Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Functional genomics and assays of regulatory activity detect mechanisms at loci for lipid traits and coronary artery disease

Tamara S Roman and Karen L Mohlke

Many genome-wide association studies (GWAS) have identified signals located in non-coding regions, and an increasing number of functional genomics annotations of regulatory elements and assays of regulatory activity have been used to investigate mechanisms. Genome-wide datasets that characterize chromatin structure help detect potential regulatory elements. Assays to experimentally assess candidate variants include transcriptional reporter assays, and recently, massively parallel reporter assays (MPRAs). Additionally, the effect of candidate regulatory elements and variants on gene expression and function can be evaluated using genomic editing with the CRISPR-Cas9 technology. We highlight some recent studies that employed these strategies to identify variant effects and elucidate molecular and/or biological mechanisms at GWAS loci for lipid traits and coronary artery disease.

Address
Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, United States

Corresponding author: Mohlke, Karen L (mohlke@med.unc.edu)

Current Opinion in Genetics & Development 2018, **50**:52–59

This review comes from a themed issue on **Molecular and genetic basis of disease**

Edited by **Ines Barroso** and **Jose Florez**

https://doi.org/10.1016/j.gde.2018.02.004

Introduction

Human genome-wide association studies (GWAS) have identified hundreds of DNA variants associated with blood lipid levels and coronary artery disease. Blood lipid levels are a risk factor for cardiovascular disease, including increased low-density lipoprotein cholesterol (LDL-C), and increased triglycerides [1,2]. GWAS have been very successful at identifying genetic variants associated with these complex metabolic diseases [3,4] however, characterizing the molecular mechanisms responsible for these associations has been challenging.

Most variants identified by GWAS are located within non-coding regions of the genome [5], suggesting that these variants do not alter the structure or function of the encoded proteins. Variants located within regulatory elements, such as enhancer or silencer regions, may act to enhance or reduce gene expression. These regulatory regions may affect multiple genes and may regulate genes located hundreds of kilobases away [6,7]. Current challenges are to identify which GWAS variants have regulatory functions and to characterize the molecular mechanisms by which allelic differences affect gene activity and disease risk. Recently, identification of regulatory elements and variants has been facilitated by technological development of genome-wide functional assays.

Experimental assays are necessary to determine which of the variants located in regulatory regions have allelic effects on regulatory activity. For example, transcriptional reporter assays are used to identify variants that alter promoter or enhancer/silencer activity, and recently, high-throughput approaches have been used to test dozens to thousands of candidate variants in a massively parallel fashion. Notably, GWAS studies identify lead variants that are most strongly associated with a trait or disease; however, these lead variants are not necessarily the functional regulatory variants due to sampling variation, technical, or stochastic reasons. Variants that are strongly linked, or inherited together (in strong linkage disequilibrium) with the lead variants may have regulatory effects, which highlights the need to test a number of variants in experimental assays. Reporter assays are often performed by cloning variant-containing regions into vectors containing a reporter gene and transfecting biologically relevant cell types including induced pluripotent stem cells (iPSCs) that can be differentiated into several target cell types [8]. Another recent strategy for experimental evaluation of regulatory elements and variants uses CRISPR-Cas9-mediated genome editing. Cells or organisms can be edited to delete or alter the effect of a regulatory element or to create specific allelic substitutions.

Functional genomics regulatory annotation data can be used to help guide the selection of candidate regulatory elements and variants for experimental testing. These data are generated through the use of high-throughput methods, including open chromatin, chromatin conformation, and chromatin immunoprecipitation assays followed by high-throughput sequencing. These assays can

be helpful at detecting candidate regulatory elements, including enhancer or silencer regions, which may ultimately aid in elucidating the mechanisms underlying the relationships between the GWAS variants and trait [9–11]. In this review, we discuss some recent approaches used to identify functional variants and mechanisms at lipid or coronary artery disease GWAS loci.

High-throughput functional genomics assays

High-throughput functional genomics assays of chromatin structure identify genomic regions characteristic of regulatory elements. Several consortia, including the Encyclopedia of DNA elements (ENCODE) consortium, the National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium, and others in the International Human Epigenome Consortium [9,10,12] have generated genome-wide maps in hundreds of cell types and tissues. Maps of open, or accessible, chromatin denote DNA regions devoid of histones and more accessible to transcription factors, as detected by DNase hypersensitivity (DNase HS) [13] formaldehyde-assisted isolation of regulatory elements (FAIRE) [14] or assays for transposase-accessible chromatin (ATAC)-sequencing [15,16] (Table 1). Maps of histone modifications, detected through chromatin immunoprecipitation (ChIP)-sequencing [17,18]), can be integrated to predict chromatin state, including promoter, enhancer, and silencer regions [19], and regions bound by transcription factors can be further annotated with sequence binding motifs and to detect transcription factor footprints [17,20]. Maps of chromatin interactions, higher order chromatin structure, topologically associated domains (TADs), and frequently interacting regions (FIREs) are based on chromosome conformation capture methods, including Hi-C and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [21–23].

Integrating these data and maps from trait-relevant cell types can be used to guide the identification of trait-associated variants located in regulatory regions. Data sets of chromatin accessibility and histone marks have differing signal strength compared to background and may be influenced by cellular environment, so multiple lines of evidence (*i.e.* peaks) can provide greater evidence of a regulatory element. Specific histone marks tend to be observed at different types of elements, such as H3K27ac marks at active enhancers. However, neither individual marks nor the absence of regulatory evidence from an assay is definitive, and mere presence of a variant in a region of histone marks does not indicate that the variant alleles alter regulatory activity [24•]. While evidence of chromatin interactions between variant positions and a transcription start site supports the potential for a regulatory effect of variants on a gene, the interaction alone is not definitive, and absence of interactions may be due to assay resolution or cellular environment. Therefore, combining one or more pieces of evidence from high-throughput regulatory assay data with other functional genomics experiments, such as reporter assays and genomic editing, enhances the likelihood of identifying candidate regulatory variants.

Allelic differences in transcriptional activity

A common approach to examine variants for effects on transcriptional activity is a reporter assay. DNA segments of tens to thousands of base pairs containing individual

Table 1

Acronyms for terms and assays used to study noncoding, regulatory regions

Acronym	Full name	Description
DNase	DNase I hypersensitivity	Assay used to detect open, accessible chromatin regions
FAIRE	Formaldehyde-assisted isolation of regulatory elements	Assay used to detect open, accessible chromatin regions
ATAC-seq	Assay for transposase-accessible chromatin with high-throughput sequencing	Assay used to detect open, accessible chromatin regions
ChIP	Chromatin immunoprecipitation	Assay used to detect protein binding or histone modifications
TAD	Topologically associated domain	Term used to describe three-dimensional chromatin organization and interacting regions
FIRE	Frequently interacting regions	Term used to describe three-dimensional chromatin organization
3C, 4C, 5C, HiC, ChIA-PET	Chromosome conformation capture, '-on-chip,' -carbon copy, chromatin interaction analysis by paired-end tag sequencing	Assays used to detect chromatin contacts, chromatin organization
EMSA	Electrophoretic mobility shift assay	Assay used to detect proteins bound to a nucleotide sequence
MPRA	Massively parallel reporter assay	Assay used to test candidates for effects on transcriptional activity
CRISPR	Clustered regularly interspaced short palindromic repeats	Used typically with Cas9 protein and guide RNAs for genomic editing