Short Communication

Recent Advance in Biomedical Omics Data Analysis

Xue Jiang*

Department of Artificial Intelligence, Nankai University, China

*Corresponding author: Xue Jiang, Department of Artificial Intelligence, Nankai University, China.

Short Communication

Sequencing technology is dominant in the study of complex human genetic diseases. Identified risk genes and other biomarkers will make genetic counselling, risk assessment and avoidance, and disease diagnosis and treatment possible. Benefit from the discovered biomarkers, many diseases could be accurate diagnosed and prevented, as well as accurate treated. Nevertheless, the molecular mechanisms of many neuropsychiatric disorders are still unclear and there are no effective treatments for these neuropsychiatric disorders. The rapid development of whole genome sequencing technology, large amounts of genetic data are generated concerning larger cohorts with diverse neuropsychiatric disorders [1], making it possible to detect various genetic variants, i.e. SNVs, InDels, and SVs. Genetic research of these disorders are focused on genome-wide association analysis (GWAS), pedigree analysis, twin family studies, trio family studies, etc., Over the last decades, GWAS have identified several hundred loci for neuropsychiatric disorders successfully [2,3]. Whole exome/genome sequencing have greatly facilitated the identification of risk de novo variants [4]. Besides, with the increasing amount of patient cohorts, more and more risk mutations have been discovered [5].

However, there is still a deep gap between the explaining the molecular mechanisms and thousands of risk gene locis identified in the past years using GWAS. For example, it has been reported that 142 risk mutation locis contribute to schizophrenia, while most of these locis do not contain any genes and it is not clear how they related to the disorder. Traditional single-dimensional genomic study are less useful in explaining the mechanisms of these disorders [6-7]. However, with the decreasing cost of sequencing technologies, large amounts of omics data become available using RNA-seq [8]. ChIp-seq [9]. Methyl-seq [10] Hi-C[11-12] and ATAC-seq [13-14] etc Whole exome/genome sequencing could generate the sequence information of genes. RNA-seq technology could be used for genotyping and identification of transcription factor binding sites, as well as generating gene expression data. ChIp-seq (Chromatin immunoprecipitation followed by sequencing) could be used for genome-wide profiling of DNA-binding proteins, histone modifications or nucleosomes. Methyl-seq could be used for detecting and quantify DNA methylation. Hi-C (high throughput chromatin conformation capture) technology is a method to study the three-dimensional architecture of genomics and could be used to comprehensively detect chromatin interactions in the mammalian nucleus. And ATAC-Seq (assay for transposase-accessible chromatin with high throughput sequencing) could be used for mapping chromatin accessibility genome-wide. With all these biotechnologies, we could detect mutations in gene sequence, gene regulation and epigenetic mechanisms, reactivities of methylated and nonmethylated cytosines, chromatin interactions, and variations in gene expressions under complex disease phenotypes.

Recently, the accumulation of multi-omics data and single-cell sequencing data have greatly promoted the development of machine learning algorithms, statistical learning methods, and deep learning methods. Integrative analysis of multi-omics data generated by various sequencing technologies is helpful for screening robustness biomarkers related to diseases. Diagnosis and treatment should be further personalized according to multi-dimensional information [15]. Nowadays, applications of machine learning and deep learning methods are becoming ubiquitous in biology and encompass not only risk gene identification, but also protein binding prediction, disease subtypes classification, and transcriptional regulatory networks characterization, etc. [16,17]. In the near future, these intelligent computational methods should be further used to multi-omics data integrated analysis.

References

1. Van Dijk E L, Auger, Helene, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genet 30(9): 418-426.

2. Gandal M J, Leppa V, Won H, Parikshak NN, Geschwind DH, et al. (2016) The road to precision psychiatry: translating genetics into disease mechanisms. Nat Neurosci 19(11): 1397-1407.

3. Gandal M J, Haney J R, Parikshak N N, Leppa V, Ramaswami G, et al. (2018) Shared Molecular Neuropathology across Major Psychiatric Disorders Parallels Polygenic Overlap. Science 359(6376): 693-697.

4. Turner T N, Yi Q, Krumm N, Huddleston J, Hoekzema K, et al. (2016) Denovo-db: a compendium of human de novo variants. Nucleic Acids Res 45: 804-811.

5. Wu Y, Zeng J, Zhang F, Futao Zhang, Zhihong Zhu, et al. (2018) Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nature Communications 9(1): 918.

6. Geschwind D H, Flint J (2015) Genetics and genomics of psychiatric disease. Science 349(6255): 1489-1494.

7. Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, et al. (2009) GWAS Consortium Coordinating Committee Genomewide Association Studies: History, Rationale, and Prospects for Psychiatric Disorders. Am J Psychiatry 166(5): 540-556.

8. Marioni J C , Mason C E , Mane S M , Stephens M, Gilad Y, et al. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18(9): 1509-1517.

9. Park P J (2013) ChIP-Seq: advantages and challenges of a maturing technology. Nat Rev Genet 10(10): 669-680.

10. Khanna A, Czyz A, Syed F (2013) EpiGnome[trade] Methyl-Seq Kit: a novel post-bisulfite conversion library prep method for methylation analysis. Nature Methods 10(10).

11. Van B N L, Erez L A (2010) Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. Journal of Visualized Experiments 39(39): 292-296.

12. Belton J M, Mccord R P, Gibcus J, Naumova N, Zhan Y, et al. (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58(3): 268-276.

13. Buenrostro JD, Wu B, Litzenburger U M, Ruff D, Gonzales ML, et al. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523(7561): 486-490.

14. Buenrostro J D, Wu B, (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol 109(1): 21.29.1-21.29.9.

15. Deo R C (2015) Machine Learning in Medicine. Circulation 132(20): 1920-1930.

16. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ, et al. (2018) Next-Generation Machine Learning for Biological Networks. Cell 173(7): 1581–1592.

17. Zou James, Huss M, Abid A, Mohammadi P, Torkamani A, et al. (2019) A Primer on Deep Learning in Genomics. Nat Genet 51(1): 12–18.