Data in Brief 65 (2026) 112459

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Single-cell RNA-seq data of wild type and fli1b mutant zebrafish embryos

Luiza N. Loges[1], Ricardo DeMoya[1], Valentina Laverde, Saulius Sumanas*

*Department of Pathology and Cell Biology, USF Heart Institute, 560 Channelside Drive Tampa, FL 33602, USA*

ARTICLE INFO

*Article history:*
Received 24 September 2025
Revised 9 December 2025
Accepted 5 January 2026
Available online 10 January 2026

Dataset link: Single-cell transcriptomic analysis of zebrafish fli1b mutants (Original data)

*Keywords:*
Myeloid
Vascular endothelial
Red blood cell
Transcriptome

ABSTRACT

Fli1b is an ETS transcription factor, which has been previously implicated in zebrafish vascular and hematopoietic development. Here we present single cell RNA sequencing data from wild-type and maternal zygotic *fli1b* mutant zebrafish embryos at 24 h post fertilization. Single-cell suspensions were obtained from approximately 40 whole maternal-zygotic (MZ) *fli1b* mutant and sibling parent wild-type embryos and subjected to RNA sequencing using the 10X Genomics Chromium platform. Following bioinformatic analysis, 34 distinct cell clusters were identified in the integrated wild-type and *fli1b* mutant dataset. The clusters were subsequently annotated based on expression of marker genes. These data will be valuable for further studies of the molecular mechanisms involved in vascular and hematopoietic development. In addition, the obtained transcriptomes of multiple cell types will be useful to investigate other developmental mechanisms in zebrafish and other models.



* Corresponding author.
*E-mail address:* ssumanas@usf.edu (S. Sumanas).
*Social media:* @ssumanas.bsky.social, @ssumanas1 (S. Sumanas)
[1] Equal contribution.

https://doi.org/10.1016/j.dib.2026.112459

Specifications Table

| | |
|---|---|
| Subject | Biology |
| Specific subject area | Vascular development and cellular biology. |
| Type of data | Table, Figures, Graphs<br>Processed, Filtered, Analyzed |
| Data collection | The data were collected by loading about 16,000 dissociated cells onto a 10 × Genomics Single Cell 3′ chip and using the Illumina NextSeq 2000 instrument (Illumina, San Diego, CA, USA) to sequence the sample at the USF Genomics Core. |
| Data source location | The cells were collected in October and November of 2024 from pooled embryos at approximately 24 h post fertilization at the University of South Florida, Health Morsani College of Medicine, Taneja College of Pharmacy, and Heart Institute 560 Channelside Dr, Tampa, FL 33602. The same library preparation was used for all single cell sequencing runs. |
| Data accessibility | Repository name: NCBI Gene Expression Omnibus (GEO)<br>Data identification number: GSE290550<br>Direct URL to data:<br>https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE290550<br>Instructions for accessing these data: Open Source |
| Related research article | Valentina Laverde, Luiza Loges, Saulius Sumanas; Zebrafish ETS transcription factor Fli1b functions upstream of Scl/Tal1 during embryonic hematopoiesis. Biol Open 2025; bio.061948. 10.1242/bio.061948. |

## 1. Value of the Data

- These data provide insight into the downstream effects of *fli1b* mutation in zebrafish.
- These data provide insight into the mechanisms that regulate vascular development and hematopoiesis.
- These data will be of specific interest to researchers focused on transcriptional regulation, cell–cell interactions, hematopoietic, vascular and developmental biology.
- These data are valuable because they can be used to interrogate multiple tissue types in wild-type and *MZfli1b* mutants. This allows other researchers to find novel roles for *fli1b* in their tissue of choice and produce hypotheses for testing based on this data. The single cell resolution of this data adds more value as others can discover cell subtypes that have not been discovered previously.

## 2. Background

Friend leukemia integration 1 (Fli1) is an essential ETS transcription factor that regulates the development of both hematopoietic and vascular endothelial lineages [1]. Zebrafish have two homologs of the mammalian *Fli1* gene, *fli1a* and *fli1b*. Previous studies have demonstrated that zebrafish *fli1b* is important for embryonic vascular development [2]. However, its role in hematopoiesis had not been previously studied. We have recently demonstrated that zebrafish *fli1b* mutant embryos show defects in both primitive and definitive hematopoiesis [3]. This dataset was collected to gain insight into the downstream transcriptional effects of *fli1b* and to elucidate the impacts of *fli1b* knockout on transcriptomic expression of different cell populations. This data article provides a thoroughly annotated version of the single-cell RNAseq data used to support the paper by Laverde et al. Here we utilized a different bioinformatic approach to reanalyze the original data, which resulted in a higher number of identified cell types, and we performed a thorough annotation of all cell types, which was absent in the original publication. This dataset can be therefore utilized for future studies in elucidating *fli1b* function and will help to identify molecular mechanisms that regulate vascular and hematopoietic development.

## 3. Data Description

The data includes single cell RNA sequencing data from pooled WT and maternal-zygotic (*MZ*) *fli1b* mutant embryos isolated at 23–24 h post fertilization. The dataset is available in the

**Table 1**
Sequencing data statistics. This table reveals the statistics from the sequencing run, including similar mean reads per cell and median genes per cell detected. The similarity between the wild type and MZfli1b mutant data suggests that they are sufficient for comparison to one another avoiding high rates of sequencing bias.

| Sequencing QC | WT | *MZfli1b* |
|---|---|---|
| Estimated # of cells | 21,007 | 26,665 |
| Mean Reads per cell | 23,828 | 19,453 |
| Median Genes per cell | 1,945 | 1,670 |
| Number of reads | 500,544,737 | 518,725,252 |
| Q30 Bases in RNA read | 93.80% | 94.70 % |
| Fraction of reads mapped to genome | 96.70% | 96.10 % |
| Total genes detected | 24,313 | 24,265 |

NCBI Gene Expression Omnibus (GEO) under the accession number GSE290550. The dataset includes aggregated sequences from WT and *fli1b* mutant embryos in a single matrix file in .h5 format, and a related aggregation.csv file. These files can be uploaded directly into common bioinformatic programs such as Seurat for further analysis. Raw sequence reads are available from SRA archive and include two sets of sequencing runs (150 bp paired end, Illumina platform) for each sample. Each library was sequenced twice to get a higher sequence coverage, and the sequences were then integrated as described in the Methods. The sequencing statistics in Table 1 shows the quality report for WT and *fli1b* mutant samples. The high Q30 scores highlight the quality of the data among samples, reporting a similar number of genes detected as expected. The high percentage of reads mapped to the genome emphasize the high alignment quality of the data. We utilized the FastQC program to further validate the high quality of our scRNAseq data. In Fig. 1, the sequence duplication levels per sample reveal that there are very few duplication events indicating high quality library uniformity and complexity. Fig. 2 highlights the quality of the data with Phred scores above 30-35, indicating excellent read quality. Based on these quality scores the data provides reliable transcriptomic information for further analysis. Fig. 3 is an annotation of cell types present in the data and is open to interpretation by other researchers. We identified 34 cell clusters using Seurat's default clustering resolution and the cell type atlas described in [4]. Table S1 lists marker genes that show significant enrichment in each of the clusters, which were used for cell type assignments.

## 4. Experimental Design, Materials and Methods

This study was performed according to the animal protocols approved by the University of South Florida IACUC, aprroval number IS00012133. Previously established *fli1b*$^{tpl50Gt}$ line (further labeled as *fli1b)* was used in the study [2]. 44 *fli1b-/-* and 39 wild type zebrafish embryos were obtained by the incrossing of homozygous mutant (*fli1b-/-)* and homozygous wild type (*fli1b+/+*), sibling parents, respectively. At approximately 23–24 h post fertilization, the embryos were manually dechorionated and transferred to 1.5 ml microcentrifuge tubes for dissociation into single cell suspension by cold protease [5] and any excess water was removed. All steps were performed on ice. To either sample, 1 ml of deyolking buffer [55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$ in 1X Phosphate Buffered Saline (PBS)] was added and pipetted up and down until full dissolution of the yolk was achieved and the samples were centrifuged at 300G for 1 min. The supernatant was aspirated, and the samples resuspended in 500 µL 0.5x Danieau Buffer [58 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5mM HEPES pH 7.6 in DI water] before centrifugation at 300G for 1 minute and subsequent aspiration of supernatant. Resuspension in 0.5x Danieau Buffer, centrifugation, and supernatant aspiration were repeated once before 1ml of *Bacillus lichenformis* enzyme mix [10mg/ml *B. lichenformensis,* 125 U/ml DNAse I, 0.5 mM EDTA in 10 % Fetal Bovine Serum (FBS) in PBS; prepared immediately prior to addition] was added to each sample and immediately triturated 15 times using a 1ml pipette

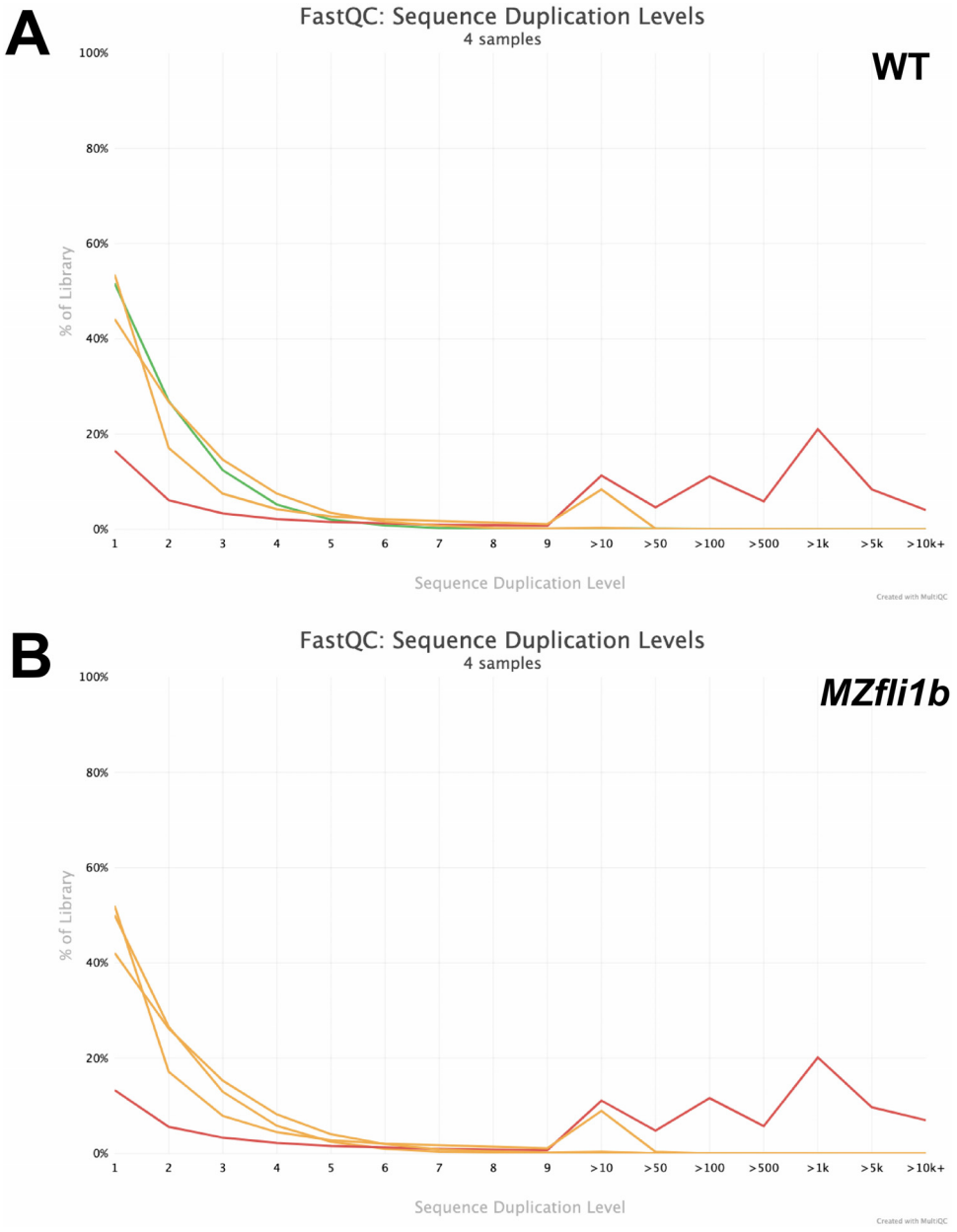


**Fig. 1.** Distribution of sequence duplication levels. A) Shows the FastQC results for duplication levels of the wild type (WT) fastq files (samples). B) Shows the same FastQC results for the *fli1b* mutant fastq files. These reveal a low level of duplication which suggests quality data without amplification bias.

with subsequent trituration 15 times approximately every 2 min for approximately 20 min, until dissociation was confirmed by visualization of a 1–2 µl aliquot of each sample with minimal clumps of cells remaining. Samples were then centrifuged at 1200 G for 5 min. Supernatant was aspirated and pellets were resuspended in 1 ml 10 % FBS in PBS before being passed through a 20 µm strainer into a clean 1.5 mL tube and the strainer was rinsed with 500 ml of 10 % FBS

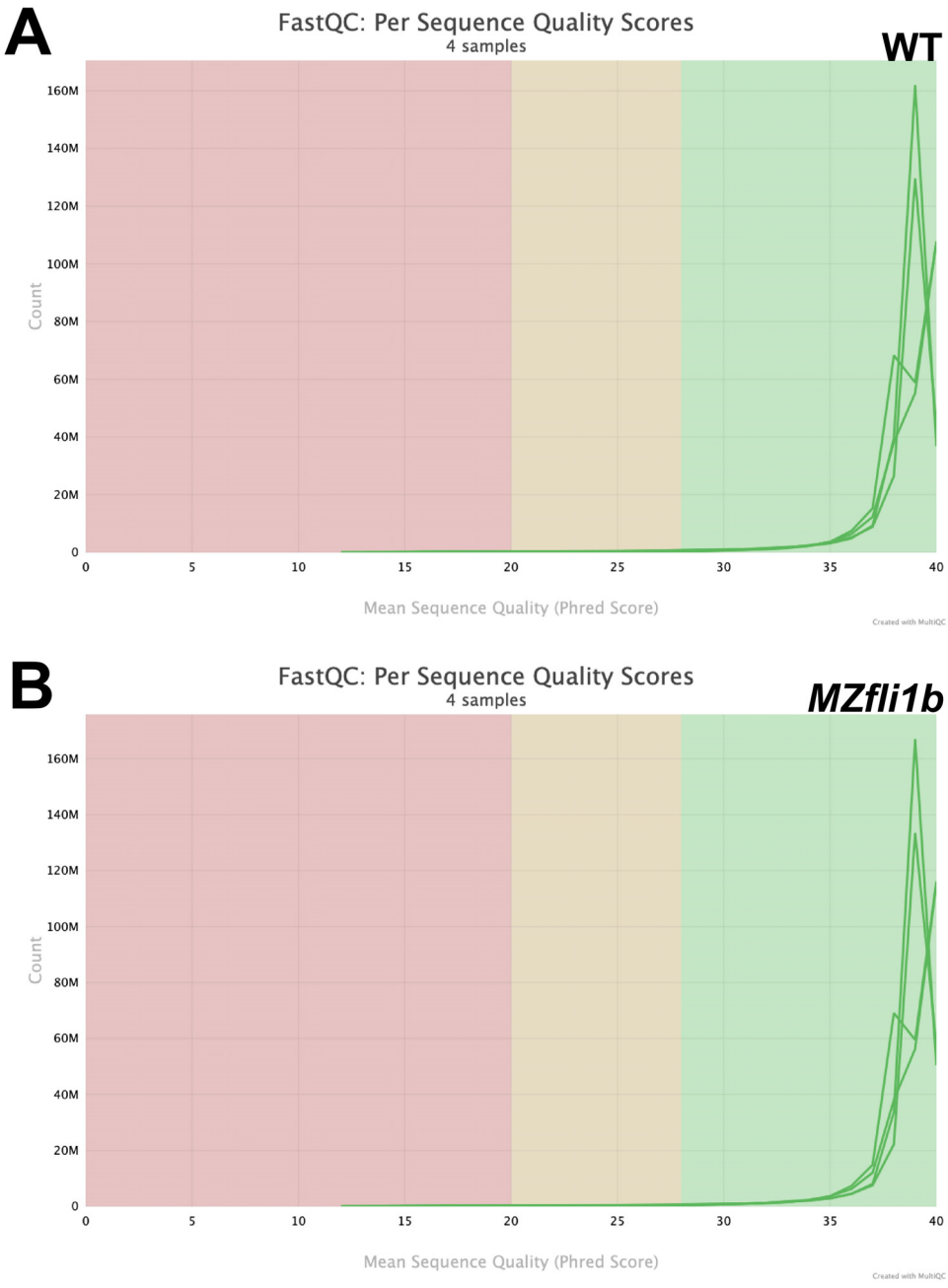


**Fig. 2.** Phred scores for WT and *MZfli1b* samples A) Shows the Phred scores for the wild type (WT) fastq files. B) Shows the Phred scores for the *MZfli1b* mutants. Scores above 25 are normally considered good quality reads. This data shows majority of scores to be higher than 30.

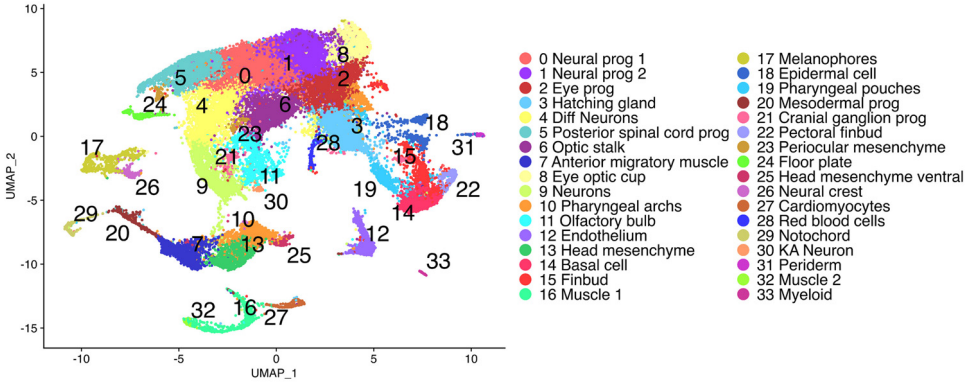


**Fig. 3.** Predicted cell types per cluster. Cell type annotations were performed based on Saunders et al, 2023 [4]. Abbreviations: prog, progenitors; diff, differentiated; KA Neuron, Kolmer-Agduhr neurons.

in PBS. Cells were pelleted at 1200G for 5 min, resuspended in 1 ml 10 % FBS in PBS and pelleted again at 1200G for 5 min. Cells were then resuspended in 250 µl of 10 % FBS in PBS and an aliquot stained with trypan blue was counted on a hemocytometer to determine total live cell number. A suspension of approximately 16,000 cells was loaded onto a 10x Genomics Single Cell 3' chip at the USF Genomics Core for each genotype and sequencing was then performed on these chips using an Illumina NextSeq 2000 instrument (Illumina, San Diego, CA).

The fastq files for each sample were received from 10X Genomics and run through FastQC to check read quality. Once quality was assessed we utilized the Cell Ranger (8.0.1) command line tool, cellranger count to process each sample separately, this produced a .h5 file per sample. Utilizing R packages; Seurat (5.3.1), ggplot2 (4.0.0), dplyr (1.1.4), tidyverse (2.0.0), hdf5r (1.3.12) and reshape2 (1.4.5) we processed the .h5 files [6–14]. First, we created Seurat objects and filtered the data according to violin plots revealing the distribution of features, counts and percent mitochondrial genes present. We filtered to remove the cells with a number of features less than 200 and greater than 4500, and also removed cells with mitochondrial gene contents greater than 5 %. Next, we normalized the data and used Seurat v4 method of integration, finding anchor genes and using these to ensure similar cells clustered together. Then, we scaled the data and ran dimensionality reductions PCA and UMAP on the data. Using the default clustering resolution of 0.8 we found 34 cell clusters. We used the cell atlas gene lists of 152 cell types from [4] to plot heatmaps of genes common in our data set and the atlas, assigning cell types based on expression patterns. If cell types were not clear from this overlap, we inferred them from differentially expressed genes using Seurat's FindAllMarkers function and a negative binomial assumption.

## Limitations

*Limitations of this work include the small sample size and the lack of independent repeats.*

## Ethics Statement

All procedures involving animals in this study were conducted in accordance with the National Institutes for Health guide for the care and use of laboratory animals and were approved by the Institutional Animal Care and Use Committee of the University of South Florida.

## Credit Author Statement

**Luiza Loges**: Investigation, Writing- Original Draft, Writing- Review and Editing. **Ricardo DeMoya**: Methodology, Software, Validation, Formal Analysis, Data Curation, Writing- Original Draft, Writing- Review and Editing, Visualization. **Valentina Laverde**: Investigation, Resources, Methodology. **Saulius Sumanas**: Conceptualization, Methodology, Resources, Writing-Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition.

## Acknowledgements

*We thank Matthew Mercurio and the USF Genomics Core for their assistance in performing single-cell RNA-seq.*

*This work was supported by the* National Institutes of Health *(R01HL153005 award to S.S.)*

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2026.112459.

## Data Availability

Single-cell transcriptomic analysis of zebrafish fli1b mutants (Original data) (GEO)

## References

[1] Y. Ben-David, et al., Current insights into the role of Fli-1 in hematopoiesis and malignant transformation, Cell. Mol. Life Sci. 79 (3) (2022) 163.
[2] M.P. Craig, et al., Etv2 and fli1b function together as key regulators of vasculogenesis and angiogenesis, Arterioscler. Thromb. Vasc. Biol. 35 (4) (2015) 865–876.
[3] V. Laverde, L. Loges, S. Sumanas, The zebrafish ETS transcription factor Fli1b functions upstream of Scl/Tal1 during embryonic hematopoiesis, Biol. Open 14 (4) (2025).
[4] L.M. Saunders, et al., Embryo-scale reverse genetics at single-cell resolution, Nature 623 (7988) (2023) 782–791.
[5] A.S. Potter, S.S. Potter, Dissociation of tissues for single-cell analysis, Methods Mol. Biol. (2019) 55–62 1926.
[6] Y. Hao, et al., Integrated analysis of multimodal single-cell data, Cell 184 (13) (2021) 3573–3587 e29.
[7] Y. Hao, et al., Dictionary learning for integrative, multimodal and scalable single-cell analysis, Nat. Biotechnol. 42 (2) (2024) 293–304.
[8] Hoefling, H. and M. Annau, hdf5r: Interface to the HDF5 binary data format. 2022, R package version.
[9] Posit team. RStudio: integrated development for R. Posit software, PBC, Boston, MA. 2025.
[10] R Core Team, R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. 2021.
[11] H. Wickham, Reshaping data with the reshape package, J. Stat. Software 21 (2007) 1–20.
[12] H. Wickham, in: ggplot2: elegant graphics for data analysis, Springer, New York. NY, 2009, p. 10.
[13] Wickham, H., dplyr: a grammar of data manipulation. R package version 04., 2015. 3: p156.
[14] H. Wickham, et al., Welcome to the Tidyverse, J. Open Source Softw. 4 (43) (2019) 1686.