

# Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles

Benjamin P Fairfax<sup>1</sup>, Seiko Makino<sup>1</sup>, Jayachandran Radhakrishnan<sup>1</sup>, Katharine Plant<sup>1</sup>, Stephen Leslie<sup>2</sup>, Alexander Dilthey<sup>3</sup>, Peter Ellis<sup>4</sup>, Cordelia Langford<sup>4</sup>, Fredrik O Vannberg<sup>1,5</sup> & Julian C Knight<sup>1</sup>

*Trans*-acting genetic variants have a substantial, albeit poorly characterized, role in the heritable determination of gene expression. Using paired purified primary monocytes and B cells, we identify new predominantly cell type-specific *cis* and *trans* expression quantitative trait loci (eQTLs), including multi-locus *trans* associations to *LYZ* and *KLF4* in monocytes and B cells, respectively. Additionally, we observe a B cell-specific *trans* association of rs11171739 at 12q13.2, a known autoimmune disease locus, with *IP6K2* ( $P = 5.8 \times 10^{-15}$ ), *PRIC285* ( $P = 3.0 \times 10^{-10}$ ) and an upstream region of *CDKN1A* ( $P = 2 \times 10^{-52}$ ), suggesting roles for cell cycle regulation and peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) signaling in autoimmune pathogenesis. We also find that specific human leukocyte antigen (HLA) alleles form *trans* associations with the expression of *AOAH* and *ARHGAP24* in monocytes but not in B cells. In summary, we show that mapping gene expression in defined primary cell populations identifies new cell type-specific *trans*-regulated networks and provides insights into the genetic basis of disease susceptibility.

Defining the genetic determinants of gene expression is crucial to understanding the biological and medical significance of genetic variation. This is particularly relevant in the drive to identify functional variants underlying observed disease associations from genome-wide association studies (GWAS)<sup>1</sup>. It is increasingly clear that the functional activity of many genetic polymorphisms is dependent on the context of relevant cell or tissue types in a particular biological state<sup>2–4</sup>. This context-dependent specificity means that, although studies of lymphoblastoid cell lines (LCLs) and other tissues have provided important insights, they may fail to capture the *in vivo* activity of particular variants in disease-relevant tissues<sup>5,6</sup>. Recent cell- and tissue-specific studies highlight the importance of context in the identification of expression-associated genetic variants<sup>3,4,7–10</sup>. In umbilical cord-derived cultured cells, up to 80% of regulatory variants act in a cell type-specific manner<sup>3</sup>. Comparison of skin, fat and LCLs showed that only 30% of eQTLs are shared among tissues<sup>4</sup>. The basis for this specificity remains unresolved but may relate to variation at tissue-specific distal enhancers instead of at conserved promoter elements<sup>3</sup>. Analyses performed on non-cultured primary tissue have typically used sources with a heterogeneous cell composition, such as peripheral blood leukocytes<sup>11</sup> or fat<sup>4</sup>. Although this approach provides general insights into tissue-specific eQTLs, cell type-specific eQTLs may be missed due to signal saturation from other cell types where the eQTL is absent. This is

especially pertinent in the elucidation of *trans*-acting eQTLs, where context specificity may be of greater relevance<sup>12</sup>.

Here we sought to determine cell type-specific eQTLs relevant to immunity and inflammation in paired samples of primary monocytes and B cells, purified by positive selection directly from healthy individuals. Our analysis highlights both the extent of cellular specificity, especially for *trans*-acting variants, and the inherent complexity of eQTL action. We observe multiple examples of genes associated with eQTLs in both cell types, but to different genomic loci, and of eQTLs showing opposing, cell type-dependent directional effects. Mapping genetic determinants of gene expression in immune cells is highly informative for understanding the function of reported GWAS-identified genes involved in immune, infectious and inflammatory diseases.

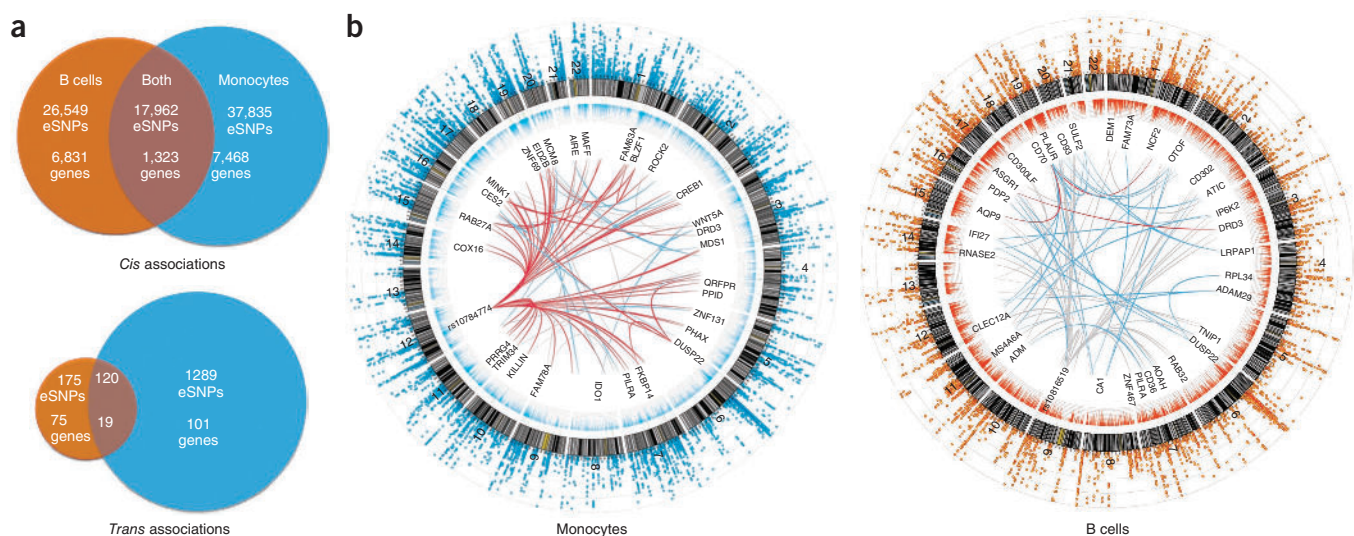
## RESULTS

### Defining eQTLs in purified B-cell and monocyte populations

B cells are lymphocytes with crucial roles in adaptive and humoral immunity, whereas monocytes constitute an innate myeloid-derived cell population that initiates an inflammatory, cytokine-mediated response upon microorganism invasion. Their divergent functions and origins ensure these cell populations are highly informative for analysis of immune and inflammatory diseases. Furthermore, whereas multiple LCL eQTL analyses have been performed, there are as yet no large studies focused on B cells, the cells immortalized to derive LCLs.

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>2</sup>Department of Oncology, University of Oxford, Oxford, UK. <sup>3</sup>Department of Statistics, University of Oxford, Oxford, UK. <sup>4</sup>Wellcome Trust Sanger Institute, University of Cambridge, Cambridge, UK. <sup>5</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA. Correspondence should be addressed to B.P.F. (bfaifax@well.ox.ac.uk) or J.C.K. (julian@well.ox.ac.uk).

Received 7 October 2011; accepted 31 January 2012; published online 25 March 2012; doi:10.1038/ng.2205



**Figure 1** Shared and cell type-specific *cis* and *trans* associations in B cells and monocytes. **(a)** Venn diagrams showing the number of eSNPs unique to each cell subset and shared between data sets in *cis* (top) and in *trans* (bottom). **(b)** Circos plots for the monocyte (left) and B-cell (right) data sets. The outermost rim shows a Manhattan plot for eQTLs from the respective data set, the second rim shows relative expression of genes, the third rim depicts an arbitrary selection of genes (constrained due to space) with significant *trans* eQTLs ( $P < 1 \times 10^{-11}$ ), and the innermost network depicts spokes between nodal eSNPs and their *trans*-regulated genes. Such nodal eSNPs include rs10784774, which marks a monocyte-specific master regulatory region, whereas rs10816519 represents a B cell-specific master regulatory region. Red spokes,  $>10$  eSNPs from this locus map to the *trans* eQTLs; blue spokes,  $>1$  and  $<10$ ; gray spokes, 1 eSNP.

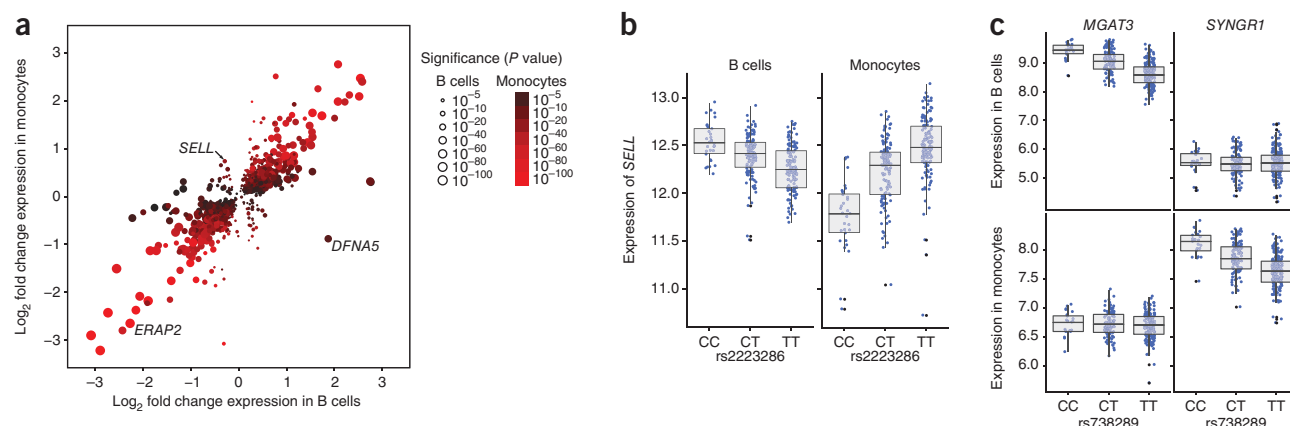
To investigate eQTLs in these primary cell types, we used positive selection, a method that has been shown to result in superior cell purity for microarray analysis<sup>13</sup>, to separate CD19<sup>+</sup> B cells and CD14<sup>+</sup> monocytes from peripheral blood mononuclear cells (PBMCs) prepared from the whole blood of 288 healthy European volunteers (Online Methods). Sample purity was confirmed with flow cytometry and was 90–95% for B cells and approached 99% for monocytes. Genome-wide gene expression profiling and genotyping were performed using HumanHT-12 v4 Expression BeadChips and Human OmniExpress-12v1.0 BeadChips, respectively (both from Illumina). Following processing and quality control filtering, we performed eQTL mapping at 651,210 markers for each of 283 individuals.

### Cell type-specific *cis* eQTLs show directional effects

Identification of locally acting eQTLs (referred to here as *cis* acting) was performed by testing SNPs that fell within a 2.5-Mb interval on either side of the probe for association with expression in each cell type, using linear and Spearman rank models. In this large, highly purified paired-sample set, we found little difference between the significance values from the two approaches—however, only eQTLs that reached a permuted  $P < 1.0 \times 10^{-3}$  in both analyses were carried forward. We identified 82,346 eQTLs (SNP-probe interactions, referred to hereafter as eSNPs) at a permuted  $P < 0.001$ , 32.2% of which were unique to B cells, 45.9% of which were unique to monocytes and 21.8% of which were shared between cell types (Fig. 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2). This corresponded to 7,468, 6,831 and 1,323 genes (8,441, 7,589 and 1,466 probes) in which there was at least one *cis* eSNP with permuted  $P < 0.001$  specific to B cells or monocytes or shared by both cell types, respectively (a full description of associations at different significance thresholds is given in Supplementary Table 3). Comparative analysis of this data set revealed a high concordance between eSNPs identified here that were shared between cell types and those previously identified in eQTL analysis in primary cells<sup>10,14</sup>. This degree of commonality was

not as clear in comparing these eSNPs with those identified in LCLs<sup>15</sup>, although such comparisons may be confounded by differences in the array platform used (Supplementary Fig. 2). Cell-specific eSNPs may exist secondary to the cell-specific function of genetic variants or, alternatively, when transcript expression is restricted to one cell type only. In order to identify eSNPs formed as a result of cell-specific functional variants, we identified genes expressed at similar levels across cell types but which showed robust cell-specific effects. Even after excluding genes whose mean  $\log_2$  expression across the cohort differed in magnitude by  $>0.5$  between cell types, we identified 70,149 *cis* eSNPs in almost identical proportions by cell type. In general, whereas cell type-specific eQTLs were observed more frequently, they tended to have smaller effect sizes than those shared between cells (median  $r^2 = 5.1\%$ ,  $5.5\%$  and  $9.8\%$  for B cell-specific, monocyte-specific and shared eSNPs, respectively). Nonetheless, in a small proportion of genes, more than 50% of the variance in expression was associated with eSNPs, involving 14, 37 and 61 genes in B cells, monocytes and both cell types, respectively. Using additional RNA purified from randomly selected individuals within the cohort, we were able to replicate the cell type specificity for selected genes by RT-PCR ( $n = 14$ –29 homozygous individuals per gene) (Supplementary Fig. 3). When RNA from crude PBMCs from the same individuals was analyzed, however, the signal was frequently absent (Supplementary Fig. 3b,d,i), providing evidence of the importance of cell purification in the identification of subtle primary cell-specific eQTLs.

Consistent with other reported eQTL analyses, we found that the effect size and statistical significance of an eQTL varied as a function of distance from the transcription start site (TSS) of the relevant gene (Supplementary Fig. 4). Notably, the density distribution of eSNPs with cell type-specific effects on expression was more dispersed around the TSS in comparison to that of eQTLs that affected expression in both cell types. This finding supports the notion that cell type-specific eQTLs are enriched in more distant enhancer elements involved in cell type-specific expression<sup>3</sup>.



**Figure 2** eSNPs shared between cell types may lead to opposing directional effects on gene expression or associate with expression of different genes in a cell type-specific manner. (a) For eSNPs shared between cell types, the most significant eSNP per probe is plotted with the fold change in expression this eSNP is associated with in the homozygous form between major and minor alleles. Whereas the majority of eSNPs shared between data sets were associated with the same directional change, there are examples of eSNPs that associated with opposing directional changes in expression, depending on cell type. Only one eSNP is plotted per probe, and only eSNPs with examples of  $>2$  individuals homozygous in the minor allele with permuted  $P < 0.001$  in both B-cell and monocyte data sets are annotated. (b) rs2223286 is associated with marked directional effects in the expression of *SELL*, depending on genotype, with the minor C allele associated with increased expression of *SELL* in B cells and reduced expression of *SELL* in monocytes ( $P_{\text{B cell}} = 4.6 \times 10^{-11}$ ,  $P_{\text{monocyte}} = 1.1 \times 10^{-22}$ ). (c) rs738289 is an example of an eSNP that constitutes an eQTL to differing genes, depending on cell type. In B cells, this eSNP is strongly associated with *MGAT3* expression ( $P_{\text{B cell}} = 9.8 \times 10^{-26}$ ) with no association to *SYNGR1* expression, whereas, in monocytes, this eSNP is significantly associated with *SYNGR1* expression ( $P_{\text{monocyte}} = 1.2 \times 10^{-17}$ ), with no association to *MGAT3* expression. Box plots show median, upper and lower quartiles, with whiskers denoting maximal and minimal data within  $1.5 \times$  interquartile range (IQR). Outlying values are shown in black.

Previously reported eQTLs that are shared between cell types have the same directional effect in each analyzed cell type<sup>3</sup>. Although in general this holds true for our analysis of primary monocytes and B cells, we observed several eQTLs with cell type-dependent directional effects, where the same eSNP was associated with opposing directional effects in B cells and monocytes. We identified 197 eSNPs in 35 genes with significant ‘directional eQTL’ activity in B cells and monocytes (permuted  $P < 0.001$  in both cell types; total opposing directional effect  $>0.5$  in terms of magnitude of mean  $\log_2$  expression difference between major and minor alleles) (Fig. 2a and Supplementary Table 4). To further investigate the implications of these directional findings, we estimated differences in slope using  $z$  scores and showed that for 31 of the 35 genes the differences were highly significant (Bonferroni-corrected  $P < 1 \times 10^{-38}$ ) (Supplementary Table 4), thus strongly suggesting that these were not chance observations. The most significant directional eQTLs were found to affect the *DFNA5* gene, which was previously implicated in familial deafness<sup>16</sup> and is a target of promoter methylation in gastric cancer<sup>17</sup>. Other notable genes with directional effects included *MCOLN2*, encoding a cation channel involved in type I interferon responses<sup>18</sup>, and *SELL* (also known as L-selectin or CD62L) (Fig. 2b). The latter is of particular interest, given the role of this cell surface receptor in monocyte recruitment to lymphoid tissues during inflammation<sup>19</sup> and previous association of this region with amyotrophic lateral sclerosis<sup>20</sup>. We did not detect an association with *KIFAP3* expression that was reported in LCLs<sup>20</sup>, but we note that disease-associated variants showed a cell type-dependent directional effect on the neighboring *SELL* gene. RT-PCR quantification of *DFNA5* and *SELL* expression in additional RNA purified from randomly selected homozygous individuals (*DFNA5*,  $n = 16$ ; *SELL*,  $n = 20$ –28) confirmed that these directional effects were not an array-derived artifact (Supplementary Fig. 3h,i).

The extent to which a particular genetic variant may have diverging effects according to cell type is unclear. Here we find that many variants have multiple cell type-specific actions, with 6.0% of cell type-specific eSNPs additionally constituting eQTLs to different

genes in the other cell type (Supplementary Fig. 5a). For example, rs738289 and linked eSNPs in the first intron of *MGAT3*, a gene encoding a glycosyltransferase<sup>21</sup>, showed association with *MGAT3* in B cells ( $P = 9.8 \times 10^{-26}$ ); by contrast, in monocytes, this eSNP was only associated with the neighboring *SYNGR1* gene, encoding a vesicle-associated protein ( $P = 1.2 \times 10^{-17}$ ) (Fig. 2c). Just as cell type may define the gene a particular variant regulates, we observed certain genes whose expression was modulated by different variants in a cell type-specific manner. Examples of this class of genes include *TSPAN3* (Supplementary Fig. 5b) and *AKAP7* (data not shown), and this finding is in keeping with the functional effect of genetic variants being dependent on the regulatory environment defined by the cell type.

### Identification of multiple cell type-specific *trans*-associated eQTLs

Although a considerable element of heritable gene expression is proposed to act in *trans*, *trans* eQTLs have proven difficult to define in the cell populations studied to date. It is probable that *trans*-acting variants are involved in a more multifactorial process than are eQTLs acting in *cis*, reflecting contributions both of prevailing environmental cues and the cellular context. To explore this topic, we mapped *trans*-associated eSNPs using a conservative significance threshold that was based on a Bonferroni-corrected  $P$  value of  $1 \times 10^{-11}$ . For autosomal genes and SNPs, we identified 1,704 eSNPs involving 75 genes specific to B cells, 101 specific to monocytes and 19 shared by both cell types (Fig. 1 and Supplementary Table 2). It is increasingly clear that particular variants act as master regulators of transcription as *trans* eQTLs<sup>12,14</sup>. Many of these eSNPs are *cis* eQTLs to genes that probably define the nodal gene of these subsets (Supplementary Fig. 6 and Supplementary Table 2). Two highly cell type-specific examples of such genes are *KLF4* in B cells (Supplementary Figs. 7 and 8) and *LYZ* in monocytes. We found that only 7% of *trans* eSNPs are shared between cell types, implying increased cell type specificity for *trans* eQTLs. Given that B cells constitute ~5% of PBMCs and monocytes 10–15% (ref. 13), saturation of cell type-specific signals by expression signatures from other



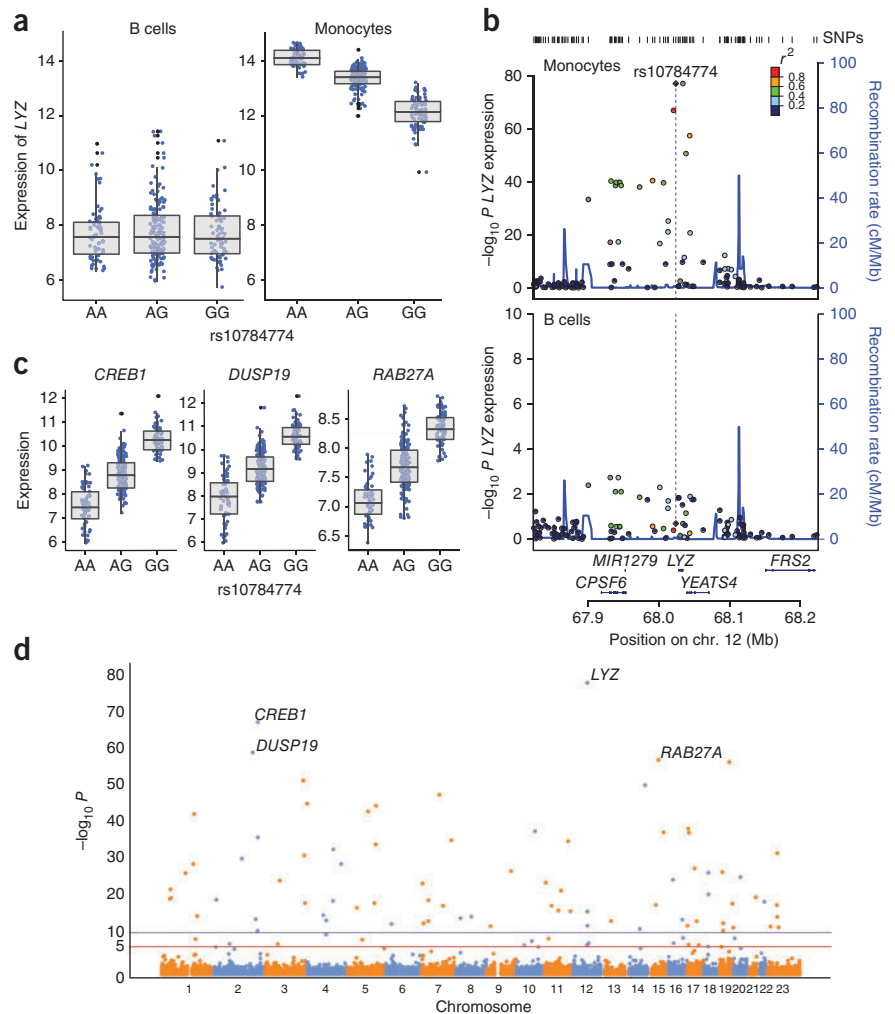
**Figure 3** rs10784774 marks a monocyte-specific *cis* eSNP to *LYZ* and constitutes a monocyte-specific master regulator of multiple genes, including *CREB1*. (a,b) rs10784774 is significantly associated with the expression of a probe mapping to the 3' UTR of *LYZ* only in monocytes ( $P_{\text{monocyte}} = 8.9 \times 10^{-78}$ ) (a) and not in B cells (b). (c) rs10784774 is also an eSNP to 72 probes mapping across the genome, the most significantly associated being in the genes *CREB1* ( $P = 2.0 \times 10^{-67}$ ), *DUSP19* ( $P = 2.2 \times 10^{-59}$ ) and *RAB27A* ( $P = 2.0 \times 10^{-57}$ ). (d) Manhattan plot showing chromosomal location of all genes with probes mapping to rs10784774. Blue line,  $1.0 \times 10^{-11}$ ; red line,  $5.0 \times 10^{-8}$ . Box plots are depicted as in Figure 2.

cell types may render many cell type-specific eQTLs undetectable when analyzing heterogeneous PBMC samples.

### *LYZ* as a monocyte-specific master regulator of a large gene set

Monocytes scavenge and present antigens through a highly conserved endosomal pathway that utilizes hydrolyzing enzymes released from secretory granules. Lysozyme is a key constituent of monocyte secretory granules and degrades bacterial cell wall-derived peptidoglycan. We found evidence of a new master regulatory region involving *cis* and *trans* eQTLs at the 12q15 locus for rs10784774 and associated eSNPs spanning *LYZ*, which encodes lysozyme (Fig. 3). This locus forms a monocyte-specific *cis* eQTL (Supplementary Fig. 9) to a probe mapping to the 3' UTR of *LYZ* ( $P_{\text{monocyte}} = 8.9 \times 10^{-78}$ ) and forms a *trans* eQTL to a total of 62 annotated genes (72 probes) throughout the genome at  $P < 1 \times 10^{-11}$ , with the most significant association being with expression of the *CREB1* gene ( $P_{\text{monocyte}} = 2.03 \times 10^{-67}$ ) (Fig. 3, Supplementary Fig. 10 and Supplementary Table 5). Interrogation of previously published primary cell eQTL analyses of negatively selected monocytes from whole blood<sup>10</sup> and of whole blood-derived RNA from a mixed disease and control cohort<sup>14</sup> also showed this strong association between 12q15 and expression of both *LYZ* and *CREB1* (*CREB1*;  $P = 5.1 \times 10^{-22}$  (rs11177644) in monocytes<sup>10</sup> and  $P = 3.2 \times 10^{-67}$  (rs2168029) in whole blood<sup>14</sup>). The multifaceted transcriptional properties of *CREB1* in immune cells<sup>22</sup> make it a highly plausible candidate to regulate expression of many downstream genes, consistent with recent reports of other transcription factors underlying multiple eQTLs from one locus<sup>12</sup>. Notably, *LYZ* expression was most strongly correlated to that of *CREB1* ( $r^2 = 0.82$ ,  $P = 5.7 \times 10^{-107}$ ), and *CREB1* expression was highly correlated with the majority of *trans*-associated genes (Supplementary Table 5). These findings support a pathway whereby *cis* modulation of *LYZ* results in differential expression of *CREB1*, with resultant regulation of a network of *trans*-associated genes.

rs10784774 is also a *cis* eSNP that acts on *YEATS4* in both B cells and monocytes ( $P_{\text{B cell}} = 1.55 \times 10^{-7}$  and  $P_{\text{monocyte}} = 1.26 \times 10^{-13}$ ). Although *YEATS4* encodes a transcription factor<sup>23</sup> and could therefore potentially account for the observed *trans* eQTL activity, the exclusivity of the *trans* eQTL to monocytes argues against this. Furthermore,



correlation analyses between *YEATS4*, *CREB1* and *LYZ* expression and that of other probes in monocytes showed that the strongest coefficient of correlation was to *LYZ* and *CREB1* expression for all *trans*-associated genes (Supplementary Table 5).

Imputation revealed that rs10784774 is the 5' SNP in a haplotype that spans the *LYZ* locus that is composed of four SNPs in almost complete linkage disequilibrium (LD) (Supplementary Fig. 11). Analysis of ENCODE ChIP-seq data<sup>24</sup> identified a p300-binding site overlapping rs10784774 and additional sites overlapping the promoter (Supplementary Fig. 12). To investigate the allelic relationship between p300 and *LYZ* expression, genotype-conditioned correlation analysis between basal expression of p300 and *LYZ* was performed. We observed a strong inverse correlation between expression of *EP300* (encoding p300) and *LYZ* in individuals homozygous for the ancestral A allele at rs10784774, but we saw no association with expression with the homozygous G allele (AA:  $r^2 = 0.41$ ,  $P = 4.4 \times 10^{-8}$ ; AG:  $r^2 = 0.22$ ,  $P = 1.4 \times 10^{-9}$ ; GG:  $r^2 = 0.02$ ,  $P = 0.11$ ), supporting the idea of allele-specific regulation of *LYZ* via p300.

### B cell-specific *trans* eQTLs involving 12q13.2

The generation of self-reactive auto-antibodies is a characteristic feature of autoimmune disease. In unaffected individuals, B cells expressing self-reactive antibody are removed before their maturation to antibody-secreting plasma cells to prevent this self-reactive activity. Here we show that SNPs at 12q13.2, a locus showing reproducible

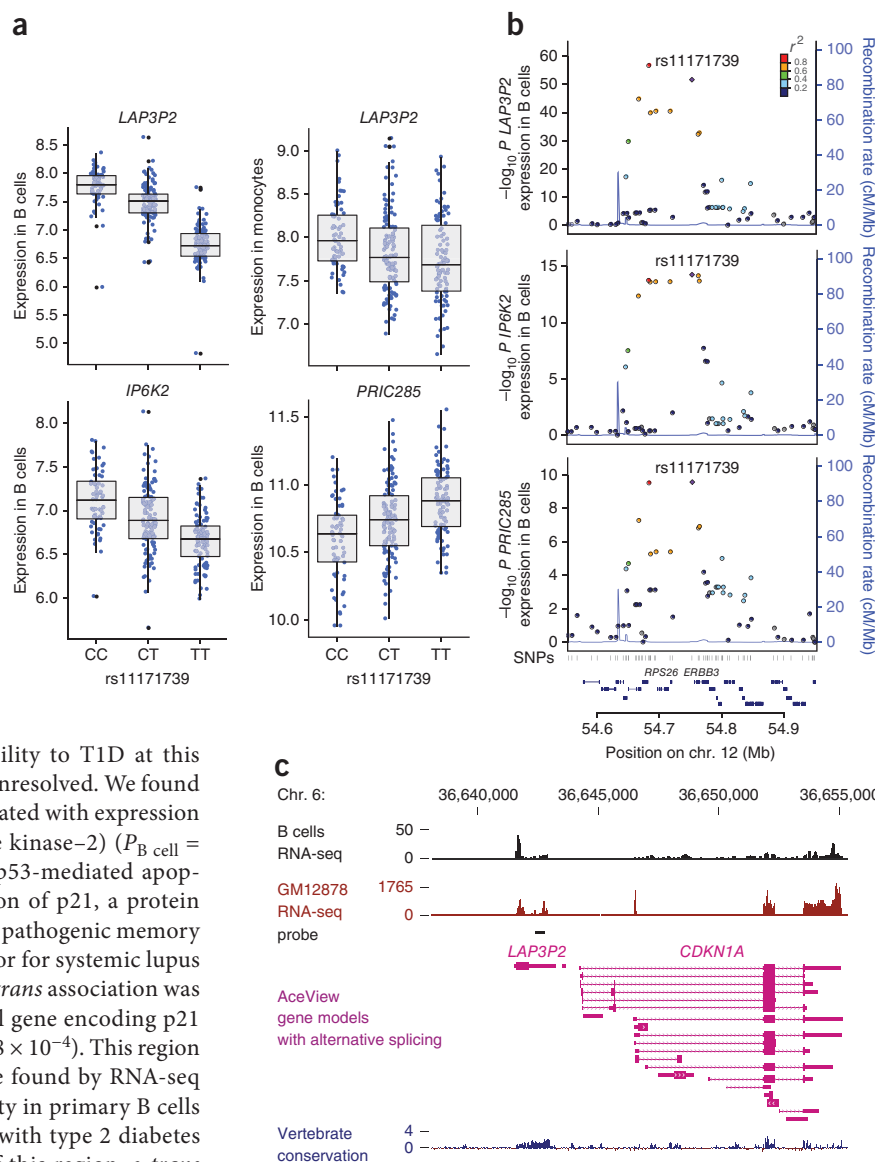
**Figure 4** rs11171739, a SNP at 12q13.2

with strong autoimmune association, is a B cell-specific *trans* eQTL to three genes. (a) rs11171739 specifically associates with the B-cell *trans* expression of the genes *LAP3P2* ( $P_{\text{B cell}} = 2.0 \times 10^{-52}$ ,  $P_{\text{monocyte}} = 1.7 \times 10^{-4}$ ), *IP6K2* ( $P_{\text{B cell}} = 5.8 \times 10^{-15}$ ) and *PRIC285* ( $P_{\text{B cell}} = 3.0 \times 10^{-10}$ ). Box plots are depicted as in **Figure 2**. (b) B-cell regional association plots of these genes showing their expression by SNP marker across chromosome 12q13.2, a known autoimmune risk locus. (c) *LAP3P2* lies ~1 kb upstream of *CDKN1A* and is denoted as a pseudogene. We observed a high degree of transcriptional activity in our RNA-seq data from CD19<sup>+</sup> primary B cells, which was also seen in LCLs (GM12878)<sup>24</sup>. This region is highly conserved among vertebrates.

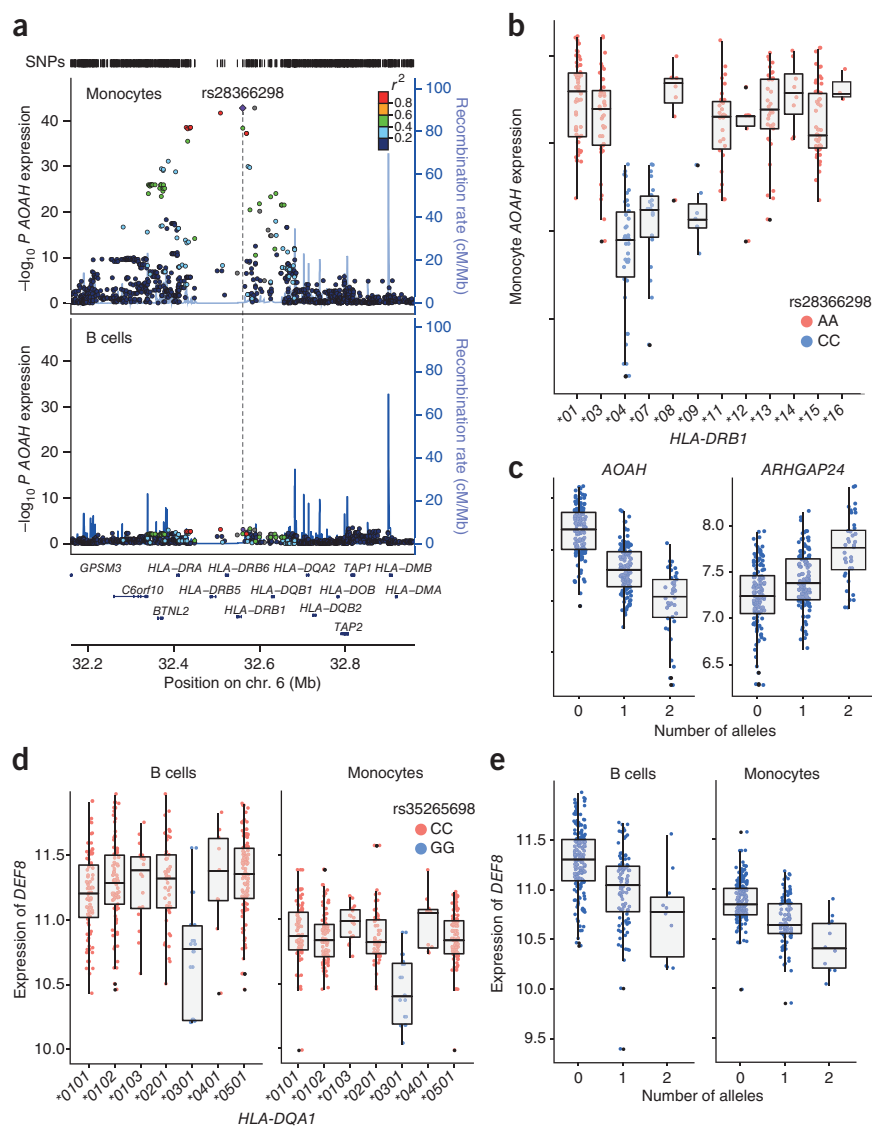
association with multiple autoimmune diseases, including type 1 diabetes (T1D)<sup>25–28</sup>, regulate the expression of genes implicated in cell cycle control in *trans* in B cells only. Previous studies have shown that rs11171739 at 12q13.2 is a *cis* eQTL that affects *RPS26* in liver cells<sup>8</sup>, LCLs<sup>15</sup> and leukocytes<sup>11</sup>. Differential *RPS26* expression, however, seems to be insufficient to explain susceptibility to T1D at this locus<sup>29</sup>, and the identity of the causal genes is unresolved. We found a highly significant *trans* eQTL in B cells associated with expression of *IP6K2* (encoding inositol hexakisphosphate kinase-2) ( $P_{\text{B cell}} = 5.8 \times 10^{-15}$ ) (**Fig. 4a**). *IP6K2* is required for p53-mediated apoptosis<sup>30</sup> and is able to regulate the accumulation of p21, a protein that is involved in cell division and apoptosis of pathogenic memory B cells<sup>31</sup> and that is a known susceptibility factor for systemic lupus erythematosus (SLE)<sup>32</sup>. Of note, an additional *trans* association was found for a transcript mapping 5' to the actual gene encoding p21 (*CDKN1A*) ( $P_{\text{B cell}} = 2 \times 10^{-52}$  and  $P_{\text{monocyte}} = 1.8 \times 10^{-4}$ ). This region is denoted as the pseudogene *LAP3P2*, but we found by RNA-seq that it is a region of high transcriptional activity in primary B cells (**Fig. 4**). rs11171739 has also been associated with type 2 diabetes (T2D)<sup>33</sup>, and we noted, on further analysis of this region, a *trans* eQTL between this SNP and *PRIC285* ( $P_{\text{B cell}} = 3.0 \times 10^{-10}$ ), which encodes a transcriptional coactivator involved in PPAR $\gamma$  signaling. Given the role of PPAR $\gamma$  agonists in T2D management, this association may be of notable biological relevance<sup>34</sup>.

### *Trans* eQTLs defined by disease-associated HLA alleles

We detected several robust associations to the major histocompatibility (MHC) region ( $P < 1 \times 10^{-11}$ ) that segregated by HLA allele, which may provide insights into the basis of autoimmune disease associations to this region. We identified a *trans* association of *AOAH*, encoding acyloxyacyl hydrolase lipase, which degrades bacterially derived lipopolysaccharide<sup>35</sup>, to the MHC class II region. An association between the MHC and *AOAH* has recently been described in a study of RNA from whole blood of a mixed population of disease and control cohorts (peak eSNP rs2395185,  $7.0 \times 10^{-38}$ )<sup>14</sup>. We replicated the finding at this SNP (rs2395185,  $P_{\text{monocyte}} = 5.7 \times 10^{-39}$ ), with the most strongly associated eSNP in our data set being rs28366298 ( $P_{\text{monocyte}} = 1.6 \times 10^{-43}$ ). Notably, we found only very weak association of this region in B cells ( $P_{\text{B cell}} > 1 \times 10^{-3}$ ) (**Fig. 5a**), suggesting that much of the signal in blood is monocyte derived and providing potential mechanistic insight. To further define this *trans*



association, we inferred the underlying HLA alleles to two- and four-digit resolution in all individuals through genotype imputation<sup>36,37</sup>. This enabled us to refine the association of *AOAH* expression and the MHC region to the presence or absence of the HLA class II alleles *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* ( $P < 2.2 \times 10^{-16}$ , two-way ANOVA) (**Fig. 5b** and **Supplementary Fig. 13**). Specifically, individuals homozygous for any of these alleles expressed 0.6-fold (95% confidence interval (CI) = 0.57–0.67) the level of *AOAH* as individuals homozygous for all other alleles at *HLA-DRB1*. We observed a second monocyte-specific *trans* association between *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* and *ARHGAP24* (peak eSNP rs28366298:  $P_{\text{monocyte}} = 4.59 \times 10^{-14}$ ). *ARHGAP24* encodes a negative regulator of Rho GTPase-activating protein that is implicated in cell migration<sup>38</sup>. In contrast to *AOAH*, individuals homozygous for the *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* alleles expressed 1.4-fold (95% CI = 1.30–1.54) the level of *ARHGAP24* relative to all other alleles. These alleles are associated with autoimmune disease susceptibility and the presence of *HLA-DRB4* (encoding the supertypic HLA-DR53 antigen)<sup>39,40</sup>. Our study shows that specifically monocytes from individuals



**Figure 5** Imputation of HLA status resolves cell type-specific *trans*-associated gene expression to the carriage of specific HLA alleles.

(a) Regional association plots from monocyte and B-cell data showing the monocyte-specific *trans* association of the chromosome 7 gene *AOAH* to the class II MHC gene *HLA-DRB1*. There was no association in the B-cell data set, whereas the peak eSNP from the monocyte data set was rs28366298 ( $P_{\text{monocyte}} = 1.6 \times 10^{-43}$ ). (b) Imputation of class II alleles to two-digit resolution showed that the C allele of rs28366298 was specific to the *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* alleles, and these alleles are associated with reduced *AOAH* expression. For clarity, only homozygotes are plotted, with two *AOAH* expression values per individual (corresponding to each *HLA-DRB1* allele). (c) The number of *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* alleles carried by an individual is significantly associated with reduced expression of *AOAH* ( $P_{\text{B cell}} < 2.2 \times 10^{-16}$ , one-way ANOVA) and increased expression of *ARHGAP24* ( $P_{\text{monocyte}} = 3.0 \times 10^{-14}$ , one-way ANOVA). (d) Imputation of *HLA-DQA* status to four digits shows that the G allele of rs35265698 is significantly associated with reduced expression of *DEF8* in both B cells and monocytes ( $P_{\text{B cell}} = 6.2 \times 10^{-13}$ ,  $P_{\text{monocyte}} = 9.0 \times 10^{-17}$ ). This SNP allele is unique to *HLA-DQA1\*0301*. For clarity, only homozygotes of rs35265698 are plotted, with two *DEF8* expression values per individual (corresponding to each *HLA-DQA1* allele). (e) Higher numbers of the *HLA-DQA1\*0301* allele in an individual is significantly associated with reduced *DEF8* expression ( $P_{\text{B cell}} = 6.2 \times 10^{-13}$ ,  $P_{\text{monocyte}} < 2.2 \times 10^{-16}$ , one-way ANOVA). Box plots are depicted as in Figure 2.

data identified a significant *cis* eQTL, common to both cell types, that was potentially confounded by SNPs within the probe-annealing sequence. We therefore performed quan-

titative RT-PCR across the cohort using primers predicted to anneal outside of the probe-binding site in regions with minimal SNP density. We thereby identified a significant eQTL for rs9264942 ( $P = 7.4 \times 10^{-9}$ ) in PBMCs, consistent with previous findings<sup>43</sup>, but the most significant association for *HLA-C* mapped to rs10484554 ( $P = 5.2 \times 10^{-16}$ ) (Supplementary Fig. 14a), a GWAS-identified SNP showing significant association with AIDS progression<sup>45</sup> and psoriasis<sup>46,47</sup>. HLA imputation analysis showed that the *cis* eQTL can be defined in terms of the presence of the *HLA-C\*0602* and *HLA-C\*1203* alleles (Supplementary Fig. 14b,c). Further investigation is required to resolve the role of *HLA-C* and genetic variation at this locus in terms of disease susceptibility. *HLA-C\*0602* has been strongly implicated in susceptibility to psoriasis<sup>48</sup>, although the association with HIV control and progression is complex, with a key role for HLA-B-mediated peptide presentation<sup>49</sup>.

These *trans* associations to class II MHC alleles represent the first observation to our knowledge of differential expression of genes defined in *trans* by the presence of specific HLA alleles. Although the implications of these allele-specific molecular signatures are unclear, they are likely to provide new insight into the mechanistic basis of MHC-associated disease susceptibility.

Finally, although carrying an HLA allele is associated with haplo-type-specific expression patterns across the MHC<sup>42</sup>, the high density of SNPs in this region, especially within the classical HLA genes, results in exclusion of many probes. For *HLA-C*, an eQTL has been reported for the rs9264942 SNP<sup>43</sup> located 35 kb upstream of the gene associated with human immunodeficiency virus (HIV)-1 viral load and acquired immunodeficiency syndrome (AIDS) progression<sup>43,44</sup>. Our array

with *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* have significantly reduced *AOAH* expression, which encodes a protein with anti-inflammatory properties, in tandem with increased expression of *ARHGAP24* (Fig. 5c). Finally, in both cell types, we observed a *trans* eQTL to *DEF8* (rs2760985:  $P_{\text{monocyte}} = 9 \times 10^{-17}$  and  $P_{\text{B cell}} = 6.2 \times 10^{-13}$ ), a poorly characterized chromosome 16-encoded gene that is expressed in PBMCs<sup>41</sup>. This association resolves to the presence of the *HLA-DQA1\*0301* allele (Fig. 5d,e).

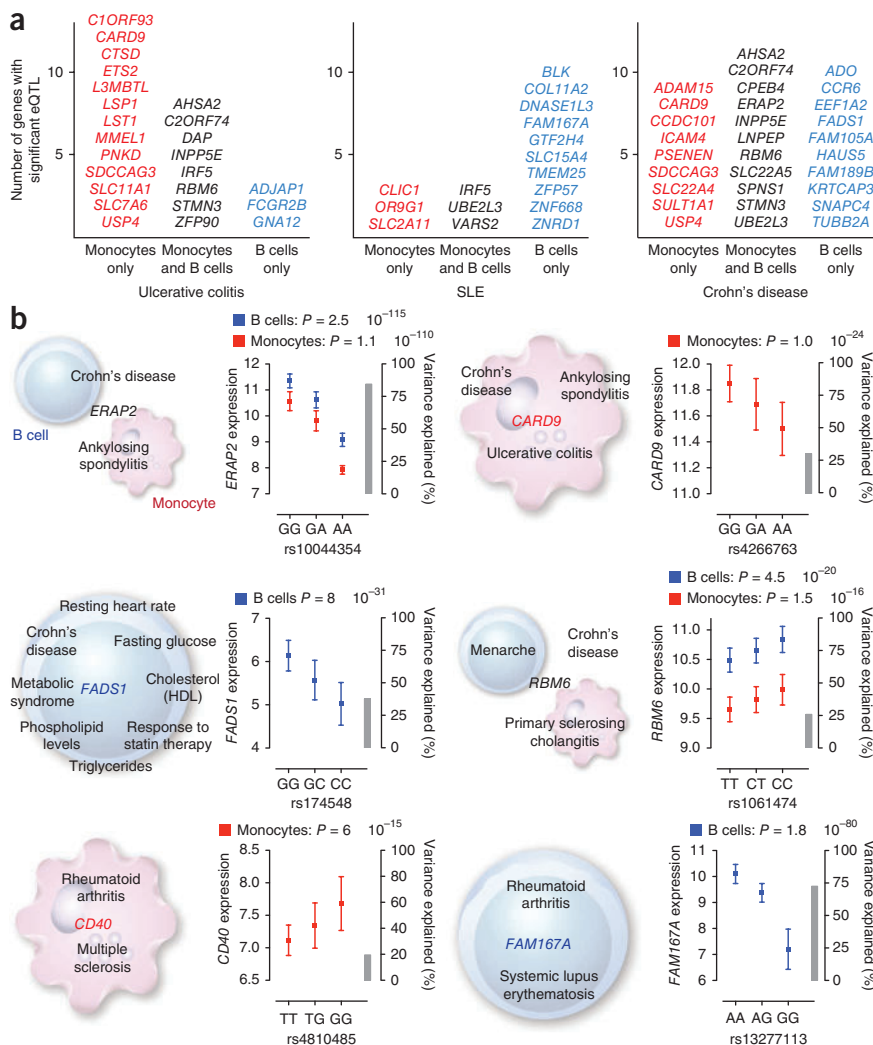
#### GWAS-identified SNPs form cell type-specific eQTLs

Disease susceptibility loci are frequently associated with differential gene expression<sup>1,50</sup>. The degree to which this effect may be cell type-specific is unclear. It is probable that certain disease-associated eQTLs only affect expression under physiological conditions of mixed cell populations



**Figure 6** Cell type-specific *cis* eQTLs involve disease-associated SNP markers. Genes showing significant *cis* eQTLs that involve SNPs reported at genome-wide significance ( $P < 5 \times 10^{-8}$ ) in the Catalog of Published Genome-Wide Association Studies (accessed 10 September 2011) or proxy SNPs identified for these disease markers from the 1000 Genomes Project (CEU cohort;  $r^2 > 0.8$ ) are shown. **(a)** List of genes with shared *cis* eQTL and GWAS SNPs for ulcerative colitis, SLE and celiac disease grouped by cell type and excluding HLA genes. Specific examples are described further in the **Supplementary Note**.

**(b)** Examples of genes showing eQTLs involving SNPs associated with multiple GWAS traits. Mean expression and s.d. are shown (left y axis) by SNP allele (x axis) for each of six genes significant in a given cell type together with variance explained (right y axis).



that permit appropriate cellular interactions and antigen exposure. We hypothesized that analysis of freshly isolated primary B cells and monocytes might allow the identification of such eQTLs, with the possibility of attributing functional activity to the innate or adaptive arms of the immune response. We investigated eQTLs involving SNPs identified as disease risk alleles at genome-wide significance ( $P < 5 \times 10^{-8}$ ) for 548 traits listed in the Catalog of Published Genome-Wide Association Studies (10 September 2011). We found that 49.4% of traits (represented by 17.3% of reported GWAS SNPs) were associated with one or more significant *cis* eQTL on the basis of the listed GWAS markers or identified proxy SNPs from the 1000 Genomes Project (Utah residents of Northern and Western European ancestry (CEU) cohort;  $r^2 > 0.8$ ). An integrated data set indexed by trait is provided (**Supplementary Table 6**). Overall, we found that 730 genes have significant *cis* eQTLs related to GWAS, of which 33.5% are B-cell specific, 45.5% are monocyte specific and 21% are found in both cell types.

Of note, we found that 4.6% of reported GWAS SNPs have *cis* eQTLs to differing genes in a cell type-dependent manner, supporting the notion of diverging effects of risk alleles across different cell types. For example, rs10781500, a SNP associated with ulcerative colitis<sup>51</sup> and ankylosing spondylitis<sup>52</sup>, showed contrasting *cis* eQTL activity for three genes at 9q34.3, depending on cell type. In B cells, rs10781500 was an eSNP to *SNAPC4* ( $P = 3.1 \times 10^{-5}$ ), encoding a small nuclear RNA-activating complex, whereas, in monocytes, the associations were with *SDCCAG3* ( $P = 1.9 \times 10^{-12}$ ) and *CARD9* ( $P = 1.0 \times 10^{-24}$ ); *SDCCAG3* encodes a protein implicated in the cell surface presentation of tumor necrosis factor (TNF) receptor 1 (TNFR1)<sup>53</sup>, and *CARD9* encodes a caspase recruitment domain-containing signaling protein with apoptotic and immunoregulatory roles<sup>54</sup>. Our analysis highlights the fact that certain traits are characterized by a preponderance of cell type-specific eQTLs, as shown by the association between eQTLs in B cells and SLE and between eQTLs in monocytes and ulcerative colitis (**Fig. 6a** and **Supplementary Note**). These diseases have contrasting and complex etiologies involving adaptive and innate immunity, with the role of B cells and auto-antibody production being paramount in SLE, whereas, in ulcerative colitis, dysregulated mucosal

immunity and monocyte lineage cells are critically important. Our analysis defines a catalog of genes showing cell type-specific or shared *cis* eQTLs for genetic variants implicated in many common diseases (**Fig. 6**, **Supplementary Fig. 15** and **Supplementary Table 6**), facilitating the fine mapping and functional characterization of specific genes and variants within GWAS disease intervals.

## DISCUSSION

In this study, we show that eQTL analysis of expression data obtained from highly purified primary cells enhances eQTL identification and reveals underlying cell type-specific complexity that is not appreciable in studies of mixed tissue. It is clear that, although many eQTLs with large effect sizes are shared across cells, the majority of eQTLs in primary tissue are cell type-specific. The degree to which a single variant can associate with the expression of different genes in monocytes and B cells is particularly notable. It is probable that analysis of additional divergent primary cell populations will increase the number of examples of such cell type-specific eQTLs. The delineation of cell type-specific directionality of eQTLs is also noteworthy; in the context of *SELL*, which encodes a membrane protein involved in cellular recruitment to areas of inflammation, the possibility is raised that directional eSNPs may influence the subset of cells recruited.

Results from this study show that *trans* eQTLs possess a higher degree of cellular autonomy than *cis* eQTLs, possibly reflecting the

fact that they represent products of upstream cell type-specific gene expression. Notably, we identified new master regulatory regions specific to monocytes that regulate *LYZ* and specific to B cells that regulate *KLF4*. The functional implications with respect to infectious disease susceptibility of the multi-locus *LYZ* eQTL remain to be explored, but the existence of several genes, including *ERAP2*, encoding an aminopeptidase involved in the loading of class I MHC molecules<sup>55</sup>, as well as *RAB27A*, encoding a GTPase critical to the membranous tethering of secretory lysosomes<sup>56</sup>, suggests that this eQTL may have an impact on antigen presentation. The identification of B cell-specific *trans* eQTLs to rs11717139, a locus with reproducible association with autoimmune diseases, in genes with putative roles in the cell cycle and apoptosis provides new candidate genes for analysis in autoimmune disease susceptibility. B cells are intrinsically prone to apoptosis, thereby providing protection against the generation of self-reactive antibodies. Regulation of genes involved in apoptosis would therefore be predicted to alter this process and predispose to autoimmunity<sup>57</sup>. It is notable that these genes do not have strong *cis* eQTLs that otherwise might be expected to be disease associated.

Finally, we show for the first time that carrying specific HLA alleles regulates gene expression in *trans* in a cell type-specific manner. The monocyte-specific *trans* association of *AOAH* and *ARHGAP24* expression to the presence of the *HLA-DRB1\*04*, *HLA-DRB1\*07* and *HLA-DRB1\*09* alleles is particularly intriguing. These *HLA-DRB1* alleles are associated with expression of *HLA-DRB4*, encoding the *HLA-DR53* superantigen<sup>58</sup>. The reduced expression in monocytes of *AOAH*, an enzyme that hydrolyzes the potent innate immune stimulant lipopolysaccharide, might be anticipated to predispose to inflammation. Whether the additional association of *ARHGAP24*, encoding a protein involved in actin remodeling<sup>59</sup>, a requirement for monocyte spreading and locomotion<sup>60</sup>, is secondary to reduced *AOAH* expression will require further investigation. This study shows that healthy individuals possess cell type-specific expression signatures attributable to the presence of HLA alleles, reflecting the overwhelming importance that HLA status has in everyday immune interactions.

Comparative analysis of highly purified cell subsets has the potential to unearth added intricacy in the nature of eQTLs, underscoring the fact that the majority act in a cell type-specific manner and suggesting that many eQTLs will likely remain undiscovered if analysis focuses on heterogeneous tissues. This study adds strong evidence to support the idea that a large number of eQTLs function in a highly context-specific manner *in vivo* and provides a plausible explanation for a degree of disease susceptibility.

**URLs.** Catalog of Published Genome-Wide Association Studies, <http://www.genome.gov/GWastudies/>; European Genome-Phenome Archive (EGA), <http://www.ebi.ac.uk/ega/>; ArrayExpress Archive, <http://www.ebi.ac.uk/arrayexpress/>; R Project for Statistical Computing, <http://www.r-project.org/>.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

**Accession codes.** Gene expression data are available through ArrayExpress (E-MTAB-945). Genotyping data have been deposited at the European Genome-Phenome Archive (EGA; EGAS00000000109) and are available on request.

Note: Supplementary information is available on the Nature Genetics website.

## ACKNOWLEDGMENTS

We are very grateful to all the volunteers who participated in this study, together with members of the Knight laboratory and A. Hill for their support. We thank A. Auton for his assistance in the identification of probes containing SNPs with European MAF of >0.01% using the 1000 Genomes Project data set and to J. Broxholme for bioinformatic support and mapping of all probes to the reference sequence. We thank G. Gibson, A. Hill and colleagues for critical reading of the manuscript and helpful suggestions. This work was supported by the Wellcome Trust (074318 to J.C.K., 088891 to B.P.F. and 075491/Z/04 to the core facilities at the Wellcome Trust Centre for Human Genetics), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) (281824 to J.C.K.) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre.

## AUTHOR CONTRIBUTIONS

B.P.F. and J.C.K. conceived, designed and initiated the study. B.P.F., J.C.K., K.P. and S.M. established the volunteer cohort and sample collection. B.P.F., K.P., P.E. and S.M. performed the experimental work. B.P.F., J.C.K., J.R. and S.M. analyzed the data. C.L., F.O.V. and S.L. contributed reagents and expertise. A.D. and S.L. performed the HLA imputation. B.P.F. and J.C.K. wrote the paper, with contributions to manuscript editing from other authors. All authors read and approved the manuscript before submission.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
2. Gerrits, A. *et al.* Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* **5**, e1000692 (2009).
3. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
4. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
5. Cheung, V.G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
6. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
7. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
8. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
9. Kwan, T. *et al.* Tissue effect on genetic control of transcript isoform variation. *PLoS Genet.* **5**, e1000608 (2009).
10. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
11. Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat. Genet.* **42**, 62–67 (2010).
12. Small, K.S. *et al.* Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
13. Lyons, P.A. *et al.* Microarray analysis of human leukocyte subsets: the advantages of positive selection and rapid purification. *BMC Genomics* **8**, 64 (2007).
14. Fehrmann, R.S. *et al.* *Trans*-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197 (2011).
15. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
16. Van Laer, L. *et al.* Nonsyndromic hearing impairment is associated with a mutation in *DFNA5*. *Nat. Genet.* **20**, 194–197 (1998).
17. Akino, K. *et al.* Identification of *DFNA5* as a target of epigenetic inactivation in gastric cancer. *Cancer Sci.* **98**, 88–95 (2007).
18. Schoggins, J.W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
19. Xu, H., Manivannan, A., Crane, I., Dawson, R. & Liversidge, J. Critical but divergent roles for CD62L and CD44 in directing blood monocyte trafficking *in vivo* during inflammation. *Blood* **112**, 1166–1174 (2008).
20. Landers, J.E. *et al.* Reduced expression of the Kinesin-Associated Protein 3 (*KIFAP3*) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. USA* **106**, 9004–9009 (2009).
21. Benson, V., Grobarova, V., Richter, J. & Fiserova, A. Glycosylation regulates NK cell-mediated effector function through PI3K pathway. *Int. Immunol.* **22**, 167–177 (2010).
22. Wen, A.Y., Sakamoto, K.M. & Miller, L.S. The role of the transcription factor CREB in immune function. *J. Immunol.* **185**, 6413–6419 (2010).



23. Fischer, U. *et al.* Cloning of a novel transcription factor-like gene amplified in human glioma including astrocytoma grade I. *Hum. Mol. Genet.* **6**, 1817–1822 (1997).
24. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
25. Hakonarson, H. *et al.* A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* **57**, 1143–1146 (2008).
26. Nair, R.P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF- $\kappa$ B pathways. *Nat. Genet.* **41**, 199–204 (2009).
27. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
28. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
29. Plagnol, V., Smyth, D.J., Todd, J.A. & Clayton, D.G. Statistical independence of the colocalized association signals for type 1 diabetes and *RPS26* gene expression on chromosome 12q13. *Biostatistics* **10**, 327–334 (2009).
30. Koldobskiy, M.A. *et al.* p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. *Proc. Natl. Acad. Sci. USA* **107**, 20947–20951 (2010).
31. Lawson, B.R. *et al.* Deficiency of the cyclin kinase inhibitor p21(WAF-1/CIP-1) promotes apoptosis of activated/memory T cells and inhibits spontaneous systemic autoimmunity. *J. Exp. Med.* **199**, 547–557 (2004).
32. Kim, K. *et al.* A regulatory SNP at position –899 in *CDKN1A* is associated with systemic lupus erythematosus and lupus nephritis. *Genes Immun.* **10**, 482–486 (2009).
33. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
34. Tomaru, T. *et al.* Isolation and characterization of a transcriptional cofactor and its novel isoform that bind the deoxyribonucleic acid-binding domain of peroxisome proliferator-activated receptor- $\gamma$ . *Endocrinology* **147**, 377–388 (2006).
35. Hagen, F.S. *et al.* Expression and characterization of recombinant human acyloxyacyl hydrolase, a leukocyte enzyme that deacylates bacterial lipopolysaccharides. *Biochemistry* **30**, 8415–8423 (1991).
36. Dilthey, A.T., Moutsianas, L., Leslie, S. & McVean, G. HLA\*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* **27**, 968–972 (2011).
37. Leslie, S., Donnelly, P. & McVean, G. A statistical method for predicting classical HLA alleles from SNP data. *Am. J. Hum. Genet.* **82**, 48–56 (2008).
38. Lavelin, I. & Geiger, B. Characterization of a novel GTPase-activating protein associated with focal adhesions and the actin cytoskeleton. *J. Biol. Chem.* **280**, 7178–7185 (2005).
39. Rioux, J.D. *et al.* Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl. Acad. Sci. USA* **106**, 18680–18685 (2009).
40. Mukai, R., Suzuki, M., Yabe, T., Hamaguchi, H. & Maeda, H. Identification of the MT3 molecule from HLA-DR4, 7, and w9 homozygous cell lines. *J. Immunol.* **133**, 3211–3219 (1984).
41. Hotfildler, M., Baxendale, S., Cross, M.A. & Sablitzky, F. *Def-2*, -3, -6 and -8, novel mouse genes differentially expressed in the haemopoietic system. *Br. J. Haematol.* **106**, 335–344 (1999).
42. Vandiedonck, C. *et al.* Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. *Genome Res.* **21**, 1042–1054 (2011).
43. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944–947 (2007).
44. Thomas, R. *et al.* HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat. Genet.* **41**, 1290–1294 (2009).
45. Limou, S. *et al.* Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J. Infect. Dis.* **199**, 419–426 (2009).
46. Liu, Y. *et al.* A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* **4**, e1000041 (2008).
47. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nat. Genet.* **42**, 985–990 (2010).
48. Ho, P.Y. *et al.* Investigating the role of the *HLA-Cw\*06* and *HLA-DRB1* genes in susceptibility to psoriatic arthritis: comparison with psoriasis and undifferentiated inflammatory arthritis. *Ann. Rheum. Dis.* **67**, 677–682 (2008).
49. Pereyra, F. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551–1557 (2010).
50. Nica, A.C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
51. Barrett, J.C. *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the *HNF4A* region. *Nat. Genet.* **41**, 1330–1334 (2009).
52. Evans, D.M. *et al.* Interaction between *ERAP1* and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* **43**, 761–767 (2011).
53. Neznanov, N., Neznanova, L., Angres, B. & Gudkov, A.V. Serologically defined colon cancer antigen 3 is necessary for the presentation of TNF receptor 1 on cell surface. *DNA Cell Biol.* **24**, 777–785 (2005).
54. Ruland, J. CARD9 signaling in the innate immune response. *Ann. NY Acad. Sci.* **1143**, 35–44 (2008).
55. Andrés, 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).
56. Elstak, E.D. *et al.* The munc13–4-rab27 complex is specifically required for tethering secretory lysosomes at the plasma membrane. *Blood* **118**, 1570–1578 (2011).
57. Gray, M., Miles, K., Salter, D., Gray, D. & Savill, J. Apoptotic cells protect mice from autoimmune inflammation by the induction of regulatory B cells. *Proc. Natl. Acad. Sci. USA* **104**, 14080–14085 (2007).
58. Knowles, R.W. *et al.* Complexity of the supertypic HLA-DRw53 specificity: two distinct epitopes differentially expressed on one or all of the DR  $\beta$ -chains depending on the HLA-DR allotype. *J. Immunol.* **137**, 2618–2626 (1986).
59. Ohta, Y., Hartwig, J.H. & Stossel, T.P. FilGAP, a Rho- and ROCK-regulated GAP for Rac binds filamin A to control actin remodelling. *Nat. Cell Biol.* **8**, 803–814 (2006).
60. Stossel, T.P. *et al.* Filamins as integrators of cell mechanics and signalling. *Nat. Rev. Mol. Cell Biol.* **2**, 138–145 (2001).

## ONLINE METHODS

**Volunteer collection and PBMC purification.** This study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55). A total of 288 volunteers were recruited in the Oxfordshire area following written informed consent. The median age of the population was 33.1 years (range: 18–62 years), with 125 males and 163 females. Whole blood (50 ml) was collected into anticoagulant EDTA-containing blood collection tubes (Vacutainer System, Becton Dickinson). PBMCs were purified by the Ficoll gradient method using Ficoll-paque from 50 ml of whole blood. PBMCs were washed three times in Hanks-buffered saline solution (HBSS) without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  (Invitrogen), and the number of cells was determined using a hemocytometer.

**Immune cell separation.** Magnetic activating cell sorting (MACS) methods (Miltenyi) were used to positively separate  $\text{CD14}^+$  and  $\text{CD19}^+$  cells as purified monocyte and B-cell populations, respectively<sup>13</sup>, and negatively separate an enriched natural killer cell population ( $\text{CD56}^+$  and  $\text{CD3}^-$ ) from 20 million PBMCs for each cell type, according to the manufacturer's instructions. In circulating blood, the CD19 marker is unique to B cells, being lost upon maturation to antibody-secreting plasma cells<sup>61</sup>. It is also found on follicular dendritic cells, but these reside in secondary and tertiary lymphoid organs. Ficoll gradient purification of PBMCs ensures all polymorphonuclear cells, including neutrophils, are removed before positive selection, and PBMCs were washed after Ficoll separation to remove possible platelet contamination. Cells were kept chilled or on ice after PBMC purification, and purified cells were lysed immediately in RLT reagent (Qiagen).

**Genomic DNA extraction.** Genomic DNA was extracted from 1–2 ml whole blood following the manufacturer's instruction (Gentra Puregene Blood kit) (Qiagen). The DNA was quantified by the PicoGreen dsDNA quantification assay (Invitrogen).

**RNA extraction.** Total RNA was extracted using the RNeasy mini kit from cells collected in the RLT reagent following the manufacturer's instruction (Qiagen). Total RNA was quantified by Nanodrop and Bioanalyzer for a subset following the manufacturers' instruction (Bioanalyzer RNA 6000 Nano kit, Agilent).

**Quantitative PCR.** Single-strand complementary DNA was synthesized by reverse transcription with the SuperScriptIII First-Strand Synthesis System (Invitrogen). qPCR was performed using SYBR Green Supermix on a CFX96 Real-Time PCR Detection System (Bio-Rad). Primer sequences are available on request. Relative gene transcript levels were determined by the  $\Delta\text{C}_T$  method expressed relative to *ACTB*.

**Gene expression array.** Total RNA from monocytes and B cells from each of 288 healthy volunteers were quantified using the Illumina HumanHT-12 v4 BeadChip gene expression array platform that included 47,231 probes. Data backgrounds were subtracted by using the R packages lumi and limma. Principal-component analysis was performed to exclude underlying structure within the data set, including for potential batch and array effects. The raw data were transformed and normalized using the robust spline normalization (RSN) method. Normalized expression data were analyzed for 29,022 probes in B-cell and monocyte samples from 286 and 287 volunteers, respectively, with biological and technical replicate samples showing high concordance ( $r^2 = 0.94$  and  $0.97$ ). All probe sequences were mapped by BLAST to the reference genome (hg18), and probes found to map to more than one location were not used. We used a comprehensive compendium of SNPs in Europeans based on the 1000 Genomes Project<sup>62</sup> (Interim Release 23/11/2010) to remove an additional 6,137 probes found to anneal in regions with SNPs present at a MAF of 1% or greater from our analysis, to minimize

any potential confounding effects. Similarly, probes mapping to non-autosomal locations were excluded from further analysis.

**Genomic DNA genotyping.** SNPs and genetic variants were determined using the Illumina Infinium high-density genotyping bead arrays (Illumina Human OmniExpress-12v1.0 BeadChips, NCBI36 Build) at 733,202 genetic markers. After standard quality control, a total of 651,210 markers were available for analysis (SNP call rate > 96%; MAF > 1%). Underlying genetic stratification in the population was assessed by multi-dimensional scaling using data from the International HapMap Project (CEU, Yoruba from Ibadan (YRI) and Han Chinese from Beijing (CHB) samples) combined with identity-by-state (IBS) cluster analysis (complete linkage agglomerative clustering based on pairwise IBS distance). Four individuals showing potential admixture were identified and removed from the analysis, together with one individual with low genotyping call rates, resulting in a final analysis of 283 individuals (122 males and 161 females). To investigate potential confounding effects from population stratification, the principal components of variance of the genotyping data set were separately correlated with B-cell and monocyte expression profiles. In only 65 probes (0.001% total) that formed 0.001% of *cis* and 0% of *trans* associations was expression significantly correlated with population substructure ( $r^2 > 0.01$ ,  $P < 0.001$ ), which is similar to the number expected by chance and showed that the gene expression profile of monocytes and B cells was not significantly influenced by population stratification in this European cohort.

**Data analysis.** The quality control filtering for the genotyping data and association analysis was performed using PLINK<sup>63</sup> and Haploview<sup>64</sup>, with imputation performed using Impute2 (ref. 65). For *cis* associations, permutation analysis ( $n = 1,000$ ) was performed by switching the phenotype labels. This number corresponds to an approximation of the number of SNPs tested per probe. The distribution of minimum  $P$  values in each permutation was used to identify significance thresholds. For *cis* associations, a permutation  $P$  value of 0.001 was used as the significance threshold for both monocytes and B cells. All significance values presented are based on the linear model unless otherwise stated. In addition to linear analysis, we performed Spearman rank analysis for all *cis* associations and eQTLs not observed to be additionally significant to permutation  $P$  value < 0.001 using this analysis were excluded. For *trans* associations, Wald tests were used to identify the genome-wide associations, and significance thresholds were determined by Bonferroni correction of commonly accepted significance levels. To ensure *trans* associations were only called if robust expression was detected, associations with maximum normalized expression of <6 were excluded. Due to power considerations, only *cis* and *trans* associations to SNPs and genes on autosomes are reported. Statistics were analyzed using R and appropriate packages. Graphs were generated using ggplot2 (ref. 66), and local association plots were generated with LocusZoom<sup>67</sup>. HLA allele imputation was performed using HLA\*IMP<sup>36,37</sup>.

61. Tedder, T.F. & Isaacs, C.M. Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes. A new member of the immunoglobulin superfamily. *J. Immunol.* **143**, 712–717 (1989).
62. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
63. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
64. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
65. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
66. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
67. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).