Supplemental Material

Material and Methods

External Transcriptomic Datasets

We searched the LINCS database, a catalog of transcriptomics assays that directly measured 978 landmark transcripts and computationally inferred >12,200 additional genes,[1] which contains transcriptional responses of >50 human cell lines to each of ~20,000 compounds across a range of conditions.[2] The >22,200 transcripts generated using Affimetrix Human Genome HGU133A microarray platform for the Library of Integrated Cellular Signatures (LINCS) database were annotated using the Gene Expression Omnibus (GEO) *probeTOentrez_gene* mapping file dated 10/13/2016.

We retrieved gene expression data on 15 available ART drugs before and after 6- and 24-hour treatments at a dose of 10 μM each across up to 59 different human cell lines (the number of total experimental instances per drug ranging between 37 and 174), for a total of 1,127 experiments (Supplemental Table S1). The 15 available ART drugs belonged to five classes: NNRTIs (efavirenz and nevirapine), NRTIs (zalcitabine, zidovudine, didanosine, stavudine, tenofovir, and lamivudine), PIs (indinavir, nelfinavir, ritonavir, and saquinavir), integrase inhibitors (raltegravir and elvitegravir), and an entry inhibitor (maraviroc).

To identify novel mechanisms by which ART may promote CAD, we looked for co-expression enrichments of ART-induced transcripts in CAD-causal regulatory gene networks (RGNs) and their key drivers (Supplemental Table S2) reported in the STAGE study.[3] In that study, seven tissues - atherosclerotic arterial wall, non-atherosclerotic arterial wall (internal mammary artery), liver, skeletal muscle, visceral fat, subcutaneous fat, and whole blood - were obtained during coronary artery bypass surgery from 114 patients with advanced CAD.[4] RGNs, their key drivers, and co-expression modules were inferred by applying weighted co-expression network and Gaussian Bayesian network algorithm to genotype data and 612 gene expression profiles obtained from these patients. Genes with expression single nucleotide polymorphisms

(eSNPs), transcription factors, CAD-associated genome-wide association hits, or genes for established CAD biological processes and molecular functions according to Gene Ontology (GO)[5] were used as the priors to defining parent node of other genes.[3] The Bayesian Information Criterion (BIC) was applied to score models; a multiple-restart greedy hill-climbing algorithm, with edge additions, deletions, and reversals, was used to search locally optimal models.[6] Key drivers were identified from the RGNs as described elsewhere.[7] The 30 RGNs examined in the present study were identified by the STAGE study[3]as being CAD-causal based on their association with several key CAD phenotypes, including extent of coronary atherosclerosis, plasma lipid levels, glucose metabolism biomarkers, and the inflammatory marker C-reactive protein.

Enrichment of CAD-related Genes in ART-Induced Transcriptional Signatures

Using LINCS gene expression profiles before and after treatment with each drug, we first generated ART-induced gene signatures by employing a heuristic based on selection of the 100 most over or under-expressed genes in response to ART from each experiment across different cell types and time exposures. Due to the experimental complexity of the LINCS collection, and the distribution of experimental replicates across multiple treatment plates, standard differential gene expression methodologies are not well suited to defining individual drug signatures. Therefore, the drug-induced transcriptomic profiles available within LINCS were best suited to "signature-level" connectivity mapping, rather than as a compendium of individual drug-gene associations with corresponding statistical significance levels. This approach was more comparable to a gene-set enrichment analysis type paradigm, than a conventional differential expression analysis, and was not based on statistical significance. Signatures were normalized with reference to within-plate dimethyl sulfoxide (DMSO) controls, which allowed comparison and combination of signatures across disparate treatment conditions, and treatment plates. To define individual drug signatures, we ranked gene expression by the mean robust Z-score (averaged across all experimental instances for the same drug) in descending order, and selected the genes with the top and bottom 100 most extreme robust Z-scores.

We only included the genes that were up or downregulated in 2 or more experiments. We then ranked gene expression by the mean robust Z-score (averaged across all experimental instances for the same

drug) in descending order, and selected the genes with the top and bottom 100 most extreme robust Z-scores. We selected this threshold for consistency with existing connectivity mapping tools,[8] and also previous reports of optimal 'rediscovery' of known clinical indications for drug signatures based on the 100 most extreme genes.[9] To assess the overrepresentation of the genes from the 30 CAD-causal RGNs among the individual ART-induced genes, we applied a one-tailed Fisher's exact test assessing whether the probability of observing the genes on both lists was greater than expected by chance. Of 7,378 genes in the 171 co-expression modules identified in the STAGE study,[3] 4,576 genes were present in LINCS and used in our analysis. The Benjamini-Hochberg method was applied globally to account for multiple hypotheses testing across all ART-RGN combinations for up and downregulated genes (900 tests).[10]

Co-expression Analysis of CAD-related Genes in ART-induced Transcriptional Signatures

To determine co-expression patterns of the genes associated with ART response and those linked to CAD, for each ART transcriptional profile extracted from LINCS, we assessed which of the 30 CAD-causal RGNs were more strongly co-expressed with ART-induced gene signatures than with random sets of genes using the *WGCNA R* package.[11] We calculated the connection strengths for all pairs of ART-induced genes within each of the 30 CAD-causal RGNs using Pearson correlation analysis and compared those to the connection strengths from 1,000 random gene sets of the same size. Since in a weighted gene co-expression network analysis (as *WGCNA*), the soft-threshold power is recommended as a noise filtering, soft-threshold power β was used to calculate adjacency based on the criterion of approximate scale-free topology. We used the guidelines at https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html to choose an appropriate threshold for network construction. We generated a connectivity measure, *k*, which is an average of adjacency values derived from weighted Pearson correlation coefficients using scaling parameter $\beta = 6$, that assessed the co-expression of the ART-induced genes within each RGN versus within the random set of genes. The strength of co-expression of the ART-induced genes within the CAD-causal RGN genes compared to those within random sets of genes, was assessed using a Wilcoxon rank-sum test implemented

in the *coin R* package.[12] We used the *NetWeaver* R package[13; 14] to generate a circular layout summarizing all significant results across the 15 ART drugs with adjusted median P<0.05 across the 1,000 tests.

Experimental Validation

Cell incubation and phenotyping

For experimental validation, we selected the three PIs (ritonavir, nelfinavir and saquinavir) whose drug-induced gene signatures were both enriched and co-expressed in atherosclerosis-related RGN (AR-RGN). We also included an integrase inhibitor raltegravir as a negative control since it was previously reported to not induce CE accumulation in THP-1 cells and an entry inhibitor maraviroc as it induced downregulation of *PQBP1* in THP-1 cells in LINCS and was associated with beneficial lipid profiles in HIV patients treated with the drug. To validate these 5 drugs, we used the *in vitro* atherosclerosis model based on human THP-1 monocytic cell line. In this model, THP-1 cells are differentiated into macrophages and loaded with acetylated low-density lipoproteins (AcLDLs) to induce cholesteryl ester (CE) accumulation and foam cell formation. In brief, human THP-1 cells (American Type Culture Collection, ATCC) were cultured in RPMI-40 medium (Corning) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Sigma) at 37°C, 5% CO_2. Macrophage differentiation was induced *in vitro* with phorbol-12-myristate-13-acetate (Sigma) (50 ng/mL) for 72 hours at 37°C in 5% CO_2, as described.[15] Macrophages were first pre-treated for 24 hours with vehicle (DMSO) or with either saquinavir (Sigma), ritonavir, nelfinavir, maraviroc, or raltegravir (all from Selleckchem) at concentrations of 5, 15, or 30 µM. The cells were then incubated with AcLDL (50 µg/mL; Alfa Aesar) for 48 hours to generate foam cells; drug treatment was maintained throughout the experiment. The effect of each drug treatment on intracellular CE accumulation was measured in cell lysates with the Total Cholesterol and Cholesteryl Ester Colorimetric/Fluorometric Assay kit (BioVision) as recommended by the manufacturer. For all treatments, 3 independent experiments per drug were performed with 3 technical replicates for each condition (n=9); values are reported as mean ± SEM. Differences in CE levels in all experiments were assessed on raw data using analysis of variance with Dunnett's adjustment for multiple comparisons. We used GraphPad Prism 6 (La Jolla, CA) and considered adjusted P<0.05 to be statistically significant.

RNA sequencing analyses

We performed RNA-sequencing (RNA-Seq; 30 million read depth, single-end reads, 100 base pairs) using the Illumina HiSeq 2500 platform on 32 RNA samples isolated from AcLDL-loaded macrophages after treatment with vehicle (DMSO) or with ARTs (ritonavir at 30 μM; nelfinavir, saquinavir and maraviroc at 15 μM). The ART with no effect on CE accumulation (raltegravir) was excluded. We first aligned the RNA-Seq data with HISAT2[16] to GRCh38, and quantified gene expression with Subread[17] and StringTie[18] to create two main assembled count matrices—one for genes and one for transcripts. For normalization and estimation of differentially expressed genes while comparing drug perturbations to DMSO treatment, we used DESeq2[19] after adjustment for experimental batch effects. To assess whether genes in a particular RGN were over-represented among the differentially expressed genes from the four drug perturbations, we used an empirical null distribution including only protein-coding genes (17,157 in total) to avoid any biases introduced by noncoding RNA present in our RNA-Seq data but absent in the microarray data from LINCS.

Transfection of small interfering RNA

To test the effect of the key driver *PQBP1* on foam cell formation induced by the 3 ART drugs associated with CE accumulation, we silenced *PQBP1* by transfecting THP-1-derived macrophages with a pool of 3 *PQBP1* small interfering RNA (siRNA) oligonucleotides (Ambion, Thermo Fisher Scientific, siRNA IDs: s19619, s19621, s527195), using Lipofectamine RNAiMax (Invitrogen) according to the manufacturer's instructions. Another pool of 2 control siRNA oligonucleotides (Ambion, Thermo Fisher Scientific, cat. no. 4390843 and 4390846) was used as control. The efficiency of target gene silencing by TaqMan analyses (Hs00172888_m1) was affirmed for up to 72 hours. The CE accumulation after *PQBP1* silencing was quantified in 3 independent experiments with 3 technical replicates for each condition (n=9) and 2 technical replicates for each condition of *PQBP1* RNA levels quantification (n=6).

Supplemental Tables

Supplemental Table S1. The list of the 15 antiretroviral drugs (ART) with transcriptional data available for various cell types in the LINCS database.

Drug name	Drug class	Cell line
Maraviroc	Entry inhibitor	A375,A549,A673,AGS,CL34,CORL23,COV644,DV90,EFO27,HA1E,HCC15,HCC515, HCT116,HEC108,HEPG2,HT115,HT29,JHUEM2,LOVO,MCF7,MDST8,NCIH1694, NCIH1836,NCIH2073,NCIH508,NCIH596,NOMO1,OV7,PC3,PL21,RMGI, RMUGS,SKLU1,SKM1,SKMEL1,SKMEL28,SNGM,SNU1040,SNUC4,SNUC5, SW480,SW620,SW948,T3M10,THP1,TYKNU,U937,VCAP,WSUDLCL2,H1299,RKO
Raltegravir	Integrase inhibitor	A375,A549,ASC,HA1E,HCC515,HEPG2,HT29,MCF7,NEU,NPC,PC3,SKB,VCAP
Elvitegravir	Integrase inhibitor	A375,A549,HA1E,HCC515,HEPG2,HT29,MCF7,NEU,PC3,VCAP
Efavirenz	NNRTI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP
Nevirapine	NNRTI	A375,VCAP,A549,HA1E,HT29,MCF7,PC3,HCC515
Zalcitabine	NRTI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP
Zidovudine	NRTI	MCF7,PC3,VCAP,A375,A549,HA1E,HCC515,HT29
Didanosine	NRTI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP
Stavudine	NRTI	HT29,VCAP,A375,A549,HA1E,MCF7,PC3,HCC515
Tenofovir	NRTI	A375,A549,ASC,HA1E,HCC515,HEPG2,HT29,MCF7,NPC,PC3,PHH,SKB,VCAP
Lamivudine	NRTI	A375,A549,HA1E,HCC515,HEPG2,HT29,MCF7,PC3,VCAP
Indinavir	PI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP
Nelfinavir	PI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP,ASC,HCC515,HEPG2,NEU,NPC,PHH,SKB
Ritonavir	PI	A375,A549,HA1E,HT29,MCF7,PC3,VCAP,HCC515
Saquinavir	PI	HA1E,HCC515,PC3,VCAP,A375,A549,HT29,MCF7,ASC,HEPG2,NPC,PHH,SKB

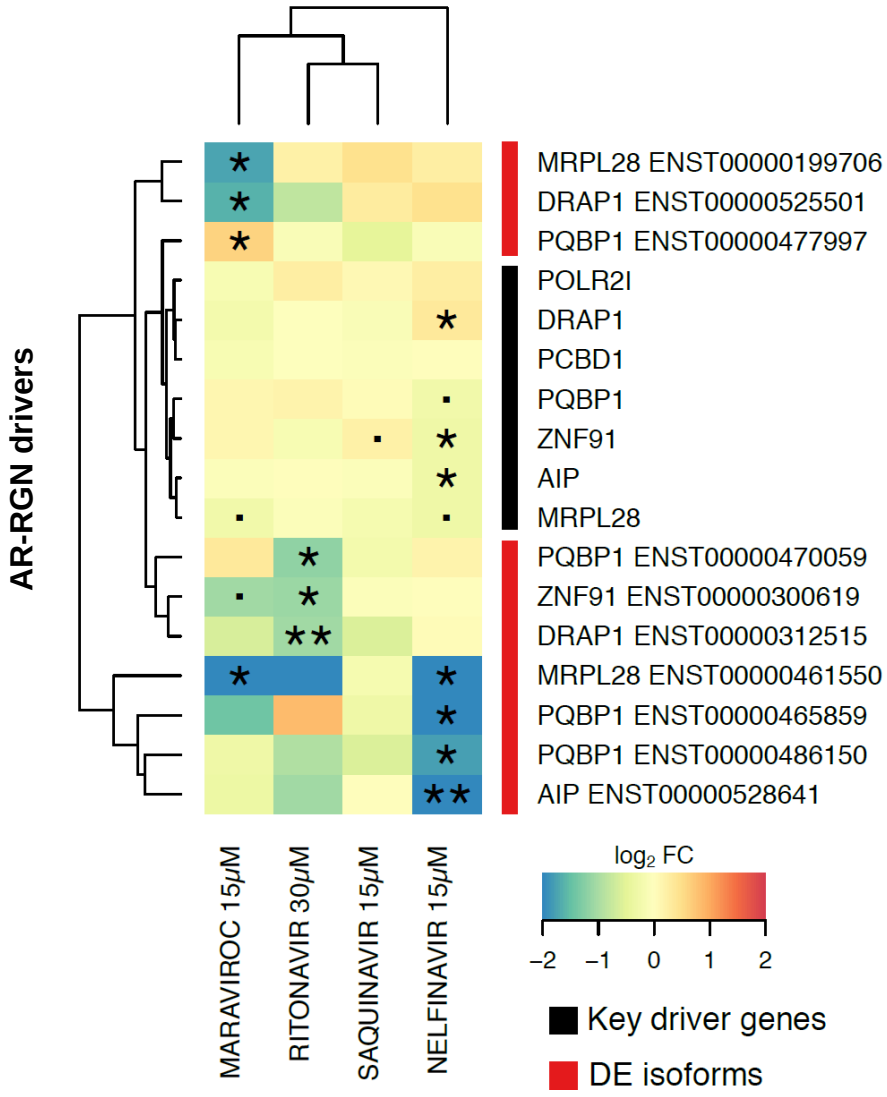
NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor, PI, Protease inhibitor.

Supplemental Tables as Excel files

Supplemental Table S2. List of coronary artery disease (CAD)-causal regulatory gene networks (RGN) derived in the STAGE study (Talukdar et al.).

Supplemental Table S3. The adjusted p-values for co-expression between 30 regulatory gene networks (RGN) and 15 antiretroviral drugs.

Supplemental Figure



Supplemental Figure S1. Heat map of differentially expressed key driver genes from the atherosclerosis-related regulatory gene network (AR-RGN) and their isoforms comparing treatments with DMSO versus maraviroc, ritonavir, saquinavir, and nelfinavir, respectively, in THP-1 foam cells incubated with acetylated low density lipoprotein.

Supplemental References

1. Subramanian, A., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 171, 1437-1452 e1417.
2. Keenan, et al. (2018). The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. Cell Syst 6, 13-24.
3. Talukdar, H.A., et al. (2016). Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. Cell systems 2, 196-208.
4. Hagg, S., et al. (2009). Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. PLoS Genet 5, e1000754.
5. Ashburner, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-29.
6. Schmidt, M., et al. (2007). Learning graphical model structure using L1-regularization paths. AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence 2, 1278–1283.
7. Zhang, B. and Zhu, J. (2013). Identification of Key Causal Regulators in Gene Networks. Proceedings of the World Congress on Engineering 2013 II, 1309-1312.
8. Duan, Q., et al. (2014). LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. Nucleic Acids Res 42, W449-460.
9. Cheng, J., et al. (2013). Evaluation of analytical methods for connectivity map data. Pac Symp Biocomput, 5-16.
10. Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series 57, 289–300.
11. Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.
12. Hothorn, T., et al. (2008). Implementing a class of permutation tests: The coin package. J Stat Softw 28, 1-23.
13. Wang, M., et al. (2016). Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Medicine 8, 104.
14. Wang, M., and Zhang, B. (2016). NetWeaver: Graphic Presentation of Complex Genomic and Network Data Analysis. In. (
15. Tsuchiya, S., et al. (1982). Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. Cancer Res 42, 1530-1536.
16. Kim, D., et al. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods 12, 357-360.
17. Liao, Y., et al. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41, e108.
18. Pertea, M., et al. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11, 1650-1667.
19. Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106.