GENETICS OF GENE EXPRESSION IN PRIMARY IMMUNE CELLS IDENTIFIES CELL-SPECIFIC MASTER REGULATORS AND ROLES OF HLA ALLELES

Benjamin P Fairfax, Seiko Makino, Jayachandran Radhakrishnan, Katharine Plant, Stephen Leslie, Alexander Dilthey, Peter Ellis, Cordelia Langford, Fredrik O Vannberg, Julian C Knight

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY NOTE

Examples of genes showing significant eQTL that involve reported GWAS markers for three autoimmune conditions highlighted in Figure 6

Ulcerative colitis (UC)

17.6% of reported GWAS markers for UC from the Catalog of Published Genome-Wide Association Studies (www.genome.gov/GWA studies) (accessed 10th September 2011) involve significant cis-eQTL (24 genes) of which 54% are monocyte specific. In monocytes these include the transcription factor gene *ETS2* rs2836878 (p_{mono} $1.4x10^{-5}$) (p_{GWAS} $2x10^{-22}$)[1], an important mediator of inflammation implicated in animal models of colitis[2]; and *LSP1* rs907611 (p_{mono} $8.9x10^{-14}$) (p_{GWAS} $1x10^{-10}$)[1] which regulates leukocyte recruitment to sites of inflammation[3]. In B-cells, eQTL include *FCGR2B* rs10800309 ($p_{B\text{-}cell}$ $6.6x10^{-4}$) (p_{GWAS} $3x10^{-9}$)[4] critical to phagocytosis of immune complexes. Shared eQTL include the zinc finger protein gene *ZFP90* rs3203684 ($p_{B\text{-}cell}$ $6.1x10^{-31}$) (p_{mono} $1.7x10^{-24}$) (r^2 1.0 with rs1728785, p_{GWAS} $3x10^{-8}$)[5]; *DAP* encoding death-associated protein rs267939 ($p_{B\text{-}cell}$ $4.2x10^{-8}$, p_{mono} $1.5x10^{-4}$) (p_{GWAS} $6x10^{-12}$)[5] previously reported in whole blood[6]; and a strong eQTL for *IRF5* (interferon regulatory factor 5) rs4728142 ($p_{B\text{-}cell}$ $4.1x10^{-20}$, $p_{monocyte}$ $4.4x10^{-16}$) (p_{GWAS} $2x10^{-8}$)[1] which resolves the weak potential association previously reported in PBMC[7] and validates the eQTL seen in most LCL datasets[8].

Systemic lupus erythematosus (SLE)

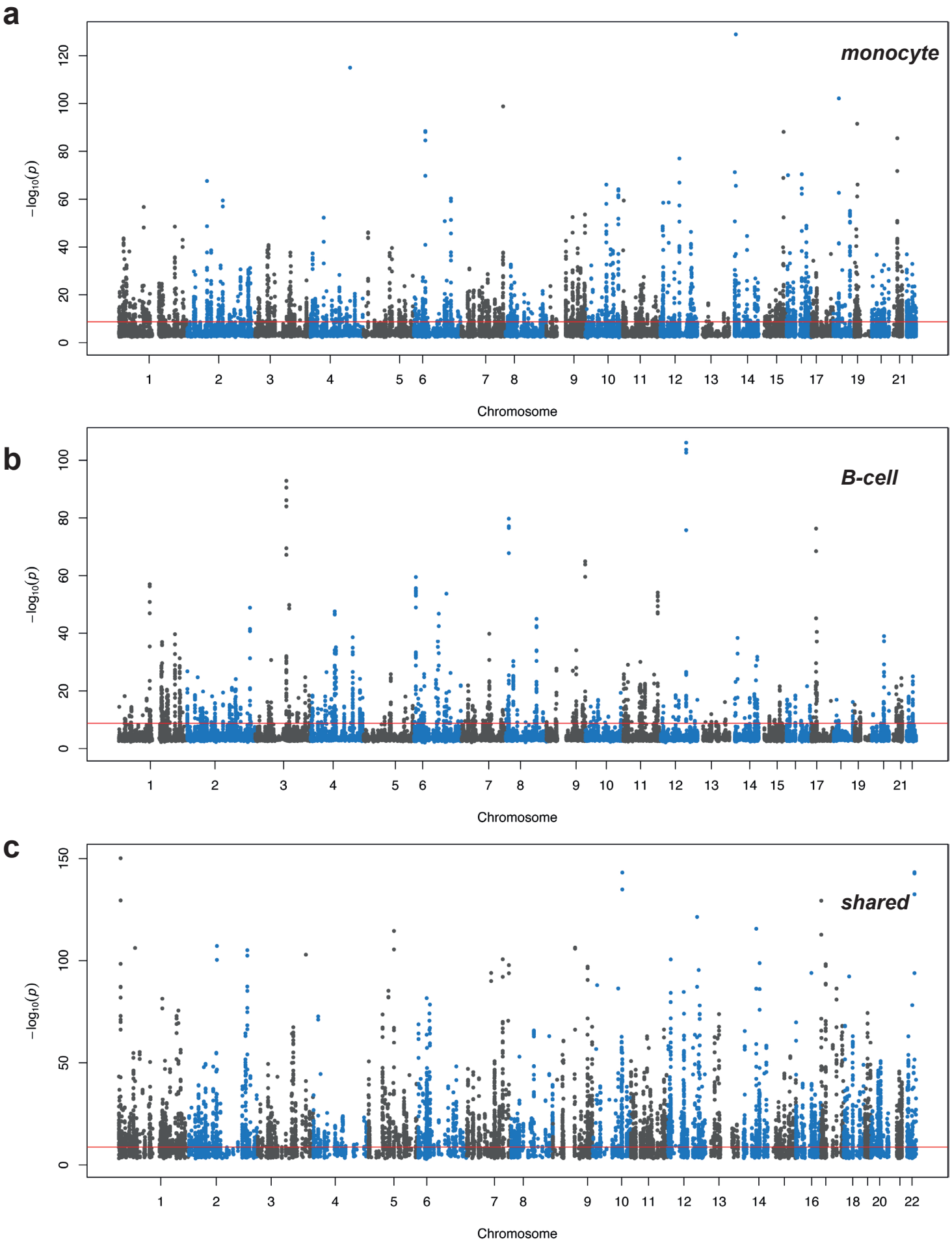
16.4% of reported GWAS markers for SLE associate with significant cis-eQTL (16 genes), 62.5% are B-cell specific. The latter include rs13277113 for *BLK* (encoding B lymphoid tyrosine kinase) ($p_{B\text{-}cell}$ $3.6x10^{-14}$); *FAM167A* ($p_{B\text{-}cell}$ $1.3x10^{-80}$) (p_{GWAS} $1x10^{-10}$)[70,71] consistent with reports in LCLs[9]; and an intronic SNP in *PXK* which shows a strong cis-eQTL with expression of a flanking gene *DNASE1L3* rs6445975 ($p_{B\text{-}cell}$ $1.3x10^{-5}$) (p_{GWAS} $7x10^{-9}$)[10] deficiency of which is associated with SLE in mice[11], consistent with a role for DNase in removal of DNA from nuclear antigens in sites undergoing cellular proliferation. Shared eQTL include *IRF5* rs4728142 ($p_{B\text{-}cell}$ $7.7x10^{-24}$, $p_{monocyte}$ $1.7x10^{-20}$) (p_{GWAS} $8x10^{-19}$)[12]; and *UBE2L3* (encoding a ubiquitin conjugating enzyme) rs5754217 ($p_{B\text{-}cell}$ $6.1x10^{-6}$, $p_{monocyte}$ $2.9x10^{-16}$) (p_{GWAS} $2x10^{-6}$)[13] which is also associated with several other autoimmune diseases and involves a previously validated eQTL[14].

Crohn's disease (CD)

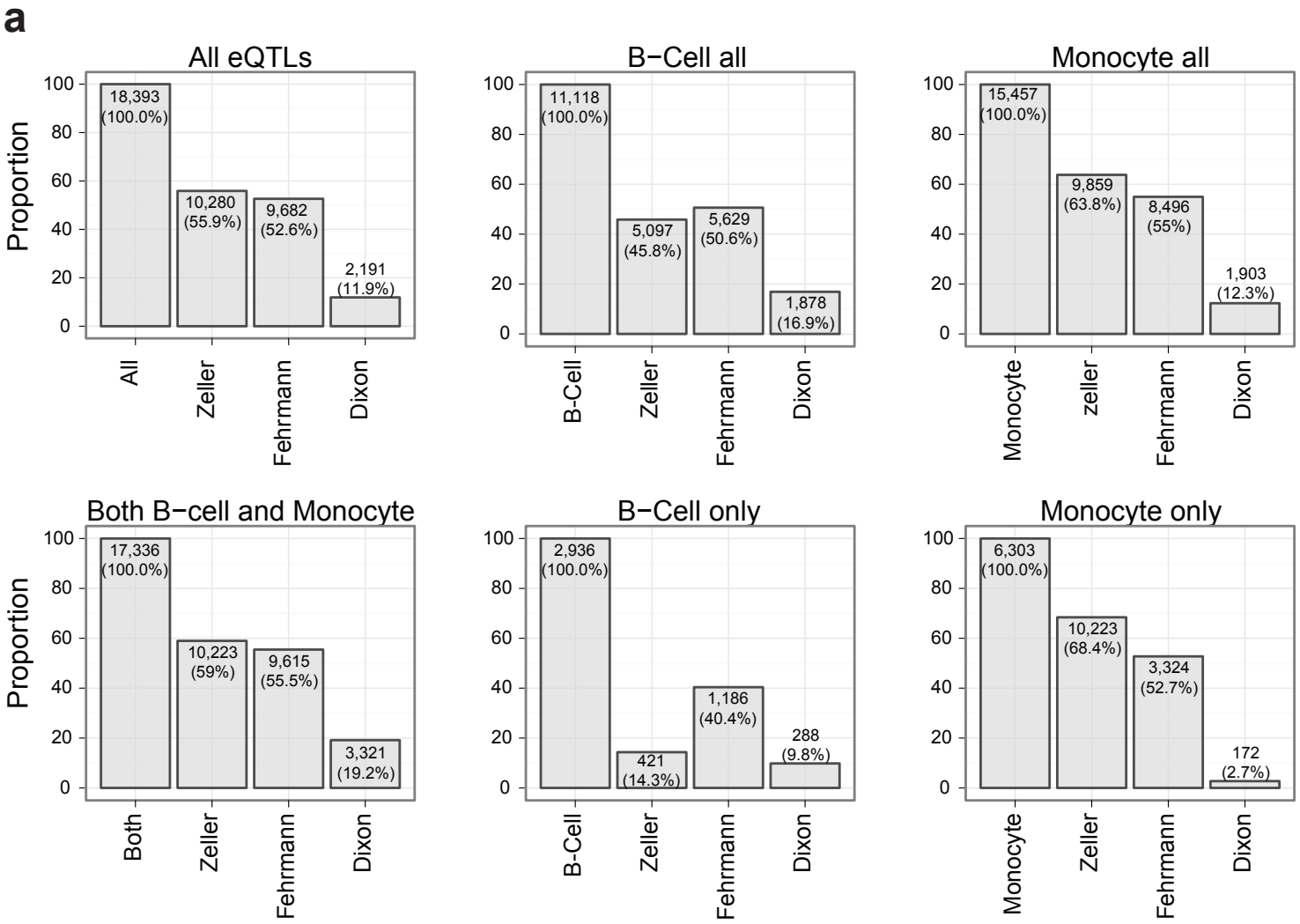
34.4% of reported GWAS SNPs for CD show significant cis-eQTL (30 genes). Monocyte specific eQTL include *CARD9* rs4077515 ($p_{monocyte}$ $1.04x10^{-24}$) (p_{GWAS} $1x10^{-36}$)[15]. A B-cell-specific eQTL for *FADS1* (encoding fatty acid desaturase 1) was associated with several traits including CD rs102275 ($p_{B\text{-}cell}$ $3.7x10^{-22}$) (p_{GWAS} $2x10^{-11}$)[15] consistent with the role of dietary fats in CD[16] and bowel ulceration in *FADS2* knockdown mice[16]. Shared eQTL include *ERAP2* rs7719705 ($p_{monocyte}$ $1.71x10^{-102}$, $p_{B\text{-}cell}$ $2.99x10^{-106}$) (r^2 0.94 rs2549794 (p_{GWAS} $1x10^{-10}$)[15] encoding an aminopeptidase involved in peptide antigen processing[17]; *SLC22A5* rs12521868 ($p_{B\text{-}cell}$ $2.8x10^{-10}$, $p_{monocyte}$ $2x10^{-30}$) (p_{GWAS} $1x10^{-20}$)[15] encoding a plasma membrane transporter protein strongly implicated in intestinal inflammation[18,19]; and *INPP5E* rs4077515 ($p_{B\text{-}cell}$ $4.3x10^{-32}$, $p_{monocyte}$ $1.2x10^{-31}$) (p_{GWAS} $1x10^{-36}$)[15] encoding an inositol polyphosphate critical to mobilising intracellular calcium, mutations of this gene are associated with ciliopathies[20].

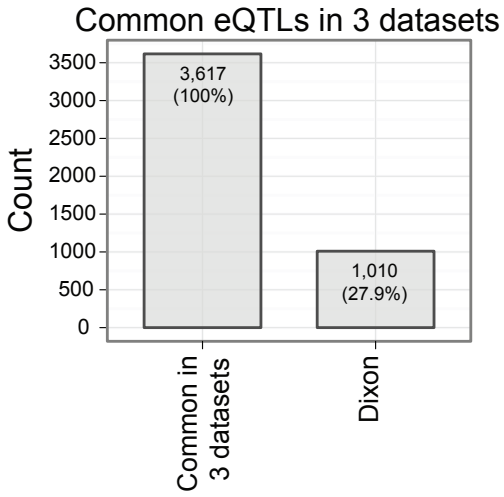
References

1. Anderson, C.A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* **43**, 246-52 (2011).
2. van der Pouw Kraan, T.C., Zwiers, A., Mulder, C.J., Kraal, G. & Bouma, G. Acute experimental colitis and human chronic inflammatory diseases share expression of inflammation-related genes with conserved Ets2 binding sites. *Inflamm Bowel Dis* **15**, 224-35 (2009).
3. Liu, L. *et al.* LSP1 is an endothelial gatekeeper of leukocyte transendothelial migration. *J Exp Med* **201**, 409-18 (2005).
4. McGovern, D.P. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet* **42**, 332-7 (2010).
5. Barrett, J.C. *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* **41**, 1330-4 (2009).
6. Dubois, P.C. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* **42**, 295-302 (2010).
7. Lofgren, S.E. *et al.* Promoter insertion/deletion in the IRF5 gene is highly associated with susceptibility to systemic lupus erythematosus in distinct populations, but exerts a modest effect on gene expression in peripheral blood mononuclear cells. *J Rheumatol* **37**, 574-8 (2010).
8. Alonso-Perez, E. *et al.* Cis-regulation of IRF5 expression is unable to fully account for systemic lupus erythematosus association: analysis of multiple experiments with lymphoblastoid cell lines. *Arthritis Res Ther* **13**, R80 (2011).
9. Hom, G. *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* **358**, 900-9 (2008).
10. Harley, J.B. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* **40**, 204-10 (2008).
11. Wilber, A., O'Connor, T.P., Lu, M.L., Karimi, A. & Schneider, M.C. Dnase1l3 deficiency in lupus-prone MRL and NZB/W F1 mice. *Clin Exp Immunol* **134**, 46-52 (2003).
12. Han, J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* **41**, 1234-7 (2009).
13. Chung, S.A. *et al.* Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet* **7**, e1001323 (2011).
14. Fransen, K. *et al.* Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. *Hum Mol Genet* **19**, 3482-8 (2010).
15. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-25 (2010).
16. Stroud, C.K. *et al.* Disruption of FADS2 gene in mice impairs male reproduction and causes dermal and intestinal ulceration. *J Lipid Res* **50**, 1870-80 (2009).
17. Andres, A.M. *et al.* Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet* **6**, e1001157 (2010).
18. Fujiya, M. *et al.* Cytokine regulation of OCTN2 expression and activity in small and large intestine. *Inflamm Bowel Dis* **17**, 907-16 (2011).
19. Kato, Y. *et al.* Gene knockout and metabolome analysis of carnitine/organic cation transporter OCTN1. *Pharm Res* **27**, 832-40 (2010).
20. Jacoby, M. *et al.* INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. *Nat Genet* **41**, 1027-31 (2009).



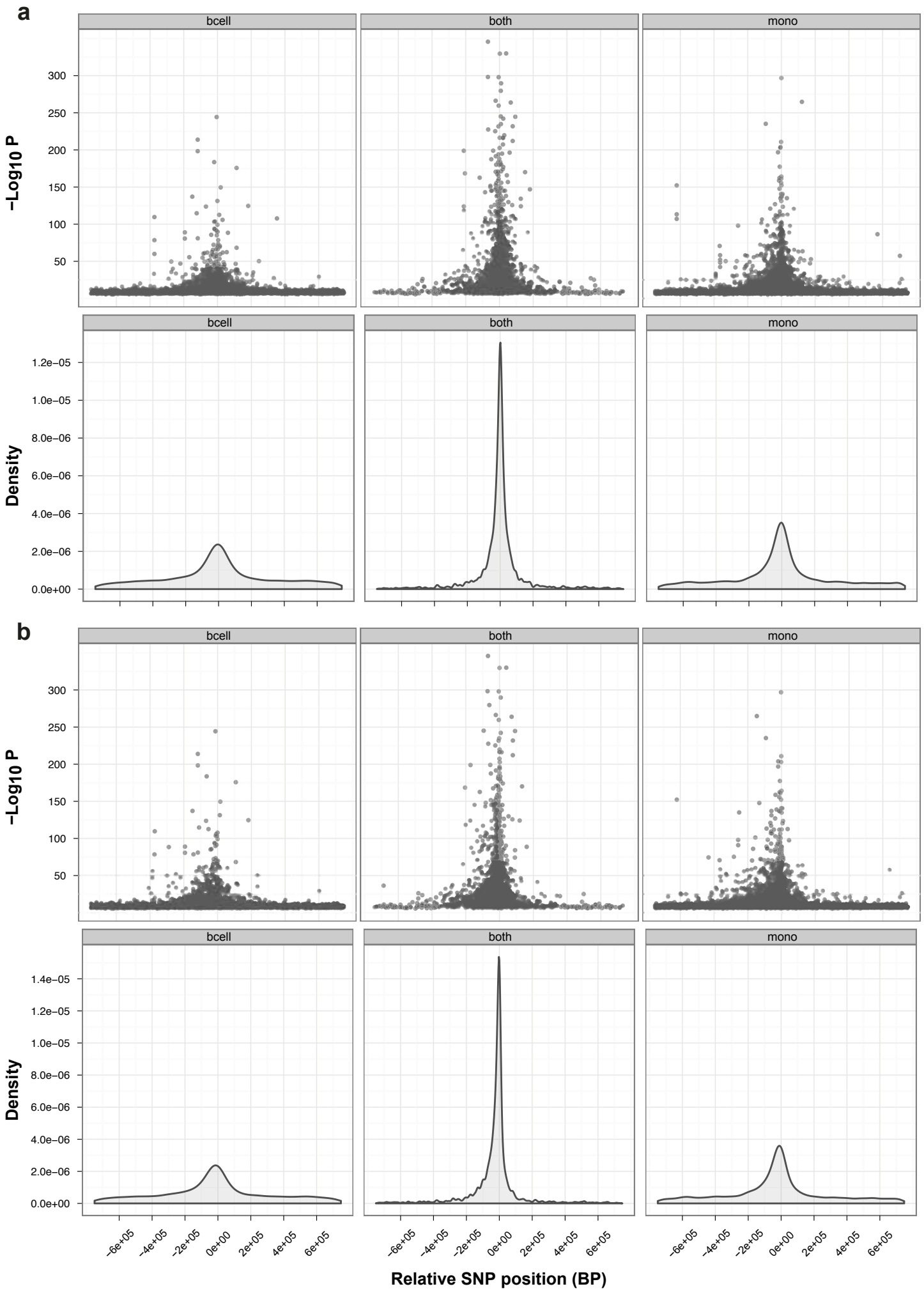
Supplementary Figure 1. Cell specificity of cis-eQTL. Manhattan plots denoting cis-associations only that (a) are unique to monocytes, (b) unique to B-cells and (c) shared between datasets. Red line indicates $p=1.2 \times 10^{-9}$.



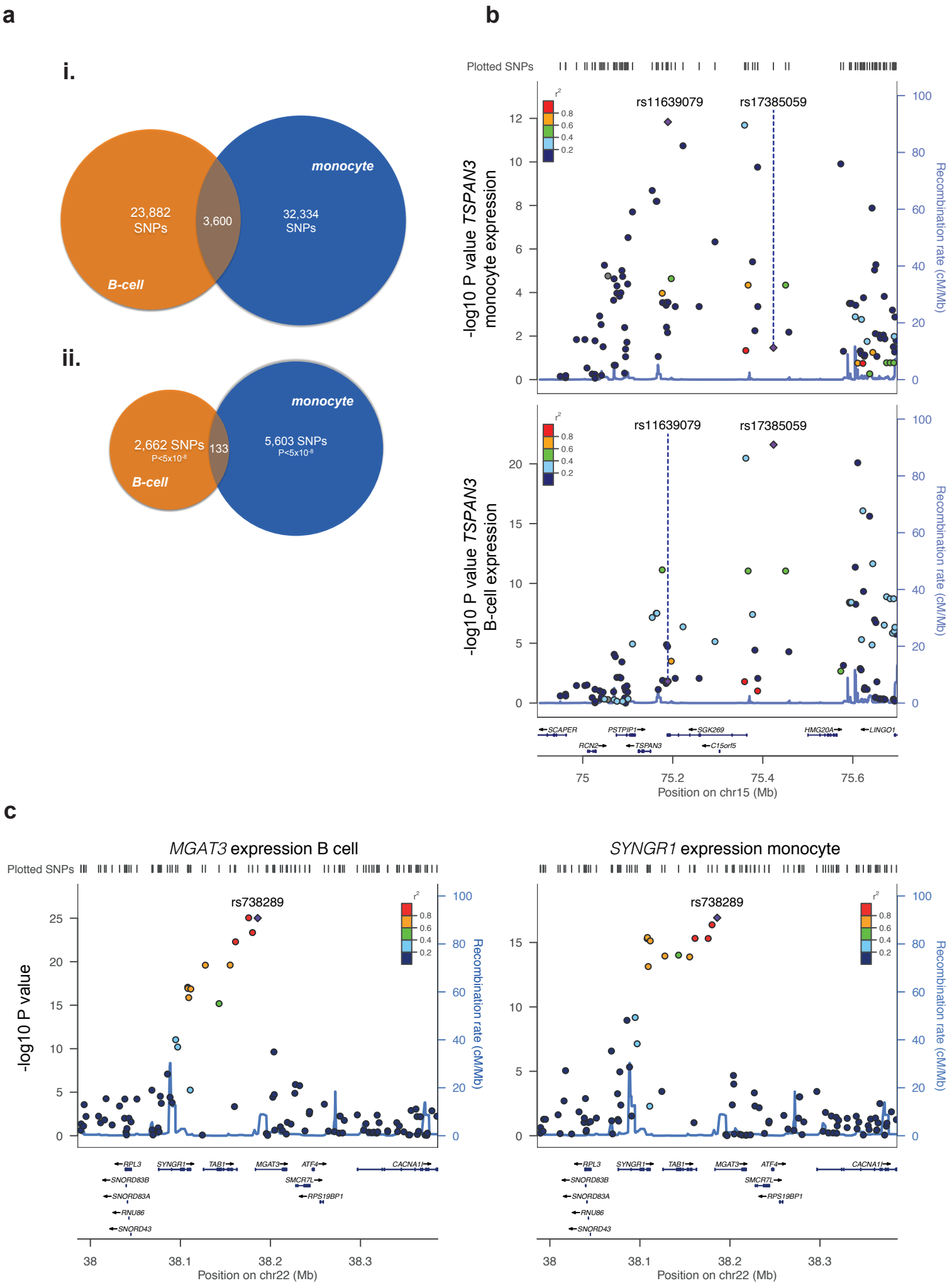


Supplementary Figure 2. Proportion of replicated eQTLs in this study to published eQTL studies. (a) Comparison of the study dataset with the publically available eQTL datasets. Zeller et al as a human primary monocyte dataset, Fehrmann et al as a human primary whole blood cell dataset, Dixon et al as a LCL dataset. The counts of matches in the SNP-gene relationships were compared with the study dataset at $p < 5x10^{-8}$. (b) Common eQTLs among our dataset, Zeller and Fehrmann datasets were compared with the LCL dataset (Dixon et al). The cutoff of the datasets were $p < 5x10^{-8}$ in the study dataset, $p < 0.05$ in Zeller and Fehrmann datasets).

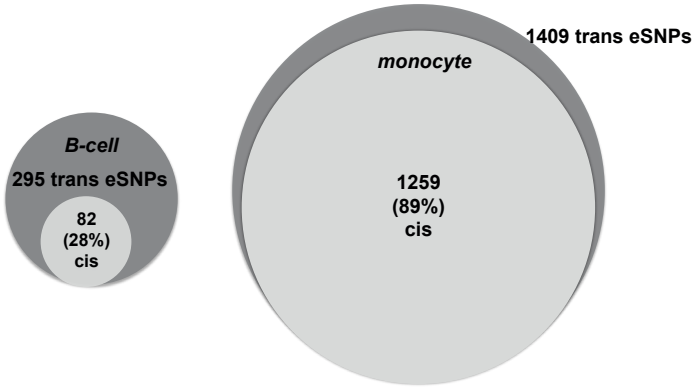
Supplementary Figure 3. Real-time PCR validates eQTL findings. Points represent individuals selected at random from homozygotes from each allele for validatory real-time PCR. Numbers per gene: *CD52* n=23; *CD40* n=15, *CD40* $_{PBMC}$ n=29; *BCL2L*13 n=16; *LITAF* n=16; *LITAF* $_{PBMC}$ n=32; *VWA5A* n=14; *BLK* n=14; *LYZ* n=14; *DFNA5* n=16; *SELL* n=21; *SELL* $_{PBMC}$ n=28. Significance calculated using Welch's modified t-test. (a-d & g) Monocyte specific eQTLs. (e,f) B-Cell specific eQTLs. (h, i) Cell-directional eQTLs.



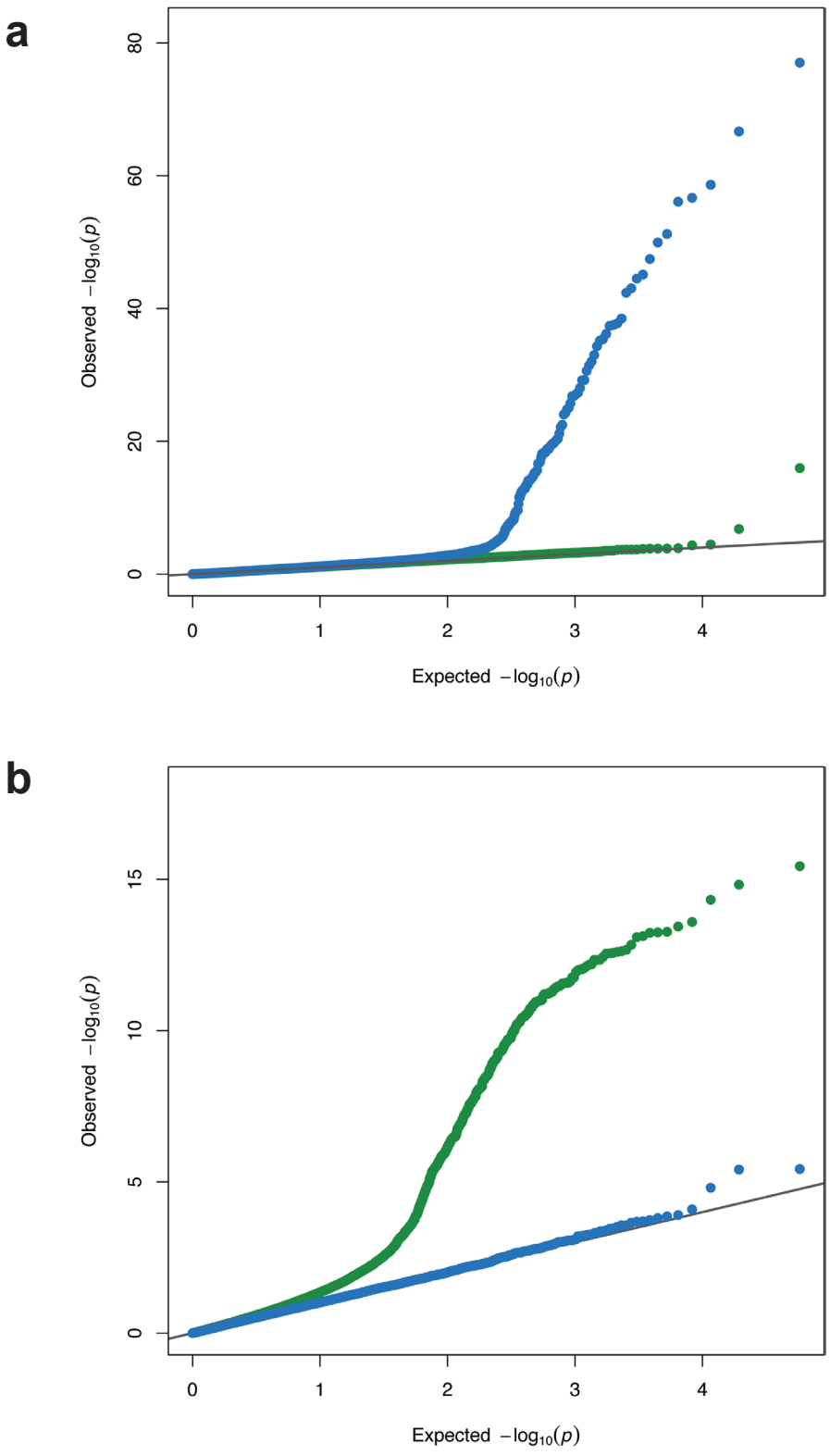
Supplementary Figure 4. Location of cell specific eSNPs with respect to genomic landscape. Cell type specificities in SNP positions relative to transcription start sites of genes (a) and (b) probes.


a
i.
monocyte
23,882 SNPs
3,600
32,334 SNPs
B-cell
ii.
monocyte
2,662 SNPs P<5x10-8
133
5,603 SNPs P<5x10-8
B-cell
b
Plotted SNPs
rs11639079
rs17385059
-log10 P value TSPAN3 monocyte expression
-log10 P value TSPAN3 B-cell expression
Recombination rate (cM/Mb)
SCAPER
PSTPIP1
SGK269
HMG20A
LINGO1
RCN2
TSPAN3
C15orf5
Position on chr15 (Mb)
c
MGAT3 expression B cell
SYNGR1 expression monocyte
Plotted SNPs
rs738289
-log10 P value
Recombination rate (cM/Mb)
RPL3
SYNGR1
TAB1
MGAT3
ATF4
CACNA1I
SNORD83B
SMCR7L
SNORD83A
RPS19BP1
RNU86
SNORD43
Position on chr22 (Mb)

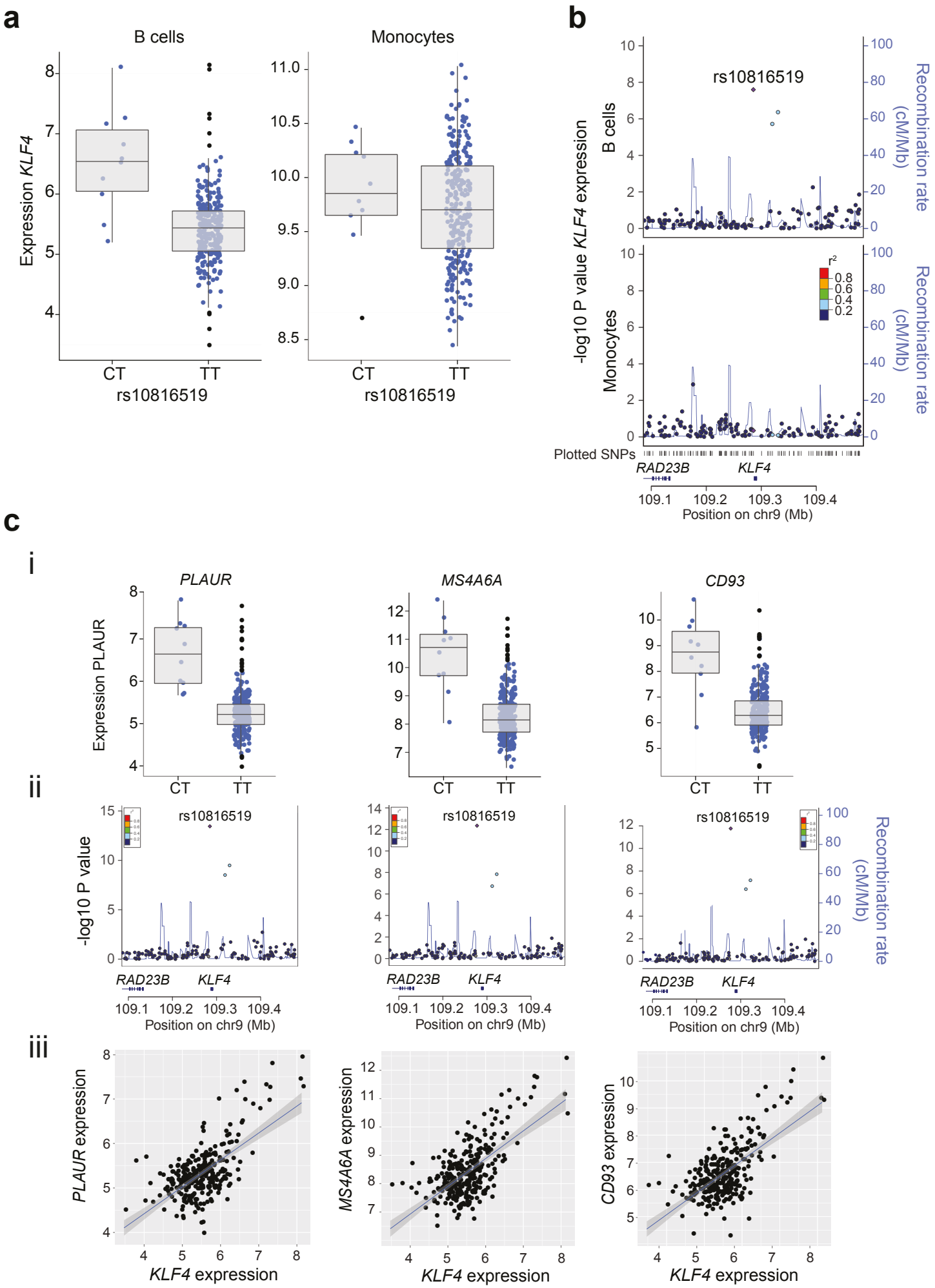
Supplementary Figure 5. The same locus may associate with expression of different genes in alternative cell types. Conversely certain genes are regulated by differing loci in a cell dependent manner (a) Venn diagrams illustrate that cell-specific eSNPs may associate with the expression of a different gene in alternative cell type (i) 3600 SNPs form eQTL to different genes at Perm p<0.001 in both datasets (ii) 133 eSNPs associate with the expression of different genes in different cell types at $p<5x10^{-8}$ in each cell type. (b) Demonstration that the eSNP for a given gene may depend upon the cell type under analysis. Here *TSPAN3* forms highly significant eQTL to different SNPs in monocytes and B-cells: rs11639079 p_{mono} 4.7×10^{-13}, $p_{B\text{-}cell}$ 0.016; rs17385059 p_{mono} 0.034, $p_{B\text{-}cell}$ 2.6×10^{-22}. (c) Local association plots for rs738289 to *MGAT3* in B-cell and *SYNGR1* in monocyte (see figure 2c for boxplots).



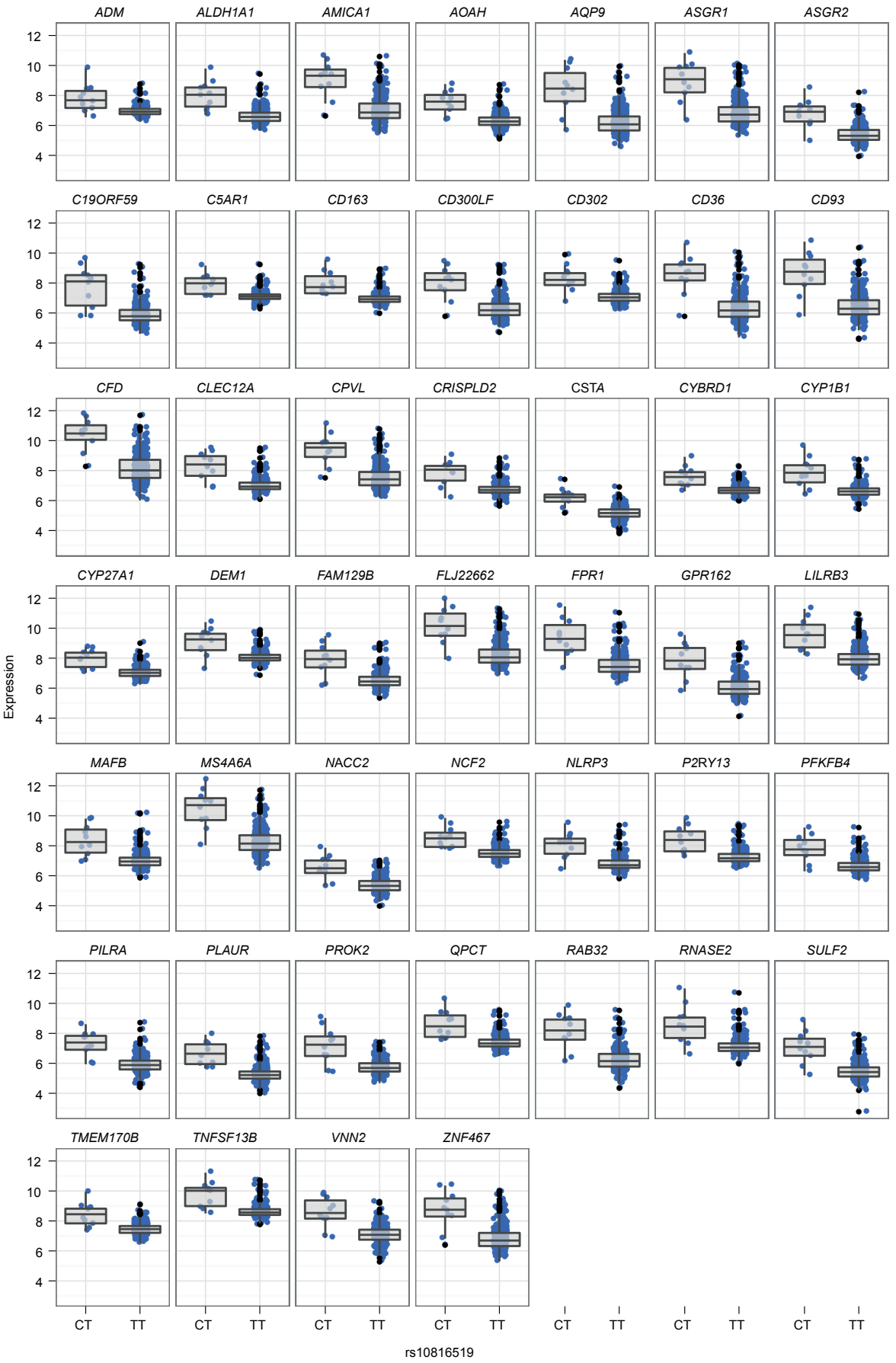
Supplementary Figure 6. Trans eSNPs frequently form cis eSNPs.
Of a total of 295 trans eSNPs in the B-cell dataset, 82 (28%) additionally form cis-eSNPs in B-cells. In the monocyte dataset this proportion is markedly different with 1259 (89%) of trans-eSNPs also forming cis-eSNP in monocytes. When Chromosome 12 is excluded (containing the *LYZ* region) this proportion falls to 77% (277/358).



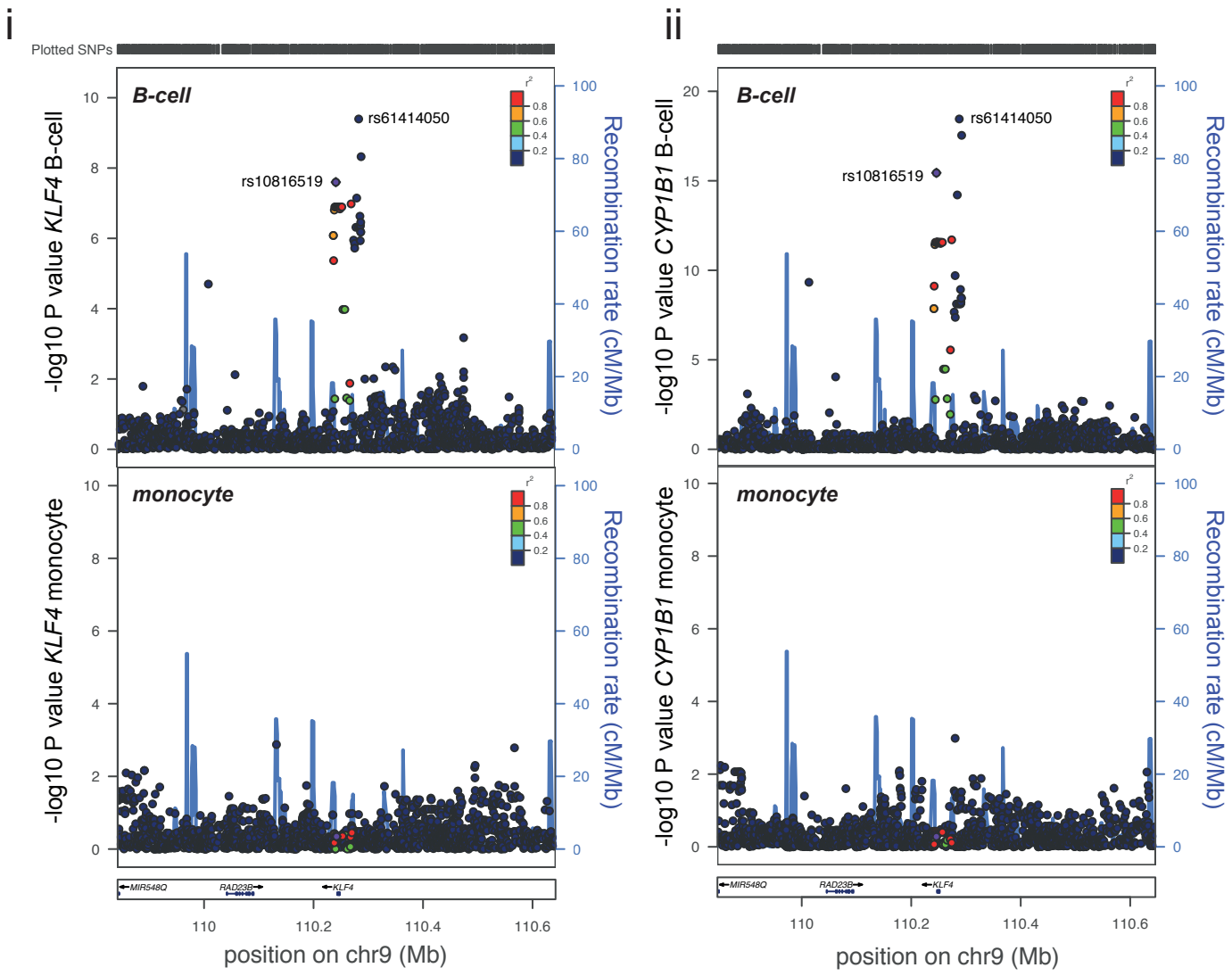
Supplementary Figure 7. Quantile-quantile plots for cell-specific multi-loci eSNPs. (a) QQ-plot of observed versus expected p-values of probe expression associated to rs10784774. Values obtained from monocytes in blue and those from B-cells in green. The two points above the observed versus expected line in B-cells correspond to *LYZ* probes. (b) QQ-plot of p-values of probe expression associated to rs10816519, a B-cell specific eQTL to the transcription factor *KLF4*. B-cell values depicted in green, monocyte blue.


a
B cells
Monocytes
Expression KLF4
CT
TT
rs10816519
b
rs10816519
B cells
Monocytes
-log10 P value KLF4 expression
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
Plotted SNPs
RAD23B
KLF4
109.1 109.2 109.3 109.4
Position on chr9 (Mb)
c
i
PLAUR
MS4A6A
CD93
Expression PLAUR
ii
-log10 P value
iii
PLAUR expression
MS4A6A expression
CD93 expression
KLF4 expression

d


ADM
ALDH1A1
AMICA1
AOAH
AQP9
ASGR1
ASGR2
C19ORF59
C5AR1
CD163
CD300LF
CD302
CD36
CD93
CFD
CLEC12A
CPVL
CRISPLD2
CSTA
CYBRD1
CYP1B1
CYP27A1
DEM1
FAM129B
FLJ22662
FPR1
GPR162
LILRB3
MAFB
MS4A6A
NACC2
NCF2
NLRP3
P2RY13
PFKFB4
PILRA
PLAUR
PROK2
QPCT
RAB32
RNASE2
SULF2
TMEM170B
TNFSF13B
VNN2
ZNF467
Expression
12
10
8
6
4
CT
TT
rs10816519

e



Supplementary Figure 8. rs10816519, a cis-eQTL to *KLF4,* forms a B-cell specific multi-loci trans-eSNP

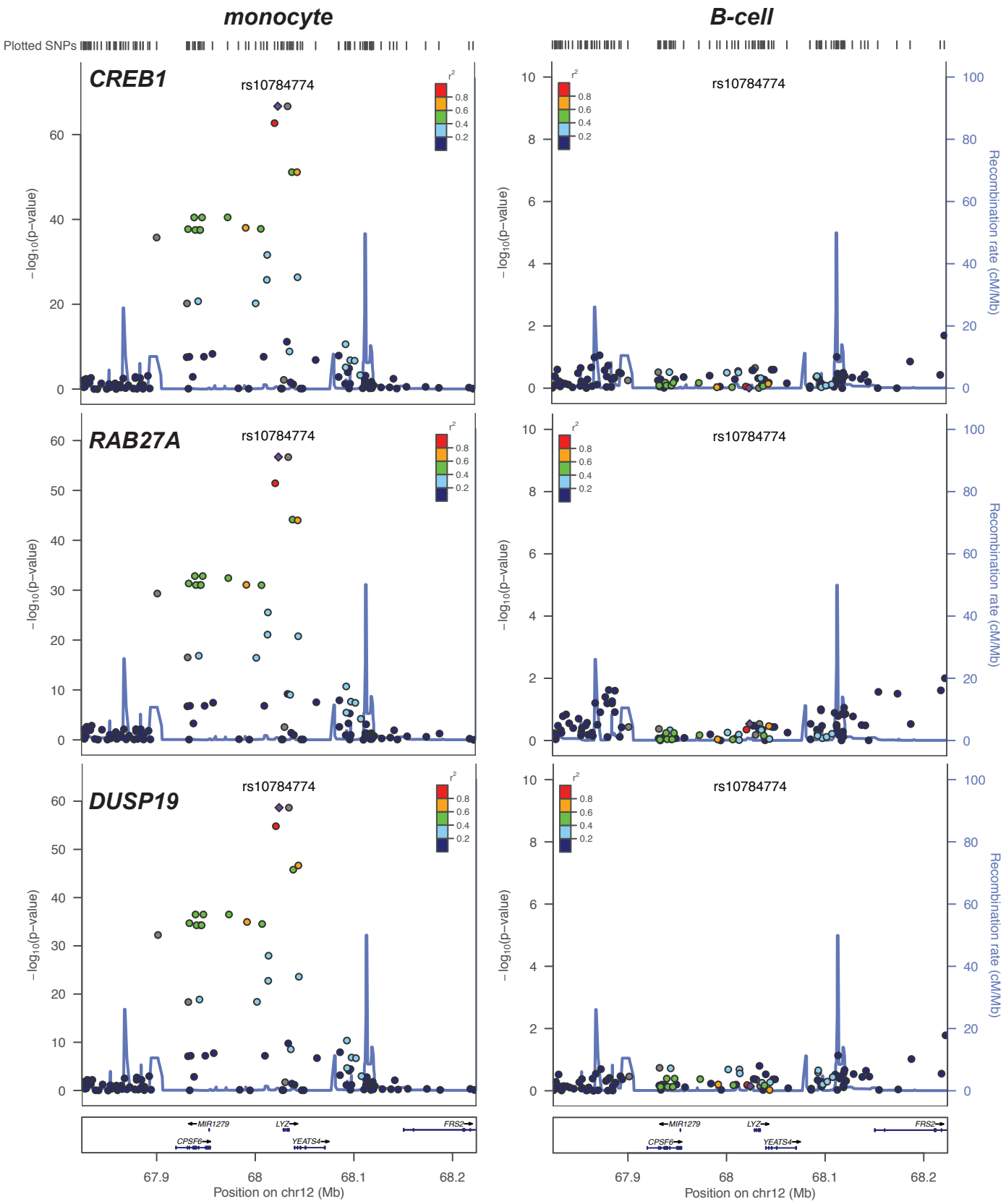
a) Heterozygotes at rs10816519 have a significantly increased expression of *KLF4* in B-cells only (CT vs. TT; $P_{B\text{-}cell}$= 0.00011, $P_{monocyte}$= 0.327, Kruskal-Wallis non-parametric ANOVA)

b) Linkage disequilibrium plots of markers associated with expression at *KLF4*, upper panel- B-cells, lower panel - monocytes. Plotted using LocusZoom using hg18 and HapMap CEU for recombination.

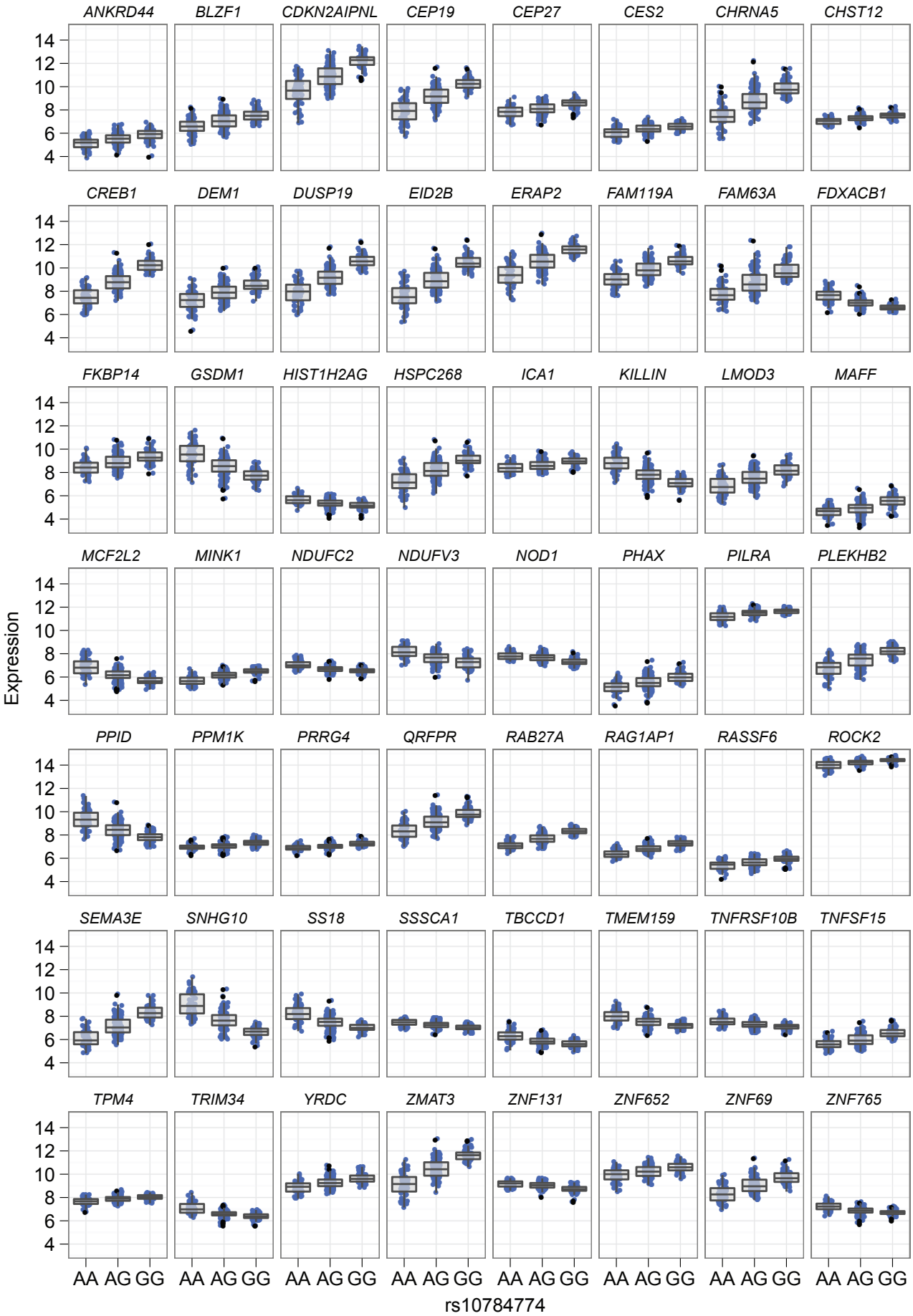
c) Plots of the three most significantly associated loci in trans to rs10816519 (i) *PLAUR* has previously been demonstrated to be downregulated in *KLF4* knockout mice and has a *KLF4* binding site in it's promoter. Of note, the expression of 3 separate probes mapping to *MS4A6A*, a gene encoding a four-transmembrane protein associated with Factor VII levels and Alzheimer's susceptibility, was associated to rs10816519. (ii) Linkage disequilibrium plots for these three genes- as per (b). (iii) B-cell *KLF4* expression is significantly associated with expression of these genes across the cohort (Spearman's rho KLF~PLAUR =0.46, KLF~CD93 =0.48, KLF~MS4A6A=0.50, $P<2.2x10^{-16}$ all three tests).

d) Boxplots demonstrating expression of all 46 genes with expression values associated to rs10816519. Where more than 1 probe is found per gene, the most significant is plotted.

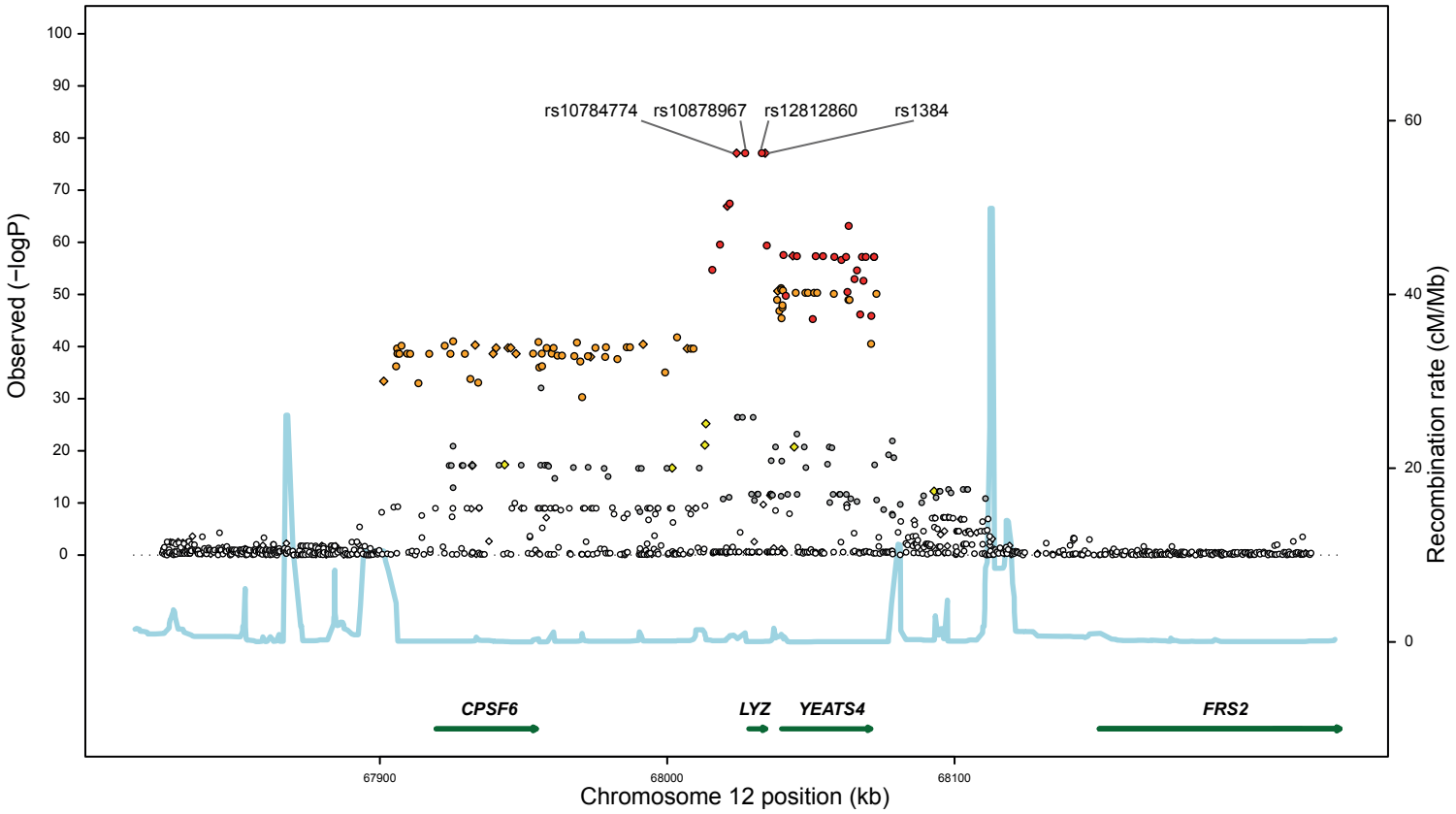
e) Imputation of *KLF4* locus using 1000 Genome data (EUR Nov. 2010) demonstrates a more significant association to *KLF4* expression at rs61414050, a SNP 5' to the gene (i) Similarly, this locus again demonstrates significance with the expression of trans associated probes as depicted in this local association plot for the gene *CYP1B1* to the region (ii) Again these associations are exclusive to B-cells.



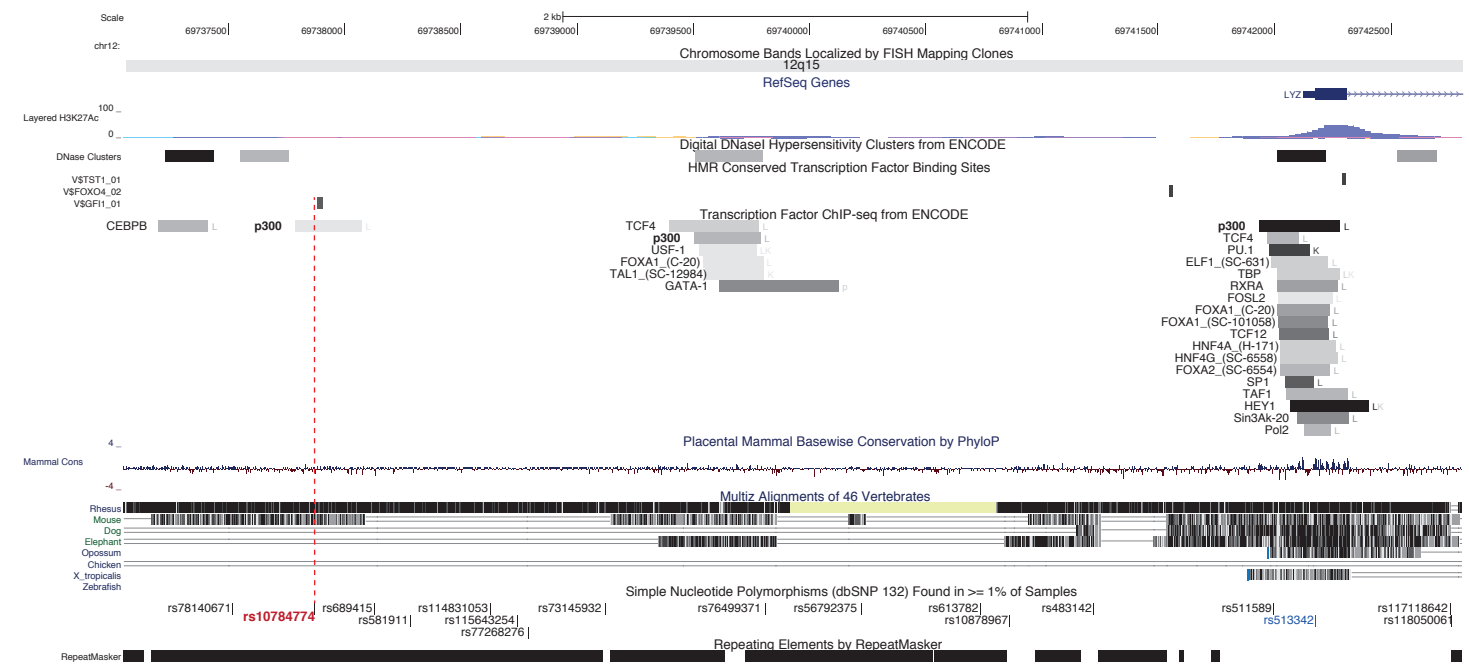
Supplementary Figure 9. Absence of trans-association in B-cells to rs10784774. Regional association plots for the most significantly associated genes in trans to the *LYZ* locus. It is apparent that whilst there is significant association in monocytes, there is none in B-cells.

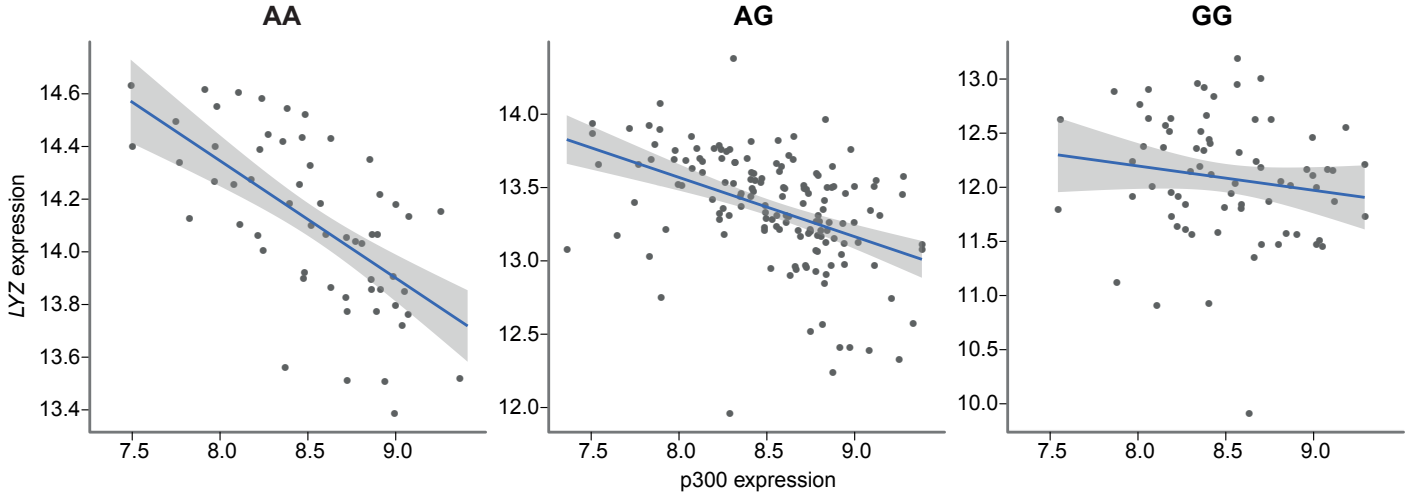


Supplementary Figure 10. Genes with regulation of expression in trans to rs10784774. A total of 72 probes demonstrate association in trans to rs10784774 at the significance threshold. This figure illustrates 56 of their corresponding genes.

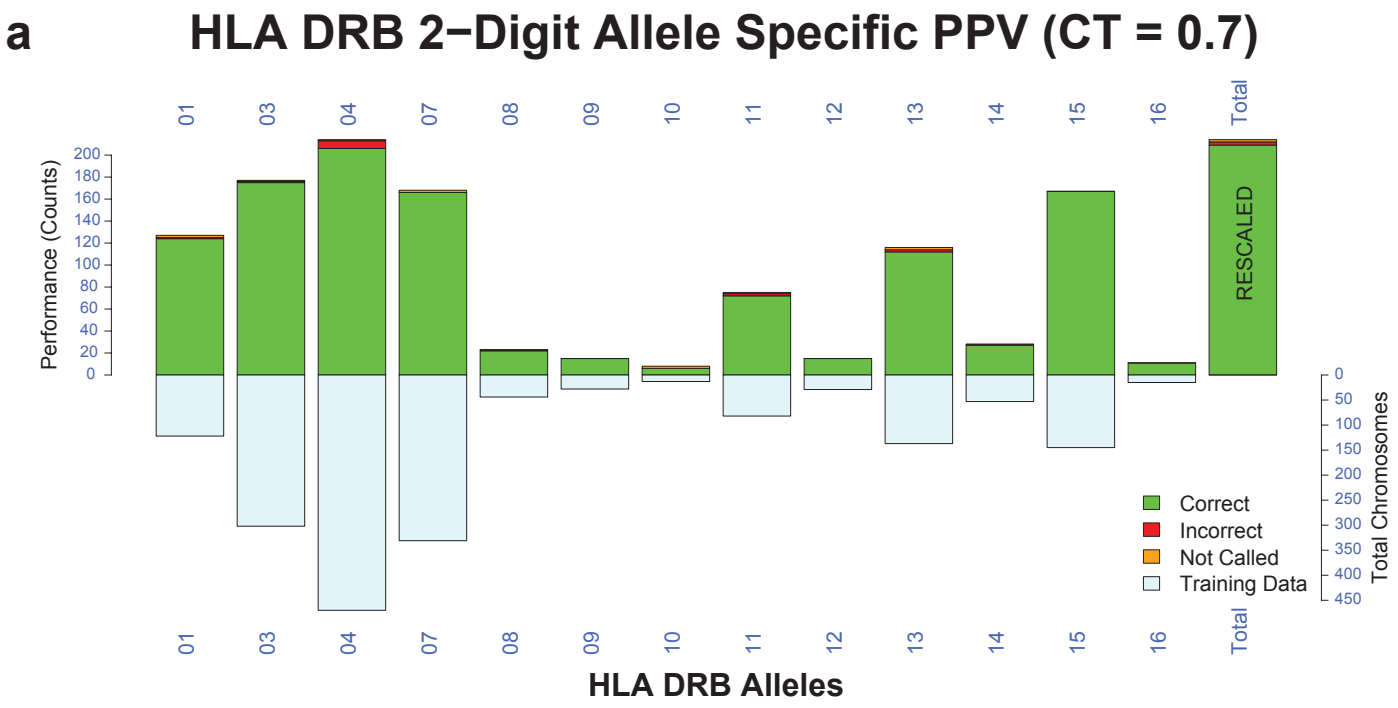


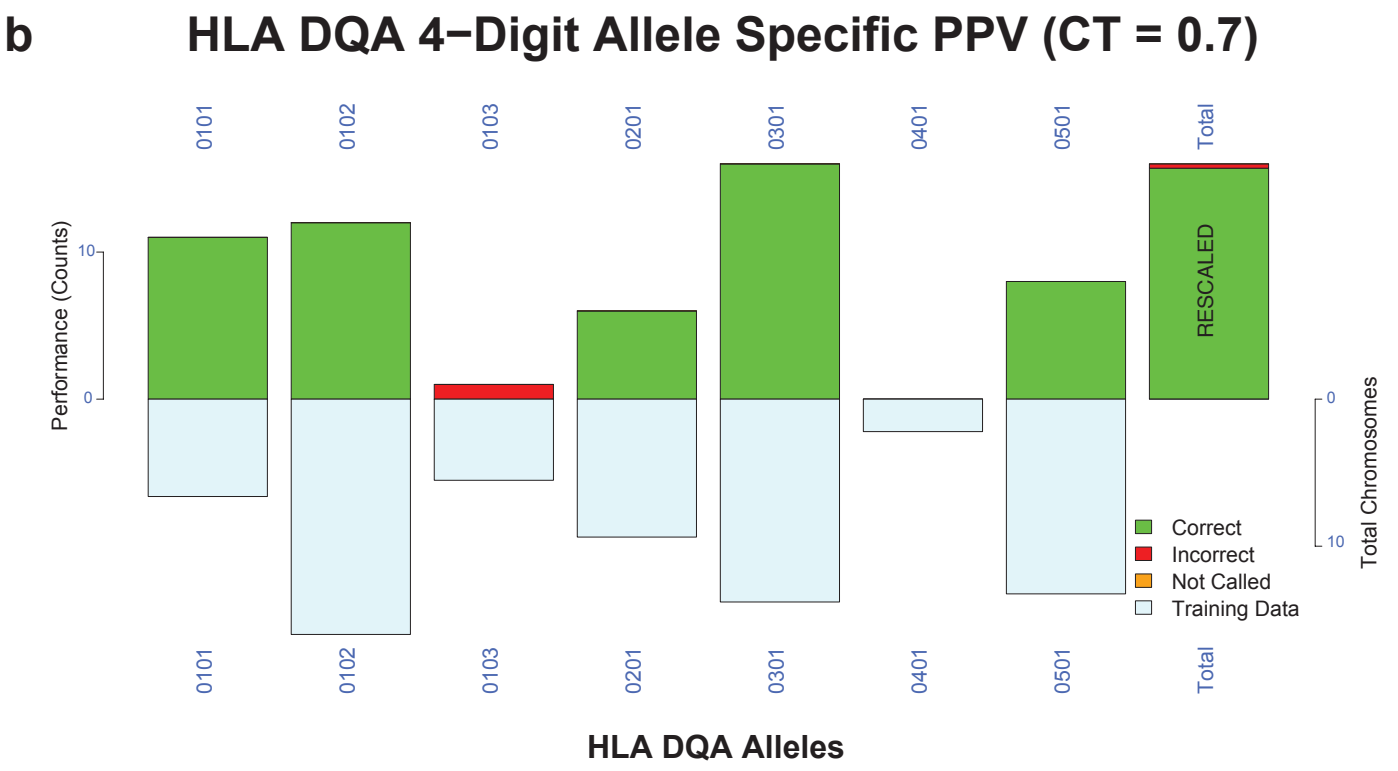
Supplementary Figure 11. Linkage-disequilibrium plot of *LYZ* expression locus.
Plot illustrates both genotyped (diamonds) and imputed (circles) markers associated with expression of *LYZ*. The most significant SNP is rs10784774 and is in complete linkage disequilibrium with 3 other SNPs across *LYZ*. This haplotype is separately associated to the *trans* expression of 67 further genes.
r^2 scale: >0.8, red; 0.5-0.8, orange; 0.2-0.5, yellow (genotyped), grey (imputed); <0.2 white.
Mapped using genome build hg18. Recombination rate from CEU HapMap data.



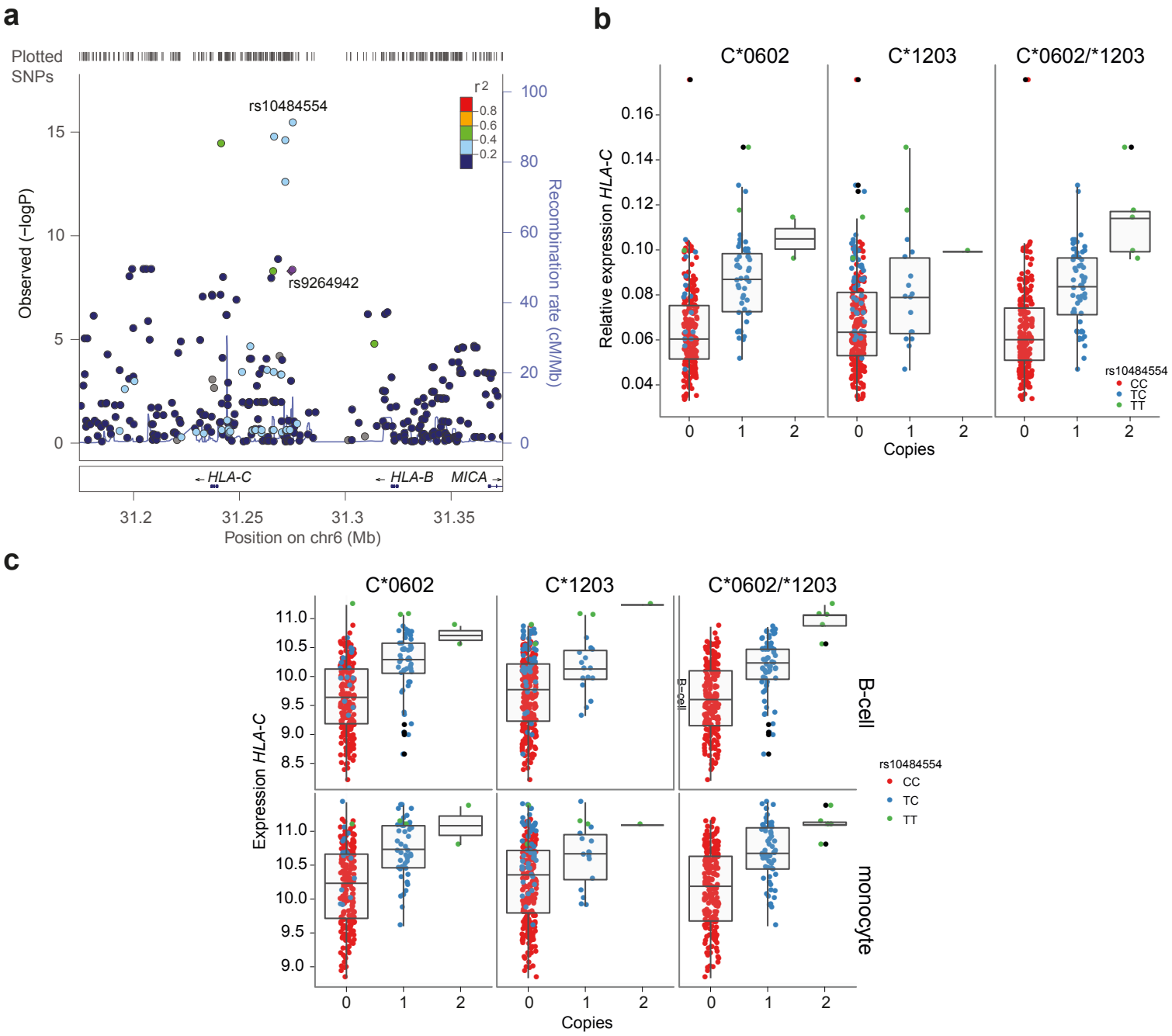


Supplementary Figure 12. Expression of p300 regulates the expression of *LYZ* in an allele specific manner. (a) UCSC screenshot indicating sites of p300 binding across the *LYZ* promoter from ChIP-seq datasets. Note multiple p300 binding sites, including one overlying rs10784774. (b) Linear regression of basal *LYZ* expression vs. p300 expression according to rs10784774 genotype. Increasing expression of p300 is associated with a significant inhibition of *LYZ* expression in individuals carrying the ancestral A allele for rs10784774 only.
AA: r^2 0.41 $p=4.4\times10^{-8}$, AG: r^2 0.22 $p=1.4\times10^{-9}$, GG: r^2 0.02 $p=0.11$





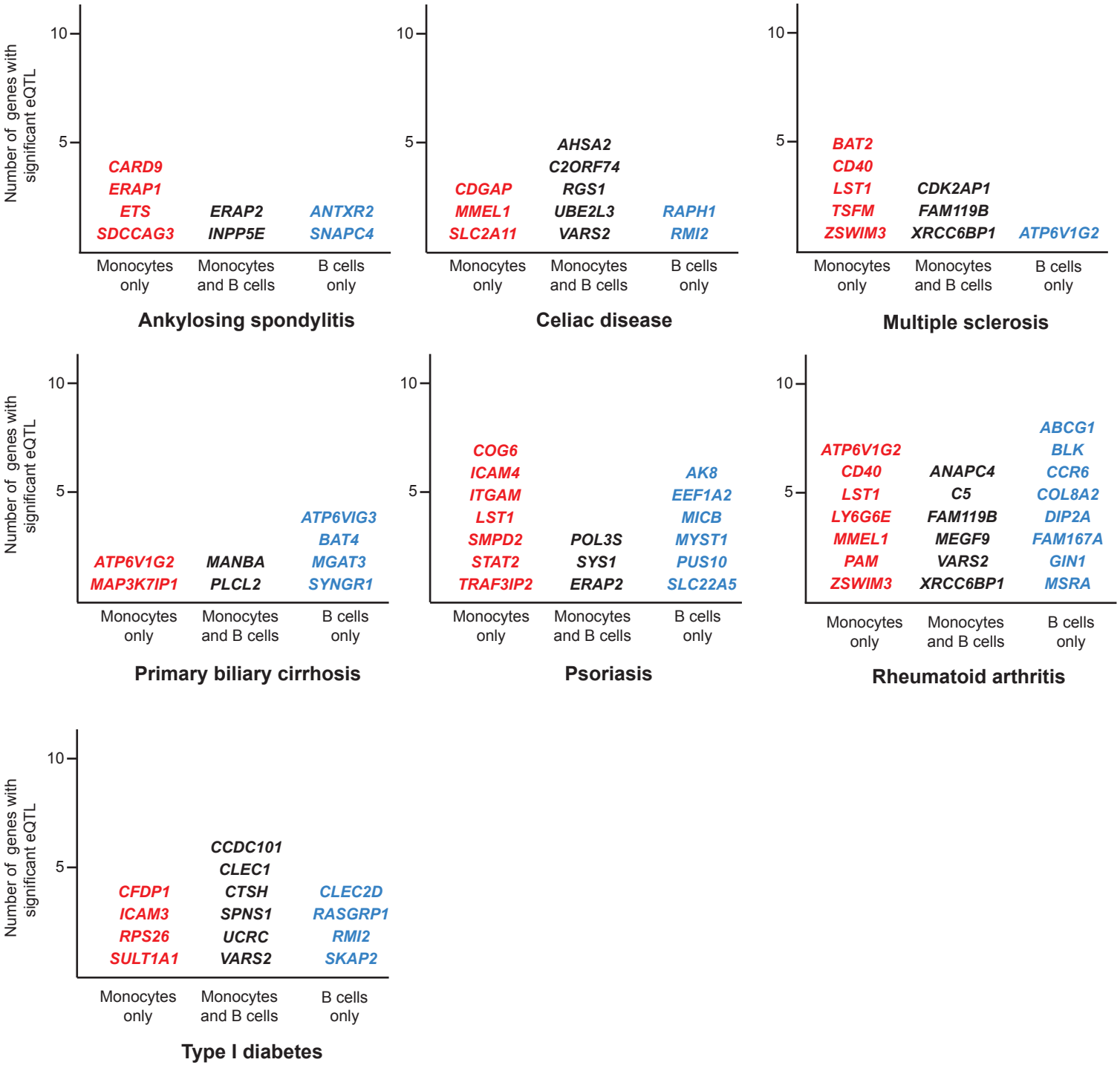
Supplementary Figure 13. Validation of imputation on HLA types. PPV: positive predicted value. CT: cutoff threshold. (a) HLA DRB 2-digit imputation. (b) HLA DQA 4-digit imputation.



Supplementary Figure 14. HLA imputation is informative for MHC cis associations. Primers designed to anneal to SNP free regions, non-overlapping with probe annealing site, were used to amplify *HLA-C* from PBMC derived cDNA (a) Local association plot showing cis-eQTL with previously reported eSNP rs9264942. The strongest association we observe is with rs10484554. (b) Imputation demonstrates that the minor allele of rs10484554 is only found in association with carriage of either C*0602 or C*1203. Carriage of C*0602 or C*1203 is associated with increased expression of *HLA-C* (C*0602, P=2.5x10^{-14}, C*1203, P=0.008, C*0602/*1203, P<2.2x10^{-16}, one-way ANOVA). (c) In B-cells and monocytes according to the array data carriage of C*0602 or C*1203 is also associated with increased expression of *HLA-C* ($P_{B\text{-}cell}$=1.3x10^{-13}, $P_{monocyte}$=3.2x10^{-12}, one-way ANOVA).


a
B cell
Celiac disease
Crohn's disease
AHSA2
Ulcerative colitis
Tuberculosis
Monocyte
Systemic sclerosis
Multiple sclerosis
Primary biliary cirrhosis
ATP6V1G2
Schizophrenia
Rheumatoid arthritis
Rheumatoid arthritis
BLK
Systemic lupus erythematosus
Early onset inflammatory bowel disease
Crohn's disease
Type I diabetes
CCDC101
Weight
Body mass index
Rheumatoid arthritis
CCR6
Crohn's disease
Vitiligo
Waist circumference
Metabolic syndrome
Cholestrol (HDL and LDL)
CETP
Triglycerides
Age-related macular degeneration
Caffeine consumption
CSK
Diastolic blood pressure
Coronary heart disease
Coronary heart disease
CTSH
Type I diabetes
Inflammatory disease disease
Ankylosing spondylitis
ETS2
C-reactive protein
Ulcerative colitis
Chronic obstructive pulmonary disease
Nicotine dependence
IREB2
Lung cancer
Smoking behaviour
Ulcerative colitis
IRF5
Systemic lupus erythematosus
Chronic kidney disease
Waist circumference
C-reactive protein
Serum calcium
KRTCAP3
Crohn's disease
Phospholipid levels
Fasting glucose
Cholestrol (LDL)
Triglycerides
Serum urate
AIDS progression
Rheumatoid arthritis
Nephropathy
LST1
Psoriasis
Ulcerative colitis
Telomere length
Multiple sclerosis
AIDS progression
Psoriasis
MICB
HIV control
Hepathocellular carcinoma
Celiac disease
MMEL1
Ulcerative colitis
Rheumatoid arthritis
Breast cancer
OCEL1
Ovarian cancer
Blood lipid traits
Coronary heart disease
PSRC1
Cholestrol (total, LDL)
Progranulin levels
Myocaridal infaction (early onset)
Crohn's disease
Ulcerative colitis
RBM6
Menarche
Primary chlerosing cholangitis
Crohn's disease
Ankylosing spondylitis
SDCCAG3
Ulcerative colitis
Ankylosing spondylitis
SNAPC4
Crohn's disease
Ulcerative colitis
Body mass index
Crohn's disease
SPNS1
Type I diabetes
Weight
Glioma
STMN3
Crohn's disease
Ulcerative colitis
Body mass index
Weight
SULT1A1
Crohn's disease
Type I diabetes
Celiac disease
Crohn's disease
UBE2L3
Systemic lupus erythematosus
Primary sclerosing cholangitis
USP4
Ulcerative colitis
Crohn's disease
HIV control
Nephropathy
Vitiligo
Celiac disease
Type I diabetes
Lung cancer
VARS2
Rheumatoid arthritis
Neonatal lupus
Immunoglobulin A
Schizophrenia
Hepatocellular carcinoma
Systemic lupus erythematosus
Rheumatoid arthritis
ZSWIM3
Multiple sclerosis

b



Supplementary Figure 15. **Cell specific cis-eQTL involving SNP markers associated with disease.** Genes showing significant cis-eQTL that involve SNPs reported at genome-wide significance ($p < 5x10-8$) in the Catalog of Published Genome-Wide Association Studies (www.genome.gov/GWA studies) (accessed 10th September 2011) or proxy SNPs identified for these disease markers from the 1000 Genomes Project (CEU cohort, $r2 > 0.8$) are shown. (a) Examples of genes showing eQTL involving SNPs associated with multiple GWAS traits. (b) List of genes with shared cis-eQTL/GWAS SNPs for disease traits grouped by cell type and excluding HLA genes.