

## Metagenomic search for the protozoa in atherosclerotic plaques

Zarubin A.<sup>1\*</sup>, Markov A.<sup>1</sup>, Sharysh D.<sup>2</sup>, Puzyrev V.<sup>1,2</sup>, Nazarenko M.<sup>1,2</sup>

<sup>1</sup>Research Institute of Medical Genetics of Tomsk NRMС, Tomsk, Russia

<sup>2</sup>Siberian State Medical University, Tomsk, Russia

\* e-mail: aleksei.zarubin@medgenetics.ru

**Key words:** metagenome, protozoa, atherosclerotic plaques

*Motivation and Aim:* Several reports have shown the relevance of infection in development of atherosclerosis. However, the bacterial DNA of atherosclerotic plaques mainly was investigated by metagenomic approaches. To search for the protozoa by using genome-wide sequencing data of human atherosclerotic plaques (AP).

*Material and Methods:* We used the results of whole genome sequencing of DNA isolated from 12 atherosclerotic plaques (SRA149235) and 2 blood samples from 1000 Genomes Project (HG00096, HG00099). Sequence reads were aligned to the human reference genome (hg19) using Bowtie2. The taxonomic classification of the reads was performed with Kraken2 using a curated microbial genome database containing archaea, bacteria, fungi, protozoa, viruses reference genome sequence.

*Results:* Pre-alignment on the human genome, even with “very-sensitive” option of Bowtie2, cannot completely clean up human gene sequences. Kraken2 determined 13 to 92 % of the human gene sequences in unmatched reads of different AP samples. We found sequences of *Toxoplasma gondii* genome in 8 samples (about 1678 reads per sample). These sequences were shown to be mapped to contigs (NW\_017384310.1, NW\_017384809.1). These contigs were also found in samples from 1000 Genomes Project.

*Conclusion:* There were no protozoa in genome-wide sequencing data of human atherosclerotic plaques. The reads that mapped to the *Toxoplasma* genome are false positive. Some microbial reference genomes probably to contain human DNA sequences not presented in the human reference genome assembly.