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

Westra, Harm-Jan; Peters, Marjolein J.; Esko, Tonu; Yaghootkar, Hanieh; Schurmann, Claudia; Kettunen, Johannes; Christiansen, Mark W.; Fairfax, Benjamin P.; Schramm, Katharina; Powell, Joseph E.

Published in:
Nature Genetics

DOI:
10.1038/ng.2756

Publication date:
2013

nature genetics

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

Harm-Jan Westra[1,40], Marjolein J Peters[2,3,40], Tõnu Esko[4,40], Hanieh Yaghootkar[5,40], Claudia Schurmann[6,40], Johannes Kettunen[7,8,40], Mark W Christiansen[9,40], 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[29–31], Harald Grallert[32], Thomas Meitinger[12,13,33,34], Konstantin Strauch[35,36], Yang Li[37], Ritsert C Jansen[37], Peter M Visscher[14,15], Julian C Knight[10], Bruce M Psaty[9,38,41], Samuli Ripatti[7,8,39,41], Alexander Teumer[6,41], Timothy M Frayling[5,41], Andres Metspalu[4,41], Joyce B J van Meurs[2,3,41] & Lude Franke[1,41]

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)[1], altered gene expression of *C1QB* and five type I interferon response genes, both hallmarks of SLE[2–4]. 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 studies[5–8] 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* eQTLs[5,9–13], 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 (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 in *trans* 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 multiple-testing 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 (643 *trans* eQTLs, including 200 unique SNPs and 223 different genes with a more stringent Bonferroni correction).

We used stringent procedures for *trans*-eQTL detection (**Supplementary Note**) and various benchmarks to ensure reliability: for 26 *trans*-eQTL genes, the eQTL SNP affected multiple probes within these genes (**Supplementary Table 5**), always with consistent allelic directions, suggesting that our probe-filtering procedure was effective in preventing false-positive *trans* eQTLs. *Trans* eQTLs showed similar effect sizes across the various cohorts (**Supplementary Fig. 5**).

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

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

	Cis-eQTL analysis		*Trans*-eQTL analysis	
	FDR <0.05 significance	Bonferroni significance	FDR <0.05 significance	Bonferroni significance
Number of significant unique SNP-probe pairs	664,097	395,543	1,513	643
Number of significant unique eQTL SNPs	397,310	266,036	346	200
Number of significant unique eQTL probes	8,228	5,738	494	240
Number of significant unique eQTL-regulated genes	6,418	4,690	430	223
Number of significant unique eQTL probes not mapping to genes	636	326	35	13

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[5] 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)[22] and 79% in BSGS (n = 862 samples)[23] (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 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[9], 282 B cell samples[9] and 608 HapMap lymphoblastoid cell line (LCL) samples[24]; **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 $P < 5 \times 10^{-8}$; out of 4,542 unique SNPs that we tested). These 2,082 SNPs showed a significantly higher number of *trans*-eQTL effects compared with the 2,460 tested SNPs with reported disease associations at lower significance levels ($P = 8 \times 10^{-22}$; **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; $P = 5.6 \times 10^{-49}$; **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 ($P = 5 \times 10^{-25}$; **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 tissue[13] 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

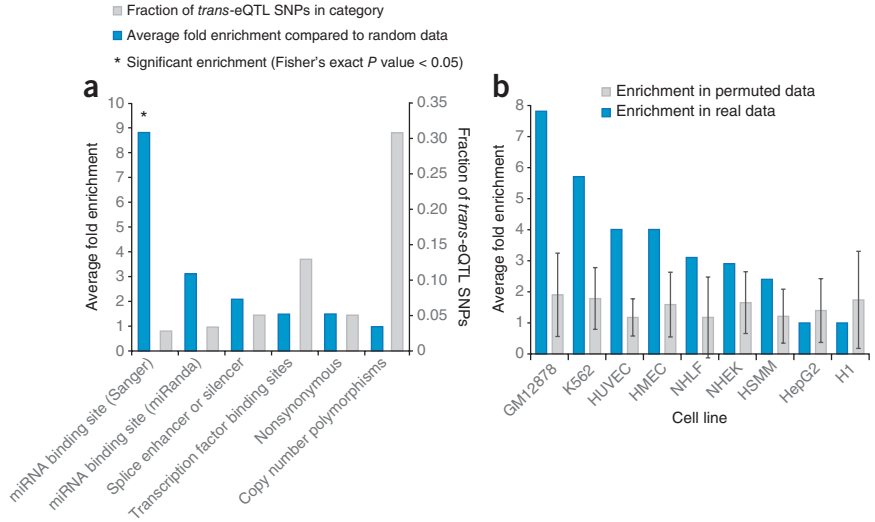

Figure 1 *Trans*-eQTL SNPs are enriched for functional elements. We investigated whether *trans*-eQTL SNPs are enriched for certain functional elements using the online tools SNPInfo, SNPNexus and HaploReg that rely on data from, among others, the ENCODE Project. (**a**) *Trans*-eQTL SNPs are enriched 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.).

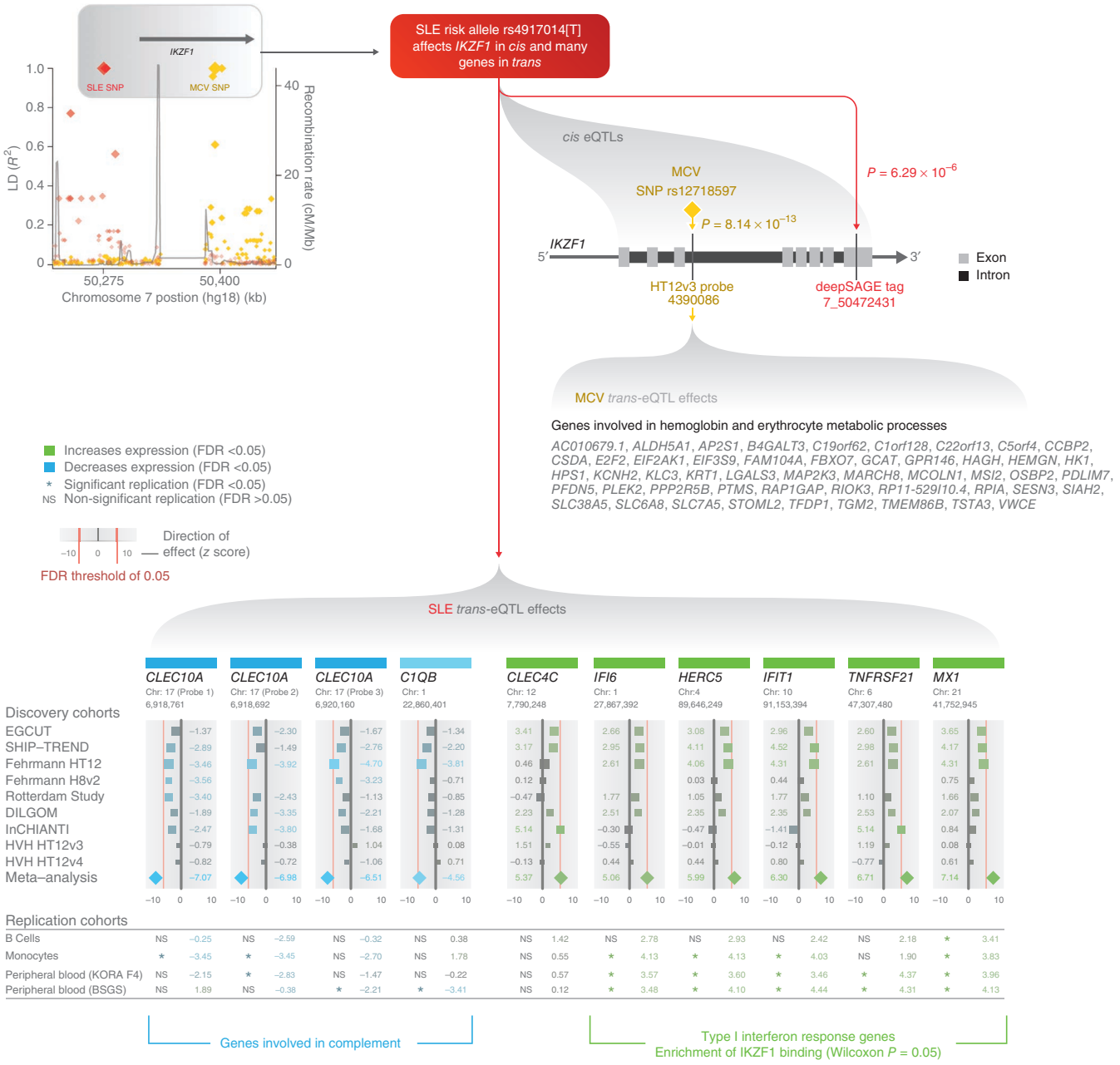

Figure 2 Independent *trans*-eQTL effects emanating from the *IKZF1* locus. SNP rs4917014, associated with SLE, and unlinked SNP rs4917014, associated with MCV, both affect the expression of *IKZF1* in *cis*. rs12718597 affects 50 genes in *trans* (mostly involved in hemoglobin metabolism), and rs4917014 affects 8 different genes in *trans*: the rs4917014[T] risk allele is associated with increased expression of genes involved in the type I interferon response. At a somewhat lower significance threshold (FDR = 0.28), rs4917014[T] is associated with decreased complement *C1QB* expression. Both processes are hallmark features of SLE.

dysregulation ultimately leads to disease or vice-versa. One example of this type of complex disease is SLE, which is an autoimmune disease causing inflammation and tissue damage. Individuals with SLE have increased type I interferon (IFN-α) levels, increased expression of IFN-α response genes[4,27,28] and decreased expression of the *C1Q* complement genes. We observed that four common SLE-associated variants affected IFN-α response genes in *cis* (*IRF5*, *IRF7*, *TAP2* and *PSMB9*; **Supplementary Table 3**). As most SLE-associated SNPs do not map near complement or IFN-α response genes, we assessed whether SLE-associated SNPs affect these 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*; **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* ($P = 5.2 \times 10^{-6}$; 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[3]. 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-α 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 (Ikaros-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[31]. The variant rs4917014[T] allele increased expression levels of the 3′ UTR of *IKZF1* (Spearman's correlation = 0.45; $P = 6.29 \times 10^{-6}$). Using Encyclopedia of DNA Elements (ENCODE) Project[32] 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^{-4}), thereby confirming the importance of *IKZF1* in SLE. *IKZF1* is also important for other phenotypes: rs12718597, an unlinked intronic variant within *IKZF1*, is associated with mean corpuscular volume (MCV)[33] and affects the expression of Illumina probe 4390086 near the 5′ end of *IKZF1* in *cis*. *Ikzf1* knockout mice show abnormal erythropoiesis[34], suggesting a causal role for human *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[35] and with low-density lipoprotein (LDL) and total cholesterol levels[36,37]) affected the expression of *TMEM258*, *FADS1* and *FADS2* in *cis* and 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[37]. *LDLR* gene expression levels correlated negatively ($P < 3.0 \times 10^{-4}$) 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 hematological (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).

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*, *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*
	β-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*

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 T1D[38,39] and the production of autoantibodies therein[38,39], celiac disease[8,40], hyperthyroidism[41], vitiligo[42] and rheumatoid arthritis[40], as well as with other complex traits such as blood pressure[43,44], chronic kidney disease[45] and eosinophil counts[46]. 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[46], we correlated

Figure 3 Two unlinked T1D risk alleles are associated with increased *STAT1* and *GBP4* expression. rs3184504[T], a risk allele for T1D (chromosome 12), affects the expression of *SH2B3* in *cis* but also affects the expression of 14 unique genes in *trans*, including 2 interferon-γ response genes, *GBP4* and *STAT1*. Another, unlinked T1D risk allele (rs4788084[C] on chromosome 16) also increases expression of these two interferon-γ response genes, suggesting that an elevated interferon-γ response is important in T1D.

the expression levels of *trans*-regulated genes with cell counts in two cohorts (Rotterdam Study and EGCUT) but did not observe significant correlations (**Supplementary Table 9**). The identified *trans* eQTLs describe different biological functions: the T1D risk allele rs3184504[T] was associated with decreased expression of nine genes, most of which are involved in Toll-like receptor signaling[47] (*C12orf75*, *FOS*, *IDS*, *IL8*, *LOC338758*, *NALP12*, *PPP1R15A*, *S100A10* and *TAGAP*) and with increased expression of five genes involved in the interferon-γ response (*GBP2*, *GBP4*, *STAT1*, *UBE2L6* and *UPP1*). We observed that another T1D risk allele, rs4788084[C][38,39] on chromosome 16, was also associated with increased expression of *GBP4* and *STAT1*, showing how different T1D risk alleles converge, with both alleles causing an increase in the expression of interferon-γ response genes.

In summary, our eQTL meta-analysis identified and replicated downstream effects for 233 trait-associated SNPs. Our analyses show that *trans*-eQTL mapping in blood for lipid-regulatory and immune-mediated disease variants yields insights into downstream pathways that are biologically meaningful. Future, larger-scale *trans*-eQTL analyses in blood will likely uncover many more of these regulatory relationships.

URLs. Catalog of Published GWAS (16 July 2011), http://www.genome.gov/gwastudies/.

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, GSE33321, GSE47729; GSE48348 and GSE48152) and ArrayExpress (E-TABM-1036, E-MTAB-945 and E-MTAB-1708).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

Experiment design and method development: H.-J.W., M.J.P., T.E., H.Y., C.S., J. Kettunen, M.W.C., B.P.F., K. Schramm, J. Karjalainen, T.L., Y.L., R.C.J., B.M.P., S.R., A.T., T.M.F., A.M., J.B.J.M. and L. Franke. Reviewing and editing of the manuscript: H.-J.W., M.J.P., T.E., H.Y., C.S., J. Kettunen, M.W.C., B.P.F., A.Z., A.G.U., A.H., F.R., V.S., J.B., T.L., Y.L., R.C.J., P.M.V., J.C.K., B.M.P., S.R., A.T., T.M.F., A.M., J.B.J.M. and L. Franke. Data collection: D.V.Z., J.H.V., J. Karjalainen, S.W., F.R., P.A.C.t.H., E.R., K.F., M. Nelis, L.M., D.M., L. Ferrucci, A.B.S., D.G.H., M.A.N., G.H., M. Nauck, D.R., U.V., M.P., A.S.-D., S.A.G., D.A.E., G.W.M., S.M., H.P., C.H., M.R., H.G., T.M., K. Strauch and L.H.V.d.B. Replication of *trans*-eQTL results: B.P.F., K. Schramm, J.E.P., P.M.V. and J.C.K.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Han, J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234–1237 (2009).
2. Bengtsson, A.A. *et al.* Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies. *Lupus* **9**, 664–671 (2000).

3. Bohlson, S.S., Fraser, D.A. & Tenner, A.J. Complement proteins C1q and MBL are pattern recognition molecules that signal immediate and long-term protective immune functions. *Mol. Immunol.* **44**, 33–43 (2007).
4. Ytterberg, S.R. & Schnitzer, T.J. Serum interferon levels in patients with systemic lupus erythematosus. *Arthritis Rheum.* **25**, 401–406 (1982).
5. Fehrmann, R.S. *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.* **7**, e1002197 (2011).
6. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
7. Pickrell, J.K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
8. Dubois, P.C. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
9. Fairfax, B.P. *et al.* Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
10. Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7**, e1002078 (2011).
11. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
12. Heinig, M. *et al.* A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* **467**, 460–464 (2010).
13. Small, K.S. *et al.* Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
14. Metspalu, A. The Estonian Genome Project. *Drug Dev. Res.* **62**, 97–101 (2004).
15. Tanaka, T. *et al.* Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet.* **5**, e1000338 (2009).
16. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.* **26**, 657–686 (2011).
17. Heckbert, S.R. *et al.* Antihypertensive treatment with ACE inhibitors or β-blockers and risk of incident atrial fibrillation in a general hypertensive population. *Am. J. Hypertens.* **22**, 538–544 (2009).
18. Psaty, B.M. *et al.* The risk of myocardial infarction associated with antihypertensive drug therapies. *J. Am. Med. Assoc.* **274**, 620–625 (1995).
19. Smith, N.L. *et al.* Esterified estrogens and conjugated equine estrogens and the risk of venous thrombosis. *J. Am. Med. Assoc.* **292**, 1581–1587 (2004).
20. Teumer, A. *et al.* Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. *Am. J. Hum. Genet.* **88**, 664–673 (2011).
21. Inouye, M. *et al.* An immune response network associated with blood lipid levels. *PLoS Genet.* **6**, e1001113 (2010).
22. Mehta, D. *et al.* Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet.* **21**, 48–54 (2013).
23. Powell, J.E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS ONE* **7**, e35430 (2012).
24. Wang, C. *et al.* Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. *Eur. J. Hum. Genet.* doi:10.1038/ejhg.2012.277 (19 December 2012).
25. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
26. Patterson, K. 1000 genomes: a world of variation. *Circ. Res.* **108**, 534–536 (2011).
27. Baechler, E.C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. USA* **100**, 2610–2615 (2003).
28. Bennett, L. *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J. Exp. Med.* **197**, 711–723 (2003).
29. McAdam, R.A., Goundis, D. & Reid, K.B. A homozygous point mutation results in a stop codon in the C1q B-chain of a C1q-deficient individual. *Immunogenetics* **27**, 259–264 (1988).
30. Botto, M. *et al.* Homozygous C1q deficiency causes glomerulonephritis associated with multiple apoptotic bodies. *Nat. Genet.* **19**, 56–59 (1998).
31. Zhernakova, D.V. *et al.* DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts. *PLoS Genet.* **9**, e1003594 (2013).
32. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
33. Ganesh, S.K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41**, 1191–1198 (2009).
34. Wang, J.H. *et al.* Selective defects in the development of the fetal and adult lymphoid system in mice with an *Ikaros* null mutation. *Immunity* **5**, 537–549 (1996).
35. Zabaneh, D. & Balding, D.J. A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS ONE* **5**, e11961 (2010).
36. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**, 35–46 (2009).
37. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
38. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
39. Plagnol, V. *et al.* Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet.* **7**, e1002216 (2011).
40. 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.* **7**, e1002004 (2011).
41. Eriksson, N. *et al.* Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS ONE* **7**, e34442 (2012).
42. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44**, 676–680 (2012).
43. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
44. Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
45. Köttgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).
46. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
47. Rotival, M. *et al.* Integrating genome-wide genetic variations and monocyte expression data reveals *trans*-regulated gene modules in humans. *PLoS Genet.* **7**, e1002367 (2011).

[1]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [2]Department of Internal Medicine, Erasmus Medical Centre Rotterdam, Rotterdam, The Netherlands. [3]Netherlands Genomics Initiative–sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden and Rotterdam, The Netherlands. [4]Estonian Genome Center, University of Tartu, Tartu, Estonia. [5]Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK. [6]Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [7]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [8]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [9]Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. [10]Wellcome Trust Centre for Human Genetics, Oxford, UK. [11]Department of Oncology, Cancer and Haematology Centre, Churchill Hospital, Oxford, UK. [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]Queensland Brain Institute, 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 Centre Rotterdam, 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, Exeter, UK. [20]Clinical Research Branch, National Institute on Aging–Advanced Studies in Translational Research on Aging (ASTRA) Unit, Harbor Hospital, Baltimore, Maryland, USA. [21]Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [22]Department of Molecular Neuroscience, Reta Lila Laboratories, Institute of Neurology, University College London, London, UK. [23]Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [24]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [25]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [26]Computational Medicine Core, Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, Department of Medicine, University of Washington, Seattle, Washington, USA. [27]Department of Statistics, University of Auckland, Auckland, New Zealand. [28]Queensland Institute of Medical Research, Herston, Queensland, Australia. [29]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [30]Department of Endocrinology and Diabetology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [31]Department of Metabolic Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [32]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [33]German Center for Cardiovascular Research (DZHK), Göttingen, Germany. [34]Munich Heart Alliance, Munich, Germany. [35]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians Universität, Neuherberg, Germany. [36]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [37]Groningen Bioinformatics Center, University of Groningen, Groningen, The Netherlands. [38]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [39]Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [40]These authors contributed equally to this work. [41]These authors jointly directed this work. Correspondence should be addressed to L. Franke (lude@ludesign.nl).

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 EGCUT[14] (n = 891), InCHIANTI[15] (n = 611), Rotterdam Study[16] (n = 762), Fehrmann[5] (n = 1,240 on the Illumina HT12v3 platform and 229 on the Illumina H8v2 platform), HVH[17–19] (n = 43 on the Illumina HT12v3 platform and 63 on the Illumina HT12v4 platform), SHIP-TREND[20] (n = 963) and DILGOM[21] (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 reference[48]. 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 $\log_2$ 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 variation[5]. 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 iteration[50]). Principal components 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 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 time 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-analysis[51]. To generate a realistic null distribution, we permuted the sample identifiers of the expression data and repeated this analysis ten 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 sets[50] (**Supplementary Note**). It has been suggested that false-positive eQTL effects can arise owing to polymorphisms in the probe sequences[52,53]. 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 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 where at least 15 bp of the probe mapped within this 5-Mb window were deemed false positives and were 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 5 independent data sets from 4 cohorts, including data obtained from LCLs (HapMap 3, n = 608)[24], B cells and monocytes (Oxford, n = 282 and 283, respectively)[9] and whole peripheral blood (KORA F4, n = 740 and BSGS, n = 862)[22,23]. All cohorts applied the same methodology as used in the discovery phase to normalize gene expression data, check for sample mix-ups and perform *trans*-eQTL mapping, including ten permutations 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 F4 and BSGS). Further details on these data sets can be found in the **Supplementary Note**.

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[54], SNPNexus[55,56] and HaploReg[57], which integrate multiple data sources (such as the ENCODE Project[32], Ensembl[58] 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 $P < 5 \times 10^{-8}$). These SNPs were subsequently pruned (using the –clump command in PLINK with $r^2 < 0.2$). We used permuted *trans*-eQTL data to generate realistic null distributions for each of these tools: we selected equally sized sets of unlinked SNPs ($r^2 < 0.2$ 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 were included in the null distribution, as it is known that trait-associated SNPs in general already have different functional properties than randomly selected SNPs[59] (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 ($r^2 < 0.2$ in 1000 Genomes Project data). We then identified perfect proxies ($r^2 = 1.0$ in 1000 Genomes Project data). For SNPInfo and SNPNexus, 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 ($r^2 < 0.2$; distance between SNPs of >5 Mb) 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

doi:10.1038/ng.2756

phenomenon was higher than expected by chance, we repeated 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[32] (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 (*MX1*, *TNFRSF21*, *IFIT1-LIPA*, *HERC5*, *CLEC4C* and *IFI6*) showed a higher ChIP-seq signal than the average signal for all other human genes.

48. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
49. Westra, H.J. *et al.* MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics* **27**, 2104–2111 (2011).
50. Breitling, R. *et al.* Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4**, e1000232 (2008).
51. Whitlock, M.C. Combining probability from independent tests: the weighted *Z*-method is superior to Fisher's approach. *J. Evol. Biol.* **18**, 1368–1373 (2005).
52. Alberts, R. *et al.* Sequence polymorphisms cause many false *cis* eQTLs. *PLoS ONE* **2**, e622 (2007).
53. Benovoy, D., Kwan, T. & Majewski, J. Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. *Nucleic Acids Res.* **36**, 4417–4423 (2008).
54. Xu, Z. & Taylor, J.A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res.* **37**, W600–W605 (2009).
55. Dayem Ullah, A.Z., Lemoine, N.R. & Chelala, C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res.* **40**, W65–W70 (2012).
56. Chelala, C., Khan, A. & Lemoine, N.R. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics* **25**, 655–661 (2009).
57. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
58. Flicek, P. *et al.* Ensembl 2012. *Nucleic Acids Res.* **40**, D84–D90 (2012).
59. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).