The rapid progress of the human genome sequencing technology allowed comparative analysis of molecular sequences from different donors. It has exposed drastic genetic diversity in the human population.

Our research conducted in collaboration with Baylor College of Medicine has contributed to changing the perception of the genome as stably coded information. Structural variations identified in genomes include deletions, duplications, inversions and balanced translocations. Molecular mechanisms of these rearrangements are not yet fully understood, and researchers are looking for reasons why certain loci on chromosomes are highly unstable. Our current research has shown that genome rearrangements can be mediated by repetitive fragments of the genome. Initially, it was shown that the crucial players here were segmental duplications, and then also much shorter, mobile genetic elements, i.e. transposons from eg. the HERV or LINE family.

The structural variants of the genome are extremely important as a tool in the analysis of molecular evolution, but mainly because of their potentially pathogenic character. A number of genetic disorders have been identified, the basis of which is e.g. the deletion of a certain DNA fragment (in previous projects we examined, among others: autism, epilepsy, developmental disorders).

Thanks to the cooperation between geneticists and computer scientists, in recent years extremely effective methods for identifying structural changes have been developed on the basis of sequential data as well as those from array comparative hybridization technology. However, the interpretation of detected changes remains a much bigger challenge. It is relatively rare that the rearrangement destroys a specific gene associated with a given disease. Most often, the interpretation of genetic variation is very difficult and for its adequate analysis, the cooperation of interdisciplinary teams consisting of biologists, physicians, mathematicians and computer scientists is necessary.

In this project, we focus on the interpretation of structural changes in the genome that integrates information from different experiments. The combination of knowledge about dose sensitivity of particular genes, information on the location of regulatory elements and the structure of chromatin in a given region of the genome, will allow to build a model that allows prediction of the pathogenicity of a given rearrangement.

In close cooperation with geneticists, we intend to create a computational tool supporting physicians in the diagnosis of genetic disorders. The probabilistic model proposed by us will allow to identify the most important features that are responsible for the phenotype. The developed algorithms will be tested on data on genome rearrangements of tens of thousands of patients gathered at Baylor College of Medicine. In addition, we plan long–read (Nanopore) sequencing for a selected group of patients, which we hope will allow to explain the occurrence of very diverse symptoms for the same genomic structural variant.