doi 10.18699/bgrs2024-1.3-15

The tissue-specific outcomes of TAD boundary disruption in Unc5b-Slc29a3 locus

Salnikov P.[1, 2]*, Stepanchuk Y.[1, 2], Yan A.[1, 2], Torgunakov N.[2], Belokopytova P.[1, 2], Fishman V.[1, 2, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Artificial Intelligence Research Institute (AIRI), Moscow, Russia*

* *paul.salnikov@gmail.com*

Key words: CTCF binding sites; cis-regulating elements; TAD boundary; *Slc29a3*; *Unc5b*

Motivation and Aim: In vertebrates, interphase chromatin is organized into structures known as Topologically Associated Domains (TADs) and smaller loop structures. These structures are highly conserved across species, and disturbances in TAD structures can drastically affect gene expression, potentially leading to disease. Researchers often focus on developmental genes that exhibit binary types of expression regulation, which may render them relatively insensitive to most TAD perturbation effects. Conversely, the impact of TAD disturbances on housekeeping genes, which are active in all cell types, is often overlooked. We propose that merging regulatory regions containing these ubiquitously expressed genes could lead to competition among promoters for enhancer activity, potentially altering gene expression levels.

In this study, we examine genes with broad activity profiles that are separated into two distinct TADs by a robust insulatory boundary. Specifically, we explore the consequences of CTCF binding sites deletion at *Slc29a3*/*Unc5b* locus, which includes genes crucial for development and viability, organized into two separate TADs by a conserved insulatory boundary. We hypothesize that disruptions at the TAD boundary could lead to misregulation of these genes, providing insight into how genome organization affects transcriptional activity across different tissues.

Methods and Algorithms: To assess gene expression alterations resulting from CTCF binding site deletions, we employed a robust experimental design involving F1 of mice with specific CTCF-bs deletions with wild-type CAST mouse strains. This allowed us to differentiate between allelic transcripts due to distinct exonic SNPs. Thus, the effects of the CTCF binding site deletions could be specifically analyzed within the same tissue samples by comparing the expression of CAST and C57BL/6 alleles. This design significantly reduces experimental biases as both alleles are subjected to the regulation except spatial cis-environment. Then we generated cDNA using UMI-barcoded, transcript-specific primers. Subsequent next-generation sequencing (NGS) analysis of these products enabled the discrimination and quantification of each allele. Using this strategy, we profiled allelic expression levels of hybrid groups carrying wild-type or deletion allele along with CAST allele, assessing differences between their levels of expression.

Results: For our experiments, we used previously obtained mouse line with deletions of CTCF binding sites (CTCF-bs) at TAD boundary region in *Slc29a3*/*Unc5b* locus. Hi-C maps derived from animal samples revealed noticeable changes at the subTAD level. We confirmed that loops within the *Slc29a3* TAD, originally anchored by the now-

deleted CTCF-bs, were disrupted in analyzed tissues (liver, kidney, and cerebellum). Typically, in wild-type animals, these loops facilitate connections from the TAD boundary to the CTCF-bs within the Cdh23 gene. Despite these disruptions, the rest of the inner loop structure orchestrating this TAD remained intact.

Significantly, the chromatin region extending from the TAD boundary to the Cdh23 promoter, which encompasses the *Slc29a3* gene, developed increased spatial interactions with the entire Unc5b TAD. This indicates that this DNA region changed its association to an adjacent TAD, with the TAD boundary shifting towards the Cdh23 promoter.

Furthermore, we observed the emergence of new, aberrant long-range interactions. These newly formed loops extend from the inner CTCF-binding sites at the Cdh23 gene to the outer borders of the *Unc5b* TAD, an area harboring silent non-coding RNA genes. Remarkably, these interactions cross the insulatory TAD boundary, suggesting that their stability might depend on mechanisms other than the standard loop extrusion, possibly involving the chromatin segregation. The pattern of these interactions also varied significantly across different tissues, likely depending on the overall epigenetic activity status of the locus.

Next, we analyzed the expression of six genes within the locus of interest (*Unc5b*, *Slc29a3*, *Psap*, *Vsir*, *Cdh23*, and *Sgpl1*) across five different organs. Out of 60 evaluations, we found 20 instances of significant gene expression alterations linked to deletions at CTCF binding sites. These changes were generally modest, with none exceeding a 50 % variation in expression levels. Notably, the alterations exhibited a tissue-specific pattern, varying in both magnitude and direction. For instance, expression of the *Slc29a3* gene decreased in the kidney but increased in the cerebellum. We confirmed these alterations using digital PCR.

We attempted to fit the observed expression changes in the existing models of interactions between promoters and regulatory elements. Given the tissue-specific nature of these expression changes, we hypothesized that these differences are likely explained by the local epigenetic environment and normal transcriptional activity of the genes in a given tissue.

Following the deletion, the expression profiles of genes divided by the TAD boundary, specifically *Slc29a3* and *Unc5b*, became more similar. In the kidney, *Slc29a3* showed a nearly 40 % decrease in expression, while *Unc5b* was upregulated by 40 %. In wild-type environment, *Slc29a3* is active while *Unc5b* is repressed in this tissue. The disruption of the insulating boundary between these genes appears to cause them to adopt each other's regulatory landscapes, leading to an "averaging" of their expressions.

In the cerebellum, the *Slc29a3* gene exhibited a 15 % increase in expression in a context where *Unc5b* is highly expressed relative to its levels in other tissues. This rise in *Slc29a3* expression might be attributed to its interaction with active cis-regulatory elements from *Unc5b*, such as the super enhancer signatures within Unc5b's first intron. This suggests that the spatial interaction with the *Unc5b* gene allows *Slc29a3* to tap into *Unc5b* active chromatin environment, enhancing its own expression. Unlike in the kidney, there was no corresponding decrease in *Unc5b* expression, likely due to the specific mechanisms of enhancer action or the underlying scale of these changes.

In the cerebellum, we observed a substantial decrease in *Cdh23* expression, which could typically be attributed to either the disconnection of the promoter from essential regulatory elements or the spreading of heterochromatin. Yet, no obvious enhancer elements altering their contact frequency with the *Cdh23* promoter upon TAD disruption were identified. Since Hi-C experiments showed the emergence of ectopic long-range

interactions between the *Cdh23* gene body and the distal regions of the *Unc5b* TAD, we suggest that these new interactions can be associated with the formation of repressive chromatin domains, leading to *Cdh23* downregulation in the cerebellum.

Conclusion: We have shown that the specific deletions of CTCF binding sites in the *Slc29a3*/*Unc5b* locus disrupted the TAD boundary and altered the internal loop architecture within the *Slc29a3* TAD. Although there was no full merging of domains across the organs we studied, the boundary itself was relocated to an inner CTCF-binding site at the promoter of the *Cdh23* gene. Additionally, we observed the formation of new long-range inter-TAD spatial contacts. The origins of these contacts are particularly intriguing as they do not align with any currently understood mechanisms, suggesting novel aspects of chromatin organization and interaction that are yet to be fully elucidated.

In this work, we have developed a sequencing-based method that allows for precise quantification of gene expression changes resulting from TAD reorganization. Accuracy of this technique is compatible with digital PCR, affirming its utility for evaluating the effects of cis-acting mutations. Using this approach, we uncovered an alterations gene expression levels caused by TAD boundary disruption.

The patterns of gene expression changes we observed were not only tissue-specific in terms of magnitude but also in the direction of these changes. Our observations suggest that actively transcribed genes generally experience a decrease in expression levels, while repressed genes tend to gain new enhancer interactions, exemplified by the cases of interactions between *Slc29a3* and *Unc5b* in kidney and cerebellum. Similarly, the *Cdh23* gene, typically highly active in the cerebellum, exhibited a reduction in expression following TAD reorganization. Therefore, the effects of TAD boundary disruptions are expected to vary not only in amplitude but also in direction based on the epigenetic state and differentiation trajectory of the locus.

Funding: The study, including animal work and NGS sequencing, was supported by RSF Project 22-14-00247.

Acknowledgments: NGS data analyses were performed on the nodes of HPC cluster of the Institute of Cytology and Genetics, supported by the project No. 121031800061-7. ChIP-seq and Hi-C assays were performed using equipment of the Novosibirsk State University, supported by the Ministry of Education and Science of Russian Federation, grant # FSUS-2024-001.