

Manuscript Title

This manuscript ([permalink](#)) was automatically generated from [codingpoppy/multiomics_review_manubot@3660c1c](#) on April 1, 2022.

Authors

- **John Doe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [johndoe](#) ·  [johndoe](#)

Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [janeroe](#)

Department of Something, University of Whatever; Department of Whatever, University of Something

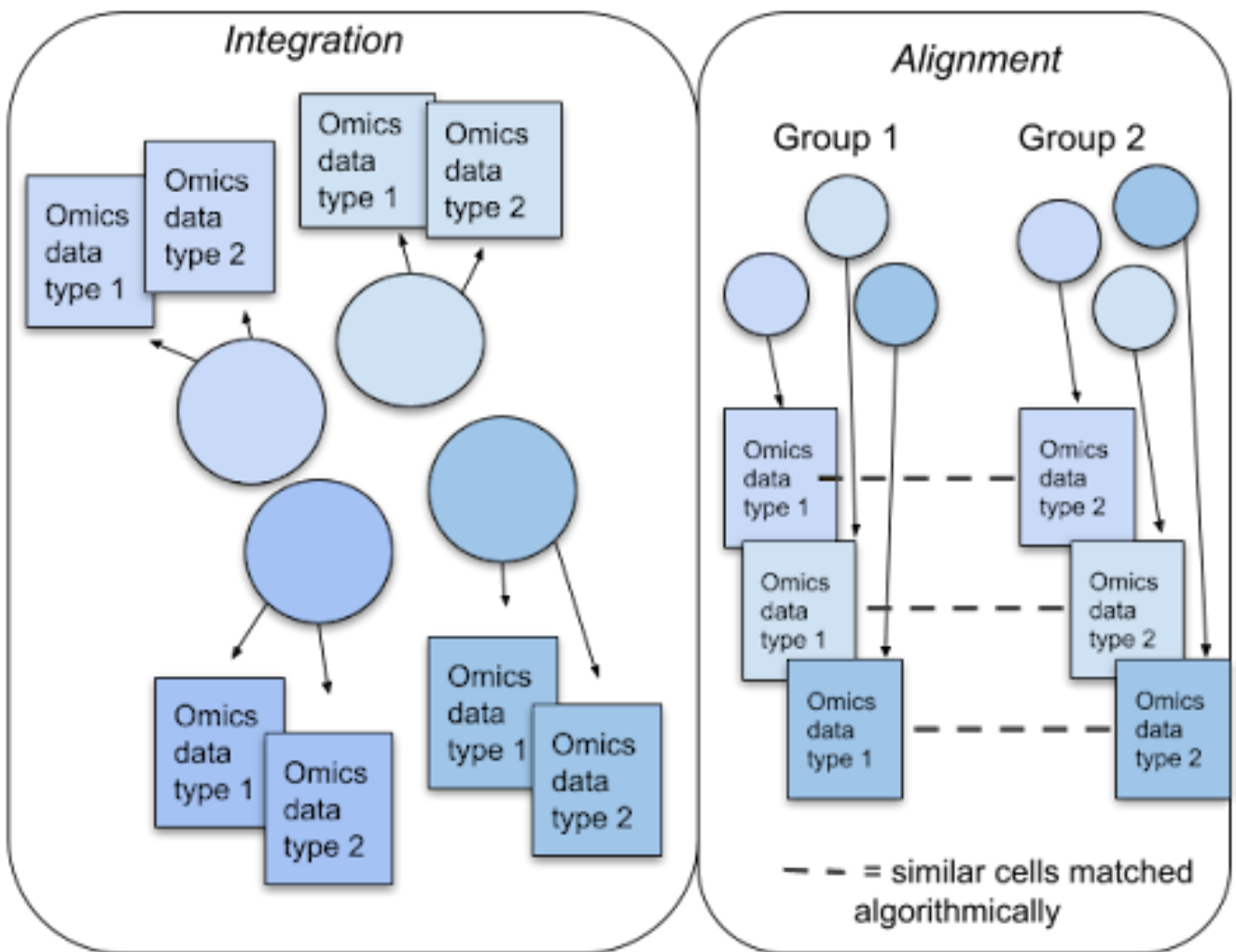
Abstract

Recently developed technologies to generate single-cell genomic data have made a revolutionary impact in the field of biology. Multi-omics assays offer even greater opportunities to understand cellular states and biological processes. However, the problem of integrating different -omics data with very different dimensionality and statistical properties remains quite challenging. A growing body of computational tools are being developed for this task, leveraging ideas ranging from machine translation to the theory of networks and representing a new frontier on the interface of biology and data science. Our goal in this review paper is to provide a comprehensive, up-to-date survey of computational techniques for the integration of multi-omics and alignment of multiple modalities of genomics data in the single cell research field.

Introduction

Single-cell sequencing technologies have opened the door to investigating biological processes at an unprecedentedly high resolution. Techniques such as DROP-seq [1] and 10x Genomics assays are capable of measuring single-cell gene expression, or scRNA-seq, in tens of thousands of single cells simultaneously. Measurements of other data modalities are also increasingly available. For example, single-cell ATAC-seq (scATAC-seq) assesses chromatin accessibility, and single-cell bisulfite sequencing captures DNA methylation, all from single cells. However, many of such techniques are designed to measure a single modality and do not lend themselves to multi-omics measurements. The way to combine information from such measurements is then to assay different -omics from different subsets of the same samples. By assuming that cells assayed by different techniques share similar properties, one can then use alignment methods to computationally aggregate similar cells across different omics assays and draw consensus biological inference.

Recently, however, a number of experimental techniques capable of assaying multiple modalities simultaneously from the same set of single cells have been developed. CITE-seq [2] and REAP-seq [3] measure proteins and gene expression. SNARE-seq [3,4], SHARE-seq [5] and sci-CAR [6] measure gene expression and chromatin accessibility, while scGEM [doi? 10.1038/nmeth.3961] measures gene expression and DNA methylation. For triple-omics data generation, scNMT [7] measures gene expression, chromatin accessibility and DNA methylation, and scTrio-seq [8,9] captures SNPs, gene expression and DNA methylation simultaneously. Integrative analysis of such data obtained from the same cells remains a challenging computational task due to a combination of reasons, such as the noise and sparsity in the assays, and different statistical distributions for different modalities. For clarity, we distinguish between integration methods that combine multiple -omics data from the set of the same single cells (Section I), from alignment methods designed to work with multi-modal data coming from the same tissue but different cells (Section II). The difference in their approaches is shown in Figure. {[fig-1?]}.



Multi-omics data can sometimes be sequenced from the same set of single cells (left); at other times, only the data sequenced from the same/similar sample, but different single cells are available (right). In the former case, we have the task of integrating the different data modalities (left); in the latter case, we need to first identify similar cells across the samples (right) - this is the computational task of alignment.

The application of data fusion algorithms for multi-omics sequencing data predates the single-cell technologies; bulk-level data have been integrated using a variety of computational tools as reviewed in [\[10\]](#). In this review, we aim to give a comprehensive, up-to-date summary of existing computational tools of multi-omics data integration and alignment in the single-cell field, for researchers in the field of computational biology. For more general surveys, the readers are encouraged to check other single-cell multi-omics reviews [\[11,12,13,14,15,16\]](#).

References

1. **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets**
Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, ... Steven A McCarroll
Cell (2015-05) <https://doi.org/f7dkxv>
DOI: [10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002) · PMID: [26000488](https://pubmed.ncbi.nlm.nih.gov/26000488/) · PMCID: [PMC4481139](https://pubmed.ncbi.nlm.nih.gov/PMC4481139/)
2. **Simultaneous epitope and transcriptome measurement in single cells**
Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, Peter Smibert
Nature Methods (2017-07-31) <https://doi.org/gfkksd>
DOI: [10.1038/nmeth.4380](https://doi.org/10.1038/nmeth.4380) · PMID: [28759029](https://pubmed.ncbi.nlm.nih.gov/28759029/) · PMCID: [PMC5669064](https://pubmed.ncbi.nlm.nih.gov/PMC5669064/)
3. **Multiplexed quantification of proteins and transcripts in single cells**
Vanessa M Peterson, Kelvin Xi Zhang, Namit Kumar, Jerelyn Wong, Lixia Li, Douglas C Wilson, Renee Moore, Terrill K McClanahan, Svetlana Sadekova, Joel A Klappenbach
Nature Biotechnology (2017-08-30) <https://doi.org/gghhzg>
DOI: [10.1038/nbt.3973](https://doi.org/10.1038/nbt.3973) · PMID: [28854175](https://pubmed.ncbi.nlm.nih.gov/28854175/)
4. **High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell**
Song Chen, Blue B Lake, Kun Zhang
Nature Biotechnology (2019-10-14) <https://doi.org/gghfzh>
DOI: [10.1038/s41587-019-0290-0](https://doi.org/10.1038/s41587-019-0290-0) · PMID: [31611697](https://pubmed.ncbi.nlm.nih.gov/31611697/) · PMCID: [PMC6893138](https://pubmed.ncbi.nlm.nih.gov/PMC6893138/)
5. **SHARE-seq reveals chromatin potential**
Dorothy Clyde
Nature Reviews Genetics (2020-11-17) <https://doi.org/gps358>
DOI: [10.1038/s41576-020-00308-6](https://doi.org/10.1038/s41576-020-00308-6) · PMID: [33204030](https://pubmed.ncbi.nlm.nih.gov/33204030/)
6. **Joint profiling of chromatin accessibility and gene expression in thousands of single cells**
Junyue Cao, Darren A Cusanovich, Vijay Ramani, Delasa Aghamirzaie, Hannah A Pliner, Andrew J Hill, Riza M Daza, Jose L McFaline-Figueroa, Jonathan S Packer, Lena Christiansen, ... Jay Shendure
Science (2018-09-28) <https://doi.org/gd4rk4>
DOI: [10.1126/science.aau0730](https://doi.org/10.1126/science.aau0730) · PMID: [30166440](https://pubmed.ncbi.nlm.nih.gov/30166440/) · PMCID: [PMC6571013](https://pubmed.ncbi.nlm.nih.gov/PMC6571013/)
7. **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells**
Stephen J Clark, Ricard Argelaguet, Chantierint-Andreas Kapourani, Thomas M Stubbs, Heather J Lee, Celia Alda-Catalinas, Felix Krueger, Guido Sanguinetti, Gavin Kelsey, John C Marioni, ... Wolf Reik
Nature Communications (2018-02-22) <https://doi.org/gc4q72>
DOI: [10.1038/s41467-018-03149-4](https://doi.org/10.1038/s41467-018-03149-4) · PMID: [29472610](https://pubmed.ncbi.nlm.nih.gov/29472610/) · PMCID: [PMC5823944](https://pubmed.ncbi.nlm.nih.gov/PMC5823944/)
8. **Single-cell multimodal profiling reveals cellular epigenetic heterogeneity**
Lih Feng Cheow, Elise T Courtois, Yuliana Tan, Ramya Viswanathan, Qiaorui Xing, Rui Zhen Tan, Daniel SW Tan, Paul Robson, Yui-Han Loh, Stephen R Quake, William F Burkholder
Nature Methods (2016-08-15) <https://doi.org/gps357>
DOI: [10.1038/nmeth.3961](https://doi.org/10.1038/nmeth.3961) · PMID: [27525975](https://pubmed.ncbi.nlm.nih.gov/27525975/)
9. **Single-cell multiomics sequencing and analyses of human colorectal cancer**

Shuhui Bian, Yu Hou, Xin Zhou, Xianlong Li, Jun Yong, Yicheng Wang, Wendong Wang, Jia Yan, Boqiang Hu, Hongshan Guo, ... Wei Fu
Science (2018-11-29) <https://doi.org/gfkwwwk>
DOI: [10.1126/science.aao3791](https://doi.org/10.1126/science.aao3791) · PMID: [30498128](https://pubmed.ncbi.nlm.nih.gov/30498128/)

10. **More Is Better: Recent Progress in Multi-Omics Data Integration Methods**
Sijia Huang, Kumardeep Chaudhary, Lana X Garmire
Frontiers in Genetics (2017-06-16) <https://doi.org/gcz6m3>
DOI: [10.3389/fgene.2017.00084](https://doi.org/10.3389/fgene.2017.00084) · PMID: [28670325](https://pubmed.ncbi.nlm.nih.gov/28670325/) · PMCID: [PMC5472696](https://pubmed.ncbi.nlm.nih.gov/PMC5472696/)
11. **Statistical single cell multi-omics integration**
M Colomé-Tatché, FJ Theis
Current Opinion in Systems Biology (2018-02) <https://doi.org/gfgkr6>
DOI: [10.1016/j.coisb.2018.01.003](https://doi.org/10.1016/j.coisb.2018.01.003)
12. **Integrative Methods and Practical Challenges for Single-Cell Multi-omics**
Anjun Ma, Adam McDermaid, Jennifer Xu, Yuzhou Chang, Qin Ma
Trends in Biotechnology (2020-09) <https://doi.org/ggzm72>
DOI: [10.1016/j.tibtech.2020.02.013](https://doi.org/10.1016/j.tibtech.2020.02.013) · PMID: [32818441](https://pubmed.ncbi.nlm.nih.gov/32818441/) · PMCID: [PMC7442857](https://pubmed.ncbi.nlm.nih.gov/PMC7442857/)
13. **Computational methods for the integrative analysis of single-cell data**
Mattia Forcato, Oriana Romano, Silvio Bicciato
Briefings in Bioinformatics (2020-08-06) <https://doi.org/gpfxqp>
DOI: [10.1093/bib/bbaa042](https://doi.org/10.1093/bib/bbaa042) · PMID: [32363378](https://pubmed.ncbi.nlm.nih.gov/32363378/) · PMCID: [PMC7820847](https://pubmed.ncbi.nlm.nih.gov/PMC7820847/)
14. **Computational principles and challenges in single-cell data integration**
Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, John C Marioni
Nature Biotechnology (2021-05-03) <https://doi.org/gjw92q>
DOI: [10.1038/s41587-021-00895-7](https://doi.org/10.1038/s41587-021-00895-7) · PMID: [33941931](https://pubmed.ncbi.nlm.nih.gov/33941931/)
15. **Computational strategies for single-cell multi-omics integration**
Nigatu Adossa, Sofia Khan, Kalle T Rytönen, Laura L Elo
Computational and Structural Biotechnology Journal (2021) <https://doi.org/gmh28b>
DOI: [10.1016/j.csbj.2021.04.060](https://doi.org/10.1016/j.csbj.2021.04.060) · PMID: [34025945](https://pubmed.ncbi.nlm.nih.gov/34025945/) · PMCID: [PMC8114078](https://pubmed.ncbi.nlm.nih.gov/PMC8114078/)
16. **Multi-omics integration in the age of million single-cell data**
Zhen Miao, Benjamin D Humphreys, Andrew P McMahon, Junhyong Kim
Nature Reviews Nephrology (2021-08-20) <https://doi.org/gmzwc3>
DOI: [10.1038/s41581-021-00463-x](https://doi.org/10.1038/s41581-021-00463-x) · PMID: [34417589](https://pubmed.ncbi.nlm.nih.gov/34417589/)