

VILNIAUS UNIVERSITETAS

Gyvybės mokslų centras



Molekulinės biologijos studijų programos magistrantė

Kotryna KELIUOTYTĖ

Magistrinis darbas

*Viso genomo sekoskaitos sistema haplotipo nustatymui*

Darbo vadovas:  
Dr. Justinas Šlikas

Vilnius 2022

VILNIAUS UNIVERSITETAS

Gyvybės mokslų centras

KOTRYNA KELIUOTYTĖ

**Viso genomo sekoskaitos sistema haplotipo nustatymui**

Molekulinės biologijos magistro studijų baigiamasis darbas

Darbas atliktas UAB „Thermo Fisher Scientific Baltics“

Mokslinių tyrimų ir eksperimentinės plėtos centre

### **SANTRAUKA**

Haplotipavimas – individualios molekulės nukleotidų sekos nustatymas, yra svarbus diploidinio genomo charakterizavimui ir sąsajai su fenotipu. Šiuo metu plačiausiai prieinami naujos kartos sekoskaitos bibliotekos paruošimo metodai generuoja trumpus nuskaitymus ir nustato genotipą – genetines variacijas nežinomos kilmės molekulėse. Dėl trumpų nuskaitymų genetinės variacijos, nutolusios toliau nei viena nuskaitymų sąranka, nėra tiesiogiai susiejamos, todėl haplotipo nustatymui taikomi *in silico* analizės metodai. Kadangi netiesioginiai metodai remiasi statistiniu genotipų palyginimu ir tiksliausiai analizuoja tik trumpus genetinius regionus, viso genomo haplotipavimui, retų ir *de novo* mutacijų paieškai reikalingi molekuliniai arba tiesioginio haplotipavimo metodai.

Šio darbo tikslas – sukurti sistemą, gebančią pateikti haplotipo informaciją paraleliai vykdant pilno ilgio genomo sekoskaitą. Tikslui įgyvendinti pritaikyti ant polistireno dalelių imobilizuoti oligonukleotidai su atsitiktinėmis sekomis bei dalelių barkodais ir pasiekta virtuali erdvėskyra tarp DNR molekulių. Darbo metu optimizuotos naujos kartos sekoskaitos bibliotekos paruošimo sąlygos – oligonukleotidu modifikuoto nukleotido įjungimo ir pratęsimo efektyvumas, optimalus dalelių, genominės DNR ir oligonukleotidu modifikuoto nukleotido kiekis reakcijoje, taip pat vertinamos šalutinės sistemos reakcijos ir jų pašalinimo efektyvumas. Optimizuotas naujos kartos sekoskaitos bibliotekos paruošimo metodas pritaikytas žmogaus genomo sekoskaitai ir įvertintas trijų skirtingų sistemų haplotipavimo efektyvumas.

VILNIUS UNIVERSITY

Life Sciences Center

KOTRYNA KELIUOTYTĖ

**Haplotype-resolved whole genome sequencing**

Master thesis

The work was performed at “Thermo Fisher Scientific Baltics” UAB  
Research and Development Center

**SUMMARY**

Haplotyping – the process of determining co-inherited alleles occurring on a single strand of DNA, has a high priority for diploid genome characterization and its association with the phenotype. Currently, the most widely used next generation sequencing workflows generate short sequencing reads and represent genotypes – genetic variations in the unknown molecular origin. Due to short reads, genetic variations farther than the contig length are not resolved directly and require additional *in silico* methods. Because indirect haplotyping is based on statistical genotype data comparison and most accurately resolve short haplotypes, molecular or direct haplotyping methods are required for whole genome analysis, as well as for rare and *de novo* mutation research.

The aim of this work was to develop a system capable of resolving haplotype information in parallel with whole genome sequencing. Oligonucleotides immobilized on polystyrene spheres with random primers and bead barcodes were used to achieve virtual space separation between DNA molecules. The conditions for next generation sequencing library preparation method were optimized - the efficiency of oligonucleotide-tethered nucleotide incorporation and extension, the optimal number of particles, amount of genomic DNA and oligonucleotide-tethered nucleotides in the reaction, as well as non-specific reactions and efficiency of side product removal. The optimized next generation sequencing library preparation method was applied to human whole genome sequencing and the haplotyping efficiency was evaluated in three different systems.