

# 1 *USP18* modulates lupus risk via negative 2 regulation of interferon response

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

- 9  
10 1. Institute of Computer Science, University of Tartu, Tartu, Estonia  
11 2. Open Targets, South Building, Wellcome Genome Campus, Hinxton, Cambridge, UK  
12 3. Department of Computational Biology, University of Lausanne, Switzerland  
13 4. Swiss Institute of Bioinformatics, Lausanne, Switzerland  
14 5. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the  
15 Netherlands  
16 6. OncoCode Institute, Amsterdam, the Netherlands  
17 7. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK  
18 8. Laboratorio Internacional de Investigación Sobre el Genoma Humano, Universidad Nacional Autónoma  
19 de México, Santiago de Querétaro, Mexico  
20 9. Internal Medicine Research Unit, Research and Development, Pfizer, Cambridge, MA, USA  
21 10. Bristol Myers Squibb, Cambridge, Massachusetts, USA  
22 11. Research Technologies, GSK, Stevenage, UK  
23 12. Research Technologies, GSK, Heidelberg, Germany  
24 13. Institute of Genomics, University of Tartu, Tartu, Estonia  
25 14. Institute of Biomedicine and Translational Medicine, Faculty of Medicine, University of Tartu, Estonia  
26 15. Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa  
27 <sup>†</sup>A list of authors and affiliations for the eQTLGen Consortium is listed in the Supplementary  
28 Materials.  
29 <sup>\*</sup>Corresponding author. Email: [kaur.alasoo@ut.ee](mailto:kaur.alasoo@ut.ee)

## 30 Abstract

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

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

## 43 Introduction

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

54  
55 *Trans*-QTLs are especially promising, because 60%-90% of gene and protein expression  
56 heritability is located in *trans*<sup>4</sup>, most associations detected in large-scale pQTL studies are  
57 located in *trans*<sup>5</sup>, and *cis*-QTL discovery is starting to saturate after 10,000 samples<sup>5</sup>.  
58 Furthermore, most complex trait heritability has been proposed to be mediated by *trans*-QTL  
59 effects<sup>4</sup>. However, current large-scale *trans*-eQTL and *trans*-pQTL studies have been limited to  
60 easily accessible bulk tissues such as whole blood<sup>6,7</sup> or plasma<sup>5,8-10</sup>. Bulk tissue studies are  
61 subject to cell type composition effects which can be difficult to distinguish from true intracellular  
62 *trans*-QTLs<sup>1,6</sup>. The whole blood and plasma studies are also likely to miss cell type and context  
63 specific regulatory effects. In contrast, *trans*-eQTL studies in other tissues and purified cell types  
64 have had limited statistical power due to small sample sizes (typically less than one thousand  
65 samples), enabling the discovery of only very large effects and potentially underestimating  
66 pleiotropic effects on multiple target genes<sup>11-18</sup>.

67  
68 A key limitation in our understanding of how *trans*-eQTLs contribute to complex traits and how  
69 they interact with *cis*-eQTL is the lack of well-characterised disease-associated *trans*-eQTL  
70 signals<sup>4</sup>. Two most prominent examples include the adipose-specific *KLF14* locus associated  
71 with type 2 diabetes<sup>16,19</sup> and the *IRX3/5* locus associated with obesity<sup>20,21</sup>. At the *KLF14* locus,  
72 the lead variant (rs4731702) is a *cis*-eQTL for the *KLF14* transcription factor and was  
73 associated with the expression of 385 target genes in *trans*, 18 of which also had independent  
74 *cis* associations for other metabolic traits<sup>19</sup>. The simultaneous regulation of multiple target genes  
75 in *trans*-eQTL regulatory networks seems to be a general property of many known *trans*-eQTL  
76 signals<sup>6,12,14</sup>. However, what proportion of *trans*-eQTL target genes directly mediate the disease  
77 or trait associations as opposed to being independent 'bystanders' with minimal direct causal  
78 effect has remained unclear.

79  
80 We performed the largest *trans*-eQTL meta-analysis in a single cell type, comprising 3,734  
81 lymphoblastoid cell line (LCL) samples across nine cohorts (MetaLCL). LCLs are obtained by  
82 transforming primary B-cells with Epstein-Barr virus<sup>22</sup>. LCLs have been widely used as a

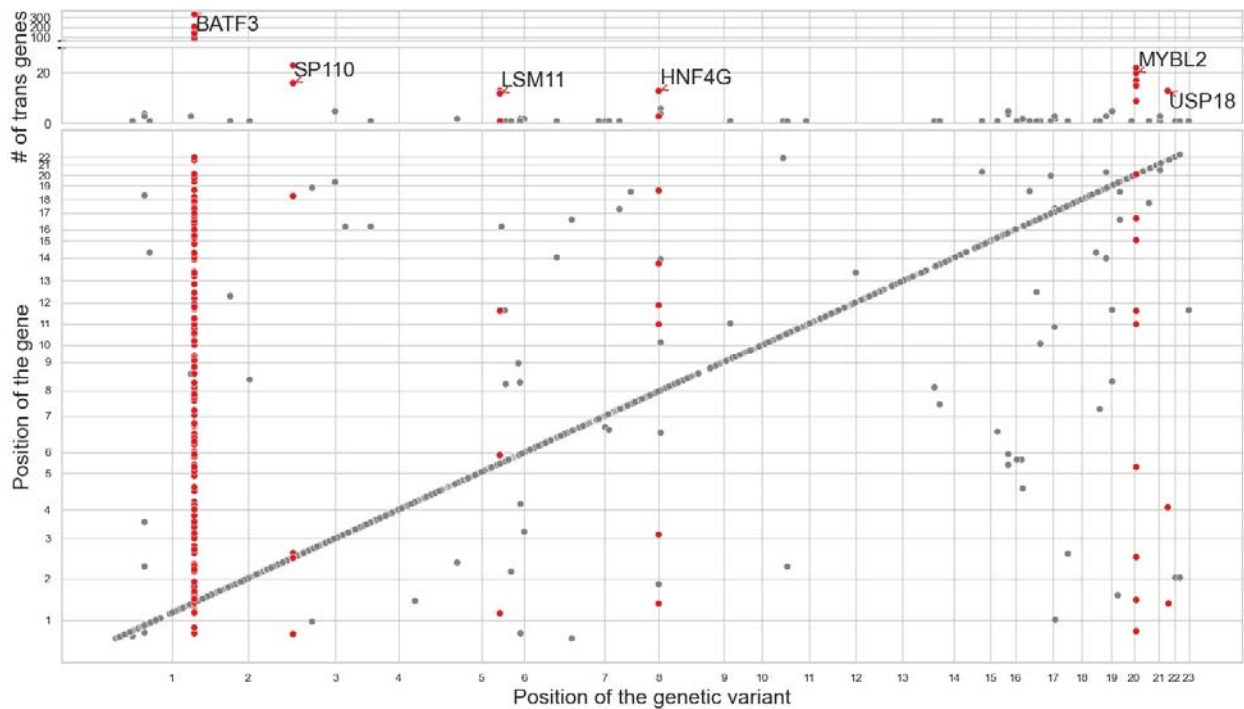
83 resource for human genetics, from banking cells from rare genetic disorders, through control  
84 material in laboratories to prevent repetitive blood sampling, to the study of tumorigenesis,  
85 mechanisms of viral latency and immune evasion<sup>22</sup>. Furthermore, Epstein-Barr virus has been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

91  
92 After stringent quality control, we identify four highly robust *trans*-eQTL associations that  
93 replicate in an independent cohort (n=682) and are associated with multiple target genes. One  
94 of these signals corresponds to a missense variant in the *USP18* gene and is also associated  
95 with increased risk of SLE. The SLE risk allele is associated with increased activity of the type I  
96 interferon signalling pathway and increased expression of several classical interferon response  
97 genes. While there is robust evidence for the potential causal role of increased interferon  
98 signalling in SLE pathogenesis, we find that the expression of most individual interferon  
99 response genes is unlikely to have a direct causal effect on SLE. Our results caution against  
100 blindly using *trans*-QTL associations for target gene prioritisation without clear understanding of  
101 the *trans*-QTL mechanism and robust genetic evidence from *cis*-acting variants implicating the  
102 same gene. To support secondary use of our data, we have made the complete MetaLCL *trans*-  
103 eQTL summary statistics for 18,792 genes publicly available via the eQTL Catalogue FTP  
104 server.

## 105 Results

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

107 We performed a large-scale *trans*-eQTL meta-analysis, utilising data from LCLs collected from  
108 3,734 donors across nine cohorts of European ancestries (Table S1). After excluding *cis*  
109 associations located within 5 Mb of the target gene, we identified 79 suggestive independent  
110 *trans*-eQTL loci at p-value <  $1 \times 10^{-11}$  threshold (Figure 1). To identify robust signals associated  
111 with multiple target genes and reduce the risk of false positives caused by cross-mappability<sup>29</sup>,  
112 we further required each locus to be associated with at least five independent target genes ( $p <$   
113  $5 \times 10^{-8}$ ) with low cross-mappability scores (see Methods). This filtering reduced the number of  
114 candidate loci to six (Figure 1), four of which replicated in an independent multi-ethnic cohort of  
115 682 individuals<sup>30</sup>. These four replicating *trans*-eQTL loci were located near the *BATF3*, *MYBL2*,  
116 *USP18*, *HNF4G* genes (Table S2, Figure S2). While the strong *trans*-eQTL signal near the  
117 *BATF3* transcription factor (2294 targets at FDR 5%) has been previously reported<sup>31</sup>, the other  
118 three seem to be novel. Remarkably, the *trans*-eQTL targets at the *MYBL2* locus were  
119 consistent with direct activation by the MYBL2 transcription factor (Supplementary Note),  
120 indicating that our analysis is identifying biologically interpretable signals.

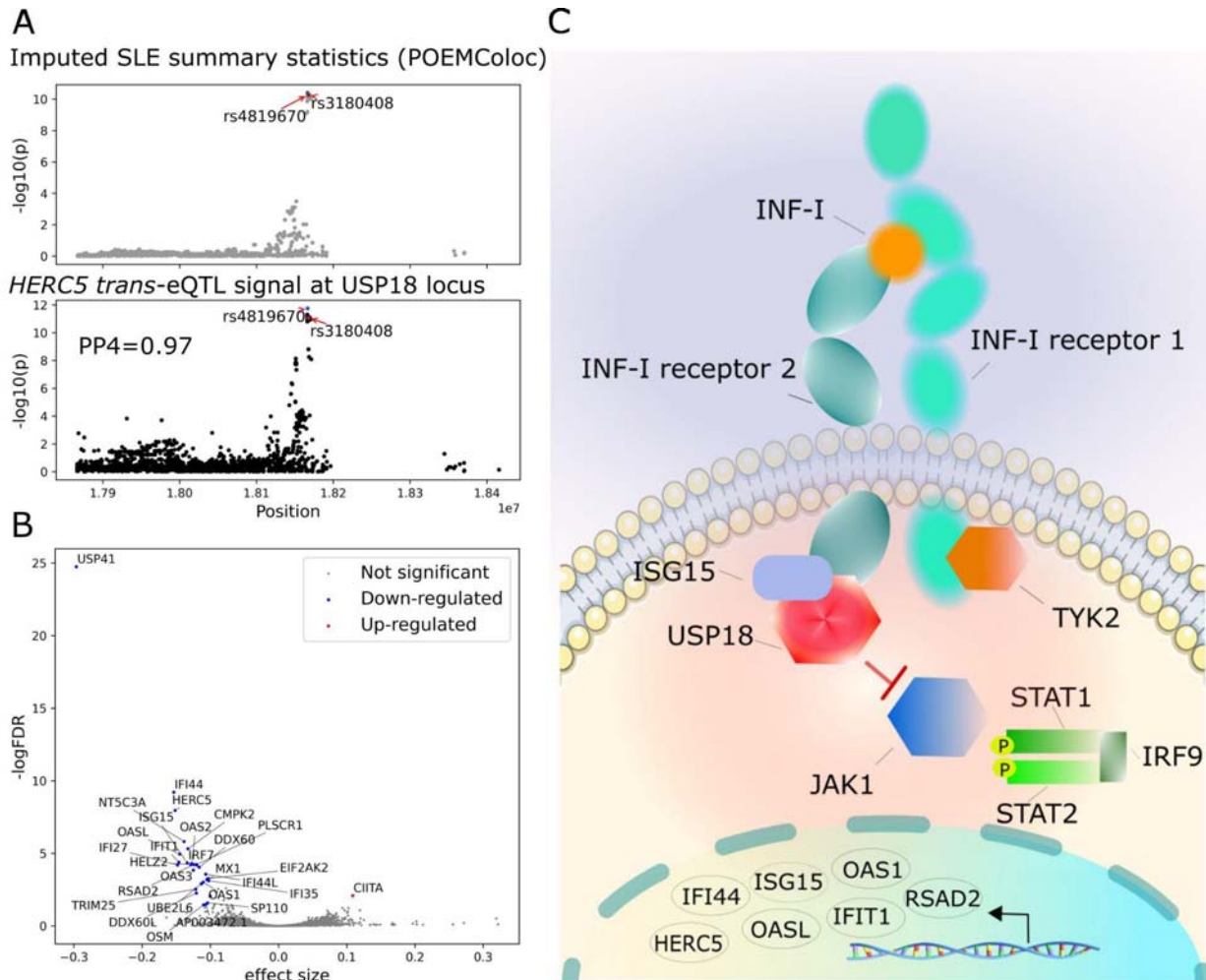


121  
122 **Figure 1. Overview of MetaLCL *trans*-eQTL results.** The upper scatter plot shows the number  
123 of *trans* associations detected at each *trans*-eQTL locus with  $p$ -values  $< 5 \times 10^{-8}$ . Six largest  
124 *trans*-eQTL loci have been labelled with the name of the closest *cis* gene. The lower scatter plot  
125 shows all significant loci for each tested gene at the more stringent  $p < 1 \times 10^{-11}$  threshold. *Cis*  
126 associations are located on the diagonal while putative *trans* associations are located off  
127 diagonal.

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

130 For the four high-confidence loci, we performed GWAS lookup using the Open Targets Genetics  
131 Portal<sup>32</sup>. We found that our *trans*-eQTL lead variant in the *USP18* locus (chr22\_18166589\_T\_C,  
132 rs4819670) was shared with a GWAS lead variant identified for SLE in East Asians<sup>33</sup>. Using the  
133 point estimation of colocalisation (POEMColoc) method, we also confirmed that the two signals  
134 colocalised (PP4 = 0.97) (Figure 2A)<sup>34</sup>. At this locus, we identified 40 *trans* target genes at FDR  
135 5% that were all strongly enriched for Reactome interferon signalling (R-HSA-913531,  $p =$   
136  $1.1 \times 10^{-26}$ ) and interferon alpha/beta signalling (R-HSA-909733,  $p = 1.7 \times 10^{-21}$ ) pathways. The  
137 rs4819670-C allele was associated with decreased expression of multiple canonical type I  
138 interferon response genes (e.g. *ISG15*, *IFI44*, *OAS1-3*) (Figure 2B). Reassuringly, we observed  
139 consistent effect sizes across nine sub-cohorts in our meta-analysis ( $I^2$  heterogeneity statistic =  
140 0.46, Figure S2). *USP18* is a known negative regulator of interferon signalling and a rare loss-  
141 of-function mutation in *USP18* causes severe type I interferonopathy (Figure 2C)<sup>35,36</sup>. The  
142 rs4819670-C was also associated with decreased risk of systemic lupus erythematosus (SLE) in  
143 East Asians<sup>33</sup>. Furthermore, the rs4819670 lead variant is in perfect linkage disequilibrium (LD)  
144 ( $r^2 = 1$ , 1000 Genomes EAS superpopulation) with a *USP18* missense variant rs3180408

145 (chr22\_18167915\_C\_T, ENSP00000215794.7:p.Thr169Met). While the GWAS association was  
 146 previously known, it remained uncertain which allele of the missense variant rs3180408 was  
 147 more likely to decrease USP18 protein function, especially because the variant was predicted to  
 148 be benign by all tested variant effect prediction tools available from Ensembl VEP<sup>37</sup>. Our results  
 149 suggest that the rs3180408-T SLE risk allele decreases USP18 protein function as USP18 is a  
 150 negative regulator of type I interferon response genes.  
 151



152  
 153 **Figure 2. SLE GWAS association at the *USP18* locus is a *trans*-eQTL for interferon**  
 154 **response genes. (A)** Regional association plot for the SLE GWAS with POEMColoc imputed  
 155 summary statistics and regional association plot for the lead *trans*-eQTL gene (*HERC5*) at the  
 156 *USP18* locus. The *trans*-eQTL lead and GWAS lead variants (rs4819670, shown in blue) are  
 157 identical and in perfect LD with a missense variant (rs3180408, shown in red) in the *USP18*  
 158 gene. The original regional association plot for the SLE GWAS is shown on Figure S4. **(B)**  
 159 Volcano plot of the *trans*-eQTL target genes. **(C)** Schematic illustration of the role of USP18 in  
 160 the regulation of interferon response genes, adapted from Alshime *et al*<sup>36</sup>.

## 161 Role of aberrant interferon signalling in lupus pathogenesis

162 Several studies have suggested that causal GWAS genes are enriched in shared pathways or  
163 biological processes<sup>38-40</sup>. To further characterise the potential role of USP18 target genes in  
164 lupus, we performed additional *trans*-eQTL meta-analysis across the nine discovery cohorts and  
165 one replication cohort (total n = 4,416). This increased the number of significant USP18 target  
166 genes to 50 (FDR < 5%). Notably, 18/50 target genes overlapped the Reactome interferon  
167 alpha/beta signalling (R-HSA-909733) pathway (hypergeometric test,  $p = 4.14 \times 10^{-24}$ ) and 26/50  
168 genes overlapped a consensus set of interferon response genes (n = 124) identified by  
169 Mostafavi *et al*<sup>41</sup> ( $p = 1.44 \times 10^{-39}$ , Table S4). Reassuringly, 39/50 genes were also more highly  
170 expressed in peripheral blood mononuclear cells from SLE cases compared to controls<sup>42</sup> (Table  
171 S4), consistent with the established role of increased interferon signalling in SLE<sup>43</sup>.

172  
173 To better understand the role of the USP18 target genes in the interferon alpha/beta signalling  
174 pathway, we focussed on the 60 genes belonging to the Reactome R-HSA-909733 interferon  
175 alpha/beta signalling pathway and divided them into three categories - category I: proteins  
176 involved in signal transduction via IFNAR1/2 receptor (n = 13 genes, including the multi-gene  
177 interferon-alpha gene cluster, Figure 3A); category II: downstream transcriptional targets of the  
178 interferon signalling (38 genes from the Reactome R-HSA-1015702 sub-pathway, Figure 3B)  
179 and 3) and category III: other pathway genes (n = 9) not belonging to the first two categories  
180 (Figure S5). We found that 16/50 USP18 targets overlapped with the 38 category II genes  
181 (transcriptional targets of interferon response) ( $p = 1.18 \times 10^{-29}$ ). In contrast, only 2/50 USP18  
182 *trans*-eQTL target genes (*STAT1* and *ISG15*) overlapped the 13 category I genes (IFNAR1/2  
183 receptor signal transduction proteins) and none overlapped the 9 category III genes. This  
184 suggests that the USP18 *trans*-eQTLs are primarily capturing the transcriptional targets of  
185 interferon response (category II), consistent with the established role of USP18 in regulating  
186 these genes (Figure 2C)<sup>36</sup>.

187  
188 Next, we assessed if there were additional lupus GWAS signals overlapping the three  
189 categories of interferon response genes defined above. We first used the Open Targets  
190 Genetics portal to extract the prioritised target genes for 108 lupus GWAS loci from Yin *et al*<sup>33</sup>.  
191 This revealed that three prioritised genes (*USP18*, *STAT1*, *IFNA1-17*) overlapped with the 13  
192 category I genes (IFNAR1/2 signal transduction proteins, Figure 3A) and four prioritised genes  
193 (*IRF1/5/8* and *OAS1*) overlapped with the 38 category II genes (transcriptional targets of  
194 interferon response, Figure 3B). Out of these, IRF1/5/8 are themselves transcription factors  
195 involved in the regulation of interferon production<sup>44,45</sup>, and stronger IRF1 binding across many  
196 GWAS loci has been associated with higher Crohn's disease risk<sup>46</sup>. Only *OAS1* represents a  
197 classical antiviral gene and here the GWAS lead variant is in perfect LD with a fine-mapped  
198 splice QTL for *OAS1* in the eQTL Catalogue (Figure S6)<sup>47</sup>. Interestingly, while the USP18 *trans*-  
199 eQTL risk allele increased *OAS1* expression (Figure 3B), the *cis* splice QTL risk allele  
200 (rs10774671-A) increased the expression of a transcript with an alternative 3' end that was  
201 associated with lower *OAS1* protein abundance (Figure S6)<sup>48,49</sup>, suggesting that *cis* and *trans*  
202 effects on the *OAS1* gene have opposite direction of effect on lupus risk.

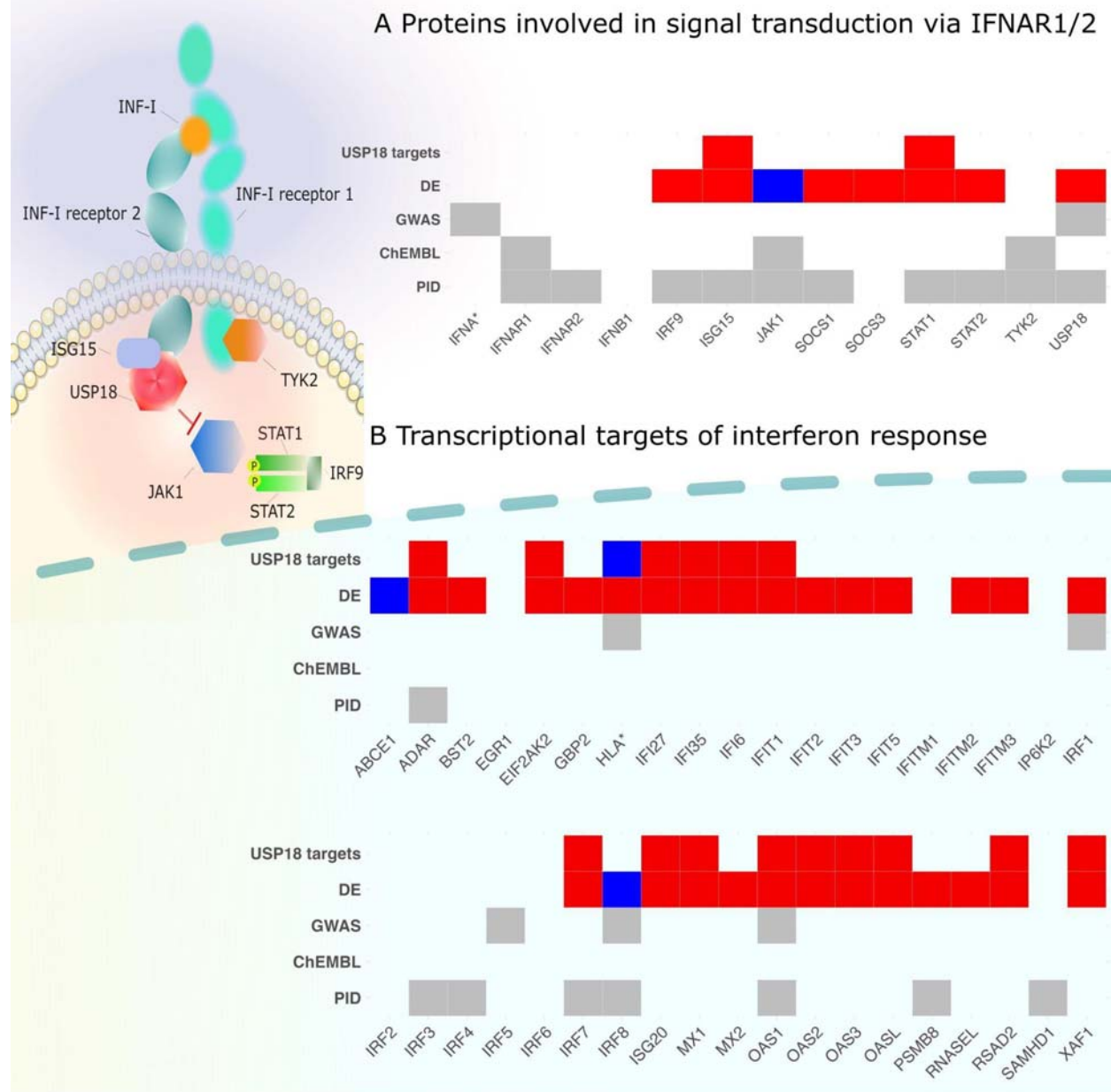
203

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

212  
213 There is an emerging consensus that rare mutations in genes prioritised for autoimmune  
214 diseases from GWAS studies can often also cause primary immunodeficiencies (PIDs)<sup>52,53</sup>. For  
215 example, loss-of-function mutations in *USP18* cause rare type I interferonopathy<sup>35,36</sup>. At the  
216 same time, GWAS studies for SLE and other autoimmune diseases are still only powered to  
217 detect variants with large effects. Thus, knowing if a gene causes PID might be a useful (if  
218 noisy) indicator that the same gene might be discovered in a future larger autoimmune GWAS  
219 study. Thus, we obtained the list of genes causing either PID or monogenic inflammatory bowel  
220 disease from Genomics England<sup>54</sup> and overlapped those with the three categories of interferon  
221 response genes defined above. We found that 10/13 category I genes (interferon signal  
222 transduction) have previously been implicated in causing PID, including *USP18* and all three  
223 phase III drug candidates for SLE (Figure 3A). In contrast, only 8 of the 38 category II genes  
224 (transcriptional targets of interferon response) have been implicated in PIDs (Figure 3B),  
225 including *OAS1* and *IRF8* also detected by SLE GWAS. Finally, none of the category III genes  
226 have been implicated in PIDs (Figure S5).

227  
228 Triangulation of evidence from prioritised lupus GWAS target genes, phase III clinical trial  
229 information and overlap with primary immunodeficiency genes highlights modulation of aberrant  
230 interferon alpha/beta signalling in B-cells as an emerging therapeutic opportunity for SLE  
231 (category I, Figure 3A). This is further supported by recent studies demonstrating that depleting  
232 autoreactive B-cells via anti-CD19 CAR T cell therapy is an effective therapy for SLE and other  
233 autoimmune diseases<sup>55,56</sup>. In contrast, most *trans*-eQTL targets of *USP18* overlap  
234 transcriptional targets of interferon response (category II, Figure 3B) and it is far less clear what  
235 is the potential causal roles of these genes in SLE pathogenesis.

236  
237  
238  
239  
240



241  
 242 **Figure 3. Role of interferon signalling in SLE pathogenesis.** (A) Upstream regulators of  
 243 interferon response genes (IFNA\* contains multi-gene interferon-alpha gene cluster). (B)  
 244 Downstream transcriptional targets of the interferon signalling (HLA\* marks the HLA region).  
 245 The increased gene expression is marked in red, while reduced gene expression is marked in  
 246 blue. The visualisation illustrates the effect on USP18 target genes in relation to the SLE risk  
 247 allele. DE - differential gene expression in SLE cases *versus* controls<sup>42</sup>; GWAS - GWAS hits for  
 248 SLE<sup>33</sup>, ChEMBL, phase III - SLE phase III clinical trials from ChEMBL<sup>50</sup>, PID - genes causing  
 249 primary immunodeficiency from Genomics England.



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

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

260  
261 To understand the potential reasons for the attenuated effect in whole blood, we compared the  
262 expression level of the *USP18* gene across 49 GTEx tissues. We found that *USP18* had the  
263 highest expression in LCLs (median transcripts per million (TPM) = 45.3) and one of the lowest  
264 in whole blood (median TPM = 0.46). Since *USP18* is itself an interferon response gene and  
265 LCLs are characterised by a strong interferon signature driven by active infection with the  
266 Epstein-Barr virus, we characterised the expression of *USP18* in naive B-cells as well as B-cells  
267 stimulated with interferon-alpha and TLR7/8 agonist R848 for 16, 40 and 64 hours. We found  
268 that the expression level of *USP18* in B-cells was upregulated by ~3.5-fold after 16 hours of  
269 stimulation and stayed elevated for at least 64 hours (Figure S7). This suggests that the very  
270 strong active interferon signalling and associated upregulation of *USP18* transcription in LCLs is  
271 required for the *trans*-eQTL signal to be detected.

## 272 Discussion

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

283  
284 Despite the strong evidence for the critical role of type I interferon response in SLE  
285 pathogenesis<sup>42,43</sup> and three active clinical trials, we were surprised to see that of the 50 *USP18*  
286 target genes, only *OAS1* had an independent *cis*-association with SLE. Expanding the analysis  
287 to interferon response genes from Reactome further implicated IRF1/5/8 genes and the HLA  
288 region, but most interferon response genes were not detected in the SLE GWAS. One potential  
289 explanation for this could be the limited statistical power of the SLE GWAS that profiled 13,377  
290 cases and 194,993 controls, identifying a total of 113 loci<sup>33</sup>. Furthermore, Liu *et al* demonstrated

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

295  
296 However, interferon response involves rapid upregulation of a broad transcriptional regulatory  
297 network of genes with diverse biological functions, only a subset of which might have a direct  
298 causal effect on SLE. This is supported by the fact that among the 38 interferon response genes  
299 (category II), only *OAS1*, *ADAR*, *PSMB8*, *SAMHD1* and the IRF transcription factors have been  
300 implicated in causing primary immunodeficiencies (Figure 3B). The remaining interferon  
301 response genes might thus be better thought of as biomarkers of the complex effect of  
302 interferon signalling on multiple parts of the immune system<sup>43,60</sup>. This could also help to explain  
303 the apparent directionally discordant *cis* and *trans* effects for the *OAS1* gene, raising an  
304 intriguing possibility that to reduce lupus risk it might be important to have high baseline levels  
305 of *OAS1* (to possibly aid with viral clearance<sup>49</sup>) rather than increase its expression long term  
306 after activation of interferon signalling. Similarly, it has been previously shown that variants in  
307 the *IL6R* region that are associated with circulating C-reactive protein (CRP) concentrations, are  
308 also associated with coronary artery disease (CAD) risk<sup>61</sup>, but variants in the *CRP* region are  
309 not<sup>62</sup>. Thus, plasma levels of CRP do not seem to have a direct causal effect on CAD risk, but  
310 can still act as a molecular readout (biomarker) of the *IL6R*-mediated inflammatory response  
311 that does seem to have a causal effect<sup>63</sup>. These observations suggest that widespread  
312 horizontal pleiotropy in gene regulatory networks could be a general property of *trans*-QTLs and  
313 could help explain why using *trans*-pQTL signals in Mendelian randomisation analysis has had  
314 low specificity for identifying known drug targets<sup>64,65</sup>. Instead, we propose that target genes  
315 identified from large-scale *trans*-QTL studies could be better thought of as drug response  
316 biomarkers for drugs targeting the *cis* gene responsible for the *trans* association<sup>8</sup>.

317  
318 A limitation of our *trans*-eQTL analysis is its susceptibility to cross-mappability artefacts (Table  
319 S3). While heuristic approaches have been developed to filter such artefacts *post hoc*, these  
320 approaches are not guaranteed to remove all cross-mappability effects and might be too  
321 conservative at other loci<sup>29</sup>. Cross-mappability artefacts also tend to replicate well in  
322 independent cohorts<sup>29</sup>. Furthermore, as the sample size of *trans*-eQTL studies increases, the  
323 power to detect subtle cross-mappability effects as putative *trans*-eQTLs also increases. To  
324 avoid these false positives, we used a very conservative strategy of requiring each *trans*-eQTL  
325 locus to have at least five independent target genes that all pass the cross-mappability filter. As  
326 a result, we likely missed many true *trans*-eQTLs regulating single or few target genes (e.g.  
327 *trans*-eQTL effect near the *CIITA* transcription factor on multiple HLA genes that has been  
328 replicated in several independent studies<sup>6,31,66-68</sup>, Table S2). Future large-scale *trans*-eQTL  
329 studies will likely require the development of novel methods to properly adjust for cross-  
330 mappability, such as explicit modelling of transcript compatibility read counts between *cis* and  
331 *trans* target genes<sup>69</sup>.

332  
333 While large-scale *trans*-eQTL studies using both bulk and single-cell measurements are likely to  
334 continue for easily accessible tissues such as whole blood (e.g. eQTLGen Phase 2<sup>70</sup>), it seems

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

## 339 Methods

### 340 Datasets, samples and ethics

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

### 365 Genotype data quality control and imputation

366 **Pre-imputation quality control.** Genotype imputation was performed as described  
367 previously<sup>47</sup>. Briefly, we lifted coordinates of the genotyped variants to the GRCh38 build with  
368 CrossMap v0.4.1<sup>82</sup>. We aligned the strands of the genotyped variants to the 1000 Genomes 30x  
369 on GRCh38 reference panel<sup>83</sup> using Genotype Harmonizer<sup>84</sup>. We excluded genetic variants with  
370 Hardy-Weinberg p-value < 10<sup>-6</sup>, missingness > 0.05 and minor allele frequency < 0.01 from  
371 further analysis. We also excluded samples with more than 5% of their genotypes missing.  
372

373 **Genotype imputation and quality control.** Most of the datasets were imputed using the 1000  
374 Genomes reference panel based on the GRCh38 genome version. CoLaus dataset was  
375 imputed using the TOPMed Imputation Server<sup>85-87</sup>, while still aligning with the same reference  
376 genome version. Additionally, GEUVADIS, GTEx and MAGE cohorts utilised whole genome  
377 sequencing data aligned to the GRCh38 reference genome.

378  
379 We pre-phased and imputed the microarray genotypes to the 1000 Genomes 30x on GRCh38  
380 reference panel<sup>83</sup> using Eagle v2.4.1<sup>88</sup> and Minimac4<sup>86</sup>. We used bcftools v1.9.0 to exclude  
381 variants with minor allele frequency (MAF) < 0.01 and imputation quality score  $R^2 < 0.4$  from  
382 downstream analysis. The genotype imputation and quality control steps are implemented in  
383 [eQTL-Catalogue/genimpute](#) (v22.01.1) workflow available from GitHub. Subsequently, we used  
384 QCTOOL v2.2.0 to convert imputed genotypes from VCF format to bgen format for *trans*-eQTL  
385 analysis with regenie.

## 386 Gene expression data

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

390  
391 **RNA-seq quantification and normalisation.** RNA-seq data were pre-processed as described  
392 previously<sup>89</sup>. Briefly, quantification of the RNA-seq data was performed using the [eQTL-](#)  
393 [Catalogue/rnaseq](#) workflow (v22.05.1) implemented in Nextflow. Before quantification, we used  
394 Trim Galore v0.5.0 to remove sequencing adapters from the fastq files. For gene expression  
395 quantification, we used HISAT2<sup>90</sup> v2.2.1 to align reads to the GRCh38 reference genome  
396 (*Homo\_sapiens.GRCh38.dna.primary\_assembly.fa* file downloaded from Ensembl). We counted  
397 the number of reads overlapping the genes in the GENCODE V30 reference transcriptome  
398 annotations with featureCounts v1.6.4.

399  
400 We excluded all samples that failed the quality control steps as described previously<sup>89</sup>. We  
401 normalised the gene counts using the conditional quantile normalisation (cqn) R package  
402 v1.30.0 with gene GC nucleotide content as a covariate. We downloaded the gene GC content  
403 estimates from Ensembl biomaRt and calculated the exon-level GC content using bedtools  
404 v2.19.0<sup>91</sup>. We also excluded lowly expressed genes, where 95 per cent of the samples within a  
405 dataset had transcripts per million (TPM)-normalised expression less than 1. Subsequently, we  
406 used the inverse normal transformation to standardise quantification estimates. Normalisation  
407 scripts together with containerised software are publicly available at [https://github.com/eQTL-](https://github.com/eQTL-Catalogue/qcnorm)  
408 [Catalogue/qcnorm](#).

409  
410 **Microarray data processing.** Gene expression from 877 individuals in the ALSPAC cohort was  
411 profiled using Illumina Human HT-12 V3 BeadChips microarray. We used the normalised gene  
412 expression matrix from the original publication<sup>31</sup>. In the MRCA cohort, gene expression from 327  
413 individuals was profiled using the Human Genome U133 Plus 2.0 microarray. We downloaded  
414 the raw CEL files from ArrayExpress (E-MTAB-1425) and normalised the data using the Robust  
415 Multi-Array Average (RMA) method from the affy Bioconductor package<sup>92</sup>. In the MRCE cohort,

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

## 421 *Trans*-eQTL mapping and meta-analysis

422 We performed independent quality control and normalisation on all datasets and only included  
423 18,792 protein coding genes in the analysis. *Trans*-eQTL analysis was conducted separately on  
424 each dataset with *regenie*<sup>93</sup>. For studies containing related samples (TwinsUK, MRCA and  
425 MRCE) and ALSPAC, both step 1 and step 2 commands were employed, while for other  
426 datasets with a smaller number of unrelated samples (Table S1), *regenie* was run in the linear  
427 regression mode (step 2 only). We used sex and six principal components of the normalised  
428 gene expression matrix and six principal components of genotype data as covariates in the  
429 *trans*-eQTL analysis. All scripts used to run *trans*-eQTL are publicly available at  
430 [https://github.com/freimannk/regenie\\_analysis](https://github.com/freimannk/regenie_analysis). Subsequently, we performed an inverse-  
431 variance weighted meta-analysis across studies. Meta-analysis workflow is available at  
432 [https://github.com/freimannk/regenie\\_metaanalyse](https://github.com/freimannk/regenie_metaanalyse).

433  
434 We used a *cis* window of  $\pm 5$ Mb to assign identified eQTLs into *cis* and *trans* eQTLs. To  
435 determine significant loci, we excluded variants proximal ( $\pm 1.5$ Mb) to the most highly associated  
436 variant per gene. This approach allowed us to identify distinct and robust signals while  
437 mitigating potential confounding effects from nearby variants. By applying these filters, we found  
438 79 *trans*-eQTLs loci at a suggestive p-value threshold of  $1 \times 10^{-11}$ .

## 439 Accounting for cross-mappability

440 A major source of false positives in *trans*-eQTL analysis is cross-mappability, whereby RNA-seq  
441 reads from gene A erroneously align to gene B, leading to very strong apparent *trans*-eQTL  
442 signals<sup>11,29</sup>. To exclude potential cross-mappability artefacts, we excluded all *trans*-eQTLs  
443 where there was high cross-mappability (cross-mappability score from Saha *et al*<sup>29</sup>  $> 1$ ) between  
444 the *trans*-eQTL target gene and at least one protein coding gene in the *cis* region ( $\pm 1.5$ Mb) of  
445 the *trans* eQTL lead variant. Since some of the strongest cross-mappability artefacts affected  
446 one or few target genes (Table S3), we further restricted our analysis to *trans*-eQTL loci that  
447 had five or more target genes with  $p < 5 \times 10^{-8}$  and cross-mappability score  $< 1$ .

## 448 Replication of *trans*-eQTL associations

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

455 dataset with concordant direction of effect. Based on this criterion, 7/10 loci replicated (Table  
456 S2). Secondly, since all of our loci had multiple target genes, we used the pi1 statistic to  
457 estimate the proportion of FDR < 5% target gene at each locus that had a non-null p-value in  
458 the replication dataset<sup>94</sup>. We used the qvalue R package<sup>95</sup> to calculate  $\pi_1 = 1 - \text{qvalue}(5\% \text{ FDR}$   
459  $\text{trans gene p-values}) \times \pi_0$ . For 3/10 loci, the proportion of non-null p-values was > 0.5 (Table  
460 S2). Note that replication in an independent cohort does not help to reduce false positives due  
461 to cross-mappability, as cross-mappability artefacts tend to be highly replicable<sup>29</sup>.

462  
463 **eQTLGen Consortium.** The eQTLGen Consortium is an initiative to investigate the genetic  
464 architecture of blood gene expression and to understand the genetic basis of complex traits. We  
465 used interim summary statistics from eQTLGen phase 2, wherein a genome-wide eQTL  
466 analysis has been performed in 52 cohorts, representing 43,301 individuals.

467  
468 All 52 cohorts performed cohort-specific analyses as outlined in the eQTLGen analysis  
469 cookbook (<https://eqtlgen.github.io/eqtlgen-web-site/eQTLGen-p2-cookbook.html>). Genotype  
470 quality control was performed according to standard bioinformatics practices and included  
471 quality metric-based variant and sample filtering, removing related samples, ethnic outliers and  
472 population outliers. Genotype data was converted to genome build hg38 if not done so already  
473 and the autosomes were imputed using the 1000 Genomes 30x on GRCh38 reference panel<sup>83</sup>  
474 (all ancestries) using the eQTLGen imputation pipeline ([eQTLGen/eQTLGenImpute](#)).

475  
476 Like the genotype data, gene expression data was processed using the eQTLGen data QC  
477 pipeline ([eQTLGen/DataQC](#)). For array-based datasets, we used the results from the empirical  
478 probe mapping approach from our previous study<sup>6</sup> to connect the most suitable probe to each  
479 gene which has previously been shown to show expression in the combined BIOS whole blood  
480 expression dataset. Raw expression data was further normalized in accordance with the  
481 expression platform used (quantile normalization for Illumina expression arrays and TMM<sup>96</sup> for  
482 RNA-seq) and inverse normal transformation was performed. Gene expression outlier samples  
483 were removed and gene summary information was collected for filtering at the central site.  
484 Samples for whom there were mismatches in genetically inferred sex, reported sex, or the  
485 expression of genes encoded from sex chromosomes were removed. Similarly, samples with  
486 unclear sex, based on genetics or gene expression were removed.

487  
488 An adaptation of the HASE framework<sup>97</sup> was used to perform genome-wide meta-analysis. For  
489 genome-wide eQTLs analysis, this limits the data transfer size while ensuring participant  
490 privacy. At each of the cohorts, the quality controlled and imputed data was processed and  
491 encoded so that the individual level data can no longer be extracted, but while still allowing  
492 effect sizes to be calculated for the linear relationship between variants and gene expression  
493 ([eQTLGen/ConvertVcf2Hdf5](#) and [eQTLGen/PerCohortDataPreparations](#)).

494 Centrally, the meta-analysis pipeline was run on the 52 cohorts. The pipeline which performs  
495 per cohort calculations of effect sizes and standard errors and the inverse variance meta-  
496 analysis is available at [eQTLGen/MetaAnalysis](#). We included 4 genetic principal components,  
497 20 gene expression principal components and other technical covariates (e.g. RNA integrity  
498 number) where available. Per every dataset, genes were included if the fraction of unique

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

### 503 Differential gene expression in SLE cases *versus* controls

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

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

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

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

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

525  
526 **Thawing and stimulation.** Cryopreserved PBMCs were thawed quickly and washed in 14mL of  
527 room temperature complete RPMI 1640 media (10% FBS, 1% Penicillin-Streptomycin, 1% L-  
528 Glutamine). Cells were incubated at  $37^{\circ}\text{C}$ , 5%  $\text{CO}_2$  for 2h. Cells were then stimulated with  
529 interferon alpha (IFN- $\alpha$ , Bio-technie) and R-848 (Resiquimod, Cambridge Bioscience) at a  
530 working concentration of 1000U/mL and 2 $\mu\text{g/ml}$ , respectively. Cells were incubated at  $37^{\circ}\text{C}$ ,  
531 5%  $\text{CO}_2$  and harvested after 16h, 40h and 64h of stimulation. Unstimulated cells were kept in  
532 culture without any stimuli for 16h (i.e., 0h of activation).

533  
534 **Multiplexing, CITE-seq staining & scRNA-seq.** Upon harvesting, cells were resuspended in a  
535 cell staining buffer (Biolegend) and cell hashing and genotype-based multiplexing was  
536 performed. Donors of the same stimulation condition were mixed at equal ratios (each pool  
537 corresponded to a mix of cells from four to five different individuals). These pools were stained

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

544  
545 Cells were next processed using the 10X Genomics Immune Profiling 5' high-throughput (HT)  
546 v2 kit, as specified by the manufacturer's instructions.  $1.15 \times 10^5$  cells were loaded into each inlet  
547 of a 10X Chromium X to create Gel Bead-in-emulsions (GEMs). Two 10X HT reactions were  
548 loaded per time point of sample processing (targeted recovery was 40,000 cells per 10X  
549 reaction). Reverse transcription was performed on the emulsion, after which cDNA and CITEseq  
550 supernatant were purified, amplified and used to construct RNA-sequencing and CSP  
551 sequencing libraries, respectively. These RNA and CSP libraries were sequenced at a 5:1 ratio,  
552 respectively, using the Illumina NovaSeq 6000 S4, with 100-bp paired-end reads and all 10X  
553 reactions were mixed at equal ratios and sequenced across two lanes.

554  
555 **Deconvolution of single cells by genotype.** Each 10X reaction comprised a mix of cells from  
556 unrelated individuals. Thus, natural genetic variation was used to assign cells to their respective  
557 individuals. First, a list of common exonic variants was compiled from the 1000 Genomes  
558 Project phase 3 exome-sequencing data (MAF > 0.05). Next, cellSNP (v1.2.1) was used to  
559 generate pileups at the genomic location of these variants. These pileups, in combination with  
560 the variants called from genotyping in each individual, were used as an input for Vireo<sup>99</sup> (v0.5.7).  
561 If any cell had less than 0.9 posterior probability of belonging to any individual or were of mixed  
562 genotypes they were labelled as 'unassigned' and 'doublets', respectively, and removed from  
563 downstream analysis.

564  
565 **Data processing and quality controls.** Raw scRNA-seq and CITE-seq data were processed  
566 using the Cell Ranger Multi pipeline (v7.0.0, 10x Genomics). In brief, RNA and CSP library  
567 reads were first assigned to cells. RNA reads were then aligned to the GRCh38 human  
568 reference genome and CSP antibody reads were matched to the provided list of known  
569 barcodes. Ensembl version 93 was used as a reference for gene annotation, and gene  
570 expression was quantified using reads assigned to cells and confidently mapped to the genome.  
571 Additionally, Cell Ranger multi was used to deconvolute samples based on HTOs. It uses an  
572 algorithm which employs a latent variable model over a state space composed of each HTO  
573 used in the experiment to assign each cell to a stimulation condition or as a doublet.

574  
575 Results from RNA and CSP quantification in Cell Ranger were imported into RStudio (v4.3.1)  
576 and analysed using Seurat (v5.0.1). Any cell identified as doublet or unassigned by Vireo and or  
577 antibody hashtag deconvolution method were excluded. 10X reactions were split by time point  
578 and stimuli condition. Cells with 1.5 - 2.5 median absolute deviations below the median of genes  
579 and counts detected were discarded. Additionally, cells with 3 - 4 median absolute deviation  
580 above the median for the percentage of mitochondrial reads detected were discarded. The



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

583

584 **Pseudobulk and normalisation.** Raw counts were pseudobulked by Azimuth annotated level 1  
585 cell types (CD4, CD8, B, Mono, DC, NK, Other and Other T) per donor, per time point and per  
586 stimulation condition, via edgeR<sup>101</sup> (v4.0.16). Pseudobulked raw counts were then counts per  
587 million (CPM) normalised and log<sub>2</sub> transformed with edgeR.

588

## 589 Data availability

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

## 615 URLs

616 MetaLCL website: <https://github.com/AlasooLab/MetaLCL>

617 MetaLCL *trans*-eQTL analysis workflow: [https://github.com/freimannk/regenie\\_analysis](https://github.com/freimannk/regenie_analysis)

618 MetaLCL meta-analysis workflow: [https://github.com/freimannk/regenie\\_metaanalyse](https://github.com/freimannk/regenie_metaanalyse)

619 eQTL Catalogue website: [https://www.ebi.ac.uk/eql/Data\\_access/](https://www.ebi.ac.uk/eql/Data_access/)

620 eQTL Catalogue genotype imputation workflow: <https://github.com/eQTL-Catalogue/genimpute>

621 eQTL Catalogue RNA-seq processing workflow: <https://github.com/eQTL-Catalogue/rnaseq>  
622 eQTL Catalogue data normalisation workflow: <https://github.com/eQTL-Catalogue/qcnorm>  
623 eQTLGen analysis cookbook: [https://eqtlgen.github.io/eqtlgen-web-site/eQTLGen-p2-  
624 cookbook.html](https://eqtlgen.github.io/eqtlgen-web-site/eQTLGen-p2-cookbook.html)  
625 eQTLGen data QC workflow: <https://github.com/eQTLGen/DataQC>  
626 eQTLGen genotype conversion workflow: <https://github.com/eQTLGen/ConvertVcf2Hdf5>  
627 eQTLGen per-cohort analysis workflow:  
628 <https://github.com/eQTLGen/PerCohortDataPreparations>  
629 eQTLGen meta-analysis workflow: <https://github.com/eQTLGen/MetaAnalysis>  
630 eQTLGen genotype imputation workflow: <https://github.com/eQTLGen/eQTLGenImpute>  
631 QCTOOL: [https://www.chg.ox.ac.uk/~gav/qctool\\_v2/](https://www.chg.ox.ac.uk/~gav/qctool_v2/)

## 632 Declaration of interests

633 The authors declare no competing interests.

## 634 Acknowledgements

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

## 647 Funding

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

659 fellowship 790972 from Consejo Nacional de Humanidades, Ciencias y Tecnologías  
660 CONAHACYT, México. The UK Medical Research Council and Wellcome (Grant ref:  
661 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. ALSPAC GWAS  
662 data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute  
663 and LabCorp (Laboratory Corporation of America) using support from 23andMe. This  
664 publication is the work of the authors and they will serve as guarantors for the contents of this  
665 paper.

## 666 Author contributions

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

## 675 Supplementary Information

676

677 1. eQTLGen Consortium - Author information

678 2. Supplementary Note

679 3. Table S1

680 4. Figure S1-S9

681

## 682 eQTLGen Consortium – Author information

683

684 Author list is ordered alphabetically.

685

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

709

710 1. Department of Public Health Sciences, University of Chicago, Illinois, USA

711 2. Pfizer–University of Granada–Junta de Andalucía Centre for Genomics and Oncological Research,  
712 Granada, Spain

713 3. Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

714 4. Ontario Institute for Cancer Research, University of Toronto

715 5. Department of Genetics, Faculty of Science, University of Granada, 18071 Granada, Spain

716 6. Bioinformatics Laboratory, Biotechnology Institute, Centro de Investigación Biomédica, PTS, Avda. del  
717 Conocimiento s/n, 18100 Granada, Spain

718 7. Department of Biomedical Engineering, Department of Computer Science, Department of Genetic  
719 Medicine, Johns Hopkins University, Baltimore, MD, USA

720 8. Department of Internal Medicine/Cardiology, Heart Center Leipzig at Leipzig University, Leipzig,  
721 Germany

722 9. Integrative Neurogenomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National  
723 Institutes of Health, Bethesda, MD, USA

724 10. Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National  
725 Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

726 11. Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud  
727 University Medical Center, Nijmegen, 6525 HP, the Netherlands

- 728 12. Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud  
729 University Medical Center, Nijmegen, 6525 HP, the Netherlands
- 730 13. Department of Human Genetics, David Geffen School of Medicine, University of California Los  
731 Angeles, Los Angeles, CA, USA
- 732 14. Amsterdam Reproduction & Development (AR&D) research institute, Amsterdam, the Netherlands
- 733 15. Amsterdam Public Health research institute, Amsterdam, the Netherlands
- 734 16. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research,  
735 Amsterdam, Vrije Universiteit Amsterdam
- 736 17. Population Health and Genomics, University of Dundee, Dundee, Scotland, UK.
- 737 18. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States
- 738 19. NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO  
739 63110, USA
- 740 20. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK
- 741 21. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
- 742 22. Precision Health Research (PRECISE), Singapore
- 743 23. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,  
744 United Kingdom
- 745 24. Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
- 746 25. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
- 747 26. Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen,  
748 The Netherlands
- 749 27. Department of Internal Medicine, Erasmus MC, Erasmus University Medical Center Rotterdam,  
750 Rotterdam, The Netherlands
- 751 28. <https://directdiabetes.org/>
- 752 29. NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, MO  
753 63110, USA
- 754 30. Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA
- 755 31. Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis,  
756 MO, United States
- 757 32. Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO,  
758 United States
- 759 33. Dominantly Inherited Alzheimer Disease Network (DIAN)
- 760 34. Department of Neurology, Faculty of Medicine, Juntendo University, Tokyo, Japan
- 761 35. Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, MD, USA
- 762 36. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
- 763 37. Human Genetics Department, Sidra Medicine, Doha, Qatar
- 764 38. Institute of Human Genetics, School of Medicine, Technische Universität München, Munich, Germany
- 765 39. Institute of Neurogenomics, Computational Health Center, Helmholtz Zentrum München, Munich,  
766 Germany.
- 767 40. Ontario Institute for Cancer Research
- 768 41. Oncode Institute, Groningen, The Netherlands
- 769 42. Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK
- 770 43. Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo,  
771 Japan.
- 772 44. Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda,  
773 MD, USA
- 774 45. Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA, USA.
- 775 46. Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University,  
776 Tampere, Finland

- 777 47. Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology,  
778 Tampere University, Tampere, Finland
- 779 48. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
- 780 49. M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan.
- 781 50. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary  
782 Care, University of Cambridge, Cambridge, UK
- 783 51. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK
- 784 52. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care,  
785 University of Cambridge, Cambridge, UK
- 786 53. Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, VIC,  
787 Australia
- 788 54. Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge,  
789 Cambridge, UK
- 790 55. British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK.
- 791 56. LIFE – Leipzig Research Center for Civilization Diseases, Leipzig University, Germany
- 792 57. Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Germany
- 793 58. Centre for Human Genetics, University of Oxford, Oxford, UK
- 794 59. Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, D-04103  
795 Leipzig, Germany
- 796 60. Department of Medicine, University of Leipzig, D-04103 Leipzig, Germany
- 797 61. Medical Faculty, University of Leipzig, Leipzig, Germany.
- 798 62. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen,  
799 9700 RB, the Netherlands
- 800 63. Nitte (Deemed to Be University), Medical Sciences Complex, Nitte University Centre for Science  
801 Education and Research (NUCSER), Deralakatte, Mangalore, 575018, India
- 802 64. Department of Clinical Physiology, Tampere University Hospital, Tampere Finland
- 803 65. Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA,  
804 USA
- 805 66. Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- 806 67. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
- 807 68. Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- 808 69. Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Leipzig, Germany
- 809 70. Leipzig Research Centre for Civilization Diseases, Leipzig University, Leipzig, Germany
- 810 71. National Skin Centre, Research Division, Singapore
- 811 72. Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore
- 812 73. Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones  
813 Científicas (IPBLN-CSIC), Granada, Spain
- 814 74. Unit of Genomic of Complex Diseases, Institut de Recerca Sant Pau (IR Sant Pau), Barcelona, Spain.
- 815 75. Centre for Population Health Research, University of Turku and Turku University Hospital, Turku,  
816 Finland
- 817 76. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,  
818 Finland
- 819 77. Institute for Medical Biometry, Epidemiology and Informatics (IMBEI), Mainz, Germany
- 820 78. Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan.
- 821 79. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald,  
822 Germany
- 823 80. Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo,  
824 Japan
- 825 81. Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan

- 826 82. Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan  
827 83. Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka  
828 University, Suita, Japan  
829 84. Premium Research Institute for Human Metaverse Medicine (WPI-PRIME), Osaka University, Suita,  
830 Japan.  
831 85. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David  
832 Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA  
833 86. Paavo Nurmi Centre, Unit of Health and Physical Activity, University of Turku, Turku, Finland  
834 87. Department of Computational Medicine, David Geffen School of Medicine, University of California Los  
835 Angeles, Los Angeles, CA, USA  
836 88. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of  
837 California Los Angeles, Los Angeles, CA, USA  
838 89. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca,  
839 Cambridge, UK.  
840 90. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care,  
841 University of Cambridge, UK.  
842 91. Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands  
843 92. Institute of Neurogenomics, Computational Health Center, Helmholtz Zentrum München, Neuherberg,  
844 Germany  
845 93. School of Medicine, Institute of Human Genetics, Technical University of Munich, Munich, Germany  
846 94. Research centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,  
847 Finland  
848 95. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland  
849 96. InFLAMES Research Flagship, University of Turku, Turku, Finland  
850 97. Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere,  
851 Finland  
852 98. Tampere University Hospital, Tampere, Finland  
853 99. Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry & Amsterdam  
854 Neuroscience -Complex Trait Genetics (VUmc) and Mood, Anxiety, Psychosis, Stress & Sleep  
855 100. Section of Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University  
856 Medical Center, Leiden, the Netherlands  
857 101. Medical Department III-Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical  
858 Center, Leipzig, Germany  
859 102. Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG), Helmholtz Zentrum  
860 München, University of Leipzig and University Hospital Leipzig, Leipzig, Germany  
861 103. Deutsches Zentrum für Diabetesforschung, Neuherberg, Germany  
862 104. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany  
863 105. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany  
864 106. Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig  
865 Medical Center, Leipzig, Germany  
866 107. Department of Neurology, UMC Utrecht Brain Center Rudolf Magnus, Utrecht, The Netherlands  
867 108. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands  
868 109. Biosciences Institute, Faculty of Medical Sciences, University of Newcastle, Newcastle upon Tyne,  
869 UK  
870 110. German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, D-17475 Greifswald,  
871 Germany  
872 111. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Felix-  
873 Hausdorff-Strasse 8, D-17475 Greifswald, Germany



- 874 112. Interfaculty Institute of Genetics and Functional Genomics, University Medicine Greifswald,  
875 Greifswald, 17475, Germany  
876 113. German Center for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, 17475,  
877 Germany.  
878 114. CARIM, Maastricht University  
879 115. Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, The  
880 Netherlands  
881 116. Department of Orthopaedics and Sportsmedicine, Erasmus MC University Medical Center,  
882 Rotterdam, The Netherlands.  
883

## 884 Supplementary Note

### 885 **MYBL2 regulates the expression of many cell cycle genes**

886 At the *MYBL2* locus, the lead variant chr20\_43721344\_C\_T was associated with the expression  
887 of 151 target genes at FDR 5% (Figure S8). The target genes were strongly enriched for the  
888 Gene Ontology mitotic cell cycle term (GO:0000278,  $p=2.607 \times 10^{-51}$ ) and the Reactome mitotic  
889 cell cycle pathway (R-HSA-69278,  $p=1.216 \times 10^{-35}$ ). Interestingly, 125/151 genes (82%) had  
890 lower expression in carriers of the alternative T allele (Figure S8). The T allele of  
891 chr20\_43721344\_C\_T was also strongly associated ( $p=8.83 \times 10^{-218}$ ) with the decreased  
892 expression of the *MYBL2* transcription factor gene in *cis* (Figure S8). Since both the *MYBL2*  
893 transcription factor located in *cis* and majority of the *trans*-genes had lower expression in the  
894 carriers of the T allele, we hypothesised that *MYBL2* might directly regulate these target genes.  
895 To test this, we download ChIP-seq data for the *MYBL2* transcription factor in the human K562  
896 myelogenous leukemia cell line from the ENCODE project (ENCSR162IEM). We then asked  
897 how many of the up- and downregulated genes had a *MYBL2* ChIP-seq peak within +/- 2kb  
898 from the annotated promoter of the gene. We found that 99/125 (78.4%) downregulated genes  
899 had a *MYBL2* peak in their promoter region (Figure S9). In contrast, only 1/26 upregulated  
900 genes had a *MYBL2* peak in their promoter region. As a negative control, we looked at the 404  
901 5% FDR target genes of the *SP140* locus (Table S2) and found that only 23/404 (5.6%) of the  
902 target genes had a *MYBL2* peak in their promoter region (Figure S9).

903  
904 To further understand which cell cycle stage these *MYBL2* target genes might be involved in,  
905 we obtained the list of genes specific to G2M and S phases of the cell cycle from Tirosh et al.  
906 2015<sup>102</sup> using Seurat R package<sup>103</sup>. We found that 33/125 genes downregulated by the *trans*-  
907 eQTL variant were markers of the G2M phase which was significantly more than expected ( $p =$   
908  $1.23 \times 10^{-53}$ ). In contrast, only 1/125 downregulated genes overlapped with markers of the S  
909 phase ( $p = 0.36$ ). Of note, 2/26 upregulated genes overlapped S-phase markers ( $p = 0.004$ ) and  
910 none of the upregulated genes overlapped G2M-phase markers.

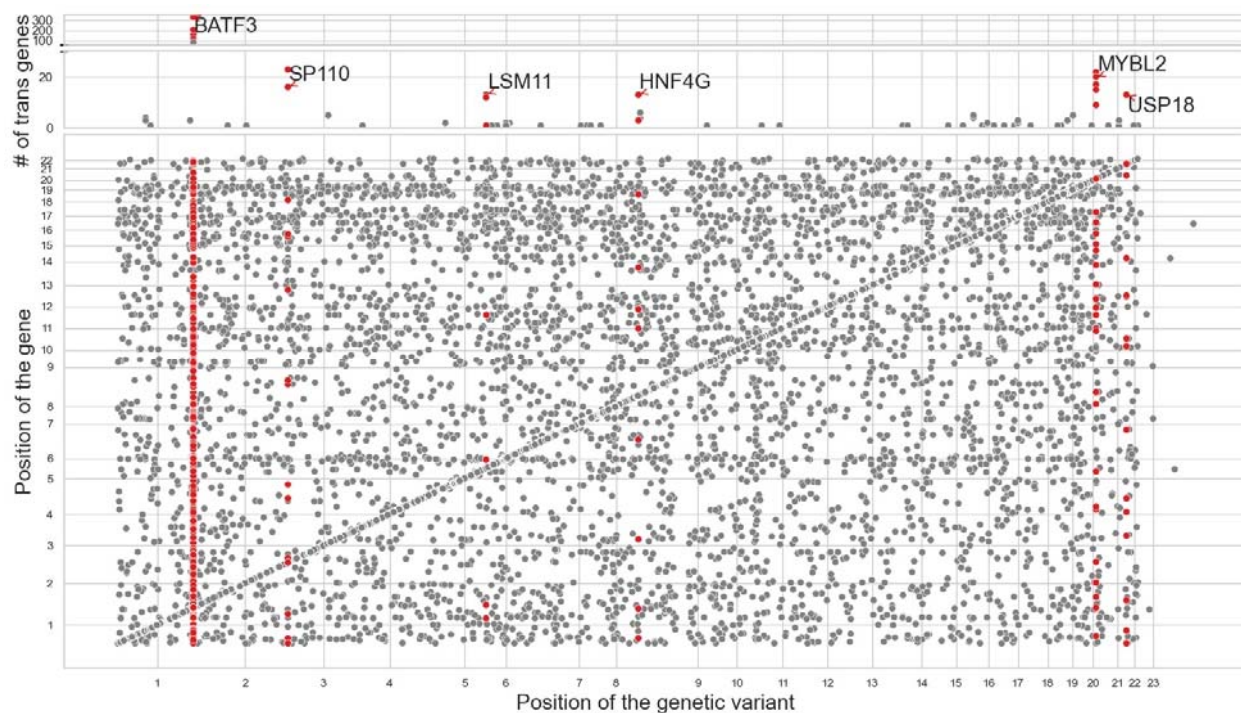
911  
912 Altogether, this evidence strongly suggests that *MYBL2* directly regulates the expression of  
913 G2M genes in *trans* by binding to their promoter sequences and is directly involved in the  
914 regulation of the expression of these target genes.

915

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

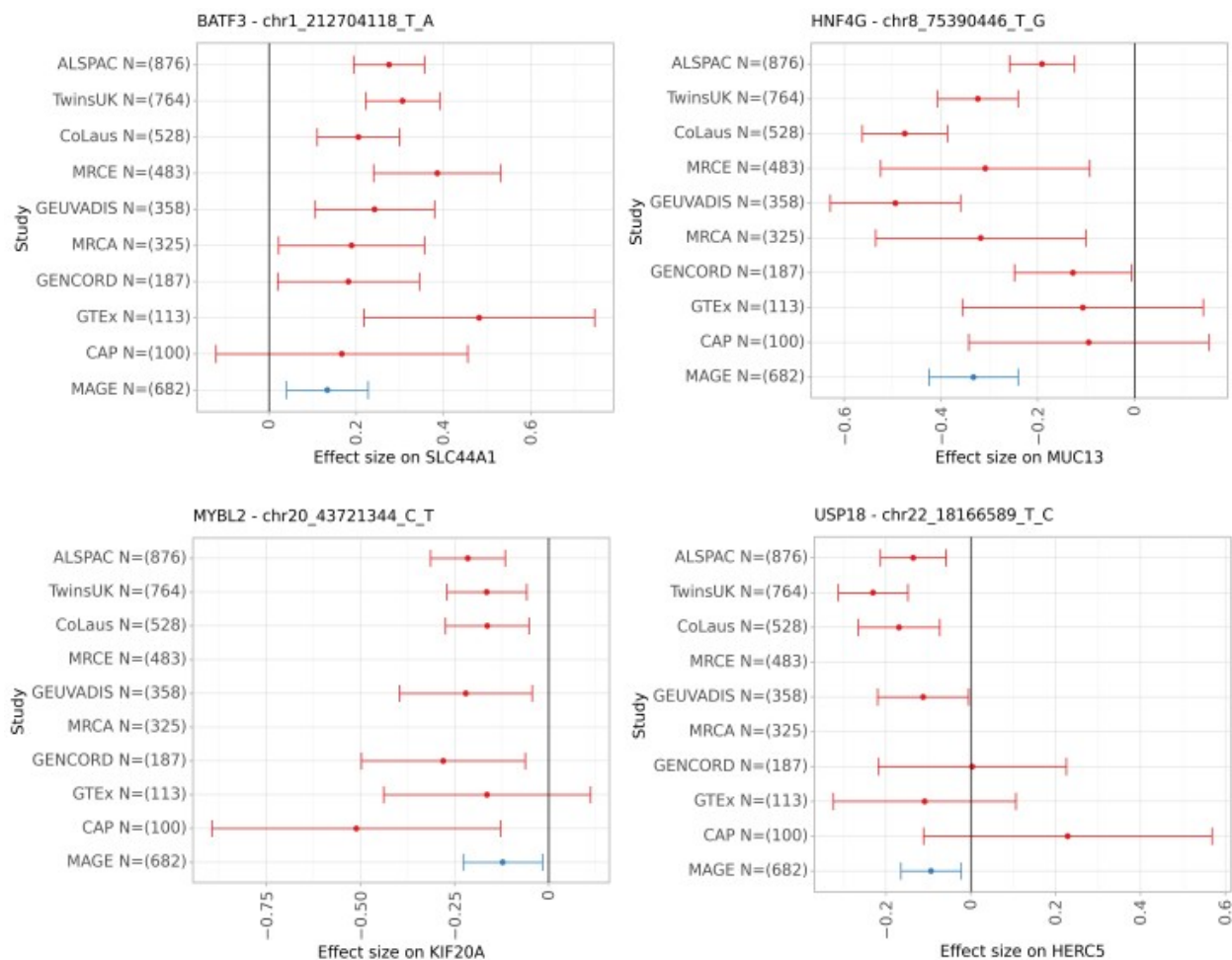
Cohort	Sample size	Expression data	Genotype data	Relatedness
ALSPAC <sup>31,73,74</sup>	877	microarray	imputed (1000G 30x on GRCh38)	unrelated
TwinsUK <sup>75</sup>	735	RNA-seq	imputed (1000G 30x on GRCh38)	twins
CoLaus <sup>76,77</sup>	553	RNA-seq	imputed (TOPMed)	unrelated
GEUVADIS <sup>78</sup>	358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

919  
920

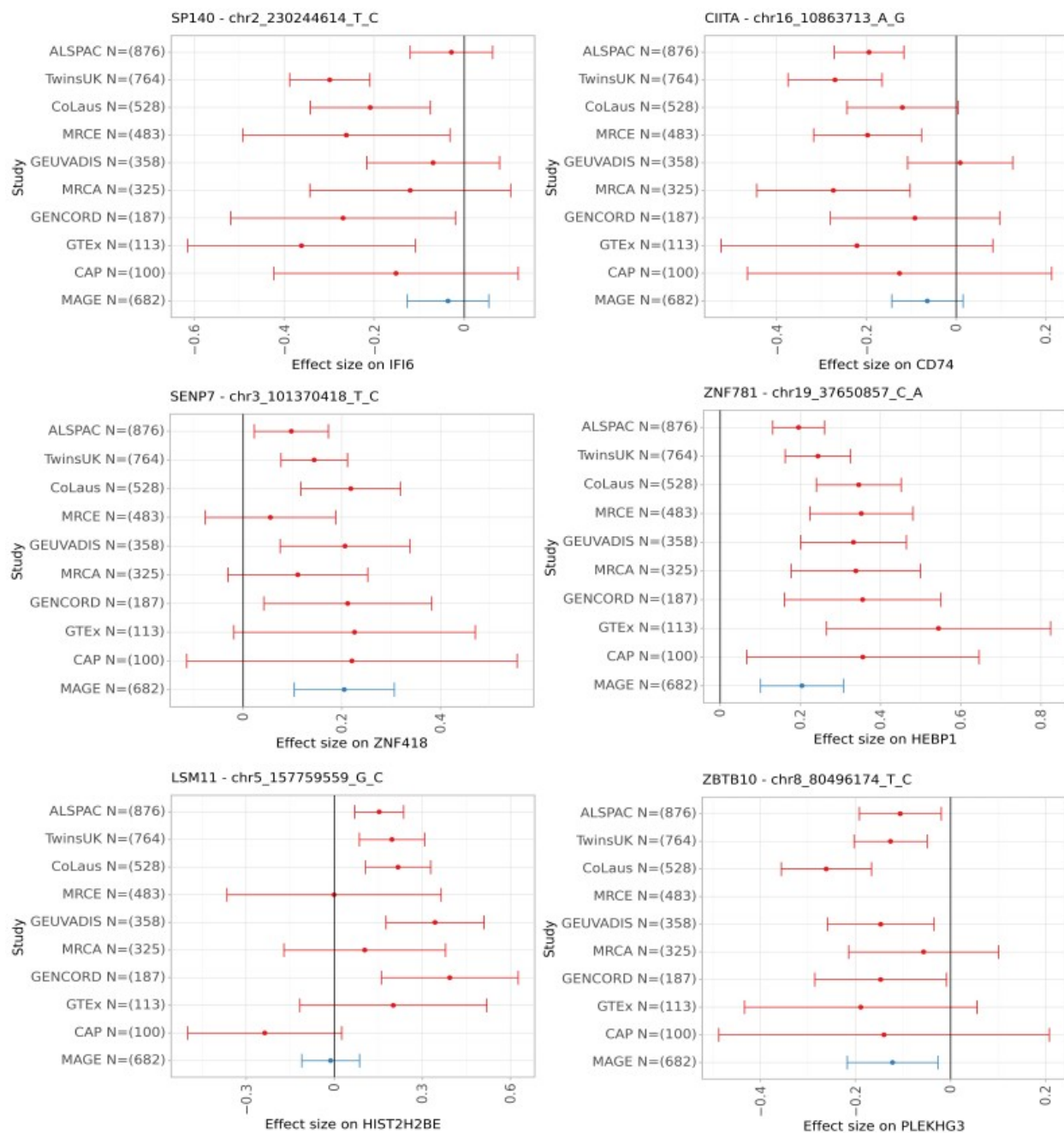


921  
922 **Figure S1.** Overview of *trans*-eQTL analysis at the relaxed  $p < 5 \times 10^{-8}$  threshold. The upper  
923 scatter plot shows the number of *trans* associations detected at each *trans*-eQTL locus with  $p$ -  
924 values  $< 5 \times 10^{-8}$ . Six largest *trans*-eQTL loci have been labelled with the name of the closest  
925 *cis* gene. The lower scatter plot shows all significant loci for each tested gene at the  $p <$   
926  $5 \times 10^{-8}$  threshold. *Cis* associations are located on the diagonal while putative *trans*  
927 associations are located off diagonal.

928  
929  
930  
931



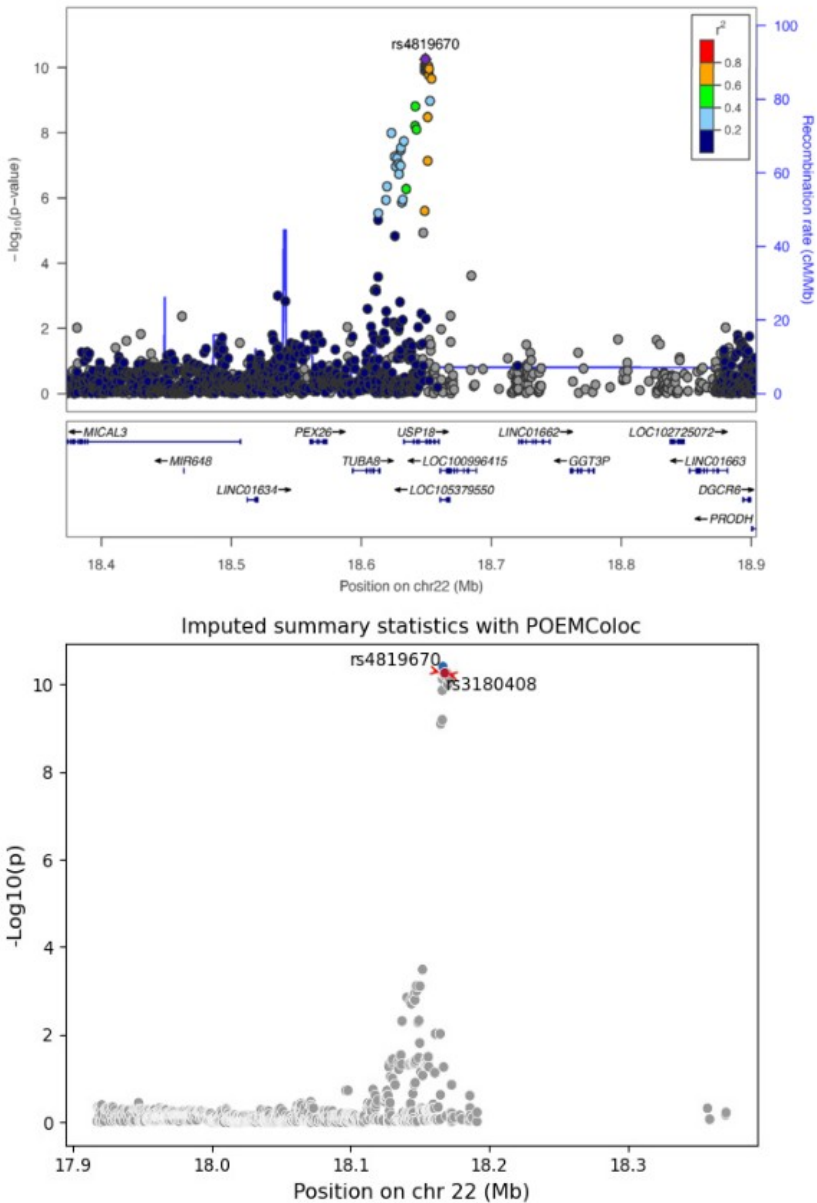
932  
933 **Figure S2. Forest plots showing cohort-specific effect size for the four *trans*-eQTL loci**  
934 **that replicated in the MAGE cohort.** The points represent the *trans*-eQTL effect size estimates  
935 from regenie and the error bars represent 95% confidence intervals.



936  
937  
938  
939  
940  
941  
942

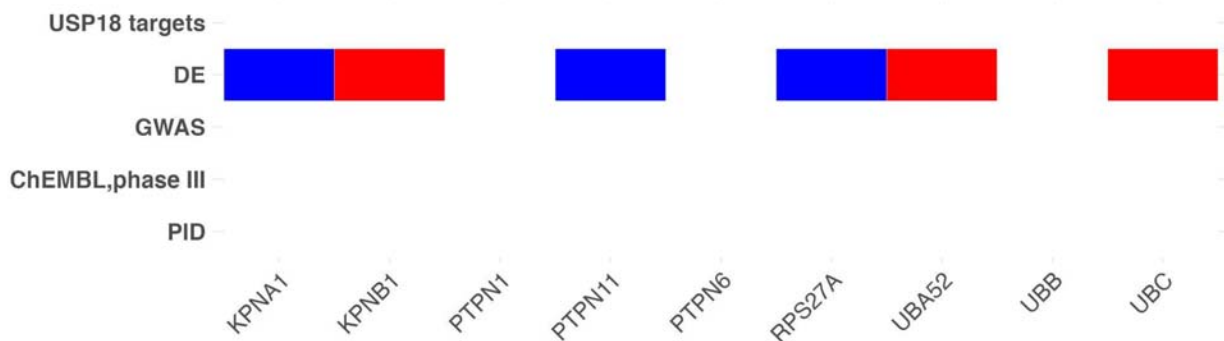
**Figure S3. Forest plots showing cohort-specific effect size for the remaining *trans*-eQTL loci that either did not replicate in the MAGE cohort or corresponded to likely cross-mappability artefacts (*SENP7*, *ZNF781* and *ZBTB10* loci).** The points represent the *trans*-eQTL effect size estimates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

## region 108: rs4819670



943  
944  
945  
946

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

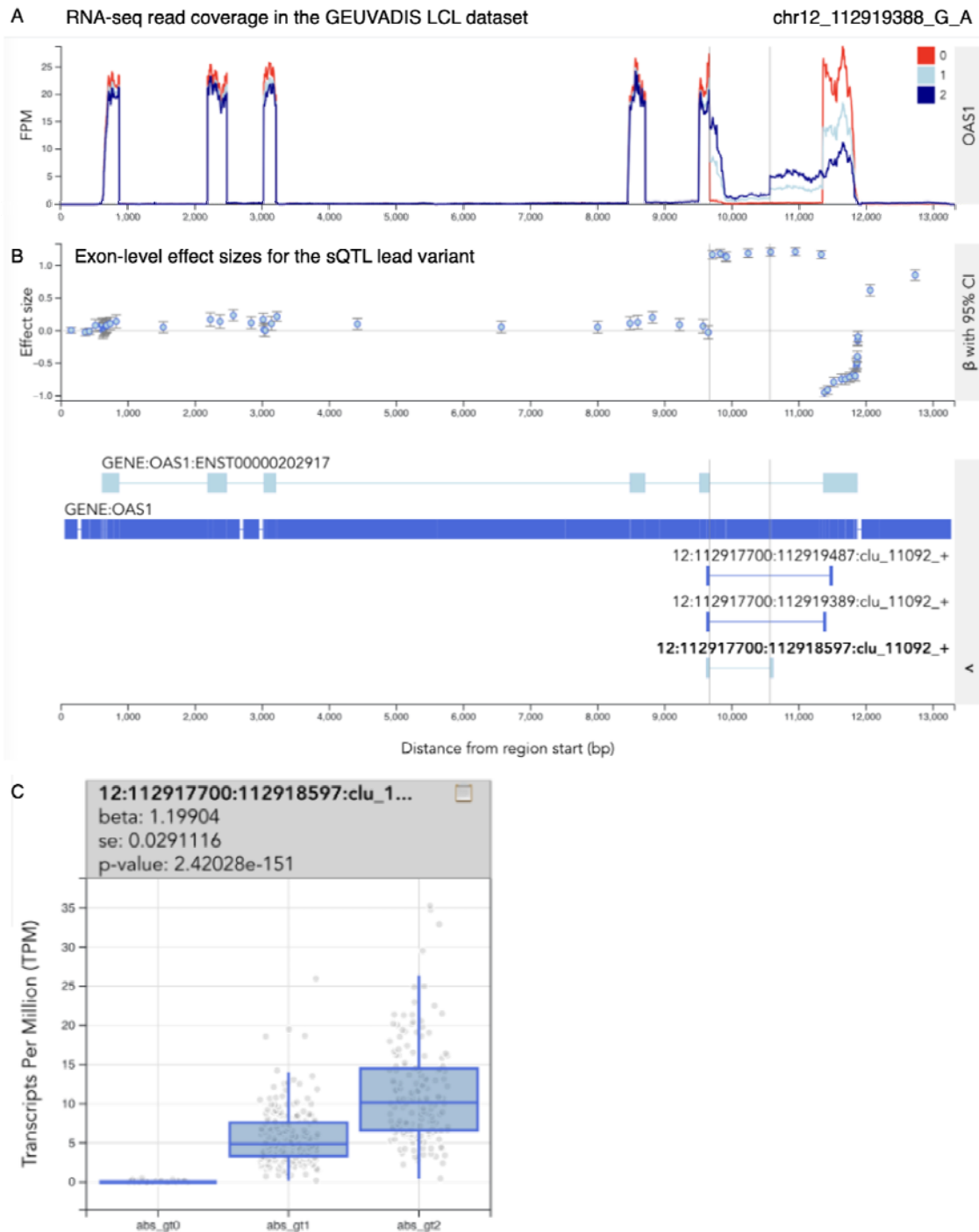


947

948

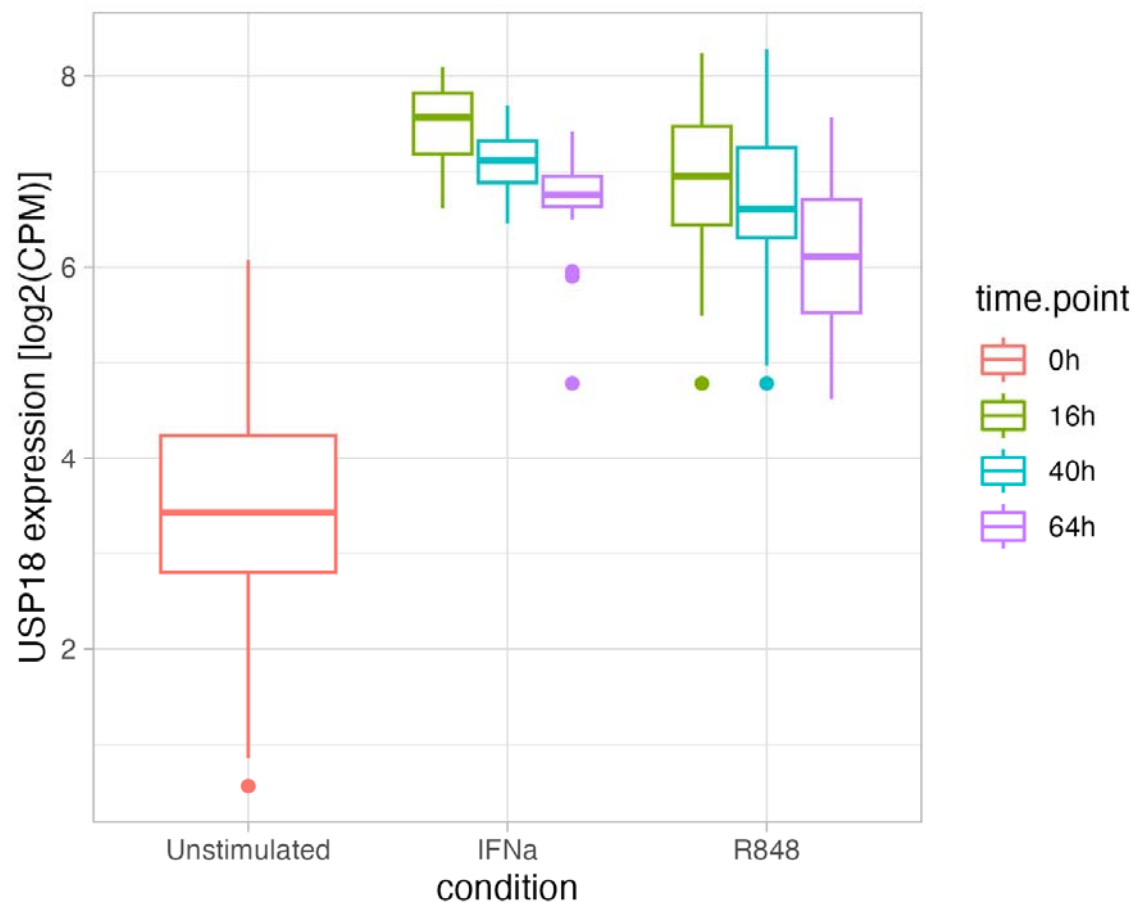
949 **Figure S5. Category III: other interferon alpha/beta signalling pathway genes that do not**  
950 **belong to categories I or II (shown in Figure 3).** The increased gene expression is marked in  
951 red, while reduced gene expression is marked in blue. The visualisation illustrates the effect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





955  
956 **Figure S6. Fine-mapped splicing QTL (sQTL) in the *OAS1* gene.** (A) RNA-seq read  
957 coverage of the *OAS1* gene in the GEUVADIS LCL dataset, stratified by the genotype of the  
958 fine-mapped sQTL variant chr12\_112919388\_G\_A (posterior inclusion probability = 1). (B)  
959 Exon-level effect sizes for the sQTL lead variant. (C) Boxplot of the absolute expression of the  
960 short last intron of the *OAS1* gene (highlighted on panel A) stratified by the genotype of the lead  
961 sQTL variant. Interactive version of the plot can be viewed [here](#).  
962

963



964

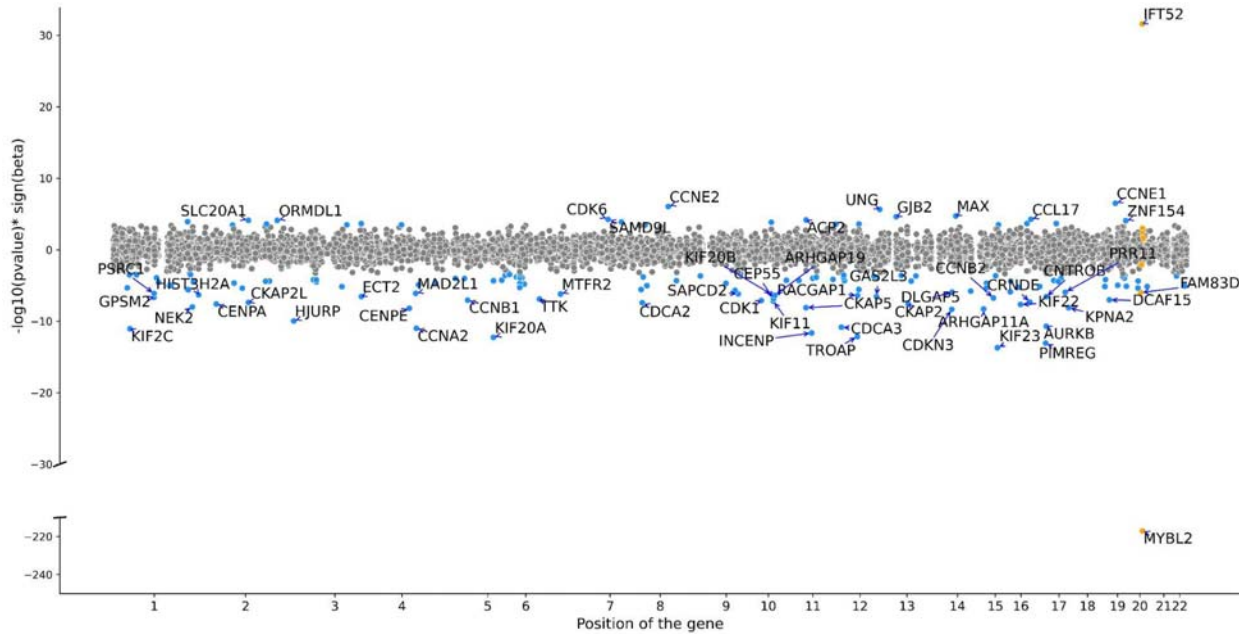
965

966

967

**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.

968



969

970

971

972

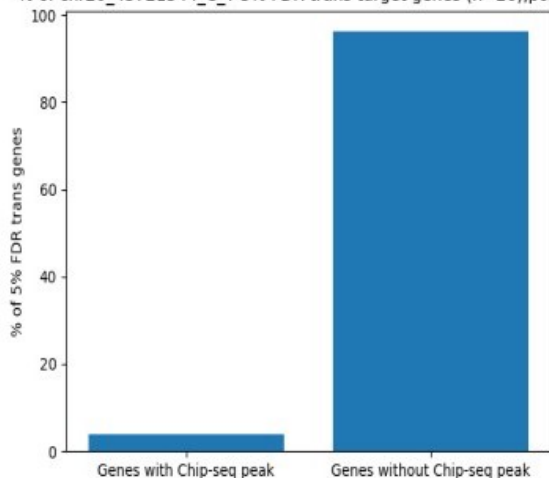
973

974

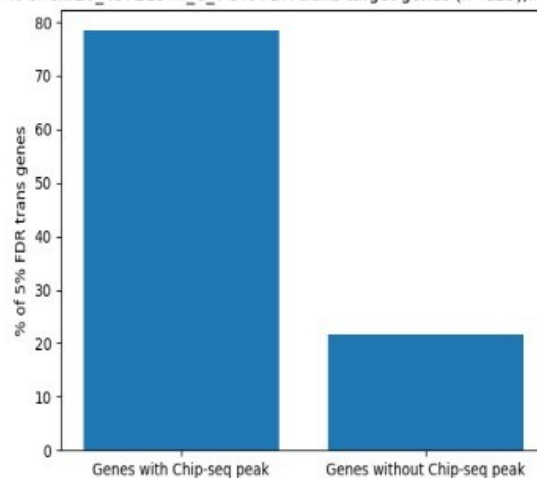
**Figure S8. MYBL2 regulates the expression of many cell cycle genes.** The scatter plot shows all genes associated with the *MYBL2* trans-eQTL lead variant (chr20\_43721344\_C\_T). Light blue points show significantly associated genes (variant-level Benjamini-Hochberg FDR 5%)

### MYBL2 - chr20\_43721344\_C\_T

% of chr20\_43721344\_C\_T 5% FDR trans target genes (n=26), pos beta

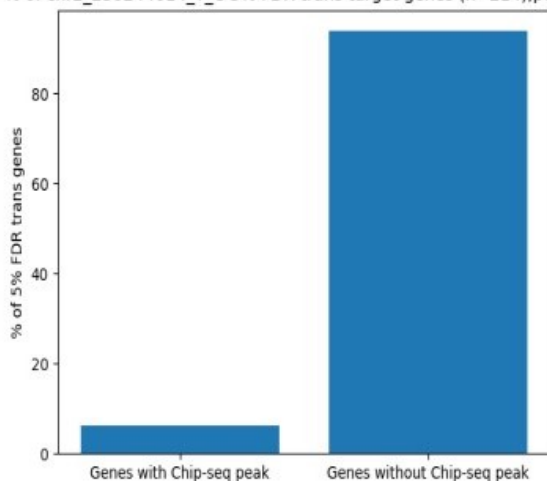


% of chr20\_43721344\_C\_T 5% FDR trans target genes (n=125), neg beta

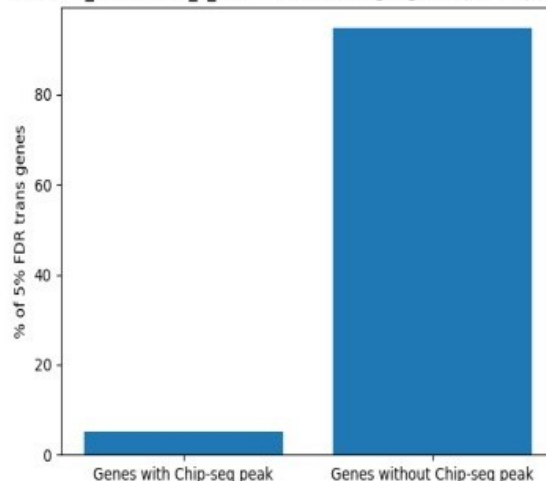


### SP140 - chr2\_230244614\_T\_C

% of chr2\_230244614\_T\_C 5% FDR trans target genes (n=214), pos beta



% of chr2\_230244614\_T\_C 5% FDR trans target genes (n=190), neg beta



975

976

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

984

## 985 References

- 986 1. Aguet, F. *et al.* Molecular quantitative trait loci. *Nature Reviews Methods Primers* **3**, 1–22  
987 (2023).
- 988 2. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines  
989 fetal hemoglobin level. *Science* **342**, 253–257 (2013).
- 990 3. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and  $\beta$ -Thalassemia.  
991 *N. Engl. J. Med.* **384**, 252–260 (2021).
- 992 4. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic  
993 Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
- 994 5. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK  
995 Biobank. *Nature* 1–10 (2023).
- 996 6. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci  
997 and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310  
998 (2021).
- 999 7. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known  
1000 disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- 1001 8. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and  
1002 disease. *Nat. Genet.* **53**, 1712–1721 (2021).
- 1003 9. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
- 1004 10. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science*  
1005 **374**, eabj1541 (2021).
- 1006 11. Bonder, M. J. *et al.* Identification of rare and common regulatory variants in pluripotent cells  
1007 using population-scale transcriptomics. *Nat. Genet.* **53**, 313–321 (2021).
- 1008 12. Kolberg, L., Kerimov, N., Peterson, H. & Alasoo, K. Co-expression analysis reveals  
1009 interpretable gene modules controlled by trans-acting genetic variants. *Elife* **9**, e58705

- 1010 (2020).
- 1011 13. Kasela, S. *et al.* Pathogenic implications for autoimmune mechanisms derived by  
1012 comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet.* **13**, e1006643  
1013 (2017).
- 1014 14. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon  
1015 monocyte gene expression. *Science* **343**, 1246949 (2014).
- 1016 15. Brandt, M. *et al.* An autoimmune disease risk variant: A trans master regulatory effect  
1017 mediated by IRF1 under immune stimulation? *PLoS Genet.* **17**, e1009684 (2021).
- 1018 16. Small, K. S. *et al.* Identification of an imprinted master trans regulator at the KLF14 locus  
1019 related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
- 1020 17. Consortium, T. G. *et al.* The GTEx Consortium atlas of genetic regulatory effects across  
1021 human tissues. *Science* **369**, 1318–1330 (2020).
- 1022 18. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of  
1023 autoimmune disease. *Science* **376**, eabf3041 (2022).
- 1024 19. Small, K. S. *et al.* Regulatory variants at KLF14 influence type 2 diabetes risk via a female-  
1025 specific effect on adipocyte size and body composition. *Nat. Genet.* **50**, 572–580 (2018).
- 1026 20. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.  
1027 *N. Engl. J. Med.* **373**, 895–907 (2015).
- 1028 21. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional  
1029 connections with IRX3. *Nature* **507**, 371–375 (2014).
- 1030 22. Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer* **4**, 757–  
1031 768 (2004).
- 1032 23. Bjornevik, K. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus  
1033 associated with multiple sclerosis. *Science* **375**, 296–301 (2022).
- 1034 24. Bar-Or, A. *et al.* Epstein-Barr Virus in Multiple Sclerosis: Theory and Emerging  
1035 Immunotherapies. *Trends Mol. Med.* **26**, 296–310 (2020).

- 1036 25. James, J. A. *et al.* An increased prevalence of Epstein-Barr virus infection in young patients  
1037 suggests a possible etiology for systemic lupus erythematosus. *J. Clin. Invest.* **100**, 3019–  
1038 3026 (1997).
- 1039 26. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated  
1040 in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).
- 1041 27. Lanz, T. V. *et al.* Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and  
1042 GlialCAM. *Nature* **603**, 321–327 (2022).
- 1043 28. Thomas, O. G. *et al.* Heightened Epstein-Barr virus immunity and potential cross-  
1044 reactivities in multiple sclerosis. *PLoS Pathog.* **20**, e1012177 (2024).
- 1045 29. Saha, A. & Battle, A. False positives in trans-eQTL and co-expression analyses arising  
1046 from RNA-sequencing alignment errors. *F1000Res.* **7**, 1860 (2018).
- 1047 30. Taylor, D. J. *et al.* Sources of gene expression variation in a globally diverse cohort. *bioRxiv*  
1048 2023.11.04.565639 (2023) doi:10.1101/2023.11.04.565639.
- 1049 31. Bryois, J. *et al.* Cis and trans effects of human genomic variants on gene expression. *PLoS*  
1050 *Genet.* **10**, e1004461 (2014).
- 1051 32. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes  
1052 at all published human GWAS trait-associated loci. *Nat. Genet.* 1–7 (2021).
- 1053 33. Yin, X. *et al.* Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for  
1054 systemic lupus erythematosus. *Ann. Rheum. Dis.* **80**, 632–640 (2021).
- 1055 34. King, E. A., Dunbar, F., Davis, J. W. & Degner, J. F. Estimating colocalization probability  
1056 from limited summary statistics. *BMC Bioinformatics* **22**, 254 (2021).
- 1057 35. Meuwissen, M. E. C. *et al.* Human USP18 deficiency underlies type 1 interferonopathy  
1058 leading to severe pseudo-TORCH syndrome. *J. Exp. Med.* **213**, 1163–1174 (2016).
- 1059 36. Alsohime, F. *et al.* JAK Inhibitor Therapy in a Child with Inherited USP18 Deficiency. *N.*  
1060 *Engl. J. Med.* **382**, 256–265 (2020).
- 1061 37. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

- 1062 38. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of  
1063 multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
- 1064 39. Schnitzler, G. R. *et al.* Convergence of coronary artery disease genes onto endothelial cell  
1065 programs. *Nature* 1–9 (2024).
- 1066 40. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes  
1067 underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
- 1068 41. Mostafavi, S. *et al.* Parsing the Interferon Transcriptional Network and Its Disease  
1069 Associations. *Cell* **164**, 564–578 (2016).
- 1070 42. Banchereau, R. *et al.* Personalized Immunomonitoring Uncovers Molecular Networks that  
1071 Stratify Lupus Patients. *Cell* **165**, 551–565 (2016).
- 1072 43. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood  
1073 cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2610–2615 (2003).
- 1074 44. Jefferies, C. A. Regulating IRFs in IFN Driven Disease. *Front. Immunol.* **10**, 325 (2019).
- 1075 45. Honda, K., Takaoka, A. & Taniguchi, T. Type I interferon [corrected] gene induction by the  
1076 interferon regulatory factor family of transcription factors. *Immunity* **25**, 349–360 (2006).
- 1077 46. Reshef, Y. A. *et al.* Detecting genome-wide directional effects of transcription factor binding  
1078 on polygenic disease risk. *Nat. Genet.* **50**, 1483–1493 (2018).
- 1079 47. Kerimov, N. *et al.* eQTL Catalogue 2023: New datasets, X chromosome QTLs, and  
1080 improved detection and visualisation of transcript-level QTLs. *PLoS Genet.* **19**, e1010932  
1081 (2023).
- 1082 48. Li, H. *et al.* Identification of a Sjögren’s syndrome susceptibility locus at OAS1 that  
1083 influences isoform switching, protein expression, and responsiveness to type I interferons.  
1084 *PLoS Genet.* **13**, e1006820 (2017).
- 1085 49. Banday, A. R. *et al.* Genetic regulation of OAS1 nonsense-mediated decay underlies  
1086 association with COVID-19 hospitalization in patients of European and African ancestries.  
1087 *Nat. Genet.* **54**, 1103–1116 (2022).



- 1088 50. Zdrzil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning  
1089 multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192  
1090 (2024).
- 1091 51. Morand, E. F. *et al.* Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N. Engl.*  
1092 *J. Med.* **382**, 211–221 (2020).
- 1093 52. Amaya-Uribe, L., Rojas, M., Azizi, G., Anaya, J.-M. & Gershwin, M. E. Primary  
1094 immunodeficiency and autoimmunity: A comprehensive review. *J. Autoimmun.* **99**, 52–72  
1095 (2019).
- 1096 53. Schmidt, R. E., Grimbacher, B. & Witte, T. Autoimmunity and primary immunodeficiency:  
1097 two sides of the same coin? *Nat. Rev. Rheumatol.* **14**, 7–18 (2017).
- 1098 54. Primary immunodeficiency or monogenic inflammatory bowel disease (Version 4.191).  
1099 <https://panelapp.genomicsengland.co.uk/panels/398/>.
- 1100 55. Mackensen, A. *et al.* Anti-CD19 CAR T cell therapy for refractory systemic lupus  
1101 erythematosus. *Nat. Med.* **28**, 2124–2132 (2022).
- 1102 56. Müller Fabian *et al.* CD19 CAR T-Cell Therapy in Autoimmune Disease — A Case Series  
1103 with Follow-up. *N. Engl. J. Med.* **390**, 687–700 (2024).
- 1104 57. Akinbiyi, T., McPeck, M. S. & Abney, M. ADELLE: A global testing method for Trans-eQTL  
1105 mapping. *bioRxiv* 2024.02.24.581871 (2024) doi:10.1101/2024.02.24.581871.
- 1106 58. Dutta, D. *et al.* Aggregative trans-eQTL analysis detects trait-specific target gene sets in  
1107 whole blood. *Nat. Commun.* **13**, 1–14 (2022).
- 1108 59. Wang, L., Babushkin, N., Liu, Z. & Liu, X. Trans-eQTL mapping in gene sets identifies  
1109 network effects of genetic variants. *bioRxiv* 2022.11.11.516189 (2022)  
1110 doi:10.1101/2022.11.11.516189.
- 1111 60. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in  
1112 infectious disease. *Nat. Rev. Immunol.* **15**, 87–103 (2015).
- 1113 61. Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium *et al.* The

- 1114 interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian  
1115 randomisation analysis. *Lancet* **379**, 1214–1224 (2012).
- 1116 62. C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC) *et al.*  
1117 Association between C reactive protein and coronary heart disease: mendelian  
1118 randomisation analysis based on individual participant data. *BMJ* **342**, d548 (2011).
- 1119 63. Burgess, S. *et al.* Using genetic association data to guide drug discovery and development:  
1120 Review of methods and applications. *Am. J. Hum. Genet.* **110**, 195–214 (2023).
- 1121 64. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the  
1122 plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
- 1123 65. Karim, M. A. *et al.* Systematic disease-agnostic identification of therapeutically actionable  
1124 targets using the genetics of human plasma proteins. *medRxiv* (2023)  
1125 doi:10.1101/2023.06.01.23290252.
- 1126 66. Wong, D. *et al.* Genomic mapping of the MHC transactivator CIITA using an integrated  
1127 ChIP-seq and genetical genomics approach. *Genome Biol.* **15**, 494 (2014).
- 1128 67. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic  
1129 Traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).
- 1130 68. Hore, V. *et al.* Tensor decomposition for multiple-tissue gene expression experiments. *Nat.*  
1131 *Genet.* **48**, 1094–1100 (2016).
- 1132 69. Ntranos, V., Kamath, G. M., Zhang, J. M., Pachter, L. & Tse, D. N. Fast and accurate  
1133 single-cell RNA-seq analysis by clustering of transcript-compatibility counts. *Genome Biol.*  
1134 **17**, 112 (2016).
- 1135 70. EQTLGen consortium. <https://www.eqtlgen.org/>.
- 1136 71. Morris, J. A. *et al.* Discovery of target genes and pathways at GWAS loci by pooled single-  
1137 cell CRISPR screens. *Science* **380**, eadh7699 (2023).
- 1138 72. Weinstock, J. S. *et al.* Gene regulatory network inference from CRISPR perturbations in  
1139 primary CD4+ T cells elucidates the genomic basis of immune disease. *bioRxiv*

- 1140 2023.09.17.557749 (2023) doi:10.1101/2023.09.17.557749.
- 1141 73. Boyd, A. *et al.* Cohort Profile: the ‘children of the 90s’--the index offspring of the Avon  
1142 Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
- 1143 74. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children:  
1144 ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
- 1145 75. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome  
1146 sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).
- 1147 76. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the  
1148 epidemiology and genetic determinants of cardiovascular risk factors and metabolic  
1149 syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
- 1150 77. Sönmez Flitman, R. *et al.* Untargeted Metabolome- and Transcriptome-Wide Association  
1151 Study Suggests Causal Genes Modulating Metabolite Concentrations in Urine. *J. Proteome*  
1152 *Res.* **20**, 5103–5114 (2021).
- 1153 78. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation  
1154 in humans. *Nature* **501**, 506–511 (2013).
- 1155 79. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived  
1156 from lymphoblastoid cell lines. *Genome Res.* **23**, 716–726 (2013).
- 1157 80. Gutierrez-Arcelus, M. *et al.* Passive and active DNA methylation and the interplay with  
1158 genetic variation in gene regulation. *Elife* **2**, e00523 (2013).
- 1159 81. Theusch, E., Chen, Y.-D. I., Rotter, J. I., Krauss, R. M. & Medina, M. W. Genetic variants  
1160 modulate gene expression statin response in human lymphoblastoid cell lines. *BMC*  
1161 *Genomics* **21**, 555 (2020).
- 1162 82. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome  
1163 assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
- 1164 83. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000  
1165 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).

- 1166 84. Deelen, P. *et al.* Genotype harmonizer: automatic strand alignment and format conversion  
1167 for genotype data integration. *BMC Res. Notes* **7**, 901 (2014).
- 1168 85. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.  
1169 *Nature* **590**, 290–299 (2021).
- 1170 86. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**,  
1171 1284–1287 (2016).
- 1172 87. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation.  
1173 *Bioinformatics* **31**, 782–784 (2015).
- 1174 88. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium  
1175 panel. *Nat. Genet.* **48**, 1443–1448 (2016).
- 1176 89. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and  
1177 splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021).
- 1178 90. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory  
1179 requirements. *Nat. Methods* **12**, 357–360 (2015).
- 1180 91. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic  
1181 features. *Bioinformatics* **26**, 841–842 (2010).
- 1182 92. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip  
1183 data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
- 1184 93. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and  
1185 binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
- 1186 94. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human  
1187 tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
- 1188 95. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl.*  
1189 *Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
- 1190 96. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression  
1191 analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

- 1192 97. Roshchupkin, G. V. *et al.* HASE: Framework for efficient high-dimensional association  
1193 analyses. *Sci. Rep.* **6**, 36076 (2016).
- 1194 98. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python.  
1195 in *Proceedings of the 9th Python in Science Conference (SciPy, 2010)*.  
1196 doi:10.25080/majora-92bf1922-011.
- 1197 99. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-  
1198 cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
- 1199 100. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29  
1200 (2021).
- 1201 101. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor  
1202 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–  
1203 4297 (2012).
- 1204 102. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation  
1205 programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
- 1206 103. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell  
1207 analysis. *Nat. Biotechnol.* **42**, 293–304 (2023).
- 1208