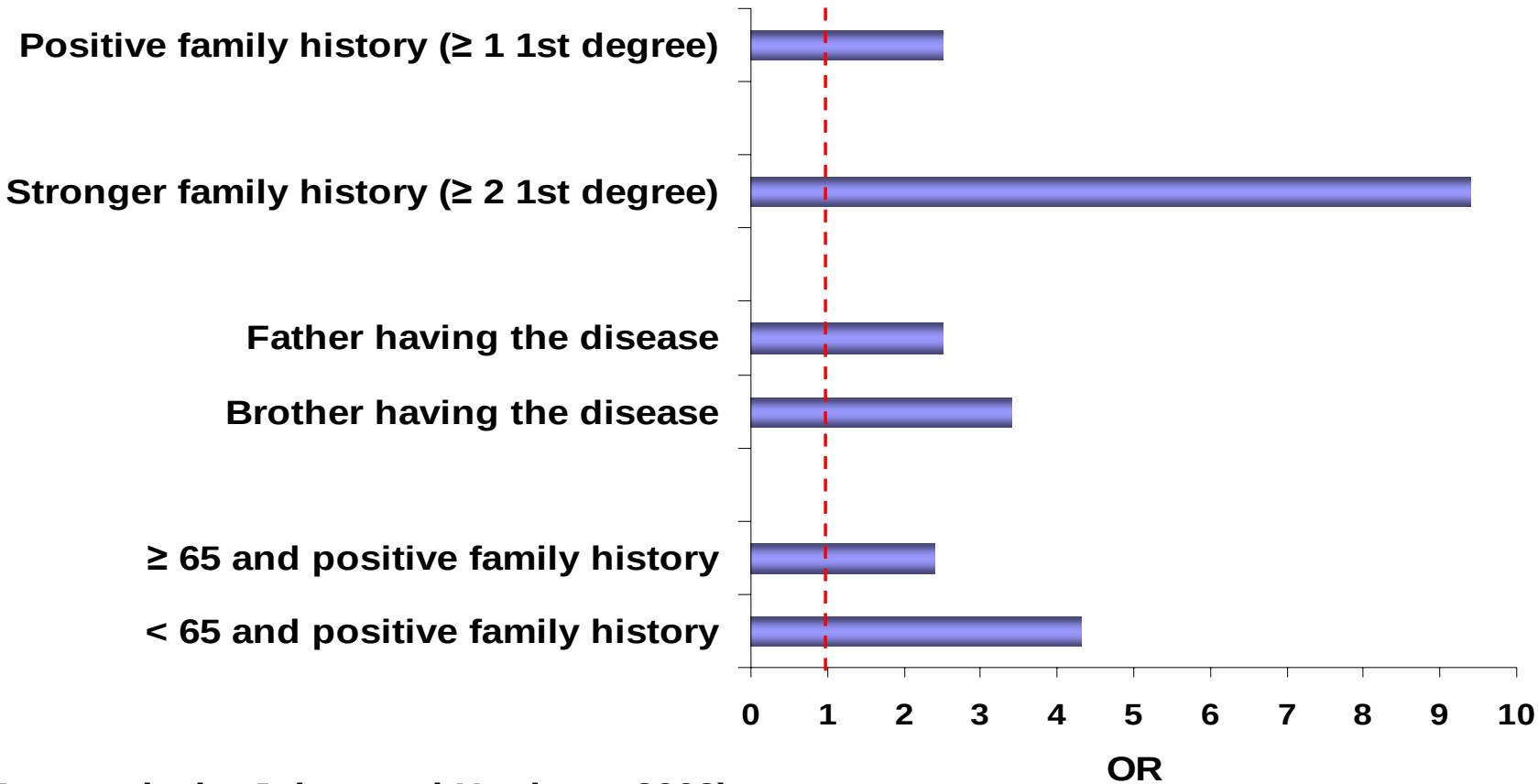
Prostate Cancer Genetics: Today and tomorrow

Henrik Grönberg
Professor Cancer Epidemiology, Deputy Chair
Department of Medical Epidemiology and Biostatistics (MEB)
Karolinska Institutet, Stockholm

IMPACT-Atanta

Familial aggregation of prostate cancer



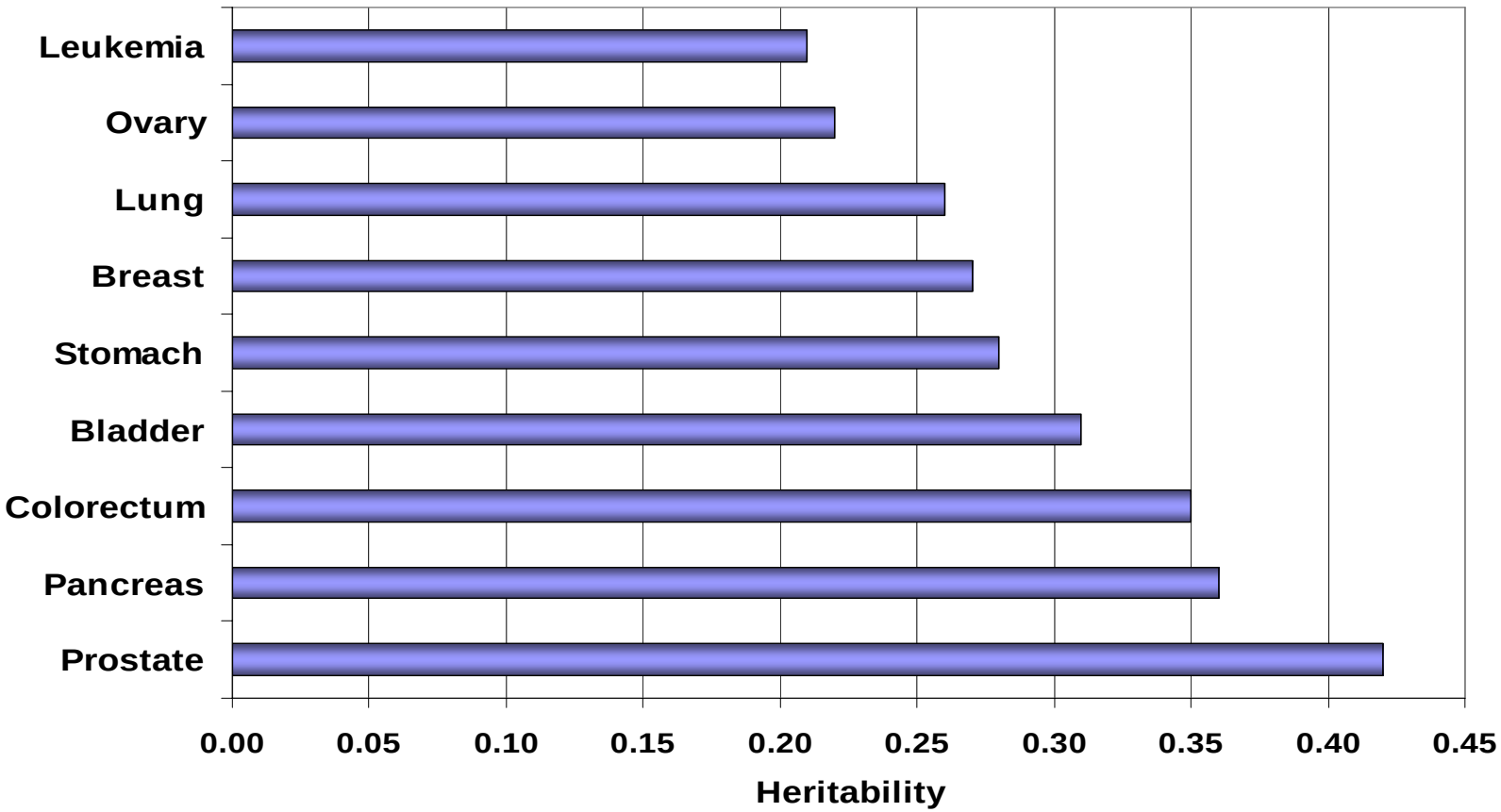
(Meta analysis, Johns and Houlston 2003)

Genes

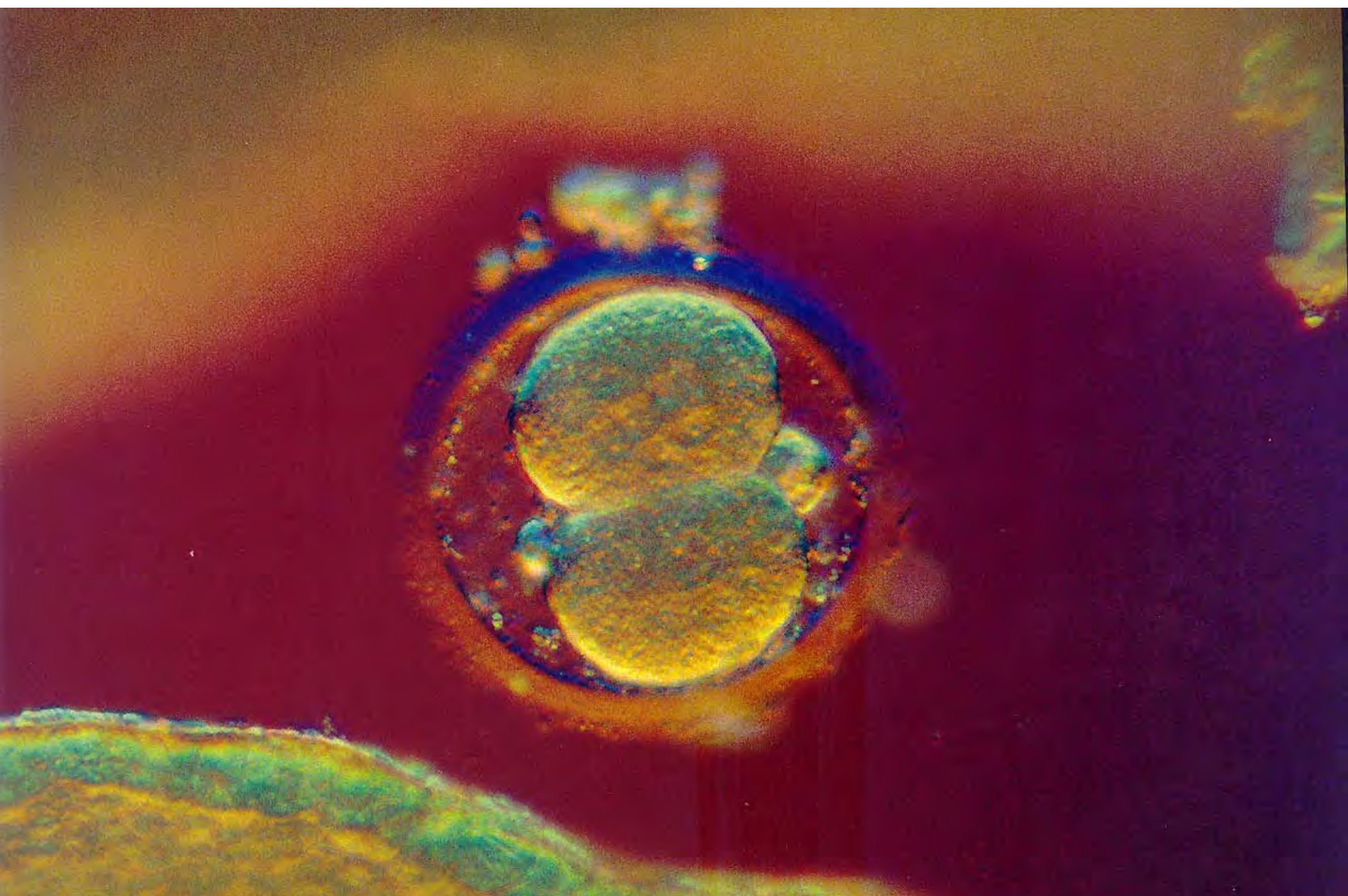
Environment

Karolinska Institutet

Familial aggregation is due to genetics



Lichtenstein et al 2000



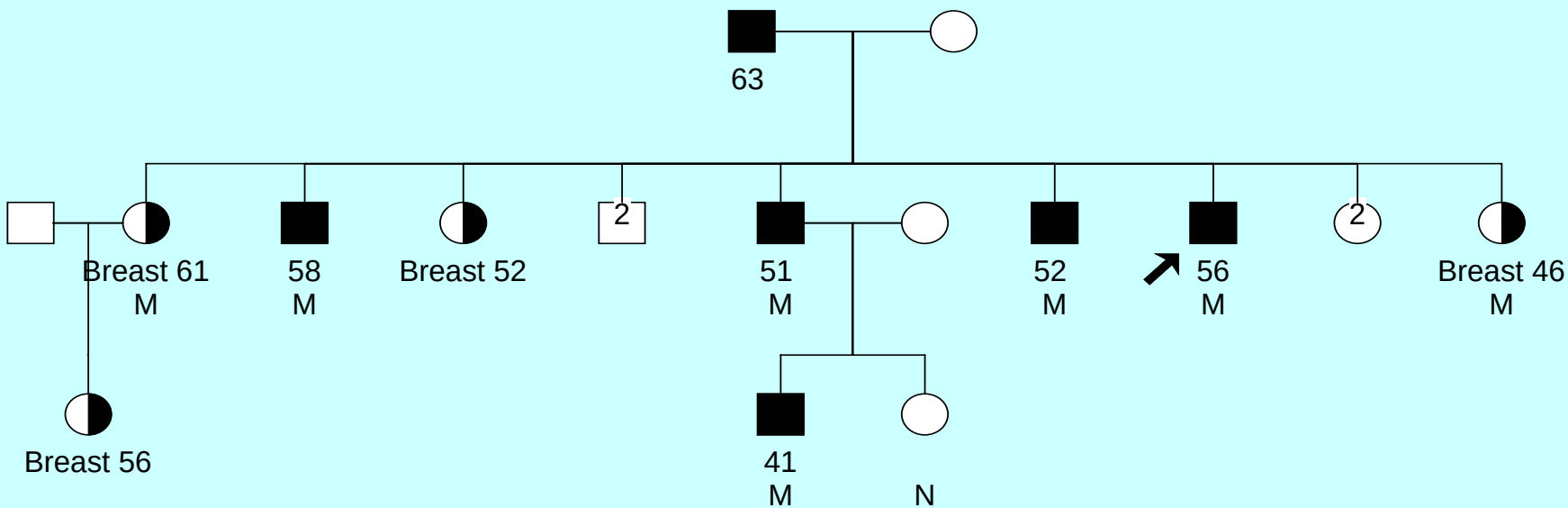
Different scenarios how germline variation effects the risk of prostate cancer

1. Rare variant High Risk (RR>5)
   - Family studies (BRCA1/2)

2. Rare variant Low risk (RR 1.2-2)
   - Sequencing/association studies

3. Common variant Low risk (RR 1.2-2)
   - Association studies/Genome Wide Association studies (GWAS)

1. Rare variant High risk



Swedish family with BRCA2 mutation

UNCOMMON, is going to account for < 0.1%

2. Rare Variant Low Risk (OR= 1.1-2.5)

- Breast cancer
  - → CHEK2, 1100delC mutation , 1.9% of all cases and 0.7% in controls which OR=2.3 (CHEK2 consortium 2004)
  - → ATM gene, OR=2.3 if combining all truncating mutations together (2,7% among cases and 0.4% among controls (Renwick 2006)

- This is difficult and time consuming

3. Common variant Low risk (RR=1.2-1.7)

- Several good examples in other complex diseases e.g. asthma, osteoporosis, stroke, AMI the last year
- Three possibilities to identify these variants
  - → Direct genetic association studies in candidate genes and pathways
  - → Linkage in family studies, fine mapping by association in case-control studies
  - → Genome wide SNP scan (300.000-500.000 SNPs)
    - BREAKTHROUGH LAST YEAR
    - Chromosome 8 and 17 in prostate cancer

nature genetics

A common variant associated with prostate cancer in European and African populations

Laufey T Amundadottir[1,12], Patrick Sulem[1,12], Julius Gudmundsson[1,12], Agnar Helgason[1], Adam Baker[1], Bjarni A Agnarsson[2], Asgeir Sigurdsson[1], Kristrun R Benediktsdottir[2], Jean-Baptiste Cazier[1], Jesus Sainz[1], Margret Jakobsdottir[1], Jelena Kostic[1], Droplaug N Magnusdottir[1], Shyamali Ghosh[1], Kari Agnarsson[1], Birgitta Birgisdottir[1], Louise Le Roux[1], Adalheidur Olafsdottir[1], Thorarinn Blondal[1], Margret Andresdottir[1], Olafia Svandis Gretarsdottir[1], Jon T Bergthorsson[1], Daniel Gudbjartsson[1], Arnaldur Gylfason[1], Gudmar Thorleifsson[1], Andrei Manolescu[1], Kristleifur Kristjansson[1], Gudmundur Geirsson[3], Helgi Isaksson[2], Julie Douglas[4], Jan-Erik Johansson[5], Katarina Bälter[6], Fredrik Wildund[6], James E Montie[7], Xiaoying Yu[8], Brian K Suarez[9], Carole Ober[10], Kathleen A Cooney[7,11], Henrik Gronberg[6], William J Catalona[8], Gudmundur V Einarsson[3], Rosa B Barkardottir[2], Jeffrey R Gulcher[1], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

Common genetic variant identified on chromosome 8q24 associated with prostate cancer

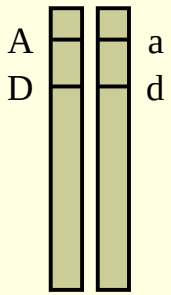
Genetic association studies

Jianfeng Xu, M.D., Dr.PH

Professor of Public Health and Cancer Biology
Director, Program for Genetic and Molecular Epidemiology of Cancer
Associate Director, Center for Human Genomics
Wake Forest University School of Medicine

What causes genetic association?

- Low degree of recombination between two loci
- A genetic association exists between two loci
  - If they are close to each other on a same chromosome, or
  - If the population is “young” or has experienced recent admixture



Basic assumptions of genetic association studies

- A disease has genetic susceptibility
  - As shown by family studies, twin studies, segregation studies

Basic assumptions of genetic association studies

- A disease has genetic susceptibility
  - As shown by family studies, twin studies, segregation studies
- The genetic susceptibility is caused by inherited variants
  - Point mutation (SNPs), indel, large-scale deletions and insertions (CNPs)

Basic assumptions of genetic association studies

- A disease (or drug response) has genetic susceptibility
  - As shown by family studies, twin studies, segregation studies
- The genetic susceptibility is caused by inherited variants
  - Point mutation (SNPs), indel, large-scale deletions and insertions (CNPs)
- Genetic markers are located near the causal variants (coverage)
  - Millions of SNPs are mapped and characterized

Basic assumptions of genetic association studies

- A disease (or drug response) has genetic susceptibility
  - As shown by family studies, twin studies, segregation studies
- The genetic susceptibility is caused by inherited variants
  - Point mutation (SNPs), indel, large-scale deletions and insertions (CNPs)
- Genetic markers are located near the causal variants (coverage)
  - Millions of SNPs are mapped and characterized
- Large enough study population (sample size)
  - Depends on the effect of risk variants

Properties of associated genetic variants

- If associated with the disease, inherited variants or nearby markers are expected to have two properties:
  - They will have a higher frequency in cases than in controls
    - Can be detected using a case-control study design
  - They are more likely to be transmitted to affected offspring
    - Can be detected using family-based study design

So, in an ideal world …

- We would have enough $$$$$$$$ (funding)
  - We could identify very large study populations
  - We could genotype all the variants
  - We would find the risk variants easily !!!

The problem is, in real life ….

- We don’t have enough $$$$$$$$ (funding)
- But we can be smarter by...
  - carefully considering the study design
    - Study populations enriched for specific genetic risk factors
    - Multiple stages
  - efficiently choosing variants to be genotyped
    - Tagging SNPs, discovery vs. confirmation
  - We can still find them and characterize them !!!!

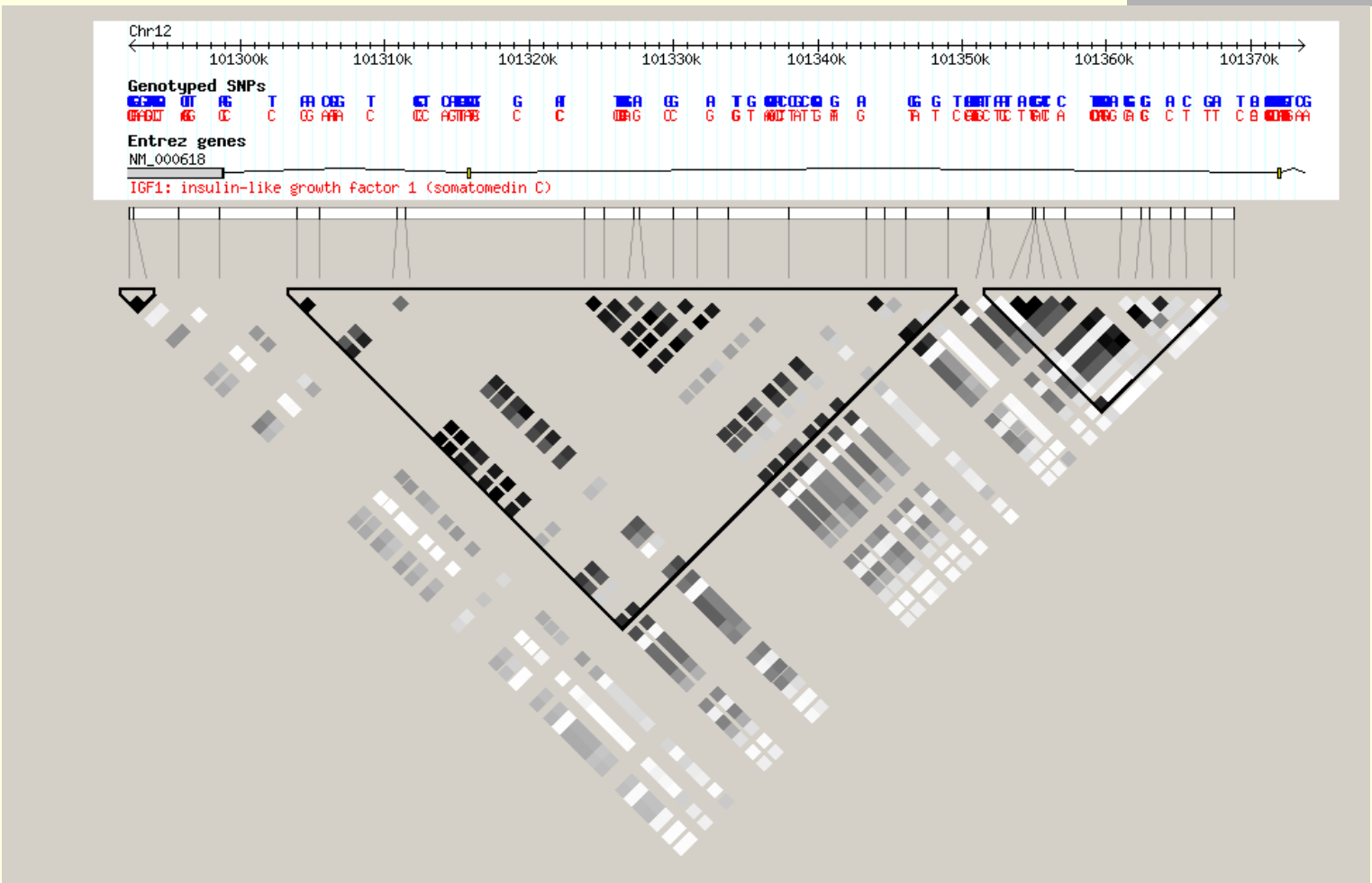
Issues in genetic association studies

- Power
  - OR, MAF, sample size, Type I error, Quanto
- Choice of study populations
  - homogeneous phenotypes
- Choice of SNPs
  - LD, block, tagging SNPs, candidate gene, pathway, and genome-wide
- Choice of analysis
  - Single SNP, haplotype analysis, and imputation
- False positive and false negative
  - multiple tests, population stratification, small effect
- Interaction
  - gene-gene, gene-environment, gene-gene-environment

Study populations

- Familial cases
- Aggressive prostate cancer
- Homogeneous population

SNPs are not independent



Haplotype blocks

- Haplotype blocks
  - Sizable regions over which there is little historical recombination
  - All (or nearly all) pairs of markers are in “strong LD”
  - “Strong LD” if upper 95% CI of D’ is > 0.98 and the lower 95% CI is > 0.7
  - “Strong evidence for historical recombination” if upper 95% CI of D’ is < 0.9
- Haplotype tagging SNPs (htSNPs)
  - Limited haplotypes within haplotype blocks (<< 2^n)
  - htSNPs are selected to capture the majority of haplotypes within blocks
  - Significantly decrease the number of SNPs need to be genotyped

Bins and tag SNPs (tSNPs)

- Bins
  - SNPs can be “binned” into groups of loci that are highly correlated with one another by the measurement of pair-wise r^2
- Tag SNPs (tSNPs)
  - tSNPs is selected from each bin, which exceeds the pre-defined threshold r^2 with any other site within the bin
  - Relatively easy to calculate and do not assume haplotype blocks

Strategies for association analysis

- Single SNP analysis using pre-specified genetic models
  - Allele test
  - 2 x 3 table (2-df)
  - Additive model (1-df), and test for additivity
  - All possible genetic models
- Haplotype analysis
  - Two-marker and three-marker slide
  - Multi-marker
  - Within haplotype block
  - Between two recombination hot spots
  - Imputation

Correction for multiple tests

- Bonferroni correction -- stringent
- Effective number of tests -- take LD into account
- Bayesian approach -- take *a priori* into account, (e.g. FPRP)
- Permutation Procedures -- permute case-control status

Population stratification

- Genomic control
- Structure (STRUCTURE)
- Principal component analysis (EIGENSTRAT)
  - Identify several eigenvectors (ancestries or geographic regions)
  - Adjust genotypes and phenotypes along each eigenvector
  - Compute association statistics using adjusted genotypes and phenotypes
  - No need for AIMs

Methods for assessing gene-gene interactions

- Gene-gene interaction is common
  - Biological relevance
  - May attribute to false negative
- Interaction with main effect
  - Logistic regression, cumulative effect
- Interaction without main effect: data mining
  - Classification and recursive tree (CART)
  - Multifactor Dimensionality Reduction (MDR)
  - Support vector machine (SVM)

Genome-wide association

- Consider costs, false negatives, and false positives
  - Platform: coverage in different ethnic groups, and cost
  - Multi-stages: power and false positives
  - Analytical approach: false positives and false negatives

Summary

- Genetic association studies are powerful
- There are many practical issues in genetic association studies
- The impact of these issues can be minimized by a well-designed study

How can we use these genetic markers in the future?

1. Insight in to new mechanisms in prostate cancer development
   - New treatments

2. Translation to direct patient care
   → Prediction of risk
      - How can a 1,3 risk have any impact on identifying men at high risk of prostate cancer??????
   → Modification of life style
   → Prognostic markers