Supplementary Information for

MULTIPLE *SCN5A* VARIANT ENHANCERS MODULATE ITS CARDIAC GENE EXPRESSION AND THE QT INTERVAL

Ashish Kapoor, Dongwon Lee, Luke Zhu, Elsayed Z. Soliman, Megan L. Grove, Eric Boerwinkle, Dan E. Arking, Aravinda Chakravarti

Aravinda Chakravarti

Email: aravinda.chakravarti@nyulangone.org

Ashish Kapoor

Email: ashish.kapoor@uth.tmc.edu

This PDF file includes:

Figures S1 to S11

Tables S1 to S5

References for SI reference citations

Other supplementary materials for this manuscript include the following:

Dataset S1-S11

www.pnas.org/cgi/doi/10.1073/pnas.1808734116

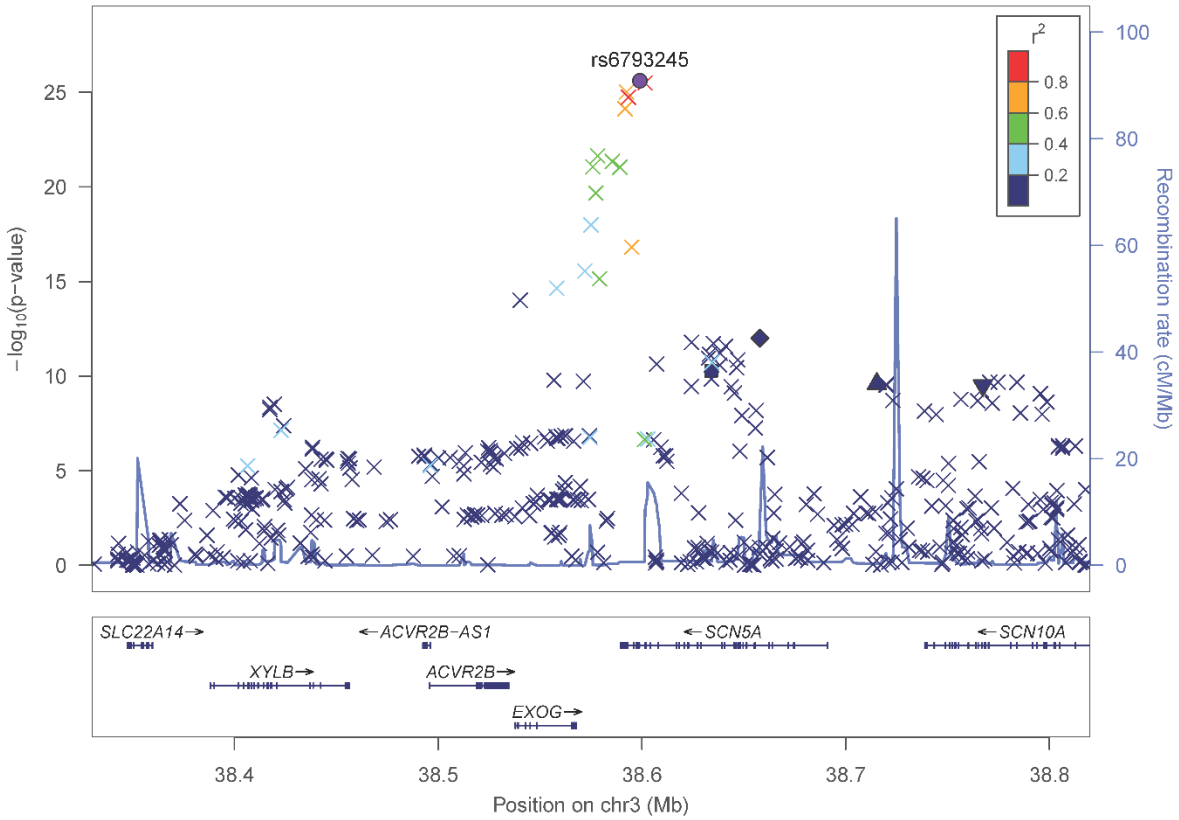

Figure S1: Regional association plot of variants at the *SCN5A-SCN10A* locus and QT Interval in the QT-IGC GWAS (1)**.** The X-axis is the genomic interval annotated with gene models, the left Y-axis the statistical significance of association ($-\log_{10}p$), and the right Y-axis the recombination rate (blue trace, cM (centiMorgan)/Mb(megabase)) based on HapMap samples (2). The most significant SNP, rs6793245, is shown as a purple circle. The four independent secondary signals are shown as square (rs11708996), diamond (rs1171007), triangle point-up (rs6599234) and triangle point-down (rs6801957); all other variants are shown as cross marks. All variants are color-coded based on their LD (r^2) with the index SNP. The plot was generated with LocusZoom (http://locuszoom.org/).

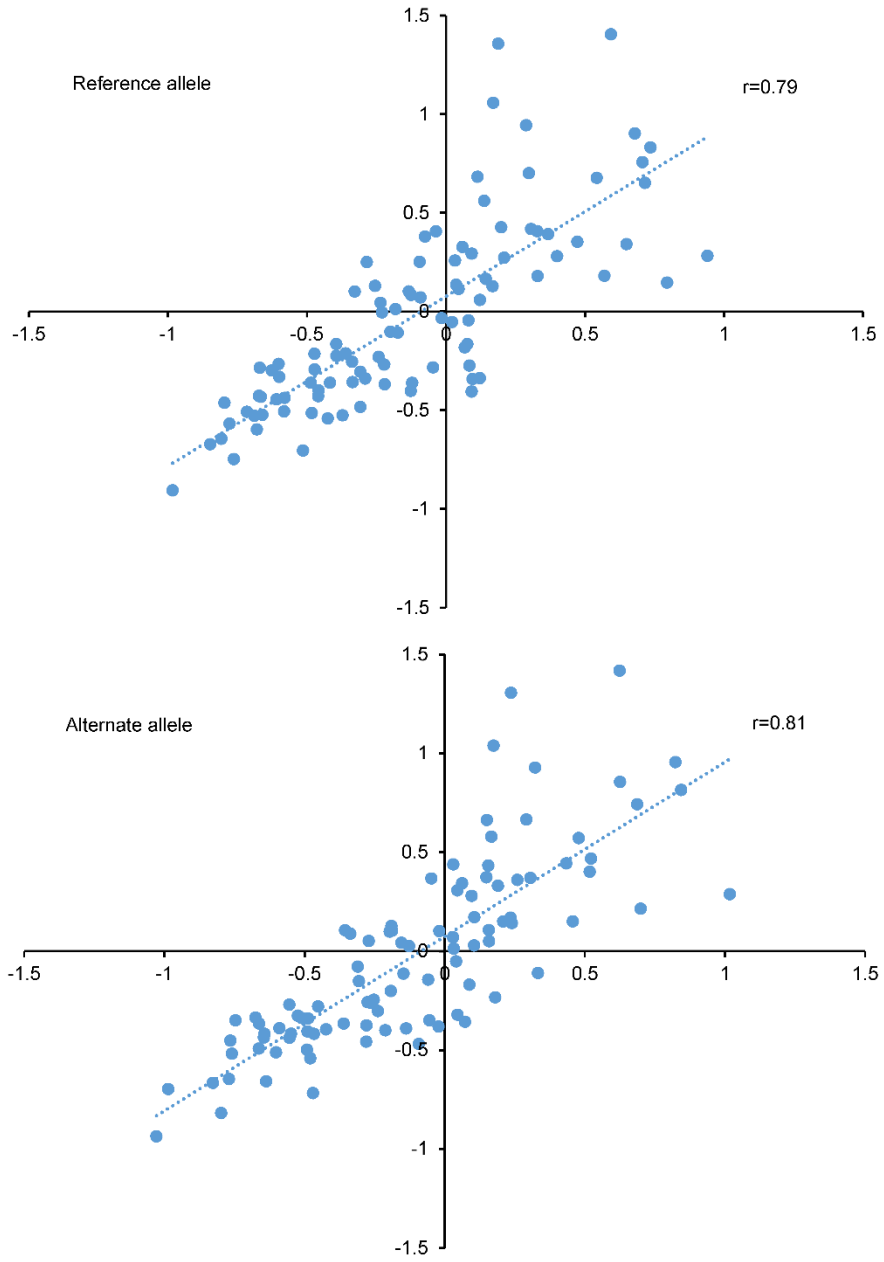

Figure S2: High correlation between allelic reporter activity of test elements in HL1 and AC16 cells. Scatter plots of $\log_{10}$-transformed mean normalized reporter activities of reference (top) and alternate (bottom) alleles containing amplicons in *in vitro* reporter assays performed in HL1 (X-axis) and AC16 (Y-axis) cells.

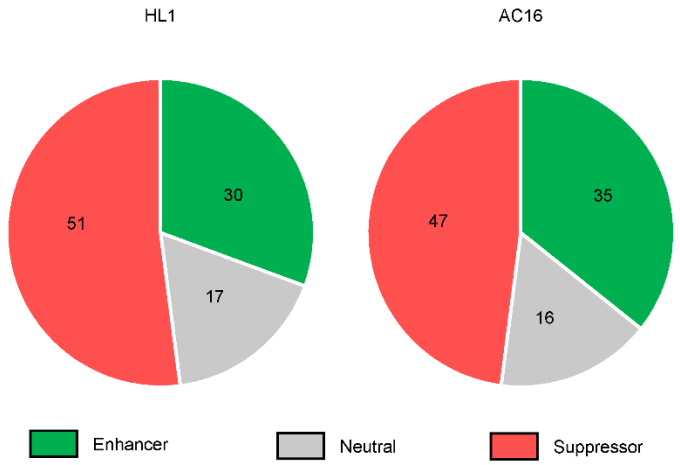

Figure S3: Distribution of the number of test elements defined as enhancers, suppressors or neutral based on mean standardized reporter activity in *in vitro* reporter assays; data from HL1 (left) and AC16 (right) cells are shown.

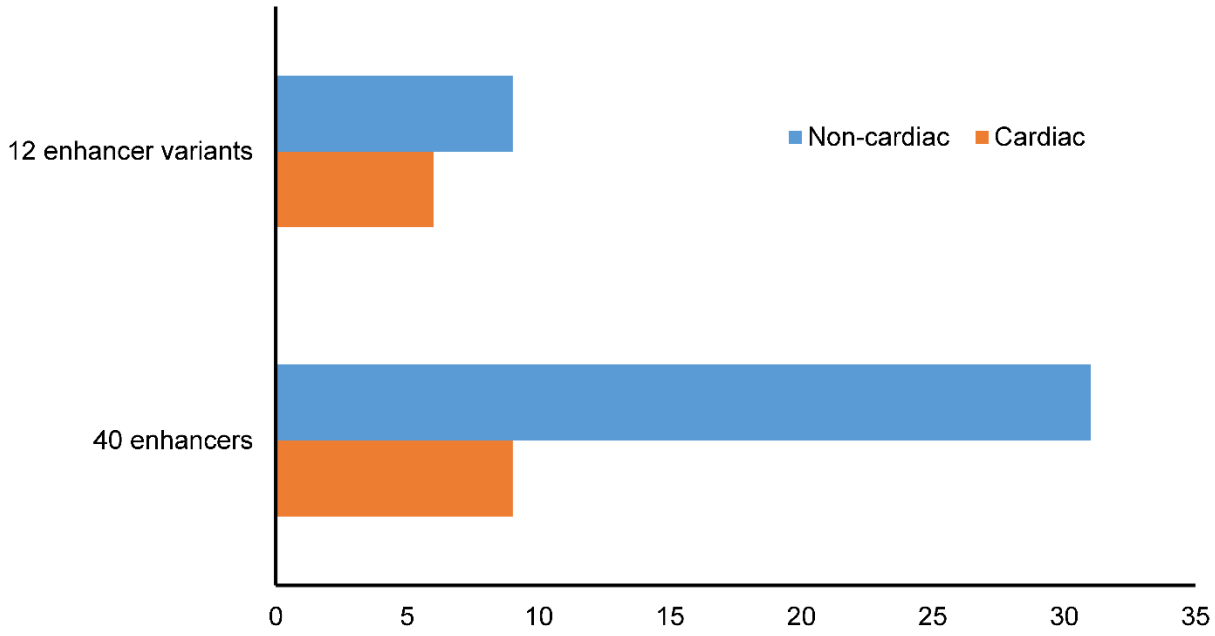

Figure S4: Overlap of *cis*-regulatory elements tested in *in vitro* reporter assays in HL1 and/or AC16 cells with open chromatin regions in human tissues and cells. Bar plots showing the number of enhancers, of the 40 *cis*-regulatory elements observed across reporter assays in HL1 and AC16 cells, or of the subset of 12 variants with significant allelic difference in reporter activity, marked as open chromatin regions in cardiac and non-cardiac human tissues and cells.

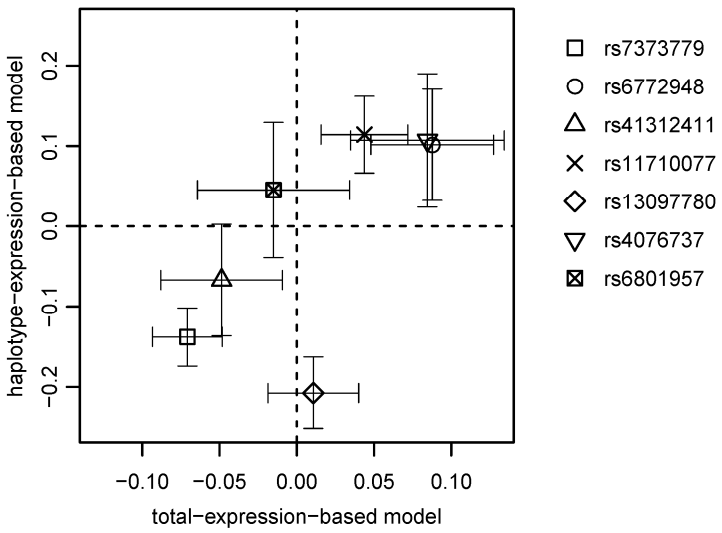

Figure S5: Two independent models capture the same underlying effects.

Regression slopes (beta values or effect sizes) from the total-expression-based model were compared to those of the haplotype-expression ratio-based model using the seven most significant causal CRE variants. Error bars are standard errors of beta.

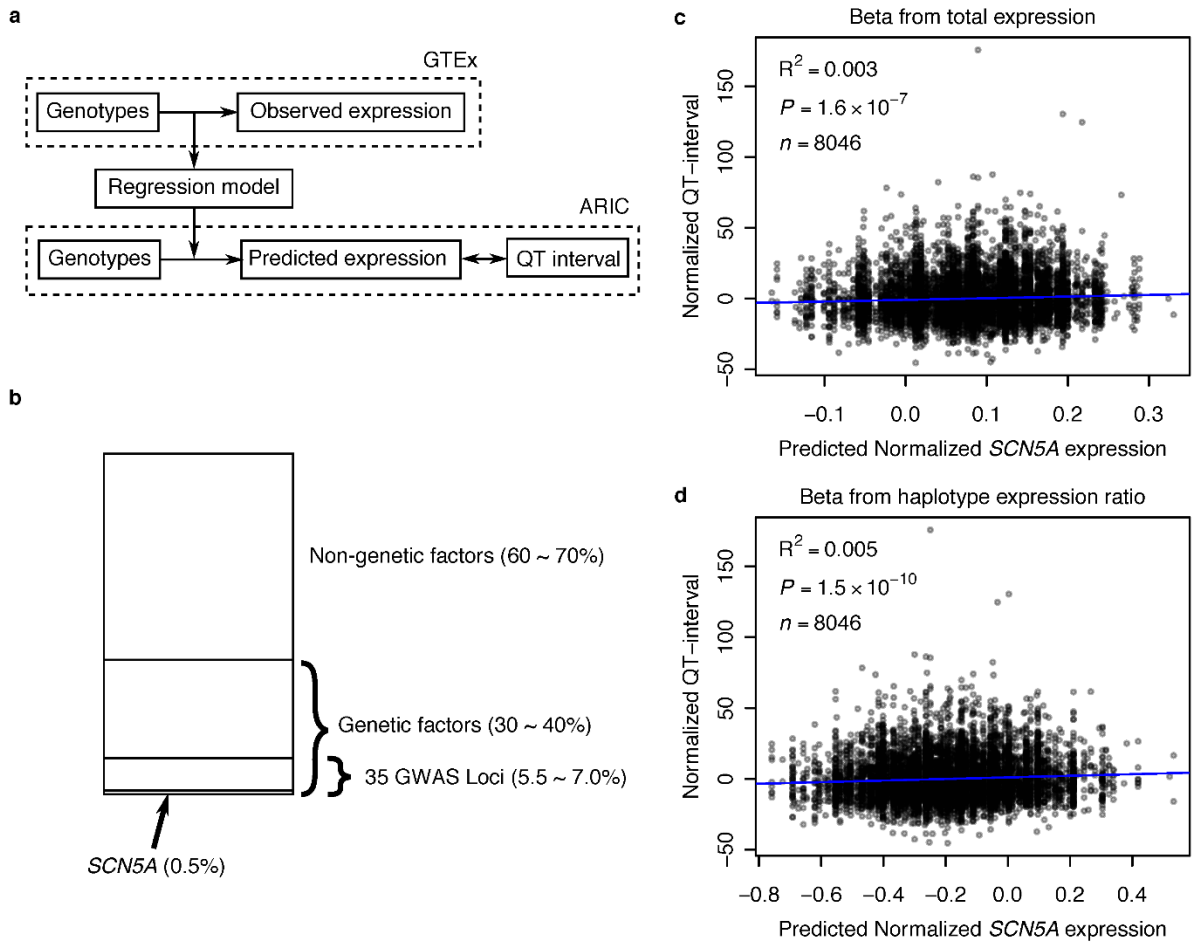

Figure S6: Significant correlation between predicted *SCN5A* cardiac expression and QT interval. a) Schematic overview of our approach. A multivariable regression model was built from GTEx data (3) and subsequently used to predict gene expression. b) Classification of QT interval variation by genetic and non-genetic factors. The total estimated heritability of QT interval is from Newton-Cheh *et al.* (4), heritability of QT interval explained by the 35 GWAS loci from Arking et al. (1), and heritability of QT interval explained by the *SCN5A* GWAS locus alone is the R^2 from the standard linear regression model between predicted *SCN5A* cardiac expression and QT interval in the ARIC cohort (5). (c-d) Comparison of predicted normalized *SCN5A* cardiac expression and normalized QT interval for all individuals in the ARIC cohort

using beta values for the 7 most significant causal CRE variants from the total-expression-based model (c) and the haplotype-expression ratio-based model (d). The linear regression line is shown in blue.

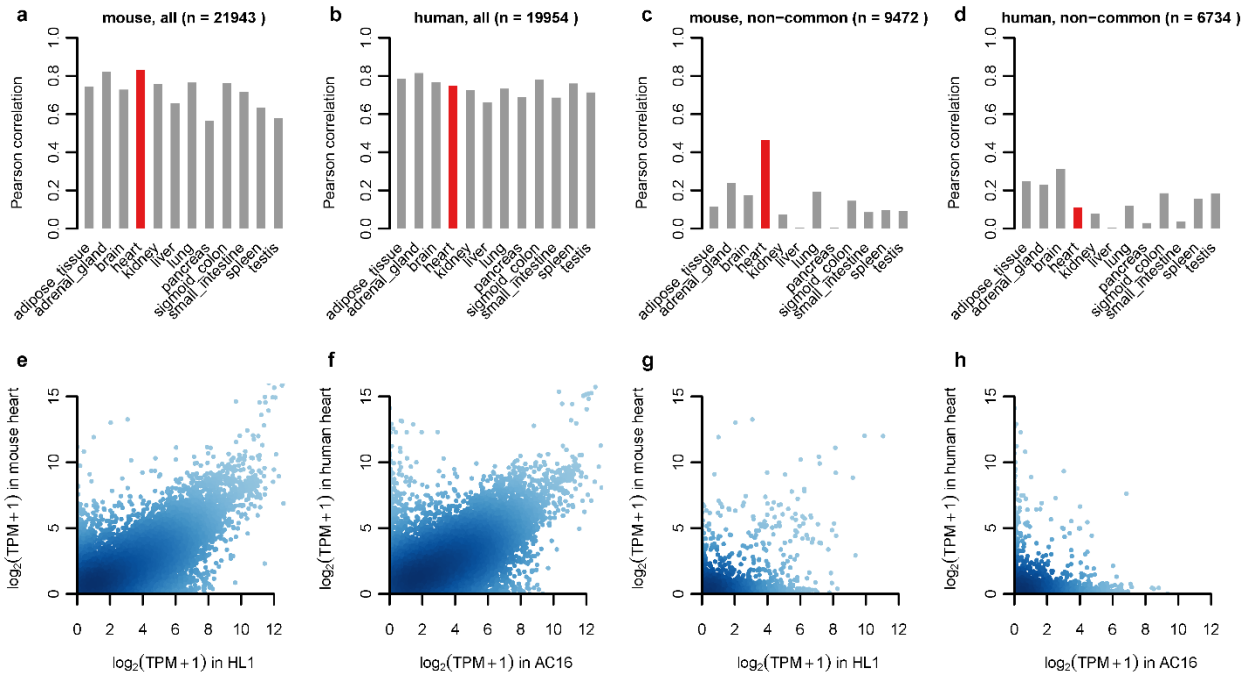

Figure S7: Gene expression profile of HL1 cells is mostly correlated with heart tissues. (a-d) Pearson correlation coefficients of gene expression profiles between HL1 (a, c) and AC16 (b, d) cell lines, and twelve mouse (a, c) and human (b, d) adult tissues using protein-coding genes (a, b) or non-common protein-coding genes (c, d). The heart tissue is highlighted in red. (e-h) Scatter plots of gene expression values between HL1 (e, g) and AC16 (f, h) cell lines, and the mouse (e, g) and human (f, h) adult heart tissues using protein-coding genes (e, f) or non-common protein-coding genes (g, h).

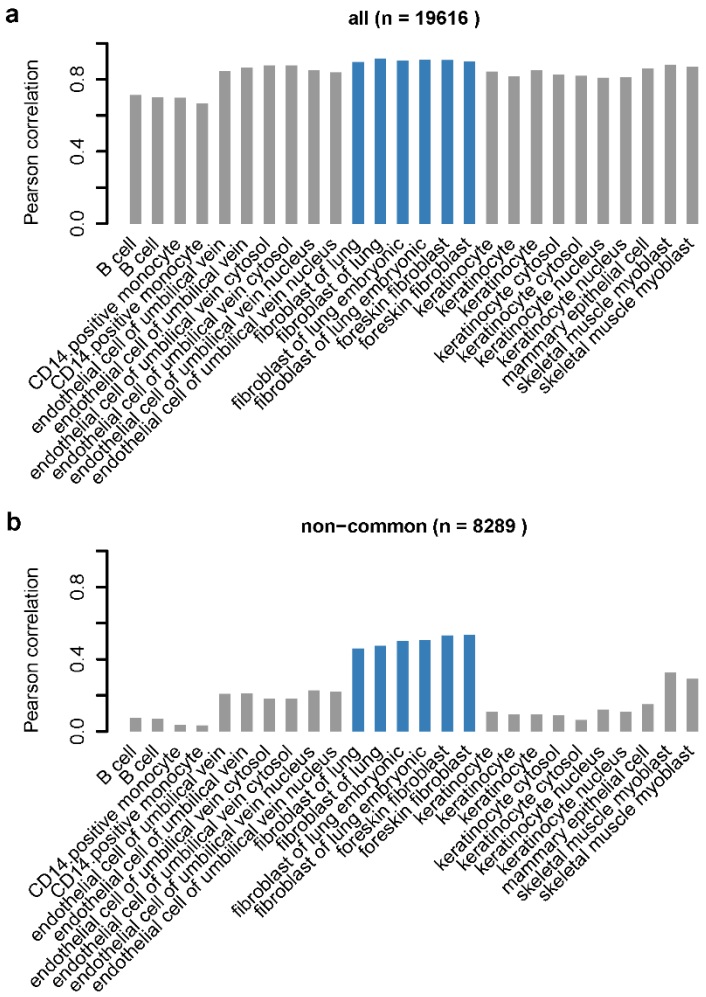

Figure S8: Gene expression profile of AC16 cells is mostly correlated with fibroblasts. Bar plots of Pearson correlation coefficients of gene expression values between AC16 and human primary cells using protein-coding genes (a) or non-common protein-coding genes (b). Fibroblast primary cells are highlighted in blue.

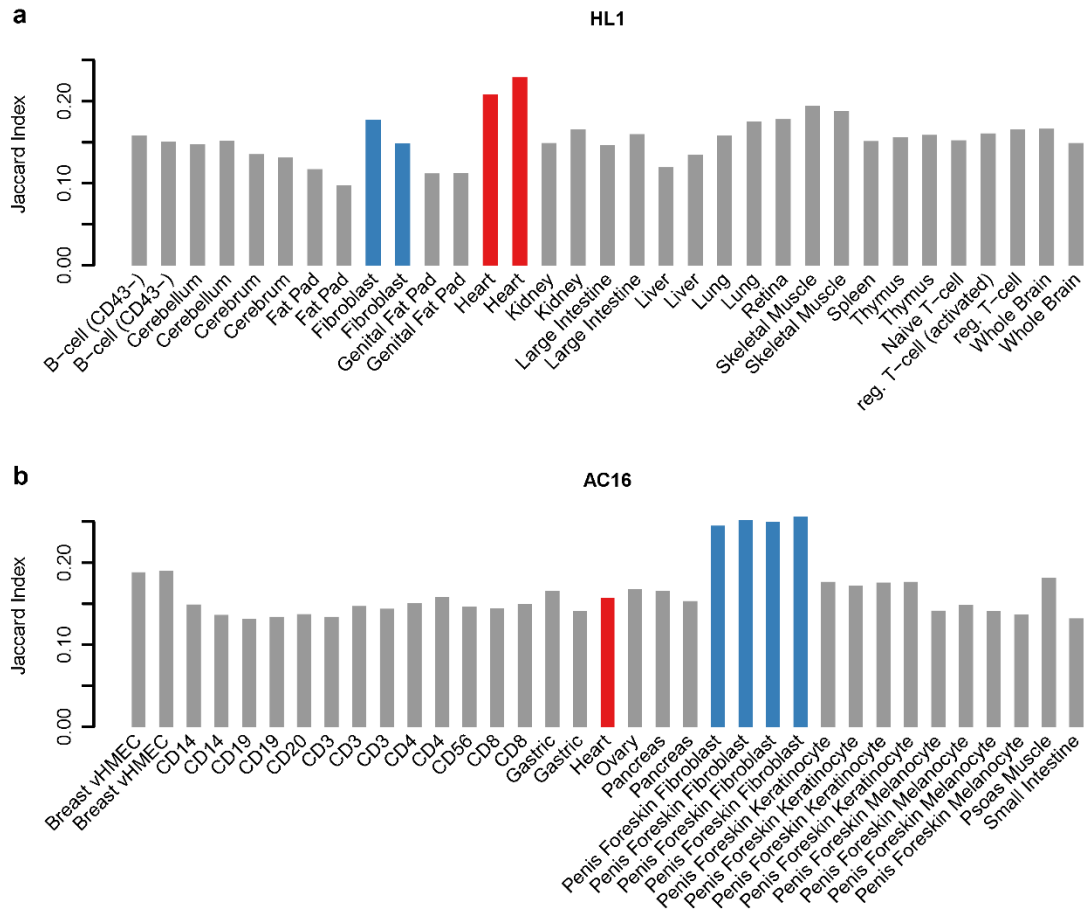

Figure S9: Comparisons of open chromatin regions in the HL1 and AC16 cell lines to tissues and primary cells. Bar plots of the Jaccard index of open chromatin between HL1 (a) or AC16 (b) and mouse (a) or human (b) tissues and primary cells. Heart tissues are highlighted in red and fibroblast primary cells in blue.

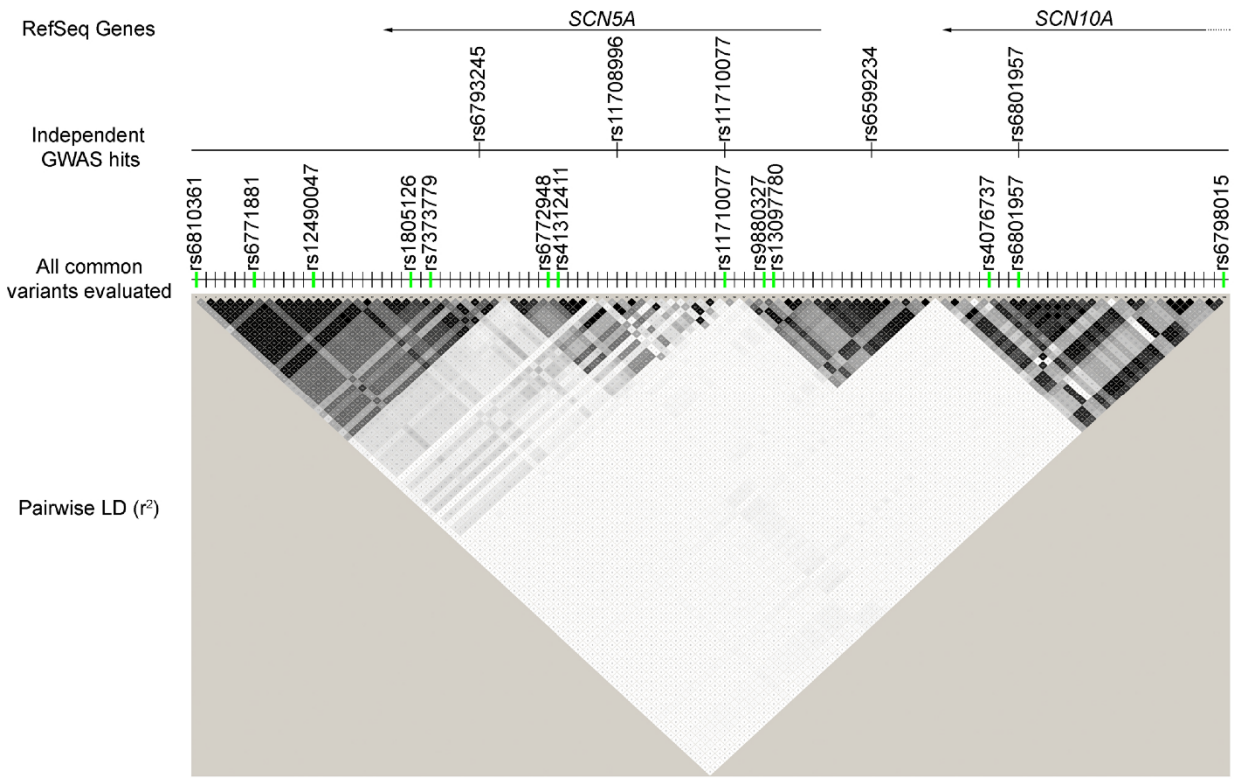

Figure S10: Multiple functional *cis*-regulatory enhancer variants underlying QT interval association at the *SCN5A-SCN10A* locus on chromosome 3p22.2. Multiple tracks showing (from top) the two protein-coding genes at the GWAS locus (RefSeq Genes), the five index/sentinel SNPs (Independent GWAS hits), all common (MAF>5%) SNPs in European ancestry subjects in moderate-to-high LD (r^2>0.3) with the five index SNPs that were evaluated in reporter assays (All common variants evaluated), and, the pairwise LD map for these variants (Pairwise LD (r^2)). Variants are shown as tick marks: *cis*-regulatory enhancer variants are in green while others are in black. LD measured by r^2 is shown on a grayscale (white: r^2=0; black: r^2=1).

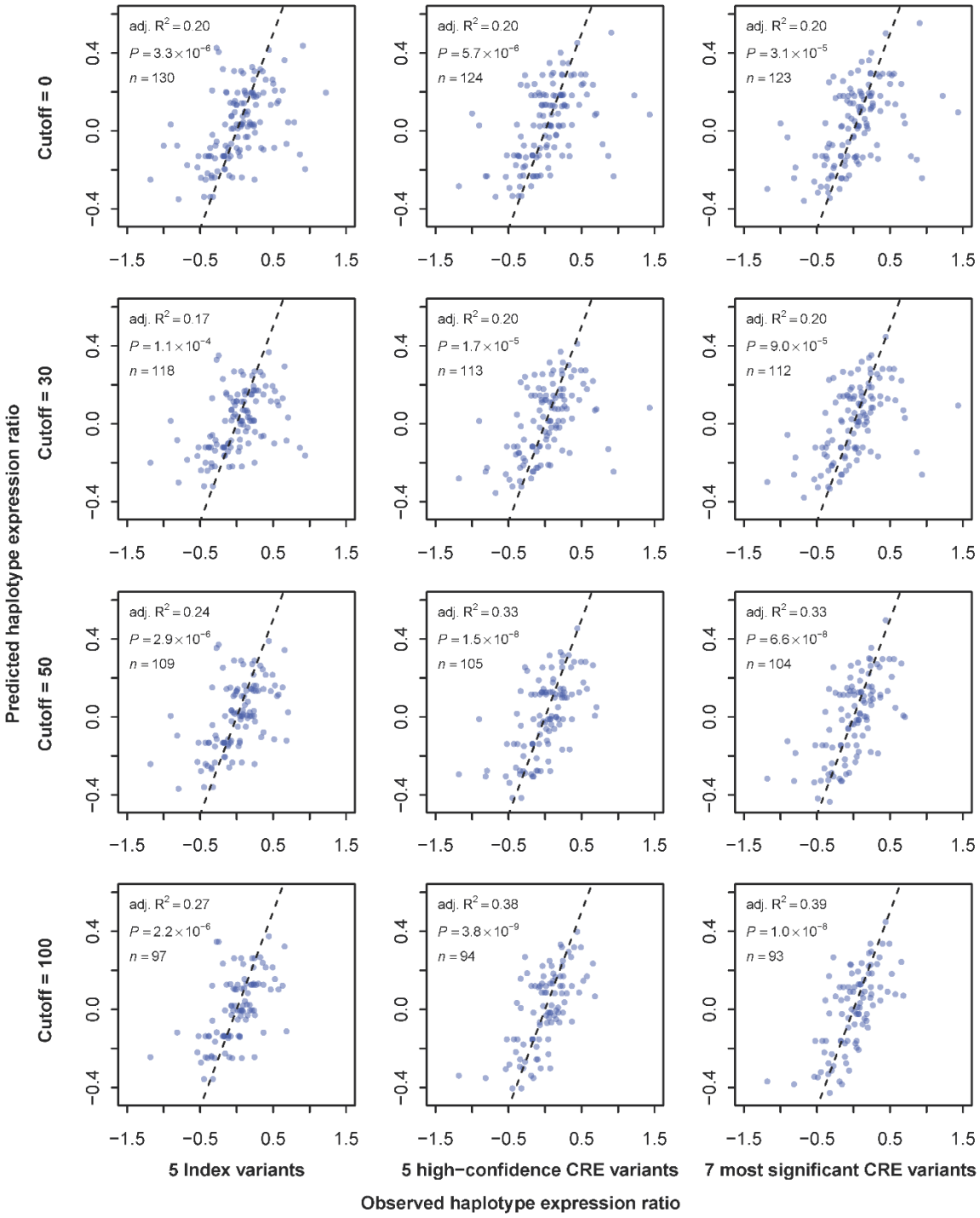

Figure S11: Removal of noisy samples using a read count filter improves the haplotype expression ratio-based models. Regression models were built by varying the cutoff of read counts (n=0, 30, 50, and 100) for each of the three sets of

variants (5 GWAS index SNPs, 5 high-confidence causal CRE variants, and 7 most significant causal CRE variants), with increasing stringency yielding better correlations.

Table S1: Target regions around each of the five QT interval associated independent variants at the *SCN5A-SCN10A* locus based on HapMap recombination hotspots (hg19 build).

Variant	Upstream hotspot position	Downstream hotspot position	Target region size
rs6793245	chr3:38352450	chr3:38601556	249,106
rs11708996	chr3:38607303	chr3:38659275	51,972
rs11710077	chr3:38607303	chr3:38659275	51,972
rs6599234	chr3:38659275	chr3:38724849	65,574
rs6801957	chr3:38726685	chr3:38803509	76,824

Table S2: Human and mouse tissues with ENCODE RNA-seq data downloaded from the ENCODE project website.

Organism	Tissue	Sex	Read1 File Accession	Read2 File Accession
Homo sapiens	Adipose tissue	female	ENCFF592VVB	ENCFF359HIQ
Homo sapiens	Adrenal gland	male	ENCFF709FHN	ENCFF681HNP
Homo sapiens	Brain	female	ENCFF456MMS	ENCFF716WNR
Homo sapiens	Heart	female	ENCFF464TEM	ENCFF221QNJ
Homo sapiens	Kidney	male	ENCFF044VER	ENCFF640PYL
Homo sapiens	Liver	male	ENCFF650JAM	ENCFF803DXA
Homo sapiens	Lung	female	ENCFF911WYB	ENCFF723QXK
Homo sapiens	Pancreas	female	ENCFF541KUW	ENCFF386MOY
Homo sapiens	Sigmoid colon	male	ENCFF322RPT	ENCFF782AHJ
Homo sapiens	Small intestine	female	ENCFF540SNP	ENCFF999PRA
Homo sapiens	Spleen	female	ENCFF926YPC	ENCFF111IRS
Homo sapiens	Testis	male	ENCFF034BHU	ENCFF788AET
Mus musculus	Adipose tissue	male	ENCFF128KGA	ENCFF510DLJ
Mus musculus	Adrenal gland	female	ENCFF307YNT	ENCFF731TMT
Mus musculus	Brain	male	ENCFF445AWP	ENCFF958CHE
Mus musculus	Heart	female	ENCFF104UFH	ENCFF126WYO
Mus musculus	Kidney	female	ENCFF214JLE	ENCFF883OGF
Mus musculus	Liver	female	ENCFF492PRP	ENCFF581OEV
Mus musculus	Lung	female	ENCFF146GNS	ENCFF280IUI
Mus musculus	Pancreas	female	ENCFF849WWH	ENCFF224IMT
Mus musculus	Sigmoid colon	female	ENCFF631QWX	ENCFF016KHB
Mus musculus	Small intestine	female	ENCFF933LCA	ENCFF684WUW
Mus musculus	Spleen	male	ENCFF432KKN	ENCFF859JTH
Mus musculus	Testis	male	ENCFF682XSC	ENCFF690HKC

Table S3: Human primary cells with ENCODE RNA-seq data downloaded from the ENCODE project website.

File Accession	Tissue	Sex	Subcellular	Replicate
ENCFF322LVK	foreskin fibroblast	male		1
ENCFF836HXE	foreskin fibroblast	male		2
ENCFF891KWL	fibroblast of lung	male		1
ENCFF388BCD	fibroblast of lung	male		2
ENCFF970EZY	keratinocyte	female	nucleus	3
ENCFF759EQF	keratinocyte	female	nucleus	4
ENCFF809ETA	endothelial cell of umbilical vein	male	nucleus	4
ENCFF492RYM	endothelial cell of umbilical vein	male	nucleus	3
ENCFF153AIN	endothelial cell of umbilical vein	male		1
ENCFF530HNB	endothelial cell of umbilical vein	male		2
ENCFF031XBL	endothelial cell of umbilical vein	male	cytosol	4
ENCFF812AVH	endothelial cell of umbilical vein	male	cytosol	3
ENCFF081MXC	CD14-positive monocyte	female		1
ENCFF669GZO	CD14-positive monocyte	female		2
ENCFF052HXO	keratinocyte	female	cytosol	3
ENCFF063XVI	keratinocyte	female	cytosol	4
ENCFF049GRB	B cell	female		1
ENCFF422SXS	B cell	female		2
ENCFF229LIX	skeletal muscle myoblast	unknown		1
ENCFF503QBP	skeletal muscle myoblast	unknown		2
ENCFF372VME	keratinocyte	female		1
ENCFF108ERJ	keratinocyte	female		2
ENCFF699KSV	keratinocyte	female		5
ENCFF458UOE	fibroblast of lung	male		2
ENCFF461AOR	fibroblast of lung	male		1
ENCFF300LEY	mammary epithelial cell	female		1

Table S4: Human tissues and primary cells with DNase-seq data downloaded from the NIH RoadMap Epigenomics project website.

File name	Tissue/Primary cell
UW.Breast_vHMEC.ChromatinAccessibility.RM035.DS18406.bed.gz	Breast vHMEC
UW.Breast_vHMEC.ChromatinAccessibility.RM035.DS18438.bed.gz	Breast vHMEC
UW.CD14_Primary_Cells.ChromatinAccessibility.RO_01689.DS17391.bed.gz	CD14
UW.CD14_Primary_Cells.ChromatinAccessibility.RO_01701.DS17889.bed.gz	CD14
UW.CD19_Primary_Cells.ChromatinAccessibility.RO_01679.DS17186.bed.gz	CD19
UW.CD19_Primary_Cells.ChromatinAccessibility.RO_01689.DS17281.bed.gz	CD19
UW.CD20_Primary_Cells.ChromatinAccessibility.RO_01778.DS17371.bed.gz	CD20
UW.CD3_Primary_Cells.ChromatinAccessibility.CB6_1-4-2011.DS17706.bed.gz	CD3
UW.CD3_Primary_Cells.ChromatinAccessibility.RO_01679.DS17198.bed.gz	CD3
UW.CD3_Primary_Cells.ChromatinAccessibility.RO_01701.DS17534.bed.gz	CD3
UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01679.DS17212.bed.gz	CD4
UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01701.DS17881.bed.gz	CD4
UW.CD56_Primary_Cells.ChromatinAccessibility.RO_01679.DS17189.bed.gz	CD56
UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01679.DS17203.bed.gz	CD8
UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01701.DS17885.bed.gz	CD8
UW.Gastric.ChromatinAccessibility.STL001.DNase.DS20260.bed.gz	Gastric
UW.Gastric.ChromatinAccessibility.STL003.DNase.DS20748.bed.gz	Gastric
UW.Heart.ChromatinAccessibility.STL001.DNase.DS20383.bed.gz	Heart
UW.Ovary.ChromatinAccessibility.STL002.DNase.DS20827.bed.gz	Ovary
UW.Pancreas.ChromatinAccessibility.STL002.DNase.DS20842.bed.gz	Pancreas
UW.Pancreas.ChromatinAccessibility.STL003.DNase.DS20753.bed.gz	Pancreas
UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin01.DS18224.bed.gz	Penis Foreskin Fibroblast
UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin01.DS18229.bed.gz	Penis Foreskin Fibroblast
UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin02.DS18252.bed.gz	Penis Foreskin Fibroblast
UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin02.DS18256.bed.gz	Penis Foreskin Fibroblast
UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18692.bed.gz	Penis Foreskin Keratinocyte
UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18695.bed.gz	Penis Foreskin Keratinocyte
UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin02.DS18714.bed.gz	Penis Foreskin Keratinocyte
UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin02.DS18718.bed.gz	Penis Foreskin Keratinocyte
UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18590.bed.gz	Penis Foreskin Melanocyte
UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18601.bed.gz	Penis Foreskin Melanocyte
UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin02.DNase.DS18668.bed.gz	Penis Foreskin Melanocyte
UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin02.DNase.DS19662.bed.gz	Penis Foreskin Melanocyte
UW.Psoas_Muscle.ChromatinAccessibility.STL001.DNase.DS20325.bed.gz	Psoas Muscle
UW.Small_Intestine.ChromatinAccessibility.STL003.DNase.DS20770.bed.gz	Small Intestine

Table S5: Mouse tissues and primary cells with ENCODE DNase-seq data downloaded from the UCSC Genome Browser.

File name	Tissue/Primary cell
wgEncodeUwDnaseBcellcd43nC57bl6MAdult8wksAlnRep1.bam	B-cell (CD43-)
wgEncodeUwDnaseBcellcd43nC57bl6MAdult8wksAlnRep2.bam	B-cell (CD43-)
wgEncodeUwDnaseCbellumC57bl6MAdult8wksAlnRep2.bam	Cerebellum
wgEncodeUwDnaseCerebellumC57bl6MAdult8wksAlnRep1.bam	Cerebellum
wgEncodeUwDnaseCerebrumC57bl6MAdult8wksAlnRep1.bam	Cerebrum
wgEncodeUwDnaseCerebrumC57bl6MAdult8wksAlnRep2.bam	Cerebrum
wgEncodeUwDnaseFatC57bl6MAdult8wksAlnRep1.bam	Fat Pad
wgEncodeUwDnaseFatC57bl6MAdult8wksAlnRep2.bam	Fat Pad
wgEncodeUwDnaseFibroblastC57bl6MAdult8wksAlnRep1.bam	Fibroblast
wgEncodeUwDnaseFibroblastC57bl6MAdult8wksAlnRep2.bam	Fibroblast
wgEncodeUwDnaseGfatC57bl6MAdult8wksAlnRep1.bam	Genital Fat Pad
wgEncodeUwDnaseGfatC57bl6MAdult8wksAlnRep2.bam	Genital Fat Pad
wgEncodeUwDnaseHeartC57bl6MAdult8wksAlnRep1.bam	Heart
wgEncodeUwDnaseHeartC57bl6MAdult8wksAlnRep2.bam	Heart
wgEncodeUwDnaseKidneyC57bl6MAdult8wksAlnRep1.bam	Kidney
wgEncodeUwDnaseKidneyC57bl6MAdult8wksAlnRep2.bam	Kidney
wgEncodeUwDnaseLgintC57bl6MAdult8wksAlnRep1.bam	Large Intestine
wgEncodeUwDnaseLgintC57bl6MAdult8wksAlnRep2.bam	Large Intestine
wgEncodeUwDnaseLiverC57bl6MAdult8wksAlnRep1.bam	Liver
wgEncodeUwDnaseLiverC57bl6MAdult8wksAlnRep2.bam	Liver
wgEncodeUwDnaseLungC57bl6MAdult8wksAlnRep1.bam	Lung
wgEncodeUwDnaseLungC57bl6MAdult8wksAlnRep2.bam	Lung
wgEncodeUwDnaseRetinaC57bl6MAdult8wksAlnRep1.bam	Retina
wgEncodeUwDnaseSkmuscleC57bl6MAdult8wksAlnRep1.bam	Skeletal Muscle
wgEncodeUwDnaseSkmuscleC57bl6MAdult8wksAlnRep2.bam	Skeletal Muscle
wgEncodeUwDnaseSpleenC57bl6MAdult8wksAlnRep1.bam	Spleen
wgEncodeUwDnaseThymusC57bl6MAdult8wksAlnRep1.bam	Thymus
wgEncodeUwDnaseThymusC57bl6MAdult8wksAlnRep2.bam	Thymus
wgEncodeUwDnaseTnaiveC57bl6MAdult8wksAlnRep1.bam	Naive T-cell
wgEncodeUwDnaseTregaC57bl6MAdult8wksAlnRep1.bam	reg. T-cell (activated)
wgEncodeUwDnaseTregC57bl6MAdult8wksAlnRep1.bam	reg. T-cell
wgEncodeUwDnaseWbrainC57bl6MAdult8wksAlnRep1.bam	Whole Brain
wgEncodeUwDnaseWbrainC57bl6MAdult8wksAlnRep2.bam	Whole Brain

References:

1. Arking DE*, et al* (2014) Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet* 46(8): 826-836.

2. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437(7063): 1299-1320.

3. GTEx Consortium (2015) Human genomics. the genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 348(6235): 648-660.

4. Newton-Cheh C*, et al* (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The framingham heart study. *Heart Rhythm* 2(3): 277-284.

5. The ARIC Investigators (1989) The atherosclerosis risk in communities (ARIC) study: Design and objectives. *Am J Epidemiol* 129(4): 687-702.