Check for updates

Editorial: Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data

Geng Chen [1,2]*

[1] Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, [2] Genecast Biotechnology Co., Ltd., Wuxi, China

Editorial on the Research Topic

Multimodal and Integrative Analysis of Single-Cell or Bulk Sequencing Data

Biological systems often involve the complex interactions among the molecules from different omics layers, including genome, methylome, transcriptome, proteome, metabolome, and even microbiome. At the genome level, diverse types of variants (e.g., single nucleotide variations, small insertions or deletions, and structural variations) that could be associated with a wide range of phenotypes or diseases may occur on the genome. For epigenome, it includes a variety of epigenetic modifications, such as covalent modifications on DNA and histones, chromatin accessibility and compaction, as well as the higher-order conformation of chromosome domains, which form an intricate regulatory network that can influence the chromatin structure and gene expression (Weinhold, 2006; Allis and Jenuwein, 2016). Exploration of the transcriptome was greatly revolutionized by RNA-seq technologies, which have gradually replaced traditional microarrays and provided unprecedented insights into the dynamics and complexity of gene expression (Costa et al., 2010; Stark et al., 2019). Specifically, many long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) were found to have critical regulatory functions in diverse biological processes (Marchese et al., 2017; Xiao et al., 2020). Proteins encoded by mRNAs are generally organized into higher-order structures and networks to perform catalytic, synthetic, and regulatory functions at specific times and locations (Aebersold and Mann, 2016). Mass spectrometry (MS)-based methods [such as liquid chromatography-MS/MS (LC-MS/MS)] greatly revolutionized proteome profiling and largely facilitated the dissection of complex biological processes and phenotypes (Angel et al., 2012). Furthermore, metabolome can theoretically link the genome, transcriptome, and proteome to phenotype (Misra et al., 2018). The levels and relative ratios of metabolites could generally reflect the metabolic functions, thus abnormal perturbations that beyond the normal range may indicate certain diseases (Hasin et al., 2017). Additionally, microbiomes may also significantly contribute to the biology and diverse phenotypes of their partner organisms, which can reveal the interactions between the genome and environment of the host organism (Knight et al., 2017; Lynch and Hsiao, 2019). Therefore, multi-omics analysis can promote the development of systems biology, which is essential for comprehensively investigating the dynamic changes and interactions of cellular molecules as well as understanding the underlying mechanisms (**Figure 1**).

In this research topic, a number of research teams conducted integrative analyses to explore the molecular mechanisms or identify potential biomarkers for certain diseases/disorders. Chen et al. systematically investigated the expression profiles of placenta accreta spectrum (PAS) at both transcriptomic and proteomic levels, which provided novel insights into the underlying molecular mechanism of PAS. Through joint analysis of the interaction networks among miRNAs, mRNAs, and lncRNAs, Wang et al. identified novel potential prognostic markers for luminal breast cancer patients. Sun et al. revealed that HIF-1α pathway-related lncRNA-HEIPP (high expression in preeclampsia placenta) could play an important role in the pathogenesis of preeclampsia based on the multi-omics exploration. Wang et al. performed an integrative analysis of the underlying mechanisms of noise-induced hearing loss (NIHL) and suggested that the inflammatory pathways are closely associated with the auditory organ changes of NIHL. Zhang et al. found that G-quadruplexes could be potential targets for the drug discovery of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by systematically analyzing the non-canonical secondary structures of G-quadruplexes in both positive and negative-sense strands of SARS-CoV-2. Wang et al. revealed that the expression profile of IGFBP7 could be a potential biomarker for vasculature in response to traumatic brain injury and TGFβ signaling might be closely correlated to the upregulation of IGFBP7. Yuan et al. systematically explored the expression patterns of major depressive disorder (MDD) and uncovered that Ephrin signaling and Ras protein signal transduction could be associated with the MDD pathogenesis. Wang et al. identified potential diagnostic and prognostic biomarkers for colorectal cancer based on an integrative analysis of the datasets from different public databases. Moreover,

Xiong et al. developed an integrative computational approach of ASDmiR for identifying the potential pathogenic genes, networks, and modules correlated with autism spectrum disorder. Overall, these studies performed joint analyses on the data from distinct omics layers, which gained novel insights into different diseases.

Although the development and innovation of single-cell approaches will gradually lead to a shift from bulk integrative analysis to a detailed exploration of individual cells, bulk strategies are complementary with single-cell approaches to get whole-system and cell-based perspectives and mechanisms for health and disease. One main limitation of bulk profiling methods is that they cannot accurately disentangle the cellular heterogeneity, thus single-cell exploration is essential for better elucidating the cellular behaviors and cell-to-cell variations for both basic and clinical research. However, compared to conventional bulk approaches, the single-cell technologies currently available for distinct omics are still in the early stages of development, which are often with relatively lower capture efficiency and higher technical noise (Chen et al., 2019). The improvement of the experimental procedures for single-cell protocols will reduce the technical noise and sparsity of multi-omics data, and increase the sensitivity and specificity of multimodal dissection. Moreover, conducting the single-cell study on a large number of samples is still expensive and time-consuming, the decreasing cost and simplified operation of single-cell profiling will make the multi-omics analysis more affordable and practicable. Since bulk strategies are feasible to study large-scale samples, combing bulk and single-cell data with deconvolution methods could be a good solution to investigate a multitude of individuals in a cell-type-specific manner (Li et al., 2020). Besides, the computational

FIGURE 1 | A simple schematic view of systems biology researches based on multi-omics data.
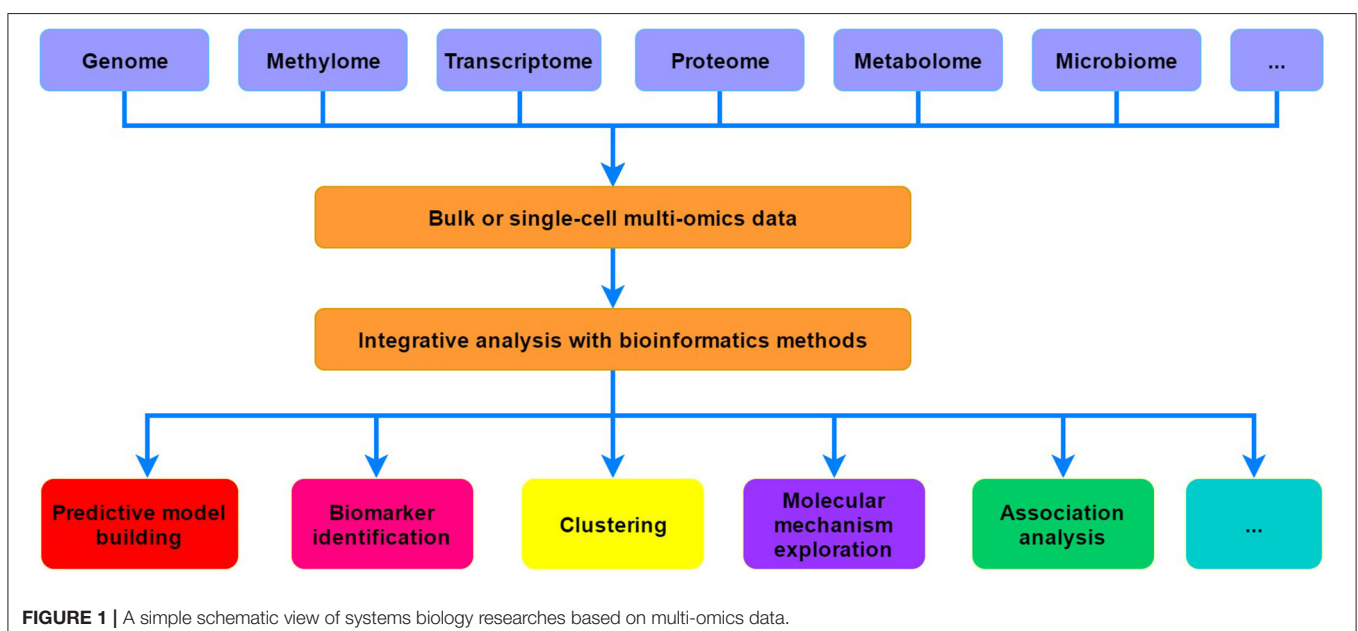
methods for joint analysis of single-cell multimodal data are just emerging in recent years, novel bioinformatics tools are required to more efficiently integrate single-cell multi-omics data. On the other hand, existing bulk and single-cell sequencing protocols are mainly based on next-generation sequencing technologies. We envision that the utilization of third-generation sequencing approaches [e.g., Nanopore (Garalde et al., 2018)] to produce super long reads in the bulk or single-cell omics studies will largely benefit the downstream data analysis and facilitate the development of systems biology.

AUTHOR CONTRIBUTIONS

GC conceived and wrote the manuscript.

FUNDING

REFERENCES

Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. doi: 10.1038/nature19949

Allis, C. D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17, 487–500. doi: 10.1038/nrg.2016.59

Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., et al. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* 41, 3912–3928. doi: 10.1039/c2cs15331a

Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-Seq technologies and related computational data analysis. *Front. Genet.* 10:317. doi: 10.3389/fgene.2019.00317

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* 2010:853916. doi: 10.1155/2010/853916

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201–206. doi: 10.1038/nmeth.4577

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18:83. doi: 10.1186/s13059-017-1215-1

Knight, R., Callewaert, C., Marotz, C., Hyde, E. R., Debelius, J. W., McDonald, D., et al. (2017). The microbiome and human biology. *Annu. Rev. Genomics Hum. Genet.* 18, 65–86. doi: 10.1146/annurev-genom-083115-022438

Li, Y., Xu, Q., Wu, D., and Chen, G. (2020). Exploring additional valuable information from single-cell RNA-Seq data. *Front. Cell Dev. Biol.* 8:593007. doi: 10.3389/fcell.2020.593007

Lynch, J. B., and Hsiao, E. Y. (2019). Microbiomes as sources of emergent host phenotypes. *Science* 365, 1405–1409. doi: 10.1126/science.aay0240

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18:206. doi: 10.1186/s13059-017-1348-2

Misra, B. B., Langefeld, C. D., Olivier, M., and Cox, L. A. (2018). Integrated omics: tools, advances, and future approaches. *J. Mol. Endocrinol.* 62, R21–R45. doi: 10.1530/JME-18-0055

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2

Weinhold, B. (2006). Epigenetics: the science of change. *Environ. Health Perspect.* 114, A160–A167. doi: 10.1289/ehp.114-a160

Xiao, M. S., Ai, Y., and Wilusz, J. E. (2020). Biogenesis and functions of circular RNAs come into focus. *Trends Cell Biol.* 30, 226–240. doi: 10.1016/j.tcb.2019.12.004