Circulation Research

EDITORIAL

Understanding Arteriosclerosis 2.0

Making Sense of Genetic Variants with scATAC

Ralf P. Brandes, Ralf Gilsbach

Arteriosclerosis is a chronic, multifactorial disease, with aging being the most important risk factor. Genetic and modifiable risk factors impact on the pace of the accumulation of the arteriosclerotic burden or influence aspects which determine the dignity of the disease like plaque rupture or capacity of outward remodeling.[1] A substantial amount of research has been dedicated to the contribution of inflammation in arteriosclerosis, as plaques are rich in immune cells, as chronic inflammatory diseases are arteriosclerotic risk factors and as anti-inflammatory interventions in mouse models of accelerated arteriosclerosis have proven effective.[2] Interestingly, the clinical benefit from anti-inflammatory therapy in arteriosclerosis are largely restricted to patients with increased systemic markers of inflammation.[3] Similarly, in genome-wide association studies (GWAS) the major arteriosclerotic risk loci are not highly enriched in inflammation-associated genes.[4] The problem of GWAS studies, however, is that most disease-relevant single nucleotide polymorphisms (SNPs) cannot be assigned to a specific gene and that even the demonstration that an individual SNP affects gene expression has been impossible for most SNPs.[4] The first aspect is consequence of the complex architecture of the genome with enhancers been often located far away from their target gene in a 2-dimensional sequence-based projection. The latter aspect might be consequence of the heterocellularity of the arteriosclerotic lesion so that a certain SNP is only relevant in one of the many different cell types expressed in a lesion.

Article, see p 240

In this issue of *Circulation Research*, Örd et al[5] address all these aspects in an elegant study combining nuclear single-cell Assay for Transposase-Accessible Chromatin (ATAC) sequencing (scATAC) of endarterectomy material with publicly available data from the internet. ATAC, using sequencing, is a technique first published in 2013, to detect accessible region of the chromatin.[6] Chromatin consists of DNA tightly wrapped around nucleosomes and associated proteins. Binding of transcription factors and RNA-polymerase II induces depletion of nucleosomes (nucleosome depleted regions) which increases the accessibility of the DNA for additional DNA binding proteins. This mechanism is essential to regulatory elements including enhancers, which are distal regulatory elements promoting target gene expression.[7]

Depletion of nucleosomes at promoters and enhancers increases also the accessibility for the Tn5 transposase, which is the enzyme used to tag DNA in the ATAC method. The transposed DNA fragments can be identified by next-generation sequencing and result in ATAC peaks, which overlap with enhancers and active promoters. The combination of ATAC with single-cell technology, therefore, provide a cell-specific and unbiased annotation of the regulatory landscape of a cell. Since the regulatory landscape is the basis for cellular identity, scATAC peaks allow clustering of cells in a similar way as in scRNA-SEQ.

In the study by Örd et al, ≈40% of the detected ATAC peaks were cell type-specific. Of those, 90% were located at distal enhancers. As gene expression is modulated by transcription factor activity at enhancers, scATAC not only allows to discriminate the different cell populations in a mixed tissue, it also provides cues to the

Key Words: Editorials ■ arteriosclerosis ■ chromatin ■ endarterectomy ■ gene expression ■ inflammation

Correspondence to: Ralf P. Brandes, MD, Institut für Kardiovaskuläre Physiologie, Fachbereich Medizin der Goethe-Universität, Theodor-Stern Kai 7, 60590 Frankfurt am Main. Email brandes@vrc.uni-frankfurt.de

For Sources of Funding and Disclosures, see page 261.

Circulation Research is available at www.ahajournals.org/journal/res

mechanistic basis of gene expression differences. Örd et al impressively document this by linking enhancers and putative target promoters using coaccessibility analysis. The validity of the scATAC approach is documented by the great overlap in cell cluster assignment between scATAC and scRNA-SEQ. More importantly, scATAC explains the scRNA-SEQ data by uncovering the transcription factors and enhancers driving gene expression of the different cell types. It also helps deciphering the disease programs active in arteriosclerosis like transdifferentiation of smooth muscle cells and fibromyocytes. The repertoire of single-cell technology is currently rapidly developing. Most recent methodology, like the combined application of scATAC and scRNA-SEQ, will allow cardiovascular researchers to assess RNA and various epigenetic read-outs simultaneously from the same cell.[8]

Compared with bulk ATAC experiments, the resolution provided by scATAC approaches for the individual cell is low. In the present study, Örd et al studied a total of 7000 cells obtained from 3 patients detecting on average 7129 ATAC fragments per cell, that is, 7129 accessible sites within the genome. Given that most cells of our body have more than hundred thousand accessible sites spanning tens of megabases of our genome, it is remarkable that it is possible to derive meaningful information on the basis of such a shallow signal. The essential trick is to perform pseudo-bulk analysis. First, ATAC information is used to cluster the cells, that is, to generate cell identity information, and subsequently all ATAC peaks from an individual cell cluster are merged to increase the coverage. Compared with bulk tissue analyses, the advantage of the single-cell approach is (among others) that it is unbiased and does not rely on historically defined cell type-specific markers. Furthermore, single-cell technology eliminates cell purification steps, which always suffer from contamination, require large sample numbers and are difficult to control, due to the lack of internal standards. By applying scATAC to endarterectomy samples, a bulk of interesting and clinically important information is obtained by Örd et al, among them the different cell types forming the plaque with the correlation of cluster-specific transcription factors and pseudo time analysis.[5]

A special focus of the article by Örd et al are distal regulatory elements known as enhancers. Unlike gene promoters, which are usually in close proximity to a gene, that is, directly upstream, enhancers mediate their

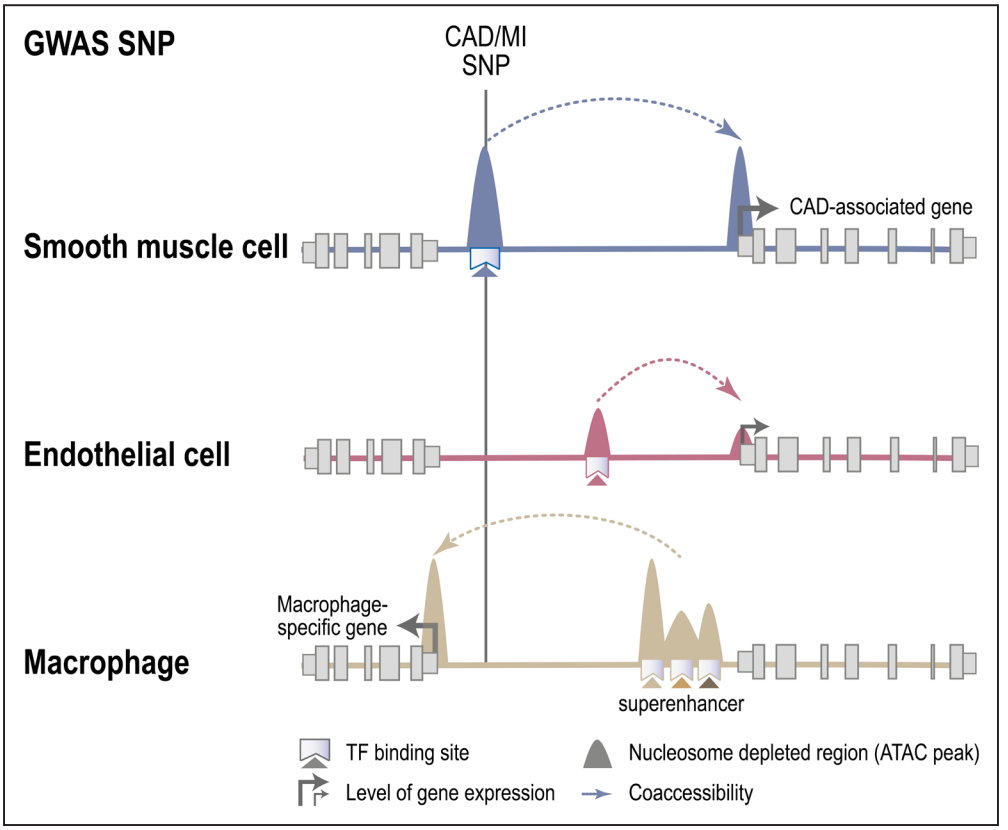

Figure. Schematic representation of the function of cell-type specific and disease-associated regulatory elements. Nucleosome depleted regions detected by pseudobulk analysis of single-cell Assay for Transposase-Accessible Chromatin (ATAC) sequencing data identify distal regulatory elements (enhancers) and active promoters in a cell-type specific manner. Coaccessibility analysis allows to link enhancer regions to their putative target genes. In combination with transcription factor motif analysis, this allows predicting regulatory networks essential for cell identity and phenotype expression in arteriosclerotic lesions. Identification of enhancer-promoter interactions further allows decoding the function of noncoding disease-associate genetic variants and to identify the affected cell types. CAD indicates coronary artery disease; MI, myocardial infarction; and SNP, relevant single nucleotide polymorphism.

function through spatial chromatin interactions. On the 2-dimensional linear DNA molecule, enhancers can be hundreds of kilobases away from their target gene, with other genes interspersed. Örd et al calculated coaccessibility scores to link enhancers with their target gene promoters. However, since enhancers are part of gene regulatory networks, additional experiments are needed to show that enhancers physically interact (Hi-C) and regulate (perturb-seq or enhancer deletion) their target genes. Comparison of genes between different cell clusters by Örd et al confirmed the general concept that cell type-specificity of enhancers is higher compared with promoter accessibility. Integration of information of cultured cells revealed that the in vivo enhancer landscape of a certain cell type is maintained to 70% to 90% in culture. It, however, also showed that the in vivo situation, potentially owing to the greater heterogeneity, contains a broader spectrum of active transcription factor binding sites.

What makes the article of Örd et al particularly interesting is that the chromatin state of the individual cell clusters is used to interpret GWAS data. Integration of both information revealed that >70% of the studied GWAS loci were expressed in a cell-specific manner. This observation highlights the need of cell type-specific epigenetic data to identify the regulatory function of disease-associated elements. In the future, this will allow to annotate the function of lead and proxy SNPs and even sub-threshold SNPs detected by whole-genome GWAS studies.[9]

Gene expression is the consequence of numerous regulatory factors acting in concert and thus, even though a single SNP may change the function of an individual enhancer, the overall consequences for the expression of the gene may be small. Of the 423 coronary artery disease/myocardial infarction GWAS SNPs studied by Örd et al, 259 fell into ATAC peaks. Combining several approaches, a gene could be assigned to 241 of the 259 open chromatin regions harboring coronary artery disease-risk SNPs. Effects on gene expression for 185 of these 259 scATAC-SEQ peaks harboring coronary artery disease SNPs could subsequently be confirmed through the integration of expression quantitative trait loci information. Collectively, the authors succeeded in assigning ≈50% of all causal SNPs to a target gene and very often a risk SNP was located in a single enhancer element. In subsequent experiments, these findings were experimentally validated and refined in endothelial and smooth muscle cells using massive parallel reporter assays (STARR-seq) with haplotype resolution. These experiments documents, that even within a single cell type, SNP-dependent effects can be context-specific with respect to the modulation of the epigenetic signature and usually affect several genes (Figure).

Collectively, the work by Örd et al illustrates how scATAC helps understanding the pathobiology of arteriosclerosis through a cell cluster-specific identification of active enhancers. Integration of this information with expression quantitative trait loci and GWAS data allows connecting disease-associated genes to regulatory elements in the genome. This information can be used for risk stratification of SNPs, development of novel therapeutic strategies and wet lab studies to uncover the gene regulatory mechanisms for the arteriosclerotic process.

ARTICLE INFORMATION

Affiliations

Institute for Cardiovascular Physiology, Faculty of Medicine, Goethe-University Frankfurt, Germany (R.P.B., R.G.). DZHK – German Center for Cardiovascular Research, Partner site Rhine-Main (R.P.B., R.G.).

Sources of Funding

This work was supported by the Goethe University Frankfurt am Main, the DZHK (German Center for Cardiovascular Research), the DFG (German Research Foundation) excellence cluster EXS2026 Cardiopulmonary System Cardio-Pulmonary Institute (CPI) and the DFG SFB834 (TP A2 to R.P. Brandes) and SFB1425 (P02 to R. Gilsbach), DFG 428316638 and 386460455 to R. Gilsbach as well as TRR 267 (TP A04&A06 to R.P. Brandes).

Disclosures

None.

REFERENCES

1. Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, Bittencourt MS, Tokgözoğlu L, Lewis EF. Atherosclerosis. *Nat Rev Dis Primers.* 2019;5:56. doi: 10.1038/s41572-019-0106-z
2. Libby P. The changing landscape of atherosclerosis. *Nature.* 2021;592:524–533. doi: 10.1038/s41586-021-03392-8
3. Lorenzatti AJ. Anti-inflammatory treatment and cardiovascular outcomes: results of clinical trials. *Eur Cardiol.* 2021;16:e15. doi: 10.15420/ecr.2020.51
4. Musunuru K, Kathiresan S. Genetics of common, complex coronary artery disease. *Cell.* 2019;177:132–145. doi: 10.1016/j.cell.2019.02.015
5. Örd T, Õunap K, Stolze L, Aherrahrou R, Nurminen V, Selvarajan I, Toropainen A, Lönnberg T, Aavik E, Yla-Herttuala S, Civelek M, et al. Single-cell epigenomics and functional fine-mapping of atherosclerosis GWAS loci. *Circ Res.* 2021;129:240–258. doi: 10.1161/CIRCRESAHA.121.318971
6. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213–1218. doi: 10.1038/nmeth.2688
7. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet.* 2014;15:272–286. doi: 10.1038/nrg3682
8. Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med.* 2020;52:1428–1442. doi: 10.1038/s12276-020-0420-2
9. Wang X, Tucker NR, Rizki G, Mills R, Krijger PH, de Wit E, Subramanian V, Bartell E, Nguyen XX, Ye J, et al. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife.* 2016;5:e10557. doi: 10.7554/eLife.10557