UNIVERSITÉ DE GENÈVE

Polymorphic variation in the human genome and susceptibility to disease

Samuel Deutsch PhD
Department of Genetic Medicine and Development
University of Geneva

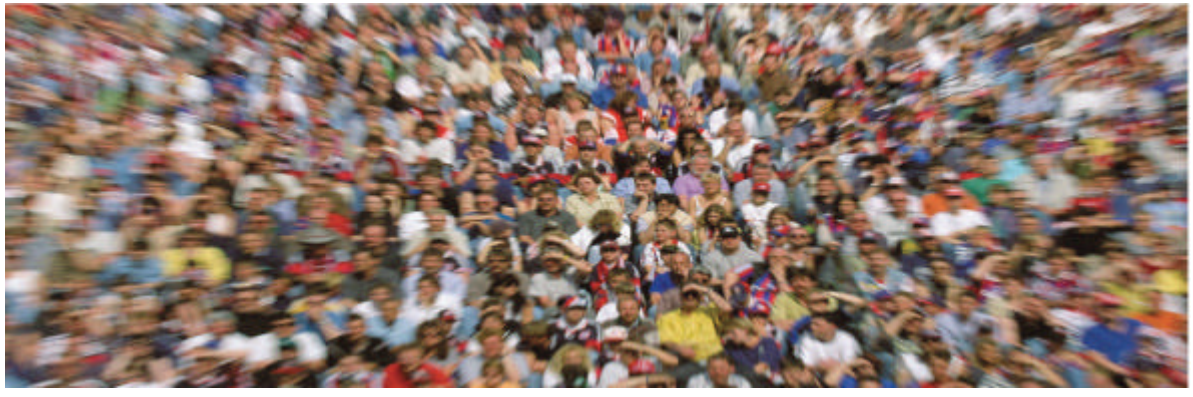
Human genome sequence

Only first Phase !

Consensus sequence for species

Annotation possible !

Human genome sequence: Diversity



Very large amount of **sequence variation** in human populations

Microsatellites

SNPs

Large-scale indels

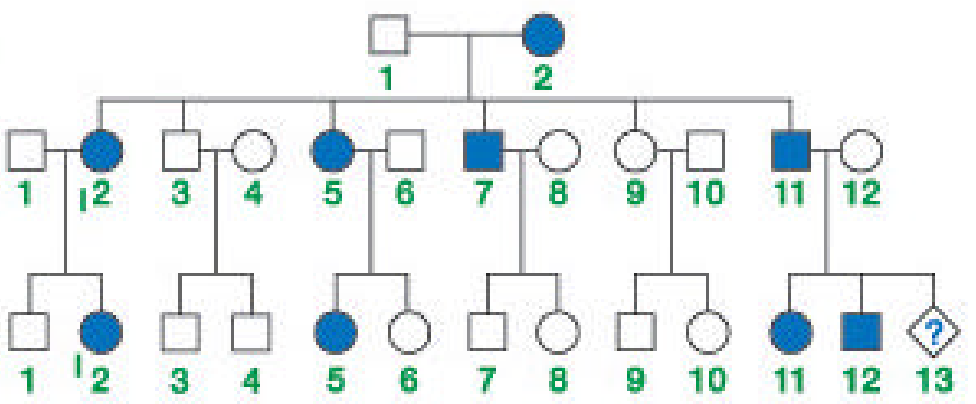
Key to **Human Genetic**

Why is sequence Diversity important ?

- Phenotype (normal variation, *Disease*)
- Evolution
- Risk prediction, Life style
- Pharmacogenomics, Personal medicine
- Forensics

Genes and disease

Is a trait **genetically** determined ?



Clear genetic effect

- **Autosomal dominant**
- **Fully penetrant**

Figure 4-2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Sequence Variation : *most traits are not monogenic !*

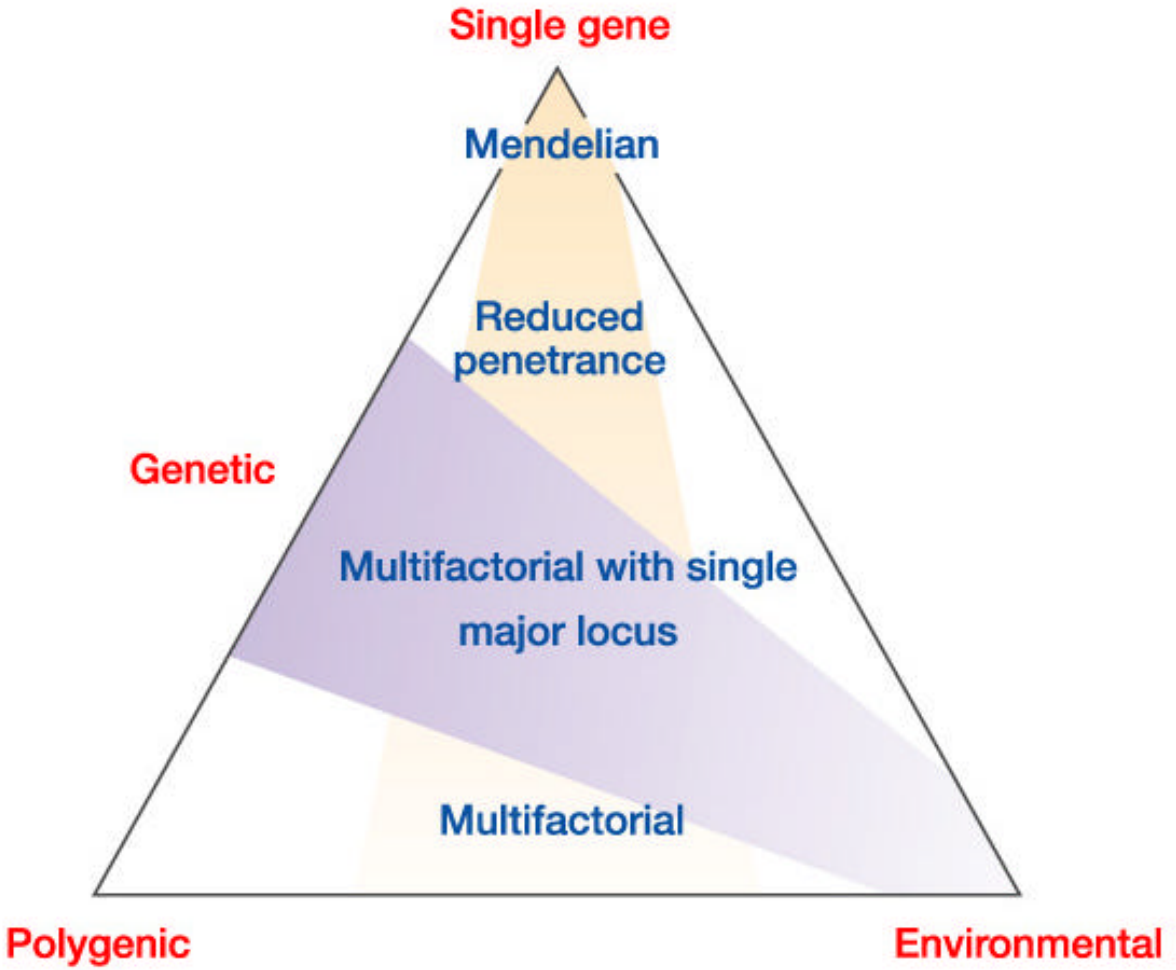


Figure 4-6 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Sequence Variation : *most traits are not monogenic !*

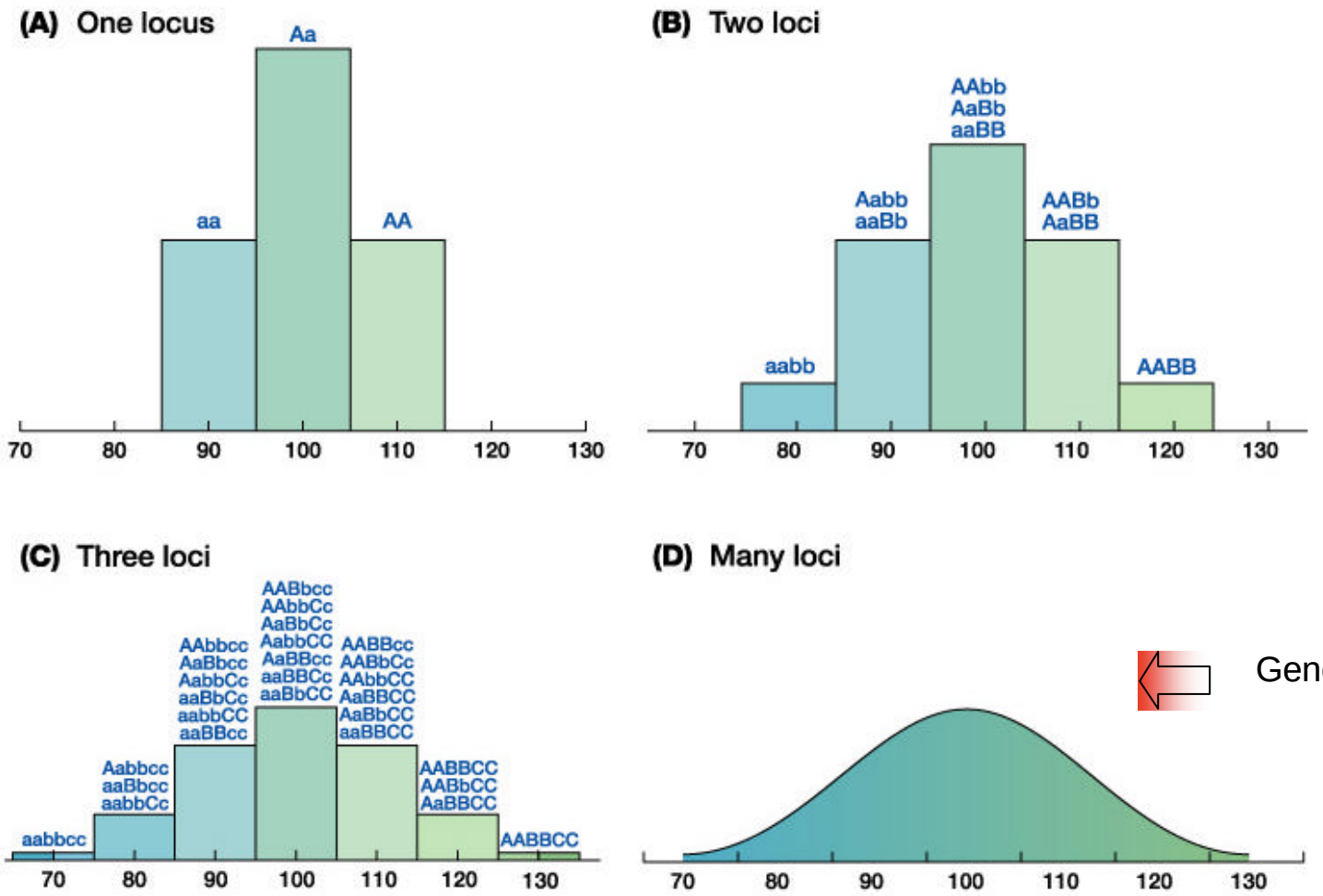
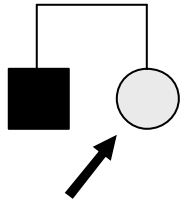


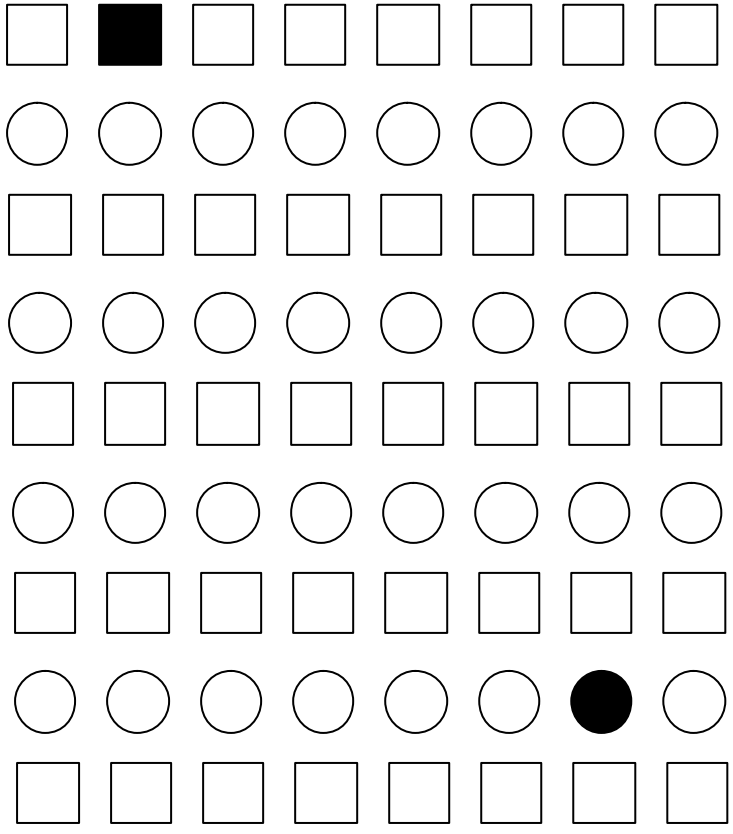
Figure 4-11 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Association studies: is the trait genetically determined ?

Affected sibling



General Population



$$l_s = \frac{f \text{ sibs of affecteds}}{f \text{ gen pop}}$$

Association studies: is the trait genetically determined ?

Disease frequency due to genome sharing

	λ_s
Schizophrenia	12
Asthma	8
Type I diabetes	12
Crohn's disease	25
Multiple sclerosis	24
Aortic stenosis	59
Ventricular septal defect	25
Cleft lip	40

λ_s broken further into multiple loci !

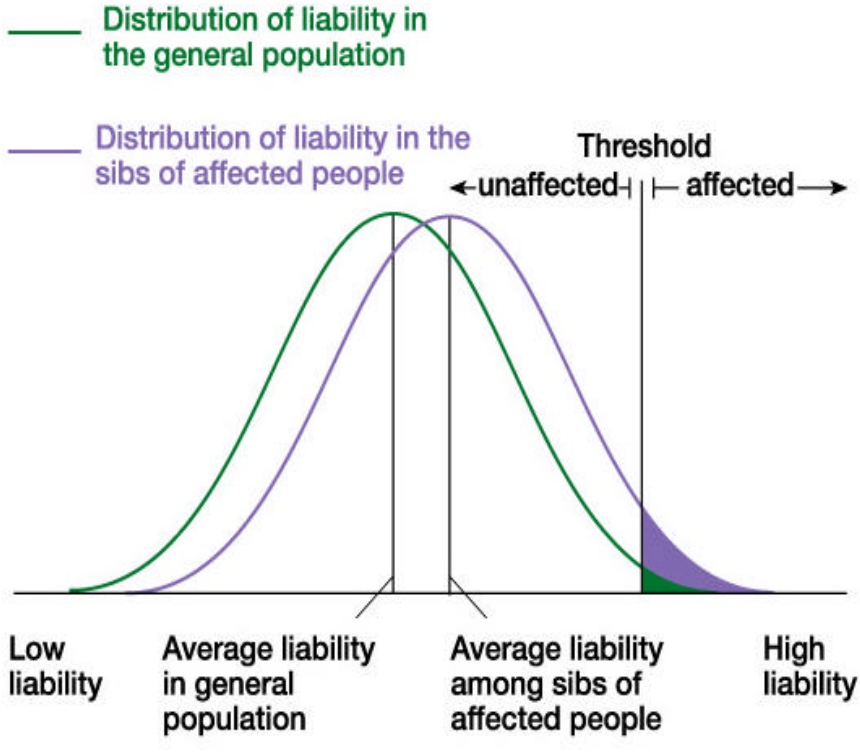


Figure 4-14 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Association studies: is the trait genetically determined ?

Disease frequency
In Monozygotic versus Dizygotic twins

Monozygotic — Share 100% of alleles
Dizygotic — Share 50% of alleles

	% concordance MZ	DZ
Epilepsy	70	6
Multiple sclerosis	18	2
Type 1 diabetes	40	5
Schizophrenia	53	15
Osteoarthritis	32	16
Rheumatoid arthritis	12	3
Psoriasis	72	15

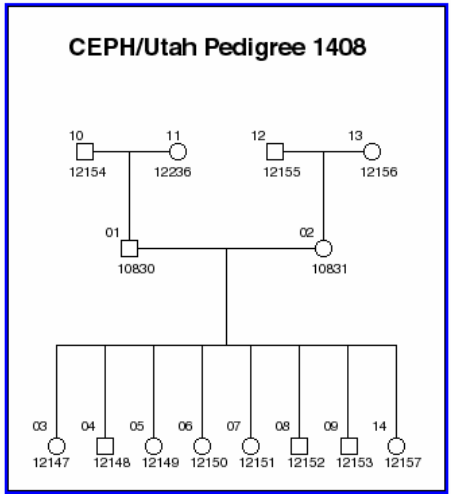
Is a quantitative trait genetically controlled ?

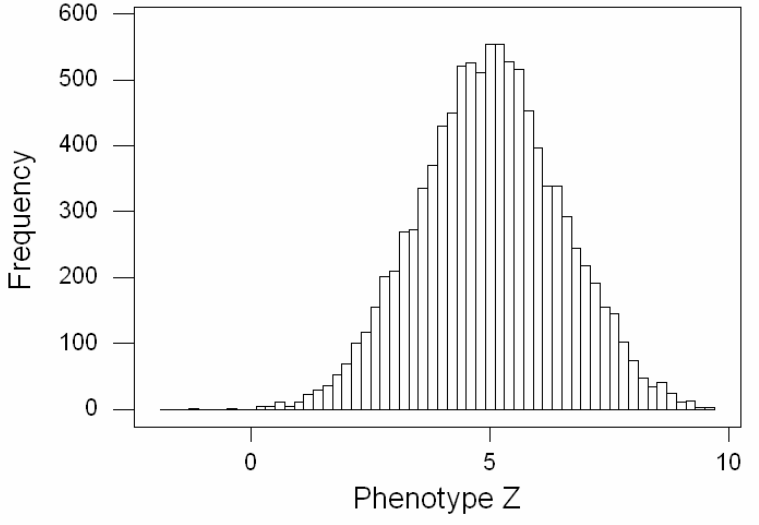
Total variance of a trait → $V_P = V_E + V_G$

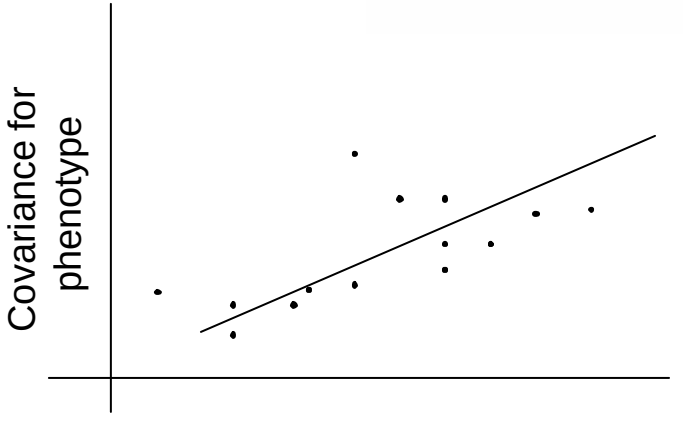
What fraction is genetic ? → $h^2 = V_G / V_P$

Can calculate **heritability** using VC methods

Heritability



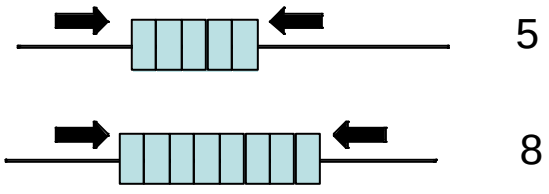




Slope = h2r

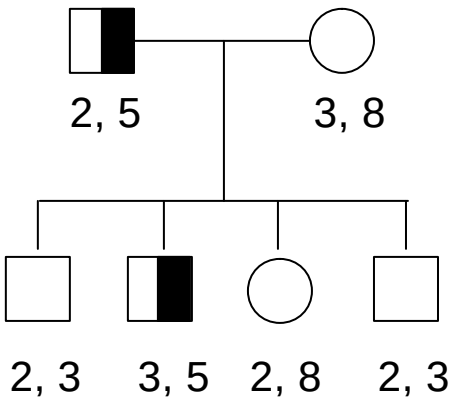
Sequence Variation : *Types and uses*

Microsatellites



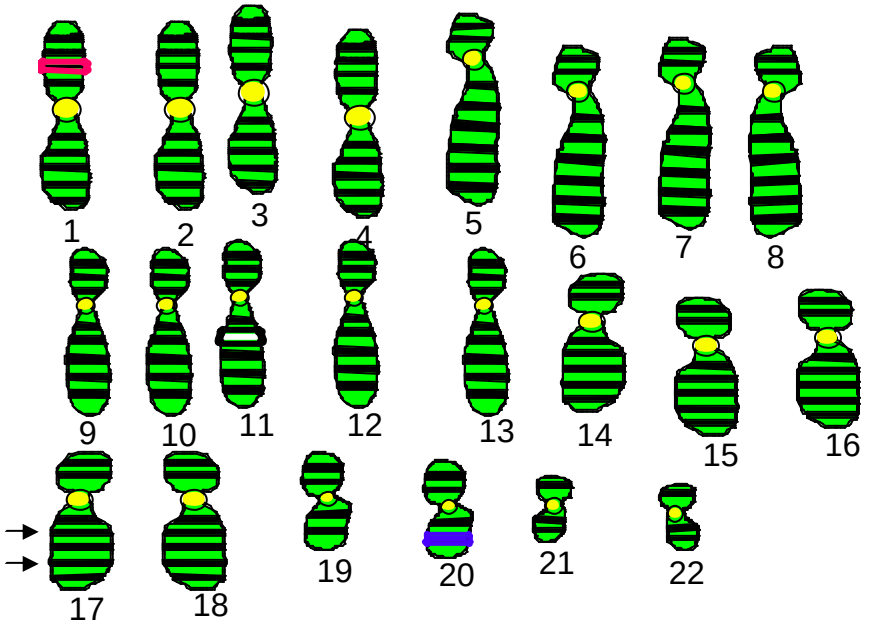
- Variation in **number** of repeats
- **Multi-allelic** in population
- Highly **informative**
- Mostly **non-functional**
- Most useful for **Family studies**

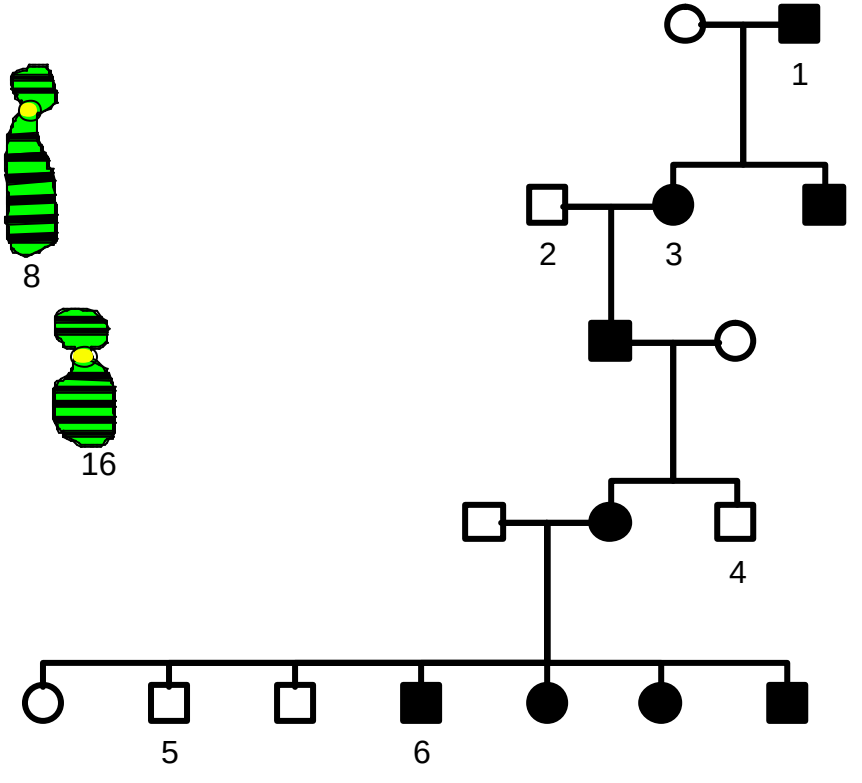
Pedigree



Can be used for
LINKAGE ANALYSIS

Sequence Variation : *Linkage Analysis*





Panel of **Microsatellites** evenly spaced throughout genome

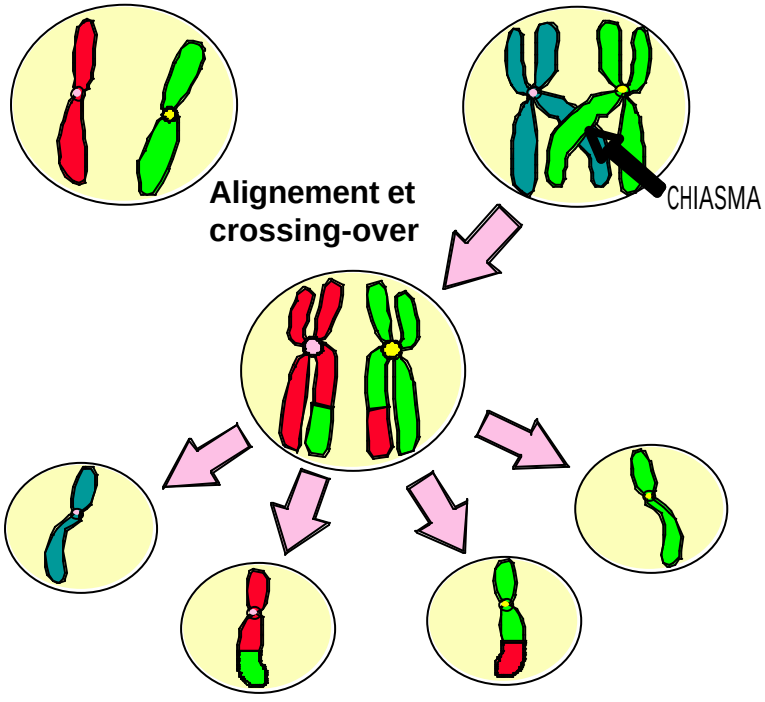
Look at **co-segregation** patterns of disease with alleles of specific markers

Sequence Variation : *Linkage Analysis*

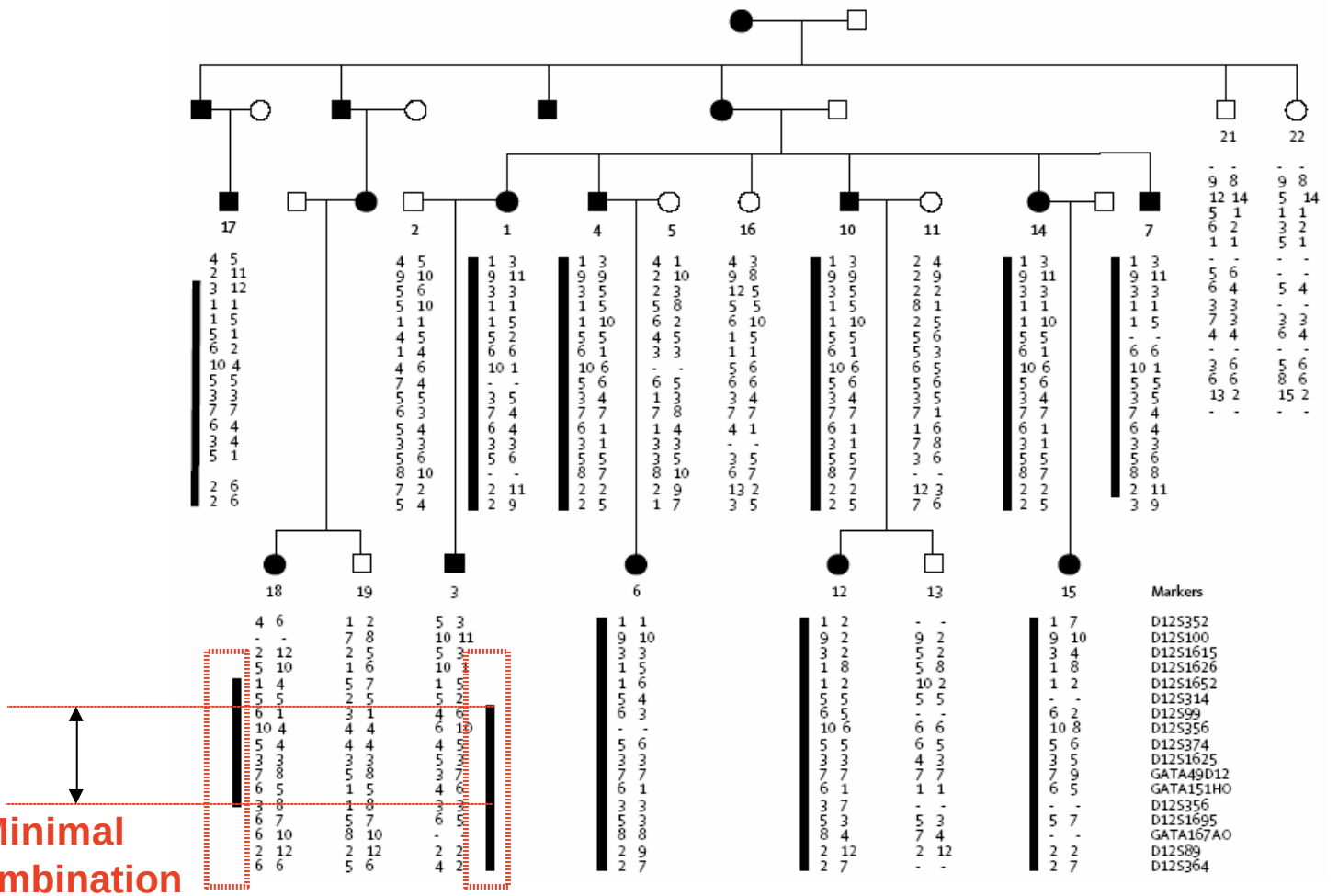
Co-segregation of alleles with disease depends on:

1.Chromosomal localisation.

2. Physical/Genetic distance between **marker** and **disease locus.**



Sequence Variation : *Linkage Analysis*



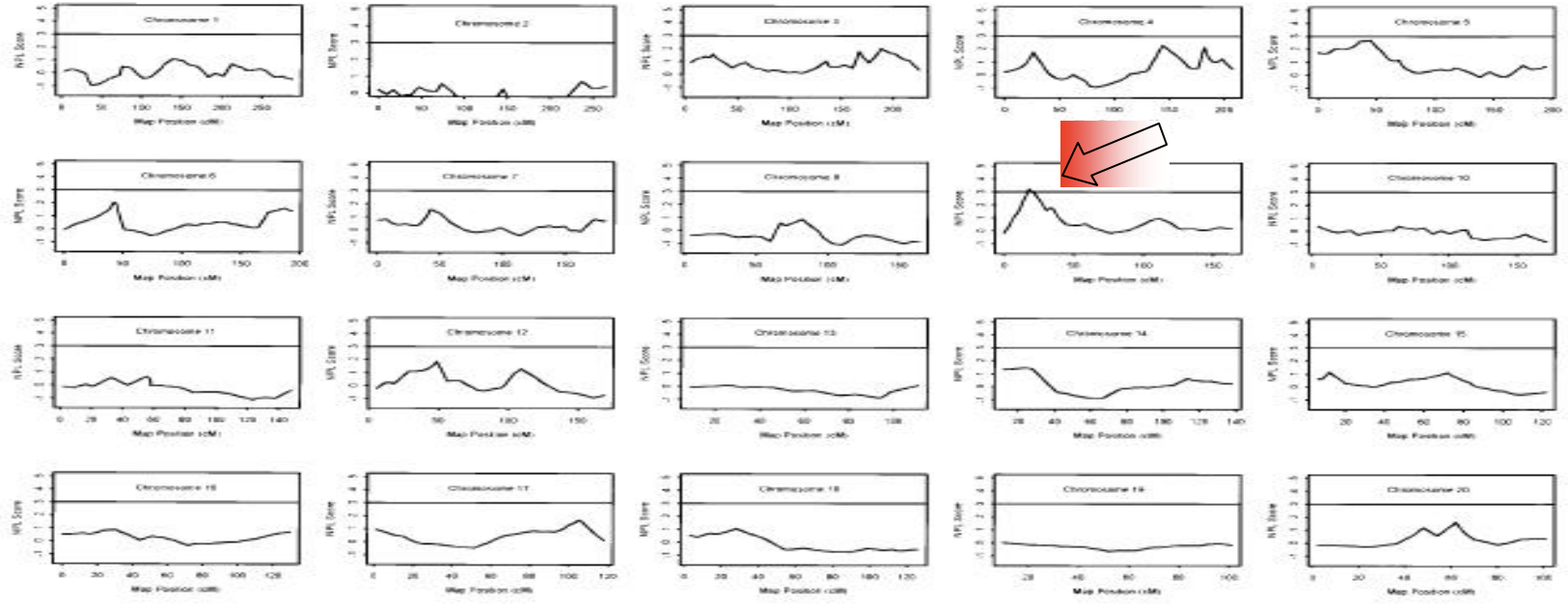
Sequence Variation : *Linkage Analysis*

$$LOD(\theta)=\log_{10}\left[\frac{Like(\theta)}{Like(\theta=\frac{1}{2})}\right]$$

LOD score calculated by maximum likelihood :

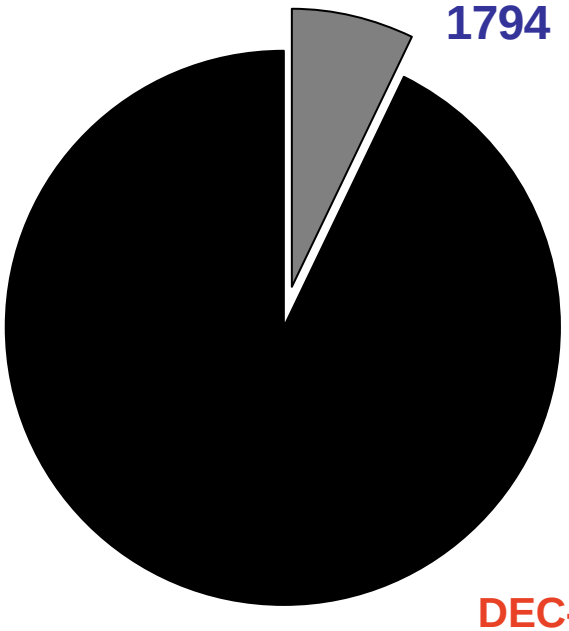
Likelihood of observation / likelihood observation by chance

LOD > 3 is usually considered to be significant on a genome-wide basis



Mapping monogenic disorders: *Great success story !*

Genes with mutations causing human disorders

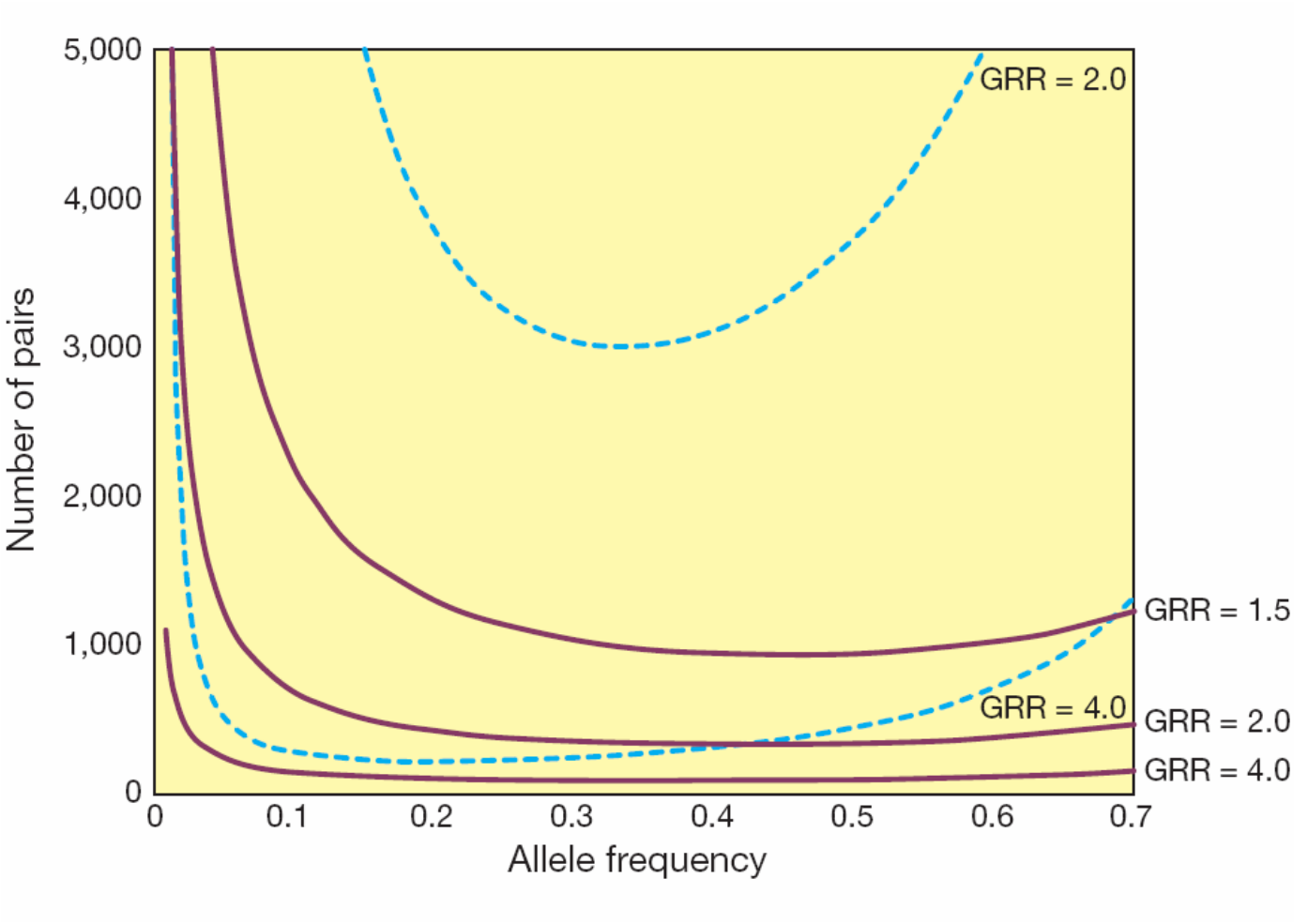


Total ~ 25,000 genes

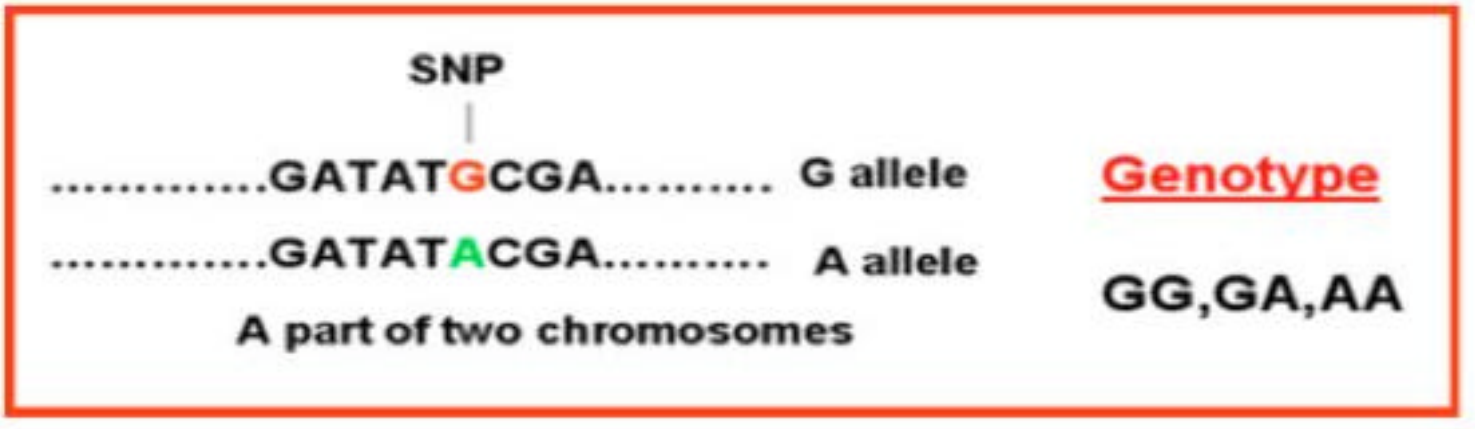
Examples include :

- **Cystic Fibrosis (7q31)**
- **Muscular dystrophy (X)**
- **Parkinson's disease (4q21)**
- **Deafness (about 45 different loci !)**

Linkage Analysis: *Limits*



Sequence Variation : *SNPs*



Most common type of variation, any two chromosomes differ every **600 bp**.

(about **10 million** genome-wide)

- Variation in **single position**
- **bi-allelic** in population
- Less **informative**
- Can be **functional**
- Most useful in **population studies**

Functional consequences of variation

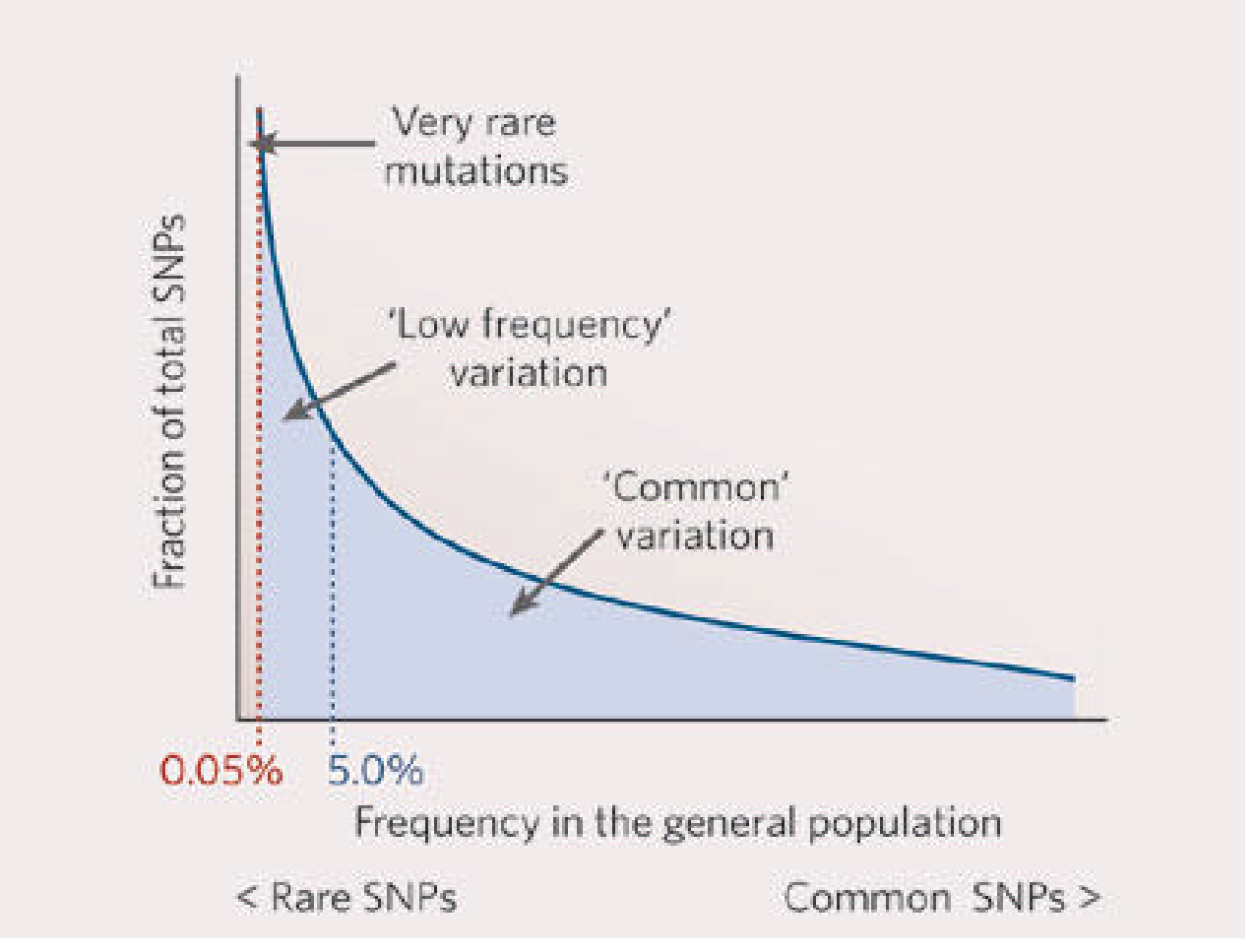
Sequence variation
(SNPs, deletions/duplications, repeats, transposable elements)

Coding variation leading to protein changes

Non coding variation affecting transcription of genes

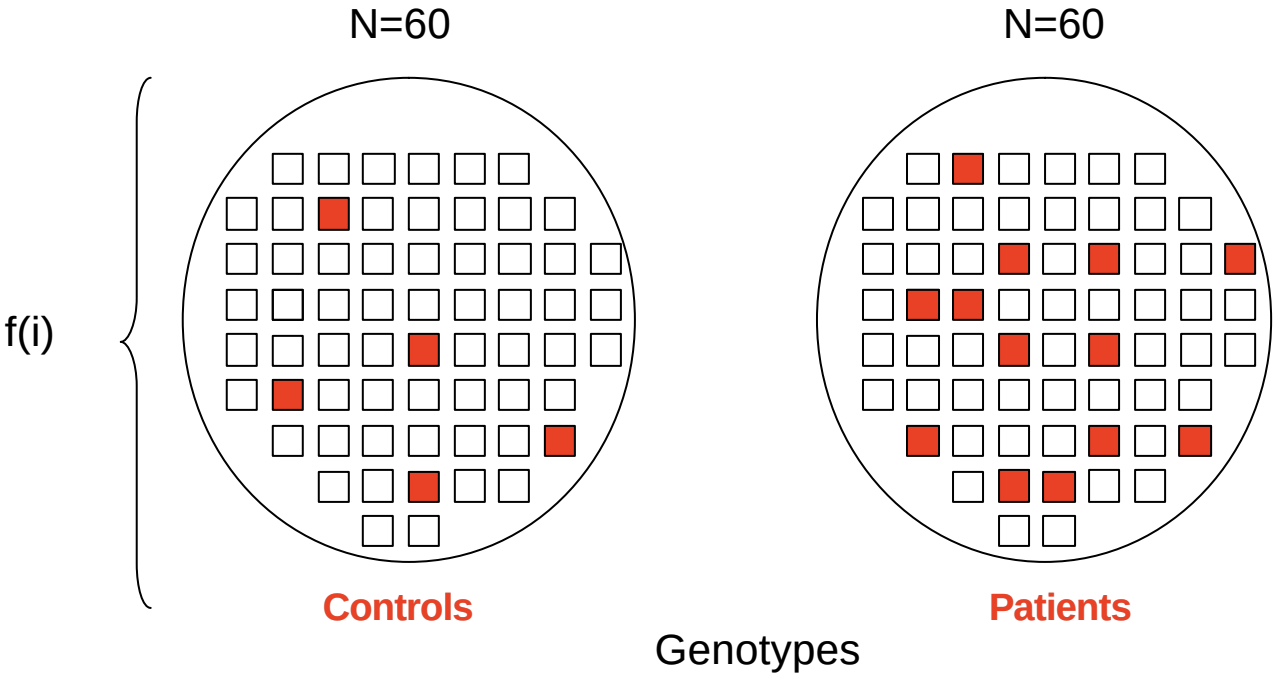
Non coding variation affecting chromatin structure

Sequence Variation : *SNPs*



Population-based association studies

• If and **allele i** in **gene x** is involved in disease pathogenesis, one expects a significant **increase in frequency** in affected groups vs. control.



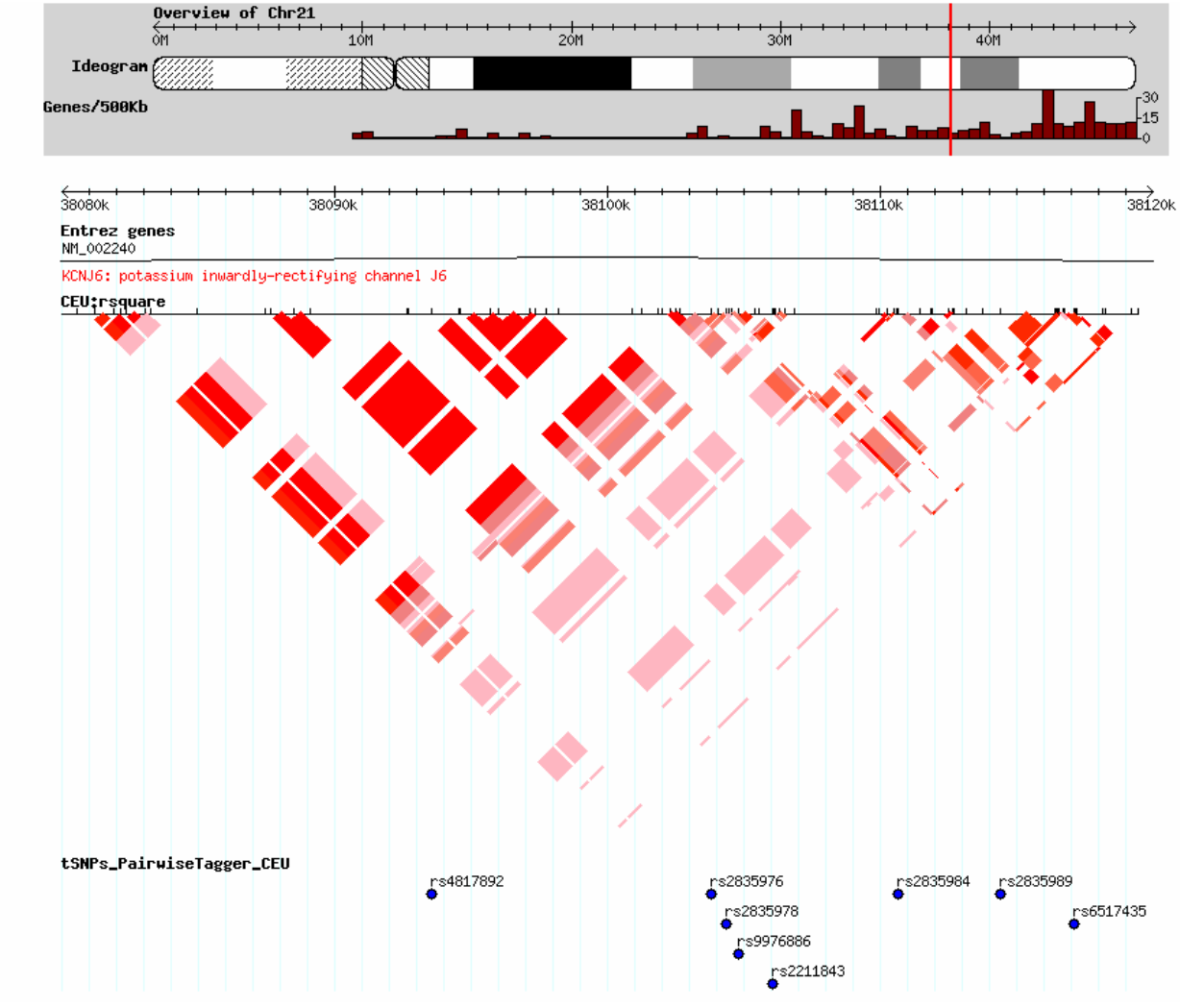
Population-based association studies

Two main approaches :

- **Candidate gene**: limited set of SNPs in set of candidate genes. In general gives a incomplete picture of phenotype determination.

- **Indirect association**: **Genome-wide** set of SNPs, no prior hypothesis, potentially could give a complete view of phenotype determination. Depends on **LD**. Only possible with **important technology advances**.

Association studies: Linkage disequilibrium

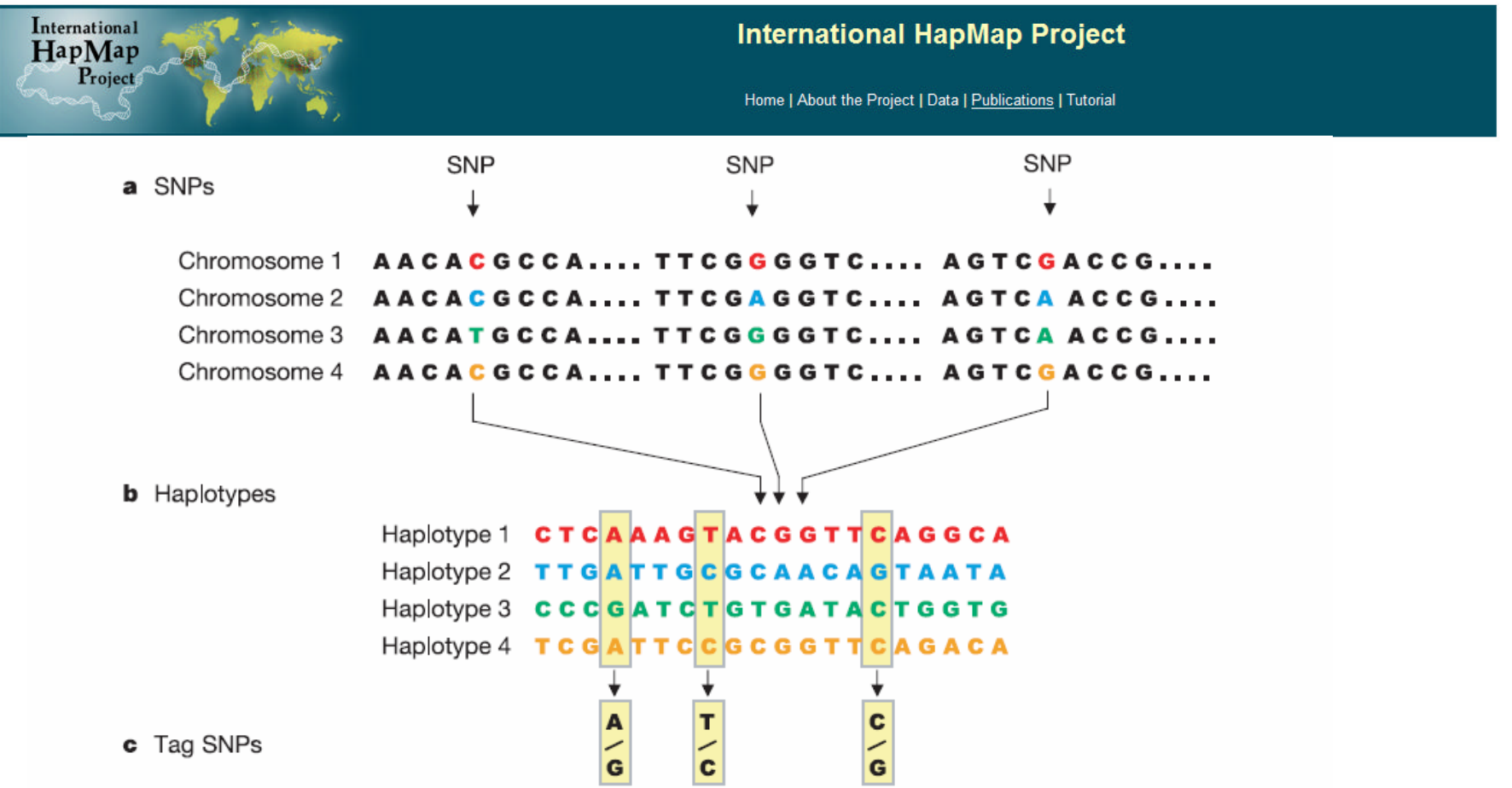


LD can be measured in several ways. For association studies **rsq** (coefficient of determination) is most common

$$r^2 = \frac{[f(AB) - f(A)f(B)]^2}{f(A)f(a)f(B)f(b)}$$

8 **'tag' SNPs** for 50 SNPs in region

Association studies: HapMap project



Ultimate goal: find the **minimal set of SNPs** that capture **most of the sequence variation** information to perform **association studies**.

www.hapmap.org

Association studies: Genotyping technologies

Table 2 | **Selected commercially available high-throughput genotyping platforms**

Company	Method of allele discrimination	Method of detection	Number of assays detected simultaneously
Third Wave	PCR, cleavase	Fluorescence; plate reader	1 (multiplexed 100-fold at PCR stage only)
Sequenom	PCR, primer extension	Mass spectrometry	7–12
ABI	PCR, primer extension	Fluorescence; gel electrophoresis	48
Illumina	Oligo ligation, generic PCR	Fluorescence; tags on beads	1,536
Parallele	Gap closure, generic PCR	Fluorescence; tags on array	10,000
Affymetrix	Generic PCR, hybridization	Fluorescence; hybridization to array	10,000–100,000
Perlegen	PCR, hybridization	Fluorescence; hybridization to array	100,000+

New 300K bead array based on HapMap

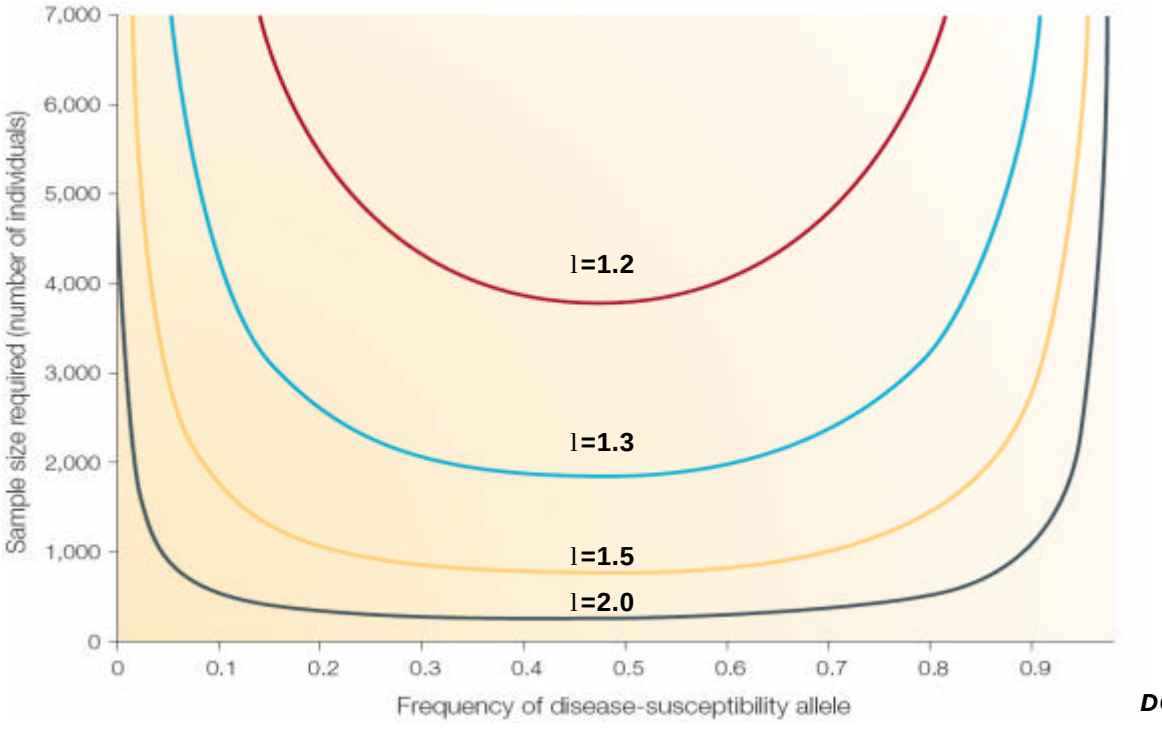
Based on affymetrix array technology

- Ilumina 300K array expected to **capture about 70%** of common variation
- Genome-wide association feasible
- Cost

Hirschhorn and Daly, Nat genet rev 2005

Association studies: Main problems

- Many studies **underpowered**. For diseases with **complex inheritance** (λs< 20)and many loci with **minor contributions** (each allele with GRR< 3.0) **1000s rather than 100s of samples needed** !



DG Clayton, JA Todd et al 2005

- How to deal with **multiple testing** problem ?

- Need new methods to extract G x G and G x E **interactions** !

Targeted drugs in the near future ?

U.S. Food and Drug Administration

Department of Health and Human Services

FDA Home Page | Search FDA Site | FDA A-Z Index | Contact FDA

FDA News

FOR IMMEDIATE RELEASE
P05-32
June 23, 2005

Media Inquiries:
Laura Alvey, 301-827-6242
Consumer Inquiries:
888-INFO-FDA

FDA Approves BiDil Heart Failure Drug for Black Patients

The Food and Drug Administration (FDA) approved BiDil (bye-DILL), a drug for the treatment of heart failure in self-identified black patients, representing a step toward the promise of personalized medicine.

Heart failure is a condition in which the heart is weakened and does not pump enough blood. It can be caused by a variety of damage to the heart, including heart attacks, high blood pressure, and infections.

The approval of BiDil was based in part on the results of the African-American Heart Failure Trial (A-HeFT). The study, which involved 1,050 self-identified black patients with severe heart failure who had already been treated with the best available therapy, was conducted because two previous trials in the general population of severe heart failure patients found no benefit, but suggested a benefit of BiDil in black

Mapping genetic susceptibility to HIV infection

Collaborative study between the labs of **S. Antonarakis**, **A. Telenti** (Corinne Loeuillet) and **J. Beckmann**

Susceptibility to HIV: Genetics role ?

- Large difference in natural history of disease, two interesting groups:
  - **Exposed non infected**
  - **Infected non progressors (rare, Familial segregation)**
- Highly concordant susceptibility in **twins**
- Several known **polymorphisms** known to play a role.

Susceptibility to HIV: known genetic factors

Box 1

The contribution of chemokine or chemokine receptor polymorphisms to HIV-1/AIDS susceptibility

CCR5 Δ32: CCR5 Δ32 homozygotes are resistant to HIV infection; heterozygotes show slower disease progression.

Other *CCR5* polymorphisms: numerous polymorphisms in *CCR5*, particularly promoter regions, affect CCR5 expression and the rate of progression to AIDS.

CCR2–64I polymorphism: associated with slower progression to AIDS.

Viral Co-receptors

CX_3CR1: rapid progression to AIDS in HIV-1-infected individuals who are homozygous for a variant of CX_3CR1. Two amino acid changes result in markedly impaired binding of CX_3CR1 to its ligand CX_3CL1.

Chemotactic molecules

CXCL12 (SDF-1): individuals who are homozygous for *SDF1–3'A* show a delayed onset of AIDS.

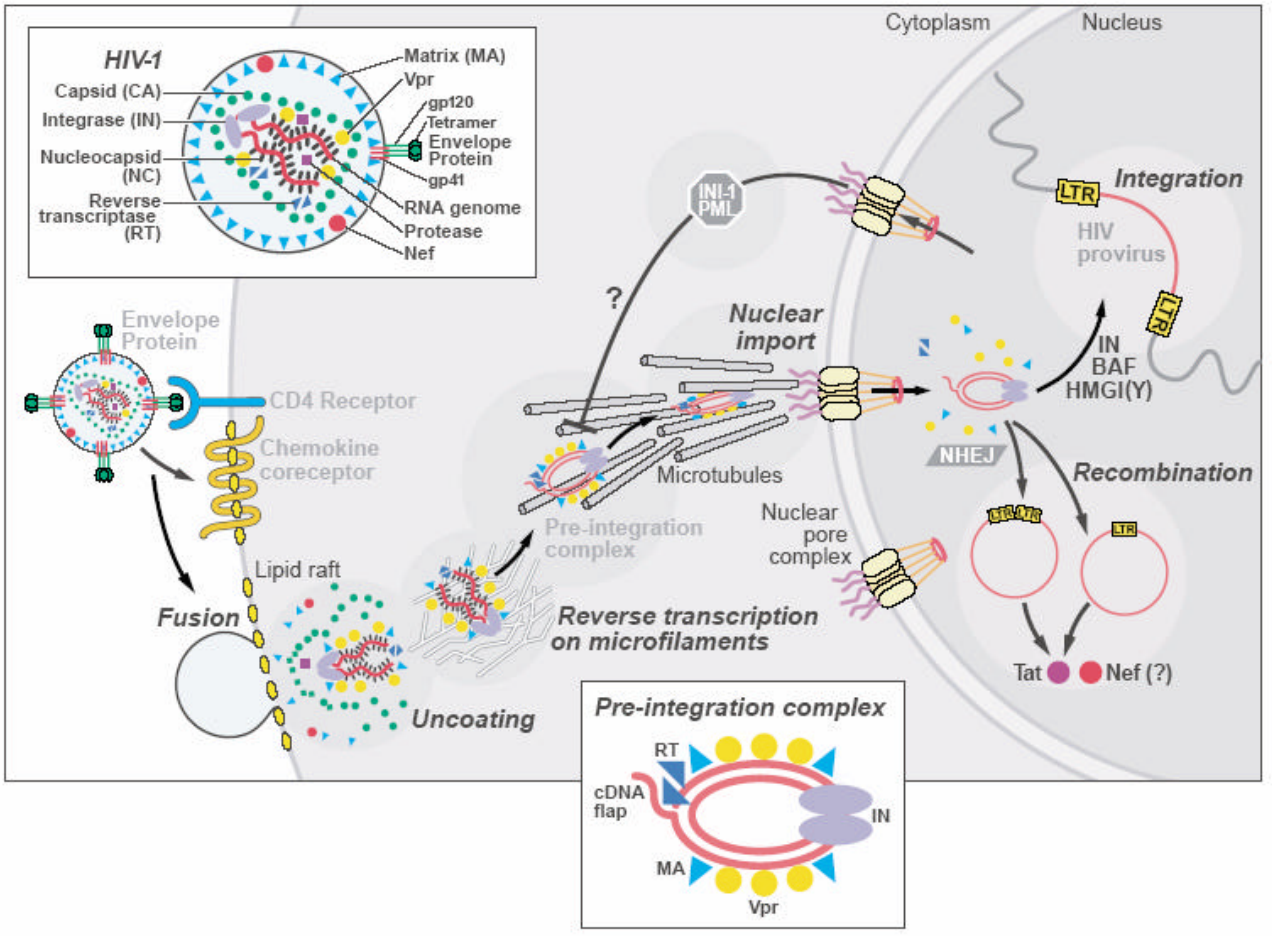
CCL2 (MCP-1): the MCP-1 –2578G allele is associated with a 50% reduction in the risk of acquiring HIV-1.

CCL5 (RANTES): the In1.1C allele is associated with a decreased expression of CCL5 and rapid progression to AIDS.

CCL3L1: low *CCL3L1* gene copy numbers, relative to the ethnic population average, is associated with markedly enhanced HIV-1 susceptibility and progression to AIDS.

Co-receptor ligands

Susceptibility to HIV: viral life cycle



Susceptibility to HIV: cellular system

Main aim :

- Develop **cellular system** in which to dissect **genetic** factors

Validation:

- Can an *in-vitro* cellular system **re-capitulate** *in-vivo* situation ?
- Would such a system be **reproducible** ?

An in vitro system for identification of lentiviral susceptibilit

SYSTEM:

- **Cell transduction of b-lymphoblastoid cells**
- **VSV-G pseudotyped lentiviral vector, expression of eGFP (CMVpromoter)**
- **infection by spinoculation (3000rpm, 3h) wash, detection of eGFP expression by FACS (72h)**

Susceptibility to HIV: Genetic analysis

15 CEPH families
=
~200 individuals

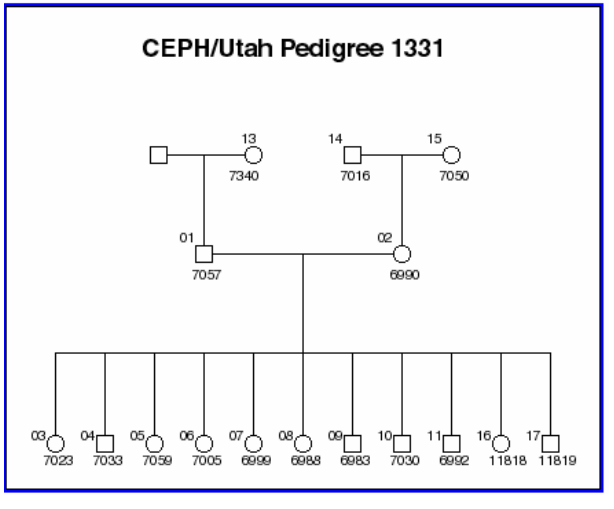
Measured **cellular phenotypes** in **triplicate**

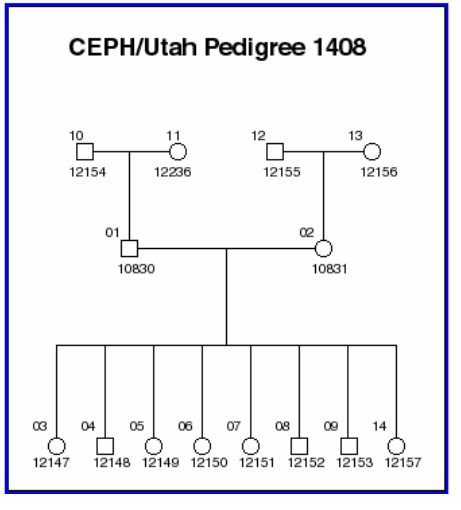
Obtained information for 2600 **SNPs** genomewide - publicly available in DBs.

CEPH families

CEPH : Centre d'Etude du Polymorphisme Humain

- Created in 1984 to provide resources for **human genome mapping**
- We used **15 families** (N=200)





Susceptibility to HIV: Genetic analysis

Heritability:

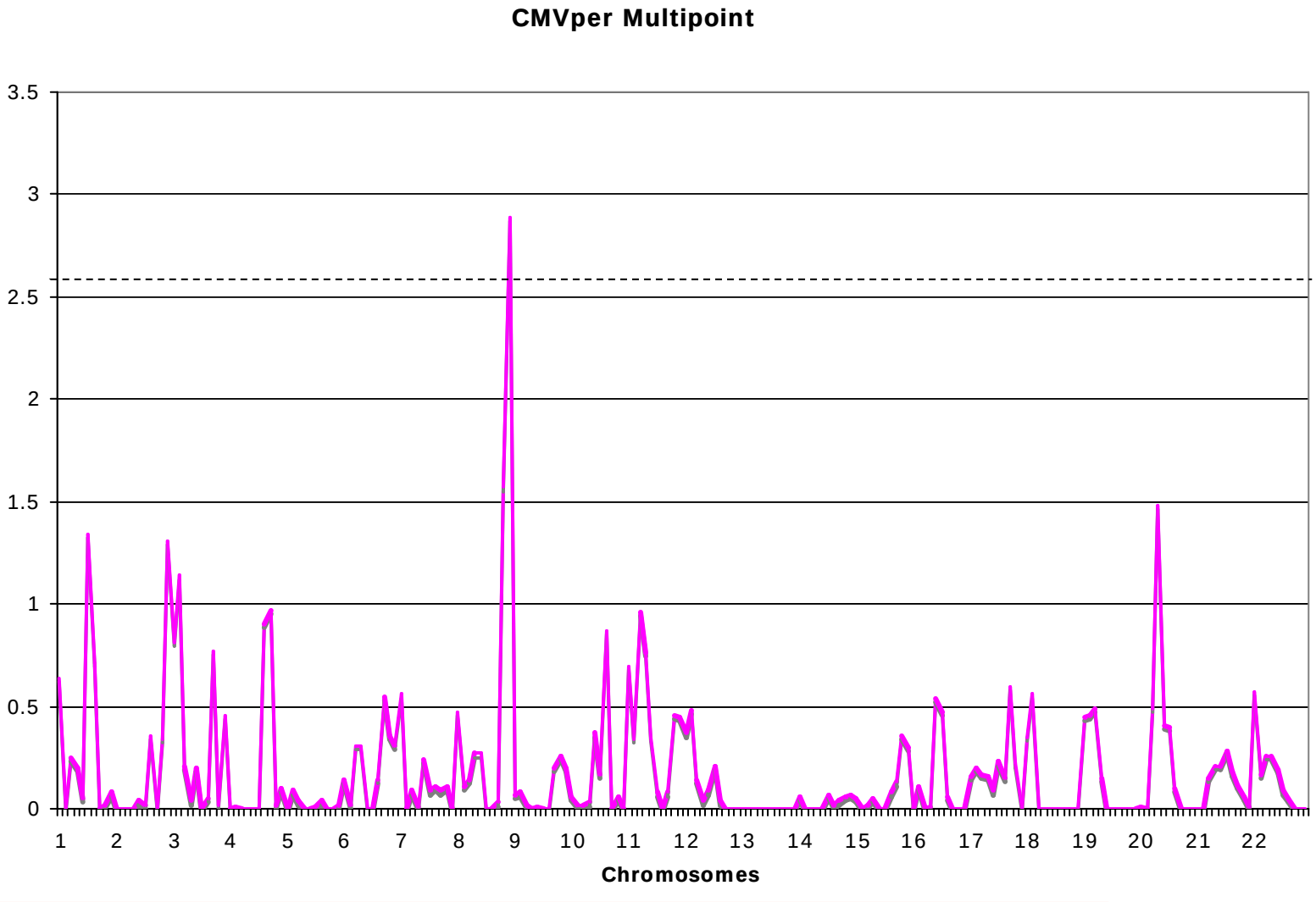
	Trait	H2r	p value
1	CMVGFPper	0.5367977	0.0000016
2	CMVGFPMFIall	0.4354729	0.0000087
10	CD39per	0.8036432	0.0003233
11	CD39MFIall	1	8.53E-68
12	CD39ratio	1	6.27E-55
13	LMP1per	0.496049	3.47E-10
14	LMP1MFIall	0.732135	1.24E-14
15	LMP1ratio	0.6194324	2.01E-17
16	CD11aMFIall	0.996293	0.0000185
17	CD11aper	0.1366545	0.1273509
18	CD11aratio	1	0.0002416
19	CD19MFIall	0.9050112	3.48E-13
20	CD19per	0.8286046	0.0000001
21	CD19ratio	1	0.0000579
22	CD21MFIall	1	0.000226
23	CD21per	1	2.41E-10
24	CD21ratio	0.4659657	0.0136537
25	CD23MFIall	0.6998939	0.0000449
26	CD23per	0.7914027	0.005566
27	CD23ratio	0.517473	0.0006306

Susceptibility

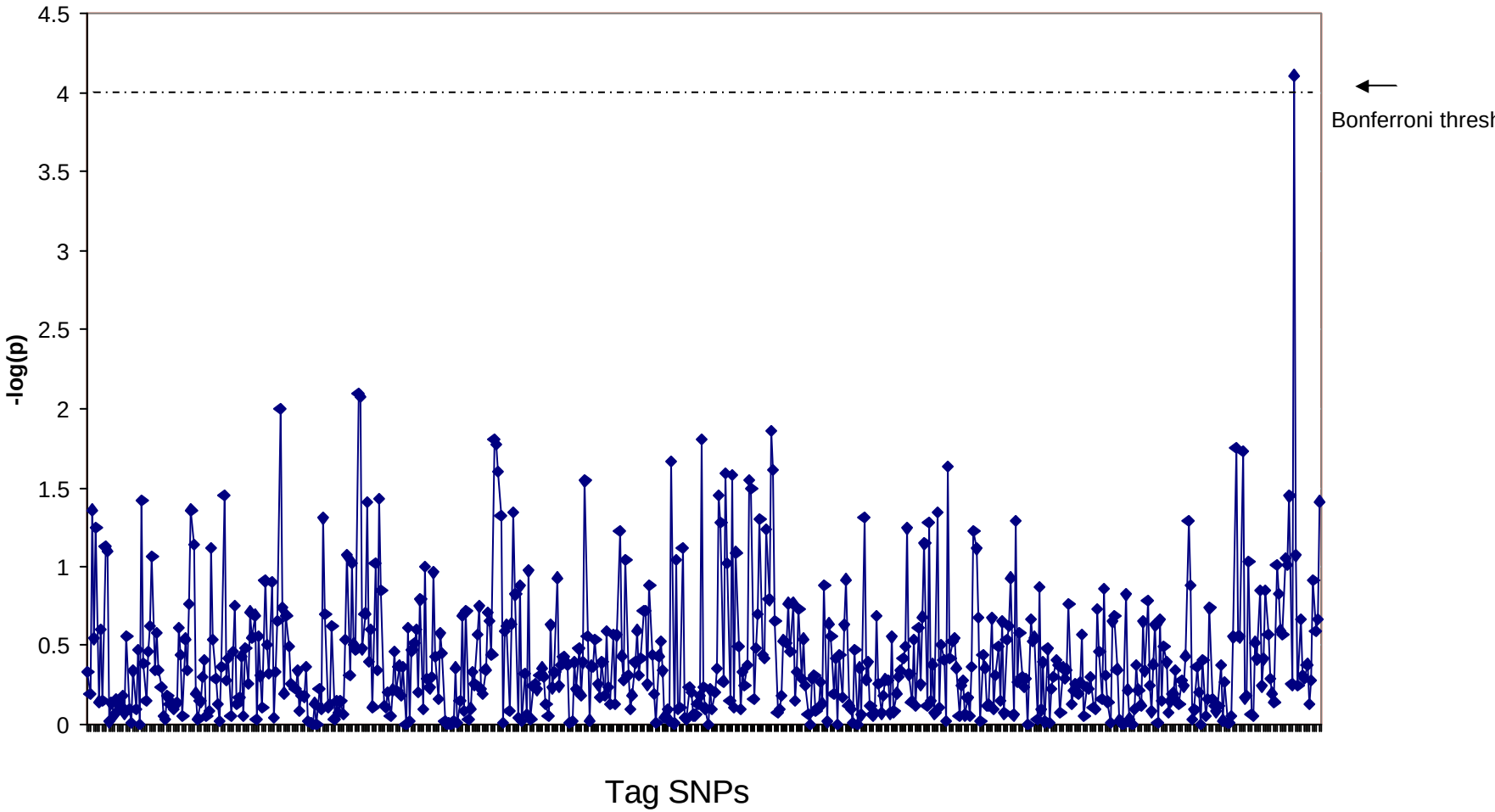
EBV marker

Innate immunity

Susceptibility to HIV: Linkage Results I



Susceptibility to HIV: Association using HapMap



Chromosome 8 CMVper association: Tag SNPs 3Mb centered on linkage finding

Trait distribution according to phenotype

Analysis of Variance for CMV GFPper

Source	DF	SS	MS	F	P
SNP	1	932.4	932.4	18.30	0.000
Error	53	2699.7	50.9		
Total	54	3632.0			

Level	N	Mean	StDev
AG	7	35.934	8.800
GG	48	23.580	6.896

Pooled StDev = 7.137

Individual 95% CIs For Mean
Based on Pooled StDev

```
-----+---------+---------+---------+-
                   (--------*--------)
 (--*---)
-----+---------+---------+---------+-
  24.0      30.0      36.0      42.0
```

Chromosome 8 CMVper association: Fine mapping using all HapMap phase 2.0 data