# Supplemental Materials

August 6, 2014

## Contents

# 1 Supplemental Methods

**eQTL Analysis**

COPD was defined as GOLD Stage 2 or greater (FEV1/FVC ratio $< 0.7$ and FEV1 % of predicted <80%[1]. RNA was extracted from sputum and whole blood cell pellets using TRIzol and amplified with the Nugen Ovation RNA Amplification kit. Gene expression profiling was performed using the Affymetrix Human U133 Plus2 array. Gene expression data were log-transformed, and background correction and normalization were performed separately for the blood and sputum samples using RMA and quantile normalization as implemented in the affy and beadarray Bioconductor packages[2, 3, 4]. eQTL analysis was performed with the Bioconductor package GGtools for window sizes ranging from 50kb to 500kb around the gene transcription start and end site[5]. Analyses were adjusted for age, gender, pack-years of smoking, the first principal component of genetic ancestry as determined by the Tracy-Widom statistic, and the first 13 principal components of the expression data[6]. False discovery rate was controlled by the plug-in FDR method[7]. eQTL analysis was performed by specifying various window sizes (from 50kb to 500kb) around genes such that all SNPs within a certain base pair distance from the start or end of transcription were tested for association with the expression level of the gene. eQTL yield can be increased by adjusting for principal components (PC) of the expression data.[8, 9] The number of expression PCs was determined by examining the effect of adjusting for a range of PCs using data from chromosome 22 in both tissues.

**eQTL and GWAS Integration** Local correlation was calculated as the Pearson correlation between GWAS and eQTL test statistics in the area of the GWAS peak (defined as SNPs in the lowest quartile of GWAS p-values). Bayesian colocalization analysis has been previously described[10]. This method calculates the posterior probability that the eQTL and GWAS signals in a given locus are caused by at least one shared causal variant. For this analysis, the locus was defined as all SNPs within a 250kb window of gene start and end boundaries.

**COPD GWAS and Conditional Association Analyses**

COPD GWAS results from a combined analysis of non-Hispanic white subjects from the NETT-NAS, Norway GenKOLS, ECLIPSE, and COPDGene studies were integrated with the eQTL results from the 121 subjects from ECLIPSE. Information and results from this GWAS meta-analysis (including African-American subjects in COPDGene) are reported separately.[11] GWAS was performed on a total of 9,767 subjects who had been genotyped on Illumina platforms (HumanHap 550, Quad 610, or OmniExpress arrays). Imputation to 1000 Genomes reference samples was performed using MaCH and minimac.[12, 13] GWAS was performed separately in each cohort adjusting for age, pack-years of smoking exposure, and principal components of genetic ancestry using plink 1.07, and results were combined via fixed-effects meta-analysis using METAL (version 2010-08-01).[14, 15] Conditional GWAS analyses were

performed using logistic regression adjusting for the same covariates used in the primary GWAS analysis and the genotype or genotypes of interest. False discovery rate for the conditional analyses was controlled using the method of Storey et al.[16, 17]

# References


[1] Rabe KH et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. American Journal of Respiratory and Critical Care Medicine, 176, 532-555.

[2] Gautier L et al. (2004) affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 20, 307-315.

[3] Ritchie ME et al. (2007) A comparison of background correction methods for two-colour microarrays. Bioinformatics, 23, 2700-2707.

[4] Dunning MJ et al. (2007) beadarray: R classes and methods for Illumina bead-based data. Bioinformatics 23: 2183-2184.

[5] Carey VJ et al. (2009) Data structures and algorithms for analysis of genetics of gene expression with Bioconductor: GGtools 3.x. Bioinformatics, 25, 1447-1448.

[6] Price A et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38, 904-909.

[7] Hastie T, Tibshirani R, Friedman J. (2001) The Elements of Statistical Learning. Springer.

[8] Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 3: 1724-1735.

[9] Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 97: 10101-10106.

[10] Giambartolomei et al. (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genetics, May 15;10(5):e1004383.

[11] Cho MH et al. (2014) Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir Med 2:214-225.

[12] Li Y et al. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34: 816-834.

[13] Howie B et al. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genetics 44: 955-959.

[14] Purcell S et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81: 559-575.

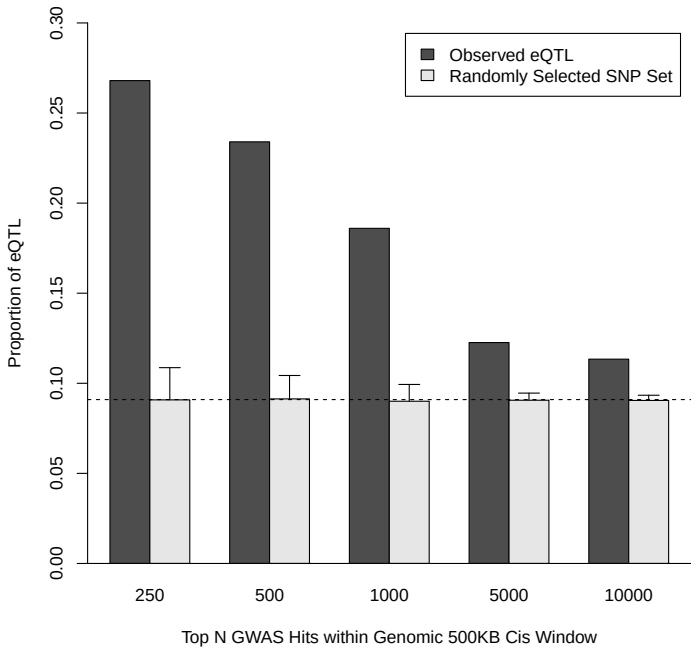
[15] Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190-2191.

[16] Storey JD (2002) A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B.

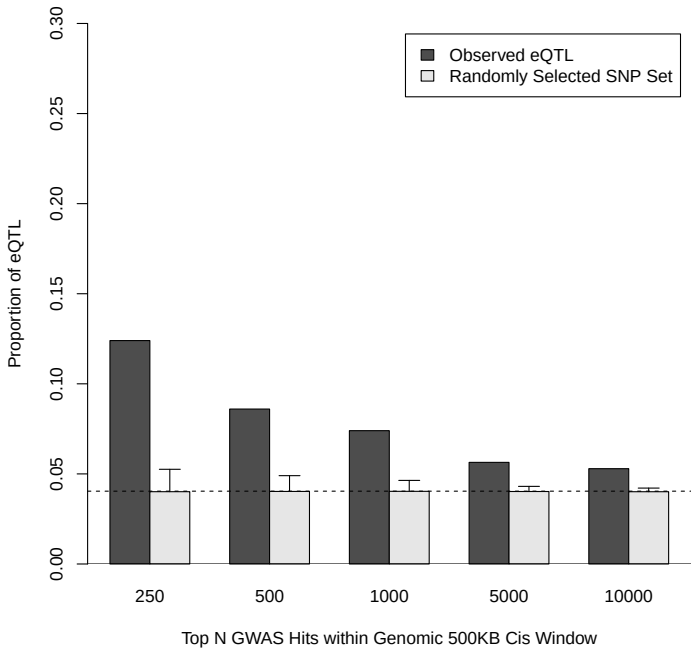
[17] Storey JD (2003) Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100: 9440-9445.

# 2 Supplemental Figures

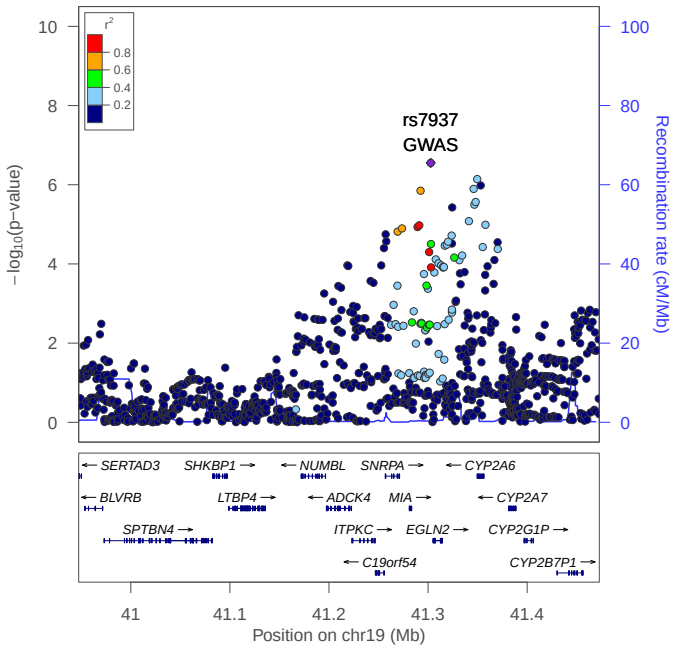
## 2.1 Supplemental Figure 1: Whole Blood eQTL Enrichment in GWAS



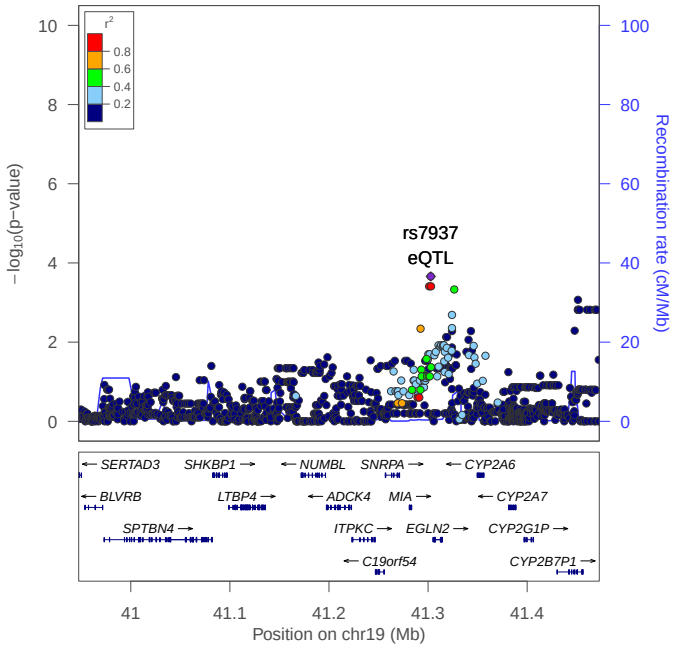
## 2.2 Supplemental Figure 2: Sputum eQTL Enrichment in GWAS



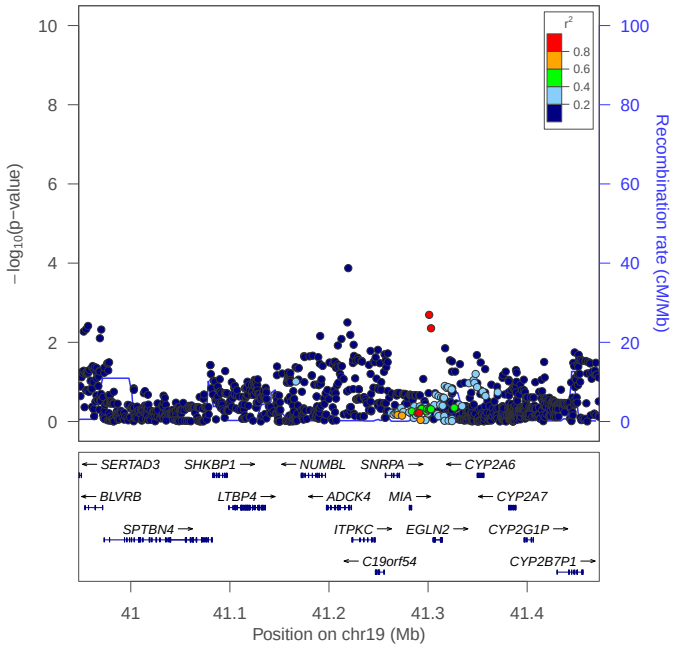
## 2.3 Supplemental Figure 3: GWAS Signal at 19q13



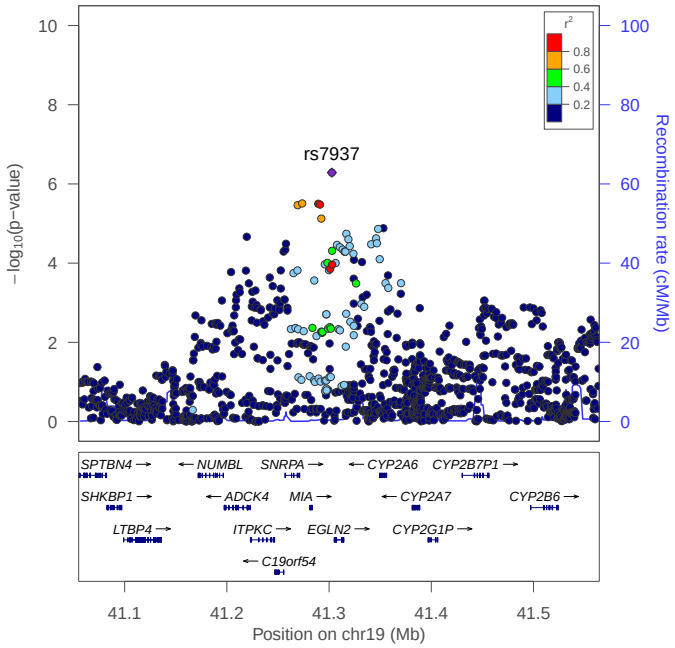
## 2.4 Supplemental Figure 4: Whole Blood eQTL Signal for ADCK4



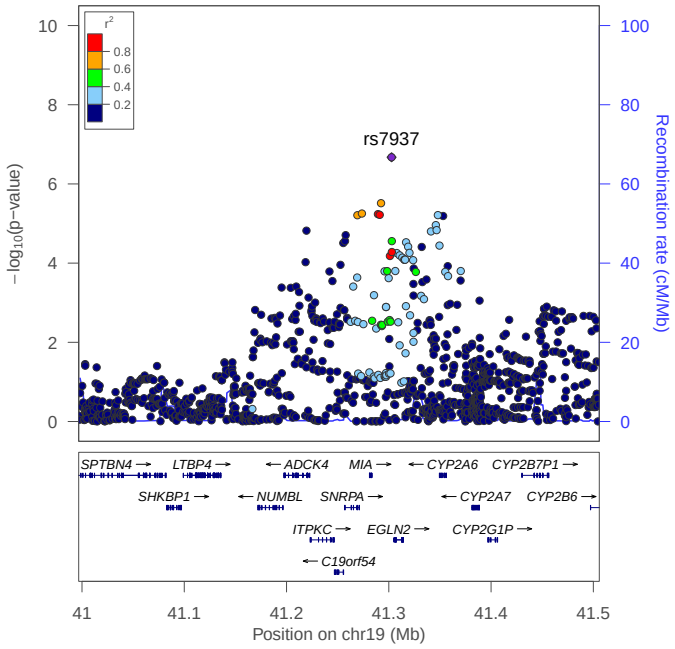
## 2.5 Supplemental Figure 5: COPD Association Plot at 19q13 Conditioning on Lead eQTL SNP for ADCK4



## 2.6 Supplemental Figure 6: COPD Association Plot at 19q13 Conditioning on Lead eQTL SNP for EGLN2

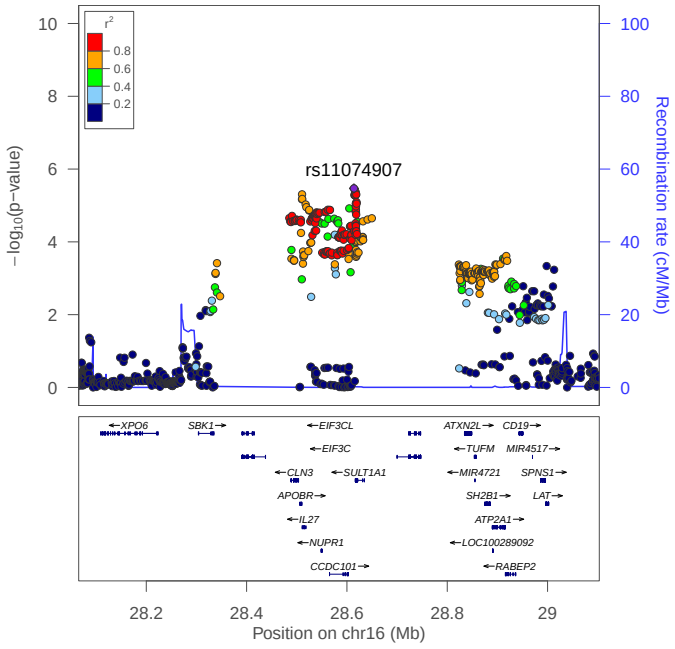


## 2.7 Supplemental Figure 7: COPD Association Plot at 19q13 Conditioning on Lead eQTL SNP for C19orf54

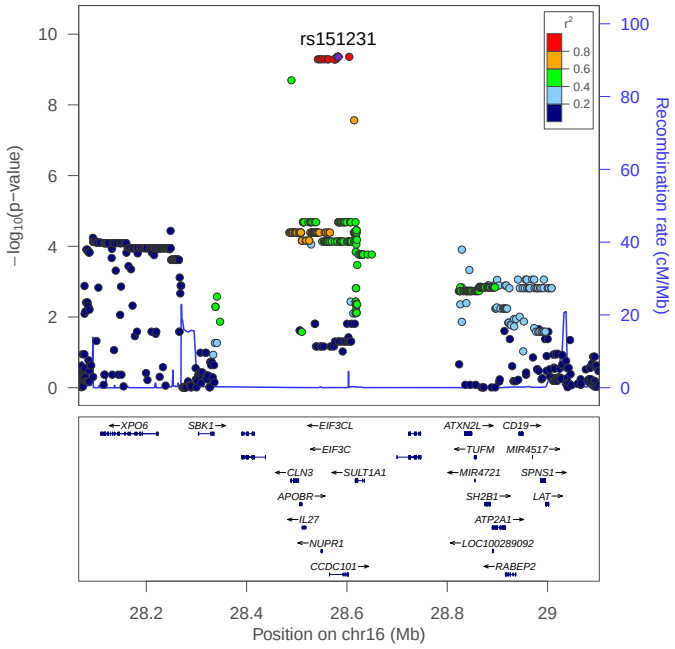


## 2.8 Supplemental Figure 8: GWAS Signal at 16q11

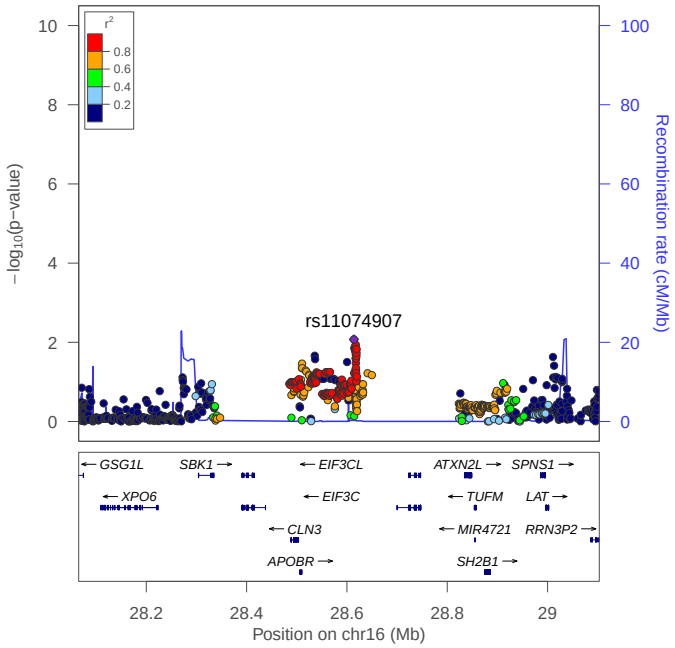
GWAS signal is bounded by two regions where 1000 Genomes reads map poorly, likely due to low genome complexity in these regions.



## 2.9 Supplemental Figure 9: Whole Blood eQTL Signal for CCDC101

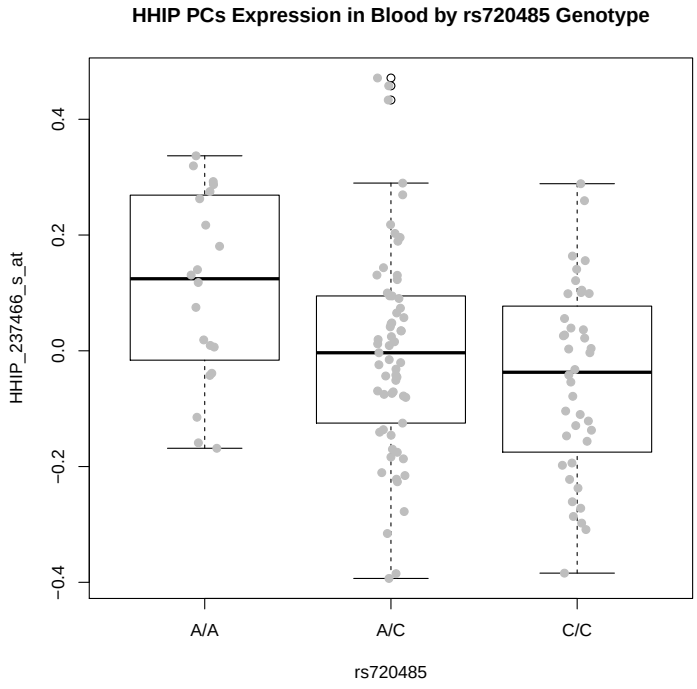


## 2.10 Supplemental Figure 10: COPD Association Plot at 16q11 Conditioning on Lead eQTL SNP for CCDC101

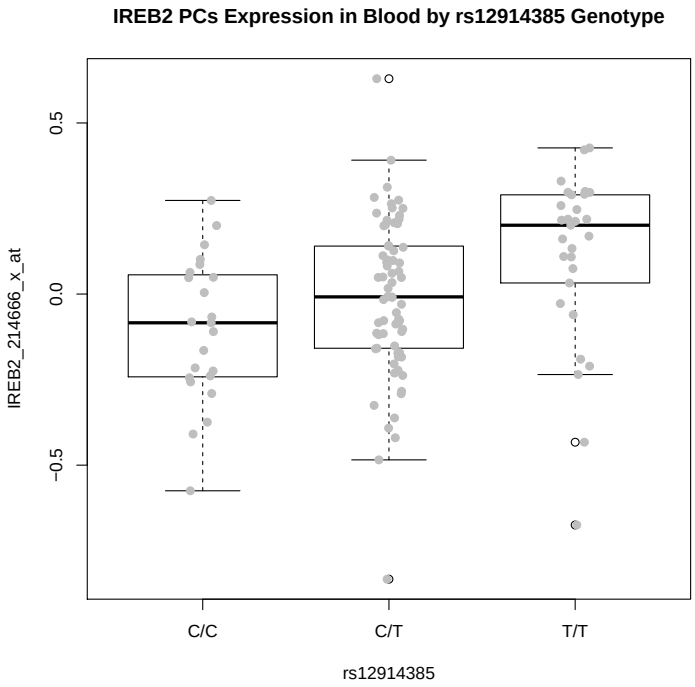


# 3 Expression-by-Genotype Plots from Top COPD eQTL Loci

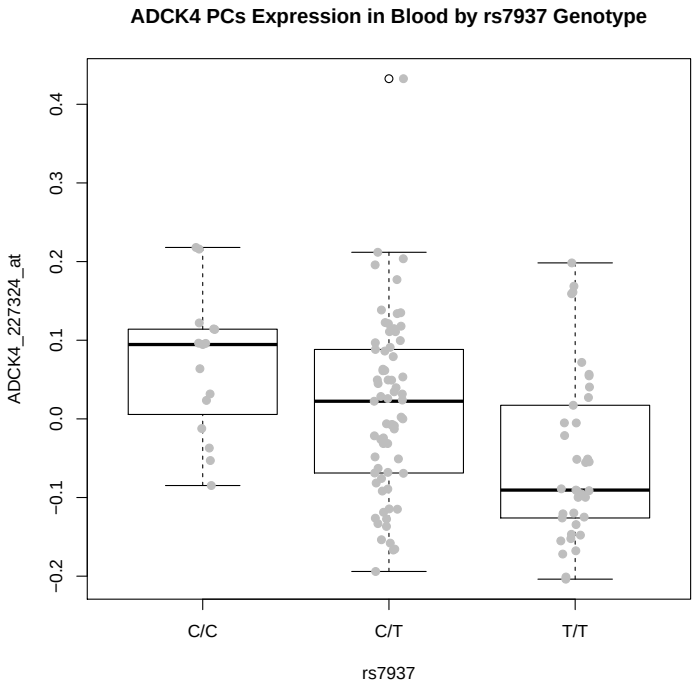
## 3.1 Supplemental Figure 11: Blood Expression by Genotype Plot for HHIP



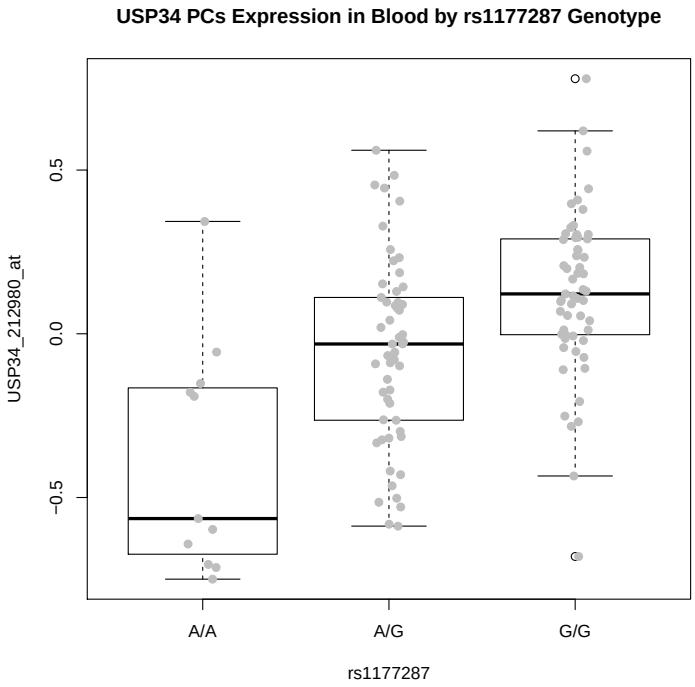
## 3.2 Supplemental Figure 12: Blood Expression by Genotype Plot for IREB2



## 3.3 Supplemental Figure 13: Blood Expression by Genotype Plot for ADCK4



## 3.4 Supplemental Figure 14: Blood Expression by Genotype Plot for USP34



## 3.5 Supplemental Figure 15: Blood Expression by Genotype Plot for CCDC101