USP18 modulates lupus risk via negative regulation of interferon response

Krista Freimann[1,2], Anneke Brümmer[3,4], Robert Warmerdam[5,6], Tarran S Rupall[2,7], Ana Laura Hernández-Ledesma[8], Joshua Chiou[9], Emily R. Holzinger[10], Joseph C. Maranville[10], Nikolina Nakic[11], Halit Ongen[12], Luca Stefanucci[2,7], Micheal C. Turchin[10], eQTLGen Consortium[†], Lude Franke[5,6], Urmo Võsa[13], Carla P Jones[2,7], Alejandra Medina-Rivera[8], Gosia Trynka[2,7], Kai Kisand[14], Sven Bergmann[3,4,15], Kaur Alasoo[1,2*]

1. Institute of Computer Science, University of Tartu, Tartu, Estonia
2. Open Targets, South Building, Wellcome Genome Campus, Hinxton, Cambridge, UK
3. Department of Computational Biology, University of Lausanne, Switzerland
4. Swiss Institute of Bioinformatics, Lausanne, Switzerland
5. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
6. Oncode Institute, Amsterdam, the Netherlands
7. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK
8. Laboratorio Internacional de Investigación Sobre el Genoma Humano, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico
9. Internal Medicine Research Unit, Research and Development, Pfizer, Cambridge, MA, USA
10. Bristol Myers Squibb, Cambridge, Massachusetts, USA
11. Research Technologies, GSK, Stevenage, UK
12. Research Technologies, GSK, Heidelberg, Germany
13. Institute of Genomics, University of Tartu, Tartu, Estonia
14. Institute of Biomedicine and Translational Medicine, Faculty of Medicine, University of Tartu, Estonia
15. Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

[†]A list of authors and affiliations for the eQTLGen Consortium is listed in the Supplementary Materials.

[*]Corresponding author. Email: kaur.alasoo@ut.ee

Abstract

Although genome-wide association studies have provided valuable insights into the genetic basis of complex traits and diseases, translating these findings to causal genes and their downstream mechanisms remains challenging. We performed *trans* expression quantitative trait locus (*trans*-eQTL) meta-analysis in 3,734 lymphoblastoid cell line samples, identifying four robust loci that replicated in an independent multi-ethnic dataset of 682 individuals. One of these loci was a missense variant in the ubiquitin specific peptidase 18 (*USP18)* gene that is a known negative regulator of interferon signalling and has previously been associated with increased risk of systemic lupus erythematosus (SLE). In our analysis, the SLE risk allele increased the expression of 50 interferon-inducible genes, suggesting that the risk allele impairs USP18's ability to effectively limit the interferon response. Intriguingly, most *trans*-eQTL targets

of USP18 lacked independent *cis* associations with SLE, cautioning against the use of *trans*-eQTL evidence alone for causal gene prioritisation.

Introduction

Genome-wide association studies (GWAS) have provided valuable insights into the genetic basis of complex traits and diseases. However, translating GWAS findings to actionable drug targets has remained challenging, particularly when the functions of the associated genes are unknown. A promising technique to identify the effector genes of GWAS variants as well as their downstream regulatory consequences is *trans* gene expression and protein quantitative trait loci (*trans*-QTL) analysis. *Trans*-QTL studies test for association between genetic variants across the genome and expression levels of all measured genes or proteins[1]. In a prominent example, an erythrocyte-specific regulatory element first identified as a *trans* protein QTL (*trans*-pQTL) for foetal haemoglobin (HbF) was used to design the first ever gene editing therapy for sickle-cell disease[2,3].

Trans-QTLs are especially promising, because 60%-90% of gene and protein expression heritability is located in *trans*[4], most associations detected in large-scale pQTL studies are located in *trans*[5], and *cis*-QTL discovery is starting to saturate after 10,000 samples[5]. Furthermore, most complex trait heritability has been proposed to be mediated by *trans*-QTL effects[4]. However, current large-scale *trans*-eQTL and *trans*-pQTL studies have been limited to easily accessible bulk tissues such as whole blood[6,7] or plasma[5,8–10]. Bulk tissue studies are subject to cell type composition effects which can be difficult to distinguish from true intracellular *trans*-QTLs[1,6]. The whole blood and plasma studies are also likely to miss cell type and context specific regulatory effects. In contrast, *trans*-eQTL studies in other tissues and purified cell types have had limited statistical power due to small sample sizes (typically less than one thousand samples), enabling the discovery of only very large effects and potentially underestimating pleiotropic effects on multiple target genes[11–18].

A key limitation in our understanding of how *trans*-eQTLs contribute to complex traits and how they interact with *cis*-eQTL is the lack of well-characterised disease-associated *trans*-eQTL signals[4]. Two most prominent examples include the adipose-specific *KLF14* locus associated with type 2 diabetes[16,19] and the IRX3/5 locus associated with obesity[20,21]. At the *KLF14* locus, the lead variant (rs4731702) is a *cis*-eQTL for the *KLF14* transcription factor and was associated with the expression of 385 target genes in *trans*, 18 of which also had independent *cis* associations for other metabolic traits[19]. The simultaneous regulation of multiple target genes in *trans*-eQTL regulatory networks seems to be a general property of many known *trans*-eQTL signals[6,12,14]. However, what proportion of *trans*-eQTL target genes directly mediate the disease or trait associations as opposed to being independent ‘bystanders’ with minimal direct causal effect has remained unclear.

We performed the largest *trans*-eQTL meta-analysis in a single cell type, comprising 3,734 lymphoblastoid cell line (LCL) samples across nine cohorts (MetaLCL). LCLs are obtained by transforming primary B-cells with Epstein-Barr virus[22]. LCLs have been widely used as a

resource for human genetics, from banking cells from rare genetic disorders, through control material in laboratories to prevent repetitive blood sampling, to the study of tumorigenesis, mechanisms of viral latency and immune evasion[22]. Furthermore, Epstein-Barr virus has been epidemiologically linked to several autoimmune diseases in which B cells are implicated to play a pathogenic role, such as multiple sclerosis (MS)[23,24] and systemic lupus erythematosus (SLE)[25] with recent studies starting to elucidate the potential molecular mechanisms underlying these associations[26–28]. Thus, *trans*-eQTLs discovered in LCLs might provide insights into the pathogenesis of these autoimmune diseases.

After stringent quality control, we identify four highly robust *trans*-eQTL associations that replicate in an independent cohort (n=682) and are associated with multiple target genes. One of these signals corresponds to a missense variant in the *USP18* gene and is also associated with increased risk of SLE. The SLE risk allele is associated with increased activity of the type I interferon signalling pathway and increased expression of several classical interferon response genes. While there is robust evidence for the potential causal role of increased interferon signalling in SLE pathogenesis, we find that the expression of most individual interferon response genes is unlikely to have a direct causal effect on SLE. Our results caution against blindly using *trans*-QTL associations for target gene prioritisation without clear understanding of the *trans*-QTL mechanism and robust genetic evidence from *cis*-acting variants implicating the same gene. To support secondary use of our data, we have made the complete MetaLCL *trans*-eQTL summary statistics for 18,792 genes publicly available via the eQTL Catalogue FTP server.

Results

Large-scale *trans*-eQTL meta-analysis in a single cell type

We performed a large-scale *trans*-eQTL meta-analysis, utilising data from LCLs collected from 3,734 donors across nine cohorts of European ancestries (Table S1). After excluding *cis* associations located within 5 Mb of the target gene, we identified 79 suggestive independent *trans*-eQTL loci at p-value < $1x10^{-11}$ threshold (Figure 1). To identify robust signals associated with multiple target genes and reduce the risk of false positives caused by cross-mappability[29], we further required each locus to be associated with at least five independent target genes ($p < 5x10^{-8}$) with low cross-mappability scores (see Methods). This filtering reduced the number of candidate loci to six (Figure 1), four of which replicated in an independent multi-ethnic cohort of 682 individuals[30]. These four replicating *trans*-eQTL loci were located near the *BATF3*, *MYBL2*, *USP18*, *HNF4G* genes (Table S2, Figure S2). While the strong *trans*-eQTL signal near the *BATF3* transcription factor (2294 targets at FDR 5%) has been previously reported[31], the other three seem to be novel. Remarkably, the *trans*-eQTL targets at the *MYBL2* locus were consistent with direct activation by the MYBL2 transcription factor (Supplementary Note), indicating that our analysis is identifying biologically interpretable signals.

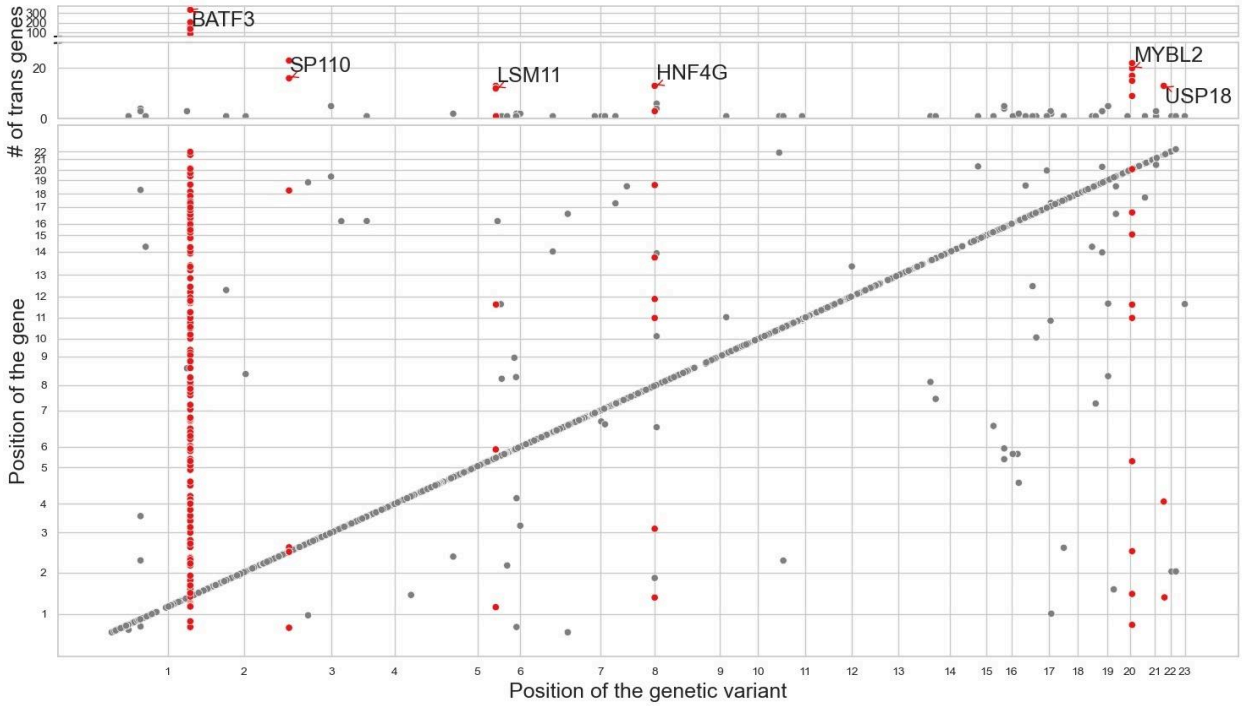

Figure 1. Overview of MetaLCL *trans*-eQTL results. The upper scatter plot shows the number of *trans* associations detected at each *trans*-eQTL locus with p-values < $5x10^{-8}$. Six largest *trans*-eQTL loci have been labelled with the name of the closest *cis* gene. The lower scatter plot shows all significant loci for each tested gene at the more stringent p < $1x10^{-11}$ threshold. *Cis* associations are located on the diagonal while putative *trans* associations are located off diagonal.

Missense variant in *USP18* affects lupus risk via negative regulation of interferon response

For the four high-confidence loci, we performed GWAS lookup using the Open Targets Genetics Portal[32]. We found that our *trans*-eQTL lead variant in the *USP18* locus (chr22_18166589_T_C, rs4819670) was shared with a GWAS lead variant identified for SLE in East Asians[33]. Using the point estimation of colocalisation (POEMColoc) method, we also confirmed that the two signals colocalised (PP4 = 0.97) (Figure 2A)[34]. At this locus, we identified 40 *trans* target genes at FDR 5% that were all strongly enriched for Reactome interferon signalling (R-HSA-913531, p = $1.1x10^{-26}$) and interferon alpha/beta signalling (R-HSA-909733, p = $1.7x10^{-21}$) pathways. The rs4819670-C allele was associated with decreased expression of multiple canonical type I interferon response genes (e.g. *ISG15*, *IFI44*, *OAS1-3*) (Figure 2B). Reassuringly, we observed consistent effect sizes across nine sub-cohorts in our meta-analysis (I^2 heterogeneity statistic = 0.46, Figure S2). *USP18* is a known negative regulator of interferon signalling and a rare loss-of-function mutation in *USP18* causes severe type I interferonopathy (Figure 2C)[35,36]. The rs4819670-C was also associated with decreased risk of systemic lupus erythematosus (SLE) in East Asians[33]. Furthermore, the rs4819670 lead variant is in perfect linkage disequilibrium (LD) (r^2 = 1, 1000 Genomes EAS superpopulation) with a *USP18* missense variant rs3180408

(chr22_18167915_C_T, ENSP00000215794.7:p.Thr169Met). While the GWAS association was previously known, it remained uncertain which allele of the missense variant rs3180408 was more likely to decrease USP18 protein function, especially because the variant was predicted to be benign by all tested variant effect prediction tools available from Ensembl VEP[37]. Our results suggest that the rs3180408-T SLE risk allele decreases USP18 protein function as USP18 is a negative regulator of type I interferon response genes.

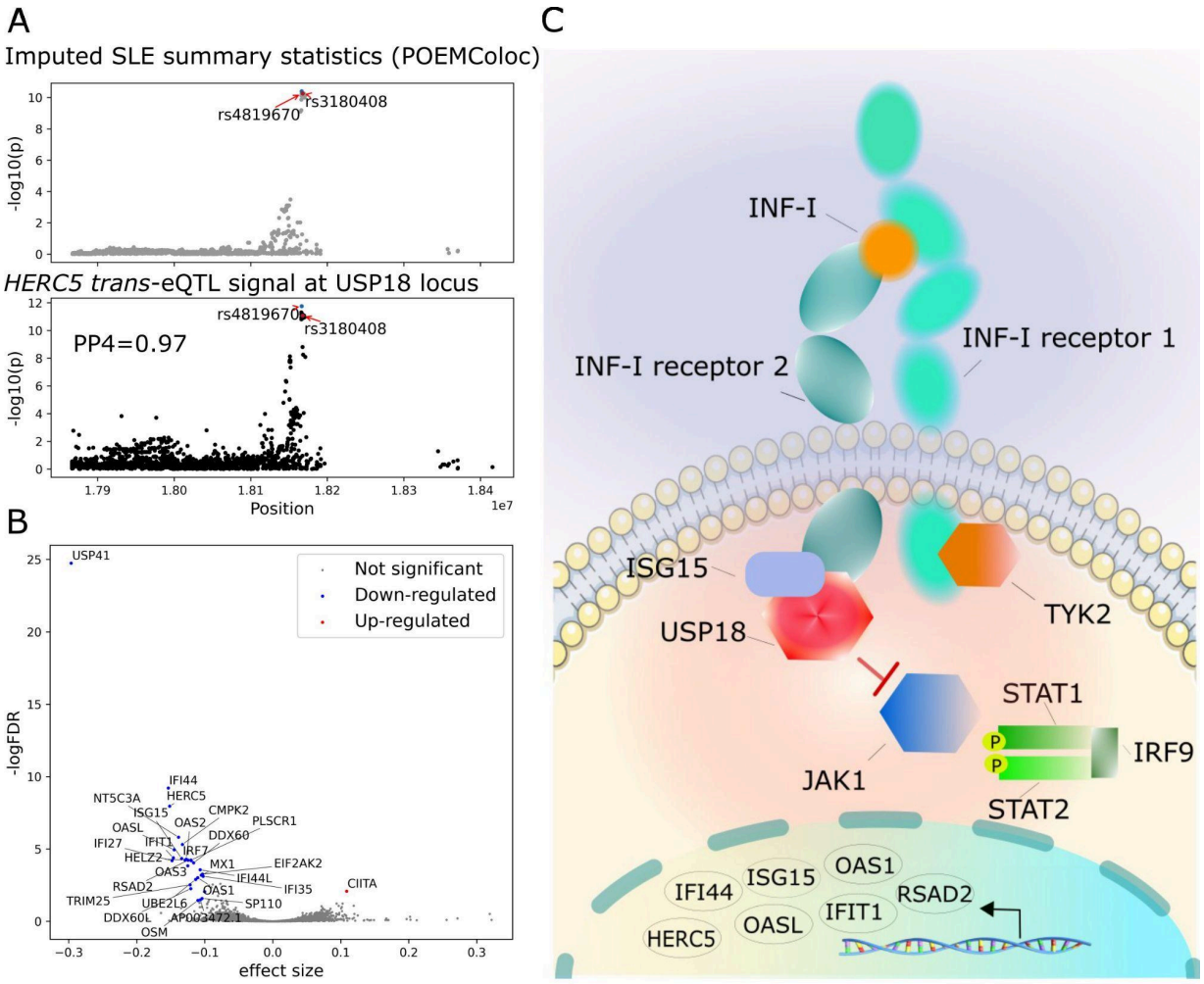

Figure 2. **SLE GWAS association at the *USP18* locus is a *trans*-eQTL for interferon response genes.** (**A**) Regional association plot for the SLE GWAS with POEMColoc imputed summary statistics and regional association plot for the lead *trans*-eQTL gene (*HERC5*) at the USP18 locus. The *trans*-eQTL lead and GWAS lead variants (rs4819670, shown in blue) are identical and in perfect LD with a missense variant (rs3180408, shown in red) in the *USP18* gene. The original regional association plot for the SLE GWAS is shown on Figure S4. (**B**) Volcano plot of the *trans*-eQTL target genes. (**C**) Schematic illustration of the role of USP18 in the regulation of interferon response genes, adapted from Alsohime *et al*[36].

Role of aberrant interferon signalling in lupus pathogenesis

Several studies have suggested that causal GWAS genes are enriched in shared pathways or biological processes[38–40]. To further characterise the potential role of USP18 target genes in lupus, we performed additional *trans*-eQTL meta-analysis across the nine discovery cohorts and one replication cohort (total n = 4,416). This increased the number of significant USP18 target genes to 50 (FDR < 5%). Notably, 18/50 target genes overlapped the Reactome interferon alpha/beta signalling (R-HSA-909733) pathway (hypergeometric test, $p = 4.14x10^{-24}$) and 26/50 genes overlapped a consensus set of interferon response genes (n = 124) identified by Mostafavi *et al*[41] ($p = 1.44x10^{-39}$, Table S4). Reassuringly, 39/50 genes were also more highly expressed in peripheral blood mononuclear cells from SLE cases compared to controls[42] (Table S4), consistent with the established role of increased interferon signalling in SLE[43].

To better understand the role of the USP18 target genes in the interferon alpha/beta signalling pathway, we focussed on the 60 genes belonging to the Reactome R-HSA-909733 interferon alpha/beta signalling pathway and divided them into three categories - category I: proteins involved in signal transduction via IFNAR1/2 receptor (n = 13 genes, including the multi-gene interferon-alpha gene cluster, Figure 3A); category II: downstream transcriptional targets of the interferon signalling (38 genes from the Reactome R-HSA-1015702 sub-pathway, Figure 3B) and 3) and category III: other pathway genes (n = 9) not belonging to the first two categories (Figure S5). We found that 16/50 USP18 targets overlapped with the 38 category II genes (transcriptional targets of interferon response) ($p = 1.18x10^{-29}$). In contrast, only 2/50 USP18 *trans*-eQTL target genes (*STAT1* and *ISG15*) overlapped the 13 category I genes (IFNAR1/2 receptor signal transduction proteins) and none overlapped the 9 category III genes. This suggests that the USP18 *trans*-eQTLs are primarily capturing the transcriptional targets of interferon response (category II), consistent with the established role of USP18 in regulating these genes (Figure 2C)[36].

Next, we assessed if there were additional lupus GWAS signals overlapping the three categories of interferon response genes defined above. We first used the Open Targets Genetics portal to extract the prioritised target genes for 108 lupus GWAS loci from Yin *et al*[33]. This revealed that three prioritised genes (*USP18*, *STAT1*, *IFNA1-17)* overlapped with the 13 category I genes (IFNAR1/2 signal transduction proteins, Figure 3A) and four prioritised genes (*IRF1/5/8* and *OAS1)* overlapped with the 38 category II genes (transcriptional targets of interferon response, Figure 3B). Out of these, IRF1/5/8 are themselves transcription factors involved in the regulation of interferon production[44,45], and stronger IRF1 binding across many GWAS loci has been associated with higher Crohn's disease risk[46]. Only *OAS1* represents a classical antiviral gene and here the GWAS lead variant is in perfect LD with a fine-mapped splice QTL for *OAS1* in the eQTL Catalogue (Figure S6)[47]. Interestingly, while the USP18 *trans*-eQTL risk allele increased *OAS1* expression (Figure 3B), the *cis* splice QTL risk allele (rs10774671-A) increased the expression of a transcript with an alternative 3' end that was associated with lower OAS1 protein abundance (Figure S6)[48,49], suggesting that *cis* and *trans* effects on the *OAS1* gene have opposite direction of effect on lupus risk.

We also overlapped interferon alpha/beta signalling pathway genes with ongoing or completed phase III clinical trials for SLE extracted from the ChEMBL database[50]. We identified three category I (interferon signal transduction) genes (*IFNAR1*, *JAK1* and *TYK2*) that have been targeted by a clinical trial for SLE (Figure 3A). While the trials targeting *JAK1* and *TYK2* are currently ongoing, a randomised control trial of anifrolumab, a human monoclonal antibody to type I interferon receptor subunit 1 (*IFNAR1)*, found it to be an effective treatment for SLE[51]. None of the category II genes (transcriptional targets of interferon signalling, Figure 3B) and category III genes (Figure S5) are currently in a phase III clinical trial for SLE (Figure 3B).

There is an emerging consensus that rare mutations in genes prioritised for autoimmune diseases from GWAS studies can often also cause primary immunodeficiencies (PIDs)[52,53]. For example, loss-of-function mutations in *USP18* cause rare type I interferonopathy[35,36]. At the same time, GWAS studies for SLE and other autoimmune diseases are still only powered to detect variants with large effects. Thus, knowing if a gene causes PID might be a useful (if noisy) indicator that the same gene might be discovered in a future larger autoimmune GWAS study. Thus, we obtained the list of genes causing either PID or monogenic inflammatory bowel disease from Genomics England[54] and overlapped those with the three categories of interferon response genes defined above. We found that 10/13 category I genes (interferon signal transduction) have previously been implicated in causing PID, including *USP18* and all three phase III drug candidates for SLE (Figure 3A). In contrast, only 8 of the 38 category II genes (transcriptional targets of interferon response) have been implicated in PIDs (Figure 3B), including *OAS1* and IRF8 also detected by SLE GWAS. Finally, none of the category III genes have been implicated in PIDs (Figure S5).

Triangulation of evidence from prioritised lupus GWAS target genes, phase III clinical trial information and overlap with primary immunodeficiency genes highlights modulation of aberrant interferon alpha/beta signalling in B-cells as an emerging therapeutic opportunity for SLE (category I, Figure 3A). This is further supported by recent studies demonstrating that depleting autoreactive B-cells via anti-CD19 CAR T cell therapy is an effective therapy for SLE and other autoimmune diseases[55,56]. In contrast, most *trans*-eQTL targets of USP18 overlap transcriptional targets of interferon response (category II, Figure 3B) and it is far less clear what is the potential causal roles of these genes in SLE pathogenesis.

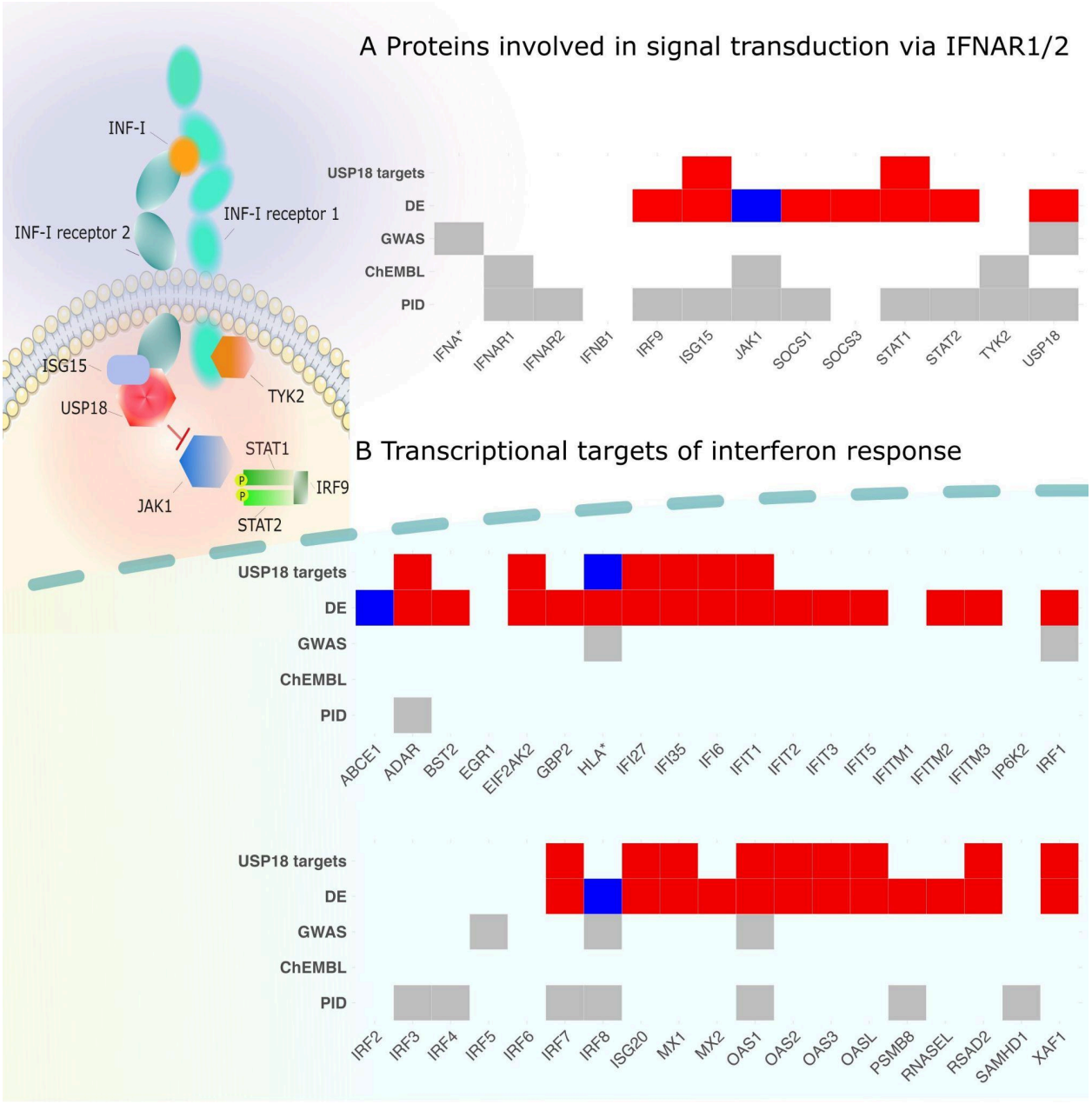

Figure 3. **A** Upstream regulators of interferon response genes (IFNA* contains multi-gene interferon-alpha gene cluster). **B** Downstream transcriptional targets of the interferon signalling (HLA* marks the HLA region). The increased gene expression is marked in red, while reduced gene expression is marked in blue. The visualisation illustrates the effect on USP18 target genes in relation to the SLE risk allele. DE - differential gene expression in SLE cases *versus* controls[42]; GWAS - GWAS hits for SLE[33], ChEMBL, phase III - SLE phase III clinical trials from ChEMBL[50], PID - genes causing primary immunodeficiency from Genomics England.

Replication of the *USP18 trans*-eQTL signal in whole blood

To understand the context-specificity of the *USP18 trans*-eQTL signal, we performed additional replication in the eQTLGen Phase 2 *trans*-eQTL meta-analysis of up to 43,301 whole blood samples. We observed that the *USP18* missense variant rs3180408 was nominally associated ($p < 0.05$) with the expression of 7/50 USP18 target genes, including our lead target gene *HERC5* ($p = 0.037$) as well as canonical interferon response genes *IFI44* and *ISG15* (Table S5). For 6/7 nominally significant associations, the effect direction was concordant between the LCL and whole blood meta-analyses, but the effect size was an order of magnitude smaller in whole blood (Table S5). Thus, even at this very large sample size, the *USP18 trans*-eQTL signal would not have been discovered in whole blood.

To understand the potential reasons for the attenuated effect in whole blood, we compared the expression level of the *USP18* gene across 49 GTEx tissues. We found that *USP18* had the highest expression in LCLs (median transcripts per million (TPM) = 45.3) and one of the lowest in whole blood (median TPM = 0.46). Since *USP18* is itself an interferon response gene and LCLs are characterised by a strong interferon signature driven by active infection with the Epstein-Barr virus, we characterised the expression of *USP18* in naive B-cells as well as B-cells stimulated with interferon-alpha and TLR7/8 agonist R848 for 16, 40 and 64 hours. We found that the expression level of *USP18* in B-cells was upregulated by ~3.5-fold after 16 hours of stimulation and stayed elevated for at least 64 hours (Figure S7). This suggests that the very strong active interferon signalling and associated upregulation of *USP18* transcription in LCLs is required for the *trans*-eQTL signal to be detected.

Discussion

We performed the largest *trans*-eQTL study in a single cell type where we profiled the expression of 18,792 genes in 3,737 individuals from nine cohorts. We then replicated these findings in an independent multi-ancestry LCL cohort of 682 individuals. After careful quality control, we identified six independent loci that were associated with five or more target genes, and that were unlikely to be driven by cross-mappability artefacts. While we primarily focussed on the SLE-associated rs3180408 missense variant in the *USP18* gene in our analysis, we have publicly released the complete genome-wide summary statistics from our MetaLCL project via the eQTL Catalogue FTP server. In addition to disease-specific colocalisation applications, we expect that our summary statistics will motivate the development and application of novel summary-based aggregative *trans*-eQTL mapping methods[57–59].

Despite the strong evidence for the critical role of type I interferon response in SLE pathogenesis[42,43] and three active clinical trials, we were surprised to see that of the 50 USP18 target genes, only *OAS1* had an independent *cis*-association with SLE. Expanding the analysis to interferon response genes from Reactome further implicated IRF1/5/8 genes and the HLA region, but most interferon response genes were not detected in the SLE GWAS. One potential explanation for this could be the limited statistical power of the SLE GWAS that profiled 13,377 cases and 194,993 controls, identifying a total of 113 loci[33]. Furthermore, Liu *et al* demonstrated

that if multiple effector genes ('core' genes) are co-regulated by shared *trans* factors, with shared directions of effects (which seems to be the case for the interferon response genes), then nearly all heritability would be due to *trans* effects, further reducing the power to detect *cis*-acting signals at individual target genes[4].

However, interferon response involves rapid upregulation of a broad transcriptional regulatory network of genes with diverse biological functions, only a subset of which might have a direct causal effect on SLE. This is supported by the fact that among the 38 interferon response genes (category II), only *OAS1*, *ADAR*, *PSMB8*, *SAMHD1* and the IRF transcription factors have been implicated in causing primary immunodeficiencies (Figure 3B). The remaining interferon response genes might thus be better thought of as biomarkers of the complex effect of interferon signalling on multiple parts of the immune system[43,60]. This could also help to explain the apparent directionally discordant *cis* and *trans* effects for the *OAS1* gene, raising an intriguing possibility that to reduce lupus risk it might be important to have high baseline levels of *OAS1* (to possibly aid with viral clearance[49]) rather than increase its expression long term after activation of interferon signalling. Similarly, it has been previously shown that variants in the *IL6R* region that are associated with circulating C-reactive protein (CRP) concentrations, are also associated with coronary artery disease (CAD) risk[61], but variants in the *CRP* region are not[62]. Thus, plasma levels of CRP do not seem to have a direct causal effect on CAD risk, but can still act as a molecular readout (biomarker) of the *IL6R*-mediated inflammatory response that does seem to have a causal effect[63]. These observations suggest that widespread horizontal pleiotropy in gene regulatory networks could be a general property of *trans*-QTLs and could help explain why using *trans*-pQTL signals in Mendelian randomisation analysis has had low specificity for identifying known drug targets[64,65]. Instead, we propose that target genes identified from large-scale *trans*-QTL studies could be better thought of as drug response biomarkers for drugs targeting the *cis* gene responsible for the *trans* association[8].

A limitation of our *trans*-eQTL analysis is its susceptibility to cross-mappability artefacts (Table S3). While heuristic approaches have been developed to filter such artefacts *post hoc*, these approaches are not guaranteed to remove all cross-mappability effects and might be too conservative at other loci[29]. Cross-mappability artefacts also tend to replicate well in independent cohorts[29]. Furthermore, as the sample size of *trans*-eQTL studies increases, the power to detect subtle cross-mappability effects as putative *trans*-eQTLs also increases. To avoid these false positives, we used a very conservative strategy of requiring each *trans*-eQTL locus to have at least five independent target genes that all pass the cross-mappability filter. As a result, we likely missed many true *trans*-eQTLs regulating single or few target genes (e.g. *trans*-eQTL effect near the CIITA transcription factor on multiple HLA genes that has been replicated in several independent studies[6,31,66–68], Table S2). Future large-scale *trans*-eQTL studies will likely require the development of novel methods to properly adjust for cross-mappability, such as explicit modelling of transcript compatibility read counts between *cis* and trans *target* genes[69].

While large-scale *trans*-eQTL studies using both bulk and single-cell measurements are likely to continue for easily accessible tissues such as whole blood (e.g. eQTLGen Phase 2[70]), it seems

unlikely that we will be able to perform *trans*-eQTL studies comprising tens of thousands of individuals for all disease-relevant cell types and contexts. A promising alternative is to use arrayed CRISPR screens or single-cell approaches to identify downstream gene-regulatory effects of disease-associated genes or individual genetic variants[39,71,72].

Methods

Datasets, samples and ethics

We used genotype and gene expression data from ALSPAC[31,73,74], TwinsUK[75], CoLaus[76,77], GEUVADIS[78], MRCA[79], MRCE[79], GENCORD[80], GTEx v8[17] and CAP[81] studies. For replication, we used data from the MAGE cohort[30]. The RNA sequencing and genotype data from the GEUVADIS and MAGE studies was publicly available as part of the 1000 Genomes project. For the other studies, we applied for access to individual-level data via relevant data access committees (DACs), explaining the aim of our project and the intent to publicly share meta-analysis summary statistics. Informed consent was obtained when research participants joined the ten studies listed above. The use of the CAP data for this project was approved by the National Heart, Lung and Blood Institute DAC. The use of the GTEx data for this project was approved by the National Human Genome Research Institute DAC. The use of the GENCORD data for this project was approved by the GENCORD DAC. The use of the MRCA and MRCE data for this project was approved by the Gabriel Consortium DAC. The use of TwinsUK data for this project was approved by the TwinsUK Resource Executive Committee. The use of the ALSPAC data for this project was approved by the ALSPAC Executive Committee. For the ALSPAC cohort, ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). The CoLaus study was approved by the Institutional Ethics Committee of the University of Lausanne.
Single-cell RNA-seq samples were sourced ethically, and their research use was in accord with the terms of informed consent under an institutional review board/ethics committee-approved protocol (UK Regional Ethics Committee approval granted to work at Wellcome Sanger Institute, protocol reference number 15/NW/0282; project was approved by the Ethics on Research Committee of the Institute of Neurobiology at Universidad Nacional Autonoma de Mexico (UNAM), with the approval number 110.H.).

Genotype data quality control and imputation

Pre-imputation quality control. Genotype imputation was performed as described previously[47]. Briefly, we lifted coordinates of the genotyped variants to the GRCh38 build with CrossMap v0.4.1[82]. We aligned the strands of the genotyped variants to the 1000 Genomes 30x on GRCh38 reference panel[83] using Genotype Harmonizer[84]. We excluded genetic variants with Hardy-Weinberg p-value < 10^{-6}, missingness > 0.05 and minor allele frequency < 0.01 from further analysis. We also excluded samples with more than 5% of their genotypes missing.

Genotype imputation and quality control.

Most of the datasets were imputed using the 1000 Genomes reference panel based on the GRCh38 genome version. CoLaus dataset was imputed using the TOPMed Imputation Server[85–87], while still aligning with the same reference genome version. Additionally, GEUVADIS, GTEx and MAGE cohorts utilised whole genome sequencing data aligned to the GRCh38 reference genome.

We pre-phased and imputed the microarray genotypes to the 1000 Genomes 30x on GRCh38 reference panel[83] using Eagle v2.4.1[88] and Minimac4[86]. We used bcftools v1.9.0 to exclude variants with minor allele frequency (MAF) < 0.01 and imputation quality score R^2 < 0.4 from downstream analysis. The genotype imputation and quality control steps are implemented in eQTL-Catalogue/genimpute (v22.01.1) workflow available from GitHub. Subsequently, we used QCTOOL v2.2.0 to convert imputed genotypes from VCF format to bgen format for *trans*-eQTL analysis with regenie.

Gene expression data

Studies. We used gene expression data from seven RNA-seq studies (TwinsUK[75], CoLaus[76,77], GEUVADIS[78], GENCORD[80], GTEx v8[17], CAP[81], MAGE[30]) and three microarray studies (ALSPAC[31,73,74], MRCA[79] and MRCE[79]).

RNA-seq quantification and normalisation. RNA-seq data were pre-processed as described previously[89]. Briefly, quantification of the RNA-seq data was performed using the eQTL-Catalogue/rnaseq workflow (v22.05.1) implemented in Nextflow. Before quantification, we used Trim Galore v0.5.0 to remove sequencing adapters from the fastq files. For gene expression quantification, we used HISAT2[90] v2.2.1 to align reads to the GRCh38 reference genome (Homo_sapiens.GRCh38.dna.primary_assembly.fa file downloaded from Ensembl). We counted the number of reads overlapping the genes in the GENCODE V30 reference transcriptome annotations with featureCounts v1.6.4.

We excluded all samples that failed the quality control steps as described previously[89]. We normalised the gene counts using the conditional quantile normalisation (cqn) R package v1.30.0 with gene GC nucleotide content as a covariate. We downloaded the gene GC content estimates from Ensembl biomaRt and calculated the exon-level GC content using bedtools v2.19.0[91]. We also excluded lowly expressed genes, where 95 per cent of the samples within a dataset had transcripts per million (TPM)-normalised expression less than 1. Subsequently, we used the inverse normal transformation to standardise quantification estimates. Normalisation scripts together with containerised software are publicly available at https://github.com/eQTL-Catalogue/qcnorm.

Microarray data processing. Gene expression from 877 individuals in the ALSPAC cohort was profiled using Illumina Human HT-12 V3 BeadChips microarray. We used the normalised gene expression matrix from the original publication[31]. In the MRCA cohort, gene expression from 327 individuals was profiled using the Human Genome U133 Plus 2.0 microarray. We downloaded the raw CEL files from ArrayExpress (E-MTAB-1425) and normalised the data using the Robust Multi-Array Average (RMA) method from the affy Bioconductor package[92]. In the MRCE cohort,

gene expression from 484 individuals was profiled using the Illumina Human-6 v1 Expression BeadChip. As raw data was unavailable, we downloaded the processed gene expression matrix from ArrayExpress (E-MTAB-1428). In all three microarray datasets, we applied inverse normal transformation to each probe before performing *trans*-eQTL analysis. If there were multiple probes mapping to the same gene, the probe with the highest average expression was used.

Trans-eQTL mapping and meta-analysis

We performed independent quality control and normalisation on all datasets and only included 18,792 protein coding genes in the analysis. *Trans*-eQTL analysis was conducted separately on each dataset with regenie[93]. For studies containing related samples (TwinsUK, MRCA and MRCE) and ALSPAC, both step 1 and step 2 commands were employed, while for other datasets with a smaller number of unrelated samples (Table S1), regenie was run in the linear regression mode (step 2 only). We used sex and six principal components of the normalised gene expression matrix and six principal components of genotype data as covariates in the *trans*-eQTL analysis. All scripts used to run *trans*-eQTL are publicly available at https://github.com/freimannk/regenie_analysis. Subsequently, we performed an inverse-variance weighted meta-analysis across studies. Meta-analysis workflow is available at https://github.com/freimannk/regenie_metaanalyse.

We used a *cis* window of ± 5Mb to assign identified eQTLs into *cis* and *trans* eQTLs. To determine significant loci, we excluded variants in close proximity (±1.5Mb) to the most highly associated variant per gene. This approach allowed us to identify distinct and robust signals while mitigating potential confounding effects from nearby variants. By applying these filters we found 79 *trans*-eQTLs loci at a suggestive p-value threshold of $1x10^{-11}$.

Accounting for cross-mappability

A major source of false positives in *trans*-eQTL analysis is cross-mappability, whereby RNA-seq reads from gene A erroneously align to gene B, leading to very strong apparent *trans*-eQTL signals[11,29]. To exclude potential cross-mappability artefacts, we excluded all *trans*-eQTLs where there was high cross-mappability (cross-mappability score from Saha *et al*[29] > 1) between the *trans*-eQTL target gene and at least one protein coding gene in the *cis* region (±1.5Mb) of the *trans* eQTL lead variant. Since some of the strongest cross-mappability artefacts affected one or few target genes (Table S3), we further restricted our analysis to *trans*-eQTL loci that had five or more target genes with $p < 5x10^{-8}$ and cross-mappability score < 1.

Replication of *trans*-eQTL associations

MAGE. Since we used somewhat arbitrary thresholds to define the initial set of 10 loci (lead p-value $< 1x10^{-11}$, five or more targets with $p < 5x10^{-8}$), we sought to replicate our findings in an independent Multi-ancestry Analysis of Gene Expression (MAGE)[30] cohort. MAGE consisted of data from 731 lymphoblastoid cell lines from the 1000 Genomes project, 682 of which also had whole genome sequencing data available. We used two strategies to assess replication. First, we assessed if the lead variant-gene pair was nominally significant ($p < 0.05$) in the replication

dataset with concordant direction of effect. Based on this criterion, 7/10 loci replicated (Table S2). Secondly, since all of our loci had multiple target genes, we used the pi1 statistic to estimate the proportion of FDR < 5% target gene at each locus that had a non-null p-value in the replication dataset[94]. We used the qvalue R package[95] to calculate pi1 = 1-qvalue(5% FDR trans gene p-values)$pi0. For 3/10 loci, the proportion of non-null p-values was > 0.5 (Table S2). Note that replication in an independent cohort does not help to reduce false positives due to cross-mappability, as cross-mappability artefacts tend to be highly replicable[29].

eQTLGen Consortium. The eQTLGen Consortium is an initiative to investigate the genetic architecture of blood gene expression and to understand the genetic basis of complex traits. We used interim summary statistics from eQTLGen phase 2, wherein a genome-wide eQTL analysis has been performed in 52 cohorts, representing 43,301 individuals.

All 52 cohorts performed cohort-specific analyses as outlined in the eQTLGen analysis cookbook (https://eqtlgen.github.io/eqtlgen-web-site/eQTLGen-p2-cookbook.html). Genotype quality control was performed according to standard bioinformatics practices and included quality metric-based variant and sample filtering, removing related samples, ethnic outliers and population outliers. Genotype data was converted to genome build hg38 if not done so already and the autosomes were imputed using the 1000G 30x WGS reference panel (64) (all ancestries) using the eQTLGen imputation pipeline (eQTLGen/eQTLGenImpute).

Like the genotype data, gene expression data was processed using the eQTLGen data QC pipeline (eQTLGen/DataQC). For array-based datasets, we used the results from the empirical probe mapping approach from our previous study[6] to connect the most suitable probe to each gene which has previously been to show expression in the combined BIOS whole blood expression dataset. Raw expression data was further normalized in accordance with the expression platform used (quantile normalization for Illumina expression arrays and TMM[96] for RNA-seq) and inverse normal transformation was performed. Gene expression outlier samples were removed and gene summary information was collected for filtering at the central site. Samples for whom there were mismatches in genetically inferred sex, reported sex, or the expression of genes encoded from sex chromosomes were removed. Similarly, samples with unclear sex, based on genetics or gene expression were removed.

An adaptation of the HASE framework[97] was used to perform genome-wide meta-analysis. For genome-wide eQTLs analysis, this limits the data transfer size while ensuring participant privacy. At each of the cohorts, the quality controlled and imputed data was processed and encoded so that the individual level data can no longer be extracted, but while still allowing effect sizes to be calculated for the linear relationship between variants and gene expression (eQTLGen/ConvertVcf2Hdf5 and eQTLGen/PerCohortDataPreparations).
Centrally, the meta-analysis pipeline was run on the 52 cohorts. The pipeline which performs per cohort calculations of effect sizes and standard errors and the inverse variance meta-analysis is available at eQTLGen/MetaAnalysis. We included 4 genetic principal components, 20 gene expression principal components and other technical covariates (e.g. RNA integrity number) where available. Per every dataset, genes were included if the fraction of unique expression

values was equal or greater than 0.8, Variants were included based on imputation quality, Hardy-Weinberg equilibrium (HWE) and minor allele frequency (MAF) (Mach R2 ≥ 0.4, HWE p ≥ 1×10^{-6} and MAF ≥ 0.01). In an additional step, genes were filtered to include only those genes that were available in at least 50% of the cohorts and 50% of the samples.

Differential gene expression in SLE cases *versus* controls

We re-analysed the microarray gene expression data from Banchereau et al. 2016[42] to explore differential gene expression between SLE cases and controls. After downloading the processed data from GEO (GSE65391), we selected one sample from each individual for our analysis based on their earliest recorded visits. The filtered dataset comprised a total of 204 samples, including 46 samples from healthy individuals and 158 samples from individuals diagnosed with SLE. We also applied the inverse normal transformation to standardise the gene expression values. Subsequently, we used the Python statsmodels[98] module to fit a linear model to identify genes that were differentially expressed between SLE cases and controls. We included gender, age and batch as covariates in all models.

Overlap between *USP18* target genes and GWAS hits for SLE

We download the list of prioritised target genes for the Yin *et al* GWAS study (GCST011956) from the Open Targets Genetics Portal. We combined the list of genes prioritised by either the L2G or the closest gene approach, yielding n = 109 target genes. We then overlapped these target lists with the list of 50 *trans*-eQTL targets for the *USP18* locus (FDR < 5%).

Single-cell differential gene expression in resting and stimulated B-cells

Sample collection, cell isolation and cryopreservation. Blood samples were collected from five healthy Mexican individuals (three males and two females). Peripheral blood mononuclear cells (PBMCs) were isolated using Vacutainer CPT tubes, according to manufacturer instructions. Samples were cryopreserved in RPMI 1640 culture media (Sigma), Fetal Bovine Serum (FBS) and Dimethyl sulfoxide (DMSO) and stored at -80ºC for 24h, before being transferred to liquid nitrogen.

Thawing and stimulation. Cryopreserved PBMCs were thawed quickly and washed in 14mL of room temperature complete RPMI 1640 media (10% FBS, 1% Penicillin-Streptomycin, 1% L-Glutamine). Cells were incubated at 37 °C, 5% CO_2 for 2h. Cells were then stimulated with interferon alpha (IFN-α, Bio-techne) and R-848 (Resiquimod, Cambridge Bioscience) at a working concentration of 1000U/mL and 2µg/ml, respectively. Cells were incubated at 37 °C, 5% CO2 and harvested after 16h, 40h and 64h of stimulation. Unstimulated cells were kept in culture without any stimuli for 16h (i.e., 0h of activation).

Multiplexing, CITE-seq staining & scRNA-seq. Upon harvesting, cells were resuspended in a cell staining buffer (Biolegend) and cell hashing and genotype-based multiplexing was performed. Donors of the same stimulation condition were mixed at equal ratios (each pool corresponded to a mix of cells from four to five different individuals). These pools were stained

with the TotalSeq-C Human Universal Cocktail, V1.0 (137 cell surface proteins (CSP), Biolegend), in addition to a unique hashtag antibody oligonucleotide (HTO, Biolegend) which corresponds to the stimulation condition pool. After staining and washing, all stimuli condition pools were pooled together at equal ratios. This pool was then stained with live/dead dye 4,6-diamidino-2-phenylindole (DAPI, Biolegend) and dead cells were removed using fluorescence-activated cell sorting.

Cells were next processed using the 10X Genomics Immune Profiling 5' high-throughput (HT) v2 kit, as specified by the manufacturer's instructions. $1.15x10^5$ cells were loaded into each inlet of a 10X Chromium X to create Gel Bead-in-emulsions (GEMs). Two 10X HT reactions were loaded per time point of sample processing (targeted recovery was 40,000 cells per 10X reaction). Reverse transcription was performed on the emulsion, after which cDNA and CITEseq supernatant were purified, amplified and used to construct RNA-sequencing and CSP sequencing libraries, respectively. These RNA and CSP libraries were sequenced at a 5:1 ratio, respectively, using the Illumina NoveSeq 6000 S4, with 100-bp paired-end reads and all 10X reactions were mixed at equal ratios and sequenced across two lanes.

Deconvolution of single cells by genotype. Each 10X reaction comprised a mix of cells from unrelated individuals. Thus, natural genetic variation was used to assign cells to their respective individuals. First, a list of common exonic variants was compiled from the 1000 Genomes Project phase 3 exome-sequencing data (MAF> 0.05). Next, cellSNP (v1.2.1) was used to generate pileups at the genomic location of these variants. These pileups, in combination with the variants called from genotyping in each individual, were used as an input for Vireo[99] (v0.5.7). If any cell had less than 0.9 posterior probability of belonging to any individual or were of mixed genotypes they were labelled as 'unassigned' and 'doublets', respectively, and removed from downstream analysis.

Data processing and quality controls. Raw scRNA-seq and CITE-seq data were processed using the Cell Ranger Multi pipeline (v7.0.0, 10x Genomics). In brief, RNA and CSP library reads were first assigned to cells. RNA reads were then aligned to the GRCh38 human reference genome and CSP antibody reads were matched to the provided list of known barcodes. Ensembl version 93 was used as a reference for gene annotation, and gene expression was quantified using reads assigned to cells and confidently mapped to the genome. Additionally, Cell Ranger multi was used to deconvolute samples based on HTOs. It uses an algorithm which employs a latent variable model over a state space composed of each HTO used in the experiment to assign each cell to a stimulation condition or as a doublet.

Results from RNA and CSP quantification in Cell Ranger were imported into RStudio (v4.3.1) and analysed using Seurat (v5.0.1). Any cell identified as doublet or unassigned by Vireo and or antibody hashtag deconvolution method were excluded. 10X reactions were split by time point and stimuli condition. Cells with 1.5 - 2.5 median absolute deviations below the median of genes and counts detected were discarded. Additionally, cells with 3 - 4 median absolute deviation above the median for the percentage of mitochondrial reads detected were discarded. The

resulting cells were then annotated by Azimuth[100] (v0.5.0), using the Azimuth PBMC reference dataset that was generated as part of the Hao and Hao *et al*, 2021 paper[100].

Pseudobulk and normalisation. Raw counts were pseudobulked by Azimuth annotated level 1 cell types (CD4, CD8, B, Mono, DC, NK, Other and Other T) per donor, per time point and per stimulation condition, via edgeR[101] (v4.0.16). Pseudobulked raw counts were then counts per million (CPM) normalised and $\log_2$ transformed with edgeR.

Data availability

The whole genome sequencing data for the GEUVADIS and MAGE studies was downloaded from the 1000 Genomes website. The GEUVADIS RNA-seq data was downloaded from the European Nucleotide Archive (ENA) under accession PRJEB3366. The MAGE RNA-seq data was downloaded from the ENA (accession PRJNA851328). The genotype and RNA-seq data from the GENCORD study was downloaded from European Genotype-phenotype Archive (EGA) under accessions EGAD00001000425 and EGAD00001000428. The microarray gene expression data from the MRCA and MRCE studies was downloaded from ArrayExpress (E-MTAB-1425 and E-MTAB-1428) and the genotype data was downloaded from EGA (EGAS00000000137). The gene expression and genotype data from GTEx and CAP studies was downloaded from dbGaP (accessions phs000424.v8.p2 and phs000481.v3.p2). The RNA-seq data from the TwinsUK study was downloaded from EGA (EGAD00001001086) and genotype data was obtained from TwinsUK (https://twinsuk.ac.uk/resources-for-researchers/access-our-data/). The informed consent obtained from ALSPAC participants does not allow the microarray and genotype data to be made freely available through any third party maintained public repository. However, data used for this study can be made available on request to the ALSPAC Executive. The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access. Full instructions for applying for data access can be found here: http://www.bristol.ac.uk/alspac/researchers/access/. The ALSPAC study website contains details of all the data that are available (http://www.bristol.ac.uk/alspac/researchers/our-data/). The RNA-seq and genotype data from the CoLaus cohort can be accessed by directly contacting the cohort (https://www.colaus-psycolaus.ch/professionals/how-to-collaborate/). The MetaLCL full *trans*-eQTL meta-analysis summary statistics are available from the eQTL Catalogue FTP server (https://www.ebi.ac.uk/eqtl/Data_access/) and additional documentation is available on the project website (https://github.com/AlasooLab/MetaLCL).

URLs

MetaLCL website: https://github.com/AlasooLab/MetaLCL
MetaLCL *trans*-eQTL analysis workflow: https://github.com/freimannk/regenie_analysis
MetaLCL meta-analysis workflow: https://github.com/freimannk/regenie_metaanalyse
eQTL Catalogue website: https://www.ebi.ac.uk/eqtl/Data_access/
eQTL Catalogue genotype imputation workflow: https://github.com/eQTL-Catalogue/genimpute

eQTL Catalogue RNA-seq processing workflow: https://github.com/eQTL-Catalogue/rnaseq
eQTL Catalogue data normalisation workflow: https://github.com/eQTL-Catalogue/qcnorm
eQTLGen analysis cookbook: https://eqtlgen.github.io/eqtlgen-web-site/eQTLGen-p2-cookbook.html
eQTLGen data QC workflow: https://github.com/eQTLGen/DataQC
eQTLGen genotype conversion workflow: https://github.com/eQTLGen/ConvertVcf2Hdf5
eQTLGen per-cohort analysis workflow: https://github.com/eQTLGen/PerCohortDataPreparations
eQTLGen meta-analysis workflow: https://github.com/eQTLGen/MetaAnalysis
eQTLGen genotype imputation workflow: https://github.com/eQTLGen/eQTLGenImpute
QCTOOL: https://www.chg.ox.ac.uk/~gav/qctool_v2/

Declaration of interests

The authors declare no competing interests.

Acknowledgements

Most of the analysis presented in the paper were performed at the High Performance Computing Center, University of Tartu. We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. This work received support from Christian Molina-Aguilar, Carina Uribe Díaz, and Alejandra Castillo Carbajal. We thank the FACS facility and sequencing pipelines at Wellcome Sanger Institute for their assistance in data generation, and the Human Genetics Informatics for their support in sequence data processing. We thank Bess L. Chau for assistance in the experiments involving generation of single-cell RNAseq data from stimulated PBMCs. We thank Karatug Ozan Bircan, Abayomi Mosaku, Laura Harris and Helen Parkinson for assistance with hosting the full *trans*-eQTL summary statistics on the eQTL Catalogue FTP server.

Funding

K.F and L.S. were supported by a grant from Open Targets (grant no. OTAR2069). K.A was supported by the Estonian Research Council (grant no. PSG415). K.A. also received funding from the European Union's Horizon 2020 research and innovation program (grant no. 825775). GT, CPJ and TSR were supported by Open Targets (grant no. OTAR2064) and the Wellcome Grant (ref. 220540/Z/20/A, 'Wellcome Sanger Institute Quinquennial Review 2021-2026). SB was supported by the Swiss National Science Foundation (grant no. 310030_152724/1). K.K. was supported by the Estonian Research Council (grant no. PRG1117). AMR was supported by CONACYT-FORDECYT-PRONACES grant no. [11311] and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica–Universidad Nacional Autónoma de México (PAPIIT-UNAM) IN218023. ALHL is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and she receives

fellowship 790972 from Consejo Nacional de Humanidades, Ciencias y Tecnologías CONAHCYT, México. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the work of the authors and they will serve as guarantors for the contents of this paper.

Author contributions

K.F. performed *trans*-eQTL analysis on nine cohorts and also performed the meta-analysis across cohorts. A.B. performed *trans*-eQTL analysis on the CoLaus cohort. K.F., J.C., E.R.H., J.C.M., N.N., H.O., L.S., M.C.T., G.T., K.K. and K.A. interpreted the results and prioritised follow-up analyses. T.S.R. generated and analysed single-cell RNAseq data under C.P.J. and G.T. supervision. A.L.H.-L. collected blood samples for single-cell RNAseq under A.M.-R. supervision. R.W. performed replications in the eQTLGen Consortium under U.V. and L.F. supervision. S.B. and K.A. supervised the research. K.F. and K.A. wrote the manuscript with feedback from all authors.

Supplementary Information

eQTLGen Consortium – Author information

Author list is ordered alphabetically.

Habibul Ahsan[1], Marta E. Alarcón-Riquelme[2,3], Philip Awadalla[4], Guillermo Barturen[5,6,2], Alexis Battle[7], Frank Beutner[8], Cornelis Blauwendraat[9,10], Collins Boahen[11,12], Toni Boltz[13], Dorret Boomsma[14,15,16], Andrew Brown[17], John Budde[18,19], Katie L. Burnham[20], John Chambers[21,22,23], Evans Cheruiyot[24], Surya B. Chhetri[25], Annique Claringbould[26,27], PRECISESADS Clinical Consortium[2], DIRECT Consortium[28], Carlos Cruchaga[18,29,30,31,32,33], Kensuke Daida[9,10,34], Emma E. Davenport[20], Devin Dikec[18,19], Théo Dupuis[17], Diptavo Dutta[35], Tõnu Esko[36], Radi Farhad[37], Aiman Farzeen[38,39], Marie-Julie Favé[40], Lude Franke[26,41], Tim Frayling[42], Koichi Fukunaga[43], J. Raphael Gibbs[44], Greg Gibson[45], Priyanka Gorijala[18,19], Binisha Hamal Mishra[46,47,48], Takanori Hasegawa[49], Michael Inouye[50,51,52,53,54,55], Matt Johnson[18,19], Holger Kirsten[56,57], Julian C. Knight[58], Peter Kovacs[59,60], Knut Krohn[61], Viktorija Kukushkina[36], Vinod Kumar[11,12,62,63], Mika Kähönen[47,64], Sandra Lapinska[65], Terho Lehtimäki[46,47,48], Yun Li[66,67,68], Markus Loeffler[69,70], Marie Loh[21,71,72,23], Leo-Pekka Lyytikäinen[46,47,48], Javier Martin[73], Angel Martinez-Perez[74], Allan McRae[24], Lili Milani[36], Pashupati P. Mishra[46,47,48], Younes Mokrab[37], Grant Montgomery[24], Juha Mykkänen[75,76], Reedik Mägi[36], Martina Müller-Nurasyid[77], Haroon Naeem[37], Sini Nagpal[45], Ho Namkoong[78], Matthias Nauck[79], Yukinori Okada[80,81,82,83,84], Roel Ophoff[65,13,85], Katja Pahkala[76,75,86], Bogdan Pasaniuc[65,87,13,88], Dirk S. Paul[50,51,89], Elodie Persyn[50,51,90], Brandon Pierce[1], René Pool[91,15], Holger Prokisch[92,93], Laura Raffield[67], Venket Raghavan[69], Olli Raitakari[94,95,75,96], Emma Raitoharju[97,98], Jansen Rick[15,99], María Rivas-Torrubia[2], Ruth D. Rodríguez[2], Suvi P. Rovio[76,75], Jessie Sanford[18,19], Markus Scholz[69,70], Eline Slagboom[100], José Manuel Soria[74], Juan Carlos Souto[74], Michael Stumvoll[101,102,103], Yun Ju Sung[18,19], Darwin Tay[21], Alexander Teumer[104,105], Joachim Thiery[70,106], Alex Tokolyi[20], Lin Tong[1], Anke Tönjes[101], Jan Veldink[107], Joost Verlouw[108], Peter M. Visscher[24], Ana Viñuela[109], Urmo Võsa[36], Uwe Völker[110,111], Qingbo S. Wang[80,81,82], Robert Warmerdam[26,41], Stefan Weiss[112,113], Jia Wen[67], Harm-Jan Westra[26,41], Andrew Wood[42], Manke Xie[45], Dasha Zhernakova[26], Marleen van Greevenbroek[114], Joyce van Meurs[115,116]

1. Department of Public Health Sciences, University of Chicago, Illinois, USA
2. Pfizer–University of Granada–Junta de Andalucía Centre for Genomics and Oncological Research, Granada, Spain
3. Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden
4. Ontario Institute for Cancer Research, University of Toronto
5. Department of Genetics, Faculty of Science, University of Granada, 18071 Granada, Spain
6. Bioinformatics Laboratory, Biotechnology Institute, Centro de Investigación Biomédica, PTS, Avda. del Conocimiento s/n, 18100 Granada, Spain
7. Department of Biomedical Engineering, Department of Computer Science, Department of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA
8. Department of Internal Medicine/Cardiology, Heart Center Leipzig at Leipzig University, Leipzig, Germany
9. Integrative Neurogenomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
10. Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
11. Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands
12. Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

13. Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
14. Amsterdam Reproduction & Development (AR&D) research institute, Amsterdam, the Netherlands
15. Amsterdam Public Health research institute, Amsterdam, the Netherlands
16. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam, Vrije Universiteit Amsterdam
17. Population Health and Genomics, University of Dundee, Dundee, Scotland, UK.
18. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States
19. NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO 63110, USA
20. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK
21. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
22. Precision Health Research (PRECISE), Singapore
23. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom
24. Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
25. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
26. Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
27. Department of Internal Medicine, Erasmus MC, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands
28. https://directdiabetes.org/
29. NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO 63110, USA
30. Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA
31. Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, MO, United States
32. Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, United States
33. Dominantly Inherited Alzheimer Disease Network (DIAN)
34. Department of Neurology, Faculty of Medicine, Juntendo University, Tokyo, Japan
35. Division of Cancer Epidemiology & Genetics, National Cancer Insititute, Bethesda, MD, USA
36. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
37. Human Genetics Department, Sidra Medicine, Doha, Qatar
38. Institute of Human Genetics, School of Medicine, Technische Universität München, Munich, Germany
39. Institute of Neurogenomics, Computational Health Center, Helmholtz Zentrum München, Munich, Germany.
40. Ontario Institute for Cancer Research
41. Oncode Institute, Groningen, The Netherlands
42. Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK
43. Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan.
44. Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
45. Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA, USA.
46. Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
47. Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

48. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
49. M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan.
50. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
51. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK
52. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
53. Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia
54. Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
55. British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK.
56. LIFE – Leipzig Research Center for Civilization Diseases, Leipzig University, Germany
57. Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Germany
58. Centre for Human Genetics, University of Oxford, Oxford, UK
59. Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, D-04103 Leipzig, Germany
60. Department of Medicine, University of Leipzig, D-04103 Leipzig, Germany
61. Medical Faculty, University of Leipzig, Leipzig, Germany.
62. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, the Netherlands
63. Nitte (Deemed to Be University), Medical Sciences Complex, Nitte University Centre for Science Education and Research (NUCSER), Deralakatte, Mangalore, 575018, India
64. Department of Clinical Physiology, Tampere University Hospital, Tampere Finland
65. Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA
66. Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
67. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
68. Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
69. Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Leipzig, Germany
70. Leipzig Research Centre for Civilization Diseases, Leipzig University, Leipzig, Germany
71. National Skin Centre, Research Division, Singapore
72. Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore
73. Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), Granada, Spain
74. Unit of Genomic of Complex Diseases, Institut de Recerca Sant Pau (IR Sant Pau), Barcelona, Spain.
75. Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland
76. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
77. Institute for Medical Biometry, Epidemiology and Informatics (IMBEI), Mainz, Germany
78. Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan.
79. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
80. Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan
81. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
82. Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

83. Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan
84. Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan.
85. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA
86. Paavo Nurmi Centre, Unit of Health and Physical Activity, University of Turku, Turku, Finland
87. Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
88. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
89. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK.
90. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, UK.
91. Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
92. Institute of Neurogenomics, Computational Health Center, Helmholtz Zentrum München, Neuherberg, Germany
93. School of Medicine, Institute of Human Genetics, Technical University of Munich, Munich, Germany
94. Research centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
95. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
96. InFLAMES Research Flagship, University of Turku, Turku, Finland
97. Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
98. Tampere University Hospital, Tampere, Finland
99. Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry & Amsterdam Neuroscience -Complex Trait Genetics (VUmc) and Mood, Anxiety, Psychosis, Stress & Sleep
100. Section of Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands
101. Medical Department III-Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany
102. Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG), Helmholtz Zentrum München, University of Leipzig and University Hospital Leipzig, Leipzig, Germany
103. Deutsches Zentrum für Diabetesforschung, Neuherberg, Germany
104. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
105. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
106. Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig Medical Center, Leipzig, Germany
107. Department of Neurology, UMC Utrecht Brain Center Rudolf Magnus, Utrecht, The Netherlands
108. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands
109. Biosciences Institute, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne, UK
110. German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, D-17475 Greifswald, Germany
111. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Felix-Hausdorff-Strasse 8, D-17475 Greifswald, Germany
112. Interfaculty Institute of Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany

113. German Center for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, 17475, Germany.
114. CARIM, Maastricht University
115. Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands
116. Department of Orthopaedics and Sportsmedicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands.

Supplementary Note

MYBL2 regulates the expression of many cell cycle genes

At the *MYBL2* locus, the lead variant chr20_43721344_C_T was associated with the expression of 151 target genes at FDR 5% (Figure S8). The target genes were strongly enriched for the Gene Ontology mitotic cell cycle term (GO:0000278, $p=2.607\times10^{-51}$) and the Reactome mitotic cell cycle pathway (R-HSA-69278, $p=1.216\times10^{-35}$). Interestingly, 125/151 genes (82%) had lower expression in carriers of the alternative T allele (Figure S8). The T allele of chr20_43721344_C_T was also strongly associated ($p =8.83x10^{-218}$) with the decreased expression of the *MYBL2* transcription factor gene in *cis (*Figure S8). Since both the MYBL2 transcription factor located in *cis* and majority of the *trans*-genes had lower expression in the carriers of the T allele, we hypothesised that MYBL2 might directly regulate these target genes. To test this, we download ChIP-seq data for the MYBL2 transcription factor in the human K562 myelogenous leukemia cell line from the ENCODE project (ENCSR162IEM). We then asked how many of the up- and downregulated genes had a MYBL2 ChIP-seq peak within +/- 2kb from the annotated promoter of the gene. We found that 99/125 (78.4%) downregulated genes had a MYBL2 peak in their promoter region (Figure S9). In contrast, only 1/26 upregulated genes had a MYBL2 peak in their promoter region. As a negative control, we looked at the 404 5% FDR target genes of the *SP140* locus (Table S2) and found that only 23/404 (5.6%) of the target genes had a MYBL2 peak in their promoter region (Figure S9).

To further understand which cell cycle stage these *MYBL2* target genes might be involved in, we obtained the list of genes specific to G2M and S phases of the cell cycle from Tirosh et al. 2015[102] using Seurat R package[103]. We found that 33/125 genes downregulated by the *trans*-eQTL variant were markers of the G2M phase which was significantly more than expected ($p = 1.23x10^{-53}$). In contrast, only 1/125 downregulated genes overlapped with markers of the S phase ($p = 0.36$). Of note, 2/26 upregulated genes overlapped S-phase markers ($p = 0.004$) and none of the upregulated genes overlapped G2M-phase markers.

Altogether, this evidence strongly suggests that MYBL2 directly regulates the expression of G2M genes in *trans* by binding to their promoter sequences and is directly involved in the regulation of the expression of these target genes.

Table S1. **Overview of the LCL eQTL discovery cohorts.** The cohorts included in the analysis used a mixture of RNA-seq and microarray technologies and three cohorts (TwinsUK, MRCE and MRCA) contained related samples.

Cohort	Sample size	Expression data	Genotype data	Relatedness
ALSPAC[31,73,74]	877	microarray	imputed (1000G 30x on GRCh38)	unrelated
TwinsUK[75]	735	RNA-seq	imputed (1000G 30x on GRCh38)	twins
CoLaus[76,77]	553	RNA-seq	imputed (TOPMed)	unrelated
GEUVADIS[78]	358	RNA-seq	WGS (1000G 30x on GRCh38)	unrelated
Liang_2013 (MRCE)[79]	484	microarray	imputed (1000G 30x on GRCh38)	siblings
Liang_2013 (MRCA)[79]	327	microarray	imputed (1000G 30x on GRCh38)	siblings
GENCORD[80]	187	RNA-seq	imputed (1000G 30x on GRCh38)	unrelated
GTEx[17]	113	RNA-seq	WGS (GRCh38)	unrelated
CAP[81]	100	RNA-seq	imputed (1000G 30x on GRCh38)	unrelated
MAGE[30] (replication)	682	RNA-seq	WGS (GRCh38)	unrelated, diverse ancestries

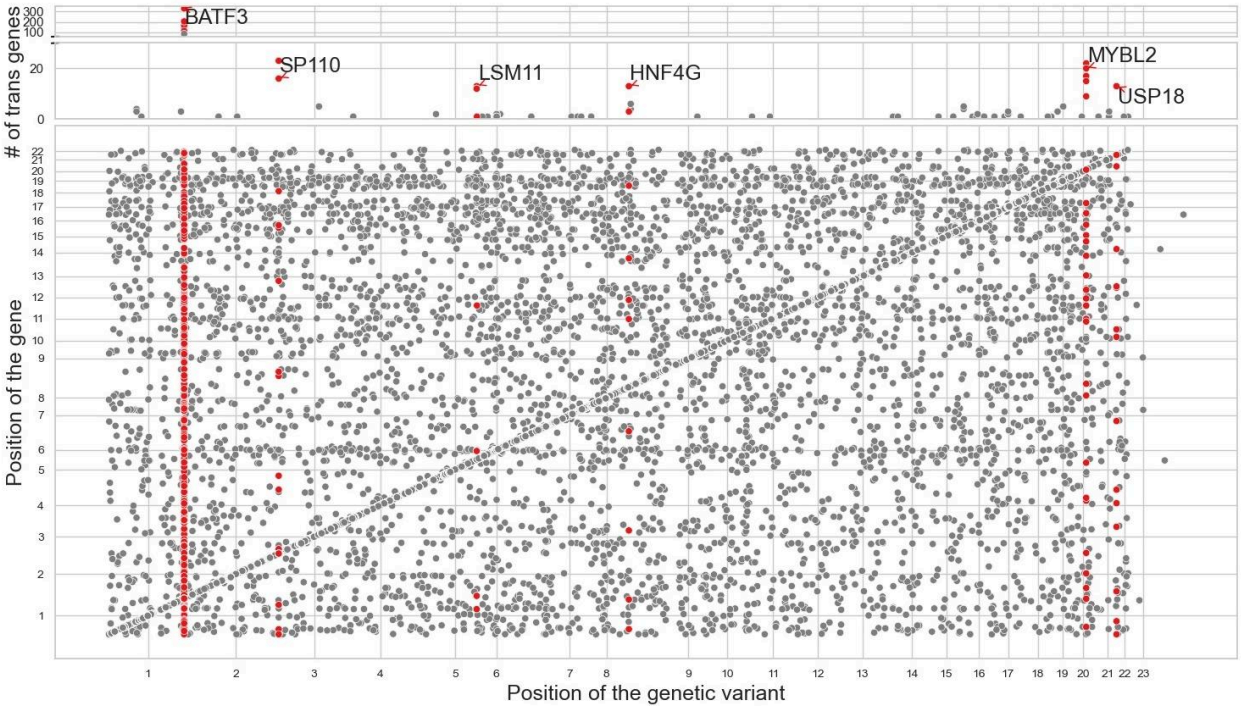

Figure S1. Overview of *trans*-eQTL analysis at the relaxed $p < 5x10^{-8}$ threshold. The upper scatter plot shows the number of *trans* associations detected at each *trans*-eQTL locus with p-values $< 5x10^{-8}$. Six largest *trans*-eQTL loci have been labelled with the name of the closest *cis* gene. The lower scatter plot shows all significant loci for each tested gene at the $p < 5x10^{-8}$ threshold. *Cis* associations are located on the diagonal while putative *trans* associations are located off diagonal.

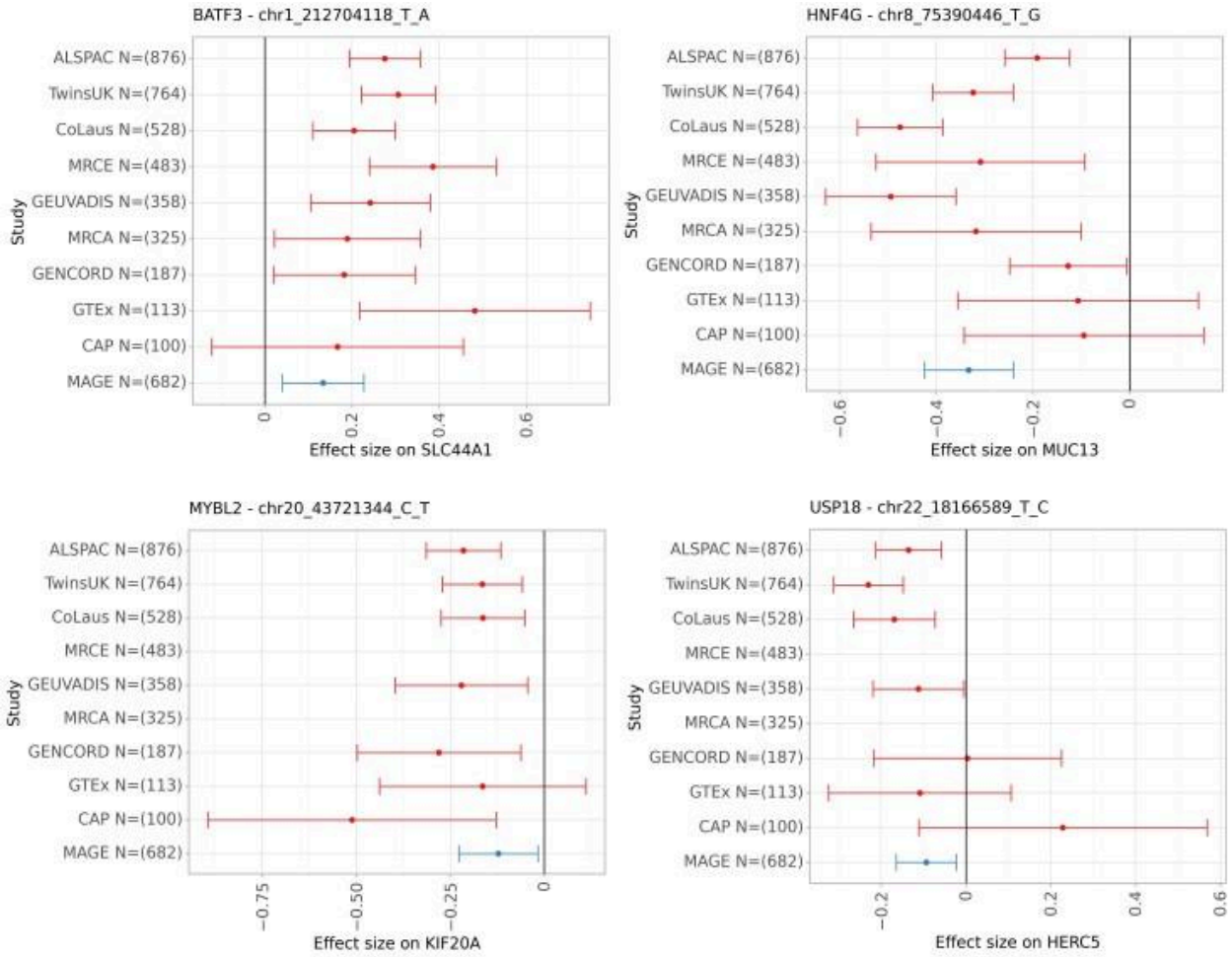

Figure S2. **Forest plots showing cohort-specific effect size for the four *trans*-eQTL loci that replicated in the MAGE cohort.** The points represent the *trans*-eQTL effect size estimates from regenie and the error bars represent 95% confidence intervals.

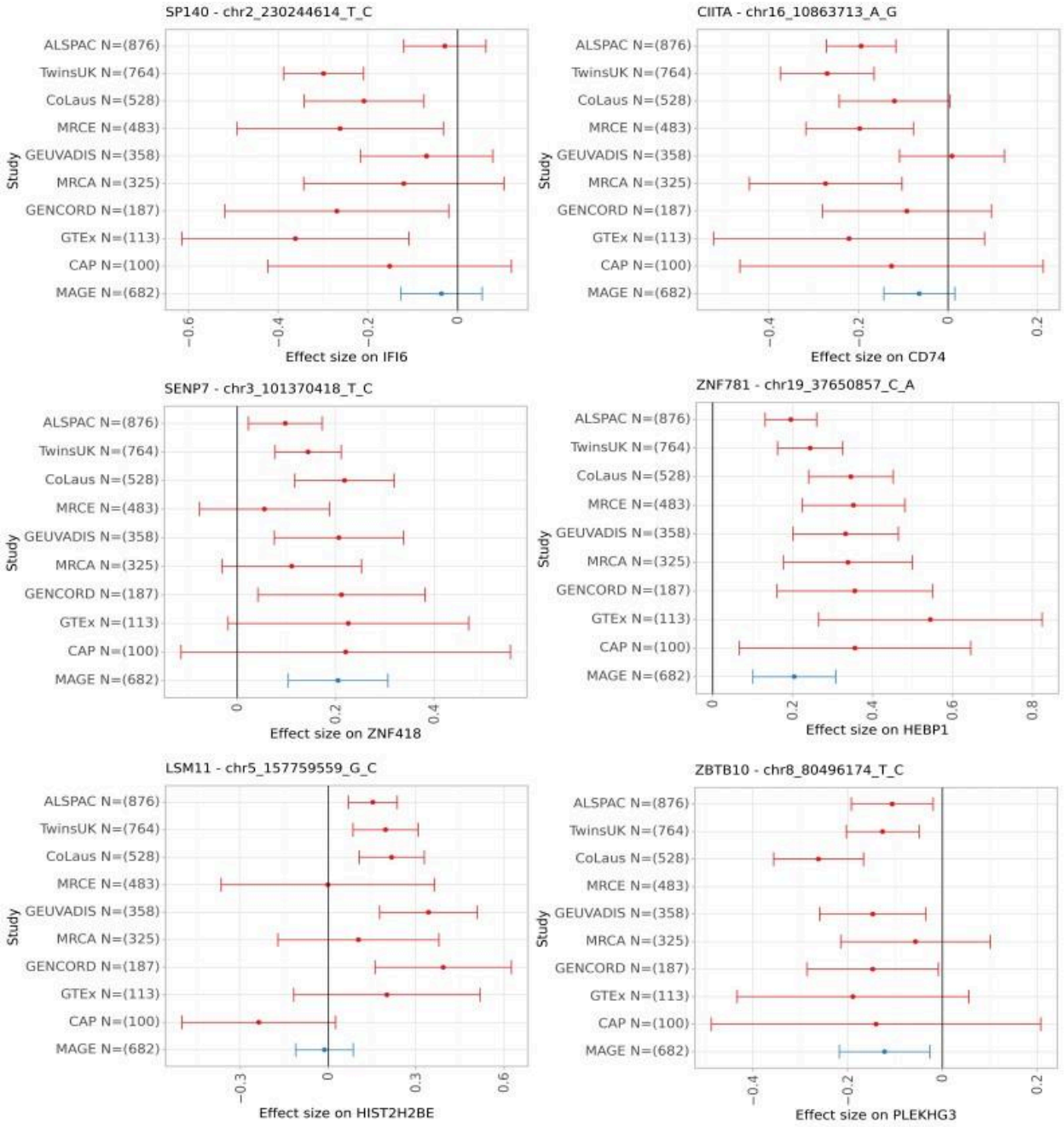

Figure S3. **Forest plots showing cohort-specific effect size for the remaining *trans*-eQTL loci that either did not replicate in the MAGE cohort or corresponded to likely cross-mappability artefacts (*SENP7*, *ZNF781* and *ZBTB10* loci).** The points represent the *trans*-eQTL effect size estimates from regenie and the error bars represent 95% confidence intervals.

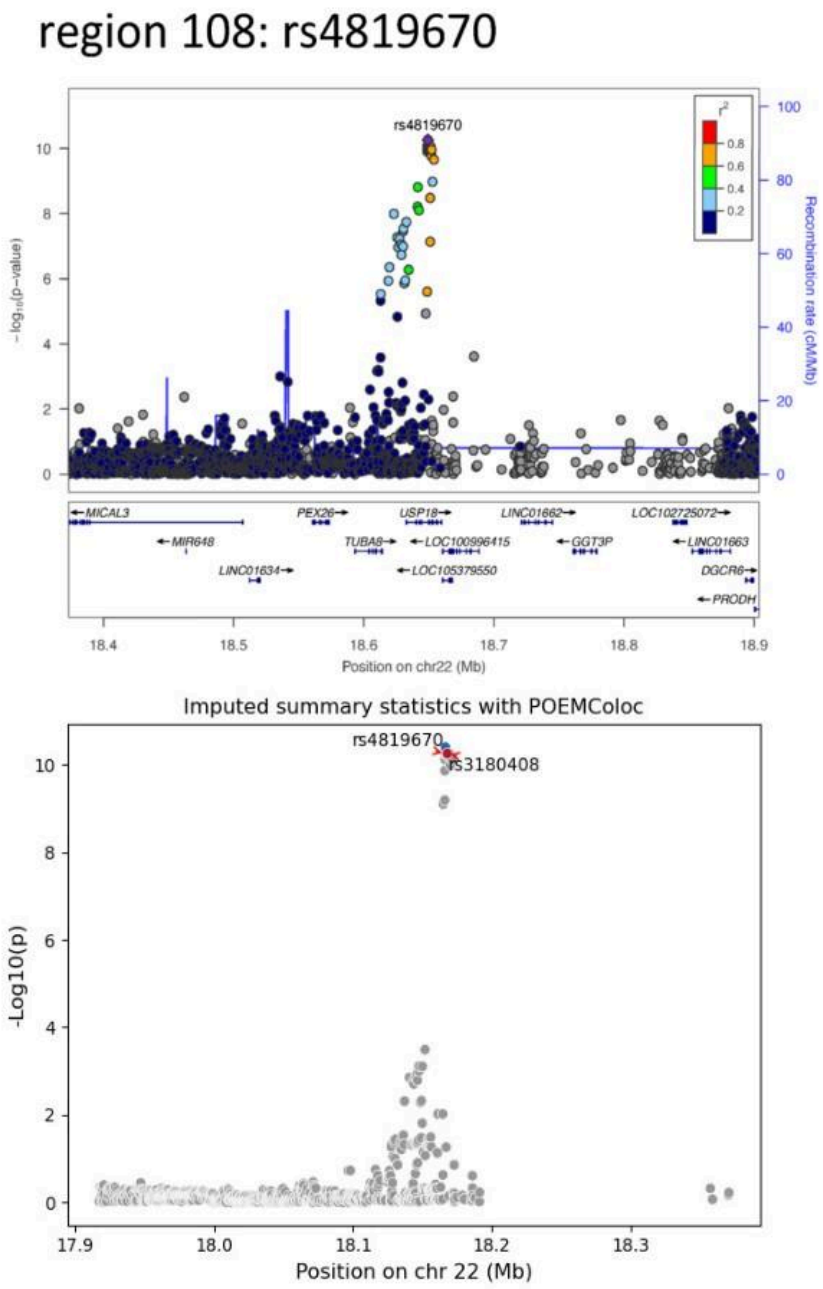

Figure S4. Original regional association plot for the *USP18* SLE GWAS locus from Yin *et al.* 2020 study and the summary statistics imputed with POEMColoc for the same locus.

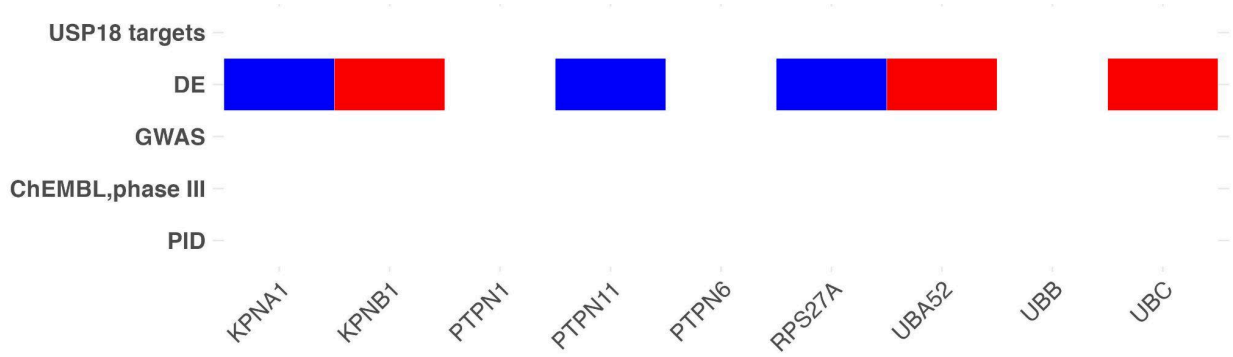

Figure S5. Category III: other interferon alpha/beta signalling pathway genes that do not belong to categories I or II (shown in Figure 3). The increased gene expression is marked in red, while reduced gene expression is marked in blue. The visualisation illustrates the effect on USP18 targets in relation to the risk allele. DE - differential gene expression in SLE cases *versus* controls[42]; GWAS - GWAS hits for SLE[33], ChEMBL, phase III - SLE phase III clinical trials from ChEMBL[50], PID - genes causing primary immunodeficiency from Genomics England.

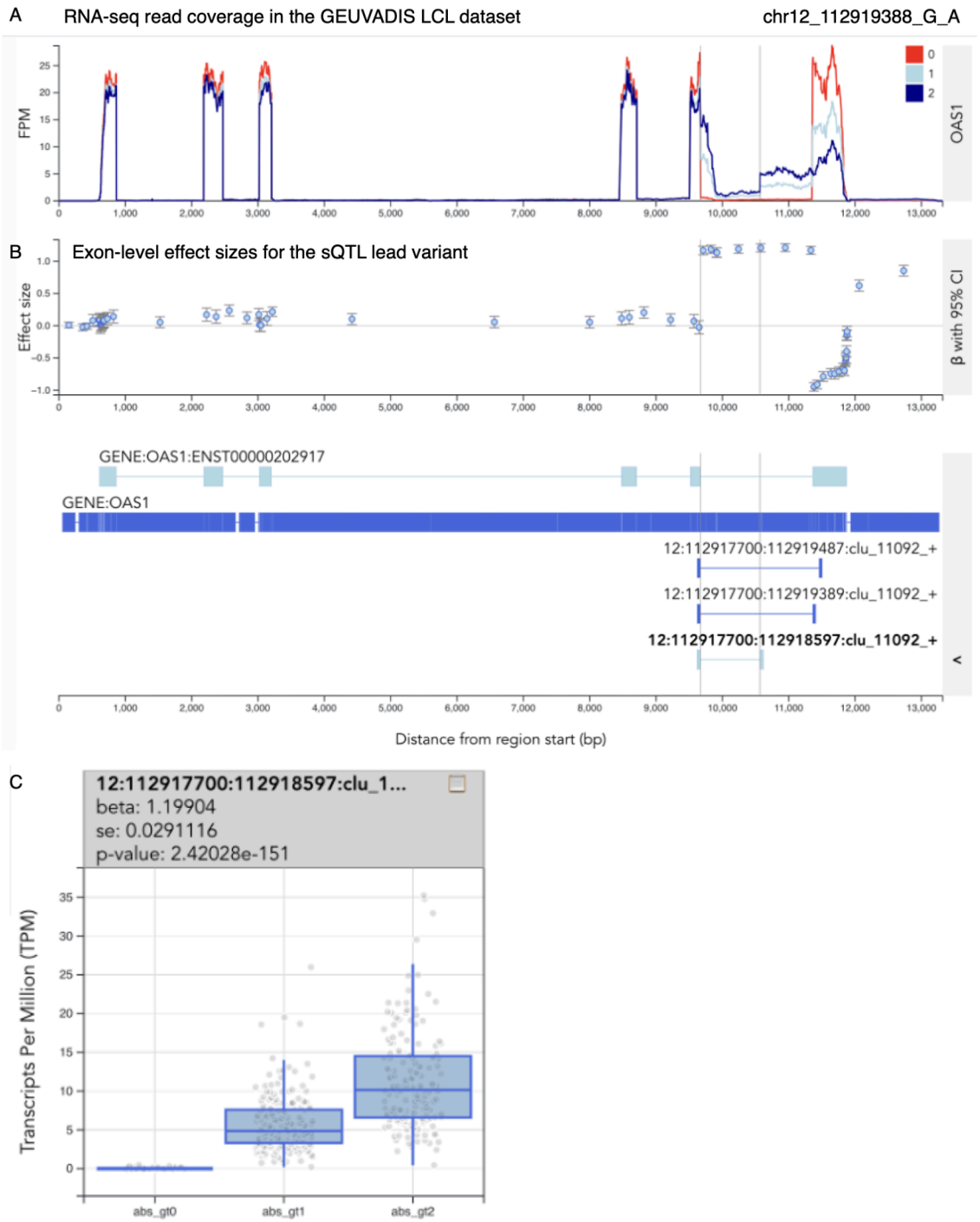

Figure S6. Fine-mapped splicing QTL (sQTL) in the *OAS1* gene. (**A**) RNA-seq read coverage of the *OAS1* gene in the GEUVADIS LCL dataset, stratified by the genotype of the fine-mapped sQTL variant chr12_112919388_G_A (posterior inclusion probability = 1). (**B**) Exon-level effect sizes for the sQTL lead variant. (**C**) Boxplot of the absolute expression of the short last intron of the *OAS1* gene (highlighted on panel A) stratified by the genotype of the lead sQTL variant. Interactive version of the plot can be viewed here.

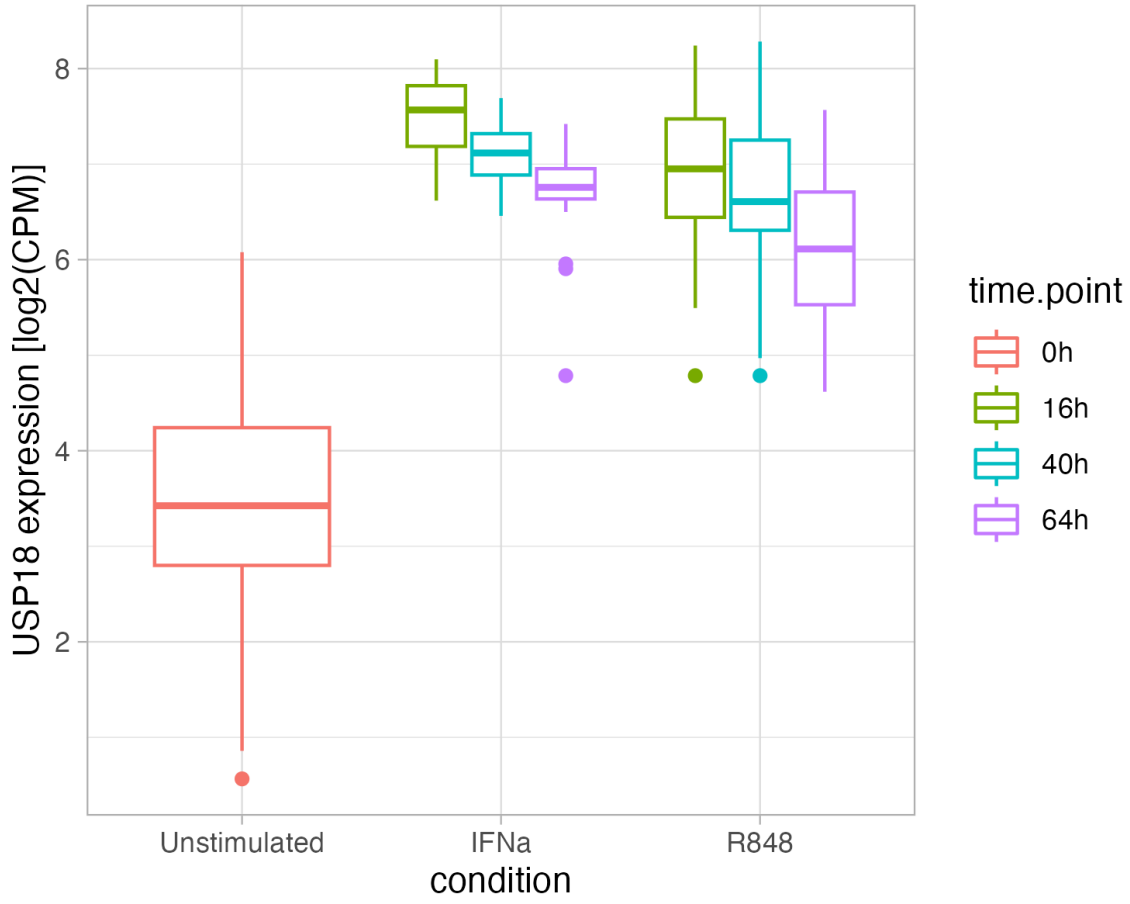

Figure S7. **Expression level of *USP18* in resting and stimulated B-cell subset of peripheral blood mononuclear cells (PBMCs)**. PBMCs were isolated from healthy donors and stimulated with interferon-alpha (IFNa) or R848 for 16, 40 and 64 hours.

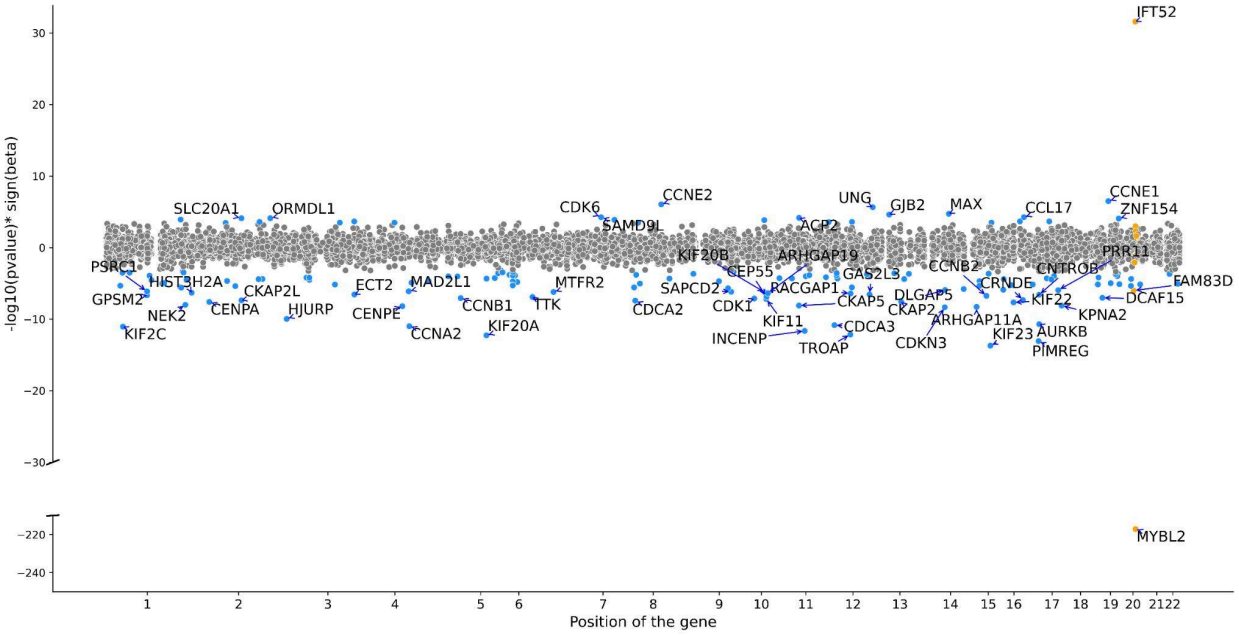

Figure S8. **MYBL2 regulates the expression of many cell cycle genes.** The scatter plot shows all genes associated with the *MYBL2 trans*-eQTL lead variant (chr20_43721344_C_T). Light blue points show significantly associated genes (variant-level Benjamini-Hochberg FDR 5%)

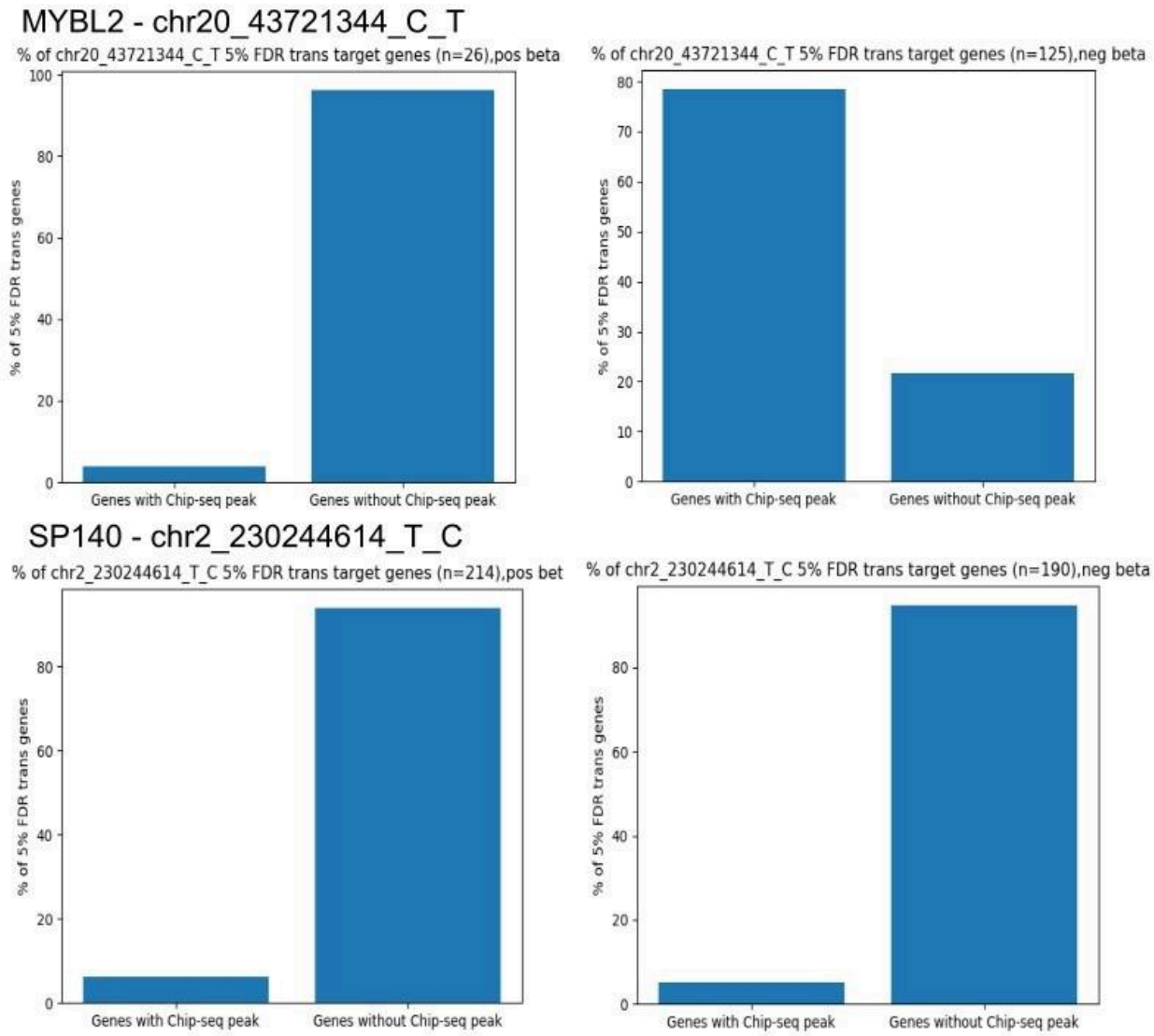

Figure S9. Overlap between *trans*-eQTL target genes and MYBL2 ChIP-seq peaks. The top panel shows the proportion of the *MYBL2 trans*-eQTL target genes upregulated (left) or downregulated (right) by the effect allele that contain a MYBL2 ChIP-seq peak within +/- 2kb from the annotated promoter. The bottom panel shows the proportion of the *SP140* trans-eQTL target genes upregulated (left) or downregulated (right) by the effect allele that contain a MYBL2 ChIP-seq peak within +/- 2kb from the annotated promoter. Only genes downregulated by the MYBL2 effect allele show a sizable overlap with MYBL2 ChIP-seq peaks.

References

1. Aguet, F. *et al.* Molecular quantitative trait loci. *Nature Reviews Methods Primers* **3**, 1–22 (2023).
2. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342**, 253–257 (2013).
3. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
4. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
5. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 1–10 (2023).
6. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
7. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* (2013) doi:10.1038/ng.2756.
8. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
9. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
10. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* eabj1541 (2021).
11. Bonder, M. J. *et al.* Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat. Genet.* (2021) doi:10.1038/s41588-021-00800-7.
12. Kolberg, L., Kerimov, N., Peterson, H. & Alasoo, K. Co-expression analysis reveals

interpretable gene modules controlled by trans-acting genetic variants. *Elife* **9**, e58705 (2020).

13. Kasela, S. *et al.* Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet.* **13**, e1006643 (2017).
14. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
15. Brandt, M. *et al.* An autoimmune disease risk variant: A trans master regulatory effect mediated by IRF1 under immune stimulation? *PLoS Genet.* **17**, e1009684 (2021).
16. Small, K. S. *et al.* Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
17. Consortium, T. G. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
18. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
19. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* 1 (2018).
20. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
21. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375 (2014).
22. Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer* **4**, 757–768 (2004).
23. Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**, 296–301 (2022).
24. Bar-Or, A. *et al.* Epstein-Barr Virus in Multiple Sclerosis: Theory and Emerging

Immunotherapies. *Trends Mol. Med.* **26**, 296–310 (2020).

25. James, J. A. *et al.* An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. *J. Clin. Invest.* **100**, 3019–3026 (1997).
26. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
27. Lanz, T. V. *et al.* Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* **603**, 321–327 (2022).
28. Thomas, O. G. *et al.* Heightened Epstein-Barr virus immunity and potential cross-reactivities in multiple sclerosis. *PLoS Pathog.* **20**, e1012177 (2024).
29. Saha, A. & Battle, A. False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. *F1000Res.* **7**, 1860 (2018).
30. Taylor, D. J. *et al.* Sources of gene expression variation in a globally diverse cohort. *bioRxiv* 2023.11.04.565639 (2023) doi:10.1101/2023.11.04.565639.
31. Bryois, J. *et al.* Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* **10**, e1004461 (2014).
32. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* 1–7 (2021).
33. Yin, X. *et al.* Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Ann. Rheum. Dis.* **80**, 632–640 (2021).
34. King, E. A., Dunbar, F., Davis, J. W. & Degner, J. F. Estimating colocalization probability from limited summary statistics. *BMC Bioinformatics* **22**, 254 (2021).
35. Meuwissen, M. E. C. *et al.* Human USP18 deficiency underlies type 1 interferonopathy leading to severe pseudo-TORCH syndrome. *J. Exp. Med.* **213**, 1163–1174 (2016).
36. Alsohime, F. *et al.* JAK Inhibitor Therapy in a Child with Inherited USP18 Deficiency. *N. Engl. J. Med.* **382**, 256–265 (2020).

37. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
38. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
39. Schnitzler, G. R. *et al.* Convergence of coronary artery disease genes onto endothelial cell programs. *Nature* 1–9 (2024).
40. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
41. Mostafavi, S. *et al.* Parsing the Interferon Transcriptional Network and Its Disease Associations. *Cell* **164**, 564–578 (2016).
42. Banchereau, R. *et al.* Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* **165**, 551–565 (2016).
43. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2610–2615 (2003).
44. Jefferies, C. A. Regulating IRFs in IFN Driven Disease. *Front. Immunol.* **10**, 325 (2019).
45. Honda, K., Takaoka, A. & Taniguchi, T. Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. *Immunity* **25**, 349–360 (2006).
46. Reshef, Y. A. *et al.* Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. *Nat. Genet.* **50**, 1483–1493 (2018).
47. Kerimov, N. *et al.* eQTL Catalogue 2023: New datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genet.* **19**, e1010932 (2023).
48. Li, H. *et al.* Identification of a Sjögren's syndrome susceptibility locus at OAS1 that influences isoform switching, protein expression, and responsiveness to type I interferons. *PLoS Genet.* **13**, e1006820 (2017).
49. Banday, A. R. *et al.* Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19 hospitalization in patients of European and African ancestries.

Nat. Genet. **54**, 1103–1116 (2022).

50. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024).
51. Morand, E. F. *et al.* Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N. Engl. J. Med.* **382**, 211–221 (2020).
52. Amaya-Uribe, L., Rojas, M., Azizi, G., Anaya, J.-M. & Gershwin, M. E. Primary immunodeficiency and autoimmunity: A comprehensive review. *J. Autoimmun.* **99**, 52–72 (2019).
53. Schmidt, R. E., Grimbacher, B. & Witte, T. Autoimmunity and primary immunodeficiency: two sides of the same coin? *Nat. Rev. Rheumatol.* **14**, 7–18 (2017).
54. Primary immunodeficiency or monogenic inflammatory bowel disease (Version 4.191). https://panelapp.genomicsengland.co.uk/panels/398/.
55. Mackensen, A. *et al.* Anti-CD19 CAR T cell therapy for refractory systemic lupus erythematosus. *Nat. Med.* **28**, 2124–2132 (2022).
56. Müller Fabian *et al.* CD19 CAR T-Cell Therapy in Autoimmune Disease — A Case Series with Follow-up. *N. Engl. J. Med.* **390**, 687–700 (2024).
57. Akinbiyi, T., McPeek, M. S. & Abney, M. ADELLE: A global testing method for Trans-eQTL mapping. *bioRxiv* 2024.02.24.581871 (2024) doi:10.1101/2024.02.24.581871.
58. Dutta, D. *et al.* Aggregative trans-eQTL analysis detects trait-specific target gene sets in whole blood. *Nat. Commun.* **13**, 1–14 (2022).
59. Wang, L., Babushkin, N., Liu, Z. & Liu, X. Trans-eQTL mapping in gene sets identifies network effects of genetic variants. *bioRxiv* 2022.11.11.516189 (2022) doi:10.1101/2022.11.11.516189.
60. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat. Rev. Immunol.* **15**, 87–103 (2015).

61. Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium *et al.* The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* **379**, 1214–1224 (2012).
62. C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC) *et al.* Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* **342**, d548 (2011).
63. Burgess, S. *et al.* Using genetic association data to guide drug discovery and development: Review of methods and applications. *Am. J. Hum. Genet.* **110**, 195–214 (2023).
64. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
65. Karim, M. A. *et al.* Systematic disease-agnostic identification of therapeutically actionable targets using the genetics of human plasma proteins. *medRxiv* (2023) doi:10.1101/2023.06.01.23290252.
66. Wong, D. *et al.* Genomic mapping of the MHC transactivator CIITA using an integrated ChIP-seq and genetical genomics approach. *Genome Biol.* **15**, 494 (2014).
67. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).
68. Hore, V. *et al.* Tensor decomposition for multiple-tissue gene expression experiments. *Nat. Genet.* **48**, 1094–1100 (2016).
69. Ntranos, V., Kamath, G. M., Zhang, J. M., Pachter, L. & Tse, D. N. Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. *Genome Biol.* **17**, 112 (2016).
70. EQTLGen consortium. https://www.eqtlgen.org/.
71. Morris, J. A. *et al.* Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science* **380**, eadh7699 (2023).
72. Weinstock, J. S. *et al.* Gene regulatory network inference from CRISPR perturbations in

primary CD4+ T cells elucidates the genomic basis of immune disease. *bioRxiv* 2023.09.17.557749 (2023) doi:10.1101/2023.09.17.557749.

73. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
74. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
75. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).
76. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
77. Sönmez Flitman, R. *et al.* Untargeted Metabolome- and Transcriptome-Wide Association Study Suggests Causal Genes Modulating Metabolite Concentrations in Urine. *J. Proteome Res.* **20**, 5103–5114 (2021).
78. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
79. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* **23**, 716–726 (2013).
80. Gutierrez-Arcelus, M. *et al.* Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* **2**, e00523 (2013).
81. Theusch, E., Chen, Y.-D. I., Rotter, J. I., Krauss, R. M. & Medina, M. W. Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics* **21**, 555 (2020).
82. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
83. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000

Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).

84. Deelen, P. *et al.* Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* **7**, 901 (2014).
85. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
86. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
87. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
88. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
89. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
90. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
91. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
92. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
93. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
94. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
95. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
96. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression

analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

97. Roshchupkin, G. V. *et al.* HASE: Framework for efficient high-dimensional association analyses. *Sci. Rep.* **6**, 36076 (2016).
98. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. in *Proceedings of the 9th Python in Science Conference* (SciPy, 2010). doi:10.25080/majora-92bf1922-011.
99. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
100. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
101. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
102. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
103. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2023).