UNIVERSITY OF TARTU
1632

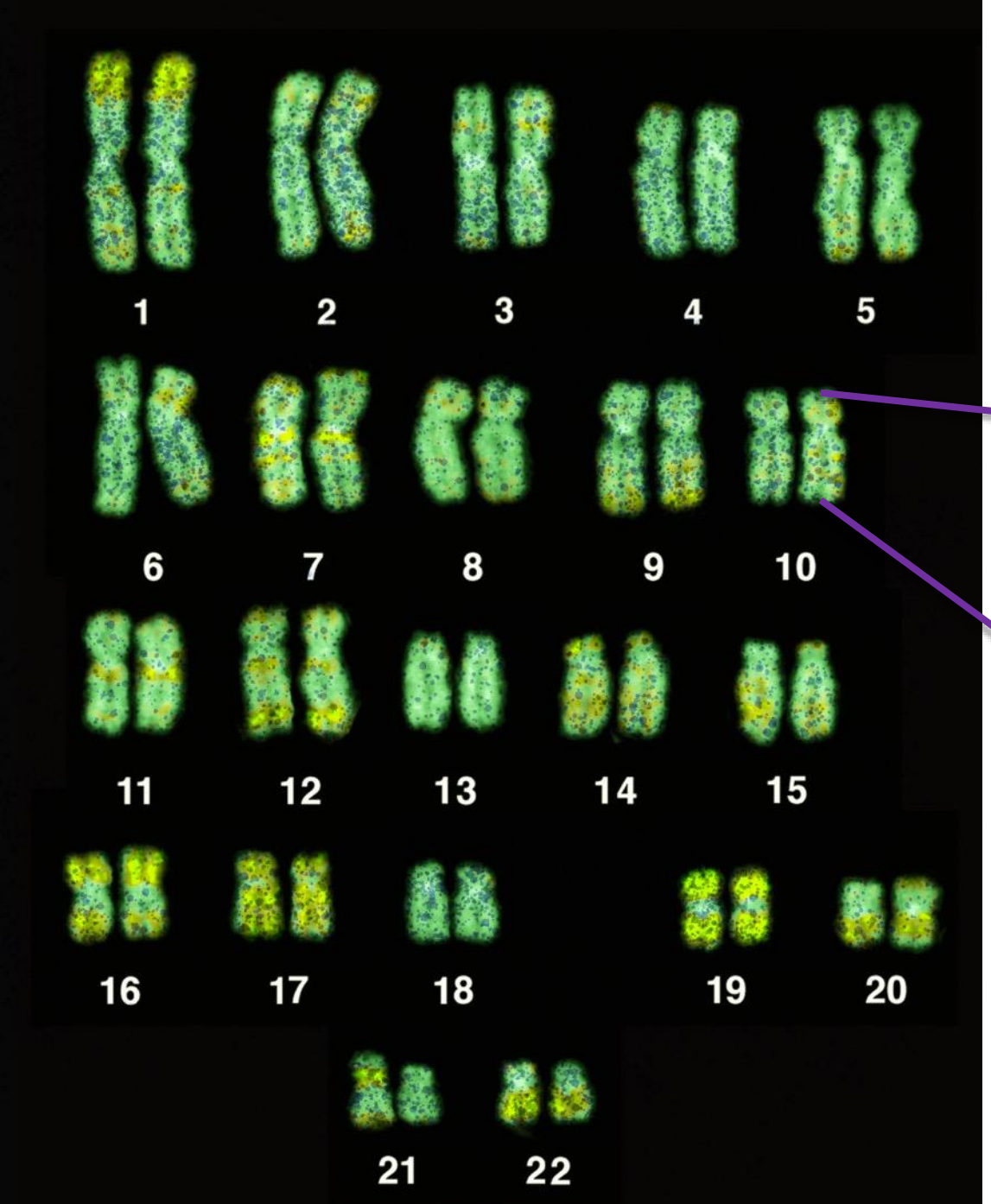
CONSIDERATIONS ON DATA SOURCES AND DATA TYPES

Sulev Reisberg, Ph.D.

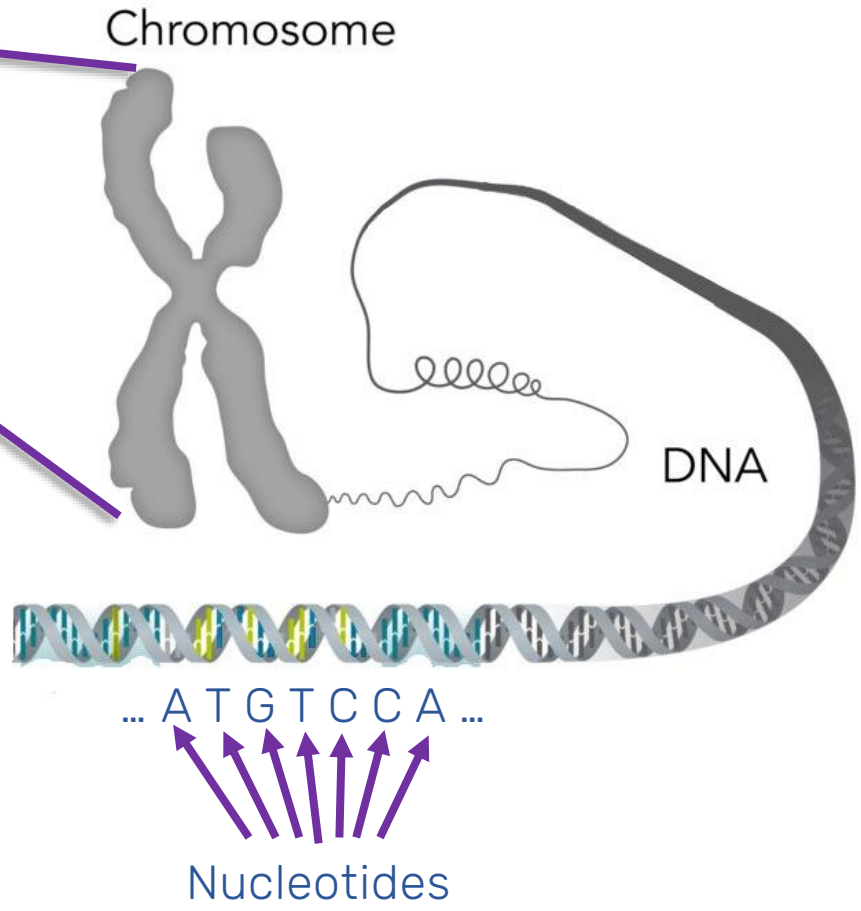
Research Fellow
of Health Informatics

September 2024

Human DNA – split into 2x23 DNA molecules called chromosomes



Source: https://commons.wikimedia.org/wiki/File:Chromosome_set_humans.jpg



Source: https://commons.wikimedia.org/wiki/File:Chromosome-DNA-gene_copy.jpg

We can illustrate this as follows

DNA of a person

…TGACAGTAACGGT…

3 billion letters

Approximately 50 m tall pile of papers when printed

At least 100-200 GB file size per person

Source: https://commons.wikimedia.org/wiki/File:Rijksmuseum.Amsterdam.jpg

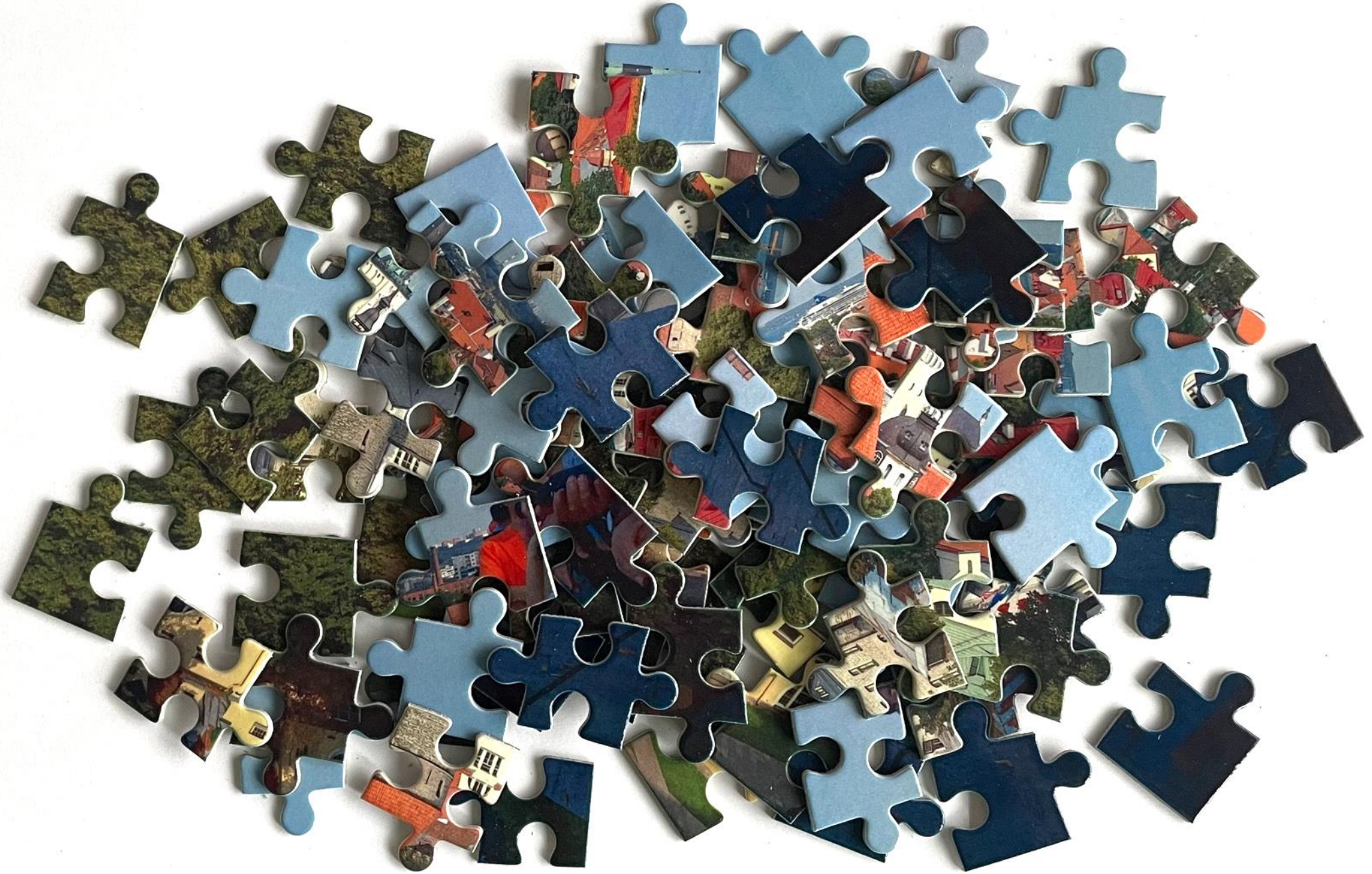
Techniques to determine DNA sequence

DNA of a person

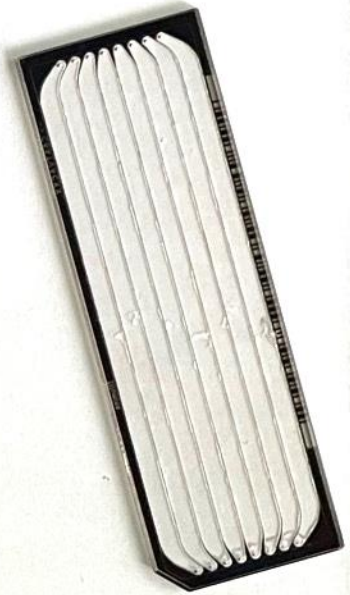
Each pixel stands for ~4000 nucleotides

Shotgun technology

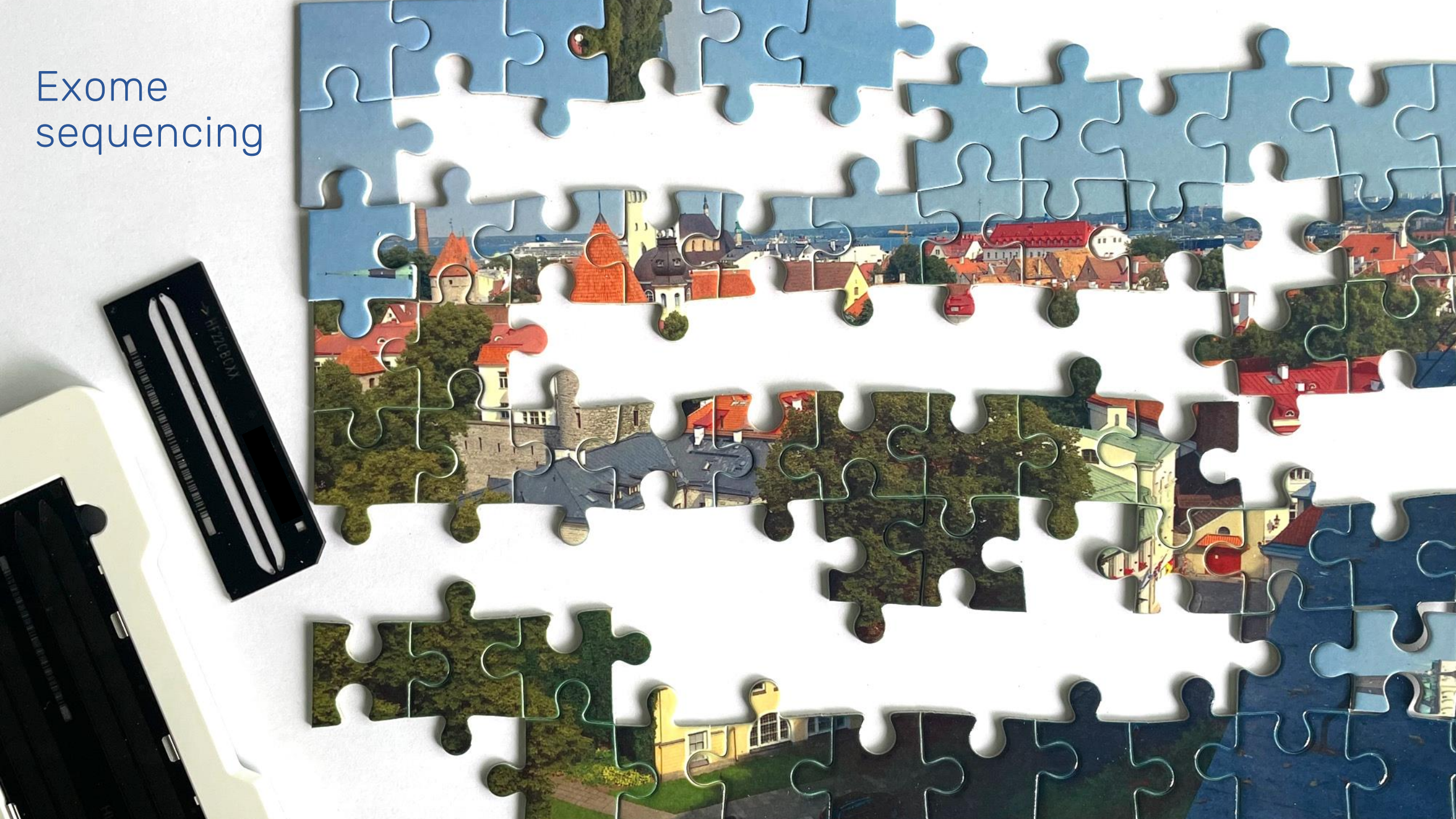
NextGen sequencing



Whole genome sequencing



Exome sequencing

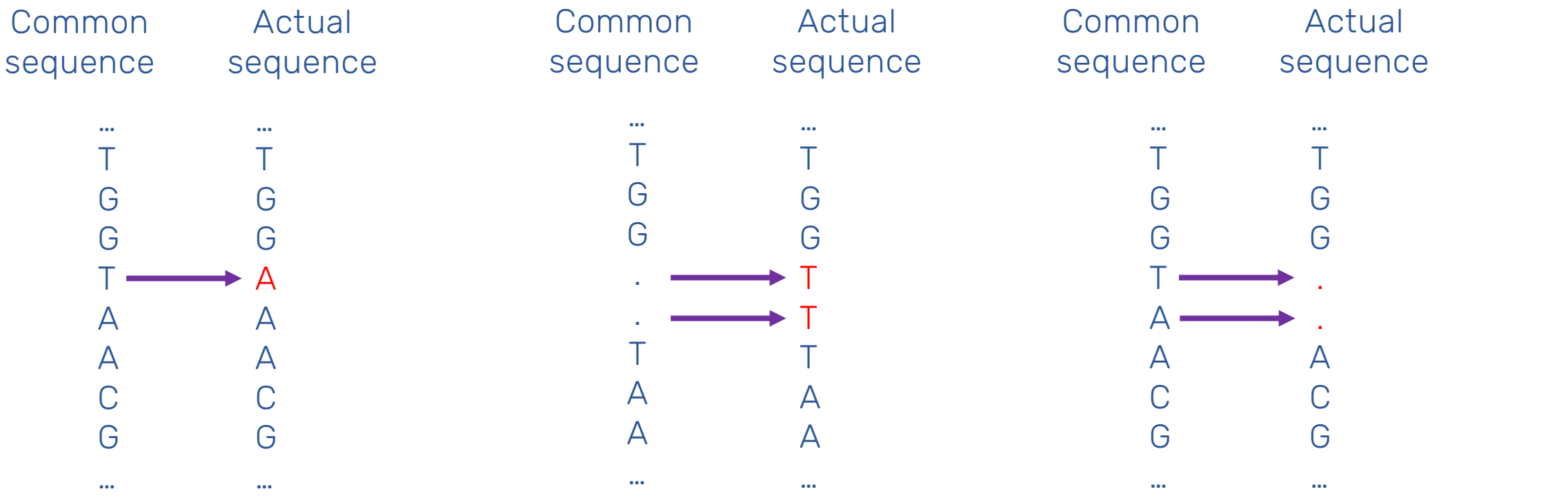


Genotyping

Long-read sequencing

Source:
https://www.pacb.com/blog/smrt-cell/

Examples of DNA variants



Single nucleotide variants (SNV)

Insertion

Deletion

etc.

Indels and larger structural variations

Easier to detect

Harder to detect

However…

There is more and more evidence, that structural variants play a big role in PGx

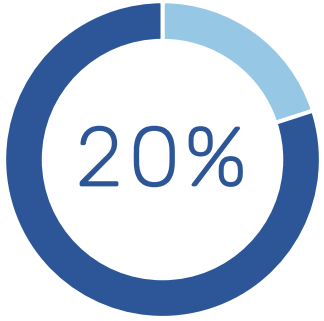
But these are hard to detect with common sequencing/genotyping methods

Big hope on long-read sequencing

ESTONIAN BIOBANK

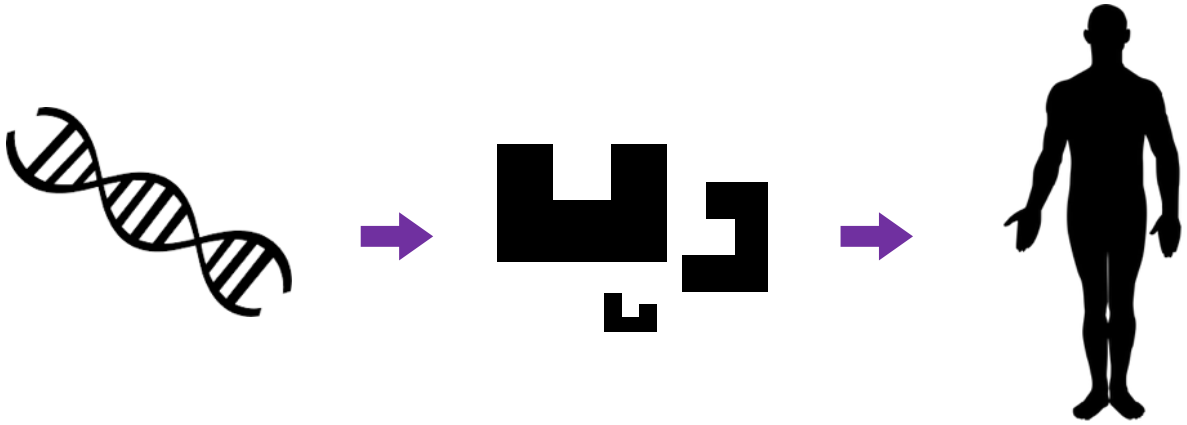
Eesti geenivaramu

A population-based biobank with a current cohort size of more than **200,000 individuals** provides data for researchers and innovators to examine how DNA affects health outcomes and to explore ancestral histories.



of Estonian adult population has joined the Estonian Biobank

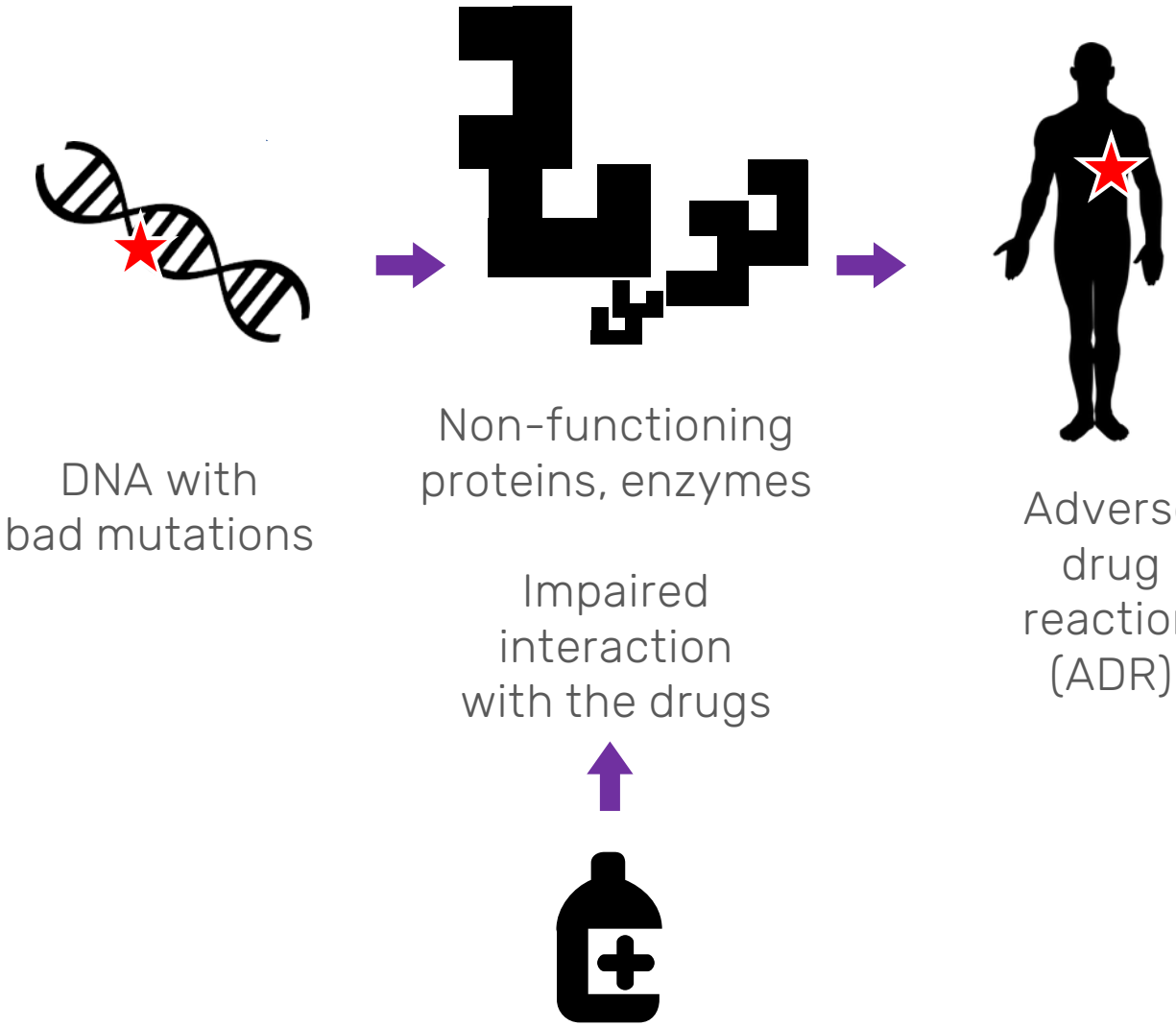
Providing PGx feedback for Biobank participants



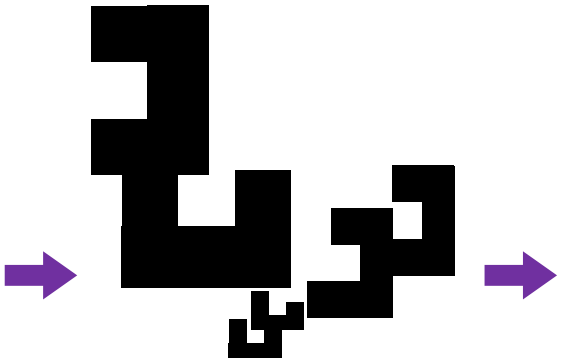
DNA

Proteins, enzymes

Body


DNA with
bad mutations
Non-functioning
proteins, enzymes
Impaired
interaction
with the drugs
Adverse
drug
reaction
(ADR)

DNA with
bad mutations

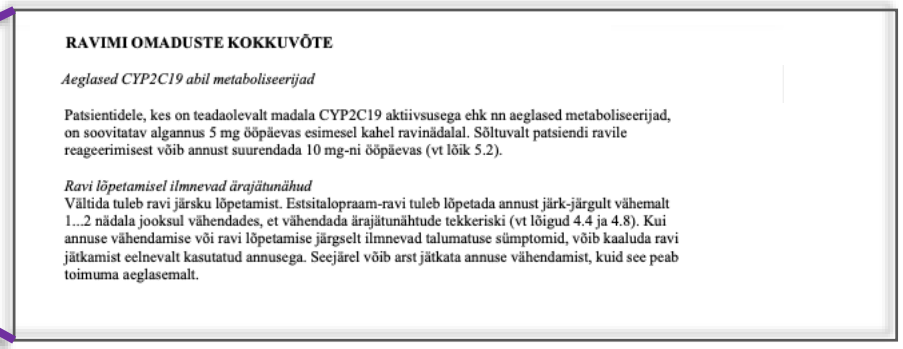


Non-functioning
proteins, enzymes

Adverse
drug
reaction
(ADR)

Impaired
interaction
with the drugs

A lot of knowledge exist already



RAVIMI OMADUSTE KOKKUVÕTE

Aeglased CYP2C19 abil metaboliseerijad

Patsientidele, kes on teadaolevalt madala CYP2C19 aktiivsusega ehk nn aeglased metaboliseerijad, on soovitatav algannus 5 mg ööpäevas esimesel kahel ravinädalal. Sõltuvalt patsiendi ravile reageerimisest võib annust suurendada 10 mg-ni ööpäevas (vt lõik 5.2).

Ravi lõpetamisel ilmnevad ärajätunähud
Vältida tuleb ravi järsku lõpetamist. Estsitalopraam-ravi tuleb lõpetada annust järk-järgult vähemalt 1...2 nädala jooksul vähendades, et vähendada ärajätunähtude tekkeriski (vt lõigud 4.4 ja 4.8). Kui annuse vähendamise või ravi lõpetamise järgselt ilmnevad talumatuse sümptomid, võib kaaluda ravi jätkamist eelnevalt kasutatud annusega. Seejärel võib arst jätkata annuse vähendamist, kuid see peab toimuma aeglasemalt.

Gene CYP2C19 example

Affects many antidepressants (for example, escitalopram):

Having T instead of C in rs12248560	→	Increased activity of CYP2C19 protein
Having A instead of G in rs4244285	→	Non-functional CYP2C19 protein
Having A instead of G in rs17884712	→	Decreased activity of CYP2C19 protein

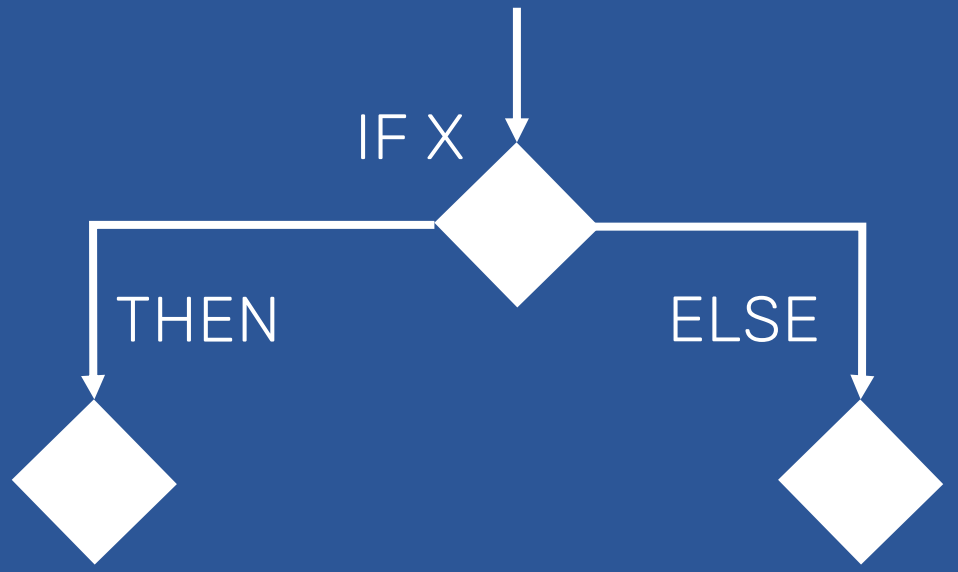
Can lead to toxicity or drug inefficancy

Problem with the vast amount of PGx knowledge

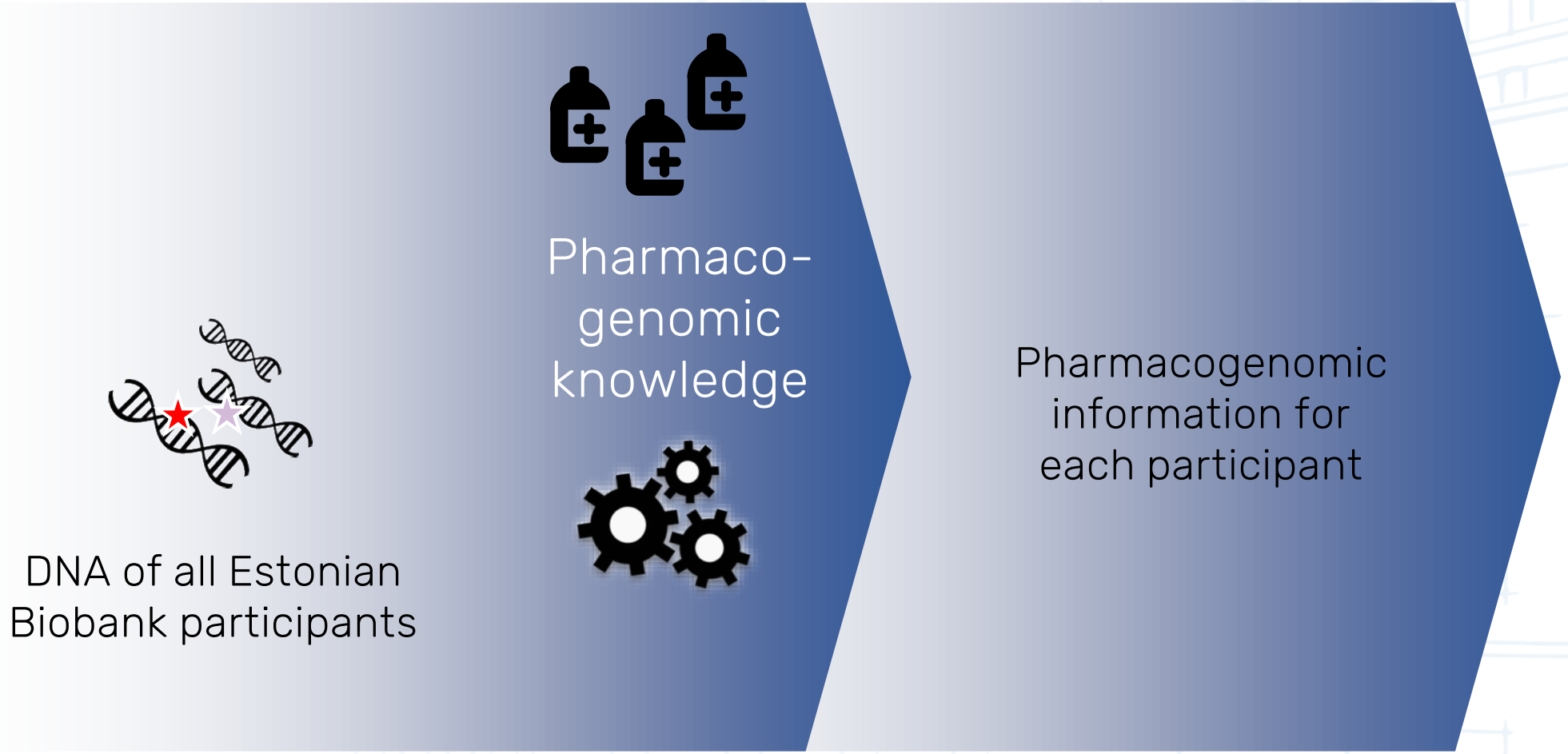
The information is

- Scattered
- With varying evidence
- Often contradicting

Very hard to create
a solid computer algorithm
for PGx



Pipeline for providing pharmacogenomic information of a person

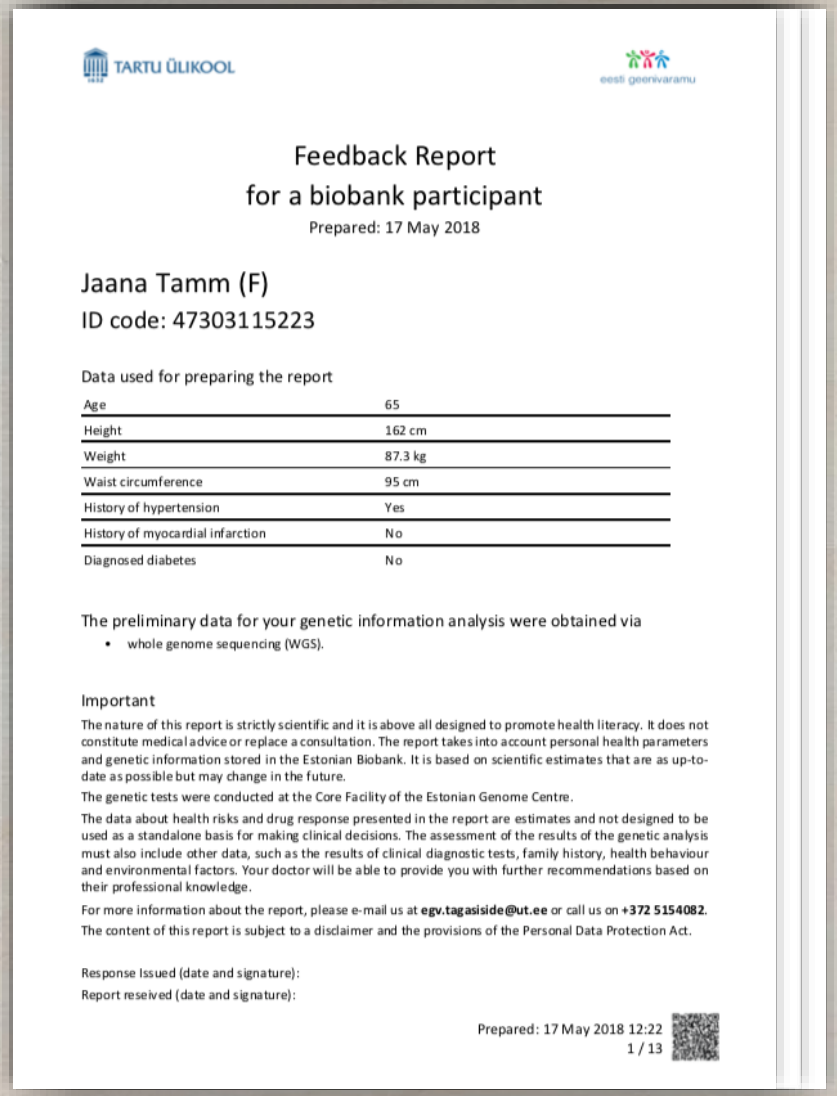


Reisberg S, Krebs K, Lepamets M, Kals M, Mägi R, Metsalu K, Lauschke VM, Vilo J, Milani L. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genetics in Medicine. 2019 Jun;21(6):1345-54. 2019

99.8%

of biobank participants need a dosage adjustment for at least one of the medications

Reisberg S, Krebs K, Lepamets M, Kals M, Mägi R, Metsalu K, Lauschke VM, Vilo J, Milani L. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genetics in Medicine. 2019 Jun;21(6):1345-54.



TARTU ÜLIKOOL

eesti geenivaramu

Feedback Report
for a biobank participant
Prepared: 17 May 2018

Jaana Tamm (F)
ID code: 47303115223

Data used for preparing the report

Age	65
Height	162 cm
Weight	87.3 kg
Waist circumference	95 cm
History of hypertension	Yes
History of myocardial infarction	No
Diagnosed diabetes	No

The preliminary data for your genetic information analysis were obtained via

- whole genome sequencing (WGS).

Important

The nature of this report is strictly scientific and it is above all designed to promote health literacy. It does not constitute medical advice or replace a consultation. The report takes into account personal health parameters and genetic information stored in the Estonian Biobank. It is based on scientific estimates that are as up-to-date as possible but may change in the future.

The genetic tests were conducted at the Core Facility of the Estonian Genome Centre.

The data about health risks and drug response presented in the report are estimates and not designed to be used as a standalone basis for making clinical decisions. The assessment of the results of the genetic analysis must also include other data, such as the results of clinical diagnostic tests, family history, health behaviour and environmental factors. Your doctor will be able to provide you with further recommendations based on their professional knowledge.

For more information about the report, please e-mail us at **egv.tagasiside@ut.ee** or call us on **+372 5154082**.

The content of this report is subject to a disclaimer and the provisions of the Personal Data Protection Act.

Response Issued (date and signature):

Report reseived (date and signature):

Prepared: 17 May 2018 12:22
1 / 13

A project of returning individual genomic results to

3000

research participants

TARTU ÜLIKOOL

eesti geenivaramu

Feedback Report
for a biobank participant

Prepared: 17 May 2018

Jaana Tamm (F)
ID code: 47303115223

Data used for preparing the report

Age	65
Height	162 cm
Weight	87.3 kg
Waist circumference	95 cm
History of hypertension	Yes
History of myocardial infarction	No
Diagnosed diabetes	No

The preliminary data for your genetic information analysis were obtained via

- whole genome sequencing (WGS).

Important

The nature of this report is strictly scientific and it is above all designed to promote health literacy. It does not constitute medical advice or replace a consultation. The report takes into account personal health parameters and genetic information stored in the Estonian Biobank. It is based on scientific estimates that are as up-to-date as possible but may change in the future.

The genetic tests were conducted at the Core Facility of the Estonian Genome Centre.

The data about health risks and drug response presented in the report are estimates and not designed to be used as a standalone basis for making clinical decisions. The assessment of the results of the genetic analysis must also include other data, such as the results of clinical diagnostic tests, family history, health behaviour and environmental factors. Your doctor will be able to provide you with further recommendations based on their professional knowledge.

For more information about the report, please e-mail us at **egv.tagasiside@ut.ee** or call us on **+372 5154082**.

The content of this report is subject to a disclaimer and the provisions of the Personal Data Protection Act.

Response Issued (date and signature):

Report reseived (date and signature):

TARTU ÜLIKOOL

eesti geenivaramu

Pharmacogenetics

Variations in genes that are necessary for drugs to take effect are one of the reasons why people have different reactions to medicines. Over the years, a number of genetic markers predictive of drug response have been identified. This test was used to determine the genetic markers of eight genes, which influence your response to 24 active substances used in medicines.

The following instructions can be given on the basis of the genetic variants tested:

Gene	Genotype	Assessment	Recommendations	Affected active substances
CYP2C19	*2/*17	Average drug metabolism	!	Escitalopram, citalopram, clopidogrel, sertaline, voriconazole, esomeprazole, lansoprazole, pantoprazole, omeprazole, clomipramine, amitriptyline
CYP2C9	*1/*1	Average drug metabolism	+	Phenytoin
CYP2C9; VKORC1	rs9923231 (GG)	Usual recommended dose	+	Warfarin
CYP3A5	*3/*3	Slow drug metabolism, Regular pattern	+	Tacrolimus
DPYD	*1/*1	Average drug metabolism	+	Capecitabine, fluoroacil
IFNL3	rs12979860 (TT)	Reduced drug efficacy	!	Peginterferon alpha-2b, ribavirin
SLCO1B1	rs4149056 (TT)	Average risk of myopathy	+	Simvastatin
TPMT	*1/*1	Average drug metabolism	+	Tioguanine, mercaptopurine, azathioprine

+ - Use a normal dose. ! - Use with caution; the dose may need to be adjusted. ! - Use with extreme caution; there is a risk of side effects.

Information for doctor

Gene	Genotype	Active substance	Influence of genotype	Recommendations
CYP2C19	*2/*17	Esomeprazole, lansoprazole, pantoprazole, omeprazole	Average drug metabolism.	Start treatment with normal dose.
CYP2C19	*2/*17	Escitalopram, citalopram	Slower than average drug metabolism	Start treatment with normal dose.
CYP2C19	*2/*17	Clomipramine, amitriptyline	Slower than average breakdown of tricyclic amines.	Start treatment with normal dose.
CYP2C19	*2/*17	Clopidogrel	Reduced inhibition of thrombocyte aggregation; increase in the residue of thrombocyte aggregation, resulting in an increased risk of cardiovascular side effects.	Start treatment with an alternative drug, e.g. those that contain prasugrel or ticagrelor as active substances.

HOW DID PARTICIPANTS FEEL?

Participants perceived the information that they received as **valuable**, even when the reporting of high risks initially caused worry.

Main challenge: **scalability**

Leitsalu L, Reigo A, Palover M, Nikopensius T, Läll K, Krebs K, Reisberg S, Mägi R, Kals M, Alavere H, Nõukas M, Kolk A, Normet I, Tammesoo ML, Käärik E, Puusepp M, Metsalu K, Allik A, Milani L, Fischer K, Tõnisson N, Metspalu A. Lessons learned during the process of reporting individual genomic results to participants of a population-based biobank. Eur J Hum Genet. 2023 Sep;31(9):1048-1056. doi: 10.1038/s41431-022-01196-6. Epub 2022 Oct 3. PMID: 36192438; PMCID: PMC10474261.

Portal for Estonian Biobank participants

Includes PGx information for 200,000 participants

Example

Estsitalopraam Escitalopram (antidepressant)

Antidepressant, selektiivsed serotoniini tagasihaarde inhibiitorid (SSRI)

Toimeainet sisaldavad ravimid on näiteks

Cipralex, Ciraset, Elicea, Elicea Q-Tab, Escitalopram Accord, Escitalopram Actavis, Escitalopram Grindeks, Escitalopram Orion, Escitalopram Teva, Eslorex, Estan

Teaduskirjanduses soovitatakse Recommendations

Tsitalopraami ja estsitalopraami kiirenenud metabolism vähem aktiivseteks ühenditeks võrreldes CYP2C19 tavapäraste metaboliseerijatega. Madalamad plasmakontsentratsioonid vähendavad kliinilise kasu tõenäosust. Alustada ravi tavapärase algannusega. Kui patsient ei reageeri piisavalt soovitatavale säilitusannusele, kaaluda tiitrimist kõrgemate säilitusannusteni või kliiniliselt sobiva alternatiivse antidepressandi kasutamist, mida ei metaboliseeri peamiselt CYP2C19.

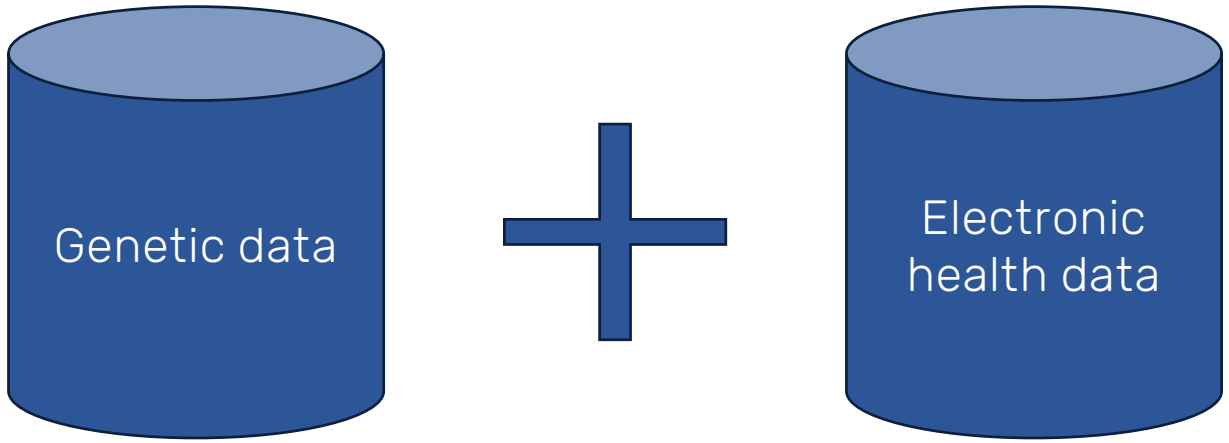
Geenide mõju toimeaine ainevahetusele

Geen CYP2C19 - kiire metaboliseerija CYP2C19 rapid metabolizer

Bousman CA, Stevenson JM, Ramsey LB, Sangkuhl K, Hicks JK, Strawn JR, Singh AB, Ruaño G, Mueller DJ, Tsermpini EE, Brown JT, Bell GC, Leeder JS, Gaedigk A, Scott SA, Klein TE, Caudle KE, Bishop JR. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6, CYP2C19, CYP2B6, SLC6A4, and HTR2A Genotypes and Serotonin Reuptake Inhibitor Antidepressants. Clin Pharmacol Ther. 2023 Jul;114(1):51-68. doi: 10.1002/cpt.2903. Epub 2023 May 30. PMID: 37032427; PMCID: PMC10564324.

Adverse drug reactions in real world

Estonian Biobank has health data also



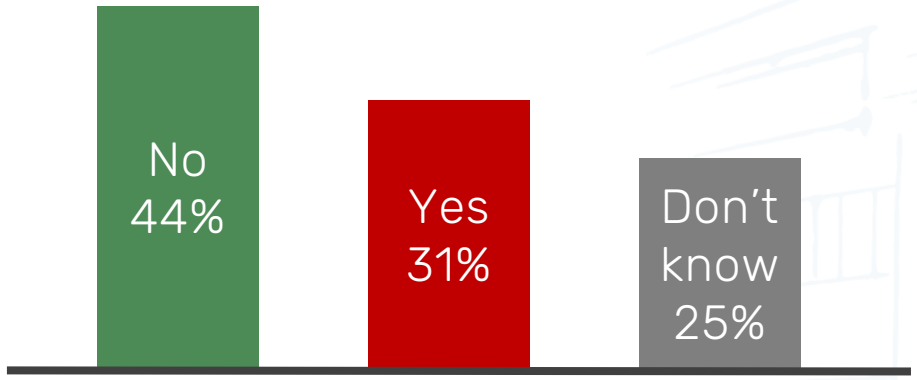
Do we see a large number of adverse drug reactions reported in routinely collected health records?

No, they are heavily underreported

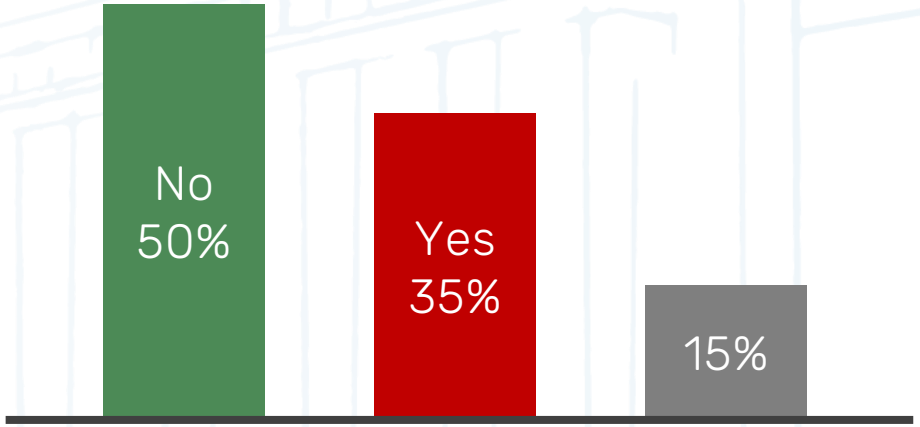
(and mostly in free-text parts)

Survey of adverse drug reactions

Approximately 50K Estonian Biobank participants replied the survey



Ever had DRUG side effects?



Ever had VACCINE side effects?

Krebs et al. to be published soon

Our DNA affects our lifelong medication-use patterns

DNA of the individual

Affects

How individual start, switch and discontinue lifelong medications

nature medicine

Article

https://doi.org/10.1038/s41591-022-02122-5

Genetic predictors of lifelong medication-use patterns in cardiometabolic diseases

Received: 19 October 2021

Accepted: 8 November 2022

Published online: 18 January 2023

Check for updates

Tuomo Kiiskinen[1,2,3,17], Pyry Helkkula[1,17], Kristi Krebs[4,17], Juha Karjalainen[1,3,5], Elmo Saarentaus[1,6], Nina Mars[1], Arto Lehisto[1], Wei Zhou[3], Mattia Cordioli[1], Sakari Jukarainen[1], Joel T. Rämö[1], Juha Mehtonen[1], Kumar Veerapen[3], Markus Räsänen[7], Sanni Ruotsalainen[1], Mutaamba Maasha[3], FinnGen*, Teemu Niiranen[2,8,9], Tiinamaija Tuomi[1,10,11,12,13], Veikko Salomaa[2], Mitja Kurki[1,3,5], Matti Pirinen[1,14,15], Aarno Palotie[1,5,16], Mark Daly[1,5,16], Andrea Ganna[1,3], Aki S. Havulinna[1,2], Lili Milani[4] & Samuli Ripatti[1,3,11]

Little is known about the genetic determinants of medication use in preventing cardiometabolic diseases. Using the Finnish nationwide drug purchase registry with follow-up since 1995, we performed genome-wide association analyses of longitudinal patterns of medication use in hyperlipidemia, hypertension and type 2 diabetes in up to 193,933 individuals (55% women) in the FinnGen study. In meta-analyses of up to 567,671 individuals combining FinnGen with the Estonian Biobank and the UK Biobank, we discovered 333 independent loci ($P < 5 \times 10^{-9}$) associated with medication use. Fine-mapping revealed 494 95% credible sets associated with the total number of medication purchases, changes in medication combinations or treatment discontinuation, including 46 credible sets in 40 loci not associated with the underlying treatment targets. The polygenic risk scores (PRS) for cardiometabolic risk factors were strongly associated with the medication-use behavior. A medication-use enhanced multitrait PRS for coronary artery disease matched the performance of a risk factor-based multitrait coronary artery disease PRS in an independent sample (UK Biobank, $n = 343{,}676$). In summary, we demonstrate medication-based strategies for identifying cardiometabolic risk loci and provide genome-wide tools for preventing cardiovascular diseases.

Cardiovascular disease (CVD) is the leading cause of excess mortality in the developed countries[1], and although approximately half of the variability in cardiometabolic diseases is heritable[2], most related harm is preventable[3,4]. Pharmacotherapies targeting cardiometabolic risk factors—type 2 diabetes (T2D), hyperlipidemia and hypertension—remain at the core of CVD prevention[5,6].

Challenges in pharmacological prevention of CVD involve identifying patients in need of therapy, setting the targets of the treatment and selecting therapies of adequate efficacy and acceptable risk profiles. In addition to socioeconomic factors, both the set and dose of medicines that patients start their treatment with and continue to use depends on factors such as cardiovascular risk profiles, disease etiology, drug responsiveness and adverse effects[5,6]. Abandoning or inadequately adhering to therapies worsens outcomes[7–10]. With limited tools to predict treatment suboptimality, pharmacotherapy is traditionally optimized in a reactive trial-and-error manner when patients experience side effects, miss their treatment targets or experience events such as myocardial infarction or stroke[5,6]. Real-world data from electronic health records and registries provide massive datasets with sufficient statistical power to explore long-term medication use.

A full list of affiliations appears at the end of the paper. e-mail: samuli.ripatti@helsinki.fi

Cascade effect of PGx

Schizophrenia

Is treated by

Antipsychotic drugs

Cause more

Metabolic syndrome

DNA of the individual

May cause

Bad adherence of antipsychotic drugs

Severe forms of metabolic syndrome

Genetic predisposition and antipsychotic treatment effect on metabolic syndrome in schizophrenia: a ten-year follow-up study using the Estonian Biobank

Maris Alver,[a,*] Silva Kasela,[a] Liina Haring,[b,c] Laura Birgit Luitva,[a] Estonian Biobank Research Team,[a] Health Informatics Research Team,[d] Krista Fischer,[a,e] Märt Möls,[d,e] and Lili Milani[a,**]

[a]Estonian Genome Centre, Institute of Genomics, University of Tartu, Riia 23b, Tartu, 51010, Estonia
[b]Department of Psychiatry, Institute of Clinical Medicine, University of Tartu, Raja 31, Tartu, 50417, Estonia
[c]Psychiatry Clinic of Tartu University Hospital, Raja 31, Tartu, 50417, Estonia
[d]Institute of Computer Science, University of Tartu, Narva mnt 18, Tartu, 51009, Estonia
[e]Institute of Mathematics and Statistics, University of Tartu, Narva mnt 18, Tartu, 51009, Estonia

The Lancet Regional Health - Europe 2024;41: 100914
Published Online 26 April 2024
https://doi.org/10.1016/j.lanepe.2024.100914

Summary

Background Schizophrenia (SCZ) patients exhibit 30% higher prevalence of metabolic syndrome (MetS) compared to the general population with its suboptimal management contributing to increased mortality. Large-scale studies providing real-world evidence of the underlying causes remain limited.

Methods To address this gap, we used real-world health data from the Estonian Biobank, spanning a median follow-up of ten years, to investigate the impact of genetic predisposition and antipsychotic treatment on the development of MetS in SCZ patients. Specifically, we set out to characterize antipsychotic treatment patterns, genetic predisposition of MetS traits, MetS prognosis, and body mass index (BMI) trajectories, comparing SCZ cases (n = 677) to age- and sex-matched controls (n = 2708).

Findings SCZ cases exhibited higher genetic predisposition to SCZ (OR = 1.75, 95% CI 1.58–1.94), but lower polygenic burden for increased BMI (OR = 0.88, 95% CI 0.88–0.96) and C-reactive protein (OR = 0.88, 95% CI 0.81–0.97) compared to controls. While SCZ cases showed worse prognosis of MetS (HR 1.95, 95% CI 1.54–2.46), higher antipsychotic adherence within the first treatment year was associated with reduced long-term MetS incidence. Linear mixed modelling, incorporating multiple BMI timepoints, underscored the significant contribution of both, antipsychotic medication, and genetic predisposition to higher BMI, driving the substantially upward trajectory of BMI in SCZ cases.

Interpretation These findings contribute to refining clinical risk prediction and prevention strategies for MetS among SCZ patients and emphasize the significance of incorporating genetic information, long-term patient tracking, and employing diverse perspectives when analyzing real-world health data.

Funding EU Horizon 2020, Swedish Research Council, Estonian Research Council, Estonian Ministry of Education and Research, University of Tartu.

Keywords: Schizophrenia; Metabolic syndrome; Antipsychotics; Treatment; Real world data; Genetics; Polygenic risk scores; Body mass index

Introduction

Schizophrenia (SCZ) is a neurodevelopmental disorder affected by genetic and environmental risk factors with symptoms manifesting in early adulthood.[1] SCZ patients have a 15-year shorter lifespan[2] and exhibit 30% higher prevalence of metabolic syndrome (MetS) compared to the general population.[3] As suboptimal management of the clustered symptoms of MetS (dyslipidaemia, abdominal obesity, hypertension, and hyperglycaemia) has a substantial impact on mortality

*Corresponding author.
**Corresponding author.
E-mail addresses: maris.alver@ut.ee (M. Alver), lili.milani@ut.ee (L. Milani).

Implementation of personalised medicine in Estonia

2019–2023

The project „Implementation of personalised medicine in Estonia“ created the capability to employ national services that use genetic data more extensively.

2025+

Implementation of first PGx services based on the new IT infrastructure

Investing in your future

European Union
European Structural and Investment Funds

National Institute for Health Development

University of Tartu

TEHIK
Health and Welfare Information Systems Centre

Estonian Health Insurance Fund

Republic of Estonia
Ministry of Social Affairs

SUMMARY

1. Pharmacogenomics is
much more important
than we previously thought

2. Structural genetic variants
play a larger role
than we previously thought

3. The inability
to reliably detect structural variants (yet)
prevents us
from using PGx services on a large scale

Acknowledgements

Lili Milani

Kristi Krebs

Maris Alver

Tuuli Reisberg

This work was supported by the Estonian Research Council grant (PRG18449) and co-funded by the European Union through the European Regional Development Fund (Project No. 2021-2027.1.01.24-0444)

Co-funded by the European Union

Investing in your future

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

imi innovative medicines initiative

efpia

Thank you!

sulev.reisberg@ut.ee

ut.ee

info@ut.ee

tartuylikool
tartuuniversity

unitartu
unitartuscience
unitartutiksu