**Hybrid-Seq data of embryonic development in human, zebrafish, mouse, chicken, Medaka and C.elegans**

**[1] Human**

OSC path: /fs/project/PCON0009/LabData/Human

**[2] Zebrafish**

**Data description:** This data contains Pacbio and Illumina data for 5 stage samples (Fertilized egg, 1-cell, 64-cell, 1K-cell and shield) of early embryonic development in zebrafish.

**OSC path:** /fs/project/PCON0009/LabData/Zebrafish

**[3] Mouse**

**Ref:** Qiao Y, Ren C, Huang S, et al. High-resolution annotation of the mouse preimplantation embryo transcriptome using long-read sequencing[J]. Nature communications, 2020, 11(1): 1-13.

**Data availability at:** <https://www.ncbi.nlm.nih.gov/sra?term=SRP225196>

**Data description:** We used an combination of long-read and short-read sequencing data to explore the transcriptome of mouse preimplantation embryo Overall design: We selected 7 stage of early embryo including sperm and oocyte from ♂DBA/2J and ♀C57/BL6J and sequenced using Pacbio and Illumina.

**OSC path:**

/fs/project/PCON0009/LabData/embryonic\_development\_raw\_data/mouse\_raw\_data

**[4] Chicken**

**Ref:** Ren J, Sun C, Clinton M, et al. Dynamic Transcriptional Landscape of the Early Chick Embryo[J]. Frontiers in cell and developmental biology, 2019, 7: 196.

**Data availability at:** <https://www.ebi.ac.uk/ena/browser/view/PRJNA488330>

**Data description:** Defining the dynamic transcriptome of the early embryo at high resolution would assist greatly in understanding vertebrate development. The chicken embryo is a long-established model system for vertebrate embryological studies, and, in combination with advanced single-molecule long-read isoform sequencing (Iso-Seq) technology, enables the dynamic transcription landscape of early vertebrate development to be studied in detail. We have generated large transcriptomic profiles reflecting the time course of chicken embryonic development from day 1 to day 8 of incubation: a period encompassing, gastrulation, somitogenesis and organogenesis.

**OSC path:**

/fs/project/PCON0009/LabData/embryonic\_development\_raw\_data/chicken\_raw\_data

**[5] Medaka**

**Ref:** Li Y, Liu Y, Yang H, et al. Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis[J]. Genome research, 2020, 30(6): 924-937.

**Long reads**

**Data availability at:** <https://www.ebi.ac.uk/ena/browser/view/PRJNA488330>

**Data description:** We collected ten embryonic stages covering cleavage, blastula, gastrula, somite, and late stages until hatching. We also collected several adult tissue samples, including brain, heart, ovary, testis, gut, andmuscle. 18 samples were pooled into five libraries and sequenced using the SMRT platform of PacBio to generate a long reads dataset.

**Short reads**

**Data availability at:** <https://www.ebi.ac.uk/ena/browser/view/PRJNA560946>

**Data description:** We collected ten embryonic stages covering cleavage, blastula, gastrula, somite, and late stages until hatching. We also collected two adult tissue samples, including ovary and testis. 13 samples were sequenced using the illumina short-read sequencing platform. Overall design: Total RNA was extracted from medaka whole embryo samples of 11 developmental stages and 2 adult tissues. Short-read strand-specific library was constructed and sequenced using Illumina platform.

**OSC path:**

/fs/project/PCON0009/LabData/embryonic\_development\_raw\_data/Medaka\_raw\_data

**[6] C.elegans**

**Ref:** Roach N P, Sadowski N, Alessi A F, et al. The full-length transcriptome of C. elegans using direct RNA sequencing[J]. Genome research, 2020, 30(2): 299-312.

**Data availability at:** <https://www.ebi.ac.uk/ena/browser/view/PRJNA560946>

**Data description:** We applied nanopore-based direct RNA sequencing to characterize the developmental polyadenylated transcriptome of C. elegans. Taking advantage of long reads spanning the full length of mRNA transcripts we provide support for 20,902 splice isoforms across 14,115 genes, without the need for computational reconstruction of gene models. Of the isoforms identified, 2,188 are novel splice isoforms not present in the Wormbase WS265 annotation. Furthermore, we identified 16,325 three-prime untranslated region (3’UTR) isoforms, 2,304 of which are novel and do not fall within 10 bp of existing 3'UTR datasets and annotations. Combining 3’ UTRs and splice isoforms we identified 25,944 full-length isoforms. We used nanopore data to measure the poly(A) tail lengths for each of our reads, and examine properties of these lengths.

**OSC path:**

/fs/project/PCON0009/LabData/embryonic\_development\_raw\_data/C.elegans\_raw\_data