

**APPLICATION OF NOVEL DATA MINING  
METHODS IN HUMAN EPIGENOME  
RESEARCH AIMED AT EARLY  
DIAGNOSIS OF COMPLEX  
NON-MENDELIAN DISEASES**

Ph.D. candidate: Ieva Meržvinskaitė

Study period: 2016 - 2020

Ph.D. advisor: Dr. Remigijus Paulavičius

Dr. Juozas Gordevičius

# Research object

DNA modifications: DNA methylation, Histone acetylation plays a major role in cell specialization

DNA modifications changes during aging

Age is the major risk factor of cancer

If aging and cancer related patterns would be identified, it would be possible to diagnose cancer before its onset

# Study plan. 1<sup>st</sup> year

| Planned  | Done  |
|--|---|
| <p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <p>1.1. Susipažinti su esančiais mašininio mokymosi algoritmais, išanalizuoti jų taikymus tiriant didelius biologinius duomenis</p> <p>1.2. Išanalizuoti statistinius metodus naudojamus tiriant didelius biologinius duomenis</p> | <p>Peržiūrėta su tyrimu susijusi literatūra.<br/>Sukurta duomenų bazė.<br/>Sukurta aplikacija gautiems rezultatams talpinti.<br/>Atlikta pirminė duomenų analizė.</p> |
|  | <p>Dalyvauta The Second NESUS Winter School 2017 (vasario mėn.)</p>   |
|  | <p>2017 m. gegužės 29 dieną skaitytas pranešimas grupės seminare.</p>   |
| <p>1 egzaminas</p>   | <p>Egzaminui ruošiamasi.</p>  |

# Related researches

Age predictor ( *Hannum et al (2013); Horvath, Genome Bio. 2013* ),

Accelerating aging ( *Horvath, Genome Bio. 2013* ),

Search for differentially methylated regions (DMR)

Epigenetic drift (*twins studies. Gregory Hannum, Justin Guinney et al (2013)*)

# Primary results

16 different case (cancer)/control datasets from TCGA. Data was imported to created database.

Each dataset has ~100 samples in case/control groups.

~450k features (DNA methylation exp. values in various genome parts)

Data pre-processing and analysis was applied.

During the analysis different linear models were applied.

Primary results revealed the existence of aging related patterns in the datasets.

# Next steps

| Planned  | Done |
|--|------|
| <p>Mokslinio tyrimo vykdymas:</p> <p>2.1. Tyrimo metodikos sudarymas:</p> <ul style="list-style-type: none"><li>2.1.1. Sudaryti didelių biologinių duomenų analizės modelius</li><li>2.1.2. Parašyti analizes įgyvendinančius skriptus/programas</li></ul> <p>2.2. Teorinis tyrimas:</p> <ul style="list-style-type: none"><li>2.2.1. Patikrinti biologinių duomenų analizei skirtų modelių teisingumą</li></ul> <p>2.3. Empirinis tyrimas:</p> <ul style="list-style-type: none"><li>2.3.1. Panaudojant sukurtus modelius ir skriptus/programas ištirti duomenis</li></ul> <p>2.4. Gautų rezultatų analizė, apibendrinimas, išvadų parengimas:</p> <ul style="list-style-type: none"><li>2.4.1. Automatizuoti ir optimizuoti analizę, rezultatų pateikimo būdus</li><li>2.4.2. Pasiūlyti naujus analizės metodus, pasiūlyti kaip pagerinti esančius</li></ul> |      |
| 2 egzaminai  |      |
| Pradėti rašyti straipsnį   |      |

**Thank you!**