TWENTY-EIGHT GENETIC LOCUS ASSOCIATED WITH HUMAN CARDIAC REPOLARIZATION
The ST segment and adjacent T wave amplitudes (ST-T wave) of the electrocardiogram are quantitative characteristics of cardiac repolarization. Repolarization abnormalities have been linked to ventricular arrhythmias and sudden cardiac death. We performed the first GWA meta-analysis of ST-T wave amplitudes in up to 37,977 individuals identifying 71 robust genotype-phenotype associations clustered within 28 independent loci. Fifty-four candidate genes were prioritized, including genes with established roles in cardiac repolarization phase (SCN5A/SCN10A, KCND3, KCNB1, NOS1AP, KCNA7, and HEY2) while others represent unknown molecular processes for the heart. These associations provide insights in the spatiotemporal contribution of genetic variation influencing cardiac repolarization and may provide novel leads for future functional follow-up.

Duration of cardiac repolarization has been previously studied by genome wide association studies and led to the discovery of 35 associated loci. However, abnormalities of repolarization are not limited to changes in duration but concern changes in amplitudes as well. Deviations of ST-T wave amplitudes can be indicative of a variety of cardiac pathologies, including myocardial ischemia, ventricular hypertrophy, long QT syndrome, early repolarization, and Brugada syndrome. For clinical use, phenotypes are dichotomized based on optimal sensitivity and specificity to predict worse outcome. However, there is no evidence that the underlying biology of ST-T wave is truly binary. Therefore, we considered that common variation involved in the biology of quantitative ST-T waves traits might provide additional biological insights in the (patho)physiological mechanisms of repolarization. Genome wide association analyses have been proven a powerful and unbiased tool to identify novel mechanisms and pathways.

Here we aim to identify variants that are associated with the heart’s repolarization phase during the ST-T wave of the ECG to further advance our knowledge on biological factors regulating cardiac repolarization.
**RESULTS**

**Genome wide discovery analysis and replication**

We performed genome-wide meta-analyses in 15,943 subjects of European descent on the ST-T wave (Table S1 and S2), with up to 2,316,136 imputed autosomal SNPs. Phenotypes for the ST segment amplitudes and T wave amplitudes were deducted from leads used in the clinic for diagnosing Brugada syndrome (septal: V1 and V2) and ER (lateral: I, aVL, V5 and V6 and inferior: II, III and aVF). Aiming to capture additional information available on the ECG during cardiac repolarization, we also investigated the other leads (anterior: V3, V4 and lead aVR), with presumed anatomical meaning during the ST-T wave. This resulted in 10 phenotypes representing the ST-T wave amplitude phenotypes: 5 ST segment amplitudes and 5 T wave amplitudes (Figure 1A).

Across the 10 phenotypes, a total of 1,340 unique SNPs passed the threshold of $P < 1 \times 10^{-6}$, 379 unique SNPs were genome wide significant ($P < 6.25 \times 10^{-9}$ based on 8 independent phenotypes). The discovery analysis identified 41 distinct genotype-phenotype associations clustered within 17 independent loci to be genome wide significant. Conditional analysis revealed a secondary signal (rs9851710 conditioned on rs6801957: $\beta_{(SE)}$ per minor allele =0.13 (0.0204), $P = 2.38 \times 10^{-10}$) in the SCN5A/SCN10A locus to remain genome wide associated with ST aVR amplitudes, this SNP was already identified to be associated with other ST-Twaves (Table S3). There was no evidence for inflation of test statistics in the final meta-analysis (Table S4) or significant heterogeneity.

To increase the robustness of the significant and suggestive ($P > 6.25 \times 10^{-9}$ and $P < 1.0 \times 10^{-6}$) associations in our discovery cohort we performed replication testing in an additional 22,034 individuals derived from 7 independent cohorts (Table S1). All genome wide significance loci in the discovery phase were all replicated. An additional 56 genotype-phenotype associations in 36 loci (20 additional loci) showed suggestive evidence for association of which 35 became genome wide significant. The 71 significant genotype-phenotype associations after replication (Table S3) are located in 28 independent loci (Table, Fig. 1B, Figure S1). Some loci were predominantly associated with ST segment amplitudes (locus 4, 5, 16, 26, and 28) while others were predominantly associated with T wave amplitudes (locus 6, 18 and 29). The test statistics for every combination of genotype-phenotype association are summarized in Table S5.
**FIG. 1B**

- **-log10(p)**

- **Fig. 1b**
- **Fig. 1c**

**INA depolarization**

- **SCN5A (locus 11)**

**I_{k1}\_fast**

- Partial repolarization
- KCND3 (locus 4)
- KGNA7 (locus 26)

**I_{k1}\_slow**

- Repolarization
- KCNB1 (locus 28)

**I_{k1}**

- Depolarization
- SCN5A (locus 11)
Figure 1a

We conducted genome wide analyses of ST-segment amplitudes at 80 milliseconds after J-point and T-wave (top) amplitudes reflecting temporal patterns in the cardiac cycle during the repolarization phase. In total 12 phenotypes were defined by taking the sum of the ST-T wave amplitudes in the lateral (I, aVL, V_6, and V_7), inferior (II, III, and aVF), septal (V_4 and V_6), anterior (V_3 and V_6), and lead aVR. These lead groups cover the combination of leads with presumed anatomical meaning of the heart, and those that are used in the clinic for diagnosing Brugada syndrome and Early repolarization.

(a) Genome wide association analyses of all ST-T wave traits identified 71 significant genotype-phenotype associations in 28 genetic loci (2MB). The x-axis represents the chromosomal position for each SNP, which was assigned the lowest P-value across the twelve traits; the y-axis represents the -log_{10} of the P-value for association. Twenty eight loci were significant for ST-T wave amplitudes.

(b) Four genes overlapping four loci are directly involved in the cardiac action potential; SNPs near SCN5A were associated with ST segment and T-top amplitudes whereas the loci containing potassium channel coding genes KCND3, KCNQ1, and KCNB1 were primarily associated with ST segment amplitudes.

Average fetal heart (n=12)
Average other tissue (n=337)
95% C.I

-log10(P) cutoff threshold
**fig. 2b**

**LEGEND**

Gene set P values

- $P < 10^{-2}$
- $P < 10^{-3}$
- $P < 10^{-4}$
- $P < 10^{-5}$

Gene set overlap (Pearson’s $r$)

- $r > 0.3$
- $r > 0.6$
**Figure 2a**

SNPs were significantly more enriched in DHs of fetal heart tissue (n=12) compared to other tissue and cells (n=337), across the full range of P-values of the discovery meta-analyses (genome wide), suggesting that functionality of regulatory DNA elements may underlie some of the associations.

**B**

DEPICT identified 56 significantly enriched gene-sets relevant for ST-T wave amplitudes. The most significant meta-term was "abnormal cardiovascular system physiology" (left), which consist of 8 individual reconstructed gene sets (right).

**C, D**

Next, we performed a meta-analysis of the 28 identified ST-T wave loci using 1000 Genomes imputed data, for this 1000 Genomes variants needed to be in LD $r^2<0.1$ with the HapMap sentinel SNP. Subsequent prioritization of potential causal annotations in these loci also suggested that regions of DHs in fetal heart are possibly underlying the associations as well as cardiac transcription factors, conserved regions, (exonic) active- and weak enhancers. While regions that are transcribed, tightly packed (heterochromatin) or function as promoters in the ventricles may be less important for the biological mechanisms of genetic variants that are associated with ST-T wave amplitudes. Subtle differences are present between the ventricles and fetal heart which could suggest that promoters that overlap potential causal ST-T wave SNPs may be active in the fetal heart but repressed in the ventricles. Percentages between parentheses indicate the amount of SNPs in the 28 loci overlapping with the annotation. Conservation (GERP & 29 Mammals), DHs of fetal heart, enhancers of the left ventricle and TB3 bound regions were used to prioritize potential causal SNPs in the 28 loci.

**Figure 2c**

Contiguous chromatin states (ChromHMM)

- ActiveEnhancer (-1.72%)
- WeakEnhancer (-3.56%)
- PooledPromoter (-0.40%)
- ZNF-GenesRepeats (-0.18%)
- TranscriptonRegulatory (-2.92%)
- BivalentPromoter (-0.67%)
- ActiveTSS (-0.34%)
- Quiescent-Low (-59.55%)
- WeakTranscription (-13.58%)
- Promoter (-1.56%)
- Transcription (-13.46%)
- RepressedPolyComb (-1.90%)
- Heterochromatin (-0.37%)

Log2 (Relative probability to be causal)
Previous genome-wide associations have studied other ECG indices such as QRS duration, PR interval, heart rate and QT interval. We intersected the previous identified SNPs with our ST-T wave loci and found overlap (within 2MB) in 13 loci. Lead SNPs from previous and current findings in 4 loci were in low LD ($r^2 < 0.02$), suggesting that 19 of the current loci are novel associations for ECG traits (Table S6).

Heritability estimates

The 28 identified sentinel SNPs collectively explained between 1.6% (T wave septal) and 5.1% (ST segment aVR) of the observed phenotypic variance (Table S7). Familial heritability estimates in the Erasmus Rucphen Family Study (ERF) varied between $h^2=30\%$ (T wave inferior) and $h^2=42\%$ (ST septal) suggesting additional genetic variants and mechanisms remain to be discovered (Table S8). The variation in proportion of variance associated with covariates varied more widely; between $r^2=0.09$ (ST segment inferior) and $r^2=0.36$ (ST segment septal).

Identification of candidate genes and pathway analyses

We prioritized 54 candidate genes in the 28 ST-T wave loci (Table). Thirty-four genes were prioritized based on the proximity criteria, 5 genes contained one or more non-synonymous SNPs (Table S9), 12 genes by eQTL analyses (Table S10), 65 genes based
on DEPICT analysis of which 25 were within genome-wide significant loci (at false discovery rate ≤ 5%; Table S1) and 3 genes based on literature mining using GRAIL (Table S12). The top 5 keywords retrieved from GRAIL were “muscle”, “channel”, “transcription”, “heart” and “channels”.

The DEPICT analysis was also applied to subsequently gain more detailed insight into the biology underlying changes in ST-T wave amplitudes of the electrocardiogram. This method identified 56 reconstituted gene sets in 9 independent meta gene sets. The top meta gene set represents various protein complexes likely related to Fascia Adherens whereas the second top meta gene set represents various protein complexes likely related to fascia adherens. The meta gene set (Abnormal Cardiovascular System Physiology) represents various expected gene sets (Figure 2, Table S13 and S14).

**Role of regulatory DNA and 1000 Genomes imputation**

ST-T wave associated SNPs were 3 fold enriched in DHSs from human fetal heart (Figure 2), which was significantly higher compared to 337 other cell types and tissues ($P = 0.0004$, Z-score statistics) in line with earlier observations. We then performed fine mapping of all significant genotype-phenotype associations (Table, Table S3) in Prevend and Lifelines imputed with 1000 Genomes, covering more of the known common variants in humans. We observed that 64 (of 71) associations were assigned to a different lead SNP and that 63 (of 71) associations were more significant in the 1000 Genomes imputed data. To facilitate future functional experiments towards the identification of causal variants and their underlying biological mechanisms, we prioritized potential causal SNPs using the probabilistic framework of PAINTOR. The 5 most significant annotations (conservation scores, DHSs of fetal heart, enhancers of the left ventricle and Tbx3 bound regions, see Figure 2) were used to prioritize potential causal SNPs in the 28 loci. This yielded 315 SNPs in the 99% confidence set and 96 SNPs in the 95% confidence set under the assumption of 1 causal SNP per locus (Table S15). As an illustration of functional follow-up we selected the two SNPs (rs34119223 ($P_{\text{posterior}}$=0.9999) intrinsic of KCND3 and rs4657166 ($P_{\text{posterior}}$=0.9999) intrinsic of NOS1AP) that were within the 10% credible set of PAINTOR and assessed differential protein binding by electrophoretic mobility shift assay’s (EMSA) on nuclear extract of adult mouse heart. For rs34119223 we observed a clear difference in nuclear protein binding of the reference allele (CC) compared to the alternative allele (the cc deletion, Figure S2). Rs34119223 is located in an conserved intronic enhancer region which may be specific for heart, brain and primary CD34+ cells (Figure S3).

Amplitudes of the ST segment and T wave are important traits that are associated with abnormal heart rhythm, conduction disturbances and ventricular arrhythmias. In this study we performed the first genome wide association analyses of ST-T wave amplitudes of the ECG in up to 37,977 individuals. Traits were defined according to the combination of ECG leads presumed to have anatomical meaning or used in the clinic for diagnosing Brugada syndrome and early repolarization. ST-T wave traits were moderately heritable ($h^2=30\%-42\%$) while the proportion of variance associated with covariates (gender, bmi and age) varied considerably (9%-36%), strongly supporting the genetic investigation of these traits. We identified 28 genome-wide significant loci for ST-T wave amplitudes and by expression QTL and bioinformatic strategies we prioritized a core set of 54 candidate genes and regions enriched for functions that are relevant to repolarization of the heart.

A recent genome wide association study on Brugada revealed the association of two loci that are shared by our ST-T wave loci. One of these signals in the SCN5A(SCN10A) locus is in complete LD with our sentinel SNP (rs10428132, $r^2=0.96$ with rs6801957) and rs9388451 near HEY2, is a ST-T wave sentinel SNP; suggesting that Brugada syndrome susceptibility loci share a common genetic background with ST-T wave traits. One of the strongest genome-wide associated signal for all amplitudes of the ST segment was in the KCND3 gene (locus 4). By means of fine mapping we identified an intronic cytosine-cytosine deletion to be a promising candidate causal SNP at the KCND3 locus that could exert its effect through disruption of a nuclear protein binding site. KCND3 encodes the Kv4.3 a-subunit that conducts the cardiac fast transient outward K+ current (I_{to}). This current is prominent in phase 1 of the action potential and contributes to the ‘notch’ of the cardiomyocyte’s action potential. Mutations in KCND3 have been implicated with Brugada syndrome, an arrhythmogenic disorder with increased risk of sudden cardiac death, as well as mutations in SCN5A and SCN10A. Here we associated loci containing potassium channels (KCND3 (locus 4), KCNB1 (locus 28) and KCNA7 (locus 26), predominantly with amplitudes of the ST segment and not with the T wave, suggesting that these potassium channels are specifically activated during the ST segment of the repolarization phase and may have clinical relevance. Future studies are required to examine the potential role of the other ST-T wave SNPs for increased arrhythmogenic risk and sudden death through an altered cardiac repolarization.
In addition to HEY2, a transcriptional regulator of cardiac electrical function in the right ventricular outflow tract, we identified six additional loci containing genes with strong evidence of being directly involved in cardiac tissue development via transcriptional regulatory pathways. Locus 5 contains a myocyte enhancer factor-2 (MEF2D) important for cardiac muscle morphogenesis and heart looping. Locus 18 and 22 contain two transcription factors from the SOX family, SOX5 and SOX8. The SOX5 is important for a functioning heart and involved in correct wnt signaling, suggesting multiple molecular-genetic mechanisms at this locus influence heart function. Transcripts of SOX8 are concentrated in the subendothelial mesenchym of the whole outflow tract, in which N-Cadherin is encoded by CDH2, locus 25, the transcription factor Tbx5 (TBX5, locus 19) and WNT4 (specific for the endocardial endothelial cushion development, part of the wnt pathway) in locus 2 also play essential developmental roles, among other cardiac cell lineages. Less characterized genes that could underlie cardiac repolarization through altered transcription, or development are: SKI, as mutations (1p36 deletion and sprintzen-Goldberg syndrome) are characterized by heart defects, brain abnormalities and muscle tone (hypotonia) in infancy; NFIA (locus 3, a transcription factor); KLF12 (locus 20, a transcription factor), VGLL2 (locus 13, a muscle specific transcription cofactor) and TNKS (locus 15, activator of the wnt pathway). Other interesting loci containing even more compelling cardiac genes are: SNPs near TMEM43 (locus 10) in which mutations cause arrhythmicogenic right ventricular cardiomyopathy and Emery-Dreifuss Muscular Dystrophy, and SNPs overlapping a sarcomeric gene MYH7B and miR499 (locus 27) which regulates a multitude of cardiac mRNA and microRNAs and promotes ventricular specification. SNPs in GALNT1, a glycosyltransferases that is required for normal heart valve development and cardiac function by regulating the extracellular matrix and altering conserved signaling pathways that regulate cell proliferation during heart development.

Five of the 28 ST-T wave sentinel SNPs were in LD with reported associations of QT-duration indicating that ST-T wave amplitudes encompass additional information of cardiac repolarization. The known functions of the candidate genes and our pathway analyses suggest that cardiac repolarization is influenced and regulated by a wide variety of molecular mechanisms, including ion-channels, structural proteins and cardiac transcription factors. However, the finding that protein complexes related to Fascia Adherens are most enriched in our pathway analyses hints that some of the biology underlying ST-T waves is less well captured in well-established pathways and (current-ly) better represented by data-driven (and not manually curated) pathways. Ventricular repolarization is a complex process, to capture this process we chose to apply a selection of composite ECG parameters with the aim to gain more insight into the biological processes underlying the ST-T wave. Based on our results it is tempting to speculate that some loci highlight spatiotemporal patterns. These surface ECG parameters are not highly specific to identify the exact anatomical region of the heart; e.g. the septal leads (V_s, V_p) may also include aspects of the right ventricular wall activity. However, there is no data available of more precise measurements such as could be derived from more sophisticated surface ECG equipment or intracardiac ECG measurements.

The main aim of this study was to elucidate and advance our understanding of cardiac repolarization, not to improve upon clinical diagnosis or risk stratification, although this should be subject to further investigation. There is some previous data on Brugada available, identifying two loci (HEY2, and SCN5A-SCN10A) both highly significant in our study. This suggest it might be of interest to perform look-ups (decrease multiple testing burden) of the ST-T wave associated variants. Unfortunately, as previous data on Brugada has not been made publically available and the authors were not willing to perform look-ups or perform a genetic risk score analysis, we are unable to translate our quantitative findings in relation to the dichotomous criterion of the Brugada syndrome.

In summary, we present a large number of genome-wide significant loci robustly associated with cardiac repolarization parameters, some with compelling biological basis for their association and a number of loci not previously implicated in cardiac function. The identified loci and selected genes have the potential to aid future studies that are focused on risk stratification or on molecular mechanisms underlying cardiac repolarization and diseases of cardiac repolarization; to facilitate these studies we have made our results (including the genome-wide association) publically available.
**Phenotype Modeling**

Summing correlated variables when expecting that genetic variants are associated to multiple of these variables will increase the power for detection; with genetic variant (Snp), and traits (Amplitude), consider the correlation: \[ \text{Corr}(\text{Snp, Amplitude} \_1 + \text{Amplitude} \_2) = \text{Corr}(\text{Snp, Amplitude} \_1) + \text{Corr}(\text{Snp, Amplitude} \_2) + 2 \times \text{Corr}(\text{Amplitude}_1, \text{Amplitude}_2) > 0, \text{the traits Amplitude measure the same latent variable} (I). \] Therefore phenotypes for the ST segment amplitudes 80 ms after J-point and T wave amplitudes were defined by taking the sum of the lateral (I, aVL, V_s) and inferior (II, III and aVF), septal (V_1 and V_s), anterior (V_2, V_3) leads and lead AVR. Individuals were excluded for bundle branch block or QRS duration >120msec, atrial fibrillation, flutter, history of myocardial infarction or electronic pacemaker rhythm and when available, heart failure and ECG altering medication. Also participants with extreme measurements (more than ±4SD from mean) were excluded on a per phenotype basis.

**Statistical Analyses**

To control for multiple testing of the 10 phenotypes while accounting for the correlation between them, we performed an eigenvalue decomposition of the correlation matrix (Table S2) of the phenotypes. The variance of the eigenvalues (Var(\lambda_\text{obs}) = 2.36) was used to estimate the effective number of independent phenotypes tested \[ a = \frac{5 \times 10^{-9}}{8} = 6.25 \times 10^{-9} \text{ as threshold for declaring genome-wide significance in order to correct for the effective number of independent phenotypes studied.} \]

Residuals of ST-T wave amplitudes were calculated using general linear regression models to adjust for age, gender and body mass index, and standardized to a mean of zero and a standard deviation of one. GWAS analyses in Prevend and Lifelines of 2,316,136 genotyped or imputed SNPs (Table S16) were performed on the standardized residuals using an additive genetic model in PLINK (v1.07). Test statistics from each cohort were then corrected for their respective genomic control inflation factor to adjust for residual population sub-structure and meta-analyzed using the inverse-variance method. SNPs with MAF<1% (weighted average across cohorts) were removed.

For conditional analyses we repeated the primary association analysis for each trait whilst conditioning on the trait-specific genome wide significant sentinel SNPs by adding the SNP genotypes as covariates. Association results for each study were again combined by inverse variance weighting. Variants were considered to be independent if the pair-wise LD (r^2) was less than 0.1 and if they were separated by at least 1 MB; this was defined a ‘locus’.

Replication of the significant genotype-phenotype associations (P<6.25×10^{-9}) and associations that did not exceed this threshold, but were suggestive (6.25×10^{-9}<P<1×10^{-6}), was performed in the Rotterdam Study (I, II and III), ERF, ARIC, CHS and YFS and combined using fixed-effects meta-analysis by inverse variance weighting (Table S16). The pre-specified statistical significance threshold for heterogeneity by Cochran’s Q (Phet) was P<0.0007 to account for multiple testing. Inverse variance weighting was used to determine meta-P-values for combined discovery and replication data. An association was considered replicated if the direction of effect was concordant with discovery, replication P<0.01 and meta-P<6.25×10^{-9}.

Heritability estimates were calculated in the Erasmus Rucphen Family Study (ERF) using the “tdist” function in SOLAR software, including gender, age and bmi as covariates.

**Data-driven Expression Prioritized Integration for Complex Traits (DEPICT)**

DEPICT systematically identifies the most likely causal gene at a given associated locus, tests gene sets for enrichment in associated SNPs, and identifies tissues and cell types in which genes from associated loci are highly expressed (see Pers et al. for a detailed description). For this work we ran DEPICT on 140 independently associated loci (association P values < 10^{-6}; PLINK parameters, ‘--clump-pl 1e-5 --clump-kb 500 --clump-r2 0.05’) resulting in 103 independent, autosomal DEPICT loci containing 363 genes (loci overlapping with the major histocompatibility complex region are by default excluded in DEPICT). We have extended the locus definition used in DEPICT (LD r^2 > 0.5) with 100kb at either side of the loci, because several genes that may be important for cardiac repolarization were outside the default DEPICT locus boundaries (e.g. SCN5A, TMEM43, VGLL2). The gene set enrichment results for this slightly extended locus definition were similar to the results based on the default locus definition used in DEPICT (see Table S17).

**Identification of Candidate Genes**

We prioritized candidate genes based on nearby genes: We considered the nearest gene and any other gene located within 100kb of the sentinel SNP. Coding variants: For identification of coding variants and LD calculations we used the 1000 Genomes (1000G) Proj.
tags on the 5-end. The labeled probe was annealed with equal amount of unlabeled reverse strand primer using standard protocol. DNA-protein binding reaction was assayed by mixing 2 to 10μg nuclear extract with 0.2 pm annealed oligonucleotides in binding buffer (10 mM Tris (pH7.5), 50 mM KCl, 1 mM DTT), 2.5% glycerol and 75 ng/μL poly(dIdC) in a final volume of 15 μL. The mixture was then incubated for 30 min at room temperature, followed by polyacrylamide gel electrophoresis at 25°C on a 4.5% polyacrylamide gel. Fluorescence was visualized with the Odyssey infrared imaging system (Li-Cor Biosciences). Competition of labeled probe was carried out by adding 100x excess unlabeled probe.

**Prioritization of Potentially Causative Variants and Enrichment of DNA Elements**

For the prioritization of genetic variants for future functional follow-up and insight of the underlying DNA elements that might be relevant for causal ST-T wave amplitude variants, we employed the PAINTOR (Probabilistic Annotation INTEGRATOR) framework. In short, this method allows us to prioritize genetic variants in each of the 28 significant associated loci by integrating LD structure, strength of association and annotations of functional DNA elements to estimate the probability for each variant to be causal but also investigate which annotations are potentially causal.

**Imputation using 1000 Genomes**

Genome positions from the Prevend and Lifelines genotypes were converted from hg18 to hg19 using the UCSC Liftover tool. Genome-wide genotype imputation was performed with SHAPEIT (v2) and IMPUTE2 (v2.3.0) using the complete 1000 Genomes v3, March 2012 haplotypes Phase 1 integrated variant set release as reference panel.

**Functional Information**

We overlapped SNPs with data from the ENCODE project and Roadmap Epigenomics Program, conservation across mammals, various cardiac transcription factor measured by ChIP-Seq and contiguous annotations of human fetal heart, left ventricle and right ventricle as determined by ChromHMM (Supplementary Note).


### Table 1: Genome-wide Significant Genotype-Phenotype Associations Clustered in 28 Genetic Loci, After Replication

<table>
<thead>
<tr>
<th>#</th>
<th>Region</th>
<th>SNP</th>
<th>Trait</th>
<th>ld (r²) with Sentinel SNP (†) in Locus</th>
<th>Trait</th>
<th>EA (FRQ) / NEA</th>
<th>β (SE)</th>
<th>P₀&lt;sub&gt;min&lt;/sub&gt;</th>
<th>N</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1p36.33</td>
<td>rs260505</td>
<td>T.ant</td>
<td>†</td>
<td>A(0.49)/C</td>
<td>0.05(0.01)</td>
<td>1.3E-11</td>
<td>37775</td>
<td></td>
<td>SKP</td>
</tr>
<tr>
<td>2</td>
<td>1p36.12</td>
<td>rs2072944</td>
<td>T.sep</td>
<td>0.37</td>
<td>T(0.55)/C</td>
<td>0.05(0.01)</td>
<td>1.7E-13</td>
<td>37821</td>
<td></td>
<td>LUZP2*, KDM1A*, WNT4*</td>
</tr>
<tr>
<td>3</td>
<td>1p31.3</td>
<td>rs2207792</td>
<td>T.aVR</td>
<td>†</td>
<td>A(0.66)/G</td>
<td>-0.06(0.01)</td>
<td>3.3E-13</td>
<td>37999</td>
<td></td>
<td>NFI4A*</td>
</tr>
<tr>
<td>4</td>
<td>1p13.2</td>
<td>rs12145374</td>
<td>ST.ant</td>
<td>†</td>
<td>A(0.70)/C</td>
<td>0.11(0.01)</td>
<td>5.9E-33</td>
<td>37753</td>
<td></td>
<td>KCND3*, FAM212B*</td>
</tr>
<tr>
<td>5</td>
<td>1q24.2</td>
<td>rs545833</td>
<td>ST.aVR</td>
<td>†</td>
<td>T(0.26)/C</td>
<td>-0.06(0.01)</td>
<td>7.0E-18</td>
<td>37753</td>
<td></td>
<td>DPT*</td>
</tr>
<tr>
<td>6</td>
<td>1q24.2</td>
<td>rs511187</td>
<td>ST.inf</td>
<td>0.79</td>
<td>A(0.72/G)</td>
<td>0.06(0.01)</td>
<td>2.6E-12</td>
<td>37770</td>
<td></td>
<td>DPT*</td>
</tr>
<tr>
<td>7</td>
<td>1q24.2</td>
<td>rs511187</td>
<td>ST.avr</td>
<td>0.79</td>
<td>A(0.72)/G</td>
<td>-0.07(0.01)</td>
<td>3.1E-17</td>
<td>37784</td>
<td></td>
<td>DPT*</td>
</tr>
<tr>
<td>8</td>
<td>2p15</td>
<td>rs7576036</td>
<td>ST.sep</td>
<td>†</td>
<td>T(0.54)/C</td>
<td>0.05(0.01)</td>
<td>2.8E-10</td>
<td>37753</td>
<td></td>
<td>XPO1*</td>
</tr>
<tr>
<td>9</td>
<td>2q33.1</td>
<td>rs1866666</td>
<td>T.aVR</td>
<td>†</td>
<td>T(0.49)/C</td>
<td>-0.06(0.01)</td>
<td>1.8E-11</td>
<td>37799</td>
<td></td>
<td>PLC11*, MAR2*, RFTN2*, MOB4*</td>
</tr>
<tr>
<td>10</td>
<td>3p25.1</td>
<td>rs4684185</td>
<td>T.aVR</td>
<td>†</td>
<td>T(0.35)/C</td>
<td>-0.05(0.01)</td>
<td>7.1E-10</td>
<td>37799</td>
<td></td>
<td>LSM3*, TPM1A4*</td>
</tr>
<tr>
<td>11</td>
<td>3p22.2</td>
<td>rs7638909</td>
<td>ST.inf</td>
<td>0.02</td>
<td>T(0.76)/G</td>
<td>-0.06(0.01)</td>
<td>2.0E-11</td>
<td>37753</td>
<td></td>
<td>SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>12</td>
<td>3p22.2</td>
<td>rs9851710</td>
<td>T.inf</td>
<td>0.03</td>
<td>A(0.66/C)</td>
<td>0.08(0.01)</td>
<td>9.3E-19</td>
<td>37810</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>13</td>
<td>3p22.2</td>
<td>rs7635988</td>
<td>T.ant</td>
<td>0.02</td>
<td>A(0.71)/T</td>
<td>0.06(0.01)</td>
<td>3.8E-11</td>
<td>37775</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>14</td>
<td>3p22.2</td>
<td>rs6783110</td>
<td>T.lat</td>
<td>0.94</td>
<td>A(0.42/G)</td>
<td>0.06(0.01)</td>
<td>3.7E-12</td>
<td>30699</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>15</td>
<td>3p22.2</td>
<td>rs4076737</td>
<td>T.aVR</td>
<td>0.93</td>
<td>T(0.57)/G</td>
<td>0.07(0.01)</td>
<td>1.2E-20</td>
<td>37799</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>16</td>
<td>3p22.2</td>
<td>rs4076737</td>
<td>ST.inf</td>
<td>0.93</td>
<td>T(0.57)/G</td>
<td>-0.07(0.01)</td>
<td>8.3E-20</td>
<td>37770</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>17</td>
<td>3p22.2</td>
<td>rs6801957</td>
<td>ST.avr</td>
<td>†</td>
<td>T(0.43)/C</td>
<td>-0.08(0.01)</td>
<td>1.1E-25</td>
<td>37784</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>18</td>
<td>3p22.2</td>
<td>rs6801957</td>
<td>ST.lat</td>
<td>†</td>
<td>T(0.43)/C</td>
<td>0.06(0.01)</td>
<td>8.5E-16</td>
<td>37771</td>
<td></td>
<td>SCN10A*, SCN5A*, ACVR2B*</td>
</tr>
<tr>
<td>19</td>
<td>6q21.31</td>
<td>rs7756236</td>
<td>ST.ant</td>
<td>†</td>
<td>A(0.70/G)</td>
<td>-0.06(0.01)</td>
<td>1.3E-14</td>
<td>37753</td>
<td></td>
<td>CDKN1A*</td>
</tr>
<tr>
<td>20</td>
<td>6q22.2</td>
<td>rs210966</td>
<td>ST.lat</td>
<td>†</td>
<td>C(0.28/G)</td>
<td>0.07(0.01)</td>
<td>5.1E-16</td>
<td>37771</td>
<td></td>
<td>ROS1*, VGLL2*</td>
</tr>
<tr>
<td>21</td>
<td>6q22.2</td>
<td>rs210966</td>
<td>ST.avr</td>
<td>†</td>
<td>C(0.28)/G</td>
<td>-0.07(0.01)</td>
<td>4.0E-18</td>
<td>37784</td>
<td></td>
<td>ROS1*, VGLL2*</td>
</tr>
<tr>
<td>22</td>
<td>6q22.2</td>
<td>rs172409</td>
<td>ST.inf</td>
<td>0.41</td>
<td>A(0.54/C)</td>
<td>-0.05(0.01)</td>
<td>6.4E-11</td>
<td>30708</td>
<td></td>
<td>ROS1*, VGLL2*</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Gene Name</td>
<td>Transcript</td>
<td>Effect</td>
<td>Strain</td>
<td>p-Value</td>
<td>Log10(p-value)</td>
<td>Fold Change</td>
<td>p-value</td>
<td>Log10(p-value)</td>
<td>Fold Change</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>---------------</td>
<td>-------------</td>
<td>--------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>14</td>
<td>6q23.31</td>
<td>rs9388451</td>
<td>ST.ant</td>
<td>†</td>
<td>T(0.52)/C</td>
<td>-0.06(0.01)</td>
<td>2.00E-18</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>8p23.1</td>
<td>rs1458942</td>
<td>ST.aVR</td>
<td>†</td>
<td>A(0.60)/G</td>
<td>0.06(0.01)</td>
<td>1.10E-13</td>
<td>37775</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>8p23.1</td>
<td>rs7011924</td>
<td>ST.inf</td>
<td>†</td>
<td>A(0.40)/G</td>
<td>-0.07(0.01)</td>
<td>3.90E-19</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12p13.32</td>
<td>rs793399</td>
<td>ST.sep</td>
<td>0.95</td>
<td>A(0.61)/G</td>
<td>-0.04(0.01)</td>
<td>2.10E-18</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>12p12.1</td>
<td>rs10842350</td>
<td>T.lat</td>
<td>-</td>
<td>A(0.54)/G</td>
<td>0.06(0.01)</td>
<td>2.40E-16</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>12p13.32</td>
<td>rs10842350</td>
<td>T.avr</td>
<td>†</td>
<td>A(0.54)/G</td>
<td>0.09(0.01)</td>
<td>1.30E-18</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>13q22.1</td>
<td>rs728926</td>
<td>T.aVR</td>
<td>-</td>
<td>T(0.38)/C</td>
<td>0.06(0.01)</td>
<td>1.50E-16</td>
<td>37799</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>15q26.3</td>
<td>rs7174918</td>
<td>ST.ant</td>
<td>0.74</td>
<td>T(0.29)/C</td>
<td>-0.06(0.01)</td>
<td>6.90E-16</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>16p13.3</td>
<td>rs712150</td>
<td>T.aVR</td>
<td>†</td>
<td>A(0.40)/G</td>
<td>0.05(0.01)</td>
<td>4.30E-13</td>
<td>37799</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>16p13.3</td>
<td>rs719330</td>
<td>T.aVR</td>
<td>0.95</td>
<td>T(0.46)/C</td>
<td>-0.08(0.01)</td>
<td>3.10E-16</td>
<td>30731</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>16q21</td>
<td>rs8057901</td>
<td>T.avr</td>
<td>0.93</td>
<td>A(0.63)/G</td>
<td>-0.06(0.01)</td>
<td>4.90E-11</td>
<td>30738</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>16q21</td>
<td>rs8057901</td>
<td>ST.inf</td>
<td>0.93</td>
<td>A(0.63)/G</td>
<td>-0.08(0.01)</td>
<td>1.60E-21</td>
<td>30713</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>16q21</td>
<td>rs8057901</td>
<td>ST.avr</td>
<td>†</td>
<td>A(0.63)/G</td>
<td>0.07(0.01)</td>
<td>3.00E-18</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>16q21</td>
<td>rs8083566</td>
<td>ST.lat</td>
<td>†</td>
<td>A(0.07)/C</td>
<td>-0.09(0.01)</td>
<td>4.30E-10</td>
<td>37771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>16q21</td>
<td>rs8083566</td>
<td>ST.inf</td>
<td>-</td>
<td>A(0.07)/C</td>
<td>-0.11(0.02)</td>
<td>1.70E-12</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>16q21</td>
<td>rs8083566</td>
<td>ST.aVR</td>
<td>†</td>
<td>A(0.07)/C</td>
<td>0.12(0.01)</td>
<td>3.60E-15</td>
<td>37784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>17q11.33</td>
<td>rs11673003</td>
<td>ST.lat</td>
<td>†</td>
<td>A(0.90)/G</td>
<td>-0.09(0.01)</td>
<td>5.10E-14</td>
<td>37771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>20q11.22</td>
<td>rs6087666</td>
<td>ST.inf</td>
<td>†</td>
<td>A(0.20)/G</td>
<td>-0.07(0.01)</td>
<td>3.10E-15</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>20q11.22</td>
<td>rs6087666</td>
<td>ST.aVR</td>
<td>†</td>
<td>A(0.20)/G</td>
<td>-0.07(0.01)</td>
<td>3.10E-15</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>20q11.22</td>
<td>rs6087666</td>
<td>ST.aVR</td>
<td>0.55</td>
<td>A(0.21)/G</td>
<td>0.06(0.01)</td>
<td>5.90E-13</td>
<td>37784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>21q13.13</td>
<td>rs11907908</td>
<td>ST.ant</td>
<td>0.83</td>
<td>T(0.07)/C</td>
<td>-0.12(0.01)</td>
<td>8.00E-18</td>
<td>37753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>21q13.13</td>
<td>rs6019750</td>
<td>ST.lat</td>
<td>1</td>
<td>C(0.93)/G</td>
<td>0.13(0.01)</td>
<td>4.30E-23</td>
<td>37771</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>21q13.13</td>
<td>rs6019750</td>
<td>ST.inf</td>
<td>0.55</td>
<td>A(0.21)/G</td>
<td>0.06(0.01)</td>
<td>5.90E-13</td>
<td>37784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>21q13.13</td>
<td>rs6019750</td>
<td>ST.avr</td>
<td>†</td>
<td>A(0.21)/G</td>
<td>-0.07(0.01)</td>
<td>3.10E-15</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>21q13.13</td>
<td>rs6019750</td>
<td>ST.aVR</td>
<td>†</td>
<td>A(0.21)/G</td>
<td>-0.07(0.01)</td>
<td>3.10E-15</td>
<td>37770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>Position</td>
<td>Gene</td>
<td>Sample Type</td>
<td>A/G</td>
<td>p-Value</td>
<td>EAF</td>
<td>rsID</td>
<td>Significant Gene(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>------</td>
<td>-------------</td>
<td>-----</td>
<td>---------</td>
<td>-----</td>
<td>------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>20q13.13</td>
<td>ST.inf</td>
<td>-</td>
<td>A(0.93)/G</td>
<td>0.12(0.01)</td>
<td>4.40E-18</td>
<td>37770</td>
<td>KCNB1, STAU1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>20q13.13</td>
<td>ST.aVR</td>
<td>↑</td>
<td>A(0.93)/G</td>
<td>-0.14(0.01)</td>
<td>5.60E-26</td>
<td>37784</td>
<td>KCNB1, STAU1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>