PhD position in context specific disease risk prediction models

We are looking for a motivated PhD student for an exciting four-year funded project on Ethnicity- and ancestry-specific disease risk prediction models. A good candidate has a master degree or equivalent in population genetics, genetic epidemiology, disease genetics, bioinformatics or other related fields. The general project description below can be tailored to reflect personal interests of a motivated candidate.

Deadline for applications: 24th March 2024. Please send your CV and motivation letter to mait.metspalu@ut.ee

Title of project in Estonian: Kultuuri ja päritoluspetsiifilised haiguste ennustusmudelid
Title of project in English: Ethnicity- and ancestry-specific disease risk prediction models

PhD Project plan

Enhancing Disease Risk Assessment

Genome sequencing can identify rare monogenic diseases that affect approximately 5-7% of the population. However, the majority of diseases are complex interplay of various genetic, environmental, and lifestyle factors. To address this complexity, researchers have devised Polygenic Risk Scores (PRS) as a means to amalgamate information from numerous common genetic variants linked to a particular disorder (discovered in Genome-Wide Association Studies or GWAS). PRS helps evaluate an individual's relative genetic risk for that disorder. Multiple studies have substantiated the predictive power of PRS in addition to conventional risk prediction models, as demonstrated in conditions such as coronary artery disease, breast cancer, colorectal cancer, and skin cancer (1).

As a result, PRS serves several vital purposes:

1. Enhancing the stratification of populations in national screening programs by offering more precise genomics-based risk assessments for early disease detection.
2. Assisting individuals in making informed health-related decisions through genomic counseling.
3. Identifying individuals at an elevated risk for co-morbitities.
4. Grouping individuals based on distinct biological pathways relevant to the progression of their diseases.

However, successful application of PRS in medical setting faces several challenges, including:

1. Ensuring the transferability of PRS across diverse populations (2,3) considering also admixed individuals (4). A recent study (5) highlighted that the accuracy of PRS-based prediction depends on the genetic distance of the target individual from the GWAS discovery cohort, and varies even within relatively genetically homogeneous groups. A likely explanation for this accuracy difference are differences in LD patterns and allele frequencies across populations, affecting marginal effects inferred in GWAS. Careful consideration of ancestral components allowed another study (6,7) to suggest that causal effects are highly consistent between ancestries.
2. Managing uncertainty in estimating an individual's PRS (8).
3. Combining polygenic risk with monogenic variants with high penetrance to enhance our understanding of disease risk and enable more precise identification of high-risk individuals for targeted interventions and screenings (9–12).
4. Adjusting PRS to reflect stratification of the population according to sex, age, socioeconomic background, and other traditional risk factors (9) and context. Underrepresented study groups often face healthcare disparities, such as gender-based misdiagnosis (13). We will extend our previous work on elucidating the interplay between genetic determinants, social factors, and lifestyle by estimating the impact of PRS on common chronic diseases within different strata of the Estonian Biobank cohort (14).

The specific goals of this PhD project are:

Ancestry specific polygenic risk scores (PRS)

PRS accuracy depends on genetic distance from the GWAS discovery cohort and varies within genetically homogeneous groups (8). To improve this:
(i) We'll enhance our recent work on ancestry-informed PRS for admixed individuals (4), combining local ancestry inference with effect sizes from relevant GWAS.
(ii) We'll assess PRS prediction accuracy in EstBB relative to Estonia's genetic diversity, using various cohorts (UKBB, FinnGen) for summary statistics.
(iii) We'll compare LD-scores between GWAS source cohorts and genetic subgroups within EstBB, addressing PRS transferability challenges.
(iv) We'll explore methods, like leveraging ancestral components (7), to mitigate PRS transferability issues related to population structure.

1) Inouye, M., *et al.* (2018) *J. Am. Coll. Cardiol.*, **72**, 1883–1893.
2) Martin, A.R., *et al.* (2017) *Am. J. Hum. Genet.*, **100**, 635–649.
3) Kachuri, L., *et al.* (2023) *Nat. Rev. Genet. 2023*, **10**, 1–18.
4) Marnetto, D., *et al.* (2020) *Nat. Commun.*, **11**.
5) Ding, Y.,*et al*. (2023) *Nat. 2023 6187966*, **618**, 774–781.
6) Hou, K., *et al.* (2023) *Nat. Genet. 2023 554*, **55**, 549–558.
7) Hu, S., *et al*. (2023) *bioRxiv*, 2023.08.08.552281.
8) Ding, Y., *et al*. (2022) *Nat. Genet.*, **54**, 30–39.
9) Mars, N., *et al.* (2020) *Nat. Commun. 2020 111*, **11**, 1–9.

10) Torkamani, A., *et al*. (2018) *Nat. Rev. Genet.*, **19**, 581–590.
11) Khera, A. V, *et al.* (2018) *Nat. Genet.*, **50**, 1219–1224.
12) Berga-Švītiņa, E., *et al*. (2023) *Cancers (Basel).*, **15**.
13) Newman-Toker, D.E., *et al*. (2014) *Diagnosis (Berlin, Ger.)*, **1**, 155–166.
14) Läll, K., *et al.* (2017) *Genet. Med.*, **19**, 322.

Supervisors:

Dr. Mait Metspalu (https://scholar.google.com/citations?user=xOPd7N0AAAAJ&hl=en) is an expert in population genetics and will lead the supervision.

Dr. Reedik Mägi (https://scholar.google.com/citations?hl=en&user=mKbe4v0AAAAJ) is a world leading expert in epidemiological genetics and polygenic risk scores in particular.

Dr. Luca Pagani (https://scholar.google.com/citations?hl=en&user=2TYX99YAAAAJ) is an expert in population genetics and has led research into ancestry specific/informad polygenic risk scores.

Dr. Vasili Pankratov (https://pubmed.ncbi.nlm.nih.gov/?term=pankratov+vasili&sort=date) is an expert in population genetics and evolutionary biology.

PhD projekti eestikeelne kokkuvõte

Kultuuri ja päritoluspetsiifilised haiguste ennustusmudelid

Juhendajad:

Dr. Mait Metspalu

Dr. Reedik Mägi

Dr. Luca Pagani

Dr. Vasili Pankratov

Genoomi järjendamine võimaldab tuvastada haruldasi monogeenseid haigusi, mis mõjutavad 5-7% inimestest. Polügeensed riskiskoorid (PRS) kombineerivad kombineerivad aga väga paljude geneenivariantide mõju suhtelise geneetilise riski hindamiseks enamuste haiguste puhul. Selliste haiguste põhjused on tavaliselt kombinatsioon keskkonnamõjust ja väga paljude geenivariantide mõjust nii, et iga üksiku geenivariandi mõju on imepisike. PRSe saab kasutada a) tõhustamaks rahvastiku stratifikatsiooni sõeluuringuteprogrammides, b) terviseotsuste langetamisel, tuvastamaks kaasuvaid haigusi ja rühmitades inimesi bioloogiliste radade järgi. Väljakutsed hõlmavad PRS-i ülekantavust erinevatele populatsioonidele, hindamise ebakindlust, polügeense ja monogeense riski integreerimist ning demograafiliste tegurite arvesse võtmist. Päritoluspetsiifilise PRS-i täpsus varieerub homogeensetes rühmades, innustades uurimistööd päritoluinformeeritud PRS-i täiustamiseks ja ennustamistäpsuse hindamist konkreetsetes andmekogudes, nagu Eesti Geenivaramu. Doktoriprojekti eesmärgid hõlmavad päritolu spetsiifilise PRS-i täpsustamist, ennustamistäpsuse hindamist mitmekesistes kogudes, LD-skooride võrdlemist ja meetodite uurimist PRS-i ülekantavuse probleemide leevendamiseks.

PhD project summary

Ethnicity- and ancestry-specific disease risk prediction models

Supervisors:

Dr. Mait Metspalu

Dr. Reedik Mägi

Dr. Luca Pagani

Dr. Vasili Pankratov

Genome sequencing uncovers rare monogenic diseases affecting 5-7% of people. Polygenic Risk Scores (PRS) amalgamate common genetic variants from Genome-Wide Association Studies (GWAS) to assess relative genetic risk for complex diseases. PRS enhances population stratification in screening programs, aids health decisions, identifies comorbidities, and groups individuals by biological pathways. Challenges include PRS transferability across diverse populations, uncertainty in estimation, integrating polygenic and monogenic risk, and adjusting for demographic factors. Ancestry-specific PRS accuracy varies within homogeneous groups, prompting research on enhancing ancestry-informed PRS and assessing prediction accuracy in specific cohorts like the Estonian Biobank. Goals of a PhD project include refining ancestry-specific PRS, assessing prediction accuracy in diverse cohorts, comparing LD-scores, and exploring methods to mitigate PRS transferability issues.