

Video Article

# Sample Preparation and Analysis of RNASeq-based Gene Expression Data from Zebrafish

Timothy L. Hosteley<sup>\*1</sup>, Jessica E. Nesmith<sup>\*1</sup>, Norann A. Zaghloul<sup>1</sup>

<sup>1</sup>Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine

<sup>\*</sup>These authors contributed equally

Correspondence to: Norann A. Zaghloul at [zaghloul@umaryland.edu](mailto:zaghloul@umaryland.edu)

URL: <https://www.jove.com/video/56187>

DOI: [doi:10.3791/56187](https://doi.org/10.3791/56187)

Keywords: Genetics, Issue 128, Zebrafish, RNAseq, transcriptome, RNA, *Danio rerio*, pathway analysis, Gene Ontology terms, flow sorting, Fluorescence Activated Cell Sorting

Date Published: 10/27/2017

Citation: Hosteley, T.L., Nesmith, J.E., Zaghloul, N.A. Sample Preparation and Analysis of RNASeq-based Gene Expression Data from Zebrafish. *J. Vis. Exp.* (128), e56187, doi:10.3791/56187 (2017).

## Abstract

The analysis of global gene expression changes is a valuable tool for identifying novel pathways underlying observed phenotypes. The zebrafish is an excellent model for rapid assessment of whole transcriptome from whole animal or individual cell populations due to the ease of isolation of RNA from large numbers of animals. Here a protocol for global gene expression analysis in zebrafish embryos using RNA sequencing (RNASeq) is presented. We describe preparation of RNA from whole embryos or from cell populations obtained using cell sorting in transgenic animals. We also describe an approach for analysis of RNASeq data to identify enriched pathways and Gene Ontology (GO) terms in global gene expression data sets. Finally, we provide a protocol for validation of gene expression changes using quantitative reverse transcriptase PCR (qRT-PCR). These protocols can be used for comparative analysis of control and experimental sets of zebrafish to identify novel gene expression changes, and provide molecular insight into phenotypes of interest.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/56187/>

## Introduction

Comparative analysis of global gene expression is a valuable tool to identify novel genes contributing to observed phenotypes. Such analyses typically rely on quantitative assessment of transcript abundance compared between experimental and control samples. Targeted approaches, such as qRT-PCR are relatively rapid and accurate for investigation of single gene expression changes. RNA sequencing (RNASeq) offers a broad, hypothesis-free approach to identify significant changes in gene expression between samples, making it now the standard for such investigations across experimental systems.

Zebrafish have emerged as a prominent model across many disease areas. Originally developed for their utility in developmental biology studies, due to their high fecundity and relatively low cost of maintenance, experimental use of zebrafish has evolved to include a broad range of phenotypes from embryonic to adult stages as well as a wide array of molecular assays<sup>1,2,3</sup>. Indeed, these advantages make molecular mechanistic studies rapid and cost-effective because of the ease of acquiring large amounts of material combined with the ease of both genetic and environmental manipulation at all stages of life. Moreover, the transparent nature of zebrafish embryos and larvae make it ideal for generating cell- and tissue-specific transgenic reporter lines allowing *in vivo* visualization of discrete cell populations<sup>4</sup>. Exploitation of such lines permits global gene expression analysis in specific isolated cell types based on reporter gene expression.

Here we present a comprehensive protocol for global gene expression analysis using RNASeq after culture of zebrafish embryos. Genetic experimental manipulations, including morpholino (MO)-based transient gene knockdown or CRISPR-mediated genome editing, have been presented elsewhere<sup>5,6,7</sup>. We therefore focus on a detailed protocol for isolation of RNA from whole embryos or sorted transgenic reporter-expressing cells followed by simple computational analysis of RNASeq results using pathway tools and gene ontology (GO) terms. Finally, we have included a strategy for validation of gene expression changes by quantitative reverse transcriptase PCR (qRT-PCR). These protocols are applicable to zebrafish embryos subjected to a wide range of experimental conditions, including comparison of genetic mutants or environmental conditions.

## Protocol

All animal protocols outlined below are in accordance with and approved by the University of Maryland Institutional Animal Care and Use Committee (IACUC).

## 1. Embryo Preparation

### 1. Generating embryos through natural mating

1. Culture embryos to 3 months of age, reproductive maturity<sup>5,8</sup>.
2. Segregate adult male and female fish from the desired strain into divided mating tanks on the evening before embryo collection, and add 2 males and 3 females to each tank.  
NOTE: Use of the transgenic *insulin2a:mCherry* fluorescent reporter strain allowed the analysis of pancreatic  $\beta$ -cells.
3. Transfer fish to the mating tank with fresh system water and remove the divider immediately after the lights come on the following morning.
4. Allow the fish to mate naturally until embryos are observed in bottom tank. Collect embryos in 30 min intervals until the desired amount is collected. Store each collected time point in separate Petri dishes in embryo media at 28.5 °C.
5. Perform microinjection of genetic material or placement in experimental culture media<sup>6</sup>, if desired, and culture the embryos in fresh Hank's embryo medium<sup>8</sup> in 10-cm Petri dishes at 28.5 °C.  
NOTE: For gene expression analysis in morpholino (MO) injected or mutant animals, note that each manipulation may have incidental impacts on gene expression. Mutants can exhibit genetic compensation at the level of transcription that is not observed by MO-based gene targeting<sup>9</sup>.

### 2. Stage embryos

1. Culture the embryos in groups of **50–75 embryos** per 10-cm Petri dish to promote consistent developmental timing of all embryos.
2. Monitor the developmental morphology using dissecting microscope at blastomere, epiboly, and somite stages to ensure developmental progression<sup>10</sup>.  
NOTE: Remove any dying or malformed embryos to prevent developmental delay in the dish.
3. Separate the embryos based on the developmental age. Measure the embryo age using somite number after segmentation (post-gastrulation, 10.33 h post fertilization (hpf)) until approximately 24 hpf. Stage the embryos and larvae using total body length after 24 hpf.  
NOTE: The somites are the chevron-shaped mesodermal tissue present on the dorsal part of the embryo.
4. Place the sorted embryos into a 28.5 °C incubator and allow development to progress to the desired age.

## 2. Single-cell Dissociation: Whole Embryo and Sorted Cell Populations

### 1. Whole embryo dissociation

1. Euthanize the zebrafish embryos at the desired stage by placing the Petri dish on ice for 5-15 min until no movement is observed.
2. Transfer a pool of 20 embryos to a labeled 1.5 mL microcentrifuge tube. Remove any excess embryo media from the microcentrifuge tube.
3. Add 200  $\mu$ L lysis reagent (see **Table of Materials**) to the tube containing embryos. Homogenize the embryos in lysis reagent mechanically using a pestle. Add 800  $\mu$ L lysis reagent to bring the total volume to 1 mL. Store at -80 °C until RNA extraction.

### 2. Flow-sorted cell isolation<sup>11</sup>

1. Transfer the embryos from the Petri dish into a 1.5 mL microcentrifuge tube and remove as much embryo medium as possible.  
NOTE: Depending on the embryo age, mechanical dissociation may also be required at this step. With embryos > 48 hpf, we recommend the use of a pestle to disrupt embryo integrity before proceeding.
2. Incubate the embryos with 1 mL of dissociation buffer 1 (see **Table of Materials**) in 1.5 mL microcentrifuge tube. Incubate for 15-30 min at room temperature. Triturate the solution by gently pipetting up and down using a P1000 tip to mix every 3-4 min during the incubation step. DO NOT VORTEX.
3. Collect the embryos by centrifugation for 3 min at 300 x g and remove the supernatant. Re-suspend the embryos in 1 mL dissociation buffer 2 (see **Table of Materials**). Incubate for 15-30 min at room temperature with regular pipet trituration.
4. Assess the digestion by diluting 1-2  $\mu$ L of suspension in fluorescence-activated cell sorting (FACS) buffer under microscope.  
NOTE: Complete digestion has occurred when the examined aliquot shows single cells with minimal cell clusters and pieces of embryonic tissue.
5. Collect the cells by centrifugation at 300 x g for 5 min. Remove supernatant. Re-suspend the cells into 1 mL FACS buffer, and gravity filter through a 40  $\mu$ m cell strainer to remove undigested tissue from the sample.
6. Count the cells using a hemocytometer and dilute with FACS buffer to approximately  $1 \times 10^6$  cells/mL in a FACS tube.  
NOTE: For cell types with a relatively small number per embryo (less than 5% of total), such as pancreatic  $\beta$ -cells, use a minimum of 1,000 stage-matched embryos.
7. Provide the FACS sorting facility with the FACS tubes containing 1 mL of cell suspension samples as well as FACS tube containing only lysis reagent or FACS Buffer. Gate the FACS flow-sorter to collect only single cells that express the fluorescent reporter.
8. Perform the FACS flow-sorting until the desired cell numbers have been collected.  
NOTE: To perform RNASeq, there is a minimum total RNA required. We have found that 3,000-5,000 cells are sufficient to isolate high-quality, high-concentration RNA.
9. Store the collected cells on ice and proceed immediately to RNA preparation.

## 3. RNA Preparation

### 1. Extract RNA

1. Incubate the collected sample (from step 2) with lysis reagent for 5 min at room temperature.

2. Add 0.2 mL of chloroform for each 1.0 mL of lysis reagent used and invert the tubes by hand for 15 s. Incubate for 2-3 min at room temperature. Centrifuge the samples at 12,000 x g for 15 min at 4 °C.
3. Transfer the separated, aqueous phase to a fresh tube and add 0.5 mL of isopropanol for each 1.0 mL of lysis reagent used. Incubate for 10 min at room temperature. Centrifuge the samples at 12,000 x g for 10 min at 4 °C.
4. Wash with 75% ethanol and centrifuge the samples at 7,500 x g for 5 min at 4 °C. Remove the supernatant completely and allow to air dry at room temperature. Re-suspend the RNA in 15-30 µL diethyl pyrocarbonate (DEPC)-treated water.

## 2. Purify high-quality RNA

1. Combine the RNA suspension with 1/10 volume 3M sodium acetate (pH 5.5) and 1 volume of isopropanol. Incubate for 20 min at room temperature and centrifuge the samples at 12,500 x g for 10 mins at 4 °C.
2. Wash the pellet with ice-cold 70% ethanol in DEPC-treated water and centrifuge the samples at 10,000 x g for 5 min at 4 °C. Repeat the wash step once.
3. Remove the supernatant and allow to dry at room temperature. Re-suspend in DEPC-treated water.
4. Assay the extracted RNA for the concentration (ng/µL) and purity (260/230 ratio) using an absorption spectrometer. Ensure that the 260/230 value is ~ 2.0 before proceeding with RNASeq. Store at -80 °C until use (up to 6 months).
5. Send the RNA samples to a vendor or core for RNASeq and gene expression change analysis based on quantification of sequencing reads. The results returned from the vendor are typically provided as a list of genes exhibiting a significant fold-change in transcript reads between samples.

NOTE: A column can be used if the above method yields poor quality RNA. The amount, concentration, and RNA Integrity Number (RIN) for RNA samples should be confirmed with the vendor or core. The vendor or core will also typically assess RIN.

## 4. Pathway and GO Term Analysis

NOTE: See **Figure 1** for representative output of the gene expression analysis after RNASeq provided by the core or vendor.

### 1. Sorting differentially expressed genes

#### 1. Comparing a single experimental condition versus control

1. Open data (results) in a spreadsheet management software (see **Table of Materials**).
2. Select the drop-down arrow next to the 'Sort' button and select "Custom Sort" within the spreadsheet containing the differentially expressed genes in the experimental versus control condition.
3. Select the box under 'Column' in the window that pops up and choose to sort by the 'LFC' column. Ensure 'Values' is selected in the 'Sort On' column. Select to sort from 'Largest to Smallest' in the 'Order' column and click "OK".

NOTE: This will sort the differentially expressed genes so that those with increased expression (positive LFC) will appear at the top of the list while those with decreased expression (negative LFC) will appear at the bottom of the list.

#### 2. Compare the multiple experimental conditions to a single control as follows.

1. Select the first cell in an empty column on the Experimental 1 versus Control spreadsheet to determine differentially expressed genes found in two experimental conditions. Type the following equation into the cell:  
`= IF(ISERROR(MATCH(A1,Experimental2_vs_Control!A:A,0)), "", "Duplicate")`  
 NOTE: This equation is a combination of the IF, ISERROR, and MATCH functions. (i) The MATCH function, MATCH(A1,Experimental2\_vs\_Control!A:A,0), will look for the value in A1 (the ENSEMBL ID) in the A column (A:A) of the spreadsheet named Experimental2\_vs\_Control (Experimental2\_vs\_Control!). The "0" indicates to look for an exact match, and if an exact match is found, the function will return a value of "1", while if no match is found, the function will return an error "N/A". (ii) The result of the MATCH function is then input into the ISERROR function, ISERROR(MATCH(A1,Experimental2\_vs\_Control!A:A,0)), where, if the input is "1" indicating a match was found, it will return "FALSE", and if the input is "N/A" indicating a match was not found, it will return "TRUE". (iii) The "TRUE" or "FALSE" result is then input into the IF function, IF(ISERROR(MATCH(A1,Experimental2\_vs\_Control!A:A,0)), "", "Duplicate"), where, if the input is "TRUE" indicating a match was not found, it will return the value in between the first set of quotation marks (""), which is empty, and therefore the cell will be empty. If the input is "FALSE" indicating a match was found, it will return the value in between the second set of quotation marks ("Duplicate"), and the cell will say "Duplicate".
2. Press "Enter" or "Return" to run the equation. Select the cell containing the equation and click the box in the bottom right corner of the cell. Keep the mouse clicked and drag the selected area all the way down the column until the last 'Feature ID', to copy the equation into each cell in the column.
3. Select "Custom Sort" again, add a second level of sorting by clicking the "+" icon in the lower left corner. In the first level, "Sort by", under Column select "Duplicate", under Sort On select 'Values', and under Order select 'Z to A'. In the second level, "Then by", under Column select 'LFC', under Sort On select 'Values', and under Order select 'Largest to Smallest' and select "OK".  
 NOTE: This will sort the list of genes into 4 groups based upon directionality of expression change. It is important to check the genes found in both experimental conditions to ensure that the change in the expression is in the same direction in both or in opposite directions.
4. Remove any parentheses from the gene symbols column before proceeding or the results will not be found in the databases. Select the entire column containing gene symbols. Select "Edit" then "Replace" in the drop-down 'File' menu. Type "(\*)" in the "Find what:" bar of the Replace window, and leave the "Replace with:" bar empty. Select 'Replace All' to remove all instances of parentheses.  
 NOTE: The \* represents any number of characters, by placing this character between two parentheses the program is prompted to find any instance in the highlighted cells and any instance of parentheses will be replaced with nothing, thereby removing them.

### 2. Determining enriched pathways

1. Copy the gene symbols of the set for which one wishes to determine the enriched pathways to the clipboard. Go to ConsensusPathDB<sup>12</sup>. Select "Gene set analysis" on the left sidebar of the webpage, followed by "Over-representation analysis".
2. Paste the list of genes in the "Paste a list of gene/protein identifiers" box. Select gene symbol in the "Gene/protein identifier type" box and click "Proceed". Select the box next to "Pathways as defined by pathway databases" under the Pathway-based sets section. NOTE: A number of options for the analysis will appear including the list of available databases to search. We recommend de-selecting all databases except the Kegg and Reactome databases as they are the most established and comprehensive. The minimum overlap with input list and p-value cutoff settings can also be adjusted based on needs. We recommend leaving these settings at the default 2 gene minimum overlap and a 0.01 p-value cutoff.
3. Select "Find enriched sets" to obtain the list of pathways containing genes in the input list. NOTE: The output will be in table format including the name of each pathway followed by "Set size" (indicating the number of total genes in the pathway), "Candidates contained" (indicating the number of genes in the input list that are in that pathway), as well as the p-value and q-value (FDR), and the database from which the pathway was identified.
3. **Generation of pathway networks**
  1. Determine enriched pathways as described above.
  2. Select all of the enriched pathways to visualize in a pathway network by either selecting each box next to the pathway names to be visualized or by clicking "All" under "Select" in the column header above the selection boxes, then select "Visualize selected sets". NOTE: We recommend visualizing all pathways if there are less than 30; if there are greater than 30 pathways we recommend selecting the top 30 most enriched pathways (those containing the greatest number of genes from the input list).
  3. Adjust the "relative overlap" and "shared candidates" filters, by selecting either box in the top center of the page and inputting desired percent relative overlap or number of shared candidates, to be more stringent in order to make the more pertinent overlaps within the pathway networks clearer, and select "apply". NOTE: We recommend a minimum relative overlap of 0.2, indicating a 20% gene overlap between 2 pathways to connect them, and a minimum of 2 shared candidates. The graph legend can be seen by clicking on the graph legend in the upper left of the page.
4. **Determination of enriched GO terms**
  1. Copy the gene symbols of the group for which one wishes to determine the enriched gene ontologies to the clipboard. Go to the 'GO Enrichment Analysis tool' in the Gene Ontology Consortium<sup>13</sup>. Paste the list of gene symbols in the box on the left side of the page under "Your gene IDs here..."
  2. Select which set of GO terms to use, listed underneath the gene IDs box: biological process, molecular function, or cellular component. Choose (recommended) 'biological process'. Select '*Danio rerio*' underneath the GO terms box and click "Submit". NOTE: We recommend using the default p-value cutoff of 0.05.

## 5. Verification by qRT-PCR

NOTE: Individual genes identified with significant gene expression changes in RNASeq should be verified by targeted qRT-PCR in replicate experiments.

1. **cDNA synthesis**
  1. Extract the RNA from embryos using the method described in Section 3.1. RNA Extraction.
  2. Convert 1 µg of RNA to cDNA using cDNA conversion kit (see **Table of Materials**). Combine RNA, dNTPs, and reverse transcriptase enzyme. Perform the thermocycler reaction according to the manufacturer's specifications.
  3. Dilute cDNA 1:3 in DEPC-treated water. Store at 4 °C for up to 1-2 months.
2. **qRT-PCR verification**
  1. Select genes to verify by qRT-PCR from RNA sequencing results. Identify genes with high fold changes in expression. Determine which of these genes also contain high read counts, as this indicates abundant expression.
  2. Select 10-12 targets from the high fold change, high read count list of genes. Design primers to amplify the target genes that span at least 1 intron-exon boundary.
  3. Enter the gene or targeted region of the gene into qPCR primer design site<sup>14</sup>. Select the "design 2 qPCR primers" and prompt the site to create primer set. NOTE: Be sure to include internal control primers, such as *β-actin*, to normalize comparisons between samples. The *β-actin* primers are F: 5'-TCGAGCTGTCTTCCCATCCA-3' and R: 5'-TCACCAACGTAGCTGTCTTTCTG-3'.
  4. Combine the cDNA, forward primer, reverse primer, and polymerase. Perform qRT-PCR amplification to determine the relative expression of genes<sup>15</sup>.
  5. Analyze the relative mRNA expression of genes of interest by relative quantification determination<sup>15</sup>.
  6. Compare the qRT-PCR to RNASeq results to ensure that the directionality of differential expression is consistent.

## Representative Results

### Sorting of Differentially Expressed Genes:

To identify differentially expressed genes in the larval stage of zebrafish models of Alström Syndrome and Bardet-Biedl Syndrome (BBS), we targeted either *alms1* or *bbs1* transcripts by injecting previously validated splice-blocking MOs into wild-type zebrafish embryos<sup>16,17</sup>. At 5 days post fertilization (dpf), two replicates of RNA were extracted from each condition, as well as embryos injected with control MO, as described in Section 3 above. RNA from each sample was sequenced to a depth of ~ 120 million reads with ~ 107 million reads aligned to the zebrafish genome. Between 85-90% of the aligned reads mapped to exonic regions of the genome.

1,348 total genes were significantly changed in the Alström model and 471 total genes were significantly changed in the BBS model. Shared changes in gene expression between the models or changes unique to one of the models were identified as described in Section 4.2. 1,084 total genes were uniquely changed in the Alström model, 612 up-regulated and 472 down-regulated, and 207 genes were uniquely changed in the BBS model, 176 up-regulated and 31 down-regulated (**Figure 2, Table S1**, and **Table S2**). 264 genes were found to be significantly changed in both models, 73 were up-regulated and 182 were down-regulated, and 9 showed opposite changes in expression between the two models (**Figure 2, Table S1**, and **Table S2** (See **Supplemental Tables.xlsx**)). The discrepant numbers of differentially expressed genes between the two models suggest that *bbs1* and *alms1* may impact different stages in development.

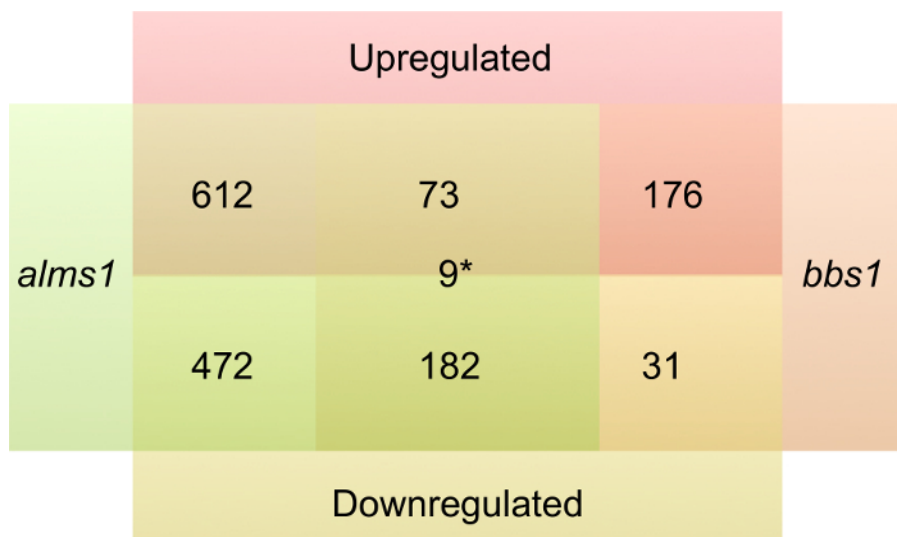
Interestingly, the large number of differentially expressed genes in the Alström model suggest that *alms1* plays an important role in development at the 5 dpf stage, while the smaller number of gene changes in the BBS model suggest that the role of *bbs1* may not be as prominent at this time point. In contrast, previous transcriptomic analyses at 2 dpf suggested the reverse<sup>18</sup>.

### Determination of Enriched Pathways and Gene Ontologies in Zebrafish Model of Alström Syndrome:

To more clearly elucidate the molecular profile of the Alström model, the pathways and gene ontologies enriched in the differentially expressed genes were identified. Enriched pathways were determined by querying the ConsensusPathDB and enriched Gene Ontologies were determined by querying the Gene Ontology Consortium as described in Sections 4.3-4.5. The most highly enriched among the genes up-regulated in the Alström model, by number of genes or fold enrichment, were analyzed. 31 total pathways were up-regulated in the Alström model. Aside from the broad grouping of "metabolism," the most highly affected pathways were the innate and adaptive immune system with 32 and 20 genes, respectively (**Figure 3A**). Downstream  $\beta$ -cell receptor signaling events were also enriched suggesting immune system up-regulation in this model. Several signaling pathways were also enriched among the up-regulated genes in the Alström model, consistent with the association of Alström Syndrome with primary cilium dysfunction, a major signaling center of the cell (**Figure 3A**). In addition, three pathways associated with insulin secretion were up-regulated: fatty acids bound to GPR40, free fatty acids, and acetylcholine (**Figure 3A**). This is consistent with the highly penetrant insulin resistance and type 2 diabetes phenotypes in Alström patients. Six GO terms were enriched among the up-regulated genes in the Alström model, including erythrocyte differentiation, erythrocyte homeostasis, myeloid cell homeostasis, and homeostatic processes (**Figure 3B**). To assess the interconnectedness of the up-regulated pathways, a pathway network of the Alström model up-regulated pathways was generated (**Figure 3C**). The major node of interconnected pathways revealed a high degree of connectivity among the pathways associated with the immune system and signaling, while one of the minor nodes revealed connectivity between the three insulin regulatory pathways. Finally, we verified the gene expression changes identified by RNASeq by assessment of 5 gene targets, which were either up- or down-regulated in the Alström model. The directionality of gene expression changes by qRT-PCR in *aste*, *gck*, *bmb*, *btr26*, and *ifi35*, relative to control, recapitulated that observed by RNASeq (**Figure 4A-B**).

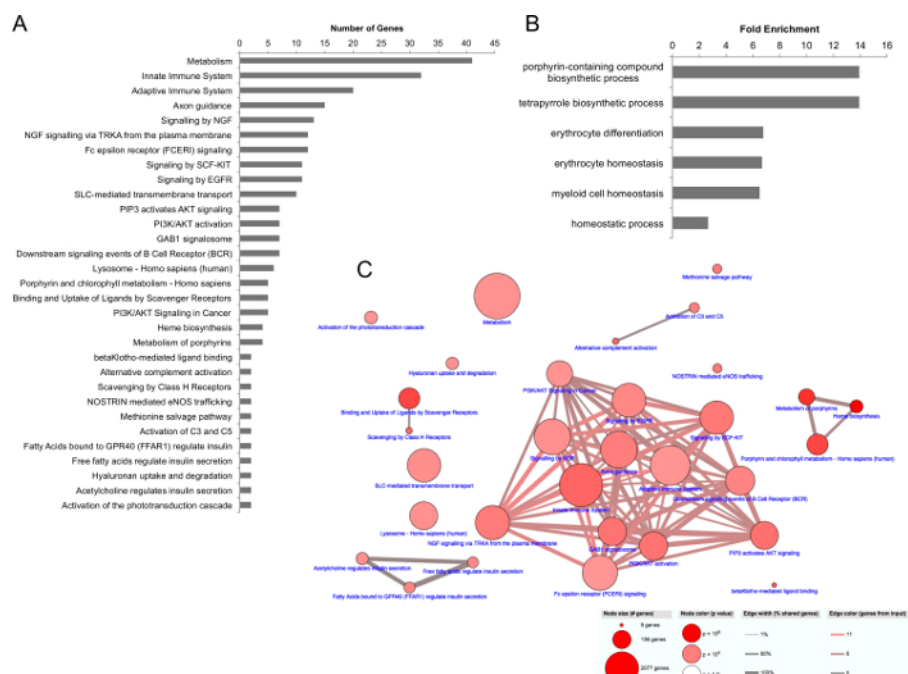
[illegible]

**Figure 1. Representative output of gene expression analysis after RNASeq provided by the core or vendor.** Feature ID indicates the ENSEMBL gene ID number. Read counts indicates the number of reads in either control or experimental samples that aligned to that gene in the genome. Fold change and the log of the fold change indicate the degree of change in expression of the gene between the experimental and control sample in either the positive (increase) in expression in the experimental sample direction, or negative (decrease) in expression in the experimental sample direction. The p-value and FDR are statistical tests to determine the significance of results; for RNASeq experiments, a more stringent FDR value of 0.05 is used to determine significance. Gene\_symbol indicates the NCBI gene ID, and gene\_name lists the full name of the gene. [Please click here to view a larger version of this figure.](#)

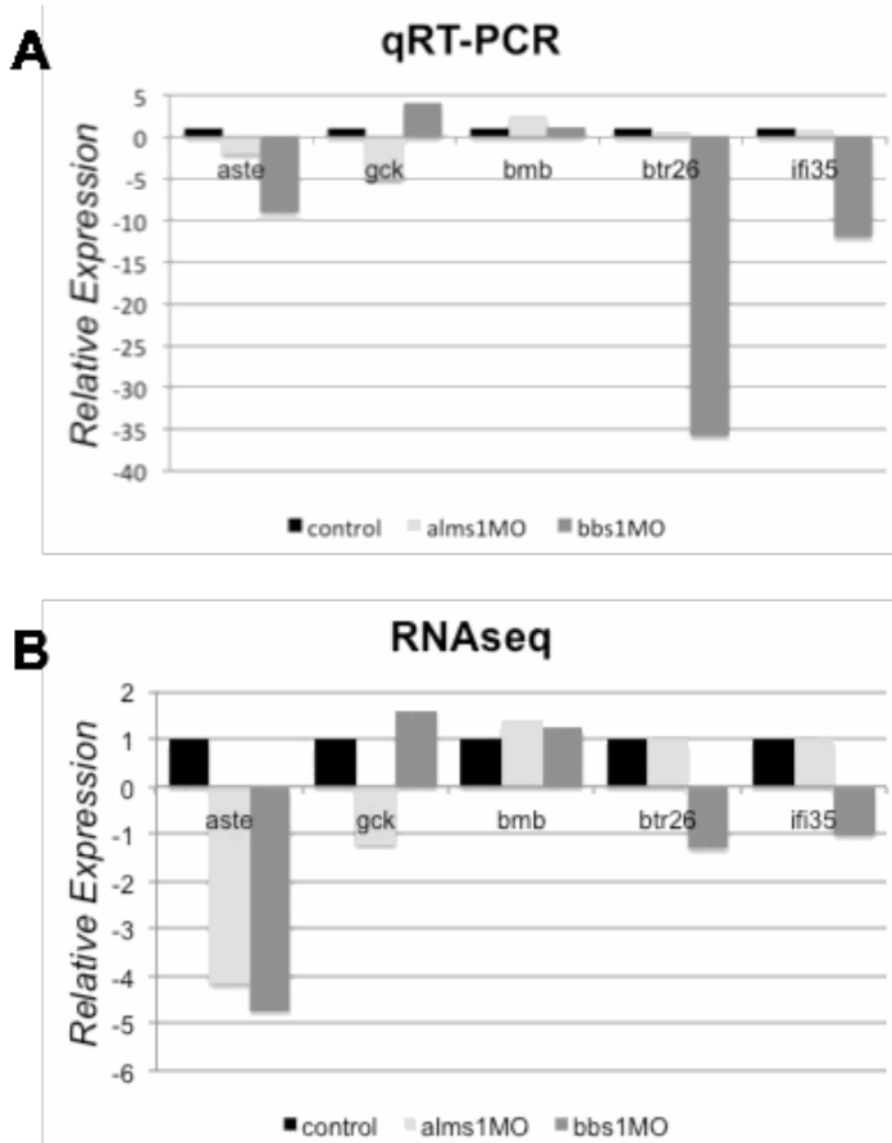


**Figure 2. Differentially expressed genes in the BBS and Alström models.** Number of genes up-regulated (red) or down-regulated (yellow) in alms1 MO (green) injected embryos, bbs1 MO (orange) injected embryos, or both compared to standard control MO injected embryos. \*Denotes number of genes changed in both but having opposite changes in expression. [Please click here to view a larger version of this figure.](#)





**Figure 3. Upregulated pathways and enriched GO terms in the Alström model. (A)** Upregulated pathways by number of genes. **(B)** Upregulated GO terms by fold enrichment. **(C)** Pathway connectivity of upregulated pathways in the Alström model. Pathway connections determined by a minimum of 20% shared genes between pathways and at least 2 genes overlap. [Please click here to view a larger version of this figure.](#)



**Figure 4. qRT-PCR validation of gene expression changes.** Changes in gene expression identified by RNAseq were confirmed in age- and treatment-matched 5 dpf embryos, injected with control, alms1, or bbs1 MO. **(A)** Subset of genes showing mRNA expression of genes using qRT-PCR. **(B)** Equivalent data showing read counts from the RNAseq dataset. All data are relative to control, and qRT-PCR were normalized to actin. aste, asteroid homolog 1; gck, glucokinase; bmb, brambleberry; btr26, bloodthirsty-related gene family, member 26; ifi35, interferon-induced protein 35. [Please click here to view a larger version of this figure.](#)

## Discussion

The approach described in this protocol offers a relatively rapid and cost-effective strategy for transcriptome-level analysis of whole animals or specific sorted cell populations. The zebrafish provides an advantageous model for this type of study because of the ease and rapidity in generating large amounts of starting material, the ease of implementing genetic or environmental experimental conditions, and the availability of a large spectrum of transgenic reporter lines allowing for isolation of cell-type specific and tissue-specific populations.

While not detailed here, this method could easily accommodate tissue sections collected from juvenile or adult fish as an alternative to FACs and collection. We also developed a basic follow-up analysis strategy that only relies on the use of basic spreadsheet commands and pre-existing pathway and GO databases. The major advantage of this approach is the ease of accessibility without requiring high-level expertise in computer programming or database management. As such, this strategy is applicable for investigators of all backgrounds for informative analysis of large gene expression data sets.

Our approach combines basic zebrafish embryo and larval culture with standard RNA extraction techniques followed by RNAseq. RNAseq requires a large quantity of high-quality starting material; some vendors require 2 µg of total RNA. As a result, rare/infrequent cell types or individual embryo analysis is out of reach. However, the last five years have demonstrated dramatic improvements in low-yield analyses, datasets generated from smaller samples, and lower RNA quantities, and we expect that these applications will be possible in the near future. To ensure accurate results, the addition of both technical replicates, *i.e.*, generating two RNA libraries from a single sample, and biological

replicates, *i.e.*, collecting embryos from multiple matings, is ideal. These steps allow the researchers to control for variation in samples, and sequences identified in all samples are more likely to be true results, with the reduced likelihood of missing a lower read-count result.

## Disclosures

The authors have nothing to disclose.

## Acknowledgements

This work was supported by R01DK102001 (N.A.Z.), P30DK072488 (N.A.Z.), and T32DK098107 (T.L.H. and J.E.N.).

## References

1. Nusslein-Volhard, C., Dahm, R. *Zebrafish*. Oxford University Press. (2002).
2. Detrich, H. W., Zon, L., Westerfield, M. *The Zebrafish: Disease Models and Chemical Screens: Edition 4*. Academic Press. (2017).
3. Detrich, H. W., Zon, L. I., & Westerfield, M. *The Zebrafish: Genetics, Genomics, and Transcriptomics, 4th Edition*. Academic Press. (2016).
4. Detrich, H. W. *The Zebrafish: Genetics, Genomics and Informatics, Edition 3*. Academic Press. (2011).
5. Avdesh, A. *et al.* Regular care and maintenance of a zebrafish (*Danio rerio*) laboratory: an introduction. *J Vis Exp.* (69), e4196 (2012).
6. Rosen, J. N., Sweeney, M. F., & Mably, J. D. Microinjection of zebrafish embryos to analyze gene function. *J Vis Exp.* (25) (2009).
7. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol.* **31** (3), 227-229 (2013).
8. Westerfield, M. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th ed. Univ. of Oregon Press, Eugene. (2000).
9. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature.* **524** (7564), 230-233 (2015).
10. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev Dyn.* **203** (3), 253-310 (1995).
11. Samsa, L. A., Fleming, N., Magness, S., Qian, L., & Liu, J. Isolation and Characterization of Single Cells from Zebrafish Embryos. *J Vis Exp.* (109) (2016).
12. ConsensusPathDB. *ConsensusPathDB*. <<http://cpdb.molgen.mpg.de/>> (2017).
13. Consortium, G. O. *Enrichment analysis Tool*. <<http://www.geneontology.org/>> (2017).
14. IDT. *IDT Primerquest Tool*. <<https://www.idtdna.com/Primerquest/Home/Index>> (2017).
15. Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6** (10), 986-994 (1996).
16. Leitch, C. C., Lodh, S., Prieto-Echague, V., Badano, J. L., & Zaghloul, N. A. Basal body proteins regulate Notch signaling through endosomal trafficking. *J Cell Sci.* **127** (Pt 11), 2407-2419 (2014).
17. Lodh, S., Hostalley, T. L., Leitch, C. C., O'Hare, E. A., & Zaghloul, N. A. Differential effects on beta-cell mass by disruption of Bardet-Biedl syndrome or Alstrom syndrome genes. *Hum Mol Genet.* **25** (1), 57-68 (2016).
18. Hostalley, T. L., Lodh, S., & Zaghloul, N. A. Whole organism transcriptome analysis of zebrafish models of Bardet-Biedl Syndrome and Alstrom Syndrome provides mechanistic insight into shared and divergent phenotypes. *BMC Genomics.* **17** 318 (2016).