SUPPLEMENT DATA

SUPPLEMENT METHODS

Collection of embryos

Zebrafish transgenic lines Tg(*nxk2.5*:EGFP), Tg(*myl7*:EGFP) in AB wild-type and *gata5*$^{tm236a/+}$ (Reiter et al. 1999), *tbx5a*$^{m21/+}$ (Garrity et al. 2002), *hand2*$^{s6/+}$ (Yelon et al. 2000) mutant background were maintained in the zebrafish facilities of the International Institute of Molecular and Cell Biology in Warsaw (License no. PL14656251), according to standard procedures and ethical practices recommended. Embryos were grown in embryo medium at 28°C, staged according to standard morphological criteria (Kimmel et al. 1995), and harvested at three different developmental stages: prim-5 (24 hpf), long-pec (48 hpf) and protruding-mouth (72 hpf).

CM collection by fluorescence-activated cell sorting (FACS)

Cell suspension was prepared from 500 zebrafish embryos and larvae as previously described (Winata et al. 2013). Cells were verified microscopically for the viability by using trypan blue solution and used for further procedures when more than 90% of viable cell were obtained in the suspension. Fluorescent (GFP+) and non-fluorescent cells (GFP-) were sorted by using FACSAria II cytometer (BD Biosciences, USA). Cells were inspected for their relative size, granularity and relative fluorescence. Cell suspension obtained from wild-type embryo was used to assess the autofluorescence. GFP+ and GFP- fractions were verified for their viability by staining with propidium iodide (Sigma-Aldrich, USA) followed by FACS.

qPCR

Total RNA was extracted from 100,000 GFP+ and GFP- cells obtained from zebrafish embryos by using TRIzol LS (Thermo Fisher Scientific, USA) according to the manufacturer protocol and followed by DNase I (Life Technologies, USA) treatment. Transcriptor first strand cDNA synthesis kit (Roche Life Science, Germany) was used to obtain cDNA. Relative mRNA expression was quantified by using FastStart SYBR green master mix on the Light Cycler 96 instrument (Roche Life Science, Germany) with specific sets of primers (Supplement Table 12).

RNA-seq

For RNA sequencing 100,000 of GFP+ and GFP- cells from zebrafish embryos were sorted directly to TRIzol LS (Thermo Fisher Scientific, USA). After ethanol precipitation RNA was depleted of DNA by using DNase I treatment and purified on columns by using RNA Clean & Concentrator™-5 (Zymo Research, USA). RNA integrity was measured by RNA ScreenTape on the Agilent 2200 TapeStation system (Agilent Technologies, USA). RNA Integrity Number (RIN) was in the range from 8.5 to 10 for all the samples used for RNA-seq. Ribosomal RNA removal from 10 ng of total RNA was performed using RiboGone Kit (Clontech Laboratories, USA). cDNA synthesis for next-generation sequencing (NGS) was performed by SMARTer Universal Low Input RNA Kit (Clontech Laboratories, USA) as recommended by the manufacturer. DNA libraries were purified with Agencourt AMPure XP PCR purification beads (Beckman Coulter, USA) and DNA fragment distribution was assessed by using D1000 ScreenTape and Agilent 2200 TapeStation system (Agilent Technologies, USA). KAPA library quantification kit (Kapa Biosystems, USA) was used for qPCR-based quantification of the libraries obtained. Paired-end sequencing (2×75bp reads) was performed with NextSeq 500 sequencing system (Illumina, USA). The sequencing coverage was at least 75 million reads and 35 million reads for GFP+ and GFP-, respectively. Pearson correlation of biological replicates and read distribution over the zebrafish genome features were performed (Supplement Fig. 6 A, B)

Assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq)

For ATAC-seq 60,000 of GFP+ cells from zebrafish embryos were sorted to Hank's solution (1× HBSS, 2mg/mL BSA, 10 mM Hepes pH 8.0), centrifuged for 5 minutes at 500 × g and prepared for chromatin tagmentation as previously described (PMID: 24097267). NEBNext High-Fidelity 2 × PCR Master Mix (New England Biolabs, USA) and custom HPLC-purified primers containing Illumina-compatible indexes were used to prepare DNA sequencing libraries as previously described (Buenrostro et al. 2015). DNA libraries were purified with Agencourt AMPure XP PCR purification beads (Beckman Coulter, USA) and DNA fragment distribution was assessed by using D1000 ScreenTape and Agilent 2200 TapeStation system (Agilent Technologies, USA). KAPA library quantification kit (Kapa Biosystems, USA) was used for qPCR-based quantification of the libraries obtained. Paired-end sequencing (2×75bp reads) was performed with NextSeq500 sequencing system (Illumina, USA). The sequencing coverage was at least 90 million reads.

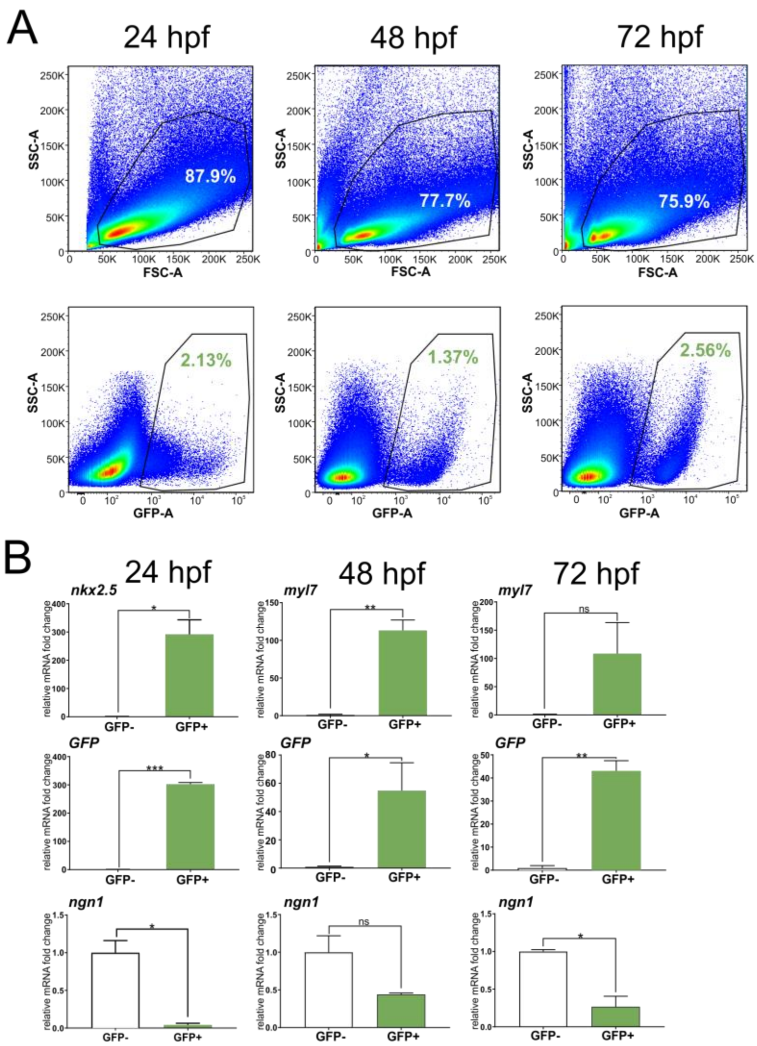
Light sheet fluorescence microscopy (LSFM)

Embryos collected from transgenic lines Tg(*nxk2.5*:GFP) and Tg(*myl7*:GFP) were maintained in embryo medium containing 0.003% 1-phenyl-2-thiourea (PTU) to inhibit the development of pigment cells. Embryos collected from wild-type at 24, 48 and 72 hpf were mounted in 1% low-melting agarose (Sigma-Aldrich, USA) in a glass capillary. LSFM was used to perform optical sectioning of the cardiomyocytes containing GFP reporter. Images were analysed with Imaris 8 software (Bitplane, Switzerland).

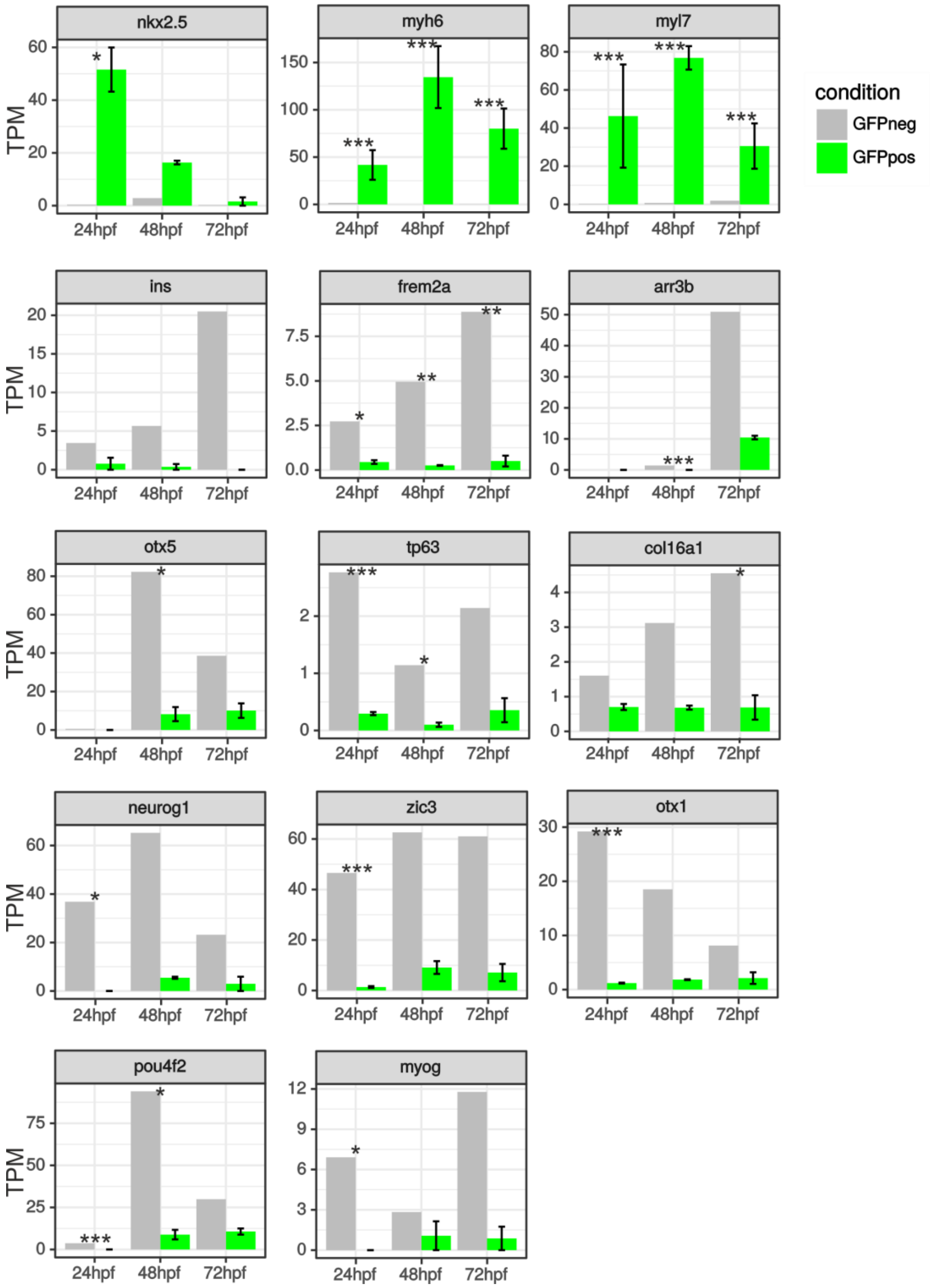
Bioinformatics analysis

Raw RNA-seq and ATAC-seq reads were quality checked using FastQC (0.11.5) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (1.1) (Ewels et al. 2016). Illumina adapters were removed using Trimmomatic (0.36) (Bolger et al. 2014). Reads matching ribosomal RNA were removed using rRNAdust (Hasegawa et al. 2014). Reads quality filtering was performed using SAMtools (1.4) (Li et al. 2009) with parameters -b -h -f 3 -F 3340 -q 30. RNA-seq reads were aligned to the zebrafish reference genome (GRCz10) using STAR (2.5) (Dobin et al. 2013) (Supplement Fig. 7). Bowtie 2 (2.2.9) (Langmead and Salzberg 2012) was used to map ATAC-seq reads to the entire GRCz10 genome except *hand2*$^{s6/s6}$ in which ~200kb region spanning hand2 gene was excluded from the analysis due to large deletion carried by those mutants as previously described (Yelon et al. 2000) (Supplement Fig. 8). Read distribution was assessed with Picard (2.10.3). NFR regions were identified as previously described (Buenrostro et al. 2013). Peaks of chromatin open regions were called using MACS2 (2.1.0) (Zhang et al. 2008) with parameters --nomodel --shift -100 --extsize 200 --broad -g 1.21e9 -q 0.05 -B --keep-dup all. Enriched motifs in NFRs were identified using the HOMER findMotifsGenome tool with parameters findMotifs.pl modules/$modir/target.fa fasta modules/$modir -mset vertebrates -p 8 -S 200 -fastaBg modules/$modir/background.fa to check against vertebrates motif collection (Heinz et al. 2010). The background collection of sequences was constructed for each investigated gene module by taking complementing set of NFRs around TSSs of that module. Downstream bioinformatics analysis were performed in R 3.4 using following Bioconductor and CRAN (Huber et al. 2015) packages: GenomicFeatures (Lawrence et al. 2013), GenomicAlignments (Lawrence et al. 2013), DESeq2 (Love et al. 2014), pheatmap, LSD, ComplexHeatmap, biomaRt (Durinck et al. 2009), dplyr, WGCNA (Langfelder and Horvath 2008), ggplot2, reshape2, org.Dr.eg.db, clusterProfiler (Yu et al. 2012), ATACseqQC (Ou et al. 2018), ChIPseeker (Yu et al. 2015), DiffBind (Ross-Innes et al. 2012), ggbio (Yin et al. 2012). RNA-seq gene counts and ATAC-seq NFR read counts for all samples were transformed to regularized log (rld) (Supplement Table 13, 14). Differentially accessible ATAC-seq peaks were quantified by DESeq2 (Supplement Table 15). Gene network visualisation and statistical analysis of gene networks was performed using Cytoscape (Cline et al. 2007). Metascape was used to visualise the output of GO enrichment analysis (Tripathi et al. 2015).

SUPPLEMENT FIGURES

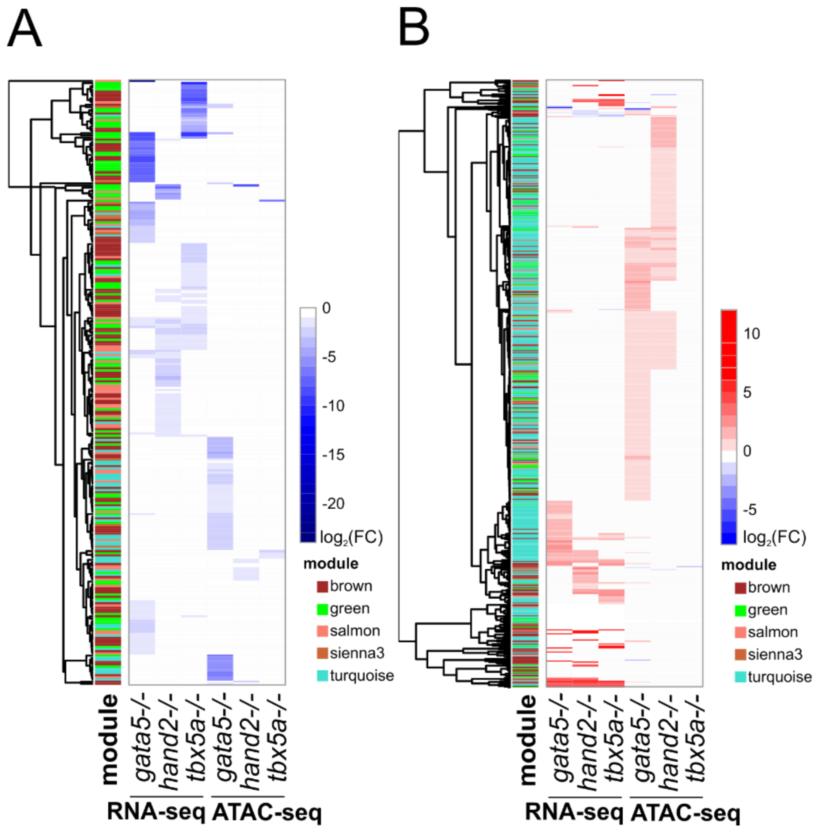


Supplement Figure 1. Biological validation of FACS-sorted cell fractions. (A) FACS analysis of cells obtained from developing zebrafish embryos. Forward scatter (FSC-A) vs side scatter (SSC-A) was used to find viable, single cell events; SSC-A vs GFP-A was applied to identify fluorescent cells. **(B)** Relative mRNA expression of nkx2.5, myl7, GFP and ngn1 in GFP+ and GFP- cells. *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.005$; by unpaired t test. Data are presented as mean ± SD, n=2/group.

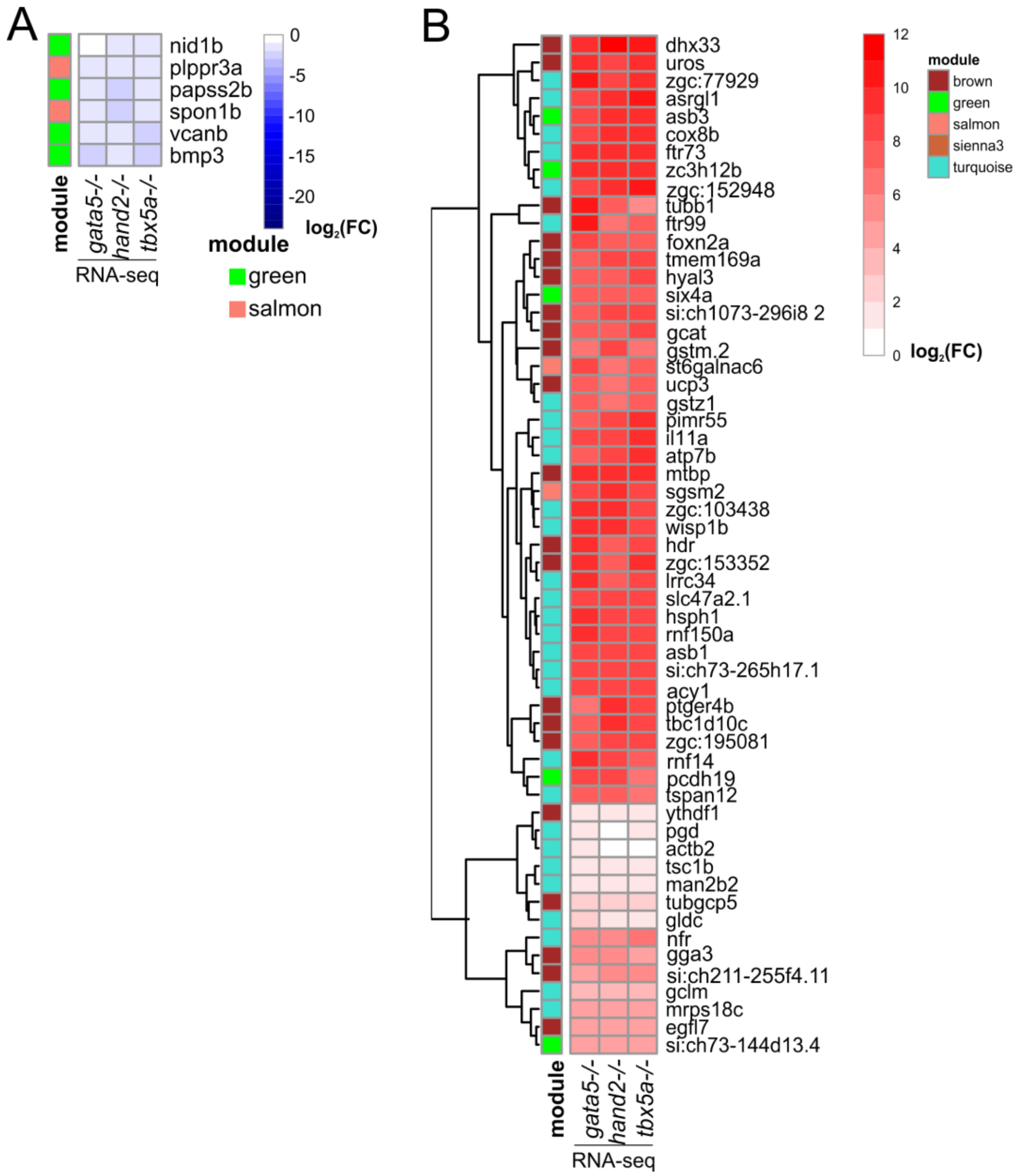


Supplement Figure 2. Assessment of CM-specific and non-CM gene markers. RNA-seq gene expression levels (transcripts per million, TPM) for CM-specific: nkx2.5 (early cardiac marker) and myl7, myh6, and non-CM markers: pancreas (ins), pharyngal arch (frem2a), retina (arr3b, otx5), skin (tp63, col16a1), neural system (neurog1, zic3, otx1), eye (pou4f2) and skeletal muscle (myog) are shown as mean ± SE, n=1/group for GFP- and n=2 for GFP+. **padj ≤ 0.01; *** padj ≤ 0.005; by Benjamini-Hochberg DESeq2 method.

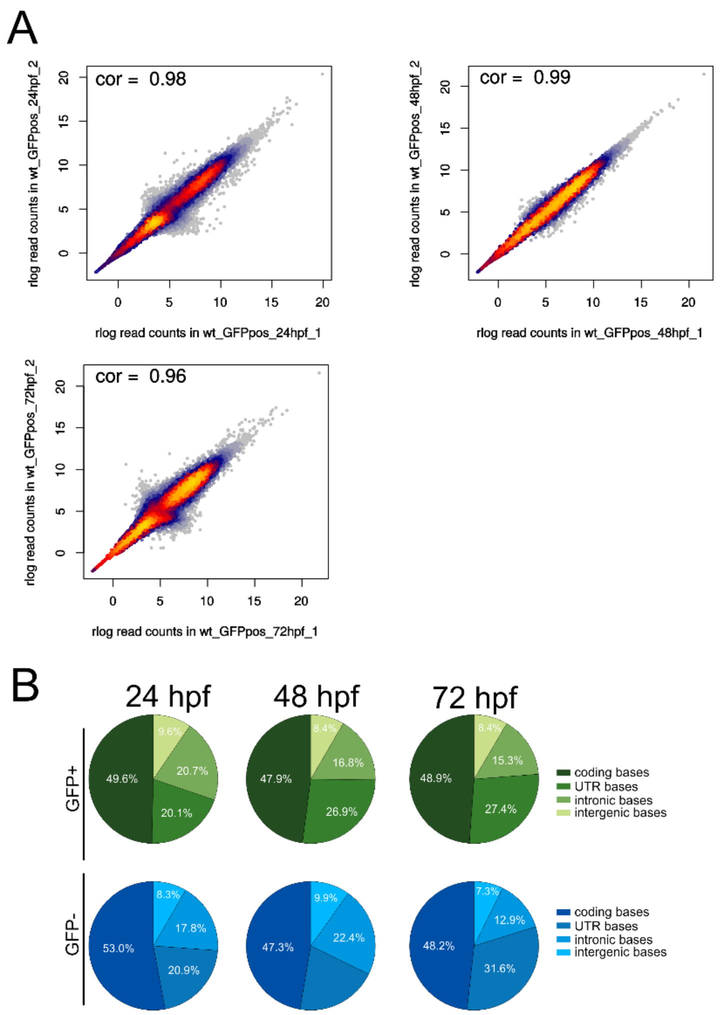
Supplement Figure 3. Cardiac module enriched HOMER-generated TF motifs. Graphical representation of Homer known motifs identified from cardiac modules. Ten motifs with p-value below 0.05 and highest number of target sequences are shown for each module.



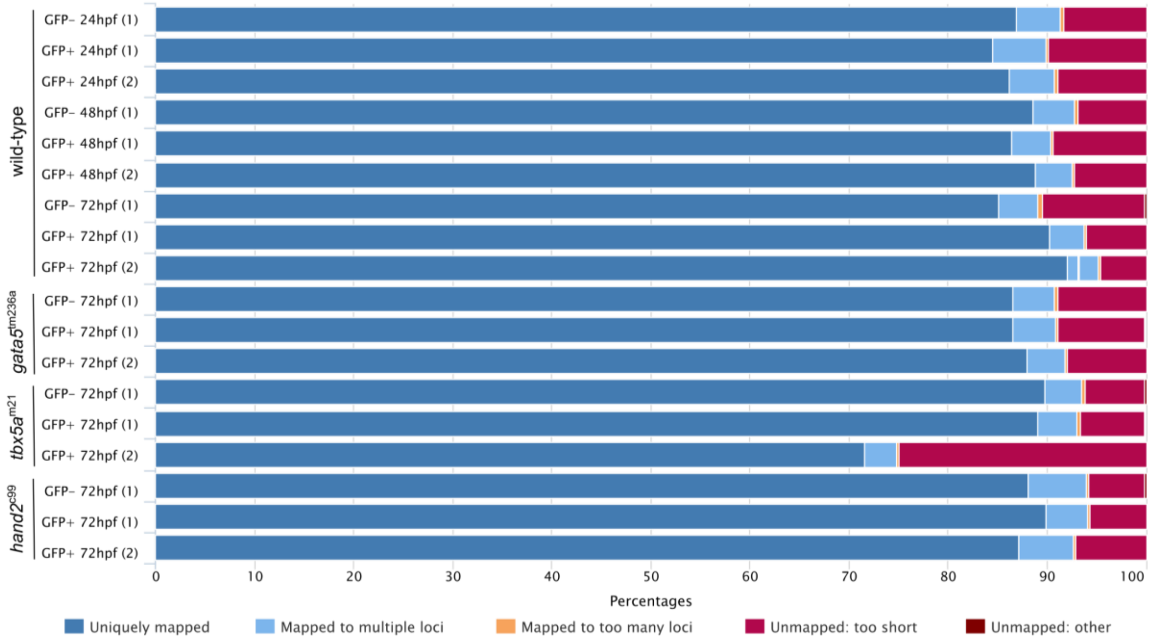
Supplement Figure 4. Hierarchical clustering of TF-differentially regulated genes and TSS proximal NFRs. (A) Hierarchical clustering of TF mutant downregulated genes and NFRs (+/- 3 kb of TSS) within cardiac regulatory modules. **(B)** Hierarchical clustering of TF mutant upregulated genes and NFRs (+/- 3 kb of TSS) within cardiac regulatory modules, padj ≤ 0.05.



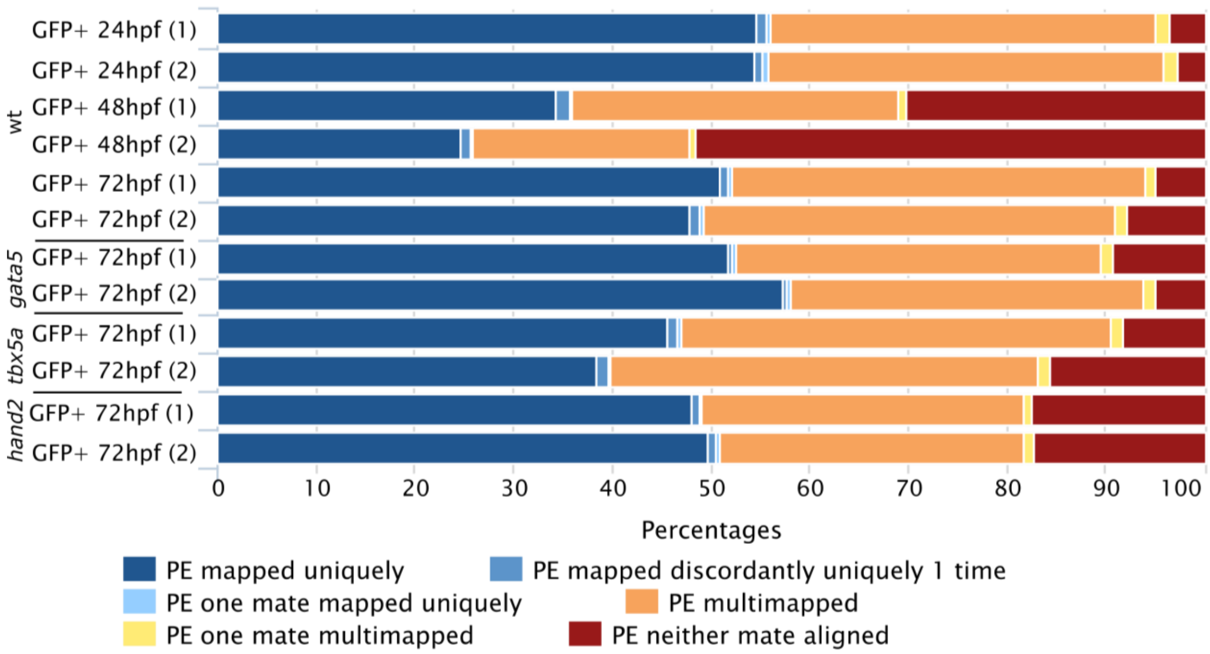
Supplement Figure 5. Hierarchical clustering of commonly TF-regulated genes. (A) Cardiac module genes commonly downregulated by TF mutants, padj ≤ 0.05. **(B)** Cardiac module genes commonly upregulated by TF mutants, padj ≤ 0.05.



Supplement Figure 6. Quality control of RNA-seq. (A) Pearson correlation of normalized reads of RNA-seq biological replicates. **(B)** RNA-seq read distribution over zebrafish genome features.



Supplement Figure 7. Alignment statistics of RNA-seq from GFP+ and GFP- cells from zebrafish cardiomyocytes.



Supplement Figure 8. Alignment statistics of ATAC-seq from zebrafish cardiomyocytes at different stage of heart development. PE (paired-end).