Genome-wide analysis of heart failure enables polygenic and monogenic prediction of heart failure risks

Table of Contents

Supplementary Notes 4

1. **Replication of previously identified HF risk loci 4**
2. **Internal and external replication for novel loci found in BBJ GWAS 4**
3. **Cohorts used for cross-ancestry meta-analyses 5**
4. **Internal and external replication of cross-ancestry meta-analyses results.. 5**
5. **Prioritization of rs10851802 in the cross-ancestry meta-analyses..... 6**
6. **Prioritized genes in MTAG 6**
7. **Detailed explanations for novel loci 7**

Supplementary methods 12

Cohort characteristics for BBJ 1st cohort and BBJ 2nd cohort 12

Cell type-specific heritability enrichment of disease associations using stratified LD score regression..... 12

Sex stratified analysis 13

Pleiotropy check 13

Links between medication and prioritized loci 13

miRNA enrichment analysis..... 14

Heritability enrichment with scDRS 14

Supplementary Figures 15

Supplementary Fig. 1 | Results of Japanese GWAS 15

Supplementary Fig. 2 | Validation of previously identified loci using BBJ 16

Supplementary Fig. 3 | Comparison of allele frequencies and allelic effects identified in Japanese GWAS 17

Supplementary Fig. 4 | Downstream analysis of Japanese GWAS 19

Supplementary Fig. 5 | Phenogram of genome-wide significant loci for HF and its subtypes 21

Supplementary Fig. 6 | Results of the cross-ancestry meta-analysis 22

Supplementary Fig. 7 | QQ plot and comparison of allele frequencies and allelic effects identified in the cross-ancestry meta-analysis 23

Supplementary Fig. 8 | Downstream analysis of cross-ancestry meta-analysis 25

Supplementary Fig. 9 | Results of multi-trait analysis of GWAS 27

Supplementary Fig. 10 | Downstream analysis of multi-trait analysis of GWAS 29

Supplementary Fig. 11 | Venn diagram for genes identified by each analysis.. 31

Supplementary Fig. 12 | miRNA enrichment analysis and Single-cell enrichment analysis 32

Supplementary Fig. 13 | Candidate drugs linked to disease susceptibility loci for HF..... 33

Supplementary Fig. 14 | Analytical scheme for PRS development 34

Supplementary Fig. 15 | PRS performance 35

Supplementary Fig. 16 | LocusZoom plots for Japanese GWAS 37

Supplementary Fig. 17 | LocusZoom plots for the cross-ancestry meta-analysis 43

Supplementary Fig. 18 | LocusZoom plots for multi-trait analysis of GWAS .. 57

Members of participating consortia 66

References 69

Supplementary Notes

1. Replication of previously identified HF risk loci

We assessed the 47 independent variants previously reported from all-cause HF GWAS[1], 13 variants from HFrEF GWAS, one variant from HFpEF GWAS[2], and one variant from NIHF[3] in our Japanese GWAS (BBJ). Out of 47 variants identified in the previous all-cause HF GWAS, 34 variants or variants tagged with the lead variant existed in BBJ, 28 had the same effect direction, and 18 were nominally significant ($P < 0.05$; **Supplementary Fig. 2, Supplementary Table 7**). Out of 13 variants identified in the European HFrEF GWAS, 11 existed in BBJ, 10 had the same effect direction, and 3 were nominally significant. The lead variant identified in the European HFpEF GWAS had the same direction and was nominally significant in BBJ, and the lead or tagged variant identified in the European NIHF GWAS did not exist in BBJ.

2. Internal and external replication for novel loci found in BBJ GWAS

We sought to replicate the HF-risk loci that reached genome-wide significance in individual GWAS cohorts. The effect size, direction (**Supplementary Fig. 3a**), and allele frequency of the identified lead variants (**Supplementary Figure 3b**) were strongly concordant in BBJ 1st and BBJ 2nd cohorts for all HF phenotypes (Pearson's correlation: all-cause HF 0.981; HFrEF 0.961; HFpEF 0.752; NIHF 0.992). Three out of the five newly identified loci were nominally significant in both BBJ 1st and BBJ 2nd cohorts (**Supplementary Fig. 3a** and **Supplementary Table 8**). Out of the five newly identified loci, three identified in all-cause HF were found in the European GWAS. rs35593046 was successfully replicated with nominal association ($P < 0.05$) with the same effect direction (**Supplementary Fig. 3c** and **Supplementary Table 9**). As for rs6471480 and rs7075837, the same effect direction was observed in the European population.

rs76704104 identified in HFrEF and rs78228190 identified in HFpEF were not observed in the European population according to the Genome Aggregation Database v4.0.0 (gnomAD) (https://gnomad.broadinstitute.org/).

3. Cohorts used for cross-ancestry meta-analyses

Cross-ancestry datasets (**Supplementary Table 23**) yielded 120,203 cases for all-cause HF (BBJ: 16,251, EUR: 95,524, KADOORIE: 1,467, AMR: 1,170, AFR: 5,791), 23,749 cases for HFrEF (BBJ 4,254, EUR: 19,495), 26,743 cases for HFpEF (BBJ: 7,154, EUR: 19,589), and 22,864 cases for NIHF (BBJ: 11,122, UKBB: 1,816, FinnGen: 9,926). The number of control samples was 1,502,645 (BBJ: 197,577, EUR: 1,270,968, KADOORIE: 75,149, AMR: 13,217, AFR: 20,883) for all-cause HF, 456,520 for HFrEF/HFpEF (BBJ: 197,577, EUR: 258,943), and 863,254 for NIHF (BBJ: 171,995, UKBB: 387,652, FinnGen: 303,607), respectively. We included variants with MAF ≥ 1% and tested a total of 11,119,536 variants for all-cause HF, 4,800,324 variants for HFrEF, 4,807,066 variants for HFpEF, and 4,998,828 variants for NIHF.

4. Internal and external replication of cross-ancestry meta-analyses results

The effect size and direction of the identified lead variants (**Supplementary Fig. 7b**) were concordant in East Asians and Europeans for all HF phenotypes (Pearson's correlation: all-cause HF 0.898; HFrEF 0.864; NIHF 0.983) except for HFpEF, for which Pearson's correlation could not be calculated with significant loci < 3.

To replicate lead variants of HFrEF and HFpEF using independent datasets, these were assessed with the European all-cause HF GWAS since there were no other available HFrEF/HFpEF GWAS datasets. As for all-cause HF and NIHF, there was no dataset for external

replication. Effect sizes were generally concordant with Pearson's correlation of 0.817 in HFrEF (**Supplementary Figure 7d and e**). The effect direction was also concordant in all of the three novel loci (**Supplementary Table 26**). As for HFpEF, both significant variants were known HF-related variants.

5. Prioritization of rs10851802 in the cross-ancestry meta-analyses

Among novel loci, we could not prioritize a gene for rs10851802, which was identified in the cross-ancestry meta-analysis for all-cause HF. rs10851802 is linked to *GLCE* according to the Open Targets database. rs10851802 was associated with rs2415040 ($r^2 = 0.548$), which was a variant for lean mass[4]. Lean mass is associated with incident HF[5]. Although we could not prioritize a gene in this locus, this evidence suggests that this locus was one of the disease susceptibility loci. Further investigations were warranted to identify the candidate gene in this locus.

6. Prioritized genes in MTAG

BDNF prioritized by PoPS (**Supplementary Table 45**) harbored pathogenic variants responsible for obesity (**Supplementary Table 38**), and its knockout showed the same phenotype (**Supplementary Table 44**). *MYBPC3* was one of the known cardiomyopathy genes (**Supplementary Table 19**, **43** and **44**), and this is the first report to show that common variants of *MYBPC3* were associated with HF. *SLC4A7*, prioritized by PoPS (**Supplementary Table 45**), was associated with abnormal vasoconstriction (**Supplementary Table 44**). Considering that *SLC4A7* splicing was enriched in fibroblasts (**Supplementary Figure 10b**), it suggested that arterial fibroblasts contributed to HF through hypertension. Common variants in *PCSK1*

influenced blood pressure[6], which was consistent with the fact that *PCSK1* was prioritized by H-MAGMA using the aorta Hi-C dataset (**Supplementary Fig. 10d**) and *PCSK1* knockout showed obesity and hypertension (**Supplementary Table 44**). *PTPRJ* knockout showed abnormal vascular development and abnormal heart development (**Supplementary Table 44**). *PLPP3* was enriched in the fibroblasts in TWAS (**Supplementary Fig. 10a**), and its knockout caused the abnormal vasculogenesis. These results were consistent with a previous report, which indicated that the *PLPP3* region was associated with hypertension[7] and coronary artery disease[8].

Among the novel loci, we could not assign a gene to rs10888955. This variant was strongly associated with reticulocyte count[9]. Given that previous Mendelian randomization analysis revealed the positive association of CAD with reticulocytes[10], this locus could contribute to HF development through CAD.

7. Detailed explanations for novel loci

CACNA1D

This locus was found in a Japanese GWAS for all-cause HF and non-ischemic HF, and cross-ancestry meta-analysis for all-cause HF. The lead SNP is rs35593046 in both the Japanese GWAS and the cross-ancestry meta-analysis. rs35593046 is more frequent in East Asians (53.8%) than in Europeans (26.8%). The alternate allele (T) was associated with protection against both HF (**Supplementary Tables 2** and **24**) and high blood pressure (**Supplementary Table 12**). The Accelerating Medicines Partnership Program for Common Metabolic Diseases also suggested that common variants in *CACNA1D* are associated with blood pressure. Taken together, rs35593046 could protect against HF by lowering blood pressure.

DPY19L4

This locus was found in a Japanese GWAS for all-cause HF, and the lead variant for this locus was rs6471480 (**Supplementary Table 2**). rs6471480 is an eQTL in the subcutaneous adipose tissue and tibial artery (**Supplementary Table 14**). *DPY19L4* was significant in the aorta and adrenal gland H-MAGMA. According to the Accelerating Medicines Partnership Program for Common Metabolic Diseases, common variants in *DPY19L4* are associated with HbA1c, glucose, and type 2 diabetes. Given that *DPY19L4* was also associated with coronary artery disease in our previous study[8], this locus was likely to contribute to HF development through ischemic heart disease.

NEBL

This locus was found in a Japanese GWAS for all-cause HF. The lead variant was rs7075837. The allele frequency of rs7075837 is 44.9% in East Asians and 36.0% in Europeans. *NEBL* is one of the known cardiomyopathy genes in both humans (**Supplementary Table 18** and **19**) and in mice (**Supplementary Table 20**).

SPRED2

This locus was found in a Japanese GWAS for HFrEF, and the lead variant was rs76704104 (**Supplementary Table 2**). The allele frequency of rs76704104 in East Asians is 1.6%, and rs76704104 has not been reported in other ancestries. *SPRED2* is one of the monogenic genes responsible for Noonan syndrome, which causes hypertrophic cardiomyopathy (**Supplementary Table 18**). *SPRED2* knockout mice showed abnormal heart morphology, which is consistent with Noonan syndrome in humans (**Supplementary Table 20**). These data indicated that this locus contributed to HF development through cardiomyopathy. Aside from cardiomyopathy, common variants in *SPRED2* are associated with type 2 diabetes according to the Accelerating Medicines Partnership Program for Common Metabolic Diseases.

CSMD1

This locus was found in a Japanese GWAS for HFpEF. The lead variant was rs78228190. *CSMD1* knockout mice showed changes in body weight and glucose tolerance (**Supplementary Table 20**).

LHX3

This locus was found in a cross-ancestry meta-analysis for non-ischemic HF. The lead SNP was rs11103379. *LHX3* knockdown mice showed decreased body weight (**Supplementary Table 33**).

CARM1

This locus was found in a cross-ancestry meta-analysis for non-ischemic HF. *CARM1* knockout mice showed congenital heart disease (**Supplementary Table 33**). The splicing-TWAS suggested that *CARM1* splicing in the heart was associated with HF (**Supplementary Table 29**), which was consistent with the knockout mouse results.

EDNRA

This locus was found in a cross-ancestry meta-analysis for HFrEF. The lead SNP was rs7659823. *EDNRA* knockout mice showed congenital arterial diseases (patent ductus arteriosus and overriding aortic valve) and heart diseases (ventricular septal defect; **Supplementary Table 33**). The splicing-TWAS suggested that *EDNRA* splicing in the heart was associated with HF (**Supplementary Table 29**). Combined with knockout mouse results and the previous literature, *EDNRA* is supposedly involved in the development of the heart and artery[11].

PDE3A

This locus was found in a cross-ancestry meta-analysis for HFrEF. The lead SNP was rs7962871. According to OMIM, *PDE3A* deficiency causes hypertension and brachydactyly syndrome. The *PDE3A* inhibitor has been used for pulmonary artery hypertension (**Supplementary Fig. 13**).

ZNF397

This locus was found in a cross-ancestry meta-analysis for HFrEF. The lead variant was rs5823966. TWAS showed that *ZNF397* in the tibial artery was associated with HF (**Supplementary Table 28**), and the splicing-TWAS suggested that *ZNF397* splicing in the heart was also associated with HF (**Supplementary Table 29**). According to the Accelerating Medicines Partnership Program for Common Metabolic Diseases, *ZNF397* is associated with left ventricular wall thickness.

PLPP3

This locus was found in MTAG for all-cause HF. The lead variant was rs10888955. TWAS suggested that *PLPP3* in fibroblasts was associated with HF (**Supplementary Table 39**). According to the Accelerating Medicines Partnership Program for Common Metabolic Diseases, *PLPP3* is associated with hypertension and coronary artery disease, which is consistent with the TWAS results.

SLC4A7

This locus was found in MTAG for all-cause HF. The lead SNP was rs582505. The splicing-TWAS suggested that *SLC4A7* splicing in fibroblasts was associated with HF (**Supplementary Table 40**). *SLC4A7* knockdown mice showed abnormal blood pressure and abnormal vasculature morphology (**Supplementary Table 44**). A previous report showed that this variant was associated with hypertension[7], which was consistent with the knockout mouse results.

PCSK1

This locus was found in MTAG for all-cause HF. The lead SNP was rs261967. *PCSK1* has been reported to be responsible for obesity (**Supplementary Table 43**), and *PCSK1* knockout led to obesity and abnormal glucose tolerance (**Supplementary Table 44**). A previous report showed that the tagged variants with rs261967 were associated with body mass index[12].

ILRUN

This locus was found in MTAG for all-cause HF. The lead SNP was rs2492863. TWAS suggested that *ILRUN* was associated with HF in various tissues, including the left atrium, left ventricle, blood, kidney, subcutaneous fat, and kidney (**Supplementary Table 39**). *ILRUN* deficiency in mice resulted in abnormal glucose and lipid metabolism (**Supplementary Table 44**). Since tagged variants with rs2492863 were associated with metabolic phenotypes, such as body mass index[13] and lipids[14], *ILRUN* is likely to contribute to HF development through metabolic disorders.

BDNF

This locus was found in MTAG for all-cause HF and HFpEF. The lead SNPs were rs10501087 and rs35038967, respectively. The splicing-TWAS suggested that *BDNF* splicing in adipose tissue was associated with HF development (**Supplementary Table 40**), and H-MAGMA using the adipose tissue dataset showed a significant association between *BDNF* and HF (**Supplementary Table 42**). Also, *BDNF* has been reported to be responsible for obesity (**Supplementary Table 43**), and *BDNF* knockout mice showed obesity as well (**Supplementary Table 44**).

MYBPC3

This locus was found in MTAG for HFrEF. The lead SNP was rs11570058. *MYBPC3* is one of the well-known cardiomyopathy genes (**Supplementary Table 19**)

PTPRJ

This locus was found in MTAG for HFpEF. The lead SNP was rs7117516. *PTPRJ* knockout mice showed abnormal heart and vascular development (**Supplementary Table 44**).

Supplementary methods

Cohort characteristics for BBJ 1st cohort and BBJ 2nd cohort

All subjects were Japanese and were registered in the BioBank Japan (BBJ) project (https://biobankjp.org/). The BBJ is a hospital-based national biobank project that collects DNA and serum samples and clinical information from 12 cooperative medical institutes throughout Japan (Osaka Medical Center for Cancer and Cardiovascular Diseases, Cancer Institute Hospital of Japanese Foundation for Cancer Research, Juntendo University, Tokyo Metropolitan Geriatric Hospital, Nippon Medical School, Nihon University School of Medicine, Iwate Medical University, Tokushukai Hospitals, Shiga University of Medical Science, Fukujuji Hospital, National Hospital Organization Osaka National Hospital, and Iizuka Hospital). Approximately 267,000 patients with any of the 47 target diseases were enrolled from 2003 to 2007, and from 2013 to 2018. All subjects were at least 18 years old. Informed consent was obtained from all participants, and our study was approved by the relevant ethical committee at each facility.

Cell type-specific heritability enrichment of disease associations using stratified LD score regression

We applied stratified LD score regression (S-LDSC)[15] (v1.0.1. https://github.com/bulik/ldsc) to the GWAS summary statistics of the cross-ancestry meta-analysis to evaluate the contribution of genetic variation in cell type-specific genes to trait heritability. Heritability enrichment per cell

type was considered significant at $P < 7.2 \times 10^{-5}$ (0.05/694) based on the number of cell types tested (694).

Sex-stratified analysis

We conducted sex-stratified analyses for 27 variant-HF phenotype pairs among 18 genome-wide significant loci. We performed Cochran's Q test and showed the *P* value for the Cochran's Q test as well as the I^2 statistic. We also assessed *P* value for the interaction term for sex, applying a Bonferroni-corrected significance threshold as interaction *P* value $< 1.85 \times 10^{-3}$ (0.05/27). In addition to the heterogeneity and *P* value for the interaction term, we showed effect estimates in males and females separately (**Supplementary Fig.1b** and **Supplementary Table 3**).

Pleiotropy check

Among genome-wide significant lead variants in each HF subtype, we assessed their effect sizes using previously reported Japanese GWAS data[12]. From a total of 220 phenotypes, we selected 41 cardiovascular-related phenotypes (heart-, vascular-, blood pressure-, diabetes-, lipid-related traits) after excluding medications and quantitative traits. We reported effect sizes for associated phenotypes when *P* was $< 5.0 \times 10^{-8}$ in each phenotype.

Links between medication and prioritized loci

We searched medications linked to prioritized genes using DrugBank and the Therapeutic Target Database. Among medications with the same pharmacological actions, we manually chose a representative one. The role of prioritized genes (e.g., protective or detrimental against HF) was

determined based on (1) knockout mice phenotyping or (2) the direction of TWAS Z-score if the knockout mice phenotypes were not reported to be associated with HF.

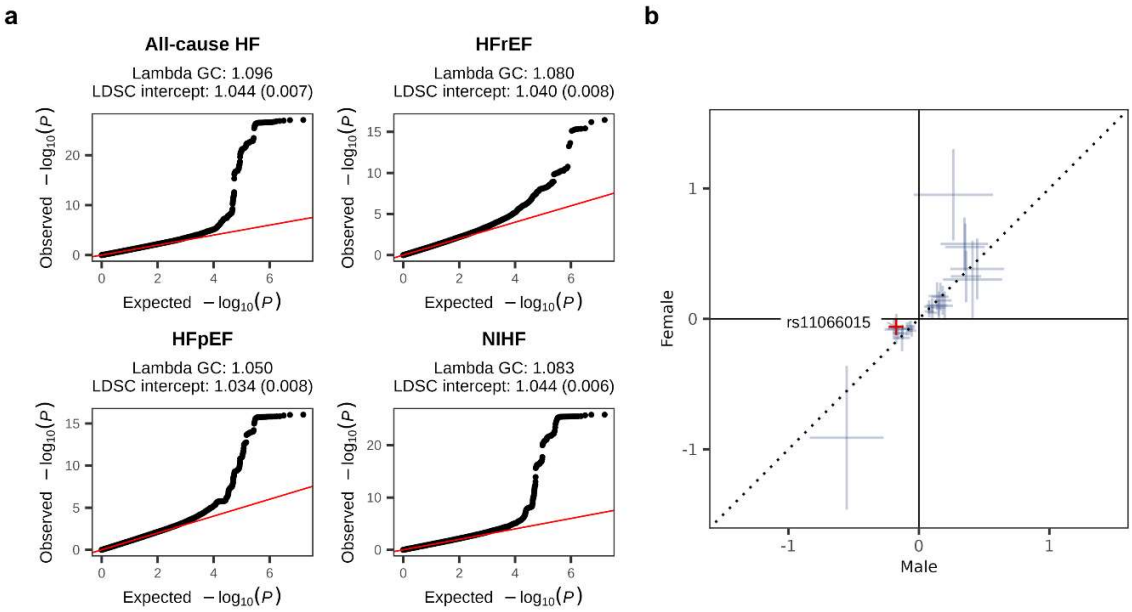
miRNA enrichment analysis

The overall polygenic contribution of miRNA–target gene network to the traits through the tissue-naïve approach was estimated using MIGWAS software with default settings[16]. The significance threshold was set at 0.0125 based on the number of HF phenotypes (0.05/4). Candidate miRNA-gene pairs were estimated by MIGWAS with a significance threshold set at FDR < 0.05 for miRNA and genes.

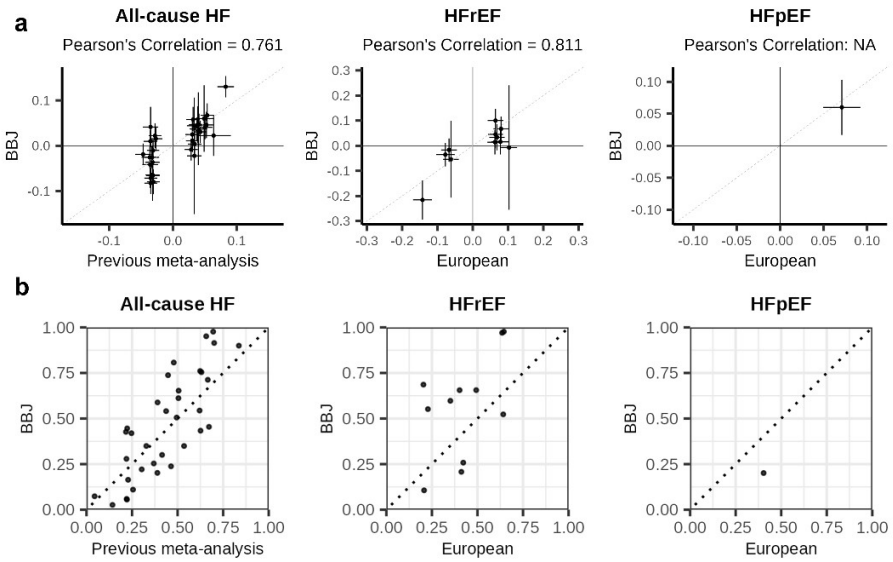
Heritability enrichment with scDRS

To determine which cell types in the heart were enriched for HF GWAS, we used scDRS (v.1.0.2)[17] with default settings. We used MAGMA v.1.10[18] to map single-nucleotide polymorphisms to genes (GRCh37 genome build from the 1000 Genomes Project) using an annotation window of 0 kb. We used the resulting annotations and GWAS summary statistics to calculate each gene's MAGMA z score (association with a given trait). The 1,000 disease genes used for scDRS were chosen and weighted based on their top MAGMA z scores. To determine trait association at the annotated cell type resolution, we used the z scores computed from scDRS's downstream Monte Carlo test. These Monte Carlo z scores were converted to theoretical P values using a one-sided test under a normal distribution. The significance threshold was set at $P < 2.4 \times 10^{-3}$ based on Bonferroni correction (0.05/21 based on the number of cell types).

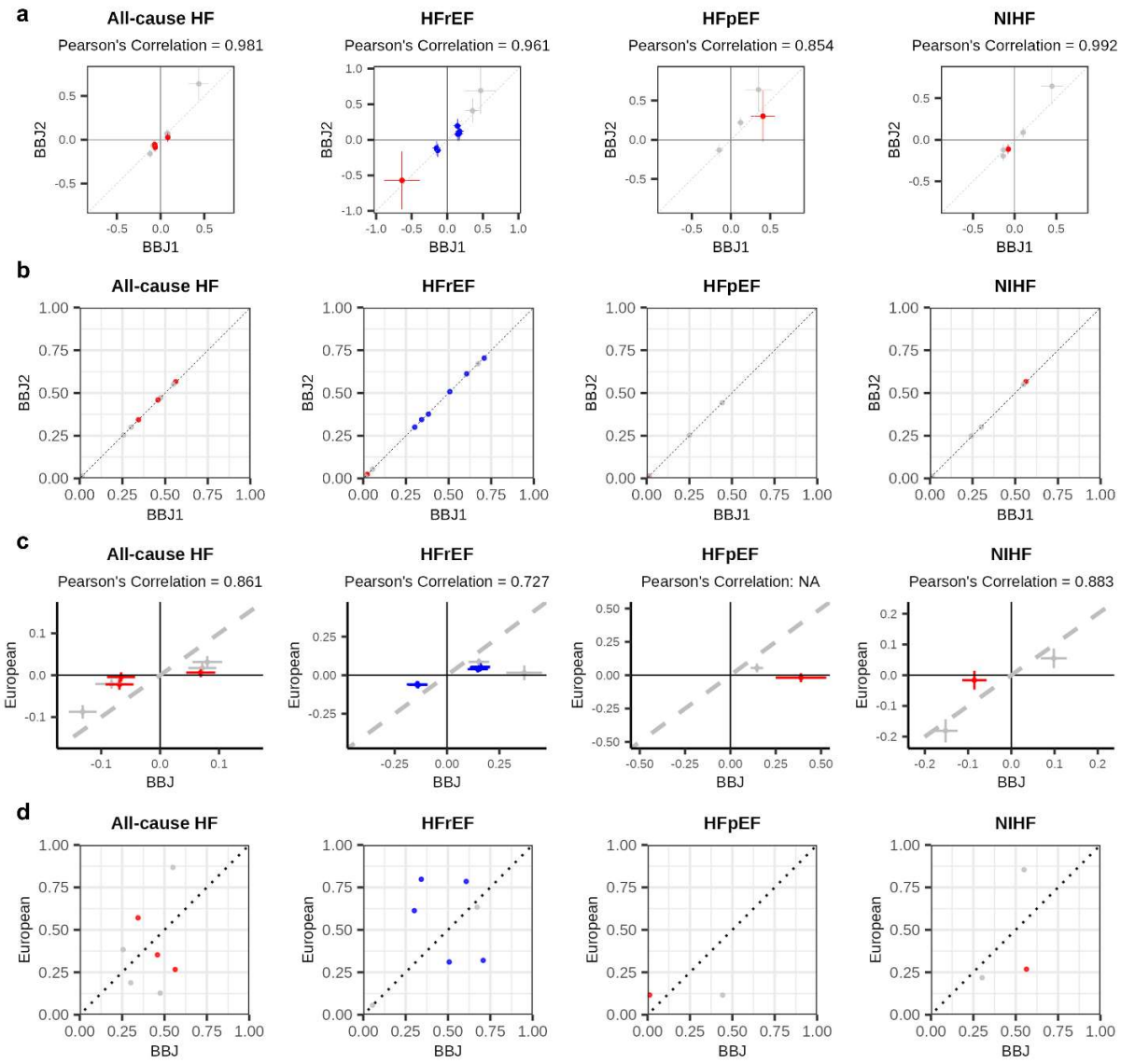
Supplementary Figures



Supplementary Fig. 1 | Results of Japanese GWAS. (a) QQ plot for each HF phenotype in Japanese GWAS. (**b**) The result of the sex-stratified analysis for 27 variant-HF phenotype pairs. Among these pairs, rs11066015 showed a significant sex difference in NIHF. Effect sizes with their 95% CI for males are shown on the x-axis, and those for females are shown on the y-axis. Two-sided uncorrected *P* values were derived from the LDSC software. The statistically significant variant (significance threshold was set at $P < 1.85 \times 10^{-3}$ by the Bonferroni correction) is highlighted in red. HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LDSC, Linkage disequilibrium score regression; NIHF; non-ischemic heart failure.

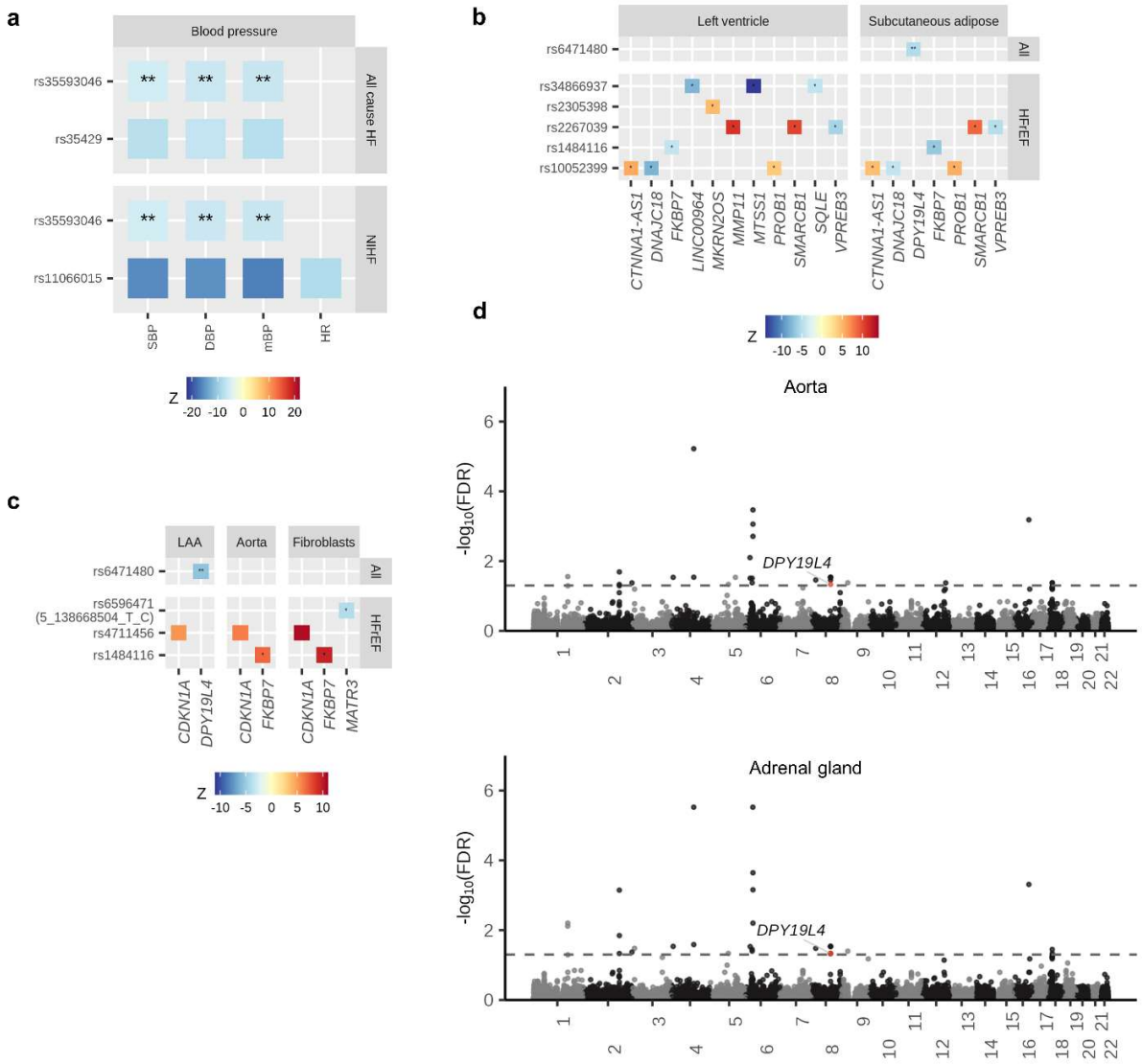


Supplementary Fig. 2 | Validation of previously identified HF loci in BBJ. Comparison of allelic effects **(a)** and allele frequency **(b)** between the previous GWAS and BBJ. Effect size with its 95% CI or allele frequency of previous GWAS are shown on the *x*-axis, and those of BBJ are shown on the *y*-axis. BBJ, BioBank Japan; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction.



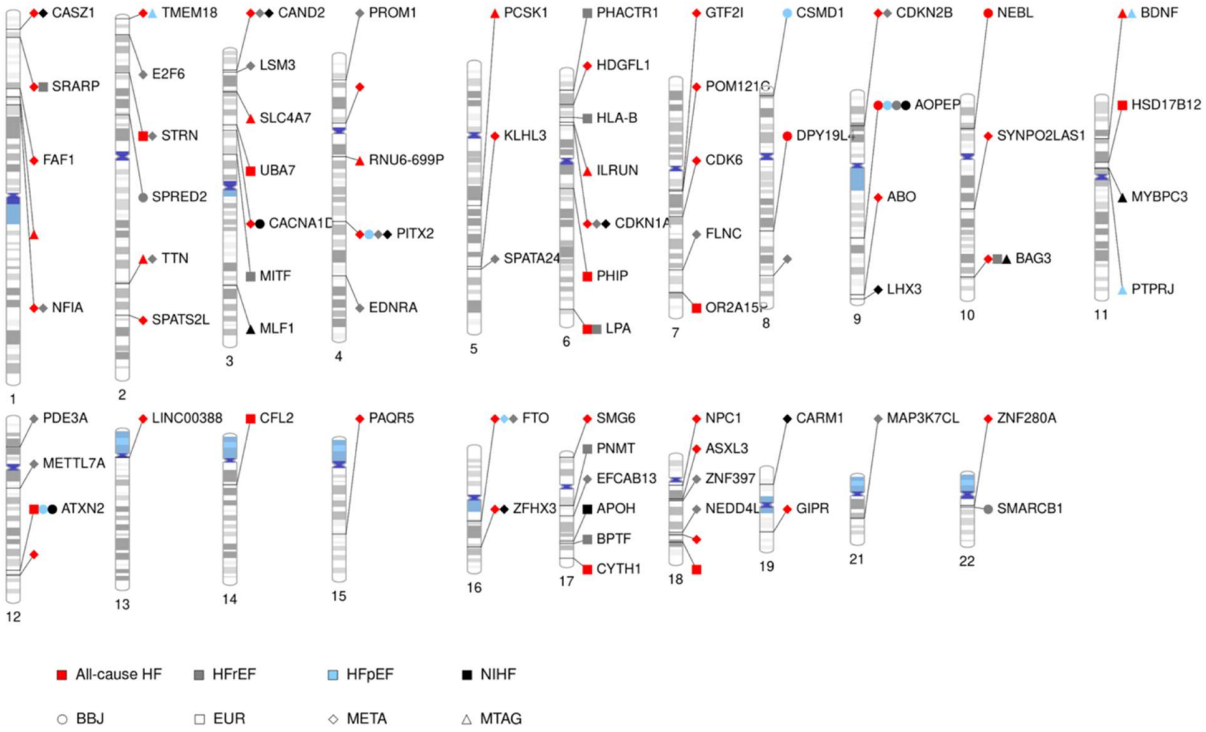
Supplementary Fig. 3 | Comparison of allele frequencies and allelic effects identified in Japanese GWAS. (**a** and **b**) Comparison of allelic effects **(a)** and allele frequency **(b)** between BBJ 1st and BBJ 2nd. Effect size with its 95% CI or allele frequency of BBJ 1st GWAS are shown on the *x*-axis, and those of BBJ 2nd are shown on the *y*-axis. **(c** and **d)** Comparison of allelic effects (**c**) and allele frequencies (**d**) between our Japanese GWAS (BBJ) and European GWAS. Effect size and its 95% CI, or allele frequencies of BBJ, are shown on the *x*-axis, and those of

European GWAS are shown on the *y*-axis. Points are shown in red if new loci and in blue if previously reported only in MTAG but not in GWAS. BBJ, BioBank Japan; CI, confidence interval; GWAS, genome-wide association study; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; MTAG, multi-trait analysis of GWAS; NIHF; non-ischemic heart failure; QQ, quantile-quantile.

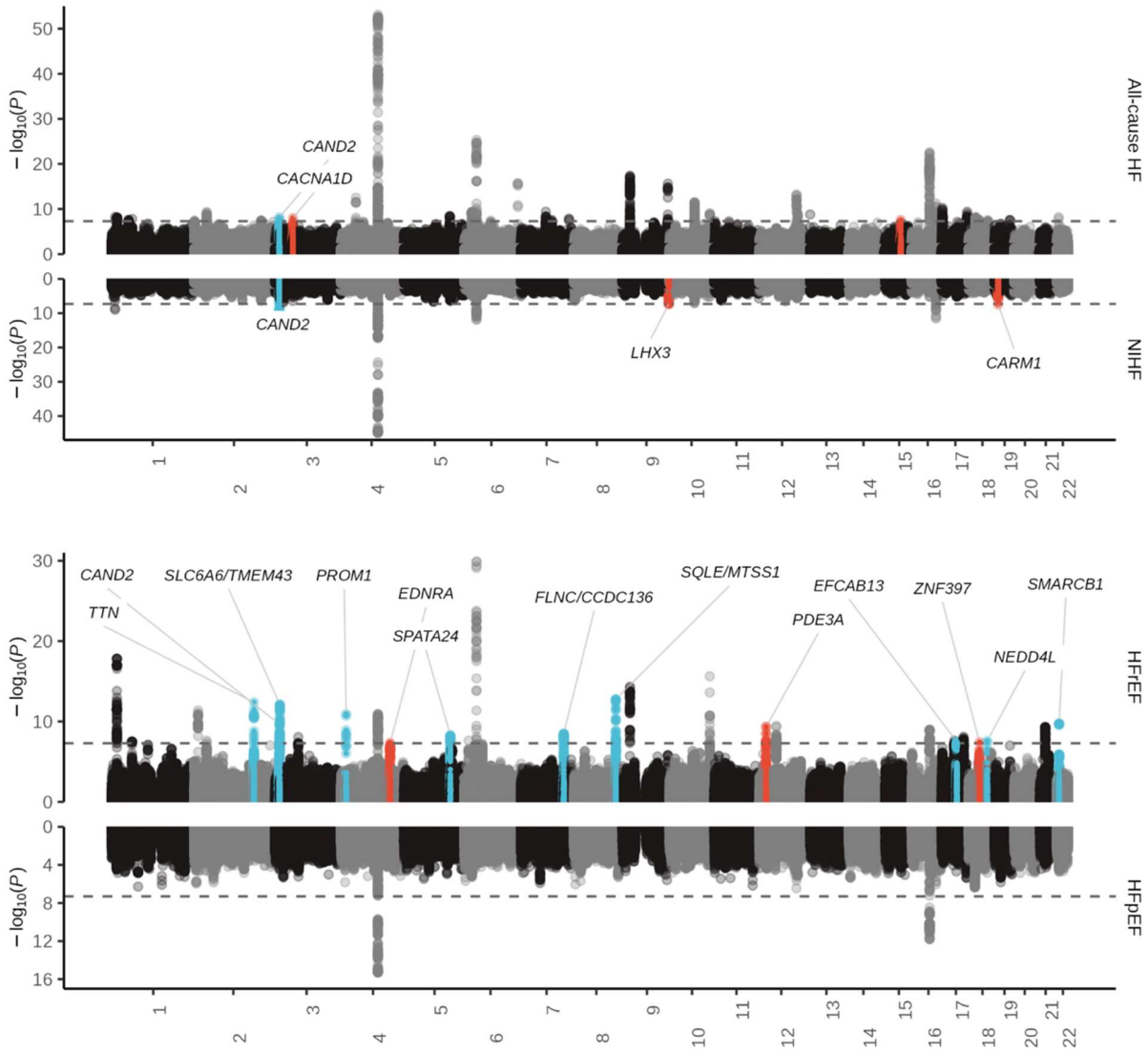


Supplementary Fig. 4 | Downstream analysis of Japanese GWAS. (a) Lead variants identified in our Japanese GWAS of each HF phenotype are searched in previously reported Japanese GWAS of cardiovascular-related phenotypes. The color indicates the Z value of the corresponding Japanese GWAS. **(b and c)** Lead variants identified by each HF phenotype Japanese GWAS are searched in GTEx eQTL **(b)** and sQTL **(c)**. The color indicates the Z value of the corresponding eQTL or sQTL. rs6596471, which was strongly correlated with lead variant chr5:138668504:T:C ($r^2 > 0.8$) is used as proxy variant since chr5:138668504:T:C was not found

in GTEx database. **(d)** Manhattan plots of H-MAGMA using aorta or adrenal gland Hi-C datasets. FDR was calculated from uncorrected *P* values using the Benjamini-Hochberg method. $-\log_{10}$(FDR) on the *y*-axis are shown against the genomic positions (hg19) on the *x*-axis. Genes that had not been identified in previous TWAS, proteome-wide MR, or gene analysis are shown in red. ** indicates novel loci, and * indicates loci previously reported only in MTAG and not in GWAS. eQTL, expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association study; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; H-MAGMA, Hi-C coupled MAGMA; LAA, left atrial appendage; MAGMA, Multi-marker Analysis of GenoMic Annotation; MTAG, multi-trait analysis of GWAS; NIHF, non-ischemic heart failure; sQTL, splicing quantitative trait loci.

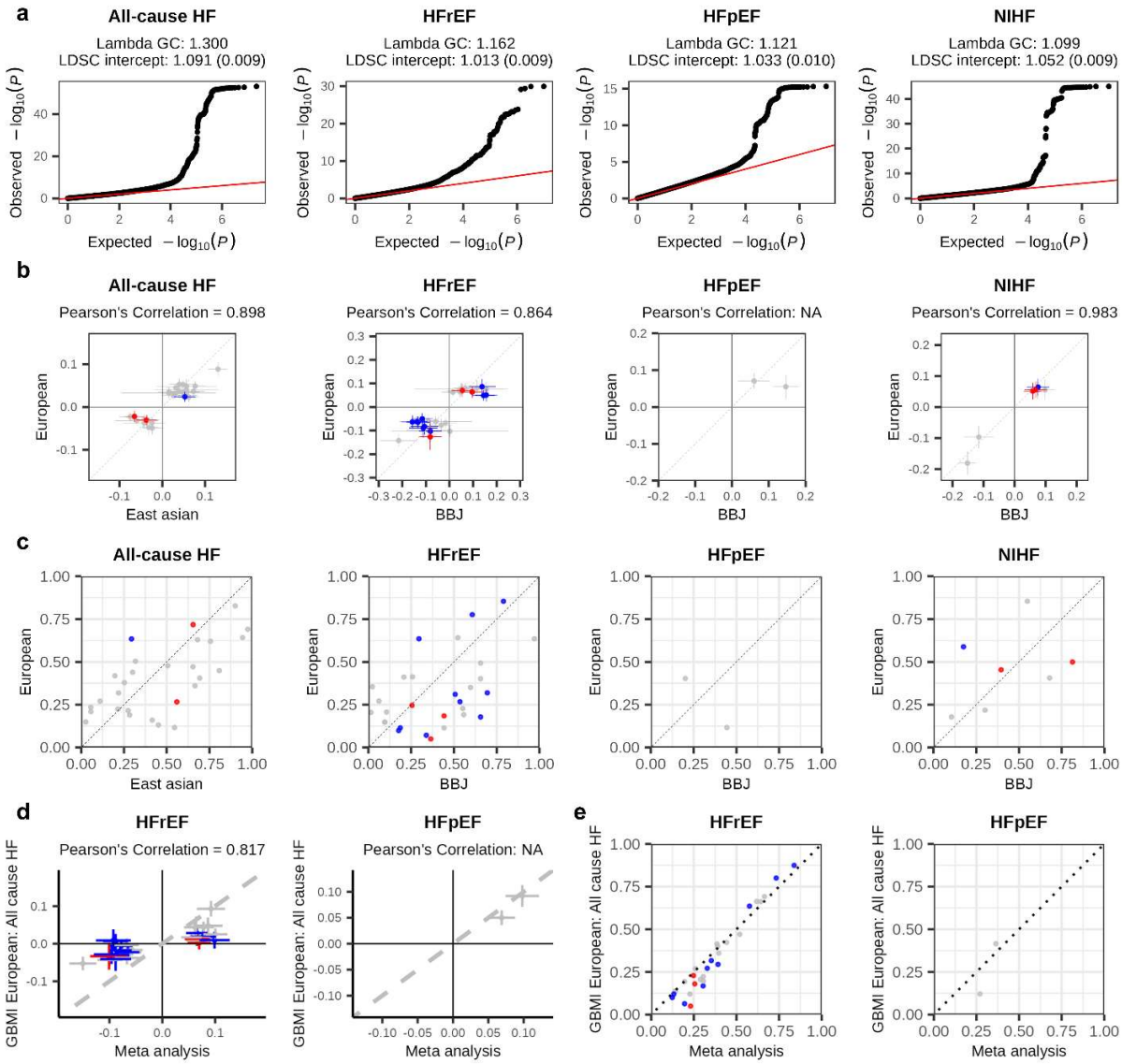


Supplementary Fig. 5 | Phenogram of genome-wide significant loci for HF and its subtypes. Indicated gene names are the prioritized gene when one gene is prioritized by our methods, the nearest gene when multiple genes are prioritized equally, and the blank when none is prioritized or the nearest gene is an RNA gene. The shapes indicate HF subtypes; circles, all-cause HF; rectangles, HFrEF; triangles, HFpEF; and diamonds, non-ischemic HF. Colors indicate HF and its subtypes; red, BioBank Japan (BBJ); gray, European (EUR); light green, trans-ancestry meta-analysis (META); and black, multi-trait analysis of GWAS (MTAG). Cytobands and annotations are based on GRCh37/hg19. Gene names were annotated if prioritized by at least 3 methods. HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction.



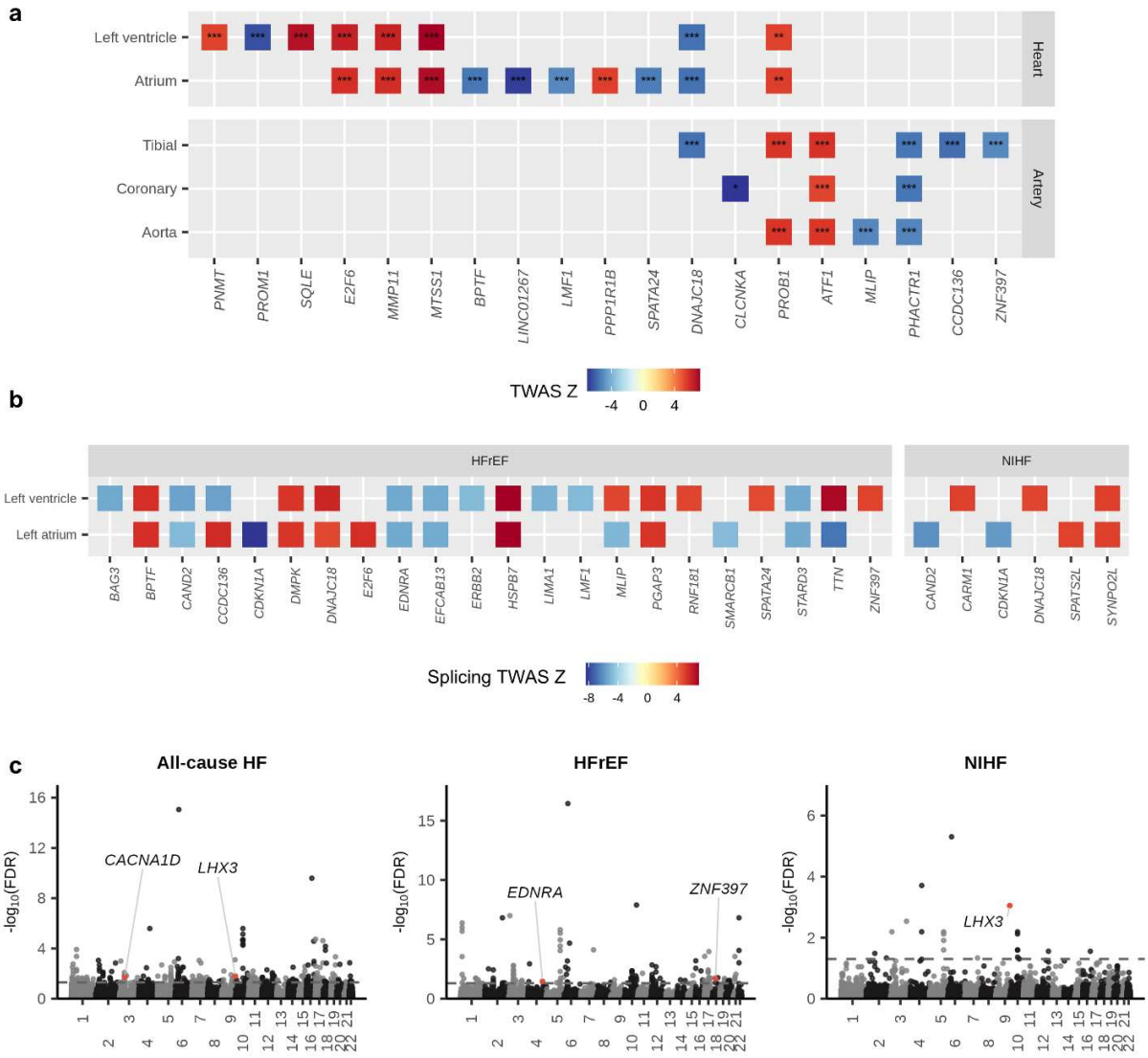
Supplementary Fig. 6 | Results of the cross-ancestry meta-analysis. Miami plot of trans-ancestry meta-analysis. -$\log_{10}$(*P* value) on the *y*-axis is shown against the genomic positions (hg19) on the *x*-axis. Two-sided uncorrected *P* values were calculated using a logistic regression model, and the genome-wide significance level was applied. Dashed lines represents the genome-wide significance threshold. Association signals that reached a genome-wide significance level ($P < 5 \times 10^{-8}$) are shown in red if new loci and in blue if previously reported only in MTAG but not in GWAS. GWAS, genome-wide association study; HF, heart failure;

HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; MTAG, multi-trait analysis of GWAS.



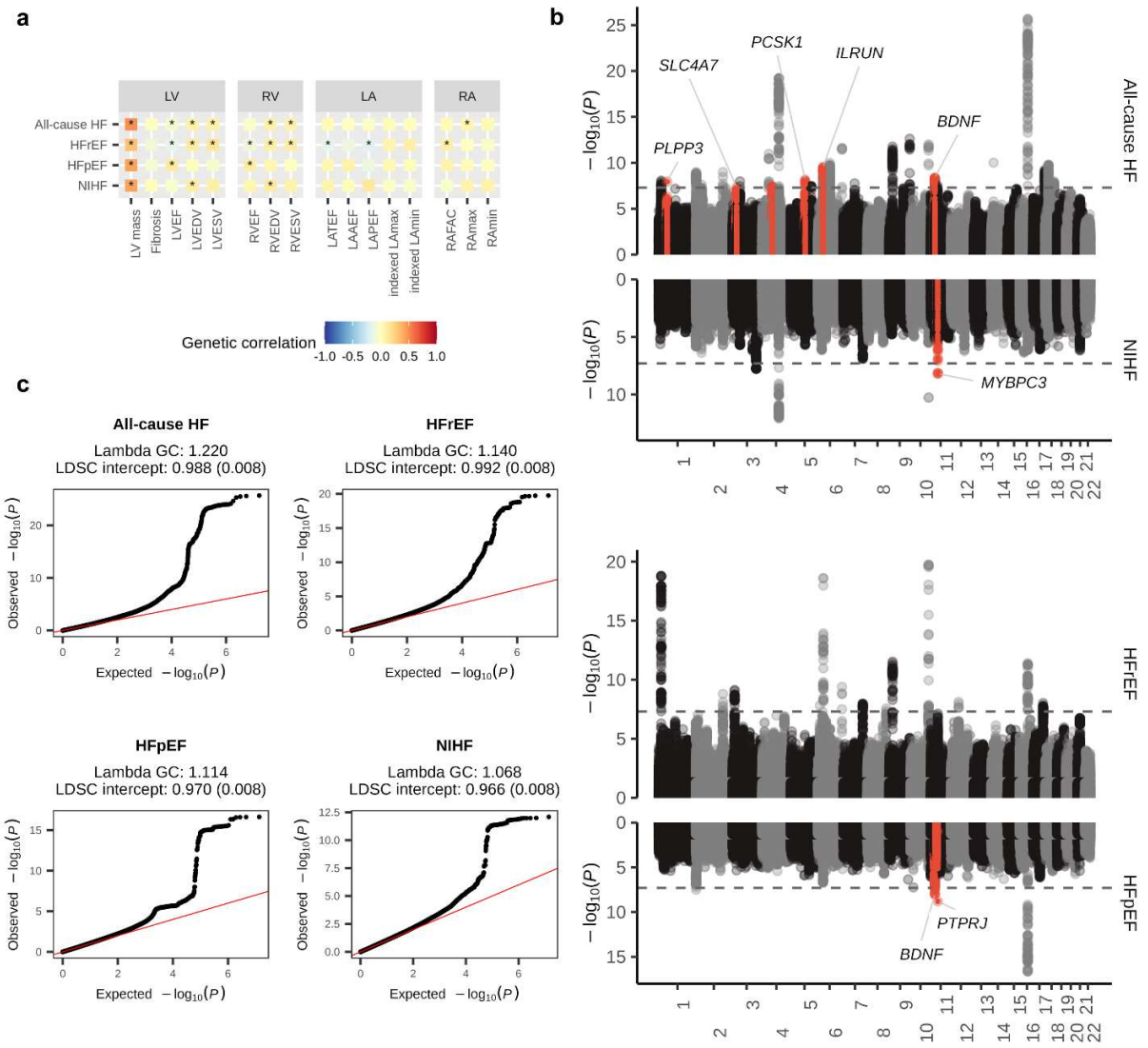
Supplementary Fig. 7 | QQ plot and comparison of allele frequencies and allelic effects identified in the cross-ancestry meta-analysis. (a) QQ plot for each HF phenotype in the cross-ancestry meta-analysis. Two-sided *P* values were calculated using a logistic regression model. **(b** and **c)** Comparison of allelic effects **(b)** and allele frequency **(c)** between East Asian GWAS and European GWAS. Effect size with its 95% CI or allele frequency of East Asian (for all-cause

HF) or BBJ (other HF phenotypes) GWAS are shown on the *x*-axis, and those of European are shown on the *y*-axis. **(d** and **e)** Comparison of allelic effects between the cross-ancestry meta-analysis (HFrEF or HFpEF) and non-overlapping European GWAS (GBMI European all-cause HF). Effect size with its 95% CI (**d**) or allele frequencies (**e**) of the cross-ancestry meta-analysis are shown on the *x*-axis, and those of European GWAS are shown on the *y*-axis. Points are shown in red if new loci and in blue if previously reported only in MTAG but not in GWAS. BBJ, BioBank Japan; CI, confidence interval; GBMI, Global Biobank Meta-analysis Initiative; GWAS, genome-wide association study; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; LDSC, Linkage disequilibrium score regression; MTAG, multi-trait analysis of GWAS; QQ, quantile-quantile.



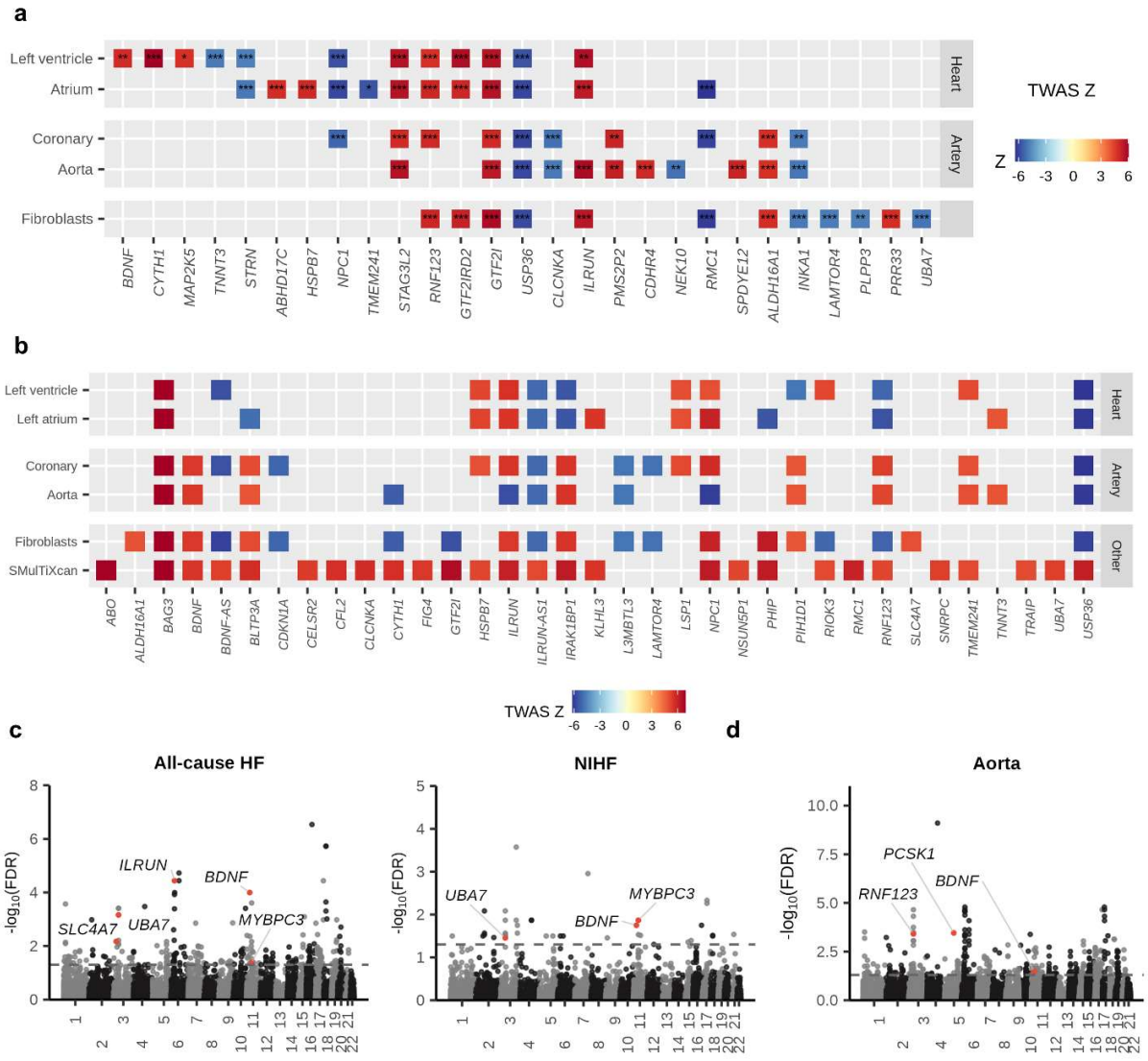
Supplementary Fig. 8 | Downstream analysis of the cross-ancestry meta-analysis. (a) Heatmap of the TWAS of HF (all-cause HF and HF subtypes) showing transcriptome-wide significant associations with supporting evidence from colocalization. Colored squares indicate TWAS significant associations based on two-sided *P* values after multiple testing corrections ($P < 1.55 \times 10^{-6}$). * indicate PP4 0.75-0.8, ** 0.8-0.9, *** 0.9-1.0. **(b)** Heatmap of the Splicing-TWAS HF showing splicing-wide significant associations. Colored squares splicing-TWAS significant associations based on two-sided *P* values after multiple testing corrections. Genes are

presented on the *x*-axis, and tissue types are on the *y*-axis in both **(a)** and **(b)**. **(c)** Manhattan plots of MAGMA for all-cause HF, HFrEF, and non-ischemic heart failure. FDR was calculated from uncorrected *P* values using the Benjamini-Hochberg method. $-\log_{10}$(FDR) on the *y*-axis are shown against the genomic positions (hg19) on the *x*-axis. Genes that have not been identified in previous TWAS, proteome-wide MR, or gene analysis are shown in red. FDR, false discovery rate; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; NIHF, non-ischemic heart failure; TWAS, transcriptome-wide association studies.



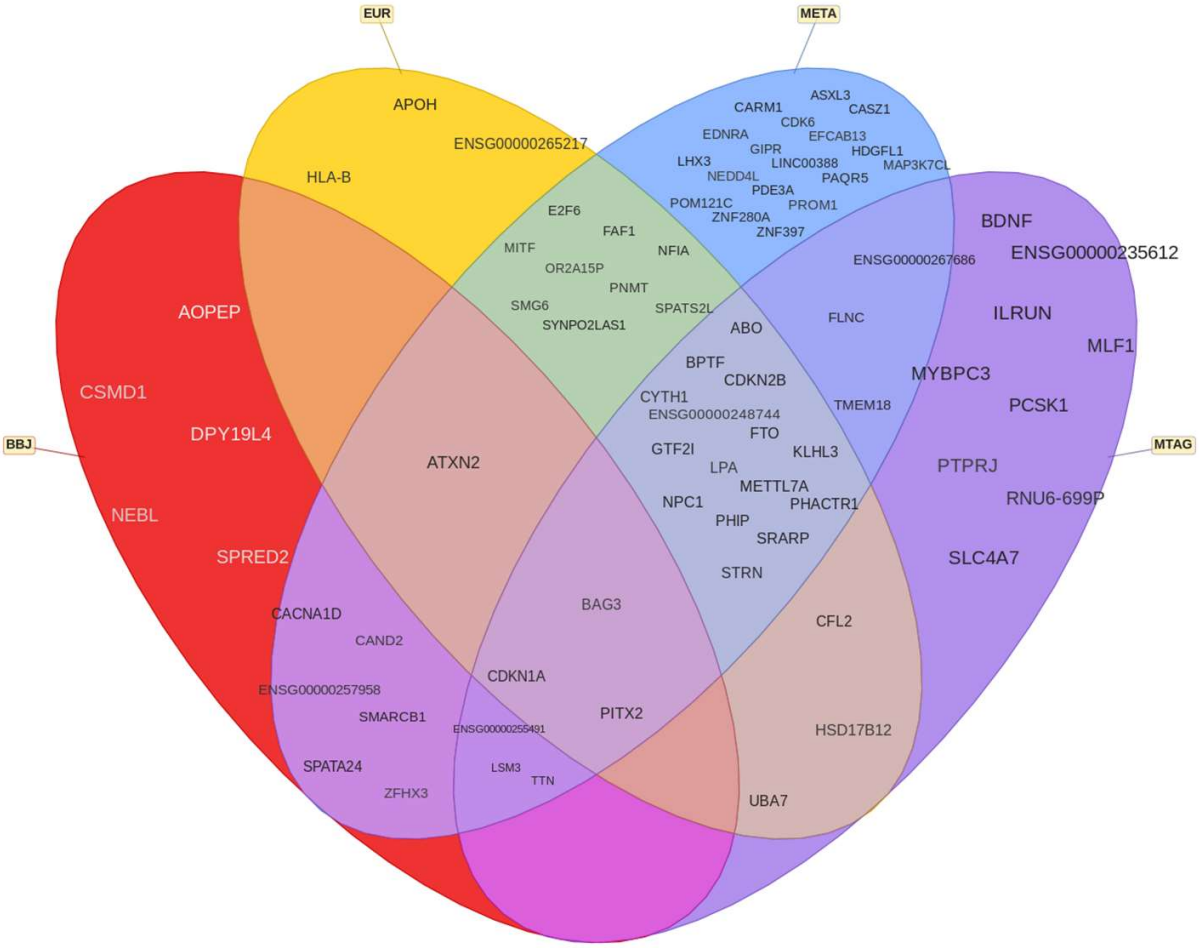
Supplementary Fig. 9 | Results of multi-trait analysis of GWAS. (a) Genetic correlation between HF and cardiac MRI-derived parameters. Significant genetic correlations (P <0.05) are marked with an asterisk. Two-sided P values were calculated using linkage disequilibrium score regression. **(b)** Miami plot of MTAG. Two-sided uncorrected P values were calculated using a logistic regression model, and the genome-wide significance level was applied. -log10(P value) on the y-axis are shown against the genomic positions (hg19) on the x-axis. Association signals that reached a genome-wide significance level ($P < 5 \times 10^{-8}$) are shown in red if new loci. **(c)** QQ

plot for each HF phenotype in MTAG. GWAS, genome-wide association study; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LDSC, Linkage disequilibrium score regression; MTAG, multi-trait analysis of GWAS; NIHF, non-ischemic heart failure; QQ, quantile-quantile.

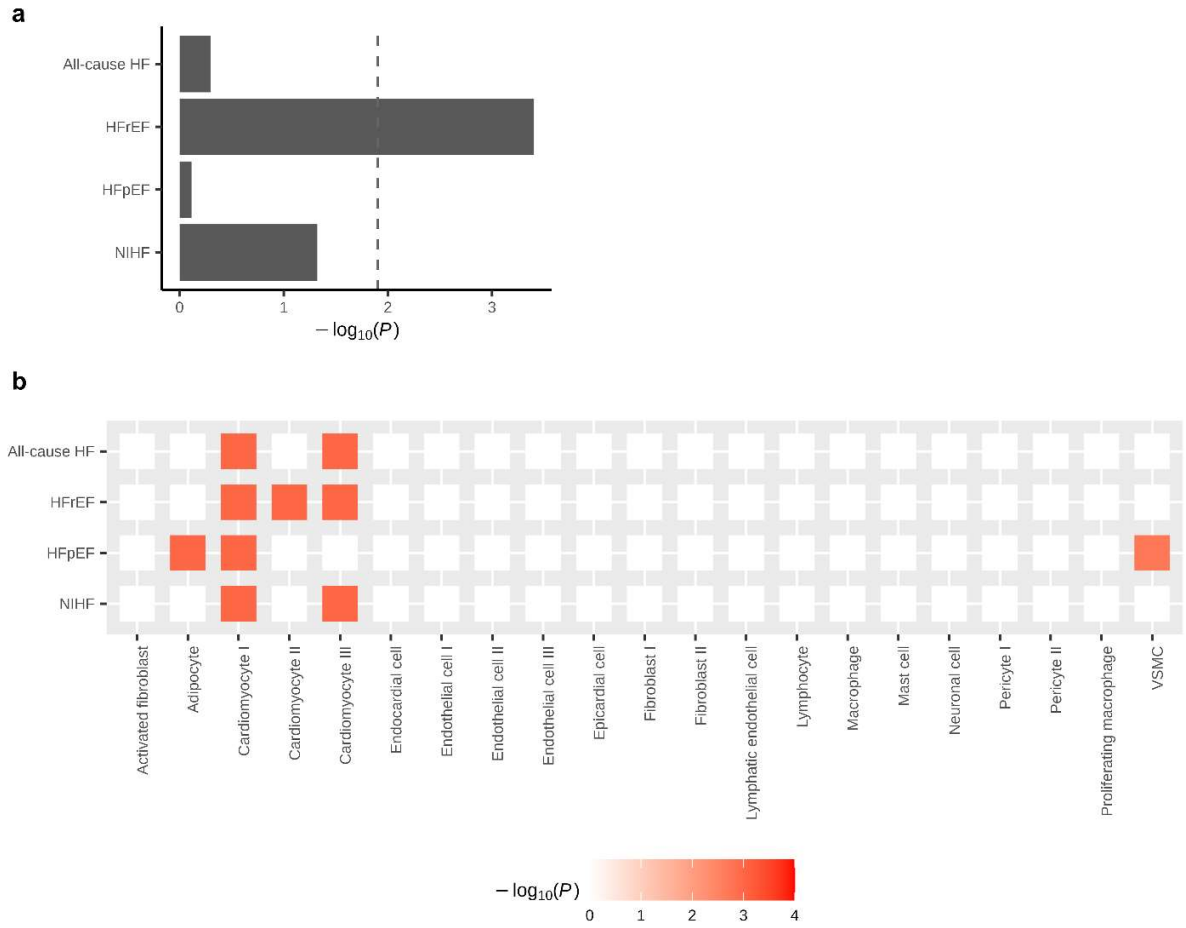


Supplementary Fig. 10 | Downstream analysis of multi-trait analysis of GWAS. (a) Heatmap of the TWAS showing transcriptome-wide significant associations with supporting evidence from colocalization. Colored squares indicate TWAS significant associations based on two-sided *P* values after multiple testing corrections ($P < 1.55 \times 10^{-6}$). * PP4 0.75-0.8, ** 0.8-0.9, *** 0.9-1.0. **(b)** Heatmap of the Splicing-TWAS HF showing splicing-wide significant associations. Colored squares indicate splicing-TWAS significant associations based on two-sided *P* values after multiple testing corrections. Genes are presented on the *x*-axis, and tissue types are on the *y*-

axis in both **(a)** and **(b)**. **(c and d)** Manhattan plots of MAGMA for all-cause HF or non-ischemic heart failure **(c)**, and H-MAGMA using aorta Hi-C datasets in all-cause HF**(d)**. FDR was calculated from uncorrected *P* values using the Benjamini-Hochberg method. $-\log_{10}$(FDR) on the *y*-axis is shown against the genomic positions (hg19) on the *x*-axis. Genes that have not been identified in previous TWAS, proteome-wide MR, or gene analysis are shown in red. FDR, false discovery rate; HF, heart failure; NIHF, non-ischemic heart failure; TWAS, transcriptome-wide association studies.



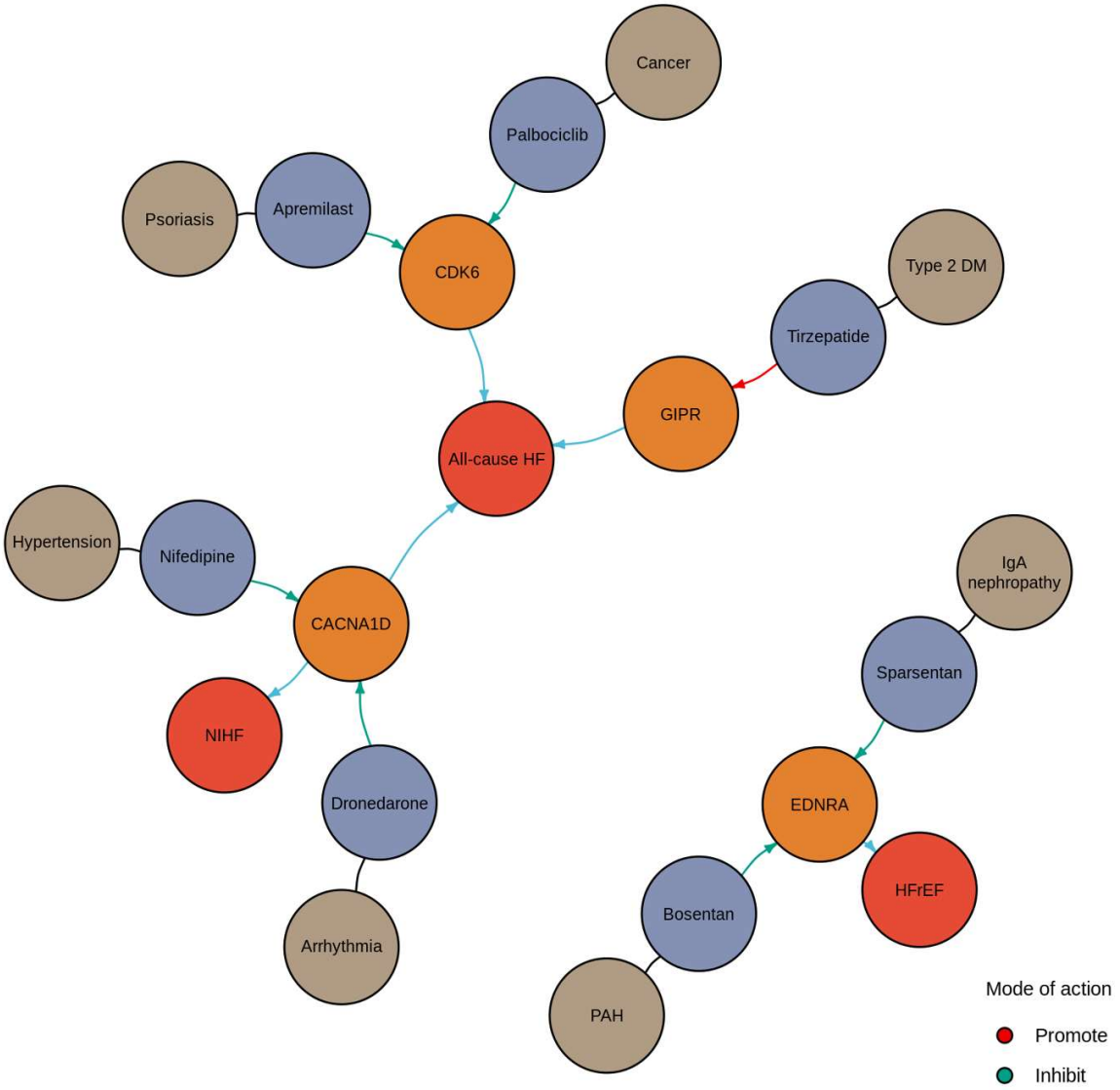
Supplementary Fig. 11 | Venn diagram for genes identified by each analysis. BBJ, BioBank Japan; EUR, European; META, meta-analysis; MTAG, multi-trait analysis of GWAS.



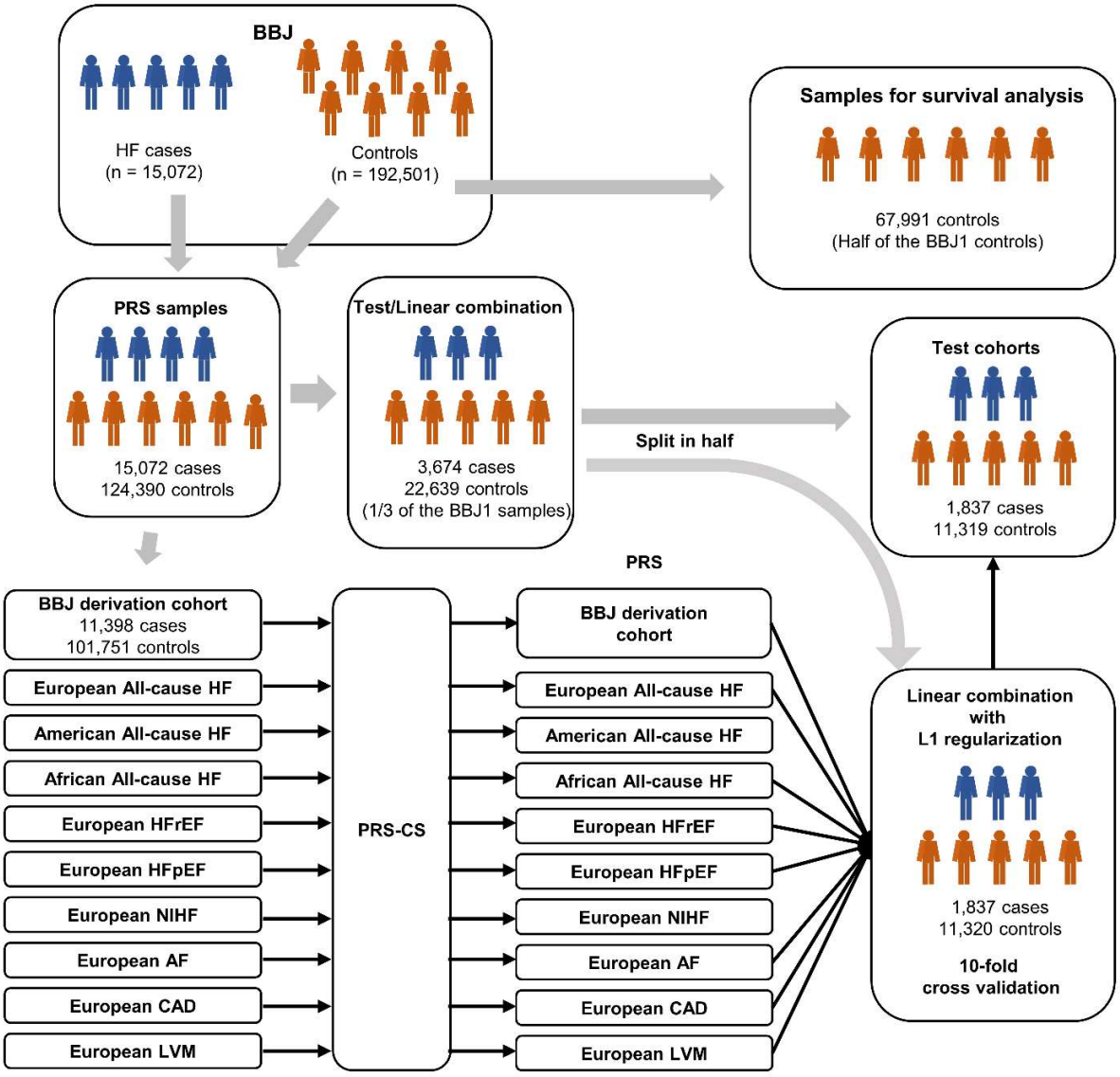
Supplementary Fig. 12 | miRNA enrichment analysis and single-cell enrichment analysis.

(a) MIGWAS results of the four HF phenotypes. The overall contribution of the miRNA-target gene network to the traits through the tissue-naïve approach. Two-sided uncorrected *P* values were calculated by the MIGWAS software. An enrichment signal is shown by $-\log_{10}(P_{\mathrm{MIGWAS}})$.

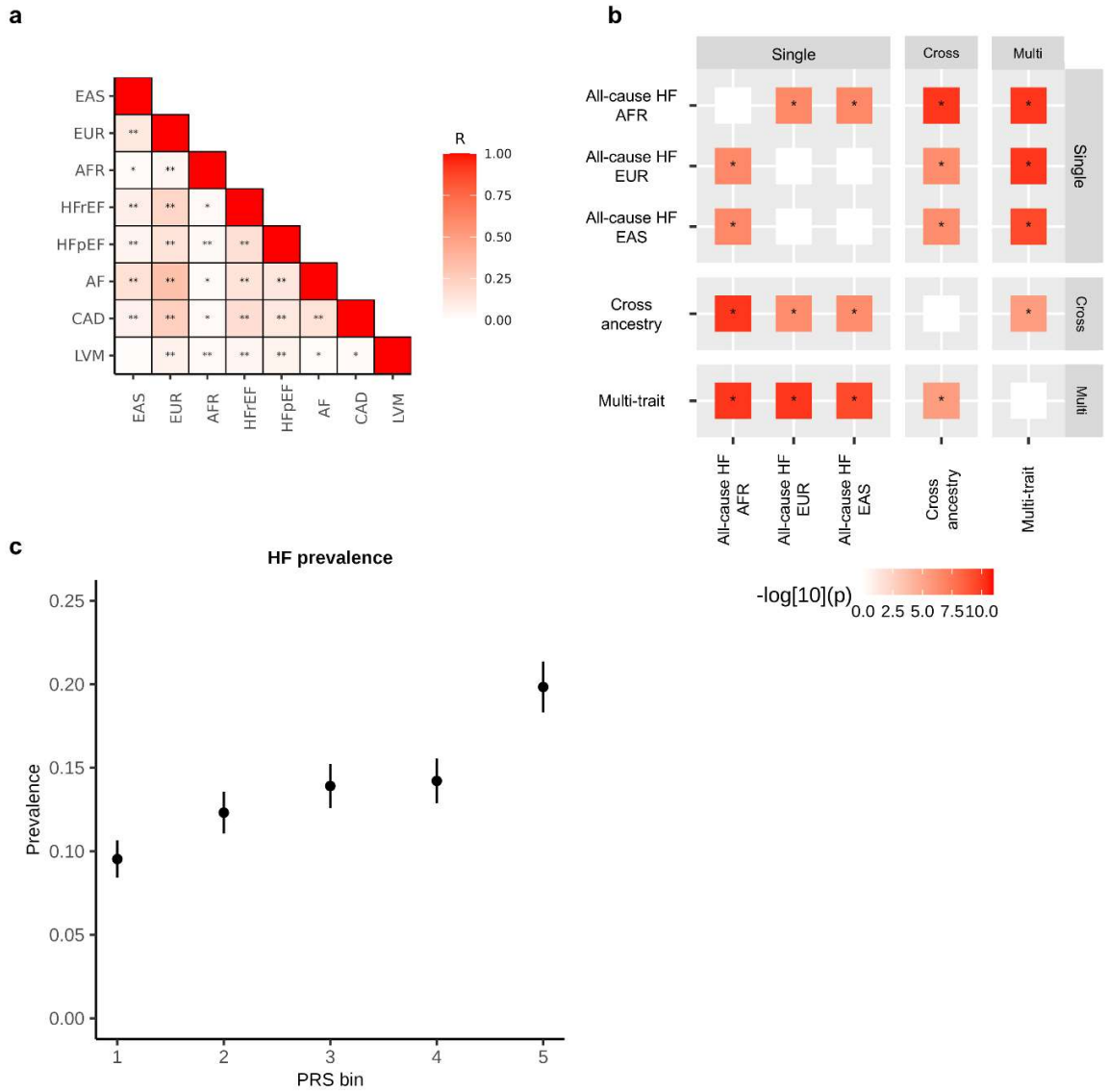
(b) Heatmaps depicting each cell type-disease association for HF phenotypes. Heatmap colors denote the two-sided uncorrected *P* value of cell-type-disease association evaluated using scDRS. Only significant associations ($P < 2.4 \times 10^{-3}$) are shown. HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; MIGWAS, miRNA–target gene networks; NIHF, non-ischemic heart failure; scDRS, single-cell disease relevance score; VSMC, Vascular smooth muscle cell.



Supplementary Fig. 13 | Candidate drugs linked to disease susceptibility loci for HF. Dark blue indicates HF subtypes; orange gene; purple medications; and brown diseases. Blue color indicates associations between genes and diseases. DM, diabetes mellitus; HF, heart failure; HFrEF, heart failure with reduced ejection fraction; NIHF, non-ischemic heart failure; PAH, pulmonary arterial hypertension.



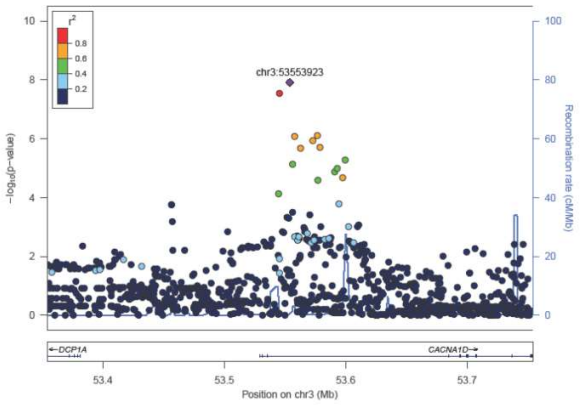
Supplementary Fig. 14 | Analytical scheme for PRS development. Schematic representation for derivation, cross-validation, performance testing of HF-PRS in the independent test cohort, and survival analysis for HF-PRS. AF, atrial fibrillation; BBJ, BioBank Japan; CAD, coronary artery disease; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LVM, left ventricular mass; NIHF, non-ischemic heart failure; PRS, polygenic risk score.

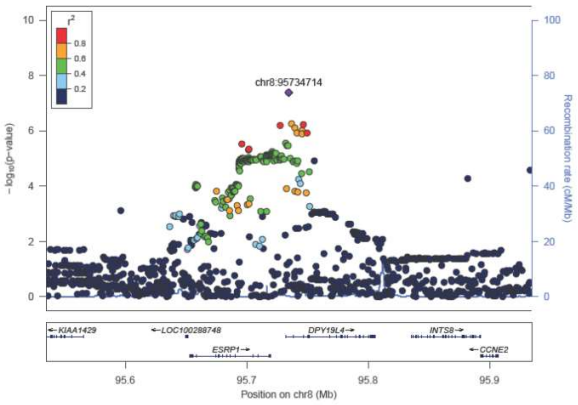


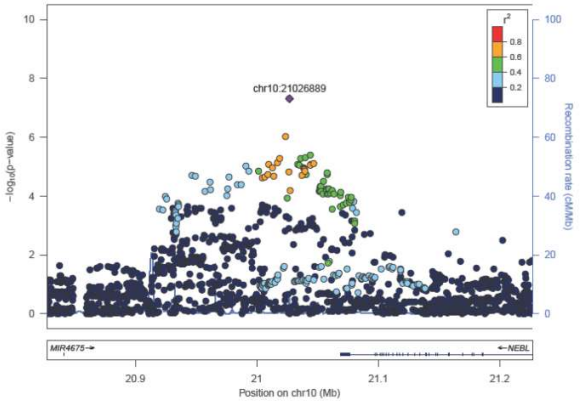
Supplementary Fig. 15 | PRS performance. **(a)** Correlation between PRS. Two-sided uncorrected *P* values were calculated using the Pearson's correlation. * $P < 0.05$, ** $P < 1.79 \times 10^{-3}$ (0.05/24). **(b)** Pairwise comparison of PRS performance. Distribution of the pairwise difference of Nagelkerke's pseudo-R^2 (Δ pseudo-R^2: Pseudo R^2_{Score} Y – Pseudo R^2_{Score} X, X, and Y are found at the axis) between each pair of PRS models. The distributions were obtained by bootstrapping 5.0×10^4 times. Two-sided bootstrap *P* values were calculated by counting the number of Δ pseudo-$R^2 \leq 0$ or Δ pseudo-$R^2 > 0$ and then multiplying the lower value by the

minimum estimated *P* value ($2 * 1 / (5.0 \times 10^4) = 4 \times 10^{-5}$: two-sided). The significance was set at $P = 5.0 \times 10{-}3$ (0.05/10). **(c)** PRS distribution and HF prevalence. Prevalence of HF based on the HF-PRS deciles in each combination of GWAS. The number of individuals in each decile is 2,631-2,632. Data are presented as medians and 95% CI. AF, atrial fibrillation; AFR, African; CAD, coronary artery disease; CI, confidence interval; Cross, cross ancestry PRS; EAS, East Asian; EUR, European; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LVM, left ventricular mass; Multi, multi-trait PRS; PRS, polygenic risk score; Single, single ancestry PRS.

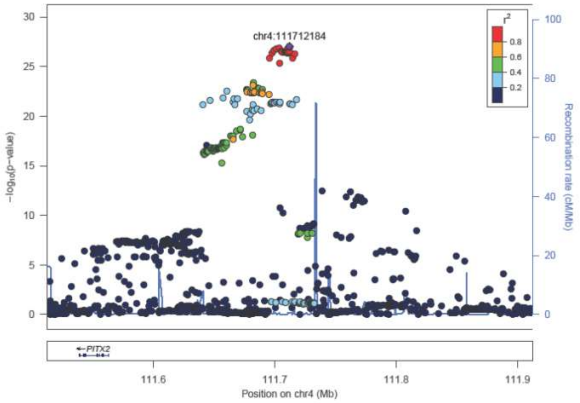
All cause heart failure (Biobank Japan): Novel loci

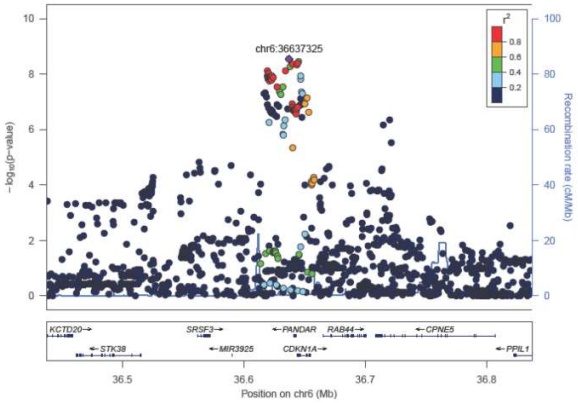




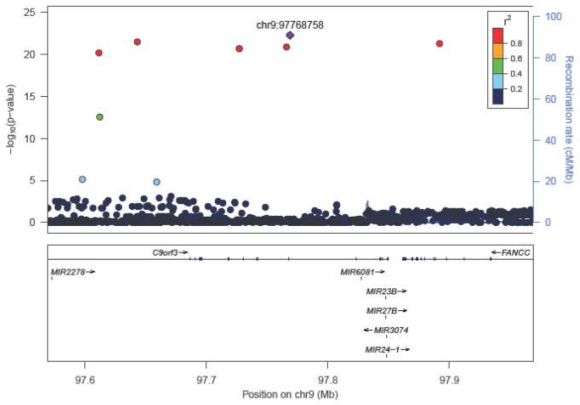


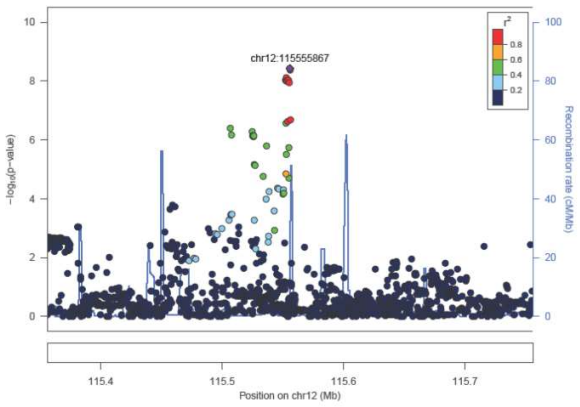
All cause heart failure (Biobank Japan): Known loci

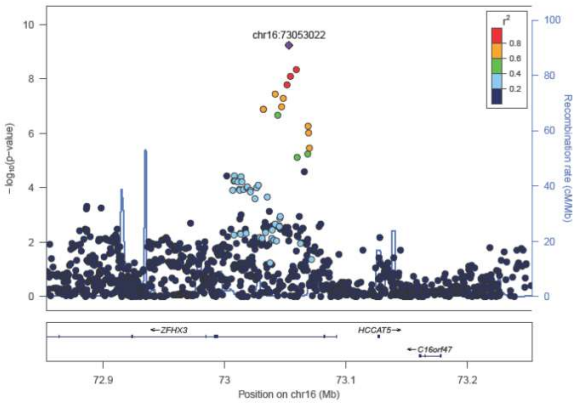




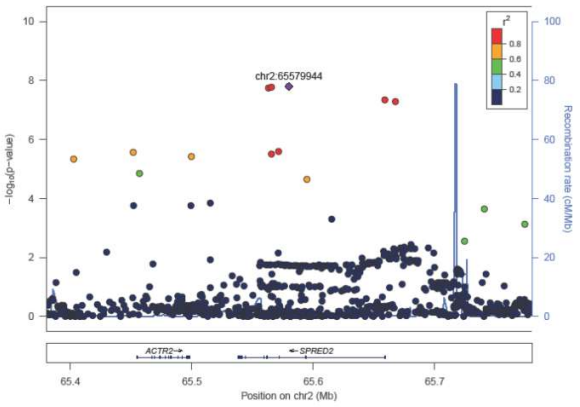
All cause heart failure (Biobank Japan): Known loci



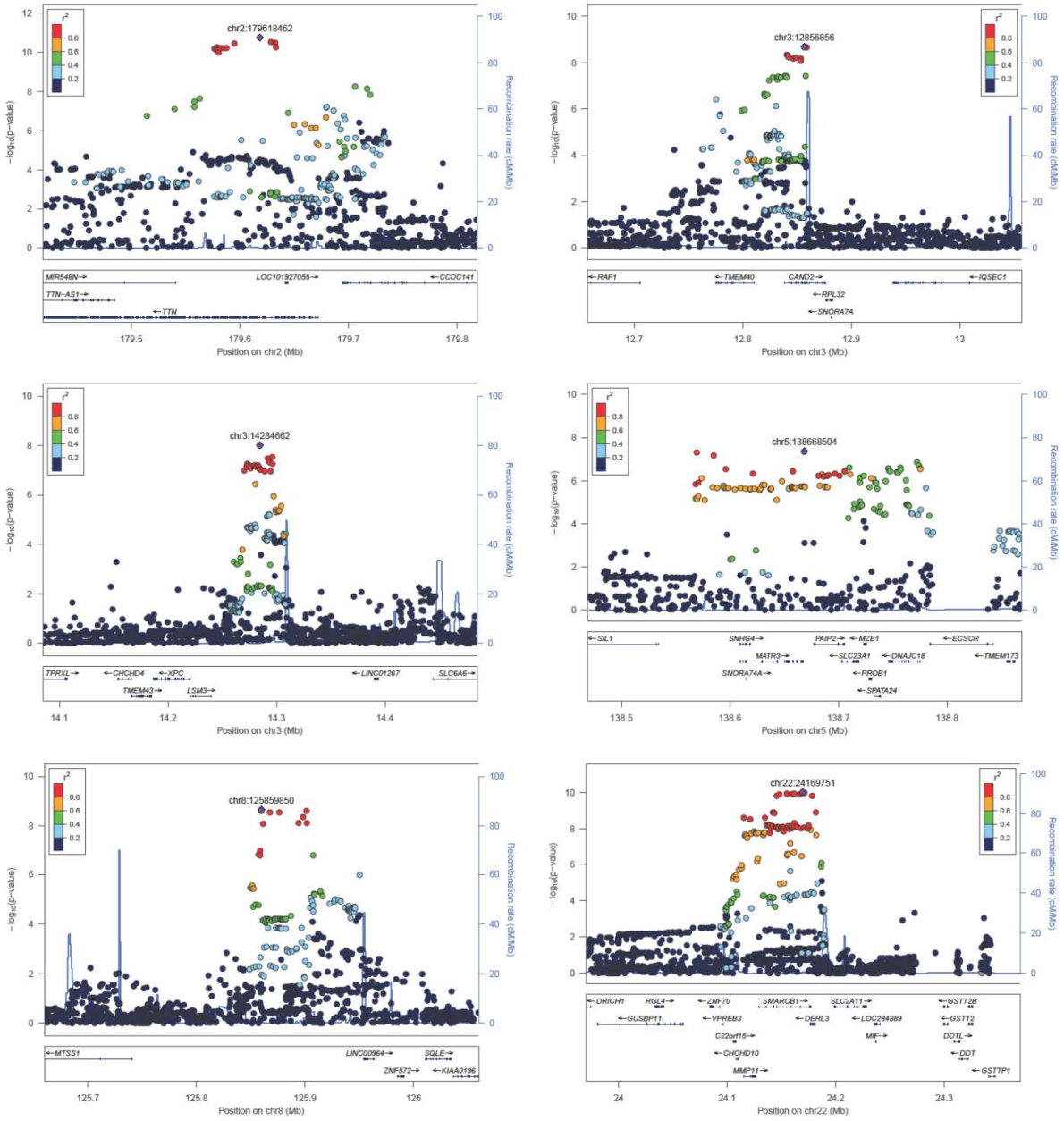




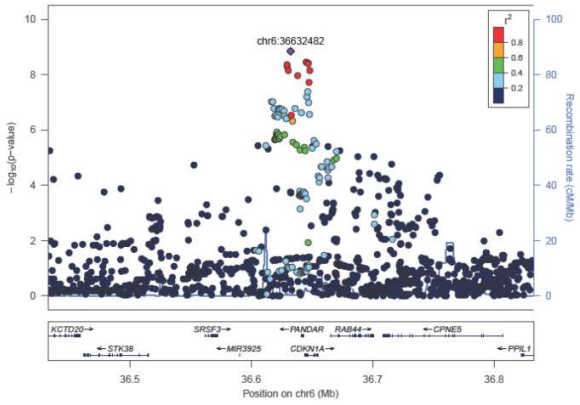
HFrEF (Biobank Japan): Novel loci

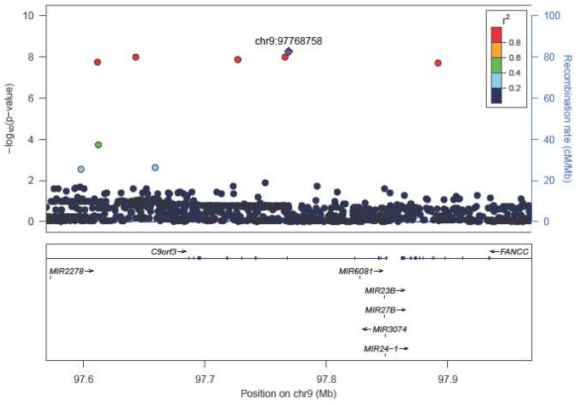


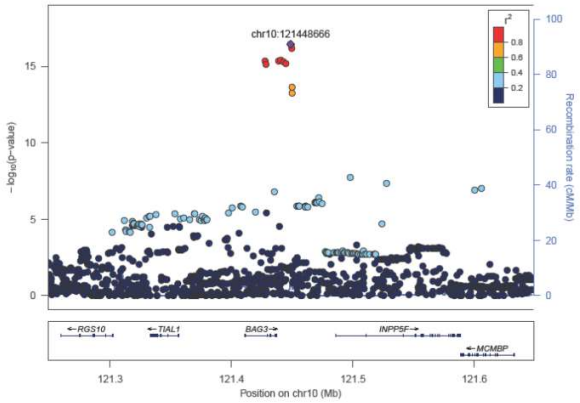
HFrEF (Biobank Japan): Previously identified only in MTAG



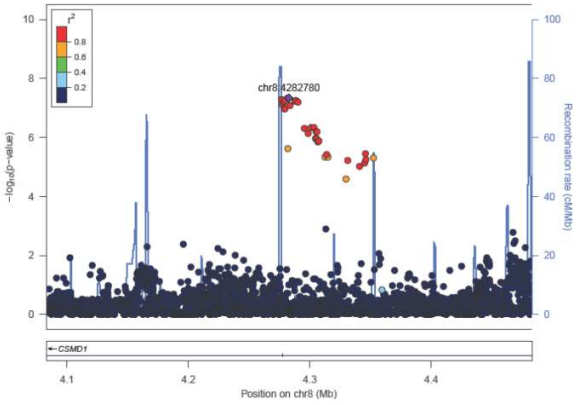
HFrEF (Biobank Japan): Known loci



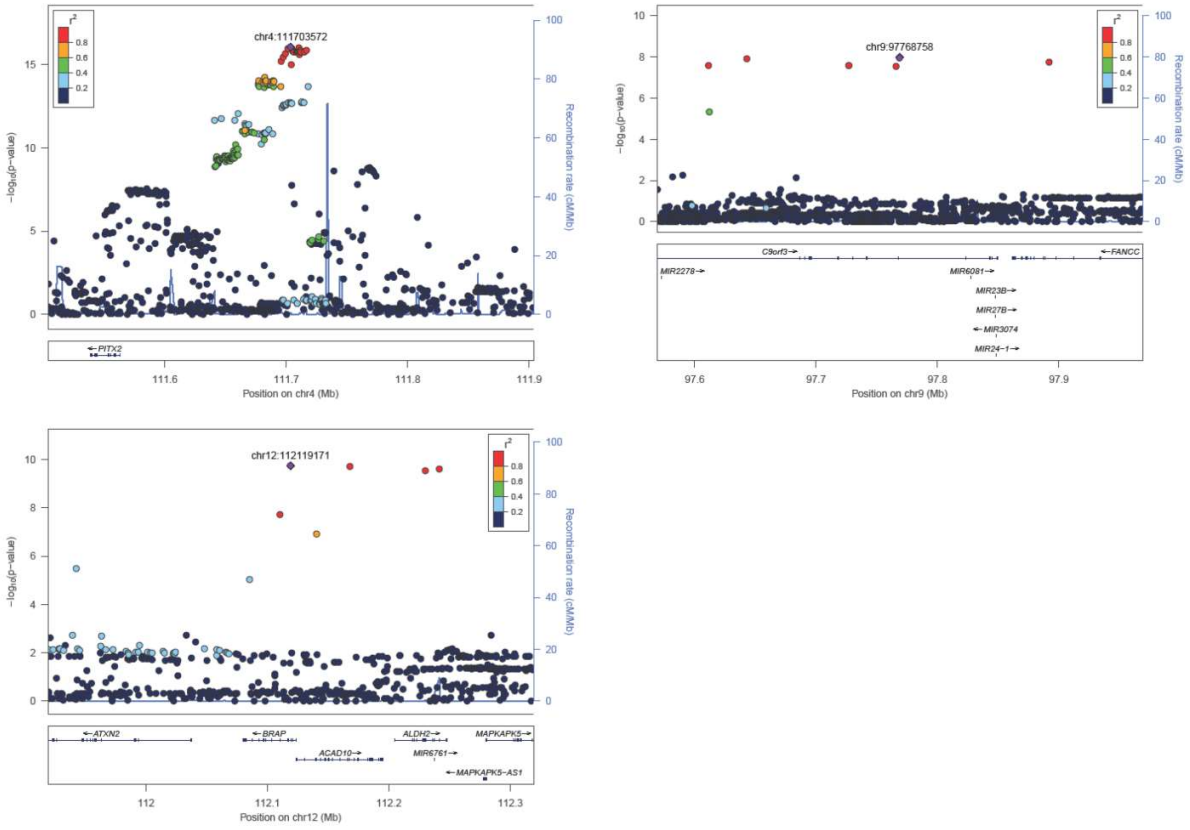




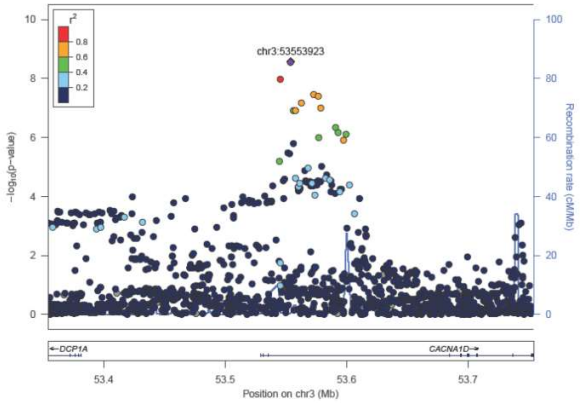
HFpEF (Biobank Japan): Novel loci

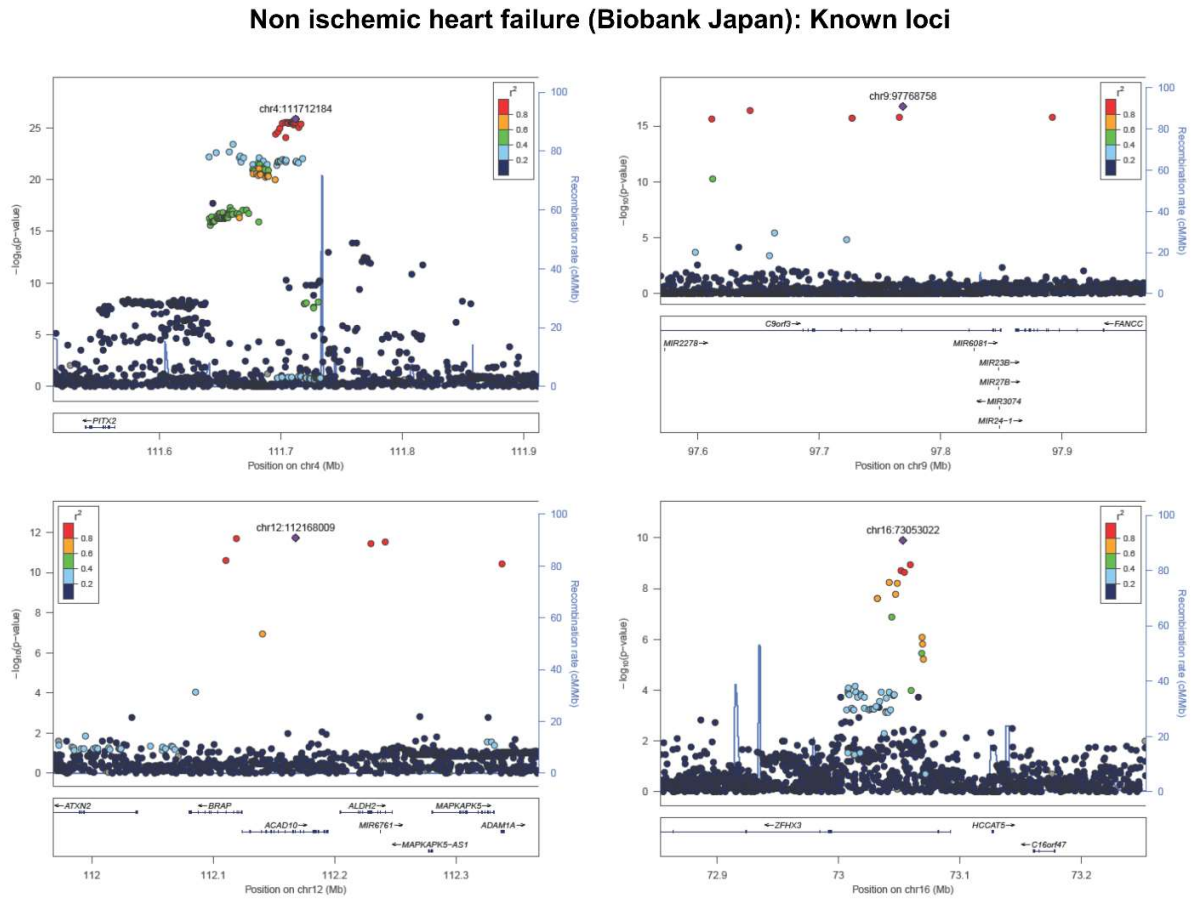


HFpEF (Biobank Japan): Known loci



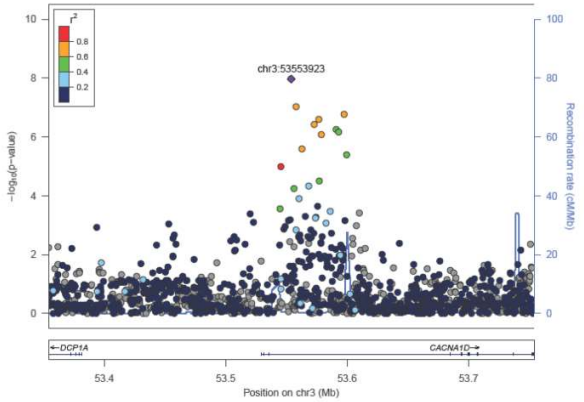
Non ischemic heart failure (Biobank Japan): Novel loci

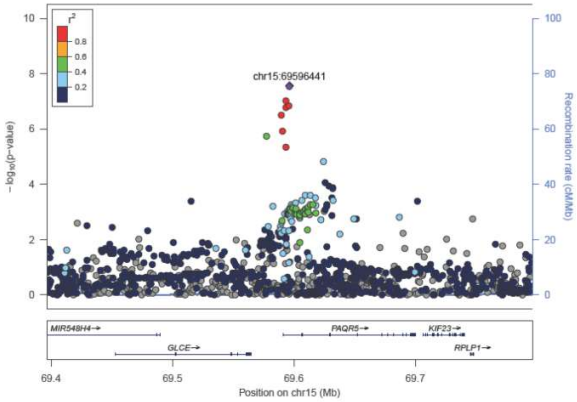




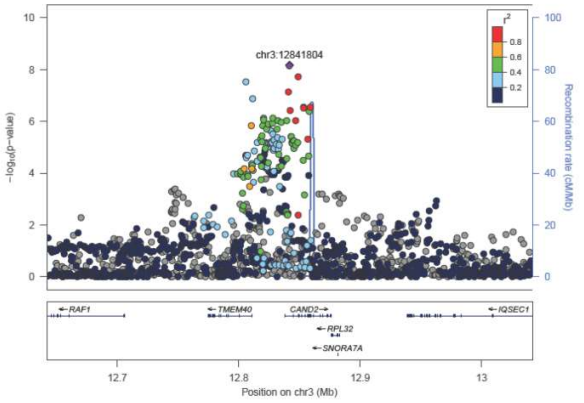
Supplementary Fig. 16 | LocusZoom plots for Japanese GWAS. -$\log_{10}$(*P* value) on the *y*-axis is shown against the genomic positions (hg19) on the *x*-axis. GWAS, genome-wide association study.

All cause heart failure (meta analysis): Novel loci

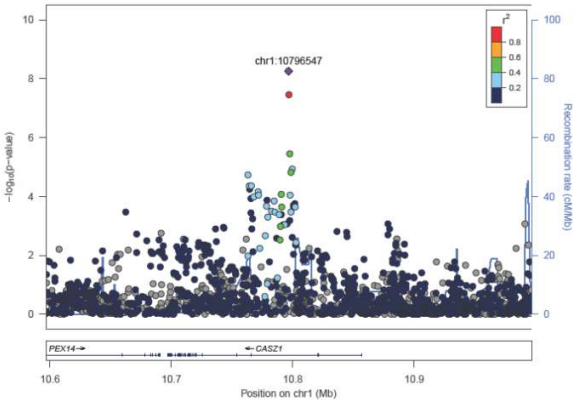


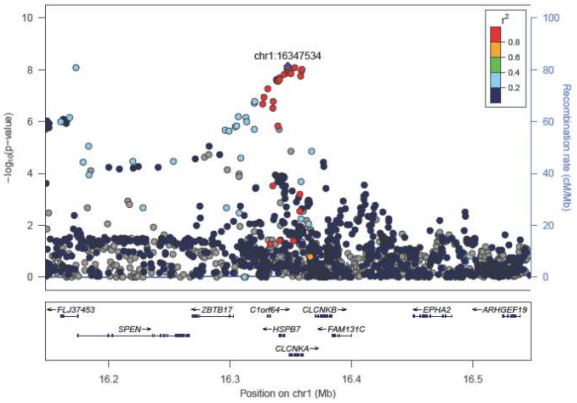


All cause heart failure (meta analysis): Previously identified only in MTAG

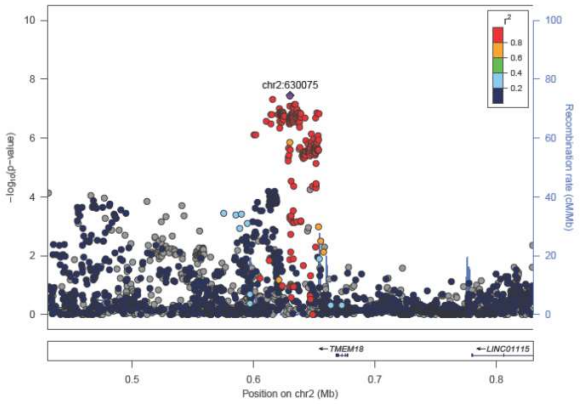


All cause heart failure (meta analysis):Known loci

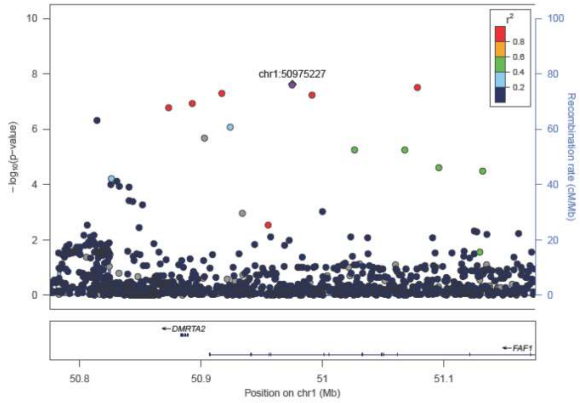


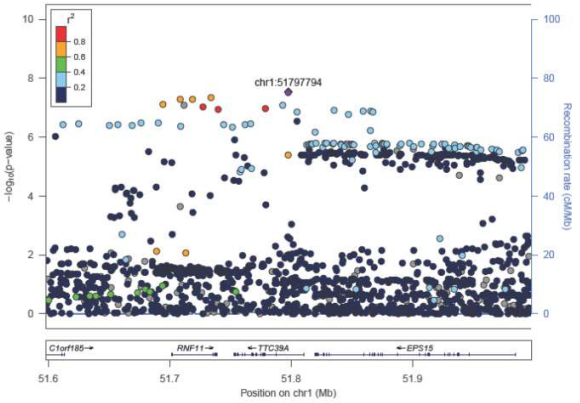


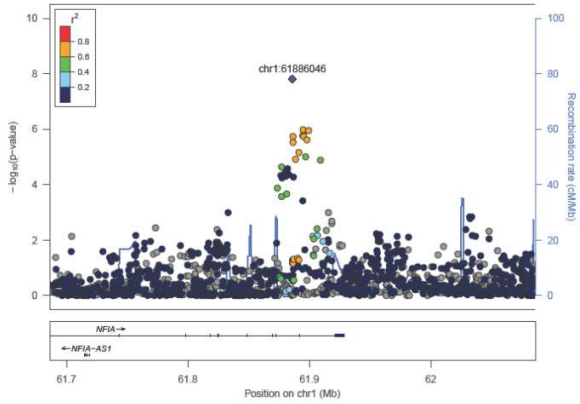
All cause heart failure (meta analysis):Known loci

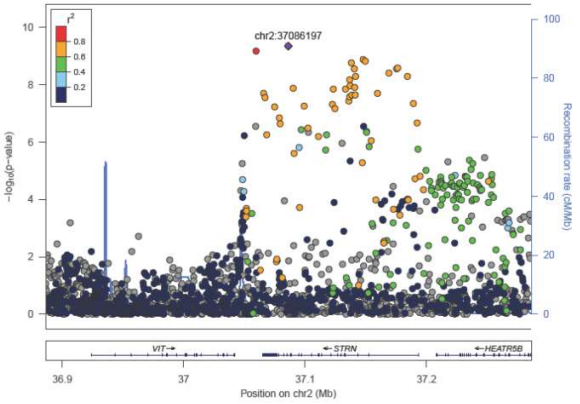


All cause heart failure (meta analysis):Known loci

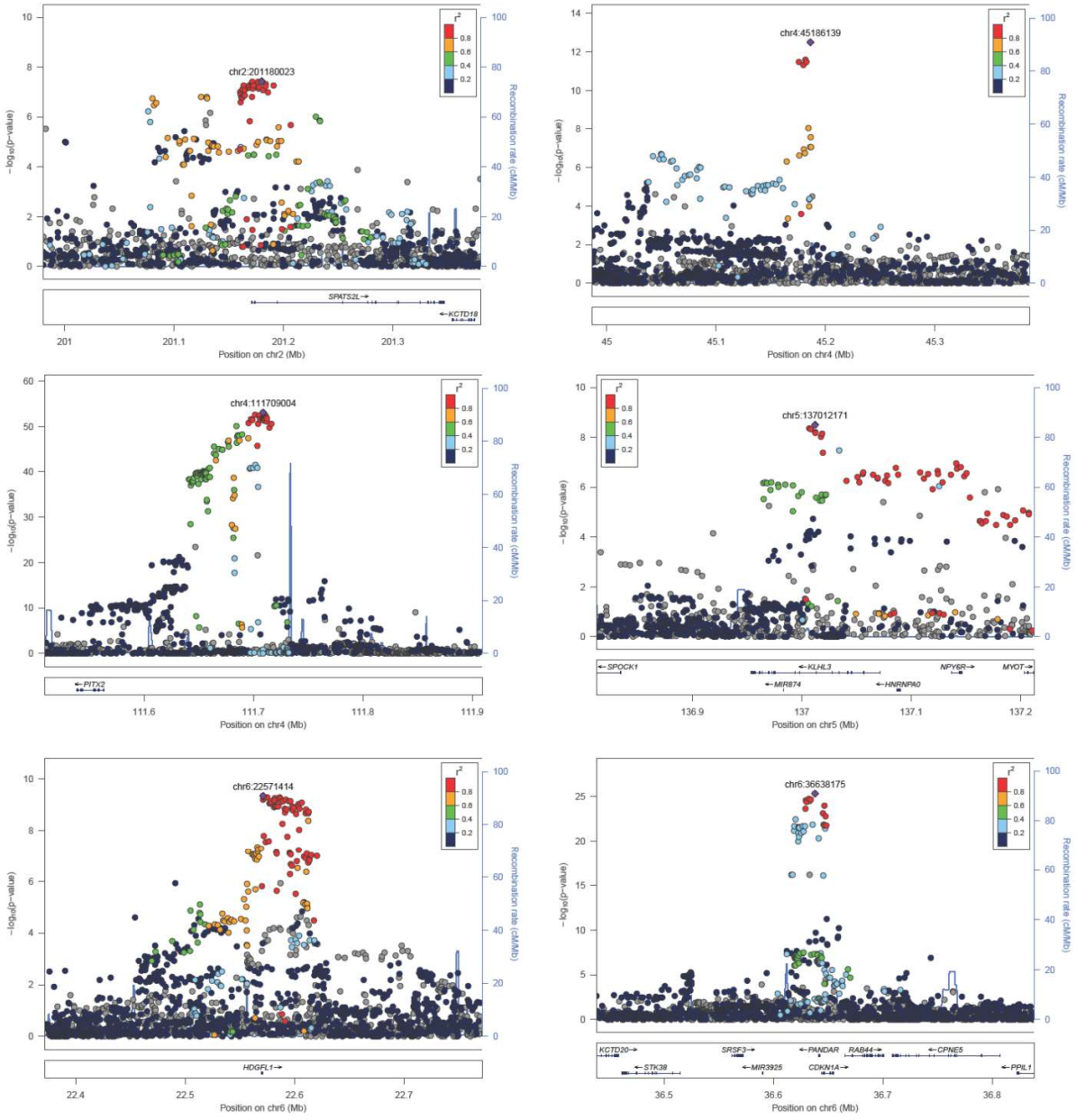




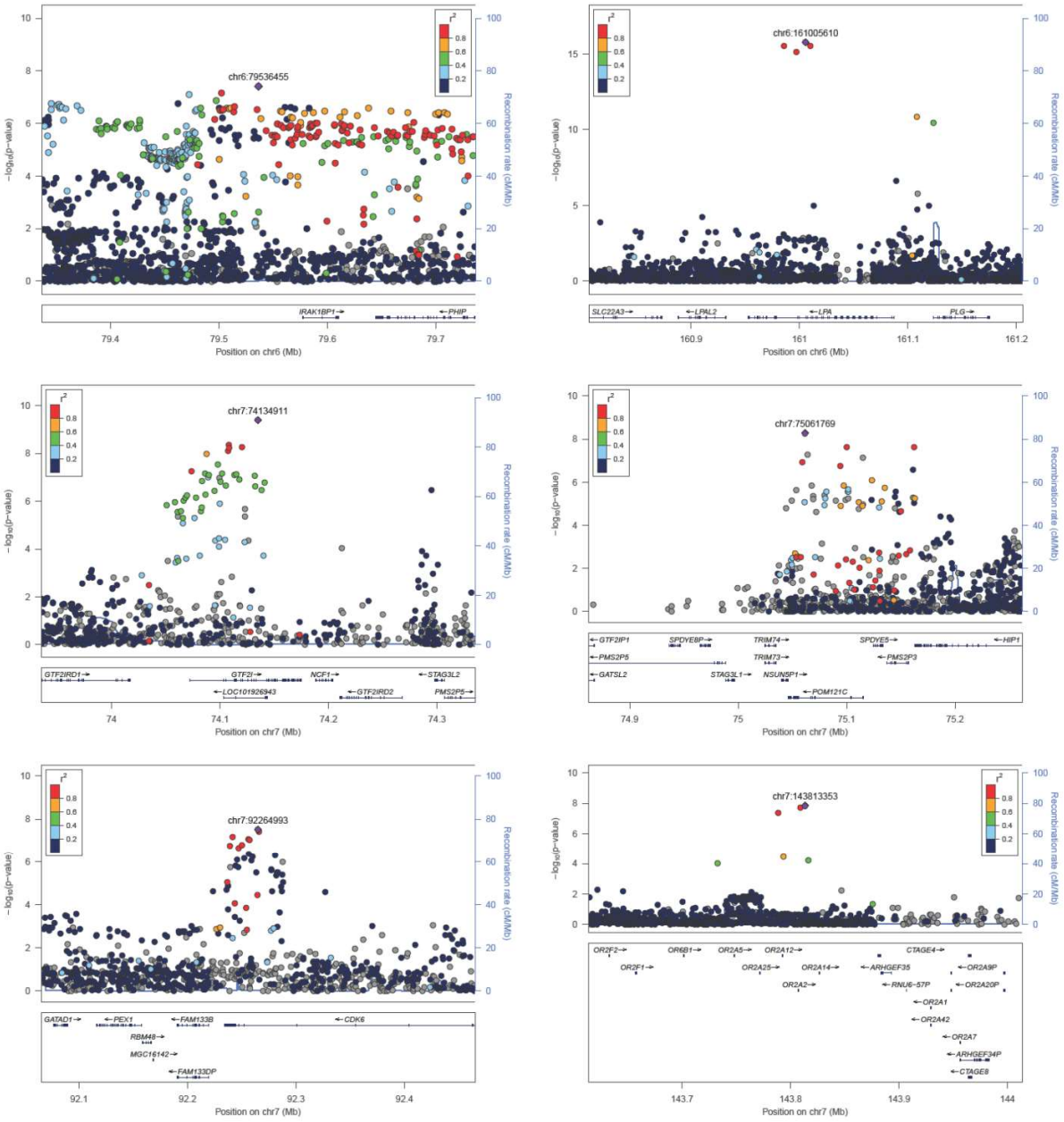




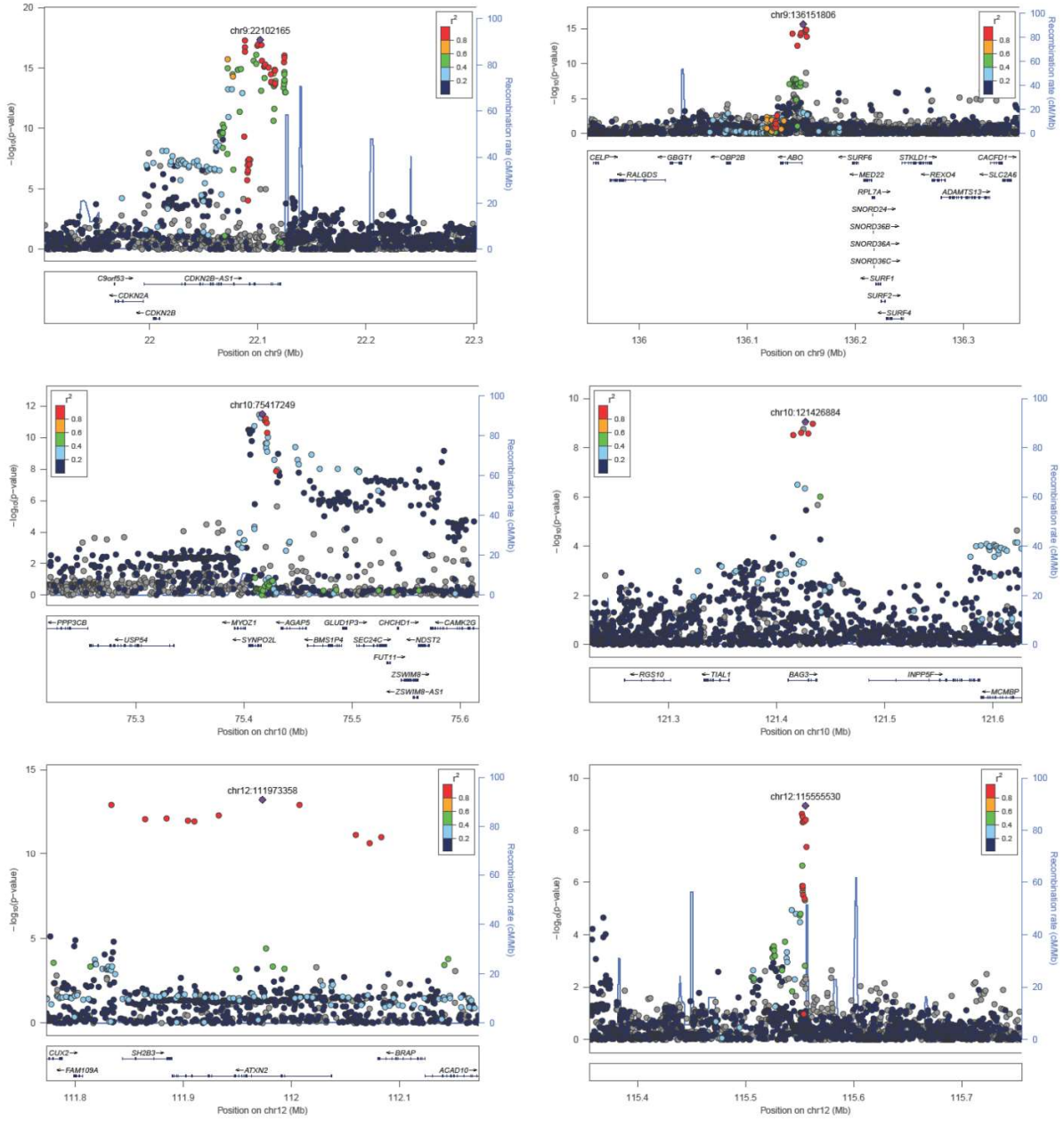
All cause heart failure (meta analysis):Known loci



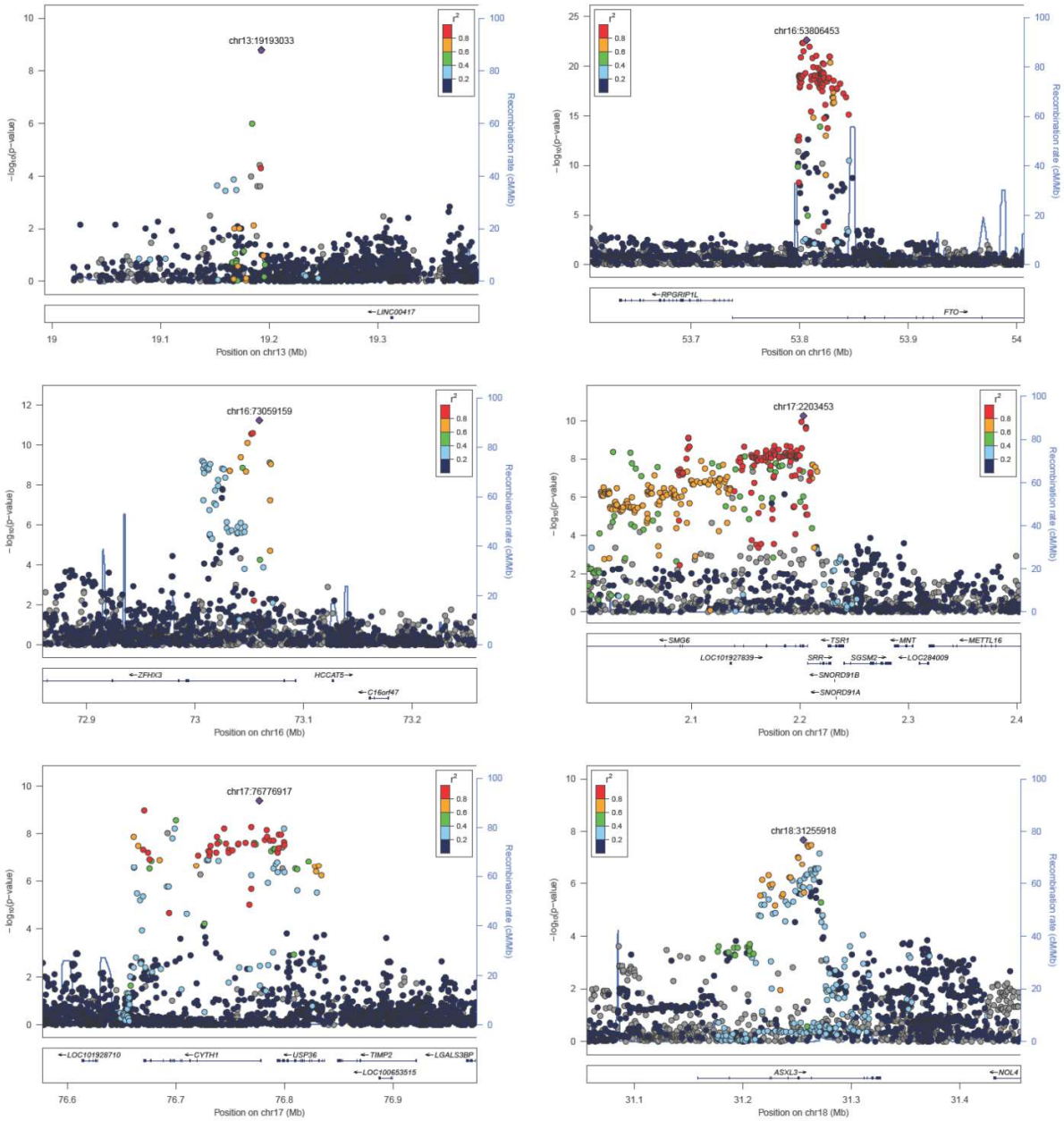
All cause heart failure (meta analysis):Known loci



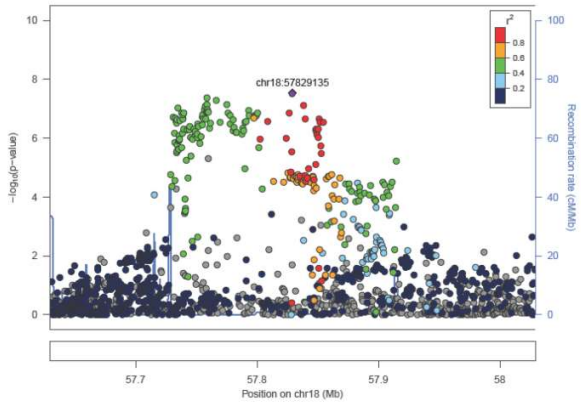
All cause heart failure (meta analysis):Known loci

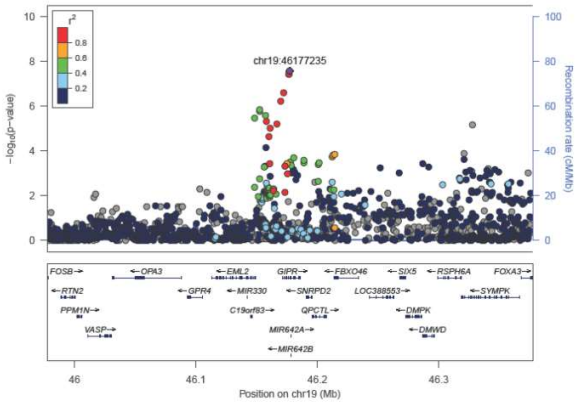


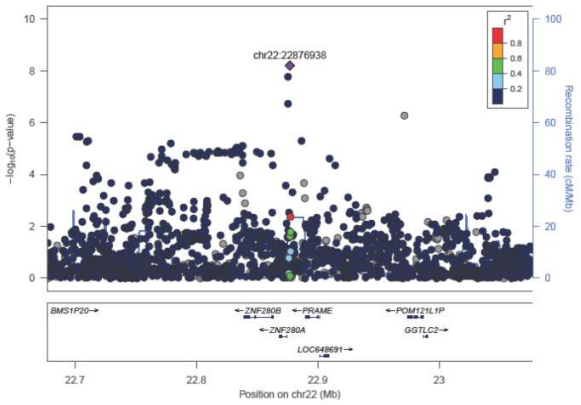
All cause heart failure (meta analysis):Known loci



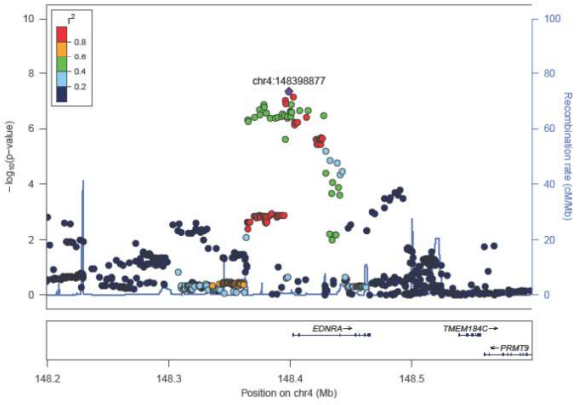
All cause heart failure (meta analysis):Known loci

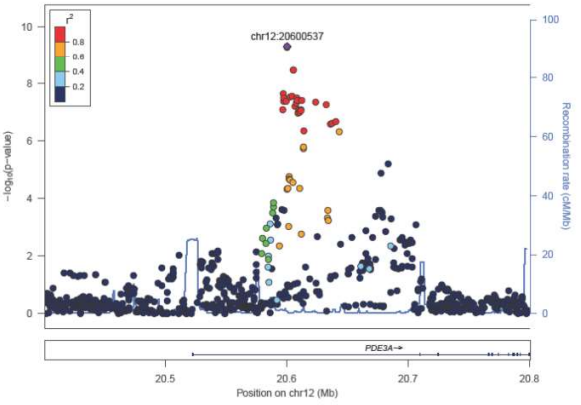




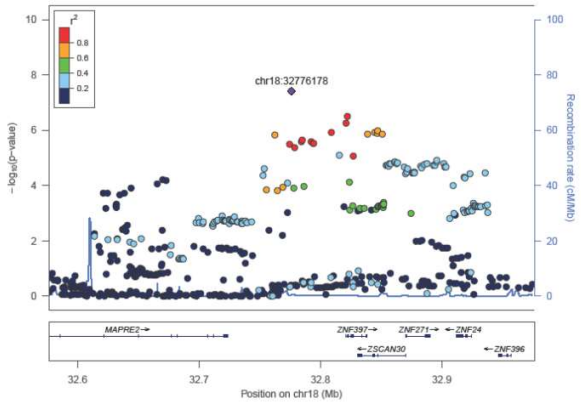


HFrEF (meta analysis): Novel loci

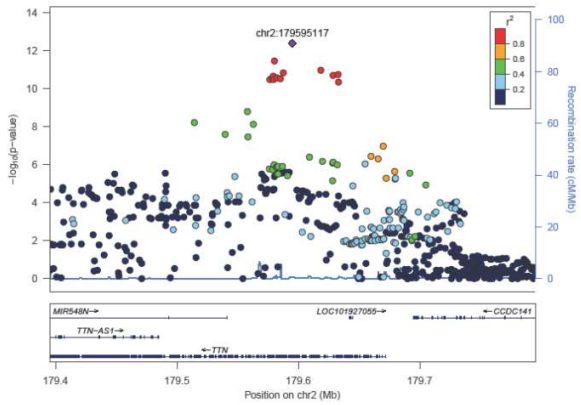


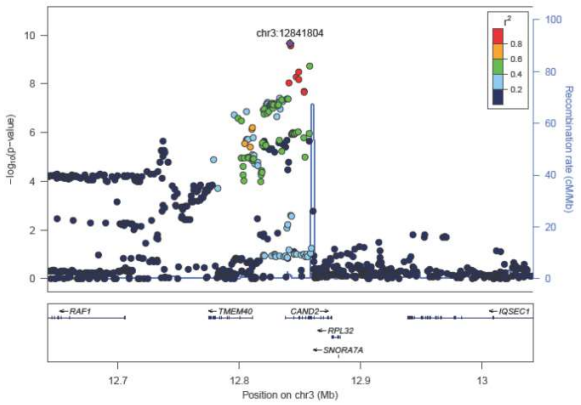


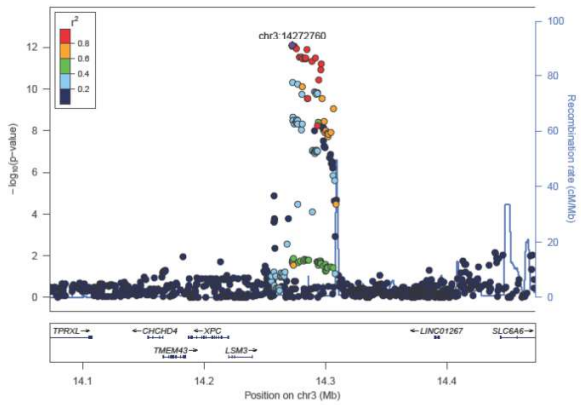
HFrEF (meta analysis): Novel loci

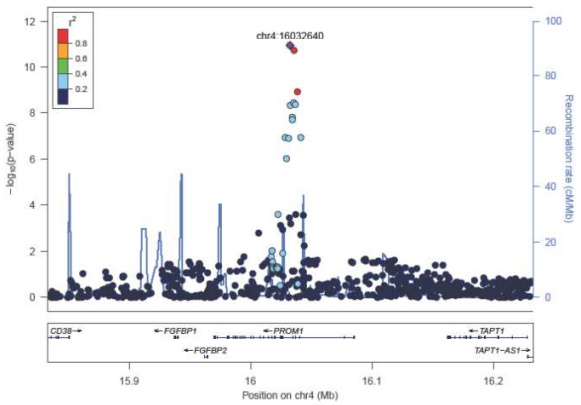


HFrEF (meta analysis): Previously identified only in MTAG

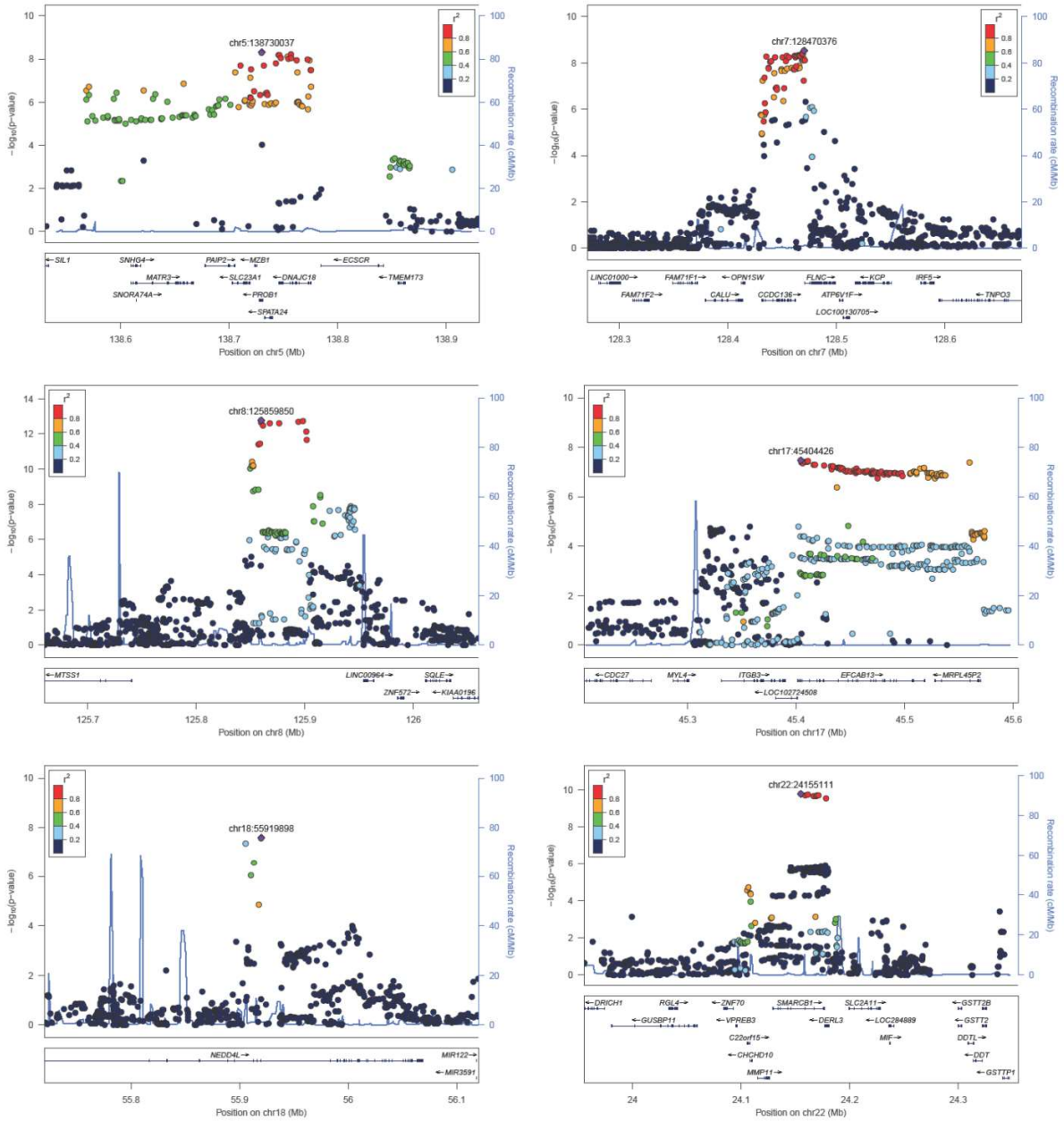




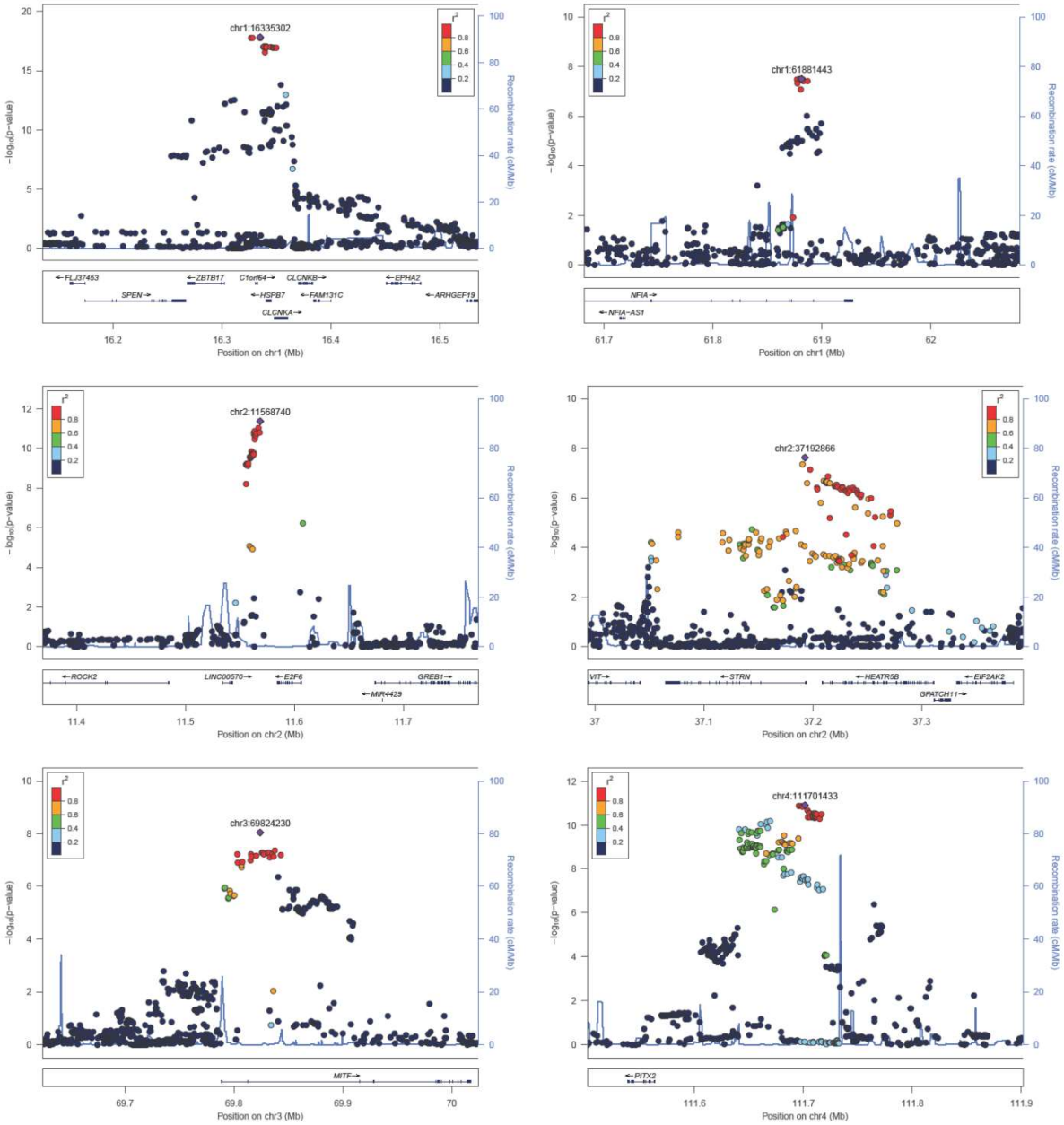




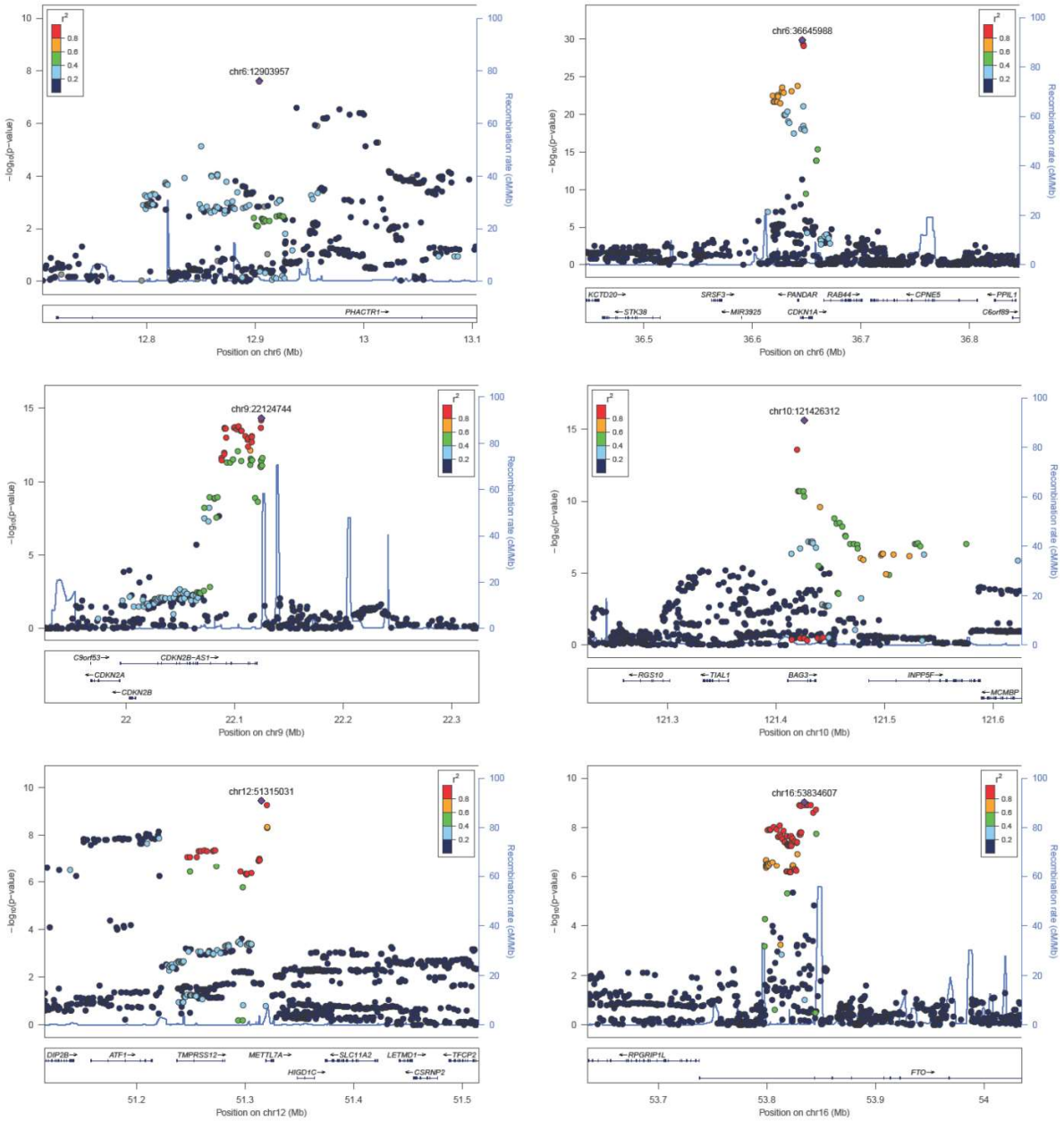
HFrEF (meta analysis): Previously identified only in MTAG



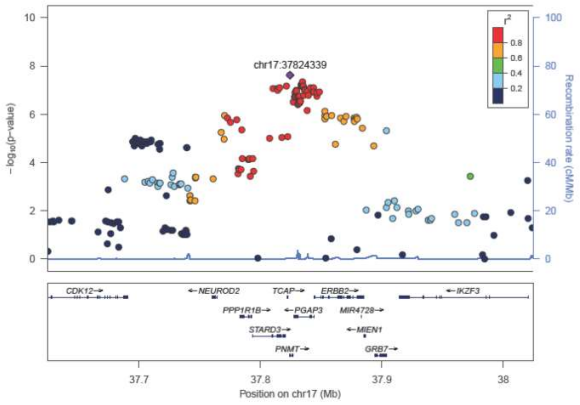
HFrEF (meta analysis): Known loci

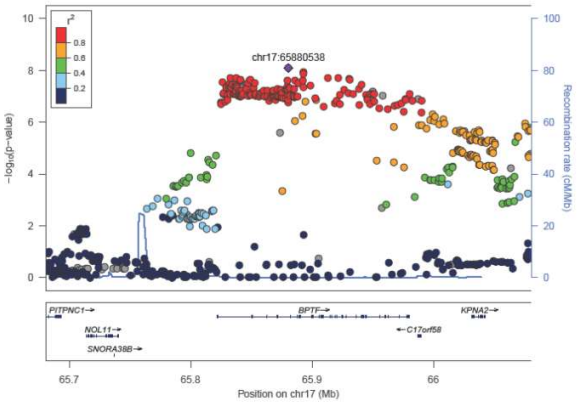


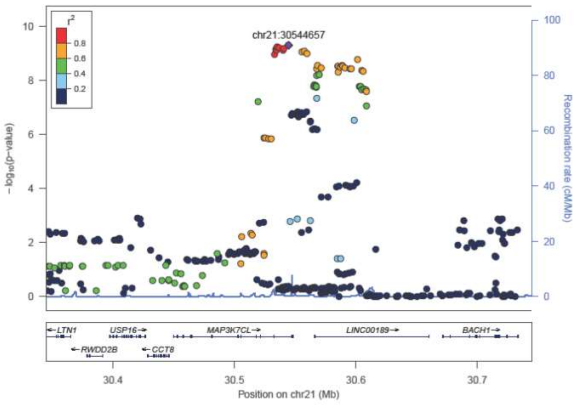
HFrEF (meta analysis): Known loci



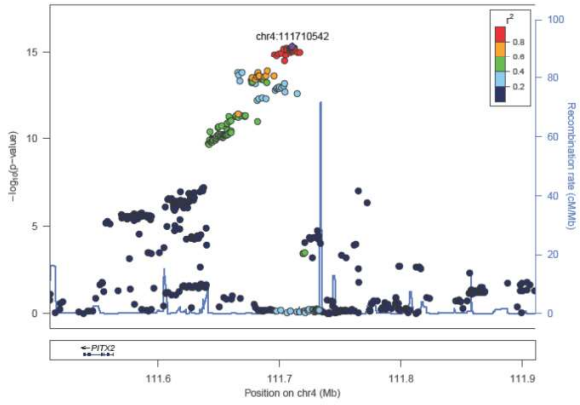
HFrEF (meta analysis): Known loci

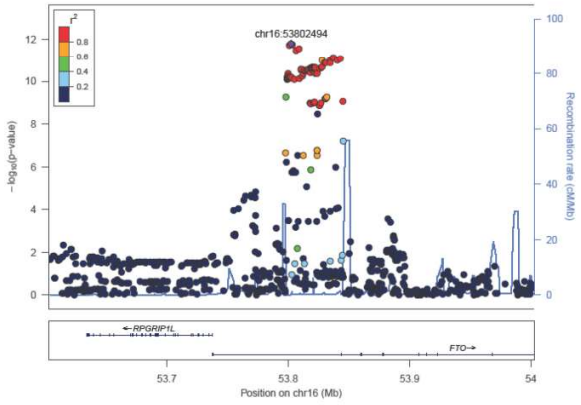




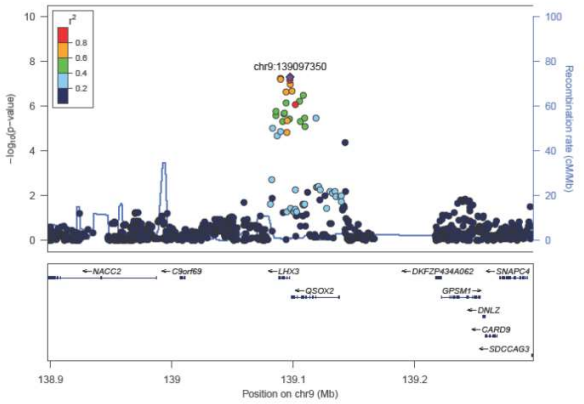


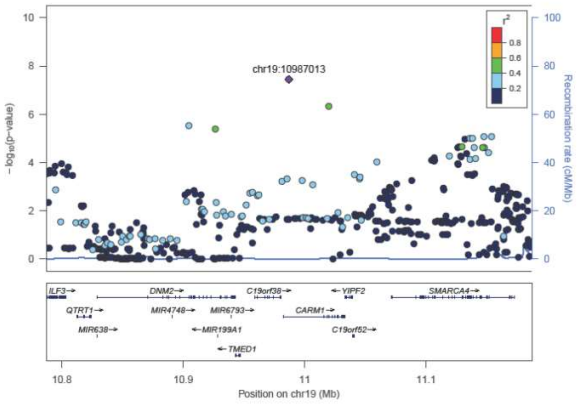
HFpEF (meta analysis): Known loci



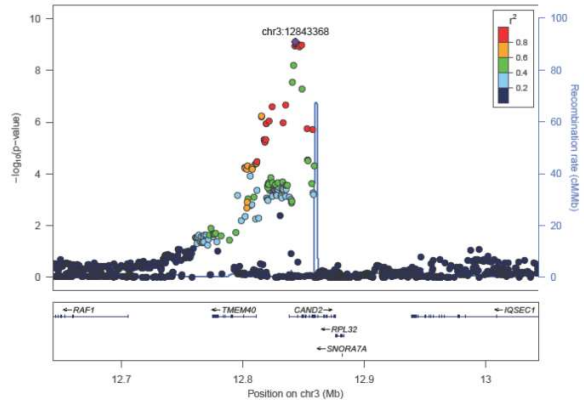


Non ischemic heart failure (meta analysis):Novel loci

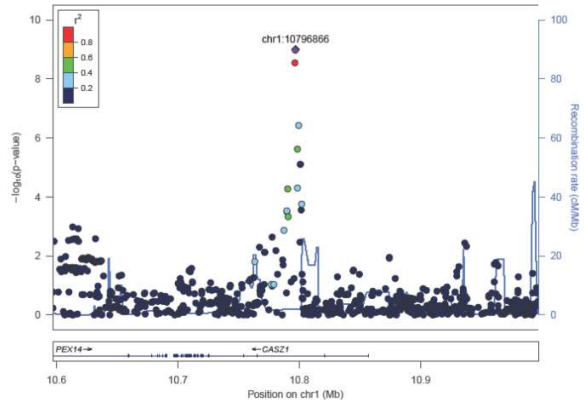


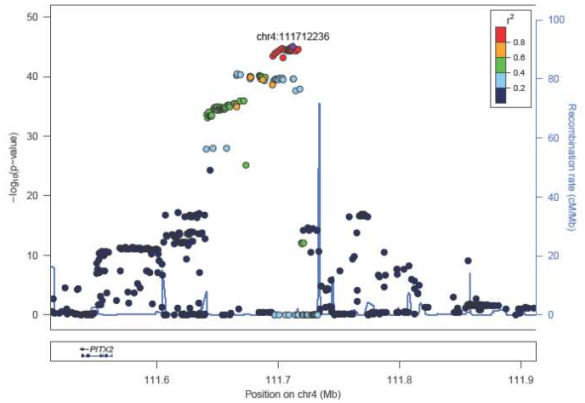


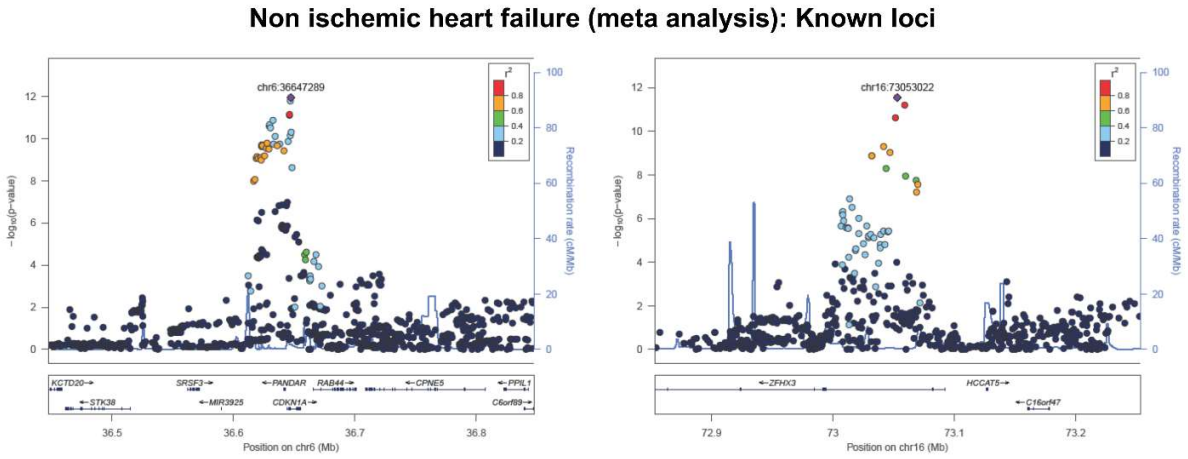
Non ischemic heart failure (meta analysis): Previously identified only in MTAG



Non ischemic heart failure (meta analysis): Known loci

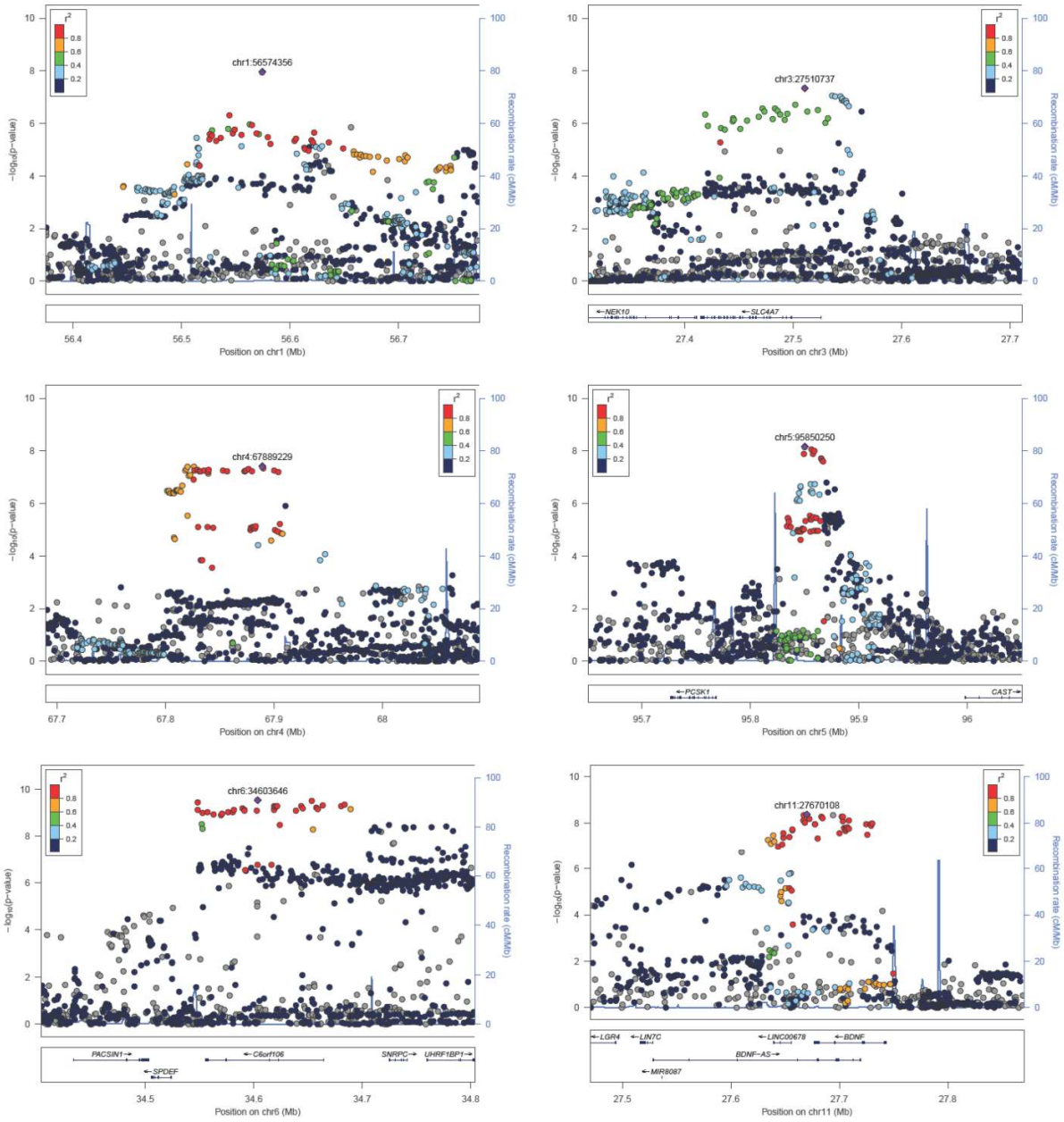




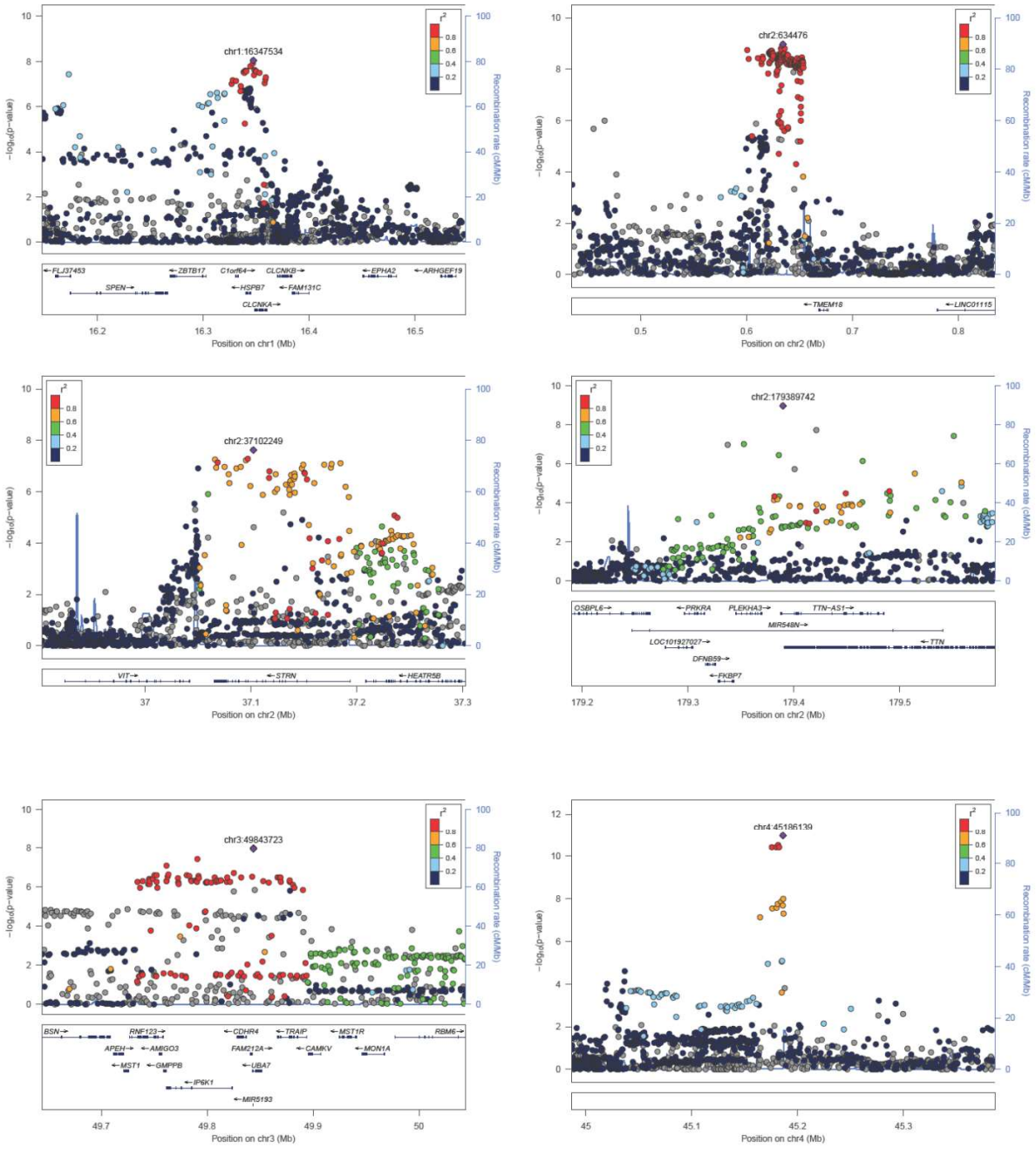


Supplementary Fig. 17 | LocusZoom plots for the cross-ancestry meta-analysis. -$\log_{10}$(*P* value) on the *y*-axis is shown against the genomic positions (hg19) on the *x*-axis.

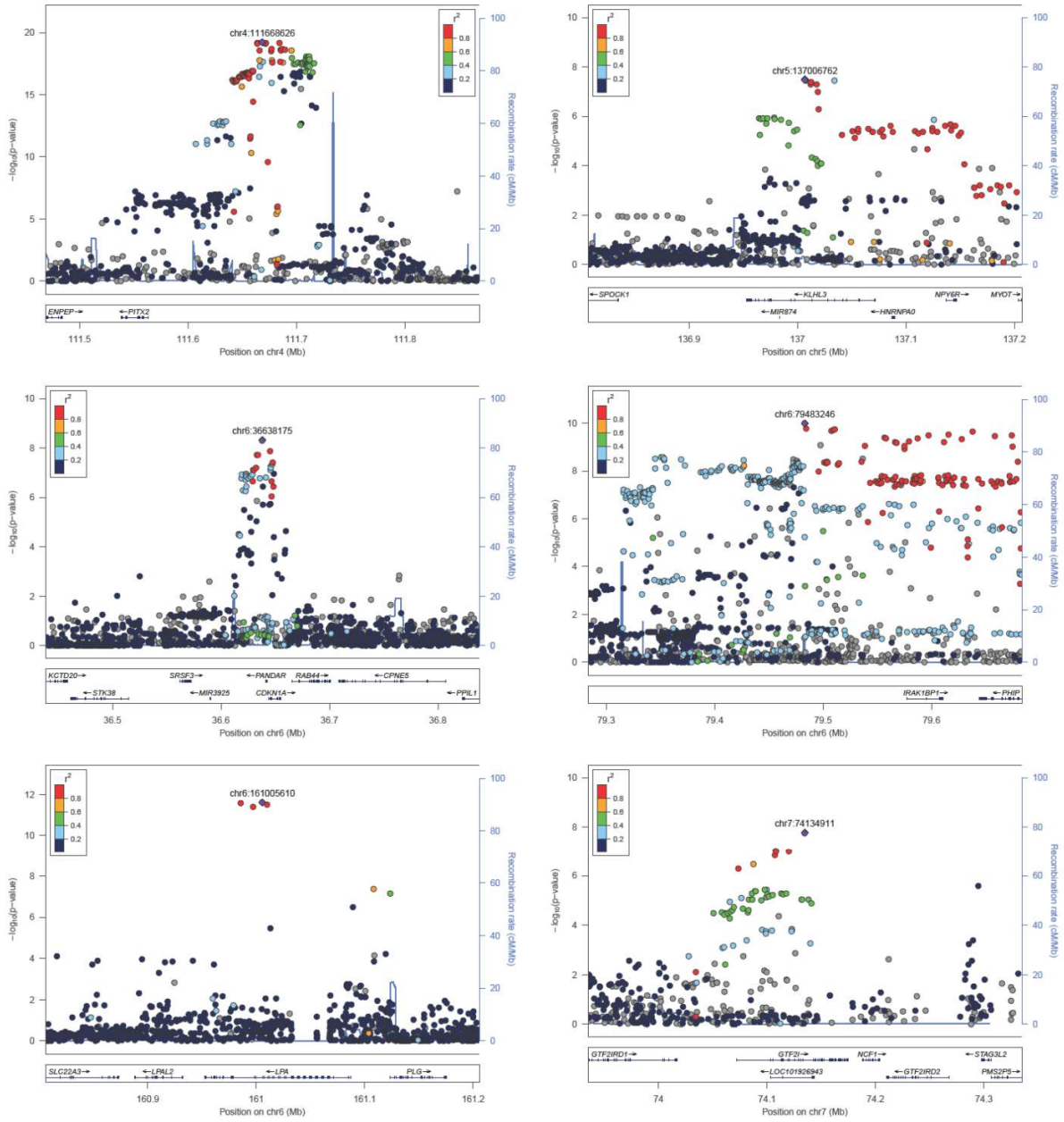
All cause heart failure (MTAG): Novel loci



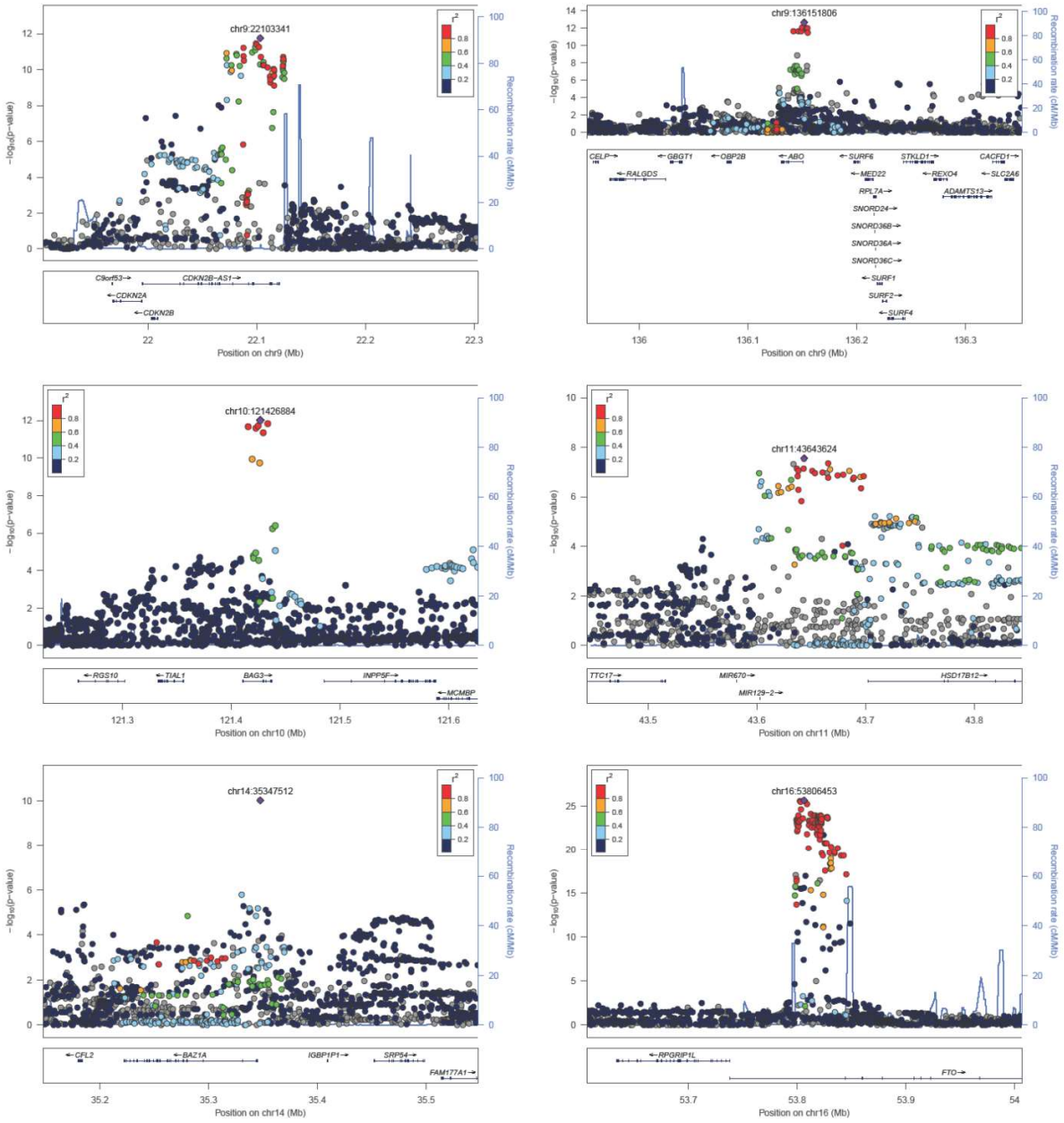
All cause heart failure (MTAG): Known loci



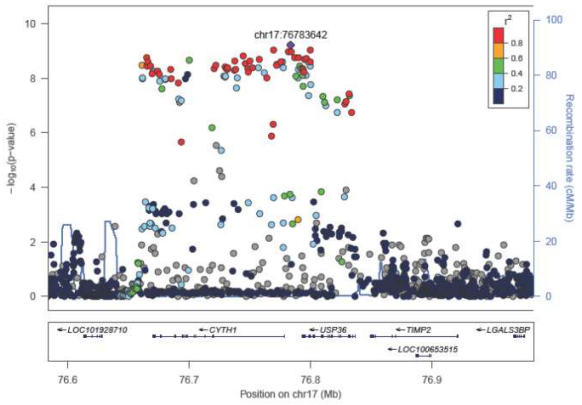
All cause heart failure (MTAG): Known loci

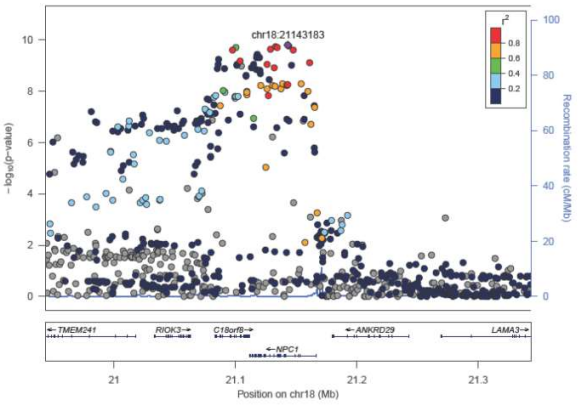


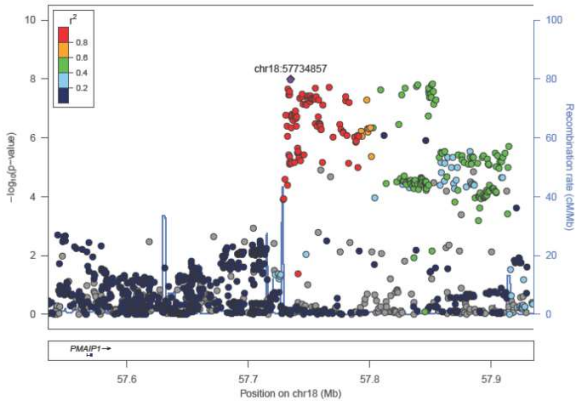
All cause heart failure (MTAG): Known loci



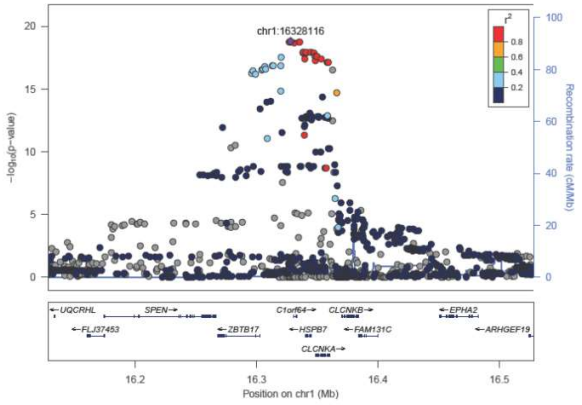
All cause heart failure (MTAG): Known loci

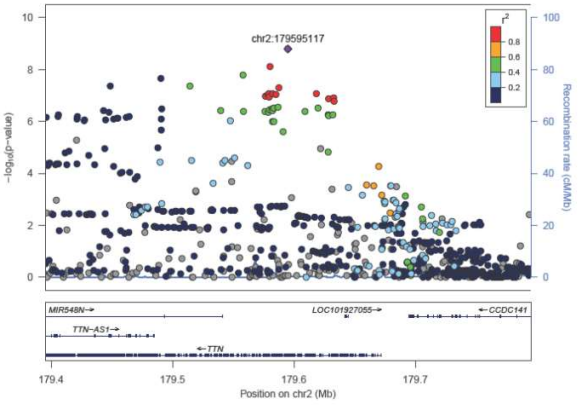




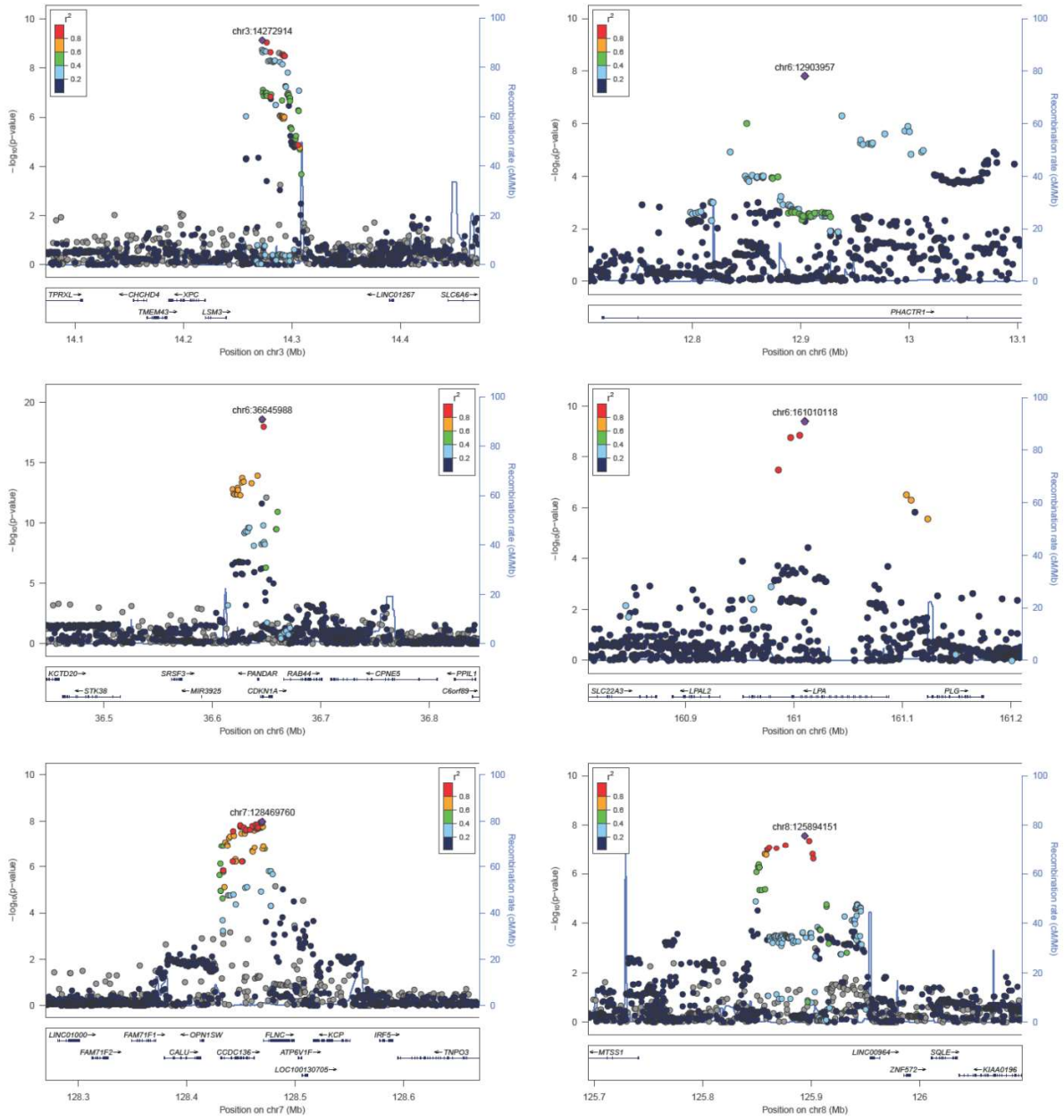


HFrEF (MTAG): Known loci

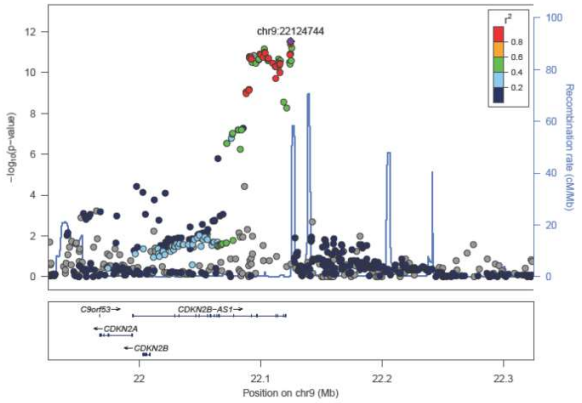


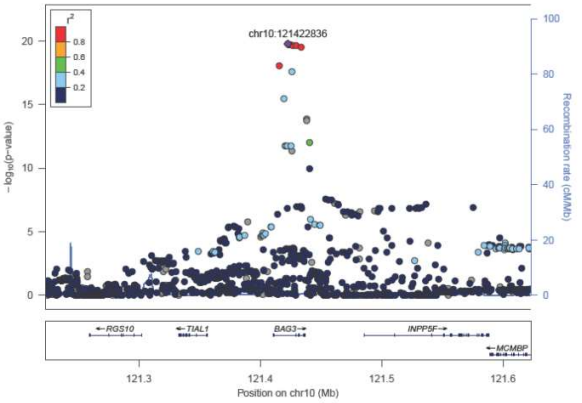


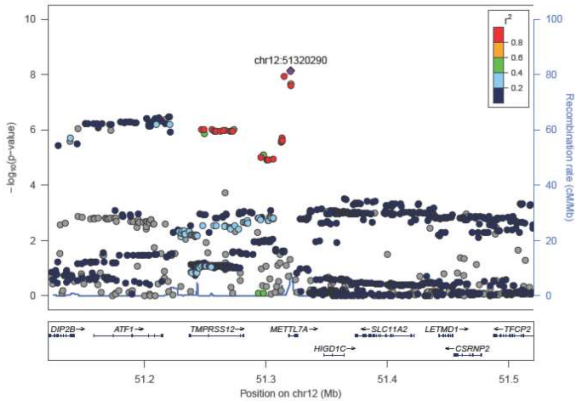
HFrEF (MTAG): Known loci

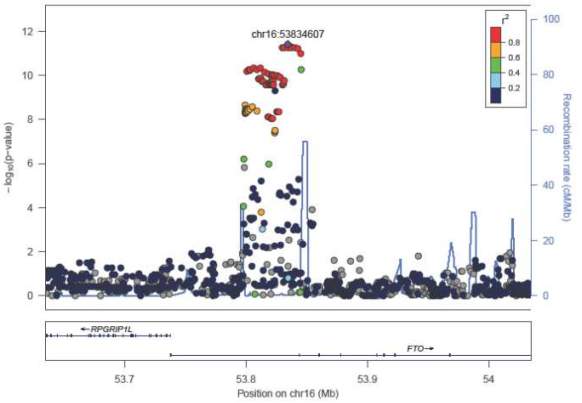


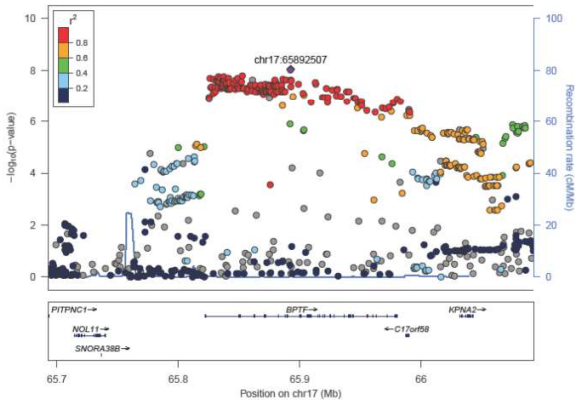
HFrEF (MTAG): Known loci



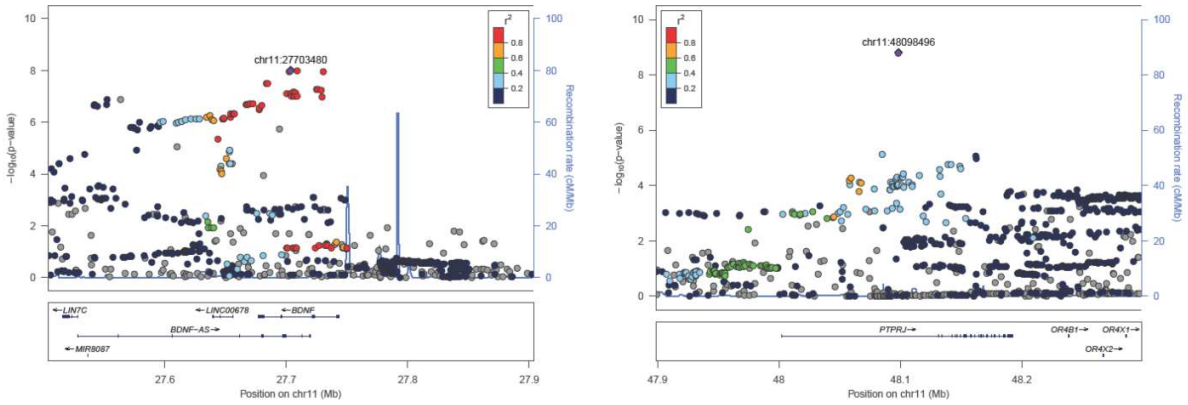




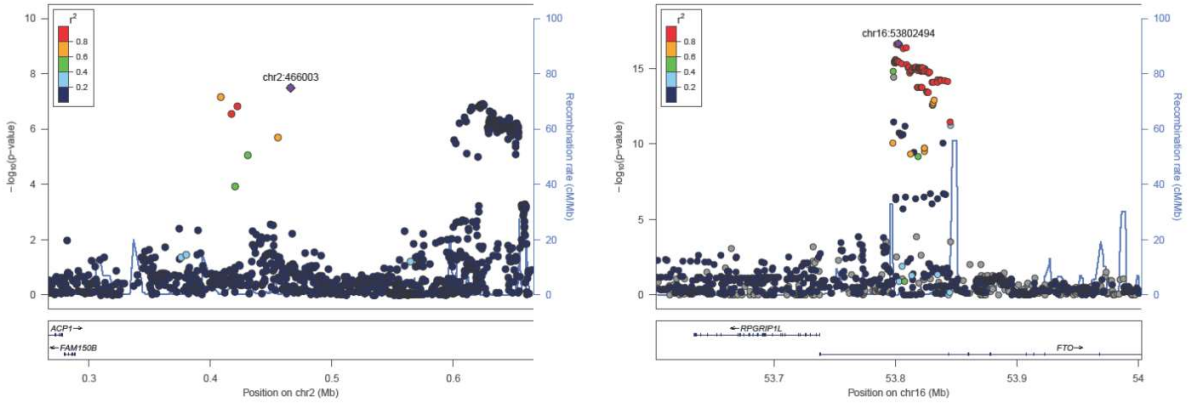




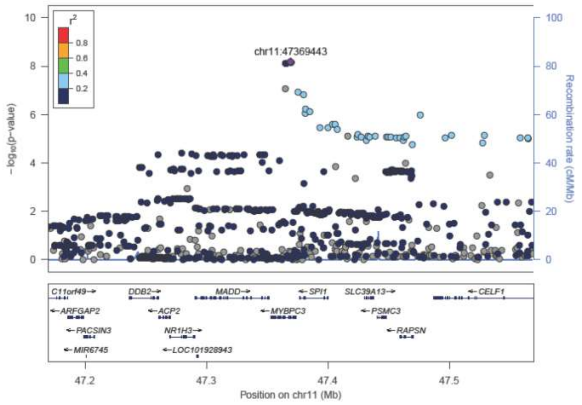
HprEF (MTAG): Novel loci

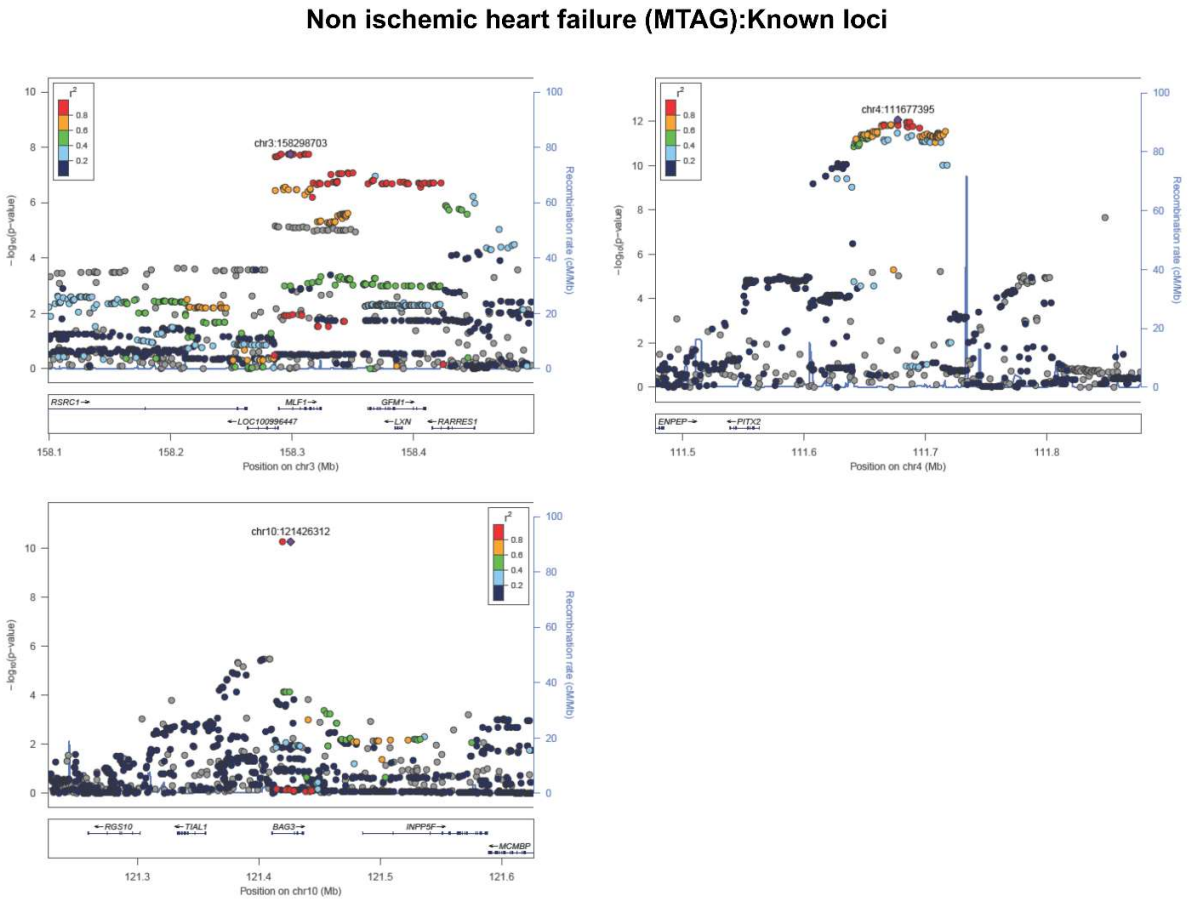


HFpEF (MTAG): Known loci



Non ischemic heart failure (MTAG):Novel loci





Supplementary Fig. 18 | LocusZoom plots for multi-trait analysis of GWAS. -log10(P value) on the y-axis is shown against the genomic positions (hg19) on the x-axis. GWAS, genome-wide association study.

Members of participating consortia

The Biobank Japan Project (biobankjapan@biobankjp.net)

Koichi Matsuda[1,2*], Takayuki Morisaki[2,3], Yukinori Okada[4], Yoichiro Kamatani[5], Kaori Muto[6], Akiko Nagai[6], Yoji Sagiya[2], Natsuhiko Kumasaka[7], Yoichi Furukawa[8], Yuji Yamanashi[3], Yoshinori Murakami[3], Yusuke Nakamura[3], Wataru Obara[9], Ken Yamaji[10], Kazuhisa Takahashi[11], Satoshi Asai[12,13], Yasuo Takahashi[13], Shinichi Higashiue[14], Shuzo Kobayashi[14], Hiroki Yamaguchi[15], Yasunobu Nagata[15], Satoshi Wakita[15], Chikako Nito[16], Yu-ki Iwasaki[17], Shigeo Murayama[18], Kozo Yoshimori[19], Yoshio Miki[20], Daisuke Obata[21], Masahiko Higashiyama[22], Akihide Masumoto[23], Yoshinobu Koga[23] & Yukihiro Koretsune[24]

[1] Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[2] Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.

[3] The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[4] Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[5] Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.

[6] Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[7] Division of Digital Genomics, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[8] Division of Clinical Genome Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[9] Department of Urology, Iwate Medical University, Iwate, Japan.

[10] Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan.

[11] Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan.

[12] Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan.

[13] Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University. School of Medicine, Tokyo, Japan.

[14] Tokushukai Group, Tokyo, Japan.

[15] Department of Hematology, Nippon Medical School, Tokyo, Japan.

[16] Laboratory for Clinical Research, Collaborative Research Center, Nippon Medical School, Tokyo, Japan.

[17] Department of Cardiovascular Medicine, Nippon Medical School, Tokyo, Japan.

[18] Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan.

[19] Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan.

[20] The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan.

[21] Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan.

[22] Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan.

[23] Iizuka Hospital, Fukuoka, Japan.

[24] National Hospital Organization Osaka National Hospital, Osaka, Japan.

* Nominated consortium representative

References

1. Levin MG*, et al.* Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. *Nat Commun* **13**, 6914 (2022).

2. Joseph J*, et al.* Genetic architecture of heart failure with preserved versus reduced ejection fraction. *Nat Commun* **13**, 7753 (2022).

3. Aragam KG*, et al.* Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation* **139**, 489-501 (2019).

4. Pei YF*, et al.* The genetic architecture of appendicular lean mass characterized by association analysis in the UK Biobank study. *Commun Biol* **3**, 608 (2020).

5. Zhang L*, et al.* Body Composition and Incident Heart Failure in Older Adults: Results From 2 Prospective Cohorts. *J Am Heart Assoc* **11**, e023707 (2022).

6. Gu Q*, et al.* Common variants in PCSK1 influence blood pressure and body mass index. *J Hum Hypertens* **29**, 82-86 (2015).

7. Donertas HM, Fabian DK, Valenzuela MF, Partridge L, Thornton JM. Common genetic associations between age-related diseases. *Nat Aging* **1**, 400-412 (2021).

8. Koyama S, *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* **52**, 1169-1177 (2020).

9. Vuckovic D, *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231 e1211 (2020).

10. Astle WJ, *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419 (2016).

11. Asai R, *et al.* Endothelin receptor type A expression defines a distinct cardiac subdomain within the heart field and is later implicated in chamber myocardium formation. *Development* **137**, 3823-3833 (2010).

12. Sakaue S, *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021).

13. Zhu Z, *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol* **145**, 537-549 (2020).

14. Richardson TG, *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med* **17**, e1003062 (2020).

15. Finucane HK*, et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).

16. Sakaue S*, et al.* Integration of genetics and miRNA-target gene network identified disease biology implicated in tissue specificity. *Nucleic Acids Res* **46**, 11898-11909 (2018).

17. Zhang MJ*, et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* **54**, 1572-1580 (2022).

18. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).