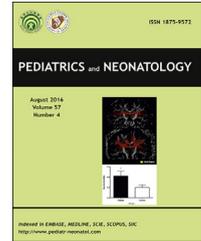


Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: <http://www.pediatr-neonatol.com>

Review Article

# Genomic technologies to improve variation identification in undiagnosed diseases

Joseph T.C. Shieh\*

University of California San Francisco, Division of Medical Genetics, Department of Pediatrics, Institute for Human Genetics, Benioff Children's Hospital, University of California San Francisco, California, USA

Received Oct 5, 2022; accepted Oct 12, 2022

Available online ■ ■ ■

## Key Words

genomic structural variation;  
inheritance patterns;  
integrative techniques;  
medical informatics;  
rare diseases (3-5, alphabetical order, MeSH)

Human genome variation has increasingly posed challenges and opportunities for patients, medical providers, and an increasing group of stakeholders including advocacy groups, disadvantaged communities, public health experts, and scientists. Here, advances in genomic sequencing and mapping technologies are discussed with particular attention to the increasing ability to detect personal and population genome variation and the potential for accurate integration of variation into health and disease-related care. Genome mapping, one technique used to create genome map scaffolds, has now been combined with long read sequencing. New technologies have led to improved variation detection, including cryptic structural variation and diverse variants with different degrees of disease association. Combined with advances in automated and medical interpretations, variation detection is increasingly being applied in healthcare. These advances promise to make disease diagnostics more rapid, and potentially more accessible, to those with medical needs. Consequentially, the need for medical genetics and genomics experts is increasing. Here, the opportunities and potential challenges for application of genome-scale variation detection in disease are examined. (<300 words).

Copyright © 2022, Taiwan Pediatric Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Genomic technologies with increased sensitivity to detect variation

While classical mapping and sequencing technologies have existed for some time, technologies that interrogate the genome are advancing rapidly. Thus, current genomic

technologies are high-throughput data generators, with more variant detection, long chromosomal phasing and increasing molecular resolution. Although traditional genomic sequencing pipelines excel at identifying small nucleotide variants, the ability to comprehensively detect all copy number variants (structural variation) has been

\* Corresponding author.

E-mail address: [Joseph.Shieh2@ucsf.edu](mailto:Joseph.Shieh2@ucsf.edu).

<https://doi.org/10.1016/j.pedneo.2022.10.002>

1875-9572/ Copyright © 2022, Taiwan Pediatric Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

more limited. Genome analysis pipelines can have limitations in resolution, leaving gaps in genome variation detection. Variants can be too large to detect using shorter read sequencing and yet too small to be seen by classic array technology and cytogenetic visualization. Advances in both sequencing platforms and software pipelines have emerged to address these limitations. First, long read genome sequencing provides an improved opportunity for copy number variant detection. Recent studies have demonstrated its utility in both small and large variant detection. In addition, the combination of long read sequencing with genome mapping technologies, such as optical mapping, have led to both efficient chromosome level assembly and sequence-level resolution (Levy-Sakin, Mostovoy, Shieh). Optical mapping relies on labeling of genomic markers, allowing for single molecule assembly or alignment, which can be used for scaffolding and more sensitive detection of inversions and mid-sized insertions (Xiao, Wong). With longer reads and new mapping techniques, phasing, or the ability to determine one chromosome from another, is now an important tool (Uppuluri). The *cis* or *trans* nature of variants is of particular importance in medical genetics or for applications for other diploid systems. Phasing can even replace parental samples in some cases of recessive disorders, if parental testing is not possible. With polyploidy or with repetitive regions of the genome, these are increasingly being resolved by these technologies (Wong, Young, Gamba). Given the increase in options for generating more complete genomic information, it is important to consider the technologies and their potential application.

## 2. Detecting different types of variants simultaneously

Disease annotation databases and genomic disease nomenclature continue to evolve as the number of diseases and genes associated with diseases increase (Biesecker). Yet, there are still patients with undiagnosed diseases. Both increased awareness of genetic disorders among clinicians as well as better tools for annotating and narrowing differential diagnoses are needed. In some cases, particularly when the differential diagnosis based on the clinical presentation suggests copy number variation or repeat variants, new technologies that detect multiple types of variation are helpful. For example, genome sequencing, with the appropriate pipeline, can resolve areas of high sequence homology or sequence repeats such as in fragile X, specific ataxias, or spinal muscular atrophy. Furthermore techniques such as optical mapping can be used to detect chromosome-scale translocations, inversions, and copy number variants (Hannon). Uniparental disomy, known to occur at some frequency in the unselected population, can also be detected and is important to identify in imprinting-region conditions including Prader Willi. Methylation, also important in imprinting, can be assessed by some long reach technologies. Optical mapping is useful in prenatal or pre-implantation contexts as it can be used to detect aneuploidy, translocation and large copy number variation simultaneously. Cancer diagnostic labs commonly assess

breakpoint detection and fusions. These may also be amenable to the new technologies. In cancer and cancer-predisposition conditions, tissue diagnostics as well as germline diagnostics have been important in patient stratification, surveillance and treatment, as well as in understanding mechanisms of disease. For example, somatic homozygosity is increasingly being recognized in tumor suppressors and in cancer-like states such as neurofibromatosis and vascular overgrowth (Tong). Higher-depth genomic sequencing technologies have also been applied to detect variants with smaller allele fractions. In cancer diagnostics, this may be particularly important to detect tumor evolution and potential therapeutic resistance. Even in normal phenotypic states the identification of constitutional mosaicism is increasing with these new technologies. Thus, the application of these new technologies in both cancer and constitutional conditions is revealing heterogeneity in genomic composition and permitting characterization of the underlying biology of variation (Kumar, Snellings). This is an important first step towards understanding the implications of germline and somatic mosaicism in normal and in disease states. The volumes of sequence data now available however have led to more unsolved questions in biology. For example, when recurrent *de novo* variants are observed in disease: Why are specific hotspots observed (e.g. in Myhre syndrome or in ferritinopathies associated neurologic disease). Is recurrent variation due to intrinsic sequence or structure? Or are these related to positive selection? Similarly, for small variants the mechanisms remain relatively unexplored despite their increasing frequent description. The integration of haplotype information from sequence data may lead to further answers.

## 3. Genome interpretation and importance of variant databases

An individual's genomic information is most useful when coupled to gene and variant level annotation. Population-based genome databases and laboratory-submitted variant annotations are two sources that have contributed substantially to the understanding of variants and their potential effects. As aggregate data on genome variation coupled with phenotypic information increases, both for small variants and copy number variants, it should lead to better medical interpretation and widespread applicability (Hsueh). For example, population variation data indicate the observed frequency of missense variants for a particular gene, while laboratory-reported variants indicate potential purported significance of specific variants. These genomic resources have facilitated the development of automated or semi-automated genome variation annotation pipelines, which have improved speed in clinical applications. One example is hospital-based rapid genome sequencing, where diagnostic variants are being identified in hours or days (e.g. intensive care unit sequencing or rapid sequencing for undiagnosed diseases), whereas more traditional pipelines with manual annotation can take significantly longer.

Generation and annotation of data is not, by itself, sufficient for medical application. Its diagnostic use

requires molecular and medical expertise, integration of clinical information, and often iterative phenotyping. Although universal access to technology and complete genomic databases are clear challenges, solving these will only provide one part of a genomic-based health care system. Equally important are clinical sequence experts and front-line clinicians who are versed in genomic medicine. Indeed, there is an apparent shortage of such experts. Although health care entities and governmental institutions are attempting to increase the workforce, they are challenged by an evolving field that touches on both the preventative and treatment-based aspect of medical care. Those working in genomic medicine and with the technologies promise to address both aspects. The realization is in its early stages.

#### 4. Opportunities to integrate genomic with additional technologies

How genomic alterations lead to diverse transcriptome effects is an important outstanding question that may require further -omics level characterization. Single cell RNA sequencing technologies (scRNA) are a potentially powerful tool to help address this question. Initial scRNA technologies relied on 5' or 3' sequence, which could be used for rapid, inexpensive, survey of short transcript sequences, ideal for highly-expressed genes and cell type description (Wang). Longer RNA sequencing technologies are increasingly being used, with full-length transcriptomes and detection of rare transcripts that were previously missed by other technologies. More complete transcriptomics at single cell resolution may be crucial to understanding genomic variation and cellular diversity. To realize this goal, further work is needed to integrate DNA and RNA technologies at single cell resolution. While these new technologies provide many novel opportunities, several challenges remain. Increasing investment in genomics, data science and relevant expertise will move these technologies toward routine practical diagnostics.

#### 5. Genomics and the future

Genome-based medical applications are often divided into two areas: rare diseases and common diseases. The division between these areas may be more historical and based on technological limitations at the time. Feasible disease mapping led to Mendelian disease studies, and SNP-association led to common disease variation studies. Increasingly, these areas of study are merging. Mendelian disease is clearly more complex than the assessment of binary traits, and genes involved in common diseases are also found in Mendelian forms. Therefore, biobanking efforts that include both genomic rare and common variants, as well as detailed trait information have had increasing utility (Wei). Indeed, knowledge of modifiers in Mendelian traits and polygenic risk scores have been active areas of research, and these may lead to further statistical methods for association or to quantify risk (Blair, Wei). Given the increasing relevance of genomics, educational institutions

need to teach and train individuals in precision medicine opportunities and complexities. Practitioners in all medical disciplines need exposure to the complexities and evolving nature of genomics. Certain specialties may become increasingly dependent on molecular testing-based or informatic-based diagnostics (Chunduru, Innes). For this transition to occur there needs to be an integration of genomics and data science with medical expertise. Similar to the integration of radiology and imaging technology, this integration will transform medicine. With thousands of genes associated with diseases, rare diseases have become collectively common, and common disease genetics are continuing to evolve. Increasingly the challenges in the field are also relevant for front-line practitioners as they may implement technologies or work with referral care centers. The equitable availability and implementation of genomics and genomics-based care is a challenge for healthcare systems with primary/tertiary care integrative efforts and distribution of health resources by geography (Penon-Portmann).

Beyond diagnosis and disease stratification, the question remains how quickly will these advances lead to new therapeutics for a growing list of disorders. Although the diagnostic and treatment technologies have great potential, their possibilities are still constrained by our knowledge of the effects of variants on cells and organ systems. While genomic editing technologies such as CRISPR have been proven in a cellular context, their application for diverse disease states requires careful disease selection and safe and effective in vivo delivery systems. Furthermore, genomic editing technologies require a solid understanding of whether normal or variant alleles require up- or downregulation, or correction, and how modulation and dosage affect the cell and organ systems. Whether these challenges can be overcome while still ensuring equity in healthcare delivery is still a concern. Indeed, since the initial completion of the first genome maps, the field has realized that the diversity of genomes from across populations is important in understanding health and disease. Whether it can also promote equity is a question. Although genomics has grown in select healthcare systems and in individual regions and countries, many populations still have limited access to basic genetic diagnostics, risk prediction and preventative risk-based care. These limitations have been observed in both high and low socioeconomic countries with disparities. Therefore, one goal of precision medicine should be to improve equitable access to genomics and genomics-based care. These challenges are therefore of increasing relevance to health care systems and economies.

#### Declaration of competing interest

The author has no conflicts of interest to declare.

#### Acknowledgments

We thank NORD (National Organization for Rare Disorders) and the NEHI Research Foundation.

## Further Reading

1. Biesecker LG, Adam MP, Alkuraya FS, Amemiya AR, Bamshad MJ, Beck AE, et al. A dyadic approach to the delineation of diagnostic entities in clinical genomics. *Am J Hum Genet* 2021;108:8–15.
2. Blair DR, Hoffmann TJ, Shieh JT. Common genetic variation associated with Mendelian disease severity revealed through cryptic phenotype analysis. *Nat Commun* 2022;13:3675.
3. Chien YH, Chiang SC, Weng WC, Lee NC, Lin CJ, Hsieh WS, et al. Presymptomatic diagnosis of spinal muscular atrophy through newborn screening. *J Pediatr* 2017;190:124–129.e1.
4. Chunduru P, Phillips JJ, Molinaro AM. Prognostic risk stratification of gliomas using deep learning in digital pathology images. *Neurooncol Adv* 2022;4:vdac111.
5. Gamba R, Mazzucco G, Wilhelm T, Velikovsky L, Salinas-Luypaert C, Chardon F, et al. Enrichment of centromeric DNA from human cells. *PLoS Genet* 2022;18(7):e1010306.
6. Hanlon VCT, Lansdorp PM, Guryev V. A survey of current methods to detect and genotype inversions. *Hum Mutat* 2022;43:1576–89.
7. Hsueh HW, Weng WC, Fan PC, Chien YH, Yang FJ, Lee WT, et al. The diversity of hereditary neuromuscular diseases: experiences from molecular diagnosis. *J Formos Med Assoc* 2022;S0929–6646.
8. Innes AM, Lynch DC. Fifty years of recognizable patterns of human malformation: insights and opportunities. *Am J Med Genet A* 2021;185:2653–69.
9. Kumar A, Mohamed E, Tong S, Chen K, Mukherjee J, Lim Y, et al. CXCL14 promotes a robust brain tumor-associated immune response in glioma. *Clin Cancer Res* 2022;28:2898–910.
10. Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AKY, McCaffrey J, et al. Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nat Commun* 2019;10:1025.
11. Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, et al. A hybrid approach for de novo human genome sequence assembly and phasing. *Nat Methods* 2016;13:587–90.
12. Penon-Portmann M, Chang J, Cheng M, Shieh JT. Genetics workforce: distribution of genetics services and challenges to health care in California. *Genet Med* 2020;22:227–31.
13. Shieh JT, Penon-Portmann M, Wong KHY, Levy-Sakin M, Verghese M, Slavotinek A, et al. Application of full-genome analysis to diagnose rare monogenic disorders. *NPJ Genom Med* 2021;6:77. Erratum in: *NPJ Genom Med*. 2021;6:88.
14. Snellings DA, Gallione CJ, Clark DS, Vozoris NT, Faughnan ME, Marchuk DA. Somatic mutations in vascular malformations of hereditary hemorrhagic telangiectasia result in Bi-allelic loss of ENG or ACVRL1. *Am J Hum Genet* 2019;105:894–906.
15. Tong S, Devine WP, Shieh JT. Tumor and constitutional sequencing for neurofibromatosis type 1. *JCO Precis Oncol* 2022;6:e2100540.
16. Uppuluri L, Wang Y, Young E, Wong JS, Abid HZ, Xiao M. Multiplex structural variant detection by whole-genome mapping and nanopore sequencing. *Sci Rep* 2022;12:6512.
17. Wang L, Babikir H, Müller S, Yagnik G, Shamardani K, Catalan F, et al. The phenotypes of proliferating glioblastoma cells reside on a single Axis of variation. *Cancer Discov* 2019;9:1708–19.
18. Wei CY, Yang JH, Yeh EC, Tsai MF, Kao HJ, Lo CZ, et al. Genetic profiles of 103,106 individuals in the Taiwan Biobank provide insights into the health and history of Han Chinese. *NPJ Genom Med* 2021;6:10.
19. Wong KHY, Ma W, Wei CY, Yeh EC, Lin WJ, Wang EHF, et al. Towards a reference genome that captures global genetic diversity. *Nat Commun* 2020;11:5482.
20. Xiao M, Phong A, Ha C, Chan TF, Cai D, Leung L, et al. Rapid DNA mapping by fluorescent single molecule detection. *Nucleic Acids Res* 2007;35:e16.
21. Young E, Abid HZ, Kwok PY, Riethman H, Xiao M. Comprehensive analysis of human subtelomeres by whole genome mapping. *PLoS Genet* 2020;16:e1008347.