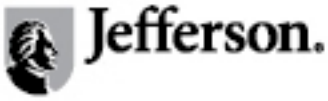


5-3-2017

Identification of a functional genetic variant driving racially dimorphic platelet gene expression of the thrombin receptor regulator, PCTP.

Xianguo Kong
Thomas Jefferson University, xianguo.kong@jefferson.edu

Lukas M. Simon
Baylor College of Medicine

Michael Holinstat
Thomas Jefferson University, michael.holinstat@jefferson.edu

Chad A. Shaw
Baylor College of Medicine

Paul F. Bray
Thomas Jefferson University, paul.bray@jefferson.edu

See next page for additional authors

Authors
Xianguo Kong, Lukas M. Simon, Michael Holinstat, Chad A. Shaw, Paul F. Bray, and Leonard C. Edelstein

Identification of a functional genetic variant driving racially dimorphic platelet gene expression of the thrombin receptor regulator, *PCTP*

Short Title: *PCTP* eQTL Functionalization

Xianguo Kong[1], Lukas M. Simon[2], Michael Holinstat[3], Chad A. Shaw[2,4], Paul F. Bray[1], and Leonard C. Edelstein[1,*]

1. The Cardeza Foundation for Hematologic Research and the Department of Medicine, Thomas Jefferson University, Philadelphia, PA 19107, USA.
2. Department of Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA
3. Department of Pharmacology, University of Michigan, Ann Arbor, MI 48109, USA
4. Department of Statistics, Rice University, Houston, TX 77251, USA

*Corresponding Author:

Leonard C. Edelstein

Department of Medicine Sidney Kimmel Medical College

Thomas Jefferson University

1020 Locust Street, Suite 394

Philadelphia, PA 19107

Phone: 215-955-1797, Fax: 215-955-9170

leonard.edelstein@jefferson.edu

Supported by United States National Institutes of Health Grants R01HL128234 and R01MD007880

What is known on this topic:

- Platelet activation in response to stimulation of the Protease Activated Receptor 4 (PAR4) receptor differs by race.
- Phosphatidylcholine Transfer Protein (*PCTP*), a regulator of PAR4 signaling, is differentially expressed by race.
- Genetic variants can affect transcriptional activity by altering regulatory regions of DNA.

What this paper adds:

- Identifies the variant rs2912553 as contributing to racially differential expression of *PCTP.*
- Identifies differential binding of a GATA1-containing complex dependent on rs2912553 genotype.

Summary

Platelet activation in response to stimulation of the Protease Activated Receptor 4 (PAR4) receptor differs by race. One factor that contributes to this difference is the expression level of Phosphatidylcholine Transfer Protein (*PCTP*), a regulator of platelet PAR4 function. We have conducted an expression Quantitative Trait Locus (eQTL) analysis that identifies single nucleotide polymorphisms (SNPs) linked to the expression level of platelet genes. This analysis revealed 26 SNPs associated with the expression level of *PCTP* at genome-wide significance ($P < 5x10^{-8}$). Using annotation from ENCODE and other public data we prioritized one of these SNPs, rs2912553, for functional testing. The allelic frequency of rs2912553 is racially-dimorphic, in concordance with the racially differential expression of *PCTP*. Reporter gene assays confirmed that the single nucleotide change caused by rs2912553 altered the transcriptional potency of the surrounding genomic locus. Electromobility shift assays, luciferase assays, and overexpression studies indicated a role for the megakaryocytic transcription factor GATA1. In summary, we have integrated multi-omic data to identify and functionalize an eQTL. This, along with the previously described relationship between *PCTP* and PAR4 function, allows us to characterize a genotype-phenotype relationship through the mechanism of gene expression.

Keywords

platelets, genetics, PAR4, PCTP, eQTL

Introduction

Genome wide association studies (GWAS) seek to find associations between regions of the genome and normal or pathological traits. There are several challenges associated with GWAS including the requirement of large sample sizes and/or effect sizes to achieve statistical significance. One approach to overcome this challenge is to first identify associations between gene expression and the phenotype of interest using expression profiling analysis. This is followed by expression quantitative trait loci (eQTL) analysis, which uncovers associations between genetic variants and gene expression. This indirect approach can lead to the identification of genetic variants that affect physiology in a manner that requires smaller sample sizes than traditional GWAS (1).

Another challenge in GWAS is deciphering the mechanism by which a genetic variant exerts its effect on the phenotype. This is complicated by the fact that 98% of SNPs identified in GWAS are located in non-protein coding portions of the genome. A relationship between genotype, gene expression, and phenotype indicates that the variant acts through a transcription-based mechanism. Efforts by the ENCODE project and others have begun to provide annotation to non-coding regions of the genome, identifying DNA-binding proteins, chromatin modifications, and other epigenetic marks located throughout the genome (2, 3). Localization of eQTLs within these annotated regions facilitates prioritization for testing, experimental design, and variant functionalization.

We have previously conducted a gene expression profiling study to identify platelet RNAs that are differentially expressed by demographic variables or associated with platelet physiological phenotypes, called the Platelet RNA And eXpression (PRAX1) study (4, 5). Using this approach, we were able to identify the expression level of the *PCTP* gene, encoding Phosphatidyl Choline Transfer Protein (PC-TP), as being positively associated with platelet reactivity in response to Protease Activated Receptor 4 (PAR4) thrombin receptor stimulation.

PCTP RNA was also more highly expressed among self-identified black subjects as compared to self-identified whites, in concordance with our finding that PAR4 stimulation induced platelet aggregation more robustly in blacks than whites. The role of PC-TP protein in regulating PAR-4 function has been confirmed both by pharmacological inhibition and by siRNA silencing (4).

By performing an eQTL analysis of the PRAX1 cohort, several single nucleotide polymorphisms (SNPs) were identified as associated with *PCTP* RNA levels (6). We now report identification of the mechanism by which a common genetic variant contributes to differential *PCTP* expression. Using a combination of publically available data, reporter gene assays and electromobility shift assays (EMSA), we have confirmed that SNP rs2912553 confers differential transcriptional potency to the locus surrounding it. rs2912553 genotype results in differential recruitment of a GATA-1 containing complex which regulates transcription at this locus.

Materials and Methods

The Platelet RNA and eXpression Study 1 (PRAX1)

154 healthy individuals (80 of European and 74 of African ancestry) were recruited between 2010 and 2011. All had platelet functional phenotyping, demographic information, genome-wide genotyping and global mRNA profiling performed. Informed consent was obtained from all participants with the approval of the institutional review boards of Baylor College of Medicine in Houston, TX, and Thomas Jefferson University in Philadelphia, PA. Research was conducted in accordance with the Declaration of Helsinki.

DNA from the buffy coats of PRAX1 subjects was hybridized to the HumanOmni5 array (Illumina Inc., San Diego, CA) as described earlier (4) and genotyped for approximately five million markers. Genotype data was restricted to 1927172 markers passing a cohort-specific minor allele frequency cutoff of 5%. Gene expression of PRAX1 samples was measured using the Human Gene 1.0 ST microarray (Affymetrix, Santa Clara, CA) as described previously (5) and can be downloaded from the Gene Expression Omnibus, accession no, GSE49921.

Expression Quantitative Trait Analysis

eQTL analysis was performed using the R statistical software (ref). We used cohort genotype and microarray based expression data to identify significant eQTL associations within 500kb of *PCTP* transcription start and stop sites. Associations were modeled using variant allele dosage (0, 1, and 2) as the independent and *PCTP* expression levels as the dependent variable in a linear regression framework as implemented in the *aov* R function (7). To account for confounding variables we included age, gender and race in the regression model. Additional details on our eQTL mapping have previously been reported (6). To assess the association between PAR4 ARS and genotype, a multivariable linear regression model was used as implemented with the R function lm.

Reporter Gene Assays

Luciferase reporter plasmids were constructed as follows: 32bp, 43bp and 56bp oligonucleotides containing sequence surrounding the rs2912553 SNP were synthesized and annealed (sticky NheI and BglII sites were formed after annealing), and ligated into pGL4.28 luciferase vector (Promega) digested with NheI and BglII. 130bp, 209bp and 454bp fragments containing sequence surrounding rs2912553 were generated by PCR using human genomic DNAs from rs2912553 T or C homozygous individuals as template. NheI and HindIII sites were incorporated into the 5' or 3' end of primers respectively. Insert and vector were digested with NheI and HindIII and ligated with T4 ligase. All the plasmids were validated by DNA sequencing. Detailed primer information is shown in Table 1.

For the assays, 2 μg of reporter plasmid was co-transfected with 0.5 μg β-gal expression plasmid into $5x10^5$ K562 chronic myelogenous leukemia cells with Lipofectamine 2000 (ThermoFisher, Waltham, MA) based on the manufacturer's instructions. For some experiments 40 ng siRNA (ThermoFisher) directed against a transcription factor or scrambled control was co-transfected. 24 hours post-transfection luciferase activity was analyzed on a FLUOstar OPIMA plate reader. β-gal activity was quantified with a β-gal assay kit (ThermoFisher). The data were presented as Luciferase/β-gal.

Electromobility Shift Assays (EMSA)

65 bp EMSA probes were generated by PCR using the reporter vectors described above as templates and 5' IRDye700 labeled oligonucleotide primers (IDT DNA, Coralville, IA) that surround rs2912553. The primer sequences were: 5'-GTTGGCCTTTCTAAGCAT-'3 and 5'-ATGTAGATAGGGAAGTTGGTG-3'. . The probes were then gel purified and quantified. K562 nuclear extracts were made with the Subcellular Protein Fractionation Kit for Cultured Cells (ThermoFisher). Probes and nuclear extract were allowed to bind in binding buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 7.5 mM $MgCl_2$, 1 mM EDTA, 5% glycerol (v/v), 5 mM DTT, 0.1 % NP-

40 (v/v), 0.5 mg/ml BSA, and 500 ng/ml poly (dI-dC)) for 15 minutes on ice. For some experiments, 4 µg super shift antibodies for GATA1 or GATA2 (Santa Cruz Biotechnology, Dallas, TX) were added to the binding reaction. After binding, the mixtures were resolved on 5% nondenaturing polyacrylamide gels, and scanned using an Odyssey Infrared Scanner (Li-Cor, Lincoln, NE). The Licor Odyssey software was used for quantification of the shifted bands.

Immunoblots

5 µg pCMV-3FLAG-GATA1 plasmid or empty vector was transfected into 293 cells in a 6 well plate using Lipofectamine 2000 (Thermo Fisher, Waltham, MA) following on manufacturer's protocol. 48 hours following transfection, 20 µg whole cell extracts in RIPA buffer (Thermo Fisher) were resolved on a 12% SDS-PAGE gel. The PVDF membrane was then probed with antibodies directed against PC-TP, GATA1, and internal loading control, GAPDH (Santa Cruz Biotechnology Dallas, TX). The blot was scanned and bands quantified using a LI-COR Odyssey imaging system (Lincoln, NE).

Results

Identification and annotation of eQTLs associated with *PCTP* expression.

To better understand the genetic mechanisms responsible for variation in *PCTP* expression we performed an eQTL analysis to identify associations between *PCTP* RNA levels and SNPs located within 500kb of the *PCTP* gene (6, 8). This resulted in the identification of 26 SNPs whose *P*-value for association with *PCTP* RNA levels met genome-wide significance ($P < 5 \times 10^{-8}$) (Table 2). This set of 23 highly significant SNPs were significantly associated with the higher level phenotype PAR4 ARS (p-value < 0.05, multiple linear regression) with multiple and adjusted R-squared values of 0.19 and 0.12, respectively.

To prioritize these variants for functional testing, we utilized publically available data to determine which of these SNPs coincided with markers of regulatory regions of the genome. The sources of data we included were: (A) ChromHMM chromatin state analysis for K562 cells, which utilizes correlations between chromatin states and known biological activity to predict biological activity (promoter, enhancer, transcription elongation, etc.) throughout the genome (9). Due to their myelogenous origin, K562 cells are the most biologically relevant among the ENCODE cell types for the study of gene expression in platelets and their precursor megakaryocytes (MKs). K562 cells have often been used to study the role of transcription factors in MK differentiation (10-12). (B) Chromatin Immunoprecipitation-Deep Sequencing (ChIP-Seq) data from CD34 hematopoietic stem cells (HSCs), and K562 cells indicating transcription factor binding. (C) ChIP-Seq data from CD34 and K562 cells indicating histone H3 monomethylated on lysine 4 (H3K4me1), a marker of enhancer regions (2, 3, 13-15) (Table 2).

We prioritized SNP rs2912553, located in the first intron of the *PCTP* gene, for further analysis for the following reasons (Table 2 and Fig 1A): (1) It was among the most significant eQTLs, by *P*-value (9.481×10^{-30}, multiple linear regression), for *PCTP*; (2) ChromHMM predicted it to be located in an enhancer region in K562 cells; (3) It was one of two candidate SNPs located

in a region containing a strong ChIP-Seq signal for H3K4me1 in CD34 cells; (4) It was the only candidate SNP which overlapped a transcription factor ChIP-Seq peaks in both K562 and CD34 cells. As shown in Fig 1B, in the PRAX1 cohort, individuals homozygous for the T allele of rs2912553 expressed the highest levels of *PCTP*, whereas C/C homozygotes expressed the least. In concordance with our observation that platelets from black PRAX1 subjects contained higher amounts of *PCTP*, the high-expressing T allele was more common in blacks (allele frequency = 72%), while the low expressing C allele was more common in whites (allele frequency = 85%) (Fig 1C), suggesting this variant may be causal for the differential expression of *PCTP* observed.

rs2912553 confers differential transcriptional potency

We cloned various sized fragments surrounding both alleles of rs2912553 into a luciferase vector containing a minimal promoter (Fig. 2A) to determine if the variant results in differential transcription enhancing potential. After transfection into K562 cells, constructs containing 56, 130, 209, and 454 bp genomic fragments demonstrated an allele-specific difference in luciferase activity (Fig. 2B). These differences replicate the observation in the PRAX1 subjects, with the constructs containing the T allele generating up to 3-fold the amount of luciferase activity compared to C allele constructs. The only difference between the two constructs is the single nucleotide change at rs2912553. Vectors containing 32 and 43 bp sized fragments resulted in similar luciferase activity in both alleles. Interestingly the 454 bp construct resulted in lower activity overall, suggesting a negative-regulatory element present in the additional sequence. 20 bp away from rs2912553 (eQTL P-value = 1.0×10^{-18}) is another common SNP, rs2912552, which could potentially contribute to the transcriptional differences observed in 56 bp and larger fragments. To test this, we generated 4 versions of the 130bp construct, which contains a large number of predicted GATA1, GATA2, and Ets factor binding sites not present in the 209bp fragment, containing all potential haplotypes of rs2912552 and rs2912553. We observed that rs2912552

genotype had no effect on transcription driven by this locus (Figure 2C). A predicted GATA1/2 binding site present in the 56bp construct, which is absent in the 43bp fragment, prompted us to further explore the role of GATA factors in regulating transcription from this genomic locus.

rs2912553 results in differential binding of a GATA1-containing protein complex

One hypothesis to explain the difference in transcriptional potency between the two alleles, is that the sequences containing the T allele had higher affinity for transcription factors (TF) and other regulatory molecules. To test this possibility, we performed electromobility shift assays (EMSAs) to investigate the interaction between DNA probes containing genomic sequence surrounding rs2912553 and nuclear extracts from K562 cells. 65 bp probes containing the T allele consistently recruited the highest molecular weight complex with higher affinity than probes containing the C allele (Fig 3A). Quantification indicated the band containing this complex was 2-fold more intense in T alleles as compared to C , indicating increased TF binding (Fig 3B).

Due to the transcription factor binding site predictions and CD34 ChIP-Seq data indicating their binding (Fig 1A), we tested whether antibodies to GATA1 or GATA2 could disrupt or alter the mobility of the DNA:protein complex formed on the rs2912553 containing probes. To control for variation in the assay, we only analyzed the highest molecular weight band formed by addition of K562 nuclear extracts to the probes. As in Fig 3A, in the absence of a specific antibody, “T”-containing probes formed this highest molecular weight complex with greater affinity as indicated by the darker intensity (Fig. 3B, lanes 3 and 4). We found that while anti-GATA1 antibodies disrupted formation of this complex, anti-GATA2 antibodies did not (Fig 3C). Quantification of the shifted bands indicated that addition of the α-GATA1 antibody abrogated this difference significantly on the T allele (P = 0.004) and trending towards significance on the C allele (P = 0.07) (Fig 3D). This indicates that the highest molecular weight complex, which is differentially recruited to the rs2912553 locus based on genotype, contains GATA1.

GATA1 Regulates Transcription from the rs2912553 locus.

Given the evidence that GATA1 antibodies disrupted complex formation on rs2912553 containing probes, we investigated the effect of silencing or overexpression of GATA1 on luciferase expression from the 130bp construct, which contains predicted GATA1 binding sites. Transfecting siRNAs directed against GATA1 resulted in a 70% reduction in GATA1 mRNA and protein level (Suppl. Figure 1 and Fig 4A). This attenuation resulted in a 90% reduction in luciferase activity on constructs containing the T allele and a 60% reduction in constructs containing the C allele compared to a scrambled siRNA control (Fig 4B). Conversely, overexpression of GATA1 resulted in a 20 and 80 fold increase in luciferase expression on the T and C alleles respectively (Fig 4C). To confirm the role of GATA1 in regulating PCTP protein levels, we transfected a FLAG-GATA1 expression plasmid into 293 cells, which contain low levels of endogenous GATA1 and homozygous for the “C” allele at rs2912553 (Fig. 4D). Over expression of GATA1 in these cells resulted in a 40% increase in PC-TP protein (Fig 4E). This data supports the hypothesis that GATA1 contributes to regulation of the rs2912553 locus.

Discussion

We have previously reported the results of a genome-wide eQTL analysis in which we identified genetic variants that are linked to the RNA level of a gene within 500kb (6). One of the identified platelet eQTLs was SNP rs2912553 which is associated with *PCTP* gene expression. Identification of this association is supported by allele-specific expression analysis of RNA-Seq data (6). In this analysis we found a variant (rs2960062) located in the transcribed region of *PCTP* and linked to rs2912553 (D'=0.9919). The C allele of rs2960062 is associated with the T allele of rs2912553 and is more frequently present in RNA-seq reads of *PCTP* transcript in rs2960062 heterozygotes, confirming higher transcriptional activity of the linked T-allele of rs2912553. In this report, we present data that supports the hypothesis that variation at rs2912553 leads to differential recruitment of a GATA1-contaning complex which regulates *PCTP* expression. This mechanism results in differential expression of *PCTP* by race and potentially contributes to dissimilar platelet PAR4 function.

The Genome-Wide Repository of Associations between SNPs and Phenotypes (GRASP) database contains the results of 2082 GWAS studies. Of the 21368 SNPs in the database having an association with a phenotype with a genome wide significance of $P < 5 \times 10^{-8}$, only 1.88% of them are categorized as causing a missense mutation in a protein coding gene (16). Therefore, the vast majority of genetic associations with phenotypes occur in non-coding, regulatory regions of the genome. This implies that gene expression levels are key determinants of human diversity and physiology. Here we demonstrate a link between platelet PCTP eQTL rs2912553 and *PCTP* expression level. This complements our previously identified link between *PCTP* levels and platelet PAR4 function. The genotype:gene expression:phenotype chain is confirmed by the observation that platelet *PCTP* eQTLs are associated with platelet PAR4 reactivity.

We have previously reported that while the Alanine/Threonine variation in position 120 of the PAR4 receptor itself determines 48% of the variance in PAR4 reactivity, PC-TP protein levels

accounted for 18%, making it a significant factor in determining PAR4 function (17). PC-TP is best understood as having a role in regulating lipid and glucose metabolism in the liver, and the mechanism by which PC-TP regulates PAR4 function is still currently unknown (18-20). *PCTP* associations with genome-wide or suggestive significance have been reported in other GWAS studies, including height in African Americans (rs1549519, $P = 5.8 \times 10^{-9}$) (21), LDL cholesterol (rs11867309, $P = 6.0 \times 10^{-9}$), total cholesterol (rs11867309, $P = 8.3 \times 10^{-7}$) (22), and collagen-stimulated platelet aggregation after aspirin and clopidogrel treatment (rs7210295 $P = 1 \times 10^{-6}$) (23). Of particular interest, the SNP associated with lipid levels, rs11867309 is located only 1600 bp away from and is linked to rs2912553 (D'=0.991).

We have made the following observations on the effect of the rs2912553 variant on *PCTP* expression: (1) The T-allele confers greater transcriptional potency than the C allele. (2) The T allele has a greater ability to recruit a GATA-1 containing protein complex than the C allele. (3) Reduction or enhancement of GATA1 levels results in a corresponding change in transcriptional activity.

A limitation of this study is a lack of the identification of the transcription factor that binds the first intron of *PCTP* and is altered by the rs2912553 genotype. Analysis of the DNA sequence predicts a number of ETS-factor binding sites which are altered by rs2912553 and attenuation of these factors with siRNA altered the enhancer potential of this locus. However, we have been unable to definitively confirm the importance of these factors to *in vivo PCTP* expression.

Better understanding of the significance behind loci identified in GWAS requires mechanistic insight as to how regulatory regions of the genome affect biology. We have utilized genotype, gene expression and physiological data to functionalize a racially dimorphic variant that regulates expression of *PCTP*, a regulator of platelet PAR4 function. More studies are needed to evaluate the role of this variant in contributing to racial differences in platelet function in both healthy and pathological states.

Acknowledgements

The authors would like to thank Edward Chen for helpful discussions regarding the manuscript.

References

1. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007; 315(5813): 848-53.
2. Encode Project Consortium, Bernstein BE, Birney E, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489(7414): 57-74.
3. Tijssen MR, Cvejic A, Joshi A, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell 2011; 20(5): 597-609.
4. Edelstein LC, Simon LM, Montoya RT, et al. Racial differences in human platelet PAR4 reactivity reflect expression of PCTP and miR-376c. Nat Med 2013; 19(12): 1609-16.
5. Simon LM, Edelstein LC, Nagalla S, et al. Human platelet microRNA-mRNA networks associated with age and gender revealed by integrated plateletomics. Blood 2014; 123(16): e37-45.
6. Simon LM, Chen ES, Edelstein LC, et al. Integrative Multi-omic Analysis of Human Platelet eQTLs Reveals Alternative Start Site in Mitofusin 2. Am J Hum Genet 2016; 98(5): 883-97.
7. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing 2016.
8. Simon LM, Chen ES, Edelstein LC, et al. Alternative start site in Mitofusin-2 revealed by integrative multi-omic analysis of human platelet eQTLs. American journal of human genetics 2016; In Press.
9. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 2012; 9(3): 215-6.
10. Kuvardina ON, Herglotz J, Kolodziej S, et al. RUNX1 represses the erythroid gene expression program during megakaryocytic differentiation. Blood 2015; 125(23): 3570-9.
11. Elagib KE, Mihaylov IS, Delehanty LL, et al. Cross-talk of GATA-1 and P-TEFb in megakaryocyte differentiation. Blood 2008; 112(13): 4884-94.
12. Elagib KE. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. Blood 2003; 101(11): 4333-41.
13. Beck D, Thoms JA, Perera D, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood 2013; 122(14): e12-22.
14. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 2010; 28(10): 1045-8.
15. Hon GC, Hawkins RD, Ren B. Predictive chromatin signatures in the mammalian genome. Hum Mol Genet 2009; 18(R2): R195-201.
16. Eicher JD, Landowski C, Stackhouse B, et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. Nucleic Acids Res 2015; 43(Database issue): D799-804.
17. Edelstein LC, Simon LM, Lindsay CR, et al. Common variants in the human platelet PAR4 thrombin receptor alter platelet function and differ by race. Blood 2014; 124(23): 3450-8.
18. Shishova EY, Stoll JM, Ersoy BA, et al. Genetic ablation or chemical inhibition of phosphatidylcholine transfer protein attenuates diet-induced hepatic glucose production. Hepatology 2011; 54(2): 664-74.
19. Kang HW, Niepel MW, Han S, et al. Thioesterase superfamily member 2/acyl-CoA thioesterase 13 (Them2/Acot13) regulates hepatic lipid and glucose metabolism. FASEB J 2012; 26(5): 2209-21.
20. Kawano Y, Ersoy BA, Li Y, et al. Thioesterase superfamily member 2 (Them2) and phosphatidylcholine transfer protein (PC-TP) interact to promote fatty acid oxidation and control glucose utilization. Mol Cell Biol 2014; 34(13): 2396-408.

21. Carty CL, Johnson NA, Hutter CM, et al. Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). Hum Mol Genet 2012; 21(3): 711-20.
22. Asselbergs FW, Guo Y, van Iperen EP, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. Am J Hum Genet 2012; 91(5): 823-38.
23. Lewis JP, Ryan K, O'Connell JR, et al. Genetic variation in PEAR1 is associated with platelet aggregation and cardiovascular outcomes. Circ Cardiovasc Genet 2013; 6(2): 184-92.
24. Chacon D, Beck D, Perera D, et al. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. Nucleic Acids Res 2014; 42(Database issue): D172-7.

Legends to Figures and Tables

Figure 1: Annotation of *PCTP* eQTLs. (A) (Top panel) Local Manhattan plot of platelet eQTL SNPs (diamonds) in *PCTP* with $P < 5 \times 10^{-8}$ (blue dashed line). X-axis: genomic coordinates, left Y-axis: $-\log_{10}$(P-value). Color indicates R^2 linkage disequilibrium value relative to rs2912553. Red line indicates recombination rate (right Y-axis). (Bottom panel) Annotation tracks (from top to bottom): Platelet eQTL SNPs in *PCTP*, colors indicate R^2 LD value relative to rs2912553. *PCTP* gene structure. ChromHMM chromatin state predictions from K562 cells, ChIP-Seq peaks from CD34 cells (H3K4me1, RUNX1, GATA1). (B) Differential expression of *PCTP* by rs2912553 genotype. Box represent interquartile range, line represents mean, whiskers represent 1.5 x interquartile range. N=154, *P*-value calculated as described in Materials and Methods. (C) Allelic frequencies of rs2912553 among black (B) and white (W) subjects in PRAX1.

Figure 2: rs2912553 genotype confers differential transcriptional potency on the surrounding locus. (A) Schematic representation of cloned fragments surrounding rs2912553. (B) Luciferase activity of reporter fragments by size containing the T or C allele of rs2912553. (C) Luciferase activity of all four potential haplotypes of rs2912552 and rs2912553. Results were normalized to cotransfected β-gal expression and presented as +/- SD. N=3 each for each condition. $P = 3 \times 10^{-22}$, 2-way ANOVA for the effect of genotype on luciferase activity.

Figure 3: Differential protein binding based on rs2912553 genotype. (A) The highest molecular weight complex protein complex from K562 nuclear extracts (arrow) bound with more affinity to probes containing a T allele than a C allele. Representative gel. (B) Quantification of the highest molecular weight complex from 3 independent gels as described in (A). N=3 for each condition $P = 0.02$, 2-way ANOVA for the effect of genotype on shifted band intensity. (C) Addition of anti-GATA1 antibodies disrupted formation of the complex (arrow) whereas anti-GATA2 antibodies had no effect. Representative gel. (D) Quantification of three independent gels as

described in (C). N=3 for each condition. *P*= 0.0004, 2-way ANOVA for effect of antibody on shifted band intensity. Error bars represent +/- SD

Figure 4: Transcriptional regulation of the rs2912553 locus by GATA1. (A) Western blot of K562 cells transfected with siRNA directed against GATA1. Blots were probed for GATA1 and GAPDH. (B) Attenuation of GATA1 expression lead to a reduction in luciferase activity driven by the rs2912553 locus. (C) Overexpression of GATA1 results in enhanced transcriptional activity from the rs2912553 locus. Luciferase results were normalized to co-transfected β-gal. (D) Western blot of 293 cells transfected with a FLAG-GATA1 expression vector or control. Blots were probed with antibodies directed against GATA1, PC-TP, and GAPDH. (E) Quantification three independent gels as described in (D). Error bars represent +/- SD N=3 for each condition. *P*-values calculated by 2-way T-test.

Table 2: Platelet eQTL SNPs for PCTP and Annotations.

Tables

Table 1

Primer	Sequence
T-32-F	5’-CTAGCTTCTAAGCATCATTTCCTTCTGTAAAGGTAGA-3’
T-32-R	5’-GATCTCTACCTTTACAGAAGGAAATGATGCTTAGAAG-3’
C-32-F	5’-CTAGCTTCTAAGCATCATTTCCTTCCGTAAAGGTAGA-3’
C-32-R	5’-GATCTCTACCTTTACGGAAGGAAATGATGCTTAGAAG-3’
T-43-F	5’-CTAGCTTCTAAGCATCATTTCCTTCTGTAAAGGTAGAGATAACACCAA-3’
T-43-R	5’-GATCTTGGTGTTATCTCTACCTTTACAGAAGGAAATGATGCTTAGAAG-3’
C-43-F	5’-CTAGCTTCTAAGCATCATTTCCTTCCGTAAAGGTAGAGATAACACCAA-3’
C-43-R	5’-GATCTTGGTGTTATCTCTACCTTTACGGAAGGAAATGATGCTTAGAAG-3’
T-56-F	5’-CTAGCTTCTAAGCATCATTTCCTTCTGTAAAGGTAGAGATAACACCAACTTCCCTATCTACA-3’
T-56-R	5’-GATCTGTAGATAGGGAAGTTGGTGTTATCTCTACCTTTACAGAAGGAAATGATGCTTAGAAG-3’
C-56-F	5’-CTAGCTTCTAAGCATCATTTCCTTCCGTAAAGGTAGAGATAACACCAA CTTCCCTATCTACA-3’
C-56-R	5’-GATCTGTAGATAGGGAAGTTGGTGTTATCTCTACCTTTACGGAAGGAAATGATGCTTAGAAG-3’
130-F	5’-CTAGCTAGCGTTGGCCTCTCTAAGCATCATTT-3’
130-R	5'- CCCAAGCTTGCATGGTCCACTTTCCAACG -3'
209-F	5’-CTAGCTAGCGAATAGAAGGGGCCAGTGTG-3’
209-R	5’-CCCAAGCTTGCATGGTCCACTTTCCAACG-3’
454-F	5'-CTAGCTAGCTCTGGAACAATGATTTGGGAATGAC-3'
454-R	5'- CCCAAGCTTATATCTTGATCCCAAGAAGAAACCAG -3'

Table 2

				TF ChIP Seq Peak Overlap		H3K4me1 ChIP Seq Peak Value*	
#	rsID	eQTL *P*-value	K562 ChromHMM Chromatin State	CD34	K562	CD34	K562
1	rs2912559	4.80E-30	11_Weak_Txn	FALSE	FALSE	0	0
2	rs35430620	4.80E-30	11_Weak_Txn	FALSE	FALSE	0	0
3	rs2332513	4.80E-30	11_Weak_Txn	FALSE	FALSE	1	0
4	rs2912553	9.48E-30	7_Weak_Enhancer	TRUE	TRUE	9	1
5	rs2912547	1.63E-26	9_Txn_Transition	FALSE	FALSE	1	1
6	rs2960067	1.63E-26	10_Txn_Elongation	FALSE	FALSE	0	0
7	rs2960070	1.63E-26	11_Weak_Txn	FALSE	FALSE	1	0
8	rs4627425	8.63E-25	6_Weak_Enhancer	FALSE	TRUE	1	4
9	rs8082034	8.63E-25	13_Heterochrom/lo	FALSE	FALSE	1	0
10	rs2303506	8.63E-25	10_Txn_Elongation	FALSE	FALSE	0	0
11	rs9911793	8.63E-25	13_Heterochrom/lo	FALSE	FALSE	1	1
12	rs2960061	9.28E-24	11_Weak_Txn	FALSE	FALSE	0	0
13	rs2912550	1.01E-18	11_Weak_Txn	FALSE	FALSE	2	0
14	rs2960064	1.01E-18	10_Txn_Elongation	FALSE	FALSE	0	0
15	rs2912558	1.43E-18	10_Txn_Elongation	FALSE	FALSE	1	0
16	rs2960058	9.79E-17	10_Txn_Elongation	FALSE	FALSE	1	0
17	rs2960060	2.77E-14	10_Txn_Elongation	FALSE	FALSE	0	0
18	rs890487	3.00E-12	6_Weak_Enhancer	FALSE	FALSE	4	2
19	rs2960081	3.00E-12	2_Weak_Promoter	FALSE	FALSE	9	3
20	rs2960088	4.30E-12	7_Weak_Enhancer	FALSE	FALSE	0	0
21	rs890488	1.51E-11	1_Active_Promoter	FALSE	TRUE	5	0
22	rs2960065	1.51E-11	10_Txn_Elongation	FALSE	FALSE	1	0
23	rs2033108	5.99E-11	11_Weak_Txn	FALSE	TRUE	0	0
24	rs1160027	5.84E-10	7_Weak_Enhancer	FALSE	TRUE	0	2
25	rs12051898	7.53E-10	13_Heterochrom/lo	FALSE	FALSE	0	0
26	rs2290433	7.53E-10	11_Weak_Txn	FALSE	FALSE	1	0

* Data derived from BloodChIP database(24).

Figure 1

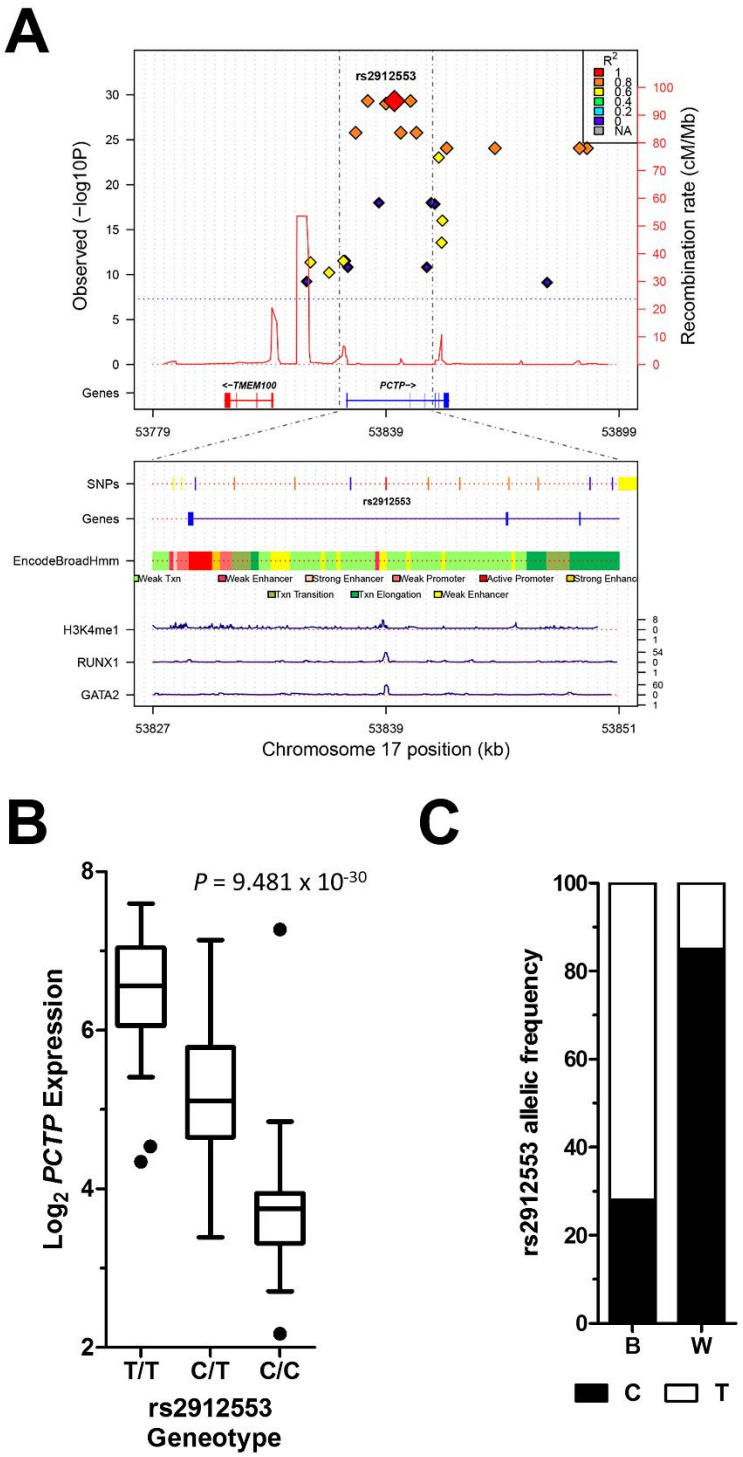

A
rs2912553
Observed (−log10P)
Recombination rate (cM/Mb)
Genes
<-TMEM100
PCTP->
53779
53839
53899
SNPs
Genes
EncodeBroadHmm
Weak Txn
Weak Enhancer
Strong Enhancer
Weak Promoter
Active Promoter
Txn Transition
Txn Elongation
Weak Enhancer
H3K4me1
RUNX1
GATA2
53827
53839
53851
Chromosome 17 position (kb)
B
P = 9.481 x 10-30
Log2 PCTP Expression
T/T
C/T
C/C
rs2912553 Geneotype
C
rs2912553 allelic frequency
B
W
C
T

Figure 2

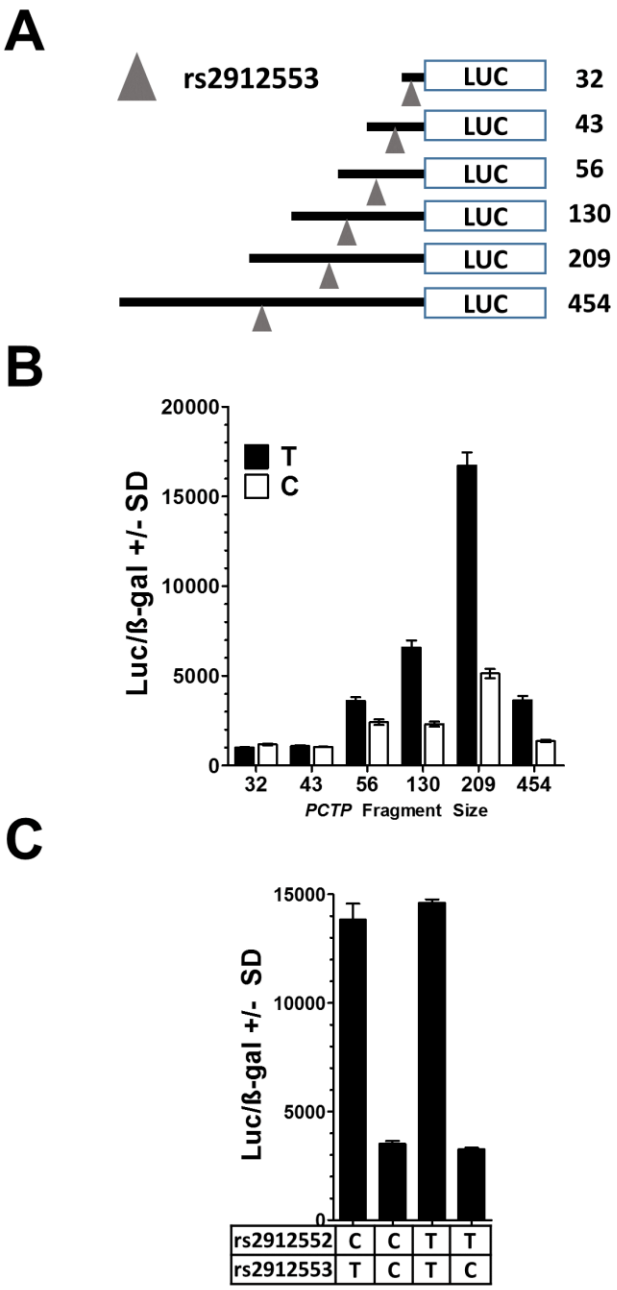

A
rs2912553
LUC
32
LUC
43
LUC
56
LUC
130
LUC
209
LUC
454
B
20000
15000
10000
5000
0
T
C
Luc/ß-gal +/- SD
32
43
56
130
209
454
PCTP Fragment Size
C
15000
10000
5000
0
Luc/ß-gal +/- SD
rs2912552 C C T T
rs2912553 T C T C

Figure 3

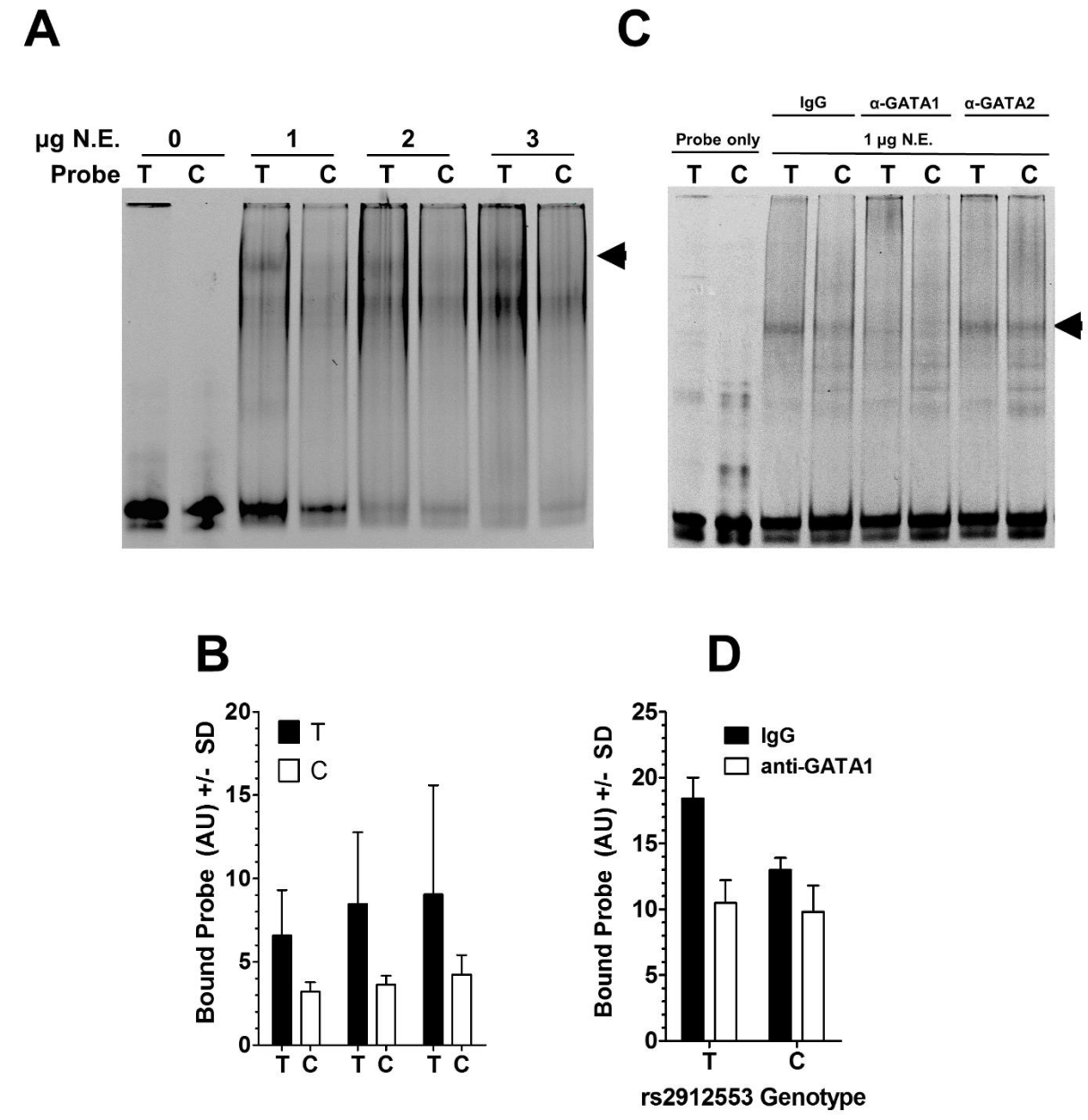
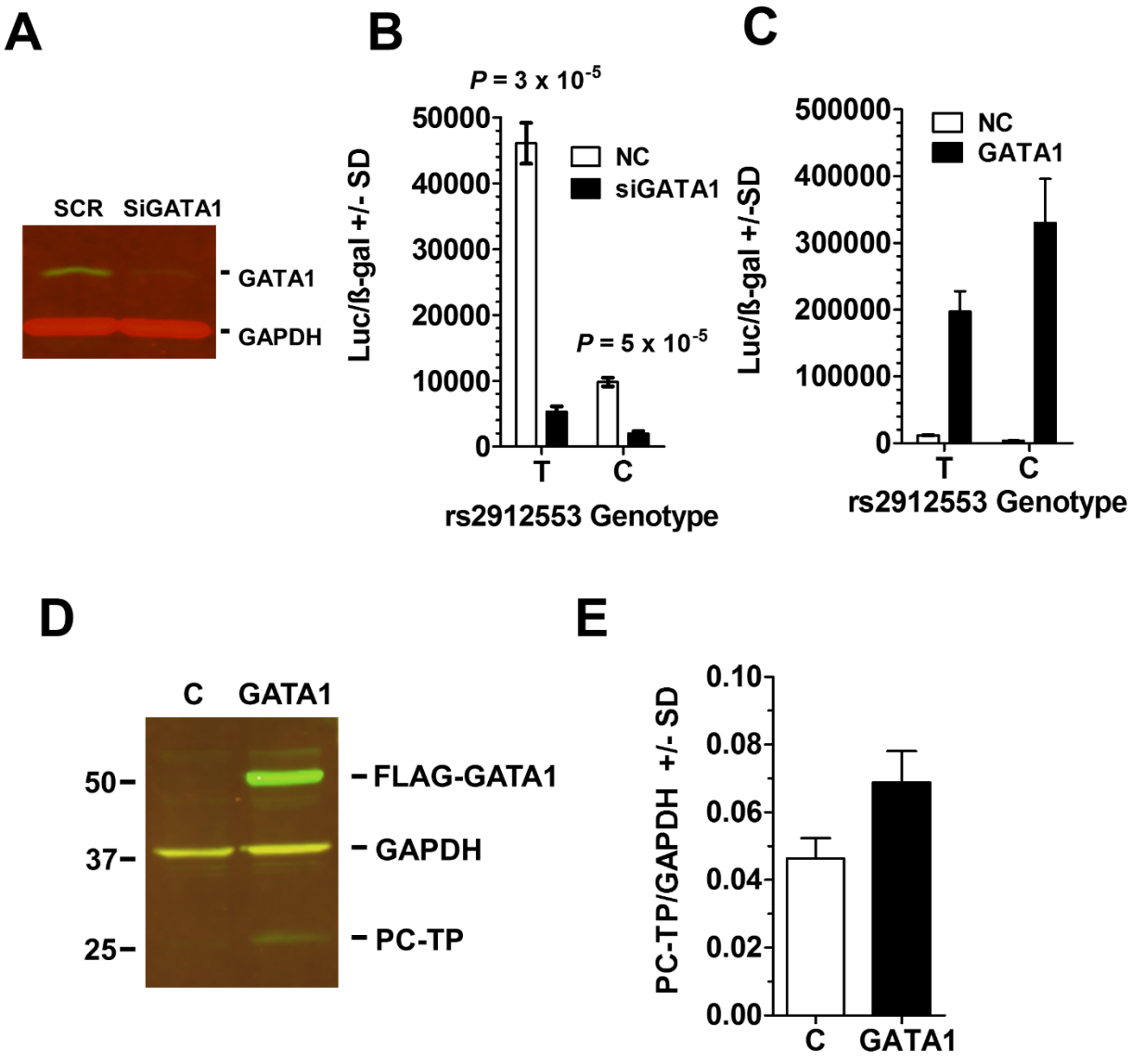
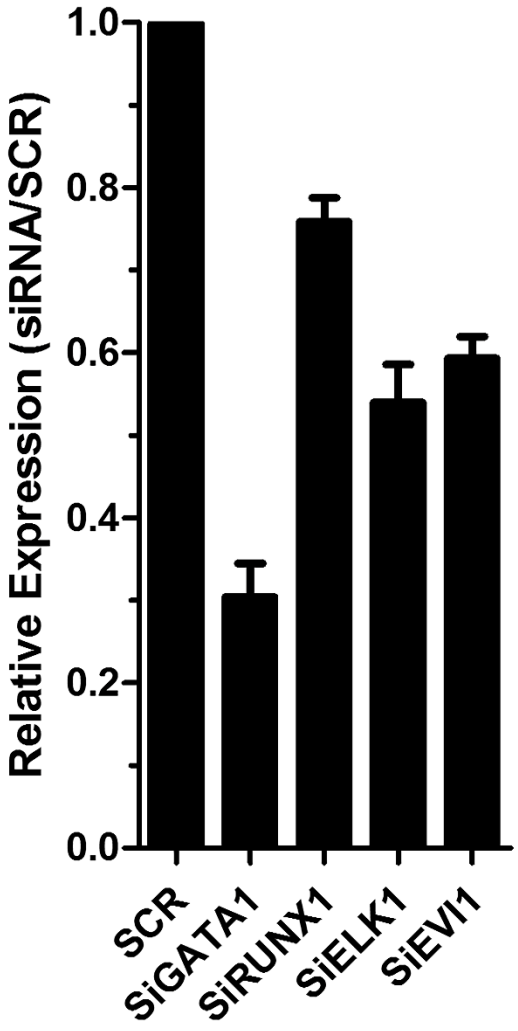

A
μg N.E. 0 1 2 3
Probe T C T C T C T C
C
IgG α-GATA1 α-GATA2
Probe only 1 μg N.E.
T C T C T C T C T C
B
Bound Probe (AU) +/- SD
20
15
10
5
0
T
C
T C T C T C
D
Bound Probe (AU) +/- SD
25
20
15
10
5
0
IgG
anti-GATA1
T C
rs2912553 Genotype

Figure 4


A
SCR
SiGATA1
GATA1
GAPDH
B
P = 3 x 10-5
P = 5 x 10-5
NC
siGATA1
Luc/ß-gal +/- SD
50000
40000
30000
20000
10000
0
T
C
rs2912553 Genotype
C
NC
GATA1
Luc/ß-gal +/-SD
500000
400000
300000
200000
100000
0
T
C
rs2912553 Genotype
D
C
GATA1
50
37
25
FLAG-GATA1
GAPDH
PC-TP
E
PC-TP/GAPDH +/- SD
0.10
0.08
0.06
0.04
0.02
0.00
C
GATA1

Suppl. Fig 1.


Relative Expression (siRNA/SCR)
1.0
0.8
0.6
0.4
0.2
0.0
SCR
SiGATA1
SiRUNX1
SiELK1
SiEVI1