任务规划

1. 任务分解：

1) 算法编写
2) 前端界面设计
3) 后端数据分析及处理
4) 测试，整合及上线

2. 时间规划

1) 算法编写：4 月——6 月中旬
2) 前端界面设计：6 月中旬——7 月中旬
3) 后端数据分析及处理：7 月中旬——8 月
4) 测试，整合及上线：8 月——9 月

3. 任务细节

任务 1：初步计划编写的算法：

1) Normalization：根据 reads count 计算 CPM，TMP，RPKM 和 FPKM(针对人类和小鼠转录组)
2) Denoising & Remove confounding factors：DCA 和 scVI[1, 2]
3) Clustering： SC3， DBSCAN， MCODE， Transitivity clustering， spectral clustering[3-7]，（NMF，HC & Kmenas）
4) Differential expression analysis：Deseq2，edgeR[8, 9]

5) Regulatory network identification：Module Networks，SCENIC，SCODE[10-12]

6) Stochasticity of transcription：[13, 14]

7) Construction of cell lineage：TSCAN，Monocle2，destiny[15-17]

任务 2-4：待规划

4. 任务分配

1) 两人一组，针对某个算法各自编写代码，完成后需互相确认以保证代码的正确。小组是动态变化的，针对某个算法组队。
2) 以单个算法为任务单位，每个小组承担单个算法编写，完成并通过小组成员及负责人确认后即可承担下一任务。

5. 协作方式及平台

1) 任务 1 采用 GitHub 平台进行协作，所有的 contribution 都会在 GitHub 上自动记录，也相对公平，请登录 OpenBiox 的 GitHub 主页查看。
2) 我会针对任务 1 的 7 项任务在小组 GitHub 项目 **scRNA-seq OnlineFlow** 中创建 master 分支及子任务文件夹，请每位成员自行创建子分支进行编写，详细请参考 GitHub 使用教程。
3) 任务代码需要进行详细注释并撰写使用说明文档

6. 编程语言与注意事项

1) 所有算法均采用 Julia 编写，请注意编程的格式规范以及代码注释。
2) 单细胞数据可以自行在网上检索下载，我也会上传公用的单细胞数据，供小

组测试使用。

3) 每项算法完成后需设计可视化展示方式，并编写代码。

4) 编写过程中可以尝试新的方法并撰写文章，完成后的代码可以包装成 Julia 包发布。

7. 参考资料

1) Julia 官网：https://julialang.org/

2) Julia computing：https://juliacomputing.com/

3) 算法主要涉及的书目：高等代数，泛函分析，概率论，微分方程，图论，贝叶斯统计，机器学习与深度学习

8. Reference

[1] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef, "Bayesian inference for a generative model of transcriptome profiles from single-cell RNA sequencing," *bioRxiv,* p. 292037, 2018.

[2] G. Eraslan, L. M. Simon, M. Mircea, N. S. Mueller, and F. J. Theis, "Single-cell RNA-seq denoising using a deep count autoencoder," *Nature communications,* vol. 10, no. 1, p. 390, 2019.

[3] V. Y. Kiselev *et al.*, "SC3: consensus clustering of single-cell RNA-seq data," *Nature methods,* vol. 14, no. 5, p. 483, 2017.

[4] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, 1996, vol. 96, no. 34, pp. 226-231.

[5] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab-an S4 package for kernel methods in R," *Journal of statistical software,* vol. 11, no. 9, pp. 1-20, 2004.

[6] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC bioinformatics,* vol. 4, no. 1, p. 2, 2003.

[7] T. Wittkop *et al.*, "Partitioning biological data with transitivity clustering," *Nature methods,* vol. 7, no. 6, p. 419, 2010.

[8] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome biology,* vol. 15, no.

12, p. 550, 2014.

[9] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics,* vol. 26, no. 1, pp. 139-140, 2010.

[10] E. Segal *et al.*, "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data," *Nature genetics,* vol. 34, no. 2, p. 166, 2003.

[11] S. Aibar *et al.*, "SCENIC: single-cell regulatory network inference and clustering," *Nature methods,* vol. 14, no. 11, p. 1083, 2017.

[12] H. Matsumoto *et al.*, "SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation," *Bioinformatics,* vol. 33, no. 15, pp. 2314-2321, 2017.

[13] A. J. Larsson *et al.*, "Genomic encoding of transcriptional burst kinetics," *Nature,* vol. 565, no. 7738, p. 251, 2019.

[14] J. K. Kim and J. C. Marioni, "Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data," *Genome biology,* vol. 14, no. 1, p. R7, 2013.

[15] Z. Ji and H. Ji, "TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis," *Nucleic acids research,* vol. 44, no. 13, pp. e117-e117, 2016.

[16] X. Qiu *et al.*, "Reversed graph embedding resolves complex single-cell trajectories," *Nature methods,* vol. 14, no. 10, p. 979, 2017.

[17] P. Angerer, L. Haghverdi, M. Büttner, F. J. Theis, C. Marr, and F. Buettner, "destiny: diffusion maps for large-scale single-cell data in R," *Bioinformatics,* vol. 32, no. 8, pp. 1241-1243, 2015.