Supplementary Materials for

Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression

Benjamin P. Fairfax,* Peter Humburg, Seiko Makino, Vivek Naranbhai, Daniel Wong, Evelyn Lau, Luke Jostins, Katharine Plant, Robert Andrews, Chris McGee, Julian C. Knight*

*Corresponding author. E-mail: bfairfax@well.ox.ac.uk (B.P.F.); julian@well.ox.ac.uk (J.C.K.)

Published 7 March 2014, *Science* **343**, 1246949 (2014)
DOI: 10.1126/science.1246949

The PDF file includes:

Other supplementary material for this manuscript includes the following:

Materials and Methods

Volunteer collection and PBMC purification

This study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55). A total of 432 volunteers of European ancestry were recruited in the Oxfordshire area following written informed consent. The cohort consisted of 194 males and 238 females with a median age of 30 years (mean, 33y; interquartile range 19-56y). Whole blood was collected into anticoagulant EDTA-containing blood collection tubes (Vacutainer System, Becton Dickinson). PBMCs, purified using Ficoll-paque density gradients from 50 ml of whole blood, were washed three times in Hanks-buffered saline solution (HBSS) without Ca2+ and Mg2+ (Invitrogen), and the number of cells was determined using a hemocytometer.

Monocyte separation

Magnetic-activated cell sorting (MACS - Miltenyi) was performed to positively separate CD14+ monocytes from 20 million PBMCs, as per manufacturer's instructions. This is designed to provide a sample of ~99% purity. Flow cytometry was performed on a subset of samples with similar purities observed. Typical monocyte yields ranged from 1-10x10^6 monocytes per individual, of which 5x10^5 untreated cells were immediately suspended in RLT reagent (Qiagen) to form the naive subset and snap frozen for RNA extraction at a later date. All other monocytes were resuspended at a concentration of 1x10^6/ml in 400µl pre-warmed RPMI-1640 complete medium, supplemented with 20% FCS, Penicillin/ Streptomycin and L-Glutamine. Cells were rested overnight (16h) at 37°C, 5% CO2 in 5ml non-adherent polypropylene cell-culture tubes (BD Biosciences) prior to stimulation assays.

Stimulation assays

For stimulation assays, 4x10^5 monocytes were exposed to 20ng/ml of ultrapure LPS (cat# tlrl-pelps, Invivogen) for either 2 or 24h. Alternatively, monocytes were exposed to Interferon-γ (IFNγ, cat# 285-IF, R&D Systems) at a concentration of 20ng/ml for 24h. The number of treated samples per individual depended on total number of cells purified, and the priority on sample accrual was collation of IFNγ treated followed by 24h LPS and then 2h LPS respectively. Experiments were terminated simultaneously ensuring identical incubation periods for all samples. A subset of samples was kept in the incubator without stimulation throughout this period to ascertain incubator effects. All experiments were completed within 48h of blood sample collection.

Genomic DNA extraction

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

RNA extraction

Total RNA was extracted in batches using the RNeasy mini kit from cells collected in the RLT reagent following the manufacturer's instruction utilizing an additional DNase step to minimize contamination with genomic DNA (Qiagen). Total RNA was quantified by

Nanodrop and Bioanalyzer for a subset of samples (Bioanalyzer RNA 6000 Nano kit, Agilent).

Genotyping

Genotyping was performed using Illumina HumanOmniExpress-12v1.0 Beadchips with coverage of 733,202 separate markers. PLINK (*54*) was used for initial genotyping QC, including identity by descent (IBD) analysis to ensure individuals were unrelated, and heterozygosity testing and multi-dimensional scaling to identify population outliers compared to HapMap populations. For SNP filtering we used sample and SNP call rates >98%. We used a MAF of 4% and a Hardy-Weinberg equilibrium threshold set at $1x10^{-6}$. We removed 11 individuals due to failure in one or more QC criteria, resulting in 421 individuals for further analysis with genotyping data available at 609,704 SNPs. Genotyping files were subsequently converted from PLINK output for use in MatrixEQTL.

Gene expression analysis

Total RNA from naive and, where available, treated monocytes from each individual was quantified using the Illumina HumanHT-12 v4 BeadChip gene expression array platform with 47,231 probes. Of these, 28688 probes correspond to coding transcripts with well established annotations (RefSeq Content NM), 11121 to coding transcript with provisional annotation (XM), 1752 to non-coding transcript with well-established annotation (NR), 2209 to non-coding transcript with provisional annotation (XR), and 3461 to experimentally confirmed mRNA sequences that align to EST clusters from UniGene. Analysis of 1488 arrays was performed in three separate batches, the first consisting of 288 naive monocyte samples as previously described (*8*). The second batch consisted of 636 samples from the same individuals randomised to individual and treatment across array. A third batch consisted of 564 samples composed of 144 further naive monocyte samples as well as 420 treated samples which were similarly randomised to individual and treatment across arrays. cRNA synthesis, clean-up and hybridization were as per the manufacturers instructions. Arrays were read and data was exported using Beadstudio software (Illumina). Samples were subsequently processed in R using the R packages lumi, limma and ComBat (*55-57*). Principal components analysis (PCA) was performed to identify samples within each treatment cohort with outlier expression for removal from further analyses. After QC and removal of duplicates and sample outliers, 1438 independent samples were suitable for further analysis. Raw data were transformed and normalized using robust spline normalization (RSN) within lumi. Samples were then corrected for batch effects using ComBat. A linear model incorporating data from 59 untreated, incubated samples and 421 untreated, non-incubated samples was used to estimate the effect of incubation on gene expression. This was then regressed from all treated samples to minimise the influence of incubator effects on the analysis. Differential gene expression analysis was performed using the limma package. Pathway analysis was performed using Ingenuity Pathway Analysis (IPA, Ingenuity Systems) to define gene networks, significant upstream regulators and relationships with canonical signaling pathways. For network analysis of trans gene lists, a global molecular network based on the IP Knowledge Base (IPKB) was used with a P-value based on the fit of the trans genes and list of biological functions within IPKB.

Probe filtering

Normalization and correction for batch and incubator effects were carried out on all probes. Probes were subsequently selected for eQTL analysis. As previously described (*8*) probe sequences mapping to more than one genomic location using BLAST were excluded from eQTL analysis. Additionally 6,137 probes found to anneal in regions with SNPs present at a MAF of at least 1% in the European population of the 1000 Genomes Project were excluded from eQTL analysis to minimize potential confounding effects, as were probes mapping to non-autosomal locations. Remaining probes were included in eQTL analysis if they were present (Illumina detection P-value <0.01) in >2.5% of samples from a specific treatment or alternatively >5% of all samples. This resulted in 15,421 probes for analysis, mapping to 11,168 independent Entrez ID.

eQTL analysis

To account for confounding variation in the gene expression data we carried out a PCA for each of the four conditions. For each set of eQTL analyses the dominant principal components (PC) were chosen for inclusion as covariates in the model such that the number of significant associations was maximised across all conditions. This resulted in the use of 30 (cis-analysis of 228 shared samples), 16 (trans-analysis of 228 shared samples), 40 (cis-analysis of full datasets) and 20 PC (trans-analysis of full datasets) as shown in Figure S2. The eQTL analysis was performed using the R package MatrixEQTL (*18*) using an additive linear model. SNPs were included in the cis-analysis if they were located within 1Mb of the gene expression probe under consideration. Correspondingly, associations between SNPs and probes outside this window were deemed trans. False discovery rates reported for the resulting associations are as reported by MatrixEQTL. For single SNP eQTL analysis, the same procedure was carried out, but with all analysis performed on a single SNP.

ChIP-seq analysis

Monocytes were purified from donor blood cones from two healthy volunteers and $5x10^6$ cells were used per stimulation assay (n=6). Cells were either untreated or treated as per LPS2 and IFNγ stimulations for the eQTL analysis. At the end of the experiment, cells were cross-linked and pellets frozen. Chromatin immunoprecipitation was performed with cross-linked cells lysed on ice in a lysis buffer (140 mM NaCl, 50 mM HEPES pH 7.5, 1 mM EDTA pH 8, 0.5 Mm EGTA pH 8, 0.5 % v/v NP-40, 0.25 % v/v Triton-X and 10 % glycerol). Lysis buffer was supplemented with 1 X Complete Protease Inhibitor cocktail (Roche), 0.5 mM Phenylmethylsulfonate Fluoride (PMSF, Sigma), 10 µg/µL pepstatin (Sigma) and 10 µg/µL leupeptin hemisulfate (Sigma) prior to usage. Nuclear pellets were collected following brief centrifugation and washed using a buffer (10 mM Tris pH 8, 1 mM EDTA pH 8, 2 % v/v Triton-X, 0.2 % v/v Na-Deoxycholate, 10 % glycerol) similarly supplemented as described above. Sonication to fragment cross-linked chromatin was carried out using the BioRuptor (Diagenode) at high power over 12 cycles of 30 s each. Fragmented chromatin was incubated overnight with Dynabeads (Life Tech) pre-coated with 10µg anti-IRF2 antibody (Santa Cruz - IRF-2 (C-19): sc-498). Progressive washing of the chromatin-Dynabead mixture was then carried out in 10 mM Tris pH 8, 1 mM EDTA pH 8, 1 % v/v NP-40, 0.7 % v/v Na-Deoxycholate, 0.5 M LiCl,

supplemented as previously but using only 0.5 X Complete Protease Inhibitor. Cross-links were reversed after an overnight incubation in elution buffer (10 mM Tris pH 8, 1 mM EDTA pH 8 and 1 % SDS). Immunoprecipitated DNA was obtained using phenol-chloroform extraction and purified following precipitation in ethanol. DNA was quantified using the High Sensitivity Qubit system (Life Tech) and the fragmentation profile was assessed using a DNA HS 2200 TapeStation chip (Agilent).

ChIP libraries were prepared using the NEBNextTM DNA Sample Prep MasterMix Set 1, according to manufacturer's specifications, but with the following amendments: end repair was carried out using 5µl of NEBNEXT End Repair Reaction Buffer and 0.5µl of NEBNext End Repair Enzyme Mix. A-tail was carried out using 5 µl of NEBNEXT A-tail Reaction Buffer and 0.3 µl of NEBNext Klenow Fragment (3´→5´ exo-) Enzyme Mix. Ligation was carried out using 10 µl of NEBNEXT Quick Ligation Reaction Buffer and 0.5ul of NEBNext Quick T4DNA Ligase Enzyme Mix. The NEBNext adapters were diluted 100-fold and the amount added was adjusted if less than 5 ng of ChIP material was available. No size selection was carried out. PCR was done using the following reaction mix with 18 cycles of PCR: DNA (36 µl), 5 X Phusion buffer (10 µl), custom paired end (PE) primer 1 (1 µl), Custom PE primer 2 (1 µl), dNTP mix (1.5µl), Phusion polymerase (0.5 µl). The clean-up after amplification was done at 1:0.85 (Ampure beads: DNA). The concentration of each library was determined by real-time PCR using Agilent qPCR Library Quantification Kit and a MX3005P instrument (Agilent). Sequencing was performed on an Illumina HiSeq2000 using 50 bp paired end reads.

Reads were mapped using Stampy (*58*) and peaks were called with MACS2 (*59*) using non-immunoprecipitated DNA as a control for enrichment. Analysis was performed using BEDTools (*60*). Only peaks found in both samples were used for comparison with trans-associated probes to rs13149699. We observed 481 independent peaks that overlapped by a minimum of 10% in the 2 biological replicates in the untreated state; 170 after 2h LPS2 and 154 after IFNγ stimulation. The number of peaks found within defined distances (as per Figure 4F) of trans associated probes was counted and compared to those found within the same distance of non-trans associated probes that had been similarly tested for eQTL. A Fisher's exact test was performed to test significance of number of occurrences per genomic window flanking the trans-associated probes versus non-associated probes.

Imputation of genotypes

Genotypes were pre-phased with SHAPEIT2 (*61*) and missing genotypes were imputed with IMPUTE2 (*62*) in 5Mb chunks against the 1000 genomes reference panel (*63*). Summary statistics for each chunk were obtained using SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) and sites with an information score of less than 0.9 or significant departure from Hardy-Weinberg equilibrium (P-value < $1x10^{-3}$) were excluded from further analysis. Genotype probabilities for all remaining sites were converted into dosage estimates and formatted for use with MatrixEQTL.

Conditional analysis

To investigate the presence of local associations in addition to the most significant SNP reported for each probe we carried out a conditional analysis with imputed genotypes. To this end the cis-eQTL analysis was repeated for all gene expression probes that had produced an association with a p-value of no more than $5x10^{-8}$, conditioning on the initial association. If no additional independent eQTL are present it is expected that this will not identify any additional significant associations. Conversely, any associations that are independent of the first one are expected to remain and may increase in significance. This process was iterated until no further significant associations remained (P-value < $5x10^{-8}$). To confirm the presence of an independent signal this process was repeated for the second and third most significantly associated SNP for a given probe. Only examples where the secondary signal was present in all instances are reported.

The resulting eQTL were further characterised by identifying the list of SNPs associated with each of the peak eSNPs. To this end all SNPs that are showing significant association with gene expression are assigned to the eSNP with which they have the highest r^2. Peaks identified in this way were matched across conditions and the list of SNPs observed in all treatments in which the peak is present are reported.

Overlap between eQTL signals and GWAS catalog

All traits listed in the Catalog of Published Genome-Wide Association Studies (http://www.genome.gov/gwastudies/ date accessed 10 Jan 2013) were extracted and expertly curated into phenotypic groupings blind to GWAS or eSNP data. Groups were defined based on organ-specificity of a particular trait, disease process or type. These included cardiovascular, respiratory, gastroenterological, urological, rheumatological, neurological, renal, endocrine, hematological, dermatological, bone, cancer, immunity and inflammation, autoimmune, allergy, genetic, viral infection, bacterial infection, parasitic disease, measurement, physiological, metabolic, chronic or degenerative disease, reproduction, drug related. We recognised that classification of a given trait was possible into multiple groups and to capture this diversity, for each trait we assigned trait membership into three possible groups.

A reported GWAS SNP was considered to coincide with an eQTL identified during the conditional analysis of cis-associations if the GWAS SNP itself or any SNPs with r2 >0.8 with this SNP were part of one of the eQTL peaks (P <5 $x10^{-8}$). Enrichment of peak eQTLs for individual GWAS categories as well as overall enrichment of GWAS SNPs was tested with Fisher's exact test by comparing the overlap obtained for peak eQTLs to that observed for all SNPs tested for cis-associations. Publicly available summary statistics for a 200kb window around *CARD9* from the latest CD GWAS meta-analysis (*46*) were extracted using the Ricopili tool (http://www.broadinstitute.org/mpg/ricopili/). LD calculations were performed using Plink (*54*) and the Phase 1 1000 Genomes data (*63*). Approximate conditional analysis, based on summary statistics and LD properties, was carried out using the method described in (*46*).

Flow cytometry

To validate the changes in class II HLA observed at transcriptomic level with IFNγ and LPS treatment, we assessed class II expression on primary monocytes by multiparametric

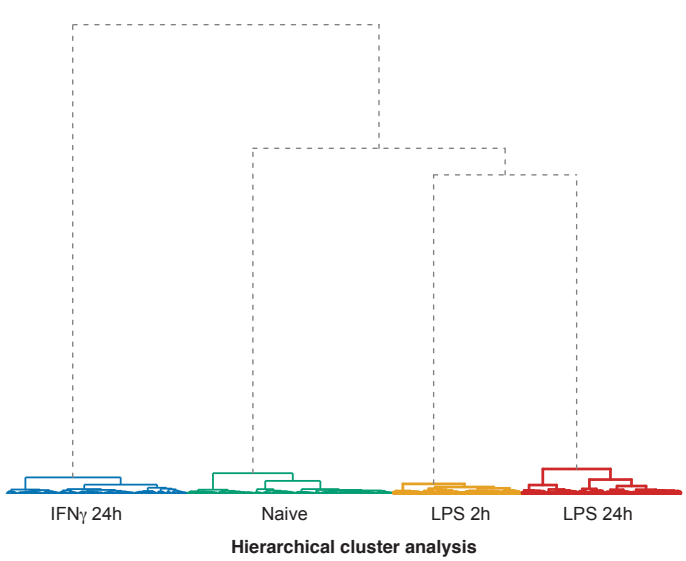
flow cytometry. Monocytes were stimulated as described prior to being washed and stained with Violet Viability Dye (Invitrogen) as per the manufacturers instructions. To reduce non-specific staining a three step procedure adapted from a protocol kindly shared by R.Apps (NCI, NIH) was used. Human IgG was added to each tube, followed by mouse anti- human antibodies directed against HLA-DP (clone B7/21, Abcam), HLA-DQ(clone SPV-L3, Abcam), HLA-DR (clone L243, BD Biosciences) or all class II alleles (clone Tu39 and SK10/SPV-L3 combined) or an isotype control. After 20 minutes, cells were washed and stained with sheep-anti-mouse PE-labelled antibody (Sigma) and sheep IgG.

After a further 20 minutes, cells were washed and stained with mouse-anti-human CD14-FITC (clone 61D3, eBiosciences) and mouse IgG (Sigma). Cells were fixed with the paraformaldehyde containing Reagent A (Life Technologies), and data was acquired immediately on a FACSCanto (BD) machine. A minimum of 10000 gated monocytes were acquired for each sample. Fluorescence-minus-one controls were used to set gate thresholds.

Data presentation

Graphs were generated using ggplot2 (*64*) and R base packages. Boxplots of gene expression data show values corrected for confounding variance using gene expression PC as per eQTL analysis. Circular plots were created using the ggbio R package (*65*) and local association plots were generated with Locus Zoom (*66*).

A



B

Canonical pathways	Significance
Naive vs. LPS2	
TREM1 Signaling	9.43E-14
Interferon Signaling	3.76E-11
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1.34E-10
Toll-like Receptor Signaling	3.92E-10
Role of PKR in Interferon Induction and Antiviral Response	1.88E-09
Death Receptor Signaling	5.70E-09
IL-10 Signaling	9.82E-07
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.66E-08
TWEAK Signaling	1.75E-08
NF-κB Signaling	1.97E-08
Naive vs. LPS24	
TREM1 Signaling	2.27E-11
IL-8 Signaling	2.50E-09
Granulocyte Adhesion and Diapedesis	2.02E-08
IL-17A Signaling in Fibroblasts	2.46E-08
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.47E-08
IL-10 Signaling	2.54E-08
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1.31E-07
Toll-like Receptor Signaling	3.91E-07
Dendritic Cell Maturation	4.04E-07
Hepatic Fibrosis / Hepatic Stellate Cell Activation	1.39E-06
Naive vs. IFNγ	
Interferon Signaling	2.46E-12
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	7.03E-10
TREM1 Signaling	3.18E-09
IL-8 Signaling	7.38E-07
Antigen Presentation Pathway	1.15E-06
Crosstalk between Dendritic Cells and Natural Killer Cells	5.78E-06
Complement System	1.95E-06
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.41E-06
HMGB1 Signaling	2.90E-06
p38 MAPK Signaling	4.19E-06

C

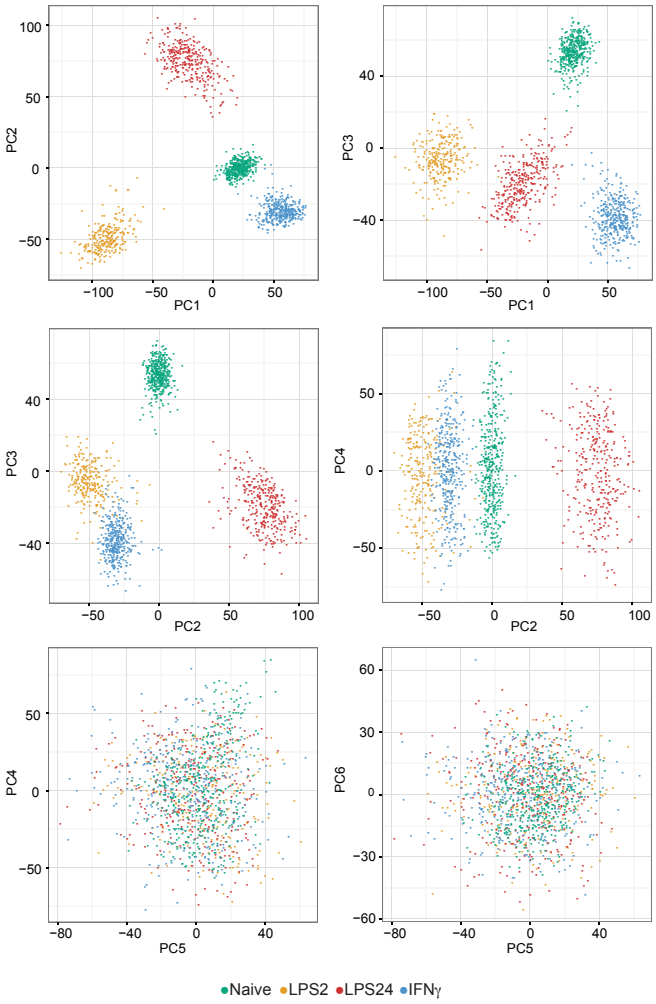


Fig. S1.

Differential gene expression following LPS and IFNγ treatment. (**A**) Hierarchical cluster analysis of expression profiles demonstrates clustering of data according to monocyte treatment. (**B**) Pathway analysis. Top canonical pathways differentially

expressed between naïve and specific stimulated cells are shown. Gene sets of 2385 (LPS 2h), 2010 (LPS 24h) and 1644 (IFNγ) significant differentially expressed genes compared to naïve using a $\log_2$ fold change cut-off of +/-0.5 (Table S1) were used in pathway analysis. For monocytes treated with IFNγ, Ingenuity Pathway Analysis (IPA) (*67*) identified IFNγ as the top predicted upstream regulator ($P= 2.3x10^{-62}$), confirming the integrity and relevance of the biological model system used. LPS was the most significant predicted regulator post 2h LPS ($P= 3.1x10^{-50}$), whilst after 24h LPS, LPS ranked second ($P= 6.7x10^{-49}$) to TNF ($P= 2.7x10^{-50}$), indicating this time point additionally represents induced cytokine associated changes. Interferon signaling was highly significant in IFNγ and 2h LPS treated cells (P_{IFNg} $2.5x10^{-12}$, P_{LPS2} $3.8x10^{-11}$). Whilst LPS stimulated cells shared TREM1 signaling as the most significant pathway, the IL8 pathway was highly significant after 24h LPS. Consistent with the known profound changes in transcription accompanying transition to an endotoxin tolerant state (*68*), marked differential gene expression between early and late responses to LPS were also apparent (Table S1). (**C**) Principal components analysis of expression data used in eQTL analysis. The first 3 PC account for inter-treatment effects, with no clustering along treatments observed in subsequent PCs.

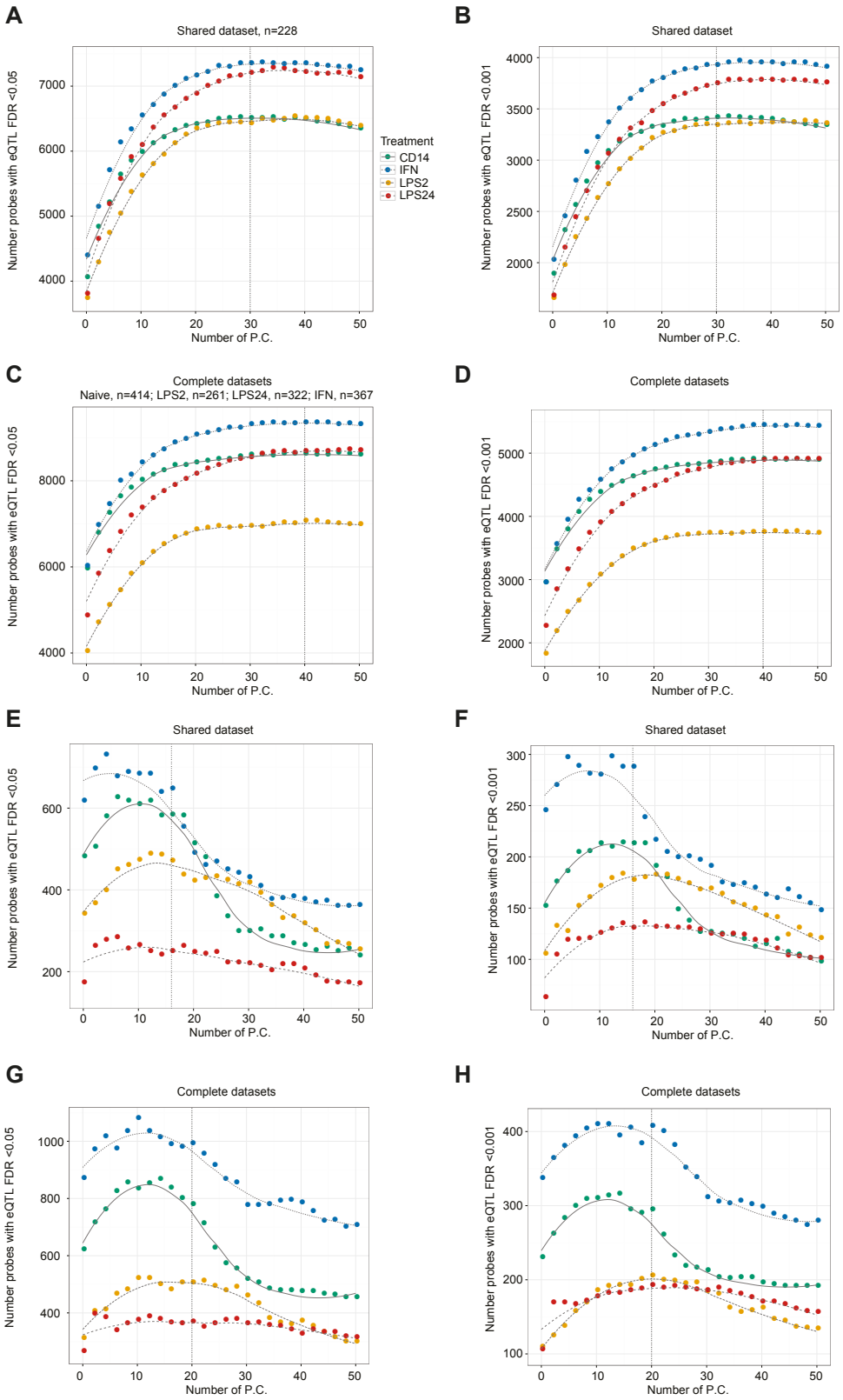


Fig. S2

Analysis of the effects of incorporation of non-genomically associated principal components (PC) as covariates on observed trans and cis eQTL. In keeping with other analyses (*18, 19*), we observe incorporating expression data PC as surrogate

variables in eQTL analysis can adjust for confounding factors. Graphs demonstrate the effect of incorporating different numbers of expression data PCs as covariates on the number of independent probes demonstrating local (cis) eQTL (**A**-**D**) and trans eQTL (**F**-**J**). Dotted lines indicate number of PC incorporated in final analyses presented, based on numbers of associations across datasets with FDR<0.001.

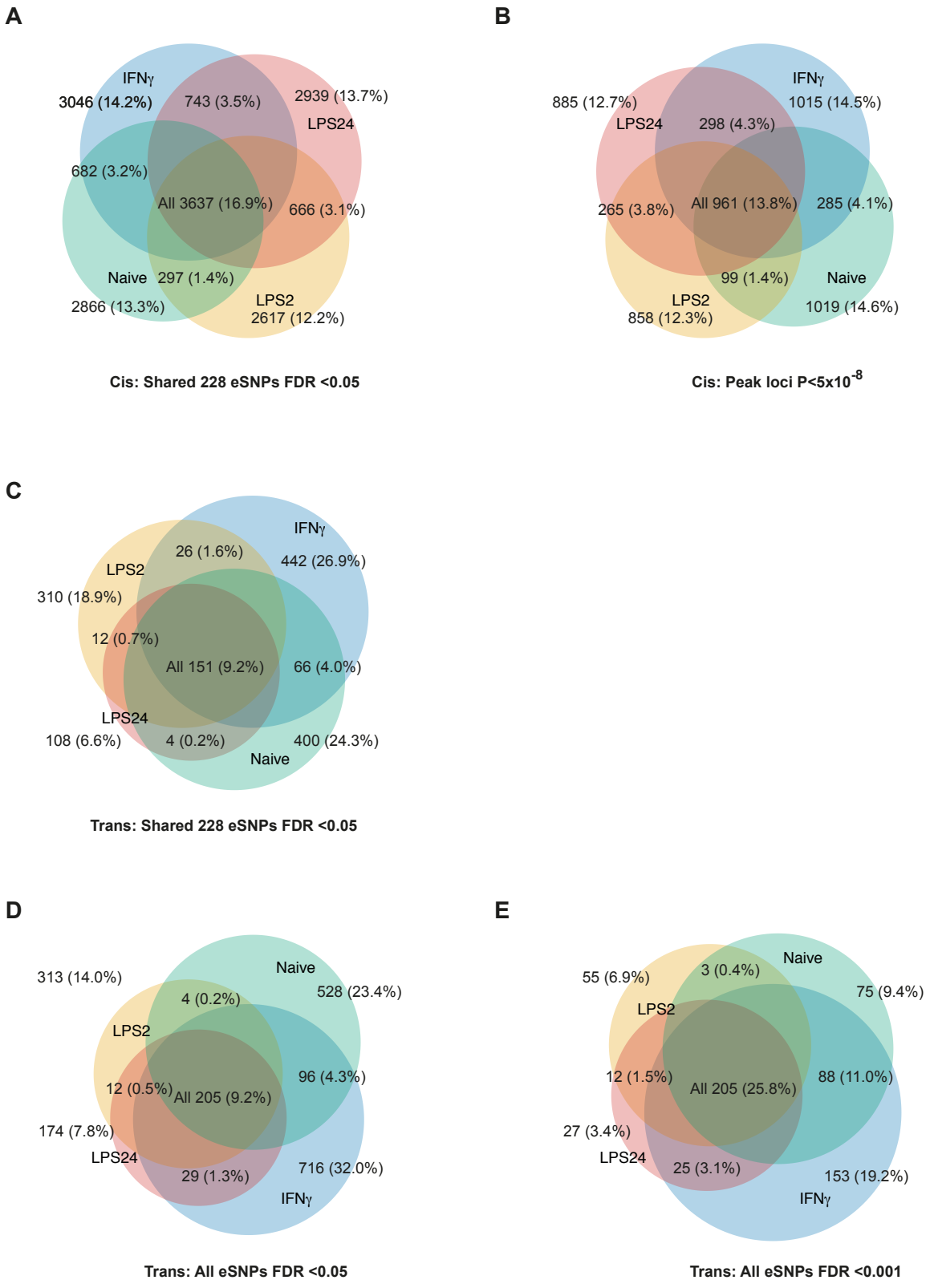


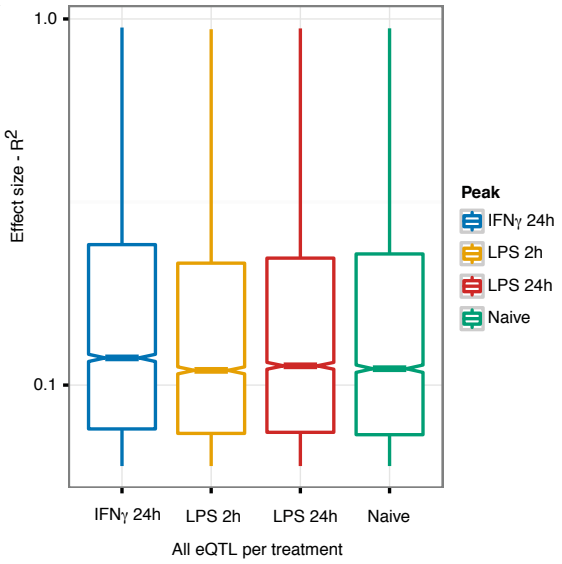
Fig. S3

Venn diagrams of associations observed across the cohort (A) Cis associations based on lead eSNP. The most significant probe-SNP combination (eSNP) was identified per matched dataset (n=228), and the overlap of eSNP significance across datasets is reported

(here and for other panels, overlaps shown for single datasets, intersection between 2 and intersection across all 4 – for full set of proportions see Table S3, sheet 6). **(B)** Cis associations based on peak genomic regions associating with probe expression ($P<5x10^{-8}$), correcting for underlying LD structure are reported using matched dataset (n=228). **(C)** Trans eQTL based on lead eSNP for matched dataset (n=228). **(D & E)** Peak trans associated eSNPs across all samples (naïve n=414; LPS2 n=261; LPS24 n=322; IFNγ n=367) at FDR 0.05 & 0.001 respectively.

B

A



Condition	Minimum	1st Qu	Median	Mean	3rd Qu	Maximum	P vs naive	P vs LPS2	P vs LPS24	P vs IFN
Naive	0.060	0.073	0.111	0.183	0.228	0.943		0.53	0.84	0.003
LPS 2h	0.060	0.074	0.110	0.176	0.215	0.937	0.53		0.53	0.00002
LPS 24h	0.060	0.074	0.113	0.180	0.222	0.942	0.53	0.53		0.003
IFNγ 24h	0.060	0.076	0.119	0.190	0.242	0.948	0.003	0.00002	0.003	

Fig. S4

Effect sizes for observed eQTL by condition. Shared dataset (n=228) analysed. **(A)** Graphical plot **(B)** Table showing distributions of effect sizes and assessment of differences between distributions (Pairwise Wilcoxon rank-sum test corrected for multiple comparisons).

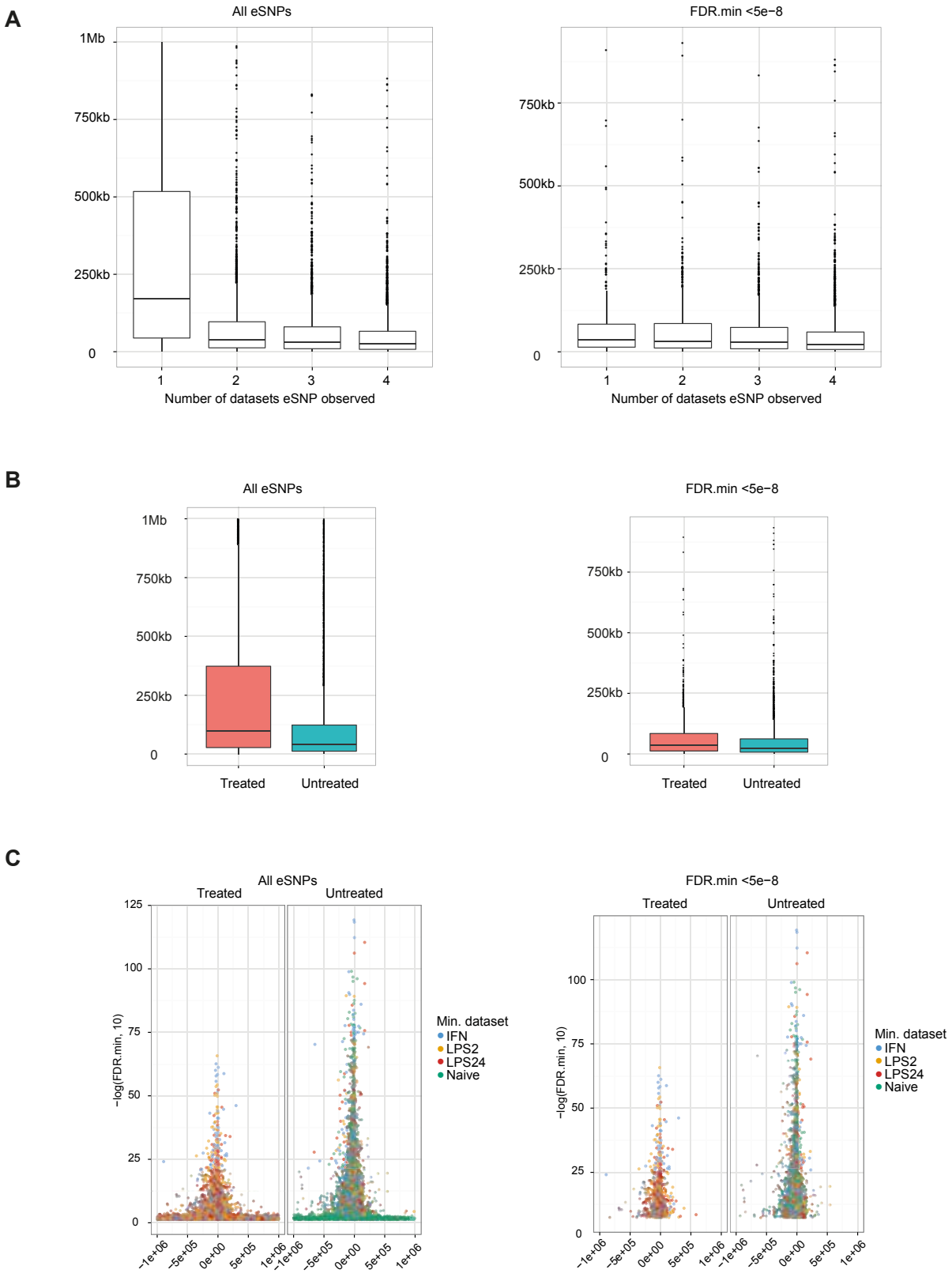


Fig. S5

Analysis of induced eQTL and distance from the TSS. (A) Distance from TSS is significantly influenced by the number of datasets the eSNP is observed in. The left hand panel demonstrates the effect across all eSNPs (Kruskal Wallis test: $P<2.2x10^{-16}$), whilst

the right panel is for eSNPs where FDR adjusted $P<5x10^{-8}$ (Kruskal Wallis test: $P<2.9x10^{-11}$). (**B**) Treatment specific eSNPs are more distal to those that are constitutively active, but not compared to those unique to the naïve state. The left panel shows for all eSNPs (median treatment observed distance 99.7kb vs. 39.6kb in naïve state, Wilcoxon-rank sum test: $P<2.2x10^{-16}$) and the right panel for eSNPs FDR$<5x10^{-8}$ (median 35.1kb vs. 23.9kb, $P<9.5x10^{-12}$). (**C**) Plots of all eSNPs (i.e. the most strongly associated SNP per probe in probes identified with eQTL), with the color reflecting the most significant dataset. Left panels illustrate eSNPs where no significant observation was made in the naïve state, whilst right panels illustrate all eSNPs where an observation was made in the naïve state. Note the closer to the TSS, typically the more significant the association.

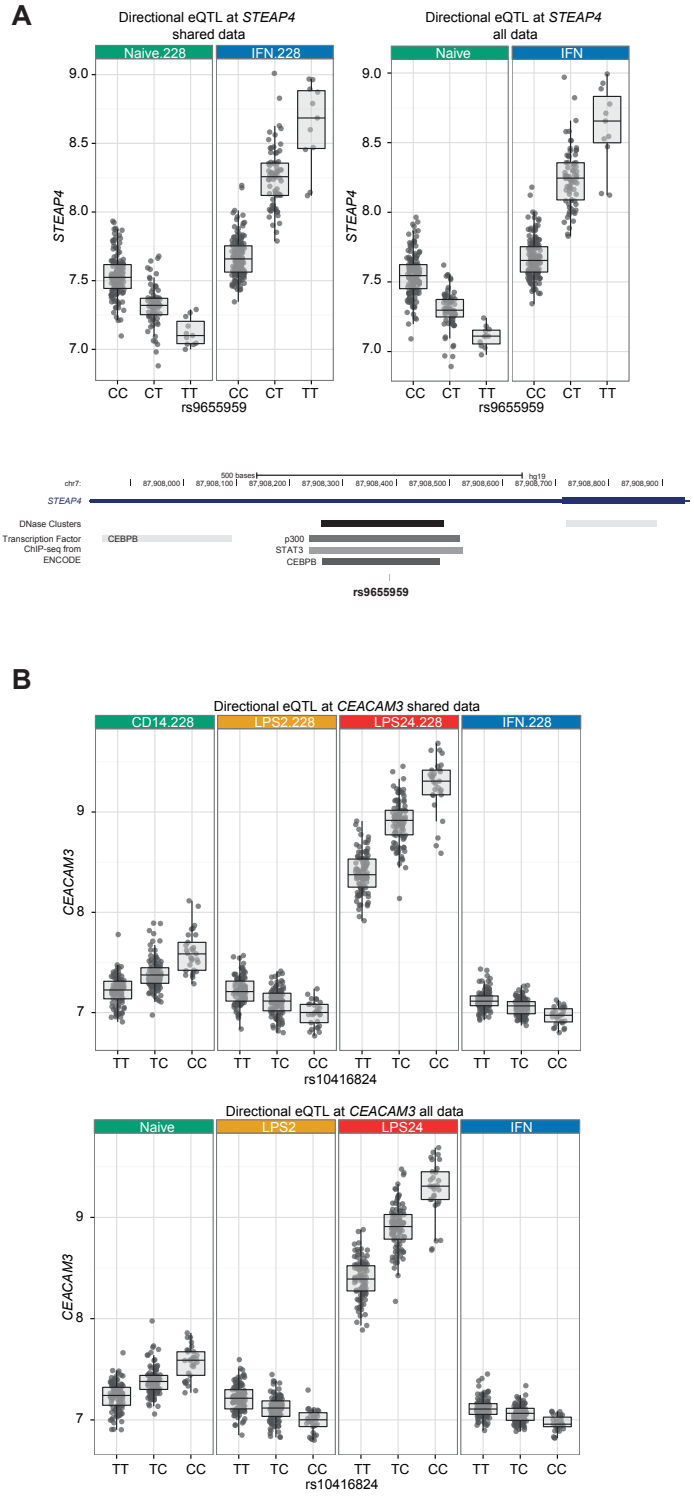


Fig. S6

Directional cis-eQTL. Directional cis-eQTL illustrated by panels (**A**) Cis associations to *STEAP4* peak at rs9655959, a SNP lying within a significant ChIP-seq peak for STAT3, P300 and CEBPB in the 3'UTR of *STEAP4*, in both naïve and IFNγ treated monocytes, but with opposing directions of effect (Shared datasets: $P_{Naïve}= 1.9x10^{-22}$, $P_{IFNg}= 3.2x10^{-56}$; all samples: $P_{Naïve}= 1.0x10^{-39}$, $P_{IFNg}= 8.4x10^{-93}$). (**B**) *CEACAM3* similarly demonstrates

highly significant directional effects across conditions. Peak eQTL in both datasets mapping to the intronic SNP rs10416824 (Shared datasets: $P_{Naïve}= 7.5x10^{-21}$, $P_{LPS2}= 5.5x10^{-12}$, $P_{LPS24}= 7.1x10^{-46}$, $P_{IFNg}= 1.5x10^{-11}$; all samples: $P_{Naïve}= 2.7x10^{-45}$, $P_{LPS2}= 2.7x10^{-14}$, $P_{LPS24}= 1.4x10^{-68}$, $P_{IFNg}= 3.1x10^{-25}$).

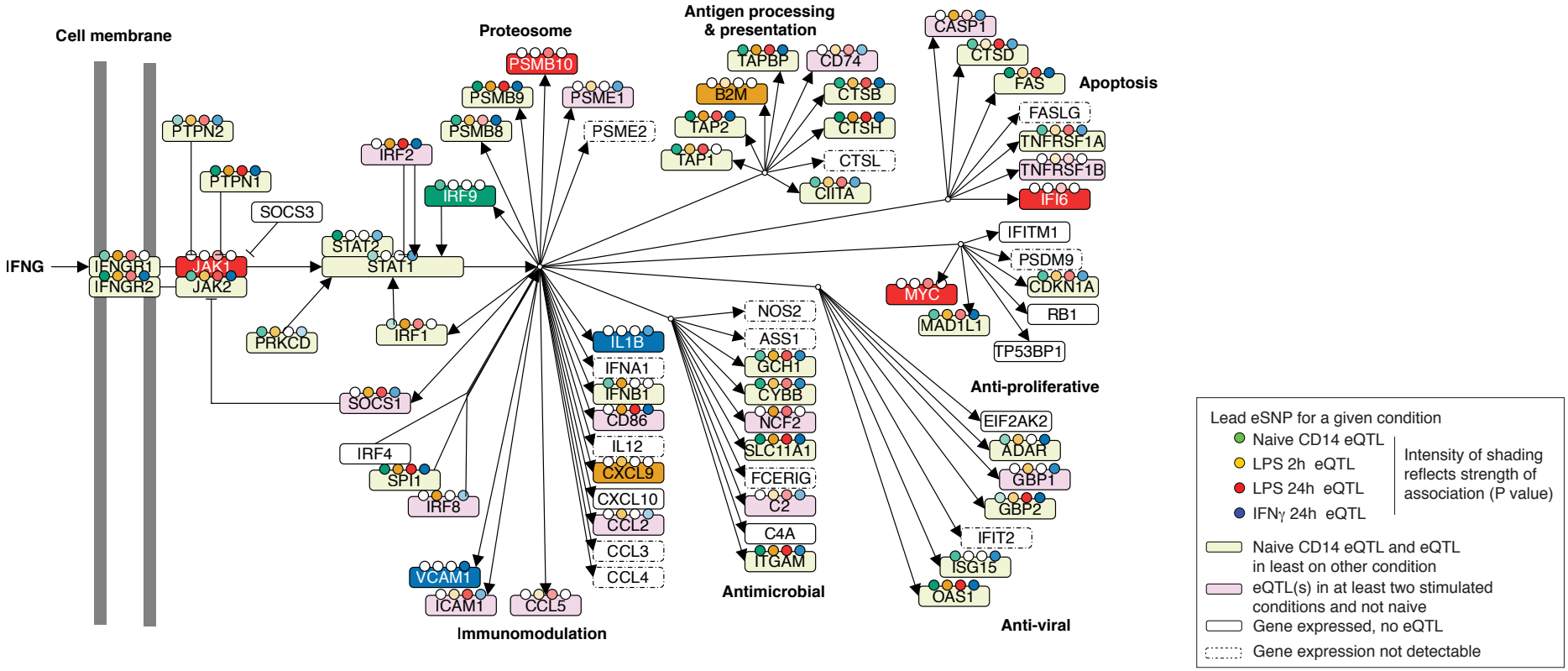


Fig. S7

Pathway analysis. eQTL described for lead SNP-probe associations observed for a given gene in 228 individuals (shared dataset). An interactive version of the figure with information about SNP identifier (rs number), P-value for association, gene information and ensemble links is available at http://galahad.well.ox.ac.uk/iQTL/.

(A) eQTL involving IFNγ signaling pathway. Genes defined and pathway drawn from curated Type II IFN signaling pathway (*69*) and list of IFNγ regulated genes (*70*). We find evidence of eQTL specific to IFNγ treatment including the cytokine gene *IL1B* (rs13005572 P=2.2x10^{-6}) and *VCAM1* (rs3176860 P=6.7x10^{-16}) (encoding a molecule critical to leukocyte-endothelial cell adhesion and signal transduction). For other genes, observed eQTL significantly increase on IFNγ treatment such as *JAK2* (rs2230724 P=8.15x10^{-16}) or show different SNP associations such as *STAT1* specific to IFNγ treatment (rs17346475 P=1.18x10^{-6}) or to the naïve state (rs12617890 P=3.5x10^{-4}). A number of constitutively present eQTL are noted, including for the *IFNGR2* gene (P=5.17x10^{-32}) which is less significant on stimulation of cells, and the protein tyrosine phosphatase *PTPN1* (rs4602269 P=6.9x10^{-14}) involved in regulating the

cellular response to interferon. There are also many striking associations specific to LPS treatment consistent with known cross-talk between pathways; for example involving the *IFNGR1* gene (LPS 2h rs7749390 $P=6.79x10^{-28}$), *IFNB1* (rs2275888 LPS 2h $P=5.29x10^{-17}$), and the suppressor of cytokine signalling gene *SOCS1* (rs2291657 LPS 2 & 24h $P=2x10^{-11}$). For *STAT2* the eQTL observed in the naïve state is lost on LPS induction at 2h (rs10876882 $P=5.75x10^{-18}$). The IRF family is found to include many genes with stimulation specific eQTL, notably in monocytes stimulated with LPS for 2h with association seen for *IRF1* (rs1050152 $P=1.3x10^{-14}$) and *IRF8* (rs17445836 $P=2.4x10^{-16}$). Strikingly, both these eSNPs are also implicated in GWAS, for multiple sclerosis (*41*) and Crohn's disease (*43*) respectively. A number of other genes show eQTL following LPS or IFNγ treatment but not in the naïve state including *IRF2* (maximal with LPS 24h rs13149699 $P=2.9x10^{-35}$); *CD86* (LPS 2h rs12491991 $3x10^{-24}$ $P=3x10^{-24}$) which is in strong LD with the GWAS signal for multiple sclerosis (*71*); and *MYC* where the eSNP (LPS 24h rs2019960 $P=1.6x10^{-5}$) is also the reported GWAS SNP for multiple sclerosis (*71*) and Hodgkin's lymphoma (*72*). A strong eQTL seen only in naïve cells for *CDKN1A* (rs1321313 $P=7x10^{-7}$) is associated with colorectal cancer on GWAS and is predicted to disrupt a Klf4 binding site.

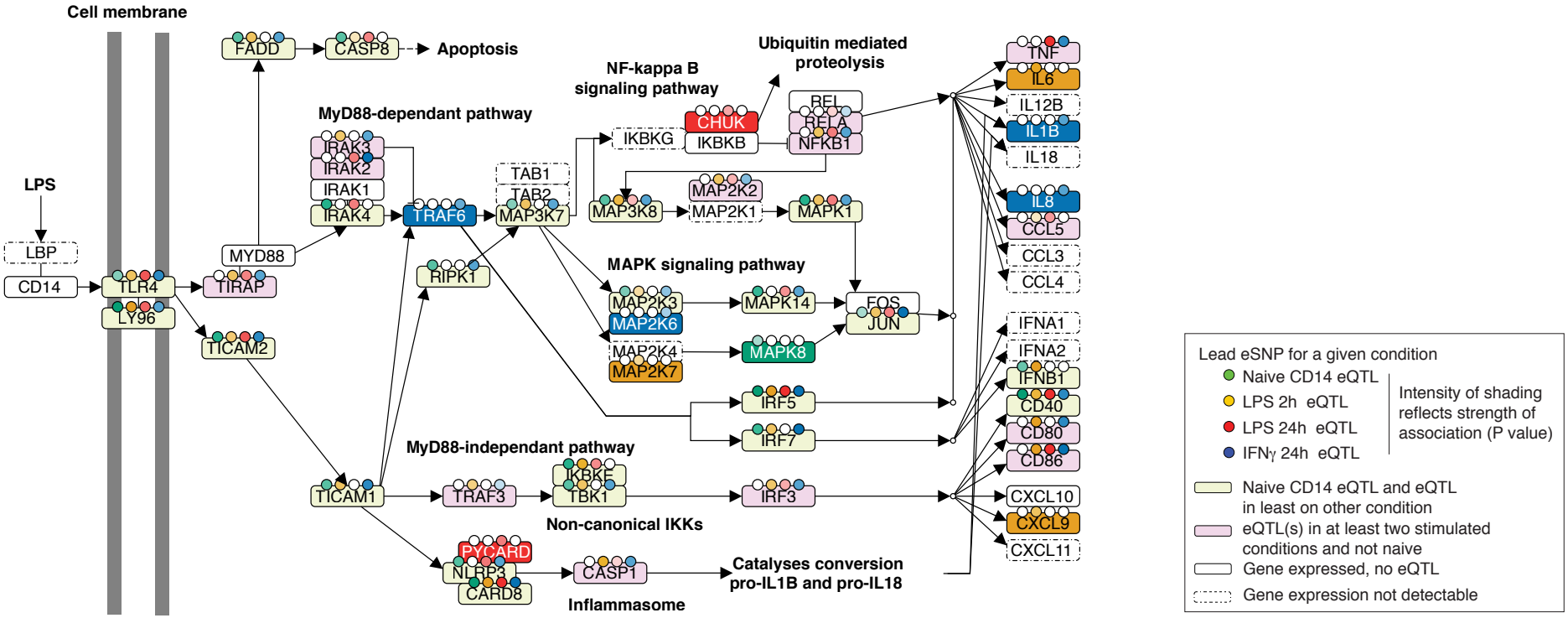


Fig. S7

Pathway analysis. (B) eQTL involving TLR4 signaling pathway. Genes defined and pathway drawn from KEGG pathway hsa04620. Context specific eQTL maximal after 2h LPS stimulation include *TIRAP*, encoding the toll-interleukin 1 receptor (TIR) domain containing adaptor protein (rs3903109 P=6.56x10^{-6}) and *CASP1*, encoding caspase 1, a key mediator of the inflammasome (rs11226623 P=2.5x10^{-10}) while *PYCARD*, encoding the PYD and capsase recruitment domain containing adaptor protein also found in the inflammasome, showed an eQTL only after 24h LPS (rs3815801 P=4.2x10^{-7}). In terms of mediating apoptotic signals, *FADD* (encoding the Fas-associated via death domain protein), was noted to have an eQTL at 2h LPS (rs1131715 P=3.15x10^{-7}). Genes encoding key downstream molecules were also found to have eQTL relating to LPS treatment, *IL6* only after 2h LPS (rs1476483 P=3.3x10^{-13}), *IFNB1* and *TNF*, which modulated the level of suppression of *TNF* gene expression at 24h (rs2516390 P=5.8x10^{-15}). LPS or IFNγ specific eQTL included *IRAK2* (maximal after IFNγ rs11465897 P=5.8x10^{-18}) and *IRAK3* (maximal LPS 2h rs10748036 P=5.5x10^{-5}); others were IFNγ specific including *TRAF6* (rs2458928 P=9.7x10^{-5}) and *IL8* (rs34945352 P=3.6x10^{-6}); or eQTL were

present in naïve cells but most significant after IFNγ including *TLR4* (rs10983736 P=1.2x10^{-10}) and *IRF7* (rs12805435 P=1.5x10^{-17}). Several treatment specific eQTL involved GWAS SNPs. Examples included *CD80*, a membrane receptor activated by binding of CD28 that leads to T cell proliferation and cytokine production (*73*) (rs12487334, in complete LD with the vitiligo associated SNP rs59374417 (*74*) which is strongly associated with differential expression in monocytes stimulated with LPS for 2h (P=8x10^{-18})); and *CD86* involving a multiple sclerosis GWAS.

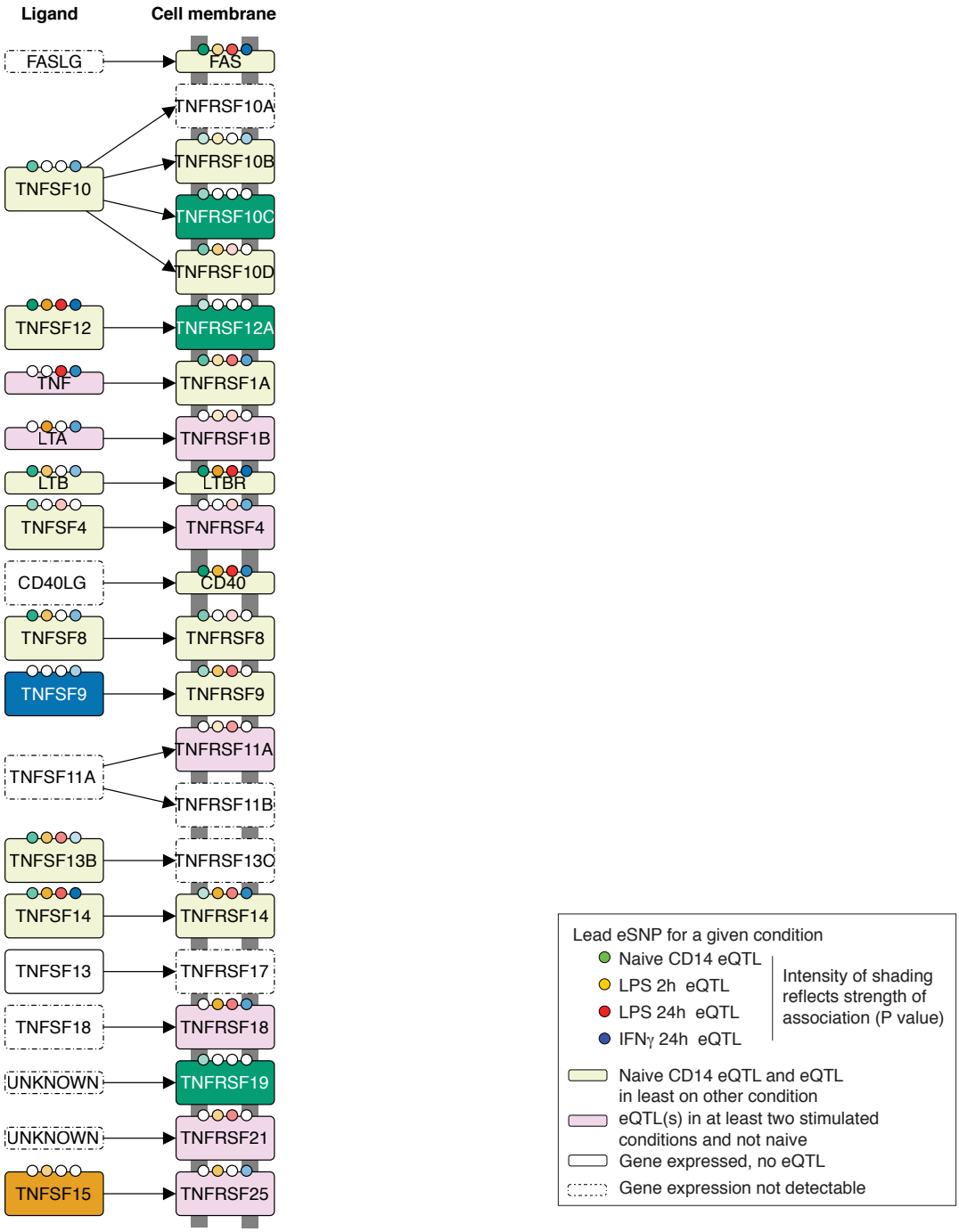


Fig. S7

Pathway analysis. (C) eQTL involving TNF ligand receptor family. Among the TNF family of genes, *LTA* showed a very strong stimulus specific eQTL after 2h LPS (P= 2.8×10^{-41}) for which the peak eSNP, rs1041981 (pThr60Asn) tags a haplotype previously shown to demonstrate allele-specific Pol II loading and expression in LCLs following mitogen stimulation (*75*), consistent with other evidence that this is a high LTA producer allele (*76, 77*). By contrast, no eQTL was apparent for *TNF* on induction at 2h but was seen on down regulation at 24h (rs2516390 P= 5.8×10^{-15}). The ligand and receptor LIGHT/LIGHTR encoded by *TNFSF14* and *TNFRSF14* respectively, have stimulus specific eQTL maximally significant after IFNγ (*TNFSF14* rs1077667 P=6.3×10^{-16}) and

LPS 2h (*TNFRSF14* rs1886730 P=5.5x10^{-11}). The lead *TNFSF14* eSNP, rs1077667, is a reported GWAS SNP for multiple sclerosis (*71*) while the stimulus specific eQTL for *TNFRSF14* is linked with a GWAS for ulcerative colitis (*46*). The TNFR1 gene, *TNFRSF1A* shows evidence of a constitutive eQTL and a stimulus specific eQTL after 24h LPS (rs4149576 P=3.6x10^{-8}); this SNP is also in strong LD (r^2 0.87) with a multiple sclerosis GWAS SNP. Other notable eQTL involving this gene family include the constitutive and the IFNγ specific eQTL for *FAS* (IFNγ rs1468063 P=4.3x10^{-33}); and the constitutive eQTL for *CD40*, *LTBR* and *TNFSF12*.

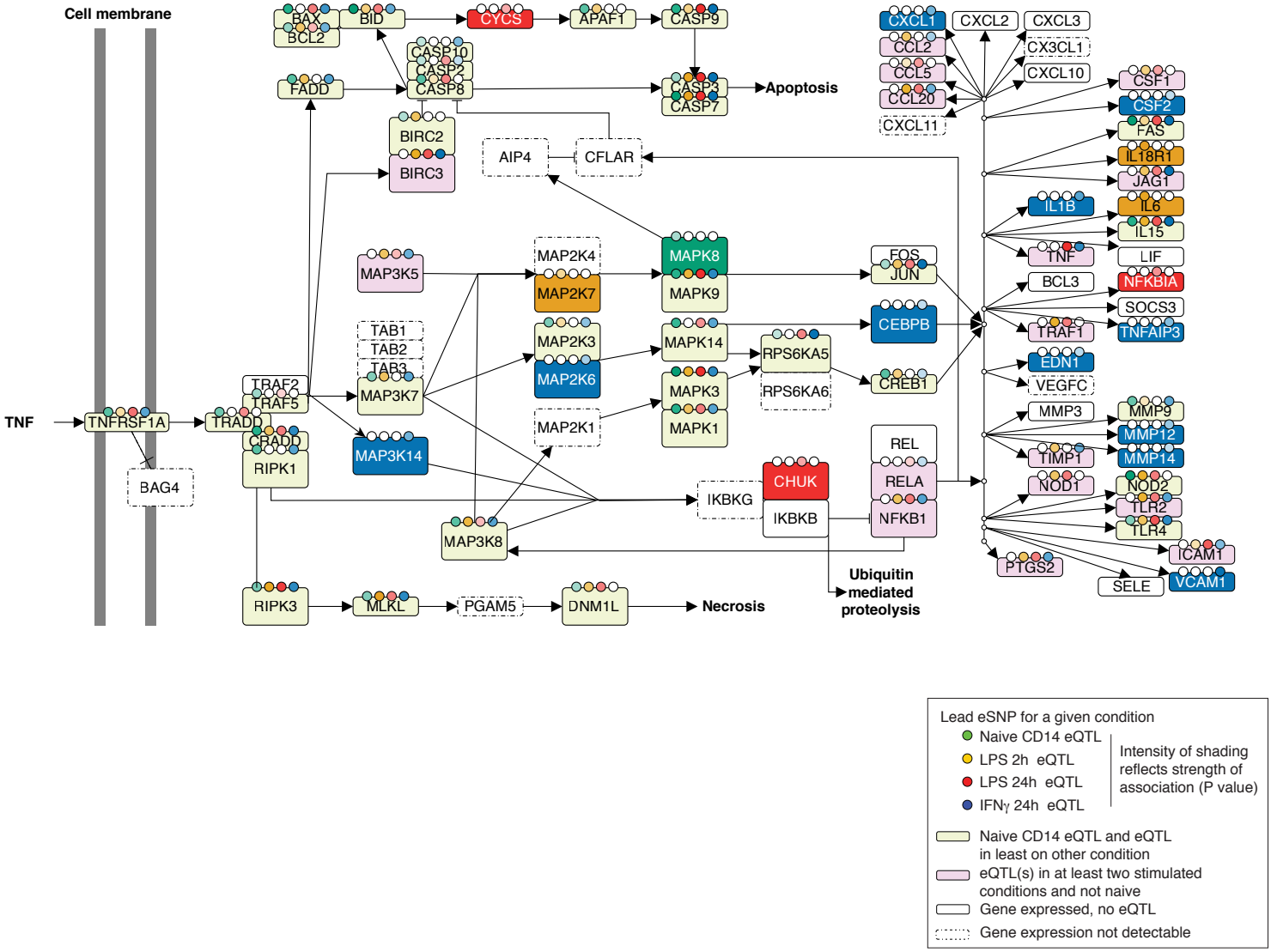


Fig. S7

Pathway analysis. ((D) eQTL involving TNF signaling pathway. Genes defined and pathway drawn from Kegg pathway hsa04668. The TNF signaling pathway also was found to contain context specific eQTL. These include a peak eSNP seen after LPS 2h for *TRADD* (the TNFR1 associated death domain protein) ($P=8.8x10^{-5}$) that is in complete LD with a reported GWAS SNP for coronary heart disease (*78*); and for the inhibitors of apoptosis genes *BIRC2* (LPS 2h rs10895290 $P=2.3x10^{-7}$) and *BIRC3* (IFNγ rs2155587 $P=4.8x10^{-15}$). A number of caspase genes including *CASP3* show very strong eQTL after IFNγ treatment (rs1314999 $P=6.2x10^{-51}$), as does *JUN* (encoding the transcription factor c-Jun) (rs4578243 $P=3.3x10^{-28}$). A strong constitutive eQTL is noted for *NOD2* (rs1981760 $P=2.5x10^{-23}$) for which the lead SNP is also a leprosy GWAS hit (*79*).

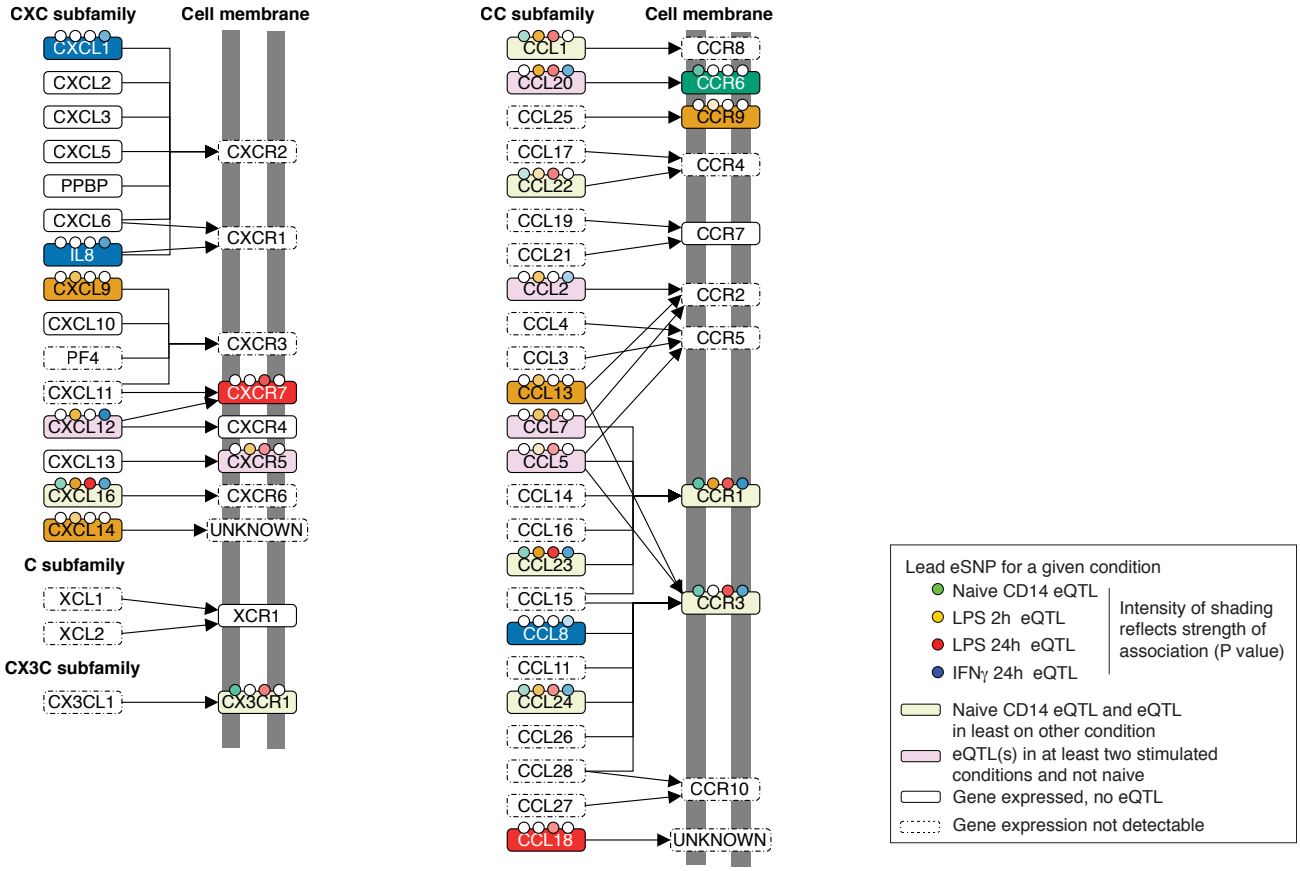


Fig. S7

Pathway analysis. (E) eQTL involving chemokine ligands and receptors. Genes defined and pathway drawn from Kegg pathway hsa04060. For chemokine ligands and receptors, only a small proportion of identified eQTL were present in naïve monocytes. The most striking association is seen upon LPS stimulation for 2h. These include examples such as *CXCR5* where the lead eSNP (rs4938573) showing association only in LPS2 ($P=9.6x10^{-5}$) is in strong LD (r^2 0.95) with a GWAS SNP for celiac disease (*80*) while for the CC-type chemokine receptor *CCR3*, rs7616215 is associated with gene expression after 24h LPS ($P=1.7x10^{-11}$) and is also the reported GWAS SNP for Behcet's disease (*42*).

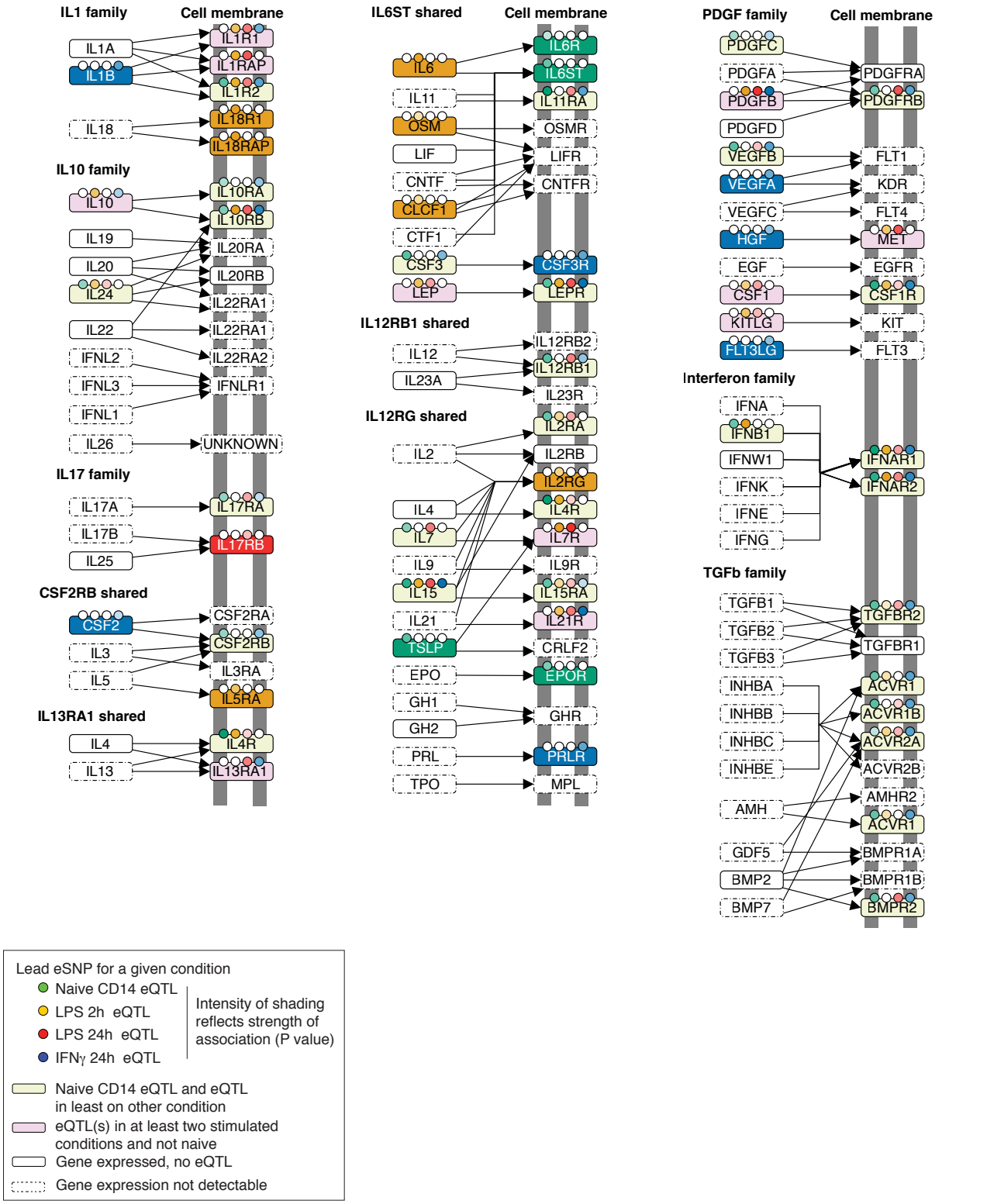


Fig. S7

Pathway analysis. (F) eQTL involving interleukins, PDGF, interferon and PDGF family ligands and receptors. Genes defined and pathway drawn from Kegg pathway hsa04060. The IL18 receptor and accessory subunit, encoded by the neighboring genes

IL18R1 and *IL18RAP*, show strong association with rs2058659 ($P= 2.8x10^{-34}$, $P=7.7x10^{-40}$ respectively) only after 2h LPS stimulation. For *IL10*, rs3024505 shows a strong LPS 2h specific eQTL ($P= 3.0x10^{-5}$); this SNP is also a GWAS hit for ulcerative colitis (*81*) and Crohn's disease (*43*) and is located in a DNase I hypersensitive site associated with multiple transcription factors, disrupts an EGR1 binding motif and has previously been associated with increased IL10 expression in Crohn's disease patients (*82*). For *IL7R*, an LPS specific eQTL maximal at 24h (rs931555 $P= 5.9x10^{-18}$) is a GWAS hit for multiple sclerosis (*83*).

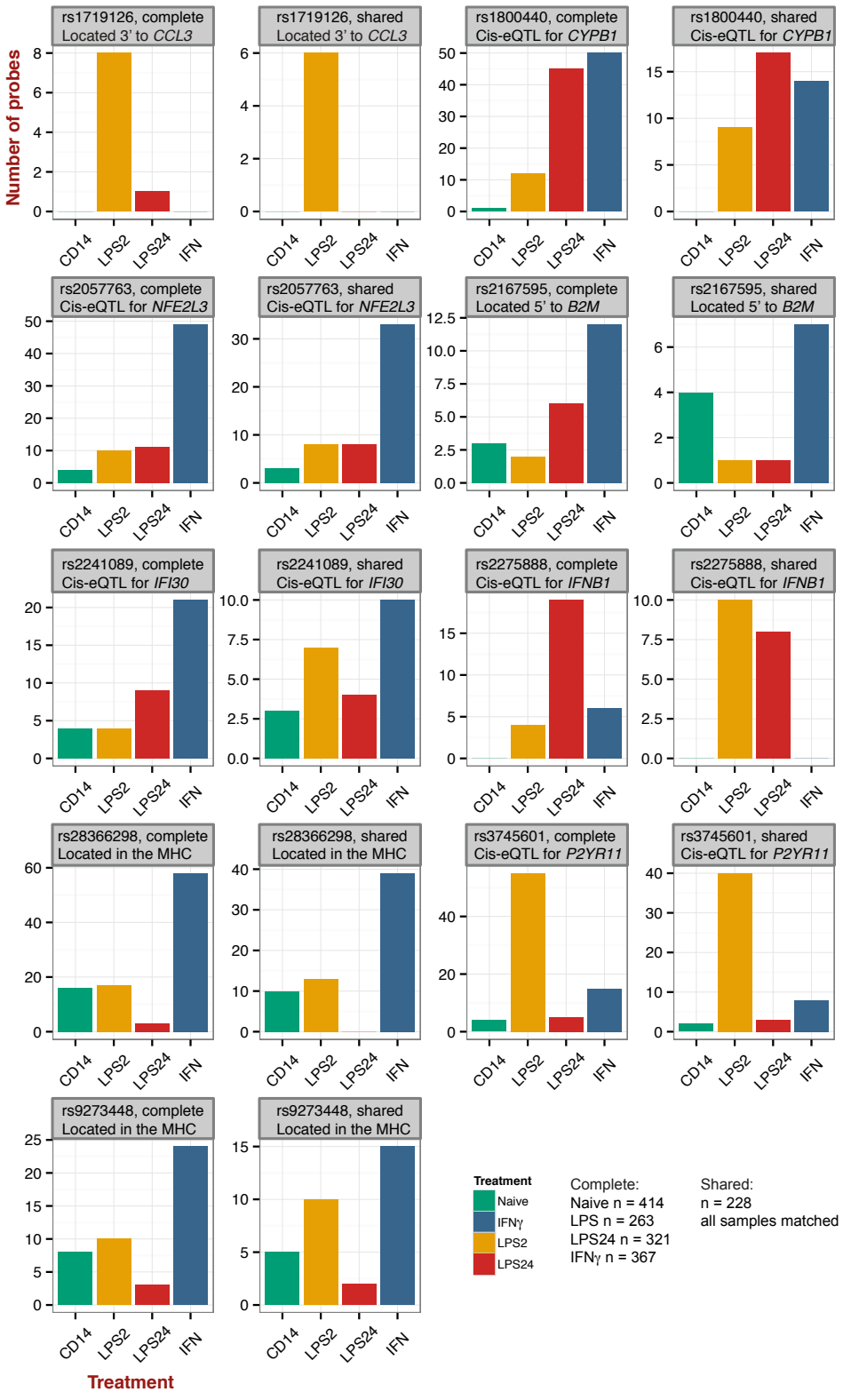


Fig. S8

Numbers of associations to observed multi-loci trans eQTL. eQTL analyses was performed on a single SNP basis to identify further trans associations in selected trans eSNPs with 4 or more separate associations in the full dataset. Bar charts illustrating

number of trans associations per SNP across the shared dataset of 228 samples and additionally in complete sample sets. rs1719126 resides within the chemokine cluster and is a context-specific trans eSNP after 2h LPS; rs1800440 is a missense variant in *CYP1B1*; rs2057763 is a cis-eQTL for *NFE2L3*, and trans-associated SNPs are highly significantly involved in proteasome function; rs2167595 is on chromosome 15 and is 5' to *B2M*; rs2241089 is immediately 3' to the ATG codon for *IFI30* and is a highly significant cis-eQTL to this gene across contexts; rs2275888 is a context specific cis-eQTL to *IFNB1* after 2h LPS, rs28366298 and rs9273448 are independent peak trans-eQTL within the MHC; rs3745601 is a missense variant within the purinoceptor encoding gene *P2YR11*.

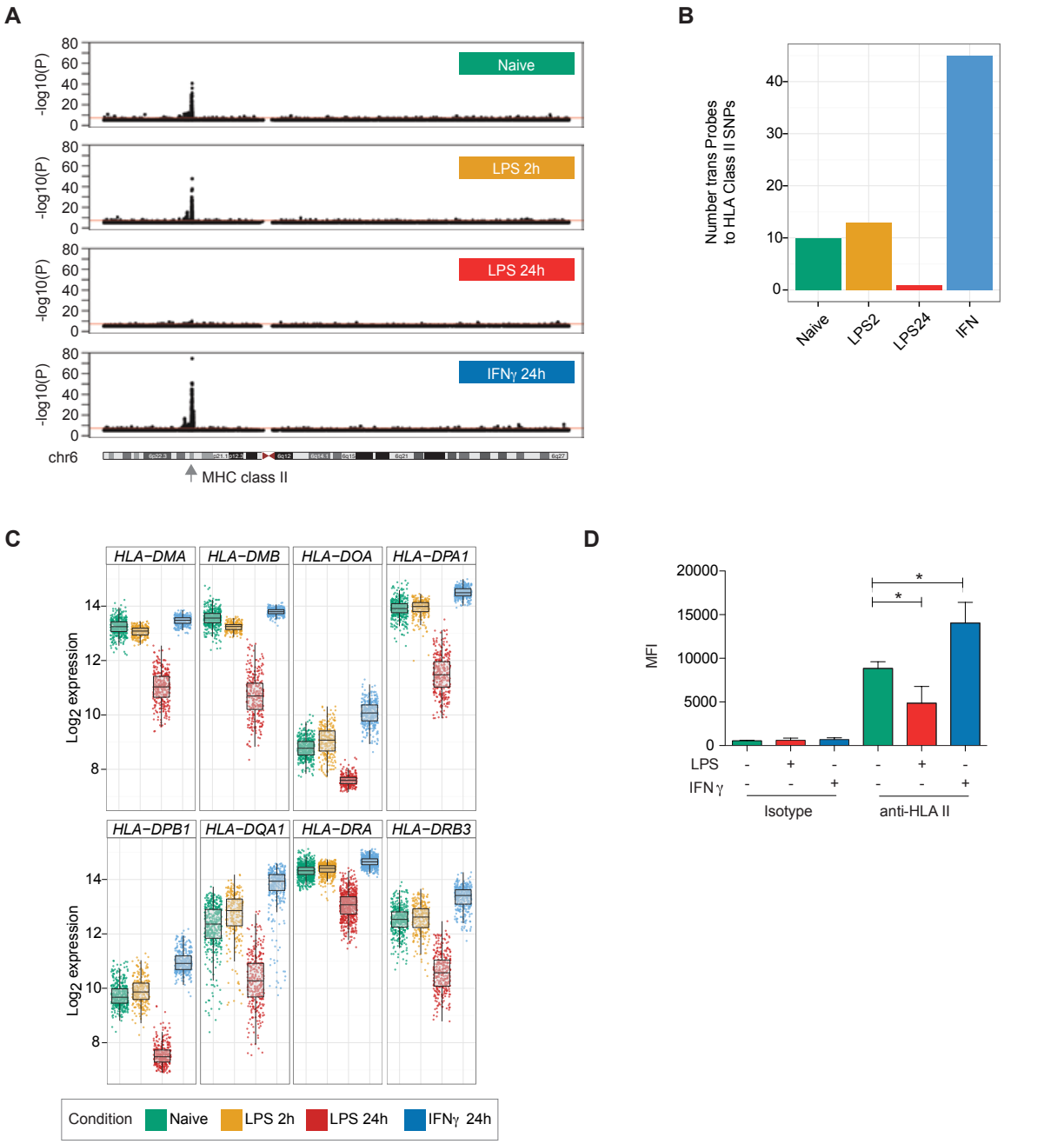


Fig. S9

Trans-eQTL involving SNPs located in the MHC. The MHC is a significant trans-eQTL locus. (**A**) Maximum strength of trans association for each SNP on chromosome 6 is shown by plotting the -log10 P-value for a given SNP association vs the genomic coordinate of that SNP. (**B**) Bar chart of the number of probes showing trans association with MHC class II SNPs by treatment for the shared dataset. Treatment with IFNγ is associated with a highly significant increase in number of *trans* loci over naïve state. In contrast, chronic exposure to LPS resolves all trans eQTL to the MHC. (**C**) Expression of specific MHC class II genes at the transcript level plotted by condition showing down regulation with LPS 24h and upregulation with IFNγ. (**D**) MHC class II expression (HLA-DR, -DP and -DQ) analysed at protein level by flow cytometry.

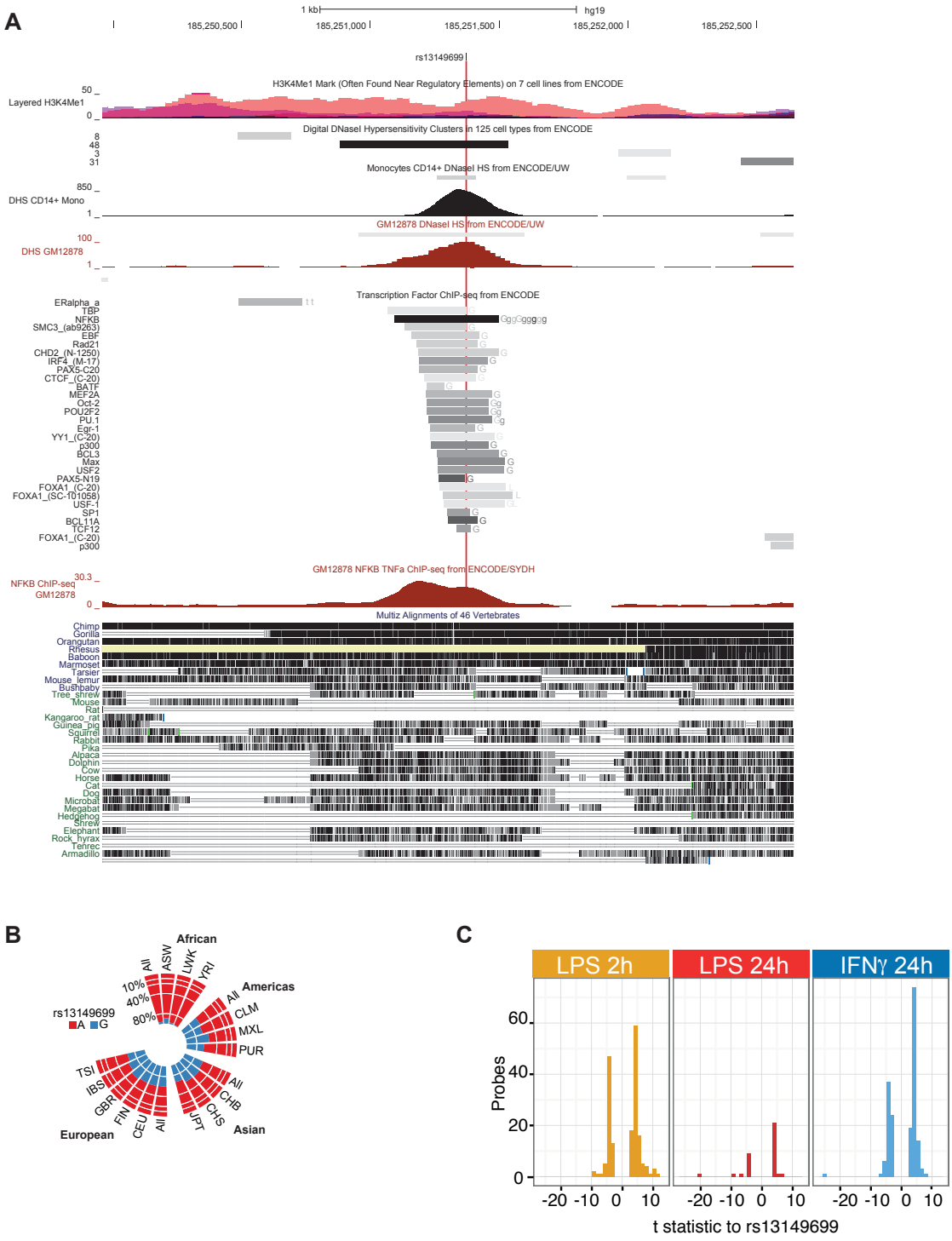


Fig. S10

Characterisation of IRF2 trans locus.
Figure continued overleaf

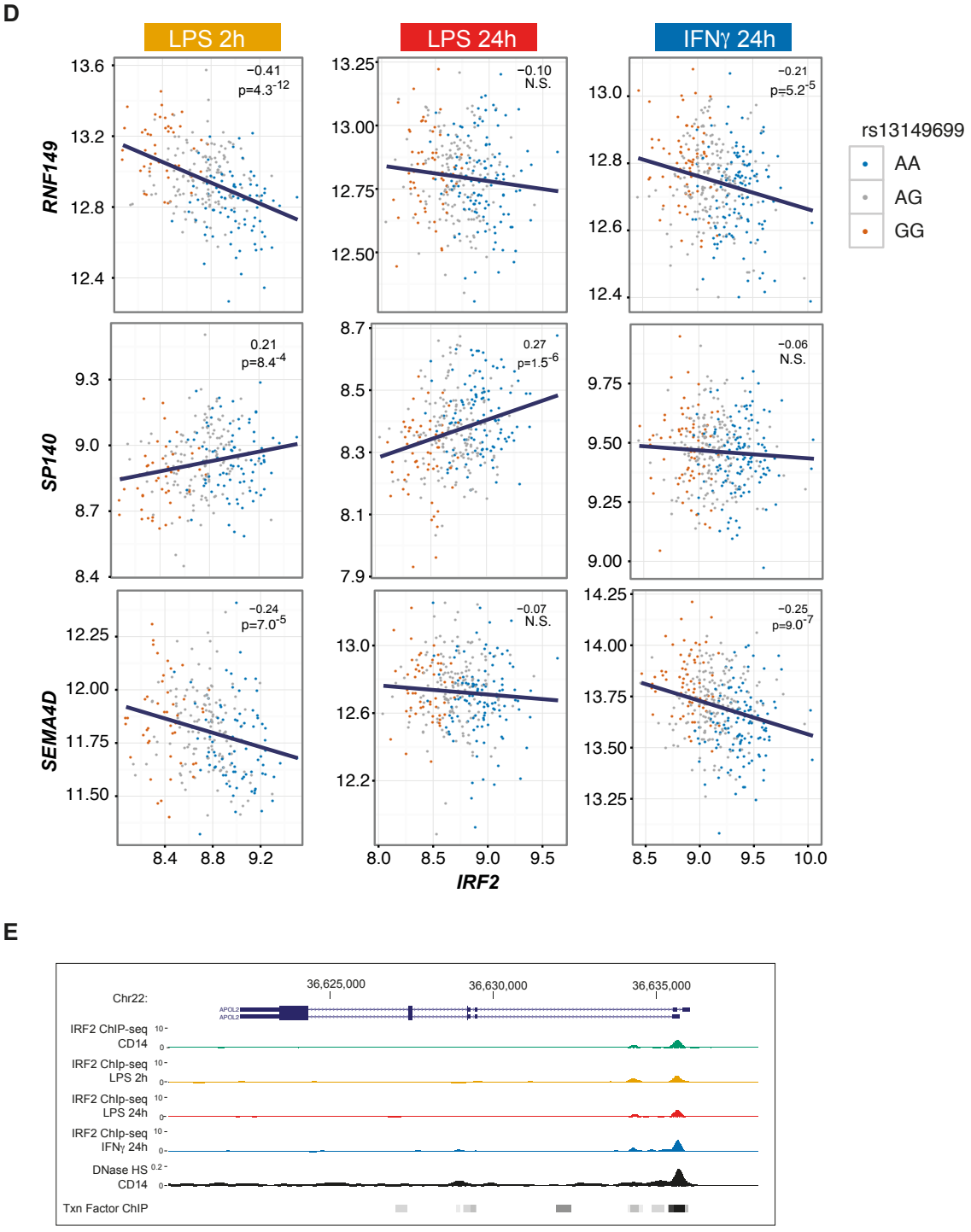


Fig. S10

Characterization of IRF2 trans locus. (A) Functional genomic landscape of rs13149699 demonstrates an intergenic region with regulatory features based on analysis of data from the ENCODE project. (**B**) Allele frequency of rs13149699 across populations. Note polymorphism is very rare in African populations (YRI MAF <1%) **(C)** Histograms of the t-statistics of probes associating in trans with rs13149699 across conditions (none observed in naïve state). Whilst the majority of probes are associated

with a positive t-statistic, in opposition to that of the effect of rs13149699 on IRF2, and suggestive of inhibition by IRF2; a number of probes are upregulated in conjunction with IRF2 expression. **(D)** *IRF2* correlation with trans associated genes is highly context specific – increasing *IRF2* expression is strongly negatively correlated with *RNF149* and *SEMAD* expression after 2h LPS or IFNγ, but no effect is observed after 24h LPS; whereas *IRF2* expression is positively correlated with *SP140* expression after 2 or 24h LPS, but not IFNγ. **(E)** *APOL2* – a further example of a trans associated gene to rs13149699 with an IRF2 ChIP-seq peak across treatments.

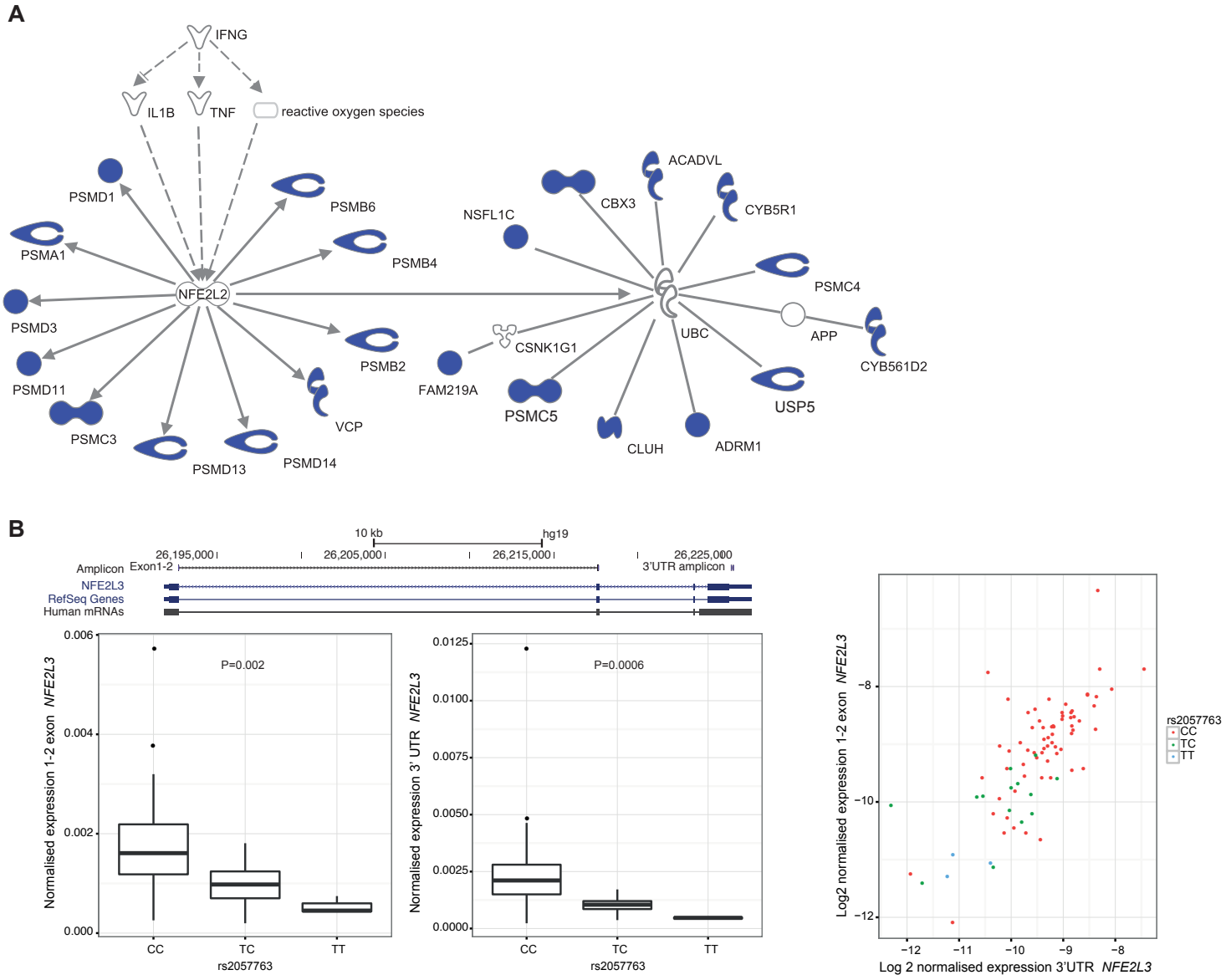


Fig. S11

A proteasomal trans network associated with differential expression of *NFE2L3*. (**A**) IPA defined a trans gene network involving proteasomal genes for rs2057763. (**B**) Quantitative real time PCR was performed on 84 individuals for *NFE2L3* using two different primer amplicons, one set spanning the first and second exon of the gene and the other targeting the 3' UTR. Box plots are shown (median +/- IQR) by allele with P values calculated using one-way ANOVA. Primer sequences used:

NFE2L3_3UTR_fwd	TCACACTTGTGGGCAATCTG
NFE2L3_3UTR_rvs	TAAAAAGGATGGCTGCCAAG
NFE2L3_ex1-2_fwd	AGCGAGGAGAATGGGGTACT
NFE2L3_ex2_rvs	GAGTTCTCCTTCTGGGCTGA

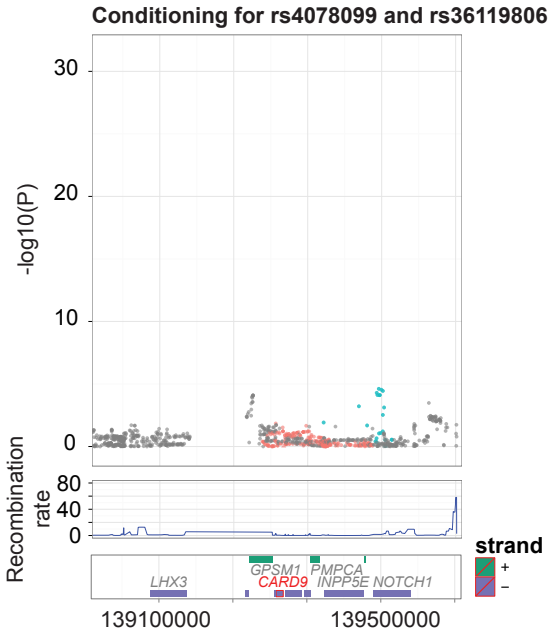


Fig. S12

Conditional analysis for *CARD9*. Association with *CARD9* expression after conditioning on rs4078099 and rs36119806.

Additional Data Table S1 (separate file)

Differentially expressed genes following treatment. (A) Analysis of differential gene expression for shared datasets (n=228). **(B)** Analysis of differential gene expression between Naïve monocytes (n=228) and those that had been incubated after correction for incubation had been applied (n=59).

Additional Data Table S2 (separate file)

List of cis-eQTL observed in naïve and induced monocytes. (A) Peak Cis-eSNP among shared datasets. **(B)** Complete dataset of cis-eQTL by individual treatment. **(C)** Directional cis-eQTL associating with one single genomic peak ($P < 5x10^{-8}$). **(D)** Summary of cis eSNP overlap.

Additional Data Table S3 (separate file)

Trans-eQTL (A) List of trans-eQTL observed in naïve and induced monocytes for complete datasets. **(B)** Summary of trans eSNP overlap.

Additional Data Table S4 (separate file)

Interrogation of trans networks. Single SNP analysis for **(A)** *IFNB1* driven trans hub (rs2275888) **(B)** *CYP1B1*(rs1800440) **(C)** *NFE2L2* (rs2057763) **(D)** *IRF2* (rs13149699).

Additional Data Table S5 (separate file)

IRF2 ChIP-seq binding intervals. (A) List of ChIP-seq binding peaks identified in Naïve monocytes (post 24 incubation). Peaks are listed if identified in 2 separate biological samples and they overlap by a minimum of 10%. Closest gene to the peak is additionally identified. **(B)** As per **(A)** but after exposure to 2h LPS (total incubation =24h). **(C)** As per **(A)** but post exposure to 24h IFNγ.

Additional Data Table S6 (separate file)

GWAS-eSNP associations. Overlap between primary peak eQTL defined for each condition and GWAS loci (www.genome.gov/gwastudies/). This was defined as present if the GWAS SNP itself or any SNPs with $r^2 > 0.8$ with this SNP were part of one of the eQTL peaks ($P < 5 x10^{-8}$).

Additional Data Table S7 (separate file)

Allele-specific correlation analysis for GSDMA (rs3859192).

References

67. S. Thomas, D. Bonchev, A survey of current software for network analysis in molecular biology. *Hum. Genom.* **4**, 353-360 (2010).
68. S. L. Foster, D. C. Hargreaves, R. Medzhitov, Gene-specific control of inflammation by TLR-induced chromatin modifications. *Nature* **447**, 972-978 (2007).
69. S. Raza *et al.*, A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst. Biol.* **2**, 36 (2008).
70. K. Schroder, P. J. Hertzog, T. Ravasi, D. A. Hume, Interferon-gamma: an overview of signals, mechanisms and functions. *J Leukoc. Biol.* **75**, 163-189 (2004).
71. S. Sawcer *et al.*, Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-219 (2011).
72. V. Enciso-Mora *et al.*, A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat. Genet.* **42**, 1126-1130 (2010).
73. R. J. Peach *et al.*, Both extracellular immunoglobin-like domains of CD80 contain residues critical for binding T cell surface receptors CTLA-4 and CD28. *J Biol. Chem.* **270**, 21181-21187 (1995).
74. Y. Jin *et al.*, Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44**, 676-680 (2012).
75. J. C. Knight, B. J. Keating, K. A. Rockett, D. P. Kwiatkowski, In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat. Genet.* **33**, 469-475 (2003).
76. K. Ozaki *et al.*, Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650-654 (2002).
77. G. Messer *et al.*, Polymorphic structure of the tumor necrosis factor (TNF) locus: an NcoI polymorphism in the first intron of the human TNF-beta gene correlates with a variant amino acid in position 26 and a reduced level of TNF- beta production. *J Exp. Med.* **173**, 209-219 (1991).
78. G. Lettre *et al.*, Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet.* **7**, e1001300 (2011).
79. F. R. Zhang *et al.*, Genomewide association study of leprosy. *N Engl. J Med.* **361**, 2609-2618 (2009).
80. A. Zhernakova *et al.*, Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7**, e1002004 (2011).
81. C. A. Anderson *et al.*, Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246-252 (2011).
82. A. H. Wang *et al.*, The effect of IL-10 genetic variation and interleukin 10 serum levels on Crohn's disease susceptibility in a New Zealand population. *Hum. Immunol.* **72**, 431-435 (2011).
83. J. H. Wang *et al.*, Modeling the cumulative genetic risk for multiple sclerosis from genome-wide association data. *Genome Med* **3**, 3 (2011).