nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0499-3

In the format provided by the authors and unedited.

Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits

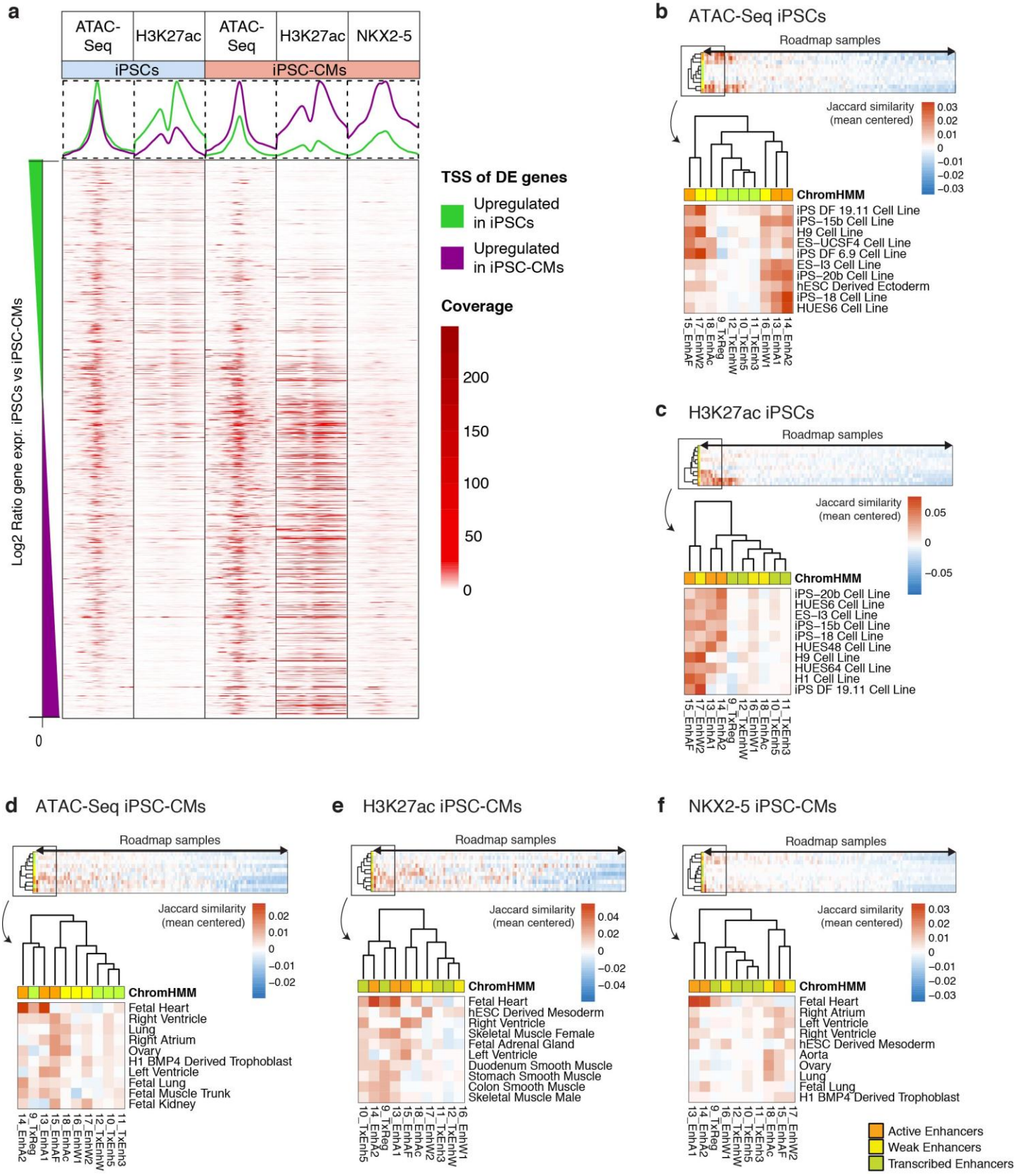
Paola Benaglio[1], Agnieszka D'Antonio-Chronowska[2,9], Wubin Ma[3,9], Feng Yang[3], William W. Young Greenwald[4], Margaret K. R. Donovan[4,5], Christopher DeBoever[4], He Li[2], Frauke Drees[2], Sanghamitra Singhal[1], Hiroko Matsui[2], Jessica van Setten[6], Nona Sotoodehnia[7,8], Kyle J. Gaulton[1], Erin N. Smith[1], Matteo D'Antonio[2], Michael G. Rosenfeld[3]* and Kelly A. Frazer[1,2]*

[1]Department of Pediatrics, Rady Children's Hospital, Division of Genome Information Sciences, University of California, San Diego, La Jolla, CA, USA. [2]Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA. [3]Howard Hughes Medical Institute, Department of Medicine, University of California, San Diego, La Jolla, CA, USA. [4]Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, CA, USA. [5]Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA. [6]Department of Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands. [7]Department of Medicine, Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA, USA. [8]Department of Epidemiology, Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA, USA. [9]These authors contributed equally: Agnieszka D'Antonio-Chronowska, Wubin Ma. *e-mail: mrosenfeld@ucsd.edu; kafrazer@ucsd.edu

Supplementary Figure 1

Characterization of iPSC-CMs by flow cytometry, immunofluorescence and gene expression.

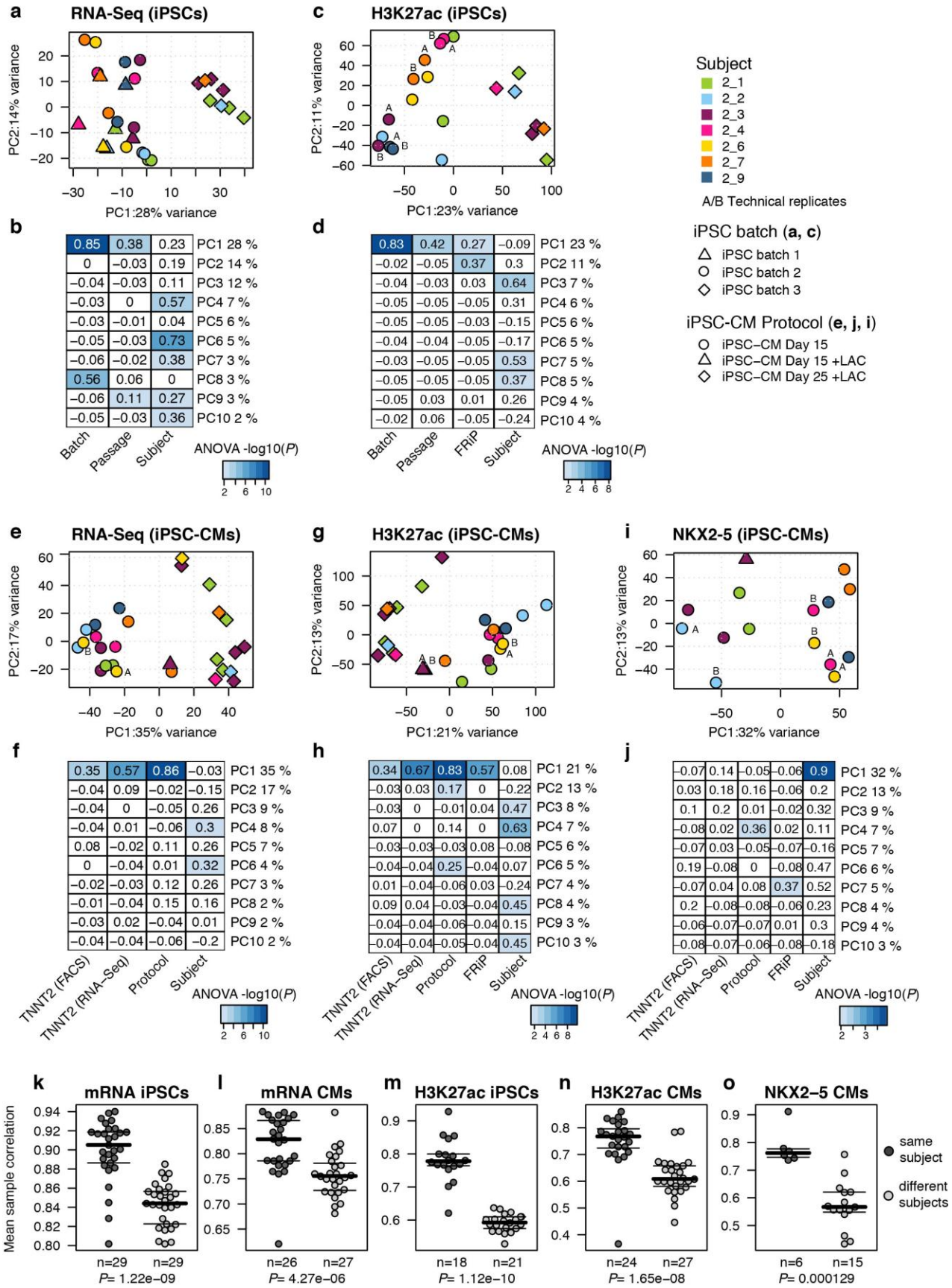
(**a**) Percentage of cells positive for the cardiomyocyte-specific marker TNNT2 analyzed by flow cytometry on iPSC-CMs generated in this study and collected at either day 15 (n = 15 independent samples) or day 25 and purified with L-lactate selection (n = 12 independent samples). The day 15 sample that was lactate purified is indicated. Plot lines indicate median, lower and upper quartiles. (**b**) Confocal microscopy of immunofluorescence for TNNT2 and MYL7 in one replicate of iPSCORE_2_1 day 15 iPSC-CMs at two different magnifications. Similar staining was observed for other two day15-iPSC-CM samples (subjects iPSCORE_2_3 and iPSCORE_2_9, not shown). (**c**) Hierarchical clustering and heatmap of RNA-seq expression data from 61 selected cell-type specific genes. In addition to the 56 samples from this study, Roadmap RNA-seq data from stem cell lines (H1, HUES64, iPS-20b and iPS-18) and human tissues (right ventricle, left ventricle, right atrium and fetal heart) are included as reference samples. Gene expression values are reported as *vst*-normalized read counts.



Supplementary Figure 2

Epigenetic profile differences between iPSCs and iPSC-CMs.

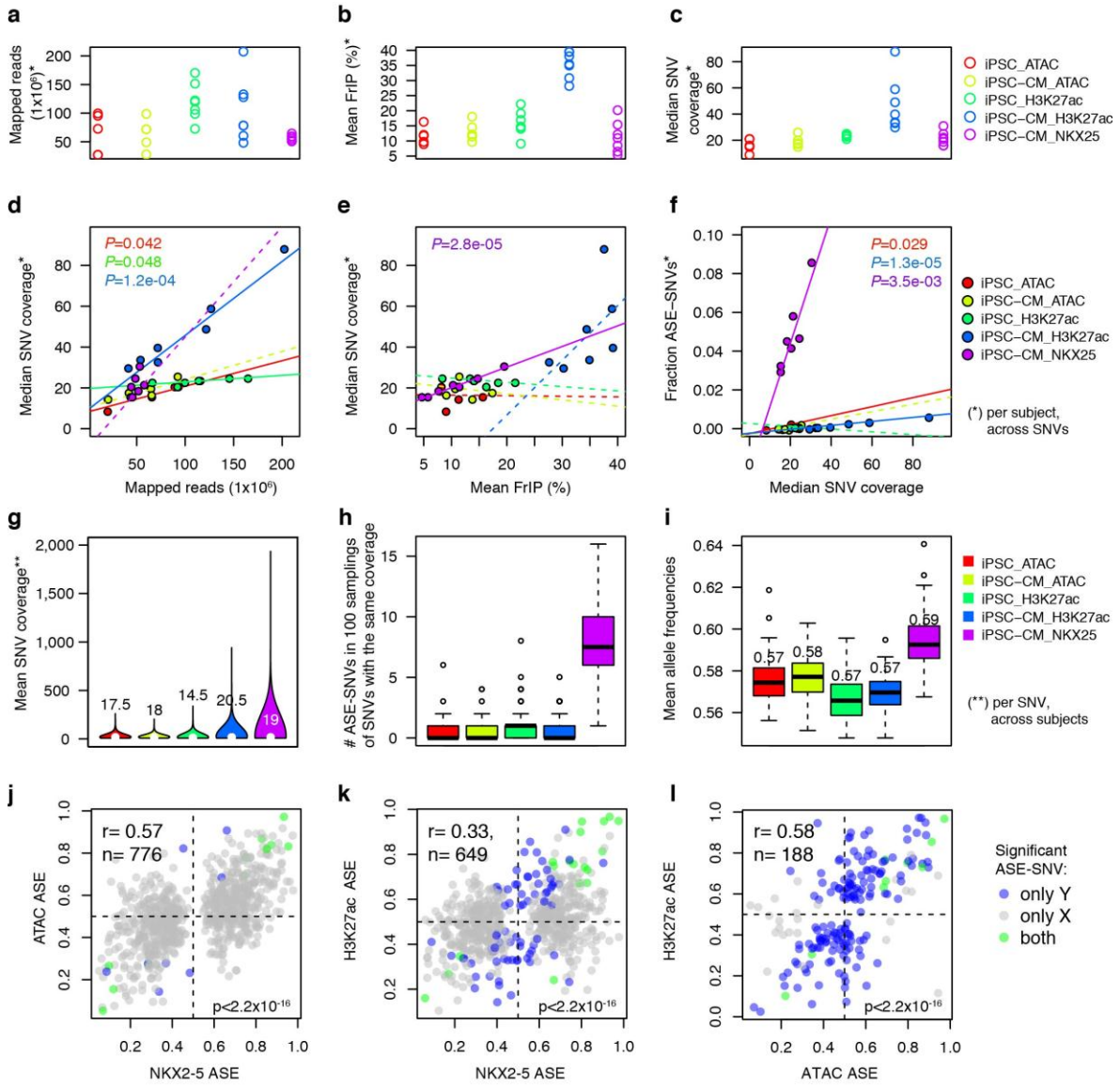
(**a**) Tag heatmap of sequence coverage at the TSSs of differentially expressed genes (*n* = 5,307 DEGs, minimum $\log_2$ fold change = 2, FDR < 0.05, DESeq2), listed in decreasing order of $\log_2$ ratio of expression in iPSCs versus iPSC-CMs. The density plot at the top shows the average normalized coverage (range 0-1) across iPSC-upregulated genes (2,444 genes, green) and iPSC-CM-upregulated genes (2,863 genes, purple). For a 2-kb window centered at the TSS, coverage values for 25-bp bins were obtained by combining all samples of each data type and normalized to a total of 10^7 reads. (**b-f**) Heatmap and hierarchical clustering of similarity based on overlap between enhancer annotations (from 25-state ChromHMM) for 127 samples from Roadmap Epigenomics (Ernst, J. and M. Kellis, *Nat. Methods* **9**, 215-216, 2012; Kundaje, A. *et al. Nature* **518**, 317-330, 2015) and ATAC-seq peaks or ChIP-Seq peaks from iPSCs (**b, c**) or iPSC-CMs (**d-f**) combined samples. Similarity was calculated using the *Jaccard* statistic (intersection/union of base pairs in each comparison), mean-centered across the 127 tissues and ordered by highest average enhancer similarity. A zoom in of the top 10 Roadmap tissues with highest average enhancer similarity is shown.


a
RNA-Seq (iPSCs)
PC2:14% variance
PC1:28% variance
c
H3K27ac (iPSCs)
PC2:11% variance
PC1:23% variance
Subject
2_1
2_2
2_3
2_4
2_6
2_7
2_9
A/B Technical replicates
iPSC batch (a, c)
iPSC batch 1
iPSC batch 2
iPSC batch 3
iPSC-CM Protocol (e, j, i)
iPSC–CM Day 15
iPSC–CM Day 15 +LAC
iPSC–CM Day 25 +LAC
b
0.85 0.38 0.23 PC1 28 %
0 −0.03 0.19 PC2 14 %
−0.04 −0.03 0.11 PC3 12 %
−0.03 0 0.57 PC4 7 %
−0.03 −0.01 0.04 PC5 6 %
−0.05 −0.03 0.73 PC6 5 %
−0.06 −0.02 0.38 PC7 3 %
0.56 0.06 0 PC8 3 %
−0.06 0.11 0.27 PC9 3 %
−0.05 −0.03 0.36 PC10 2 %
Batch
Passage
Subject
ANOVA -log10(P)
d
0.83 0.42 0.27 −0.09 PC1 23 %
−0.02 −0.05 0.37 0.3 PC2 11 %
−0.04 −0.03 0.03 0.64 PC3 7 %
−0.05 −0.05 −0.05 0.31 PC4 6 %
−0.05 −0.05 −0.03 −0.15 PC5 6 %
−0.04 −0.04 −0.05 −0.17 PC6 5 %
−0.03 −0.05 −0.05 0.53 PC7 5 %
−0.05 −0.05 −0.05 0.37 PC8 5 %
−0.05 0.03 0.01 0.26 PC9 4 %
−0.02 0.06 −0.05 −0.24 PC10 4 %
Batch
Passage
FRiP
Subject
ANOVA -log10(P)
e
RNA-Seq (iPSC-CMs)
PC2:17% variance
PC1:35% variance
g
H3K27ac (iPSC-CMs)
PC2:13% variance
PC1:21% variance
i
NKX2-5 (iPSC-CMs)
PC2:13% variance
PC1:32% variance
f
0.35 0.57 0.86 −0.03 PC1 35 %
−0.04 0.09 −0.02 −0.15 PC2 17 %
−0.04 0 −0.05 0.26 PC3 9 %
−0.04 0.01 −0.06 0.3 PC4 8 %
0.08 −0.02 0.11 0.26 PC5 7 %
0 −0.04 0.01 0.32 PC6 4 %
−0.02 −0.03 0.12 0.26 PC7 3 %
−0.01 −0.04 0.15 0.16 PC8 2 %
−0.03 0.02 −0.04 0.01 PC9 2 %
−0.04 −0.04 −0.06 −0.2 PC10 2 %
TNNT2 (FACS)
TNNT2 (RNA−Seq)
Protocol
Subject
ANOVA -log10(P)
h
0.34 0.67 0.83 0.57 0.08 PC1 21 %
−0.03 0.03 0.17 0 −0.22 PC2 13 %
−0.03 0 −0.01 0.04 0.47 PC3 8 %
0.07 0 0.14 0 0.63 PC4 7 %
−0.03 −0.03 −0.03 0.08 −0.08 PC5 6 %
−0.04 −0.04 0.25 −0.04 0.07 PC6 5 %
0.01 −0.04 −0.06 0.03 −0.24 PC7 4 %
0.09 0.04 −0.03 −0.04 0.45 PC8 4 %
−0.04 −0.04 −0.06 −0.04 0.15 PC9 3 %
−0.04 −0.04 −0.05 −0.04 0.45 PC10 3 %
TNNT2 (FACS)
TNNT2 (RNA−Seq)
Protocol
FRiP
Subject
ANOVA -log10(P)
j
−0.07 0.14 −0.05 −0.06 0.9 PC1 32 %
0.03 0.18 0.16 −0.06 0.2 PC2 13 %
0.1 0.2 0.01 −0.02 0.32 PC3 9 %
−0.08 0.02 0.36 0.02 0.11 PC4 7 %
−0.07 0.03 −0.05 −0.07 −0.16 PC5 7 %
0.19 −0.08 0 −0.08 0.47 PC6 6 %
−0.07 0.04 0.08 0.37 0.52 PC7 5 %
0.2 −0.08 −0.08 −0.06 0.23 PC8 4 %
−0.06 −0.07 −0.07 0.01 0.3 PC9 4 %
−0.08 −0.07 −0.06 −0.08 −0.18 PC10 3 %
TNNT2 (FACS)
TNNT2 (RNA−Seq)
Protocol
FRiP
Subject
ANOVA -log10(P)
k
mRNA iPSCs
Mean sample correlation
n=29 n=29
P= 1.22e−09
l
mRNA CMs
n=26 n=27
P= 4.27e−06
m
H3K27ac iPSCs
n=18 n=21
P= 1.12e−10
n
H3K27ac CMs
n=24 n=27
P= 1.65e−08
o
NKX2−5 CMs
n=6 n=15
P= 0.000129
same subject
different subjects

Supplementary Figure 3

Analysis of variation of molecular phenotypes across samples.

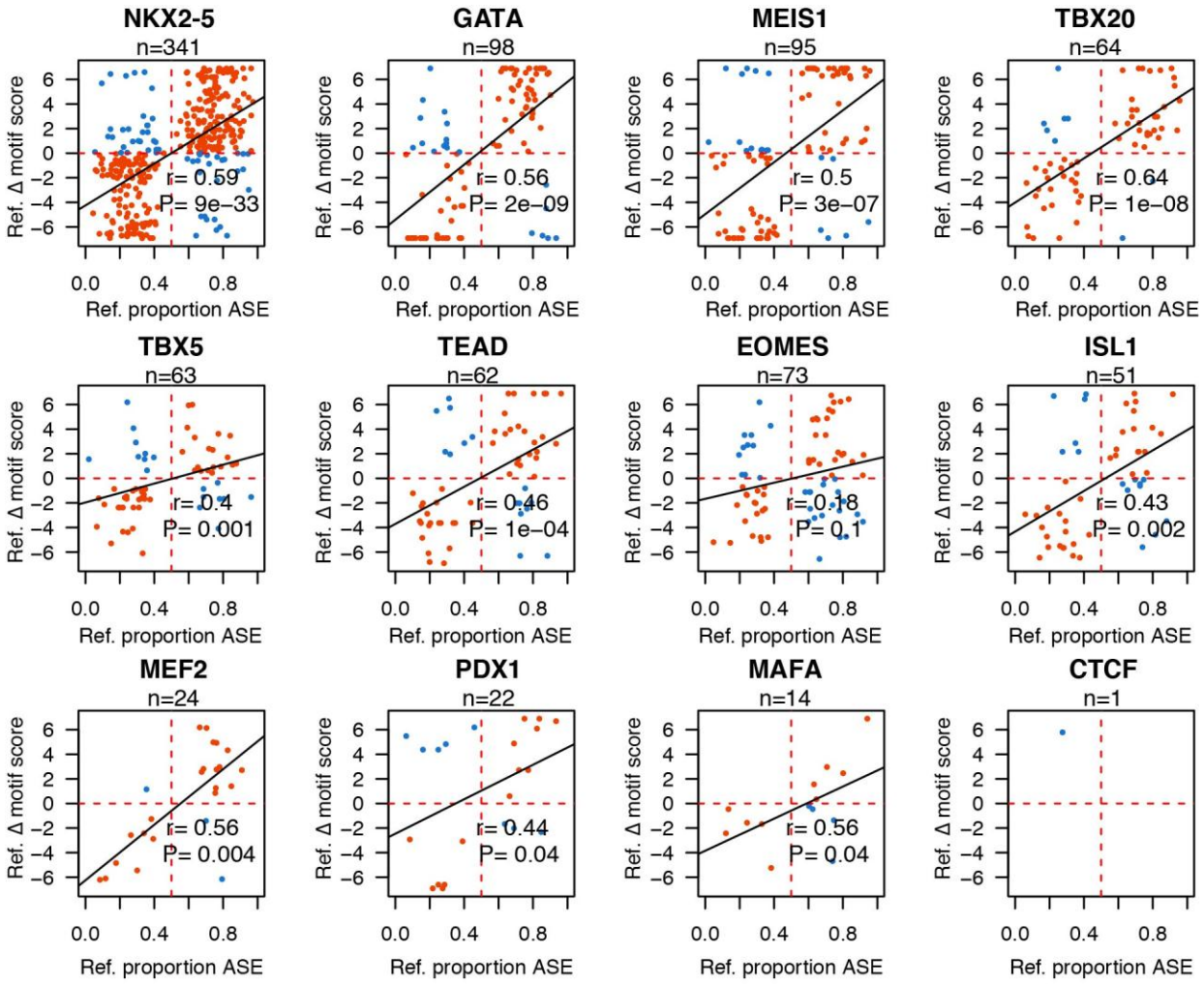
(**a, c, e, g, i**) Plots of the two first principal components (PC) calculated on all genes/peaks identified in each RNA-seq (**a**, iPSCs: 29 independent samples; **e**, iPSC-CMs: 26 independent samples and 1 technical replicate) or ChIP-seq dataset (**c**, H3K27ac in iPSCs: 17 independent samples and 4 technical replicates; **g**, H3K27ac in iPSC-CMs: 25 independent samples and 2 technical replicates; **i**, NKX2-5: 12 independent samples and 3 technical replicates). Samples are color-coded by subject and distinguished by a different symbol representing a different batch for iPSC (cultured, collected and sequenced at different times) or a different protocol (day 15 vs. day 25) for iPSC-CMs differentiation. Samples from the same individual are considered independent if iPSC lines were cultured -or iPSC-CMs were differentiated- at different times, and are considered technical replicates (indicated by A and B) if the assays were performed on cell material collected from the same sample. (**b, d, f, h, j**) Tables showing adjusted r-squared values from ANOVA tests are reported as a measure of association between PC1 to PC10 and different covariates in each dataset. Tables are color-coded according to $-\log_{10}$ of ANOVA *P*-values. (**k-o**) Plots of the average Spearman correlation coefficients between pairs of samples across the 1,000 most variable genes or peaks for the indicated molecular phenotypes. Each dot corresponds to the per-sample pairwise correlation coefficient averaged across samples from either the same subject or different subjects. Technical replicates were excluded for the comparisons between samples of the same subject. Plot lines indicate median, lower and upper quartiles. *P*-values from one-tailed Mann-Whitney test are shown.



Supplementary Figure 4

Comparison of number and effects of ASE-SNVs identified in ChIP-seq and ATAC-seq.

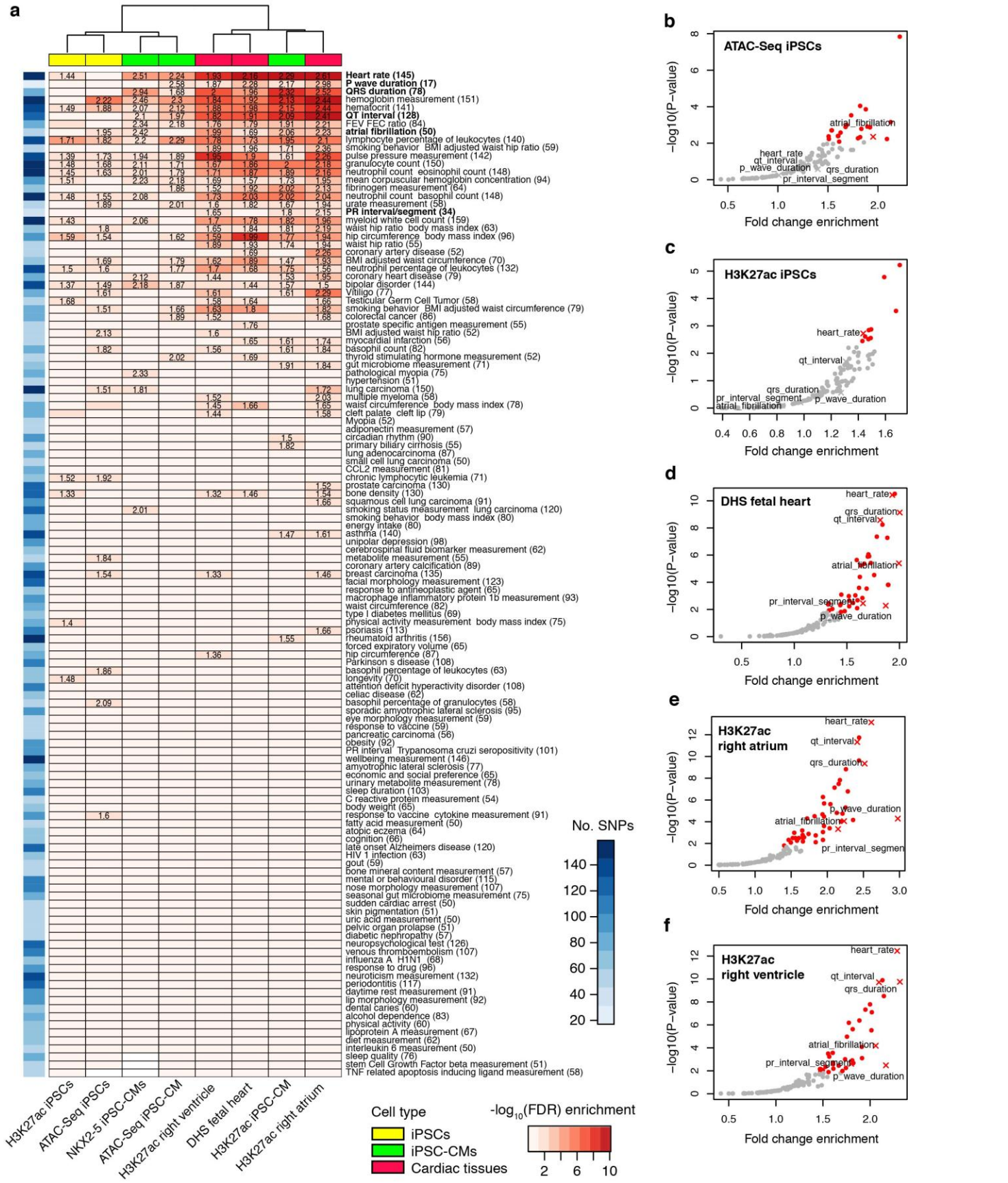
(**a**) Number of uniquely mapped reads for each data type and subject (n = 7). Each open circle corresponds to merged reads from different samples of the same subject. (**b**) Average FrIP across different samples of the same subject in each data type. (**c**) Median number of reads at all heterozygous SNVs tested for ASE in each individual. (**d,e**) Scatterplot showing increase in the median SNV coverage as a function of the number of mapped reads (**d**) and of FrIP (**e**) in each subject (n = 7). Continuous lines indicate that the relation is significant (linear regression, $P < 0.05$). (**f**) Scatterplot showing increase in the fraction of identified ASE-SNV in each subject (FDR < 0.05), as a function of the median SNV coverage. Continuous lines indicate that the relation is significant (linear regression, $P < 0.05$). (**g**) Distribution of the mean SNV coverage between subjects, across all SNVs analyzed in each data type (number of SNVs in each distribution from left to right: 30,463, 26,201, 116,898, 123,151 and 19,371). Median values are shown as white dots within the violin plot and are indicated. (**h**) Distribution of the number of ASE-SNVs (FDR < 10%) in 100 samplings of 100 SNVs with the same coverage in the different data types. (**i**) Distribution of the mean ASE effect sizes (allele frequencies) of the 100 SNVs samples shown in **h**. Median values are shown. Boxplot elements: median (thick line), lower and upper quartiles (box), maximum and minimum (wiskers). (**j-l**) Scatterplot showing correlation of ASE effects of the same SNV between peaks from different data types in iPSC-CMs: (**j**) NKX2-5 vs. ATAC-seq peaks; (**k**) NKX2-5 vs. H3K27ac peaks; (**l**) ATAC-seq vs. H3K27ac peaks. The union of significant ASE-SNVs in each pair of datasets is shown, with ASE effects expressed as the proportion of the reference allele at heterozygous SNVs. Gray dots denote ASE-SNVs significant (binomial test, FDR < 0.05) only in the peaks indicated in the x-axis, blue dots in the y-axis, and green dots in both. Pearson correlation coefficient (r), number of SNVs (n) and P-values are indicated.



Supplementary Figure 5

Correlation between ASE and motif disruption in NKX2-5 ChIP-seq.

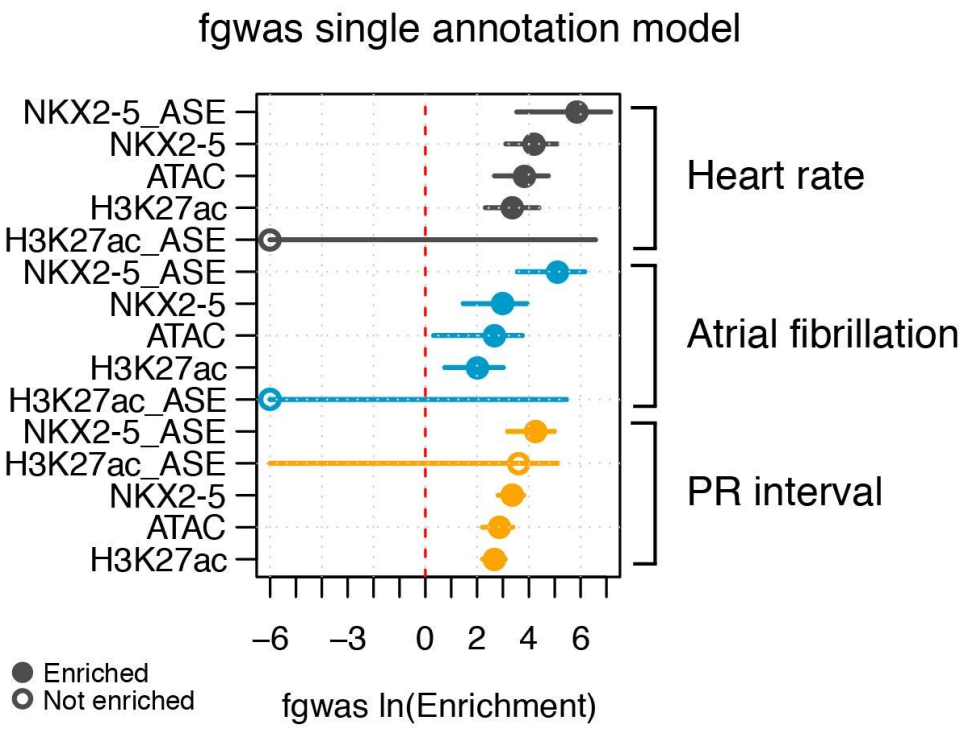
Scatterplots showing relationship between the proportion of reads for the reference allele at ASE-SNVs in the NKX2-5 ChIP-seq data and the difference in motif strength between the reference and alternate allele. Spearman correlation statistics (*r* and *P*-value) and the number of motif-altering ASE-SNVs (*n*) are indicated. The 12 most enriched families of motifs in NKX2-5 peaks (**Supplementary Table 4**) were tested. TFBS motifs that were strengthened (red) or weakened (blue) by the preferred allele of ASE-SNVs are indicated.


a
b
c
d
e
f
Heart rate (145)
P wave duration (17)
QRS duration (78)
hemoglobin measurement (151)
hematocrit (141)
QT interval (128)
FEV FEC ratio (84)
atrial fibrillation (50)
lymphocyte percentage of leukocytes (140)
smoking behavior BMI adjusted waist hip ratio (59)
pulse pressure measurement (142)
granulocyte count (150)
neutrophil count eosinophil count (148)
mean corpuscular hemoglobin concentration (94)
fibrinogen measurement (64)
neutrophil count basophil count (148)
urate measurement (58)
PR interval/segment (34)
myeloid white cell count (159)
waist hip ratio body mass index (63)
hip circumference body mass index (96)
waist hip ratio (55)
coronary artery disease (52)
BMI adjusted waist circumference (70)
neutrophil percentage of leukocytes (132)
coronary heart disease (79)
bipolar disorder (144)
Vitiligo (77)
Testicular Germ Cell Tumor (58)
smoking behavior BMI adjusted waist circumference (79)
colorectal cancer (86)
prostate specific antigen measurement (55)
BMI adjusted waist hip ratio (52)
myocardial infarction (56)
basophil count (82)
thyroid stimulating hormone measurement (52)
gut microbiome measurement (71)
pathological myopia (75)
hypertension (51)
lung carcinoma (150)
multiple myeloma (58)
waist circumference body mass index (78)
cleft palate cleft lip (79)
Myopia (52)
adiponectin measurement (57)
circadian rhythm (90)
primary biliary cirrhosis (55)
lung adenocarcinoma (87)
small cell lung carcinoma (50)
CCL2 measurement (81)
chronic lymphocytic leukemia (71)
prostate carcinoma (130)
bone density (130)
squamous cell lung carcinoma (91)
smoking status measurement lung carcinoma (120)
smoking behavior body mass index (80)
energy intake (80)
asthma (140)
unipolar depression (98)
cerebrospinal fluid biomarker measurement (62)
metabolite measurement (55)
coronary artery calcification (89)
breast carcinoma (135)
facial morphology measurement (123)
response to antineoplastic agent (65)
macrophage inflammatory protein 1b measurement (93)
waist circumference (82)
type I diabetes mellitus (69)
physical activity measurement body mass index (75)
psoriasis (113)
rheumatoid arthritis (156)
forced expiratory volume (65)
hip circumference (67)
Parkinson s disease (108)
basophil percentage of leukocytes (63)
longevity (70)
attention deficit hyperactivity disorder (108)
celiac disease (62)
basophil percentage of granulocytes (58)
sporadic amyotrophic lateral sclerosis (95)
eye morphology measurement (59)
response to vaccine (59)
pancreatic carcinoma (56)
obesity (92)
PR interval Trypanosoma cruzi seropositivity (101)
wellbeing measurement (146)
amyotrophic lateral sclerosis (77)
economic and social preference (65)
urinary metabolite measurement (78)
sleep duration (103)
C reactive protein measurement (54)
body weight (65)
response to vaccine cytokine measurement (91)
fatty acid measurement (50)
atopic eczema (64)
cognition (66)
late onset Alzheimers disease (120)
HIV 1 infection (63)
gout (59)
bone mineral content measurement (57)
mental or behavioural disorder (115)
nose morphology measurement (107)
seasonal gut microbiome measurement (75)
sudden cardiac arrest (50)
skin pigmentation (51)
uric acid measurement (50)
pelvic organ prolapse (51)
diabetic nephropathy (57)
neuropsychological test (126)
venous thromboembolism (107)
influenza A H1N1 (68)
response to drug (96)
neuroticism measurement (132)
periodontitis (117)
daytime rest measurement (91)
lip morphology measurement (92)
dental caries (60)
alcohol dependence (83)
physical activity (60)
lipoprotein A measurement (67)
diet measurement (62)
interleukin 6 measurement (50)
sleep quality (76)
stem Cell Growth Factor beta measurement (51)
TNF related apoptosis inducing ligand measurement (58)
H3K27ac iPSCs
ATAC-Seq iPSCs
NKX2-5 iPSC-CMs
ATAC-Seq iPSC-CM
H3K27ac right ventricle
DHS fetal heart
H3K27ac iPSC-CM
H3K27ac right atrium
Cell type
iPSCs
iPSC-CMs
Cardiac tissues
No. SNPs
140
120
100
80
60
40
20
-log10(FDR) enrichment
2
6
10
ATAC-Seq iPSCs
−log10(P−value)
Fold change enrichment
atrial_fibrillation
heart_rate
qt_interval
p_wave_duration
qrs_duration
pr_interval_segment
H3K27ac iPSCs
DHS fetal heart
H3K27ac right atrium
H3K27ac right ventricle

Supplementary Figure 6

Analysis of enrichment of ChIP-seq and ATAC-seq peaks for GWAS SNPs.

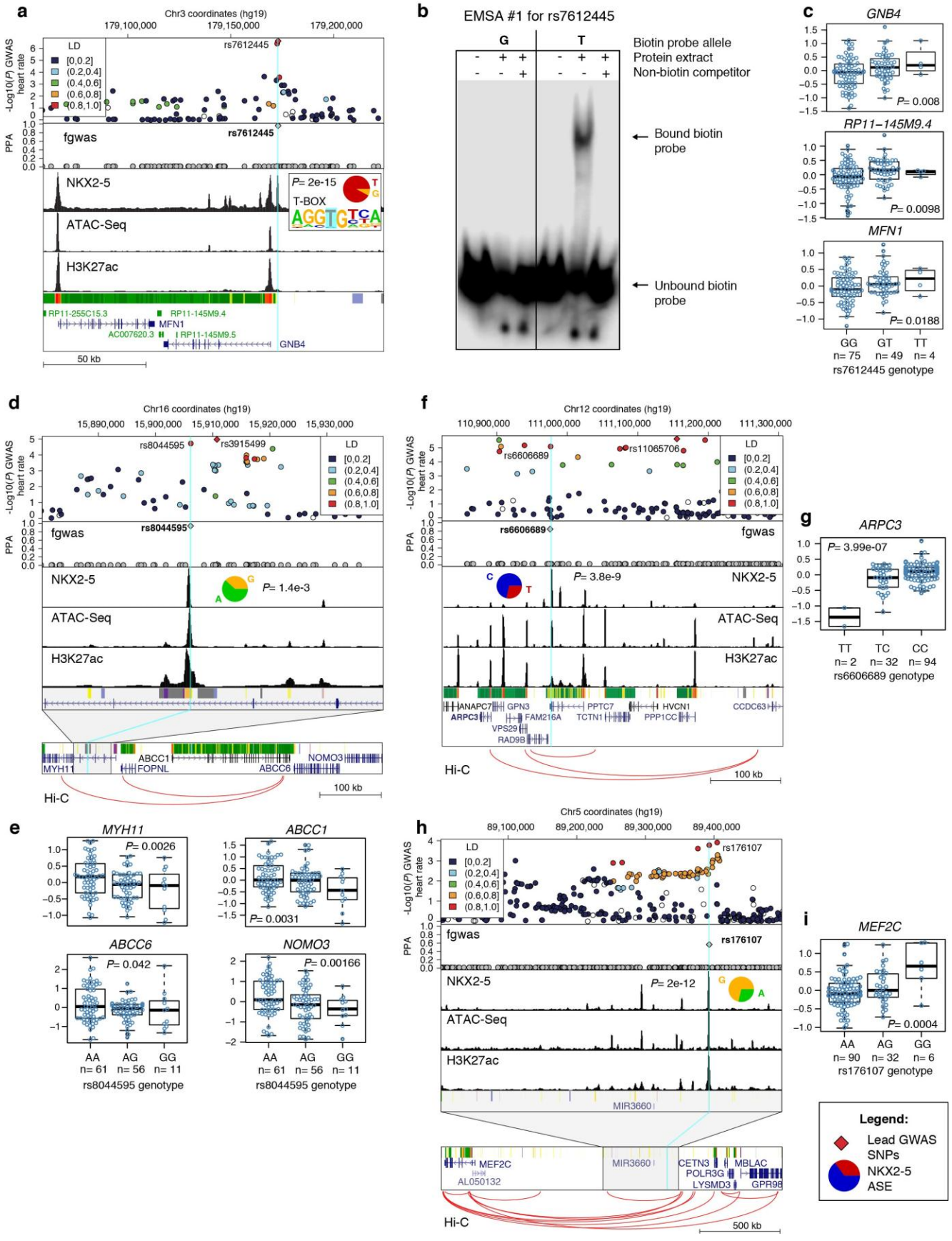
(**a**) Heatmap of enrichment for GWAS SNPs in ChIP-seq and ATAC-seq peaks from iPSCs and iPSC-CMs combined samples from this study as well as in peaks from cardiac tissues from Roadmap (DHS of fetal heart, H3K27ac of right ventricle and right atrium). Heatmap is ordered by the most enriched GWAS traits on average in the cardiac datasets (iPSC-CMs and Roadmap tissues) and shows fold change values for significant enrichment at FDR corrected *P*-value < 0.05. A total of 125 GWAS traits were tested for enrichment, and the corresponding number of independent SNPs is given in parenthesis. The statistical enrichment test was performed using GREGOR software. (**b-f**) Volcano plots showing -$\log_{10}$*P*-values (*y*-axis) and fold enrichment (*x*-axis) for GWAS loci showed in **a,** indicating the position of the 6 electrocardiographic traits. Red symbols indicate significant enrichment at FDR corrected *P*-value < 0.05. The iPSC-CMs NKX2-5, H3K27ac and ATAC-seq enrichment plots are shown in **Figure 5a-c**.



Supplementary Figure 7

Enrichment of GWAS signals within iPSC-CM functional annotations using fgwas single state models.

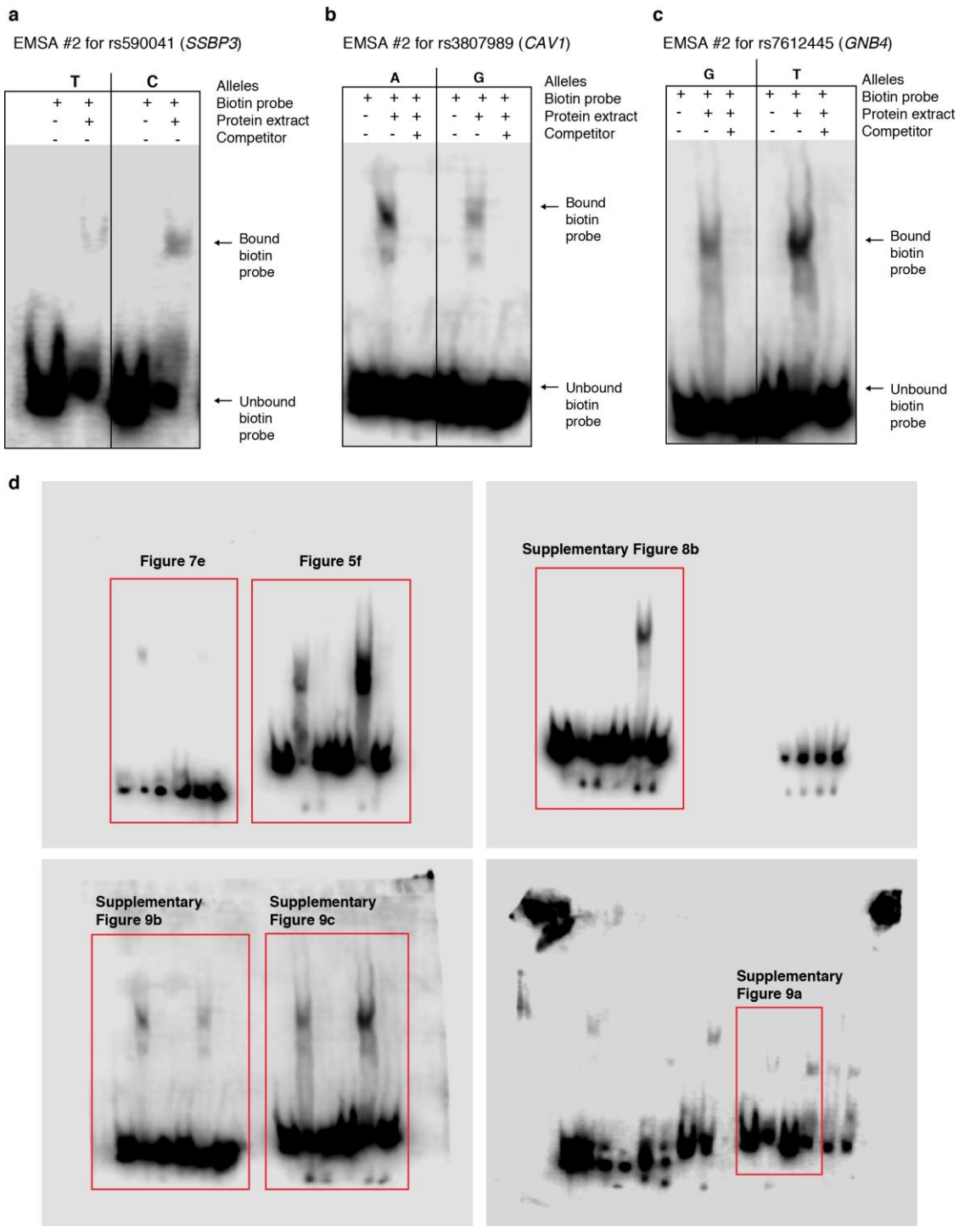
Fgwas natural log fold enrichment of iPSC-CM genomic annotations (*y*-axis) in heart rate (den Hoed, M. *et al. Nat. Genet.* **45**, 621-631, 2013), atrial fibrillation (Christophersen, I. E. *et al. Nat. Genet.* **49**, 946-952, 2017), and PR interval (van Setten, J. *et al. Nat. Commun.* **9**, 2904, 2018) GWAS signals. Solid circles indicate significant enrichment (defined as 95% CI above zero) and the bars indicate the 95% confidence intervals. The genomic annotations include NKX2-5 ASE-SNVs, NKX2-5 peaks, ATAC-seq peaks, H3K27ac peaks and H3K27ac ASE-SNVs. Peaks were called by combining all samples from each data type. The number of SNPs analyzed for each GWAS was: 2,516,407 SNPs for heart rate, 11,779,664 SNPs for atrial fibrillation, and 2,712,310 SNPs for PR interval.


a
Chr3 coordinates (hg19)
179,100,000
179,150,000
179,200,000
rs7612445
LD
[0,0.2]
(0.2,0.4]
(0.4,0.6]
(0.6,0.8]
(0.8,1.0]
-Log10(P) GWAS heart rate
PPA
fgwas
rs7612445
NKX2-5
P= 2e-15
T-BOX
ATAC-Seq
H3K27ac
RP11-255C15.3
RP11-145M9.4
MFN1
AC007620.3
RP11-145M9.5
GNB4
50 kb
b
EMSA #1 for rs7612445
G
T
Biotin probe allele
Protein extract
Non-biotin competitor
Bound biotin probe
Unbound biotin probe
c
GNB4
P= 0.008
RP11−145M9.4
P= 0.0098
MFN1
P= 0.0188
GG
n= 75
GT
n= 49
TT
n= 4
rs7612445 genotype
d
Chr16 coordinates (hg19)
15,890,000
15,900,000
15,910,000
15,920,000
15,930,000
rs8044595
rs3915499
rs8044595
NKX2-5
P= 1.4e-3
ATAC-Seq
H3K27ac
MYH11
ABCC1
FOPNL
ABCC6
NOMO3
100 kb
Hi-C
e
MYH11
P= 0.0026
ABCC1
P= 0.0031
ABCC6
P= 0.042
NOMO3
P= 0.00166
AA
n= 61
AG
n= 56
GG
n= 11
rs8044595 genotype
f
Chr12 coordinates (hg19)
110,900,000
111,000,000
111,100,000
111,200,000
111,300,000
rs6606689
rs11065706
rs6606689
fgwas
NKX2-5
P= 3.8e-9
ATAC-Seq
H3K27ac
ANAPC7
GPN3
PPTC7
HVCN1
ARPC3
FAM216A
TCTN1
PPP1CC
CCDC63
VPS29
RAD9B
100 kb
Hi-C
g
ARPC3
P= 3.99e-07
TT
n= 2
TC
n= 32
CC
n= 94
rs6606689 genotype
h
Chr5 coordinates (hg19)
89,100,000
89,200,000
89,300,000
89,400,000
rs176107
fgwas
rs176107
NKX2-5
P= 2e-12
ATAC-Seq
H3K27ac
MIR3660
MEF2C
AL050132
CETN3
POLR3G
LYSMD3
MBLAC
GPR98
500 kb
Hi-C
i
MEF2C
P= 0.0004
AA
n= 90
AG
n= 32
GG
n= 6
rs176107 genotype
Legend:
Lead GWAS SNPs
NKX2-5 ASE

Supplementary Figure 8

Functional characterization of candidate causal variants at four loci associated with heart rate.

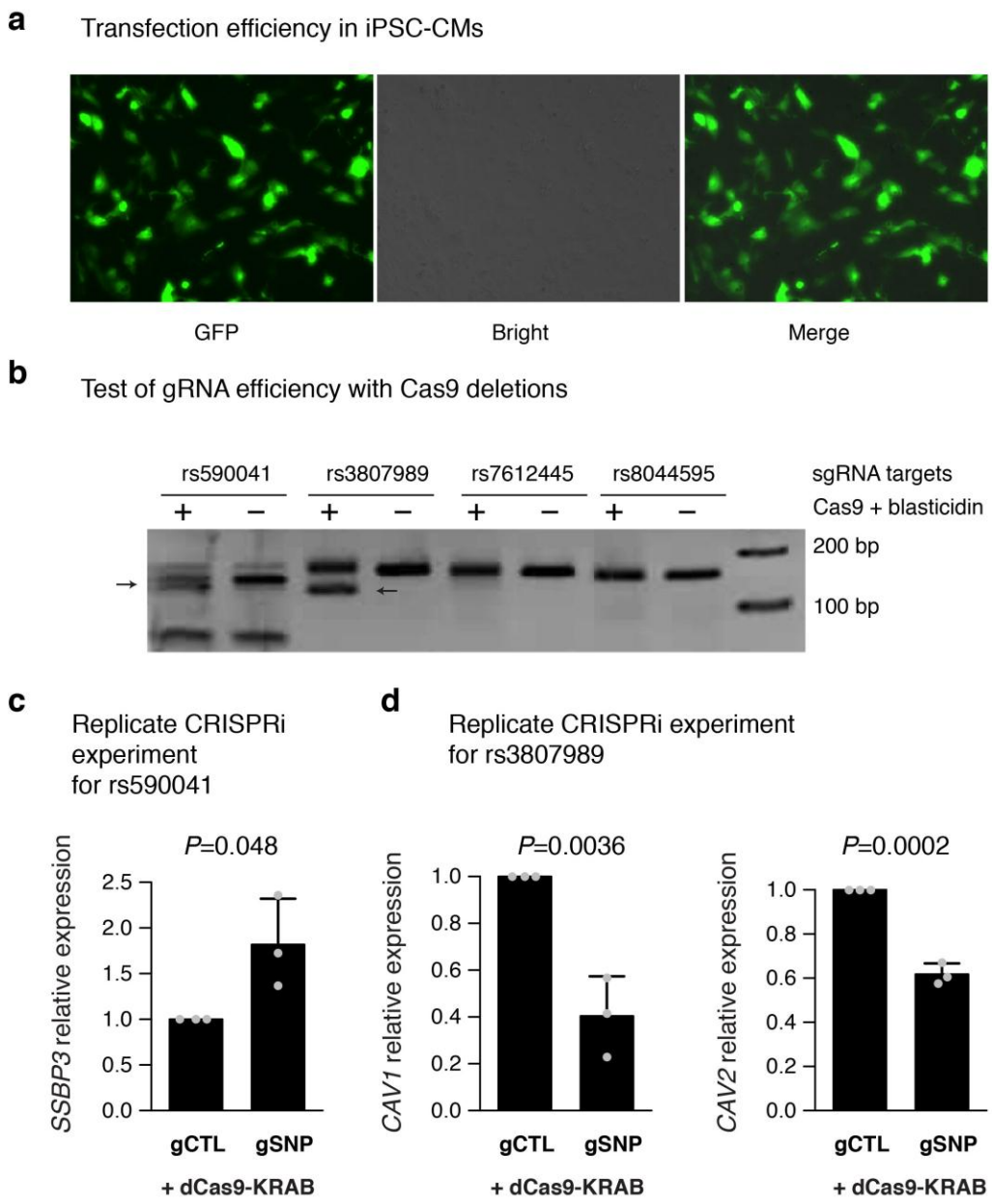
(**a, d, f, h**) For each of the four loci, the top panel shows the regional plot of association *P*-values with heart rate (den Hoed, M. *et al. Nat. Genet.* **45**, 621-631, 2013); SNPs are color coded based on r^2 values from the 1000 Genome Project CEU population (Johnson, A. D. *et al. Bioinformatics,* **24**, 2938-2939, 2008); lead GWAS variants in the locus are indicated by a diamond. The second panel shows the posterior probability of causality (PPA) of the variants in the locus calculated using fgwas, and panels three through five show epigenetic tracks from iPSC-CM combined samples (NKX2-5, ATAC-seq and H3K27a). The bottom panel shows the Roadmap fetal heart ChromHMM and genes from UCSC genome browser (conventional ChromHMM color code). For **d** and **h,** the bottom panel shows the locus at lower scale. For **d**, **f** and **h,** the locations of Hi-C loops from iPSC-CM are shown in red. For the candidate causal variants (turquoise lines), the allelic imbalance (pie chart) of NKX2-5 ASE and FRD-corrected *P*-values are shown; for **a**, the altered TF motif is shown. Significant associations ($P < 0.05$, linear regression) between putative variants genotypes and normalized gene expression of candidate genes in iPSC-CMs from 128 different individuals from iPSCORE are shown (**c, e, g**, **i**). Boxplot elements: median (thick line), lower and upper quartiles (box), maximum and minimum (whiskers). (**b**) EMSA with iPSC-CM nuclear extract using probes containing both allelic variants of rs7612445. An independent replicate is shown in **Supplementary Figure 9.**



Supplementary Figure 9

Electrophoretic mobility shift assay (EMSA) for NKX2-5 ASE-SNVs.

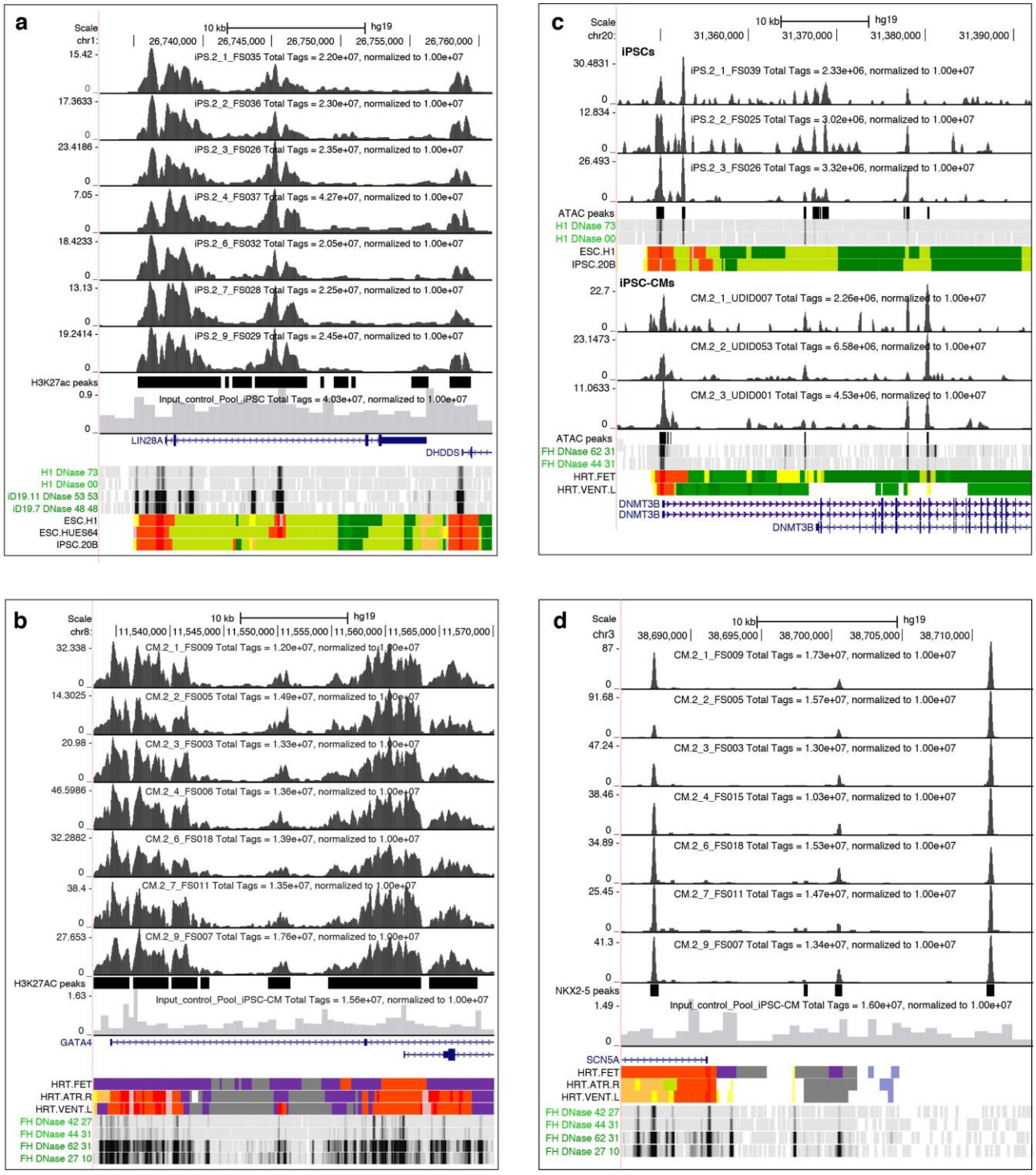
(**a-c**) Second independent replicate of EMSA with iPSC-CMs nuclear extract using probes containing two allelic variants of rs590041 (**a**), rs3807989 (**b**), and rs7612445 (**c**). (**d**) Original (not cropped) blots of all presented EMSAs. The figures and supplementary figures where we showed the corresponding cropped versions are indicated.



Supplementary Figure 10

Experimental validation using luciferase assays and CRISPRi in iPSC-CMs.

(**a**) Representative fluorescence microscopy image showing that in the luciferase assays we achieved approximately 70% transfection efficiency of a GFP-over-expressing plasmid in iPSC-CMs. Efficiency was measured once. (**b**) Test of gRNA efficiency for CRISPR system in HEK293T cells. HEK293T were transfected with two vectors: one containing two gRNAs targeting the indicated SNP (2sgRNA-ccdB-EF1a-Puromycin) and the other expressing Cas9 (Lenti-Cas9-Blast, Addgene #52962). The gRNAs for SNPs rs590041 (*SSBP3* intron) and rs3807989 (*CAV1* intron) showed additional bands corresponding to targeted deletions (arrows) and were used for CRISPRi experiments in iPSC-CMs. The gRNAs for SNPs rs7612445 (*GNB4*) and rs8044595 (*MYH11*) did not show additional bands and were not used for CRISPRi. Efficiency was tested once. (**c**,**d**) qPCR expression of *SSBP3* (**c**) or *CAV1* and *CAV2* (**d**) in iPSC-CMs (id: iPSCORE_1_57) stably expressing dCas9-KRAB (CRISPRi) and either a control guide RNA (gCTL) or two guide RNAs targeting the region encompassing rs590041 (**c**) or rs3807989 (**d**). Bars and error bars represent the mean and the standard deviation from three qPCR measurements, respectively; two-tailed t-test *P*-values are shown. Similar results were obtained in an independent cell line presented in the main manuscript (**Fig. 5i** and **Fig. 7h**).



Supplementary Figure 11

UCSC genome browser screenshots showing quality of ChIP-seq and ATAC-seq data.

Bedgraph tracks are shown for one representative sample from each of the 7 individuals for H3K27ac ChIP-seq: (**a**) iPSCs and (**b**) iPSC-CMs; (**c**) iPSCs and iPSC-CMs samples from 3 individuals are shown for ATAC-seq; and (**d**) iPSC-CMs for 7 individuals for NKX2-5.

Supplementary Information

Table of Content

Supplementary Note

Characterization of iPSC-CMs by flow cytometer and immunofluorescence

All iPSC-CMs displayed synchronous, spontaneous beating, and, on average, 89% +/- 8.4% of the cells in the lactate purified batch and 68.6% +/- 13% in the non-lactate purified batch stained positive for the cardiomyocyte-specific marker cardiac troponin T (TNNT2) by flow cytometry (**Supplementary Table 2, Supplementary Fig 1a**). For flow cytometry, 1 x 10^6 fixed iPSC-CMs were permeabilized and blocked for 30 min at room temperature in 0.5% BSA, 5% goat serum, 0.2% TX-100 in PBS and stained with anti-TNNT2 (Thermo Scientific MA5-12960) for 30 min, followed by anti-mouse Alexa Fluor 488 secondary antibody (Life Technologies A11001) for 30 min. Washes, incubation with antibodies, and flow cytometry analysis were done in 0.5% BSA in PBS. Cells were acquired on a FACSCanto system (BD Biosciences) and analyzed using FlowJo.

Immunofluorescence for TNNT2 and MYL7 on iPSC-CMs confirmed that they developed cardiomyocyte-like cytoskeletal structures (**Supplementary Fig 1b**). For immunofluorescence, three iPSC-CM lines were fixed using 4% PFA in PBS, blocked and permeabilized for 1 h at 37°C with 5% BSA, 5% serum, 0.1% Triton X-100 in PBS. Cells were then incubated with 1:200 dilutions of primary antibodies against TNNT2 (Thermo Scientific MA5-12960) and MYL7 (Synaptics Systems 311011) overnight at 4°C. Slides were washed three times with PBS and incubated for 2 h at room temperature with secondary antibody (AlexaFluor 488) and 20 min with DAPI. Cells were washed with 1× PBS and mounted using Vectashield. Sections were analyzed using a Leica SP5 confocal microscope.

Methodological details of ChIP-Seq, ATAC-Seq and Hi-C

ChIP-Seq. We generated and analyzed 48 ChIP-Seq of histone modification H3K27ac (iPSCs: 17 samples and 4 technical replicates; iPSC-CMs: 25 samples and 2 technical replicates), and 15 ChIP-seq of NKX2-5 (iPSC-CMs: 12 samples and 3 technical replicates) (**Supplementary Tables 2** and **3**). For histone modification H3K27ac, 2 x 10^6 fixed cells were lysed in 60 µl of MAGnify™ Chromatin Immunoprecipitation System Lysis Buffer (Thermo Scientific) and sonicated using Bioruptor 200 (Diagenode) for 35-45 min of 30 sec on/30 sec off cycles. H3K27ac antibodies (Abcam ab4729, lots GR183922-2 (1.75µg), GR184333-2 (1µg), or GR00324078 (1µg)) were coupled for 2-4 hours to ProteinG Dynabeads (Thermo Scientific), and used for overnight chromatin immunoprecipitation in IP buffer (1% Triton-X, 0.1% DOC, 1x TE, 1x Roche Complete Proteinase Inhibitor tablets (RCPI)). Beads were washed five times with washing buffer (50 mM Hepes pH 8, 1% NP-40, 0.7% DOC, 0.5M LiCl, 1mM EDTA and 1x RCPI) and once with TE buffer.For the subset of H3K27ac ChIP-Seq samples with Sample ID denoted by "UDID" (**Supplementary Table 2**), a slightly different protocol was used, where cells were lysed in SDS Lysis Buffer (0.5% SDS, 50mM Tris-HCl pH 8.0, 20mM EDTA, 1x RCPI) using Covaris E220 Focused-ultrasonicators (Covaris) for 14 cycles, 1 min per cycle, duty cycle 5. For NKX2-5, 1-2 x 10^7 cells were lysed in 300 µl RIPA buffer (1xPBS, 1% NP-40, 0.5% DOC, 0.1% SDS, RCPI) and sonicated for 70-80 min using Bioruptor 200 with instrument and setting as above. Five µg of NKX2-5 antibody (Santa Cruz Biotechnology, sc-8697x, lot C0113) were incubated with Dynabeads for 2 hours and washed with BSA 0.5% in PBS. Chromatin was diluted to 1 ml of RIPA buffer and added to the beads for overnight IP. Five washes were performed with washing buffer (50 mM Hepes pH 8, 1% NP-40, 0.7% DOC, 0.5M LiCl, 1mM EDTA and 1x RCPI), followed by one wash with TE. DNA was eluted

and reverse crosslinked in elution buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 1% SDS) at 65˚C, followed by RNAse A (Sigma) and Proteinase K (Thermo Scientific). DNA was purified using Qiagen MinElute PCR Purification kit (or using 2X Agencourt AMPure XP DNA beads, Beckman Coulter), quantified by Qubit (Thermo Scientific) and submitted to library preparation and barcoding using KAPA Hyper Library preparation kit (KAPA Biosystems). Libraries were sequenced on an Illumina HiSeq2500 or a HiSeq4000 to an average of 35 M 100 bp paired-end reads per sample.

ChIP-Seq reads were mapped to the hg19 reference using BWA[1]. Duplicate reads, reads mapping to blacklisted regions from ENCODE, reads mapping in chromosome other than chr1-chr22, chrX, chrY, and read-pairs with mapping quality Q<30 were filtered. Peak calling was performed using MACS2 ('macs2 callpeak -f BAMPE -g hs -B --SPMR --verbose 3 --cutoff-analysis --call-summits -q 0.01')[2] with reads derived from sonicated chromatin not subjected to IP (i.e. input chromatin) from a pool of samples used as negative control. Peak coordinates were called from combined samples of either iPSCs or iPSC-CMs, generated by pooling the BAM files of each data type across all samples of the given cell type. Quantification of the signal at peaks in each sample was performed using featureCounts[3]. Genome browser screenshots showing data quality are presented in **Supplementary Fig. 11**. Motif enrichment analysis was performed using HOMER 'findMotifsGenome.pl'[4] and, for NKX2-5, also using MEME ChIP[5]. Motif analysis of the NKX2-5 ChIP-Seq confirmed a significant enrichment (binomial test, q-value <0.0001) for the NKX2-5 homeobox motif, as well as for the motifs of other heart development TFs (GATA4, TBX5, TBX20, MEF2A/C and MEIS1, **Supplementary Table 4**). Peaks from 14 non-purified and one purified iPSC-CM line showed similar motif enrichment results. Additionally, iPSC and iPSC-CM H3K27ac ChIP-Seq and ATAC-Seq data showed enrichment of stem cell TFs (OCT4, NANOG and SOX family) and cardiac TFs, respectively (**Supplementary Table 4**).

ATAC-Seq. We generated and analyzed 37 ATAC-Seq (iPSCs: 12 samples and 5 technical replicates; iPSC-CMs: 11 samples and 9 technical replicates). The ATAC-Seq protocol has been adapted from Buenrostro et al.[6]. Frozen nuclear pellets of 2.5×10^4 cells (for iPSCs) and 10×10^4 cells (for iPSC-CMs) were thawed on ice, suspended in 25 μL transposition reaction mix (2.5 μL Tn5 transposase from Nextera DNA Library Preparation Kit (Illumina), in permeabilization buffer containing digitonin), and incubated for 30-60 min at 37°C in a thermomixer (500 RPM shaking). Reactions were purified using AMPure XP beads (Beckman Coulter), eluted in 12 μL of DNA Elution Buffer (Zymo) and amplified using the NEBNext® High-Fidelity 2X PCR Master Mix (NEB) with barcoded adaptors. PCR reactions were terminated after 10 to 12 cycles and purified and size selected using AMPure XP beads. Libraries were sequenced on an Illumina HiSeq4000 to an average depth of 20 M 100-150 bp paired end reads.

ATAC-Seq reads were aligned using STAR to hg19 and filtered using the same protocol as for ChIP-Seq. In addition, to restrict the analysis to regions spanning only one nucleosome, we required an insert size no larger than 140 bp, as we observed that this improved sensitivity to call peaks and reduced noise. Peak calling was performed using MACS2 on merged BAM files of iPSC and iPSC-CM combined samples with the command 'macs2 callpeak --nomodel --nolambda --keep-dup all --call-summits -f BAMPE -g hs –q 0.01'. Genome browser screenshots showing data quality are in **Supplementary Fig. 11**.

Hi-C. We used the Hi-C data from 13 iPSC-CMs samples generated in this study (**Supplementary Table 2**) and described in detail in Greenwald and Li et al[7]. Briefly, *in situ* Hi-C was performed on 2-5 million cells as previously described[8]. Formaldehyde-crosslinked cells were lysed and nuclei were

digested with 100U MboI overnight at 37°C. Fragmented ends were biotinylated for 90 min at 37°C, diluted and proximity ligated for 4 hours at room temperature. Crosslinks were reversed and samples were then purified by ethanol precipitation, resuspended in 100uL 1X Elution Buffer, fragmented using a Covaris S2 instrument, and size selected using AmpureXP beads. Subsequently, biotinylated ligation junctions were pulled down using T1 Streptavidin beads and used for library preparation consisting of end-repair, dA-tailing, adapter ligation, PCR amplification and purification. Libraries were sequenced on an Illumina HiSeq 4000 using 150bp paired-end reads.

For each sample, Hi-C reads were aligned to human reference genome hg19 using BWA-MEM[1] (version 0.7.15) with default parameters, with forward and reverse reads from the paired-end data being aligned independently. Paired-end reads were then reconstructed, processed, and filtered using the Juicer pipeline[9]. These filtered read pairs (contacts) were pooled across all iPSC-CM samples to obtain a higher resolution chromatin contact map. Loops were called with a combination of two distinct loop calling algorithms, HICCUPS[8] and Fit-Hi-C[10], and subsequently filtered to high confidence loops with pgltools[11], resulting in 19,003 high-resolution chromatin loops with a median size of 403 kb.

Functional genomic analyses of iPSCs and iPSC-CMs

We generated and analyzed 56 RNA-Seq (iPSCs: 29 independent samples; iPSC-CMs: 26 independent samples and 1 technical replicate), 37 ATAC-Seq (iPSCs: 12 samples and 5 technical replicates; iPSC-CMs: 11 samples and 9 technical replicates), 48 ChIP-Seq of histone modification H3K27ac (iPSCs: 17 samples and 4 technical replicates; iPSC-CMs: 25 samples and 2 technical replicates), and 15 ChIP-seq of NKX2-5 (iPSC-CMs: 12 samples and 3 technical replicates) (**Supplementary Tables 2** and **3**). To demonstrate that iPSCs and iPSC-CMs from this study showed the expected cell-type specific gene expression and epigenetic signatures, we compared their respective functional genomics profiles between the two cell types, as well as with those from the Roadmap Epigenomics project[14].

Analysis of RNA-Seq data for 61 selected signature genes (23 pluripotency genes[12], 22 regulators of cardiac differentiation factors and 16 cardiomyocyte structure factors[13]) (**Supplementary Fig. 1c**) showed that the iPSCs and iPSC-CMs clustered separately from each other, with all iPSCs showing high gene expression of pluripotency markers such as *POU5F1* (OCT4), *SOX2* and *NANOG*; and all the iPSC-CMs expressing high levels of core cardiac TFs such as *NKX2-5*, *TBX5*, and *GATA4* as well as structural proteins including *TNNT2*, *MYL7* and *MYH7*. Principal component (PC) analysis of genome-wide gene expression further showed that iPSC-CMs and iPSCs clustered separately from each other, and closer to human cardiac tissue and stem cell reference samples[14], respectively (**Fig. 1b**), with the iPSC-CMs that were lactate purified having similar PC1 values to fetal heart tissue.

To investigate if the epigenetic profiles in the iPSCs and iPSC-CMs corresponded to the observed gene expression differences, we analyzed the coverage of at the transcription start sites (TSSs) of differentially expressed genes between the two cell types (**Supplementary Fig. 2a**). Specifically, we used HOMER[4] to quantify the read density of ATAC-Seq and ChIP-Seq in iPSC or iPSC-CM combined samples at the TSS of differentially expressed genes between iPSCs and iPSC-CMs (2,444 genes upregulated in iPSCs and 2,863 in iPSC-CMs). For each combined sample, the command 'annotatePeaks.pl tss hg19 -list -size 2000 -hist 25 -ghist –d [alignment files]', was used to obtain the coverage of each combined sample normalized to 10 M reads, in 25-bp

sequential bins across a 2-kb window centered on the TSS. We observed that molecular data from iPSCs on average showed higher coverage at the TSSs of iPSC upregulated genes compared with iPSC-CMs upregulated genes, and vice-versa for the iPSC-CM molecular data. These data show that coordinated changes of gene expression, chromatin states and TF binding occurred during the transition of iPSCs into iPSC-CMs.

To identify which tissues among those in the Roadmap Epigenomics project[14,15] (127 tissues) were the most similar to the iPSC-CMs and iPSCs, we examined the enhancers of each Roadmap tissue, as enhancers are known to be highly tissue-specific. Specifically, genomic coordinates identified by ChIP-Seq and ATAC-Seq experiments from iPSC or iPSC-CM combined samples were intersected with the Roadmap Epigenomics 25-state ChromHMM annotations of 127 human tissues[15] using 'bedtools jaccard'. For each tissue and each chromatin state, the *Jaccard* similarity score measured the number of intersecting bases divided by the union of each pair of dataset. Enhancer chromatin states ('9_TxReg', '10_TxEnh5', '11_TxEnh3', '12_TxEnhW', '13_EnhA1', '14_EnhA2', '15_EnhAF', '16_EnhW1', '17_EnhW2', '18_EnhAc') were selected. Similarity scores for each enhancer chromatin state were mean-centered normalized across all 127 tissues, ordered by mean similarity and hierarchically clustered and plotted using 'pheatmap' (**Supplementary Fig. 2b-f**). As expected, the tissues with the highest degree of enhancer similarity with ChIP-Seq and ATAC-Seq were fetal heart for iPSC-CMs, and stem cells for iPSCs. Notably, cardiac and mesodermal tissues were also among the top 10 tissues most similar to iPSC-CMs. These analyses show that the genomic datasets generated for the iPSCs and iPSC-CMs recapitulate cell type-specific gene expression and regulatory elements.

Comparison of allele-specific effects between ChIP-Seq and ATAC-Seq datasets

The different data types showed over a 30-fold difference in the percent of heterozygous SNPs located within regions (genes or peaks) that displayed ASE (**Fig. 2**). NKX2-5 ChIP-Seq showed the highest fraction of ASE-SNVs (10% of tested SNVs), RNA-Seq showed intermediate values (2.5% in iPSC-CMs), while H3K27ac (0.7% in iPSC-CMs) and ATAC-seq (0.3% in iPSC-CMs) were considerably lower. We examined technical aspects of ASE analysis in ChIP-Seq and ATAC-Seq peaks, in order to elucidate potential experimental and/or biological reasons underlying these observed differences. Detecting ASE-SNVs is highly dependent on the depth of sequencing reads at heterozygous variants within the peaks. The median coverage of heterozygous SNVs ranged between 9 and 88 reads and, as expected, was dependent on the number of mapped reads in each subject (after merging different samples from the same individual), as well as on the mean FrIP per subject in each assay (**Supplementary Fig. 4a-e**). Within a given data type, samples with higher SNV coverage had higher number of significant ASE-SNVs identified at FDR <0.05 (**Supplementary Fig. 4f**). However, between the different data types, NKX2-5 had the highest fraction of ASE-SNVs identified given similar median SNV coverage per subject. Given that the distribution of SNV coverage varied across the different data types (**Supplementary Fig. 4g**), we examined if the higher efficiency of the NKX2-5 ChIP-Seq assay was due a greater fraction of SNVs having high coverage. We performed a simulation where we sampled 100 SNVs from the iPSC-ATAC-Seq list (as a reference distribution) and sampled an equal number of SNVs with the same coverage in the remaining datasets, and repeated 100 times. For each sampling, we calculated the number of SNVs with significant ASE at FDR<10% and the average allele frequency of the most represented allele (effect sizes). This analysis showed that independently from coverage, the fraction of ASE-SNVs in NKX2-5 ChIP-Seq was more than 10 fold greater than the other assays

(**Supplementary Fig. 4h**), which was consistent with higher observed effect sizes (**Supplementary Fig. 4i**). We next examined correlation of ASE effects of the same SNV in peaks from the different data types in iPSC-CMs. Although the three assays resulted in different sets of ASE-SNVs, with limited overlap, the direction of effect was consistent between them (**Supplementary Fig. 4h-j**). Overall, these data suggest that many ASE-SNVs identified in one data type but not the others may have been due to due to power differences between the assays. ChIP-Seq of NKX2-5 was a more powerful approach to detect ASE-SNVs, potentially because it directly measures differential TF binding, whereas ATAC-Seq and H3K27ac measure altered chromatin accessibility and histone modification, respectively, which are indirect consequences of differential TF binding.

Details of GWAS enrichment analyses

Enrichment of GWAS-SNPs in regulatory regions in iPSC-CMs. To calculate enrichment for GWAS-SNPs in ChIP-Seq and ATAC-Seq peaks, we extracted sets of lead SNPs associated with 6 EKG traits from the GWAS catalog[16]: QT interval (matching 'QT interval' or 'electrocardiography, QT interval', 156 SNPs); PR interval/segment (matching 'PR segment', 'PR interval', 'electrocardiography, PR interval', or 'electrocardiography, PR segment', 52 SNPs); QRS duration (matching 'QRS duration', 'QRS complex', 'QRS complex, QRS duration', 'QRS amplitude, R wave amplitude', 'QRS amplitude, QRS complex', 80 SNPs); atrial fibrillation (matching 'atrial fibrillation', 73 SNPs); 'heart rate (matching 'heart rate', 'RR interval' or 'resting heart rate', 175 SNPs); and P wave duration (matching 'P wave duration' and 'electrocardiography, P wave duration'). To ensure that SNPs within each set were independent from each other, we pruned the lists of GWAS-SNPs using the '--clump' function in PLINK v1.07[17] with cut-offs of r^2=0.5 and distance of kb=250. The SNP with lowest GWAS *P*-value was kept. We also extracted and pruned SNPs associated with 119 other traits that were associated with a similar number of SNPs (between 50 and 160 independent SNPs per trait), which are listed in **Supplementary Fig. 6.** We used GREGOR[18] to test each of these 125 SNP sets for enrichment in ChIP-Seq and ATAC-Seq peaks from iPSCs and iPSC-CMs from this study as well as in peaks from cardiac tissues from Roadmap (fetal heart DHS, right ventricle and right atrium H3K27ac), used as positive controls. An r^2> 0.8 (1000 Genomes phase 1 EUR) was used to include LD proxies and 500 random null sets matched on minor allele frequency and number of SNPs were used to calculate *P*-value and fold-enrichment over expectation based on a binomial distribution[18]. Enrichment *P*-values were BH FDR corrected for multiple testing using 'p.adjust' (R).

To calculate the enrichment for EKG GWAS-SNPs in NKX2-5 ASE-SNVs, we used GREGOR to obtain SNVs overlapping NKX2-5 peaks and associated with any of the six EKG traits (417 independent lead variants plus variants in LD r^2>0.8 in EUR). For the SNVs that could be tested for ASE (heterozygous and a minimum of 8X read depth), we calculated the proportion with and without ASE and tested their relative enrichment using the Fisher exact test.

In **Supplementary Table 5**, we also show overlap between all GWAS-SNPs in the NHGRI-EBI catalog (all traits) and 3,117 ASE-SNVs found in ChIP-Seq of NKX2-5 (1,941), H3K27ac in iPSC-CMs (803), H3K27ac in iPSCs (373), ATAC-Seq in iPSCs (129) and ATAC-Seq in iPSC-CMs (79). To annotate these variants we obtained all lead SNPs from the NHGRI-EBI GWAS catalog accessed on 10/20/2017[16] and extracted SNPs in LD with them (r^2>0.8) in Asian (ASN) and European (EUR) populations from the 1000 Genomes Project Phase 1[19].

Estimating GWAS enrichment in molecular phenotypes and prioritizing putative causal

variants. To determine the enrichment of genetic variants influencing EKG traits within the different iPSC-CM molecular phenotypes and to identify putative causal variants and novel associations, we employed the fgwas framework, as described by Pickrell et al[20]. Briefly, the fgwas model uses as input GWAS summary statistics and divides the genome into segments of approximately the same size. The model estimates the prior probability of a segment to contain an association and the prior probability of a variant to be causal, assuming that there is either one or no causal SNP per segment. The prior probabilities are weighted using enrichment estimates of different annotations across the genome using Bayes factors.

We obtained summary statistics from the den Hoed et al.[21] heart-rate GWAS meta-analysis (2,516,407 SNPs analyzed) from LD hub (http://ldsc.broadinstitute.org/ldhub/), the Christophersen et al. atrial fibrillation meta-analysis[22] (11,779,664 SNPs) from the CVD portal (http://broadcvdi.org/) and the van Setten at al. PR-interval[23] GWAS (2,712,310 SNPs) as a collaboration with the authors. For each GWAS, we first annotated each variant with the type of molecular phenotype it overlapped: peaks (ATAC-seq, H3K27ac, and NKX2-5 peaks) and/or ASE-SNVs (H3K27ac and NKX2-5). We then followed the approach recommended by Pickrell et al[20]. First, we calculated enrichment for each of the 5 annotation individually (single annotation model), using the β and standard error of the β from the GWAS as inputs, to consecutive ~1Mb intervals across the genome. The -k parameter, representing the average number of SNPs per segment was set to 900, 4,000 and 1,000 for heart rate, atrial fibrillation and PR interval GWASs, respectively. Then, we added annotations into the model one at a time (joint model) keeping the annotation that increased the likelihood of the model the most, until no more annotations remained or increased the likelihood. Next, we determined the best cross-validation penalty by trying penalties between 0.01 and 0.30 in steps of 0.01. As the models with all 5 annotations had a better likelihood than any model with 4 annotations, we used the full joint model to update the Bayes Factors for each variant using the cross-validation estimated ridge parameters, and calculated the posterior probability of association (PPA) for each variant within the same consecutive 1Mb windows used above. The PPA is the proportion of the total GWAS risk signal at a locus measured by Bayes Factors that is attributed to a particular variant, multiplied by the probability that the genomic region contained a real signal. We report all variants with final PPA>0.3 in **Supplementary Table 6**.

Characterization of NKX2-5 ASE-SNVs prioritized using fgwas as causal for heart rate association

To further investigate the mechanisms of association between heart rate and NKX2-5 ASE-SNVs identified as candidate causal variants by fgwas (**Fig. 6a,b**), we followed up three loci previously associated with heart rate (rs7612445, rs8044595 and rs6606689**)** and a potential novel locus (rs176107) with additional experimental data. These data included Hi-C chromatin conformation maps from the same iPSC-CM lines used to generate the other molecular phenotypes in this study[7] (**Supplementary Table 2A**), in order to identify long-range interactions between the candidate causal variants and target genes, and RNA-Seq data from iPSC-CMs from an additional 128 whole-genome-sequenced subjects[24], to evaluate the association of NKX2-5 ASE-SNVs genotypes with the expression of candidate genes.

rs7612445 was previously identified as a lead heart rate GWAS variant[21,25], as well as a heart-specific eQTL in GTEx for *GNB4* and two non-coding genes (*RP11-145M9.4* and AC007620.3). Our fine-mapping analysis showed that rs7612445 had a probability of 98% of being

causal, which is further supported by the fact that the variant creates a T-box binding sequence, resulting in increased NKX2-5 binding in the iPSC-CMs (**Supplementary Fig. 8a**). We further experimentally validated allele-specific protein binding by EMSA using iPSC-CMs nuclear extract. (**Supplementary Fig. 8b**). Analysis of the RNA-Seq data from the 128 iPSC-CM lines showed associations between rs7612445 and expression of *GNB4* and *RP11-145M9.4* as observed in GTEx (**Supplementary Fig. 8c)**, as well as other nearby genes: *MFN1*, *NDUFB5, ZNF639*, *PIK3CA*, and *RP11-255C15.4*. For four of these genes (*GNB4*, *RP11-145M9.4*, *MFN1* and *NDUFB5,*), the allele of rs7612445 with preferential TF binding was associated with increased gene expression (positively associated, i.e activation), while for the other three genes (*ZNF639*, *PIK3CA*, and *RP11-255C15.4*), it was associated with decreased gene expression (negatively associated, i.e. repression). Although it is unclear which gene(s) in the interval are responsible for the GWAS association, the mitochondrial fusion protein *MFN1* was previously indicated as a promising candidate, based on a reduction in heart rate when the gene was knocked-down in both Zebrafish and Drosophila models[21].

rs8044595 had a high probability of being causal (89%) in our fine-mapping analysis, but was not detected in the heart rate GWAS meta-analysis for which we used summary statistics[21]. However, it was associated with heart rate in a larger GWAS[25], with rs8044595 being in perfect LD with the lead SNP (rs3915499) (**Supplementary Fig. 8d**). Using the Hi-C data, we determined that rs8044595 was in a chromosome loop that would result in contact of the regulatory element with five genes (*MYH11*, *FOPNL*, *ABCC1*, *ABCC6* and *NOMO3*) spanning ~500kb. Gene expression analysis in the 128 iPSC-CMs revealed positive associations with *MYH11*, *ABCC1*, *ABCC6* and *NOMO3* (**Supplementary Fig. 8e**). Among the four genes, *NOMO3* is the most plausible candidate, as the protein is highly expressed in cardiomyocytes[26], is part of Nodal signaling, which plays an essential role in development, and has been implicated in with congenital heart defects through copy number alterations in the interval encoding the gene [27,28].

rs6606689 had an 86% probability of causality, and was significantly associated through LD with a lead variant (rs11065706) in the first stage of the heart rate GWAS meta-analysis (**Supplementary Fig. 8f**); however, it did not replicate in stage 2. Among the 10 genes in the same Hi-C loop (~400kb), we found that *ARPC3* was the only gene whose expression was associated with rs6606689 in iPSC-CMs (positive association observed) (**Supplementary Fig. 8g**). In GTEx, rs6606689 is an eQTL for *ARPC3* in skeletal muscle. These data combined with the fact that actin-related protein 2/3 complex subunit 3 (ARPC3) is implicated in the control of actin cytoskeleton polymerization in cells, supports our fine-mapping analysis suggesting that rs6606689 is likely a causal variant associated with heart rate, potentially by its regulation of *ARPC3*.

rs176107, while having a 35% probability of causality, was never reported as either a GWAS variant for any EKG trait, or an eQTL for any tissue; however, it had a *P*-value of 1.7×10^{-4} for association in the heart rate GWAS meta-analysis, and therefore was considered a “sub-threshold” signal (**Supplementary Fig. 8h**). Using Hi-C, we determined that the *MEF2C* gene, located ~1.2 Mb away, was brought into close proximity with the NKX2-5 binding site through a long-range interaction. Further, expression analysis of *MEF2C* showed a positive association with rs176107 (**Supplementary Fig. 8i**). MEF2C is a transcription factor critical for muscle, heart and neuronal development, it interacts with NKX2-5 itself[29]; and a KO in mice has been shown to result in early development lethality due to severe cardiovascular defects[30,31]. Thus, although MEF2C has not previously been linked to cardiac defects in humans, it is as is an excellent candidate gene for heart phenotypes.

Supplementary Tables 1, 3 and 7

Subject ID	Subject UUID	Family Relationship	Sex	Age	Recorded Ethnicity	*KCNH2* p.Trp1001* genotype	WiCell iPSC ID	Somatic CNV
iPSCORE_2_1	f772212acd9840f9bdcde75740c1f6be	Proband	F	18	European	+/-	UCSD102i-2-1	CN Loss chr5:128294900-128536973 (242 kb)
iPSCORE_2_2	83dacb1141804807b09905fd1561a722	Sister	F	21	Asian/ European	+/-	UCSD103i-2-2	
iPSCORE_2_3	e932e55659a64f709b4cef5f69dac3ce	Mother	F	48	Asian/ European	+/-	UCSD104i-2-3	
iPSCORE_2_4	bd04a8cc5d6345bca2cc91b0c7cb6e01	Mother's sister	F	47	Asian/ European	+/-	UCSD105i-2-4	
iPSCORE_2_6	ae3637d4493e482ebc486915e30ffb9c	Mother's mother	F	74	European	+/+	UCSD107i-2-6	
iPSCORE_2_7	6a784dd79c48484193f9930b9de49cc4	Mother's father	M	77	Asian	+/-	UCSD108i-2-7	
iPSCORE_2_9	f549b5faa6c049fb8a07dda4f72ff076	Father	M	52	European	+/+	UCSD110i-2-9	Allelic Imbalance chr3:30048710-30249375 (200 kb)

Supplementary Table 1. Subjects and iPSC lines. The subject ID and the subject universally unique identifier (UUID) are both provided with the family relationships, sex, age and recorded ethnicity at the time of recruitment. Five members of this family segregate the long-QT syndrome type 2 mutation *KCNH2* p.Trp1001* (rs121912509, c.3003G>A), and the genotypes are reported. However, the disease phenotype is not analyzed in this study. iPSC lines analyzed in this study are available at WiCell Research Institute and the corresponding WiCell iPSC ID is given. Analyses for somatic CNVs were performed on iPSC lines at passage 12 or 13 as described [32].

Assay	Cell type	Number of samples	Average number of uniquely mapped million reads	Average number of peaks identified +/- sd	Average FRiP +/- sd (%)	Total number of peaks or genes analyzed	Median peak size (bp)
RNA-Seq	iPSC	29	22.6 +/- 3.8	\	\	14,933	\
RNA-Seq	iPSC-CM	27	23.2 +/- 6.3	\	\	15,725	\
ATAC-Seq	iPSC	17	19.0 +/- 8.2	59,390 +/- 35,904	11.42 +/- 4.1	147,459	158
ATAC-Seq	iPSC-CM	20	21.7 +/- 14.3	73,569 +/- 47,775	11.7 +/- 5.6	110,095	147
ChIP-Seq H3K27ac	iPSC	21	40.2 +/- 17.4	101,166 +/- 44,731	15.4 +/- 4.7	110,345	500
ChIP-Seq H3K27ac	iPSC-CM	27	27.3 +/- 12.2	132,701 +/- 38,105	36.4 +/- 11.4	83,689	695
ChIP-Seq NKX2-5	iPSC-CM	15	26.3 +/- 5.5	40,519 +/- 22,748	11.2 +/- 6.1	37,994	333

Supplementary Table 3. Sequence data metrics for six functional assays. For each assay, we report separately for each cell type (iPSC or iPSC-CM) the number of samples analyzed, the average number of uniquely mapped reads (PE 100 for all assays). For ATAC-Seq and ChIP-Seq data we show the average number of peaks identified and FRiP (fraction of reads mapped within peaks) plus or minus the standard deviations (sd). We show the total number of peaks or transcripts identified in combined samples (merging the reads for each assay from all samples of a cell type) and the median peak size. Per sample metrics are given in **Supplementary Table 2B.**

Name	Oligo FOR or gRNA 1	Oligo REV or gRNA 2	Size (bp)	Experiment
gCTL	ACGGAGGCTAAGCGTCGCAA	-	-	CRISPRi gRNA - control
gSNP CAV1- rs3807989	ACTGTTCAGTGTCACTTGTTTGG	TTTGTCCTCTGGTGTCAATTTGG	71	CRISPRi gRNAs
gSNP SSBP3 - rs590041	TGAGTaTGGGCTTGATCCAGTGG	GATCAAGCCCAtACTCAACAAGG	29	CRISPRi gRNAs
CAV1 - NKX2-5 peak	CCAGGTAGAAAAACCCTGTTGTA	CATTAATGATGGAGGCAGAGTGG	1715	Luciferase region cloning
SSBP3 - NKX2-5 peak	GTCTGGGTCATATCTCCGGC	TGCAACTTACTTTCATTTTCATCAC	1675	Luciferase region cloning
CAV1 - gene	TCTTCCAACACGTAGCTGCC	CGGTGTAGAGATGTCCCTGCG	96	qPCR
CAV2 - gene	AGTTCCTGACGGTGTTCCTG	CGTCCTACGCTCGTACACAA	193	qPCR
METTL2B - gene	TGTGGTGTGGGAAACACAGT	CCAGTTCTATAGCTGTGGAAGAAA	100	qPCR
SSBP3 - gene	ACTGTGCAGCTCCTGAAAGG	CCGGAGGACCCTGAAAGAAA	168	qPCR
SSBP3 rs590041_T	CACTGGATCAAGCCCA**T**ACTCAACAAGGTCCTTC		34	EMSA
SSBP3 rs590041_C	CACTGGATCAAGCCCA**C**ACTCAACAAGGTCCTTC		34	EMSA
CAV1 rs3807989_A	GTTTGGTAACACTCAACATCAAC**A**TGTGCTACCA		34	EMSA
CAV1 rs3807989_G	GTTTGGTAACACTCAACATCAAC**G**TGTGCTACCA		34	EMSA
GNB4 rs7612445_G	GACAGATATCATTGAC**C**CTCTCTTAAGGTCACT		33	EMSA
GNB4 rs7612445_T	GACAGATATCATTGAC**A**CTCTCTTAAGGTCACT		33	EMSA

Supplementary Table 7. Sequences of primers, sgRNAs and ssDNA oligos.

References

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
2. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
3. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
4. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-89 (2010).
5. Machanick, P. & Bailey, T.L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696-7 (2011).
6. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213-8 (2013).
7. Greenwald, W.W. *et al.* Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nat Commun* **10**, 1054 (2019).
8. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
9. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-8 (2016).
10. Ay, F., Bailey, T.L. & Noble, W.S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res* **24**, 999-1011 (2014).
11. Greenwald, W.W. *et al.* Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. *BMC Bioinformatics* **18**, 207 (2017).
12. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-20 (2007).
13. Paige, S.L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221-32 (2012).
14. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
15. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-6 (2012).
16. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
17. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
18. Schmidt, E.M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601-6 (2015).
19. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
20. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559-73 (2014).
21. den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nature genetics* **45**, 621-31 (2013).
22. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet* **49**, 946-952 (2017).
23. van Setten, J. *et al.* PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun* **9**, 2904 (2018).
24. D'Antonio-Chronowska, A. *et al.* Human iPSC gene signatures and X chromosome dosage impact response to WNT inhibition and cardiac differentiation fate. *bioRxiv* https://doi.org/10.1101/644633(2019).
25. Eppinga, R.N. *et al.* Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nat Genet* **48**, 1557-1563 (2016).
26. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
27. Warburton, D. *et al.* The contribution of de novo and rare inherited copy number changes to congenital heart disease in an unselected sample of children with conotruncal defects or hypoplastic left heart disease. *Hum Genet* **133**, 11-27 (2014).
28. Harvey, R.P. Patterning the vertebrate heart. *Nat Rev Genet* **3**, 544-56 (2002).
29. Vincentz, J.W., Barnes, R.M., Firulli, B.A., Conway, S.J. & Firulli, A.B. Cooperative interaction of Nkx2.5 and Mef2c transcription factors during heart development. *Dev Dyn* **237**, 3809-19 (2008).
30. Lin, Q., Schwarz, J., Bucana, C. & Olson, E.N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**, 1404-7 (1997).
31. Naya, F.J., Wu, C., Richardson, J.A., Overbeek, P. & Olson, E.N. Transcriptional activity of MEF2 during mouse embryogenesis monitored with a MEF2-dependent transgene. *Development* **126**, 2045-52 (1999).
32. Panopoulos, A.D. *et al.* iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. *Stem Cell Reports* **8**, 1086-1100 (2017).