Genetic-epidemiological research methods

Yurii Aulchenko (Erasmus MC, Rotterdam)

Simon Heath (CNG, Paris)

Cornelia van Duijn (Erasmus MC, Rotterdam)

Overview: what, why, and how?

- What is genetic epidemiology?
- Why it is important?
- How the aim is achieved?

- Description of the course

What? ... is genetic epidemiology

- “a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations” Morton NE, 1982
- Study of relation between genetic and phenotypic variation in humans
- Questions to be answered
  - What genes and alleles are involved?
  - What is the model of genetic control?

What? ...is gene-finding

Type 2 diabetes risk factors

- Obesity
- “Western” diet
- Aging
- Presence of specific genetic variants (=alleles)

“Finding genes” = identifying risk alleles present in human populations

What? ...is gene-characterisation

Being able to predict the phenotype based on genotypic and environmental information

- Environmental risk factors
- Allelic spectrum of the locus
- How alleles interact with
  - Other alleles of the same locus
  - Alleles at other loci
  - Environmental factors

Why? ...do we need to find genes

Gene

- Generation of new biological knowledge
- Development of novel treatments

Allele

- Identification of people at high risk
- Personalised medicine
- Early diagnosis and preventive treatment

Why? ...is it important to find human genes

- Why biology can not be studied in mice?
- We do not want to know why *mice* become ill
- Large similarities, but differences in details

- Details are important (TGN1412 story):
  - Supposed to stimulate T-cells
  - Tested in rabbit and monkey
  - Phase 1 clinical trial
    - 6 people administered
    - Multiple organ failure within 24 hours
  - *Suntharalingam et al., NEJM, 2006*

How? ...do we find genes?

When we have sequence

- Which variant is correlated with phenotype?
- Molecular biology: what is the function?

Proxies to sequence

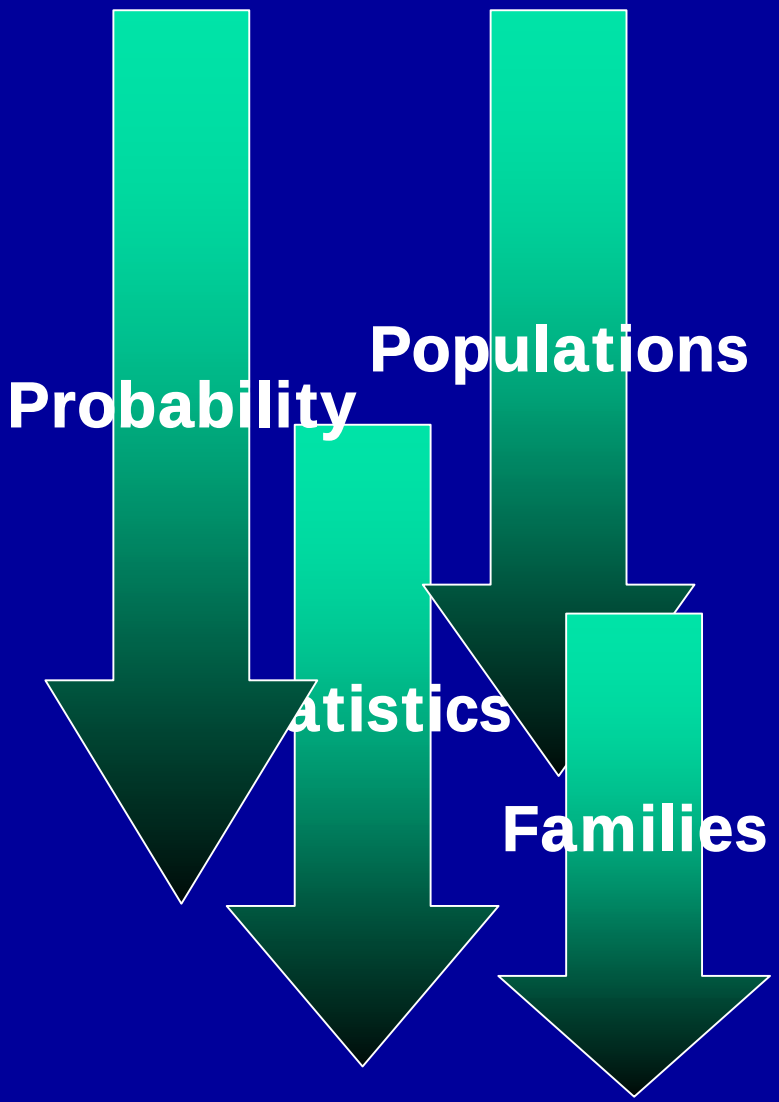
- Familial relations (relatives share sequences)
  - Proportion of variance explained by genes (heritability)
  - Risk in relatives
- Markers (sequence variants with known location)
  - Similarity of markers => similarity of sequence
  - Linkage (100s markers in families): large regions
  - Genome-Wide Association (100s of 100s of markers): small regions

Overview of the course

- Scope: models describing behaviour of genes & quantitative methods for GER
- Introduction to GER designs & methods
  - Week 1: genes in populations, binary traits
  - Week 2: genes in families, quantitative traits
  - Week 3: GER designs
  - Week 4: research assignment
- Details of methods for gene-finding
  - GE05: family-based studies (linkage & association)
  - GE03: population-based studies (association)

Overview of week 1

- Day 1
  - Basic probability, Mendel's Lows, Population genetics, Hardy-Weinberg equilibrium (HWE), departure from HWE
- Day 2
  - Conditional probability, Total probability, Bayes theorem, relative and absolute risk, mutation-selection balance
- Day 3
  - Binomial distribution, genetic drift, Normal & Poisson approximations to Binomial
- Day 4:
  - Hypothesis testing, P-value, χ^2 distribution, Likelihood Ratio test, Power
- Day 5:
  - Genetic risk calculations in families (X-linked, dominant and recessive)

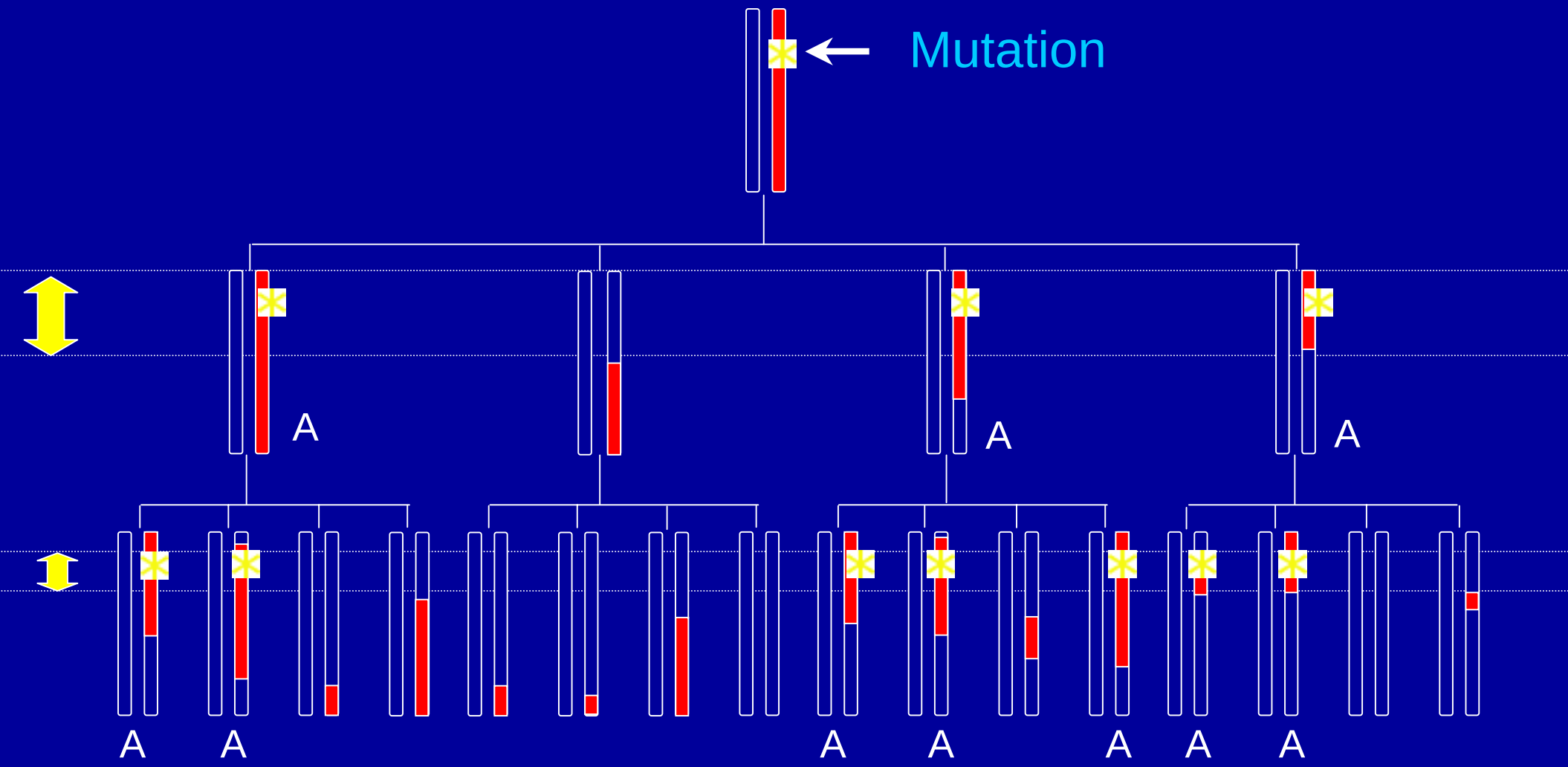


Assignments

- Each group: 20 min. presentation with answers to (not obvious) previous day's questions
- Can check answers with me at 17:00 (office Ee2200)

- Day 1-2
  - Group 1: Morning Quizzes, HWE
- Day 2-3
  - Group 2: Probability formulas, Selection
- Day 3-4
  - Group 3: Drift, Approximations to Binomial
- Day 4-5
  - Group 4: Hypothesis testing, Power

Rationale behind gene-identification



How? ...do we find genes?

Rationale: if a genetic variant is involved,

When we have sequence

- Which variant is correlated with phenotype?
- What is the function?

Proxies to sequence

- Close familial relations
  - Is risk increased in relatives?
  - What proportion of variation is explained by familial clustering?
- Close familial relations + few 100s markers
  - Poor proxy to sequence
  - Good proxy to sequence similarity
  - Large regions identified by linkage
- 100s of 1000s of markers
  - Good proxy to sequence & sequence similarity
  - Genome-Wide Association analysis identifies small regions