

Published in final edited form as:

Nat Commun. ; 6: 7545. doi:10.1038/ncomms8545.

Genomic modulators of gene expression in human neutrophils

Vivek Naranbhai^{1,*}, Benjamin P. Fairfax^{1,*}, Seiko Makino¹, Peter Humburg¹, Daniel Wong¹, Esther Ng¹, Adrian V.S. Hill¹, and Julian C. Knight^{1,*}

¹Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Abstract

Neutrophils form the most abundant leukocyte subset and are central to many disease processes. Technical challenges in transcriptomic profiling have prohibited genetical genomic approaches to date. Here, we map expression quantitative trait loci (eQTL) in peripheral blood CD16+ neutrophils from 101 healthy European adults. We identify *cis*-eQTL for 3281 neutrophil-expressed genes including many implicated in neutrophil function, with 450 of these not previously observed in myeloid or lymphoid cells. Paired comparison with monocyte eQTL demonstrates nuanced conditioning of genetic regulation of gene expression by cellular context which relates to cell type specific DNA methylation and histone modifications. Neutrophil eQTL are markedly enriched for trait-associated variants particularly autoimmune, allergy and infectious disease. We further demonstrate how eQTL in *PADI4* and *NOD2* delineate risk variant function in rheumatoid arthritis, leprosy and Crohn's disease (CD). Taken together, these data help advance understanding of the genetics of gene expression, neutrophil biology and immune-related diseases.

Variation in the human genome is a major regulatory mechanism of gene transcription¹. Regulatory variants may modulate gene expression of local genes (*cis* eQTL, likely acting on the same chromosome) or genes at a distance on non-contiguous chromosomes (*trans* eQTL). Quantitative variation in transcription frequently leads to protein variation² leading to phenotypic traits. Although cell-type specific effects were noted in early studies³, ease of sample availability may explain why the largest studies of eQTL are in cell lines⁴⁻⁸ or mixed populations of leucocytes from peripheral blood⁹⁻¹². Studies of eQTL in monocytes¹³⁻¹⁵, monocyte-derived dendritic cells¹⁶, B-lymphocytes¹³, regulatory T-cells¹⁷, CD4+ T-cells or

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding: Vivek Naranbhai, Benjamin P. Fairfax and, Julian C. Knight Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN United Kingdom
vivekn@well.ox.ac.uk bfairfax@well.ox.ac.uk julian.knight@ndm.ox.ac.uk Phone: +44 1865 287 634 .

Author contributions

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

Competing Financial Interests

The authors have no competing financial interests to declare.

Accession codes

Gene expression data is available through ArrayExpress (E-MTAB-2232 & E-MTAB-3536). Genotyping data has been deposited at the European Genome- Phenome Archive (EGA) and is available on request to EGA (EGAS0000000109).

bulk T-cells³ demonstrate that the effects of a genetic variant on gene expression differ according to cell type and these are further conditioned by cellular activation state and stimulatory environment¹⁸⁻²⁰. Similar results comparing tissues support this conclusion²¹⁻²⁴. Collectively, these studies demonstrate the need for context-specific eQTL mapping in diverse primary populations of cells informative for disease.

Neutrophils make up 40-70% of the total circulating leucocyte pool and due to their abundance in blood and tissue, they are frequently observed in tissue specimens with minimal blood contamination. There has been one recent study of eQTL in murine neutrophils²⁵, but the genetic architecture of gene expression in human neutrophils remains unclear. Challenges in isolating neutrophils²⁶ (as opposed bulk granulocytes that paradoxically bias RNA measures towards eosinophils²⁷) have limited the ability to study these cells leading to attempts at *in silico* prediction of neutrophil eQTL²⁸ from whole blood data. About 10¹¹ neutrophils are produced daily in the adult human bone marrow and constitute first-responders in the innate immune response to a variety of infectious and non-infectious insults²⁹. Neutrophils are characterised by multilobed nuclei and an abundance of primary, secondary and tertiary soluble-defence-mediator-filled granules. Although neutrophils are classically thought to be short-lived cells, they play roles in acute and chronic inflammation and can migrate into tissues or return to the blood compartment and survive³⁰. Neutrophils have extensive crosstalk with each of the major blood cell subsets (megakaryocyte, myeloid and lymphoid)³¹⁻³⁴ further magnifying their function in health and disease. Neutrophils are thus central to orchestrating immune responses and understanding the regulation of gene expression in neutrophils may, we hypothesised, offer insights into disease biology.

We identified *cis*-acting genetic modulators of gene expression in neutrophils for more than 3000 genes, including dozens involved in the development, migration and function of neutrophils. Comparison with other cell types and paired analysis of neutrophil and monocyte eQTL demonstrates how cell-type modifies the effect of eQTL. Moreover, cell-type specific epigenetic data helps resolve the mechanisms of cell-type constraint of eQTL and fine-map causal variants. We show how variants that affect gene expression are implicated in hundreds of common diseases. We interrogate two such associations. First we show how integration of neutrophil and monocyte eQTL and epigenetic data with GWAS data implicates PADI4 expression in neutrophils in rheumatoid arthritis susceptibility. Second we show how a SNP with pleiotropic association with leprosy and CD susceptibility alters neutrophil inflammatory responses to NOD2 ligands through altered STAT3 binding and consequent NOD2 expression. Finally, we observe that many neutrophil eQTL reside within regions that have been subject to selection. Collectively, our data advance the understanding of the genetics of gene expression for a pathophysiologically important cell type and role in human disease.

Results

Defining eQTL in primary human neutrophils

To identify regulatory variants for gene expression in primary human neutrophils, we enrolled 101 healthy adult European volunteers in Oxford, United Kingdom (Figure 1a).

Individuals were genotyped on the Human OmniExpress 12v1.0 chip, and we used genome-wide imputation with stringent quality checks to infer additional single-nucleotide polymorphism (SNP) or simple insertion/deletion genotypes with high confidence. CD16+ neutrophils, isolated by a two-step sequential gradient-density and immunomagnetic sorting procedure yielding a highly purified population, were subjected to whole-transcriptome characterization by hybridization of cRNA to Human HT-12 v4 Expression BeadChips (Illumina). We assessed only genes for which specific probes existed, and which were confidently detected in more than 5% of the cohort (GenomeStudio detection $p < 0.01$). We used a widely-adopted linear-additive modeling approach (MatrixeQTL)³⁵ adjusting for principal components which has been shown to enhance eQTL discovery and control for technical and demographic heterogeneity^{10,19,28,36}.

After genome-wide imputation 3281 genes (roughly (~) 30% of 9147 genes tested and equivalent to 3675 probes) had one or more identifiable loci within 1Mb (defined as *cis*-eQTL) of a probe that was significantly associated with gene expression at a false discovery rate (FDR) threshold of 5% (Figure 1b and Supplementary Data 1). The median proportion of variance explained by the most significant variant for each probe, was 16.6% (IQR 12.9%-26.0%) but notably was greater than 50% for several genes, such as *C4BPA*, involved in venous thrombosis³⁷, the anti-inflammatory monosodium urate receptor *CLECI2A*³⁸ and the antibacterial enzyme lysozyme (*LYZ*) (Additional genes with large effect sizes are highlighted in Figure 1b). Genes with an eQTL were more highly expressed (median normalized expression 8.14 vs. 7.7, $p = 1.7 \times 10^{-72}$) and had greater variance (median variance 0.049 vs 0.029, $p = 1.07 \times 10^{-122}$) than genes without an eQTL. The most significant eQTL per gene (denoted as the peak eQTL) clustered around the transcriptional start site (TSS), and the effect sizes increased with proximity to the TSS (Supplementary Figure 1a) but, annotation and inspection at higher-resolution demonstrates distribution of eQTL across gene structures with clustering around TSS and transcription end sites, consistent with previous studies³⁹ (Supplementary Figure 1b).

We did not identify association between any of 168 copy-number variants, inferred from raw genotype calls, and gene expression after correcting for multiple comparisons (Supplementary Data 2).

Comprehensive identification of loci associated with gene expression on non-contiguous chromosomes (*trans* eQTL) typically requires large sample¹² sizes due to typically reduced effect sizes of variants acting *trans* allied to an increased burden of multiple hypothesis testing. Accordingly, we identified a smaller set of 33 genes regulated in *trans*, at a false discovery rate threshold of <1% (Figure 1b and Supplementary Data 3). These *trans* eQTL include rs10012416, which we find is a *cis* eQTL for *CRIPAK* (encoding an inhibitor of p21 activated kinase Pak1 important in neutrophil cytoskeletal dynamics involved in phagocytosis⁴⁰) and has *trans* effects on *AVP*, *SPTBN3* and *IRF6*; and rs10784774, a *cis* eQTL for *LYZ* that is associated with *ZNF131* in neutrophils as we previously had reported in monocytes¹³. We note that this study is not powered to identify *trans* eQTL with weak to moderate sized effects.

eQTL in genes central to neutrophil biology

We proceeded to examine genes relevant to neutrophil function for eQTL as these may serve as a narrative resource to understanding the role of regulatory variants in neutrophil biology. We assembled a list of 164 genes described, with a verifiable reference, to be involved in a particular aspect of neutrophil function as being involved in aspects of neutrophil function in recent reviews^{29,30,41-43}. Because this approach is at risk of narrative bias, we caution against interpretation of these results as evidence of enrichment for genes important in neutrophil function. Of the 113 genes for which an expression probe existed, passed QC and was included in analysis, 104 had an identifiable variant associated with expression ($p < 0.05$) and 47 after false-discovery adjustment ($FDR < 0.05$). We observe eQTL for genes involved in most aspects of neutrophil development and function (Figure 2, and Supplementary Data 4). Interestingly, several genes with an eQTL are implicated in Mendelian disorders involving neutrophils. For example rs933222 is associated with expression of *RAC2*, a gene encoding a Rho GTPase that is part of the NADPH oxidase complex involved in initiation of phagocytosis (Figure 11) and is involved in neutrophil immunodeficiency syndrome⁴⁴ (OMIM #608203).

Ingenuity pathway analysis of 975 genes with an eQTL in neutrophils but not in monocytes (Supplementary Figure 2, and detailed further below), revealed enrichment for functions relating to cell death ($p = 1.4 \times 10^{-7}$), apoptosis (8.2×10^{-7}), necrosis (6.4×10^{-6}); and infection, notably viral infection (4.2×10^{-6}) and infection of cells (2.2×10^{-5}). The most significant upstream transcriptional regulator was TP53, which plays a critical role in cell proliferation and apoptosis as well as antimicrobial function highlighting how the consequences of TP53 for the individual may be modulated by eQTL in downstream mediator and effector genes. A number of cytokines were also identified as upstream regulators including IFNB1 ($p = 5.9 \times 10^{-4}$), IL15 (3.0×10^{-3}), IFNG (4.1×10^{-3}), CD40LG (6.8×10^{-3}) and TNF (7.2×10^{-3}).

Patterns of eQTL in neutrophils and other immune cell types

Regulation of gene expression may be constrained to specific cell-types and contexts. To delineate aspects of shared and unique regulatory genomic architecture in neutrophils, we pursued three complementary approaches.

First, we note that 63% (2069/3281) of genes with a *cis* eQTL in neutrophils are reported to have a *cis* eQTL in the largest blood eQTL meta-analysis to date¹² (obtained through the *blodeqtl* browser) and recent *in silico* predictions²⁸ attribute 20% (443/2188) to neutrophils and 90% (1969/2188) as 'generic' (note that a gene may be denoted as neutrophil *and* generic). Therefore many eQTL in neutrophils may not have been identified in whole blood studies, and even when they are, *in silico* deconvolution of cell types may not establish if a gene has an eQTL in a given cell type. Conversely, 84% of genes (415/495) with an eQTL in blood that are bioinformatically ascribed to neutrophils (and tested in our study) have an eQTL in purified neutrophils, providing evidence of cross-study validation. The estimated effect sizes in neutrophils and whole blood for genes with an eQTL in both are only moderately correlated (spearman correlation estimate for 691 genes, $\rho = 0.54$, 95% CI 0.43-0.60, $p = 1.77 \times 10^{-23}$); as we shall present later, directionally opposing eQTL amongst

different cells are not uncommon and this may plausibly affect *in silico* effect size estimates. Secondly, for 40% (261/646) of genes with a *cis* eQTL in mouse CD4+ T-cells and/or neutrophils²⁵ whose human homologues we tested in humans, we observe a significant *cis* eQTL in neutrophils (a 5.72 fold enrichment over random expectation (646/9147), $p=1.02\times 10^{-80}$). Thirdly, we compared genes with an eQTL in neutrophils to those that have been reported previously in other primary immune cells: CD4 T-cells, regulatory T-cells¹⁷ and B-cells¹³ (grouped as lymphoid cells) or monocytes and monocyte derived dendritic cells^{14-16,19} (grouped as myeloid cells). Of the 3281 genes with an eQTL in neutrophils, 1671 (51%) had an eQTL in both myeloid and lymphoid cells, 823 (25%) in non-neutrophil myeloid cells, 337(10%) in lymphoid cells, and 450 have no reported eQTL in either myeloid or lymphoid cells (Supplementary Figure 3). Although our analysis neither assesses whether the same genetic variant regulates gene expression in all cell types nor whether the effect is the same, it demonstrates that, for at least 14% of the genes in which we observe an eQTL, we have identified novel regulatory variants and shown the utility of eQTL mapping in primary neutrophils.

To further examine the effect of cell-type on eQTL we performed a detailed analysis of monocytes and neutrophils isolated contemporaneously from 99 caucasian donors in which we confined analysis to 8362 genes expressed in both cell types. We identified 400,783 unique variant-gene associations across both cell types: 87,276 variants for 1031 genes in both neutrophils and monocytes, 118,817 variants for 2847 genes in neutrophils and 194,690 variants for 3674 genes in monocytes (Figure 3a). Higher expression of a gene in neutrophils compared with monocytes was a positive predictor of whether an eQTL was present in neutrophils or not, but normalised probe intensity explains just ~2% of variance and there are many genes in which an eQTL is observed in only one of neutrophils or monocytes despite similar levels of gene expression. The effect size of regulatory variants in both monocytes and neutrophils was larger for eQTL seen in both cell types compared with those seen only in one cell type ($p=7.4\times 10^{-196}$, Supplementary Figure 4) confirming previous observations¹³. Interestingly, whereas the number of unique eQTL is greater in monocytes, eQTL effect sizes were larger in neutrophils than in monocytes regardless of whether the eQTL was seen in both cell types ($p=1.9\times 10^{-19}$) or just one ($p=9\times 10^{-200}$). This likely reflects a proportionally greater impact of genetic variation in governing gene expression in neutrophils that recapitulates their shorter lifespan and reduced exposure to environmental modifiers.

We highlight three patterns of cell environment modifying genetic effects on gene expression in addition to apparent cell-type constraint of regulatory activity. First, although 1939 genes have at least one eQTL in both cell types (and 1031 share an eQTL), for 908 genes the peak regulatory variants are different and for 840 (93%) of these, independent ($r^2<0.2$) demonstrating that the same gene may have independent regulatory variants of varying strengths and directions in different cell types. An example for the *OSCAR* gene is shown in Figure 3b. Conversely a particular variant may show pleiotropy in the gene it regulates in neutrophils and monocytes. About 4.5% (12661/278200) of variants that regulate a gene in one cell-type are associated to a different gene in the other cell type. For example, rs35244261 is associated with levels of *ATM* in neutrophils and *NPAT* in

monocytes (Figure 3c). In gene-dense regions this could be due to cell type conditioning the effect of one or more variants that have regulatory effects on nearby genes. The third pattern we highlight is pleiotropy in direction of effect of a variant in different cell types on the same gene. We identified 2823 variants with significant but directionally opposing effects on 66 genes in both cell types (Figure 3d). Several are notable for the variant being associated with a disease directly or through a linked variant such as rs8066560 (*TOMIL2*, Parkinson's disease), chr18:3375159:D (Figure 3e, *ELP2*, oesophageal squamous cell carcinoma), rs74058715 (*PADI4*, rheumatoid arthritis), and rs1981760 (*NOD2*, leprosy and CD), the latter two described in detail below. Collectively, the comparison of monocyte and neutrophil eQTL supports a model of widespread interaction between cellular milieu and genetic factors in regulation of gene expression even amongst cells of similar (myeloid) lineage. Moreover, delineation of cell-type restriction of eQTL in neutrophils may provide a tool to understand their involvement in variant-phenotype association.

Epigenetic mechanisms of eQTL in neutrophils and monocytes

The local genomic environment of a regulatory variant is cell type dependent and this may modify variant effect. Moreover, as has been shown for complex traits, leveraging epigenetic mark information may help fine-map genotype-phenotype associations⁴⁵. Therefore, to elucidate the role DNA methylation and histone modification play in genetic effects on gene expression, we compared eQTL maps in neutrophils and monocytes to maps of DNA methylation and several histone modifications generated by whole-genome bisulfite sequencing or chromatin-immunoprecipitation and sequencing (ChIP-Seq) respectively in neutrophils and monocytes from 4-8 individuals by the BLUEPRINT consortium⁴⁶. As may be predicted by their shared ontology, monocytes and neutrophils have substantial overlap in regions of the genome that are methylated or subject to histone modifications, although this differs by the specific combination of epigenetic mark (Supplementary Data 5). Relative to all imputed variants within 1MB of a gene (4812340 variants near a neutrophil-expressed gene tested for being a *cis* eQTL), we observed marked and significant enrichment of peak *cis* eQTL in hypomethylated regions ($p_{\text{neut}}=2.14\times 10^{-107}$, $p_{\text{mono}}=2.52\times 10^{-124}$) and depletion in hypermethylated regions ($p_{\text{neut}}=6.36\times 10^{-46}$, $p_{\text{mono}}=7.55\times 10^{-23}$) in the concordant cell type i.e. neutrophil eQTL in neutrophil hypomethylated regions (Figure 4a,4c and Supplementary Data 6). Similarly, as shown in Figure 4b (and Supplementary Data 6), peak *cis* eQTL were enriched in regions with histone marks associated with promoter activity (H3K4me3; $p_{\text{neut}}=8.23\times 10^{-231}$, $p_{\text{mono}}=1.31\times 10^{-165}$), active or poised enhancers (H3K27Ac; $p_{\text{neut}}=4.85\times 10^{-230}$, $p_{\text{neut}}=1.34\times 10^{-80}$, H3K4me1; $p_{\text{neut}}=2.23\times 10^{-308}$, $p_{\text{mono}}=4.84\times 10^{-229}$) or activation (H3K36me3; $p_{\text{neut}}=2.01\times 10^{-238}$, $p_{\text{mono}}=3.22\times 10^{-196}$) and depleted in regions associated with repressive histone modifications (H3K27me3; $p_{\text{neut}}=3.78\times 10^{-3}$, $p_{\text{mono}}=2.59\times 10^{-2}$ and H3K9me3; $p_{\text{neut}}=3.35\times 10^{-4}$, $p_{\text{mono}}=2.57\times 10^{-2}$). These data are replicated by an orthogonal approach, in which, despite vastly fewer regions, we observe 14-fold enrichment of neutrophil eQTL in neutrophil enhancer regions ($p=3.03\times 10^{-15}$) and 6-fold enrichment of monocyte eQTL in monocyte enhancer regions ($p=3.58\times 10^{-17}$) using enhancer maps generated by CAGE-Seq in FANTOM^{547,48}. These data support a model of epigenetic environment modifying regulatory effects in different cell types. An example for the *PADI2* gene is shown in Figure 4d-e.

Mechanisms by which regulatory variants operate include alteration in transcription factor binding. Neutrophil and monocyte peak eQTL are 3.1- and 2.87-fold enriched (2.70×10^{-118} and 1.92×10^{-123} , Supplementary Data 6) respectively in regions bound by a transcription factor in cell lines studied in the ENCODE project⁴⁹. Expression levels of 44% (415/943) of reported human transcription factors⁵⁰ that are expressed in both cell types differ by >0.5 \log_2 between neutrophils and monocytes (Supplementary Data 7), making it plausible that alteration in transcription factor binding sites in addition to differences in transcription factor abundance could lead to eQTL being observed in one cell type and not the other. About 1% (52 in neutrophils, 35 in monocytes) of peak eQTL lie within microRNA binding sites, a 3.44 ($p=1.05 \times 10^{-9}$)- and 4.14-fold (1.01×10^{-16}) enrichment for neutrophils and monocyte respectively (Supplementary Data 6 and Supplementary Data 8). For example, rs4559 is a common variant associated with *STAT6* expression (Figure 4f) in both cell types and is within the binding region of miR-18b. We note that although this analysis does not take into account whether the microRNA is expressed in the cell-type, the results are consistent with some eQTL having their effect through microRNA-directed transcriptional gene silencing⁵¹.

Neutrophil eQTL are enriched for disease-associated variants

eQTL mapping can provide functional insight into the basis for genetic association of complex traits. We systematically examined eQTL in neutrophils for association with complex traits. Of all eQTL in neutrophils 7.5% (14,390/191,767) are associated with one or more of 327 traits (of 1138 unique disease/traits curated in the NHGRI GWAS catalog⁵²) directly or through linkage ($r^2 > 0.8$) with the trait-associated variant (Supplementary Data 9). This represents a 2.7-fold (95% CI 2.67-2.77, $p < 2.2 \times 10^{-16}$) enrichment relative to all SNPs originally tested for association with gene expression (eQTL). Notable examples are highlighted in Figure 5a. Reciprocally, 7.8% (14390/183124) of trait-associated variants are, or tag, eQTL in neutrophils.

Collation of GWAS traits into disease categories demonstrates that enrichment is observed for gastrointestinal disorders (7.2-fold enrichment; $p < 2 \times 10^{-308}$), allergy (6.3 fold enrichment; $p = 1.3 \times 10^{-117}$), autoimmune (6.3-fold enrichment, $p < 2 \times 10^{-308}$) and infectious diseases (3.1-fold enrichment, $p = 7.9 \times 10^{-137}$) grouped together or disaggregated for viral (3.7-fold enrichment, $p = 1.8 \times 10^{-56}$), parasitic (5.1-fold enrichment, $p = 6.4 \times 10^{-96}$) and bacterial (3.1-fold enrichment, $p = 2 \times 10^{-34}$) diseases (Supplementary Data 10).

We note that many of the eQTL observed in neutrophils may also be observed in monocytes (Supplementary Figure 5) reinforcing the need for additional follow-up to resolve complexity of how a variant may predispose to disease through effects in one or more cell types. We therefore explored two eQTL in greater detail to demonstrate how integrated analysis can provide novel insights into disease.

An eQTL of PADI4 affects rheumatoid arthritic susceptibility

Rheumatoid arthritis (RA) is a systemic autoimmune disease afflicting ~0.5-1% of adults in which autoimmune destruction of synovial joints occurs⁵³. A major feature of disease is presence of autoantibodies directed against citrullinated proteins. We found that 13 of 101

loci (a 3-fold enrichment compared with background, 95% CI 1.5-5.4, $p=8\times 10^{-4}$) recently reported to be associated with RA risk in the largest meta-analysis to date⁵⁴ are eQTL in neutrophils. These include rs2240335, an eQTL for *PADI4* that is in near complete linkage disequilibrium with rs230188 ($r^2=0.93$ in our dataset). The A allele at rs2240335 is associated with elevated risk of RA and together with rs230188 is a genome-wide significant correlate of RA risk⁵⁴. Intriguingly, rs2240335-A is associated with increased expression of *PADI4* in neutrophils but reduced expression in monocytes (Figure 5c). rs74058715-T, a nearby SNP independent to rs2240335 ($r^2=0.02$), is associated with reduced *PADI4* expression in both cell types providing an additional instrument to probe the role of *PADI4* in neutrophils in RA risk (figure 5b). Examination of histone modifications in the region show that rs2240335 lies in a region marked by histone 3 lysine 27 acetylation (H3K27Ac) and histone 3 lysine 4 monomethylation (H3K4me1) in neutrophils but not in monocytes whereas rs74058715 is in a region marked in both cell types (Figure 5d), consistent with the cell type in which the eQTL is seen and suggesting rs2240335 as the functional variant as opposed to rs230188. Conditioning on rs2240335 does not reveal a secondary eQTL peak. Following from this prediction, rs74058715-T is nominally associated with reduced RA risk ($p=0.05$ in the largest meta-analysis of RA) showing directional consistency. We note that rs74058715 tags at least six other SNPs ($LD>0.8$) that are nominally associated with RA, and therefore may not itself be the causal SNP. Subsequent to oxidative responses by stimuli including rheumatoid factor, *PADI4* post-translationally citrullinates histones and other proteins initiating NETosis, a process shown to be central to RA pathogenesis⁵⁵⁻⁵⁷. Because *PADI4* expression is confined to neutrophils and monocytes⁵⁸, these data support a model in which rs230188 tags rs2240335-A, which alters *PADI4* expression and RA risk likely through its actions in neutrophils.

Functional basis of rs1981760 in leprosy and Crohn's disease

Leprosy, a disease caused by the infectious agent *Mycobacterium leprae*, and CD, an autoimmune inflammatory bowel disease, show profound overlap in genetic architecture⁵⁹. The ancestral T allele of rs1981760 is associated with increased susceptibility of leprosy, particularly multibacillary disease⁶⁰ yet has weak protective effects in CD independent to the major CD-associated missense mutations⁵⁹. We observed a strong association between rs1981760-T and reduced expression of both *NOD2* ($p=8\times 10^{-30}$, variance explained=39%), and the adjacent gene *SNX20* ($p=7.8\times 10^{-10}$) in neutrophils and conversely elevated expression in monocytes ($p=3.3\times 10^{-10}$) (Figure 6a-b). The frequency of rs1981760-C is strikingly differentiated in Asians compared with other populations (78% in Asians vs. 26% in Europeans from the 1000G project) (Figure 6c). To further test the mechanism of this finding and functional consequences, we enrolled an independent cohort of 23 individuals. We were able to replicate the observed eQTL by quantitative PCR with two independent probe sets (Figure 6d). The rs1981760 polymorphism is reported to alter *STAT3* binding in some ENCODE cell-lines. We note that rs1981760 is in complete LD with rs9302752, the SNP first reported to be associated with leprosy and CD, but since rs1981760 is the primary eQTL signal, and in view of it potentially modifying transcription factor binding, we pursued rs1981760. Conditioning on rs1981760 does not reveal a secondary eQTL peak. To test whether the polymorphism alters *STAT3* binding in neutrophils we performed chromatin immunoprecipitation with an antibody directed against *STAT3* in 4

heterozygotes. Using allele-specific probes in a digital droplet PCR reaction designed to accurately quantify the number of each allele, we observed significantly fewer copies of immunoprecipitated DNA containing T-alleles than C alleles after ChIP in each individual (Figure 6e and Supplementary Data 11) consistent with the hypothesis that the T allele reduces STAT3 binding either alone or potentially in a complex with other transcription factors. Moreover, *STAT3* is markedly differentially expressed in neutrophils and monocytes (Figure 6f). Finally, we stimulated neutrophils from individuals with the NOD2 ligand muramyl-dipeptide (supplemented with Pam3-CSK4 as a synergistic agonist). Individuals with the CC genotype expressed significantly greater levels of mRNA for *IFNB* ($p=0.02$) (Figure 6g) after stimulation. These data demonstrate the rs1981760 affects *NOD2* expression and subsequent *IFNB* responses to its ligand. Notably, eQTL in neutrophils are enriched for genes in the *IFNB* network ($p=5.9\times 10^{-4}$). Therefore these data suggest that type-1 interferons and neutrophils may be involved in leprosy as has been shown in tuberculosis⁶¹.

Neutrophil eQTL are enriched in regions of natural selection

Neutrophils are an evolutionarily ancient cell. In light of the rs1981760 observation, we systematically examined whether regulatory variants in neutrophils overlap regions that are identified as being subject to selection in systematic human surveys⁶². Indeed, 1527 variants that are eQTL for a total of 39 genes lie in regions that have been subject to selection in humans (Supplementary Data 12), a 2.4-fold (95%CI 2.33-2.6) enrichment ($p=1.04\times 10^{-195}$) relative to the 20464 variants tested in our study that lie in annotated regions of selection. Many of the genes showing eQTL driven by variants in such regions are known to be involved in leukocyte biology, for example the chemokine receptors *CCR1*⁶³ and *CCR3*, or *GUSB*, a gene in which mutations cause mucopolysaccharidosis VII⁶⁴ in which recurrent respiratory infections are common. Others, such as *HERC2* involved in eye colour⁶⁵ and for which the second strongest eQTL in neutrophils is observed, may suggest new hypotheses for operative selective forces and the cells to yield the patterns observed. Therefore, regulatory variants in neutrophils are likely to have been subject to selection, reinforcing their biological relevance.

In conclusion, these data illustrate the utility of eQTL mapping in different leucocyte subsets and integration with epigenetic and disease-association data to reveal mechanisms of disease.

Methods

Volunteer enrolment and cell purification

This study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55) and each individual gave informed consent to participation. The median age of the 101 included participants was 31 years (IQR 24-41), 50 were male and 51 female. For this substudy of our previously reported work¹⁹ we recruited 144 healthy Caucasian volunteers in Oxford, United Kingdom. A trained professional nurse and/or a medical doctor conducted a verbal review of clinical history to determine eligibility based on the absence of major chronic illness, current medication administration or symptoms of infection. We

excluded 43 individuals because we did not have genotyping information at the time of this report (23 individuals) or the sample was an outlier after principal components analysis (20 individuals). We used an automated outlier detection algorithm ‘detectOutlier’ (*lumi* package) based on Euclidean distance to the centre of the cluster with iterative outlier removal and normalisation. Whole blood was collected into sodium-heparin containing blood collection tubes (Becton Dickinson) and processed with 4-6 hours after collection.

We isolated peripheral blood mononuclear cells by density gradient centrifugation of blood-diluted with Hanks Buffered Saline solution (HBSS, Life Technologies, UK) layered on Lymphoprep (Axis-Shield, Norway), and sorted CD14⁺ monocytes using magnetic-activated cell sorting (MACS, Miltenyi Biotech) according to the manufacturers instructions.

We isolated granulocytes using density gradient centrifugation in which blood was layered on Polymorphprep (Axis-Shield, Norway) according to the manufacturers instructions. To remove red blood cell contamination we exposed granulocytes to endotoxin free cell-culture water (Life Technologies, UK) for 30 seconds and restored isosmosis using 2X HBSS. As neutrophils and eosinophils share density properties (1.085g/ml), and previous studies demonstrated that *ex vivo* granulocyte transcriptomes are dominated by eosinophil transcripts despite their relative paucity²⁷ (<5% of granulocytes), we further isolated CD16⁺ granulocytes as a pure population of neutrophils using CD16⁺ microbeads. Purity assessed on a representative sample was >90%. We performed full blood counts on a Sysmex haematology analyser in a certified clinical laboratory; median neutrophil counts were 3.07 (IQR 2.52-3.78)×10⁹ cells/L. We note that the use of beads directed against CD16 may, in principle, alter or stimulate neutrophils although all purification steps were carried out on ice or at 4°C. Cell viability after isolation was >80%. Lysed cell pellets were immediately cryopreserved until extraction.

Genotyping

A total of 733,202 variants were genotyped using the Illumina HumanOmniExpress-12v1.0 Beadchip. QC steps included removal of individuals that were outliers by PCA, had poor genotyping call rates or heterozygosity measures. Variant QC included removal of SNP's with MAF<5%, HWE<1×10⁻⁶ or poor call rates. Genome-wide imputation was performed by prephasing a scaffold of 588,170 SNP's with SHAPEIT⁶⁶ and imputation in IMPUTE2⁶⁷ using the 1000G phase 1 release as the reference. Imputed SNP or indels with an info score <0.9, MAF<5% or departure from HWE (1×10⁻³) were removed. In total 5,680,354 variants were included in the eQTL analysis. Locations, where reported, are according to the human genome build GRCh37. Measures of LD, where mentioned, are based on 1000G CEU⁶⁸ or, if specifically mentioned, from this cohort of volunteers.

Gene expression analysis

During optimization of methods to isolate RNA from neutrophils we found that commonly used column-based isolation techniques were vastly outperformed by phenol-chloroform extraction using Trizol LS (Life Technologies) according to the manufacturers instructions. This may be because a higher concentration of chaotropic and reducing agent is required to

fully inhibit the abundance of nucleases in neutrophils. Therefore RNA was extracted using Trizol LS and further cleaned using the RNA MinElute cleanup kit (Qiagen, UK). RNA was quantified and integrity assessed using a Bioanalyser RNA 6000 Nano kit (Agilent, UK). Gene expression was quantified using the Illumina HumanHT-12 v4 BeadChip gene expression array platform with 47,231 probes according to the manufacturers instructions. Samples were randomized across expression chips and run in a single batch.

Gene expression data were normalized using random-spline normalization, transformed by variance-stabilising transformation and sample outliers were iteratively removed and normalization repeated. We note that of the 123 samples hybridised, 1 failed, and 20 were removed as they were sample outliers-almost universally (19/20 samples) due to cRNA quality with these individuals having RNA size distribution <800 by BioAnalyser RNA 6000 Nano kit evaluation. We therefore did not obtain replacement arrays and hence all the data in this study derive from a single batch. Probe sequences mapping to more than one genomic location or regions with underlying polymorphisms frequent in >1% of the population were excluded from eQTL analysis (n=18220 probes). Only probes that were expressed and detected (GenomeStudio probe detection $p < 0.01$) in neutrophils (n=11,023 probes for 9147 genes) were included in the primary analysis. For comparisons between neutrophils and monocytes, 10,012 probes for 8362 genes were included from 99 individuals.

eQTL mapping

We tested for association between genetic variation (using dosage estimates for imputed variants) within 1Mb of the expression probe (*cis*) or more distantly (*trans*), and gene expression in a linear regression framework implemented in 'MatrixEQTL'³⁵ in *R* (Matrix_eQTL_main function), incorporating principal components as covariates to account for hidden confounders (empirically determined number of PCs =15). We filtered results by a false discovery rate of 5% for *cis* associations or 1% for *trans* associations. We caution that this study, due to its modest size, is not adequately powered to exhaustively identify eQTL in neutrophils, or to map *trans* eQTL.

Annotation of eQTL with genomic features

Identified eQTL were annotated by location relative to the transcription start site or gene structures using data from the UCSC genome browser (GRCh37/b19)⁶⁹.

We accessed whole genome bisulfite sequencing and histone chromatin immunoprecipitation data generated by the BLUEPRINT consortium in primary human monocytes and neutrophils from 4-8 individuals (depending on feature). We defined features marked by an epigenetic mark as those present in at least half of all individuals (Supplementary Table 5).

For analyses of eQTL overlap with the location of epigenetic marks⁴⁶, enhancers^{47,48}, transcription factors⁴⁹, conserved microRNA binding sites⁷⁰ and in regions that have been under selection⁶², the location of the given feature relative to an eQTL was calculated using the 'GenomicRanges' package in *R* with the 'distanceToNearest' function. Fishers exact test was used to evaluate enrichment relative to all imputed variants tested for eQTL (ie those

within 1Mb of a gene expressed in neutrophils and monocytes). We note that an implicit assumption in the Fisher's test is of independence of loci. The code used to generate the results is available from the authors.

Overlap between eQTL and trait/disease associated variants

All traits listed in the NHGRI Catalog Genome-Wide Association Studies (1138 traits as at 14 October 2014) were abstracted and expertly curated into one or more categories according to the disease system affected (28 potential categories) 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. Classification of a given trait was possible into multiple groups and to capture this diversity, for each trait we assigned trait membership into up to three possible groups.

A Fishers exact test comparing the proportion of eQTL that are trait associated variants or in linkage disequilibrium with these variants ($r^2 > 0.8$ in 1000G Caucasian populations) to the proportion of all tested variants that are associated variants was performed stratified by disease category.

Neutrophil stimulation

To identify the role of rs1981760 in NOD2 downstream effects, we stimulated $5-10 \times 10^6$ freshly isolated neutrophils resuspended at 2×10^6 cells/ml in RPMI1640 supplemented with 10% foetal calf serum and L-glutamine with $1 \mu\text{g/ml}$ muramyl dipeptide (n-acetyl) and $1 \mu\text{g/ml}$ Pam3CSK4, (both from Invivogen, UK) for two hours. Primer sequences for quantitative PCR are reported in Supplementary Data 13.

Chromatin immunoprecipitation

Crosslinked DNA from neutrophils was stored and subjected to chromatin immunoprecipitation (ChIP) after fragmentation of crosslinked DNA by sonication consisting two sets of 9 cycles of 30s each in a BioRuptor (Diagenode, Belgium) at high power. A mouse anti-human STAT3 antibody (Cell Signaling-NEB, UK, catalogue #4904S) was used at a 1:50 dilution in conjunction with magnetic beads to immunoprecipitate STAT3 bound DNA. Quality control of ChIP-ped DNA was achieved by quantification on a Qubit 2.0 fluorometer (Invitrogen) using the Quant-iT dsDNA HS Assay Kit (Invitrogen).

Assessment of allele-specific STAT3 binding in neutrophils

For allele-specific digital droplet PCR (ddPCR) quantification of rs1981760-C compared with rs1981760-T we designed allele-specific probes (Supplementary Data 13), and performed amplification using the ddPCR Supermix for Probes (BioRad) according to the supplied protocol. The probe for the T and C allele were designed to fluoresce in different channels allowing simultaneous detection of both in a single reaction. Droplets were generated on a qx100 droplet generator (BioRad, UK) and droplets read on a qx100 droplet

reader. Probes targeting SOCS3 and JAK3 were used as positive controls for ChIP of STAT3, and RAB4 as a negative control. Input DNA as well as DNA obtained after ChIP was amplified on the same plate. For allele-specific wells, we performed technical duplicates.

The proportion of positive droplets relative to negative droplets is associated with the absolute concentration of the product according to a poisson distribution. We therefore exploited this to empirically calculate the expected proportion of C vs. T droplets (so as to take into account possible probe efficiency differences) and compared this to the actual ratio after ChIP. P values using a chi-squared test with 2 degrees of freedom were used to estimate the probability of the observed vs. the expected ratio being due to chance.

Software used for analysis

Analyses were conducted in *R*, using the following packages: limma, lumi, 'annotate', 'vsn', 'GenomicRanges', 'qvalue', 'data.table'. For data presentation the 'gridExtra', 'ggbio', 'ggplot2' and 'VennDiagram' packages were used. Local association plots were created using LocusZoom. Ingenuity Pathway Analysis where used, was performed using a background set of all human genes on the Illumina array. A listing of external files and their sources is given in Supplementary Data 14. FDR where shown denotes the false discovery rate calculated using the Benjamini Hochberg procedure using the p.adjust function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Wellcome Trust (Grants 074318 [J.C.K.], 088891 [B.P.F.], and 090532/Z/09/Z [core facilities Wellcome Trust Centre for Human Genetics including High-Throughput Genomics Group]), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement no. 281824 (J.C.K.), the Medical Research Council (98082 [J.C.K]) and the NIHR Oxford Biomedical Research Centre. V.N. was supported by the Rhodes Trust. We thank the volunteers for their participation in this study, Sr Jane Cheeseman for assistance in enrolling participants, Dr Katharine Plant for administrative assistance during the conduct of this study and Dr Christina Chang for help with illustrations. We also thank Prof's Yuta Kochi, Alison Simmons, Dr's Yukinori Okada and Luke Jostins and Mr Daniel Gaughan for helpful discussions.

References

1. Wright FA, et al. Heritability and genomics of gene expression in peripheral blood. *Nature genetics*. 2014; 46:430–437. [PubMed: 24728292]
2. Wilhelm M, et al. Mass-spectrometry-based draft of the human proteome. *Nature*. 2014; 509:582–587. [PubMed: 24870543]
3. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*. 2009; 325:1246–1250. [PubMed: 19644074]
4. Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013; 501:506–511. [PubMed: 24037378]
5. Montgomery SB, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*. 2010; 464:773–777. [PubMed: 20220756]
6. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010; 464:768–772. [PubMed: 20220758]

7. Stranger BE, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS genetics*. 2012; 8:e1002639. [PubMed: 22532805]
8. Stranger BE, et al. Population genomics of human gene expression. *Nature genetics*. 2007; 39:1217–1224. [PubMed: 17873874]
9. Battle A, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome research*. 2014; 24:14–24. [PubMed: 24092820]
10. Fehrmann RS, 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 genetics*. 2011; 7:e1002197. [PubMed: 21829388]
11. Mehta D, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *European journal of human genetics : EJHG*. 2013; 21:48–54. [PubMed: 22692066]
12. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics*. 2013; 45:1238–1243. [PubMed: 24013639]
13. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature genetics*. 2012; 44:502–510. [PubMed: 22446964]
14. Raj T, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014; 344:519–523. [PubMed: 24786080]
15. Zeller T, et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PloS one*. 2010; 5:e10693. [PubMed: 20502693]
16. Lee MN, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*. 2014; 343:1246980. [PubMed: 24604203]
17. Ferraro A, et al. Interindividual variation in human T regulatory cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E1111–1120. [PubMed: 24610777]
18. Barreiro LB, et al. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:1204–1209. [PubMed: 22233810]
19. Fairfax BP, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014; 343:1246949. [PubMed: 24604202]
20. Kim S, et al. Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes. *Nature communications*. 2014; 5:5236.
21. Gibbs JR, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics*. 2010; 6:e1000952. [PubMed: 20485568]
22. Nica AC, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS genetics*. 2011; 7:e1002003. [PubMed: 21304890]
23. Ramasamy A, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience*. 2014; 17:1418–1428. [PubMed: 25174004]
24. Schadt EE, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS biology*. 2008; 6:e107. [PubMed: 18462017]
25. Mostafavi S, et al. Variation and genetic control of gene expression in primary immunocytes across inbred mouse strains. *Journal of immunology*. 2014; 193:4485–4496.
26. Grisham MB, Engerson TD, McCord JM, Jones HP. A comparative study of neutrophil purification and function. *Journal of immunological methods*. 1985; 82:315–320. [PubMed: 2995497]
27. Stejskal S, Koutna I, Rucka Z. Isolation of granulocytes: which transcriptome do we analyse - neutrophils or eosinophils? *Folia biologica*. 2010; 56:252–255. [PubMed: 21324266]
28. Westra, H., et al. Cell specific eQTL analysis without sorting cells. *bioRxiv*. 2014. doi:<http://dx.doi.org/10.1101/002600>
29. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil function: from mechanisms to disease. *Annual review of immunology*. 2012; 30:459–489.

30. Kolaczowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nature reviews. Immunology*. 2013; 13:159–175.
31. Bennouna S, Bliss SK, Curiel TJ, Denkers EY. Cross-talk in the innate immune system: neutrophils instruct recruitment and activation of dendritic cells during microbial infection. *Journal of immunology*. 2003; 171:6052–6058.
32. Diana J, et al. Crosstalk between neutrophils, B-1a cells and plasmacytoid dendritic cells initiates autoimmune diabetes. *Nature medicine*. 2013; 19:65–73. doi:10.1038/nm.3042.
33. Pelletier M, et al. Evidence for a cross-talk between human neutrophils and Th17 cells. *Blood*. 2010; 115:335–343. [PubMed: 19890092]
34. von Bruhl ML, et al. Monocytes, neutrophils, and platelets cooperate to initiate and propagate venous thrombosis in mice in vivo. *The Journal of experimental medicine*. 2012; 209:819–835. [PubMed: 22451716]
35. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012; 28:1353–1358. [PubMed: 22492648]
36. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics*. 2007; 3:1724–1735. [PubMed: 17907809]
37. Buil A, et al. C4BPB/C4BPA is a new susceptibility locus for venous thrombosis with unknown protein S-independent mechanism: results from genome-wide association and gene expression analyses followed by case-control studies. *Blood*. 2010; 115:4644–4650. [PubMed: 20212171]
38. Neumann K, et al. Clec12a is an inhibitory receptor for uric acid crystals that regulates inflammation in response to cell death. *Immunity*. 2014; 40:389–399. [PubMed: 24631154]
39. Veyrieras JB, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS genetics*. 2008; 4:e1000214. [PubMed: 18846210]
40. Dharmawardhane S, Brownson D, Lennartz M, Bokoch GM. Localization of p21-activated kinase 1 (PAK1) to pseudopodia, membrane ruffles, and phagocytic cups in activated human neutrophils. *Journal of leukocyte biology*. 1999; 66:521–527. [PubMed: 10496324]
41. Bardeel BW, Kenny EF, Sollberger G, Zychlinsky A. The balancing act of neutrophils. *Cell host & microbe*. 2014; 15:526–536. [PubMed: 24832448]
42. Borregaard N, Sorensen OE, Theilgaard-Monch K. Neutrophil granules: a library of innate immunity proteins. *Trends in immunology*. 2007; 28:340–345. [PubMed: 17627888]
43. McCracken JM, Allen LA. Regulation of human neutrophil apoptosis and lifespan in health and disease. *Journal of cell death*. 2014; 7:15–23. [PubMed: 25278783]
44. Ambruso DR, et al. Human neutrophil immunodeficiency syndrome is associated with an inhibitory Rac2 mutation. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:4654–4659. [PubMed: 10758162]
45. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature genetics*. 2013; 45:124–130. [PubMed: 23263488]
46. Adams D, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nature biotechnology*. 2012; 30:224–226.
47. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014; 507:455–461. [PubMed: 24670763]
48. Consortium F, et al. A promoter-level mammalian expression atlas. *Nature*. 2014; 507:462–470. [PubMed: 24670764]
49. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*. 2012; 22:1790–1797. [PubMed: 22955989]
50. Ravasi T, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*. 2010; 140:744–752. [PubMed: 20211142]
51. Kim DH, Saetrom P, Snove O Jr, Rossi JJ. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:16230–16235. [PubMed: 18852463]
52. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*. 2014; 42:D1001–1006. [PubMed: 24316577]

53. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet*. 2010; 376:1094–1108. [PubMed: 20870100]
54. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014; 506:376–381. [PubMed: 24390342]
55. Khandpur R, et al. NETs are a source of citrullinated autoantigens and stimulate inflammatory responses in rheumatoid arthritis. *Science translational medicine*. 2013; 5:178ra140.
56. Sohn, DH., et al. American College of Rheumatology. Wiley; Boston, USA: 2014.
57. Wright HL, Moots RJ, Edwards SW. The multifactorial role of neutrophils in rheumatoid arthritis. *Nature reviews. Rheumatology*. 2014; 10:593–601. [PubMed: 24914698]
58. Suzuki A, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature genetics*. 2003; 34:395–402. [PubMed: 12833157]
59. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012; 491:119–124. [PubMed: 23128233]
60. Zhang FR, et al. Genomewide association study of leprosy. *The New England journal of medicine*. 2009; 361:2609–2618. [PubMed: 20018961]
61. Mayer-Barber KD, et al. Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature*. 2014; 511:99–103. [PubMed: 24990750]
62. Grossman SR, et al. Identifying recent adaptations in large-scale genomic data. *Cell*. 2013; 152:703–713. [PubMed: 23415221]
63. Lionakis MS, et al. Chemokine receptor Ccr1 drives neutrophil-mediated kidney immunopathology and mortality in invasive candidiasis. *PLoS pathogens*. 2012; 8:e1002865. [PubMed: 22916017]
64. OMIM. Online Mendelian Inheritance in Man, OMIM®. MIM Number: 253220: 02/12/2014. 2014. <<http://omim.org/>>
65. Sturm RA, et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *American journal of human genetics*. 2008; 82:424–431. [PubMed: 18252222]
66. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nature methods*. 2012; 9:179–181. [PubMed: 22138821]
67. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. 2009; 5:e1000529. [PubMed: 19543373]
68. Genomes Project, C. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
69. Karolchik D, et al. The UCSC Genome Browser database: 2014 update. *Nucleic acids research*. 2014; 42:D764–770. [PubMed: 24270787]
70. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome biology*. 2010; 11:R90. [PubMed: 20799968]

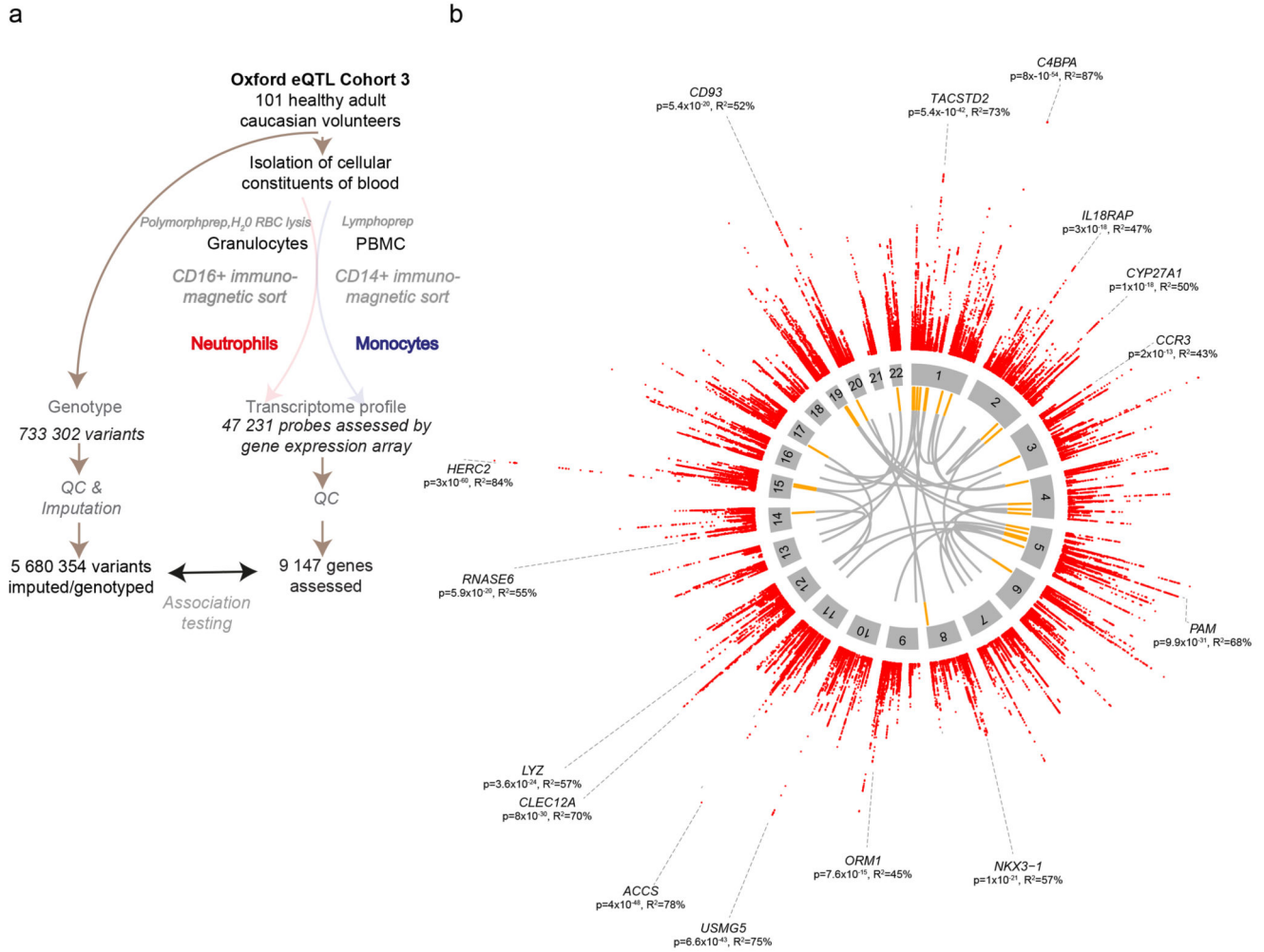


Figure 1. Study schema and overview of eQTL in primary human neutrophils

(a) Flowchart showing the schema of this study to identify eQTL in human neutrophils from healthy adult volunteers of European ancestry with genotypes determined by array-genotyping and by imputation tested for association with global gene expression in *cis* (variant <1Mb of gene) or *trans* (variant on non-contiguous chromosome to gene). This analysis was contemporaneous to our previously reported study¹⁹ of monocytes enabling direct comparison of genetic correlates of gene expression between neutrophils and monocytes. (b) Circos plots for the neutrophil dataset. Outermost rim (red dots) shows a Manhattan plot for significant *cis* eQTL (FDR<0.05) with names of genes with large effect sizes (variance explained, R²>45%) or lowest five p-values; the second rim (grey boxes) shows the chromosome ideogram with chromosome number within each box; the third rim (orange boxes) indicates a gene to which a significant (FDR<0.01) *trans*-eQTL was observed; and the innermost spokes (grey) connect *trans*-acting variants to the gene the variant is associated with.

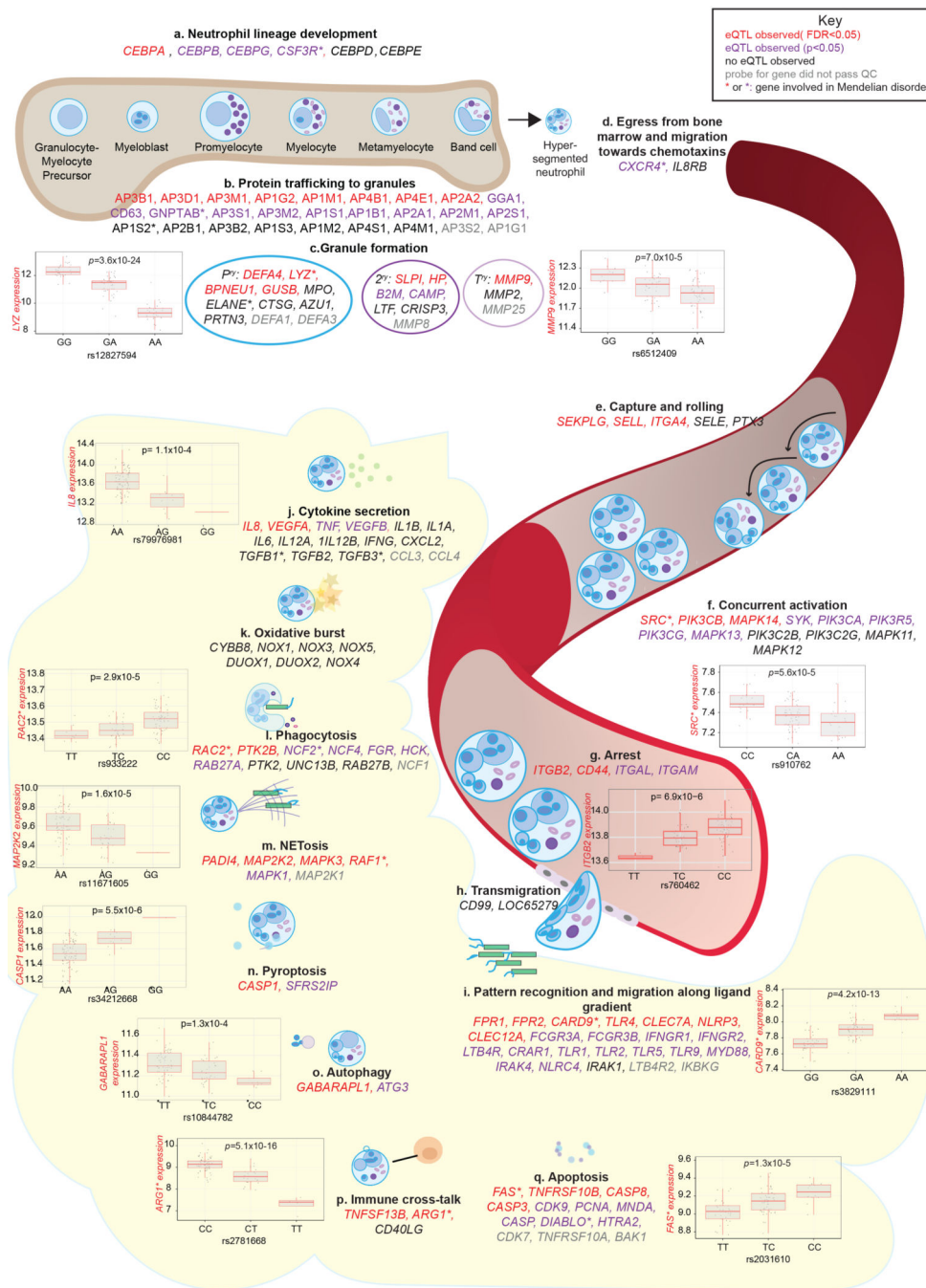


Figure 2. eQTL in genes involved in neutrophil biology

As a narrative resource, we compiled a list of 164 genes of importance in neutrophil biology^{29,30,41-43} and illustrate their role together with eQTL information (denoted by gene colour). Probes for 113 genes passed QC: 9 genes had no eQTL (black gene names) and 104 had an eQTL (47 with FDR<0.05 red gene names, 57 with p<0.05 but FDR>0.05 purple gene names). Genes involved in nearly every aspect of neutrophil biology have identifiable eQTL and several are involved in Mendelian disorders (asterisked genes) for example *RAC2* (neutrophil immunodeficiency syndrome) and *FAS* (autoimmune lymphoproliferative

syndrome type 1A). (a) Granulocyte-monocyte progenitor cells (GMP) respond to GM-CSF stimulation of a receptor encoded by *CSFR3* to form neutrophils under the instruction of CEBP α and other CEBP transcription factors. Neutrophil development through various stages is coincidental with expression of proteins that traffic to primary secondary and tertiary granules. (b) AP heterotetramers are involved in trafficking of proteins such as LYZ (involved in renal amyloidosis) and MMP9 (involved in metaphyseal osteolysis, nodulosis and arthropathy) to (c) primary, secondary and tertiary granules. (d) Hyper-segmented neutrophils egress from the bone marrow in response to IL-8 and SDF-1, and are (e) captured on endothelium within blood vessels through interaction of selectins and ligands such as SEKPLG and SELL. (f) Capture, rolling and eventual arrest on endothelial surfaces is associated with activation, mediated in part by PI-3 and MAPK pathways. (g) Arrest of neutrophils is mediated by $\alpha 1\beta 2/\alpha M\beta 2$ integrin heterodimers encoded by *ITGAL*, *ITGBA2*, *ITGAM* and (h) rapid para- and trans-cellular migration ensues. (i) Neutrophils express pattern recognition receptors and receptors for host derived molecules allowing detection of bacteria and effective migration. In tissues, neutrophil response includes (j) cytokine secretion, (k) oxidative burst, (l) phagocytosis, (m) production of neutrophil extracellular traps (NETosis), (n) pyroptosis, (o) autophagy, (p) cross-talk with other immune, and eventual cell death via apoptosis should cell death not have eventuated from another response. Box lower and upper border denote 25th and 75th centiles respectively, central line denotes median and whiskers extend to 1.5*IQR. In all cases 101 donor replicates shown.

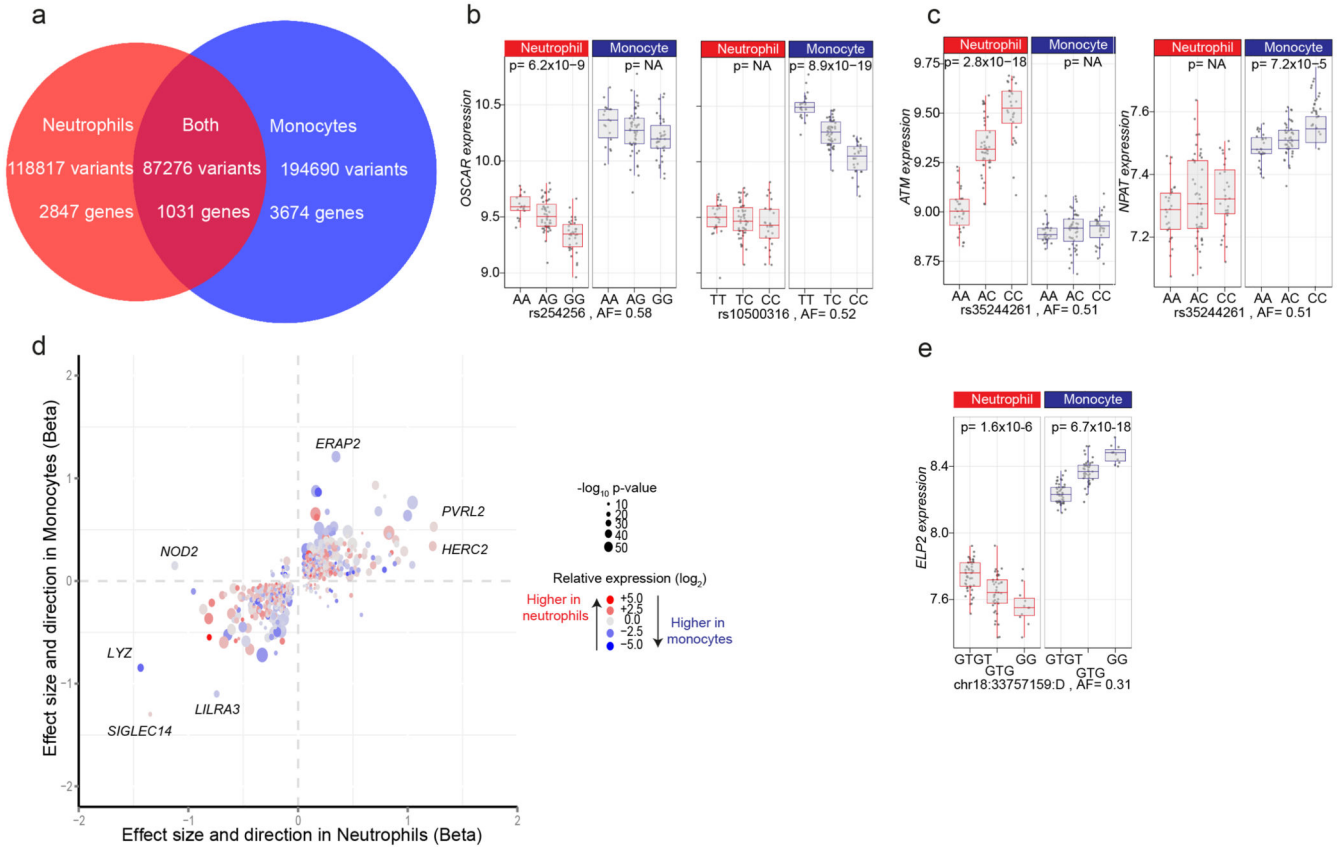


Figure 3. Shared and cell-type specific *cis* eQTL in neutrophils and monocytes

(a) Venn diagrams showing the number of variants and genes with *cis*-sQTL by cell type. Examples of cell-type constraint include (b) independent variants associated with expression of the same gene, *OSCAR*, specific to either neutrophils or monocytes (lead eSNPs rs254256 and rs1050031 respectively, $r^2 < 0.02$). Conversely, (c) a single variant may be associated with different genes in each cell type as shown for rs35244261 (associated with elevated expression of *ATM* in neutrophils, and *NPAT* in monocytes). (d) Amongst 1939 genes that have an eQTL in both cell types, 1031 involve the same variant and are plotted with the effect size in neutrophils and monocytes (shown on x- and y-axes respectively). Color denotes the relative expression in neutrophils and monocytes, and size denotes the minimum p value of an eQTL for that gene. Several eQTL show divergent direction of effects on the gene they regulate including *NOD2*, *THBD*, *TSTD1*, *TOM1L2* and *ELP2* (latter shown in e). Genes with large effect sizes are highlighted, including *HERC2*, the gene responsible for iris colour, which has amongst the most significant eQTL of all genes in neutrophils. P values greater than 0.05 are denoted as p=NA. Box lower and upper border denote 25th and 75th centiles respectively, central line denotes median and whiskers extend to 1.5*IQR. In all cases 101 donor replicates shown.

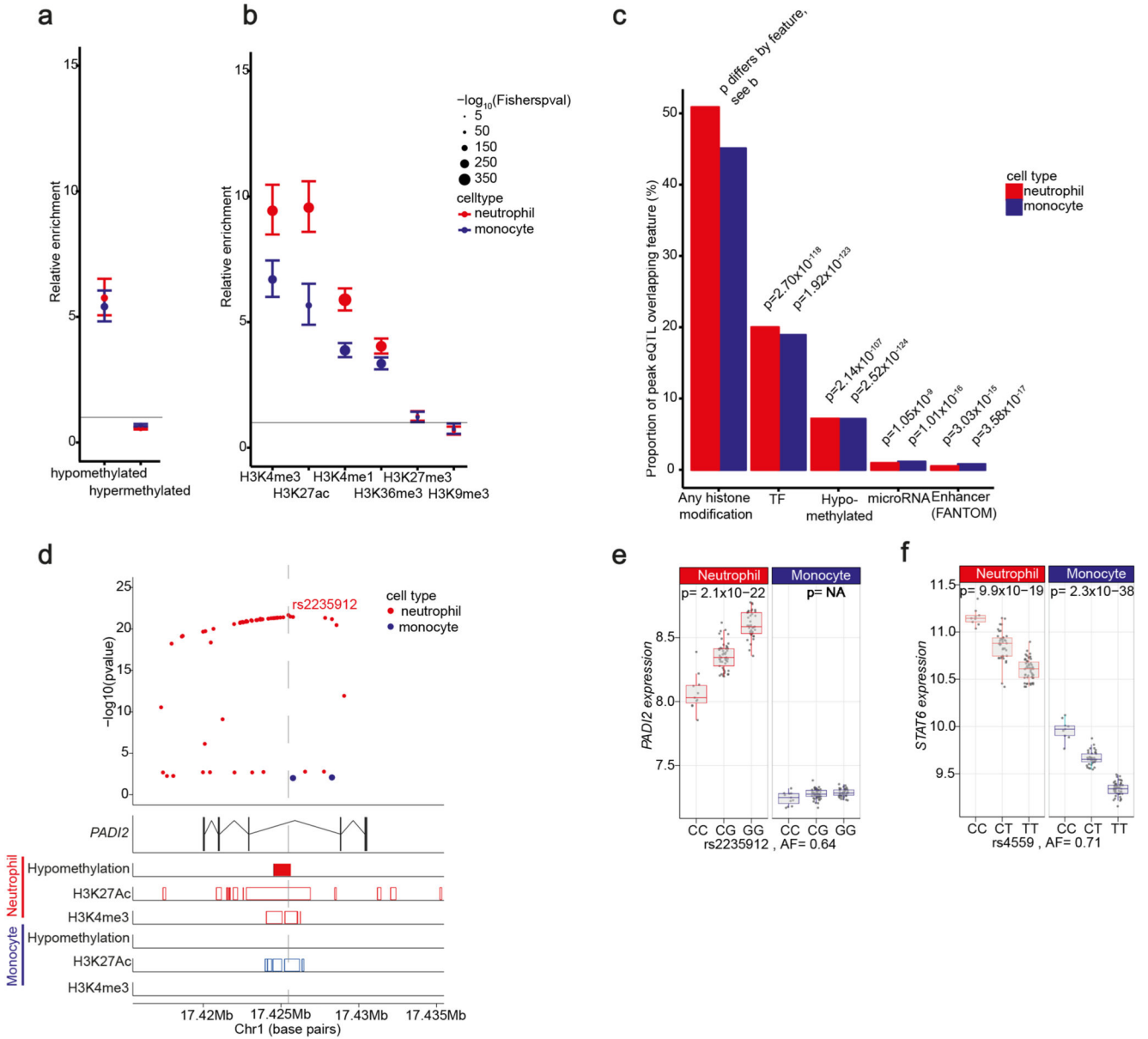


Figure 4. Epigenetic basis of eQTL in neutrophils and monocytes

We integrated eQTL data with whole-genome bisulfite sequencing and chromatin-immunoprecipitation sequencing data from primary neutrophils and monocytes from 4-8 individuals from the BLUEPRINT Consortium. (a) Enrichment of *cis* eQTL in neutrophils and monocytes, in hypo- and hyper-methylated regions of the genome in concordant cell types relative to all SNP's tested for eQTL. (b) Enrichment of neutrophil and monocyte *cis* eQTL relative to all variants tested for colocalisation in regions of the genome associated with modified histones based on ChIP-Seq of modified histones in neutrophils and monocytes. (c) Proportion of *cis* eQTL in neutrophils (n=3774) and monocytes (n=4668) that overlie modified histones in the concordant cell type (BLUEPRINT), transcription factor binding sites in cell lines (ENCODE), sites that are hypomethylated in the concordant

cell type (BLUEPRINT), target sites of conserved miRNAs with high mirSVR scores (microRNA.org) or CAGE-seq defined enhancer sites in neutrophils and monocytes (FANTOM5). Note that an eQTL may overlap more than one feature. (d-e) An example of a cell type specific eQTL possibly attributable to methylation differences in cell types is rs2235912, a site which is in an intron of *PADI2* that is hypomethylated in neutrophils but not monocytes. This locus is marked by the activating histone marks H3K27Ac and H3K4me3 in neutrophils but in monocytes has a smaller region marked by H3K27Ac only. As shown in (e), the gene is more highly expressed in neutrophils than monocytes, perhaps due to the greater activating histone marks in neutrophils. The G allele creates an additional putative CpG site which may explain the allele being associated with higher *PADI2* expression relative to the C allele. (f) *STAT6* is an example of a gene with an eQTL in both monocytes and neutrophils where the eQTL lies in a predicted binding region for mir-18b, a microRNA. Tails in panels A and B show 95% CI of the enrichment estimate. In panel E and F, box lower and upper border denote 25th and 75th centiles respectively, central line denotes median and whiskers extend to 1.5*IQR. In all cases 101 donor replicates shown. P values greater than 0.05 are denoted as p=NA.

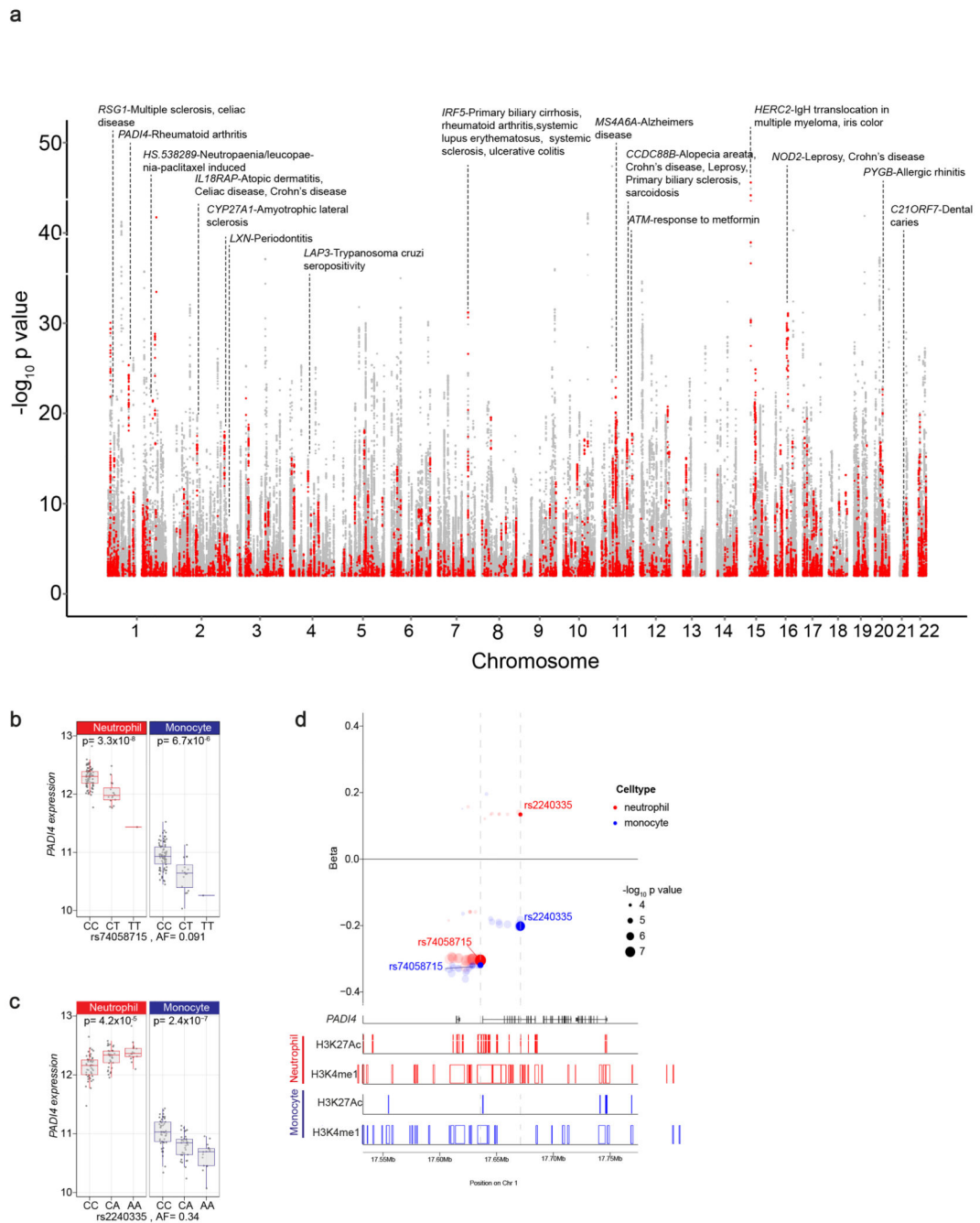


Figure 5. eQTL in neutrophils and their association with complex disease or trait
 (a) Manhattan plot showing eQTL in neutrophils highlighting those associated with or in linkage disequilibrium ($r^2 > 0.8$) with a disease/trait associated variant listed in the NHGRI GWAS catalog. Each point denotes a single eQTL and is coloured red if the locus is associated with a complex trait or grey if not. (b-d) An example in *PADI4* demonstrates how integrated analysis of an eQTL informs understanding of rheumatoid arthritis (RA) risk. Expression level of *PADI4* is associated with two independent variants, (b) rs2240335 and (c) rs74058715 ($r^2 = 0.02$). rs2240335-A, the derived allele, is associated with elevated

expression of *PADI4* in neutrophils and diminished levels in monocytes. rs7405871-T is associated with reduced *PADI4* expression in neutrophils and monocytes. rs2240335 is in near complete linkage with rs2301888, and this locus is associated with RA risk⁵⁴ (d) Compiled plot showing effect size estimate (beta) for variants associated with *PADI4* expression in neutrophils and monocytes (upper panel), relative to genic structures (track two), and BLUEPRINT ChIP-Seq reads for two histone marks in neutrophils (tracks three and four) and monocytes (tracks 5 and 6) for H3K27Ac and H3K4me1 demonstrating that rs2240335 lies in a region marked by both H3K27Ac and H3K4me1, a marker of active enhancers, in neutrophils but not in monocytes. In panel B and D box lower and upper border denote 25th and 75th centiles respectively, central line denotes median and whiskers extend to 1.5*IQR. In all cases 101 donor replicates shown.

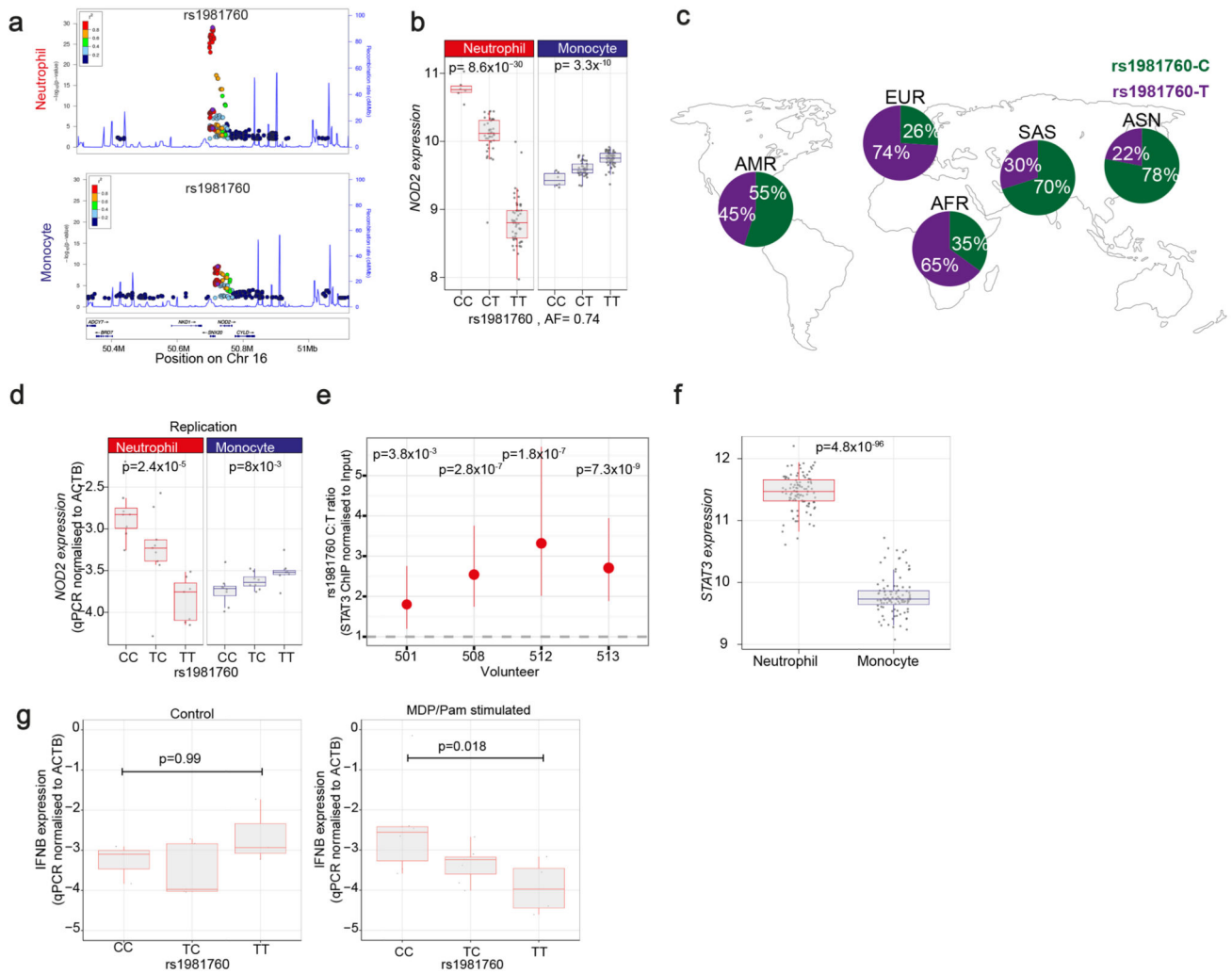


Figure 6. eQTL in *NOD2* and their involvement in Leprosy and Crohn's Disease (CD)

(a) Regional association plots for rs1981760 and *NOD2* in Neutrophils (top) and monocytes (bottom). (b) The C (derived) allele at rs1981760, a variant located in an intergenic region between *NOD2* and *SNX20*, is associated with elevated *NOD2* expression in neutrophils and reduced *NOD2* expression in monocytes, n=101. (c) rs1981760 is remarkable for the profound differentiation this allele appears to have undergone in Asian populations with the derived C allele predominating in Asians (data from the 1000 genomes project⁶⁸). rs1981760 and rs9302752 are in high linkage disequilibrium ($r^2=0.98$ in our cohort). rs9302752-C is associated with leprosy risk in GWAS reports from Chinese populations⁶⁰. The C variant has a mild protective effect in CD independent of the deleterious 3020insC frameshift mutation in *NOD2*⁵⁹. (d) Quantitative PCR of *NOD2* with two different primer sets performed blinded to genotype status in an independent set of 23 donors replicates the finding. Data representative of one primer-pair. (e) The rs1981760 SNP is located within a STAT3 transcription factor ChIP-seq binding site in MCF10A-Er-Src, an epithelial cell line (ENCODE). We investigated whether rs1981760 alters STAT3 binding in neutrophils by ChIP of STAT3 in four rs1981760 heterozygotes (screened to not have the 3020insC

frameshift mutation in NOD2). Allele-specific probes in a digital droplet PCR assay demonstrate that the C allele is significantly more frequently immunoprecipitated with STAT3 than the T-allele. p values denote a chi-squared test comparing expected vs. actual ratio of positive and negative droplets for each donor. Tails show 95% CI of the enrichment estimate. (f) *STAT3* expression levels differ between neutrophils and monocytes as measured by gene expression array, n=101. (g) After stimulation with muramyl dipeptide (MDP), the ligand for NOD2 and Pam3CSK4 (a co-stimulant) neutrophils from rs1981760-CC homozygous individuals express significantly higher levels of *IFNB* mRNA than TT homozygotes, and heterozygotes express intermediate levels (left panel). No such association is apparent in mock-treated control neutrophils (right panel). In panels B, D, F and G, box lower and upper border denote 25th and 75th centiles respectively, central line denotes median and whiskers extend to 1.5*IQR.