonland, S.O. (2008) Evaluation of the cytogenetic aberration pattern in amyloid light chain amyloidosis as compared with monoclonal gammopathy of undetermined significance reveals common pathways of karyotypic instability. *Blood*, **111,** 4700–4705.

Bochtler, T., Hegenbart, U., Heiss, C., Benner, A., Moos, M., Seckinger, A., Pschowski-Zuck, S., Kirn, D., Neben, K., Bartram, C.R., Ho, A.D., Goldschmidt, H., Hose, D., Jauch, A. & Schonland, S.O. (2011) Hyperdiploidy is less frequent in AL amyloidosis compared with monoclonal gammopathy of undetermined significance and inversely associated with translocation t(11;14). *Blood*, **117,** 3809–3815.

Bochtler, T., Hegenbart, U., Kunz, C., Benner, A., Kimmich, C., Seckinger, A., Hose, D., Goldschmidt, H., Granzow, M., Dreger, P., Ho, A.D., Jauch, A. & Schönland, S.O. (2016) Prognostic impact of cytogenetic aberrations in AL amyloidosis patients after high-dose melphalan: a long-term follow-up study. *Blood*, **128,** 594–602.

Bochtler, T., Hegenbart, U., Kunz, C., Granzow, M., Benner, A., Seckinger, A., Kimmich, C., Goldschmidt, H., Ho, A.D., Hose, D., Jauch, A. & Schönland, S.O. (2015) Translocation t(11;14) is associated with adverse outcome in patients with newly diagnosed AL amyloidosis when treated with bortezomib-based regimens. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **33,** 1371–1378.

Botto, M., Hawkins, P.N., Bickerstaff, M.C., Herbert, J., Bygrave, A.E., McBride, A., Hutchinson, W.L., Tennent, G.A., Walport, M.J. & Pepys, M.B. (1997) Amyloid deposition is delayed in mice with targeted deletion of the serum amyloid P component gene. *Nature Medicine*, **3,** 855–859.

Comenzo, R.L., Zhang, Y., Martinez, C., Osman, K. & Herrera, G.A. (2001) The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V(L) germ line gene use and clonal plasma cell burden. *Blood*, **98,** 714–720.

Consortium, T. 1000 G.P. (2015) A global reference for human genetic variation. *Nature*, **526,** 68–74.

Corcoran, A.E., Riddell, A., Krooshoop, D. & Venkitaraman, A.R. (1998) Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature*, **391,** 904–907.

Cramer, S.D., Aplan, P.D. & Durum, S.K. (2016) Therapeutic targeting of IL-7Rα signaling pathways in ALL treatment. *Blood*, **128,** 473–478.

Cramer, S.D., Hixon, J.A., Andrews, C., Porter, R.J., Rodrigues, G.O.L., Wu, X., Back, T., Czarra, K., Michael, H., Cam, M., Chen, J., Esposito, D., Senkevitch, E., Negi, V., Aplan, P.D., Li, W. & Durum, S.K. (2018) Mutant IL-7Rα and mutant NRas are sufficient to induce murine T cell acute lymphoblastic leukemia. *Leukemia*.

Granzow, M., Hegenbart, U., Hinderhofer, K., Hose, D., Seckinger, A., Bochtler, T., Hemminki, K., Goldschmidt, H., Schönland, S.O. & Jauch, A. (2017) Novel recurrent chromosomal aberrations detected in clonal plasma cells of light chain amyloidosis patients show potential adverse prognostic effect: first results from a genome-wide copy number array analysis. *Haematologica*, **102,** 1281–1290.

Hall, A., Marshall, C.J., Spurr, N.K. & Weiss, R.A. (1983) Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1. *Nature*, **303,** 396–400.

Havelange, V., Pekarsky, Y., Nakamura, T., Palamarchuk, A., Alder, H., Rassenti, L., Kipps, T. & Croce, C.M. (2011) IRF4 mutations in chronic lymphocytic leukemia. *Blood*, **118,** 2827–2829.

Kim, M.S., Yoo, N.J. & Lee, S.H. (2011) Very low mutation frequency of exons 5-9 in the CARD11 gene from 186 adult acute leukemia and 31 multiple myeloma samples. *Leukemia Research*, **35,** e46-47.

Kim, S.J., Shin, H.-T., Lee, H.-O., Kim, N.K.D., Yun, J.W., Hwang, J.H., Kim, K. & Park, W.-Y. (2016) Recurrent mutations of MAPK pathway genes in multiple myeloma but not in amyloid light-chain amyloidosis. *Oncotarget*, **7,** 68350–68359.

Kryukov, F., Kryukova, E., Brozova, L., Kufova, Z., Filipova, J., Growkova, K., Sevcikova, T., Jarkovsky, J. & Hajek, R. (2016) Does AL amyloidosis have a unique genomic profile? Gene expression profiling meta-analysis and literature overview. *Gene*, **591,** 490–498.

Kyle, R. & Rajkumar, S. (2009) Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, **23,** 3–9.

Landau, D.A., Sun, C., Rosebrock, D., Herman, S.E.M., Fein, J., Sivina, M., Underbayev, C., Liu, D., Hoellenriegel, J., Ravichandran, S., Farooqui, M.Z.H., Zhang, W., Cibulskis, C., Zviran, A., Neuberg, D.S., Livitz, D., Bozic, I., Leshchiner, I., Getz, G., Burger, J.A., et al (2017) The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nature Communications*, **8,** 2185.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., Kiezun, A., Hammerman, P.S., McKenna, A., Drier, Y., Zou, L., Ramos, A.H., Pugh, T.J., Stransky, N., Helman, E., Kim, J., et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, **499,** 214–218.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536,** 285–291.

Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., Novak, A.J., Dogan, A., Ansell, S.M., Link, B.K., Zou, L., Gould, J., Saksena, G., Stransky, N., Rangel-Escareño, C., Fernandez-Lopez, J.C., et

al (2012) Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **109,** 3879–3884.

Merlini, G. & Bellotti, V. (2003) Molecular mechanisms of amyloidosis. *The New England Journal of Medicine*, **349,** 583–596.

Mikulasova, A., Wardell, C.P., Murison, A., Boyle, E.M., Jackson, G.H., Smetana, J., Kufova, Z., Pour, L., Sandecka, V., Almasi, M., Vsianska, P., Gregora, E., Kuglik, P., Hajek, R., Davies, F.E., Morgan, G.J. & Walker, B.A. (2017) The spectrum of somatic mutations in monoclonal gammopathy of undetermined significance indicates a less complex genomic landscape than that in multiple myeloma. *Haematologica*, **102,** 1617–1625.

Morgan, G.J., Walker, B.A. & Davies, F.E. (2012) The genetic architecture of multiple myeloma. *Nature Reviews Cancer*, **12,** 335–348.

Otokida, K., Yoshida, H., Mizunuma, Y., Yamada, M. & Hiramori, K. (1990) Multiple myeloma associated with amyloidosis and t(1;20)(q21;q11) translocation. *The Tohoku Journal of Experimental Medicine*, **162,** 363–365.

Paiva, B., Martinez-Lopez, J., Corchete, L.A., Sanchez-Vega, B., Rapado, I., Puig, N., Barrio, S., Sanchez, M.-L., Alignani, D., Lasa, M., Coca, A.G. de, Pardal, E., Oriol, A., Garcia, M.-E.G., Escalante, F., González-López, T.J., Palomera, L., Alonso, J., Prosper, F., Orfao, A., et al (2016) Phenotypic, transcriptomic, and genomic features of clonal plasma cells in light-chain amyloidosis. *Blood*, **127,** 3035–3039.

Perfetti, V., Palladini, G., Casarini, S., Navazza, V., Rognoni, P., Obici, L., Invernizzi, R., Perlini, S., Klersy, C. & Merlini, G. (2012) The repertoire of λ light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, IGLV1-44. *Blood*, **119,** 144–150.

Put, N., Lierman, E., Wlodarska, I., Hagemeijer, A., Vandenberghe, P. & Michaux, L. (2012) Translocation t(1;6)(p35.3;p25.2) Involves RCC1 and IRF4 and Is Not Restricted to Unmutated Chronic Lymphocytic Leukemia. *Blood*, **120,** 4584–4584.

Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., et al (2014) International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet. Oncology*, **15,** e538–e548.

Rossi, A., Voigtlaender, M., Janjetovic, S., Thiele, B., Alawi, M., März, M., Brandt, A., Hansen, T., Radloff, J., Schön, G., Hegenbart, U., Schönland, S., Langer, C., Bokemeyer, C. & Binder, M. (2017) Mutational landscape reflects the biological continuum of plasma cell dyscrasias. *Blood Cancer Journal*, **7,** e537.

Sanchorawala, V., Blanchard, E., Seldin, D.C., O'Hara, C., Skinner, M. & Wright, D.G. (2006) AL amyloidosis associated with B-cell lymphoproliferative disorders: frequency and treatment outcomes. *American Journal of Hematology*, **81,** 692–695.

da Silva Filho, M.I., Försti, A., Weinhold, N., Meziane, I., Campo, C., Huhn, S., Nickel, J., Hoffmann, P., Nöthen, M.M., Jöckel, K.-H., Landi, S., Mitchell, J.S., Johnson, D., Morgan, G.J., Houlston, R., Goldschmidt, H., Jauch, A., Milani, P., Merlini, G., Rowcieno, D., et al (2017) Genome-wide association study of immunoglobulin light chain amyloidosis in three patient cohorts: comparison with myeloma. *Leukemia*, **31,** 1735–1742.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. & Narechania, A. (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, **13,** 2129–2141.

Tsuneoka, M., Nakano, F., Ohgusu, H. & Mekada, E. (1997) c-myc activates RCC1 gene expression through E-box elements. *Oncogene*, **14,** 2301–2311.

Walker, B.A., Boyle, E.M., Wardell, C.P., Murison, A., Begum, D.B., Dahir, N.M., Proszek, P.Z., Johnson, D.C., Kaiser, M.F., Melchor, L., Aronson, L.I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J.R., Brioli, A., Mikulasova, A., Cairns, D.A., Gregory, W.M., Quartilho, A., et al (2015) Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018) A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*.

Walker, B.A., Wardell, C.P., Melchor, L., Brioli, A., Johnson, D.C., Kaiser, M.F., Mirabella, F., Lopez-Corral, L., Humphray, S., Murray, L., Ross, M., Bentley, D., Gutiérrez, N.C., Garcia-Sanz, R., San Miguel, J., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2014) Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*, **28,** 384–390.

Warsame, R., Kumar, S.K., Gertz, M.A., Lacy, M.Q., Buadi, F.K., Hayman, S.R., Leung, N., Dingli, D., Lust, J.A., Ketterling, R.P., Lin, Y., Russell, S., Hwa, L., Kapoor, P., Go, R.S., Zeldenrust, S.R., Kyle, R.A., Rajkumar, S.V. & Dispenzieri, A. (2015) Abnormal FISH in patients with immunoglobulin light chain amyloidosis is a risk factor for cardiac involvement and for death. *Blood Cancer Journal*, **5,** e310.

Zenatti, P.P., Ribeiro, D., Li, W., Zuurbier, L., Silva, M.C., Paganin, M., Tritapoe, J., Hixon, J.A., Silveira, A.B., Cardoso, B.A., Sarmento, L.M., Correia, N., Toribio, M.L., Kobarg, J., Horstmann, M., Pieters, R., Brandalise, S.R., Ferrando, A.A., Meijerink, J.P., Durum, S.K., et al (2011) Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nature Genetics*, **43,** 932–939.

Chapter 4: Long-Term Follow-up Identifies That *BRAF* and *DIS3* Mutations Impact Outcome In Multiple Myeloma

4.1. Summary

Copy number changes and translocations have been studied extensively in many datasets with long term follow-up. The impact of mutations remains debated given the short time to follow-up of most datasets. We performed targeted panel sequencing covering 125 myeloma-specific genes and the loci involved in translocations in 223 newly diagnosed myeloma samples recruited into one of the Total Therapy Trials (TT). As expected, the most commonly mutated genes were *NRAS*, *KRAS*, and *BRAF* making up 44% of patients. Double-Hit, *BRAF* and *DIS3* mutations had an impact on outcome alongside classical risk factors in the context of an intensive treatment approach. We were able to identify both V600E and non-V600E *BRAF* mutations, 58% of which were predicted to be hypoactive or kinase dead. Interestingly, 44% of the hypoactive/kinase dead *BRAF* mutated patients showed co-occurring alterations in *KRAS, NRAS* or activating *BRAF* mutations suggesting they play a role in the oncogenesis of multiple myeloma (MM) by facilitating MAPK activation and may lead to chemo resistance. Overall, these data highlight the importance of mutational screening to better understand newly diagnosed MM (NDMM) and may lead to patient specific mutation-driven treatment approaches.

4.2. Introduction

Multiple Myeloma (MM) is a haematological malignancy of plasma cells that afflicts around 30,000 people in the US per year with a five-year survival rate of 47% (SEER data, 2018). High risk MM (HRMM) is seen in up to 30% of newly diagnosed cases whose

outcome, in contrast to the majority of MM cases, has seen very little improvement over the past 15 years (Lonial *et al*, 2015) with a median progression free survival (PFS) of 1.8 years and overall survival (OS) of 2.6 years (Walker *et al*, 2018a). There is, therefore, a clear need to identify these patients in order to apply relevant new approaches in their management.

The study of MM has identified many genetic events that are associated with event free survival (EFS) and overall survival (OS). Some of these features occur with a greater frequency in HRMM cases, including translocations into the immunoglobulin (Ig) loci involving chromosomes 4 and 16, which define two etiological subgroups [t(4;14), 15%, and t(14;16), 5%].(Walker *et al*, 2015a) Other instability mechanisms associated with HRMM include additional structural variations such as del(1p), and del(17p), jumping translocations of 1q and secondary translocations to *MYC* at 8q24 (Magrangeas *et al*, 2013; Egan *et al*, 2012; Sawyer *et al*, 2014).

A great deal is known about the genetics of MM with over 800 genomes and 2000 exomes sequenced (Walker *et al*, 2015a; Egan *et al*, 2012; Walker *et al*, 2018b; Hoang *et al*, 2018; Bolli *et al*, 2014, 2018, 2017; Johnson *et al*, 2017). However, the prognostic impact of mutations has not been widely evaluated and available datasets have generally had a relatively short follow-up ranging from 22 to 25 months, with one dataset being up to 5.4 years (Walker *et al*, 2015a; Bolli *et al*, 2018). These analyses have identified a diverse range of mutations that are associated with outcome, making it important to extend these observations over time in larger studies with robust diagnostic technologies.

Despite falling prices, it is not feasible to run whole exome sequencing (WES) for every MM patient, and even then, many translocations outside of the capture region would be missed, requiring a combination of fluorescence in situ hybridization (FISH) or gene expression profiling. Since this approach is time and cost prohibitive it has led to the adoption of targeted sequencing to generate data in a timely, cost-effective manner. This led us to design a custom myeloma targeted panel which provided rapid, fiscally responsible characterization of patient subgroups.

We evaluated this panel on 223 newly diagnosed MM (NDMM) patients included in the Total Therapy trials (TT) and correlated results to both gene expression and clinical data, with the ultimate aim of this work being to effectively characterize NDMM patients and identify the impact of mutations long-term.

4.3. Results

4.3.1. Patients characteristics

A total of 223 patients were sequenced and included in the study. Overall, they were representative of a fit-newly diagnosed population. The median follow-up time from diagnosis was 8.14 years (95% CI 7.39-9.02). The median OS was not met at the time of analysis and the 8-year OS was 61% (95% CI 54-69%). The median EFS was 6.16 years (95% CI 5.18-7.75). A summary of patient characteristics may be found in **Table 4.3.1.1.**

The most commonly mutated genes were *KRAS* (23% of patients), *NRAS* (17% of patients) and *BRAF* (12% of patients), **Figure 4.3.1.1.,** in keeping with previously published datasets. The incidence of mutations was similar to the MGP study, **Table 4.3.1.2**.

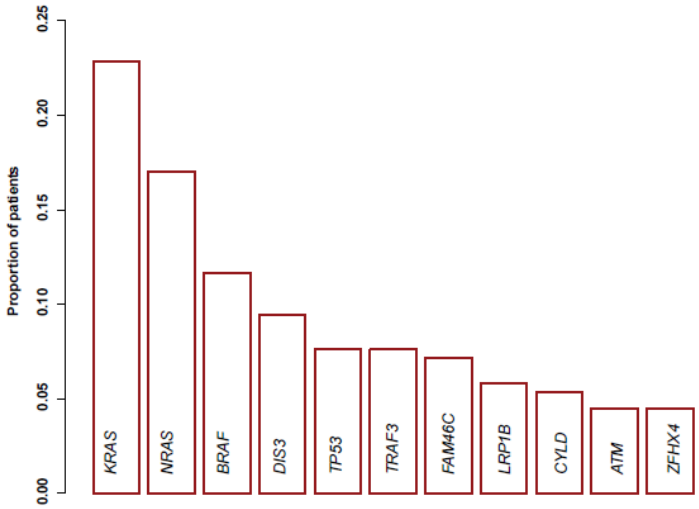


Figure 4.3.1.1. Most frequent mutations seen in MM.

	223-baseline study	Combined TT3a-3b-4-4like-5a-5b-6
Number of patients	223	1039
Inclusion dates	02/2004 to 08/2017	02/2004 to 08/2017
Median Follow Up	8.14 years (95% CI 7.39-9.02)	8.35 years (95% CI 8.00-8.63)
Median EFS	6.16 years (95% CI 5.18-7.75)	4.8 years (95% CI 52%-58%)
8- year OS	61% (95% CI 54-69%)	42% (95% CI 39%-45%)
Median age (years)	59 (range: 30-75)	61 (range: 30-76)
Sex ratio M:F	1.8:1	1.6:1
Ethnicity % - **White** - **African-American** - **Other**	 88% (n=197) 10% (n=22) 2% (n=4)	 86.8% (n=902) 9.7% (n=101) 3.5% (n=36)
ISS % - **I** - **II** - **III**	 26.5% (n=59) 43.5% (n=97) 30.0% (n=67)	 34.0% (n=352) 40.2% (n=416) 25.8% (n=267)
R-ISS % - **I** - **II** - **III**	 17.0% (n=38) 67.7% (n=151) 15.2% (n=34)	
GEP70 high risk %	16.1% (n=36)	15.9% (n=165)

Table 4.3.1.1. Summary of patient's characteristics and comparability to the complete TT trial population.

Gene	Percentage in this study (n=223)	Percentage in MGP (n=1273)
KRAS	22.87%	21.84%
NRAS	17.04%	17.44%
BRAF	11.66%	8.01%
DIS3	9.42%	9.98%
TP53	7.62%	5.66%
TRAF3	7.62%	5.26%
FAM46C	7.17%	9.35%
LRP1B	5.83%	7.31%
LRRK2	5.83%	1.18%*
CYLD	5.38%	3.38%
ATM	4.48%	4.32%
ZFHX4	4.48%	4.70%

Table 4.3.1.2. Incidence of the most frequently mutated genes in this TT baseline study and comparison to MGP study. *p<0.05.

The incidence of translocation and copy number were in keeping with previously published data. *MYC* translocations were identified in 26% of cases: they involved the Ig locus in 39% of cases and non-Ig partners in 61% of cases. The breakpoints were within the previously published hotspots **Figure 4.3.1.2.A**. *MYC* deletions and gain were seen in 16% and 28% of patients, respectively. Overall, *MYC* events were seen in 47% of patients. A summary of the translocations and comparison to the MGP(Walker *et al*, 2018a) data may be found in **Table 4.3.1.3.**

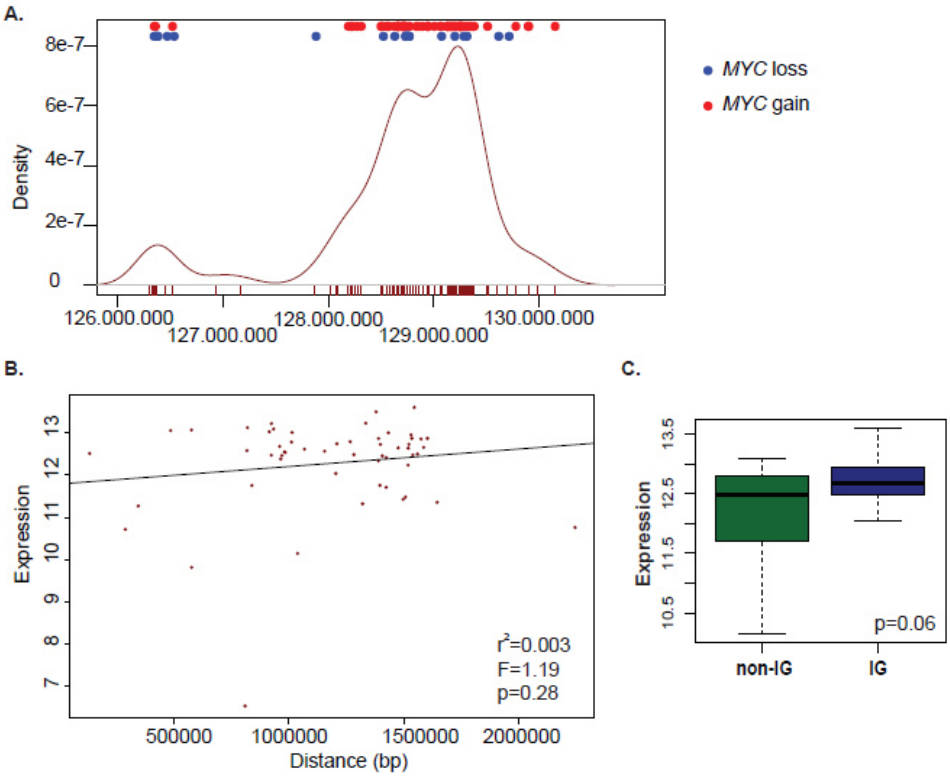


Figure 4.3.1.2. MYC rearrangements A. Density plot featuring the breakpoint on 8q24 consistent with the three hotspots previously published. Above, correspondence between the breakpoints and co-occurring *MYC* loss or gain. B. Correlation between distance from *MYC* and expression suggesting there is no correlation C. Difference in expression between IG and non-IG translocations.

Gene	Percentage in this study (n=223)	Percentage in MGP (n=1273)	p-value
CCND1	13% (n=29/223)	18% (n=234/1273)	p=0.051
CCDN3	3.5% (n=8/223)	1% (n=14/1273)	p=0.005
MAF or MAFB	6.6% (n=15/223)	5% (n=62/1273)	p= 0.24
MMSET	14% (n=30/223)	12% (n=155/1273)	p=0.60

Table 4.3.1.3. Incidence of translocations in this study and in the MGP dataset.

Exome sequencing identifies an APOBEC-derived mutational signature in approximately 80% of t(14;16) samples (Walker *et al*, 2015b). We performed nNMF analysis on our targeted sequencing in order to determine if we can identify an APOBEC signature. In an attempt to maximise the cophenetic and the dispersion, we determined that the best fit for the number of signatures was three, **Figure 4.3.1.3.** We were to visualise, three homogeneous groups on the rank plot, **Figure 4.3.1.4.A.** to identify which previously defined signatures our three signatures (**Figure 4.3.1.4.C-E**) best fitted, we performed ward clustering and identified one signature group dominated by signatures 2 and 13 (APOBEC signatures) and two signatures that clustered within the same clade dominated by signature 5, consistent with a them being background signatures, **Figure 4.3.1.4.B.** An APOBEC signature was seen in seven patients (3.2%), five of which had a t(14;16) and one a t(14;20) translocation, **Figure 4.3.1.5..** Both frequency and enrichment for the MAF subgroups were in keeping with previous reports (Walker *et al*, 2015b).

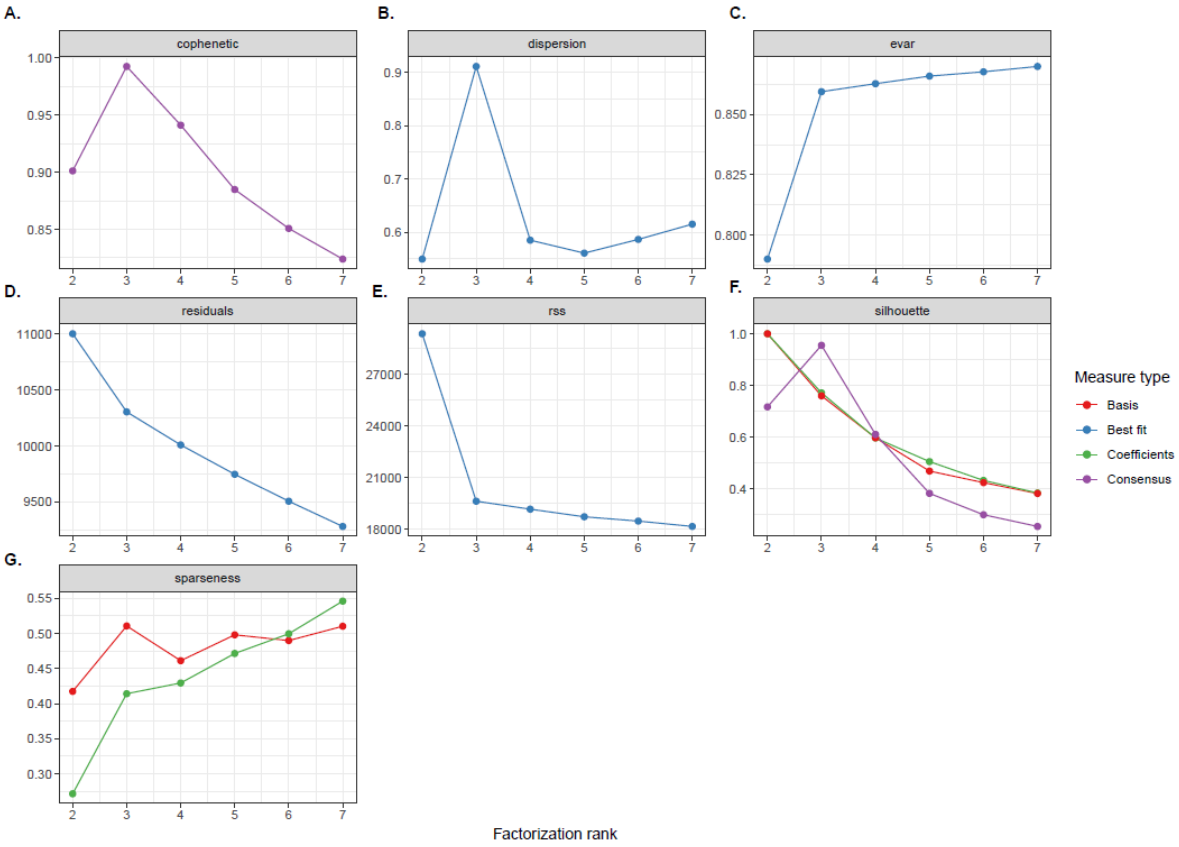


Figure 4.3.1.3. Result of the fitness of the nNMF analysis.

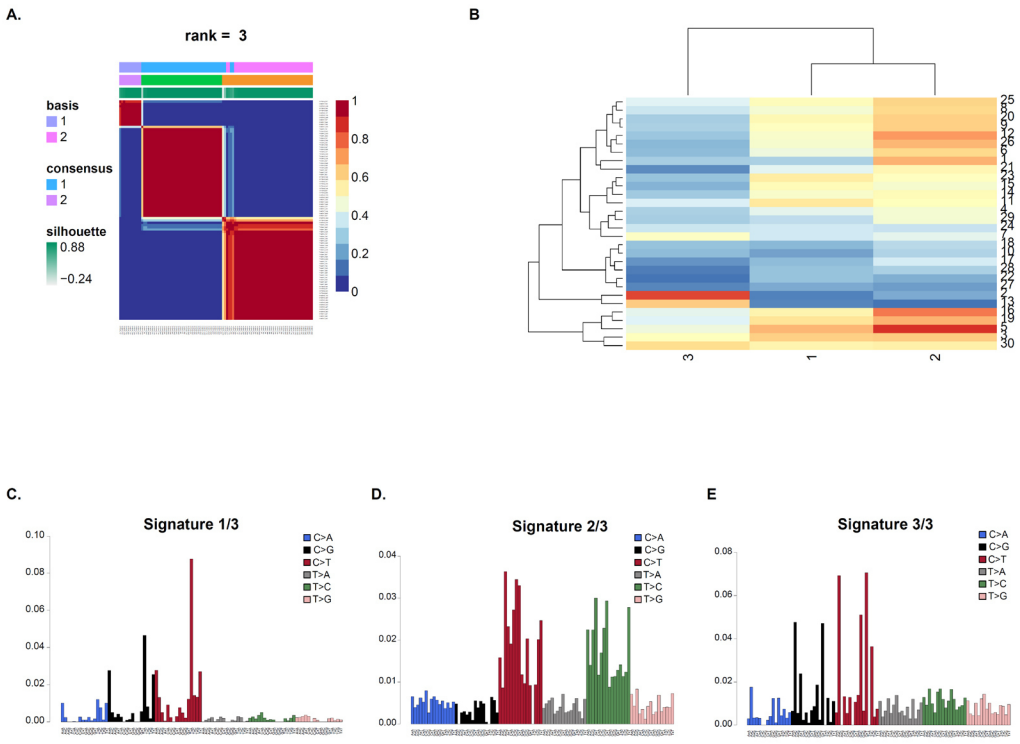


Figure 4.3.1.4. nNMF: signatures. A. Rank=3 was the best ranking option. B There are two background signatures within the same clade and an APOBEC (2,13) signature C. Background signature, D. Background signature, E. APOBEC signature (2 and 13).

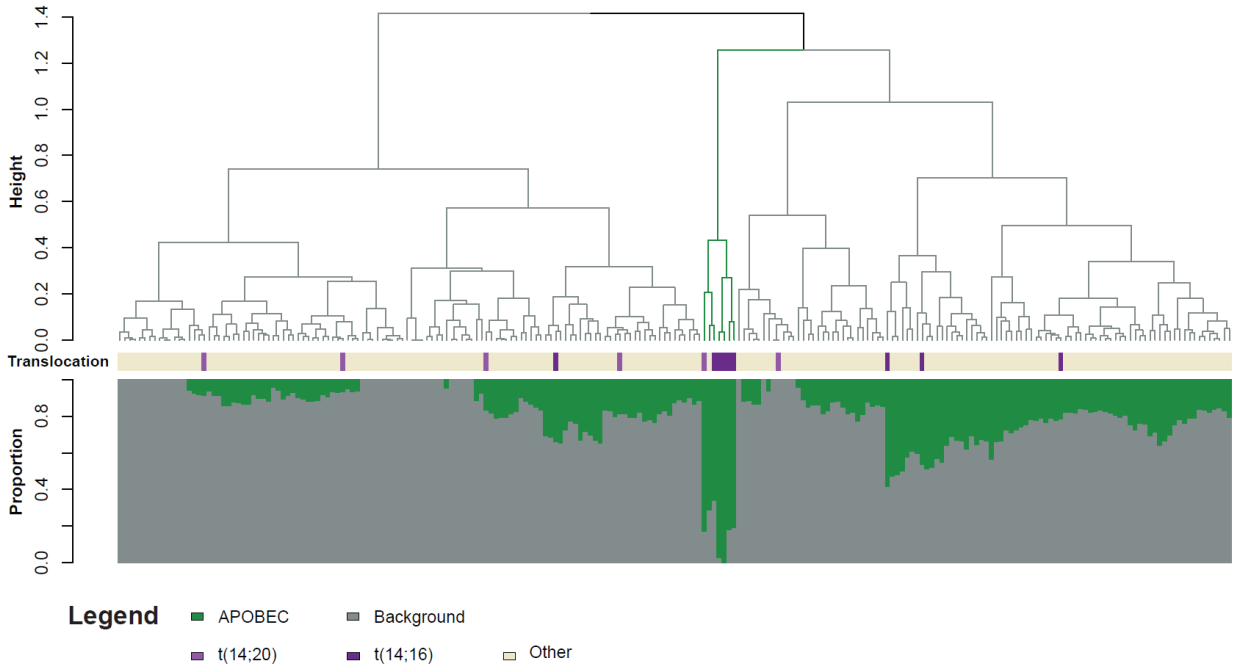


Figure 4.3.1.5. Dendrogram representation of the result of the nNMF showing an APOBEC cluster composed on predominantly *MAF* and *MAFB* patients.

4.3.2. Interactions between genomic abnormalities and Double-Hit myeloma

Pearson's correlation identified a significant correlation between *CYLD* mutations and deletions (r=0.31, p=2.07x10^{-7}), *TRAF3* mutations and deletions (r=0.34, p=1.5x10^{-7}), and *TP53* mutations and deletions (r=0.32, p=1.12x10^{-6}), as previously reported.(Walker *et al*, 2015a, 2018b) *ATM* mutations were positively correlated with the t(14;16) subgroup (r=0.40, p=3.33 x10^{-7}) and the presence of an APOBEC signature (r=0.62, p=8.66 x10^{-10}). There was no significant negative correlation between *DIS3* mutations and hyperdiploidy (HRD) (r=-0.11, p=0.09) in this dataset although they were positively correlated to the presence of t(4;14) (r=0.21, p=0.0005). On the other hand, del(13q) was negatively correlated to HRD (r=-0.26, p=0.0001), **Figure 4.3.2.1.**

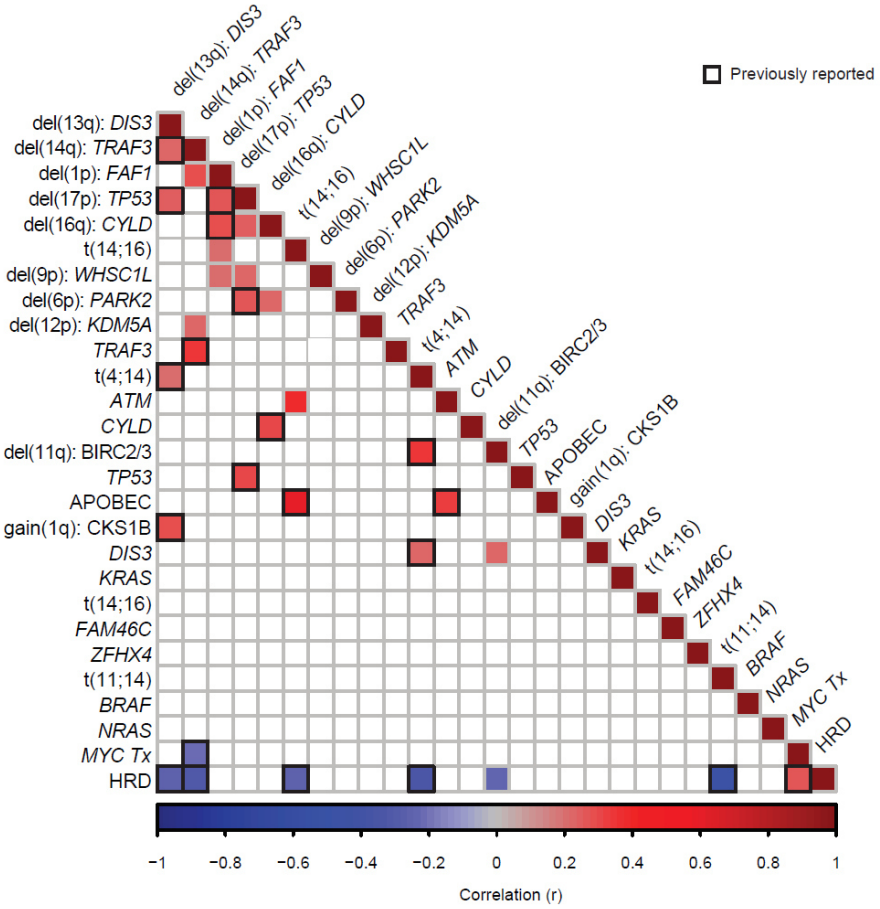


Figure 4.3.2.1. Corrplot representing the significant Pearson correlations seen in this dataset

We identified 8.1% patients with Double-Hit (ISS III on a background of amp(1q) or biallelic inactivation of *TP53*) which is not significantly different to the 6.1% of patients previously described. Double-Hit was associated with both an adverse EFS (median: 24.6 months (95% CI 10.6-42.7) versus and 6.77 years ((95% CI 5.60- ∞), p<0.0001) and OS 37.3 months ((95% CI 32.3-∞) versus 64% at 8 years ((95% CI 57%-73%), p<0.0001). Interestingly, when analysing the impact of Double-Hit in the TT population, which received two autologous stem cell transplants (ASCTs), the impact of Double-Hit was still significant. This is of particular interest in a double-ASCT population, **Figure 4.3.2.2.**, identifying a population who still has a dire outcome despite intensive treatment.

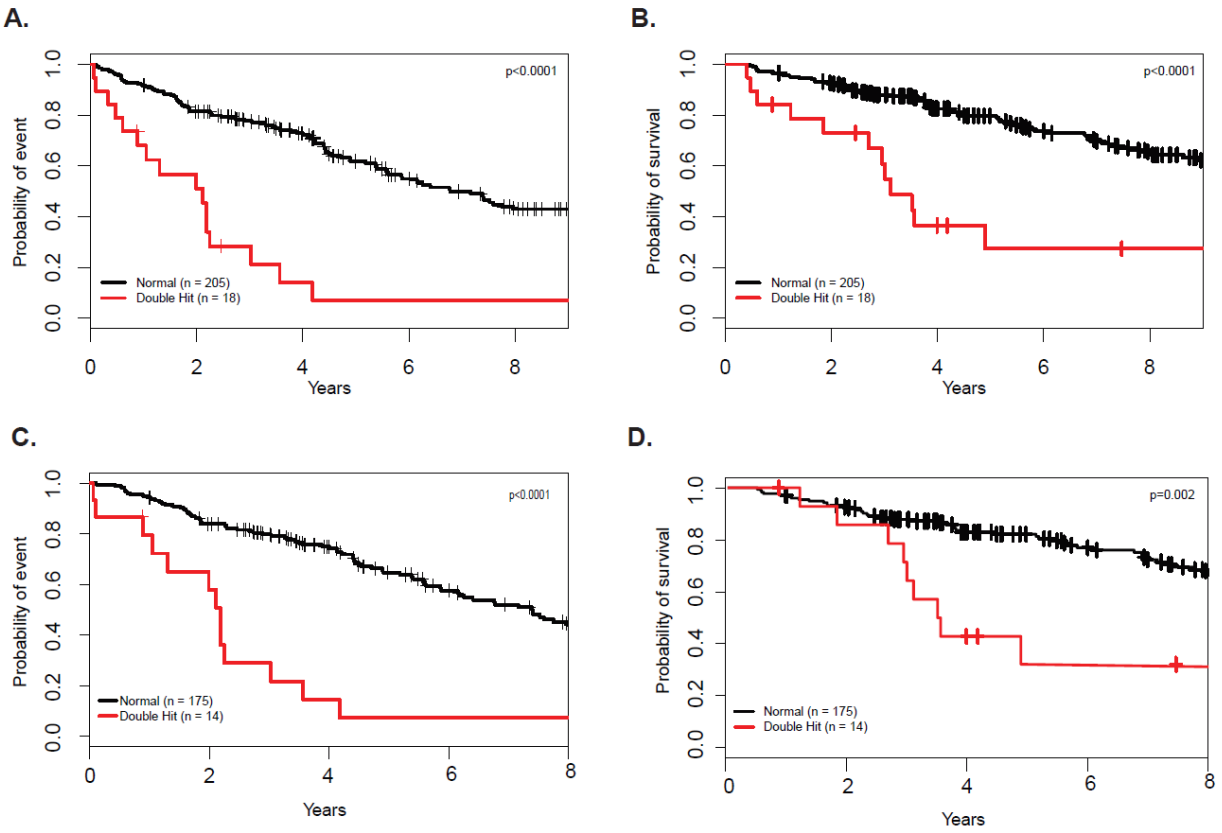


Figure 4.3.2.1.: Impact of Double-Hit on outcome. A. EFS B. OS C. EFS in the Intention to treat (ITT) population and D. OS in the ITT population.

4.3.3. Survival analysis identifies that *BRAF* and *DIS3* mutations are associated with an adverse outcome with long-term follow-up.

4.3.3.1. Univariate analysis

The results of univariate analyses for EFS and OS for molecular features are shown, **Figure 4.3.3.1.1.** Overall, this dataset behaved as expected with del(12p), del(17p), gain/amp(1q), and del(1p) being significantly associated with both adverse EFS and OS. Trisomy(9) and trisomy(19) were associated with a better EFS and OS and trisomy(2) and trisomy(5) resulted in a better EFS. The other trisomies (3, 15 and 21) had no impact on outcome. Sixteen percent of patients were considered as high risk according to the GEP70 score and they had a worse outcome than standard risk patients both in terms of EFS (HR=2.5 ((95% CI 1.6-3.9), p<0.0001) and OS (HR=3.5 ((95% CI 2.1-6), p<0.0001), **Figure 4.3.3.1.2.** The PR subgroup, was also associated with both short EFS and OS whereas the MF and MS subgroup were associated with a short EFS. Based on their ISS, 26.5%, 43.5% and 30% of patients were considered ISS I, ISS II and ISS III respectively with a HR of death of 2.7 ((95% CI 1.2-5.9), p=0.01) and 6.04 ((95% CI 2.8-13) p<0.0001) for ISS II and III respectively in comparison to ISS I. *MYC* translocations, gains and deletions, were associated with a difference in OS in this dataset, but not EFS **Figure 4.3.3.1.3.** In terms of mutations, *BRAF* mutations were associated with an adverse EFS (HR=2 (95% CI 1.2-3.3), p=0.009) and OS (HR=2.7 (95% CI 1.5-4.7), p=0.0007), **Figure 4.3.3.1.4.** *TP53* and *DIS3* mutations were associated with a worse EFS (HR=2.3 (95% CI 1.3-4.2), p=0.0065 and HR=2 (95% CI 1.2-3.5), p=0.009 respectively) but not OS (HR=1.5 (95% CI 0.67-3.2), p=0.34 and HR=1.2 (95% CI 0.56-2.4), p=0.68) respectively). We went on to test combinations of markers previously published such as DNA repair pathway mutations(Walker *et al*, 2015a) and bi-allelic *TP53*(Walker *et al*, 2018a). As previously shown(Walker *et al*, 2015a), DNA repair pathway mutations defined by the presence of an *ATM* or *ATR* mutation were associated with an adverse outcome in terms of EFS (HR=2.1 (95% CI 1.1-4.1), p=0.023) and OS (HR=2.2 (95% CI 1.1-4.6, p=0.033)), as was bi-allelic *TP53* (HR=4.3 (95% CI 2.4-7.7), p<0.0001) and OS (HR=2.8 (95% CI 1.4-5.6), p=0.004). There was no significant impact of IGL translocations on outcome, **Figure 4.3.3.5.** These data are summarized in **Table 4.3.3.1.1**.

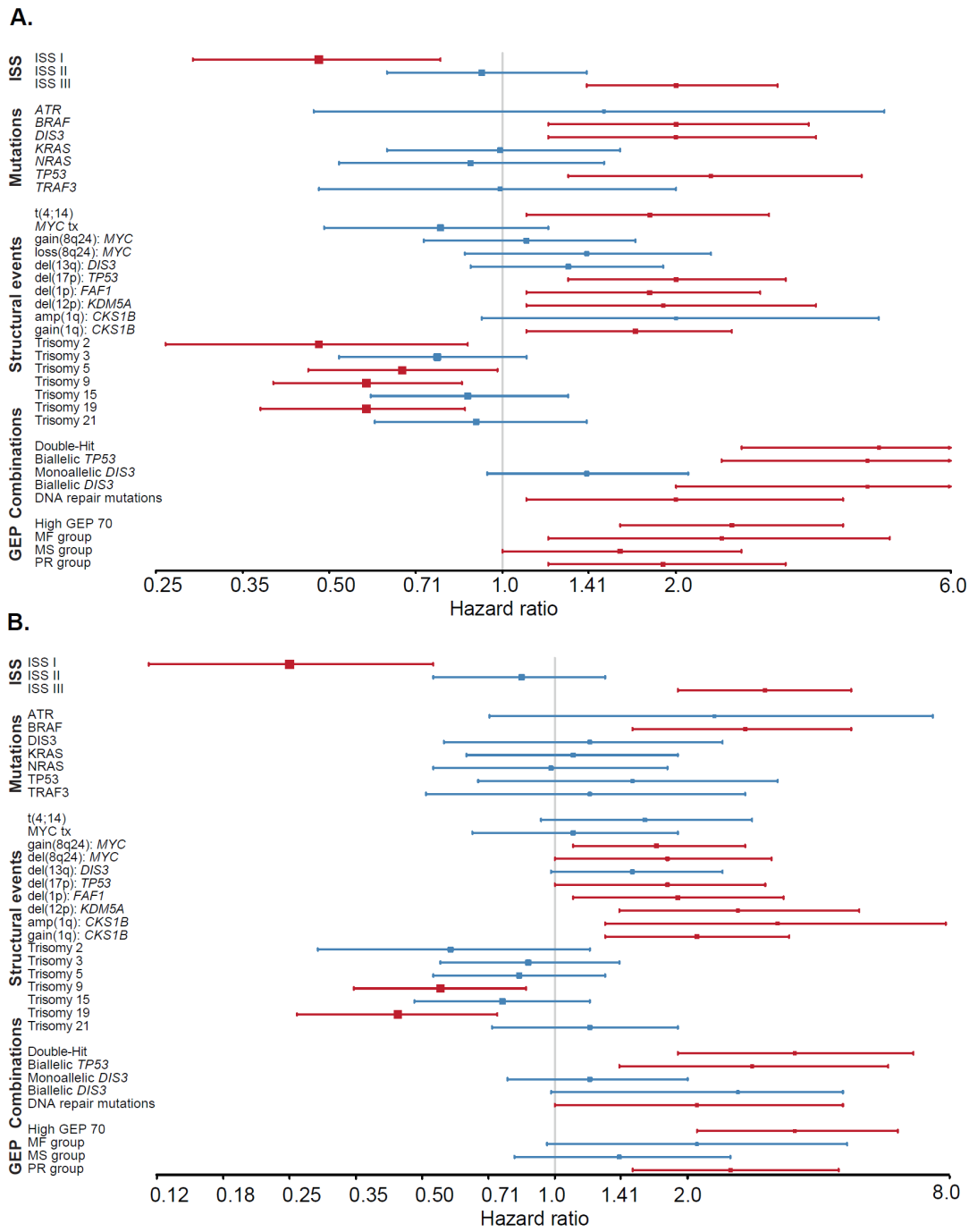


Figure 4.3.3.1.1.: Summary of the univariate analysis. A. EFS. B. OS. Forest plots representing the result of the univariate analysis. In red, those that were significantly associated with outcome at the level of $p<0.05$, in blue the non-significant variables.

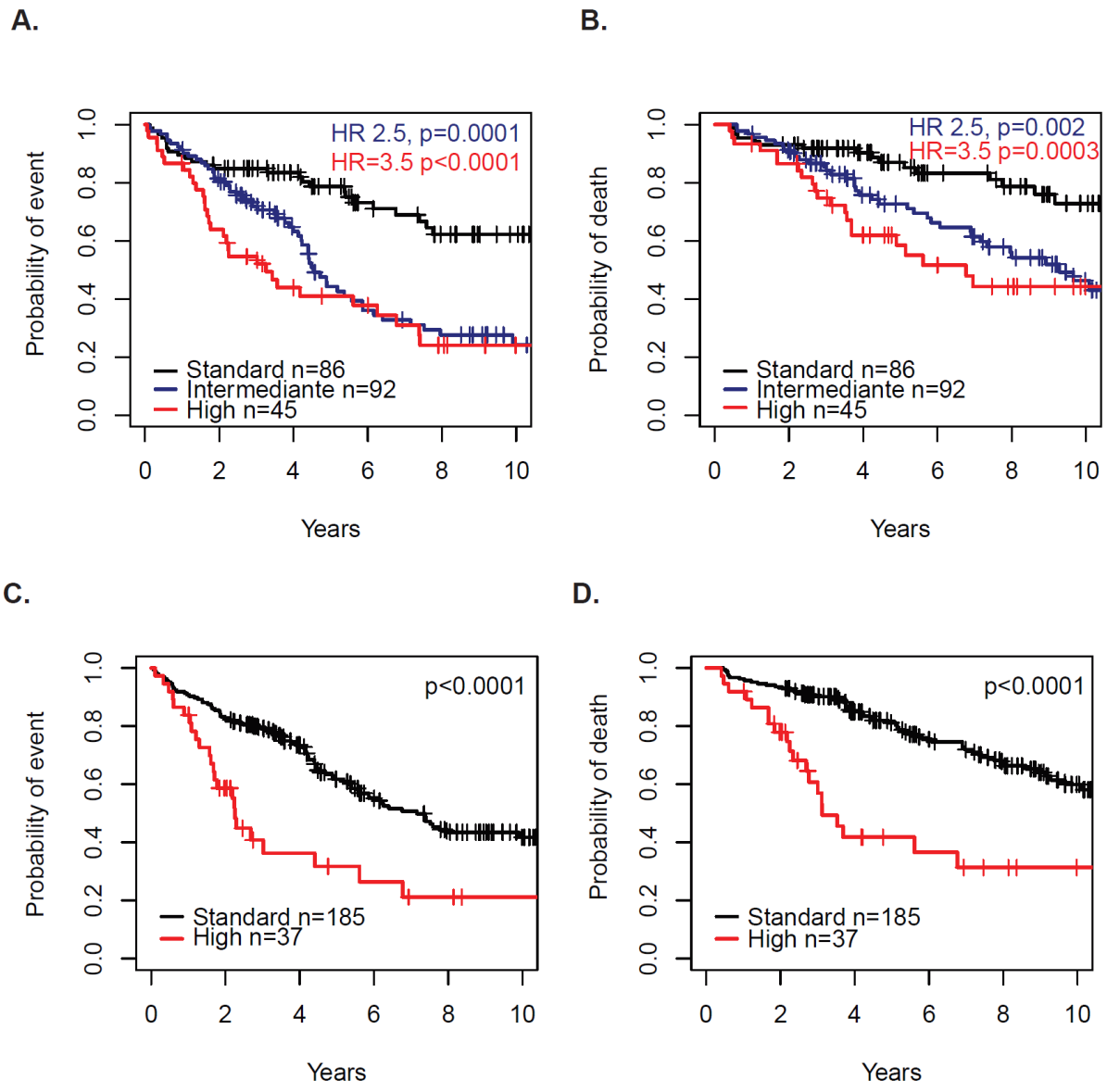


Figure 4.3.1.1.2: Impact of current survival models on outcome in this dataset. IFM2009 model on EFS (A.) and OS (B). GEP70 score on EFS (C) and OS (D).

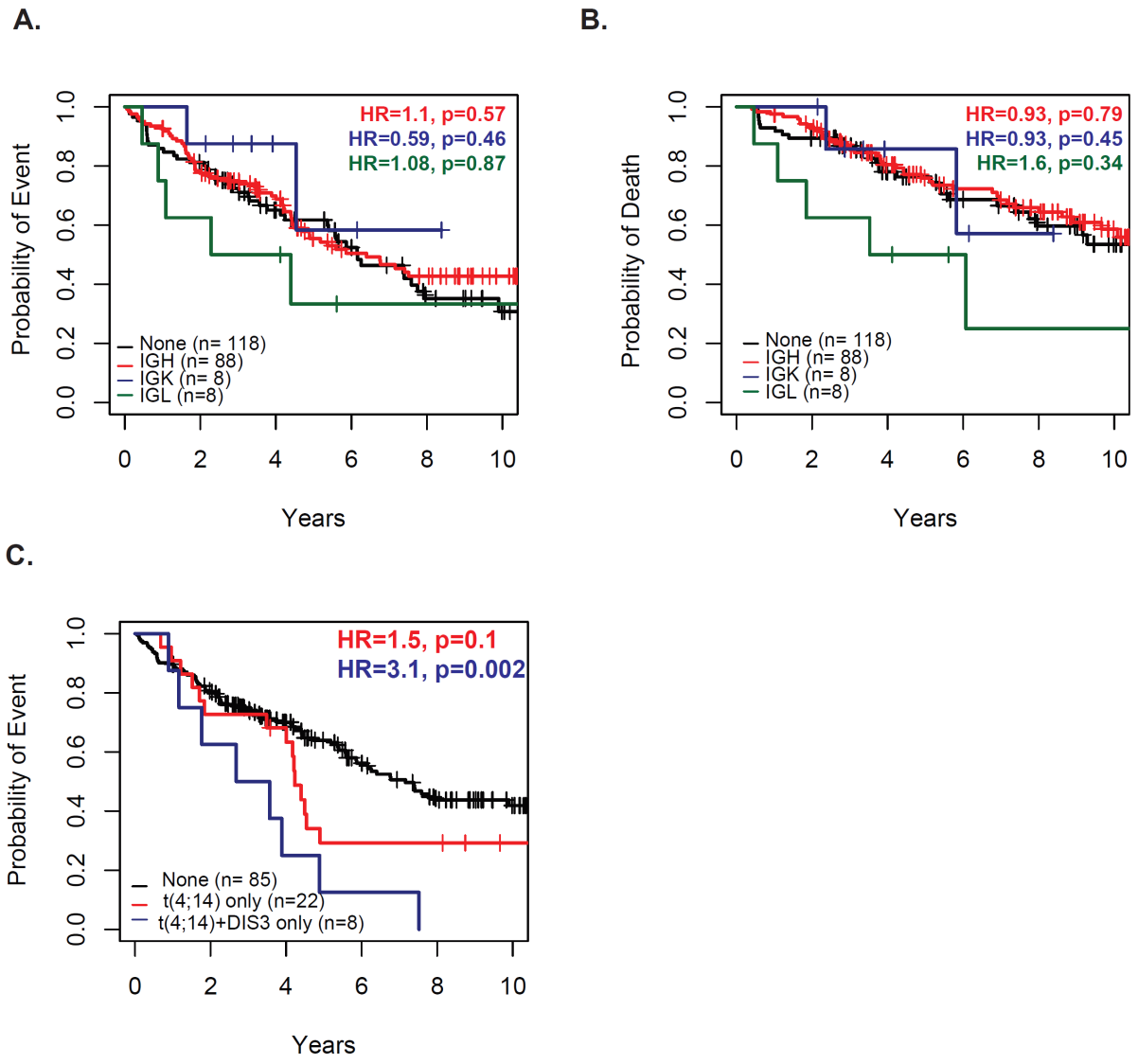


Figure 4.3.1.1.3: Impact of translocation partners. A EFS B. OS C. Impact of t(4;14) depending on the presence of a DIS3 mutation on EFS

	EFS		OS	
Covariate	**HR (95% CI for HR)**	**p-value**	**HR (95% CI for HR)**	**p-value**
Double hit	4.6 (2.7-7.9)	<0.0001	3.5 (1.9-6.5)	<0.0001
Biallelic *TP53*	4.3 (2.4-7.7)	<0.0001	2.8 (1.4-5.6)	0.004
High GEP70	2.5 (1.6-3.9)	<0.0001	3.5 (2.1-6)	<0.0001
Bi-allelic DIS3	3.6 (1.8-7.2)	0.00033	2.2 (0.89-5.6)	0.088
ISS3	2 (1.4-3)	0.00027	3 (1.9-4.7)	<0.0001
del(17p): *TP53*	2.3 (1.5-3.6)	0.0015	1.8 (1.3-3.3)	0.034
ISS1	0.48 (0.29-0.78)	0.0036	0.26 (0.12-0.53)	0.00028
Gain or amp(1q)	1.8 (1.2-2.6)	0.0037	2.3 (1.4-3.5)	0.00043
Trisomy 9	0.58 (0.4-0.84)	0.0042	0.54 (0.35-0.85)	0.0078
Gain 8q	2.3 (1.3-4.1)	0.003	2.9 (1.5-5.5)	0.0012
TP53	2.3 (1.3-4.2)	0.0065	1.5 (0.67-3.2)	0.34
Trisomy 19	0.75 (0.63-0.9)	0.0022	0.65 (0.53-0.79)	<0.0001
BRAF	2 (1.2-3.3)	0.0094	2.7 (1.5-4.7)	0.00076
DIS3	2 (1.2-3.4)	0.0096	1.2 (0.56-2.4)	0.68
PR subgroup GEP	1.9 (1.1-3.1)	0.0096	2.5 (1.5-4.4)	0.00085
gain(1q)	1.6 (1.1-2.3)	0.02	1.9 (1.2-3)	0.0056
***MMSET* translocation**	1.8 (1.2-2.9)	0.0092	1.6 (0.94-2.8)	0.082
MF cluster	2.4 (1.2-4.7)	0.015	2.1 (0.96-4.6)	0.063
del(1p): *FAF1*	1.8 (1.1-2.8)	0.018	1.9 (1.1-3.3)	0.018
Trisomy 19: *KMT2B*	0.64 (0.44-0.94)	0.023	0.56 (0.35-0. 9)	0.016
del(1p): *CDKN2C*	1.7 (1.1-2.7)	0.023	1.9 (1.1-3.2)	0.022
del(12p): *KDM5A*	1.9 (1.1-3.5)	0.026	2.6 (1.4-4.9)	0.0024
Trisomy 19: *PRKD2*	0.66 (0.45-0.97)	0.035	0.61 (0.38-0.97)	0.036
DNA repair mutations	2 (1.1-4.1)	0.023	2.2 (1.1-4.6)	0.033
Trisomy 5	0.67 (0.46-0.98)	0.037	0.83 (0.53-1.3)	0.41
MS cluster by GEP	1.7 (1-2.7)	0.034	1.4 (0.82-2.5)	0.21
del(1p): *FAM46C*	1.5 (0.99-2.3)	0.061	1.6 (1-2.7)	0.051
del(13q): telomere	1.4 (0.97-2.1)	0.072	1.7 (1-2.6)	0.03
amp(1q)	1.9 (0.89-4.2)	0.096	2.6 (1.1-6)	0.026
Monoallelic *DIS3*	1.2 (0.85-1.8)	0.27	1.2 (0.74-1.8)	0.53
***MYC* deletion**	1.4 (0.88-2.3)	0.15	1.8 (1-3.1)	0.043
Trisomy 3	0.75 (0.51-1.1)	0.14	0.84 (0.53-1.3)	0.44
del(13q): centromere	1.3 (0.88-1.9)	0.19	1.6 (0.99-2.4)	0.056
del(17p): telomere	1.4 (0.79-2.3)	0.27	1.2 (0.6-2.3)	0.64
***MYC* translocation**	0.77 (0.49-1.2)	0.26	1.1 (0.65-1.8)	0.7
CYLD	1.4 (0.61-3.2)	0.6	1.4 (0.58-3.6)	0.59
Trisomy 15	0.87 (0.6-1.3)	0.47	0.76 (0.48-1.2)	0.23
***MYC* gain**	1.1 (0.73-1.7)	0.64	1.7 (1.1-2.7)	0.03
NRAS	0.88 (0.52-1.5)	0.65	0.98 (0.53-1.8)	0.96
Trisomy 21	0.92 (0.6-1.4)	0.17	1.2 (0.73-1.9)	0.45
ISS2	0.92 (0.63-1.3)	0.66	0.84 (0.53-1.3)	0.46
KRAS	0.99 (0.63-1.6)	0.96	1.1 (0.63-1.9)	0.79
TRAF3	0.99 (0.48-2)	0.98	1.2 (0.51-2.7)	0.7

Table 4.3.3.1.1. Univariate analysis for EFS and OS. In red all variables with p<0.05. HR=hazard ratio, CI= confidence intervals.

4.3.3.2. Multivariate Analysis Identifies Mutations of *BRAF* and *DIS3* As Independently Associated With Prognosis

We went on to perform a multivariate analysis using all the genetic features with $p<0.1$.

For EFS, a protective effect was associated with trisomy(19). An adverse association was seen for Double Hit, del(1p)(*FAF1*), t(4;14), del(12p)(*KDM5A*) and mutations of *BRAF* and *DIS3* (Corrected C-index=0.689). Similarly, for OS, trisomy(19) was associated with a positive effect, whereas an adverse association was seen with Double Hit, del(1p)(*FAF1*), del(12p)(*KDM5A*), gain8q24 (*MYC*) and mutations of *BRAF* (Corrected C-index=0.73). With the long follow-up there was consistency between markers in the multivariate analysis of EFS and OS, with the exception of *MYC* gains and mutation of *DIS3*, indicating a high reliability in the dataset. A summary of the multivariate may be found in **Figure 4.3.3.2.1 and Table 4.3.3.2.1.**

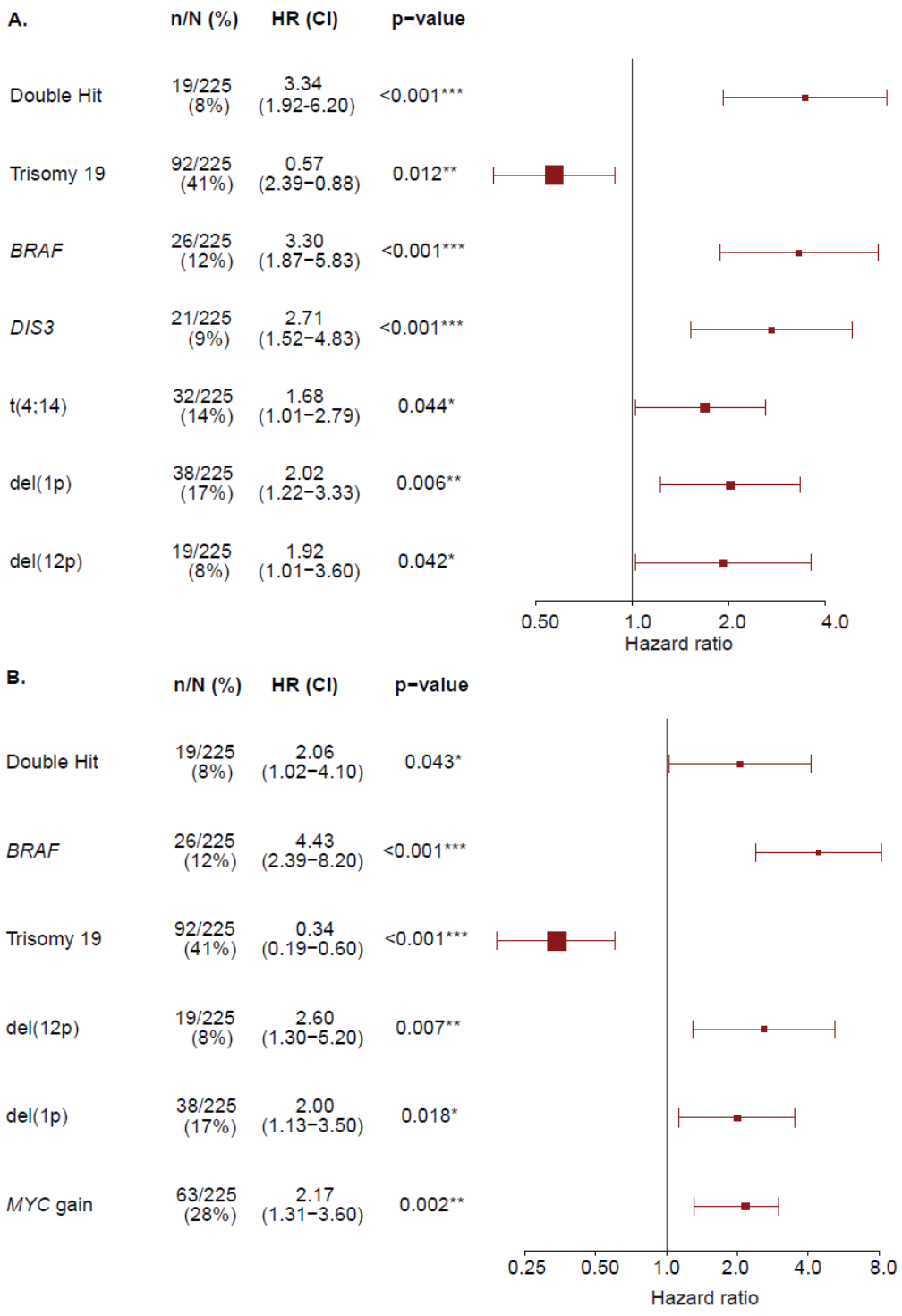


Figure 4.3.3.2.1. Multivariate analysis. Forest plots representing the results of the multivariate analysis for A. EFS and B. OS

		n/N	Coef	S.E.	Wald	Pr(>|Z|)
PFS	Double-Hit	19/223	1.2592	0.2984	4.22	<0.0001
	Trisomy 19	92/223	-0.5467	0.2208	-2.475	0.013307
	BRAF	26/223	0.6361	0.3213	1.98	0.047723
	DIS3	21/223	0.5447	0.259	2.103	0.035474
	t(4;14)	32/223	0.6983	0.2555	2.733	0.006283
	del(1p): *FAF1*	38/223	1.1822	0.2908	4.065	<0.0001
	del(12p) *KDM5A*	19/223	0.9837	0.2953	3.332	0.000864
	Concordance= 0.689 (se = 0.027), r^2= 0.223 (max possible= 0.991), Likelihood ratio test= 56.99 on 7 df, p=6e^{-10}, Wald test= 62.03 on 7 df, p=6e^{-11}, Score (logrank) test = 72.96 on 7 df, p=4e^{-13}					
OS	**Double-Hit**	**n/N**	**Coef**	**S.E.**	**Wald**	**Pr(>|Z|)**
	Double-Hit	19/223	0.7258	2.0664	0.3559	2.04
	BRAF	26/223	-1.0758	0.341	0.2853	-3.771
	Trisomy 19	92/223	0.9501	2.586	0.352	2.699
	del(12p): *KDM5A*	19/223	0.6879	1.9896	0.2912	2.363
	del(1p): *FAF1*	38/223	1.485	4.4151	0.3138	4.732
	MYC gain	63/223	0.7724	2.165	0.2555	3.024
	Concordance= 0.73 (se = 0.032) r^2= 0.22 (max possible= 0.962) Likelihood ratio test= 52.7 on 6 df, p=1e^{-09} Wald test = 55.11 on 6 df, p=4e^{-10} Score (logrank) test = 62.33 on 6 df, p=2e^{-11}					

Table 4.3.3.2.1. Multivariate analysis for EFS and OS. Multivariate model uses stepwise selection with entry level 0.1 and variable remain if meets the 0.05 level**.**

We tested the solidity of this analysis by repeating the analysis using classical risk factors and previously published models such as the IFM2009 model(Perrot *et al*, 2019) and GEP70(Zhou *et al*, 2012), **Figure 4.3.3.2.2**. *DIS3* mutations and *BRAF* mutations retained their prognostic significance, irrespective of other high-risk features.

4. Impact of mutations in NDMM

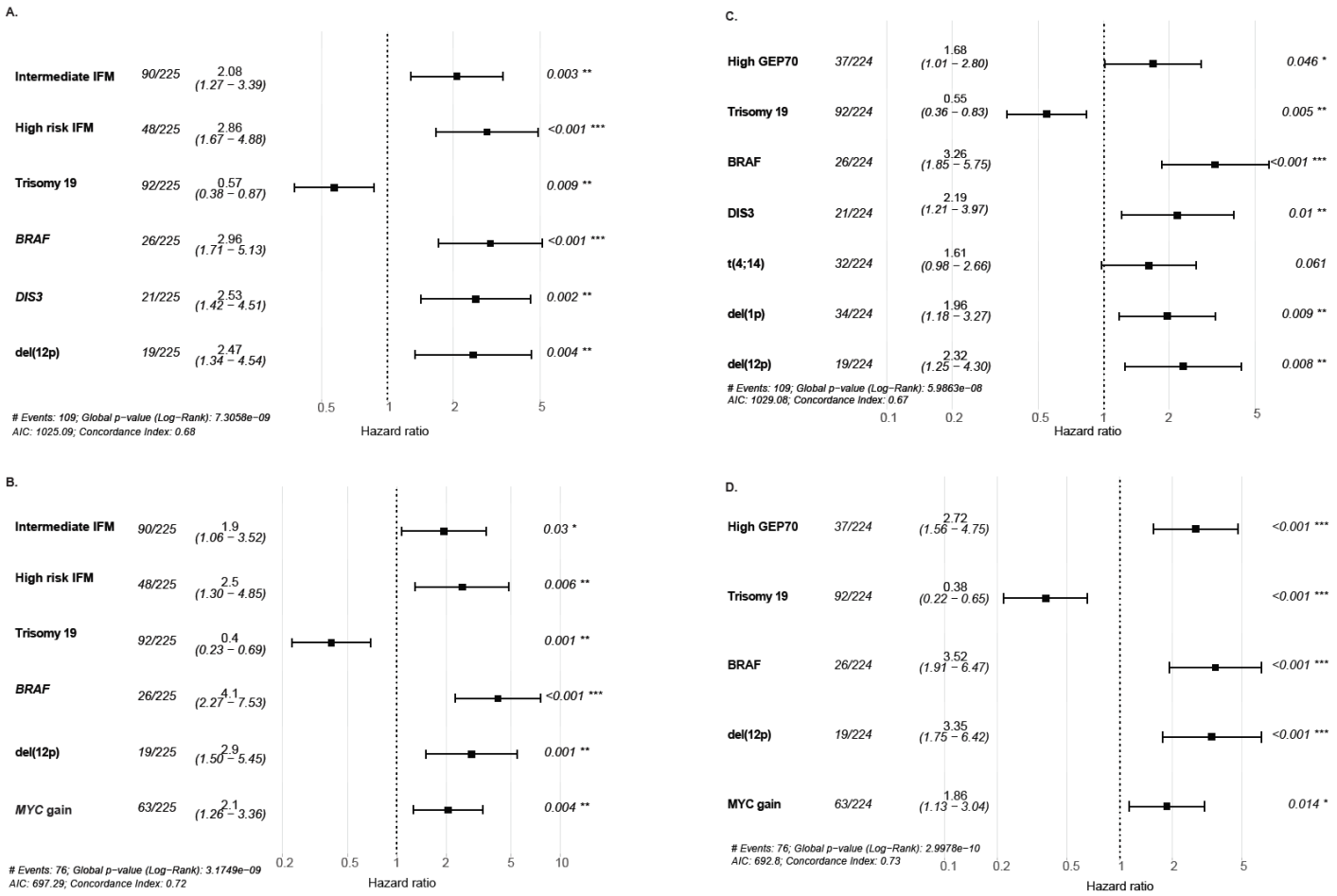


Table 4.3.3.2.2: Multivariate analysis for EFS and OS. Multivariate model uses stepwise selection with entry level 0.1 and variable remain if meets the 0.05 level**.** IFM 2009 score (figure A/B) and GEP70 (figure C/D)

4.3.4. *DIS3* mutations and biallelic *DIS3* events are associated with poor prognosis in MM

In our dataset we identified 21 patients (9.4%) with a *DIS3* mutation. The majority of the mutations were missense and were located throughout the gene suggesting they were inactivating although a hotspot is present at amino acid R780. Mutations in *DIS3* were associated with a worse EFS (HR=2 (1.2-3.4), p=0.01) but not OS. A similar trend was seen in the Myeloma XI dataset and the MGP, Biallelic events were seen in 11 (5%) patients, mostly consisting of deletions and mutations (91%, n=10/11). There was no case of biallelic deletion. Bialleleic *DIS3* events had a stronger association with EFS (HR for progression of 3.6 (1.8-7.2), p<0.0001) than monoallelic events (HR of 1.2 (0.85-1.8), p=0.27), **Figure 4.3.4.1.**

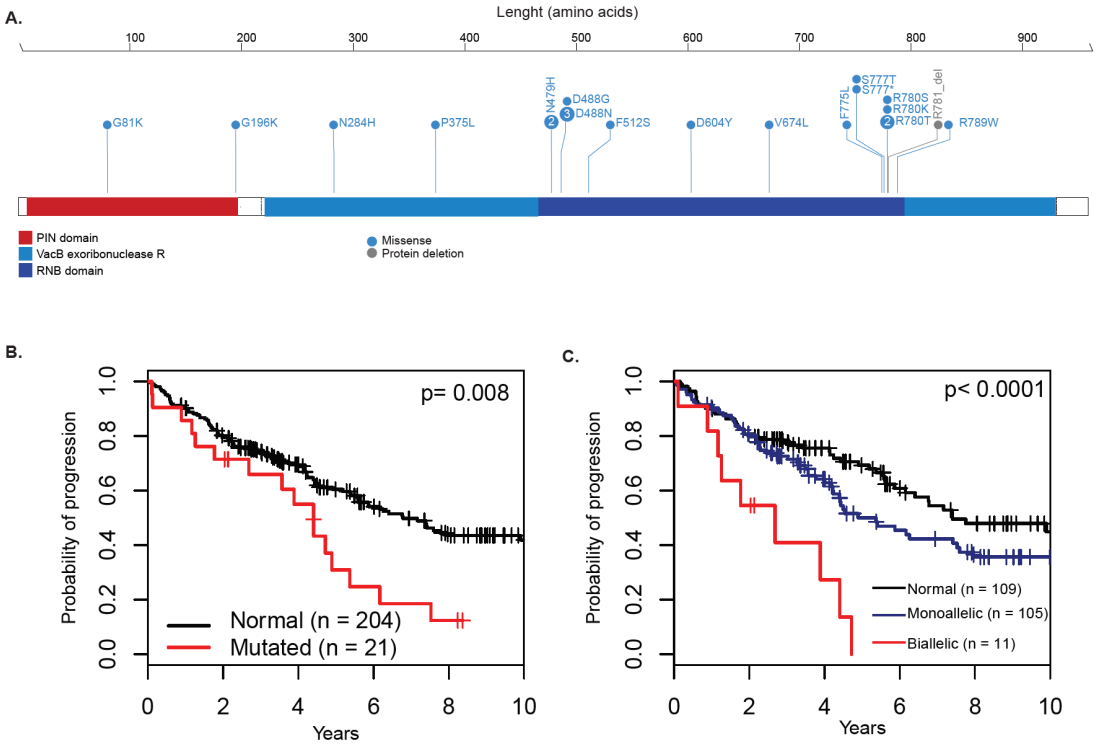


Figure 4.3.4.1. A. Distribution of DIS3 mutations throughout the gene. B. DIS3 mutations are associated with an adverse EFS. C. Biallelic DIS3 inactivation is associated with a worse outcome than monoallelic inactivation.

4.3.5. *BRAF* non-V600E mutations comprise kinase dead variants which were associated with adverse outcome, and may lead to increase MAPK activation through CRAF via co-occurring *KRAS* and *NRAS* mutations

BRAF mutations were associated with an adverse outcome in this cohort, **Figure 4.3.5.1-A-B**. Forty-six percent (n=12/26) were at the classical V600E hotspot, **Figure 4.3.5.2-A**. When comparing the impact of the non-V600E versus the V600E patients, for EFS especially, most of the prognostic impact appeared to driven by non-V600E mutations, (1.3 years (0.58-∞) versus 5.6 years (2.46-∞), p=0.02 for EFS; 3.12 years (1.21-∞) versus 8.62 years (5.73-∞), p=0.08 for OS), **Figure 4.3.5.1-C-D**.

From a functional perspective, *BRAF* mutations can be sub-divided into activating or non-activating, based on information from other cancers. Using the CKB database, we were able to dissect the non-V600E mutations into activating (n=5), inactivating (n=8), and unknown (n=1) **Table 4.3.5.2.** The outcome of patients with the inactivating mutations was worse (HR=6.4 (2.74-15), p<0.0001) than those who had an activating mutation (HR=2.1 (1.05-4.2), p=0.04), which in turn was worse than those who did not have a *BRAF* mutation, **Figure 4.3.5.2-B-C**. Similar trends were confirmed in the MGP dataset subset of patients who received an autologous stem cell transplant (OS, p=0.008), but not the Myeloma XI cohort, **Figure 4.3.5.3.**

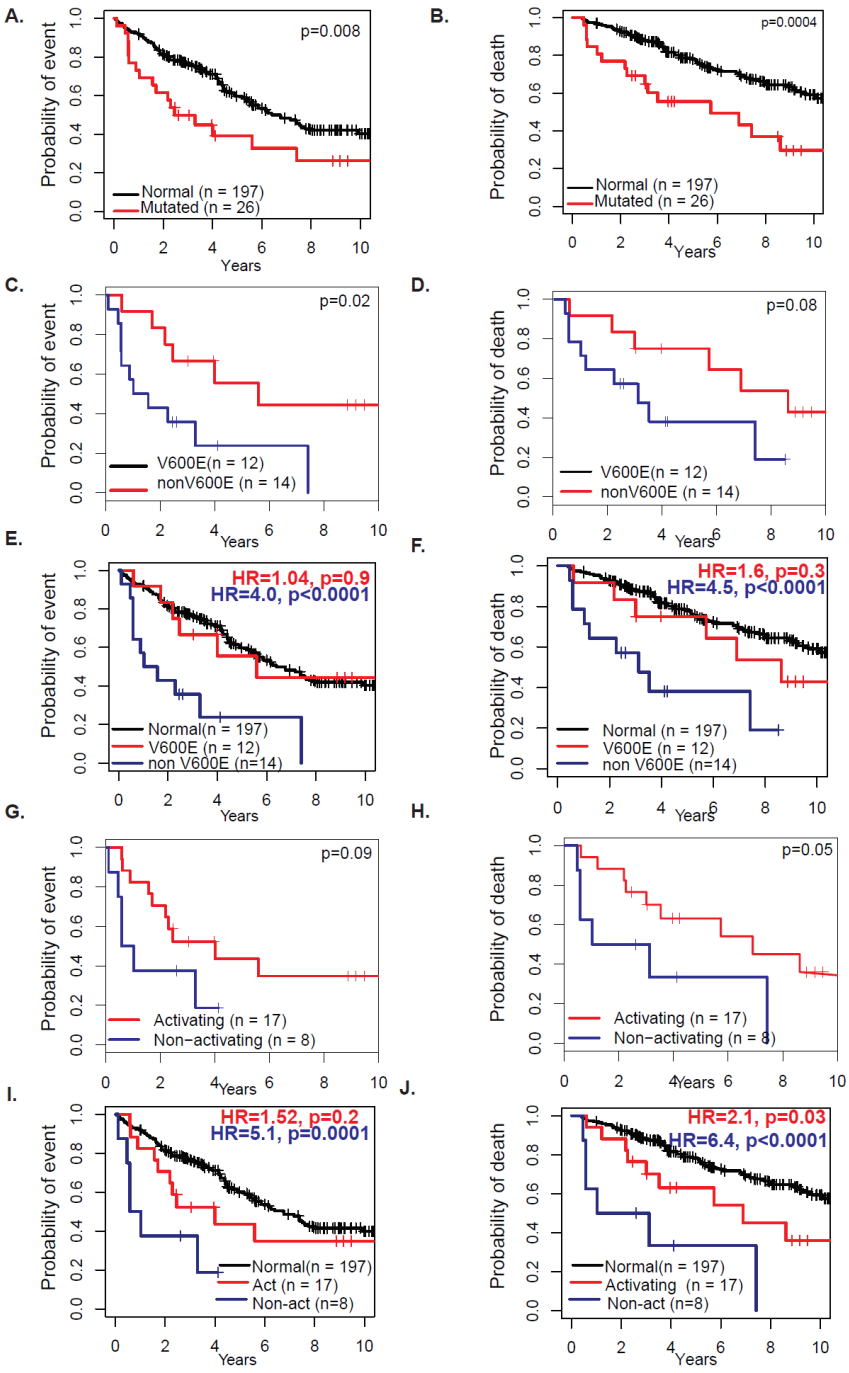


Figure 4.3.5.1 Impact of BRAF mutations on outcome. A. Impact of BRAF mutations on EFS (A) and OS (B). Differential impact of V600E and non-V600E mutations on EFS (C-E) and OS (D-F) Differential impact based on predicted BRAF function on EFS (G-I) and OS (H-J).

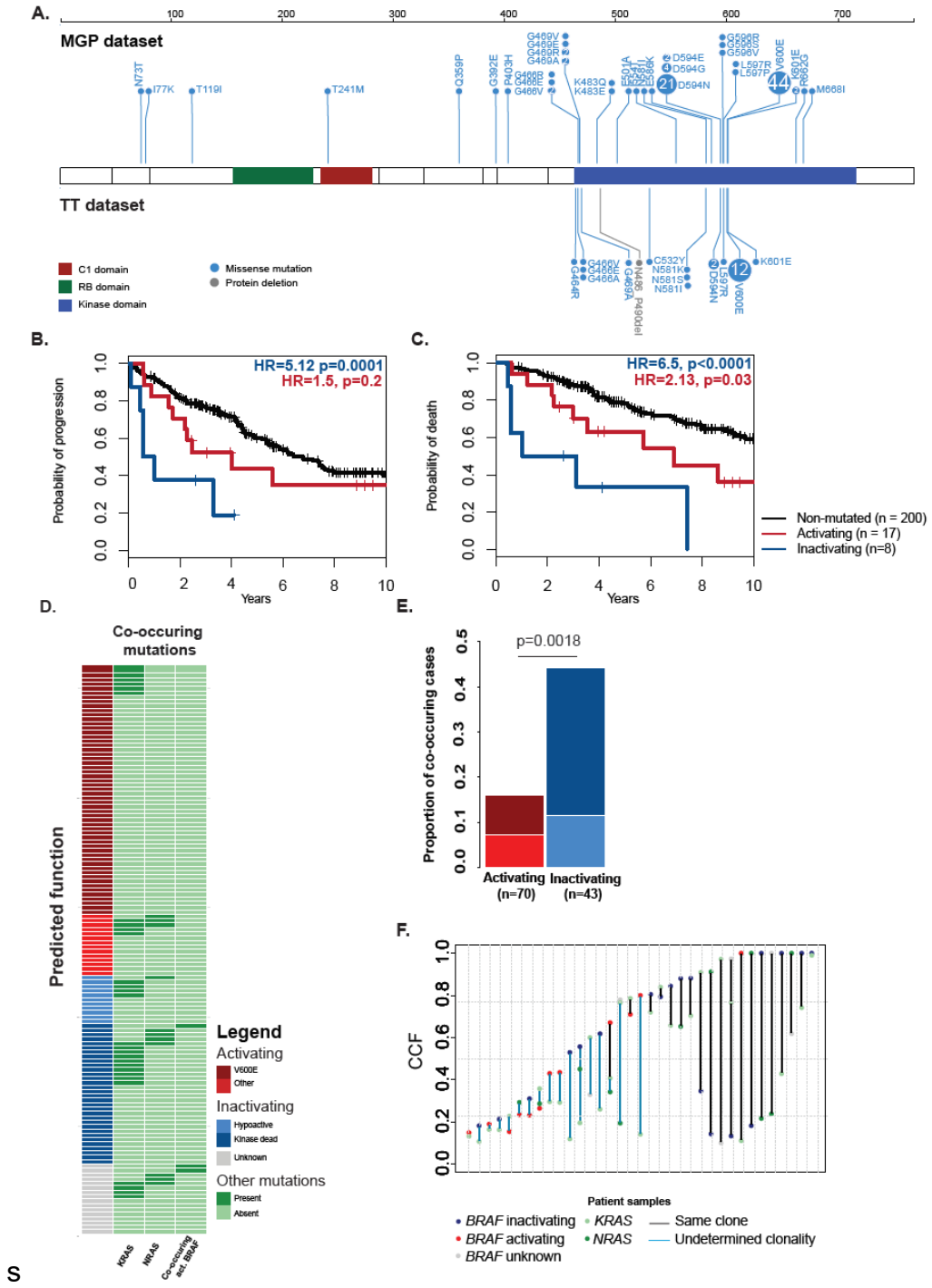


Figure 4.3.5.2. Inactivating *BRAF* mutations affect outcome and co-occur with *NRAS* or *KRAS* mutations. A. Stick plot representing the locations of the different *BRAF* mutations in the MGP dataset (above) and this dataset (below). Differential impact of *BRAF* mutations depending on predicted function on EFS (B) and OS (C). D. The spectrum of *BRAF* mutations with co-occurring mutations. E. The proportion of cases with co-occurring

S

Protein	Description
E586K	*BRAF* E586K lies within the protein kinase domain of the BRAF protein (UniProt.org). E586K results in increased BRAF kinase activity, and activation of Mek and Erk in cell culture (PMID: 15035987, PMID: 22510884), and increases cell proliferation and viability compared to wild-type *BRAF* in one of two cell lines (PMID: 29533785).
G464R	*BRAF* G464R lies within the protein kinase domain of the BRAF protein (UniProt.org). G464R results in increased *BRAF* kinase activity, increased downstream Erk signaling (PMID: 15046639), and induces cell proliferation and cell viability in culture (PMID: 29533785).
G469A	*BRAF* G469A is a hotspot mutation within the protein kinase domain of the *BRAF* protein (UniProt.org). G469A results in increased *BRAF* kinase activity and downstream activation of Erk, and is transforming in cell culture (PMID: 19010912, PMID: 12068308, PMID: 29533785).
G469V	*BRAF* G469V is a hotspot mutation within the protein kinase domain of the BRAF protein (UniProt.org). G469V results in increased BRAF kinase activity and activation of downstream MEK and ERK in cell culture (PMID: 28947956, PMID: 26343582, PMID: 28783719), and in one of two cell lines, increased cell proliferation and cell viability compared to wild-type BRAF (PMID: 29533785).
K601E	*BRAF* K601E lies within the activation segment in the kinase domain of the BRAF protein (PMID: 15343278). K601E results in increased BRAF kinase activity and downstream activation of MEK and ERK in cell culture (PMID: 22798288, PMID: 28783719) and induces cell proliferation and cell viability in culture (PMID: 29533785).
L597R	*BRAF* L597R lies within the protein kinase domain of the BRAF protein (UniProt.org). L597R results in activation of BRAF as indicated by increased phosphorylation of Mek and Erk in cell culture (PMID: 22798288, PMID: 26343582), is associated with Erk activation in a patient tumour sample (PMID: 23715574), and in one of two cell lines, increased cell proliferation and cell viability compared to wild-type BRAF (PMID: 29533785).
N486_P490del	*BRAF* N486_P490del results in the deletion of five amino acids near the alphaC-helix region of the kinase domain (PMID: 26732095). N486_P490del confers a gain of function to the BRAF protein as indicated by activation of the MAPK signalling pathway and increased cell proliferation in culture (PMID: 26732095).
V600E	*BRAF* V600E lies within the activation segment of the kinase domain of the BRAF protein (PMID: 15035987). V600E confers a gain of function to the BRAF protein as demonstrated by increased BRAF kinase activity, downstream signalling, and the ability to transform cells in culture (PMID: 15035987, PMID: 29533785).
G469R	*BRAF* G469R is a hotspot mutation within the protein kinase domain of the BRAF protein (UniProt.org). G469R demonstrates intermediate BRAF kinase activity (PMID: 28783719) and results in constitutive ERK activation in cell culture (PMID: 24920063), and in one of two cell lines leads to increased cell proliferation and cell viability compared to wild-type *BRAF* (PMID: 29533785), and is therefore predicted to confer a gain of function to the BRAF protein.

Table 4.3.5.1. Justification of activating classification of BRAF mutations.

Protein	Description
D594E	BRAF D594E lies within the protein kinase domain of the Braf protein (UniProt.org). D594E results in impaired Braf kinase activity, however, results in increased Mek and Erk phosphorylation in the presence of CRAF in cell culture (PMID: 28783719).
G466V	BRAF G466V lies within the protein kinase domain of the Braf protein (UniProt.org). G466V results in impaired Braf kinase activity, but paradoxically activates MEK and ERK through transactivation of CRAF in cell culture (PMID: 22649091, PMID: 28783719), and in one of two cell lines, G466V decreased cell proliferation and cell viability as compared to wild-type Braf (PMID: 29533785).
D594N	BRAF D594N lies within the protein kinase domain of the Braf protein (UniProt.org). D594N results in impaired Braf kinase activity, but leads to activation of Erk signalling through CRAF in cell culture (PMID: 28783719), and demonstrates decreased transforming ability compared to wild-type Braf in one of two cell lines in culture (PMID: 29533785), and therefore is predicted to confer a loss of function to the Braf protein.
G466R	BRAF G466R (previously reported as G465R) lies within the glycine-rich loop in the protein kinase domain of the Braf protein (PMID: 14681681). G466R results in impaired Braf kinase activity, but activates Erk signalling in cell culture (PMID: 15046639), and in one of two cell lines, G466R decreased cell proliferation and cell viability compared to wild-type Braf (PMID: 29533785), and is therefore predicted to confer a loss of function
G596R	BRAF G596R lies within the protein kinase domain of the Braf protein, within the DFG motif (PMID: 19735675). G596R results in impaired Braf kinase activity and decreased Mek and Erk phosphorylation, including in the presence of BRAF V600E, is not transforming in culture and does not promote tumour formation in mouse models, but results in activation of Erk in the presence of CRAF (PMID: 19735675, PMID: 28783719), however, in another study demonstrates similar cell proliferation and viability levels to wild-type Braf (PMID: 29533785), and is predicted to confer a loss of function to the Braf protein.
G596V	BRAF G596V lies within the protein kinase domain of the Braf protein (UniProt.org). G596V results in impaired Braf kinase activity and does not activate downstream MEK and ERK in cell culture (PMID: 16439621), but leads to activation of ERK in zebrafish models (PMID: 19376813), and is therefore predicted to lead to a loss of function.
N581K	BRAF N581K lies within the protein kinase domain of the Braf protein (UniProt.org). N581K has not been biochemically characterized, but demonstrates decreased transformation ability in cell culture (PMID: 29533785), and is therefore predicted to confer a loss of function to the Braf protein.
D594G	BRAF D594G lies within the protein kinase domain of the Braf protein (UniProt.org). D594G has been demonstrated to result in impaired Braf kinase activity, but leads to increased activation of Erk signalling through CRAF in cell culture (PMID: 18794803, PMID: 28783719), however, has increased transforming ability in one of two cell lines in culture (PMID: 29533785), and therefore its effect on Braf protein function is unknown.
G466A	BRAF G466A (also reported as G465A) lies within the protein kinase domain of the Braf protein (UniProt.org). The functional effect of G466A on Braf is unclear as it has been characterized both as having intermediate Braf kinase activity (PMID: 15035987) and low Braf kinase activity (PMID: 28783719), leads to Ras-dependent activation of downstream Erk in cell culture (PMID: 28783719), and in one of two cell lines increased cell proliferation and cell viability compared to wild-type Braf (PMID: 29533785).
G466E	BRAF G466E lies within the protein kinase domain of the Braf protein (UniProt.org). G466E results in impaired Braf kinase activity, but paradoxically increases Erk signalling through C-raf activation in cell culture and Xenopus embryos (PMID: 15035987).
G469E	BRAF G469E is a hotspot mutation within the protein kinase domain of the Braf protein (UniProt.org). The functional effect of G469E on Braf is unclear as it has been characterized as having both low Braf kinase activity (PMID: 28783719) and intermediate Braf kinase activity (PMID: 15035987), results in Ras-dependent activation of ERK signaling in cell culture (PMID: 28783719), and in one of two cell lines, G469E increased cell proliferation and cell viability as compared to wild-type Braf (PMID: 29533785).
N581I	BRAF N581I lies within the protein kinase domain of the Braf protein (UniProt.org). N581I results in low Braf kinase activity and Ras-dependent activation of Erk signalling in cell culture (PMID: 28783719) but, induces similar cell proliferation and cell viability as wild-type Braf (PMID: 29533785).
N581S	BRAF N581S lies within the protein kinase domain of the Braf protein (UniProt.org). N581S has been demonstrated to result in intermediate Braf kinase activity (PMID: 15035987), as well low Braf kinase activity (PMID: 28783719), and results in Ras-dependent activation of ERK signalling in cell culture (PMID: 28783719), however in another study, N581S demonstrated increased transformation ability in one of two different cell lines as compared to wild-type Braf (PMID: 29533785).

Table 4.3.5.2. Justification of inactivating classification of BRAF mutations.

Protein	Description
G596S	BRAF G596S lies within the protein kinase domain of the Braf protein (UniProt.org). G596S has not been biochemically characterized, but results in increased transformation ability compared to wild-type Braf in one of two different cell lines in culture (PMID: 29533785).
K483E	BRAF K483E lies within the protein kinase domain of the Braf protein (UniProt.org). K483E has not been biochemically characterized, but results in increased transformation ability compared to wild-type Braf in one of two cell lines in culture (PMID: 29533785).
L597P	BRAF L597P lies within the protein kinase domain of the Braf protein (UniProt.org). L597P has been identified in sequencing studies (PMID: 24714776), but has not been biochemically characterized and therefore, its effect on Braf protein function is unknown.
C532Y	NA
E501A	NA
G392E	NA
I554T	NA
I77K	NA
K483Q	NA
M668I	NA
N73T	NA
P403H	NA
Q359P	NA
R662G	NA
T119I	NA
T241M	NA

Table 4.3.5.3. Justification of undetermined classification of *BRAF* mutations.

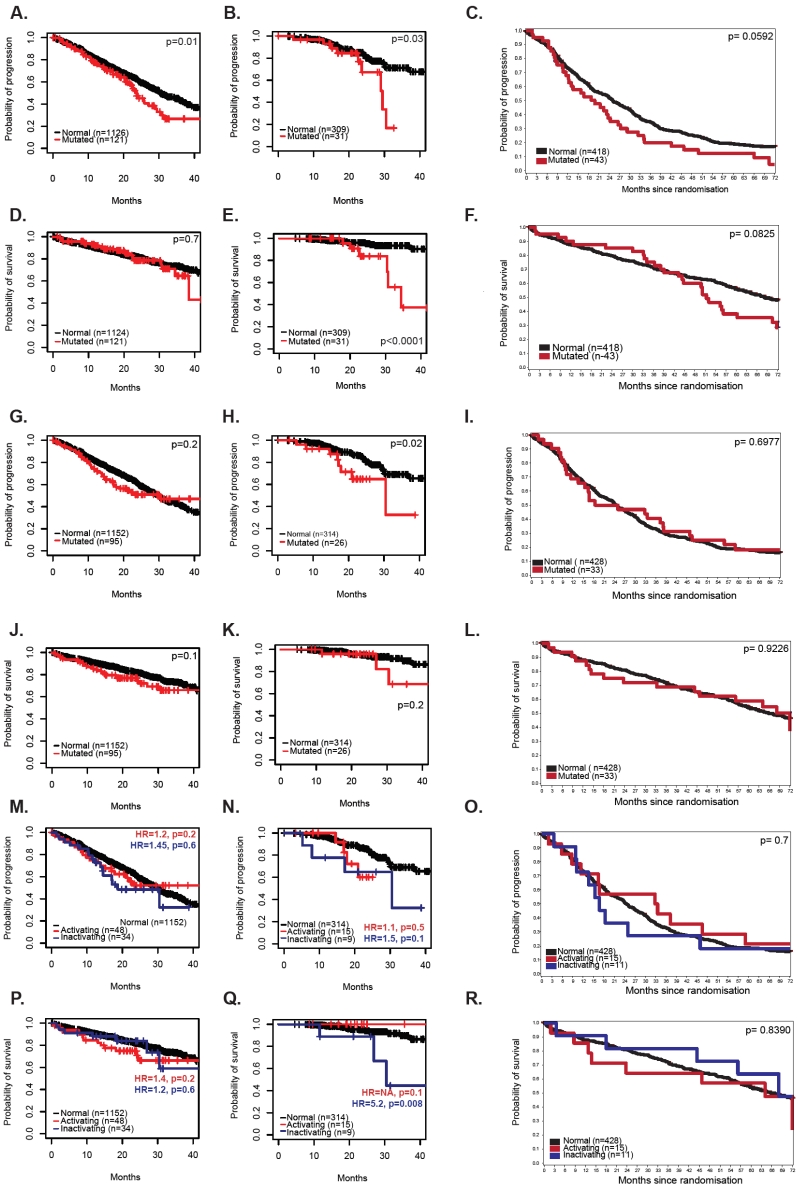


Figure 4.3.5.3. Validation of mutations in the MGP complete dataset (n=1274), MGP intensively treated patients (n=340) and Myeloma XI patients only (n=463). Impact of DIS3 mutations on PFS in the A. MGP complete dataset, B MGP intensively treated patients and C Myeloma XI patients; OS in D. MGP complete dataset, E MGP intensively treated patients and F. Myeloma XI patients. Impact of BRAF mutations on PFS in the G. MGP complete dataset, H MGP intensively treated patients and I Myeloma XI patients; OS in J. MGP complete dataset, K MGP intensively treated patients and L. Myeloma XI patients. Impact of BRAF mutation's function on PFS in the M. MGP complete dataset, N MGP intensively treated patients and O Myeloma XI patients, OS in P. MGP complete dataset, Q MGP intensively treated patients and R. Myeloma XI patients.

To explore this further, we expanded this analysis (n=26) using the MGP dataset (n=103). Combined, forty-three percent (56/129) were V600E mutations, 11% (14/129) were predicted to be activating, 8.5% with hypoactive (11/129), 25% kinase dead (32/129) and 12.5% unknown (16/129).

In melanoma inactivating mutations often co-occur with other MAPK alterations such as *NRAS or KRAS* mutations*, NF1* biallelic inactivation or *PTPN11* mutations, and contribute to increased MAPK signalling through enhanced binding and recruitment of CRAF.(Dankner *et al*, 2018) We hypothesized that the adverse outcome associated with inactivating *BRAF* mutations in myeloma is due to increased MAPK signalling, in which case co-occurring *NRAS* or *KRAS* mutations would need to be present in the same clone.

NRAS, *KRAS*, and *BRAF* mutations are believed to be mutually exclusive(Walker *et al*, 2015a, 2018b) in MM. In most cases, these three mutations were also mutually exclusive (**Figure 4.3.5.2 A-B**), however, inactivating *BRAF* mutants co-occurred more frequently with *NRAS*, *KRAS*, or activating *BRAF* mutations than expected, reaching 44% of patients with inactivating *BRAF* mutations (p=0.0018), **Figure 4.3.5.1-E**. To determine if the co-occurring mutations were present in the same clone we calculated the cancer clonal fraction (CCF) of the mutations and, based on the resulting proportions, assessed whether they were in the same clone. We identified that of the 68% (n=13/19) of samples with an inactivating *BRAF* mutation had a co-occurring *NRAS/KRAS* mutation in the same clone and 32% (n=6/19) could not be determined from the data, **Figure 4.3.5.1-F and Figure 4.3.5.3-C-D.**

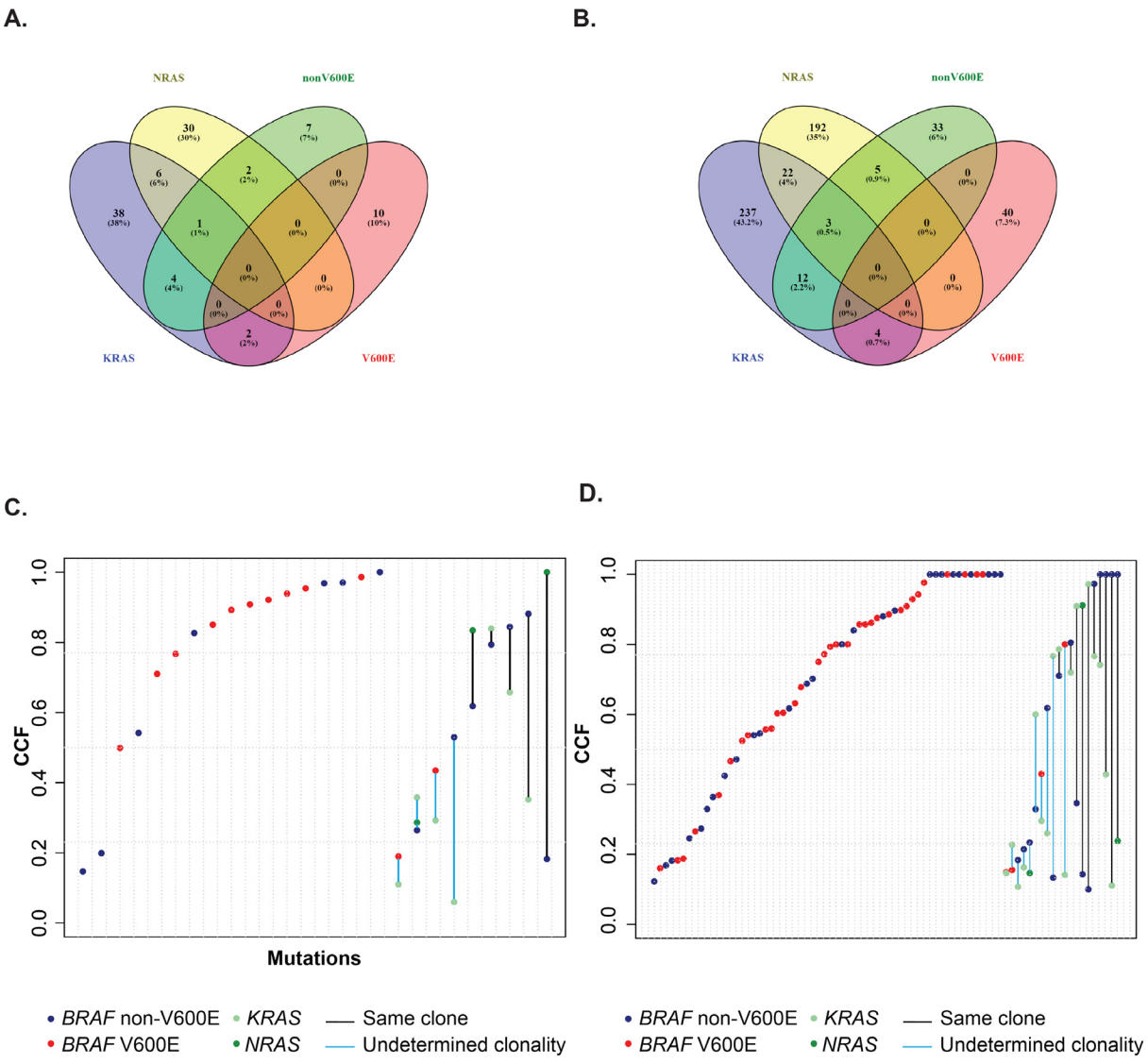


Figure 4.3.5.3 Co-segregation of *BRAF*, *KRAS*, and *NRAS* mutations. Venn diagrams representing the co-segregation of mutations in this (A) and the MGP (B) datasets. Respective CCFs of each mutation suggest at least half of them are in the same clones in this (C) and the MGP (D) datasets.

4.4. Discussion

The incorporation of the complete spectrum of genomic lesions from translocation to mutations including copy number changes is required to gain insight and accurately predict outcome in MM. Using multiple techniques to determine these factors is both

labour intensive, time consuming, and yields high failure rates given the amount of tumour cells required (Boyle *et al*, 2015). Next-generation sequencing has helped unravel the genetic complexity of MM but cost and time are often setbacks in a clinical setting. Many targeted approaches have been developed, some specific to MM (Barrio *et al*, 2018; Bolli *et al*, 2016) and some applicable to MM (He *et al*, 2016) but most do not take into account the Ig loci, thus requiring combinations with other tests such as FISH or GEP to identify translocations. Like Bolli et al,(Bolli *et al*, 2016) our approach offers a complete view of translocations, CNA, and mutations. The set of genes in this capture largely overlaps the Sanger capture but given our wide *MYC* tiling we offer a better understanding of the complex rearrangements that occur on 8q24. Finally, given the size of this capture, we are also able to identify an APOBEC signature, that has also demonstrated prognostic significance in MM (Walker *et al*, 2015b).

Here we show that an increased follow-up of patient outcome data combined with targeted sequencing can identify consistent genomic markers associated with inferior outcome. We identified the classical cytogenetic abnormalities, such as t(4;14), del(1p) and del(12p), as affecting outcome, as well as the more recently defined Double-Hit myeloma. In addition, we identify that mutations in *BRAF* and *DIS3* are prognostically implicated in the outcome of myeloma patients in an intensive setting.

In the MGP dataset we have previously shown that Double-Hit myeloma resulted in a worse outcome for PFS and OS, irrespective of the treatment used, but these patients mostly had a single autologous stem cell transplant (ASCT). ASCT is the standard of care for all NDMM aged ≤65 and before the era of novel agents ASCT proved beneficial on OS (Attal *et al*, 1996). Double transplants were subsequently investigated and deemed safe,(Harousseau *et al*, 1992) and two randomized trials confirmed the benefit of double versus single transplant in terms of OS (Attal *et al*, 1996). In this dataset, all patients received a double ASCT, but despite this Double-Hit patients still perform badly, although they have a slightly better outcome than was seen in the MGP data (20.7 months (95% CI 17.4-20.6)), with a 16-month improvement in their outcome.

BRAF mutations were seen in 11% of myeloma patients and were associated with a poor outcome. *BRAF* is mutated in numerous cancers and the substitution of a valine

(V) for a glutamic acid (E) residue at position 600 in the kinase domain is the most common *BRAF* mutation.(Cosmic) This mutation mimics the phosphorylation of the activation loop, thereby inducing constitutive BRAF kinase activation. *BRAF*^{V600E} mutations are present in 50% of melanoma patients and 2% of non-small cell lung cancer (NSCLC) patients. In NSCLC, they are associated with a shorter OS and resistance to cisplatinum chemotherapy (Marchetti *et al*, 2011). The clinical significance of *BRAF*^{V600E} in multiple myeloma has been characterized in two previous studies where seven myeloma patients with *BRAF*^{V600E} had significantly shorter OS and an increased incidence of extra medullary disease (57% vs. 17%) compared with wild-type *BRAF* (Andrulis *et al*, 2013).More recently Rustad *et al*.(Rustad *et al*, 2015) reported a good response to broad acting drugs and no relation to prognosis among eleven *BRAF*^{V600E} mutant patients.

Fifty-four percent of the *BRAF* mutations seen in this dataset were non-V600E and is a similar rate to that seen in NSCLC (Litvak *et al*, 2014; Tissot *et al*, 2016) and other MM datasets (Lionetti *et al*, 2015). The biology of these non-V600E mutants is heterogeneous, with some leading to high kinase activity (class I and II) and Ras independence while others are hypoactive or kinase dead (class III) variants but nonetheless these still impact on the MAPK pathway through CRAF heterodimerization (Wan *et al*, 2004; Haling *et al*, 2014).In our dataset the hypoinactive/kinase dead variants were more likely to co-occur with a *KRAS* or *NRAS* mutation which has also has been previously described by Lionetti et al (Lionetti *et al*, 2015). In melanoma, biallelic *NF1* inactivation or *PTPN11* activation have also been linked to MAPK activation through inactive *BRAF* mutants,(Lionetti *et al*, 2015; Richtig *et al*, 2017) but these are rare events in MM and they were not associated with non-BRAF^{V600E} mutants, **Figure 4.4.1**.

BRAF^{V600E} mutations in melanoma are sensitive to the BRAF inhibitors vemurafenib and dabrafenib. Case reports (Bohn *et al*, 2014; Sharman *et al*, 2014) and clinical trials (Hyman *et al*, 2015) also support the use of vemurafenib in this setting. These drugs are not effective against non-BRAF^{V600E} mutations, but could be targeted using MEK inhibitors.

In other cancers, the presence of concomitant NRAS/KRAS and BRAF kinase dead mutations results in chemoresistance (Litvak *et al*, 2014; Sánchez-Torres *et al*, 2013). Identifying these non-BRAF^{V600E} mutants would not only help identify patients who should not receive BRAF inhibitors but also patients who would not benefit from intensive alkylator heavy regimens. The adverse outcome of these patients in this dataset could suggest that heavily treating these patients may be deleterious and may suggest they would benefit from alkylator free regimens, thus explaining some of the outcome discrepancies in the literature (Rustad *et al*, 2015).

The prognostic impact of chromosome 13 has been long debated. Forty percent of patients have a del(13q) either by a monosomy 13 (35%) or a simple loss of 13q (Binder *et al*, 2017). Del(13q) was found to be associated with a short outcome in many studies, before the associations between t(4;14) and del(13q) were made.(Avet-Loiseau *et al*, 2003) *DIS3* is located on chromosome 13 and as such is frequently deleted, as well as being mutated in MM (Walker *et al*, 2015a, 2018b). More recently, *DIS3* germline variants have been described in familial cases of plasma cell disorders (Pertesi *et al*, 2019). Combined, we saw *DIS3* events in 53% of cases and bi-allelic events in 5%. We identified an association with poor outcome and bi-allelically affected *DIS3*, which may suggest that *DIS3* is a tumour suppressor gene. However, the majority of *DIS3* mutations are missense and not nonsense or frameshift mutations, further the mutations are clustered at particular codons which is not typical of a tumour suppressor gene and may suggest an oncogenic potential for DIS3. In this respect, the mutations may cause a change of function, as has recently been suggested in yeast where point mutations are associated with genome instability (Milbury *et al*, 2019). Given the role of DIS3 in RNA processing (Tomecki *et al*, 2014) it is possible that complete inactivation of both alleles is lethal, as has been seen for *SF3B1* (Milbury *et al*, 2019).

BRAF and *DIS3* mutations have an impact on outcome alongside classical risk markers in the context of the TT trials. We were able to identify both *BRAF*^{V600E} mutations and non-V600E *BRAF* mutations, 58% of which were predicted to be hypoactive or kinase dead. Interestingly, 44% of the hypoactive/kinase dead *BRAF* patients showed co-occurring mutations in *KRAS* or *NRAS*, suggesting they play a role in the oncogenesis of

MM by facilitating MAPK activation by upstream mutated factors through CRAF. These data highlight the importance of mutational screening to better understand NDMM and may lead to patient specific mutation-driven treatment approaches.

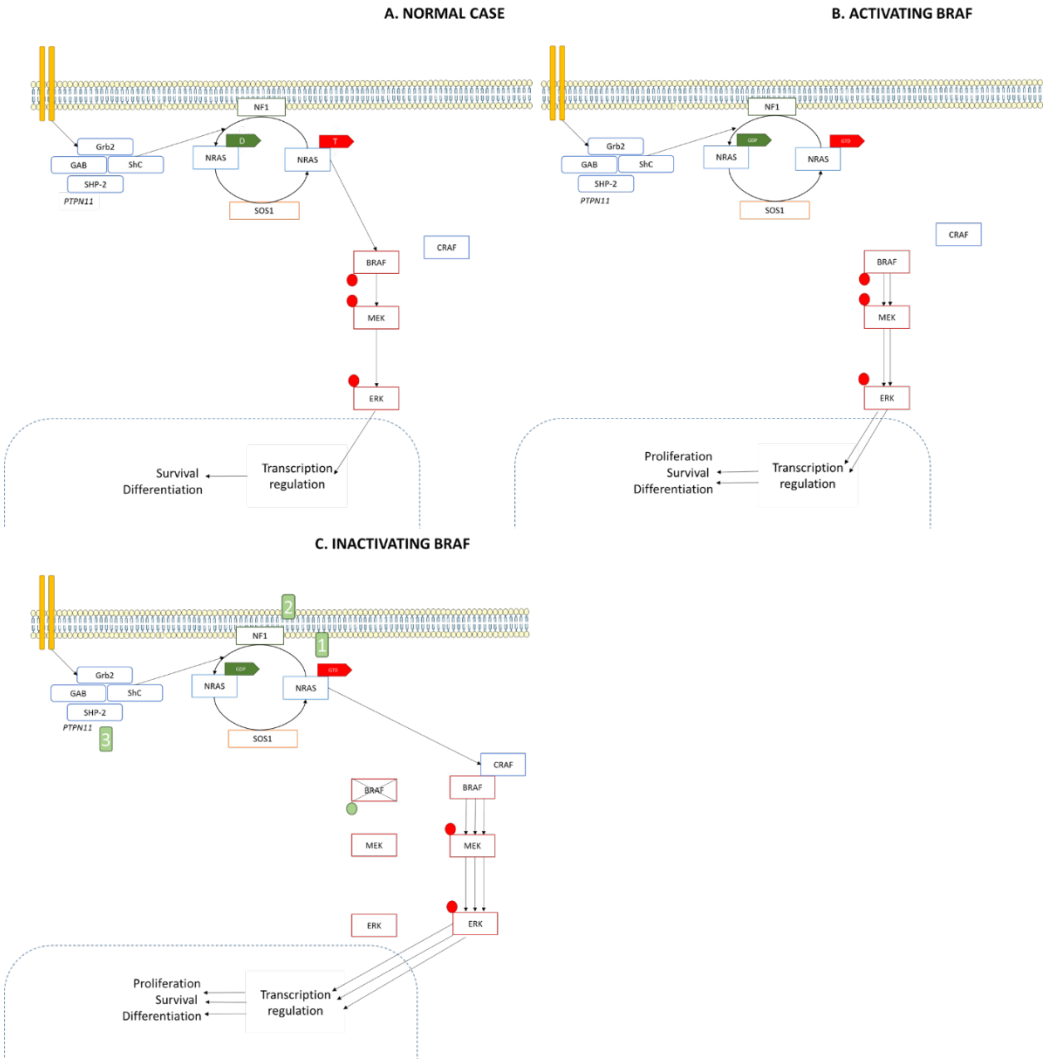


Figure 4.4.1: MAPK pathway. A. In the physiological case, upon a signal from the receptor, SHP2 proteins, NRAS or KRAS, hydrolyse to phosphorylate BRAF and activate the downstream MAPK pathway leading to survival and differentiation. This reaction is regulated by GAPS and GEFs such as RASA1 and NF1. B. In case of an activating BRAF mutation, MAPK activation becomes independent from upstream signal and regulation. C. In the case of an inactivating mutation, BRAF can dimerize with CRAF and lead to hyperactivation of the MAPK pathway, thus leading to a survival advantage. A second "hit" is believed to be required. In melanoma it has been shown to be bi-allelic loss of *NF1* or *PTPN11*, in NSCL it can be *EGFR* mutations, in both diseases like what we hypothesize in MM it is *KRAS* or *NRAS* mutations.

4.5. References

Andrulis, M., Lehners, N., Capper, D., Penzel, R., Heining, C., Huellein, J., Zenz, T., von Deimling, A., Schirmacher, P., Ho, A.D., Goldschmidt, H., Neben, K. & Raab, M.S. (2013) Targeting the BRAF V600E mutation in multiple myeloma. *Cancer Discovery*, **3,** 862–869.

Attal, M., Harousseau, J.-L., Facon, T., Guilhot, F., Doyen, C., Fuzibet, J.-G., Monconduit, M., Hulin, C., Caillot, D., Bouabdallah, R., Voillat, L., Sotto, J.-J., Grosbois, B. & Bataille, R. (2003) Single versus Double Autologous Stem-Cell Transplantation for Multiple Myeloma. *New England Journal of Medicine*, **349,** 2495–2502.

Attal, M., Harousseau, J.-L., Stoppa, A.-M., Sotto, J.-J., Fuzibet, J.-G., Rossi, J.-F., Casassus, P., Maisonneuve, H., Facon, T., Ifrah, N., Payen, C. & Bataille, R. (1996) A Prospective, Randomized Trial of Autologous Bone Marrow Transplantation and Chemotherapy in Multiple Myeloma. *New England Journal of Medicine*, **335,** 91–97.

Avet-Loiseau, H., Garand, R., Lodé, L., Harousseau, J.-L. & Bataille, R. (2003) Translocation t(11;14)(q13;q32) is the hallmark of IgM, IgE, and nonsecretory multiple myeloma variants. *Blood*, **101,** 1570–1571.

Barlogie, B., Jagannath, S., Desikan, K.R., Mattox, S., Vesole, D., Siegel, D., Tricot, G., Munshi, N., Fassas, A., Singhal, S., Mehta, J., Anaissie, E., Dhodapkar, D., Naucke, S., Cromer, J., Sawyer, J., Epstein, J., Spoon, D., Ayers, D., Cheson, B., et al (1999) Total therapy with tandem transplants for newly diagnosed multiple myeloma. *Blood*, **93,** 55–65.

Barrio, S., DáVia, M., Bruins, L., Stühmer, T., Steinbrunn, T., Bittrich, M., Einsele, H., Stewart, A.K., Braggio, E. & Kortüm, K.M. (2018) Protocol for M3P: A Comprehensive and Clinical Oriented Targeted Sequencing Panel for Routine Molecular Analysis in Multiple Myeloma. *Methods in Molecular Biology (Clifton, N.J.)*, **1792,** 117–128.

Binder, M., Rajkumar, S.V., Ketterling, R.P., Greipp, P.T., Dispenzieri, A., Lacy, M.Q., Gertz, M.A., Buadi, F.K., Hayman, S.R., Hwa, Y.L., Zeldenrust, S.R., Lust, J.A., Russell, S.J., Leung, N., Kapoor, P., Go, R.S., Gonsalves, W.I., Kyle, R.A. & Kumar, S.K. (2017) Prognostic implications of abnormalities of chromosome 13 and the presence of multiple cytogenetic high-risk abnormalities in newly diagnosed multiple myeloma. *Blood Cancer Journal*, **7,** e600.

Björkstrand, B. (2001) European Group for Blood and Marrow Transplantation Registry studies in multiple myeloma. *Seminars in Hematology*, **38,** 219–225.

Bohn, O.L., Hsu, K., Hyman, D.M., Pignataro, D.S., Giralt, S. & Teruya-Feldstein, J. (2014) BRAF V600E mutation and clonal evolution in a patient with relapsed refractory myeloma with plasmablastic differentiation. *Clinical Lymphoma, Myeloma & Leukemia*, **14,** e65-68.

Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., Hinton, J.W., Li, Y., Tubio, J.M.C., McLaren, S., O' Meara, S., Butler, A.P., Teague, J.W., Mudie, L., Anderson, E., Rashid, N., et al (2014) Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature Communications*, **5,** Available at: http://www.nature.com.gate2.inist.fr/ncomms/2014/140116/ncomms3997/full/ncomms3997.html [Accessed February 8, 2015].

Bolli, N., Biancon, G., Moarii, M., Gimondi, S., Li, Y., de Philippis, C., Maura, F., Sathiaseelan, V., Tai, Y.-T., Mudie, L., O'Meara, S., Raine, K., Teague, J.W., Butler, A.P., Carniti, C., Gerstung, M., Bagratuni, T., Kastritis, E., Dimopoulos, M., Corradini, P., et al (2017) Analysis of the genomic

landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia*.

Bolli, N., Biancon, G., Moarii, M., Gimondi, S., Li, Y., de Philippis, C., Maura, F., Sathiaseelan, V., Tai, Y.-T., Mudie, L., O'Meara, S., Raine, K., Teague, J.W., Butler, A.P., Carniti, C., Gerstung, M., Bagratuni, T., Kastritis, E., Dimopoulos, M., Corradini, P., et al (2018) Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia*.

Bolli, N., Li, Y., Sathiaseelan, V., Raine, K., Jones, D., Ganly, P., Cocito, F., Bignell, G., Chapman, M.A., Sperling, A.S., Anderson, K.C., Avet-Loiseau, H., Minvielle, S., Campbell, P.J. & Munshi, N.C. (2016) A DNA target-enrichment approach to detect mutations, copy number changes and immunoglobulin translocations in multiple myeloma. *Blood Cancer Journal*, **6,** e467–e467.

Boyle, E.M., Proszek, P.Z., Kaiser, M.F., Begum, D., Dahir, N., Savola, S., Wardell, C.P., Leleu, X., Ross, F.M., Chiecchio, L., Cook, G., Drayson, M.T., Owen, R.G., Ashcroft, J.M., Jackson, G.H., Anthony Child, J., Davies, F.E., Walker, B.A. & Morgan, G.J. (2015) A molecular diagnostic approach able to detect the recurrent genetic prognostic factors typical of presenting myeloma. *Genes, Chromosomes & Cancer*, **54,** 91–98.

Cosmic COSMIC - Catalogue of Somatic Mutations in Cancer. Available at: https://cancer.sanger.ac.uk/cosmic [Accessed June 4, 2018].

Dankner, M., Rose, A.A.N., Rajkumar, S., Siegel, P.M. & Watson, I.R. (2018) Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations. *Oncogene*, **37,** 3183.

Egan, J.B., Shi, C.-X., Tembe, W., Christoforides, A., Kurdoglu, A., Sinari, S., Middha, S., Asmann, Y., Schmidt, J., Braggio, E., Keats, J.J., Fonseca, R., Bergsagel, P.L., Craig, D.W., Carpten, J.D. & Stewart, A.K. (2012) Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood*, **120,** 1060–1066.

Haling, J.R., Sudhamsu, J., Yen, I., Sideris, S., Sandoval, W., Phung, W., Bravo, B.J., Giannetti, A.M., Peck, A., Masselot, A., Morales, T., Smith, D., Brandhuber, B.J., Hymowitz, S.G. & Malek, S. (2014) Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell*, **26,** 402–413.

Harousseau, J.L., Milpied, N., Laporte, J.P., Collombat, P., Facon, T., Tigaud, J.D., Casassus, P., Guilhot, F., Ifrah, N. & Gandhour, C. (1992) Double-intensive therapy in high-risk multiple myeloma. *Blood*, **79,** 2827–2833.

He, J., Abdel-Wahab, O., Nahas, M.K., Wang, K., Rampal, R.K., Intlekofer, A.M., Patel, J., Krivstov, A., Frampton, G.M., Young, L.E., Zhong, S., Bailey, M., White, J.R., Roels, S., Deffenbaugh, J., Fichtenholtz, A., Brennan, T., Rosenzweig, M., Pelak, K., Knapp, K.M., et al (2016) Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood***,** blood-2015-08-664649.

Hoang, P.H., Dobbins, S.E., Cornish, A.J., Chubb, D., Law, P.J., Kaiser, M. & Houlston, R.S. (2018) Whole-genome sequencing of multiple myeloma reveals oncogenic pathways are targeted somatically through multiple mechanisms. *Leukemia*, **32,** 2459–2470.

Hyman, D.M., Puzanov, I., Subbiah, V., Faris, J.E., Chau, I., Blay, J.-Y., Wolf, J., Raje, N.S., Diamond, E.L., Hollebecque, A., Gervais, R., Elez-Fernandez, M.E., Italiano, A., Hofheinz, R.-D., Hidalgo, M., Chan, E., Schuler, M., Lasserre, S.F., Makrutzki, M., Sirzen, F., et al (2015) Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. *New England Journal of Medicine*, **373,** 726–736.

Johnson, D.C., Lenive, O., Mitchell, J., Jackson, G., Owen, R., Drayson, M., Cook, G., Jones, J.R., Pawlyn, C., Davies, F.E., Walker, B.A., Wardell, C., Gregory, W.M., Cairns, D., Morgan, G.J., Houlston, R.S. & Kaiser, M.F. (2017) Neutral tumor evolution in myeloma is associated with poor prognosis. *Blood*, **130,** 1639–1643.

Lionetti, M., Barbieri, M., Todoerti, K., Agnelli, L., Marzorati, S., Fabris, S., Ciceri, G., Galletti, S., Milesi, G., Manzoni, M., Mazzoni, M., Greco, A., Tonon, G., Musto, P., Baldini, L. & Neri, A. (2015) Molecular spectrum of BRAF, NRAS and KRAS gene mutations in plasma cell dyscrasias: implication for MEK-ERK pathway activation. *Oncotarget*, **6,** 24205–24217.

Litvak, A.M., Paik, P.K., Woo, K.M., Sima, C.S., Hellmann, M.D., Arcila, M.E., Ladanyi, M., Rudin, C.M., Kris, M.G. & Riely, G.J. (2014) Clinical characteristics and course of 63 patients with BRAF mutant lung cancers. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, **9,** 1669–1674.

Lonial, S., Boise, L.H. & Kaufman, J. (2015) How I treat high-risk myeloma. *Blood*, **126,** 1536–1543.

Magrangeas, F., Avet-Loiseau, H., Gouraud, W., Lodé, L., Decaux, O., Godmer, P., Garderet, L., Voillat, L., Facon, T., Stoppa, A.M., Marit, G., Hulin, C., Casassus, P., Tiab, M., Voog, E., Randriamalala, E., Anderson, K.C., Moreau, P., Munshi, N.C. & Minvielle, S. (2013) Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*, **27,** 473–481.

Marchetti, A., Felicioni, L., Malatesta, S., Grazia Sciarrotta, M., Guetti, L., Chella, A., Viola, P., Pullara, C., Mucilli, F. & Buttitta, F. (2011) Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, **29,** 3574–3579.

Milbury, K.L., Paul, B., Lari, A., Fowler, C., Montpetit, B. & Stirling, P.C. (2019) Exonuclease domain mutants of yeast DIS3 display genome instability. *Nucleus (Austin, Tex.)*, **10,** 21–32.

Perrot, A., Lauwers-Cances, V., Tournay, E., Hulin, C., Chretien, M.-L., Royer, B., Dib, M., Decaux, O., Jaccard, A., Belhadj, K., Brechignac, S., Fontan, J., Voillat, L., Demarquette, H., Collet, P., Rodon, P., Sohn, C., Lifermann, F., Orsini-Piocelle, F., Richez, V., et al (2019) Development and Validation of a Cytogenetic Prognostic Index Predicting Survival in Multiple Myeloma. *Journal of Clinical Oncology*, **37,** 1657–1665.

Pertesi, M., Vallée, M., Wei, X., Revuelta, M.V., Galia, P., Demangel, D., Oliver, J., Foll, M., Chen, S., Perrial, E., Garderet, L., Corre, J., Leleu, X., Boyle, E.M., Decaux, O., Rodon, P., Kolb, B., Slama, B., Mineur, P., Voog, E., et al (2019) Exome sequencing identifies germline variants in DIS3 in familial multiple myeloma. *Leukemia***,** 1.

Richtig, G., Hoeller, C., Kashofer, K., Aigelsreiter, A., Heinemann, A., Kwong, L.N., Pichler, M. & Richtig, E. (2017) Beyond the BRAFV600E hotspot: biology and clinical implications of rare BRAF gene mutations in melanoma patients. *The British Journal of Dermatology*, **177,** 936–944.

Rustad, E.H., Dai, H.Y., Hov, H., Coward, E., Beisvag, V., Myklebost, O., Hovig, E., Nakken, S., Vodák, D., Meza-Zepeda, L.A., Sandvik, A.K., Wader, K.F., Misund, K., Sundan, A., Aarset, H. & Waage, A. (2015) BRAF V600E mutation in early-stage multiple myeloma: good response to broad acting drugs and no relation to prognosis. *Blood Cancer Journal*, **5,** e299.

Sánchez-Torres, J.M., Viteri, S., Molina, M.A. & Rosell, R. (2013) BRAF mutant non-small cell lung cancer and treatment with BRAF inhibitors. *Translational Lung Cancer Research*, **2,** 244–250.

Sawyer, J.R., Tian, E., Heuck, C.J., Epstein, J., Johann, D.J., Swanson, C.M., Lukacs, J.L., Johnson, M., Binz, R., Boast, A., Sammartino, G., Usmani, S., Zangari, M., Waheed, S., van Rhee, F. & Barlogie,

B. (2014) Jumping translocations of 1q12 in multiple myeloma: a novel mechanism for deletion of 17p in cytogenetically defined high-risk disease. *Blood*, **123,** 2504–2512.

SEER data (2018) Myeloma - Cancer Stat Facts. Available at: https://seer.cancer.gov/statfacts/html/mulmy.html [Accessed July 19, 2018].

Sharman, J.P., Chmielecki, J., Morosini, D., Palmer, G.A., Ross, J.S., Stephens, P.J., Stafl, J., Miller, V.A. & Ali, S.M. (2014) Vemurafenib response in 2 patients with posttransplant refractory BRAF V600E-mutated multiple myeloma. *Clinical Lymphoma, Myeloma & Leukemia*, **14,** e161-163.

Tissot, C., Couraud, S., Tanguy, R., Bringuier, P.-P., Girard, N. & Souquet, P.-J. (2016) Clinical characteristics and outcome of patients with lung cancer harboring BRAF mutations. *Lung Cancer (Amsterdam, Netherlands)*, **91,** 23–28.

Tomecki, R., Drazkowska, K., Kucinski, I., Stodus, K., Szczesny, R.J., Gruchota, J., Owczarek, E.P., Kalisiak, K. & Dziembowski, A. (2014) Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Research*, **42,** 1270–1290.

Walker, B.A., Boyle, E.M., Wardell, C.P., Murison, A., Begum, D.B., Dahir, N.M., Proszek, P.Z., Johnson, D.C., Kaiser, M.F., Melchor, L., Aronson, L.I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J.R., Brioli, A., Mikulasova, A., Cairns, D.A., Gregory, W.M., Quartilho, A., et al (2015a) Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018a) A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018b) Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*, **132,** 587–597.

Walker, B.A., Wardell, C.P., Murison, A., Boyle, E.M., Begum, D.B., Dahir, N.M., Proszek, P.Z., Melchor, L., Pawlyn, C., Kaiser, M.F., Johnson, D.C., Qiang, Y.-W., Jones, J.R., Cairns, D.A., Gregory, W.M., Owen, R.G., Cook, G., Drayson, M.T., Jackson, G.H., Davies, F.E., et al (2015b) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nature Communications*, **6,** 6997.

Wan, P.T.C., Garnett, M.J., Roe, S.M., Lee, S., Niculescu-Duvaz, D., Good, V.M., Jones, C.M., Marshall, C.J., Springer, C.J., Barford, D., Marais, R. & Cancer Genome Project (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*, **116,** 855–867.

Zhou, Y., Zhang, Q., Stephens, O., Heuck, C.J., Tian, E., Sawyer, J.R., Cartron-Mizeracki, M.-A., Qu, P., Keller, J., Epstein, J., Barlogie, B. & Shaughnessy, J.D. (2012) Prediction of cytogenetic abnormalities with gene expression profiles. *Blood*, **119,** e148-150.

Chapter 5: The Molecular Structure of Smouldering Myeloma Highlights Evolutionary Trajectories Leading to Myeloma

5.1. Summary

To identify the molecular keys of progression of smouldering myeloma (SMM), we studied 82 patients with both targeted sequencing that included a capture of the immunoglobulin and *MYC* regions together with ultra-low-pass whole genome sequencing. We compared these results to newly diagnosed myeloma (MM) patients. We found fewer *NRAS* and *FAM46C* mutations together with fewer adverse translocations, del(1p), del(14q), del(16q), and del(17p) in SMM consistent with their role as drivers of the transition to MM. There were fewer biallelic events in SMM than in MM. We performed an nNMF analysis to identify the signature makeup of these samples and show that SMM samples had less APOBEC mutations consistent with it being associated with disease progression. *KRAS* mutations were associated with a shorter time to progression (HR 3.5 (1.5-8.1), p= 0.001). In a unique analysis of change in clonal structure over time, we studied 53 samples from nine patients at multiple time points. Branching evolutionary patterns, novel mutations, biallelic hits in crucial tumour suppressor genes, and segmental copy number changes are key mechanisms underlying the transition to MM, which can precede progression and may be used to guide early intervention strategies.

5.2. Introduction

Smouldering multiple myeloma (SMM) is an asymptomatic plasma cell disorder, the behaviour of which is distinguished from monoclonal gammopathy of undetermined significance (MGUS) by a higher rate of progression to symptomatic multiple myeloma (MM)(Rajkumar *et al*, 2011; Larsen *et al*, 2013; Hillengass *et al*, 2010). The nature of SMM as a discreet disease entity has been called into question because of the triphasic

shape of the Kaplan-Meier curves associated with its progression to MM (Landgren, 2017). These curves have been interpreted as being consistent with a low-risk group similar to MGUS, a high-risk group “early myeloma” who progress quickly, and an intermediate risk group the nature of which is uncertain.

The International Myeloma Working Group (IMWG) definitions of SMM (International Myeloma Working Group, 2003) adapted to this idea by taking into account additional risk stratification factors such as serum free light chain (SFLC) ratio (Sørrig *et al*, 2016; Pérez-Persona *et al*, 2007), tumour burden in the bone marrow (Rajkumar *et al*, 2014), and the presence of bone lesions on magnetic resonance imaging (MRI) to improve the disease definition (Siontis *et al*, 2015; Zamagni *et al*, 2016; Hillengass *et al*, 2010). Using these criteria, a group associated with an 80% chance of progression at two years was recognised and redefined as early myeloma (EM), suitable for therapeutic intervention (Rajkumar *et al*, 2014). However, the newly defined residual SMM entity still retains significant variation in the time to progression to MM, the molecular basis for which is uncertain. Efforts to recognize acquired genetic factors that can explain this variation focused on those recognised in MM including del(17p), t(4;14), *MYC* translocations (MMRF CoMMpass Network *et al*, 2020), gain(1q)(Rosiñol *et al*, 2005; Rajkumar *et al*, 2013), and GEP risk scores (Khan *et al*, 2015).

The application of next generation sequencing allowed us to better understand the molecular basis of MM and to define a limited set of disease drivers. However, until recently this technology has not been widely applied to the asymptomatic phases of MM (MMRF CoMMpass Network *et al*, 2020; Bolli *et al*, 2018; Bustoros *et al*, 2020). It has been suggested based on current data that the majority of the genetic changes necessary to give rise to MM are present at the SMM stage and that there is either no change or limited changes in the sub-clonal architecture in the transition of SMM to MM (Bolli *et al*, 2018; Walker *et al*, 2014). In contrast, in the symptomatic stages of MM spatial-temporal genetic variation has clearly been demonstrated consistent with an important role for evolutionary mechanisms in the development of MM. However, currently there are little data available for us to understand the processes active during the early evolutionary stages of disease when the so-called “trunk” of the MM “tree” is developing, which lead

to the heterogeneous evolutionary “branches” of the disease that underlie variation in outcome (Rasche *et al*, 2017).

To understand in more detail the genetics underlying these asymptomatic precursor stages of disease and their contribution to risk stratification we have carried out an analysis of eighty-two SMM cases. We combined this cross sectional with a study of multiple sequential samples derived from the same patients. The aim of the study is to determine differences in terms of genetic makeup and their impact on the sub-clonal structure from smouldering phases to progression to MM.

5.3. Results

5.3.1. Identifying Significant Genomic Differences Between SMM and MM

Genetic abnormalities that are more frequent in MM have been suggested to be drivers of progression (Walker *et al*, 2018). To guarantee comparability in terms of mutation detection, given the low tumour burden in SMM and the possible contamination by normal plasma-cells, a targeted sequencing approach was used to ensure greater depth of coverage of crucial regions. Thus, in order to determine the spectrum of genomic differences characterizing SMM we performed targeted sequencing of a set of 82 previously untreated SMM patients and compared the results to a the previously analysed (Chapter 4) dataset of 223 MM patients assessed using the same sequencing approach, **Table 5.3.1.1**.

Characteristic	SMM (n=82)	MM (n=223)
Median age at diagnosis (range)	63 years (37-85)	59 years (30-75)
Median age at sample collection (range)	65 years (38-85)	59 years (30-75)
Median Follow-up (95% CI)	5.18 years (95%CI 3.53-6.89)	8.14 years (95% CI 7.35-8.96).
5-year Progression Free Survival (95% CI) to symptomatic MM	69.9% (95%CI 58-83%)	NA
5-year Overall Survival (95% CI)	96.5% (95%CI 92-100%)	6.17 years (95% CI 5.18-7.75)
Time point	Diagnosis: 44% Follow-up: 56% after a median time 2.9 years from diagnosis	Diagnosis: 100%
Gender ratio (Female: Male)	1:1.15	1:1.8
Ethnicity	24% African-American 71% White-Caucasian 5% Other (Native American, Pacific Islander or unknown)	10% African-American 88% White-Caucasian 2% Other (Native American, Pacific Islander or unknown)
Isotype	IgA: 30% (n=25) IgG: 60% (n=49) IGD: 1.2% (n=1) LCO: 8% (n=7)	

Table 5.3.1.1.1. Summary of patient's characteristics

5.3.1.1. Copy Number Abnormalities and Translocation Frequency Differs Between SMM and MM.

We first looked at translocations to the *IGH* and *MYC* region and copy number changes. Overall, 35% (n=29/82) of SMM samples had a canonical translocation, a frequency that is identical to that seen in MM, 37% (n=82/223, non-Yates adjusted, χ^2=0.008, p=0.82). The most common translocation in the SMM dataset was the t(11;14) seen in 23% of cases (n=19/82) followed by the t(4;14) (4.9%, n=4/82), t(6;14) (3.7%, n=3/82), t(14;16) (2.4%, n=2/82). Two samples were

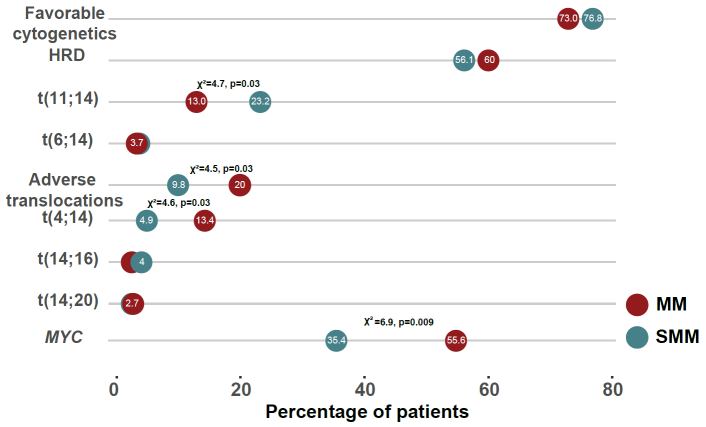


Figure 5.3.1.1.1. Frequency of cytogenetic subgroups and main translocations in SMM and MM suggesting that fewer SMM have high risk features.

identified with a *MAFB* translocation [one t(14;20) and one t(20;22)]. There were significantly fewer t(4;14) among the SMM patients (non-Yates adjusted χ^2=4.4, p=0.03) and more t(11;14) (non-Yates adjusted χ^2=4.6, p=0.03), suggesting that SMM carried fewer adverse risk cytogenetic groups [t(4;14), t(14;16), t(14;20)] than MM (9.8% versus 20%, non-Yates adjusted χ^2=4.5, p=0.03), **Figure 5.3.1.1.1.**

The genomic partners of translocations and secondary rearrangements detected at the *MYC* locus were similar in the two datasets but differed in frequency. Rearrangements involving *MYC* were seen in 35% (n=29/82) of SMM cases, which is fewer than were found in MM (55%, n=124/223, χ^2=6.9, p=0.009). Fifty-five percent of the rearrangements (n=16/29) involved a translocation with the remainder being either duplications, inversions, or gain of a region surrounding 8q24. Most translocations to *MYC* involved non-Ig partners (11/16, 69%), with the remainder involving *IGH* (3/5) or *IGL* (2/5). This frequency is similar to that seen in the MM (63% non-Ig).

In SMM, among the other translocations (non-canonical Ig translocation, and non-Ig *MYC* translocations), the two recurrent partners identified were *FAM46C* (n=2) and *TXNDC5* (n=2), which are also the most common recurrent partners in MM (n=6 and n=4, respectively). Some translocation partners were specific to SMM or MM, but they were non-recurrent and not reflective of a significant difference, **Figure 5.3.1.1.2.**

Copy number abnormalities (CNA) have been shown to be prognostically important in MM and to accurately determine their spectrum in SMM we performed ultra-low pass whole genome sequencing (ULP-WGS) on a subset of patients (68 patients with SMM and 116 NDMM patients). The recurrent regions identified were similar to those seen in MM, **Figure 5.3.1.1.3.** Using the targeted panel on the whole dataset CNAs which were less frequent in SMM included del(1p) (*FAF1/CDKN2C;* 2% vs. 17%, χ^2=11,2, p=0.0008), del(8p) (*NSD3*) (3.6% vs. 15%, χ^2=6.5, p=0.01), del(14q) (*TRAF3*) (7% vs. 19%, χ^2=5.5, p=0.02), del(16q) (*CYLD*) (13% vs. 28%, χ^2=5.8, p=0.01), and del(17p) (6% vs. 15%, χ^2=4.0, p=0.04), **Table 5.3.1.1.2.**

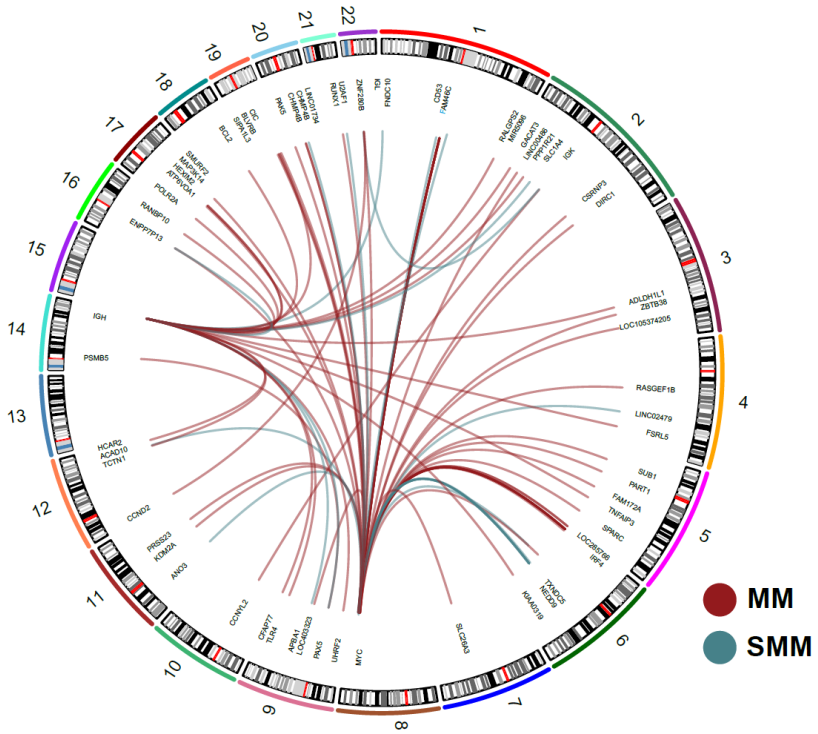


Figure 5.3.1.1.2. Circos plot of the landscape of non-Ig *MYC* and non-canonical Ig translocations highlighting *FAM46C* and *TXNDC5* as current recurrent partners.

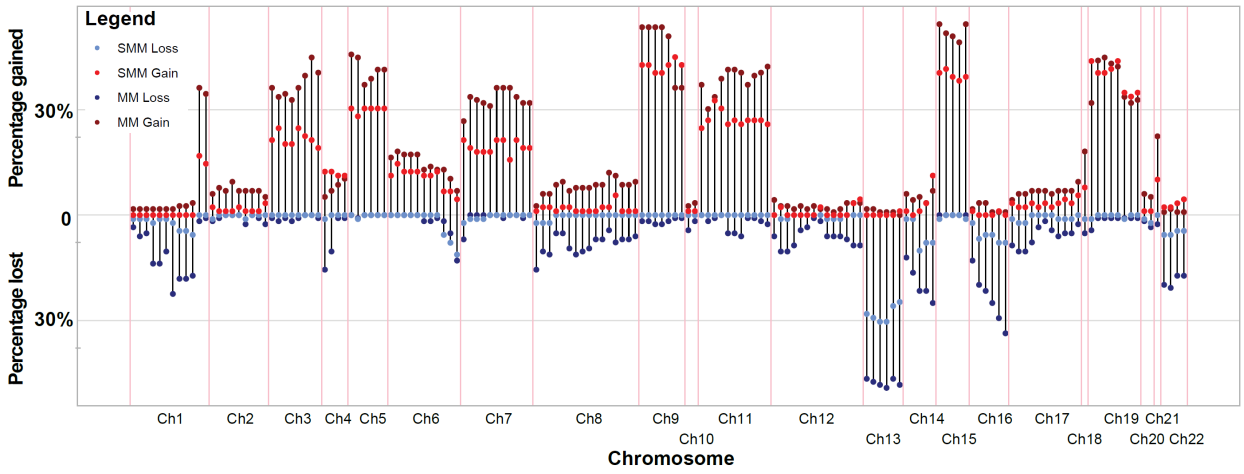


Figure 5.3.1.1.3. Copy number changes in 160 recurrently altered regions covering the known drivers in SMM and MM suggesting the landscape is similar but there are fewer copy number changes.

	MM	SMM	X^2	p=value
gain(1q)	33.63%	26.83%	0.9	0.3
amp(1q)	5.38%	4.88%	0.01	0.9
del(11q)	4.93%	3.66%	0.2	0.6
Trisomy 11	49.33%	40.24%	1.9	0.1
del(1p): *CKS1B*	17.04%	2.44%	11.2	0.0008
Trisomy 9	56.95%	46.34%	2.7	0.12
del(16q): *CYLD*	26.01%	12.20%	5.8	0.01
del(2p)	2.44%	4.48%	0.05	0.8
del(1p): FAM46C	22.87%	8.54%	7.8	0.008
Trisomy 6	21.52%	18.29%	0.2	0.6
del(16q): *MAF*	28.25%	13.41%	6.3	0.01
del(6q)	15.70%	13.41%	0.06	0.2
del(13q)	48.43%	41.46%	0.9	0.3
del(17p)	15.70%	6.10%	4	0.04
del(14q)	19.28%	7.32%	5.5	0.02

Table 5.3.1.1.2. The incidence of copy number changes in SMM (n=82) and MM (n=223) based on CN estimates from the targeted panel.

5.3.1.2. The Frequency of Known Driver Gene Mutations is Greater in MM.

We next investigated the impact and frequency of important mutations in SMM and MM using the same targeted sequencing panel. We applied an analysis of a catalogue of key MM mutations (Boyle *et al*, 2020) to cases with MGUS, SMM, EM, and MM. There was a significant difference in the overall number of mutations identified at various disease stages, **Figure 5.3.1.2.1**. Overall, 61% and 82% of SMM and MM had a mutation in one of the previously described mutational drivers that were present on the panel. There were no *IDH2, MAX, XBP1, CDKN2C, RB1,* and *SET2D* mutations in the SMM group. The most frequently mutated gene in SMM was *KRAS* (n=11, 13.4%). In comparison to NDMM fewer *NRAS* (χ^2=6,4, p=0.01), and *FAM46C* (Fisher's, p=0.008) mutations were detected, **Figure 5.3.1.2.2**. Overall, we identified a significant difference in the frequency of mutations in the MAPK pathway (*BRAF*, *NRAS,* or *KRAS* mutation) between SMM and MM (24% vs. 44%, χ^2=9.2, p=0.002). There were also fewer patients with NF-κB pathway mutations (*BIRC2*, *BIRC3*, *BCL10*, *CYLD*, *IRF4*, *MAP3K14*, *TRAF2*, or *TRAF3*; 5% vs. 16%, χ^2=5.7, p=0.02) and a trend

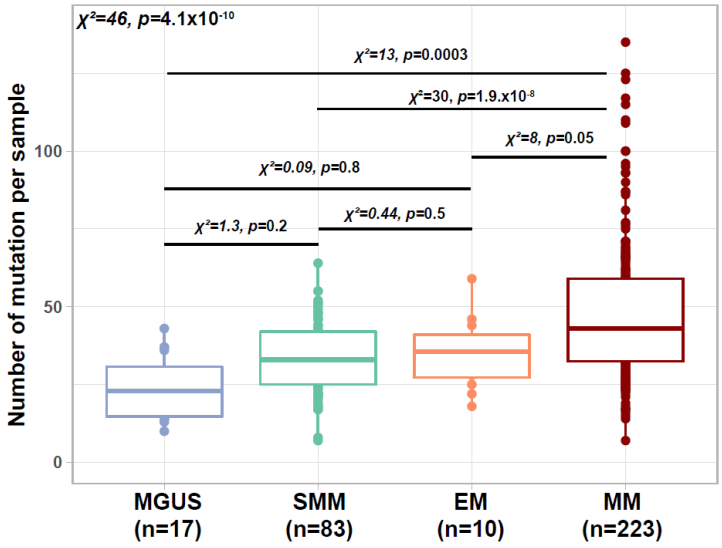


Figure 5.3.1.2.1: The number of mutations per sample identified on the targeted panel is lower in SMM than in MM

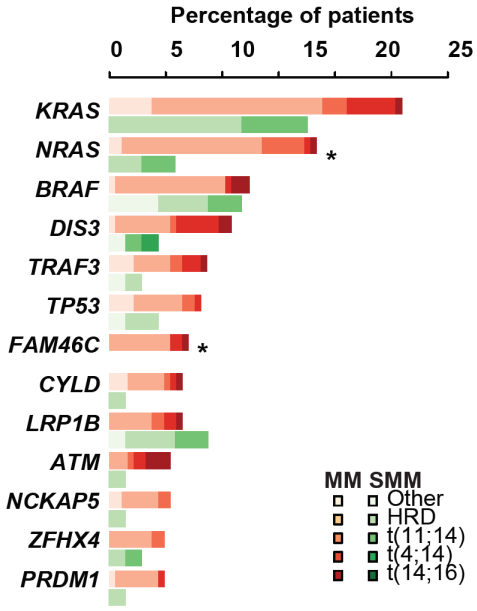


Figure 5.3.1.2.2. The most frequently mutated genes and their distribution across the most common molecular subgroups

suggesting fewer DNA repair pathway mutated patients (*TP53*, *ATM*, *ATR*, *ATRX*, *BRCA1*, or *BRCA2*; 7% vs. 17%, χ^2=3.8, p=0.0504).

We defined mutational signatures using an nNMF (as per Chapter 4). In terms of mutational signatures, we show that the contribution of APOBEC is significantly lower in SMM than in MM (11% (0-43%) versus 17% (0-48%), χ^2=5.2, p = 0.02) consistent with it having a role later in disease progression, **Figure 5.3.1.2.3**.

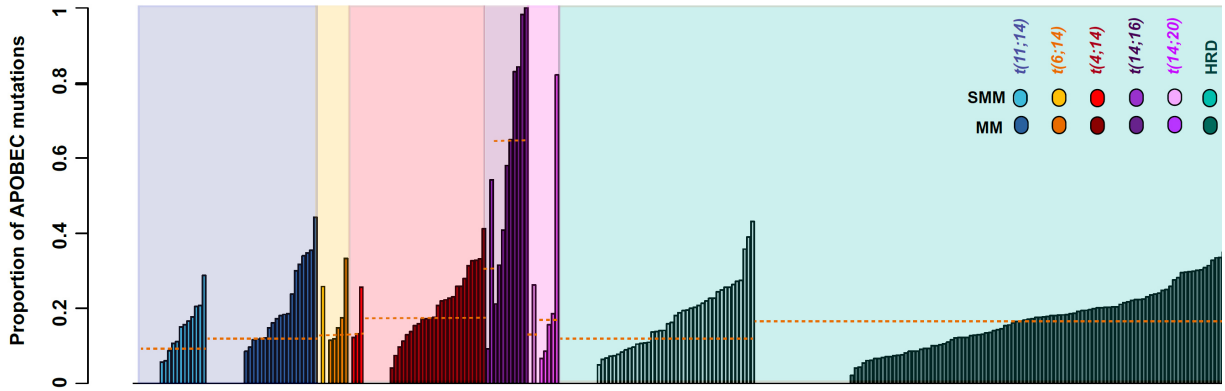


Figure 5.3.1.2.3. Contribution of APOBEC signatures to the mutational landscape of SMM and MM. Contribution of the APOBEC signature in SMM and MM by cytogenetic subgroup (yellow lines=median).

Furthermore, in SMM patients with either a *MAF* or *MAFB* translocation (termed maf), the median APOBEC contribution is 18% (0-54%, n=4), compared to 11% (0-43%, χ^2=0.5, p=0.4, n=78) in non-maf samples. Therefore, unlike observations in MM, we do not identify any significant difference between the two subgroups in SMM. Finally, the APOBEC contribution also seems lower in maf-SMM than it does in maf-MM (16% (0-44%, n=4) vs. 41% (0-100%, n=15)). Despite the small sample size, these data suggest that APOBEC is associated with disease progression, **Figure 5.3.1.2.4**.

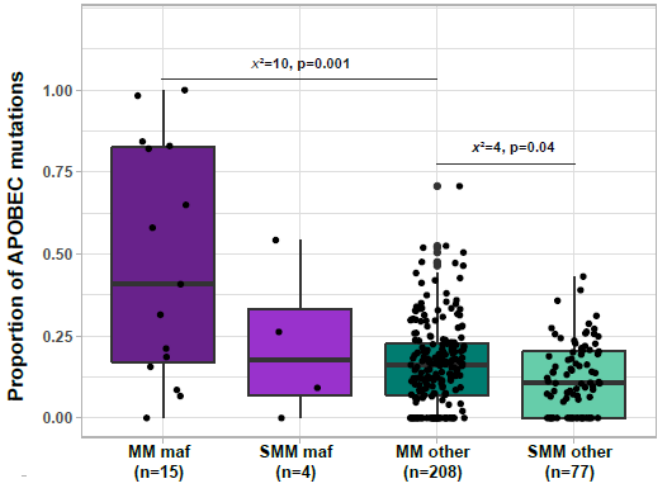


Figure 5.3.1.2.4. Contribution of the APOBEC signature in SMM and MM in maf and non-maf samples. X²=chi-squared, Kruskal Wallis test

5.3.1.3. Biallelic Inactivation of Tumour Suppressor Genes is Less Frequent in SMM.

Tumour suppressor gene inactivation is an important mechanism of relapse in NDMM with biallelic events having the most penetrant effects (Berger *et al*, 2011). To determine the role of this mechanism in SMM we defined a list of 20 previously identified tumour suppressor loci of relevance in MM relapse and investigated them for biallelic inactivation using combined mutation and CNA analysis. In total there were 103 biallelic events identified in MM (103 in 64/223 patients) versus only eight

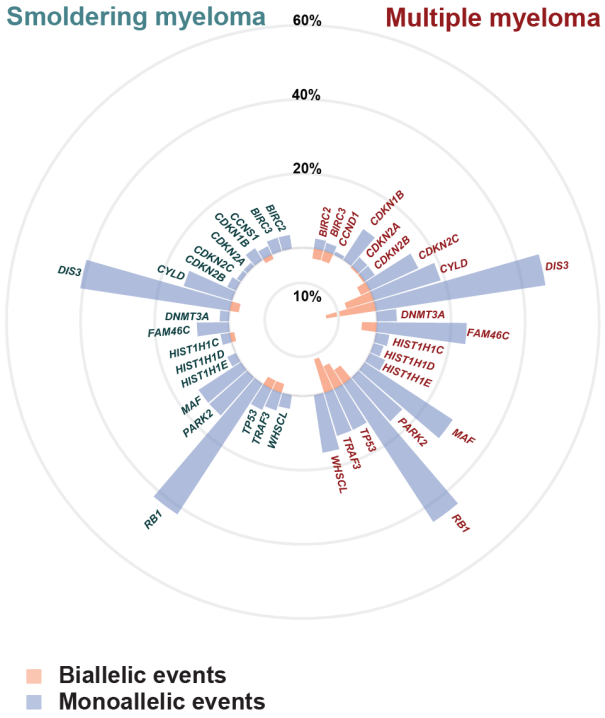


Figure 5.3.1.3.1. Frequency of biallelic and monoallelic events per driver locus in SMM and MM showing fewer biallelic drivers in SMM.

in SMM (8 in 8/82 patients) (χ^2=10.9, p=0.001) suggesting second hits at the same locus are a hallmark mechanism of transition to MM, **Figure 5.3.1.3.1.** Two Double-Hit patients, defined by a bi-allelic *TP53* inactivation were identified in this series of SMM patients compared to 18 in MM. Other key biallelic events included *DIS3, RB1, FAM46C* and *TRAF3*. Interestingly, biallelic *DIS3*, which has been associated with an adverse outcome in MM (Boyle *et al*, 2020) was present in 5% of MM cases vs only 2% of SMM cases, again consistent with it being associated with a more aggressive disease phenotype. These results are consistent with data previously generated at relapse after treatment and is indicative of biallelic inactivation of tumour suppressor genes being an important mechanism of progression at all disease stages.

5.3.1.4. VDJ rearrangement in SMM and MM

VDJ rearrangement are present in all malignant plasmacells. Little is known to their contribution to disease progression. To this intent, we were able to analyse the light chain loci rearrangement using VIDJIL (Giraud *et al*, 2014) in both the SMM and MM samples.

We confirm that the five different Jκ gene segments are used at similar frequencies in MM. We confirm the preferential use of Vκ1 over Vκ3 and over other segments that is seen in both normal B-cells and B-cell malignancies. Some segments seem to be used more often: Vκ1-33/1D-33 Vκ1-39/1D-39 Vκ3-20. The use of Jκ1 appeared higher in SMM than in MM (p=0.02, χ^2=5.4) and Jκ2 appeared lower in SMM than in MM (p=0.03, χ^2=4.5). A summary of the frequencies of usage of Jκ segments may be found in **Table 5.3.1.4.1**

Gene	JK1	JK2	JK3	JK4	JK5
MM	14.8 %	27.6 %	15. %	28.7%	13.3%
MDE	11.7 %	58.8 %	0.0 %	17.6%	11.7%
SMM	25.2 %	17.3 %	17.%	28.7%	9.57%
MGUS	23.08%	19.23%	7.69%	34.6%	15.3%

Table 5.3.1.4.1: Frequency of usage of Jκ segment per disease stage.

Similarly, we show that the overall frequency of usage of each Vκ segment is similar in SMM and MM, **Figure 5.3.1.4.1**.

As previously described, each J and V segment can rearrange with each other. The frequency of each rearrangement appears balance. With the exception of Vκ1 segments that rearrange with Jκ2 (p=0.02, χ^2=5.4), **Figure 5.3.1.4.2.**

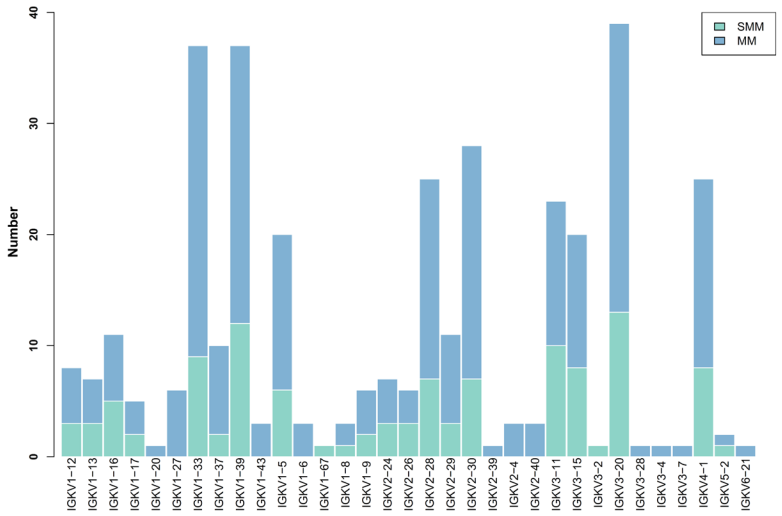


Figure 5.3.1.4.1. Comparison of Vκ usage in SMM and MM

5. Smouldering Myeloma

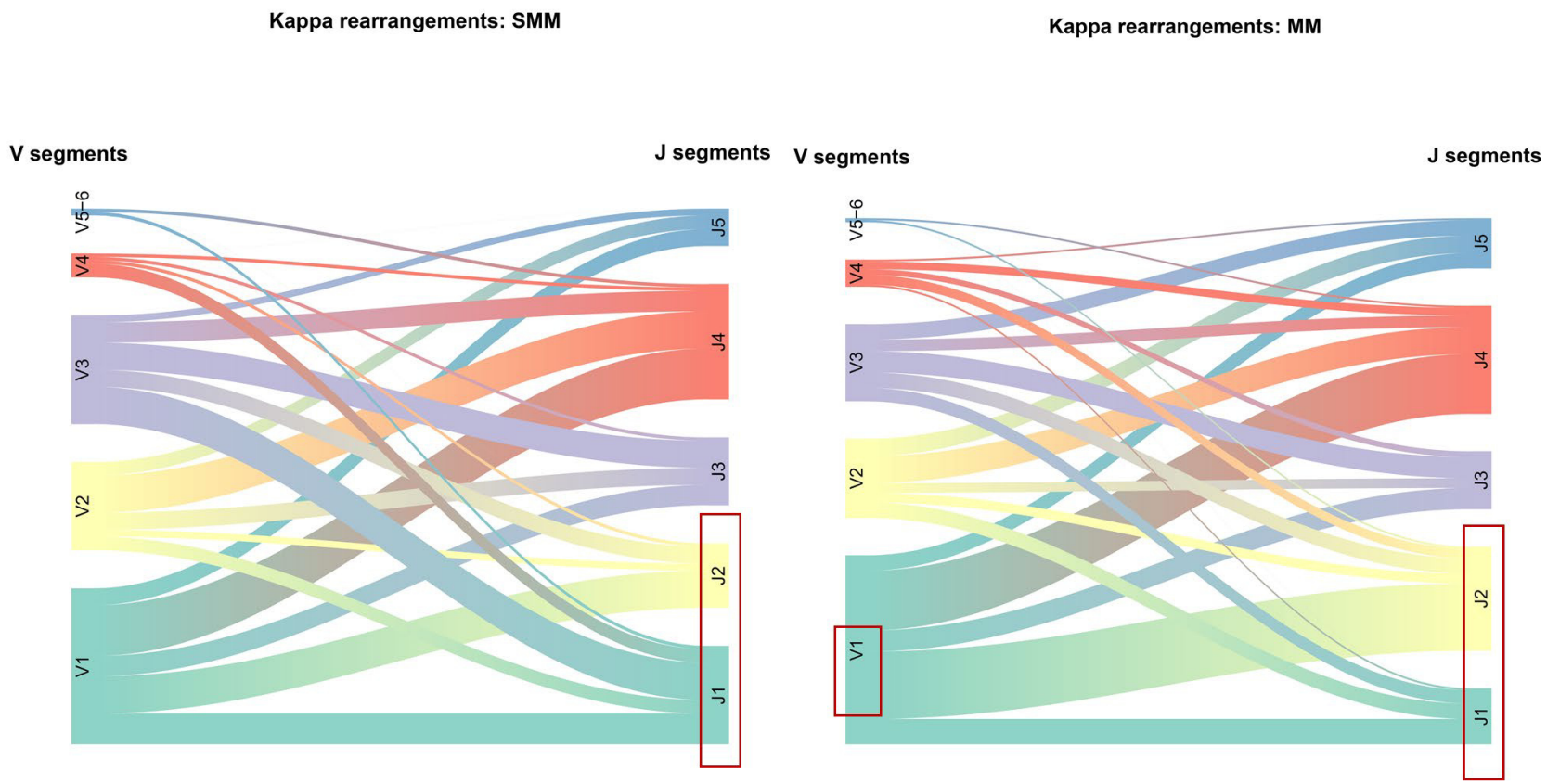


Figure 5.3.1.4.2. Riverplot showing the distributions of kappa segment usages between SMM and MM

5. Smouldering Myeloma

When considering the lambda segments, we confirm the preferential use of Vλ3 over other segments that is seen in both normal B-cells and B-cell malignancies. There is no statistical difference between the usage of Vλ segments between the different stages of disease. Some segments seem to be used more often: Vλ 3-1 and Vλ 3-21. This has previously not been described. We confirm that the Jλ1, Jλ2, Jλ3 gene segments are the most frequently used. Jλ5 was not seen in SMM, **Table 5.3.1.4.3.**

Gene	Jλ1	Jλ2	Jλ3	Jλ4	Jλ5
MM	29.09%	23.64%	29.09%	0.00%	18.18%
EM	0.00%	50.00%	50.00%	0.00%	0.00%
SMM	20.00%	40.00%	37.14%	2.86%	0.00%
MGUS	25.00%	50.00%	12.50%	0.00%	12.50%
Total	24.51%	32.35%	31.37%	0.98%	10.78%

Table 5.3.1.4.3. Frequency of usage of Jλ segment per disease stage

Overall, all Vλ can rearrange with all J λ. The frequency of each rearrangement appears balanced with the exception of Vλ1 segments that rearrange with Jκ2 at a lower frequency than expected ($p=0.01$, $\chi^2=5.9$), **Figure 5.3.1.4.3.**

We then went on to consider light chain translocations. *IGL* translocations are more frequent than IGK. IGL translocations are not restricted to λ+ patients but there is no evidence of primary/canonical *IGL* translocation in κ+ patients suggesting they are either secondary events or follow abnormal receptor revision. *IGL* breakpoints are within expected regions. There is no significant difference between SMM and MM in terms of *IGL* and *IGK* translocation, **Table 5.3.1.4.4 and 5.3.1.4.5.**

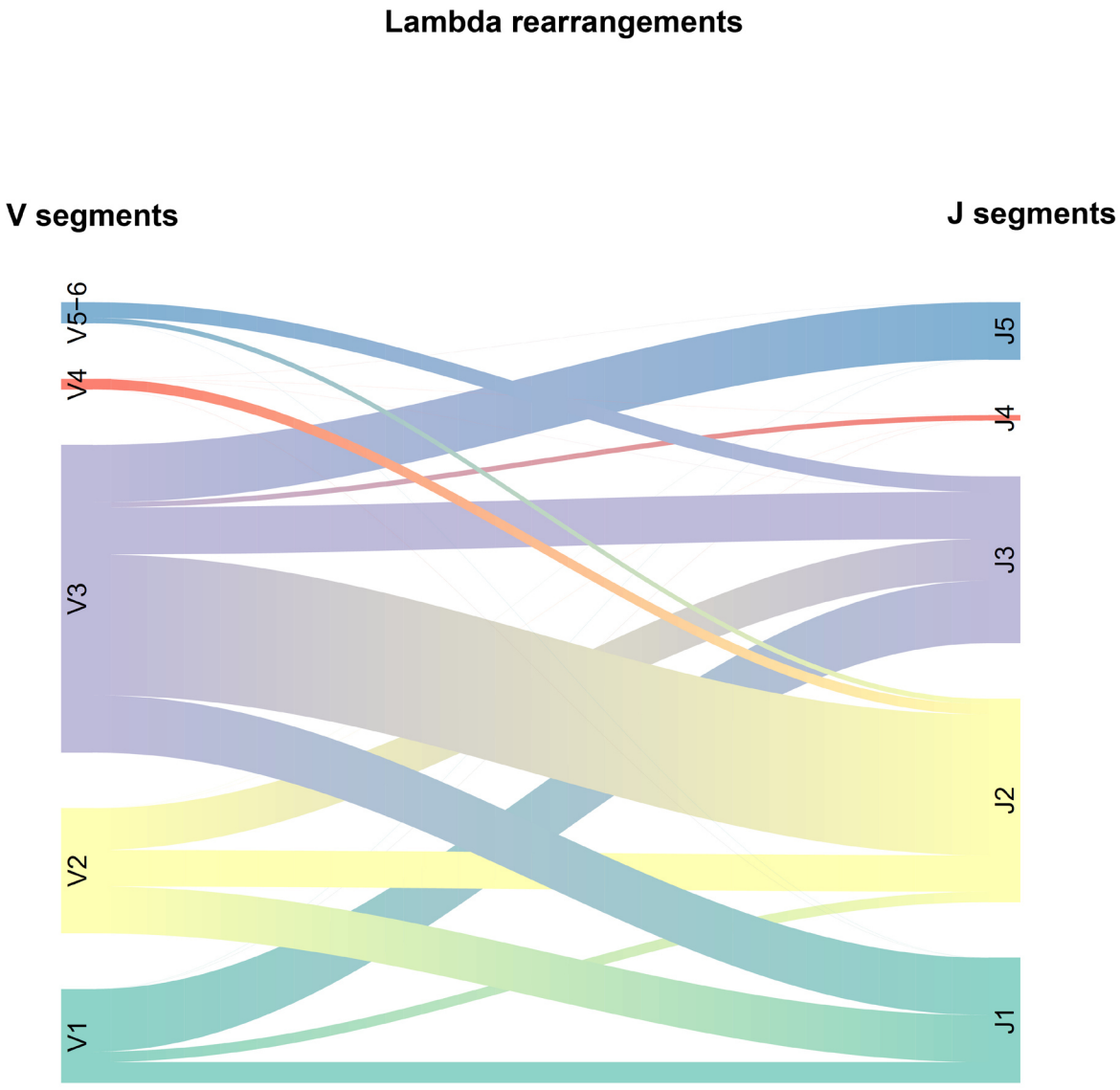


Figure 5.3.1.4.3. Lambda VJ rearrangements in SMM.

5. Smouldering Myeloma

PatID	Isotype	KV	KJ	IGHV	IGHJ	λV	λJ	Tx
21645	IGAK	IGKV1-33	J2	IGHV4-30-2	J6			Enhancer
21694	IGDL	IGKV1-37	J2			IGLV3-1	J2	C3-J4 & Enhancer
21694	IGDL	IGKV2-30	J3					
22949	IGAL	IGKV1-33	J4			IGLV3-1	J1	Enhancer
26757	IGAK	IGKV3-20	J5					C2-J3 & Enhancer
33675	IGAL	IGKV3-20	J5			IGLV1-44	J2	C3-J4 & Enhancer
34304	IGAL	IGKV2-40	J1			IGLV5-48	J3	IGLV3-4 & IGLV4-3
34304	IGAL	IGKV4-1	J4					

Table 5.3.1.4.4 *IGL* translocation

PatID	Isotype	KV	KJ	IGHV	IGHJ	λV	λJ	Partner	IGL
16631	IGAL	IGKV1-12	J4					*MYC*	*IGLC6-J7* & Enhancer
23115	IGAL	IGKV2-30	J3	IGHV3-66	J4	IGLV3-21	J3	*MYC*	*IGLC3-J4*
23115	IGAL	IGKV2-30	J4						
30966	IGGL	IGKV1-33	J3					*MAFB*	*IGLV6-57*
30966	IGGL	IGKV2-28	J4						
30966	IGGL	IGKV2-OR22	J4						
35704	LCOK	IGKV1-39	J4					*MYC*	*IGLV8-61*
35704	LCOK	IGKV1-8	J3	IGHV1-18	J5				
35704	LCOK	IGKV2-30	J4						
35892	IGGK	IGKV3-20	J2					*VAMP1*	*IGLC4-J5*
40455	IGGK	IGKV1-39	J2					*IGKJ3*	*IGLV4-45*
40455	IGGK	IGKV3-2	J2						

Table 5.3.1.4.5: *IGK* translocation

5. Smouldering Myeloma

We attempted to analyse the IGH rearrangement using the same approach. Unfortunately, given the length of the reads, we failed to accurately reconstruct the correct mapping of the region in the majority of samples.

Overall, these data highlight that there are subtle but probably not clinically significant differences in choice of κ segment in SMM and MM and that light chain translocation respect the normal rearrangement sequences with no evidence of λ rearrangement in κ patient.

5.3.1.5. Mutations in KRAS are Associated with a Shorter Time to Progression.

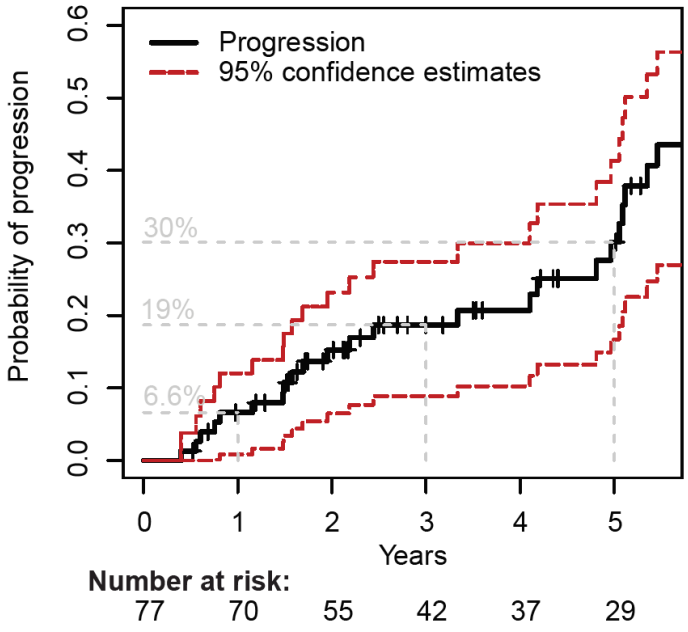


Figure 5.3.1.5.1. Progression free survival with a 30% progression rate at 5 years with no plateau, suggesting an ongoing risk.

In order to determine the molecular factors associated with time to progression to MM we assessed outcome in 77 patients evaluable for progression. Thirteen percent of patients were identified as high-risk according to the GEP4 risk-score[15], and 23% were high-risk and 35% intermediate risk according to IMWG criteria. With a median follow-up of 5.18 years we identified a five-year progression rate of 30.1%, with no plateau consistent with there being an ongoing risk of progression, **Figure 5.3.1.5.1**.

5. Smouldering Myeloma

To determine the associations with high-risk SMM we analysed the distribution of mutations across the IMWG and GEP4 risk groups, **Figure 5.3.1.5.3**. Although the numbers were low, we note that *TP53* (n=3/3) mutations and del(17p) (n=3/4) were enriched in the high risk GEP4 patients. Mutations in *DIS3* (n=3), *CCND1* (n=2), and *ATM* (n=1), as well as the t(4;14) were associated with high-risk patients according to the IMWG definitions**.**

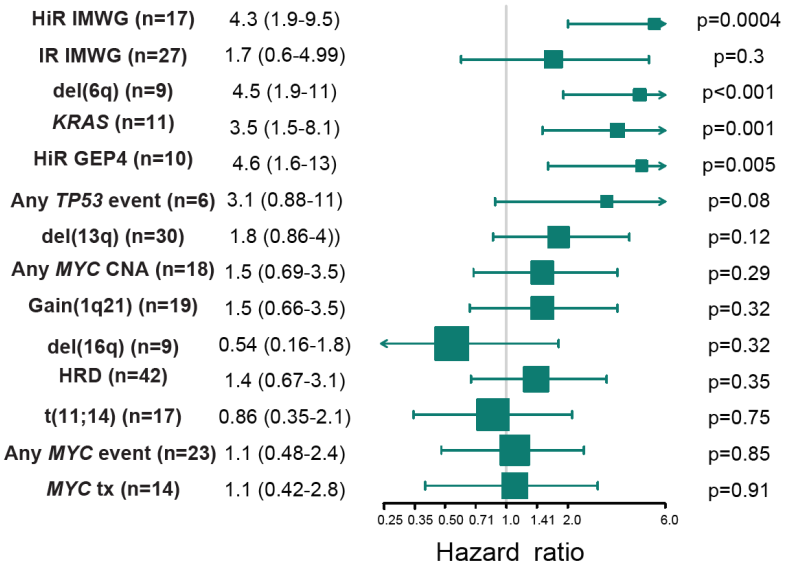


Figure 5.3.1.5.2. Univariate analysis of molecular features in SMM.

In a univariate analysis of time to progression to MM including all abnormalities present in at least seven samples, **Figure 5.3.1.5.2,** we show that high-risk IMWG status and the GEP4 risk were associated with an adverse outcome, **Figure 5.3.1.5.4A-B**. Importantly, the presence of a *KRAS* mutation (n=11) was associated with a short time to progression (HR 3.5 (1.5-8.1) p= 0.001), **Figure 5.3.1.5.4.C**. Del(6q) (n=9) was also associated with a short time to progression (HR 4.5 (1.9-11), p=0.0005)), **Figure 5.3.1.5.4.D.** Of note, four of these patients had high-risk status according to GEP4, and two of them carried a del(17p). Biallelic events were not associated with outcome (data not shown)**.** Combined *BRAF, KRAS,* and *NRAS* did not significantly contribute to progression, **Figure 5.3.1.5.4.E**. The presence of a translocation to the *MYC* locus has been identified previously as being associated with adverse clinical outcomes in SMM. (MMRF CoMMpass Network *et al*, 2020; Bustoros *et al*, 2020) Here we did not find that *MYC* translocations associated with outcome in this set of patients, defined by the current IMWG criteria.(Rajkumar *et al*, 2014) Finally there was a trend suggesting that patients

with a small contribution of APOBEC signatures (<5%) had a better outcome than the others **Figure 5.3.1.5.4.F**.

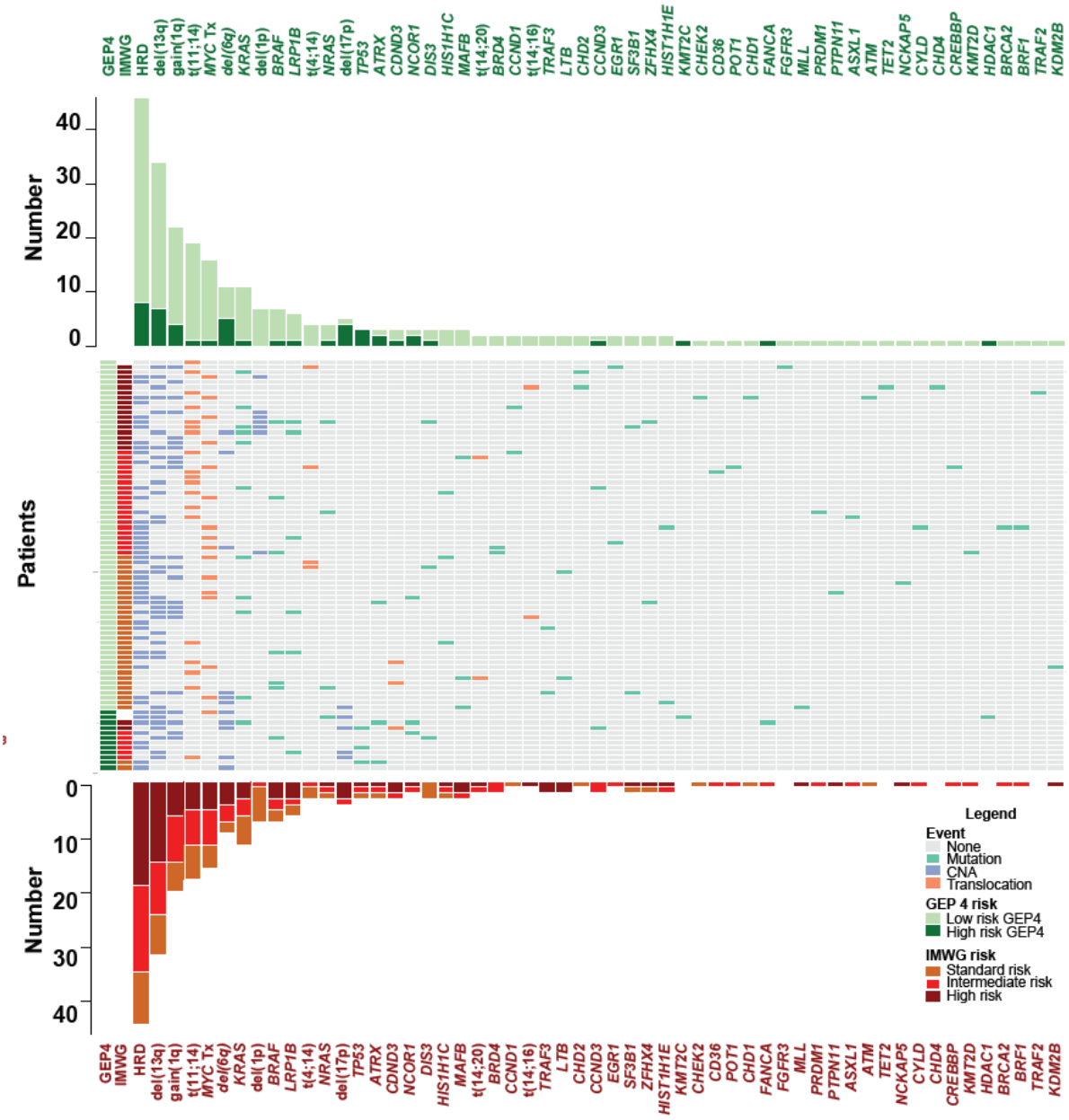


Figure 5.3.1.5.3: The distribution of somatic abnormalities per sample and risk group

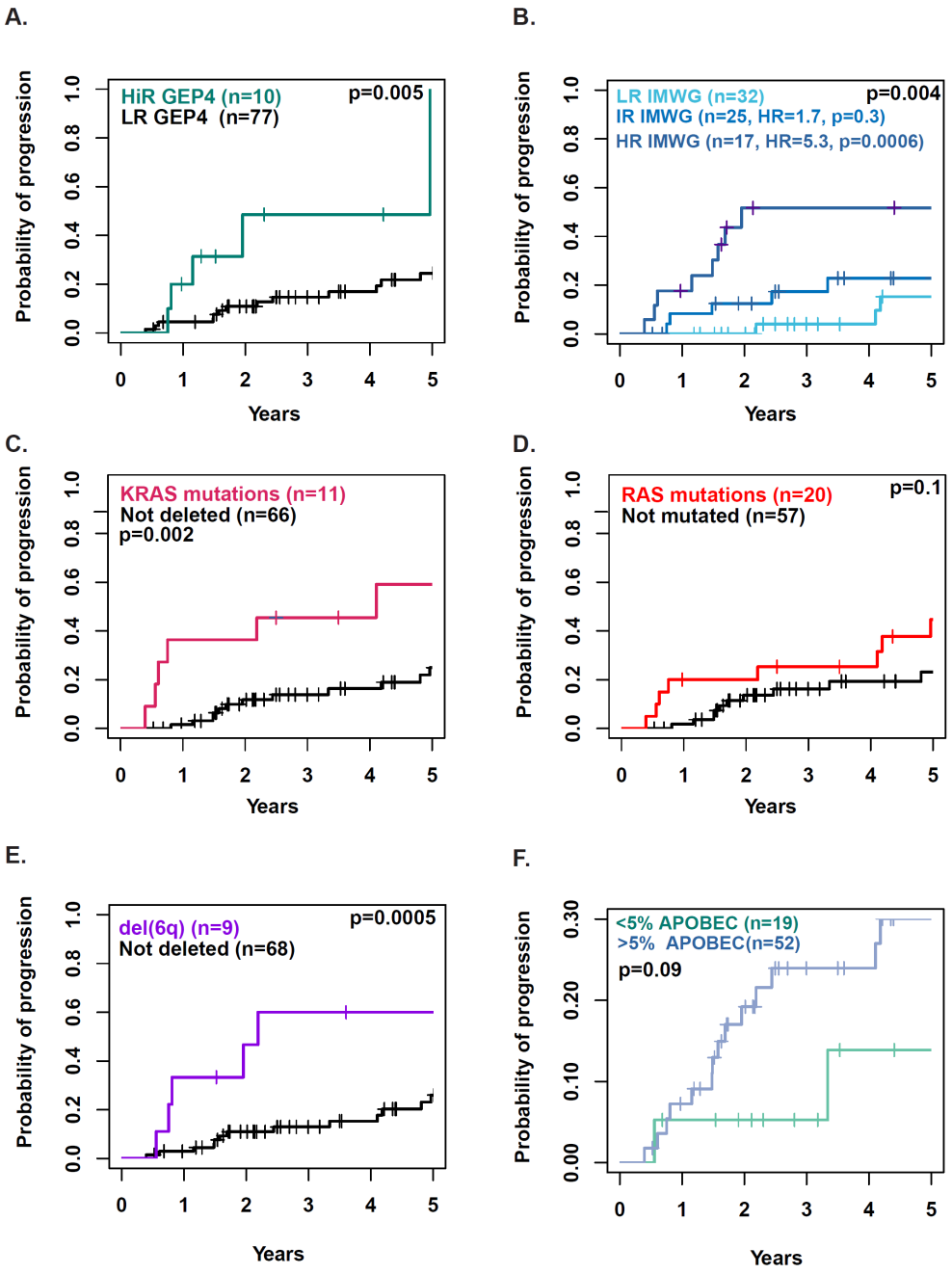


Figure 5.3.1.5.4. Probability of progression according to GEP4 (A), IMWG score (B), *KRAS* (C), RAS mutations (D), del(6q) (E), and APOBEC signature (G)

5.3.1.6. Multivariate Analysis of Molecular Markers Involved in Time to Progression.

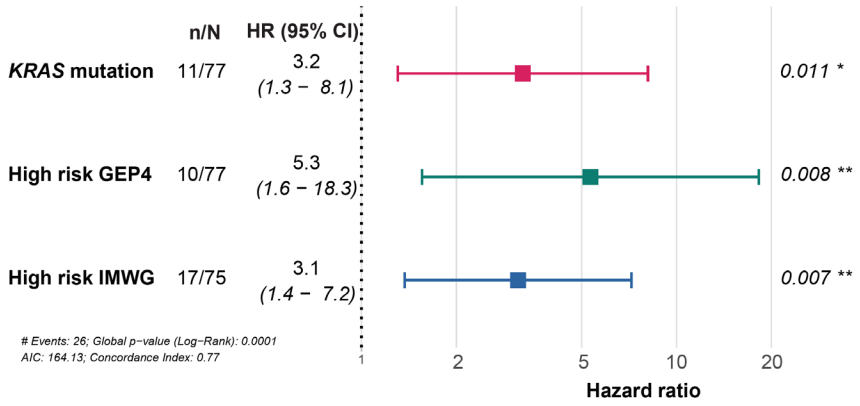


Figure 5.3.1.6.1. Multivariate analysis suggests GEP4, high-risk IMWG, and *KRAS* mutations are independent prognostic markers.

In order to identify independent prognostic markers a multivariate analysis using previously published risk scores (GEP4 and IMWG) alongside factors with n≥10 and at least a borderline impact on univariate (*KRAS* mutations and del(13q)) was performed. The presence of a *KRAS* mutation, a high-risk GEP4 or high-risk IMWG retained their impact on outcome, **Figure 5.3.1.6.1**. Indeed, the presence a *KRAS* mutation, a high-risk GEP4, and high-risk IMWG was associated with a hazard ratio for progression of 3.2 (CI 1.3−8.1, p=0.011), 5.3 (CI 1.6-18.2, p=0.008), and 3.1 (CI 1.4-7.2, p=0.007), respectively, consistent with their independent contribution to risk at this disease stage. Thus, we show in a significant number of cases the important prognostic contribution of mutations in *KRAS* at the SMM disease stage. This observation contrasts significantly with the lack of association with prognosis that Ras mutations have in NDMM, which underlines the importance of these mutations in disease biology.

5. Smouldering Myeloma

To determine whether *KRAS* mutations have the potential to contribute to novel risk scores we compared the *KRAS* mutated cases with the two previously used risk models and show that only 13% of patients overlapped between one of the three risk models, with five patients only having *KRAS* mutations, **Figure 5.3.1.6.2.** No patient fell into the intersection of all three risk factors suggesting that they may contribute useful clinical information and as such should be included in future clinical trials. We have shown previously that an accumulation of risk factors leads to higher risk status in MM,(Boyd *et al*, 2012) and therefore, we performed a similar analysis with the SMM risk factors. When more than one lesion was present, a HR of 11.2 (4-31) for progression was seen with an associated median time to progression of 1.16 years, **Figure 5.3.1.6.3,** compared to a HR of 3.7 (1.4-9.5) and 5-year PFS of 57% (35-94%) when only one risk factor is present and an 88% (78-99%) 5-year PFS when none are present. This was clearer when considering the patients whose samples were taken within the first 90 days of diagnosis, **Figure 5.3.1.6.4**. both for *KRAS* (A) and the three-group risk model. We also performed the analysis with variants present at a number of n≥7 and found similar results, **Figure 5.3.1.6.5**.

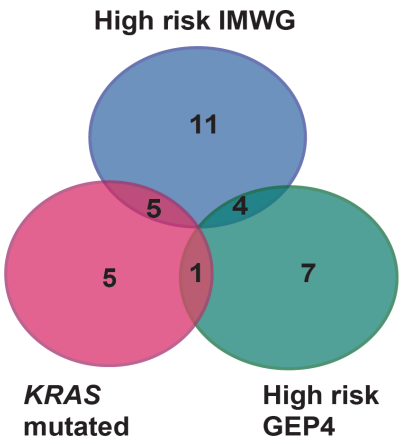


Figure 5.3.1.6.2. Venn diagram showing the overlap between high risk GEP4, high risk IMWG, and KRAS mutations

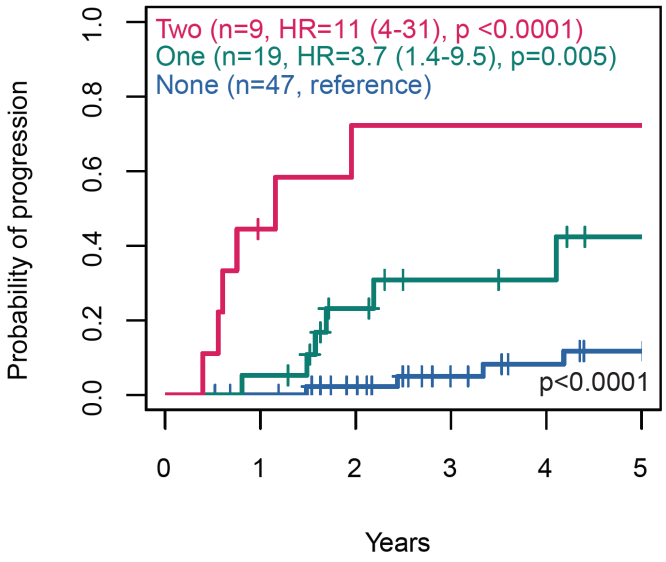


Figure 5.3.1.6.3. PFS of patients with either no, one, or two adverse factors defined buy high risk GEP4, high risk IMWG, and KRAS mutations

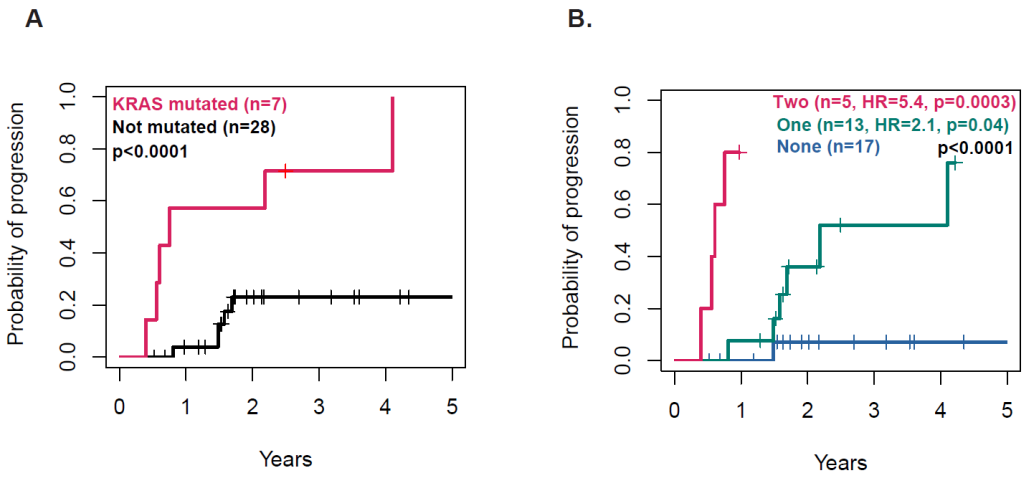


Figure 5.3.1.6.4. Analysis of the patients that were within 3 months from initial diagnosis suggesting A. KRAS mutations have a stronger impact on progression among the NDSMM patients. B. GEP4, KRAS mutations, and HR mutations segregate patients effectively.

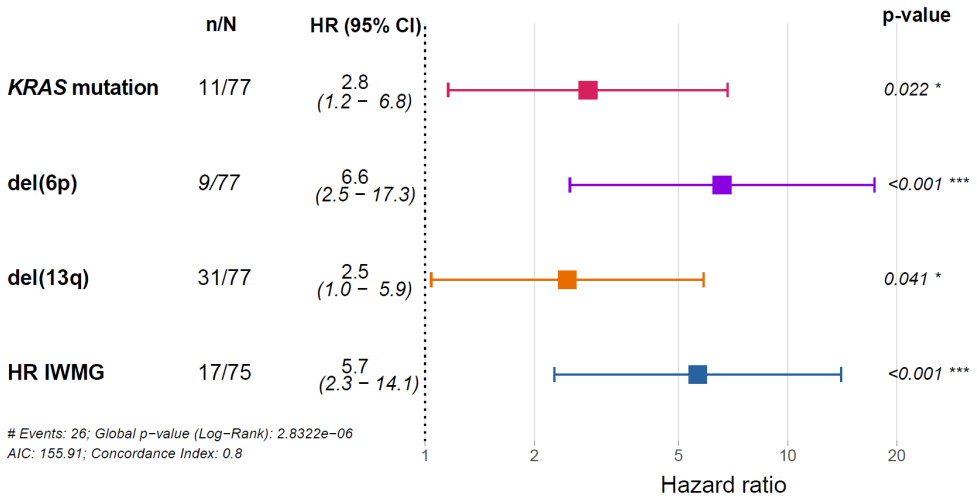


Figure 5.3.1.6.5. Multivariate analysis performed using all factors with n≥7 (IMWG, GEP4, del(6q), del(13q), mut/del *TP53*, and *KRAS* mutations).

5.3.1.7. Pathway Deregulation Associated with Progression.

One of cornerstone pathways in MM is the NF-κB pathway that is constitutionally active in MM. Mutations of the NF-κB pathway are amongst the most frequent in plasma cell disorders making an understanding of their impact in SMM critical. Mutations affecting

5. Smouldering Myeloma

NF-κB were seen in 5% of SMM patients, which is lower than the 16% seen in MM patients (χ^2=5.7, p=0.02). Mutations in the NF-κB pathway did not associate with outcome in this dataset (data not shown). Based on previously published work (Annunziata *et al*, 2007), we generated gene signatures, from gene expression array data, that correlated with NF-κB p65 activity that can be used as a surrogate for changes in NF-κB pathway activity. These results show that the score is not statistically different in SMM compared to MM [SMM-10.7 (IQR: 9.6 -11.9) versus MM -10.57 (IQR (8.68-12.6) p=0.8)]. However, it was lower in normal CD138$^+$ cells (11.4 (11-11.7), p<0.0001). These data suggest that NF-κB dependency is similar in SMM and MM, consistent with it being an early event associated with progression from MGUS and not with the transition of SMM to MM, **Figure 5.3.1.7.1.**

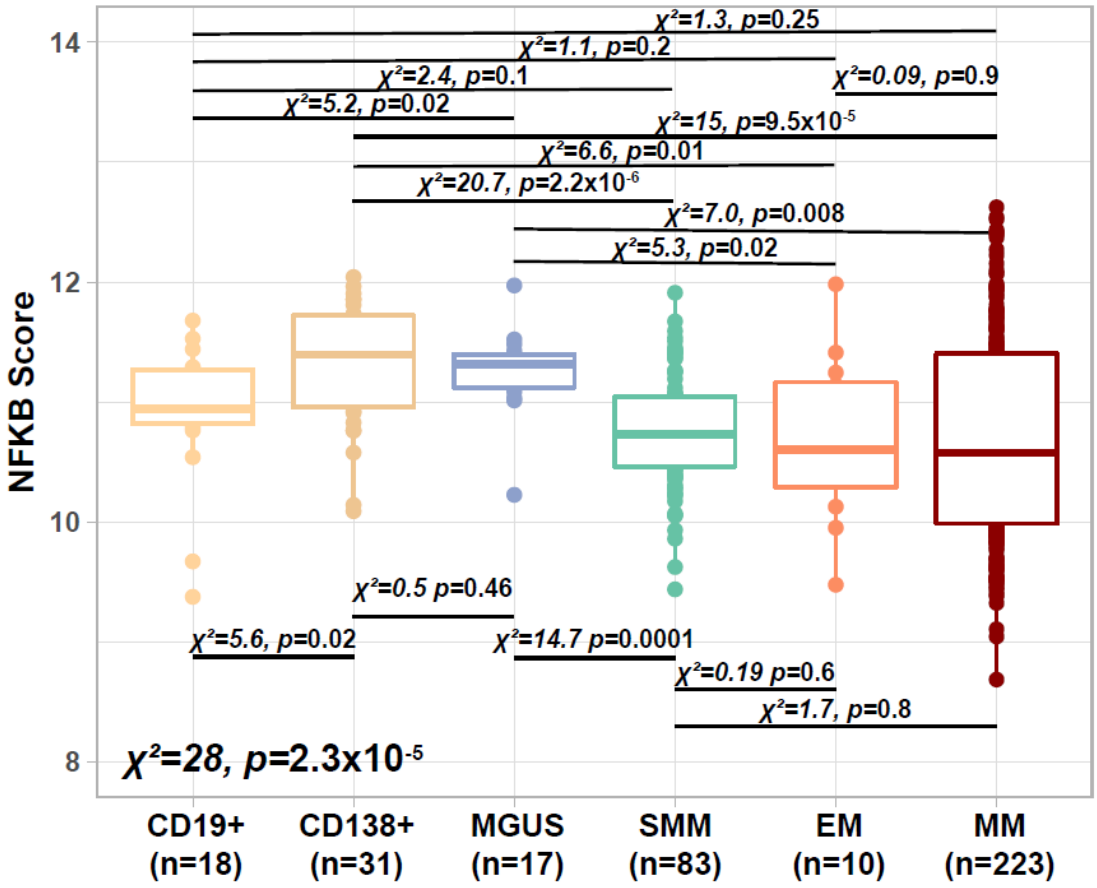


Figure 5.3.1.7.1. NFκB score in SMM is similar to MM and not MGUS

5.3.2. Sequential Molecular Changes Identified Within Individual Cases Over Time.

In order to investigate the temporal relationship of genetic changes associated with progression we studied 53 sequential samples obtained from nine patients with whole exome sequencing derived over a median follow-up of 7.26 years (5.17-∞). Six of the patients progressed during the time of follow up, with progression samples being available for 5/6 patients. Two patients with SMM had not progressed at the time of the analysis. Additionally, one of the 9 patients, patient G, who had features of early myeloma, progressed to MM within the first six months, **Figure 5.3.2.1.1.**

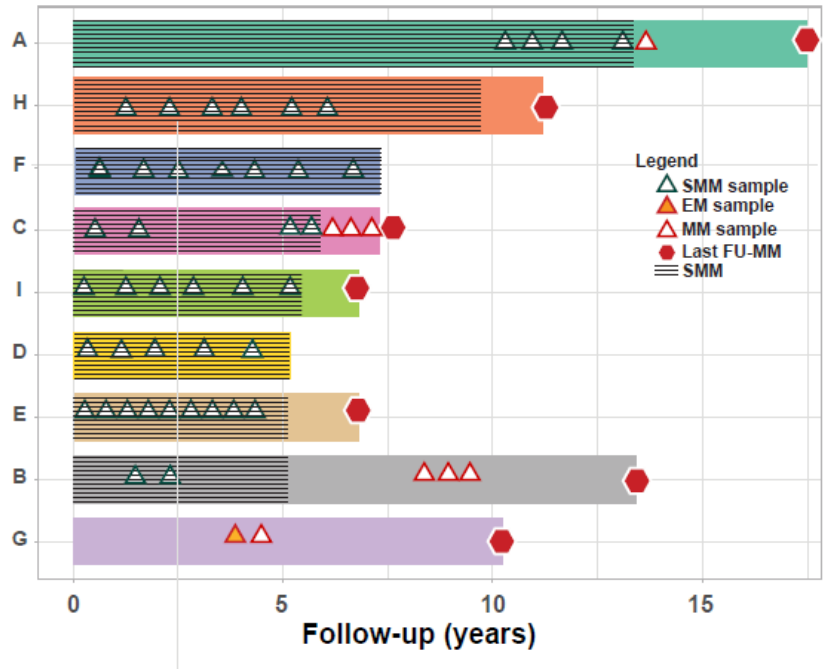


Figure 5.3.2.1.1. Swimmer plot of the group of 9 patients followed overtime. The colour bars represent the progression free survival

5.3.2.1. Structural abnormalities.

We identified four patients with a t(11;14) (patient C,D,E,F) and one with a t(4;14) (patient I). These molecular events were detected as being clonal in each of the sequential samples. Five patients had HRD (patient B, E, F, G, I). Neither an etiological translocation or hyperdiploidy could be detected in patient A but there was evidence of segmental copy number changes. We have previously defined a catalogue of segmental

CNA in NDMM that contribute significantly to prognosis. In this series of sequential samples, we show that hyperdiploidy is an early event and is stable over time and does not constitute a significant mediator of progression. In contrast, we identified a number of regions of segmental copy number change that may contribute to disease progression. While only limited numbers of segmental copy number change were identified, some provide informative examples. One patient for example acquired a *DIS3* and *RB1* deletion through loss of chromosome 13, that gradually become more clonal over time, as well as a gain(1q) and a del(11q) at progression, **Figure 5.3.2.1.2.C**. The sequential data provided additional insights into the role played by copy number change allowing the assessment of changes over time. In this fashion, we were able to identify fluctuations in the levels of individual clones defined by copy number at specific sites that are not seen at the transition to MM. These include a loss of chromosome 5 in patient C that was not sustained at progression and a loss of 16q that was only seen in patient B's initial sample, **Figure 5.3.2.1.2.**

Importantly we identified a case where a t(8;14) appeared during follow-up allowing us to determine in an individual case its impact on sub-clonal size. The translocation involved a rearrangement between *MYC* and the *IGHA2* switch region developing at a relatively early stage of disease; post-initial immortalisation but before clinical myeloma. Using ddPCR, we quantified the rearranged Ig-MYC allele and the productive IGHG3 allele. This analysis confirmed that the translocation was not present at diagnosis, appeared in a small fraction (1%) 3.0 years after initial diagnosis and steadily increased over time,

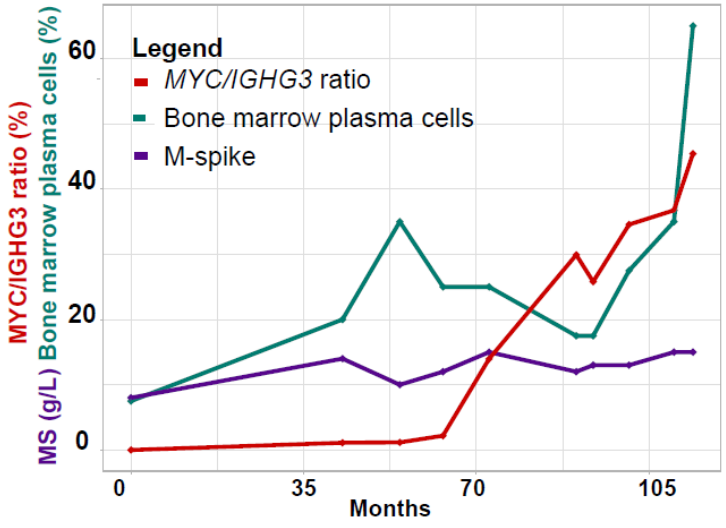


Figure 5.3.2.1.3. The acquisition of a t(8;14) takes over the tumour population and is associated with disease progression

reaching 45% of cells on the last sample, 8.2 years from the initial diagnosis. This increase was matched by an increase in bone marrow plasma cells and M-spike, **Figure 5.3.2.1.3.** These findings are consistent with it having contributed to clonal expansion potentially via an increased proliferation rate.

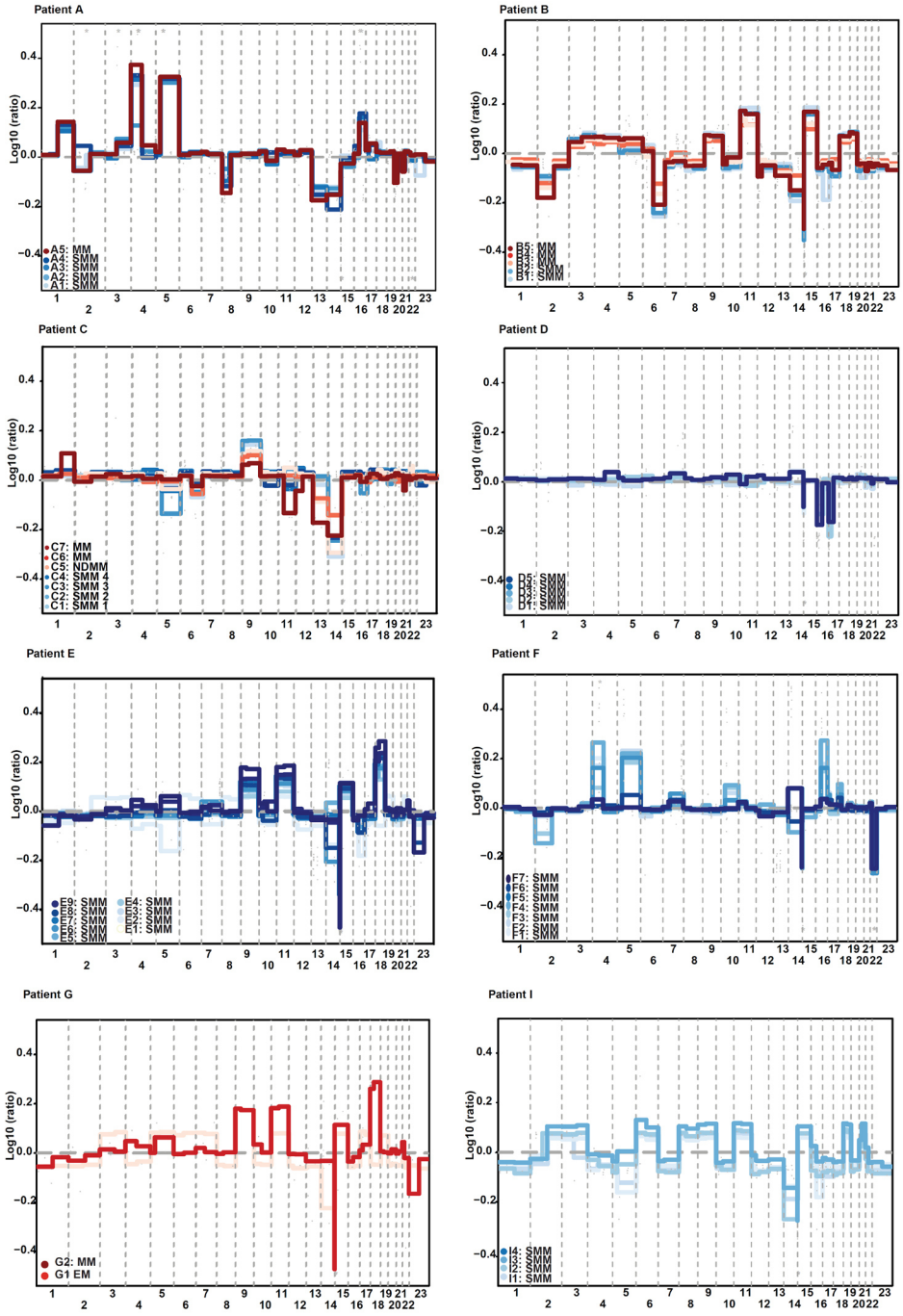


Figure 5.3.2.1.2. Evolution of CNA changes overtime

5.3.2.2. Mutational Load in SMM Patients Increases Over Time.

Mutational load has the potential to both contribute to progression and to be a marker of changes in mutational mechanisms. To investigate changes in mutational load in this set of serial samples and to account for variation in depth of coverage, we normalised the number of mutations per sample according to the depth of coverage and analysed trends in mutational load over time. At the SMM stage, the mutational load increased with time. Samples with hyperdiploidy (HRD) seemed to have a higher mutational rate than other subgroups (nHRD) but their follow-up was longer in this group of patients**.** The mutational rate of patients that eventually progressed was not statistically different from the mutation rate of those that have not progressed, but there was a trend suggesting the mutation rate increased around progression in SMM, **Figure 5.3.2.2.1.**

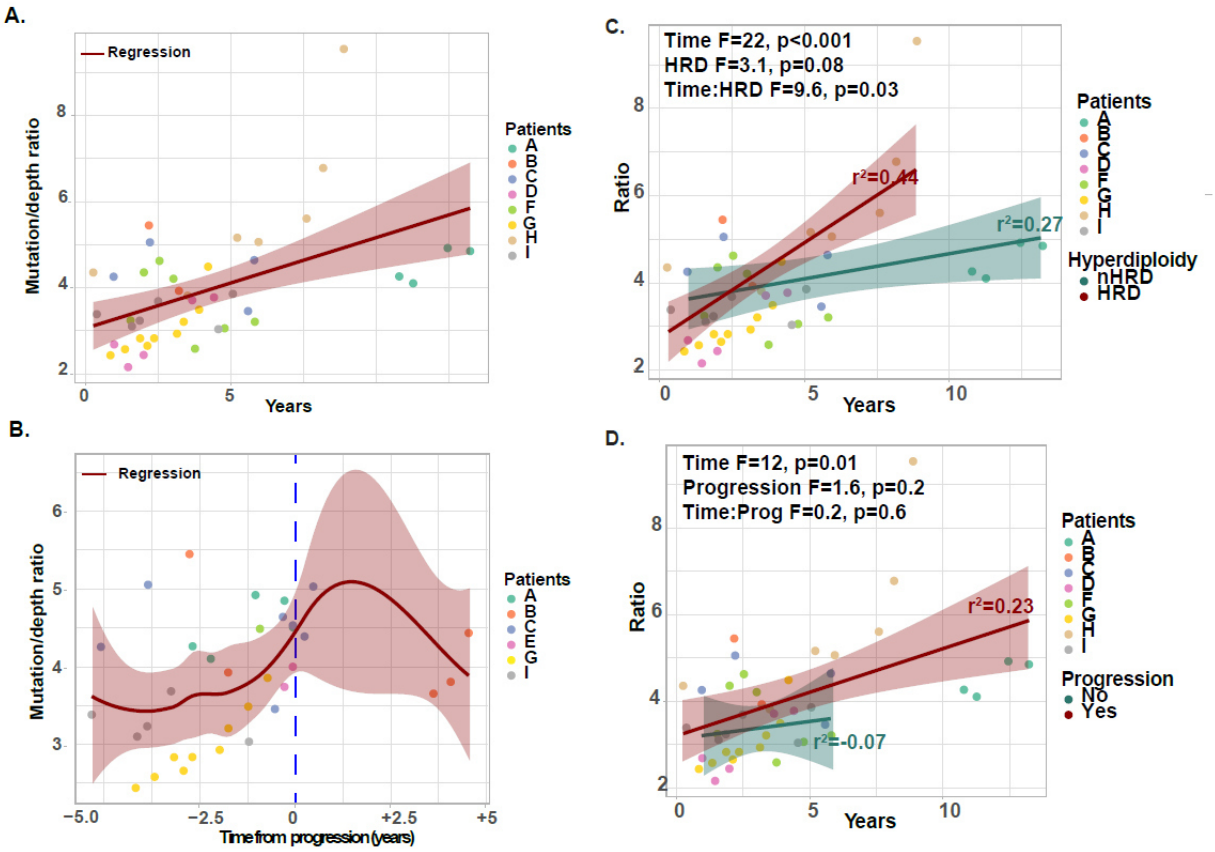


Figure 5.3.2.2.1. Evolution of mutational burden overtime. A. Overall, the mutation burden increases overtime. B. The mutational rate accelerates around progression. Correlation between Time and tumour burden according to C. HRD status. D. Progression status.

Focussing on previously described (Walker *et al*, 2018) mutational drivers, we identified a median of 1.5 (range 0-4) drivers per SMM sample with 21 genes being mutated in 31/53 of samples. We identified a correlation between the number of drivers present and time from clinical presentation but this was driven by one sample only (A), **Figure 5.3.2.2.A**. The CCF of the driver mutations increased over time, consistent with them being actively selected, **Figure 5.3.2.2.B**. These findings are consistent with prior data derived from paired SMM/MM studies where only limited numbers of new mutational drivers at progression were seen**.**

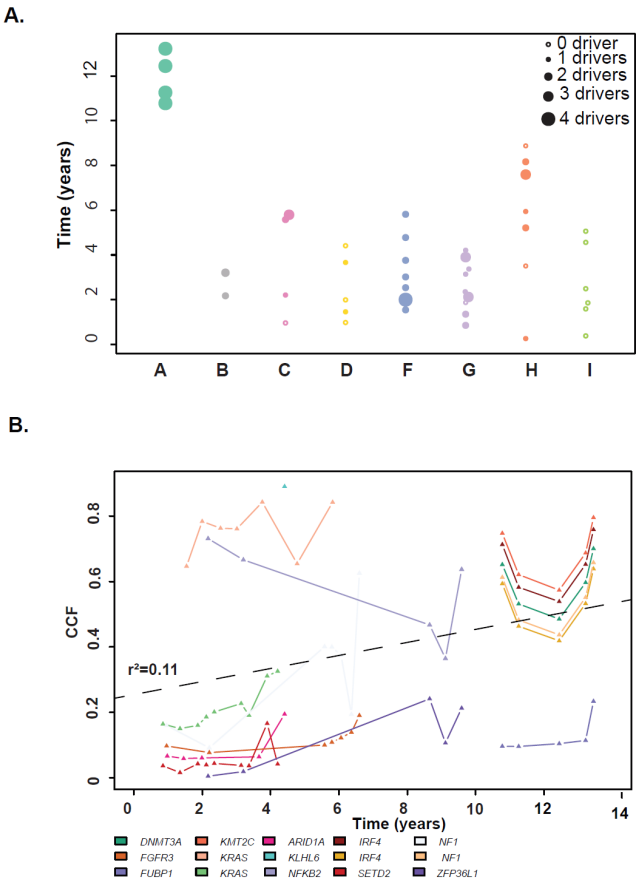


Figure 5.3.2.1.2. Evolution of Driver genes overtime. A. Evolution of the number of driver genes per sample overtime suggesting fluctuations but no steady increase. B. Evolution of the CCF of driver mutations in each patient overtime.

5.3.2.3. Mutational Processes are Stable at the SMM/MM Interface.

Signatures of mutational processes have emerged as an important tool to determine both intrinsic and extrinsic mechanistic factors mediating cancer aetiology and progression. Here, we examined the role played by mutational signatures at the SMM disease phase in serial samples. The background signatures (SBS1 and SBS5)(Alexandrov *et al*, 2020) are the major contributors to the mutational patterns seen in SMM. A single patient had a non-canonical AID signature (SBS9) seen in more than one sample, on a HRD background (data not shown). To determine whether changes in mutational processes correspond with the transition to NDMM we analysed the contribution of each signature in relation to the time of progression to MM. Samples taken more than two years prior to progression had the same signatures as both those taken within two years of progression and those cases that did not progress, **Figure 5.3.2.3.1.A.**This observation supports the idea that mutational processes are stable over time and do not define the transition to MM. In order to further dissect out the contribution of mutational signatures we clustered mutations into five groups based on their respective CCF. Overall, the contribution of mutational signatures did not vary substantially according to clonality, **5.3.2.3.1.B**. We identified mutations in genes with a CCF that increased from intermediate (CCF 0.4-0.8) to clonal (0.8-1) such as *KRAS, CHD2, ABCC2, TNIP1, TRPS1, RCCD1, MTRR, ERCC6,* and from low CCF (0-0.4) to medium (0.4-0.8) such as *RMI1, PSMA8 CAMK1D, GALNT2, PON3, SALL4*, that could be potential drivers **5.3.2.3.1.C.**

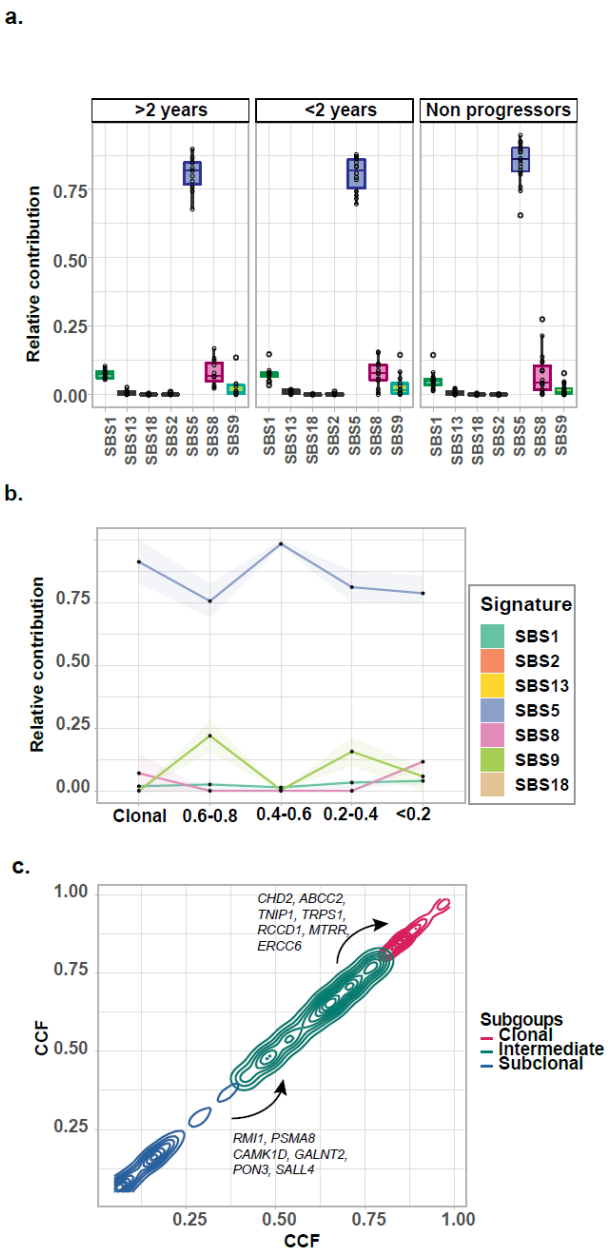


Figure 5.3.2.3.1. Mutational signatures. A. The mutational signature composition does not differ between patients that progress more or less than two years after the sample was taken and those who do not progress. B. The contribution of mutational signatures did not vary substantially according to clonality. C. The CCF of genes increased overtime suggesting they may be drivers.

5.3.2.4. Changes in Sub-Clonal Architecture Precede Progression.

The clinical management of SMM is different from NDMM in that it is possible to monitor progression over time and to reassess the disease at a number of occasions without impairing clinical outcomes. Therefore, we determined whether it may be clinically relevant to monitor sub-clonal structure as a means of predicting progression. We reconstructed sub-clonal structure and followed the size of each clone over time using Pyclone in eight patients. A median of eight clones per patient was identified, most related via branching patterns (7/8) with only one case being associated with a linear pattern. Five clones made up 90% of the tumour in 84% of cases, and six clones in 16% of cases. In the year prior to biochemical progression significant changes in sub-clonal structures were seen in all evaluable patients.

When analysing the five samples from Patient A, we see evidence, a year before progression, of an increase in the CCF of a cluster containing a *SOX2* mutation. Using Pyclone, to further cluster the data, and adjust for contamination and depth variation, we identify a branching evolutionary pattern with 6/7 clones being present at a significant level at the earliest timepoint. Several mutations in known drivers were seen in patient including in genes such as *KMT2C*, *IRF4*, *NF1*, and *DNMT3A.* Finally, in the MM sample, we see the expansion of a clone containing an *FUB1* mutation that encodes for a ubiquitin-like protein. Interestingly, these changes in sub clonal structure are matched by progressive changes in the paraprotein levels, **Figure 5.3.2.4.1**.

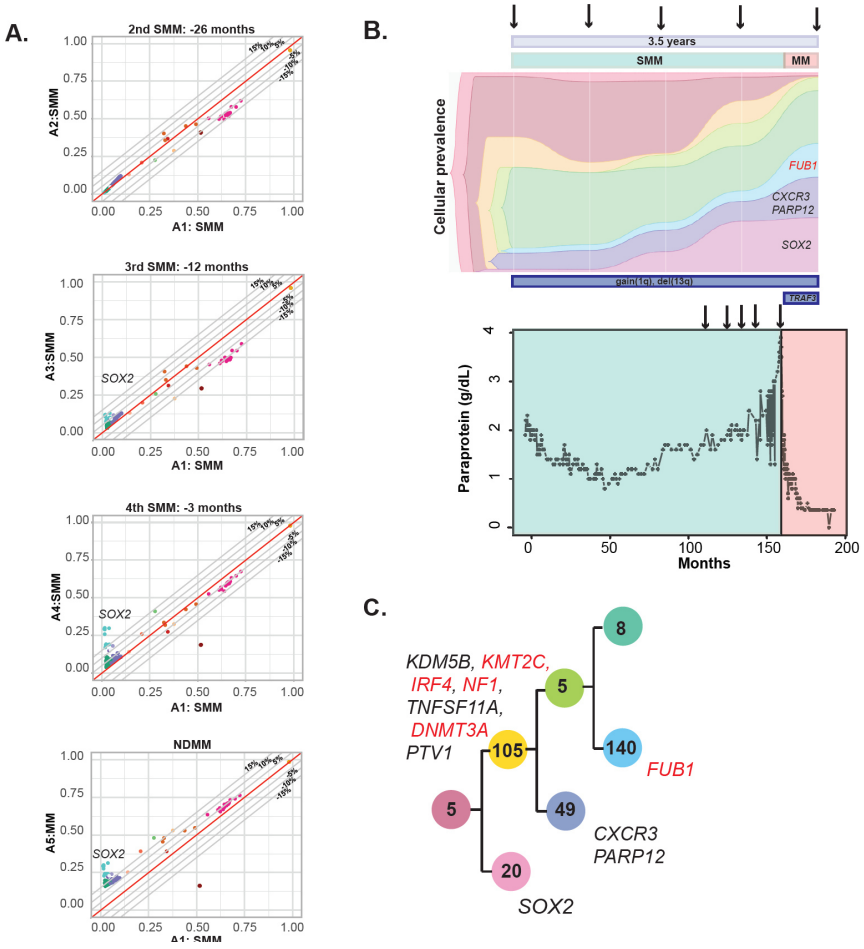


Figure 5.3.2.4.1. A. CCF plot of patient A showing the emergence of a *SOX2* cluster before progression. Patient and sample number indicated on axis. The dotted lines represent the percent of difference in CCF from the angle bisector line which represents perfect identity between the samples. B. Fish-plot summarizing the clonal evolution in parallel to the paraprotein evolution in patient A. C. Phylogeny tree showing branching evolution of patient A.

The five samples from patient B span over a seven-year period including two SMM samples. At the copy number level, that patient had a del(16q) that was lost after the second sample. There was no change in the CCF of mutations between the pre-progression samples but the last SMM sample was taken five years prior to the progression. These samples also displayed a branching evolutionary pattern, with multiple clones present from the first sample. The relapse samples showed an increase in CCF of clusters containing mutations in driver genes such as *ZFP36L1* or known oncogenes such as *MAPK3*, **Figure 5.3.2.4.2.**

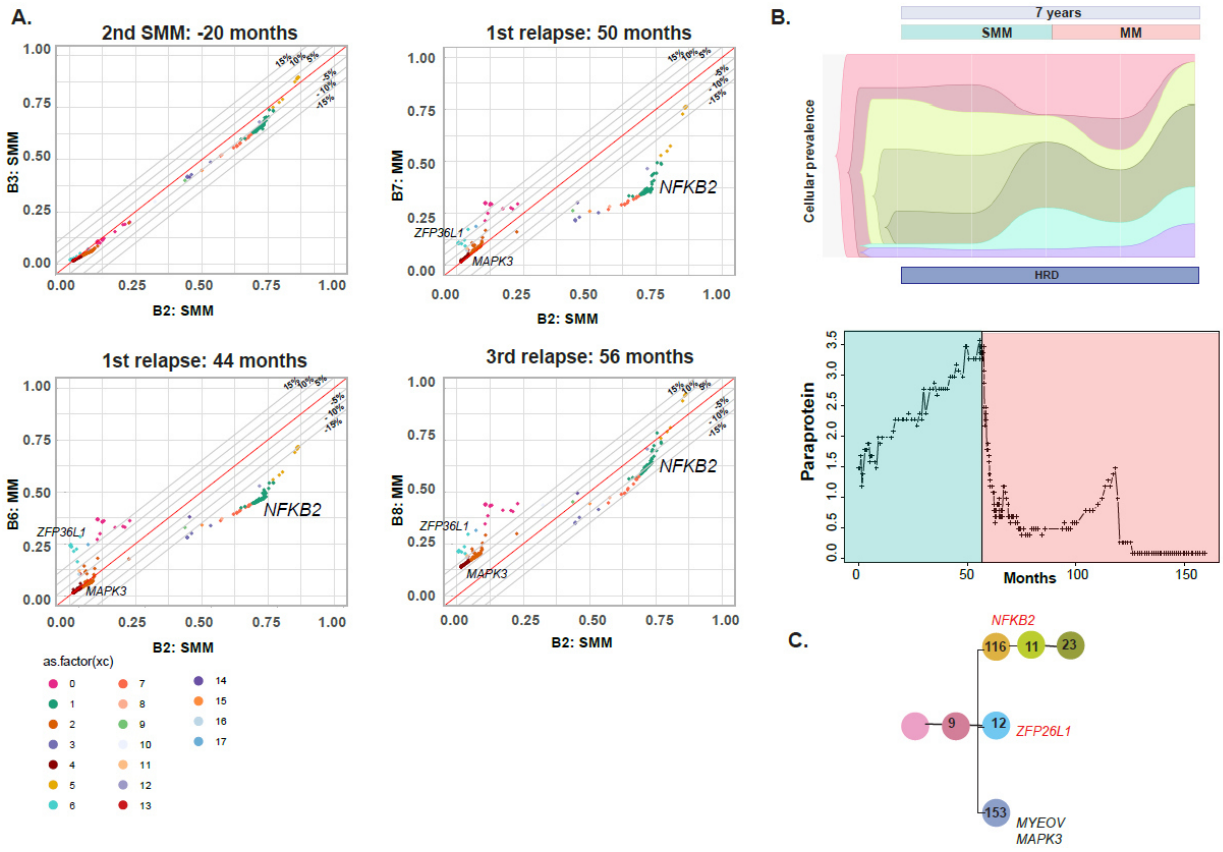


Figure 5.3.2.4.2. Genomic evolution of Patient B. A. CCF plot showing the emergence of *ZFP36L1* clone. B. Fishplot summarising the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution

5. Smouldering Myeloma

We analysed seven samples from Patient C that span over a 5.5 years period including four SMM samples. That patient acquired a gain(1q) and del(11q) in the 7th sample, and a del(13) from the 6th samples. Patient C also presented a del(5q) in the 4th sample that was not carried through to the next samples. Before progression, a cluster containing mutations in genes such as *NF1, RMI1, SLL4, TGM4*, and *ZDHHC19* increased in size, and a clone containing the *KDM2A* mutation decreased in size suggesting this mutation in this Lysine Demethylase did not drive progression. Like the first two patients, we saw a branching evolutionary pattern with 4/7 clones been present at a significant level early on, **Figure 5.3.2.4.3.**

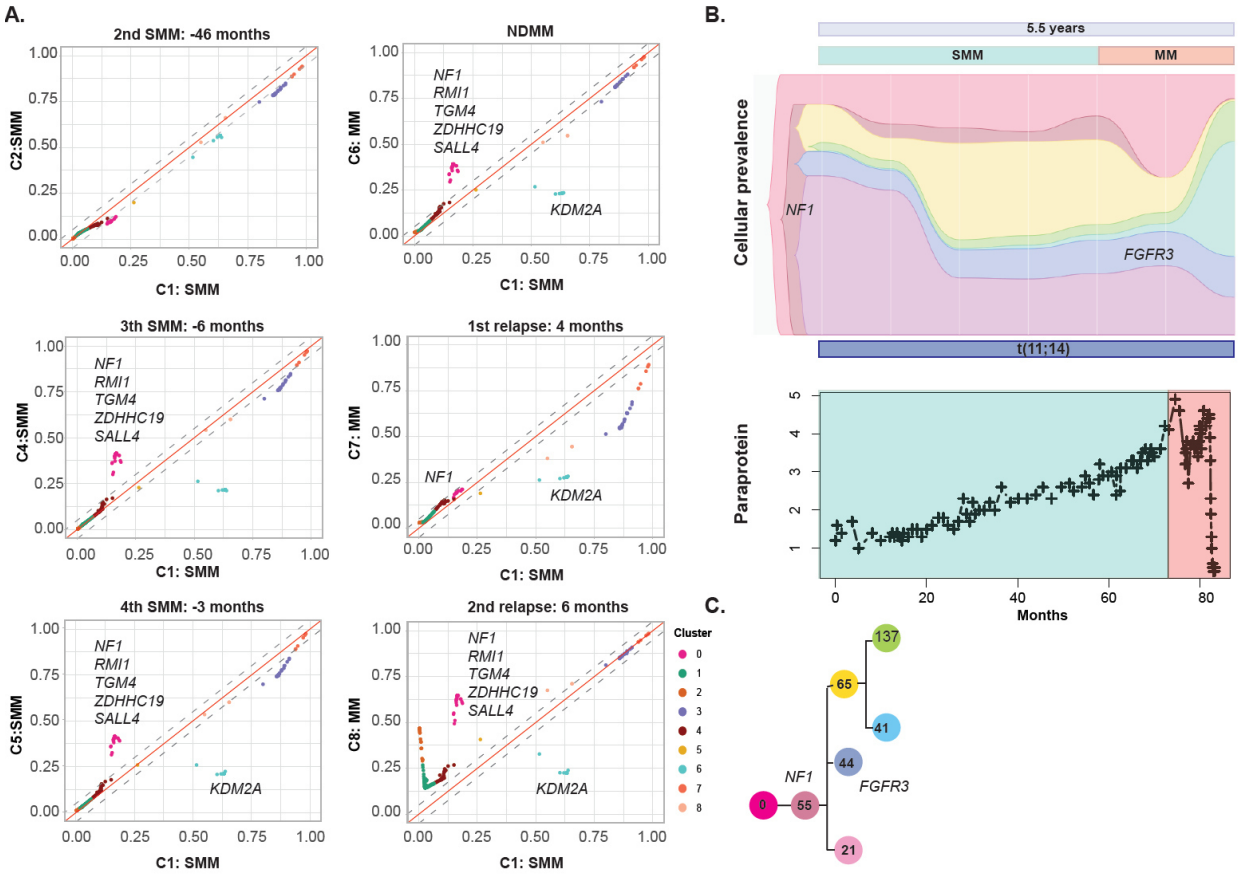


Figure 5.3.2.4.3. Genomic evolution of Patient C. A. CCF plot showing the emergence of an *NF1* clone. B. Fishplot summarizing the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution

We analysed five SMM samples from Patient D, that span over a 4.4 years period. Patient D has not progressed and did not display any changes in copy numbers. The proportion of the founding clone seen at the second timepoint, would suggest a significant normal plasmacell contamination and should be excluded from the analysis. We found both a linear with a secondary branching evolutionary pattern and a clone bearing a *ARID1A* mutation emerged in the last sample that did not correlate to any significant change in the paraprotein, **Figure 5.3.2.4.4.**

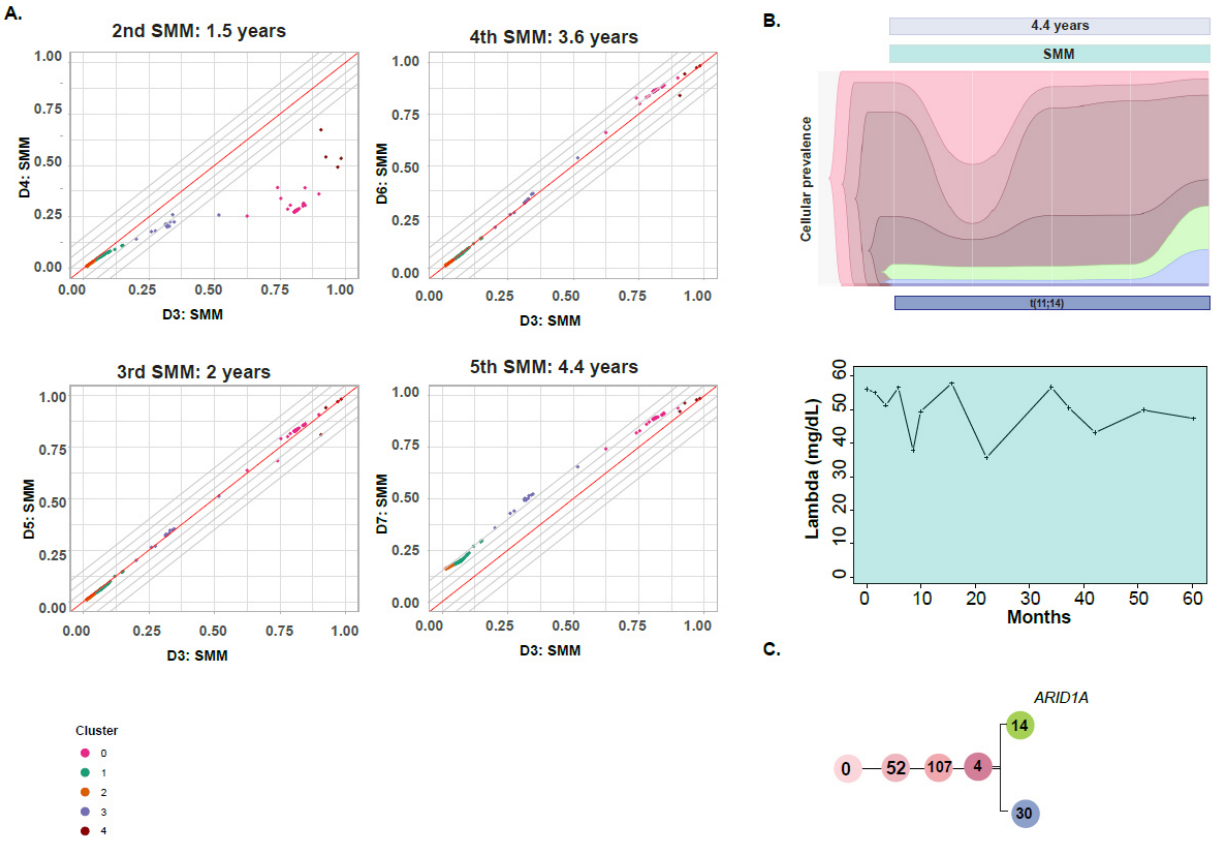


Figure 5.3.2.4.4. Genomic evolution of Patient D. A. CCF plot showing the absence of clonal selection in a patient that has yet to progress. B. Fishplot summarizing the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution.

5. Smouldering Myeloma

The nine samples from patient E were all taken at the SMM stage, spanning over a four-year period, the last one taken ten months before progression to MM. There was evidence supporting the acquisition of a del(22p) from the 8th sample and a del(5q) in the 2nd sample that was not carried through to the following samples. At the mutational level we identified the emergence of two clusters: the first containing a *KRAS* mutation the second containing *PREX1* and *MAP3K6* mutations that increases in CCF from the 6th sample. The Pyclone analysis, suggests a combined linear and branching evolutionary pattern, **Figure 5.3.2.4.5.**

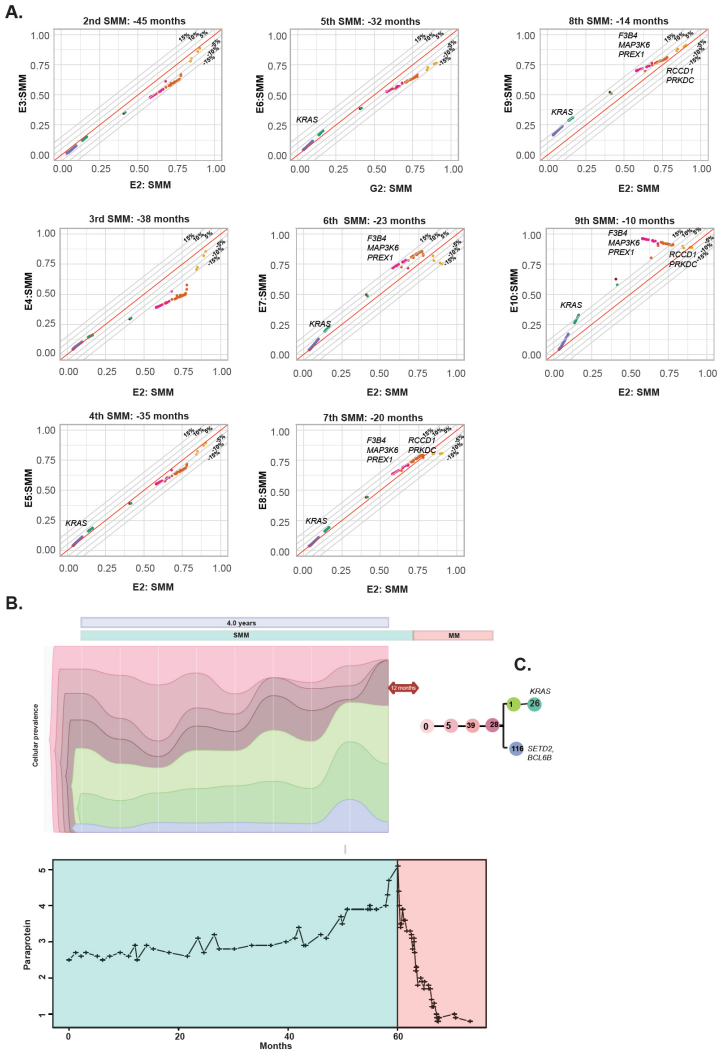


Figure 5.3.2.4.5. Genomic evolution of Patient E. A. CCF plot showing the emergence of a *KRAS* and *MAPK6* clone. B. Fishplot summarizing the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution.

5. Smouldering Myeloma

The seven samples from patient F were all taken at the SMM stage, spanning over a 5.5-year period. To date, that patient has not progressed. There was no evidence for changes in CNA across these samples. There was no obvious change in subclonal structure until the last sample that displayed an increase in three clones: the first containing a *KRAS* mutation, the second containing both the *KRAS* and *PARP10* mutation, and the third containing *CD40*, *IRAK1*, and *HDAC11* mutation. The Pyclone analysis, also suggest a combined linear and branching evolutionary pattern, **Figure 5.3.2.4.6.**

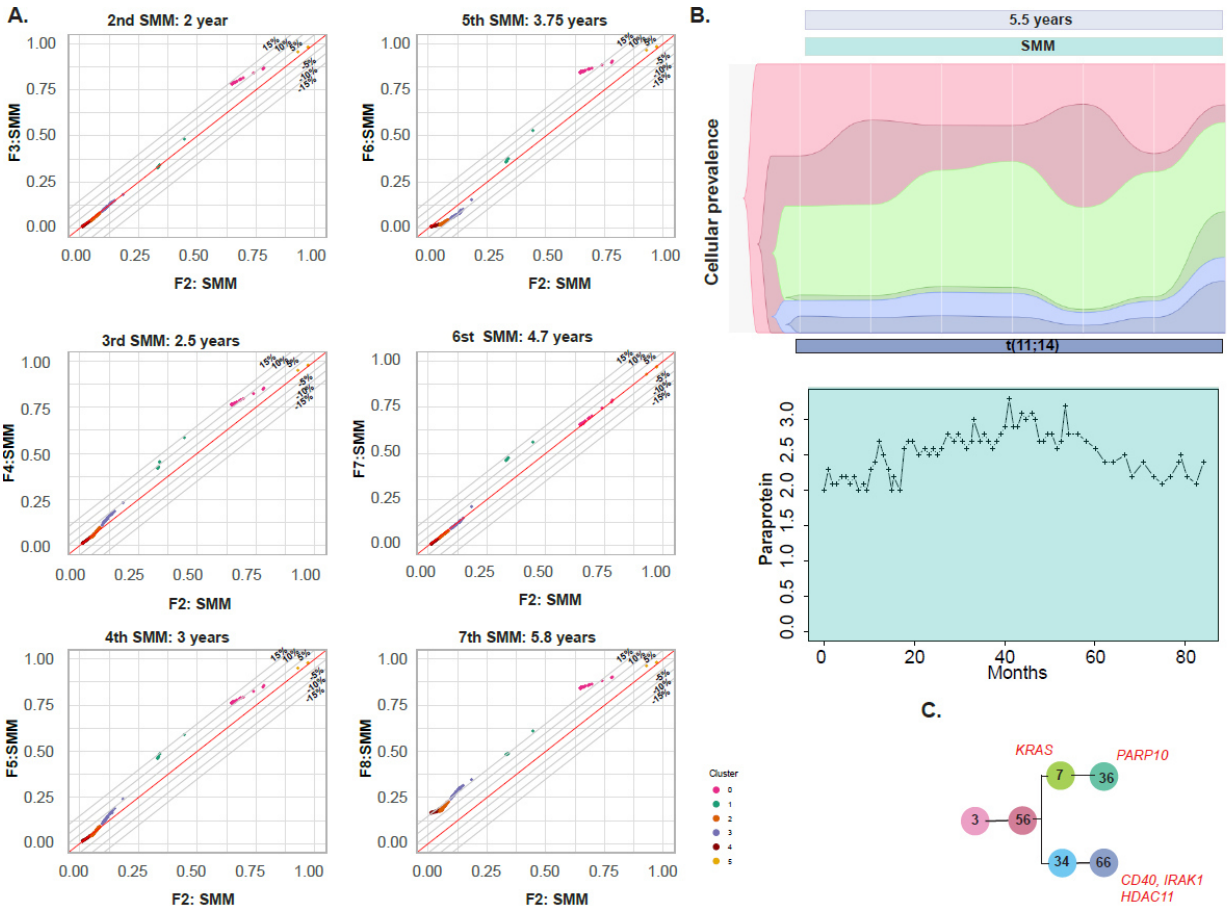


Figure 5.3.2.4.6. Genomic evolution of Patient F. A. CCF plot showing stable clonal composition. B. Fishplot summarizing the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution0

Patient G had two samples taken six months apart: one at the EM stage and one at the MM stage. That patient had a del(14) that was only seen at the EM stage. There was evidence supporting the increase of a clone caring a *TNFSFR1B* mutation. The Pyclone analysis, also suggest a combined linear and branching evolutionary pattern, **Figure 5.3.2.4.7.**

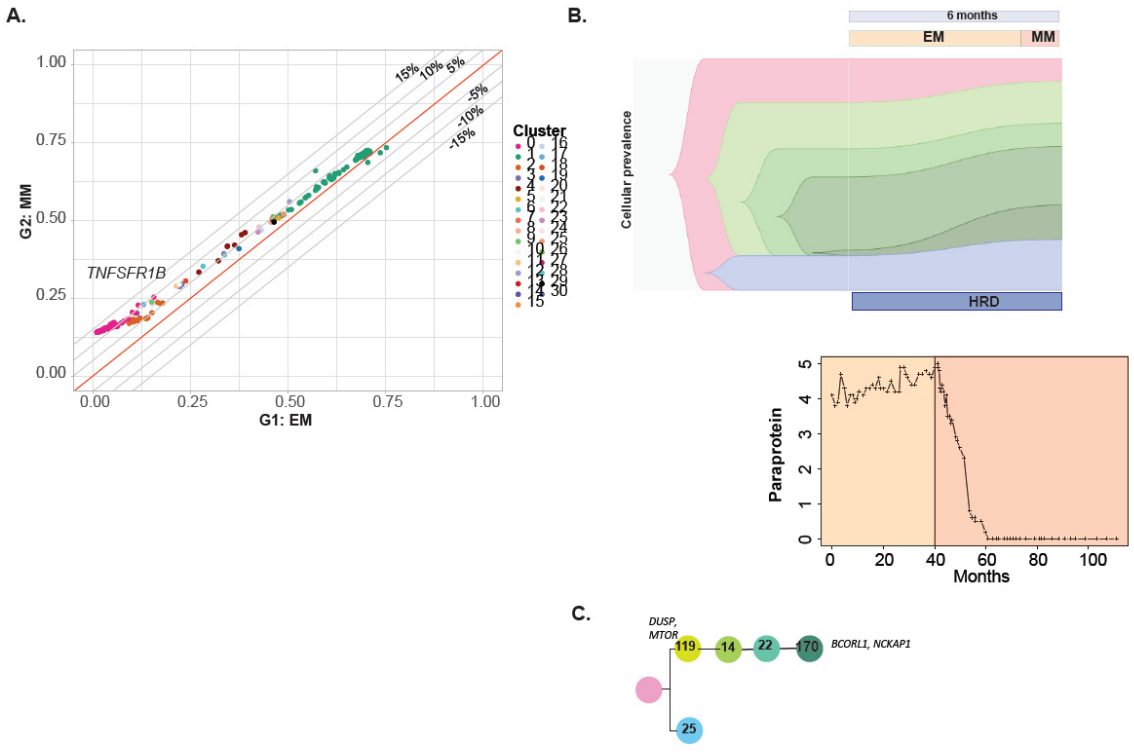


Figure 5.3.2.4.7. Genomic evolution of Patient G. A. CCF plot showing the emergence of a *TNSFR1B* clone. B. Fishplot summarizing the clonal evolution in parallel to the paraprotein evolution. C. Phylogeny tree showing branching evolution

The four samples from patient I were all taken at the SMM stage, spanning over a 4.5-year period, the last taken eight months before progression to MM. There was no evidence for changes in CNA across these samples. There seems to be an increase in the CCF of a cluster (A) and clone (B) carrying a *CDKN3* mutation. Interestingly the Pyclone analysis, suggested a linear evolutionary pattern, **Figure 5.3.2.4.8.**

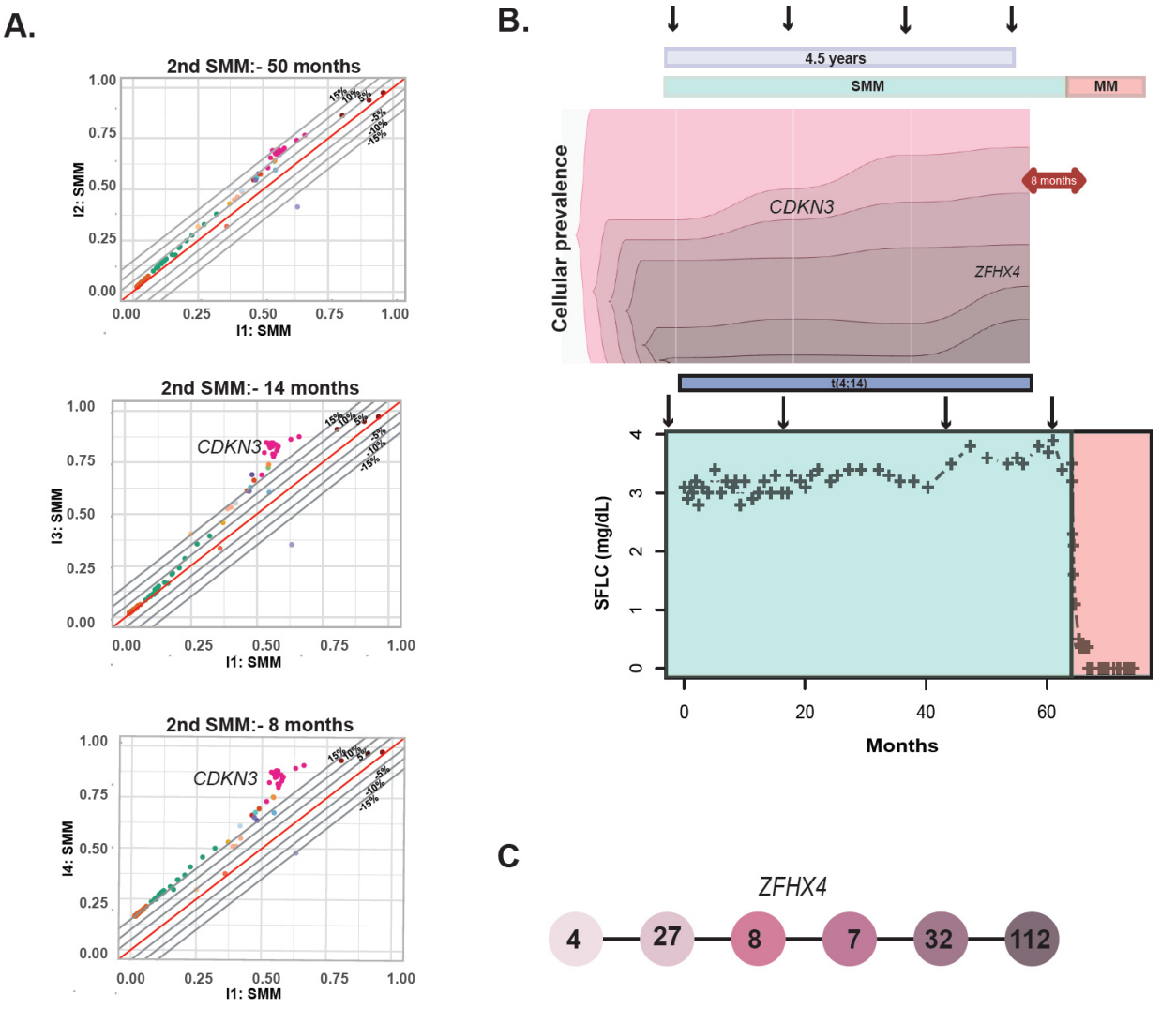


Figure 5.3.2.4.8. A. CCF plot showing the emergence of a *CDKN3* cluster before progression in patient I. B. Fish-plot summarizing the clonal evolution in parallel to the paraprotein evolution in patient I. V. Phylogeny tree showing linear evolution in patient I.

Overall, these data suggest that monitoring sub-clonal structure as a means of disease assessment is possible and could direct clinical intervention at a time long before end organ damage or clinical symptoms develop, highlighting the importance of bone marrow monitoring of SMM patients.

5.3.2.5. Clonal Diversity is a Marker of Time to Progression.

These early disease stages can be considered as forming a distinct ecosystem and as such can be studied using ecological tools. The Shannon Diversity index (Allaby, 2009) is one such tool that has been extensively used and we applied this approach to each of the sequential SMM samples. This analysis showed that samples from patients that progressed had a higher index than those that did not (χ^2=11, p=0.0006). Further, we showed a trend where patients that progressed had a stable index while those that did not progress had evidence of ongoing change. These observations are consistent with a high diversity index that is stable being associated already transformed disease, **Figure 5.3.2.5.1**. In contrast, in cases of SMM increasing diversity is seen over time reflecting ongoing diversification of the disease. These observations suggest that applying the Shannon diversity index to serial samples may be able to identify cases at high risk of progression.

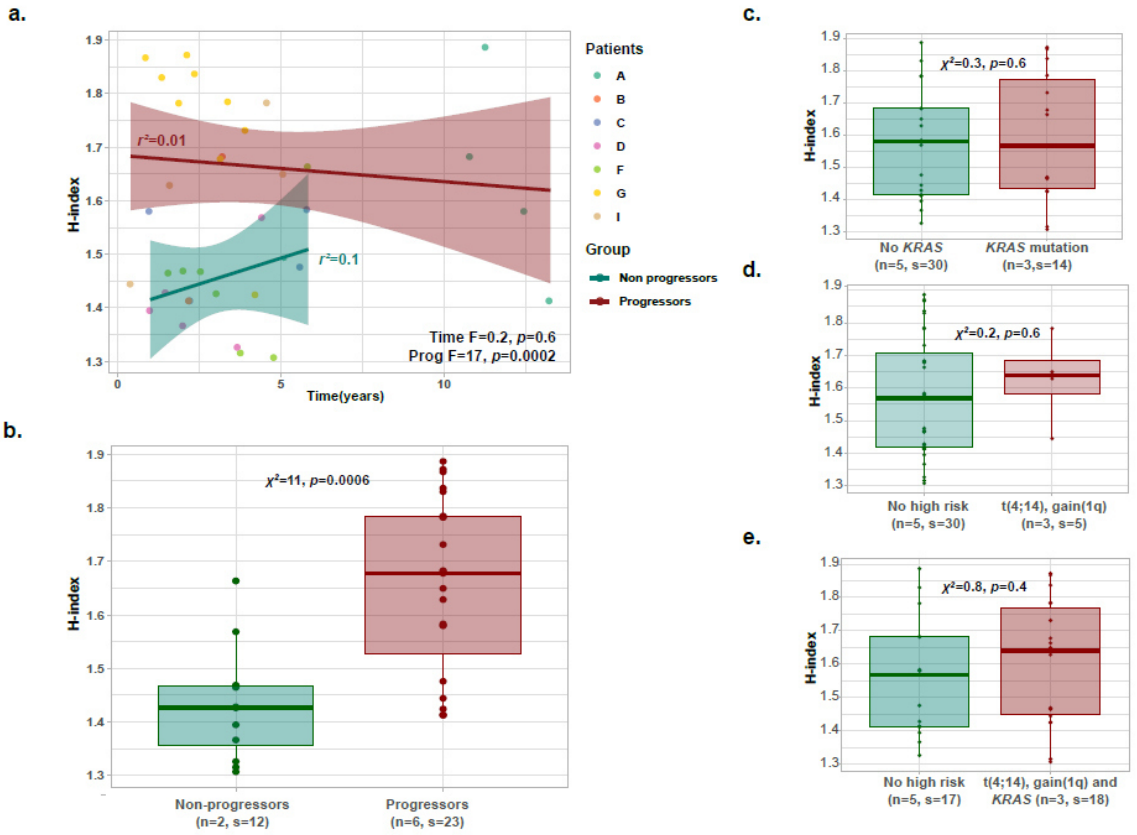


Figure 5.3.2.5.1. Shannon diversity indices overtime. There was no linear correlation with time but patients that progressed had a stable H index whereas those that did not seemed to present changes in H index (A). Patients that progressed had significantly higher H indices than those that did not (B). There were no differences in H index according to *KRAS* mutations (C), high risk cytogenetics (D) or both (E)

5.3.3. At the SMM Stage, Driver Processes Vary by Molecular Subgroup.

In order to elucidate mechanisms of progression, we investigated 225 SMM and 1348 MM from UAMS for which we had baseline gene expression profiling (GEP) using the Illumina microarrays.

The time to therapy did not differ between Translocation Cyclin-D (TC) subgroups, **Figure 5.3.3.1.A**. Due to a low number of samples and the infrequency of some subgroups, particularly in SMM, we did not study the *CCND3* and *MAFB* groups. The breakdown of the different TC subgroups was similar between SMM and MM with a trend suggesting fewer *MMSET* and D1 patients in SMM, **Figure 5.3.3.1.B**. We selected five hundred TC-matched MM patients and compared them to 225 SMM and looked at the patterns of gene expression within each of the major TC molecular subgroup to test the

hypothesis that at the SMM stage that differential pathway controls distinguish disease subtypes. We identified three pathways significantly upregulated in a comparison of SMM with MM. A major feature unifying these groups was increased proliferation at the MM stage with important changes in DNA conformation change, and DNA metabolic processes being seen, **Figure 5.3.3.1.C.**

Among the *CCND1* (t(11;14)) patients, pathways relating to cell division (chromosome segregation, nuclear division, DNA replication, centromere assembly, DNA replication, cell division, spindle organisation, checkpoints and cycle transition) together with amino acid metabolic processes were the most upregulated. Pathways related to normal plasma cell function such as defence response, chemotaxis and inflammation response were downregulated, **Figure 5.3.3.1.D**.

In contrast to the t(11;14) subgroup, the expression profile of MMSET patients was very different with the main dysregulated pathways being those of cell adhesion (upregulated), and downregulation of the DNA replication. These results support previous hypotheses that this group is rapidly driven to a clinical disease phenotype and that even at the SMM stage that they already have features of aggressive disease and as such could be considered at high-risk of rapidly transforming to MM, **Figure 5.3.3.1.F.** This observation supports observations made by the IMWG where the t(4;14) is considered a high-risk subgroup.

The MAF group differs from the *MMSET* group in that while it is considered high-risk in NDMM paradoxically in MGUS it is a good prognostic feature. In the MAF subgroup of SMM we show that cell division was upregulated and normal plasma cell function was downregulated. Interestingly, the DNA repair pathway was upregulated in MM, **Figure 5.3.3.1.E**. In the two hyperdiploidy groups (D1 and D2) no downregulated pathways were seen. In the D1 subgroup, pathways related to cell cycle and DNA repair were upregulated and in the D2 cell cycle and protein ubiquitination pathways, **Figure 5.3.3.1.G-H.** We show, consistent with previous data showing different oncogene dependencies in NDMM, that subgroup specific differences exist even at the SMM stage of disease with different biological features driving progression within each TC molecular subgroup. Importantly,

if intervention strategies are developed, they should potentially be subtype specific or at least subtype specific therapeutic effects may be expected.

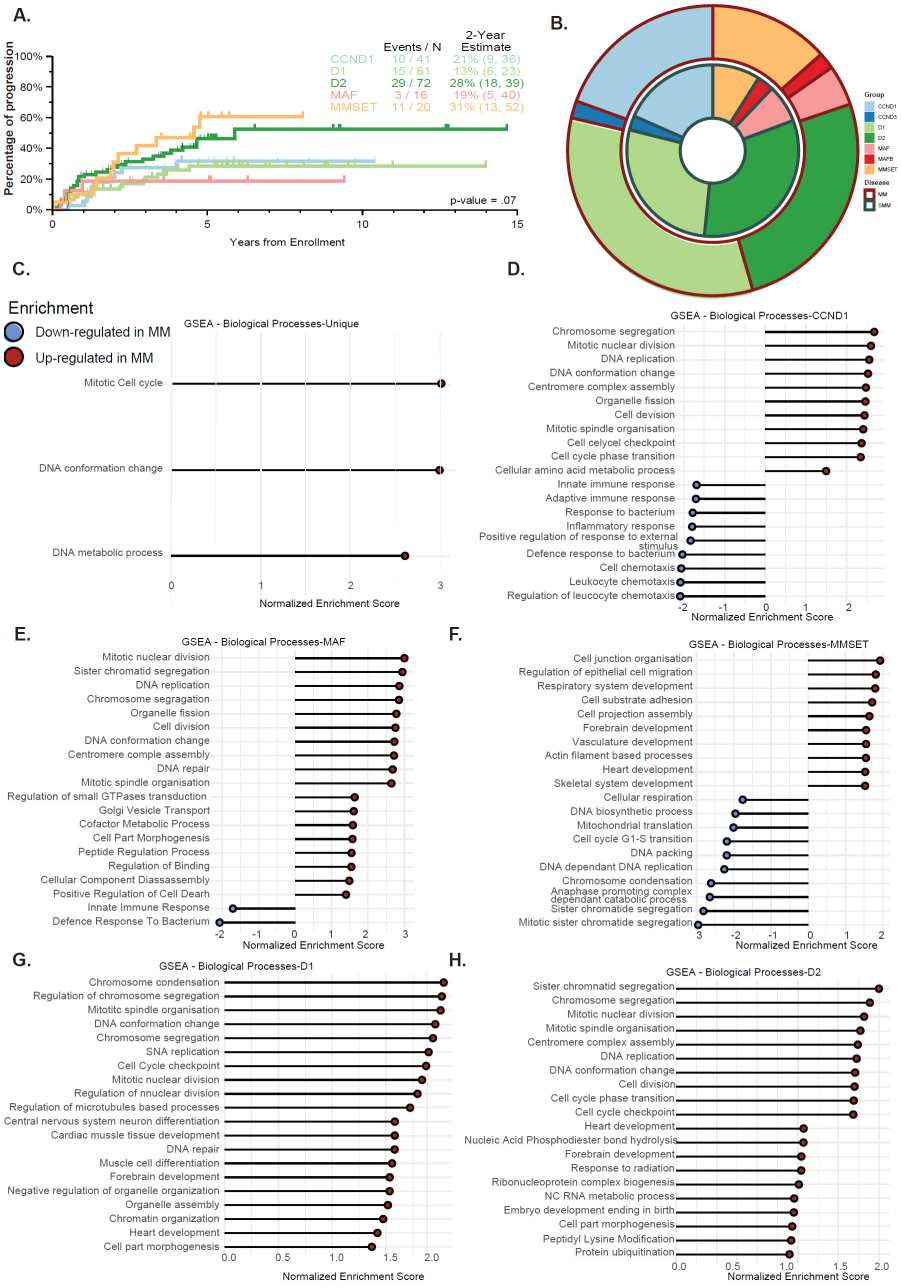


Figure 5.3.3.1: Patterns of expression changes differ among cytogenetic subgroups. A. Outcome of patients depending on their TC subgroup B. Proportion of TC groups in the different disease states (centre SMM, periphery MM), C. Pathways enriched overall between SMM and MM. D. Pathways enriched in the *CCND1* group between SMM and MM. E. Pathways enriched in the MAF group between SMM and MM. F. Pathways enriched in the *MMSET* group between SMM and MM. G. Pathways enriched in the D1 group between SMM and MM. H. Pathways enriched in the D2 group between SMM and MM

5.4. Discussion.

` Managing the premalignant phases of MM is becoming a major clinical goal and the key to doing this safely and effectively is to define the optimum time for intervention. Using genetic variants to identify cases of SMM at substantial risk of transformation before the development of end organ damage is a time point where early intervention may be important. Further, strengthening the definitions of SMM cases suitable for intervention could improve the health care value of early intervention. Utilizing a-priori information on drivers of MM and analysing their impact in SMM we suggest that a subset of these molecular events, including *KRAS, NRAS* and *FAM46C* (Zhu *et al*, 2017) define symptomatic behaviour and when present in the asymptomatic disease phase represent cases in the process of transformation to MM. This conclusion is supported by the analysis of the interaction of the variants with sub-clonal structure based on a unique series of sequential samples from SMM where there is ongoing sub-clonal expansion ultimately leading to a change in clinical behaviour. An alternative interpretation may be that high-risk MM evolves more rapidly through a shorter smoldering phase thus explaining why it is under-represented in SMM. In contrast to the genetic hits outlined above, NF-κB dependency at this disease stage do not have an impact on time to progression and are not associated with sub- clonal expansion. Thus, while these events occur at early molecular timepoints and contribute to the genetic complexity of MGUS, they are not associated with symptomatic behaviour.

Further, from a clinical perspective we identify an important role for *KRAS* mutations as a prognostic factor that can significantly contribute to risk assignment in SMM (Walker *et al*, 2015). With a hazard ratio for progression of 3.8, *KRAS* mutations are a critically important molecular prognostic factor. A recent study (Bustoros *et al*, 2020) identified a similar impact confirming the prognostic relevance of these molecular variants. It is important to consider why such a prognostic effect is seen in SMM but not in MM and this may reflect the impact on cell-cycle progression promoting a clonal sweep and transition to MM. Thus, cases with KRAS mutations may be better defined as MM in the process of transition and may, therefore, be better grouped with MM rather than MGUS. Consistent with this hypothesis, prior studies using the plasma cell labelling index

have found significant differences between precursor phases and NDMM.(Madan *et al*, 2010)

Further evidence for the role of a proliferation advantage in symptomatic behaviour and with time to progression in SMM comes from our identification of a case of SMM which acquired a translocation at the *MYC* locus. The cancer clonal fraction of the subclone, defined by this abnormality, sequentially increased over time. We were, however, unable to identify a prognostic impact of *MYC* rearrangements, as has been suggested previously (MMRF CoMMpass Network *et al*, 2020; Bustoros *et al*, 2020). We speculate that this difference is a reflection of the exclusion of many such events in this study by the application of the stringent new diagnostic definitions of SMM. Consistent with this idea, the current study had a lower incidence of *MYC* translocated cases than was seen in other previous series (Misund *et al*, 2020).

Knowledge of the role of structural events in the precursor stages of disease has been limited. Here, we show that one of the major initiating events of MM, HRD, is stable and does not constitute a significant mechanism impacting time to progression. In contrast, segmental copy number gains and losses together are seen fluctuate overtime consistent “clonal tiding” of sub-clones defined by these abnormalities. In this dataset the sites of these events seem to be restricted to the sites of known tumour suppressor genes identified in MM such as on chromosome 16p, which is one of the most common features in MM at 30%% but only in 18% of SMM. We identified significant differences in the frequency of copy number changes between SMM and MM providing further support with their role as drivers. However, these differences may be explained by the different composition of molecular subgroups in the asymptomatic disease stages. For example, there are fewer t(4;14) patients in the SMM dataset compared to MM, and therefore the frequency of copy number changes associated with this translocation, such as del(16q), may be impacted by this.

Knowledge of pathway dysregulation within specific TC groups highlight disease specificities where each subgroup behaves and progresses differently. The biggest changes were seen in the MAF subgroups: this could be the resultant of the APOBEC activity responsible for high levels of somatic mutations, DNA instability, and probably

drug resistance at the MM level. This may explain the adverse prognosis of the maf MM and the relative indolent phenotype of the maf SMM. The MMSET SMM patients display very few differences with their MM counterparts suggesting they have already in transformation, consistent with them being high-risk.

Knowledge of the molecular events underlying the genetic complexity of the early truncal stages of MM has largely been extrapolated from the study of NDMM. Here we have used sequential samples from the early disease stage of SMM to improve the resolution of the definitions of sub-clonal structure. In order to understand the mechanistic basis driving mutational abnormalities at these early truncal stages of disease we used the mutation data to identify signatures and studied their relationship overtime. We identified the background signature SBS1 and 5 and show that these are stable over time and were not influenced by their relationship to the time of transformation. In 2015, we first reported a significant contribution of APOBEC hypermutation signatures in the *MAF* and *MAFB* translocated newly diagnosed subgroups, where they made-up a median of 58% and 44% of the total mutation in the t(14;16) and t(14;20) respectively. We show that the proportion of APOBEC mutations is higher in MM than in SMM with a trend toward lower levels in SMM mafs than MM mafs. These observations are consistent with there being two distinct levels of APOBEC signatures in MM. These data suggest that an APOBEC mutational signature may be associated with progression from SMM to MM, in keeping with this we observe a trend towards more rapid progression to MM in cases with an APOBEC signature over the level of 5%.

The evolutionary relationship of subclones in precursor stages is critical to disease development and its clinical behaviour. While prior reports have studied paired pre- and post-progression samples none have had access to multiple sequential samples. The analysis of multiple samples in this analysis has allowed us to more precisely define the clonal architecture using Pyclone. The results of this analysis highlight the importance of branching evolutionary pathways at these asymptomatic disease stages. We show that these changes in clonal substructure can be used to monitor the asymptomatic stage of SMM end organ damage has developed (Rasche *et al*, 2017). An increase in a sub-clonal fraction occurs before biochemical progression and the development of end organ

damage. This observation is a critical advance because with such changes preceding clinically relevant events by more than one or two years a safe therapeutic window can be defined. Such an approach could prevent the potential adverse impacts of therapeutic intervention and restrict such risk to a time point with the maximal risk benefit ratio is present**.**

A knowledge of the genetic basis of SMM can allow us to better define and monitor early disease stages. The results of this study suggest that the presence of one of a limited number of the typical drivers of MM can identify asymptomatic cases in the process of transition that are destined to behave a clinical MM within a limited time. The evidence supports the use of *KRAS* mutations as markers of more rapid transition to MM, which could readily be incorporated into clinical definitions, **Figure 5.4.1.** From a clinical perspective, monitoring sub-clonal structure over time offers a clinically useful means of directing early clinical interventions. Going forward with improved definitions of SMM it may be possible to identify a group of monoclonal gammopathy and early MM where treatment is indicated.

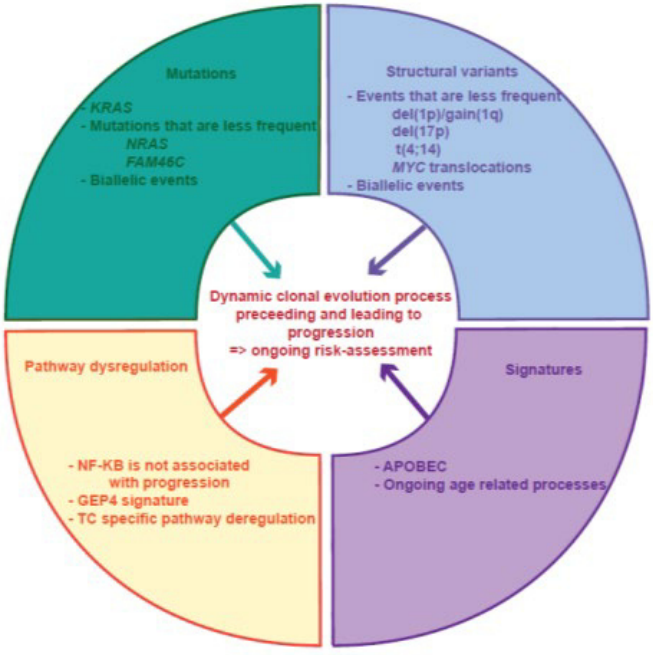


Figure 5.4.1. Summary figure of the different changes associated with SMM progression

5.5. References

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., Islam, S.M.A., Lopez-Bigas, N., Klimczak, L.J., McPherson, J.R., Morganella, S., Sabarinathan, R., Wheeler, D.A., Mustonen, V., Getz, G., Rozen, S.G., et al (2020) The repertoire of mutational signatures in human cancer. Nature, 578, 94–101.

Allaby, M. (2009)A Dictionary of Zoology Oxford University Press Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199233410.001.0001/acref-9780199233410.

Annunziata, C.M., Davis, R.E., Demchenko, Y., Bellamy, W., Gabrea, A., Zhan, F., Lenz, G., Hanamura, I., Wright, G., Xiao, W., Dave, S., Hurt, E.M., Tan, B., Zhao, H., Stephens, O., Santra, M., Williams, D.R., Dang, L., Barlogie, B., Shaughnessy, J.D., Jr, et al (2007) Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. Cancer cell, 12, 115–130.

Berger, A.H., Knudson, A.G. & Pandolfi, P.P. (2011) A continuum model for tumour suppression. Nature, 476, 163–169.

Bolli, N., Maura, F., Minvielle, S., Gloznik, D., Szalat, R., Fullam, A., Martincorena, I., Dawson, K.J., Samur, M.K., Zamora, J., Tarpey, P., Davies, H., Fulciniti, M., Shammas, M.A., Tai, Y.T., Magrangeas, F., Moreau, P., Corradini, P., Anderson, K., Alexandrov, L., et al (2018) Genomic patterns of progression in smoldering multiple myeloma. Nature Communications, 9, 3363.

Boyd, K.D., Ross, F.M., Chiecchio, L., Dagrada, G., Konn, Z.J., Tapper, W.J., Walker, B.A., Wardell, C.P., Gregory, W.M., Szubert, A.J., Bell, S.E., Child, J.A., Jackson, G.H., Davies, F.E. & Morgan, G.J. (2012) A novel prognostic model in myeloma based on co-segregating adverse FISH lesions and the ISS: analysis of patients treated in the MRC Myeloma IX trial. Leukemia, 26, 349–355.

Boyle, E.M., Ashby, C., Tytarenko, R., Deshpande, S., Wang, Y., Sawyer, J., Tian, E., Johnson, S., Rutherford, M.W., Wardell, C.P., Bauer, M.A., Thanendrarajan, S., Schinke, C., Zangari, M., van Rhee, F., Wang, H., Rosenthal, A., Hoering, A., Flynt, E., Thakurta, A., et al (2020) BRAF and DIS3 Mutations Associate with Adverse Outcome in a Long-term Follow-up of Patients with Multiple Myeloma. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research.

Bustoros, M., Sklavenitis-Pistofidis, R., Park, J., Redd, R., Zhitomirsky, B., Dunford, A.J., Salem, K., Tai, Y.-T., Anand, S., Mouhieddine, T.H., Chavda, S.J., Boehner, C., Elagina, L., Neuse, C.J., Cha, J., Rahmat, M., Taylor-Weiner, A., Van Allen, E., Kumar, S., Kastritis, E., et al (2020) Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. Journal of Clinical Oncology, JCO.20.00437.

Giraud, M., Salson, M., Duez, M., Villenet, C., Quief, S., Caillault, A., Grardel, N., Roumier, C., Preudhomme, C. & Figeac, M. (2014) Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. BMC genomics, 15, 409.

Hillengass, J., Fechtner, K., Weber, M.-A., Bäuerle, T., Ayyaz, S., Heiss, C., Hielscher, T., Moehler, T.M., Egerer, G., Neben, K., Ho, A.D., Kauczor, H.-U., Delorme, S. & Goldschmidt, H. (2010) Prognostic significance of focal lesions in whole-body magnetic resonance imaging in patients with asymptomatic multiple myeloma. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28, 1606–1610.

International Myeloma Working Group (2003) Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. British Journal of Haematology, 121, 749–757.

Khan, R., Dhodapkar, M., Rosenthal, A., Heuck, C., Papanikolaou, X., Qu, P., van Rhee, F., Zangari, M., Jethava, Y., Epstein, J., Yaccoby, S., Hoering, A., Crowley, J., Petty, N., Bailey, C., Morgan, G. & Barlogie, B. (2015) Four genes predict high risk of progression from smoldering to symptomatic multiple myeloma (SWOG S0120). Haematologica, 100, 1214–1221.

Landgren, O. (2017) Shall we treat smoldering multiple myeloma in the near future? Hematology. American Society of Hematology. Education Program, 2017, 194–204.

Larsen, J.T., Kumar, S.K., Dispenzieri, A., Kyle, R.A., Katzmann, J.A. & Rajkumar, S.V. (2013) Serum free light chain ratio as a biomarker for high-risk smoldering multiple myeloma. Leukemia, 27, 941–946.

Madan, S., Kyle, R.A. & Greipp, P.R. (2010) Plasma Cell Labeling Index in the Evaluation of Smoldering (Asymptomatic) Multiple Myeloma. Mayo Clinic Proceedings, 85, 300 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2843123/ [Accessed April 16, 2020].

Misund, K., Keane, N., Stein, C., Asmann, Y., Day, G., Welsh, S., Wier, S., Riggs, D., Ahmann, G., Chesi, M., Viswanatha, D., Kumar, S., Dispenzieri, A., Gonzalez-Calle, V., Kyle, R., O'Dwyer, M., Rajkumar, S., Kortüm, K., Keats, J., Fonseca, R., et al (2020) MYC dysregulation in the progression of multiple myeloma. Leukemia, 34, 322–326.

MMRF CoMMpass Network, Misund, K., Keane, N., Stein, C.K., Asmann, Y.W., Day, G., Welsh, S., Van Wier, S.A., Riggs, D.L., Ahmann, G., Chesi, M., Viswanatha, D.S., Kumar, S.K., Dispenzieri, A., Gonzalez-Calle, V., Kyle, R.A., O'Dwyer, M., Rajkumar, S.V., Kortüm, K.M., Keats, J.J., et al (2020) MYC dysregulation in the progression of multiple myeloma. Leukemia, 34, 322–326.

Pérez-Persona, E., Vidriales, M.-B., Mateo, G., García-Sanz, R., Mateos, M.-V., de Coca, A.G., Galende, J., Martín-Nuñez, G., Alonso, J.M., de Las Heras, N., Hernández, J.M., Martín, A., López-Berges, C., Orfao, A. & San Miguel, J.F. (2007) New criteria to identify risk of progression in monoclonal gammopathy of uncertain significance and smoldering multiple myeloma based on multiparameter flow cytometry analysis of bone marrow plasma cells. Blood, 110, 2586–2592.

Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., et al (2014) International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. The Lancet. Oncology, 15, e538-548.

Rajkumar, S.V., Gupta, V., Fonseca, R., Dispenzieri, A., Gonsalves, W.I., Larson, D., Ketterling, R.P., Lust, J.A., Kyle, R.A. & Kumar, S.K. (2013) Impact of primary molecular cytogenetic abnormalities and risk of progression in smoldering multiple myeloma. Leukemia, 27, 1738–1744.

Rajkumar, S.V., Larson, D. & Kyle, R.A. (2011) Diagnosis of Smoldering Multiple Myeloma. New England Journal of Medicine, 365, 474–475 Available at: https://doi.org/10.1056/NEJMc1106428 [Accessed March 27, 2020].

Rasche, L., Chavan, S.S., Stephens, O.W., Patel, P.H., Tytarenko, R., Ashby, C., Bauer, M., Stein, C., Deshpande, S., Wardell, C., Buzder, T., Molnar, G., Zangari, M., van Rhee, F., Thanendrarajan, S., Schinke, C., Epstein, J., Davies, F.E., Walker, B.A., Meissner, T., et al (2017) Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. Nature Communications, 8, 268.

Rosiñol, L., Carrió, A., Bladé, J., Queralt, R., Aymerich, M., Cibeira, M.T., Esteve, J., Rozman, M., Campo, E. & Montserrat, E. (2005) Comparative genomic hybridisation identifies two variants of smoldering multiple myeloma. British Journal of Haematology, 130, 729–732.

Siontis, B., Kumar, S., Dispenzieri, A., Drake, M.T., Lacy, M.Q., Buadi, F., Dingli, D., Kapoor, P., Gonsalves, W., Gertz, M.A. & Rajkumar, S.V. (2015) Positron emission tomography-computed tomography in the diagnostic evaluation of smoldering multiple myeloma: identification of patients needing therapy. Blood Cancer Journal, 5, e364.

Sørrig, R., Klausen, T.W., Salomo, M., Vangsted, A.J., Østergaard, B., Gregersen, H., Frølund, U.C., Andersen, N.F., Helleberg, C., Andersen, K.T., Pedersen, R.S., Pedersen, P., Abildgaard, N., Gimsing, P. & Danish Myeloma Study Group (2016) Smoldering multiple myeloma risk factors for progression: a Danish population-based cohort study. European Journal of Haematology, 97, 303–309.

Walker, B.A., Boyle, E.M., Wardell, C.P., Murison, A., Begum, D.B., Dahir, N.M., Proszek, P.Z., Johnson, D.C., Kaiser, M.F., Melchor, L., Aronson, L.I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J.R., Brioli, A., Mikulasova, A., Cairns, D.A., Gregory, W.M., Quartilho, A., et al (2015) Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018) Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. Blood, 132, 587–597.

Walker, B.A., Wardell, C.P., Melchor, L., Brioli, A., Johnson, D.C., Kaiser, M.F., Mirabella, F., Lopez-Corral, L., Humphray, S., Murray, L., Ross, M., Bentley, D., Gutiérrez, N.C., Garcia-Sanz, R., San Miguel, J., Davies, F.E., Gonzalez, D. & Morgan, G.J. (2014) Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. Leukemia, 28, 384–390.

Zamagni, E., Nanni, C., Gay, F., Pezzi, A., Patriarca, F., Bellò, M., Rambaldi, I., Tacchetti, P., Hillengass, J., Gamberi, B., Pantani, L., Magarotto, V., Versari, A., Offidani, M., Zannetti, B., Carobolante, F., Balma, M., Musto, P., Rensi, M., Mancuso, K., et al (2016) 18F-FDG PET/CT focal, but not osteolytic, lesions predict the progression of smoldering myeloma to active disease. Leukemia, 30, 417–422.

Zhu, Y.X., Shi, C.-X., Bruins, L.A., Jedlowski, P., Wang, X., Kortüm, K.M., Luo, M., Ahmann, J.M., Braggio, E. & Stewart, A.K. (2017) Loss of FAM46C Promotes Cell Survival in Myeloma. Cancer Research, 77, 4317–4327 Available at: https://cancerres.aacrjournals.org/content/77/16/4317 [Accessed April 9, 2020].

Chapter 6: Telomere Maintenance and Aging Processes Impact the Outcome of Myeloma Patients

6.1. Summary

Multiple myeloma (MM) is a disease of the elderly with a median age at diagnosis of 69 years with nearly a third being diagnosed above the age 75, however, little is known about the impact of aging processes on outcomes. Frailty scores have been used to stratify elderly MM patients by functional status but could be improved by biological markers to enhance decision making, and guide treatment choices. In this work, we characterize the genetics of older MM patients compared to younger patients, and determine the associations of clonal haematopoiesis and telomere length with age. We characterised 980 NDMM patients with whole genome long insert sequencing samples for genetic variants, telomere length and clonal hemopoiesis and studied how they impact outcome with age. The results show that tumour telomere significantly impacts clinical outcome and can aid in personalized treatment decisions in older patients with MM.

6.2. Introduction

Multiple myeloma (MM) constitutes 1.5% of all malignant diseases, with an annual incidence rate of 6.9/100,000 accounting for 4,500 new cases each year in the UK (Home - Office for National Statistics) and 32,270 cases in the USA (Myeloma - Cancer Stat Facts). The median age at diagnosis is 69 years with 34.8% diagnosed above the age 75, and 9.6% above 85 years (Home - Office for National Statistics) (Ageing Europe — 2019 edition; Population Projections). Outcome in older populations is impaired compared to younger populations with an estimated overall survival (OS) at 5-years being approximately 60% for the 50-64 age group compared to 40% for those over 65. Outcome is even further impaired as patients age with the similar figures for the 70-79 age groups being 26% and for the 80-99 group of 14%(Home - Office for National Statistics). These

figures highlight the importance of optimizing therapeutic choices in older populations where outcomes reflect not only therapeutic efficacy but also the potential for toxicity as patients age.

Developing predictors to enhance therapeutic choices in older age groups is an important aim. Tumour acquired genetic variable has been used to define variability in clinical outcome and a number of well-known features have been incorporated into the R-ISS. However, only a few studies have looked at the impact of these variables in older populations where the spectrum of cytogenetic subgroups is also potentially different. Previous work has shown that the t(4;14) subgroup is less prevalent in older age groups while the t(11;14), gain(1q), HRD and del(17p) subgroups are constant (Avet-Loiseau *et al*, 2013; Pawlyn *et al*, 2020). Interpreting the clinical value of acquired genetic features in older age groups is compromised because of the impact of clinical factors such as age, ISS, performance status, and frailty. Previous work aimed at improving outcome prediction in older age groups has focussed on integrating quantitative variables of the aging process (Facon *et al*, 2018, 2020). Important examples of such variables include the study of clonal haematopoiesis (CH) which broadly describes the expansion of a clonal population of blood cells with one or more somatic mutations. Individuals with CH have been shown to be at greater risk of haematological malignancies, cardiovascular disease, and increased mortality from non-haematological cancers. However, little is known of role of clonal haematopoiesis in MM but in cancer mutations in a select number of drivers (*DNMT3A, TET2, ASXL1, JAK2, SF3B1, TP53, PPM1D, ATM, and CHEK2*) were found to be associated with adverse outcome which may be relevant to outcome prediction in aging populations with MM (Coombs *et al*, 2017).

Telomeres are protective tandem-repeated hexamers (TTAGGG) located at the termini of mammalian chromosomes and their attrition is a known hallmark of aging that associated with genomic instability, increasing risk of cancer and cellular senescence (López-Otín *et al*, 2013). Telomerase catalytically adds TTAGGG hexameric nucleotide repeats to the 3′-hydroxyl end of the telomeric leading strand, using a specific sequence in the RNA component as the template. The Telomerase Complex is composed of the enzyme telomerase reverse transcriptase (*TERT*), its RNA component (*TERC*), the

protein dyskerin (encoded by *DKC1*) and other associated proteins (*NHP2, NOP10*, and *GAR1*). Telomere length is a variable that is known to be impacted by age and where short telomeres have been associated with adverse outcomes in cancer (Wu *et al*, 2003a). Telomere shortening below a critical level can initiate cellular senescence or apoptosis or alternatively in the absence of these responses it can initiate chromosomal instability and neoplastic transformation (Shay & Bacchetti, 1997; Wu *et al*, 2003b). We reasoned that telomere shortening may reflect premature cellular aging and in contrast to the standard MM genetic prognostic events may contain significant information on how the tumour will behaves and thus constitute a significant prognostic factor.

The mutational spectrum of NDMM was consolidated via the publication of the 1273 patient myeloma genome project series (Walker *et al*, 2018). This dataset also contributed to the identification of the high-risk “double hit” population and the adverse prognostic impact of age over 65 (Walker *et al*, 2019). Despite its size the study data was predominantly focused on the exome and consequently has limits in the detection of other genomic features. Here we take advantage of a set of the data where WGS data is available such that we can impute structural variants, clonal hemopoiesis and telomere length all of which may contribute significantly to outcome in MM patients including in the older patients. Here, we have characterized the genetics of older MM patients, and determined their associations with age, clonal haematopoiesis and telomere length. We show that biological features that contribute to the aging processes such as telomere attrition and mutagenesis play an important role in the pathogenesis of MM and can be used to improve clinical decision making in older patients.

In this chapter, we characterized the genetics of older MM patients, and determined the associations of age with clonal haematopoiesis and telomere length, both of which have been shown to be impacted by aging. We show that biological features that contribute to the aging processes such as telomere attrition and mutagenesis play an important role in the pathogenesis of MM and can be used to improve clinical decision making in older patients.

6.3. Results

6.3.1. The MMRF data set is representative of modern myeloma patients

Although they are fewer elderly patients, the age distribution, treatment and outcome of the 980 patient case series with paired WGS data is comparable to previously published datasets, **Figure 6.3.1.1**. The median age at diagnosis was 63 (range 27-93) with 43 percent being aged 65 or over, 13% 75 or over and 2% 85 or over, **Table 6.3.1.1**. Major cytogenetic subgroups comprised 19% t(11;14), 12% t(4;14), 5% t(14;16)/t(14;20) with 57% HRD. The frequency of mutation showed 27% *KRAS*, 22% *NRAS*, 9.8% *DIS3*, 9.8% *FAM46C*, and 7.0% *BRAF*. Complex events such as chromoplexy and chromothrypsis were seen in 19%, and 9% respectively.

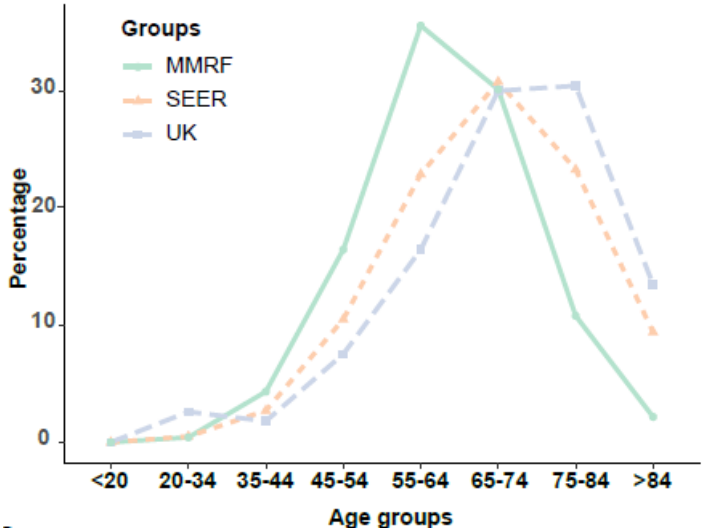


Figure 6.3.1.1. The age distribution of patients is similar in the USA and the UK but there are fewer older patients in the MMRF CoMMpass study.

	Summary
Median age at diagnosis	63 (range 27-93)
Age breakdown	<55: 21% (n=209) 55-64: 36% (n=349) 65-74: 30% (n=295) >74: 13% (n=127)
Proportion per age group that received an ASCT	<55: 56% (n=117) 55-64: 51% (n=179) 65-74: 36% (n=108) >74: 7% (n=9)
Sex ratio F:M	0.65:1
Race	Asian 1.8% (n=15) Black: 17% (n=137) White: 76% (n=622) Other: 5% (n=42)
ISS	I 36% (n=344) II 36% (n=348) III 27% (n=261)
Performance status (ECOG)	0: 35% (n=252) 1: 49% (n=348) 2: 23% (n=84) 3: 5% (n=34) 4: 1% (n=6)
Median follow-up	3.84 (95%CI 3.71-3.93)
3-year overall survival	77% (95%CI 74-80%)

Table 6.3.1.1. Summary of patient's characteristics

6.3.2. Age of diagnosis has a significant impact on genetic markers at presentation

We determined the impact of age on the distribution of major cytogenetic and clinical determinants of risk using age as a continuous variable. We did not see any difference in the frequency of t(4;14), del(17p), gain(1q), del(1p), and ISS. However, poor performance status did associate with older age**, Figure 6.3.2.1.**

6. Elderly

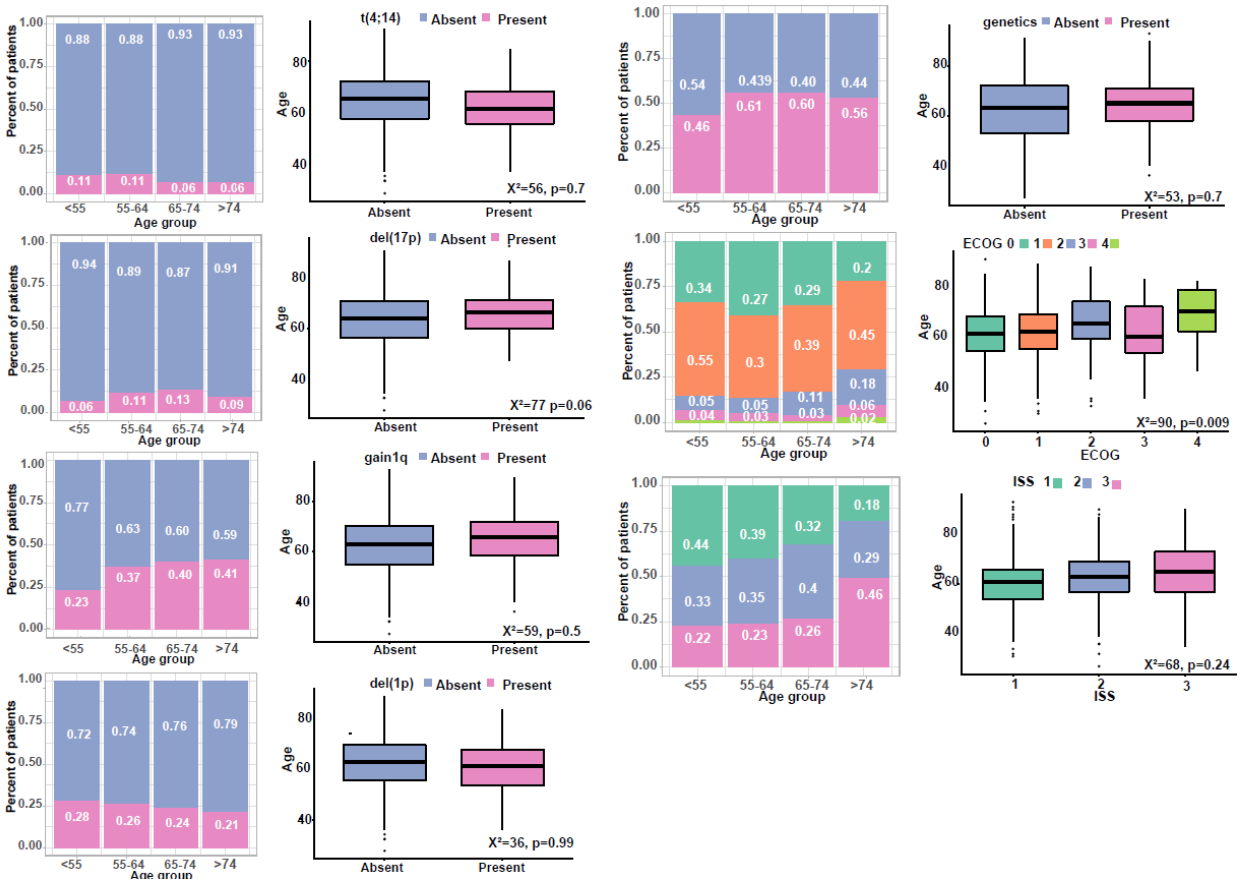


Figure 6.3.2.1. Age differences between subgroups in NDMM.

Importantly for the structural chromosomal variants, we identified a significant increase in chromothrypsis with increasing age (χ^2=10, p=0.001), **Figure 6.3.2.2**.

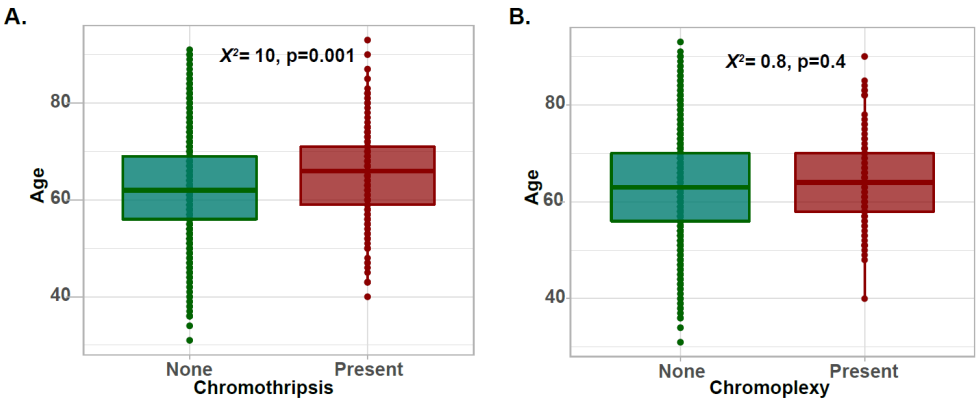


Figure 6.3.2.2. Chromothripsis and chromoplexy by age. Patients with chromthrypsis were significantly older than the others but there was no difference with chromoplexy.

To complete this analysis, we determined the impact of age on the distribution of genetic features seen at presentation using a dataset of 888 patients for which cytogenetic subgroup, genome wide copy number abnormalities, and mutations were available. Using age as a continuous variable and a Bayesian inference strategy, we did not see a significant impact of age on the major cytogenetic subgroups (t(11;14), t(4;14) t(14;16). Furthermore, we did not detect an impact of age on mutational burden or mutational processes with no significant correlation between the proportion of APOBEC, SBS2 and SBS5 signatures with age. We did identify some variability with age of *DNAH5* mutations (Corr=0.10, BF=3.9), *KMT2B* mutations (Corr=0.1, BF=2.4), del(6q) (Corr=0.1, BF=2) and del(1p) (Corr 0.11, BF=5.2) with a significant decrease with increase age.

Overall, this analysis did not identify a clinically relevant significant difference in the distribution of the common clinically relevant genetic markers in MM but a novel finding of an increased prevalence of chromothripsis was identified which is of relevance to the potential genetic drivers of progression in patients of older age.

6.3.3. Cytogenetic subgroups impact on clinical outcome in an age-related fashion

From a treatment perspective the study group was treated in a standard clinically relevant fashion. All patients received an IMiD (e.g., lenalidomide, pomalidomide, thalidomide) and/or proteasome inhibitor (e.g., bortezomib, carfilzomib) as part of their initial treatment regimens. The group as a whole had a median follow-up of 3,84 years (95% CI 3.71-3.93), with a 3-year OS of 77% (95% CI 74-80%). The dataset behaved as we expected with ISS, performance status, del(17p), and *TP53* mutations being associated with adverse outcomes**.**

Age at diagnosis was associated with a significant impact on outcome with the OS of patients aged 65 and above being significantly worse than those below this age (HR 2.1 (95%CI 1.6-2.7), p<0.0001). This difference in outcome was more pronounced when patients were grouped by decades, with a negative impact on OS being seen for patients age 65-74 (HR=2.1 (95%CI 1.3-3.3), p=0.0007)) and ≥75 (HR=4.3 (95%CI 2.7-6.9), p<0.0001)), **Figure 6.3.3.1.**

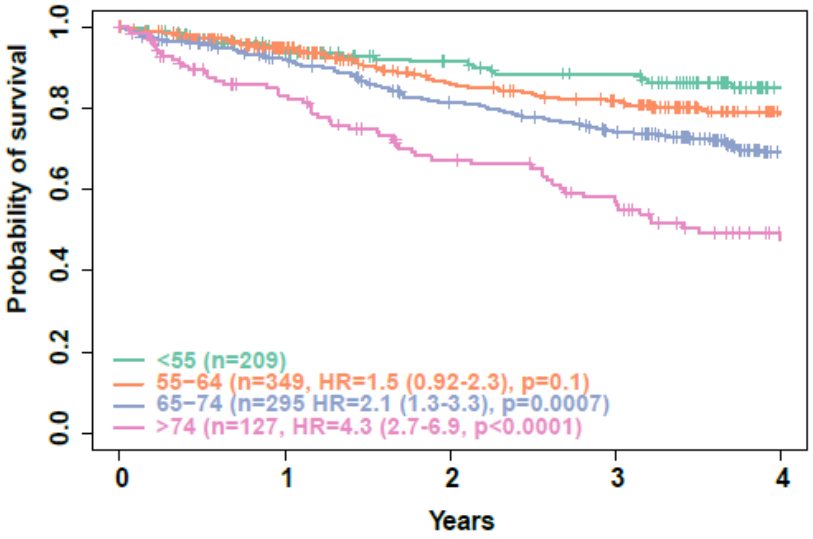


Figure 6.3.3.1. Overall survival in MM by age group.

It has been argued that the contribution of tumour genetic to outcome in elderly patients with MM was negligeable. To test this hypothesis, we performed a multivariate analysis per age group using ISS ≥2, performance status >2, and a composite genetic score (presence of a del(1p), gain(1q), t(4;14), or del(17p)). This analysis shows a reduction in the contribution of genetics overtime, **Figure 6.3.3.2**

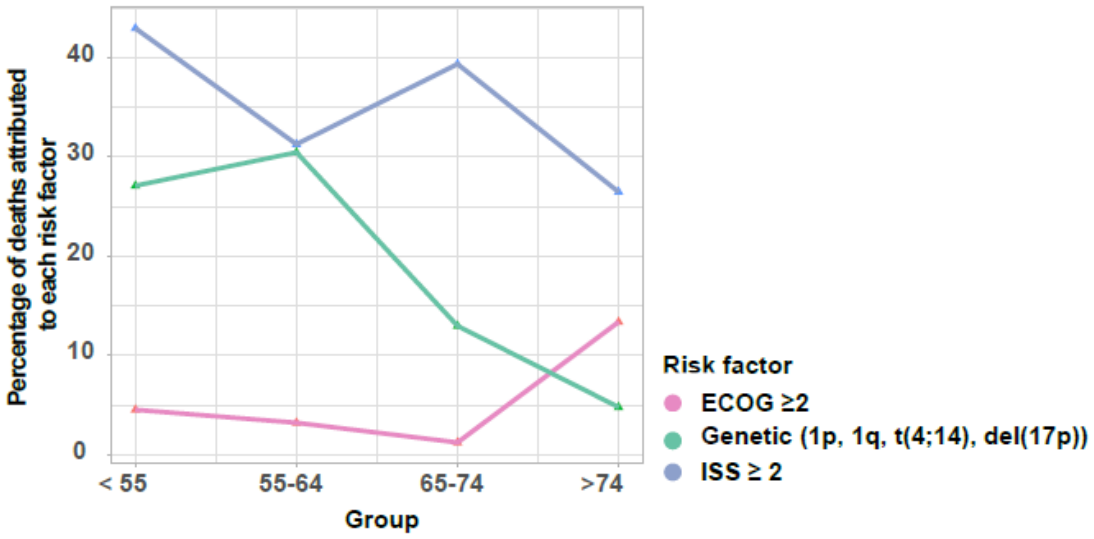


Figure 6.3.3.2. Contribution of genetics to risk in the elderly.

6.3.4. Aging markers in the haematopoietic compartment do not associate with outcome in MM

The haematopoietic compartment is particularly sensitive to aging processes. We hypothesised that clonal haematopoiesis and telomere attrition were associated with disease biology and outcome.

6.3.4.1. Clonal haematopoiesis

The prevalence of clonal haematopoiesis increased with increasing age in MM patients, based on an analysis 980 somatic DNA control samples. We identified 169 mutations in 156 (16%) patients. The median age of patients with clonal haematopoiesis was 66 years (range: 34-85) versus 63 (range: 27-93) (χ^2= 3.9, df = 1, p-value = 0.048), **Figure 6.3.4.1.1.**

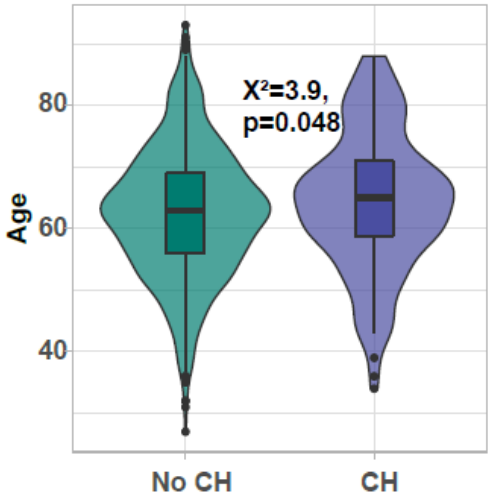


Figure 6.3.4.1.1. Clonal haematopoiesis was associated with age.

The most commonly mutated genes were *DNMT3A*, *ASXL1*, and *TET2*, **Figure 6.3.4.1.2**. The median VAF of these variants was 5.6% (range: 1.3-29.6%). A signature analysis on the mutational variants detected using MMSIG (Rustad *et al*, 2020) identified the signatures SBS1 and SBS5 suggesting that they were generated by conventional clock-like processes, consistent with their age-based aetiology and not via a MM specific mechanism such as AID or APOBEC, **Figure 6.3.4.1.3.**

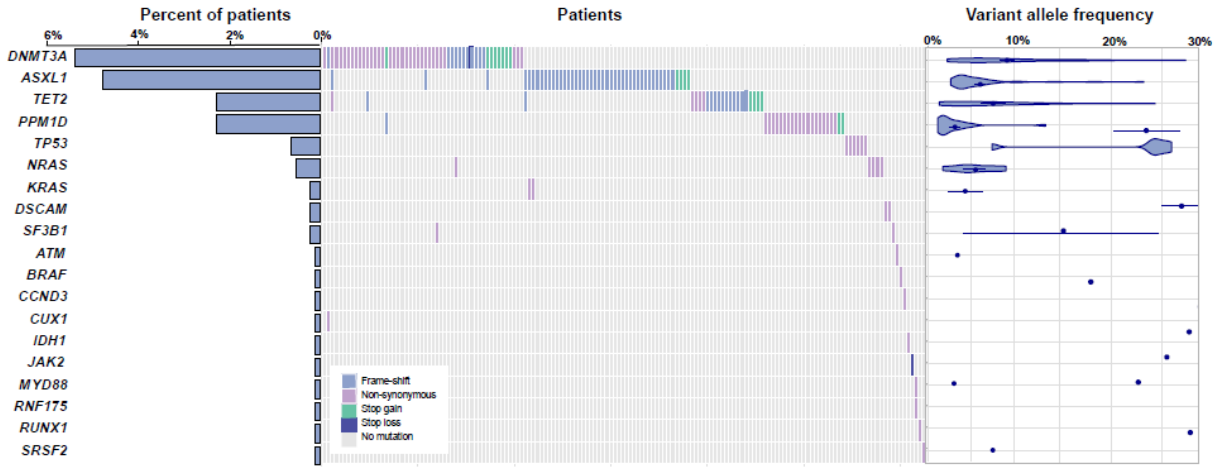


Figure 6.3.4.1.2. *DNMT3A*, *ASXL1*, and *TET2* are the most frequently mutated genes. Waterfall plot summarizing the CH mutations. Left panel showing the percentage of patients. Right panel showing the VAD distribution. B. Boxplot suggesting patients with CH are older than the ones without CH.

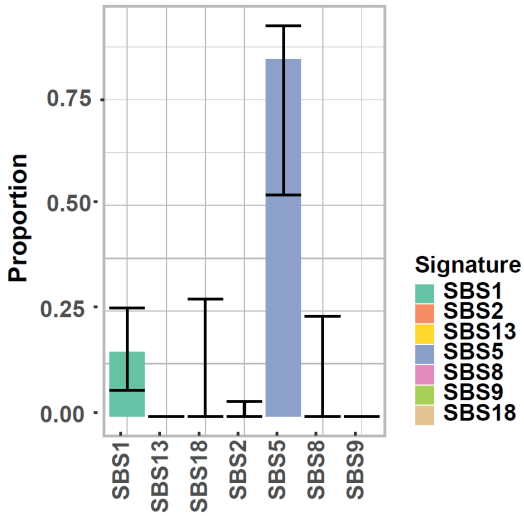


Figure 6.3.4.1.3. Signature analysis suggesting CH mutations are all generated via an age-related process.

The 3-year OS of patients with clonal haematopoiesis mutations was 77% (95% CI 74-80%) compared to 79% (95% CI 72-86%) in the remainder consistent with them having a neutral impact on outcome, **Figure 6.3.4.1.4**. To investigate this further, we analyzed the impact of CH in the ≥65 and those that received an ASCT, and found no difference in outcome

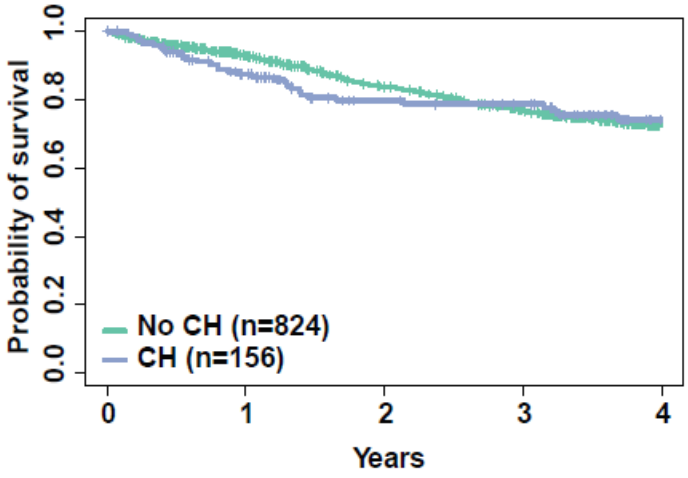


Figure 6.3.4.1.4. Impact of CH on outcome.

6.3.4.2. Leucocyte telomere attrition

Telomere shortening below a critical level can initiate cellular senescence or apoptosis or alternatively in the absence of these responses it can initiate chromosomal instability and neoplastic transformation (Shay & Bacchetti, 1997; Wu *et al*, 2003b). Thus, we reasoned that telomere shortening may reflect premature cellular aging and thus in contrast to single genetic events may contain significant information on tumour cell behaviour. To this aim, we determined the Leucocyte telomere length (LTL) from either a pool of whole blood cells (n=848) or CD3 positive selected cells (n=63) in 911 NDMM. There was no difference in the estimated LTL between CD3+ and whole blood samples (χ^2=3.2, p=0.07), they were therefore considered together.

The median LTL was 1.6kB (range: 1.1-9.5 kB). African Americans has significantly shorter LTL than Whites despite being younger on average (1.6 (1.1-2.5) vs. 1.7 (1.2-2.6), χ^2=5.2, p=0.02), the clinical significance of which remains to be determined.

There were no differences according to sex, **Figure 6.3.4.2.1**. When considering CH patients, there was no difference in the LTL of patients that had CH and those that did not (χ^2=2.9, p=0.09) suggesting different mechanisms are involved**.**

LTL did not affect the presence of translocations, HRD, or complex rearrangements. An association of LTL with prognosis was not identified, **Figure 6.3.4.2.2.**

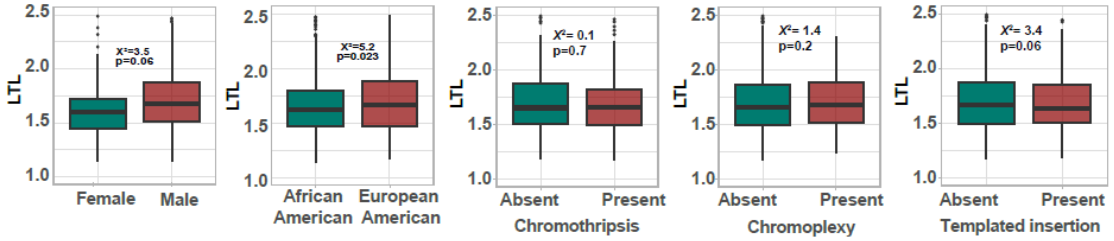


Figure 6.3.4.2.1. LTL was significantly shorter in African Americans.

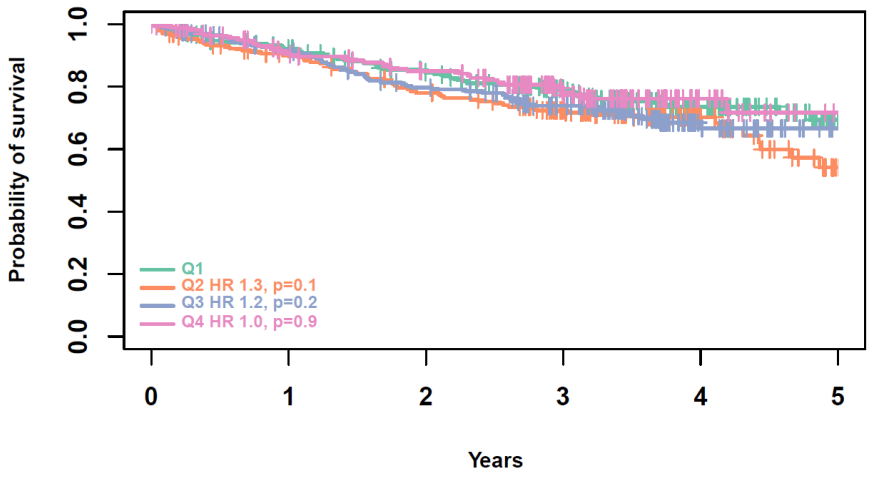


Figure 6.3.4.2.2. Impact of LTL by quartile on outcome.

6.3.5. Tumour telomere length (TTL) has a number of specific correlations and is prognostic

The correlate of looking at LTL was to analyse tumour telomere length (TTL) and evaluate its impact on the biology of MM especially in the elderly.

The median TTL was 5.8kB (1.3-53) and we demonstrate that the absolute length and distribution of LTL and TTL were very different with LTL being significantly shorter and spanning a smaller range than TTL, **Figure 6.3.5.1.A.**

We identified a small but significant negative correlation between TTL and age (r=-0.04, F=9.5, p=0.002) with shorter telomeres being seen in the older populations, **Figure 6.3.5.1.B**. Short TTL was associated with the presence of the complex structural event chromothrypsis (χ^2=4.8, p=0.03), and templated insertion (χ^2=4.2, p=0.04) but not chromoplexy (χ^2=1.0, p=0.3), **Figure 6.3.5.1.C.** TTL was significantly shorter in the *TP53* and *ATM* groups; identified using Pearson correlation (χ^2=9, p=0.002) and (χ^2=7.2, p=0.007) respectively consistent with TTL shortening being associated with markers of DNA instability.

A negative correlation between TTL and gain(1q) (Corr=-0.15, BF=209), del(1p) (Corr=-0.11, BF=12), del(12p) (Corr=-0.10, BF=1), and del(15q26) (Corr=0.1, BF=1.2) was identified using a Bayesian approach. A positive correlation with mutations in *MAGI1* (Corr=0.14, BF=40), *MDGA2* (Corr=0.12, BF=5.8), *DNAH17* (Corr=0.11, BF=3), *PTPN13* (Corr=0.11, BF=1.9), *SVIL* (Corr=0.10, BF=1), and *NRK* (Corr=0.10, BF=1) was identified.

6. Elderly

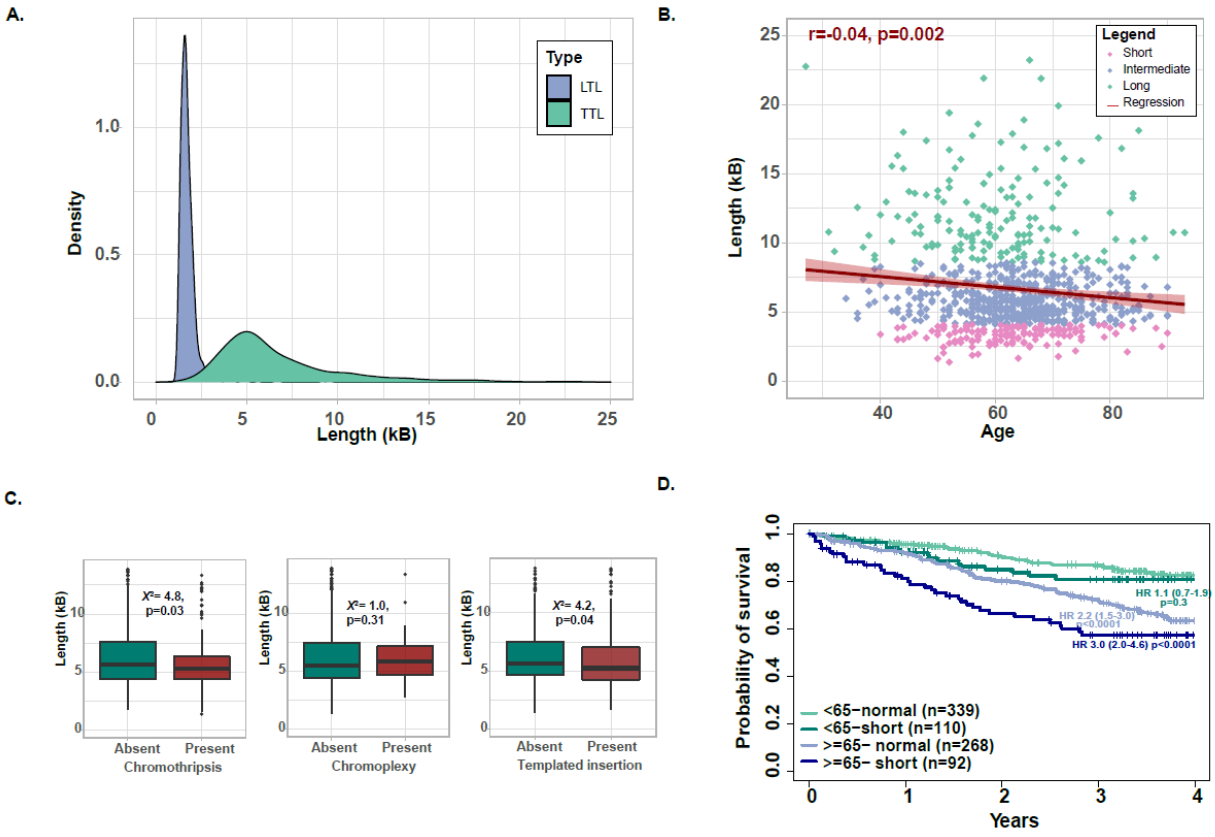


Figure 6.3.5.1. Age defined genetic makeup of myeloma. A. Distribution of TTL and LTL showing that LTL are much less variable than TTL. B. Correlation between age and TLL. C. Correlations between TTL and complex rearrangement suggesting patients with CT and TI have shorter TTL. D. Impact on tumor TTL on outcome

To determine the relationship between TTL and outcome we first analysed TTL by quartiles: the shorter the TTL were the worse the outcome, **Figure 6.3.5.2.A.** A small fraction (n=35, 4%) had significantly longer TTL (>15kB), but there was no clear relationship to outcome in these patient, **Figure 6.3.5.2.B**. To determine further the association of short TTL in malignant plasma cells with adverse outcome we performed a predictive cut-off based on maximally ranked statistic. The cut-offs determined was 4.1 and 8.5 kB. There was a significant correlation with adverse OS especially in the older patients where having short TTL is associated with a HR for death of 3.9 (2.5-6.2), p<0.001, **Figure 6.3.5.2.C.** Interestingly this seemed to be seen in most age groups, **Figure 6.3.5.3.**

6. Elderly

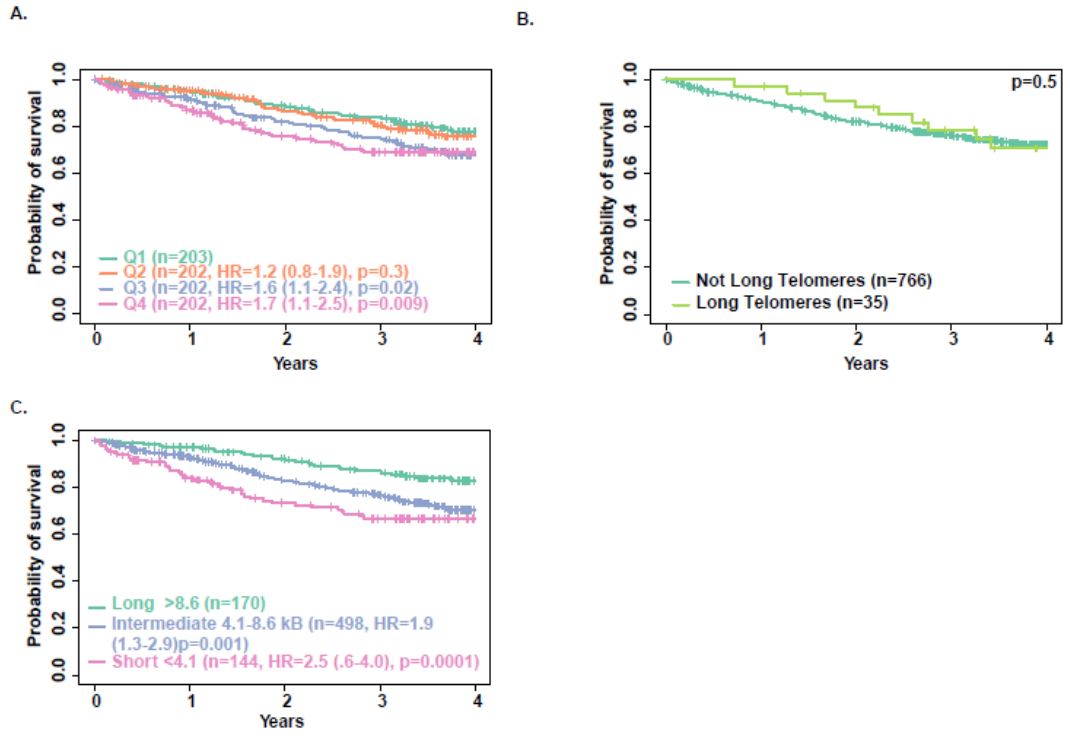


Figure 6.3.5.2. Impact of TTL on outcome A. Survival by quartile B. Impact of long TTL, and C. Impact of TTL by ranking groups.

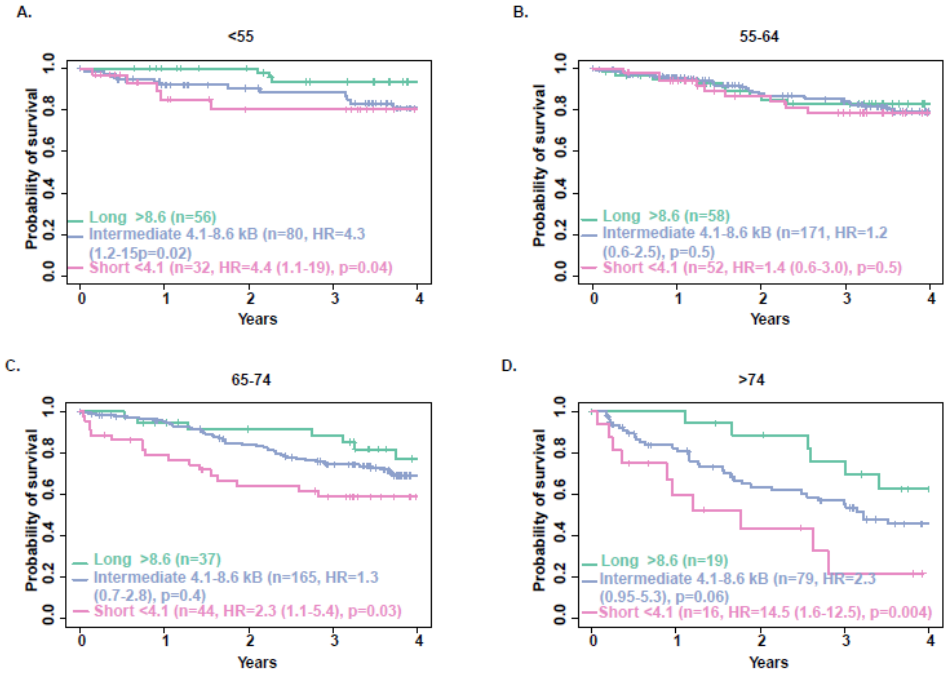


Figure 6.3.5.3. Impact of TTL by age group A. <55, B. 55-64, C . 65-74 D. >75

6. Elderly

To understand how both clinical and molecular factors contribute to outcome we carried out a multivariate analysis combining age (as a continuous variable), chromothripsis, TTL (<4.1 kB), ISSII, ISSIII and performance status (ECOG≥2) both overall and in the age group over 65. This analysis showed that all factors retained their independent adverse predictive significance. Importantly these features are additive for risk with patients lacking any of them having the best outcome and patients with 3 factors having the worst outcome. This observation is most penetrant in the over 65 group where 30% have chromothripsis, 14% short TTL, 36% ISSII, 32% ISSIII and 17% ECOG2, **Figure 6.3.5.4**. TTL is an independent adverse predictive marker in the dataset overall, **Figure 6.3.5.6.**

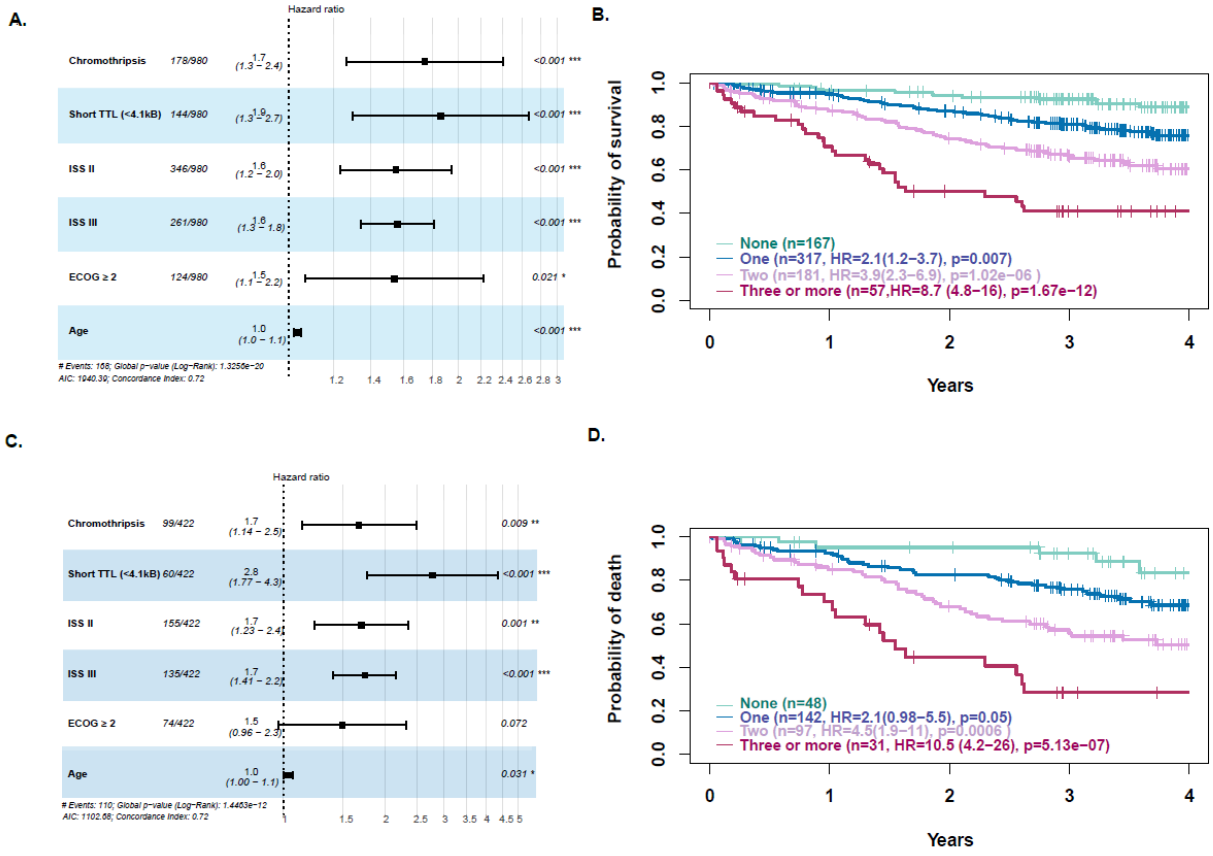


Figure 6.3.5.4. Multivariate analysis incorporating age, ISS, ECOG, TTL and chromothripsis. The results suggest that ISS, PS, Age and short TTL are independent markers of outcome. In A. we present the results overall and in B. we present the results for the patients over the age of 65. In C-D we present the impact of carrying none or more than one adverse marker.

6.3.6. Mechanistic basis of telomere length maintenance in plasma cells

To address the mechanisms by which tumour telomere attrition occurred we assessed the expression of components of the telomerase complex and its regulators in the RNAseq data derived from selected tumour plasma cells (n=680). Telomerase catalytically adds TTAGGG hexameric nucleotide repeats to the 3′-hydroxyl end of the telomeric leading strand, using a specific sequence in the RNA component as the template. The Telomerase Complex is composed of the enzyme telomerase reverse transcriptase (*TERT*), its RNA component (*TERC*), the protein dyskerin (encoded by *DKC1*) and other associated proteins (*NHP2, NOP10*, and *GAR1*), **Figure 6.3.6.1**. Expression of *TERT, DKC1, NHP2, NOP10*, and *GAR1* were all negatively correlated to TTL but not *TERC***.** The Pontin/Reptin complex (encoded by *RUVBL1* and *RUVBL2*) are essential components for the assembly of the telomerase complex are also negatively correlated to TTL. *SRSF11* that mediates the loading of the TC onto the telomeres is on the contrary positively correlated to TTL, **Figure 6.3.6.2.** The protective proteins associated with telomere DNA collectively termed shelterin (*TERF1, TERF2, TINF2, POT1, ACD,* and *TERF2IP*) were not correlated with TTL**.** Other proteins such as the members of the CST complex (*CTC1) and regulators RTEL1, CBX3, BLM,* and *ATM* were positively correlated to TTL, **Figure 6.3.6.2**. Expression levels are summarised in **Figure 6.3.6.3**.

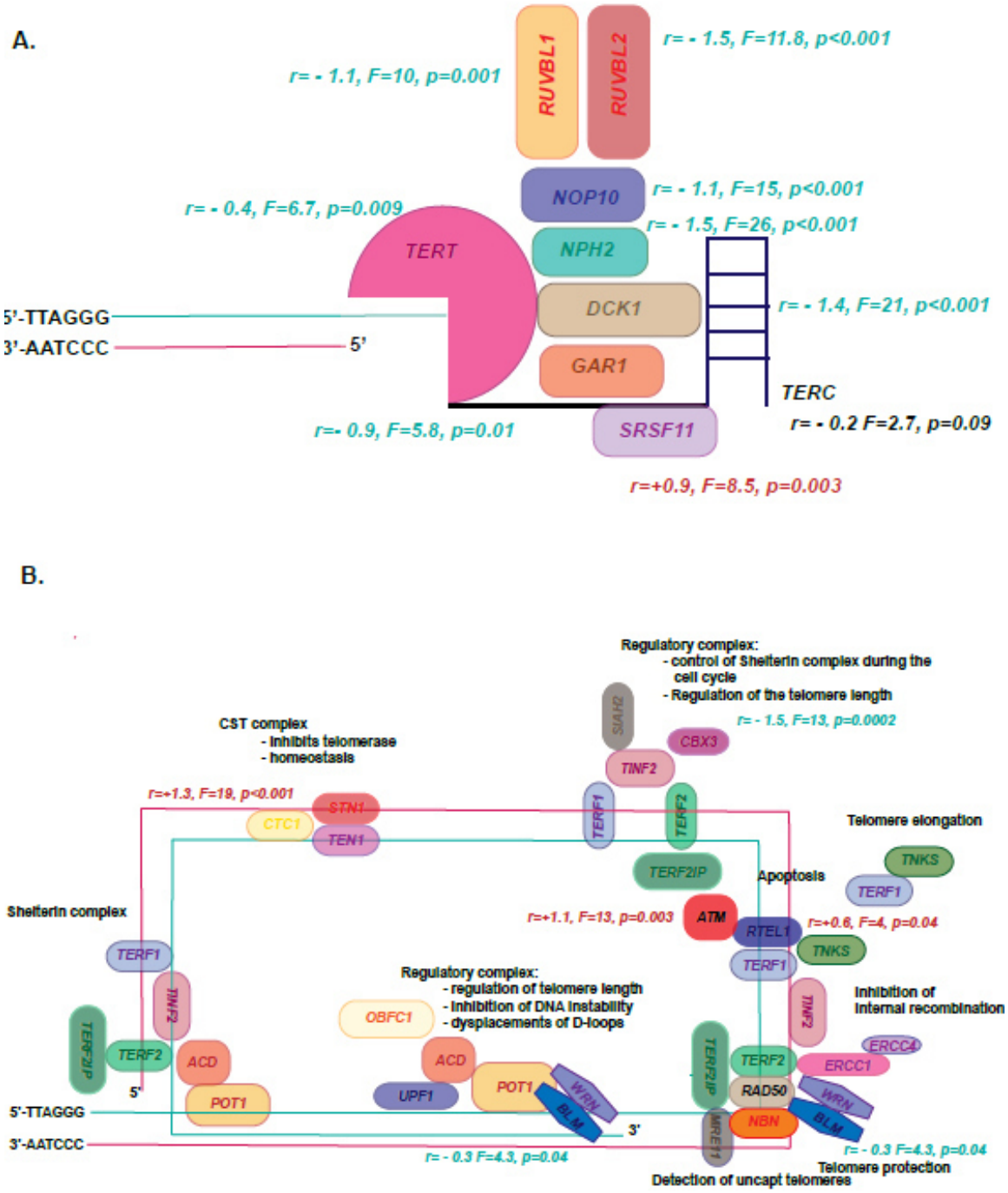
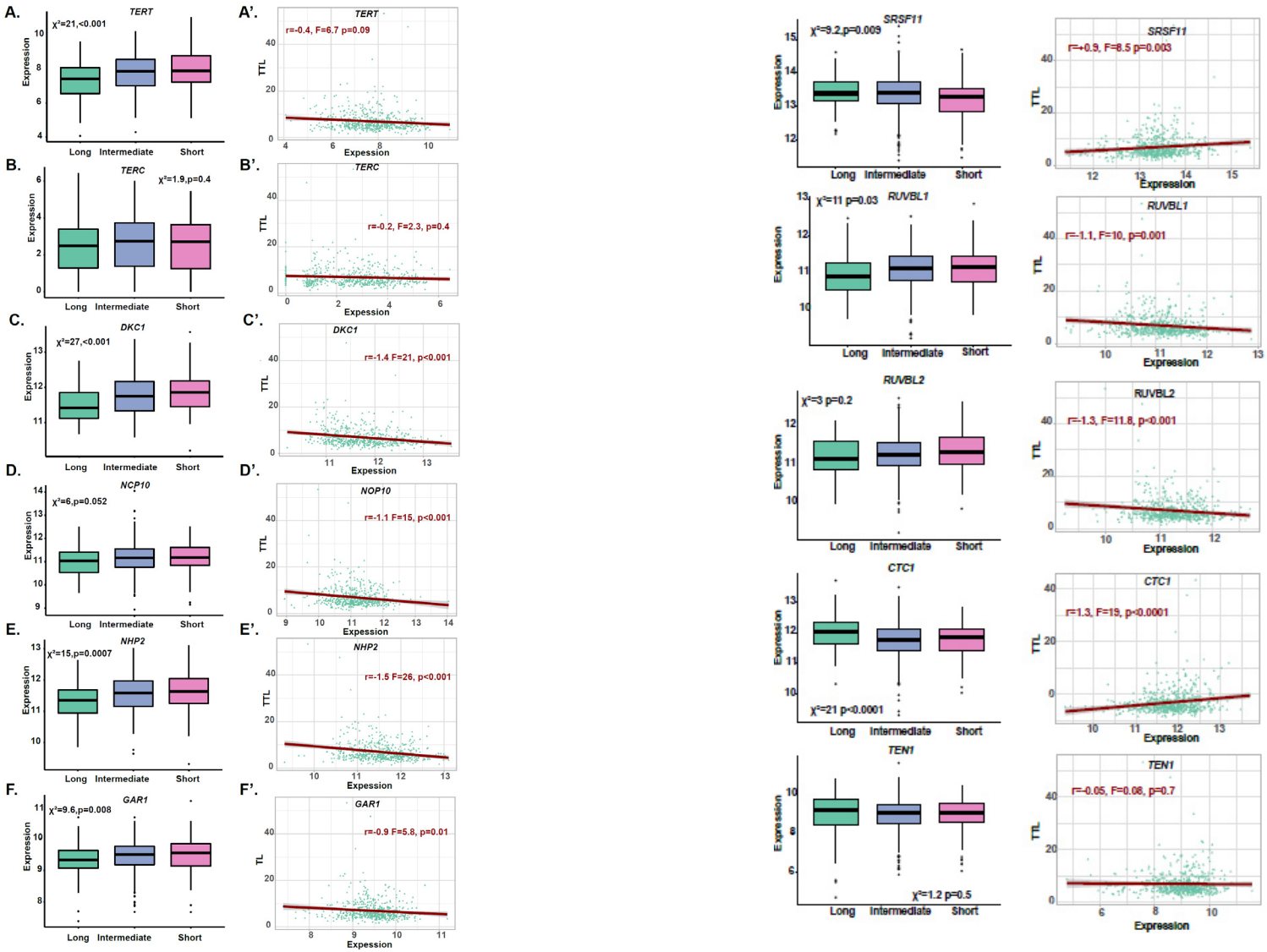
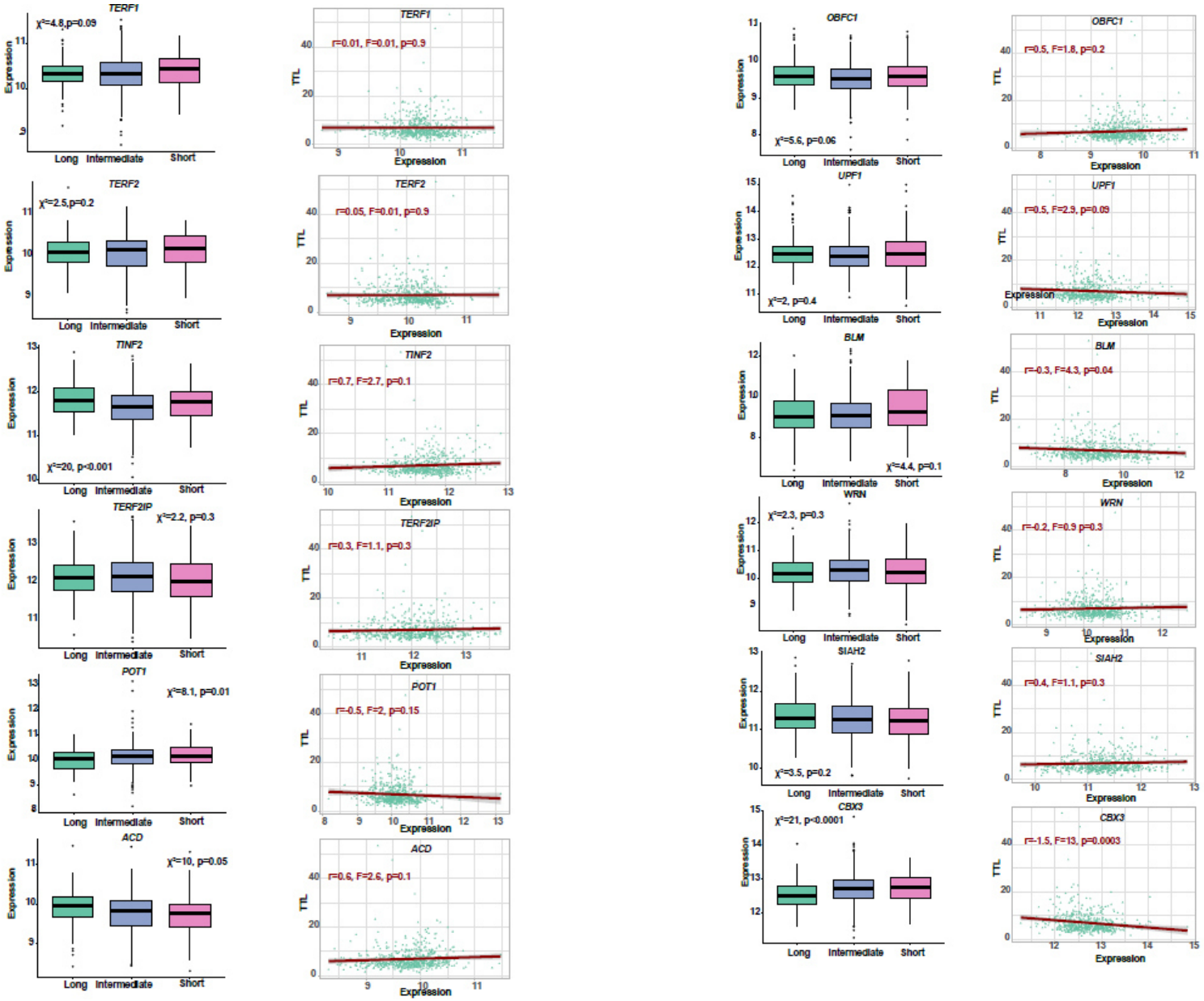


Figure 6.3.6.2. Telomerase complex and regulatory complex expression. A. Expression of the Telomerase complex genes suggesting they are negatively correlated to TTL B. Expression of the regulatory complexes suggesting they are mostly unchanged except *CTC1* and *RTEL1* that are upregulated and *BML* and *CBX3* that are downregulated.


A.
TERT
χ²=21,<0.001
Expression
Long Intermediate Short
A'.
TERT
r=-0.4, F=6.7 p=0.09
TTL
Expression
B.
TERC
χ²=1.9,p=0.4
Expression
Long Intermediate Short
B'.
TERC
r=-0.2, F=2.3, p=0.4
TTL
Expression
C.
DKC1
χ²=27,<0.001
Expression
Long Intermediate Short
C'.
DKC1
r=-1.4 F=21, p<0.001
TTL
Expression
D.
NCP10
χ²=6,p=0.052
Expression
Long Intermediate Short
D'.
NOP10
r=-1.1 F=15, p<0.001
TTL
Expression
E.
NHP2
χ²=15,p=0.0007
Expression
Long Intermediate Short
E'.
NHP2
r=-1.5 F=26, p<0.001
TTL
Expression
F.
GAR1
χ²=9.6,p=0.008
Expression
Long Intermediate Short
F'.
GAR1
r=-0.9 F=5.8, p=0.01
TL
Expression
SRSF11
χ²=9.2,p=0.009
Expression
Long Intermediate Short
SRSF11
r=+0.9, F=8.5 p=0.003
TTL
Expression
RUVBL1
χ²=11 p=0.03
Expression
Long Intermediate Short
RUVBL1
r=-1.1, F=10, p=0.001
TTL
Expression
RUVBL2
χ²=3 p=0.2
Expression
Long Intermediate Short
RUVBL2
r=-1.3, F=11.8, p<0.001
TTL
Expression
CTC1
χ²=21 p<0.0001
Expression
Long Intermediate Short
CTC1
r=1.3, F=19, p<0.0001
TTL
Expression
TEN1
χ²=1.2 p=0.5
Expression
Long Intermediate Short
TEN1
r=-0.05, F=0.06, p=0.7
TTL
Expression

6. Elderly


TERF1
χ²=4.8,p=0.09
Expression
Long Intermediate Short
TERF1
r=0.01, F=0.01, p=0.9
TTL
Expression
TERF2
χ²=2.5,p=0.2
TERF2
r=0.05, F=0.01, p=0.9
TINF2
χ²=20, p<0.001
TINF2
r=0.7, F=2.7, p=0.1
TERF2IP
χ²=2.2, p=0.3
TERF2IP
r=0.3, F=1.1, p=0.3
POT1
χ²=8.1, p=0.01
POT1
r=-0.5, F=2, p=0.15
ACD
χ²=10, p=0.05
ACD
r=0.6, F=2.6, p=0.1
OBFC1
χ²=5.6, p=0.06
OBFC1
r=0.5, F=1.8, p=0.2
UPF1
χ²=2, p=0.4
UPF1
r=0.5, F=2.9, p=0.09
BLM
χ²=4.4, p=0.1
BLM
r=-0.3, F=4.3, p=0.04
WRN
χ²=2.3, p=0.3
WRN
r=-0.2, F=0.9 p=0.3
SIAH2
χ²=3.5, p=0.2
SIAH2
r=0.4, F=1.1, p=0.3
CBX3
χ²=21, p<0.0001
CBX3
r=-1.5, F=13, p=0.0003

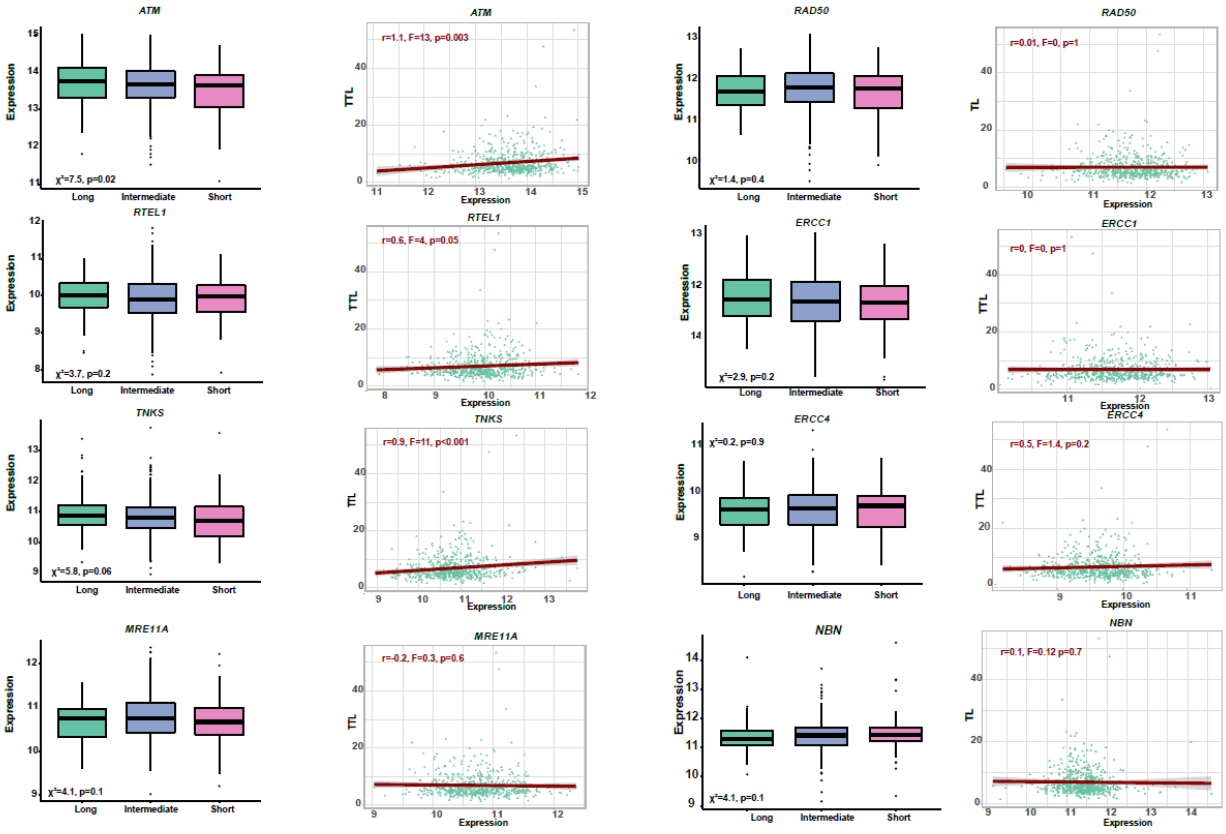


Figure 6.3.6.3. Summary panels of the expression of the telomerase related genes either by length group or continuous variables.

Overall, these data are consistent with the idea that the loss of the Telomerase Complex components despite some increase in a number of regulatory components (*CTC1, RTEL*) that is responsible to short TTL.

6.4. Discussion

In this first WGS analysis of an older population we show that del(6q), and del(1p) significantly decrease in frequency with increasing age whereas in contrast chromothripsis increased. In a novel and important observation we show that chromothripsis which has reported frequencies of 10%-20% in NDMM (Ashby *et al*, 2019; Rustad *et al*, 2020) is an age related molecular abnormality. Thus, while the genetic make-up of MM does vary with increasing age these variations are small and are not seen in the aetiologic translocations as has been suggested previously.

Surprisingly despite finding clonal hemopoiesis at anticipated frequencies in a population of aging individuals we did not identify a significant impact on the clinical outcome of patients with MM. In another series of cases it was detected at the same frequency (20%) but there it was associated with impaired outcome (Mouhieddine *et al*, 2020). A second study, Maia et al (Maia *et al*, 2020), also identified clonal hemopoiesis at these frequencies consistent with this being truly reflective of a population of MM patients. Taken together we suggest that the frequency of clonal haematopoiesis is approximately 6% overall but that the full impact on treatment outcome is currently unclear but based on our results it seems unlikely that it exerts a significant negative impact on survival such that it could be considered for incorporation as a prognostic factor useful in clinical practice.

The association of telomere attrition with outcome in other disease areas suggested that it may integrate significant molecular information and so constitute a relevant clinical prognostic marker that is both prevalent and penetrant (Ernst *et al*, 2016; Fukami *et al*, 2017; Weischer Maren *et al*, 2012). In breast and oesophageal cancer telomere attrition has been correlated with *TP53* mutations (Hao *et al*, 2013). In CLL for instance TTL has been correlated with both high-risk disease and poor outcome (Wu *et al*, 2003b). In those studies, telomere length was assessed initially by Southern blotting and subsequently by PCR based approaches. More recently, the availability of WGS datasets has allowed the inference of telomere length in large datasets. Here we have inferred telomere length in a large series of MM cases. We show that short telomere length is associated with adverse OS and this effect is independent from ISS, PS, and

age. Mechanistically we speculate that it constitutes a good marker because it can integrate information on DNA instability, telomeric copy number loss, chromoplexy, and aging. In MM we identified a strong correlation between TTL in the malignant plasma cells and the presence of *TP53* and *ATM* mutations. Mechanistically we propose that impaired telomere function activates the canonical DNA damage response pathway that engages p53 to initiate apoptosis or replicative senescence (Deng *et al*, 2008). In cases with dysfunctional DNA pathways such as *TP53* or *ATM* mutants this pathway may not be activated enabling MM cells to evade programmed cell death or senescence.

For its application in the clinic, Single Telomere Length Analysis (STELA) which provides high-resolution chromosome specific telomere length profiles could offer an interesting alternative. STELA is, a single-molecule PCR based telomere length analysis technology that can determine the full spectrum of telomere lengths from specific chromosome ends (Baird *et al*, 2003). Whilst STELA is comparatively low throughput, its high-resolution can detect the presence of telomeres within the length ranges that can lead to senescence, apoptosis and telomere fusion and may help not only translate our findings into the clinic but also add precisions to them.

We describe the genetic changes present in the elderly patients and identify markers of Aging such as CH and telomere attrition. We highlight the importance of tumour telomere length, a composite factor that takes into account DNA instability, copy number losses, and aging as a potential novel biological marker to assess outcome and aid personalized treatment decisions in older patients with MM.

6.5. Reference

Ageing Europe — 2019 edition Available at: https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-02-19-681 [Accessed August 7, 2020].

Ashby, C., Boyle, E.M., Walker, B.A., Bauer, M.A., Ryan, K.R., Dent, J., Thakurta, A., Flynt, E., Davies, F.E. & Morgan, G. (2019) Chromoplexy and Chromothripsis Are Important Prognostically in Myeloma and Deregulate Gene Function By a Range of Mechanisms. *Blood*, **134,** 3767–3767.

Avet-Loiseau, H., Hulin, C., Campion, L., Rodon, P., Marit, G., Attal, M., Royer, B., Dib, M., Voillat, L., Bouscary, D., Caillot, D., Wetterwald, M., Pegourie, B., Lepeu, G., Corront, B., Karlin, L., Stoppa, A.-M., Fuzibet, J.-G., Delbrel, X., Guilhot, F., et al (2013) Chromosomal Abnormalities Are Major Prognostic Factors in Elderly Patients With Multiple Myeloma: The Intergroupe Francophone du Myélome Experience. *Journal of Clinical Oncology*, **31,** 2806–2809.

Baird, D.M., Rowson, J., Wynford-Thomas, D. & Kipling, D. (2003) Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nature Genetics*, **33,** 203–207.

Coombs, C.C., Zehir, A., Devlin, S.M., Kishtagari, A., Syed, A., Jonsson, P., Hyman, D.M., Solit, D.B., Robson, M.E., Baselga, J., Arcila, M.E., Ladanyi, M., Tallman, M.S., Levine, R.L. & Berger, M.F. (2017) Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*, **21,** 374-382.e4.

Deng, Y., Chan, S. & Chang, S. (2008) Telomere dysfunction and Tumor Suppression-the Senescence Connection. *Nature reviews. Cancer*, **8,** 450–458.

Ernst, A., Jones, D.T.W., Maass, K.K., Rode, A., Deeg, K.I., Jebaraj, B.M.C., Korshunov, A., Hovestadt, V., Tainsky, M.A., Pajtler, K.W., Bender, S., Brabetz, S., Gröbner, S., Kool, M., Devens, F., Edelmann, J., Zhang, C., Castelo-Branco, P., Tabori, U., Malkin, D., et al (2016) Telomere dysfunction and chromothripsis. *International Journal of Cancer*, **138,** 2905–2914.

Facon, T., Dimopoulos, M.A., Dispenzieri, A., Catalano, J.V., Belch, A., Cavo, M., Pinto, A., Weisel, K., Ludwig, H., Bahlis, N.J., Banos, A., Tiab, M., Delforge, M., Cavenagh, J.D., Geraldes, C., Lee, J.-J., Chen, C., Oriol, A., De La Rubia, J., White, D., et al (2018) Final analysis of survival outcomes in the phase 3 FIRST trial of up-front treatment for multiple myeloma. *Blood*, **131,** 301–310.

Facon, T., Dimopoulos, M.A., Meuleman, N., Belch, A., Mohty, M., Chen, W.-M., Kim, K., Zamagni, E., Rodriguez-Otero, P., Renwick, W., Rose, C., Tempescul, A., Boyle, E., Manier, S., Attal, M., Moreau, P., Macro, M., Leleu, X., Lorraine Chretien, M., Ludwig, H., et al (2020) A simplified frailty scale predicts outcomes in transplant-ineligible patients with newly diagnosed multiple myeloma treated in the FIRST (MM-020) trial. *Leukemia*, **34,** 224–233.

Fukami, M., Shima, H., Suzuki, E., Ogata, T., Matsubara, K. & Kamimaki, T. (2017) Catastrophic cellular events leading to complex chromosomal rearrangements in the germline. *Clinical Genetics*, **91,** 653–660.

Hao, X.-D., Yang, Y., Song, X., Zhao, X.-K., Wang, L.-D., He, J.-D., Kong, Q.-P., Tang, N.L.S. & Zhang, Y.-P. (2013) Correlation of telomere length shortening with TP53 somatic mutations, polymorphisms and allelic loss in breast tumors and esophageal cancer. *Oncology Reports*, **29,** 226–236.

Home - Office for National Statistics Available at: https://www.ons.gov.uk/ [Accessed August 7, 2020].

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. (2013) The Hallmarks of Aging. *Cell*, **153,** 1194–1217.

Maia, C., Group, on behalf of the P.C., Puig, N., Group, on behalf of the P.C., Cedena, M.-T., Group, on behalf of the P.C., Goicoechea, I., Group, on behalf of the P.C., Valdes-Mas, R., Group, on behalf of the P.C., Vazquez, I., Group, on behalf of the P.C., Chillon, M.-C., Group, on behalf of the P.C., Aguirre, P., Group, on behalf of the P.C., Sarvide, S., Group, on behalf of the P.C., Gracia-Aznárez, F.J., Group, on behalf of the P.C., et al (2020) Biological and clinical significance of dysplastic hematopoiesis in patients with newly diagnosed multiple myeloma. *Blood*, **135,** 2375–2387.

Mouhieddine, T.H., Sperling, A.S., Redd, R., Park, J., Leventhal, M., Gibson, C.J., Manier, S., Nassar, A.H., Capelletti, M., Huynh, D., Bustoros, M., Sklavenitis-Pistofidis, R., Tahri, S., Hornburg, K., Dumke, H., Itani, M.M., Boehner, C.J., Liu, C.-J., AlDubayan, S.H., Reardon, B., et al (2020) Clonal hematopoiesis is associated with adverse outcomes in multiple myeloma patients undergoing transplant. *Nature Communications*, **11,** 2996.

Myeloma - Cancer Stat Facts *SEER* Available at: https://seer.cancer.gov/statfacts/html/mulmy.html [Accessed August 7, 2020].

Pawlyn, C., Cairns, D., Kaiser, M., Striha, A., Jones, J., Shah, V., Jenner, M., Drayson, M., Owen, R., Gregory, W., Cook, G., Morgan, G., Jackson, G. & Davies, F. (2020) The relative importance of factors predicting outcome for myeloma patients at different ages: results from 3894 patients in the Myeloma XI trial. *Leukemia*, **34,** 604–612.

Population Projections Available at: https://www.census.gov/programs-surveys/popproj.html [Accessed August 7, 2020].

Rustad, E.H., Yellapantula, V., Leongamornlert, D., Bolli, N., Ledergor, G., Nadeu, F., Angelopoulos, N., Dawson, K.J., Mitchell, T.J., Osborne, R.J., Ziccheddu, B., Carniti, C., Montefusco, V., Corradini, P., Anderson, K.C., Moreau, P., Papaemmanuil, E., Alexandrov, L.B., Puente, X.S., Campo, E., et al (2020) Timing the initiation of multiple myeloma. *Nature Communications*, **11,** 1917.

Shay, J.W. & Bacchetti, S. (1997) A survey of telomerase activity in human cancer. *European Journal of Cancer (Oxford, England: 1990)*, **33,** 787–791.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2019) A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia*, **33,** 159–170.

Walker, B.A., Mavrommatis, K., Wardell, C.P., Ashby, T.C., Bauer, M., Davies, F.E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., et al (2018) Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*, **132,** 587–597.

Weischer Maren, Bojesen Stig E., Cawthon Richard M., Freiberg Jacob J., Tybjærg-Hansen Anne & Nordestgaard Børge G. (2012) Short Telomere Length, Myocardial Infarction, Ischemic Heart Disease, and Early Death. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **32,** 822–829.

Wu, K.-D., Orme, L.M., Shaughnessy, J., Jacobson, J., Barlogie, B. & Moore, M.A.S. (2003a) Telomerase and telomere length in multiple myeloma: correlations with disease heterogeneity, cytogenetic status, and overall survival. *Blood*, **101,** 4982–4989.

Wu, K.-D., Orme, L.M., Shaughnessy, J., Jacobson, J., Barlogie, B. & Moore, M.A.S. (2003b) Telomerase and telomere length in multiple myeloma: correlations with disease heterogeneity, cytogenetic status, and overall survival. *Blood*, **101,** 4982–4989.

Chapter 7. Conclusions and perspectives

7.1. The genetic complexity of plasma cell disorders

Combined these data confirm that plasma cell disorders are heterogeneous in terms of mutations, structural events, mutational signatures, and sub clonal architecture with no unifying mutations. When considering genetics alone, not one factor seems sufficient to distinguish one subtype from the other.

In the case of light-chain amyloidosis, there are no significant differences with MGUS and MM suggesting the specificity of the clinical phenotype lies somewhere else. AL is not a different disease in terms of tumour biology: it can arise from the entire spectrum of plasma cell disorders. This, by itself, is an important clinical message that could easily be integrated in the current treatment decision algorithms alongside classical variables such as frailty and organ involvement.

When comparing SMM and MM, differences at the DNA level are seen. These changes do not lie within the nature of the genetic events but in their frequency, especially when considering specific mutational drivers such as *NRAS*, *KRAS*, or *FAM46C*. The lack of clear separation between SMM and MM confirms the rather artificial classification currently in use and highlights the importance of further studies in this setting. Unfortunately, before any shift in paradigm can occur, prospective trials enrolling many patients are required to model strong predictors that may be used to accurately replace low-risk cases with MGUS and high-risk cases with MM requiring treatment.

There are other levels of complexity that ought to be considered. Epigenetic patterns could indeed also impact the behaviour and phenotype of these disorders (Rahmat *et al*, 2018). There is also a growing body of evidence supporting the role of the immune and non-immune microenvironment in disease phenotypes. Single-cell RNA sequencing (Zavidij *et al*, 2020) or high dimensional flow approaches (Woods *et al*, 2019),

could help to comprehensively grasp the cellular architecture, activation status, and demonstrate the impact of factors such as chronic inflammation or senescence on the disease phenotype. Only once these multilevel steps are understood, will we be able to address the full complexity of plasma cell disorders.

7.2. The evolving complexity of plasma cell disorders

One of the key concepts that has become apparent with time is that plasma cell disorders, like many haematological malignancies, are not static thus adding an additional level of complexity.

Studying these dynamic processes requires multiple samples to identify reliable changes in the genetic architecture where often two does not suffice (Dentro *et al*, 2017). Furthermore, the relative differences in the incidence of biallelic events suggests a multi-hit model thus leaving a window for interception strategies. If increasing complexity is acquired overtime- even more under the influence of treatment- many changes are already apparent in SMM suggesting it is already “late” in the disease evolution. A clear understanding of when this phenomenon occurs could be accomplished through the study of MGUS or even germinal centre B-cells which could be achieved through progress in the novel sequencing approaches requiring low-input DNA (Moore *et al*, 2020).

But genetics by itself is unlikely to account for the complete clonal competition patterns seen. Understanding the relative importance of both tumour factors and microenvironmental factors is required to comprehend these mechanisms and offer effective inception strategies.

7.3. The outstanding complexity of inception strategies

Prevention is difficult in MM, as it is in many diseases with an unclear aetiology. When considering lung cancer and smoking for instance it is well established that 80-90% of deaths can be avoided by smoking cessation only (Doll, 2000). Is there an equivalent for MM?

Mechanistically a role for APOBEC was identified in the progression from SMM to MM. To date, most of our understanding comes from the study of HIV. Small molecules are available for APOBEC3G but effectively targeting APOBEC in MM would probably

require pan-APOBEC antagonist targeting 2D, 3F, 2G, 2H (Venkatesan *et al*, 2018). The contrary of this approach would paradoxically be to give APOBEC agonist in an effort to create ‘therapy by hypermutations’ that would lead to cell death (Olson *et al*, 2018). This continues to be highly controversial. Overall, these approaches remain in their infancy but offer some promise

Overall, our data offer some insight into the genetic complexity of plasma cell disorders and suggest future research strategies that hopefully could benefit patient’s management in the future.

7.4. References

Dentro, S.C., Wedge, D.C. & Van Loo, P. (2017) Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harbor Perspectives in Medicine*, **7,** Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5538405/ [Accessed October 23, 2020].

Doll, S.R. (2000) Smoking and Lung Cancer. *American Journal of Respiratory and Critical Care Medicine*, **162,** 4–6.

Moore, L., Leongamornlert, D., Coorens, T.H.H., Sanders, M.A., Ellis, P., Dentro, S.C., Dawson, K.J., Butler, T., Rahbari, R., Mitchell, T.J., Maura, F., Nangalia, J., Tarpey, P.S., Brunner, S.F., Lee-Six, H., Hooks, Y., Moody, S., Mahbubani, K.T., Jimenez-Linan, M., Brosens, J.J., et al (2020) The mutational landscape of normal human endometrial epithelium. *Nature*, **580,** 640–646.

Olson, M.E., Harris, R.S. & Harki, D.A. (2018) APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chemical Biology*, **25,** 36–49.

Rahmat, M., Haradhvala, N., Sklavenitis-Pistofidis, R., Park, J., Huynh, D., Bustoros, M., Berrios, B., Reidy, M., Perilla-Glen, A., Rivotto, B., Hornburg, K., Dorfman, D.M., Licht, J.D., Getz, G. & Ghobrial, I.M. (2018) Dissecting the Epigenetic Landscape of Smoldering, Newly Diagnosed and Relapsed Multiple Myeloma Revealed IRAK3 As a Marker of Disease Progression. *Blood*, **132,** 3896–3896.

Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Bartek, J., Quezada, S.A., Hare, J., Harris, R.S. & Swanton, C. (2018) Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Annals of Oncology*, **29,** 563–572.

Woods, D.M., Laino, A., Winters, A., Alexandre, J., Weber, J. & Chattopadhyay, P. (2019) High-dimension flow cytometry reveals comprehensive deviations in immunophenotypes associated with non-response to melanoma immunotherapies. *The Journal of Immunology*, **202,** 194.21-194.21.

Zavidij, O., Haradhvala, N.J., Mouhieddine, T.H., Sklavenitis-Pistofidis, R., Cai, S., Reidy, M., Rahmat, M., Flaifel, A., Ferland, B., Su, N.K., Agius, M.P., Park, J., Manier, S., Bustoros, M., Huynh, D., Capelletti, M., Berrios, B., Liu, C.-J., He, M.X., Braggio, E., et al (2020) Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nature Cancer*, **1,** 493–506.

Annex

Annex 1: ISS and R-ISS

International Staging System (ISS)

Stage[1]	Values (β2M = Serum β2 microglobulin; ALB = serum albumin
I	β2M < 3.5 mg/L; ALB ≥ 3.5 g/dL
II	β2M < 3.5 mg/L; ALB ≥ 3.5 g/dL; or β2M 3.5 – 5.5 mg/L
III	β2M > 5.5 mg/L

Revised International Staging System (R-ISS)

Stage[2]	Values
I	Serum β2 microglobulin < 3.5 mg/l Serum albumin ≥ 3.5 g/dl Standard-risk chromosomal abnormalities (CA) Normal LDH
II	Neither R-ISS stage I or III
III	Serum β2 microglobulin ≥ 5.5 mg/L and either High-risk CA by FISH OR High LDH

1. Greipp, P. R. *et al.* International staging system for multiple myeloma. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **23**, 3412–3420 (2005).
2. Palumbo, A. *et al.* Revised International Staging System for Multiple Myeloma: A Report From1. Greipp, P. R. et al. International staging system for multiple myeloma. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 23, 3412–3420 (2005).

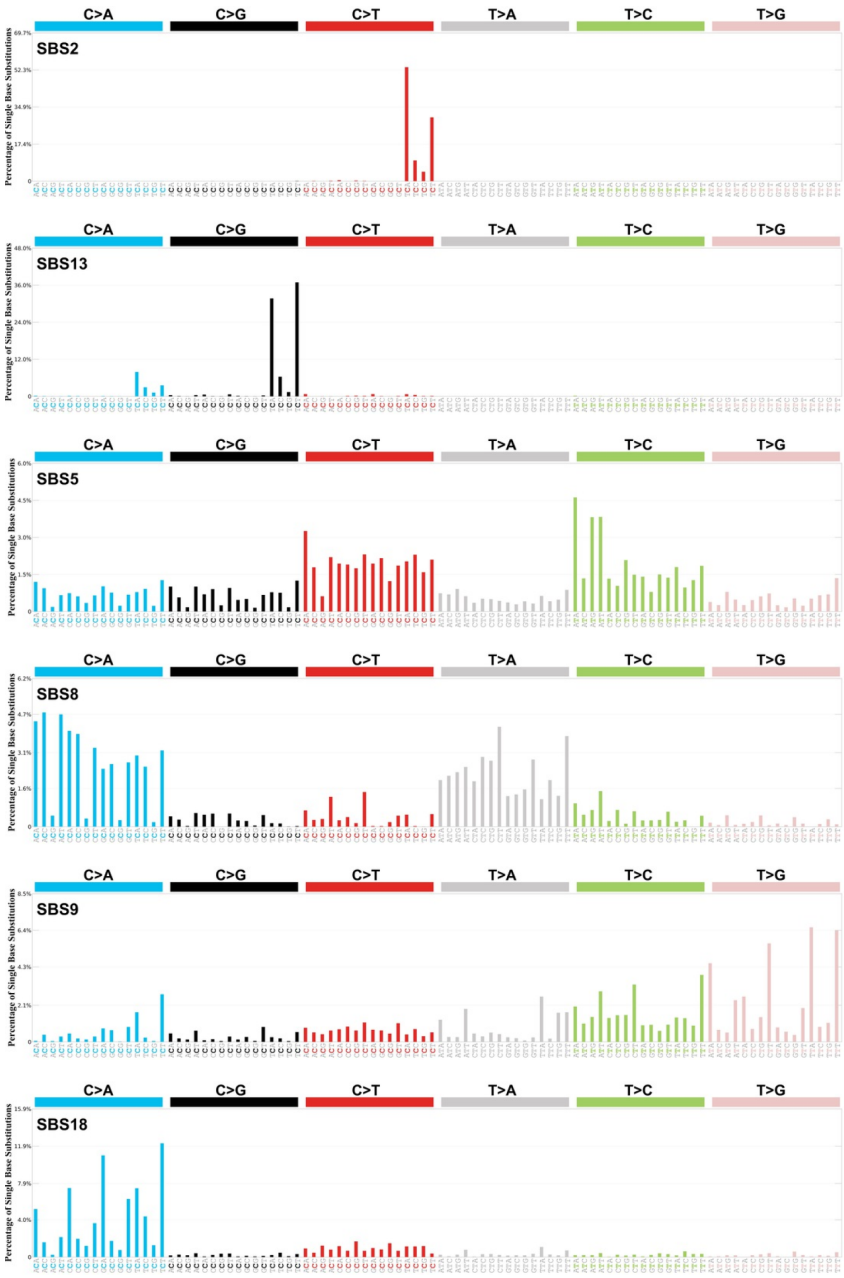
Annex

Annex 2: The 20/2/20 SMM staging system

Stage[1]	>20% BMPC, MS>2g/dL, SFLC ratio >20
1 criteria	Low risk (5% risk of progression at 2 years)
2 criteria	Intermediate progression (17% risk of progression at 2 years)
3 criteria	High risk (46% risk progression at 2 years)

1. Lakshman, A. *et al.* Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. *Blood Cancer J* **8**, 59 (2018).

Annex 3: Myeloma signatures



SBS2 and **SBS13** are attributed to activity of the AID/APOBEC family of cytidine deaminases on the basis of similarities in the sequence context of cytosine mutations caused by APOBEC enzymes in experimental systems. They are usually found in the same samples. A dominant SBS2 and SBS13 is found 80% of the t(14;16) and t(14;20).

SBS5 is clock-like in that the number of mutations in most cancers and normal cells correlates with the age of the individual.

SBS8 is of unknown aetiology but probably age-related.

SBS9 ay be due in part to mutations induced during replication by polymerase eta as part of somatic hypermutation in lymphoid cells. In other words it is “AID” related .

SBS18 is possibly the result of damage induced by reactive oxygen species.