Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression

Võsa, Urmo; Claringbould, Annique; Westra, Harm-Jan; Bonder, Marc Jan

Published in:
Nature Genetics

DOI:
10.1038/s41588-021-00913-z

Publication date:
2021

Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression

Urmo Võsa*[1,2#], Annique Claringbould*[1,3#], Harm-Jan Westra**[1,3], Marc Jan Bonder**[1,4], Patrick Deelen**[1,5,3,6], Biao Zeng[7], Holger Kirsten[8], Ashis Saha[9], Roman Kreuzhuber[10,11], Seyhan Yazar[12], Harm Brugge[1,3], Roy Oelen[1,3], Dylan H. de Vries[1,3], Monique van der Wijst[1,3], Silva Kasela[2], Natalia Pervjakova[2], Isabel Alves[13], Marie-Julie Favé[13], Mawussé Agbessi[13], Mark W. Christiansen[14], Rick Jansen[15], Ilkka Seppälä[16], Lin Tong[17], Alexander Teumer[18], Katharina Schramm[19,20], Gibran Hemani[21], Joost Verlouw[22], Hanieh Yaghootkar[23,24,25], Reyhan Sönmez[26,27], Andrew Brown[28,29], Viktorija Kukushkina[2], Anette Kalnapenkis[2], Sina Rüeger[30], Eleonora Porcu[30], Jaanika Kronberg[2], Johannes Kettunen[31], Bernett Lee[32], Futao Zhang[33], Ting Qi[33], Jose Alquicira Hernandez[12], Wibowo Arindrarto[34], Frank Beutner[35], BIOS Consortium[†], i2QTL Consortium[†], Julia Dmitrieva[36], Mahmoud Elansary[36], Benjamin P. Fairfax[37], Michel Georges[36], Bastiaan T. Heijmans[34], Alex W. Hewitt[38,39], Mika Kähönen[40], Yungil Kim[41,42], Julian C. Knight[37], Peter Kovacs[43], Knut Krohn[44], Shuang Li[1], Markus Loeffler[8], Urko M. Marigorta[7,45,46], Hailang Mei[47], Yukihide Momozawa[36,48], Martina Müller-Nurasyid[19,20,49], Matthias Nauck[50,51], Michel G. Nivard[52], Brenda Penninx[15], Jonathan Pritchard[53], Olli T. Raitakari[54], Olaf Rotzschke[32], Eline P. Slagboom[34], Coen D.A. Stehouwer[55], Michael Stumvoll[56], Patrick Sullivan[57], Peter-Bram A.C. 't Hoen[58], Joachim Thiery[59], Anke Tönjes[56], Jenny van Dongen[60], Maarten van Iterson[34], Jan H. Veldink[61], Uwe Völker[62], Robert Warmerdam[1,3], Cisca Wijmenga[1], Morris Swertz[5], Anand Kumar Andiappan[32], Grant W. Montgomery[33], Samuli Ripatti[63], Markus Perola[64], Zoltan Kutalik[30], Emmanouil Dermitzakis[27,28,65], Sven Bergmann[26,27], Timothy Frayling[23], Joyce van Meurs[22], Holger Prokisch[66,67], Habibul Ahsan[17], Brandon L. Pierce[17], Terho Lehtimäki[16], Dorret I. Boomsma[60], Bruce M. Psaty[68], Sina A. Gharib[14,69], Philip Awadalla[13], Lili Milani[2], Willem Ouwehand[10,70], Kate Downes[10], Oliver Stegle[4,11,71], Alexis Battle[9,72], Peter M. Visscher[33], Jian Yang[33,73], Markus Scholz[8], Joseph Powell***[12,74], Greg Gibson***[7], Tõnu Esko***[2], Lude Franke***[1,3#]

* These authors contributed equally
** These authors contributed equally
*** These authors contributed equally
† Full author list for consortium authors appears in Supplementary Information
# Correspondence to: Urmo Võsa (urmo.vosa@gmail.com), Annique Claringbould (anniqueclaringbould@gmail.com) and Lude Franke (lude@ludesign.nl)

1. Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands
2. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
3. Oncode Institute, Amsterdam, The Netherlands
4. European Molecular Biology Laboratory, Genome Biology Unit, 69117 Heidelberg, Germany
5. Genomics Coordination Center, University Medical Centre Groningen, Groningen, The Netherlands
6. Department of Genetics, University Medical Centre Utrecht, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands
7. School of Biological Sciences, Georgia Tech, Atlanta, United States of America
8. Institut für Medizinische InformatiK, Statistik und Epidemiologie, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
9. Department of Computer Science, Johns Hopkins University, Baltimore, United States of America

10. Department of Haematology, University of Cambridge and NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom
11. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom
12. Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, Australia
13. Computational Biology, Ontario Institute for Cancer Research, Toronto, Canada
14. Cardiovascular Health Research Unit, University of Washington, Seattle, United States of America
15. Amsterdam UMC, Vrije Universiteit, Department of Psychiatry, Amsterdam Public Health research institute and Amsterdam Neuroscience, The Netherlands
16. Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
17. Department of Public Health Sciences, University of Chicago, Chicago, United States of America
18. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
19. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
20. Department of Medicine I, University Hospital Munich, Ludwig Maximilian's University, München, Germany
21. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
22. Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands
23. Genetics of Complex Traits, University of Exeter Medical School, Royal Devon & Exeter Hospital, Exeter, United Kingdom
24. School of Life Sciences, College of Liberal Arts and Science, University of Westminster, 115 New Cavendish Street, London, United Kingdom
25. Division of Medical Sciences, Department of Health Sciences, Luleå University of Technology, Luleå, Sweden
26. Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland
27. Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
28. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
29. Population Health and Genomics, University of Dundee, Dundee, United Kingdom
30. Lausanne University Hospital, Lausanne, Switzerland
31. University of Helsinki, Helsinki, Finland
32. Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore
33. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
34. Leiden University Medical Center, Leiden, The Netherlands
35. Heart Center Leipzig, Universität Leipzig, Leipzig, Germany
36. Unit of Animal Genomics, WELBIO, GIGA-R & Faculty of Veterinary Medicine, University of Liege, 1 Avenue de l'Hôpital, Liège 4000, Belgium
37. Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom
38. Menzies Institute for Medical Research, School of Medicine, University of Tasmania, Hobart, Australia
39. Centre for Eye Research Australia, Department of Surgery, University of Melbourne, Australia
40. Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
41. Department of Computer Science, Johns Hopkins University, Baltimore, United States of America
42. Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, United States of America
43. IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany
44. Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany
45. Integrative Genomics Lab, CIC bioGUNE, Bizkaia Science and Technology Park, Derio, Bizkaia, Basque Country, Spain
46. IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
47. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
48. Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa 230-0045, Japan
49. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
50. Institute of Clinical Chemistry and Laboratory Medicine, Greifswald University Hospital, Greifswald, Germany
51. German Center for Cardiovascular Research (partner site Greifswald), Greifswald, Germany
52. Department of Biological Psychology, Faculty of Behaviour and Movement Sciences, VU, Amsterdam, The Netherlands
53. Stanford University, Stanford, United States of America
54. Centre for Population Health Research, Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital and University of Turku, Turku, Finland
55. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands
56. Department of Medicine, Universität Leipzig, Leipzig, Germany
57. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
58. Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands
59. Institute for Laboratory Medicine, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
60. Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam Public Health research institute and Amsterdam Neuroscience, the Netherlands
61. UMC Utrecht Brain Center, University Medical Center Utrecht, Department of Neurology, Utrecht University, Utrecht, The Netherlands
62. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
63. Statistical and Translational Genetics, University of Helsinki, Helsinki, Finland
64. National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland
65. Institute of Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland
66. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
67. Institute of Human Genetics, Technical University Munich, Munich, Germany
68. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA 98101 USA
69. Department of Medicine, University of Washington, Seattle, United States of America

70. Human Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton Cambridge, United Kingdom
71. Division of Computational Genomics and Systems Genetics, German Cancer Research Center, 69120 Heidelberg, Germany
72. Departments of Biomedical Engineering, Johns Hopkins University, Baltimore, United States of America
73. Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China
74. UNSW Cellular genomics Futures Institute, University of New South Wales, Sydney

Abstract

Genetic variants identified by genome-wide association studies (GWAS) primarily affect complex phenotypes via regulatory mechanisms on the transcriptome. To comprehensively investigate the effect of genetics on human gene expression, we performed *cis*- and *trans*- expression quantitative trait locus (eQTL) analyses using blood-derived bulk gene expression profiles from 31,684 individuals through the eQTLGen Consortium.

We detected local *cis*-eQTL effects for 88% of the 19,942 genes studied, and these effects were replicable in multiple cell types and tissues. In contrast, distal *trans*-eQTLs (detected in whole blood for 37% of the 10,317 trait-associated variants studied) showed lower replication rates in individual cell types, partially due to statistical power and confounding effects of cell-type-composition differences across individuals. We therefore performed extensive replication analyses using single-cell RNA-seq eQTL data on 1,139 individuals. These *trans*-eQTLs exert their effects via several mechanisms of action, with regulation through transcription factors (TFs) being the most prevalent. In some cases, multiple unlinked variants associated with the same complex trait converged on *trans*-genes that are known to play central roles in disease etiology. These converging patterns were recapitulated when ascertaining the effect of polygenic scores (PGS) calculated for 1,263 GWAS traits. Expression levels of 13% of the studied genes correlated with PGS, and many resulting genes are known to be associated with those traits.

This work represents the largest effort to date aimed at systematically identifying the local and distal transcriptional consequences of human genetic variation. The resource we

present here serves as a starting point for more in-depth interpretative studies of complex traits.

Main text

Expression quantitative trait loci (eQTLs) have become a common tool to interpret the regulatory mechanisms of variants associated with complex traits by genome-wide association studies (GWAS). In particular, *cis*-eQTLs, where gene expression levels are affected by a gene-proximal single nucleotide polymorphism (SNP) (<1 megabases; Mb), have been widely used for this purpose. However, the expression of *cis*-eQTL genes generally explains only a modest proportion of disease heritability[1], suggesting additional routes of regulation leading to disease.

Trans-eQTLs, where the SNP is located distal to the gene (>5 Mb) or on other chromosomes, generally have smaller effect sizes than *cis*-eQTLs and thus require larger sample sizes for detection. However, we reasoned that *trans*-eQTLs could also be relevant for complex traits because, compared to stronger *cis*-eQTL effects, each individual *trans*-effect is less likely to be dampened by compensatory post-transcriptional buffering or removed from population by negative selection[2,3]. Indeed, genes regulated by weak eQTL effects are estimated to have more impact on the phenotype as compared to those regulated by strong eQTL effects[4]. At the same time, individual *trans*-eQTL SNPs can affect many genes and collectively have a widespread impact on regulatory networks. Consequently, weak *trans*-eQTLs have the potential to identify trait-relevant genes, and *trans*-eQTLs[1,5–10] have already been used to prioritize genes that are likely to contribute to disease[5].

While *trans*-eQTLs are useful for the identification of the distal effects of a single variant, a different approach is required to determine the combined consequences of all variants associated with a polygenic trait. Polygenic scores (PGSs) summarize genome-wide combined risk for a complex disease into a single metric that may be of clinical use for the stratification of individuals in groups of high and low genetic risk[11,12]. The recently proposed omnigenic model[13,14] postulates that the heritability of most complex traits is dominated by numerous weak *trans*-effects and hypothesizes that those effects converge on a smaller set of trait-relevant 'core' genes. This suggests that associations between PGSs and gene expression (expression quantitative trait scores, eQTS) could help to prioritize putative trait-relevant genes (**Supplementary Equations**, Liu et al.[14]). While it remains unclear what fraction of the genome affects complex traits, we here systematically investigated *trans*-eQTLs and eQTS to determine how genetic effects influence and converge on genes and pathways and whether these effects could be informative about the biology of the respective traits.

To maximize the statistical power to detect eQTL and eQTS effects, we performed a large-scale meta-analysis in up to 31,684 blood samples from 37 cohorts (assayed using three gene expression platforms) in the context of the eQTLGen Consortium. This allowed us to identify *cis*-eQTLs for 16,987 genes, *trans*-eQTLs for 6,298 genes and eQTS effects for 2,568 genes (false discovery rate (FDR) <0.05, determined by permutations (**Methods**); 15,073, 2,666 and 905 genes, respectively, after more conservative Bonferroni correction; out of 19,942 tested genes; **Figure 1**) that revealed complex regulatory effects of trait-associated variants. We then replicated these eQTLs across gene expression platforms, in other tissues and in single cell data. What we found was that, while the overall concordance was good, formal replication remained limited, possibly due to the effects of genetics on blood cell composition, the limited power of the available replication datasets and the cell-type-specific nature of distal effects. To demonstrate the utility of our resource, we combined the associations we identified with additional data layers to gain biological insights into the mechanisms of blood eQTLs and complex traits.

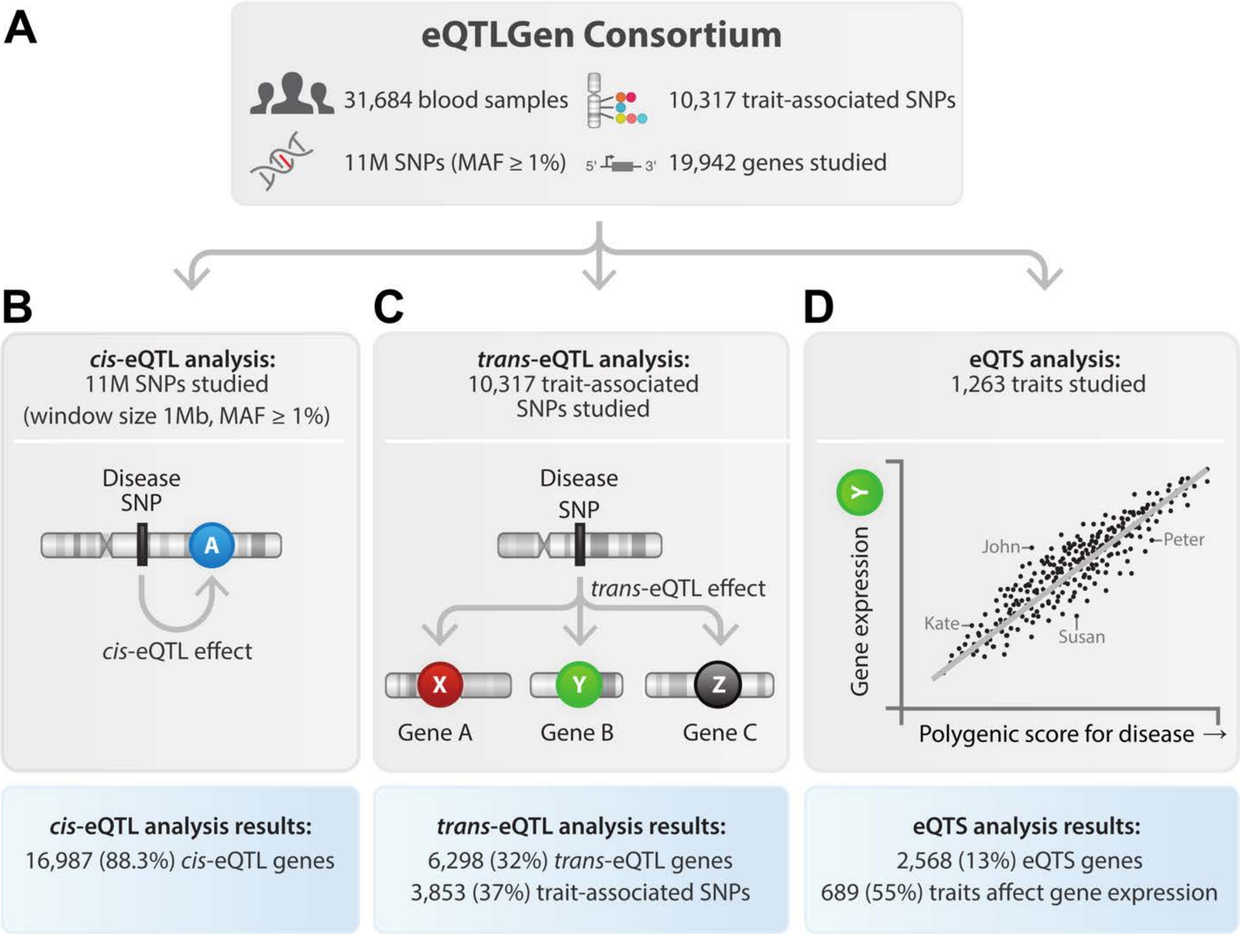


Figure 1. Overview of the study. Overview of discovery analyses and their results.

Results

Meta-analyses on local and distal gene expression

We performed *cis*-eQTL, *trans*-eQTL and eQTS meta-analyses using eQTLGen Consortium data from 31,684 individuals (**Figure 1A**, **Supplementary Table 1, Supplementary Information**). Our consortium contains datasets profiled using different expression profiling platforms, including several Illumina and Affymetrix expression array versions and RNA-seq, making a direct meta-analysis impossible. We therefore made use of co-regulation patterns between genes to assign the best-matching expression probe from each expression array type to each gene (**Methods**). After applying this method, we meta-analysed the different expression profiling platforms on gene-level. We then performed eQTL and eQTS discovery and replication analyses between each combination of platforms. Because the different platforms had variable sample sizes, which resulted in differences in replication power, replication rates varied from 86.3% (among *cis*-eQTLs in the largest replication dataset) to 13% (among *trans*-eQTLs in the smallest replication dataset) **(Supplementary Figure 1A-C**). However, effects that were replicated (FDR<0.05) showed consistent allelic directions for *cis*-eQTLs (average over all comparisons 93.23%), *trans*-eQTLs (average over all comparisons 99.2%) and eQTS (average over all comparisons 99.4%). This demonstrates that our integration method enabled us to combine different expression profiling platforms and, importantly, that the eQTLs and eQTSs identified by our approach are replicable between different whole blood datasets (**Methods**, **Supplementary Results**, **Supplementary Figure 1A-C**). In all the analyses, we accounted for unknown technical confounders (such as batch effects) and biological confounders (such as interindividual differences in cell-type-composition) by correcting the expression data per cohort for up to 25 expression principal components (PCs) that were not associated with genetic variation (**Methods**). When testing for cell-type-composition effects in a subset of samples (N up to 3,831) from the BIOS cohort, this correction adjusted for the majority (**Supplementary Note, Supplementary Figure 2**). Nevertheless, we acknowledge that our dataset may still include residual cell-type-composition effects.

As our analysis tested nearly 20,000 genes, our study required a strategy to correct for multiple testing. Bonferroni correction is overly stringent for eQTL analysis due to many correlating genes and extensive linkage between genetic variants. Instead, permutation-based approaches[5,15–17] or Benjamini-Hochberg FDR[18,6,1] are often used for multiple testing correction in eQTL studies. Here, we adopted a permutation-based strategy[5,15,19] where each cohort performed the regular analyses and 10 permutations in which the links between gene expression and genotypes were shuffled in each permutation (**Methods**). As with the non-permuted results, we meta-analyzed the results from each permutation and compared the P-value distributions across all tests between the non-permuted and permuted data to determine an FDR estimate for each association (methodology varies slightly between *cis*-eQTL, *trans*-eQTL and eQTS analyses, see details in **Methods**). We have previously shown that these FDR estimates stabilize after only a few permutations, demonstrating that 10 permutations is sufficient[5]. By evaluating the FDR estimates over all tests performed, our approach yields an analysis-wide estimate of FDR (i.e. genome-wide for *cis*-eQTLs), rather than a specific FDR estimate per gene, which would require many more permutations. For all the discovery analyses, we observed that our strategy was more conservative than Benjamini-Hochberg FDR and less stringent than the Bonferroni method (**Supplementary Figure 3**). Because users of our resource may require different levels of stringency, we provide both permutation-based FDRs and Bonferroni-corrected P-values for all the reported effects.

Local genetic effects on gene expression in blood are widespread and replicable in other tissues

We identified *cis*-eQTLs (SNP gene distance <1 Mb, FDR<0.05; **Methods**) for 16,987 unique genes (88.3% of autosomal genes expressed in blood and tested in *cis*-eQTL analysis; **Figure 1B**; 15,073 genes attained the more conservative Bonferroni threshold of 3.9×10^{-10}).

After we observed that *cis*-eQTLs replicated between whole blood datasets (**Supplementary Figure 1A**), we investigated the replicability of *cis*-eQTLs in other tissues. We considered an eQTL replicated when it was significant in the replication dataset (Benjamini-Hochberg FDR<0.05) and had the same allelic direction. In general, *cis*-eQTLs showed directional consistency across tissues. In 47 postmortem tissues[17], we observed an average replication rate of 14.8% (discovery analysis without GTEx, replication FDR<0.05 in GTEx; median 15.0%, range 3.6–29.6% when excluding whole blood) and, on average, a 94.9% concordance in allelic directions (median 95.2%, range 86.7–99.2%, when excluding whole blood) among the *cis*-eQTLs for which the lead SNP effect replicated in GTEx (**Supplementary Figure 4**, **Supplementary Information** and **Supplementary Table 3**).

Genes highly expressed in blood that did not have a detectable *cis*-eQTL effect were more likely (two-sided Wilcoxon rank sum test, $P=2\times10^{-6}$; **Figure 2A**) to be intolerant to loss-of-function mutations in their coding region[20], suggesting that eQTLs on such genes are selectively constrained, as has been recently proposed[21].

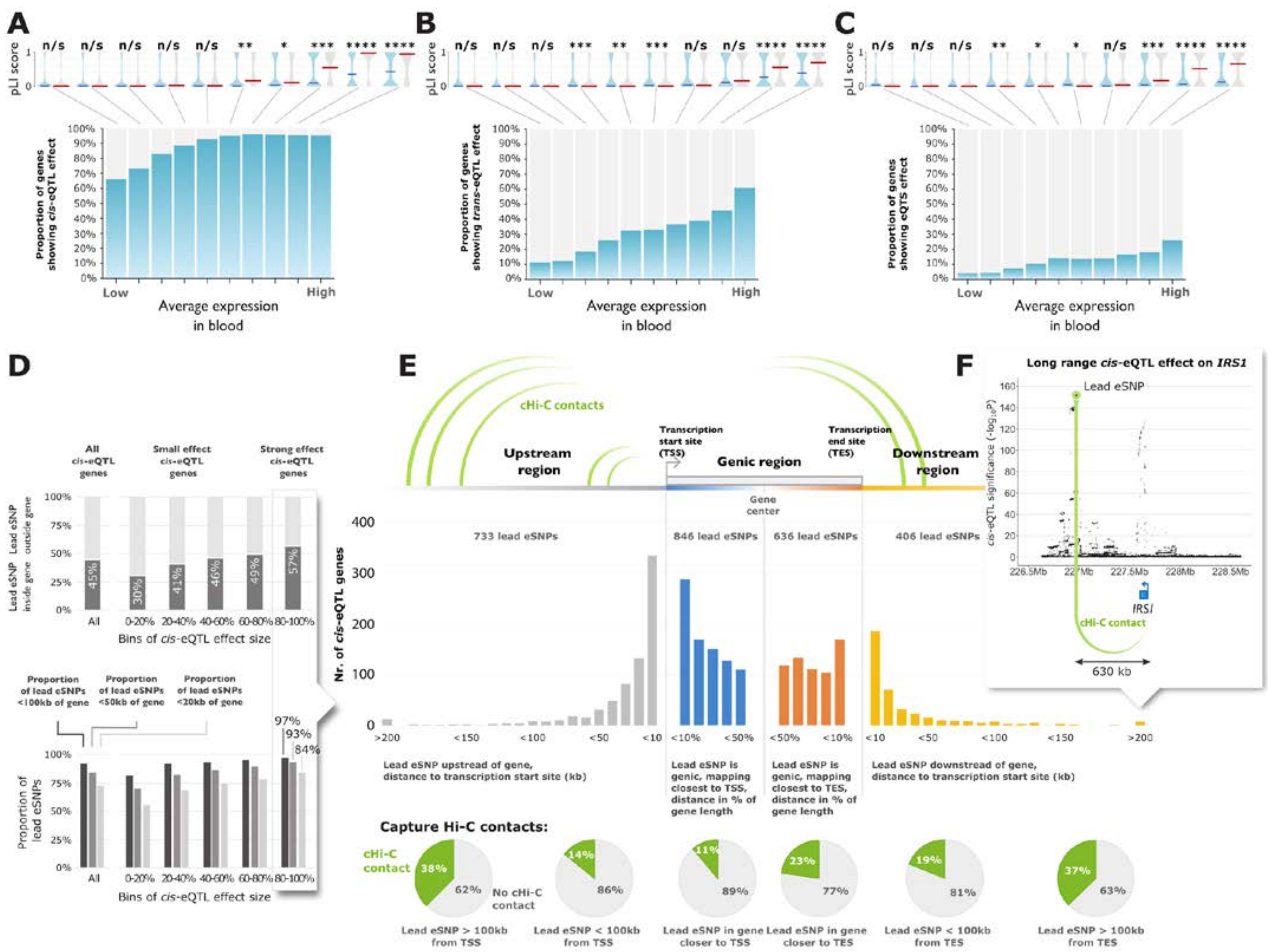


Figure 2. Results of the *cis*- and *trans*-eQTL analysis. All genes tested in (**A**) *cis*-eQTL analysis, (**B**) *trans*-eQTL analysis, and (**C**) eQTS analysis were divided into 10 bins based on their average expression levels in blood (BIOS Cohort). Highly expressed genes without any eQTL effect (grey bars) were less tolerant to loss-of-function variants (two-sided Wilcoxon rank sum test on pLI scores). Indicated are median pLIs per bin. n/s (not significant) P>0.05; * P<0.05; ** P<0.01; *** P<0.001; **** P<1×10^{-4}. (**D**) Genes with strong effect sizes are more likely to have a lead SNP fall within (top panel) or close to the gene (bottom panel) (**E**) Lead *cis*-eQTL SNPs overlap with capture Hi-C contacts with transcription start sites (TSS).

We observed that 92% of lead *cis*-eQTL SNPs were located within 100kb of the gene (**Figure 2D**) and that stronger *cis*-eQTL effects were more likely to map closeby (within 20kb for 84.1% of the top 20% strongest eQTLs).

The lead *cis*-eQTL SNPs which located >100kb from the transcription start site (TSS) or transcription end site (TES) of the *cis*-eQTL gene were more likely to overlap with capture Hi-C contacts than expected by chance (2.0-fold enrichment compared to when location of Hi-C target was flipped relative to the TSS; $P<3.3\times10^{-12}$; two-tailed two-sample test of equal proportions; **Methods**, **Figure 2E, Supplementary Results**). This suggests that some long-range *cis*-eQTLs are caused by physical interactions between the genomic regions of the SNP and gene. For example, a capture Hi-C contact for *IRS1* overlapped the lead eQTL SNP, mapping 630kb downstream from *IRS1* (**Figure 2F**). Similarly, we observed an enriched overlap with Hi-C contacts for short-range *cis*-eQTL effects (<100kb, 1.3-fold; $P<9.1\times10^{-16}$; two-tailed two-sample test of equal proportions; **Figure 2E, Supplementary Results**).

When comparing our results to the 5,440 protein-coding *cis*-eQTL genes that we had previously identified in 5,311 samples[5], the lead SNPs in the current study typically mapped closer to the *cis*-eQTL gene (**Supplementary Figure 5**). In GWAS studies, larger sample sizes and more dense imputation panels generally increase the resolution of signals in associated loci, especially for weaker effects. Additionally, GWAS simulations have indicated that lead GWAS signals generally map near the causal variant (within 33.5kb in 80% of cases)[22]. Since the majority of the *cis*-eQTL variants identified in our study map within 100kb of the TSS and TES, we consider it highly likely that causal variants affecting gene expression are also generally within these regions.

One third of trait-associated variants have *trans*-eQTL effects

An alternative strategy to gain insight into the molecular functional consequences of disease-associated genetic variants is to ascertain *trans*-eQTL effects. Due to the extensive computational burden that genome-wide *trans*-eQTL analyses would impose on participating cohorts, we constrained our analyses to a subset of variants that have previously been associated with complex phenotypes. We tested 10,317 trait-associated SNPs (GWAS $P \leq 5\times10^{-8}$; **Methods**, **Supplementary Table 2**) and identified 59,786 *trans*-eQTLs (SNP-gene distance >5 Mb; $P<8.3\times10^{-6}$, corresponding to an FDR<0.05; 17,395 *trans*-eQTLs were below the Bonferroni threshold of $P<2.4\times10^{-10}$; **Supplementary Table 4**, **Supplementary Figure 6**), representing 3,853 unique SNPs (37% of tested GWAS SNPs) and 6,298 unique genes (32% of tested genes; **Figure 1C**). The largest previous *trans*-eQTL meta-analysis in blood[5] (N=5,311) identified *trans*-eQTLs for only 8% of the trait-associated SNPs tested, indicating that a larger sample size is beneficial for the identification of distal effects. Similar to what we saw for *cis*-eQTLs, highly expressed genes without detectable *trans*-eQTL effects were more likely to be intolerant to loss-of-function variants (two-sided Wilcoxon rank sum test, $P=6.4\times10^{-7}$; **Figure 2B**), suggesting constrained expression of these genes.

While blood-cell-composition SNPs[23] comprised 21% of all the trait-associated SNPs tested, they represented the majority (64%) of *trans*-eQTL SNPs. This could be due to the fact that many of the identified *trans*-eQTL SNPs regulate the abundance of a specific blood cell type and could thus result in *trans*-eQTL effects on genes specifically expressed in that cell type. Although we corrected the individual expression datasets for cell-type composition effects using PCs (**Methods**, **Supplementary Note**), the fact that numerous *trans*-eQTLs emanate from known blood-cell-composition SNPs indicated that there was likely a residual effect of cell composition. We therefore aimed to distinguish *trans*-eQTLs caused by intracellular molecular mechanisms from eQTLs induced by blood cell type-composition.

To do so, we investigated a subset of up to 1,858 whole blood samples from the BIOS Consortium for which 49 measured and predicted blood cell metrics were available

(**Methods**, **Supplementary Note**). We first reasoned that if a *trans*-eQTL is intracellular (i.e. not driven solely by cell-type-composition), the main *trans*-eQTL effect should remain after correcting for cell-type-composition differences. We constructed a linear model incorporating all 49 available cell metrics (**Methods**) and tested whether a residual main effect remained for each *trans*-eQTL. We were able to test 55,311 *trans*-eQTLs in this subset (minor allele frequency (MAF) >0.05 in each BIOS cohort) and found that 4,241 (7.67%) were below the P-value threshold ($P<8.3\times10^{-6}$, threshold determined in discovery meta-analysis) in a linear model without any cell type metrics. Out of these, 2,952 (69.6% of 4,241 effects) *trans*-eQTLs remained below the significance threshold when all 49 cell metrics were included in the model (**Supplementary Figure 7; Supplementary Table 5**). Here we need to acknowledge that cell-type-composition may lead to false positive *trans*-eQTL effects, but we also note that large-scale cell count measures were not available for any of the included cohorts, which precluded us from drawing definite conclusions about this issue. We next reasoned that, if a *trans*-eQTL is generic (i.e. it has similar effect sizes within each individual cell type), the main *trans*-eQTL effect would also remain after correcting for cell-type-composition differences and their interactions with the *trans*-eQTL SNP. When we included all the interaction terms between cell-type metric and genetic variant in the model, only 33 (0.06%) out of 4,241 *trans*-eQTLs remained below the P-value threshold ($P<8.3\times10^{-6}$), suggesting that most *trans*-eQTLs have variable effect sizes in different blood cell types (**Methods; Supplementary Figure 7**). We also aimed to assign each of the *trans*-eQTLs to the cell type it most likely manifests in by testing the interaction between genotype and each cell metric (**Methods**). However, no individual interaction effects were below the FDR threshold (Benjamini-Hochberg FDR>0.05; smallest $P=1.37\times10^{-7}$; **Supplementary Table 6**), likely due to the extensive multiple testing burden and limited power.

Our replication analyses between different expression platforms suggest that *trans*-eQTLs are replicable between blood datasets (**Supplementary Figure 1B**) but cannot identify cell-type-composition effects. To estimate the fraction of *trans*-eQTLs that constitute intracellular *trans*-eQTLs, we performed replication analyses in bulk RNA-seq datasets derived from specific cell types: lymphoblastoid cell lines (LCL), induced pluripotent cells (iPSCs) and several purified blood cell types (CD4+, CD8+, CD14+, CD15+, CD19+, monocytes and platelets). Additionally, we used blood DNA methylation QTL data to support the validity of *trans*-eQTLs. In total, 4,018 (6.7% of the total) *trans*-eQTLs showed replication in at least one cell type (Benjamini-Hochberg FDR<0.05; 93.3% with same allelic direction, on average) or were supported by the methylation data (Benjamini-Hochberg FDR<0.05; meQTL effect direction supporting the discovery eQTL effect, see **Supplementary Information**, **Supplementary Figure 8**, **Supplementary Table 4**). We then investigated whether *trans*-eQTLs are shared across tissues from GTEx[16]. We repeated our discovery meta-analysis while excluding whole blood samples from GTEx, performed replication analyses in all GTEx tissues, and observed that the replication rate was very low (0.07% of *trans*-eQTLs replicated in any non-blood tissue, 0.09% in blood, Benjamini-Hochberg FDR<0.05). However, the allelic concordance of significant effects was, on average, 66% in non-blood tissues and 100% in blood (**Supplementary Table 4**). Despite these low replication rates, *trans*-eQTLs showed an inflation of replication signal in the majority of tissues (**Supplementary Figure 9A**), most notably in whole blood, esophagus muscularis, liver, heart atrial appendage and non-sun-exposed skin.

Ideally, replication of individual *trans*-eQTLs should be performed using single-cell (sc)RNA-seq eQTL datasets, since such datasets are less impacted by the cell-type-composition differences present in bulk eQTL datasets. Currently available scRNA-seq eQTL datasets are still relatively small, but by meta-analysing two different PBMC-based scRNA-seq cohorts using the 10X Chromium platform (OneK1K, N=982 and 1M-scBloodNL, N=157), we were able to perform *trans*-eQTL replication analysis in B-cells, CD4+ T-cells, CD8+ T-cells, classical monocytes, non-classical monocytes, dendritic cells, natural killer (NK) cells and plasma cells from up to 1,139 individuals (up to 3.6% of the discovery sample size, **Supplementary Note**). For each of the 59,786 discovery *trans*-eQTLs, we tested the association within each cell type, but only if the *trans*-eQTL gene was sufficiently expressed (i.e. had a missing sample fraction of at most 20% in the larger OneK1K dataset). We did this because the expression of only a few thousand genes per cell were quantified in scRNA-seq data.

Since scRNA-seq eQTL data is noisier than bulk RNA-seq data, fewer eQTLs can be identified when using the same number of samples[24]. Moreover, *trans*-eQTLs in eQTLGen were identified using 31,684 samples, while the single-cell replication cohort was limited to 1,139 individuals. Therefore, since the statistical power to formally replicate *trans*-eQTLs was limited, we first studied whether there was any inflation of replication test statistics. For 7 out of the 8 cell types examined, we observed inflation of signal (**Supplementary Table 7**, **Supplementary Figure 9A**; for the least abundant cell type, plasma cells (**Figure 3A**), no inflation of signal was observed) and greater than expected allelic concordance with the discovery analysis (**Figure 3A; Supplementary Table 7;** two-sided binomial test P<0.05). Similarly, by correlating the effect sizes of independent *trans*-eQTLs using the r_b method (**Methods**)[25], we observed that blood *trans*-eQTL effect sizes correlate significantly with replication effects in the scRNA-seq data (**Figure 3A**; two-sided P<0.05) for 4 out of 8 cell types (classical monocytes ($P=3.36\times10^{-8}$, $r_b=0.514$, S.E.=0.093), NK cells ($P=3.24\times10^{-4}$, $r_b=0.185$, S.E.=0.051), CD8+ lymphocytes ($P=3.41\times10^{-3}$, $r_b=0.454$, S.E.=0.155) and B cells ($P=5.98\times10^{-3}$, $r_b=0.049$, S.E.=0.018)).

More abundant cell types showed higher *trans*-eQTL effect size correlations with whole blood (**Figure 3A**, Pearson R^2=0.53, two-sided P=0.04). When conducting r_b analysis on the bulk expression profiles from purified blood cell types (**Supplementary Figure 10**; average r_b=0.55), we observed r_b metrics similar to scRNA-seq data for several cell types, demonstrating that there is concordance between scRNA-seq and bulk expression data from specific cell types.

These correlations and inflations of signal show that some of the *trans*-eQTLs identified in blood are also present in the cell types in our scRNA-seq data, although it remains challenging to prioritize individual effects. Still, we aimed to formally replicate individual *trans*-eQTLs. Depending on the cell type, we could reliably test between 1,917 and 27,582 of the *trans*-eQTLs identified in the discovery analysis (**Figure 3A**). We replicated 35 *trans*-eQTLs at FDR<0.05 (**Supplementary Table 8**), with two effects appearing in more than one cell type. For *trans*-eQTLs which replicated, the allelic concordance between the discovery and the replication analysis was very high (97% concordance), providing additional support for valid replication of these eQTLs.

Lastly, to increase the statistical power to replicate individual *trans*-eQTLs in the noisy scRNA-seq data, we combined the summary statistics from 8 cell types by averaging the Z-scores per *trans*-eQTL over the available cell types. When confining the analysis to the 729 *trans*-eQTLs with an absolute average Z>1.96 (corresponding to a nominal P<0.05, **Supplementary Table 8**), we observed a relatively high concordance of 84% (**Figure 3A**, **Supplementary Table 7**, two-sided binomial test; $P=1.25\times10^{-84}$) suggesting that many of these *trans*-eQTLs represent effects that are independent of cell-type-composition. Among the 729 *trans*-eQTLs, we observed a strong enrichment for genes involved in cytokine-mediated signalling (hypergeometric test from ToppGene[26], $P=3.3\times10^{-12}$, Benjamini-Hochberg FDR<0.05).

The *trans*-eQTL effect sizes we observed are generally small (median r=0.033; **Supplementary Figure 24E**, **Supplementary Results**). Considering the small sample sizes of the bulk and scRNA-seq *trans*-eQTL datasets available for replication, statistical power to replicate these effects was low. Consequently, this most likely limited our ability to replicate individual *trans*-eQTLs, and our ability to reliably distinguish cell-type-composition effects from intracellular effects. We did observe that ~70% of the *trans*-eQTLs remain significant after correcting for all available cell metrics, and that the highest significant correlation of effects (i.e. r_b correlation) was 0.5 (classical monocytes) in the scRNA-seq replication data. We hope that large-scale single-cell eQTL studies will attain more statistical power in the near future, such that we can more reliably differentiate intracellular *trans*-eQTLs from those driven by cell composition. Because the replication results in external datasets did not enable such a distinction, we decided to use all *trans*-eQTLs for the following interpretive analyses.

To evaluate which *trans*-eQTL SNPs also have *cis*-eQTL effects, we conducted locus-wide *trans*-eQTL analyses in a subset of samples (N=4,339; EGCUT and BIOS cohorts; **Supplementary Figure 11**; **Supplementary Methods**). For this analysis, we focused on *trans*-eQTLs identified in the discovery meta-analysis. We extracted the *trans*-eQTL SNPs that showed significant effect in this subset of samples ($P<8.3\times10^{-6}$; P-value threshold estimated using discovery *trans*-eQTL meta-analysis) and constructed 12,911 *trans*-eQTL loci (±1 Mb from tested GWAS SNP) (**Methods, Supplementary Figure 11**). We then performed conditional *trans*-eQTL analyses to identify independent lead *trans*-eQTL SNPs for each locus (**Supplementary Table 9**). For each of these lead *trans*-eQTL SNPs, we then calculated linkage disequilibrium (LD) with lead *cis*-eQTL SNPs identified in the discovery meta-analysis. Out of 12,911 *trans*-eQTL loci, 3,786 (29.3%) were in LD with at least one lead *cis*-eQTL SNP (R^2>0.8 between *cis*-eQTL and *trans*-eQTL lead

SNPs, 1kG p1v3 EUR, **Supplementary Table 10**). Since the discovery *cis*-eQTL and *trans*-eQTL analyses were performed in the same set of samples, we note that this estimated proportion might be somewhat biased. However, corresponding *cis*-eQTL genes were strongly enriched for having transcription factor (TF) activity ("RNA polymerase II regulatory region sequence-specific DNA binding (GO:0000977)"; one-sided Fisher's exact test $P=9.15\times10^{-6}$, Benjamini-Hochberg FDR=0.043; **Supplementary Figure 12**).

These LD-based lead-SNP-overlap analyses identify loci where two association signals likely overlap. We next formally tested whether local genes within 100kb of the *trans*-eQTL SNP affect the expression of the *trans*-eQTL gene, limiting the analysis to non-HLA *trans*-eQTLs detected in the discovery meta-analysis. We used a subset of 4,339 samples from the BIOS and EGCUT cohorts and included the local gene in a linear model as a gene-environment (G × E) interaction term. We considered *trans*-eQTLs with a Benjamini-Hochberg FDR<0.05 for an interaction term to be driven by the expression of a *cis*-acting gene. We observed interaction effects for 615 out of 201,106 SNP–*cis*–*trans*–gene combinations tested (**Supplementary Table 11**), reflecting 585 *trans*-eQTLs. For instance, for rs7045087 (associated to red blood cell counts[23]), we observed that the expression of the interferon gene *DDX58* (mapping 38bp downstream from rs7045087) interacted with *trans*-eQTL effects on *HERC5, OAS1, OAS3, MX1, IFIT1, IFIT2, IFIT5, IFI44, IFI44L, RSAD2* and *SAMD9* (**Supplementary Figure 13**), most of which are known to be in involved in interferon signaling. These results indicate that *trans*-eQTL effects can be affected by the expression of local genes, but comprehensive characterization of such interaction effects requires larger sample sizes.

We then conducted enrichment analyses to evaluate which biological mechanisms might lead to *trans*-eQTLs (**Supplementary Methods, Supplementary Results, Supplementary Figure 14, Figure 3B**). The most intuitive interpretation of how a *trans*-eQTL might arise is that a SNP affects the gene expression of a nearby TF, which leads to up- or downregulation of its target genes. To test if our *trans*-eQTLs adhere to this mechanism, we overlapped our results with known TF–target gene pairs in blood cell lines[27] (**Supplementary Methods, Supplementary Results)** and found that pairs of *cis*- and *trans*-eQTL genes emerging from the same SNP were 1.28-fold enriched in TF target genes as compared to all other gene pairs tested in eQTLGen ($P=4.0\times10^{-21}$; two-sided Fisher's exact test; **Supplementary Figure 14**). This limited enrichment could be due to different molecular mechanisms involved in *cis*- versus *trans*-regulation, or mechanisms not directly involving TFs. To investigate this further, we reasoned that, even if we did not observe a *cis*-eQTL, a *trans*-eQTL SNP would usually act via a gene located near the *trans*-eQTL SNP. For this reason, we linked the *trans*-eQTL SNPs to nearby genes using the Pascal method[28] (**Supplementary Figure 15, Supplementary Methods**), which allowed us to calculate a score representing how likely it is that a local gene is an intermediate of a *trans*-eQTL effect. We connected local genes to distal *trans*-eQTL genes and, using these local–distal gene pairs, performed several enrichment analyses to reveal mechanisms that can result in *trans*-eQTL effects. Using this procedure, we observed a 1.40-fold enrichment for TFs (**Figure 3B**). Interestingly, there was also a clear enrichment when we tested genes co-regulated with known TFs (1.38-fold, $P=5.8\times10^{-72}$; two-sided Fisher's exact test; **Figure 3B**), genes co-regulated with known target genes (3.57-fold, $P<1.0\times10^{-308}$; two-sided Fisher's exact test; **Figure 3B**), and genes co-regulated with both (4.37-fold, $P<1.0\times10^{-308}$; two-sided Fisher's exact test; **Figure 3B**). This suggests that many *trans*-eQTL genes are not direct TF targets themselves, but might still represent an indirect consequence of transcriptional regulation. Additionally, we observed a strong 22.3-fold enrichment ($P<1.0\times10^{-308}$; two-sided Fisher's exact test) of co-regulated gene pairs and a 1.45-fold enrichment of protein–protein interaction (PPI)[29] pairs ($P=3.5\times10^{-17}$; two-sided Fisher's exact test), including co-regulated subunits of the same protein complex (e.g. *CPSF1* and *CPSF7*) and receptor-ligand pairs (e.g. *CSF3* and *CSF3R*). We note that cell-type composition effects are likely to contribute to this high enrichment of co-regulated gene pairs: we observed that co-regulated gene pairs were depleted by *trans*-eQTL effects replicating in scRNA-seq data (OR=0.5, P=0.015, two-sided Fisher's exact test). However, for 11 *trans*-eQTLs we observed both *cis*-*trans* co-regulation and nominal replication in scRNA-seq. These pairs included celiac disease-associated rs6498114, affecting *CIITA* which is co-regulated with genes *CD74* and *HLA-DMB*. It has been previously described that CIITA affects both CD74 and HLA-DMB[30]. We also observed an enrichment of Hi-C chromatin contacts[31] among local–distal gene pairs across and within chromosomes (OR=1.47; $P=2.4\times10^{-153}$; two-sided Fisher's exact test), suggesting that some *trans*-eQTLs are driven by physical contact (**Supplementary Figure 14**). When we combined all potential mechanisms, 30,579 (51%) of the reported *trans*-eQTLs could be assigned a putative biological mechanism, i.e. these *trans*-eQTLs could be driven by TF activity, PPI, or co-regulation patterns (**Figure 3C, Supplementary Table 12**). While the enrichment of some of these mechanisms (like co-regulation) may also be a result of the cell-type-composition in blood, the observed enrichment among known TF-target pairs supports the validity for a subset of our *trans*-eQTLs. Finally, we note that, since our *trans*-eQTL analysis was limited to trait-associated variants, the enrichment results presented here may not reflect *trans*-eQTLs for genetic variants that have not been associated with complex traits or diseases.

Despite these enrichments, most individual blood *trans*-eQTL effects remain unexplained. We have made all *trans*-eQTLs publicly available (irrespective of their statistical significance) to facilitate follow-up research into the regulatory mechanisms of trait-associated SNPs. In the **Supplementary Results**, we highlight examples involving *trans*-eQTL variants previously associated with age of menarche[32] (*ZNF131* locus), lipid levels[33] (*FADS1/2* locus), IBD[34] and SLE[35] (*IFIH1* locus), asthma[36] (*GSDMB* locus), and height[37] (*CLOCK* locus), and explore their potential biological mechanisms to show how this resource can serve as a starting point to generate hypotheses for further research (**Supplementary Figure 16A-E**).

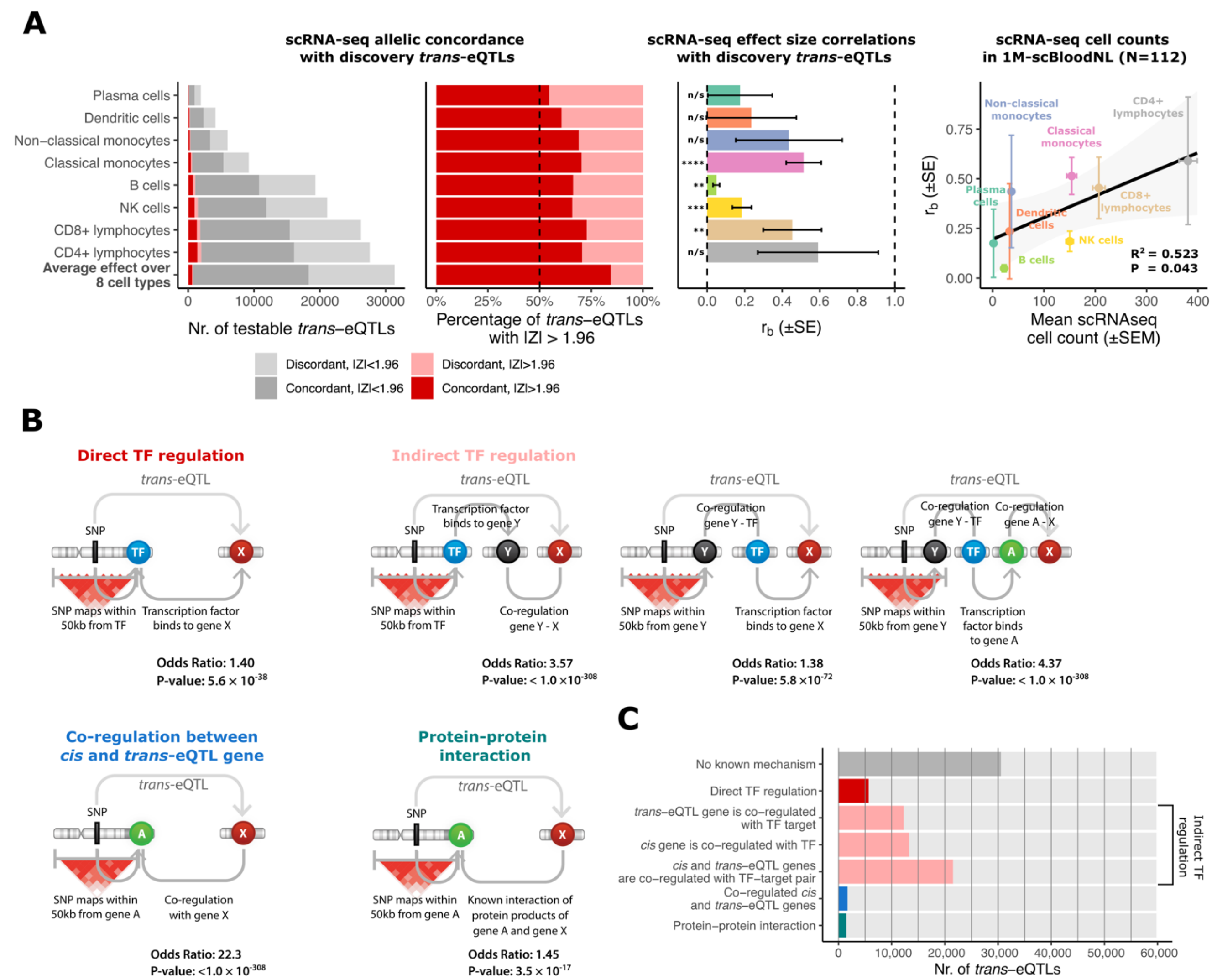


Figure 3. *Trans*-eQTL replication in scRNA-seq cell types and mechanisms leading to *trans*-eQTLs. (A) Replication analyses in scRNA-seq of 8 cell types in up to 1,139 unrelated individuals. Left panels: allelic concordances relative to *trans*-eQTL effect direction in the discovery *trans*-eQTL analysis. Middle panel: correlation estimates (r_b) of *trans*-eQTL effects between the discovery analysis in blood and scRNA-seq blood cell types. A subset of independent *trans*-eQTL effects was used to calculate r_b estimates (**Methods**). n/s P>0.05; * P<0.05; ** P<0.01; *** P<0.001; **** P<1×10^{-4}. Right panel: correlation between cell-type counts for each cell type in a subset of samples from the 1M-scBloodNL cohort (N=112) and the r_b estimates. Values shown are the squared Pearson correlation coefficient and the two-sided P-value from the Pearson correlation test. (**B**) Enrichment analyses for known transcription factor (TF)

associations, gene co-regulation and protein–protein interactions (PPIs). *Cis*-acting genes were determined by *cis*-eQTLs or assigned by the Pascal method (**Methods**, **Supplementary Methods**). Enrichment analyses were conducted using the two-sided Fisher's exact test. (**C**) All 59,786 *trans*-eQTLs stratified by putative mechanism of action. Hi-C enrichment results are not shown as we only observed enrichment when using a lenient (>0) threshold for Hi-C contacts. Full results are shown in **Supplementary Figure 13.**

Next, for each GWAS phenotype, we interrogated whether *trans*-eQTL genes were enriched for Gene Ontology (GO) terms. In total, we observed 347 enriched GO terms for 208 out of 345 (60%) traits (one-sided Fisher's exact test, Benjamini-Hochberg FDR<0.05; **Supplementary Table 13**). We observed that several of the enriched GO terms were relevant for the tested trait. For example, *trans*-eQTL SNPs associated with celiac disease and inflammatory bowel disease showed the strongest enrichments for GO terms associated with response to cytokine stimulus (e.g. celiac disease: "cellular response to cytokine stimulus", FDR=1.06×10^{-5}), platelet count was enriched for "platelet degranulation" (FDR=2.6×10^{-10}), and heart rhythm traits were most enriched for cholesterol-related terms (e.g P-wave duration was enriched for "regulation of cholesterol biosynthetic process", FDR=4.6×10^{-14}).

As discussed above, one putative mechanism driving *trans*-eQTLs could be the action of a TF regulated in *cis*, resulting in many potential *trans*-eQTL effects. We reasoned that such a variant would act as a master regulator: a *'hub'* SNP. Indeed, we identified 1,050 (10.2%) '*hub'* SNPs that regulated the expression of >10 genes (**Supplementary Table 14**). Of these, 196 (18.6%) had a global up- or down-regulating effect on the expression levels of downstream genes (two-sided binomial test, Bonferroni-corrected P<0.05, **Supplementary Table 14**). We identified 507 (48%) '*hub'* SNPs showing enrichment for TF- or miRNA-binding sites (one-sided Fisher's exact test, Benjamini-Hochberg

FDR<0.05; **Supplementary Table 15**) and observed that the respective TF was encoded by a gene positioned <1 Mb from the *'hub'* SNP for 9 of these (5 independent loci), which supports a mechanism of TF binding.

For example, rs17087335 (which is associated with coronary artery disease[38]) affects the expression of 88 genes in *trans* (FDR<0.05, Bonferroni corrected P<0.05 for 39 genes; **Figure 4, Supplementary Table 16**) that are highly expressed in brain (one-sided Fisher's exact test, ARCHS4 database, Benjamini-Hochberg FDR=6.43×10^{-14}; **Figure 4**). Eighty-five out of the 88 (96.6%) *trans*-eQTL genes were upregulated by the minor allele of rs17087335 and strongly enriched for the targets of REST (RE-1 silencing transcription factor; one-sided Fisher's exact test for ENCODE[39,40] project REST ChIP-seq, Benjamini-Hochberg FDR=8.84×10^{-38}, **Figure 4**). While the minor allele of rs17087335 was associated with lower expression of *REST*, it was not in LD (R^2<0.2, 1kG p1v3 EUR) with the lead *cis*-eQTL SNP (rs13353552). A SNP in high LD with rs17087335, rs3796529 (R^2=0.91, 1kG p1v3 EUR), is a missense variant for REST, suggesting that these *trans*-eQTLs could also arise from a post-transcriptional mechanism of action. Because REST is a TF that downregulates the expression of neuronal genes in non-neuronal tissues[41,42], we speculate that the observed *trans*-eQTLs reflect the impact of genetic variation on the effectiveness of downregulation, although experimental follow-up is required to confirm this hypothesis. Nevertheless, this example illustrates that blood *trans*-eQTL effects can help to prioritize the putatively causal *cis*-eQTL gene among multiple genes in a locus (here *REST*).

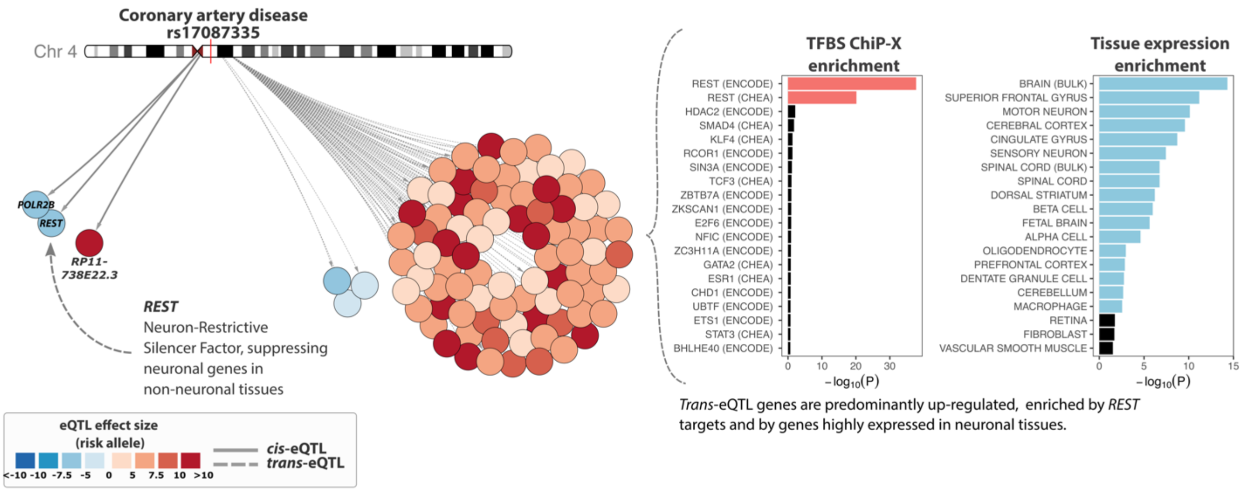
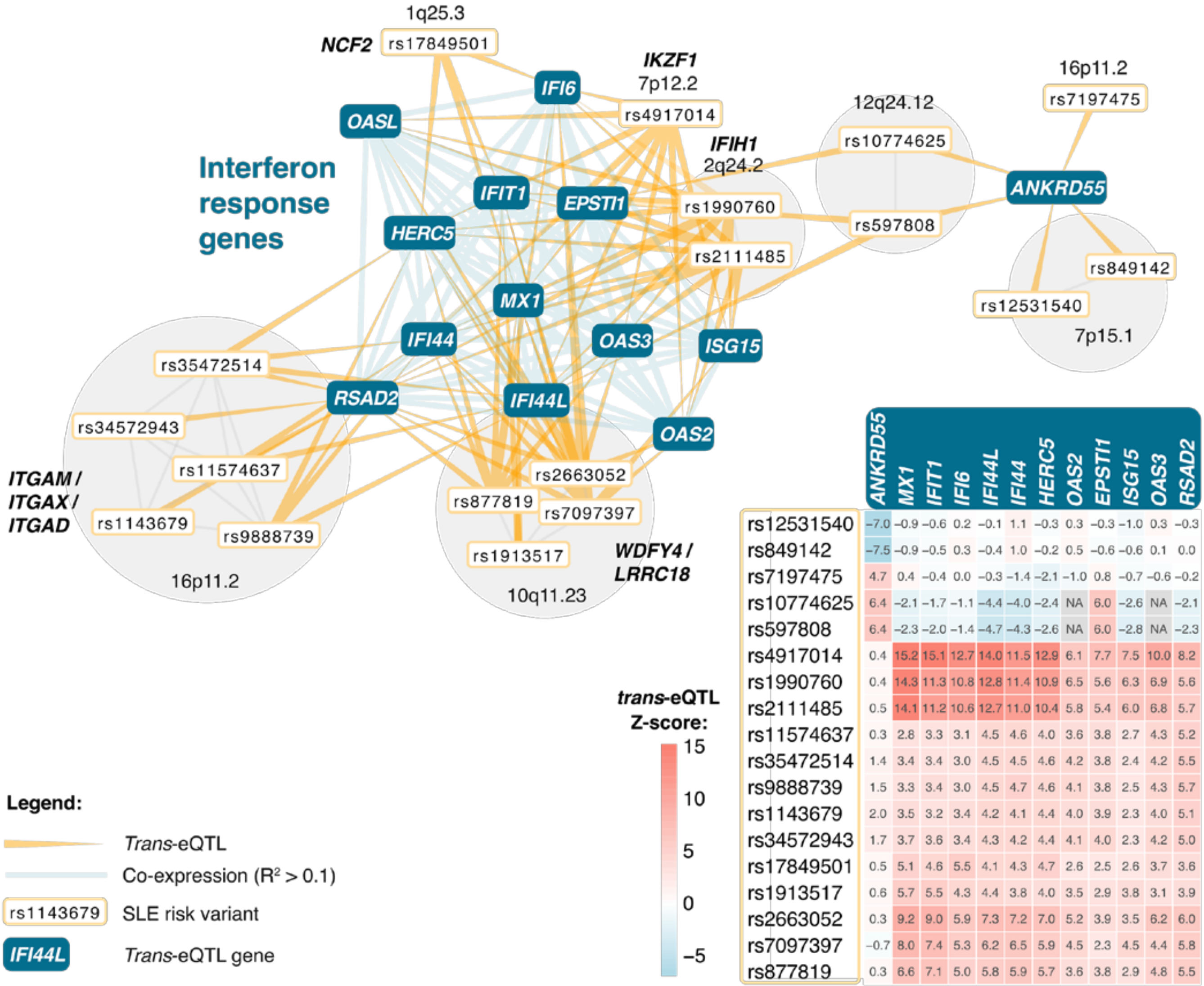


Figure 4. *REST* locus regulates the expression of 88 *trans*-eQTL genes. *Trans*-eQTL genes for the *REST* locus are highly enriched for REST transcription factor targets and for expression of neuronal genes.

Next, we investigated whether *trans*-eQTLs can also identify genes relevant to the biology of the corresponding complex trait. We grouped the *trans*-eQTL SNPs by GWAS trait and tested whether unlinked trait-associated variants showed *trans*-eQTL effects on the same gene. This revealed 47 different traits for which at least four independent variants affected the same gene in *trans* (**Supplementary Table 17)**, which is 3.4-times higher than expected by chance (P=0.001; two-tailed two-sample test of equal proportions). For systemic lupus erythematosus (SLE)[43], the gene expression levels of *IFIT1, IFI44L*, *HERC5*, *IFI6*, *IFI44*, *RSAD2*, *MX1*, *ISG15*, *ANKRD55*, *OAS3*, *OAS2*, *OASL* and *EPSTI1* were affected by at least three SLE-associated genetic variants (FDR<0.05, all genes except *OAS2* also had at least one *trans*-eQTL that reached Bonferroni significance). These genes include nearly all known interferon genes in the well-described SLE interferon signature[44–46] (**Supplementary Table 18**), reflecting the involvement of interferon signaling as a key component of SLE pathophysiology (**Figure 5**). While our *trans*-eQTL analysis did not identify novel interferon signature genes, it helped to pinpoint SLE GWAS loci that collectively affect SLE interferon signature genes.



	ANKRD55	*MX1*	*IFIT1*	*IFI6*	*IFI44L*	*IFI44*	*HERC5*	*OAS2*	*EPSTI1*	*ISG15*	*OAS3*	*RSAD2*
rs12531540	−7.0	−0.9	−0.6	0.2	−0.1	1.1	−0.3	0.3	−0.3	−1.0	0.3	−0.3
rs849142	−7.5	−0.9	−0.5	0.3	−0.4	1.0	−0.2	0.5	−0.6	−0.6	0.1	0.0
rs7197475	4.7	0.4	−0.4	0.0	−0.3	−1.4	−2.1	−1.0	0.8	−0.7	−0.6	−0.2
rs10774625	6.4	−2.1	−1.7	−1.1	−4.4	−4.0	−2.4	NA	6.0	−2.6	NA	−2.1
rs597808	6.4	−2.3	−2.0	−1.4	−4.7	−4.3	−2.6	NA	6.0	−2.8	NA	−2.3
rs4917014	0.4	15.2	15.1	12.7	14.0	11.5	12.9	6.1	7.7	7.5	10.0	8.2
rs1990760	0.4	14.3	11.3	10.8	12.8	11.4	10.9	6.5	5.6	6.3	6.9	5.6
rs2111485	0.5	14.1	11.2	10.6	12.7	11.0	10.4	5.8	5.4	6.0	6.8	5.7
rs11574637	0.3	2.8	3.3	3.1	4.5	4.6	4.0	3.6	3.8	2.7	4.3	5.2
rs35472514	1.4	3.4	3.4	3.0	4.5	4.5	4.6	4.2	3.8	2.4	4.2	5.5
rs9888739	1.5	3.3	3.4	3.0	4.5	4.7	4.6	4.1	3.8	2.5	4.3	5.7
rs1143679	2.0	3.5	3.2	3.4	4.2	4.1	4.4	4.0	3.9	2.3	4.0	5.1
rs34572943	1.7	3.7	3.6	3.4	4.3	4.2	4.4	4.1	4.0	2.3	4.2	5.0
rs17849501	0.5	5.1	4.6	5.5	4.1	4.3	4.7	2.6	2.5	2.6	3.7	3.6
rs1913517	0.6	5.7	5.5	4.3	4.4	3.8	4.0	3.5	2.9	3.8	3.1	3.9
rs2663052	0.3	9.2	9.0	5.9	7.3	7.2	7.0	5.2	3.9	3.5	6.2	6.0
rs7097397	−0.7	8.0	7.4	5.3	6.2	6.5	5.9	4.5	2.3	4.5	4.4	5.8
rs877819	0.3	6.6	7.1	5.0	5.8	5.9	5.7	3.6	3.8	2.9	4.8	5.5

Figure 5. SNPs associated with systemic lupus erythematosus (SLE) converge on a shared cluster of interferon-response genes. The genes shown are those affected by at least three independent GWAS SNPs. SNPs in the HLA region are not visualised and SNPs in partial LD are grouped together. The heatmap indicates the direction and strength of individual *trans*-eQTL effects (Z-scores), relative to the SLE risk allele.

Very recently, Vuckovic *et al.*[47] used our *trans*-eQTL data to interpret SNPs that affect blood-cell traits and observed that *trans*-eQTL genes are strongly enriched for genes known to cause stem cell and myeloid disorders; bleeding, thrombotic and platelet disorders; and bone-marrow failure syndromes, a finding that underscores the value of using *trans*-eQTLs to identify trait-relevant genes. To more comprehensively query for the genes affected by several trait-associated loci, we next systematically investigated the relationships between PGSs and gene expression.

eQTSs identify potential key driver genes for polygenic traits

To ascertain the coordinated effects of trait-associated variants on gene expression, we used GWAS summary statistics to calculate PGSs for 1,263 traits in 28,158 samples (**Methods, Supplementary Table 19**). We reasoned that when the PGS for a specific trait correlates with the expression levels of a gene, the *trans*-eQTL effects of the individual risk variants (**Figure 6A**) converge on that gene, and it can be prioritised as a putative driver of the disease (**Figure 6B**).

Our meta-analysis identified 18,210 eQTS effects (FDR<0.05) representing 689 unique traits (55% of tested traits) and 2,568 unique genes (13% of tested genes; 285 traits and 905 genes were Bonferroni significant; **Supplementary Table 20**, **Figure 1D**). Of these genes, 719 (28%) were not identified in the *trans*-eQTL analysis, emphasizing the added value of analyzing eQTS in addition to *trans*-eQTLs (**Figure 6A-B**). We observed that median eQTS effect sizes were smaller than *cis*-eQTL effect sizes and similar to *trans*-eQTLs (**Supplementary Figure 24A, E, I**).

When calculating PGSs, the P-value threshold for including the SNPs that corresponds to most explained variation is likely to be trait-dependent. We therefore calculated PGSs using clumped GWAS lead SNPs at five significance levels ($P<0.01$; 1×10^{-3}; 1×10^{-4}; 1×10^{-5}; 5×10^{-8}). While we could detect the majority of eQTSs (70.5%) at the most conservative threshold ($P<5\times10^{-8}$), the total number of results was higher than for each P-value threshold separately (**Supplementary Table 21**), suggesting that our analysis captured different genetic architectures. Unsurprisingly, we identified more eQTSs for GWAS with larger sample sizes (Spearman r=0.42–0.59 at different P-value cut-offs). Traits with few eQTS associations typically also had lower average (Spearman r=0.42–0.72) and maximum eQTS effect sizes (Spearman r=0.69–0.85; **Supplementary Table 22**).

As in the previous analyses, the cross-platform replication rates showed high allelic concordance between blood datasets (average concordance rate was 99.2% for effects reaching FDR<0.05 in replication dataset, **Supplementary Figure 1C**), although the replication rates were quite low in the platforms with fewer samples (21.35-26.4% of tested effects reached FDR<0.05 in 1,549 FHS samples, **Supplementary Figure 1C**). We next ascertained to what extent eQTS associations can be replicated in independent datasets by studying 1,460 LCL samples, 762 iPSC samples and all GTEx tissues[17]. We were able to replicate 10 eQTSs in the LCL dataset, and 9 out of 10 (FDR<0.05) had the same effect direction as in the discovery dataset (**Supplementary Figure 17A, Supplementary Table 20**). Seventy-eight eQTSs replicated in the iPSCs dataset (FDR<0.05), with 71 (91%) showing the same direction of effect (**Supplementary Figure 17B, Supplementary Table 20**). Since polygenic risk scores can differ substantially between populations, we performed GTEx replication analyses while confining ourselves to Europeans and identified 19 replicating eQTSs with FDR<0.05 and same direction of effect (eQTS discovery performed without GTEx; 66 replicated when also including non-European samples, **Supplementary Table 20**). We observed the inflation of replication signal in some tissues, primarily in blood (**Supplementary Figure 9B**). Because only a few eQTS associations were replicated, there was no strong replication signal in non-blood tissues, and the majority of identified eQTS associations were observed for blood-related traits, we speculate that these effects are highly tissue- or cell-type-specific. However, as suggested by the power analyses, the limited replication in other tissues could also be a result of the small effect size of eQTS effects (median r=0.037;

Supplementary Figure 24I) causing a lack of statistical power in the replication datasets due to their small sample size, or because of variability in PGS estimates caused by differences in sample characteristics (e.g. age, sex, socio-economic status, etc) of the included datasets[48].

Similar to our analysis of *trans*-eQTLs, we next investigated whether eQTS could be driven by interindividual differences in cell-type-composition. We fitted linear models with and without cell-type metrics as covariates in a subset of 1,858 samples (**Methods**). Out of 18,210 eQTSs, 2,313 (12.7%) were below the P-value threshold in the original model ($P<3.02\times10^{-6}$, threshold determined by discovery meta-analysis). When all 49 cell metrics were included, 618 (3.39%) out of 2,313 eQTSs remained below the P-value threshold (**Supplementary Table 24, Supplementary Figure 7**). Twenty-one (3.4%, affecting 7 genes) replicated in at least one of our replication datasets. However, the majority of replicating effects originated from PGSs of erythrocyte- and platelet-related GWAS traits, while also affecting several blood-related genes such as *HBG1* and *HBG2*. This suggests that some strong cell-type-composition effects might still be detectable after correcting the data for all main effects. When including all interaction terms between cell-type metric and PGS, only two eQTSs (0.01%) remained below the P-value threshold ($P<3.02\times10^{-6}$), demonstrating the cell-type-specific nature of eQTSs. In line with the *trans*-eQTL effects, none of the eQTS effects could be reliably assigned to any of the cell-type metrics when testing individual PRS-cell metric interaction effects (**Methods**, Benjamini-Hochberg FDR>0.05; smallest $P=1.31\times10^{-6}$; **Supplementary Table 25**).

As expected based on the replication results, most eQTS associations (72.8%) represented blood-cell traits (**Supplementary Figure 18, Supplementary Table 20**). For instance, the PGS for mean corpuscular volume[49] correlated positively with the expression levels of genes specifically expressed in erythrocytes, e.g. genes coding for hemoglobin subunits (*HBG1*, FDR<0.05, smallest Bonferroni-corrected P=2.7×10^{-38} and *HBG2* FDR<0.05, smallest Bonferroni-corrected P=1.14×10^{-28}). eQTS genes were most strongly enriched by GO terms involved in cellular secretion, blood cell traits and intercellular signalling (**Supplementary Table 23**). There was no enrichment for TFs from the FANTOM5[50] database (one-sided Fisher exact test, P>0.05). Moreover, we observed a smaller number of InBio[29] PPIs for eQTS genes as compared to non-eQTS genes (two-sided Wilcoxon rank sum test, P=1.98×10^{-5}; median over eQTS genes 20, median over non-eQTS genes 25; **Supplementary Methods, Supplementary Results**). This suggests that transcriptional regulation and PPIs are not the main mechanisms by which eQTS genes convey their effect on the phenotype. When stratifying eQTS effects by GWAS phenotype, we identified 90 phenotypes showing enrichment with any GO term, and these often reflected known biology (one-sided Fisher's exact test, Benjamini-Hochberg FDR<0.05; **Supplementary Table 23**). For instance, platelet count showed the strongest enrichment for the process "platelet degranulation" (FDR=6×10^{-17}), monocyte count for "neutrophil degranulation" (FDR=4.7×10^{-16}) and total lipids in large HDL for "cholesterol metabolic process" (FDR=1.6×10^{-6}).

We expect that any eQTS analysis would yield the most informative genes if conducted in the trait-relevant tissue type. Still, in our blood data, we also identified eQTS associations for non-blood PGS, including with metabolite- and lipid-levels, anthropometric traits and several diseases such as asthma, celiac disease and coronary artery disease (**Supplementary Results; Supplementary Figure 19A-C; Supplementary Table 20**).

For example, 11 out of the 26 eQTS genes that were associated with the PGS for high-density lipoprotein levels (HDL[51,52]; FDR<0.05; 11 out of 26 were Bonferroni significant; **Figure 6C**) have previously been linked to lipid or cholesterol metabolism

(**Supplementary Table 26**). *ABCA1* and *ABCG1*, which positively correlated with the PGS for high HDL (r=0.05-0.07 for both genes, r derived from Z-score, both Bonferroni significant), mediate the efflux of cholesterol from macrophage foam cells and participate in HDL formation. In macrophages, downregulation of both *ABCA1* and *ABCG1* reduces reverse cholesterol transport into the liver by HDL[53] (**Figure 6D**). The PGS for high HDL was also negatively correlated with the expression of the low-density lipoprotein receptor *LDLR* (strongest eQTS $P=3.35\times10^{-20}$, r=0.06, r derived from Z-score), which is known to cause hypercholesterolemia[54]. Similarly, *SREBF2,* the gene encoding the TF SREBP-2, which is known to increase the expression of *LDLR*, was downregulated (strongest eQTS $P=3.08\times10^{-7}$ (not Bonferroni significant), r=0.03, r derived from Z-score). The negative correlation between *SREBF2* expression and measured HDL levels has been described before[15], indicating that the eQTS reflects an association with an actual phenotype. Zhernakova *et al*.[15] proposed a model where down-regulation of *SREBF2* results in lower expression of its target gene, *FADS2*. However, we did not observe an HDL eQTS effect on *FADS2* (all eQTS P>0.07), possibly because the indirect effect was too small to detect. We hypothesize that higher blood HDL levels can result in stronger reverse cholesterol transport into the liver, which may result in downregulation of *LDLR*[55]*.*

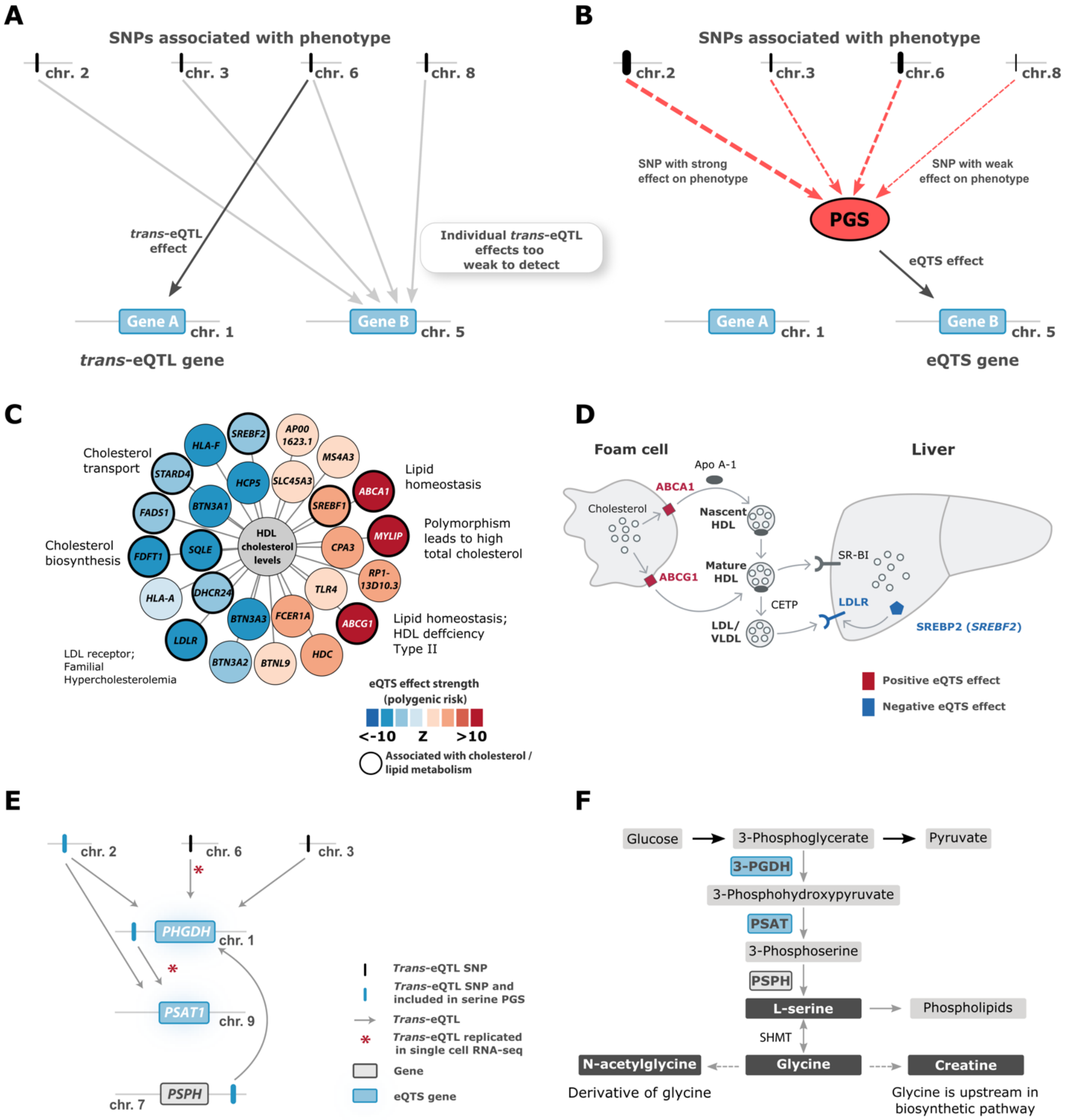


Figure 6. eQTS analyses. (**A**) In *trans*-eQTL analysis, individual SNPs are associated with gene expression. (**B**) In eQTS analysis, the effect sizes and directions of individual trait-associated SNPs are combined into a polygenic score (PGS) that is associated with gene expression. Here, we outline the case where eQTS analysis identifies a gene not detectable in the *trans*-eQTL analysis. Other scenarios we observed include: Gene A also being identified by eQTS analysis, Gene B being identified by both methods, or the combined effect of PGS yielding no significant eQTS. (**C**) The PGS for high density lipoprotein (HDL) associates to lipid metabolism genes. (**D**)

The role of *ABCA1*, *ABCG1*, *LDLR* and *SREBF2* in cholesterol transport. (**E**) Both *trans*-eQTLs and the serine PGS associate with the known serine biosynthesis genes *PHGDH* and *PSAT1*. (**F**) Serine biosynthesis pathway.

eQTS can also identify pathways known to be associated with monogenic diseases**.** For example, the PGSs for serine, glycine, the glycine derivative N-acetylglycine and creatine[56,57] were all negatively associated with the gene expression levels of *PHGDH*, *PSAT1* and *AARS* ($P<5.3\times10^{-7}$, $-0.05<r<-0.03$; $-0.08<r<-0.03$; $-0.05<r<-0.03$, respectively, r derived from Z-score, not Bonferroni significant). The PGSs for these traits are driven by SNPs near *CPS1* (2q34), *PHGDH* (1p12) and *PSPH* (7p11.2) (**Supplementary Table 27**) that influence expression of *PHGDH* and *PSAT1* in *trans*. We nominally replicated these *trans*-eQTLs in scRNA-seq data (average $|Z|>1.96$ across tested cell types, part of the 729 *trans*-eQTLs; **Supplementary Table 8, Figure 6E**), indicating that this eQTS is indeed driven by multiple genetic loci, but independent of cell-type composition. *PHGDH* and *PSAT1* encode crucial enzymes that regulate the synthesis of serine and, in turn, glycine[58], while N-acetylglycine and creatine form downstream of glycine[59] (**Figure 6F**). Mutations in *PSAT1* and *PHGDH* can result in serine biosynthesis defects including phosphoserine aminotransferase deficiency[60], phosphoglycerate dehydrogenase deficiency[61], and Neu-Laxova syndrome[62], which are all characterized by low concentrations of serine and glycine in blood and severe neuronal manifestations. Unexpectedly, the PGS for higher levels of these amino acids was associated with lower expression of *PHGDH*, *PSAT1* and *AARS*, implying the presence of a negative feedback loop that controls serine synthesis.

Discussion

We performed *cis*-eQTL, *trans*-eQTL and eQTS analyses in 31,684 blood samples - a six-fold increase in sample size over earlier studies[5,9]. Of the genes expressed in blood, 88.3% showed a *cis*-eQTL effect, 32% showed a *trans*-eQTL effect and 13% showed an eQTS effect.

Most of the studies prioritizing genes for complex traits have considered only *cis*-eQTL effects and thus our catalogue of blood *cis*-eQTLs can be used to prioritize genes in genetic loci for various phenotypes. However, *cis*-eQTL effects have been estimated to explain only a limited fraction of the heritability of gene expression, while the combination of many weak *trans*-eQTL effects is estimated to explain the majority[63], emphasizing the importance of distal effects. At the same time, the interpretation of *trans*-eQTLs in blood remains challenging: limited replication and the influence of cell composition suggest that the effects are highly cell-type-specific. Nevertheless, the replication analyses we carried out in PBMC scRNAs-seq data prioritized 729 *trans*-eQTLs, and more than half of the *trans*-eQTLs identified by our study were assigned a putative biological mechanism of action, with transcriptional regulation through TF activity being the most prevalent.

To identify genes that are coordinately affected by multiple independent trait-associated SNPs, we performed eQTS analysis. By calculating PGSs at multiple GWAS significance thresholds, we included not only genome-wide significant SNPs but also SNPs reaching less stringent significance thresholds, potentially leading to additional information representing differences in polygenicity between traits. We identified eQTS associations for 2,568 genes and have outlined several examples where the associated genes point to interpretable biology. One possible interpretation of these eQTS associations is in the context of the recently proposed omnigenic model[13,14]. As explained by Liu et al[14], many weak distal effects could converge on the trait-relevant '*core*' genes, and eQTS analysis might help to prioritise such genes (**Supplementary Equations**). However, an important limitation is that eQTS analysis can also identify genes which are merely co-regulated with the trait-relevant ones. Therefore, it remains challenging to systematically evaluate which fraction of the detected eQTS genes is actually causal. Whereas our analysis does not formally prove or disprove the validity of the model by Liu et al[14], and the true implications of this model remain to be investigated, our results can serve as a future starting point to follow up on the identified eQTS genes and to ascertain their role in complex traits. To our knowledge, our eQTS analysis provides the first comprehensive resource in blood that can be used to interpret the effects of PGS on a molecular level.

There are some important limitations that require consideration when using our resource for hypothesis generation. First, we limited our *cis*-eQTL analysis to variants within 1 Mb of the gene center, and limited our *trans*-eQTL analysis on variants >5 Mb from genes on the same chromosome. We acknowledge the possibility that these thresholds may have excluded discovery of distal *cis*-eQTLs (e.g. those caused by distal enhancers or chromatin loops), and *trans*-eQTLs on nearby genes. We chose these thresholds to ensure that the *trans*-eQTLs we observed were not driven by long-range *cis*-eQTLs. While our approach might have excluded long-range *cis*-eQTLs, we observed that for 95.6% of genes, the lead *cis*-eQTL SNP maps within 100kb of the gene, suggesting that long-range *cis*-eQTLs reflect only a small proportion of all *cis*-eQTLs. Second, we confined our *trans*-eQTL analyses to a subset of 10,317 variants previously associated with complex phenotypes. As such, a significant *trans*-eQTL for a trait-associated variant does not necessarily mean that there is the same underlying variant affecting both the phenotype and gene expression. Third, PGS estimates have been shown to have variable prediction accuracy even when evaluated within the same ancestry. This variability may be caused by differences in sample characteristics (e.g. age, sex, socio-economic status) in the original GWAS as well as the dataset in which the PGS is calculated[48]. Such variability may therefore have caused either inflation or deflation of our eQTS effect sizes. Although per-phenotype enrichment analyses for *trans*-eQTL and eQTS genes resulted in several examples of GO terms that were interpretable in the context of the respective traits, caution is needed when drawing conclusions on higher-level phenotypes. Instead, our resource should serve as a starting point for further in-depth studies that can reliably connect the reported eQTL and eQTS associations to phenotypes.

Although putative biological mechanisms of action could be assigned to more than half of the *trans*-eQTLs we identified, significant replication in different scRNA-seq, purified cell type and cell line datasets was very limited. Such low replication rates suggest two likely causes. First, a number of the distal effects are likely driven by inter-individual blood-cell-type composition differences, which occur in any bulk tissue. While such effects could be informative in the context of some complex traits (i.e. for autoimmune diseases), the most interesting information lies in the intracellular effects. Furthermore, while we corrected for unknown confounders in our analyses, some residual cell-type-composition effects remain in the data. Therefore, it was not possible to reliably distinguish cell-type-dependent effects from intracellular ones. Instead we present a catalogue of blood eQTLs that should serve as a prioritised list for in-depth functional studies.

Second, our discovery analyses were conducted in a sample >10 times larger than the largest replication datasets available. Because *trans*-eQTL effects are generally weak, this lack of statistical power is likely to cause low replication rates. Additionally, *trans*-eQTL effects are widely considered to be more cell-type- and tissue-specific than local *cis*-eQTL effects[17]. Although this belief might be partly caused by variable eQTL strengths in different tissue contexts and the limited power of current *trans*-eQTL studies, it would also lead to lower replication rates of blood *trans*-eQTLs in specific cell types.

Compared to the gene expression from bulk tissues, scRNA-seq datasets are less likely to be affected by cell-type composition and currently serve as the best available source for replicating, prioritizing and annotating *trans*-eQTLs. While we have compiled, to the best of our knowledge, the largest available blood scRNA replication dataset, it was still only 3.6% of the sample size of the discovery study. It is therefore unsurprising that only 35 *trans*-eQTLs reached the significance threshold (FDR<0.05). None-the-less, 84% of the 729 *trans*-eQTLs attaining nominal significance (P<0.05) also showed allelic concordance with our discovery analysis. This over-representation of concordant effects suggests that there are intracellular effects among our catalogue of *trans*-eQTL effects, even if comprehensive distinction of cell-type-composition and intracellular effects is not yet possible. Upcoming large-scale single cell eQTL studies[64] (e.g. https://www.eqtlgen.org/single-cell.html), as well as highly-powered eQTL analyses conducted in non-blood tissues[65] and cell lines, will be instrumental in distinguishing intracellular effects from cell-type composition.

Full summary statistics for our *cis*-eQTL, *trans*-eQTL and eQTS studies are publicly available (www.eqtlgen.org) and can be used to interpret GWAS, to prioritize putative trait-related genes for in-depth functional studies and to develop new methods to perform those tasks. We envision that upcoming statistical tools and frameworks that will enable federated analyses in large consortia will make it possible to conduct highly powered global *trans*-eQTL studies. This will expand the work presented here and enable researchers to better connect distal effects on gene expression with complex phenotypes.

Methods

Cohorts

The eQTLGen Consortium data consists of 31,684 blood and peripheral blood mononuclear cell (PBMC) samples from 37 datasets, pre-processed in a standardized way and analyzed by each cohort analyst. 25,482 (80.4%) of the samples added to discovery analysis were whole blood samples and 6,202 (19.6%) were PBMCs, and the majority of samples were of European ancestry (**Supplementary Table 1**). The gene expression levels of the samples were profiled by the Illumina (N=17,421; 55%), Affymetrix U219 (N=2,767; 8.7%), and Affymetrix HuEx v1.0 ST (N=5,075; 16%) expression arrays and by RNA-seq (N=6,422; 20.3%). A summary of each dataset is outlined in **Supplementary Table 1**. Detailed cohort descriptions can be found in the **Supplementary Information**. All cohorts participating in this study enrolled participants with informed consent, collected and analyzed data in accordance with ethical and institutional regulations, and provided summary statistics for the meta-analyses. The information about individual institutional review board approvals is available in the original publications for each cohort (**Supplementary Information**) or in the cohort-specific **Supplementary Information**.

Each of the cohorts carried out genotype and expression data pre-processing, PGS calculation and *cis*-eQTL-, *trans*-eQTL- and eQTS-mapping following the steps outlined in the online analysis plans, specific for each platform (see **URL**s), or with slight alterations as described in **Supplementary Table 1** and the **Supplementary Information**. All but one cohort (Framingham Heart Study), included unrelated individuals into the analysis.

Information about replication datasets is detailed in the **Supplementary Information**.

Genotype data preprocessing

The primary pre-processing and quality control of genotype data was conducted by each cohort, as specified in the original publications and in the **Supplementary Information**. The majority of cohorts used genotypes imputed to 1kG p1v3 or a newer reference panel. GenotypeHarmonizer[66] was used to harmonize all genotype datasets to match the GIANT 1kG p1v3 ALL reference panel and to fix potential strand issues for A/T and C/G SNPs. Each cohort tested SNPs with MAF >0.01, Hardy-Weinberg P-value >0.0001, call rate >0.95, and MACH r^2>0.5.

Expression data preprocessing

Illumina arrays

Illumina array expression datasets were profiled by HT-12v3, HT-12v4 and HT-12v4 WGDASL arrays. Before analysis, all the probe sequences from the manifest files of those platforms were re-mapped to GRCh37.p10 human genome and transcriptome using SHRiMP v2.2.3 aligner[67] and allowing two mismatches. Probes mapping to multiple locations in the genome were removed from further analyses.

For Illumina arrays, the raw unprocessed expression matrix was exported from GenomeStudio. Before any pre-processing, the first two PCs were calculated on the expression data and plotted to identify and exclude outlier samples. The data was normalized in several steps: quantile normalization, $\log_2$ transformation, probe centering and scaling by the equation $\text{Expression}_{\text{Probe,Sample}}$ = $(\text{Expression}_{\text{Probe,Sample}}$ - $\text{Mean}_{\text{Probe}})$ /

$Std.Dev._{Probe}$. Genes showing no variance were removed. Next, the first four multidimensional scaling (MDS) components, calculated based on non-imputed and pruned genotypes using plink v1.07[68], were regressed out of the expression matrix to account for population stratification. We further removed up to 20 of the first expression-based PCs that were not associated with any SNPs, as these capture non-genetic variation in expression. After regressing out these covariates, the residual gene expression matrix was used for eQTL mapping. Each cohort also ran MixupMapper[69] software to identify incorrectly labeled genotype–expression combinations, and remove identified sample mix-ups.

Affymetrix arrays

Affymetrix-array-based datasets used expression data previously pre-processed and quality controlled as indicated in the **Supplementary Information**.

RNA-seq

Alignment, initial quality control and quantification differed slightly across datasets, as described in the **Supplementary Information**. Each cohort removed outliers as described above, and then used Trimmed Mean of M-values normalization[70] and a counts per million (CPM) filter to include genes with >0.5 CPM in at least 1% of the samples. Subsequent steps were identical to the Illumina processing, with some exceptions for the BIOS Consortium datasets (**Supplementary Information**). For BIOS Consortium datasets, up to 25 of the first expression PCs that were not associated with any SNPs were removed instead of 20.

Empirical probe matching

To integrate the different expression platforms (four different Illumina array models, RNA-seq, Affymetrix U219 and Affymetrix Hu-Ex v1.0 ST) for the purpose of meta-analysis, we developed an empirical probe-matching approach. We used the pruned set of SNPs to conduct per-platform meta-analyses for all Illumina arrays, for all RNA-seq datasets, and for each Affymetrix dataset separately, using summary statistics from analyses without any gene expression correction for PCs. For each platform, this yielded an empirical *trans*-eQTL Z-score matrix, as well as 10 permuted Z-score matrices in which links between genotype and expression files were permuted. These permuted Z-score matrices reflect the gene–gene or probe–probe correlation structure.

We used RNA-seq permuted Z-score matrices as a gold standard reference and calculated, for each gene, the Pearson correlation coefficients with all the other genes, yielding a correlation profile for each gene. We then repeated the same analysis for the Illumina meta-analysis and the two different Affymetrix platforms. Finally, we correlated the correlation profiles from each array platform with the correlation profiles from RNA-seq. For each array platform, we selected the probe showing the highest Pearson correlation with the corresponding gene in the RNA-seq data and treated those as matching expression features in the combined meta-analyses. This yielded 19,942 genes that were detected in RNA-seq datasets and tested in the combined meta-analyses. Genes and probes were matched to Ensembl v71[71] (see **URLs**) stable gene IDs and HGNC symbols in all the analyses.

Meta-analysis procedure

The results presented in this study were meta-analyzed using a weighted Z-score method[72], where the Z-scores are weighted by the square root of the sample size of the cohort. For *cis*-eQTL and *trans*-eQTL meta-analyses, this resulted in a final sample size of up to 31,684. The combined eQTS meta-analysis included the subset of unrelated individuals from the Framingham Heart Study, resulting in a combined sample size of up to 28,158. Considering that our analysis contained many different gene expression and genotyping platforms, we limited our meta-analysis to associations present in at least two cohorts in order to reduce platform-specific effects. Specifics for each meta-analysis (*cis*-eQTL, *trans*-eQTL, eQTS) are detailed below.

Cross-platform replications

To test the performance of the empirical probe-matching approach, we conducted discovery *cis*-, *trans*- and eQTS meta-analyses for each expression platform (RNA-seq, Illumina, Affymetrix U219 and Affymetrix Hu-Ex v1.0 ST arrays; array probes matched to 19,942 genes by empirical probe matching). For each discovery analysis, we conducted replication analyses in the three remaining platforms, observing strong replication of both *cis*-eQTLs, *trans*-eQTLs and eQTS in the different platforms with very good concordance in allelic direction.

Quality control of the meta-analyses

For quality control of the overall meta-analysis results, MAFs for all tested SNPs were compared between eQTLGen and 1kG p1v3 EUR (**Supplementary Figure 20**), and the effect direction of each dataset was compared against the meta-analyzed effect (**Supplementary Figure 21A-C**).

Cis-eQTL mapping

Cis-eQTL mapping was performed in each cohort using a pipeline described previously[5]. In brief, the pipeline takes a window of 1 Mb upstream and 1 Mb downstream around each SNP to select genes or expression probes to test, based on the center position of the gene or probe. The associations between these SNP–gene combinations are then calculated using Spearman correlation. Next, every cohort performed 10 permutations. In each permutation, the links between genotype and expression identifiers were shuffled prior to re-calculating all associations. Both the non-permuted results and each round of permuted results were then meta-analyzed over cohorts.

Multiple testing correction for *cis*-eQTL mapping

For our multiple testing procedure, we used the meta-analyzed permutations across all cohorts to calculate the overall FDR, as previously described[5]. In short, we reasoned that the large numbers of correlated SNPs and genes present in the *cis*-eQTL results might cause inflated estimates (i.e. highly correlated SNPs associated with a specific gene would result in equal permuted P-values for that particular gene). To circumvent this issue, we first selected the lowest association P-value per gene in both the permuted and non-permuted meta-analyses. The resulting lists of P-values were then sorted and, per given P-value in the non-permuted data, we determined the proportion of P-values equal to or below this value in both the permuted and non-permuted data. We then determined our FDR estimate as the proportion of permuted P-values over the proportion of non-permuted P-values. If a specific eQTL from the full set was not among the set of per-gene lowest association P-values, this eQTL was assigned the higher FDR value corresponding to the next eQTL available among the set of lead variants per gene. We refer to this procedure as ‘gene-level’ FDR, but note that the FDR estimates should be evaluated as ‘analysis-wide’, since the ultimate distribution of permuted P-values used to calculate our FDR estimates was derived for all tested genes, rather than per-gene. *Cis*-eQTLs with a gene-level FDR<0.05 (corresponding to $P<2.02\times10^{-5}$) that were tested in more than one cohort were deemed significant.

Trans-eQTL mapping

Trans-eQTL mapping was performed using a previously described pipeline[5] while testing a subset of 10,317 SNPs associated with complex traits. We required the distance between the SNP and the center of the gene or probe to be >5 Mb. To maximize the power to identify *trans*-eQTL effects, the results of the summary-statistics-based or iterative conditional *cis*-eQTL mapping analyses (**Supplementary Methods**) were used to correct the expression matrices before *trans*-eQTL mapping. For that, lead SNPs for significant (FDR<0.05) conditional *cis*-eQTLs were regressed out from the expression matrix. Finally, we removed potential false positive *trans*-eQTLs caused by reads cross-mapping with *cis* regions (**Supplementary Methods**).

Genetic risk factor selection

Genetic risk factors were downloaded from three public repositories: the EBI GWAS Catalog[73] (downloaded 21 November 2016), the NIH GWAS Catalogue and Immunobase (www.immunobase.org; accessed 26 April 2016), applying a significance threshold of $P\leq5\times10^{-8}$. Additionally, we added 2,706 genome-wide significant GWAS SNPs from a recent blood trait GWAS[23]. SNP coordinates were lifted to hg19 using the *liftOver* command from R package rtracklayer v1.34.1[74] and subsequently standardized to match the GIANT 1kG p1v3 ALL reference panel. This yielded 10,562 SNPs (**Supplementary Table 2**). We tested associations between all risk factors and genes that were at least 5 Mb away to ensure that that they did not tag a *cis*-eQTL effect. In total, 10,317 trait-associated SNPs were tested in *trans*-eQTL analyses.

Conditional *trans*-eQTL analyses

We aimed to estimate how many *trans*-eQTL SNPs were likely to drive both the *trans*-eQTL effect and the GWAS phenotype. The workflow of this analysis is shown in **Supplementary Figure 11**. We used the discovery *trans*-eQTL analysis results as an input, confined ourselves to those effects that were present in the datasets we had direct access to (BBMRI-BIOS+EGCUT; N=4,339), and showed nominal $P<8.3\times10^{-06}$ in the meta-analysis of those datasets. This P-value threshold was the same as in the full combined *trans*-eQTL meta-analysis and was based on the FDR=0.05 significance threshold identified from the analysis run on the pruned set of GWAS SNPs after removal of cross-mapping effects. We used the same methods and SNP filters as in the full combined *trans*-eQTL meta-analysis, aside from the FDR calculation, which was based on the full set of SNPs instead of the pruned set of SNPs.

For each significant *trans*-eQTL SNP (FDR<0.05), we defined the locus by adding a ±1 Mb window around it. Next, for each *trans*-eQTL gene, we ran iterative conditional *trans*-eQTL analysis using all loci for a given *trans*-eQTL gene. We then evaluated the LD between all conditional lead *trans*-eQTL SNPs and lead *cis*-eQTL SNPs using a 1 Mb window and $R^2>0.8$ (1kG p1v3 EUR) as a threshold for LD overlap.

cis-eQTL - *trans*-eQTL interaction analyses

We aimed to identify local *cis*-eQTL genes that affect the *trans*-eQTL effect by changing its strength or direction and might therefore serve as potential mediators. We used a G × E interaction model to test this:

$t = \beta_0 + \beta_1 \times s + \beta_2 \times m + \beta_3 \times s \times m$

where t is the expression of the *trans*-eQTL gene, s is the *trans*-eQTL SNP, and m is the expression of a potential mediator gene within 100kb of the *trans*-eQTL SNP. We omitted trans-eQTL SNP locating to HLA region from those analyses because of the complex structure of this region. On top of the gene expression normalization that we used for discovery analyses, we used a rank-based inverse normal transformation to enforce a normal distribution before fitting the linear model. This is identical to the normalization used by Zhernakova et al[15] in their G × E interaction eQTL analyses. We fitted this model separately to each of the cohorts that are part of the BIOS consortium and to EGCUT. We transformed the interaction P-values to Z-scores and used the weighted Z-score method[72] to perform a meta-analysis of 4,339 samples. The Benjamini-Hochberg procedure[18] was used to limit the FDR to 0.05. The plots in **Supplementary Figure 13** were created with the default normalization, and the regression lines are the best-fitting lines between the mediator gene and the *trans*-eQTL gene, stratified by genotype.

eQTS mapping

PGS trait inclusion

Full association summary statistics were downloaded from several publicly available resources (**Supplementary Table 19**). Based on the information presented on the web sites or abstracts of corresponding publications, GWAS performed exclusively in non-European cohorts were omitted. Filters applied to the separate data sources are indicated in the **Supplementary Information**. All the dbSNP rs numbers were standardized to match GIANT 1kG p1v3, and the directions of effects were standardized to correspond to the GIANT 1kG p1v3 minor allele. SNPs with opposite alleles compared to GIANT alleles were flipped. SNPs with A/T and C/G alleles, tri-allelic SNPs, indels, SNPs with different alleles in GIANT 1kG p1v3 and SNPs with unknown alleles were removed from the analysis. Genomic control was applied to all the P-values for the datasets not genotyped by Immunochip or Metabochip. Additionally, genomic control was skipped for one dataset that did not have full associations available[75] and for all the datasets from the GIANT consortium because genomic control had already been applied for these. In total, 1,263 summary statistics files were added to the analysis. Information about the summary statistics files can be found in the **Supplementary Information** and **Supplementary Table 19**.

PGS calculation

A custom Java program, GeneticRiskScoreCalculator-v0.1.0c, was used for calculating several PGS in parallel. Independent effect SNPs for each summary statistics file were identified by double-clumping, first using a 250kb window and then subsequently a 10Mb window with an LD threshold R^2=0.1. Weighted PGS were calculated by summing the risk alleles for each independent SNP, weighted by its GWAS effect size (beta or log(OR) from the GWAS study). Five GWAS P-value thresholds ($P<5\times10^{-8}$, 1×10^{-5}, 1×10^{-4}, 1×10^{-3} and 1×10^{-2}) were used for constructing PGSs for each summary statistics file. PGS were scaled to fall between 0 and 2, for compatibility with QTL mapping pipeline.

Pruning SNPs and PGSs

To identify a set of independent genetic risk factors, we conducted LD-based pruning as implemented in PLINK 1.9[76] with the setting --indep-pairwise 50 5 0.1. This yielded 4,586 uncorrelated SNPs (R^2<0.1, GIANT 1kG p1v3 ALL).

To identify the set of uncorrelated PGS, 10 permuted *trans*-eQTL Z-score matrices from the combined *trans*-eQTL analysis were first confined to the pruned set of SNPs. Those matrices were then used to identify 3,042 uncorrelated genes based on Z-score correlations (absolute Pearson R<0.05). Next, permuted eQTS Z-score matrices were confined to uncorrelated genes and used to calculate pairwise correlations between all genetic risk scores to define a set of 1,873 uncorrelated genetic risk scores (Pearson R^2<0.1).

Multiple testing correction in *trans*-eQTL and eQTS mapping

To calculate FDR estimates for *trans*-eQTLs and eQTS, we compared each P-value from the non-permuted meta-analysis with all P-values from 10 meta-analyzed permutation rounds. We note that this differs from the permutation strategy used in the *cis*-eQTL analysis, because here we used the P-values from all SNP-gene combinations, not just the smallest P-value for each gene. Nevertheless, we note that the 10,317 SNPs tested for *trans*-eQTLs contained many linked variants. To establish a conservative FDR estimate, we therefore first pruned this list of variants, leaving 4,586 SNPs. Using this list of SNPs, we then performed a focused meta-analysis for both the non-permuted and permuted datasets. We derived FDR estimates from these limited meta-analyses by sorting the resultant lists of P-values and determining the proportion of P-values in the non-permuted and permuted datasets for each given P-value in the non-permuted dataset. We then applied these FDR estimates to the *trans*-eQTL results from all 10,317 genetic trait-associated SNPs. If a specific eQTL from the full set was not tested in the meta-analysis conducted on the pruned set, this eQTL was assigned the higher FDR value corresponding to the next eQTL tested in the pruned set. We used an FDR threshold of 0.05 to declare a *trans*-eQTL effect significant. Similarly, in the eQTS analysis, we used a set of 1,873 uncorrelated (Pearson R^2<0.1) PGSs and performed an analogous FDR calculation. We analyzed only SNP/PGS–gene pairs tested in at least two cohorts.

Replication of *trans*-eQTL and eQTS effects in cell lines and purified cell types

Information about replication cohorts and their respective settings for replication analyses is outlined in the **Supplementary Information**. If applicable, summary statistics from different replication datasets for the same specific cell type or tissue were meta-analysed using a weighted Z-score method[72]. Benjamini-Hochberg FDR[18] was used to adjust replication analysis P-values for multiple testing. We required FDR<0.05 and the same effect direction with discovery to declare effect replicating. R package *pwr* (**URLs**) was used to conduct power analyses for replication datasets.

Cell-type-composition effects of *trans*-eQTLs and eQTS

Dataset

We used data from a subset 3,831 BIOS individuals to which we had direct access. We further narrowed our sample set down to 1,858 individuals for whom the measured cell metric data was available for at least ⅔ of measured cell metrics. All samples were part of discovery meta-analyses.

Measured cell metrics

Several cell types were counted in peripheral blood from each of the BIOS cohort participants, but cohorts differed in the availability. Cells were counted as an absolute number in a liter of blood (white blood cell count, red blood cell count, platelet count), or as a percentage of the white blood cell count (neutrophil percentage, lymphocyte percentage, etc.). Out of 24 cell metrics, we excluded eight (LUC, LUC%, RBC, RDW, MCH, MPV, MCHC, MCV) because these measurements were not available for the large majority of samples, hindering the estimation of the combined effect of measured cell metrics on *trans*-eQTLs and eQTS. All measured cell metrics are summarized in **Supplementary Table 28**.

Estimated cell counts

We estimated the cell counts of 33 different cell types using Decon-cell, part of the Decon2 method[77]. Decon-cell was trained using information from the independent 500FG cohort, which includes detailed cell type measures as well as RNA-seq expression profiles[78]. Next, the prediction model was used to impute cell proportions based on the BIOS gene expression matrix. Predicted cell metrics are summarized in **Supplementary Table 28**.

Cell type interaction analyses

Here we used data from a subset of up to 1,858 BIOS Consortium samples for which 49 measured and predicted cell type metrics were available. For these analyses we tested only effects where the SNP had a MAF>0.05 in each BIOS cohort.

All 49 cell metrics were transformed by inverse normal transformation prior to analyses. For gene expression, we used the same preprocessing as in the discovery meta-analyses, including correction for expression PCs and regression of *cis*-eQTL effects. In addition to the standard preprocessing, the expression of each gene was transformed using inverse normal transformation.

For multivariate linear models, analyses were conducted using R v3.4.4, data.table v1.12, tidyverse v1.2.1, broom v0.5.1 and the pheatmap v1.0.12 packages (**URLs**). For each BIOS cohort, linear models were fitted for each *trans*-eQTL identified in meta-analysis (FDR<0.05), using lm() function from R. For eQTS analyses, PGS was used instead of SNP.

Three different interaction models were fitted for each *trans*-eQTL and eQTS:

$$t = \beta_0 + \beta_1 \times s$$
$$t = \beta_0 + \beta_1 \times c_1 + \beta_2 \times c_2 + \ldots + \beta_{49} \times c_{49} + \beta_{50} \times s$$
$$t = \beta_0 + \beta_1 \times c_1 + \beta_2 \times c_2 + \ldots + \beta_{49} \times c_{49} + \beta_{50} \times c1 \times s + \ldots + \beta_{99} \times c_{49} \times s + \beta_{100} \times s$$

where t is the expression of *trans*-eQTL/eQTS gene, c is cell-type metric, and s is a dosage of *trans*-eQTL SNP or scaled value of polygenic score. P-values from each term of the linear model (main effects and interaction effects) were converted to signed Z-scores and effects were meta-analyzed by weighted Z-score method, using the square root of per-cohort sample size as weight.

To determine the effect of cell-type composition on *trans*-eQTLs, we applied models and assessed the SNP main effect. Here we used the same significance thresholds as determined by the permutation-based FDR in the discovery meta-analyses.

To determine the likely cell types where *trans*-eQTLs or eQTSs can manifest, we applied the third model with the difference that no PCs were removed from gene expression data prior to analysis and queried the individual interaction term for each cell metric. A Benjamini-Hochberg FDR[18] across all interaction P-values was used to determine significance in this analysis.

scRNA-seq analyses

scRNA-seq cohorts and data

For the replication of *trans*-eQTLs in scRNA-seq, we used unpublished data of PBMCs from 1,139 unrelated individuals in two cohorts generated using the 10X Chromium platform: OneK1K (N=982) and 1M-scBloodNL (N=157). The data was processed using the Cell Ranger Single Cell Software Suite (see **URLs**) and aligned using STAR[79]. Cells were demultiplexed and doublets removed before performing cell type classification. We combined the data in a meta-analysis within each of the eight available cell types: B-cells, CD4+ T-cells, CD8+ T-cells, classical monocytes, non-classical monocytes, dendritic cells, natural killer cells and plasma cells.

Replication of *trans*-eQTL effects

We tested the replication of the 59,786 discovery *trans*-eQTLs only if the *trans*-eQTL gene was sufficiently expressed (i.e. had a missing sample fraction that was at most 20% in the large OneK1K dataset), leaving between 1,917 and 27,582 eQTLs to be studied, depending on cell type. We estimated the inflation of signal by calculating the lambda inflation relative to the inverse chi-square cumulative distribution function of 0.5. *Trans*-eQTLs with FDR<0.05 in any cell type were deemed significantly replicating. To get a better idea of replication across cell types, we calculated the average Z-score across cell types. We selected effects with an absolute average Z-score>1.96 (equivalent to P<0.05) to calculate the allelic concordance with the discovery *trans*-eQTLs.

Correlation of *trans*-eQTL effects

To test the correlation of *trans*-eQTL effects, we used the r_b approach[25], which accounts for the errors in the estimated eQTL effects so that the estimate of correlation is less dependent on sample sizes. First, we derived the estimate of the *trans*-eQTL effect (beta) and the standard error of the beta (SE(beta)) from the Z-score and the MAF of the significant *trans*-eQTLs, using the following formulae from Zhu et al., 2016[80]

$$beta = z / (\sqrt{(2p(1-p)(n+z^2))})$$

$$SE(beta) = 1 / (\sqrt{(2p(1-p)(n+z^2))})$$

where p is the MAF, n is the sample size and z is the meta-analysis Z-score. MAF was computed from 26,609 eQTLGen samples (excluding FHS) for discovery analysis and from 1,309 replication samples for scRNA-seq replication analyses. For analyses in purified cell types and cell lines (LCL, iPSC) where allele frequencies were not available, we used the MAF as observed in eQTLGen instead.

In order to include independent effects in the analysis, for each *trans*-eQTL gene, we included only the strongest significant discovery effect in each 2 Mb window. Statistics of r_b and $SE(r_b)$ were calculated as detailed in[25] assuming no sample overlap between discovery and replication datasets. Because we were only seeking to correlate the effects of identified *trans*-eQTLs, we did not use any reference discovery dataset for selecting *trans*-eQTLs to estimate r_b, and hence did not consider potential ascertainment bias, although such bias is likely to be small. To calculate a P-value, the Z-score was first calculated by dividing r_b by SE(r_b) and then squared to calculate the χ^2 statistic. The P-value was then derived from the χ^2 distribution with one degree of freedom.

TF and tissue enrichment analyses for *REST* locus

We downloaded curated sets of known TF-targets and tissue-expressed genes from the Enrichr website[81,82]. TF-target gene sets included TF-targets as assayed by ChIP-X experiments from the ChEA[83] and ENCODE[39,40] projects. Tissue-expressed genes were based on the ARCHS4 database[84]. Gene sets were processed and mapped to entrez IDs with R package ClusterProfiler v3.10.1[85]. Those gene sets were then used to conduct over-representation analyses by one-sided Fisher's exact test as implemented into the R package GeneOverlap v1.18.0 (see **URLs**).

Trans-eQTL enrichment analyses

To better understand the biological mechanisms underlying the *trans*-eQTLs, we performed a number of enrichment analyses. We converted *trans*-eQTLs to a gene-by-gene matrix via three methods: using Pascal, using *cis*-eQTL information and combining both (**Supplementary Methods**). For the enrichments, we calculated whether there was significant overlap with known TF–target pairs[27]; gene co-regulation patterns (**Supplementary Methods**); PPIs[29] and Hi-C contacts in LCL cells[31] using a two-sided Fisher's exact test.

Capture Hi-C overlap for *cis*-eQTLs

To assess whether *cis*-eQTL lead SNPs overlapped with chromosomal contact as measured by Hi-C data, we used promoter capture Hi-C data[86] downloaded from CHiCP[87] (see **URLs**). We took the lead eQTL SNPs, overlapped these with the capture Hi-C data, and studied the 10,428 *cis*-eQTL genes for which this data was available. We then checked whether the capture Hi-C target maps within 5kb of the lead SNP. Of the 803 *cis*-eQTL genes for which the lead SNP mapped more than 100 kb away from the TSS or TES, 223 overlapped with capture Hi-C data (27.8%). Of 9,625 *cis*-eQTL genes for which the lead SNP mapped within 100kb from the TSS or TES, 1,641 overlapped with capture Hi-C data (17.0%). To test if these observed overlaps were not happening by chance, we performed the same analysis while flipping the location of the capture Hi-C target relative to the location of the bait. To test the difference between observed Hi-C overlap compared to flipped analysis, we used a two-tailed two-sample test of equal proportions.

Data availability

Primary genotype and gene expression data was analyzed by individual cohorts participating in the study and current study analyzed summary statistics. Full summary statistics of the eQTLGen *cis*-eQTL, *trans*-eQTL and eQTS meta-analyses are available on the eQTLGen website, www.eqtlgen.org, which was built using the MOLGENIS framework[88]. We also provide *cis*-eQTL files formatted for use in SMR, MAFs, and replication statistics for *cis*-eQTLs, *trans*-eQTLs and eQTS.

Code availability

Individual cohorts participating in the study followed the analysis plans as specified in the **URL**s or with slight alterations as described in the **Methods** and **Supplementary Information**. All tools and source codes used for genotype harmonization, identification of sample mix-ups, eQTL mapping, meta-analyses and calculation of PGS are freely available at https://github.com/molgenis/systemsgenetics/.

Author contributions

U.V. and A.C. coordinated the consortium analyses, ran the meta-analyses, interpreted the data, performed downstream analyses and drafted the manuscript. H-J.W, MJ.B and P.D. developed the software used in the analyses, did downstream analyses and participated in manuscript writing. L.F. and T.E. conceived the study. L.F. supervised the project, participated in the manuscript writing and ran downstream analyses. B.Z., H.K., A.S., S.K., N.P., I.A., M-J.F, M.A., M.C., R.J., I.S., L.T., A.T., K.S., J.V., H.Y., V.K., A.K., J.K., J.P., B.L. ran consortium analyses in their respective cohorts. A.S., R.K., S.K., G.H., R.S., A.B. ran replication analyses in their respective cohorts. A.A., G.W.M., S.R., M.P., E.D., S.B., T.F., J.v.M, H.P., H.A., B.P., T.L., D.B., B.M.P., S.A.G., P. A., L.M., W.O., K.D., O.S., A.B., M.Sc., G.G., T.E., W.A., F.B., J.D., M.E., B.P.F, M.G., B.T.H., M.K., Y.K., J.C.K, P.K., K.K., M.L., U.M.M., H.M., Y.M., M.M-N., M.Na., M.Ni., B.P., O.Ra., O.Ro., E.P.S, C.D.A.S., M.St., P.S., P.A.C.'tH, J.T., A.T., J.v.D., M.v.I, J.V., U.Vö., C.W. provided the data used in the study. B.Z., H.K., Z.K., J.K-G, S.R., E.P., S.L., J.Y., F.Z., P.M.V., J.P., T.Q. and G.G. participated in downstream analyses. S.Y., H.B., R.O., R.W., D.d.V. and M.v.d.W. ran replication analyses in scRNA-seq cohorts. A.H. and J.A.H. generated scRNA-seq cohort data. H.B. and M.Sw. created the website to host the results. U.V. and A.C. contributed equally to this work. H-J.W, MJ.B and P.D. contributed equally to this work. L.F., T.E., G.G and J.P contributed equally to this work.

Acknowledgments

The cohorts participating in this study list the acknowledgments in the cohort-specific supplemental information in **Supplementary Information**.

The acknowledgments for GWAS consortia whose summary statistics were used in this study are listed in consortium-specific supplementary information in **Supplementary Information**.

We thank the i2QTL CONSORTIUM for providing the iPSC replication results.

This work is supported by a grant from the European Research Council (ERC, ERC Starting Grant agreement number 637640 ImmRisk) to L.F and a VIDI grant (917.14.374) from the Netherlands Organisation for Scientific Research (NWO) to L.F. This work has also been supported by the European Regional Development Fund and the programme Mobilitas Pluss (MOBTP108) to U.V. The project has also been supported by the Foundation “De Drie Lichten” in the Netherlands with a grant to A.C. M.G.N. is supported by ZonMw grants 849200011 and 531003014 from The Netherlands Organisation for Health Research and Development, a VENI grant from NWO (VI.Veni.191G.030) and a Jacobs foundation research fellowship. H.Y. is funded by a Diabetes UK RD Lawrence fellowship (17/0005594). This project also received funding from the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 772376 - EScORIAL) to J.V. This work was supported by the Estonian Research Council grant PUT (PUT1660) to T.E. This research was supported by the European Union through Horizon 2020 grant no. 810645 to T.E. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 847776 to L.M. A.B. was supported by NIH grant 1R01MH109905, NIH grant R01HG008150 (NHGRI; Non-Coding Variants Program) and NIH grant R01MH101814 (NIH Common Fund; GTEx Program). M.v.d.W was funded by Nederlandse Organisatie voor Wetenschappelijk onderzoek, NWO-Veni 192.029. This work was supported by National Institutes of Health grants R21ES024834 (B.P. and M.A.), R01ES020506 (B.P.), R01ES023834 (B.P.), R35ES028379 (B.P.), R01 GM108711 (L.C.), and R01CA107431 (H.A.). EGCUT analyses were funded by EU H2020 grant 692145, Estonian Research Council Grant IUT20-60, IUT24-6, and European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012GENTRANSMED. HVH was supported in part by grants R01 HL085251 and R01 HL073410 from the National Heart, Lung, and Blood Institute (NHLBI). The provision of genomic data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Research Center. This study was supported in part by the Intramural Research Program, National Institute on Ageing. UK-based work was generously supported by a Wellcome Trust Institutional Strategic Support Award (WT097835MF), plus internal University of Exeter Medical School funding. This work has made use of the resources provided by the University of Exeter Science

Strategy and resulting Systems Biology initiative. The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative (project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). This work was supported by grants from the German Research Foundation (SFB-1052 ‘‘Obesity mechanisms’’ A01, B03, SPP 1629 TO 718/2- 1), from the German Diabetes Association, from the DHFD (Diabetes Hilfs- und Forschungsfonds Deutschland) and from IFB Adiposity Diseases (AD2-060E, AD2-06E95, AD2-06E99). IFB Adiposity Diseases is supported by the Federal Ministry of Education and Research (BMBF), Germany, FKZ: 01EO1501. 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 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). This study was supported by grants from the Singapore Immunology Network (SIgN-06-006, SIgN-08-020 and SIgN-10-029); the National Medical Research Council (NMRC/1150/2008), Singapore; the Biomedical Research Council, Singapore; SIgN core funding from the Agency for Science, Technology and Research (A*STAR); and the National University of Singapore for the Graduate Research Scholarship for students involved in the study. The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS); and European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation. We acknowledge support from BBMRI–NL (Biobanking and Biomolecular Resources Research Infrastructure 184.021.007 and 184.033.111); Spinozapremie (NWO- 56-464-14192), the European Research Council (ERC Advanced 230374) and KNAW Academy Professor Award (PAH/6635) to DIB, the National Institutes of Health (NIH, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA) and NWO-Groot 480-15-001/674: Netherlands Twin Registry Repository. A.B. was supported by NIH grant 1R01MH109905, NIH grant R01HG008150 (NHGRI; Non-Coding Variants Program), and NIH grant R01MH101814 (NIH Common Fund; GTEx Program). The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (NIH). Additional funds were provided by the National Cancer Institute; National Human Genome Research Institute (NHGRI); National Heart, Lung, and Blood Institute; National Institute on Drug Abuse; National Institute of Mental Health; and National Institute of Neurological Disorders and Stroke. Donors were enrolled at the Biospecimen Source Sites funded by

Leidos Biomedical, Inc. (Leidos) subcontracts to the National Disease Research Interchange (10XS170) and Roswell Park Cancer Institute (10XS171). The LDACC was funded through a contract (HHSN268201000029C) to The Broad Institute. Biorepository operations were funded through a Leidos subcontract to the Van Andel Institute (10ST1035). Additional data repository and project management were provided by Leidos (HHSN261200800001E). The Brain Bank was supported by a supplement to University of Miami grant DA006227. For the DGN data, we gratefully acknowledge the resources supported by National Institutes of Health/National Institute of Mental Health grants 5RC2MH089916 (PI: Douglas F. Levinson, M.D.; Co- investigators: Myrna M. Weissman, Ph.D., James B. Potash, M.D., MPH, Daphne Koller, Ph.D., and Alexander E. Urban, Ph.D.) and 3R01MH090941 (Co-investigator: Daphne Koller, Ph.D.). The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). This research was specifically funded by ERC 260927, Swiss National Foundation 130342 and Wellcome Trust and MRC 092731. S.B was supported by the Swiss National Science Foundation (310030-152724). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Research in the Ouwehand laboratory receives funding from the British Heart Foundation, European Commission (TrainMALTA), International Society on Thrombosis and Haemostasis, National Institute for Health Research England, Medical Research Council, NHS Blood and Transplant and the Rosetrees Trust.

We thank the UMCG Genomics Coordination Center, the MOLGENIS team, the UG Center for Information Technology, and the UMCG research IT program and their sponsors in particular BBMRI-NL for data storage, high performance computing and web hosting infrastructure. BBMRI-NL is a research infrastructure financed by the Netherlands Organization for Scientific Research (NWO) [grant number 184.033.111].

We thank Kate McIntyre for editing the final text.

URLs

Full summary statistics from this study, www.eqtlgen.org

ExAC pLI scores, http://exac.broadinstitute.org/downloads;

Ensembl v71 annotation file, ftp://ftp.ensembl.org/pub/release-71/gtf/homo_sapiens;

Reference for genotype harmonizing, ftp://share.sph.umich.edu/1000genomes/fullProject/2012.03.14/GIANT.phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.ALL.vcf.gz.tgz

eQTLGen analysis plan for Illumina array datasets, https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook;

eQTLGen analysis plan for RNA-seq datasets, https://github.com/molgenis/systemsgenetics/wiki/eQTL-mapping-analysis-cookbook-for-RNA-seq-data;

eQTLGen analysis plan for Affymetrix array datasets, https://github.com/molgenis/systemsgenetics/wiki/QTL-mapping-analysis-cookbook-for-Affymetrix-expression-arrays;

GenotypeHarmonizer, https://github.com/molgenis/systemsgenetics/wiki/Genotype-Harmonizer;

Protocol to resolve sample mixups, https://github.com/molgenis/systemsgenetics/wiki/Resolving-mixups;

Enrichr gene set enrichment libraries, http://amp.pharm.mssm.edu/Enrichr/;

GeneOverlap package for enrichment analyses, https://www.bioconductor.org/packages/release/bioc/html/GeneOverlap.html;

ToppGene web tool used for enrichment analyses,
https://toppgene.cchmc.org/;

SHRiMP aligner used for re-mapping Illumina probes,
http://compbio.cs.toronto.edu/shrimp/;

EBI GWAS Catalog,
https://www.ebi.ac.uk/gwas/;

Immunobase,
http://www.immunobase.org/;

ClusterProfiler package used for tissue enrichment analyses,
http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html;

Capture Hi-C data,
https://www.chicp.org/

SNiPA, used to acquire proxy SNPs,
http://snipa.helmholtz-muenchen.de/snipa3/

Regulatory Circuits, used to acquire TF data,
www.RegulatoryCircuits.org

FANTOM5 TF annotations,
http://fantom.gsc.riken.jp/5/sstar/Browse_Transcription_Factors_hg19

PPI interactions,
https://www.intomics.com/inbio/map/api/get_data?file=InBio_Map_core_2016_09_12.tar.gz

R v3.4.4 software,
https://cran.r-project.org/

data.table v1.12 package,
https://cran.r-project.org/web/packages/data.table/index.html

tidyverse v1.2.1 package,
https://cran.r-project.org/web/packages/tidyverse/index.html

broom v0.5.1 package, https://cran.r-project.org/web/packages/broom/index.html

pheatmap v1.0.12 package, https://cran.r-project.org/web/packages/pheatmap/index.html

pwr v1.3-0 package, https://cran.r-project.org/web/packages/pwr/index.html

Cell Ranger software, https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger

References

1. Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *American Journal of Human Genetics* **100**, 571–580 (2017).
2. O'Connor, L. J. *et al.* Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *American Journal of Human Genetics* **105**, 456–476 (2019).
3. Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics* **50**, 746–753 (2018).
4. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics* **52**, 626–633 (2020).
5. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics* **45**, 1238–1243 (2013).
6. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Human Molecular Genetics* **24**, 4746–4763 (2015).
7. Lloyd-Jones, L. R. *et al.* The Genetic Architecture of Gene Expression in Peripheral Blood. *American Journal of Human Genetics* **100**, 228–237 (2017).
8. Jansen, R. *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Human Molecular Genetics* **26**, 1444–1451 (2017).
9. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biology* **18**, 16 (2017).
10. Brynedal, B. *et al.* Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation. *American Journal of Human Genetics* **100**, 581–591 (2017).

11. Lewis, C. M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Medicine* **9**, (2017).
12. Natarajan, P. *et al.* Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* **135**, 2091–2101 (2017).
13. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
14. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
15. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nature Genetics* **49**, 139–145 (2017).
16. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
17. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
18. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
19. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nature Genetics* **49**, 131–138 (2017).
20. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
21. Glassberg, E. C., Gao, Z., Harpak, A., Lan, X. & Pritchard, J. K. Evidence for weak selective constraint on human gene expression. *Genetics* **211**, 757–772 (2019).

22. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biology* **18**, 86 (2017).

23. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).

24. Van Der Wijst, M. G. P., De Vries, D. H., Brugge, H., Westra, H. J. & Franke, L. An integrative approach for building personalized gene regulatory networks for precision medicine. *Genome Medicine* **10**, 96 (2018).

25. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nature Communications* **9**, 2282 (2018).

26. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* **37**, W305–11 (2009).

27. Marbach, D. *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nature Methods* **13**, 366–370 (2016).

28. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Computational Biology* **12**, 1–20 (2016).

29. Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nature Methods* **14**, 61–64 (2016).

30. Scharer, C. D. *et al.* Genome-wide CIITA-binding profile identifies sequence preferences that dictate function versus recruitment. *Nucleic Acids Research* **43**, 3128–3142 (2015).

31. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

32. Perry, J. R. B. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).

33. Lemaitre, R. N. *et al.* Genetic loci associated with plasma phospholipid N-3 fatty acids: A Meta-Analysis of Genome-Wide association studies from the charge consortium. *PLoS Genetics* **7**, e1002193 (2011).

34. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47**, 979–986 (2015).

35. Gateva, V. *et al.* A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature Genetics* **41**, 1228–1233 (2009).

36. Moffatt, M. F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *New England Journal of Medicine* **363**, 1211–1221 (2010).

37. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**, 1173–1186 (2014).

38. Nikpay, M. *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47**, 1121–1130 (2015).

39. Feingold, E. A. *et al.* The ENCODE (ENCyclopedia of DNA Elements) Project. *Science* **306**, 636–640 (2004).

40. Myers, R. M. *et al.* A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biology* **9**, (2011).

41. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).

42. Chong, J. A. *et al.* REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).
43. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature Genetics* **47**, 1457–1464 (2015).
44. Davenport, E. E. *et al.* Discovering in vivo cytokine-eQTL interactions from a lupus clinical trial. *Genome Biology* **19**, (2018).
45. McBride, J. M. *et al.* Safety and pharmacodynamics of rontalizumab in patients with systemic lupus erythematosus: Results of a phase I, placebo-controlled, double-blind, dose-escalation study. *Arthritis and Rheumatism* **64**, 3666–3676 (2012).
46. Yao, Y. *et al.* Development of Potential Pharmacodynamic and Diagnostic Markers for Anti-IFN-α Monoclonal Antibody Trials in Systemic Lupus Erythematosus. *Human Genomics and Proteomics* **1**, (2009).
47. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214–1231.e11 (2020).
48. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, (2020).
49. Van Der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
50. Abugessaisa, I. *et al.* FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. *Database : the journal of biological databases and curation* **2016**, baw105 (2016).
51. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

52. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nature Genetics* **45**, 1274–1285 (2013).

53. Wang, X. *et al.* Macrophage ABCA1 and ABCG1, but not SR-BI, promote macrophage reverse cholesterol transport in vivo. *Journal of Clinical Investigation* **117**, 2216–2224 (2007).

54. Goldstein, J. L. & Brown, M. S. Binding and degradation of low density lipoproteins by cultured human fibroblasts. Comparison of cells from a normal subject and from a patient with homozygous familial hypercholesterolemia. *Journal of Biological Chemistry* **249**, 5153–5162 (1974).

55. Singh, A. B., Kan, C. F. K., Shende, V., Dong, B. & Liu, J. A novel posttranscriptional mechanism for dietary cholesterol-mediated suppression of liver LDL receptor expression. *Journal of Lipid Research* **55**, 1397–1407 (2014).

56. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature Communications* **7**, 11122 (2016).

57. Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nature Genetics* **46**, 543–550 (2014).

58. El-Hattab, A. W. Serine biosynthesis and transport defects. *Molecular Genetics and Metabolism* **118**, 153–159 (2016).

59. Leuzzi, V., Alessandrì, M. G., Casarano, M., Battini, R. & Cioni, G. Arginine and glycine stimulate creatine synthesis in creatine transporter 1-deficient lymphoblasts. *Analytical Biochemistry* **375**, 153–155 (2008).

60. Hart, C. E. *et al.* Phosphoserine aminotransferase deficiency: A novel disorder of the serine biosynthesis pathway. *American Journal of Human Genetics* **80**, 931–937 (2007).

61. Klomp, L. W. J. *et al.* Molecular characterization of 3-phosphoglycerate dehydrogenase deficiency - A neurometabolic disorder associated with reduced L-serine biosynthesis. *American Journal of Human Genetics* **67**, 1389–1399 (2000).
62. Shaheen, R. *et al.* Neu-laxova syndrome, an inborn error of serine metabolism, is caused by mutations in PHGDH. *American Journal of Human Genetics* **94**, 898–904 (2014).
63. Price, A. L. *et al.* Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genetics* **7**, e1001317 (2011).
64. van der Wijst, M. G. P. *et al.* Single-cell eQTLGen Consortium: a personalized understanding of disease. *arXiv e-prints* arXiv:1909.12550 (2019).
65. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
66. Deelen, P. *et al.* Genotype harmonizer: Automatic strand alignment and format conversion for genotype data integration. *BMC Research Notes* **7**, 901 (2014).
67. Rumble, S. M. *et al.* SHRiMP: Accurate mapping of short color-space reads. *PLoS Computational Biology* **5**, e1000386 (2009).
68. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575 (2007).
69. 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).
70. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25 (2010).
71. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754–D761 (2018).

72. Zaykin, D. V. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of Evolutionary Biology* **24**, 1836–1841 (2011).
73. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research* **45**, D896–D901 (2017).
74. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
75. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031–1036 (2016).
76. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, (2015).
77. Aguirre-Gamboa, R. *et al.* Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* **21**, 243 (2020).
78. Netea, M. G. *et al.* Understanding human immune function using the resources from the Human Functional Genomics Project. *Nature Medicine* **22**, 831–833 (2016).
79. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
80. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481–487 (2016).
81. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
82. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, W90–W97 (2016).

83. Lachmann, A. *et al.* ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
84. Lachmann, A. *et al.* Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications* **9**, 1366 (2018).
85. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology* **16**, 284–287 (2012).
86. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384.e19 (2016).
87. Schofield, E. C. *et al.* CHiCP: A web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* **32**, 2511–2513 (2016).
88. Swertz, M. A. *et al.* The MOLGENIS toolkit: Rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics* **11**, S12 (2010).