Genetic architecture of natural variation of cardiac performance in flies

Saswati Saha[1†], Lionel Spinelli[1†], Jaime A Castro-Mondragon[2], Anaïs Kervadec[3], Laurent Kremmer[1,5], Laurence Roder[1], Sallouha Krifa[1], Magali Torres[1], Christine Brun[1,4], Georg Vogler[3], Rolf Bodmer[3], Alexandre R. Colas[3]*, Karen Ocorr[3]*, Laurent Perrin[1,4]*

[†] Equal contribution

[1]Aix Marseille Univ, INSERM, TAGC, UMR_S_1090, Turing Centre for Living systems, 13288 Marseille, France

[2] Centre for Molecular Medicine Norway (NCMM), P.O. Box 1137 Blindern, 0318 Oslo, Norway

[3] Development, Aging and Regeneration Program, Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA, 92037, USA

[4] CNRS, 13288 Marseille, France

[5] Present address: INRAE, Institut Sophia Agrobiotech, Université Côte d'Azur, CNRS, 06903 Sophia Antipolis, France

Short title:

Genetics of cardiac disorders: from flies to humans.

*Authors for correspondance (acolas@sbpdiscovery.org, kocorr@sbpdiscovery.org; laurent.perrin@univ-amu.fr) - Laurent Perrin, TAGC UMR 1090, Campus de Luminy, Case 928, 13288 Marseille Cedex 9, France. Tel : +33 4 91828727

Abstract

Deciphering the genetic architecture of human cardiac disorders is of fundamental importance but their underlying complexity is a major hurdle. We investigated the natural variation of cardiac performance in the sequenced inbred lines of the *Drosophila* Genetic Reference Panel (DGRP)[1]. Genome Wide Associations Studies (GWAS) identified genetic networks associated with natural variation of cardiac traits which were used to gain insights as to the molecular and cellular processes affected. Non-coding variants that we identified were used to map potential regulatory non-coding regions, which in turn were employed to predict Transcription Factors (TFs) binding sites. Cognate TFs, many of which themselves bear polymorphisms associated with variations of cardiac performance, were also validated by heart specific knockdown. Additionally, we showed that the natural variations associated with variability in cardiac performance affect a set of genes overlapping those associated with average traits but through different variants in the same genes. Furthermore, we showed that phenotypic variability was also associated with natural variation of gene regulatory networks. More importantly, we documented correlations between genes associated with cardiac phenotypes in both flies and humans, which supports a conserved genetic architecture regulating adult cardiac function from arthropods to mammals. Specifically, roles for PAX9 and EGR2 in the regulation of the cardiac rhythm were established in both models, illustrating that the characteristics of natural variations in cardiac function identified in *Drosophila* can accelerate discovery in humans.

Introduction

Heart diseases is a major cause of mortality[2]. Although a large number of genome wide association studies (GWAS) have identified hundreds of genetic variants related to cardiovascular traits[3–6], we are very far from a comprehensive understanding of the genetic architecture of these complex traits. Deciphering the impact of genetic variations on quantitative traits is however critical for the prediction of disease risk. But disentangling the relative genetic and environmental contributions to pathologies is challenging due to the difficulty in accounting for environmental influences and disease comorbidities. Underlying epistatic interactions may also contribute to problems with replication in human GWAS performed in distinct populations which rarely take epistatic effects into account. In addition, linking a trait associated locus to a candidate gene or a set of genes for prioritization is not straightforward[7,8]. Furthermore, the analysis of genetic factors related to cardiac traits is

complicated by their interactions with several risk factors, such as increasing age, hypertension, diabetes mellitus, ischemic and structural heart diseases[9].

These pitfalls can be overcome using animal models. Model organisms allow precise controlling of the genetic background and environmental rearing conditions. They can provide generally applicable insights into the genetic underpinnings of complex traits and human diseases, due to the evolutionary conservation of biological pathways. Numerous studies have highlighted the conservation of cardiac development and function from flies to mammals. Indeed, orthologous genes control the early development as well as the essential functional elements of the heart. The fly is the simplest genetic model with a heart muscle and is increasingly used to identify the genes involved in heart disease and ageing[10–12]. Although a large number of genes are implicated in establishing and maintaining cardiac function in *Drosophila*[13], the extent to which genes identified from mutant analysis reflect naturally occurring variants is neither known, nor do we know how allelic variants at several segregating loci combine to affect cardiac performance.

We previously showed that wild populations of flies' harbor rare polymorphisms of major effects that predispose them to cardiac dysfunction[14]. Here, we analysed the genetic architecture of the natural variation of cardiac performance in *Drosophila*. Our aims were to *(i)* identify the variants associated with cardiac traits found in a natural population, *(ii)* decipher how these variants interact with each other and with the environment to impact cardiac performance, and *(iii)* gain insights into the molecular and cellular processes affected. For this, we used the *Drosophila* Genetic Reference Panel (DGRP) [1,15], a community resource of sequenced inbred lines. Previous GWAS performed in the DGRP indicate that inheritance of most quantitative traits in *Drosophila* is complex, involving many genes with small additive effects as well as epistatic interactions[16]. The use of inbred lines allows us to assessthe effects of genetic variations in distinct but constant genetic backgrounds and discriminate genetic and environmental effects.

We demonstrated substantial among-lines variations of cardiac performance and identified genetic variants associated with the cardiac traits together with epistatic interactions among polymorphisms. Candidate loci were enriched for genes encoding transcription factors (TFs) and signaling pathways, which we validated *in vivo*. We used non-coding variants - which represented the vast majority of identified polymorphisms - for predicting transcriptional regulators of associated genes. Corresponding TFs were further validated *in vivo* by heart specific RNAi mediated knockdown. This illustrates that natural variations of gene regulatory networks have widespread impact on cardiac function. In addition, we analyzed the

phenotypic variability of cardiac traits between individuals within each of the DGRP lines (ie, with the same genotype), and we documented significant diversity in phenotypic variability among the DGRP lines, suggesting genetic variations influenced phenotypic variability of cardiac performance. Genetic variants associated with this phenotypic variability were identified and shown to affect a set of genes that overlapped with those associated with trait means, although through different genetic variants in the same genes.

Comparison of human GWAS of cardiac disorders with results in flies identified a set of orthologous genes associated with cardiac traits both in *Drosophila* and humans, supporting the conservation of the genetic architecture of cardiac performance traits, from arthropods to mammals. siRNA mediated gene knockdowns were performed in human derived cardiomyocytes (hiPSC-CM) to indeed show that *dmPox-meso*/hPAX9 and *dmStripe*/hEGR2 have conserved functions in cardiomyocytes from both flies and humans. These new insights into the fly's genetic architecture and the connections between natural variations and cardiac performance permits the accelerated identification of essential cardiac genes and pathways in humans.

Results

Quantitative genetics of heart performances in the DGRP

In this study, we aimed to evaluate how naturally occurring genetic variations affect cardiac performance in young *Drosophila* adults and identify variants and genes involved in the genetic architecture of cardiac traits. To assess the magnitude of naturally occurring variations of the traits, we measured heart parameters in one week-old females for 167 lines from the DGRP, a publicly available population of sequenced inbred lines derived from wild caught females (Figure 1A). Briefly, semi-intact preparations of individual flies[17] were used for high speed video recording combined with Semi-automated Heartbeat Analysis (SOHA) software (www.sohasoftware.com) which allows precise quantification of a number of cardiac function parameters[18 19]. Fly cardiac function parameters are highly influenced by sex[20]. Due to the experimental burden of analyzing individual cardiac phenotypes, we focused on female flies only and designed our experiment in the following way: we randomly selected 14 lines out of 167 that were replicated twice. The remaining 153 lines were replicated once. Each replicate was composed of 12 individuals. No block effect was observed due to the replicates in the 14 selected lines (see Table S1). This allowed us to perform our final analysis on one replicate of each of the 167 lines. A total sample of 1956 individuals was analyzed. Heart period (HP), Diastolic Interval (DI), Systolic Interval (SI), Arrhythmicity Index (AI), End Diastolic Diameter (EDD), End Systolic Diameters (ESD) and Fractional Shortening (FS) (Figure 1A) were measured across the whole population (Dataset S1 and Table 1). We found significant genetic variation for all traits (Figure 1B and Figure S1) with broad sense heritability ranging from 0,30 (AI) to 0,56 (EDD) (Table 1). Except for EDD/ESD and HP/DI, quantitative traits were poorly correlated with each other (Figure S1B).

GWAS analyses of heart performance

To identify candidate variants associated with cardiac performance variation, we performed GWAS analyses and evaluated single marker associations of line means with common variants using a linear mixed model[21] and after accounting for effects of Wolbachia infection and common polymorphism inversions (see M&M). Genotype-phenotype associations were performed separately for all 7 quantitative traits and variants were ranked based on their p-values. For most of the phenotypes analyzed, QQ plots were uniform (Figure S2) and none of the variants reached the strict Bonferroni correction for multiple tests ($2\ 10^{-8}$), which is usual in the DGRP given the size of the population. However, the decisive advantage of the

Drosophila system is that we can use GWA analyses as primary screens for candidate genes and mechanisms that can be subsequently validated by different means. We therefore chose to analyze the 100 top ranked variants for each quantitative trait. This choice is based on our strategy to test the selected SNPs and associated genes by a variety of approaches - data mining and experimental validation (see below)- in order to provide a global validation of association results and to gain insights into the characteristics of the genetic architecture of the cardiac traits. This cut-off was chosen in order to be able to test a significant number of variants while being globally similar to the nominal cut-off (10^{-5}) generally used in DGRP analyses. A large proportion of the variants retained have indeed a p-value below 10^{-5}. Selected variants were further filtered on the basis of minor allele frequency (MAF> 4 %) (Dataset S2, Table S2). Among the 7 quantitative traits analyzed, we identified 530 unique variants. These variants were associated to genes if they were within 1 kb of transcription start site (TSS) or transcription end sites (TES). Using these criteria, 417 variants were mapped to 332 genes (Table S2). We performed a chi-squared test to determine if the genomic location of variants associated with cardiac traits are biased towards any particular genomic region when compared with the whole set of variants with MAF>4% in the DGRP population and obtained a p-value of 2.778e-13. Genomic locations of the variants were biased towards regions within 1 kb upstream of genes TSS, and, to a lesser extent, to genes 5' UTR (Figure 1C). Variants not mapped to genes (located at >1kb from TSS or TES) were slightly depleted in our set.

In GWAS analyses, loci associated with a complex trait collectively account for only a small proportion of the observed genetic variation [22] and part of this 'missing heritability' is thought to come from interactions between variants[23 22 24 25]. As a first step towards identifying such interactions, we used FastEpistasis[26]. Single Nucleotide Polymorphism (SNP) identified by GWAS were used as focal SNPs and were tested for interactions with all other SNPs in the DGRP. FastEpistasis reports best ranked interacting SNP for each starting focal SNP, thus extending the network of variants and genes associated to natural variation of cardiac performance, which were used for hypothesis generation and functional validations. 288 unique SNPs were identified, that were mapped to 261 genes (Dataset S3, Table S3). While none of the focal SNPs interacted with each other, there is a significant overlap between the 332 genes associated with single marker GWAS and the 261 genes identified by epistasis (n= 31, Figure 1D and Table S3, Fold Change (FC)= 6; hypergeometric pval = $6{,}8 \times 10^{-16}$). This illustrates that the genes that contribute to quantitative variations in cardiac performance have a tendency to interact with each other, although through distinct alleles.

Taken together, single marker GWAS and epistatic interactions performed on the 7 cardiac phenotypes identified a compendium of 562 genes associated with natural variations of heart performance (Table S4). In line with the correlation noted between their phenotypes (Figure S1B), the GWAS for HP and DI identified partially overlapping gene sets (overlap index 0.23, Figure 1E). The same was true, to a lesser extent, for ESD and EDD (0.15). Otherwise, the sets of genes associated with each of the cardiac traits are poorly correlated with each other.

Functional annotations and network analyses of association results

Our next objective was to identify the biological processes potentially affected by natural variation in cardiac performance. Gene Ontology (GO) enrichment analysis of the combined single marker GWAS and epistatic interactions analyses indicated that genes encoding signaling receptors, transcription factors (TFs) and cell adhesion molecules were over-represented among these gene sets (pval= $1{,}4x10^{-9}$ (FC = 2,9), $5x10^{-4}$ (FC = 2) and $3x10^{-3}$ (FC = 4,6) respectively). There was also a bias for genes encoding proteins located at the plasma membrane, at ion channel complexes as well as components of contractile fibers (pval = $3{,}4x10^{-10}$ (FC= 3), $7x10^{-5}$ (FC= 4,2) and $4x10^{-2}$ (FC= 3,6); Figure 2A; Table S5). Of note, although a number of genes have previously been identified as being required during heart development or for the establishment and maintenance of cardiac function by single gene approaches, we found no enrichment for these gene categories in our analysis. In addition, genes identified in a global screen for stress induced lethality following heart specific RNAi knockdown[13] were also not enriched in GWAS detected genes (FC = 1; Table S5). This indicates that genes associated to natural variations of cardiac traits are typically missed by traditional forward or reverse genetics approaches, which highlights the value of our approach.

In order to gain additional knowledge about the cellular and molecular pathways affected by natural variations of cardiac traits, we have mapped the associated genes and gene products onto characterized interaction networks. Of the 562 identified genes, 419 were mapped to the fly interactome that includes both physical (protein-protein) and genetic interactions from both DroID[27] and flybi databases (see Materials and Methods and Dataset S4). Remarkably, a high proportion (148) of the GWAS identified genes were directly connected within the fly interactome and formed a large network of interacting genes/proteins (Table S6, Figure 2B) suggesting that they participate in common biological processes. This network encompasses several TFs and ion channel complex genes, consistent with their

potential role in the genetic architecture of natural variation of heart performance. Several components of signaling pathways are also present in the network, including members of the FGF and TGFb pathways (see below).

Functional validations of candidate genes

To assess whether mutations in genes harboring SNPs associated with variation in cardiac traits contributed to these phenotypes, we selected 42 GWAS associated genes for cardiac-specific RNAi knockdown and assessed the effects on cardiac performance. We selected genes that were identified in at least two independent GWAS for two traits or that were known to be dynamically expressed in the adult heart[28] and for which inducible RNAi lines were available. Genes were tested in 1 week old adult female flies, using the heart specific Hand>Gal4 driver[29] and the same semi-intact heart preparations and SOHA analyses as for DGRP lines screening. Notably, 38 of the 42 selected genes led to various levels of cardiac performance defects following heart-specific knock down (KD) (Figure 2C). In parallel, we tested the effect on cardiac performance of knocking down 18 genes randomly selected in the genome - the GWAS associated genes being excluded from the selection (see Materials and Methods and Figure S3). Although a number of these genes lead to cardiac phenotypes when inactivated -which is consistent with published observations that quantitative traits can be influenced by a large number of genes[30]- when inactivated in the heart, the genes selected from GWAS lead to significantly more frequent phenotypes compared to the randomly selected genes (Figure S3). These results therefore supported our association results.

We further focused on the TGFb pathway, since members of both BMP and activin pathways were identified in our analyses. We tested different members of the TGFb pathway for cardiac phenotypes using cardiac specific RNAi knockdown (Figure 2C), and confirmed the involvement of the activin agonist *snoo* (Ski orthologue) and the BMP antagonist *sog* (chordin orthologue). Notably, Activin and BMP pathways are usually antagonistic (Figure 2D). Their joint identification in our GWAS suggest that they act in a coordinated fashion to regulate heart function. Alternatively, it may simply reflect their involvement in different aspects of cardiac development and/or functional maturation. In order to discriminate between these two hypotheses, we tested if different components of these pathways interacted genetically. Single heterozygotes for loss of function alleles show dosage-dependent effects of *snoo* and *sog* on several phenotypes, providing an independent confirmation of their involvement in several cardiac traits (Figure S4). Importantly, compared to each single

heterozygotes, *snoo*^{BSC234}/ *sog*^{U2} double heterozygotes flies showed non additive SI phenotypes (two-way ANOVA p val: 2,1 10-7) suggesting a genetic interaction (Figure 2E and Figure S4). This result supported a coordinated action of activin and BMP in the establishment and/or maintenance of cardiac activity. Synergistic interaction of BMP and activin pathways was further suggested by the analysis of cardiac performance in loss of function mutants for *babo*, the type 2 activin receptor, and *dpp*,the BMP ligand, supporting a genetic interaction and affecting fractional shortening (Figure 2E and Figure S4).

Taken together, the cardiac phenotypes observed following RNAi mediated gene KD of GWAS selected genes and the interactions characterized between different Activin and BMP pathways members are strong validation of the GWAS results that are further supported by the functional analysis of non-coding variants and by the correlations we observed between identified fly genes and human genes associated to cardiac dysfunction (discussed below).

Natural variations of heart performance primarily affect Gene Regulatory Networks

The over-representation of genes coding for TFs found in GWAS for cardiac performance traits (p-val=5,559.10^{-4}, Figure 2A, Table S5) suggested that perturbations of gene regulatory networks (GRN) play a central role in the genetic architecture of natural variations of cardiac performance. In addition, most variants (95%) were localized in non-coding regions of the genome. Remarkably, their distribution was enriched in gene proximal (0-1kb) upstream regions (Figure 1C) suggesting that a high proportion of variants affects cis-regulatory sequences.

Thus, we sought to determine whether potential TF binding motifs are enriched in the vicinity of variants identified by GWAS. Analysis of the sites of the variants themselves did not reveal any particular pattern. However, oligo analysis[31] of the 150 bp sequences around these non-coding variants detected significant over-represented k-mers (Evalue <= 1e-4). These were further assembled and converted into Positions Specific Scoring Matrices (PSSM) and compared with PSSM databases of known fly TFs binding motifs were used for annotations (see methods, Dataset S6 and Figure S5A). Variants associated with quantitative traits linked to heart size and contractility (ESD, EDD and FS; 367 variants) and those related to heart rhythmicity (AI, HP, DI, SI; 435 variants) were analyzed separately. We identified enrichment of potential DNA binding motifs for both sets of traits (Figure S5B). Although this approach predicted motifs that were specific for each group of variants, a significant proportion of potential DNA binding motifs were predicted by both groups, suggesting commonalities in the genetic architecture of the corresponding traits.

Remarkably, we identified DNA binding motifs for TFs that were themselves associated with natural variations of the cardiac traits. Some examples are *klumpfuss (klu/ZBTB7A), luna (KLF5-7), longitudinals lacking (lola/BTB18), serpent (srp/GATA1-6), mirror (mirr/IRX3-6), Helix Loop Helix protein 3B (HLH3B/TAL2)* and *defective proventiculus (dve/SATB1)* which were also identified by GWAS (see Table S4). Potential binding sites for Mother against dpp, Mad (SMAD) were enriched in the vicinity of the non-coding variants associated with heart size and contractility. SMAD is the downstream effector of the BMP signaling pathway that we showed was also involved in regulating cardiac contractility (Figure 2D-E). Motifs predicted to be bound by the NK TFs Tinman (Tin/NKX2.5) and Bagpipe (Bap/NKX3.2) are over-represented in both sets of variants. *tin* has a well characterized function in the cardiac GRN, both during embryogenesis[32] and at adulthood [33] and *bap* is also a member of the cardiac network[34]. Lastly, we found *Jun* related antigen (Jra/JUN) TFBS enriched in sequences associated with natural variations of cardiac rhythm. We previously reported that *djun* knockdown leads to altered HP and AI in young flies[28], supporting a function for this stress-activated TF in the cardiac GRN.

The effects of cardiac specific RNAi mediated gene KD of *luna*, *lola* and *srp* KD had already been analyzed in semi-intact preparations and shown to induced strong cardiac phenotypes (Figure 2C). We repeated these heart specific manipulations employing additional, independent RNAi lines targeting these genes and an *in vivo* approach that uses the red fluorophore TdTomato driven by a cardiac enhancer[35] to monitor heart wall mouvements. To validate roles for the predicted TF Binding Sites (TFBSs) we performed heart spcific KD of *mirr*, *mad*, *klu* and *HLH3B* using this same *in vivo* approach. As illustrated in Figure S5C, all tested TFs are associated with changes in multiple cardiac traits. *mad* knockdown affected FS but also impacted several traits related to rhythmicity. Consistent with the predictions of its binding sites, *klu* knockdown altered HP and SI but also EDD and ESD. A PSSM for one isoform of *lola* (Lola-RC) was predicted from variants associated with rhythmicity, and another one (for Lola-RL) from variants related to contractility. Of note, *lola* KD strongly impacted all analyzed cardiac phenotypes. *luna* cardiac specific RNAi KD had an impact on EDD, FS, HP and SI, in line with its binding sites predictions. Similarly, *srp* KD impacted HP, DI and SI and also modified cardiac diameters. Lastly, *mirr* KD affected both SI and ESD and *HLH3B* impacted EDD, HP and SI (Figure S5C and Table S7).

Taken together, the high rate of non-coding SNPs among variants identified in GWAS, the TFBS predictions with *in vivo* validation of cognate TFs, and enrichment for TFs among

genes associated with natural variations of heart performance, support a central role of GRN deviations in the natural variation of cardiac performance.

Identifying natural variations of phenotypic variability of cardiac performance

The use of inbred lines, offers the possibility of studying the heritability of the within-line variance of the phenotypes. Several studies have indeed shown that the within-line environmental variance of inbred lines is heritable[36–38]. These genetic variations in phenotypic variability suggest that different genotypes respond differently in response to micro-environmental changes[39,40]. We used the Levene test[41] to examine heterogeneity of the "micro-environmental" variance among lines for the 7 analyzed cardiac traits. We found evidence of significant heterogeneity among the lines in the variance for all analyzed cardiac traits (Table 1). This suggests that the lines differ in their ability to adapt to micro-environmental variations, and supports the heritability of this differing adaptability. The coefficient of within line variation (CVe) was used as the metric for within-line environmental variance. We found significant genetic variation for all CVe traits (Figure 3A, B and Figure S6A) indicating that within some lines, individuals display relatively constant cardiac performance traits, whereas within others, individuals display widely varying cardiac traits. As for trait means, CVe-traits were poorly correlated with each other (Figure S6B), except for HP/DI and EDD/ESD. Correlations of CVe traits with their respective trait means ranged from moderate (FS: -0,53) to none (DI: 0,064, SI: 0,09, ESD: 0,08) (Figure S6C). This suggested that distinct loci may affect natural variations of CVe and means of cardiac performance parameters.

We performed GWAS analysis to identify candidate variants and genes associated with variation of CVe of cardiac performance using both single marker and epistasis GWAS analyses. As for trait means we retained the 100 top ranked variants for reporting single marker associations. These variants were further used as focal variants for epistasis detection and the best ranked interacting SNP for each starting focal SNP was identified. Overall, this led to the identification of 887 variants and 566 genes associated with natural variations of CVe of cardiac performance (Dataset S7, table S8), which were used for hypothesis generation and functional annotations. Although the number of individuals we were able to analyze -due to the experimental burden of analyzing individual cardiac phenotypes- reduced the statistical power of the analysis of micro environmental variance, the functional enrichment analyses of variants and genes described below provide support for our association results and highlight important features of the genetic architecture of cardiac traits.

Consistent with the weak correlation between cardiac trait CVe and trait means, we found only 3 variants associated with both (Figure 3C, D and Figure S7; Table S8). Two variants in total linkage disequilibrium (5bp apart) are localized in an intron of *Ca-β*, encoding a voltage gated calcium channel subunit, and are associated for both mean and CVe of fractional shortening (Figure S7, Table S8). The other variant, highlighted by both GWAS, was situated at more than 25kb away from any gene. Despite unraveling largely distinct set of variants, GWAS analyses on CVe and trait means are associated with two sets of genes that overlap significantly (Figure 3C, Table S8). Among 566 detected respectively from trait CVe and 562 genes detected by trait mean GWAS, 92 genes are found associated with both (FC= 5, hypergeometric pval = $2,2 \times 10^{-39}$). Therefore, a significant number of genes contributes to natural variation of both mean and CVe traits, although through distinct variants.

Within the CVe-associated genes we observed an over-representation of cell adhesion processes, transmembrane signaling and ion channel activity (pval= $1,6 \times 10^{-8}$ (FC = 7), $1,1 \times 10^{-3}$ (FC = 2,2) and $4,5 \times 10^{-2}$ (FC = 2,5) respectively). Enriched cellular component annotations included cell-cell junction (pval= $5,7 \times 10^{-3}$ (FC = 3,2), receptor complexes (pval= $3,3 \times 10^{-4}$ (FC = 4,6), sarcolemma (pval= $1,3 \times 10^{-2}$ (FC = 25) as well as several annotations related to neuronal projections suggesting both cell autonomous and non-autonomous activity of related genes (Figure 3F, Table S9). In addition, genes leading to stress-induced lethality upon heart specific knockdown[13] were significantly underrepresented in this gene set (FC= 0,6; p= 0,05; Table S9). This may suggest that a significant proportion of CVe GWAS associated genes affect cardiac trait variance non-cell autonomously. As for trait means, variants localized in the proximal region of the genes (0-1kb upstream of TSS) are over-represented among SNPs associated with traits CVe (Figure 3E). This suggested that natural variations of cardiac performance variance are also mainly affecting cis-regulatory regions. We also identified over-represented PSSMs in sequences within 150pb of non-coding CVe variants (Figure 3G). Some of these identified PSSMs were specifically enriched in the sequences around CVe variants. This is the case for instance for Ladybird-early, Lbe, the fly orthologue of the mammalian LBX transcription factor. However, several identified CVe PSSMs are similar to those identified in the vicinity of variants associated with trait means. This is the case for motifs predicted to be bound by Luna and Lola - variants of which are associated with CVe traits (Table S8)-, and it is also the case for Klu and Mad. Therefore, similar to the situation described for trait means, natural variations of phenotypic plasticity of cardiac performance appears, for a large part, to be driven by modifications of gene regulatory networks.

Conserved features of natural variations of heart performance between flies and humans

Next, we compared our GWAS results for cardiac performance in flies with similar data in humans. A literature survey identified a comprehensive set of human genes associated with cardiac disorders and coronary artery diseases (CAD) in human populations (Table S10). Using DIOPT[42], we identified human orthologues of the genes associated with cardiac trait means and CVe in flies (Table S11). Compared to the whole fly genome gene set (17500 genes), both sets of GWAS associated genes were significantly enriched for genes that have a human orthologue (FC= 1,4, pval= 4,3 10^{-21} for genes associated with trait means and FC= 1,25; pval = 7 10^{-11} for genes associated with traits CVe, Table 2 and S11), indicating that, overall, the loci associated with natural variations in cardiac traits have evolutionary conserved functions. More importantly, GWAS loci were enriched for genes whose human orthologues were associated with cardiac disorders (FC= 1,66 (trait means) and 1,7 (traits CVe), pval = 7 10^{-6}, Table 2 and S11). Lower enrichment was found for orthologues of human genes associated with Coronary Artery Disease (FC= 1,35, pval=0,04 (trait means) and 1,25, pval = 0,07 (traits CVe)). This analysis strongly supports the view that the cellular and molecular processes of heart development and physiology - which is already well characterized between fly and humans[10–12,43] - are not only conserved but also have a common genetic basis underlying their natural variation.

Building on these observations we next wanted to determine whether our data in flies could guide the identification of novel players in the human cardiac system. We focused on the paired box TF gene *pox-meso* (*poxm*/PAX9) and on the zinc finger containing TF *stripe* (*sr*/EGR2) – both were associated with natural variations of cardiac trait means (Table S4). Notably, neither the fly genes nor their human orthologues have a known function in the cardiac system. In flies, *poxm* contributes to the development and specification of somatic muscles[44], while *sr* is required for induction of tendon cell fate[45]. We first tested the effects of heart specific KD on cardiac performance in flies using the *in vivo* assay. Both induced an increased heart rate (reduced HP and SI Figure 4A) but did not affect other cardiac traits analyzed (not shown). The function of their human orthologues PAX9 and EGR2 was then tested in human induced pluripotent stem cells derived cardiomyocytes (hiPSC-CM). We asked whether siRNA-mediated KD affected action potential kinetics using a recently developed *in vitro* assay (Figure 4B)[46]. As shown in Figure 4C-E, both gene KDs induced shortening of the AP duration. The effect was stronger for PAX9 KD and seems specific to

the repolarization phase, as suggested by APD50 and 75 shortening (Figure 4F-I). Increased AP kinetics correlated with shorter calcium transient (Figure S8A-D and H-K), including faster upstroke and downstroke calcium kinetics and increased beat rate (peak frequency) (Figure S8E-G and L, M), which confirmed that both PAX9 and EGR2 KD impacted cardiomyocyte contractility by shortening AP and calcium transient duration, thereby increasing beat frequency. Our data therefore show comparable effects on cardiac rhythm of the inactivation of these conserved TFs both in the context of the whole organ in flies, and in the *in vitro* assay in human iPSC-CM.

Discussion

Disentangling the relative contribution of genetics and environmental factors to cardiac pathologies and the influence of comorbidities is of fundamental importance from a medical perspective. From a fundamental perspective, natural selection is thought to operate on complex traits rather than on single genes, further illustrating the need to characterizing and understanding the genetic architecture of natural variations in cardiac performance. Although mammalian studies provide important information about the molecular and cellular basis of heart development and physiology, these systems are less suited to dissecting the relationship between naturally occurring genetic variations and variations in individual cardiac performance.

Here we report on the quantitative genetics of cardiac performance in *Drosophila*. We performed a standard GWAS of variants associated with cardiac trait looking at both within line and intra-line variance for 7 weakly correlated traits that also included an epistasis analysis. The cornerstone of our approach is a series of downstream analyses, which strongly implied that top hits are indeed genetic polymorphisms influencing quantitative variation in cardiac performance. These findings include: (*i*) the enrichment of pathways known to be related to cardiac development and physiology; (*ii*) the generation of an extensive network of protein-protein/genetic interactions that plausibly describes the architecture of cardiac performance; (*iii*) identification of significant overlap between gene occurrences for trait mean and variance ; (*iv*) enrichment of likely causal variants near binding motifs for a set of cardiac-related transcription factors; (*v*) observation that several of these TFs are themselves GWAS candidate genes; and (*vi*) enrichment of fly genes for overlap with known human

cardiac-related genes. In addition, we performed validation experiments that showed enrichment of cardiac effects for GWAS identified genes compared to randomly selected control genes. Importantly we confirmed synergistic genetic interactions between the BMP and Activin pathways; and defined 2 novel regulators of cardiac function that may well be relevant to human disease.

Our results therefore validate *Drosophila* as a unique model that permits the analysis of the etiology of the genetic foundations of cardiac function. Importantly, we identified many genes not previously associated with cardiac traits at the genetic, cellular or molecular levels. This suggested that the space for natural variation associated with cardiac performance extends far beyond the genes and gene networks already identified as participating in the establishment and maintenance of cardiac function and illustrates the value of complementing classical genetic studies with an in-depth analysis of the genetic architecture of corresponding complex traits. A large number of candidate genes were validated *in vivo* using mutations or RNAi knockdown generated in unrelated genetic backgrounds, suggesting that the underlying biological processes are common to most genotypes.

Most of the variants identified affected non-coding regions of the genome. Recent studies in humans have highlighted that non-coding SNPs may affect the function of enhancers or promoters, thereby influencing the expression levels of surrounding or distant genes and which would explain how some non-coding SNPs may contribute to pathogenesis[47,48]. Although we failed to find enriched potential TFBS at SNPs positions, extending the search space to sequences of +/- 75bp around the variants revealed enriched DNA binding motifs, and pointed to potential roles for several TFBS. Our results indicate that variants likely affect sequences outside the core TF binding motif (e.g. flanking sequences) that in turn affect TF binding[49]. Supporting our findings, different reports on human GWAS analyses have shown that many risk-associated SNPs are not precisely located in conserved TF binding motifs but are in nearby regions[50,51]. Several observations indicate that our predictions are accurate: *(i)* a large number of similar motifs were independently predicted from distinct sets of variants, *(ii)* some predicted motifs correspond to TFs whose role during cardiogenesis or for the maintenance of cardiac function has already been characterized, *(iii)* several discovered motifs correspond to TFs for which SNPs were themselves associated with the variation of some of the cardiac traits studied and *(iv)* a high percentage of predicted motifs were validated by testing their cognate TFs *in vivo* for cardiac performance. Of note, while the cardiac GRN has been extensively studied in flies[34,52], a large proportion of the TFBS and cognate factors identified in this study have not been previously

characterized as cardiac TFs, again illustrating the complementary contribution represented by the analysis of the genetic architecture of these phenotypes compared to more classical genetic, molecular and cellular studies.

Although rarely studied, the genetic control of phenotypic variability is of primary importance. Indeed, variability offers adaptive solutions to environmental changes and may impact developmental robustness to microenvironmental perturbations[39]. From a medical genetics point of view, highly variable genotypes may produce individuals beyond a phenotypic threshold, leading to disease state[36,53]. Having assessed the cardiac traits for multiple individuals from inbred lines, we were able to examine phenotypic variability among individuals with the same genotypes and therefore to analyze how natural variations impinge on such complex traits. Our results indeed highlight the important contribution that genetic control of variability can play in our understanding of the cause of phenotypic variation. An important feature concerns the nature of the variants and associated genes compared to trait means. In particular, genes identified by a global screen for genes cell-autonomously affecting cardiac function[13] are under-represented among the genes associated with CVe traits. This suggests that this subset of genes influences the variance of cardiac traits in a cell non-autonomous manner. The separation between variants associated with trait means and those associated with the variance of the same traits was another remarkable feature of the genetic architecture of cardiac phenotypic plasticity. Although based on non-overlapping sets of non-coding regions, predictions of TFBS in the vicinity of non-coding variants identified overlapping sets of motifs. This suggests that natural variation of trait means and trait variance affect partially overlapping GRNs, albeit by distinct polymorphisms. Our analysis allowed us to identify common variants associated with phenotypic variability, and to identify essential characteristics of the genetic architecture of these traits. However, we cannot exclude that part of the observed variability is due also to rare alleles in partial linkage disequilibrium (LD) with the lead associations[54] or to within-line accumulation of mutations. Nevertheless, the former is likely to be marginal, given the rapid decay of LD with local physical distance in *Drosophila*[16].

Several previous studies have established that the principles identified on the DGRP population are universal and can be applied across phyla[40]. In particular, the identification of human genes orthologous to those identified by GWAS in the DGRP may allow identification of a translational blueprint and prioritize candidate genes for focused studies in human populations[55,56]. Our results provide additional support that the genetic architecture of quantitative traits is conserved. Indeed, a remarkable characteristic of the gene sets associated

either with trait means or with trait variance is that they are both enriched for orthologs of human genes that are also associated with disorders affecting the cardiovascular system. This is a strong argument for considering that the genetic architecture of cardiac performance traits is conserved, at least in part, from flies to humans. Hence, the features learned from our study in flies should help reveal the characteristics of the natural variations of cardiac performance in human populations. One intriguing observation is that genes associated with intra-genotypic variance are highly significantly enriched for orthologues of human genes associated with cardiac disorders. Yet GWAS on human populations are not designed to test the variants involved in intra-genotypic variations. One possible explanation may be that GWAS in humans can pick up variants that do not necessarily affect the mean of cardiac traits, but that cause extreme phenotypes in some individuals by affecting genes that function as buffers.

The high proportion of orthologous genes identified in GWAS in both flies and humans - for cardiac traits in *Drosophila* and for cardiac disorders in humans - suggested that our results in flies may accelerate discovery of causal genes in humans. We suggest that our study may be used as an aid in prioritizing genes at loci identified by GWAS for cardiac pathologies in humans. Indeed, we show that our data allow identification of new actors that have a conserved role in cardiac function from flies to humans. Although not previously revealed by conventional approaches, we show that *poxm* and *sr* - two genes identified by GWAS as being associated with natural variations of cardiac traits - have a conserved function in the regulation of cardiac activity.

In summary, the analysis of natural variations of cardiac performance in *Drosophila*, combined with functional validations *in vivo* and in human iPSC-CM, represents a major advance in understanding the mechanisms underlying the genetic architecture of these complex traits, and holds promise for the identification of the genes and mechanisms underlying cardiac disorders in humans.

Materials and Methods

- Flies' strains and husbandry

Fly strains were obtained from the Bloomington *Drosophila* Stock Center or the Vienna *Drosophila* Resource Center. Flies were reared at a controlled density on corn flour/yeast/sucrose-agar medium at 25°C, 50% relative humidity and 12h light/dark. For movies acquisitions, newly emerged adults (0-24h) were collected and aged for 7 days. For screening of the DGRP population 12 flies were analyzed for each DGRP strain. For validations (RNAi mediated gene knockdown or mutant analyses) 30 to 40 flies per genotype were analyzed.

- Fly heart-specific RNAi knockdown and mutant analyses:

Driver-line (Hand>Gal4 4.2)[29] virgins were crossed to UAS-RNAi males or corresponding isogenic control males. BMP mutants were back-crossed into the same isogenic control background. Flies were raised on standard fly food and kept at 25°C. Female progeny were collected and aged to 1 week at 25°C, at which point they were imaged and analyzed.

- Semi-intact preparations and SOHA analyses

To measure cardiac function parameters, denervated, semi-intact *Drosophila* females adult fly hearts were prepared according to previously described protocols[13]. High-speed ~3000 frames movies were taken at a rate of 140 frames per second using a Hamamatsu ORCAFlash4.0 digital CMOS camera (Hamamatsu Photonics) on a Zeiss Axioplan microscope with a 10X water immersion lens. The live images of the heart tube within abdominal A3-4 segments were captured using HCI imaging software (Hamamatsu Photonics). M-modes and cardiac parameters were generated using SOHA (www.sohasoftare.com), as described previously[13]. M-mode records provide a snapshot of the movement of heart wall over time. Cardiac parameters were calculated as below: Diastolic interval (DI) is the duration for each heart relaxation phase (diastole). Systolic interval (SI) is the duration of the contraction phase of the heart. Heart period (HP) is the time between the two consecutive diastoles (i.e., DI+SI) and heart rate is calculated from HP (1/HP). Arrhythmia index (AI) is the standard deviation of the HP mean for each fly normalized to the median HP. Fractional shortening was calculated as (diastolic diameter – systolic diameter)/diastolic diameter).

- In vivo analysis of heart function (tdtk):

R94C02::tdTomato (attP2)[35] was combined with Hand-Gal4 as stable stock. Flies were then crossed to RNAi lines and reared to 3 weeks of age. To assay fluorescent hearts *in vivo* flies were mounted as described[35] and were illuminated with green light (3mW power). 5 sec movies of the beating heart were acquired at 270fpsand high-speed recordings were then processed using a custom R script[57].

– Source code for analyses :

We have developed a github project containing the instructions and material to reproduce the analyses. Source code are available in the github repository. Required data and built Singularity images are available on download. Instructions to reproduce the analyses are provided. (https://doi.org/10.5281/zenodo.5582846)

- Quantitative genetic analyses:

Phenotype data were analyzed separately to perform quality control. In each line, data outside the range [Q1-1.5*IQR; Q3+1.5*IQR] were identified (where Q1/3 are first/third quartile and IQR=Q3-Q1). Most of the films corresponding to these outlier data were re-analyzed. It was identified that these extreme results corresponded to technical problems (poor quality film, bad placement of the measurement points...) and were removed from the analysis to avoid technical bias.

Mean and Coefficient of Variation (CV) were computed for each line and each phenotype. Only lines with at least 7 remaining observations after QC were kept for the computation.

Correlation between phenotypes were computed using Spearman correlation between mean/CV of lines. Genetic indicators, like broad sense heritability, were computed from the QC data.

- Genotype-phenotype associations: GWAS and Epistasis study

We used mean and CV summary statistics from our GWAS analysis to study the leverage effect of extreme values. We identified strong leverage effects in some phenotypes and treated them for outliers. We decided to remove lines with large leverage effect from GWAS analysis of the phenotype the, since they imply a strong bias in the linear model. Hence, mean/CV values outside the range of [Q1-1.5*IQR; Q3+1.5*IQR] were discarded from GWAS analysis of the corresponding phenotype, thus removing all leverage effects.

GWAS analysis was performed using FastLMM[21]. Linear mixed model was built using Wolbachia infection and common polymorphic inversions (data provided by the DGRP Consortium) as covariates. A github project containing the instructions and material to reproduce the analyses reported is available at: https://doi.org/10.5281/zenodo.5582846.

A quantile-quantile (QQ plot) was generated to display the quantile distribution of observed marker-phenotype association P-values versus the distribution of expected P-values by chance. The QQ plot was plotted to check if some polymorphisms are more associated with the phenotype under a null hypothesis, which would be represented by a uniform distribution of P-values. For each phenotype, we selected the 100 variants with the lowest GWAS p-value. These variants were mapped to genes using the following rules: (i) if the variant was located in a gene (INTRON, EXON, 3'or 5' UTR), the SNP was then associated with this gene, (ii) if not, the variant was associated with a gene if its distance from the boundary of the gene (Transcription Start Site (TSS) and Transcription End Site (TES) was less than 1kb (iii) otherwise the variant was not mapped to a gene. Of the mapped variants we then examined variants with a Minor Allele Frequency (MAF) greater than 4% and discarded SNPs with synonymous effects. Epistasis analysis was performed using FastEpistasis[26]. For each phenotype, we used the SNP selected from the GWAS analysis as focal SNP and the interaction was tested with all the SNPs with MAF greater than 5% in the DGRP population. The best matching test SNP was selected for each focal SNP. Each resulting SNP was mapped to genes using the same rules as for GWAS variants.

Miami plots were designed according to[58]

- <u>Interaction Network analysis:</u>

Drosophila Interactome: A network generated from genetic and physical interaction databases has been built. Briefly, known protein-protein interactions, identified using Y2H and AP-MS techniques, and genetic interactions, are curated in the DroID database (http://www.droidb.org/)[27], to which the Y2H FlyBi dataset (https://flybi.hms.harvard.edu/results.php) has been added. Overall, this *Drosophila* interaction network contains 79,698 interactions between 11,022 genes/proteins. Network representations were made using Cytoscape[59].

- <u>Statistics</u>

Enrichment p-values were based on a test following the hypergeometric distribution. Enrichment (fold change: FC) were calculated as the ratio between observed and expected successes in samples (based on the observations in the population). Gene Ontology enrichment analyzes were conducted using the online version of G:profiler tool[60] with default parameters, including g:SCS method for computing multiple testing corrections for p-values gained from GO and pathway enrichment analysis.
(https://biit.cs.ut.ee/gprofiler/page/docs#significance_threhshold).

Statistical analysis of cardiac parameters in RNAi and mutant experiments were performed using R. The non-parametric Wilcoxon-Mann-Whitney test was used to compare the median of the difference between two groups (tested genes/respective control).

Comparing Z scores (Figure S3B): For each phenotype and each gene, Z-scores summarizes the difference in the phenotype distribution from tested genes to their respective control. Z scores were calculated for all phenotypes (7 phenotypes) and all tested genes (GWAS associated genes: 42, randomly selected genes: 18). The Z-scores for all the phenotypes of the tested genes were pooled together (GWAS associated genes: 294 Z scores; randomly selected genes: 126 Z scores) and the distribution of absolute Z scores between the two groups was analyzed using Mann Whitney-Wilcoxon Test.

Tests for genetic interactions (Figure 2E and Figure S4): Two way ANOVA with interactions was conducted to test if the interaction of two-genes is significantly different from the two way ANOVA model with the additive effects of the two genes[61].

Phenotypic correlation between each trait pair was computed using spearman correlation. Pearson's chi squared test was applied to test if the genomic location of variants associated with cardiac traits are biased towards any particular genomic region when compared against the whole set of variants in the entire genome of the DGRP population. The overlap coefficient/ overlap index is a similarity measure that measures the overlap between two finite sets. This is used to quantify the overlap between the gene sets associated with the different cardiac traits.

- <u>Motif's analysis</u>

For every set of SNPs (Rhythmicity CVe and Mean, contractility CVe and Mean), the coordinates (dm3 genome assembly) were extended 75 nt to both sides and their corresponding fasta sequences were retrieved. TF binding motifs were identified using RSAT

peak-motifs[62] (default parameter) using two distinct background models: intrinsic (frequencies of short k-mers in the input sequences to calculate the enrichment of longer k-mers) and random (set of random genomic regions matching length and nucleotide composition using biasaway[63]). To reduce the redundancy of the found motifs (PSSMs), we applied the RSAT matrix-clustering[64] algorithm (default parameters) to find clusters of motifs. The clustered motifs were manually annotated by comparison with PSSM databases (Cis-BP[65], JASPAR[66] and FlyFactorSurvey[67]) using the tool RSAT compare-matrices[68] (default parameters).

- Generation of hiPSC-ACMs

Id1 overexpressing hiPSCs[69] were dissociated with 0.5 mM EDTA (ThermoFisher Scientific) in PBS without $CaCl_2$ and $MgCl_2$ (Corning) for 7 min at room temperature. hiPSC were resuspended in mTeSR-1 media (StemCell Technologies) supplemented with 2 µM Thiazovivin (StemCell Technologies) and plated in a Matrigel-coated 12-well plate at a density of 3 x 10^5 cells per well. After 24 hours after passage, cells were fed daily with mTeSR-1 media (without Thiazovivin) for 3-5 days until they reached ≥ 90% confluence to begin differentiation. hiPSC-ACMs were differentiated as previously described[70]. At day 0, cells were treated with 6 µM CHIR99021 (Selleck Chemicals) in S12 media[71] for 48 hours. At day 2, cells were treated with 2 µM Wnt-C59 (Selleck Chemicals), an inhibitor of WNT pathway, in S12 media. 48 hours later (at day 4), S12 media is fully changed. At day 5, cells were dissociated with TrypLE Express (Gibco) for 2 min and blocked with RPMI (Gibco) supplemented with 10% FBS (Omega Scientific). Cells were resuspended in S12 media supplemented with 4 mg/L Recombinant Human Insulin (Gibco) (S12+ media), 2 µM Thiazovivin and 300 nM Retinoic Acid (Sigma) and plated onto a Matrigel-coated 12-well plate at a density of 9 x 10^5 cells per well. S12+ media was changed at day 8 and replaced at day 10 with RPMI (Gibco) media supplemented with 213 µg/µL L-ascorbic acid (Sigma), 500 mg/L BSA-FV (Gibco), 0.5 mM L-carnitine (Sigma) and 8 g/L AlbuMAX Lipid-Rich BSA (Gibco) (CM media). Typically, hiPSC-ACMs start to beat around day 10. At day 15, cells were purified with lactate media (RPMI without glucose, 213 µg/µL L-ascorbic acid, 500 mg/L BSA-FV and 8 mM Sodium-DL-Lactate (Sigma)), for 4 days. At day 19, media was replaced with CM media.

- Voltage Assay in hiPSC-ACMs

Voltage assay is performed using labeling protocol described in McKeithan et al., 2017[46]. Briefly, hiPSC-ACMs at day 25 of differentiation were dissociated with TrypLE Select 10X for up to 10 min and action of TrypLE was neutralized with RPMI supplemented with 10% FBS. Cells were resuspended in RPMI with 2% KOSR (Gibco) and 2% B27 50X with vitamin A (Life Technologies) supplemented with 2 µM Thiazovivin and plated at a density of 6,000 cells per well in a Matrigel-coated 384-well plate. hiPSC-ACMs were then transfected with siRNAs directed against PAX9 and EGR2 (ON-TARGETplus Human) using lipofectamine RNAi Max (ThermoFisher). Each siRNA was tested individually in 8-plicates. Three days post-transfection, cells were first washed with pre-warmed Tyrode's solution (Sigma) by removing 50 µL of media and adding 50 µL of Tyrode's solution 5 times using a 16-channel pipette. After the fifth wash, 50 µL of 2x dye solution consisting in voltage sensitive dye Vf2.1 Cl (Fluovolt, 1:2000, ThermoFisher) diluted in Tyrode's solution supplemented with 1 µL of 10% Pluronic F127 (diluted in water, ThermoFisher) and 20 µg/mL Hoescht 33258 (diluted in water, ThermoFisher) was added to each well. The plate was placed back in the 37°C 5% CO_2 incubator for 45 min. After incubation time, cells were washed 4 times with fresh pre-warmed Tyrode's solution using the same method described above. hiPSC-ACMs were then automatically imaged with ImageXpress Micro XLS microscope at an acquisition frequency of 100 Hz for a duration of 5 sec with excitation wavelength of 485/20 nm and emission filter 525/30 nm. A single image of Hoescht was acquired before the time series. Fluorescence over time quantification and trace analysis were automatically quantified using custom software packages developed by Molecular Devices and Colas lab. Two independent experiments were performed.

- <u>Calcium Assay in hiPSC-ACMs</u>

Calcium assay is performed using labeling protocol as described[72]. Cells were prepared and transfected with siRNAs against PAX9 and EGR2 as described above in the voltage assay. Three days post-transfection, 50 µL of media was removed and replaced in each well by a 2x calcium dye solution consisting in Fluo-4 NW dye (Invitrogen), 1.25 mM Probenecid F-127 (Invitrogen) and 100 µg/mL Hoescht 33258 (diluted in water, ThermoFisher) diluted in warm Tyrode's solution (Sigma). The plate was placed back in the 37°C 5% CO2 incubator for 45 min. After incubation time, cells were washed 4 times with fresh pre-warmed Tyrode's solution and imaged as described above.

Acknowledgments and funding sources:

The project leading to this publication has received funding from Fondation de France (N°00071034) and from Excellence Initiative of Aix-Marseille University - A*MIDEX (A*Midex International) to LP, from the National Institute of Health (R01 HL148827-03, R01 HL149992-02, R01 AG071464-01) and from the Department of Defence ((W81XWH-21-1-0104-01) to AC and from National Institute of Health (R01 HL132241 and R21 AG061598) to KO. We thank the Bloomington *Drosophila* Stock Center (BDSC) and the Vienna *Drosophila* Research Center (VDRC) for fly stocks. Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources.

Contributions:

Experiments, A. K., L. K., L. R., S. K., M. T. and L. P. Methodology and data analysis, S.S., L. S., J. A. C., C. B., G. V. and L. P.; writing manuscript, L.P.; editing manuscript, S.S., L.S., J.A.C., C.B., G. V., A. C. and K. O.; funding acquisition, C. B., R. B., A. C., K. O. and L. P.

References

1. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–8 (2012).
2. Bezzina Connie R., Lahrouchi Najim, & Priori Silvia G. Genetics of Sudden Cardiac Death. *Circ. Res.* **116**, 1919–1936 (2015).
3. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* **50**, 1225–1233 (2018).
4. 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**, 2904–2904 (2018).
5. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
6. Verweij, N. *et al.* The Genetic Makeup of the Electrocardiogram. *Cell Syst.* **11**, 229-238.e5 (2020).
7. Mackay, T. F. C. Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions. *Nat. Rev. Genet.* **15**, 22–22 (2014).
8. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
9. Paludan-Müller, C., Svendsen, J. H. & Olesen, M. S. The role of common genetic variants in atrial fibrillation. *J. Electrocardiol.* **49**, 864–870 (2016).
10. Ocorr, K. *et al.* Genetic Control of Heart Function and aging. *Trends Cardiovasc. Med.* **17**, 177–182 (2007).
11. Diop, S. B. & Bodmer, R. Gaining Insights into Diabetic Cardiomyopathy from Drosophila. *Trends Endocrinol. Metab.* **26**, 618–627 (2015).
12. Rosenthal, N., Harvey, R. P., Bodmer, R. & Frasch, M. Development and Aging of the Drosophila Heart. in *Heart Development and Regeneration* 47–86 (2010).
13. Neely, G. G. *et al.* A global in vivo Drosophila RNAi screen identifies NOT3 as a conserved regulator of heart function. *Cell* **141**, 142–53 (2010).
14. Ocorr, K. a, Crawley, T., Gibson, G. & Bodmer, R. Genetic variation for cardiac dysfunction in Drosophila. *PloS One* **2**, e601–e601 (2007).
15. Huang, W. *et al.* Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res.* **24**, 1193–208 (2014).
16. Mackay, T. F. C. & Huang, W. Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, (2018).
17. Ocorr, K. *et al.* KCNQ potassium channel mutations cause cardiac arrhythmias in Drosophila that mimic the effects of aging. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3943–8 (2007).
18. Fink, M. *et al.* A new method for detection and quantification of heartbeat parameters in Drosophila, zebrafish and embryonic mouse heart. *Biotechniques* **46**, 101–113 (2010).
19. Cammarato, A., Ocorr, S. & Ocorr, K. Enhanced assessment of contractile dynamics in Drosophila hearts. *BioTechniques* **58**, 77–80 (2014).
20. Wessells, R. J., Fitzgerald, E., Cypser, J. R., Tatar, M. & Bodmer, R. Insulin regulation of heart function in aging fruit flies. *Nat. Genet.* **36**, 1275–81 (2004).
21. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
22. Manolio, T. a *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).

23. Flint, J. & Mackay, T. F. C. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
24. Huang, W. *et al.* Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15553–9 (2012).
25. Shorter, J. *et al.* Genetic architecture of natural variation in Drosophila melanogaster aggressive behavior. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3555-63 (2015).
26. Schüpbach, T., Xenarios, I., Bergmann, S. & Kapur, K. FastEpistasis: a high performance computing solution for quantitative trait epistasis. *Bioinforma. Oxf. Engl.* **26**, 1468–9 (2010).
27. Murali, T. *et al.* DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Res.* **39**, D736-43 (2011).
28. Monnier, V. *et al.* dJun and Vri/dNFIL3 Are Major Regulators of Cardiac Aging in Drosophila. *PLoS Genet.* **8**, (2012).
29. Popichenko, D., Sellin, J., Bartkuhn, M. & Paululat, A. Hand is a direct target of the forkhead transcription factor Biniou during Drosophila visceral mesoderm differentiation. *BMC Dev. Biol.* **7**, 49 (2007).
30. Zhang, W., Reeves, G. R. & Tautz, D. Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association. *Curr. Biol.* **31**, 1092-1098.e6 (2021).
31. van Helden, J., André, B. & Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281**, 827–842 (1998).
32. Bodmer, R. The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Dev. Camb. Engl.* **118**, 719–729 (1993).
33. Qian, L. *et al.* Tinman/Nkx2-5 acts via miR-1 and upstream of Cdc42 to regulate heart function across species. *J. Cell Biol.* **193**, 1181–96 (2011).
34. Seyres, D. *et al.* Identification and in silico modeling of enhancers reveals new features of the cardiac differentiation network. *Dev. Camb. Engl.* **143**, 4533–4542 (2016).
35. Klassen, M. P. *et al.* Age-dependent diastolic heart failure in an in vivo Drosophila model. *eLife* **6**, 209–218 (2017).
36. Morgante, F., Sørensen, P., Sorensen, D. A., Maltecca, C. & Mackay, T. F. C. Genetic Architecture of Micro-Environmental Plasticity in Drosophila melanogaster. *Sci. Rep.* **5**, 9785–9785 (2015).
37. Harbison, S. T., McCoy, L. J. & Mackay, T. F. C. Genome-wide association study of sleep in Drosophila melanogaster. *BMC Genomics* **14**, 281–281 (2013).
38. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–11 (2015).
39. Geiler-Samerotte, K. A. *et al.* The details in the distributions: why and how to study phenotypic variability. *Curr. Opin. Biotechnol.* **24**, 752–759 (2013).
40. Anholt, R. R. H. & Mackay, T. F. C. The road less traveled: from genotype to phenotype in flies and humans. *Mamm. Genome* **29**, 5–23 (2018).
41. Levene, H. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling,. in 278–292 (Stanford Univ Press, Stanford, CA, 1960).
42. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).
43. Bier, E. & Bodmer, R. Drosophila, an emerging model for cardiac disease. *Gene* **342**, 1–11 (2004).
44. Duan, H. *et al.* A key role of Pox meso in somatic myogenesis of Drosophila. *Dev. Camb. Engl.* **134**, 3985–3997 (2007).

45. Gunage, R. D., Dhanyasi, N., Reichert, H. & VijayRaghavan, K. Drosophila adult muscle development and regeneration. *Semin. Cell Dev. Biol.* **72**, 56–66 (2017).
46. McKeithan, W. L. *et al.* An Automated Platform for Assessment of Congenital and Drug-Induced Arrhythmia with hiPSC-Derived Cardiomyocytes. *Front. Physiol.* **8**, 766 (2017).
47. Tak, Y. G. *et al.* Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57–57 (2015).
48. Kikuchi, M. *et al.* Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med. Genomics* **12**, 128 (2019).
49. Grossman, S. R. *et al.* Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc. Natl. Acad. Sci.* **114**, E1291–E1300 (2017).
50. Heinz, S. *et al.* Effect of natural genetic variation on enhancer selection and function. *Nature* **503**, 487–492 (2013).
51. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
52. Seyres, D., Röder, L. & Perrin, L. Genes and networks regulating cardiac development and function in flies: Genetic and functional genomic approaches. *Brief. Funct. Genomics* **11**, (2012).
53. Gibson, G. Decanalization and the origin of complex disease. *Nat. Rev. Genet.* **10**, 134–140 (2009).
54. Ek, W. E. *et al.* Genetic variants influencing phenotypic variance heterogeneity. *Hum. Mol. Genet.* **27**, 799–810 (2018).
55. Zhou, S. *et al.* The Genetic Basis for Variation in Sensitivity to Lead Toxicity in Drosophila melanogaster. *Environ. Health Perspect.* **124**, 1062–1070 (2016).
56. Carbone, M. A. *et al.* Genetic architecture of natural variation in visual senescence in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6620–E6629 (2016).
57. Vogler, G. R-script to analyze fluorescent Drosophila melanogaster hearts.
58. https://github.com/kaustubhad/fastman.
59. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
60. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
61. Kim, H.-Y. Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors. *Restor. Dent. Endod.* **39**, 143–147 (2014).
62. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40**, e31–e31 (2012).
63. Worsley Hunt, R., Mathelier, A., Del Peso, L. & Wasserman, W. W. Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC Genomics* **15**, 472 (2014).
64. Castro-Mondragon, J. A., Jaeger, S., Thieffry, D., Thomas-Chollier, M. & van Helden, J. RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res.* **45**, e119 (2017).
65. Weirauch, M. T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431–1443 (2014).
66. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
67. Zhu, L. J. *et al.* FlyFactorSurvey: a database of Drosophila transcription factor binding

specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39**, D111-117 (2011).
68. Nguyen, N. T. T. *et al.* RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res*. **46**, W209–W214 (2018).
69. Cunningham, T. J. *et al.* Id genes are essential for early heart formation. *Genes Dev.* **31**, 1325–1338 (2017).
70. Burridge, P. W., Holmström, A. & Wu, J. C. Chemically Defined Culture and Cardiomyocyte Differentiation of Human Pluripotent Stem Cells. *Curr. Protoc. Hum. Genet.* **87**, 21.3.1-21.3.15 (2015).
71. Pei, F. *et al.* Chemical-defined and albumin-free generation of human atrial and ventricular myocytes from human pluripotent stem cells. *Stem Cell Res.* **19**, 94–103 (2017).
72. Cerignoli, F. *et al*. High throughput measurement of Ca^{2+} dynamics for drug risk assessment in human stem cell-derived cardiomyocytes by kinetic image cytometry. *J. Pharmacol. Toxicol. Methods* **66**, 246–256 (2012).

Table 1: Quantitative genetics of cardiac traits in the DGRP.

Summary statistics over all DGRP genotypes assayed. Number of lines and individuals (after outlier removal, see Materials and Methods) analyzed for each cardiac trait is indicated. Mean, standard deviation (Std dev.) and coefficient of variation (Coef. Var) among the whole population are indicated. Genetic, environment and phenotypic variance (respectively Genet. var, Env. var and Phen. var) were calculated for each trait. Broad sense heritability of traits means (H2) suggested heritability of corresponding traits. Levene test indicated significant heterogeneity of the variance among the lines.

	Diastolic Intervals	**Systolic Intervals**	**Heart Period**	**Diastolic Diameter**	**Systolic Diameter**	**Fractional Shortening**	**Arhythmia Index**
total.nb.lines	167	167	167	167	167	167	167
mean	0.4638	0.2166	0.6883	79.4200	51.0500	0.3538	0.2475
Std dev.	0.26330	0.03216	0.27690	14.09000	9.49300	0.06837	0.29230
Coef. var	0.5677	0.1485	0.4022	0.1774	0.1860	0.1933	1.1810
lines (mean)	165	166	165	159	157	158	166
Indiv. (mean)	1914	1911	1920	1779	1753	1767	1832
lines (Cve)	165	166	165	159	157	158	166
Indiv. (Cve)	1914	1911	1920	1779	1753	1767	1832
Genet. var	2.59e-02	5.03e-04	2.87e-02	1.13e+02	4.39e+01	1.57e-03	2.21e-02
Env. var	4.36e-02	5.35e-04	4.82e-02	8.64e+01	4.65e+01	3.11e-03	6.35e-02
Phen. var	6.95e-02	1.04e-03	7.68e-02	1.99e+02	9.04e+01	4.68e-03	8.56e-02
H2	0.373	0.485	0.373	0.566	0.485	0.335	0.258
F value	7,6864	11,686	7,4715	4,6950	15,041	11,164	6,5308
Pr(F)	8,8e-120	2,3e-187	5,8e-116	7,1e-62	1,9e-231	8,8e-175	1,8e-96
Levene test	1,9 e-10	1,9 e-10	1,7 e-08	1,6 e-05	2,1 e-13	1,6 e-05	2,1 e-13

Table 2: Conserved genes associated with natural variations of cardiac traits from flies to humans.

Enrichment analyses for genes conserved in human and for genes whose human ortholog is associated with either coronary artery diseases (CAD) or with cardiac disorders. Numbers in parenthesis indicate the total number of genes in each analysed gene set (whole fly genome/ mean and CVe GWAS associated genes). First row (Human orthologue): Number of fly genes that display a human orthologue according to DIOPT (high and moderate rank). Second and third rows:Number of genes whose human orthologue (high and moderate rank) has been associated with CAD or cardiac disorders by GWAS in human populations (Table S10, Table S11). FC: Ratio between expected (based on successes observed on fly genome) and observed number of successes in respective GWAS gene sets. pval: hypergeometric p value.

	Fly Genome (17500)	**GWAS mean (562)**	**FC**	**pval**	**GWAS Cve (566)**	**FC**	**pval**
Human orthologue	9463	410	1,4	$4,3\ 10^{-21}$	379	1,25	$7\ 10^{-11}$
CAD	321	19	1,35	0,04	16	1,25	0,07
Cardiac disorders	944	68	1,66	$7\ 10^{-6}$	64	1,7	$7\ 10^{-6}$

Figure legends

Figure 1:

Quantitative genetics and GWAS for cardiac traits in the DGRP.

(A) Left: Cardiac performance traits were analyzed in 167 sequenced inbred lines from the DGRP population. Approximately 12 females per line were analyzed. Right panels: Schematic of the *Drosophila* adult heart assay and example of M mode generated from video recording of a beating fly heart. Semi intact preparations of 1 week old adult females were used for high-speed video recording followed by automated and quantitative assessment of heart size and function. The representative M-Mode trace illustrate the cardiac traits analyzed. DI: diastolic interval; SI: systolic interval; HP: heart period (duration of one heartbeat); EDD: End diastolic diameter (fully relaxed cardiac tube diameter); ESD: end systolic diameter (fully contracted cardiac tube diameter). Fractional shortening (FS =EDD-ESD)/EDD) and Arrythmia Index (AI = Std-Dev (HP)/HP) were additionally calculated and analyzed. (B) Distribution of line means and within lines variations (box plots) from 167 measured DGRP lines for HP and EDD. DGRP lines are ranked by their increasing mean phenotypic values. For both phenotypes, representative M-Modes from extreme lines are shown below (other traits are displayed in Figure S1). (C) Pearson residuals of chi-square test from the comparison of indicated SNPs categories in the DGRP and among variants associated with cardiac traits. According to DGRP annotations, SNPs are attributed to genes if they are within the gene transcription unit (5' and 3' UTR, synonymous and non-synonymous coding, introns) or within 1kb from transcription start and end sites (1kb upstream, 1 kb downstream). NA: SNPs not attributed to genes (> 1kb from TSS and TES). (D) Comparison of gene sets identified by single marker using Fast-LMM (LMM) and in interaction using FastEpistasis (Epistasis). The Venn diagram illustrates the size of the two populations and their overlap. (E) Overlap coefficient of gene sets associated with the different cardiac traits analyzed.

Figure 2:

Functional annotations and validations of genes associated with GWAS for cardiac performance.

(A) GO enrichment analyses. Selected molecular functions (MF, Left) and cellular components (CC, right) associated with cardiac performances at FDR < 0.05 are shown. Enrichment analysis was performed using G:profiler with a correction for multitesting (see

Materials and Methods) (B) Interaction network of genes associated with natural variations of cardiac performance. Direct genetic and physical interactions between cardiac fly GWAS genes are displayed. Nodes represent genes and/or proteins, edges represent interactions (blue: genetic; black: physical). Node shapes refer to Single marker and/or epistasis GWAS, node color to the cardiac performances phenotype(s) for which associations were established. Genes and proteins highlighted in pink point to transcription factors, in green and blue to signaling pathways (FGF and TGFb, respectively) and in yellow to ion channels. (C) Heat map representing the effects on indicated cardiac traits of heart specific RNAi mediated knockdown of 42 genes identified in GWAS for cardiac performance. Results of Wilcoxon rank sum test of the effects of indicated heart-specific RNAi mediated gene KD (rows) for cardiac performance traits (columns), analyzed on semi-intact 1week females flies. Detailed data are presented in Dataset S5. 38 (out of 42) genes tested lead to significant effects on cardiac performance traits upon KD. Black dots indicate the trait(s) for which the corresponding gene was associated in GWAS. ns: not significant; *: pval< 0,05; **: pval<0,01; ***: pval< 0,005; ****pval<0,0001 (p values were adjusted for multiple testing using Bonferroni correction). Comparison with heart specific effect of random selected genes is displayed in Figure S3. (D) Schematic drawing of BMP and Activin pathways in *Drosophila*. (E) Genetic interactions between BMP and Activin pathway genes. Genetic interactions tested between $snoo^{BSC234}$ and sog^{U2} for SI and between dpp^{d14} and $babo^{32}$ for FS (other phenotypes are shown in Figure S4). Cardiac traits were measured on each single heterozygotes and on double heterozygotes flies. Two-way ANOVA reveals that the interaction between $snoo^{BSC234}$ and sog^{U2} for SI and between dpp^{d14} and $babo^{32}$ for FS are significant. Detailed data for interaction effect corresponding to all phenotypes are displayed in Figure S4.

Figure 3:

<u>Quantitative genetics and GWAS for cardiac traits CVe in the DGRP.</u>

(A) Distribution of ESD CVe values among DGRP lines. Lines are arranged by increasing CVe values (other phenotypes are displayed in Figure S6). (B) Distribution of ESD values across individuals for two representative DGRP lines with low and high intra-genotypic variability. Each dot represents ESD value of a single fly. (Individual values scaled by line mean). (C) Venn diagram illustrating the overlap between SNPs (left) and genes (right) associated with mean (yellow) and CVe traits (pink). While only 3 SNPs were retrieved from GWAS of both mean and CVe traits, genes associated with CVe traits are largely overlapping

with those associated with mean traits. (D) Miami plots showing the results of GWAS performed on mean of ESD (top, red) and ESD CVe (bottom, blue). No overlap between sets of variants is observed. Miami plots for other traits are displayed in Figure S7). (E) Pearson residuals of chi-square test from the comparison of indicated SNPs categories in the DGRP and among variants associated with cardiac traits CVe (see legend to Figure 1C for detailed description of SNPs categories). (F) GO enrichment analyses of genes associated with CVe traits. Selected molecular functions and cellular components associated with variance of cardiac performances at FDR < 0.05 are shown. (G) Logo representation of Position Specific Scoring Matrices (PSSMs) identified in corresponding sets of non-coding variants associated with cardiac traits CVe. Potential cognate TFs indicated below were identified by comparing PSSMs to TF binding sites databases.

Figure 4:

<u>*In vivo* and *in vitro* assays for *poxm*/Pax9 and *sr*/Egr2 in flies and human-iPSC-CM</u>

(A) Heart specific KD of *poxm/Pax9* and *sr/Egr2* lead to increased heart rate in flies. Cardiac specific RNAi knock-down phenotypes of tested TFs. Hand>Gal4; UAS RNAi 1week adult females (n>30 per cross) were analyzed for cardiac performance using the *in vivo* assay on intact flies. Scatter plots of individual data, mean and SD are represented. Wilcoxon test results are indicated (***: pval< 0,005). Data were normalized to genotype matched controls. Right panel: representative M modes (5 s) of RNAi KD and their respective control illustrating the increased heart rate observed upon *poxm* and *sr* inactivation. Number of heartbeats counted from M modes are indicated.

(B-I) Effects of Pax9 and Egr2 siRNA on human derived iPSC-CM action potentials. (B) Schematic overview of single cell and high throughput voltage assay. (C) Two-dimensional graph for APD75 and Kolgomorov Smirnov distance (KS-D) representing screen results for PAX9 and EGR2 knockdown. (D-E) Population distribution of APD75 measurements for siPAX9 and siEGR2 vs siControl-transfected hiPSC-ACMs respectively. (F) Median action potential traces for siPAX9, siEGR2 and siControl-transfected hiPSC-ACMs. (G-I) Histograms showing median APD50, APD75 and APD90 for siPAX9, siEGR2 and siControl transfected-hiPSC-ACMs. T-test was used to calculate p-values. *p<0.05, ***p<0.001, ****p<0.0001.

Supplemental material

Figure S1:

Quantitative genetics of cardiac traits means.

(A) Distribution of line means and within lines variations (box plots) from 167 measured DGRP lines for Diastolic Interval (DI), Systolic Interval (SI), Arrhythmia Index (AI), End Systolic Diameter (ESD) and Fractional shortening (FS). DGRP lines are ranked by their mean phenotypic values. (B) Overall correlations between phenotypes in the DGRP population. Spearman correlations were calculated based on average of line means for each cardiac trait.

Figure S2:

Quantile-Quantile plots of association P-values

QQ plot display the quantile distribution of observed marker-phenotype association P-values versus the distribution of expected P-values assuming the null hypothesis (uniformly distributed). The first part shows the QQ plot for trait means and the second part shows the QQ plot for trait CVe's. QQ plots fail to capture deviation from uniformity for most of the phenotypes.

Figure S3

Heart specific RNAi mediated knockdown of random selected genes

(A) Heat map representing the effects on indicated cardiac traits of heart specific RNAi mediated knockdown of 20 genes selected randomly (below the top 100 threshold for reporting associations). Results of Wilcoxon rank sum test of the effects of indicated heart-specific RNAi mediated gene KD (rows) for cardiac performance traits (columns), analyzed on semi-intact 1week females' flies.

(B) For each phenotype and each gene, Z-scores summarizes the difference in the phenotype distribution from tested genes to their respective control (detailed data are in Dataset S5)). The density curve highlighted in pink shows the distribution of the absolute Z-scores of all the phenotypes for the GWAS associated genes and the density curve in blue highlights the corresponding distribution for randomly selected genes. Using the Mann-Whitney-Wilcoxon Test, we showed that the Z-scores of

GWAS associated genes were significantly different (higher) from the randomly selected genes.

Figure S4:

Genetic interactions between BMP and Activin pathway genes.

A. Cardiac performances of controls, *snoo*^{BSC234}, *sog*^{U2} heterozygotes, and *snoo*^{BSC234}; *sog*^{U2} double heterozyotes. Box plots and wilcoxon test results are indicated (*: pval< 0,05; **: pval<0,01; ***: pval< 0,005; ****: pval<0,0001). n: number of individual flies analyzed for each genotype. p-values from 2-way ANOVA, testing the significance of the genetic interaction between *snoo*^{BSC234} and *sog*^{U2}, are indicated for each phenotypes.

B: Cardiac performances phenotypes of heterozygotes for *babo*^{32} and *dpp*^{d14} and of *babo*^{32}; *dpp*^{d14} double heterozyotes. Box plots and wilcoxon test results are indicated (*: pval< 0,05; **: pval<0,01; ***: pval< 0,005; ****: pval<0,0001). n: number of individual flies analyzed for each genotype. The 2-way ANOVA p-values testing the significance of the genetic interaction between *babo*^{32} and *dpp*^{d14} are reported.

Figure S5:

TFBS predictions from sequences surrounding non-coding variants and *in vivo* validation of cognate TFs:

(A) Schematic representation of the work flow used for motifs analysis in the vicinity of the non-coding variants and comparison with TF binding sites databases. (B) Logo representation of Position Specific Scoring Matrices (PSSMs) identified in corresponding sets of non-coding variants. Potential cognate TFs are indicated and were identified by comparing PSSMs to TF binding sites databases. (C) Cardiac specific RNAi knock-down phenotypes of TFs tested *in vivo*. Hand>Gal4; UAS RNAi 1-week adult females (n>30 per cross) were analyzed for cardiac performance in intact flies. Scatter plots of individual data, mean and SD are represented. Wilcoxon test results are indicated (*: pval< 0,05; ** : pval<0,01; ***: pval< 0,005 ; ****pval<0,0001). Data were normalized to genotype matched controls. Raw data are presented in Dataset S5. Detailed statistical analyses are in Table S7.

Figure S6:

Quantitative genetics of cardiac traits CVe.

(A) Barplot distribution of HP, DI, SI, AI, EDD and FS CVe values among DGRP lines. Lines are arranged by increased CVe values. (B) Overall correlations between CVe phenotypes. Spearman correlations were calculated on line CVe for each cardiac trait. (C) Overall correlations among mean and CVe phenotypes. Spearman correlations were calculated among line mean and line CVe for each cardiac trait.

Figure S7:

Comparison of GWAS results for mean and CVe cardiac traits.

Miami plots showing the results of GWAS performed on mean (top, red) and CVe (bottom, blue) of indicated cardiac traits. The arrows point the *Ca-β* SNPs associated to both mean FS and CVe FS.

Figure S8:

Calcium transient measured in iPSC-ACM upon Pax9 and EGR2 siRNA-mediated KD.

(A) Schematic overview of single cell and high throughput calcium assay. (B) Two-dimensional graph for CTD75 and Kolgomorov Smirnov distance (KS-D) representing screen results for PAX9 and EGR2 knockdown. (C-D) Population distribution of CTD75 measurements for siPAX9 and siEGR2 vs siControl-transfected hiPSC-ACMs respectively. (E) Two-dimensional graph for Peak Frequency measurements and Kolgomorov Smirnov distance (KS-D) representing screen results for PAX9 and EGR2 knockdown. (F-G) Population distribution of Peak Frequency measurements for siPAX9 and siEGR2 vs siControl-transfected hiPSC-ACMs respectively. (H) Median action potential traces for siPAX9, siEGR2 and siControl-transfected hiPSC-ACMs. (I-O) Histograms showing median CTD50, CTD75, CTD90, Peak frequency, Peak count, Peak rise time and Peak decay time for siPAX9, siEGR2 and siControl transfected-hiPSC-ACMs. T-test was used to calculate p-values. *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.

Table S1:

Block effect evaluation

Tests for block effect within the 14 DGRP lines analyzed twice

Table S2:

Summary table of variants with MAF>4% mapped to genes in single marker GWA analyses for traits means

- Sheet 'Variants_MAF4%_merged': Summary table of variants with MAF>4% identified by single marker GWAS for the 7 traits, and their associated genes.
- Sheet 'mergesummary_genes': Summary table of genes associated with cardiac traits.
- Sheet 'GeneOverlap_between_GWAS': Analysis of genes found in at least two independent GWA analyses.

Table S3:

Summary table of genes associated with SNPs identified by FastEpistasis for traits means. Fly Base gene numbers (FBgn), gene symbol and corresponding quantitative traits are indicated. Gene Ontology (Molecular function) annotations are provided. The 31 genes also found in single marker GWA (Table S2) are labeled in red.

Table S4:

Summary of all trait mean GWA analyses associated genes. Summary table of genes associated with cardiac traits means, computing data from Fast LMM (single marker GWA) and Fast -Epistasis (Epistasis GWA). Fly Base gene numbers (FBgn), gene symbol and method from which the associated gene was identified are indicated. Corresponding quantitative traits and Gene Ontology (Molecular function) annotations are provided. The 31 genes found in both single marker and epistasis GWA are labeled in red.

Table S5:

Enriched annotations within GWA associated genes traits means.

- gProfiler_allGWAS: Functional enrichment analysis performed on the 562 genes from FastLMM and FastEpistasis analyses, all traits merged (see Table S4), using g:GOSt from G profiler (https://biit.cs.ut.ee/gprofiler/gost). Parameters used were as follow: Background: all annotated genes / Selected GO terms size >5 and <500.
- Enrichment_Neely_et_al: Enrichment analysis (hypergeometric distribution calculations) for genes identified in a global RNAi screen for heart function[13].

Table S6

Principal network Component (PNC) formed by genes and gene products from cardiac traits means GWAS (refers to figure 2B).

- Sheet 1: Details of the 148 genes forming the principal network components of direct genetic and physical interactions among the 562 cardiac performances associated genes.

Gene symbol, FBgn and cardiac trait(s) for which they have been associated are indicated, together with corresponding GO annotations. Annotations for signaling pathways (in biological processes) are highlighted in yellow, annotations for transcription factors (in molecular function) are highlighted in blue.

- Sheet 2: Details on the interactions depicted in figure 2B.

Table S7:

Statistics of Transcription factors RNAi cardiac knockdown effects on heart performance traits.

TFs were inactivated in the heart by hand-Gal4 and heart parameters monitored *in vivo* by tdtk tomato high speed video recording (see methods). Graphic representation of the data is displayed in Figure S5C. Wilcoxon-Mann-Whitney rank sum test results for indicated cardiac parameters are shown.

Table S8:

Summary of all trait CVe GWA analyses associated variants and genes.

- merge_allGWAS_Cve_annotatedGene: Summary table of genes associated with cardiac traits CVe, computing data from Fast LMM (single marker GWA) and Fast-Epistasis (Epistasis GWA). Raw data are provided in Dataset S7. Fly Base gene numbers (FBgn), gene symbol and method from which the associated gene was identified are indicated. Corresponding quantitative traits and Gene Ontology (Molecular function) annotations are provided. The genes found in both single marker and epistasis GWA are labeled in red.
- overlap Cve_Mean (genes): List of the 92 genes found associated with both Mean and CVe of cardiac performance traits
- overlap Cve_Mean (Variants): Lists of all variants associated with CVe and Mean identified by FastLMM and FastEpistasis and overlap between both lists.

Table S9:

Enriched annotations within genes associated with GWA on CVe

- gProfiler_allGWAS_CVe: Functional enrichment analysis performed on the 566 genes from FastLMM and FastEpistasis analyses, all CVe traits merged (Table S8), using g:GOSt from G profiler. Parameters used were as follow: Background: all annotated genes / Selected GO terms size >5 and <500.

enrichment_Neely_et_al: Enrichment analysis (hypergeometric distribution calculations) for genes identified in a global RNAi screen for heart function[13].

Table S10:

Human genes found associated with coronary artery diseases (CAD) and cardiac traits in GWAS studies, and their fly orthologues.

- Sheet 1 (Human_CAD_genes): Human Gene ID, SNP ID (when available) and original article describing the association are indicated.
- Sheet 2 (CAD-genes_fly_orthologs): *Drosophila* orthologues of human CAD genes identified using the DIOPT software.
- Sheet 3 (human_cardiac_traits_genes): Human Gene ID, SNP ID (when available), quantitative trait analyzed for association, and original article describing the association are indicated.
- Sheet 4 (Cardiactraitsgenes_fly_orthologs): *Drosophila* orthologues of human cardiac traits genes identified using the DIOPT software.

Table S11:

Human orthologues of fly GWAS genes, and enrichment for genes associated with cardiac pathologies in humans.

Drosophila genes identified in GWAS (from Table S4 and S8), corresponding human orthologues as determined using DIOPT (rank high and moderate) and enrichment calculations with human genes associated with CAD and cardiac disorders (from Table S10)

- Sheet 1 (mergeMean_human_orthologues): Fly cardiac Mean GWAS genes and corresponding human orthologues identified with DIOPT.
- Sheet 2 (mergeMean_human_Heart_GWAS_orth): Fly cardiac Mean GWAS genes and corresponding human orthologues associated with cardiac disorders.
- Sheet 3 (MergeMean_human_CAD_GWAS_orthol): Fly cardiac MeanGWAS genes and corresponding human orthologues associated with CAD
- Sheet 4 (mergeCVe_human_orthologues): Fly cardiac CVe GWAS genes and corresponding human orthologues identified with DIOPT.
- Sheet 5 (mergeCVe_human_Heart_GWAS_orth): Fly cardiac CVe GWAS genes and corresponding human orthologues associated with cardiac disorders.
- Sheet 6 (MergeCVe_human_CAD_GWAS_orthol): Fly cardiac Cve GWAS genes and corresponding human orthologues associated with CAD

- Sheet 7 (Enrichments): headcounts of Fly Mean GWAS genes for which human orthologues have been identified in Human GWAS studies (moderate and high rank), and hypergeometric distribution calculations.

Datasets

A repository containing all the original raw data (individual phenotypes of DGRP lines/variants identified by GWAS/individual phenotypes from validation experiments) and large-scale datasets used for analyses (PPI and genetic interactions/regulatory variants) is available here:
https://doi.org/10.5281/zenodo.5582846

Dataset S1:

Individual values for cardiac traits analyzed across the 167 DGRP lines. Individual and DGRP line number are indicated. Phenotypic values were determined from high-speed video recording on dissected flies and movie analysis using SOHA[14].

Dataset S2:

Variants identified by FastLMM as associated to indicated phenotypes. Among the 100 best ranked associations, only variants with MAF>4% were retained. Tables for variants mapped to genes and for variants that are not within gene mapping criteria (>1kb from TSS and TTS) are indicated.

Dataset S3:

All Fast-Epistasis data on mean phenotypes, per quantitative trait: SNPs ID, position, associated genes and statistics are indicated for both focal SNPs (left) and their interacting SNPs (right). Each sheet displays the results for indicated quantitative traits, except for the first one which is a merge of all quantitative traits association analyses.

Dataset S4:

Collection of physical (IP, Y2H) and genetic interactions identified in *Drosophila*.

Dataset S5:

Data from validation experiments. (Heart specific RNAi validations and tests for genetic interactions among BMP members).

Dataset S6:

Variants in non-coding regions used for motif discovery, associated with EDD, ESD and FS ('structure') and to AI, DI, SI and HP ('rhythm') for both mean and CVe of cardiac traits.

Dataset S7:

- Sheets 'phenotypes_CVe': Variants among the 100 best ranked associations for each cardiac trait, with information about the associated genes and the number of DGRP lines in which they are present. Variants among the 100 best ranked associations for each cardiac trait that are not within gene mapping criteria (>1kb from TSS and TTS) are shown below.
- sheets 'FastEpistasis_CVe_...': All Fast-Epistasis data (on traits CVe), per phenotypes. SNPs ID, position, associated genes and statistics are indicated for both focal SNPs (left) and their interacting SNPs (right). Each sheet displays the results for indicated quantitative traits, except for the first one which is a merge of all quantitative traits association analyses.

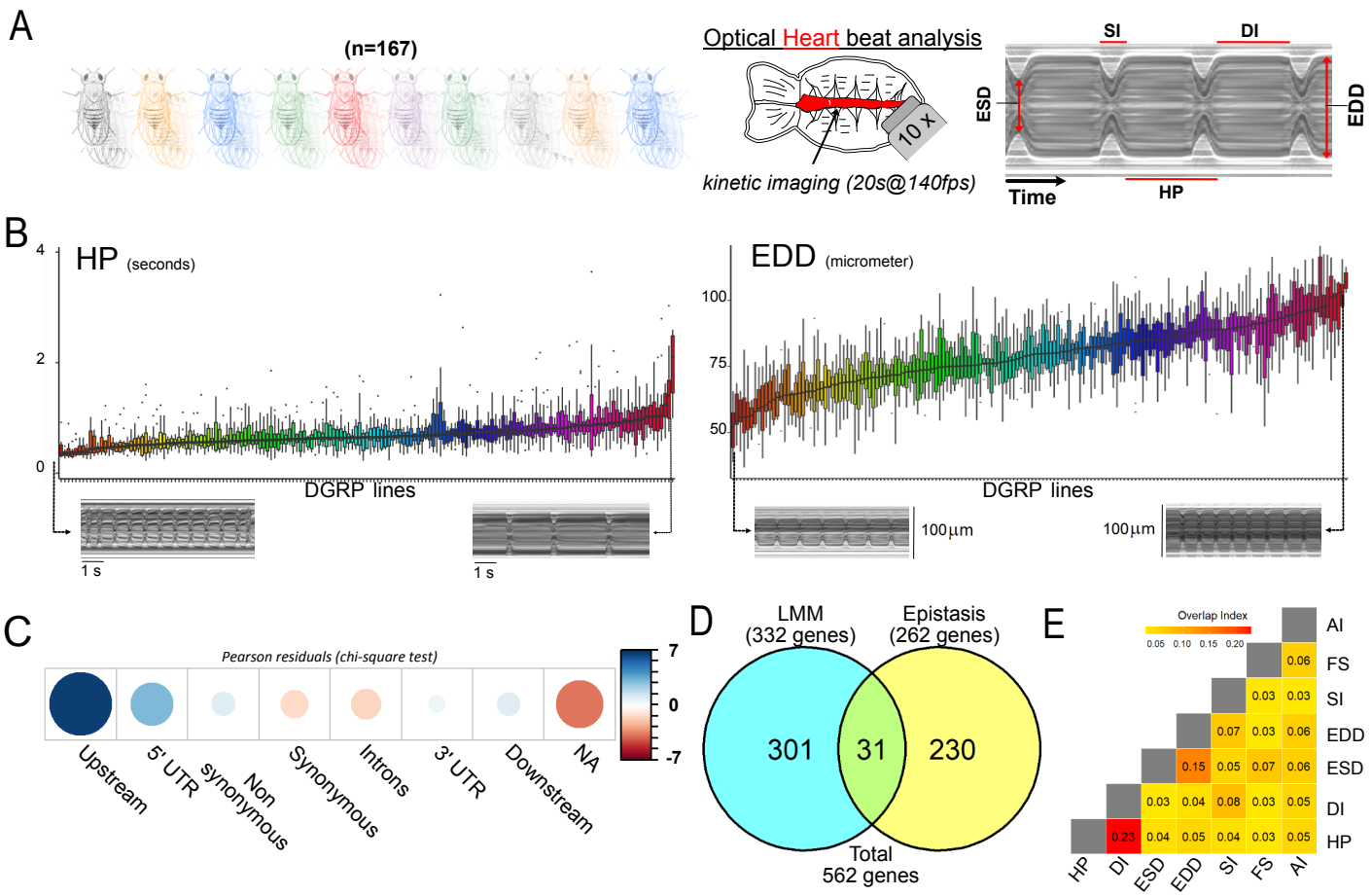

A
(n=167)
Optical Heart beat analysis
10 x
kinetic imaging (20s@140fps)
SI
DI
ESD
EDD
Time
HP
B
HP (seconds)
DGRP lines
1 s
EDD (micrometer)
100µm
C
Pearson residuals (chi-square test)
Upstream
5' UTR
Non synonymous
Synonymous
Introns
3' UTR
Downstream
NA
D
LMM (332 genes)
Epistasis (262 genes)
301
31
230
Total 562 genes
E
Overlap Index
AI
FS
SI
EDD
ESD
DI
HP

Figure 1

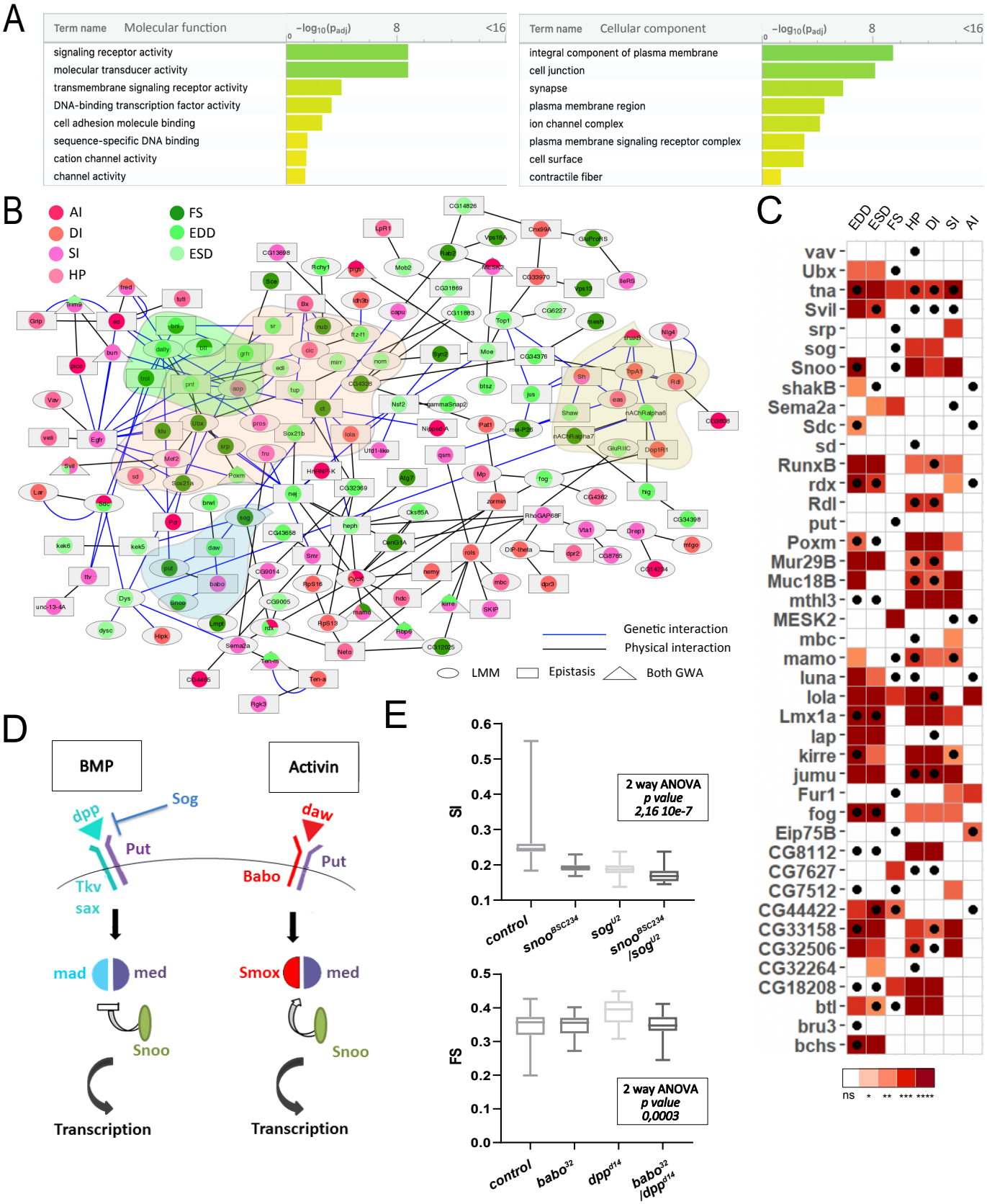

A
Term name
Molecular function
$-\log_{10}(p_{adj})$
0
8
<16
signaling receptor activity
molecular transducer activity
transmembrane signaling receptor activity
DNA-binding transcription factor activity
cell adhesion molecule binding
sequence-specific DNA binding
cation channel activity
channel activity
Term name
Cellular component
$-\log_{10}(p_{adj})$
0
8
<16
integral component of plasma membrane
cell junction
synapse
plasma membrane region
ion channel complex
plasma membrane signaling receptor complex
cell surface
contractile fiber
B
AI
DI
SI
HP
FS
EDD
ESD
Genetic interaction
Physical interaction
LMM
Epistasis
Both GWA
C
EDD
ESD
FS
HP
DI
SI
AI
vav
Ubx
tna
Svil
srp
sog
Snoo
shakB
Sema2a
Sdc
sd
RunxB
rdx
Rdl
put
Poxm
Mur29B
Muc18B
mthl3
MESK2
mbc
mamo
luna
lola
Lmx1a
lap
kirre
jumu
Fur1
fog
Eip75B
CG8112
CG7627
CG7512
CG44422
CG33158
CG32506
CG32264
CG18208
btl
bru3
bchs
ns
*
**
***
****
D
BMP
Activin
dpp
Sog
Put
Tkv
sax
daw
Babo
Put
mad
med
Smox
med
Snoo
Snoo
Transcription
Transcription
E
SI
0.6
0.5
0.4
0.3
0.2
0.1
2 way ANOVA
p value
2,16 10e-7
control
snoo^{BSC234}
sog^{U2}
snoo^{BSC234}/sog^{U2}
FS
0.5
0.4
0.3
0.2
0.1
0.0
2 way ANOVA
p value
0,0003
control
babo^{32}
dpp^{d14}
babo^{32}/dpp^{d14}

Figure 2

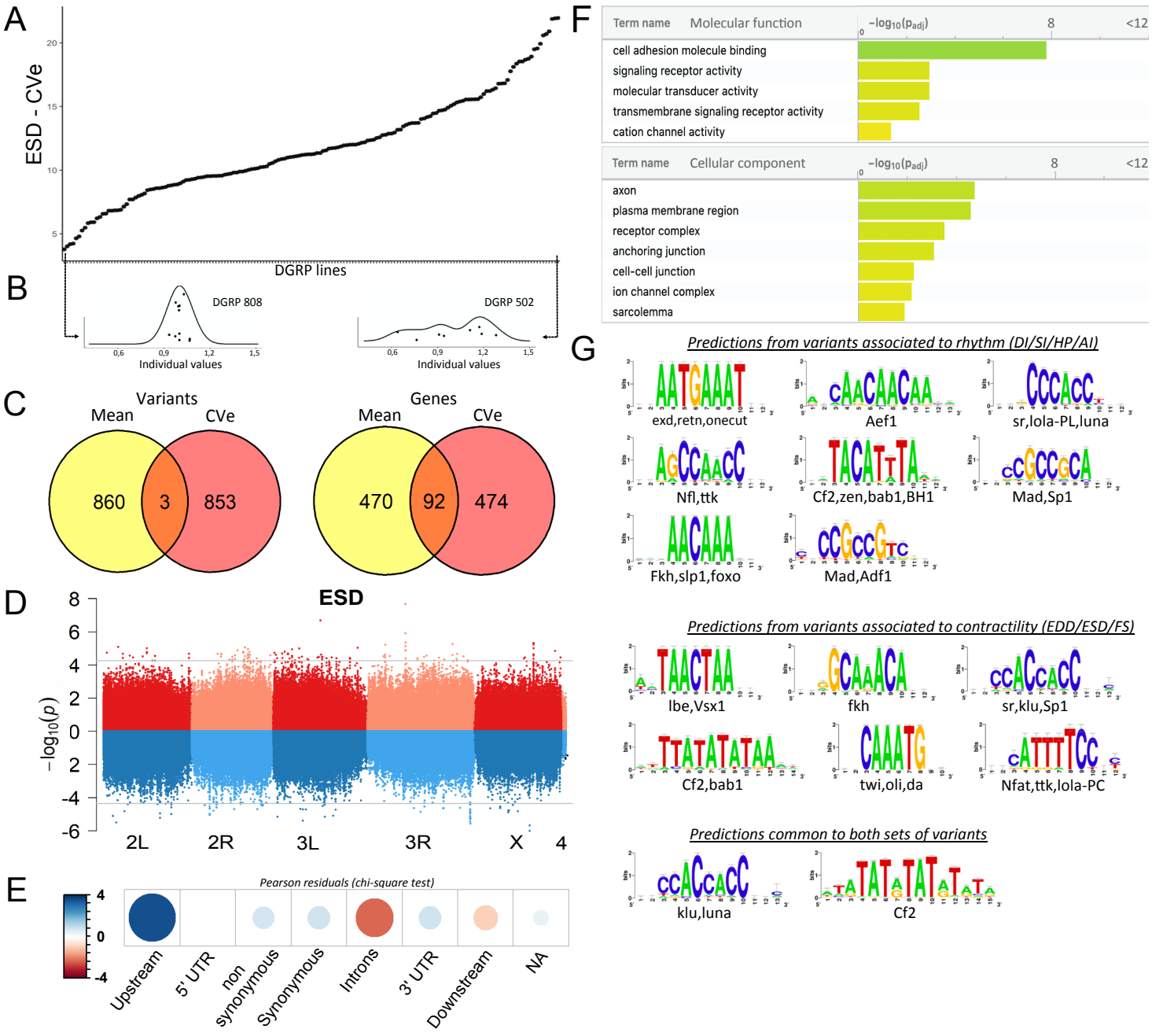

A
ESD - CVe
DGRP lines
B
DGRP 808
DGRP 502
Individual values
C
Variants
Mean
CVe
860
3
853
Genes
470
92
474
D
ESD
2L
2R
3L
3R
X
4
E
Pearson residuals (chi-square test)
Upstream
5' UTR
non synonymous
Synonymous
Introns
3' UTR
Downstream
NA
F
Term name
Molecular function
cell adhesion molecule binding
signaling receptor activity
molecular transducer activity
transmembrane signaling receptor activity
cation channel activity
Cellular component
axon
plasma membrane region
receptor complex
anchoring junction
cell-cell junction
ion channel complex
sarcolemma
G
Predictions from variants associated to rhythm (DI/SI/HP/AI)
exd,retn,onecut
Aef1
sr,lola-PL,luna
Nfl,ttk
Cf2,zen,bab1,BH1
Mad,Sp1
Fkh,slp1,foxo
Mad,Adf1
Predictions from variants associated to contractility (EDD/ESD/FS)
lbe,Vsx1
fkh
sr,klu,Sp1
Cf2,bab1
twi,oli,da
Nfat,ttk,lola-PC
Predictions common to both sets of variants
klu,luna
Cf2

Figure 3

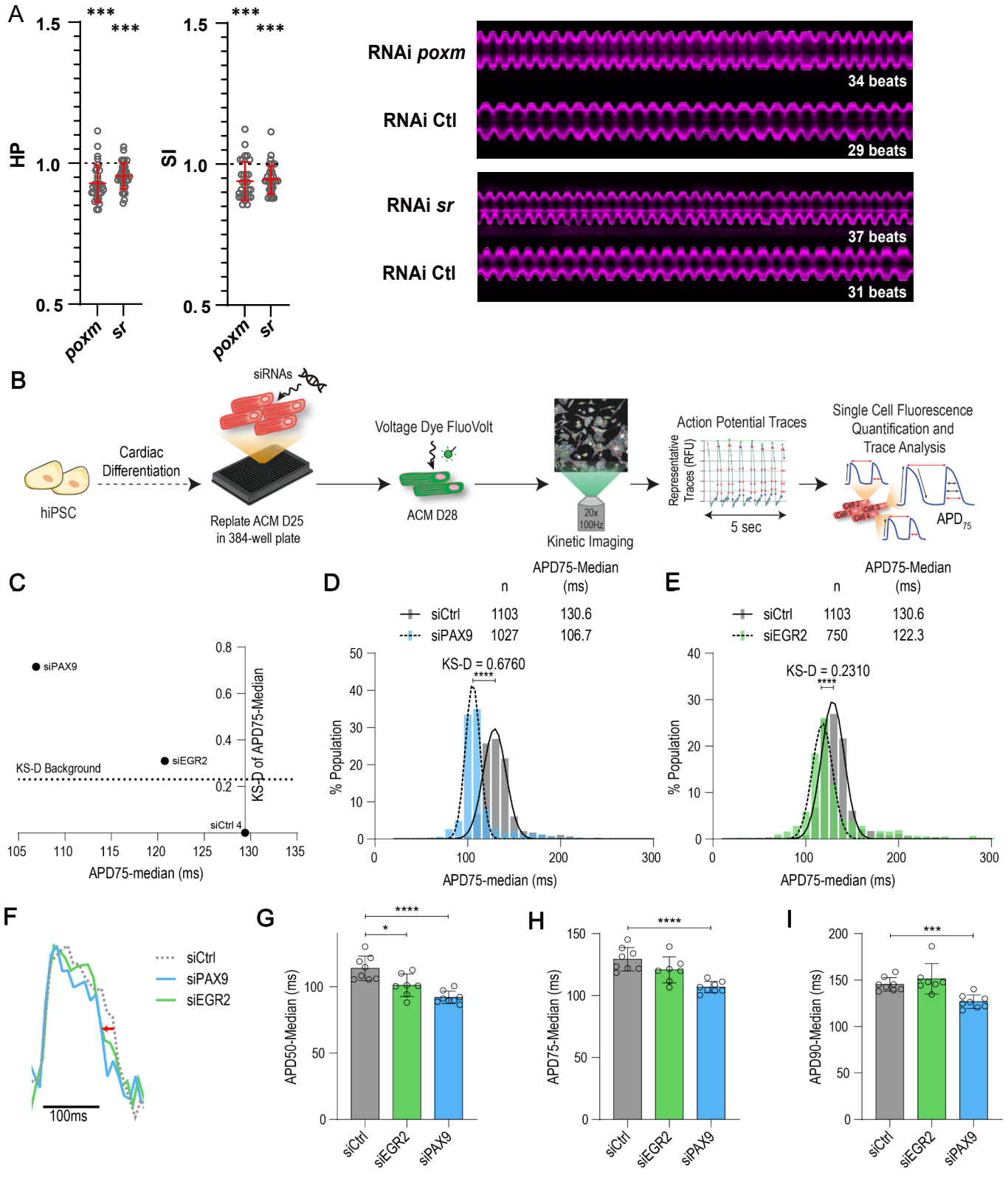

A
HP
SI
poxm
sr
RNAi poxm
RNAi Ctl
RNAi sr
RNAi Ctl
34 beats
29 beats
37 beats
31 beats
B
siRNAs
Cardiac Differentiation
hiPSC
Replate ACM D25 in 384-well plate
Voltage Dye FluoVolt
ACM D28
Kinetic Imaging
Action Potential Traces
Representative Traces (RFU)
5 sec
Single Cell Fluorescence Quantification and Trace Analysis
APD75
C
siPAX9
siEGR2
KS-D Background
siCtrl 4
KS-D of APD75-Median
APD75-median (ms)
D
n
APD75-Median (ms)
siCtrl 1103 130.6
siPAX9 1027 106.7
KS-D = 0.6760
% Population
APD75-median (ms)
E
n
APD75-Median (ms)
siCtrl 1103 130.6
siEGR2 750 122.3
KS-D = 0.2310
% Population
APD75-median (ms)
F
siCtrl
siPAX9
siEGR2
100ms
G
APD50-Median (ms)
siCtrl
siEGR2
siPAX9
H
APD75-Median (ms)
siCtrl
siEGR2
siPAX9
I
APD90-Median (ms)
siCtrl
siEGR2
siPAX9

Figure 4

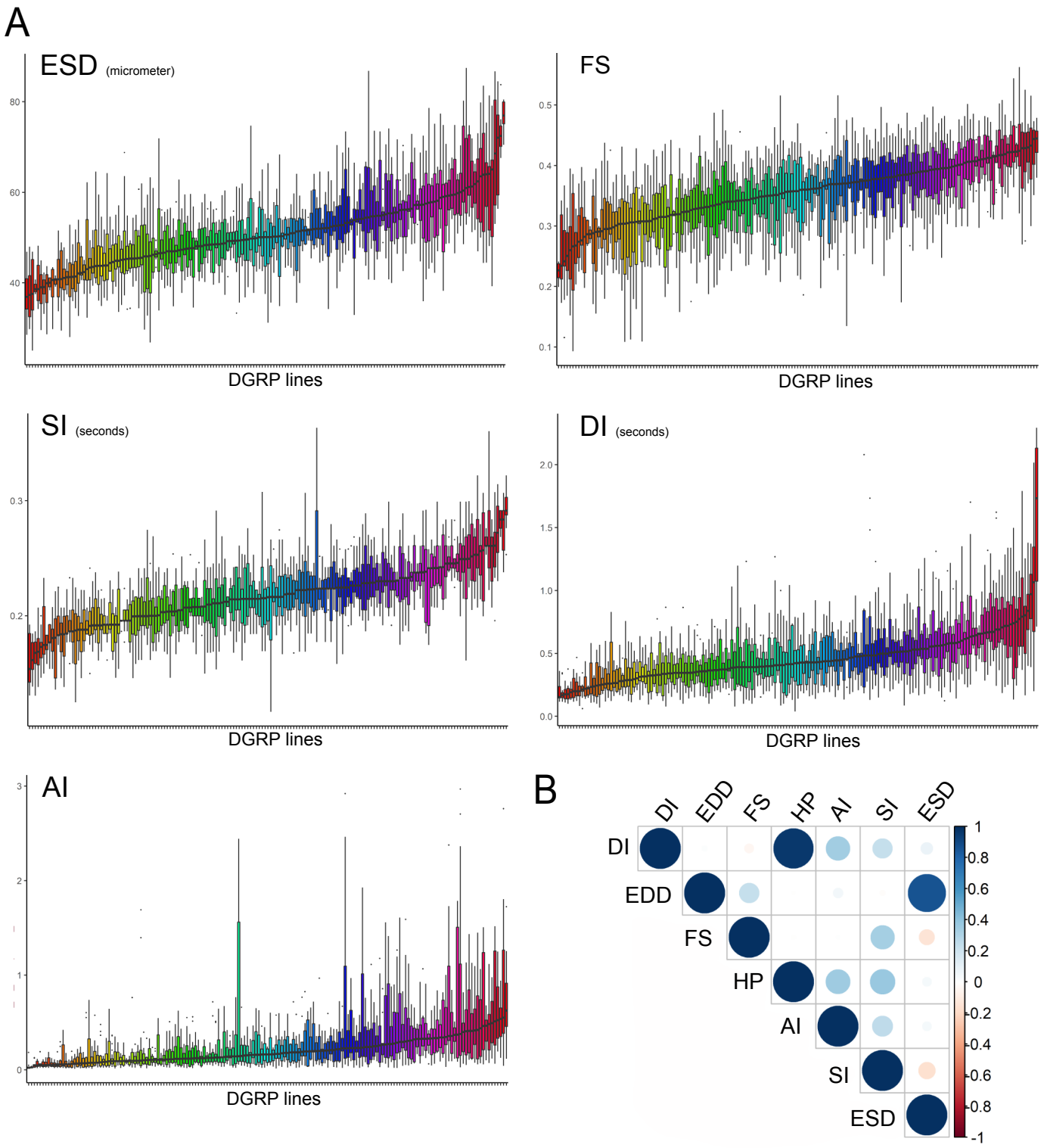
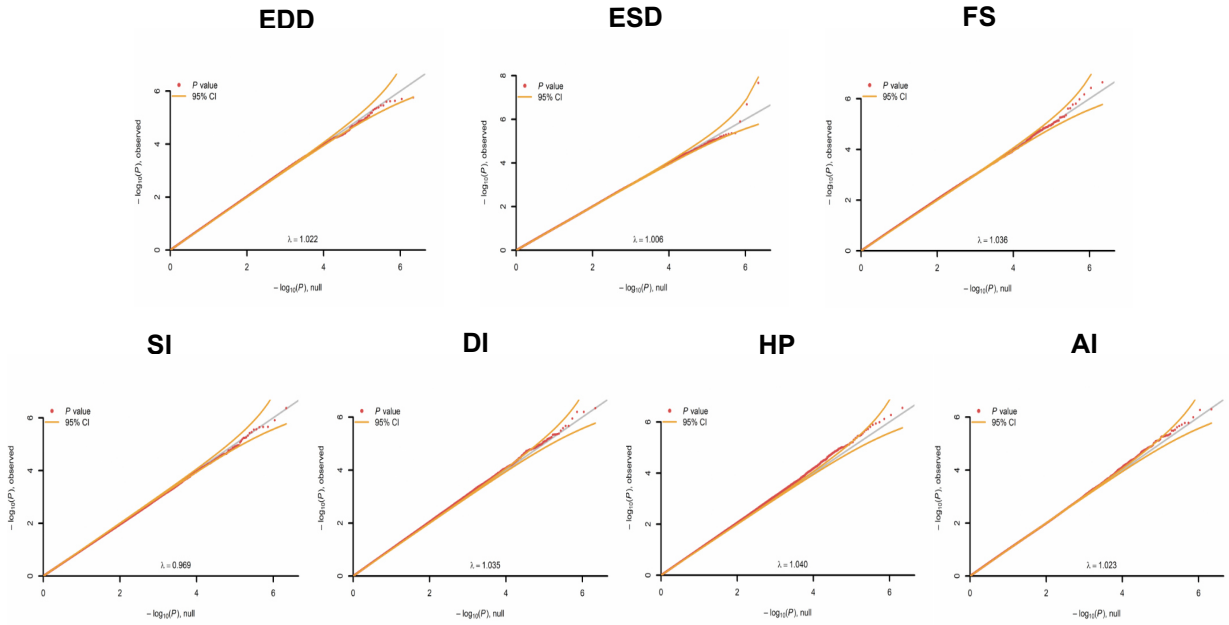

A
ESD (micrometer)
80
60
40
DGRP lines
FS
0.5
0.4
0.3
0.2
0.1
DGRP lines
SI (seconds)
0.3
0.2
DGRP lines
DI (seconds)
2.0
1.5
1.0
0.5
0.0
DGRP lines
AI
3
2
1
0
DGRP lines
B
DI
EDD
FS
HP
AI
SI
ESD
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Figure S1

A

QQ plots for GWAS on trait means



B

QQ plots for GWAS on trait CVe

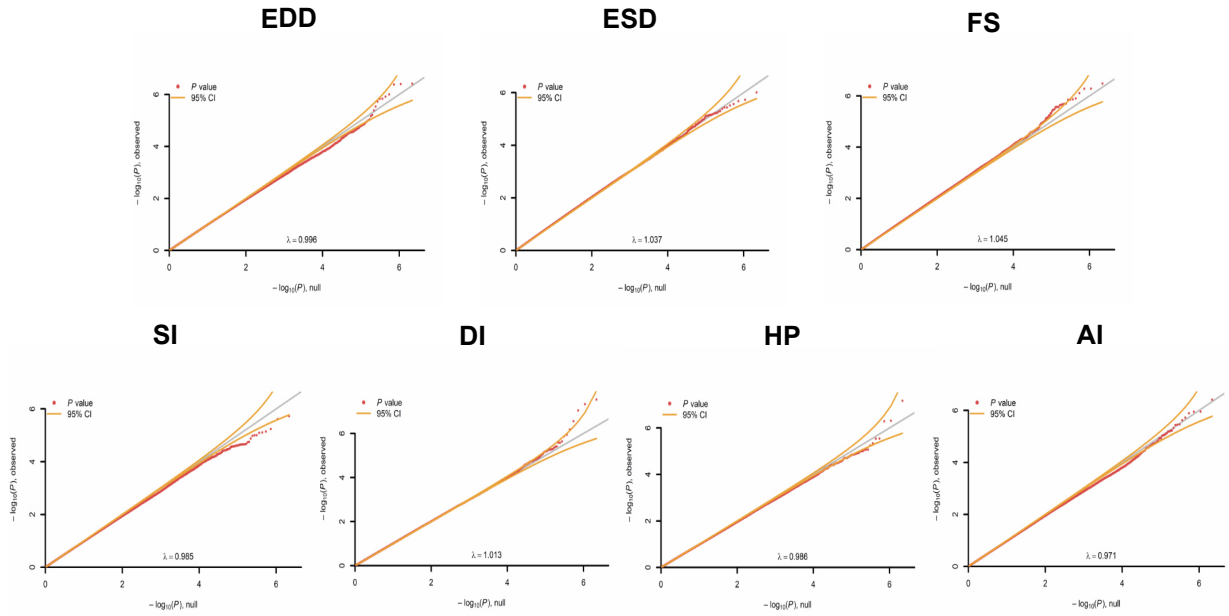


Figure S2

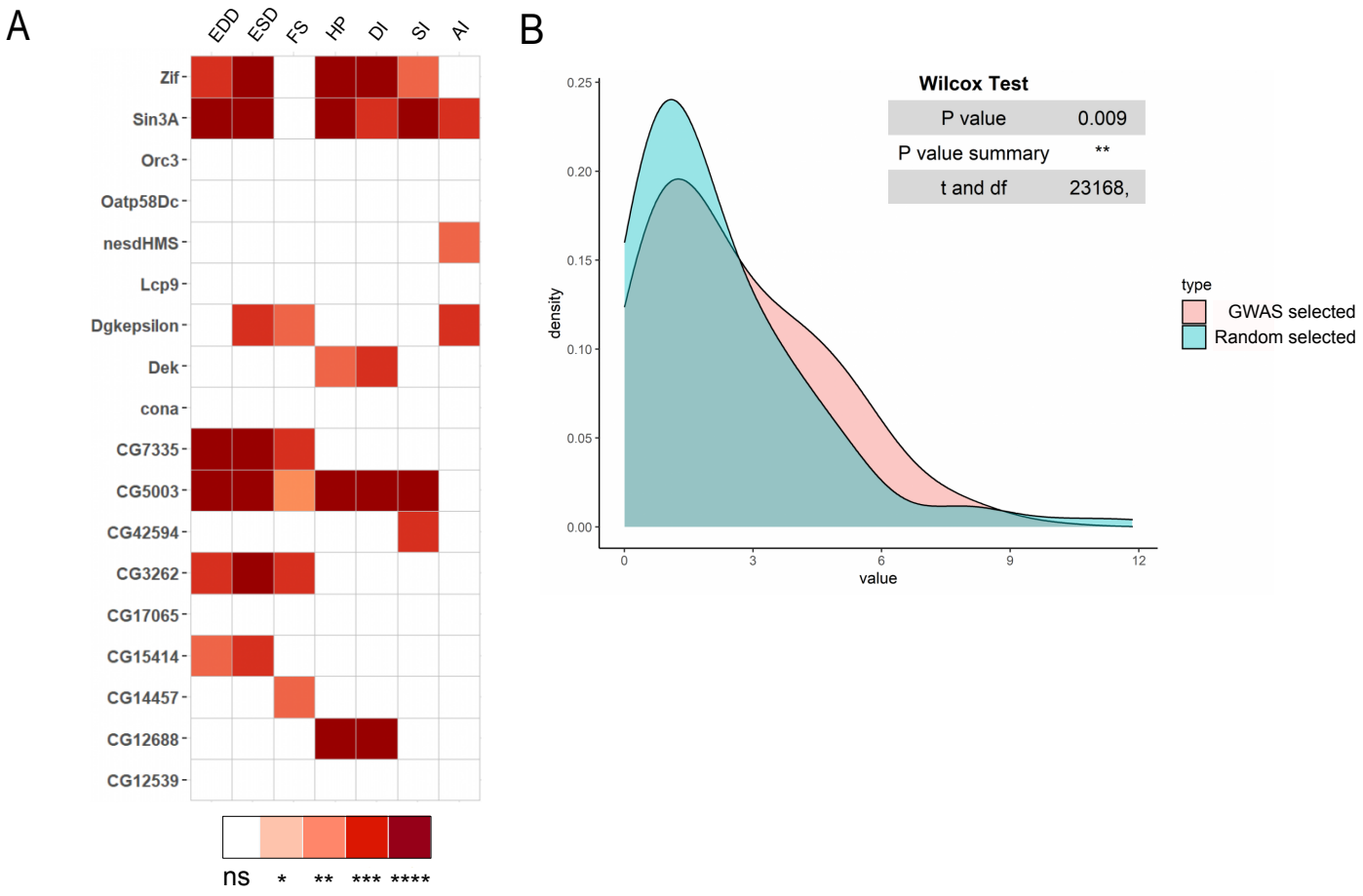
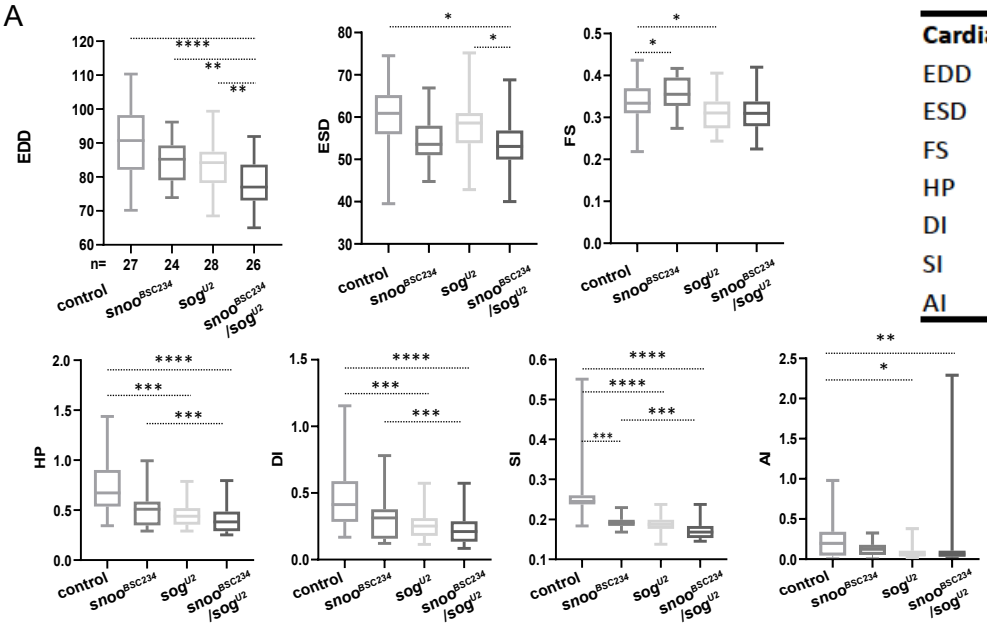

A
EDD
ESD
FS
HP
DI
SI
AI
Zif
Sin3A
Orc3
Oatp58Dc
nesdHMS
Lcp9
Dgkepsilon
Dek
cona
CG7335
CG5003
CG42594
CG3262
CG17065
CG15414
CG14457
CG12688
CG12539
ns
*
**
***
****
B
Wilcox Test
P value 0.009
P value summary **
t and df 23168,
density
value
type
GWAS selected
Random selected

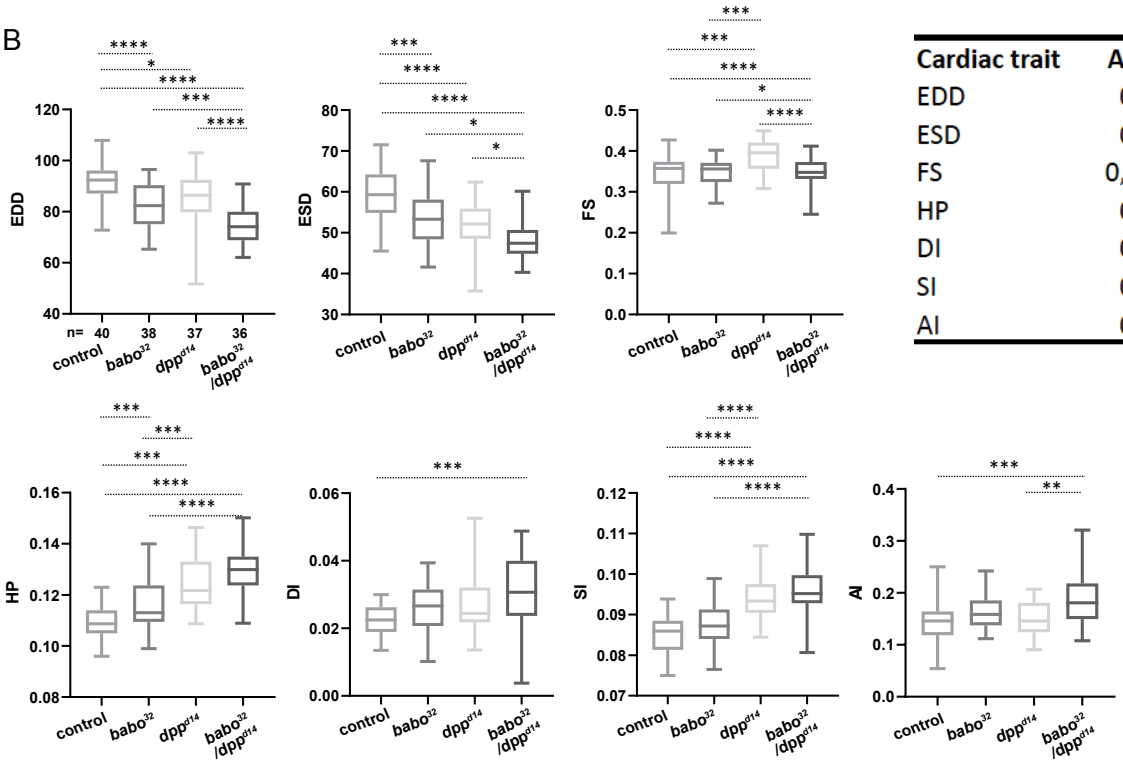
Figure S3

A



Cardiac trait	Anova
EDD	0,77
ESD	0,21
FS	0,53
HP	0,01
DI	0,075
SI	2,16E-07
AI	0,13

B



Cardiac trait	Anova
EDD	0,34
ESD	0,16
FS	0,0003
HP	0,47
DI	0,75
SI	0,48
AI	0,32

Figure S4

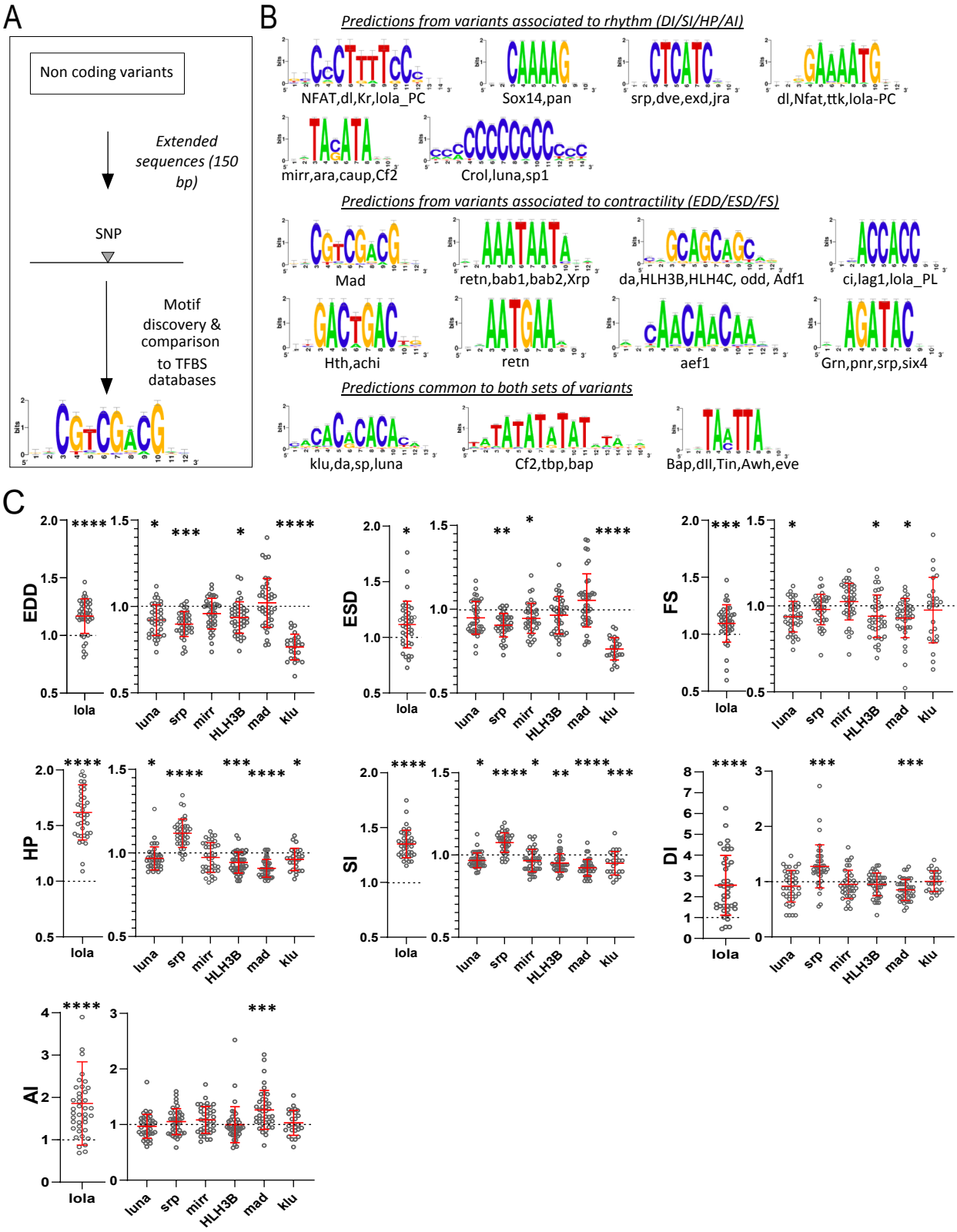

A
Non coding variants
Extended sequences (150 bp)
SNP
Motif discovery & comparison to TFBS databases
B
Predictions from variants associated to rhythm (DI/SI/HP/AI)
NFAT,dl,Kr,lola_PC
Sox14,pan
srp,dve,exd,jra
dl,Nfat,ttk,lola-PC
mirr,ara,caup,Cf2
Crol,luna,sp1
Predictions from variants associated to contractility (EDD/ESD/FS)
Mad
retn,bab1,bab2,Xrp
da,HLH3B,HLH4C, odd, Adf1
ci,lag1,lola_PL
Hth,achi
retn
aef1
Grn,pnr,srp,six4
Predictions common to both sets of variants
klu,da,sp,luna
Cf2,tbp,bap
Bap,dll,Tin,Awh,eve
C
EDD
ESD
FS
HP
SI
DI
AI
lola
luna
srp
mirr
HLH3B
mad
klu

Figure S5

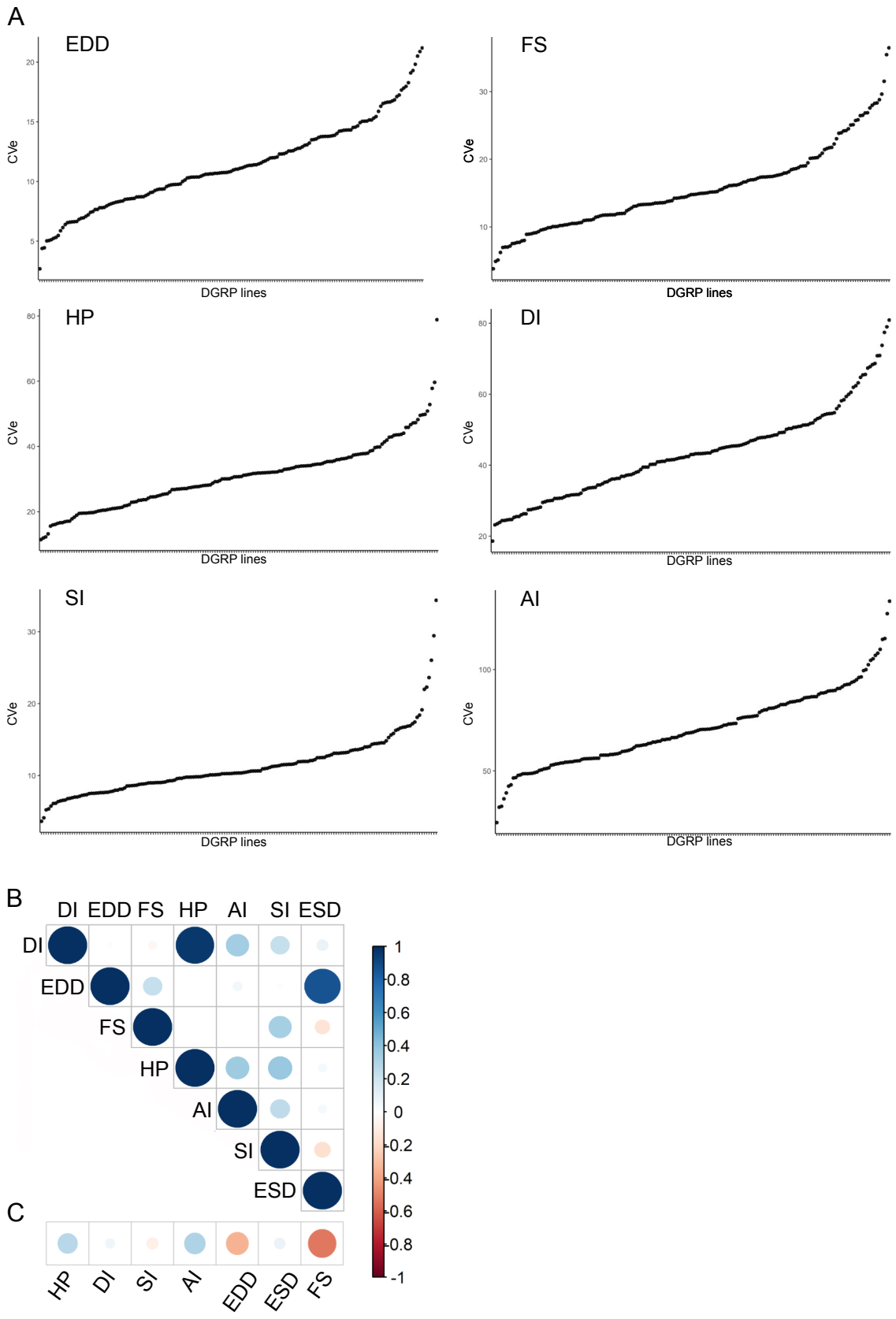

A
EDD
FS
HP
DI
SI
AI
CVe
DGRP lines
B
DI EDD FS HP AI SI ESD
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
C
HP DI SI AI EDD ESD FS

Figure S6

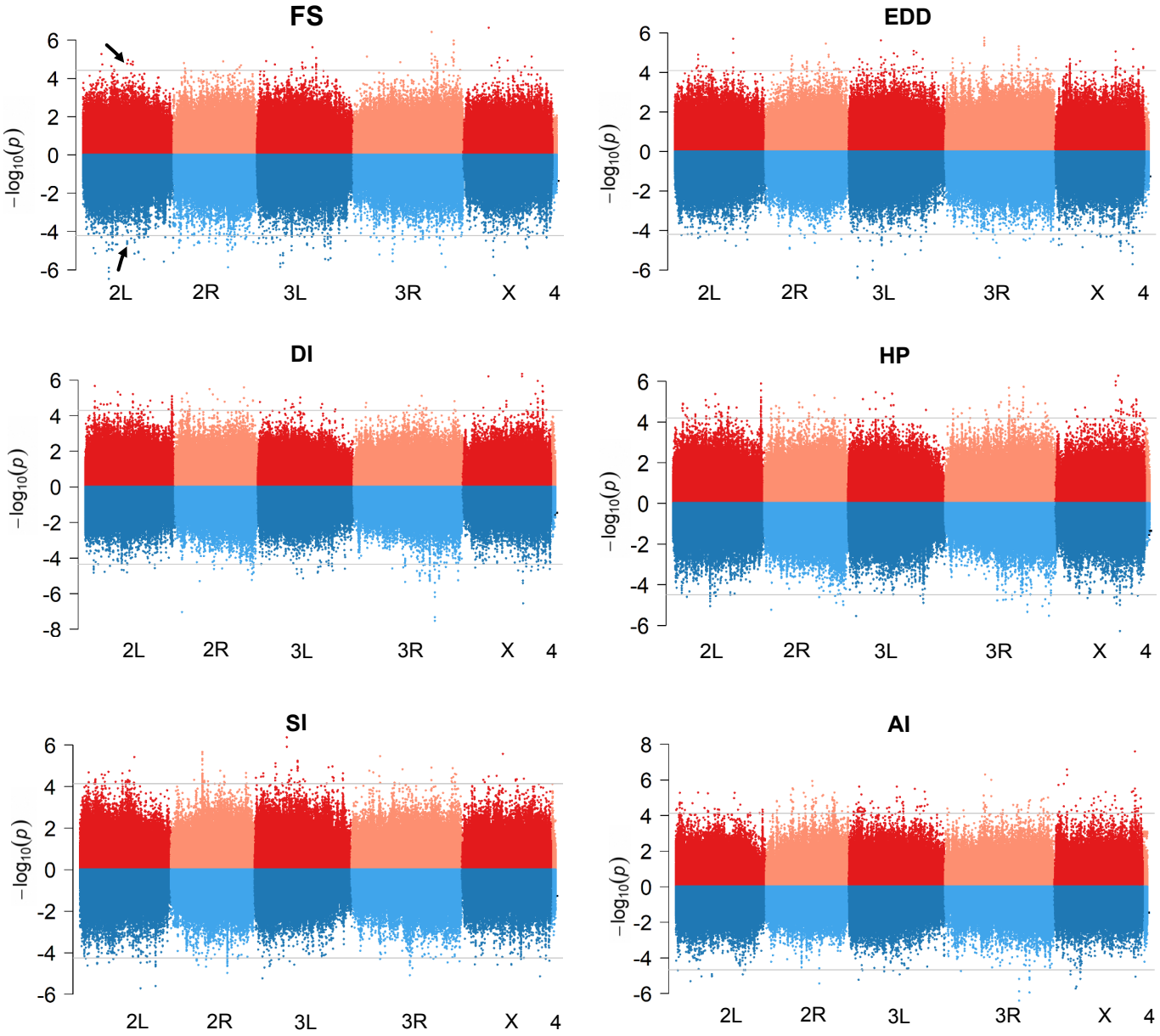


Figure S7

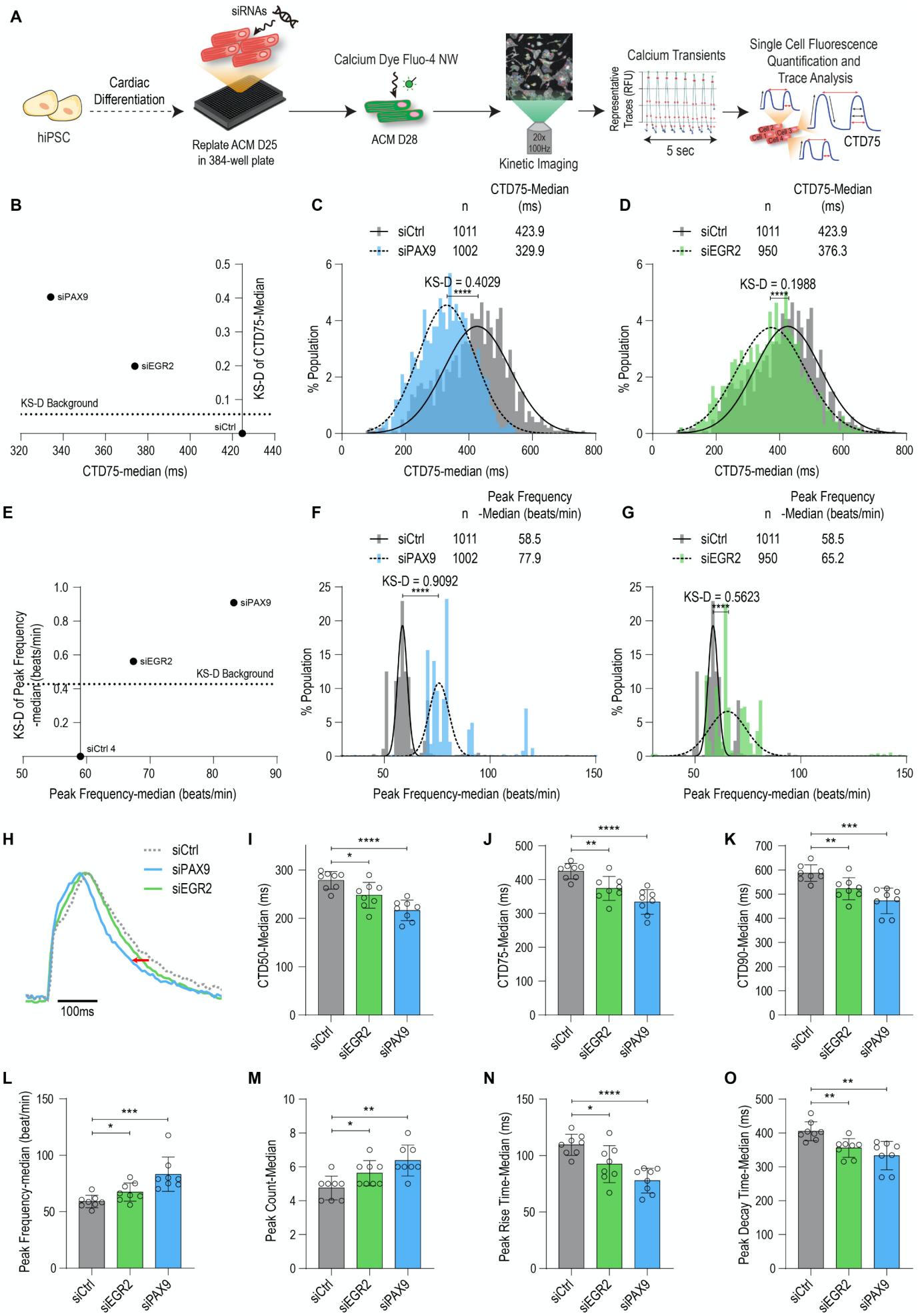

A
siRNAs
Cardiac Differentiation
hiPSC
Replate ACM D25 in 384-well plate
Calcium Dye Fluo-4 NW
ACM D28
20x 100Hz
Kinetic Imaging
Calcium Transients
Representative Traces (RFU)
5 sec
Single Cell Fluorescence Quantification and Trace Analysis
CTD75
B
siPAX9
siEGR2
KS-D Background
siCtrl
KS-D of CTD75-Median
CTD75-median (ms)
C
n
CTD75-Median (ms)
siCtrl 1011 423.9
siPAX9 1002 329.9
KS-D = 0.4029
****
% Population
CTD75-median (ms)
D
n
CTD75-Median (ms)
siCtrl 1011 423.9
siEGR2 950 376.3
KS-D = 0.1988
****
% Population
CTD75-median (ms)
E
siPAX9
siEGR2
KS-D Background
siCtrl 4
KS-D of Peak Frequency -median (beats/min)
Peak Frequency-median (beats/min)
F
n
Peak Frequency -Median (beats/min)
siCtrl 1011 58.5
siPAX9 1002 77.9
KS-D = 0.9092
****
% Population
Peak Frequency-median (beats/min)
G
n
Peak Frequency -Median (beats/min)
siCtrl 1011 58.5
siEGR2 950 65.2
KS-D = 0.5623
****
% Population
Peak Frequency-median (beats/min)
H
siCtrl
siPAX9
siEGR2
100ms
I
CTD50-Median (ms)
siCtrl
siEGR2
siPAX9
J
CTD75-Median (ms)
siCtrl
siEGR2
siPAX9
K
CTD90-Median (ms)
siCtrl
siEGR2
siPAX9
L
Peak Frequency-median (beat/min)
siCtrl
siEGR2
siPAX9
M
Peak Count-Median
siCtrl
siEGR2
siPAX9
N
Peak Rise Time-Median (ms)
siCtrl
siEGR2
siPAX9
O
Peak Decay Time-Median (ms)
siCtrl
siEGR2
siPAX9

Figure S8