Journal of the American Heart Association

EDITORIAL

Identifying Structural Variants and Their Contribution to Cardiovascular Disease Risk: The Long and the Short of It

Hsiao-Huei Chen, PhD; Alexandre F. R. Stewart, PhD

The notion that genetic variants contribute to common diseases is established beyond a doubt. Rare mutations with large effects cause congenital diseases, such as cystic fibrosis or Duchenne muscular dystrophy. Genome-wide association studies have shown that common single-nucleotide polymorphisms, those that occur at a frequency >1% in a population arising from transitions (substitutions of one purine for another, A→G or G→A) or transversions (substitution of a purine for a pyrimidine, or vice versa) with modest effects can influence the risk of many common diseases, including coronary artery disease.[1] Single-nucleotide polymorphisms occasionally change protein coding sequences, but more often they alter transcription factor–binding sites to affect gene expression[2] or splicing of alternative exons or affect RNA translation or stability. Microarrays are ideally suited to interrogate single-nucleotide polymorphisms but less well suited to interrogate structural variants, including insertions, deletions, duplications, and inversions. Structural variants arise from rearrangements at mobile genetic elements or repetitive DNA sequences, of which there are >1200 different kinds scattered all over the human genome.[3] These sequences often contain binding sites for transcriptional regulators,[4] and their rearrangement can affect local gene regulation.

See article by Iyer et al.

Whole-genome sequencing can identify structural variants and, depending on their frequency, can reveal new associations with disease risk. This approach has identified new relatively common structural variants conferring risk for non-Alzheimer dementias, including a deletion in the *TPCN1* gene associated with Lewy body dementia and replication of known structural variants at the *C9orf72* and *MAPT* loci associated with frontotemporal dementia/amyotrophic lateral sclerosis.[5] In this issue of the *Journal of the American Heart Association* (*JAHA*), Iyer et al used whole-genome sequencing to identify structural variants and to carry out the first large-scale association study of structural variants with coronary artery disease.[6] The TopMed group, the National Heart, Lung, and Blood Initiative's Trans-Omics for Precision Medicine program led by Themistocles Assimes at Stanford University, carried out whole-genome sequencing of various individuals

Key Words: Editorials ■ cardiovascular disease ■ genetic association ■ structural genetic variants

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

Correspondence to: Hsiao-Huei Chen, Ottawa Hospital Research Institute, Ottawa, ON K1H 8M5, Canada. Email: hchen@uottawa.ca and Alexandre F. R. Stewart, University of Ottawa Heart Institute Ottawa, ON K1Y 4W7, Canada. Email: astewart@ottawaheart.ca

This manuscript was sent to Barry London, MD, PhD, Senior Guest Editor, for editorial decision and final disposition.

For Sources of Funding and Disclosures, see page 2.

JAHA is available at: www.ahajournals.org/journal/jaha

to yield single-nucleotide and structural variant information from short (≈150 base pair) end reads.

Sixty-six authors contributed to the study, which included 11 556 coronary artery disease cases and 42 907 controls and yielded a single genome-wide significant locus out of 58 706 structural variants tested for association (P<1.54E-9). This locus contains a 1200-base-pair duplication of an enhancer at 6q21; this 6q21 region contains few genes, a so-called gene desert, so it was not obvious how this structural variant would affect gene expression and disease risk. Using a deep learning algorithm to predict chromatin folding based solely on DNA sequence, they found that the structural variant at 6q21 can modify chromatin folding and bring the 6q21 enhancer to the vicinity of the *LAMA4* gene. *LAMA4* encodes the laminin subunit α4, a gene involved in cell proliferation and tumor metastasis. Intriguingly, reduced *LAMA4* expression was reported in human atherosclerotic plaques.[7]

The structural variants ranged in size from 11 base pairs to ≈18.9 kilobase pairs and most (≈75%) of the structural variants tested were of low frequency (<1%), necessitating a "sliding window" algorithm to test for association. When the data were "tortured" by the sliding window approach, 1 other locus panned out, overlapping the *USP36* gene (P=1.03E-10). The *USP36* locus is interesting, as it not only codes for the USP36 (ubiquitin-specific peptidase 36) protein expressed in the vasculature, it also encodes a circular RNA, circUSP36/hsa_circ_0003204, which also has been implicated in atherosclerosis.[8,9] A further examination yielded an additional 24 structural variants with a nominal association with coronary artery disease risk (P<0.0001), far from the threshold of genome-wide significance (P<8.52E-07).

Was the discovery of 2 structural variants and 24 putative loci worth the effort? At first glance, it seems to have been almost futile. However, it is important to highlight what this exercise has taught us. In other studies for common diseases like non-Alzheimer dementia, investigators also identified few structural variants contributing to disease, so the TOPMed study here is in the same ballpark, in terms of outcomes. One thing to bear in mind, however, is that short-read sequences may underestimate complex structural variants, and future studies with long-read sequences are likely to add to the repertoire of structural variants that associate with disease. Most importantly, replication of the findings from the TOPMed study in an independent cohort is needed to validate these discoveries. It seems likely that this will happen, if prior studies on dementia-related variants are any indication.[5] The TOPMed study has identified a number of putative rare variants and some are likely to replicate in other studies. As the cost of whole-genome long-read sequencing becomes more affordable and the numbers of whole-genome sequencing data being deposited from multiple research institutions grows, additional loci will be discovered, and true loci should be validated by future studies. Determining how these structural variants confer disease risk will be the real challenge.

ARTICLE INFORMATION

Affiliations

Neuroscience, Ottawa Hospital Research Institute, Ottawa, ON, Canada (H-H.C.); Brain and Mind Institute (H-H.C.), Departments of Cellular and Molecular Medicine and Medicine (H-H.C.) and Centre for Infection, Immunity & Inflammation (H-H.C., A.F.S.), University of Ottawa, Ottawa, ON, Canada; Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada (A.F.S.); and University of Ottawa Heart Institute, Ottawa, ON, Canada (A.F.S.).

Sources of Funding

Drs Stewart and Chen are supported by operating grants from the Canadian Institutes of Health Research (Dr Stewart: 497628, 376503; Dr Chen: 376403, 479803), discovery grants from the Natural Sciences and Engineering Research Council of Canada (Dr Chen: RGPIN-2019-03942; Dr Stewart: RGPIN-2016-04985), grants-in-aid by the Heart and Stroke Foundation of Canada (Dr Stewart: G-16-00014085; Dr Chen: G-18-0022157), a grant from the Weston Family Foundation (Dr Chen), and a midcareer salary award by the Heart and Stroke Foundation of Ontario (Dr Chen).

Disclosures

None.

REFERENCES

1. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang M, Hindy G, Zhou W, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54:1803–1815. doi: 10.1038/s41588-022-01233-6
2. Almontashiri NA, Antoine D, Zhou X, Vilmundarson RO, Zhang SX, Hao KN, Chen HH, Stewart AF. 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent transforming growth factor beta regulation of p16 expression in human aortic smooth muscle cells. *Circulation*. 2015;132:1969–1978. doi: 10.1161/CIRCULATIONAHA.114.015023
3. Kojima KK. Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob DNA*. 2018;9:2. doi: 10.1186/s13100-017-0107-y
4. Karttunen K, Patel D, Xia J, Fei L, Palin K, Aaltonen L, Sahu B. Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nat Commun*. 2023;14:5313. doi: 10.1038/s41467-023-41081-4
5. Kaivola K, Chia R, Ding J, Rasheed M, Fujita M, Menon V, Walton RL, Collins RL, Billingsley K, Brand H, et al. Genome-wide structural variant analysis identifies risk loci for non-Alzheimer's dementias. *Cell Genom*. 2023;3:100316. doi: 10.1016/j.xgen.2023.100316
6. Iyer KR, Clarke SL, Guarischi-Sousa R, Gjoni K, Heath AS, Young EP, Stitziel NO, Laurie C, Broome JG, Khan AT, et al. Unveiling the genetic landscape of coronary artery disease through common and rare structural variants. *J Am Heart Assoc*. 2025;14:e036499. doi: 10.1161/JAHA.124.036499
7. Zegeye MM, Matic L, Lengquist M, Hayderi A, Grenegard M, Hedin U, Sirsjo A, Ljungberg LU, Kumawat AK. Interleukin-6 trans-signaling induced laminin switch contributes to reduced trans-endothelial migration of granulocytic cells. *Atherosclerosis*. 2023;371:41–53. doi: 10.1016/j.atherosclerosis.2023.03.010
8. Liu H, Ma X, Mao Z, Shen M, Zhu J, Chen F. Circular RNA has_circ_0003204 inhibits oxLDL-induced vascular endothelial cell proliferation and angiogenesis. *Cell Signal*. 2020;70:109595. doi: 10.1016/j.cellsig.2020.109595
9. Huang JG, Tang X, Wang JJ, Liu J, Chen P, Sun Y. A circular RNA, circUSP36, accelerates endothelial cell dysfunction in atherosclerosis by adsorbing miR-637 to enhance WNT4 expression. *Bioengineered*. 2021;12:6759–6770. doi: 10.1080/21655979.2021.1964891