

Abbreviations

ASE	allele-specific expression
ASW	African ancestry in Southwest USA
BMI	body mass index
bp	base pairs
c-dSNP	causal disease SNP
c-eQTL	causal eQTL SNP
CD	Crohn's disease
cDNA	copy DNA
CEPH	Centre d'Étude du Polymorphisme Humain
CEU	Utah residents with Northern of Western European ancestry
CHB	Han Chinese in Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado, USA
ChIP	chromatin immunoprecipitation
cM	centimorgan
CNV	copy number variant
cRNA	copy RNA
cSNP	causal SNP
dSNP	disease SNP
DZ	dizygotic
EBV	Epstein-Barr virus
ENCODE	encyclopedia of DNA elements
eQTL	expression quantitative trait locus
EST	expressed sequence tag
FA	factor analysis
FDR	false discovery rate
GIH	Gujarati Indians in Houston, Texas, USA
GWAS	genome-wide association study/studies
IVT	in vitro transcription
JPT	Japanese in Tokyo, Japan
kb	kilobase
KCL	King's College London
LCLs	lymphoblastoid cell lines (EBV-transformed B-cells)
LCR	locus control region
LD	linkage disequilibrium
LR	linear regression
LWK	Luhya in Webuye, Kenya
MAF	minor allele frequency
Mb	megabase
MCTA	matched co-twin analysis
MEMN	macrophage enriched metabolic network
MEX	Mexican ancestry in Los Angeles, California, USA

miRNA	microRNA
MKK	Maasai in Kinyawa, Kenya
MS	multiple sclerosis
MuTHER	multiple tissue human expression resource
MZ	monozygotic
NHGRI	National Human Genome Research Institute
PCA	principal component analysis
PT	permutation threshold
QC	quality control
QTL	quantitative trait locus
RNA-seq	RNA sequencing
RTC	regulatory trait concordance
SAM	sentrix array matrix
SNP	single nucleotide polymorphism
SRC	Spearman rank correlation
SSAHA	sequence search and alignment by hashing algorithm
T2D	type 2 diabetes
TF	transcription factor
TP	true positives
TSI	Toscans in Italy
TSS	transcription start site
UGMS	University of Geneva Medical School
UTR	untranslated region
WTCCC	Wellcome Trust Case Control Consortium
WTCHG	Wellcome Trust Centre for Human Genetics
WTSI	Wellcome Trust Sanger Institute
YRI	Yoruban in Ibadan, Nigeria

List of Figures

Figure 1.1 The process of eukaryotic gene expression.....	2
Figure 1.2. The haplotype structure of the human genome	10
Figure 1.3. <i>Cis</i> and <i>trans</i> effects on transcript levels.	13
Figure 1.4. Different trichome patterns among <i>Drosophila</i> species.....	18
Figure 1.5. Geographic distribution of lactose intolerance.	19
Figure 1.6. Common disorders as quantitative traits.....	23
Figure 1.7. Interpreting GWAS results with eQTL data.....	25
Figure 2.1. Illumina II assay protocol	32
Figure 2.2. Direct hybridization assay overview and workflow.....	34
Figure 2.3. SNP-gene association example	38
Figure 3.1. Excess of regulatory variants among GWAS signals.....	49
Figure 3.2. <i>Cis</i> regulatory enrichment stratified by immunity relatedness	50
Figure 3.3. RTC score distribution following simulations	53
Figure 3.4. Properties of the RTC score when varying r^2	54
Figure 3.5. Properties of the RTC score when varying D'	55
Figure 3.6. Properties of the RTC score when varying the median r^2 of the hotspot interval.....	56
Figure 3.7. Overrepresentation of immunity-related high RTC scoring <i>cis</i> signals	60
Figure 3.8. The RTC method compared to standard LD measurements in the observed data.....	64
Figure 3.9. The fraction of eQTL variance explained away by the dSNP versus the RTC score.....	65
Figure 4.1. P-value distribution of <i>cis</i> eQTLs (10^{-3} PT) after FA correction in the uncorrected data.....	74
Figure 4.2. Proportion of tissue shared and tissue-specific eQTLs from the SRC and SRC-FA analysis....	75
Figure 4.3. Shared eQTLs (10^{-2} PT, SRC) have the same direction of effect (SRC rho) across tissues	78
Figure 4.4. Cumulative SRC rho distribution across tissues for tissue-specific and shared eQTLs	79
Figure 4.5. Distribution of independent <i>cis</i> eQTLs (10^{-3} PT, SRC) around the transcription start site... 81	
Figure 4.6. Fold change within twins and across tissues for LCL eQTLs (10^{-3} PT, SRC)	84
Figure 5.1. Distribution of RTC scores across tissues (GenCord)	89
Figure 5.2. Distribution of RTC scores across tissues for shared intervals-disease combinations (GenCord).....	89

List of Tables

Table 2.1. Summary of resources (samples, SNPs and transcripts) used throughout the thesis.	29
Table 2.2. Factor analysis weight and noise prior values applied to each tissue. Analysis performed on MuTHER pilot samples.	46
Table 3.1. Candidate <i>cis</i> results. Candidate genes (RTC score ≥ 0.9) for <i>cis</i> regulatory mediated GWAS effects. RTC applied on 976 GWAS SNPs from NHGRI and HapMap 3 CEU expression data in LCLs. The higher the score, the more likely it is that the GWAS SNP and the eQTL for the gene shown are tagging the same functional variant.....	58
Table 3.2. Candidate <i>trans</i> results. Candidate <i>trans</i> genes likely involved in the same biological pathways, relevant to the GWAS SNPs (GWAS SNP and the genes it affects in <i>trans</i> often reside on different chromosomes, as indicated in the SNP Chr and Genes Chr fields respectively). Signals related to the same hotspot interval separated by a horizontal line. Regulatory <i>trans</i> effects RTC applied in <i>trans</i> on 976 GWAS SNPs from NHGRI and HapMap 3 CEU expression data in LCLs. Table contains only the confident results (RTC Score ≥ 0.9) for the six immunity related intervals.	61
Table 3.3. Hotspot intervals with overlapping <i>cis</i> and <i>trans</i> effects as indicated by the high RTC score. Candidate regulatory effects explaining GWAS signals were detected for two of the 15 intervals tested for both <i>cis</i> and <i>trans</i> effects.	63
Table 4.1. Successfully genotyped twin pairs (MZ and DZ) with available gene expression data. Number of twin pairs per tissue sharing both genotypic and expression information.	69
Table 4.2. <i>Cis</i> eQTL associations detected with SRC analysis. Significant discoveries (number of genes with eQTLs) are shown at different permutation thresholds for each tissue. Within each tissue, two independent eQTL analyses were performed after separating related individuals in two subsets (Twin1, Twin2).....	71
Table 4.3. Replication of <i>cis</i> eQTL discoveries (number of significant genes per tissue at 10^{-3} permutation threshold). Results from both the Spearman Rank Correlation (SRC) and Factor Analysis (SRC-FA) are presented. Proportion of replicating signals calculated as the mean co-twin π_1 estimates from the p-value distribution of same SNP-gene associations in the reciprocal twin set.....	72
Table 4.4. <i>Cis</i> eQTL associations detected with SRC-FA analysis. Number of genes with a significant eQTL is shown for each co-twin analysis per tissue at different permutation thresholds.	73
Table 4.5. Recovery of SRC eQTLs (10^{-3} PT gene associations) with factor analysis correction. In each tissue and for both co-twins, 80-90% of eQTLs detected before correction (standard analysis - Std) are recovered with SRC-FA.....	73
Table 4.6. Tissue-shared and tissue-specific gene associations (10^{-3} PT), SRC analysis	76
Table 4.7. Tissue-shared and tissue-specific gene associations (10^{-3} PT), SRC-FA analysis.	77
Table 4.8. Tissue-shared and tissue-specific interval-gene associations (10^{-3} PT), SRC analysis	80
Table 4.9. Continuous estimates of tissue sharing by enrichment of low p-values (π_1) of reference eQTLs (SNP-genes 10^{-3} PT) in the secondary tissues.	82

Table 5.1. Number of nonredundant recombination hotspot intervals with co-localizing GenCord eQTLs and GWAS SNPs. Unique count of intervals (#Intervals) and interval-disease combinations (#Intervals with multiple disease GWAS loci - #Interval-diseases) per tissue tested for causal regulatory effects with the RTC. A subset of intervals (Shared) harbour eQTLs detected in all three tissues..... **88**

Table 5.2. Tissue shared and tissue-specific interval-disease combinations with high RTC score. RTC results with a score ≥ 0.9 are overlapped and compared across tissues. The significant tissue-specific component (~70% of results are found in one tissue only) implies that predicting causal regulatory effects for GWAS loci is highly dependent on the tissue where expression is determined..... **91**

Table 5.3. Candidate B-cell results. Candidate genes (RTC score ≥ 0.9) for *cis* regulatory mediated GWAS effects. RTC applied on NHGRI GWAS SNPs and B-cell expression data from the GenCord. **92**

Table 5.4. Candidate T-cell results. Candidate genes (RTC score ≥ 0.9) for *cis* regulatory mediated GWAS effects. RTC applied on NHGRI GWAS SNPs and T-cell expression data from the GenCord. **95**

Table 5.5. Candidate fibroblast results. Candidate genes (RTC score ≥ 0.9) for *cis* regulatory mediated GWAS effects. RTC applied on NHGRI GWAS SNPs and fibroblast expression data from the GenCord..... **99**

Appendix

1. Biopsy technique protocol

1. The lower abdominal biopsy site is cleaned. Local anesthetic with adrenaline is infiltrated into the pre-inked skin.
2. Stretching the skin perpendicular to the relaxed skin tension lines between thumb and finger either side of the area to be sampled, the punch blade is placed on the skin and rotated under gentle pressure by rolling it between the thumb and finger using a twisting drilling action.
3. One should penetrate to the level of the fat layer to achieve a full thickness skin biopsy specimen. The specimen should be weighed, cut in half, and stored immediately in liquid nitrogen
4. The specimen will either float up on the fat layer or can be gently lifted using a skin hook or gently applied forceps to allow specimen collection by cutting through the fat layer using a scalpel or sharp scissors.
5. Further fat can be obtained by careful dissection of the fat layer using forceps and scalpel. The fat sample should be weighed and immediately stored in liquid nitrogen.
6. Haemostasis can be achieved with direct pressure and/or interrupted sutures. Both absorbent and non-absorbent sutures can be used in a layered closure for larger punch defects.
7. The resultant defect can be allowed to heal by secondary intention as an alternative method although optimal haemostasis and cosmesis as well as reduced healing time are usually seen with sutured wounds.

