Supplementary Online Material

"Twenty-eight genetic loci associated with ST-T wave amplitudes of the electrocardiogram"

Contents

1. Supplementary Note

1a. Cohort methods, details on participating studies

The Prevention of REnal and Vascular ENd stage Disease (PREVEND, discovery)

The PREVEND study is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Inhabitants 28 to 75 years of age (n=85,421) in the city of Groningen, The Netherlands, were asked to complete a short questionnaire, 47% responded, and individuals were then selected with a urinary albumin concentration of at least 10 mg/L (n = 7,768) and a randomly selected control group with a urinary albumin concentration less than 10 mg/L (n = 3,395). Details of the protocol have been described elsewhere (www.prevend.org). Standard 12-lead electrocardiograms were recorded with CardioPerfect equipment (Cardio Control; currently Welch Allyn, Delft, The Netherlands), the digital measurements of the ST-T wave amplitudes were extracted and processed using the Modular ECG Analysis System (MEANS)(1).

Lifelines (discovery)

LifeLines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity and complex genetics. Details of the protocol have been described elsewhere (https://www.lifelines.nl/lifelines-research/news). Standard 12-lead electrocardiograms were recorded with CardioPerfect equipment (Cardio Control; currently Welch Allyn, Delft, The Netherlands), the digital measurements of the ST-T wave amplitudes were extracted and processed using the Modular ECG Analysis System (MEANS)(1).

Rotterdam study I, II and III (RS I,II,III, replication)

The Rotterdam Study is a prospective population-based cohort study comprising 7,983 subjects aged 55 years or older (RS-I), which started in 1990. In 2000-2001, an additional 3,011 individuals aged 55 years or older were recruited (RS-II). In the RS-I and RS-II, electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QRS intervals were made using the Modular ECG Analysis System (MEANS). MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the ST-T wave amplitudes among the other parts of the signal.

Erasmus Rucphen Family study (ERF, replication)

The Erasmus Rucphen Family study is comprised of a family-based cohort embedded in the Genetic Research in Isolated Populations (GRIP) program in the southwest of the Netherlands. The aim of this program is to identify genetic risk factors for the development of complex disorders. In ERF, twenty-two families that had a minimum of five children baptized in the community church between 1850 and 1900

were identified with the help of detailed genealogical records. All living descendants of these couples, and their spouses, were invited to take part in the study. Comprehensive interviews, questionnaires, and examinations were completed at a research center in the area; approximately 3,200 individuals participated. Examinations included 12 lead ECG measurements. Electrocardiograms were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements were made using the Modular ECG Analysis System (MEANS). The QRS detector of MEANS operates on multiple simultaneously recorded leads, which are transformed to a detection function that brings out the QRS complexes among the other parts of the signal. Data collection started in June 2002 and was completed in February 2005.

Atherosclerosis Risk in Communities study (ARIC, replication)

The Atherosclerosis Risk in Communities (ARIC) Study is a prospective community-based study of cardiovascular disease and its risk factors. At baseline (1987-89), 15,792 men and women age 45-64 were recruited from 4 communities in the US (Washington County, Maryland; Forsyth County, North Carolina; Jackson, Mississippi; Minneapolis suburbs, Minnesota). Participants were mostly white in the Minnesota and Washington County field centers, white and African American in Forsyth County, and exclusively African American in the Jackson field center. ECGs were obtained from the MAC Personal Computer (Marquette Electronics, Milwaukee, WI, USA) and processed with the GE Marquette 12-SL program (GE Marquette, Milwaukee, WI, USA)

Cardiovascular Health Study (CHS, replication)

The CHS is a population-based cohort study of risk factors for CHD and stroke in adults ≥65 years conducted across four field centers [PMID: 1669507]. The original predominantly Caucasian cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons were enrolled for a total sample of 5,888. DNA was extracted from blood samples drawn on all participants at their baseline examination in 1989-90. In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system on 3980 CHS participants who were free of CVD at baseline, consented to genetic testing, and had DNA available for genotyping.

A total of 1908 persons were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack or lack of available DNA. Because the other cohorts were predominantly white, the African American participants were excluded from this analysis to reduce the possibility of confounding by population structure. Participants were excluded from analysis for sex mismatch, discordance with prior genotyping, or call rate < 95%. Genotypes and ECG phenotypes were available for 2,953 European ancestry participants; these individuals constitute the CHS sample for this study.

Study electrocardiograms were recorded using MAC PC ECG machines (Marquette Electronics, Milwaukee, Wisconsin) in all clinical centers. ECGs were initially processed in a central laboratory at the EPICORE Center (University of Alberta, Edmonton, Alberta, Canada) and during later phases of the study, at the EPICARE Center (Wake Forest University, Winston-Salem, North Carolina). All ECGs were visually inspected for technical errors and inadequate quality. All measurements are from the baseline ECG for eligible

subjects. Initial ECG processing was done by the Dalhousie ECG program, and processing was later repeated with the 2001 version of the GE Marquette 12-SL program (GE Marquette, Milwaukee, Wisconsin).

Young Finns Study (YFS, replication)

The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent. ECGs were measured using Mac 800 & Mac 5500/Magellan systems.

1c. Regulatory footprint analyses for candidate cell lines

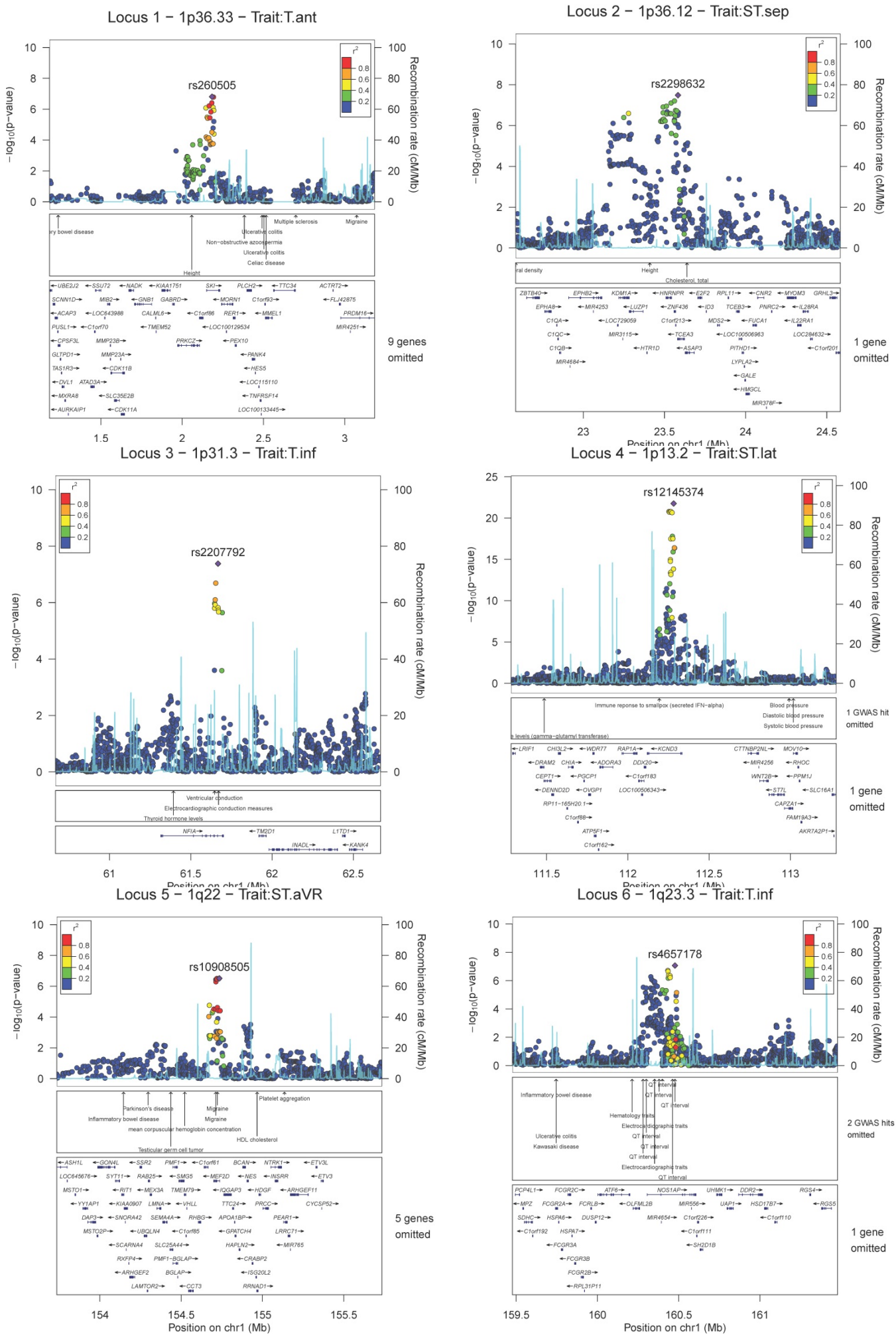
We used data on DNAse I hypersensitivity sites (DHSs) from 349 tissues (GEO accession numbers GSE29692 and GSE18927) of the ENCODE project(2) and Roadmap Epigenomics Program(3) with hotspots identified using the hotspot algorithm and peaks were called at 5% false discovery rate in a uniform manner, as previously described (4). Overlap of SNPs with regulatory DNA elements were performed after the genomic coordinates were converted from hg18 to hg19 using the UCSC liftOver tool. We compared the ratio of SNPs in peaks of regulatory DNA elements that meet increasing P-value cutoffs to the ratio of all SNPs in regulatory DNA elements of the heart (4). Next we tested whether SNPs in DHSs of fetal heart tissue (n=12) were enriched compared to other tissues (n=337) using z-score statistics.

Data (Aligned sequence reads, bed files) of 7 distinct histone modifications was obtained from the Roadmap Epigenomics Project release 8. Only samples with matching input DNA samples were included. If replicate experiments were available we aggregated the sequence reads. MACS (v1.4) software was used to identify significant peaks (1×10^{-3}) using a fixed DNA fragment size of 146 (5). For annotation of other functional elements relevant for the heart we used the called peak data from GEO (GSE35151, GSE32587 and GSE21529). This includes cardiac transcription factor data on Tbx3, Gata4 and Nkx2-5 from mouse heart(6); p300 marks in human adult and fetal heart and RNAP2 from human fetal heart (7); and Gata4, Mef2, Nkx2-5, Srf and Tbx5 from the atrial HL-1 cell line (8). Peaks from mice were lift-over to human using the UCSC Genome Browser liftOver tool with the options "-minMatch=0.1 -multiple" after extending the regions by 1kb(9).

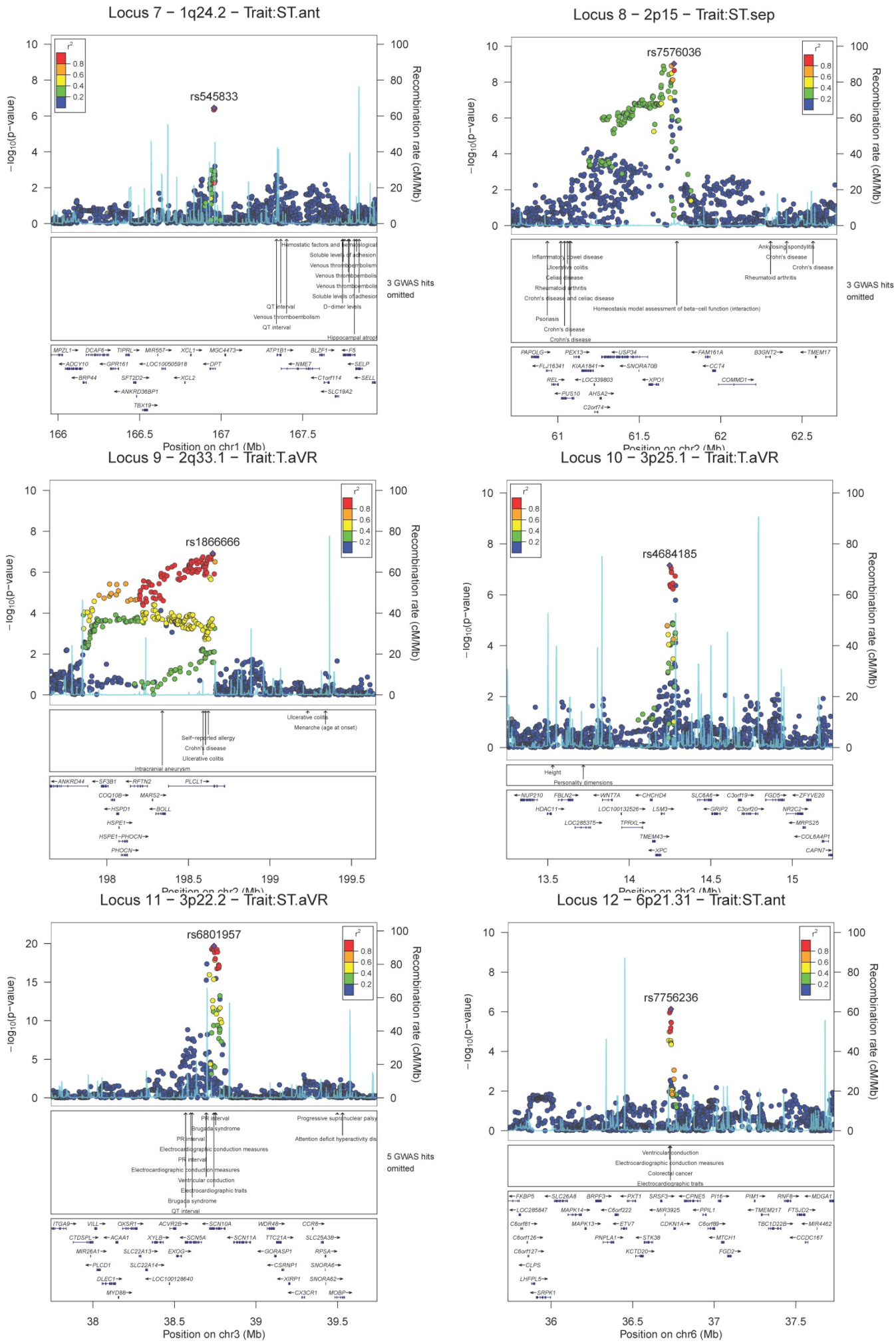
We intersected SNPs with regulatory elements relevant for the heart to provide a list of candidate functional SNPs underlying ST-T waves.

Supplementary Figures

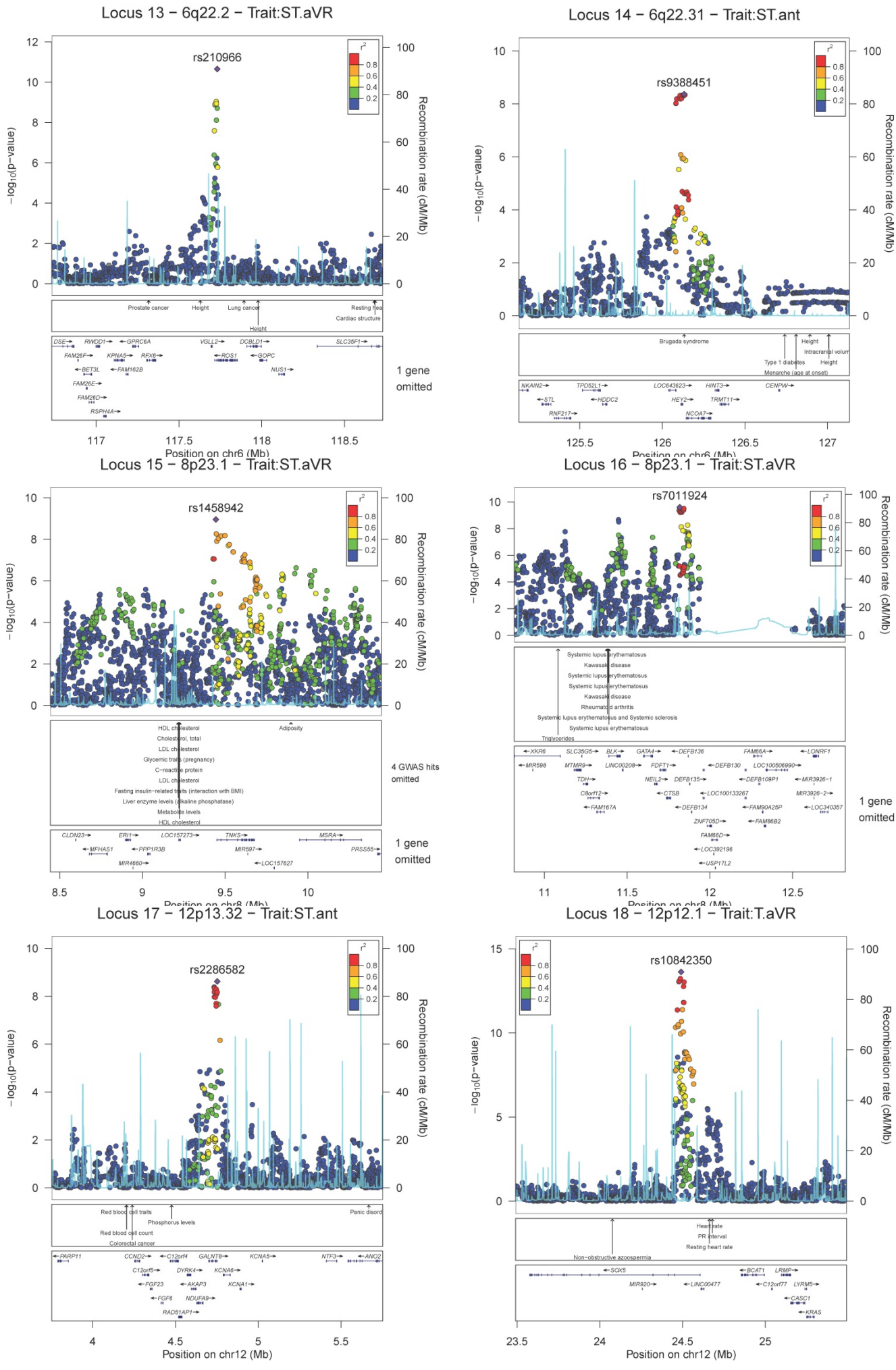
Supplementary Fig. 1 Regional plots from the 28 genome wide significant loci, of discovery. At each region pairwise LD with the sentinel SNP (strongest associated SNP across all genotypes and phenotypes) is indicated with color.



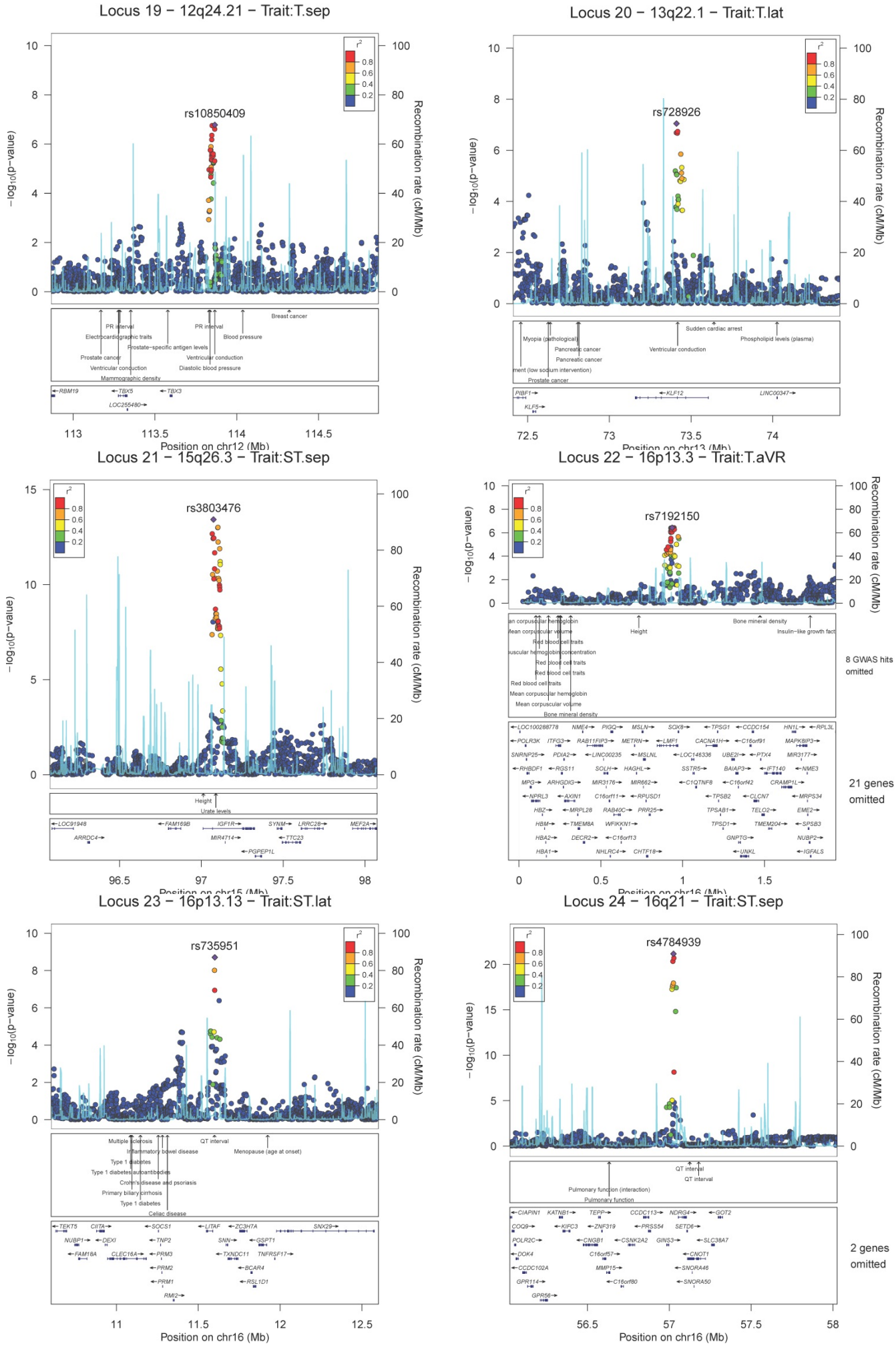
Supplementary Figure 1



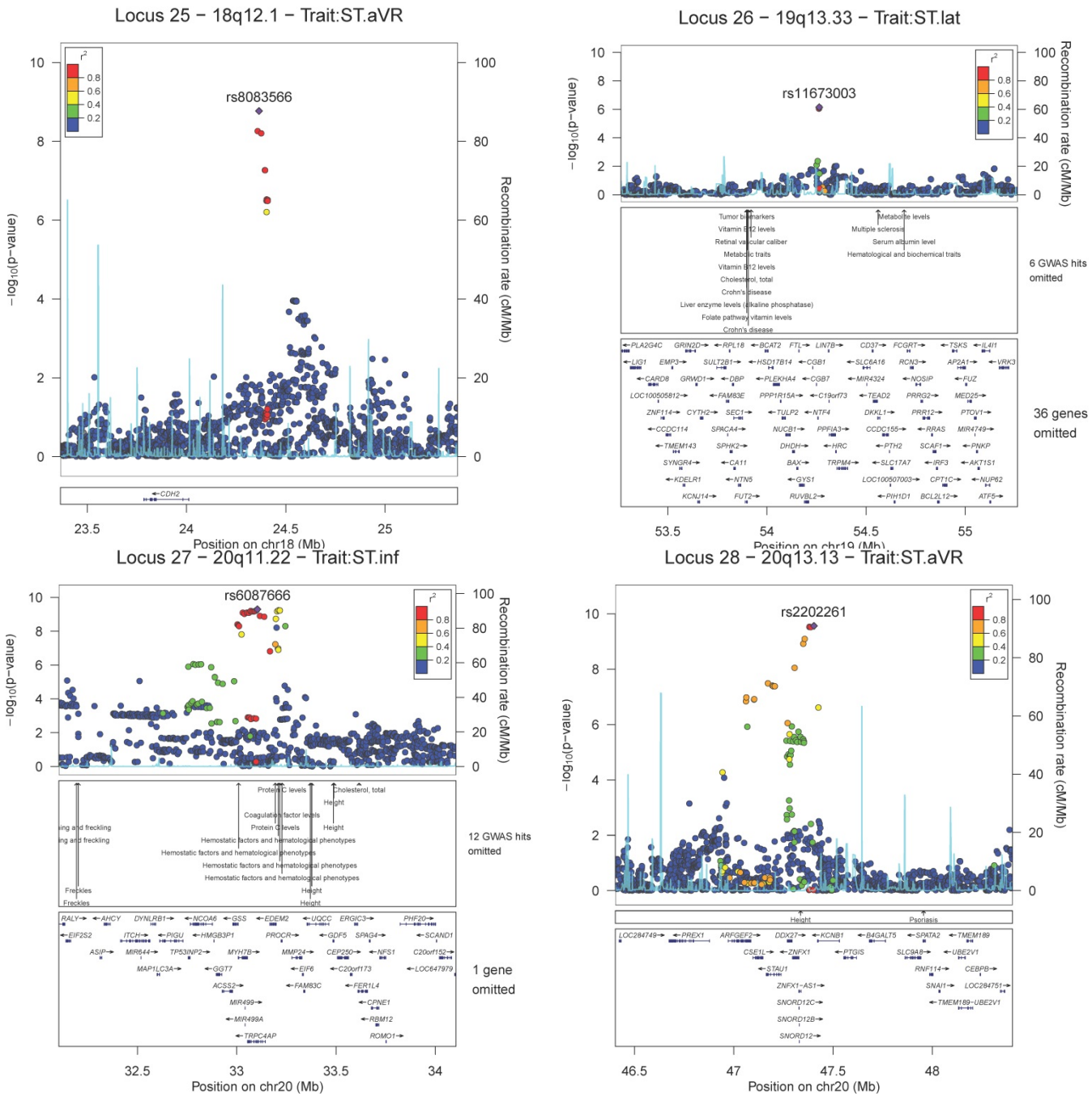
Supplementary Figure 1, continued



Supplementary Figure 1, continued



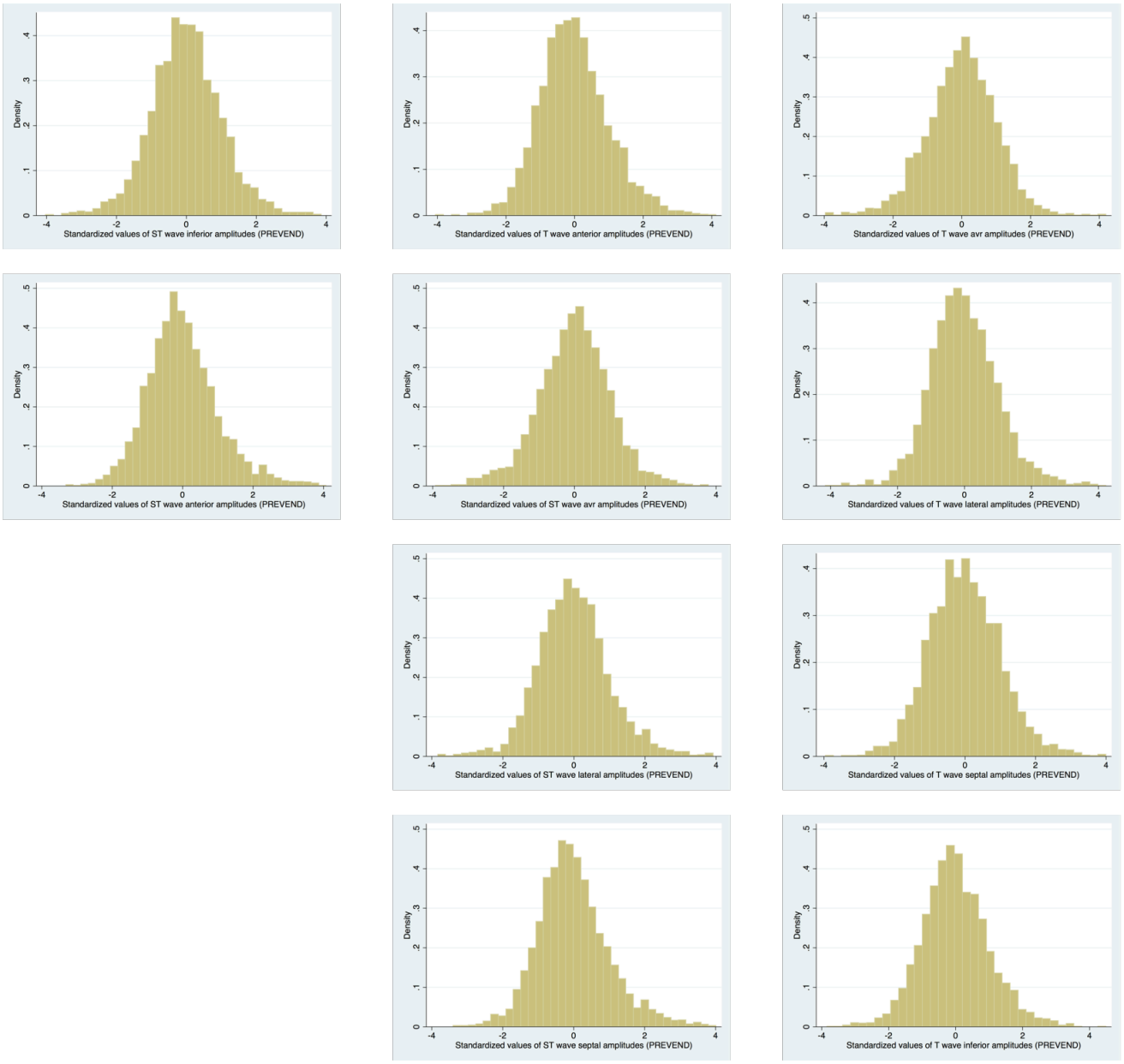
Supplementary Figure 1, continued



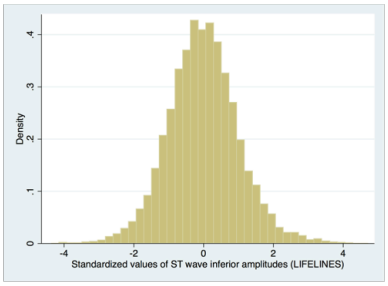
Supplementary Figure 1, continued

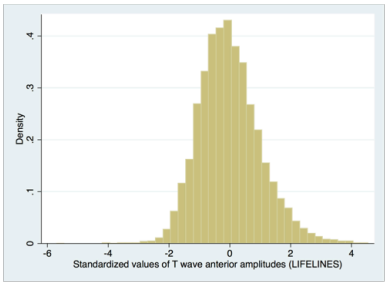
Supplementary Fig. 2 Histogram of each ST-T wave amplitude phenotype in (A) Prevend and (B) Lifelines.

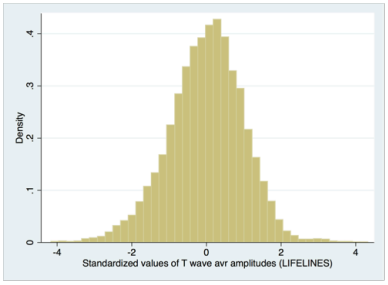
(A)

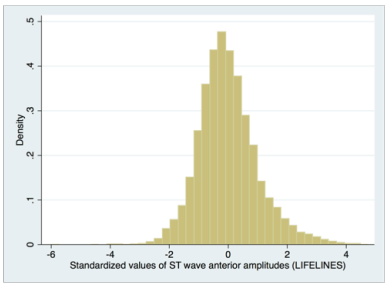


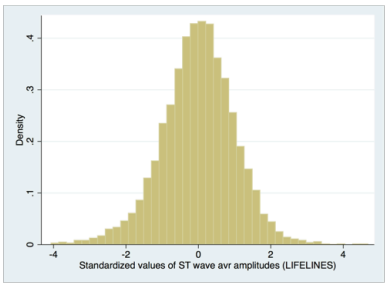
(B)

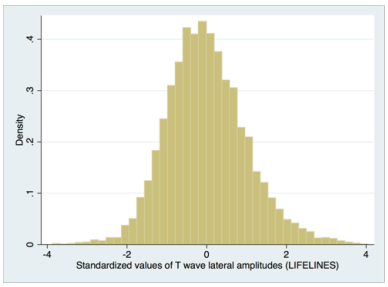

Density
Standardized values of ST wave inferior amplitudes (LIFELINES)

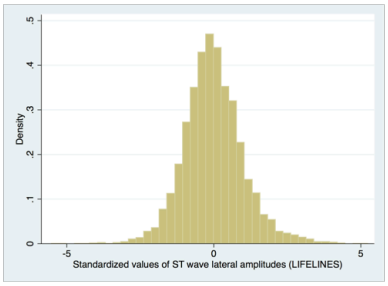

Density
Standardized values of T wave anterior amplitudes (LIFELINES)

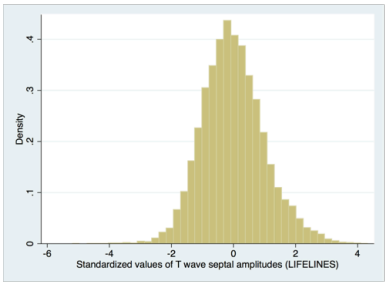

Density
Standardized values of T wave avr amplitudes (LIFELINES)

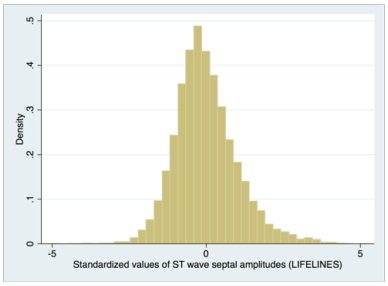

Density
Standardized values of ST wave anterior amplitudes (LIFELINES)

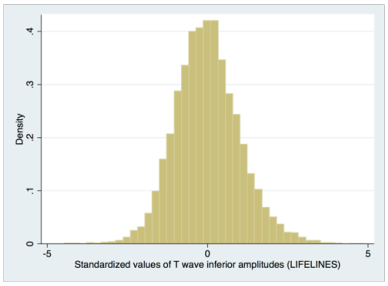

Density
Standardized values of ST wave avr amplitudes (LIFELINES)


Density
Standardized values of T wave lateral amplitudes (LIFELINES)


Density
Standardized values of ST wave lateral amplitudes (LIFELINES)


Density
Standardized values of T wave septal amplitudes (LIFELINES)


Density
Standardized values of ST wave septal amplitudes (LIFELINES)


Density
Standardized values of T wave inferior amplitudes (LIFELINES)

2. Literature

1 Kors, J.A. and van Herpen, G. (2009) Methodology of QT-interval measurement in the modular ECG analysis system (MEANS). *Annals of noninvasive electrocardiology : the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc*, **14 Suppl 1**, S48-53.

2 Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75-82.

3 Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. *et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*, **28**, 1045-1048.

4 Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.

5 Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nature protocols*, **7**, 1728-1740.

6 van den Boogaard, M., Wong, L.Y., Tessadori, F., Bakker, M.L., Dreizehnter, L.K., Wakker, V., Bezzina, C.R., t Hoen, P.A., Bakkers, J., Barnett, P. *et al.* (2012) Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *The Journal of clinical investigation*, **122**, 2519-2530.

7 May, D., Blow, M.J., Kaplan, T., McCulley, D.J., Jensen, B.C., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C. *et al.* (2012) Large-scale discovery of enhancers from human heart tissue. *Nature genetics*, **44**, 89-93.

8 He, A., Kong, S.W., Ma, Q. and Pu, W.T. (2011) Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5632-5637.

9 Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. *et al.* (2010) ChIP-Seq identification of weakly conserved heart enhancers. *Nature genetics*, **42**, 806-810.