doi 10.18699/bgrs2024-1.2-10

Tissue-specific and allele-specific activity of transcribed gene regulatory elements in human skeletal muscle

Buyan A.I.[1, 3]*, Nozdrin V.A.[2], Gazizova G.R.[3], *Meshcheryakov* G.A.[1], Hayashizaki Y.[3], Shagimardanova E.I.[3, 4], Gusev O.A.[3, 4], Kulakovskiy I.V.[1, 3]

[1] *Institute of Protein Research, RAS, Pushchino, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*

[3] *Research Laboratory "Regulatory Genomics", Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

* *andreybuyanchik123@yandex.ru*

Key words: cap analysis of gene expression; transcribed regulatory elements; single-nucleotide variants; allele-specific activity; skeletal muscles

Motivation and Aim: The rapid development of sequencing technologies over the last decades brought molecular genetics close to its medical applications. However, genome-based personalized medicine is still in its early age as its implementation requires reliable tools for annotating and interpreting individual genome variants, including widespread single-nucleotide variants (SNVs). Hundreds of thousands of genome-wide association studies (GWAS) have been conducted to trace the links between genetic variants present in the human population and various phenotypes. Unfortunately, GWAS have limited resolution, strongly depend on the sample size to detect the associated variants, and do not directly differentiate between causal and linked variants. Post-GWAS functional annotation and identification of the causal variants remains challenging and critical for non-coding variants that usually comprise about 90 % of the GWAS credible sets [1]. Considering the regulatory non-coding regions, there are reliable maps of promoters and enhancers in immortalized cell lines and human tissues [2], but they do not describe the actual diversity of particular tissues, such as skeletal muscle. Among the current efforts in validation of regulatory variants detected in GWAS, skeletal muscle-associated phenotypes also remain among the least studied [1].

In this study, we utilized the cap analysis of gene expression (CAGE) data to construct an atlas of transcribed regulatory elements of diverse human skeletal muscles. Next, we assessed the functional effect of SNPs by estimating the allelic imbalance in the transcriptional activity of promoters and enhancers.

Methods and Algorithms: Here we used the results of 228 CAGE experiments from 76 muscles of 4 individuals. The basic bioinformatics pipeline included read mapping with HISAT2 [3], peak calling with CAGEfightR [4], SNV calling using bcftools and PySam [5], and estimating the allelic imbalance with MIXALIME [6].

Results: We identified more than 20 thousand CAGE clusters, including more than 1.5 thousand bidirectional enhancers. More than 3/4 of promoters and enhancers exhibited tissue-specific activity across different skeletal muscles, highlighting very specific gene expression patterns in eye, tongue, and diaphragm. The atlas yielded multiple interesting individual cases, such as tongue-specific keratin and bactericidal SPRR expression, or *CFAP61*, whose expression is positively correlated with aging [7] and was consistently low in eye samples. Furthermore, we compared the promoter

activity of skeletal muscles in different human body parts, considering Hox genes and myosin light and heavy chains, whose expression was in agreement with the muscle origin and functioning. Finally, we performed the differential promoter usage analysis to identify promoters driving different transcript isoforms depending on the muscle type. Among these genes, we found *SEPTIN9* that encodes a cytoskeleton protein containing microtubule-associated domain in most muscles, except for the eyes which use its shorter isoform [8]. Next, we performed the SNV calling which yielded more than 15 thousand reliable variants suitable for the allele-specific analysis. As a result, we detected more than 6 thousand allele-specific events (ASEs) with a significant allelic imbalance in at least one skeletal muscle type. More than half of these events were annotated as expression quantitative trait loci from GTEx [9] or allele-specific transcription factor binding sites from ADASTRA [10]. Besides the events that exhibited allele-specificity in some particular muscles, we also identified nearly thousand of SNVs demonstrating a significant difference in their allelic imbalance across the samples. Annotating these variants by phenotypes using stratified linkage disequilibrium score regression [11] revealed significant enrichment for the normalized muscle volume, suggesting the regulatory nature of the genetic basis of this trait [12].

Conclusion: Taken together, we constructed an atlas of transcribed regulatory elements in diverse human muscles (FANTOMUS: functional annotation of muscle promoters and enhancers), and performed the allele-specific analysis of promoters and enhancers. We believe that FANTOMUS will be useful for studying gene expression patterns and variant effects in healthy muscles as well as myopathies.

Funding: The study is supported by the Ministry of Science and Higher Education of the Russian Federation (No. 075-15-2021-601) and Russian Science Foundation (No. 20-74-10075).

References

1. Alsheikh A.J., Wollenhaupt S., King E.A. et al. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genomics*. 2022;15(1):74. doi 10.1186/s12920-022-01216-w
2. Andersson R., Gebhard C., Miguel-Escalada I. et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455-461. doi 10.1038/nature12787
3. Kim D., Paggi J.M., Park C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4
4. Thodberg M., Thieffry A., Vitting-Seerup K. et al. CAGEfightR: analysis of 5′-end data using R/Bioconductor. *BMC Bioinformatics.* 2019;20(1):487. doi 10.1186/s12859-019-3029-5
5. Danecek P., Bonfield J.K., Liddle J. et al. Twelve years of SAMtools and BCFtools. *GigaScience. 2021;10(2):giab008. doi 10.1093/gigascience/giab008*
6. *Meshcheryakov G., Abramov S., Boytsov* A. et al. MIXALIME: MIXture models for allelic imbalance estimation in high-throughput sequencing data. *arXiv.* 2024. doi 10.48550/arXiv.2306.08287
7. Tumasian R.A., Harish A., Kundu G. et al. Skeletal muscle transcriptome in healthy aging. *Nat Commun.* 2021;12(1):2014. doi 10.1038/s41467-021-22168-2
8. Kuzmić M., Linares G.C., Fialová J.L. et al. Septin-microtubule association via a motif unique to isoform 1 of septin 9 tunes stress fibers. *J Cell Sci.* 2022;135(1):jcs258850. doi 10.1242/jcs.258850
9. Lonsdale J., Thomas J., Salvatore M. et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585. doi 10.1038/ng.2653
10. Abramov S., Boytsov A., Bykova D. et al. Landscape of allele-specific transcription factor binding in the human genome. *Nat Commun.* 2021;12(1):2751. doi 10.1038/ng.2653
11. Finucane H., Bulik-Sullivan B., Gusev A. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228-1235. doi 10.1038/ng.3404
12. van der Meer D., Gurholt T.P., Sønderby I.E. et al. The link between liver fat and cardiometabolic diseases is highlighted by genome-wide association study of MRI-derived measures of body composition. *Commun Biol.* 2022;5(1):1271. doi 10.1038/s42003-022-04237-4