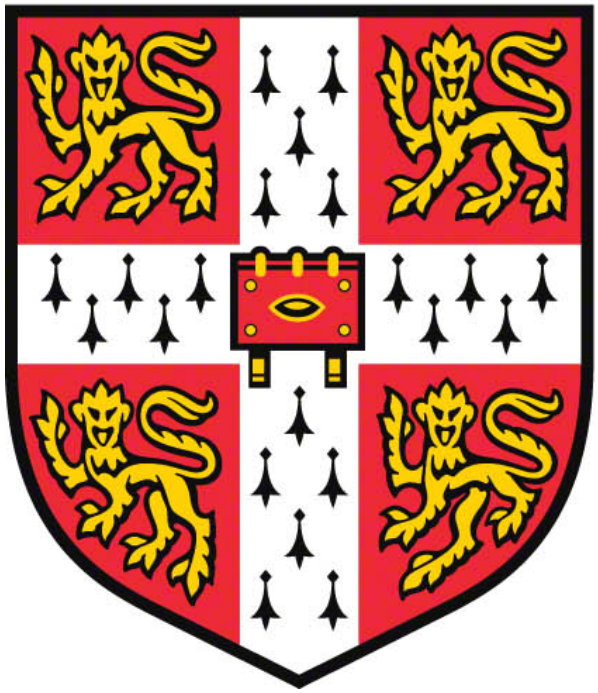
Using genetic and genomic approaches to understand haematopoietic cellular biology and dysregulation in disease

Alice Louise Mann

St John's College

Wellcome Sanger Institute



University of Cambridge

November 2017

This dissertation is submitted for the degree of Doctor of Philosophy

Using genetic and genomic approaches to understand haematopoietic cellular biology and dysregulation in disease

Alice Louise Mann
St John's College, University of Cambridge
Professor Nicole Soranzo

Summary

Genetic and genomic approaches have revolutionised the way we address disease aetiology, potential treatment and methods to understand fundamental biology. Many different approaches can be applied to attempt to resolve the mechanisms through which sequence variation disrupts downstream biological processes, which I discuss and apply in this thesis. Specifically, I use tractable haematopoietic cellular systems focusing mainly on neutrophils but also extending these analyses to monocytes and naïve $CD4^+$ cells. First, I introduce the fundamental principles of human genetic variation and associated challenges in resolving functional mechanisms. I then discuss how immune functions are dysregulated in classical autoimmune diseases and emerging evidence for the role of these cells in complex disorders not previously considered immune-mediated. I then integrate molecular phenotypes from resting monocytes, neutrophils and $CD4^+$ T cells with disease-risk loci. Molecular data have the advantage of enabling measurement in larger cohorts and have therefore been used in quantitative trait loci studies to identify variants influencing processes such as gene expression, histone modification or splicing. Using these data, I map molecular mechanisms acting at risk loci associated with a range of complex disorders.

Following this, I highlight recent efforts in applying systematic genome-wide association approaches to cellular and functional traits, many of which can represent intermediate processes disrupted by complex disease. I then apply such approaches to novel neutrophil functional phenotypes to ascertain whether such population-based approaches can be used to gain insight into neutrophil biology. Finally, I discuss studies of haematological blood cell count traits and immunophenotyping and apply a targeted recall-by-genotype study to dissect the relationship between these traits, specifically neutrophil count and surface receptor expression.

In summary, I demonstrate how describing biological mechanisms of genetic variants requires the integration of multiple and complementary datasets and offers insight into fundamental biology, disease risk and therapeutic utility.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed word limit for the Biology Degree Committee.

A. L. Mann
November 2017

Acknowledgements

First, I would like to thank my supervisor, Nicole, for the opportunity of working in a stimulating field but also the unbelievable amount that I learnt from the mentorship and the advice. Also thank you for the time, space and patience that enabled me to become an independent scientist and a more confidence person all-round. Second, to the rest of Team Soranzo. There are no words that can properly express my gratitude for the help and advice over these past four years. It is incredible to think of how much I have learnt from those around me and this thesis is testament to what a fantastic team it really is. Particular thanks to Klaudia, who really taught me how to do computational and statistical analysis and for always answering questions even in the busy periods! Thank you to Louella, also for all the questions, the teaching and the unwavering positivity. The same goes for Valentina, Kousik, Lorenzo and Heather- for teaching me some great command line tricks. To Yumi for the enduring support both in work and in life. To Lisa for fantastic pointers on how to make figures look brilliant. Finally, thank you to Stephen for keeping the gene regulation interest going and for putting up with being my desk neighbour and my actual neighbour!

Thank you to all my collaborators. Kate, for enabling me to keep my lab skills going and for the mentorship. To Taco and Anton, for great neutrophil discussions, for the chance to visit Amsterdam and also for your hard work in organising all the replication studies and offering opportunities to follow up interesting results. Thank you to my thesis adviser, Edwin, for stimulating discussions and always a fresh perspective on the challenges. The same to my second adviser, Dan, for useful discussions and pointers. To Dirk, for general advice and also for feedback on my thesis. Thank you to the graduate programme committee, Julian, Carl, Annabel and Christina. Also to Paris, Carol and Sally for all the help and support. To everyone in pipelines at WTSI for helping everything run smoothly. I also thank and acknowledge my funders, the Medical Research Council and the Wellcome Trust.

One of the greatest things about being a PhD student at WTSI, is the fantastic student support network and the friends you make along the way, the PhD life would not have been the same without you all.

To my family, my biggest support network. Being a PhD student in my hometown meant that my family went through the same experiences that I did and were there to get me through it when I needed a little extra push.

To my husband, Tim, there would absolutely be no thesis without you, there are not enough ways for me to thank you for the past four years.

The work described in this thesis was funded by the Medical Research Council and the Wellcome Trust. I thank both of my funding bodies for enabling me to undertake my PhD research.

This thesis is dedicated to my husband,

Timothy Mann

Table of Contents

List of Figures

List of Tables

Abbreviations

AAT	Anti-inflammatory alpha-1-antitrypsin
AAV	Antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis
ABCA1	ATP-binding cassette transporter
AID	Autoimmune disease
AMD	Age-related macular degeneration
AML	Acute myeloid leukaemia
ANCA	Antineutrophil cytoplasmic antibodies
APC	Allophycocyanin
APOE	Apolipoprotein E
APP	Amyloid precursor protein
ARHGEF26	Rho guanine nucleotide exchange factor 26
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
ATP	Adenosine triphosphate
ATRA	All trans-retinoic acid
Aβ	Amyloid β
BAFF	B cell activating factor
BBB	Blood-brain barrier
BM	Bruch's membrane
BPI	Bactericidal/permeability-increasing protein
BRE	TFIIB recognition element
C/EBP	CCAAT/enhancer binding protein
CAD	Coronary artery disease
CANTOS	Canakinumab Antiinflammatory Thrombosis Outcome Study
CBR	Cambridge BioResource
CD	Crohn's disease
CDCV	Common disease-common variant
CEL	Celiac disease
CETP	Cholesterylester transfer protein
CFH	Complement factor H
CFI	Complement factor I
CHD	Coronary heart disease
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP-seq	Chromatin immunoprecipitation with next-generation sequencing
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CNV	Choroidal neovascular membranes

COPD	Chronic obstructive pulmonary disorder
CR1	Complement factor 1
CRC	Colorectal cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRP	C-reactive protein
CTCF	Transcriptional repressor CTCF
CVD	Cardiovascular disease
CytoB	Cytochalasin B
DC	Dendritic cell
DG	Diacylglycerol
DHS	Dnase I hypersensitive site
DMSO	Dimethyl sulfoxide
DNMT	DNA methyltransferase
DPE	Downstream promoter element
DTT	Dithiothreitol
EA	Effect allele
EAE	Experimental autoimmune encephalomyelitis
EAF	Effect allele frequency
eQTL	QTL for gene expression
ERK	Extracellular signal-related kinase
FACs	Fluorescence-activated cell sorting
FBC	Full blood count
FEV1	Forced expiratory volume
FITC	Fluorescein isothiocyanate
fMLP	N-formylmethionine-leucyl-phenylalanine
FPKM	Fragments per kilobase of transcript per million fragments sequenced
FS	Forward scatter
FVC	Forced vital capacity
GARFIELD	GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction
GCSFR	Granulocyte colony-stimulating factor receptor
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GMP	Granulocyte/macrophage progenitor
GPCR	G-protein-coupled receptors
GPI	Glycosyl phosphatidylinositol anchor
GTFs	General transcription factors
GWAS	Genome-wide association studies
HAT	Histone acetyltransferase

HDAC	Histone deacetylase
HDM	Histone demethylase
Hep3B	Human hepatocellular carcinoma cells
HFGP	Human Functional genomics project
HLA	Human Leukocyte antigen
hQTL	QTL for histone modification
HRP	Horseradish peroxidase
HSC	Haematopoietic stem cells
IBD	Inflammatory bowel disease
ICAM	Intercellular-adhesion molecules
IFN	Interferon
IGAP	International Genomics of Alzheimer's Project
IL	Interleukin
Inr	Initiator element
iPSC	Induced pluripotent stem cell
LBP	Lipopolysaccharide binding protein
LCL	Lymphoblastoid cell lines
lcRNA	Long non-coding RNA
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDL-C	Low-density lipoprotein cholesterol
LFA-1	Lymphocyte function-associated antigen 1
LIPC	Lipase C
LOAD	Late-onset Alzheimer's disease
LPS	Lipopolysaccharides
MAC	Membrane attack complex
MAC-1	Macrophage-1 antigen
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinase
MD	Maximum difference (effect size estimate for isotype QTLs)
MDP	Muramyl dipeptide
MEP	Megakaryocyte/erythroid progenitor
MFI	Median fluorescence intensity
MHC	Major histocompatibility complex
miRNA	micro RNA
MPO	Myeloperoxidase
MR	Mendelian randomization
MS	Multiple sclerosis

NE-FSC	Neutrophil forward scatter parameter
NE-SFL	Neutrophil side fluorescence
NET	Neutrophil extracellular traps
NK cells	Natural killer cells
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
nvAMD	Neovascular AMD
OA	Other allele
OR	Odds ratio
P-TEFb	Positive transcription elongation factor b
PAF	Platelet-activating factor
PBMC	Peripheral blood mononuclear cells
PBPC	Peripheral blood progenitor cells
PBPCT	Peripheral blood progenitor cells transplantation
PBS	Phosphate-buffered saline
PcHiC	Promoter-capture HiC
PD	Parkinson's disease
PDAC	Pancreatic ductal adenocarcinoma
PE	Phycoerythrin
PI3K	Phosphoinositide 3-kinase
PKC	Protein kinase C
PLAUR/uPAR	Urokinase receptor
PLC	Phospholipase C
PMA	Phorbol myristate acetate
PMNs	Polymorphonuclear leukocytes
Pol II	RNA polymerase II
PP	Posterior probability
PR3	Proteinase 3
PRCS	Peripheral retinal pigment epithelium/choroid/sclera
PSGL-1	P-selectin glycoprotein ligand-1
QTL	Quantitative trait locus
RA	Rheumatoid arthritis
RbG	Recall-by-genotype
RCT	Randomised clinical trial
RDW	Red cell distribution width
RFU	Relative fluorescence unit
RISC	RNA-induced silencing complex
RNAP	RNA polymerase
ROS	Reactive oxygen species

RPE	Retinal pigment epithelium
SCN	Severe congenital neutropenia
SD	Standard deviation
SE	Standard error
siRNA	Small interfering RNA
SJIA	Systemic juvenile idiopathic arthritis
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SS	Side scatter
STZ	Serum-treated zymosan
T1D	Type 1 Diabetes
T2D	Type 2 diabetes
TAD	Topologically-associated domain
TF	Transcription factor
TLR	Toll-like receptor
TM	Transmembrane domain
TNF	Tumour necrosis factor
TNFRSF10A	Tumour necrosis factor receptor superfamily 10A
TRAIL	Tumour necrosis factor-related apoptosis-inducing ligand
TRAILR	Tumour necrosis factor-related apoptosis-inducing ligand receptor
T_{REG}	Regulatory T cells
TSS	Transcription start site
UTR	Untranslated region
VCM	Variable chromatin modules
VEGFA	Vascular endothelial growth factor A
VLDL	Very-low density lipoprotein
WGS	Whole-genome sequencing
YRI	Yoruba in Ibadan