The scientific goal of the project is to analyze the impact of structural variants (SVs, such as deletions, duplications, insertions, inversions, translocations, etc.) and single nucleotide polymorphisms (single molecule polymorphism, SNP) on the three-dimensional structure of the human genome. We want to understand how rewiring (i.e. changing the set of chromatin loops) of the genome in different cell types disrupts gene expression, and thus leads to the tumor formation process. Public and private datasets from the large-scale next generation sequencing (NGS) experiments will be used to implement the project, and computational methods developed by Principal Investigator (PI), such as statistical data analysis, machine learning and computer simulations. The project will finish with the publicizing the web server and the source code of the proprietary 3D-OME algorithm predicting the three-dimensional structure of the mammalian genome based on DNA sequencing data, biopolymer theory and biophysical properties of chromatin. We want to show that modeling allows to describe the transformation of the local three-dimensional structure of chromatin, and the disruption of expression of genes in the genomic domain after a change in the DNA sequence (single nucleotide mutation, deletion, duplication, insertion, etc.).

According to current knowledge, structural changes are a major factor in the variability of the human genome. This means that most of the changes taking place in the human genome are associated with the rearrangement of its spatial structure. Analysis of the changes in the spatial conformation of the genome in response to the appearance of structural changes can bring us closer to understanding the mechanisms of formation of structural variants and the way in which they modify the functioning of the genome. Such knowledge would enable a better understanding the mechanisms leading to pathological conditions and could improve the effectiveness of treatment of genetic diseases and support the development of personalized medicine. It could also help in understanding the mechanisms of genome self-organization.

The project focuses on analyzing the DNA sequences for the human population from the '1000 Genomes Project' and three-dimensional data for various cell types from the 4DNucleome project in which the principal investigator participates. We will build the mechanistic model linking the DNA sequence, chromatin structure and the expression of genes within topologically associating domain (TAD). Next, we plan to use the developed model to better describe evolutionary processes leading to local spatial conformation changes. To this end, we will analyze the structural variants and changes that they cause in the spatial organization of the genome and expression of genes found in the representatives of the human population. It also provides deep biological analysis, based on the knowledge gained in the study of structural variants in healthy people, structural changes occurring in cancer cells.