Published in final edited form as:

Nat Genet. 2013 October ; 45(10): 1238–1243. doi:10.1038/ng.2756.

Systematic identification of *trans*-eQTLs as putative drivers of known disease associations

A full list of authors and affiliations appears at the end of the article.

[#] These authors contributed equally to this work.

Abstract

Identifying the downstream effects of disease-associated single nucleotide polymorphisms (SNPs) is challenging: the causal gene is often unknown or it is unclear how the SNP affects the causal gene, making it difficult to design experiments that reveal functional consequences. To help overcome this problem, we performed the largest expression quantitative trait locus (eQTL) meta-analysis so far reported in non-transformed peripheral blood samples of 5,311 individuals, with replication in 2,775 individuals. We identified and replicated *trans*-eQTLs for 233 SNPs (reflecting 103 independent loci) that were previously associated with complex traits at genome-wide significance. Although we did not study specific patient cohorts, we identified trait-associated SNPs that affect multiple *trans*-genes that are known to be markedly altered in patients: for example, systemic lupus erythematosus (SLE) SNP rs4917014[1] altered *C1QB* and five type 1 interferon response genes, both hallmarks of SLE[2-4]. Subsequent ChIP-seq data analysis on these *trans*-genes implicated transcription factor *IKZF1* as the causal gene at this locus, with DeepSAGE RNA-sequencing revealing that rs4917014 strongly alters 3' UTR levels of *IKZF1*. Variants associated with cholesterol metabolism and type 1 diabetes showed similar phenomena, indicating that large-scale eQTL mapping provides insight into the downstream effects of many trait-associated variants.

Genome-wide association studies (GWAS) have identified thousands of variants that are associated with complex traits and diseases. However, because most variants and their proxies are non-coding, it is generally difficult to identify the causal genes. Recently, several eQTL-mapping studies[5-8] have now shown that the majority of disease-predisposing variants actually affect gene expression levels of nearby genes (i.e. *cis*-eQTLs). A few recent studies have also identified *trans*-eQTLs[5,9-13], revealing the downstream consequences of some variants. However, the total number of reported *trans*-eQTLs is fairly low, mainly due to the severe burden of multiple testing. To improve statistical power, we performed an eQTL meta-analysis in 5,311 peripheral blood samples, from seven studies

Corresponding author: Lude Franke (lude@ludesign.nl).

Author contributions

Experiment design and method development: H-J.W., M.J.P., T.E., H.Y., C.S., J.K., M.W.C., Y.L., J.K., R.C.J., T.L., B.M.P., T.M.F., J.B.J.M., L.F.

Reviewing and editing of the manuscript: H-J.W., M.J.P., T.E., H.Y., C.S., J.K., M.W.C., R.C.J., B.P.F., A.Z., A.H., V.S., C.W., A.G.U., F.R., J.B., S.A.G., Y.L., J.C.K., P.M.V., T.L., B.M.P., S.R., A.T., T.M.F., A.M. J.B.J.M., L.F.

Data collection: B.P.F., K.H., J.E.P., J.V.H., L.H.B., A.H., E.R., K.F., M.N., L.M., D.M., L.F., A.B.S., D.G.H.,M.A.N., G.H., M.N., D.R., U.V., M.P., V.S., J.B., A.S-D., S.A.G, D.A.E., T.L., S.M., H.P., C.H.,H.G., T.M., K.S., P.M.V., J.C.K., B.M.P., A.M.

Replication of *trans*-eQTL results: B.P.F., K.H., J.E.P., G.W.M.

using Illumina gene expression arrays (EGCUT[14], InCHIANTI[15], Rotterdam Study[16], Fehrmann[5], HVH[17-19], SHIP-TREND[20], and DILGOM[21]) and replication analysis in another 2,775 samples. We aimed to ascertain to what extent SNPs affect genes in *cis* and *trans* and whether eQTL mapping in peripheral blood could reveal important downstream pathways that may be putative drivers of disease processes.

Our genome-wide analysis identified *cis*-eQTLs for 44% of all tested genes (6,418 genes at probe-level false discovery rate (FDR) < 0.05 and 4,690 genes with more stringent Bonferroni multiple testing correction, Table 1, Supplementary Table 1, Supplementary Figures 1-3). Our *trans*-eQTL analysis focused on 4,542 SNPs that have been implicated in complex disease or traits (derived from the "Catalog of Published GWAS"). In the discovery dataset, we detected *trans*-eQTLs at FDR < 0.05 for 1,513 significant *trans*-eQTLs that include 346 unique SNPs (8% of all tested SNPs, Table 1, Supplementary Table 2, Supplementary Figure 4 and 5). These SNPs affect the expression of 430 different genes (a more stringent Bonferroni correction revealed 643 significant *trans*-eQTLs, including 200 unique SNPs and 223 different genes).

We used stringent procedures for *trans*-eQTL detection (Supplementary methods), and various benchmarks to ensure reliability: for 26 *trans*-eQTL genes the eQTL SNP affected multiple probes within these genes (Supplementary Table 3), always with consistent allelic directions, suggesting that our probe filtering procedure was effective in preventing false-positive *trans*-eQTLs. We observed uniform directionality for 90% of the tested *trans*-eQTLs across all studies within our discovery meta-analysis (Supplementary Figure 5). We did not find evidence that the eQTLs were driven by differences in age or blood cell-counts between individuals (Supplementary Results and Supplementary Table 4, Supplementary Figure 6). However, we cannot exclude this possibility entirely because FACS analyses on individual cell-types had not been conducted.

To ensure reproducibility of the *trans*-eQTLs of our current meta-analysis, we performed various analyses. We replicated previously reported blood *trans*-eQTLs[5] (Supplementary Table 5, Supplementary Results and Supplementary Figure 7) and replicated *trans*-eQTLs from our discovery meta-analysis in two independent studies of peripheral blood gene expression (52% in KORA F4[22], N = 740 samples and 79% in BSGS[23], N = 862 samples, FDR < 0.05, Supplementary Figure 8). Irrespective of significance, 91% and 93% of all 1,513 significant *trans*-eQTL SNP-probe combinations showed consistent allelic direction in these replication cohorts as compared to the discovery analysis. A meta-analysis of these two replication studies improved replication rates: 89% of the 1,513 *trans*-eQTLs were significantly replicated (FDR < 0.05), 99.7% of which showed a consistent allelic direction. Irrespective of significance, 97% of the *trans*-eQTLs showed a consistent allelic direction in this replication meta-analysis (Supplementary Figure 8). We found that some *trans*-eQTLs could also be detected in three cell-type-specific datasets (283 monocyte samples[9], 282 B-cell samples[9] and 608 HapMap lymphoblastoid cell-line (LCL) samples[24]; Supplementary Figures 9 and 10). Despite the different tissue of these three studies, we were still able to significantly replicate 7%, 4% and 2% of the *trans*-eQTLs (FDR < 0.05), respectively. As 95% of the *trans*-eQTL SNPs explained less than 3% of the total expression variance

(Supplementary Figure 11), we lack statistical power to replicate most *trans*-eQTLs in these smaller replication cohorts.

We subsequently confined further analyses to 2,082 different SNPs that have been found associated with complex traits at genome-wide significant levels ('trait-associated SNPs', reported $P < 5 \times 10^{-8}$, out of 4,542 unique SNPs that we tested). These 2,082 SNPs showed a significantly higher number of *trans*eQTL effects as compared to the 2,460 tested SNPs with reported disease associations at lower significance levels ($P = 8 \times 10^{-22}$, Supplementary methods and results, Supplementary Figure 12): 254 of these 2,082 SNPs show a *trans*-eQTL effect in the discovery analysis (reflecting 1,340 SNP-probe combinations, of which we significantly replicated 1,201 SNP-probe combinations, reflecting 233 different SNPs and 103 independent loci in blood). For 671 out of these 1,340 *trans*-eQTLs (50%) the trait-associated SNP was either the strongest *trans*-eQTL SNP within the locus (or in strong LD with the strongest *trans*-eQTL SNP) or unlinked to the strongest *trans*-eQTL SNP (Supplementary results and Supplementary Table 6). We observed that the 2,082 trait-associated SNPs were six times more likely to cause *trans*-eQTL effects than randomly selected SNPs (matched for distance to gene and allele frequency, $P = 5.6 \times 10^{-49}$, Supplementary methods and results, Supplementary Figure 13). SNPs, associated with (auto)immune or hematological traits were twice as likely to cause *trans*-eQTLs, as compared to other trait-associated SNPs ($P = 5 \times 10^{-25}$, Supplementary methods and results). We observed that trait-associated SNPs that also cause *trans*-eQTLs more often affect the expression levels of nearby transcription factors in *cis*, as compared to trait-associated SNPs that do not affect genes in *trans* (Fisher's exact P = 0.032; Supplementary results), suggesting that some of the *trans*-eQTLs arise due to altered *cis* gene expression levels of nearby transcription factors.

We also examined genomic SNP properties of the *trans*-eQTLs: these SNPs (and their perfect proxies based on data from the 1000 Genomes Project[25,26]) are significantly enriched (Fisher's exact $P < 0.05$) for mapping within miRNA binding sites (Figure 1A). They map to regions showing strong enrichment (fold-change > 2.5) of histone enhancer signals in K562 (myeloid) and GM12878 (lymphoid) cell-lines (Figure 1B), when compared to six non-blood cell-lines. This myeloid and lymphoid enhancer enrichment supports the validity of our blood-derived *trans*-eQTLs. These enrichment results suggest tissue specificity, which is supported by our inability to replicate a strong *trans*-eQTL that was previously identified in adipose tissue for SNP rs4731702[13] that is associated with both type 2 diabetes and lipid levels.

These *trans*-eQTLs can provide insight into the pathogenesis of disease. Although RNA microarray studies have revealed dysregulated pathways for many complex diseases, it is often unclear what comes first: whether the associated SNPs first cause defects in the pathways whose dysregulation ultimately leads to disease, or whether the SNPs first cause disease that then perturbs these pathways. One example is SLE, an auto-immune disease resulting in inflammation and tissue damage. It is known that SLE patients show markedly increased type 1 interferon (IFN-α) levels, increased expression of IFN-α response genes[4,27,28] and decreased complement *C1q* expression. We observed that four common SLE associated variants do indeed affect IFN-α response genes in *cis* (*IRF5, IRF7, TAP2*

and *PSMB9*;Supplementary Table 1). However, as most SLE-associated SNPs do not map near complement or IFN-α response genes, we assessed whether these SNPs might affect complement or IFN-α response genes in *trans*. This was the case for rs4917014, for which the SLE risk allele (rs4917014*T, showing genome-wide significance in Asian populations and nominal significance in European populations[1,24]) not only increased expression of five different IFN-α response genes (*HERC5, IFI6, IFIT1, MX1* and *TNFRSF21*; Figure 2), but also decreased expression of three different probes in *CLEC10A*. In addition, we observed a nominal significant association of rs4917014*T with decreased expression of C1QB (P = 5.2 x 10^{-6}, FDR 0.28), a subunit of the first component of complement C1q, which has an established protective role in lupus. The complete deletion of C1q practically assures the development of SLE[29,30]. *CLEC10A* and *CLEC4C* belong to the C-type lectin family, which also includes mannose-binding lectins (MBL). While, to our knowledge, *CLEC10A* and *CLEC4C* have not been studied in the context of SLE, the role of MBL is similar to C1q and is a risk factor for the development of autoimmunity in both humans and mice[3]. The rs4917014 *trans*-eQTLs were well replicated in the peripheral blood and monocyte replication datasets and reinforce the role of altered IFN-α mediated pathway, C-type lectin and *C1q* gene expression in SLE. In addition, people who do not have SLE, but who carry the rs497014*T risk allele already show these pathway alterations, which indicates these affected pathways are not solely a consequence of SLE, but could well precede SLE onset.

We next investigated the underlying mechanisms of the effects exerted by rs4917014. *IKZF1* is the only gene residing within the rs4917014 locus. Being a transcription factor (Ikaros family zinc finger 1), *cis*-regulatory effects of rs4917014 on *IKZF1*, that would translate in altered IKZF1 protein levels, could provide a working mechanism for the detected *trans*-eQTLs. However, since our meta-analysis initially did not show a *cis*-eQTL on the Illumina probe for *IKZF1* that is located near the 5' untranslated region (UTR) of *IKZF1*, we investigated the 3'-UTR by using DeepSAGE next-generation RNA-sequencing data of 94 peripheral blood samples. The variant rs4917014*T strongly increased the 3'-UTR expression levels of *IKZF1* (Spearman correlation = 0.45, P = 6.29 x 10^{-6}, Zhernakova et al, submitted). We then used ChIP-seq data from the ENCODE-project[31] and observed significantly increased IKZF1 protein binding to the genomic DNA locations where the upregulated *trans*-eQTL genes map (Wilcoxon P-value = 0.046), compared to IKZF1 binding to all other genic DNA. We also observed increased IKZF1 binding to the other SLE *cis*-genes outside of the *IKZF1* locus (Wilcoxon P-value = 4.3×10^{-4}), thereby confirming the importance of *IKZF1* in SLE. *IKZF1* is important for other phenotypes as well: another, unlinked intronic variant within *IKZF1*, rs12718597, is associated with mean corpuscular volume (MCV) [32] and affects the 5' end of *IKZF1* in *cis*. As *IKZF1* knock-out mice show abnormal erythropoiesis[33], this suggests a causal role for *IKZF1* in MCV as well. However, although rs12718597*A increases expression of 31 *trans*-genes and decreases expression of another 19 *trans*-genes, none of the SLE *trans*-genes overlap the MCV *trans*-genes. The latter are mainly involved in hemoglobin metabolism and do not show an increased IKZF1-binding signal, Wilcoxon P = 0.35. In summary, these results indicate that *IKZF1* has multiple functions and that different SNPs near *IKZF1* elicit function-specific effects.

We identified other *trans*-eQTLs showing similar phenomena: we observed that rs174546 (located in the 3'-UTR of *FADS1*, and associated with metabolic syndrome[34], LDL and total cholesterol levels[35,36]) affects *C11orf10, FADS1* and *FADS2* in *cis* and *LDLR* in *trans*. *LDLR* encodes the LDL receptor and contains common variants that are also associated with lipid levels[36] (Figure 3). *LDLR* gene expression levelscorrelated negatively ($P < 3.0 \times 10^{-4}$) with total, HDL and LDL cholesterol levels in the tested cohorts (Rotterdam Study and EGCUT, Supplementary Table 7), indicating that peripheral blood is a useful tissue for gaining downstream insight into the effects of lipid SNPs.

For 21 different complex traits, we found that at least two unlinked variants that are associated with these diseases, affected exactly the same gene in *trans*. When taking an equally sized, but permuted list of *trans*-eQTLs we would on average find only one complex trait where two unlinked SNPs affected the same gene in *trans* (Figure 4, Supplementary Table 8, Online methods). Although most of these traits are hematological (e.g. mean platelet volume or serum iron levels) we also observed this convergence for blood pressure, celiac disease, multiple sclerosis, and type 1 diabetes (T1D). rs3184504 (located in an exon of *SH2B3*) and its near-perfect proxy rs653178 (located in an intronic region of *ATXN2* on chromosome 12) are associated with several auto-immune diseases including T1D[37,38], T1D auto-antibodies[37,38], celiac disease[8,39], hyperthyroidism[40], vitiligo[41], rheumatoid arthritis[39] and other complex traits such as blood pressure[42,43], chronic kidney disease[44], and eosinophil counts[45].

We observed a *cis*-eQTL on *SH2B3* (FDR < 0.05) and fourteen *trans*-eQTL genes (FDR < 0.05, Figure 5), all highly expressed in neutrophils. Since these *trans*-eQTLs could potentially appear due to the known effect of rs3184504 on differences in cell-count proportions[45], we correlated *trans*-gene expression levels with cell counts in two cohorts (the Rotterdam Study and EGCUT) but did not observe significant correlations (Supplementary Table 6). These fourteen *trans*-eQTLs describe different biological functions: T1D disease risk allele rs3184504*T decreases expression levels of nine genes, most of which are involved in toll-like receptor signaling[46] (C12orf75, FOS, IDS, IL8, LOC338758, NALP12, PPP1R15A, S100A10 and *TAGAP*) and increases expression of five genes involved in interferon-γ response (*GBP2, GBP4, STAT1, UBE2L6* and *UPP1)*. We observed that another T1D risk allele, rs4788084*C[37,38] on chromosome 16, increases expression of *GBP4* and *STAT1* as well (Figure 5), revealing how different T1D risk alleles converge: they both cause an increase of interferon-γ response gene expression.

In summary, our eQTL meta-analysis revealed and replicated downstream effects for 233 trait-associated SNPs. We have highlighted only a few here and shown that *trans*-eQTL mapping in blood for lipid and immune-mediated disease variants yields downstream insight which is biologically meaningful. Our results on *IKZF1* show that the two unlinked SLE and MCV variants near this gene give strikingly different yet biologically meaningful *trans*-regulatory effects. Future, larger-scale *trans*-eQTL analysis in blood will likely uncover many more of these regulatory relationships.

Online methods

Study populations

We performed a whole-genome eQTL meta-analysis of 5,311 samples from peripheral blood, divided over a total of nine datasets from seven cohorts, including EGCUT[14] (N = 891), InCHIANTI[15] (N = 611), Rotterdam Study[16] (N = 762), Fehrmann[5] (N = 1,240 on the Illumina HT12v3 platform and N = 229 on the Illumina H8v2 platform), HVH[17-19] (N = 43 on the Illumina HT12v3 platform and N = 63 on the Illumina HT12v4 platform) SHIP-TREND[20] (N = 963), and DILGOM[21] (N = 509). Gene expression data for each dataset was obtained using either PAXGene (Becton Dickinson) or Tempus tubes (Life Technologies), followed by hybridization to Illumina whole-genome Expression BeadChips (HT12v3, HT12v4 or H8v2 arrays). The gene expression platforms were harmonized by matching probe sequences across the different platforms. Mappings for these sequences were obtained by mapping the sequences against the human genome build 36 (Ensembl build 54, Hg18) using BLAT, BWA and SOAPv2 sequence alignment programs. Highly stringent alignment criteria were used to ensure that probes map unequivocally to one single genomic position. Genotype data was acquired using different genotyping platforms, and harmonized by imputation, using the HapMap2[47] Central European population as a reference. Each dataset was individually checked for sample mix-ups using *MixupMapper*[48]. For a full description of the individual datasets, results of the sample mix-up analysis, specifics on the gene expression platforms and probe mapping procedure and filtering, see Supplementary methods.

Gene expression normalization

Gene expression data was quantile-normalized to the median distribution, and subsequently $\log_2$ transformed. The probe and sample means were centered to zero. Gene expression data was then corrected for possible population structure by removal of four multi-dimensional scaling components using linear regression. We reasoned earlier that normalized gene-expression data still contains large amounts of non-genetic variation[5]. After population stratification correction, principal component analysis (PCA) was therefore performed on the sample correlation matrix. We performed a separate QTL analysis for each principal component (PC), to ascertain whether genetic variants could be detected that affect the PC. If we found an effect on the PC, we did not correct the expression data for these components, to ensure we would not unintentionally remove genetic effects from the expression data. Significance of these associations was established by controlling the false discovery rate (FDR), testing each association against a null-distribution created by repeating the analysis 100 times (permuting the sample labels for each iteration[49]). PCs that did not show significance at the FDR threshold of 0.0 were removed from the gene expression data by linear regression. In all but two very small datasets, the first 40 PCs were removed (excluding those components per cohort that showed a QTL effect). We observed that the removal of these 40 components revealed the highest number of eQTLs in each dataset. Although PC correction may remove some eQTL effects, we observed that the majority (95% when removing 35 PCs and 90% when removing 40 PCs) of *trans*-eQTL effects was independent of the number of PCs removed (Supplementary Figure 14).

eQTL mapping

After normalization of the data, we performed both *cis*- and *trans*-eQTL mapping. eQTLs were deemed *cis*-eQTLs, when the distance between the SNP chromosomal position and the probe midpoint was less than 250 kilobases (kb), while eQTLs with a distance greater than five megabases (mb) were defined as *trans*-eQTLs. Only SNPs with a minor allele-frequency (MAF) > 0.05 and a Hardy-Weinberg equilibrium p-value > 0.001 were included in the analyses. Since most cohorts had generated the gene expression data using the HT12v3 platform, we chose to only include probes that were present on this platform. We only tested SNP-probe pairs when the SNP passed quality control in at least three cohorts. Furthermore, in order to reduce issues with respect to computational time and multiple testing, we confined our *trans*eQTL analysis to those SNPs present in the "*Catalog of Published GWAS*" (http://www.genome.gov/gwastudies/, accessed July 16th, 2011). We reasoned that for genes with strong *cis*-eQTL effects, the *cis*-eQTL effect may obscure the detectability of *trans*-eQTL. Therefore, we used linear regression to remove *cis*-eQTL effects prior to *trans*-eQTL mapping and observed a 12% increase in the number of detected *trans*-eQTLs (Supplementary Figure 15). For each cohort, eQTLs were mapped using a Spearman's rank correlation on the imputed genotype dosage values. We used a weighted Z-method for subsequent meta-analysis[50]. To get a realistic null-distribution, we permuted the sample identifiers labels of the expression data and repeated this analysis ten times (Supplementary Figure 16). In each permutation the sample labels were permuted. We then corrected for multiple testing by controlling the FDR at 0.05, by testing each p-value in the real data against a null-distribution created from the permuted datasets[49] (see Supplementary methods). It has been suggested that false-positive eQTL effects can arise due to polymorphisms in the probe sequences[51,52]. Therefore, we tested whether a significant *cis*-eQTL SNP was in LD ($r^2 > 0.2$) with any SNP in the *cis*-probe sequence, using the Western European subpopulations of the 1000 genomes project[25] (2011-05-21 release, 286 individuals, excluding Finnish individuals) as a reference. If we observed this to be the case the respective *cis*-eQTLs were removed. Furthermore, for each *trans*-eQTL we investigated whether portions of the probe sequence could map in the vicinity of the *trans*-eQTL SNP (which in fact would imply a *cis*-eQTL, rather than a *trans*-eQTL effect). Therefore, we tried to map the *trans*-eQTL probe sequences, using very permissive settings, within a 5 Mb window of the *trans*-eQTL SNP. SNP-probe combinations where the probe mapped with at least 15 bp within the 5 Mb window, were deemed false-positive and removed from further analysis. After this filtering we recalculated the FDR for both the *cis*- and *trans*-eQTL results.

Trans-eQTL replication

Replication of the *trans*-eQTL results was carried out in five independent datasets from four cohorts, including data obtained from lymphoblastoid cell lines (HapMap3, N = 604[24]), B-cells and monocytes (Oxford[9], N = 282 and N = 283, respectively), and whole peripheral blood (KORA F4[22], N = 740, and BSGS[23], N = 862). All the cohorts applied the same methodology as used in the discovery phase to normalize the gene expression data, check for sample mix-ups and perform *trans*-eQTL mapping, including 10 permutations in order to establish the FDR threshold at 0.05. Finally, we performed a sample-size weighted Z-score

meta-analysis on the two peripheral blood replication cohorts (KORA and BSGS). Further details on these datasets can be found in the Supplementary methods.

Enhancer enrichment and functional annotation

To determine whether the significant *trans*-eQTL SNPs were enriched for functional regions on the genome, we annotated the *trans*-eQTL SNPs using SNPInfo[53], SNPNexus[54,55], and HaploReg[56], which integrate multiple data sources (such as ENCODE project data[31], Ensembl[57], and several micro-RNA databases). We limited these analyses to those *trans*-eQTL SNPs that were previously shown to be associated with complex traits at genome-wide significance levels ('trait associated SNPs', reported $P < 5 \times 10^{-8}$). These SNPs were subsequently pruned (using PLINK's --clump command, using an $r^2 < 0.2$). We used the permuted *trans*-eQTL data to get realistic null-distributions for each of these tools: we selected equally sized sets of unlinked SNPs ($r^2 < 0.2$ in the Western-European subpopulations of the 1000 genomes project[25], 2011-05-21 release, 286 individuals, excluding Finnish individuals) that showed the highest significance in the permuted data, ensuring that only trait-associated SNPs are included in the null-distribution, as it is known that trait-associated SNPs in general already have different functional properties than randomly selected SNPs[58] (e.g. trait-associated SNPs typically map in closer proximity to genes than random SNPs). We also ensured that none of the SNPs in the null-distribution were affecting genes in *trans*, or were linked to those SNPs ($r^2 < 0.2$ in 1000 genomes). We then identified perfect proxies ($r^2 = 1.0$ in 1000 genomes). For SNPInfo and SNPNexus, we calculated the enrichment for each functional category using a Fisher's exact test. We determined the enhancer enrichment in nine different cell-types using HaploReg, where we averaged the enhancer enrichment over the ten permutations.

Convergence analysis

We determined which unlinked trait-associated SNPs show eQTL effects on exactly the same gene: per trait, we analyzed the SNPs that are known to be associated with this trait and assessed whether any unlinked SNP pair ($r^2 < 0.2$, distance between SNPs > 5Mb) showed a *cis*- and/or *trans*-eQTL effect on exactly the same gene, as previously described[5]. To determine whether the number of traits for which we observed this phenomenon was higher than expected by chance, we re-ran this analysis 20 times, each time using a different set of permuted *trans*-eQTLs, equal in size to the non-permuted set of *trans*-eQTLs.

SLE IKZF1 ENCODE ChIP-seq Analysis

We used IKZF1 ChIP-seq signal data obtained from the ENCODE-project[31] (IKZF1 ChIP-seq data acquired and processed by UCSC, ENCODE March 2012 Freeze). For every human gene we determined the average signal (corrected for gene size), corrected for GC-content bias, and performed a Wilcoxon Mann-Whitney test to see whether the upregulated genes (*MX1, TNFRSF21, IFIT1/LIPA, HERC5, CLEC4C, IFI6*) showed a higher ChIP-seq signal compared to all other human genes.

Data availability

We have made a browser available for all significant *trans*-eQTL and *cis*-eQTL at http://www.genenetwork.nl/bloodeqtlbrowser. This browser also provides all *trans*-eQTLs that we detected at a somewhat less stringent false discovery rate of 0.5, to enable more in-depth *post-hoc* analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Harm-Jan Westra[#1], Marjolein J. Peters[#2,3], Tõnu Esko[#4], Hanieh Yaghootkar[#5], Claudia Schurmann[#6], Johannes Kettunen[#7,8], Mark W. Christiansen[#9], Benjamin P. Fairfax[10,11], Katharina Schramm[12,13], Joseph E. Powell[14,15], Alexandra Zhernakova[1], Daria V Zhernakova[1], Jan H. Veldink[16], Leonard H. Van den Berg[16], Juha Karjalainen[1], Sebo Withoff[1], André G. Uitterlinden[2,3,17], Albert Hofman[3,17], Fernando Rivadeneira[2,3,17], Peter A C 't Hoen[18], Eva Reinmaa[4], Krista Fischer[4], Mari Nelis[4], Lili Milani[4], David Melzer[19], Luigi Ferrucci[20], Andrew B. Singleton[21], Dena G. Hernandez[21,22], Michael A. Nalls[21], Georg Homuth[6], Matthias Nauck[23], Dörte Radke[24], Uwe Völker[6], Markus Perola[4,8], Veikko Salomaa[8], Jennifer Brody[9], Astrid Suchy-Dicey[25], Sina A. Gharib[26], Daniel A. Enquobahrie[25], Thomas Lumley[27], Grant W. Montgomery[28], Seiko Makino[10], Holger Prokisch[12,13], Christian Herder[29], Michael Roden[29,30], Harald Grallert[31], Thomas Meitinger[12,13,32], Konstantin Strauch[33,34], Yang Li[35], Ritsert C. Jansen[35], Peter M. Visscher[14,15], Julian C. Knight[10], Bruce M. Psaty[#9,36], Samuli Ripatti[#7,8,37], Alexander Teumer[#6], Timothy M. Frayling[#5], Andres Metspalu[#4], Joyce B.J. van Meurs[#2,3], and Lude Franke[#1]

Affiliations

[1] Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen, the Netherlands [2] Department of Internal Medicine, Erasmus Medical Centre Rotterdam, the Netherlands [3] The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden/Rotterdam, the Netherlands [4] Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia [5] Genetics of Complex Traits, University of Exeter Medical School, Exeter, EX1 2LU, UK [6] Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, D-17487 Greifswald, Germany [7] Institute for Molecular Medicine Finland FIMM, FI-00014 University of Helsinki, Helsinki, Finland [8] Department of Chronic Disease Prevention, National Institute for Health and Welfare, FI-00271 Helsinki, Finland. [9] Cardiovascular Health Research Unit, University of Washington, Seattle, WA [10] Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK [11] Department of Oncology, Cancer and Haematology Centre, Churchill Hospital, Oxford, OX3 7LJ [12] Institute of Human Genetics, Helmholz Zentrum München - German Research Center for Environmental Health, Neuherberg,

Germany [13] Institute of Human Genetics, Technical University Munich, Munich, Germany [14] University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia [15] The Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia [16] Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, the Netherlands [17] Department of Epidemiology, Erasmus Medical Center Rotterdam, the Netherlands [18] Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands [19] Institute of Biomedical and Clinical Sciences, University of Exeter Medical School, Barrack Road, Exeter, EX2 5DW, UK [20] Clinical Research Branch, National Institute on Aging NIA-ASTRA Unit, Harbor Hospital, MD, USA [21] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, 35 Lincoln Drive, Bethesda, MD, USA [22] Department of Molecular Neuroscience and Reta Lila Laboratories, Institute of Neurology, UCL, Queen Square House, Queen Square, London WC1N 3BG, UK. [23] Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany [24] Institute for Community Medicine, University Medicine Greifswald, D-17487 Greifswald, Germany [25] Department of Epidemiology, University of Washington, Seattle, WA, USA [26] Computational Medicine Core, Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, Department of Medicine, University of Washington, Seattle, WA, USA [27] Department of Statistics, University of Auckland, Auckland, New Zealand [28] Queensland Institute of Medical Research, Herston, Brisbane, Australia [29] Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, University Düsseldorf, Düsseldorf, Germany [30] Departments of Endocrinology & Diabetology & Metabolic Diseases, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. [31] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany [32] German Center for Cardiovascular Research (DZHK), Germany; Munich Heart Allience, Munich, Germany [33] Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Neuherberg, Germany [34] Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany [35] Groningen Bioinformatics Centre, University of Groningen, Groningen, the Netherlands. [36] Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA. [37] Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,CB10 1SA, Hinxton, UK

Acknowledgments

DILGOM

J.K. and S.R. were supported by funds from The European Community's Seventh Framework Programme (FP7/2007-2013) BioSHaRE, grant agreement 261433, S.R. was supported by funds from The European Community's Seventh Framework Programme (FP7/2007-2013) ENGAGE Consortium, grant agreement

HEALTH-F4-2007- 201413", the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 213506 and 129680), Academy of Finland (grant 251217), the Finnish foundation for Cardiovascular Research and the Sigrid Juselius Foundation. V.S. was supported by the Academy of Finland, grant number 139635 and Finnish Foundation for Cardiovascular Research. MP was partly financially supported for this work by the Finnish Academy SALVE program "Pubgensense" 129322 and by grants from the Finnish Foundation for Cardiovascular Research. The DILGOM-study was supported by the Academy of Finland, grant # 118065.

SHIP-TREND

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Deutsche Forschungsgemeinschaft (DFG GRK840-D2), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg, West Pomerania. Whole-body MR imaging was supported by a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH. The SHIP authors thank Mario Stanke for the opportunity to use his Server Cluster for the SNP imputation.

EGCUT

EGCUT received financing by FP7 grants (201413, 245536), also received targeted financing from the Estonian Government (SF0180142s08) and direct funding from the Ministries of Research and Science and Social Affairs. EGCUT studies are funded by the University of Tartu in the framework of the Center of Translational Genomics and by the European Union through the European Regional Development Fund, in the framework of the Centre of Excellence in Genomics. We thank EGCUT personnel, especially Ms. M. Hass and Mr V. Soo. EGCUT data analyses were carried out in part in the High Performance Computing Center of the University of Tartu.

HVH

HVH was supported in part by grants R01 HL085251 and R01 HL073410 from the National Heart, Lung, and Blood Institute (NHLBI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI or the National Institutes of Health, USA. Dr. Psaty serves on a DSMB for a clinical trial of a device funded by Zoll LifeCor and on the Steering Committee for the Yale Open Data Access Project funded by Medtronic. Both activities are unrelated to this work.

Rotterdam Study

We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, and Marjolein Peters for their help in creating the GWAS database; Karol Estrada and Maksim Struchalin for their support in creation and analysis of imputed data; Tobias A. Knoch, Anis Abuseiris, Karol Estrada, and Rob de Graaf as well as their institutions, the Erasmus GRID Office, Erasmus MC Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. The authors thank the study participants and staff from the Rotterdam Study, the participating general practitioners and the pharmacists.

The Rotterdam Study was funded by the European Commission (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008 35627, TREAT-OA 200800), the Netherlands Organization of Scientific Research NWO Investments (nos 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) (project nr. 050-060-810), an NWO VIDI grant (#917103521).

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

Fehrmann

L.F.,H-J.W.: This study was supported by grants from the Celiac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government (grant BSIK03009), the Netherlands Organization for Scientific Research (NWO-VICI grant 918.66.620, NWO-VENI grant 916.10.135

to L.F.), the Dutch Digestive Disease Foundation (MLDS WO11-30), and a Horizon Breakthrough grant from the Netherlands Genomics Initiative (grant 92519031 to L.F.). This project was supported by the Prinses Beatrix Fonds, VSB fonds, H. Kersten and M. Kersten (Kersten Foundation), The Netherlands ALS Foundation, and J.R. van Dijk and the Adessium Foundation. The research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013) under grant agreement 259867. We especially thank Cisca Wijmenga for helpful comments and support and Jackie Senior for critically reading the manuscript.

InCHIANTI

InCHIANTI was supported by the Wellcome Trust 083270/Z/07/Z. The InCHIANTI study was supported by contract funding from the U.S. National Institute on Aging (NIA), and the research was supported in part by the Intramural Research Program, NIA, and National Institute of Health (NIH). A.R.W. was supported by the Peninsula NIHR Clinical Research Facility. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

KORA F4

The KORA authors acknowledge the contributions of Peter Lichtner, Gertrud Eckstein, Guido Fischer, Norman Klopp, Nicole Spada, and all members of the Helmholtz Zentrum München genotyping staff for generating the SNP data and Katja Junghans and Anne Löschner (Helmholtz Zentrum München) for generating gene expression data from both KORA and SHIP-TREND samples.

The KORA research platform and the KORA Augsburg studies are financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, which is funded by the BMBF and by the State of Bavaria. We thank the field staff in Augsburg who were involved in the studies. The German Diabetes Center is funded by the German Federal Ministry of Health and the Ministry of School, Science and Research of the State of North-Rhine-Westphalia. The Diabetes Cohort Study was funded by a German Research Foundation project grant to W.R. (DFG; RA 459/2-1). This study was supported in part by a grant from the BMBF to the German Center for Diabetes Research (DZD e.V.), by the DZHK (Deutsches Zentrum für Herz-Kreislauf-Forschung – German Centre for Cardiovascular Research) and by the BMBF funded Systems Biology of Metabotypes grant (SysMBo#0315494A). Additional support was given by the BMBF (National Genome Research Network NGFNplus Atherogenomics, 01GS0834) and the Leibniz Association (WGL Pakt für Forschung und Innovation). We thank Maren Carstensen, Gabi Gornitzka and Astrid Hoffmann (German Diabetes Center) for excellent technical assistance.

Oxford cell-specific

This work was supported by the Wellcome Trust (Grants 074318 [J.C.K.], 088891 [B.P.F.], and 075491/Z/04 [core facilities Wellcome Trust Centre for Human Genetics]), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement no. 281824 (J.C.K.) and the NIHR Oxford Biomedical Research Centre.

BSGS

This research was supported by Australian National Health and Medical Research Council (NHMRC) grants 389892, 496667, 613601, 1010374 and 1046880 and National Institutes of Health (NIH) grant GM057091. G.W.M. and P.V. are supported by the NHMRC Fellowship Scheme. J.E.P. is supported by the Australian Research Council (ARC) fellowship scheme. The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

References

1. Han JW, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009; 41:1234–7. [PubMed: 19838193]
2. Bengtsson AA, et al. Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies. Lupus. 2000; 9:664–71. [PubMed: 11199920]
3. Bohlson SS, Fraser DA, Tenner AJ. Complement proteins C1q and MBL are pattern recognition molecules that signal immediate and long-term protective immune functions. Mol Immunol. 2007; 44:33–43. [PubMed: 16908067]
4. Ytterberg SR, Schnitzer TJ. Serum interferon levels in patients with systemic lupus erythematosus. Arthritis Rheum. 1982; 25:401–6. [PubMed: 6176248]

5. 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 Genet. 2011; 7:e1002197. [PubMed: 21829388]
6. Nicolae DL, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010; 6:e1000888. [PubMed: 20369019]
7. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464:768–72. [PubMed: 20220758]
8. Dubois PC, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
9. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012; 44:502–10. [PubMed: 22446964]
10. Innocenti F, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011; 7:e1002078. [PubMed: 21637794]
11. Grundberg E, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012; 44:1084–9. [PubMed: 22941192]
12. Heinig M, et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. 2010; 467:460–4. [PubMed: 20827270]
13. Small KS, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 2011; 43:561–4. [PubMed: 21572415]
14. Metspalu A. The Estonian Genome Project. Drug Development Research. 2004; 62:97–101.
15. Tanaka T, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. PLoS Genet. 2009; 5:e1000338. [PubMed: 19148276]
16. Hofman A, et al. The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol. 2011; 26:657–86. [PubMed: 21877163]
17. Heckbert SR, et al. Antihypertensive treatment with ACE inhibitors or beta-blockers and risk of incident atrial fibrillation in a general hypertensive population. Am J Hypertens. 2009; 22:538–44. [PubMed: 19265792]
18. Psaty BM, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. JAMA. 1995; 274:620–5. [PubMed: 7637142]
19. Smith NL, et al. Esterified estrogens and conjugated equine estrogens and the risk of venous thrombosis. JAMA. 2004; 292:1581–7. [PubMed: 15467060]
20. Teumer A, et al. Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. Am J Hum Genet. 2011; 88:664–73. [PubMed: 21565293]
21. Inouye M, et al. An immune response network associated with blood lipid levels. PLoS Genet. 2010; 6
22. Mehta D, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. Eur J Hum Genet. 2012
23. Powell JE, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PLoS One. 2012; 7:e35430. [PubMed: 22563384]
24. Wang C, et al. Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. Eur J Hum Genet. 2012
25. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]
26. Patterson K. 1000 genomes: a world of variation. Circ Res. 2011; 108:534–6. [PubMed: 21372290]
27. Baechler EC, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A. 2003; 100:2610–5. [PubMed: 12604793]
28. Bennett L, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med. 2003; 197:711–23. [PubMed: 12642603]
29. McAdam RA, Goundis D, Reid KB. A homozygous point mutation results in a stop codon in the C1q B-chain of a C1q-deficient individual. Immunogenetics. 1988; 27:259–64. [PubMed: 2894352]

30. Botto M, et al. Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies. Nat Genet. 1998; 19:56–9. [PubMed: 9590289]
31. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
32. Ganesh SK, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. 2009; 41:1191–8. [PubMed: 19862010]
33. Wang JH, et al. Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation. Immunity. 1996; 5:537–49. [PubMed: 8986714]
34. Zabaneh D, Balding DJ. A genome-wide association study of the metabolic syndrome in Indian Asian men. PLoS One. 2010; 5:e11961. [PubMed: 20694148]
35. Sabatti C, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 2009; 41:35–46. [PubMed: 19060910]
36. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–13. [PubMed: 20686565]
37. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–7. [PubMed: 19430480]
38. Plagnol V, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. PLoS Genet. 2011; 7:e1002216. [PubMed: 21829393]
39. Zhernakova A, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011; 7:e1002004. [PubMed: 21383967]
40. Eriksson N, et al. Novel associations for hypothyroidism include known autoimmune risk loci. PLoS One. 2012; 7:e34442. [PubMed: 22493691]
41. Jin Y, et al. Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat Genet. 2012; 44:676–80. [PubMed: 22561518]
42. Newton-Cheh C, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009; 41:666–76. [PubMed: 19430483]
43. Wain LV, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet. 2011; 43:1005–11. [PubMed: 21909110]
44. Kottgen A, et al. New loci associated with kidney function and chronic kidney disease. Nat Genet. 2010; 42:376–84. [PubMed: 20383146]
45. Gudbjartsson DF, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet. 2009; 41:342–7. [PubMed: 19198610]
46. Rotival M, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet. 2011; 7:e1002367. [PubMed: 22144904]
47. The International HapMap Project. Nature. 2003; 426:789–96. [PubMed: 14685227]
48. Westra HJ, et al. MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. Bioinformatics. 2011; 27:2104–2111. [PubMed: 21653519]
49. Breitling R, et al. Genetical genomics: spotlight on QTL hotspots. PLoS Genet. 2008; 4:e1000232. [PubMed: 18949031]
50. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol. 2005; 18:1368–73. [PubMed: 16135132]
51. Alberts R, et al. Sequence polymorphisms cause many false cis eQTLs. PLoS One. 2007; 2:e622. [PubMed: 17637838]
52. Benovoy D, Kwan T, Majewski J. Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. Nucleic Acids Res. 2008; 36:4417–23. [PubMed: 18596082]
53. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009; 37:W600–5. [PubMed: 19417063]
54. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012; 40:W65–70. [PubMed: 22544707]

55. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics. 2009; 25:655–61. [PubMed: 19098027]

56. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–4. [PubMed: 22064851]

57. Flicek P, et al. Ensembl 2012. Nucleic Acids Res. 2012; 40:D84–90. [PubMed: 22086963]

58. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–7. [PubMed: 19474294]

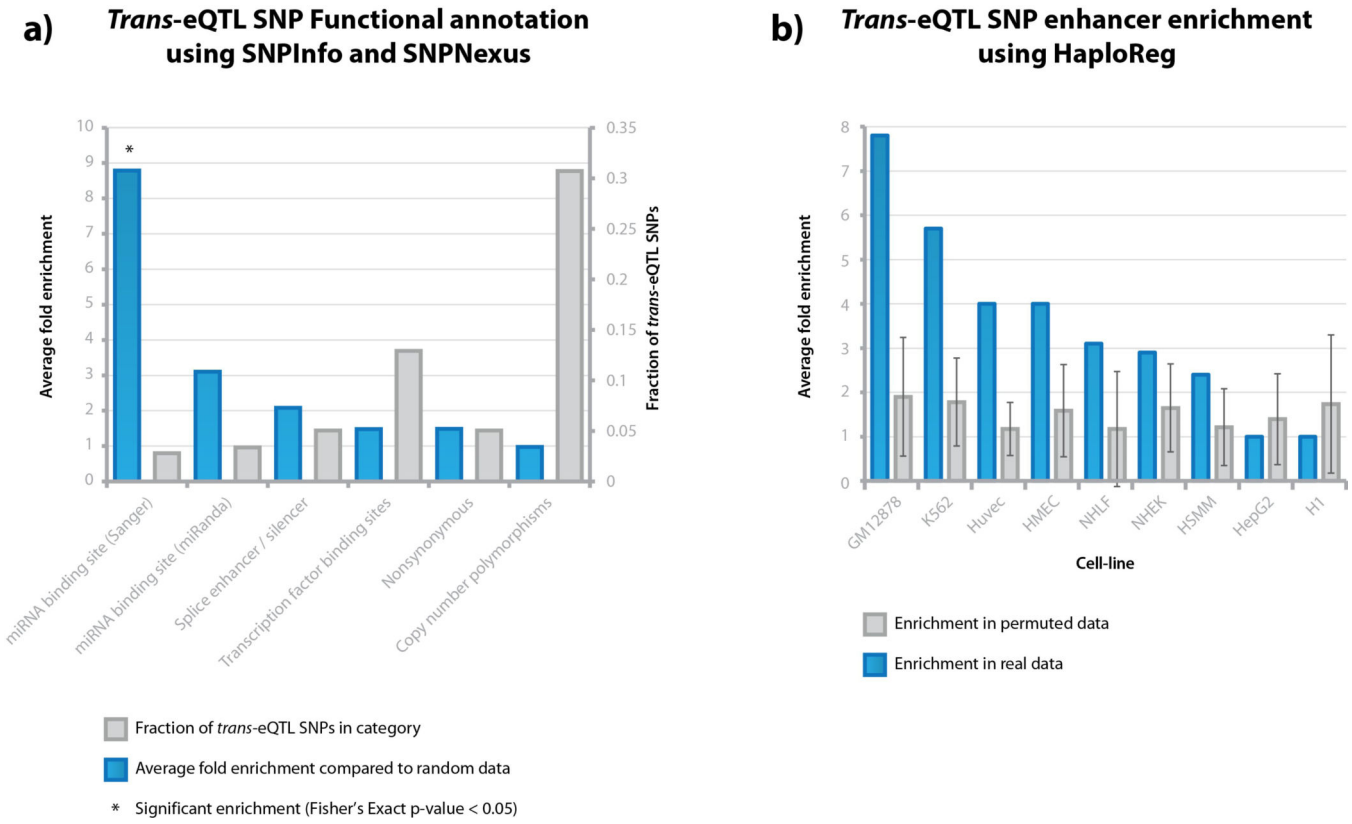


Figure 1. *Trans*-eQTL SNPs are enriched for functional elements

We investigated whether the *trans*-eQTL SNPs are enriched for certain functional elements. We used the online tools SNPInfo, SNPNexus, and HaploReg that rely upon data from, amongst others, the ENCODE project. (a) We observed that *trans*-eQTL SNPs are enriched for mapping within miRNA binding sites (b) *trans*-eQTL SNPs show strong enrichment (as annotated using HaploReg) for enhancer regions that are present in K562 (myeloid) and GM12878 (lymphoid) cell-lines (error bars represent one standard deviation).

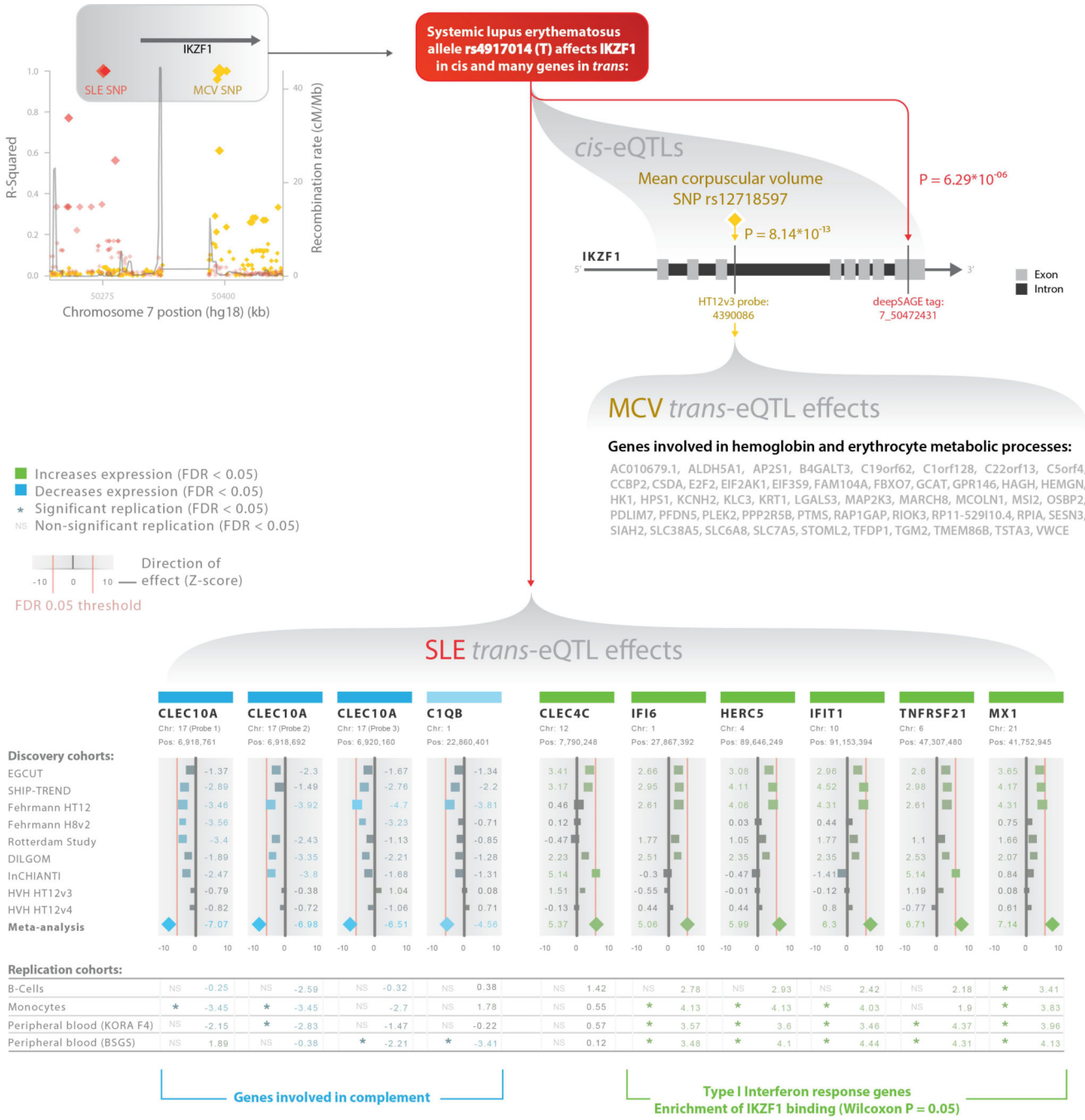


Figure 2. Independent *trans*-eQTL effects emanating from the *IKZF1* locus

Systemic lupus erythematosus SNP rs4917014 and unlinked mean corpuscular volume SNP rs4917014 both affect expression of *IKZF1* in *cis*. rs12718597 affects 50 *trans*-genes (mostly involved in hemoglobin metabolism) while rs4917014 affects eight different genes in *trans*: the rs4917014*T risk allele increases expression of genes involved in type I interferon response. At a somewhat lower significance threshold of FDR 0.28 rs4917014*T decreases complement *C1QB* expression. Both processes are hallmark features of SLE.

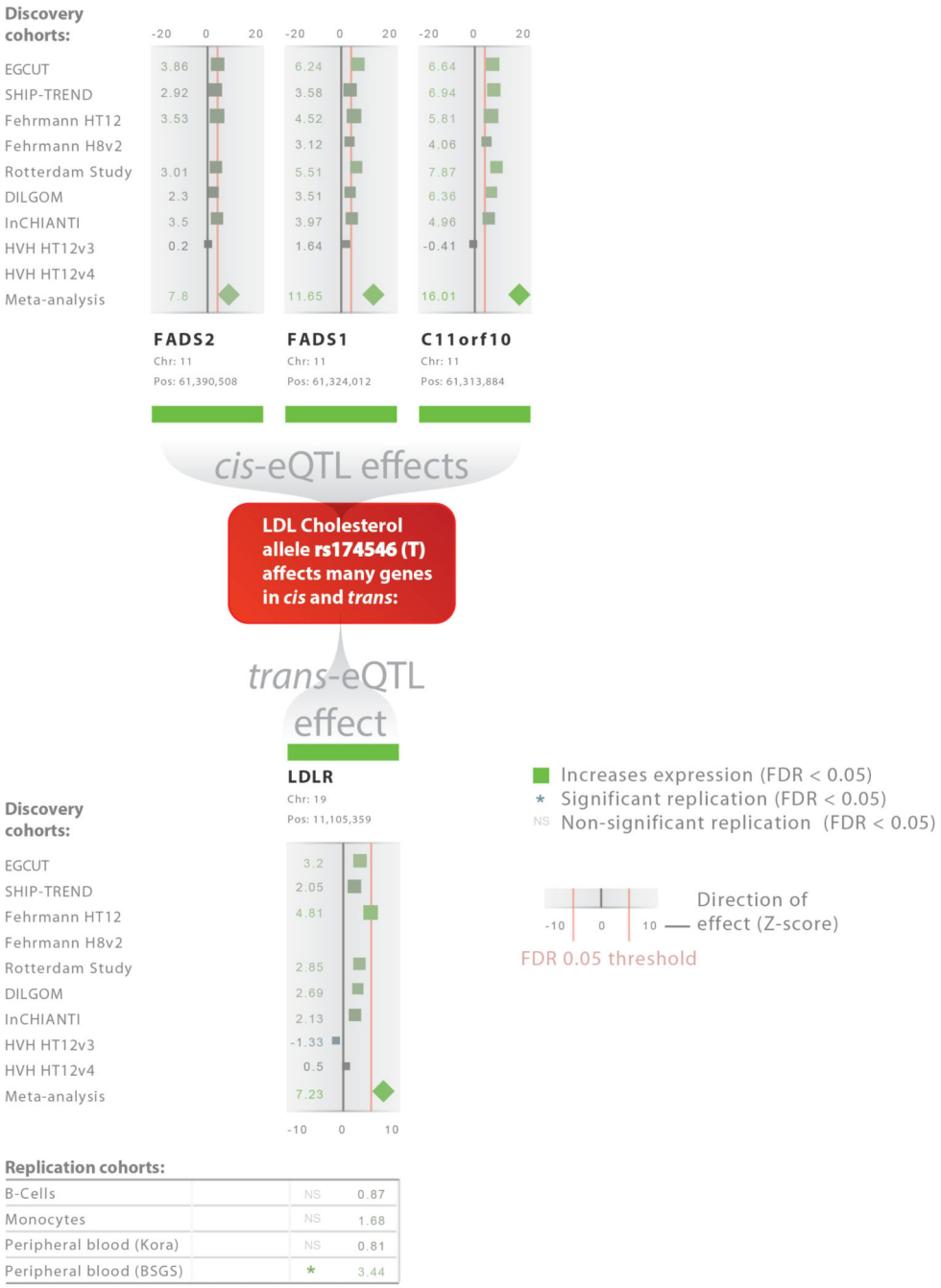
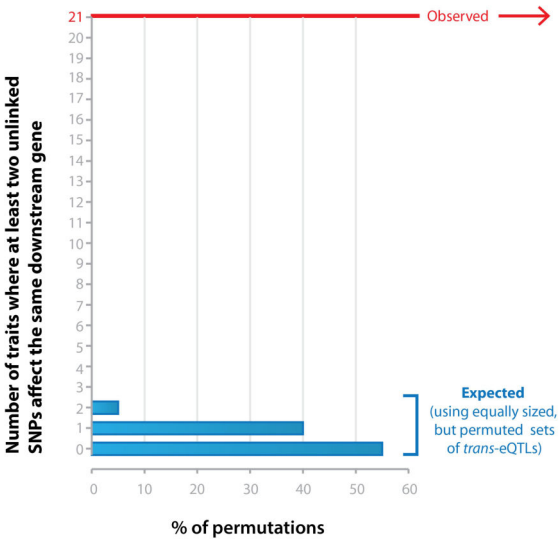


Figure 3. Cholesterol SNP rs174546 affects *LDLR* in *trans*

The rs174546*T allele is known to be associated with a decrease in serum LDL cholesterol and triglycerides levels. It increases the expression levels of three genes in *cis*, but also increases gene expression levels of *LDLR* that encodes the LDL receptor.

A)



B)

Trait type	Complex trait	At least two unlinked trait SNPs both affect gene(s):
Immune related traits	Type 1 diabetes	*GBP4, STAT1*
	Type 1 diabetes autoantibodies	*GBP4, STAT1*
	Celiac disease	*CXCR6, FYCO1*
	Multiple sclerosis	*CD5*
Blood pressure traits	Diastolic blood pressure	*LOC338758*
	Systolic blood pressure	*LOC338758*
Hematological traits	Hemoglobin	*ALAS2*
	Hematological Parameters	*FBXO7*
	F-cell distribution	*ESPN, PHOSPHO1, GNAS, TSPAN13, VWCE,*
	Hematocrit	*ALAS2*
	Serum markers of iron status	*ALAS2*
	Red blood cell traits	*ALAS2*
	Serum iron levels	*ALAS2*
	Glycated hemoglobin levels	*ALAS2*
	Hematology traits	*ALAS2*
	Serum hepcidin	*ALAS2*
	Beta-thalassemia	*PHOSPHO1, VWCE, TSPAN13, ESPN*
	Hematological and biochemical traits	*AL109955.37-3, RBM38, TRIM58*
	Mean corpuscular hemoglobin	*ALAS2, C18orf10, DNAJB2, ESPN, HBM, KEL, PDZK1IP1, PIM1, PRDX5 RAP1GAP, UBXN6, VWCE, XK*
	Mean corpuscular volume	*ALAS2, B4GALT3, C18orf10, C1orf128, C22orf13, C5orf4, CCBP2, CSDA, DNAJB2, EIF2AK1, ESPN, FBXO7, HAGH, HBM, HPS1, KEL, KLC3, KRT1, LGALS3, MARCH8, MCOLN1, OSBP2, PDZK1IP1, PHOSPHO1, PIM1, PLEK2, PPP2R5B, PRDX5, PTMS, RAP1GAP, RIOK3, TGM2, TSTA3, UBXN6, VWCE, XK*
	Mean platelet volume	*ABCC3, AL353716.18, AQP10, C19orf33, C6orf152, CABP5, CTDSPL, CTTN, CXCL5 ESAM, F13A1, GNB5, GNG11, GP9, GUCY1A3, ITGA2B, ITGB5, LIMS1, LY6G6F, MMRN1, MPL, NRGN, PARVB, PRDX6, PTCRA, RAB27B, RBPMS2, SAMD14, SH3BGRL2, TSPAN9, VCL*

Figure 4. For 21 complex traits, pairs of unlinked trait-associated SNPs affect the same downstream genes

We observed that for 21 different traits, there were pairs of unlinked SNPs that have previously been reported to be associated with these traits and which also affect exactly the same downstream genes in *trans*, whereas this is rarely observed when using an equally sized, but permuted list of *trans*-eQTLs.

Figure 5. Two unlinked type-1 diabetes risk alleles both increase *STAT1* and *GBP4* expression

rs3184504*T, a risk allele for type 1 diabetes (chromosome 12), affects the expression of *SH2B3* in *cis*, but also affects the expression levels of fourteen unique genes in *trans*, including two interferon-γ response genes *GBP4* and *STAT1*. Another unlinked type-1 diabetes risk allele (rs4788084*C on chromosome 16) also increases expression levels of these two interferon-γ response genes, indicating that an elevated interferon-γ response is important in type 1 diabetes.

Table 1

Results of the *cis*- and *trans*-eQTL mapping analyses.

Summary statistics	***Cis*-eQTLs**	***Trans*-eQTLs**
Number of SNPs tested that pass QC	1,962,237	4,542 (of which 2,082 are associated with complex traits at genome-wide significance, $P < 5 \times 10^{-8}$)
Number of probes tested that pass QC	29,891	34,061
Number of genes tested	14,542	16,332
Number of probes not mapping to genes	9,260	18,018
Number of statistical tests performed	11,172,453	153,134,630

Significance thresholds	***Cis*-eQTLs**		***Trans*-eQTLs**	
	Meta-analysis z-score	**Meta-analysis p-value**	**Meta-analysis z-score**	**Meta-analysis p-value**
FDR < 0.05 significance	3.824	1.31×10^{-4}	5.022	5.12×10^{-7}
Bonferroni significance	5.867	4.5×10^{-9}	6.287	3.3×10^{-10}

***cis*-eQTL analysis**	**FDR < 0.05 significance**	**Bonferroni significance**
Number of significant unique SNP-Probe pairs	**664,097**	**395,543**
Number of significant unique eQTL SNPs	397,310	266,036
Number of significant unique eQTL probes	8,228	5,738
Number of significant unique eQTL genes	6,418	4,690
Number of significant unique eQTL probes not mapping to genes	636	326

***trans*-eQTL analysis**	**FDR < 0.05 significance**	**Bonferroni significance**
Number of significant unique SNP-Probe pairs	**1,513**	**643**
Number of significant unique eQTL SNPs	346	200
Number of significant unique eQTL probes	494	240
Number of significant unique eQTL genes	430	223
Number of significant unique eQTL probes not mapping to genes	35	13