BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Multi-omics data analysis and integration
Title	ASTRA enables allele specific expression and/or chromatin accessibility analysis from sequencing data.
All Authors	Mattevi S(1), Mazzarotto F(1), Martini P(1).
Affiliation	
(1) Department of Molecular and Translational Medicine, University of Brescia, Brescia	

Motivation

Diploid organisms have two almost identical copies of each gene, with both alleles being transcribed. Some exceptions are present, with two of them being genes on the X chromosome, subject to inactivation mechanism in females, and imprinted genes, expressed from a specific parental allele. Recently, it has been shown that also many autosomal genes are monoallelically expressed and that this behavior is transient during developmental stages (Naik et al, iScience, 2021).
This evidence suggests the importance of re-thinking expression studies in terms of allele specific expression. In fact, many genetic variants affecting human phenotypes and disease are heterozygous and the study of allele specific expression could untangle the functional significance of these sites as well as provide insights in terms of genotype-phenotype correlations.
Epigenetic factors are increasingly believed to play a role also on the differential allelic expression. In particular, it has been demonstrated that allele specificity of both expression (ASE) and chromatin accessibility (ASA) can have an important joint influence on several disorders (Zhang et al, Science, 2020).
However, given the high heterogeneity of available analytical approaches, the scientific community would benefit from an integrated bioinformatic pipeline enabling joint ASA and ASE analysis.

Methods

We developed a computational approach called ASTRA (Allelic Specific Transcriptional Regulation Analysis), organized as a Snakemake workflow, enabling integrated analysis of Allele Specific Expression and/or Allele Specific Accessibility from bulk or single-cell RNA-seq data. More specifically, the pipeline adheres to the current best practices for different NGS technologies to process raw sequencing data and obtain variant calling and allele specific read count results. Moreover, we implemented a single-sample phasing step, allowing derivation of a consensus haplotype across two different phasing tools and a refinement based on the minimization of the intra-genic allelic expression variation. The final data analysis is implemented as an R package providing the chance to integrate results and to filter and annotate genes by their allelic specific layout. In particular, each gene’s status is inferred on the basis of the phased allelic counts of variants within its sequence and is categorized as expressed/accessible in a biallelic/monoallelic way.

Results

In this report, we analyzed with ASTRA two isogenic bulk RNA-seq samples derived from a BJ fibroblast and its naive iPSC line for which RNA-seq, ATAC-seq and methylation data are available. Focusing on the expression of imprinted genes, our pipeline confirmed the Loss Of Imprinting (LOI) of key genes like MEG3 and H19 as published in Martini et al, Comm. Biol, 2022.
Besides the imprinted genes, our pipeline enabled identification of additional features acquirable from this type of data. We detected 66% (5545 genes) of the analyzed autosomal genes to be biallelically expressed in the naive iPSC, with 4% (420 genes) of these having a skewed allelic expression in BJ fibroblasts toward one of the parental alleles, including specific imprinted genes that had lost their imprinting status in the iPSC stage (LOI).
Moreover, we found 2149 genes with monoallelic expression in naive iPSCs and 1484 with monoallelic expression in BJ fibroblasts. Notably, the overlap was limited to 512 genes (16%).
To validate the results of our pipeline, we performed a specific analysis of naive pluripotency markers. Doing so, we identified the DPPA5 gene to be characterized by strong monoallelic expression (2 informative heterozygous SNPs with 1780 RNA-seq reads supporting the expression of a single allele) in naive iPSC, but not expressed in BJ fibroblasts. This result is in line with recent findings concerning transcriptional burst and random monoallelic expression of lineage gene markers (Naik et al, iScience, 2021).
To investigate possible causes of monoallelic expression, we attempted ASA analysis but depth unfortunately resulted suboptimal and did not enable confident ASA calling at the DPPA5 locus. As far as the methylation status is concerned, we observed high methylation rates in BJ and markedly lower values in naive iPCs.
Overall, ASTRA can serve as a useful tool to provide a genome-wide snapshot of the transcriptional landscape and to enhance our understanding of determinants of allelic expression and chromatin accessibility.

Info

-

filename	-

Figure	
-	
Availability	-

Corresponding Author

Name, Surname	Stefania, Mattevi
Email	s.mattevi@unibs.it
Submitted on	21.04.2023