Design of Genetic Studies

Dan Koller, Ph.D.
Research Assistant Professor
Medical and Molecular Genetics

Genetics and Medicine

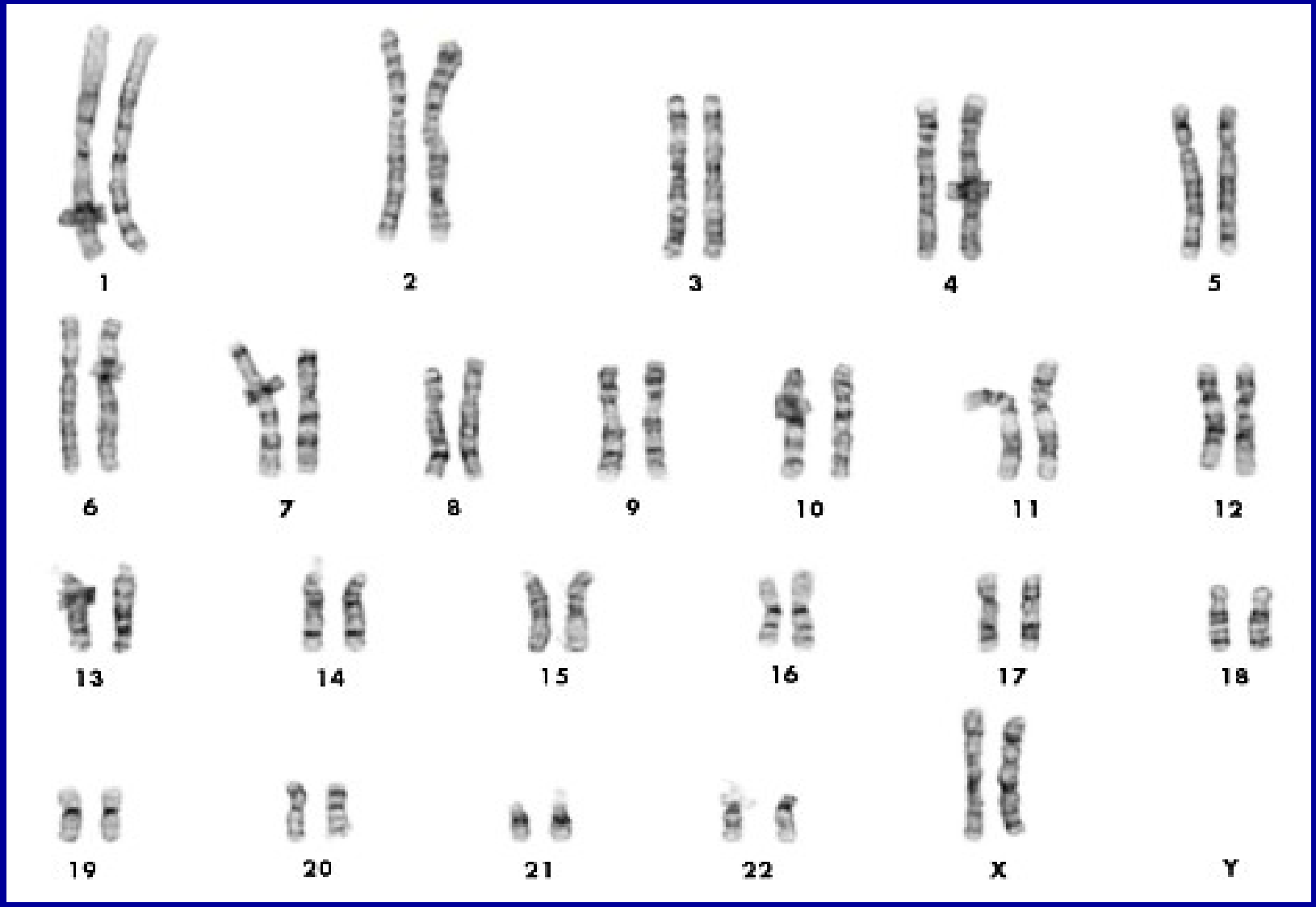
- Over the past decade, advances from genetics have permeated medicine
  - Identification of genes causing disease
  - Identification of genetic risk factors that may modulate disease risk

Objectives

- Review basic genetic concepts
- Review the study designs and statistical methods that led to genetic advances in medicine
- Review concepts in genetics which are being widely used in current studies seeking to dissect the genetic contribution to disease

Basic Genetic Concepts

23 Pairs of Chromosomes



Tools of Genetic Studies

- Molecular markers
  - Microsatellite markers
  - Widely distributed in the genome
  - Variable number of copies of a tandemly repeated segment
  - Typically, this segment is 2 (di-), 3 (tri-) or 4 (tetra-) base pairs long

Allele 1 AGCT**CACACACACACACACACA**ATCG
Allele 2 AGCT**CACACACACACACA**ATCGTCGA
Allele 3 AGCT**CACACACACA**ATCGTCGACCGC
Allele 4 AGCT**CACACACA**ATCGTCGACCGCGG

Tools of Genetic Studies

- Molecular markers
  - Single Nucleotide Polymorphisms (SNPs)

Allele 1 AGCT**C**ACACACACACAC

Allele 2 AGCT**A**ACACACACACAC

Several Important Questions

- What is the evidence that a disease or trait is genetic?
- Do I have the patient and family resources to perform genetic studies?

Is it Genetic?

- Single gene (Mendelian) disorders
  - Obvious they are genetic
  - Reviewing pedigrees makes the mode of inheritance clear
- Genetically complex disorders
  - There may be NO recognizable pattern of inheritance

How to prove a disease has a genetic component?

- Twin Studies
- Familial Aggregation

Twin Studies

- Compare Monozygotic and Dizygotic Twins
  - Monozygotic Twins
    - Genetically identical
  - Dizygotic Twins
    - Like siblings (1/2 genome shared)
- Compare concordance rates of MZ and DZ twins

Twin Studies

- If disease entirely genetic:
  - MZ disease concordance = 100%
  - DZ disease concordance = 50%
- If disease only partly genetic:
  - MZ concordance < 100%
  - DZ concordance < 50%
  - MZ concordance > DZ concordance

Familial Aggregation

- Increased risk for disease among family members of an affected individual
- Compare frequency of disease among first degree relatives of affected individuals with the frequency of the disease in the general population.

Familial Aggregation

Heart disease:

- 3 fold increased risk of disease among the offspring of an affected individual

Parkinson disease:

- 2-3 fold increased risk of disease among the siblings of an affected person

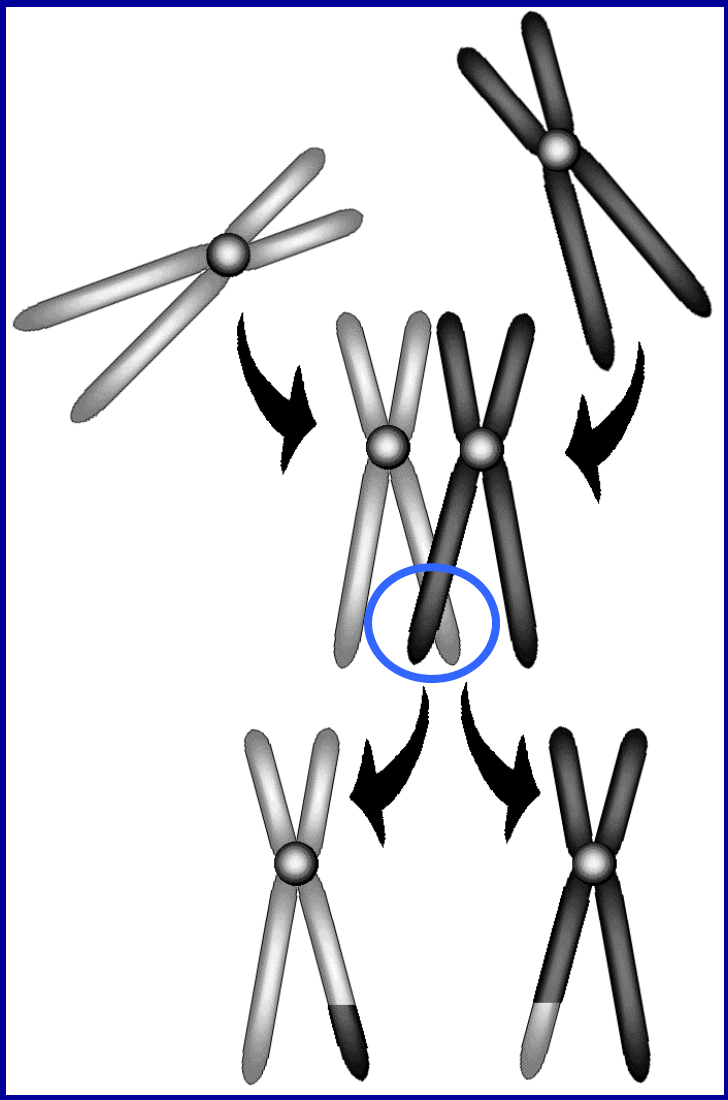
How were genes found for Simple Mendelian Disease

Simple Mendelian Disease

- Most single gene, Mendelian disorders have been identified
  - Examples: cystic fibrosis, Huntington's disease
  - Caused by gross changes in the DNA sequence of a gene
- A few disorders still remain
  - Often found in only a few families

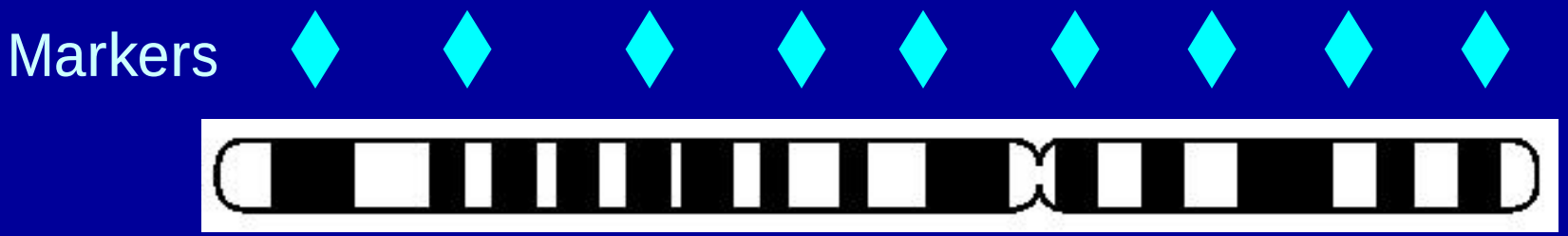
Meiosis and Linkage

- Gamete formation
  - Meiosis I: Homologous chromosomes pair
- Crossing over occurs
  - Genes that are physically close together are more likely to be coinherited
  - Genes that are physically far apart on the chromosome are less likely to be coinherited

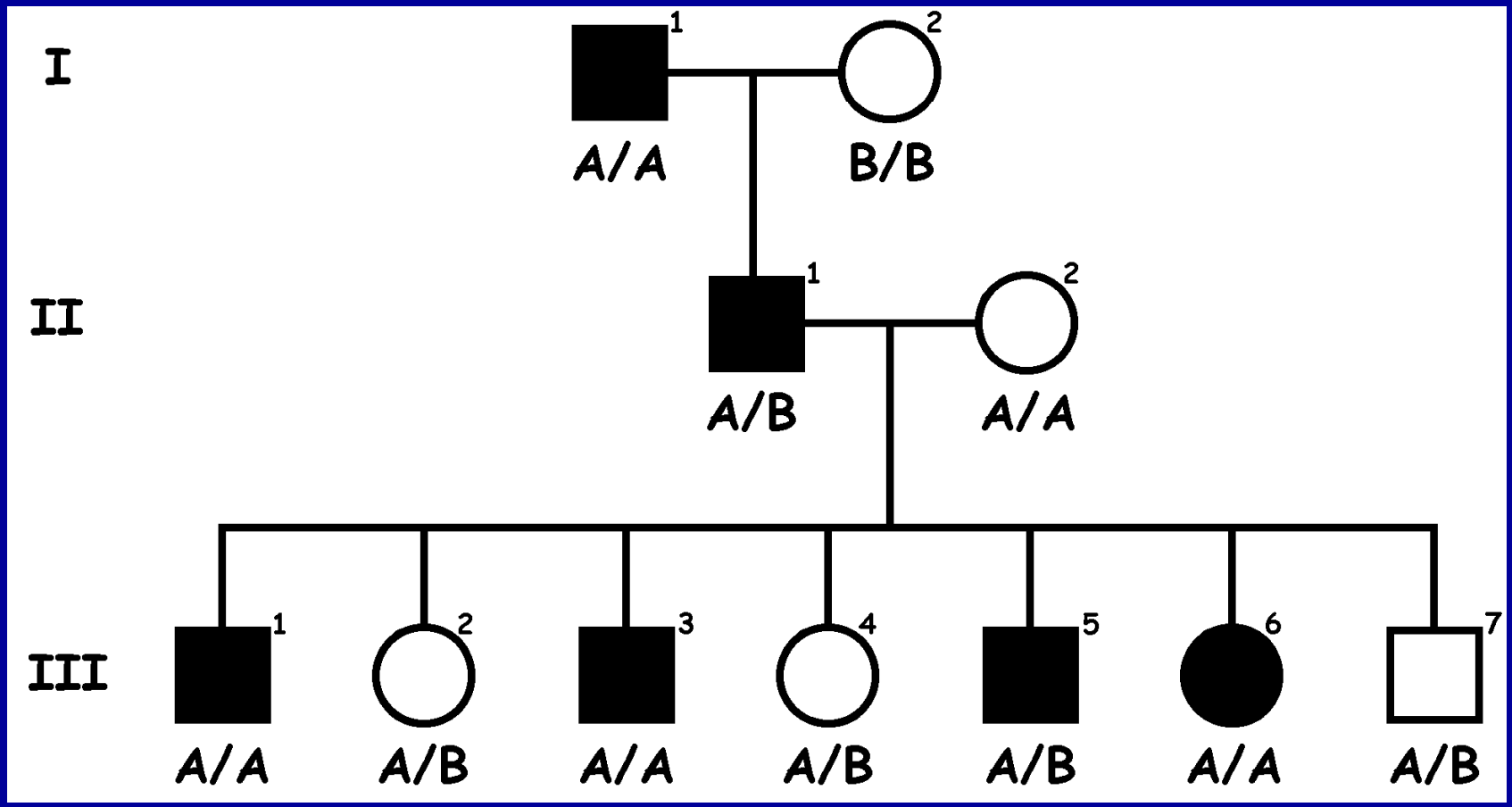


Genome Screen Approach

- Seeks to identify, IN FAMILIES, chromosomal regions that are consistently transmitted to affected individuals.
- Analyze markers located at regular intervals throughout the genome



Linkage: Autosomal Dominant



Compute a LOD score as a statistical test for linkage

Positional Candidate Approach

- Once linkage to a particular chromosomal region has been detected the following steps are followed:
  - Narrow the critical region by adding more families or more members of existing families to the analysis
  - Once the region is reduced to a few centimorgans, identify all genes in the interval
  - Sequence candidate genes in affected and unaffected family members to identify DNA sequence alterations

Simple Mendelian Disease

- Even with the identification of the mutant disease gene, important questions often remain
  - Why is there clinical variability among individuals with the same mutation?
  - Why do individuals with the same mutation develop disease at variable ages?
  - When does disease onset?

Variability in Simple Mendelian Diseases

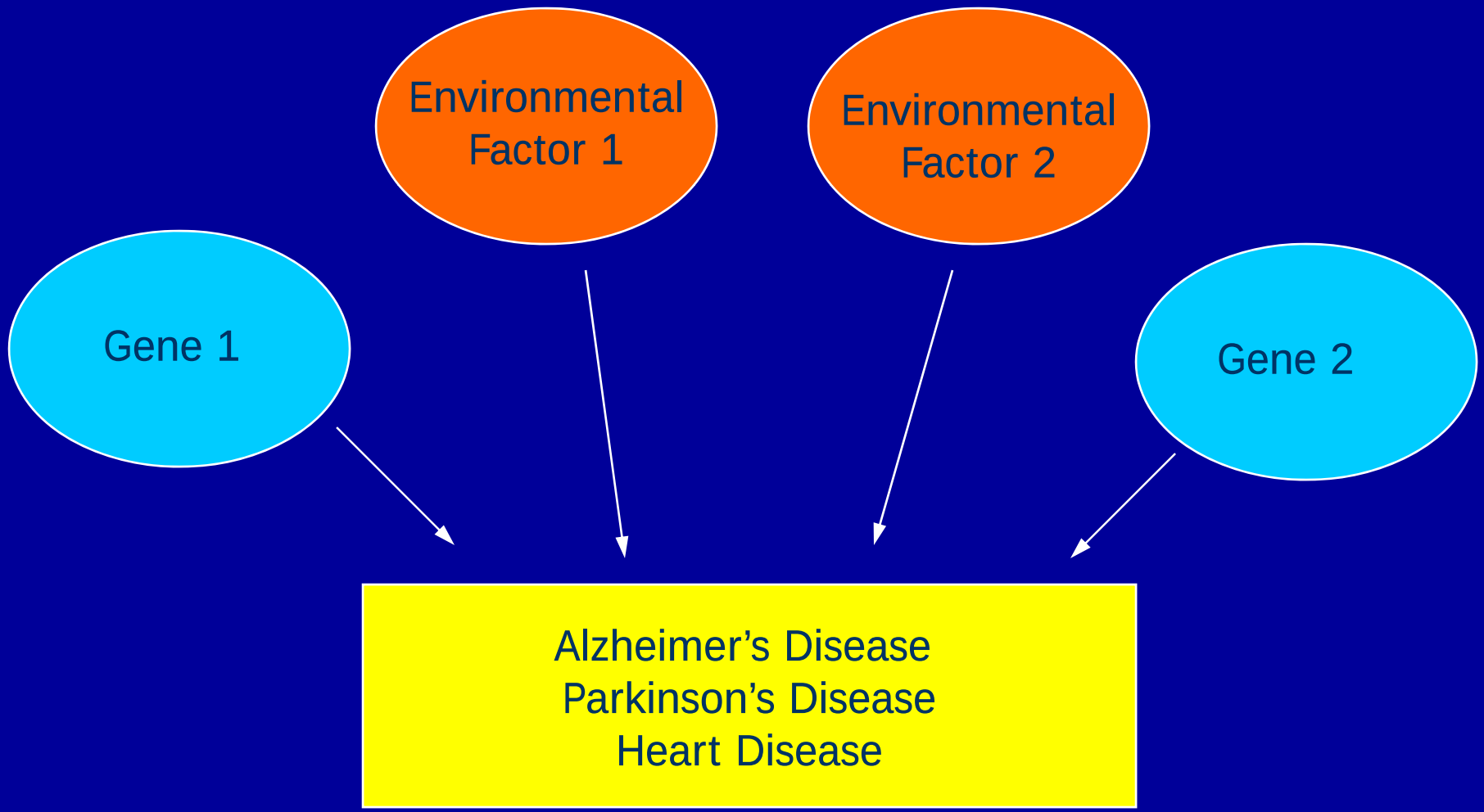
- Currently, there is great interest in determining whether polymorphisms in other genes might contribute to phenotypic variability in what otherwise seems to be a genetically simple disease

Genetically Complex Disease

What is a complex disease?

- Disorders with complex inheritance
  - Likely due to the action of multiple genes
  - Genes may be interacting with each other to result in disease phenotype (epistasis)
  - Affected individuals may have different genetic mutations/polymorphisms leading to the same disease phenotype
  - Environmental factors may be important

Genetics of a Complex Disease



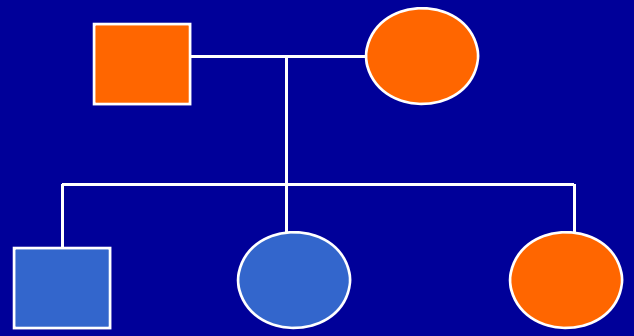
Identifying genes for complex disease

Association

- Test candidate gene
- Collect sample of affected and control subjects
- Compare frequency of a genetic polymorphism in 2 samples

Linkage

- Test entire genome
- Collect families with multiple affected members



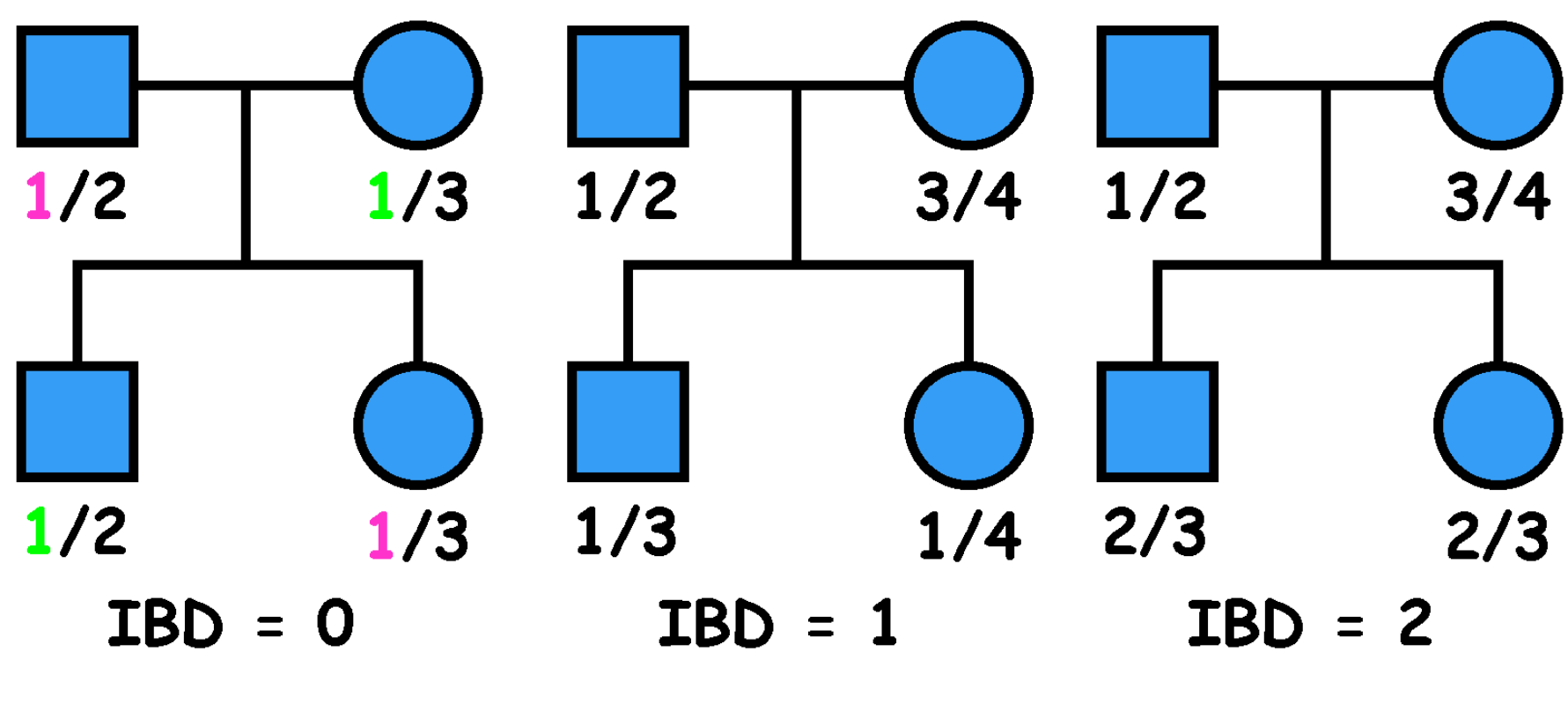
Linkage vs. Association

- Linkage
  - Measures the segregation of alleles and a phenotype within a family
  - Detected over large physical distances
- Association
  - Measures preferential segregation of a particular allele with a phenotype across families
  - Detected over shorter distances

Linkage in Complex Disease

- Identify families with multiple affected members
  - Increases the likelihood that genes are important in disease susceptibility in that family
- Pattern of inheritance less certain
  - Collect family members to follow segregation of disease and marker alleles

Identity By Descent (IBD)



Allele 1 AGCT**CACACACACACACACACA**ATCG
Allele 2 AGCT**CACACACACACACA**ATCGTCGA
Allele 3 AGCT**CACACACACA**ATCGTCGACCGC
Allele 4 AGCT**CACACACA**ATCGTCGACCGCGG

Linkage Analysis

- Employ nonparametric linkage methods
  - Identify chromosomal regions that are preferentially transmitted within a family to the affected individuals.
  - Method is not based on recombination but on IBD marker allele sharing
  - It is often used in the analysis of complex diseases (ex. heart disease, Alzheimer's disease, diabetes)

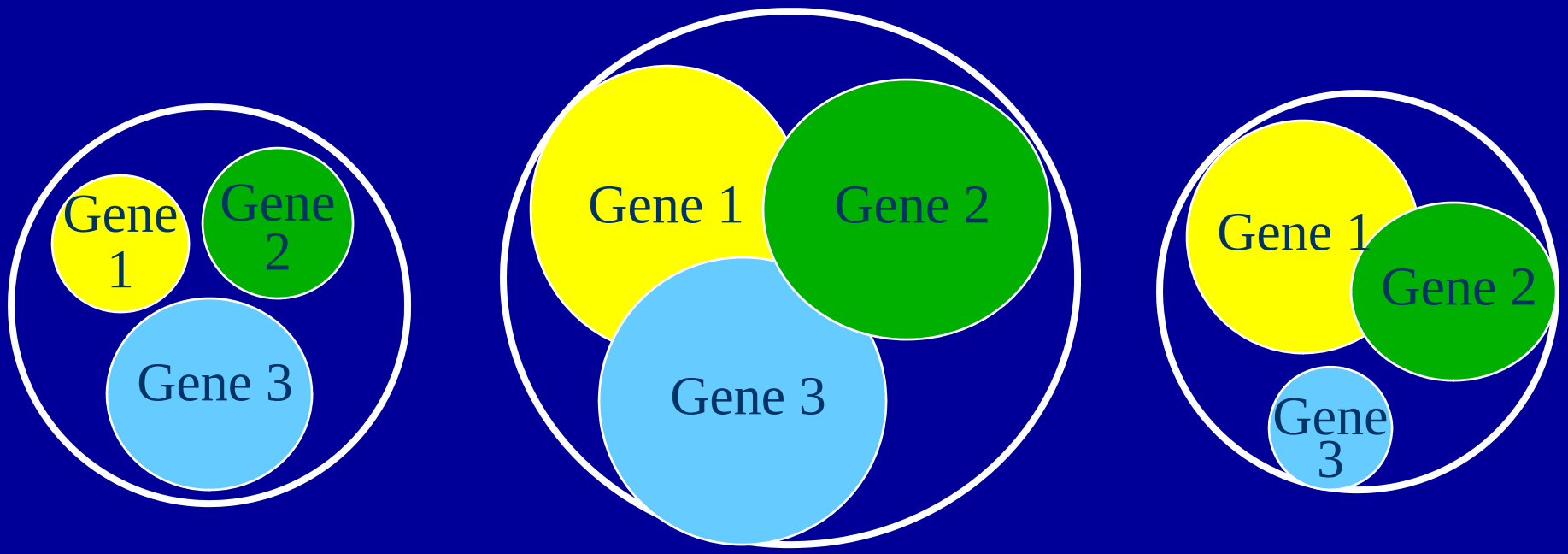
Linkage Analysis in Complex Disease

- This approach often leads to the identification of broad chromosomal regions shared by affected family members
- Often, there can be a lack of replication of linkage results between studies

Replication of Linkage

- Lack of replication may be due to:
  - Initial linkage was a false positive result
  - A different proportion of contributory genes were sampled in the 2 groups
  - Insufficient sample size (power) to detect loci of small to moderate effect size
  - Unique risk genes in certain populations
  - Differences in sample recruitment
  - Differences in environmental risk factors

Replication of Linkage



Initial Sample

Population

Replication Sample

Linkage Approaches in Complex Disease

- This technique has been widely used to identify chromosomal regions linked to
  - Diabetes
  - Inflammatory bowel disease
  - Cancer
  - Alzheimer's disease
  - Bipolar disorder

Linkage vs. Association

- Linkage
  - Measures the segregation of alleles and a phenotype within a family
  - Detected over large physical distances
- Association
  - Measures preferential segregation of a particular allele with a phenotype across families
  - Detected over shorter distances

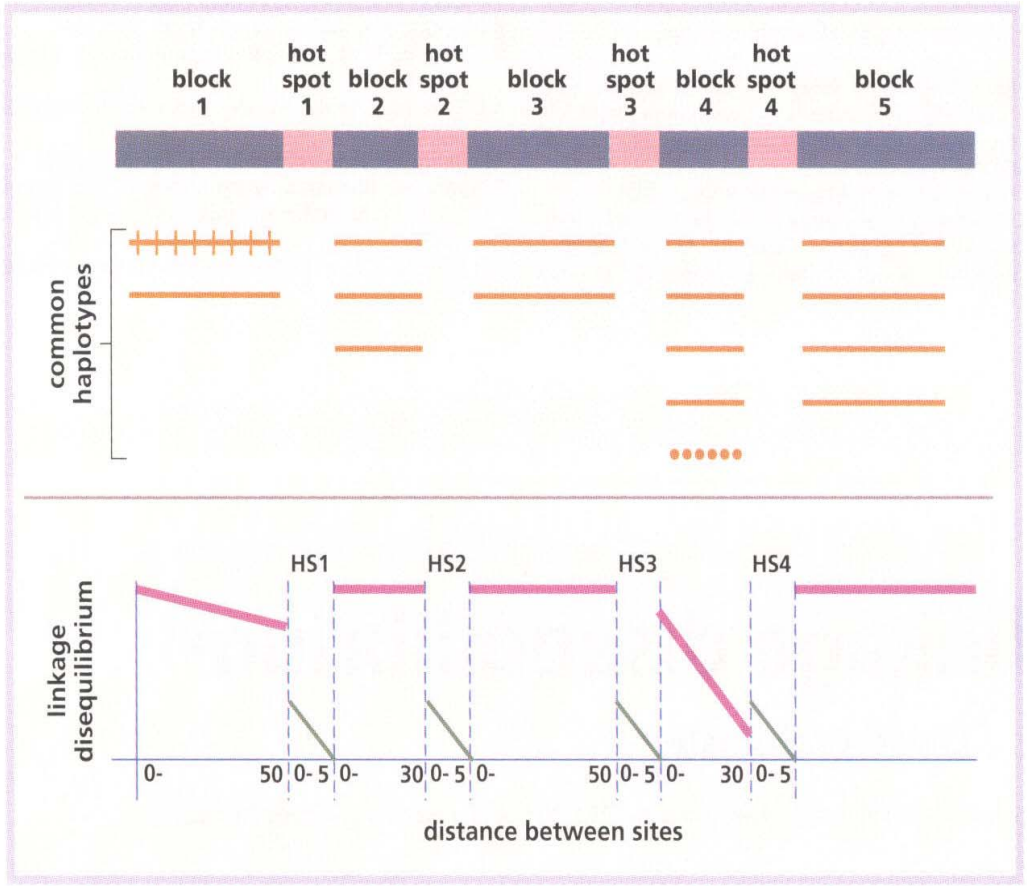
Association Studies

- Typically employed to test the role of a candidate gene
- Candidate gene may be nominated based on:
  - Pathophysiology
  - Genomic location
  - Similarity to other important genes

Association Studies

- Most tests of association are evaluating the evidence of linkage disequilibrium between polymorphisms in a candidate gene and a disease risk allele

Linkage Disequilibrium Studies



BOB CRIMI

- LD is defined as associations between alleles at different loci within the population
- Measure LD:
  - between SNPs
  - between SNP and phenotype

Association Studies

- Two commonly used statistical tests employed to test for association between a SNP and a disease
  - Population based approach
    - Case control design
  - Family based approach
    - Transmission Disequilibrium Test (TDT)

Population Based Association

- For a disease risk, the most commonly applied design to test for association is the case control design
  - Compare allele frequencies of a polymorphism in a candidate gene between the cases and controls
  - Can be quite powerful to detect relatively small genotypic effects, even in modest samples of cases and controls (ex. 100-500 of each)

Population Based Association

- For a quantitative phenotype (ex. Bone density, a-beta levels, etc), the most commonly applied design to test for association is analysis of variance
  - Evaluate the evidence of association using a regression model with the SNP genotype as the main effect.

Population Based Association

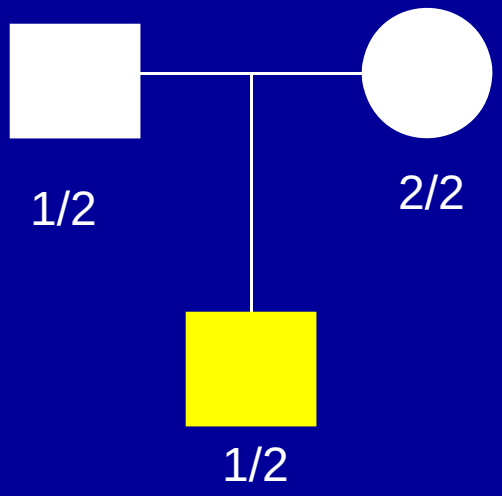
- Advantages
  - Quite powerful to detect relatively small genotypic effects, even in modest samples of cases and controls (ex. 100-500 of each)
  - Relatively easy to collect the cases and controls or general population samples

Population Based Association

- Disadvantages
  - Population stratification – if there are underlying differences in the cases and controls that are unrelated to disease risk, false positive results are more likely

Family Based Association

- Employ a trio design which includes both parents and an affected offspring
  - Compare the frequency of alleles transmitted to affected offspring to the frequency of alleles not transmitted to the affected offspring



Allele	Transmitted	Not transmitted
1	1	0
2	0	1

Family based Association

- Advantages
  - Resistant to potential bias from population stratification since alleles not transmitted in the family are used as the ‘control alleles’

Family based Association

- Disadvantages
  - Requires at least one parent to be heterozygous at the marker being tested, therefore power of this approach is significantly lower than population based approaches
  - Can be more difficult to find 2 generational families willing and able to participate

Association Approaches in Complex Disease

- This technique has begun to be used to test candidate genes for
  - Alzheimer's disease and APOE
  - Diabetes and Calpain
  - Inflammatory bowel disease and NOD2

Summary

- Past success of genetics in medicine has led to the identification of a number of genes which when mutated lead to disease
- Focus of many current studies is to identify risk factors for disease
  - Various approaches can be employed including Linkage and Association

Where to next

- Things to consider when designing a genetic study….
  - Is it clear that the disease/trait is genetic?
  - Do I have the sample base to support this type of research which typically requires large numbers of families or patients?