Genetic architecture of natural variations 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], Krifa Sallouha[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

*Authors for correspondence (acolas@sbpdiscovery.org, kocorr@sbpdiscovery.org; laurent.perrin@univ-amu.fr)

[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, UMR ISA 1355, 06903 Sophia Antipolis, France

Abstract

Background: Deciphering the genetic architecture of cardiac disorders is of fundamental importance but their underlying complexity is a major hurdle. Drosophila has gained importance as a useful model to study heart development and function and allows the analysis of organismal traits in a physiologically relevant and accessible system. Our aim was to *(i)* identify in flies the loci associated to natural variations of cardiac performances among a natural population, *(ii)* decipher how these variants interact with each other and with the environment to impact cardiac traits, *(iii)* gain insights about the molecular and cellular processes affected, *(iv)* determine whether the genetic architecture of cardiac disorders is conserved with humans.

Methods and Results: We investigated the genetic architecture of natural variations of cardiac performance in the sequenced inbred lines of the Drosophila Genetic Reference Panel (DGRP). Genome Wide Associations (GWA) for single markers and epistatic interactions identified genetic networks associated with natural variations of cardiac traits that were extensively validated *in vivo.* Non-coding variants 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 validated by heart specific knockdown. We also analyzed natural variations of cardiac traits variance that revealed unique features of their micro-environmental plasticity. More importantly, correlations between genes associated with cardiac phenotypes both in flies and in humans support the conserved genetic architecture of cardiac functioning from arthropods to mammals. The characteristics of natural variations in cardiac function established in Drosophila may thus guide the analysis of cardiac disorders in humans. Using human iPSC-derived cardiomyocytes, we indeed characterized a conserved function for PAX9 and EGR2 in the regulation of the cardiac rhythm

Conclusion: In-depth analysis of the genetic architecture of natural variations of cardiac performance in flies combined with functional validations *in vivo* and in human iPSC-CM represents a major achievement in understanding the mechanisms underlying the genetic architecture of these complex traits and a valuable resource for the identification of genes and mechanisms involved in cardiac disorders in humans.

Introduction

Deciphering the impact of genetic variations on quantitative traits is critical for the prediction of disease risk. In human populations Genome Wide Associations Studies (GWAS) typically aim at identifying genetic variations associated with disorders in order to unravel the underlying genetic architecture of associated diseases. However, 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 cause lack of replication in human GWAS performed in distinct populations which rarely take epistatic effects into account. In addition, linking a complex trait associated locus to a candidate gene or a set of genes for prioritization is not straightforward[1,2].

Heart diseases is a major cause of mortality[3]. Although a large number of studies have identified hundreds of genetic variants related to cardiovascular traits, we are very far from a comprehensive understanding of the genetic architecture of these complex traits. In addition to the inherent limitations of human GWAS, 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[4]. 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[5–7]. Although a large number of genes are implicated in establishing and maintaining cardiac function in Drosophila[8], 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..

In this study, we describe the genetic architecture of the natural variations 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) [9,10], 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[11]. The use of inbred lines allows assessing the effect 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 transcription factors (TFs) and signaling pathways encoding genes, that we validated *in vivo*. We used non-coding variants - which represented the vast majority of identified polymorphisms - for predicting transcriptional regulators of associated gene. Corresponding TFs were further validated *in vivo* by heart specific RNAi mediated knockdown. This illustrates that natural variations of cardiac function have widespread impact on gene regulatory networks. 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 between DGRP lines, suggesting genetic variations influencing microenvironmental plasticity of cardiac performance. Genetic variants associated with phenotypic variability were identified and shown to affect a set of genes that overlap 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 human, 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 in flies and human. These new insights into the fly's genetic architecture and how natural variations affect cardiac performance therefore allows accelerated identification of essential cardiac components 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[12] 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 [13] [14]. A total sample of 1956 individuals was analyzed (approximately 12 individuals per line, see methods). 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[15]. Genotype-phenotype associations were performed separately for all 7 quantitative traits and variants were ranked based on their p-values. We considered the 100 top ranked variants for reporting candidate associations for each quantitative trait and further filtered out the ones with a minor allele frequency (MAF) < 4 % (Dataset 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 S1). 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 [16] and part of this 'missing heritability' is thought to come from interactions between variants[17 16 18 19]. We therefore performed GWAS analyses for pairwise interactions. For all cardiac traits, Single Nucleotide Polymorphism (SNP) identified by single marker GWAS were used as focal SNPs and were tested for interactions with all other SNPs in the DGRP using FastEpistasis[20]. The best ranked interacting SNP was retained for each starting focal SNP. This screen for epistatic interactions identified 288 unique SNPs that were mapped to 261 genes (Dataset S3, Table S2). While none of the focal SNPs interacted with each other (not shown), 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 S2, Fold Change (FC)= 6; hypergeometric pval = 6,8 x 10^{-16}). This illustrates that the genes that contribute to quantitative variations in cardiac performance have the 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 S3). In line with the correlation noted between their phenotypes (Figure S1F), the GWAS for HP and DI identify partially common gene sets (overlap index 0.23, Figure 1E). The same is 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 analyze the biological processes potentially affected by natural variations in cardiac performance. We considered overall results of GWAS (combining single marker GWAS and epistatic interactions) for the functional analysis of associated genes. Gene Ontology (GO) enrichment analysis pointed out that genes encoding signaling receptors, transcription factors (TFs) and cell adhesion molecules were over-represented among this gene set (pval= $1{,}4 \times 10^{-9}$ (FC = 2,9), 5×10^{-4} (FC = 2) and 3×10^{-3} (FC = 4,6) respectively). Among cellular components annotations, we noticed a biased representation of genes encoding for proteins located at the plasma membrane, at ion channel complexes as well as parts of contractile fibers (pval = $3{,}4 \times 10^{-10}$ (FC= 3), 7×10^{-5} (FC= 4,2) and 4×10^{-2} (FC= 3,6); Figure 2A; Table S4). Of note, although some genes have previously been identified by single gene approaches, as required during heart development or for the

establishment and maintenance of cardiac function, we found no enrichment for these gene categories in our analysis. In addition, GWAS detected genes were not enriched in genes identified in a global screen for stress induced lethality following heart specific RNAi knockdown[8] (FC = 1; Table S4). This indicates that 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 by both physical (protein-protein) and genetic interactions (see Materials and Methods and dataset S4). Remarkably, a high proportion (148) of the GWAS associated genes were directly connected within the fly interactome and formed a large network of interacting genes/proteins (Table S5, 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 TGFβ pathways (see below).

Functional validations of candidate genes

To assess whether mutations in genes that harbor SNPs associated with variation in cardiac traits could contribute to these phenotypes, we selected 42 candidate genes as target genes for cardiac-specific RNAi knockdown and assessed to what extent gene knockdown affected cardiac performance. The tested genes were selected when they were identified in at least two independent GWAS for two traits or characterized as being dynamically expressed in the adult heart[21] and according to the availability of inducible RNAi lines. Genes were tested in 1 week adult female flies, using the heart specific Hand>Gal4 driver[22] and the same semi-intact heart preparations and SOHA analyses as for DGRP lines screening. Notably, 40 of the 42 selected genes led to various levels of cardiac performance defects following heart-specific knock down (KD) (Figure 2C). This high rate of validations supports our association results.

We further focused on the TGFβ pathway, of which members of both BMP and activin pathways were identified. We have tested different members of the TGFβ pathway for cardiac phenotypes after cardiac specific RNAi knockdown (Figure 2C), and confirmed the

involvement of the activin agonist *snoo* (Ski orthologue), the BMP antagonist *sog* (chordin orthologue) and *put* -a type 2 TGFβ receptor- in cardiac performance. Notably, Activin and BMP pathways are usually antagonistic (Figure 2D). Their joint identification in our GWAS could indicate that they have, uniquely for these complex traits, coordinated activity. 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 would genetically interact. 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 S2). Single heterozygotes phenotypic values were then used to model expected double heterozygous phenotypic values, which were compared to the actually measured parameters from double heterozygotes flies. Remarkably $snoo^{BSC234}$/ sog^{U2} double heterozygotes show significantly higher than expected SI parameters, illustrating a genetic interaction (Figure 2E and S2). This result supports a coordinated action of activin and BMP in the establishment and/or maintenance of cardiac activity. Synergistic action of BMP and activin pathways for setting cardiac performance was further supported by the analysis of cardiac performance of loss of function mutants for *babo* and *dpp*, which demonstrated genetic interactions between the type 2 activin receptor and the BMP ligand in setting fractional shortening (Figure 2E and S3).

Taken together, the high rate (40/42) of cardiac phenotypes observed following RNAi mediated gene KD and the genetic interactions characterized between different activin and BMP pathways members are strong arguments that the GWAS results on cardiac traits are indeed valid.

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) 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 and, remarkably, their distribution was enriched in gene proximal (0-1kb) upstream regions (Figure 1C). This may indicate that a high proportion of variants affects cis-regulatory sequences. Hypothesizing that variants associated with natural variations in cardiac performance might allow identification of shared transcriptional

regulatory modalities, we sought to determine whether potential TF binding motifs are enriched in the vicinity of variants identified by GWAS. We analyzed potential regulatory regions encompassing these variants by performing motif discovery in 150 bp sequences around non-coding variants associated with the various cardiac traits. Using oligo analysis[23] we detected significant over represented k-mers (Evalue <= 1e-4) that were further assembled and converted into Positions Specific Scoring Matrices (PSSM). Finally, comparisons with PSSM databases of known fly TFs binding motifs were used for annotations (see methods, dataset S6 and Figure 3A). 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 3B). Although this approach predicted motifs that were specific for each group of variants, a significant proportion of potential DNA binding motifs were predicted from 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 S3). Potential binding sites for Mother against dpp, Mad (SMAD) - the downstream effector of the BMP signaling pathway that we show involved in cardiac contractility (Figure 2D-E) - were found enriched in the vicinity of non-coding variants associated with heart size and contractility. 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[24] and at adulthood [25] and *bap* is also a member of the cardiac network[26]. 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[21], supporting a function for this stress-activated TF in the cardiac GRN.

The cardiac function of several predicted TFs was tested by analyzing cardiac specific RNAi mediated gene KD. *luna*, *lola* and *srp* KD have been already analyzed on semi-intact preparations (Figure 2C) and we demonstrated that their heart specific inhibition induces strong cardiac phenotypes. Validations experiments of TF Binding Sites (TFBSs) predictions were further performed by analyzing *mirr*, *mad*, *klu* and *HLH3B* heart specific RNAi KD *in*

vivo, in intact animals using the red fluorophore TdTomato driven by a cardiac enhancer[27]. Additional independent RNAi lines for *lola*, *luna* and *srp* were also analyzed using this approach. As illustrated in Figure 3C, 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. *lola* encodes a TF with multiple isoforms. A PSSM for one isoform (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.

Altogether, the high rate of non-coding SNPs among variants identified in GWAS, TFBS predictions, *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, for which several individuals that share the same genotype have been analyzed, offers the possibility of studying the heritability of the within-line variance of the phenotypes, thus allowing analysis of the genetics of phenotypic variability of cardiac traits. Several studies have indeed shown that the within-line environmental variance of inbred lies is heritable[28–30]. These genetic variations in phenotypic variability suggest that different genotypes respond differently in response to micro-environmental changes[31,32]. We used the Levene test[33] to test the heterogeneity of the "micro-environmental" variance among lines for the 7 analyzed cardiac traits. We found evidence of significant heterogeneity of the variances among the lines for all analyzed cardiac traits (Table 1). This indicated that different lines have distinct phenotypic variability of cardiac traits, suggesting that the lines differ in their ability to adapt to micro-environmental variations, and supporting heritability of these. 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 4A, B and Figure S4A) 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 S4B),

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 S4C). 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, the 100 top ranked variants were retained following single variant GWAS analyses and were further used as focal variants for epistasis. Overall, this led to the identification of 887 variants and 566 genes associated with natural variations of coefficient of within line variation of cardiac performance (dataset S7, table S7).

Consistent with the weak correlation between cardiac traits CVe and trait means, only 3 variants were found associated with both trait means and trait CVe (Figure 4C; Table S7). Two variants in total linkage disequilibrium (5bp apart) are localized in an intron of *Ca-β*, encoding a voltage gated calcium channel subunit. 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 analyzes on CVe and trait means are associated with two sets of genes that overlap significantly (Figure 4C, Table S7): among 566 and 562 genes detected respectively from trait CVe and trait mean GWAS, 92 genes are found associated with both CVe and mean of cardiac traits (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.

We observed an over-representation of cell adhesion processes, transmembrane signaling and ion channel activity (pval= $1,6\times10^{-8}$ (FC = 7), $1,1\times10^{-3}$ (FC = 2,2) and $4,5\times10^{-2}$ (FC = 2,5) respectively) for CVe associated genes. Enriched cellular component annotations comprised cell-cell junction (pval= $5,7\times10^{-3}$ (FC = 3,2), receptor complexes (pval= $3,3\times10^{-4}$ (FC = 4,6), sarcolemma (pval= $1,3\times10^{-2}$ (FC = 25) as well as several annotations related to neuronal projections suggesting both cell autonomous and non-autonomous activity of related genes (Figure 4D, Table S8). In addition, genes leading to stress-induced lethality upon heart specific knockdown[8] were significantly underrepresented in this gene set (FC= 0,6; p= 0,05; Table S8). This may suggest that a significant proportion of CVe GWAS associated genes affect cardiac traits 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 4E). This suggested that natural variations of cardiac performance variance are also mainly affecting cis-regulatory regions. Using the strategy

followed for trait means, we identified over-represented PSSMs in sequences of 150pb around non-coding CVe variants (Figure 4F). Compared to the analysis performed on variants associated with trait means, some identified PSSMs are 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 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 S7)-, 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 S9). Using DIOPT[34], we identified human orthologues of the genes associated with cardiac traits means and CVe in flies (Table S10). 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 S10), 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 S10). 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[5–7,35] - are not only conserved but also can be largely be extended to the genetic basis of 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) – which are both associated with natural variations of cardiac trait means (Table

S3). Notably, to our knowledge, none of them and their human orthologues have a function in the cardiac system. In flies, *poxm* contributes to the development and specification of somatic muscles[36] while *sr* is required for tendon cell fate induction[37]. We first tested the effects of their heart specific KD on cardiac performance in flies using the *in vivo* assay. Both specifically induced an increased heart rate with a reduced HP and SI (i.e., increased heart rate; Figure 5A) and did not affected other cardiac traits analysed (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 5B)[38]. As shown in figure 5C-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 (Fig.5F-K and N, O). Increased AP kinetics correlated with shorter calcium transient (Figure S5A-D and H-K), including faster upstroke and downstroke calcium kinetics and increased beat rate (peak frequency) (Figure S5E-G and L, M), which confirmed that both PAX9 and EGR2 KD impacted cardiomyocytes 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

Characterizing the genetic architecture of natural variations of cardiac disorders is of fundamental importance from a medical perspective, as it may allow disentangling the relative contribution of genetics and environmental factors to cardiac pathologies, and the influence of comorbidities. In addition, natural selection is thought to operate on these complex traits rather than on single gene mutations, which further illustrates the interest in describing and understanding their genetic architecture. 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 phenotypic variations in individual cardiac performance.

Our results validate Drosophila as a unique model to analyze the etiology of the genetic foundations of cardiac function. In this study we identified many genes that were

previously not associated with cardiac traits, either at 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. This 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[39,40]. Leveraging on non-coding variants, we failed to find enriched potential TFBS at SNPs positions (not shown). However, extending the search space to sequences of +/- 75bp around variants revealed enriched DNA motifs, highlighting the potential role of several TFBS. Our results indicate that variants can affect sequences outside the core TF binding motif (e.g. flanking sequences) that in turn would affect TF binding[41]. Supporting our findings, different reports on human GWAS analyses have shown that many risk-associated SNPs are not precisely located in the conserved binding motifs of TFs but are in nearby regions[42,43]. 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 number of predicted motifs were challenged by testing their cognate TFs *in vivo* for cardiac performance and a high rate of validation was obtained. Of note, while the cardiac GRN has been extensively studied in flies[26,44], a large proportion of identified TFBS and cognate factors 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[31]. From a medical

genetics point of view, high variability genotypes may produce individuals beyond a phenotypic threshold, leading to disease state [28,45]. Having assessed the cardiac traits on 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[8] are under-represented among the genes associated with CVe traits. This may suggest that part of this gene set influence 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 - even though the sets of associated genes are significantly overlapping - is another remarkable feature of the genetic architecture of cardiac phenotypic plasticity. However, although based on non-overlapping sets of non-coding regions, TFBS predictions in the vicinity of non-coding variants identify partially overlapping sets of motifs. This suggests that natural variation of trait means and trait variance affect partially overlapping GRNs, albeit by distinct polymorphisms.

Several previous studies have established that the principles identified on the DGRP population are universal and can be applied across phyla[32]. 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[46,47]. 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 may help reveal the characteristics of the natural variations of cardiac performances 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 great 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 except when indicated in the Fly strain resource table. 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)[22] virgins were crossed to RNAi males and corresponding isogenic control males. Flies were raised on standard fly food and kept at 25°C. Female progeny was collected and aged to 1 week at 25°C, at which point they were imaged and analyzed. Mutant analyses followed the same procedure.

- 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)[27] 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[27]. Hearts were illuminated with green light (3mW power) and 5sec movies of the beating heart were acquired at 270fps. High-speed recordings were then processed using a custom R script[48].

- 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 data have been reanalyzed. It was identified that these extreme results corresponded to technical problems (poor quality film, bad placement of the measurement points...). We then decided to remove these data 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

Since we have used mean and CV summary statistics for our GWAS analysis, it is worth studying the leverage effect of extreme values. We identified strong leverage effect in some phenotypes and treated them for outliers. We decided to remove from GWAS analysis of the phenotype the lines with large leverage effect, 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, removing all leverage effects.

GWAS analysis was performed using FastLMM[15]. Analysis of lineages effect reveal no family/block effect between lines as stated by the DGRP consortium. Linear mixed model was built using contamination by Wolbachia and Principal Genetic Inversions (data provided by the DGRP Consortium) as covariates.

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 is located in a gene (INTRON, EXON, 3'or 5' UTR), the SNP is associated with this gene, (ii) if not, the variant is associated with a gene if its distance from the boundary of the gene (Transcription Start Site (TSS) and Transcription End Site (TES) is lesser than 1kb (iii) if no previous rules was applied, the variant is not mapped to gene. We then considered the variants with a Minor Allele Frequency (MAF) greater than 4% and discarded the SNP with synonymous effect. Epistasis analysis was performed using FastEpistasis[20]. For each phenotype, we used the SNP selected from the GWAS analysis as focal SNP and all the SNP with MAF greater than 5% as test SNP. 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.

- Interaction Network analysis:

Drosophila Interactome: A network made of genetic and physical interactions has been built. Briefly, it contains protein-protein interactions identified using Y2H and AP-MS techniques, and genetic interactions, all available in the DroID database (http://www.droidb.org/)[49], 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 have been made using Cytoscape[50].

- Genetic interaction test

To test for genetic interaction, we first modeled for each double-heterozygous genotype and heart parameter a range of expected values. We calculated these as the product of the single heterozygote measurements relative to the control values. Specifically, we sampled for each parameter single values from the control genotype and both single genotypes, and multiplied the single values normalized to the control value, and then multiplied the result by the control value. This was done 50 times for each parameter to produce an expected value dataset. We then tested if the measured values obtained from the double-heterozygotes were statistically different from the artificial expected value data using both t-test and Wilcoxon test. We repeated this 10 times for all parameters and calculated the mean p-value of each parameter (and standard deviation).

- Statistics

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 sample (based on the observations in the population). Gene Ontology enrichment analyses were conducted using the online version of G:profiler tool[51] with default parameters. 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. 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.

- Motif's analysis

For every set of SNPs (Rhythmicity CVe and Mean, contractility CVe and Mean), the coordinates (dm3 genome assembly) where extended 75 nt to both sides and their corresponding fasta sequences were retrieved. TF binding motifs were found using RSAT peak-motifs[52] (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[53]). To reduce the redundancy of the found motifs (PSSMs), we applied the RSAT matrix-clustering[54] algorithm (default parameters) to find clusters of motifs. The clustered motifs were manually annotated by comparison with PSSM databases (Cis-BP[55], JASPAR[56] and FlyFactorSurvey[57]) using the tool RSAT compare-matrices[58] (default parameters).

- Generation of hiPSC-ACMs

Id1 overexpressing hiPSCs[59] 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[60]. At day 0, cells were treated with 6 µM CHIR99021 (Selleck Chemicals) in S12 media[61] 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[38]. 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.

- Calcium Assay in hiPSC-ACMs

Calcium assay is performed using labeling protocol as described[62]. 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:

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 KK. We thank the Bloomington Drosophila Stock Center (BDSC) and the Vienna Drosophila Research Center (VDRC) for fly stocks.

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
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 that have a human ortholog and for genes with a human ortholog associated with cardiac disorders and Coronary Artery Diseases (CAD) within genes associated to traits mean (562 genes) and to traits CVe (566 genes) compared to the whole fly genome gene set (17500 genes). First row: Number of fly genes that display a human orthologue according to DIOPT (high and moderate rank). Second and third rows (CAD/Cardiac disorders): Number of genes whose human orthologue (high and moderate rank) has been associated with either CAD or cardiac disorders by GWAS in human populations (Table S9, Table S10). 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 individuals per line were analyzed. Right panels: Schematic of the Drosophila adult heart assay and example of M mode generated from video recording of 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. (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. (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 TGFβ, 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. Except *vav* and *sd*, all genes tested lead to significant effects on cardiac performance traits upon KD. ns: not significant; *: pval< 0,05; **: pval<0,01; ***: pval< 0,005; ****pval<0,0001. (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 S2 and Figure S3, respectively). Cardiac traits were measured on indicated genotypes and values measured on double heterozygotes flies were compared to expected ones which were modeled from the single mutant values. Statistical significance is indicated (**** pval<0,0001). In both cases, double heterozygotes measured phenotypes are significantly different from expected values.

Figure 3:

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

(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 below 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 S6.

Figure 4:

Quantitative genetics and GWAS for cardiac traits CVe in the DGRP.

(A) Barplot distribution of ESD CVe values among DGRP lines. Lines are arranged by increasing CVe values. (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) 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. (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) 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 5:

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

(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 1 week 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. N = number of heartbeats counted from M modes.

(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: (complement to figure 1)

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. Spearman correlations were calculated on average of line means for each cardiac trait.

Figure S2 (complement to Figure 2):

Cardiac performances phenotypes of heterozygotes for *snoo*^{BSC234}, *sog*^{U2} and *snoo* ^{BSC234}; *sog*^{U2} double heterozyotes.

A: 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.

B-C: Genetic interactions tested between *snoo*^{BSC23} and *sog*^{U2}. B: Cardiac traits measured on indicated genotypes (same experiments as in A) were used to model values expected for double heterozygotes (expected_data, see Materials and Methods). C: Mean p-values of the difference between observed and artificial expected value data for double heterozygotes (from 10 artificial value predictions, see Materials and Methods).

Figure S3 (complement to Figure 2):

Cardiac performances phenotypes of heterozygotes for babo^{32}, dpp^{d14} and babo^{32}; dpp^{d14} double heterozyotes.

A: 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.

B-C: Genetic interactions tested between *dpp*^{d14} and *babo*^{32}. B: Cardiac traits measured on indicated genotypes (same experiments as in A) were used to model values expected for double heterozygotes (expected_data, see Materials and Methods). C: Mean p-values of

difference between observed and artificial expected value data for double heterozygotes (from 10 artificial value predictions, see Materials and Methods).

Figure S4: (complement to figure 4)

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 mean and CVe for each cardiac trait.

Figure S5: (complement to figure 5)

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:

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 S2:

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 S1) are labeled in red.

Table S3:

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 S4:

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 S3), 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[8].

Table S5

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 S6:

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 3C. Wilcoxon-Mann-Whitney rank sum test results for indicated cardiac parameters are shown.

Table S7:

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 S8:

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 S7), 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[8].

Table S9:

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 S10:

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

Drosophila genes identified in GWAS (from Table S3 and S7), 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 S9)

- 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

Dataset S1:

Phenotypic values for the 7 traits analysed across the 167 DGRP lines and the 1956 individuals.

Dataset S2:

FastLMM_GWASresults_signif100th_genemap:

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.

FastLMM_GWASresults_signif100th_genemapmissing:

Variants among the 100 best ranked associations for each cardiac trait that are not within gene mapping criteria (>1kb from TSS and TTS) and their frequency in the DGRP population.

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:

Raw data and detailed statistical analyses from validation experiments.

- RNAi validation figure 2: data from heart specific RNAi knockdown presented in figure 2C.

- interactions BMP: data from test of BMP and activin mutant interactions (Figure 2E, Figure S2 and Figure S3).
- TFs validation: data from heart specific RNAi knockdown of TFs presented in figure 3 and 5.

Dataset S6:

- non coding variants mean: variants in non-coding regions used for motif discovery, associated with EDD, ESD and FS ('structure') and to AI, DI, SI and HP ('rhythm').
- Non coding variants CVe

Dataset S7:

- FastLMM_GWAS_CVe_results_signif100th_genemap:
- 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.
- FastLMM_GWAS_CVe_results_signif100th_genemapmissing:
- Variants among the 100 best ranked associations for each cardiac trait that are not within gene mapping criteria (>1kb from TSS and TTS.
- Epistasis_on_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.

Bibliography

1. Mackay, T. F. C. Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions. *Nat. Rev. Genet.* **15**, 22–22 (2014).
2. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
3. Bezzina Connie R., Lahrouchi Najim, & Priori Silvia G. Genetics of Sudden Cardiac Death. *Circ. Res.* **116**, 1919–1936 (2015).
4. Paludan-Müller, C., Svendsen, J. H. & Olesen, M. S. The role of common genetic variants in atrial fibrillation. *J. Electrocardiol.* **49**, 864–870 (2016).
5. Ocorr, K. *et al.* Genetic Control of Heart Function and aging. *Trends Cardiovasc. Med.* **17**, 177–182 (2007).
6. Diop, S. B. & Bodmer, R. Gaining Insights into Diabetic Cardiomyopathy from Drosophila. *Trends Endocrinol. Metab.* **26**, 618–627 (2015).
7. Rosenthal, N., Harvey, R. P., Bodmer, R. & Frasch, M. Development and Aging of the Drosophila Heart. in *Heart Development and Regeneration* 47–86 (2010).
8. 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).
9. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–8 (2012).
10. Huang, W. *et al.* Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res.* **24**, 1193–208 (2014).
11. 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).
12. 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).
13. 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).
14. Cammarato, A., Ocorr, S. & Ocorr, K. Enhanced assessment of contractile dynamics in Drosophila hearts. *BioTechniques* **58**, 77–80 (2014).
15. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
16. Manolio, T. a *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
17. Flint, J. & Mackay, T. F. C. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
18. Huang, W. *et al.* Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15553–9 (2012).
19. 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).
20. 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).
21. Monnier, V. *et al.* dJun and Vri/dNFIL3 Are Major Regulators of Cardiac Aging in Drosophila. *PLoS Genet.* **8**, (2012).
22. 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).
23. 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).
24. Bodmer, R. The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Dev. Camb. Engl.* **118**, 719–729 (1993).
25. 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).
26. 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).
27. Klassen, M. P. *et al.* Age-dependent diastolic heart failure in an in vivo Drosophila model. *eLife* **6**, 209–218 (2017).
28. 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).
29. 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).
30. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6706–11 (2015).
31. 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).
32. 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).
33. Levene, H. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling,. in 278–292 (Stanford Univ Press, Stanford, CA, 1960).
34. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).
35. Bier, E. & Bodmer, R. Drosophila, an emerging model for cardiac disease. *Gene* **342**, 1–11 (2004).
36. Duan, H. *et al.* A key role of Pox meso in somatic myogenesis of Drosophila. *Dev. Camb. Engl.* **134**, 3985–3997 (2007).
37. Gunage, R. D., Dhanyasi, N., Reichert, H. & VijayRaghavan, K. Drosophila adult muscle development and regeneration. *Semin. Cell Dev. Biol.* **72**, 56–66 (2017).
38. 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).
39. 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).
40. Kikuchi, M. *et al.* Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med. Genomics* **12**, 128 (2019).
41. 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).
42. Heinz, S. *et al.* Effect of natural genetic variation on enhancer selection and function. *Nature* **503**, 487–492 (2013).
43. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
44. 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).
45. Gibson, G. Decanalization and the origin of complex disease. *Nat. Rev. Genet.* **10**, 134–140 (2009).
46. Zhou, S. *et al.* The Genetic Basis for Variation in Sensitivity to Lead Toxicity in Drosophila melanogaster. *Environ. Health Perspect.* **124**, 1062–1070 (2016).
47. 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).
48. Vogler, G. R-script to analyze fluorescent Drosophila melanogaster hearts.
49. 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).
50. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
51. 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).
52. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40**, e31–e31 (2012).
53. 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).
54. 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).
55. Weirauch, M. T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431–1443 (2014).
56. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
57. 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).
58. Nguyen, N. T. T. *et al.* RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res.* **46**, W209–W214 (2018).
59. Cunningham, T. J. *et al.* Id genes are essential for early heart formation. *Genes Dev.* **31**, 1325–1338 (2017).
60. 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).
61. 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).
62. 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).

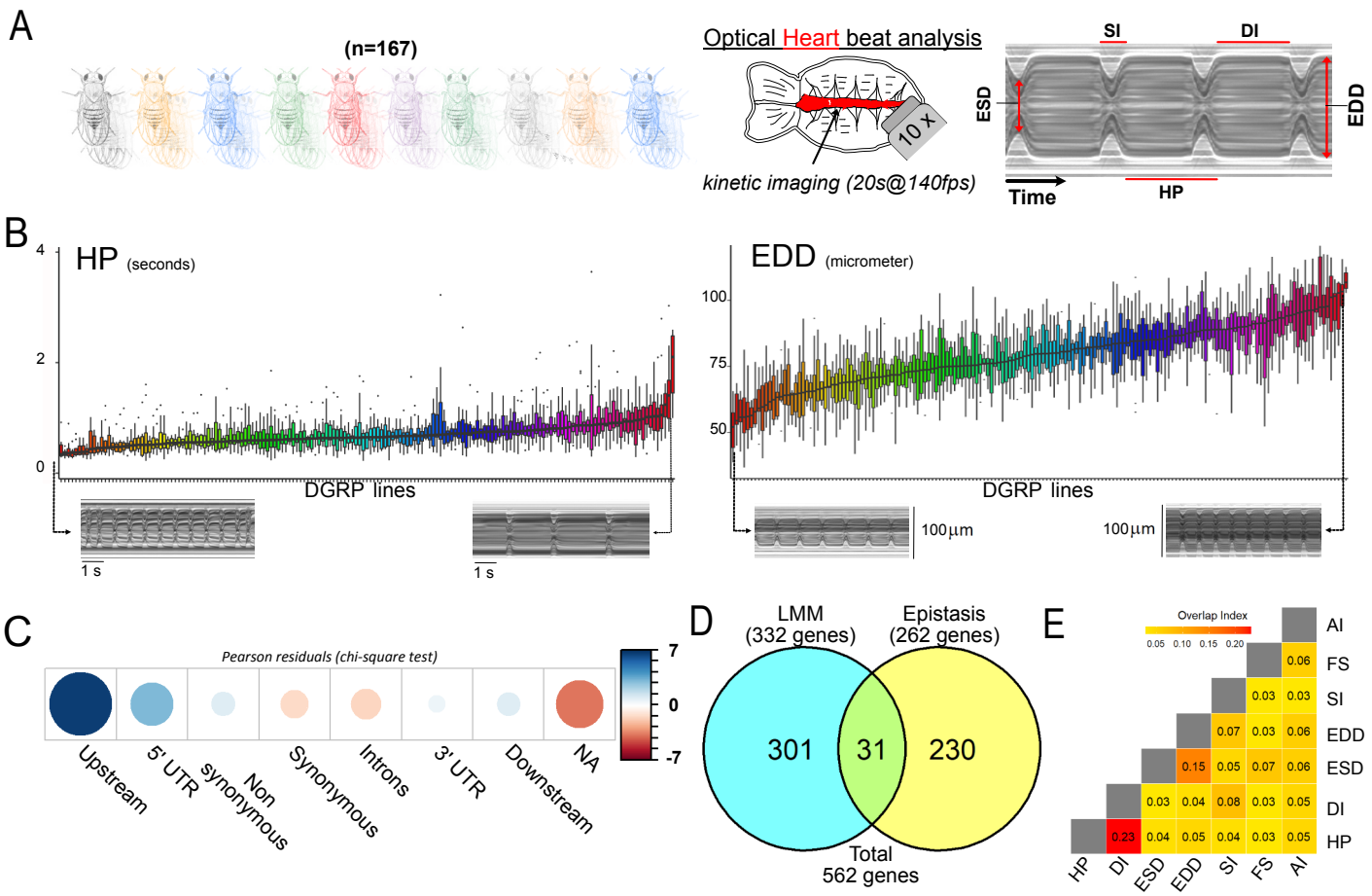

A
(n=167)
Optical Heart beat analysis
10 x
kinetic imaging (20s@140fps)
SI
DI
ESD
EDD
Time
HP
B
HP (seconds)
DGRP lines
1 s
EDD (micrometer)
100µm
C
Pearson residuals (chi-square test)
Upstream
5' UTR
Non synonymous
Synonymous
Introns
3' UTR
Downstream
NA
D
LMM (332 genes)
Epistasis (262 genes)
301
31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

Figure 1

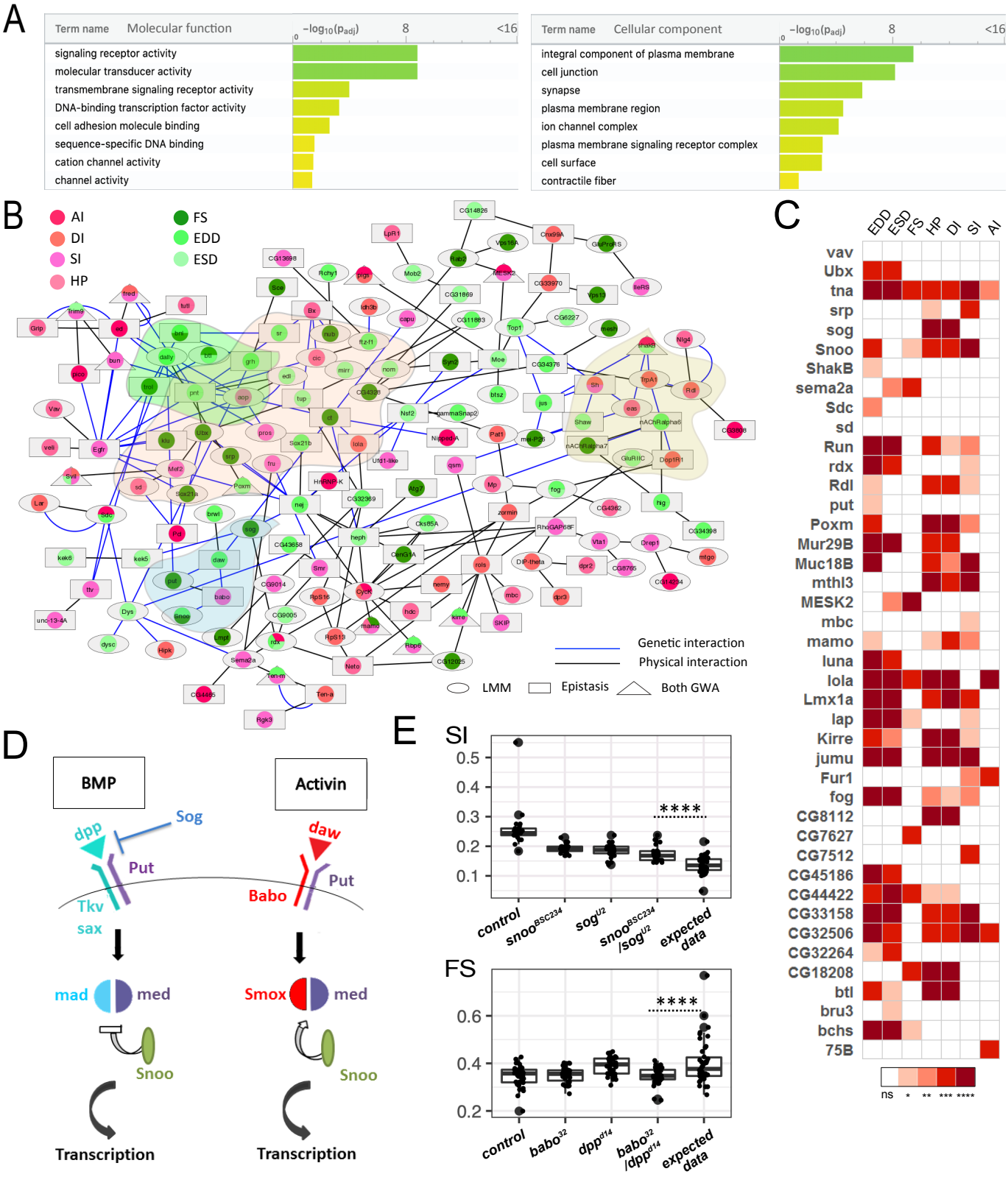

A
Term name
Molecular function
$-\log_{10}(p_{adj})$
0
8
<16
signaling receptor activity
molecular transducer activity
transmembrane signaling receptor activity
DNA-binding transcription factor activity
cell adhesion molecule binding
sequence-specific DNA binding
cation channel activity
channel activity
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
srp
sog
Snoo
ShakB
sema2a
Sdc
sd
Run
rdx
Rdl
put
Poxm
Mur29B
Muc18B
mthl3
MESK2
mbc
mamo
luna
lola
Lmx1a
lap
Kirre
jumu
Fur1
fog
CG8112
CG7627
CG7512
CG45186
CG44422
CG33158
CG32506
CG32264
CG18208
btl
bru3
bchs
75B
ns
*
**
***
****
D
BMP
Activin
dpp
Sog
Put
Tkv
sax
daw
Babo
Put
mad
med
Smox
med
Snoo
Snoo
Transcription
Transcription
E
SI
0.5
0.4
0.3
0.2
0.1
****
control
snoo^{BSC234}
sog^{U2}
snoo^{BSC234}/sog^{U2}
expected data
FS
0.6
0.4
0.2
****
control
babo^{32}
dpp^{d14}
babo^{32}/dpp^{d14}
expected data

Figure 2

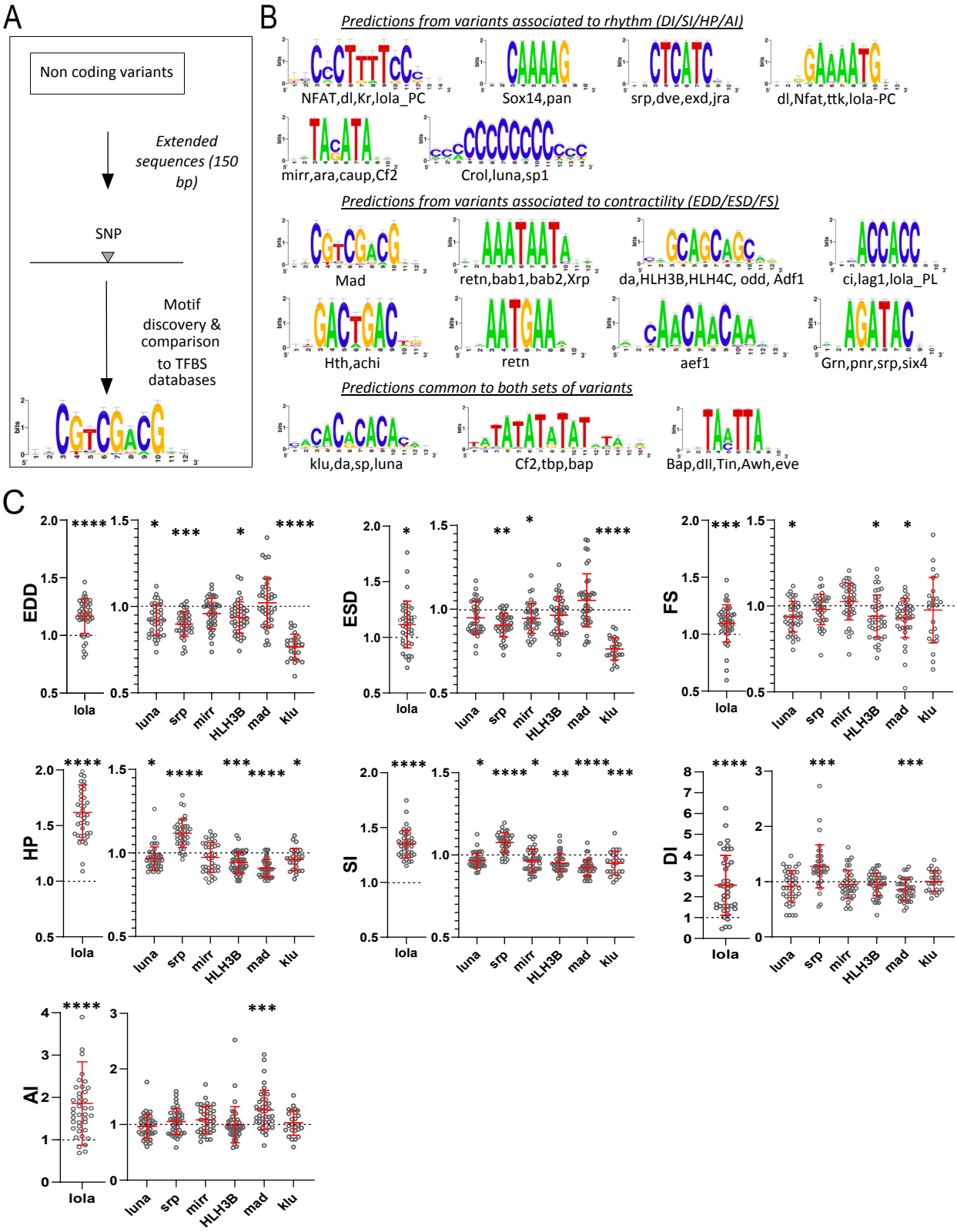

A
Non coding variants
Extended sequences (150 bp)
SNP
Motif discovery & comparison to TFBS databases
B
Predictions from variants associated to rhythm (DI/SI/HP/AI)
NFAT,dl,Kr,lola_PC
Sox14,pan
srp,dve,exd,jra
dl,Nfat,ttk,lola-PC
mirr,ara,caup,Cf2
Crol,luna,sp1
Predictions from variants associated to 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 3

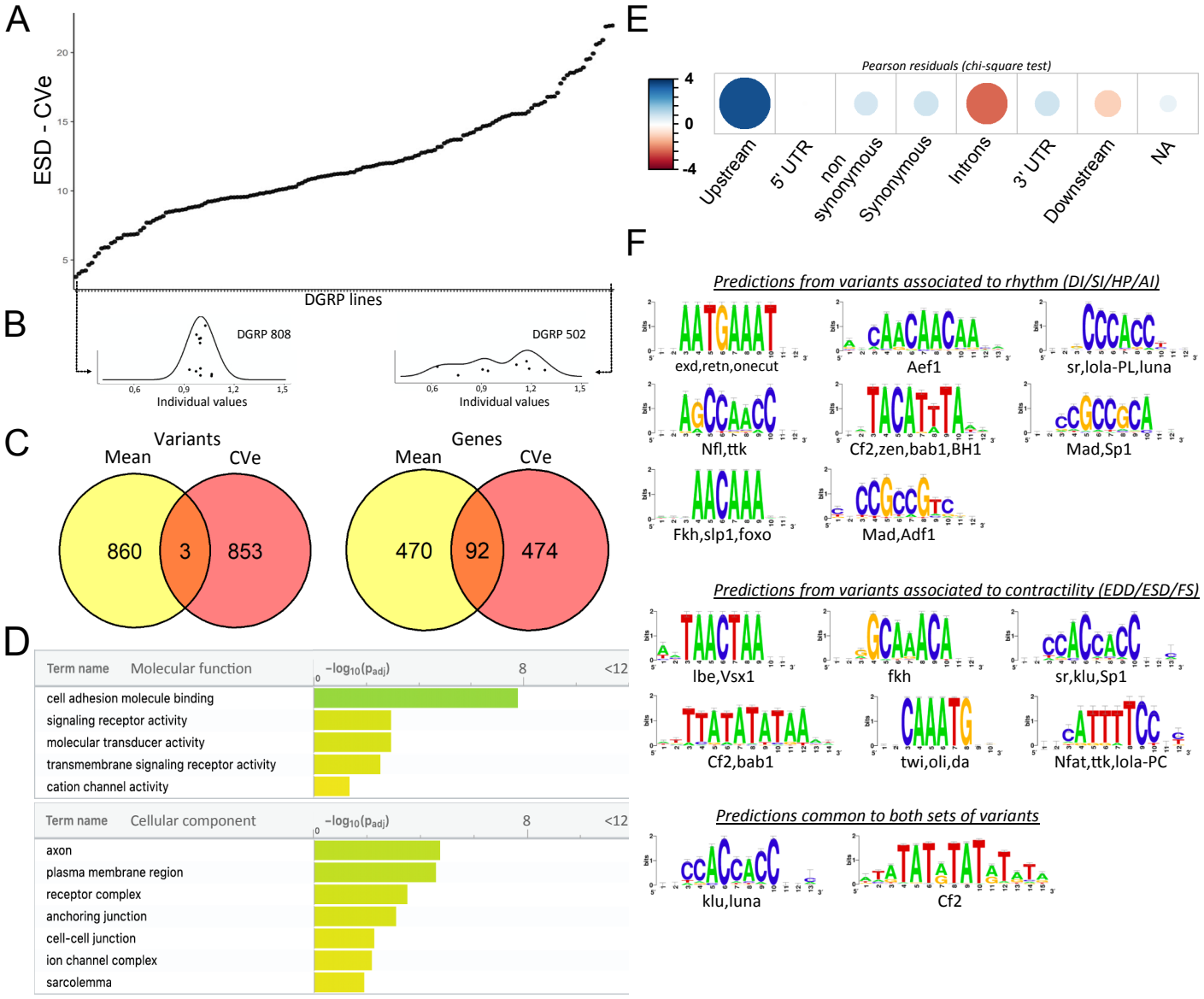

A
ESD - CVe
DGRP lines
B
DGRP 808
DGRP 502
Individual values
C
Variants
Mean
CVe
860
3
853
Genes
470
92
474
D
Term name
Molecular function
Cellular component
cell adhesion molecule binding
signaling receptor activity
molecular transducer activity
transmembrane signaling receptor activity
cation channel activity
axon
plasma membrane region
receptor complex
anchoring junction
cell-cell junction
ion channel complex
sarcolemma
E
Pearson residuals (chi-square test)
Upstream
5' UTR
non synonymous
Synonymous
Introns
3' UTR
Downstream
NA
F
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 4

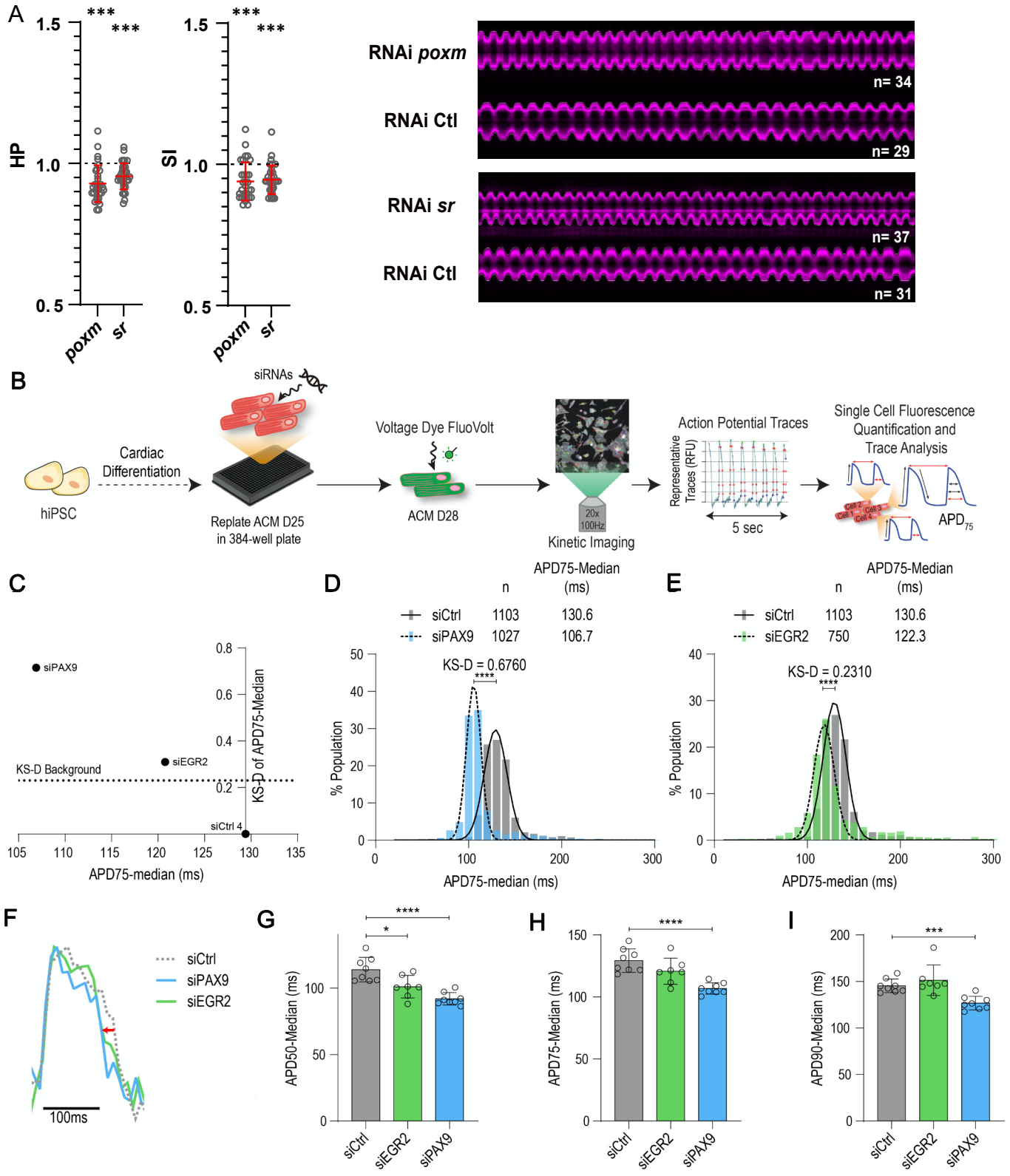

A
HP
SI
poxm
sr
RNAi poxm
RNAi Ctl
RNAi sr
RNAi Ctl
n= 34
n= 29
n= 37
n= 31
B
hiPSC
Cardiac Differentiation
siRNAs
Replate ACM D25 in 384-well plate
Voltage Dye FluoVolt
ACM D28
Kinetic Imaging
Action Potential Traces
Representative Traces (RFU)
5 sec
Single Cell Fluorescence Quantification and Trace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

Figure 5

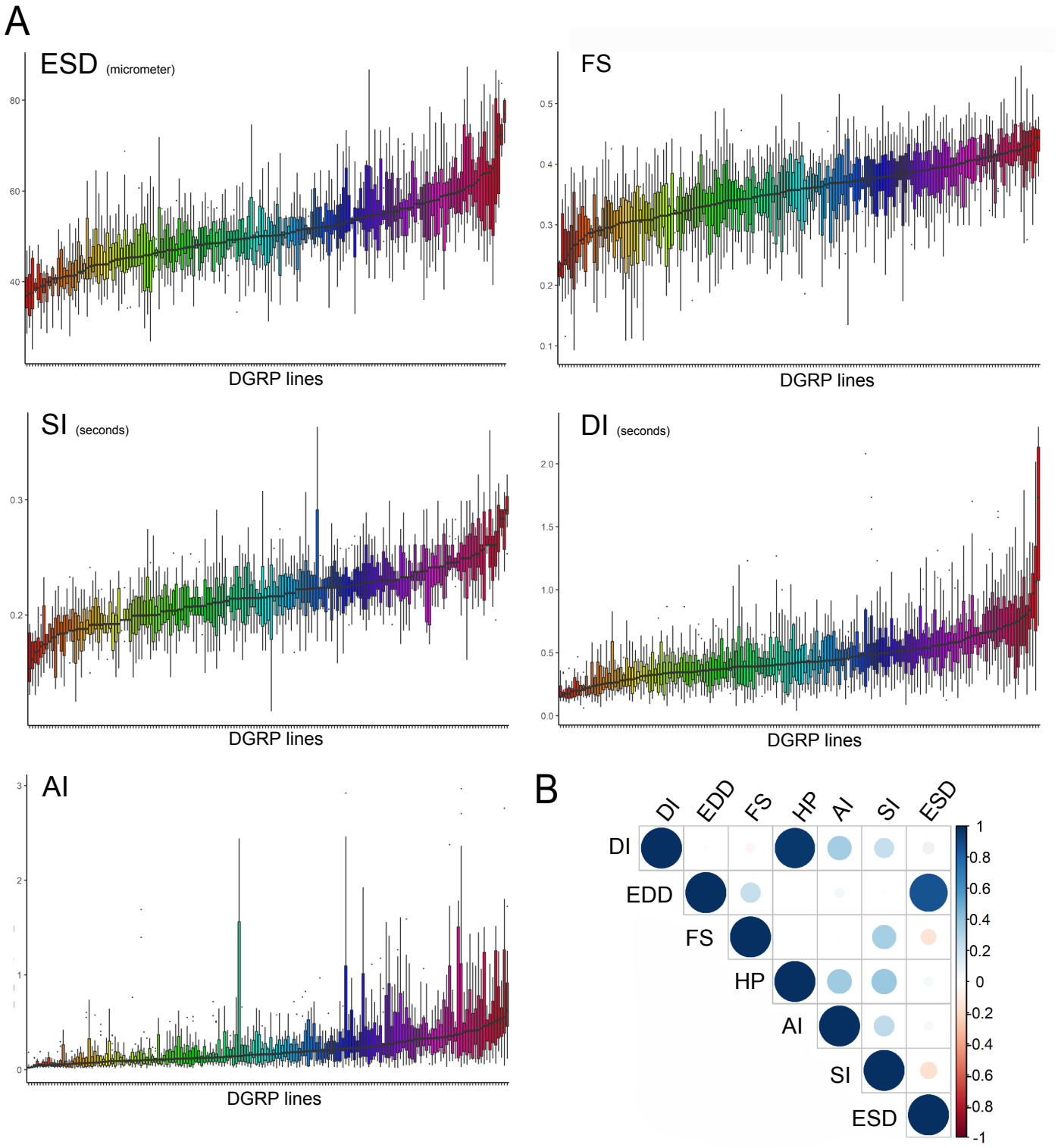
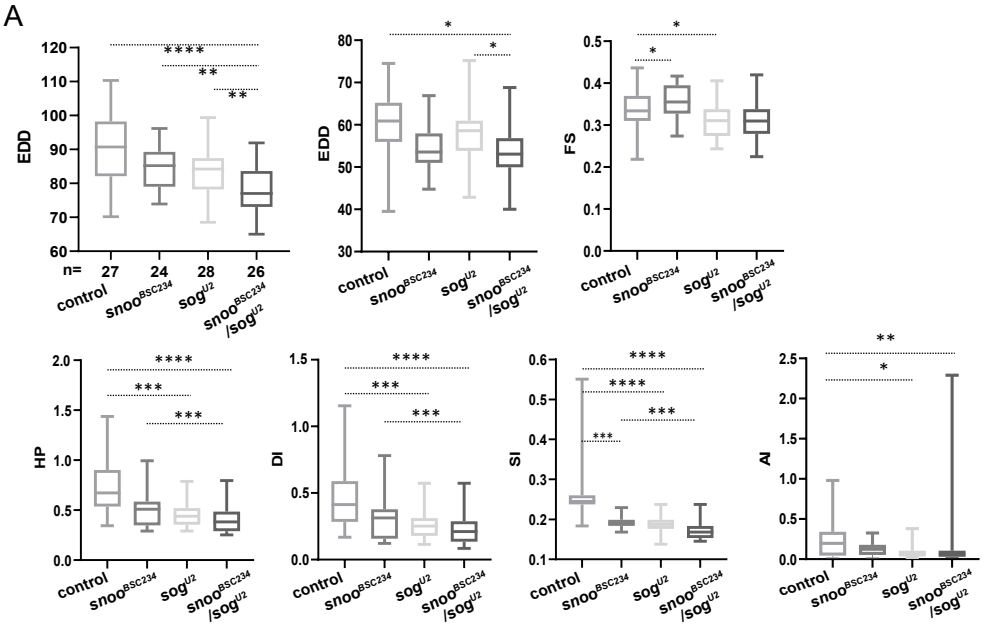
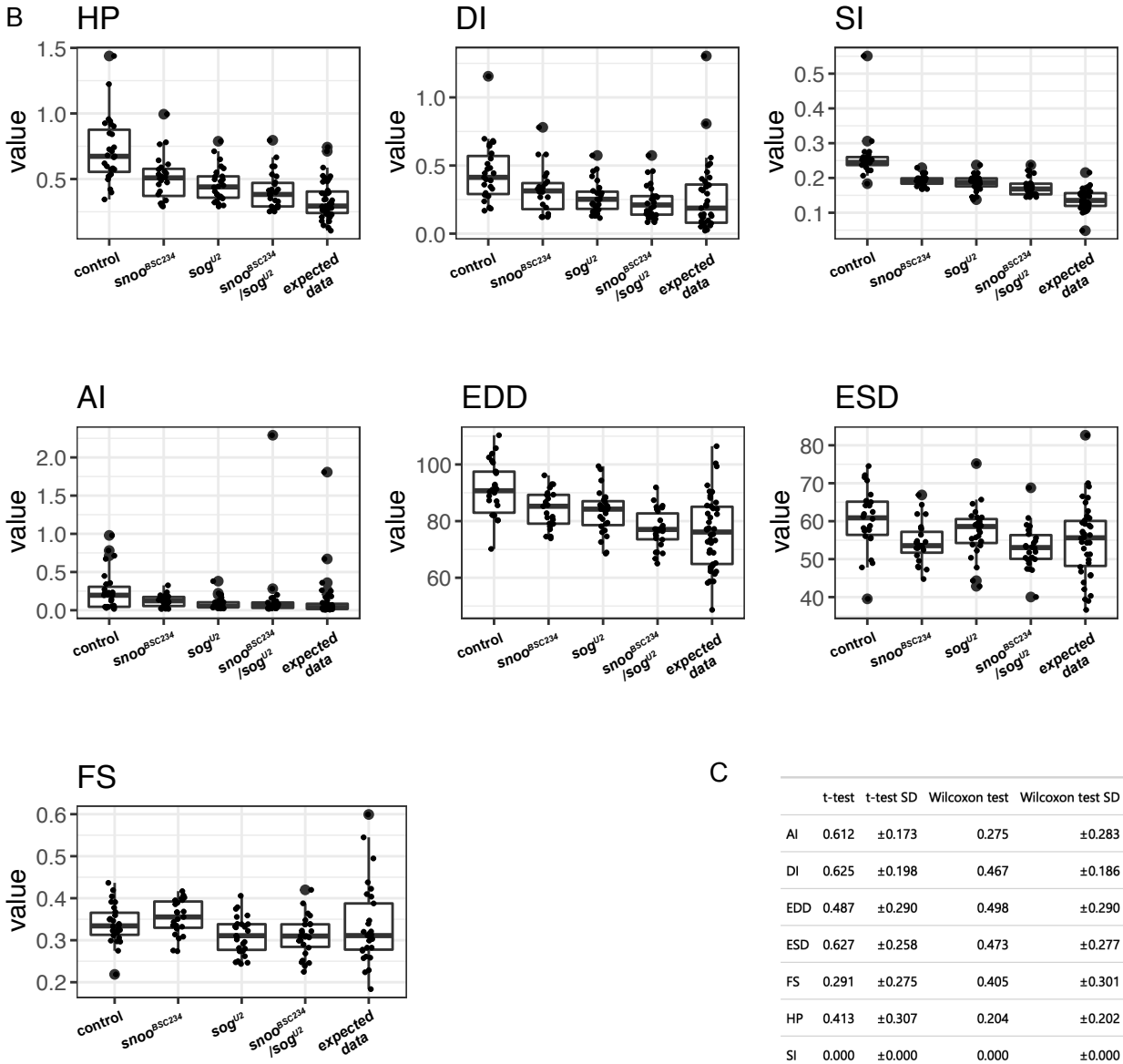

A
ESD (micrometer)
80
60
40
DGRP lines
FS
0.5
0.4
0.3
0.2
0.1
DGRP lines
SI (seconds)
0.3
0.2
DGRP lines
DI (seconds)
2.0
1.5
1.0
0.5
0.0
DGRP lines
AI
3
2
1
0
DGRP lines
B
DI
EDD
FS
HP
AI
SI
ESD
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Figure S1

A



B

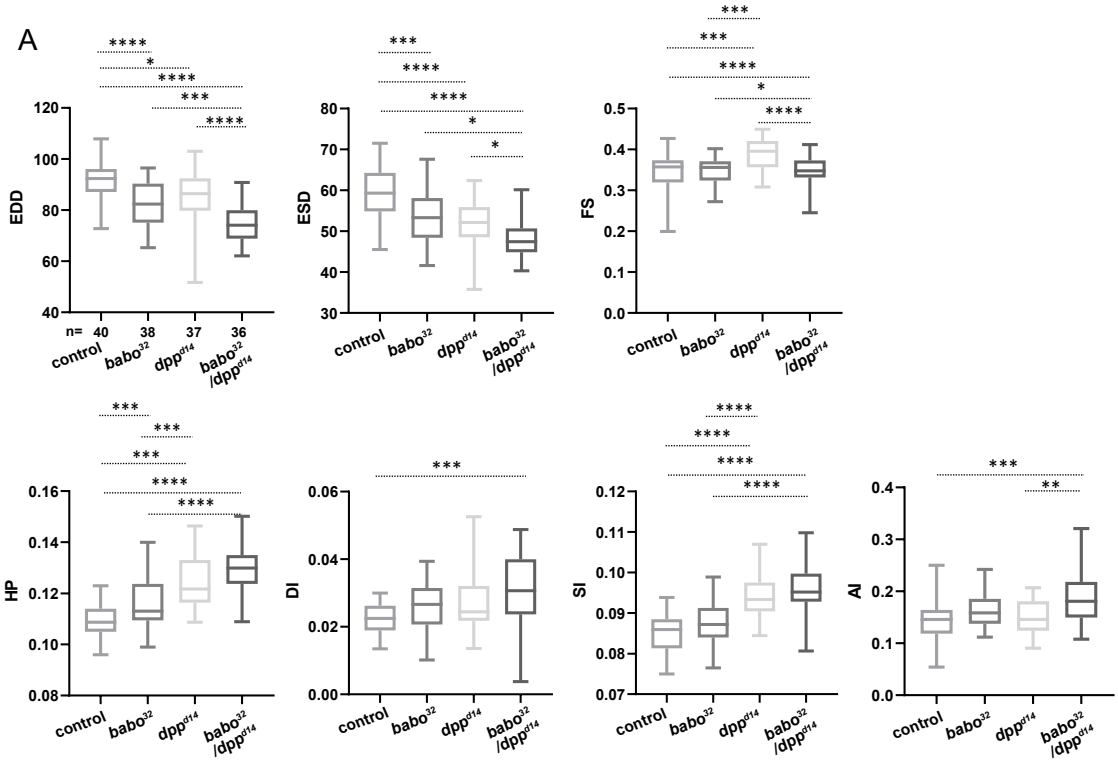


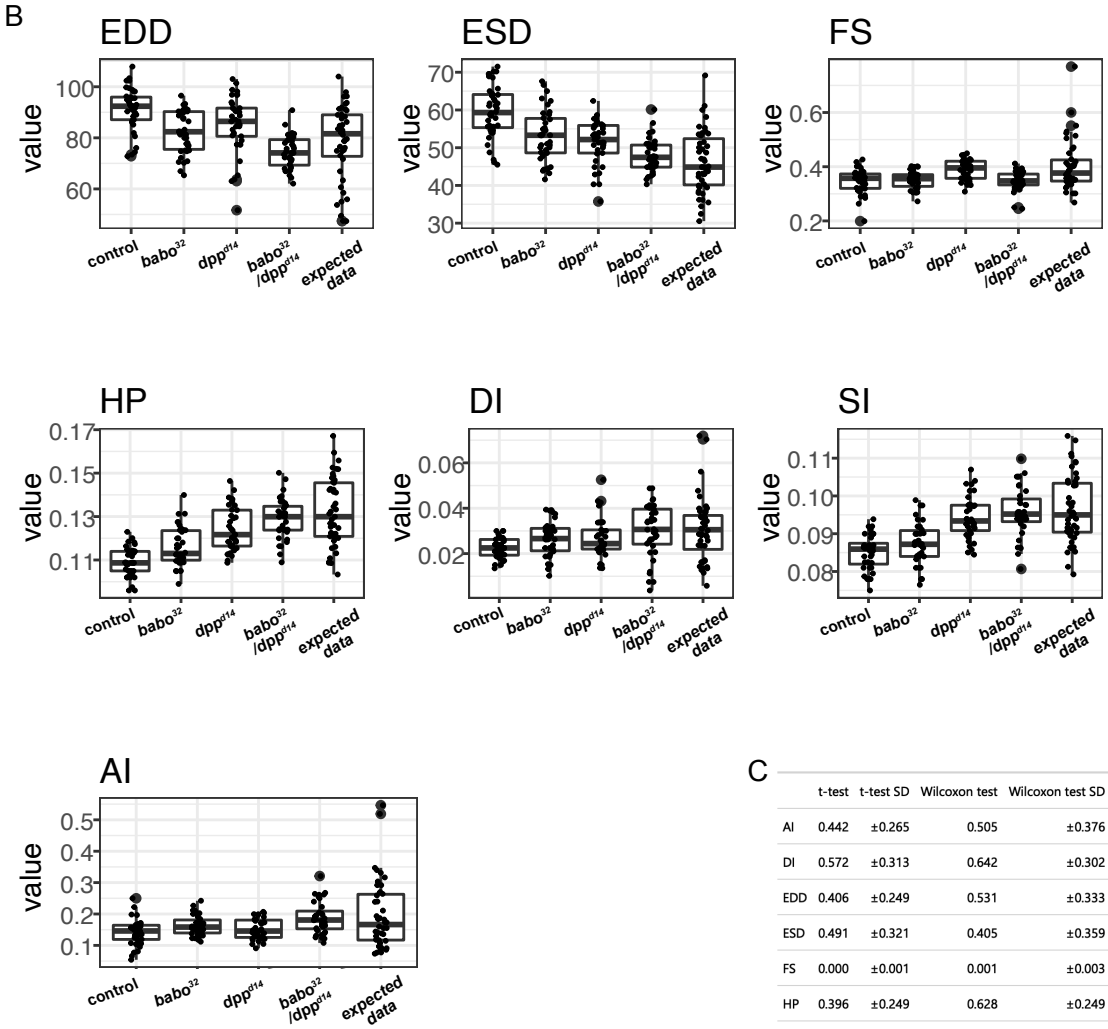
C

	t-test	t-test SD	Wilcoxon test	Wilcoxon test SD
AI	0.612	±0.173	0.275	±0.283
DI	0.625	±0.198	0.467	±0.186
EDD	0.487	±0.290	0.498	±0.290
ESD	0.627	±0.258	0.473	±0.277
FS	0.291	±0.275	0.405	±0.301
HP	0.413	±0.307	0.204	±0.202
SI	0.000	±0.000	0.000	±0.000

Mean p-values after 50 iterations

Figure S2





C

	t-test	t-test SD	Wilcoxon test	Wilcoxon test SD
AI	0.442	±0.265	0.505	±0.376
DI	0.572	±0.313	0.642	±0.302
EDD	0.406	±0.249	0.531	±0.333
ESD	0.491	±0.321	0.405	±0.359
FS	0.000	±0.001	0.001	±0.003
HP	0.396	±0.249	0.628	±0.249
SI	0.439	±0.278	0.550	±0.284

Mean p-values after 50 iterations

Figure S3

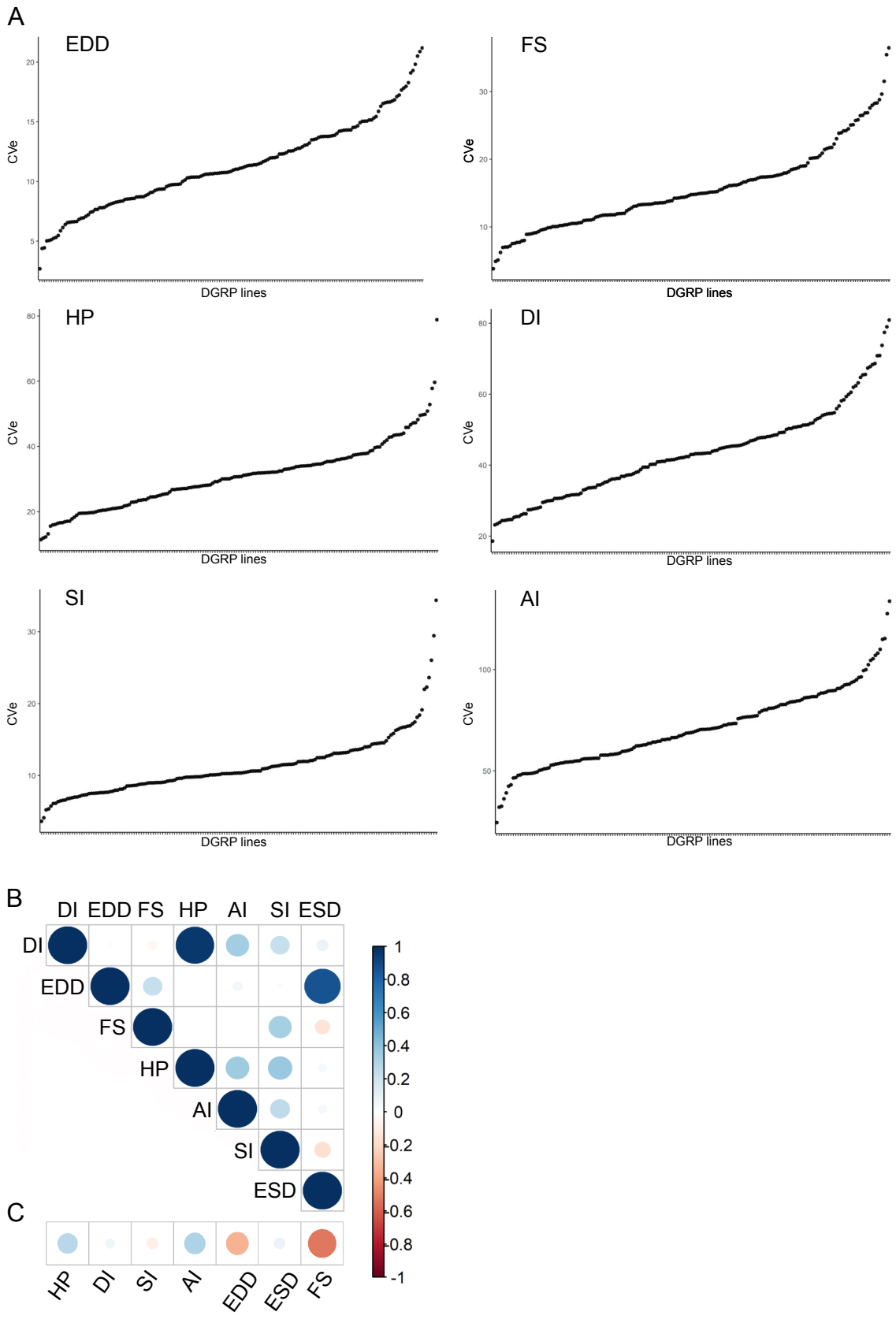

A
EDD
FS
HP
DI
SI
AI
CVe
DGRP lines
B
DI EDD FS HP AI SI ESD
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
C
HP DI SI AI EDD ESD FS

Figure S4

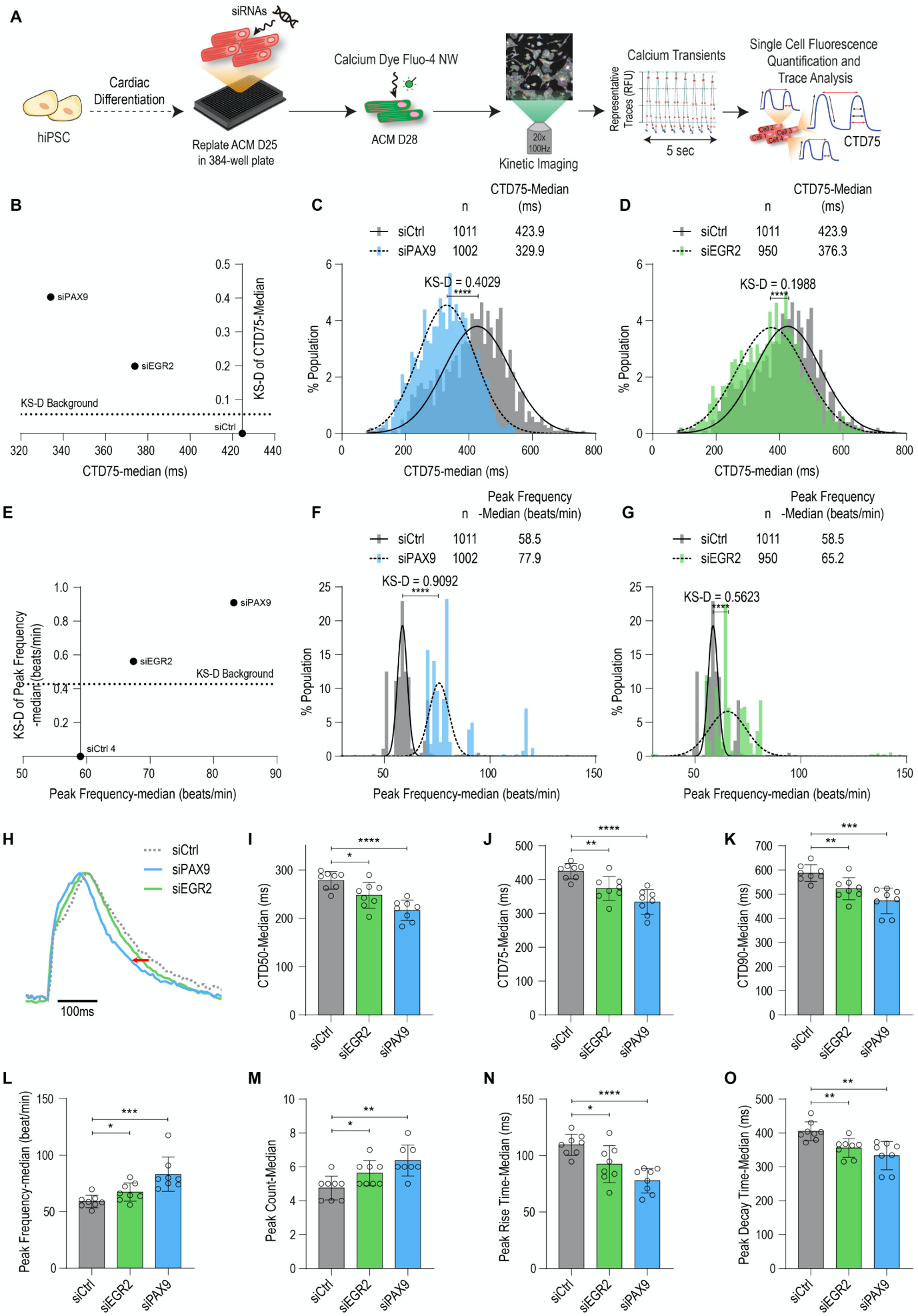

A
siRNAs
Cardiac Differentiation
hiPSC
Replate ACM D25 in 384-well plate
Calcium Dye Fluo-4 NW
ACM D28
20x 100Hz
Kinetic Imaging
Calcium Transients
Representative Traces (RFU)
5 sec
Single Cell Fluorescence Quantification and Trace Analysis
CTD75
B
siPAX9
siEGR2
KS-D Background
siCtrl
KS-D of CTD75-Median
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 S5