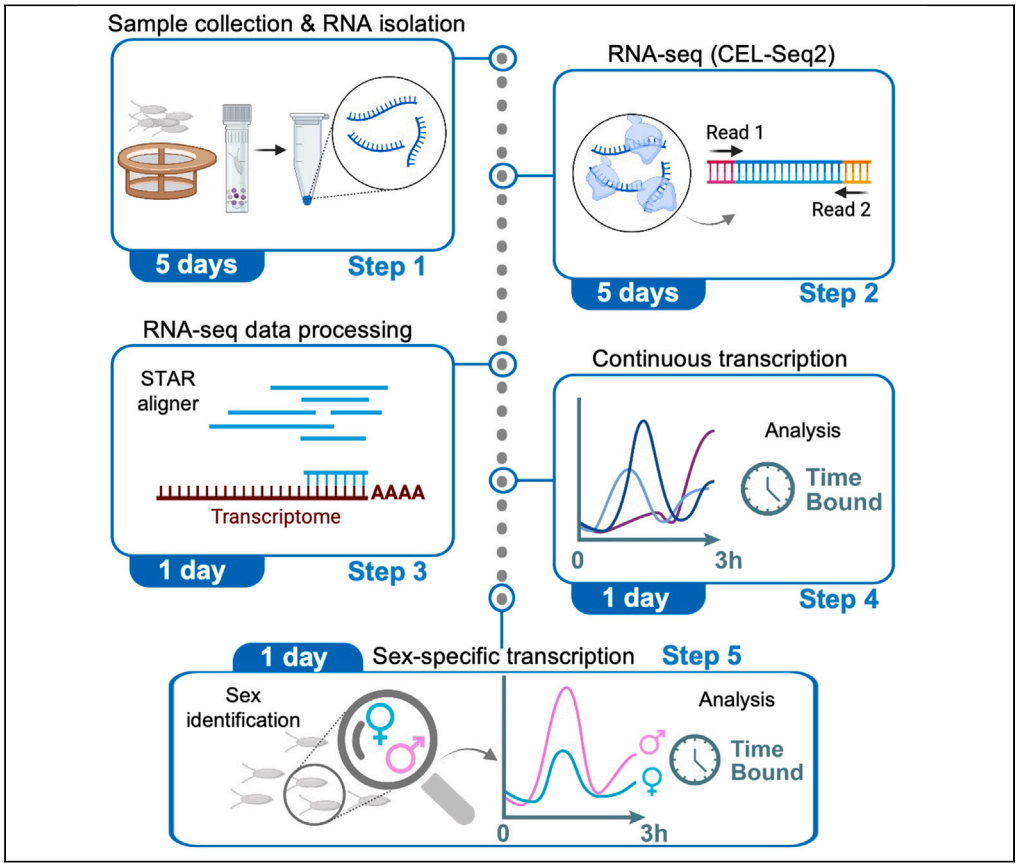
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Single-embryo RNA sequencing for continuous and sex-specific gene expression analysis on *Drosophila*



Exploring early embryonic gene expression is challenging due to the rate of development and the limited material available. Here, we present a protocol for ordering *Drosophila* embryos along a developmental pseudo-time trajectory and determining the sex of the embryos using RNA-seq data. We describe steps for sample collection, RNA isolation, RNA-seq, and RNA-seq data processing. We then detail the establishment of a continuous transcriptome dataset for assessing gene expression throughout early development and in a sex-specific manner.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

J. Eduardo Pérez-Mojica, Lennart Enders, Kin H. Lau, Adelheid Lempradl

eduardo.perezmojica@vai.org (J.E.P.-M.)
heidi.lempradl@vai.org (A.L.)

Highlights

An accessible RNA sequencing method for single-*Drosophila* embryos

A complete workflow for continuous expression analysis for thousands of genes

High-resolution transcriptome profiling during early embryo development

Identification of male and female embryos enables sex-specific expression analysis

Pérez-Mojica et al., STAR Protocols 4, 102535
September 15, 2023
https://doi.org/10.1016/j.xpro.2023.102535

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Single-embryo RNA sequencing for continuous and sex-specific gene expression analysis on *Drosophila*

J. Eduardo Pérez-Mojica,[1,5,*] Lennart Enders,[2,4] Kin H. Lau,[3] and Adelheid Lempradl[1,6,*]

[1]Department of Metabolism and Nutritional Programming, Van Andel Institute, Grand Rapids, MI 49503, USA

[2]Department of Epigenetics, Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg im Breisgau, Germany

[3]Bioinformatics and Biostatistics Core, Van Andel Institute, Grand Rapids, MI 49503, USA

[4]Present address: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria

[5]Technical contact

[6]Lead contact

*Correspondence: eduardo.perezmojica@vai.org (J.E.P.-M.), heidi.lempradl@vai.org (A.L.)
https://doi.org/10.1016/j.xpro.2023.102535

SUMMARY

Exploring early embryonic gene expression is challenging due to the rate of development and the limited material available. Here, we present a protocol for ordering *Drosophila* embryos along a developmental pseudo-time trajectory and determining the sex of the embryos using RNA-seq data. We describe steps for sample collection, RNA isolation, RNA-seq, and RNA-seq data processing. We then detail the establishment of a continuous transcriptome dataset for assessing gene expression throughout early development and in a sex-specific manner. For complete details on the use and execution of this protocol, please refer to Pérez-Mojica et al.[1]

BEFORE YOU BEGIN

The protocol below describes the specific steps for sequencing 192 *Drosophila* embryos during the first 3 h of development using the I.DOT (Dispendix) liquid handler. However, this protocol also works for zebrafish and mouse embryos and can also be accomplished with other liquid handlers that are capable of accurately dispensing volumes down to 120 nL (we have also successfully used the mosquito LV from sptlabtech). Here, we detail an approach where 192 *Drosophila* embryos are processed and pooled to create two sequencing libraries. This protocol also includes a procedure for computational analysis. While we have included an in-silico approach for determining the sex of embryos between nuclear cycle 12 and 14 (1.5–3 h into development), we also describe a protocol for sex chromosome genotyping. Before beginning to collect samples for sequencing, we recommend optimizing the egg laying procedure for the specific *Drosophila* strain being used.

Optimization of fly egg laying (optional but highly recommended)

Timing: 1–7 days

1. Evaluate the egg laying behavior of the fly strain of interest to ensure that 64 embryos per 1-h collection interval and condition are obtained. Adjust the age and number of flies, hour of the day, and time interval for egg laying as necessary.
   a. Transfer 1-day-old flies into embryo collection cages (approximately 200 females and 100 males) and change the food plates at least every 2 days.
   b. Run egg laying tests, using 1-h collection intervals, to evaluate egg laying behavior.

i. Before each test, exchange the food plate twice after a 1-h interval and discard. This step allows females to lay withheld (older) embryos before the actual collection begins.
ii. Exchange the plate after 1-h of egg laying and count the number of embryos.
iii. Evaluate different conditions and compare the egg laying rate.
iv. Determine the optimal egg laying conditions for your fly strain and experimental setup.

Alternatives: Flies of any age can be used, but keep in mind that age impacts egg laying activity. Some fly strains also show reduced egg laying activity when cages are flipped too often. In our experience, a 1-h interval will work for most strains. However, egg laying times might have to be extended to 1.5 h or more in some cases. While the number of female flies can be increased to improve embryo yield for some strains with low fecundity, in general, we recommend adding more cages instead, as overcrowding can also cause stress and further reduce yield.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Pure Bright Bleach (6% sodium hypochlorite)	KIK International	Cat# 11008635042
Supelco NaCl (sodium chloride)	MilliporeSigma	Cat# SX0420-3
Triton X-100	Thermo Fisher Scientific	Cat# BP151-100
Gibco PBS (phosphate-buffered saline), pH 7.2	Thermo Fisher Scientific	Cat# 20012027
TRIzol Reagent	Thermo Fisher Scientific	Cat# 15596026
Chloroform	MilliporeSigma	Cat# 319988
Isopropanol	MilliporeSigma	Cat# 190764
GlycoBlue Co-precipitant	Thermo Fisher Scientific	Cat# AM9516
Invitrogen Nuclease-free water	Thermo Fisher Scientific	Cat# AM9938
ERCC RNA Spike-In Mix	Thermo Fisher Scientific	Cat# 4456740
dNTP Mix (10 mM/each)	Thermo Fisher Scientific	Cat# 18427088
Vapor-Lock	Qiagen	Cat# 981611
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	Cat# 10777019
Second Strand Buffer	Thermo Fisher Scientific	Cat# 10812014
DNA Polymerase I	Thermo Fisher Scientific	Cat# 18010025
E. coli DNA Ligase	Thermo Fisher Scientific	Cat# 18052019
Ambion RNase H, from *E. coli*, 10 U/μL	Thermo Fisher Scientific	Cat# AM2293
Pharmco Ethanol	Greenfield Global	Cat# 111000200CSPP
AMPure XP reagent	Beckman Coulter	Cat# A63881
RNAClean XP	Beckman Coulter	Cat# A63987
ExoSAP-IT PCR Product Cleanup Reagent	Thermo Fisher Scientific	Cat# 78200.200.UL
Phusion High-Fidelity PCR Master Mix with HF Buffer	New England Biolabs	Cat# M0531S
Supelco NaOH (sodium hydroxide)	MilliporeSigma	Cat# SX0590-1
Critical commercial assays		
SuperScript II Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18064014
MEGAscript T7 Transcription Kit	Thermo Fisher Scientific	Cat# AM1333
NEBNext Magnesium RNA Fragmentation Module	New England Biolabs	Cat# E6150S
Agilent RNA 6000 Pico Kit	Agilent Technologies	Cat# 5067-1513
Qubit dsDNA HS (High Sensitivity) Assay Kit	Thermo Fisher Scientific	Cat# Q32851
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	Cat# **1725271**
Deposited data		
Single-embryo RNA-seq fastq files and raw reads	Pérez-Mojica et al.[1]	GEO: GSE214118
Oligonucleotides		
CEL-Seq2 RT primers: GCCGGTAATACGACTCA CTATAGGGAGTTCTACAGTCCGACGATCNNN NNN[6-nucleotide barcode]TTTTTTTTTTTTTTT TTTTTTTTV	Hashimshony et al.[2]	https://doi.org/10.1186/s13059-016-0938-8

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CEL-Seq2 library RT primer: GCCTTGGCACCC GAGAATTCCANNNNNN	Hashimshony et al.[2]	https://doi.org/10.1186/s13059-016-0938-8
RPI primer: AATGATACGGCGACCACCGAGAT CTACACGTTCAGAGTTCTACAGTCCGA	Sagar et al.[3]	https://doi.org/10.1007/978-1-4939-7768-0_15
RPI1 primer: CAAGCAGAAGACGGCATACGAGA TCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	Sagar et al.[3]	https://doi.org/10.1007/978-1-4939-7768-0_15
X chromosome forward primer: TTGGGCTGCTTCAGGTTTGA	Pérez-Mojica et al.[1]	https://doi.org/10.1016/j.xgen.2023.100265
X chromosome reverse primer: GAAGAGACACGCCAAGGCTA	Pérez-Mojica et al.[1]	https://doi.org/10.1016/j.xgen.2023.100265
Y chromosome forward primer: TCATAAGGAGTGAAGCGGTCC	Pérez-Mojica et al.[1]	https://doi.org/10.1016/j.xgen.2023.100265
Y chromosome reverse primer: AATTGTGTGCATCGGTGGGTC	Pérez-Mojica et al.[1]	https://doi.org/10.1016/j.xgen.2023.100265
Software and algorithms		
STAR (v2.7.8a)	Dobin et al.[4]	https://github.com/alexdobin/STAR
R (v4.2.2)	The R Foundation	https://www.r-project.org
RStudio (v2022.12.0 + 353)	Posit Software, PBC	https://posit.co/download/rstudio-desktop/
readr (v2.1.4)	Wickham et al.[5]	https://readr.tidyverse.org
dplyr (v1.1.2)	Wickham et al.[5]	https://dplyr.tidyverse.org
tibble (v3.2.0)	Wickham et al.[5]	https://tibble.tidyverse.org
DropletUtils (v1.20.0)	Lun et al.[6]	https://doi.org/10.18129/B9.bioc.DropletUtils
scuttle (v1.10.1)	McCarthy et al.[7]	https://doi.org/10.18129/B9.bioc.scuttle
RaceID (v0.3.0)	Herman et al.[8]	https://github.com/dgrun/RaceID3_StemID2_package
DESeq2 (v1.38.3)	Love et al.[9]	https://doi.org/10.18129/B9.bioc.DESeq2
splineTimeR (v1.26.0)	Michna et al.[10]	https://doi.org/10.18129/B9.bioc.splineTimeR
Prism (v9.4.1)	GraphPad Software	https://www.graphpad.com
Other		
Embryo collection cage	Genesee Scientific	Cat# 59-105
Petri dishes (60 × 15 mm)	Corning	Cat# 430166
Fly food M	LabExpress	Cat# 7002-Bulk
pluriStrainer 150 μm (cell strainer)	pluriSelect	Cat# 43-50150-03
Paintbrush, angular, size 1/8	Royal & Langnickel	Cat# SG160
Liner brush, size 20/0	Royal & Langnickel	Cat# Z73L
2-mL screw cap, skirted tubes	Thermo Fisher Scientific	Cat# 3488
Lysing Matrix D (1.4 mm ceramic beads)	MP Biomedicals	Cat# 116540434
FastPrep-24 Classic bead beating system	MP Biomedicals	Cat# 6004500
I.DOT non-contact liquid handler	Dispendix	https://dispendix.com/idot-non-contact-liquid-handler/
I.DOT source plates, orifice size 100 μm	Dispendix	Cat# S.100
Micropipette Transferpette S, 0.1–1 μL	BrandTech Scientific	Cat# 705868
PIPETMAN L pipette 1–10 μL	Gilson	SKU# FA1002M
PIPETMAN L pipette 2–20 μL	Gilson	SKU# FA1003M
PIPETMAN L pipette 20–200 μL	Gilson	SKU# FA1005M
PIPETMAN L pipette 100–1000 μL	Gilson	SKU# FA1006M
Tacta Mechanical Pipette, 8 Channel, 0.5–10 μL	Sartorius	Cat# LH-729120
Tacta Mechanical Pipette, 8 Channel, 5–100 μL	Sartorius	Cat# LH-729130
Reagent reservoir (25 mL)	Thermo Fisher Scientific	Cat# 8093-11
Pipette filter tips, 0.1–1 μL	BrandTech Scientific	Cat# 732722
VWR Pipette tips, low-retention, filtered, 0.1–10 μL	Avantor	Cat# 76322-528
VWR Pipette tips, low-retention, filtered, 1–20 μL	Avantor	Cat# 76322-134
VWR Pipette tips, low-retention, filtered, 20–200 μL	Avantor	Cat# 76322-150
VWR Pipette tips, low-retention, filtered, 100–1000 μL	Avantor	Cat# 76322-154
Microseal 'B' PCR seal (for reactions)	Bio-Rad	Cat# MSB1001
Microseal 'F' PCR seal (for storage at −80°C)	Bio-Rad	Cat# MSF1001
Hard-Shell 384-well PCR plates, thin wall, skirted.	Bio-Rad	Cat# HSP3801
Eppendorf twin.tec 96-well PCR plates, LoBind	Eppendorf	Cat# 0030129504
Invitrogen DynaMag-96 side skirted magnet	Thermo Fisher Scientific	Cat# 12027
DNA LoBind Tubes, 1.5 mL	Eppendorf	Cat# 022431021

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
2100 Bioanalyzer Instrument	Agilent Technologies	Cat# G2939BA
Syringe (for Bioanalyzer)	Agilent Technologies	Cat# G2938-68706
Qubit 4 Fluorometer	Thermo Fisher Scientific	Cat# Q33238
Qubit assay tubes	Thermo Fisher Scientific	Cat# Q32856

MATERIALS AND EQUIPMENT

I.DOT program settings

- **General settings**: For all programs below, use a 96-well plate in the source plate position and a 384-well plate in the target plate position. Use 'deionisation' and 'reorder and parallel dispensing (level 2)' options. The latter instructs the instrument to dispense volume in every other row in the target plates. This is the fastest way to dispense volumes from a 96-well plate onto a 384-well plate. Select 'dispense to waste' to prime wells of the source plate when instructed to do so by the specific program. Run dispensing simulations to ensure that the program is working as expected before use.
- **Barcoded-primer mix #1:** Program I.DOT to dispense 240 nL from each well in a source plate onto columns 1–12 rows A, C, E, G, I, K, M, and O of a target plate. Prime each well in the source plate by dispensing 100 nL for 3 cycles. Each barcoded-primer mix in the source plate must be dispensed only one time to a different well of the target plate. Record which well in the source plate is transferred to which well in the target plate to ensure that you know the arrangement of the barcoded primers in the target plate. Save this to I.DOT as barcoded-primer mix #1 protocol.
- **Barcoded-primer mix #2:** Program I.DOT to dispense 240 nL from each well in a source plate onto columns 1–12 rows B, D, F, H, J, L, N, and P of a target plate. Prime each well in the source plate by dispensing 100 nL for 3 cycles. Each barcoded-primer mix in the source plate must be dispensed only one time to a different well of the target plate. Record which well in the source plate is transferred to which well in the target plate to ensure that you know the arrangement of the barcoded primers in the target plate. Save this to I.DOT as barcoded-primer mix #2 protocol.
- **Vapor-lock:** Program I.DOT to dispense 1.2 µL from all wells in column 1 of the source plate onto each well of columns 1–12 of the target plate. Save this as vaporlock protocol.
- **First-strand synthesis (FSS):** Program I.DOT to dispense 160 nL from well A2 or B2 of the source plate onto each well in columns 1–12 of the target plate. Prime each well in the source plate by dispensing 200 nL for 3 cycles. Save this as FSS protocol.
- **Second strand synthesis (SSS):** Program I.DOT to dispense 2.2 µL from column 3 of the source plate onto each well in columns 1–12 of the target plate. Prime each well in the source plate by dispensing 200 nL for 3 cycles. Save this as SSS protocol.

Alternatives: Any liquid handler capable of accurately dispense 120 nL can be used. The steps described in this protocol are specific for the I.DOT non-contact liquid handler. Depending on the instrument, RNA could directly be dispensed into working plates using the liquid handler.

Alternatives: Any pipette that can confidently dispense 120 nL can be used in place of the Transferpette S (0.1–1 µL) to transfer RNA samples into working plates.

Short recipes.

- **Egg-wash solution** (120 mM NaCl, 0.03% Triton X-100): dissolve 7.13 g NaCl and 300 µL Triton X-100 in 996.5 mL MilliQ-grade H_2O.

Stored at room temperature indefinitely.

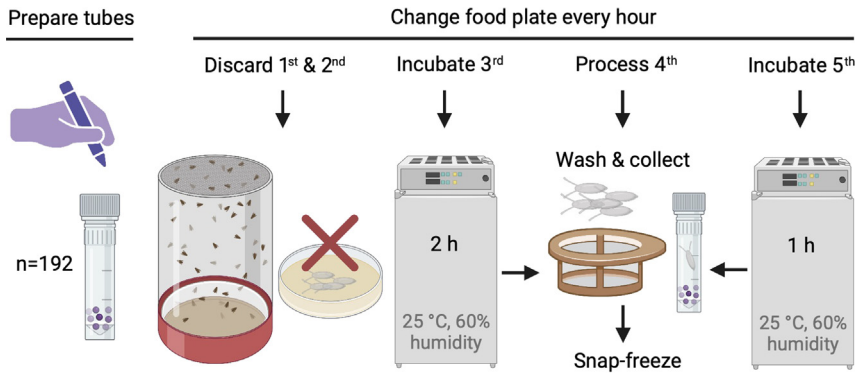


Figure 1. Embryo collection diagram

- **3% sodium hypochlorite**: combine 50 mL 6% sodium hypochlorite (household bleach) with 50 mL MilliQ-grade H_2O.

Store at room temperature for up to 1 month.

- **1 M NaOH (w/v):** dissolve 10 g NaOH in 200 mL MilliQ-grade H_2O and add additional MilliQ-grade H_2O to a total volume of 250 mL.

Store at room temperature for up to 1 year.

- **8 mM NaOH (w/v):** add 100 µL of 1 M NaOH to 12.4 mL MilliQ-grade H_2O.

Store at room temperature for up to 1 month.

STEP-BY-STEP METHOD DETAILS

Embryo collection and RNA isolation from single embryos

Timing: 5 days

This major step describes the collection and processing of embryos and RNA isolation. Embryos are collected in 1-h time intervals to ensure an equal distribution of embryos across the developmental time window. In this protocol, we detail the steps to study the first 3 h of embryonic development. Embryos are dechorionated, washed, and frozen. To obtain total RNA from individual embryos, RNA is isolated using a miniaturized version of a standard TRIzol protocol. At the end of these steps, total RNA from 192 individual embryos will be obtained.

1. Prepare tubes for embryo collection.
   a. Weight 0.2 g ± 0.02 of 1.4 mm ceramic beads (lysing matrix D) into 192 2 mL screw-cap tubes.
   b. Label each tube on both the side and the lid, as later steps might remove the labels.
      i. Label each of the 192 tubes with a prefix for the collection time interval (1, 2, or 3), and the embryo number (1–64). For example, label 2.15 corresponds to the 15th embryo collected from the 2-h collection window.

   Alternatives: 2 mL screw-cap tubes already pre-filled with 1.4 mm ceramic beads can be purchased instead.

2. Collect embryos in three 1-h time intervals (Figure 1).

a. Before collecting the first study timepoint, exchange the food plate twice, each after a 1-h interval, and discard embryos. This step allows females to lay withheld (older) embryos, before the actual study begins.
b. Replace the first study plate with a new food plate and incubate the embryos on the first plate for 2 h at 25°C and 60% humidity, before proceeding to step 3. These will be your 2–3 h old embryos.
c. Exchange the next food plate and proceed to step 3 immediately. These will be your 0–1 h old embryos.
d. Remove the final food plate and incubate for 1 h at 25°C and 60% humidity, before proceeding to step 3. These will be your 1–2 h old embryos.

Note: To make food plates, heat fly food in a beaker using a microwave and transfer 10 mL to petri dishes (60 × 15 mm). Let them cool uncovered for 20 min at room temperature in a clean room. Cover petri dishes and store at 4°C for up to 1 month.

Optional: To assess the fertility rate of the parents, collect additional embryos the same day. The fertility rate is specific for your fly strain and experimental conditions, and it is used to estimate the number of unfertilized eggs expected in the experiment. Transfer at least 150 eggs to a petri dish with enough MilliQ-grade water to cover all embryos. Incubate at 25°C with 60% humidity. Count the number of hatched larvae after 48-h. Calculate the fertility rate by dividing the number of larvae hatched by the total number of eggs and multiply by 100. For example, if 132 larvae hatched from 150 eggs (132/150*100) this would equal to 88% fertility. This percentage will allow you to estimate the number of unfertilized eggs expected in your final dataset. Unfertilized eggs will be excluded during data analysis (see step 36).

3. Wash embryos.
   a. Add some egg-wash solution to the food plate and use a paintbrush (size 1/8) to gently loosen the embryos from the food. Pour the egg-wash solution, including the embryos, from the food plate onto a 150 μM mesh-size cell strainer (pluriStrainer).
   b. Rinse the embryos on cell strainer with tap water for at least 15 s. Use a wash bottle to control the water pressure and avoid losing embryos.
   c. Place the strainer into the lid of a 50 mL centrifuge tube and add 3% sodium hypochlorite so that all embryos are completely submerged, and incubate for 4 min. Use only freshly prepared solution (maximum storage time: 1 month).
   d. Rinse the embryos with tap water for at least 15 s using a wash bottle.
   e. Soak the embryos in egg-wash solution for 1 min, as described for step 3.c.
   f. Wash the embryos with MilliQ-grade water for 15 s using a wash bottle.
4. Snap-freeze embryos using dry ice.
   a. Transfer each single embryo into a pre-labeled 2 mL tube (prepared in step 1) using a liner brush (size 20/0). Place embryos in the middle of the tube to avoid accidental lysis with the ceramic beads during storage.
   b. Place the tube on dry ice. Avoid shaking.
   c. After processing all samples, store embryos at −80°C.

Pause point: Embryos collected can be stored at −80°C until RNA isolation. We have stored samples for up to 1 year, however, RNA is prone to degradation and samples should be used as soon as possible. To avoid possible batch effects due to different storage times, embryos from all experimental conditions need to be included in each batch.

⚠ CRITICAL: TRIzol is a mixture of volatile hazardous chemicals. All steps involving an opened tube containing TRIzol should be performed using a chemical fume hood.

Note: We recommend processing no more than 24 samples for RNA isolation at the same time.

5. Homogenize embryos.
   a. Remove the embryo samples from −80°C freezer and place them on ice.
   b. Add 500 μL TRIzol reagent into each tube. After this step, there is no need to keep samples on ice unless otherwise specified.
   c. Add 50 μL PBS, pH 7.2.
   d. Ensure that all tubes are tightly closed.
   e. Lyse embryos using a benchtop homogenizer (FastPrep-24) at 6 m/s for 30 s at room temperature.
6. Recover aqueous RNA-containing phase.
   a. Incubate samples for 5 min at room temperature.
   b. Add 100 μL chloroform.
   c. Vortex vigorously for 15 s and incubate for 2 min at room temperature.
   d. Centrifuge at 12,000 × *g* for 15 min at 4°C.
      i. In the meantime, place isopropanol (250 μL per sample) on ice to cool and label new 1.5 mL tubes for RNA samples.
   e. Carefully transfer 300 μL of the upper aqueous phase (colorless) to the previously labeled 1.5 mL tubes.
      i. Use 200 μL tips and transfer only 150 μL at a time to better control pipetting. Hold origin tubes at a 45° angle during transfer, placing the tip at the top of the aqueous layer to avoid touching and transferring any of the interphase or lower organic phase.
7. Store the interphase and organic phase.
   a. Ensure all tubes are tightly closed.
   b. Store samples at −80°C for optional DNA extraction for sex chromosome genotyping (see steps 44 and 45, respectively).

Note: If you touch or pull up any of the interphase or lower organic phase (red phenol-chloroform phase) with the pipette tip, transfer everything back into the tube, repeat the centrifugation step (step 6.d), and try again. Traces of phenol will inhibit downstream reactions.

8. RNA precipitation.
   a. Add 250 μL ice-cold isopropanol to the recovered aqueous phase (see step 6).
   b. Add 2 μL GlycoBlue co-precipitant (15 mg/mL).
   c. Mix by hand, inverting the tube 5 times.
   d. Incubate samples for 10 min at room temperature.
   e. Centrifuge at 12,000 × *g* for 10 min at 4°C.
   f. Remove and discard the supernatant. Remove as much liquid as possible without disturbing the RNA-containing blue pellet.
   g. Add 1 mL 75% ethanol to RNA pellets.
   h. Centrifuge at 7,500 × *g* for 10 min at 4°C.
   i. Remove and discard the supernatant without disturbing the pellet.
   j. Air-dry samples until completely dry (approximately 20 min).
   k. Store RNA pellets at −80°C.

Note: To practice and confirm that RNA can be successfully isolated, use several spare samples and quantify the total RNA yield. Total RNA from a single embryo in the first 3 h of development should be ~50 ng (~10 ng/μL from a 5 μL RNA elution).

CEL-Seq2 protocol

Timing: 5 days

This major step describes a modification of a previously published single-cell sequencing protocol (CEL-seq2)[2,3] that has been modified to work with single embryos. Major adjustments to the original protocol include: 1) using RNA as a starting point, instead of intact single cells; 2) diluting cDNA

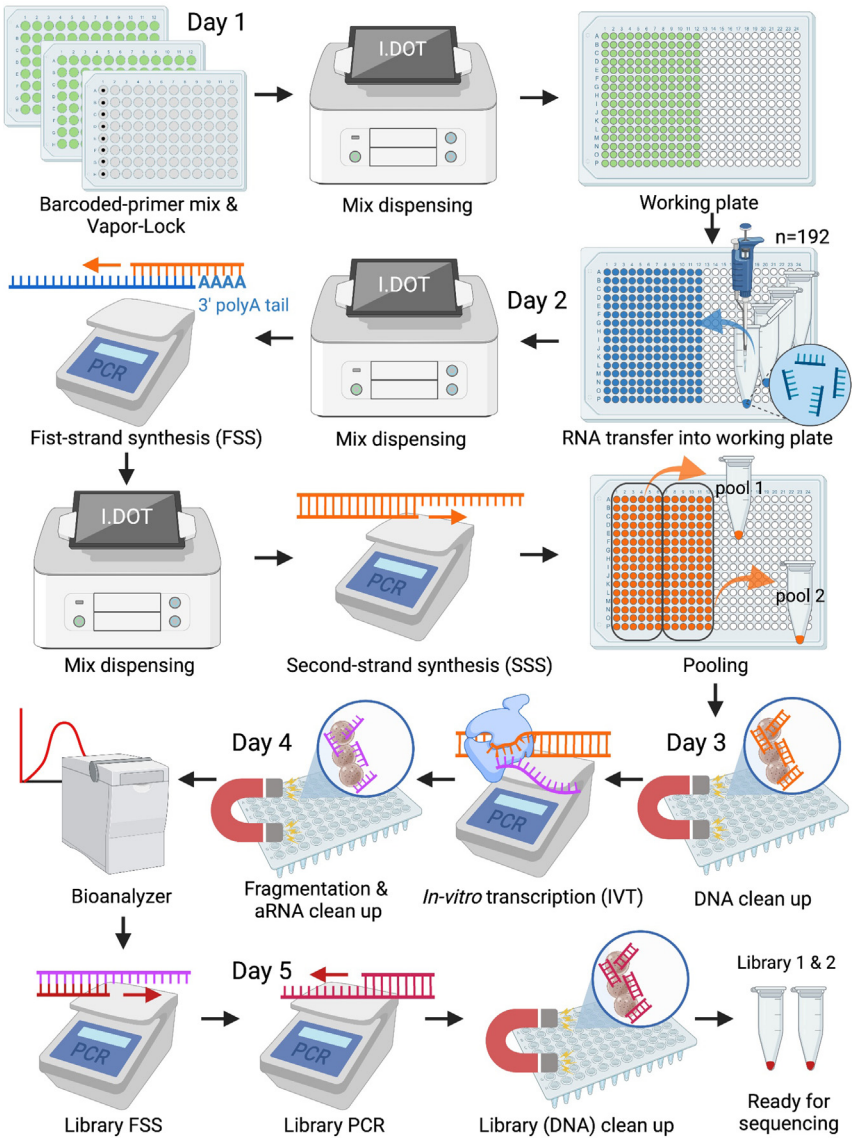


Figure 2. Modified CEL-Seq2 protocol adjusted to work with single embryos

concentration by 20-fold prior to library PCR. The protocol described below is suited for working with the I.DOT non-contact liquid handler and 96-well source plates (S.100, 100 μm pore). At the end of the protocol, two DNA libraries will have been created, each containing 96 individual embryo samples. An overview of this major step is shown in Figure 2.

> *Note:* The following oligonucleotides are needed for the RNA-seq protocol: 1) 192 RT primers containing a T7 promoter, a part of the Illumina 5′ adaptor, a 6-nucloetide unique molecular identifier (UMI), a 6-nucleotide barcode, and an anchored poly-T tail; 2) a library RT primer containing a random hexamer with the Illumina 3′ adaptor sequence in the 5′-tail; and 3) one RNA PCR primer (RPI) and one RPI index (here we used RPI index 1) (The RPI primer and indexes described are the same as those provided by Illumina, Inc).

9. Acquire primers. We recommend Integrated DNA Technologies (IDT, USA) for primer synthesis. Primer sequences are listed in the key resources table and Table 1.

Table 1. Barcode sequences for CEL-seq2 RT primers

#	Barcode	#	Barcode	#	Barcode	#	Barcode	#	Barcode
1	AGTGTC	41	GTACTC	81	TGTACG	121	CGGTTA	161	GTTAGG
2	ACCATG	42	TGTCGA	82	GACGAA	122	CTAACG	162	TACGCA
3	GAGTGA	43	ACGTGA	83	CCACAA	123	CACGTA	163	AGATGC
4	CACTCA	44	AGGATC	84	CACCAA	124	TTGGAG	164	TTGCAC
5	CATGTC	45	CTCATG	85	CTAAGC	125	GCAATG	165	CAGGAA
6	ACAGGA	46	AGACTC	86	GATACG	126	TATCCG	166	TCTAGG
7	GTACCA	47	GACAAC	87	ACAAGC	127	ATGCAG	167	GCTTCA
8	ACAGAC	48	AGGACA	88	TGAAGG	128	GCTCTA	168	TTGGTC
9	ACGTTG	49	ACTCAC	89	TAACGG	129	ATTCGC	169	TACCGA
10	ACCAAC	50	GTACAG	90	AACCTC	130	TGTTGG	170	CATTGG
11	GTGAAG	51	AGGAAG	91	CGTCTA	131	ATGACC	171	CTCGAA
12	CACTTC	52	AGTGCA	92	CCATAG	132	CCGTAA	172	GCTTAC
13	GAGTTG	53	TGGTGA	93	TTCCAG	133	TGATCG	173	ATACGG
14	GAAGAC	54	AGACAG	94	GGACAA	134	TACAGG	174	GTATGC
15	TGCAGA	55	AGCTTC	95	ACTTCG	135	AGAACG	175	TGTAGC
16	CTAGGA	56	TGAGGA	96	TTGTGC	136	GCCATA	176	CGTAAG
17	ACCAGA	57	ACGTAC	97	GGTATG	137	ACGGTA	177	TTACGC
18	GTGACA	58	TCACAG	98	CTGCTA	138	AAGCAC	178	TACCAC
19	CTAGAC	59	ACAGTG	99	ATGAGG	139	CGAACA	179	CGCATA
20	AGCTCA	60	CGATTG	100	GGTAGA	140	ATGCTC	180	GCTAAG
21	ACTCGA	61	TCTTGC	101	ATCGTG	141	GGCTTA	181	ATCCAC
22	CTGTTG	62	GGTAAC	102	ATGGAC	142	ATCGCA	182	CCTTGA
23	CATGCA	63	TCATCC	103	AGTAGG	143	GGATCA	183	AGCGAA
24	CAGAAG	64	TAGGAC	104	CCATCA	144	CCAATC	184	GGTTAG
25	GTCTCA	65	TTCACC	105	AGTACC	145	AAGGTG	185	GATTGC
26	GTGATC	66	AACGAG	106	CCAGTA	146	ATCTCC	186	CGTTCA
27	TGTCTG	67	GTGGAA	107	CGTTAC	147	GTATCG	187	ATCCGA
28	GACAGA	68	ATGTCG	108	GAGGTA	148	TGTTCC	188	GCATTC
29	ACTCTG	69	ATCACG	109	TTGGCA	149	GGTGTA	189	CCTGAA
30	TGCAAC	70	GAATCC	110	CAATGC	150	TACTCC	190	GGAATC
31	GAAGGA	71	CGATGA	111	GCGTTA	151	ATCAGC	191	TCAACG
32	GTTGAG	72	GAATGG	112	TAGCTC	152	AAGTGC	192	AACACC
33	AGACCA	73	GCAACA	113	TTCGAC	153	AGGCTA		
34	TGGTTG	74	TTCTCG	114	GAGCAA	154	CCTATG		
35	GATCTG	75	ATTGCG	115	TTGCTG	155	TATCGC		
36	CTAGTG	76	TAGTGG	116	TTGCGA	156	CGCTAA		
37	CTCAGA	77	AAGCCA	117	GCAGAA	157	GTAACC		
38	CTTCGA	78	CTATCC	118	CCTACA	158	ACATGG		
39	AGCTAG	79	TCCGAA	119	GCATGA	159	CCGATA		
40	GATCGA	80	TGAACC	120	AACTGG	160	GGATAC		

a. Replace each '[6-nucleotide barcode]' in the RT primer sequence list in the key resources table with the barcode sequences given in Table 1.
b. Order the 192 CEL-seq2 RT primers in 96-well plates at 50 μM concentration. Keep a detailed record of the position of each barcode in each plate.
c. Order all oligonucleotides as ultramer DNA oligos with standard desalting.
d. Order the CEL-seq2 library RT primer, RPI, and RPI1 primer at 100 μM concentration.
e. Ensure that each random nucleotide in the CEL-Seq2 library RT primer has an equal proportion in each position. When ordering replace 'NNNNNN' in oligonucleotides listed in the key resources table with '(N:25252525)(N)(N)(N)(N)(N)'.

10. Dilute the 192 CEL-seq2 RT primers and prepare aliquots with a final working concentration of 1 μM to avoid repeated freeze-thaw cycles.
a. Thaw on ice the two 96-well plates containing 50 μM primers.

b. Spin plates for 30 s at ≤850 × *g* and place them back on ice.
c. As a starting point, make two sets of diluted primers using low-binding 96-well plates (DNA LoBind plates).
    i. Using a multi-channel pipette, transfer 10 μL of each 50 μM primer solution to the corresponding wells in a new plate containing 90 μL nuclease-free water to reach an intermediate concentration of 5 μM.
    ii. Mix by pipetting the entire volume up and down 5 times.
    iii. Spin plates for 30 s at ≤850 × *g* and place them on ice.
    iv. Transfer 10 μL of the 5 μM intermediate dilution to the corresponding wells in two new 96-well plates containing 40 μL nuclease-free water to reach a final concentration of 1 μM.
    v. Mix by pipetting the entire volume up and down 5 times.
    vi. Spin plates for 30 s at ≤850 × *g* and place them on ice.
d. Store undiluted and 5 μM primers at −80°C, and 1 μM primers at −20°C.

11. Dilute the CEL-seq2 library RT primer, RPI, and RPI1 primers to 10 μM and aliquot to avoid repeated freeze-thaw cycles.
    a. Thaw the 100 μM stock primers on ice.
    b. Spin tubes for 5 s at ≤2000 × *g* and place them back on ice.
    c. As a starting point, make 4 sets of 50 μL diluted primers using 1.5 mL low-binding tubes (DNA LoBind tubes).
        i. Using filter tips, transfer 20 μL of each 100 μM primer solution to new 1.5 tubes containing 180 μL nuclease-free water to reach 10 μM concentration.
        ii. Mix by pipetting the entire volume up and down 5 times.
        iii. Spin tubes for 5 s at ≤2000 × *g* and place them on ice.
        iv. Transfer 50 μL of each diluted primer to 3 other new 1.5 mL tubes.
        v. Spin tubes for 5 s at ≤2000 × *g* and place them on ice.
    d. Store undiluted primers at −80°C, and 10 μM aliquots at −20°C.

Note: Primer stocks can be stored at −80°C for several years. Avoid freeze/thaw cycles.

Note: Throughout this protocol two different types of plates will be used with the I.DOT, source plates and target plates. Source plates contain the solution to be dispensed, have a pore at the bottom of each well, and are used exclusively with the I.DOT. In this protocol, we used 96-well plates with a 100 μm pore as our source plate. Target plates are the recipient plates for solutions and/or reagents. They are generic. Here, we describe the steps using a 384-well plate as the target plate. Pipetting steps using the I.DOT need to be pre-programmed to ensure that the exact amount of liquid is transferred from the proper wells in the source plate into the correct wells in the target plate. You can find dispensing program specifications in the materials and equipment section.

12. Prepare working plates for RNA-seq.
    a. Prepare a serial dilution of ERCC RNA Spike-In Mix in nuclease-free water to reach a 1:100,000 final dilution. For example, add 1 μL ERCC RNA Spike-In Mix in 999 μL nuclease-free water to get a 1:1,000 dilution. Then, add 10 μL diluted ERCC RNA Spike-In (1:1,000) to 990 μL nuclease-free water to get a 1:100,000 dilution.
        i. To avoid unnecessary freeze-thaw cycles, prepare two 2 μL aliquots of the undiluted ERCC RNA Spike-In mix the first time you use it.
        ii. Make three 200 μL aliquots of 1:1,000 diluted mix for storage and discard the rest. Note the expiration date for the original reagent on the lid of the tube.
        iii. Store undiluted reagent and aliquots at −80°C. Any remaining diluted mix can be discarded.
    b. Prepare 3600 μL Spike-In-dNTPs master mix.
        i. Thaw reagents on ice, vortex for 5 s, and spin them for 5 s at ≤2000 × *g*.
        ii. Prepare Spike-In-dNTPs master as described in the table below.
        iii. Vortex for 5 s and spin it for 5 s at ≤2000 × *g*.

iv. Place master mix on ice.

Spike-In-dNTPs master mix	
Reagent	Amount
Nuclease-free water	2520 μL
ERCC RNA Spike-In Mix (1:100,000)	720 μL
dNTP Mix (10 mM/each)	360 μL
Total	3600 μL

c. Transfer Spike-In-dNTPs master mix into source plates.
   i. Pour the 3.6 mL Spike-In-dNTPs master mix into a multi-channel reagent reservoir.
   ii. Pipette 15 μL Spike-In-dNTPs master mix into each well of two new 96-well I.DOT source plates, using a multi-channel pipette.
   iii. Keep samples on ice during transfer. Use a clean empty 96-well plate as a holder for the source plates to avoid direct contact of the pore with the ice.
d. Transfer 3 μL of each 1 μM CEL-Seq2 RT primer into a different well of the two 96-well I.DOT source plates containing the Spike-In-dNTP mix. The resulting two 96-well plates are your barcoded-primer source plate #1 and #2 for the next steps.
   i. Pipette primers using a multichannel pipette. Change tips every time.
   ii. Mix by gently tapping plate on the side.
   iii. Be sure to keep a record of the position of each of the 192 barcoded primers in the two 96-well source plates.
   iv. Keep plates on ice during transfer.

Note: The 18 μL barcoded-primer mix in each well of the source plates is the minimum volume we recommend using. Lower volumes are more prone to produce errors when using the I.DOT in the next steps. While errors can be corrected for after the first dispensation, and the I.DOT software offers an automatic correction, this can be a time-consuming process. To address dispensing volume errors, see problem 1 in the troubleshooting section.

e. Using the I.DOT, dispense 240 nL of barcoded-primer mix from the source plates into a 384-well target plate.
   i. Run previously programmed instructions to dispense 240 nL of each barcoded-primer mix from source plate #1 (96 primers) into a different well of columns 1–12 rows A, C, E, G, I, K, M, and O of a 384-well plate.
   ii. Switch barcoded-primer source plate #1 with the plate #2.
   iii. Run previously programmed instructions to dispense 240 nL of each barcoded-primer mix from source plate #2 (remaining 96 primers) into a different well of columns 1–12 rows B, D, F, H, J, L, N, and P of the same 384-well plate. All wells in columns 13–24 will stay empty.
   iv. Keep a record of the position of each one of the 192 barcoded primers on the target plate.
   v. Keep source and target plates on ice before and after using the I.DOT.

Alternatives: Here we describe how to prepare plates for 192 samples. You can also prepare plates using all 384 wells, depending on how many embryos you plan to process in your experiments. Keep plates on ice in between steps to avoid evaporation.

f. Using the I.DOT, dispense 1.2 μL of Vapor-Lock into each well of the 384-well target plate that contains barcoded-primer mix.
   i. Place the 384-well target plate, containing the primer mix, on the target position and insert it in the I.DOT.
   ii. Place a new I.DOT source plate in the source position without inserting the deck into the instrument.

iii. Quickly pipette 30 μL Vapor-Lock into each well of source plate column 1 while on I.DOT deck and immediately insert it into the instrument.
iv. Run previously programmed instructions to dispense 1.2 μL Vapor-Lock into all 192 wells containing the primer mix.
v. Visually check that the volume is equal in all wells. If it appears that Vapor-Lock dispensing was not equal, consult problem 2 in the troubleshooting section.
vi. Keep target plate on ice before and after using the I.DOT.
vii. Seal target plate with a Microseal 'F' PCR seal.
viii. Spin plate for 30 s at ≤850 × *g*.
ix. Use immediately or store plate at −80°C until it is used as the working plate in step 13.

⚠ CRITICAL: Due to its low-viscosity Vapor-Lock can leak out of the 100 μm pores of the source plate. This can cause the formation of small droplets at the pores over time that lead to inaccurate dispensing. Therefore, Vapor-Lock dispensing with the I.DOT should be done as quickly as possible. For more details, see problem 2 in the troubleshooting section.

Alternatives: If more than one working plate is being prepared at the same time, we recommend dispensing the 240 nL barcoded-primer mix into all plates prior to adding Vapor-Lock. All plates should be kept on ice when not in the I.DOT and immediately used in step 13 or stored at −80°C.

Pause point: Working plate can be stored at −80°C for up to 6 months, but should only be thawed once, immediately before use.

13. Pipette RNA into working plate.
    a. Take dried RNA samples out of the freezer and place them on ice.
    b. Visually inspect tubes and ensure the blue RNA pellets are at the bottom of the tube. Otherwise, spin tubes down until pellets are located at the bottom of the tube.
    c. Add 2 μL nuclease-free water directly onto each of the blue RNA pellets.
    d. Incubate RNA samples at 60°C for 5 min in a heat block or water bath.
    e. Vortex RNA samples for 10 s, and spin them for 5 s at ≤2000 × *g*.
    f. Visually inspect RNA samples to ensure that pellets were solubilized.
        i. If necessary, mix by vortex again or pipette the entire volume up and down to aid solubilization.
    g. Remove working plate prepared in step 12 from the freezer and place it on ice.
    h. Using a Transferpette S (0.1–1 μL), transfer 120 nL of each of the 192 solubilized RNA samples into a different well of columns 1–12 of the working plate (wells containing barcoded-primer mix)
        i. Keep a record of the position of each RNA sample in the working plate.

Note: To avoid batch effects, make sure that an equal number of samples from all treatment groups and timepoints is included in each of the final libraries of 96 embryos.

    i. Maintain the RNA samples and plates on ice during all steps.
    j. Seal working plate with a Microseal 'F' PCR seal and spin for 30 s at ≤850 × *g*.
    k. Store leftover RNA at −80°C.

Pause point: The working plate, containing the RNA samples, can be stored at −20°C for up to 1 week.

14. Carry out first-strand synthesis (FSS).
    a. Denature RNA secondary structures.
        i. Remove working plate prepared in step 13 from freezer (if needed).
        ii. Spin for 30 s at ≤850 × *g*.

iii. Incubate at 95°C for 3 min using a thermocycler (lid temperature 105°C). After incubation, immediately chill on ice for at least 5 min.
iv. Spin for 30 s at ≤850 × *g* and place it on ice.

b. Thaw reagents and prepare 56 µL FSS reaction master mix in a 1.5 mL tube according to the table below.

FSS master mix	
Reagent	Amount
5× First-strand buffer	28 µL
DTT (100 mM)	14 µL
RNaseOUT (40 U/µL)	7 µL
Superscript II (200 U/µL)	7 µL
Total	56 µL

c. Vortex FSS reaction master mix for 5 s and spin tube for 5 s at ≤2000 × *g*.
d. Pipette 25 µL FSS master mix into both well A2 and well B2 of a source plate, and insert the plate into the I.DOT source plate position.
e. Unseal the working plate, containing the denatured RNA, and insert it into the I.DOT target plate position.
f. Run previously programmed instructions to dispense 160 nL FSS master mix into each well in columns 1–12 in the working plate.
g. Seal working plate with a Microseal 'B' PCR seal and mix by gently tapping plate on the side.
h. Spin working plate for 30 s at ≤850 × *g*.
i. Incubate working plate in a thermocycler with lid temperature set at 85°C and perform FSS as follows.

FSS reaction conditions		
Steps	Temperature	Time
Reverse transcription	42°C	1 h
Inactivation	70°C	10 min
Hold	4°C	forever

j. Spin working plate for 30 s at ≤850 × *g* and place it on ice.

15. Carry out second strand synthesis (SSS).
a. Thaw reagents and prepare 463.5 µL SSS master mix in a 1.5 mL tube.

SSS master mix	
Reagent	Amount
Nuclease-free water	324 µL
5× Second strand buffer	106 µL
dNTP Mix (10 µM/each)	10.5 µL
E. coli DNA Pol. I (10 U/µL)	15 µL
E. coli DNA ligase (10 U/µL)	4 µL
E. coli RNase H (10 U/µL)	4 µL
Total	463.5 µL

b. Vortex SSS reaction master mix for 5 s and spin tube for 5 s at ≤2000 × *g*.
c. Pipette 55 µL SSS master mix into each well in column 3 of the source plate and insert it into the I.DOT source plate position.
d. Unseal the working plate after FSS and insert it into the I.DOT target plate position.
e. Run the previously programmed instructions to dispense 2.2 µL SSS into each well in columns 1–12 of the working plate.

f. Seal working plate with a Microseal 'B' PCR seal, mix by gently tapping the side of the plate, and spin for 30 s at ≤850 × *g*.
g. Incubate the working plate in a thermocycler with unheated lid, and perform SSS as follows.

SSS reaction conditions

Steps	Temperature	Time
Second strand synthesis	16°C	2 h
Hold	4°C	forever

h. Remove the working plate from the thermocycler and place on ice.

16. Pool 96 samples.
    a. Prepare AMPure XP reagent for further steps.
        i. Take AMPure XP reagent from refrigerator (4°C).
        ii. Mix by shaking and/or vortex until reagent is homogenous.
        iii. Make two aliquots, one of 250 μL and another of 500 μL using 1.5 mL tubes. This volume is enough for 2 libraries (192 samples).
        iv. Store the 250 μL aliquot and AMPure XP reagent bottle at 4°C.
        v. Keep the 500 μL aliquot at room temperature for at least 20 min before its use.
    b. Spin the working plate from step 15 for 30 s at ≤850 × *g*.
    c. Remove the seal and keep the plate on ice.
    d. Transfer 96 samples from columns 1–6 into a single 1.5 mL low-binding tube. This pool 1 will later become your Library 1.
        i. Use filter tips for transferring samples. There is no need to change the tips unless they get contaminated (e.g., touching surfaces other than wells in columns 1–6).
    e. Transfer 96 samples from columns 7–12 into a single 1.5 mL low-binding tube. This pool 2 will later become your Library 2.
        i. Use filter tips for transferring samples. There is no need to change the tips unless they get contaminated (e.g., touching surfaces other than wells in columns 7–12).

    Alternatives: When preparing more than 192 samples, process all libraries up to this point before continuing with the protocol.

    f. Centrifuge pool 1 and 2 tubes at ≥ 16,000 × *g* for 1 min.
    g. Transfer the lower aqueous phase from each pool into new 1.5 mL tubes.
        i. Press down the plunger while moving the pipette tip to the bottom of the tube to avoid pulling up any of the upper Vapor-Lock layer.
        ii. Repeat centrifugation and transfer, if Vapor-Lock was accidentally transferred.

    ⏸ **Pause point:** Pooled samples can be stored at −20°C for up to 2 weeks.

17. Clean up cDNA using AMPure XP reagent.
    a. Measure the total volume recovered in pools 1 and 2 (approximately 220 μL is expected for 96-samples).
    b. Split the volume of each pool into 2 equal parts (pool 1A, 1B, 2A, and 2B) and transfer into 4 different wells of a low-binding 96-well plate.
        i. Change tips between each pool.
    c. Vortex the room temperature AMPure XP reagent until homogeneous.
    d. Add AMPure XP reagent at a volume correspondent to 0.8× sample volume. For example, for a 110 μL pool volume, add 88 μL AMPure XP reagent to each pool.
    e. Mix AMPure XP reagent with samples by pipetting up and down until homogenous.
    f. Incubate plate on the bench for 10 min at room temperature.
    g. Place plate on a 96-well side magnetic stand (DynaMag-96) and incubate for 10 min at room temperature.

i. Prepare 2 mL 80% ethanol in nuclease-free water while incubating samples.

h. Perform ethanol washes.
   i. With plate on magnetic stand, remove and discard all liquid without disturbing the beads (brown pellet on the side). The cDNA is attached to the beads.
   ii. Add 150 µL freshly prepared 80% ethanol to each sample well.
   iii. Incubate for 30 s and remove ethanol without disturbing the beads.
   iv. Repeat ethanol wash a second time.
   v. Air-dry samples until completely dry (approximately 5–10 min).

i. Remove plate from magnet.

j. Elute cDNA
   i. Resuspend beads in pool 1A with 14 µL nuclease-free water and mix thoroughly by pipetting up and down until homogeneous.
   ii. Transfer the 14 µL bead mix from pool 1A into the well with pool 1B and mix thoroughly by pipetting up and down until homogeneous. The total volume should be ~14 µL.
   iii. Repeat steps i and ii for pools 2A and 2B.
   iv. Incubate both pooled samples for 2 min at room temperature.
   v. Place plate back on the magnet and incubate for at least 5 min at room temperature.
   vi. With the plate on the magnet, transfer the pool 1 and pool 2 liquid containing cDNA (~14 µL/each) into two 1.5 mL low-binding tubes labeled pool 1 and pool 2.
   vii. Place samples on ice.

18. Perform *in vitro* transcription (IVT).

a. Thaw reagents and prepare 43.2 µL IVT master mix in a 1.5 mL tube.

IVT master mix

Reagent	Amount
ATP (75 mM)	7.2 µL
GTP (75 mM)	7.2 µL
UTP (75 mM)	7.2 µL
CTP (75 mM)	7.2 µL
10× T7 reaction buffer	7.2 µL
T7 enzyme mix	7.2 µL
Total	43.2 µL

b. Vortex IVT master mix for 5 s, spin tube for 5 s at ≤2000 × *g*, and place it on ice.

c. Pipette 19 µL IVT master mix into 2 wells of a low-binding 96-well plate.

d. Add ~14 µL cDNA from each pool to a well containing the IVT mix.

e. Mix by gently pipetting up and down 5 times.

f. Seal with a Microseal 'B' PCR seal and spin plate for 1 min at ≤850 × *g*.

g. Incubate the samples in a thermocycler with lid temperature set at 70°C and run the following program.

IVT reaction conditions

Steps	Temperature	Time
In vitro transcription	37°C	16 h
Hold	4°C	forever

h. Remove the plate containing the amplified RNA (aRNA) from the thermocycler (aRNA 1 from pool 1 and aRNA 2 from pool 2).

i. Spin the plate for 30 s at ≤850 × *g* and place it on ice.

19. Prepare reagents.

a. Thaw ExoSAP-IT reagent (Exonuclease I and shrimp alkaline phosphatase) and the NEBNext Magnesium RNA Fragmentation Module on ice. Keep them on ice until use.

b. Vortex the RNAClean XP beads and pipette 150 µL into a new 1.5 mL tube.

c. Equilibrate the RNAClean XP aliquot at room temperature for at least 20 min before use.

20. Treat samples with ExoSAP-IT.
    a. Vortex the thawed ExoSAP-IT for 5 s and spin reagent for 5 s at ≤2000 × *g*.
    b. Add 12 μL ExoSAP-IT reagent to each of the two aRNA samples and mix by gently pipetting up and down 5 times.
    c. Spin plate for 30 s at ≤850 × *g*.
    d. Incubate plate in a thermocycler with lid temperature set at 50°C and run the following program.

ExoSAP-IT reaction conditions

Steps	Temperature	Time
Treatment	37°C	15 min
Hold	4°C	forever

    e. Remove the plate from the thermocycler and incubate it on ice for at least 5 min.
    f. Spin plate for 30 s at ≤850 × *g* and put it back on ice.

21. Treat samples with NEBNext Magnesium RNA Fragmentation Module.
    a. Vortex the thawed fragmentation buffer and stop solution, included in the kit, for 5 s.
    b. Spin reagents for 5 s at ≤2000 × *g* and put them back on ice.

    Note: The Fragmentation Module stop solution can precipitate during storage. Mix until completely dissolved.

    c. Add 5 μL fragmentation buffer to both aRNA samples and mix by gently pipetting up and down 5 times. Keep samples on ice.
    d. Seal plate with a Microseal 'B' PCR seal and spin for 5 s at ≤2000 × *g*.
    e. Incubate plate in a thermocycler with lid temperature set at 105°C and run the following program.

Fragmentation reaction conditions

Steps	Temperature	Time
Fragmentation	94°C	3 min
Hold	4°C	forever

    f. Quickly transfer plate to ice, after 3 min at 94°C, and cool the top of the plate with ice, as well. The thermocycler lid will be hot, use caution.
    g. Cool the plate on ice for 30 s.
    h. Spin plate for 30 s at ≤850 and place it back on ice.
    i. Immediately add 5 μL stop solution and mix by gently pipetting the entire volume (~54 μL) up and down 5 times. Keep plate on ice during this step.

    Note: Fragmentation of aRNA should be performed for exactly 3 min at 94°C to avoid under or over fragmentation. For consistent results, we recommend putting the plate into the thermocycler after the lid is already hot (105°C) and removing the plate immediately after the 3 min incubation time is completed.

22. Clean up aRNA using RNAClean XP.
    a. Measure the total volume of each sample after fragmentation. The volume should be approximately 54 μL.
    b. Vortex the 150 μL RNAClean XP aliquot, previously equilibrated at room temperature (step 19), until homogeneous.
    c. Add RNAClean XP to each sample at a volume corresponding to 0.8× sample volume. For example, for a 50 μL sample add 40 μL RNAClean XP.

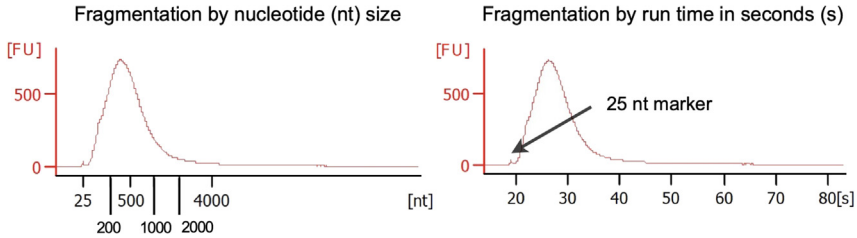


Figure 3. Expected size distribution after fragmentation of aRNA
FU, height threshold.

d. Mix samples by pipetting up and down until homogeneous.
e. Incubate samples for 10 min at room temperature.
f. Place the plate on a 96-well side magnet (DynaMag-96) and incubate for 10 min at room temperature.
   i. Prepare 1 mL 70% ethanol in nuclease-free water while incubating samples.
g. Perform ethanol washes.
   i. With plate on magnetic stand, remove and discard all liquid without disturbing the beads (brown pellet on the side). The aRNA is attached to the beads.
   ii. Add 150 µL freshly prepared 70% ethanol to each sample well.
   iii. Incubate for 30 s and remove ethanol without disturbing the beads.
   iv. Repeat steps ii and iii two more times.
   v. Air-dry sample pellets until completely dry (approximately 5–10 min).
h. Remove plate from magnet.
i. Elute aRNA.
   i. Resuspend beads from each sample in 22 µL nuclease-free water.
   ii. Mix thoroughly by pipetting up and down until homogeneous.
   iii. Incubate samples for 2 min at room temperature.
   iv. Place plate back on the side magnet and incubate for at least 5 min.
   v. Transfer each sample (~22 µL of aRNA 1 and aRNA 2) separately into two new labeled 1.5 mL low-binding tubes.
   vi. Place samples on ice.

23. Check fragment size distribution of the samples.
   a. Use the Agilent 2100 Bioanalyzer and Agilent RNA 6000 Pico Kit, according to manufacturer instructions (https://www.agilent.com/cs/library/usermanuals/Public/G2938-90046_RNA600Pico_KG_EN.pdf).
   b. Ensure that each sample has a peak between 200–1000 nucleotides. See Figure 3 for an example.
   c. If peak is present, proceed with protocol. Otherwise, refer to problem 3 in the troubleshooting section.

Note: If the ladder fails and fragment size cannot be directly determined, it is sufficient to look for a peak between 20–30 s on the time scale. See Figure 3 for an example.

Pause point: aRNA samples can be stored at −20°C for up to 1 week.

24. Perform FSS on aRNA samples.
   a. Prepare reagents.
      i. Thaw the 10 µM CEL-seq2 library RT primer, 10 µM dNTPs, and aRNA samples (if needed) on ice.
      ii. Vortex reagents for 5 s, spin them for 5 s at ≤2000 × *g*, and place them on ice.
   b. Denature RNA secondary structures.
      i. Add 4 µL 10 µM CEL-seq2 library RT primer into 2 wells of a low-binding 96-well plate.

ii. Add 2 μL 10 μM dNTPs to the same wells.
iii. Add 20 μL aRNA 1 to one of the wells containing the primer-dNTPs mix.
iv. Add 20 μL aRNA 2 to the other well containing the primer-dNTPs mix.
v. Incubate the plate at 65°C for 5 min using a thermocycler (lid temperature 80°C). After incubation, quickly chill on ice for at least 5 min.
vi. Spin working plate for 30 s at ≤850 × *g* and place it back on ice.

c. Thaw the listed reagents on ice and prepare 40 μL FSS master mix in a 1.5 mL tube according to the table below.

FSS master mix for aRNA samples

Reagent	Amount
5× First-strand buffer	20 μL
DTT (100 mM)	10 μL
RNaseOUT (40 U/μL)	5 μL
Superscript II (200 U/μL)	5 μL
Total	40 μL

d. Vortex FSS master mix for 5 s and spin tube for 5 s at ≤2000 × *g*.
e. Pipette 16 μL FSS master mix into the two wells containing the primer-dNTPs-aRNA mix.
f. Mix by gently pipetting the entire volume (42 μL) up and down 5 times.
g. Seal the plate with a Microseal 'B' PCR seal.
h. Place the plate in a thermocycler with lid at 50°C and run the following program.

FSS reaction conditions for aRNA samples

Steps	Temperature	Time
Incubation	25°C	10 min
Reverse transcription	42°C	1 h
Hold	4°C	forever

i. Remove plate from thermocycler.
j. Spin plate for 30 s at ≤850 × *g* and place it on ice.
k. Transfer each sample (~42 μL cDNA 1 and cDNA 2) separately into two new 1.5 mL low-binding tubes and place them on ice.

25. Perform a 20-fold dilution of cDNA.
    a. For each of the two cDNA samples, add 1 μL cDNA to 19 μL nuclease-free water in new 1.5 mL low-binding tubes.
    b. Vortex samples for 5 s and spin them for 5 s at ≤2000 × *g*.
    c. Place and keep the diluted cDNA samples on ice.
    d. Store undiluted cDNA samples at −80°C.
26. Perform PCR on diluted cDNA
    a. Thaw reagents on ice and prepare library PCR master mix in a 1.5 mL tube, according to the table below. Keep reagents on ice until use.

Library PCR master mix

Reagent	Amount
Nuclease-free water	25.3 μL
2× Phusion High-Fidelity PCR Master Mix with HF Buffer	57.5 μL
RPI primer (10 μM)	4.6 μL
RPI1 primer (10 μM)	4.6 μL
Total	92 μL

b. Vortex library PCR master mix for 5 s and spin tube for 5 s at ≤2000 × *g*.

c. Pipette 10 μL of each diluted cDNA sample (cDNA 1 and cDNA 2) into different wells on a low-binding 96-well plate. Keep plate on ice.
d. Add 40 μL PCR reaction master mix into the same two wells as the samples.
e. Mix by gently pipetting the entire volume (50 μL) up and down 5 times.
f. Incubate plate in a thermocycler with lid temperature set at 105°C and run the following protocol.

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	10 s	11 cycles
Annealing	60°C	30 s	
Extension	72°C	30 s	
Final extension	72°C	10 min	1
Hold	4°C	Forever	

Alternatives: Indexes and barcodes allow you to identify each individual sample after sequencing. When preparing more than two libraries, use a different RPI index for every two libraries (192 samples). For example, to prepare 4 libraries, use two different RPI indexes.

27. Clean up DNA libraries using AMPure XP reagent.
    a. Equilibrate the 250 μL AMPure XP aliquot (prepared in step 16) at room temperature for 20 min.
    b. Remove the plate (step 26) from the thermocycler.
    c. Spin plate for 30 s at ≤850 × *g* and place it on ice.
    d. Measure the total volume recovered of each sample after PCR (should be ~50 μL).
    e. Vortex the equilibrated AMPure XP until homogeneous.
    f. Add AMPure XP reagent at a volume corresponding to 0.8× of the sample volume. For example, for a 50 μL sample add 40 μL AMPure XP reagent.
    g. Pipette samples up and down until homogenous.
    h. Incubate samples at room temperature for 10 min.
    i. Place the plate on a side magnet (DynaMag-96) and incubate for 10 min at room temperature.
        i. Prepare 1 mL of 80% ethanol in nuclease-free water while incubating samples.
    j. Perform ethanol washes.
        i. With plate on magnetic stand remove and discard all liquid without disturbing the beads (brown pellet on the side). Amplicons are attached to the beads.
        ii. Add 150 μL freshly prepared 80% ethanol to each sample.
        iii. Incubate for 30 s and remove ethanol without disturbing the beads.
        iv. Repeat steps ii and iii two more times.
        v. Air-dry sample pellets until completely dry (approximately 5–10 min).
    k. Remove the plate from the magnet.
    l. Elute libraries.
        i. Resuspend beads with 25 μL nuclease-free water.
        ii. Incubate samples for 2 min at room temperature.
        iii. Place the plate on the magnet and incubate for 5 min at room temperature.
        iv. Transfer ~25 μL from each of the 2 libraries to new wells on the same plate.
    m. Carry out a second clean up.
        i. Vortex equilibrated AMPure XP until homogeneous.
        ii. Add AMPure XP reagent at a volume correspondent to 0.8× sample volume. For example, if 25 μL were successfully recovered, add 20 μL to each library.
        iii. Repeat steps 27.d to 27.h (see above).
    n. Elute libraries.

i. Resuspend dried beads in 11 μL nuclease-free water.
ii. Incubate samples for 2 min at room temperature.
iii. Place the plate on the magnet and incubate for 5 min at room temperature.
iv. Transfer ~11 μL of each of the 2 libraries to two separate labeled, 1.5 mL low-binding tubes.
v. Place libraries on ice.

28. Measure libraries DNA concentration.
    a. Use the Qubit Fluorometer and Qubit dsDNA HS Assay Kit, according to manufacturer instructions (https://www.thermofisher.com/order/catalog/product/Q32851).
    b. Concentration should be approximately 7–9 ng/μL per library. Consult problem 4 in the troubleshooting section if your sample concentration is lower than this.
    c. Keep samples on ice during the process.
    d. Store the samples at −80°C.

Pause point: the libraries can be stored at −80° for up to 6 months before sequencing.

Note: DNA concentration in each library is dependent on the cDNA concentration and the number of PCR cycles. However, adding too much cDNA can lead to a reduced PCR yield and lower detection of transcripts. A 20-fold cDNA dilution and 11 PCR cycles worked well for 96 Drosophila embryos. Modifications may be needed if more or less embryos are being analyzed.

29. Perform paired-end sequencing of DNA libraries.
    a. Use the Illumina NovaSeq 6000 sequencer and correspondent flow cells.
    b. CEL-Seq2 libraries should be less than 20% of the pool to ensure base diversity during sequencing.
        i. Pool libraries onto lanes with other diverse libraries prepared without CEL-Seq2 primers and different Illumina indexes.
        OR.
        ii. Spike in no less than 40% PhiX Control (Illumina In., San Diego, CA USA)
    c. Sequence 2 × 150 bp, plus a 6 bp i7 index.
    d. Sequence each library at a depth of 3 M reads per sample or higher. For example, you should aim for 288 M reads for a library containing 96 samples.

Note: Illumina sequencers use the first 25 bases to perform % pass filter calculations and other cluster metrics. Due to the poly-T region after 15 bp in Read 1, the base diversity must be increased or the clusters will not register and the sequencing run can fail. Adequate base diversity can be easily achieved by loading CEL-Seq2 libraries with other diverse libraries, as indicated above.

Alternatives: Perform a custom read length run and sequence only 15 bp in Read 1; this ensures base diversity by avoiding the poly-T region. A short Read 1 is not supported by Illumina; the demultiplexing requires a special command (–mask-short-adapter-reads 0) in bcl2fastq in order for the reads to process properly.

Sequencing data processing

Timing: 1 day

This major step describes RNA-seq data alignment and demultiplexing using STAR (Spliced Transcripts Alignment to a Reference)[4] and the CEL-Seq2 barcode sequences, respectively. This step creates a table with the deduplicated reads per gene per barcode (embryo). The code for each step can be found here, on Zenodo, and on our GitHub repository.

Note: When doing read alignment and processing to obtain a read counts matrix, working on a Linux-based system with >100Gb of memory and >16 available cores, such as is generally found on a High Performance Computing (HPC) cluster, is recommended.

30. Align sequencing data to reference genome using STAR.
    a. Index a reference genome.

Note: Before STAR can be used to align reads to a reference genome, the genome must first be indexed. Genome indexing needs to be done only once for a given genome version. Here we download the BDGP6.28 (dm6) reference genome and gene annotations.

```
> wget ftp://ftp.ensembl.org/pub/release-100/fasta/drosophila_melanogaster/dna/Drosophila_melanogaster.BDGP6.28.dna.toplevel.fa.gz> wget ftp://ftp.ensembl.org/pub/release-100/gtf/drosophila_melanogaster/Drosophila_melanogaster.BDGP6.28.100.gtf.gz
```

    b. Decompress the fasta and gtf files.

```
> gunzip Drosophila_melanogaster.BDGP6.28.dna.toplevel.fa.gz

> gunzip Drosophila_melanogaster.BDGP6.28.100.gtf .gz
```

    c. Run STAR.
        i. Here, the output folder is specified as 'dm6_star_index' but it can be named to any unused filename.

```
> STAR --runMode genomeGenerate --runThreadN 4 --outFileNamePrefix ./dm6_star_index/ --genomeDir ./dm6_star_index/ --genomeFastaFiles Drosophila_melanogaster.BDGP6.28.dna.toplevel.fa --sjdbGTFfile Drosophila_melanogaster.BDGP6.28.100.gtf --sjdbOverhang 100
```

Note: The value for '–runThreadN' should be adjusted based on the number of available cores/threads. A higher value can reduce runtime. The optimal value will depend on the specific computational system; a good starting point is 4 or 8.

Note: The ideal value for '–sjdbOverhang' is 1 bp less than the read length. However, a value of 100 will work well across a range of read lengths.

31. Download the barcode sequences.

```
> wget https://raw.githubusercontent.com/vari-bbc/scRNAseq/main/whitelists/cellseq192/celseq_barcodes.192.1col.txt
```

Optional: Download our sample dataset fastq files to replicate our analysis.[1] This is not necessary if you already have your own fastq files.

```
> wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR216/059/SRR21697059/SRR21697059_1.fastq.gz

> wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR216/059/SRR21697059/SRR21697059_2.fastq.gz

> wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR216/060/SRR21697060/SRR21697060_1.fastq.gz

> wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR216/060/SRR21697060/SRR21697060_2.fastq.gz
```

32. Align the reads to the indexed reference genome and count deduplicated reads per gene for each barcode.

a. The code below aligns a library (run SRR21697060) from Gene Expression Omnibus (GEO) sample GSM6599295, specified using the '–readFilesIn' parameter, to the genome index created in step 30. The file specified using the '–soloCBwhitelist' parameter is downloaded from the previous step. When analyzing your own data, the names for the Read 2 and Read 1 files will have to be changed accordingly.

```
> STAR --runThreadN 16 --limitBAMsortRAM 137438953472 --genomeDir ./dm6_star_index/
--readFilesIn SRR21697060_2.fastq.gz SRR21697060_1.fastq.gz --outSAMattributes NH HI AS nM
CB UB --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --outFileNamePrefix STARso-
lo_out/sample1. --soloType CB_UMI_Simple --soloCBwhitelist celseq_barcodes.192.1col.txt
--soloUMIstart 1 --soloUMIlen 6 --soloCBstart 7 --soloCBlen 6 --soloBarcodeReadLength
0 --soloCBmatchWLtype Exact --soloStrand Forward --soloUMIdedup Exact
```

Note: Here, the '–limitBAMsortRAM' parameter is requesting ~137 Gb of memory. This parameter is sometimes needed to avoid memory issues at the BAM sorting stage in STAR. Alternatively, one could try omitting this option and not sorting the output BAM by setting '–outSAMtype BAM Unsorted' instead of '–outSAMtype BAM SortedByCoordinate'.

Note: The order of the files for the '–readFilesIn' parameter must be Read 2 then Read 1.

Note: The parameters, '–soloUMIstart 1 –soloUMIlen 6 –soloCBstart 7 –soloCBlen 6', tell STAR where the barcode and UMI are located on each read. '–soloCBmatchWLtype Exact' and '–soloUMIdedup Exact' request exact sequence matches for matching barcodes with the whitelist and deduplicating UMIs, respectively.

b. Repeat previous step for your second library (sample 2).
   i. Change the '–readFilesIn' parameter to match the file names of Read 2 and Read 1 of your second library.
   ii. Change the '–outFileNamePrefix' parameter to 'STARsolo_out/sample2.'

Optional: In our sample dataset, run SRR21697059 from GEO GSM6599296 is our library 2.

33. Output the read counts into a simple table.
    a. The output of STAR will be in the style of 10× CellRanger. Thus, downstream tools designed to process 10× CellRanger data should be readily adaptable to these data. The R code below demonstrates the use of the DropletUtils package to import the counts and output a simple counts table with the columns representing each barcoded-embryo (renamed from the barcode sequence to the sample number based on the table from https://raw.githubusercontent.com/vari-bbc/scRNAseq/main/whitelists/cellseq192/celseq_barcodes.192.txt) and the rows representing genes.[5–7]

```
> library(readr)
> library(dplyr)
> library(tibble)
> library(DropletUtils)
> library(scuttle)
>
> decoder_file <- "celseq_barcodes.192.txt"
> samples <- c("sample1", "sample2")
```

```
>
> dec <- read_tsv(decoder_file, col_names = c("bc_num","bc"))
>
> starsolo_to_table <- function(starsolo_dir, decoder){
>
>   sce <- read10xCounts(starsolo_dir, col.names=TRUE)
>   rownames(sce) <- uniquifyFeatureNames(rownames(sce), rowData(sce)$Symbol)
>   counts <- as.matrix(counts(sce))
>
>   # rename columns with bc_num
>   colnames(counts) <- decoder$bc_num[match(colnames(counts), decoder$bc)]
>
>   tibble::as_tibble(counts, rownames="Gene") %>%
>     dplyr::select(Gene, as.character(sort(as.integer(colnames(counts)))))
> }
> for (i in 1:length(samples)){
>   dir <- paste0("STARsolo_out/", samples[i], ".Solo.out/Gene/raw/")
>   out_file <- paste0(samples[i], "_counts.tsv")
>
>   df <- starsolo_to_table(dir, dec)
>   write_tsv(df, out_file)
> }
```

Note: At the end of major step 3 you will have created two files, sample1_counts.tsv and sample2_counts.tsv. Each file will have the raw read counts for each individual embryo in each library. These raw reads counts are deduplicated reads per gene per barcode.

Continuous transcriptome analysis

Timing: 1 day

This major step describes normalization of raw reads, sample quality control, continuous data analysis, and differential expression analysis using RaceID.[8] This R package allows for the pseudo-temporal ordering of transcriptomes along an inferred trajectory of differentiation. For our experimental conditions, the pseudo-temporal order (or pseudo-time) indicates the embryonic developmental trajectory and replaces visual embryo staging. At the end of these steps, each embryo will be assigned a position on a developmental timeline according to its transcriptome. The code for each step can be found here, on Zenodo, and on our GitHub repository.

34. Install software. The specific version for each software/package used to analyze our sample dataset is found in the key resources table.
    a. Install R and RStudio, if necessary.
    b. Install the RaceID R package. The latest version of this package can be found at https://github.com/dgrun/RaceID3_StemID2_package.

c. Ensure the R package is installed correctly by running the sample data.
   i. Run function *vignette* in R to get the step-by-step instructions for running the sample data (intestinalData).

```
> vignette("RaceID")
```

35. Carry out analysis on RNA-seq data.
   a. Get the raw read counts. These files were generated in the major step 3 (steps 30–33). The same files can be downloaded directly from GEO: GSE214118 at the following link: https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE214118&format=file
      i. Rename files if downloaded from the link provided. Rename 'GSM6599295_Sample1.STARsolo_raw.counts.txt' file to 'sample1_counts.tsv' and 'GSM6599295_Sample2.STARsolo_raw.counts.txt' file to 'sample2_counts.tsv'.
   b. Read the two raw read count files (96 embryos each) and create a single data frame.

```
> library1 <-read.csv("sample1_counts.tsv", sep="\t", header=TRUE, row.names = 1)
>
> library2 <-read.csv("sample2_counts.tsv", sep="\t", header=TRUE, row.names = 1)
>
> data <- cbind(library1[,1:96], library2[,97:192])
```

   c. Analyze data using the RaceID workflow.

Note: The code below filters out samples with < 250,000 total read counts and transcripts with < 3 read counts in < 5 embryos. We recommend using these parameters initially and then adjusting them according to your specific dataset. For a description of each specific function, see the RaceID manual.

```
> library(RaceID)
> sc <- SCseq(data)
> sc <- filterdata(sc, minexpr = 3, minnumber = 5, LBatch = NULL, mintotal=250000)
> sc <- compdist(sc,metric="spearman", FSelect = FALSE,knn = NULL,alpha = 3)
> sc <- clustexp(sc, rseed = 12345, samp = 1000 , FUNcluster = "kmedoids")
> sc <- findoutliers(sc, probthr = 0.001, outlg = 3, outminc = 5)
```

36. Identify unfertilized eggs in dataset.
   a. Plot t-SNE maps and export them as a pdf file.
      i. Plot the log transformed expression of *scw*, *sc*, and *esg*. These transcripts are associated with the minor wave of zygotic genomeactivation.[1,11]
      ii. Plot the log transformed expression of *pnr*, *ths*, and *tin*. These transcripts are associated with the major wave of zygotic genome activation.[1,12]
      iii. Plot the labels map.

```
> pdf(file = "01tsne_maps.pdf",width = 11, height = 7.5)
> sc <- comptsne(sc,perplexity = 16, rseed = 420)
> plotmap(sc,cex=3)
```

```
> plotexpmap(sc, g="scw", n="scw", logsc = TRUE, cex = 3)

> plotexpmap(sc, g="sc", n="sc", logsc = TRUE, cex = 3)

> plotexpmap(sc, g="esg", n="esg", logsc = TRUE, cex = 3)

> plotexpmap(sc, g="pnr", n="pnr", logsc = TRUE, cex = 3)

> plotexpmap(sc, g="ths", n="ths", logsc = TRUE, cex = 3)

> plotexpmap(sc, g="tin", n="tin", logsc = TRUE, cex = 3)

> plotlabelsmap(sc, cex = 0.2)

> dev.off()
```

b. Determine developmental trajectory.
   i. Identify samples not expressing *scw*, *sc*, *esg*, *pnr*, *ths*, and *tin* in the t-SNE map. Those samples represent embryos without evidence of zygotic transcription (i.e., < 1 h into development).
   ii. Identify samples showing increasing levels of expression of early expressed genes such as *scw*, *sc*, and *esg* in the t-SNE map. Those embryos are approximately > 1 h into development.
   iii. Identify samples showing increasing levels of expression of *pnr*, *ths*, and *tin* in the t-SNE map, genes expressed during the major ZGA. Those embryos are approximately > 2 h into development.
   iv. Determine the direction of the developmental trajectory by comparing the groups identified. For example, the embryo development increases from left to right and from up to down in the sample dataset, as shown by the expression pattern of *scw* and *ths* in Figure 4.

c. Identify unfertilized embryos by visual inspection of the t-SNE maps.
   i. Identify any samples that have no or very low expression of *scw*, *sc*, and *esg*, but that are in the midst of embryos that are expressing these genes. These samples are unfertilized eggs. An example is shown in Figure 5.
   ii. Use the labels map to get the sample ID for the unfertilized eggs that are identified (Figure 5).

Note: Samples with a similar transcriptome cluster together in a t-SNE map. In our data, the extensive transcriptional changes during development led samples to assemble into a linear pattern. At the beginning of development, embryos and unfertilized eggs follow the same linear pattern in the t-SNE map because maternal mRNA decay is the dominant process. The outlined approach only identifies unfertilized eggs once transcription starts.

Note: It is possible to miss some unfertilized eggs due to overlap with other samples in the t-SNE map. Be sure to view the map in different layouts to ensure that all unfertilized eggs in the dataset are identified.

Note: Sample IDs equal the CEL-Seq2 barcode numbers in our sample dataset. Note that a letter 'X' is added as a prefix when importing raw read count files into R. For example, sample ID '1' becomes 'X1'.

37. Identify embryos that are older than 3 h in the dataset.
    a. Plot a t-SNE map with the combined linear expression of *pnr*, *ths*, and *tin*.

```
> pdf(file = "02pnr_ths_tin.pdf",width = 11, height = 7.5)

> plotexpmap(sc, g=c("pnr","ths","tin"), n="pnr+ths+tin", logsc = FALSE, cex = 3)

> dev.off()
```

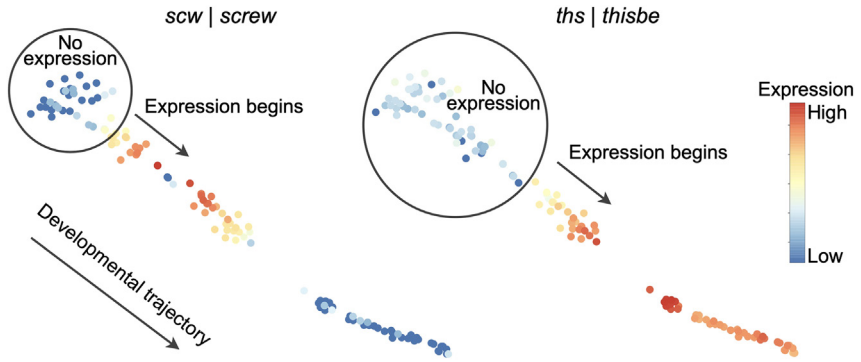


Figure 4. Gene expression levels of *scw* and *ths* for each embryo on the t-SNE map
Gene expression patterns of candidate genes are used for visual identification of the developmental trajectory.

b. Export the sample IDs from each cluster into a table.

```
> write.csv(sc@cpart, file = "03sampleid_by_cluster.csv")
```

c. Identify samples after the highest expression of *pnr*, *ths*, and *tin* in the t-SNE map. These are embryos that are older than 3 h (Figure 6).
   i. Use the "03sampleid_by_cluster.scv" file to get Sample IDs.

Note: Female flies can withhold embryos. Any *Drosophila* dataset will have embryos that are older than the defined collection time interval chosen.

38. Exclude unfertilized eggs and embryos older than 3 h from dataset and re-run RaceID workflow.

```
# The first 5 are the identified unfertilized eggs

> exclude <- c("X75","X132","X136","X141","X166","X1","X4","X9","X10","X25","X30","X31",
"X32","X46","X50","X51","X55","X72","X77","X80","X85","X94","X95","X102","X108","X124",
"X125","X127","X130","X153","X158","X171","185","X188")

> embryos_3h <- data[,!(names(data) %in% exclude)]

>

> sc <- SCseq(embryos_3h)

> sc <- filterdata(sc, minexpr = 3, minnumber = 5, LBatch = NULL, mintotal=250000)

> sc <- compdist(sc,metric="spearman", FSelect = FALSE,knn = NULL,alpha = 3)

> sc <- clustexp(sc, rseed = 12345, samp = 1000 , FUNcluster = "kmedoids")

> sc <- findoutliers(sc, probthr = 0.001, outlg = 3, outminc = 5)
```

39. Perform quality assessment (Figure 7).
    a. Plot the following t-SNE maps and export them to your working directory.

```
> pdf(file = "04tsne_maps_3h_embryos.pdf",width = 11, height = 7.5)

> sc <- comptsne(sc,perplexity = 10, rseed = 420)

> plotmap(sc,cex=3)
```

```
> plotexpmap(sc, g=c("scw","sc","esg"), n="scw+sc+esg", logsc = TRUE, cex = 3)

> plotexpmap(sc, g=c("pnr","ths","tin"), n="pnr+ths+tin", logsc = TRUE, cex = 3)

> dev.off()
```

b. Inspect the expression of *scw*, *sc*, *esg*, *pnr*, *ths*, and *tin* in the t-SNE maps to ensure all unfertilized eggs and older embryos were excluded.
c. Determine the developmental trajectory in the new dataset. For example, in our sample dataset the developmental trajectory changed, and it is now from right to left (Figure 7).

Note: The *perplexity* argument in the *comptsne* function can be adjusted to get a stable configuration. In addition to the t-SNE map, we recommend using the Fruchterman-Rheingold layout that is included in the RaceID package to confirm the relationship between samples.

40. Create a pseudo-temporal ordering of embryos using the StemID workflow (included in the RaceID package).
    a. Use the StemID workflow below to infer the inter-cluster links.

```
> ltr <- Ltree(sc)

> ltr <- compentropy(ltr)

> ltr <- projcells(ltr,cthr=2,nmode=T,knn=3)

> ltr <- lineagegraph(ltr)

> ltr <- comppvalue(ltr,pthr=0.05, sensitive = T)
```

Note: StemID is an algorithm included in the RaceID package. For details about what is computed in each function, see the RaceID vignette.

b. Plot and export the inter-cluster links onto the t-SNE map.

```
> pdf(file = "05intercluster_links.pdf", width = 11, height = 7.5)

> plotspantree(ltr,cex = 3, projections = T)

> dev.off()
```

c. Order clusters according to the inter-cluster links, computed by StemID, and the developmental trajectory previously defined. For example, the cluster order is 1, 5, 4, 6, 2, 9, 3, 7 in the sample dataset.

```
> n <- cellsfromtree(ltr,c(1,5,4,6,2,9,3,7))
```

41. Analyze the continuous transcriptome changes during development.
    a. Generate a gene expression matrix where rows represent transcripts and columns represent each embryo, in ascendent pseudo-time order.
        i. Obtain the list of sample IDs, according to pseudo-time, from the StemID workflow.
        ii. Get the normalized read counts from the RaceID workflow.
        iii. Assign the normalized read counts to each embryo, in order according to pseudo-time.
        iv. Export the gene expression matrix.

```
> list_pseudotime <- row.names(as.data.frame(ltr@sc@cpart[n$f]))

> norm_counts <- as.matrix(getfdata(sc))

> norm_counts <- norm_counts[, list_pseudotime]

> write.csv(norm_counts, file = "06normalized_counts_by_pseudotime.csv")
```

b. Smooth the gene expression patterns of any genes of interest.
   i. Copy and paste the normalized read counts by pseudo-temporal order into Prism. Use an XY table format with a single Y value for each point, where the x-axis represents pseudotime (ascending numbers) and the y-axis represents gene expression.
   ii. Analyze the data using the 'Smooth, differentiate or integrative curve' option in the XY analyses drop-down menu, using the following settings: 1) 'Don't differentiate or integrate'; 2) 'Smoothing of 5 neighbors on each size'; and 3) '2nd order of the smoothing polynomial'.
   iii. Plot both the normalized read counts and the smoothed line on a graph (see Figure 8).

c. Compare and analyze continuous transcriptional patterns.
   i. Use these graphs to analyze gene expression patterns. For example, gene expression of *ato* (*atonal*) and *yl* (*yolkless*) show very distinct patters (Figure 8).

42. Perform differential expression analysis between clusters.
   a. Install the DESeq2 R package[9] available from https://bioconductor.org/packages/release/bioc/html/DESeq2.html.
   b. Carry out differential expression analysis between each set of clusters, comparing two clusters at a time. For example, to get differentially expressed genes with p-value adjusted (padj) < 0.05 and log2 fold change >1 or < −1 between cluster #1 and cluster #4, run the code below.

```
> library("DESeq2")

> A <- names(sc@cpart)[sc@cpart %in% c(1)] #cluster 1

> B <- names(sc@cpart)[sc@cpart %in% c(4)] #cluster 4

> x <- diffexpnb(sc@expdata,n=c(A,B),DESeq = TRUE, A=A, B=B, method = "per-condition")

>plotdiffgenesnb(x,pthr=.05,lthr=,mthr=-1,  Aname="Cluster1",  Bname="Cluster4",  show_
names=TRUE, padj=TRUE, main="Clusters 1vs4")
```

   c. Export results to your working directory.

```
> write.table(x$res, "05results_1vs4.xls", col.names=TRUE, sep="\t", quote=FALSE)
```

   d. Replace the cluster numbers in the code and repeat the process to see what genes are differentially expressed between two other clusters. You should repeat this process until all pairs of clusters have been compared.

Sex-biased transcriptome analysis

Timing: 1 day (+ 4 days if genotyping)

This major step describes how to determine the sex of each embryo and how to perform differential expression analysis between male and female embryos on continuous data. Here, we describe a computational method for sex determination that relies on the normalized read counts of *Sxl* and

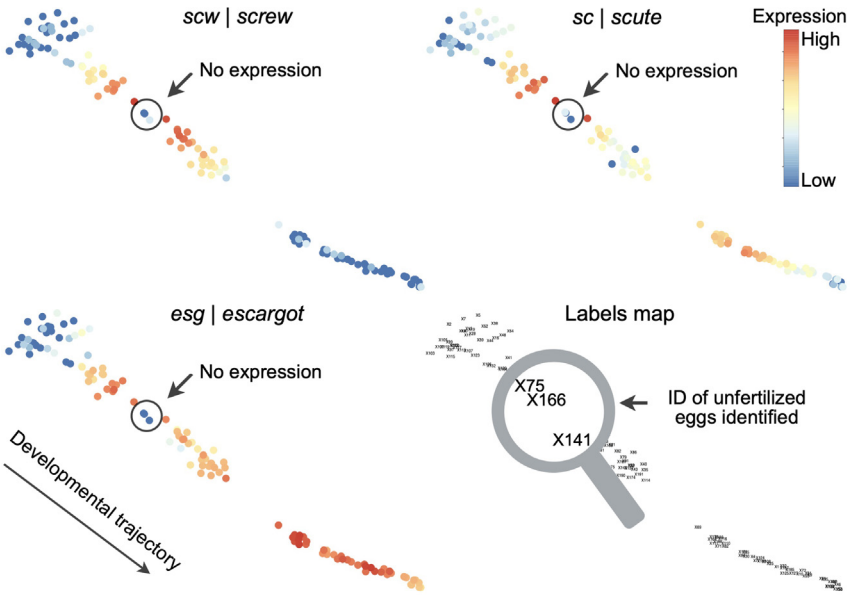


Figure 5. Identification of unfertilized eggs in the dataset

Eggs lacking expression of early transcribed genes, such as *scw*, *sc*, and *esg*, are identified as unfertilized eggs in the t-SNE map.

msl-2. We also provide an optional method that uses sex chromosome specific genotyping PCR. The splineTimeR R package is utilized to identify statistical differences between male and female embryos during the first 3 h of development. At the end of this step, a list of transcripts that are differentially expressed in male and female embryos as well as sex-specific transcriptional patterns are generated.

43. Determine embryo sex using *Sxl* and *msl-2* gene expression.
    a. Obtain normalized read counts of *Sxl* by pseudo-time order and plot them in Prism (see step 41 for details).
    b. Determine when during development the expression of *Sxl* begins to diverge into low and high expression groups. For example, expression of *Sxl* begins to diverge at pseudo-time #47 in the sample data (Figure 9A).
    c. Define the sex of each embryo.
        i. Define embryos with high *Sxl* expression relative to its immediate neighbors as females and those with low *Sxl* expression as males (Figure 9A).

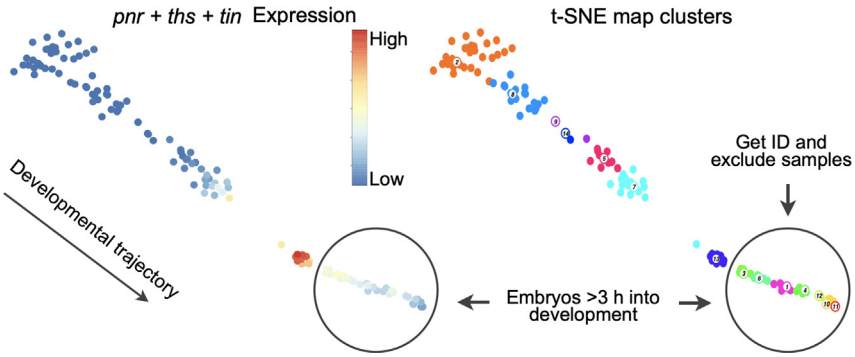


Figure 6. Identification of embryos that are more than 3 h into development

Embryos beyond the combined linear expression of *pnr* (*pannier*), *ths* (*thisbe*) and *tin* (*tinman*) peaks (approximately 3 h into development) should be removed from the dataset.

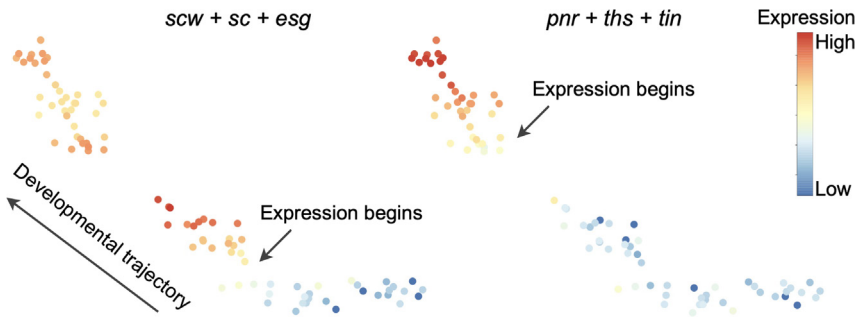


Figure 7. Quality assessment after exclusion of unfertilized eggs and embryos older than 3 h

Plotting the expression of *scw*, *sc*, *esg*, *pnr*, *ths*, and tin ensures that all unfertilized eggs have been removed prior to determining any changes in the developmental trajectory on the t-SNE map.

Note: The high differential expression of *Sxl* between male and female embryos makes it possible to determine the embryo's sex. Such differential expression is only observed during the second half of the pseudo-time window (from approximately 1.5–3 h). The expression of *Sxl* follows a bell-shaped pattern in female embryos. Be particularly cautious when assigning embryo sex at the extremes of the bell shape as differences in *Sxl* expression between males and females are smaller in this region.

d. Confirm embryo sex using *msl-2* expression.
   i. Obtain the normalized read counts of *msl-2* by pseudo-time order.
   ii. Compare the expression levels between male and female embryos, as defined previously by *Sxl* expression. Embryos with higher *msl-2* expression are males and those with lower *mls-2* expression are females. Rely only on the gene expression values of immediate neighbors to establish high or low *msl-2* expression.
   iii. Plot an XY graph with a single Y value for each point, where the x-axis represents *Sxl* normalized read counts and the y-axis represents *msl-2* normalized read counts. All female embryos should be located between a 0° and 45° angle on the plot, while male embryos appear between a 45° and 90° angle (Figure 9B).
   iv. Re-assess the sex of embryos when males or females do not cluster together by their *Sxl* and *msl-2* expression. Table 2 shows the sex of each embryo as determined by both sex chromosome genotyping and *Sxl*|*msl-2* expression, and sample IDs in pseudo-temporal order of the sample data.

Note: The differential expression of *msl-2* between males and females helps to determine the sex of embryos only at high pseudo-times in the sample data (from approximately 2–3 h). The *msl-2* expression follows a bell-shaped pattern in both male and female embryos, however, higher expression levels are observed for males.

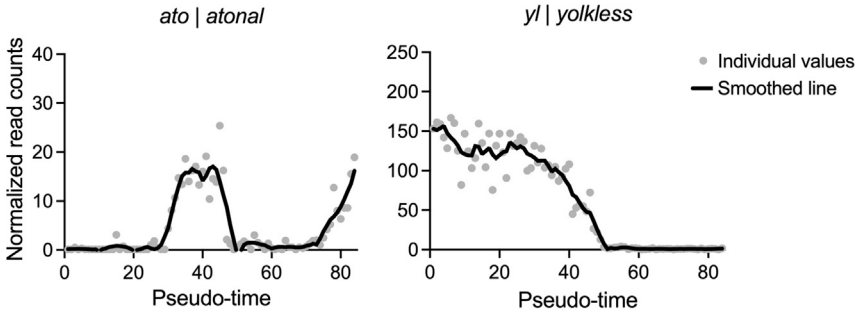


Figure 8. Continuous analysis of transcriptional patterns

Example data showing the gene expression of *ato* and *yl* during the first 3 h of embryonic development.

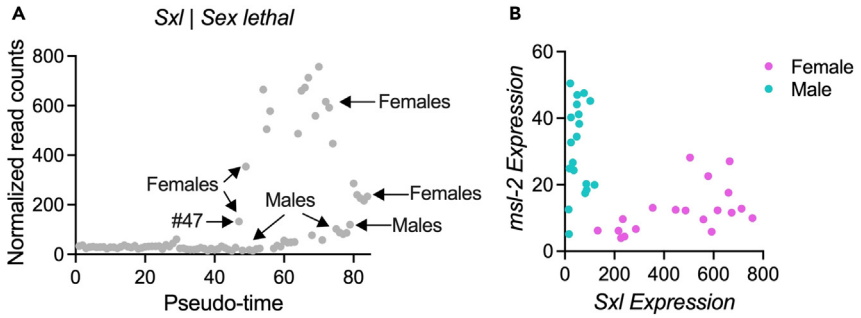


Figure 9. Identification of embryo sex by *Sxl* and *msl-2* gene expression

(A) Differential expression of *Sxl* between male and female embryos makes it possible to identify embryo sex; and (B) *msl-2* expression further differentiates between the sexes. Plotting expression of *Sxl* and *msl-2* on an XY graph, positions female embryos between a 0° and 45° angle and male embryos between a 45°–90° angle.

Note: Embryo determination using *Sxl* and *msl-2* levels avoids the time-consuming process of sex chromosome genotyping. Nevertheless, this method only works for embryos during the developmental timeframe when there are differential expression patterns for *Sxl* and *msl-2* between male and female embryos. Genotyping of the sex chromosomes is needed to determine the sex of embryos outside this timeframe.

44. Extract DNA after RNA isolation (optional).

Note: This protocol is an adaptation of the manufacturer's protocol (Doc. Part No. 15596026.PPS, Pub. No. MAN0001271) available from https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf.

a. Select only samples included in the pseudo-time analysis for DNA extraction (Table 2). The total number of samples included in the pseudo-time analysis of our sample data is 84, after excluding embryos that were either unfertilized or that were older than 3 h.
b. Remove any remaining aqueous phase.
   i. Take samples (from step 7) out of the freezer (−80°C).
   ii. Incubate for 3 min at room temperature.
   iii. Centrifuge at 12,000 × *g* for 10 min at 4°C.
   iv. Remove and discard any remaining upper aqueous phase using a pipette.
c. Precipitate DNA.
   i. Transfer the interphase and the lower phenol-chloroform phase to a new 1.5 tube.
   ii. Add 2 μL GlycoBlue co-precipitant (15 mg/mL).
   iii. Invert the tube 5 times.
   iv. Add 150 μL of 100% ethanol.
   v. Invert the tube 5 times.
   vi. Incubate for 3 min at room temperature.
   vii. Centrifuge at 7,000 × *g* for 5 min at 4°C to pellet the DNA.
   viii. Remove and discard the supernatant without disturbing the pellet.
d. Wash with sodium citrate.
   i. Resuspend the pellet in 500 μL 0.1 M sodium citrate in 10% ethanol.
   ii. Incubate samples for 30 min at room temperature. Mix by gently inverting tubes every 10 min during incubation.
   iii. Centrifuge at 7,000 × *g* for 5 min at 4°C.
   iv. Discard the supernatant.
   v. Repeat steps i-iv.
e. Wash with ethanol.

i. Resuspend the pellet in 1 mL 75% ethanol in MilliQ-grade water.
ii. Add 2 μL GlycoBlue co-precipitant (15 mg/mL).
iii. Incubate samples for 10 min at room temperature. Mix samples by gently inverting tubes every 2–5 min during incubation.
iv. Centrifuge at 7,000 × *g* for 5 min at 4°C.
v. Discard supernatant.
vi. Air-dry samples until they are completely dry. This should take approximately 15 min.

f. Solubilize the DNA.
i. Resuspend the pellet in 23 μL 8 mM NaOH (w/v).
ii. Incubate DNA for 10 min at 80°C. Vortex briefly every 2 min.
iii. Chill samples on ice for 5 min.
iv. Vortex samples for 5 s and spin them for 5 s at ≤2000 × *g*.
v. Centrifuge at 12,000 × *g* for 10 min at 4°C.
vi. Transfer the supernatant (containing the DNA) to a new 1.5 mL tube.
vii. Store DNA samples at −80°C until quantitative PCR (qPCR) is performed.

45. Perform qPCR on DNA to determine the sex of embryos.
a. Thaw the reagents on ice and prepare 800 μL qPCR reaction master mix in a 1.5 mL tube, according to the table below. Keep the reagents on ice until use. This volume is enough for 96 reactions (44 samples and 4 no template control, all in duplicate).
b. Use the primers to amplify the X-chromosome first. The amplification of the Y chromosome is done separately. The primer sequences for the X and the Y chromosome are listed on the key resources table.

qPCR reaction master mix

Reagent	Amount
Nuclease-free water	250 μL
2× SsoAdvanced Universal SYBR Green Supermix	500 μL
Forward primer (10 μM)	25 μL
Reverse primer (10 μM)	25 μL
Total	800 μL

c. Vortex qPCR reaction master mix for 5 s and spin tube for 5 s at ≤2000 × *g*.
d. Pipette 8 μL qPCR reaction master mix into each well of a 96-well PCR plate on ice.
e. Add 2 μL nuclease-free water to 8 wells distributed across the plate.
i. Use a new pipette tip for each control and duplicate.
ii. Keep the plate on ice during this process.
iii. Carefully note the location of these control wells.
f. Add 2 μL of DNA sample to the rest of the wells (one sample per well).
i. Include a technical duplicate for every sample. A maximum of 44 different samples can be process per 96-well plate.
ii. Use a new pipette tip for each sample and duplicate.
iii. Keep plate on ice during this process.
g. Seal the plate with a Microseal 'B' PCR seal, gently tap to mix, and spin plate for 30 s at ≤850 × *g*.
h. Incubate the plate in a thermocycler with the lid temperature set at 105°C and perform qPCR using the following program:

qPCR cycling conditions

Steps	Temperature	Time	Cycles
Activation	98°C	3 min	1
Denaturation	98°C	10 s	45 cycles
Annealing/Extension	60°C	30 s	
Melting curve	65°C–95°C	0.5°C increment each 5 s	1

Table 2. Pseudo-time order and sex of embryos in the filtered data up to 3 h into development

			Sex by					Sex by	
Pt	Sample	Barcode	Geno	Exp	Pt	Sample	Barcode	Geno	Exp
1	X1A.22	X14	UND	UND	43	X2A.32	X144	FEMALE	UND
2	X1B.04	X20	UND	UND	44	X2A.30	X142	MALE	UND
3	X2A.07	X39	UND	UND	45	X2A.15	X135	MALE	UND
4	X1B.20	X28	UND	UND	46	X2B.20	X60	FEMALE	UND
5	X1A.19	X11	UND	UND	47	X2A.21	X45	FEMALE	FEMALE
6	X1A.02	X2	UND	UND	48	X2A.18	X42	MALE	MALE
7	X1A.20	X12	UND	UND	49	X2B.24	X64	FEMALE	FEMALE
8	X1A.25	X105	UND	UND	50	X3B.04	X84	MALE	MALE
9	X1A.11	X99	UND	UND	51	X3A.03	X67	MALE	MALE
10	X1B.03	X19	UND	UND	52	X3A.11	X163	MALE	MALE
11	X1B.13	X117	UND	UND	53	X3B.03	X83	MALE	MALE
12	X1A.12	X100	UND	UND	54	X3B.15	X183	FEMALE	FEMALE
13	X1B.15	X119	UND	UND	55	X3A.28	X172	FEMALE	FEMALE
14	X1A.05	X5	UND	UND	56	X3A.25	X169	FEMALE	FEMALE
15	X2A.06	X38	UND	UND	57	X1B.05	X21	MALE	MALE
16	X1A.13	X101	UND	UND	58	X3B.01	X81	MALE	MALE
17	X1A.07	X7	UND	UND	59	X3B.02	X82	MALE	MALE
18	X1A.15	X103	UND	UND	60	X3B.19	X91	MALE	MALE
19	X1B.09	X113	UND	UND	61	X3A.31	X175	MALE	MALE
20	X1A.32	X112	UND	UND	62	X3A.09	X161	MALE	MALE
21	X2B.04	X52	UND	UND	63	X3B.06	X86	MALE	MALE
22	X1A.09	X97	UND	UND	64	X3B.20	X92	FEMALE	FEMALE
23	X1B.27	X123	FEMALE	UND	65	X3A.23	X79	FEMALE	FEMALE
24	X1A.24	X16	MALE	UND	66	X2B.09	X145	FEMALE	FEMALE
25	X2A.20	X44	MALE	UND	67	X3B.30	X190	FEMALE	FEMALE
26	X2A.24	X48	MALE	UND	68	X2B.19	X59	MALE	MALE
27	X1A.27	X107	MALE	UND	69	X2A.25	X137	FEMALE	FEMALE
28	X1B.11	X115	UND	UND	70	X2A.08	X40	FEMALE	FEMALE
29	X2B.06	X54	FEMALE	UND	71	X2A.03	X35	MALE	MALE
30	X2A.17	X41	FEMALE	UND	72	X3A.30	X174	FEMALE	FEMALE
31	X2B.26	X154	FEMALE	UND	73	X3B.31	X191	FEMALE	FEMALE
32	X2A.27	X139	MALE	UND	74	X2A.19	X43	FEMALE	FEMALE
33	X2A.23	X47	FEMALE	UND	75	X1B.10	X114	MALE	MALE
34	X2A.26	X138	MALE	UND	76	X3B.11	X179	MALE	MALE
35	X2A.11	X131	MALE	UND	77	X3A.12	X164	MALE	MALE
36	X2A.28	X140	FEMALE	UND	78	X1A.31	X111	MALE	MALE
37	X2B.28	X156	FEMALE	UND	79	X3B.29	X189	MALE	MALE
38	X2B.12	X148	MALE	UND	80	X3B.26	X186	FEMALE	FEMALE
39	X2B.18	X58	MALE	UND	81	X3A.07	X71	FEMALE	FEMALE
40	X2B.21	X61	MALE	UND	82	X1B.12	X116	FEMALE	FEMALE
41	X2B.15	X151	FEMALE	UND	83	X2B.22	X62	FEMALE	FEMALE
42	X2A.04	X36	FEMALE	UND	84	X1A.30	X110	FEMALE	FEMALE

Pt, Pseudo-time; Geno, genotyping of the Y and the X chromosome; Exp, expression of *Sxl* and *msl-2*; UND, undetermined.

i. Remove plate from the thermocycler and discard.
j. Get qPCR data from the instrument and analyze it to get the cycle threshold (Ct) values.
k. Repeat the process for the rest of the X and all of the Y chromosome reactions. A total of four 96-well plates are needed to amplify both sex chromosomes on 84 samples in duplicate.
l. Determine the sex of each embryo using the qPCR data.

i. Ensure the melting curve for each sample has one peak at 83.5°C–84°C for the X-chromosome and one at 80°C–80.5°C for the Y-chromosome. An example of the melt peak for each chromosome is shown in Figure 10. If these peaks are not observed, the data should be excluded or/and the qPCR should be repeated for that sample.
ii. Make the threshold for Ct values the same for all samples and chromosomes.
iii. Compare the Y-chromosome Ct values to the X-chromosome Ct values in each sample. Male embryos have similar Ct values (±2.5) for both sex chromosomes while female embryos only amplify the X-chromosome. Refer to problem 5 in the troubleshooting section if the difference between Ct values is more than 2.5 and amplification has occurred.
iv. Designate all reactions with a difference >1.5 between the Ct values of technical replicates as an inconsistent result. See problem 6 in the troubleshooting section for options to address this.

m. Repeat qPCR for any samples with an inconsistent result between replicates.

Note: No amplification of the Y chromosome indicates the embryo is female. The X-chromosome amplification reaction is performed to verify the successful extraction of DNA for all samples. In addition, the X- and Y-chromosome Ct values should be the same in male embryos, however, in our experience we consider differences of up to 2.5 in Ct values between the X- and Y-chromosome amplification reaction as normal due to differences in primer efficiencies and batch effects. A greater difference likely indicates contamination in one of the reactions.

46. Determine differentially expressed transcripts between male and female embryos using splineTimeR.[10]
    a. Install the splineTimeR R package.
    b. Create a csv file with the name 'metadata'.
        i. Use the 1st column for sample IDs in pseudo-time order. Set the column name to 'SampleName'.
        ii. Use the 2nd column for cluster number reorganized in ascending order according to the developmental trajectory. Set the column name to 'Time'.
        iii. Use the 3rd column for the sex of each embryo. Set the column name to 'Treatment'.
        iv. Use the 4th column for the different number of females or males included in each cluster. Set the column name to 'Replicate'.
        v. Exclude all samples with undetermined sex from the metadata file. For our sample data, 61 out of 84 samples were included in the analysis (Table 2).
    c. Create a file with the normalize read counts for each sample included in the metadata file.

```
> norm.counts <-read.csv("06normalized_counts_by_pseudotime.csv", sep=",",
header=TRUE, row.names = 1)
>
> m.data <- read.csv("metadata.csv", sep=",", header=TRUE)
>
> sample_list <- m.data[,1]
> norm.counts <- norm.counts[,sample_list]
> row.names(m.data) <- m.data[,1]
```

Note: The '06normalized_counts_by_pseudotime.csv' file was previously created following the StemID workflow (see step 41.a). It contains the normalized read counts for all samples included in the pseudo-time trajectory.

d. Create an ExpressionSet object using the normalized read counts and metadata.

```
> phenoData <- new("AnnotatedDataFrame", data=m.data)

> minimalSet <- ExpressionSet(assayData=as.matrix(norm.counts), phenoData = phenoData)
```

e. Perform differential expression analysis using the *splineDiffExprs* function. The example code below is to get differentially expressed genes with padj value < 0.01 using 7 degrees of freedom.

```
> diffExprs <- splineDiffExprs(eSetObject = minimalSet, df = 7, cutoff.adj.pVal = 0.01,
reference = "MALE", intercept = TRUE)
```

f. Export results with *write.csv* function.

```
> write.csv(diffExprs, file = "07diffExp_males_females.csv")
```

Note: splineTimeR uses transcript abundance in each RaceID cluster of the pseudo-time trajectory to create curves of gene expression over time. Comparisons of the shape of the curves and expression levels between male and female embryos allows for the identification of differentially expressed genes. Further details about splineTimeR can be found in the splineTimeR vignette.

47. Plot the gene expression of male and female embryos in pseudo-temporal order.
    a. Smooth the expression of genes over pseudo-time using Prism, as previously described (see step 41.b).
        i. Use Y-axis (columns) for gene expression and x-axis (rows) for pseudo-time number.
        ii. Use two columns, one for expression in male embryos and the other for female embryos.
        iii. Enter the normalized read counts of male and female embryos into the respective columns.
        iv. Leave cells blanks for those embryos with undetermined sex. For example, embryo #28 in the pseudo-time in the sample dataset (Table 2).

EXPECTED OUTCOMES

This single-embryo sequencing method generates high-quality RNA-seq data, allowing for the accurate characterization of early transcriptional events. The continuous collection of data allows us to determine gene expression patterns and the precise start of transcription for each detectable gene during the zygote genome activation. The ability to determine the sex of each embryo, directly from the transcription profiles, allows for sex specific analysis of gene expression from the very beginning of zygotic genome activation.

At the end of this protocol, researchers will achieve 3 main outcomes: 1) A matrix with normalized gene counts by samples, according to the developmental embryo trajectory (pseudo-time) from 1–3 h; 2) Lists of differentially expressed transcripts between different clusters along the developmental trajectory; and 3) A list of differentially expressed transcripts between males and females during the first 3 h of embryonic development.

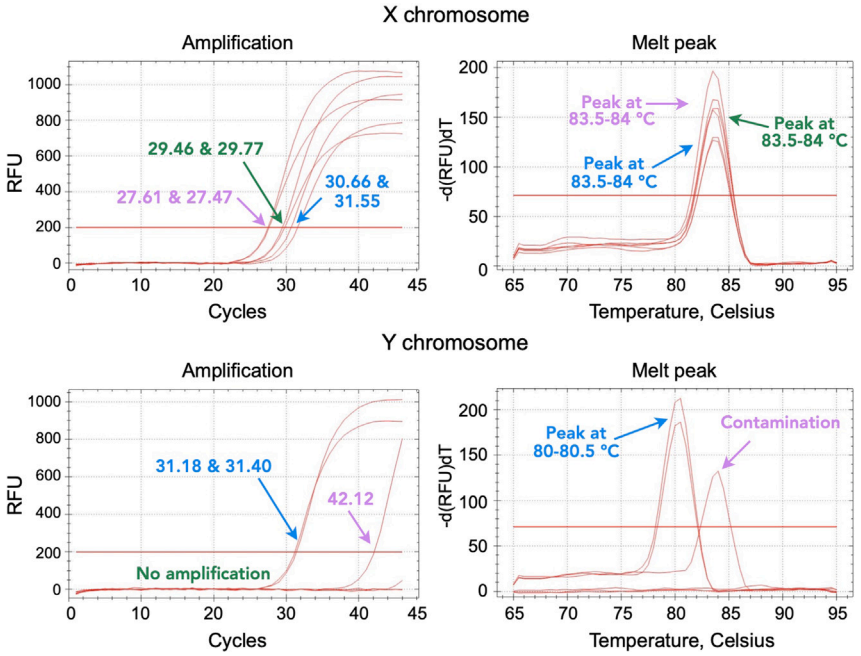


Figure 10. Identification of embryo sex by sex chromosome genotyping

Example of expected qPCR results for female embryos, male embryos, and a contaminated sample. DNA from female embryos amplifies only the X chromosome (green arrow, top). DNA from male embryos amplifies both the X and the Y chromosome at similar Ct values (blue arrows, top and bottom). Melt peak is 83.5°C–84°C for the X chromosome and 80°C–80.5°C for the Y chromosome. Contamination of a qPCR reaction can be identified by a melt peak at a time other than these. The example data shows contamination only during the Y chromosome amplification (purple arrows).

LIMITATIONS

Although this method identifies approximately 7,600 different transcripts in single embryos, it may not detect genes with very low expression levels. Further, the use of whole single embryos prevents drawing any conclusions about cell-type-specific gene expression or the spatial distribution of transcripts. However, cell-type-specific information is likely of minor relevance during early embryonic development, as nuclei share the same cytoplasm until cellularization. It is well known that the spatial distribution of transcripts is important for early embryogenesis. While the methodology presented here cannot determine the spatial distribution of a transcript, it can be used for selecting candidate genes with interesting temporal expression patterns for further characterization of spatial distribution using other methods.

TROUBLESHOOTING

Problem 1

I.DOT is dispensing lower volumes than directed (step 12.e).

Potential solution

- Be sure to prime the wells before dispensing the desired volume onto the target plates.
- Increase the priming volume before dispensing onto the target plates.
- Low volumes in the I.DOT source plates (<10 µL) lead to more errors. Increase the volume in source-plate wells (if possible).
- Allow for automatic correction. This means I.DOT will dispense the missing volume after you 'accept' the option to do so.

Problem 2

I.DOT is dispensing inadequate volumes of Vapor-Lock (step 12.f).

Potential solution

- Try to reduce the time between when you add the Vapor-Lock to the source plate and when it is dispensed onto the target plate.
- Use new wells every time you want to dispense Vapor-Lock.
- Use the source plates with a 60 nm pore size.
- Manually pipette Vapor-Lock into the wells.

Problem 3

No peak detected by Agilent Bioanalyzer on aRNA (step 23).

Potential solution

- Run the instrument using a spare RNA sample with known RNA concentration to ensure that the instrument and reagents are working properly.
- Run the sample again using another 1 μL aRNA to eliminate the possibility that there was an issue with a specific run or chip.
- Run blanks (no RNA, only the gel-matrix with the marker) in wells #1 and #2. Place your sample on the 3rd well on the chip.

Problem 4

Low DNA yield in sequencing libraries (step 28).

Potential solution

- The yield for sequencing libraries depends on the developmental stage of the embryos included, RNA isolation yield, RNA elution volume (2 μL), cDNA dilution, and the number of PCR cycles. Because both high and low cDNA input into the PCR reactions can be responsible for low yields, start by increasing and/or decreasing the dilution of cDNA and repeat the library PCR.
- Ask your sequencing facility about library yields. There are low input sequencing protocols that can be perform.
- Repeat the library PCR after increasing the total reaction volume to 100 μL (original volume is 50 μL). Adjust the volumes in the following steps accordingly, except for library elution volume. After the clean-up procedure, you will have twice as much material in each library.

Problem 5

Ct value differences > 2.5 between the X-chromosome and the Y-chromosome in the same sample, yet amplification is detected, making it unlikely that the sample belongs to a female embryo (step 45.L).

Potential solution

- Ensure that the melting curve shows only one peak at the right temperature. Multiple peaks, or a peak at the wrong melting temperature, indicate sample contamination.
- Ensure the threshold for Ct values is the same in all samples during analysis. The automatic threshold might be different for each plate.
- Repeat qPCR with a diluted or more concentrated sample. For older embryos, it is likely that a diluted sample is required, as they have more nuclei/DNA. To dilute samples, simply add more 8 mM NaOH (w/v) into DNA. For younger embryos, it is likely a more concentrated sample will be required.
- Repeat the denaturing step (10 min at 80°C), mix thoroughly, spin, and repeat qPCR.

Problem 6

Ct value differences > 1.5 between technical replicates (step 45.l).

Potential solution

- Ensure that the melting curve shows only one peak at the right temperature. Multiple peaks, or a peak at the wrong melting temperature, indicate sample contamination.
- Reduce pipetting errors by using 2 μL pipette tips to transfer DNA.
- Reduce pipetting errors by increasing the qPCR reaction total volume to 30 μL (3×). Pipetting 6 μL DNA sample (30 μL reaction) has less pipetting error compared with pipetting 2 μL (10 μL reaction).
- Incubate DNA sample for 10 min at 80°C again. Mix thoroughly, spin, and repeat qPCR.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Adelheid Lempradl (heidi.lempradl@vai.org).

Materials availability

This study did not generate unique reagents.

Data and code availability

- Single-embryo RNA-seq data from this study has been deposited at NCBI Gene Expression Omnibus (GEO). The accession number is listed in the key resources table.
- All code used is available in each step on this protocol, on Zenodo at https://doi.org/10.5281/zenodo.8190517, and on our GitHub repository at https://github.com/Lempradl/PerezMojica2023_STARProtocols.
- Any additional information required to reanalyze the data reported in this protocol is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Carolyn Anderson and Joe Roy for critical reading and editorial feedback on this manuscript. We thank the Van Andel Institute Genomics Core (RRID:SCR_022913), especially Marc Wegener and Marie Adams, for their assistance with RNA-seq. We thank the VAI Bioinformatics and Biostatistics Core for supporting the analyses. Graphical abstract, Figure 1, and Figure 2 were created using BioRender.com. We thank J. Andrew Pospisilik for supporting this project. This research was funded by the Van Andel Research Institute.

AUTHOR CONTRIBUTIONS

J.E.P.-M., Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Visualization, Supervision; L.E., Investigation, Methodology; K.H.L., Software, Formal Analysis, Writing – Original Draft; A.L., Conceptualization, Methodology, Validation, Data Curation, Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community.

REFERENCES

1. Pérez-Mojica, J.E., Enders, L., Walsh, J., Lau, K.H., and Lempradl, A. (2023). Continuous transcriptome analysis reveals novel patterns of early gene expression in Drosophila embryos. Cell Genom. *3*, 100265. https://doi.org/10.1016/j.xgen.2023.100265.

2. Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*, 77. https://doi.org/10.1186/s13059-016-0938-8.

3. Sagar, Herman, J.S., Pospisilik, J.A., and Grün, D. (2018). High-Throughput Single-Cell RNA Sequencing and Data Analysis. In CpG Islands: Methods and Protocols, T. Vavouri and M.A. Peinado, eds. (Springer New York), pp. 257–283. https://doi.org/10.1007/978-1-4939-7768-0_15.

4. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

5. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

6. Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and Marioni, J.C.; participants in the 1st Human Cell Atlas Jamboree (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63. https://doi.org/10.1186/s13059-019-1662-y.

7. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777.

8. Herman, J.S., Sagar, and Grün, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat. Methods *15*, 379–386. https://doi.org/10.1038/nmeth.4662.

9. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

10. Michna, A., Braselmann, H., Selmansberger, M., Dietz, A., Hess, J., Gomolka, M., Hornhardt, S., Blüthgen, N., Zitzelsberger, H., and Unger, K. (2016). Natural cubic spline regression modeling followed by dynamic network reconstruction for the identification of radiation-sensitivity gene association networks from time-course transcriptome data. PLoS One *11*, e0160791. https://doi.org/10.1371/journal.pone.0160791.

11. Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2019). Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res. *29*, 1188–1197. https://doi.org/10.1101/gr.242164.118.

12. Sandler, J.E., and Stathopoulos, A. (2016). Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics *202*, 1575–1584. https://doi.org/10.1534/genetics.116.186783.