

Regulation of gene expression in macrophage immune response



Kaur Alasoo

Wellcome Trust Sanger Institute
Hughes Hall

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy
September 2016

Abstract

Regulation of gene expression in macrophage immune response

Kaur Alasoo

Gene expression quantitative trait loci (eQTL) mapping studies can provide mechanistic insights into the functions of disease-associated variants. However, many eQTLs are cell type and context specific. This is particularly relevant for immune cells, whose cellular function and behaviour can be substantially altered by external cues. Furthermore, understanding mechanisms behind eQTLs is hindered by the difficulty of identifying causal variants. We differentiated macrophages from induced pluripotent stem cells from 86 unrelated, healthy individuals derived as part of the Human Induced Pluripotent Stem Cells Initiative. We generated RNA-seq data from these cells in four experimental conditions: naïve, interferon-gamma (IFN γ) treatment (18h), *Salmonella* infection (5h), and IFN γ treatment followed by *Salmonella* infection. We also measured chromatin accessibility with ATAC-seq in 31-42 individuals in the same four conditions. We detected gene expression QTLs (eQTLs) for 4326 genes, over 900 of which were condition-specific. We also detected a similar number of transcript ratio QTLs (trQTLs) that influenced mRNA processing and alternative splicing. Macrophage eQTLs and trQTLs were enriched for variants associated with Alzheimer's disease, multiple autoimmune disorders and lipid traits. We also detected chromatin accessibility QTLs (caQTLs) for 14,602 accessible regions, including hundreds of long-range interactions. Joint analysis of eQTLs with caQTLs allowed us to greatly reduce the set of credible causal variants, often pinpointing to a single most likely variant. We found that caQTLs were less condition-specific than eQTLs and ~50% of the stimulation-specific eQTLs manifested on the chromatin level already in the naive cells. These observations might help to explain the discrepancy between strong enrichment of diseases associations in regulatory elements but only modest overlap with current eQTL studies, suggesting that many regulatory elements are in a 'primed' state waiting for an appropriate environmental signal before regulating gene expression.

Declaration of Originality

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 beginning of each chapter. 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. 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 of similar institution. This dissertation does not exceed the word limit set by the Degree Committee for the Faculty of Biology.

Signature:

Date:

Kaur Alasoo

September 2016

Acknowledgements

First, I am grateful to many people who have inspired and nurtured my passion for science. I thank Hedi Peterson for drawing me into the field before I had even finished my first semester at the university. I thank Phaedra Agius for introducing me into machine learning and teaching me that it is completely normal in research not to know beforehand what works and what does not. I thank Prof. Jaak Vilo for supporting and guiding me during my undergraduate years, Harri Lähdesmäki for letting me play with new and exciting data, and Isabel Sá-Correia for giving me an independent experimental project with no prior experience. All of these experiences gave me the necessary skills and confidence to help me realize that research was what I wanted to do.

Many people have made the four years of my PhD studies such a rewarding and stimulating experience. Over the past four years, fair and constructive feedback from my supervisor Daniel Gaffney has greatly improved the clarity and focus of my research and writing. I also greatly value the independence that I had in shaping the projects that I was working on. I thank my co-supervisor Gordon Dougan for accepting me to his lab even though I had never seen neither human nor a bacterial cell before. Subhankar Mukhopadhyay was very patient in explaining both intricacies of immunology as well as basics of cell culture. Finally, most of the experimental work presented in this thesis would not have been possible without the dedication and hard work of Julia Rodrigues. Thank you, Julia, for not giving up even when cells were dying and differentiations failing. I will always remember the fruitful discussions with other members of the Gaffney on a wide range of experimental on computational topics.

I have received great advice from many other people on the campus. Leopold Parts has been a great mentor by listening to my ideas, thinking about them hard and asking important questions. Conversation with Leo have made me to reconsider my experimental design more than once, ultimately allowing me to answer more important questions.

Last but not least, I am most grateful to my family. Aitäh isale nende õhtuste matemaatika ja füüsika olümpiaadiülesannete lahendamiste eest, mis olid küll väga põnevad, aga millega ma ise alati esimesel katsel hakkama ei saanud. Aitäh vennale, õele, vanaemadele, vanaisadele, tädile ja teistele suurtele ja väikestele sugulastele pidevalt meelde tuletamast, kui hea on kodus olla. Aitäh, Maret, et oled minu kõrval olnud ja mind toetanud kogu selle aja.

Table of Contents

1	Introduction	13
1.1	Regulation of cell type and condition specific gene expression	14
1.1.1	Principles of cell type specific TF binding	15
1.1.2	Signal dependent TFs bind to established enhancers	16
1.1.3	Role of signal dependent TFs in establishing new enhancers	17
1.1.4	Long range interactions between cell type specific and signal dependent TFs	17
1.2	Macrophage biology in the context of immune response	19
1.2.1	Signalling pathways activated by lipopolysaccharide and interferon-gamma	19
1.2.2	Macrophage response to <i>Salmonella</i> infection	21
1.3	Tissue culture models of macrophage biology	21
1.3.1	Differentiating macrophages from human induced pluripotent stem cells	22
1.4	Genome-wide profiling of gene expression and chromatin accessibility	23
1.4.1	RNA sequencing	23
1.4.2	Chromatin state profiling	25
1.5	Genetics of molecular traits	27
1.5.1	Genetics of gene expression	27
1.5.2	Genetics of chromatin states	29
1.5.3	Using eQTLs to interpret GWAS associations	32
1.6	Outline of the thesis	33
2	Comparison of monocyte-derived and iPSC-derived macrophages	35
2.1	Introduction	35
2.2	Methods	36
2.2.1	Samples	36
2.2.2	Cell culture and reagents	37
2.2.3	Flow cytometry	39
2.2.4	RNA extraction and sequencing	40
2.2.5	RNA-seq data analysis	41
2.3	Gene expression variation between iPSCs, IPSDMs and MDMs	45
2.3.1	Global patterns of gene expression	45
2.3.2	Differential expression analysis of IPSDMs vs MDMs	47
2.3.3	Mechanisms underlying differences between MDMs and IPSDMs	51
2.4	Global variation in alternative transcript usage	54
2.4.1	Identification and characterisation of alternative transcription events	56
2.5	Discussion	61
3	Large-scale differentiation of macrophages from human iPSCs	65
3.1	Introduction	65
3.2	Methods	67
3.2.1	Cell culture and reagents	67
3.2.2	Macrophage stimulation assays	69

3.2.3	RNA sequencing.....	69
3.2.4	Flow cytometry.....	72
3.3	Large-scale differentiation of macrophages for genomics assays	75
3.3.1	Variability in success rate	76
3.3.2	Variability in the duration of the differentiation	77
3.3.3	Variability in cell numbers	79
3.3.4	Variability in macrophage purity	79
3.4	Variability in gene expression data	81
3.4.1	Technical variability between RNA-seq samples	81
3.4.2	Variance component analysis of the RNA-seq data	82
3.4.3	Detecting hidden sources of variation.....	85
3.4.4	Reproducibility of differentiation	87
3.5	Variability in cell surface marker expression	89
3.6	Discussion	90
4	Genetics of gene expression in macrophage immune response	93
4.1	Introduction	93
4.2	Methods.....	95
4.2.1	Gene expression analysis.....	95
4.2.2	Gene expression QTL mapping	97
4.2.3	Alternative transcription analysis.....	99
4.2.4	Transcript ratio QTL mapping	104
4.2.5	Overlap analysis with the NHGRI-EBI GWAS catalogue.....	105
4.2.6	QTL replicability between conditions	105
4.3	Quantifying gene expression and alternative transcription.....	105
4.3.1	Differential expression analysis reveals expected pathways	107
4.4	Genetics of gene expression	110
4.4.1	Gene expression QTL mapping	110
4.4.2	Transcript ratio QTL mapping	111
4.4.3	Concordance of QTLs detected by different methods	112
4.4.4	Condition specificity of eQTLs and trQTLs.....	113
4.5	Case study: genetics of IRF5 transcription.....	117
4.6	Overlap with GWAS hits	121
4.7	Discussion	123
5	Genetics of chromatin accessibility in macrophage immune response	125
5.1	Introduction	125
5.2	Methods.....	128
5.2.1	ATAC-seq	128
5.2.2	ChIP-seq data analysis	130
5.2.3	Chromatin accessibility QTL mapping.....	133
5.3	Quantifying chromatin accessibility	135
5.3.1	Differential chromatin accessibility between conditions	136

5.3.2	Overlap with ChIP-seq signals	138
5.4	Genetics of chromatin accessibility.....	140
5.4.1	Fine mapping putative causal variants	142
5.4.2	Assessing condition-specificity of caQTLs.....	145
5.4.3	Condition-specific dependent peaks	149
5.5	Linking chromatin accessibility to the transcriptome	150
5.5.1	Linking caQTLs to eQTLs	151
5.5.2	Using caQTLs to fine map causal variants for GWAS hits	154
5.6	Discussion	155
6	Conclusions	159
6.1	Using iPSC-derived cells to map QTLs for molecular traits.....	159
6.2	Alternative transcription QTLs.....	160
6.3	Information flow from DNA to protein	162
6.4	What are we going to do with all of the QTLs?.....	163
7	References	167

