



Editorial: Deciphering Non-Coding Regulatory Variants: Computational and Functional Validation

Li Chen^{1,2*} and Mulin Jun Li^{3*}

¹Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, United States,

²Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, United States,

³Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

Keywords: non-coding variants, regulatory variants, variant annotation, functional prediction, functional validation, non-coding variants, functional validation

Editorial on the Research Topic

Deciphering Non-Coding Regulatory Variants: Computational and Functional Validation

OPEN ACCESS

Edited and reviewed by:

Richard D. Emes,
University of Nottingham,
United Kingdom

*Correspondence:

Li Chen
chen61@iu.edu
Mulin Jun Li
mulinli@connect.hku.hk

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 02 September 2021

Accepted: 11 October 2021

Published: 03 November 2021

Citation:

Chen L and Li MJ (2021) Editorial:
Deciphering Non-Coding Regulatory
Variants: Computational and
Functional Validation.
Front. Bioeng. Biotechnol. 9:769614.
doi: 10.3389/fbioe.2021.769614

1 INTRODUCTION

Non-coding DNA sequences play an important role in an organism's genome. Genome-wide association studies (GWASs) have produced tens of thousands of genetic variants that are associated with complex phenotypes from different species, and the majority of these associations fall into the non-coding regions and are suggested to be mediated by context-specific regulatory codes. As large-scale genomic sequencing together with constantly evolving biotechnology revolutionize the functional genomics field, we still face great challenges to accurately predict, interpret, and evaluate the biological functions of non-coding regulatory variants in gene regulation. This research topic aims to discuss recent advances made in understanding the regulatory potentials of non-coding genetic variants/somatic mutations (e.g., affecting transcription, splicing, or epigenetic activities) in various organisms using either computational or experimental methods. Novel approaches, tools, and results for accurate regulatory variant prediction and classification, non-coding variant annotation, deleterious prioritization, and functional assays for regulatory mechanism are recommended. Moreover, the research topic is also expanded to seek recent advances in understanding the regulatory potentials of non-coding RNAs.

2 BIOINFORMATICS METHODS AND ANALYSES FOR IDENTIFYING FUNCTIONAL EFFECTS OF NON-CODING REGULATORY VARIANTS

To pinpoint the non-coding causal variants, Xu et al. developed a computational framework called "regSNPs-ASB" for identifying regulatory single nucleotide polymorphisms (SNPs) in allele-specific transcription factor binding sites from ATAC-seq data. Specifically, regSNPs-ASB simultaneously infers both the genotype information and open chromatin regions from ATAC-seq mapped reads. Then, regSNPs-ASB defines candidate allele-specific transcription factor binding sites by overlapping TF motif binding sites with heterozygous SNPs and open chromatin region. Next, for each candidate allele-specific transcription factor binding site, regSNPs-ASB adopts a generalized linear model (GLM) to model the relationship between ATAC-seq mapped reads as response variable and covariates, which include SNP genotype,

the transposase-cleavage event, and the interaction of SNP genotype and transposase-cleavage event. The GLM model aims to test the interaction term to identify heterozygous SNPs with significant allelic imbalance between reference and alternative alleles in open chromatin region, and the heterozygous SNPs with the significant allelic imbalance will be considered as regulatory SNPs. The authors further applied regSNPs-ASB to human MCF-7 breast cancer cells and human mesenchymal stem cells (MSC) and identified 53 and 125 regulatory SNPs respectively. Validated identified regulatory SNPs can affect the expression of their target genes, and the authors found that most of the target genes of regulatory SNPs in promoter regions and target genes of regulatory SNPs in putative enhancer regions in MCF-7 showed significant allele-specific expression. They also found the identified regulatory SNPs in MCF-7 and MSC are enriched in GTEx eQTLs. Based on these discoveries, the author claims that the regSNPs-ASB can be a useful bioinformatics tool to identify candidate causal SNPs from ATAC-seq data only. To study one of the possible mechanisms for non-coding variants affecting gene regulation, Jin et al. leveraged position weight matrixes of transcription factor (TF) and gkm-SVM algorithm to perform a comprehensive evaluation for the functional effects of SNPs on TF binding affinity because the functional consequence of non-coding variants is believed to disrupt the transcription factor binding sites. By exemplifying the predicted binding affinity changes for 18 selected TFs, the authors revealed that the impact of SNPs on TF binding affinity varies substantially and only around 20% of SNPs within putative TF binding sites could have a significant effect on TF binding.

3 BIOINFORMATICS ANALYSES FOR STUDYING THE FUNCTIONAL ROLES OF NON-CODING RNAs

The research topic also expands its collection for research works in the non-coding era, which focus on studying the functions of non-coding RNAs. Morenikeji et al. identified 22 conserved long non-coding RNAs (lncRNAs) potentially regulating gene expression in cytokine storm during COVID-19 and found the lncRNA activated by DNA damage is evolutionarily conserved across multiple species, which may serve as

potential as targets for intervening in SARS-CoV-2 pathogenesis. Sun et al. identified differential expressed (DE) lncRNAs in salt-tolerant sweet sorghum line (M-81E) and the salt-sensitive line (Roma) and found DE lncRNAs can potentially function as competitive endogenous RNAs to influence plant responses to salt stress. You et al. applied qualitative transcriptional signature to predict CpG island methylator phenotype of right-sided colon cancer. Wang et al. constructed a lncRNA-miRNA-mRNA network and found lncRNA EPB41L4A-AS1 could function as a regulator in the pathogenesis of Non-small Cell Lung Cancer. Chen et al. deciphered the cooperative interactions among these regulatory factors, which include transcription factors, lncRNAs, and microRNAs in Diabetic Nephropathy Progression. All these discoveries demonstrate the importance of non-coding RNAs in gene regulation and biological processes.

AUTHOR CONTRIBUTIONS

LC and ML were guest associate editors of the research topic and wrote the paper text.

FUNDING

This work was supported by the Indiana University Precision Health Initiative to LC and Natural Science Foundation of Tianjin under Award Number 19JCJQC63600 to ML.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.