Genetic Epidemiology and Personalized Medicine

Prof. Dr. Konstantin Strauch

IBE - Lehrstuhl für Genetische Epidemiologie
Ludwig-Maximilians-Universität

Institut für Genetische Epidemiologie
Helmholtz-Zentrum München

E-Mail: strauch@helmholtz-muenchen.de

What is genetic epidemiology?

Genetic epidemiology

deals with

the identification and characterization of genes

*that are **responsible for diseases***

in humans.

Genetic epi is searching genes ***causing diseases***,

and NOT (the larger set of) genes ***involved in diseases***!

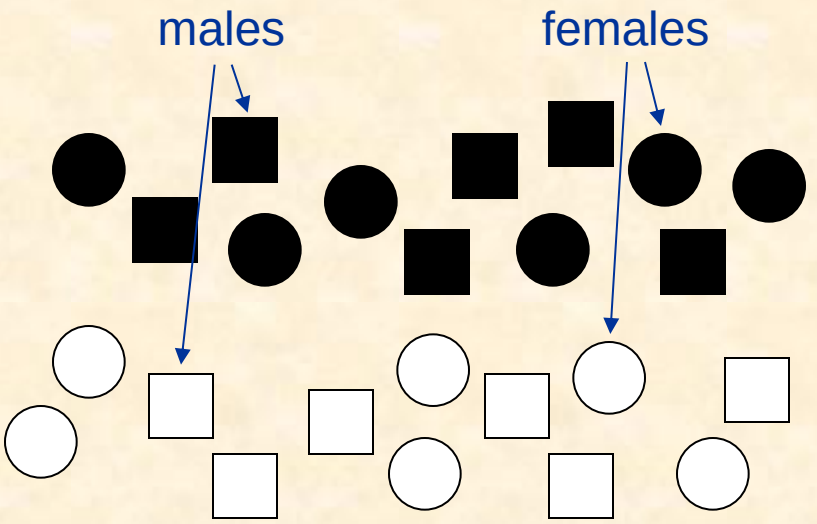
➞ Difference between genetic epidemiology and gene-expression studies / clinical genomics

How can we find disease-causing genes?

We look at…

… affected (sick) persons

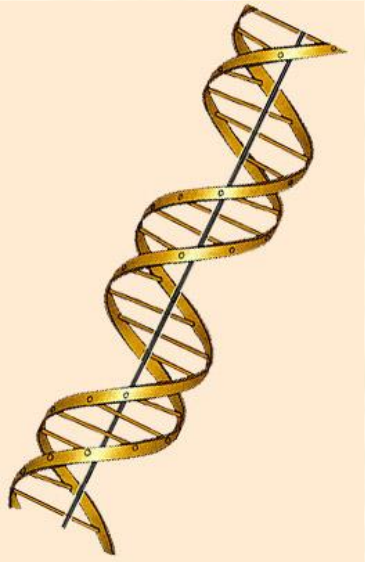
… unaffected (well) persons



How can we identify the genomic region(s) causing the trait?

- ➞ regions that are identical among affected individuals
- ➞ regions that differ between affecteds and unaffecteds

Deoxyribonucleic acid (DNA)



The double helix

James D. Watson & Francis Crick (1953)
Nobel prize 1962

Auf deutsch:
Desoxyribonukleinsäure (DNS)

Many techniques in molecular biology / genetics:

- Polymerase chain reaction (PCR)
- Chip-based genotyping (~1 million markers)
- Large-scale (next-generation) sequencing

Genetics and inheritance

- 23 chromosomes in humans (22 autosomes, 1 sex chro.), an individual carries two copies of each sort (father/mother)
- Genetic locus: refers to certain position on a chromosome
- Alleles: different possible DNA variants at a certain locus
- Genetic markers / polymorphisms:
  positions (loci) in the genome at which the DNA sequence varies between individuals
  - microsatellite markers: short tandem repeats (STRs); a few base pairs (2-4) repeated with varying # of copies
  - single nucleotide polymorphism (SNP): exchange of a single base pair; “letters” of the genetic code: adenine (A), cytosine (C), guanine (G), thymine (T).

Different levels of molecular markers

- DNA: genetics, genomics
- Methylation, chromatin structure: epigenetics, epigenomics
- RNA: amount related to gene expression, transcriptomics
- Proteins, proteomics
- Metabolites, metabolomics

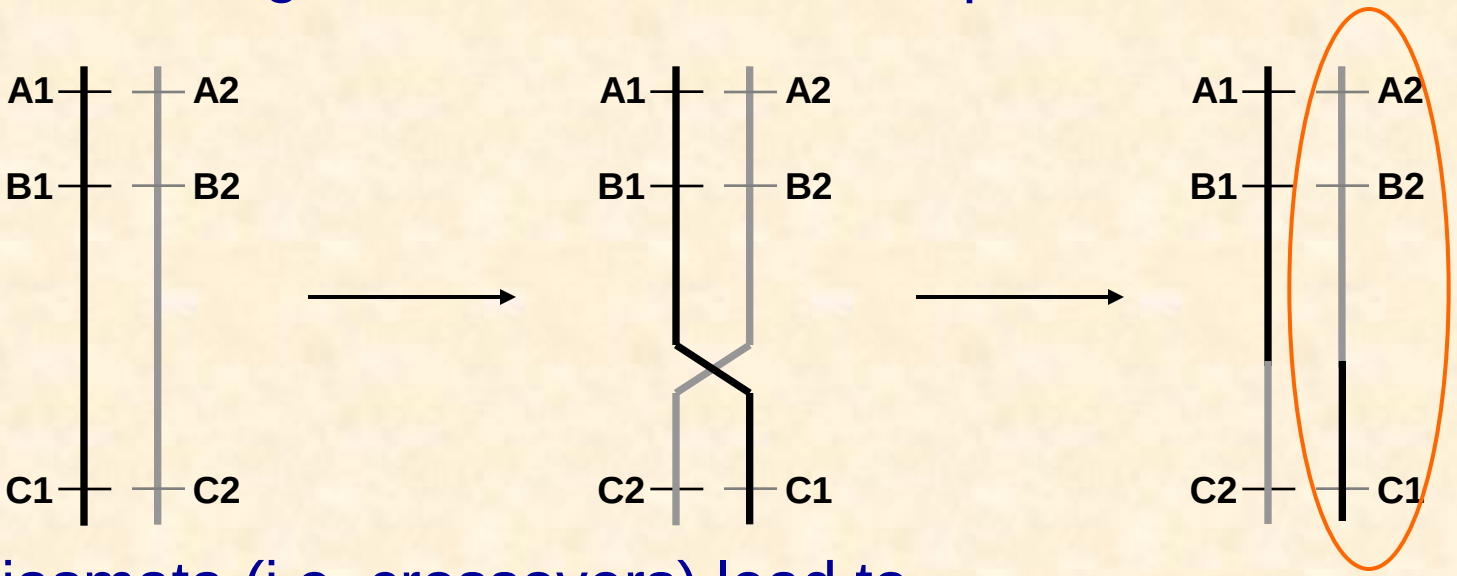
  - altogether often called ‘omics’ data

- Central dogma: DNA makes RNA makes protein

  - Double-stranded DNA is transcribed into single-stranded RNA which is then translated into protein.

Meiosis and recombination

- Two homologous chromosomes in a parent

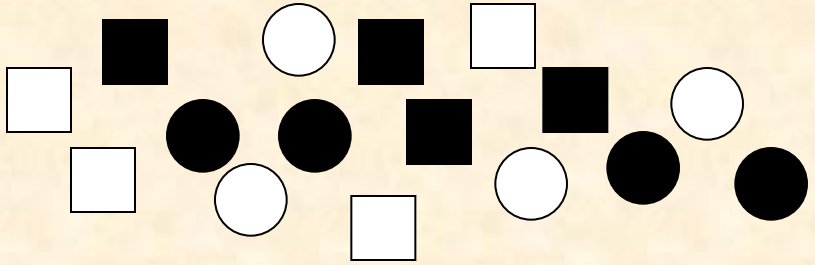


- Chiasmata (i.e. crossovers) lead to *recombinations between genetic loci* (e.g., here, between loci B and C).
- A recombination is more likely if the two loci are more distant from each other.

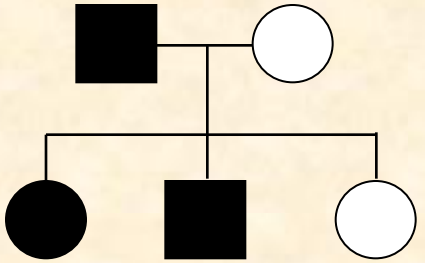
transmitted to the child

Two paradigms of statistical analysis in genetic epidemiology

- **Association analysis:**
  correlation of alleles
  at different genetic loci
  on the population level



- **Linkage analysis:**
  co-segregation (i.e., joint inheritance)
  of alleles at different genetic loci
  within families



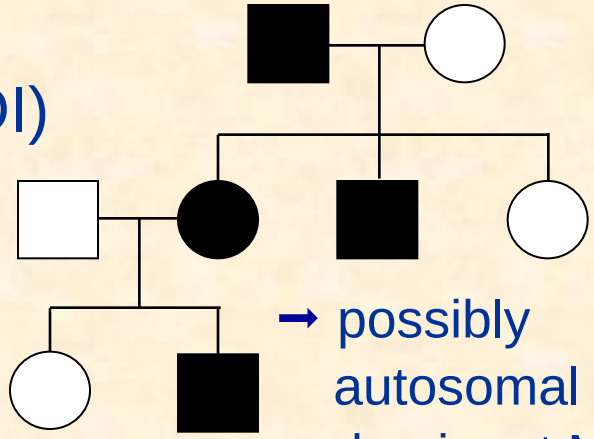
Association analysis is based on assumptions
related to population genetics;
linkage analysis does not rely on these assumptions.

How does a gene act on the disease?

- **Segregation analysis:**

  what is the *mode of inheritance* (MOI)
  of the disease –
  dominant, recessive, X-linked, …?

  What is the penetrance
  of the high-risk genotype?



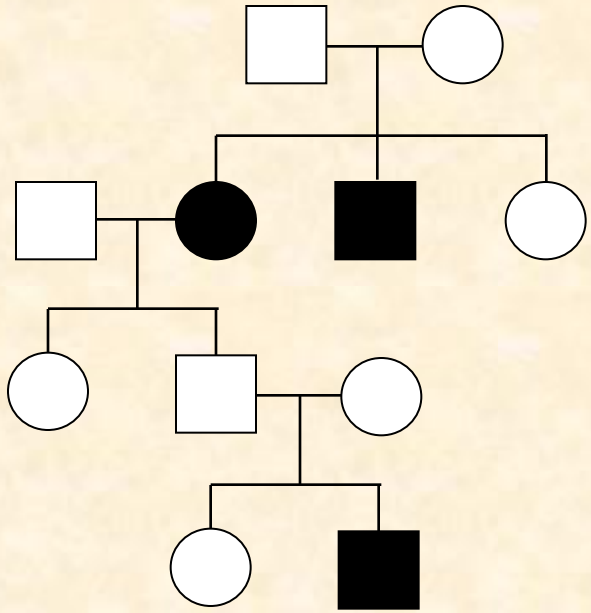
➞ possibly autosomal dominant MOI

Penetrance: Probability of an individual being affected given a certain genotype at disease locus (conditional probability)

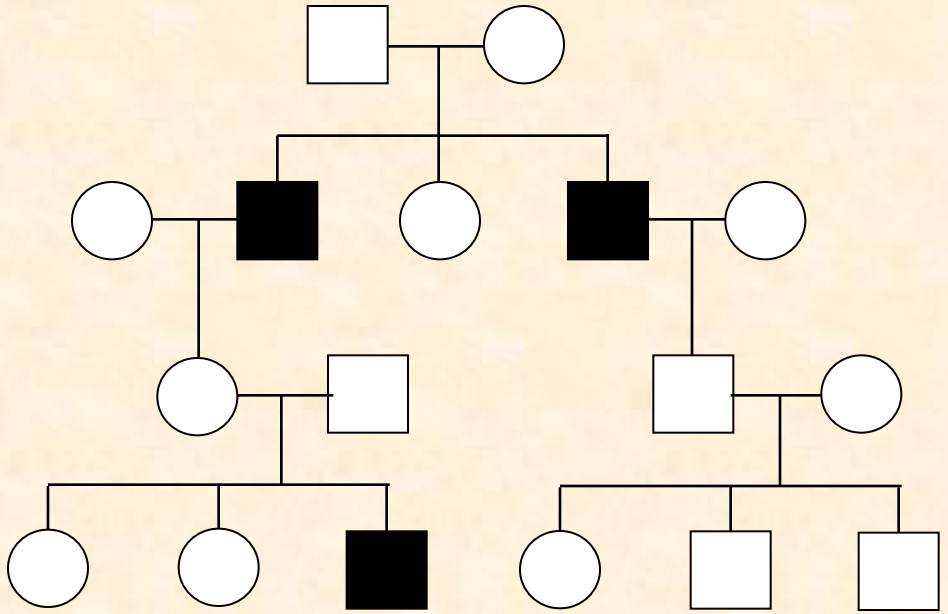
➞ Pure segregation analysis: no markers, only phenotypes

➞ Joint segregation and linkage analysis: uses also markers

Exercise: pedigree examples of Mendelian diseases



possibly autosomal recessive MOI



possibly X-chromosomal recessive MOI

Konstantin Strauch, LMU / HMGU München

“Non-Mendelian“ inheritance

Other modes of inheritance:

- Additive (intermediate) model
- Imprinting, parent-of-origin effect
- Metabolic interference (“overdominance”)

} Single locus

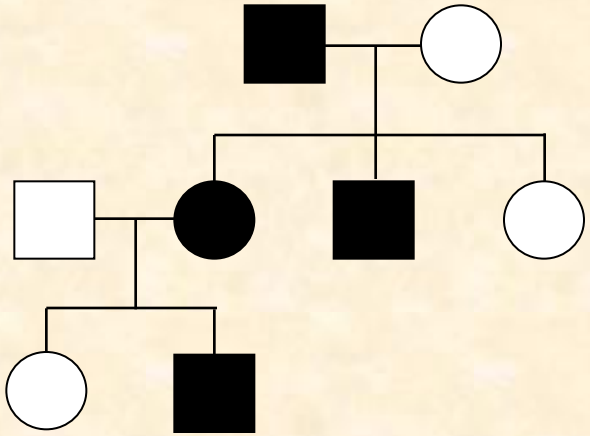
- Genetic heterogeneity: several genetic loci; each of them acting independently, with the same MOI
- Polygenic inheritance: trait is only expressed with mutations at several loci

Segregation analysis with polygenic inheritance:
difficult – each locus may act under a different MOI; hence, segregation analysis not performed with complex diseases.

Ascertainment bias

- **Segregation analysis:**
  Ascertained pedigrees have often been selected to include several affected individuals

  ➞ Ascertainment bias:
  Penetrance will be overestimated



  Problem can be accounted for by specific methods (ascertainment correction)

- **The same problem occurs with case-control studies:**
  here, it is not possible to estimate the absolute risks (nor relative risks) – only odds ratios can be estimated!

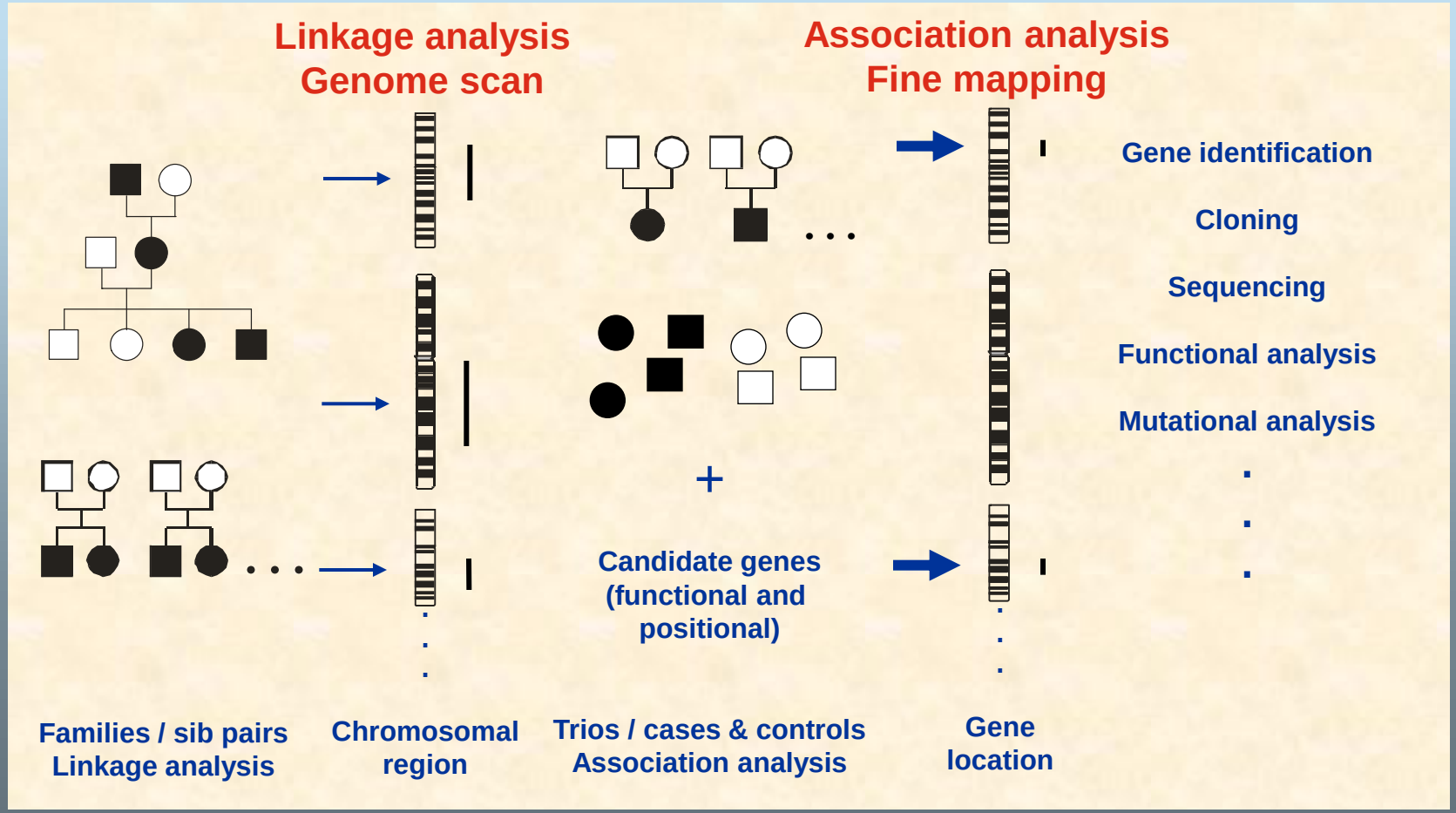
Population-based studies

- Principally, the results obtained from a sample apply only to similar types of samples –
  i.e., subsets of patients or families to which the same mode of ascertainment, inclusion and exclusion criteria apply.
- If transferred to other types of samples (e.g. from families with several affecteds to the general population) without ascertainment correction, the results will be biased!
- **Population-based studies** offer an opportunity to estimate risks (of genetic as well as environmental factors) that are directly applicable to the general population.

Does the trait have genetic causes at all?

- **Look for familial aggregation**
  however, familial aggregation may be due to environmental rather than genetic causes
- **Perform heritability estimates, e.g. with twin studies**
  monozygotic and dizygotic twins,
  both share a large portion of their environment
  - ➞ A higher phenotypic concordance (equality) in monozygotic than dizygotic speaks in favor of genetic effects, i.e., heritability.
- **Like segregation analysis, these steps are usually skipped nowadays when studying complex diseases.**

Scientific research in complex diseases
Positional cloning



Genome-wide association studies (GWAS)

- GWAS with various chips for genotyping (~1 million SNPs)
- Large database (including genetic & other ‘omics’ data)
- Extensive quality control
- Additional ‘genotyping’ by imputation: 10 million SNPs
- Covariables, gene-environment interactions

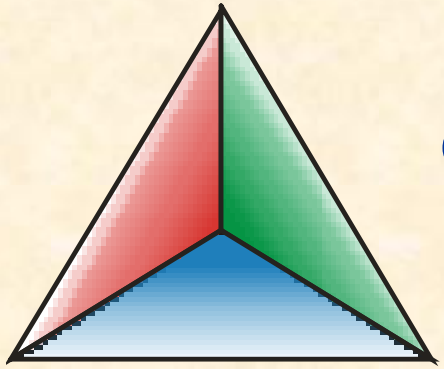
- Numerous collaborations on national and international level
- Primary analyses in KORA, meta analyses in consortia

➞ KORA-gen platform

Scientific research in complex diseases: collaboration

Disciplines involved:

Clinical genetics (patients)



Genetic epidemiology (planning, statistics)

Molecular genetics (DNA analysis)

Methods: application — further development

Mendelian vs. complex diseases

- Mendelian traits: dominant or recessive,
  autosomal or sex-linked

  ➞ Clear genotype-phenotype relation:
  individuals with high-risk genotype: always affected
  individuals with low-risk genotype: never affected

  ➞ Can infer disease-locus genotype from the phenotype!

- Complex traits: often several genes responsible for the disease,
  most likely environmental factors as well

  ➞ Small effect of a single gene, mapping a particular gene is difficult – low power, **large sample sizes needed**!

How to address the difficulties arising with complex diseases...?

- Complex traits:
  Small effect of a single gene,
  mapping a particular gene is difficult – low power!

  ➞ Collect large samples (i.e., many individuals)

  ➞ Sample affected persons with (several) affected relatives

  ➞ Sample affected persons with early onset of the disease

  ➞ Obtain well-characterized phenotypes

  ➞ Use a dense set of genetic markers with accurate typing

  ➞ Avoid confounding (population stratification, batch effect)

  ➞ Perform calculations with adequate statistical models

The missing heritability

- Gene-gene interactions, two-locus modeling
  - ➞ High computational and memory demands
- Copy-number variants (CNVs)
- Epigenetics
- Gene-environment interactions / sex-specific effects
- Rare variants:
  only identifiable by (genome/exome wide) sequencing
  - ➞ Large amounts of data – hard disk storage
  - ➞ Lower power as with common variants
    - – Comeback of family studies

Application of existing methods — Further development

The missing heritability

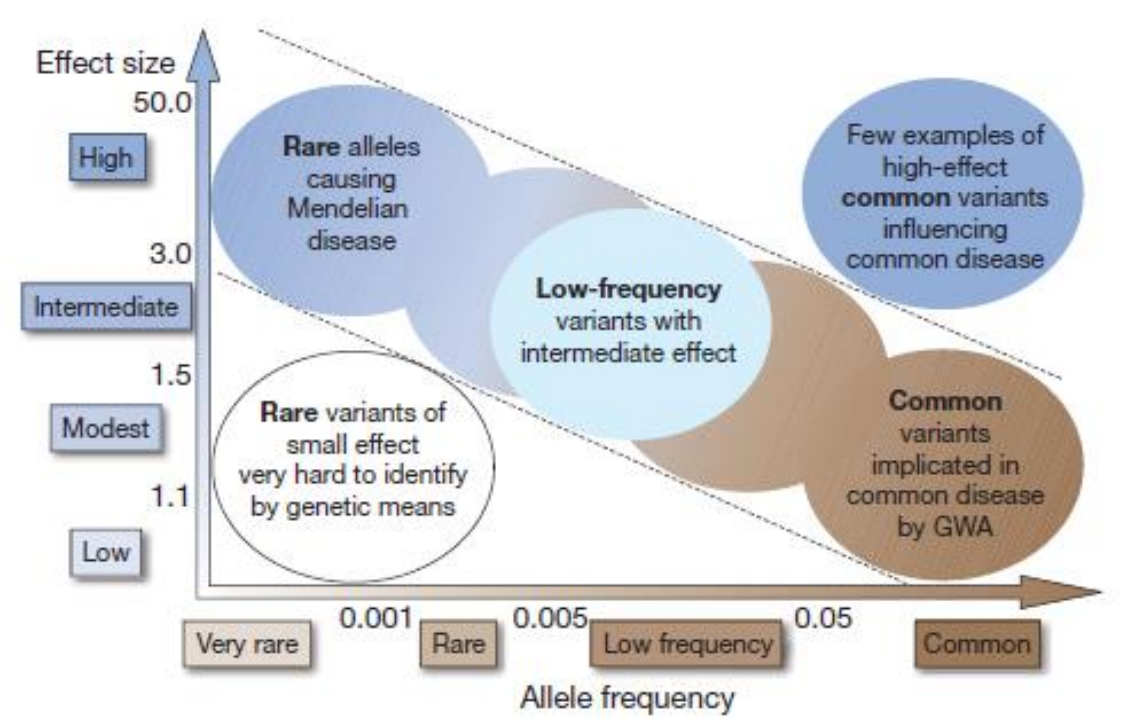


Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.

Manolio et al. (2009) Finding the missing heritability of complex diseases. Nature 461:747-753.

Phenotype definition

- Dichotomous (binary) or quantitative phenotype?
  ➞ depends on the physiological basis of the disease (truly quantitative or dichotomized with cut-off value)
- More broad or more narrow phenotype definition?
  ➞ depends on the way the disease is governed by the genetic factor(s)

  Genetic marker data can be used to **refine the definition of the (sub-)phenotype**
  – concept of “reverse phenotyping” (Schulze & McMahon, Hum Hered 58:131-138, 2004)

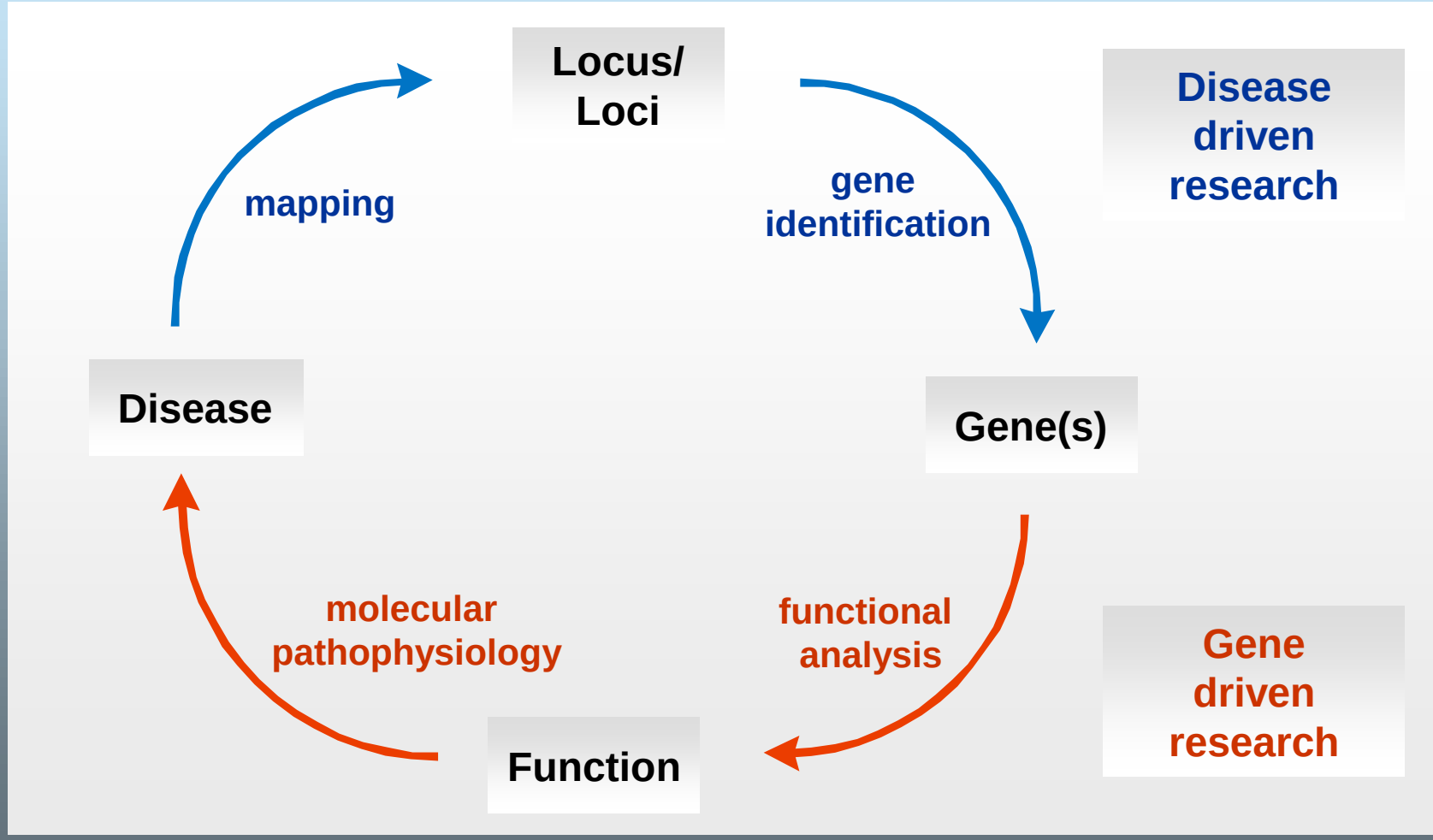
Genetic epidemiology and understanding of disease

- The genetic mapping of Mendelian and complex traits has not only contributed to unraveling the molecular basis of a disease, but also to a more profound understanding of the disease per se.

  By genetic analysis, a disease may have turned out to be not just one disease, but rather several subforms or even completely different disease entities.

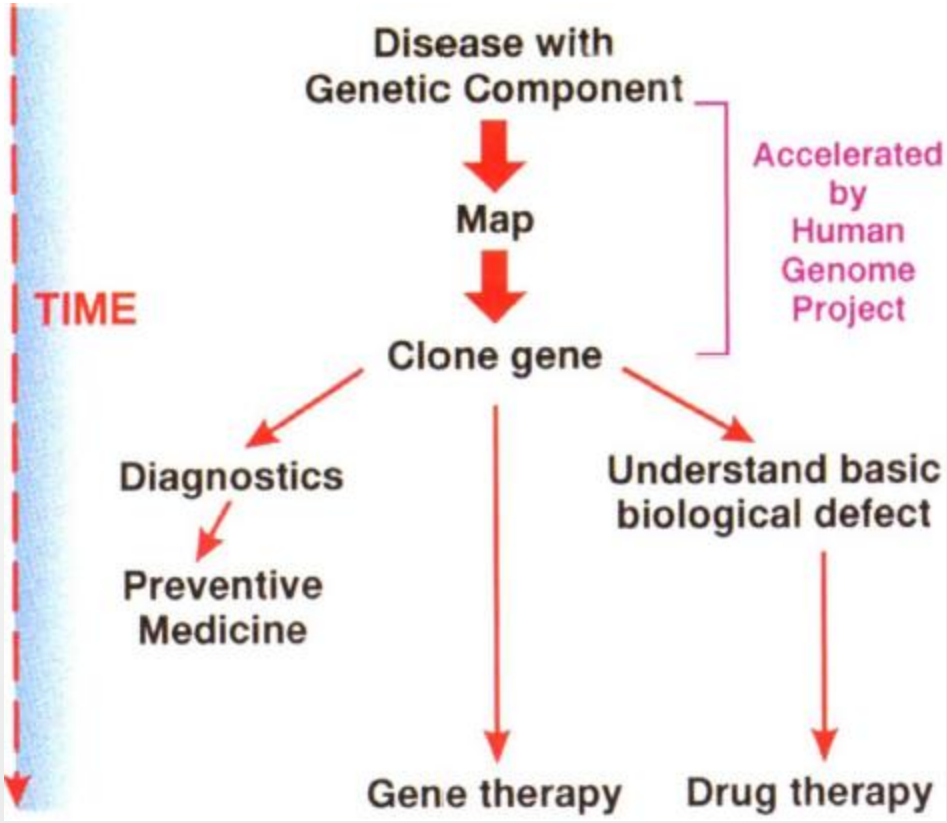
  They possibly require different ways of treatment or prevention → **personalized medicine**
  - *combination of (molecular) diagnosis and intervention suitable for the particular subgroup of patients*

Scientific research in complex diseases
Positional cloning



Francis Collins adapted by Thomas F. Wienker

Why map genes?



Francis Collins:
Positional cloning moves from perditional to traditional. Nature Genetics 9:347-350 (1995).

Types of research questions

- Diagnostic study
- Therapeutic study – interventional study
- Prognostic study
- Risk factor study
- Clinical epidemiology

Clinical trials (in broad sense)

- Genetic epidemiology (Linkage analysis, association analysis)

Environmental epidemiology

- Gene expression study
- Functional analysis & mouse models

Basic research

Types of study designs

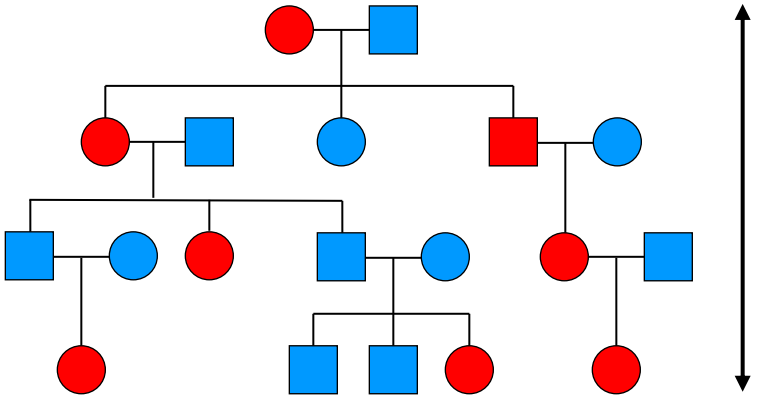
- Cross-sectional study
- Cohort study
- Interventional study, experimental study
- Case-control study
- Cohort study with embedded case-control study
- ... …
- Family study
- Twin study
- Experimental crosses (with animals)

time direction of *reasoning*: prospective or retrospective

way of *data collection*: prolective or retrolective

Study Designs

1

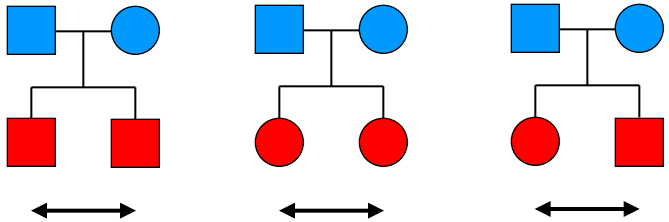


Linkage analysis:

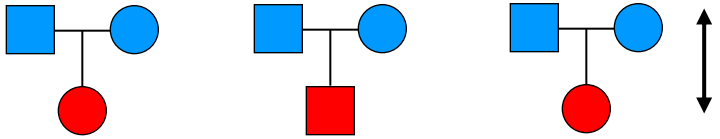
large pedigrees with
several generations

LOD - Scores
NPL Analysis

100

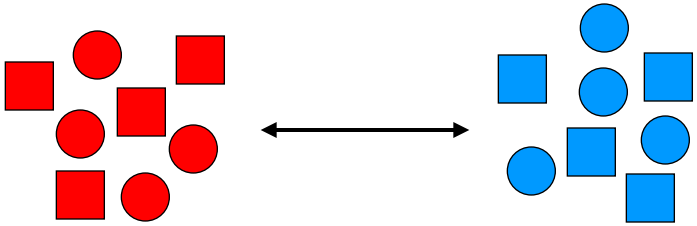


Affected Sib-Pairs
ASP-Analysis
Nuclear families



Association analysis
with family controls

1000



Case-control
association analysis
with genomic controls

adapted from Max P. Baur, IMBIE Bonn

Research questions / Study designs

- Diagnostic study: who is **momentarily** sick vs. well?
- Risk factor study: who will become sick **in the future**? (often retrospective case-control study)
- Prognostic study: in whom will the disease have a favourable course **in the future**? (often prospective cohort study)

– with these types of study: (final) goal of **prediction**

➞ Development of **prediction models** with genetic and other molecular markers is a prerequisite for an **individualized prevention or treatment**

Genetic factors exist…

1. …for the **occurrence of the disease** per se.

2a. …for the **age** at which the disease develops (age at onset).

2b. …for the **severity** of the disease.

2c. …for the **prognosis** of the course of the disease.

2a,b,c: modifying genes

Options for action

➞ Idea: if predictive markers are present,

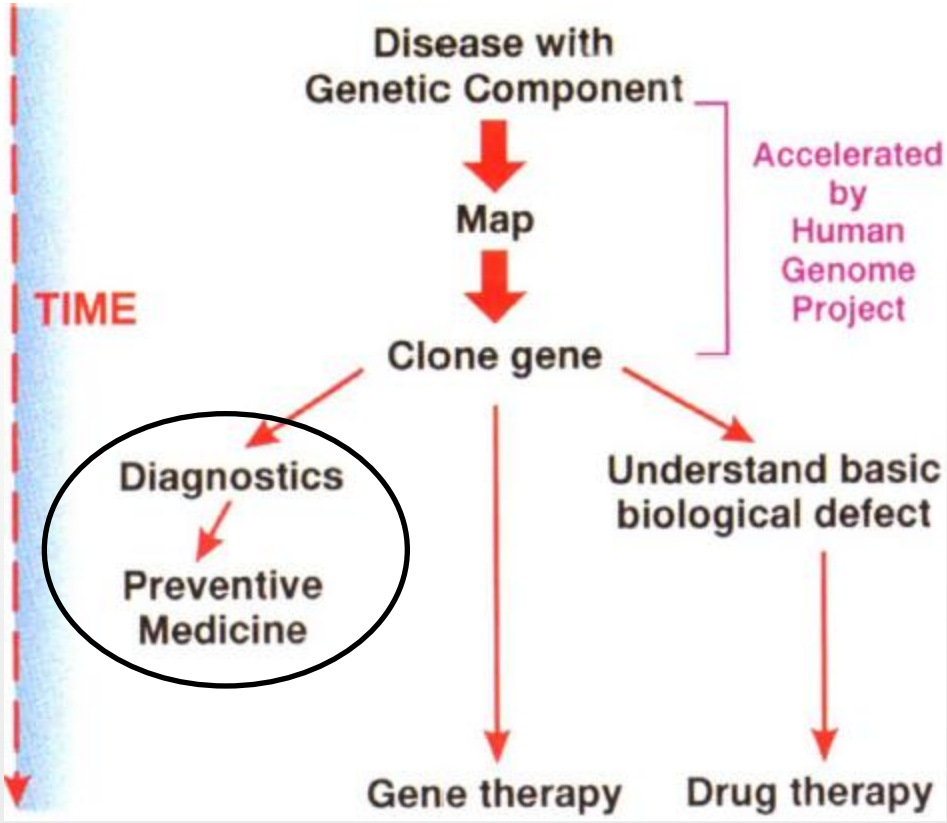
…**preventive measures** are taken

…or it is attempted by **screening** to diagnose a disease as early as possible to have better treatment options, e.g. with cancer.

➞ If diagnostic criteria for a disease are present,

…a suitable **therapy** will be initiated that accounts for the **particular form of the disease**.

Why map genes?



Francis Collins:
Positional cloning moves from perditional to traditional. Nature Genetics 9:347-350 (1995).

Genetic factors influence…

3. …the **efficacy** of therapies or drugs

4. …and their **side effects (safety)**.

Example – pharmacogenetics or pharmacogenomics:
variant of the CYP2C9 gene influences the warfarin metabolism –
a third of the population should receive a lower dose (otherwise higher risk of bleeding).

Molecular markers and personalized medicine

- Predictive markers for the **occurrence** (or earlier onset, more severe phenotype/course, faster progression) **of the disease per se**
  ➞ preventive / diagnostic measures for these patients
- For persons who **are affected by the disease**:
  - Predictive markers for a therapy being **effective**
  - Predictive markers for a therapy being **safe**
  (also applies to preventive measures in the above case, that is, to persons with high disease risk)
- Generally different genes for **development of disease**, **efficacy** and **safety** of a therapy!

Personalized medicine – is reality!

➞ Examples of personalized medicine…

…for therapeutic interventions:

- not simply apply temperature-lowering drug, but specific therapy after thorough physical examination
- not simply apply rad. therapy / surgery, but specific chemo. according to mutations / gene expression

…for preventive measures (primary or secondary):

- Regular routine examinations or change of lifestyle depending on whether further risk factors (familial or environmental) are present

Personalized medicine – clinical and genetic epidemiology

- Genetic loci do not only influence **diseases**, but also **effects** and **side effects** of therapies or drugs.
- Generally not new, but **understanding of molecular basis** and high dimensional genotypes / “omics” data new
- Combination of diagnosis and therapeutic intervention
- Large amounts of data: memory & data management
- Biobanking: ethical, legal and social aspects
- Statistics: Who will be treated and who not ➞ error; Subgroup analyses / multiplicity