

BITS :: Call for Abstracts 2019 - Oral communication

<i>Type</i>	Oral communication
<i>Session</i>	Genome Organization and Gene Regulation
<i>Title</i>	The length of the expressed 3' UTR is an intermediate molecular phenotype linking genetic variants to complex diseases
<i>All Authors</i>	Mariella E (1), Marotta F (1), Grassi E (1), Gilotto S (1), Provero P (1,2).
<i>Affiliation</i>	(1) Department of Molecular Biotechnolgy and Health Sciences, University of Turin, Torino (2) Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Milano

Motivation

In the last decades, genome wide association studies (GWAS) have uncovered thousands of common genetic variants that are associated with complex human traits, including several pathological conditions. However, GWAS results usually provide little information from a functional point of view, especially because the majority of the identified genetic variants are located within non-coding regions. The investigation of the genetic determinants of intermediate molecular phenotypes, such as gene expression and transcript structure, can be helpful to illuminate the road that leads from genetic variants to diseases. In particular, past studies have shown that genetic variants frequently cause changes in the 3' regions of transcripts, suggesting us to devise a computational strategy able to discover genetic variants that specifically affect the relative expression of alternative 3' untranslated region (UTR) isoforms. These isoforms are generated through alternative polyadenylation (APA), a post-transcriptional regulatory mechanism that affect many human genes and can have relevant functional consequences. In the present study, we decided to focus on the most common and simplest situation in which the alternative isoforms have the same coding sequence and a 3'UTR of different length.

Methods

The quantification of alternative 3'UTR isoforms is a common preliminary task for any investigation of alternative polyadenylation and several tools are now available. In particular, some years ago we implemented an algorithm that is able to catch APA events from standard RNA sequencing (RNA-Seq) data. Basically, for each gene in a certain condition we can compute a quantity (the m/M value) that is proportional to the ratio between the expression of the short and the long isoform. This algorithm was originally developed to compare different cell types, but we reasoned that it can also be exploited to explore the variation of the APA landscape across individuals in the same cell type. In this project we took advantage of the GEUVADIS dataset that includes whole genome sequencing and RNA-Seq data for 373 European individuals. RNA-Seq data were obtained in a lymphoblastoid cell line and they were used, together with a custom annotation of alternative 3'UTR isoforms, to compute an m/M value for 6,256 genes in each individual. Then, the association between the m/M values of a gene and the genotype of the individuals was evaluated for cis-acting genetic variants by linear regression. In addition, several enrichment analyses were done to characterize the pinpointed genetic variants and investigate their possible mechanisms of action.

Results

The proposed strategy allowed to identify 2,530 genes with alternative polyadenylation quantitative trait loci (apaQTLs) and we observed that apaQTLs are significantly enriched within genomic regions that are active in lymphoblastoid cells (i.e. promoters, enhancers and transcribed regions), consistently with the fact that otherwise they could not have any regulatory effect.

From a mechanistic point of view, we propose a classification of the significant genetic variants in two main classes. Intragenic apaQTLs are located within the genes for which the models were done and probably exert their effects at the RNA level. On the contrary, extragenic apaQTLs fall inside other genes or within intergenic regions and they may influence the APA regulation through the perturbation of DNA regulatory elements.

The APA regulation is primarily determined by a short RNA motif called polyadenylation signal (PAS) and, as expected, we observed that the perturbation of putative functional PAS motifs is a common mechanism of action for intragenic apaQTLs. For example, we found that the variant rs2842980 is significantly associated with the lengthening of the 3'UTR of SOD2, as a consequence of the creation of a new canonical PAS near the end of the transcript. This variant is in strong linkage disequilibrium with the intergenic variant rs2842992 that has been associated with atrophic macular degeneration, suggesting that

a higher risk for the disease can result from the alteration of the APA regulation.

More in general, we found an overall striking enrichment among apaQTLs of genetic variants reported in the NHGRI-EBI GWAS Catalog. When the same evaluation was performed independently for each trait category, the strongest enrichment was observed for immune system disorders, in line with the fact that the apaQTL mapping was performed in lymphoblastoid cells. However, a strong enrichment was also detected for almost all the other tested categories, including neurological disorders and cancer.

In conclusion, our results point to a widespread effect of genetic variants on the relative expression of alternative 3'UTR isoforms in human and they suggest new ways to extract functional information from GWAS data.

Info

-

Figure

-

Availability <https://doi.org/10.1101/540088>

Corresponding Author

Name, Surname Elisa, Mariella

Email elisa.mariella@unito.it

Submitted on 18.04.2019

Società Italiana di Bioinformatica

C.F. / P.IVA 97319460586

E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma

Website bioinformatics.it

message generated by sciencedev.com for bioinformatics.it 12:55:32 18.04.2019
