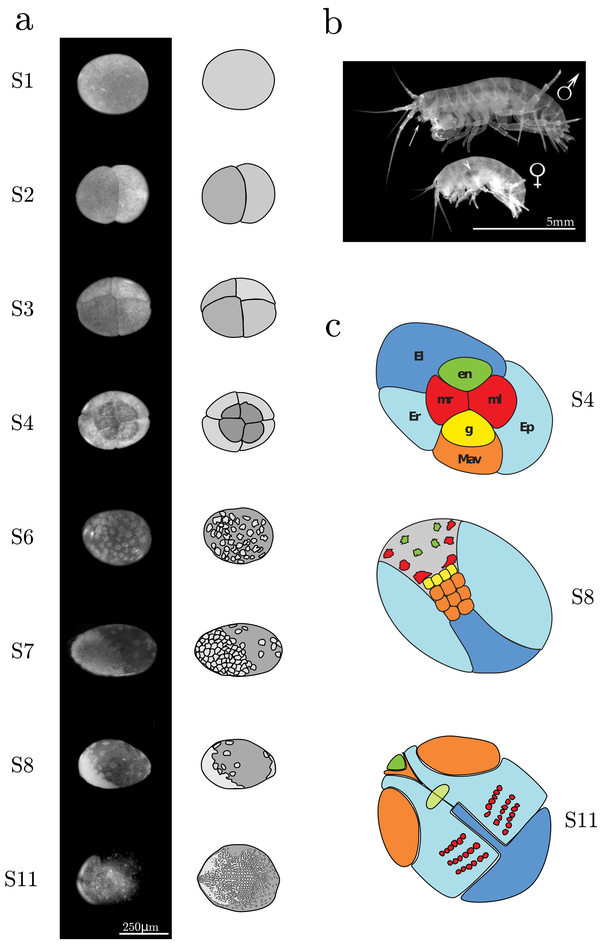
A study of germ layer specification of the crustacean Parhyale hawaiensis

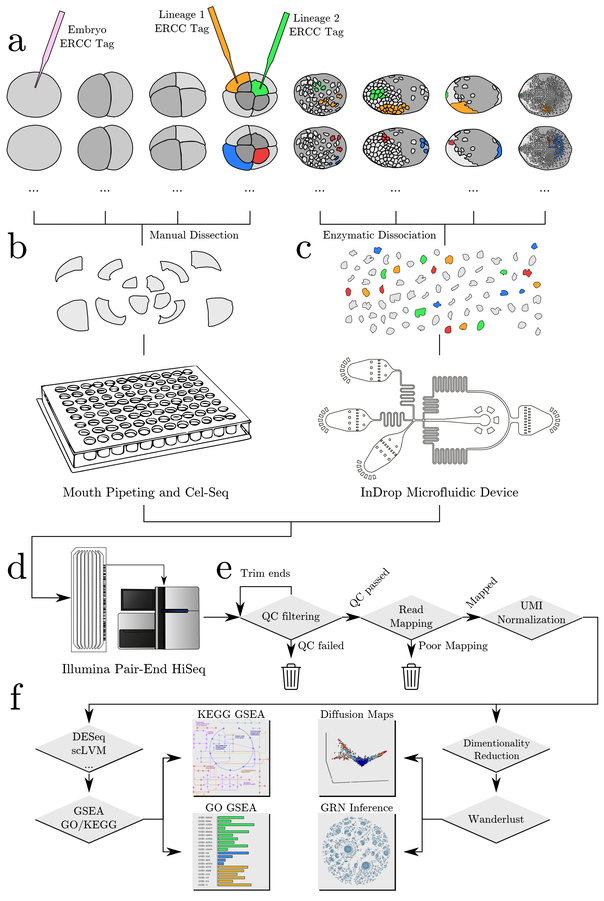
Leo Blondel, Cassandra Extavour

Since the first observations of embryogenesis, the specification of cell fates, from an oocyte to a terminal cell type, has been a central question. In 1940, Waddington proposed a model named the **"Epigenetic Landscape"** in which he described the specification of cell fate as a marble rolling down a landscape shaped by different gene expression patterns. Thanks to the advances of modern "omics" technics, it is now possible to observe the gene landscape of cells during development. However, most studies focus on classical model organisms, or in vitro models. Here, I propose to study the specification of the different germ layers of an emerging model organism, Parhyale hawaiensis, using single cell RNA sequencing measurement of development from a single cell to the so-called germ band stage, allowing for the exploration of gene regulatory networks topologies, shifts and rearrangements along the epigenetic landscape. Finally, using a Gene Positioning System (GPS) and light sheet imaging, a 3D virtual embryo will be created where one can visualize gene expression throughout the development rendered in time and space, on a single cell resolution.

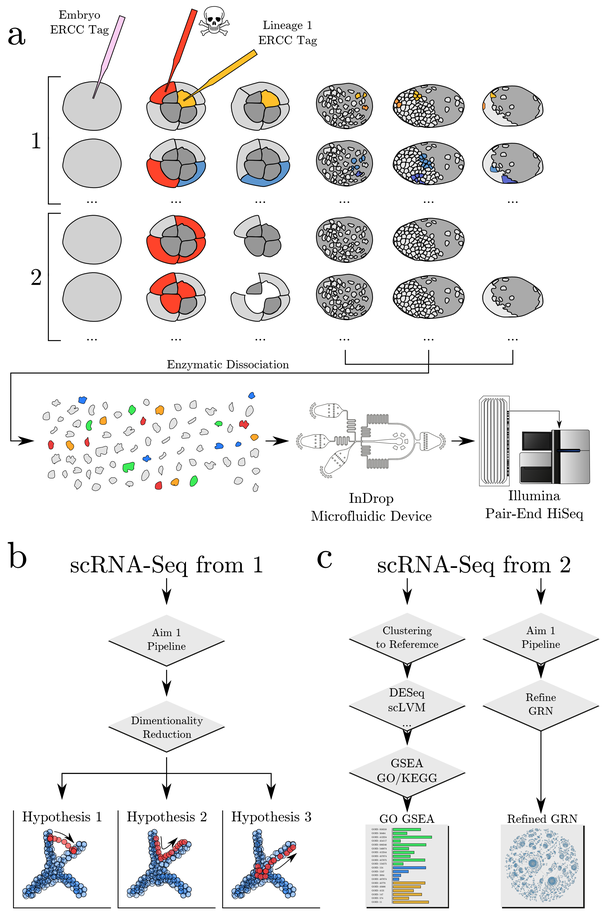
Introduction



Aim 1: How does *P. hawaiensis* specifies the 3 germ layers?



Aim 2: How do the Gene Regulatory Networks remodel themselves upon, following and throughout regeneration of the ablation of a germ layer ?



Aim 3: How consistent are the spatial relationships between cells during development, and how does this influence their GRNs?

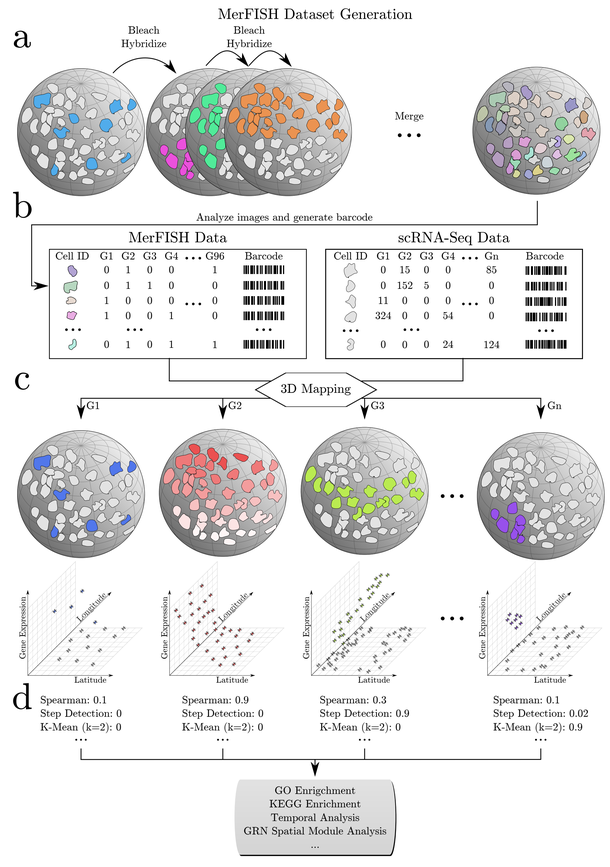


Figure 1: Different stages of *P. hawaiensis* picture and schematics

Overview of P. hawaiensis embryogenesis. In (a), incident light illumination micrographs of selected stages during embryonic development. In between S1 and S4, holoblastic cleavage leads to 8 cells with defined lineage commitment (c). From S4 to S6, yolk segregation occurs and leads to a layer of cells on the surface of the embryo, named the soccer ball stage. During S7, cells migrate towards to anterior pole of the embryo. At S8, gastrulation starts with the invagination of the mesoderm and germ cells. From S8 to S11, gastrulation movements lead to the formation of the germ band, a structure classic to arthropod development where elongation and segmentation will happen. In (b), adult P. hawaiensis, male and female. In (c), lineage specification of the 8 cell stage (S4). Each of the eight blastomeres is committed toward a given lineage as early as S4 (the eight cell stage): three macromeres will give rise to the Ectoderm, two micromeres and one macromere will give rise to the Mesoderm, one micromere to the Endoderm, and one micromere is the germ line precursor. Each cell gives rise to a predictable spatial pattern of descendants in later stages of the embryos, as shown here for S8, where the descendant of Mav will start the invagination; and for S11 where the body plan have been patterned and each lineage participating in a specific portion of the embryo.

Figure 2: Aim 1 experimental and analysis workflow

Overview of the experimental and analysis workflow proposed in Aim 1. In (a), each embryo is tagged with an ERCC mRNA, and 2 different blastomeres per embryo are tagged as well. Then (b) up to S4, embryos will be manually dissected and mouth pipetted into individual wells, followed by barcoding and amplification using CelSeq. After S4 (c), enzymatic dissociation will be performed to recover single cells, which will be inputted into the InDrop microfluidic device for reverse transcription and barcoding. Libraries from (b) and (c) will be sequenced using illumina paired-end HiSeq sequencing. Reads will be filtered (e) by sequencing quality control, followed by read mapping thresholding to remove cells that have a low quality library, and finally reads will be normalized by UMI to obtain mRNA counts for each cell. The dataset will be analyzed (f) using diffusion map dimensionality reduction and clustering, followed by GRN inference using Wanderlust. Differential Expression analysis using DESeq2 and scLVM will be performed in between each previously clustered lineage. Finally, genes differentially expressed will be subjected to Gene Set Enrichment Analysis (GSEA) for GO term and KEGG pathways.

Figure 3: Aim2 experimental and analysis workflow

Overview of the experimental and analysis workflow proposed in Aim 2. In (a), the two ablation experiments are presented. First (1), a blastomere will be ablated from the embryo, and the blastomere which will regenerate the missing lineage will be tagged with an ERCC mRNA. For example, the Mav macromere is ablated, and each of the ml and mr micromeres will be tagged to study the regeneration process. Second (2), all the blastomeres of a given lineage will be ablated and embryos will be allowed to develop, until S8 for ablation of the Ectoderm, and until S9 for all other germ layers. In both cases, cells will be dissociated and their transcriptomes sequenced using InDrop and Illumina Paired End HiSeq as described in Aim 1. The dataset obtained by experiment 1 will be analysed as in Aim 1 (b) and the trajectories of regenerating cells will be analysed in the diffusion map space. Here are shown the three hypotheses discussed in Aim 2. The dataset obtained in the second experiment will serve two purposes (c). First, cells′ transcriptomes will be assigned to a a lineage by clustering them against the reference transcriptome obtained in Aim 1, and differential gene expression analysis will be performed between the ablated embryo and the reference embryo. Genes down- or up-regulated will be subjected to GSEA. Second, cells transcriptomes will be filtered using the pipeline defined in Aim 1, and inferred GRN will be refined using the perturbation created by lack of a germ layer. Finally, the refined GRN will be analysed.

Figure 4: Aim 3 experimental and analysis workflow

Overview of the experimental and analysis workflow proposed in Aim 3. Embryos of different stages will be fixed and hybridized against 96 selected genes (as explained in Aim 3) using the MERFISH protocol. Each set of transcripts corresponds to a unique probe set code (using ON/ OFF states). Following successive rounds of hybridization and photobleaching, the number of transcripts for each cell at a given stage will be recorded. In (b), using the approximate transcript count from MERFISH and the cells′ transcriptomes from the scRNA-seq data of Aim 1, a unique transcriptomic barcode will be assigned to each cell, which will be used to bin the cells together. Using this correspondence between the two datasets, the expression of each gene in space and time will be mapped (c) onto a reference 3D embryo obtained by lightsheet microscopy, allowing for the visualisation of gene expression in time and space and spatial analysis. Finally (d), Spearman correlation between the different dimensions, along with step detection and K-mean clustering will allow for the selection of patterned gene expressions. Gene following similar activation patterns will be further analysed using GO and KEGG GSEA, along with the extraction of GRN submodule activated in space.