# 23rd International Workshop on Statistical Genetics and Methodology of Twin and Family Studies: Advanced course

- Pak Sham (director)
- Lindon Eaves
- Jeff Barrett
- David Evans
- Goncalo Abecasis
- Mike Neale
- Hermine Maes
- Sarah Medland
- Dorret Boomsma
- Danielle Posthuma
- Meike Bartels
- John Hewitt (host)
- Jeff Lessem
- Marleen de Moor
- Stacey Cherny
- Ben Neale
- Shaun Purcell
- Manuel Ferreira
- Nick Martin
- Kate Morley
- Allan McRae

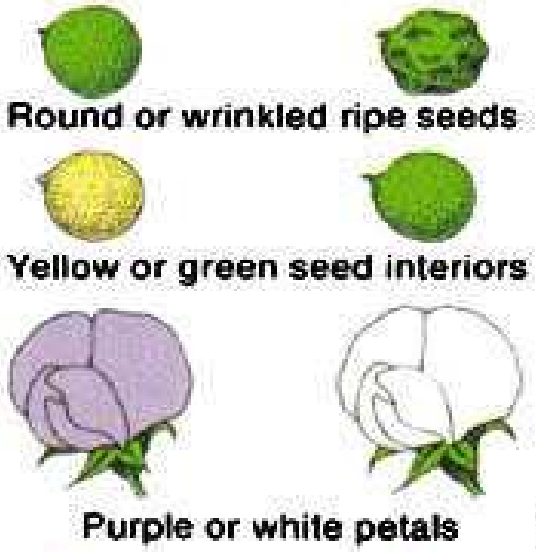
# Hunting QTLs – an introduction

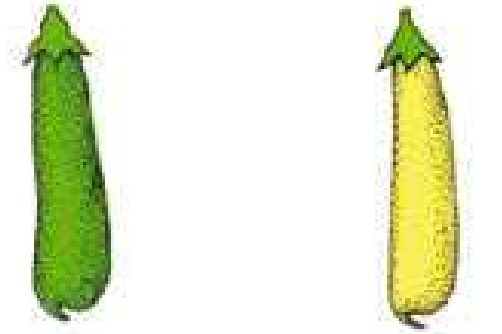
Nick Martin

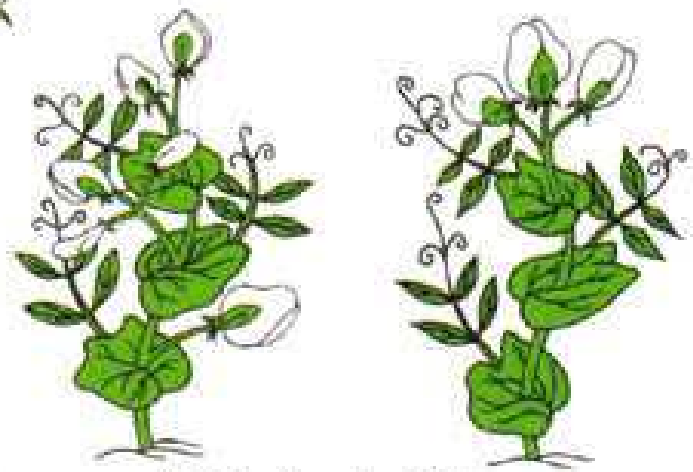
Queensland Institute of Medical Research

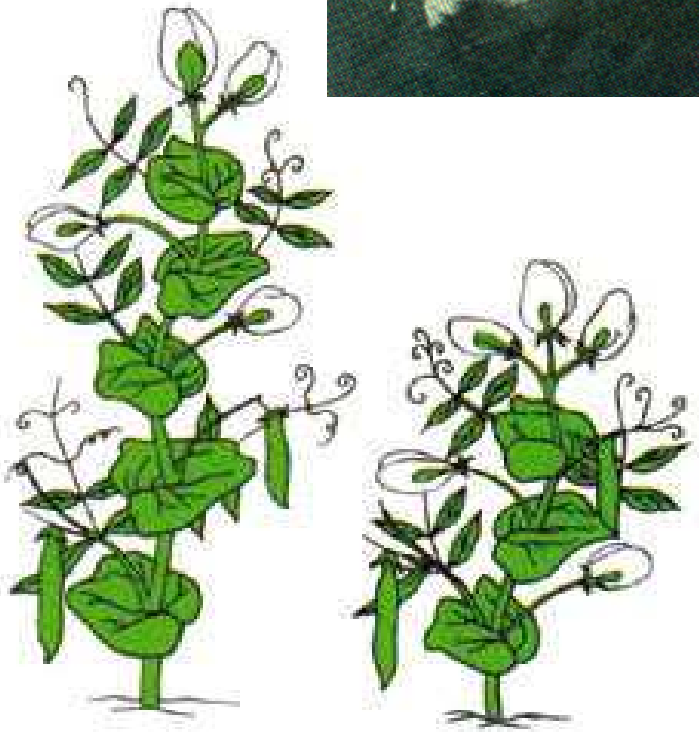
Boulder workshop: March 2, 2009

# Mendel 1865 – genetics of discrete traits

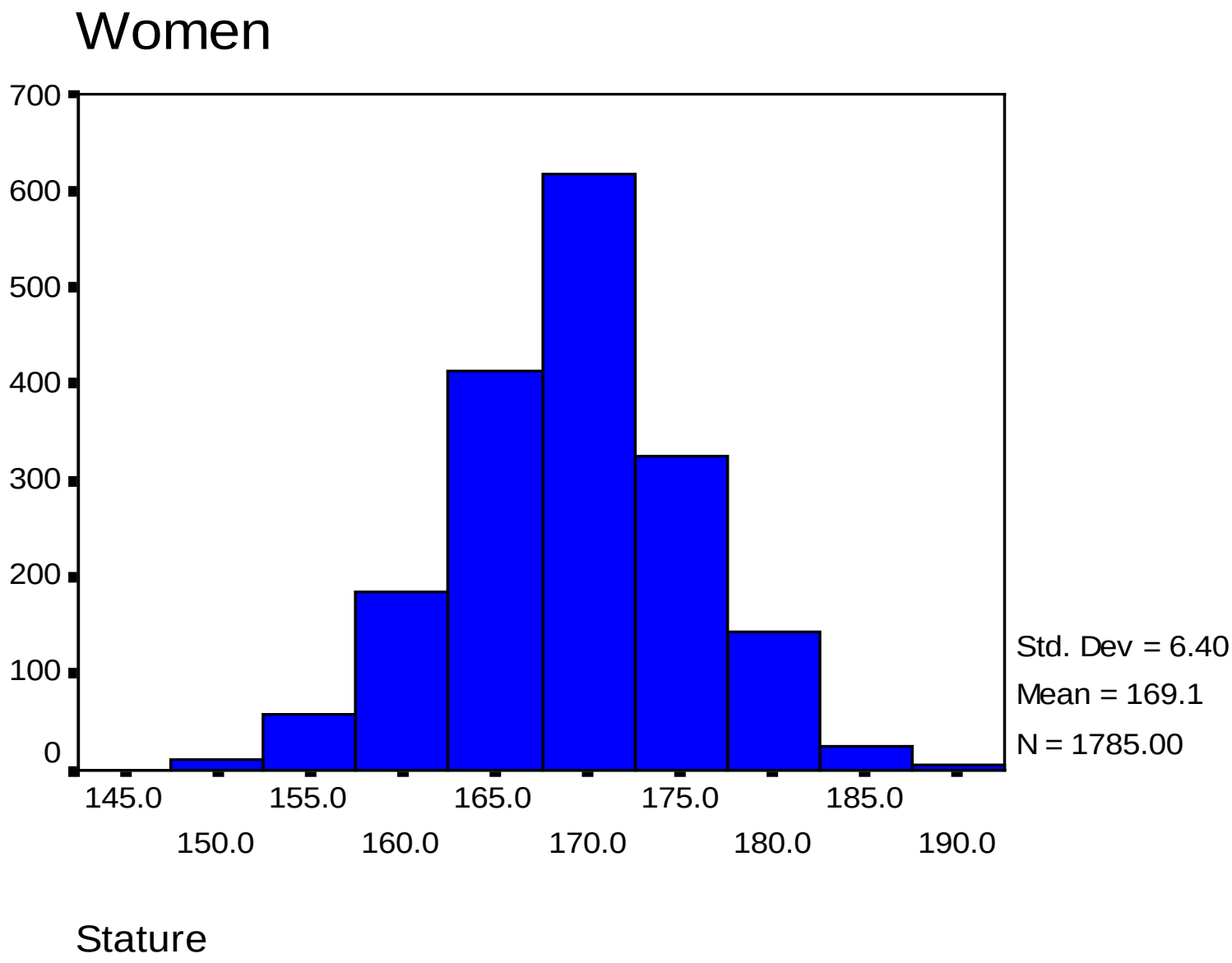


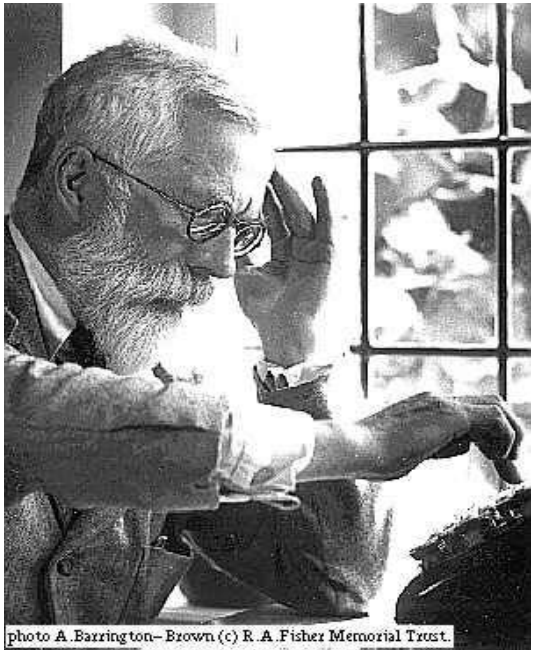



# Stature in adolescent twins

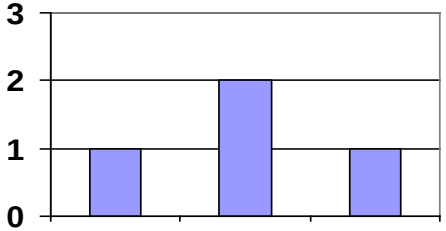


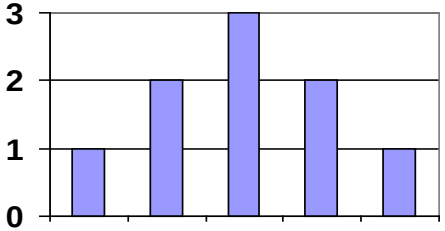
# R.A. Fisher, 1918

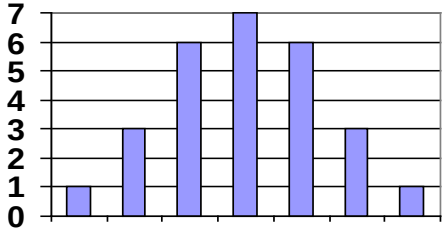
## The explanation of quantitative inheritance in Mendelian terms

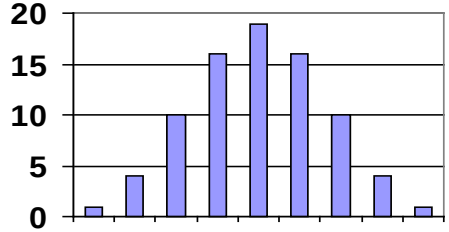



| 1 Gene | 2 Genes | 3 Genes | 4 Genes |
|---|---|---|---|
| → 3 Genotypes | → 9 Genotypes | → 27 Genotypes | → 81 Genotypes |
| → 3 Phenotypes | → 5 Phenotypes | → 7 Phenotypes | → 9 Phenotypes |

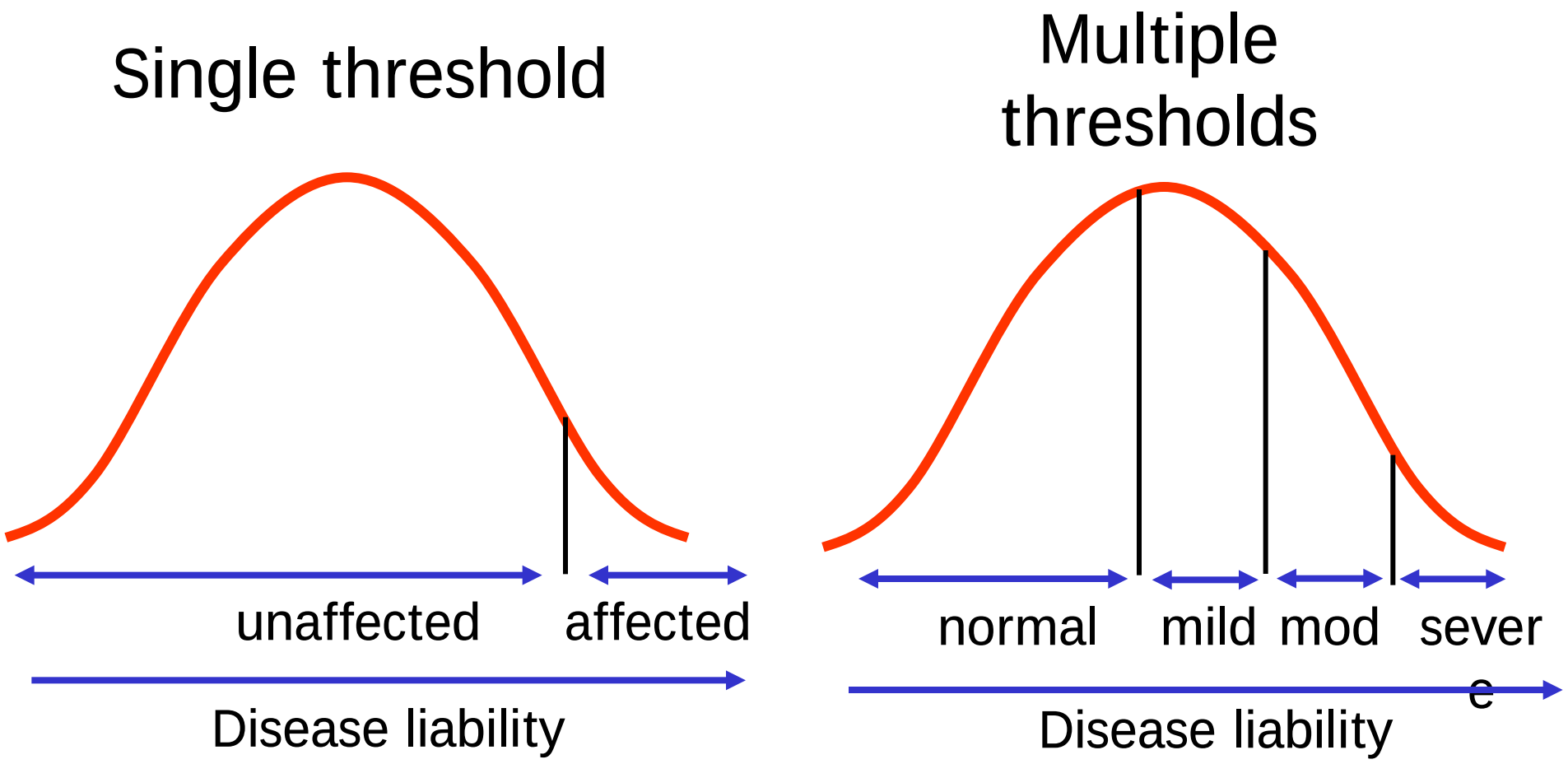


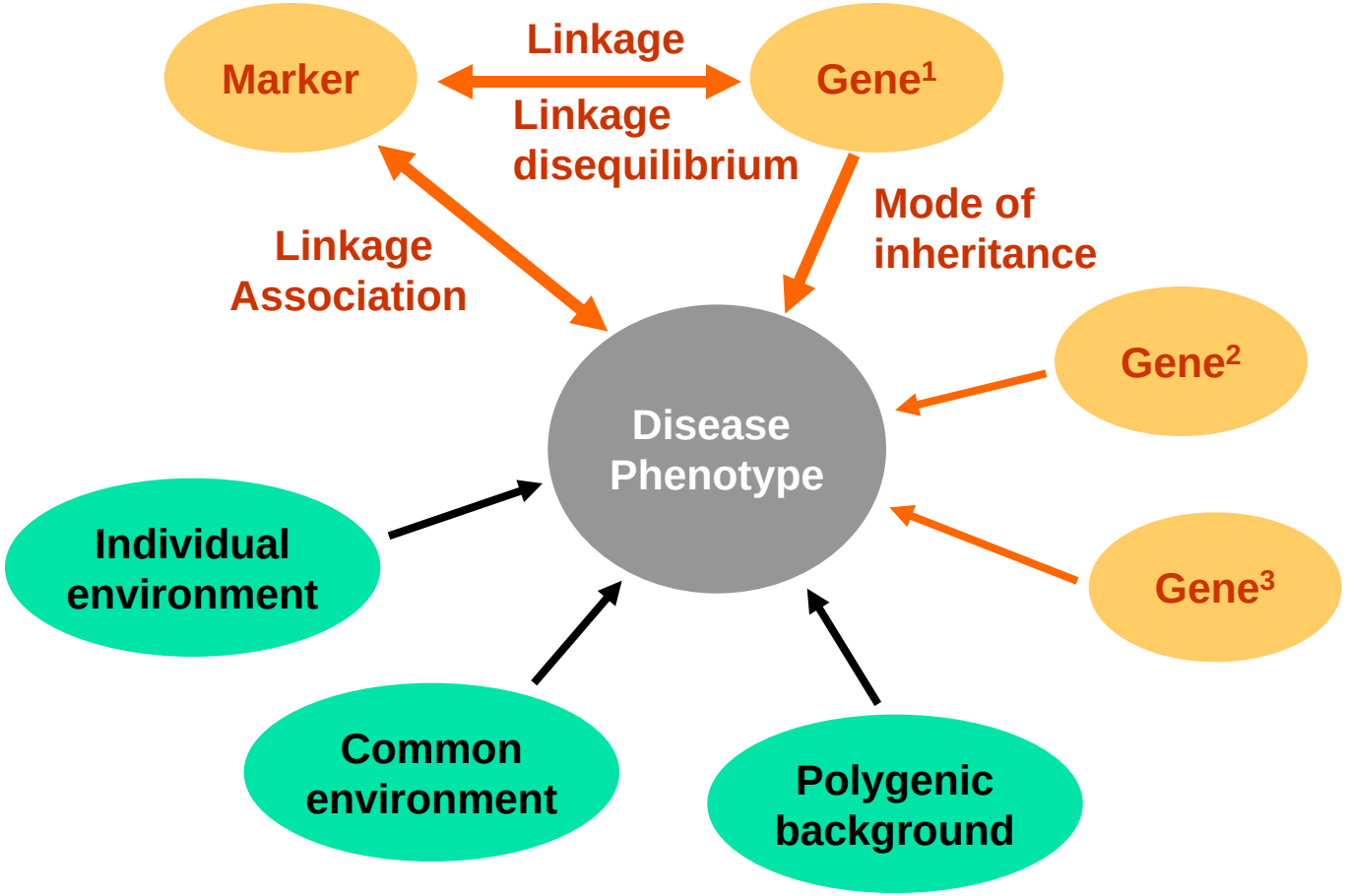



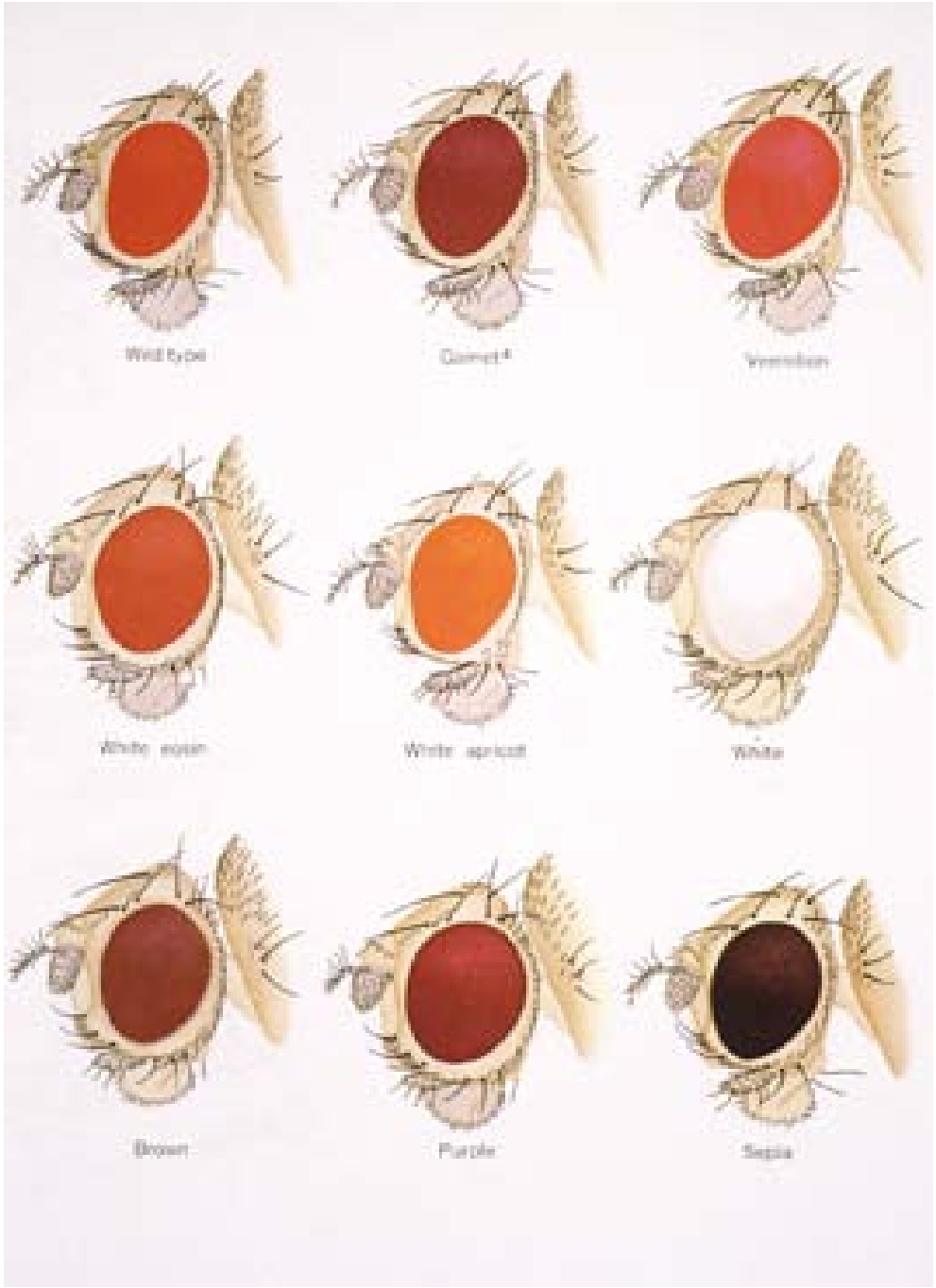



# Multifactorial Threshold Model of Disease



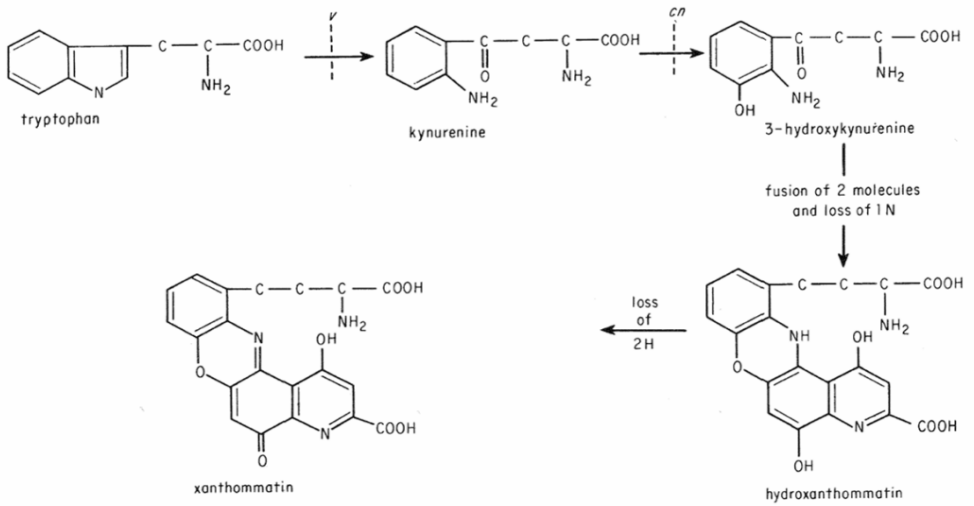
# Complex Trait Model






# Using genetics to dissect metabolic pathways: Drosophila eye color

## Beadle & Ephrussi, 1936



# Finding QTLs

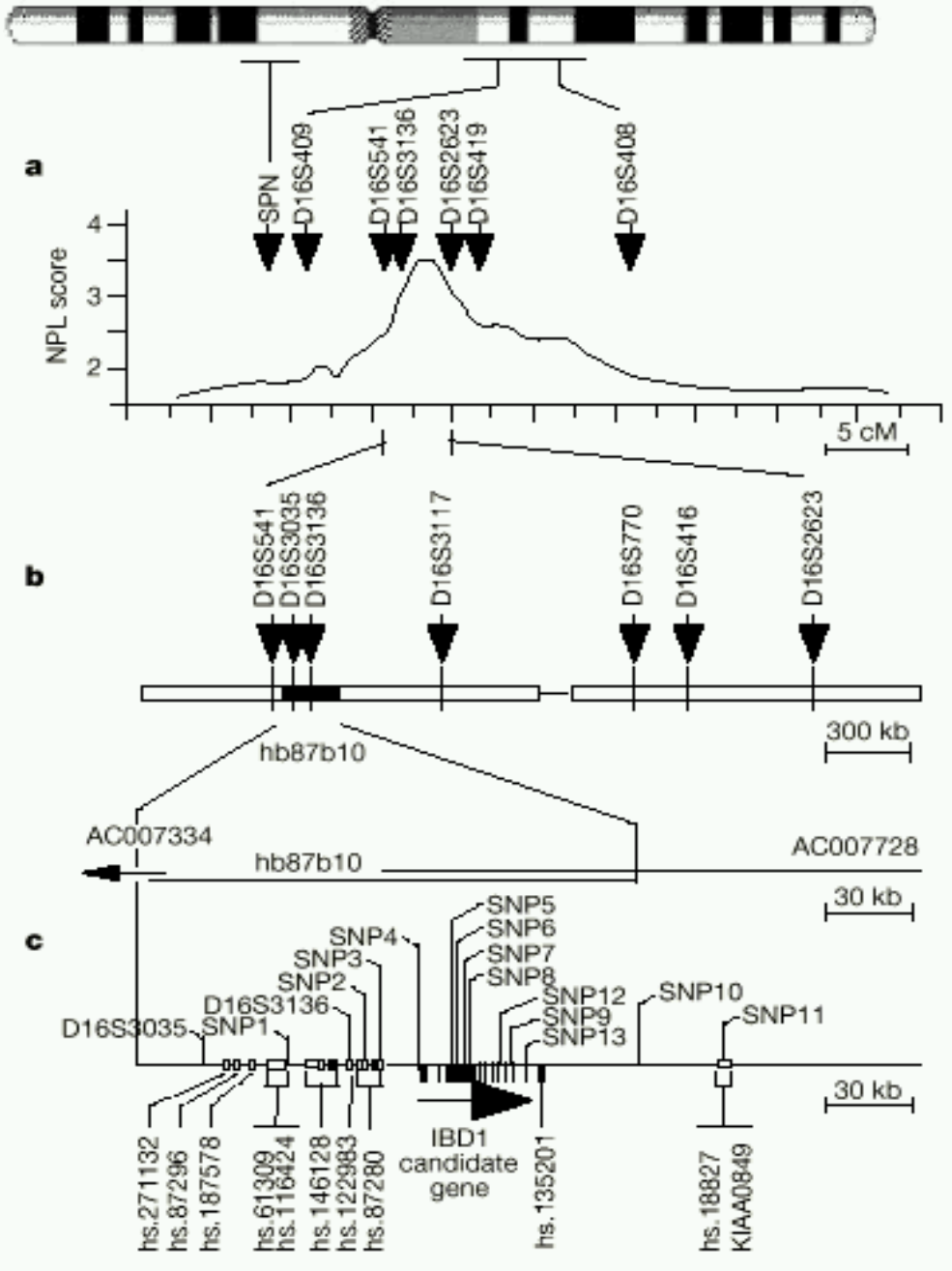
- Linkage
- Association

# Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

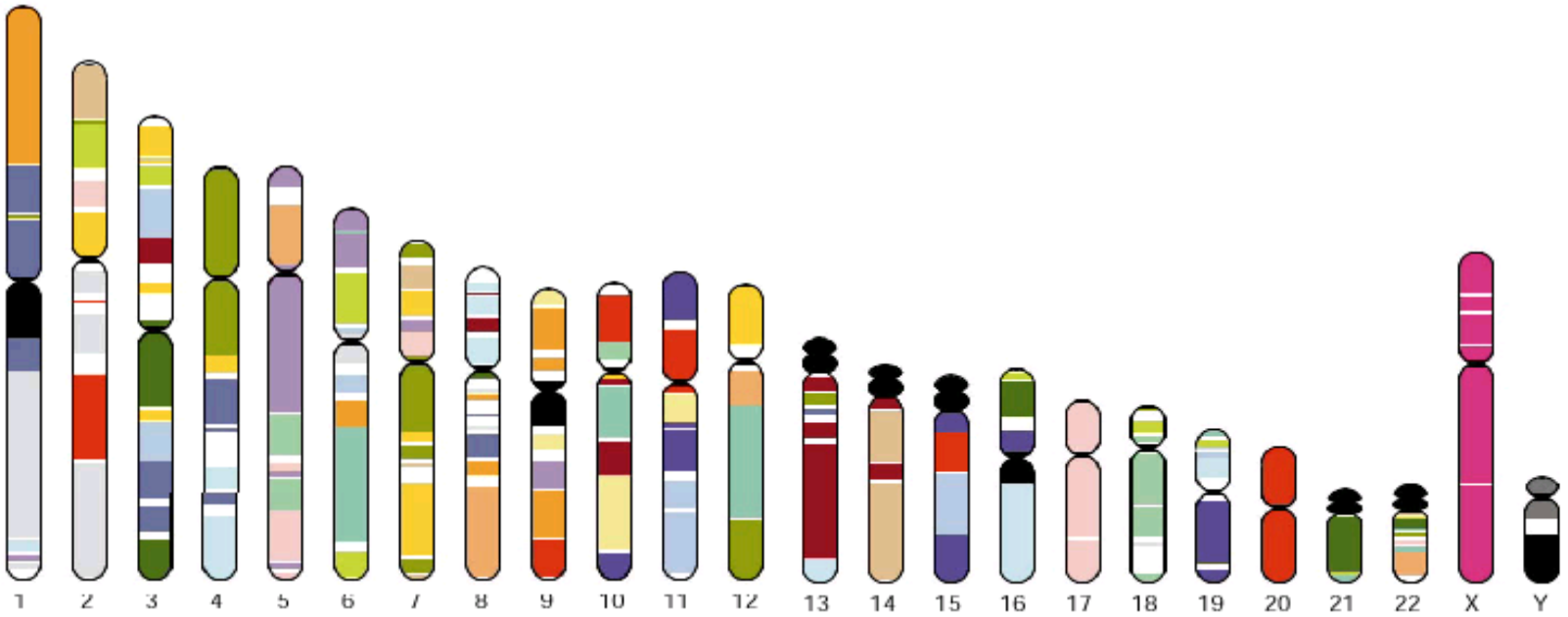
Jean-Pierre Hugot*†‡, Mathias Chamaillard*†, Habib Zouali*, Suzanne Lesage*, Jean-Pierre Cézard‡, Jacques Belaiche§, Sven Almer‖, Curt Tysk¶, Colm A. O'Morain#, Miquel Gassull☆, Vibeke Binder**, Yigael Finkel††, Antoine Cortot‡‡, Robert Modigliani§§, Pierre Laurent-Puig†, Corine Gower-Rousseau‡‡, Jeanne Macry‖‖, Jean-Frédéric Colombel‡‡, Mourad Sahbatou* & Gilles Thomas*†¶¶

NATURE | VOL 411 | 31 MAY 2001

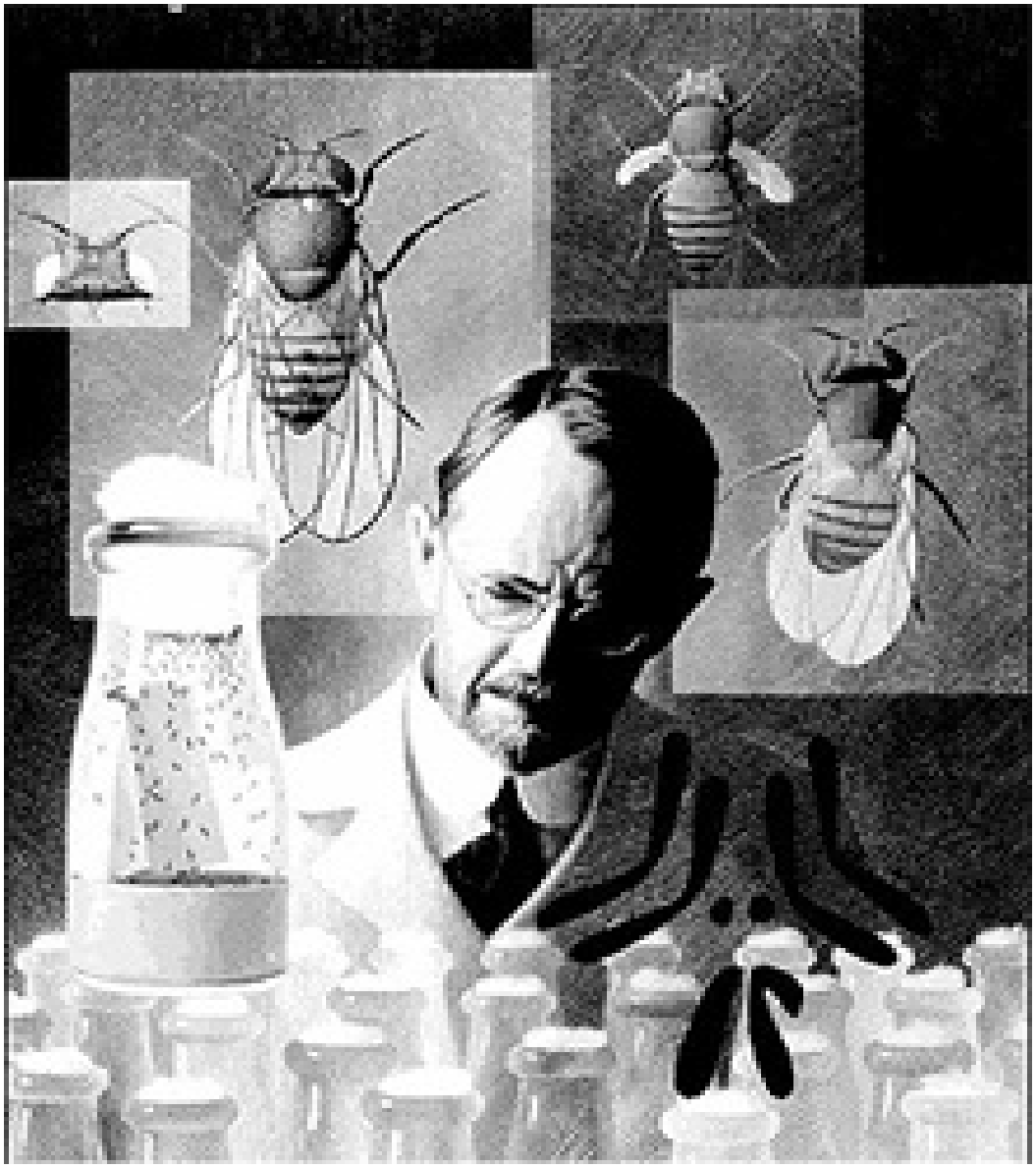
First (unequivocal) positional cloning of a complex disease QTL !



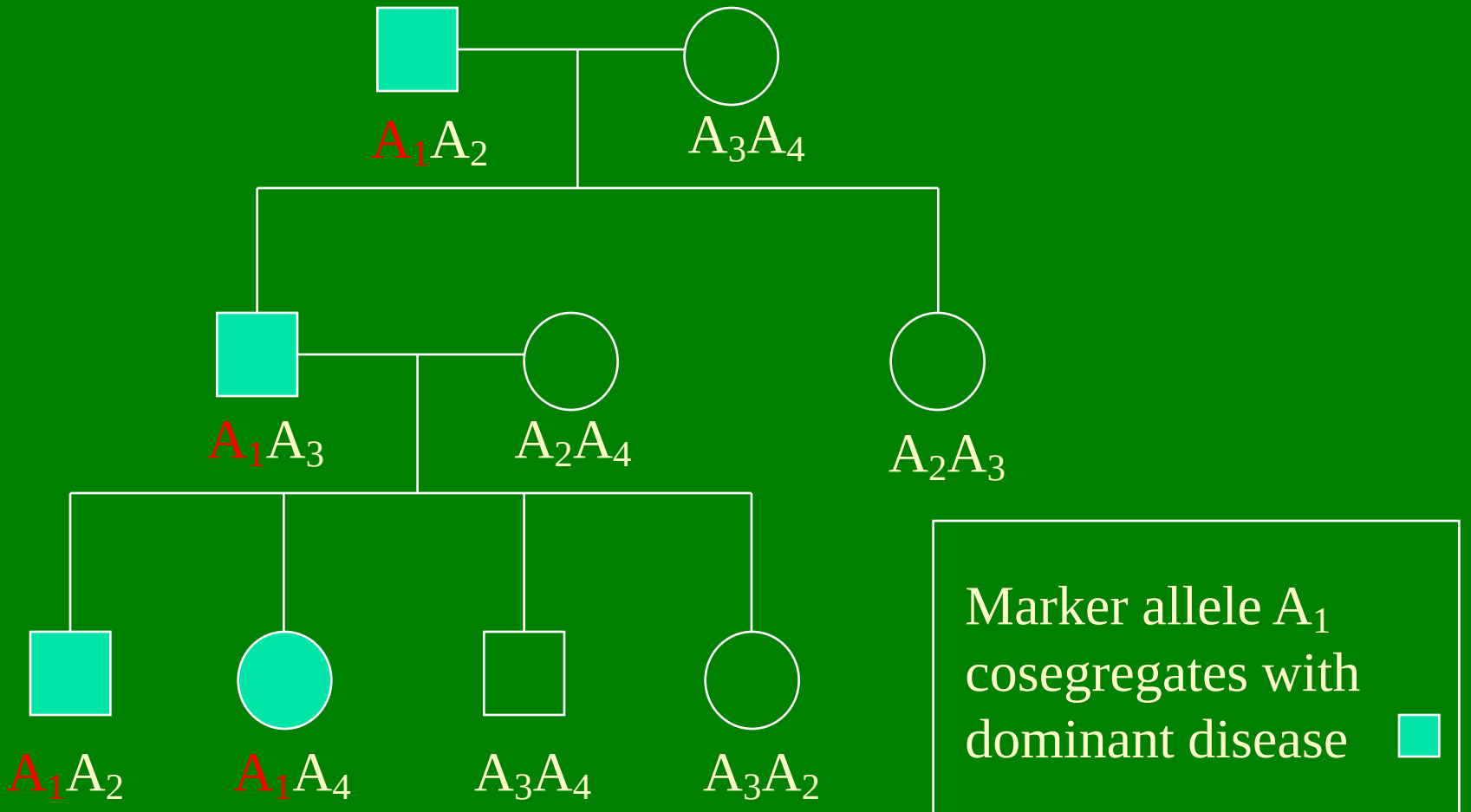
# Linkage analysis



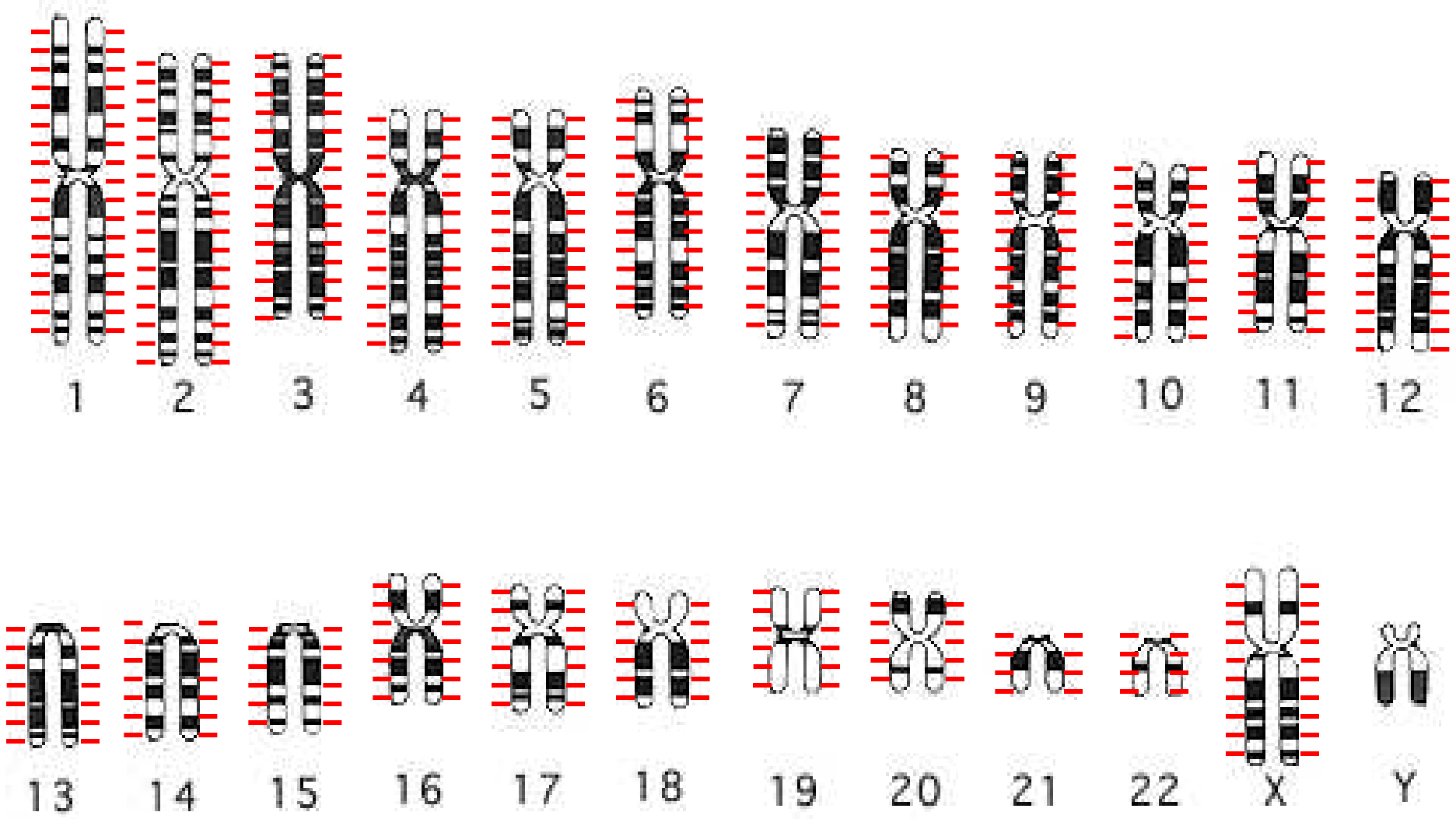
# Thomas Hunt Morgan – discoverer of linkage

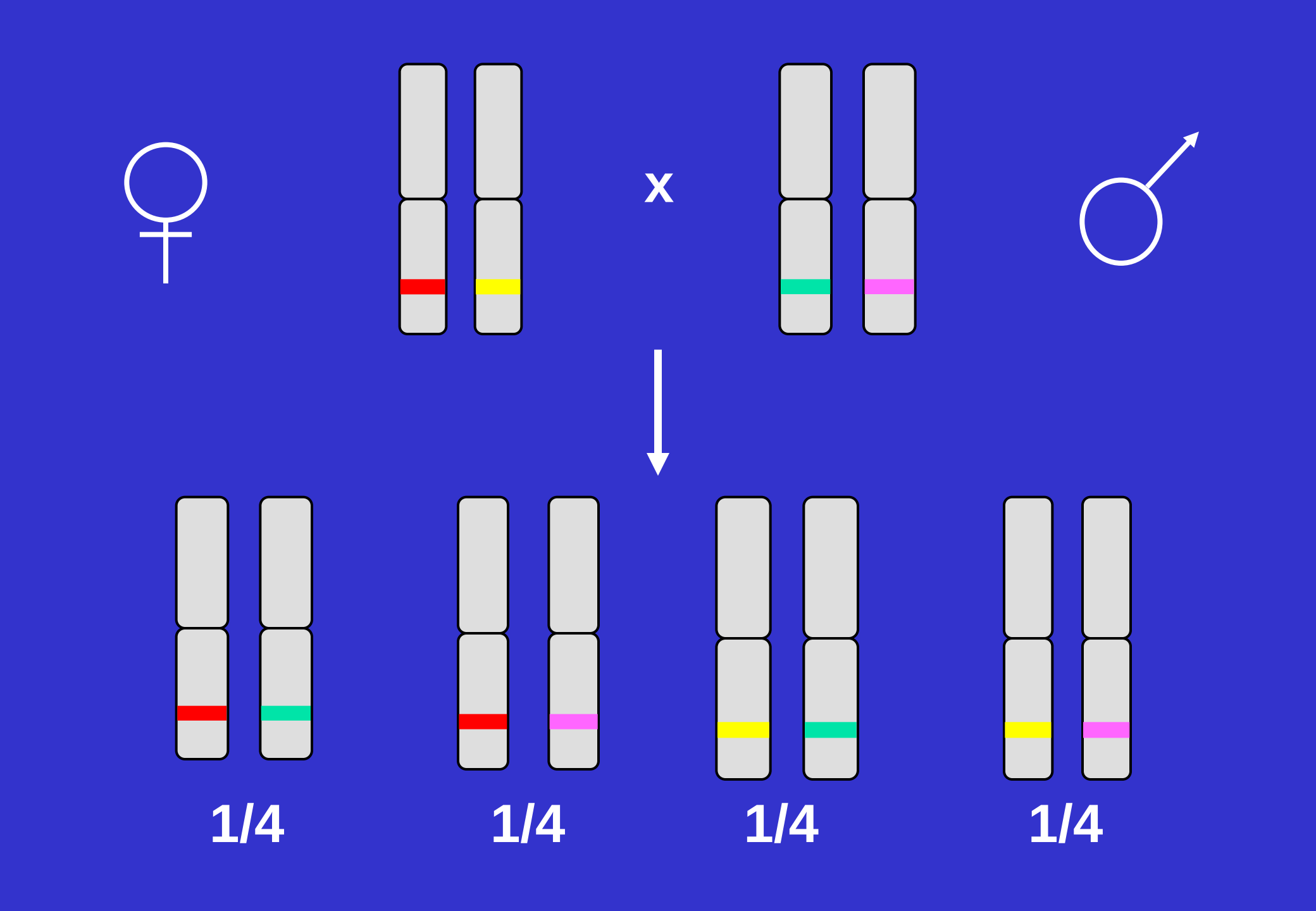


# Linkage = Co-segregation

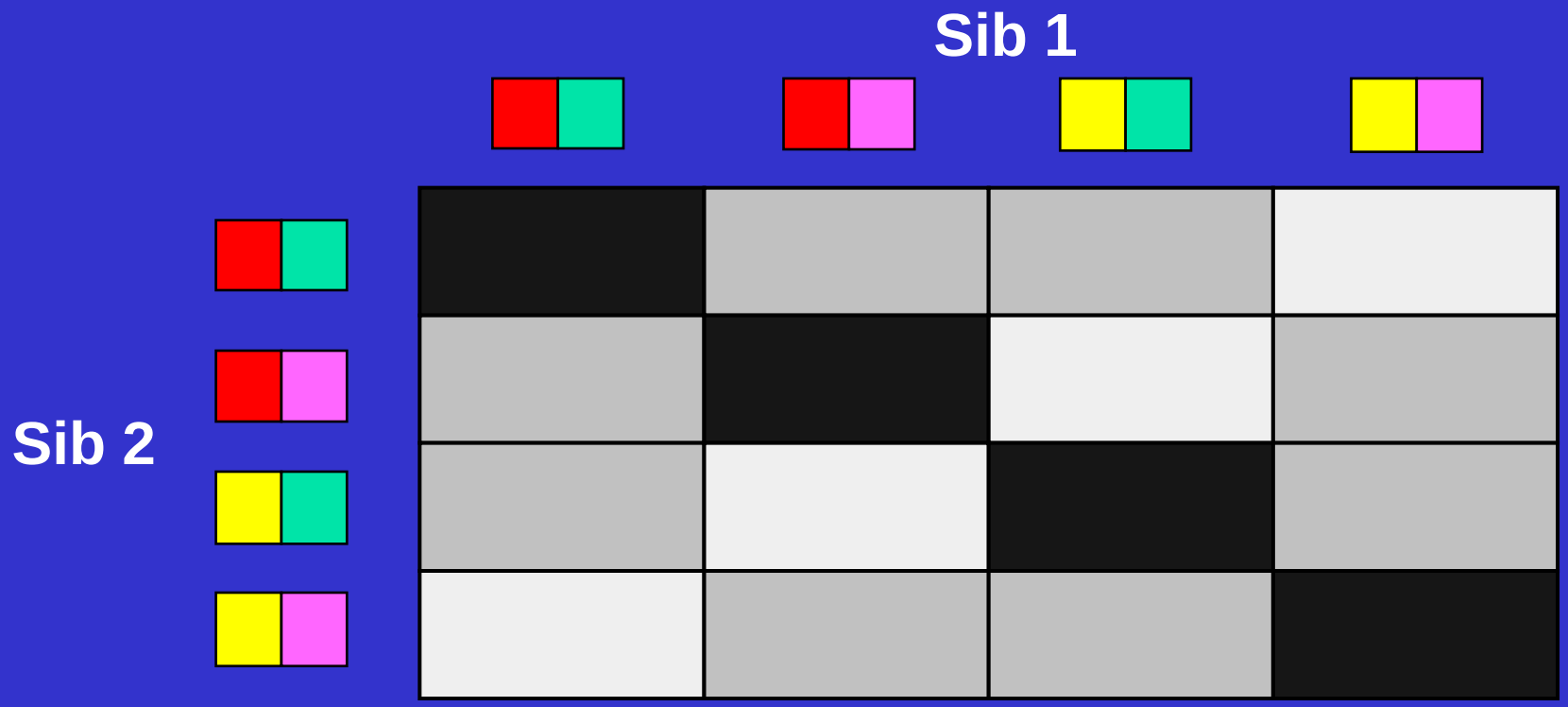


# Linkage Markers...




x
1/4
1/4
1/4
1/4

# IDENTITY BY DESCENT



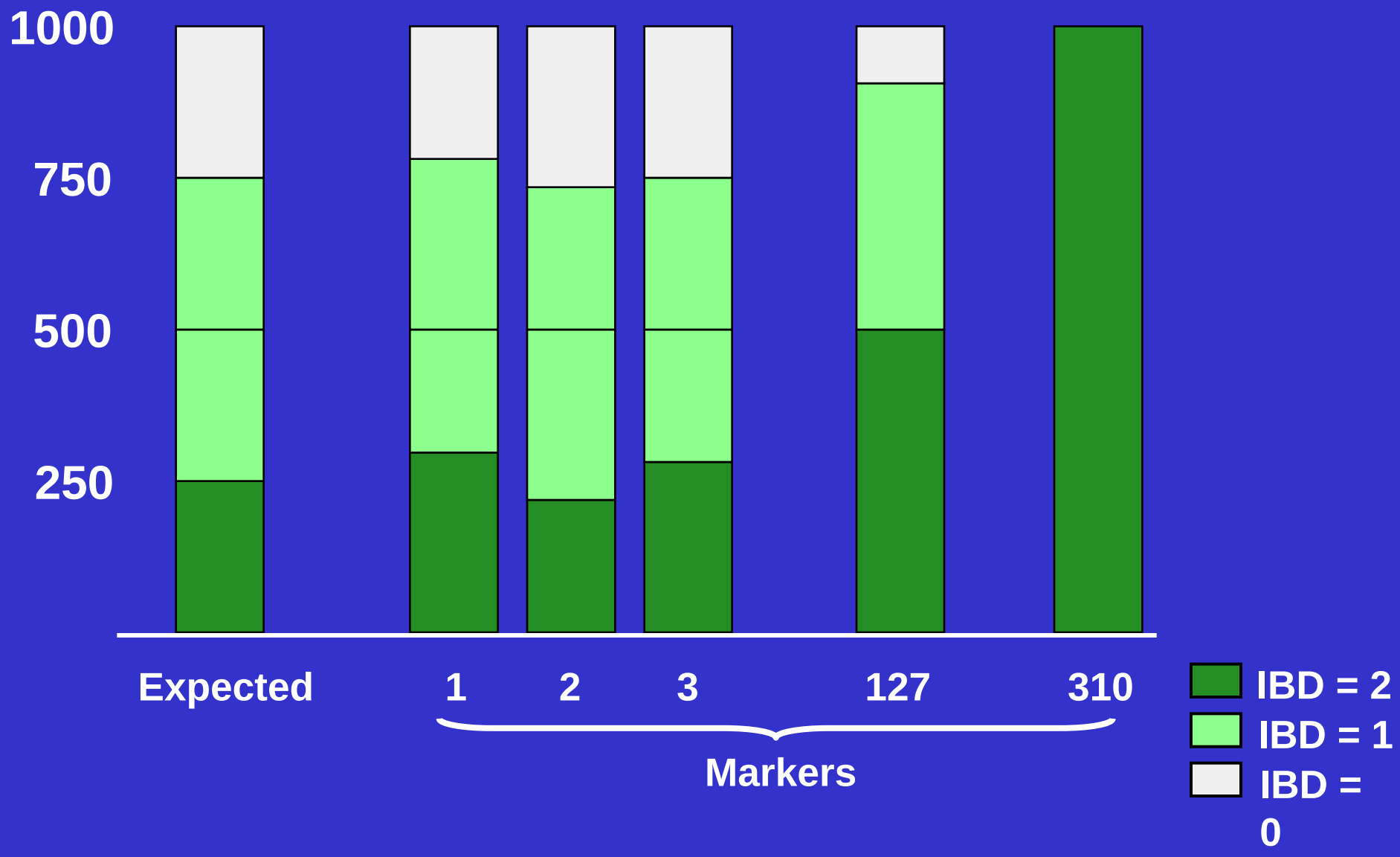

4/16 = 1/4 sibs share BOTH parental alleles IBD = 2

8/16 = 1/2 sibs share ONE parental allele IBD = 1

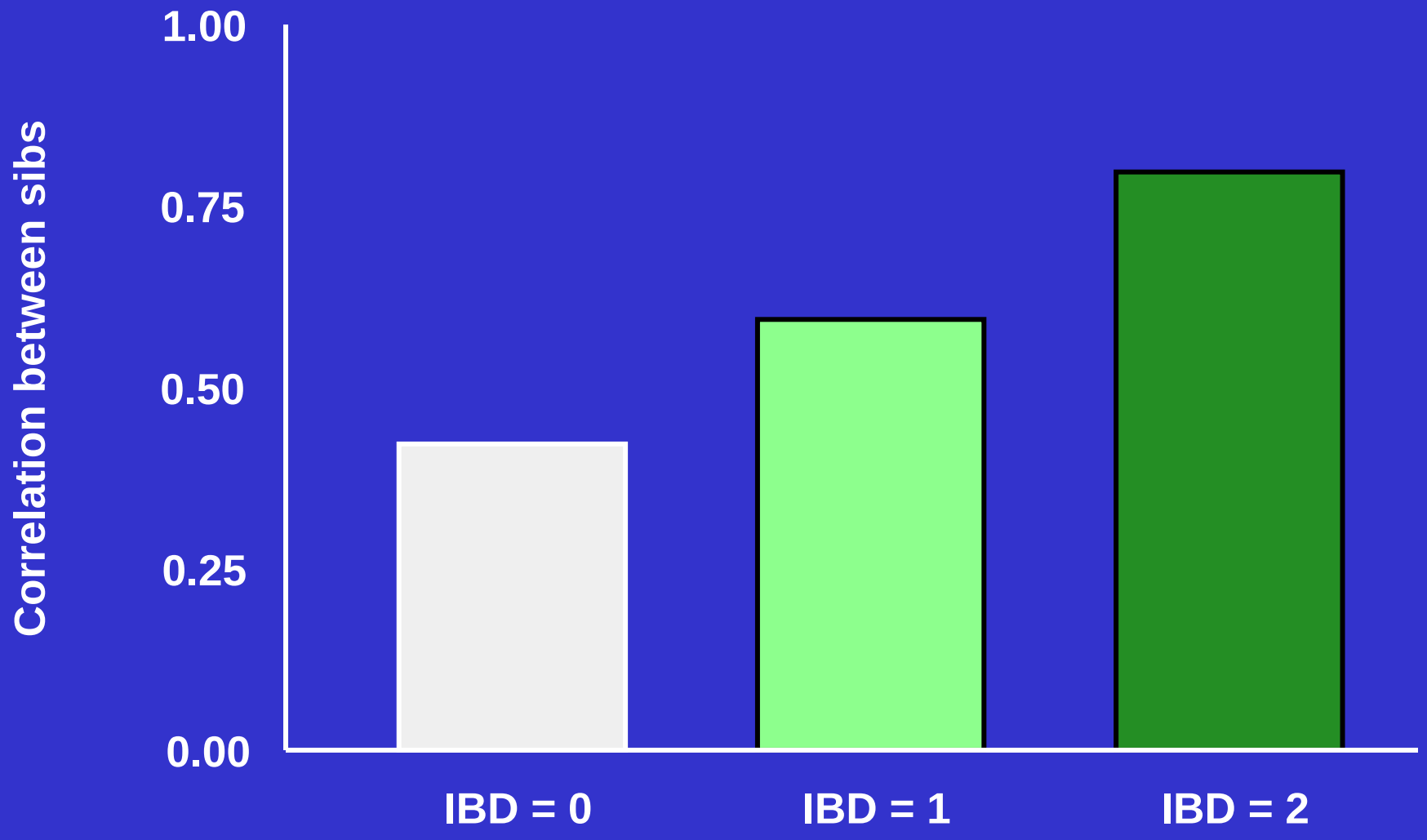
4/16 = 1/4 sibs share NO parental alleles IBD = 0

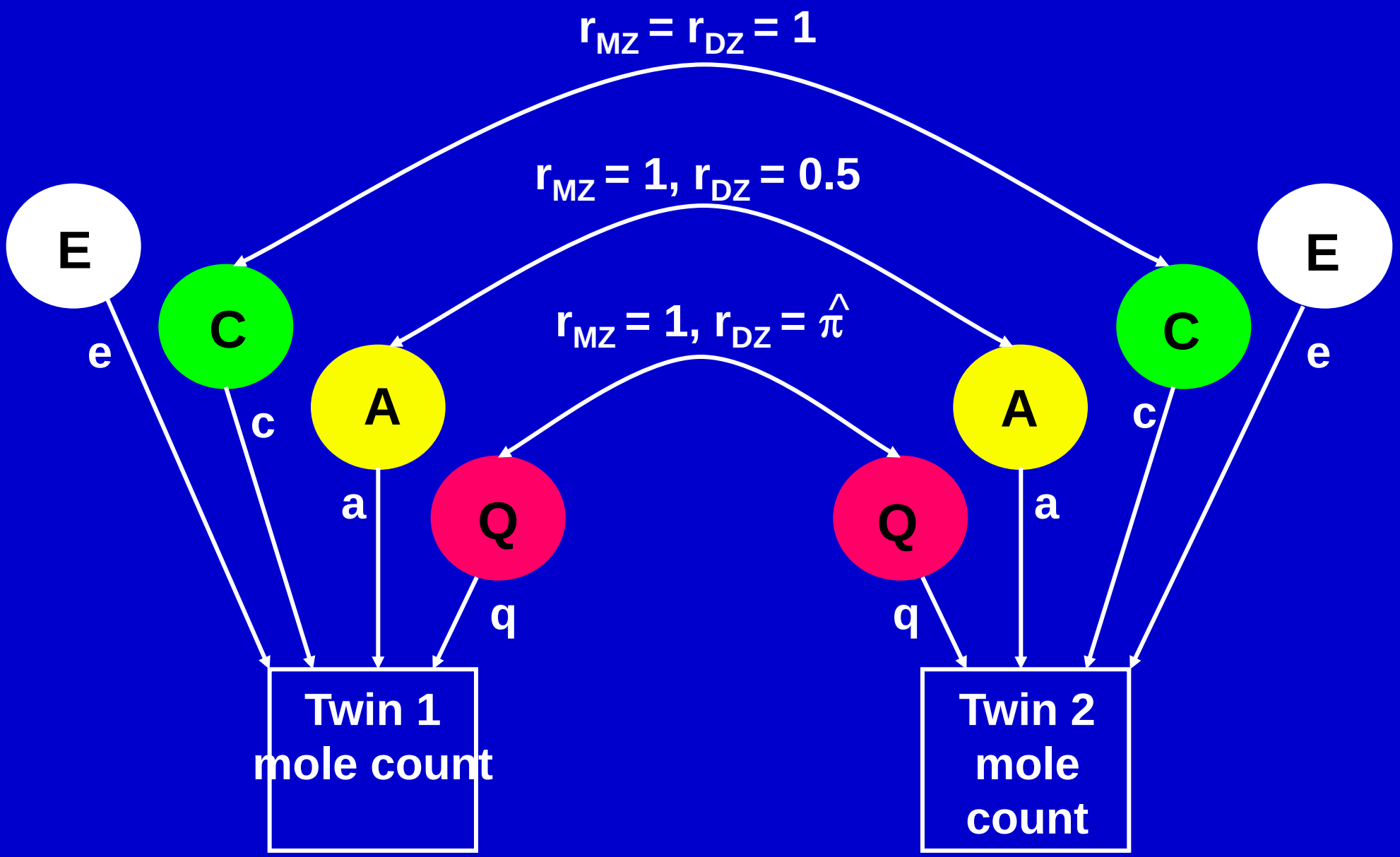
# For disease traits (affected/unaffected)

## Affected sib pairs selected

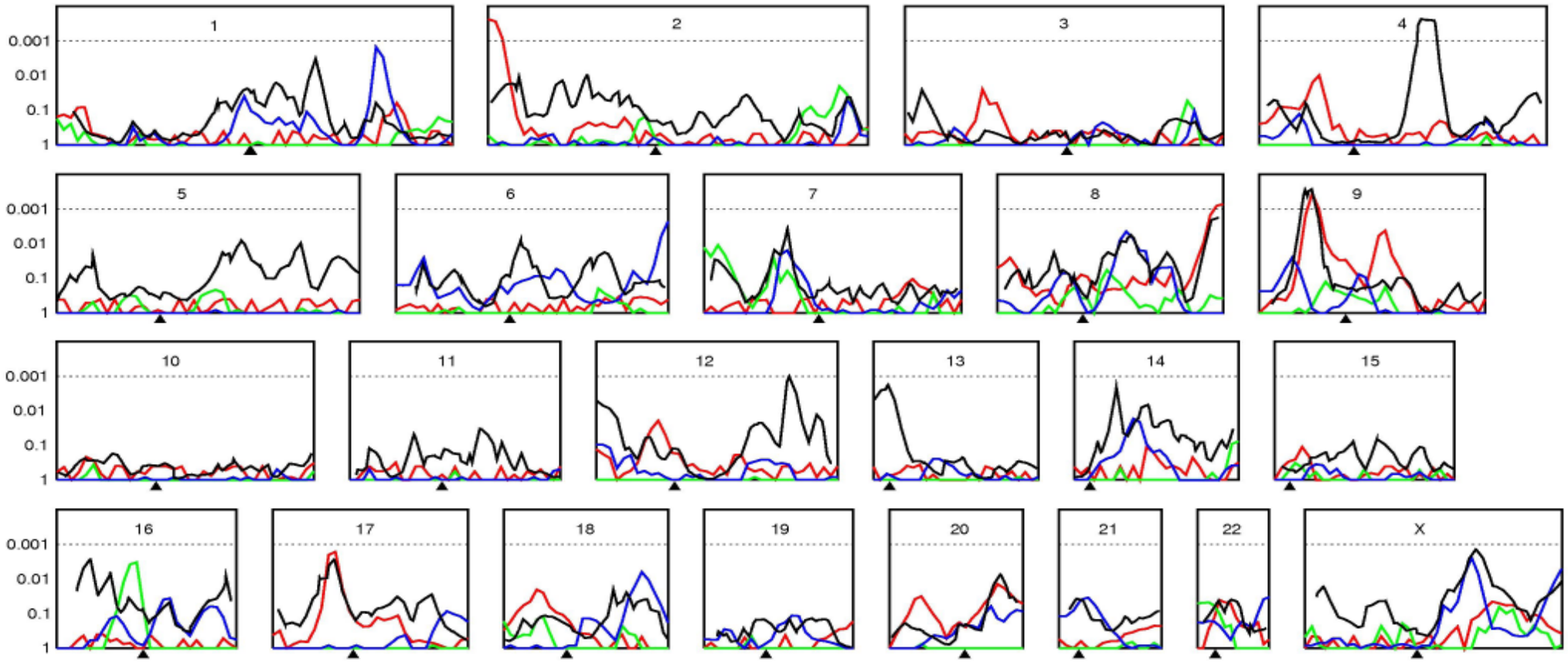


# For continuous measures
# Unselected sib pairs




$r_{MZ} = r_{DZ} = 1$
$r_{MZ} = 1, r_{DZ} = 0.5$
$r_{MZ} = 1, r_{DZ} = \hat{\pi}$
E
C
A
Q
E
C
A
Q
e
c
a
q
q
a
c
e
Twin 1 mole count
Twin 2 mole count

# Linkage for mole counts in Australian twin families



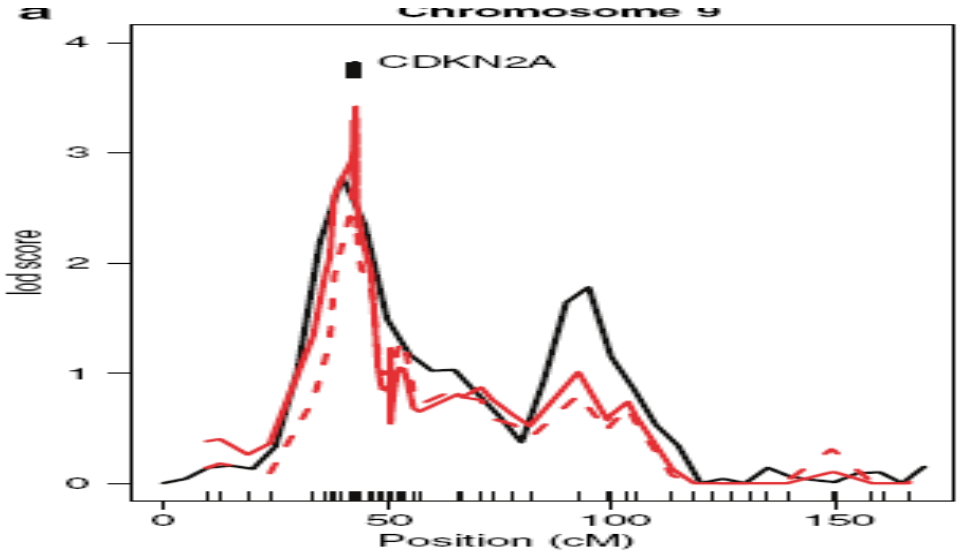

**A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions**

Gu Zhu[1], Grant W Montgomery[1], Michael R James[1], Jeff M Trent[2], Nicholas K Hayward[1], Nicholas G Martin[1] and David L Duffy*,[1]

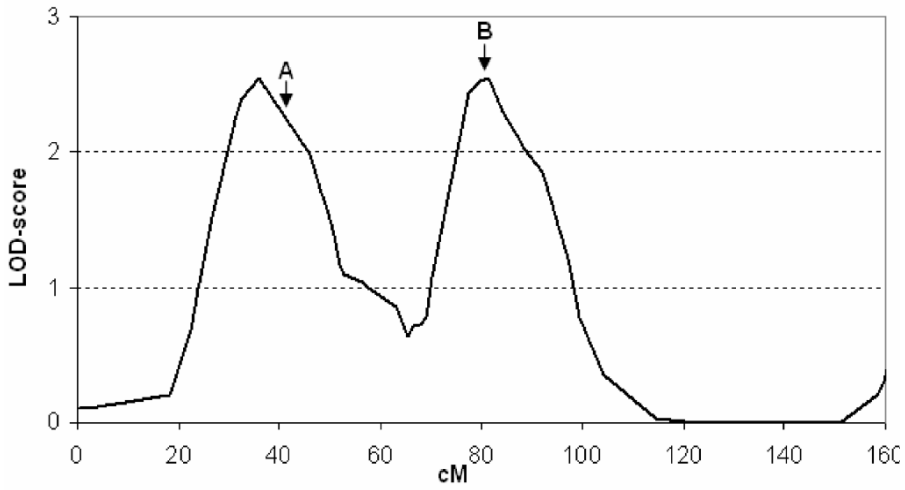
European Journal of Human Genetics (2007) 15, 94–102

# Flat mole count: chromosome 9 linkage in Australian and UK twins

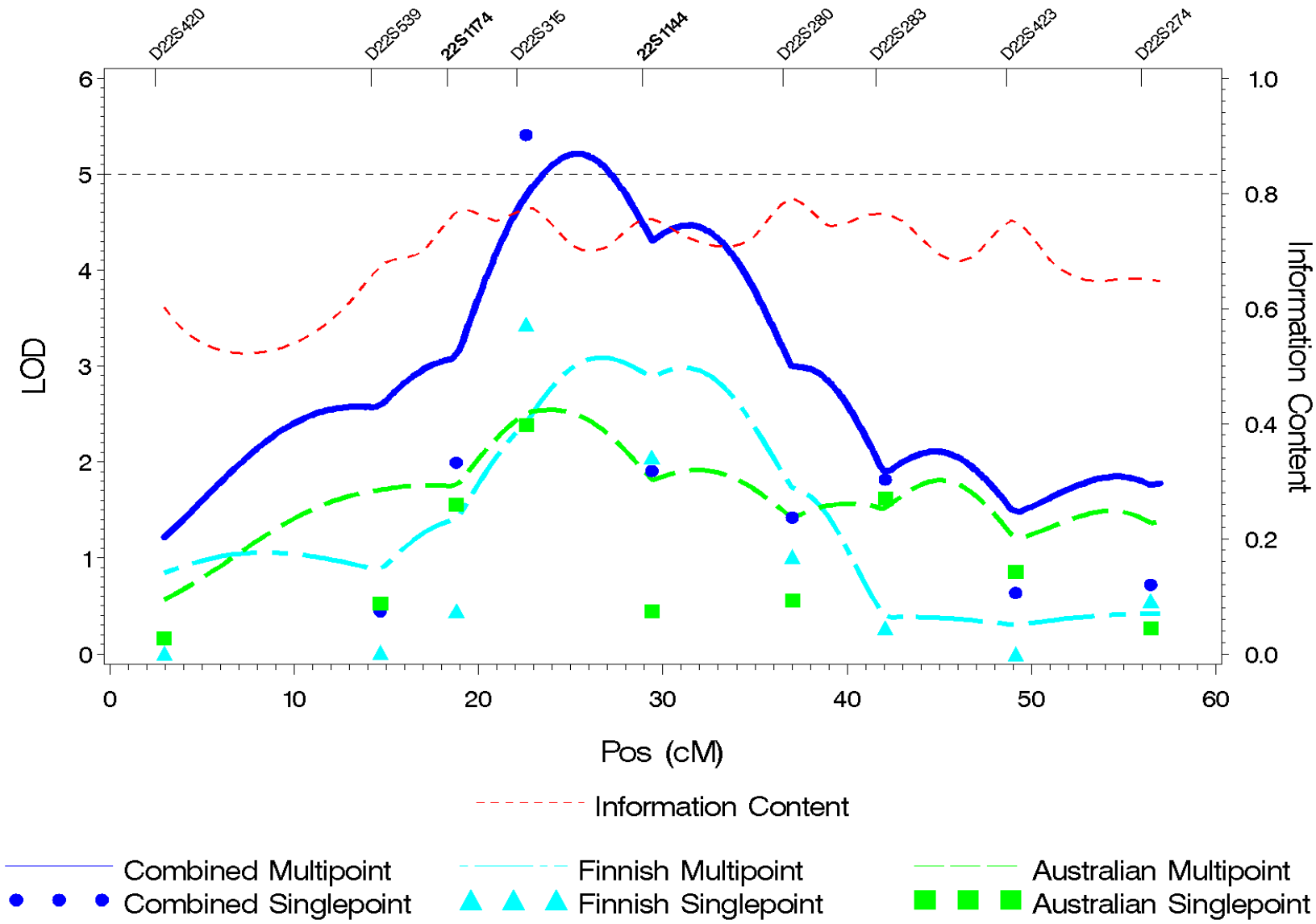
Australia




UK



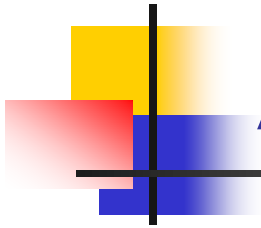
# Linkage for MaxCigs24 in Australia and Finland




AJHG, in press

# Linkage

- Doesn’t depend on “guessing gene”
- Needs family data !
- Works over broad regions and whole genome
- Only detects large effects (>10%)
- Requires large samples (10,000’s?)
- Can’t guarantee close to gene
- On the whole, has failed to deliver for complex traits



# Association

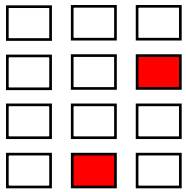
- Looks for correlation between specific alleles and phenotype (trait value, disease risk)

# Association studies

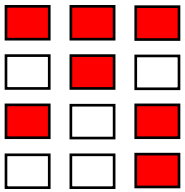
```
                 1         2         3         4         5
           1234567890 1234567890 1234567890 1234567890 1234567890

191,044,935  GTGAGTATGG CTTTCTTCCC TCTCCCGCCA CCCCCTGCCC CACACTGCAA Intron 1
191,044,985  GCTGCAAACG CGGTACTTTC GGGCTCGCCT TTGACGTTAG GAAACTAGCC
191,045,035  TGAGCCTATG CAGGGAAAAA AAATCGAAAA GGTCAATTTG TTAAGTAAGG
191,045,085  TTAAATCTGG GTGATGCTCG GGTACAGTTT AAGAACCGAG GGAGACAGTT
191,045,135  GATATGAGGG CGGTGGTTGA TGCGCTAAGA AATTGCGGGT TGGCTTTTTG
191,045,185  TCCTCCTGCA TTCAAAATGA CATCAGAATC CTGCGGCTGA AGCGCGTCCC  rs34774923
191,045,235  CAGCATTCAT ACGTTGCATG ATGAGTTCTC ATCAGCTTAC ACAGCTACTG
191,045,285  GAAGGTGATG CTCTTGCTGG TTCTGAATAT ACTCGTTTAA AATCCATTTT
191,045,335  TGTTTTTTAA TTATAGAGCA GATCTCACCC AGTCCGAATG TGGAACAAAT
191,045,385  AATTGTTATG CAGCGTCTGC TTAAAAGAAG TGTCGTAGGT GGGGAAGGAA
191,045,435  GAGCGCAGGG GAATCAGTCA CCCACCTCTT TGTACAGTCT CTGGCGTGGT  rs10218513
191,045,485  CCAGAACCTC CTGCTCTAAA GAGAGAAGCG TGGGCCGGCT CCAGACAGTT
191,045,535  CCATGTCTGT CCTTTTCATT AAAGTGCAAA ACGTCTCGGA ATTGTAATTA
191,045,585  ACCTTGCAAA CAAACTGATG CCCTTTGTGA GCCAGAAATA GTGTCTGCCT
191,045,635  TTTGAACTAA ATTCATTAAC AATTCTTTAA AATACCCTAG TGATTATAGG
191,045,685  TAGCCCTGCC CTTAGTTGTA AAACTAGTAG ATACGGTCAG ATTAATGGAC
191,045,735  GAAACTGCTC AGTACATGAG GTTTAAATGT TAGGTGGATA AGACTTATTT
191,045,785  GAAGAGTTCT TGCTTTGCCT TATGCGGTTT GTCTCTAGTT ACTGGGTGAC
191,045,835  TTTATTTGGT AAAAATGCGT TCAGCTGCAG TAGCATATTC AAGTGTTGCT  rs2746073
191,045,885  AGTTAGTAAT TATCTTTTTA ATTTTTTGTT TTAG
```

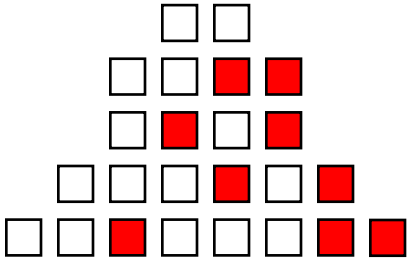
Single Nucleotide Polymorphisms



Controls



Cases

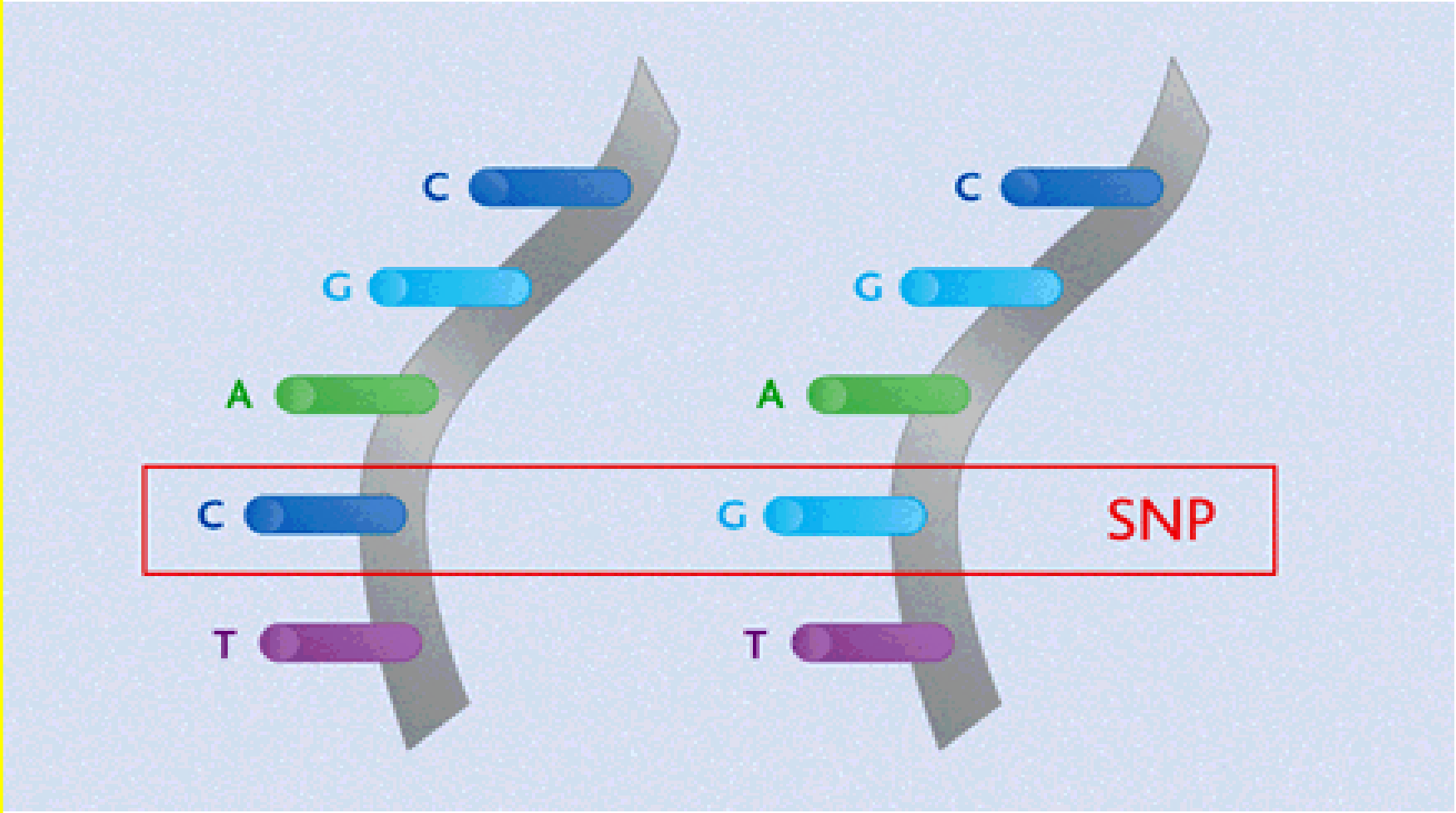


Quantitative Traits

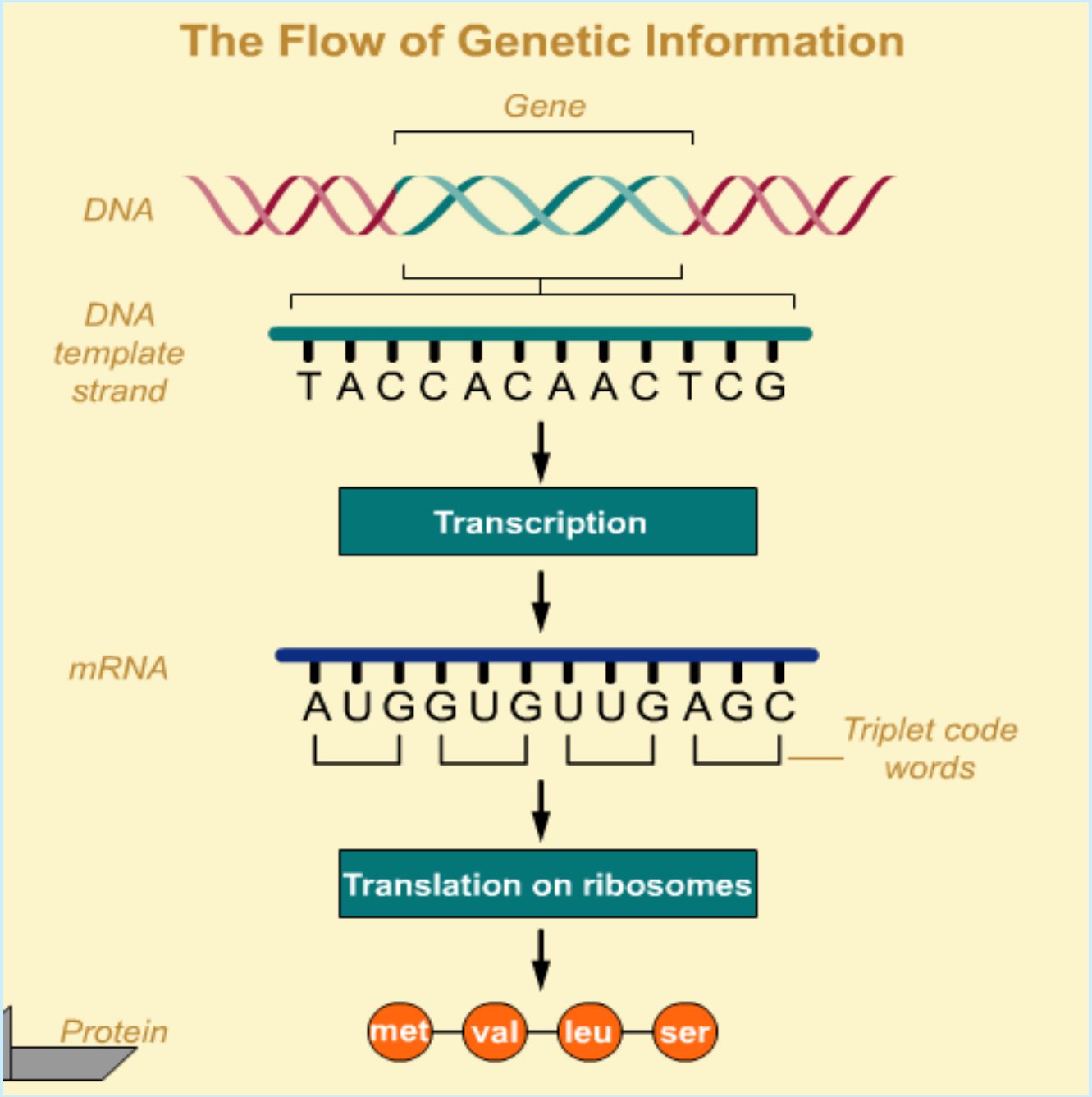
# Association

- More sensitive to small effects
- Need to “guess” gene/alleles (“candidate gene”) or be close enough for linkage disequilibrium with nearby loci
- May get spurious association (“stratification”) – need to have genetic controls to be convinced

# Variation: Single Nucleotide Polymorphisms

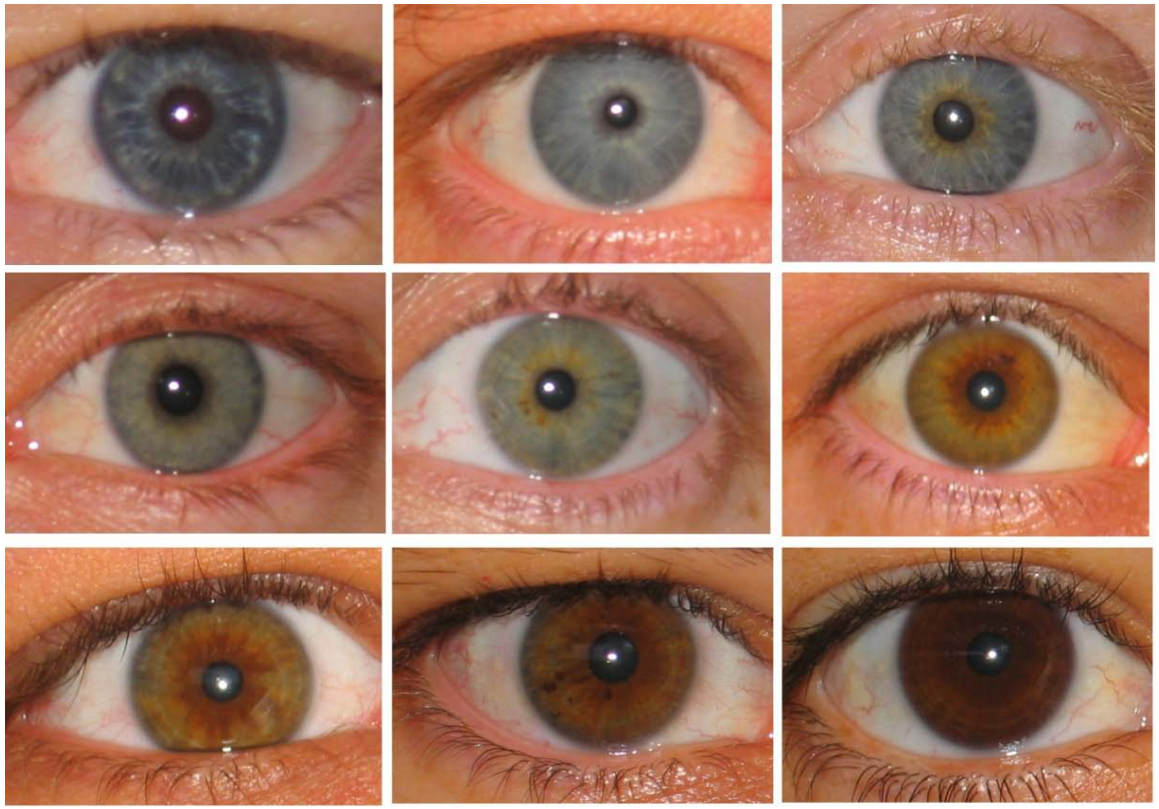


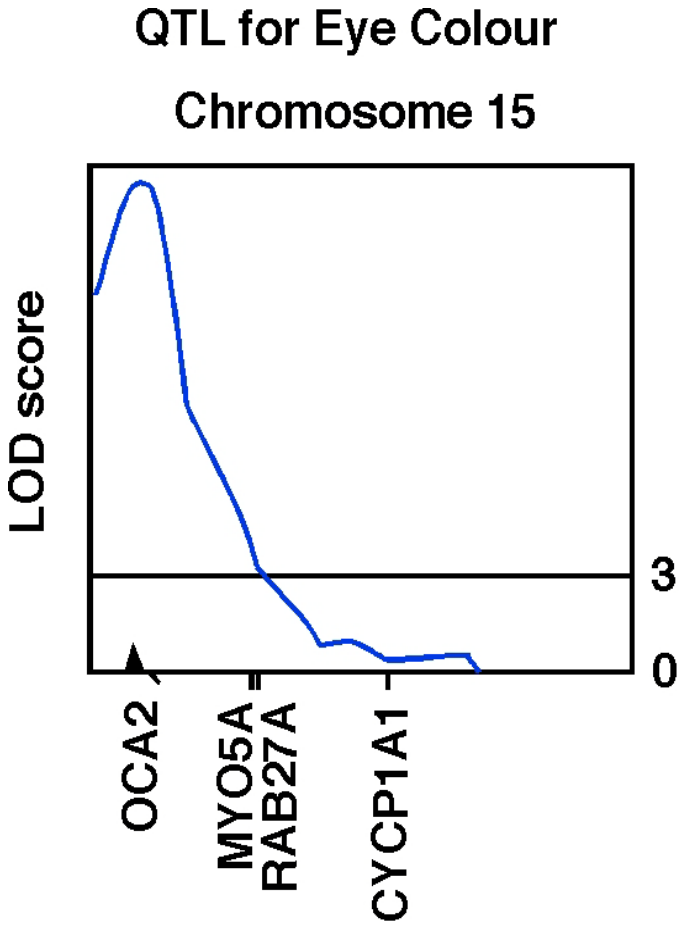

**Complex disease marker?** SNPs are single-base differences in DNA




Differences (between subjects) in DNA sequence are responsible for (structural) differences in proteins.

# Human OCA2 and eye colour






Zhu et al., *Twin Research* 7:197-210 (2004)

# Oculotaneous Albinism type 2

## Missense Mutations

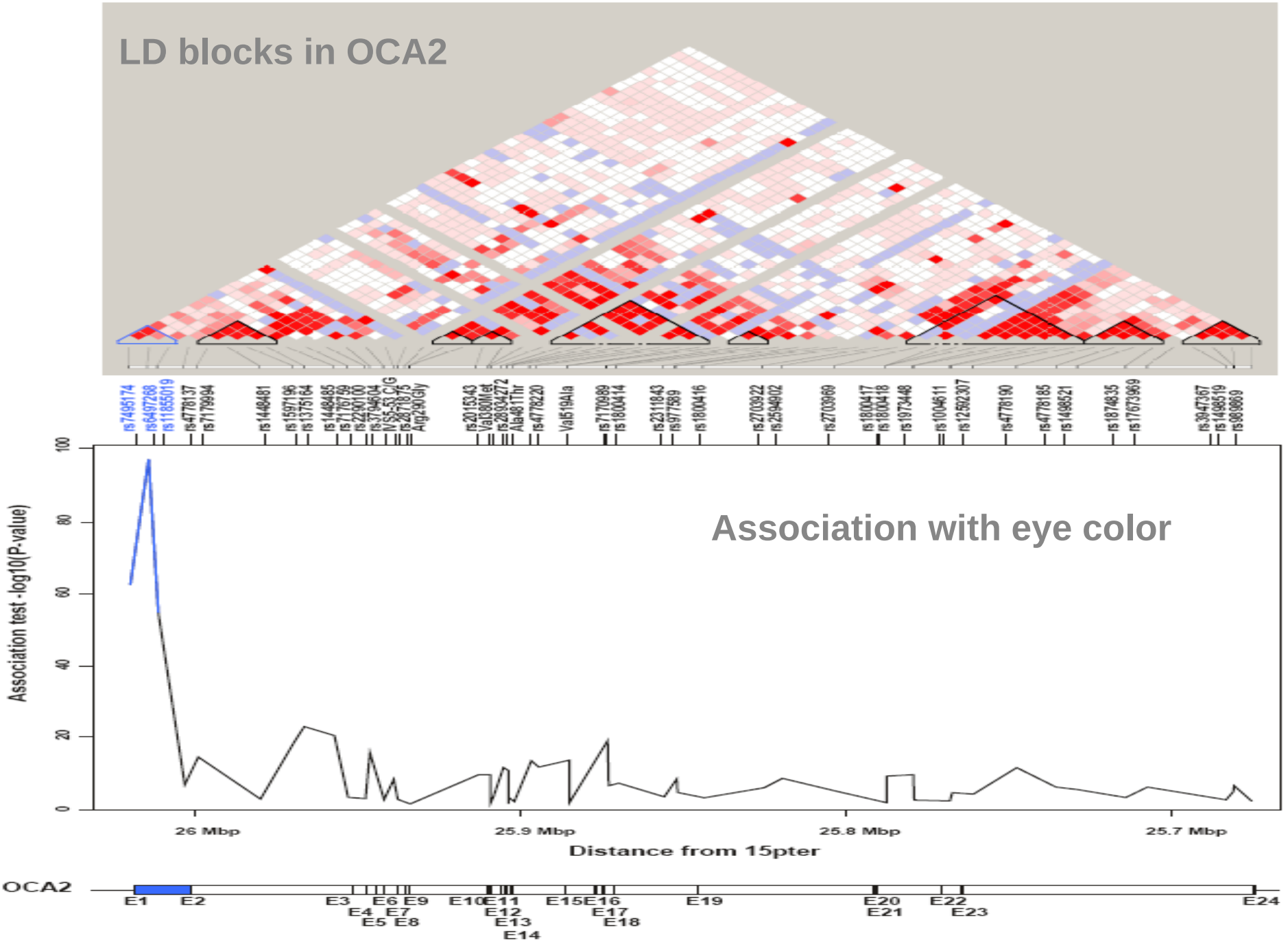
G27R
S86R
C112F
NW273KV
R290G
A368V
P385I
M395L
T404M
V443I
M446V
I473S
A481T
N489D
H549Q
T592I
K614E
K614N
I617L
W652R
W679C
W679R
R720C
A724P
S736L
P743L
A787V
G795R

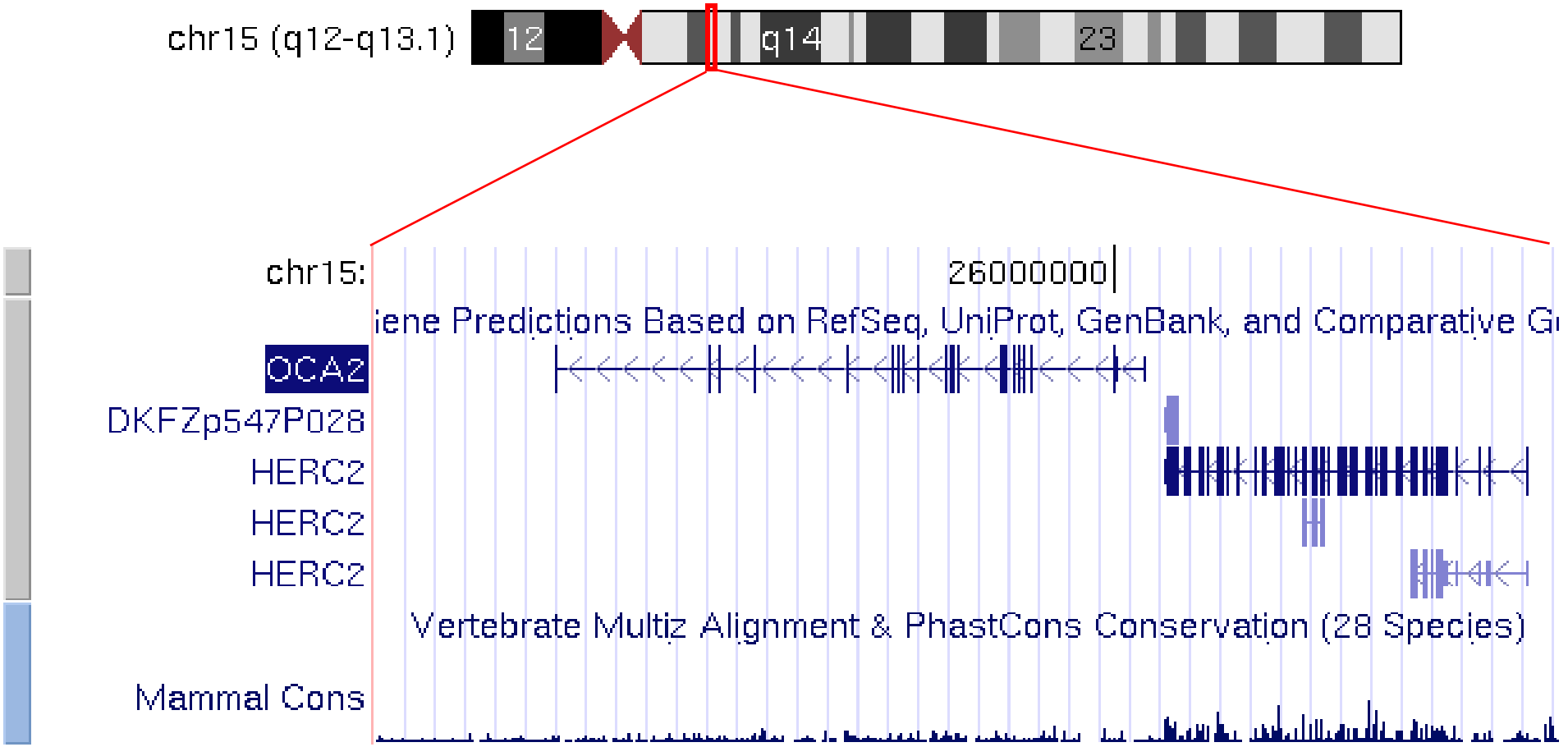
## Polymorphisms

D/A257
R/W305
R/Q419
L/F440
H/R615
I/T722

# OCA2 Mutations

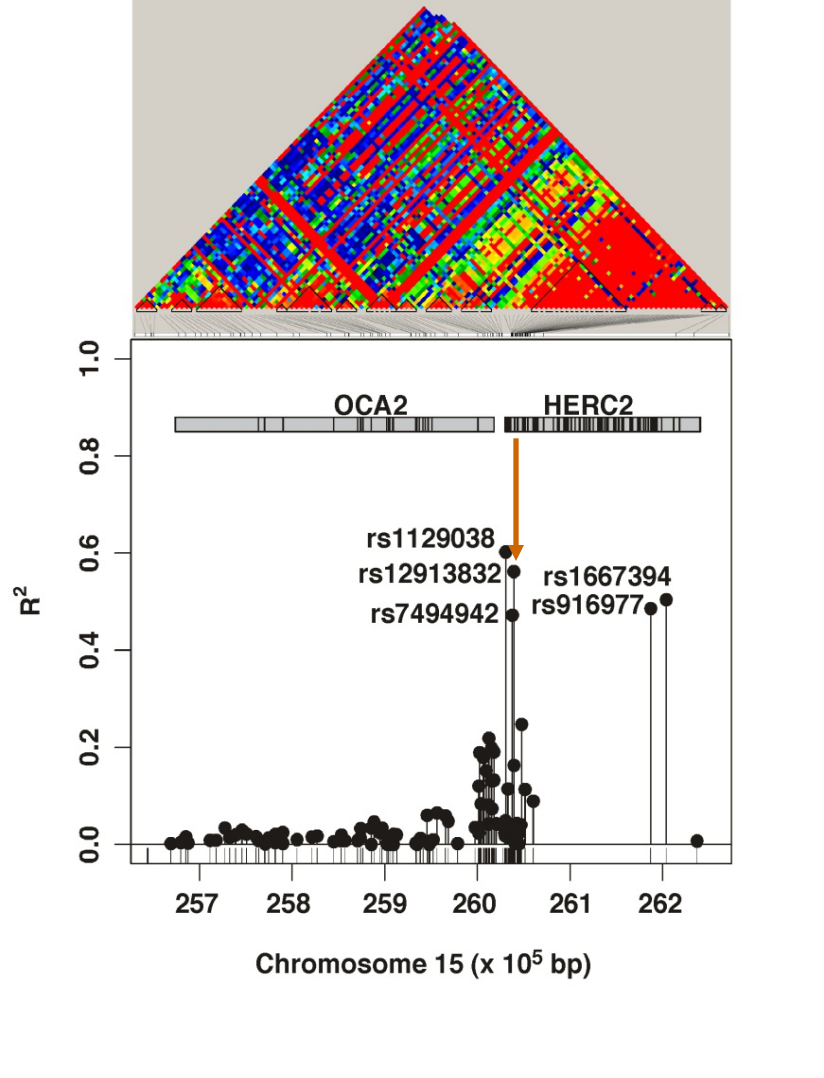
Albinism data base www.cbc.umn.e


LD blocks in OCA2
rs7495174
rs6497268
rs11855019
rs4778137
rs7179994
rs1448481
rs1597196
rs1375164
rs1448485
rs7176769
rs2290100
rs3794604
IVS5-53 C/G
rs2871875
Arg290Gly
rs2015343
Val380Met
rs2934272
Ala481Thr
rs4778220
Val519Ala
rs7170989
rs1800414
rs2311843
rs977589
rs1800416
rs2703922
rs2594802
rs2703969
rs1800417
rs1800418
rs1973448
rs1004611
rs12592307
rs4778190
rs4778185
rs1498521
rs1874835
rs17673969
rs3947367
rs1498519
rs989869
Association with eye color
Association test -log10(P-value)
0
20
40
60
80
100
26 Mbp
25.9 Mbp
25.8 Mbp
25.7 Mbp
Distance from 15pter
OCA2
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11
E12
E13
E14
E15
E16
E17
E18
E19
E20
E21
E22
E23
E24


chr15 (q12-q13.1)
12
q14
28
chr15:
26000000
ene Predictions Based on RefSeq, UniProt, GenBank, and Comparative G
OCA2
DKFZp547P028
HERC2
HERC2
HERC2
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Mammal Cons

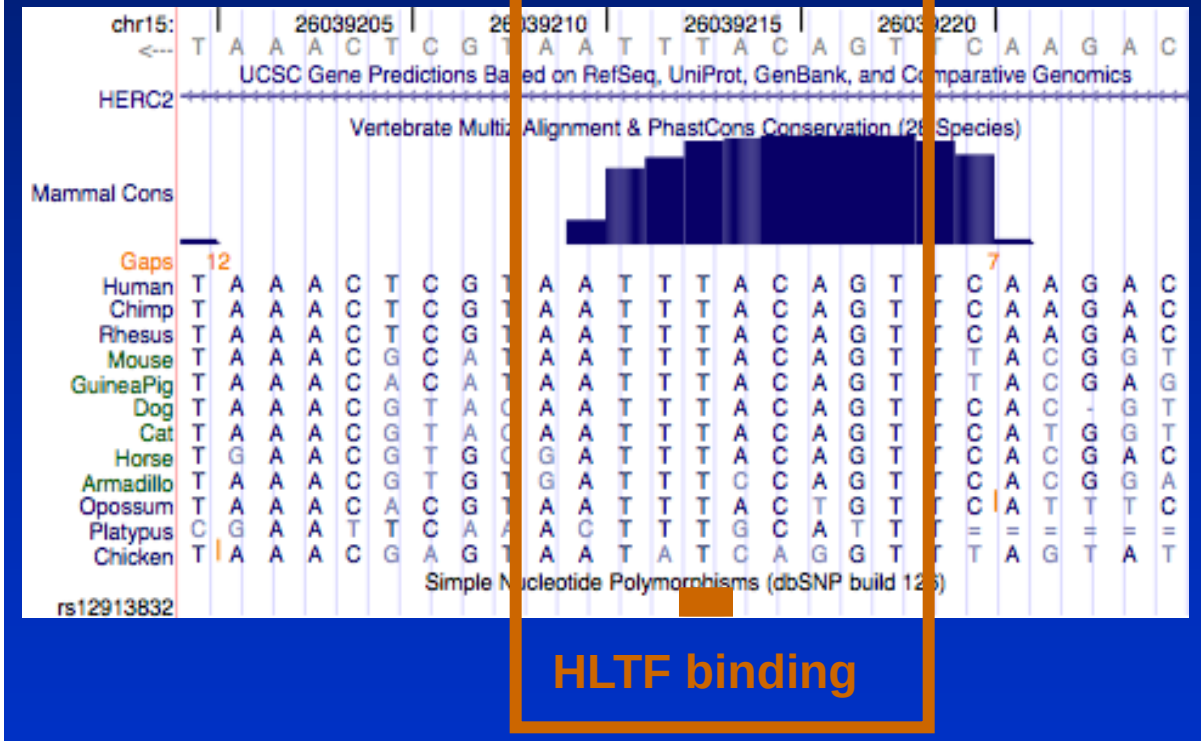
# A single SNP within intron 86 of HERC2 determines Blue-Brown eye colour

## Sturm et al., AJHG 82: 424-431, 2008




**rs12913832 C = Blue**

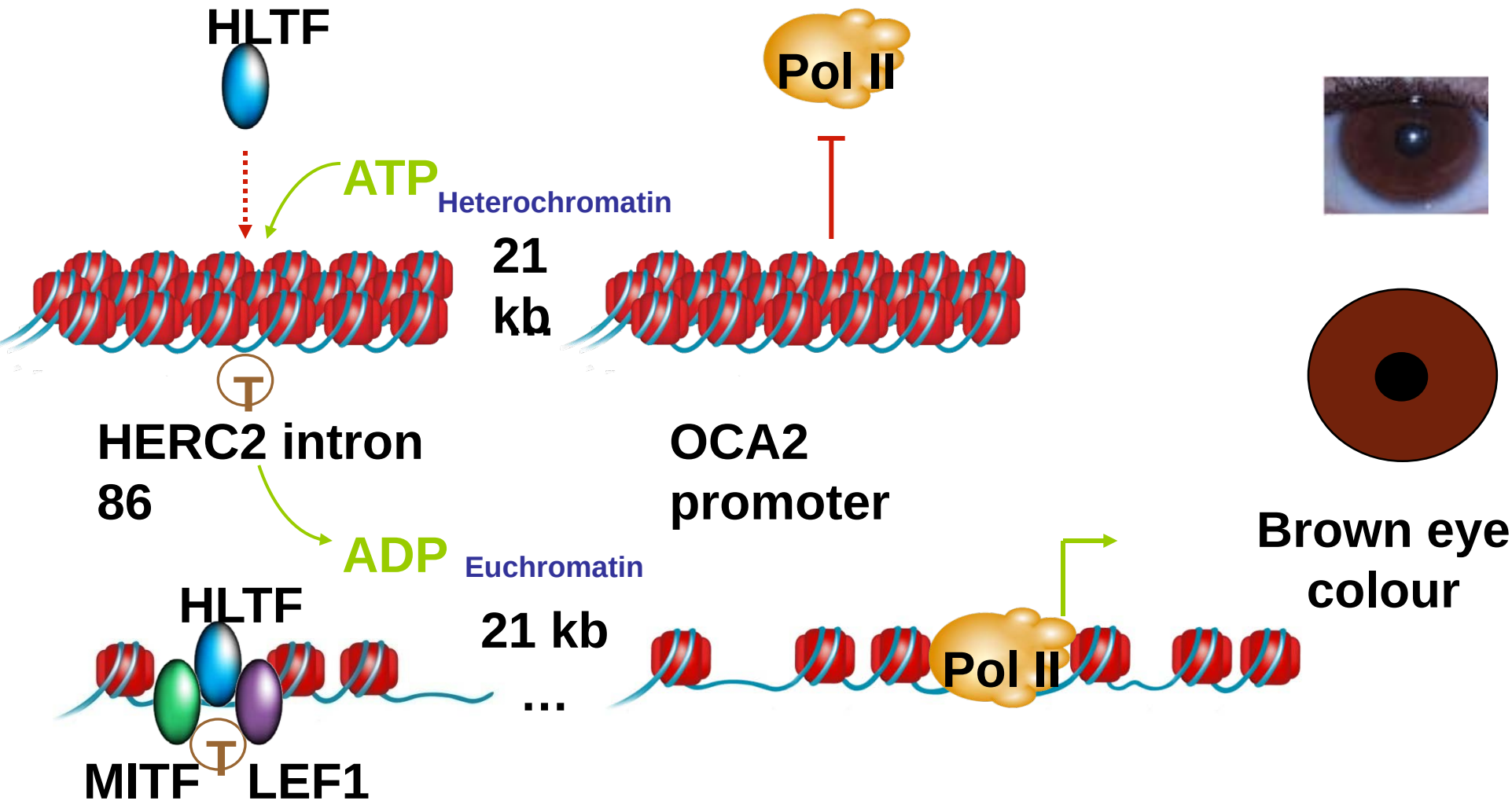
**rs12913832 T = Brown**



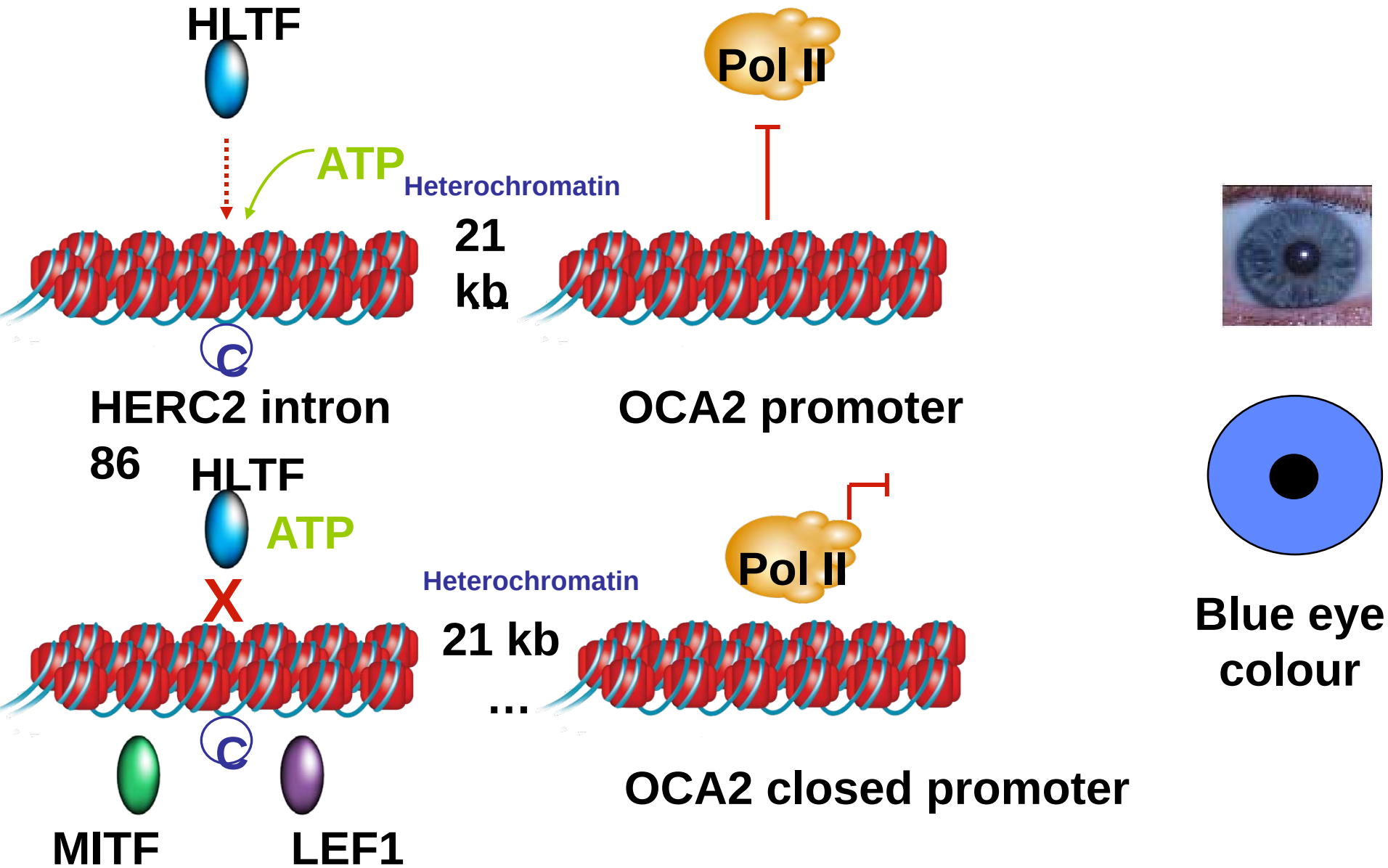

HLTF = Helicase-like transcription factor
Uses ATP for energy to initiate unwinding of DNA, so allowing transcription

Sulem et al, Nat Genet 39: 1443, 2007; Kayser et al, AJHG 82: 411-423, 2008; Eiberg et al, Hum Genet 123: 177-187, 2008

# SWI/SNF Chromatin remodelling model for OCA2 gene activation



# SWI/SNF Chromatin remodelling model for OCA2 gene activation







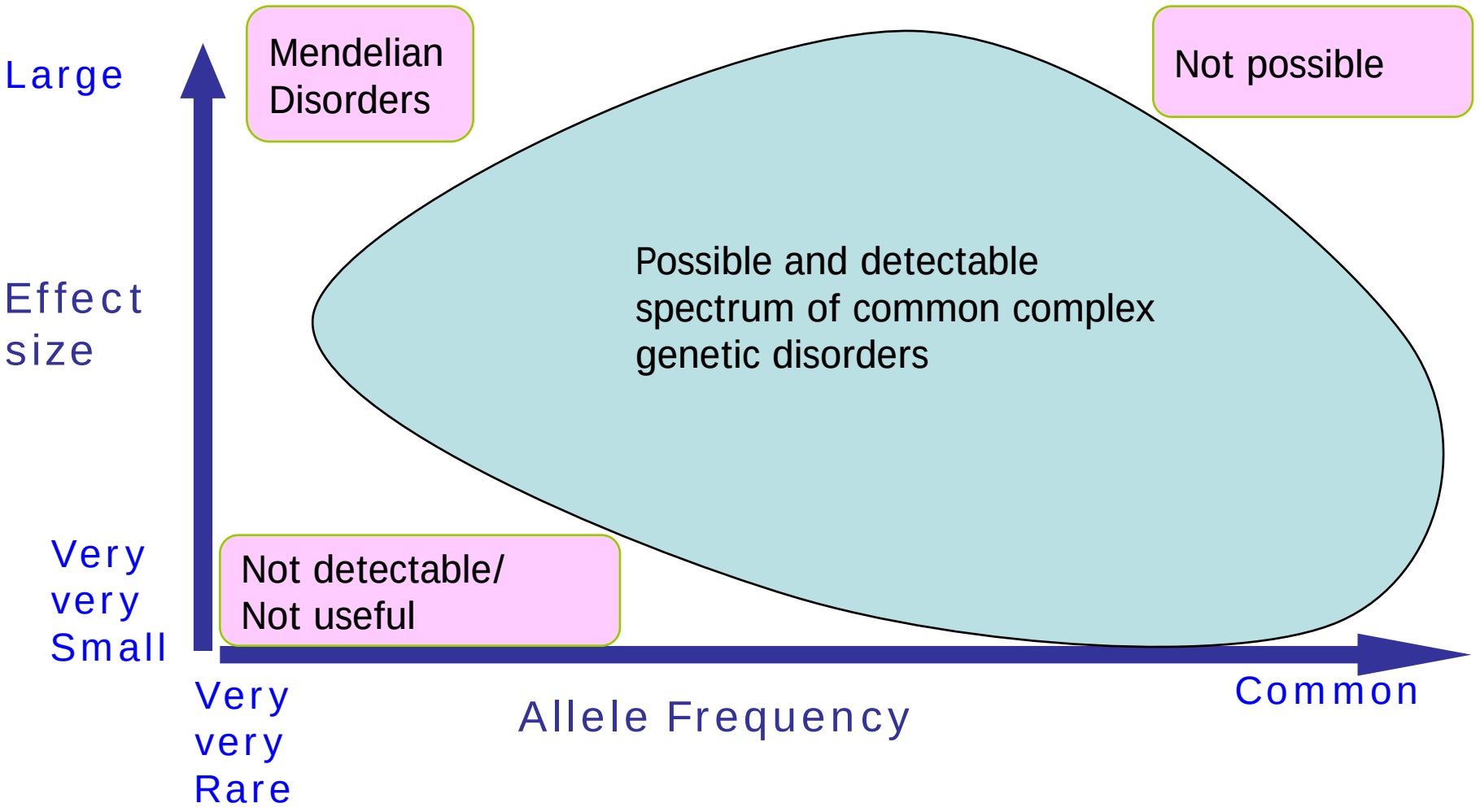

Blue eye colour

# Genetic architecture

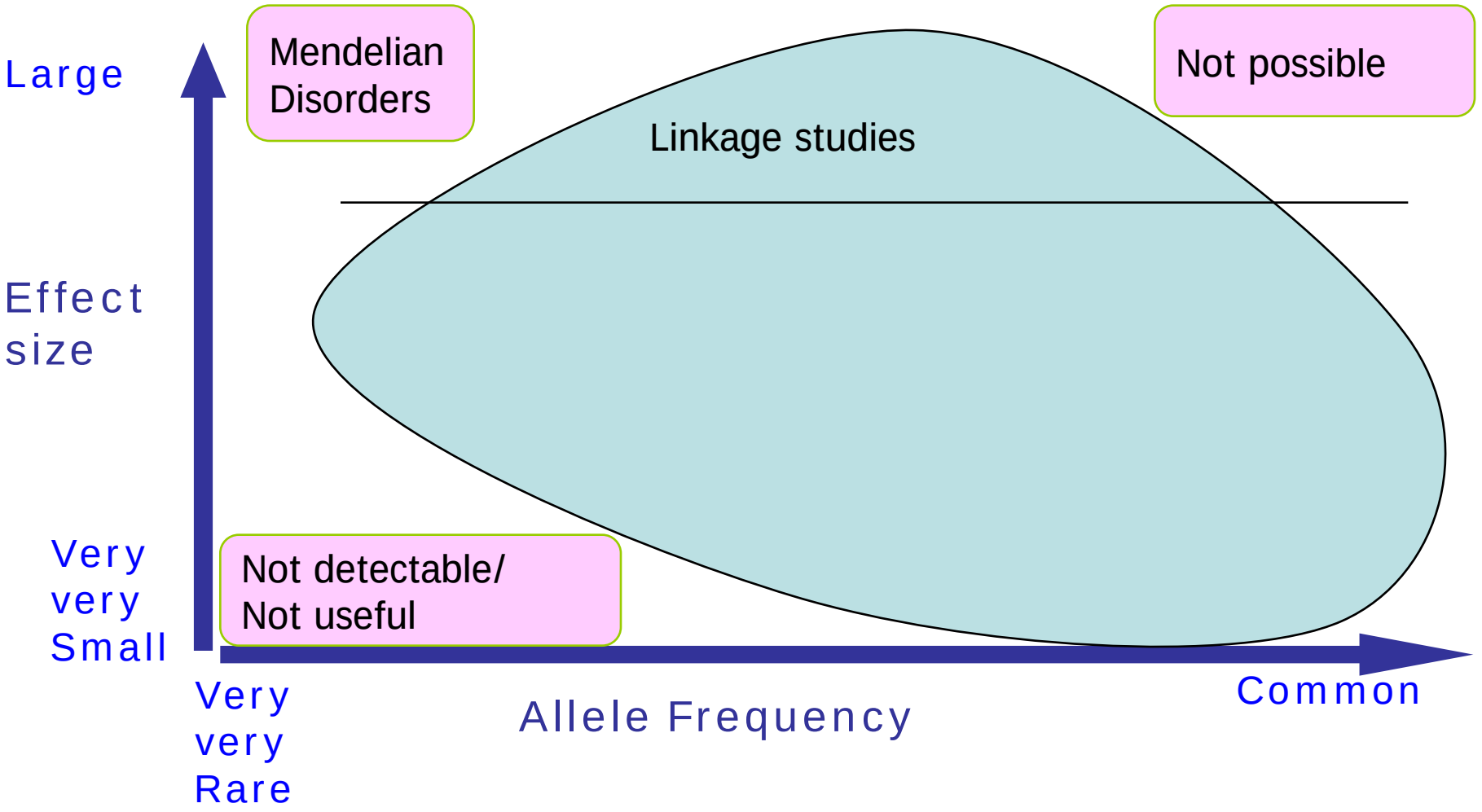
- Number of variants
- Effect size of variants
- Frequency of risk allele
- Genetic mode of action

**Genetic variance**

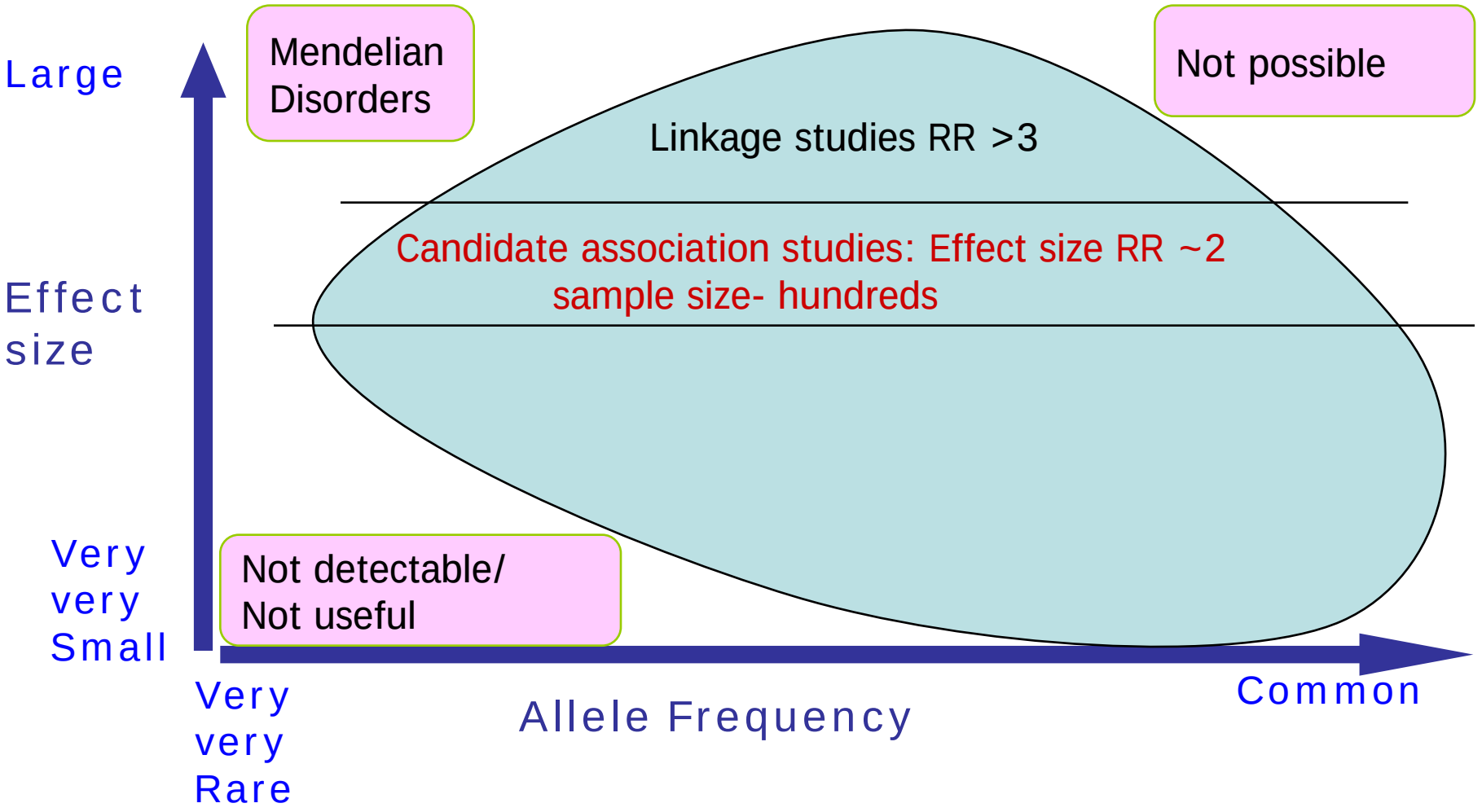
# What is the genetic architecture of complex genetic disorders?



# A potted history of genetic studies



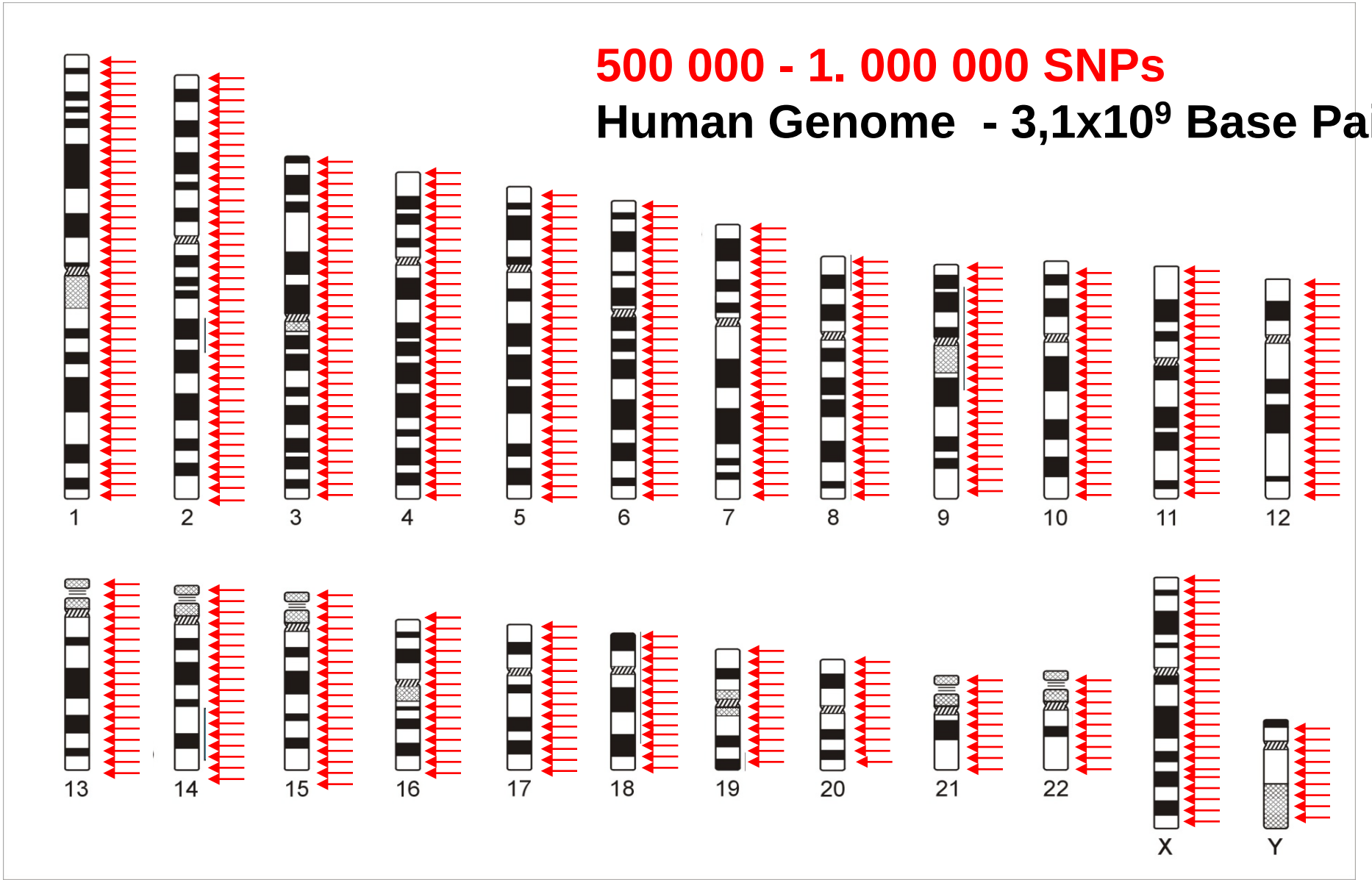
# A potted history of genetic studies



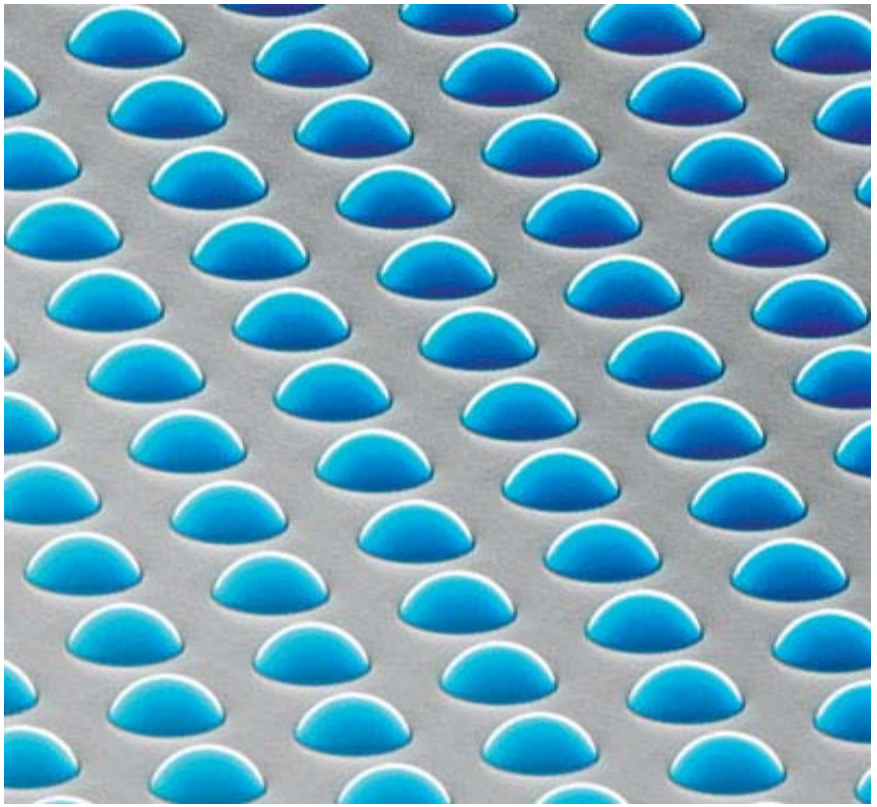
# Association studies until 2007

- Candidate regions
- Some successes but generally:
  - Poor replication
  - Small sample sizes
- Conclusion
  - effect sizes are smaller than expected
  - Poor selection of candidate regions

# Genome-Wide Association Studies



# High density SNP arrays – up to 1 million SNPs

# Genome-wide Association Studies

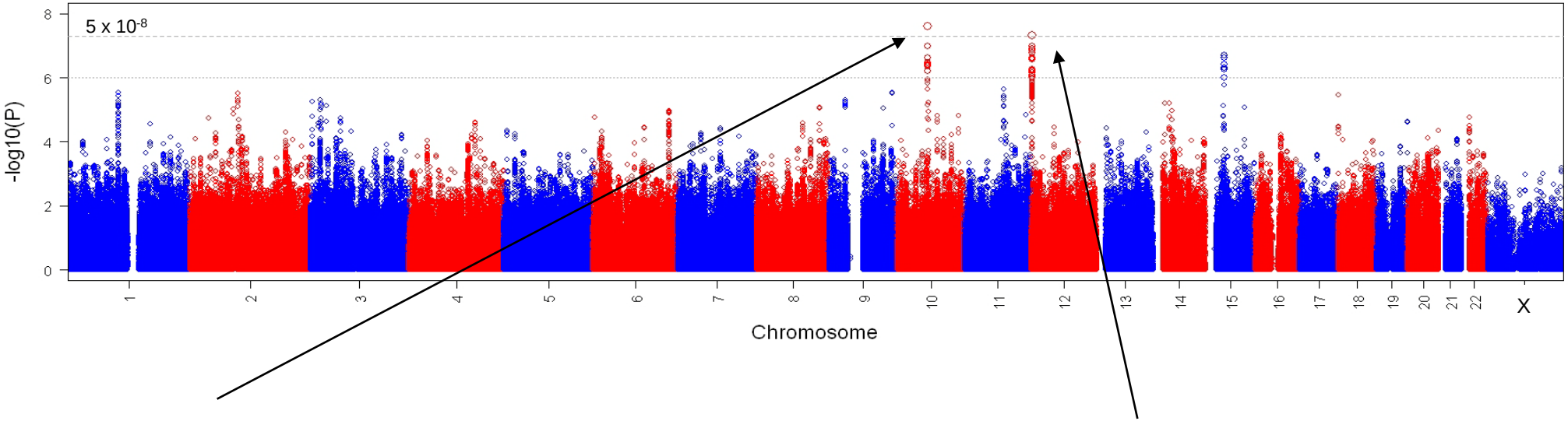
- Multiple testing !
  - Declare significance at $0.05/500{,}000 = 10^{-8}$

- Follow-up significant SNPs by genotyping in independent replication samples

- Validated associations:
  - significant across replication samples
  - same SNP
  - same allele

# Bipolar GWAS of 10,648 samples

***>1.7 million genotyped and (high confidence) imputed SNPs***




## *Ankryin-G (ANK3)*

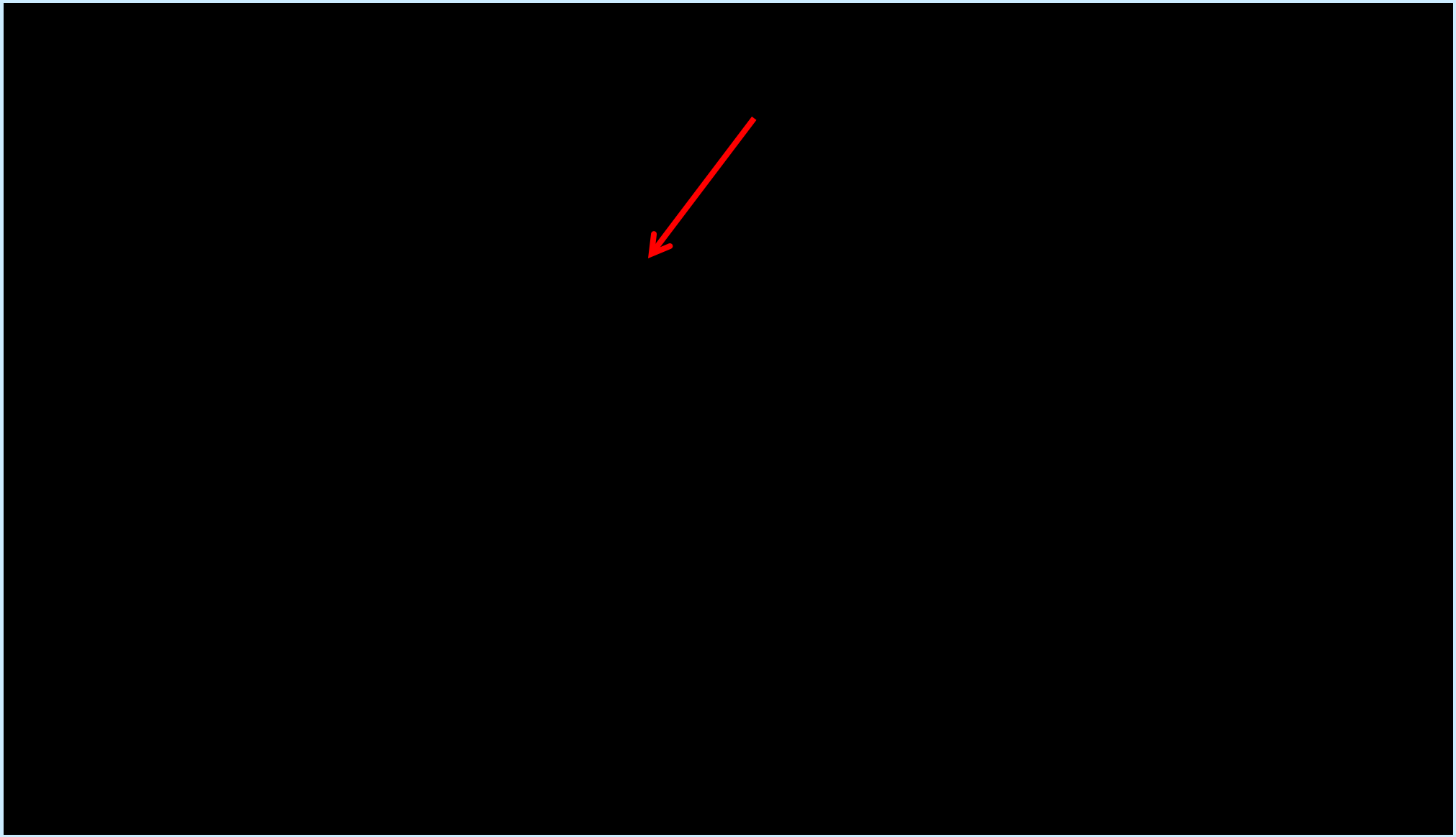
| Sample | Cases | Controls | *P*-value |
|---|---|---|---|
| STEP | 7.4% | 5.8% | 0.0013 |
| WTCCC | 7.6% | 5.9% | 0.0008 |
| EXT | 7.3% | 4.7% | 0.0002 |
| **Total** | **7.5%** | **5.6%** | **$9.1 \times 10^{-9}$** |

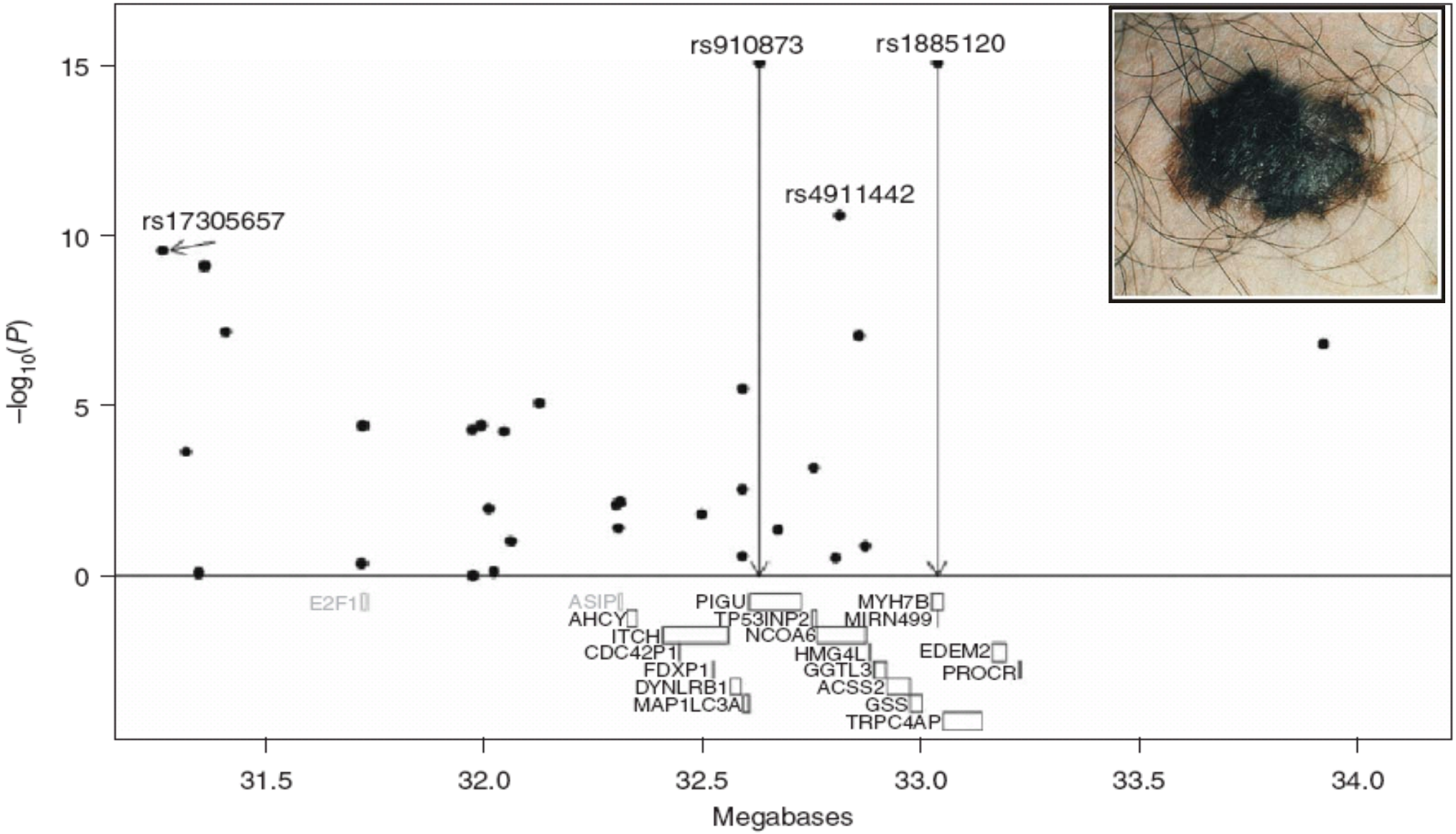
## *CACNA1C*

| Sample | Case | Controls | *P*-value |
|---|---|---|---|
| STEP | 35.7% | 32.4% | 0.0015 |
| WTCCC | 35.7% | 31.5% | 0.0003 |
| EXT | 35.3% | 33.7% | 0.0108 |
| **Total** | **35.6%** | **32.4%** | **$7 \times 10^{-8}$** |

**Ferreira *et al* (*Nature Genetics*, 2008)**

# GWAS for major depression






**GWAS for Melanoma** Association analysis of SNPs across a region of chromosome 20q11.22 for the combined sample. The x-axis is chromosomal position, the left y-axis $-\log_{10}(p)$ for genotyped SNPs. *Nature Genetics* 2008 Jul;40(7):838-40.

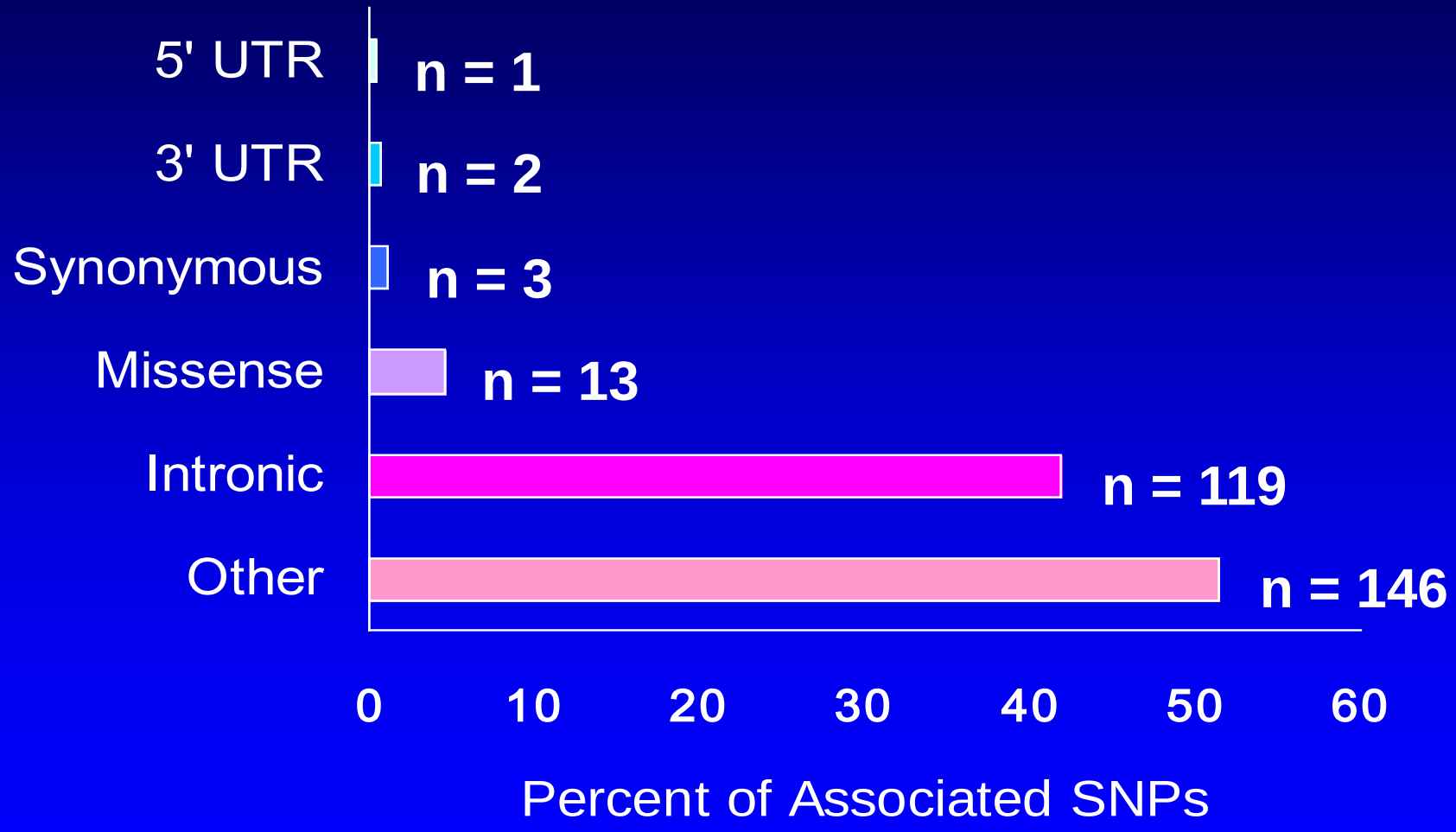
# Susceptibility variants for male-pattern baldness on chromosome 20p11

# First quarter 2008


Manolio, Brooks, Collins, J. Clin. Invest., May 2008

Stephen Channock

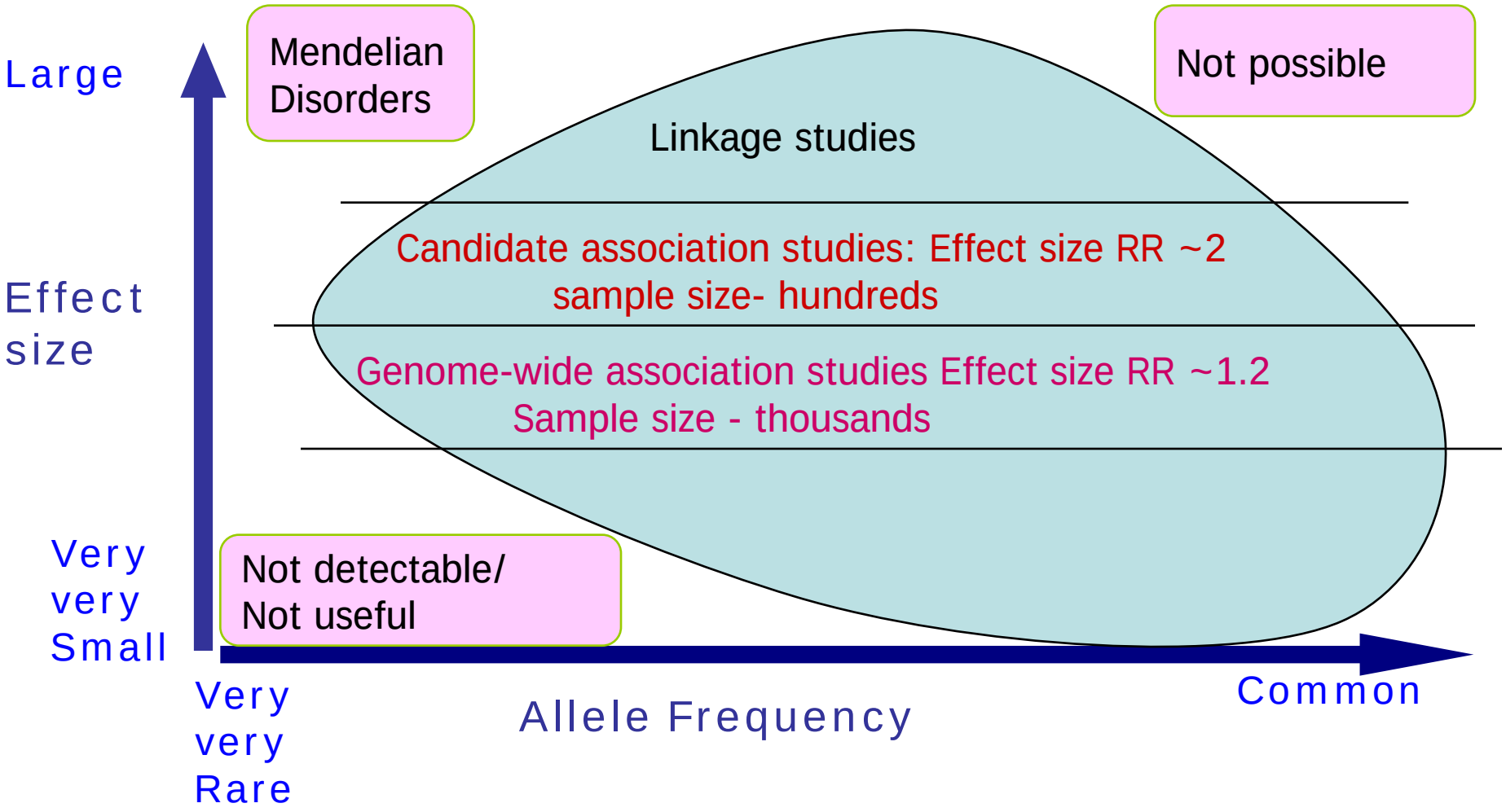
# Functional Classification of 284 SNPs Associated with Complex Traits




http://www.genome.gov/gwastudies/

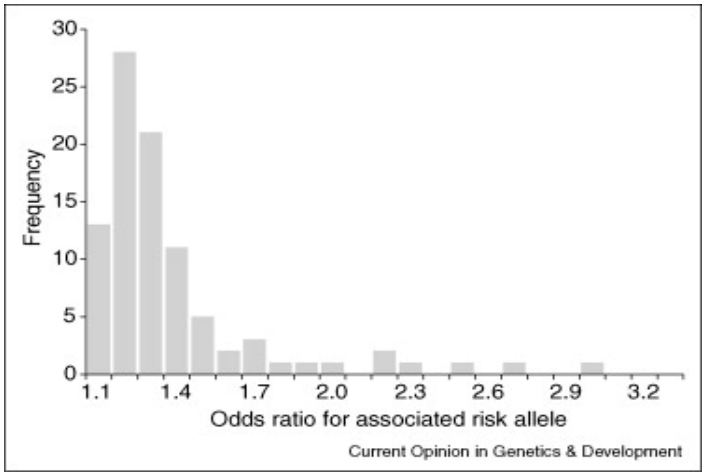
Stephen Channock

# A potted history of genetic studies



# Conclusions about genetic architecture of complex genetic traits based on GWAS results

Effects sizes of validated variants from 1st 16 GWAS studies




Genetic variation is not tagged by GWAS SNPs

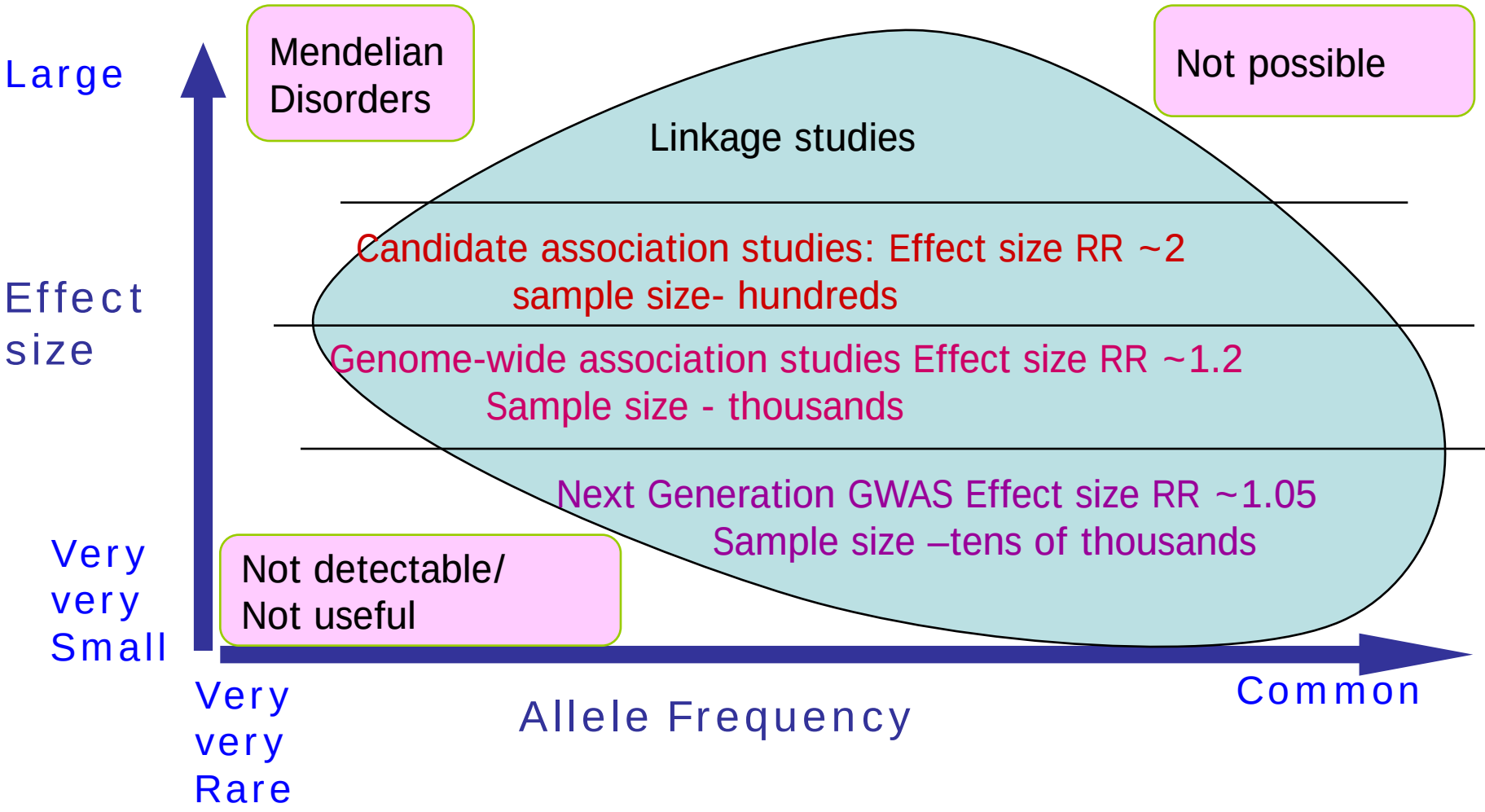
Effect sizes are smaller

**Prediction of individual genetic risk of complex disease**

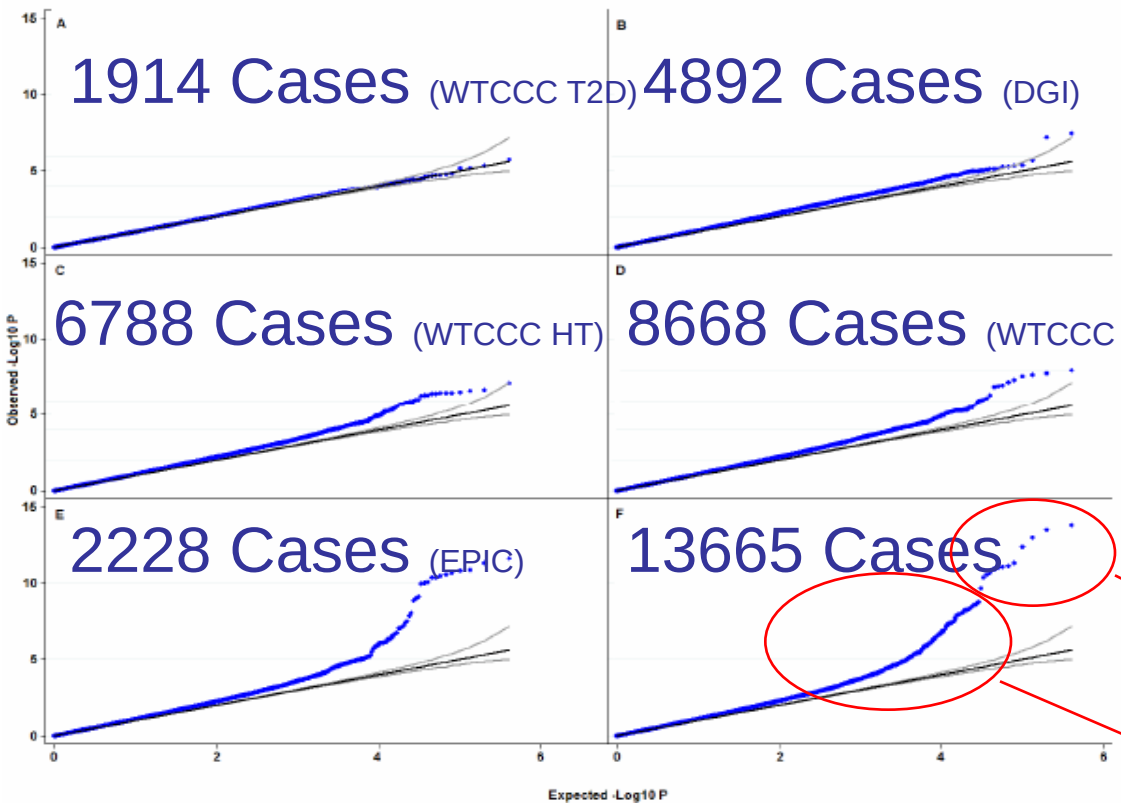
Naomi R Wray[1], Michael E Goddard[2] and Peter M Visscher[1]

Current Opinion in Genetics & Development 2008, 18:257–263

# A potted history of genetic studies



# GWA of Height




Significant result

Other loci?

- Large numbers are needed to detect QTLs !!!
- Collaboration is the name of the game !!!

Weedon et al. (in press) *Nat Genet*

# Usual paradigm of association studies

- Identify significant SNPs (not many)
- Genotype in independent replication samples
- Find causal variants to inform on biology

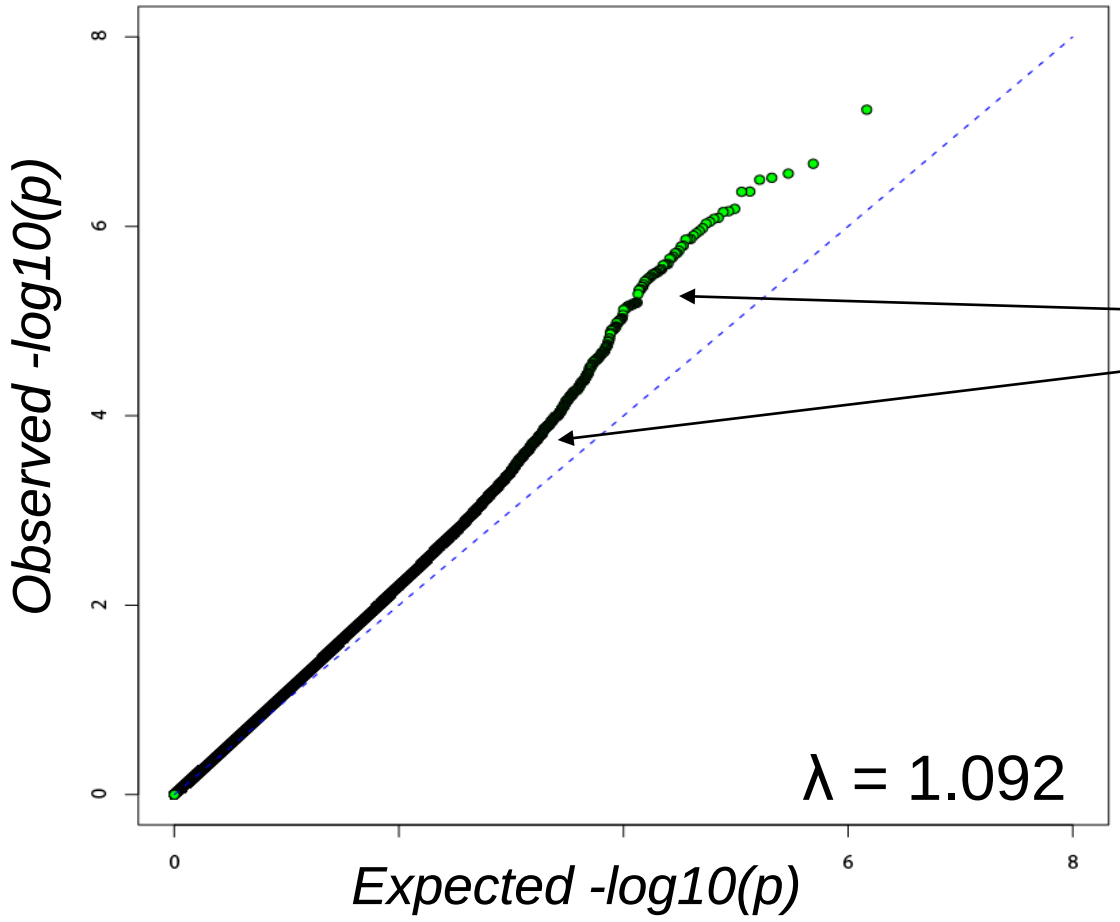
But only a small proportion of the known genetic variance is explained

What would we expect under a polygenic model of many, many variants each of small effect?

# A 5% risk increasing common variant

| | |
|---|---|
| Relative risk | 1.05 |
| Population MAF | 0.2 |
| MAF in cases | 0.2079 |
| MAF in controls | 0.1999 |
| Power $\alpha=1\times10^{-6}$<br>(*N*=10,000 cases, 10,000 controls) | 0.2% |
| Proportion ranked in top half<br>(*N*=ISC) | 72% |

# Schizophrenia (ISC) Q-Q plot



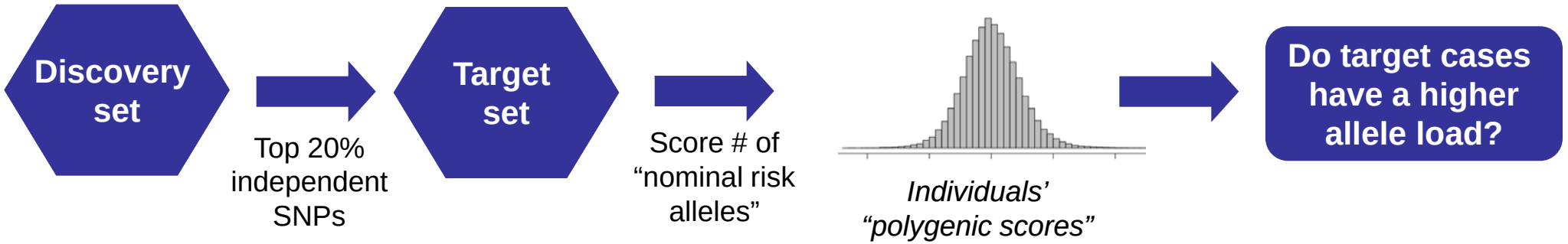

**Consistent with:**

*Stratification?*

*Genotyping bias?*

*Distribution of true polygenic effects?*

# Indexing polygenic variance with large sets of weakly associated alleles



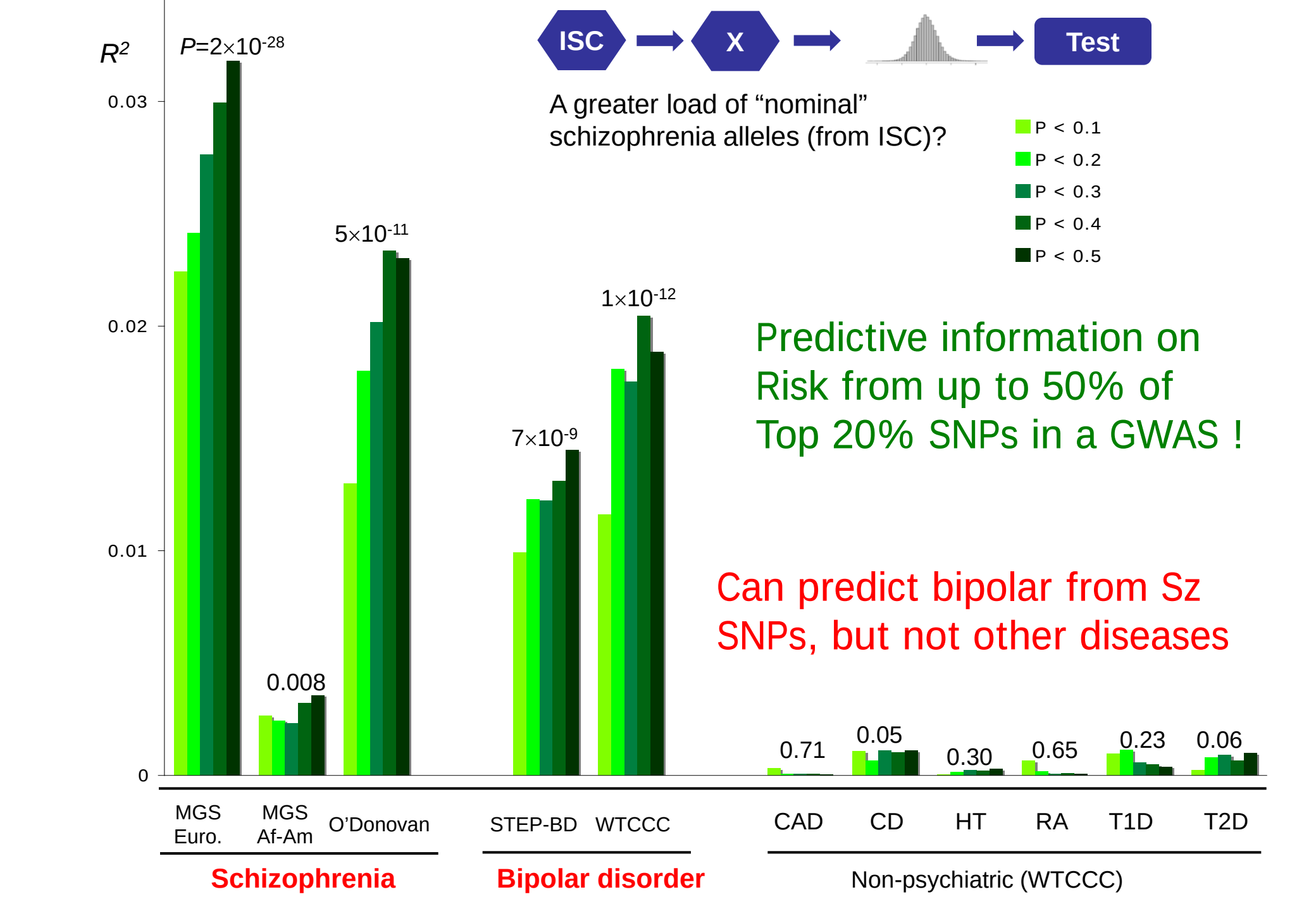

ISC → ISC

→ Independent SCZ studies (MGS, O'Donovan)

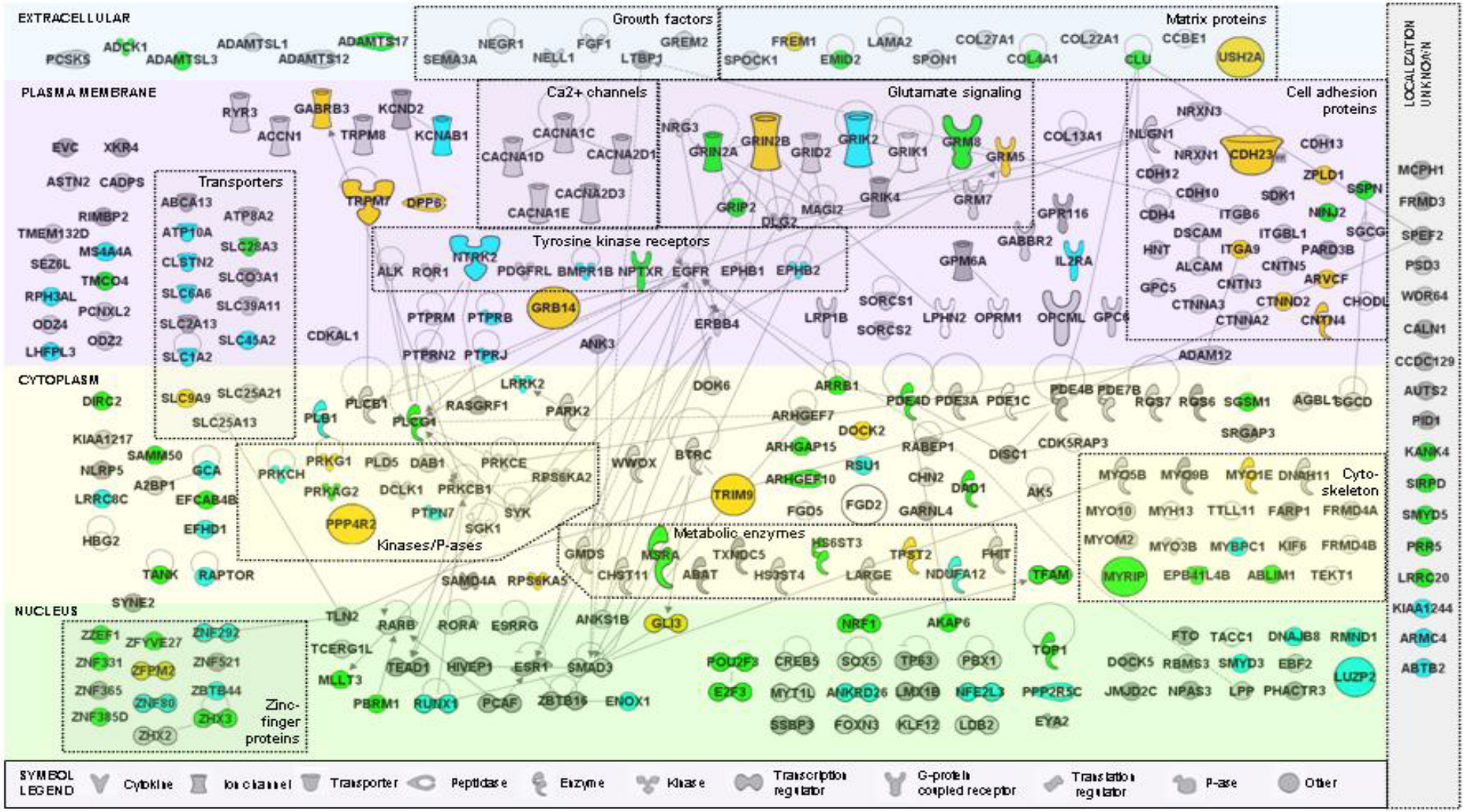
→ Bipolar disorder (STEP-BD, WTCCC)

→ Non-psychiatric disease (WTCCC)

**Douglas Levinson, Pablo Gejman,** Jianxin Shi and colleagues


ISC
X
Test
A greater load of “nominal” schizophrenia alleles (from ISC)?
P < 0.1
P < 0.2
P < 0.3
P < 0.4
P < 0.5
R2
0.03
0.02
0.01
0
P=2×10-28
0.008
5×10-11
7×10-9
1×10-12
0.71
0.05
0.30
0.65
0.23
0.06
Predictive information on Risk from up to 50% of Top 20% SNPs in a GWAS !
Can predict bipolar from Sz SNPs, but not other diseases
MGS Euro.
MGS Af-Am
O’Donovan
STEP-BD
WTCCC
CAD
CD
HT
RA
T1D
T2D
Schizophrenia
Bipolar disorder
Non-psychiatric (WTCCC)

# Pathway (Ingenuity) analysis of GWAS for smoking




Vink et al, 2009 *AJHG* in press

# Even for “simple” diseases the number of alleles is large

- Ischaemic heart disease (LDR) >190
- Breast cancer (BRAC1) >300
- Colorectal cancer (MLN1) >140

# Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol

Jonathan C. Cohen,[1,2,3*†] Robert S. Kiss,[5*]
Alexander Pertsemlidis,[1] Yves L. Marcel,[5†] Ruth McPherson,[5]
Helen H. Hobbs[1,3,4]

Heritable variation in complex traits is generally considered to be conferred by common DNA sequence polymorphisms. We tested whether rare DNA sequence variants collectively contribute to variation in plasma levels of high-density lipoprotein cholesterol (HDL-C). We sequenced three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) that cause Mendelian forms of low HDL-C levels in individuals from a population-based study. Nonsynonymous sequence variants were significantly more common (16% versus 2%) in individuals with low HDL-C (<fifth percentile) than in those with high HDL-C (>95th percentile). Similar findings were obtained in an independent population, and biochemical studies indicated that most sequence variants in the low HDL-C group were functionally important. Thus, rare alleles with major phenotypic effects contribute significantly to low plasma HDL-C levels in the general population.

Complex disease: common or rare alleles?

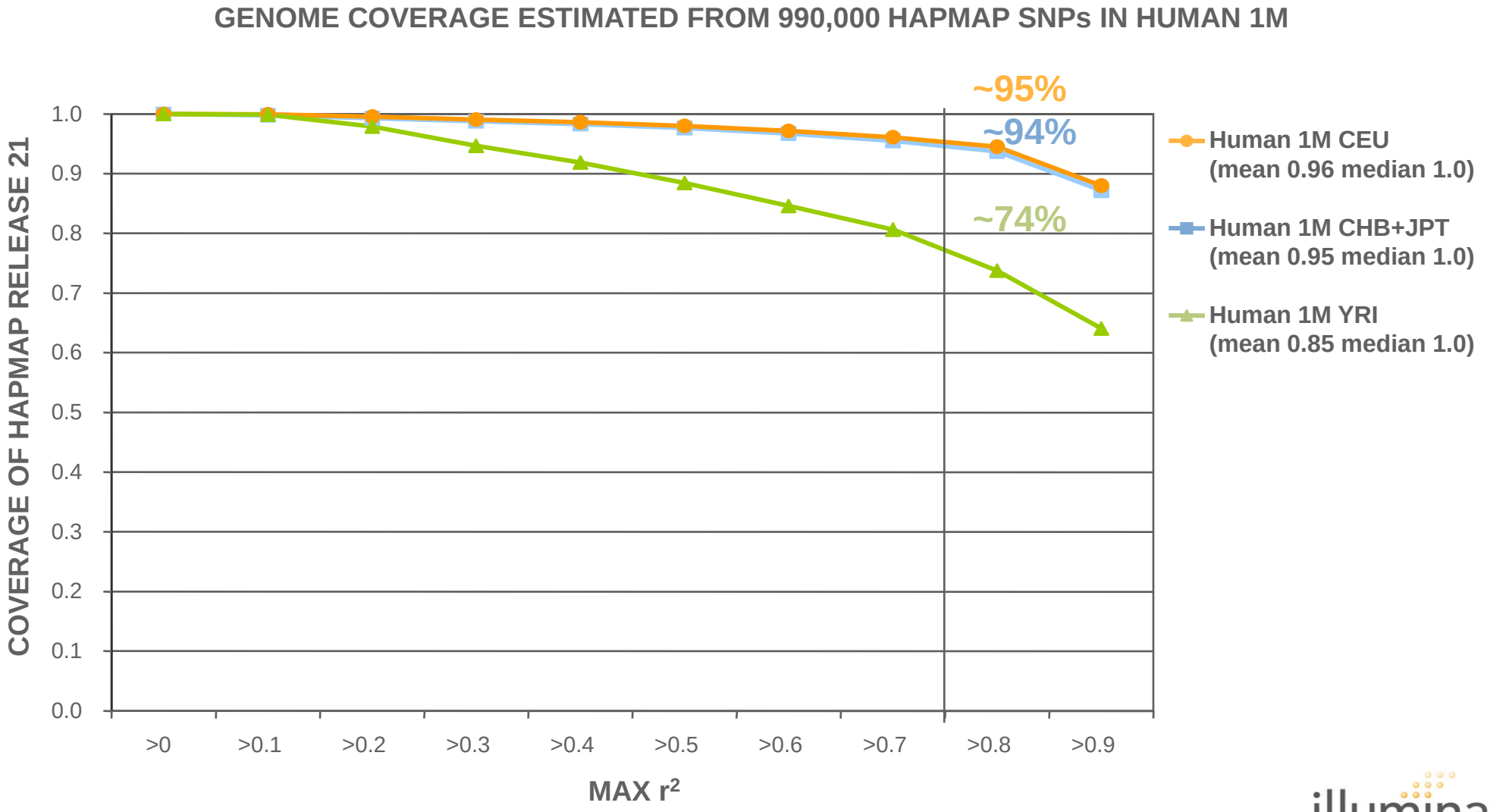
Increasing evidence for Common Disease – Rare Variant hypothesis (CDRV)

**Table 1.** Sequence variations in the coding regions of *ABCA1*, *APOA1*, and *LCAT*. Values represent the numbers of sequence variants identified in 256 individuals from the Dallas Heart Study (DHS) (128 with low HDL-C and 128 with high HDL-C) and 263 Canadians (155 with low HDL-C and 108 with high HDL-C) (*17*). NS, nonsynonymous (nucleotide substitutions resulting in an amino acid change); S, synonymous (coding sequence substitutions that do not result in an amino acid change). GenBank accession numbers for DHS *ABCA1*, *APOA1*, and *LCAT* sequences are NM_005502, NM_000039, and NM_000229, respectively.

| | Sequence variants unique to one group | | | | Sequence variants common to both groups | |
|---|---|---|---|---|---|---|
| | Low HDL-C | | High HDL-C | | | |
| | NS | S | NS | S | NS | S |
| | | | DHS | | | |
| *ABCA1* | 14 | 6 | 2 | 5 | 10 | 19 |
| *APOA1* | 1 | 0 | 0 | 1 | 0 | 1 |
| *LCAT* | 0 | 1 | 1 | 0 | 1 | 1 |
| | | | Canadians | | | |
| *ABCA1* | 14 | 2 | 2 | 3 | 7 | 5 |
| *APOA1* | 0 | 1 | 0 | 0 | 2 | 0 |
| *LCAT* | 6 | 1 | 0 | 0 | 0 | 0 |

[Science 2004]

# Human 1M HapMap Coverage by Population



products & services

- overview
- systems & software
- dna analysis solutions
- rna analysis solutions
- solexa applications
- services
- product literature

- print this page

## solexa sequencing applications

Illumina's Solexa Sequencing technology offers a powerful new approach to some of today's most important applications for genetic analysis and functional genomics, including:

### sequencing and resequencing

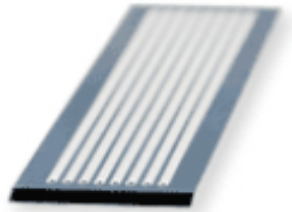
Whether you need to sequence an entire genome or a large candidate region, the Illumina Genome Analyzer System is today's most productive and economical sequencing tool. Solexa sequencing technology and reversable terminator chemistry deliver unprecedented volumes of high quality data, rapidly and economically.

### expression profiling

Sequencing millions of short cDNA tags per sample, the Genome Analyzer allows you to generate digital expression profiles at costs comparable to current analog methods. Because our protocol does not require any transcript-specific probes, you can apply the technology to discover and quantitate transcripts in any organisms, irrespective of the annotation available on the organism.

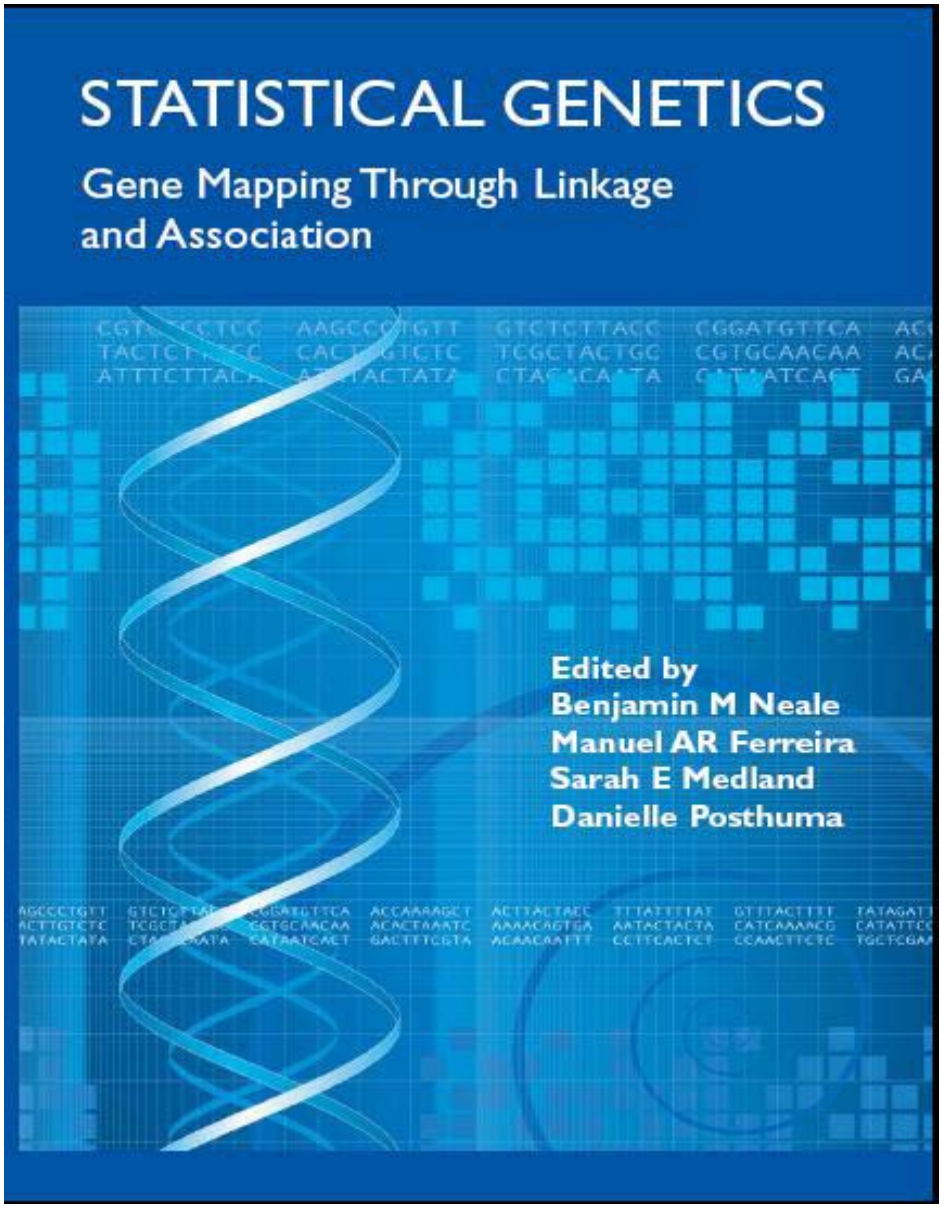
### small rna identification and quantification

Solexa sequencing technology also offers a unique and powerful solution for the comprehensive discovery and characterization of small RNAs in a wide range of species. The massively parallel sequencing protocol allows researchers to discover and analyze genome-wide profiles of small RNA in any species. With the potential to generate several million sequence tags economically, the Illumina Genome Analyzer offers investigators the opportunity to uncover global profiles of small RNA at an unprecedented scale.



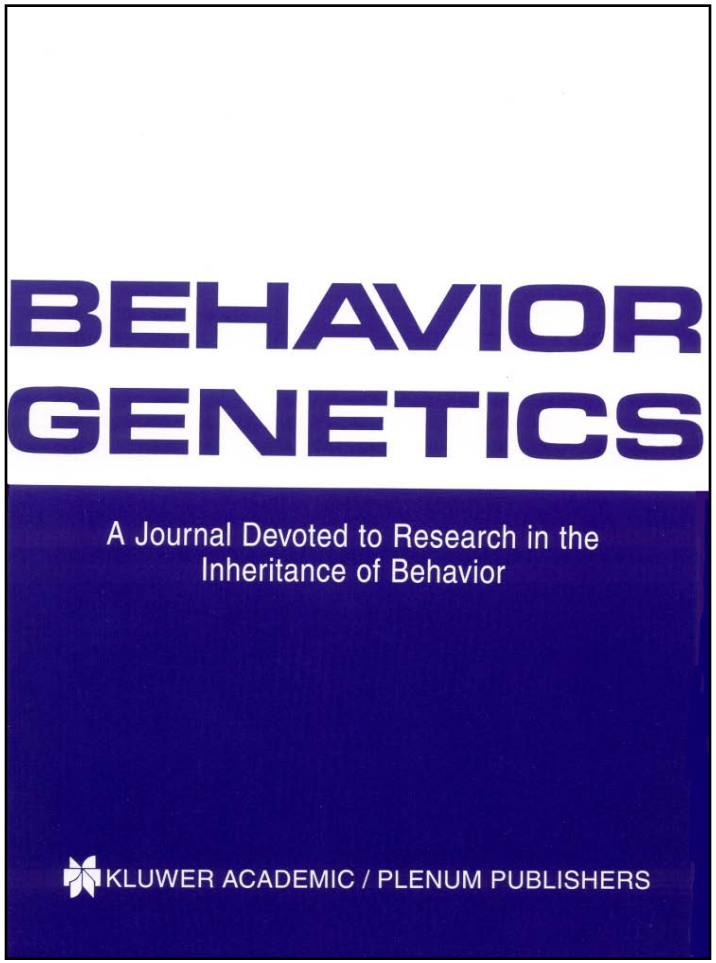
important information

- product literature
- publications
- faqs
- have a rep contact me




http://www.genemapping.org/

# We also run two journals (1)




- Editor: John Hewitt
- Editorial assistant Christina Hewitt
- Publisher: Kluwer /Plenum
- Fully online
- http://www.bga.org

# We also run two journals (2)


- Editor: Nick Martin
- Editorial assistant + subscriptions: Lorin Grey
- Publisher: Australian Academic Press
- Fully online
- http://www.ists.qimr.edu.au/journal.html