



# Estonian Genome Project

**7 October 2002**

**ECCB 2002**

**Andres Metspalu**

Dept. of Biotechnology, University of Tartu,

EGL. Inc



eesti geenivaramu



# Top 4 causes of death in US in 1994

- 1. Heart 734 000**
- 2. Cancer 536 000**
- 3. Stroke 154 000**
- 4. Adverse Drug 106 000**

**Reactions (ADR)**



eesti geenivaramu

## Estimated Number of Hospital Patients in 1994 with Adverse Drug Reactions (ADRs)

- 4,986,000 ADRs of all severities
- 2,216,000 serious ADRs
- 106,000 fatal ADRs



eesti geenivaramu

# Getting the right medicine to the right patient

Patients who respond without a serious side effect



Patients who respond but experienced a serious side effect



ATTGCATGCCAGTAGG

SNP Profile

TATGATTGCCGC TAGG



Diagnostic

Right Medicine

Right Patient

eesti geenivaramu

**GP is thinking hard!**



eesti geenivaramu



# What is Estonian Genome Project (EGP)?

- Goal is to collect 1 000 000 individuals into health and genetic database.
  - Better healthcare delivery – short term benefit
- Application of DNA based diagnosis and personalized treatment methods to achieve better service at lower cost
  - Better healthcare availability – long term benefit
- Development of local IT-based gene technology and medical sector infrastructure
  - Economic benefits (jobs, investments, education, etc.) – short and long term

**Here I stand and....**



# Legal background

## Human Genome Research Act (HGRA)

- Approved by Government on Aug 8, 2000
- Passed by Parliament on Dec 13, 2000 (42 yes, 3 no)
- Enforce since January 08, 2001

## HGRA regulates

- scientific research on human genetics
- the establishment and maintenance of the GeneBank
- the use of genetic information (informed consent, based on principle of open consent)

## HGRA protects

- the confidentiality of gene donor
- the public from the misuse of genetic information
- the gene donor from genetic discrimination



# General Principles of HGRA

- voluntary participation in genetic research studies
- confidentiality of gene donor
- gene donor's right to apply for destruction of his/her identification data
- gene donor's right to know or not to know his/her genetic data
- non-discrimination by employers and insurers
- the institution of the Ethics Committee and SAB
- EGP Foundation

# Estonian Genome Project Foundation

- Established by the Government of Estonia in 2001
- Non-profit private entity
- Legal owner of database – Chief Processor
- Supervisory Board - 9 members of the Parliament and Government of Estonia and Academy of Sciences
- Ethics Committee and Scientific Advisory Board
- 28 employees during the Pilot Project

# Scientific research

- Free access for Estonian Scientific Institutes (and together with Estonian academic researchers)
- Approved by Ethics Committee of EGPF and IRB
- Only anonymous records (not less than 5)
- IP rights will be shared with EGPF

# Timeline

## I **Pilot Project** – 3 counties, 10 000 gene donors, ~100 data collectors GPs (Primary Care Physicians)

- Preparations 2001-2002

Laboratory, depository of tissue samples, coding center, IT solutions, logistics and transport, security systems, data collectors' training, informing the public

- Carrying out in 6 months started in September 2002  
Testing of the elaborated model, the quality and the security of the processes

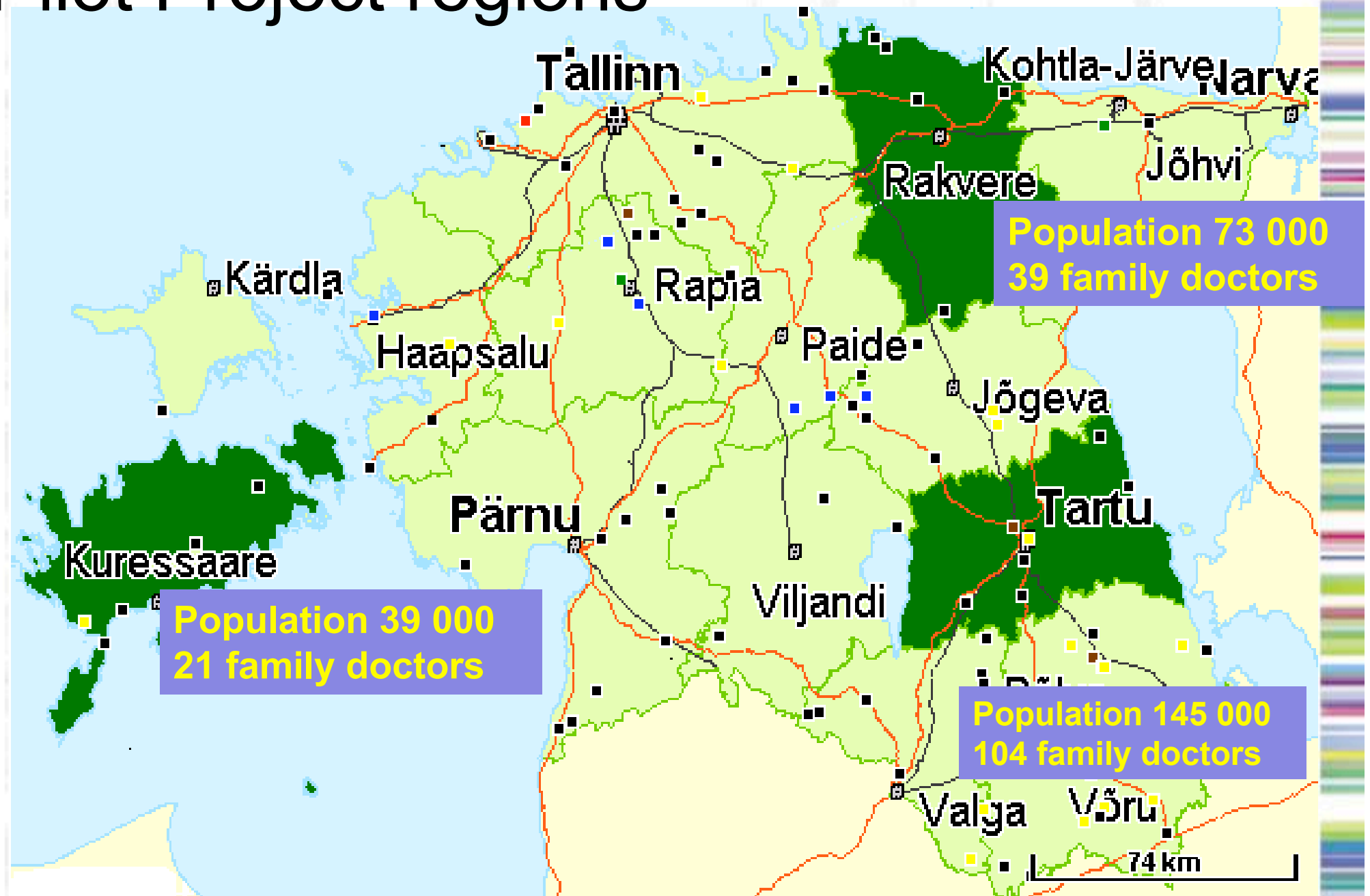
## II **Main Project** 2003-2007 - ~3/4 of the population of Estonia, 800 data collectors

# Some facts about Estonia

- Population on
- 1 January 2000 - 1 439 197
- Males - 669 583
- Females - 769 614
- Life expectancy / males - 65 years
- Life expectancy / females – 75 years
- GDP per capita \$5900
- GDP growth 6% (for2001)



# Pilot Project regions



# Collection of Gene Bank Data

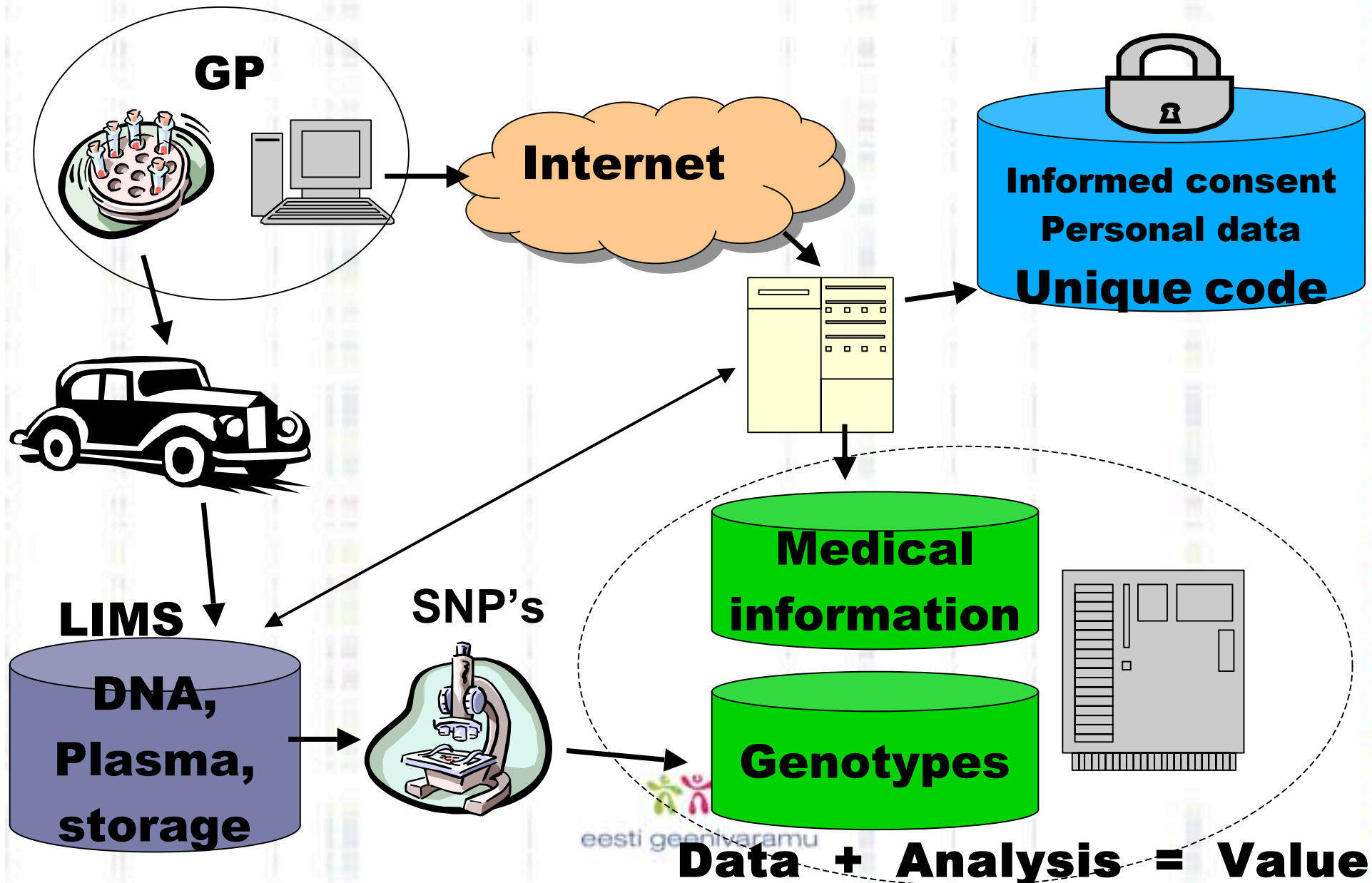
- **Data Collectors (GP)** - training, IT base necessary for the participation, contract, future delivery of information back to patients

Co-operation with the Estonian Association of Family Physicians

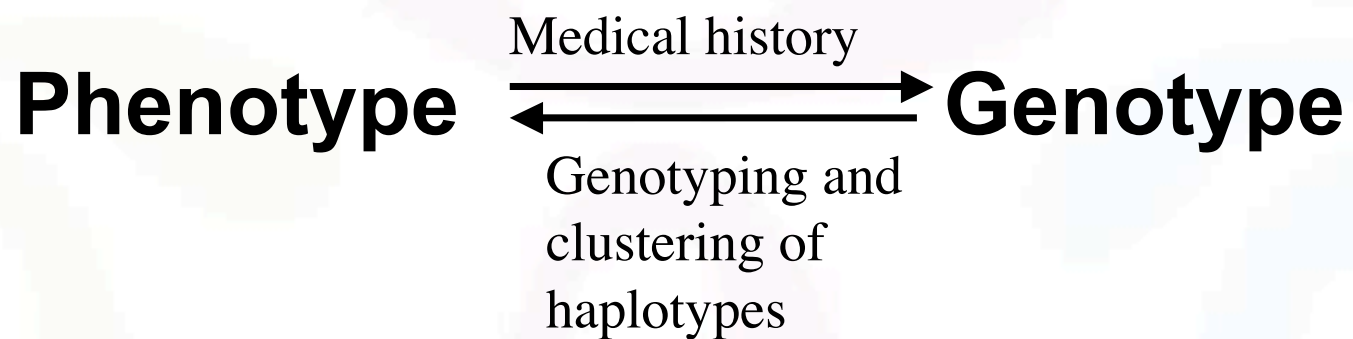
- **Health data questionnaire** – personal data, environment, diet, workplace, lifestyle, clinical data as in ICD-10, genealogical data (3 generations)
- **Data monitoring**
- **Logistics and transport** – transport of 10 000 blood samples to EGPF in 36 h
- **Laboratory** – 50 ml of blood, ~2 mg of DNA in 7 straws, also plasma in liquid N2 - 600l containers a' for 20 000 individuals, lab is 10 people, 300 m<sup>2</sup>



# Process of data collection and handling







eesti geenivaramu



## IT tasks for the beginning

- 1. Message passing centre, standardized message formats and encryption procedures that allow to use public Internet infrastructure for passing sensitive information related to the project**
- 2. Questionnaire tool used by GP's for patient interviews and a tool for developing questionnaires, user interfaces and procedures for questionnaire tool, represented in XML format**
- 3. Coding and decoding centre for *deidentification* and *reidentification* of personal identities, and storage of informed consent forms**
- 4. General information system that handles data and sample collection logistics, and a LIMS system for the central laboratory information management**



eesti geenivaramu

# Clear Legal-Ethical Structure Facilitates Commercialization

**Estonian population**



**EGP Foundation**

(non-profit entity)



**EGeen Ltd.**

For-profit company

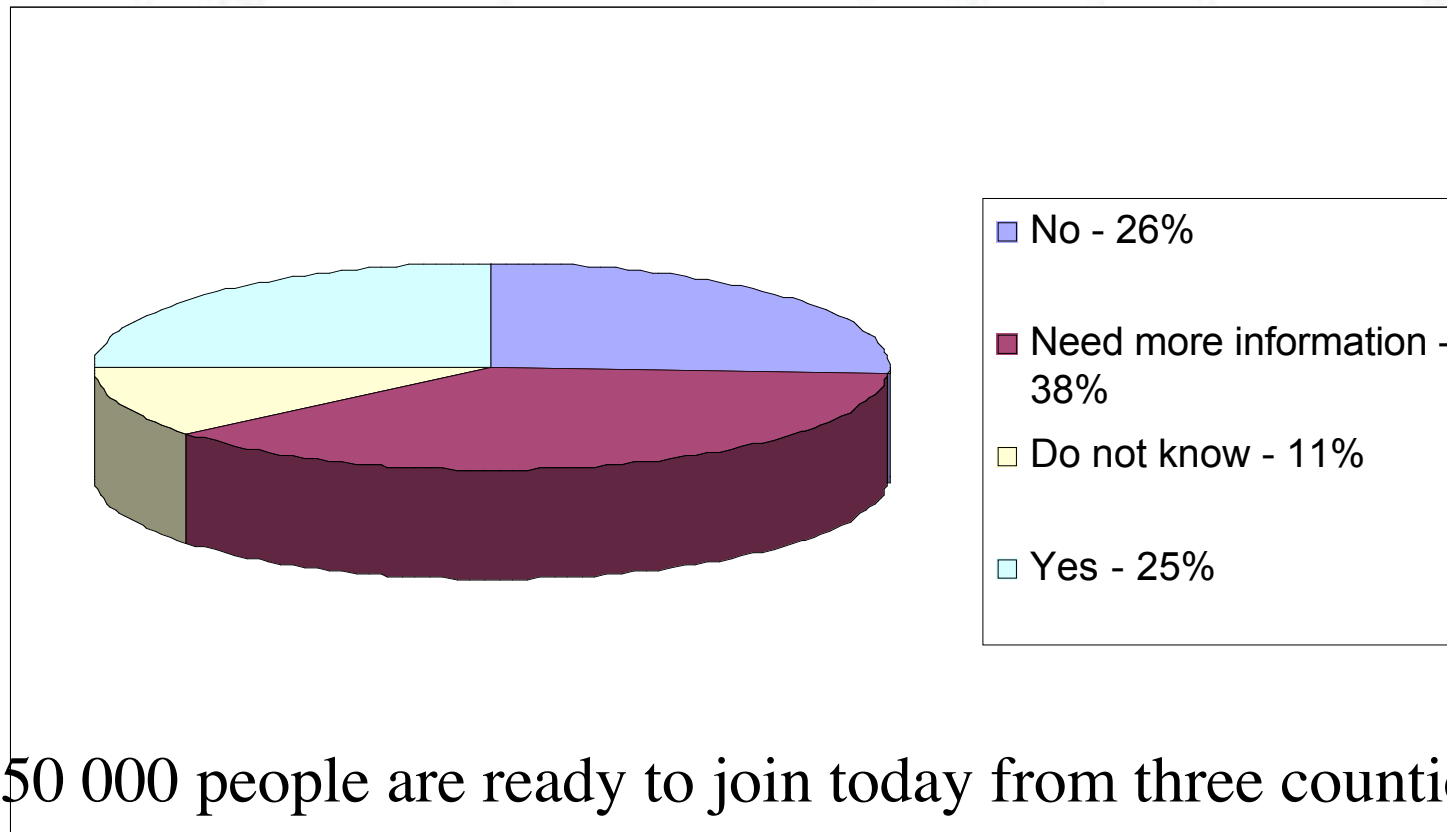
- Provides health and genotype data
- Gains access to individual records (if desired)
- Reserves right to destroy personal data
- Performs data collection and storage
- Legal owner of the database
- Grants exclusive license for all commercial activity via EGeen Ltd.(EST) To EGeen, Inc. (USA)
- Retains a minority stake in EGeen, Inc. (earnings go for earmarked public good)
- Monitors database use (privacy shelter)
- Privately held
- Gets exclusive license for all commercial activity from the EGP Foundation
- DNA genotyping, creation of genetic map
- Sale of database subscription
- Identification of drug and diagnostic targets
- Pharmacogenetics



eesti geenivaramu

# Public opinion in the counties of the pilot project – August 2002

Do you want to become a gene donor today?



# GENETICS

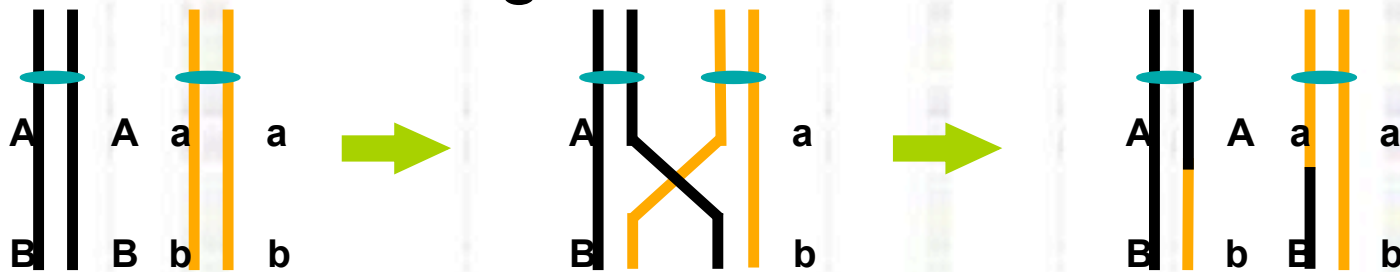


eesti geenivaramu



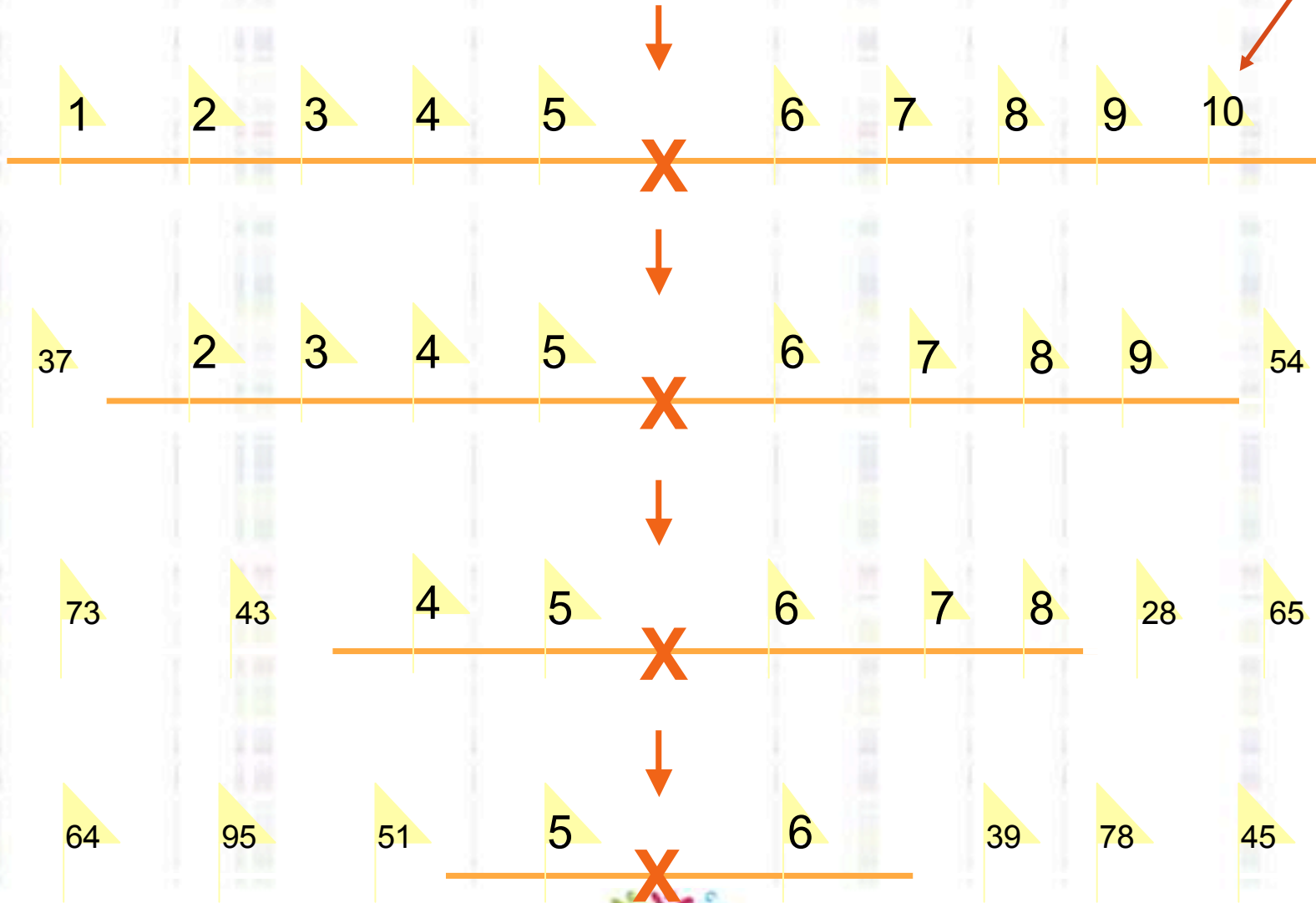
# Recombination

- Recombination is the exchange of genetic information between paired chromosomes
- Recombination occurs primarily during meiosis, and produces gametes with new combinations of genetic information
- Recombination frequency is proportional to the distance between genetic markers



# Linkage disequilibrium

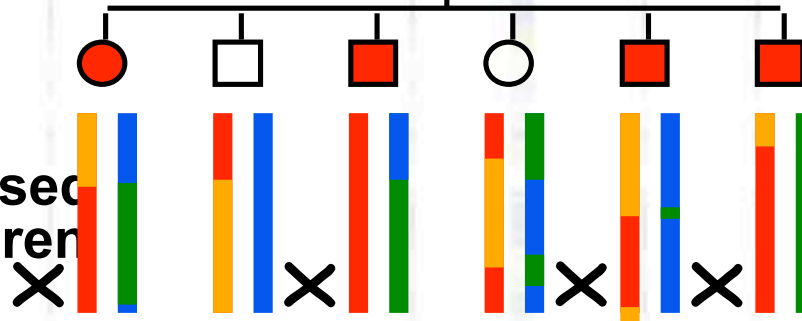
SNP



# Linkage disequilibrium



The mutation is passed to the affected children



After several generations..



.. many apparently unrelated individuals share the mutation and small parts of the ancestral chromosome...



...whereas unaffected individuals did not inherit the region harbouring the mutation.

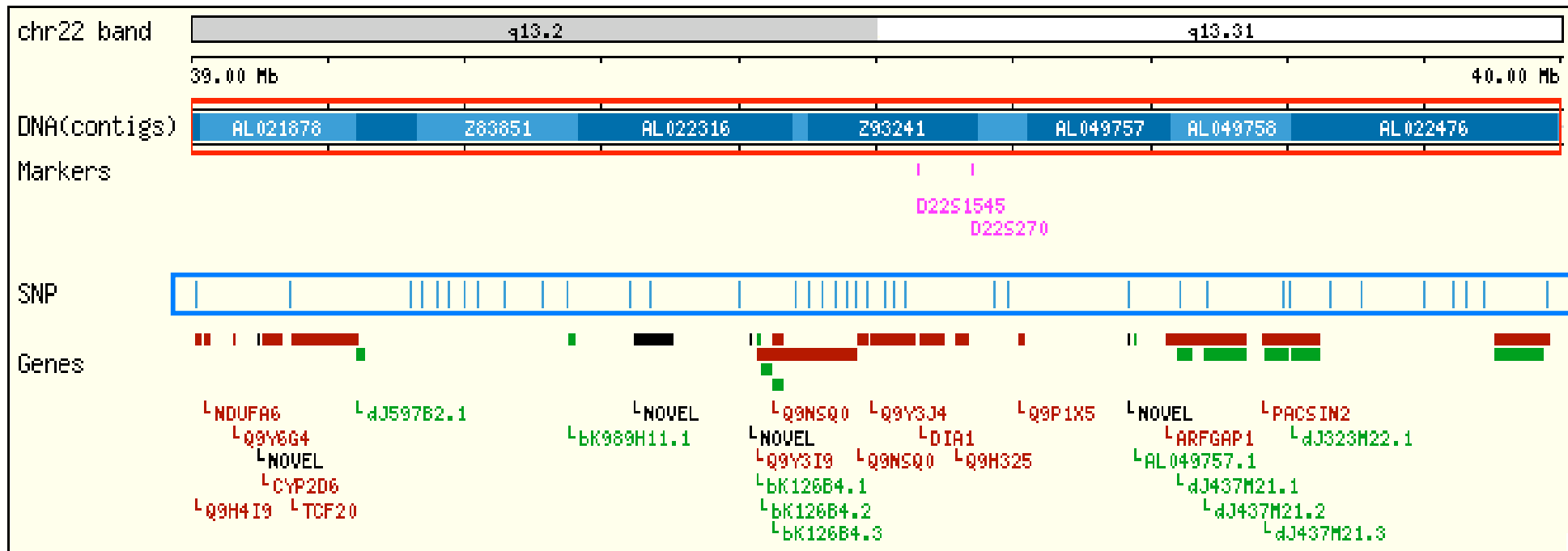


## SNPs characteristics:

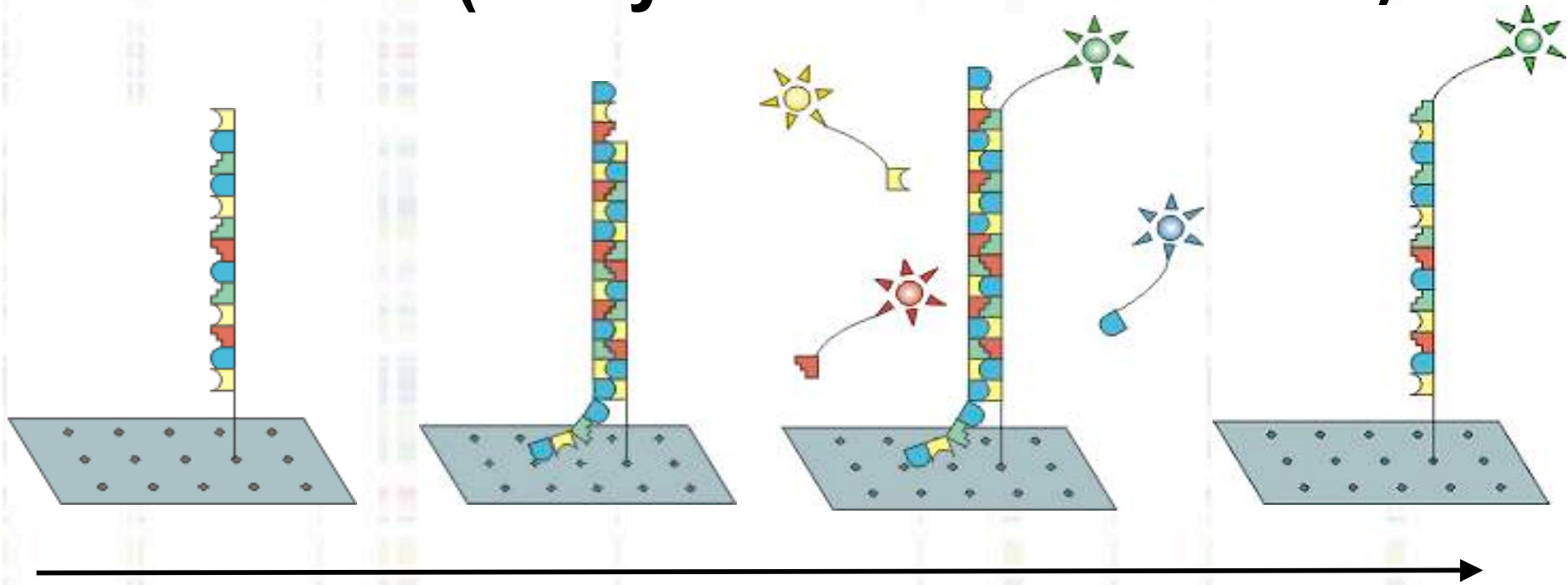
- SNPs are highly abundant; about 90 % from all human DNA polymorphisms are SNPs, occurring with a frequency of more than 1 per 500-1000 base pairs
- SNPs are located over the whole genome (also in genes and their regulatory elements).
- SNPs are genetically stably inherited (mutation rate is about  $10^{-8}$ )
- They are biallelic markers □ can be used for high throughput genotyping methods (DNA chip)

# Chromosome 22

Location of SNPs and genes (between 26-27 Mb)



# SNP Genotyping – APEX (Arrayed Primer Extension)



Up to 60000 25 mer oligos are immobilized via their 5' end onto activated glass surface

Complementary PCR amplified sample DNA is annealed to oligos

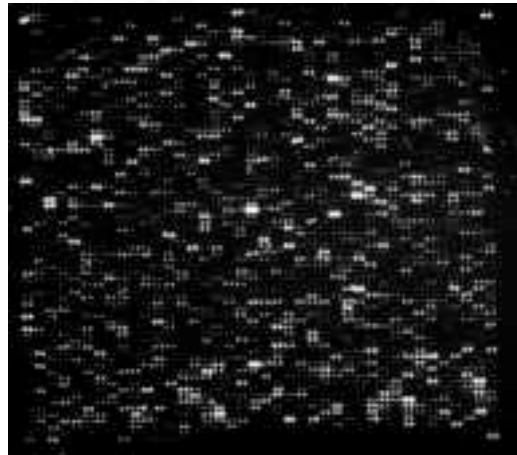
The Thermosequenase extends each oligo on slide by one of four fluorescently labeled bases according to the template DNA

The template DNA fragments and unused dye are washed off. Signals are detected by four-laser imaging system

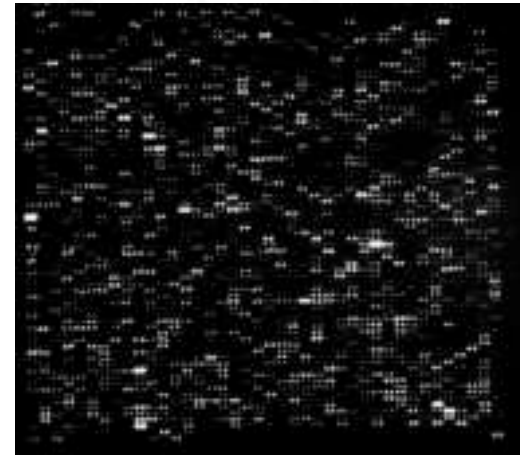
# Chromosome 22 microarray

5600 spots

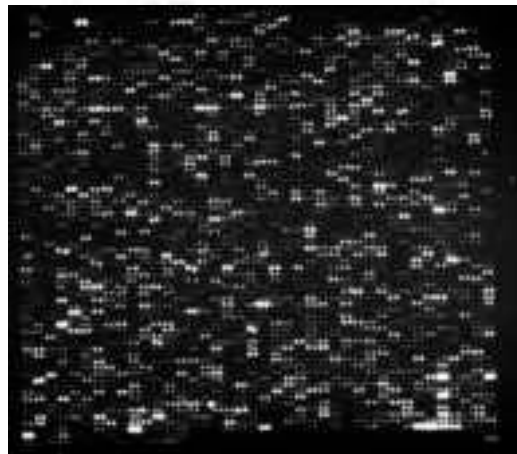
A



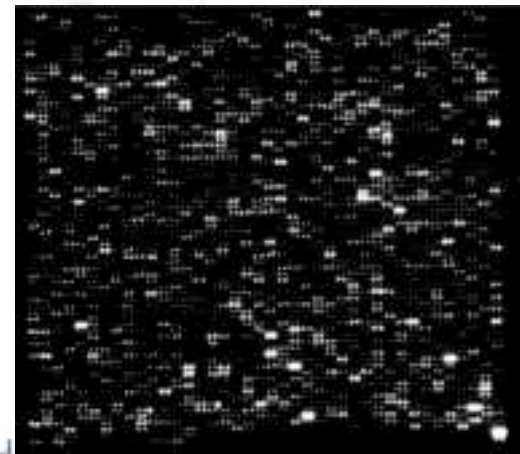
C



G

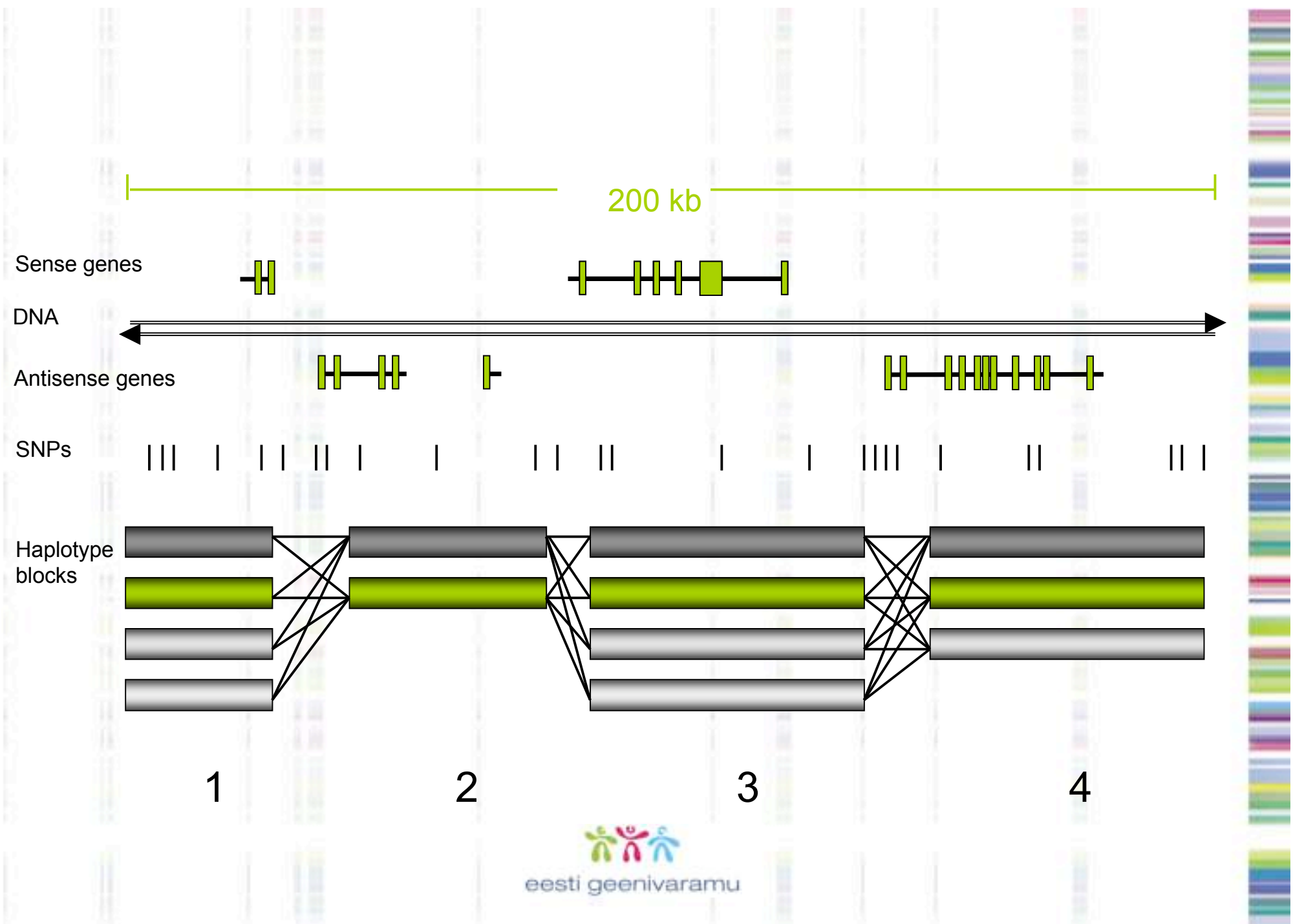


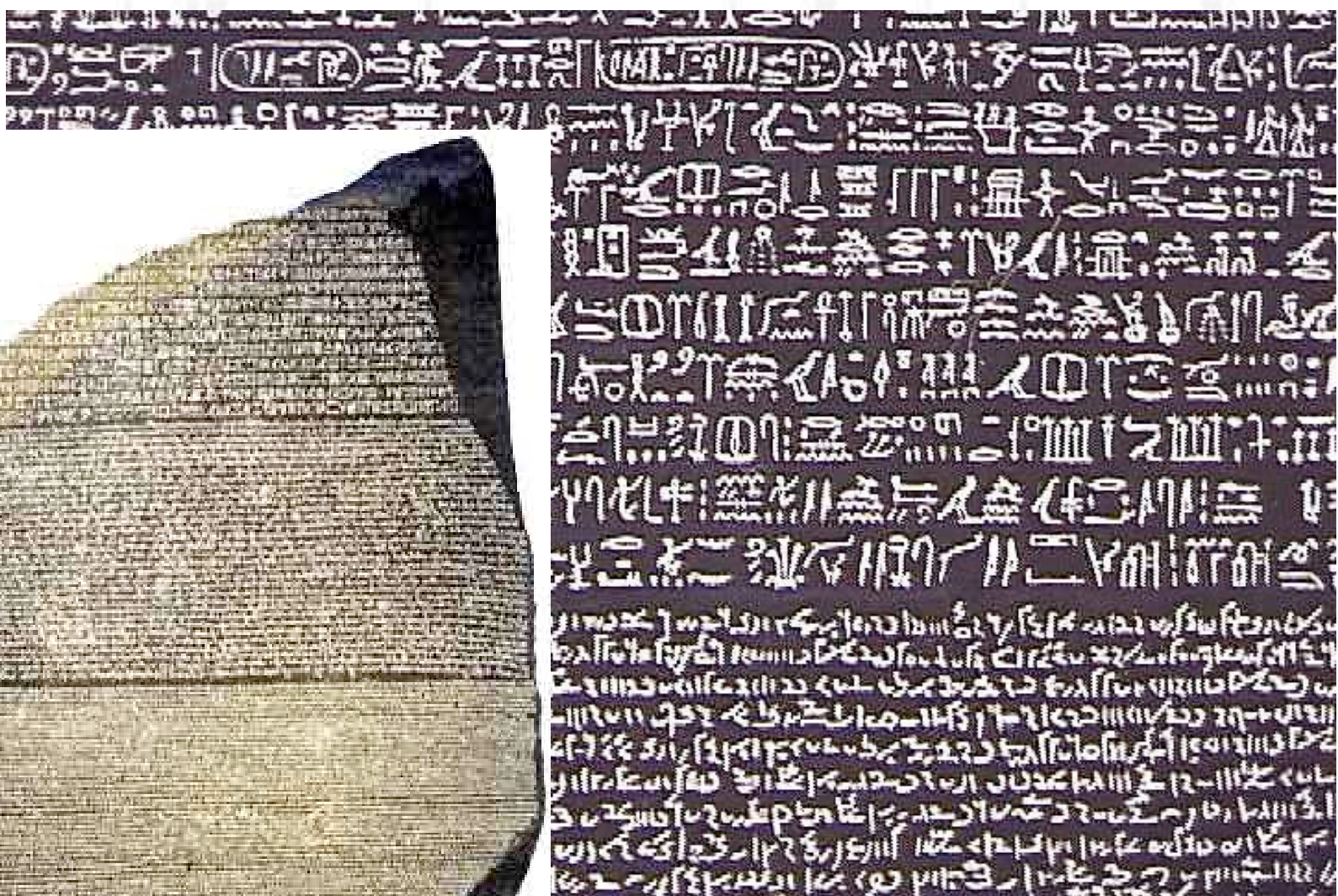
T



esti geenivaramu







eesti geenivaramu





*Dušan  
Toronto, Canada*

