OXFORD

Databases and ontologies

zfRegeneration: a database for gene expression profiling during regeneration

Rocío Nieto-Arellano (iD) *,† **and Héctor Sánchez-Iranzo** (iD) *,‡

Independent Researcher, Boxbergring 107, Heidelberg 69126, Germany

*To whom correspondence should be addressed.
†Present address: Proteome Organisation and Dynamics Group, Institute of Molecular Biology, Mainz 55128, Germany
‡Present address: Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany

Abstract

Motivation: Zebrafish is a model organism with the ability to regenerate many different organs. Although RNA-Seq has been used extensively to study this process, there are no databases that allow easy access to data.

Results: Here we develop the first regeneration database that provides easy access to a large number of RNA-Seq datasets through custom-made plots of expression levels, differential expression analyses, correlations of genes and comparisons of the different datasets. zfRegeneration has a user-friendly web interface designed to enhance regeneration studies and to overcome the barriers between different research groups that study the regeneration of distinct organs. Using several case studies, we demonstrate that zfRegeneration provides a unique platform to analyse and understand gene expression during regeneration.

Availability and implementation: zfRegeneration is freely available at www.zfregeneration.org.

Contact: hsancheziranzo@gmail.com or rocionietoarellano@gmail.com.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Zebrafish is a widely used vertebrate organism in research. Its easy transgenic manipulation, short generation time and ability to regenerate different organs make it one of the preferred model organisms to study regeneration.

RNA-Seq provides a robust way to analyse gene expression quantitatively. Several experiments have been performed in the context of regenerating different organs in zebrafish. However, the major data portals, including Gene Expression Omnibus (GEO) and Sequence Read Archive, provide this information as raw data archives that are not easily accessible. While there are some useful databases that include zebrafish gene expression or regeneration data (Kapushesky *et al.*, 2012; King *et al.*, 2018; Westerfield *et al.*, 1999), none of them provide extensive access to RNA-Seq zebrafish regeneration data.

This article presents the 'Zebrafish regeneration database' (www.zfregeneration.org), a public database that provides access to regeneration RNA-Seq data. As regeneration has been related to development (Galdos *et al.*, 2017; Mercola *et al.*, 2011; Witman *et al.*, 2011), it also includes embryonic development-related datasets to allow their comparison. This database is an intuitive way to perform open analyses, including differential gene expression analysis and correlations of genes across selected datasets. The database also provides a way to compare different datasets by representing Venn diagrams and highlight genes that are differentially expressed across experiments. Here we present some examples to show how this database can extend our understanding of regeneration.

2 Results

2.1 Database navigation

The zfRegeneration database is presented as a user-friendly web interface that contains a menu sidebar with four applications that perform different analyses. To allow an easy interpretation of the results, in each application we provide a brief description of them. Additional information on how the data is processed can be found in Figure 1A and Supplementary Methods. In the analysis where the output is a table that
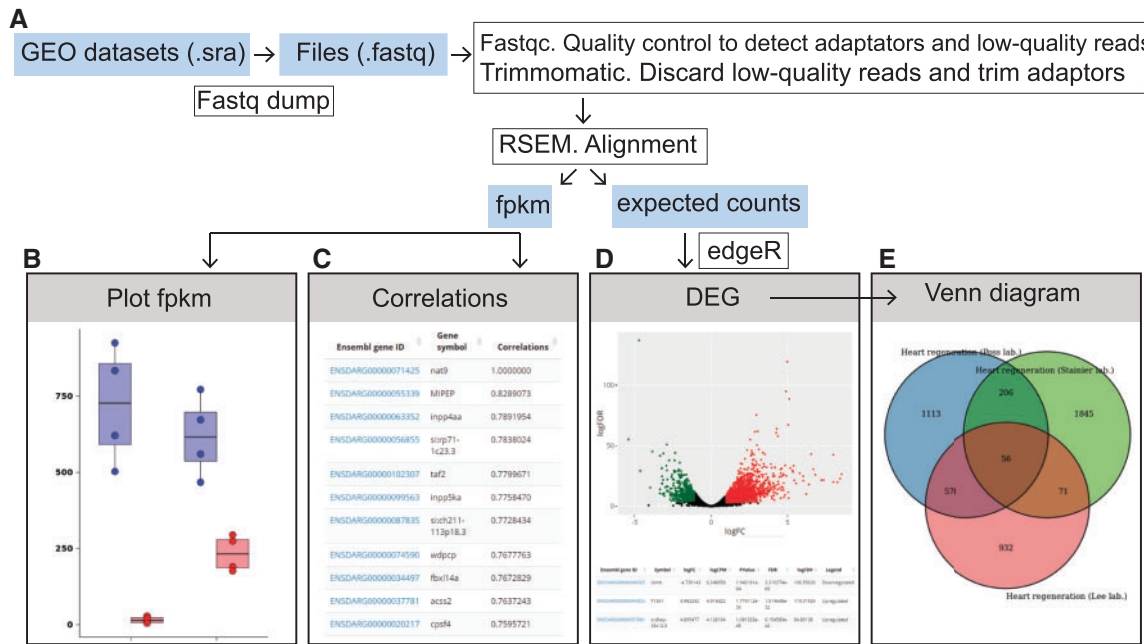
Fig. 1. Development and use of the zfRegeneration database. (**A**) Regeneration RNA-Seq datasets were decompressed by using fastq-dump, quality-controlled with fastqc and low-quality reads were filtered or trimmed with trimmomatic. These files were aligned with RSEM, and the fpkm and expected counts were inputted into the database for further analyses. (**B–E**) Analysis that can be performed: (B) plot of expression data, (C) list of highly correlated genes, (D) differentially expressed genes, (E) Venn diagram

contains genes, we provide an Ensembl link to each gene. Moreover, it is possible to download the outputs for further analysis by users.

2.2 Plot of expression data (fpkm)

This application allows the user to visualize the expression levels of 1–5 selected genes either in one or all zebrafish datasets (Fig. 1B). Genes can be found by either their gene symbol or their ID. Next we offer an example of how this information can be used to better understand previously published results.

Govindan *et al.* have showed by immunohistochemistry that Fibronectin protein is up-regulated during zebrafish fin regeneration (Govindan and Iovine, 2015). As the zebrafish genome underwent a genome duplication event (Taylor *et al.*, 2001; Wittbrodt *et al.*, 1998) the results were not enough to pinpoint which of the two paralogues were responsible for this signal. By plotting the expression of both *fn1a* and *fn1b* in two different datasets, we can now qualitatively detect the expression of the two genes. As expected, it can be observed how both *fn1a* and *fn1b* are up-regulated upon injury. However, we can also detect that *fn1b* increases much more than *fn1a* in two different datasets (Supplementary Fig. S1). This supports their results as well as adds new information that could have otherwise been unappreciated. This knowledge can be very helpful, for example, to decide whether to generate a *fn1a* or a *fn1b* transgenic reporter line.

2.3 Correlations

In the correlation application, the user can calculate which genes better correlate to their selected gene (Fig. 1C). It is very versatile, as it allows the selection of the datasets to be used to calculate the correlation.

As an example, to better understand the fibrotic response to injury, we analysed the correlation of *col1a2* across all the regeneration datasets. We found that most of the highly correlated genes encoded for extracellular matrix (ECM) and ECM processing proteins, including *col1a1b*, *col1a1a*, *col11a1b* and *mmp14a*. This suggests that the

fibrotic response is conserved across organs, and that ECM proteins are produced in a coordinated manner (Supplementary Fig. S2).

2.4 Differential expression

This application allows a comparison of different groups of samples in a dataset selected by the user. It calculates the list of differentially expressed genes and shows it in an interactive and downloadable table. It also plots the results as an interactive volcano plot, where the user can see to which gene each dot corresponds (Fig. 1D), and by using the 'lasso' or box select' tools, users can generate a new table containing only chosen genes.

We provide an example of this application with a heart regeneration dataset (Bednarek *et al.*, 2015). This article focuses on the role that telomerase plays during regeneration. However, here we show that general information about the heart regeneration process can be extracted by comparing uninjured and 3-day post-injury hearts. We found genes related to ECM (*fn1a*, *col11a1b*), genes related to sarcomeric proteins (*ankrd1a*) and signalling molecules (*mdka*) (Supplementary Fig. S3).

2.5 Venn diagrams

To better understand the similarities between regeneration in different organs, we developed an application called Venn diagrams that calculates the differentially expressed genes in different datasets and determines the genes in common between them. The result is displayed as both a Venn diagram and a downloadable table, which includes the list of genes in common in all the selected datasets (Fig. 1E).

An important question in regenerative biology is whether there are general regenerative responses to an injury conserved between different organs (Strähle and Schmidt, 2012). As an example of how this database can be used to understand this, we compared the genes that were up-regulated after the injury of three different organs, heart, fin and spinal cord, using datasets from the same laboratory (Goldman *et al.*, 2017; Kang *et al.*, 2016; Mokalled *et al.*, 2016).

We found 44 common genes (Supplementary Fig. S4) that were up-regulated upon different injury types. Interestingly, most of these genes were ECM-related genes (*TNC*, *serpinh1b*, *ogn*, *loxa*, *col11a1a*), which further support their important role during regeneration. Additionally, we found some genes related to the immune response (*mpeg1.1*) and, more importantly, genes whose function has not been related to regeneration (*slc8a4b*, *fam46b*, *ywhag2*). Our results indicate that they could play an important role for regeneration and that its response is conserved across different organs. Moreover, since different comparisons can be made, there is the possibility of finding new genes related to their process of interest.

2.6 Datasets are linked to the original papers and raw data

We provide a short description of all the datasets included in this database (Supplementary Fig. S5). Furthermore, the reference link is connected to both the database where the raw data is stored and the publication where the RNA-Seq experiment is described. In the contact section, an email is provided to allow researchers to suggest the inclusion of new datasets.

3 Discussion

There has been a great advance in sequencing technology in recent years and its application has extended to many different fields, regeneration among them. Generated datasets are normally deposited in public repositories such as GEO. However, obtaining meaningful information from raw data is time-consuming and requires specialized expertise. Here we develop an easy-to-use regeneration RNA-Seq database. Apart from being a way to access data, it provides researchers with the possibility of running open analyses that allows them to find differentially expressed genes between the selected groups of samples, and also to compare various datasets among them.

In this work, we offer some examples of its applicability (additional examples in Supplementary Material). However, unlimited additional analyses can be performed depending on the user's needs. For example, the application provides helpful information to decide which genes are to be studied and, furthermore, can be a time-saving tool for *in situ* hybridization experiments as it can determine which genes are not expressed.

We expect this database to help find the similarities and differences in the regeneration processes in different organs and promote the collaboration of different research groups. We are confident that this database will be a valuable tool in the research community by aiding in the design of experiments, testing hypotheses and validating results.

Conflict of Interest: none declared.

References

Bednarek,D. *et al.* (2015) Telomerase is essential for zebrafish heart regeneration. *Cell Rep.*, **12**, 1691–1703.

Galdos,F.X. *et al.* (2017) Cardiac regeneration: lessons from development. *Circ. Res.*, **120**, 941–959.

Goldman,J.A. *et al.* (2017) Resolving heart regeneration by replacement histone profiling. *Dev. Cell*, **40**, 392–404.

Govindan,J. and Iovine,M.K. (2015) Dynamic remodeling of the extra cellular matrix during zebrafish fin regeneration. *Gene Expr. Patterns*, **19**, 21–29.

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

Kapushesky,M. *et al.* (2012) Gene Expression Atlas update–a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **40**, D1077–D1081.

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.

Mercola,M. *et al.* (2011) Cardiac muscle regeneration: lessons from development. *Genes Dev.*, **25**, 299–309.

Mokalled,M.H. *et al.* (2016) Injury-induced ctgfa directs glial bridging and spinal cord regeneration in zebrafish. *Science*, **354**, 630–634.

Strähle,U. and Schmidt,R. (2012) A universal program for tissue regeneration? *Dev. Cell*, **23**, 1123–1124.

Taylor,J.S. *et al.* (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. B Biol. Sci.*, **356**, 1661–1679.

Westerfield,M. *et al.* (1999) Chapter 19 zebrafish informatics and the ZFIN database. *Methods Cell Biol.*, **60**, 339–355.

Witman,N. *et al.* (2011) Recapitulation of developmental cardiogenesis governs the morphological and functional regeneration of adult newt hearts following injury. *Dev. Biol.*, **354**, 67–76.

Wittbrodt,J. *et al.* (1998) More genes in fish? *BioEssays*, **20**, 511–515.