Title: Discovering *in vivo* eQTL interactions with interferon status and drug exposure from a lupus clinical trial

Authors: Emma E. Davenport,[1,2,3] Tiffany Amariuta,[1,2,3,4] Maria Gutierrez-Arcelus,[1,2,3] Kamil Slowikowski,[1,2,3,4] Harm-Jan Westra,[1,2,3] Ying Zhang,[5] Stephen Pearson,[6] David von Schack,[5] Jean S. Beebe,[5] Nan Bing,[5] Michael S. Vincent,[5] Baohong Zhang[5] and Soumya Raychaudhuri[1,2,3,4,7*]

Affiliations:
[1]Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA.
[2]Partners Center for Personalized Genetic Medicine, Boston, MA, 02115, USA.
[3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA.
[5]Pfizer Inc., Cambridge, MA, 02139, USA.
[6]Pfizer New Haven Clinical Research Unit, New Haven, CT, 06511, USA.
[7]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:

If an expression quantitative trait locus (eQTL) effect is modulated by an environmental stimulus, such as drug exposure or disease status, it can point to key regulatory mediators. In a clinical trial for anti-IL-6 in 157 patients with systemic lupus erythematosus we measured cell counts, interferon (IFN) status, drug exposure and genome-wide gene expression at three time points. First, we confirmed an increase in power using repeat transcriptomic measurements. Then, after detecting 4,976 *cis* eQTLs, we discovered that 154, 185 and 126 had evidence of significant eQTL interactions with T cell proportion, IFN status and anti-IL-6 drug exposure respectively. Next, we found an enrichment of transcription factor binding motifs interrupted by eQTL interaction SNPs, pointing to regulatory mediators of these environmental stimuli and therefore potential therapeutic targets for autoimmune diseases. In particular, IFN interactions are enriched for IRF1 binding site motifs, while anti-IL-6 interactions are enriched for IRF4 motifs. Finally, we used the drug-eQTL interactions to define an informative drug exposure score, reflecting a drug's effect in an individual patient, thus highlighting the potential for utilizing drug-eQTL interactions in a pharmacogenetic framework.

Main Text:

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[1] and while most are stable across tissues and conditions, environmental perturbations can alter the effects of some of them[2–8]. If a perturbation disrupts upstream regulatory mechanisms for a gene then it could magnify or dampen an eQTL effect, resulting in a genetic by environment interaction. Therefore, observing a set of eQTL interactions due to a perturbagen (such as drug exposure or disease status) can identify shared upstream regulatory mechanisms. These may represent key transcription factors and affected pathways that inform our understanding of both disease and drug mechanisms. Defining the molecular effects of a drug's action is particularly critical as these effects could help to classify likely non-responders, indicate off-target effects, predict the toxicities of the medication, find more accurate biomarkers and identify diseases for which a drug might be repurposed.

However, *cis* eQTL interactions with environmental factors in humans have been challenging to discover *in vivo*[9–13] even with large cohorts[1,7]. Success at finding *cis* eQTL interactions has largely been found in studies using model organisms[14,15] or treating cells *in vitro* with non-physiologic conditions[16]. Thus far, these studies might be limited in power since they either map eQTLs separately across conditions and fail to exploit the power of repeat measurements[17]. Or they test for genetic variants associated with differential expression and miss information about the magnitude of the eQTL effect in a specific condition[18].

88

89 Here, we predicted that if RNA is queried at multiple time points under different
90 exposure states, repeat measurements could increase power to not only detect eQTLs,
91 but also their interactions with environment.

92

93 Clinical trials, with their structured study design, may be the ideal setting to detect eQTL
94 interactions with drugs or other environmental perturbations. In clinical trials, it is
95 becoming increasingly common to collect transcriptional and genetic data alongside
96 clinical and physiological data. This extensive phenotyping of the same individuals at
97 multiple time points across conditions can be leveraged to identify eQTL interactions.
98 The profiling of an individual both before and after exposure to a drug provides a unique
99 opportunity to identify *in vivo* drug-eQTL interactions.

100

101 As a proof of principle, we examined the modulation of eQTL effects by environmental
102 factors, including cell count, interferon (IFN) status and drug exposure, using data from
103 a phase II clinical trial to evaluate the safety and efficacy of a neutralizing IL-6
104 monoclonal antibody (PF-04236921) in 157 systemic lupus erythematosus (SLE)
105 patients[19] (**Methods**). IL-6 is a cytokine elevated in SLE and other autoimmune
106 diseases such as rheumatoid arthritis (RA). In SLE, IL-6 is thought to play a role in the
107 observed B cell hyperactivity and autoantibody production[20]. The IL-6 receptor has been
108 successfully targeted with tocilizumab in RA[21] and has shown promise in a phase I trial
109 for SLE[22]. IL-6 itself has been targeted with siltuximab for the treatment of Castleman's
110 disease[23] and this has led to the development of other biologics targeting IL-6, such as
111 PF-04236921. While this drug was not significantly different from placebo for the

112 primary efficacy endpoint (proportion of patients achieving the SLE Responder Index
113 (SRI-4) at week 24), biologically it effectively reduced free IL-6 protein levels. Given the
114 key role of IL-6 in a range of diseases, IL-6 blockade and mechanism of action are of
115 great interest to study.
116
117 In this study, we leverage the power of repeat transcriptional and environmental
118 measurements from a lupus clinical trial to identify eQTL interactions. We discover
119 eQTL interactions with cell count, IFN status and drug exposure. We find the eQTL
120 interaction SNPs are enriched for transcription factor binding motifs (such as IRF1 and
121 IRF4 for IFN and drug exposure respectively) highlighting regulatory mediators of these
122 interactions and potential therapeutic targets.
123
124 **Results**
125 We conducted whole blood high-depth RNA-seq profiling at 0, 12, and 24 weeks in anti-
126 IL-6 exposed and unexposed individuals with the Illumina TruSeq protocol. We
127 observed and quantified 20,253 gene features and genotyped 608,017 variants
128 genome-wide (**Methods**). Along with each RNA-seq assay, we documented drug
129 exposure and quantified cell counts with FACS and IFN signature status with real-time
130 PCR.
131
132 *Mapping eQTL in SLE patients*
133 We first mapped *cis* eQTLs (SNPs within 250kb of the transcription start site of the
134 gene) and then tested those eQTLs for interactions with cell counts, IFN status and drug

135 exposure. eQTL interactions can be explored using our interactive visualization tool
136 (http://baohongz.github.io/Lupus_eQTL).
137
138 We used a linear mixed model, including repeat measurements with up to three RNA-
139 seq assays per patient (**Fig. 1A,** 379 samples from 157 patients, **Methods**). To
140 maximize power, we adjusted for 5 population and 25 gene expression principal
141 components (**Methods**). To ensure a set of highly confident eQTLs, we applied a
142 stringent multiple hypothesis testing correction and identified 4,976 *cis* eQTL genes with
143 $p_{eqtl}<2.3x10^{-8}$ (0.05/2,177,889 tests, **Fig. 1B,1C, Supplementary Table 1**).
144
145 *Repeat measurements increase power to detect eQTL*
146 We observed that repeat samples increased our power by detecting 63% more *cis*
147 eQTLs compared to using a single sample per individual (3,061 genes **Fig. 1B**). Our
148 results are highly concordant with the BIOS cohort, a much larger dataset of 2,166
149 healthy individuals[1]. 85.4% of our SLE eQTL SNP-gene pairs are reported as eQTLs in
150 BIOS (FDR<0.05). Of these, 99.1% showed consistent direction of effect ($p<5x10^{-16}$,
151 binomial test, **Fig. 1D**). For each of the 4,976 *cis* eQTL genes, we tested the most
152 significantly associated SNP for environmental interactions.
153
154 *eQTL interactions with T cell proportions*
155 We first tested a type of eQTL interaction that has been examined previously: cell
156 counts[24,25]. We obtained FACS data for 320 samples for which we had
157 contemporaneous RNA-seq profiles (n=152 subjects). We determined the percentage of
158 total lymphocytes that were T cells by gating (**Fig. 2A**). We found 154 T cell-eQTL

159 interactions with nominal evidence ($p_{interact}$<0.01, **Supplementary Table 1**), whereas
160 from 4,976 tests we would expect ~50 from chance alone. To ensure that our statistics
161 were not inflated, we conducted 1,000 stringent permutations, where we reassigned T
162 cell percentages across samples and retested. This permutation preserved the main
163 eQTL effect, while disrupting interactions that might be present in the data. In no
164 instance did we observe 154 or more interactions at $p_{interact}$<0.01 (maximum=133),
165 suggesting that the number of observed interactions is highly unlikely to have happened
166 by chance (**Fig. 2B**, $p_{permute}$~0/1,000 = <0.001).
167
168 Interactions can be divided into magnifiers, where environmental exposure and eQTL
169 effects are in the same direction, and dampeners where the effects work in the opposite
170 direction (**Fig. 2C, Supplementary Fig. 1**). The *NOD2* rs1981760 eQTL is an example
171 of an interaction that is dampened by increased T cell count (**Fig. 2D**, $p_{interact}$=6.5x10^{-5}),
172 and has separately been shown to vary across cell types[24,25]. For each 10% increase in
173 T cell proportion, the eQTL effect is reduced by about 7%.
174
175 *IFN status eQTL interactions*
176 Many patients with SLE exhibit high levels of genes induced by IFN alpha; these genes,
177 known as the IFN signature, are a marker of disease severity[26,27] and a pathogenic
178 feature of SLE. We explored the influence of IFN alpha on gene regulation after
179 determining the IFN status of every patient at each time point using real-time PCR of 11
180 IFN-inducible genes[28] (**Methods, Fig. 3A**). We identified 185 IFN-eQTL interactions
181 with $p_{interact}$<0.01 (**Supplementary Table 1**). Following the same permutation
182 procedures as above, our observed interactions were unlikely to have occurred by

183 chance (maximum permutation interactions=112, $p_{permute}$<0.001) (**Supplementary Fig.**
184 **2**). We note that interactions with a proxy gene for IFN status have been described[1] and
185 we find overlap of genes with those reported interactions (**Supplementary Fig. 3**). For
186 example, *SLFN5* expression is influenced by the rs12945522 SNP ($p_{interact}$ =1.3x10^{-10},
187 **Fig. 3B**). This effect is dampened in IFN low samples.
188
189 To define transcription factors that drive the response to IFN alpha, we sought to
190 identify motifs that explain the differences between magnifier (n=75) and dampener
191 (n=110) eQTLs (**Supplementary Fig. 4**). We applied HOMER[29] to assess overlap
192 between transcription factor binding motifs and the eQTL interaction SNPs (and SNPs
193 in high linkage disequilibrium (LD, r^2>0.8) in the *cis* window) (**Methods**). We found
194 significant enrichment of motifs for key transcription factors involved in IFN signaling
195 including the IRF1 motif. IRF1 motif disruption occurred for 11 genes with an eQTL
196 dampened in IFN low samples but for only 2 genes with an eQTL magnified (HOMER
197 p=0.001, permutation p<0.048, **Methods, Fig. 3C**, **Supplementary Table 2**). An
198 example is the *GTF2A2* rs2306355 eQTL ($p_{interact}$=8.7x10^{-3}, **Fig. 3D**); rs2306355 is in
199 tight LD (r^2=0.85 in Europeans) with rs6494127, which interrupts the GAAA core of the
200 IRF1 motif (**Fig. 3C**), and likely disrupts IRF1 binding[30]. We observe greater expression
201 of *GTF2A2* in individuals with the rs2306355 A allele compared to G; this difference is
202 dampened in IFN low individuals (**Fig. 3D**).
203
204 *Discovery of eQTL interactions with drug exposure*
205 We then examined whether IL-6 blockade alters the relationship between genomic
206 variation and gene expression and induces drug-eQTL interactions. We observed 126

207 drug-eQTL interactions with $p_{interact}$<0.01 (**Supplementary Table 1**). Following the same
208 permutation strategy as above, we found a median of 77 interactions with $p_{interact}$<0.01
209 (maximum=117) from 1,000 permutations. This suggests that about half of our drug-
210 eQTL interactions likely represent real biological phenomena, and not statistical artifact
211 (**Supplementary Fig. 5**). These drug-eQTL interactions showed little overlap with the
212 interactions observed for T cell count or IFN status (**Supplementary Fig. 6**). We note
213 biologically relevant drug-eQTL interactions for *IL10* (**Supplementary Fig. 7**), an anti-
214 inflammatory cytokine, *CLEC4C* which has previously been associated in *trans* with an
215 SLE risk allele[31] and *CLEC18A* (**Fig. 4A**) another member of the C-type lectin domain
216 family.
217
218 Again, the drug-eQTL interactions can be divided into magnifiers and dampeners
219 (**Supplementary Fig. 8**) and a similar approach as described above can be applied to
220 define transcription factors driving the response to IL-6 blockade. The most enriched
221 motif for eQTLs magnified after drug treatment was IRF4. Strikingly, the IRF4 motif
222 disruption occurred for 16 genes, including *CLEC18A*, with an eQTL magnified after
223 drug treatment compared to 6 genes with an eQTL dampened ($p=1x10^{-5}$, permutation
224 p<0.01, **Fig. 4B**, **Methods**, **Supplementary Table 3**).
225
226 Searching for transcription factor binding motifs in DNA regions of interest may not be
227 the only way of identifying downstream mediators of IL-6. From chromatin
228 immunoprecipitation sequencing (ChIP-seq), our drug-eQTL interaction SNP
229 (rs2270843) for *CLEC18A* lies in a MAFK binding site. SNPs in high LD (r^2>0.8) in the

230 *cis* window also lie in JUND, CTCF, RAD21, SMC3 and ZNF143 binding sites. Hence
231 one or more of these factors could also be acting to translate changes in serum IL-6
232 levels to regulation of *CLEC18A* gene expression.
233
234 A more common strategy to determine the effect of a perturbagen is to use differential
235 gene expression. For differential expression following drug treatment, we identified
236 1,161 genes with nominal statistical evidence ($p<0.01$) but modest effects (max fold
237 change=1.3, **Supplementary Fig. 9**). Only 8/126 drug-eQTL interaction genes also
238 show evidence of differential gene expression. This suggests that eQTL interactions
239 offer independent information from differential expression, which might contribute to
240 defining mechanisms.
241
242 *Concordance of drug-eQTL interactions with protein level interactions*
243 To validate these interactions, we hypothesized that interactions due to drug exposure
244 are likely driven by free IL-6 cytokine levels (our key clinical biomarker of interest). If this
245 is the case, for eQTLs dampened by drug exposure, an increase in free IL-6 should
246 elicit an opposite interaction effect and result in eQTL magnification. We assessed
247 whether eQTL interactions with free IL-6 protein levels measured in the patient serum
248 samples were consistent with those following IL-6 blockade. We observed enrichment in
249 the overlap between cytokine interactions and drug interactions (91/126 interactions in
250 consistent direction, **Fig. 4C,** $p=3.2\times10^{-7}$, binomial test). We were concerned that free
251 IL-6 and drug exposure are not independent in this dataset, and that this concordance
252 might be in part due to the connection between free IL-6 protein levels and IL-6
253 blockade (**Supplementary Fig. 10**). To assess whether free IL-6 offers independent

254 interaction effects that were consistent after accounting for drug effect, we first modeled
255 free IL-6 levels to account for the presence or absence of drug, and then we assessed
256 interactions with residual IL-6 levels. Again, we observed a significant number of
257 interactions in a consistent direction with residual IL-6 levels, which are independent of
258 drug exposure (p=0.03, **Supplementary Fig. 11**).
259
260 *Drug score for assessing drug exposure*
261 We speculated that these drug-eQTL interactions could be used in a clinical
262 pharmacogenetic context to assess effective drug exposure for patients. We defined a
263 simple drug exposure score using the 126 drug-eQTL interactions (**Methods**). For each
264 RNA-seq sample, we assessed whether the expression of the interaction gene was
265 more consistent with the drug exposed or unexposed state for the corresponding
266 interaction SNP genotype. Samples more consistent with the drug-exposed state are
267 assigned a larger drug exposure score. Unsurprisingly, we found a difference in drug
268 exposure score between the unexposed and exposed samples (**Supplementary Fig.
269 12**) (r_s=0.79, p=6.9x10^{-81}); these differences reflect the fact that the eQTLs were
270 themselves identified by examining samples with and without drug exposure. However,
271 while we did not utilize the administered drug dose to identify drug-eQTL interactions,
272 we found a significant correlation between drug dose (10, 50 or 200mg) and drug
273 exposure score (r_s=0.16, p=0.018) in the drug-exposed samples (**Fig. 4D**).
274
275 **Discussion**
276 Here, we mapped eQTLs in a cohort of SLE patients and discovered interactions with
277 physiological (T cell abundance) and clinical (IFN status) data, but most importantly,

278 with drug exposure. We hypothesized that the eQTL effects modulated following
279 treatment with anti-IL-6 were likely being driven by changes in the free IL-6 protein
280 levels. Remarkably, we were able to replicate 91/126 of our drug-eQTL interactions at
281 the protein level, providing additional support that the vast majority of interactions are
282 biologically relevant.
283
284 Our ability to identify eQTL interactions in humans *in vivo* stems from the use of a
285 structured study design with repeat measurements of gene expression across different
286 conditions in the same individual. By allowing the same individual to be assessed under
287 different circumstances, the noise in transcriptomic measurements caused by non-
288 genetic factors is reduced. This is apparent in our increased power to detect main eQTL
289 effects with repeat measurements. We observed 4,976 eQTL genes with repeat
290 measurements versus 3,061 with only single measures.
291
292 Observing the same individuals when they are both unexposed and exposed to an
293 environmental perturbagen further increases power to find interactions. Structured study
294 designs such as clinical trials provide this design setup and therefore offer excellent
295 opportunities to identify interactions to drug exposure, thereby potentially illuminating
296 mechanisms. This strategy may be particularly informative for research conducted in a
297 clinical setting where the size of a cohort can be limited by the ability to recruit suitable
298 patients.
299

300 eQTL interactions with drug interventions or other environmental stimuli are important
301 because they indicate mechanisms of action by highlighting key regulatory factors, such
302 as transcription factors or subclasses of enhancers. These regulatory factors act
303 downstream of the environmental condition of interest and drive groups of eQTL
304 interactions. The IFN status eQTL interactions we identified provide support for this
305 approach. By making use of the direction of effect for the eQTL interaction, we were
306 able to identify an enrichment of dampening eQTL interaction SNPs interrupting the
307 binding sites of transcription factors known to be important in the response to IFN, such
308 as IRF1.
309
310 We were intrigued to discover an enrichment of magnifying drug-eQTL interaction SNPs
311 interrupting the binding site of IRF4. It has been suggested that IRF4 works downstream
312 of IL-6 by binding BATF and coordinately regulating the production of IL10 and other
313 genes[32]. Consistent with this, we observed that the *IL10* eQTL does indeed interact with
314 presence of anti-IL-6 (**Supplementary Fig. 7**). Previous studies have highlighted a role
315 for IRF4 in the pathogenesis of autoimmune diseases in mouse and humans. For
316 example in a murine model of SLE, *IRF4* knockout mice did not develop lupus
317 nephritis[33]. In humans, IRF4 is associated with RA[34], a disease in which anti-IL-6
318 treatment has been successful[21]. Our findings provide further support that IRF4 could
319 be a potential therapeutic target for autoimmune diseases such as RA where anti-IL-6 is
320 effective[35]. As the regulatory genome continues to be mapped[36,37] and available binding
321 sites for different regulator proteins are specifically defined, these approaches to define
322 environmental mechanisms will become more potent. For example, it will become easier

323 to connect groups of interaction eQTLs being driven by specific transcription factors or a
324 subset of regulatory elements.
325
326 In addition to defining drug interactions, the ability to focus on interactions with specific
327 patient phenotypes might point to key targets for disease intervention. For example, our
328 discovery of many IFN interactions was likely enabled by the prevalence of this key
329 immunophenotype in SLE patients compared to healthy controls[26,27]. Our study largely
330 replicates the IFN interactions discovered in a much larger study of healthy controls[1].
331 The IFN status immunophenotype is now a target for therapy. A recent phase II clinical
332 trial has shown that an antagonist to the type I IFN receptor, acting upstream of IRF1,
333 reduced severity of symptoms in SLE. Interestingly, the antagonist was more effective in
334 the patients with a high baseline IFN status[38]. This example provides a compelling case
335 study for how understanding master regulators of key disease phenotypes might lead to
336 promising new therapeutic strategies. We speculate that this provides a mechanism for
337 stratified medicine for future studies, which may be applicable to other diseases.
338
339 Finally, we argue that drug-eQTL interactions can augment pharmacogenetic strategies
340 and may be informative for patient response. For many biologic medications, predictive
341 pharmacogenetics has been challenging; for example, studies to define genetic or non-
342 genetic biomarkers of anti-TNF response have not been successful[39,40]. Our drug
343 exposure score, based on a novel approach using SNP-gene pairs from drug-eQTL
344 interactions, might reflect the biological activity that a medication is having upon an
345 individual, and may be modeling an effective medication activity level. This score may

346 therefore be helpful in stratifying individuals when assessing response to a medication
347 for example those with a higher drug exposure score may have a better response to
348 treatment. We note a limitation of this study is that the drug itself did not achieve its
349 primary efficacy endpoint of improving SLE outcomes. Hence, while the drug exposure
350 score for this study tracked with the biological effect of the drug (reducing free IL-6
351 protein levels), it might not be useful for SLE specifically. However, such a scoring
352 system could be implemented easily in most phase III trials for a broad range of
353 therapeutics, where the numbers of samples are far in excess of this phase II trial,
354 ensuring better powered and more accurate eQTL-interaction mapping.
355

356 **Methods**

357 *Study design*

358 The objectives of this study were to map eQTLs in a cohort of lupus patients and
359 identify eQTL interactions with environmental perturbations such as drug treatment to
360 shed light on drug and disease mechanisms. SLE patients were recruited to a phase II
361 clinical trial to test the efficacy and safety of an IL-6 monoclonal antibody (PF-
362 04236921). The patient population recruited to this trial have been detailed extensively
363 by Wallace et al.[19] 183 patients (forming a multiethnic cohort) were randomized to
364 receive three doses of drug (10, 50 or 200mg) or placebo at three time points during the
365 trial (weeks 0, 8 and 16).
366

367 *RNA-sequencing*

368 We collected peripheral venous blood samples in PAXgene Blood RNA tubes
369 (PreAnalytiX GmbH, BD Biosciences) for high-depth RNA-seq profiling at 0, 12, and 24
370 weeks. We extracted total RNA from blood samples using the PAXgene Blood RNA kit
371 (Qiagen) at a contract lab using a customized automation method. We assessed the
372 yield and quality of the isolated RNA using Quant-iT™ RiboGreen® RNA Assay Kit
373 (ThermoFisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies),
374 respectively. Following quality assessment, we processed an aliquot of 500-1000 ng of
375 each RNA with a GlobinClear-Human kit, (ThermoFisher Scientific) to remove globin
376 mRNA. We then converted RNA samples to cDNA libraries using TruSeq RNA Sample
377 Prep Kit v2 (Illumina) and sequenced using Illumina HiSeq 2000 sequencers. We
378 generated an average of 40M 100bp pair-end reads per sample for downstream
379 analysis.
380
381 We successfully obtained 468 RNA-seq profiles from 180 patients. We aligned reads to
382 the reference genome and quantified gene expression using Subread[41] and
383 featureCounts[42] respectively. We included genes with at least 10 reads (CPM>0.38) in
384 at least 32 samples (minimum number of patients with both unexposed and exposed
385 RNA-seq assays in a drug group) prior to normalization. Following quality control (QC),
386 we removed 4 samples as outliers. We then normalized 20,253 transcripts using the
387 trimmed mean of M-values method and the edgeR R package[43]. Expression levels are
388 presented as $\log_2$(cpm +1).
389
390 *Genotyping*

391 We genotyped 160 individuals across 964,193 variants genome-wide with the Illumina
392 HumanOmniExpressExome-8v1.2 beadchip. We removed SNPs if they deviated from
393 Hardy-Weinberg Equilibrium ($p < 1x10^{-7}$), had a minor allele frequency <5%,
394 missingness >2% or a heterozygosity rate greater than 3 standard deviations from the
395 mean (PLINK[44,45]). For mapping eQTLs, we removed SNPs on the Y chromosome.
396 Following QC, we used 608,017 variants for further analysis. We removed one sample
397 with high missingness and outlying heterozygosity rate from further analysis.
398
399 *Cell counts*
400 We collected blood samples for cytometry analysis at weeks 0, 12 and 24. Samples
401 were subjected to flow cytometry for T cell immunophenotyping. We counted T cells
402 (CD3+) as a percentage of lymphocytes (CD45+). FACS data were available for 320
403 samples from 152 subjects.
404
405 *Interferon status*
406 We classified the interferon (IFN) status of each sample at each time point from the
407 expression of 11 IFN response genes (*HERC5*, *IFI27*, *IRF7*, *ISG15*, *LY6E*, *MX1*, *OAS2*,
408 *OAS3*, *RSAD2*, *USP18*, *GBP5*)[28] using TaqMan Low Density Arrays. We classified
409 samples as high or low IFN based on the first PCA score for the expression of these
410 genes. IFN status was available for 376 samples from 157 subjects.
411
412 *Drug exposure*

413 Samples were assigned as unexposed (placebo or week 0 samples) or drug exposed
414 (week 12 and week 24 samples in the drug groups).
415
416 *Free IL-6 protein levels*
417 We determined free IL-6 protein levels from serum using a commercial sandwich ELISA
418 selected for binding only free IL-6. The assay was validated according to FDA
419 biomarker and fit-for purpose guidelines. Free IL-6 protein levels were available for 311
420 samples from 145 subjects. Since IL6 levels were highly nonparametric, we ranked
421 samples in order of IL-6 protein levels and included in the model to identify cytokine-
422 eQTL interactions.
423
424 *Statistical analysis*
425 *eQTL and interaction analysis*
426 In total, 157 patients (with 379 RNA-seq samples) had good quality gene expression
427 and genotyping data for eQTL analysis. All statistical analyses were carried out in R[46].
428
429 We defined a *cis* eQTL as the SNP within 250kb either side of the GENCODE[47]
430 transcription start site of the gene. We first applied a linear model for the first available
431 time point to identify each eQTL using the first 25 principal components of gene
432 expression and the first 5 principal components of genotyping as covariates. SNPs were
433 encoded as 0, 1 and 2. To adjust for multiple testing during eQTL discovery we used a
434 stringent Bonferroni corrected *p*-value threshold of 2.3×10^{-8} (0.05/ 2,177,889 tests).
435

To map eQTLs using multiple samples for each individual, we applied a random intercept linear mixed model using the first 25 principal components of gene expression and the first 5 principal components of genotyping as covariates and patient as a random effect:

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

Where $E_{i,j}$ is gene expression for the ith sample from the jth subject, θ is the intercept, β_{geno} is the genotype effect (eQTL), $(\kappa_j|i)$ is the random effect for the ith sample from the jth subject, $pc_{i,l}$ is principal component l of gene expression for sample i, $pc_{j,m}$ is principal component m of genotyping for subject j.

We used the most significant SNP (with $p<2.3\times10^{-8}$) from the 4,976 identified eQTL genes to explore eQTL interactions. For each environmental interaction analysis, we further filtered these eQTLs to include only those with at least two individuals homozygous for the minor allele of the SNP being tested in each of the environmental factor groups. For example we required two of these individuals in each of the drug exposed and drug unexposed groups. To identify eQTL interactions, we added an additional covariate to the model for example drug exposure, and an interaction term between this covariate and the genotype of the SNP:

457
$$E_{i,j} = \theta + \beta_{geno} \cdot g_j + (\kappa_j | i) + \sum_{l=1}^{25} \phi_l \cdot pc_{i,l} + \sum_{m=1}^{5} \gamma_m \cdot pc_{j,m} + \beta_{drug} \cdot d_i + \beta_x \cdot d_i \cdot g_j$$

458

459 Where $E_{i,j}$ is gene expression for the ith sample from the jth subject, θ is the intercept,
460 β_{geno} is the genotype effect (eQTL), $(\kappa_j|i)$ is the random effect for the ith sample from the
461 j^{th} subject, $pc_{i,l}$ is principal component l of gene expression for sample i, $pc_{j,m}$ is principal
462 component m of genotyping for subject j, β_{drug} is the drug effect (differential gene
463 expression) and β_x is the interaction effect.

464

465 We determined the significance of the interaction term with a likelihood ratio test.

466

467 To rigorously confirm the relative enrichment of eQTL interactions, we shuffled the
468 interaction covariate (for example drug exposure) 1,000 times and calculated the
469 number of significant interactions observed in each permutation. For T cell counts and
470 IFN high/low status, we shuffled across all samples. For drug interaction permutation
471 analysis, we maintained the number of individuals in the drug group and the number of
472 samples with exposure to drug.

473

474 *Concordance with an eQTL study in healthy individuals*

475 In the SLE cohort, we classified 4,976 *cis* eQTL genes ($p<2.3x10^{-8}$). The *z*-score for the
476 most associated SNP for each of these genes was compared to the *z*- score from a
477 previously published eQTL dataset from whole blood from 2,166 healthy individuals[1].
478 4,250/4976 SNP-gene pairs (85.4%) were also reported in the BIOS dataset

479 (FDR<0.05). After removing 60 SNPs, which could not be mapped to a strand
480 4,154/4,190 (99.1%) had a *z*-score (eQTL effect) in a consistent direction.
481
482 *Magnifiers and Dampeners*
483 An eQTL interaction can either magnify or dampen the original eQTL effect. We
484 multiplied the interaction *z*-score by the sign of the original eQTL effect (genotype beta)
485 and defined magnifiers as interactions with an adjusted *z*-score > 0 and dampeners as
486 interactions with an adjusted *z*-score < 0.
487
488 *Differential gene expression analysis*
489 To identify differentially expressed genes following drug exposure, we applied a random
490 intercept linear mixed model using the first 25 principal components of gene expression
491 and the first 5 principal components of genotyping as covariates and patient as a
492 random effect.
493
494 *Conditional analysis for IL-6 protein levels*
495 We modeled the relationship between free IL-6 protein levels and drug exposure using
496 a linear model. We used the residuals from this model in our interaction linear mixed
497 model to identify IL-6 protein interactions independent of drug exposure.
498
499 *Drug exposure score*
500 We used linear discriminant analysis to assign a drug exposure score for each sample.
501 A score was calculated for each gene (see equation below) and then the final drug
502 exposure score is the average across the 126 drug-eQTL genes.

503

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

504

505 Where G is gene expression for a given sample, G_{Unexp} is predicted mean gene
506 expression for unexposed samples of the relevant SNP genotype, G_{Exp} is predicted
507 mean gene expression for exposed samples of the relevant SNP genotype and SE is
508 standard error for the intercept term of the model (unexposed expression for genotype
509 0).

510

511 *HOMER analysis for transcription factor binding motif enrichment*

512 We used the HOMER software suite[29] to look for enrichment of transcription factor
513 binding motifs in the 185 IFN-eQTL interactions ($p<0.01$) and the 126 drug-eQTL
514 interactions ($p<0.01$). Each eQTL interaction was identified using the most highly
515 associated SNP for that eQTL. However, as this SNP is not necessarily the functional
516 SNP, we additionally considered all those with an $r^2 \geq 0.8$ in the 1000 Genomes
517 European population[48] within 250kB of the transcription start site of the gene. We
518 defined our motif search window as 20 bp on either side of each SNP (i.e. 41 bp wide).

519

520 For each environmental factor, we divided the eQTL interactions into magnifiers or
521 dampeners and conducted two separate HOMER analyses: one with magnifiers in the
522 foreground and dampeners in the background; the other with dampeners in the
523 foreground and magnifiers in the background. HOMER reported the transcription factor

524 motifs that were significantly enriched in the foreground relative to background. Motifs
525 were plotted using the SeqLogo R library[49].
526
527 We determined permutation p values for enrichment of the IRF1 and IRF4 transcription
528 factor binding sites as follows. For IRF1, the motif is interrupted by interaction SNPs (or
529 SNPs in LD) corresponding to 11 dampening genes and 2 magnifying genes. We
530 permuted which genes were labeled as magnifiers or dampeners 10,000 times and
531 counted the number of genes in each category with an IRF1 motif interrupted. We found
532 478 occurrences from 10,000 trials with at least 11 dampening genes ($p<0.0479$). For
533 IRF4 the motif is interrupted by SNPs corresponding to 16 magnifying genes and 6
534 dampening genes. Using the same permutation approach, we found 133 occurrences
535 from 10,000 trials with at least 16 magnifying genes ($p<0.0134$).
536
537
538 **References**

539 1. Zhernakova, D. V *et al.* Identification of context-dependent expression quantitative
540 trait loci in whole blood. *Nat. Genet.* **49,** 139–145 (2017).
541 2. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory
542 Variants upon Monocyte Gene Expression. *Science* **343,** 1246949 (2014).
543 3. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies
544 cell type-specific master regulators and roles of HLA alleles. *Nat Genet* **44,** 502–
545 510 (2012).
546 4. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative
547 Risk Alleles in Leukocytes. *Science* **344,** 519–523 (2014).
548 5. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis:
549 Multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
550 6. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple
551 human tissues: the MuTHER study. *PLoS Genet.* **7,** e1002003 (2011).
552 7. Kukurba, K. R. *et al.* Impact of the X chromosome and sex on regulatory variation.
553 *Genome Res.* **26,** 768–777 (2016).
554 8. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by

555 transcriptome sequence analysis in twins. *Nat. Genet.* **47,** 88–91 (2015).
556 9. Maranville, J. C., Luca, F., Stephens, M. & Di Rienzo, A. Mapping gene-
557 environment interactions at regulatory polymorphisms: insights into mechanisms
558 of phenotypic variation. *Transcription* **3,** 56–62 (2012).
559 10. Idaghdour, Y. & Awadalla, P. Exploiting gene expression variation to capture
560 gene-environment interactions for disease. *Front. Genet.* **4,** 1–7 (2013).
561 11. Idaghdour, Y. *et al.* Evidence for additive and interaction effects of host genotype
562 and infection in malaria. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16786–16793 (2012).
563 12. Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression
564 variation in southern Morocco. *Nat. Genet.* **42,** 62–67 (2010).
565 13. Peters, J. E. *et al.* Insight into Genotype-Phenotype Associations through eQTL
566 Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS*
567 *Genet.* **12,** e1005908 (2016).
568 14. Li, Y. *et al.* Mapping determinants of gene expression plasticity by genetical
569 genomics in C. elegans. *PLoS Genet.* **2,** 2155–2161 (2006).
570 15. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene
571 expression. *PLoS Biol.* **6,** 810–824 (2008).
572 16. Maranville, J. C. *et al.* Interactions between glucocorticoid treatment and Cis-
573 regulatory polymorphisms contribute to cellular response phenotypes. *PLoS*
574 *Genet.* **7,** e1002162 (2011).
575 17. Barreiro, L. B. *et al.* Deciphering the genetic architecture of variation in the
576 immune response to Mycobacterium tuberculosis infection. *Proc. Natl. Acad. Sci.*
577 *U. S. A.* **109,** 1204–1209 (2012).
578 18. Smirnov, D., Morley, M. & Shin, E. Genetic analysis of radiation-induced changes
579 in human gene expression. *Nature* **459,** 587–591 (2009).
580 19. Wallace, D. J. *et al.* Efficacy and safety of an interleukin 6 monoclonal antibody
581 for the treatment of systemic lupus erythematosus: a phase II dose-ranging
582 randomised controlled trial. *Ann. Rheum. Dis.* **76,** 534–542 (2017).
583 20. Tackey, E., Lipsky, P. E. & Illei, G. G. Rationale for interleukin-6 blockade in
584 systemic lupus erythematosus. *Lupus* **13,** 339–343 (2004).
585 21. Tanaka, Y. & Mola, E. M. IL-6 targeting compared to TNF targeting in rheumatoid
586 arthritis: studies of olokizumab, sarilumab and sirukumab. *Ann. Rheum. Dis.* **73,**
587 1595–1597 (2014).
588 22. Illei, G. G. *et al.* Tocilizumab in systemic lupus erythematosus: Data on safety,
589 preliminary efficacy, and impact on circulating plasma cells from an open-label
590 phase I dosage-escalation study. *Arthritis Rheum.* **62,** 542–552 (2010).
591 23. Van Rhee, F. *et al.* Siltuximab for multicentric Castleman's disease: A
592 randomised, double-blind, placebo-controlled trial. *Lancet Oncol.* **15,** 966–974
593 (2014).
594 24. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLoS*
595 *Genet.* **11,** e1005223 (2015).
596 25. Naranbhai, V. *et al.* Genomic modulators of gene expression in human
597 neutrophils. *Nat. Commun.* **6,** 7545 (2015).
598 26. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral
599 blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A.* **100,**
600 2610–2615 (2003).

601 27. Bennett, L. *et al.* Interferon and granulopoiesis signatures in systemic lupus
602 erythematosus blood. *J. Exp. Med.* **197,** 711–723 (2003).
603 28. Hill, A. A. *et al.* FRI0003 Determination of interferon (IFN) signatures for sle
604 patients may be critical for optimal treatment selection but depends on the genes
605 chosen: report from the bold (biomarkers of lupus disease) study. *Ann. Rheum.*
606 *Dis.* **72,** A369–A370 (2013).
607 29. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription
608 Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell
609 Identities. *Mol. Cell* **38,** 576–589 (2010).
610 30. Escalante, C. R., Yie, J., Thanos, D. & Aggarwal, A. K. Structure of IRF-1 with
611 bound DNA reveals determinants of interferon regulation. *Nature* **391,** 103–106
612 (1998).
613 31. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of
614 known disease associations. *Nat. Genet.* **45,** 1238–43 (2013).
615 32. Koch, S. *et al.* IL-6 activated integrated BATF/IRF4 functions in lymphocytes are
616 T-bet-independent and reversed by subcutaneous immunotherapy. *Sci. Rep.* **3,**
617 1754 (2013).
618 33. Lech, M. *et al.* IRF4 Deficiency Abrogates Lupus Nephritis Despite Enhancing
619 Systemic Cytokine Production. *J. Am. Soc. Nephrol.* **22,** 1443–1452 (2011).
620 34. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug
621 discovery. *Nature* **506,** 376–81 (2014).
622 35. Xu, W. D., Pan, H. F., Ye, D. Q. & Xu, Y. Targeting IRF4 in autoimmune diseases.
623 *Autoimmun. Rev.* **11,** 918–924 (2012).
624 36. ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements
625 in the human genome. *Nature* **489,** 57–74 (2012).
626 37. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference
627 human epigenomes. *Nature* **518,** 317–330 (2015).
628 38. Furie, R. *et al.* Anifrolumab, an Anti-Interferon Alpha Receptor Monoclonal
629 Antibody, in Moderate to Severe Systemic Lupus Erythematosus (SLE). *Arthritis*
630 *Rheumatol* **69,** 376–386 (2017).
631 39. Cui, J. *et al.* Genome-wide association study and gene expression analysis
632 identifies CD84 as a predictor of response to etanercept therapy in rheumatoid
633 arthritis. *PLoS Genet.* **9,** e1003394 (2013).
634 40. Cui, J. *et al.* The role of rare protein-coding variants to anti-TNF treatment
635 response in rheumatoid arthritis. *Arthritis Rheumatol.* **69,** 735–741 (2017).
636 41. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable
637 read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108–e108 (2013).
638 42. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose
639 program for assigning sequence reads to genomic features. *Bioinformatics* **30,**
640 923–930 (2014).
641 43. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package
642 for differential expression analysis of digital gene expression data. *Bioinformatics*
643 **26,** 139–140 (2010).
644 44. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
645 richer datasets. *Gigascience* **4,** 7 (2015).
646 45. Purcell, S. & Chang, C. PLINK 1.9. at <https://www.cog-genomics.org/plink2>

46. R Core Team. R: A Language and Environment for Statistical Computing. (2015). at <https://www.r-project.org>
47. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).
48. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
49. Bembom, O. seqLogo: Sequence logos for DNA sequence alignments. (2016).

Acknowledgements

This work is supported in part by funding from the National Institutes of Health (U01GM092691, UH2AR067677, U19AI111224 (SR)), the Doris Duke Charitable Foundation Grant #2013097 and the Ruth L. Kirschstein National Research Service Award (F31AR070582) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (KS). This work is also supported by unrestricted funding from Pfizer, Inc.

Author Contributions

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

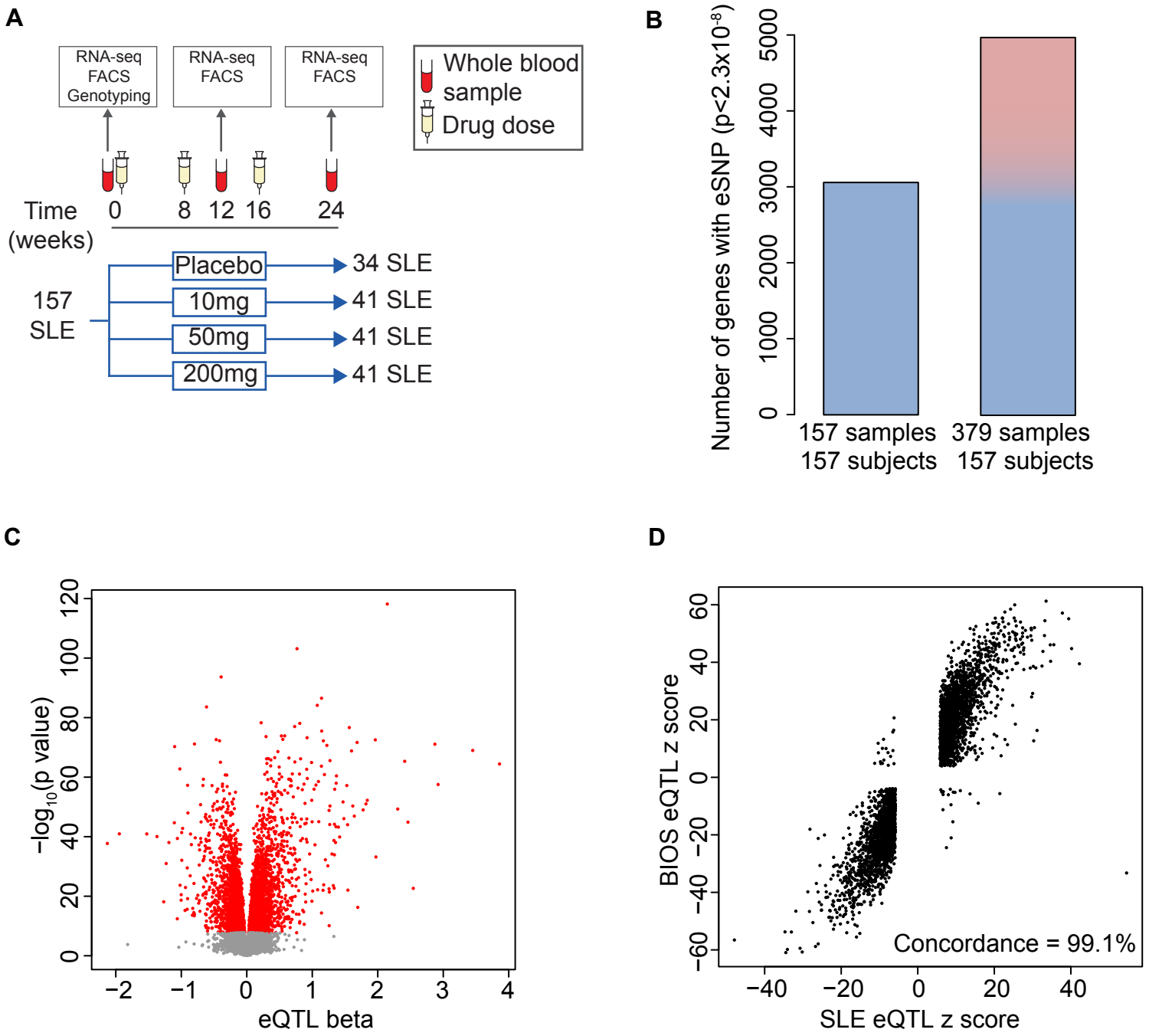
A
RNA-seq
FACS
Genotyping
RNA-seq
FACS
RNA-seq
FACS
Whole blood sample
Drug dose
Time (weeks)
0
8
12
16
24
157 SLE
Placebo
10mg
50mg
200mg
34 SLE
41 SLE
41 SLE
41 SLE
B
Number of genes with eSNP (p<2.3x10^-8)
0
1000
2000
3000
4000
5000
157 samples
157 subjects
379 samples
157 subjects
C
−log10(p value)
eQTL beta
D
BIOS eQTL z score
SLE eQTL z score
Concordance = 99.1%

Figure 1

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). (**C**) Volcano plot of eQTL effects for the most significantly associated SNP for each gene (red color indicates $p<2.3x10^{-8}$). (**D**) Concordance of SLE eQTL effects ($p<2.3x10^{-8}$) with eQTLs observed in the BIOS cohort[1] of healthy individuals (FDR <0.05). Each point represents the most significant SNP-gene pair for the SLE eQTL.

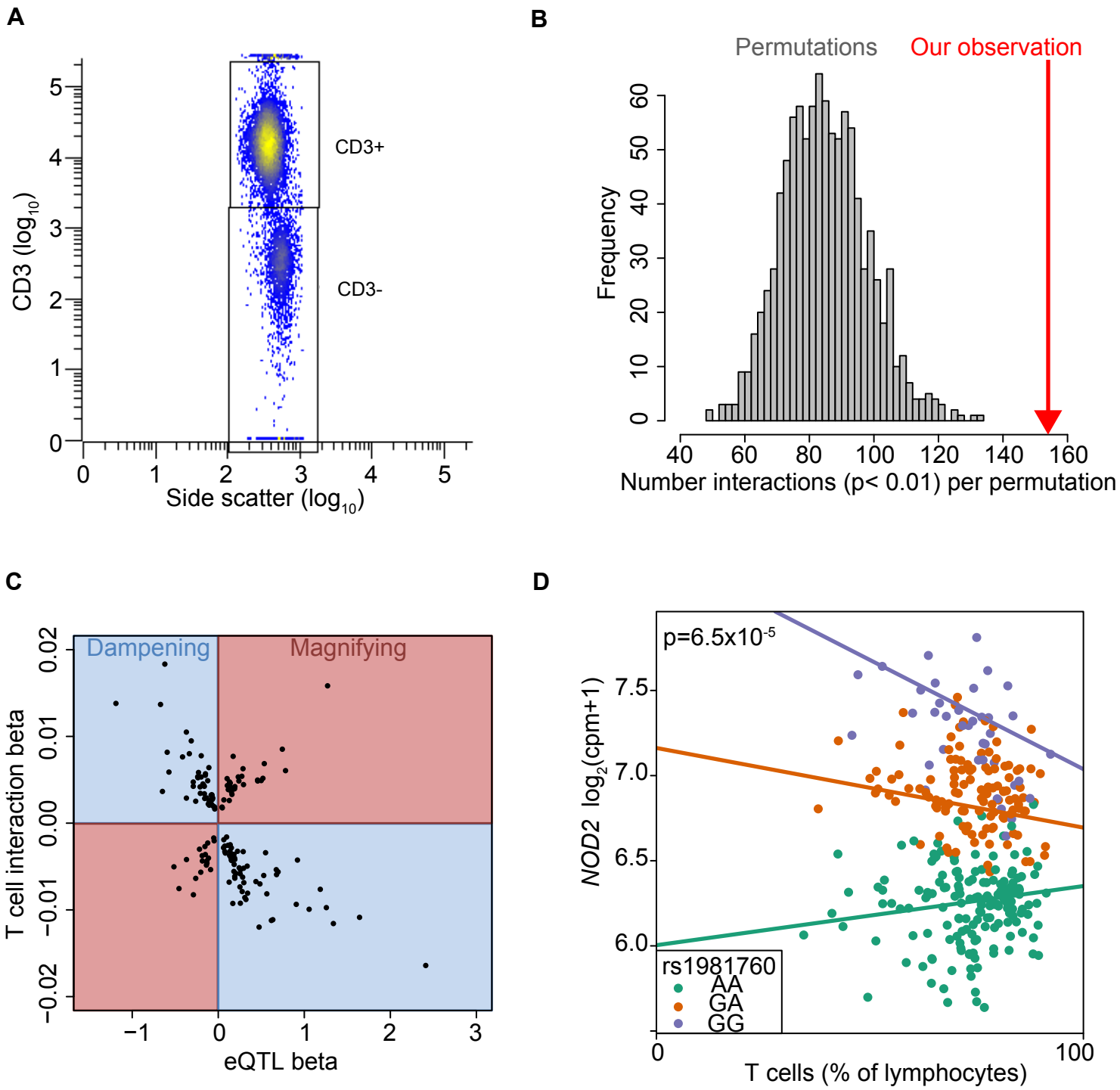
A
CD3+
CD3-
CD3 (log10)
Side scatter (log10)
B
Permutations
Our observation
Frequency
Number interactions (p< 0.01) per permutation
C
Dampening
Magnifying
T cell interaction beta
eQTL beta
D
p=6.5x10-5
NOD2 log2(cpm+1)
rs1981760
AA
GA
GG
T cells (% of lymphocytes)

Figure 2

Figure 2. eQTL interactions with T cell percentages (**A**) Flow cytometry gating strategy to determine T cells as a percentage of lymphocytes. (**B**) Number of significant interactions ($p<0.01$) from 1,000 permutations of T cell counts (median=85). (**C**) Effect of the T cell interaction on the original eQTL effect. (**D**) T cell interaction with the *NOD2* eQTL. The eQTL effect is dampened as T cell percentage increases.

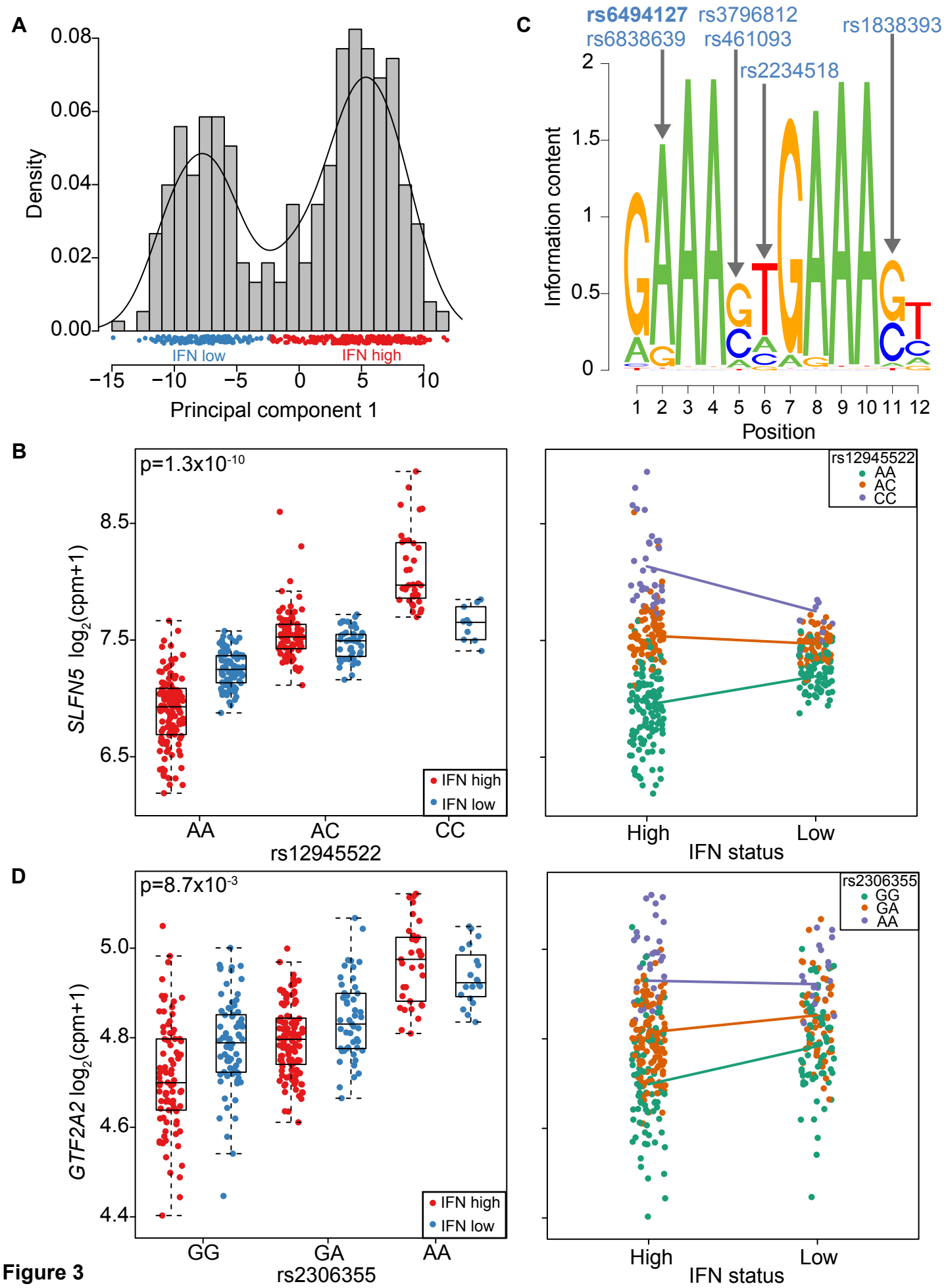
A
Density
0.08
0.06
0.04
0.02
0.00
IFN low
IFN high
−15
−10
−5
0
5
10
Principal component 1
C
rs6494127
rs3796812
rs6838639
rs461093
rs1838393
rs2234518
Information content
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12
Position
B
p=1.3x10-10
SLFN5 log2(cpm+1)
8.5
7.5
6.5
IFN high
IFN low
AA
AC
CC
rs12945522
rs12945522
AA
AC
CC
High
Low
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
High
Low
IFN status

Figure 3

Figure 3. 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 rs12945522 genotype (left) and IFN status of the sample (right). (**C**) The IRF1 motif enriched among eQTLs dampened in IFN low samples. Arrows indicate positions of the motif interrupted by interaction SNPs (or SNPs in strong LD). Blue indicates these SNPs correspond to dampened eQTLs. (**D**) IFN status interaction with the *GTF2A2* eQTL plotted with respect to rs2306355 genotype (left) and IFN status of the sample (right).

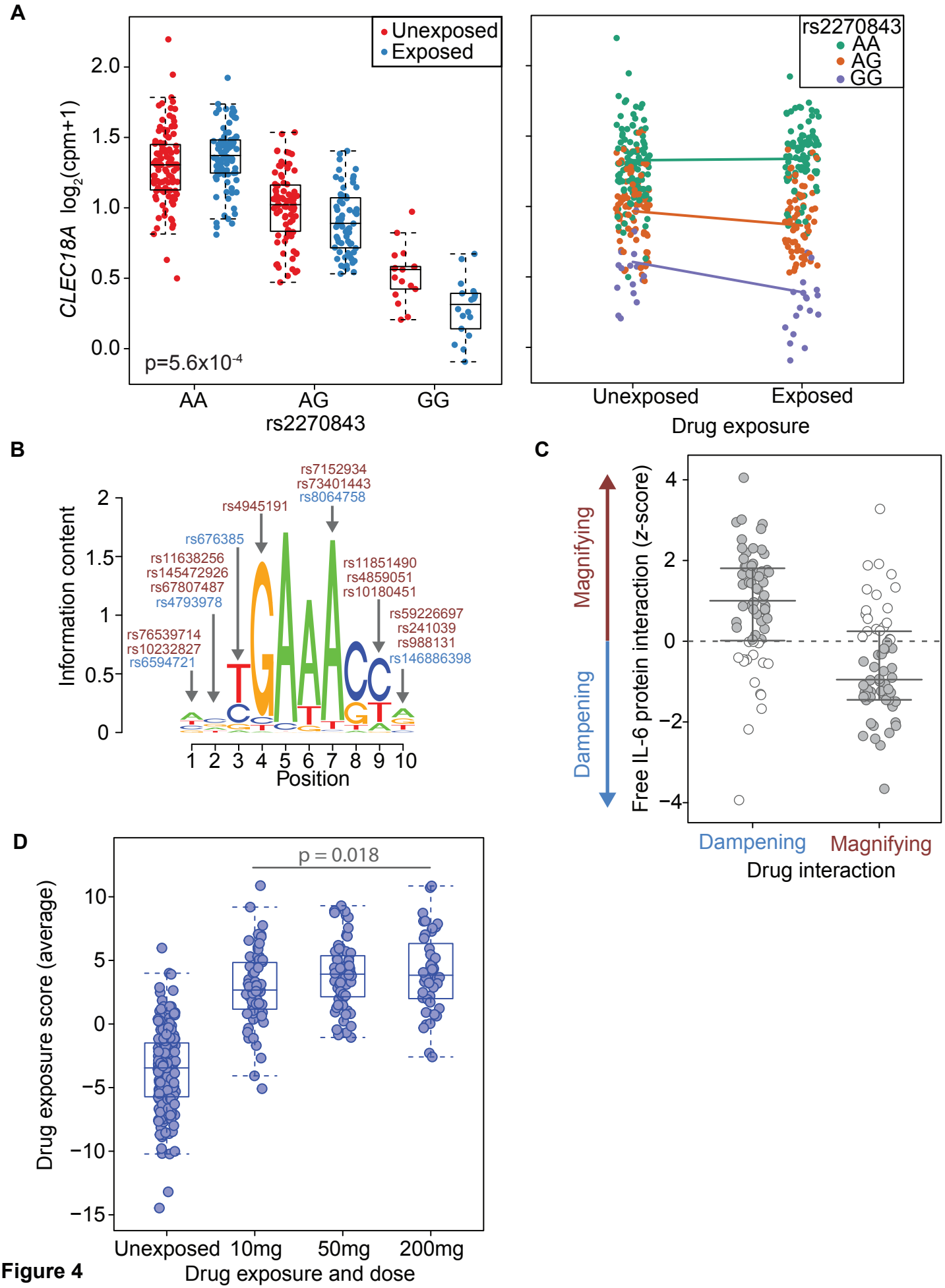
A
Unexposed
Exposed
CLEC18A log2(cpm+1)
p=5.6x10-4
AA
AG
GG
rs2270843
rs2270843
AA
AG
GG
Unexposed
Exposed
Drug exposure
B
Information content
rs7152934
rs73401443
rs8064758
rs4945191
rs676385
rs11638256
rs145472926
rs67807487
rs4793978
rs11851490
rs4859051
rs10180451
rs59226697
rs241039
rs988131
rs146886398
rs76539714
rs10232827
rs6594721
Position
C
Magnifying
Dampening
Free IL-6 protein interaction (z-score)
Dampening
Magnifying
Drug interaction
D
p = 0.018
Drug exposure score (average)
Unexposed
10mg
50mg
200mg
Drug exposure and dose

Figure 4

Figure 4. eQTL interactions with drug exposure (**A**) Drug exposure interaction with the *CLEC18A* eQTL plotted with respect to rs2270843 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 consistent direction). (**D**) Drug exposure score calculated from 126 drug-eQTL interactions.