Title: Glucocorticoids Unmask Silent Non-Coding Genetic Risk Variants for Common Diseases

One Sentence Summary:

Genomic and epigenomic fine-mapping of ligand-dependent genetic variants unmasks novel disease risk genes

Authors:

Thanh Thanh Le Nguyen[1†], Huanyao Gao[1†], Duan Liu[1†], Zhenqing Ye[3], Jeong-Heon Lee[2], Geng-xian Shi[2], Kaleigh Copenhaver[1], Lingxin Zhang[1], Lixuan Wei[1], Jia Yu[1], Cheng Zhang[1], Hu Li[1], Liewei Wang[1], Tamas Ordog[2,4,5], Richard M. Weinshilboum[1]*

Affiliations:

[1]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic; Rochester, MN, USA.

[2]Epigenomics Program, Center for Individualized Medicine, Mayo Clinic; Rochester, MN, USA.

[3]Department of Health Sciences Research, Mayo Clinic; Rochester, MN, USA.

[4]Department of Physiology and Biomedical Engineering, Mayo Clinic; Rochester, MN, USA.

[5]Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic; Rochester, MN, USA.

[†]These authors contributed equally to this work

*Correspondence to: Weinshilboum.Richard@mayo.edu

Abstract:

Understanding the function of non-coding genetic variants represents a formidable challenge for biomedicine. We previously identified genetic variants that influence gene expression only after exposure to a hormone or drug. Using glucocorticoid signaling as a model system, we have now demonstrated, in a genome-wide manner, that exposure to glucocorticoids triggered disease risk variants with previously unclear function to influence the expression of genes involved in autoimmunity, metabolic and mood disorders, osteoporosis and cancer. Integrating a series of genomic and epigenomic assays, we identified the *cis*-regulatory elements and 3-dimensional interactions underlying the ligand-dependent associations between those genetic variants and distant risk genes. These observations increase our understanding of mechanisms of non-coding genetic variant-chemical environment interactions and advance the fine-mapping of disease risk and pharmacogenomic loci.

Main Text:

Many genetic sequence variants associated with human disease have been discovered, but the task of understanding the functions of those variants remains challenging since most of them map to non-coding regions of the genome (*1*). One approach to address this challenge has been to associate these variants with gene transcription, identifying so-called expression quantitative trait loci (eQTLs). While the multitissue Genotype-Tissue Expression (GTEx) project (*2*) focuses on characterizing variants at steady-state conditions, it is increasingly clear that eQTLs can be dynamic. For example, eQTLs can be modulated by pathogens (*3*) or cellular differentiation (*4*) (**Fig. 1A**). However, much less is known regarding sequence variants that interact with the chemical environment to influence gene expression variation for either disease risk or drug response. We (*5-7*) and others (*8*) recently observed a series of uniquely "pharmacologic" eQTLs, referred to here as "pharmacogenomic (PGx)-eQTLs", for which eQTL behavior is elicited or significantly amplified in the presence of a drug or hormone. In the present study, we have moved beyond "single variants and single genes" and beyond association studies to mechanistically interrogate PGx-eQTLs genome-wide by generating and integrating "pharmaco-omic" datasets. Using an important pharmacological target, the glucocorticoid receptor (GR), as a model, we uncovered a series of novel insights into how glucocorticoids might interact with genetic risk variants in predisposing individuals to a wide range of pathology involving inflammation and immunity, bone disease, metabolic disorder, psychiatric disorder and cancer.

Our study design (see **Fig. 1B, Materials & Methods**) began with the identification of PGx-eQTLs using genome-wide single-nucleotide polymorphisms (SNPs) and RNA-sequencing **(Figs. S1A, Supplementary Text)** before and after exposure to glucocorticoids in immortalized human lymphoblastoid cell lines (LCLs) from 30 individuals (**Table S1**). We focused on SNPs that mapped within or near (+/-500bp) a GR binding site (**Figs. S1B-E, Supplementary Text**),

suggesting possible direct interference with GR signaling. To identify and confirm drug-dependent effects, we treated the cells with cortisol, an endogenous GR agonist, and CORT108297 (C297), a selective GR modulator which, in our observation, displayed antagonist properties against cortisol and partial agonist activity by itself. A series of epigenomic techniques (see **Fig. 1B under "Mechanisms"**) including integrative chromatin state prediction (ChromHMM), a massively parallel enhancer reporter assay (STARR-seq), and a 3D chromatin conformation capture assay (H3K27ac HiChIP), were then performed to determine mechanism(s) underlying the observed PGx-eQTLs. We then also identified which disease risk variants might behave as PGx-eQTLs using data from publicly available genome-wide and phenome-wide association studies (GWAS and PheWAS). **Figures S2-3** show the positive controls for these "-omic" datasets worked as expected**.**

Twenty-five percent of the over 100 glucocorticoid-dependent PGx-eQTLs that we identified, as discussed in detail below, had already been associated with clinical phenotypes but usually without a clear mechanism based on the results of previous GWA/PheWA studies. For example, a mechanistically unexplained variant that had been associated with breast cancer risk, rs1697139 (*9*), was a cortisol-dependent PGx-eQTL for the Microtubule Associated Serine/Threonine Kinase Family Member 4 (*MAST4*) gene. Specifically, *MAST4* expression was repressed by cortisol in subjects with the G/G but not the A/A genotype, and that repression was reversed after antagonist treatment (**Fig. 2A**). This SNP, in a genotype-dependent fashion, modulated a GR-responsive intergenic enhancer (**Fig. S4A**) that "looped" over 40,000 base pairs to *MAST4*, transcriptionally regulating that gene (**Fig. 2B)**. Interestingly, in breast cancer cell lines, GR bound to the same rs1697139 locus (**Fig. S4B-C**) and dramatically repressed *MAST4* expression in MDA-MB-231 cells, a triple-negative breast cancer cell line, as well as cell lines for other breast cancer subtypes (**Fig. 2C, S5**). Although the function of *MAST4* in breast cancer

is unknown, it may be a novel risk gene since its expression was significantly repressed in breast tumor tissue when compared with normal tissue (**Fig**. **2D)**. Furthermore, its expression appeared to be a predictor of treatment response since decreased expression of *MAST4* was associated with decreased relapse-free survival (**Fig. 2E**). These observations agree with the fact that the rs1697139-G/G genotype was associated with increased risk for breast cancer since it was associated with decreased *MAST4* expression after exposure to cortisol, a hormone that promotes breast cancer heterogeneity and metastasis (*10*). In another example involving the rs11678116 SNP and the Lipid-Droplet Associated Hydrolase (*LDAH*) gene (**Fig. S6A**), cortisol was shown to bring together SNP-gene loci that otherwise would not have been in proximity (**Fig. S6C**), "unmasking" the impact of rs11678116 on *LDAH* transcription. Specifically, the rs11678116 SNP created a GR binding motif (**Data S1**), which led to decreased cortisol-responsive enhancer activity of the T/T genotype (**Fig. S6B**) and lower *LDAH* expression after cortisol treatment of subjects with the T/T but not the G/G genotype. Based on PheWAS results, rs11678116 is associated with migraine, providing an intriguing genotype-phenotype link for functional investigation given the elevated levels of cholesterol and triglycerides observed in migraine patients (*11*).

Beyond these illustrative examples, **Table 1** lists a total of 33 disease risk loci that we found to behave as PGx-eQTLs. These loci span many different disease and drug response categories involving adverse response to corticosteroids, inflammation and immunity, osteoporosis, neuropsychiatric disorders and cancer (see also **Data S1**). Of importance, glucocorticoids were either known risk factors or therapeutic agents used to treat most of these diseases (**Supplementary Text**), demonstrating a genetic risk × hormone/drug risk interaction in disease predisposition. These SNPs often affected relevant gene expression only after glucocorticoid exposure, providing functional explanations for the GWAS signals (**Table 1**). For

instance, SNPs previously associated with immune-related phenotypes were found to interact with cortisol to influence the expression of *RUNX1* (**Fig. S7A-D)**, a master regulator of hematopoiesis, which influences the generation of different types of immune cells (*12*), or *NLRC5*, a key regulator of adaptive immune responses (*13*). SNPs previously associated with obesity-related phenotypes were found to interact with cortisol to influence genes located as far as 150,000bp away such as *FBXL19*, an adipogenesis-controlling gene (*14*), *HSD3B7*, a cholesterol metabolizing enzyme (**Fig. S7E-F**), or *CADM1*, a gene that regulates body weight via neuronal modulation (*15*). These examples demonstrate that glucocorticoid-dependent PGx-eQTLs identified in LCLs uncovered functional SNPs related not only to immune-related diseases but also diseases reflecting dysfunctions of various other cell types. This is consistent with previous observations that eQTLs that are shared across tissues comprise a larger fraction of trait associations than do tissue-specific eQTLs (*16*).

Overall, we identified 102 cortisol-dependent and 32 C297-dependent *cis* PGx-eQTL SNP-gene pairs (**Figs. 3A-B, S8, S9**), the majority of which lost their eQTL property after C297 was introduced to antagonize cortisol, thus demonstrating the drug-dependent properties of this type of eQTL (**Fig. S9C)**. Furthermore, these PGx-eQTLs were not significant baseline eQTLs based on data from 174 LCLs in the GTEx database (**Fig. S9A-B**). We then sought to determine the molecular mechanisms of PGx-eQTLs. We found that the SNPs themselves could be located within known GR binding motifs, in tight linkage disequilibrium with SNPs within GR motifs (**Fig. S9D, Data S1**), or distant from GR motifs. However, in all cases, they still influenced GR-dependent transcriptional activity (**Data S1**). Using ChromHMM, a software that makes it possible to annotate the putative regulatory function of the noncoding genome using epigenomic information (*17*), we integrated 15 LCL epigenomic datasets from the Encyclopedia of DNA Element (ENCODE) portal (*18*) (**Data S2**) with our GR-targeted ChIP-seq data (**Fig. 3C, S10,**

Table S2). That effort demonstrated that GR-dependent PGx-eQTLs were enriched in a variety of chromatin states but predominantly in enhancers, with the primary site of enrichment being long-range enhancers that displayed promoter-looping properties (**Fig. 3D, Table S3, Data S1**). We then applied STARR-seq, a massively parallel reporter gene assay that can capture enhancer activity in a high-throughput fashion (*19*), to verify the effect of SNPs and drugs on transcriptional activity associated with the identified PGx-eQTLs (**Fig. S11, S12, S13, Supplementary Text**). As anticipated, the majority of allele- and drug-dependent loci identified by STARR-seq were ChromHMM-predicted enhancers. These loci displayed high consistency between the two cell lines in which STARR-seq was applied (**Fig. S12C)**, validating up to 81% of the cloned PGx loci that we had identified (**Fig. 3E, Data S1**). Finally, to determine the nature of physical interactions between PGx SNP loci and eQTL genes, we applied H3K27ac HiChIP, an assay that can capture chromatin conformation of enhancer-promoter and enhancer-enhancer interactions in a high-resolution manner (*20*), before and after cortisol exposure (**Fig. S14, Supplementary Text**). As expected, most identified SNPs that were associated with the eQTL genes with H3K27ac loops were ChromHMM-predicted active enhancers (**Fig. 3F**). Publicly available data (*20*) showed that 71 of the PGx-eQTLs that we identified displayed H3K27ac connecting "loops" at baseline. We observed a similar number of connected PGx-eQTLs before and after cortisol treatments in our data (78 and 76, respectively) (**Fig. 3F, Data S1)**, with 68 being constitutive loops connecting the PGx SNP loci and eQTL genes but requiring glucocorticoids to initiate transcriptional activity. In other cases, cortisol was shown to induce loops for 5 SNP-gene pairs and repress loops for 6, bringing together SNP-gene loci that otherwise would not have been in proximity or separating them (**Data S1**).

In summary, we have systematically interrogated genome-wide a type of dynamic eQTL, so-called "PGx-eQTLs" or genetic sequence variants with pharmacologically dependent effects

on gene transcription. Due to limited power, this study could not identify all possible relevant SNPs, but it does provide a comprehensive mechanistic framework and proof of concept for the study of ligand-dependent variants that influence transcription. By systematically fine-mapping genotype-phenotype interactions in which measurable environmental factors were taken into consideration, we uncovered novel risk genes for a wide range of diseases in which the pharmacologic or physiologic agents played a role. The function of these SNPs was usually unknown, i.e., it was "masked" in the absence of exposure to hormones or related drugs. Exposure to these compounds could either initiate transcriptional activity between connected loci or elicit a conformational change in the epigenomic landscape to disrupt or bring the SNP locus into contact with distal gene(s) that often were relevant to the observed clinical phenotypes. As a result, this series of studies has provided an in-depth investigation into the molecular signature of gene x hormone-drug environment interactions, which could be applicable to a broad range of diseases.

References and Notes

1. J. MacArthur *et al.*, The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
2. G. T. Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
3. M. N. Lee *et al.*, Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014).
4. B. J. Strober *et al.*, Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287-1290 (2019).
5. J. N. Ingle *et al.*, Selective estrogen receptor modulators and pharmacogenomic variation in ZNF423 regulation of BRCA1 expression: individualized breast cancer prevention. *Cancer Discov* **3**, 812-825 (2013).
6. D. R. Neavin *et al.*, Single Nucleotide Polymorphisms at a Distance from Aryl Hydrocarbon Receptor (AHR) Binding Sites Influence AHR Ligand-Dependent Gene Expression. *Drug Metab Dispos* **47**, 983-994 (2019).
7. D. Liu *et al.*, TCF7L2 lncRNA: a link between bipolar disorder and body mass index through glucocorticoid signaling. *Mol Psychiatry*, (2021).
8. J. Arloth *et al.*, Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. *Neuron* **86**, 1189-1202 (2015).

9. K. Michailidou *et al.*, Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
10. M. M. S. Obradovic *et al.*, Glucocorticoids promote breast cancer metastasis. *Nature* **567**, 540-544 (2019).
11. P. M. Rist, C. Tzourio, T. Kurth, Associations between lipid levels and migraine: cross-sectional analysis in the epidemiology of vascular ageing study. *Cephalalgia* **31**, 1459-1465 (2011).
12. R. Mevel, J. E. Draper, A. L. M. Lie, V. Kouskoff, G. Lacaud, RUNX transcription factors: orchestrators of development. *Development* **146**, (2019).
13. K. S. Kobayashi, P. J. van den Elsen, NLRC5: a key regulator of MHC class I-dependent immune responses. *Nat Rev Immunol* **12**, 813-820 (2012).
14. A. Acharya *et al.*, miR-26 suppresses adipocyte progenitor differentiation and fat production by targeting Fbxl19. *Genes Dev* **33**, 1367-1380 (2019).
15. T. Rathjen *et al.*, Regulation of body weight and energy homeostasis by neuronal cell adhesion molecule 1. *Nat Neurosci* **20**, 1096-1103 (2017).
16. E. R. Gamazon *et al.*, Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).
17. J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216 (2012).
18. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
19. C. D. Arnold *et al.*, Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**, 1074-1077 (2013).
20. M. R. Mumbach *et al.*, Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612 (2017).
21. S. A. Gagliano Taliun *et al.*, Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat Genet* **52**, 550-552 (2020).
22. I. G. Ovsyannikova *et al.*, Genome-wide association study of antibody response to smallpox vaccine. *Vaccine* **30**, 4182-4189 (2012).
23. D. Vuckovic *et al.*, The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231 e1211 (2020).
24. C. D. Langefeld *et al.*, Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat Commun* **8**, 16021 (2017).
25. T. F. Andlauer *et al.*, Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv* **2**, e1501678 (2016).
26. M. H. Chen *et al.*, Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213 e1114 (2020).
27. K. Matsunami *et al.*, Genome-Wide Association Study Identifies ZNF354C Variants Associated with Depression from Interferon-Based Therapy for Chronic Hepatitis C. *PLoS One* **11**, e0164418 (2016).
28. J. Ward *et al.*, The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. *Mol Psychiatry*, (2019).
29. J. R. I. Coleman *et al.*, The Genetics of the Mood Disorder Spectrum: Genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. *Biol Psychiatry* **88**, 169-184 (2020).

30. U. Styrkarsdottir *et al.*, Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* **358**, 2355-2365 (2008).
31. Y. Wu *et al.*, Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nature Communications* **10**, (2019).
32. S. Vogelezang *et al.*, Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. *PLoS Genet* **16**, e1008718 (2020).
33. N. Niu *et al.*, Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Res* **20**, 1482-1492 (2010).
34. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
35. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
36. K. D. Hansen, R. A. Irizarry, Z. Wu, Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204-216 (2012).
37. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
38. J. Zhong *et al.*, Purification of nanogram-range immunoprecipitated DNA in ChIP-seq application. *BMC Genomics* **18**, 985 (2017).
39. H. Yan *et al.*, HiChIP: a high-throughput pipeline for integrative analysis of ChIP-Seq data. *BMC Bioinformatics* **15**, 280 (2014).
40. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
41. A. A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358 (2012).
42. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
43. C. A. Davis *et al.*, The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794-D801 (2018).
44. B. E. Bernstein *et al.*, The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048 (2010).
45. J. Ernst, M. Kellis, Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* **33**, 364-376 (2015).
46. T. E. Reddy *et al.*, Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* **19**, 2163-2171 (2009).
47. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
48. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
49. M. Lawrence *et al.*, Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
50. I. Juric *et al.*, MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput Biol* **15**, e1006982 (2019).
51. B. Gyorffy, Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J* **19**, 4101-4109 (2021).

52. B. Gyorffy *et al.*, An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-731 (2010).
53. F. Luca *et al.*, Genetic, functional and molecular features of glucocorticoid receptor binding. *PLoS One* **8**, e61654 (2013).
54. D. W. Cain, J. A. Cidlowski, Immune regulation by glucocorticoids. *Nat Rev Immunol* **17**, 233-247 (2017).
55. D. S. Goodin, Glucocorticoid treatment of multiple sclerosis. *Handb Clin Neurol* **122**, 455-464 (2014).
56. M. Mosca, C. Tani, L. Carli, S. Bombardieri, Glucocorticoids in systemic lupus erythematosus. *Clin Exp Rheumatol* **29**, S126-129 (2011).
57. C. E. Neunert, Management of newly diagnosed immune thrombocytopenia: can we change outcomes? *Blood Adv* **1**, 2295-2301 (2017).
58. L. Buckley, M. B. Humphrey, Glucocorticoid-Induced Osteoporosis. *N Engl J Med* **379**, 2547-2556 (2018).
59. M. Kennis *et al.*, Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. *Mol Psychiatry* **25**, 321-338 (2020).
60. M. Belvederi Murri *et al.*, The HPA axis in bipolar disorder: Systematic review and meta-analysis. *Psychoneuroendocrinology* **63**, 327-342 (2016).
61. D. Borsook, N. Maleki, L. Becerra, B. McEwen, Understanding migraine through the lens of maladaptive stress responses: a model disease of allostatic load. *Neuron* **73**, 219-234 (2012).
62. A. Vegiopoulos, S. Herzig, Glucocorticoids, metabolism and metabolic diseases. *Mol Cell Endocrinol* **275**, 43-61 (2007).

Acknowledgments: We thank the GTEx and ENCODE Consortiums for their generation of transcriptomic and epigenomic datasets used in this study. We want to acknowledge the participants and investigators of the FinnGen study.

Funding:

National Institutes of Health grant U19GM61388 (RMW, LWang)

National Institutes of Health grant R01GM28157 (RMW)

National Institutes of Health grant R01AA027486 (RMW)

National Institutes of Health grant R01DK126827 (TO)

National Institutes of Health grant R01DK058185 (TO)

The Noble Foundation (RMW)

Author contributions:

Conceptualization: TTLN, DL, TO, RMW

Investigation: TTLN, DL, LZ

Data Curation: HG, KC, LWei, HL, CZ

Formal Analysis: HG, TTLN, ZY

Methodology: JHL, GS, TO

Resources: LWang, TO

Supervision: RMW, TO

Funding Acquisition: RMW and LWang

Visualization: TTLN, HG, KC

Data Interpretation: TTLN, DL, HG, JY, LWang, TO, RMW

Writing – original draft: TTLN, RMW

Writing – review & editing: DL, HG, JHL, LZ, JY, TO, RMW

Competing interests: Drs. Weinshilboum and Wang are co-founders and stockholders in OneOme, LLC. Other authors declare no conflict of interests

Data and materials availability: All sequencing data was deposited on Gene Expression Omnibus under accession number GSE185941.

Supplementary materials

Materials and Methods

Supplementary Text

Figures S1-S14

Tables S1-S3

References (*33-62*)

Data S1-S3

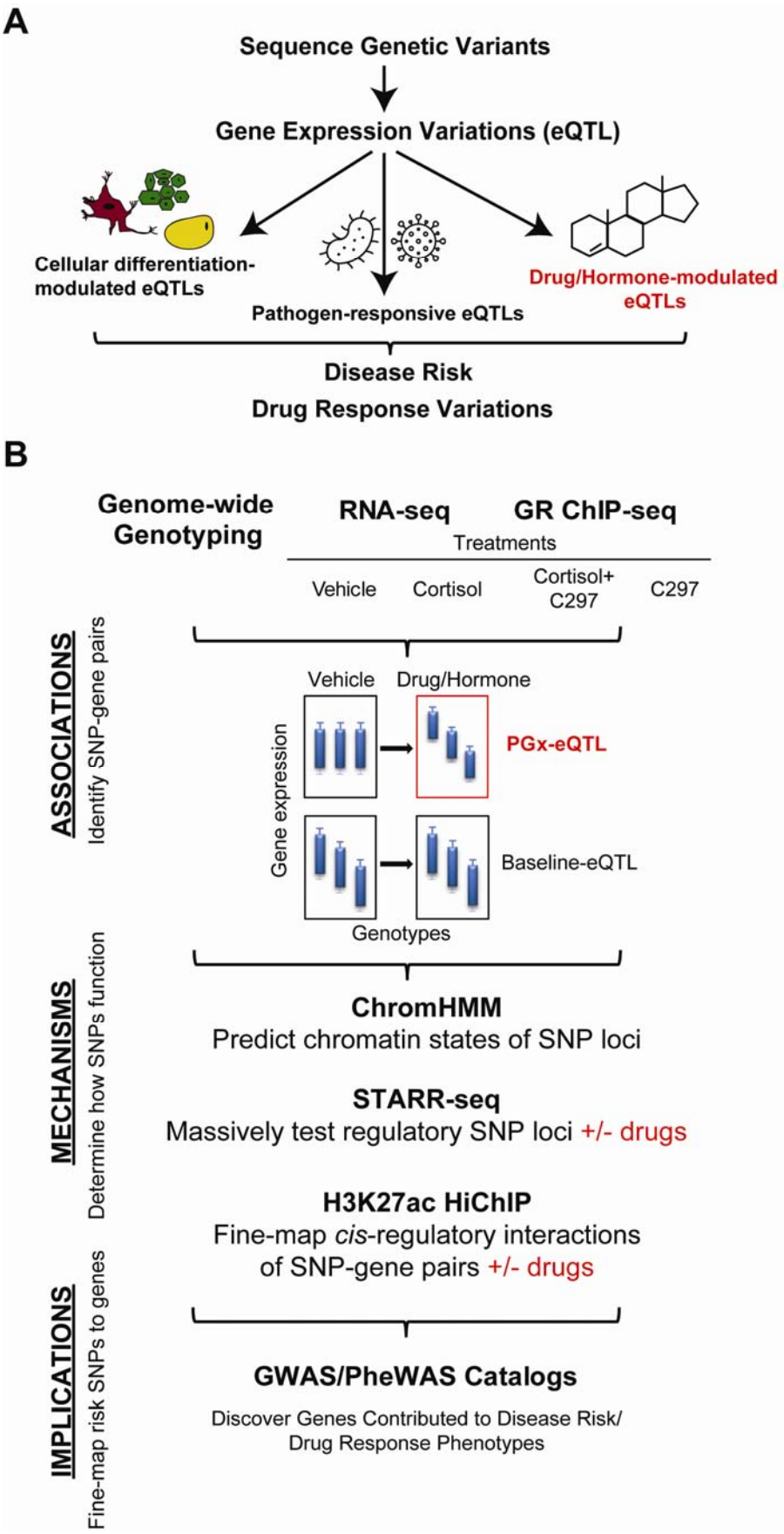
A
Sequence Genetic Variants
Gene Expression Variations (eQTL)
Cellular differentiation-modulated eQTLs
Pathogen-responsive eQTLs
Drug/Hormone-modulated eQTLs
Disease Risk
Drug Response Variations
B
Genome-wide Genotyping
RNA-seq
GR ChIP-seq
Treatments
Vehicle
Cortisol
Cortisol+ C297
C297
ASSOCIATIONS
Identify SNP-gene pairs
Vehicle
Drug/Hormone
Gene expression
PGx-eQTL
Baseline-eQTL
Genotypes
MECHANISMS
Determine how SNPs function
ChromHMM
Predict chromatin states of SNP loci
STARR-seq
Massively test regulatory SNP loci +/- drugs
H3K27ac HiChIP
Fine-map cis-regulatory interactions of SNP-gene pairs +/- drugs
IMPLICATIONS
Fine-map risk SNPs to genes
GWAS/PheWAS Catalogs
Discover Genes Contributed to Disease Risk/ Drug Response Phenotypes

Fig.1. Conceptual Framework for the Study. (**A**) PGx-eQTLs represent a type of dynamic genetic sequence variants that influence transcriptional variation during exposure to hormones or drugs. (**B**) Experimental design used in this study. ChromHMM is a software used to annotate chromatin states from epigenomic data. STARR-seq: self-transcribing active regulatory region sequencing. Cortisol is an agonist, C297 is a modulator which acts as an antagonist when administered together with cortisol and a partial agonist by itself.

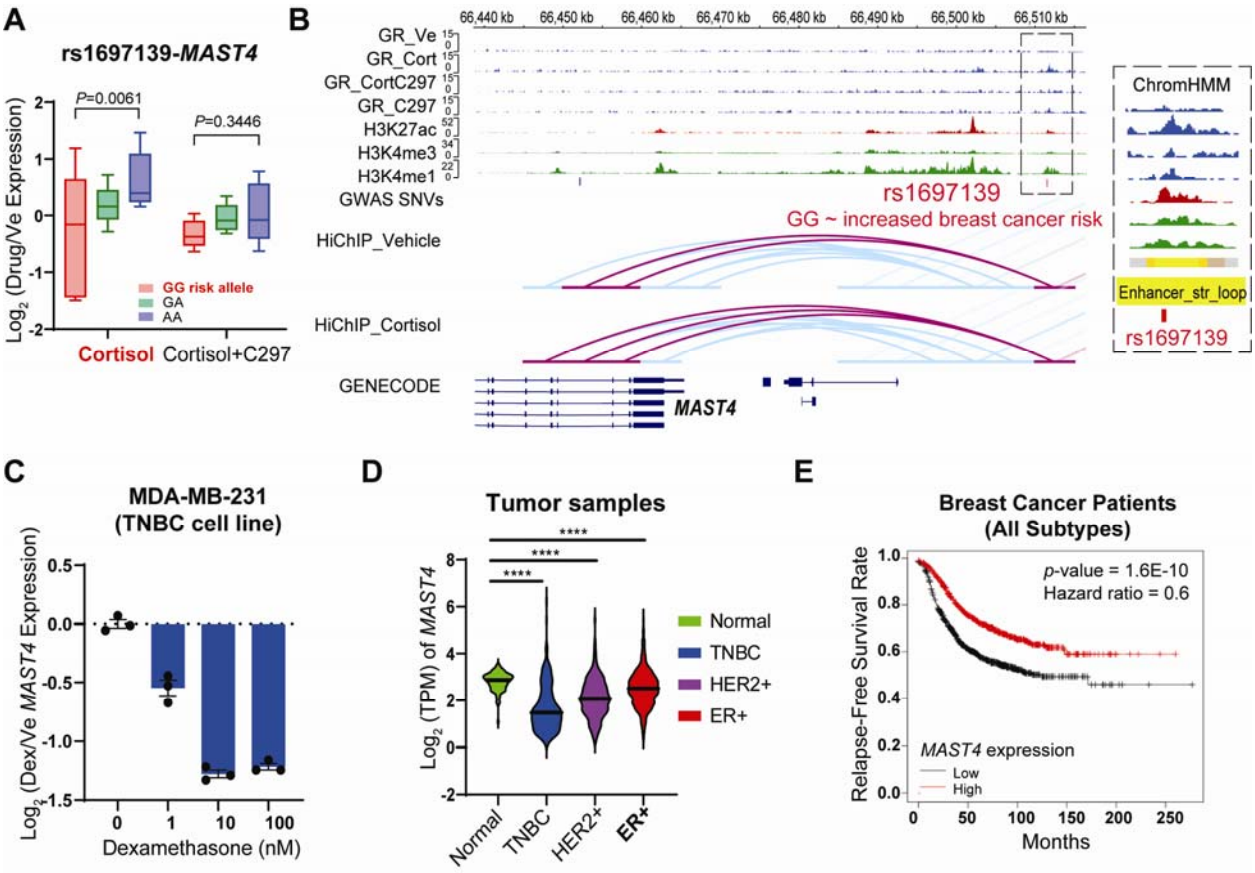

Fig.2. A Selected Example of GR-dependent PGx-eQTLs. (A) PGx-eQTL SNP-gene pair. Cortisol induced eQTL, and C297 antagonized cortisol effect, normalizing eQTL expression. (**B**) Integrative Genomics Viewer (IGV) plot of the PGx SNP-eQTL gene locus. For the H3K27ac HiChIP tracks, loops directly interacting with the PGx SNP loci are highlighted in pink and others in blue. (**C**) Dose-dependent repression of *MAST4* by glucocorticoids in a triple-negative breast cancer cell line. (**D**) Expression of *MAST4* in tumors from breast cancer patients in the TCGA database. TPM = transcripts per million. (**E**) Kaplan-Meier curve of relapse-free survival rate in 1764 breast cancer patients predicted based on *MAST4* expression.

Table 1. GR-dependent PGx-eQTLs which we identified that have been associated with a clinical phenotype by previous GWAS/PheWAS but for which the gene responsible was unclear

Index SNP	Ligands	PGx-eQTL Gene(s) Identified	SNP-Associated Phenotype(s) in GWAS/PheWAS	
rs11633087/ rs4843073/ rs4843074/ rs4843075/ rs7162168/	Cortisol	*KLHL25*	Adrenal cortical steroids causing adverse effects in therapeutic use (*21*)	**Pharmaco-genomics**
rs2834655	Cortisol	*RUNX1*	Immune response to smallpox vaccine (*22*) Monocyte percentage of white cells (*23*)	**Inflammation/ Immunity**
rs2984920/ rs7535818	Cortisol	*RGS1*	Systemic lupus erythematosus risk (*24*) Intestinal malabsorption/Celilac disease (*21*) Multiple sclerosis (*25*) Dermatitis and eczema*	
rs4735336	Cortisol	*PLEKHF2*	Viral hepatitis C (*21*)	
rs2297539	Cortisol	*IKBKE/ SRGAP2*	Fasciitis Celilac disease (*21*)	
rs2399594	Cortisol	*NLRC5*	Lichen Planus (inflammation of the skin)*	
rs12440899	Cortisol	*SRP14-AS1*	Inflammatory liver diseases*	
rs12053126	CORT108297	*AFF3*	Otosclerosis*	
rs9594738	CORT108297	*DGKH*	Otosclerosis*	
rs2051541	Cortisol	*HIST1H2AC*	Celilac disease (*21*) Ankylosing spondylitis (inflammation of the spine)* Acute and subacute iridocyclitis (inflammation of the iris)*	
rs7356	Cortisol	*EYA3*	Platelet count (*26*)	
rs4843073	Cortisol	*KLHL25*	Neutrophil count (*23*)	
rs4984913	CORT108297	*CCDC78/WDR90/FAM195A/ WFIKKN1*	Platelet count (*23*)	

rs2779180	Cortisol	*ARRDC5*	Depression in response to interferon-based therapy of chronic hepatitis C (*27*)	**Psychiatry & Neurology**
rs1050863	Cortisol	*ZCCHC14/JPH3*	Mood Instability (*28*)	
rs696284	CORT108297	*NFX1*	Mood disorders*	
rs7288411	Cortisol	*RRP7BP*	Bipolar disorder or major depressive disorder (*29*)	
rs11678116	Cortisol	*LDAH*	Migraine*	
rs1697139	Cortisol	*MAST4*	Breast cancer risk (*9*)	**Cancer**
rs9594738	CORT108297	*DGKH*	Bone mineral density (*30*) Medication use (drugs affecting bone structure and mineralization) (*31*) Otosclerosis*	**Bone-related diseases**
rs12053126	Cortisol	*AFF3*	Otosclerosis*	
rs1539330	Cortisol	*KIF11*	Fracture of hand or wrist (*21*)	
rs7406439	Cortisol	*HEXDC*	Disorders of continuity of bone*	
rs1978487	Cortisol	*FBXL19/HSD3B7*	Obesity/hypertension/Body mass index (*21*)	**Metab-olism**
rs11215427	Cortisol	*CADM1*	Childhood body mass index (*32*)	
rs4735336/ rs4735337	Cortisol	*PLEKHF2*	Hypertensive diseases*	
rs2051541	Cortisol	*HIST1H2AC*	Disorder of iron metabolism (*21*) Hereditary hemochromatosis (*21*)	**Others**
rs10053292	Cortisol	*SLC26A2*	Acquired deformities of finger (*1*)	
rs12440899	Cortisol	*SRP14-AS1*	Endometriosis*	
rs2313167	CORT108297	*PDLIM2*	Height (*1*)	

* Associations from FinnGen PheWAS

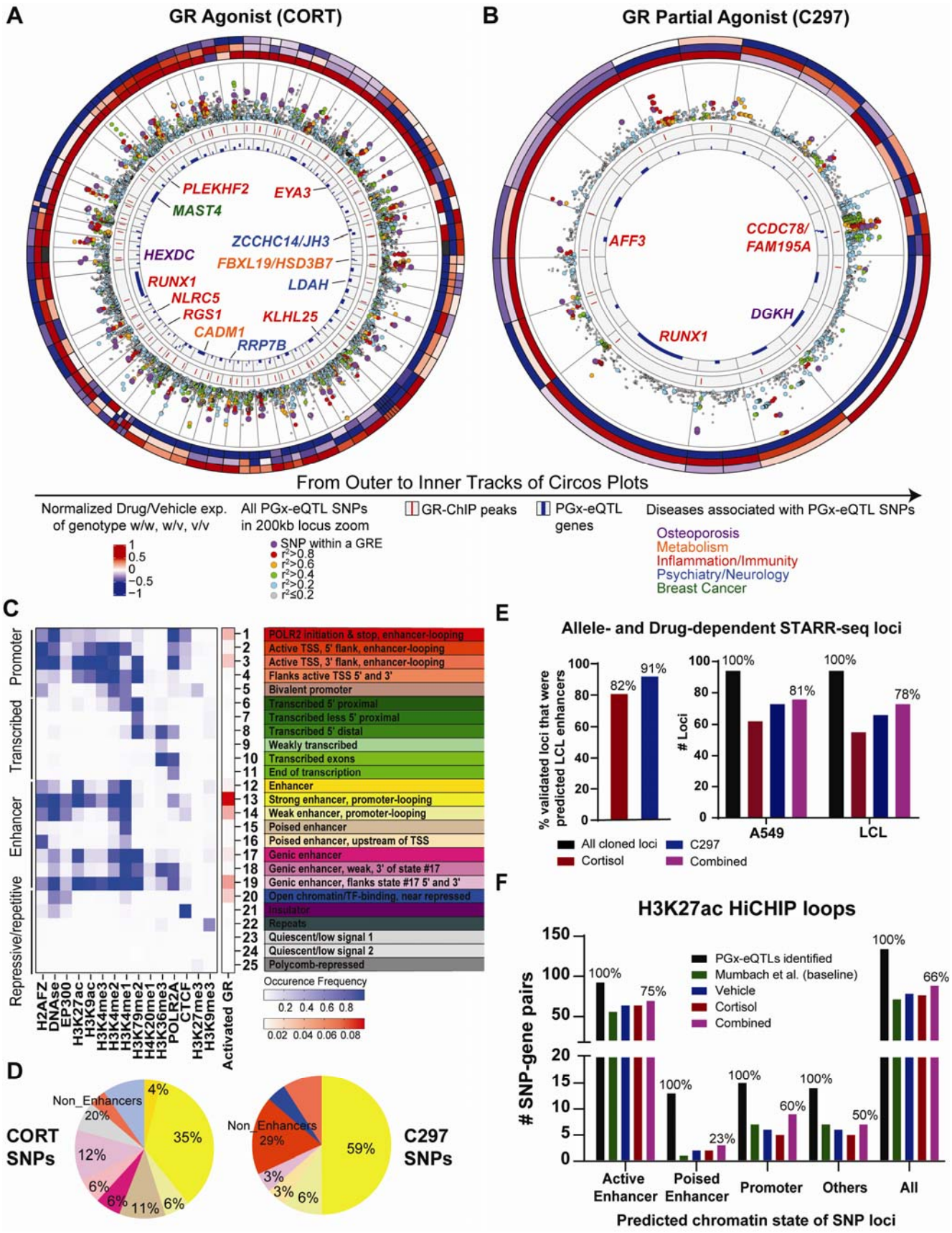
A
GR Agonist (CORT)
B
GR Partial Agonist (C297)
PLEKHF2
EYA3
MAST4
ZCCHC14/JH3
HEXDC
FBXL19/HSD3B7
RUNX1
LDAH
NLRC5
RGS1
KLHL25
CADM1
RRP7B
AFF3
CCDC78/
FAM195A
DGKH
RUNX1
From Outer to Inner Tracks of Circos Plots
Normalized Drug/Vehicle exp. of genotype w/w, w/v, v/v
All PGx-eQTL SNPs in 200kb locus zoom
GR-ChIP peaks
PGx-eQTL genes
Diseases associated with PGx-eQTL SNPs
SNP within a GRE
Osteoporosis
Metabolism
Inflammation/Immunity
Psychiatry/Neurology
Breast Cancer
C
Promoter
Transcribed
Enhancer
Repressive/repetitive
H2AFZ
DNAse
EP300
H3K27ac
H3K9ac
H3K4me3
H3K4me2
H3K4me1
H3K79me2
H4K20me1
H3K36me3
POLR2A
CTCF
H3K27me3
H3K9me3
Activated GR
POLR2 initiation & stop, enhancer-looping
Active TSS, 5' flank, enhancer-looping
Active TSS, 3' flank, enhancer-looping
Flanks active TSS 5' and 3'
Bivalent promoter
Transcribed 5' proximal
Transcribed less 5' proximal
Transcribed 5' distal
Weakly transcribed
Transcribed exons
End of transcription
Enhancer
Strong enhancer, promoter-looping
Weak enhancer, promoter-looping
Poised enhancer
Poised enhancer, upstream of TSS
Genic enhancer
Genic enhancer, weak, 3' of state #17
Genic enhancer, flanks state #17 5' and 3'
Open chromatin/TF-binding, near repressed
Insulator
Repeats
Quiescent/low signal 1
Quiescent/low signal 2
Polycomb-repressed
Occurence Frequency
D
CORT SNPs
Non_Enhancers
C297 SNPs
E
Allele- and Drug-dependent STARR-seq loci
% validated loci that were predicted LCL enhancers
# Loci
A549
LCL
All cloned loci
Cortisol
C297
Combined
F
H3K27ac HiCHIP loops
SNP-gene pairs
PGx-eQTLs identified
Mumbach et al. (baseline)
Vehicle
Cortisol
Combined
Active Enhancer
Poised Enhancer
Promoter
Others
All
Predicted chromatin state of SNP loci

Fig.3. Global Characterization of GR-dependent PGx-eQTLs. (**A**), (**B**) Global views of GR-dependent PGx-eQTLs identified in 30 LCLs with results for cortisol (CORT) and the partial agonist (C297) shown in (**A**) and (**B**), respectively. The three outermost circles represent relative gene expression values (drug/vehicle) for each genotype, with each column depicting a PGx SNP-gene pair. The inner tracks are explained in the figure. (**C**) 25 LCL chromatin states predicted from the occupancy of 15 epigenetic marks of the reference LCL from ENCODE and the enrichment of GR peaks within each state. (**D**) Distribution of GR-dependent PGx-eQTL SNPs among different chromatin states, which have been color coded as in (**C**). (**E**) STARR-seq results. (**F**) Number of PGx-eQTLs that displayed physical interactions between SNP loci (categorized by enhancer/non-enhancer states) and eQTL genes as demonstrated by H3K27ac HiChIP.