## 1 Journal list

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Read in depth</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Pub 2021-2022</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Pub 2020-2021</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>Best3 in semester1-2020</td>
<td>31</td>
</tr>
</tbody>
</table>
HLab PubMon: Publication monitoring in Huang Lab @ HKU.

This repository aims to keep track of our reading habits on recent publications.

Codes: https://github.com/StatBiomed/HLab-pubmon
Topics are mainly about but not limited to computational biology, bioinformatics, systems biology, biostatistics, machine learning, artificial intelligence.

Here are some journals that we often find papers interesting to many of us:

- Nature: https://www.nature.com
- Science: https://www.sciencemag.org
- PNAS: https://www.pnas.org
- Genome Biology: https://genomebiology.biomedcentral.com
- Genome Research: https://genome.cshlp.org
- Bioinformatics: https://academic.oup.com/bioinformatics
- Plos comp bio: https://journals.plos.org/ploscompbiol/
- Nucleic Acids Research: https://academic.oup.com/nar
- Nature biotech: https://www.nature.com/nbt/
- Nature methods: https://www.nature.com/nmeth/
- Nature genetics: https://www.nature.com/ng/
- Nature communications: https://www.nature.com/ncomms/
- Conferences in ML, e.g., UAI, NeurIPS …
- and more …

1.1 Read in depth

1.1.1 2020-09-14

We read CHISEL paper for single-cell copy number variation calling [zaccaria2020characterizing] in detail today.
1.1.2 2020-10-11

We read variational auto-encoder paper [kingma2013auto] in detail today, with brief discussion on its single cell application scVI [lopez2018deep].

1.1.3 2021-05-17

We read three papers: RPCI [liu2021robust], Cell-ID: [cortal2021gene], scPred [alquicira2019scpred], all about linear projection of single-cell transcriptome to lower dimensional space with singular value decomposition (SVD), the implementation technique for Principal component analysis (PCA).

We also referred more details on PCA and SVD by reading the following book chapters:


1.1.4 References


1.2 Pub 2021-2022

1.2.1 2021-12-10


### 1.2.2 2021-11-26


### 1.2.3 2021-11-12


### 1.2. Pub 2021-2022
Chapter 1. Journal list


1.2.6 2021-09-17


1.3 Pub 2020-2021

1.3.1 2021-07-26


1.3.3 2021-05-31


9. Jonathan R. Belyeu, Murad Chowdhury, Joseph Brown, Brent S. Pedersen, Michael J. Cormier, Aaron R. Quinlan, and Ryan M. Layer. Samplot: a platform for structural variant visual validation and auto-


1.3.4 2021-05-03


1.3.5 2021-04-12


### 1.3.6 2021-03-29


2. Juexin Wang, Anjun Ma, Yuzhou Chang, Jiantong Gong, Yue Xu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scGNN is a novel graph neural network framework for single-cell RNA-seq analyses. *Nature Communications*, mar 2021. URL: https://doi.org/10.1038/s41467-021-22197-x, doi:10.1038/s41467-021-22197-x.


1.3.7 2021-03-15


1.3.9 2021-02-01


1.3.11 2020-12-14


1.3.14 2020-11-02


1.3.16 2020-10-05


1.3.17 2020-09-21


1.4 Best3 in semester1-2020

1.4.1 Best 3


1.4.2 Rongting

1. CHISEL: It is really the relevant one to what I’m doing now. The integration of BAF and RDR of scDNA-seq data in chisel is a step forward in calling CNV more accurately. Though CaSpER also uses BAF and RDR information of scRNA-seq at the same time, the phasing strategy in chisel can be more helpful for getting more information.

2. Methods Review article in single cell data: The review outlines the major challenges in analyzing single-cell cancer genomics data and surveys the current computational tools available to tackle these. They further outline unsolved problems that may be considered as major opportunities for future methods development to help interpret the vast quantities of data being generated.

3. AlphaFold2: As the title of the share says, “game has changed in the field of solving protein structures!” It can be helpful in many research works. However, “This isn’t the end of something,” “It’s the beginning of many new things.”


1.4.3 Xianjie

1. Proteins prediction-Deep learning Cool stuff

2. scVelo Life process is assumed to be dynamically regulated. ScVelo uses kinetics to build the dynamical model trying to characterize and predict the change of gene expression level. I learned kinetics from the course of physical chemistry and it’s pleasantly surprised to find that how knowledge from textbook could apply to the real-world problem and benefit to the community.

3. Benefit from the CHISEL model, especially the EM part.


1.4.4 Yuanhua

1. AlphaFold2: it addressed a long-history problem on protein structure prediction with near perfect performance, and opens many exciting opportunities for functional and interaction analysis in biology. It is also a very encouraging example that data driven (often black-box) methods have great potentials on revealing complex patterns in science.

2. CHISEL: the multi-step phasing of the gremline SNPs effectively increases the power of using the allelic information to identify copy number variations. The modelling of the two-module data further motivates the integration pivot of multiple mutations and multiple omics data for cancer mutations.

3. scVelo: it offers an interesting extension of RNA velocity and nice implementation. The inferred splicing kinetic parameters through the dynamic model also inspires further investigation on its regulatory patterns.


1.4.5 Julia

1. It provided insight on factors affecting tumor development and progression, which was very useful to me to reference to when I was writing my literature review.

2. Also used as a reference when i was writing my literature review. cause it provided insight on different computational analysis method used for analysing cell heterogeneity.

3. Even though it is not really relevant to what I’m doing for my fyp, but it is probably one of the article that i understand the most regarding the methodology.


1.4.6 Jerry

1. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. I like it because this paper provided a new angle – Network Analysis to look into single cell BCR repertoire, which can clearly visualize the condition of diseased and healthy subjects. This paper also explains quite a bit on the mechanism of BCR differentiation and mutation.

2. A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. This paper described the B cell clonal expansion of multiple sclerosis and proved the effectiveness of B cell depletion therapy. This paper inspired me to compare the B cell expansion between diseased patient and healthy subject.

3. Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. I like it because this paper applied transcriptomic analysis into many parts of our body, which can reveal a clear pathway of immune cell differentiation. This paper also provide marker genes of many cell types, which are useful for future analysis.


1.4.7 Weizhong

1. Momentum Contrast for Unsupervised Visual Representation Learning. Reason: this work improve the contrastive learning by a simple queuing technique, intuitively and at the same time effectively.

2. Dissecting heterogeneous cell populations across drug and disease conditions with PopAlign. Reason: This paper focuses on an important question: how to map cell clusters across different condition. The creativity is mainly that we only build one probability density function, making the changes of coefficient explainable.

3. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. Reason: Of great biological significance, this paper covers a large scale population genetics analysis on 5′ UTR sequences, showing the strength of purification selection on the sequence motif.


1.4.8 Qiaochen

1. Strategies for Pre-training Graph Neural Networks Strategies for pretraining graph neural networks are summarized and validated, pretrained models can also be downloaded. Like Bert-triggered pretrained language models in the NLP filed, this work has the potential to benefit many downstream tasks.

2. scVelo: Generalizing RNA velocity to transient cell states through dynamical modeling An interesting extension of steady-state estimation of RNA velocity to maximum likelihood estimation.

3. Rethinking the Value of Labels for Improving Class-Imbalanced Learning: A lite-weighted yet effective strategy to tackle class-imbalance problems in classification.


1.4.9 Yoyo

1. Jointly defining cell types from multiple single-cell datasets using LIGER The tool may be useful for analyzing scRNA datasets, although I haven’t used it.

2. Identification of a novel cancer stem cell subpopulation that promotes progression of human fatal renal cell carcinoma by single-cell rna-seq analysis. This paper involves a couple of fancy tools I’m interested in, e.g. RNA velocity, cell-cell communication analysis.

3. Reprogramming roadmap reveals route to human induced trophoblast stem cells The scRNA-seq data is relevant to my research, and I am exploring this.

