

Curriculum Vitae

Buhan Feng(馮步翰)

Postgraduate, College of Life Science, Sun Yat-sen University

Higher Education Mega Center, Guangzhou, China

Email: buhan_feng@foxmail.com

TEL: (+86)19928362840

Education background

- | | |
|----------------------------|--|
| Sept. 2020
to Present | Mphil in Bioinformatics, College of Life Science, Sun Yat-Sen University, Guangzhou, China <ul style="list-style-type: none">GPA 3.12The Third Prize ScholarshipCommand skills: Python, R, statistic, other practical skills in bioinformaticBasic biological courses: Molecular Biology, Cell Biology, Biochemistry |
| Sept. 2012
to Jun. 2016 | B.E. in Electronic and Information Engineering, College of Engineering, South China Agricultural University, Guangzhou, China <ul style="list-style-type: none">GPA 3.27The Second Prize Scholarship (1/5)CHN Patent 《低成本的農業航空機載多光譜相機成像與採集系統》Command skills: Embedding System designing, one-chip computer, C/C++ |

Working experience

- | | |
|----------------------------|--|
| Sept. 2017
to Jul. 2020 | Software Engineer in Huizhi Telecommunications Technology CO., LTD (匯智通信技術有限公司), Guangzhou, China <ul style="list-style-type: none">Familiar with developing web applicationCommand skills: Java, JavaScript, Database, SQL, Spring Boot, VueLeading role of development crew |
| Jun. 2016
to Aug. 2017 | Associated Software Engineer in HSBC Development Center (匯豐銀行環球開發中心), Guangzhou, China <ul style="list-style-type: none">Support banking business of HSBC around the worldDesign workflow of server using Shell Script |

Academic practice

1. A Comprehensive Development Cells Atlas for Zebrafish

Zebrafish has long been an excellent organism for development research. More than forty thousand papers that involved in zebrafish are published for the past ten years. Single cell RNA sequencing(scRNA-seq) data generated by 10x Chromium platform was boost since 2017. Up to now, lots of zebrafish scRNA-seq data are available online, ranging from early embryo to matured adult, covering various organs. Therefore, I came out with an idea of using these data to construct a comprehensive development cells atlas for zebrafish. Note that, 10x scRNA-seq data has less variance cross different samples that may come from different labs, that is, less batch effects, particularly comparing to bulk RNA sequencing data. This argument was already proved. One more point emphasize, this character (less batch effects) may be benefited from UMI technology, which mark every single RNA sequence inside a cell with an unique molecule identifier, i.e. UMI. It simply counts the

UMI to determine a gene's expression, which different from using metrics like RPKM or FRKM. Through this instrument, noising come from PCR phase are avoided. It is a foundational point for this research.

I collected scRNA-seq data of zebrafish ranging from 12hpf to 8dpf, which almost cover all zebrafish embryo development periods. in addition, data of various organs are also collected. I am reaching this research goal from five aspects:

1. Processing and annotating each sample individually. Here, I use a subsample strategy to get highest resolution of clustering. then combining annotated cell cluster of all samples manually, after that, we get a general outline of a development tree. Clustering base on Nearest Neighbor is used by Seurat (a popular scRNA data analysis tool), but was proved to be not suitable for this type of data. Therefore, I consider to use hierarchical cluster or Bayes Classifiers, while the former can naturally reflect a tree structure and the latter can start from a priori distribution (the manually constructing tree).
2. Curating a known marker genes list. As mentioned above, there are plenty of papers involved in zebrafish. I used a scraping program to obtain all of the known marker genes from these published papers. Then, I use these known marker genes to verify the tree that was constructed in the first step. Also, we can use a subset of the marker genes to perform a heuristic clustering process.
3. Dissecting new cell type or new marker gene. Through step one and two, we obtain a reliable tree. In some circumstance, we can't annotate cell type in an individual sample confidentially, but can obtain extra information from the development tree. thus, we can find new cell type or new marker gene from the tree. Such as I discovered that Ig gene was expression at 3dpf embryo (standard genome are lack of annotation of Ig gene). This might indicate that naive B cell are Differentiated at very early stage of embryo.
4. Cross species analysis. Some research declare that 70 percent paralogous genes has found between zebrafish and human in hematopoiesis. besides, more other paralogous genes are discuss in different physiological processes. we therefore consider to perform a cross species analysis.
5. Developing a UI friendly website to exhibit the results for all results mentioned above. including development tree, curated marker genes, cell cluster annotations.

2. Malnutrition and Short Life Span in Zebrafish Genetically Devoid of Chitin-based Peritrophic Membrane

Chitin-based peritrophic membrane (PM), a structure involved in the gut barrier immunity and digestion in insects, was recently found to exist in fishes. So far almost nothing is known about the PM functions in fishes. Here we generated zebrafish mutants genetically devoid of PM. We found that PM creates at least two different niches for gut microbiota, but they are both destroyed by PM loss. I compared the gut microbiota between the wildtypes and the mutants to investigate if the balance of the bacterial population is disturbed by the PM loss. Six wildtypes and six mutants 16S rRNA libraries were prepared from the isolated zebrafish gut. Beta diversity metrics show that the mutant samples are highly divergent from those wildtype samples. This compositional change could be also reflected by the taxonomy bar plots and the alpha diversity metrics. I further identified a set of marker operational taxonomic units (OTU) which has significant deference between wildtype and mutant. These OTUs can offer a direction for research of bacterium in fish gut.

In this research, I finish analysing microbiome data. Firstly, I compared all of the popular tools for microbiome data analysis, including Usearch, Dada2, Vsearch etc. Then I build up a highly automatic pipeline with best performance. Secondly, because microbiome data has high heterogeneity of variance, I made a great many of trials and found a solid trend from a tranche of results. This step is benefited from the highly automatic pipeline built up from first step. The achievement of this research is about to publishing.

3. Training Received from College of Engineering, South China Agricultural University

From 2012 to 2016, I took undergraduate course from South China Agricultural University, majoring in

Embedding System (one of the subclass in Electronic and Information Engineering). I had lots of Embedding System designing practice during this period. Including Smart Home System (Undergraduate thesis), Airborne Multispectral Camera in Agriculture (CHN Patent), National Youth Campus Radio Education Competition (First Prize) etc.

To design an Embedding System, we usually starting with an one-ship computer, which is typically the foundational device for Embedding System. Then we load a scalable system to the chip, we usually choose Linux system for this highly scalable character. Then we develop application for this system. In most circumstance, we use C/C++ programming language to finish developing job because it is an interpreter-free language. Most Embedding Systems are sensitive to power dissipation and volume of the application.

Working practice

1. Developing Web Application for Security Department in Huizhi

Huizhi is an information technology company. Its principal business is developing various applications for security department of the government. During this period, I took part in several website developing projects (features of the application was not described here due to confidentiality agreement).

I was leading role of the development crew in two projects. We develop front end use JavaScript programming language, at the meantime, we make good use of some popular framework of front end, including Vue, Angular etc. These frameworks facilitate our development and benefit for tackling complex application demands. besides, I have good command of rear end developing skills. We use Java programming language to develop rear end. Also, we use Spring Boot or other framework to standardize our development. The core problem for rear end is its big data volume and real time requirement. To handle these problems, we design a high availability data transfer scheme. overall, I capable of developing both front end and rear end application independently.

2. Supporting Banking Business of HSBC Around the World

HSBC Development Center is an IT service department of HSBC. we deliver IT development service to business department of HSBC around the world.

My counterpart clients come from Malaysia. I am responsible for collecting demand from them and developing practical tools for them. for example, barcode extractor, extracting barcodes from high throughput trading letters. Besides, server job flow, archiving or delivering files automatically.