Identification of drug eQTL interactions from repeat transcriptional and environmental measurements in a lupus clinical trial

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]

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

We hypothesize that regulatory mechanisms influenced by an environmental perturbagen may be identified with eQTL (expression quantitative trait locus) interactions, which alter the relationship between a genetic variant and transcript levels. In an anti-IL-6 clinical trial of 157 patients with systemic lupus erythematosus (SLE) we measured cell counts, interferon (IFN) signature, and drug exposure at three time points alongside genome-wide transcription. Repeated transcriptomic measurements detected 4,976 *cis* eQTLs, 63% more than detectable from single measurements. We identified 154, 185 and 126 nominal eQTL interactions with T cell count, IFN status and anti-IL-6 drug exposure respectively (more than expected by chance, p<0.001). IFN interactions are enriched for IRF1 motifs, and 91/126 drug-eQTL interactions are consistent with interactions using free IL-6 protein levels. This same approach can be easily used to define informative drug exposure scores, and can be applied to larger drug trials to further our understanding of the drug mechanisms.

A *cis* eQTL contains a genetic variant that alters expression of a nearby gene. eQTLs are ubiquitous across the genome[1] and while most are stable across tissues and conditions, environmental perturbations can alter a minority 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 an environmental interaction. Therefore, if we observe a set of eQTL interactions due to a perturbagen (such as a drug), shared upstream regulatory mechanisms may represent key affected pathways. As the regulatory genome is mapped[9,10], this approach to define environmental mechanisms will become more potent.

A clinical trial is an ideal cohort to detect eQTL interactions with drugs or other environmental perturbations and could lead to key insights about mechanisms of action. In a clinical trial, homogenous groups of subjects are randomly exposed to a drug in a structured study design that frequently includes biological samples collected at different time points with extensive clinical and physiological phenotyping. If RNA is queried at multiple time points under different exposure states, we speculated that repeat measurements could increase power to detect not only eQTLs, but also their interactions.

As a proof of principle, we looked at the modulation of eQTL effects by anti-IL-6 exposure using data from a phase II clinical trial to evaluate the safety and efficacy of a neutralizing IL-6 monoclonal antibody (PF-04236921) in 157 SLE patients[11] (**Online Methods**). Here, our clinical outcome of interest is a reduction in free IL-6 protein levels. We conducted whole blood high-depth RNA-seq profiling at 0, 12, and 24 weeks in anti-IL-6 exposed and unexposed individuals with the Illumina TruSeq protocol. We observed and quantified 20,253 gene features and genotyped 608,017 variants genomewide (**Online Methods**). Along with each RNA-seq assay, we documented drug exposure and quantified cell counts with FACS and IFN signature status with real-time PCR.

We first mapped *cis* eQTLs (SNPs within 250kb of the transcription start site of the gene) and then tested those eQTLs for interactions with cell counts, IFN status and drug

exposure. eQTL interactions can be explored using our interactive visualization tool (http://baohongz.github.io/Lupus_eQTL).

We used a linear mixed model, including repeat measurements with up to three RNA-seq assays per patient (**Fig. 1a,** 379 samples from 157 patients, **Online Methods**). To maximize power, we adjusted for 5 population and 25 gene expression principal components (**Online Methods**). To ensure a set of highly confident eQTLs, we applied a stringent multiple hypothesis testing correction and identified 4,976 *cis* eQTL genes with $p<2.3x10^{-8}$ (0.05/2,177,889 tests, **Fig. 1b,1c, Supplementary Table 1**). We observed that repeat samples increased our power by detecting 63% more *cis* eQTLs compared to using a single sample per individual (3,061 genes **Fig. 1b**). Our results are highly concordant with the BIOS cohort, a much larger dataset of 2,166 healthy individuals[1]. 85.4% of our SLE eQTL SNP-gene pairs are reported as eQTLs in BIOS (FDR<0.05). Of these, 99.1% showed consistent direction of effect (**Fig. 1d**). For each of the 4,976 *cis* eQTL genes, we tested the most significantly associated SNP for environmental interactions.

We first tested a type of eQTL interactions that have been examined previously: cell counts[12,13]. We obtained FACS data for 320 samples for which we had contemporaneous RNA-seq profiles (n=152 subjects). We determined the percentage of total lymphocytes that were T cells by gating (**Fig. 2a**). We found 154 T cell-eQTL interactions with nominal evidence (p<0.01, **Supplementary Table 1**), whereas from 4,976 tests we would expect ~50 from chance alone. To ensure that our statistics were not inflated, we conducted 1,000 stringent permutations, where we reassigned T cell

percentages across samples and retested. This permutation preserved the main eQTL effect, while disrupting interactions that might be present in the data. In no instance did we observe 154 or more interactions at $p<0.01$ (maximum=133), suggesting that the number of observed interactions is highly unlikely to have happened by chance (**Fig. 2b**, p=1/1,001 = <0.001).

Interactions can be divided into magnifiers, where environmental exposure and eQTL effects are in the same direction, and dampeners where the effects work in the opposite direction (**Supplementary Fig. 1**, **Fig. 2c**). The *NOD2* rs1981760 eQTL is an example of an interaction that is dampened by T cell count (**Fig. 2d**, $p=6.5\times10^{-5}$), and has separately been shown to vary across cell types[12,13]. For each percent increase in T cell proportion, the eQTL effect is reduced by 0.008 $\log_2$(cpm+1).

Many patients with SLE exhibit high levels of genes induced by IFN alpha; these genes, known as the IFN signature, are a marker of disease severity[14,15]. We explored the influence of IFN alpha on gene regulation after determining the IFN status of every patient at each time point using real-time PCR of 11 IFN-inducible genes[16] (**Online Methods, Fig. 3a**). We identified 185 IFN-eQTL interactions with $p<0.01$ (**Supplementary Table 1**). Following the same permutation procedures as above, our observed interactions were unlikely to have occurred by chance (maximum permutation interactions=112, $p<0.001$) (**Supplementary Fig. 2**). We note that interactions with a proxy gene for IFN status have been described[1] and we find overlap of genes with those reported interactions (**Supplementary Fig. 3**). For example, *SLFN5* expression is

influenced by the rs12945522 SNP ($1.3x10^{-10}$, **Fig. 3b**). This effect is magnified in IFN high samples.

To define transcription factors that drive the response to IFN alpha, we sought to identify motifs that explain the differences between magnifier and dampener eQTLs (**Supplementary Fig. 4**). We applied HOMER[17] to assess overlap with the interaction SNPs (or SNPs in high linkage disequilibrium (LD, r^2>0.8)) in the *cis* window (**Online Methods**). We found significant enrichment of motifs for key transcription factors involved in IFN signaling including the IRF1 motif in eQTLs dampened in IFN low individuals (p=0.001, permutation p<0.002, **Online Methods, Fig. 3c**, **Supplementary Table 2**). An example is the *GTF2A2* rs2306355 eQTL ($p=8.7x10^{-3}$, **Fig. 3d**); rs2306355 is in tight LD (r^2=0.82 in Europeans) with rs6494127, which interrupts the GAAA core of the IRF1 motif (**Fig. 3c**), and likely disrupts IRF1 binding[18]. We observe greater expression in individuals with the rs2306355 A allele compared to G; this difference is dampened in IFN low individuals (**Fig. 3d**).

We then examined whether IL-6 blockade alters the relationship between genomic variation and gene expression and induces drug-eQTL interactions. We observed 126 drug-eQTL interactions with p<0.01 (**Supplementary Table 1**). Following the same permutation strategy as above, we found a median of 77 interactions with p<0.01 (maximum=117) from 1,000 permutations. This suggests that about half of our drug-eQTL interactions likely represent real biological phenomena, and not statistical artifact (**Supplementary Fig. 5**). These drug-eQTL interactions showed little overlap with the interactions observed for T cell count or IFN status (**Supplementary Fig. 6**). Again,

these interactions can be divided into magnifiers and dampeners (**Supplementary Fig. 7**). **Figure 4a** highlights one of the most significant drug-eQTL interactions ($p=6.5x10^{-4}$) for the gene *KIAA2013*. Of particular biological relevance is an interaction with *IL10*, an anti-inflammatory cytokine (**Supplementary Fig. 8**).

A more common strategy to determine the effect of a perturbagen is to use differential gene expression. For differential expression following drug treatment, we identified 1,161 genes with nominal statistical evidence ($p<0.01$) but modest effects (max fold change=1.3, **Supplementary Fig. 9**). Furthermore, only 8/126 drug-eQTL interaction genes also show evidence of differential gene expression. This suggests that eQTL interactions offer independent information from differential expression, which might contribute to defining mechanisms.

To validate these interactions, we hypothesized that interactions due to drug exposure are likely driven by free IL-6 cytokine levels (our key clinical biomarker of interest). If this is the case, for eQTLs dampened by drug exposure, an increase in free IL-6 should elicit an opposite interaction effect and result in eQTL magnification. We assessed whether eQTL interactions with free IL-6 protein levels measured in the patient serum samples were consistent with those following IL-6 blockade. We observed enrichment in the overlap between cytokine interactions and drug interactions (91/126 interactions in consistent direction, **Figure 4b,** $p=3.2x10^{-7}$, binomial test). We were concerned that free IL-6 and drug exposure are not independent in this dataset, and that this concordance might be in part due to the connection between free IL-6 protein levels and IL-6 blockade (**Supplementary Fig. 10)**. To assess whether free IL-6 offers independent

interaction effects that were consistent after accounting for drug effect, we modeled free IL-6 levels based on the presence or absence of drug and then assessed interactions with residual IL-6 levels. Again, we observed a significant number of interactions in a consistent direction (p=0.03 **Supplementary Fig. 11)**.

For many biologic medications, predictive pharmacogenetics has been challenging; for example, studies to define genetic or non-genetic biomarkers of anti-TNF response have not been successful[19,20]. Identifying relevant SNP-gene pairs from eQTL interactions could offer an alternative molecular strategy to assess effective drug response in patients. We speculated that these drug-eQTL interactions could be used in a clinical pharmacogenetic context to assess effective drug exposure for patients.

We defined a simple drug exposure score using the 126 drug-eQTL interactions (**Online Methods**). For each RNA-seq sample, we assessed whether the expression of the interaction gene was more consistent with the drug exposed or unexposed state for the corresponding interaction SNP genotype. Samples more consistent with the drug-exposed state are assigned a larger drug exposure score. Unsurprisingly, we found a difference in drug exposure score between the unexposed and exposed samples (**Supplementary Fig. 12**) (r_s=0.79, p=6.9x10^{-81}); these differences reflect the fact that the eQTLs were themselves identified by examining samples with and without drug exposure. However, while we did not utilize actual drug dose to identify drug-eQTL interactions, we found a significant correlation between drug dose (10, 50 or 200mg) and drug exposure score (r_s=0.16, p=0.018) in the drug-exposed samples (**Fig. 4c**).

Hence scores based on eQTL interactions might reflect the biological activity that a medication is having upon an individual, and may be modeling an effective medication activity level. Such a scoring system could be implemented easily in most phase III trials, where the numbers of samples are far in excess of this phase II trial, ensuring better powered and more accurate eQTL-interaction mapping. Such a strategy might be even more effective if appropriate cell-types or affected tissues were queried and could be modeled in conjunction with differential expression to obtain more power.

Defining the downstream mechanism influenced by a drug is critical and might be a path to classifying individuals as responders or non-responders, understanding off-target effects, and finding more accurate biomarkers. Clinical trial data offers an excellent opportunity to identify mechanisms by examining eQTL effects and their environmental interactions. In this relatively modest phase II trial, we have identified eQTL interactions with clinical and physiological data, but more importantly, with drug exposure. It will become possible to connect interaction eQTLs as the specific transcription factors driving those interactions are defined by high throughput techniques such as ChIP-seq.

Online methods

Patient recruitment

SLE patients were recruited to a phase II clinical trial to test the efficacy and safety of an IL-6 monoclonal antibody (PF-04236921). The patient population recruited to this trial have been detailed extensively by Wallace et al[11]. 183 patients (forming a multiethnic cohort) were randomized to receive three doses of drug (10, 50 or 200mg) or placebo at three time points during the trial (weeks 0, 8 and 16).

RNA-sequencing

Peripheral venous blood samples were collected in PAXgene Blood RNA tubes (PreAnalytiX GmbH, BD Biosciences) for high-depth RNA-seq profiling at 0, 12, and 24 weeks. Total RNA was extracted from blood samples using the PAXgene Blood RNA kit (Qiagen) at a contract lab using a customized automation method. The yield and quality of the isolated RNA were assessed using Quant-iT™ RiboGreen® RNA Assay Kit (ThermoFisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. Following quality assessment, the RNA was sent to another contract lab for RNA sequencing: an aliquot of 500-1000 ng of each RNA was processed with a GlobinClear kit, Human (ThermoFisher Scientific) to remove globin mRNA. After globin mRNA depletion, the RNA samples were converted to cDNA libraries using TruSeq RNA Sample Prep Kit v2 (Illumina) and sequenced using Illumina HiSeq 2000 sequencers. An average of 40M 100bp pair-end reads were generated per sample for downstream analysis.

468 RNA-seq profiles were generated from 180 patients. Data were aligned to the reference genome and gene expression quantified using Subread[21] and featureCounts[22] respectively. The genes with mapped reads were filtered so only those with at least 10 reads (CPM>0.38) in at least 32 samples (minimum number of patients with both unexposed and exposed RNA-seq assays in a drug group) were retained prior to normalization. 20,253 transcripts were normalized using the trimmed mean of M-values method and the edgeR R package[23]. Following quality control (QC), 4 samples were removed as outliers. Expression levels are presented as $\log_2$(cpm +1).

Genotyping

160 individuals were genotyped across 964,193 variants genome-wide with the Illumina HumanOmniExpressExome-8v1.2 beadchip. SNPs were removed if they deviated from Hardy-Weinberg Equilibrium ($p < 1x10^{-7}$), had a minor allele frequency <5%, missingness >2% or a heterozygosity rate greater than 3 standard deviations from the mean (PLINK[24,25]). For mapping eQTLs, SNPs on the Y chromosome were removed. Following QC, 608,017 variants were used for further analysis. One sample had high missingness and was an outlier for heterozygosity rate so was removed from further analysis.

Cell counts

Blood samples were collected for cytometry analysis at weeks 0, 12 and 24. Samples were subjected to flow cytometry for T cell immunophenotyping. T cells (CD3+) were

counted as a percentage of lymphocytes (CD45+). FACS data were available for 320 samples from 152 subjects.

Interferon status

The IFN status was classified using the expression of IFN response genes at each time point. The expression of 11 genes (*HERC5*, *IFI27*, *IRF7*, *ISG15*, *LY6E*, *MX1*, *OAS2*, *OAS3*, *RSAD2*, *USP18*, *GBP5*)[16] was measured using TaqMan Low Density Arrays. Samples were classified as high or low IFN based on the first PCA score for the expression of these genes (**Fig. 3a**). IFN status was available for 376 samples from 157 subjects.

Drug exposure

Samples were assigned as unexposed (placebo or week 0 samples) or drug exposed (week 12 and week 24 samples in the drug groups). When included in the model, unexposed samples were assigned as 0 and exposed samples as 1.

Free IL-6 protein levels

Free IL-6 protein levels were determined from serum using a commercial sandwich ELISA selected for binding only free IL-6. The assay was validated according to FDA biomarker and fit-for purpose guidelines. Free IL-6 protein levels were available for 311 samples from 145 subjects. Samples were ranked in order of IL-6 protein levels and included in the model to identify cytokine-eQTL interactions.

Statistical analysis

eQTL and interaction analysis

157 patients (with 379 RNA-seq samples) had good quality gene expression and genotyping data for eQTL analysis. All statistical analyses were carried out in R[26].

We defined a *cis* eQTL as the SNP within 250kb either side of the GENCODE[27] transcription start site of the gene. We first applied a linear model for the first available time point to identify each eQTL using the first 25 principal components of gene expression and the first 5 principal components of genotyping as covariates. SNPs were encoded as 0, 1 and 2. To adjust for multiple testing during eQTL discovery we used a corrected *p*-value threshold of 2.3×10^{-8} (0.05/ 2,177,889 tests).

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 $p<2.3x10^{-8}$) 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:

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

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, β_{drug} is the drug effect (differential gene expression) and β_x is the interaction effect.

A *p*-value for the interaction term was determined with a likelihood ratio test.

To confirm the relative enrichment of eQTL interactions, we shuffled the interaction covariate (for example drug exposure) 1,000 times and calculated the number of

significant interactions observed in each permutation. For T cell counts and IFN high/low status, we shuffled across all samples. For drug interaction permutation analysis, we maintained the number of individuals in the drug group and the number of samples with exposure to drug.

Concordance with an eQTL study in healthy individuals

4,976 genes were classified as having a *cis* eQTL in the SLE cohort ($p<2.3x10^{-8}$). The *z*-score for the most associated SNP for each of these genes was compared to the *z*-score from a previously published eQTL dataset from whole blood from 2,166 healthy individuals[1]. 4,250/4976 SNP-gene pairs (85.4%) were also reported in the BIOS dataset (FDR<0.05). After removing 60 SNPs, which could not be mapped to a strand 4,154/4,190 (99.1%) had a *z*-score (eQTL effect) in a consistent direction.

Magnifiers and Dampeners

An eQTL interaction can either magnify or dampen the original eQTL effect. We multiplied the interaction *z*-score by the sign of the original eQTL effect (genotype beta) and defined magnifiers as interactions with an adjusted *z*-score > 0 and dampeners as interactions with an adjusted *z*-score < 0.

Differential gene expression analysis

To identify differentially expressed genes following drug exposure, 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.

Conditional analysis for IL-6 protein levels

We modeled the relationship between free IL-6 protein levels and drug exposure using a linear model. We used the residuals from this model in our interaction linear mixed model to identify IL-6 protein interactions independent of drug exposure.

Drug exposure score

We used linear discriminant analysis to assign a drug exposure score for each sample. A score was calculated for each gene (see equation below) and then the final drug exposure score is the average across the 126 drug-eQTL genes.

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

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

HOMER analysis for transcription factor binding motif enrichment

We used the HOMER software suite[17] to look for enrichment of transcription factor binding motifs in the 185 IFN eQTL interactions ($p<0.01$). Each eQTL interaction was identified using the most highly associated SNP for that eQTL. However, as this SNP is

not necessarily the functional SNP, we additionally considered all those with an $r^2 \geq 0.8$ in the 1000 Genomes European population[28] within 250kB of the transcription start site of the gene. We defined our motif search window as 20 bp on either side of each SNP (i.e. 41 bp wide).

We divided the eQTL interactions into magnifiers or dampeners and conducted two separate HOMER analyses: one with magnifiers in the foreground and dampeners in the background; the other with dampeners in the foreground and magnifiers in the background. HOMER reported the transcription factor motifs that were significantly enriched in the foreground relative to background. Motifs were plotted using the SeqLogo R library[29].

A permutation p value for enrichment of the IRF1 transcription factor binding motif in IFN dampeners was determined as follows. The motif is interrupted by interaction SNPs (or SNPs in LD) corresponding to 11 dampening genes and 2 magnifying genes. We permuted which genes were labeled as magnifiers or dampeners 10,000 times and counted the number of genes in each category with an IRF1 motif interrupted. We found 18 occurrences from 10,000 trials with at least 11 dampening genes ($p<0.002$).

Acknowledgements

This work is supported in part by funding from the National Institutes of Health (U01GM092691, UH2AR067677, U19AI111224 (SR)) and the Doris Duke Charitable Foundation Grant #2013097. 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.

References

1. Zhernakova, D. V *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49,** 139–145 (2017).
2. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343,** 1246949 (2014).
3. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* **44,** 502–510 (2012).
4. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. *Science* **344,** 519–523 (2014).
5. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
6. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7,** e1002003 (2011).
7. Kukurba, K. R. *et al.* Impact of the X chromosome and sex on regulatory variation. *Genome Res.* **26,** 768–777 (2016).
8. Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47,** 88–91 (2015).
9. ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
10. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
11. Wallace, D. J. *et al.* Efficacy and safety of an interleukin 6 monoclonal antibody for the treatment of systemic lupus erythematosus: a phase II dose-ranging randomised controlled trial. *Ann. Rheum. Dis.* **76,** 534–542 (2017).
12. Westra, H.-J. *et al.* Cell Specific eQTL Analysis without Sorting Cells. *PLoS Genet.* **11,** e1005223 (2015).
13. Naranbhai, V. *et al.* Genomic modulators of gene expression in human neutrophils. *Nat. Commun.* **6,** 7545 (2015).
14. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2610–2615 (2003).
15. Bennett, L. *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J. Exp. Med.* **197,** 711–723 (2003).
16. Hill, A. A. *et al.* FRI0003 Determination of interferon (IFN) signatures for sle patients may be critical for optimal treatment selection but depends on the genes chosen: report from the bold (biomarkers of lupus disease) study. *Ann. Rheum. Dis.* **72,** A369.3-A370 (2013).
17. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38,** 576–589 (2010).
18. Escalante, C. R., Yie, J., Thanos, D. & Aggarwal, A. K. Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature* **391,** 103–106 (1998).
19. Cui, J. *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.* **9,** e1003394 (2013).
20. Cui, J. *et al.* The role of rare protein-coding variants to anti-TNF treatment response in rheumatoid arthritis. *Arthritis Rheumatol.* (2016). doi:10.1002/art.39966
21. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108–e108 (2013).
22. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
23. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
24. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
25. Purcell, S. & Chang, C. PLINK 1.9. at <https://www.cog-genomics.org/plink2>
26. R Core Team. R: A Language and Environment for Statistical Computing. (2015). at <https://www.r-project.org>
27. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).
28. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
29. Bembom, O. seqLogo: Sequence logos for DNA sequence alignments. (2016).

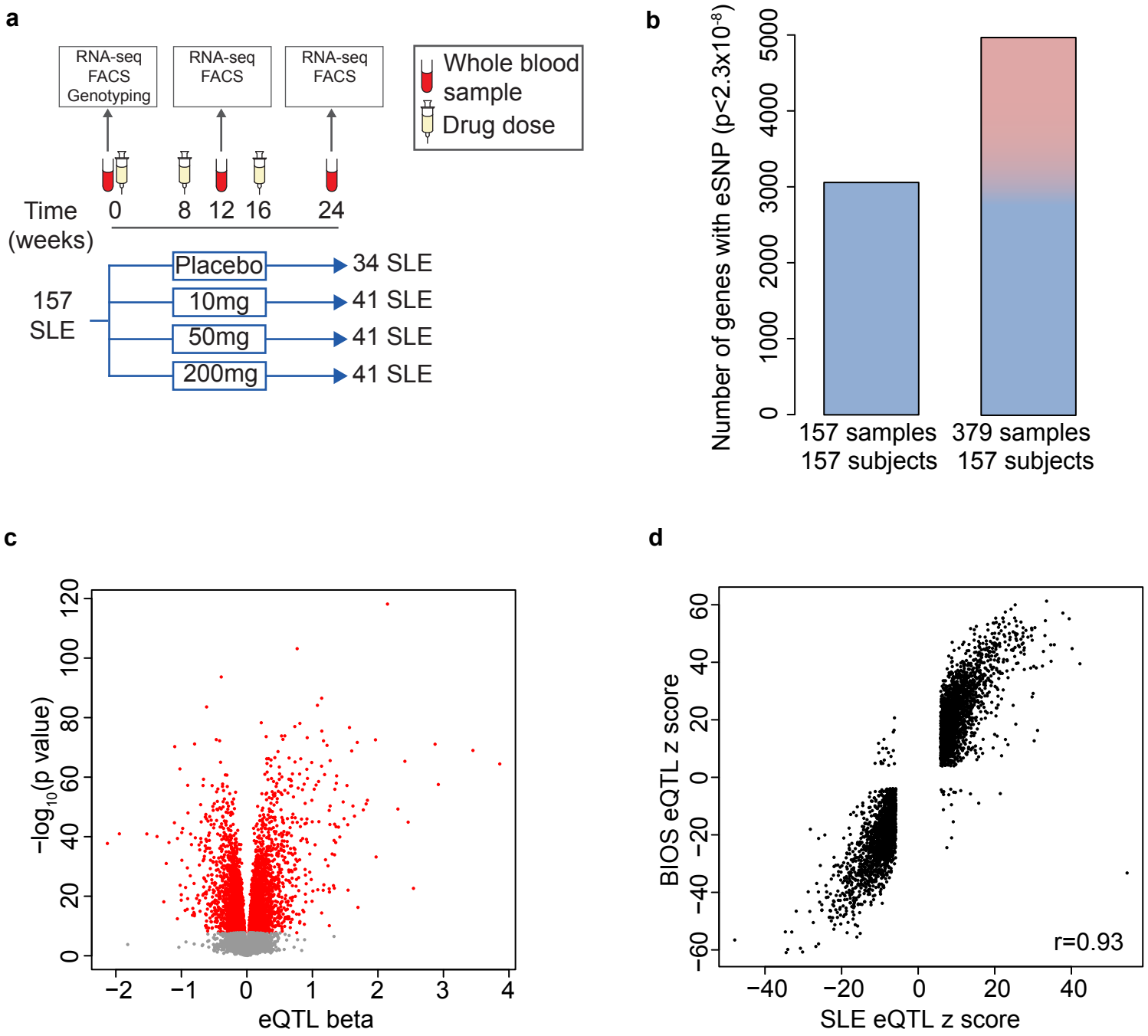

Figure 1. Identifying eQTLs in SLE patients (**a**) Clinical trial structure and sampling strategy. (**b**) Number of eQTL genes identified using a linear model (left) and a linear mixed model (right). (**c**) Volcano plot of eQTL effects for the most significantly associated SNP for each gene (red color indicates $p<2.3\times10^{-8}$) (**d**) Correlation of SLE eQTL effects ($p<2.3\times10^{-8}$) 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.

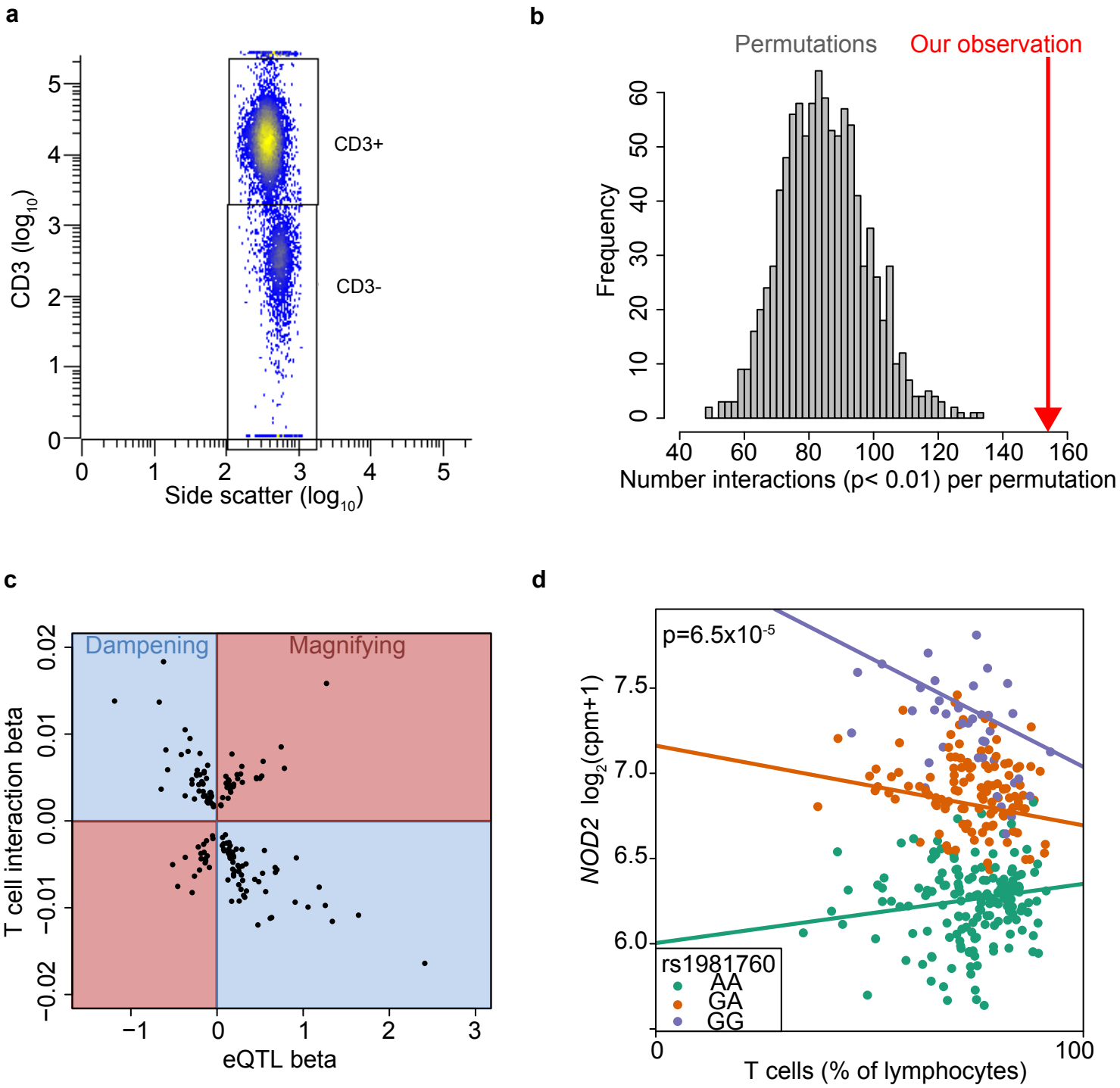

Figure 2. eQTL interactions with T cell percentages (**a**) Flow cytometry gating strategy to determine T cells as a percentage of lymphocytes. (**b**) Number of significant interactions (p<0.01) from 1,000 permutations of T cell counts (median=85). (**c**) Effect of the T cell interaction on the original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

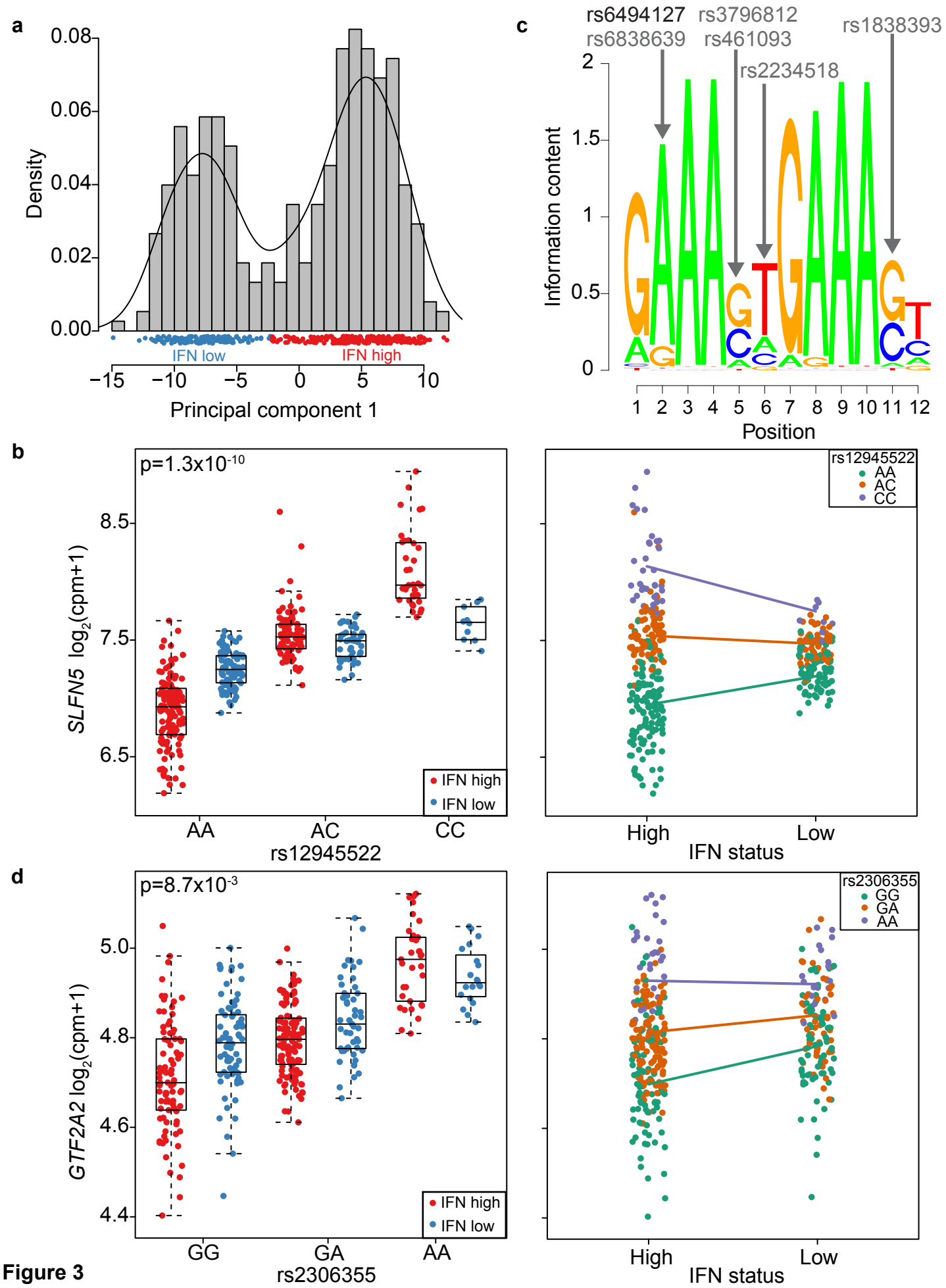
a
Density
0.08
0.06
0.04
0.02
0.00
IFN low
IFN high
−15
−10
−5
0
5
10
Principal component 1
c
rs6494127
rs6838639
rs3796812
rs461093
rs2234518
rs1838393
Information content
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12
Position
b
p=1.3x10-10
SLFN5 log2(cpm+1)
8.5
7.5
6.5
IFN high
IFN low
AA
AC
CC
rs12945522
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 *SLNF5* 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). (**d**) IFN status interaction with the *GTF2A2* eQTL plotted with respect to rs2306355 genotype (left) and IFN status of the sample (right).

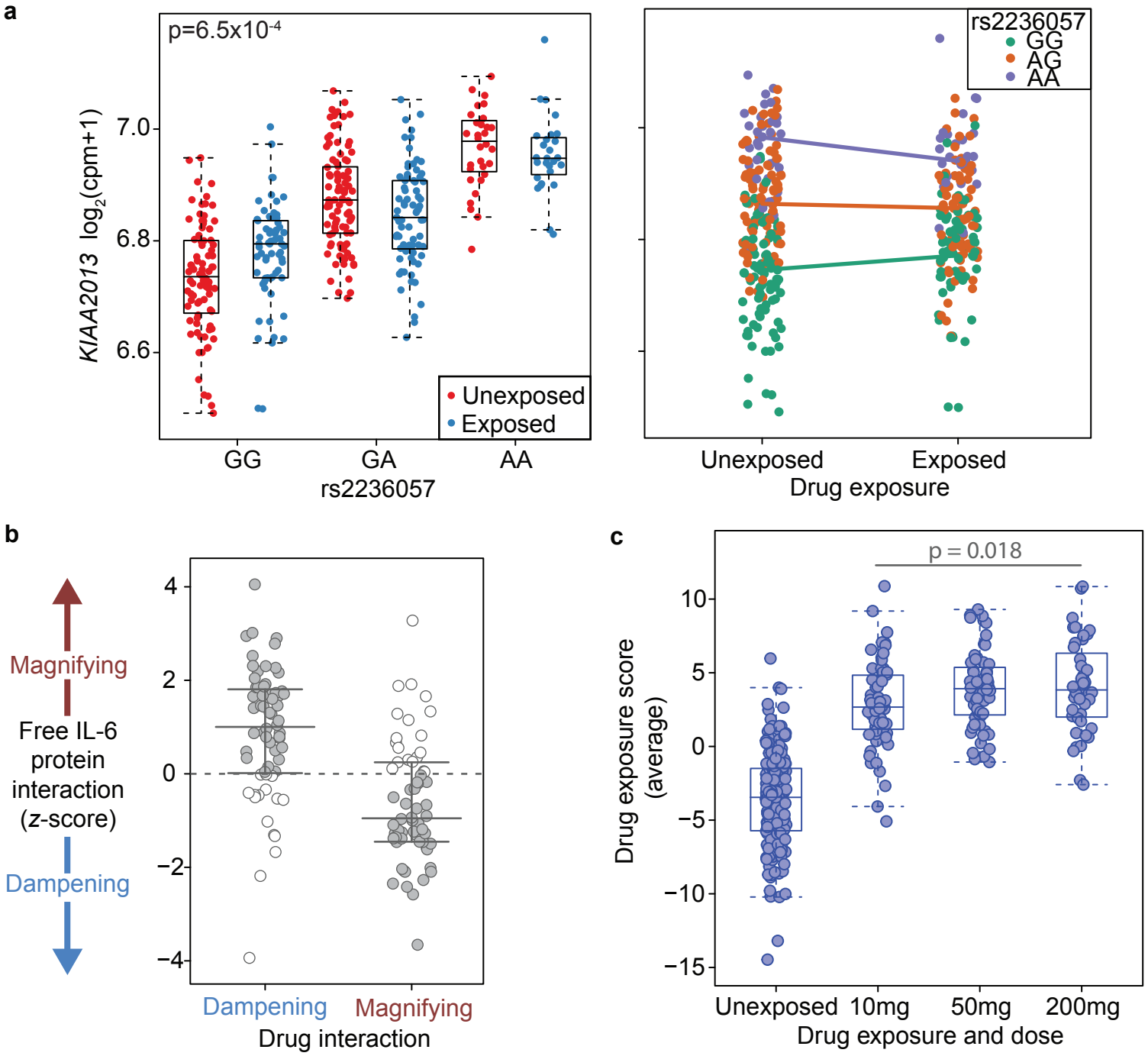

Figure 4. eQTL interactions with drug exposure (**a**) Drug exposure interaction with the *KIAA2013* eQTL plotted with respect to rs2236057 genotype (left) and drug exposure (right). (**b**) Concordance of free IL-6 protein interaction effects with drug exposure interaction effects (grey indicates consistent direction). (**c**) Drug exposure score calculated from 126 drug-eQTL interactions.