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

Harm-Jan Westra1,40, Marjolein J Peters2,3,40, Tõnu Esko4,40, Hanieh Yaghoobkar5,40, Claudia Schurmann6,40, Johannes Kettunen7,8,40, Mark W Christiansen9,40, Benjamin P Fairfax10,11, Katharina Schramm12,13, Joseph E Powell14,15, Alexandra Zhernakova1, Daria V Zhernakova1, Jan H Veldink16, Leonard H Van den Berg16, Juha Karjalainen1, Sebo Withoff1, André G Uitterlinden2,3,17, Albert Hofman3,17, Fernando Rivadeneira2,3,17, Peter A C’t Hoen18, Eva Reimmaa4, Krista Fischer4, Mari Nelis4, Lili Milani4, David Melzer19, Luigi Ferrucci20, Andrew B Singleton21, Dena G Hernandez21,22, Michael A Nalls21, Georg Homuth6, Matthias Nauck23, Dörte Radke24, Uwe Völker6, Markus Perola4,8, Veikko Salomaa8, Jennifer Brody9, Astrid Suchy-Dicey25, Sina A Gharib26, Daniel A Enquobahrie25, Thomas Lumley27, Grant W Montgomery28, Seiko Makino10, Holger Prokisch12,13, Christian Herder29, Michael Roden29–31, Harald Grallert32, Thomas Meitinger12,13,33,34, Konstantin Strauch35,36, Yang Li37, Ritsert C Jansen37, Peter M Visscher14,15, Julian C Knight10, Bruce M Psaty9,38,41, Samuli Ripatti7,8,41, Alexander Teumer6,41, Timothy M Frayling5,41, Andres Metspalu4,41, Joyce B J van Meurs23,41 & Lude Franke1,41

Identifying the downstream effects of disease-associated SNPs is challenging. To help overcome this problem, we performed expression quantitative trait locus (eQTL) meta-analysis in non-transformed peripheral blood samples from 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. Some of these SNPs affect multiple genes in trans that are known to be altered in individuals with disease: rs4917014, previously associated with systemic lupus erythematosus (SLE)4, altered gene expression of C1QB and five type I interferon response genes, both hallmarks of SLE2–4. DeepSAGE RNA sequencing showed that rs4917014 strongly alters the 3’UTR levels of IKZF1 in cis, and chromatin immunoprecipitation and sequencing analysis of the trans-regulated genes implicated IKZF1 as the causal gene. 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 are noncoding, it is difficult to identify causal genes. Several eQTL-mapping studies5–8 have shown that disease-predisposing variants often affect the gene expression levels of nearby genes (cis eQTLs). A few recent studies have also identified trans eQTLs5,9–11, showing the downstream consequences of some variants. However, the total number of reported trans eQTLs is low, mainly owing to the multiple-testing burden. To improve statistical power, we performed an eQTL meta-analysis in 5,311 peripheral blood samples from 7 studies (EGCUT14, InCHIANTI15, Rotterdam Study16, Fehrman17–19, SHIP-TREND20 and DILGOM21) and replication analysis in another 2,775 samples. We aimed to ascertain to what extent SNPs affect genes in cis and to determine whether eQTL mapping in peripheral blood could identify downstream pathways that might be 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 a more stringent Bonferroni correction; Table 1, Supplementary Figs. 1–3 and Supplementary Tables 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; see URLs). In the discovery data set, we detected trans eQTLs for 1,513 significant trans eQTLs that included 346 unique SNPs (FDR <0.05; 8% of all tested SNPs; Table 1, Supplementary Fig. 4 and Supplementary Table 4) affecting the expression of 430 different genes. A recent study has also identified trans eQTLs for 233 SNPs that are known to be altered in individuals with disease: rs4917014, previously associated with systemic lupus erythematosus (SLE)4, altered gene expression of C1QB and five type I interferon response genes, both hallmarks of SLE2–4. DeepSAGE RNA sequencing showed that rs4917014 strongly alters the 3’UTR levels of IKZF1 in cis, and chromatin immunoprecipitation and sequencing analysis of the trans-regulated genes implicated IKZF1 as the causal gene. 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.

A full list of authors affiliation appears at the end of the paper.

Received 24 December 2012; accepted 14 August 2013; published online 8 September 2013; doi:10.1038/ng.2756
We did not find evidence that trans eQTLs were driven by differences in age or blood cell counts between individuals (Supplementary Fig. 6, Supplementary Table 6 and Supplementary Note). However, we cannot exclude this possibility entirely because FACS analyses on individual cell types had not been conducted. We also detected previously reported blood trans eQTLs in this study (Supplementary Fig. 7, Supplementary Table 7 and Supplementary Note).

To ensure reproducibility of the detected trans eQTLs, we replicated trans eQTLs from our discovery meta-analysis in 2 independent studies of peripheral blood gene expression: 52% in KORA F4 (n = 740 samples) and 79% in BSGS (n = 862 samples) (FDR < 0.05; Supplementary Fig. 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 compared with those detected in the discovery analysis. A meta-analysis of the two replication studies improved replication rates: 89% of the 1,513 trans eQTLs were significantly replicated (FDR < 0.05), with 99.7% showing a consistent allelic direction. Irrespective of significance, 97% of the trans eQTLs showed a consistent allelic direction in this replication meta-analysis (Supplementary Fig. 8). We found that some trans eQTLs could be detected in three cell type–specific data sets (283 monocyte samples, 282 B cell samples and 608 HapMap lymphoblastoid cell line (LCL) samples; Supplementary Figs. 9 and 10). Despite the different tissues analyzed in these three studies, we were 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 Fig. 11 and Supplementary Table 6), 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 to be associated with complex traits at genome-wide significance (trait-associated SNPs; reported P < 5 × 10−8), out of 4,542 unique SNPs that we tested. These 2,082 SNPs showed a significantly higher number of trans-eQTL effects compared with the 2,460 tested SNPs with reported disease associations at lower significance levels (P = 8 × 10−2; Supplementary Fig. 12 and Supplementary Note): 254 of these 2,082 SNPs showed a trans-eQTL effect in the discovery analysis (reflecting 1,340 SNP-probe combinations; 1,201 of these were significantly replicated in blood, reflecting 233 different SNPs and 103 independent loci). For 671 of these 1,340 trans eQTLs (50%), the trait-associated SNP (or a SNP in strong linkage disequilibrium, LD) was the strongest trans-eQTL SNP within the locus or was unlinked to the strongest trans-eQTL SNP (Supplementary Table 8 and Supplementary Note). The 2,082 trait-associated SNPs were 6 times more likely to cause trans-eQTL effects than were randomly selected SNPs (matched for distance to the gene and allele frequency; P = 5.6 × 10−49; Supplementary Fig. 13 and Supplementary Note). SNPs associated with (auto)immune or hematological traits were twice as likely to underlie trans eQTLs compared with other trait-associated SNPs (P = 5 × 10−27; Supplementary Note). Trait-associated SNPs that also caused trans eQTLs affected the expression levels of nearby transcription factors in cis more frequently than trait-associated SNPs that did not affect genes in trans (Fisher’s exact P = 0.032; Supplementary Note), suggesting that some trans eQTLs arise owing to altered cis gene expression levels of nearby transcription factors.

We examined the genomic properties of the trans-eQTL SNPs (and their perfect proxies identified using data from the 1000 Genomes Project)25,26; these SNPs were significantly enriched for mapping within microRNA (miRNA) binding sites (Fisher’s exact P < 0.05; Fig. 1a). They mapped to regions in K562 (myeloid) and GM12878 (lymphoid) cell lines showing enrichment of histone enhancer signals (fold change ≥2.5; Fig. 1b) compared to the signals observed in six non-blood cell lines. Enhancer enrichment in myeloid and lymphoid cells supports the validity of our blood-derived trans eQTLs. These results suggest that trans-eQTL effects are tissue specific, a notion that is supported by our inability to replicate a trans eQTL that was previously identified in adipose tissue13 for SNP rs4731702, associated with both type 2 diabetes (T2D) and lipid levels.

These trans eQTLs can provide insight into the pathogenesis of disease. Although RNA microarray studies have identified dysregulated pathways for many complex diseases, it is often unclear whether associated SNPs first cause defects in the pathways whose

---

**Table 1 Results of cis- and trans-eQTL mapping analyses**

<table>
<thead>
<tr>
<th></th>
<th>Cis-eQTL analysis</th>
<th>Trans-eQTL analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of significant unique SNP-probe pairs</td>
<td>FDR &lt;0.05</td>
<td>Bonferroni</td>
</tr>
<tr>
<td></td>
<td>significance</td>
<td>significance</td>
</tr>
<tr>
<td>Number of significant unique eQTL SNPs</td>
<td>664,097</td>
<td>395,543</td>
</tr>
<tr>
<td>Number of significant unique eQTL probes</td>
<td>397,310</td>
<td>266,036</td>
</tr>
<tr>
<td>Number of significant unique eQTL-regulated genes</td>
<td>8,228</td>
<td>5,738</td>
</tr>
<tr>
<td>Number of significant unique eQTL probes not mapping to genes</td>
<td>6,418</td>
<td>4,690</td>
</tr>
</tbody>
</table>

---

**Figure 1** Trans-eQTL SNPs are enriched for functional elements. We investigated whether trans-eQTL SNPs are enriched for certain functional elements using the online tools SNPInfo, SNP Nexus and HaploReg that rely on data from, among others, the ENCODE Project. (a) 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, 1 s.d.).
whether SLE-associated SNPs affect these genes in αPSMB9 expression. Both processes are hallmark features of SLE. Variants affected IFN-α complement genes. We observed that four common SLE-associated variants affected IFN-α response genes: 4,27,28 and decreased expression of the complement complex, which has a protective role in lupus: complete deletion of the genes encoding the C1q subunits practically ensures the development of SLE.29,30 CLEC10A and CLEC4C belong to the C-type lectin family, which includes mannose-binding lectins (MBLs).

Figure 2 Independent trans-eQTL effects emanating from the IKZF1 locus. SNP rs4917014, associated with SLE, and unlinked SNP rs4917014, affects IKZF1 in cis and many 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 populations1,24) not only increased expression of five different IFN-α response genes (HERC5, IFI6, IFIT1, MX1 and TNFRSF21; Fig. 2) but also decreased expression of three different probes in CLEC10A. We also observed a nominally significant association of rs4917014[T] with decreased expression of C1QB (P = 5.2 × 10−6; FDR = 0.28), encoding a subunit of the C1q complement complex, which has a protective role in lupus: complete deletion of the genes encoding the C1q subunits practically ensures the development of SLE.29,30 CLEC10A and CLEC4C belong to the C-type lectin family, which includes mannose-binding lectins (MBLs).
Although, to our knowledge, CLEC10A and CLEC4C have not been studied in the context of SLE, the role of MBLs is similar to that of the C1q complex, and MBLs are a risk factor for the development of autoimmunity in humans and mice. The rs4917014 trans eQTLs replicated well in the peripheral blood and monocyte replication data sets and reinforce the role of altered expression of the IFN-γ pathway, C-type lectin and C1q genes in SLE. Individuals without SLE but who carry the rs497014[T] risk allele show these pathway alterations, indicating that these affected pathways are not solely a consequence of SLE but could precede SLE onset.

We investigated the underlying mechanisms of the effects exerted by rs4917014. IKZF1 is the only gene overlapping the rs4917014 locus. As this gene encodes a transcription factor (Ikars-family zinc finger 1), cis- regulatory effects of rs4917014 on IKZF1 and consequent altered IKZF1 protein levels could constitute a mechanism for the detected trans-eQTL effects. However, because our meta-analysis did not initially detect a cis eQTL on the Illumina probe for IKZF1 located near the 5′ UTR of the gene, we investigated the 3′ UTR using Deep Serial Analysis of Gene Expression (DeepSAGE) next-generation RNA sequencing data from 94 peripheral blood samples. The variant rs4917014[T] allele increased expression levels of the 3′ UTR of IKZF1 (Spearman’s correlation = 0.45; P = 6.29 × 10^{-6}). Using Encyclopedia of DNA Elements (ENCODE) Project chromatin immunoprecipitation and sequencing (ChIP-seq) data, we observed significantly increased IKZF1 protein binding within genomic locations corresponding with trans-eQTL- upregulated genes compared with all other genic DNA (Wilcoxon P value = 0.046) and with SLE cis eQTL-upregulated genes outside of the IKZF1 locus (Wilcoxon P value = 4.3 × 10^{-3}), thereby confirming the importance of IKZF1 in SLE. IKZF1 is also important for other phenotypes: rs12718597, an unlinked intronic variant within the importance of IKZF1 in MCV as well. However, although rs12718597[A] was associated in trans with the upregulation of 31 genes and with the downregulation of 19 genes, none of the SLE trans-regulated genes overlapped with the MCV trans-regulated genes. The latter were mainly involved in hemoglobin metabolism and did not show increased IKZF1 binding (Wilcoxon P value = 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. For example, rs174546 (located in the 3′ UTR of FADS1 and associated with metabolic syndrome) and with low-density lipoprotein (LDL) and total cholesterol levels affected the expression of TMEM258, FADS1 and FADS2 in cis with the expression of LDLR in trans (Supplementary Fig. 14). LDLR encodes the LDL receptor and contains common variants that are also associated with lipid levels. LDLR gene expression levels correlated negatively (P < 3.0 × 10^{-4}) with total, high-density lipoprotein (HDL) and LDL cholesterol levels in the tested cohorts (Rotterdam Study and EGCUT; Supplementary Table 9), indicating that peripheral blood is a useful tissue for gaining insight into the downstream effects of lipid-regulating SNPs.

For 21 different complex traits, at least 2 unlinked variants that have been associated with these diseases affected exactly the same gene in trans (compared with 1 complex trait similarly affected by variants from equally sized but permuted lists of trans-eQTLs; Table 2, Supplementary Fig. 15 and Supplementary Table 10). Although most of these traits are hematomal (for example, 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 proxy rs653178 (located in an intronic region of ATXN2 on chromosome 12) have been associated with several autoimmune diseases, including T1D39,40 and the production of autoantibodies therein38,39, celiac disease4,40, hyperthyroidism41, vitiligo42 and rheumatoid arthritis43, as well as with other complex traits such as blood pressure43,44, chronic kidney disease45 and eosinophil counts. We observed a cis-eQTL effect for this SNP on SH2B3 (FDR < 0.05) and trans-eQTL effects on 14 genes (FDR < 0.05; Fig. 3), all of which are highly expressed in neutrophils. Because the trans-eQTLs effects could be explained by known effect of rs3184504 on differences in cell count proportions, we correlated

| Table 2 Complex traits where multiple unlinked SNPs affect the same downstream genes |
|---------------------------------|-----------------|-----------------|-----------------|
| Trait type | Complex trait | Genes affected by at least two unlinked trait-associated SNPs |
| Immune-related traits | T1D | GBP4, STAT1 |
| | T1D autoantibodies | GBP4, STAT1 |
| | Celiac disease | CXCR6, FXYD1 |
| | Multiple sclerosis | CDS5 |
| Blood pressure traits | Diastolic blood pressure | LOC338758 |
| Hematological traits | Systolic blood pressure | LOC338758 |
| | Hemoglobin | ALAS2 |
| | Hematological parameters | FBX07 |
| | F cell distribution | ESPN, PHOSPH01, GNAS, TSPAN13, VWCE, |
| | Hemotocrit | 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 |
| | β-thalassemia | PHOSPH01, VWCE, TSPAN13, ESPN |
| Mean corpuscular volume | ALAS2, C18orf10, DNAJ12B, ESPN, HMB, KEL, PD2K1P1, PIM1, PRDX5, RAP1GAP, UBNX6, VWCE, XY |
| Mean corpuscular hemoglobin | ALAS2, C1orf128, C22orf13, C5orf4, CCBP2, CSNA, DNAJ12B, EIF2AK1, ESPN, FBX07, HAGH, HMB, HPS1, KEL, ILK3, KRT1, LCGAL53, MARCHE8, MCOLN1, OSMBP2, PD2K1P1, PHOSPH01, PIM1, PLEK2, PPP2R5B, PRDX5, PTMS, RAP1GAP, RIOK3, TGM2, TSTA3, UBNX6, VWCE, XY |
| Mean platelet volume | ABC3C, AL353716.18, AQP10, C1orf93, C6orf152, C2BP5, CTDSPL, CTNT, CXCL5, ESAM, F13A1, GNB5, GNG11, G6P, GUCY1A3, ITGA2B, ITGB5, LIM51, LGF166, MMRN1, MPL, NRGN, PARVb, PRDX6, PTOR, RAB27B, RBPM12, SAMD14, SH3BGR1L2, TSPAN9, DLC1 |

NATURE GENETICS VOLUME 45 | NUMBER 10 | OCTOBER 2013

© 2013 Nature America, Inc. All rights reserved.
METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. We have made a browser available for all significant trans-eQTLs and cis-eQTLs at http://www.genenetwork.nl/bloodeqtlbrowser. This browser also provides all trans-eQTLs that we detected at a somewhat less stringent FDR of 0.5 to enable more in-depth post hoc analyses. Gene expression data are available for download at the Gene Expression Omnibus (GEO) (GSE36382, GSE20142, GSE20332, GSE33828, GSE3321, GSE47729; GSE48348 and GSE48152) and ArrayExpress (E-TABM-1036, E-MTAB-945 and E-MTAB-1708).

ACKNOWLEDGMENTS

Acknowledgments for each participating cohort can be found in the Supplementary Note.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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


ONLINE METHODS

Study populations. We performed a whole-genome eQTL meta-analysis of 5,311 samples from peripheral blood divided over a total of 9 data sets from 7 cohorts, including EGCGUT24 (n = 891), InCHIANTI25 (n = 611), Rotterdam Study26 (n = 762), Fehrmann27 (n = 1,240 on the Illumina HT12v3 platform and 229 on the Illumina H8v2 platform), HVT1-19 (n = 43 on the Illumina HT12v3 platform and 63 on the Illumina HT12v4 platform), SHIP-TREND28 (n = 963) and DILGOM29 (n = 509). Gene expression data for each data set were obtained by isolating RNA using either PAXGene (Becton Dickinson) or Tempus (Life Technologies) tubes and then hybridizing RNA to Illumina whole-genome Expression BeadChips (HT12v3, HT12v4 or H8v2 arrays). Gene expression platforms were harmonized by matching probe sequences across the different platforms. Mappings for these sequences were obtained by mapping the sequences against Build 36 of the human genome (Ensembl Build 54, hg18) using the BLAT, BWA and SOAPv2 sequence alignment programs. Highly stringent alignment criteria were used to ensure that probes mapped unequivocally to a single genomic position. Genotype data were acquired using different genotyping platforms and were harmonized by imputation, using the HapMap 2 CEU population as a reference48. Each data set was individually checked for sample mix-ups using MixupMapper 49. For a full descriptions of the individual data sets, the results of the sample mix-up analysis, specifics on the gene expression platforms used and probe mapping and filtering procedures, see the Supplementary Note.

Gene expression normalization. Gene expression data were quantile normalized to the median distribution and were subsequently log2 transformed. Probe and sample means were centered to zero. Gene expression data were then corrected for possible population structure through the removal of four multidimensional scaling components using linear regression. We reasoned earlier that normalized gene expression data still contain large amounts of non-genetic variation3. Therefore, after correction for population stratification, we performed principal-component analysis (PCA) on the sample correlation matrix. We performed a separate QTL analysis for each principal component to ascertain whether genetic variants could be detected that affected each principal component. If we found an effect on the principal component, we did not correct the expression data for this component to ensure that we would not unintentionally remove genetic effects from the expression data. We established the significance of these associations by controlling the FDR, testing each association against a null distribution created by repeating the analysis 100 times (permuting the sample labels for each iteration53). Principal components that did not show significance at the FDR threshold of 0.05 were removed from the gene expression data by linear regression. In all but 2 very small data sets, the first 40 principal components were removed (excluding those components for each cohort that showed a QTL effect). We observed that the removal of these 40 components resulted in the identification of the highest number of eQTLs in each data set. Although principal-component correction might remove some eQTL effects, we observed that the majority of trans-eQTL effects (95% when removing 35 principal components and 90% when removing 45 principal components) were independent of the number of principal components removed (Supplementary Fig. 16).

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 kb, whereas eQTLs with a distance greater than 5 Mb were defined as trans-eQTLs. Only SNPs with a minor allele frequency (MAF) of >0.05 and a Hardy-Weinberg equilibrium P-value of <0.001 were included in the analyses. Because most cohorts had generated 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, to address issues with respect to computational power and multiple testing, we confined our trans-eQTL analysis to those SNPs present in the Catalog of Published GWAS (see URLs; accessed 16 July 2011). We reasoned that, for genes with strong cis-eQTL effects, a cis-eQTL effect might obscure the detectability of trans-eQTLs. Therefore, we used linear regression to remove cis-eQTL effects before trans-eQTL mapping and observed a 12% increase in the number of detected trans-eQTLs (Supplementary Fig. 17). For each cohort, eQTLs were mapped using a Spearman’s rank correlation on imputed genotype dosages. We used a weighted z-score method for subsequent meta-analysis51. To generate a realistic null distribution, we permuted the sample identifiers of the expression data and repeated this analysis 10 times (Supplementary Fig. 18). In each permutation, the sample labels were permuted. We then corrected for multiple testing by setting the FDR at 0.05, testing each P value in the real data against a null distribution created from the permuted data sets52 (Supplementary Note). It has been suggested that false-positive eQTL effects can arise owing to polymorphisms in the probe sequences52,53. Therefore, we tested whether a significant cis-eQTL SNP was in LD (r2 > 0.2) with any SNP in the cis probe sequence, using the Western European subpopulations of the 1000 Genomes Project25 (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 be mapped to the vicinity of the trans-eQTL SNP (which would imply a cis-eQTL rather than a trans-eQTL effect). For this analysis, we tried to map the trans-eQTL probe sequences, using very permissive settings, within a 5-Mb window centered on the trans-eQTL SNP. SNP-probe combinations were defined as false positives and were removed from further analysis. After this filtering, we recalculated the FDR for both the cis- and trans-eQTL results.

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 SNPInfo54, SNP Nexus55,56 and HaploReg57, which integrate multiple data sources (such as the ENCODE Project58, Ensembl59 and several miRNA databases). We limited these analyses to those trans-eQTL SNPs that were previously shown to be associated with complex traits at genome-wide significance (trait-associated SNPs; reported P < 5 × 10−8). These SNPs were subsequently pruned (using the --clump command in PLINK with r2 < 0.2). We used permuted trans-eQTL data to generate realistic null distributions for each of these tools: we selected equally sized sets of unlinked SNPs (r2 < 0.2 in the Western European subpopulations of the 1000 Genomes Project25, 2011-05-21 release; 286 individuals, excluding Finnish individuals) that showed the highest significance in the permuted data, ensuring that only trait-associated SNPs were included in the null distribution, as it is known that trait-associated SNPs in general already have different functional properties than randomly selected SNPs50 (for example, trait-associated SNPs typically map in closer proximity to genes than randomly selected SNPs). We also ensured that none of the SNPs in the null distribution were affecting genes in trans or were linked to those SNPs (r2 = 0.2 in 1000 Genomes Project data). We then identified perfect proxies (r2 = 1.0 in 1000 Genomes Project data). For SNPInfo and SNP Nexus, we calculated the enrichment for each functional category using a Fisher’s exact test. We examined enhancer enrichment in nine different cell types using HaploReg, averaging enhancer enrichment over the ten permutations.

Convergence analysis. We determined which unlinked trait-associated SNPs showed eQTL effects on exactly the same gene: for each trait, we analyzed the SNPs that are known to be associated with the trait and assessed whether any unlinked SNP pair (r2 < 0.2; distance between SNPs of >5 Mb) showed a cis- and/or trans-eQTL effect on exactly the same gene, as previously described5. To determine whether the number of traits for which we observed this...
phenomenon was higher than expected by chance, we repeated this analysis 20 times, each time using a different set of permuted \textit{trans} eQTLs, equal in size to the non-permuted set of \textit{trans} eQTLs.

**SLE IKZF1 ENCODE ChIP-seq analysis.** We used IKZF1 ChIP-seq signal data obtained from the ENCODE Project\(^{12}\) (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 and bias in GC content) and performed a Wilcoxon Mann-Whitney test to determine whether the upregulated genes (\textit{MX1}, \textit{TNFRSF21}, \textit{IFIT1-LIPA}, \textit{HERC5}, \textit{CLEC4C} and \textit{IFI6}) showed a higher ChIP-seq signal than the average signal for all other human genes.