# SUPPLEMENTARY INFORMATION

## Additional examples of biological applications

### Plot of expression data

An example of how this database can also be applied to understand development is the study of the role of *ezrin* (*ezr*) genes. Diz-Muñoz *et al*. have shown that during zebrafish early development, inhibition of *ezrb* is sufficient to affect the mechanical properties of precordal plate progenitors (Diz-Muñoz *et al.*, 2010). However, the *ezra* gene was not analysed. Using this database, we can observe that the only paralogue expressed during the first hours of development is *ezrb* (Supplementary Fig. S6), which supports their results and explains why *ezra* gene inhibition was not necessary.

### Correlations

An example on how this database can be used to obtain general results that are not specific to regeneration is the study of proteins involved in caveolae formation. To do that, we used all the datasets in the database to find which genes are co-expressed with *cav1*. Interestingly, some of the first genes on the list, such as *ehd2b* (Moren *et al.*, 2012), *sdprb* (Hansen *et al.*, 2009), *myo1ca and myo1cb* (Hernandez *et al.*, 2013) are known to form part, or are closely associated with, caveolae. However, in the top part of this list, we find other genes, such as *ccdc187* and *gypc*, which have not been previously related to caveolae. Our results suggest a possible relationship between these genes and the caveolae (Supplementary Fig. S7).

### Venn diagrams

This database includes different datasets that study the regeneration of each organ. However, there are differences in their experimental setup, including the injury model used and the stages analysed. Using this application, we analysed three heart regeneration datasets from the laboratories of Poss (Goldman *et al.*, 2017), Stainier (Lai *et al.*, 2017) and Flores (Bednarek *et al.*, 2015). Besides the experimental differences, we found 269 differentially expressed genes in common in the three different datasets. This provides a high degree of confidence of the behaviour of these genes in response to injury (Supplementary Fig. S8).

### Methods

#### Sources of RNA-Seq datasets

We collected a total of 22 datasets and 340 samples from public platforms Gene Expression Omnibus (GEO) (Barrett *et al.*, 2012), Sequence read Archive (SRA) and Bioproject (BioProject) (Goldman *et al.*, 2017; Lai *et al.*, 2017; Han *et al.*, 2014; Bednarek *et al.*, 2015; Natarajan *et al.*, 2018; King *et al.*, 2018; Sánchez-Iranzo, Galardi-Castilla, Minguillón, *et al.*, 2018; Sánchez-Iranzo, Galardi-Castilla, Sanz-Morejón, *et al.*, 2018; Mokalled *et al.*, 2016; Kang *et al.*, 2016; King and Yin, 2016; Rabinowitz *et al.*, 2017; Jiang *et al.*, 2014; Sifuentes *et al.*, 2016; Oosterhof *et al.*, 2017; Feng *et al.*, 2015; Louie *et al.*, 2017; PRJEB1986; Pauli *et al.*, 2012; Levin *et al.*, 2016; Herman *et al.*, 2018).

Datasets were selected according to the following criteria: they can be found in GEO or Bioproject by using the words “zebrafish regeneration” or “fish regeneration”. The model organism should be zebrafish or any other fish species, and the datasets should be of general interest, excluding any experiments that are restricted to study the effect of a specific gene. In addition, developmental datasets that are representative of a wide developmental range (defined as including at least 5 different developmental stages) were included.

#### SRA to Fastq conversion

SRA files were converted into fastq files by using fastq-dump (Sequence Read Archive Handbook) using its default parameters, or --split-files for all datasets that are paired-end.

**Quality control and trim of low-quality reads and adapters**

A quality control was performed with fastqc (Andrews). Trimmomatic (Bolger *et al.*, 2014) was used to trim low-quality reads and adaptor sequences when necessary. The following parameters were used: "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36".

Specifically, Trimommatic was applied to a whole dataset when, in "per base sequence quality" section of fastqc, there was at least one of the samples with at least one base with more than 10% of the reads with a quality value (Q in the fastq format) lower than 20 (lower whisker in the red area). This applied to the following datasets: "Heart regeneration (Poss lab.)", "Heart regeneration (Stainier lab)", "Heart regeneration (Xiong lab)", "Heart regeneration (Yin lab.)", "Muller glia (Raymond lab.)", "Fin regeneration (Yin lab)", "Developmental stages (Schier lab)", "Developmental stages (PRJEB1986)".

In addition, Trimmomatic was also applied when the percentage of adapters was too high. We defined too high adapter content when they were present in more than 5% of the reads of at least one of the samples of a dataset (warning in the "adapter content" section of fastqc). This applied to the "Cardiac fibrosis (Mercader lab)", where the ILLUMINACLIP:TrueSeq3-SE.fa:2:30:10" parameter was added; and for the "Liver regeneration (Cui lab)" dataset, where the parameter "ILLUMINACLIP:TruSeq2-PE.fa:2:30:10" was added.

In all the datasets that were paired-end, the option "PE" was included and both paired files were trimmed simultaneously, while when they were single-end, the option "SE" was included.

**Alignment to the reference genome**

RSEM 1.2.25 (Li and Dewey, 2011), calling Bowtie2 2.2.6 (Langmead and Salzberg, 2012), was utilised to align the reads to the *Danio rerio* reference transcriptome (ENSEMBL release 89). Default parameters were used, except for "--estimate-rspd" in RSEM that was set to "on". The parameter "--paired-end" was included in RSEM when necessary.

RSEM aligned files for the dataset "Spinal cord regeneration P. marinus (Bloom lab.)" were directly downloaded from GEO.

"Expected counts" and "fragments per kilobase of transcript per million mapped reads" (fpkm) were used for further analyses. Fig. 1 shows a schematic representation of this workflow.

**Web interface and R analysis**

Data was analysed and displayed using R (Team., 2014). Specifically, the Shiny R (Chang W *et al.*, 2015) package was used to build the web interface.

**Plot fpkm application**

fpkm values were plotted by using ggplot2 (Wickham, 2009). In the box plots, the hinges correspond to the first and third quartiles ($25^{th}$ and $75^{th}$ percentiles). The whiskers extends from the hinges to the highest and lowest values as long as they don't extend more than 1.5 interquartile ranges from the hinges.

In datasets where there are no replicates, boxplots are substituted by lines.

Moreover, individual data points were overlaid.

In the "Developmental stages (Yanai lab.)" dataset, as there is a big number of points with only one replicate and equally spaced in time, a continuous line was displayed.

**Correlations application**

Every sample present in the selected datasets is used to calculate the correlations.

Correlations were calculated by using the ***cor*** function (stats package) with its default parameters. Correlations are calculated using the Pearson method.

t-statistic was calculated by using the following equation:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

where ***r*** is the correlation coefficient, and ***n*** is the number of samples used to calculate a correlation.

The two-tailed p-value was calculated as 2*pt(t, df = (n-1)), where ***t*** is the t-statistic, ***n*** is the number of samples used to calculate a correlation and ***pt*** the R function (stats package).

**Differentially expressed genes application**

Differentially expressed genes were calculated with the edgeR R package (Robinson *et al.*, 2010).
First, a dataframe is generated, which includes the expected counts of only the samples present in the two groups that are being compared.
Then, low-expressed genes are filtered. To keep a gene for the differential expression analysis, it has to be expressed at more than 1 count per million (cpm) level in at least as many samples as the size of the smallest group that is being compared. For example, if we are comparing a condition including 4 biological replicates to other condition including 3 biological replicates, only genes that are expressed at least at 1 cpm in 3 of the samples are kept.
Normalization is done by using the calcNormFactors function (edgeR package) with its default parameters, which includes the TMM-normalization method (trimmed mean of M-values).
Dispersion is calculated by using the estimateDisp function (edgeR package) with its defaults parameters. For datasets where there were no replicates, a bcv value of 0.4 was used.
exactTest function (edgeR package) with its default parameters is used to calculate differentially expressed genes. When there were no replicates, the parameter dispersion was set to 0.4^2.
Volcano plots are made using plotly (Plotly Technologies, 2015).

**Datasets information**

Information on each sample and links to publications were manually introduced and displayed in the database, according to the information available in GEO.

## References


Andrews,S. FastQC: A quality control tool for high throughput sequence data.

Barrett,T. *et al.* (2012) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.

Bednarek,D. *et al.* (2015) Telomerase Is Essential for Zebrafish Heart Regeneration. *Cell Rep.*, **12**, 1691–1703.

BioProject *https://www.ncbi.nlm.nih.gov/bioproject*.

Bolger,A.M. *et al.* (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Chang W *et al.* (2015) Shiny: Web Application Framework for R. *http://cran.r-project.org/package=shiny*.

Diz-Muñoz,A. *et al.* (2010) Control of Directed Cell Migration In Vivo by Membrane-to-Cortex Attachment. *PLoS Biol.*, **8**, e1000544.

Feng,G. *et al.* (2015) Transcriptomic characterization of the dorsal lobes after hepatectomy of the ventral lobe in zebrafish. *BMC Genomics*, **16**, 1–11.

Goldman,J.A. *et al.* (2017) Resolving Heart Regeneration by Replacement Histone Profiling. *Dev. Cell*, **40**, 392–404.e5.

Han,P. *et al.* (2014) Hydrogen peroxide primes heart regeneration with a derepression mechanism. *Cell Res.*, **24**, 1091–1107.

Hansen,C.G. *et al.* (2009) SDPR induces membrane curvature and functions in the formation of caveolae. *Nat. Cell Biol.*, **11**, 807–814.

Herman,P.E. *et al.* (2018) Highly conserved molecular pathways, including Wnt signaling, promote functional recovery from spinal cord injury in lampreys. *Sci. Rep.*, **8**, 742.

Hernandez,V.J. *et al.* (2013) Cavin-3 dictates the balance between ERK and Akt signaling. *Elife*, **2**, e00905.

Jiang,L. *et al.* (2014) Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. *Proc. Natl. Acad. Sci.*, **111**, E1383–E1392.

Kang,J. *et al.* (2016) Modulation of tissue repair by regeneration enhancer elements. *Nature*, **532**, 201–206.

King,B.L. *et al.* (2018) RegenDbase: a comparative database of noncoding RNA regulation of tissue regeneration circuits across multiple taxa. *npj Regen. Med.*, **3**, 10.

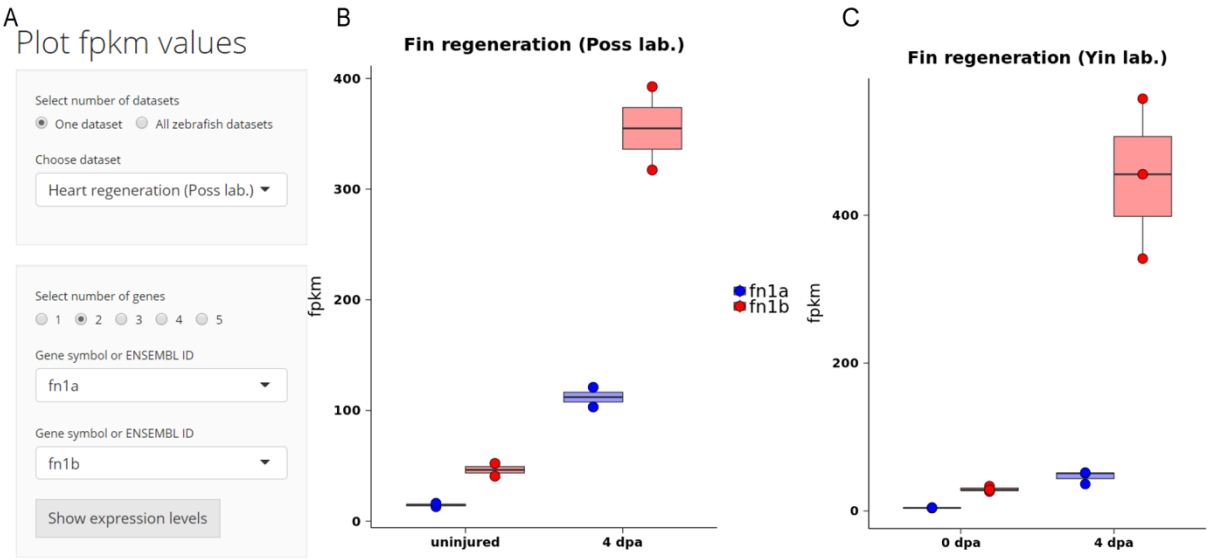
King,B.L. and Yin,V.P. (2016) A conserved microRNA regulatory circuit is differentially controlled during limb/appendage regeneration. *PLoS One*, **11**, 1–25.

Lai,S.L. *et al.* (2017) Reciprocal analyses in zebrafish and medaka reveal that harnessing the immune response promotes cardiac regeneration. *Elife*, **6**, 1–20.

Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Levin,M. *et al.* (2016) The mid-developmental transition and the evolution of animal body plans. *Nature*, **531**, 637–641.
Li,B. and Dewey,C.N. (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
Louie,K.W. *et al.* (2017) Temporally distinct transcriptional regulation of myocyte dedifferentiation and Myofiber growth during muscle regeneration. *BMC Genomics*, **18**, 854.
Mokalled,M.H. *et al.* (2016) Injury-induced ctgfa directs glial bridging and spinal cord regeneration in zebrafish. *Science (80-. ).*, **354**, 630–634.
Moren,B. *et al.* (2012) EHD2 regulates caveolar dynamics via ATP-driven targeting and oligomerization. *Mol. Biol. Cell*, **23**, 1316–1329.
Natarajan,N. *et al.* (2018) Complement Receptor C5aR1 Plays an Evolutionarily Conserved Role in Successful Cardiac Regeneration. *Circulation*, **137**, 2152–2165.
Oosterhof,N. *et al.* (2017) Identification of a conserved and acute neurodegeneration-specific microglial transcriptome in the zebrafish. *Glia*, **65**, 138–149.
Pauli,A. *et al.* (2012) Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.*, **22**, 577–591.
Plotly Technologies,I. (2015) Collaborative data science Publisher: Plotly Technologies Inc.
PRJEB1986 *https://www.ncbi.nlm.nih.gov/bioproject/PRJEB1986*.
Rabinowitz,J.S. *et al.* (2017) Transcriptomic, proteomic, and metabolomic landscape of positional memory in the caudal fin of zebrafish. *Proc. Natl. Acad. Sci.*, **114**, E717–E726.
Robinson,M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–40.
Sánchez-Iranzo,H., Galardi-Castilla,M., Minguillón,C., *et al.* (2018) Tbx5a lineage tracing shows cardiomyocyte plasticity during zebrafish heart regeneration. *Nat. Commun.*, **9**, 428.
Sánchez-Iranzo,H., Galardi-Castilla,M., Sanz-Morejón,A., *et al.* (2018) Transient fibrosis resolves via fibroblast inactivation in the regenerating zebrafish heart. *Proc. Natl. Acad. Sci.*, **115**, 4188–4193.
Sifuentes,C.J. *et al.* (2016) Rapid, dynamic activation of Müller Glial stem cell responses in Zebrafish. *Investig. Ophthalmol. Vis. Sci.*, **57**, 5148–5160.
SRA *https://www.ncbi.nlm.nih.gov/sra/*.
Team.,R.C. (2014) R: a language and environment for statistical computing. *http://www.r-project.org*.
Wickham,H. (2009) ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

## Supplementary Figures




**Supplementary Figure 1. Expression of the *fn1a* and *fn1b* genes by using the "Plot fpkm" application.** (A) Screenshot of the selection menu. Screenshot of the plots showing the expression levels of *fn1a* and *fn1b* in the "Fin regeneration (Poss lab.)" dataset (B) and the "Fin regeneration (Yin lab.)" dataset (C).

A

## Calculate correlations

Choose datasets

- [x] Heart regeneration (Poss lab.)
- [x] Heart regeneration (Stainier lab.)
- [x] Heart regeneration (Xiong lab.)
- [x] Heart regeneration (Flores lab.)
- [x] Heart regeneration (Lee lab.)
- [x] Heart regeneration (Yin lab.)
- [ ] Cardiomyocyte subtypes (Mercader lab.)
- [x] Cardiac fibrosis (Mercader lab.)
- [x] Cardiac fibroblast inactivation (Mercader lab.)
- [x] Spinal cord regeneration (Poss lab.)
- [x] Fin regeneration (Poss lab.)
- [x] Fin regeneration (Yin lab.)
- [ ] Fin proximodistal (Moon lab.)
- [x] Lateral line (Piotrowski lab.)
- [ ] Muller glia (Raymond lab.)
- [ ] Microglia (van Ham lab.)
- [x] Liver regeneration (Cui lab.)
- [x] Skeletal muscle (Kahana lab.)
- [ ] Developmental stages (PRJEB1986)
- [ ] Developmental stages (Schier lab.)
- [ ] Developmental stages (Yanai lab.)

Gene symbol or EMSEMBL ID

col1a2

Calculate correlated genes

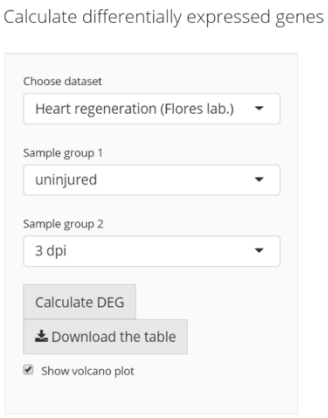
Download the table

B

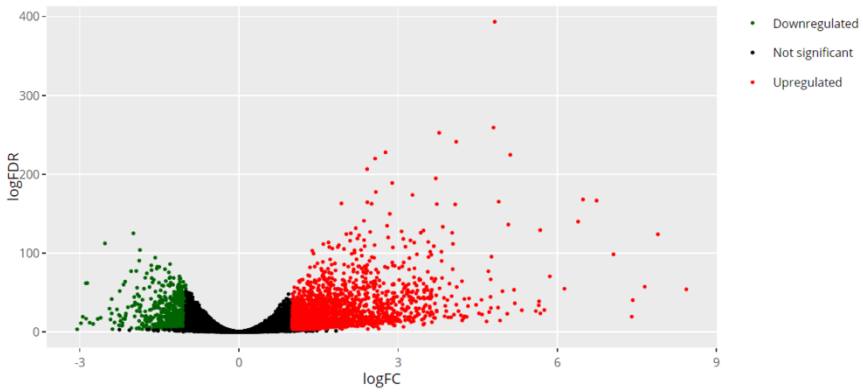
| Ensembl gene ID | Gene symbol | Correlations |
|---|---|---|
| ENSDARG00000020007 | col1a2 | 1.0000000 |
| ENSDARG00000035809 | col1a1b | 0.9843652 |
| ENSDARG00000002235 | mmp14a | 0.9346462 |
| ENSDARG00000012405 | col1a1a | 0.9170508 |
| ENSDARG00000009014 | col11a1b | 0.9130619 |
| ENSDARG00000059367 | mfap2 | 0.9095764 |
| ENSDARG00000031678 | col5a2a | 0.8895317 |
| ENSDARG00000025641 | gli2a | 0.8766321 |
| ENSDARG00000079049 | cercam | 0.8713788 |
| ENSDARG00000044074 | loxl2b | 0.8639827 |
| ENSDARG00000028071 | bmp1a | 0.8596621 |
| ENSDARG00000094752 | rpe65b | 0.8456597 |
| ENSDARG00000021948 | tnc | 0.8444393 |
| ENSDARG00000104267 | postnb | 0.8397433 |
| ENSDARG00000089162 | afap1l1a | 0.8387874 |
| ENSDARG00000019353 | sparc | 0.8331497 |
| ENSDARG00000068036 | tmem119b | 0.8322230 |
| ENSDARG00000076862 | fam198a | 0.8303862 |
| ENSDARG00000036558 | col18a1 | 0.8297215 |
| ENSDARG00000078494 | ADAMTS14 | 0.8265787 |
| ENSDARG00000059693 | adam19a | 0.8255913 |
| ENSDARG00000026165 | col11a1a | 0.8253985 |
| ENSDARG00000032831 | htra1a | 0.8221051 |
| ENSDARG00000020072 | thbs4b | 0.8220285 |
| ENSDARG00000102464 | wnt5b | 0.8214248 |

**Supplementary Figure 2. Correlation of the *col1a2* expression with all the other genes in the zebrafish genome.** All the regeneration data sets were used to calculate correlations. (A) Selection panel of the datasets and the gene to be used to calculate the correlations. (B) Genes showing a higher correlation value with the selected gene.

A




B

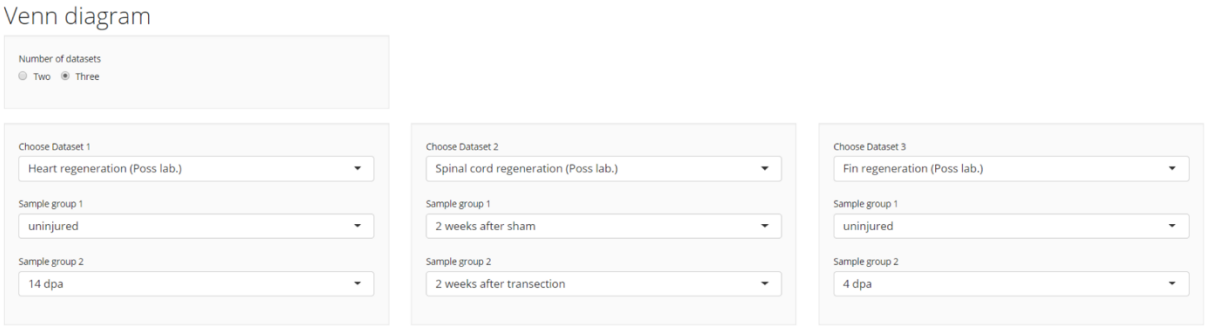



C

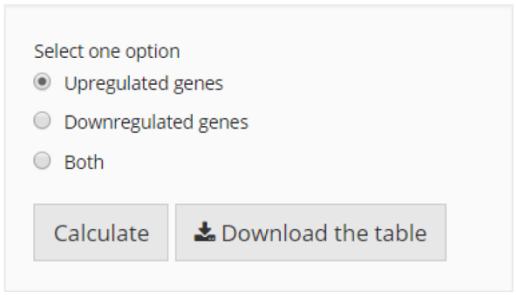
| Ensembl gene ID | Symbol | logFC | logCPM | PValue | FDR | logFDR | Legend |
|---|---|---|---|---|---|---|---|
| ENSDARG00000075263 | ankrd1a | 4.821193 | 5.324303 | 7.309077e-176 | 1.196204e-171 | 393.5629 | Upregulated |
| ENSDARG00000019815 | fn1a | 4.797434 | 9.346662 | 2.904799e-117 | 2.376997e-113 | 259.3263 | Upregulated |
| ENSDARG00000037997 | tubb5 | 3.773298 | 5.001189 | 3.530555e-114 | 1.926036e-110 | 252.6289 | Upregulated |
| ENSDARG00000009014 | col11a1b | 4.094623 | 4.436410 | 3.917284e-109 | 1.602757e-105 | 241.2997 | Upregulated |
| ENSDARG00000053091 | DAB2 | 2.762112 | 7.425201 | 3.472724e-103 | 1.136692e-99 | 227.8278 | Upregulated |
| ENSDARG00000044125 | txn | 5.114308 | 5.721516 | 9.187156e-102 | 2.505950e-98 | 224.7347 | Upregulated |
| ENSDARG00000011821 | plod2 | 2.565402 | 6.474365 | 1.235004e-99 | 2.887440e-96 | 219.9878 | Upregulated |
| ENSDARG00000013415 | lmna | 2.414015 | 6.835250 | 8.625862e-94 | 1.764636e-90 | 206.6647 | Upregulated |
| ENSDARG00000036036 | mdka | 3.709869 | 8.300990 | 1.230288e-88 | 2.237211e-85 | 194.9145 | Upregulated |
| ENSDARG00000055226 | slc7a7 | 2.888318 | 5.142137 | 5.067317e-86 | 8.293170e-83 | 188.9991 | Upregulated |
| ENSDARG00000010434 | clu | 2.578411 | 5.508846 | 5.159267e-81 | 7.676051e-78 | 177.5635 | Upregulated |
| ENSDARG00000059049 | zgc:174904 | 3.270931 | 6.183739 | 2.504511e-79 | 3.415736e-76 | 173.7681 | Upregulated |
| ENSDARG00000089362 | grn1 | 6.484748 | 6.833140 | 8.949943e-77 | 1.126729e-73 | 167.9694 | Upregulated |
| ENSDARG00000088641 | grn2 | 6.740608 | 5.171708 | 3.951698e-76 | 4.619534e-73 | 166.5584 | Upregulated |
| ENSDARG00000096979 | NPC2 | 4.895844 | 4.979053 | 1.513245e-75 | 1.651051e-72 | 165.2847 | Upregulated |
| ENSDARG00000043081 | ctsz | 2.417885 | 6.743727 | 3.317210e-75 | 3.393091e-72 | 164.5644 | Upregulated |
| ENSDARG00000040251 | ctsk | 1.931087 | 7.422065 | 1.685078e-74 | 1.622235e-71 | 162.9997 | Upregulated |
| ENSDARG00000095627 | c1qc | 2.499741 | 5.928172 | 2.734149e-74 | 2.485949e-71 | 162.5729 | Upregulated |
| ENSDARG00000093748 | si:ch211-217k17.11 | 3.729388 | 4.689974 | 3.933477e-74 | 3.388173e-71 | 162.2633 | Upregulated |
| ENSDARG00000001452 | adam8a | 4.074256 | 5.062048 | 6.434118e-74 | 5.265039e-71 | 161.8225 | Upregulated |
| ENSDARG00000040178 | havcr1 | 2.842657 | 4.354924 | 9.905156e-69 | 7.719418e-66 | 149.9269 | Upregulated |
| ENSDARG00000044613 | c1qa | 2.354285 | 5.550918 | 7.425902e-65 | 5.524196e-62 | 141.0511 | Upregulated |
| ENSDARG00000019601 | col12a1b | 6.392603 | 6.613110 | 2.217030e-64 | 1.577562e-61 | 140.0018 | Upregulated |
| ENSDARG00000008803 | marcksb | 5.077147 | 3.730108 | 9.914065e-63 | 6.760566e-60 | 136.2440 | Upregulated |
| ENSDARG00000093440 | tnfaip6 | 2.790729 | 5.375003 | 4.719312e-62 | 3.089450e-59 | 134.7245 | Upregulated |

**Supplementary Figure 3. Screenshot of the "Differential expression" application.** (A) Selection panel. (B) Interactive volcano plot. (C) Differentially expressed genes between the selected samples.

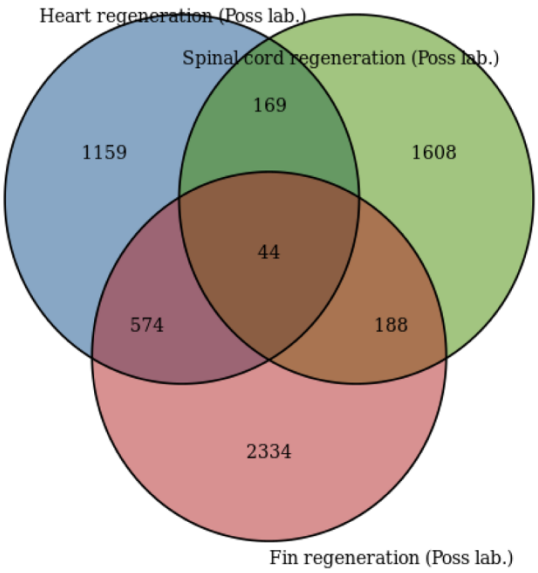
A




B




C




D

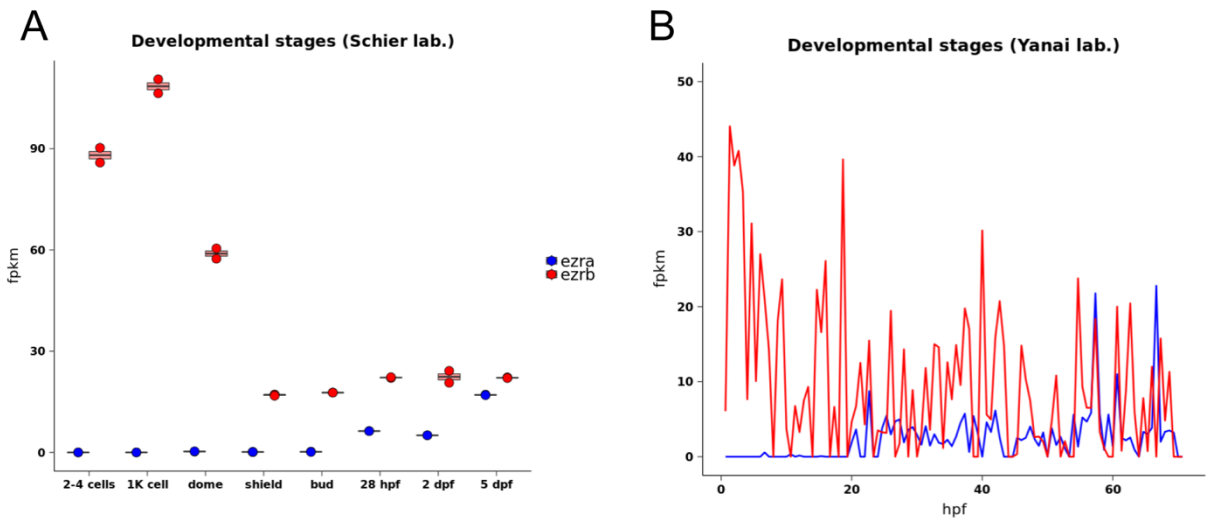
Common genes in all comparisons

| Ensembl gene ID | Symbol | Mean FDR |
|---|---|---|
| ENSDARG00000078362 | TNC | 1.202395e-60 |
| ENSDARG00000019949 | serpinh1b | 1.129996e-44 |
| ENSDARG00000011797 | fam46bb | 5.170351e-38 |
| ENSDARG00000037145 | slc8a4b | 5.507456e-37 |
| ENSDARG00000044132 | ogn | 2.722156e-35 |
| ENSDARG00000070597 | prelp | 6.923048e-32 |
| ENSDARG00000077356 | sypl1 | 1.347770e-30 |
| ENSDARG00000003259 | loxa | 8.251567e-30 |
| ENSDARG00000070141 | si:ch211-191i18.2 | 1.318002e-25 |
| ENSDARG00000026165 | col11a1a | 2.516242e-23 |
| ENSDARG00000068036 | tmem119b | 4.544060e-20 |
| ENSDARG00000076357 | CABZ01102039.1 | 5.244201e-20 |
| ENSDARG00000071658 | ywhag2 | 1.518708e-18 |
| ENSDARG00000045932 | cpeb1a | 3.242530e-17 |
| ENSDARG00000035809 | col1a1b | 7.142868e-17 |
| ENSDARG00000021720 | col7a1 | 1.066579e-16 |
| ENSDARG00000045453 | f13a1a.1 | 2.568778e-16 |
| ENSDARG00000006526 | fn1b | 5.604598e-15 |
| ENSDARG00000078494 | ADAMTS14 | 1.115966e-14 |
| ENSDARG00000032831 | htra1a | 7.459753e-14 |
| ENSDARG00000069017 | elnb | 4.630516e-13 |
| ENSDARG00000006901 | AEBP1 | 5.480482e-13 |
| ENSDARG00000074908 | col6a1 | 1.294978e-12 |
| ENSDARG00000062880 | cntn3a.1 | 1.894471e-12 |
| ENSDARG00000007219 | actn1 | 5.637597e-12 |

**Supplementary Figure 4. Up-regulated genes in three different organs during regeneration: heart, spinal cord and fin.** (A, B) Screenshot of the selection menu. (C) Venn diagram. (D) List of the genes in common among the three lists. Geometric mean of the FDR values is shown and genes are ordered according to the descending order of this value.

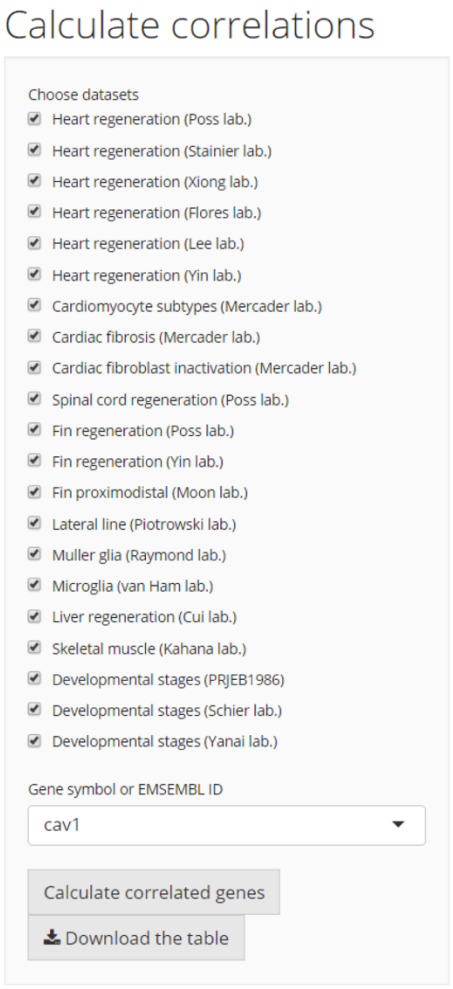
| Dataset | GEO | Pubmed | Samples description |
|---|---|---|---|
| Heart regeneration (Poss lab.) | GSE81865 | Resolving Heart Regeneration by Replacement Histone Profiling. Goldman JA, Kuzu G, Lee N, Karasik J et al. Dev Cell 2017 Feb 27;40(4):392-404. | Gene expression profile of heart ventricles after cardiomyocyte genetic ablation. RNA-Seq from uninjured hearts (uninjured) and 14-day post-injured (14 dpi). Two replicates per condition. |
| Heart regeneration (Stainier lab.) | GSE94617 | Reciprocal analyses in zebrafish and medaka reveal that harnessing the immune response promotes cardiac regeneration. Lai SL, Marn-Juez R, Moura PL, Kuenne C et al. Elife 2017 June 20;6. | RNA-Seq of four heart ventricles pooled at time 0 (uninjured), 6 h, and 1, 2, 3, 4, and 5 days post-sham (6 hps, 1 dps, 2 dps, 3 dps, 5 dps) and after cryoinjury (6 hpi, 1 dpi, 2 dpi, 3 dpi, 5 dpi). One replicate per condition. |
| Heart regeneration (Xiong lab.) | GSE50203 | Hydrogen peroxide primes heart regeneration with a derepression mechanism. Han P, Zhou XH, Chang N, Xiao CL et al. Cell Res 2014 Sep;24(9):1091-107. | RNA-Seq of one heart after sham (sham) or 7 days after amputation (7 dpa). |
| Heart regeneration (Flores lab.) | GSE71755 | Telomerase Is Essential for Zebrafish Heart Regeneration. Bednarek D, Gonzalez-Rosa JM, Guzman-Martinez G, Gutierrez-Gutierrez O et al. Cell Rep 2015 Sep 8;12(10):1691-703. | Transcriptome of hearts under the control conditions (uninjured) and 3 days after cryoinjury (3 dpi). Four replicates per condition. |
| Heart regeneratoin (Lee lab.) | GSE108493 | Complement Receptor C5aR1 Plays an Evolutionarily Conserved Role in Successful Cardiac Regeneration. Natarajan N, Abbas Y, Bryant DM, Gonzalez-Rosa JM et al. Circulation 2018 Jan 18. | RNA-Seq of the lower halves of heart ventricles at 12, 24 and 48 h post-cryoinjury (12 hpi, 24 hpi, 48 hpi) or sham (12 hps, 24 hps, 48 hps). Three pools of three halves of ventricles were used per condition. |
| Heart regeberation (Yin lab.) | GSE106884 | | mRNA gene expression profiling during 0, 1, 3, 7, 14, 21 and 30 days post ventricular resection. |
| Cardiomyocyte subtypes (Mercader lab.) | GSE87596 | Tbx5a lineage tracing shows cardiomyocyte plasticity during zebrafish heart regeneration. Sanchez-Iranzo H, Galardi-Castilla M, Minguillon C, Sanz-Morejon A et al. Nat Commun 2018 Jan 30;9(1):428. | RNA-Seq of the tbx5a-postive (tbx5a-positive) or -negative (tbx5a-negative) cardiomyocytes (CM) from an uninjured heart. Four replicates per condition obtained from the same heart. |
| Cardiac fibrosis (Mercader lab.) | GSE101200 | Transient fibrosis resolves via fibroblast inactivation in the regenerating zebrafish heart. Sanchez-Iranzo H, Galardi-Castilla M, Sanz-Morejon A, Gonzalez-Rosa JM et al. Proc Natl Acad Sci U S A 2018 Apr 17;115(16):4188-4193. | Three to six biological replicates consisting of different cell types obtained from the ventricular apex. |
| Cardiac fibroblast inactivation (Mercader lab.) | GSE101199 | Transient fibrosis resolves via fibroblast inactivation in the regenerating zebrafish heart. Sanchez-Iranzo H, Galardi-Castilla M, Sanz-Morejon A, Gonzalez-Rosa JM et al. Proc Natl Acad Sci U S A 2018 Apr 17;115(16):4188-4193. | postnb-derived cells were FAC sorted from a pool of three to five biological samples. Four pools were collected at 7 dpi and three at 60 dpi. RNA was extracted from those pools and further processed for transcriptome analysis. |
| Spinal cord regeneration (Poss lab.) | GSE77025 | Injury-induced ctgfa directs glial bridging and spinal cord regeneration in zebrafish. Mokalled MH, Patra C, Dickson AL, Endo T et al. Science 2016 Nov 4;354(6312):630-634. | Gene expression analysis of spinal cord after sham (2 weeks after sham) and after injury (2 weeks after transection). Two replicates per condition. |
| Fin regeneration (Poss lab.) | GSE76564 | Modulation of tissue repair by regeneration enhancer elements. Kang J, Hu J, Karra R, Dickson AL et al. Nature 2016 Apr 14;532(7598):201-6. | Transcriptional profiles of uninjured caudal fin (uninjured) and 4 days after amputation (4 dpa). Two pools of 10 fins per condition. |
| Fin regeneration (Yin lab.) | GSE74415 | A Conserved MicroRNA Regulatory Circuit Is Differentially Controlled during Limb/Appendage Regeneration. King BL, Yin VP. PLoS One 2016;11(6):e0157106. | RNA-Seq of uninjured (uninjured) or 4 days after amputation (4 dpa). Three replicates per condition. |
| Fin proximodistal (Moon lab.) | GSE92760 | Transcriptomic, proteomic, and metabolomic landscape of positional memory in the caudal fin of zebrafish. Rabinowitz JS, Robitaille AM, Wang Y, Ray CA et al. Proc Natl Acad Sci U S A 2017 Jan 31;114(5):E717-E726 | RNA-Seq of three locations along the proximo-distal axis of the caudal fin, proximal (prox), middle (mid) and distal (dist). Five replicates per condition, each replicate is a pool of two males and two females. |
| Lateral line (Piotrowski lab.) | GSE56176 | Gene-expression analysis of hair cell regeneration in the zebrafish lateral line. Jiang L, Romero-Carvajal A, Haug JS, Seidel CW et al. Proc Natl Acad Sci U S A 2014 Apr 8;111(14):E1383-92. | RNA-Seq of the GFP-postive cells from the sqET20 transgenic line, which labels inner and mantle support cells (GFP) and the rest of the cells (control) after neomycin treatment (neo) or non-treated (nt). Neomycin chemically ablates hair cells. |
| Muller glia (Raymond lab.) | GSE86872 | Rapid, Dynamic Activation of Mller Glial Stem Cell Responses in Zebrafish. Sifuentes CJ, Kim JW, Swaroop A, Raymond PA. Invest Ophthalmol Vis Sci 2016 Oct 1;57(13):5148-5160. | RNA-Seq from Muller glia 0, 8 and 16 h post-injury. Three replicates per condition. |
| Microglia (van Ham lab.) | GSE86921 | Identification of a conserved and acute neurodegeneration-specific microglial transcriptome in the zebrafish. Oosterhof N, Holtman IR, Kuil LE, van der Linde HC et al. Glia 2017 Jan;65(1):138-149. | Transcriptome analysis of brain cell (brain), homeostatic microglia (microglia control) (triplicates per condition) and activated microglia after cell ablation by nitroreductase at 24 (active microglia 24 h) and 48 h after treatment (active microglia 48 h). Two replicates per condition. |
| Liver regeneration (Cui lab.) | SRP053395 | FTranscriptomic characterization of the dorsal lobes after hepatectomy of the ventral lobe in zebrafish. eng G, Long Y, Peng J, Li Q, Cui Z. BMC Genomics 2015. Nov 19;16:979 | RNA-Seq transcriptomic analysis of 12 female livers after 6 (6 h) and 12 (12 h) h after partial hepatectomy, and from 24 h post-sham surgery (sham). One replicate per condition. |
| Skeletal muscle | GSE92489 | Temporally distinct transcriptional regulation of myocyte dedifferentiation and Myofiber growth during muscle regeneration. BMC Genomics 2017 Nov 9;18(1):854.Loule KW, Saera-Vila A, Kish PE, Colacino JA et al. | RNA-seq of zebrafish lateral eye muscle 0 (0 hpi), 9 (0 hpi) and 18 (18 hpi) hours post injury. Four replicates per condition. |
| Developmental stages (PRJEB1986) | PRJEB1986 | | RNA-Seq from embryos after 1, 2, 3, 5 and 14 days post-fertilisation. 1dpf (2 replicates), 2dpf (2 replicates), 3dpf (3 replicates), 5dpf (3 replictates) and 14dpf (1 replicate). |
| Developmental stages (Schier lab.) | GSE32898 | Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Pauli A, Valen E, Lin MF, Garber M et al. Genome Res 2012 Mar;22(3):577-91. Chew GL, Pauli A, Rinn JL, Regev A et al. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Development 2013 Jul;140(13):2828-34. | RNA-Seq of embryos at 2-4 cells (2-4 cells), one thousand cells (1K cell), dome (dome) and bud (bud) stages, and after 28 h post-fertilisation (28 hpf) and 2 (2 dpf) and 5 days post-fertilisation. Two replicates per condition. |
| Developmental stages (Yanai lab.) | GSE60619 | The mid-developmental transition and the evolution of animal body plans. Levin M, Anavy L, Cole AG, Winter E et al. Nature 2016 Mar 31;531(7596):637-641. | RNA-Seq from one single embryo every 40 minutes from fertilisation. No replicates. |
| Spinal cord regeneration P.marinus (Bloom lab.) | GSE60619 | Highly conserved molecular pathways, including Wnt signaling, promote functional recovery from spinal cord injury in lampreys. Herman PE, Papatheodorou A, Bryant SA, Waterbury CKM et al. Sci Rep 2018 Jan 15;8(1):742. | RNA-Seq from spinal cord (SC) and brain (BR) uninjured and 6 hours (6 hpi), 1 and 3 days (1 dpi, 3 dpi), 1, 2, 3, 4, 5, 6, and 12 weeks (1 wpi, 2 wpi, 3 wpi, 4 wpi, 5wpi, 6 wpi, 12 wpi) after spinal cord injur. One replicate per condition. RNA from 4-6 animals pooled. |

**Supplementary Figure 5.** Screenshot of the information and references provided about each dataset. Links to the publication that describes the datasets and the raw data repository are provided. To facilitate the rapid understanding of the data, a short summary of the experiment is provided.




**Supplementary Figure 6.** Expression of the *ezra* and *ezrb* genes using the "Plot fpkm" application. Expression of *ezra* and *ezrb* are shown in different development stages datasets from Schier lab. (A), and Yanain lab. (B).

A

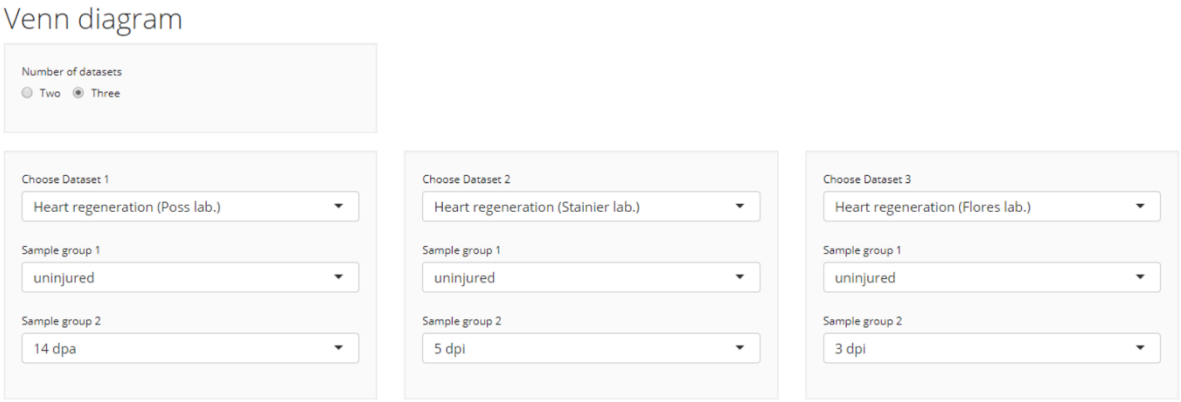



B

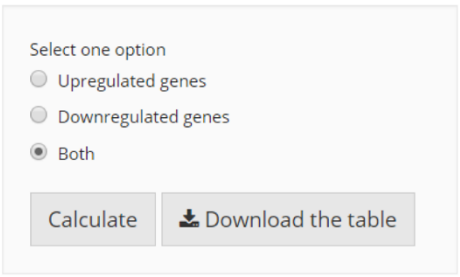
| Ensembl gene ID | Gene symbol | Correlations |
|---|---|---|
| ENSDARG00000103747 | cav1 | 1.0000000 |
| ENSDARG00000020924 | myo1ca | 0.7861321 |
| ENSDARG00000071196 | sdprb | 0.7658510 |
| ENSDARG00000100968 | si:ch211-1a19.3 | 0.7616283 |
| ENSDARG00000040362 | ehd2b | 0.7473702 |
| ENSDARG00000054451 | loxl1 | 0.7335184 |
| ENSDARG00000061579 | myo1cb | 0.7245319 |
| ENSDARG00000053857 | ccdc187 | 0.7219719 |
| ENSDARG00000041546 | gypc | 0.7050950 |
| ENSDARG00000040133 | ackr4b | 0.7015259 |
| ENSDARG00000105653 | BX323797.4 | 0.7011573 |
| ENSDARG00000061941 | trpv4 | 0.7000284 |
| ENSDARG00000073711 | mmrn2b | 0.6947155 |
| ENSDARG00000019367 | tgfb3 | 0.6922020 |
| ENSDARG00000105065 | dock9a | 0.6896928 |
| ENSDARG00000070391 | tspan4b | 0.6861500 |
| ENSDARG00000004451 | tnfrsfa | 0.6856331 |
| ENSDARG00000103020 | si:dkey-237i9.8 | 0.6841137 |
| ENSDARG00000040920 | si:dkey-49n23.1 | 0.6825790 |
| ENSDARG00000099891 | si:ch211-57n23.1 | 0.6822988 |
| ENSDARG00000029072 | klf6a | 0.6816982 |
| ENSDARG00000039881 | cemip | 0.6810626 |
| ENSDARG00000034718 | tfpia | 0.6800483 |
| ENSDARG00000103774 | limch1b | 0.6762323 |
| ENSDARG00000035858 | cnn2 | 0.6753525 |

**Supplementary Figure 7.** Correlation of the *cav1* expression with all the other genes in the zebrafish genome. All the datasets were used to calculate correlations. (A) Selection panel of the datasets and the gene to be used to calculate the correlations. (B) Genes showing a higher correlation value with the selected gene.

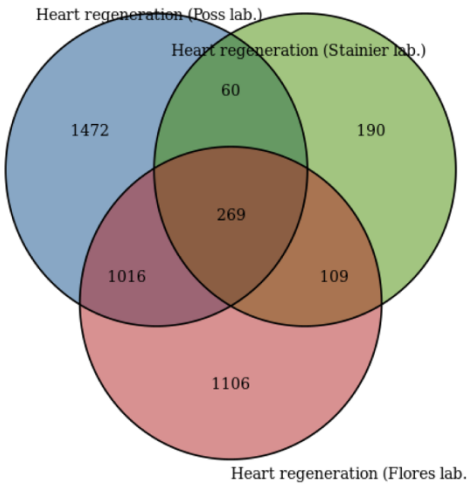
A




B




C




D

Common genes in all comparisons

| Ensembl gene ID | Symbol | Mean FDR |
|---|---|---|
| ENSDARG00000075263 | ankrd1a | 1.517492e-63 |
| ENSDARG00000019815 | fn1a | 6.116973e-42 |
| ENSDARG00000009014 | col11a1b | 2.143921e-41 |
| ENSDARG00000037997 | tubb5 | 2.220153e-39 |
| ENSDARG00000044125 | txn | 5.268458e-37 |
| ENSDARG00000036036 | mdka | 6.205232e-33 |
| ENSDARG00000059049 | zgc:174904 | 2.816338e-31 |
| ENSDARG00000088641 | grn2 | 4.312869e-31 |
| ENSDARG00000010434 | clu | 8.625036e-30 |
| ENSDARG00000089362 | grn1 | 1.214575e-29 |
| ENSDARG00000019601 | col12a1b | 1.971203e-29 |
| ENSDARG00000001452 | adam8a | 2.993890e-29 |
| ENSDARG00000093748 | si:ch211-217k17.11 | 6.578861e-29 |
| ENSDARG00000040178 | havcr1 | 9.265716e-29 |
| ENSDARG00000096979 | NPC2 | 8.633365e-28 |
| ENSDARG00000008803 | marcksb | 5.151995e-26 |
| ENSDARG00000030449 | crabp2b | 1.036045e-25 |
| ENSDARG00000044613 | c1qa | 9.850418e-25 |
| ENSDARG00000075891 | sall1b | 1.698623e-24 |
| ENSDARG00000055439 | adamtsl7 | 1.525654e-23 |
| ENSDARG00000061120 | slc43a2b | 1.630468e-23 |
| ENSDARG00000060917 | anln | 2.901153e-22 |
| ENSDARG00000005789 | enpp1 | 5.205089e-22 |
| ENSDARG00000075261 | timp2b | 9.417912e-22 |
| ENSDARG00000037859 | il11a | 3.982114e-21 |

**Supplementary Figure 8.** Genes differentially expressed in three heart regeneration datasets each of them performed by a different laboratory. (A, B) Screenshot of the selection menu. (C) Venn diagram. (D) List of the genes in common among the three lists. Geometric mean of the FDR values is shown and genes are ordered according to the descending order of this value.