SUPPLEMENTAL MATERIAL

Supplemental Methods:

Human liver tissues and human primary hepatocytes:

Human liver tissues were obtained from the Cooperative Human Tissue Network. Human primary hepatocytes (Hep4, Hep8, and Hep10) were obtained from the Liver Tissue Cell Distribution System (LTCDS, Pittsburgh, PA). The demographic information of liver tissues and hepatocytes is shown in Supplemental Table 6.

Preparation of 4C template: The 4C template for human primary hepatocytes was prepared as described [1]. Briefly, 10 million primary cultured human hepatocytes were detached with trypsin to obtain a single cell suspension. Then cells were cross-linked with 2% formaldehyde for 10 min before lysis to release cell nuclei. Cell nuclei were digested with DpnII at 37°C overnight followed by proximity ligation with T4 ligase after a >20-fold dilution. After reverse cross-linking, the purified DNA was subjected to secondary digestion with NlaIII, followed again by proximity ligation in a diluted condition. The resulting DNA was purified by phenol/chloroform extraction.

PCR amplification of 4C template and Ion Torrent sequencing: We used our modified protocol [2] for 4C template amplification followed by Ion Torrent sequencing. Inverse primers were designed as described [3]. PCR was performed over 25 cycles with inverse primers (Supplemental Table 3) specific to the *CYP7A1* promoter using 25 ng 4C templates. A total of 8 PCR reactions were performed and the PCR products combined. We used 100 ng DNA for Ion Torrent library preparation using the NEB Next Fast DNA library Preparation kit (New England BioLabs, Ipswich, MA, USA). After adaptor ligation, the DNA was purified with four columns of High

Pure PCR Product Purification columns (Roche, Madison, WI, USA) to avoid adaptor dimer contamination. The adaptor ligated library was PCR amplified for 8 cycles followed by DNA purification with 2 columns of High Pure PCR Product Purification columns. The amplified library was diluted and subjected to emulsion PCR using an Ion OneTouch 2 instrument (Life Technology, Grant Island, NY, USA). The resulting library was sequenced on Ion Torrent PGM using the Ion PGM 200 Sequencing kit (Life Technology). The sequencing data were analyzed with CLC Genomics Workbench 4.8 (CLC bio, Denmark) as described [2]. Since our goal is to identify *cis*-acting regulatory regions for *CYP7A1*, we focused our analysis on chromosome 8. To identify robust interacting fragments, we selected only replicated signals from two independent experiments, and focused on signals with >100 reads (two-fold genome-wide noise background).

Chromatin immuno-precipitation (ChIP) assays:

ChIP was performed in primary human hepatocytes using ChIP-IT Express Enzymatic kit (Active Motif, Carlsbad, CA, USA) with anti-p300 antibody (Active Motif #61401) [2]. P300-enriched DNA fragments were measured with real-time PCR using SYBR Green, and fold enrichment was calculated relative to a negative control without the p300 antibody. Multiple primer pairs (Supplemental Table 3) were designed for each test region (surrounding the highest 4C signal), and results with the highest enrichment are shown. Data are from three donors, each performed in triplicates.

Deletion of enhancer regions using CRISPR-mediated genome editing in HepG2 cells [2]:

A lentiviral based single vector (LentiCRISPR V2) that simultaneously delivers Cas9, single guide RNA (sgRNA), and a puromycin selection marker engineered by the Zhang Laboratory [4],

was purchased from Addgene (http://www.addgene.org/). Oligonucleotides (20 bp in length) corresponding to guide RNA sequence were designed using an online tool (tools.genome-engineering.org) (Supplemental Table 3). Experimental details and analysis are published [2].

Quantitative analysis of allelic ratios in genomic DNA and mRNA using SNaPshot [2]: A fragment of DNA or RNA (after conversion to cDNA) surrounding a marker SNP was PCR amplified. The unincorporated dNTPs and excess primers were inactivated and degraded with exonuclease I and antarctic alkaline phosphatase. The PCR products were then subjected to a primer extension assay (SNaPshot, Life Technology) using a primer designed to anneal to the amplified DNA adjacent to the SNP site. After addition of a single fluorophore-labeled dideoxyribonucleoside triphosphate (ddNTP) complementary to the nucleoside at the polymorphic site, the resulting primer extension products were run on an ABI 3730 capillary electrophoresis DNA analyzer, and the data were analyzed using Gene Mapper software (Life Technology). To measure the expression ratios of the two alleles, we used heterozygous samples, which yield two differently labeled peaks with similar retention times. The peak area is proportional to the number of amplified alleles. To account for different fluorescence yields and migration rates, the cDNA ratios of heterozygous samples were normalized by setting the corresponding genomic DNA ratio = 1. SNaPshot assays were performed twice, each in duplicates. The association between allelic RNA expression imbalance (AEI) and heterozygosity of candidate SNPs was examined using K-analysis, which tests the agreement between AEI status (RNA ratio deviating from 1) and heterozygosity of each SNP.

Total CYP7A1 mRNA level: Total CYP7A1 mRNA was measured using real time PCR with specific primers (Supplemental Table 3) and SYBR Green PCR master mix (Life Technologies) using β-actin as the internal control [2].

Candidate SNP selection for association with allelic expression imbalance:

We used genotype data from the 1000 genome project to search for SNPs located in or nearby the R2 and R3 regions. We then searched current literature, GWAS hits (GRASP, GWAS catalogue), gene expression (GTEx), and functional genomics (ENCODE, etc.) information to prioritize candidate SNPs.

Genotyping: All SNPs were genotyped using multiplexed primer extension assays [2]. Genotyping primers are listed in Supplemental Table 3.

Cell culture and transfection: Primary hepatocytes were incubated for 24 hrs in serum-free William's E media supplemented with penicillin/streptomycin/fungizone (100 U/100 µg/0.25 µg per ml), 100 nM dexamethasone, 2 mM L-glutamine, 15 mM HEPES, and ITS (0.55 mg/ml human transferrin, 1 mg/ml bovine insulin and 0.5 µg/ml sodium selenite, from Sigma). Then cells were used for 4C (Hep10 and Hep8) and ChIP (Hep4, Hep8 and Hep10) assays. HepG2 cells were cultured at 37°C in a humidified incubator at 5% CO_2 in DMEM supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 µg/ml streptomycin. Transfection was performed using lipofectamine 2000 (Invitrogen Life Technologies, Carlsbad, CA, USA) and luciferase activity was measured with the Dual-Glo luciferase assay kit (Promega, Madison, WI, USA) on a Fusion Universal Microplate Analyzer (PerkinElmer Life and Analytical Science, Waltham, MA, USA) with *Renilla* luciferase constructs (Promega #E2231) driven by a TK promoter as the internal control. Data shown are from three independent experiments, each with duplicates.

Reporter gene assays: Regulatory regions (R2a, R2b, R2aR2b and R3) were PCR amplified and cloned into the pGL4.23 vector, using the In-Fusion HD cloning kit (Takara Bio USA,

California, USA) (see Supplemental Table 3 for PCR primers and cloning sites). Enhancers with different genotype combinations were generated by PCR amplification from gDNA with corresponding genotypes/haplotypes. *CYP7A1* promoter regions (~700 bp) carrying rs3808607 *T* or *G* alleles were also PCR amplified from gDNA and cloned into pGL3 vector. To test combined effects of enhancer and promoter, we joined enhancer fragments to the 5' end of the *CYP7A1* promoter fragment in the pGL3 vector. All constructs were sequenced to ensure the absence of random mutations during PCR amplification. To avoid variability from colonized bacterial clones, the plasmids were re-transformed into DH5α competent cells, and for each construct, three clones were selected for plasmid DNA preparation. Cells were transfected with 1 μg plasmid DNA and luciferase activity measured 48 h later with Dual-Glo luciferase assay kits (Promega, Madison, WI, USA) on a Fusion Universal Microplate Analyzer (PerkinElmer Life and Analytical Science, Waltham, MA, USA). As internal control, *Renilla* luciferase constructs (Promega #E2231) driven by a TK promoter were co-transfected with promoter constructs at a 1:5 ratio.

Data analysis:

Data are shown as mean ± SD. Statistical analyses were performed using Prism (GraphPad Software, San Diego, CA, USA).

Association between CYP7A1 genotype and liver gene expression: A multiple linear regression model was used to test the associations between *CYP7A1* genotypes and liver gene expressions. Log transformed CYP7A1 mRNA expression data followed a normal distribution (P=0.61, one-sample Kolmogorov-Smirnov test). Since the sample size is small in at least one of the genotype groups (n=2), it is impossible to test for normality within each group shown in Figure 3b-3d.

Therefore, we also performed a non-parametric Kruskal-Wallis test. The p values for Figures 3b, 3c, and 3d are 0.03, 0.07, and 5.1E-06, respectively, consistent with multiple linear regression analysis.

Association between the CYP7A1 genotype-defined CYP7A1 activity status and clinical phenotypes: A multiple linear regression model was used to test the associations between the genotype-defined CYP7A1 activity status, and lipid levels using Minitab software. We used forward and backward stepwise regression to select the best set of predictors in the multiple linear regression models with a cutoff p-value ≤ 0.05. The association between inferred CYP7A1 activity status and statin goal reaching, risk of CAD, hypertension, diabetes, and hypertension was analyzed using a logistics regression model, adjusting for covariates.

Statistical analysis in CATHGEN, diseased vessels: The association between the number of diseased vessels and reduced activity alleles was assessed using a logistic model (with the indicator of the event that the number of vessels exceeded 1 as the dependent variable) and age, smoking status, and sex as additional predictors. This analysis was done in participants of Caucasian ancestry only.

Statistical analysis in the Framingham cohort: Owing to an extensive family structure, the association analysis between myocardial infarctions and the sum of reduced activity alleles was performed following a Generalized Linear Mixed Model framework as proposed by Chen et *al*. [5] and implemented in the 'GENESIS' package in R. Thus, the Genetic Relatedness Matrix was estimated from the genome-wide profiles of polymorphisms, and a logistic mixed model was fitted, with myocardial infarction status as the dependent variable and age (at recruitment), sex, BMI at the first visit, number of times that the given person was recorded to take lipid lowering

drugs, and sum of reduced activity alleles as independent variables. We assessed statistical the significance of predictions be means using the Wald's test.

Acknowledgements

Framingham Heart study and CATHGEN:

The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI. Additional funding for SABRe was provided by the Division of Intramural Research, NHLBI, and Center for Population Studies, NHLBI. Funding for SHARe Affymetrix genotyping was provided by the NHLBI Contract N02-HL- 64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University. The data used for the analyses described in this manuscript were obtained from the dbGaP accession number phs000007.v23.p1 on 21 July 2014.

For CATHGEN, clinical data originated from the Duke Databank for Cardiovascular Disease (DDCD) and biological samples originated from the Duke Cardiac CATHeterization (CATHGEN) study. Funding support for the Genetic Mediators of Metabolic CVD Risk was provided by NHLBI grant RC2 HL101621 (William E. Kraus). The data used for the analyses described in this manuscript were obtained from the dbGaP accession number phs0000703.v1.p1 on 25 March 2015.

These analyses were made possible by computing time provided by the Ohio Supercomputer Center, GRANT #: PAS0885-2 PROJECT: COLLABORATION ENVIRONMENT FOR

PHARMACOGENOMICS. Liver Tissue Cell Distribution System (LTCDS, Pittsburgh, PA) is funded by NIH Contract # HHSN276201200017C.

Availability of data and materials:

The datasets supporting the conclusions of this article are available in the dbGaP repository, phs0000703 ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000703(CATHGEN), phs000007 ftp://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000007 (FHS),

Supplemental References:

1. Simonis M, et al. An evaluation of 3c-based methods to capture DNA interactions. *Nat Methods*. 2007;4:895-901
2. Wang D, et al. Functional characterization of cyp2d6 enhancer polymorphisms. *Hum Mol Genet*. 2015;24:1556-1562
3. van de Werken HJ, et al. Robust 4c-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. 2012;9:969-972
4. Shalem O, et al. Genome-scale crispr-cas9 knockout screening in human cells. *Science*. 2014;343:84-87
5. Papp AC, et al. Cholesteryl ester transfer protein (cetp) polymorphisms affect mrna splicing, hdl levels, and sex-dependent cardiovascular risk. *PLoS One*. 2012;7:e31930
6. Teslovich TM, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713
7. Willer CJ, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283
8. Surakka I, et al. The impact of low-frequency and rare variants on lipid levels. *Nat Genet*. 2015;47:589-597
9. Joshi AD, et al. Four susceptibility loci for gallstone disease identified in a meta-analysis of genome-wide association studies. *Gastroenterology*. 2016;151:351-363 e328
10. Shin SY, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46:543-550

11. Kooner JS, et al. Genome-wide association study in individuals of south asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet*. 2011;43:984-989

12. Kim HJ, et al. Common cyp7a1 promoter polymorphism associated with risk of neuromyelitis optica. *Neurobiol Dis*. 2010;37:349-355

13. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44:981-990

Supplemental Table 1. Association between *CYP7A1* SNPs and clinical phenotypes in GWAS.

#	Disease/Trait	SNP	LD with GWAS SNPs and functional SNPs (r2/D')				P value	Ref
			rs3808607		rs9297994			
			EUR	AA	EUR	AA		
1	Total cholesterol	rs2081687	0.66/0.92	0.05/0.48	0.93/0.97	0.11/1	9E-13	[6]
2	Total cholesterol	rs2081687	0.66/0.92	0.05/0.48	0.93/0.97	0.11/1	9E-12	[7]
	Total cholesterol	rs4738684	0.72/0.95	0.02/1	1/1	1/1	3E-11	[8]
3	LDL cholesterol	rs9297994	0.72/0.95	0.02/1	1/1	1/1	2E-11	[8]
4	LDL cholesterol	rs2081687	0.66/0.92	0.05/0.48	0.93/0.97	0.11/1	4E-9	[6]
5	LDL cholesterol	rs2081687	0.66/0.92	0.05/0.48	0.93/0.97	0.11/1	1E-7	[7]
6	gallstone	rs6471717	0.66/0.92	0.01/0.62	0.92/0.97	0.13/1	8.8E-9	[9]
7	Deoxycholate blood level	rs8192870	0.79/1	0.11/1	0.91/0.95	0.19/0.94	4E-8	[10]
8	Type 2 diabetes	rs16923500	0.11/0.75	0.00/0.03	0.12/0.87	0.02/1	2.8E-5	[11]
9	Neuromyelitis optica	rs1457043	0.93/1	0.98/0.99	0.66/0.95	0.02/1	1.5E-5	[12]
10	Type 2 diabetes	rs8192870	0.79/1	0.11/1	0.91/0.95	0.19/0.94	8E-3	[13]

Note: EUR, European; AA, African American; Ref, references.

Supplemental Table 2 4C signal. Chromosome position is based on GRCh37/hg19 assembly

chromosome position	FDR (%)	# Reads	5' gene
58118873..58118973	2.31E-14	110	IMPAD1 (212443)
58122203..58122274	1.19E-10	112	IMPAD1 (215773)
58124246..58124263	6.74E-03	304	IMPAD1 (217816)
58416869..58416969	1.24E-04	178	RPL30P10 (111428)
58438010..58438140	2.47E-16	159	RPL30P10 (132569)
58853995..58854017	3.16E-05	374	LOC286178 (191239)
58861802..58861902	6.27E-07	125	LOC286178 (199046)
58882086..58882189	1.59E-13	108	LOC286178 (219333)
59314439..59314539	1.91E-09	146	RPL26P26 (290924)
59359558..59359658	5.42E-18	132	LOC100421822 (20314)
59367804..59367906	2.24E-26	183	LOC100421822 (28560)
59367806..59367907	6.22E-06	198	LOC100421822 (28562)
59369140..59369158	5.40E-29	249	LOC100421822 (29896)
59375795..59375897	1.15E-28	204	LOC100421822 (36551)
59380550..59380923	2.01E-32	475	LOC100421822 (41306)
59381932..59382051	2.48E-56	448	LOC100421822 (42688)
59382457..59382557	3.56E-13	105	LOC100421822 (43213)
59382704..59382939	0.01	338	LOC100421822 (43460)
59384478..59385507	1.36E-53	550	LOC100421822 (45234)
59386733..59386847	3.68E-19	184	LOC100421822 (47503)
59387280..59387360	7.36E-10	127	LOC100421822 (48036)
59390664..59391750	8.08E-58	858	LOC100421822 (51420)
59391834..59393000	4.35E-25	782	LOC100421822 (53950)
59393023..59393238	4.05E-20	791	LOC100421822 (53979)
59395001..59395101	6.44E-11	120	LOC100421822 (55757)
59395880..59395980	7.91E-07	206	LOC100421822 (56636)
59398034..59399034	1.08E-13	105	LOC100421822 (57290)
59399092..59399192	7.55E-15	267	LOC100421882 (59848)
59402632..59404864	6.40E-24	243	LOC100421822 (63388)
59406409..59406479	7.22E-27	225	LOC100421822 (67165)
59409438..59409446	2.30E-04	1449	LOC100421822 (70194)
59413772..59413851	0	6215	CYP7A1 (1052)
59419922..59420022	0.01	915	LOC100996389 (3294)
59421980..59422050	1.92E-59	680	LOC100996389 (5354)
59430669..59430769	3.13E-25	223	LOC100996389 (14043)
59434738..59434838	1.57E-49	338	LOC100996389 (18112)
59445312..59445488	1.15E-18	256	LOC100996389 (28689)
59446081..59446181	6.63E-61	556	LOC100996389 (29455)
59448890..59448990	6.48E-12	186	LOC100996389 (32264)
59512895..59512995	8.63E-20	143	SDCBP (17476)
59516055..59516064	1.32E-03	134	SDCBP (20636)

Supplemental Table 3. Sequence of gRNAs and primers		
		hg19 assembly
gRNAs	**Sequence**	chromosome position
R1 5'end, #1	GCATTGTGGGCAGTTGGTTA	59385131..59385050
R1 3'end, #1	GGTTGTGTGCCATAAATAGG	59385430..59385449
R1 3'end, #2	AACTGAAACCCTATACCAAT	59385911..59385930
R2a 5'end, #1	CGGAAATCTCAACGGAATTC	59390273..59390292
R2a 5'end, #2	GAGGTCTGAGCCACGGAAAT	59390261..59390280
R2a 3'end, #1	AGAGCGTTAAGCAATCCCTT	59391062..59391081
R2a 3'end, #2	CATAACAACCCTCTGAGTTC	59391150..59391169
R2b 5'end, #1	TTTAGTTTGCCGTCTAGAGT	59392832..59392851
R2b 5'end, #2	GCCTCACTTTTAGTTTGCCG	59392824..59392843
R2b 3'end, #1	ACCTTGTGATCTGCCCGCCT	59393511..59393530
R2b 3'end, #2	GGGTTCAAGAGATTCTCGGG	59393373..59393392
R3 5'end, #1	GCGTTTGCAACTCACTAGAG	59398535..59398554
R3 3'end, #1	GGAGCACCAAATGTTTCACC	59398899..59398918
none target control gRNA	GCTTAGTTACGCGTGGACGA	
4C inverse primers		
forward	TCTGATACCTGTGGACTTAGTTCAAGG	
reverse	GCTTTGCCAGAGAGACGGTG	
ChIP qPCR primers		
R1		
primer 1, forward	TGCCCTGCTGTCACCTATTTATG	
primer1, reverse	CCCTGTGGATGGAAGCCTC	
primer 2, forward	AAGAGGAATGGATCCCCAAAA	
primer2, reverse	TACCCAAATGTCAGTACTATGCCATAC	
R2		
primer 1, forward	TAAGAAGTTGGCGGTTTGGC	
primer1, reverse	GGCTGGGAATGTAAAAACAAGAGA	
primer 2, forward	TGAAGGGCAAAGTTAAGAGATTTAGC	
primer2, reverse	CTGATGTCCCCATTCCTATTCC	
primer 3, forward	GGAAACCAAGGCATAGAGAGAGC	
primer3, reverse	TCCTCCACTTCCTAGCCTCATG	
primer 4, forward	GCTCTGTCACCAGGCTGGAG	
primer4, reverse	GAAGTTGCAGTGAGCCGAGAT	
R3		
primer 1, forward	CTGAGCACTAGCCAGCTGTGTT	
primer1, reverse	ATTGGGAAAGAGCTACACACAGAA	
primer 2, forward	TGGCACACTGTTACCTGGAGC	
primer2, reverse	CACGGGATATCATCTCGGTGA	
primer 3, forward	AACTGTCATTGATTTAATCCATGAGG	
primer3, reverse	AACGCTTAACAGTTCTGGAGCAA	
negative control region		
forward primer	GGTGCCAAATAAGCAGTGCATA	
reverse primer	TGATCTCTCCAGCCTCCAAATTC	
Cloning primers		
promoter, add XhoI site	CAGATGCCAAATTGTTACTAGTGGTT	59407935..59407961
	CTGATTAGAAAGGGAAGGATGCCA	59408761..59408785
R1, add KpnI site	TGGAGTTGTATACCCAAATGTCAGTACT	59384316..59384943
	CAAACCAAACAGAACAAAGAGAGGA	59385906..59385932
R2a, add KpnI site	AATTACAACAGCTTCCAGGTGATG	59390324..59390347

	GTCAACAATTTCAGTCCCCACAC	59391351..59391372
R2b, add KpnI site	GTCCCCCAGTGCTCATAACATT	59392510..59392531
	CAAGGACACAACGGCTGCTT	59393665..59393684
R3, add KpnI site	AAGACTGTGTCTGCTTCCCCC	59398237..59398257
	CGCCACTATCACCAACAAGACA	59399129..59399150
AEI primers		
forward PCR primer	ACTCGAGAGGCCAAAGAAGG	
reverse PCR primer	TCACACCCGAAGAACCCC	
Snapshot primer	CAGCTTAGTGAGATCCCGTCTCC	
CYP7A1 quantitation primer		
forward	CTGAATGACCTGCCAGTATTAGATAGT	
reverse	AAAGCCTCAGCGATTCCTTG	
Genotyping		
rs3808607	PCR primers	SNaPshot primer
Forward	CAGATGCCAAATTGTTACTAGTGGTT	CGAATGTTAAGTCAACATATATTTGAGAGA
Reverse	CTGATTAGAAAGGGAAGGATGCCA	
rs9297994		
Forward	GTATGTGTCCATTTGCGATCTTCT	GCTGAGGGACAATAATATGATCTTGTTT
Reverse	GATGACTCACCGATTAAATACGTCC	
rs10107182		
Forward	GCCAACTGAGGTTTGTGTACATG	CATATTGCAGCAGTATTTTATAAATTTAGTAGGTC
Reverse	AGTGAGGCAGGCCAAAAGAAT	
rs10504255		
Forward	TTGTCACGCAGCGTGGG	CAGCTGTGTTCTCAATTCTGTGTGTA
Reverse	GCAAACGCTGGTTGCTTTACTAG	
rs2081687		
Forward	GGAACAAAAGTGGAGGAATCACA	GAACAAAAGTGGAGGAATCACACT
Reverse	GGCTCTCTATTCTGTTCCATTGCT	
rs4738684		
Forward	GTTTTTAATTTGGAGCACTGAGCC	GAGTGATGATAAGGGCTAAAGAGTTGT
Reverse	GAAGTTGCAGTGAGCCGAGAT	
rs13263105		
Forward	CTCAGGAGAAAGATGCTATAACTTTCC	GCCAAGAATCAGGTGGTCATTAA
Reverse	TTTGATCCACTGGATTAAGACTGC	
rs7005978		
Forward	AGACATTCAACACCAGCCTATGAAG	GAGCCCACCTCTTGCATCAAT
Reverse	AATGGGAGAAATTGGCAAAAAC	
rs983812		
Forward	GAACTTTCTCCACACTCCAGCC	CATGAGGACCCAAGATTGGTTCTAT
Reverse	TGTTGACTTTGGAAGCTTTGTAGC	
rs7007181		
Forward	TTGGGTGAAACCATCTCTACAATGT	GGCTCCGTGACAGACCCG
Reverse	GAAGGAGTTTAAATAACTTGCCAATTG	
rs6985620		
Forward	GTACCCACAAGCATCTCAATCCT	GATGTAATGAAAATAGCAAAAAGAAGCCT
Reverse	TGAACCATTAAAAGTGGTAAAAAGTGG	
rs8192870		
Forward	CTTTAAATTGCTCTTTGATTCATCCAT	ATCATTCTATTACTGTTTTTAAAT
Reverse	AAGAAACTGCAAAAGGCAGAAAA	
rs4738679		
Forward	GTACCCACAAGCATCTCAATCCT	TTAAATTAGGAGGAGCTACCTTAATTTAATCT
Reverse	TGAACCATTAAAAGTGGTAAAAAGTGG	
rs3824260		
Forward	GTGCTTTGCCAGAGAGACGG	CAGAGGAAAGAGAACTGGGAAAAAC
Reverse	AAGCACTGAAACATGAAGCAGC	
rs7018333		

Forward	TGCCCAACATGGTGAAACC	GATATGAAAATTAGCTAGGCATGGTG
Reverse	ACTGCCACCTCCTGAGTTCAA	
rs3808609		
Forward	GCGCAGCACCACACCTAACT	ACTTGCTGCCCTAGTCTCACTCTC
Reverse	CCGTGCCACCCACTCACTA	
rs13251671		
Forward	TACATTGCCACCAACAATGTACAAG	CTAGCTTTTGTCTTTTGATAATAGCCATCTTA
Reverse	CAAAAGGCTAACAGGTATATGAAAAGG	
rs7845104		
Forward	AAGGAGGGCAAGGCTTCTTC	GCTTTATTTCTGGCCTAACAAGCC
Reverse	TGGACAATGAAATTGTGGCAG	
rs34828061		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	GACAAGGGCCCCCTCTCC
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs17202790		
Forward	AATGTCCTTTATGGCTGAGTTCATTA	CTGAGGATCAGGAAGGAAAGGAT
Reverse	CCCCATGGAAAAGGCCAG	
rs34231701		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	TTTTTTTGCAGCCTCAGGGCATGGTT
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs6471720		
Forward	CAAGGAGGGCAAGGCTTCTT	CTCTCCTTGAGTCTGGGCTGG
Reverse	AGTATTACTCCTCTCATTGAGAAGTGGA	
rs10105411		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	CTCCAGCCCCTGGTAACCACT
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs10086874		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	GTTCTCCTGGTATATCCATATTGTTGTAAATAA
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs10087236		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	CATCATACTGTCTTCATTTTCTTTGGATATA
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs113895159		
Forward	TGACGGTCCCTGGGCC	CCCGCGCGTTTCGGG
Reverse	CGTGGCCAGAGGCTCTTTC	
rs59985577		
Forward	CTTGTTGAGAACTGGAGTAAAGGTCA	TTTTTTTGAAAATGTCTTCAGGGCATGTCA
Reverse	CACACTGCCTTAACAGAGGTTTTC	
rs28612536		
Forward	AAACCAAGGCATAGAGAGAGCTAAGA	GAGGCTAGGAAGTGGAGGAGCAG
Reverse	GGGCGTGAACAGCAGAAGTC	
rs10435592		
Forward	ATCTCGGCTCACTGCAACTTC	CCGGCTAATTTTTTTGTATTTTTAGTAGAGA
Reverse	AGCACTTTGGGAGGCCG	
rs28514538		
Forward	ACCTCTCTGTGTCTCTGAGCCCT	AGCGTGGGCCTGAGCAC
Reverse	GTTATTTGTGCGGAGGCGTT	
rs6471722		
Forward	TAAGAAGTTGGCGGTTTGGCT	CTTGTTTTTACATTCCCAGCCTGAC
Reverse	AGCTATGCCTCCCTAACAGACCT	
rs10087499		
Forward	AGGACACTTGGGTTGTTTCCA	CAACATGGGAGCGCAGG
Reverse	GTGGGAAGGTAAATTAGTACAGCCA	
rs62512933		
Forward	AGCGCACCAGCATGGC	GTTGAAGCCACCTGCTCCC
Reverse	GGGCTGACGGCACCATAC	

Supplemental Table 4. Association between AEI and candidate SNPs, using k analysis.

SNP#	rs ID	K	P	region	chr posi, hg19	bp change
1	rs9297994	0.765	<0.0001	R2b	59392324	G>A
2	rs10107182	0.765	<0.0001	R2b	59392737	C>T
3	rs10504255	0.765	<0.0001	R3	59398461	G>A
4	rs2081687	0.726	<0.0001	GWAS SNP	59388565	T>C
5	rs4738684	0.726	<0.0001	R2b	59393273	A>G
6	rs13263105	0.687	<0.0001	down R1	59325595	C>G
7	rs7005978	0.687	<0.0001	down R1	59382715	A>G
8	rs983812	0.687	<0.0001	R1	59384181	C>T
9	rs7007181	0.569	<0.0001	down R1	59339279	T>C
10	rs6985620	0.569	<0.0001	down R1	59370159	T>C
11	rs8192870	0.567	<0.0001	promoter	59412066	T>G
12	rs4738679	0.531	<0.0001	down R1	59370320	G>A
13	rs3808607	0.45	<0.01	promoter	59412924	G>T
14	rs3824260	0.448	0.001	promoter	59413190	A>G
15	rs7018333	0.302	0.006	3'end	59402570	A>G
16	rs3808609	0.255	0.068	up prom	59465377	G>C
17	rs13251671	0.058	0.676	up R1	59387855	A>G
18	rs7845104	0.02	0.886	R1	59384671	A>G
19	rs34828061	0.018	0.895	down R1	59382169	A>G
20	rs17202790	0.018	0.895	up R1	59386544	T>A
21	rs34231701	-0.018	0.895	down R1	59382205	C>A
22	rs6471720	-0.02	0.886	R1	59384675	G>A
23	rs10105411	-0.02	0.886	up R1	59387357	G>A
24	rs10086874	-0.02	0.886	up R1	59387495	A>T
25	rs10087236	-0.02	0.886	up R1	59387670	T>C
26	rs113895159	-0.022	0.877	up prom	59466106	T>C
27	rs59985577	-0.493	0.001	down R1	59382092	G>T
28	rs28612536	-0.562	<0.0001	R2a	59391255	A>G
29	rs10435592	-0.645	<0.0001	R2b	59393469	T>C
30	rs28514538	-0.645	<0.0001	R3	59398432	A>G
31	rs6471722	-0.683	<0.0001	R2a	59390540	A>G
32	rs10087499	-0.683	<0.0001	R2b	59394969	T>G
33	rs62512933	-0.686	<0.0001	R2b	59397262	G>C
marker	rs8192879			3’UTR	59403576	A>G

Supplemental Table 5a. Basic statistics of the OSU CAD cohort

	Higher activity of CYP7A1, n=87			**Lower activity of CYP7A1, n=398**				P value
	Copy # of reduced activity allele			**Copy # of reduced activity allele**				
	0	**1**	**0+1**	**2**	**3**	**4**	**2+3+4**	
	Count		count (%)	Count			count (%)	
Sex(male)	12	45	57(65.5%)	252	9	2	263(66.1%)	>0.05
Race(white)	3	47	50(57.5%)	364	10	2	376(94.5%)	**<0.000**
Hypertension	13	43	46(64.3%)	266	10	1	277(69.6%)	>0.05
Diabetes	8	24	32(36.8%)	160	4	0	164(41.2%)	>0.05
Family history	10	29	39(44.8%)	170	5	1	176(44.2%)	>0.05
Tobacco use	13	40	53(60.9%)	204	6	1	211(53.0%)	>0.05
On statin	9	29	38(43.6%)	226	8	0	234(58.8%)	**0.014**
MI	7	22	29(33.3%)	109	4	0	113(28.4%)	>0.05
	mean ±SE		mean ± SE	mean ±SE			mean ± SE	
Age	57±2	60±1	60±1	62±1	63±5	67±10	62±1	>0.05
Total cholesterol level	178±6	167±5	169±4	165±3	160±18	203*	166±3	>0.05
HDL level	36±3	38±2	37±1	35±1	34±3	27*	35±1	>0.05
Triglycerol level	132±16	156±16	151±13	170±11	131±22	234*	169±10	>0.05
LDL level	116±6	99±4	102±4	98±2	100±16	129*	99±2	>0.05

* single value

Supplemental Table 5b. Basic statistics of CATHGEN cohort

	Higher activity of CYP7A1, n=306			**Lower activity of CYP7A1, n=898**			P value
	Copy # of reduced activity allele			**Copy # of reduced activity allele**			
	0	**1**	**0+1**	**2**	**3**	**2+3**	
	Count		count (%)	Count		count (%)	
Sex (male)	35	122	157(51%)	601	7	608(68%)	**<0.0001**
Race (white)	3	86	89(29%)	739	6	745(83%)	**<0.0001**
Hypertension	63	170	233(76%)	572	5	577(64%)	**<0.0001**
Hypercholesterolemia	46	133	179(60%)	556	6	562(63%)	>0.05
Diabetes	38	85	123(41%)	272	2	274(31%)	**0.002**
Tobacco use	30	101	131(43%)	418	4	422(47%)	>0.05
MI	54	136	190(62%)	662	5	667(74%)	**<0.0001**
Death	16	50	66(22%)	249	1	250(28%)	**0.024**
	mean ± SE		mean ± SE	mean ±SE		mean ± SE	
Age	60 ± 1	60 ± 1	60±12	63±0.3	55±4	63±12	**0.002**
CADINDEX	23±3	25±2	24±1	34±1	39±11	34±1	**<0.0001***
BMI	31±1	30±0.5	31±7	30±0.2	28±2	30±7	>0.05

Supplemental Table 5c. Basic statistics of Flamingham cohort

	Rapid metabolizer, n=15			**Slow metabolizer, n=1873**				P value
	copy # of reduced activity allele			**copy # of reduced activity allele**				
	0	**1**	**0+1**	**2**	**3**	**4**	**2+3+4**	
	Count		count (%)	Count			count (%)	
sex(male)	0	6	40.00%	766	77	2	45.11%	>0.05
Hypertension treatment	0	8	53.30%	655	85	1	39.56%	**<0.001**
Diabetes treatment	0	1	6.67%	123	17	0	7.47%	>0.05
on statin	0	7	46.67%	599	70	0	35.72%	**<0.001**
MI	0	0	0.00%	103	11	0	6.09%	**<0.0001**
	mean ± SE		mean ± SE	mean ±SE			mean ± SE	
age	0	33.5±7.8	33.5 ±7.8	34.7 ±9.1	34.5± 8.7	31.5 ±16.2	34.7± 9.1	>0.05
Total cholesterol	0	212.9 ± 21.4	212.9± 21.4	199.1± 26.5	198.1± 25.3	183.6± 45.4	198.9± 26.4	>0.05
HDL	0	54.1 ±12.5	54.1± 12.5	53.3± 14.3	52.4± 13.3	60.6± 22.6	53.2± 14.2	>0.05
Triglycerol	0	111.4± 42.0	111.4 ±42.0	121.4± 68.5	117.9 ±55.1	80.4± 23.2	120.9± 67.2	>0.05
LDL	0	136.5± 21.1	136.5± 21.1	121.9± 24.2	122.5± 22.7	106.9± 23.7	121.9± 24.1	>0.05

Supplemental Table 6. Demographic information of donors

Table 6a. Liver tissues

Sample ID	RACE	SEX	AGE
L01	W	F	49
L02	W	F	14
L03	W	M	74
L04	W	F	53
L05	W	F	63
L06	W	M	51
L07	W	M	53
L08	W	M	49
L09	W	F	62
L10	W	F	76
L11	W	M	83
L12	U	M	78
L13	U	U	U
L14	U	M	67
L15	W	M	48
L16	U	U	U
L17	W	M	68
L18	B	M	73
L19	W	M	56
L20	W	F	55
L24	W	F	57
L25	U	F	44
L26	U	F	67
L27	W	F	65
L28	W	F	65
L29	W	M	78
L30	W	M	71
L31	W	F	57
L32	W	F	76
L33	W	F	74
L34	B	M	78
L35	W	M	70
L36	W	F	59
L37	U	F	67
L38	W	M	56
L39	W	M	75
L40	W	F	80
L41	W	M	63
L42	W	F	58
L43	W	F	57
L44	W	F	76

L45	W	F	60
L46	W	M	69
L47	W	F	64
L48	W	F	60
L49	B	M	75
L50	W	F	54
L51	W	F	46
L52	W	F	55
L53	W	F	53
L54	W	M	50
L55	W	F	67
L56	W	M	69
L57	W	F	64
L58	W	F	48
L59	U	M	37
L60	W	M	79
L61	U	F	77
L62	W	M	63
L63	W	F	75
L64	W	M	66
L65	W	M	54
L66	W	M	80
L67	W	F	42
L68	B	F	63
L69	W	F	54
L70	W	M	72
L71	W	F	79
L72	U	F	65
L73	W	F	52
L74	W	M	64
L75	W	F	50
L76	W	M	73
L77	U	F	78
L78	W	M	72
L79	W	F	74
L80	W	F	63
L81	W	M	66
L82	W	M	71
L83	W	M	38
L84	W	M	66
L85	W	F	72
L86	W	F	58
L87	W	M	51
L88	W	F	58
L89	W	F	71
L90	B	F	78
L91	W	M	28

L92	W	M	51
L93	W	M	73
L94	W	F	19
L95	W	M	66
L96	W	F	61
L97	W	F	38
L98	U	M	68
L99	W	M	72
L100	W	F	67
L101	W	M	81
L102	B	M	71
L103	W	M	64
L104	U	F	66
L105	W	F	57
L106	W	F	53
L107	B	M	56
L108	W	F	44
L109	U	F	27
L110	W	F	58
L111	W	F	58
L112	W	F	39
L113	W	M	52
L114	W	F	67
L115	W	M	44
L116	W	M	73
L117	W	M	54
L118	W	F	46
L119	W	M	67
L120	W	M	58
L121	W	M	68
L122	W	F	65
L123	B	M	50
L124	W	M	65
L125	W	F	77

Table 6b. Human hepatocytes

	Sex	Age	Race	Disease and Treatment
Hep4	male	50	unknown	metastatic cancer, chenotherapy
Hep8	male	31	European	metastatic cancer, chenotherapy
Hep10	female	31	unknown	benign liver tumor

Supplemental Figure legends:

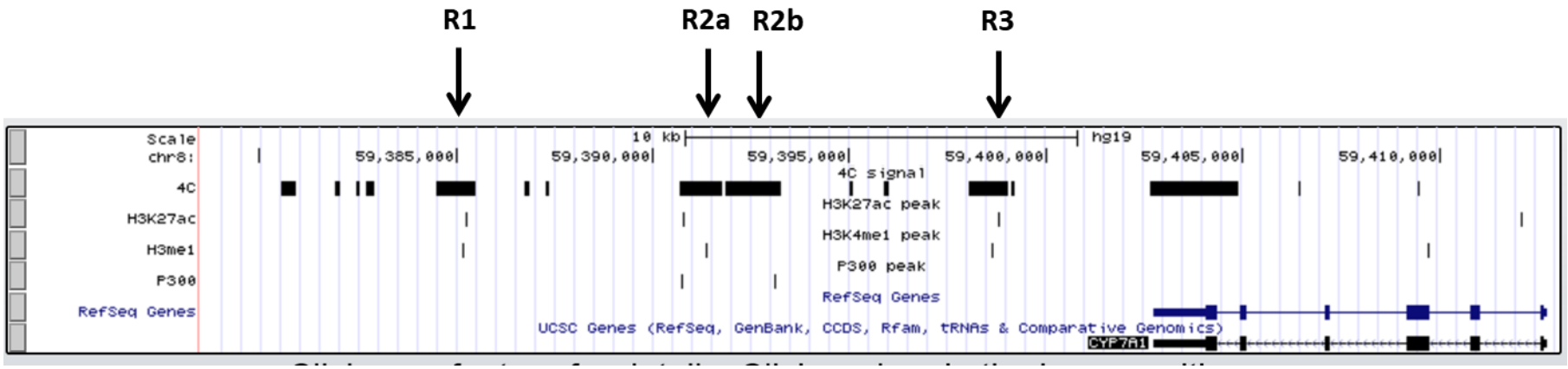
Supplemental Figure 1. Overlap of 4C signals with published ChIP-seq signals for H3K4me1, H3K27ac and P300 in human hepatocytes.

Supplemental Figure 2. ChIP-qPCR with p300 antibody in human hepatocytes performed with different sets of primers (see Supplemental Table 3 for primer sequence and location).

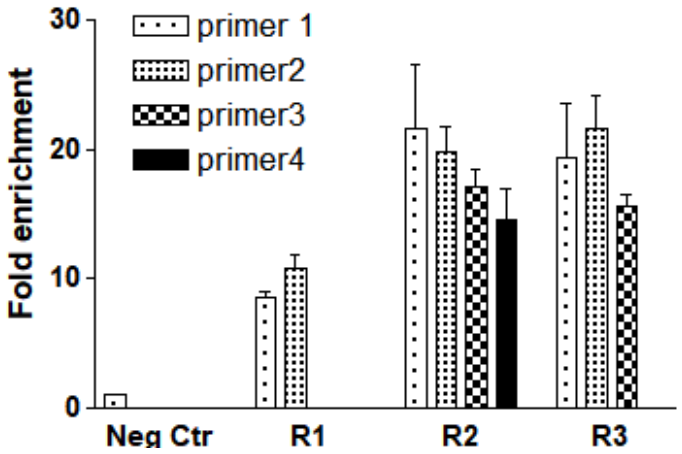
Supplemental Figure 3. Gel image of PCR products amplified from genomic DNA prepared from cells with combinations of two gRNAs targeting 5' and 3' of R1 (a), R2a/R2b (b) or R3 (c) regions. Successful deletion of a targeted region is shown as a band with lower molecular weight as compared to no-target control. See Supplemental Table 3 for sequence of gRNAs.

Supplemental Figure 4. CYP7A1 mRNA expression in HepG2 cells after CRISPR-deletion of regulatory regions, using different combinations of gRNAs.

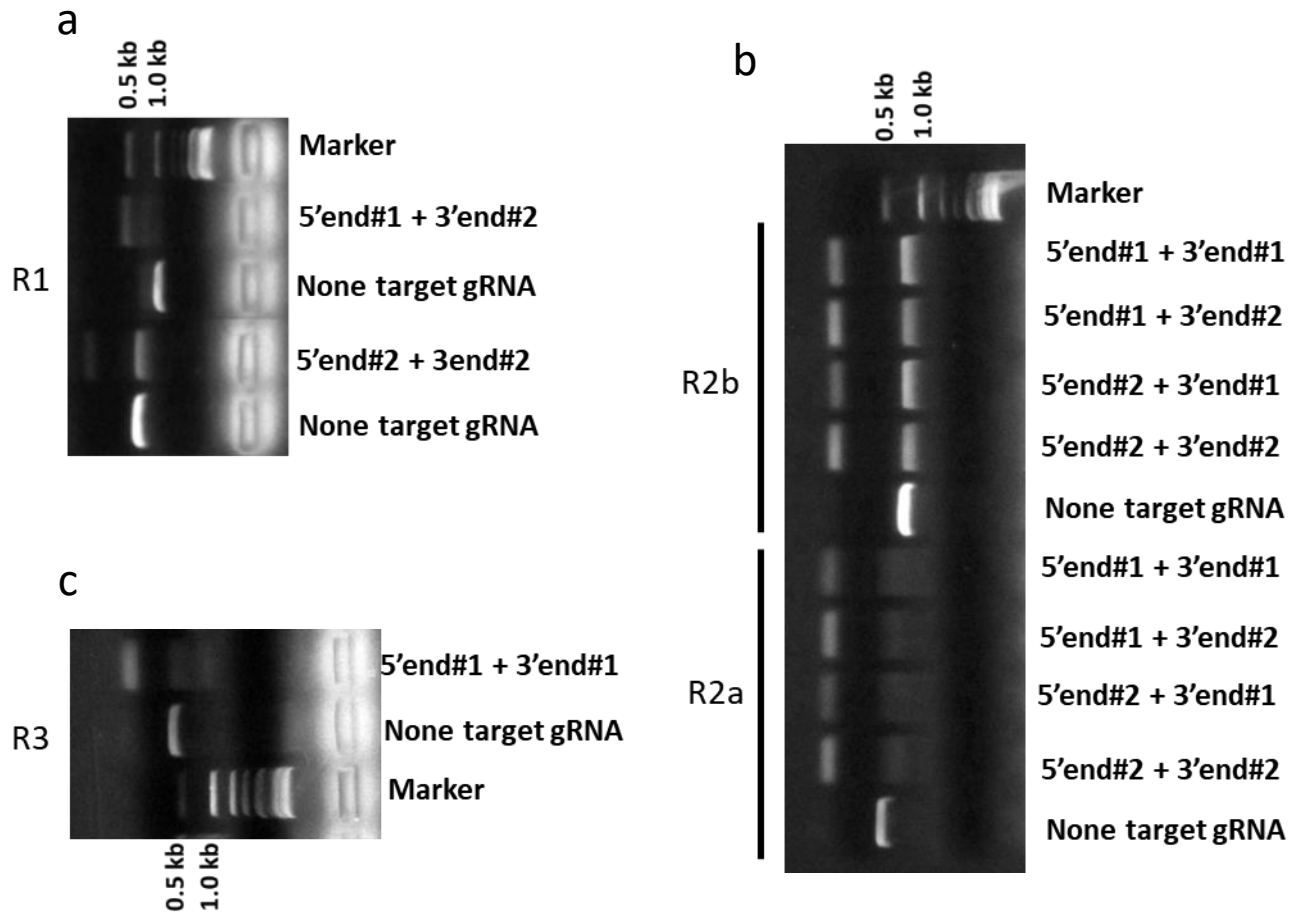
Supplemental Figure 5. Haplotype structure of CYP7A1. See Supplemental Table 4 for SNP rs numbers.



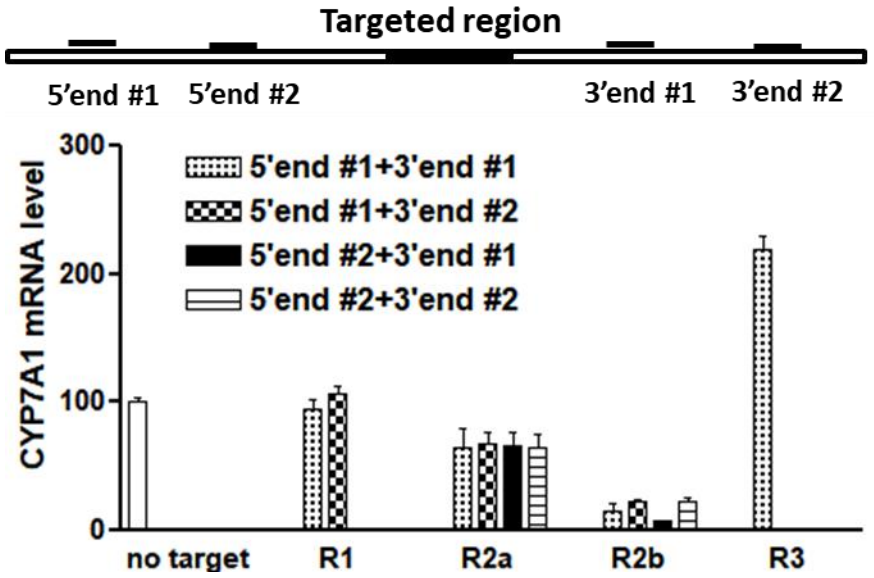
Supplemental Figure 1.



Supplemental Figure 2.



Supplemental Figure 3.



Supplemental Figure 4.

Supplemental Figure 5.