

# Genomic Map of Poland project—how the human reference genome is constructed

Aleksandra Świercz

Institute of Computing Science, Poznań University of Technology  
&

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań

22 November 2022

## Abstract

The first draft of the human reference genome was published more than two decades ago. Due to the technological limitations, it was full of gaps, especially in highly repetitive regions of centromeres and telomeres. Since that time, the reference genome has been updated and corrected several times; until now, the newest version is GRCh38. DNA sequencing has become much cheaper in the recent years, and large sequencing projects are more popular. Many countries are sequencing large populations of individuals from their countries, which results in population genetic maps, with the most common variants that appear in the population. However, people noticed that the reference genome GRCh38, although well-annotated, is not sufficient to represent a given population. Several attempts were made in this area, and a few reference genomes have been already published (e.g. Korea, China, Japan).

In our project called Genomic Map of Poland one of the goals is to create *de novo* a reference genome, based on the trio: mother, father, and child. In the pipeline, we used several technologies: short read, long high-quality reads (PacBio HIFI), artificial long reads (stLFR), long-distance contact reads (HiC), and ultralong reads (Nanopore). The resulting scaffolds, spanning the whole chromosome, were compared with the GRCh38 reference genome, showing differences between the references.

## About the presenter

Aleksandra Świercz received a PhD degree and habilitation degree in bioinformatics at the Institute of Computing Science at Poznań University of Technology. She works as an Associate Professor at the same university and also at the Institute of Bioorganic Chemistry, Polish Academy of Sciences. She gives courses for students of Computer Science and of Bioinformatics on Programming, Algorithms Design, Computational Complexity, Microarray Data Analysis, High Throughput Sequencing Data Analysis, and Structural Variations in Genomes. Her research interests include algorithms design, hyper-heuristics, combinatorial problems in molecular biology, especially high-throughput sequencing analysis. She cooperates with many universities in bioinformatics analysis of NGS data. She

led and took part in different national grants, most recently at the Genomic Map of Poland project. She gave courses on NGS data analysis at the workshops for the Romanian Bioinformatics Society in Timisoara and Bucharest in 2017 and 2018. She is a board member of the Polish Bioinformatics Society, and member of the Operational Research in Computational Biology, Bioinformatics and Medicine Working Group of EURO.