Title: Discovering *in vivo* cytokine eQTL interactions from a lupus clinical trial

Authors: Emma E. Davenport,[1,2,3,4] Tiffany Amariuta,[1,2,3,4,5] Maria Gutierrez-Arcelus,[1,2,3,4] Kamil Slowikowski,[1,2,3,4,5] Harm-Jan Westra,[1,2,3,4] Yang Luo,[1,2,3,4] Ciyue Shen,[6] Deepak A. Rao,[7] Ying Zhang,[8] Stephen Pearson,[9] David von Schack,[8] Jean S. Beebe,[8] Nan Bing,[8] Sally John,[10] Michael S. Vincent,[8] Baohong Zhang[8] and Soumya Raychaudhuri[1,2,3,4,5,11*]

Affiliations:
[1]Center for Data Sciences, Brigham and Women's Hospital, Boston, MA, 02115, USA.
[2]Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA.
[3]Partners Center for Personalized Genetic Medicine, Boston, MA, 02115, USA.
[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.
[5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA.
[6]Department of Cell Biology, Harvard Medical School, Boston, MA, 02115, USA
[7]Division of Rheumatology, Allergy, Immunology, Brigham and Women's Hospital, Harvard Medical School, MA 02115, USA.
[8]Pfizer Inc., Cambridge, MA, 02139, USA.
[9]Pfizer New Haven Clinical Research Unit, New Haven, CT, 06511, USA.
[10]Biogen, Cambridge, MA, 02142, USA.
[11]Faculty of Medical and Human Sciences, University of Manchester, Manchester, M13 9PL, UK.

*Correspondence to:
Soumya Raychaudhuri
77 Avenue Louis Pasteur
Harvard New Research Building, Suite 250D
Boston, MA 02446, USA.
soumya@broadinstitute.org; 617-525-4484 (tel); 617-525-4488 (fax)

Abstract:

Background: Cytokines are critical to human disease and are attractive therapeutic targets given their widespread influence on gene regulation and transcription. Defining the downstream regulatory mechanisms influenced by cytokines is central to defining drug and disease mechanisms. One promising strategy is to use interactions between expression quantitative trait loci (eQTLs) and cytokine levels to define target genes and mechanisms.

Results: In a clinical trial for anti-IL-6 in patients with systemic lupus erythematosus we measured interferon (IFN) status, anti-IL-6 drug exposure and genome-wide gene expression at three time points (379 samples from 157 individuals). First, we show that repeat transcriptomic measurements increases the number of *cis* eQTLs identified compared to using a single time point by 64%. Then, after identifying 4,818 cis-eQTLs, we observed a statistically significant enrichment of *in vivo* eQTL interactions with IFN status ($p<0.001$ by permutation) and anti-IL-6 drug exposure ($p<0.001$). We observed 210 and 72 interactions for IFN and anti-IL-6 respectively (FDR<20%). Anti-IL-6 interactions have not yet been described while 99 of the IFN interactions are novel. Finally, we found transcription factor binding motifs interrupted by eQTL interaction SNPs, pointing to key regulatory mediators of these environmental stimuli and therefore potential therapeutic targets for autoimmune diseases. In particular, genes with IFN interactions are enriched for ISRE binding site motifs, while those with anti-IL-6 interactions are enriched for IRF4 motifs.

Conclusion: This study highlights the potential to exploit clinical trial data to discover *in vivo* eQTL interactions with therapeutically relevant environmental variables.

Keywords: eQTL, interactions, clinical trials, cytokines

Background

Cytokines are critical signals used by the immune system to coordinate inflammatory responses. These factors bind to specific receptors to induce widespread transcriptional effects. Cytokines and their receptors are not only genetically associated with susceptibility to a range of human diseases, they have also emerged as effective therapeutic targets[1]. Blockade of tumour necrosis factor (TNF) was the first cytokine-directed therapy to achieve widespread use and is now used broadly to treat multiple inflammatory diseases including rheumatoid arthritis (RA), psoriasis, and inflammatory bowel disease[2]. More recently, IL-6 has emerged as a compelling therapeutic target. IL-6 levels are elevated in autoimmune diseases such as systemic lupus erythematosus (SLE) and RA. The IL-6 receptor has been successfully targeted with tocilizumab in RA[3] and giant cell arteritis[4], while IL-6 has been targeted directly with siltuximab for successful treatment of Castleman's disease[5]. In SLE, IL-6 is thought to play a role in the observed B cell hyperactivity and autoantibody production[6]. Targeting IL-6-R in SLE has shown promise in phase I trials[7] and this has led to the development of other biologics targeting IL-6 such as PF-04236921[8]. Interferon (IFN)-α, produced primarily by plasmacytoid dendritic cells, has pleiotropic effects on the immune system. It has been implicated as a key mechanism in SLE development and pathogenesis, and is being investigated as a therapeutic target[9]. Agents targeting other inflammatory cytokines, including Interleukin-1 (IL-1), IL-12, and IL-17A and IL-23 are also in clinical use to treat autoimmune conditions. Interestingly, IL-1 blockade with canakinumab has

also been recently reported to reduce risk of heart attacks, stroke and cardiovascular disease[10]. Therefore, defining the regulatory consequences of physiologic perturbations of cytokine levels will inform our understanding of both disease and drug mechanisms.

A *cis* expression quantitative trait locus (eQTL) contains a genetic variant that alters expression of a nearby gene. *Cis* eQTLs are ubiquitous across the genome[11] and while most are stable across tissues and conditions, environmental variables can alter the effects of some of them[12–18]. If an environmental change leads to disruption of regulators upstream of a gene, then it could magnify or dampen an eQTL effect, resulting in a genotype-by-environment interaction (**Figure S1**). Therefore, observing a set of eQTL interactions due to a perturbagen, such as a cytokine, can identify shared upstream regulatory mechanisms, such as transcription factors and key pathways. Even a single eQTL interaction where we can define mechanism can lead to insights about the action of the perturbagen.

However, *cis* eQTL interactions with physiologic environmental factors in humans have been challenging to discover *in vivo*[19–23] even with large cohorts[11, 17]. Success at finding *cis* eQTL interactions has largely been found in studies using model organisms[24, 25] or treating cells *in vitro* with non-physiologic conditions[26]. Thus far, these studies might be limited in power since they often map eQTLs separately across conditions and fail to exploit the power of repeat measurements[27]. In other instances,

111 they test for genetic variants associated with differential expression and miss
112 information about the magnitude of the eQTL effect in a specific condition[28].
113
114 We predicted that if the transcriptome is assayed at multiple time points under different
115 exposure states, then the repeat measurements could lead to an increase in power to
116 detect eQTLs and their interactions with environmental perturbations. If the same
117 individual is assessed at multiple times, then the noise in transcriptomic measurements
118 is reduced. Furthermore, repeat measurements from the same individuals when they
119 are both unexposed and exposed to an environmental perturbagen allow for more
120 accurate modelling of the effect of the perturbagen within those subjects.
121
122 Clinical trials, with their structured study design, may be the ideal setting to detect eQTL
123 interactions with therapeutically important variables. In clinical trials, it is becoming
124 increasingly common to collect transcriptional and genetic data alongside clinical and
125 physiological data[29]. This extensive phenotyping of therapeutically important variables
126 and biomarkers within the same individual at multiple time points provides a unique
127 opportunity to identify *in vivo* eQTL interactions.
128
129 Here, we examined the modulation of eQTL effects by environmental factors that alter
130 cytokine levels using data from a phase II clinical trial to evaluate the safety and efficacy
131 of a neutralizing IL-6 monoclonal antibody (PF-04236921) in 157 SLE patients[8]
132 (**Methods**). Many patients with SLE exhibit high levels of genes induced by type I IFN;
133 these genes, known as the IFN signature, are a marker of disease severity[30, 31] and
134 a pathogenic feature of SLE. This feature of the disease, together with exposure to anti-

135 IL-6 leads to cytokine fluctuations in this cohort yielding opportunities to assess the
136 impact of cytokine levels on eQTL effects. While this drug was not significantly different
137 from placebo for the primary efficacy endpoint (proportion of patients achieving the SLE
138 Responder Index (SRI-4) at week 24), biologically it effectively reduced free IL-6 protein
139 levels (**Figure S2**). Given the key role of IL-6 and IFN in a range of diseases, the
140 downstream regulatory effects of these cytokines are of great interest to study.
141
142 In this study, we leverage the power of repeat transcriptional and environmental
143 measurements from a lupus clinical trial to identify *in vivo* eQTL interactions with IFN
144 status and anti-IL-6 exposure. In the process, we define novel eQTL interactions for
145 both IFN and IL-6.
146

147 **Results**

148 We conducted whole blood high-depth RNA-seq profiling at 0, 12, and 24 weeks in anti-
149 IL-6 exposed and unexposed individuals with the Illumina TruSeq protocol. We
150 quantified 20,253 gene features and examined 1,595,793 genotyped and imputed
151 common variants genome-wide (**Methods**). Along with each RNA-seq assay, we
152 documented anti-IL-6 exposure and quantified IFN signature status with real-time PCR.
153

154 *Mapping eQTL in SLE patients*

155 We first mapped *cis* eQTLs and then tested them for interactions with IFN status and
156 anti-IL-6 exposure. eQTL interactions can be explored using our interactive visualisation
157 tool (http://baohongz.github.io/Lupus_eQTL, **Figure S3**).
158

To identify *cis* eQTLs, we examined the association between gene expression and SNPs within 250kb upstream of the transcription start site and 250kb downstream of the transcription end site. In order to account for repeat measurements, with up to three RNA-seq assays per patient (**Figure 1A,** 379 samples from 157 patients, **Methods**), we used a linear mixed model. We included 25 gene expression principal components to maximise the number of eQTL detected and 5 genotyping principal components to account for the heterogeneity in ethnicity in our cohort (**Methods**). We observed that the multi-ethnic nature of our study did not confound our results, consistent with Stranger et al.[32] (**Figure S4**).

To ensure we only tested for interactions in a set of highly confident eQTLs, we applied a stringent correction for the total number of hypotheses tested. We recognized that this approach might arguably be overly stringent for eQTL discovery, but we wanted to be certain that we were only testing eQTLs for interactions that had a convincing main effect. Since we tested a total of 5,872,001 SNP-gene pairs genomewide, we set a significance threshold of $p_{eqtl}<8.5\times10^{-9}$ (0.05/5,872,001 tests). We identified 4,818 *cis* eQTL genes (**Figure 1B,1C, Table S1**). The summary statistics for all the gene SNP pairs tested are available in **Table S2**.

To confirm the validity of our eQTLs, we compared them to a larger dataset. In the BIOS cohort, consisting of 2,166 healthy individuals[11], we observed that 85.4% of our SLE eQTL SNP-gene pairs are reported as eQTLs (FDR<0.05). Of these, 98.9% showed

181 consistent direction of effect ($p<5x10^{-16}$, binomial test, **Figure 1D**), suggesting that our
182 results were highly concordant with those in this substantially larger study.
183
184 *Repeat measurements increase power to detect eQTL*
185 Under reasonable assumptions, we would expect repeat samples to increase our
186 power. Supporting that expectation, we detected 64% more *cis* eQTLs compared to the
187 2,934 genes from using a single sample (first available time point) per individual (**Figure**
188 **1B**). An alternative might have been to identify eQTLs separately from each of the three
189 time points; however, this approach identified only a total of 3,050 eQTL genes (**Figure**
190 **S5**). Modelling all three time points together results in 58% more *cis* eQTLs than
191 modelling each time point separately.
192
193 We speculated that while repeat measures did increase power over single measures,
194 that given a fixed number of samples, independent samples would lead to more power.
195 To this end, we conducted an analysis fixing the number of samples at 157 and using
196 53 individuals with repeat measures (with two missing samples). Unsurprisingly, we
197 found fewer eQTLs (2,215 genes) with the repeat measures alone compared to an
198 analysis with the same number of independent samples (2,934 genes).
199
200 *IFN status eQTL interactions*
201 For each of the 4,818 *cis* eQTL genes, we tested the most significantly associated SNP
202 for environmental interactions with our linear mixed model framework. We first explored
203 the influence of type I IFN on gene regulation after determining the IFN status of every

204 patient at each time point. We classified each sample as either IFN high or IFN low
205 using real-time PCR of 11 IFN-inducible genes[33] (**Methods, Figure 2A**).
206
207 We first wanted to assess whether our results were indeed enriched for interactions. To
208 do this, we identified those eQTLs with nominally significant interaction effects at
209 $p_{interact}$<0.01. We would expect ~48 out of 4,818 from chance alone. Surprisingly we
210 observed 182 IFN-eQTL interactions (**Table S1**) that were nominally significant at
211 $p_{interact}$<0.01 suggesting that there was evidence of enrichment for eQTL interactions.
212 We conducted permutations to ensure that these results were not the consequence of
213 potentially inflated statistics, which might be the result for example of low frequency
214 alleles, genes violating normality assumptions, or other technical artefacts. In each of
215 1,000 stringent permutations, we simply reassigned IFN status across samples and
216 retested for eQTL interactions. This permutation preserves the main eQTL effect, since
217 it maintains genotypes of the individuals with the associated expression data, but
218 disrupts any real interactions that might be present in the data. In 0 out of 1,000
219 instance did we observe 182 or more interactions at $p_{interact}$<0.01 suggesting that the
220 number of observed interactions is enriched and highly unlikely to have happened by
221 chance (**Figure S6**, $p_{permute}$~0/1,000 = <0.001).
222
223 We then went on to identify those specific IFN-eQTL interactions of greatest interest by
224 calculating a false discovery rate or q value for each interaction using the q value
225 package[34] (**Methods**). We observed a total of 210 interactions with an FDR<0.2
226 threshold (**Table S1**). We note that 11 of these genes have already been described as

227 having an interaction with a proxy gene for type I IFN signalling in the much larger BIOS
228 study[11]. For example, *SLFN5* expression is influenced by the rs12602407 SNP
229 ($p_{interact}$ =1.3x10^{-10}, FDR <9.9x10^{-8}, **Figure 2B**) and this effect is magnified in IFN high
230 samples. Of these 210 IFN-eQTL interactions, 99 were not reported in the BIOS
231 study[11]. Indeed, applying a more stringent cut off of FDR<0.01, 27/34 of our
232 interactions are not previously reported and therefore are almost certainly novel IFN-
233 eQTL interactions with high confidence (**Figure S7**).
234
235 We speculated that groups of eQTL interactions might be driven by the same common
236 regulatory factor. We divided interactions into magnifiers, where the environmental
237 exposure increases the size of the eQTL effect, and dampeners where the
238 environmental exposure decreases the eQTL effect (**Figure S8**). We hypothesized that
239 the transcription factors driving the response to type I IFN may be different for the eQTL
240 interactions defined as magnifiers (n=127, FDR<0.2) and dampeners (n=83, FDR<0.2).
241
242 We applied HOMER[35] to assess overlap between transcription factor binding motifs
243 and the eQTL interaction SNPs (and SNPs in high linkage disequilibrium (LD, r^2>0.8) in
244 the *cis* window, **Methods**). We conducted two separate analyses: magnifying eQTL
245 interactions in the foreground with dampening interactions in the background and vice
246 versa. We found enrichment of motifs for key transcription factors involved in IFN
247 signalling including a statistically significant enrichment for the ISRE motif (HOMER
248 p=1x10^{-4}, **Table S3**). The ISRE motif disruption occurred for 11 genes with an eQTL
249 magnified in IFN high samples but for only 1 gene with an eQTL dampened

250 (permutation p<0.019, **Methods, Figure 2C**). An example is the *GTF2A2* rs2306355
251 eQTL ($p_{interact}$=8.7x10^{-3}, FDR<0.15, **Figure 2D**); rs2306355 is in tight LD (r^2=0.83 in
252 Europeans) with rs6494127, which interrupts the TTCNNTTT core of the ISRE motif
253 (**Figure 3C**). This SNP likely disrupts IRF9 and STAT2 binding in the ISGF3
254 complex[36], which binds to the ISRE motif. We observe greater expression of *GTF2A2*
255 in individuals with the rs2306355 A allele compared to G; this difference is magnified in
256 IFN high individuals (**Figure 2D**).
257
258 We considered that the principal components included as covariates in our model might
259 be mitigating power. For example, the 4th principal component of gene expression is
260 correlated with the IFN signature status of the sample (r_s= -0.7, **Table S4**) so we
261 repeated the IFN interaction analysis without correcting for principal component 4. For
262 all the eQTLs tested for an IFN interaction, we observed very similar results with highly
263 correlated z-scores (r_s=0.94, **Figure S9**).
264
265 *Discovery of eQTL interactions with anti-IL-6 drug exposure*
266 We then examined whether IL-6 blockade alters the relationship between genomic
267 variation and gene expression and induces drug-eQTL interactions. We wanted to first
268 test if there was evidence of such interactions in our data set. Again, using a threshold
269 of $p_{interact}$<0.01 for nominal significance for interactions, we observed 121 drug-eQTL
270 interactions with anti-IL-6 out of 4,818 eQTLs tested (**Table S1**); similar to IFN
271 interactions, this is far in excess of the ~48 we would expect by chance. As above, to
272 ensure that these results were not the consequence of statistical artefact, we applied
273 the same stringent permutation strategy, reassigning which samples were exposed or

274 not to anti-IL-6. After 1,000 permutations, we never observed as many as 121 drug-
275 eQTL interactions with $p_{interact}$<0.01 (**Figure S10**), suggesting that our eQTLs were
276 indeed highly enriched for those interacting with anti-IL-6 ($p_{permute}$~0/1000<0.001).
277
278 To identify specific eQTL events that interact with anti-IL-6, we again calculated a false
279 discovery rate. We observed that 72 of these interactions have an FDR<0.2. Only 8 of
280 these drug-eQTL interactions overlap with the interactions observed for IFN status
281 (**Table S1**). We note biologically relevant drug-eQTL interactions for *IL10*
282 ($p_{interact}$=2.6x10^{-3}, FDR<0.19, **Figure S11**), an anti-inflammatory cytokine, *CLEC4C*
283 ($p_{interact}$=2.9x10^{-3}, FDR<0.19) which has previously been associated in *trans* with an SLE
284 risk allele[37] and *CLEC18A* ($p_{interact}$=5.1x10^{-4}, FDR<0.14, **Figure 3A**) another member
285 of the C-type lectin domain family.
286
287 Similar to the IFN-eQTL interactions, we divided the drug-eQTL interactions into
288 magnifiers (n=33, FDR<0.2) and dampeners (n=39, FDR<0.2) (**Figure S12**) and used
289 the approach as described above to define transcription factors potentially driving the
290 response to IL-6 blockade (**Table S5**). One of the motifs enriched for eQTLs magnified
291 after drug treatment (with dampeners in the background) was IRF4 (HOMER p=1x10^{-3}).
292 The IRF4 motif disruption occurred for 9 genes, including *CLEC18A*, with an eQTL
293 magnified after drug treatment compared to 4 genes with an eQTL dampened (**Figure**
294 **3B**, **Methods**). We permuted the magnifying and dampening genes and found this ratio
295 for enrichment is suggestive at the gene level (p=0.058) but additional eQTL
296 interactions will be necessary to confirm.
297

298 *Comparing differential expression to eQTL interactions*

299 A more common strategy to determine the effect of an environmental variable is to use
300 differential gene expression. For differential expression following anti-IL-6 treatment, we
301 identified 415 genes with FDR<0.05 but modest effects (max fold change=1.3, **Figure**
302 **S13**). Intriguingly, only 1/72 drug-eQTL interaction genes also show evidence of
303 differential gene expression. This suggests that eQTL interactions offer independent
304 information from differential expression, which might contribute to defining mechanisms.

305

306 *Concordance of drug-eQTL interactions with protein level interactions*

307 We hypothesized that interactions due to drug exposure are likely driven by free IL-6
308 cytokine levels (our key clinical biomarker of interest). If this is the case, for eQTLs
309 dampened by drug exposure, an increase in free IL-6 should elicit an opposite
310 interaction effect and result in eQTL magnification. We assessed whether eQTL
311 interactions with free IL-6 protein levels measured in the patient serum samples were
312 consistent with those following IL-6 blockade. We observed enrichment in the overlap
313 between cytokine interactions and drug interactions (53/72 interactions in expected
314 opposite direction, **Figure 3C,** $p=3.8x10^{-5}$, binomial test).

315

316 **Discussion**

317 In this study we mapped eQTLs in a clinical trial of SLE patients and discovered
318 interactions with IFN and IL-6, two clinically important cytokines. Our study had
319 dramatic variation in IL-6 that was therapeutically induced, and variation in IFN due to
320 the disease status of the SLE patients. This, together with the structured study design

321 with repeat measurements of gene expression across different conditions in the same
322 individual, allowed us to identify *in vivo* eQTL interactions.
323
324 eQTL interactions with drug interventions or other therapeutically relevant physiologic
325 variables are important to identify. We note that these results are independent of
326 differentially expressed genes, which are more likely to represent second order effects,
327 rather than primary genetic effects. Therefore, these eQTL interactions can point to
328 regulatory mechanisms, such as transcription factors or subclasses of enhancers,
329 acting downstream of the environmental condition of interest and driving groups of
330 eQTL interactions. The IFN status eQTL interactions we identified provide support for
331 this approach. By making use of the direction of effect for the eQTL interaction, we were
332 able to identify an enrichment of magnifying eQTL interaction SNPs interrupting the
333 binding sites of transcription factors known to be important in the response to IFN, such
334 as ISGF3 (the STAT1, STAT2 and IRF9 complex), which binds ISRE. Once we are able
335 to recognize the downstream drivers of therapeutically relevant clinical variables, then it
336 may become possible to define more mechanisms of action for drugs and more precise
337 drug targets.
338
339 As a powerful example, we note enrichment of magnifying anti-IL-6-eQTL interaction
340 SNPs interrupting the binding site of IRF4. It has been suggested that IRF4 works
341 downstream of IL-6 by binding BATF and co-ordinately regulating the production of IL10
342 and other genes[38]. Consistent with this, we observed that the *IL10* eQTL does indeed
343 interact with presence of anti-IL-6 (**Figure S11**). Previous studies have highlighted a

344 role for IRF4 in the pathogenesis of autoimmune diseases in mouse and humans. For
345 example in a murine model of SLE, *IRF4* knockout mice did not develop lupus
346 nephritis[39]. In humans, IRF4 is associated with RA[40], a disease in which anti-IL-6
347 treatment has been successful[3]. Our findings provide further support that IRF4 could
348 be a potential therapeutic target for autoimmune diseases such as RA where anti-IL-6 is
349 effective[41].
350
351 The ability to focus on interactions with specific patient phenotypes might point to key
352 targets for disease intervention. For example, IFN is a key immunophenotype in SLE
353 patients, and elevated in SLE compared to healthy controls[30, 31]. The IFN status
354 immunophenotype is already itself driving interest in therapeutic targets. A recent phase
355 II clinical trial has shown that an antagonist to the type I IFN receptor, acting upstream
356 of ISRE, reduced severity of symptoms in SLE. Interestingly, the antagonist was more
357 effective in the patients with a high baseline IFN status[42]. This example provides a
358 compelling case study for how understanding master regulators of key disease
359 phenotypes might lead to promising new therapeutic strategies. We speculate that this
360 provides a mechanism for stratified medicine for future studies, which may be applicable
361 to other diseases.
362
363 We recognized that computing eQTL interactions requires a robust statistical model that
364 accounts for genotype, environmental factor, RNA expression levels, repeat
365 measurements, and technical covariates. We were sensitive to the possibility that pre-
366 processing and normalisation of these factors could potentially have an impact on our

367 results. For this reason, we used stringent filtering and examined only variants that were
368 common and where the minor allele was present for each of the exposure groups. Next,
369 to confirm enrichment of eQTL interactions, we used a stringent permutation-based
370 strategy that preserved the distribution of genotypes and corresponding expression
371 values. Finally, we also utilized a standard normal transformation[43] (**Methods**) and
372 observed that this had little effect on the primary eQTL analysis (r_s=0.99 for z scores,
373 **Figure S14**) and interaction analyses (IFN r_s=0.84, drug r_s=0.76 for z scores, **Figure**
374 **S15**), or the observed enrichment over the null in our stringent permutation analysis
375 (**Figure S16**).
376
377 We speculate that drug-eQTL interactions might offer an alternative pharmacogenetic
378 strategy to assess drug response. For many biologic medications, predictive
379 pharmacogenetics through typical association studies has been challenging; for
380 example, studies trying to define genetic or transcriptomic biomarkers of anti-TNF
381 response have not been successful[44, 45]. An eQTL interaction approach can be used
382 to define a genotype-aware score reflecting the biological activity that a medication is
383 having upon an individual, given their allelic combination of multiple genetic markers.
384 For example, we can define a simple anti-IL-6 exposure score based on 7 anti-IL-6
385 eQTL interactions with a more stringent FDR (FDR<0.1). This score is based on
386 assessing whether the expression of the eQTL target gene was more consistent with
387 the drug exposed or the unexposed state for the corresponding interaction SNP
388 genotype. Unsurprisingly, we found a difference in drug exposure score between the
389 unexposed and exposed samples (**Figure S17**) (r_s=0.40, p=2.1×10^{-16}); these differences

390 reflect the fact that the eQTLs were themselves identified by examining samples with
391 and without drug exposure. However, while we did not utilize the administered drug
392 dose to identify drug-eQTL interactions; we observed a significant correlation between
393 drug dose (10, 50 or 200mg) and drug exposure score (r_s=0.16, p=0.02) in the drug-
394 exposed samples (**Figure S18**). A simple eQTL interaction score may therefore have
395 the potential to stratify individuals when assessing response to a medication for
396 example those with a higher drug exposure score may have a better response to
397 treatment. Similarly, this score could be correlated with adverse effects to capture
398 informative gene expression signatures.
399
400 We note a limitation of this study is that the drug itself did not achieve its primary
401 efficacy endpoint of improving SLE outcomes. Hence, while the drug exposure score for
402 this study tracked with the biological effect of the drug (reducing free IL-6 protein levels),
403 it might not be useful for SLE specifically. However, such a scoring system could be
404 implemented easily in most phase III trials for a broad range of therapeutics, where the
405 numbers of samples are far in excess of this phase II trial, ensuring better powered and
406 more accurate eQTL-interaction mapping.
407
408 **Conclusion**
409 We devised a framework for identifying *in vivo* eQTL interactions with therapeutically
410 relevant variables, exploiting repeat measurements from a clinical trial. We have applied
411 this approach to demonstrate how downstream regulatory effects of cytokine biology
412 can be elucidated. This same approach can be applied to a wide range of other

413 clinically important cytokines, their antagonists, or indeed other targeted biologic
414 therapies. We speculate that this approach might even be applied to the presence or
415 absence of disease, or disease activity. However, given the multifaceted nature of
416 disease effects, interpreting an eQTL interaction in that context might be more
417 challenging. Modern clinical cohorts and clinical trial data sets with RNA-seq data that
418 has been collected will make this approach easily applicable on a wide scale.
419
420

421 **Methods**

422 *Study design*

423 The objectives of this study were to map eQTLs in a cohort of lupus patients and
424 identify eQTL interactions with environmental perturbations such as drug treatment to
425 shed light on drug and disease mechanisms. SLE patients were recruited to a phase II
426 clinical trial to test the efficacy and safety of an IL-6 monoclonal antibody (PF-
427 04236921). The patient population recruited to this trial have been detailed extensively
428 by Wallace et al.[8] 183 patients (forming a multi-ethnic cohort) were randomized to
429 receive three doses of drug (10, 50 or 200mg) or placebo at three time points during the
430 trial (weeks 0, 8 and 16).
431

432 *RNA-sequencing*

433 We collected peripheral venous blood samples in PAXgene Blood RNA tubes
434 (PreAnalytiX GmbH, BD Biosciences) for high-depth RNA-seq profiling at 0, 12, and 24
435 weeks. We extracted total RNA from blood samples using the PAXgene Blood RNA kit

436 (Qiagen) at a contract lab using a customized automation method. We assessed the
437 yield and quality of the isolated RNA using Quant-iT™ RiboGreen® RNA Assay Kit
438 (ThermoFisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies),
439 respectively. Following quality assessment, we processed an aliquot of 500-1000 ng of
440 each RNA with a GlobinClear-Human kit, (ThermoFisher Scientific) to remove globin
441 mRNA. We then converted RNA samples to cDNA libraries using TruSeq RNA Sample
442 Prep Kit v2 (Illumina) and sequenced using Illumina HiSeq 2000 sequencers. We
443 generated an average of 40M 100bp pair-end reads per sample for downstream
444 analysis.
445
446 We successfully obtained 468 RNA-seq profiles from 180 patients. We aligned reads to
447 the reference genome and quantified gene expression using Subread[46] and
448 featureCounts[47] respectively. We included genes with at least 10 reads (CPM>0.38)
449 in at least 32 samples (minimum number of patients with both unexposed and exposed
450 RNA-seq assays in a drug group) prior to normalization. Following quality control (QC),
451 we removed 4 samples as outliers. We then normalized 20,253 transcripts using the
452 trimmed mean of M-values method and the edgeR R package[48]. Expression levels
453 are presented as $\log_2$(cpm +1) (**Table S6**).
454
455 *Genotyping*
456 We genotyped 160 individuals across 964,193 variants genome-wide with the Illumina
457 HumanOmniExpressExome-8v1.2 beadchip. We removed SNPs if they deviated from
458 Hardy-Weinberg Equilibrium (HWE) ($p < 1x10^{-7}$), had a minor allele frequency <5%,

459 missingness >2% or a heterozygosity rate greater than 3 standard deviations from the
460 mean (PLINK[49, 50]). For mapping eQTLs, we removed SNPs on the Y chromosome.
461 Following QC, we used 608,017 variants for further analysis. We removed one sample
462 with high missingness and outlying heterozygosity rate from further analysis.
463
464 *Imputation*
465 We pre-phased the genotypes with SHAPEIT v2[51]. We imputed missing genotypes
466 and untyped SNPs using Impute2[52] in 5Mb chunks against the 1000 Genomes Phase
467 3[53] reference panel. To ensure only highly quality genotypes, and to avoid artefacts
468 that can be induced by imputation uncertainty, we removed SNPs with an info score <1,
469 MAF<0.05 or HWE $p<1x10^{-7}$ leaving 1,595,793 SNPs for further analysis.
470
471 *Interferon status*
472 We classified the interferon (IFN) status of each sample at each time point from the
473 expression of 11 IFN response genes (*HERC5*, *IFI27*, *IRF7*, *ISG15*, *LY6E*, *MX1*, *OAS2*,
474 *OAS3*, *RSAD2*, *USP18*, *GBP5*) using TaqMan Low Density Arrays. These 11 genes
475 were selected by identifying transcripts for which there was both a measureable
476 response to IFN treatment *in vitro*, as well as differential expression (reduction in
477 expression level) between baseline and visits with clinical improvement in the BOLD
478 study[33]. There is no consensus set of genes to determine the IFN status of SLE
479 patients but these 11 genes do overlap with other published gene sets. For example
480 4/11 genes are also used in the 7-gene set defined by McBride et al[54] and 9/11 genes
481 overlap with the 21-gene set defined by Yao et al[55].

482

483 The first principal component of the expression of the 11-gene set captured 91.7% of
484 the variation (**Figure S19**). The distribution of this first principal component is nearly
485 bimodal with good separation (**Figure 2A**) and we classified samples as high or low IFN
486 based on this first principal component score. In our dataset, we see excellent
487 correlations (r_s=0.86-0.98) between the real-time PCR expression and the RNA-seq
488 expression for these 11 genes (**Figure S20**). The first PC of the IFN signature of RNA-
489 seq data is also strongly correlated with the first PC of the IFN signature of real-time
490 PCR (r_s=0.96, **Figure S21**). IFN status was available for 376 samples from 157
491 subjects.

492

493 *Drug exposure*

494 Samples were assigned as unexposed (placebo or week 0 samples) or drug exposed
495 (week 12 and week 24 samples in the drug groups).

496

497 *Free IL-6 protein levels*

498 We determined free IL-6 protein levels from serum using a commercial sandwich ELISA
499 selected for binding only free IL-6. The assay was validated according to FDA
500 biomarker and fit-for purpose guidelines. Free IL-6 protein levels were available for 311
501 samples from 145 subjects. Since the distribution of IL-6 levels was highly skewed, we
502 ranked samples in order of IL-6 protein levels and included in the model to identify drug-
503 eQTL interactions.

504

505 *Statistical analysis*

506 *eQTL and interaction analysis*

507 In total, 157 patients (with 379 RNA-seq samples) had good quality gene expression
508 and genotyping data for eQTL analysis. All statistical analyses were carried out in R[56].
509
510 We defined a *cis* eQTL as the SNP within 250kb upstream of the GENCODE[57]
511 transcription start site of the gene or 250kb downstream of the transcription end site.
512 We first applied a linear model for the first available time point (week 0 sample for
513 n=152, week 12 sample for n=5) to identify each eQTL using the first 25 principal
514 components of gene expression and the first 5 principal components of genotyping as
515 covariates.
516
517 To select the number of gene expression principal components to include, we counted
518 the number of eQTL genes identified after incrementally increasing the number of
519 principal components accounted for in the model from 0 to 50 by increments of five
520 (**Figure S22**). We selected 25 principal components of gene expression to maximise the
521 number of eQTL genes detected while minimising the number of principal components
522 we corrected for. We included 5 principal components of genotyping to account for the
523 heterogeneity in ethnicity in our cohort (**Figure S23**).
524
525 SNPs were encoded as 0, 1 and 2 with respect to the number of copies of the minor
526 allele. To adjust for multiple testing during eQTL discovery we used a stringent
527 Bonferroni corrected *p*-value threshold of 8.5×10^{-9} (0.05/ 5,872,001 tests). The

528 Bonferroni adjustment assumes independence among the tests and we therefore note
529 that it is a conservative multiple comparisons adjustment.
530
531 To map eQTLs using multiple samples for each individual, we applied a random
532 intercept linear mixed model using the first 25 principal components of gene expression
533 and the first 5 principal components of genotyping as covariates and patient as a
534 random effect:
535

536 $$E_{i,j} = \theta + \beta_{geno} \cdot g_j + (\kappa_j | i) + \sum_{l=1}^{25} \phi_l \cdot pc_{i,l} + \sum_{m=1}^{5} \gamma_m \cdot pc_{j,m}$$

537
538 Where $E_{i,j}$ is gene expression for the ith sample from the jth subject, θ is the intercept,
539 β_{geno} is the genotype effect (eQTL), $(\kappa_j|i)$ is the random effect for the ith sample from the
540 j^{th} subject, $pc_{i,l}$ is principal component l of gene expression for sample i, $pc_{j,m}$ is principal
541 component m of genotyping for subject j.
542
543 We fitted the linear mixed models using the lme4 R package[58]. We assumed
544 covariance between samples from the same individual, but did not assume any
545 structure in this covariance.
546
547 We used the most significant SNP (with $p<8.5\times10^{-9}$) from the 4,818 identified eQTL
548 genes to explore eQTL interactions. For each environmental interaction analysis, we
549 further filtered these eQTLs to include only those with at least two individuals
550 homozygous for the minor allele of the SNP being tested in each of the environmental

551 factor groups. For example, we required two of these individuals in each of the drug
552 exposed and drug unexposed groups. To identify eQTL interactions, we added an
553 additional covariate to the model for example drug exposure, and an interaction term
554 between this covariate and the genotype of the SNP:
555

556
$$E_{i,j} = \theta + \beta_{geno} \cdot g_j + (\kappa_j | i) + \sum_{l=1}^{25} \phi_l \cdot pc_{i,l} + \sum_{m=1}^{5} \gamma_m \cdot pc_{j,m} + \beta_{drug} \cdot d_i + \beta_x \cdot d_i \cdot g_j$$

557
558 Where $E_{i,j}$ is gene expression for the ith sample from the jth subject, θ is the intercept,
559 β_{geno} is the genotype effect (eQTL), $(\kappa_j|i)$ is the random effect for the ith sample from the
560 j^{th} subject, $pc_{i,l}$ is principal component l of gene expression for sample i, $pc_{j,m}$ is principal
561 component m of genotyping for subject j, β_{drug} is the drug effect (differential gene
562 expression) and β_x is the interaction effect.
563
564 We determined the significance of the interaction term with a likelihood ratio test.
565
566 To rigorously confirm the relative enrichment of eQTL interactions, we shuffled the
567 interaction covariate (for example drug exposure) 1,000 times and calculated the
568 number of significant interactions observed in each permutation. Our primary goal for
569 the permutation analysis was to retain the main eQTL effect while examining only the
570 effect of the environmental factor on the interaction. In this study, the main purpose of
571 the covariates included in the model is to ensure the main eQTL effect is found. For IFN
572 high/low status, we shuffled across all samples. For drug interaction permutation

573 analysis, we maintained the number of individuals in the drug group and the number of
574 samples with exposure to drug. We calculated a qvalue for each interaction using the q
575 value package[34]. **Figure S24** shows the observed versus the expected p values for
576 the interaction analyses.
577
578 The expression of the majority of genes followed a normal distribution (**Figure S25**) but
579 to assess whether non-normality could be causing an inflation of our test statistic, we
580 repeated the identification of eQTLs and eQTL interactions following the standard
581 normal transformation. We transformed the expression values of each gene to their
582 respective quantiles of a normal distribution using the qqnorm function in R, breaking
583 any ties (for example expression levels of zero in some individuals) randomly.
584
585 *Concordance with an eQTL study in healthy individuals*
586 In the SLE cohort, we classified 4,818 *cis* eQTL genes ($p<8.5x10^{-9}$). The *z*-score for the
587 most associated SNP for each of these genes was compared to the *z*- score from a
588 previously published eQTL dataset from whole blood from 2,166 healthy individuals[11].
589 4,113/4818 SNP-gene pairs (85.4%) were also reported in the BIOS dataset
590 (FDR<0.05). After removing 301 SNPs, which could not be mapped to a strand
591 3,770/3,812 (98.9%) had a *z*-score (eQTL effect) in a consistent direction.
592
593 *Magnifiers and Dampeners*
594 An eQTL interaction can either magnify or dampen the original eQTL effect. We
595 multiplied the interaction *z*-score by the sign of the original eQTL effect (genotype beta)

596 and defined magnifiers as interactions with an adjusted *z*-score > 0 and dampeners as
597 interactions with an adjusted *z*-score < 0.
598
599 *Differential gene expression analysis*
600 To identify differentially expressed genes following drug exposure (unexposed or
601 exposed), we applied a random intercept linear mixed model using the first 25 principal
602 components of gene expression and the first 5 principal components of genotyping as
603 covariates and patient as a random effect. We calculated a q value using the q value
604 package[34].
605
606 *Drug exposure score*
607 We assigned a drug exposure score to each sample. We calculated a score for each
608 gene (see equation below) and then averaged across the 7 drug-eQTL genes
609 (FDR<0.1) to give the final drug exposure score.
610

$$Drug\ exposure\ score\ for\ gene\ = \frac{1}{2}(\frac{G\ -\ G_{Unexp}}{SE})^2 - \frac{1}{2}(\frac{G\ -G_{Exp}}{SE})^2$$

611
612 Where G is gene expression for a given sample, G_{Unexp} is predicted mean gene
613 expression for unexposed samples of the relevant SNP genotype, G_{Exp} is predicted
614 mean gene expression for exposed samples of the relevant SNP genotype and SE is
615 standard error for the intercept term of the model (unexposed expression for genotype
616 0).
617

618 *HOMER analysis for transcription factor binding motif enrichment*

619 We used the HOMER software suite[35] to look for enrichment of transcription factor
620 binding motifs in the 210 IFN-eQTL interactions (FDR<0.2) and the 72 drug-eQTL
621 interactions (FDR<0.2). Each eQTL interaction was identified using the most highly
622 associated SNP for that eQTL. However, as this SNP is not necessarily the functional
623 SNP, we additionally considered all those with an $r^2 \geq 0.8$ in the 1000 Genomes
624 European population[53] within the *cis* eQTL window. We defined our motif search
625 window as 20 bp on either side of each SNP (i.e. 41 bp wide).
626
627 For each environmental factor, we divided the eQTL interactions into magnifiers or
628 dampeners and conducted two separate HOMER analyses: one with magnifiers in the
629 foreground and dampeners in the background; the other with dampeners in the
630 foreground and magnifiers in the background. HOMER reported the transcription factor
631 motifs that were significantly enriched in the foreground relative to background. Motifs
632 were plotted using the SeqLogo R library[59].
633
634 We determined permutation p values for enrichment of the ISRE and IRF4 transcription
635 factor binding sites as follows. For ISRE, the motif is interrupted by interaction SNPs (or
636 SNPs in LD) corresponding to 11 magnifying genes and 1 dampening genes. We
637 permuted which genes were labelled as magnifiers or dampeners 100,000 times and
638 counted the number of genes in each category with an ISRE motif interrupted. We
639 found 1,855 occurrences from 100,000 trials with at least 11 magnifying genes
640 ($p<0.019$). For IRF4 the motif is interrupted by SNPs corresponding to 9 magnifying

641 genes and 4 dampening genes. Using the same permutation approach, we found 5,801
642 occurrences from 100,000 trials with at least 9 magnifying genes ($p<0.058$).
643

644 **Declarations**

645 **Ethics approval and consent to participate**

646 The protocol was approved by each institutional review board subject to applicable laws
647 and regulations and ethical principles consistent with the Declaration of Helsinki.
648

649 **Availability of data and material**

650 Data are available in the supplementary tables and at
651 http://baohongz.github.io/Lupus_eQTL.
652

653 **Acknowledgements**

654 This work is supported in part by funding from the National Institutes of Health
655 (U01GM092691, UH2AR067677, U19AI111224 (SR)), the Doris Duke Charitable
656 Foundation Grant #2013097, the Ruth L. Kirschstein National Research Service Award
657 (F31AR070582) from the National Institute of Arthritis and Musculoskeletal and Skin
658 Diseases (KS) and the Rheumatology Research Foundation Tobé and Stephen E.
659 Malawista, MD Endowment in Academic Rheumatology (D.A.R). This work is also
660 supported by unrestricted funding from Pfizer, Inc.
661

662 **Author Contributions**

663 The project was conceived and designed by EED, MSV, BZ and SR. Statistical analysis
664 was conducted by EED, TA, MG-A, KS, H-JW, YL and CS. Molecular data was
665 obtained, organized and analysed by YZ, SP, DvS, JSB, NB, MSV, BZ and DAR. The
666 initial manuscript was written by EED and SR. All authors edited and approved the
667 manuscript.
668

669 **Competing Interests**

670 The authors declare that they have no competing interests.
671

672 **References**

673 1. Rider P, Carmi Y, Cohen I. Biologics for Targeting Inflammatory Cytokines, Clinical
674 Uses, and Limitations. Int J Cell Biol. 2016;2016 iv.
675 2. Cessak G, Kuzawińska O, Burda A, Lis K, Wojnar M, Mirowska-Guzel D, et al. TNF
676 inhibitors - Mechanisms of action, approved and off-label indications. Pharmacol
677 Reports. 2014;66:836–44.
678 3. Tanaka Y, Mola EM. IL-6 targeting compared to TNF targeting in rheumatoid arthritis:
679 studies of olokizumab, sarilumab and sirukumab. Ann Rheum Dis. 2014;73:1595–7.
680 4. Stone JH, Tuckwell K, Dimonaco S, Klearman M, Aringer M, Blockmans D, et al. Trial
681 of Tocilizumab in Giant-Cell Arteritis. N Engl J Med. 2017;377:317–28.
682 doi:10.1056/NEJMoa1613849.
683 5. Van Rhee F, Wong RS, Munshi N, Rossi JF, Ke XY, Fosså A, et al. Siltuximab for
684 multicentric Castleman's disease: A randomised, double-blind, placebo-controlled trial.
685 Lancet Oncol. 2014;15:966–74.
686 6. Tackey E, Lipsky PE, Illei GG. Rationale for interleukin-6 blockade in systemic lupus
687 erythematosus. Lupus. 2004;13:339–43.
688 7. Illei GG, Shirota Y, Yarboro CH, Daruwalla J, Tackey E, Takada K, et al. Tocilizumab
689 in systemic lupus erythematosus: Data on safety, preliminary efficacy, and impact on
690 circulating plasma cells from an open-label phase I dosage-escalation study. Arthritis
691 Rheum. 2010;62:542–52.
692 8. Wallace DJ, Strand V, Merrill JT, Popa S, Spindler AJ, Eimon A, et al. Efficacy and
693 safety of an interleukin 6 monoclonal antibody for the treatment of systemic lupus
694 erythematosus: a phase II dose-ranging randomised controlled trial. Ann Rheum Dis.
695 2017;76:534–42. doi:10.1136/annrheumdis-2016-209668.
696 9. Bronson PG, Chaivorapol C, Ortmann W, Behrens TW, Graham RR. The genetics of
697 type I interferon in systemic lupus erythematosus. Curr Opin Immunol. 2012;24:530–7.
698 doi:10.1016/j.coi.2012.07.008.

699 10. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al.
700 Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med.
701 2017;377:119–1131. doi:10.1056/NEJMoa1707914.
702 11. Zhernakova D V, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W,
703 et al. Identification of context-dependent expression quantitative trait loci in whole blood.
704 Nat Genet. 2017;49 November:139–45. doi:10.1038/ng.3737.
705 12. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate
706 Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene
707 Expression. Science. 2014;343:1246949. doi:10.1126/science.1246949.
708 13. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics
709 of gene expression in primary immune cells identifies cell type-specific master
710 regulators and roles of HLA alleles. Nat Genet. 2012;44:502–10. doi:10.1038/ng.2205.
711 14. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, et al. Polarization of
712 the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. Science.
713 2014;344:519–23.
714 15. Ardlie KG, Deluca DS, Segre A V., Sullivan TJ, Young TR, Gelfand ET, et al. The
715 Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in
716 humans. Science. 2015;348:648–60. doi:10.1126/science.1262110.
717 16. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture
718 of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS
719 Genet. 2011;7:e1002003. doi:10.1371/journal.pgen.1002003.
720 17. Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z, Knowles DA, et al. Impact
721 of the X chromosome and sex on regulatory variation. Genome Res. 2016;26:768–77.
722 18. Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng H-F, et al. Gene-
723 gene and gene-environment interactions detected by transcriptome sequence analysis
724 in twins. Nat Genet. 2015;47:88–91. doi:10.1038/ng.3162.
725 19. Maranville JC, Luca F, Stephens M, Di Rienzo A. Mapping gene-environment
726 interactions at regulatory polymorphisms: insights into mechanisms of phenotypic
727 variation. Transcription. 2012;3:56–62. doi:10.4161/trns.19497.
728 20. Idaghdour Y, Awadalla P. Exploiting gene expression variation to capture gene-
729 environment interactions for disease. Front Genet. 2013;4 MAY:1–7.
730 21. Idaghdour Y, Quinlan J, Goulet J-P, Berghout J, Gbeha E, Bruat V, et al. Evidence
731 for additive and interaction effects of host genotype and infection in malaria. Proc Natl
732 Acad Sci U S A. 2012;109:16786–93. doi:10.1073/pnas.1204945109.
733 22. Idaghdour Y, Czika W, Shianna K V, Lee SH, Visscher PM, Martin HC, et al.
734 Geographical genomics of human leukocyte gene expression variation in southern
735 Morocco. Nat Genet. 2010;42:62–7. doi:10.1038/ng.495.
736 23. Peters JE, Lyons PA, Lee JC, Richard AC, Fortune MD, Newcombe PJ, et al.
737 Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell
738 Types in Health and Immune-Mediated Disease. PLoS Genet. 2016;12:e1005908.
739 doi:10.1371/journal.pgen.1005908.
740 24. Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, et al. Mapping
741 determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS
742 Genet. 2006;2:2155–61.
743 25. Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression.
744 PLoS Biol. 2008;6:810–24.

745 26. Maranville JC, Luca F, Richards AL, Wen X, Witonsky DB, Baxter S, et al.
746 Interactions between glucocorticoid treatment and Cis-regulatory polymorphisms
747 contribute to cellular response phenotypes. PLoS Genet. 2011;7:e1002162.
748 27. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering the
749 genetic architecture of variation in the immune response to Mycobacterium tuberculosis
750 infection. Proc Natl Acad Sci U S A. 2012;109:1204–9.
751 28. Smirnov D, Morley M, Shin E. Genetic analysis of radiation-induced changes in
752 human gene expression. Nature. 2009;459:587–91. doi:10.1038/nature07940.Genetic.
753 29. Walsh AM, Whitaker JW, Huang CC, Cherkas Y, Lamberth SL, Brodmerkel C, et al.
754 Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell
755 type associations. Genome Biol. 2016;17:79. doi:10.1186/s13059-016-0948-6.
756 30. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann W a, Espe KJ, et al.
757 Interferon-inducible gene expression signature in peripheral blood cells of patients with
758 severe lupus. Proc Natl Acad Sci U S A. 2003;100:2610–5.
759 31. Bennett L, Palucka a K, Arce E, Cantrell V, Borvak J, Banchereau J, et al.
760 Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp
761 Med. 2003;197:711–23.
762 32. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, et al.
763 Patterns of Cis regulatory variation in diverse human populations. PLoS Genet. 2012;8.
764 33. Hill AA, Immermann FW, Zhang Y, Reddy PS, Zhou T, O'Toole M, et al. FRI0003
765 Determination of interferon (IFN) signatures for sle patients may be critical for optimal
766 treatment selection but depends on the genes chosen: report from the bold (biomarkers
767 of lupus disease) study. Ann Rheum Dis. 2013;72:A369–70. doi:10.1136/annrheumdis-
768 2013-eular.1131.
769 34. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl
770 Acad Sci. 2003;100:9440–5. doi:10.1073/pnas.1530509100.
771 35. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple
772 Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory
773 Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010;38:576–89.
774 doi:10.1016/j.molcel.2010.05.004.
775 36. Au-Yeung N, Mandhana R, Horvath CM. Transcriptional regulation by STAT1 and
776 STAT2 in the interferon JAK-STAT pathway. Jak-Stat. 2013;2:e23931.
777 doi:10.4161/jkst.23931.
778 37. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al.
779 Systematic identification of trans eQTLs as putative drivers of known disease
780 associations. Nat Genet. 2013;45:1238–43. doi:10.1038/ng.2756.
781 38. Koch S, Mousset S, Graser A, Reppert S, Übel C, Reinhardt C, et al. IL-6 activated
782 integrated BATF/IRF4 functions in lymphocytes are T-bet-independent and reversed by
783 subcutaneous immunotherapy. Sci Rep. 2013;3:1754. doi:10.1038/srep01754.
784 39. Lech M, Weidenbusch M, Kulkarni OP, Ryu M, Darisipudi MN, Susanti HE, et al.
785 IRF4 Deficiency Abrogates Lupus Nephritis Despite Enhancing Systemic Cytokine
786 Production. J Am Soc Nephrol. 2011;22:1443–52. doi:10.1681/ASN.2010121260.
787 40. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid
788 arthritis contributes to biology and drug discovery. Nature. 2014;506:376–81.
789 doi:10.1038/nature12873.
790 41. Xu WD, Pan HF, Ye DQ, Xu Y. Targeting IRF4 in autoimmune diseases. Autoimmun

Rev. 2012;11:918–24. doi:10.1016/j.autrev.2012.08.011.
42. Furie R, Merrill J, Werth V, Khamashta M, Kalunian K, Brohawn P, et al. Anifrolumab, an Anti-Interferon Alpha Receptor Monoclonal Antibody, in Moderate to Severe Systemic Lupus Erythematosus (SLE). Arthritis Rheumatol. 2017;69:376–86.
43. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501:506–11. doi:10.1038/nature12531.
44. Cui J, Stahl EA, Saevarsdottir S, Miceli C, Diogo D, Trynka G, et al. Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. PLoS Genet. 2013;9:e1003394. doi:10.1371/journal.pgen.1003394.
45. Cui J, Diogo D, Stahl EA, Canhao H, Mariette X, Greenberg MPH JD, et al. The role of rare protein-coding variants to anti-TNF treatment response in rheumatoid arthritis. Arthritis Rheumatol. 2017;69:735–41.
46. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41:e108–e108. doi:10.1093/nar/gkt214.
47. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
48. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40. doi:10.1093/bioinformatics/btp616.
49. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. doi:10.1186/s13742-015-0047-8.
50. Purcell S, Chang C. PLINK 1.9. https://www.cog-genomics.org/plink2.
51. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011;9:179–81. doi:10.1038/nmeth.1785.
52. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genet. 2009;5:e1000529. doi:10.1371/journal.pgen.1000529.
53. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. doi:10.1038/nature15393.
54. McBride JM, Jiang J, Abbas AR, Morimoto A, Li J, MacIuca R, et al. Safety and pharmacodynamics of rontalizumab in patients with systemic lupus erythematosus: Results of a phase I, placebo-controlled, double-blind, dose-escalation study. Arthritis Rheum. 2012;64:3666–76.
55. Yao Y, Higgs BW, Morehouse C, de Los Reyes M, Trigona W, Brohawn P, et al. Development of Potential Pharmacodynamic and Diagnostic Markers for Anti-IFN-α Monoclonal Antibody Trials in Systemic Lupus Erythematosus. Hum Genomics Proteomics. 2009;2009:374312-. doi:10.4061/2009/374312.
56. R Core Team. R: A Language and Environment for Statistical Computing. 2015. https://www.r-project.org.
57. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project.

837 Genome Res. 2012;22:1760–74. doi:10.1101/gr.135350.111.
838 58. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
839 lme4. J Stat Softw. 2015;67:1–48. doi:10.18637/jss.v067.i01.
840 59. Bembom O. seqLogo: Sequence logos for DNA sequence alignments. 2016.

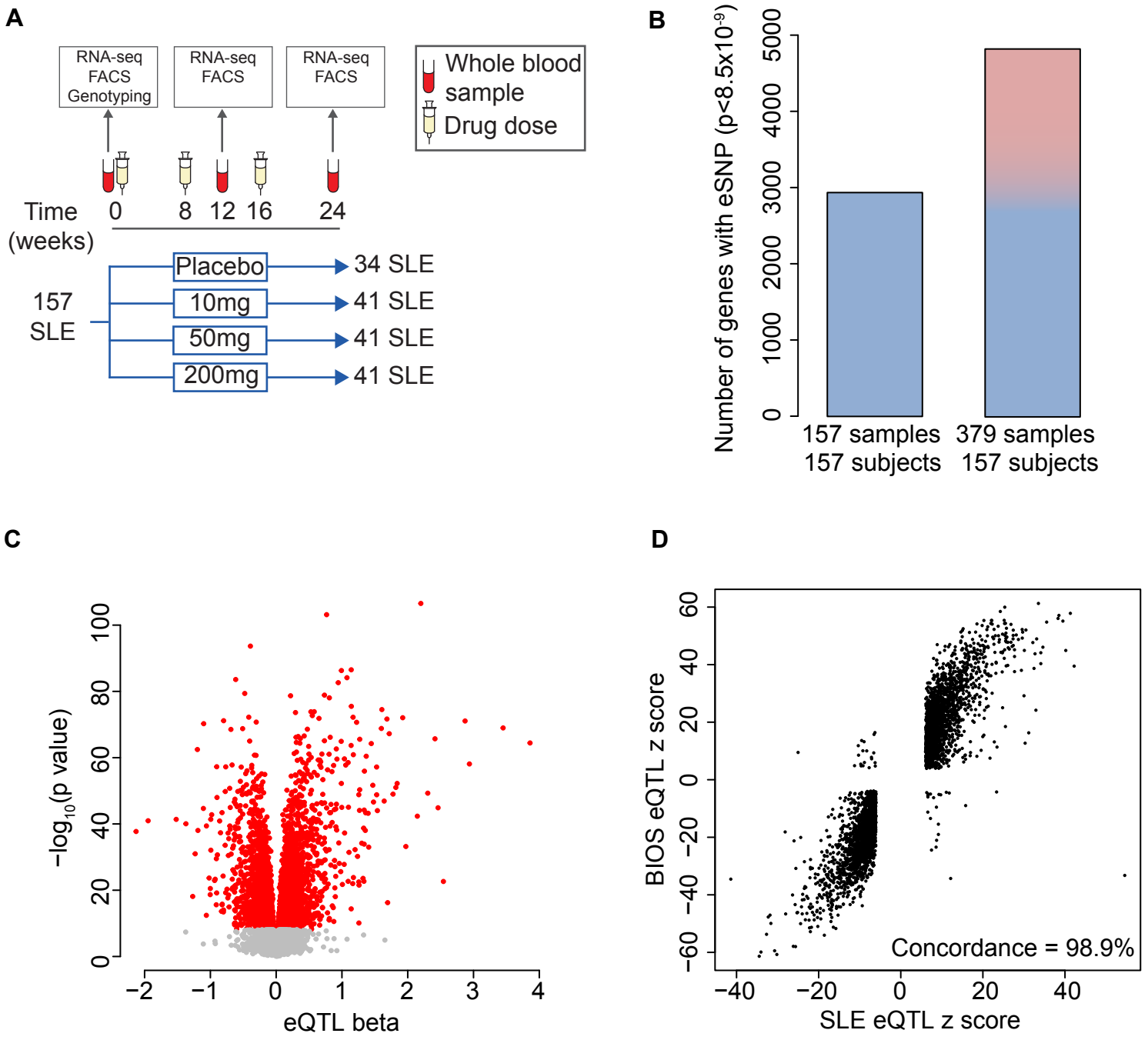

Figure 1. Identifying eQTLs in SLE patients (A) Clinical trial structure and sampling strategy. (B) Number of eQTL genes identified using a linear model (left) and a linear mixed model (right). For the linear model, we used the first available time point for each individual (week 0 sample for n=152, week 12 sample for n=5). (C) Volcano plot of eQTL effects for the most significantly associated SNP for each gene (red color indicates $p<8.5\times10^{-9}$). (D) Concordance of SLE eQTL effects ($p<8.5\times10^{-9}$) with eQTLs observed in the BIOS cohort[11] of healthy individuals (FDR <0.05). Each point represents the most significant SNP-gene pair for the SLE eQTL.

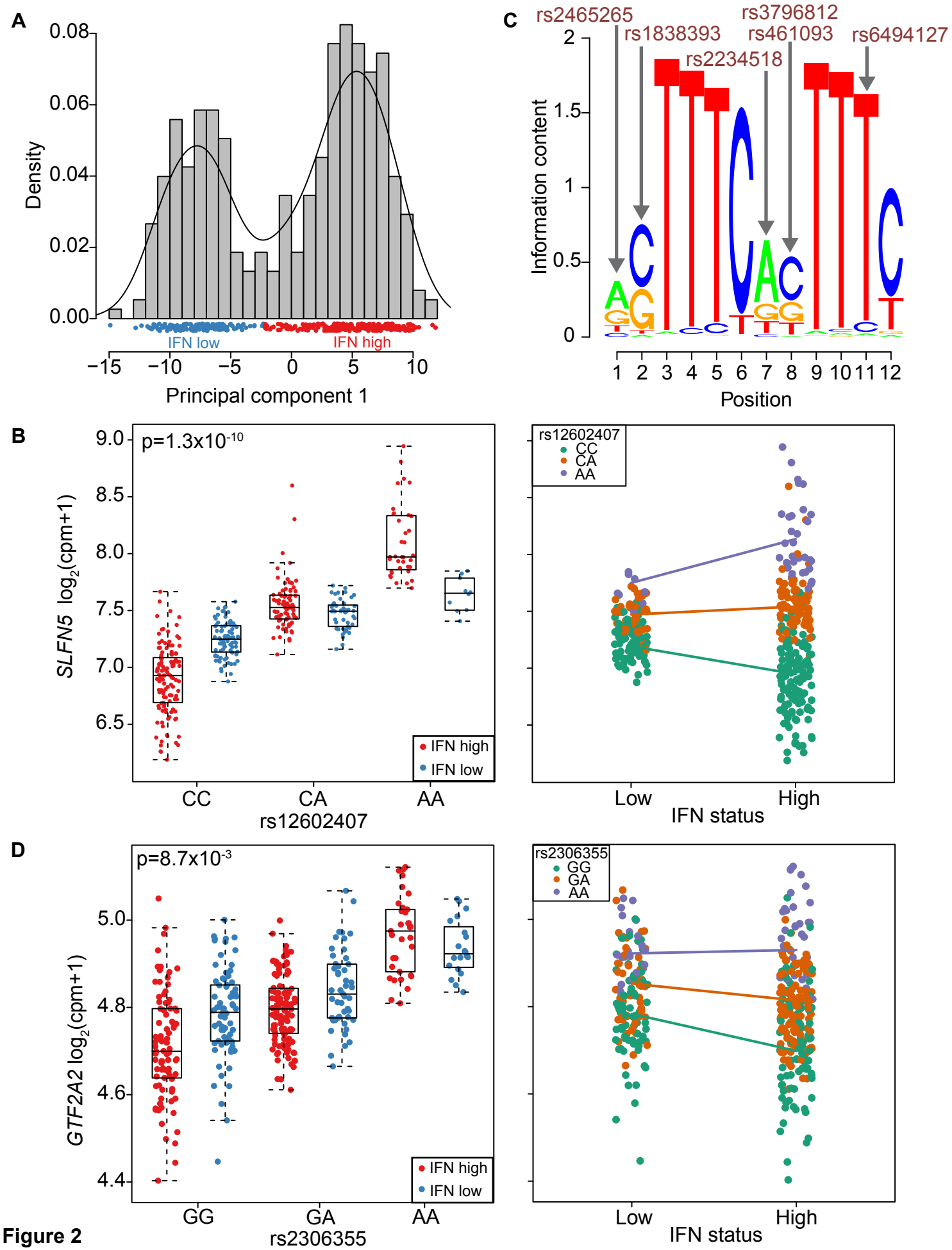
A
Density
0.08
0.06
0.04
0.02
0.00
IFN low
IFN high
−15
−10
−5
0
5
10
Principal component 1
C
rs2465265
rs1838393
rs2234518
rs3796812
rs461093
rs6494127
Information content
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12
Position
B
p=1.3x10-10
SLFN5 log2(cpm+1)
9.0
8.5
8.0
7.5
7.0
6.5
IFN high
IFN low
CC
CA
AA
rs12602407
rs12602407
CC
CA
AA
Low
High
IFN status
D
p=8.7x10-3
GTF2A2 log2(cpm+1)
5.0
4.8
4.6
4.4
IFN high
IFN low
GG
GA
AA
rs2306355
rs2306355
GG
GA
AA
Low
High
IFN status

Figure 2

Figure 2. eQTL interactions with IFN status (**A**) Designation of IFN status for each sample from the real-time PCR expression of 11 genes (first principal component). (**B**) IFN status interaction with the *SLFN5* eQTL plotted with respect to rs12602407 genotype (left) and IFN status of the sample (right). (**C**) The ISRE motif enriched among eQTLs magnified in IFN high samples. Arrows indicate positions of the motif interrupted by interaction SNPs (or SNPs in strong LD). Red indicates these SNPs correspond to magnified eQTLs. (**D**) IFN status interaction with the *GTF2A2* eQTL plotted with respect to rs2306355 genotype (left) and IFN status of the sample (right).

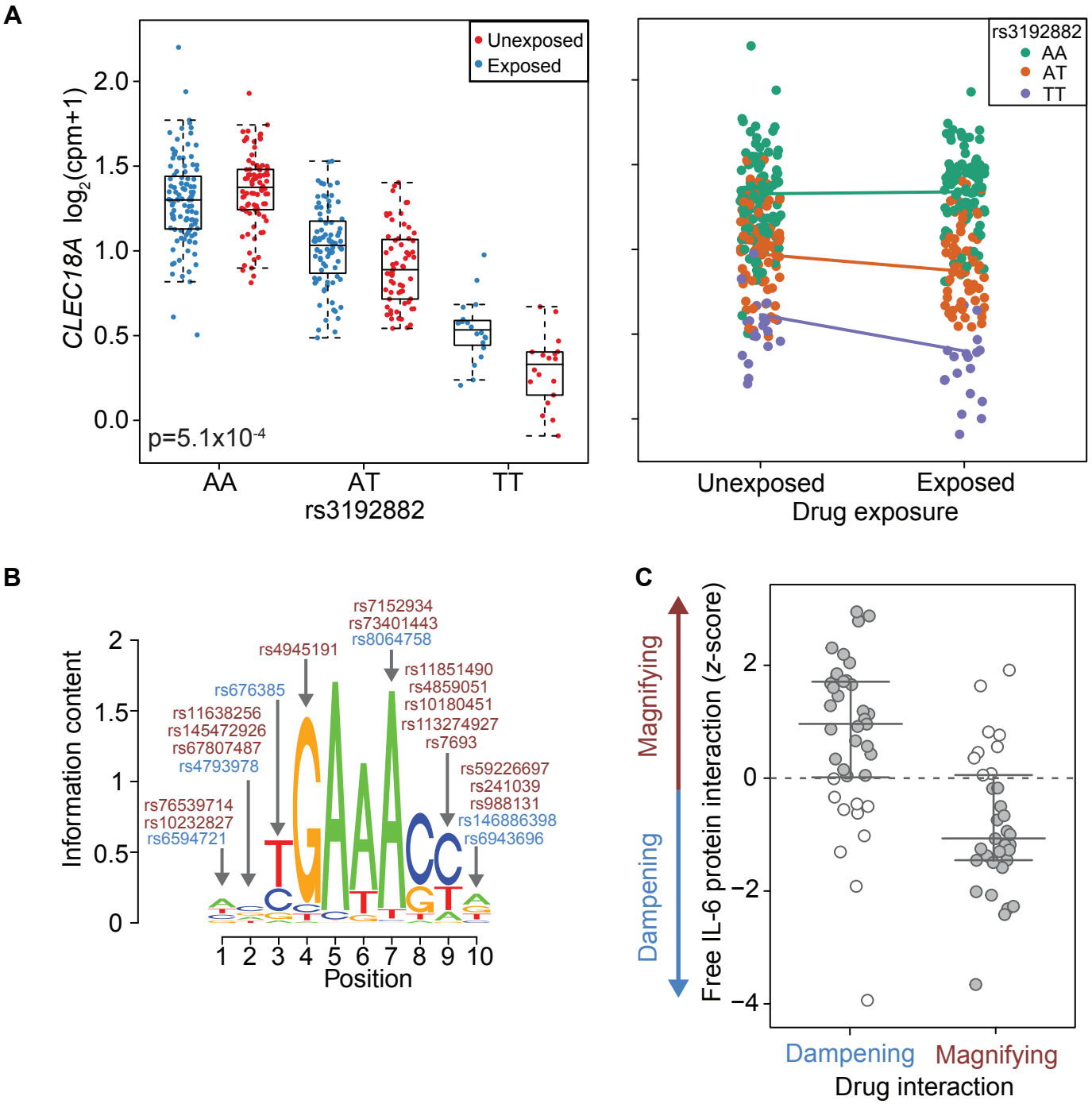

Figure 3. eQTL interactions with drug exposure (A) Drug exposure interaction with the *CLEC18A* eQTL plotted with respect to rs3192882 genotype (left) and drug exposure (right). (B) The IRF4 motif enriched among eQTLs magnified following drug treatment. Arrows indicate positions of the motif interrupted by interaction SNPs (or SNPs in strong LD). Red and blue indicate SNPs corresponding to magnified and dampened eQTLs respectively. (C) Concordance of free IL-6 protein interaction effects with drug exposure interaction effects (grey indicates expected opposite direction).