Biological Psychiatry

Commentary

Mosaic Chromosomal Alterations/Somatic Copy Number Variations: A New Frontier in Genetic Association Studies of Complex Diseases

Dawei Li

Genome-wide association studies have identified numerous common and rare germline genetic variants associated with complex diseases, including single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and other constitutional structural variations. However, a significant portion of disease susceptibility remains unexplained, often referred to as missing heritability. One area of growing interest is genetic variation arising postfertilization, known as mosaic mutation, which occurs during cellular division. Whereas cells carrying detrimental mutations may be eliminated through repair mechanisms, apoptosis, or immune surveillance, others can pass their mutations to daughter cells. Consequently, during early embryonic development, one or more postzygotic mutations are retained with each cell division. As development progresses, these mutations accumulate, leading to diverse genomic landscapes across cells. Consequently, most cells end up carrying a unique genome. While many mosaic mutations may be neutral, certain mutations can be pathogenic. Mosaicism can occur in both somatic and germline cells, with somatic mosaicism recently gaining attention for its potential role in neurogenetic diseases.

Postzygotic mutations encompass all major mutation classes, including chromosomal aneuploidy, large-scale structural abnormalities, CNVs, small insertions/deletions, and single nucleotide variants. Among these, mosaic chromosomal alterations, also known as somatic CNVs (sCNVs), often result from chromosomal instability during embryogenesis. These mutations primarily occur postzygotically or in early embryonic development, occasionally arising from the partial postzygotic rescue of meiotic errors, leading to subsets of cells carrying these mutations. Notably, sCNVs are abundant in human neurons (1). The brain develops primarily from the ectoderm, while blood cells originate from the mesoderm. Somatic mutations with high cellular fractions are more likely to occur early in development. If these mutations arise early enough, such as during or before gastrulation, they may be present in both brain and blood cells. As individuals age, clonal hematopoiesis can lead to the accumulation of numerous high cellular fraction somatic mutations in blood cells that may not be present in other tissues. Thus, analyzing genomic data from the blood of younger individuals could identify somatic mutations shared with the brain, providing valuable insights into genetic predisposition to brain diseases (Figure 1).

At least 8 experimental platforms are currently available for detecting sCNVs. Table 1 compares the resolutions, strengths, and weaknesses of these molecular assays. Among them, signal intensity data from SNP genotyping arrays have been used to call germline CNVs based on allele frequencies and signal strengths of predesigned SNP probes (2,3). The same data can also detect mosaicism for CNVs by searching for slight deviations from expected allele frequencies (0, 0.5, and 1). Compared with smaller whole-genome sequencing datasets, larger SNP genotyping array datasets are publicly available for many complex diseases. Since somatic mosaic mutations are rare, larger sample sizes are essential. Thus, larger SNP genotyping array datasets are more desirable than sequencing data. The resolution of SNP arrays typically allows for the detection of somatic mutations with larger cellular fractions. Indeed, higher cellular fraction mutations are more likely to be shared between blood and the brain, making them suitable for brain disease research.

In the current issue of *Biological Psychiatry*, Chang *et al.* (4) investigated the link between sCNVs and schizophrenia using blood-derived SNP genotyping array data from 9715 patients with schizophrenia and 28,822 control individuals of Chinese descent. They reported an association between sCNVs and schizophrenia, identifying 166 autosomal sCNVs ranging from 124 kb to 249 Mb (median = 2 Mb) in 109 participants. The overall sCNV burden was 1.12% in patients versus 0.73% in control individuals, with an odds ratio of 1.54 (95% CI, 1.21–1.95). The maximum number of sCNVs in an individual was 5 (3 individuals had 5 sCNVs, 4 had 4, 9 had 3, 15 had 2, and 78 had 1). Approximately 92% of these post–quality control sCNVs were deletions. The estimated cellular fractions of these sCNVs ranged from 2% to 60% (median = 45%), suggesting that most identified sCNVs may occur during early-stage development. No sCNVs shared identical start and end coordinates; thus, sCNVs within the same cytoband locus were merged as recurrent sCNVs for case-control comparisons. Eight sCNVs were identified as associated with schizophrenia at a threshold of raw $p < .05$. Only up to 6 patients (0.06%) carried an sCNV of the same locus, indicating that sCNVs occur sporadically across the genome. Chang *et al.* also analyzed brain tissue and found that all sCNVs detected in brain cells had significantly low cellular fractions (i.e., 3%–7%). Consequently, only 2 of the 8 sCNV loci identified in blood samples were also observed in brain samples. Technical validation using a targeted assay, such as interphase fluorescent in situ hybridization or digital polymerase chain reaction, is desirable for understanding the roles of sCNVs in schizophrenia (Table 1). Furthermore, it may be worthwhile to explore whether specific blood cell subsets harboring certain sCNVs

SEE CORRESPONDING ARTICLE ON PAGE 198

https://doi.org/10.1016/j.biopsych.2024.10.023
Biological Psychiatry January 15, 2025; 97:104–106 www.sobp.org/journal
ISSN: 0006-3223

Biological Psychiatry

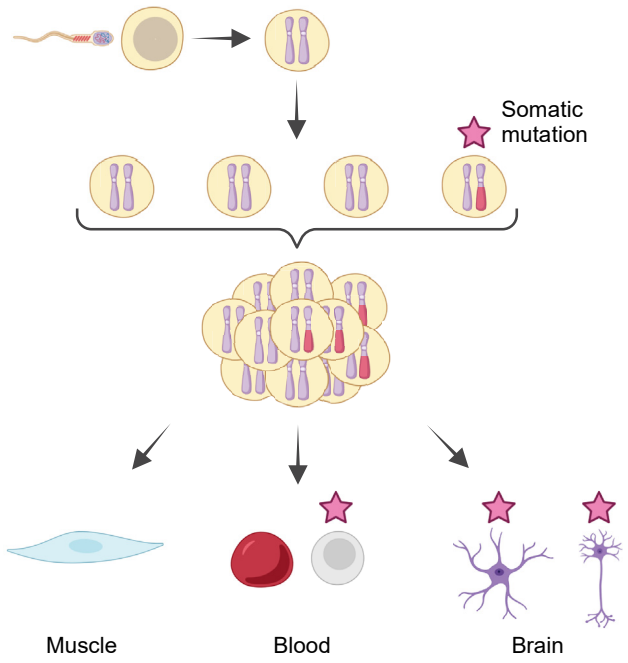

Figure 1. Somatic mosaicism in human tissues. Postfertilization mutations can occur during embryogenesis, postzygotically, or during the early stages of embryonic development. A second mosaic mutation may occur after the initial mutation. These somatic mutations may be shared across various combinations of tissue types. Only 3 tissue types are shown. (Figure created with BioRender.com.)

were favorably selected and underwent clonal amplification, leading to higher cellular fractions in blood compared with brain tissue.

A similar 2023 study by Maury *et al.* (5) analyzed blood-derived SNP genotyping array data of 12,834 patients with schizophrenia and 11,648 control individuals from the Psychiatric Genomics Consortium. They identified at least 2 rare sCNVs more common in schizophrenia, including exonic deletions in *NRXN1*, a known schizophrenia gene. Both studies indicated an increased sCNV burden in schizophrenia and suggested that sCNVs potentially contribute to approximately 1% of schizophrenia cases. However, like many other genetic studies of complex diseases, replicating individual variants across independent studies is challenging, particularly for extremely rare variants like sCNVs. For instance, among the top signals identified in European samples (5), *NRXN1* and *ABCB11*, only one Chinese sample (4) carried the *NRXN1* mutation. Various factors could contribute to discrepancies, such as differences in participant races and ethnicities, disease heterogeneity, signal variations between Illumina and Affymetrix arrays, and differing analytic parameters. Compared with the Maury *et al.* study (5), the Chang *et al.* study (4) identified significantly fewer gain events, and their sCNVs had higher cellular fractions and longer lengths, suggesting ongoing debate regarding sCNV detection and filtering parameters.

Several computational tools are available for identifying sCNVs from array and sequencing data. One widely used tool is the MOsaic CHromosomal Alterations (MoChA) caller (6,7), which detects mosaic chromosomal alterations (sCNVs) based on long-range haplotype phased chromosomal information. It accepts Illumina and Affymetrix genotyping array data as well as whole-genome sequencing data. Users need to generate phased VCF files with B allele frequency and log R ratio for array data or with allelic depth for sequencing data. MoChA provides scripts for preparing such files. Although the software can process whole-exome sequencing data, meaningful results may not be achieved due to insufficient chromosomal coverage. Low-coverage whole-genome sequencing may also be ineffective, particularly for detecting sCNVs with lower cellular fractions. As indicated in the Chang *et al.* study (4), sCNV detection rates correlated with the number of predesigned probes on microarrays, suggesting that analyzing high-

Table 1. Experimental Platforms to Detect Mosaic CNVs

Technology	Cellular Fractions	Resolution	Strengths	Weaknesses
Chromosome Analysis/ Karyotyping	>10%–15% (≥2 of 20 metaphase cells)	>5–10 Mb	Whole genome coverage	Cell culturing required and low resolution
SNP-Based Chromosomal Microarray	>10%–20% (depending on CNV size)	>10–100 kb	Whole genome coverage, high resolution, and SNP information available	–
Array Comparative Genomic Hybridization	>10%–20% (depending on CNV size)	>10–100 kb	Whole genome coverage and high resolution	No SNP information
Optical Genome Mapping	>10%–20% (depending on CNV size and sequencing depth)	>10–100 kb	Whole genome coverage and high resolution	No SNP information
Massively Parallel Sequencing	>10%–20% (depending on CNV size and sequencing depth)	>10–100 kb	Whole genome coverage, high resolution, and SNP information available	Cost and bioinformatics challenges
Interphase FISH	1%–5%	100 kb	High sensitivity for low-level mosaicism	Target assay
Digital PCR	1%–5%	>0.5 kb	High resolution and high sensitivity for low-level mosaicism	Target assay
Multiplex Ligation-Probe Amplification	>30%	>1 kb	High resolution	Target assay

CNV, copy number variation; FISH, fluorescent in situ hybridization; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism.

depth whole-genome sequencing data may improve detection accuracy. Additional caution is warranted depending on the underlying disease. For instance, clonal mosaicism in blood cells becomes more common with age, especially in individuals exposed to certain environmental triggers. Specific sCNVs are known to be associated with leukemia or precancerous conditions. When assessing genetic risk in schizophrenia, it is important to exclude known cancer-related sCNVs (gains and losses of the same sequence should be treated as distinct variants, as they can have different functional consequences). Other age-related sCNVs may also require cautious interpretation, as the typical ages of onset for schizophrenia are late adolescence and the early twenties. However, it may not be necessary to exclude age-related sCNVs when studying other diseases.

Current studies are limited to detecting variants arising during early development that are shared between blood and brain cells. Mutations occurring in later stages of brain development often present lower cellular fractions and may be detectable only in brain tissue. To better assess the impact of sCNVs on brain disease, it is important to analyze brain-derived genomic data to identify later-stage somatic mutations as well. Findings from the 2 schizophrenia studies suggest that sCNVs may contribute to various complex diseases, highlighting the need for further investigation into their roles in the genetic architecture of other diseases. Compared with SNP genotyping arrays, the higher resolution of whole-genome sequencing may allow for detecting somatic fusion genes. Single-cell analysis may reveal cell type–specific somatic mosaicism, presenting another exciting research opportunity.

Recent studies have also shown that transposable elements (TEs) actively retrotranspose during neurogenesis, adding another layer of genomic diversity among neurons and contributing to individual somatic mosaicism. Somatic TE insertions have been identified in normal human brain tissue (8,9). However, TEs may be misregulated in some neurological and psychiatric disorders, and the extent of somatic TEs remains under debate. Accurately detecting somatic TEs is even more challenging than detecting sCNVs, suggesting the need for new bioinformatics pipelines to improve the detection of these understudied sequences. Benchmarking various bioinformatics approaches will be valuable for quantifying false discovery rates and guiding the selection of appropriate pipelines and parameters for analyzing different datasets and diseases. Profiling a broader spectrum of somatic mosaic variations can deepen our understanding of their biological implications and contributions to human diseases.

Acknowledgments and Disclosures

This work was partially supported by the National Institute of Allergy and Infectious Diseases (Grant Nos. AI147084 and AI159710), Solve ME/CFS Initiative Ramsay Research Grant Program, Open Medicine Foundation, Department of Defense Lung Cancer Research Program (Grant No. LC190467), The Cornell ME/CFS Collaborative Research Center (Grant No. U54AI178855), and The Interdisciplinary Canadian Collaborative Myalgic Encephalomyelitis Research Network.

I thank Jian Zhao and Xiaoxu Yang for their helpful suggestions and comments on the manuscript.

The author reports no biomedical financial interests or potential conflicts of interest.

Article Information

From the Department of Immunology and Molecular Microbiology, Texas Tech University Health Sciences Center, Lubbock, Texas.

Address correspondence to Dawei Li, Ph.D., at dawei.li@ttuhsc.edu.

Received and accepted Oct 22, 2024.

References

1. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, *et al.* (2013): Mosaic copy number variation in human neurons. Science 342:632–637.
2. Sulovari A, Liu Z, Zhu Z, Li D (2018): Genome-wide meta-analysis of copy number variations with alcohol dependence. Pharmacogenomics J 18:398–405.
3. Li D, Zhao H, Kranzler HR, Li MD, Jensen KP, Zayats T, *et al.* (2015): Genome-wide association study of copy number variations (CNVs) with opioid dependence. Neuropsychopharmacology 40:1016–1026.
4. Chang K, Jian X, Wu C, Gao C, Li Y, Chen J, *et al.* (2025): The contribution of mosaic chromosomal alterations to schizophrenia. Biol Psychiatry 97:198–207.
5. Maury EA, Sherman MA, Genovese G, Gilgenast TG, Kamath T, Burris SJ, *et al.* (2023): Schizophrenia-associated somatic copy-number variants from 12,834 cases reveal recurrent *NRXN1* and *ABCB11* disruptions. Cell Genom 3:100356.
6. Loh PR, Genovese G, McCarroll SA (2020): Monogenic and polygenic inheritance become instruments for clonal selection. Nature 584:136–141.
7. Loh PR, Genovese G, Handsaker RE, Finucane HK, Reshef YA, Palamara PF, *et al.* (2018): Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. Nature 559:350–355.
8. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, *et al.* (2012): Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell 151:483–496.
9. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, *et al.* (2009): L1 retrotransposition in human neural progenitor cells. Nature 460:1127–1131.