# Role of chromatin structure in yeast gene expression regulation

Sasha Galitsyna

Kristina Perevoshchikova

Olga Kuzmich

## Introduction

Yeast is a model organism that has made it possible to study many aspects of the molecular biology of eukaryotes. However, in the relatively young field of structural biology of chromatin, this organism remains little studied. For example, the yeast genome forms structural domains[1], which are relatively compact compared to other eukaryotes and include a small number of genes. At the same time, small coregulatory clusters of neighboring genes are formed in yeast.[2] The correspondence between the structural domains of yeast chromatin and the regulatory gene domains is yet to be established. For example, for humans, the genome is split into coregulatory clusters, and coregulatory clusters poorly correspond to structural domains.[3] In this project, we aim to study a similar relationship between the structural organization of the genome and gene expression in yeast.

## Materials and methods

We collected and analyzed several data types from different sources. Firstly, we predicted structural domains from Hi-C maps with 500 bp resolution[4]. Domain borders were defined by calculating insulation score[5] (diamond window size 2000*2000 bp, boundary confirmed if stronger than 40% of all identified boundaries).

I. To calculate the correlation of gene expression, we aggregated RNA-seq datasets from several sources.[6] Gene expression values were normalized and presented as Transcripts Per Million (TPM), so it was possible to compare expressions between datasets. A total of 26 expression states were obtained. The mean correlation of expressions in a domain was taken as the average of all pairwise Spearman correlations of genes belonging to the domain. Gene correlations were also calculated according to the all-with-all principle within the same chromosome. Gene-gene correlations were converted to a bin-bin matrix to plot heat maps and compare them with Hi-C maps. Coregulatory domains were identified in the whole-genome using the k-means[7] method (cluster number = 2).

II. To study the effect of inserting a KanMX[8], only deletions of those genes were selected that did not lead to a significant change in gene expression on other chromosomes (log2(FC) > 1.5). Thus, global regulators were excluded, allowing focusing our study on the local changes that the cassette makes. There are 1054 deletions left to study. Next, all pairwise distances between the start of the cassette and the start of genes within the same chromosome were calculated, and the type of mutual orientation was assigned: “head-to-head”, “tail-to-tail”, “head-to-tail,” or “tail-to-head”.

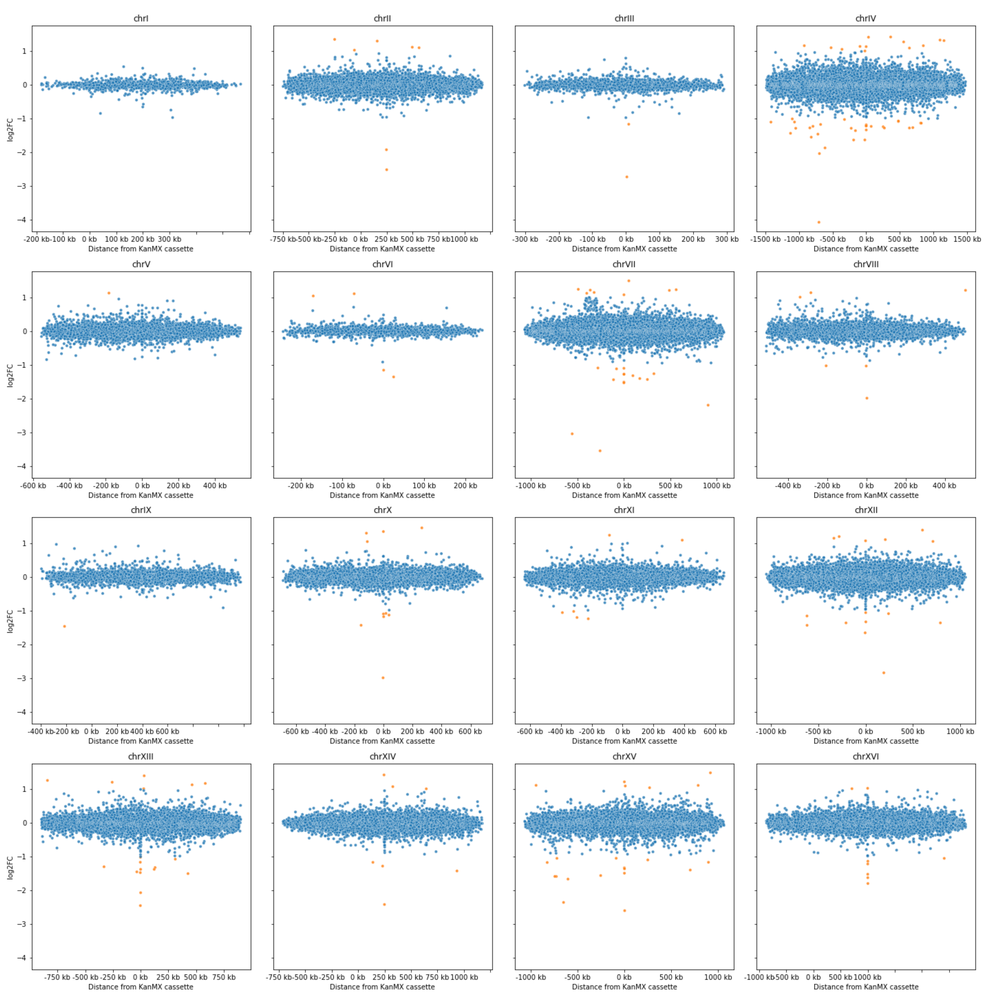



Fig. 4: Aggregated effect of KanMX cassette insertion on the expression of surrounding genes per each chromosome. The x-axis shows the distances between the start of the considered gene and the target cassette. Thus, the genes located above have a negative distance. The y-axis shows the binary logarithm of changes in gene expression. There is a noticeable decrease in the effect with increasing distance from the insert. Points with |log2(FC)| > 1 are highlighted in orange. Note the predominant accumulation of such points in the vicinity of the cassette. The reasons for the local and long-range effect of the insert on gene expression remain unclear.

## Future plans

Next, we plan to study the local clustering of genes in regulatory domains and compare them with structural domains. In addition, it remains interesting to study the mechanism of changes in the expression of genes adjacent to the KanMX cassette. We plan to analyze the relationship between changes in gene expression and their position in the same structural and/or regulatory domain. In the future, it is also planned to study the relationship between epigenetic marks and the level of chromatin openness. The question of the activity of transcription factors in the structural and regulatory domains remains of interest to us.

### I. Genes coexpression is not connected to yeast structural domain organization

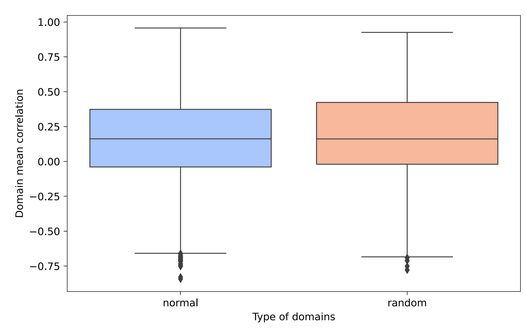



Fig. 1: Distribution of mean Spearman correlations over structural domains. Comparison of the observed domains isolated by the insulation method and random domains. We kept the number of domains per chromosome the same. Random domains have the same distribution of lengths as observed ones. According to the Mann-Whitney test, the average correlation does not differ between domain types (p_value = 0.90)

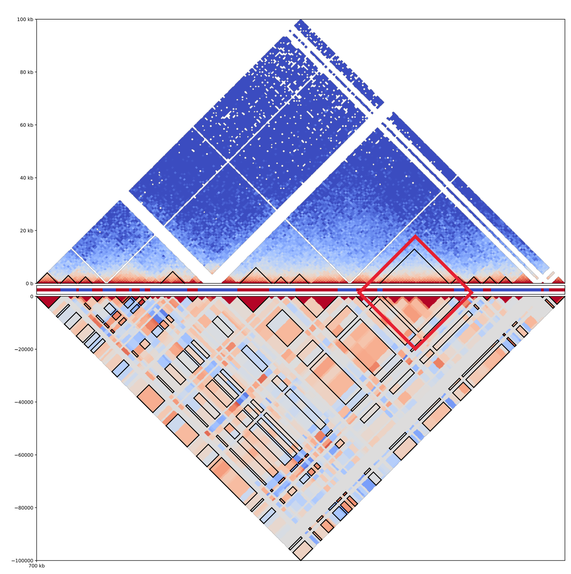


Fig. 2: Fragment of chromosome 12. Top: Hi-C map with segmentation into domains obtained by insulation score method. Bottom: heatmap of the correlations of genes located on the same region of the chromosome. Global coregulatory clusters were identified using the k-means method (cluster number = 2). It is noteworthy that segmentation into clusters demonstrates some correspondence between coregulatory clusters and structural domains (highlighted in red). In between the heatmaps, there is a representation of 2 clusters of the genes: blue and red. Only the borders of the red cluster are highlighted

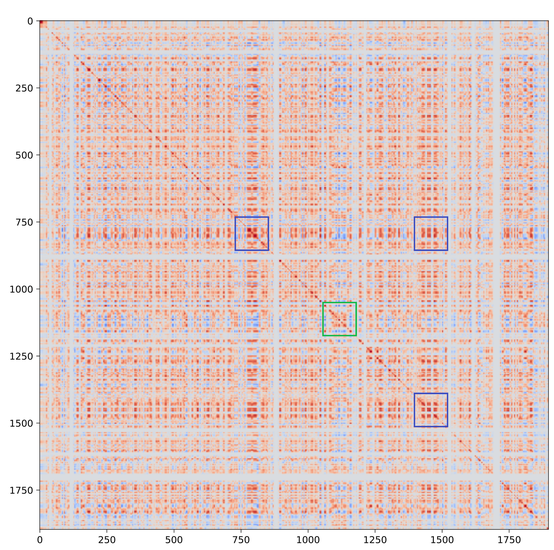
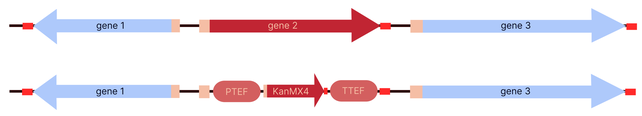


Fig. 3: Heatmap of the correlation of gene expression on the example of chromosome 16. The map is presented with a resolution of 500 nucleotides. The bright diagonal shows the correlation of 1 gene with itself. The correlation ranges from -1 to 1 (blue to red). Some local coregulatory clusters are distinguishable along the diagonal and their interaction with each other. It is noteworthy that some clusters interact positively with each other (highlighted in blue), and some (highlighted in green) have the opposite effect on other parts of the chromosome.

### II. Insertion of KanMX cassette impacts local gene expression

The KanMX cassette is inserted at the site of the deleted gene. Thus, it becomes possible to study yeast phenotypes with gene deletions. It has been shown, however, that cassette insertion leads to a significant change in nearest neighbor gene expression, potentially leading to misannotation of the function of the deleted gene.[9] In our project, we wanted to explore the broader local changes caused by the insertion.



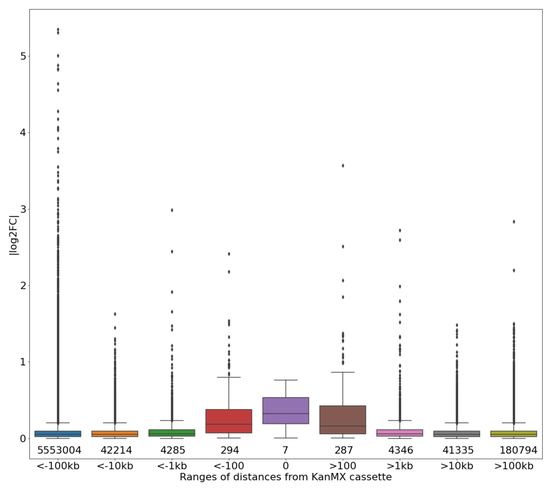




Fig. 5: Demonstration of the effect of KanMX cassette insertion on gene expression as a function of distance. The x-axis highlights the ranges of distances from the cassette start. The y-axis represents the log2(FC) module. It is easy to see the peak of the insert effect at a distance of [-1kb, 1kb] from the start of the tape. The effect decreases with distance from the insert. However, significant changes in gene expression over a long distance are also of interest.

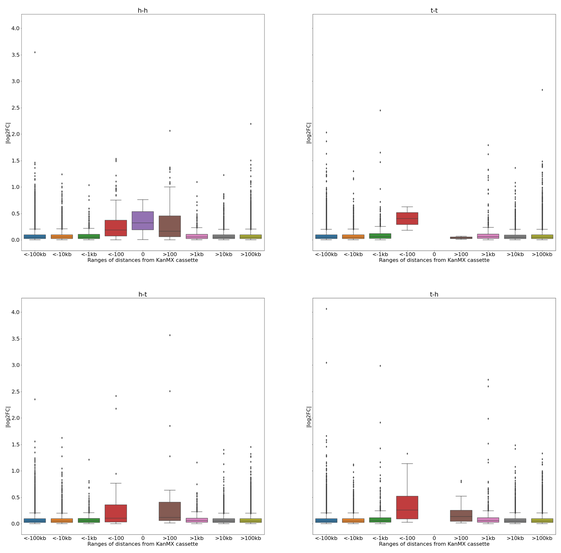



Fig. 6: Demonstration of the effect of insertion of KanMX cassette by the types of genes orientations: 'head-to-head', 'tail-to-tail', 'head-to-tail', 'tail-to-head'. The x-axis represents the ranges of genomic distances. The y-axis represents the absolute values of log2(FC).

We find out that KanMX insertion affects expression of adjacent genes and the effect on gene expression decreases with increasing distance from the KanMX cassette. It may be a consequence of local changes in the chromatin structure caused by the expression of the cassette, since it is under a strong promoter. Our work can serve as a basis for testing the hypothesis that changes in gene expression are associated with localisation in the same structural and/or regulatory domain.

## References

1. Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell. 2015;162: 108–119.
2. Cohen BA, Mitra RD, Hughes JD, Church GM. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 2000;26: 183–186.
3. Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, Reyes JC. Analysis of the relationship between coexpression domains and chromatin 3D organization. PLoS Comput Biol. 2017;13: e1005708.
4. Hsieh T-HS, Fudenberg G, Goloborodko A, Rando OJ. Micro-C XL: assaying chromosome conformation at length scales from the nucleosome to the entire genome. doi:10.1101/071357
5. Venev S, Abdennur N, Goloborodko A, Flyamer I, Fudenberg G, Nuebler J, et al. open2c/cooltools: v0.4.1. 2021 [cited 5 Aug 2022]. doi:10.5281/zenodo.5214125
6. GEO datasets: GSE114077,GSE115556,GSE135568,GSE153609,GSE175794
7. Grisel O, Mueller A, Lars, Gramfort A, Louppe G, Prettenhofer P, et al. scikit-learn/scikit-learn: scikit-learn 1.1.1. 2022 [cited 5 Aug 2022]. doi:10.5281/zenodo.6563718
8. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752.
9. Egorov AA, Alexandrov AI, Urakov VN, Makeeva DS, Edakin RO, Kushchenko AS, et al. A standard knockout procedure alters expression of adjacent loci at the translational level. Nucleic Acids Res. 2021;49: 11134–11144.