LUND UNIVERSITY

Genetic predisposition for Multiple Myeloma. Identification and functional characterization of risk variants

Duran Lozano, Laura

2022

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Duran Lozano, L. (2022). *Genetic predisposition for Multiple Myeloma. Identification and functional characterization of risk variants*. [Doctoral Thesis (compilation), Department of Laboratory Medicine]. Lund University, Faculty of Medicine.

Total number of authors:
1

Creative Commons License:
CC BY-NC

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

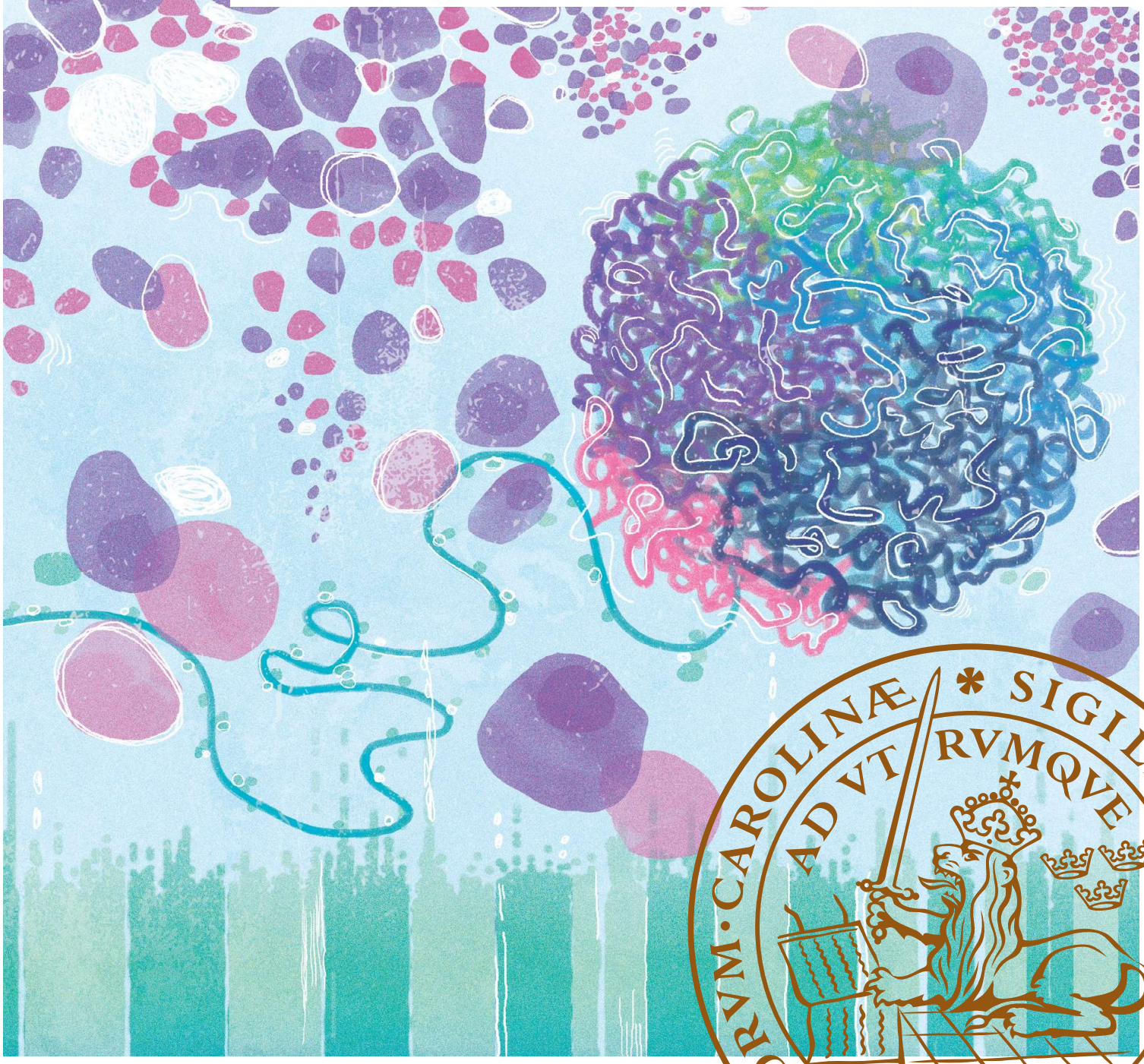
PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 15. Oct. 2026

Genetic predisposition for Multiple Myeloma

Identification and functional characterization of risk variants

LAURA DURAN LOZANO
LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY



Genetic predisposition for Multiple Myeloma

Genetic predisposition for Multiple Myeloma

Identification and functional characterization of risk variants

Laura Duran Lozano

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on the 21st of October at 13:00 in Belfragesalen, Klinkgatan 32, BMC D15

Supervisors:

Professor Björn Nilsson and Dr. Aitzkoa Lopez de Lapuente Portilla
Department of Laboratory Medicine, Lund, Sweden

Faculty opponent:

Dr. Gosia Trynka
Wellcome Sanger Institute, Cambridge, UK

Organization LUND UNIVERSITY Faculty of Medicine	**Document name** Doctoral dissertation
	Date of issue: 21-10-2022
Author: Laura Duran Lozano	Sponsoring organization: Lund University

Title and subtitle: Genetic predisposition for Multiple Myeloma. Identification and functional characterization of risk variants

Abstract

Multiple myeloma (MM) is a blood malignancy originating from plasma cells. First-degree relatives of patients with MM have two- to four-fold higher risk of MM. However, the molecular basis remains largely unknown. This Ph.D. project aims to identify novel DNA sequence variants predisposing to MM through genome-wide association studies (GWAS) and, subsequently, characterize identified variants functionally.

Article I describes a systematic study where we screened for causal gene-regulatory variants at 21 MM risk loci. Article II describes a Nordic GWAS identifying the *SOHLH2* as a novel MM risk locus. Article III describes a novel international meta-analysis of GWAS data totalling 10 906 cases and 366 221 controls, identifying twelve new risk variants for MM accounted for by nine loci: 5q35.2 *CPEB4*, 6p22.2 *BTN3A2*, 9q21.33 *DAPK1*, 10q24.33 *STN1*, 10q25.2 *MXI1*, 13q13.3 *SOHLH2*, 19p13.3 *NFIC*, 21q11.2, *SAMSN1* and a rare variant at 13q13.1 *BRCA2*. Finally, in Article IV, we explore the possibility of identifying transcription factors that mediate allele-specific gene-regulatory effects through combined use of CRISPR/Cas9 screening and epistasis analysis of gene expression data.

The work presented in this thesis provides new insight into the mechanisms underlying genetic predisposition for multiple myeloma.

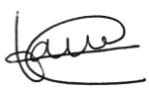
Key words: GWAS, Multiple Myeloma, CRISPR/Cas9, MPRA, Cancer genetics, functional characterization

Classification system and/or index terms (if any)

Supplementary bibliographical information	**Language:** English
ISSN and key title: 1652-8220 Lund University faculty of medicine doctoral dissertation series 2022:132	**ISBN:** 978-91-8021-294-6

Recipient's notes	**Number of pages:** 89	Price: -
	Security classification:	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature  Date 2022-09-21

Genetic predisposition for Multiple Myeloma

Identification and functional characterization of risk variants

Laura Duran Lozano

Cover illustration by Laura Moreno Caparrós
Thesis proofreading by Andrew Hudson

Copyright pp 1-89 Laura Duran Lozano
Article 1 © Springer Nature (Open Access)
Article 2 © Springer Nature (Open Access)
Article 3 © by the Authors (Unpublished manuscript)
Article 4 © by the Authors (Unpublished manuscript)

Faculty of Medicine
Department of Laboratory Medicine

ISBN 978-91-8021-294-6
ISSN 1652-8220
Lund University Faculty of Medicine doctoral dissertation series 2022:132

Printed in Sweden by Media-Tryck, Lund University
Lund 2022

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
MADE IN SWEDEN

Per la Margarida

I pel Yayo Juan

Perquè les guerres es lluitin als laboratoris i les trinxeres siguin els hospitals i les poyates.

A totes les dones que no han pogut tirar, estudiar i ser independents pel simple fet de ser dones.

Table of Contents

Table of Contents **8**

List of articles **10**

Abstract **11**

Lay Summaries **13**

Lay Summary in English 13

Populärvetenskaplig sammanfattning 15

Resum divulgatiu en català 17

Resumen divulgativo en castellano 19

Abbreviations **20**

Aims of the thesis **23**

Introduction **25**

Multiple Myeloma 25

Plasma cells 25

Benign and malignant pre-stages 27

Treatment 29

Risk factors of Multiple myeloma 30

Genetic predisposition to human diseases 31

Heritability 32

Linkage disequilibrium 33

Polygenic risk scores 34

Gene expression and transcription factors 35

Genetic predisposition to Multiple Myeloma 35

Methods **43**

Genome-wide association studies 43

Genotyping and imputation 43

Statistical analysis: association testing and multiple testing correction 44

Functional fine-mapping: 46

ATAC-seq: 46

Chromosome conformation capture 47

Chip-Seq 48

Luciferase assays 48

Massive Parallel Reporter Assay (MPRA) 49
Expression quantitative trait loci 50
Chromatin availability quantitative trait loci (caQTL) 50
Electrophoretic mobility shift assay (EMSA) 50
Doxyclicine inducible promoters and overexpression 50
CRISPR/Cas9 51
Lentiviral library 52
Intracellular staining 54
Interaction modelling 55

Summary of results and discussion 57

Article I 57
Article II 58
Article III 58
Article IV 60

Conclusions 61

Future perspectives 63

Acknowledgements 65

References 71

List of articles

The thesis is based on the following articles and manuscripts:

I. Ajore R, Niroula A, Pertesi M, Cafaro C, Thodberg M, Went M, Bao E.L., **Duran-Lozano L**, Lopez de Lapuente Portilla A, Olafsdottir T, Ugidos-Damboriena N, Magnusson O, Samur M, Lareau C.A., Halldorsson G.H., Thorleifsson G, Norddahl G.L., Gunnarsdottir K, Försti A, Goldschmidt H, Hemminki K, van Rhee F, Kimber S, Sperling A.S., Kaiser M, Anderson K, Jonsdottir I, Munshi N, Rafnar T, Waage A, Weinhold N, Thorsteinsdottir U, Sankaran V.G., Stefansson K, Houlston R and Nilsson, B. "Functional dissection of inherited non-coding variation influencing multiple myeloma risk". *Nature Communications* 2022:13(1), 1–15.

II. **Duran-Lozano L**, Thorleifsson G, Lopez de Lapuente Portilla A, Niroula A, Went M, Thodberg M, Pertesi M, Ajore R, Cafaro C, Olason P, Stefansdottir L, Walters G.B., Halldorsson G.H., Turesson I, Kaiser M.F., Weinhold N, Abildgaard N, Andersen N.F., Mellqvist U-H, Waage A, Juul-Vangsted A, Thorsteinsdottir U, Hansson M, Houlston R, Rafnar T, Stefansson K, Nilsson B. "Germline variants at *SOHLH2* influence multiple myeloma risk". *Blood Cancer Journal* 2021:11(4).

III. Went M*, **Duran-Lozano L***, Halldorsson G, Gunnel A, Lopez de Lapuente Portilla A, Ekdahl L, Olafsdottir T, Ali Z, Law P, Sud A, Thorleifsson G, Niroula A, Pertesi M, Sulem P, Juul-Vangstedt A, Abildgaard N, Frost-Andersen N, Weinhold N, Mellqvist UH, Goldschmidt H, Hemminki K, Hansson M, Thorsteinsdottir U, Rafnar T, Stefansson K, Houlston R**, Nilsson B**. "Deciphering the genetics of multiple myeloma predisposition". *Shared first-authors; **Shared last-authors. *Manuscript in preparation*.

IV. **Duran-Lozano L**, Cafaro C, Mattsson J, Pertesi M, Ekdahl L, Lopez de Lapuente Portilla A, Nilsson B. "Mechanistic dissection of non-coding variation through computational and CRISPR-based analysis of allele-specific transcription interactions". *Manuscript in prep*aration.

Abstract

Multiple myeloma (MM) is a blood malignancy originating from plasma cells. First-degree relatives of patients with MM have two- to four-fold higher risk of MM. However, the molecular basis remains largely unknown. This Ph.D. project aims to identify novel DNA sequence variants predisposing to MM through genome-wide association studies (GWAS) and, subsequently, characterize identified variants functionally.

Article I describes a systematic study where we screened for causal gene-regulatory variants at 21 MM risk loci. Article II describes a Nordic GWAS identifying the *SOHLH2* as a novel MM risk locus. Article III describes a novel international meta-analysis of GWAS data totalling 10 906 cases and 366 221 controls, identifying twelve new risk variants for MM accounted for by nine loci: 5q35.2 *CPEB4*, 6p22.2 *BTN3A2*, 9q21.33 *DAPK1*, 10q24.33 *STN1*, 10q25.2 *MXI1*, 13q13.3 *SOHLH2*, 19p13.3 *NFIC*, 21q11.2, *SAMSN1* and a rare variant at 13q13.1 *BRCA2*. Finally, in Article IV, we explore the possibility of identifying transcription factors that mediate allele-specific gene-regulatory effects through combined use of CRISPR/Cas9 screening and epistasis analysis of gene expression data. The work presented in this thesis provides new insight into the mechanisms underlying genetic predisposition for multiple myeloma.

Lay Summary in English

All cancers are caused by the uncontrolled division and growth of a specific cell type, and a microenvironment that hosts and protects this malignant growth.

In multiple myeloma (MM), the plasma cells grow uncontrollably in the bone marrow. Under healthy conditions, these cells are part of our immune system and produce antibodies. In MM, malignant plasma cells outcompete normal blood cell formation and produce a monoclonal immunoglobulin ("M-protein"), leading to anaemia, thrombocytopenia, immunodeficiency, kidney failure and bone lesions.

But what makes the plasma cells divide with no control? Our aim is to understand which genes are involved and how they drive the cells towards the development of the disease. We use the term 'mutation' to refer to genetic changes that increase the risk of having a disease. Known mutations only explain a small proportion of the cases, and treatments are still ineffective in the long term.

This Ph.D. thesis focuses on finding genetic variants that predispose for multiple myeloma. We want to find new mutations involved in the development of this disease and study their effects.

In Article I we performed a systematic functional study to understand the molecular mechanisms by which known genes cause MM.

In Articles II and III we conducted genetic association studies that compare the genome of thousands of patients from the Nordic Region, USA, Germany, the Netherlands and UK and found 10 new genes that had not been previously reported to affect MM development.

In Article IV we wanted to study the transcription factors that are responsible for the expression of a gene, and we designed a CRISPR/Cas9 library that can *turn off* the expression of all the transcription factors one by one, to investigate which are the most relevant for the expression of our gene of interest.

We would like to contribute to a better understanding of MM, which in the future can lead to finding new therapeutic targets for better clinical management and to an earlier detection of this malignancy.

Populärvetenskaplig sammanfattning

Multipelt myelom (MM) är en av de allra vanligaste blodcancerformerna. Vid MM växer s k plasmaceller okontrollerat i benmärgen. Under normala förhållanden är plasmaceller en del av vårt immunsystem och producerar antikroppar som bidrar till vårt immunförsvar. Vid MM tar elakartade plasmaceller över benmärgen, vilket ger finns mindre utrymme för normal blodcellsbildning, vilket bl a orsakar anemi, immunbrist och skelettskador. Trots att behandlingen blivit allt bättre är MM fortfarande en obotlig och dödlig sjukdom.

Denna avhandling fokuserar på att hitta nedärvda, genetiska varianter som ökar risken att drabbas av MM. Bakgrunden är att epidemiologiska familjestudier visat att nära släktingar till patienter med MM har högre risk att själva drabbas av sjukdomen. Vilka gener och genvarianter som ligger bakom är emellertid bara delvis känt. Syftet med avhandlingen är att hitta nya gener och genvarianter som ökar risken att drabbas av MM samt att studera deras molekylära effekter.

Delarbete I utgör en systematisk studie där vi undersökte de molekylära effekterna för en rad genvarianter som ökar risken att drabbas av MM. I delarbete II och III genomförde vi stora genetiska associationsstudier syftande till att upptäcka nya genvarianter som ökar risken att drabbas MM, och hittade totalt 13 sådana varianter. I delarbete IV undersökte vi en ny metod för att förstå de molekylära effekterna av genvarianter som ökar risken att drabbas av sjukdom.

Mitt avhandlingsarbete bidrar förhoppningsvis till en bättre fördjupad förståelse av hur MM utvecklas, vilket på sikt skulle kunna bidra till bättre metoder för prevention och behandling av sjukdomen.

Resum divulgatiu en català

Tots els càncers tenen en comú el creixement descontrolat d'un tipus de cèl·lules que estan envoltades per un microambient que modifiquen per a què els hi doni energia i protecció del sistema immunitari.

Al mieloma múltiple, són les cèl·lules plasmàtiques les que es divideixen descontroladament a la medul·la òssia. Sota condicions normals o saludables, aquestes cèl·lules fabriquen anticossos molt variants que aboquen a la sang contribuint a les defenses del nostre cos. En el context del mieloma múltiple en canvi, la divisió i invasió incontrolada de les cèl·lules plasmàtiques a la medul·la òssia causa anèmia perquè no hi ha prou espai per a fabricar glòbuls vermells, desequilibri en els nivells de calci que causa fragilitat òssia, hipercalcèmia, i també disfunció renal perquè fragments d'anticossos defectuosos s'acumulen als túbuls renals.

Tot i que aquestes les teràpies utilitzades avui dia redueixen el número de cèl·lules plasmàtiques canceroses, en la majoria dels casos els i les pacients recauen o desenvolupen resistència als fàrmacs. És per això que el mieloma múltiple encara es considera generalment incurable, i és important trobar noves dianes terapèutiques.

Aquesta tesi doctoral se centra en l'estudi de les causes genètiques del mieloma múltiple. Per una banda, el descobriment de noves variants genètiques (mutacions) que incrementen el risc de patir la malaltia. I per l'altra, l'estudi molecular dels gens implicats per a entendre els mecanismes que fan que ser portardor/a d'aquestes variants incrementi el risc de patir la malaltia. En resum, l'objectiu és entendre millor la malaltia per a poder lluitar-hi de manera més eficient i tenir millors eines per detectar qui té més risc de patir-la.

Als articles II i III vam realitzar estudis d'associació genètica en què comparem el genoma de milers de pacients dels Països Nòrdics (Suècia, Dinamarca, Noruega i Islàndia) o a nivell internacional (incloent EEUU, Alemanya, Paisos Baixos i Anglaterra) amb el de milers de controls dels pateixos països. Així, vam trobar variants en 10 gens que fins ara no es relacionaven amb el mieloma múltiple.

Per altra banda, hem fet estudis funcionals per estudiar la capacitat reguladora de les variants associades a la malaltia. És a dir, l'efecte quantitatiu que aquestes variacions del genoma tenen en la quantitat de gen transcrit i traduït a proteïna. A l'article I vam fer servir una tècnica que es diu MPRA per estudiar molts gens a la vegada, i vam trobar el mecanisme molecular que fa que 6 gens incrementin el risc de mieloma.

També s'ha dissenyat una llibreria KO CRISPR/Cas9 que permet anul·lar l'expressió de tots els factors de transcripció, un a un, per veure quins són els més rellevants per a l'expressió d'un gen d'interès. Els factors de transcripció són proteïnes que s'uneixen al DNA en regions promotores i *enhancers*.

Resumen divulgativo en castellano

Todos los cánceres tienen en común el crecimiento descontrolado de un tipo de células, rodeadas por un microambiente que modifican para que les ofrezca energía y protección del sistema inmune.

En el mieloma múltiple, son las células plasmáticas las que se dividen descontroladamente en la médula ósea. Bajo condiciones normales o saludables, estas células fabrican anticuerpos que contribuyen a las defensas de nuestro cuerpo. En el contexto de mieloma múltiple en cambio, la división e invasión descontrolada que ejercen las células plasmáticas causa anemia, ya que disminuye la producción de glóbulos rojos, desequilibrio en los niveles de calcio que causan fragilidad ósea, fracturas recurrentes e hipercalcemia, y disfunción renal, por acumulación de fragmentos de anticuerpos en los túbulos renales.

Aunque hoy en día existen muchas terapias altamente dirigidas y capaces de reducir el número de células malignas, en la mayoría de casos los pacientes recaen o desarrollan resistencia. Por esa razón, el mieloma múltiple aún se considera una enfermedad incurable, y es importante encontrar nuevas dianas terapéuticas para atacar individualmente o en combinación con las terapias existentes.

Esta tesis doctoral se centra en el estudio de las causas genéticas del mieloma múltiple. Por un lado, el descubrimiento de nuevos genes implicados en el inicio de esta enfermedad. Y por otro, el estudio molecular y funcional de los genes implicados ya conocidos, para entender los mecanismos asociados que ocurren en personas portadoras de las variantes de riesgo, como hicimos en el artículo I. En resumen, el objetivo es poder entender mejor la enfermedad para poder luchar contra ella de manera más eficiente y tener herramientas para detectar quién tiene más riesgo de desarrollarla.

En los artículos II y III realizamos estudios de asociación genéticas que comparan el genoma de miles de pacientes de los países nórdicos (Suecia, Dinamarca, Noruega e Islandia) o a nivel internacional (incluyendo EEUU, Alemania, Países Bajos e Inglaterra) y hemos encontrado 10 nuevos genes que hasta ahora no se relacionaban con el riesgo de mieloma múltiple.

En el artículo IV, hemos realizado estudios funcionales para estudiar la capacidad reguladora de las variantes genéticas asociadas a la enfermedad. Es decir, el efecto que estas variantes tienen en la expresión genética y cantidad de RNAm y proteína generada por los genes que regulan. Con ese fin, se han utilizado herramientas como ensayo de luciferasa o MPRA. Y también hemos desarrollado una librería CRISPR/Cas9 que permite *apagar* la expresión de los factores de transcripción de uno en uno para investigar cuáles son los más relevantes para la expresión de un gen. Ya que preguntarse qué factores de transcripción regulan la expresión de un gen es muy recurrente en laboratorios de genética molecular.

Abbreviations

ATAC-Seq	Assay for transposase-accessible chromatin
ATP	Adenosine triphosphate
ALL	Acute lymphoblastic leukaemia
AMP	Adenosine monophosphate
bp	Base pairs
BSA	Bovine serum albumin
caQTL	Chromatic accessibility quantitative trait loci
CAR-T	Chimeric antigen receptor T-cell
cDNA	Complementary DNA
Chip-Seq	Chromatin immunoprecipitation and sequencing
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
eQTL	Expression quantitative trait loci
FISH	Fluorescence in situ hybridization
FPKM	Fragments per kilobase of transcript per million of mapped reads
GFP	Green fluorescent protein
GTEx	Genotype-tissue expression project
GWAS	Genome-Wide Association Study
HR	Homologous recombination
HSC	Haematological stem cell
Ig	Immunoglobulins
IMiDS	Immunomodulatory drugs
KD	Knock down
KO	Knock out
LD	Linkage disequilibrium
LTL	Leukocyte telomere length
lncRNA	Long non-coding RNA

meQTL	Methylation quantitative trait locus
MGUS	Monoclonal gammopathy of unknown significance
MM	Multiple Myeloma
MPRA	Massively parallel reporter assay
mRNA	Messenger RNA
nts	Nucleotides
NHEJ	Non-homologous end joining
NK	Natural Killer
OR	Odds Ratio
PAM	Protospacer adjacent motif
PBS	Phosphate-Buffered Saline
PC	Plasma cell
PCA	Principal components analysis
PC Hi-C	Promoter Capture Hi-C
PCR	Polymerase chain reaction
PPi	Pyrophosphate
PRS	Polygenic risk score
RAF	Risk allele frequency
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
SEC	Super elongation complex
siRNA	Small interfering RNA
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
sMM	Smouldering Multiple Myeloma
TF	Transcription factor
WGS	Whole genome sequencing
WT	Wild Type

Aims of the thesis

This PhD thesis seeks two main objectives: finding genetic variants that predispose for MM and understanding the molecular effects of DNA sequence variants that predispose for MM.

These main objectives have been fractioned into the following specific aims:

- Functionally dissecting and clarifying the mechanism of already known MM risk variants (Article I)
- Finding new variants in a homogenous population, the Nordic countries (Article II)
- Identifying novel MM risk variants through meta-analysis of association data in a broader set of European populations (Article III)
- Exploring a combined CRISPR/Cas9 and computational approach to identify causal transcription factors underlying GWAS signals (Article IV)

Introduction

Multiple Myeloma

Multiple myeloma is a hematologic malignancy caused by a clonal expansion of plasma cells, usually in the bone marrow. It is preceded by monoclonal gammopathy of undetermined significance (MGUS) and smouldering MM (sMM). It is the second most common hematologic malignancy after Non-Hodgkin's Lymphoma, representing approximately 10% of all hematologic malignancies. The average 5-year survival rate is 38.6 % (Baris *et al*., 2013) and 70% of the cases are older than 65 years old at diagnosis (Rajkumar & Kumar, 2016). The worldwide incidence is estimated at 160 000 new cases per year, but it is slightly variable among countries due to differences in genetic risk, lifestyle and access to health care for early diagnosis (Cowan *et al.*, 2018; Hemminki *et al.*, 2021; Ludwig *et al.*, 2020).

Plasma cells

Plasma cells are the terminally differentiated cells of the B cell lineage. They are a key component of the adaptive humoral immune system as they produce and secrete mature immunoglobulins.

Activated B cells can differentiate into transitional preplasmablasts, a cell population with high proliferation activity that migrates to the bone marrow and differentiates into quiescent long lived plasma cells (R. Das *et al.*, 2016; Jourdan *et al.*, 2011; Kassambara *et al.*, 2017; Nutt *et al.*, 2015). The high transcriptional and translational activity required to produce the necessary amounts of antibodies is sustained by an expanded Golgi apparatus and prominent nucleus, which give these cells their characteristic fried egg morphological appearance in the microscope (Fujino, 2018).

Bone marrow stromal cells release CXCL12, which recruits plasma cells to the bone marrow through binding to the CXCR4 plasma cell receptors. Other molecules like VLA4, CD44, and CD28 promote plasma cell retention in the bone marrow niche (Nutt *et al.*, 2015). Other molecular factors that are required for plasma cell function in the bone marrow include CD138, which mediates the selection of mature plasma cells by regulating their survival and is used as the main plasma cell marker, CD38, a highly expressed marker of long-lived plasma cells, and BCMA/CD269 (B cell maturation antigen) that promotes PC survival when activated by APRIL or BAFF ligands (Slamanig & Nolte, 2021).

IRF4, BLMP1(*PRDM1*) and XBP1 are key transcription factors for plasma cell differentiation and homeostasis (Perini *et al.*, 2021). *IRF4* is highly expressed in B cells and plasma cells and is essential for Ig class switching and differentiation of plasma cells and also supports cell survival and proliferation (Agnarelli *et al.*, 2018; Shaffer *et al.*, 2008).

Both BLIMP1 and XBP1 are involved in endoplasmatic reticulum functionality and expansion, which allows Ig production in plasma cells. BLIMP1 is a transcriptional repressor with a key role in the terminal differentiation of B cells to plasma cells (Shapiro-Shelef & Calame, 2005; Turner *et al.*, 1994) and XBP1 is required for the terminal differentiation of plasma cells and reacts to endoplasmic reticulum stress by regulating the unfolded protein response (Reimold *et al.*, 2001).

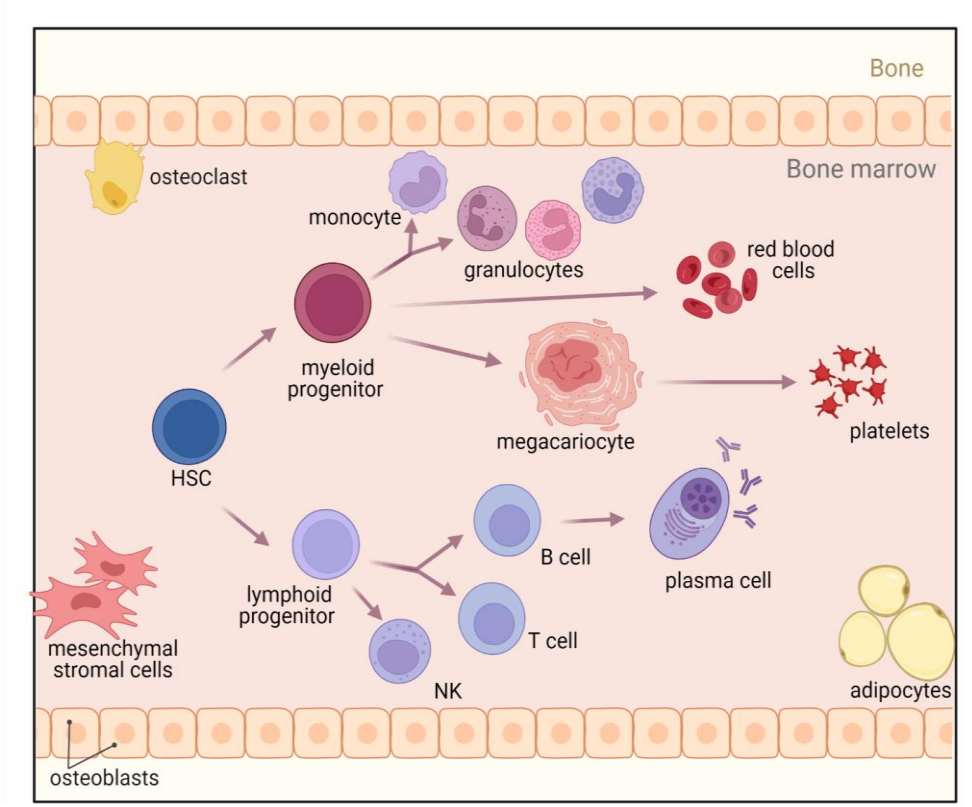


Figure 1:hematopoietic tree in the context of human bone marrow. Created with Biorender.

Benign and malignant pre-stages

Disease stage	MGUS	SMM	Active MM
Serum M-protein	<3 g/dL	≥3 g/dL	≥1 **myeloma defining events** + (1) or (2): End-organ damage (CRAB): any one of • Hypercalcemia, renal insufficiency, anemia, bone lesions Biomarkers of malignancy: • ≥60% clonal BM plasma cells, • Serum involved/uninvolved free light chain ratio ≥100 • >1 focal lesion on MRI ≥5mm in size (1) Clonal bone marrow plasma cells ≥10% or (2) Biopsy proven plasmacytoma
Urine M-protein	N/A	≥500 mg/day	
% BM plasma cells	<10%	10-60%	
Myeloma defining events	Absence of myeloma defining events or amyloidosis		
Progression risk	1% per year	10% per year (1st 5y) 3% per year (next 5y)	

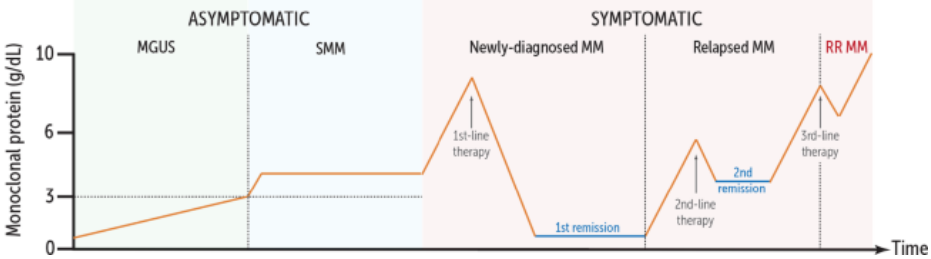


Figure 2: Progress from MGUS to relapsed disease.
Reprint with permission from Ho *et al.,* 2020

Monoclonal gammopathy of undetermined significance

Monoclonal gammopathy of undetermined significance (MGUS) is a common condition defined by the presence of a plasma cell clone that does not yet satisfy the criteria for MM. MGUS is usually diagnosed by a general practitioner, when plasma cells represent up to 10% of the bone marrow cell burden (instead of the normal 2-3%) and there is presence of M protein in blood or urine (Moser-Katz *et al.*, 2021; Mouhieddine *et al.*, 2019).

The monoclonal immunoglobulin produced by malignant plasma cells is called M protein or monoclonal component, and the presence of M protein or light chains in urine is called Bence-Jones proteinuria (Kyle *et al.*, 2014).

MGUS is considered a benign and common condition in wealthy countries, affecting ~3% of individuals older than 50 years old, and the prevalence increases with age (Kyle & Rajkumar, 2007). Prevalence is two to three times higher in African descendent population (Landgren & Weiss, 2009). Multiple myeloma is always preceded by MGUS but not all MGUS cases progress to MM. The risk of progression to MM or other malignancies is 1% a year. Long-term follow-up is recommended (Go & Rajkumar, 2018).

MGUS can also derive to other conditions like light-chain amyloidosis, and Waldenström macroglobulinemia. In light chain amyloidosis, light chains of Ig are

produced in excess and form amyloid fibrillary aggregates that can lead to organ dysfunction (Merlini *et al.,* 2018). Waldenström macroglobulinemia is defined by overproduction of monoclonal IgM, a pentameric or (macro)immunoglobulin that can derive into blood hyperviscosity and cause serious complications (Hunter *et al.,* 2017).

sMM

Smoldering Multiple Myeloma (sMM) is a plasma cell proliferative disorder usually asymptomatic but malignant. The percentage of malignant plasma cells in the bone marrow is greater than 10% (but lower than 60%) and M proteins levels are higher than 3g/dL, with no manifestation of the CRAB symptoms, explained in the next section (Kyle & Rajkumar, 2007).

The general recommendation is to follow up until the development of symptomatic disease. Risk of progression of sMM to MM is 10% per year during the first 5 years after diagnosis but then goes down to 1% 10 years after diagnosis (Raje & Yee, 2020).

A randomized trial showed that early intervention in high-risk sMM cases increased both overall and disease-free survival (Mateos *et al.,* 2013). Discussions in the clinical setting regarding the best moment to begin treatment and to define high-risk sMM cases are ongoing. (Kyle & Rajkumar, 2007; Landgren *et al.,* 2009; Pérez-Persona *et al.,* 2007).

Diagnostics and pathophysiology

The clinical presentation of MM can include hypercalcemia, renal failure, anaemia and lytic bone lesions which are referred to as the "CRAB" symptoms. A diagnosis of MM requires and at least one of the four CRAB myeloma defining events (Rajkumar, 2020). Hypercalcemia is caused by a disequilibrium between osteoblast and osteoclast activity, which in turn also causes lytic bone lesions in the bones, visible by X-ray. Malignant plasma cells release osteoclast activating factor, which stimulates osteoclast-mediated bone resorption and thereby Ca^{2+}release into the blood stream, causing lytic bone lesions and even bone fracture. The clonal growth of plasma cells in the bone marrow outcompetes the production of normal blood cells, leading to anemia, thrombocytopenia and immunodeficiency due to a lack of polyclonal immunoglobulins. Finally deposition of immunoglobulin light chains in the kidneys may lead to renal failure (S. Kumar *et al.,* 2016; Kyle *et al.,* 2014)(S. K. Kumar & Rajkumar, 2018).

MM patients suffer from seriously compromised immunity caused by the disease and also adverse effects of the medication. Infections are frequent and often result in serious complications (S. K. Kumar & Rajkumar, 2018).

All of this translates into the main clinical symptoms of MM chronic back pain, frequent bone fractures, fatigue and shortness of breath and proneness to infection.

In advanced stages, myeloma cells can extravasate from the bone marrow leading to extramedullary plasmacytomas and circulating plasma cells in the blood, which is associated with poor outcome and resistance to treatment. (Ocio *et al.,* 2022).

Treatment

Current treatment of MM includes proteasome inhibitors, immunomodulatory drugs, corticosteroids, monoclonal antibodies and autologous stem cell transplantation.

Bortezomib (trade name Velcade) is the most common proteasome inhibitor. The proteasome function is essential for malignant plasma cells, and its inhibition causes the accumulation of misfolded protein, endoplasmic reticulum stress and NF-κβ pathway inhibition (Gandolfi *et al.*, 2017). The approval of proteasome inhibitors in the treatment of MM contributed to the improvement in overall survival during the last decade (Field-smith, 2006). Some studies suggest that it also acts by increasing oxidative stress to toxic levels in malignant cells (Lipchick *et al.,* 2016).

Immunomodulatory drugs (IMiDs) are angiogenic and cytotoxic, and can modify the immune system response. The most frequently used being lenalidomide, commercially distributed as Revlimid (Holstein & McCarthy, 2017). Lenalidomide and Pomalidomide are further development of Thalidomide, the centrepiece of a historical scandal in pharmacology but which also presented an opportunity to strengthen the responsibility of drug agencies and clinical trials[1].

Corticosteroids (mainly dexamethasone) glucocorticoid receptor agonists that are used in combination with other antimyeloma regimens to help easing inflammation and immune system inhibition (Burwick & Sharma, 2019).

High-dose therapy followed by autologous stem cell transplantation is the treatment of choice for those patients who are up to 65-70 years and have no major comorbidities, which represented approximately half of the patients in studies by Chim *et al.,* 2018 and Hemminki *et al.,* 2021.

Immunotherapies, particularly anti-CD38 Daratumumab (Frerichs *et al.*, 2018), but also anti-BCMA CAR-T cells (George *et al.,* 2021; Lancman *et al.,* 2021; U. A. Shah & Mailankody, 2020) show great promise for further therapy development and improvement. Additionally, bispecific antibodies, with dual specificity to a plasma

[1] Thalidomide was initially introduced in the market 1956. After showing no toxicity in mice, it was an over-the-counter *wonder drug* for insomnia, coughs, headaches and also morning sickness for pregnant women. More than 10 000 children were born with teratogenic deformities and this event caused the United States Congress to pass a historical amendment in 1962 (Greene & Podolsky, 2012). However, women's hormonal cycles are still not well represented in clinical trials (Liu & Dipietro Mager, 2016).

cell antigen and the CD3 antigen on T cells, are in promising clinical trials (e.g., anti-BCMA/CD3) (Caraccio *et al.,* 2020; Verkleij *et al.,* 2020)

In Sweden, current first-line therapy for transplantable patients are the d-VRD combination in 21-day cycles: daratumuab, Velcade, Revlimid and Dexamethasone), followed by auto stem cell transplantation.

Despite huge advances in the last decades regarding new treatment development and efficiency, multiple myeloma is still incurable and ultimately fatal. Myeloma is relapsing/remitting cancer with a median of 3.1 years from diagnosis to relapse (S. K. Kumar & Rajkumar, 2018). 5-year survival after diagnosis has increased from 28% in 1975 to around 60% nowadays (Hemminki *et al.,* 2021)

Risk factors of Multiple myeloma

There are no clearly validated environmental factors other than MGUS and family history of MM. Several studies have shown significant association between increased prevalence of MM and ionic radiation, obesity, certain organic solvents and agricultural work could be risk factors. But similar studies have been inconclusive (Baris *et al.,* 2013).

A study in atomic bomb survivors reported a higher MM mortality among MGUS patients (2 284 /100 000 people-years in exposed population and 14.6/100 000 people-years in non-exposed) but also showed that MGUS incidence was not significantly associated with radiation dose (P = 0.91) (Neriishi *et al.,* 2003). Obesity has also been associated with increased risk of MM and physiological alterations such as oxidative stress, abnormal immunologic response, metabolic response and altered hormonal levels have been proposed to contribute to MM development. A meta-analysis of 13 120 MM cases reported significant association between BMI and MM risk RR=1.27, 95% CI, 1.15–1.41) (Larsson & Wolk, 2007; Morgan *et al.,* 2014).

Prevalence and differences of MM worldwide

Like most malignancies, MM is a complex genetic disease. It is more common in men (54.3) than in women and more common in African populations diagnosed 4 years earlier on average in population studies. Asian populations have the lowest prevalence. The annual global incidence is estimated to be around 155 700 (Waxman *et al.,* 2010; Zhou *et al.,* 2021).

Available data is contaminated with different kinds of bias and uneven representation (Martin *et al.,* 2019; Stepanikova & Oates, 2017) but social, ethnic and geographical determinants continue to influence multiple myeloma treatment access and clinical outcomes (Ailawadhi *et al.,* 2019; Hungria *et al.,* 2017; Ludwig *et al.,* 2020; Obeng-Gyasi *et al.,* 2022).

Several studies performed in the USA have established a higher incidence of MGUS and MM in individuals with African ancestry than in those of European ancestry (Greenberg, Rajkumar, *et al.*, 2012; Janz *et al.*, 2019; Landgren *et al.*, 2017). A study analysing WES and RNA-seq data from 721 MM patients from the CoMMpass cohort who self-reported as African American (n = 128) or *Caucasian*[2] (n = 593) concluded that African American MM patients had a higher mutation frequency in 15 of the 17 genes that were analyzed. The prevalence of MGUS is also two- to three-fold higher among African Americans than in individuals of European ancestry (D. D. Alexander *et al.*, 2007; Greenberg, Vachon, *et al.*, 2012).

The reported prevalence is slightly lower than expected in Latin American countries (Curado *et al.*, 2018). Modern drugs are not available or affordable for a considerable proportion of MM patients (Ludwig *et al.*, 2020). A recent study in 16 countries in Latin America reported that the primary standard treatment based on bortezomib and the autologous transplant is frequently not available or even logistically possible (Pessoa de Magalhães Filho *et al.*, 2019). Moreover, stem cell transplantation carries a social stigma in some countries (Garg *et al.*, 2016).

Genetic predisposition to human diseases

The genetic risk for a given disease phenotype can be explained by variable amounts and types of genetic variation. The simplest examples are highly penetrant monogenic traits, where the disease is caused by a single mutation in all carriers. At the other end of the spectrum, high numbers of variants with modest effects contribute to the risk of complex diseases and quantitative traits, alongside other factors such as environmental exposure (Manolio *et al.*, 2009).

These definitions are based on the simplistic assumption that phenotypic variation is the consequence of genetic variation + environmental exposure + interaction between genetic and environmental factors, which are understood as absolutely anything that is not encoded in the genome.

[2]The term *Caucasian* should not be used in scientific writing. It was coined by the anthropologist and craniologist Johann Friedrich Blumenbach, who described a skull found in the Caucasus mountains as the "most beautiful" human skull. It was larger than the Ethiopian and Mongolian skulls he had studied, which he translated into larger brain and thus intellectual superiority. He ascribed the term *Caucasian* to define Europeans with lighter skin ("Johann Friedrich Blumenbach (1753–1840)," 1940; Popejoy, 2021; Shamambo & Henry, 2022).

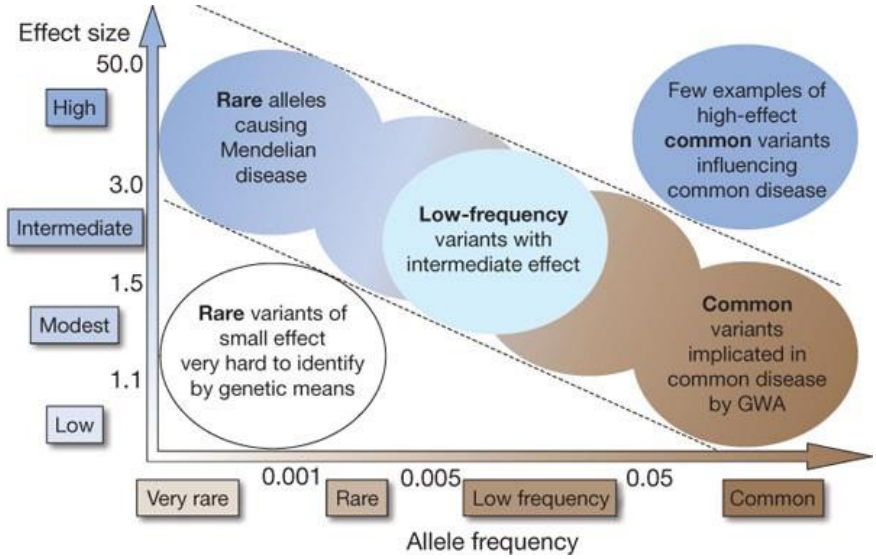


Figure 3: genetic variants by risk allele frequency and strength of genetic effect (OR). Reprint with permission from Manolio *et al.,* 2009 CCC license # 5382500422560

Heritability

Heritability is a population statistic that estimates the proportion of a trait that is attributable to variation in genetic factors. Broad-sense heritability (H^2) is the proportion of variance of a trait that can be attributed to all type of genetic variation (additive, dominance and genetic interaction) whereas narrow-sense heritability (h^2) is the proportion of variation of a trait attributed to additive genetic factors.The heritability of a trait is calculated as the ratio of variances. Variance of additive genetic effects divided by the variance of the observable phenotypes for narrow-sense heritability (Visscher *et al.*, 2008) and total genetic variance divided by the variance of the observable phenotypes for broad-sense heritability (Hill *et al.*, 2008).

The term missing heritability refers to the gap between total estimated heritability and the proportion of heritability explained by known variants.

Genome wide association studies (GWAS) have pinpointed thousands of risk loci, previously unknown relevant pathways, and potential drug targets. However, at this point, it has become apparent that GWAS-identified variants only account for a modest proportion of the estimated heritability (between one third and one half for most complex traits) can be explained by single nucleotide polymorphisms (SNPs). Some of the suggested explanations are too conservative significance thresholds, variation other than SNPs being responsible for disease risk, effect of gene-environment interaction and gene-gene interactions (Manolio *et al.*, 2009; Tam *et*

al., 2019). In our most recent meta-analysis, we estimated the total narrow-sense SNP heritability for MM at 15.7% (Article III).

Linkage disequilibrium

Linkage disequilibrium (LD) is the amount of non-independent association of two alleles in a population (Uffelmann *et al.*, 2021). LD is commonly measured by r^2. For two biallelic loci, locus 1 with alleles *a* and *A* and locus 2 with alleles *b* and *B*, with frequencies for alleles *a* and *A* being respectively p_a and 1-p_a, and the frequencies for alleles *b* and *B* being p_b and 1-p_b r^2 is defined as:

$$r^2(p_a, p_b, p_{ab}) = \frac{(p_{ab} - p_a p_b)^2}{p_a(1-p_a)p_b(1-p_b)}$$

where p_{ab} is the frequency of haplotypes having allele *a in* locus 1 and allele *b* in locus 2 (VanLiere & Rosenberg, 2008). By definition, LD is therefore population-dependent. LD is caused by the chromosomal breakpoints created during meiotic recombination are not random and create haplotype blocks that are inherited together.

This phenomenon has been known for a long time (Hill & Robertson, 1968), but its relevance relies on the fact that most of the association testing methods used assume independence. It is also relevant because genetic association studies rely on LD for imputation. This topic is discussed more extensively in the methods sections.

Some risk loci show differences in frequency and effect size among different ethnic groups as the structure of LD blocks differs across ancestries, hindering the extrapolation of GWAs findings. An indication of that could be, for example, having different loci as the most significantly associated with a trait or disease. In many other cases, common variation is shared across ethnicities (Tam *et al.*, 2019).

On the one hand, genomic studies tend to contain bigger and bigger cohorts to increase their statistical power. On the other, grouping individuals from very different ancestral origins dilutes the effect of not-so-frequent variants contributing to missing heritability.

In genetic association studies, we very often speak about lead variants. The lead variant is the SNP from a given LD block or genomic risk locus that is taken as the one that could better explain that association (Uffelmann *et al.*, 2021). The lead SNP is therefore a genetic marker, although it is not necessarily responsible for the effect or risk. Association does therefore not mean causation.

Polygenic risk scores

Polygenic risk score (PRS) is calculated by adding the effect of risk alleles that one individual carries, weighted by their effect size. Both identification of associated variants and calculations of their weight (or odds ratio, OR) discussed in the Methodology section of this thesis.

A study from 2018 showed that PRSs for certain common diseases such as type 2 diabetes and breast cancer could predict disease risk with the same reliability as highly penetrant monogenic variants (Khera *et al.*, 2018).

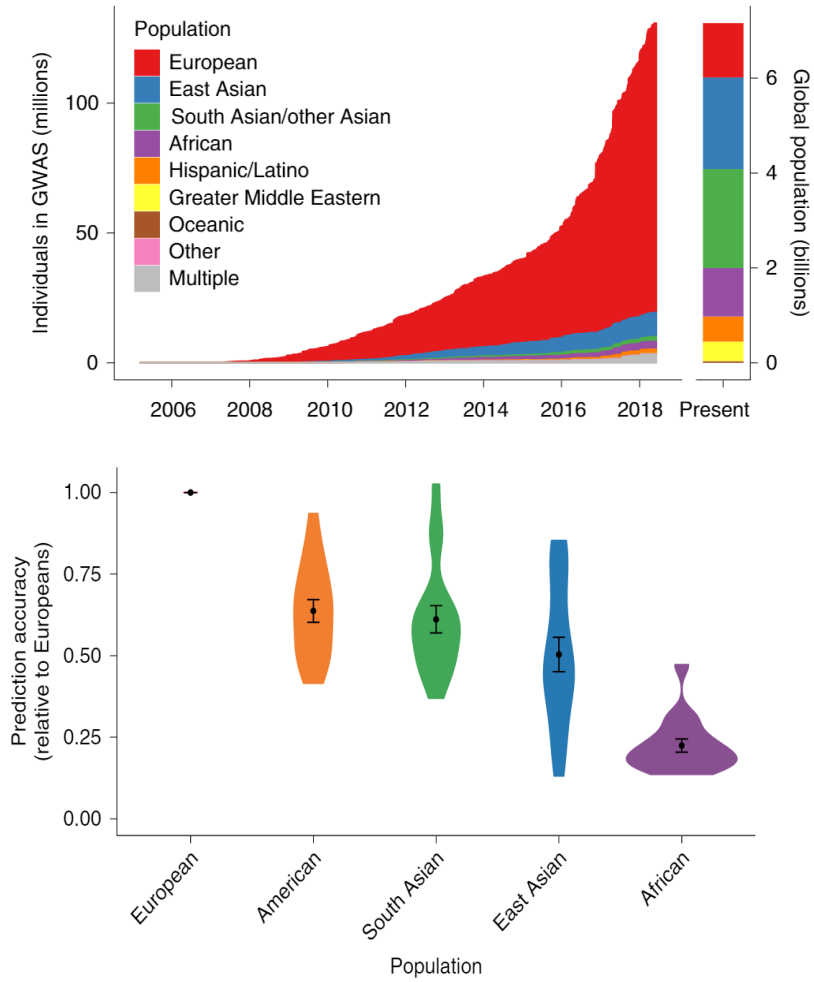


Figure 4: Representation of different populations in published GWAS (top left) compared to the proportioj of total human population (top right) and the consecuent differences in prediction accuracy of polygenic risk scores in non-European individuals (bottom). Reprint with permission from Martin *et al.,* 2019.

Differences in LD structure imply differences in effect-size estimates. Associations found in a given population might therefore not apply to other populations or ethnic groups. Eighty percent of the individuals included in GWAS were considered of European ancestry, which had dramatic effects on risk prediction, with PRS several times less accurate for populations of non-European ancestry (Figure 4, Martin *et al.,* 2019). It will be interesting to see how PRS is implemented in the clinical setting. Current recommendations suggest defining a set of genetic variations that, if present, would be medically actionable – thereby raising the issue of how to act with incidental findings and the ethical requirement to inform family members (Lewis & Green, 2021).

Gene expression and transcription factors

A very big part of this thesis is dedicated to understanding how do GWAS appointed variants affect the expression of neighbouring genes. Decades ago DNA was thought of as a linear molecule encoding genes, and also containing some regions of *junk DNA* the functionality of which was not understood and therefore underrated. However, we now know non-coding DNA is highly functional and can influence traits in many different ways. And we also know that the three-dimensional conformation of the DNA in the nucleus is highly regulated, dynamic and cell type-specific (Dixon *et al.*, 2012; Hafner & Boettiger, 2022).

Genetic predisposition to Multiple Myeloma

Early studies including a case reports of MM in three siblings (L. Alexander & Benninghoff, 1965) and monozygotic twins (Judson *et al.*, 1985) and a case-control study from the Swedish Cancer registry (Eriksson & Hållberg, 1992) initially suggested a genetic component in MM aetiology. Familial aggregation and shared genetic risk factors of MM and MGUS have been confirmed by several studies since then (Kristinsson *et al.,* 2009; Landgren *et al.,* 2006, 2009; Landgren & Weiss, 2009). It is estimated that first degree relatives of MM patient have 2 to 4 fold increased risk of getting MM (Altieri *et al.,* 2006).

Familial MM represents 1 to 2% of all MM cases (Pertesi *et al.*, 2020) and familial studies have found high-risk rare germline mutations in CDKN2A (Dilworth *et al.,* 2000; V. Shah *et al.,* 2017), LSD1/KDM1A (Wei *et al.,* 2018), ARID1A, USP45 (Waller *et al.,* 2018) and DIS3 (Pertesi, Vallée, *et al.,* 2019). A study by our lab reported high burden of common MM risk alleles in familial MM cases (Halvarsson *et al.,* 2017). The main focus of this thesis is on common germline variation that increases the risk for sporadic multiple myeloma, representing up to 98% of all the MM cases. When this thesis began in 2018, 25 loci had been associated with MM. A table with the updated list of MM associated MM loci including the advanced from Article II and III is available in the conclusions section.

Table 1: MM-associated loci identified prior tho this thesis. RAF:risk allele frequency, extracted from 1000 genome phase 3, European population.

						Discovery study		
Candidate gene	locus	rsID	coding effect	Risk allele*	RAF	Reference	OR	*P* - value
DTNB, DNMT3A	2p23.3	rs6746082	NM_183361.2: c.1168-2380T>G	A	0,48	Broderik *et al.,*	1.29	1.22×10^{-7}
SP3	2q31.1	rs4325816			0.77	Went *et al.,*	1,12	7.37×10^{-9}
ULK4	3p22.1	rs1052501	NP_060356.2: p.Ala542Thr	T	0.19	Broderik *et al.,*	1,32	7.47×10^{-9}
ACTRT3, MYNN, LRRC34	3q26.2	rs10936599	NM_018657.5: c.18C>G	T	0.76	Chubb *et al.,*	1,26	8.7×10^{-14}
ELL2	5q15	rs56219066	NM_012081.5: c.482-445A>G	T	0.72	Swaminathan *et al.,*	1,25	9.6×10^{-10}
CEP120	5q23.2	rs6595443		T	0.45	Went *et al.,*	1.11	1.20×10^{-8}
JARID2	6p22.3	rs34229995		G	0.02	Mitchell *et al.,*	1,37	1.31×10^{-8}
HLA region	6p21.3	rs2285803		T	0.26	Chubb *et al.,*	1.19	1.65×10^{-9}
ATG5	6q21	rs9372120	NM_004849.3: c.574-17571A>C	G	0.19	Mitchell *et al.,*	1.18	9.09×10^{-15}
CDCA7L	7p15.3	rs4487645		C	0.66	Broderik *et al.,*	1.38	3.33×10^{-15}
CCDC71L	7q22.3	rs17507636		C	0.74	Went *et al.,*	1.12	9.20×10^{-9}
POT1, POT1-AS1	7q31.33	rs58618031		T	0.73	Went *et al.,*	1.12	2.73×10^{-8}
SMARCD3	7q36.1	rs7781265	NM_003078.3: c.40-8183C>T	A	0.09	Mitchell *et al.,*	1.19	9.79×10^{-9}
CCAT1	8q24.21	rs1948915		C	0.33	Mitchell *et al.,*	1.13	4.20×10^{-11}
CDKN2A	9p21.3	rs2811710		C	0.64	Mitchell *et al.,*	1.15	1.72×10^{-13}
WAC	10p12.1	rs2790457		G	0.74	Mitchell *et al.,*	1.12	1.77×10^{-8}
CCND1	11q13.3	rs603965 (rs9344)		G	0.50	Weinhold *et al.,*	1.82	2.92×10^{-10}
PRR14, SRCAP, FBRS	16p11.2	rs13338946		C	0.28	Went *et al.,*	1.15	1.02×10^{-13}
RFWD3	16q23.1	rs7193541		T	0.61	Mitchell *et al.,*	1.13	5.00×10^{-12}
TNFRSF13B	17p11.2	rs4273077	NM_012452.2: c.445+2913T>C	G	0.10	Chubb *et al.,*	1.26	7.67×10^{-9}
KLF2	19p13.11	rs11086029		T	0.23	Went *et al.,*	1,14	6.79×10^{-11}
PREX1	20q13.13	rs6066835	NM_020820.3: c.415-3822A>G	C	0.09	Mitchell *et al.,*	1.26	1.36×10^{-13}
HMGXB4, TOM1	22q13.1	rs138740		C	0.34	Swaminathan *et al.,*	1.18	2.80×10^{-9}
CBX7	22q13.1	rs877529	NM_175709.3: c.113+3502C>T	A	0.44	Chubb *et al.,*	1.23	7.63×10^{-16}

2p23.3 *DTNB, DNMT3A:*

rs6746082 was first reported as a borderline promising association (*P* = 1.22E-07) in the first MM GWAS in 2011, in a study that analysed 1 675 individuals with multiple myeloma and 5 903 control subjects from Germany and UK (Broderick *et al.,* 2011) and rs7577599 was later validated in later GWAS (*P* = 2.28E-14 in Mitchell *et al.,* 2015; *P* = 7.37E-09 in Went *et al.,* 2018;). They are both intronic variants in *DTNB* but the neighbour gene *DNMT3A* has also been suggested as a candidate, and it is frequently mutated somatically in AML and clonal haematopoiesis. There is no certain causal gene in this locus yet.

2q31.1 *SP3:*

rs4325816 maps to *SP3,* a transcription factor involved in Antigen-stimulated B lymphocytes specific expression at the germinal centre (Steinke *et al.,* 2004). SP3 can act as an activator or repressor depending on the isoform and possible post-translational modifications. Phosphorilation, acetylation, glycosylation and sumolation allow immediately effective regulation of this TFs (Waby *et al.,* 2008). *SP3* and its paralog *SP1* are overexpressed in MM and have also shown a significant reduction under the effect of Bortezomib, one of the main therapeutical agents used to treat MM (Ghosal & Banerjee, 2022).

3p22.1 *ULK4:*

The mutation that confers risk for MM at 3p22.1, rs1052501, is a missense variant (NP_060356.2: p.Ala542Thr) in *ULK4* that confers risk to MM but is predicted to be benign (Broderick *et al.,* 2011). This gene encodes a serine/threonine-protein kinase, involved in cytoskeletal remodelling (Preuss *et al.,* 2020) a key regulator of mTOR-mediated autophagy (Jung *et al.,* 2010).

3q26.2 *LRRC34, TERC, MYNN:*

The MM-associated LD block covers an area that comprises three coding genes: *TERC*, *LRRC34* and *MYNN*, delimited by two recombination hotspots. *TERC* is the RNA component of telomerase, and telomeric function has shown to be affected in multiple myeloma. *LRRC34* is involved in ribosome biogenesis in pluripotent stem cells. It is mostly expressed in pluripotent embryonic stem cells and premeiotic germ cells in adult mice testis (Lührig *et al.,* 2014). The lead variant of this locus, rs10936599-G, has also been associated with colorectal cancer (Houlston *et al.,* 2010) and longer leucocyte telomeres (Jones *et al.,* 2012).

5q15 *ELL2:*

The high LD (r^2=0.8) block at 5q15 is composed of 73 SNPs overlapping the gene body and 3'UTR of ELL2. The encoded protein is part of the super elongation complex (SEC), a multiproteic complex allows expression of high amounts of immunoglobulins. In PC, this complex is required to increase the catalytic rate of RNA polymerase II transcription by suppressing transient pausing by the polymerase at multiple sites along the DNA (Martincic *et al.*, 2009; K. S. Park *et al.*, 2014). The MM-associated locus is also associated with lower gene expression and reduced levels of IgA and IgG (Ali *et al.*, 2018; Swaminathan *et al.*, 2015). This is supported by *ELL2* KO mice showing impaired Ig production and reduction of mature plasma cells in the bone marrow (Park *et al.*, 2014).

5q23.2 *CEP120:*

A *cis*-eQTL effect suggests that *CEP120* is the causal gene at 5q23.2 (Went *et al.*, 2018). *CEP120* (centrosomal protein 120) is necessary for microtubule elongation and centriole formation. Microtubules and the centriole are required for cytoskeleton formation, cell division, shape, transport and polarization (Badano & Katsanis, 2006; Borys *et al.*, 2020). *CEP120* involvement in the organization of the mitotic spindle can affect chromosome segregation and promote genetic instability (Mahjoub *et al.*, 2010) which is relevant given the high proportion of hyperploid MM cases.

6p21.3 HLA region:

The 6p21.3 association signal maps to the HLA region, a complex region that harbours multiple genes implicated in the immune system and is associated with more than 100 diseases (Shiina *et al.*, 2009). The LD block associated with MM risk maps to the 3'region of *PSORS1C2* and gene body of *CCHCR1*. The MM risk could be associated with one or more specific HLA haplotypes, HLA-DRB5*01 was suggested in the discovery GWAS for this MM risk locus (Chubb *et al.*, 2013).

6p22.3 *JARID2:*

Unlike most of the GWAS identified MM association, this variant has low frequency (RAF=0.02). rs34229995 lies in the promoter region of *JARID2* and even though no eQTL effect has been reported *JARID2* remains the main candidate due to its central role in coordinating hematopoietic stem and progenitor cell function (Kinkel *et al.*, 2015). *JARID2* recruits the Polycomb repressive complex 2 (PRC2), a protein complex with histone methyltransferase activity, mainly H3K27me2/3, which has a chromatin silencing effect. PRC2-mediated gene silencing control transcriptional programs during plasma cell differentiation (Margueron & Reinberg, 2011). This gene is frequently deleted in chronic myeloid malignancies (Puda *et al.*, 2012).

6q21 *ATG5:*

The 29 SNPs in high LD map to the Autophagy protein 5 (*ATG5*) gene and its promoter. This gene is essential for plasma cells homeostasis and Ig production and it is also required for the formation of autophagic vesicles. (Conway *et al.*, 2013). Autophagy is highly important for malignant plasma cells, alone and in cooperation with the proteasome system. Autophagy has also been pointed out as a mechanism of drug resistance in multiple myeloma (Yun *et al.*, 2017).

7p15.3 *CDCA7L, DNAH11:*

The SNPs in high LD at this locus lie in an extensive region of open chromatin for most hematopoietic cell types, with enhancer histone marks. It corresponds to last introns and 3'UTR of *DNAH11* and *CDCA7L*, encoded in opposite directions. The 7p15.3 lead variant, rs4487645, maps to intron 79 of the *DNAH11* gene and the promoter region of *CDCA7L* and associates with increased *CDCA7L* expression in plasma cells (Weinhold *et al.*, 2015). The rs4487645-G risk allele creates a new IRF4 binding site. The authors showed that suppression of *CDCA7L* reduces MM proliferation through apoptosis, and *CDCA7L* expression is associated with adverse patient survival (N. Li *et al.*, 2016). Weinhold *et al.*, showed that rs4487645 had the strongest an eQTL effect on the gene, and claimed that the risk association effect is mediated by rs4487645 and involves IRF4 binding and c-Myc (Weinhold *et al.*, 2015).

Cell division cycle-associated 7-like protein (*CDCA7L*) is involved in apoptotic signalling pathways, and the downregulation of *CDCA7L* expression decreases *CCND1* expression too (Ji *et al.*, 2019).

7q22.3 *CCDC71L*

The 7q22.3 risk locus maps to the 3' of *CCDC71L*. This gene promotes cell proliferation, migration and invasion. It is regulated at an mRNA level by miR-6504-5p and miR-3139 that are, at its turn, sponged by the lncRNA *GREP1* (Luo & Wang, 2021).

7q31.33 *POT1:*

The associated SNPs at 7q31.33 are located in the lncRNA *POT1-AS1* (*POT* antisense 1) which has been reported to increase the glucose metabolism enzyme *PDK3* expression by a sponging miR-497-5p and to have an oncogenic role in gastric cancer (W. M. Chen *et al.*, 2021). *POT-AS1* is encoded upstream of the protein-coding gene *POT1*, that has been suggested as the causal gene candidate for this locus. Protection of telomeres protein 1 (*POT1*) is part of the shelterin complex that protects telomeres, contributing to chromosome stability and a negative regulator of the telomerase (Kelleher *et al.*, 2005). No eQTL effect has been shown for POT1.

7q36.1 SMARCD3:

The lead variant at 7q36.1, rs7781265, is an intronic variant, with low risk allele frequency (RAF= 0.09). *SMARCD3* encodes a subunit of the SWI/SNF chromatin remodelling and transcriptional regulating complex. SMARCD3 recruits other proteins of the complex to specific target regions allowing access to the transcriptional machinery (Lickert *et al.,* 2004) *SMARCD1, SMARCD2* and *SMARCD3* (also called BAF60a, BAF60b and BAF60c). These homolog proteins compete as alternative subunits of the SWI/SNF complex and are differentially expressed in a tissue-specific manner (Mashtalir *et al.*, 2018).

8q24.21 *CCAT1:*

This gene produces a long non-coding RNA. *CCAT1* is significantly upregulated in MM patients' plasma cells and cell lines compared with plasma cells from healthy donors and high expression of this gene correlates with shorter overall survival of MM patients (L. Chen *et al.,* 2018). In addition, the 8q24.12 locus is involved in long-range chromosomal interactions acting as an enhancer for *MYC* (Jia *et al.,* 2009).

9p21.3 *CDKN2A:*

The lead SNP of this locus maps to intron 1 of *CDKN2A*. Hi-C data shows a loop that connects it with the neighbouring gene *MTAP* gene in KMS11 cells, an MM cell line. Both *CDKN2A* and *MTAP* are frequently deleted in cancer cells (Kryukov *et al.,* 2016). The cyclin dependent kinase inhibitor 2A, *CDKN2A*, is tumour suppressor gene through negative regulation of cell proliferation. Interestingly, expression levels of *CDKN2A* are partly regulated by SP1 and SP3 transcription factors (Ghosal & Banerjee, 2022). *CDKN2A* expression was reported to be upregulated in glucocorticoid resistant MM patients (Ghosal & Banerjee, 2022).GWAS studies have indicated that a SNP in this gene (but not in LD with the MM loci) is associated with several kinds of cancer such as breast cancer, lung cancer, melanoma or ALL (Sherborne *et al.,* 2010).

10p12.1 *WAC:*

This association signal maps to *WAC*, and the rs2790457-G is significantly associated with decreased gene expression (eQTL P = 6.58E-24) and also has a *cis*-meQTL effect (P = 1.42E-6) (Mitchell *et al.,* 2016). WAC's interaction with the E3 ligase RNF20/40 is necessary for histone H2B monoubiquitination. The N-terminal of WAC interacts with the RNA polymerase II transcriptional machinery. WAC is also involved in autophagy by inducing amino acid starvation-induced autophagy and regulates the cell-cycle checkpoint in response to DNA damage (Joachim *et al.,* 2012; Zhang & Yu, 2012).

11q13.3 *CCND1:*

The rs9344 SNP is associated with risk for a specific subtype of multiple myeloma with the specific chromosomal translocation t(11;14)(q13;q32), in which the *CCND1* gene (usually at 11q13.3) is placed under the transcriptional control of the immunoglobulin heavy chain enhancer (at 14q3) (Fonseca *et al.,* 2002; Weinhold *et al.,* 2013). The oncogene *CCND1* encodes cyclin D1, which controls the G1/S checkpoint together with CDK4/6. This protein is overexpressed in several cancer types (Gao *et al.,* 2020; Landi *et al.,* 2020; Moreno-Bueno *et al.,* 2003).

16p11.2 PRR14, FBRS, SRCAP:

The association signal at 16p11 covers an area of 81kb containing *PRR14, FBRS, SRCAP* a small nucleolar RNA and two pseudogenes. The lead variant, rs8058928 is at 5' of *SRCAP*, a helicase involved in transcriptional regulation by chromatin remodelling. It mediates the exchange of histone H2AZ/H2B dimers for nucleosomal H2A/H2B, which enhances promoter accessibility of target genes, which is important for multipotent progenitors (MPP) commitment into lymphoid or myeloid lineage (Ye *et al.,* 2017). This gene is also known to be mutated and act as a driver gene in clonal haematopoiesis (Beauchamp *et al.,* 2021). *PRR14* interacts with heterochromatin reattaching it to the nuclear lamina and is also involved in the positive regulation of the PI3K-Akt-mTOR signalling pathway and in promoting cell proliferation.

16q23.1 *RFWD3:*

The 16q23.2 association maps to *RFWD3*. This gene has been associated with leucocyte telomere length (LTL) in different GWAS studies (C. Li *et al.,* 2020; Taub *et al.,* 2022). The RFWD3 protein also protects p53 from MDM2 degradation and is required for DNA interstrand cross-links repair (Elia *et al.,* 2015; Inano *et al.,* 2017; Mitchell *et al.,* 2016). Biallelic mutations in *RFWD3* cause Fanconi anaemia, a chromosomal instability syndrome that leads to bone marrow failure and very high cancer risk (Knies *et al.,* 2017).

17p11.2 *TNFRSF13B:*

This locus has one of the strongest association signals for MM and Ig levels (Chubb *et al.,* 2013; Jonsson *et al.,* 2017; Liao *et al.,* 2012). The LD block is composed by one coding and 16 non-coding variants.

TNFRSF13B gene encodes TACI, a receptor of the APRIL and BAFF ligands. TACI is a key regulator of B-cell and plasma cell homeostasis. This gene is primarily expressed in switch memory B cells (Salzer *et al.,* 2005), showing lower expression in plasma cells, which suggest they could act in B cells. *TNFRSF13B* is an obvious candidate to explain the MM risk at this locus for its involvement in B-cell and PC functions but the exact mechanism of action and specific causal variant is not known yet.

TNFRSF13B encodes three different isoforms; one long, one short and a soluble isoform. The metalloproteinase ADAM10 can excise the long and short membrane bound isoforms generating soluble TACI (Hoffmann *et al.,* 2015).

19p13.11 *KLF2*

KLF2 is the only gene overlapping the association signal, that maps a genomic location with widely accessible chromatin in most hematopoietic cell types, including plasma cells. About half of the SNPs in the LD block overlap ATAC-seq peaks, and one of them, rs3745318 is also a missense variant (NP_057354.1:p.Leu104Pro). Kruppel like factor 2 (*KLF2*) is a transcription factor that mediates induction of pluripotency (Bourillot & Savatier, 2010). It belongs to the Specificity protein/Kruppel-like factor (Sp/KLF) family of transcription factors, and *SP3* is a member too. The transcription factors of this family share conserved zinc finger domain DNA-binding motifs and the DNA recognition sites (Waby *et al.,* 2008).

20q13.13 *PREX1*

Expression and methylation QTLs for *PREX1* colocalize with the 20q13 MM-risk association at rs6066832 (with p-values of 3.85E-5, 1.12E-4 and 1.36E-13 respectively) (Mitchell *et al.,* 2016). PC Hi-C data shows contact between a regulatory region containing rs6066832 and gene promoter, which could explain the expression upregulation.

22q13.1 *CBX7:*

The LD block at 22q13.1 overlaps *CBX7* and the regulatory region GH22J039143 with promoter histone marks and Hi-C looping to the neighbouring genes *APOBEC3G* and *PDGFB*. Chromobox protein homolog 7, *CBX7*, is a component of a the canonical Polycomb repressive complex (PRC1). CBX proteins recognize H3K27me3 in histones that have been methylated by the PRC2 complex and recruit PRC1 E3 ubiquitin ligase (Margueron & Reinberg, 2011; Vidal & Starowicz, 2017). No clear causal variant has been reported to explain the mechanism that increases risk for MM at this locus.

Methods

Genome-wide association studies

Genome-wide association studies (GWAS) are a standard tool to identify DNA sequence variants that contribute to a phenotype of interest. Essentially, the allelic frequencies of millions of variants are analysed to detect genotypes that are statistically over-represented in the cases compared with controls and thereby to define associations of genetic variation with a given trait (Uffelmann *et al.,* 2021).

The genetic architecture of most complex traits is highly polygenic and the associated variants are mostly non-protein-coding. Typically, the list of phenotype-associated genetic variants that results from GWAS studies are common variants and each has modest effects. Each of the independent significant association signals is represented by a group of variants inherited together as an LD block, often spanning more than one gene or in intergenic regions.

GWAS has identified risk loci, previously unknown relevant pathways, and potential drug targets. However, at this point, it has become clear that GWAS-identified variants only account for a modest proportion of the estimated heritability (between one-third and one-half for most complex traits) that can be explained by SNPs. Some of the suggested explanations are having too conservative significance thresholds, variation other than SNPs being responsible for disease risk, and effect of gene-environment interaction and gene-gene interactions (Manolio *et al.*, 2009; Tam *et al.,* 2019).

Result validation in different cohorts and functional work to explain the gene function, cell type of action and potential role in disease development are essential.

Genotyping and imputation

The genotypes detected with SNP arrays are not phased, and although they provide information of which nucleotides are present for each selected genomic location, they lack information on which DNA strand each allele is from. For this reason, mathematical reconstruction of the sequence belonging to each chromosome at a local level is necessary to infer the haplotype. This process is called phasing. (Kong *et al.*, 2008) and it is a necessary step that precedes imputation.

Imputation increases genomic resolution and increases statistical power by testing association of more variants than the originally typed. It is very cost effective

because it enriches the input genotyping data bringing it near WGS, but with lower sequencing costs.

It is based on the principle that variants in high linkage disequilibrium will be inherited together, so unsequenced variants can be inferred (imputed) by using a reference genome panel (S. Das *et al.*, 2018).

These reference genome panels are created by whole genome sequencing (WGS) data of a large numbers of individuals. Some examples of commonly used population panel references are The Icelandic reference panel, created from WGS of 49,962 Icelanders (Jónsson *et al.*, 2017), an updated version of which was employed in Articles II and IIV, and the HapMap project (Belmont *et al.*, 2005).

Statistical analysis: association testing and multiple testing correction

Genetic association tests are run for each genetic variant, using an appropriate model. Linear models are usually used to test for associations continuous phenotypes and logistic regression models are used for binary traits, as it would be presence or absence of disease (Uffelmann *et al.*, 2021). Covariates such as age, sex and ancestry are included to account for stratification and avoid confounding effects from demographic factors.

Testing association of millions of variants provides a lot of information but also carries some limitations.

A real problem in the field of genetics is that high throughput genetic association studies performed nowadays generate long lists of associations. There are always association results, and the challenge is where to draw the line for trustworthy results and to correct for multiple testing burden and false positive associations that arise from population structure. There are different multiple testing methods. Bonferroni correction is the most widely used in GWAS. The threshold of significance after Bonferroni correction is calculated by dividing the P-value threshold for significance by the number of tests performed, which correspond to the number of genetic variants analysed in the GWAS. This method is over-conservative, assumes independence of hypothesis (which is not true considering LD) and fails at modelling intergenic interactions (Stringer *et al.*, 2011).

Genetic principal components analysis (PCA) is performed to avoid false discoveries caused by differences in population structure, relatedness and ethnic distance, ancestry differences between cases and controls and other causes of cryptic population stratification (Price *et al.*, 2006).

Summary statistics are the main output of a GWAS study and contain the results of association testing for every variant analysed in GWAS. Summary statistics contain variant identifiers, variant coordinates, association p-value, effect allele, effect (calculated as beta or OR) and standard error. Summary statistics is a format that

contains very concentrated information. Genomic sequencing and typing data are highly sensitive, bulky and unanonymizable. Summary statistics cannot be traced back to individual genetic information, is easy to share and publically available.

Two very relevant aspects when planning a GWAS are sample size and genetic homogeneity. Population stratification is the presence of genetically distinct subpopulations in the studied cohort (Uffelmann *et al.*, 2021). Sample size is critical for the study to reach high enough statistical power. The larger the sample size, the higher statistical power will be reached but meta-analysis of various populations usually result in association of common variants (as these transcend multiple populations), and have very limited power to detect association with rare variants (as these are more often population-specific).

Figure 5 shows the number of loci identified in GWAS with different sample sizes and for three different traits. There is a trait-specific threshold, affected by the size of the genetic component of each trait, above which the rate of locus discovery increases promptly with small sample size increases (Tam *et al.*, 2019).

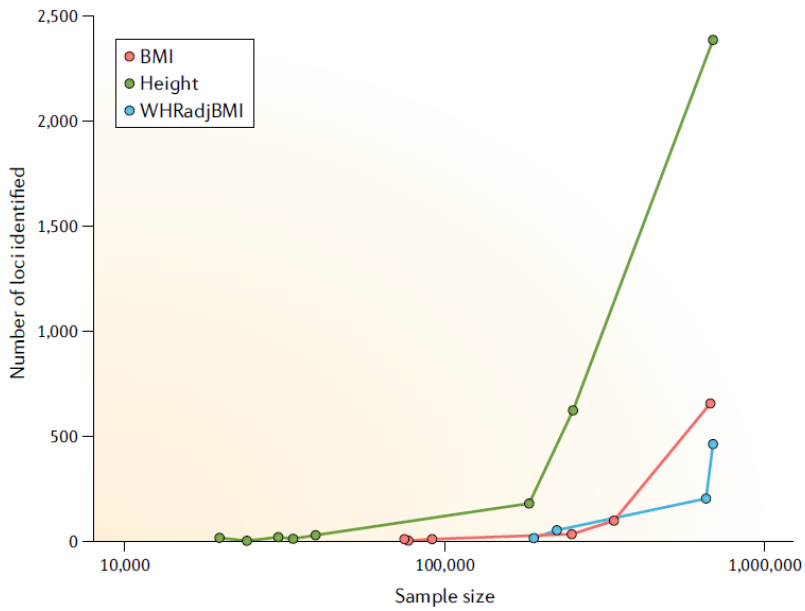


Figure 5: Number of loci identified as a function of GWAS sample size. **N**umber of genome- wide significant loci reported from GWAS for three anthropometric traits: body mass index, height and waist-to-hip ratio adjusted for BMI. Reprint with permission from Tam *et al.,* 2019.

For this reason, meta-GWAS analysis are a common strategy. The idea behind it is that increasing the sample size to overcome the multiple testing limitation, increases statistical power and detects previously unseen associations. It often involves collaborations and international consortia because it requires summary statistics from several cohorts, usually already published GWAS.

Along with the exposed arguments, in Article II we were able to identify a variant with low frequency (RAF=0.035) with a very homogeneous case group comprised of Scandinavian individuals. Additionally, in the meta-analysis presented in Article III we found 8 new associations with risk allele frequencies ranging from 0.11 to 0.81.

Functional fine-mapping:

Once GWAS has provided a list of reliable risk loci the next challenge is interpreting the results in a biological and genomic context to identify the causal variant within the LD block and their mechanism of action. This process is referred to as fine-mapping and it usually combined computational and experimental approaches to fill the gap between statistical association and disease predisposition.

There are different approaches to characterize the function of non-protein-coding variation. The more relevant for this thesis are described below.

Following the classification suggested by Ray *et al.*, 2020; this approaches can be divided into the following four groups: 1) observational assays that characterize the genomic region like ATAC-seq and PC Hi-C; 2) observational assays that characterize the impact of naturally occurring genetic differences at the variant: eQTL, caQTL and also meQTL; 3) engineered perturbational assays that test the impact of the variant itself like CRISPR-directed activation or KO and finally 4) engineered perturbational assays that test the impact of the variant itself, usually tested in a synthetic context, like the reporter assays luciferase and MPRA.

This Ph.D. focuses is on functional fine-mapping of multiple myeloma GWAS variants using techniques such as CRISPR/Cas9, Luciferase or EMSA on different MM cell lines and data on ATAC-seq, eQTL, caQTL and Hi-C looping.

ATAC-seq:

ATAC-seq, is the assay for transposase-accessible chromatin using sequencing, which maps chromatin accessibility at a whole genome level. Briefly, ATAC-seq is based on sequencing the fragments generated by the transposase Tn5 enzyme, which cuts unprotected, hence accessible DNA (Grandi *et al.*, 2022).

The resulting data is visualized as cut density peaks across the genome. The cut density is inferred from the reads of the fragments generated by Tn5. High peaks are seen in highly accessible chromatin regions (Figure 6).

Chromatin accessibility has a strong tissue and cell type specificity. This higher-order structure of chromatin controls interaction of regulatory regions and allows tissue-specific transcription factors regulate several aspects of development and differentiation.

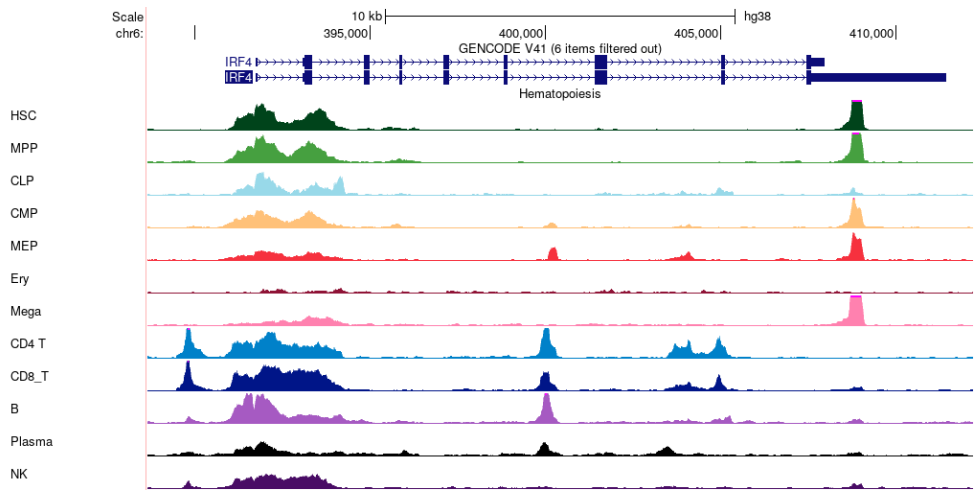


Figure 6: *Example of ATAC-seq data visualization. The image shows different patterns of chromatin accessibility in a genomic area comprising IRF4 for different hematopoietic cell types. HSC:hematopoietic stem cell, MPP: multipotent progenitor, CLP: common lymphoid progenitor, CMP: common myeloid progenitor, MEP: Megakaryocyte erythroid progenitor, Ery: Erythroblast, Mega: Megakaryocyte, CD4_T: CD4+ T cell, CD8_T: CD8+ T cell, B: B cell, Plasma: Plasma cell, NK: natural killer.*

The ATAC-Seq data used to identify regions of open chromatin in plasma cells was obtained by analyzing $CD138^{+}$ cells isolated from bone marrow aspirates from 185 MM patients from the MM Lund Biobank. ATAC-seq data from 18 other hematopoietic cell types was downloaded from the Sequence Read Archive (Ulirsch *et al.*, 2019).

Chromosome conformation capture

Chromosome conformation capture is a tool used to identify three-dimensional looping interactions between regulatory regions, and their target genes located in in neighbouring or very distant genomic regions.

Chromosome conformation capture methods such as Hi-C (Lieberman-Aiden *et al.*, 2009) are based on the sequencing of DNA that has previously been cross-linked fragmented and ligated with its extremes. After being fragmented, DNA is purified to remove any bound protein and interacting DNA strands ligated, creating circular molecules that can be sequenced (Eagen, 2018).

The resulting data is highly valuable as it informs of the regions of the genome that interact in the nucleus (for a given cell type, and under determined conditions). And it could point, for example, to how a SNP in a regulatory region is in contact with the promoter region of a gene.

Chip-Seq

Chip-seq is the combination of immunoprecipitation and sequencing. In chromatin immunoprecipitation-based methods, a DNA binding protein of interest is detected by an antibody and retrieved after being cross-linked with the DNA. DNA sequencing generates a collection of sequence reads that correspond to the DNA that the protein was bound to and that can be graphically presented as corresponding to the density of reads. Alignment of those reads provides qualitative (genomic location) and semi-quantitative measurements of protein binding (P. J. Park, 2009).

Chip-seq can be performed with proteins that are ubiquitously/broadly present in the chromatin, like CTCF, or proteins with specific and much more limited binding like transcription factors. Some limitations of this technique are its dependency on antibody specificity. Large amounts of Chip-seq data are available in public repositories like GEO, a public functional genomics data repository from NCBI (https://www.ncbi.nlm.nih.gov/geo/) and Cistrome Data Browser (http://cistrome.org/db/#/ ; Zheng *et al.*, 2019).

Newer alternatives to Chip-Seq are available, CUT&RUN (Cleavage Under Targets and Release Using Nuclease) for example, requires lower amounts of input material, and does not involve the use of strong detergents to permeabelised cells(Skene & Henikoff, 2017). On the negative side, there is less available data in the public domain.

Luciferase assays

Luciferase assays are reporter assays designed to measure the ability of a given DNA sequence (usually tens to hundreds of base pairs) to promote genetic expression. This sequence is usually synthetised and inserted in a plasmid upstream the firefly luciferase reporter gene. In the presence of O_2, Mg^{2+} and ATP, the recombinant firefly luciferase transforms luciferin to oxyluciferin producing AMP, PPi and CO_2 in a bioluminescent reaction. The amount of light is directly proportional to luciferase gene expression and it can be quantified. This signal is

usually normalised by renilla luciferase fluorescence to control for transfection efficiency (Schagat *et al.,* 2007). The renilla luciferase transforms coelenterazine into coelenteramide, CO_2 in a reaction that is also bioluminescent. The two reactions are measured consecutively in the same sample using a Luminometer. (Marques *et al.*, 2009; Sherf *et al.,* 1996). Normalised reporter activity is then compared between different promoter sequences or experimental conditions. Performing the same experiment on two different alleles of a variant is commonly used to assess the differences in transcriptional regulation control.

Advantages of enzymatic assays are their specificity and sensitivity. One the other hand, a limitation of the luciferase assay is that only one sequence can be tested at a time.

Massive Parallel Reporter Assay (MPRA)

MPRA consists on a high throughput reporter assay that allows simultaneous assessment of the genetic regulatory function of several hundreds or thousands of DNA sequence variants at the same time and in a pooled assay. The results can be traced back to each of the studied variants' construct thanks to a DNA barcoding system. MPRA has been used for some years to study the mechanistic basis for regulatory activity behind risk variants (Gordon *et al.,* 2020; Tewhey *et al.,* 2016).

In Article I we designed inserts containing three sliding windows 20 bp apart from each other containing each of the two alleles and for both strands totalling 12 constructs per variant, and therefore 12 468 contructs in total. Each construct also contained adaptors for PCR amplification and 20 nts random barcode. The resulting fragments were cloned into a vector containing a minimal promoter, the GFP open reading frame and a partial 3'UTR.

Two different multiple myeloma cell lines were transfected and after 48h of culture total mRNA was extracted and GFP mRNA was pulled down using biotin-labelled GFP probes and streptavidin beads. The recovered mRNA was DNase-treated and retrotranscribed to prepare sequencing libraries compatible with NextSeq 1x 75bp sequencing. The *MPRAscore* is computational tool developed in our group to analyse this kind of data (Niroula *et al.,* 2019).

This method offers several advantages, like the possibility of testing a high number of variants at the same time and in a matter of days but one of the limitations is the synthetic context in which the regulatory effect of variants is tested.

Expression quantitative trait loci

Expression quantitative trait loci (eQTLs) are DNA sequence variants that influence the transcription of one or more genes (Albert & Kruglyak, 2015). In eQTL analysis, the input data is expression data such mRNA-sequencing or gene expression microarrays and paired genotype data from the same samples. This data is used to identify association between genotype and mRNA levels for the genes of interest.

Examples of MM risk loci with eQTL effects are rs2488002 influencing the expression of *STN1/OFBC1* in B cells (Article IV) and rs75712673 influencing the expression of *SOHLH2* in malignant plasma cells (Article II).

Chromatin availability quantitative trait loci (caQTL)

Chromatin accessibility quantitative trait loci (caQTL) are defined based on the allele-specific abundance of sequencing reads from ATAC-seq data (Broekema *et al.*, 2020).

To identify genomic regions with allele-dependent chromatin accessibility we tested for association between genotype and Tn5 cut density of 150bp fragments surrounding a variant every 10 bps in a given locus.

Examples of MM risk locus with caQTL are rs78740585 at *SMARCD3*, rs4487645 at *CDCA7L* and rs7922679 at *MXI1*.

Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assay is used to detect protein-DNA complexes.

The supershift assay consists on adding an antibody against the candidate transcription factor. If the protein binds the DNA probe there will be a secondary mobility shift because the complex formed by DNA probe, the TF and the antibody migrate slower than the probe bound to the TF alone (Hellman & Fried, 2007)

It provides very useful and specific information; the binding of a TF to a given sequence, represented by a DNA probe. Different versions of the DNA probe can be made, for example containing the different alleles of a variant.

Doxyclicine inducible promoters and overexpression

For the overexpression studies of SMARCD3, we used Tet-On 3G tetracycline-inducible gene expression system (Clontech PR053540). This inducible expression system is based on aa two vector system. The first vector, pEF1α, encodes the Tet-ON 3G trans activator under the control of the EF1α constitutive promoter and P_{TRE3G} that contains the inducible promoter.

The *SMARCD3* coding region was cloned into the P_{TRE3G} vector, downstream of the P_{TRE3G} inducible promoter, that contains 7 repeats of the tet operator. In the presence of doxycycline, the Tet-ON 3G trans-activator binds the P_{TRE3G} promoter and induces high transcription of the downstream gene (SMARCD3 in this case). Electroporation of 5M L363 cells with PTRE3G-Luciferase, as Doxycycline induction control or PTRE3G-SMARCD3 was performed. L363 cell line was chosen because its *SMARCD3* expression levels are inherently very low, and it was previously transfected with pEF1α.

Cells were incubated for 4.5, 17.5 and 24 hours, half of the samples with 20μL of [10ng/μL] Doxycycline. After that time, PTRE3G-Luciferase treated and untreated controls luminescence was analysed to confirm induction and RNA and protein was extracted from the PTRE3G-SMARCD3 treated and untreated samples for Western Blot validation and RNA-Seq for differential expression analysis.

CRISPR/Cas9

This technology builds on the fundamental chemistry of the bacterial Cas9 proteins that use single guide RNAs (sgRNAs) molecules containing ~20 nucleotides that are complementary to the targeted DNA region and contain a PAM sequence at the 3' (the most commonly used being NGG for *Streptococcus pyogenes*). This sgRNAs direct the Cas9 proteins make double-strand breaks. This breaks lead to viral DNA destruction in bacteria but trigger DNA repair in eukaryotic cells. Two of the fundamental DNA repair pathways that happen in humans, animals and plants cells are non-homologous (NHEJ) which produces small insertions or deletions at the site of the repair or Homology-directed repair (HR) that can lead to the integration of new DNA sequences.

CRISPR/Cas9 solved a great challenge in molecular genetics, cutting and editing the genome at virtually any position. Accordingly, we can direct the Cas9 enzymes to cut a region of interest with single nucleotide resolution through sgRNA custom design. The only requirement for the selection of Cas9 target sites is the presence of a PAM sequence at 3' of the targeted sequence by the sgRNA.

CRISPR/Cas9 was be used to cut and edit DNA, but the technique has rapidly evolved and various other applications are being developed in the fast-growing field. These include activation or knock-in (La Russa & Qi, 2015), transcriptional suppression or KO (Joung *et al.*, 2017) epigenetic modification with dead Cas9 (dCas) (Xie *et al.*, 2017); silencing or activating mRNA transcription, using dead Cas9 linked to a transcription activator or repressor; base editing (Huang *et al.*, 2021; Newby & Liu, 2021); prime editing, which does not involve any double-strand break (Anzalone *et al.*, 2019; Newby & Liu, 2021); and fusing an error-prone polymerase to Cas9 to cause continuous mutagenesis and enable directed evolution (Halperin *et al.*, 2018). Another potential use is the introduction of huge fragments

of DNA using CRISPR-assisted transposases (Klompe *et al.*, 2019; Strecker *et al.*, 2019) or even imaging, where a fluorescent form of the protein can be used to illuminate specific sequences for live cell imaging (Clow *et al.*, 2022).

The origins, challenges and multiple applications are elegantly discussed by the pioneer scientists that developed this revolutionizing tool here: Charpentier *et al.*, 2019; Doudna, 2020; Mojica *et al.*, 2016; *Interview to F. Zhang*, 2021.

In Article I, we used CRISPR to delete specific and relatively small regions containing the variant of interest, to study the downstream effects on gene expression.

In Article IV, we designed a library to knock out all the transcription factors. In the case of CRISPR KO, sgRNAs direct Cas9 nuclease to make double-strand breaks in key areas of the genes that are inefficiently repaired by non-homologous end-joining and introduce deleterious insertions/deletions. This leads to the knock-out of each targeted transcription factor, which further impacts the expression of downstream genes (C. H. Chen *et al.*, 2018; Joung *et al.*, 2017). We switched off the expression of all transcription factors, one at a time, and observed the effects of that on the expression of our gene of interest. The assay was designed to be broadly applicable to explore the transcription factors driving the expression of any gene of interest that can be detected in a flow cytometer.

Lentiviral library

Library design

The first challenge of Article IV was to assemble a list of all exiting transcription factors. Gene transcription is a highly regulated process and can act in many different ways. The function of those proteins can vary between directly recruiting the RNA polymerase, by binding DNA and not allowing others to so, or by recruiting factors and forming large multi-subunit complexes (Lambert *et al.*, 2018). Briefly, we combined existing lists of transcription factors from diverse resources were combined. We merged the datasets from Jaspar, HOCOMOCO, SMILE SEQ and human C2H2ZF databases, and the lists from two published articles (Lambert *et al.*, 2018; Novershtern *et al.*, 2011). We then annotated Gene Ontology terms and manually curated to ensure that all the components would fulfil two premises: binding DNA and affecting expression.

Four pre-designed sgRNAs from the whole genome KO Brunello library were then selected for all the transcription factors present in the curated list.

Clonning and bacterial replication

We chose the LentiCRISPRv2 vector for the TF KO library. This vector encodes the Cas9 enzyme gene, the sgRNA scaffold and amplycilin and puromycin

resistance genes that are required in the following steps of lentiviral production and cell infection.

Two different cloning approaches were tested in parallel to clone the inserts containing the TF sgRNAs into the CRISPR vector; Gibsson assembly and Golden Gate cloning. The first one assembles different fragments containing homology ends in a PCR-like reaction resulting in a final molecule containing all the fragments. Golden gate cloning relies on Type IIS restriction enzymes that cut DNA outside the recognition sites and a T4 DNA ligase that ligates the complementary overhangs (Engler *et al.,* 2008). We obtained better results by using Golden Gate cloning, by incubating undigested empty lentiCRISPRv2 vector and PCR amplified oligos containing the sgRNAs at a 1:3 ratio, together with T4 DNA ligase and Fast digest Esp3I restriction enzyme for 2h at 25°C.

Electrocompetent *E. coli* cells were transformed with the ligated product to generate a high number of copies of all the vectors in the library, since representation of all sgRNA was crucial in future steps of the experiment and high amounts of the plasmid library are required for virus production. Transformed bacteria was plated in Ampicillin LB plates and incubated overnight at 37°C. Plasmid extraction was performed with the Maxiprep kit. We checked sgRNA variability in the resulting transcription factor KO lentiCRISPR v2 library by Sanger sequencing 33 randomly-picked bacterial colonies and later sent a sample of the vector pool for next generation sequencing (Nova Seq 6000, single-end read x200 cycles**)**

Lentiviral production

HEK293T/17 cells were transfected with the transcription factor KO lentiCRISPRv2 library, PM2.G encoding the viral envelope protein and psPAX2 encoding the retroviral polyprotein and reverse transcriptase. We used Lipofectamine 3000 (ThermoFisher #L3000001), a reagent that enhances transfection and cell survival by creating liposomal structures with a positive surface that facilitate DNA delivery into cells.

The supernatant of the transfected cells was collected 24, 48 and 72h after transfection and ultracentrifugated at 25 000rpm for 90 minutes to purify the virus parties containing the TF KO library.

A detailed protocol of the library generation and viral production can be found in the methods section of Article IV which was adapted and optimized from existing protocols (Elegheert *et al.,* 2018 and Joung *et al.,* 2017).

Cell transduction

Both cell lines (KMS20 and L363) were tested to identify the minimal puromycin concentration that would select the transduced cells, and virus titration experiments were performed to determine the concentration of virus that would result in an

approximate infection efficiency of 30%, to reduce the possibility of more than one sgRNA acting in the same cell.

Cells were transduced with the TF KO library-containing lentivirus in the presence of lentiboost (Sirion Biotech), a non-ionic amphiphilic poloxamer that reduces electrostatic repulsion increasing virus infection.

Transduced cells were incubated with puromycin from day 2 to 10 after infection to discard untransduced cells, and then cultured until the total cell number was close to 80 million cells, enough for sorting of 8 replicates.

Intracellular staining

The ELL2 protein acts in the cell nucleus, so in order for the antibodies to detect and mark the protein, cells had to be previously fixed and permeabelised.

For the optimization of the stating, that included numerous variables, we used ELL2 KO L363 cells and always selected the conditions that showed better separation between KO and ELL2 WT L363 cells in the flow cytometer analysis (performed with BioRad Ze5)

We tested different methods. First a two-step process consisting on fixation with Paraformaldehyde and permeabilization with saponin, tween 20 or triton. The lasts being the better but still unsatisfactory. So we then tested incubation with different antibody concentrations and incubation times with -20°C methanol, which fixes and permeabelises the cells simultaneously and selected 30 min incubation with cold methanol as the best method.
We tested different blocking strategies and a 4-hour incubation at room temperature with PBS 3% BSA was the best, but with small improvement in comparison with 2h which was finally selected for the sake of time.

We tested three different mouse anti human ELL2 primary antibodies: two monoclonal (Santa Cruz Biotech #SC-376611 and Antibodies online ABIN930983) and one polyclonal (Abnova ABIN364967). We also tested five different secondary anti mouse antibodies with different fluorophores: FITC, APC (Jackson ImmunoResearch #115135164), Brilliant Violet IgG1 (Biolegend #406615), PE IgG1 Rat anti mouse (Biolegend #406607) and PE-Dazzle IgG1 Rat anti mouse (Biolegend #406627). After a process of stating optimization, we concluded that the best combination was the primary monoclonal antibody from Santa Cruz Biotech #SC-376611 diluted 1:200 and PE Biolegend 406607 diluted 1:60, offering the most specific and bright signal.

We used life/dead markers (Invitrogen fixable scarlet live/dead marker #17468262 and Live/dead fixable Violet #L34955) since KMS20 tend to have low viability (70-80%) in our hands and we did not want dead cells to interfere in the experiment.

Fluorescence-activated cell sorting

The strategy to find the transcription factor that drive ELL2 expression was to sort the cells with minimal and maximal expression and look for sgRNAs that were enriched in the first group and not present in the second one. For that, we sorted the lowest 2% and highest 20% extremes of the fluorescent emitting cell distribution.

Illumina sequencing library preparation

DNA was extracted from fixed cells and following amplification, samples were run on bioanalyzer to verify amplification and fragment size. We used a custom made Illumina compatible primer collection that had been used in previous projects of the lab (Mattsson *et al.*, 2021; Pertesi, Ekdahl, *et al.*, 2019). The forward primers contained the Illumina P5 adaptor sequence, the Illumina PCR2 forward primer sequence, a stagger of variable length (1 to 12 nts, different in each forward primer) and a priming site complementary to the lentiCRIPSRv2 vector fragment upstream of the sgRNA. The reverse primers contain the Illumina P7 adaptor sequence, an index sequence (different for each primer), the Illumina PCR2 reverse primer sequence and a priming site complementary to the lentiCRIPSRv2 vector fragment 140 nts downstream of the sgRNA sequence. A mixed of 12 forward primers pooled and a unique reverse primer was used to amplify the DNA from each of the samples, and the plasmid library was sequenced with NextSeq 500/550 (single-end read x150 cycles) with 25% of PhiX to add sequence diversity.

Interaction modelling

A clear limitation of the TF KO screening is that not all the differences between the KMS20 and L363 models will be due the their *ELL2* genotype. We developed a model to include MM PC RNA-seq data from the Dana Farber Cancer Institute dataset, the CoMMPass project and in house data to add co-expression analysis to our pipeline.

To identify functional, allele-specific interactions among the TFs identified in the CRISPR/Cas9 screen, we used the following regression model:

$$y = x_0 + a_1x_1 + a_2x_2 + a_2x_1x_2 + a_3x_3$$

Where $\boldsymbol{y}$ is the expression of the gene of interest (ELL2 in this pilot study) x_0 represents a bias term, x_1 is expression of the candidate TF in FPKM, x_2 is the number of minor alleles at rs3815768 (0 for homozygous major CC, 1 for heterozygotes CT, and 2 for homozygous minor TT) and x_3 is ELL2 copy number, since copy number aberrations are common in malignant cells in general and in addition hyperploid is the most common of MM types.

$a_2x_1x_2$ models interactions between genotype and candidate TF expression. A positive interaction indicates that the correlation between the TF and target gene expression is stronger with the minor allele.

Summary of results and discussion

Article I

The objective of this project was to systematically search for causal regulatory variants among variants in high LD with MM lead variants. With that aim, we developed a massive parallel reporter assay (MPRA) and tested it in two MM cell lines. We assessed a total of 1 039 variants comprising the high LD (h^2>0.8) blocks of the lead variants of MM risk loci known at the time (Pertesi *et al.*, 2020; Went *et al.*, 2018). Following MPRA, we integrated data from luciferase assays, eQTL, caQTL, PCHiC, a GWAS on Ig levels and CRIPSR to further the characterise the identified candidate variants.

We identified putative causal variants with MPRA activity, luciferase signal and concordant plasma cell *cis*-eQTLs, in which the allele that was associated with higher MPRA and luciferase activity was also associated with higher expression of the eQTL target gene at the locus at six risk loci: 5q15 *ELL2*, 5q23 *CEP120*, 7p15 *CDCA7L*, 7q36.1 *SMARCD3*, 10p12.1 *WAC*, and 20q13 *PREX1*. We also identified concordant MPRA activity, luciferase signal and *cis*-eQTLs at the 17p11.2 *TNFRSF13B* locus, with the eQTL effect significant in B cells.

Additional evidence for a gene-regulatory role of the identified variants was obtained through PC Hi-C and ATAC-seq, showing that some of the variants mapped to regulatory regions with three-dimensional looping to the eQTL gene promoters. The activity of these regions was further investigated in dual-sgRNA-CRISPR/Cas9 experiments where the resulting deletion of the variant-harbouring regions impacted the expression of the eQTL genes. Furthermore, we validated the regulatory role of rs4487645 at CDCA7L using CRISPR/Cas9 precision editing.

Finally, to obtain evidence for regulatory activity at the precise positions of the identified variants in an endogenous chromosomal context, we generated a caQTL data set for MM plasma cells as well as a new computational tool (caQTLseg), to identify genomic regions with allele-dependent chromatin accessibility. That analysis provided evidence for colocalized regulatory activity at *CDCA7L*, *SMARCD3* and *ELL2*.

The results of Article I represent the first systematic functional dissection of risk loci for a hematologic malignancy.

Article II

In Article II, we conducted a GWAS on MM from the Nordic region. We discovered a new risk locus at 13q13.3. The association is represented by a group of 16 SNPs in high LD (r^2>0.8) spanning the *SOHLH2* gene. The lead variant was rs200203825, with a relatively low risk allele frequency compared to most MM risk variants (RAF=0.035).

Within the 14 variants in the LD block, we identified rs75712673 as a putative causal variant as it maps to a genomic region with accessible chromatin in plasma cells and a H3K4me3 histone mark. Additionally, Hi-C data suggested a looping interaction with the gene promoter. Finally, we identified a *cis*-eQTL for *SOHLH2* at rs75712673 in MM plasma cells and luciferase assays showed higher signal for the rs75712673 *SOHLH2*-high-expressing allele in plasma cell lines tested. We explored whether there was differential binding of the FOX transcription factors by EMSA but did not get any conclusive results.

The *SOHLH2* gene is mainly expressed in testis. Our results suggest that the effects of the SOHLH2 MM risk allele could be mediated by ectopic *SOHLH2* expression in plasma cells.

Article III

In Article III, we performed an international meta-analysis of nine published and one previously unpublished GWAS of MM, totalling 10 906 cases and 366 221 controls. We identified nine new risk loci. In addition, conditional analysis identified underlying independent signals at three of the loci, bringing the total number of new signals to twelve.

5q35.2 *CPEB4*: *CPEB4* (5q35.2) mediates translational activation and repression through cytoplasmic changes in the poly(A) tail length of mRNA molecules and regulates activation of unfolded protein response (Guillén-Boixet *et al.*, 2016). We found a significant *cis*-eQTL in granulocytes and nominally significant TWAS association**.**

6p22.2 *BTN3A2*: the 6p22.2 risk locus maps to the HLA region. We identified *BTN3A*2 as a candidate based on proximity and a significant association between cross-tissue gene expression data and genotype. *BTN3A2* is involved in T-cell response and its expression in PC is generally low.

6p25.3 *IRF4*: The *IRF4* gene encodes a transcription factor essential for the development and survival of plasma cells (Perini *et al.*, 2021; Shaffer *et al.*, 2008).

In Article I, we showed differential binding between the two alleles of *SMARCD3* and *CD7AL1* risk SNPs.

9q21.33 *DAPK1*: our data pointed at rs1329600 (at *DAPK1*) as putative causal variant in this locus. This variant lies in a region of open chromatin and we showed concordant *cis*-eQTL and luciferase in at rs1329600. *DAPK1* is involved in apoptosis and autophagy (Singh *et al.,* 2016).

10q24.3 *STN1/OBFC1*: rs4387287 is the promoter region of *STN1* (also referred to as *OBFC1)*. We identified an eQTL in B cells with the proxy SNP in very high LD (r^2=0.97) rs2488002 coherent with significant luciferase differences in the B cell line U266B1. *STN1* is part of the CST complex that protects telomeres from DNA degradation. This same locus had been previously associated with LTL (Levy *et al.,* 2010), blood pressure (Surendran *et al.,* 2016), and different cancer types (Rashkin *et al.,* 2020).

10q25.2 *MXI1*: we found a new association at 10q25.2 represented by 15 SNPs in high LD mapping to the *MXI1* gene. We detected a co-localizing caQTL effect at the intronic variant rs7922679, with the minor allele showing higher chromosome availability.

19p13.3 *NFIC:* the association signal at 19p13.3 overlaps the intronic region of *NF1C*, in a region of open chromatin for hematopoietic progenitors.

21q11.2 *SAMSN1*: we detected a caQTL in the *SAMSN1* intronic region containing the MM risk SNPs rs2822745, rs2822746 and rs2822747. *SAMSN1* is co-expressed with *IRF4* and is a negative regulator of B cell activation. It is highly expressed in the bone marrow and peripheral blood. Our data showed open chromatin in this locus for plasma cells and other differentiated cell types but not in earlier progenitors.

13q13.1 *BRCA2:* we found an association between *BRCA2**c.9976A>T; p.Lys3326* (also referred to as K3326X) and MM risk. It is the variant with the highest effect (OR=1.57) and lowest frequency (RAF=0.0048) among the reported MM associations.

This new meta-GWAS identified twelve new risk variants accounted for by nine loci and provided new information on previously unsolved MM associated loci. It does not provide further support for the previously reported association at 22qq13.1 (rs138747, *TOM1*) with a P-value = 0.001. This locus was originally reported in Swaminathan *et al.,* 2015) with a borderline significant *P*-value of 6E-08. Moreover, the TWAS results from in Article III suggest that *KIF3C* could be the gene causing the association at 2p23.2, instead of the candidate genes *DTNB* and *DNMT3A*.

This work provides further inside in the functional mechanisms underlying the in the germline genetic landscape of multiple myeloma. We observed that several MM risk genes are known to be involved in telomere maintenance: *TERC* is the RNA

component of the telomerase enzyme, *POT1* and *STN1* protect telomeres and *RFWD3* had been associated with telomere length in GWAS. We performed a Mendelian randomization analysis and reported association between leukocyte telomere length (LTL) and MM risk which showed significant results for causality from LTL to MM.

Article IV

In Article IV, we explored a new approach to identify causal transcription factors underlying altered gene regulation at GWAS loci. We developed a tool that can be used regularly in the lab, to point at the key transcription factors regulating expression of GWAS-appointed genes.

As a proof of concept, we studied *ELL2*. The *ELL2* risk allele for multiple myeloma is associated with lower expression of the gene. Our working hypothesis in this project was that differences in transcription factor binding depending on the ELL2 genotype could be responsible of this effect.

We designed a pooled CRIPSR KO library to test the effect of knock out all the transcription factors in parallel and detect which sgRNAs caused loss of expression. We studied the effects of this knock out in two cell lines that were homozygous the major or minor alleles, sorted the cells that had lost *ELL2* expression to identify the responsible transcription factors. For further filtering, and to overcome some limitations of our over simplified model, we developed a computational method to integrate malignant plasma cells expression patterns from 1247 patients and developed a multi-variable regression model, where *ELL2* expression is predicted as a linear combination of candidate transcription factor expression and *ELL2* genotype.

Integrating our CRISPR screening data with computational interaction modelling, we identified *ARID3A* and *MAFF* as putative causal transcription factors at the *ELL2* MM risk locus.

Conclusions

Article I

By integrating MPRA with additional genomic techniques we provided functional evidence that identify putative causal variants at six multiple myeloma risk loci.

Article II

We have identified a new genetic association for MM at 13q13 *SOHLH2*, and we provided evidence of its genetic regulatory effects in plasma cells.

Article III

In this new meta-GWAS, we discovered 9 new Multiple myeloma risk loci: 5q35.2, *CPEB4*, 6p22.2 *BTN3A2*, 9q21.33 *DAPK1*, 10q24.33 *STN1*, 10q25.2 *MXI1*, 19p13.3 *NFIC*, 21q11.2, *SAMSN1* and a rare variant at 13q13.1 *BRCA2*.

Article IV

We have developed a screening method to identify allele-specific transcription factor interaction by combining a CRISPR/Cas9 screening and a computational epistasis analysis based on gene expression data. We applied this approach to dissect the ELL2 MM risk allele, identifying *ARID3A* and *MAFF* as the transcription factors driving the effects on *ELL2* expression.

Table 2: Table of Multiple Myeloma GWAS associated loci, Combining the previously known Multiple Myeloma risk loci and the advances occurred during the period that this thesis has been carried out,

Locus	candidate gene	*lead* rsID	RA/OA	OR	RAF	P-value	Discovery study
2p23.3	*DTNB*	rs7577599	T/C	1.27	0,48	8.00E-28	(Broderick *et al.,* 2011)
2q31.1	*SP3*	rs16862227	G/T	1.12	0.77	3.89E-09	(Went *et al.,* 2018)
3p22.1	*ULK4*	rs9856633	A/G	1.23	0.19	1.65E-20	(Broderick *et al.,* 2011)
3q26.2	*LRRC34, TERC*	rs7621631	C/A	1.18	0.76	8.58E-18	(Chubb *et al.,* 2013)
5q15	*ELL2*	rs11744881	A/T	1.15	0.72	6.08E-13	(Swaminathan *et al.,* 2015)
5q23.2	*CEP120*	rs2162826	C/A	1.12	0.45	6.58E-09	(Went *et al.,* 2018)
5q35.2	*CPEB4*	rs6864880	C/T	1.11	0.33	1.85E-08	Article IV
6p21.33	*HLA region*	rs3132535	A/G	1.19	0.26	1.52E-22	(Chubb *et al.,* 2013)
6p22.2	*BTN3A2*	rs34565965	T/A	1.13	0.22	1.65E-09	Article IV
6p22.3	*JARID2*	rs74875586	A/G	1.45	0.02	7.12E-08	(Mitchell *et al.,* 2016)
6p25.3	*IRF4*	rs1050976	T/C	1.10	0.49	2.33E-08	Article IV
6q21	*ATG5*	rs9386514	C/T	1.18	0.19	1.51E-16	(Mitchell *et al.,* 2016)
7p15.3	*CDCA7L*	rs75341503	A/C	1.25	0.66	1.47E-35	(Broderick *et al.,* 2011)
7q22.3	*CCDC71L*	rs11762574	A/G	1.14	0.74	8.18E-13	(Went *et al.,* 2018)
7q31.33	*POT1*	rs10954065	C/A	1.10	0.73	1.22E-07	(Went *et al.,* 2018)
7q36.1	*SMARCD3*	rs10233479	T/C	1.25	0.09	4.93E-19	(Mitchell *et al.,* 2016)
8q24.21	*CCAT1, MYC*	rs1948915	C/T	1.15	0.33	1.54E-15	(Mitchell *et al.,* 2016)
9p21.3	*CDKN2A*	rs3731222	T/C	1.23	0.64	2.84E-16	(Mitchell *et al.,* 2016)
9q21.33	*DAPK1*	rs10746812	C/T	1.12	0.35	5.13E-11	Article IV
10p12.1	*WAC*	rs2993984	T/A	1.12	0.74	7.32E-10	(Mitchell *et al.,* 2016)
10q24.33	*STN1*	rs11813268	T/C	1.15	0.11	1.30E-09	Article IV
10q25.2	*MXI1*	rs3737315	T/G	1.11	0.36	7.62E-10	Article IV
13q13.1	*BRCA2*	rs11571833	T/A	1.57	0.0048	2.95E-08	Article IV
13q13.3	*SOHLH2*	rs75712673	G/T	1.29	0.035	3.26E-10	Article II
16p11.2	*FBRS, SRCAP*	rs8058928	G/T	1.14	0.28	3.82E-12	(Went *et al.,* 2018)
16q23.1	*RFWD3*	rs8050262	T/C	1.11	0.61	7.83E-10	(Mitchell *et al.,* 2016)
17p11.2	*TNFRSF13B*	rs34562254	A/G	1.30	0.10	2.82E-23	(Chubb *et al.,* 2013)
19p13.11	*KLF2*	rs4808046	G/A	1.13	0.23	4.62E-10	(Went *et al.,* 2018)
19p13.3	*NFIC*	rs11085015	T/G	1.19	0.20	1.73E-13	Article IV
20q13.13	*PREX1*	rs6090899	G/A	1.22	0.09	3.45E-13	(Mitchell *et al.,* 2016)
21q11.2	*SAMSN1*	rs2822736	C/G	1.11	0.74	2.795E-09	Article IV
22q13.1	*CBX7*	rs5995688	G/A	1.21	0.44	2.71E-31	(Chubb *et al.,* 2013)

Future perspectives

Following our work, there are now 35 known MM risk variants accounted for by 32 loci. For most of these, the causal variant and their mechanisms of action have not yet been established.

Most of the research is focused on characterizing transcriptional *cis*-regulatory functions in plasma cells. However, DNA methylation studies in MM PC, and eQTL analysis and single cell RNA-seq on non-hematopoietic cell types of the bone marrow could be explored to explain some of the still unknown mechanisms.

Better follow-up of MGUS cases could contribute to early diagnosis, which will hopefully happen in parallel with improved therapeutic approaches that reduce relapse.

MGUS and sMM are generally untreated to delay the development of resistance to first line treatment. Studies need to be performed to predict the MGUS cases with higher risk of evolving to MM. The clinical use of PRS could potentially be implemented in this setting.

Having established the most relevantly dysregulated biological processes in MM predisposition, with a number of putative causal genes interfering in each process, it could be interesting to investigate the epistatic effects of mutations affecting the same pathways. The pleiotropic effect of variants affecting telomeric function and MM predisposition needs to be further studied. For example, telomere length has been proposed as a prognostic and survival marker for MM patients (Hyatt *et al.*, 2017; Rangel-Pozzo *et al.*, 2021).

Acknowledgements

The acknowledgments of this thesis won't be brief. After all, this is the most read part of the thesis and I do have a lot to be thankful for. And I was never known for being *concise*. I started writing them long ago because these four and a half years I have been constantly thankful to a lot of people and I now finally have an opportunity to tell you. These words have been written with a smile on my face and a casual tear in some occasions, and they were written in several moments of my PhD, as it gave me peace to be thankful in stressful times. So here I go:

In first place, to my PI and supervisor **Björn**; he never came to my office asking for results but asked if I was calm and happy or looked at me with very opened eyes so I'd tell him how I was. I take with me a lot of valuable lessons and ways of working methodologically that I have learned from you. I am immensely grateful for these years in the lab.

To my co-supervisor **Aitzkoa**, thank you for believing in me and supporting me. I admire you as a scientist and activist and I treasure all the conversations we have had in your office trying to fix the world. You make both the academic world and the real world are better.

I have been incredibly lucky to work in a lab with a group of people that are very committed to science and that I blindly trust. Integrity. My absolute favourite are the endless and extremely interesting discussions about all kinds of topics during lunch. I don't have the words to explain how much I treasure those moments and how much I have learned from all of you. I can't imagine any context where I could have learned so much and from all these different cultures and ways of understanding the world.

So thank you to all my lab mates. **Cat,** you could be a la bon your own. You are so resourceful and energized and I know I can count on you when there is any trouble in the lab. Always ready for a meme, or a tweet, or to add some sense of humour to the worst of the situations. **Nerea**, thank you for all the office talk I learned a lot from you. And thank you for all the cakes! To **Maroulitsa**, with whom I share a "scientific taste", this last year I have looked at you as a third mentor, and you have always been there when it was more necessary, reading a manuscript or running a script. To **Jenny**, the PhD student that could be a PI, thanks for all the times you've

helped me with different projects, I always try to absorb as much as possible from your knowledge. I admire your professional, easy-going and kind energy. **Abhishek**, the kindness in person, thank you for the peace you bring. And for checking all those eQTLs and caQTLs and teaching me how to do it. It has been a pleasure to learn from you. To **Ludvig**, I have great memories of the time we discussed the most random topics in the big glass PhD office. Thanks for always explaining and teaching while helping with any project. To **Zain,** for all the interesting points of view in those lunch discussions (as I have said, one of my favourite moments of the PhD) for keeping us all updated with the latest scientific advances and interesting Articles. To **Daniela** for listening to all my jabbering, joking, being honest and caring about me. I wish the brightest future for you and I have learned a lot from you too. To **Antton** for always having a moment to help, for your calmness, your genuine interest in science and for starting the shared soy milk club. To **Ram**, I have learned soo much working with you. To **Yuwen** for brining joy to the office, I wish you the best in your PhD. And to the past members of the lab, To **Britt-Marie** for asking the important questions in the meetings, and for the nice discussion about genetics. To **Mina**, I have always admired your bravery and the path you took to get to our lab. And to **Ellinor** that showed me around when I first arrived at the lab.

To **Sara** my Nutella buddy, for all the homemade sweets you have left on my table and for all the time we have shared.

For my time at **deCODE**, thanks to **Thórunn**, for taking care of me and trusting me while I was in Rejkyavík. You are a role model for me. And to **Guðmar** for your immense patience and the conversations about genetic statistics during my internship.

To all the **patients** represented in the **Lund MM biobank**, and to **Anna Collin** in particular for maintaining this valuable biobank that my PhD has nourished so much from.
To the "***37 years old woman"*** that L363 cells were extracted from and the "***65 years-old woman"*** KMS20 cell line donor, I have thought a lot about how terrible but different it had to be for you as patients, one with such fast growing cells and the other one with cells that I struggled to expand.

I en aquests moments no puc oblidar-me d'on vinc. Gràcies a la **Sara** i l'**Orland**, els mes pares en el món de la recerca. A la **Sandra** que em va ensenyar a fer PCRs (amb temperatures d'annealing sempre parells) la **Gemma** un pilar fonamental del lab i la primera PhD de referència que vaig tenir, treballadora incansable. Aquests

últims mesos m'ha salvat la vida tenir un calaix amb menjar a l'oficina. A l'**Ale** (a.k.a. **el Malagueño**) por todos estos años de amistad, por tu sentido del humor y estar siempre allí, para mirar un splicing o reírnos de la vida. Empezamos esta aventura de juntos en el mòdul blau y me hace inmensamente haber seguido recorriendo juntos desde distintos labs y países.

A la **Vicky Reyes** i la **Meritxell Mollà** perquè em van fer sentir valorada en un moment i àmbit en el que em veia molt petita. I per ser, juntament amb moltes de les professionals que enumero en aquests agraïments, referents de dones amb carreres científiques exitoses.

Gracias a **Benja** por abrirme las puertas a la genética ese verano de prácticas. Estic també molt agraïda al **José García Valero** per creure en mi quan ni jo hi creia i animar-me a apostar fort. Aquelles converses em van donar molta força per arribar fins aquí.

To **Andrew** for his contribution to improving this thesis (and all the amazing Christmas meals) y **Patri** por darme siempre buenos consejos,

I am truly thankful to the international scientific community, for all the previous knowledge that allowed me to do write this thesis. To **Lund University** for providing excellent conditions during my PhD and to **MDR** that has fought during many years for us to have those good. To the **Marie Skłodowska-Curie Actions** (**MSCA**) and the **European Commission** because they changed my life by giving me the opportunity of performing this PhD. To **Jana** and **Sara**, because it is only thanks to the CanFaster doctoral program that I came to Lund and I am here today.

To all the hard moments that have taught me important lessons. We could have skipped the pandemic to be honest, but we also learned from that. And to **sense of humour**! In which I have found shelter in the best and worst moments.

To all the people that has been by my side outside of the lab. I feel enormously lucky to have you in my life. My big little bubble of PhD students, you really became my family during these years. A PhD has lots of good moments but tonnes of challenging ones too, being through them all together has created the deepest bond that I know will keep us close for the rest of our lives and honestly I can't wait to see you all grow and go in different directions. **Kat,** I don't even know where to start, so many moments and conversations are crossing your mind right now. And it is exactly because of how you get me that you've become so important. And for those million details, coffees, texts, screenshots. **Marta** podría pasarme horas

contigo y no dejariamos de hablar ni un minuto. Por favor no perdamos esto nunca. Por pensar en los detalles, por tu integridad personal y científica. **Myriam** no em cansaré de decirlo nunca y es que estoy aquí gracias a ti. Hiciste que mi vida diera una voltereta y has sido una parte fundamental de mi vida aquí, presente siempre que se te necesita, con un sobre de sopa o un *atpc*. A tu **Ana** que tens el poder d'irradiar alegria i positivitat. Admiro molt la teva fortalesa personal i rigor científic. És una pena això de la "xocolata" blanca jajajaj pero si a tu et fa feliç, a mi també. **Maria** he perdido la cuenta de cuantas veces me has salvado la vida en el lab dejándonos unas placas, medio, cell counting chambers. Gracias por la dulzura y magia que pones en todo lo que haces. **Martina** thank you for your kindness, I have learned so much from you and your way of seeing life. For those nice evenings in Lund, Malmö and Vienna and for all the things that are to come.

Mo you have been an essential support, always there in the best and worst moments understanding me as very few people do. **Alex** thanks for all the scientific discussions, excellent food, and delicious cocktails. And thanks for all the chocolate balls you've dropped by my office that have fuelled me during the good and bad moments of this PhD. **Lavanya** thank you for taking care of me, for the way you show you trust me, for the incredible trip to India and for all the interesting conversations on all the companies we've wanted to start**.** To **Oscar** for the lovely days at the summer house, simple and deep conversations and for the Populärvetenskaplig sammanfattning. **Sid** for your sense of humour when speaking about very serious things, and for your kindness. **Jordi** has sigut un molt bon fitxatge, gràcies per tots els moments compartits, per donar un cop demà sempre que cal. **Martino** for those magic *spaghetti aglio e olio*, how we easily had some deep conversations and for the good times partying. **Matilde**, thank you for all the good times and conversations and for the peaceful energy you bring to the group. **Alba** por haber estado, por las cafés y birras al estilo "comedia dramática" no sé si sabes la de vida que me han dado estas mini desconexiones en nuestros afterworks. **David** gràcies per totes les bones estones que vam passar a Lilla Södergatan, per totes les festes i confidències. And thank you **Claudio** for all the nice moments and conversations, food and Gigi D'Agostino dancing.

And thank you to all the people that I have shared nice moments with here in Lund, the lovely bubble made of mostly PhD students in which I have been living this years: **Roberta, Eliška, Edo, Julie, Lucho, Marcus, Antonio, Laura TG, Isa, Jessica, Albert, Andreas, Marja, Isaak, Fábio, Nic, Scott, TiagoS, Joana, Maria Neves, Roger, James, Aastha, Sergio, Kelin** and **all the other CanFasters**.

Viure fora és ultra enriquidor i és el que sempre havia volgut, però també comporta perdre's molts moments i tenir relacions a distància amb molta gent. Espero haver sabut cuidar-les prou, en especial les de tota aquesta gent que ve ara:

A la **Garazi** per les hores i hores d'audios en les que hem seguit estant tan aprop com quan compartiem pis, tot i estar a punta i punta d'Europa. Per aquesta incondicionalitat i haver-hi estat absolutatment sempre. I per tot el que ve :) **Monica y Ainhoa** nos hemos visto crecer, y me conocéis como poca gente. Me hace muy feliz que sigáis en mi vida y segur en las vuestras. Yo siempre con el equipo enfermeras! A la **Judit, Berta** i **Nerea** per seguir aquí i ser part important de l'ancla que em recorda on és casa. A la **Raquel** per ser-hi quan més m'ha calgut i sobretot per haver sabut que em calies. Pel cap d'any a Cadiz que em va salvar, per tants recors d'adolescència i per tanta bondat. **Leiva,** tu també hi has estat sempre, i t'he tingut sempre molt present. **Mercè** i **Paula**, els orígens, no sabeu com d'important, energitzant va ser i com de feliç em va fer aquell cap de setmana a Copenhagen. Per seguir veient-nos avançar a la vida i pel món.

I finalment, el punt de suport incondicional des de que tinc memòria. I absolutament incondicional; des d'assajar portar aigua a la boca per fer boles de sorra fina al pati quan tenia 4 anys a enviar cyber-abraçades mentre escribía i pensar en el jamón per després de la defensa. Atacs de riure com els que tinc a casa (o a hotels de Carretera, aeroports francesos, un aparcament de mirador qualsevol de les Açores) no els tinc amb ningú més, és la millor sensación del món. A hores d'ara sé que un dels privilegis més grans dels que gaudeixo ha estat créixer amb vosaltres i aprenre-ho tot de vosaltres; gràcies **Mama** i **Papa** per haver-me donat les eines i els recursos que m'han permès arribar aquí i haver-me ensenyat a no posar-me litmits. Per "la llei del minim esforç", "no facis el que no t'agradaria que et féssin" "m'ho passaré be i el monitor de piscina serà simpàtic", per organitzar la vida fent llistes (absolutament clau durant tota la tesi). A la **Marta** per ser la millor germana, un suport també incondicional i constant, a través de les petites coses del dia a dia tot i estar tan lluny. Super diferent i super sembalnt a mi, amb qui puc comptar sempre i puc parlar del més profund i el més superflu. Siusplau compta amb mi per absoltament el que calgui.

De la **Iaia Maria** he heredat l'ordre i l'organització (ojalà hagués heredat també una mica de calma!), gràcies per cuidar-nos, per totes les escudelles, el fricandó. I a la

resta de família que se que em té en bona consideració i estima que és absolutament recíproca: **Tieta Sílvia, Tiet Miquel, Gerard, Mireia, Tieta Maria Teresa, Tiet Ramón, Astrid & family, Marc & family i Tieta Montse.** No heu col·laborat directament en els experiments de la tesi però en qualsevol moment d'aquests 4 anys he sabut que us tenia i he notat el vostre support, com el de la **Mercedes, Juan Manuel, Alex, Paula, David,** muchas gracias por querer scompartir congmigo un momento tan especial. A la **Yaya Alicia**, hace dos generaciones empezaste a luchar para que yo haya podido defender la tesis. Lo que daría por que tu pudieras haber tenido estas oportunidades.

References

Agnarelli, A., Chevassut, T., & Mancini, E. J. (2018). IRF4 in multiple myeloma—Biology, disease and therapeutic target. *Leukemia Research*, *72*(May), 52–58. https://doi.org/10.1016/j.leukres.2018.07.025

Ailawadhi, S., Parikh, K., Abouzaid, S., Zhou, Z., Tang, W., Clancy, Z., Cheung, C., Zhou, Z. Y., & Xie, J. (2019). Racial disparities in treatment patterns and outcomes among patients with multiple myeloma: A SEER-Medicare analysis. *Blood Advances*, *3*(20), 2986–2994. https://doi.org/10.1182/bloodadvances.2019000308

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, *16*(4), 197–212. https://doi.org/10.1038/nrg3891

Alexander, D. D., Mink, P. J., Adami, H. O., Cole, P., Mandel, J. S., Oken, M. M., & Trichopoulos, D. (2007). Multiple myeloma: A review of the epidemiologic literature. *International Journal of Cancer*, *120*(SUPPL. 14), 40–61. https://doi.org/10.1002/ijc.22718

Alexander, L., & Benninghoff, D. (1965). *Familial Multiple Myeloma*. *57*(6), 471–475.

Ali, M., Ajore, R., Wihlborg, A. K., Niroula, A., Swaminathan, B., Johnsson, E., Stephens, O. W., Morgan, G., Meissner, T., Turesson, I., Goldschmidt, H., Mellqvist, U. H., Gullberg, U., Hansson, M., Hemminki, K., Nahi, H., Waage, A., Weinhold, N., & Nilsson, B. (2018). The multiple myeloma risk allele at 5q15 lowers ELL2 expression and increases ribosomal gene expression. *Nature Communications*, *9*, 1649. https://doi.org/10.1038/s41467-018-04082-2

Altieri, A., Chen, B., Bermejo, J. L., Castro, F., & Hemminki, K. (2006). Familial risks and temporal incidence trends of multiple myeloma. *European Journal of Cancer*, *42*(11), 1661–1670. https://doi.org/10.1016/j.ejca.2005.11.033

Anzalone, A. V, Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *August*. https://doi.org/10.1038/s41586-019-1711-4

Badano, J. L., & Katsanis, N. (2006). Life without Centrioles: Cilia in the Spotlight. *Cell*, *125*(7), 1228–1230. https://doi.org/10.1016/j.cell.2006.06.013

Baris, D., Brown, L. M., Andreotti, G., & Devesa, S. S. (2013). Neoplastic diseases of the blood. *Neoplastic Diseases of the Blood*, 1–1431. https://doi.org/10.1007/978-1-4614-3764-2

Beauchamp, E. M., Leventhal, M., Bernard, E., Hoppe, E. R., Todisco, G., Creignou, M., Gallì, A., Castellano, C. A., McConkey, M., Tarun, A., Wong, W., Schenone, M., Stanclift, C., Tanenbaum, B., Malolepsza, E., Nilsson, B., Bick, A. G., Weinstock, J. S., Miller, M., … Ebert, B. L. (2021). ZBTB33 Is Mutated in Clonal Hematopoiesis and Myelodysplastic Syndromes and Impacts RNA Splicing. *Blood Cancer Discovery*, *2*(5), 500–517. https://doi.org/10.1158/2643-3230.bcd-20-0224

Belmont, J. W., Boudreau, A., Leal, S. M., Hardenbol, P., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., … Stewart, J. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi.org/10.1038/nature04226

Borys, F., Joachimiak, E., Krawczyk, H., & Fabczak, H. (2020). Intrinsic and Extrinsic Factors Affecting Microtubule Dynamics in Normal and Cancer Cells. *Molecules*, *25*(16), 1–36. https://doi.org/10.3390/molecules25163705

Bourillot, P. Y., & Savatier, P. (2010). Krüppel-like transcription factors and control of pluripotency. *BMC Biology*, *8*(1), 8–10. https://doi.org/10.1186/1741-7007-8-125

Broderick, P., Chubb, D., Johnson, D. C., Weinhold, N., Försti, A., Lloyd, A., Olver, B., Ma, Y. P., Dobbins, S. E., Walker, B. A., Davies, F. E., Gregory, W. A., Child, J. A., Ross, F. M., Jackson, G. H., Neben, K., Jauch, A., Hoffmann, P., Mühleisen, T. W., … Houlston, R. S. (2011). Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nature Genetics*, *44*(1), 58–61. https://doi.org/10.1038/ng.993

Broekema, R. V., Bakker, O. B., & Jonkers, I. H. (2020). A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biology*, *10*(1). https://doi.org/10.1098/rsob.190221

Burwick, N., & Sharma, S. (2019). Glucocorticoids in multiple myeloma: past, present, and future. *Annals of Hematology*, *98*(1), 19–28. https://doi.org/10.1007/s00277-018-3465-8

Caraccio, C., Krishna, S., Phillips, D. J., & Schürch, C. M. (2020). Bispecific Antibodies for Multiple Myeloma: A Review of Targets, Drugs, Clinical Trials, and Future Directions. *Frontiers in Immunology*, *11*(April), 1–25. https://doi.org/10.3389/fimmu.2020.00501

Charpentier, E., Elsholz, A., & Marchfelder, A. (2019). CRISPR-Cas: more than ten years and still full of mysteries. *RNA Biology*, *16*(4), 377–379. https://doi.org/10.1080/15476286.2019.1591659

Chen, C. H., Xiao, T., Xu, H., Jiang, P., Meyer, C. A., Li, W., Brown, M., & Liu, X. S. (2018). Improved design and analysis of CRISPR knockout screens. *Bioinformatics*, *34*(23), 4095–4101. https://doi.org/10.1093/bioinformatics/bty450

Chen, L., Hu, N., Wang, C., Zhao, H., & Gu, Y. (2018). Long non-coding RNA CCAT1 promotes multiple myeloma progression by acting as a molecular sponge of miR-181a-5p to modulate HOXA1 expression. *Cell Cycle*, *17*(3), 319–329. https://doi.org/10.1080/15384101.2017.1407893

Chen, W. M., Chen, Y. M., Jiang, S. Y., Tao, Y. P., Hong, Y. G., Yang, L., Zheng, H., & Chen, J. Q. (2021). LncRNA POT1-AS1 accelerates the progression of gastric cancer by serving as a competing endogenous RNA of microRNA-497-5p to increase PDK3 expression. *Journal of Gastrointestinal Oncology*, *13*(6), 2728–2742. https://doi.org/10.21037/jgo-21-709

Chim, C. S., Kumar, S. K., Orlowski, R. Z., Cook, G., Richardson, P. G., Gertz, M. A., Giralt, S., Mateos, M. V., Leleu, X., & Anderson, K. C. (2018). Management of relapsed and refractory multiple myeloma: Novel agents, antibodies, immunotherapies and beyond. *Leukemia*, *32*(2), 252–262. https://doi.org/10.1038/leu.2017.329

Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D. C., Försti, A., Vijayakrishnan, J., Migliorini, G., Dobbins, S. E., Holroyd, A., Hose, D., Walker, B. A., Davies, F. E., Gregory, W. A., Jackson, G. H., Irving, J. A., Pratt, G., Fegan, C., Fenton, J. A. L., … Goldschmidt, H. (2013). Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nature Genetics*, *45*(10), 1221–1225. https://doi.org/10.1038/ng.2733

Clow, P. A., Du, M., Jillette, N., Taghbalout, A., Zhu, J. J., & Cheng, A. W. (2022). CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus. *Nature Communications*, *13*(1), 1–10. https://doi.org/10.1038/s41467-022-29343-z

Conway, K. L., Kuballa, P., Khor, B., Zhang, M., Shi, H. N., Virgin, H. W., & Xavier, R. J. (2013). ATG5 regulates plasma cell differentiation. *Autophagy*, *9*(4), 528–537. https://doi.org/10.4161/auto.23484

Cowan, A. J., Allen, C., Barac, A., Basaleem, H., Bensenor, I., Curado, M. P., Foreman, K., Gupta, R., Harvey, J., Dean Hosgood, H., Jakovljevic, M., Khader, Y., Linn, S., Lad, D., Mantovani, L., Nong, V. M., Mokdad, A., Naghavi, M., Postma, M., … Fitzmaurice, C. (2018). Global burden of multiple myeloma: A systematic analysis for the global burden of disease study 2016. *JAMA Oncology*, *4*(9), 1221–1227. https://doi.org/10.1001/jamaoncol.2018.2128

Curado, M. P., Oliveira, M. M., Silva, D. R. M., & Souza, D. L. B. (2018). Epidemiology of multiple myeloma in 17 Latin American countries: an update. *Cancer Medicine*, *7*(5), 2101–2108. https://doi.org/10.1002/cam4.1347

Das, R., Dimitrova, N., Xuan, Z., Rollins, R. A. R., Haghighi, F., Edwards, J. J. R., Ju, J., Bestor, T. H., Zhang, M. Q. M., Merry, C. R., Forrest, M. E., Sabers, J. N., Beard, L., Gao, X.-H., Hatzoglou, M., Jackson, M. W., Wang, Z., Markowitz, S. D., Khalil, A. M., … Bird, A. A. (2016). An ImprovedSimple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities Sven. *Nature*, *7*(1), 705–708. https://doi.org/10.1016/j.molcel.2010.05.004

Das, S., Abecasis, G. R., & Browning, B. L. (2018). Genotype imputation from large reference panels. *Annual Review of Genomics and Human Genetics*, *19*, 73–96. https://doi.org/10.1146/annurev-genom-083117-021602

Dilworth, D., Liu, L., Stewart, A. K., Berenson, J. R., Lassam, N., & Hogg, D. (2000). Germline CDKN2A mutation implicated in predisposition to multiple myeloma. *Blood*, *95*(5), 1869–1871. https://doi.org/10.1182/blood.v95.5.1869.005k09_1869_1871

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Doudna, J. A. (2020). The promise and challenge of therapeutic genome editing. *Nature*, *578*(7794), 229–236. https://doi.org/10.1038/s41586-020-1978-5

Eagen, K. P. (2018). Principles of Chromosome Architecture Revealed by Hi-C. *Trends in Biochemical Sciences*, *43*(6), 469–478. https://doi.org/10.1016/j.tibs.2018.03.006

Elegheert, J., Behiels, E., Bishop, B., Scott, S., Woolley, R. E., Griffiths, S. C., Byrne, E. F. X., Chang, V. T., Stuart, D. I., Jones, E. Y., Siebold, C., & Aricescu, A. R. (2018). Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nature Protocols*, *13*(12), 2991–3017. https://doi.org/10.1038/s41596-018-0075-9

Elia, A. E. H., Wang, D. C., Willis, N. A., Boardman, A. P., Hajdu, I., Adeyemi, R. O., Lowry, E., Gygi, S. P., Scully, R., & Elledge, S. J. (2015). RFWD3-Dependent Ubiquitination of RPA Regulates Repair at Stalled Replication Forks. *Molecular Cell*, *60*(2), 280–293. https://doi.org/10.1016/j.molcel.2015.09.011

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, *3*(11). https://doi.org/10.1371/journal.pone.0003647

Eriksson, M., & Hållberg, B. (1992). Familial occurrence of hematologic malignancies and other diseases in multiple myeloma: a case-control study. *Cancer Causes and Control*, *3*(1), 63–67. https://doi.org/10.1007/BF00051914

Field-smith, A. (2006). *Bortezomib (Velcade ™) in the treatment of multiple myeloma*. *2*(3), 271–279.

Fonseca, R., Blood, E. A., Oken, M. M., Kyle, R. A., Dewald, G. W., Bailey, R. J., Van Wier, S. A., Henderson, K. J., Hoyer, J. D., Harrington, D., Kay, N. E., Van Ness, B., & Greipp, P. R. (2002). Myeloma and the t(11;14)(q13;q32); evidence for a biologically defined unique subset of patients. *Blood*, *99*(10), 3735–3741. https://doi.org/10.1182/blood.V99.10.3735

Frerichs, K. A., Nagy, N. A., Lindenbergh, P. L., Bosman, P., Marin Soto, J., Broekmans, M., Groen, R. W. J., Themeli, M., Nieuwenhuis, L., Stege, C., Nijhof, I. S., Mutis, T., Zweegman, S., Lokhorst, H. M., & van de Donk, N. W. C. J. (2018). CD38-targeting antibodies in multiple myeloma: mechanisms of action and clinical experience. *Expert Review of Clinical Immunology*, *14*(3), 197–206. https://doi.org/10.1080/1744666X.2018.1443809

Fujino, M. (2018). The histopathology of myeloma in the bone marrow. *Journal of Clinical and Experimental Hematopathology*, *58*(2), 61–67. https://doi.org/10.3960/jslrt.18014

Gandolfi, S., Laubach, J. P., Hideshima, T., Chauhan, D., Anderson, K. C., & Richardson, P. G. (2017). The proteasome and proteasome inhibitors in multiple myeloma. *Cancer and Metastasis Reviews*, *36*(4), 561–584. https://doi.org/10.1007/s10555-017-9707-8

Gao, X., Leone, G. W., & Wang, H. (2020). Cyclin D-CDK4/6 functions in cancer. In *Advances in Cancer Research* (1st ed., Vol. 148). Elsevier Inc. https://doi.org/10.1016/bs.acr.2020.02.002

Garg, A., Morgunskyy, M., Belagali, Y., Gupta, N., & Akku, S. P. (2016). Management of Multiple Myeloma and usage of bortezomib: Perspective from India and Ukraine. *Frontiers in Oncology*, *6*(NOV), 1–8. https://doi.org/10.3389/fonc.2016.00243

George, L. L., Deshpande, S. R., Cortese, M. J., Kendall, E. K., Chattaraj, A., Shah, Z., Zhao, J., & Anwer, F. (2021). Emerging Targets and Cellular Therapy for Relapsed Refractory Multiple Myeloma: A Systematic Review. *Clinical Lymphoma, Myeloma and Leukemia*, *21*(11), 741–751. https://doi.org/10.1016/j.clml.2021.06.003

Ghosal, S., & Banerjee, S. (2022). In silico bioinformatics analysis for identification of differentially expressed genes and therapeutic drug molecules in Glucocorticoid-resistant Multiple myeloma. *Medical Oncology*, *39*(5), 1–11. https://doi.org/10.1007/s12032-022-01651-w

Go, R. S., & Rajkumar, V. S. (2018). How i manage monoclonal gammopathy of undetermined significance. *Blood*, *131*(2), 163–173. https://doi.org/10.1182/blood-2017-09-807560

Gordon, M. G., Inoue, F., Martin, B., Schubach, M., Agarwal, V., Whalen, S., Feng, S., Zhao, J., Ashuach, T., Ziffra, R., Kreimer, A., Georgakopoulous-Soares, I., Yosef, N., Ye, C. J., Pollard, K. S., Shendure, J., Kircher, M., & Ahituv, N. (2020). lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nature Protocols*, *15*(8), 2387–2412. https://doi.org/10.1038/s41596-020-0333-5

Grandi, F. C., Modi, H., Kampman, L., & Corces, M. R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nature Protocols*, *17*(6), 1518–1552. https://doi.org/10.1038/s41596-022-00692-9

Greenberg, A. J., Rajkumar, S. V., & Vachon, C. M. (2012). Familial monoclonal gammopathy of undetermined significance and multiple myeloma: Epidemiology, risk factors, and biological characteristics. *Blood*, *119*(23), 5359–5366. https://doi.org/10.1182/blood-2011-11-387324

Greenberg, A. J., Vachon, C. M., & Rajkumar, S. V. (2012). Disparities in the prevalence, pathogenesis and progression of monoclonal gammopathy of undetermined significance and multiple myeloma between blacks and whites. *Leukemia*, *26*(4), 609–614. https://doi.org/10.1038/leu.2011.368

Greene, J. A., & Podolsky, S. H. (2012). Reform, Regulation, and Pharmaceuticals — The Kefauver– Harris Amendments at 50. *New England Journal of Medicine*, *367*(16), 1479–1481. https://doi.org/10.1056/nejmp1210763

Guillén-Boixet, J., Buzon, V., Salvatella, X., & Méndez, R. (2016). CPEB4 is regulated during cell cycle by ERK2/Cdk1-mediated phosphorylation and its assembly into liquid-like droplets. *ELife*, *5*(NOVEMBER2016), 1–26. https://doi.org/10.7554/eLife.19298

Hafner, A., & Boettiger, A. (2022). The spatial organization of transcriptional control Antonina. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-022-00526-0

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., & Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, *560*(7717), 248–252. https://doi.org/10.1038/s41586-018-0384-8

Halvarsson, B.-M., Wihlborg, A.-K., Ali, M., Lemonakis, K., Johnsson, E., Niroula, A., Cibulskis, C., Weinhold, N., Försti, A., Alici, E., Langer, C., Pfreundschuh, M., Goldschmidt, H., Mellqvist, U.-H., Turesson, I., Waage, A., Hemminki, K., Golub, T., Nahi, H., … Nilsson, B. (2017). Direct evidence for a polygenic etiology in familial multiple myeloma. *Blood Advances*, *1*(10), 619–623. https://doi.org/10.1182/bloodadvances.2016003111

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.1038/nprot.2007.249

Hemminki, K., Försti, A., Houlston, R., & Sud, A. (2021). Epidemiology, genetics and treatment of multiple myeloma and precursor diseases. *International Journal of Cancer*, *149*(12), 1980–1996. https://doi.org/10.1002/ijc.33762

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, *4*(2). https://doi.org/10.1371/journal.pgen.1000008

Hill, W. G., & Robertson, A. (1968). Linkage Disequilibrium in Finite Populations. *Theoretical and Applied Genetics 38*, *38*, 226--23 t. https://doi.org/10.1080/03071848408522227

Ho, M., Patel, A., Goh, C. Y., Moscvin, M., Zhang, L., & Bianchi, G. (2020). Changing paradigms in diagnosis and treatment of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Leukemia*, *34*(12), 3111–3125. https://doi.org/10.1038/s41375-020-01051-x

Hoffmann, F. S., Kuhn, P.-H., Laurent, S. A., Hauck, S. M., Berer, K., Wendlinger, S. A., Krumbholz, M., Khademi, M., Olsson, T., Dreyling, M., Pfister, H.-W., Alexander, T., Hiepe, F., Kümpfel, T., Crawford, H. C., Wekerle, H., Hohlfeld, R., Lichtenthaler, S. F., & Meinl, E. (2015). The Immunoregulator Soluble TACI Is Released by ADAM10 and Reflects B Cell Activation in Autoimmunity. *The Journal of Immunology*, *194*(2), 542–552. https://doi.org/10.4049/jimmunol.1402070

Holstein, S. A., & McCarthy, P. L. (2017). Immunomodulatory Drugs in Multiple Myeloma: Mechanisms of Action and Clinical Experience. *Drugs*, *77*(5), 505–520. https://doi.org/10.1007/s40265-017-0689-1

Houlston, R. S., Cheadle, J., Dobbins, S. E., Tenesa, A., Jones, A. M., Howarth, K., Spain, S. L., Broderick, P., Domingo, E., Farrington, S., Prendergast, J. G. D., Pittman, A. M., Theodoratou, E., Smith, C. G., Olver, B., Walther, A., Barnetson, R. A., Churchman, M., Jaeger, E. E. M., … Tomlinson, I. P. M. (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nature Genetics*, *42*(11), 973–977. https://doi.org/10.1038/ng.670

Huang, T. P., Newby, G. A., & Liu, D. R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nature Protocols*, *16*(2), 1089–1128. https://doi.org/10.1038/s41596-020-00450-9

Hungria, V. T. M., Maiolino, A., Martinez, G., Duarte, G. O., Bittencourt, R., Peters, L., Colleoni, G., Oliveira, L. C. O., Crusoé, E., Coelho, É. O. D. M., Pasquini, R., Magalhães, S. M. M., Nunes, R., Neto, J. V. P., Faria, R. M. O., Souza, M., Hamerschlak, N., Flantl, D., Navarro, J. R., … Durie, B. G. M. (2017). Observational study of multiple myeloma in Latin America. *Annals of Hematology*, *96*(1), 65–72. https://doi.org/10.1007/s00277-016-2866-9

Hunter, Z. R., Yang, G., Xu, L., Liu, X., Castillo, J. J., & Treon, S. P. (2017). Genomics, signaling, and treatment of waldenström macroglobulinemia. *Journal of Clinical Oncology*, *35*(9), 994–1001. https://doi.org/10.1200/JCO.2016.71.0814

Hyatt, S., Jones, R. E., Heppel, N. H., Grimstead, J. W., Fegan, C., Jackson, G. H., Hills, R., Allan, J. M., Pratt, G., Pepper, C., & Baird, D. M. (2017). Telomere length is a critical determinant for survival in multiple myeloma. *British Journal of Haematology*, *178*(1), 94–98. https://doi.org/10.1111/bjh.14643

Inano, S., Sato, K., Katsuki, Y., Kobayashi, W., Tanaka, H., Nakajima, K., Nakada, S., Miyoshi, H., Knies, K., Takaori-Kondo, A., Schindler, D., Ishiai, M., Kurumizaka, H., & Takata, M. (2017). RFWD3-Mediated Ubiquitination Promotes Timely Removal of Both RPA and RAD51 from DNA Damage Sites to Facilitate Homologous Recombination. *Molecular Cell*, *66*(5), 622-634.e8. https://doi.org/10.1016/j.molcel.2017.04.022

Janz, S., Zhan, F., Sun, F., Cheng, Y., Pisano, M., Yang, Y., Goldschmidt, H., & Hari, P. (2019). Germline risk contribution to genomic instability in multiple myeloma. *Frontiers in Genetics*, *10*(MAY), 1–20. https://doi.org/10.3389/fgene.2019.00424

Ji, Q. K., Ma, J. W., Liu, R. H., Li, X. S., Shen, F. Z., Huang, L. Y., Hui, L., Ma, Y. J., & Jin, B. Z. (2019). CDCA7L promotes glioma proliferation by targeting CCND1 and predicts an unfavorable prognosis. *Molecular Medicine Reports*, *20*(2), 1149–1156. https://doi.org/10.3892/mmr.2019.10349

Jia, L., Landan, G., Pomerantz, M., Jaschek, R., Herman, P., Reich, D., Yan, C., Khalid, O., Kantoff, P., Oh, W., Manak, J. R., Berman, B. P., Henderson, B. E., Frenkel, B., Haiman, C. A., Freedman, M., Tanay, A., & Coetzee, G. A. (2009). Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genetics*, *5*(8), 1–10. https://doi.org/10.1371/journal.pgen.1000597

Joachim, J., Wirth, M., McKnight, N. C., & Tooze, S. A. (2012). Coiling up with SCOC and WAC. *Autophagy*, *8*(9), 1397–1400. https://doi.org/10.4161/auto.21043

Johann Friedrich Blumenbach (1753–1840). (1940). *Nature*, *3663*, 3663. https://doi.org/https://doi.org/10.1038/145063a0

Jones, A. M., Beggs, A. D., Carvajal-Carmona, L., Farrington, S., Tenesa, A., Walker, M., Howarth, K., Ballereau, S., Hodgson, S. V., Zauber, A., Bertagnolli, M., Midgley, R., Campbell, H., Kerr, D., Dunlop, M. G., & Tomlinson, I. P. M. (2012). TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. *Gut*, *61*(2), 248–254. https://doi.org/10.1136/gut.2011.239772

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M. T., Hjorleifsson, K. E., Helgason, E. A., Helgason, H., Gylfason, A., & Jonasdottir, A. (2017). Whole genome characterization of sequence diversity of 15,220 Icelanders. *Scientific Da*, *4*, 170115. https://doi.org/10.1038/sdata.2017.115

Jonsson, S., Sveinbjornsson, G., De Lapuente Portilla, A. L., Swaminathan, B., Plomp, R., Dekkers, G., Ajore, R., Ali, M., Bentlage, A. E. H., Elmér, E., Eyjolfsson, G. I., Gudjonsson, S. A., Gullberg, U., Gylfason, A., Halldorsson, B. V., Hansson, M., Holm, H., Johansson, Å., Johnsson, E., … Stefansson, K. (2017). Identification of sequence variants influencing immunoglobulin levels. *Nature Genetics*, *49*(8), 1182–1191. https://doi.org/10.1038/ng.3897

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Jourdan, M., Caraux, A., Caron, G., Robert, N., Fiol, G., Rème, T., Bolloré, K., Vendrell, J.-P., Le Gallou, S., Mourcin, F., De Vos, J., Kassambara, A., Duperray, C., Hose, D., Fest, T., Tarte, K., & Klein, B. (2011). Characterization of a Transitional Preplasmablast Population in the Process of Human B Cell to Plasma Cell Differentiation. *The Journal of Immunology*, *187*(8), 3931–3941. https://doi.org/10.4049/jimmunol.1101230

Judson, I. R., Wiltshaw, E., & Newland, A. C. (1985). Multiple myeloma in a pair of monozygotic twins: the first reported case. *British Journal of Haematology*, *60*(3), 551–554. https://doi.org/10.1111/j.1365-2141.1985.tb07452.x

Jung, C. H., Ro, S. H., Cao, J., Otto, N. M., & Kim, D. H. (2010). MTOR regulation of autophagy. *FEBS Letters*, *584*(7), 1287–1295. https://doi.org/10.1016/j.febslet.2010.01.017

Kassambara, A., Jourdan, M., Bruyer, A., Robert, N., Pantesco, V., Elemento, O., Klein, B., & Moreaux, J. (2017). Global miRNA expression analysis identifies novel key regulators of plasma cell differentiation and malignant plasma cell. *Nucleic Acids Research*, *45*(10), 5639–5652. https://doi.org/10.1093/nar/gkx327

Kelleher, C., Kurth, I., & Lingner, J. (2005). Human Protection of Telomeres 1 (POT1) Is a Negative Regulator of Telomerase Activity In Vitro. *Molecular and Cellular Biology*, *25*(2), 808–818. https://doi.org/10.1128/mcb.25.2.808-818.2005

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kinkel, S. A., Galeev, R., Flensburg, C., Keniry, A., Breslin, K., Gilan, O., Lee, S., Liu, J., Chen, K., Gearing, L. J., Moore, D. L., Alexander, W. S., Dawson, M., Majewski, I. J., Oshlack, A., Larsson, J., & Blewitt, M. E. (2015). Jarid2 regulates hematopoietic stem cell function by acting with polycomb repressive complex 2. *Blood*, *125*(12), 1890–1900. https://doi.org/10.1182/blood-2014-10-603969

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219–225. https://doi.org/10.1038/s41586-019-1323-z

Knies, K., Inano, S., Ramírez, M. J., Ishiai, M., Surrallés, J., Takata, M., & Schindler, D. (2017). Biallelic mutations in the ubiquitin ligase RFWD3 cause Fanconi anemia. *Journal of Clinical Investigation*, *127*(8), 3013–3027. https://doi.org/10.1172/JCI92069

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H., & Stefansson, K. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genetics*, *40*(9), 1068–1075. https://doi.org/10.1038/ng.216

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 Sigurdur. *Molecular and Cellular Biochemistry*, *125*(9), 2147–2150. https://doi.org/10.1002/ijc.24514.Patterns

Kryukov, G. V., Wilson, F. H., Ruth, J. R., Paulk, J., Tsherniak, A., Marlow, S. E., Vazquez, F., Weir, B. A., Fitzgerald, M. E., Tanaka, M., Bielski, C. M., Scott, J. M., Dennis, C., Cowley, G. S., Boehm, J. S., Root, D. E., Golub, T. R., Clish, C. B., Bradner, J. E., … Garraway, L. A. (2016). MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science*, *351*(6278), 1214–1218. https://doi.org/10.1126/science.aad5214

Kumar, S. K., & Rajkumar, S. V. (2018). The multiple myelomas - Current concepts in cytogenetic classification and therapy. *Nature Reviews Clinical Oncology*, *15*(7), 409–421. https://doi.org/10.1038/s41571-018-0018-y

Kumar, S., Paiva, B., Anderson, K. C., Durie, B., Landgren, O., Moreau, P., Munshi, N., Lonial, S., Bladé, J., Mateos, M. V., Dimopoulos, M., Kastritis, E., Boccadoro, M., Orlowski, R., Goldschmidt, H., Spencer, A., Hou, J., Chng, W. J., Usmani, S. Z., … Avet-Loiseau, H. (2016). International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *The Lancet Oncology*, *17*(8), e328–e346. https://doi.org/10.1016/S1470-2045(16)30206-6

Kyle, R. A., Larson, D. R., Therneau, T. M., Dispenzieri, A., Melton, L. J., Benson, J. T., Kumar, S., & Rajkumar, S. V. (2014). Clinical course of light-chain smouldering multiple myeloma (idiopathic Bence Jones proteinuria): A retrospective cohort study. *The Lancet Haematology*, *1*(1), e28–e36. https://doi.org/10.1016/S2352-3026(14)70001-8

Kyle, R. A., & Rajkumar, S. V. (2007). Monoclonal gammopathy of undetermined significance and smouldering multiple myeloma: Emphasis on risk factors for progression. *British Journal of Haematology*, *139*(5), 730–743. https://doi.org/10.1111/j.1365-2141.2007.06873.x

La Russa, M. F., & Qi, L. S. (2015). The New State of the Art: Cas9 for Gene Activation and Repression. *Molecular and Cellular Biology*, *35*(22), 3800–3809. https://doi.org/10.1128/mcb.00512-15

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Lancman, G., Sastow, D. L., Cho, H. J., Jagannath, S., Madduri, D., Parekh, S. S., Richard, S., Richter, J., Sanchez, L., & Chari, A. (2021). Bispecific Antibodies in Multiple Myeloma: Present and Future. *Blood Cancer Discovery*, *2*(5), 423–433. https://doi.org/10.1158/2643-3230.bcd-21-0028

Landgren, O., Graubard, B. I., Kumar, S., Kyle, R. A., Katzmann, J. A., Murata, K., Costello, R., Dispenzieri, A., Caporaso, N., Mailankody, S., Korde, N., Hultcrantz, M., Therneau, T. M., Larson, D. R., Cerhan, J. R., & Rajkumar, S. V. (2017). Prevalence of myeloma precursor state monoclonal gammopathy of undetermined significance in 12372 individuals 10-49 years old: a population-based study from the National Health and Nutrition Examination Survey. *Blood Cancer Journal*, *7*(10), e618. https://doi.org/10.1038/bcj.2017.97

Landgren, O., Kristinsson, S. Y., Goldin, L. R., Caporaso, N. E., Blimark, C., Mellqvist, U. H., Wahlin, A., Bjorkholm, M., & Turesson, I. (2009). Risk of plasma cell and lymphoproliferative disorders among 14 621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. *Blood*, *114*(4), 791–795. https://doi.org/10.1182/blood-2008-12-191676

Landgren, O., Linet, M. S., McMaster, M. L., Gridley, G., Hemminki, K., & Goldin, L. R. (2006). Familial characteristics of autoimmune and hematologic disorders in 8,406 multiple myeloma patients: A population-based case-control study. *International Journal of Cancer*, *118*(12), 3095–3098. https://doi.org/10.1002/ijc.21745

Landgren, O., & Weiss, B. M. (2009). Patterns of monoclonal gammopathy of undetermined significance and multiple myeloma in various ethnic/racial groups: Support for genetic factors in pathogenesis. *Leukemia*, *23*(10), 1691–1697. https://doi.org/10.1038/leu.2009.134

Landi, M. T., Bishop, D. T., MacGregor, S., Machiela, M. J., Stratigos, A. J., Ghiorzo, P., Brossard, M., Calista, D., Choi, J., Fargnoli, M. C., Zhang, T., Rodolfo, M., Trower, A. J., Menin, C., Martinez, J., Hadjisavvas, A., Song, L., Stefanaki, I., Scolyer, R., … Law, M. H. (2020). Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nature Genetics*, *52*(5), 494–504. https://doi.org/10.1038/s41588-020-0611-8

Larsson, S. C., & Wolk, A. (2007). Body mass index and risk of multiple myeloma: A meta-analysis. *International Journal of Cancer*, *121*(11), 2512–2516. https://doi.org/10.1002/ijc.22968

Levy, D., Neuhausen, S. L., Hunt, S. C., Kimura, M., Hwang, S. J., Chen, W., Bis, J. C., Fitzpatrick, A. L., Smith, E., Johnson, A. D., Gardner, J. P., Srinivasan, S. R., Schork, N., Rotter, J. I., Herbig, U., Psaty, B. M., Sastrasinh, M., Murray, S. S., Vasan, R. S., … Aviv, A. (2010). Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9293–9298. https://doi.org/10.1073/pnas.0911494107

Lewis, A. C. F., & Green, R. C. (2021). Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues. *Genome Medicine*, *13*(1), 1–10. https://doi.org/10.1186/s13073-021-00829-7

Li, C., Stoma, S., Lotta, L. A., Warner, S., Albrecht, E., Allione, A., Arp, P. P., Broer, L., Buxton, J. L., Da Silva Couto Alves, A., Deelen, J., Fedko, I. O., Gordon, S. D., Jiang, T., Karlsson, R., Kerrison, N., Loe, T. K., Mangino, M., Milaneschi, Y., … Codd, V. (2020). Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. *American Journal of Human Genetics*, *106*(3), 389–404. https://doi.org/10.1016/j.ajhg.2020.02.006

Li, N., Johnson, D. C., Weinhold, N., Studd, J. B., Orlando, G., Mirabella, F., Mitchell, J. S., Meissner, T., Kaiser, M., Goldschmidt, H., Hemminki, K., Morgan, G. J., & Houlston, R. S. (2016). Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. *Nature Communications*, *7*, 1–9. https://doi.org/10.1038/ncomms13656

Liao, M., Ye, F., Zhang, B., Huang, L., Xiao, Q., Qin, M., Mo, L., Tan, A., Gao, Y., Lu, Z., Wu, C., Zhang, Y., Zhang, H., Qin, X., Hu, Y., Yang, X., & Mo, Z. (2012). Genome-wide association study identifies common variants at TNFRSF13B associated with IgG level in a healthy Chinese male population. *Genes and Immunity*, *13*(6), 509–513. https://doi.org/10.1038/gene.2012.26

Lickert, H., Takeuchi, J. K., Von Both, I., Walls, J. R., McAuliffe, F., Adamson, S. L., Henkelman, R. M., Wrana, J. L., Rossant, J., & Bruneau, B. G. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*, *432*(7013), 107–112. https://doi.org/10.1038/nature03071

Lieberman-Aiden, E., Berkum, N. L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., & Mirny, L. A. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome Erez. *Science*, *326*, 289–294.

Lipchick, B. C., Fink, E. E., & Nikiforov, M. A. (2016). Oxidative stress and proteasome inhibitors in multiple myeloma. *Pharmacological Research*, *105*, 210–215. https://doi.org/10.1016/j.phrs.2016.01.029

Liu, K. A., & Dipietro Mager, N. A. (2016). Women's involvement in clinical trials: Historical perspective and future implications. *Pharmacy Practice*, *14*(1), 1–9. https://doi.org/10.18549/PharmPract.2016.01.708

Lopez de Lapuente Portilla, A., Ekdahl, L., Cafaro, C., Ali, Z., Miharada, N., Thorleifsson, G., Žemaitis, K., Lamarca Arrizabalaga, A., Thodberg, M., Pertesi, M., Dhapola, P., Bao, E. L., Niroula, A., Bali, D., Norddahl, G. L., Ugidos Damboriena, N., Sankaran, V. G., Karlsson, G., Thorsteinsdottir, U., … Nilsson, B. (2022). Genome-wide association study on 13,167 individuals identifies regulators of blood CD34+ cell levels. *Blood*. https://doi.org/10.1182/blood.2021013220

Ludwig, H., Novis Durie, S., Meckl, A., Hinke, A., & Durie, B. (2020). Multiple Myeloma Incidence and Mortality Around the Globe; Interrelations Between Health Access and Quality, Economic Resources, and Patient Empowerment. *The Oncologist*, *25*(9), e1406–e1413. https://doi.org/10.1634/theoncologist.2020-0141

Lührig, S., Siamishi, I., Tesmer-Wolf, M., Zechner, U., Engel, W., & Nolte, J. (2014). Lrrc34, a novel nucleolar protein, interacts with Npm1 and Ncl and has an impact on pluripotent stem cells. *Stem Cells and Development*, *23*(23), 2862–2874. https://doi.org/10.1089/scd.2013.0470

Luo, X., & Wang, H. (2021). LINC00514 upregulates CCDC71L to promote cell proliferation, migration and invasion in triple-negative breast cancer by sponging miR-6504-5p and miR-3139. *Cancer Cell International*, *21*(1), 1–14. https://doi.org/10.1186/s12935-021-01875-2

Mahjoub, M. R., Xie, Z., & Stearns, T. (2010). Cep120 is asymmetrically localized to the daughter centriole and is essential for centriole assembly. *Journal of Cell Biology*, *191*(2), 331–346. https://doi.org/10.1083/jcb.201003009

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. https://doi.org/10.1038/nature08494

Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, *469*(7330), 343–349. https://doi.org/10.1038/nature09784

Marques, S. M., Peralta, F., & Esteves da Silva, J. C. G. (2009). Optimized chromatographic and bioluminescent methods for inorganic pyrophosphate based on its conversion to ATP by firefly luciferase. *Talanta*, *77*(4), 1497–1503. https://doi.org/10.1016/j.talanta.2008.09.032

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L., & Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nature Immunology*, *10*(10), 1102–1109. https://doi.org/10.1038/ni.1786

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullow, H. J., McKenzie, Z. M., Kubiak, R. L., St. Pierre, R., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, *175*(5), 1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032

Mateos, M.-V., Hernández, M.-T., Giraldo, P., de la Rubia, J., de Arriba, F., Corral, L. L., Rosiñol, L., Paiva, B., Palomera, L., Bargay, J., Oriol, A., Prosper, F., López, J., Olavarría, E., Quintana, N., García, J.-L., Bladé, J., Lahuerta, J.-J., & San Miguel, J.-F. (2013). Lenalidomide plus Dexamethasone for High-Risk Smoldering Multiple Myeloma. *New England Journal of Medicine*, *369*(5), 438–447. https://doi.org/10.1056/nejmoa1300439

Mattsson, J., Ekdahl, L., Junghus, F., Ajore, R., Erlandsson, E., Niroula, A., Pertesi, M., Frendéus, B., Teige, I., & Nilsson, B. (2021). Accelerating target deconvolution for therapeutic antibody candidates using highly parallelized genome editing. *Nature Communications*, *12*(1), 1–8. https://doi.org/10.1038/s41467-021-21518-4

Merlini, G., Dispenzieri, A., Sanchorawala, V., Schönland, S. O., Palladini, G., Hawkins, P. N., & Gertz, M. A. (2018). Systemic immunoglobulin light chain amyloidosis. *Nature Reviews Disease Primers*, *4*(1). https://doi.org/10.1038/s41572-018-0034-3

Mitchell, J. S., Li, N., Weinhold, N., Försti, A., Ali, M., Van Duin, M., Thorleifsson, G., Johnson, D. C., Chen, B., Halvarsson, B. M., Gudbjartsson, D. F., Kuiper, R., Stephens, O. W., Bertsch, U., Broderick, P., Campo, C., Einsele, H., Gregory, W. A., Gullberg, U., … Houlston, R. S. (2016). Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nature Communications*, *7*(May). https://doi.org/10.1038/ncomms12050

Mojica, F. J. M., & Montoliu, L. (2016). On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. *Trends in Microbiology*, *24*(10), 811–820. https://doi.org/10.1016/j.tim.2016.06.005

Moreno-Bueno, G., Rodríguez-Perales, S., Sánchez-Estévez, C., Hardisson, D., Sarrió, D., Prat, J., Cigudosa, J. C., Matias-Guiu, X., & Palacios, J. (2003). Cyclin D1 gene (CCND1) mutations in endometrial cancer. *Oncogene*, *22*(38), 6115–6118. https://doi.org/10.1038/sj.onc.1206868

Morgan, G. J., Johnson, D. C., Weinhold, N., Goldschmidt, H., Landgren, O., Lynch, H. T., Hemminki, K., & Houlston, R. S. (2014). Inherited genetic susceptibility to multiple myeloma. *Leukemia*, *28*(3), 518–524. https://doi.org/10.1038/leu.2013.344

Moser-Katz, T., Joseph, N. S., Dhodapkar, M. V., Lee, K. P., & Boise, L. H. (2021). Game of Bones: How Myeloma Manipulates Its Microenvironment. *Frontiers in Oncology*, *10*(February), 1–16. https://doi.org/10.3389/fonc.2020.625199

Mouhieddine, T. H., Weeks, L. D., & Ghobrial, I. M. (2019). Monoclonal gammopathy of undetermined significance. *Blood*, *133*(23), 2484–2494. https://doi.org/10.1182/blood.2019846782

Mullard, A. (2021). CRISPR technologies are going to need a bigger toolbox. *Nature Reviews. Drug Discovery*, *20*(11), 808–809. https://doi.org/10.1038/d41573-021-00177-6

Neriishi, K., Nakashima, E., & Suzuki, G. (2003). Monoclonal gammopathy of undetermined significance in atomic bomb survivors: Incidence and transformation to multiple myeloma. *British Journal of Haematology*, *121*(3), 405–410. https://doi.org/10.1046/j.1365-2141.2003.04287.x

Newby, G. A., & Liu, D. R. (2021). In vivo somatic cell base editing and prime editing. *Molecular Therapy*, *29*(11), 3107–3124. https://doi.org/10.1016/j.ymthe.2021.09.002

Niroula, A., Ajore, R., & Nilsson, B. (2019). MPRAscore: Robust and non-parametric analysis of massively parallel reporter assays. *Bioinformatics*, *35*(24), 5351–5353. https://doi.org/10.1093/bioinformatics/btz591

Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N., Yosef, N., Chang, C. Y., Shay, T., Frampton, G. M., Drake, A. C. B., Leskov, I., Nilsson, B., Preffer, F., Dombkowski, D., Evans, J. W., Liefeld, T., Smutko, J. S., … Ebert, B. L. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*, *144*(2), 296–309. https://doi.org/10.1016/j.cell.2011.01.004

Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M., & Corcoran, L. M. (2015). The generation of antibody-secreting plasma cells. *Nature Reviews Immunology*, *15*(3), 160–171. https://doi.org/10.1038/nri3795

Obeng-Gyasi, S., Graham, N., Kumar, S., Lee, J. W., Jacobus, S., Weiss, M., Cella, D., Zhao, F., Ip, E. H., O'Connell, N., Hong, F., Peipert, D. J., Gareen, Ii. I. F., Timsina, L. R., Gray, R., Wagner, L. I., & Carlos, R. C. (2022). Examining allostatic load, neighborhood socioeconomic status, symptom burden and mortality in multiple myeloma patients. *Blood Cancer Journal*, *12*(4), 1–8. https://doi.org/10.1038/s41408-022-00648-y

Ocio, E. M., Montes-Gaisán, C., Bustamante, G., Garzón, S., González, E., Pérez, E., Sirvent, M., Arguiñano, J. M., González, Y., Ríos, R., de Miguel, D., Grande, M., Fernández, A., Naves, A., & Rosiñol, L. (2022). Clinical and Sociodemographic Characteristics of Patients With Relapsed and/or Refractory Multiple Myeloma and Their influence on Treatment in the Real-World Setting in Spain: The CharisMMa Study. *Clinical Lymphoma, Myeloma and Leukemia*, *22*(4), e241–e249. https://doi.org/10.1016/j.clml.2021.10.001

Park, K. S., Bayles, I., Szlachta-McGinn, A., Paul, J., Boiko, J., Santos, P., Liu, J., Wang, Z., Borghesi, L., & Milcarek, C. (2014). Transcription Elongation Factor ELL2 Drives Ig Secretory-Specific mRNA Production and the Unfolded Protein Response. *The Journal of Immunology*, *193*(9), 4663–4674. https://doi.org/10.4049/jimmunol.1401608

Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

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 Heras, N. Las, 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*(7), 2586–2592. https://doi.org/10.1182/blood-2007-05-088443

Perini, T., Materozzi, M., & Milan, E. (2021). The Immunity-malignancy equilibrium in multiple myeloma: lessons from oncogenic events in plasma cells. *FEBS Journal*, *289*, 4383–4397. https://doi.org/10.1111/febs.16068

Pertesi, M., Ekdahl, L., Palm, A., Johnsson, E., Järvstråt, L., Wihlborg, A. K., & Nilsson, B. (2019). Essential genes shape cancer genomes through linear limitation of homozygous deletions. *Communications Biology*, *2*(1), 1–11. https://doi.org/10.1038/s42003-019-0517-0

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. (2019). Exome sequencing identi fi es germline variants in DIS3 in familial multiple myeloma. *Leukemia*, *33*, 2324–2330.

Pertesi, M., Went, M., Hansson, M., Hemminki, K., Houlston, R. S., & Nilsson, B. (2020). Genetic predisposition for multiple myeloma. *Leukemia*, *34*, 697–708. https://doi.org/10.1038/s41375-019-0703-6

Pessoa de Magalhães Filho, R. J., Crusoe, E., Riva, E., Bujan, W., Conte, G., Navarro Cabrera, J. R., Garcia, D. K., Vega, G. Q., Macias, J., Oliveros Alvear, J. W., Royg, M., Neves, L. A., Lopez Dopico, J. L., Espino, G., Ortiz, D. R., Socarra, Z., Fantl, D., Ruiz-Arguelles, G. J., Maiolino, A., … Durie, B. (2019). Analysis of Availability and Access of Anti-myeloma Drugs and Impact on the Management of Multiple Myeloma in Latin American Countries. *Clinical Lymphoma, Myeloma and Leukemia*, *19*(1), e43–e50. https://doi.org/10.1016/j.clml.2018.08.005

Popejoy, A. B. (2021). Too many scientists still say Caucasian. *Nature*, *596*(7873), 463–463. https://doi.org/10.1038/d41586-021-02288-x

Preuss, F., Chatterjee, D., Mathea, S., Shrestha, S., St-Germain, J., Saha, M., Kannan, N., Raught, B., Rottapel, R., & Knapp, S. (2020). Nucleotide Binding, Evolutionary Insights, and Interaction Partners of the Pseudokinase Unc-51-like Kinase 4. *Structure*, *28*(11), 1184-1196.e6. https://doi.org/10.1016/j.str.2020.07.016

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Puda, A., Milosevic, J. D., Berg, T., Klampfl, T., Harutyunyan, A. S., Gisslinger, B., Rumi, E., Pietra, D., Malcovati, L., Elena, C., Doubek, M., Steurer, M., Tosic, N., Pavlovic, S., Guglielmelli, P., Pieri, L., Vannucchi, A. M., Gisslinger, H., Cazzola, M., & Kralovics, R. (2012). Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. *American Journal of Hematology*, *87*(3), 245–250. https://doi.org/10.1002/ajh.22257

Raje, N., & Yee, A. J. (2020). How We Approach Smoldering Multiple Myeloma oncology grand rounds. *Journal of Clinical Oncology*, 15–18. https://doi.org/10.1200/JCO.19.02834

Rajkumar, S. V. (2020). Multiple myeloma: 2020 update on diagnosis, risk-stratification and management. *American Journal of Hematology*, *95*(5), 548–567. https://doi.org/10.1002/ajh.25791

Rajkumar, S. V., & Kumar, S. (2016). Multiple Myeloma: Diagnosis and Treatment. *Mayo Clinic Proceedings*, *91*(1), 101–119. https://doi.org/10.1016/j.mayocp.2015.11.007

Rangel-Pozzo, A., Yu, P. L. I., Lal, S., Asbaghi, Y., Sisdelli, L., Tammur, P., Tamm, A., Punab, M., Klewes, L., Louis, S., Knecht, H., Olujohungbe, A., & Mai, S. (2021). Telomere architecture correlates with aggressiveness in multiple myeloma. *Cancers*, *13*(8), 1–16. https://doi.org/10.3390/cancers13081969

Rashkin, S. R., Graff, R. E., Kachuri, L., Thai, K. K., Alexeeff, S. E., Blatchins, M. A., Cavazos, T. B., Corley, D. A., Emami, N. C., Hoffman, J. D., Jorgenson, E., Kushi, L. H., Meyers, T. J., Van Den Eeden, S. K., Ziv, E., Habel, L. A., Hoffmann, T. J., Sakoda, L. C., & Witte, J. S. (2020). Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18246-6

Ray, J. P., de Boer, C. G., Fulco, C. P., Lareau, C. A., Kanai, M., Ulirsch, J. C., Tewhey, R., Ludwig, L. S., Reilly, S. K., Bergman, D. T., Engreitz, J. M., Issner, R., Finucane, H. K., Lander, E. S., Regev, A., & Hacohen, N. (2020). Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15022-4

Reimold, A. M., Iwakoshi, N. N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallese, E. M., Friend, D., Grusby, M. J., Alt, F., & Glimcher, L. H. (2001). Plasma cell differentiation requires the transcription factor XBP-1. *Nature*, *412*(6844), 300–307. https://doi.org/10.1038/35085509

Salzer, U., Chapel, H. M., Webster, A. D. B., Pan-Hammarström, Q., Schmitt-Graeff, A., Schlesier, M., Peter, H. H., Rockstroh, J. K., Schneider, P., Schäffer, A. A., Hammarström, L., & Grimbacher, B. (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nature Genetics*, *37*(8), 820–828. https://doi.org/10.1038/ng1600

Schagat, T., Paguio, A., Kopish, K., & Promega corporation. (2007). *Normalizing genetic reporter assays : approaches and considerations for increasing consistency and statistical significance*. *17*, 9–12.

Shaffer, A. L., Emre, N. C. T., Lamy, L., Ngo, V. N., Wright, G., Xiao, W., Powell, J., Dave, S., Yu, X., Zhao, H., Zeng, Y., Chen, B., Epstein, J., & Staudt, L. M. (2008). IRF4 addiction in multiple myeloma. *Nature*, *454*(7201), 226–231. https://doi.org/10.1038/nature07064

Shah, U. A., & Mailankody, S. (2020). Emerging immunotherapies in multiple myeloma. *The BMJ*, *370*(fig 1). https://doi.org/10.1136/bmj.m3176

Shah, V., Boyd, K. D., Houlston, R. S., & Kaiser, M. F. (2017). Constitutional mutation in CDKN2A is associated with long term survivorship in multiple myeloma: A case report. *BMC Cancer*, *17*(1), 2–7. https://doi.org/10.1186/s12885-017-3715-5

Shamambo, L. J., & Henry, T. L. (2022). Rethinking the Use of "Caucasian" in Clinical Language and Curricula: a Trainee's Call to Action. *Journal of General Internal Medicine*, *37*(7), 1780–1782. https://doi.org/10.1007/s11606-022-07431-6

Shapiro-Shelef, M., & Calame, K. C. (2005). Regulation of plasma-cell development. *Nature Reviews Immunology*, *5*(3), 230–242. https://doi.org/10.1038/nri1572

Sherborne, A. L., Hosking, F. J., Prasad, R. B., Kumar, R., Koehler, R., Vijayakrishnan, J., Papaemmanuil, E., Bartram, C. R., Stanulla, M., Schrappe, M., Gast, A., Dobbins, S. E., Ma, Y., Sheridan, E., Taylor, M., Kinsey, S. E., Lightfoot, T., Roman, E., Irving, J. A. E., … Houlston, R. S. (2010). Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nature Genetics*, *42*(6), 492–494. https://doi.org/10.1038/ng.585

Sherf, B. A., Navarro, S. L., Hannah, R. R., Wood, K. V., & Promega Corporation. (1996). Dual-Luciferase™ Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase. *Promega Notes Magazine*, *2*(57).

Shiina, T., Hosomichi, K., Inoko, H., & Kulski, J. K. (2009). The HLA genomic loci map: Expression, interaction, diversity and disease. *Journal of Human Genetics*, *54*(1), 15–39. https://doi.org/10.1038/jhg.2008.5

Singh, P., Ravanan, P., & Talwar, P. (2016). Death associated protein kinase 1 (DAPK1): A regulator of apoptosis and autophagy. *Frontiers in Molecular Neuroscience*, *9*(JUNE), 1–11. https://doi.org/10.3389/fnmol.2016.00046

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *ELife*, *6*, 1–35. https://doi.org/10.7554/eLife.21856

Slamanig, S. A., & Nolte, M. A. (2021). The bone marrow as sanctuary for plasma cells and memory t-cells: Implications for adaptive immunity and vaccinology. *Cells*, *10*(6). https://doi.org/10.3390/cells10061508

Steinke, J. W., Hodsdon, W., Parenti, S., Ostraat, R., Lutz, R., Borish, L., & Hagman, J. (2004). Identification of an Sp factor-dependent promoter in GCET, a gene expressed at high levels in germinal center B cells. *Molecular Immunology*, *41*(12), 1145–1153. https://doi.org/10.1016/j.molimm.2004.06.031

Stepanikova, I., & Oates, G. R. (2017). Perceived Discrimination and Privilege in Health Care: The Role of Socioeconomic Status and Race. *American Journal of Preventive Medicine*, *52*(1), S86–S94. https://doi.org/10.1016/j.amepre.2016.09.024

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, *364*(6448), 48–53. https://doi.org/10.1126/science.aax9181

Stringer, S., Wray, N. R., Kahn, R. S., & Derks, E. M. (2011). Underestimated effect sizes in GWAS: Fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS ONE*, *6*(11), 1–7. https://doi.org/10.1371/journal.pone.0027964

Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J. P., Manning, A. K., Grarup, N., Sim, X., Barnes, D. R., Witkowska, K., Staley, J. R., Tragante, V., Tukiainen, T., Yaghootkar, H., Masca, N., Freitag, D. F., Ferreira, T., Giannakopoulou, O., Tinker, A., … Munroe, P. B. (2016). Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature Genetics*, *48*(10), 1151–1161. https://doi.org/10.1038/ng.3654

Swaminathan, B., Thorleifsson, G., Jöud, M., Ali, M., Johnsson, E., Ajore, R., Sulem, P., Halvarsson, B. M., Eyjolfsson, G., Haraldsdottir, V., Hultman, C., Ingelsson, E., Kristinsson, S. Y., Kähler, A. K., Lenhoff, S., Masson, G., Mellqvist, U. H., Månsson, R., Nelander, S., … Nilsson, B. (2015). Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nature Communications*, *6*(May), 1–8. https://doi.org/10.1038/ncomms8213

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1

Taub, M. A., Conomos, M. P., Keener, R., Iyer, K. R., Weinstock, J. S., Yanek, L. R., Lane, J., Miller-Fleming, T. W., Brody, J. A., Raffield, L. M., McHugh, C. P., Jain, D., Gogarten, S. M., Laurie, C. A., Keramati, A., Arvanitis, M., Smith, A. V., Heavner, B., Barwick, L., … Mathias, R. A. (2022). Genetic determinants of telomere length from 109,122 ancestrally diverse whole-genome sequences in TOPMed. *Cell Genomics*, *2*(1), 100084. https://doi.org/10.1016/j.xgen.2021.100084

Tewhey, R., Kotliar, D., Park, D. S., Liu, B., Winnicki, S., Reilly, S. K., Andersen, K. G., Mikkelsen, T. S., Lander, E. S., Schaffner, S. F., & Sabeti, P. C. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*, *165*(6), 1519–1529. https://doi.org/10.1016/j.cell.2016.04.027

Turner, C. A., Mack, D. H., & Davis, M. M. (1994). Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell*, *77*(2), 297–306. https://doi.org/10.1016/0092-8674(94)90321-2

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Ulirsch, J. C., Lareau, C. A., Bao, E. L., Ludwig, L. S., Guo, M. H., Benner, C., Satpathy, A. T., Kartha, V. K., Salem, R. M., Hirschhorn, J. N., Finucane, H. K., Aryee, M. J., Buenrostro, J. D., & Sankaran, V. G. (2019). Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature Genetics*, *51*(4), 683–693. https://doi.org/10.1038/s41588-019-0362-6

VanLiere, J. M., & Rosenberg, N. A. (2008). Mathematical properties of the r2 measure of linkage disequilibrium. *Theoretical Population Biology*, *74*(1), 130–137. https://doi.org/10.1016/j.tpb.2008.05.006

Verkleij, C. P. M., Frerichs, K. A., Broekmans, M., Absalah, S., Maas-Bosman, P. W. C., Kruyswijk, S., Nijhof, I. S., Mutis, T., Zweegman, S., & van de Donk, N. W. C. J. (2020). T-cell redirecting bispecific antibodies targeting BCMA for the treatment of multiple myeloma. *Oncotarget*, *11*(45), 4076–4081. https://doi.org/10.18632/oncotarget.27792

Vidal, M., & Starowicz, K. (2017). Polycomb complexes PRC1 and their function in hematopoiesis. *Experimental Hematology*, *48*, 12–31. https://doi.org/10.1016/j.exphem.2016.12.006

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era - Concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. https://doi.org/10.1038/nrg2322

Waby, J., Bingle, C., & Corfe, B. (2008). Post-Translational Control of Sp-Family Transcription Factors. *Current Genomics*, *9*(5), 301–311. https://doi.org/10.2174/138920208785133244

Waller, R. G., Darlington, T. M., Wei, X., Madsen, M. J., Thomas, A., Curtin, K., Coon, H., Rajamanickam, V., Musinsky, J., Jayabalan, D., Atanackovic, D., Rajkumar, S. V., Kumar, S., Slager, S., Middha, M., Galia, P., Demangel, D., Salama, M., Joseph, V., … Camp, N. J. (2018). Novel pedigree analysis implicates DNA repair and chromatin remodeling in multiple myeloma risk. *PLoS Genetics*, *14*(2), 1–19. https://doi.org/10.1371/journal.pgen.1007111

Waxman, A. J., Mink, P. J., Devesa, S. S., Anderson, W. F., Weiss, B. M., Kristinsson, S. Y., McGlynn, K. A., & Landgren, O. (2010). Racial disparities in incidence and outcome in multiple myeloma: A population-based study. *Blood*, *116*(25), 5501–5506. https://doi.org/10.1182/blood-2010-07-298760

Wei, X., Calvo-Vidal, M. N., Chen, S., Wu, G., Revuelta, M. V., Sun, J., Zhang, J., Walsh, M. F., Nichols, K. E., Joseph, V., Snyder, C., Vachon, C. M., McKay, J. D., Wang, S.-P., Jayabalan, D. S., Jacobs, L. M., Becirovic, D., Waller, R. G., Artomov, M., … Lipkin, S. M. (2018). Germline Lysine-Specific Demethylase 1 (LSD1/KDM1A) Mutations Confer Susceptibility to Multiple Myeloma. *Cancer Research*, *78*(10), 2747–2759. https://doi.org/10.1016/j.physbeh.2017.03.040

Weinhold, N., Johnson, D. C., Chubb, D., Chen, B., Försti, A., Hosking, F. J., Broderick, P., Ma, Y. P., Dobbins, S. E., Hose, D., Walker, B. A., Davies, F. E., Kaiser, M. F., Li, N. L., Gregory, W. A., Jackson, G. H., Witzens-Harig, M., Neben, K., Hoffmann, P., … Hemminki, K. (2013). The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nature Genetics*, *45*(5), 522–525. https://doi.org/10.1038/ng.2583

Weinhold, N., Meissner, T., Johnson, D. C., Seckinger, A., Moreaux, J., Försti, A., Chen, B., Nickel, J., Chubb, D., Rawstron, A. C., Doughty, C., Dahir, N. B., Begum, D. B., Young, K., Walker, B. A., Hoffmann, P., Nöthen, M. M., Davies, F. E., Klein, B., … Hemminki, K. (2015). The 7p15.3 (rs4487645) association for multiple myeloma shows strong allele-specific regulation of the myc-interacting gene cdca7l in malignant plasma cells. *Haematologica*, *100*(3), e110–e113. https://doi.org/10.3324/haematol.2014.118786

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., … Thibodeau, S. N. (2018). Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. *Nature Communications*, *9*(1), 1–10. https://doi.org/10.1038/s41467-018-04989-w

Xie, S., Duan, J., Li, B., Zhou, P., & Hon, G. C. (2017). Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Molecular Cell*, *66*(2), 285-299.e5. https://doi.org/10.1016/j.molcel.2017.03.007

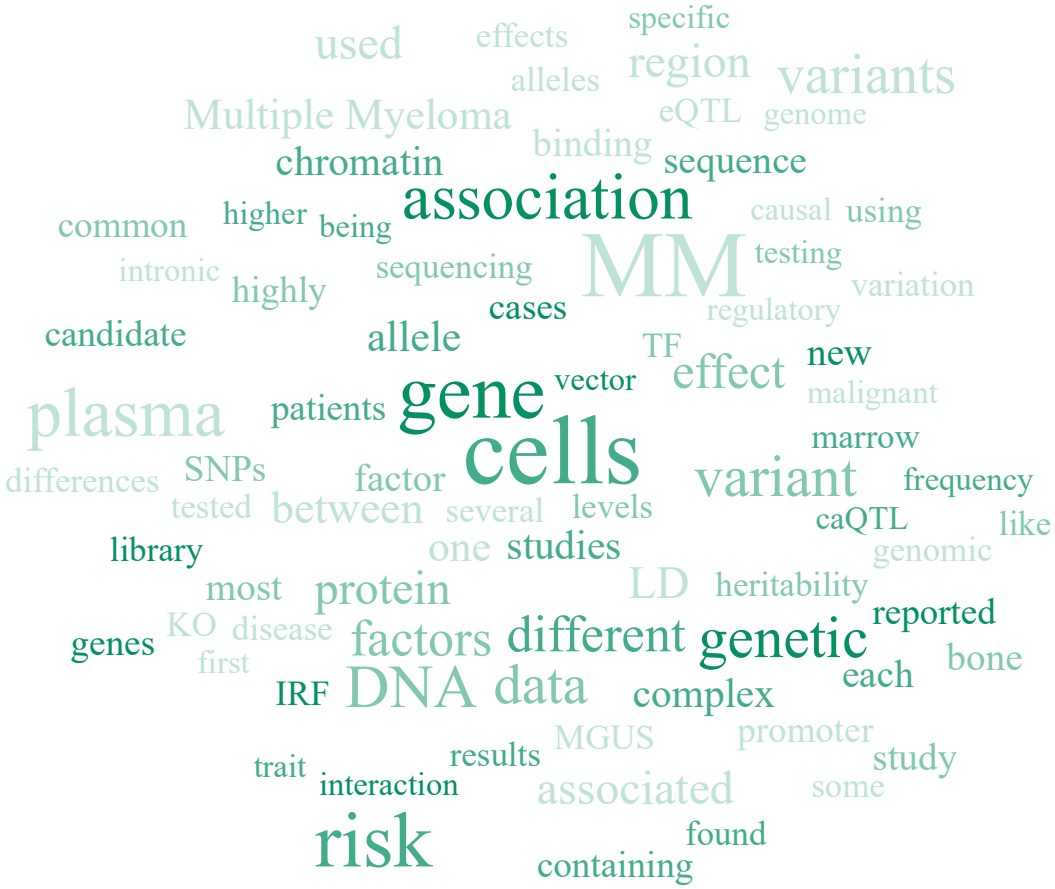
Ye, B., Liu, B., Yang, L., Huang, G., Hao, L., Xia, P., Wang, S., Du, Y., Qin, X., Zhu, P., Wu, J., Sakaguchi, N., Zhang, J., & Fan, Z. (2017). Suppression of SRCAP chromatin remodelling complex and restriction of lymphoid lineage commitment by Pcid2. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01788-7

Yun, Z., Zhichao, J., Hao, Y., Ou, J., Ran, Y., Wen, D., & Qun, S. (2017). Targeting autophagy in multiple myeloma. *Leukemia Research*, *59*(155), 97–104. https://doi.org/10.1016/j.leukres.2017.06.002

Zhang, F., & Yu, X. (2012). WAC, a functional partner of RNF20/40, regulates histone H2B ubiquitination and gene transcription. *Molecular Cell*, *41*(4), 384–397. https://doi.org/doi:10.1016/j.molcel.2011.01.024.

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C. H., Brown, M., Zhang, X., Meyer, C. A., & Liu, X. S. (2019). Cistrome Data Browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Research*, *47*(D1), D729–D735. https://doi.org/10.1093/nar/gky1094

Zhou, L., Yu, Q., Wei, G., Wang, L., Huang, Y., Hu, K., Hu, Y., & Huang, H. (2021). Measuring the global, regional, and national burden of multiple myeloma from 1990 to 2019. *BMC Cancer*, *21*(1), 1–13. https://doi.org/10.1186/s12885-021-08280-y



Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2022:132
ISBN 978-91-8021-294-6
ISSN 1652-8220