## ONLINE METHODS

### Volunteer collection and PBMCs 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), 125 male and 163 female. 50 ml whole blood 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 $Ca^{2+}$ and $Mg^{2+}$ (Invitrogen) and the number of cells was determined using a haemocytometer.

### Immune cell separation

Magnetic activating cell sorting methods (MACS) (Miltenyi) were used to positively separate CD14 and CD19 positive cells as purified monocyte and B-cell populations respectively[13], and negatively separate an enriched NK cell population ($CD56^+$ and $CD3^-$) 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[61]. It is additionally 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 prior to positive selection, and PBMCs were washed post Ficoll 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 polymerase chain reaction**

Single strand complementary DNA was synthesized by reverse transcription reaction with SuperScript III First-Strand Synthesis System (Invitrogen). Q-PCR was performed using SYBR Green Supermix on a CFX96 Real-Time PCR Detection System (BioRad). Primer sequences are available on request. Relative gene transcript levels were determined by the delta Ct 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 47231 probes. Data backgrounds were subtracted by using the R packages lumi and limma. PCA was performed to exclude underlying structure within the dataset including for potential batch and array effects. The raw data was transformed and normalized using the robust spline normalization method (RSN). Normalized expression data was analysed for 29022 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, 0.97). All probe sequences were BLAST mapped 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 Genome Project[62] (Interim Release 23/11/2010) to remove a further 6137 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 genetics variants were determined using the Illumina Infinium high-density genotyping bead arrays (Illumina HumanOmniExpress-12v1.0 Beadchips, NCBI36 Build) at 733202 genetic markers. After standard QC, a total of 651210 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, YRI and CHB samples) combined with IBS cluster analysis (complete linkage agglomerative clustering based on pairwise identity-by-state (IBS) distance). Four individuals demonstrating 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 dataset 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 demonstrated 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 of the genotyping data and association analysis were performed using PLINK[63] and Haploview[64] with imputation performed using Impute2[65]. For cis-associations, permutation analysis (n = 1000) 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 eQTL 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-association were only called if robust expression was detected, associations where maximum normalized expression <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 (http://www.r-project.org/). Graphs were generated using ggplot2[66] and local association plots were generated with Locus Zoom[67]. HLA allele imputation was performed using HLA*IMP[36,37].