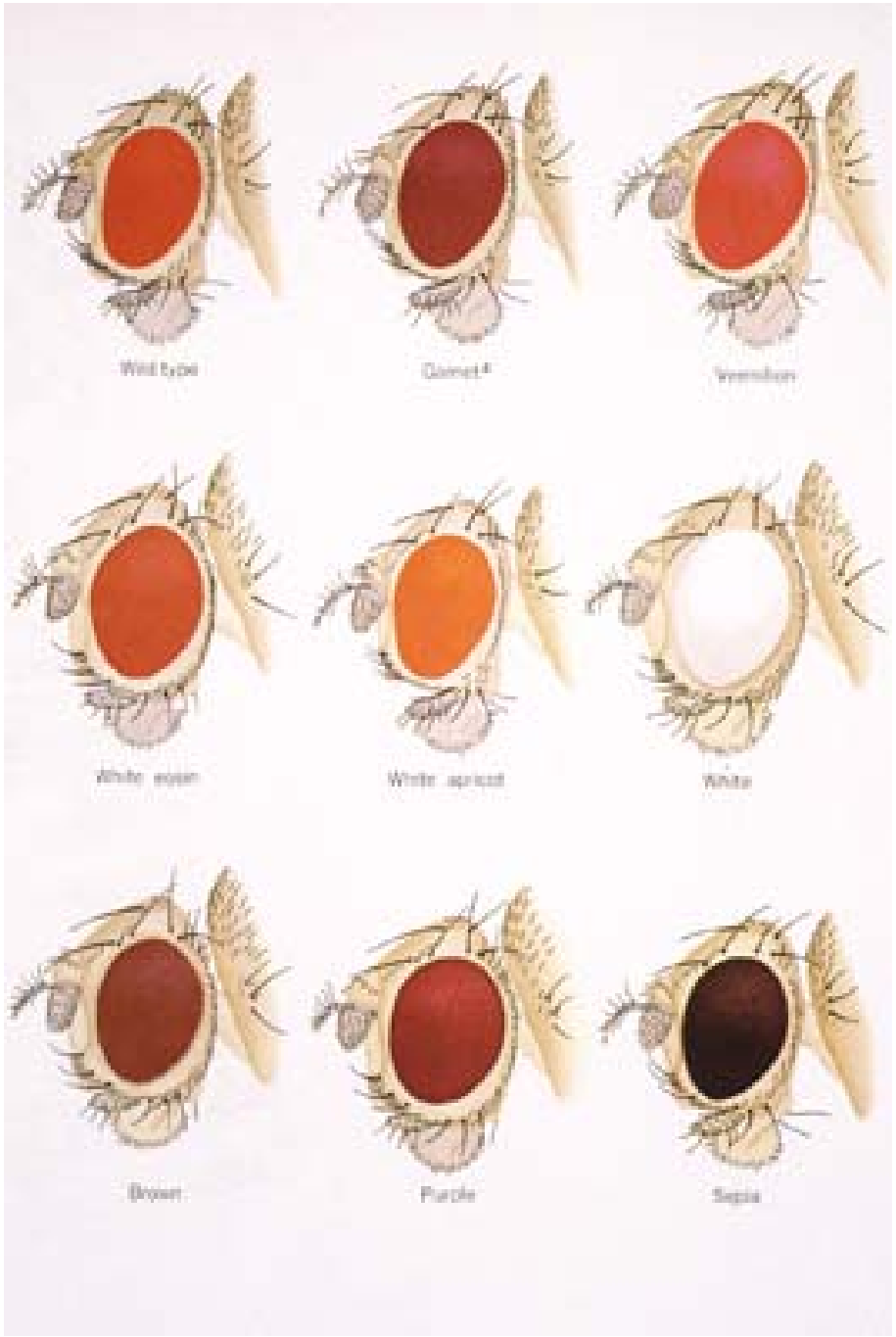
QTL studies: past, present and future

Nick Martin

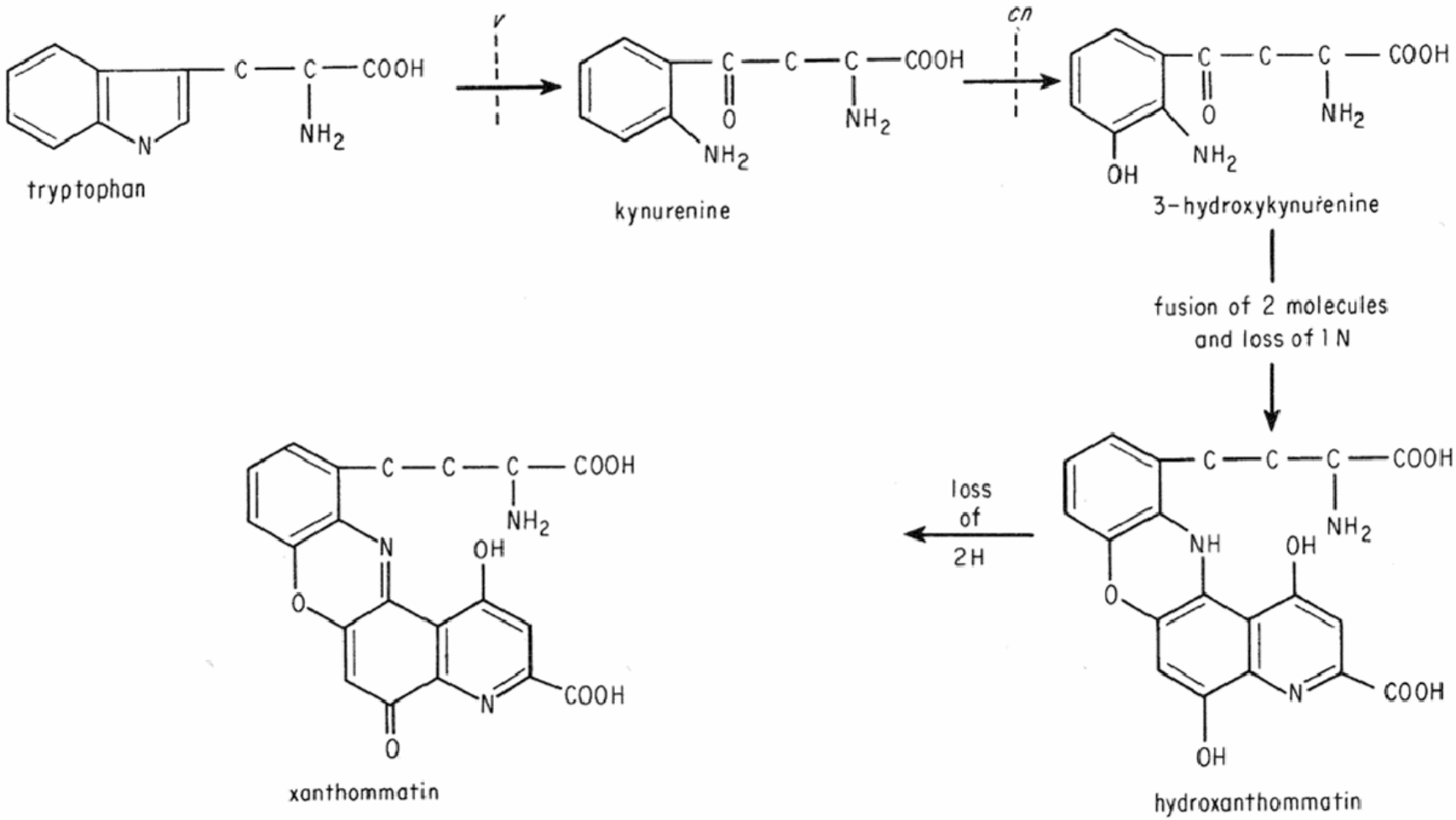
Queensland Institute of Medical Research

Boulder workshop: March 10, 2006



Using genetics to dissect metabolic pathways: Drosophila eye color

Beadle & Ephrussi, 1936



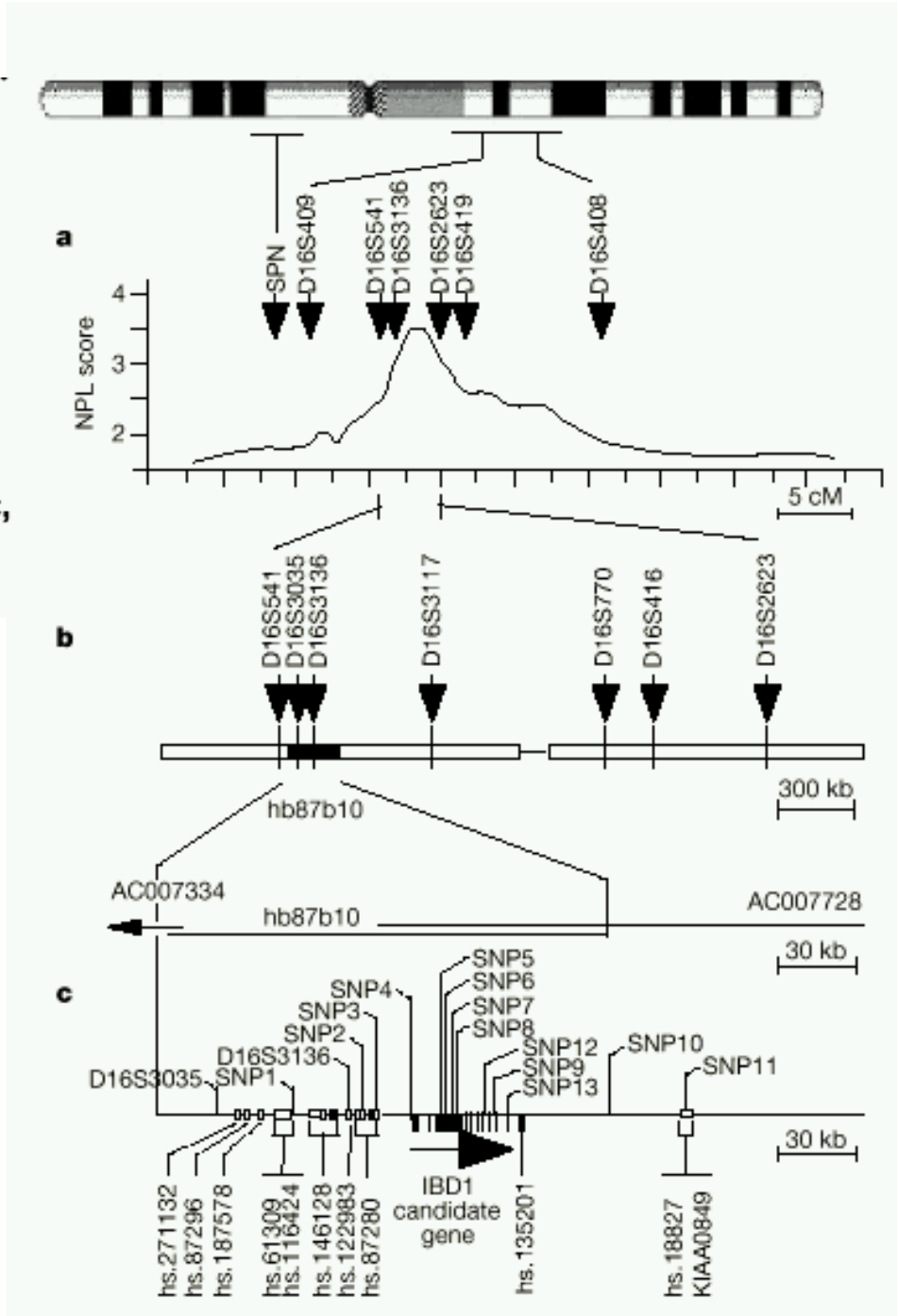
Beadle and Ephrussi, 1936

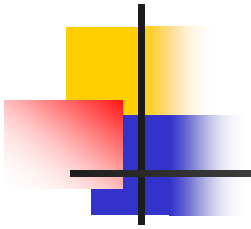
Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

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 !

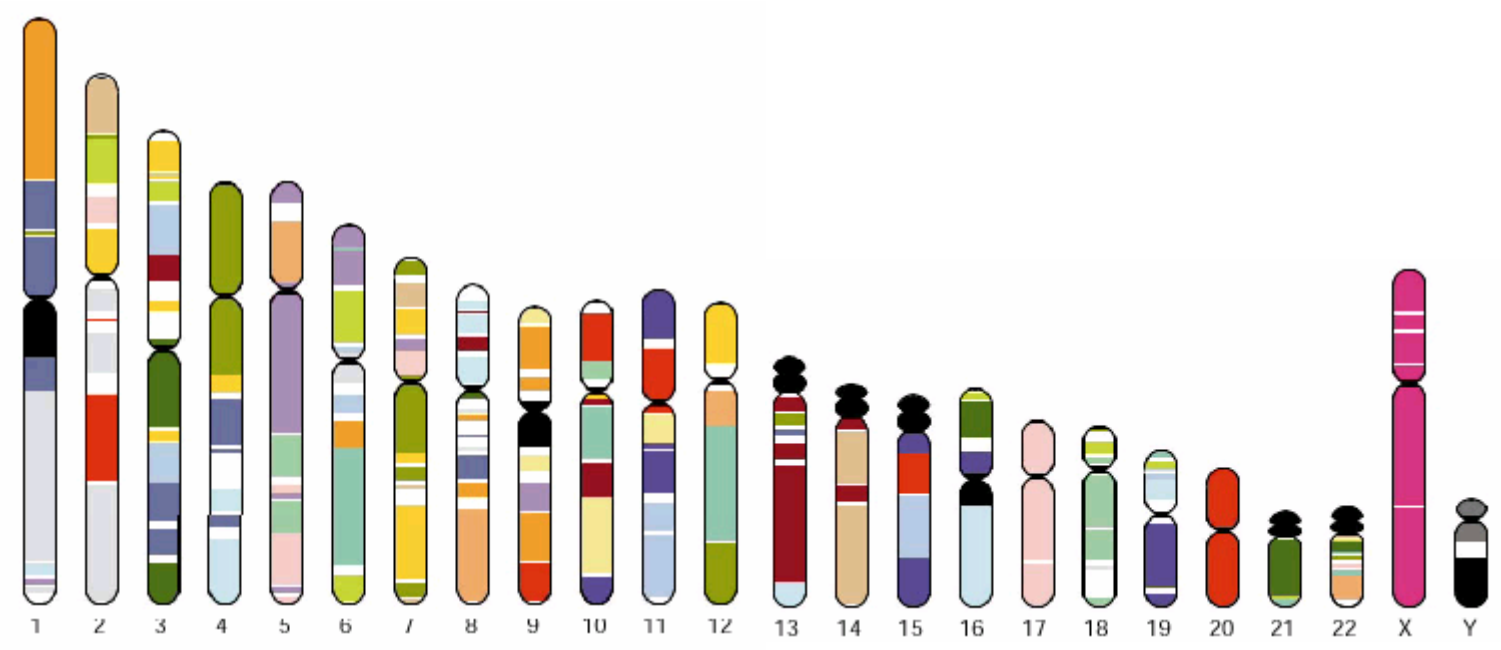




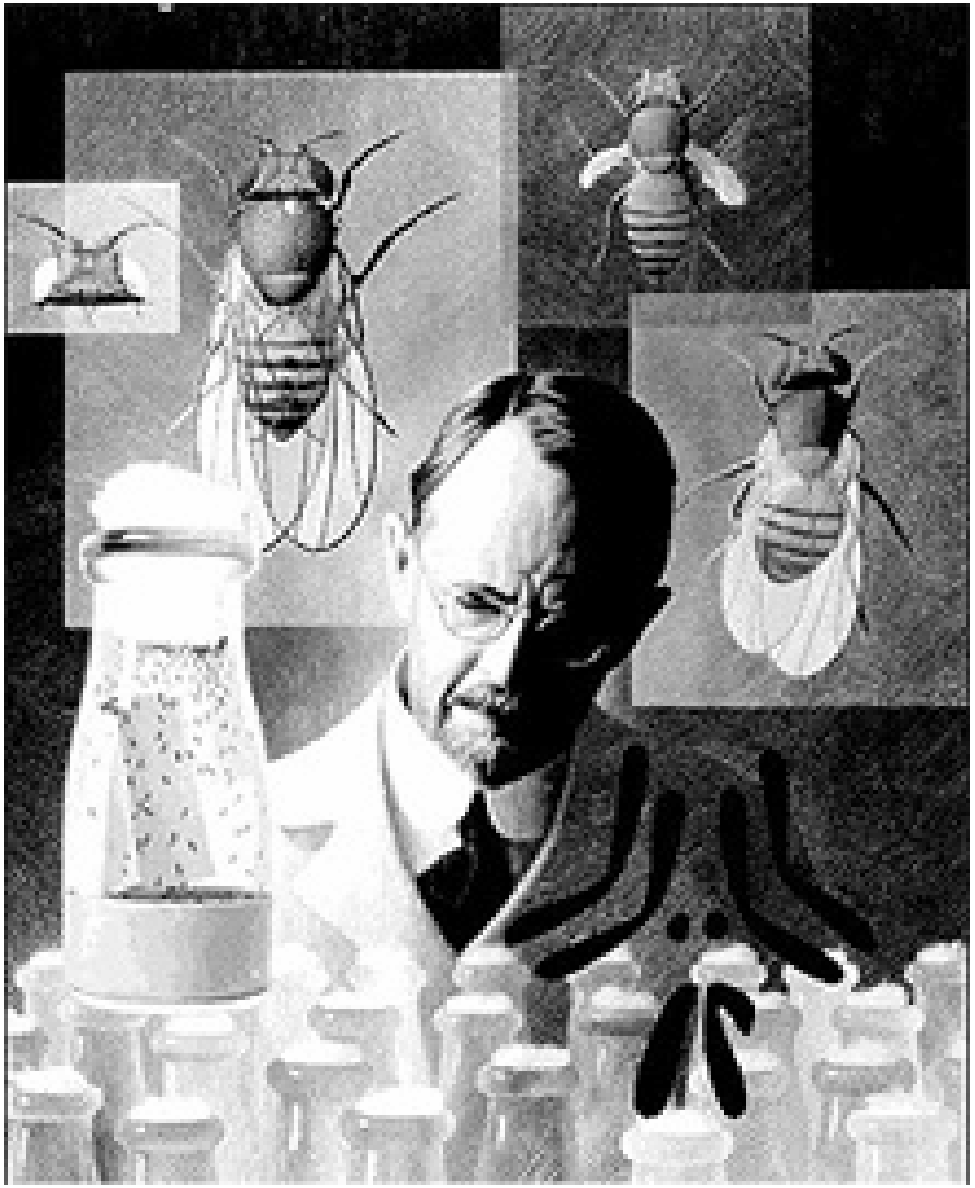
Finding QTLs

- Linkage
- Association

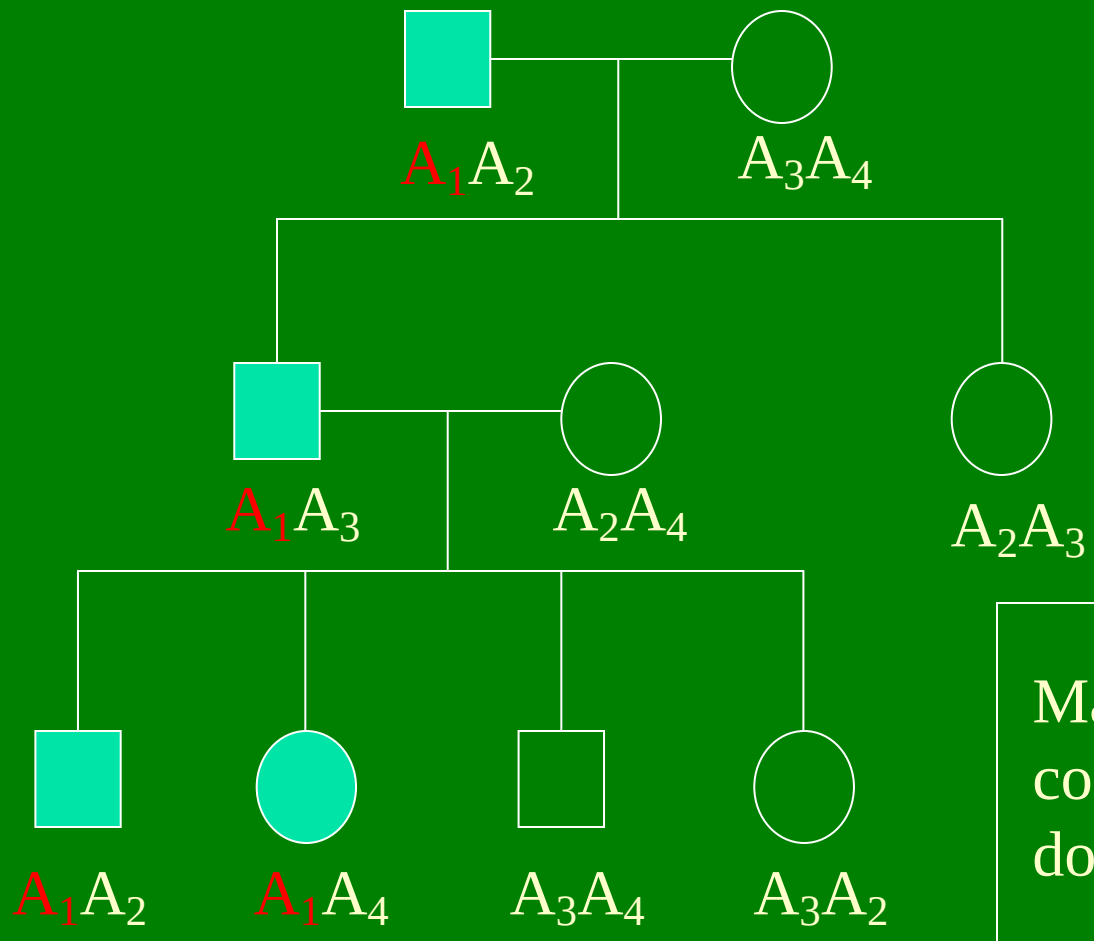
Linkage analysis



Thomas Hunt Morgan – discoverer of linkage

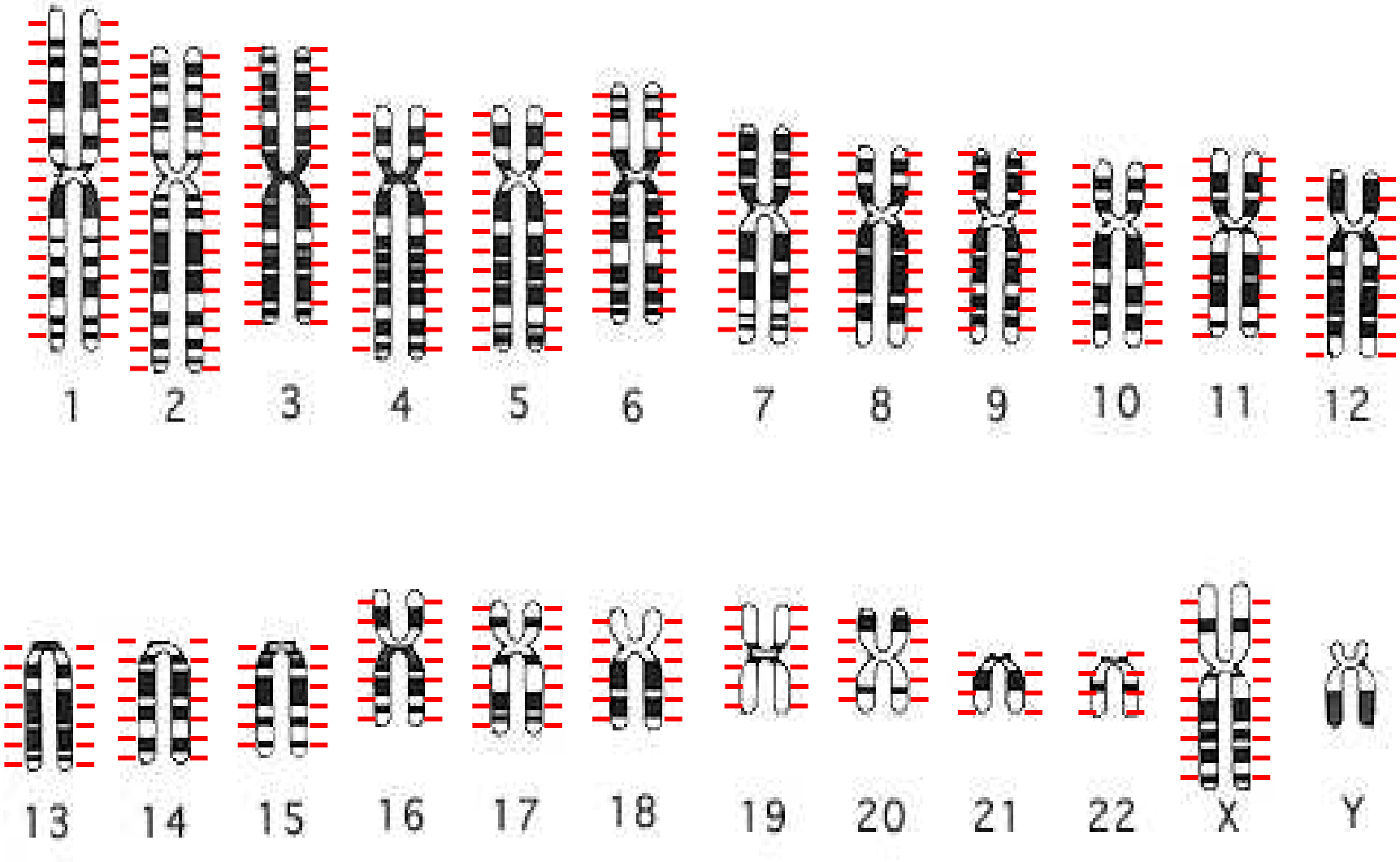


Linkage = Co-segregation

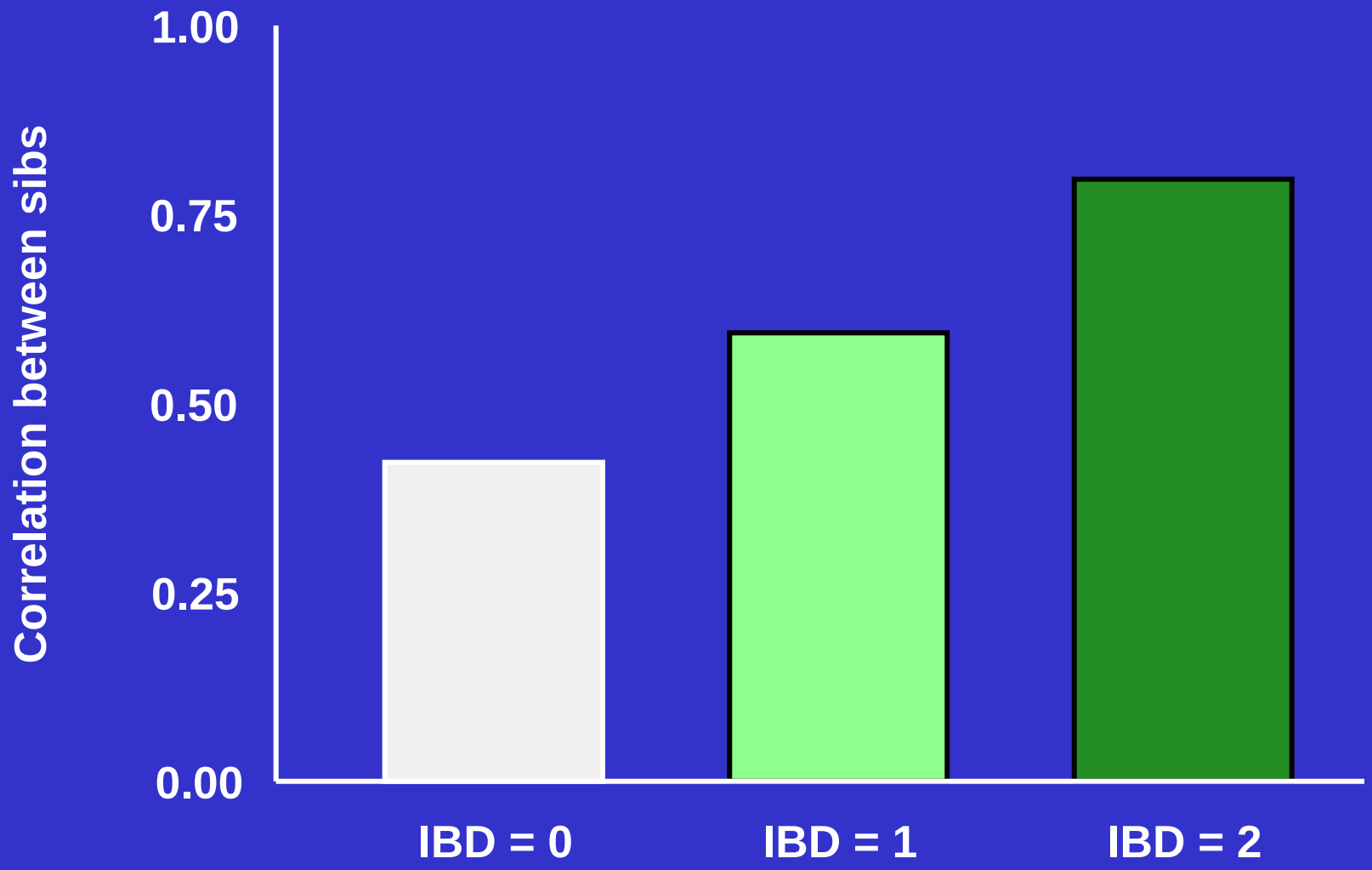


Marker allele A_1 cosegregates with dominant disease

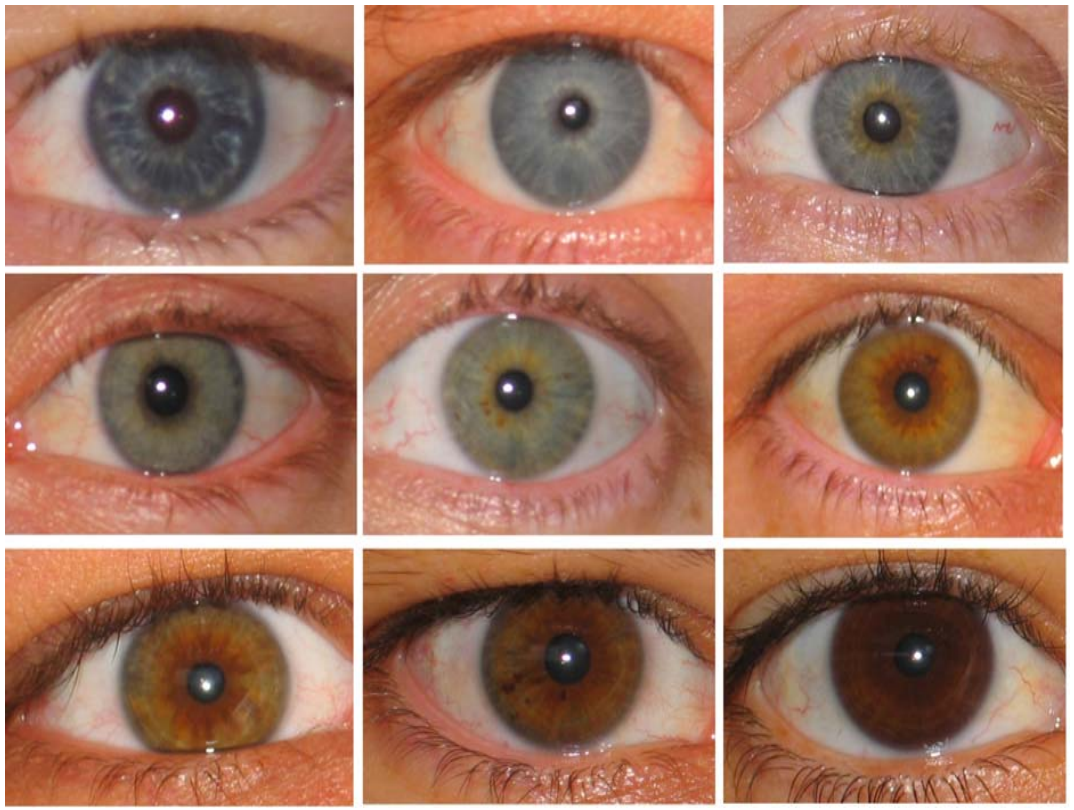
Linkage Markers...

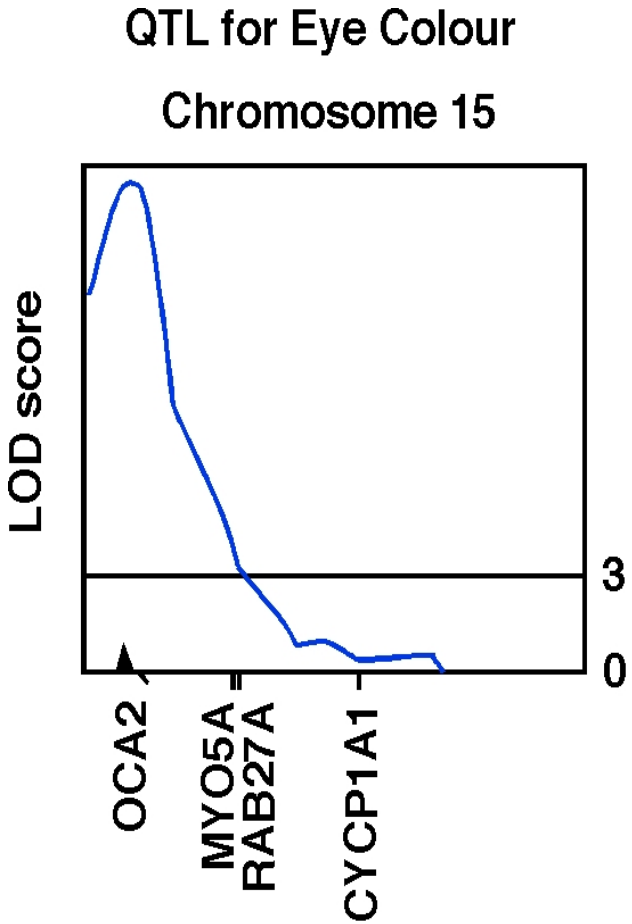


For continuous measures
Unselected sib pairs



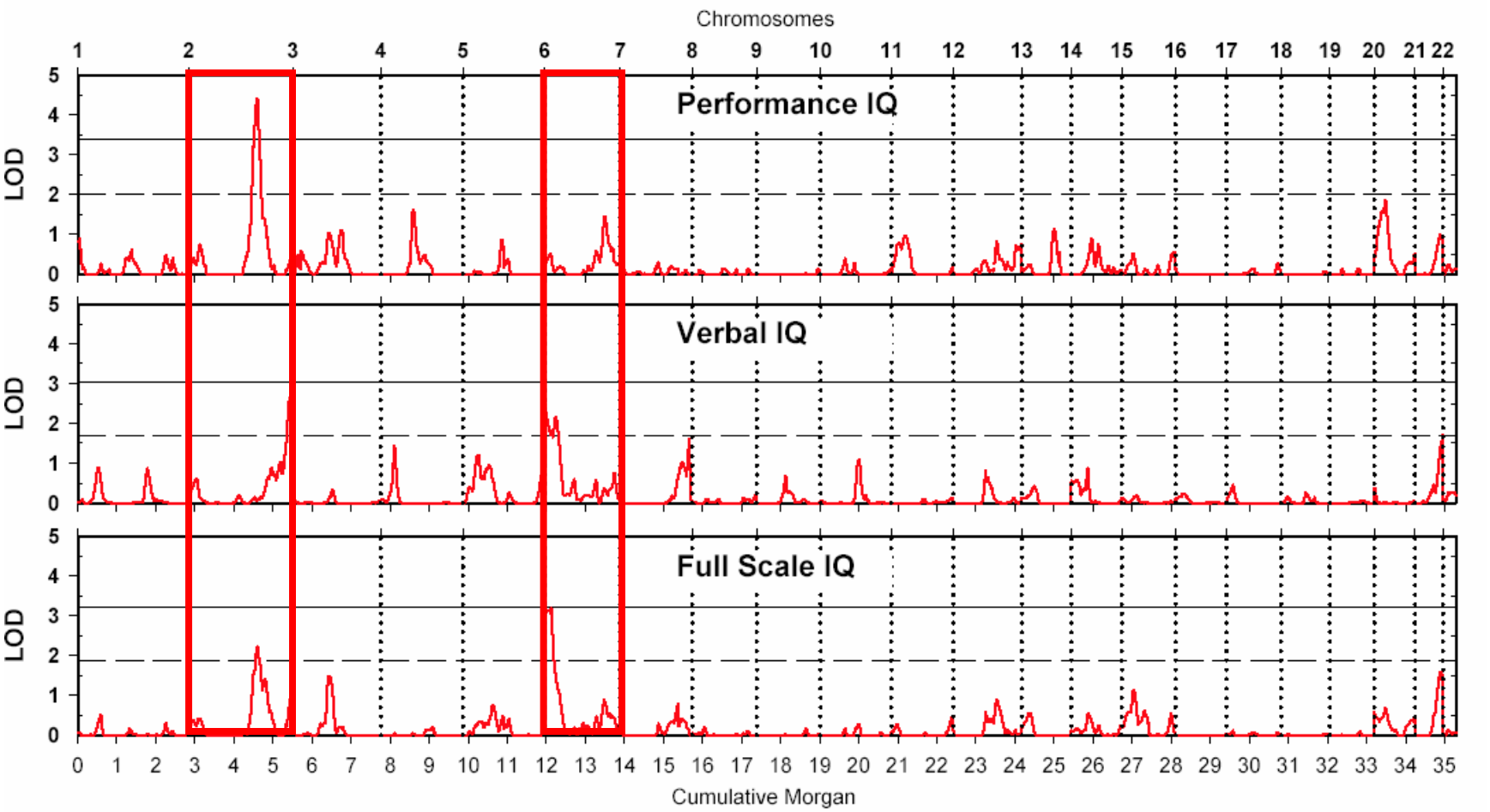
Human OCA2 and eye colour



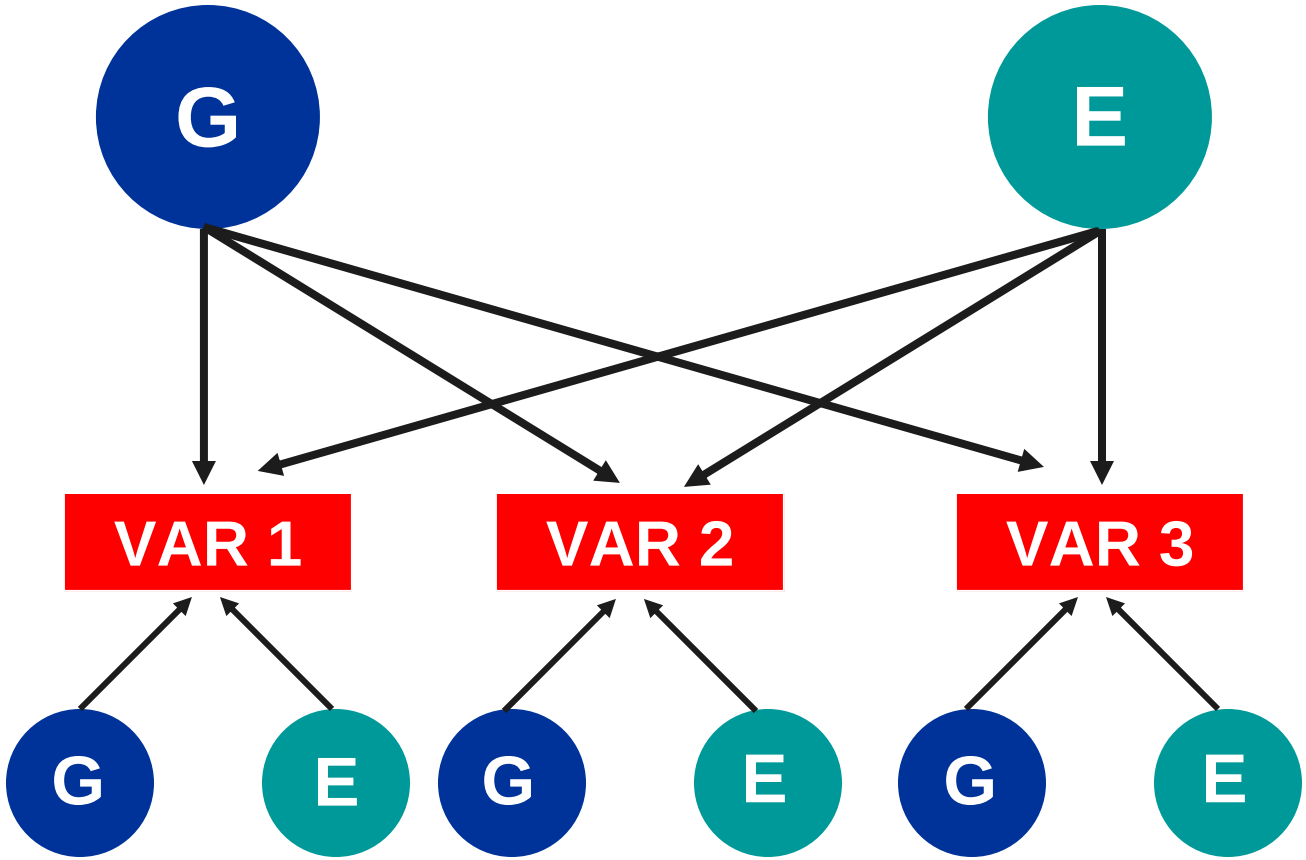


Zhu et al., *Twin Research* 7:197-210 (2004)

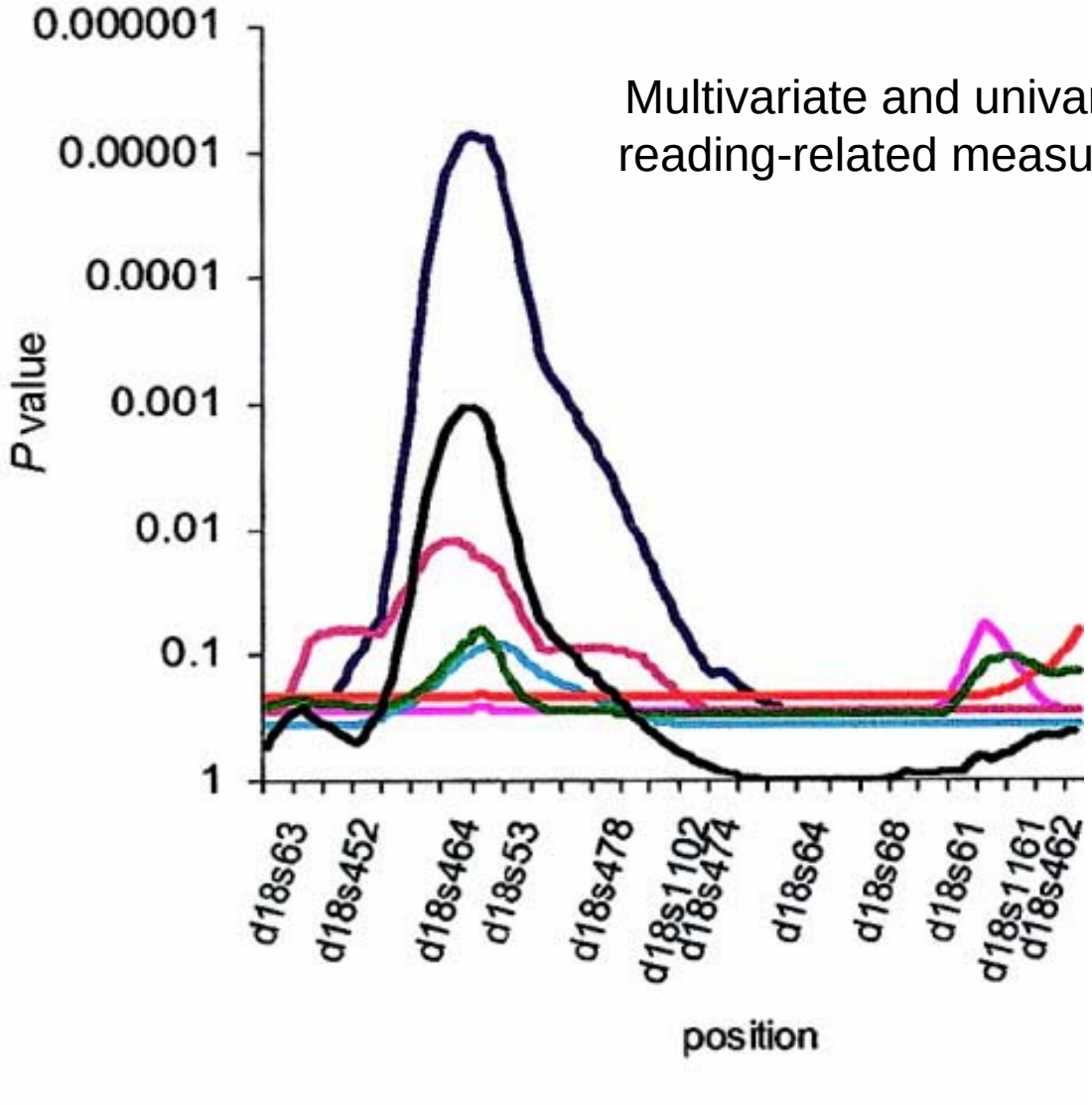
First genome-wide linkage scan for Intelligence



Posthuma et al., AJHG, in press


G
E
VAR 1
VAR 2
VAR 3
G
E
G
E
G
E

Effect of multivariate analysis on linkage power

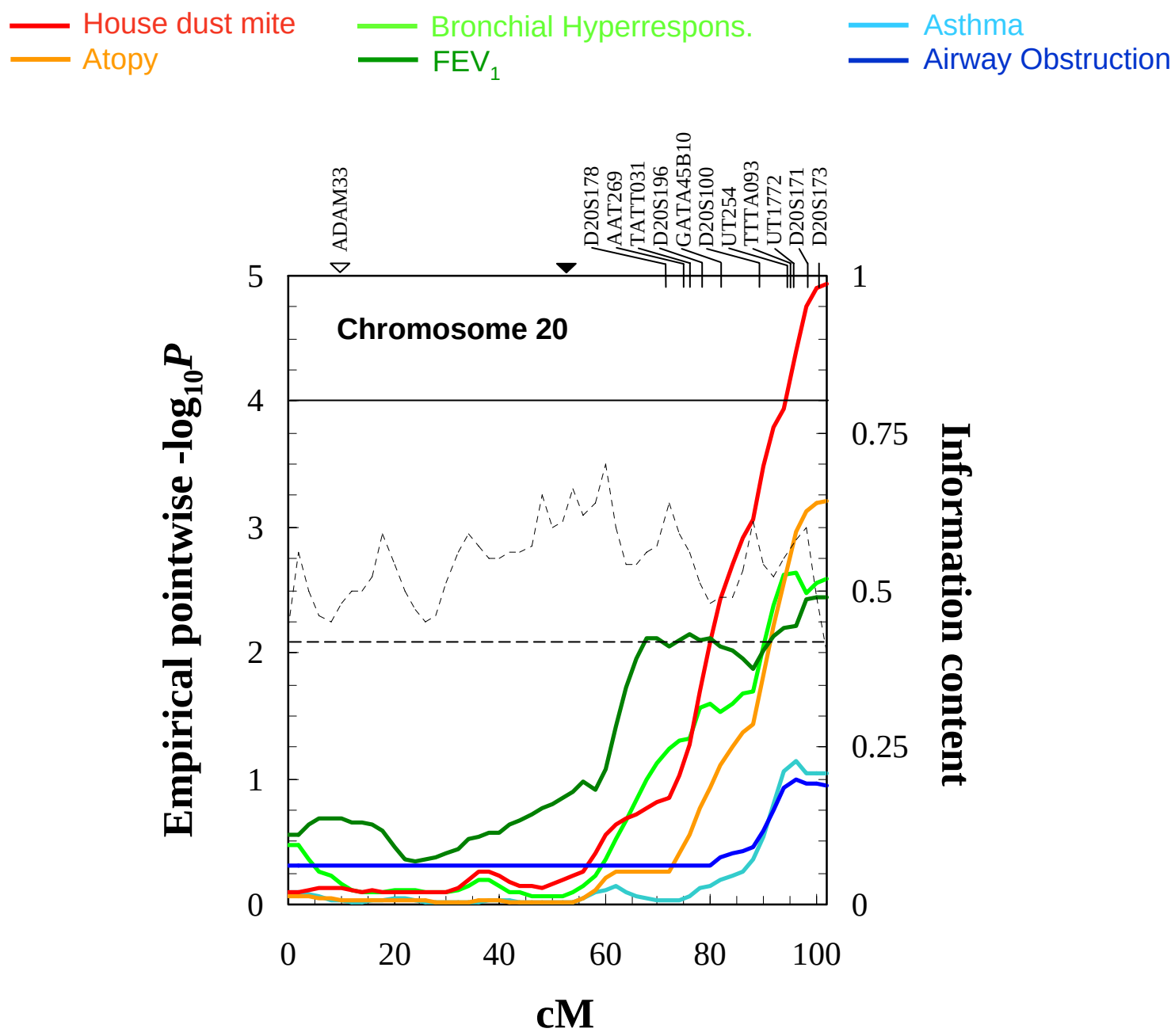


Multivariate and univariate linkage analysis of six reading-related measures on chromosome 18

Am. J. Hum. Genet., 72:561-570, 2003

Use of Multivariate Linkage Analysis for Dissection of a Complex Cognitive Trait

Angela Marlow, Simon Fisher, Clyde Francks, Laurence MacPhie, Stacey Cherny, Alex Richardson, Joel Talcott, John Stein, Anthony Monaco, and Lon Cardon


House dust mite
Atopy
Bronchial Hyperrespons.
FEV1
Asthma
Airway Obstruction
ADAM33
D20S178
AAT269
TATT031
D20S196
GATA45B10
D20S100
UT254
TTTA093
UT1772
D20S171
D20S173
Chromosome 20
Empirical pointwise -log10P
Information content
cM
0
20
40
60
80
100
1
2
3
4
5
0.25
0.5
0.75

A simple method to localise pleiotropic or multiple clustered quantitative trait loci using univariate linkage analyses of correlated traits

Manuel A. R. Ferreira, Peter M. Visscher, Nicholas G. Martin and David L. Duffy

Queensland Institute of Medical Research, Brisbane, Australia.

European Journal of Human Genetics (***almost*** *in press)*

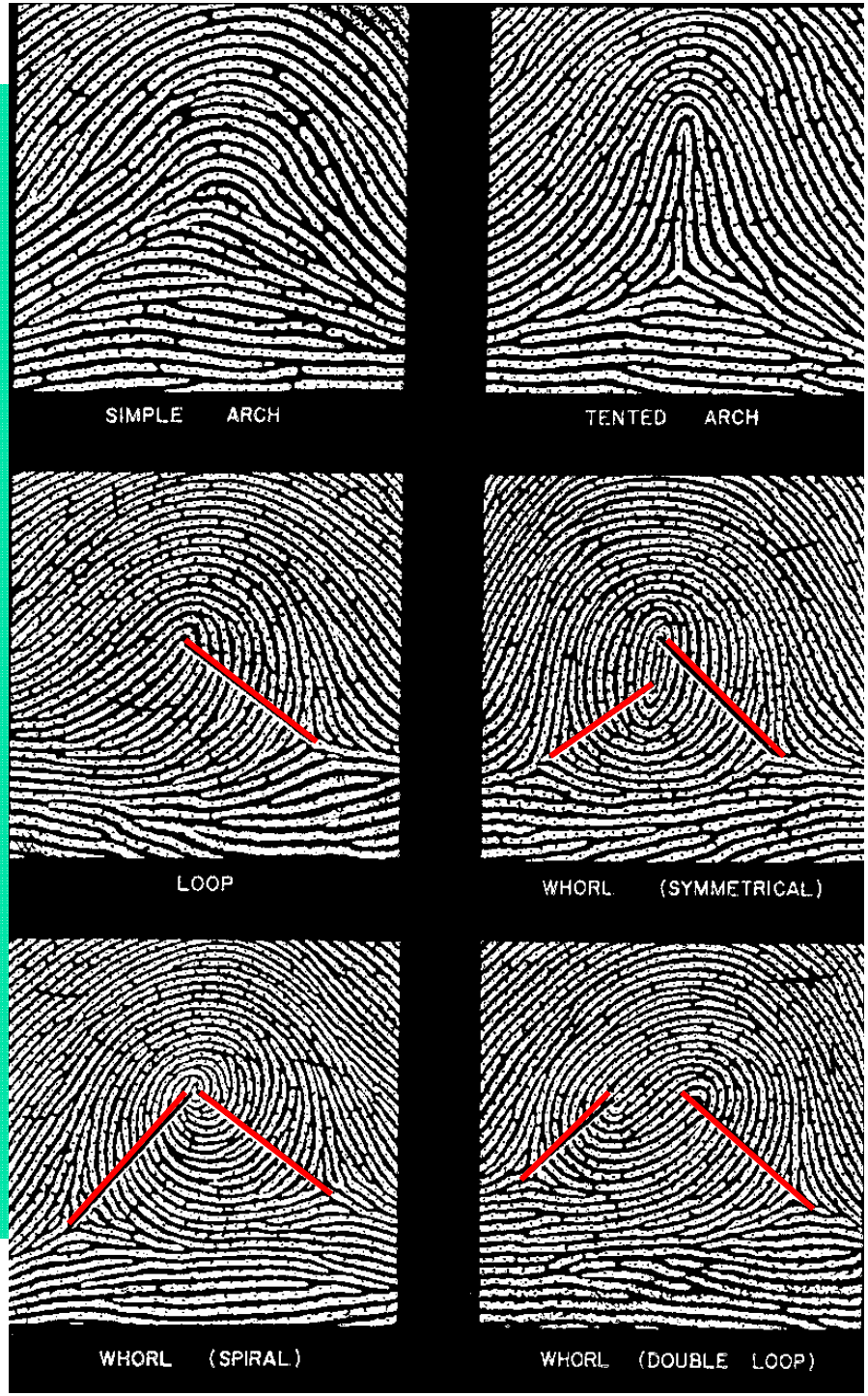
Ridge count

The size of prints can be measured by counting the number of ridges from the triradii to the core

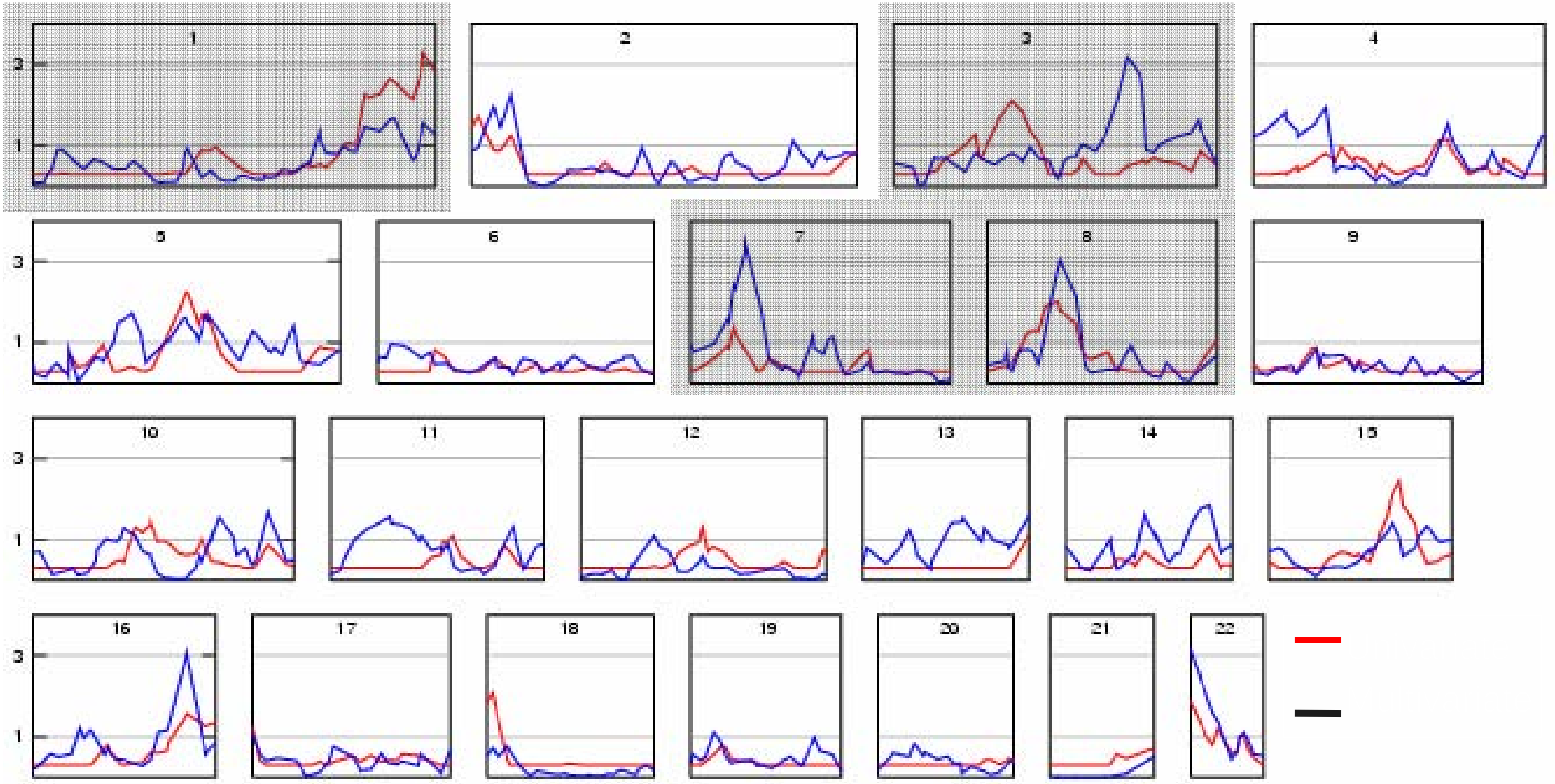
Ridge count can be summed over all fingers to give a total ridge count

Diagram from Holt, 1968

- Highly heritable:
  - MZ r = .94 CI .89 - .96
  - DZ r = .42 CI .34 - .50
  - A .82 CI .56 - .95
  - D .11 CI .00 - .37
  - E .07 CI .05 - .10



TRC vs Multivariate ($-LOG_{10}p$)

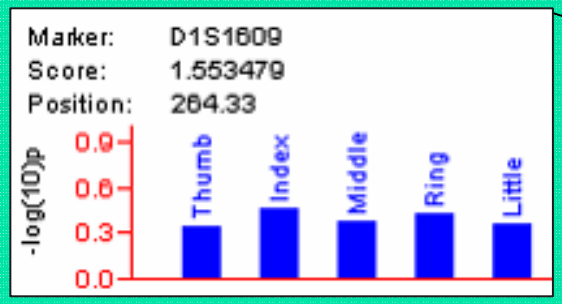


χ_1^2 χ_5^2

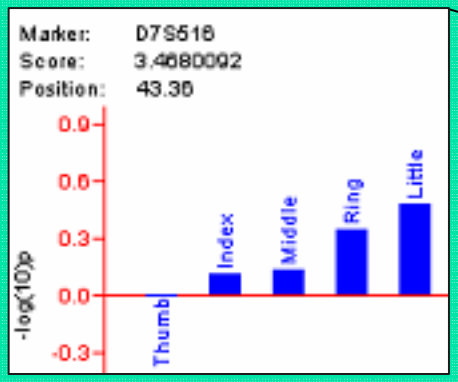
Chromosome 1

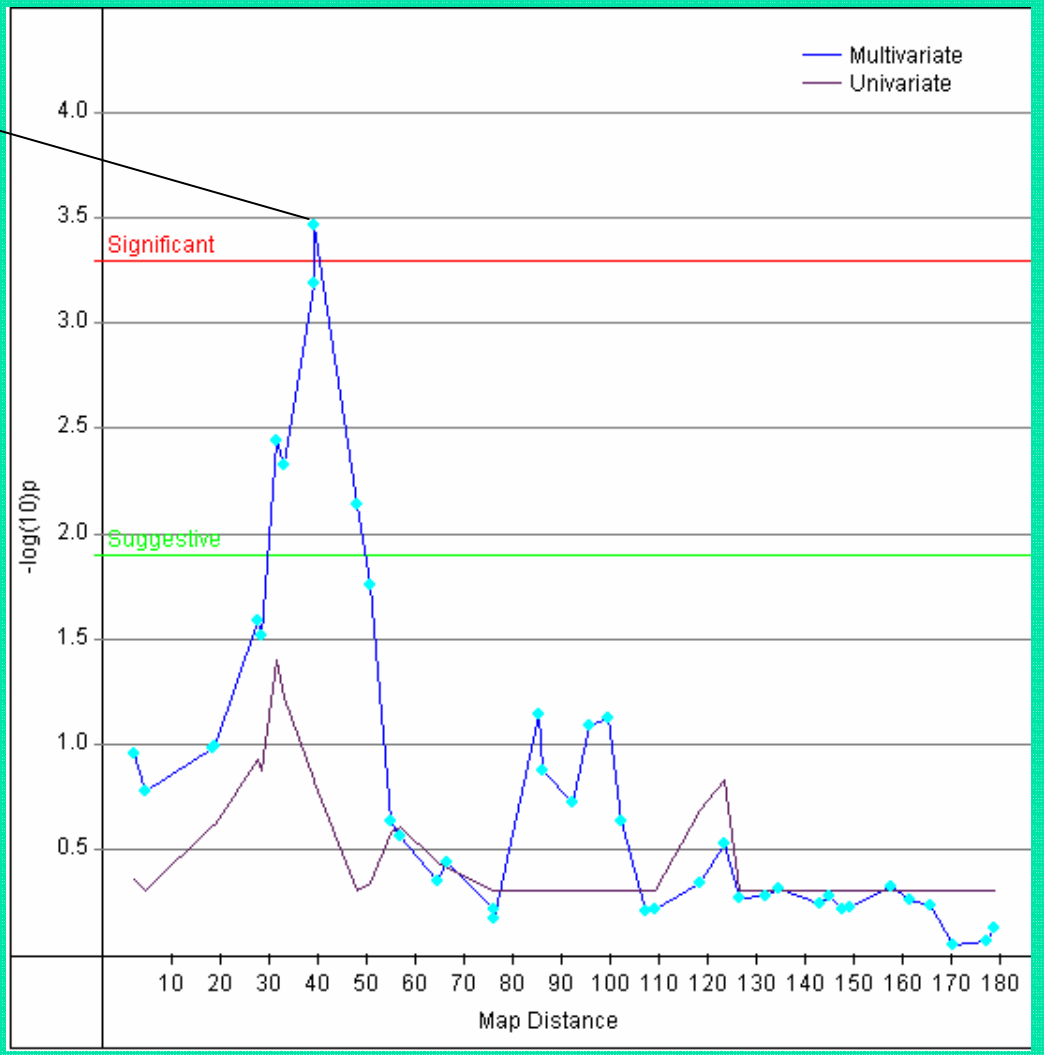
Similar ‘drop chi-squares’ for pleiotropic QTLs

Resulting in a very conservative test



Chromosome 7 ...



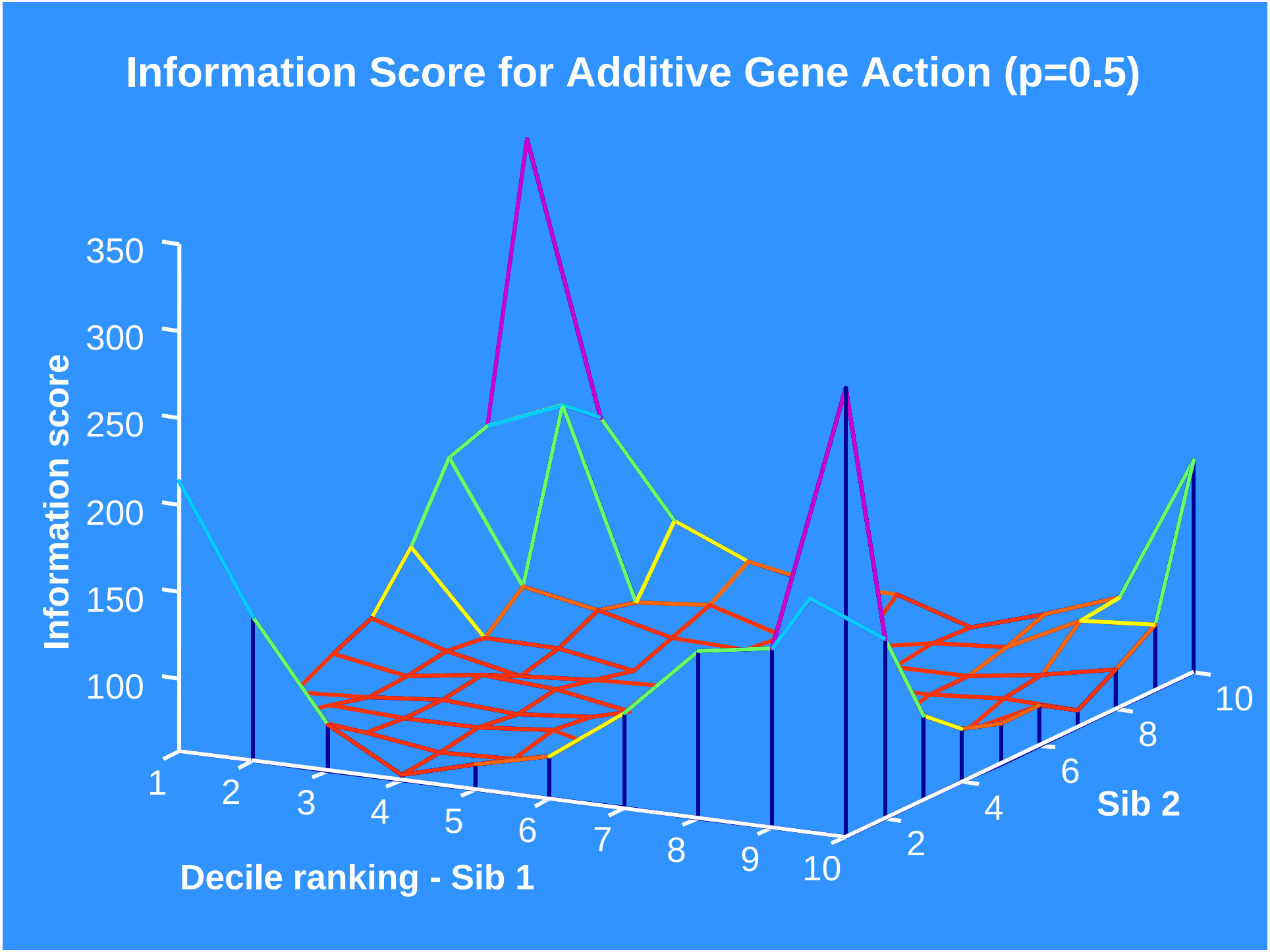


Evidence of developmental fields?

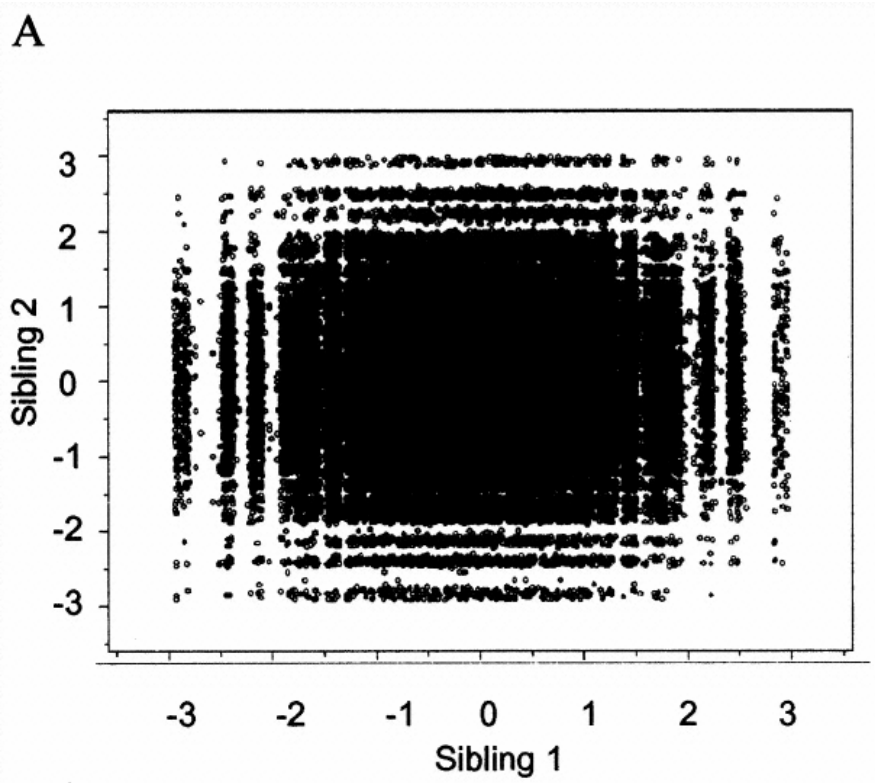
Extreme Discordant Sib Pairs for Mapping Quantitative Trait Loci in Humans

Neil Risch* and Heping Zhang

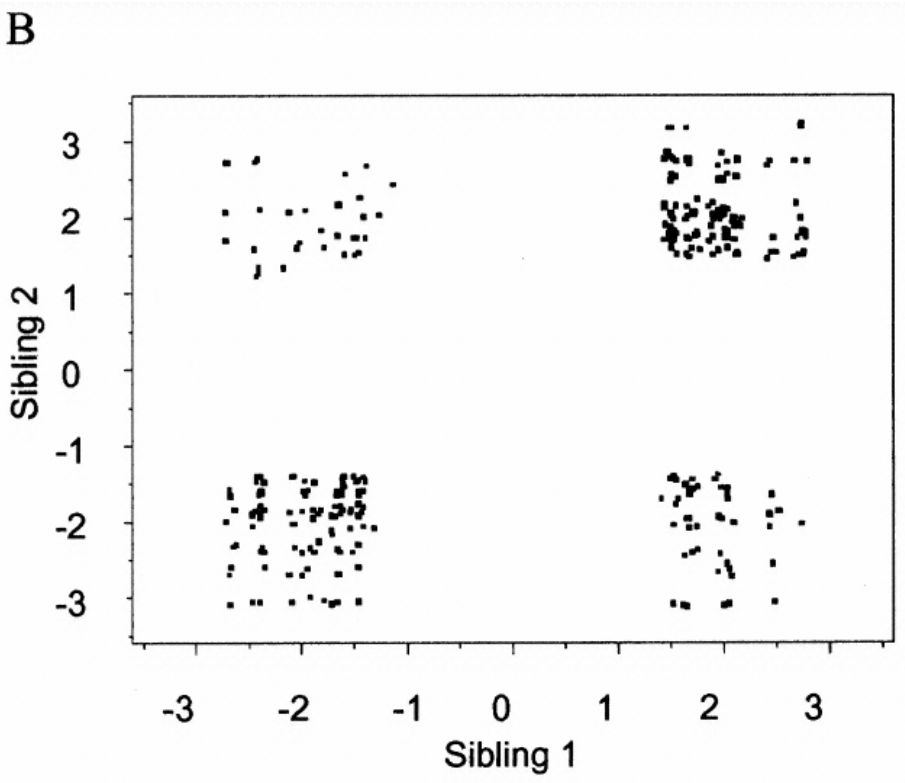
Analysis of differences between siblings (sib pair analysis) is a standard method of genetic linkage analysis for mapping quantitative trait loci, such as those contributing to hypertension and obesity, in humans. In traditional designs, pairs are selected at random or with one sib having an extreme trait value. The majority of such pairs provide little power to detect linkage; only pairs that are concordant for high values, low values, or extremely discordant pairs (for example, one in the top 10 percent and the other in the bottom 10 percent of the distribution) provide substantial power. Focus on discordant pairs can reduce the amount of genotyping necessary over conventional designs by 10- to 40 -fold.


Information Score for Additive Gene Action (p=0.5)
Information score
350
300
250
200
150
100
1
2
3
4
5
6
7
8
9
10
Decile ranking - Sib 1
2
4
6
8
10
Sib 2

Scatterplots of the distribution of neuroticism scores for each sibling pair

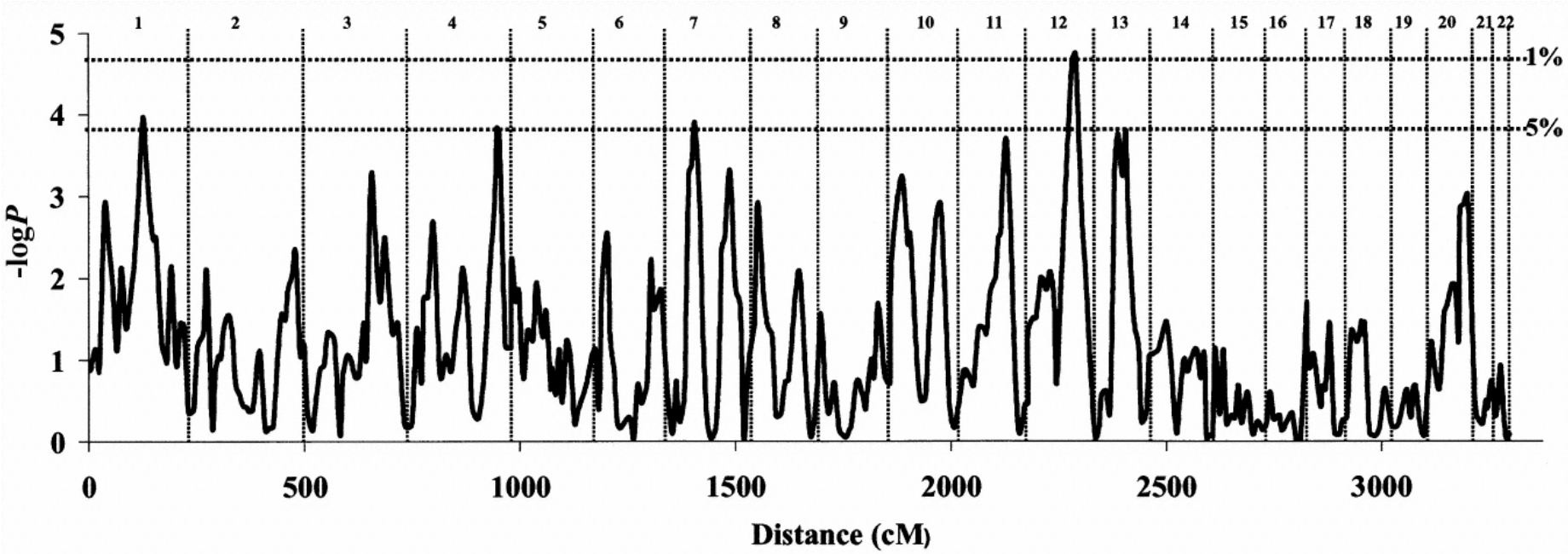


Distribution of entire sample



Distribution of selected sample

Multipoint linkage analysis of the genome for individual variation in neuroticism



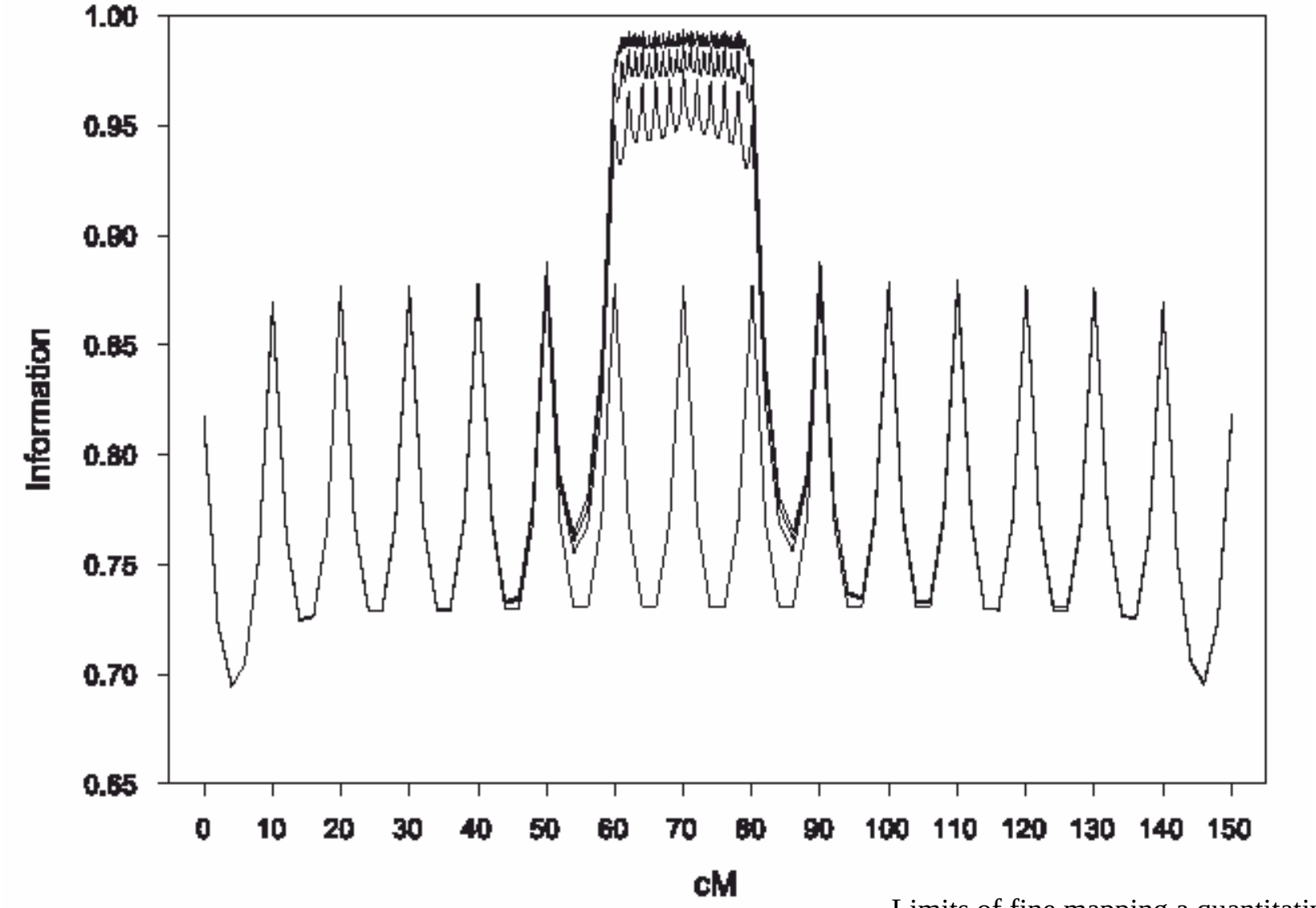
The -logP values (vertical axis) for the Visscher-Hopper regression are shown. The cumulative distance is given at the bottom, and chromosome numbers are given at the top. The two dotted, horizontal lines represent the empirically derived genome wide significance thresholds (5% and 1%).

Am. J. Hum. Genet., 72:000, 2003

Linkage Analysis of Extremely Discordant and Concordant Sibling Pairs Identifies Quantitative-Trait Loci That Influence Variation in the Human Personality Trait Neuroticism

Jan Fullerton, Matthew Cubin, Hemant Tiwari, Chenxi Wang, Amarjit Bomhra, Stuart Davidson, Sue Miller, Christopher Fairburn, Guy Goodwin, Michael Neale, Simon Fiddy, Richard Mott, David B. Allison, and Jonathan Flint

Information for marker density 0.5, 1, 2, 10cM scan



Limits of fine mapping a quantitative trait
Attwood LD & Heard-Costa NL.
Genetic Epidemiology 24:99-106, 2003

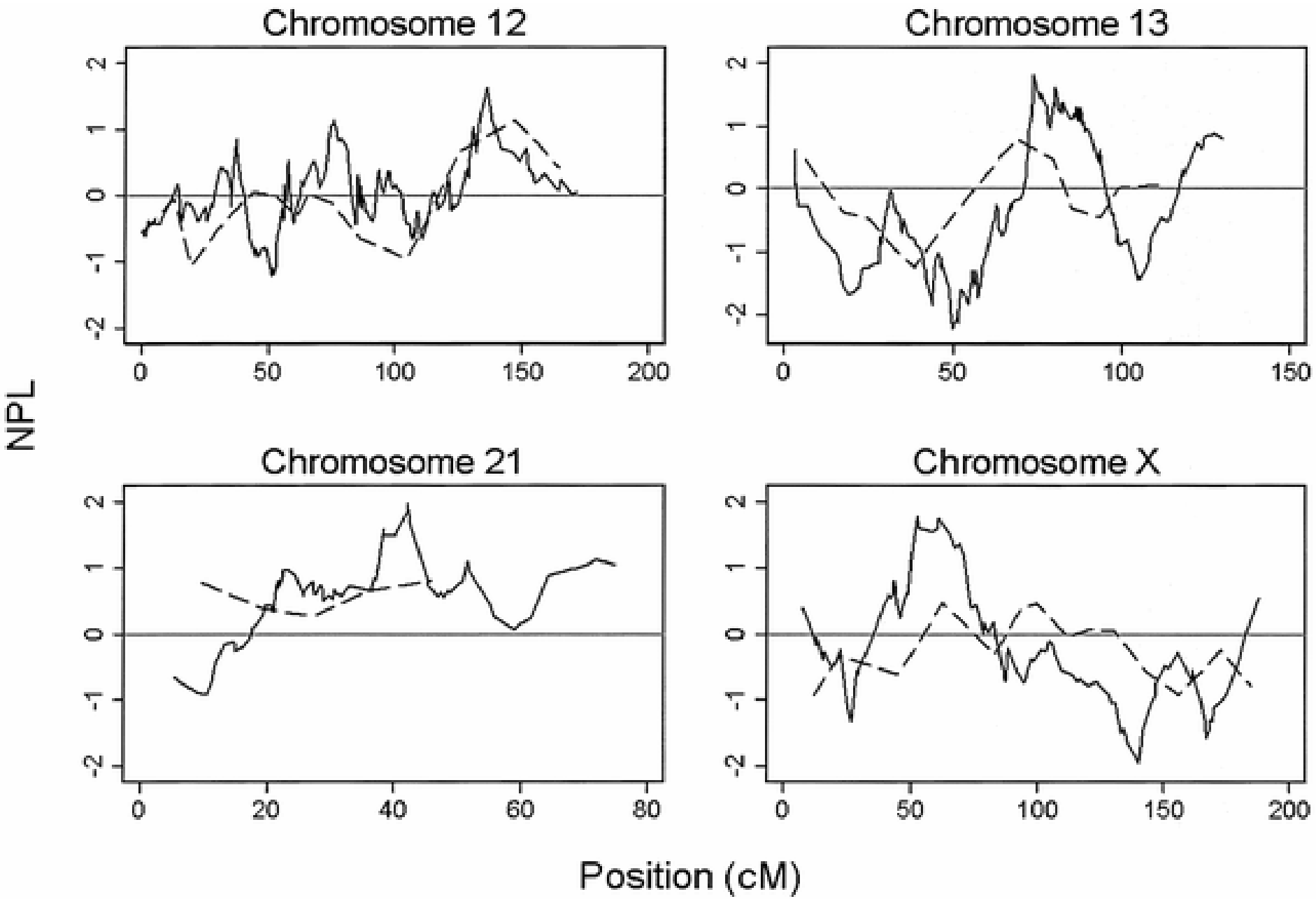
Am. J. Hum. Genet. 75:54–64, 2004

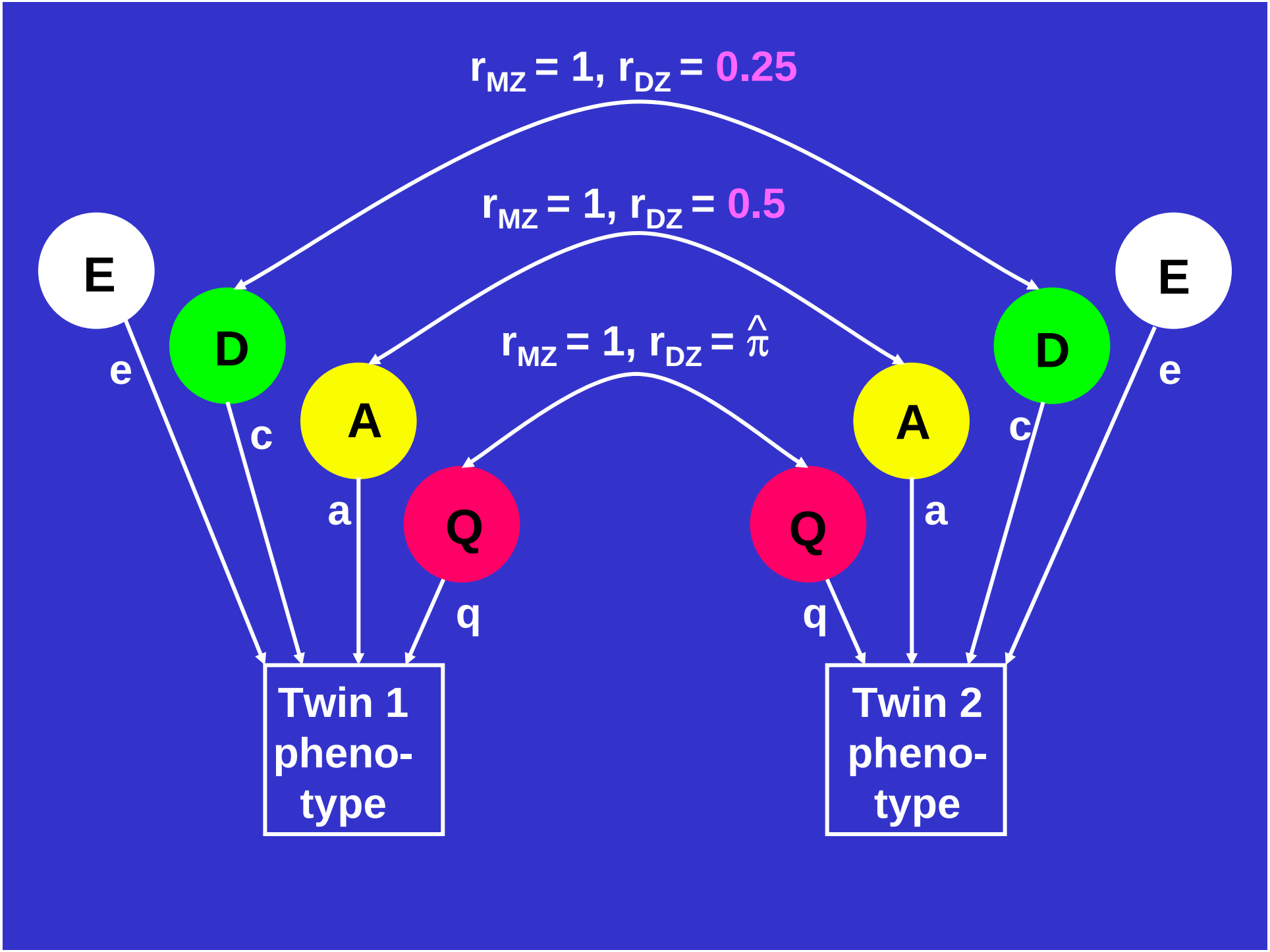
Whole-Genome Scan, in a Complex Disease, Using 11,245 Single-Nucleotide Polymorphisms: Comparison with Microsatellites

Sally John,[1] Neil Shephard,[1] Guoying Liu,[2] Eleftheria Zeggini,[1] Manqiu Cao,[2] Wenwei Chen,[2] Nisha Vasavda,[3] Tracy Mills,[3] Anne Barton,[1] Anne Hinks,[1] Steve Eyre,[1] Keith W. Jones,[2] William Ollier,[1] Alan Silman,[1] Neil Gibson,[3] Jane Worthington,[1] and Giulia C. Kennedy[2]

[1]University of Manchester, Manchester, United Kingdom; [2]Affymetrix, Santa Clara, CA; and [3]AstraZeneca, Macclesfield, United Kingdom

Despite the theoretical evidence of the utility of single-nucleotide polymorphisms (SNPs) for linkage analysis, no whole-genome scans of a complex disease have yet been published to directly compare SNPs with microsatellites. Here, we describe a whole-genome screen of 157 families with multiple cases of rheumatoid arthritis (RA), performed using 11,245 genomewide SNPs. The results were compared with those from a 10-cM microsatellite scan in the same cohort. The SNP analysis detected HLA*DRB1, the major RA susceptibility locus ($P = .00004$), with a linkage interval of 31 cM, compared with a 50-cM linkage interval detected by the microsatellite scan. In addition, four loci were detected at a nominal significance level ($P < .05$) in the SNP linkage analysis; these were not observed in the microsatellite scan. We demonstrate that variation in information content was the main factor contributing to observed differences in the two scans, with the SNPs providing significantly higher information content than the microsatellites. Reducing the number of SNPs in the marker set to 3,300 (1-cM spacing) caused several loci to drop below nominal significance levels, suggesting that decreases in information content can have significant effects on linkage results. In contrast, differences in maps employed in the analysis, the low detectable rate of genotyping error, and the presence of moderate linkage disequilibrium between markers did not significantly affect the results. We have demonstrated the utility of a dense SNP map for performing linkage analysis in a late-age-at-onset disease,


Chromosome 12
Chromosome 13
Chromosome 21
Chromosome X
NPL
Position (cM)


$r_{MZ} = 1, r_{DZ} = 0.25$
$r_{MZ} = 1, r_{DZ} = 0.5$
$r_{MZ} = 1, r_{DZ} = \hat{\pi}$
E
D
A
Q
E
D
A
Q
e
c
a
q
q
a
c
e
Twin 1 pheno-type
Twin 2 pheno-type

But why do we use the average sib values of

$$r_a = 0.5$$
$$r_d = 0.25$$

when we can estimate the (almost) exact values for each sib pair from marker data ?

Are there any advantages in doing so ?

Mean IBD sharing across the genome for the *j*th sib pair was based on IBD estimated from Merlin every centimorgan and averaged at all 3491 points

additive

$$\overline{\hat{\pi}}_{a(j)} = \sum_{i=1}^{3491} \hat{\pi}_{a(ij)} / 3491$$

dominance

$$\overline{\hat{\pi}}_{d(j)} = \sum_{i=1}^{3491} p_{2(ij)} / 3491$$

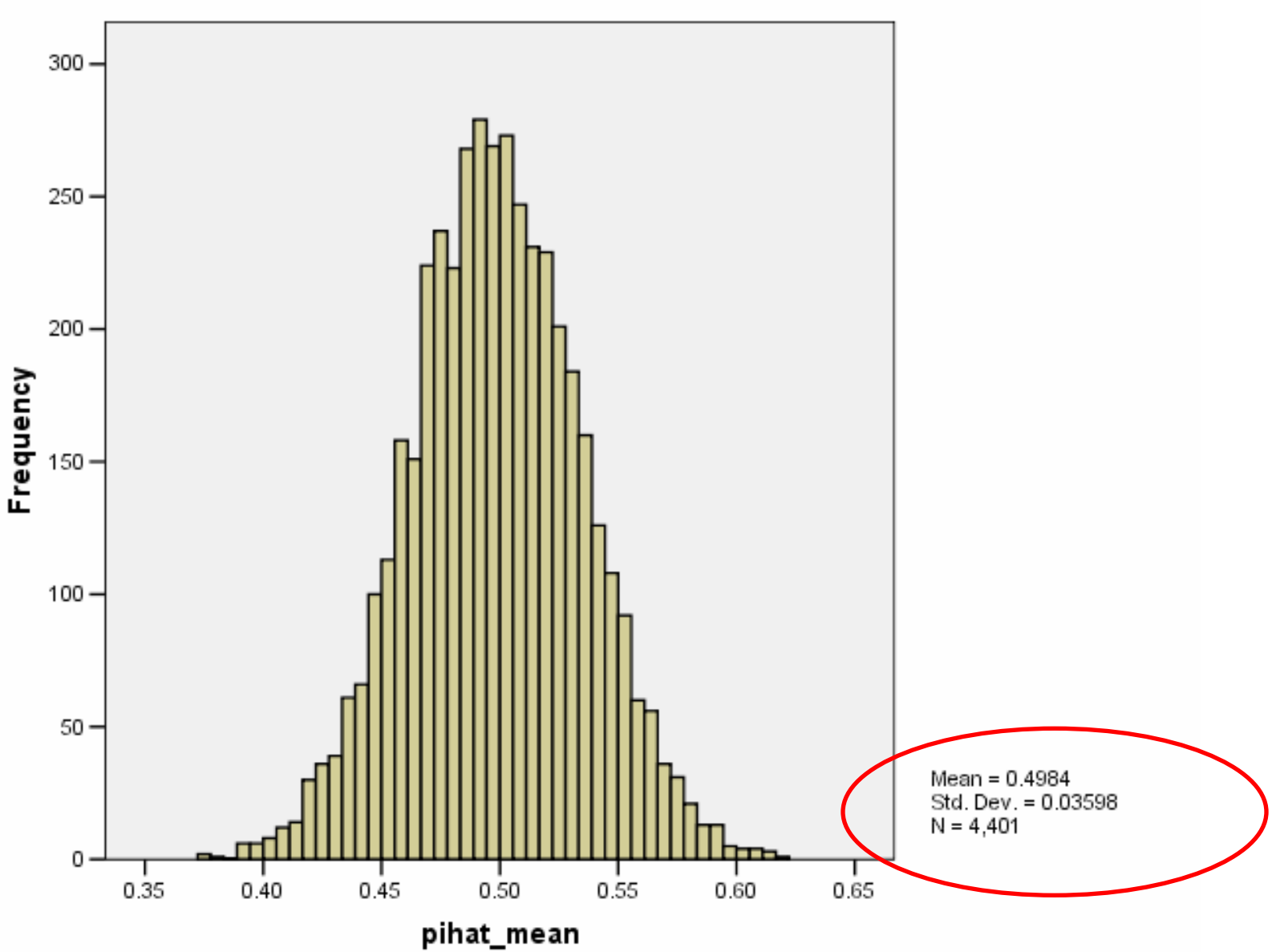
Application

- Phenotype = height

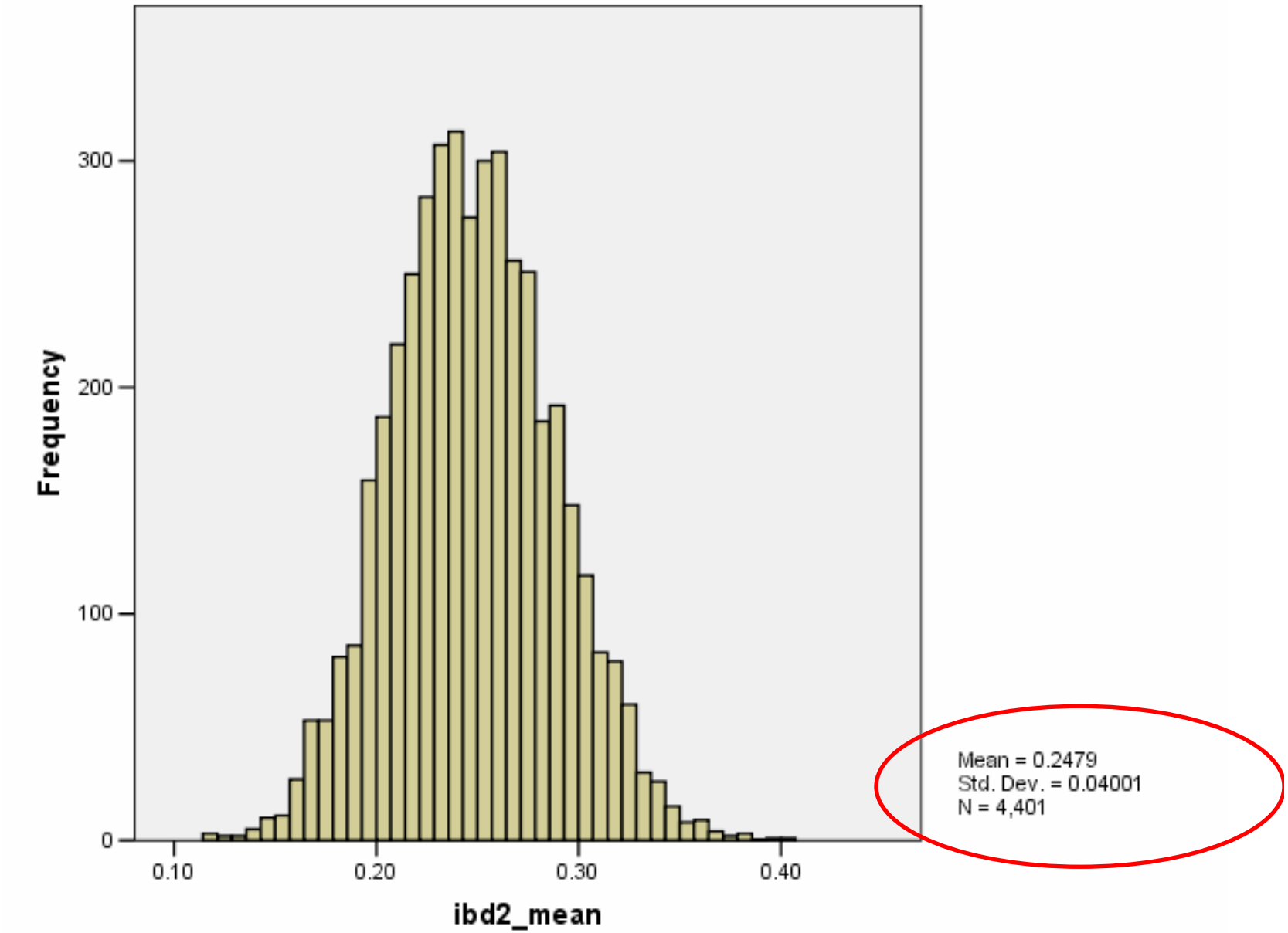
Number of sibpairs with phenotypes and genotypes

Adolescent cohort	931
Adult cohort	2444
Combined	3375

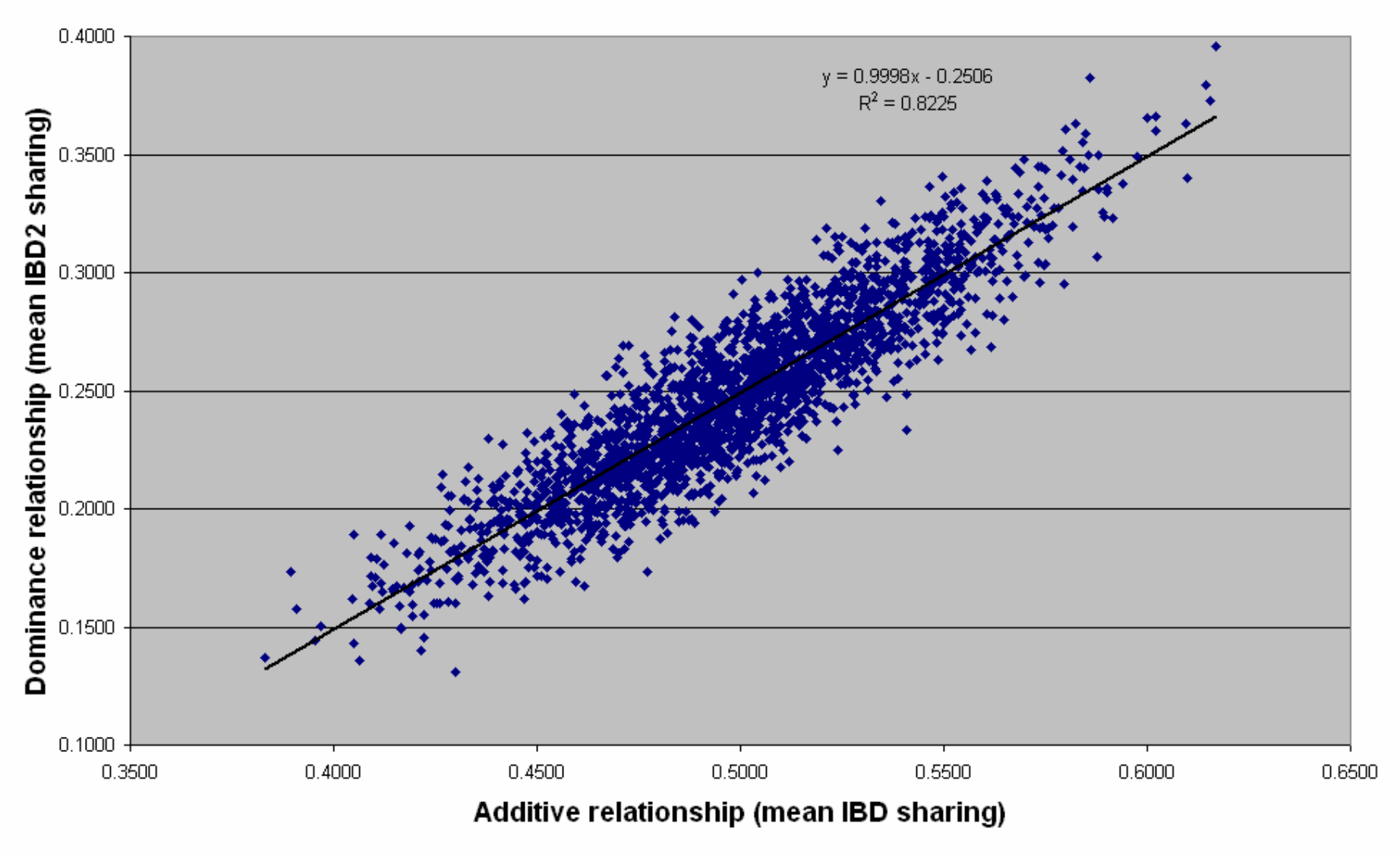
Mean and SD of genome-wide additive relationships



Mean and SD of genome-wide dominance relationships



Additive and dominance relationships correlation = 0.91 (*n*= 4401)



Models

F = Family effect

A = Genome-wide additive genetic

E = Residual

Full model $F + \bar{\hat{\pi}}_{a(j)}A + E$

Reduced model $F + E$

Sampling variances are large

Cohort	F+A (95% CI)
Adolescent	0.80 (0.36 – 0.90)
Adult	0.80 (0.61 – 0.86)
Combined	0.80 (0.62 – 0.85)

►***Estimates of MZ correlation from fullsibs!***

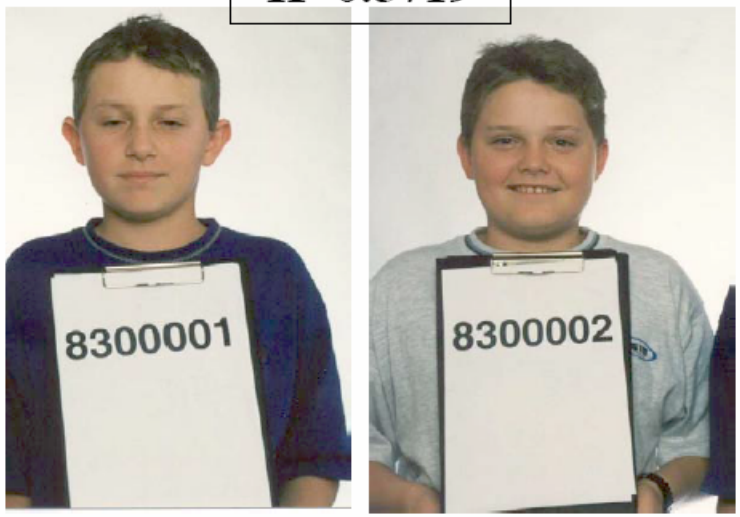
PLOS Genetics, *in press*

And now for IQ! Anyone got sibpairs with IQ + genome scan?

8188001,02
H=0.5677

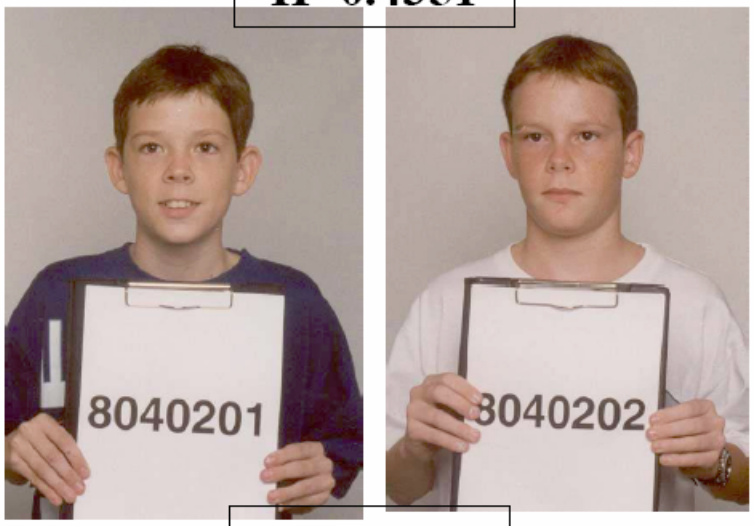
8473001,02
H=0.5577

8300001,02
H=0.5719



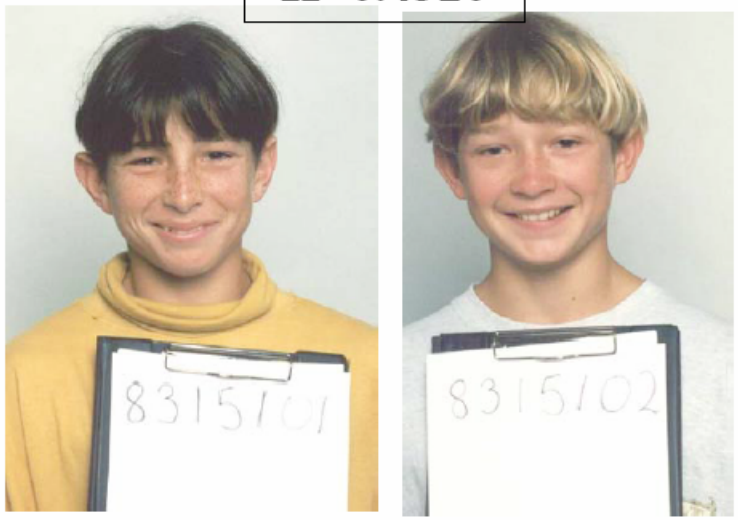
8582601,02
H=0.5640

8040201,02
H=0.4351

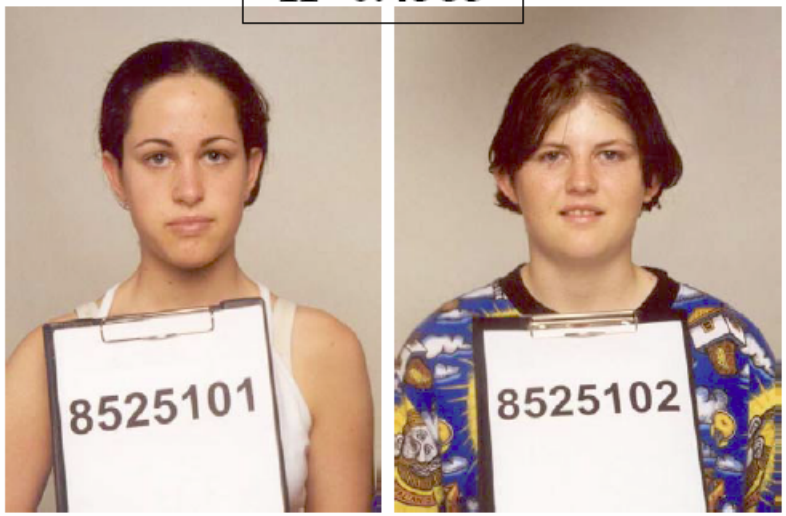


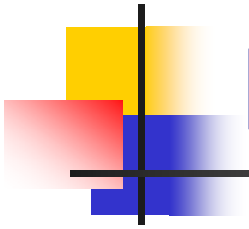
8069101,02
H=0.4291

8315101,02
H=0.4320



8525101,02
H=0.4385





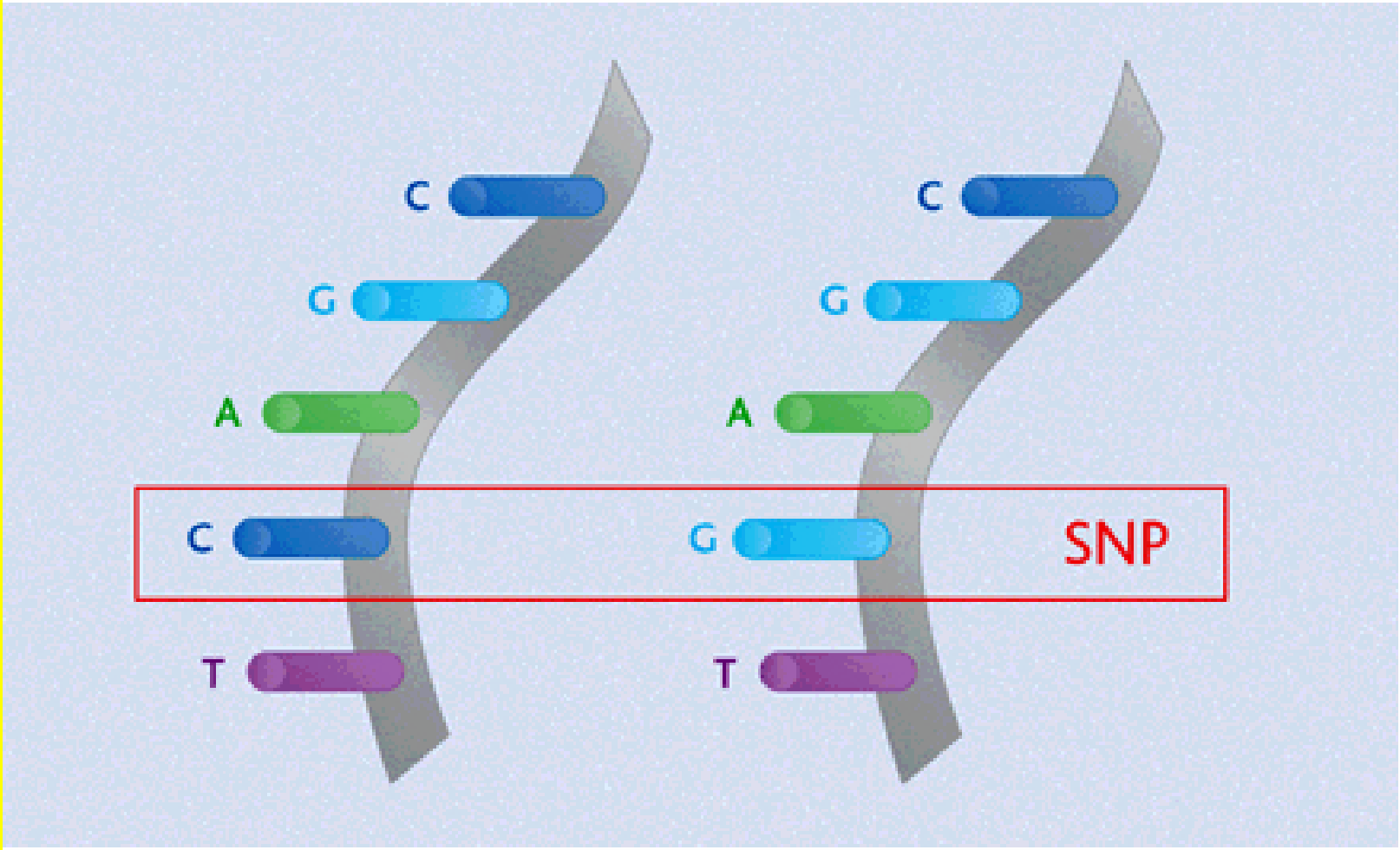
Linkage

- Doesn’t depend on “guessing gene”
- Works over broad regions (good for getting in right ball-park) and whole genome (“genome scan”)
- Only detects large effects (>10%)
- Requires large samples (10,000’s?)
- Can’t guarantee close to gene

Association

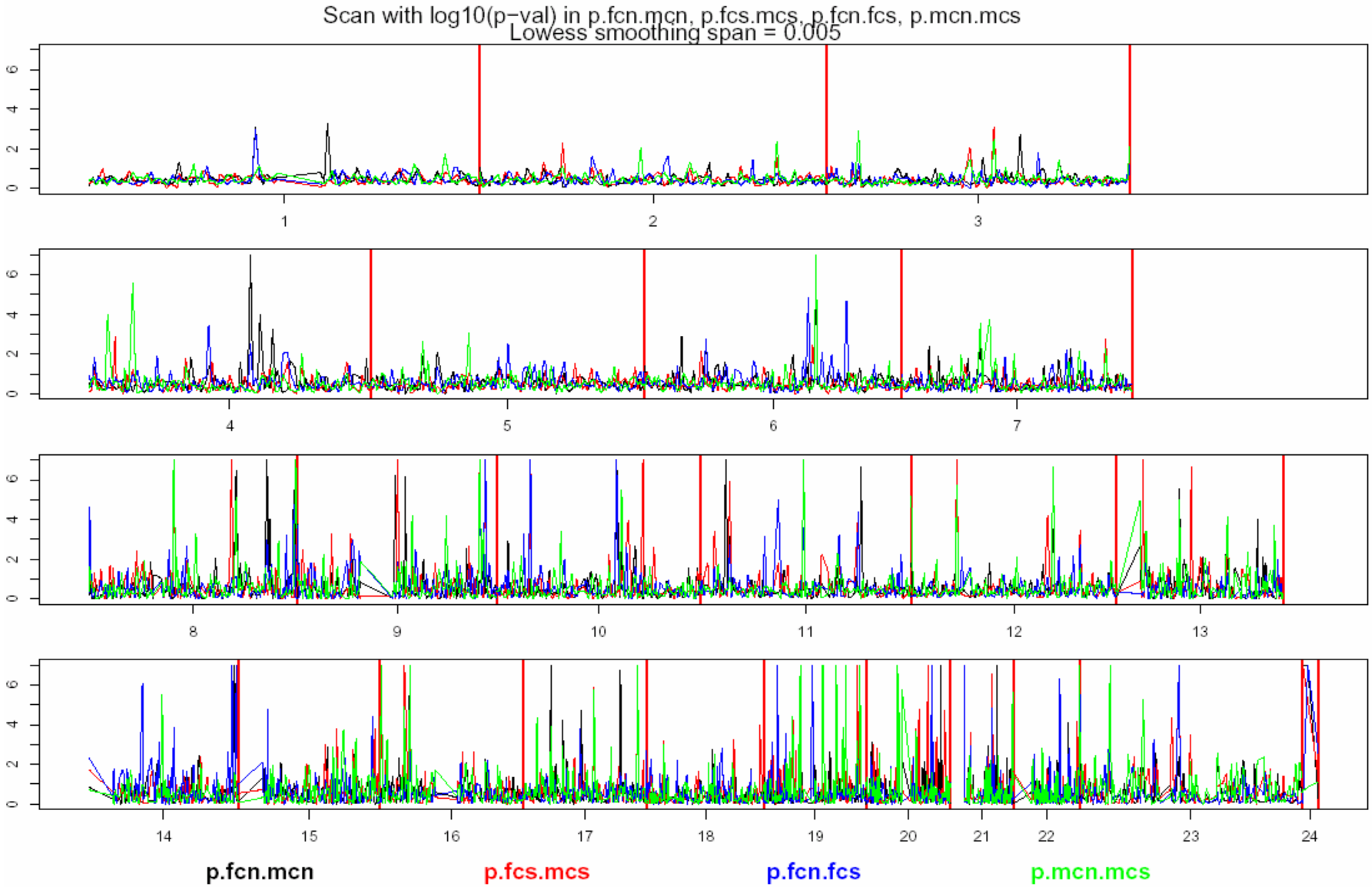
- More sensitive to small effects
- Need to “guess” gene/alleles (“candidate gene”) or be close enough for linkage disequilibrium with nearby loci
- May get spurious association (“stratification”) – need to have genetic controls to be convinced

Variation: Single Nucleotide Polymorphisms

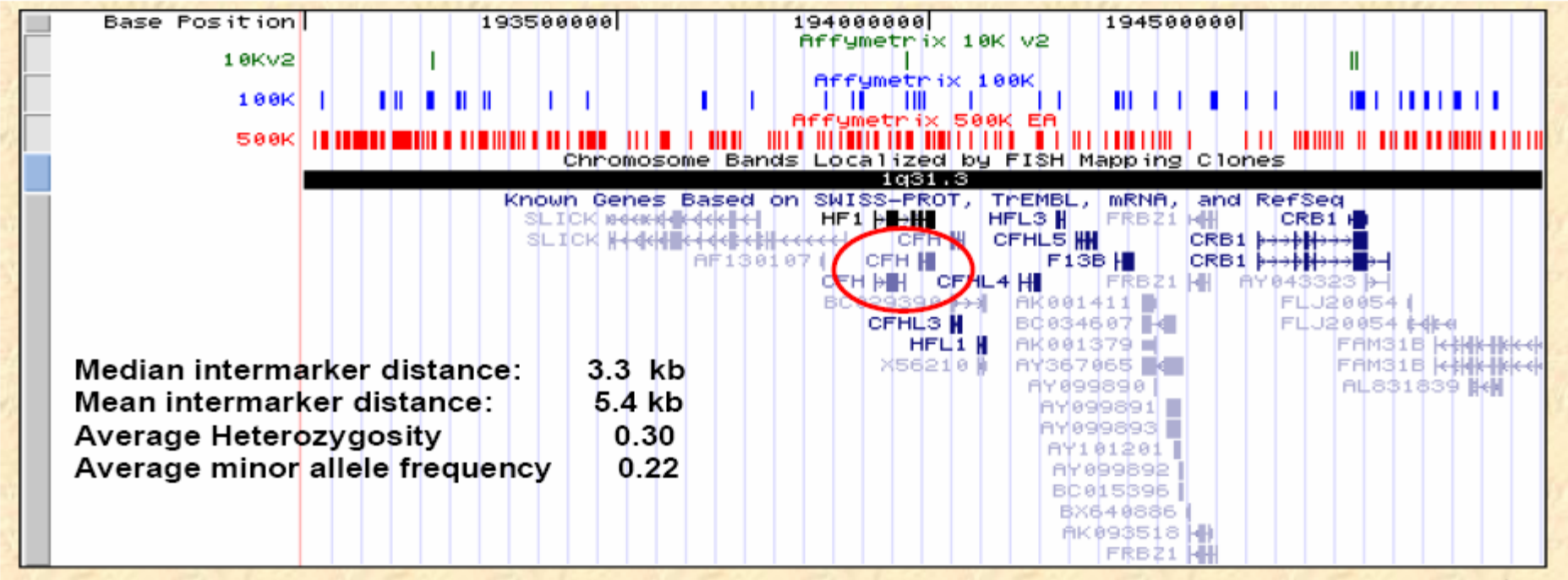


Complex disease marker? SNPs are single-base differences in DNA

Melanoma genome-wide association study



Comparison of Affymetrix 10k, 100k, 500k SNP chips



Genetic analysis of genome-wide variation in human gene expression

Michael Morley[1,3*], Cliona M. Molony[2*], Teresa M. Weber[1,3], James L. Devlin[2], Kathryn G. Ewens[2], Richard S. Spielman[2] & Vivian G. Cheung[1,2,3]

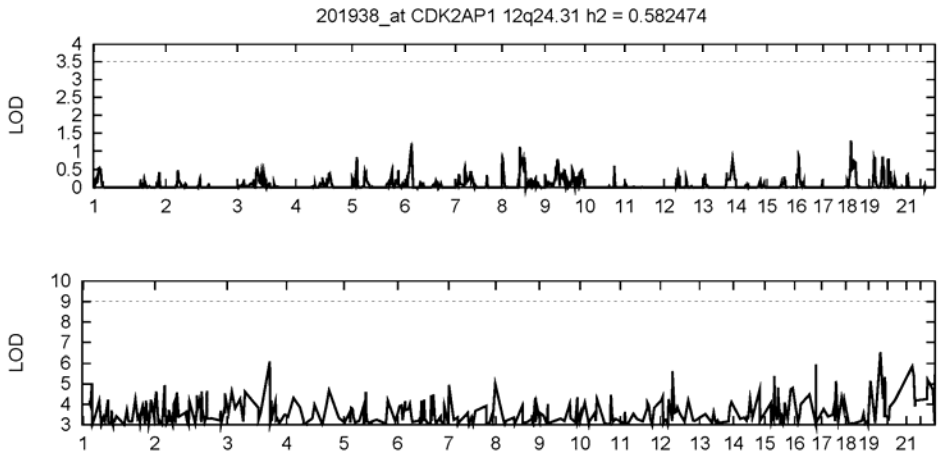
[1]Department of Pediatrics and [2]Department of Genetics, University of Pennsylvania,
[3]The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

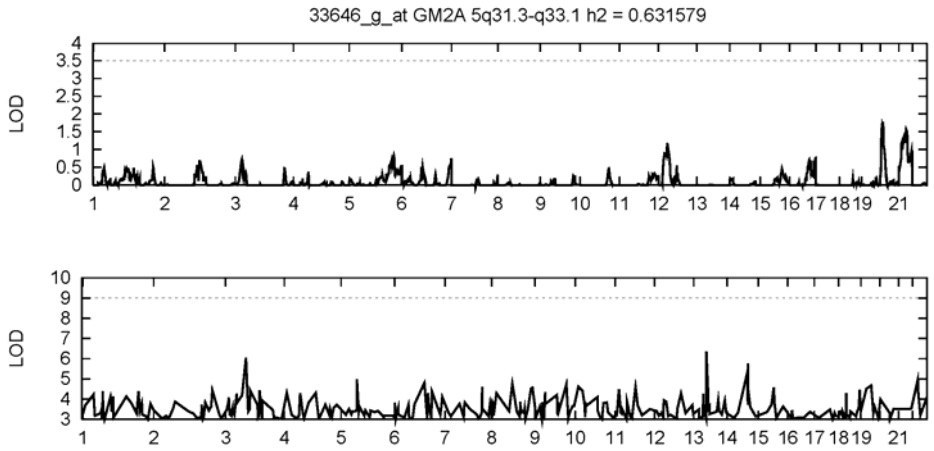
*These authors contributed equally to this work

- Examined expression levels of ~8000 genes on CEPH families
  - Used expression levels as 'phenotypes'
  - Linked expression phenotypes with CEPH microsatellites
- Found evidence for linkage for many phenotypes
- Follow-up SNP genotyping also showed some association

- Found many cis- linkages (linkage region overlaps location of gene whose expression is phenotype), but also many trans

No Linkage
No Association





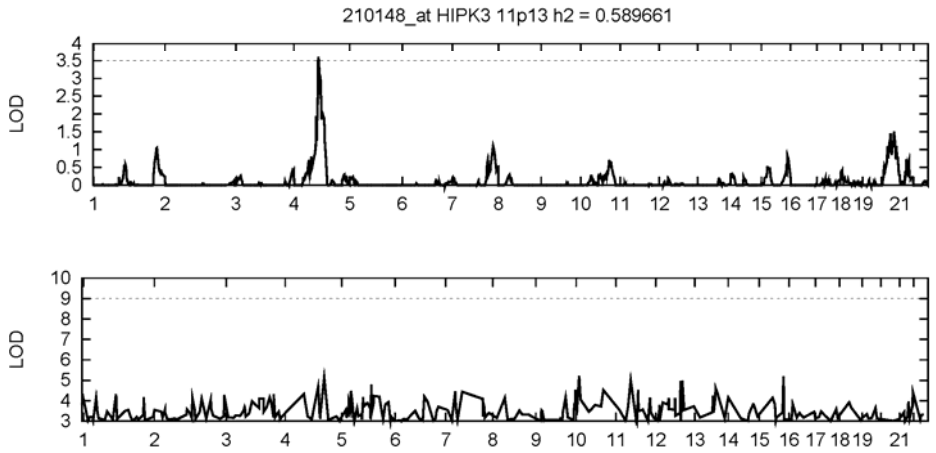
Linkage genome scan

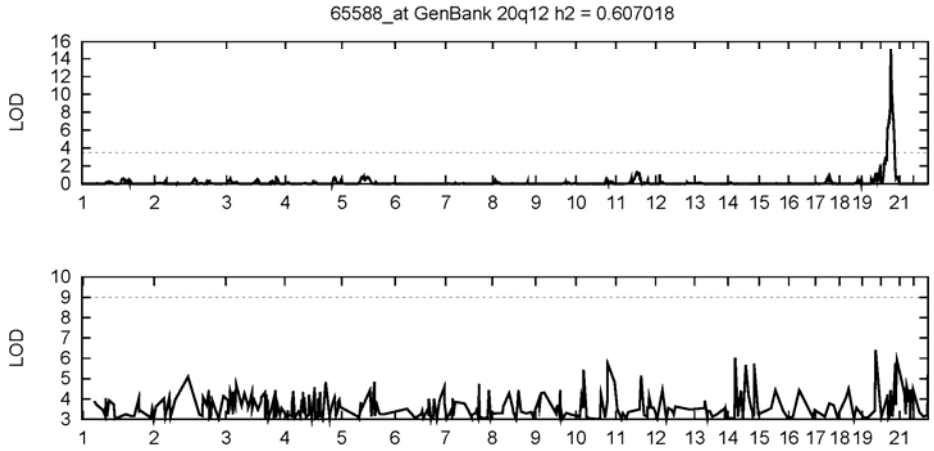
4,000 highly polymorphic markers

Association genome scan

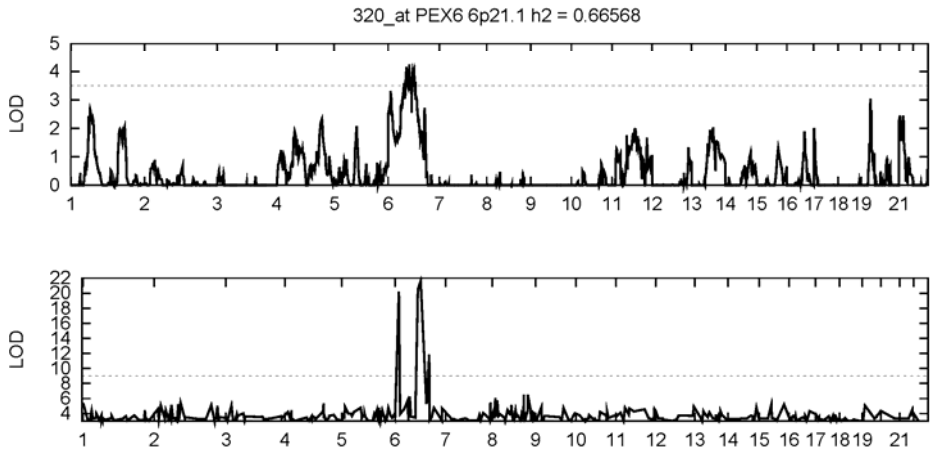
1,000,000 diallelic markers

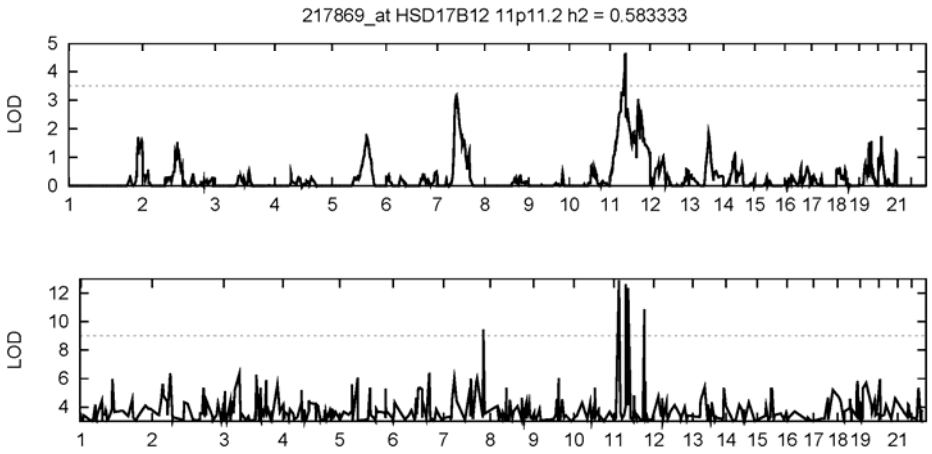
+ Linkage
No Association



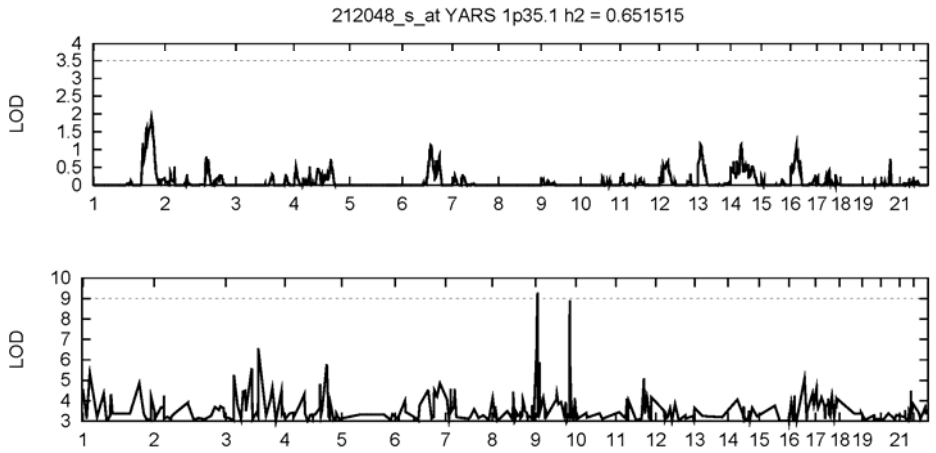


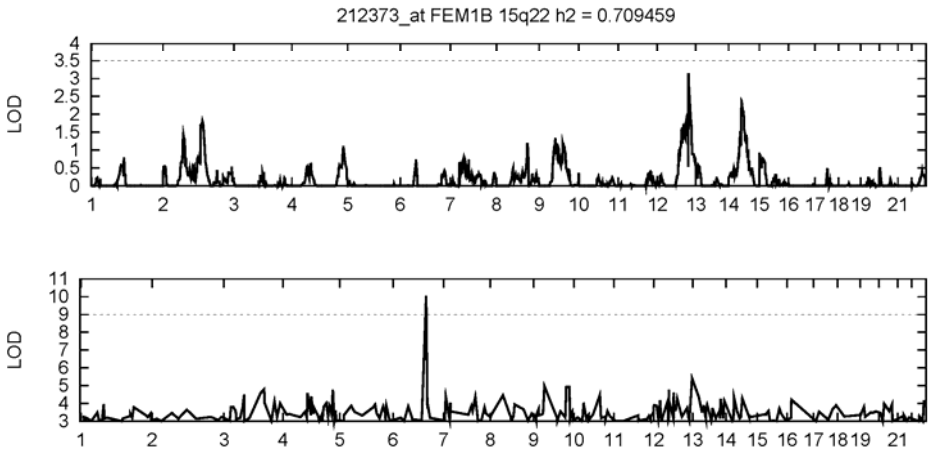
+ Linkage
+ Association





No Linkage
+ Association

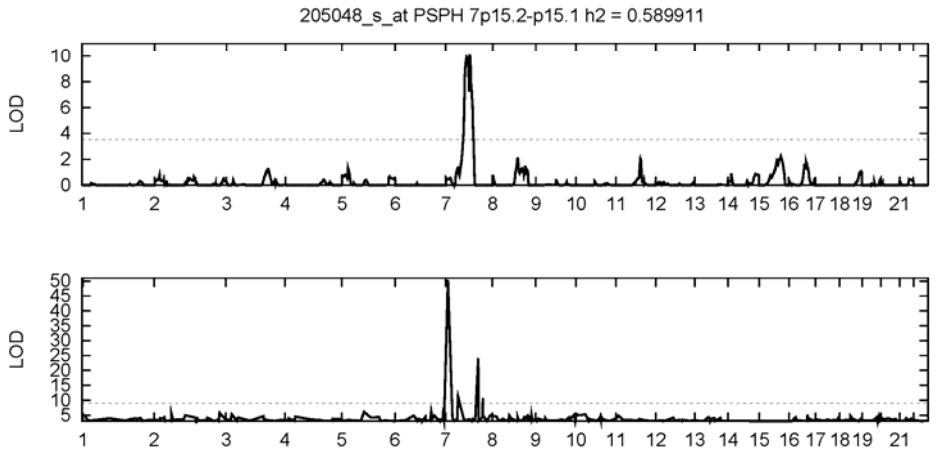


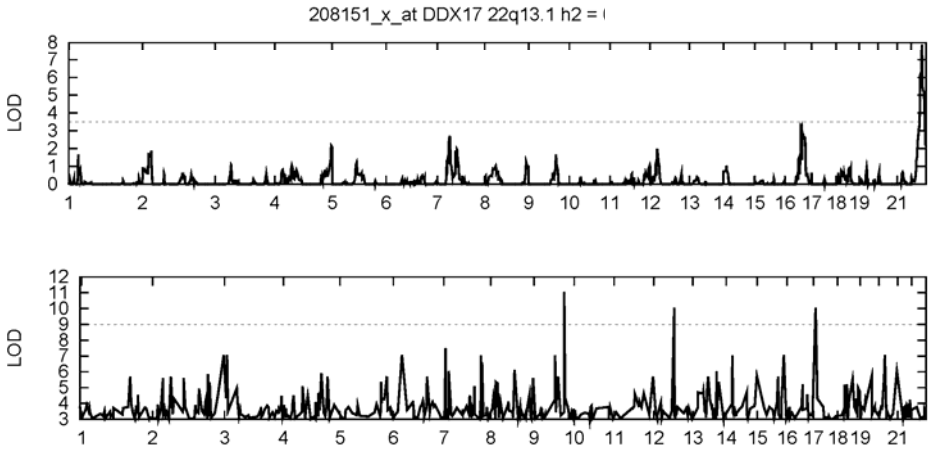


Yes, genome-wide association will work

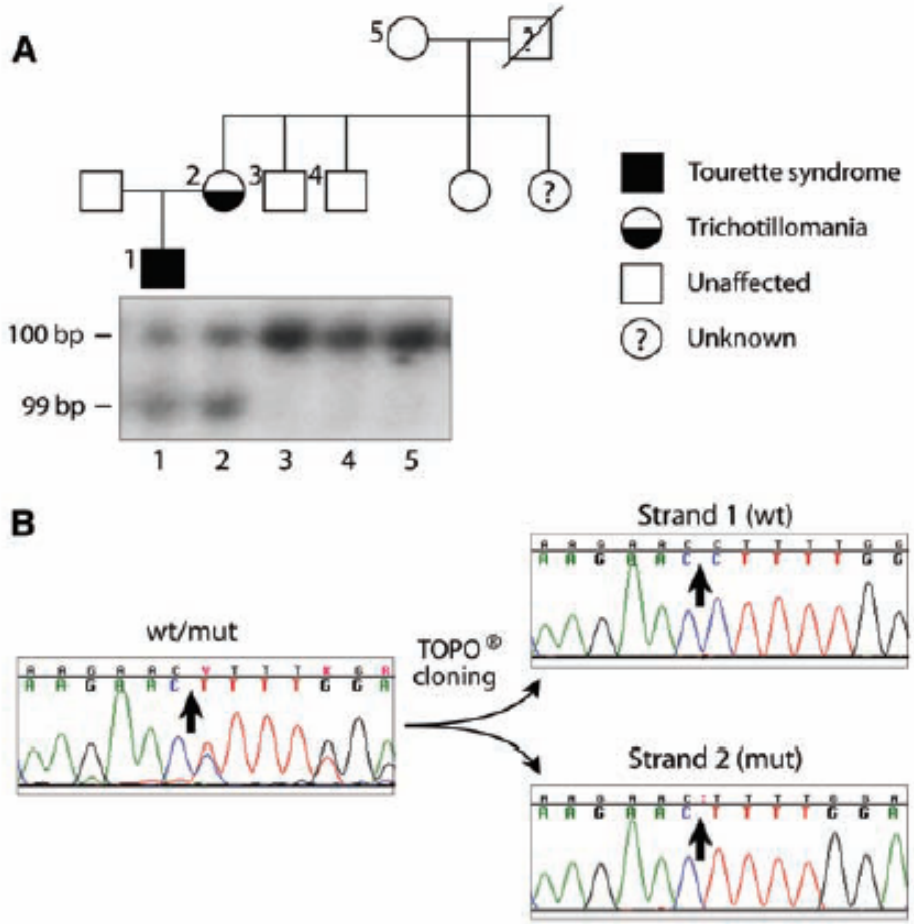
(…sometimes…)

Challenges to come?





Role of miRNA (binding sites) in disease ?



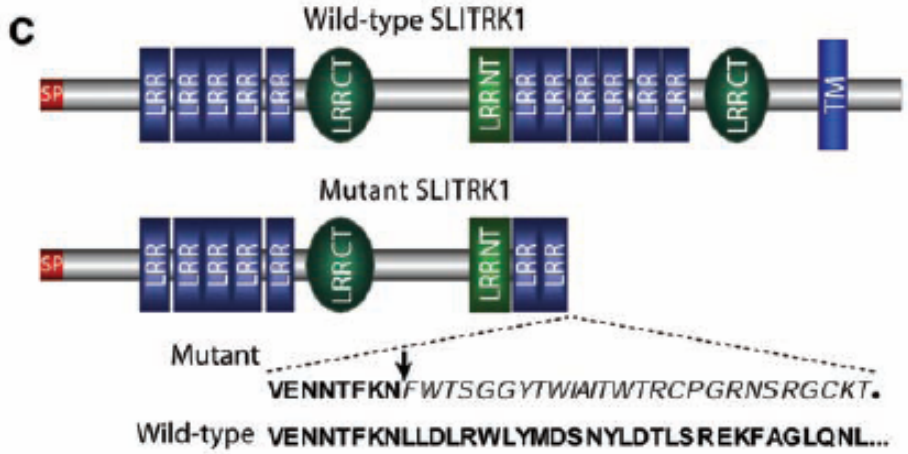


Fig. 2. Identification of a truncating frameshift mutation in *SLITRK1*. (**A**) Pedigree of Family 2 showing the proband (individual 1) diagnosed with TS and ADHD. The patient's mother (individual 2) was retrospectively diagnosed with TTM. Individuals 3 to 5 are unaffected. The affected individuals possess a predicted 100–base pair as well as a mutant 99–base pair fragment amplifying with the same polymerase chain reaction primer pair analyzed by denaturing polyacrylamide gel electrophoresis (*16*). The unaffected individuals in the pedigree carry only the single expected homozygous 100–base pair band. (**B**) A heterozygous sequence trace from the proband shows the overlap of normal and frameshift

Sequence Variants in *SLITRK1* Are Associated with Tourette's Syndrome

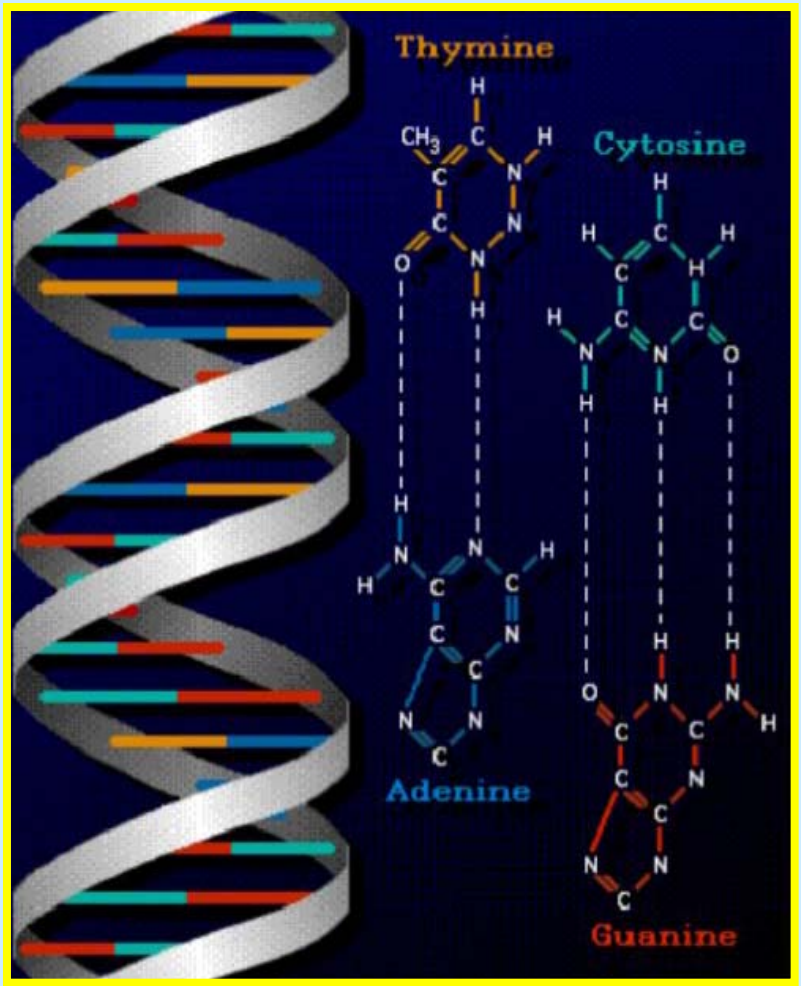
Jesse F. Abelson,[1,2*] Kenneth Y. Kwan,[3,4*] Brian J. O'Roak,[2*]

SCIENCE VOL 310 14 OCTOBER 2005

Comparative Genomics

= differences in DNA sequence

Human-Human 1:1000 = 0.1%



Human-Chimp 1:100 = 1%

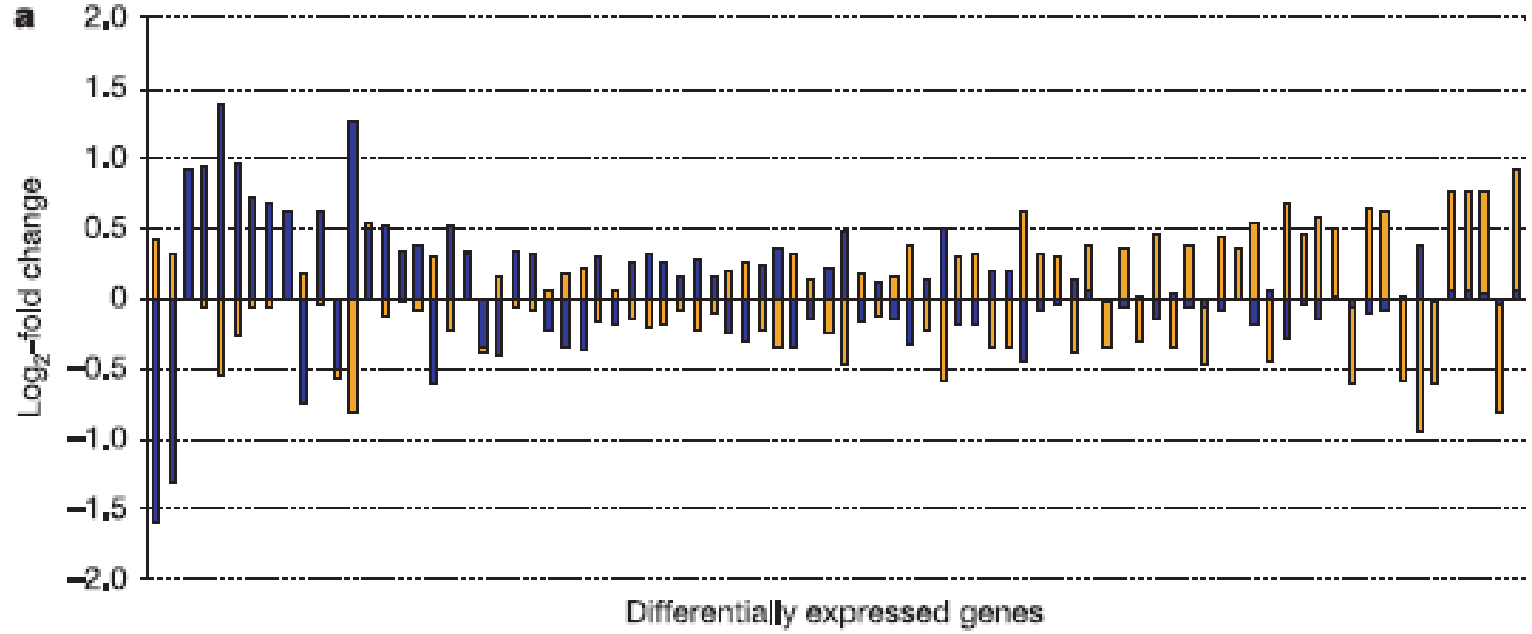
Human-Mouse 1:8 = 15%

nature Vol 440|9 March 2006|doi:10.1038/nature04559

LETTERS

Expression profiling in primates reveals a rapid evolution of human transcription factors

Yoav Gilad[1]†, Alicia Oshlack[2], Gordon K. Smyth[2], Terence P. Speed[2,3] & Kevin P. White[1]



features that point to the action of directional selection. Among the gene set with a human-specific increase in expression, there is an excess of transcription factors; the same is not true for genes with increased expression in chimpanzee.

Which genes have evolved fastest?

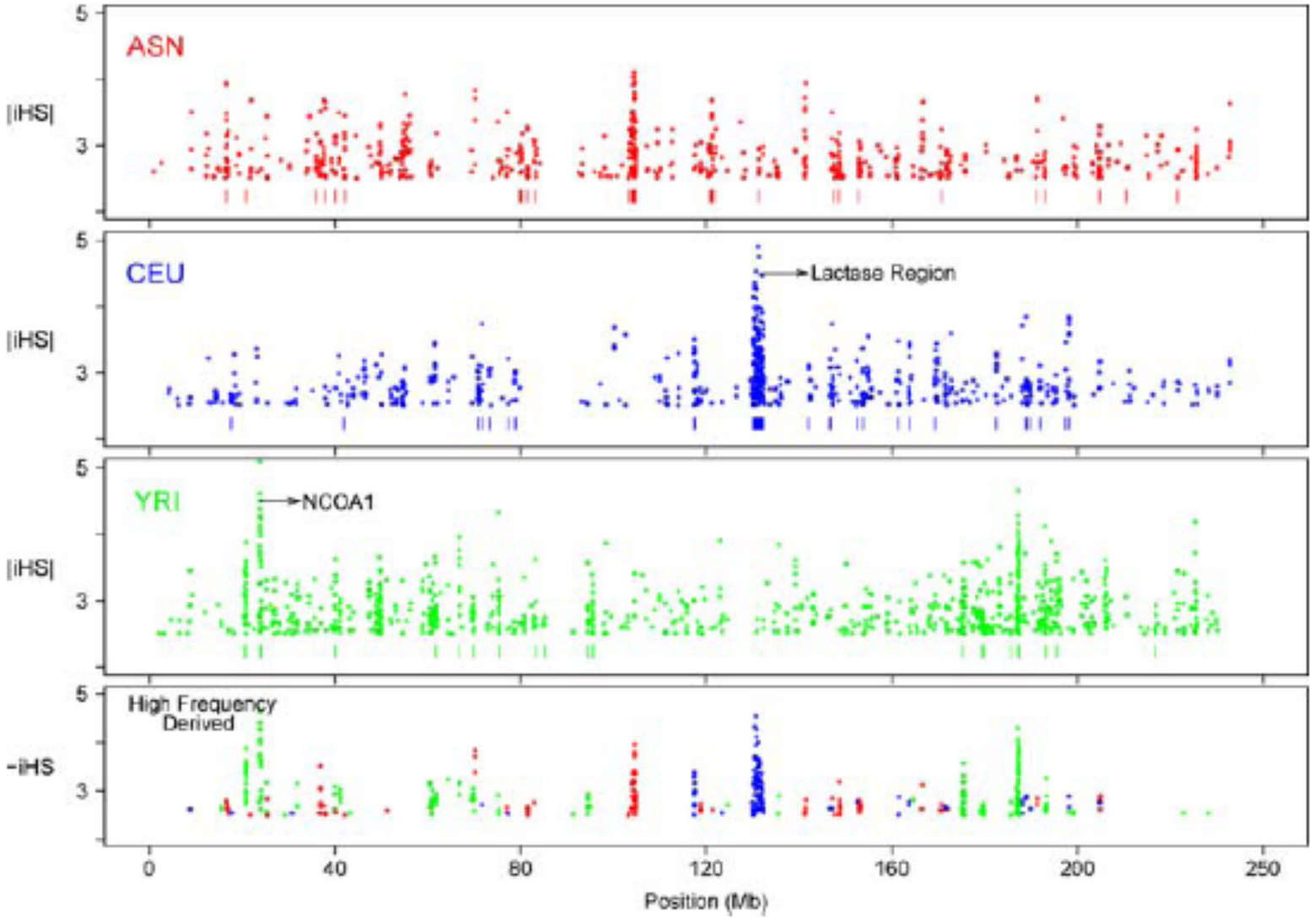


Figure 3. Plots of Chromosome 2 SNPs with Extreme iHS Values Indicate Discrete Clusters of Signals

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]

The $1,000 Human Genome - Implications for Life Science, Healthcare, and IT

Published by:
IDC

EPIGENETIC DISCORDANCE IN IDENTICAL TWINS

The missing “environment” ?

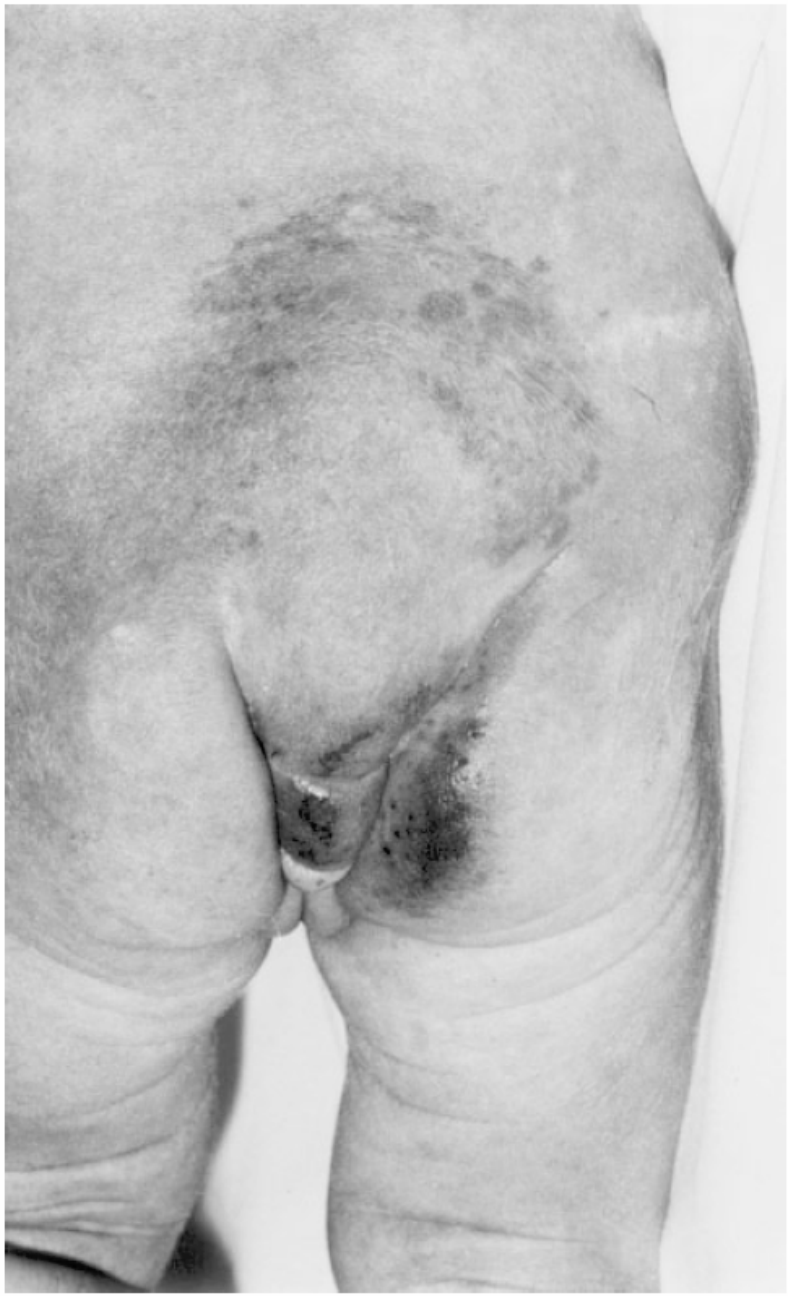


Fig. 1. Patient 1. Soft tumor and abnormal aspect in the lumbosacral area.

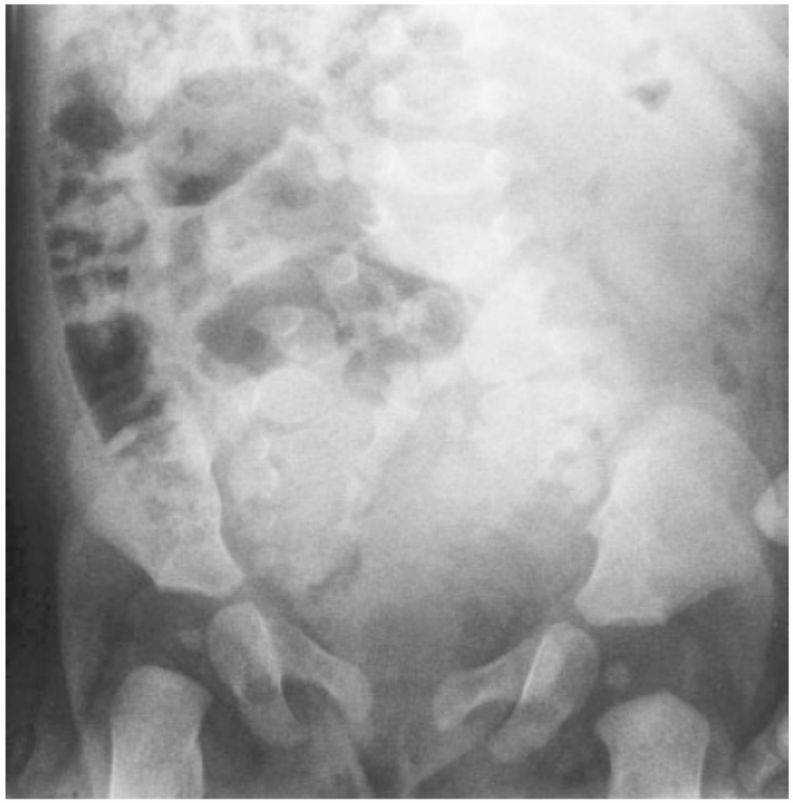
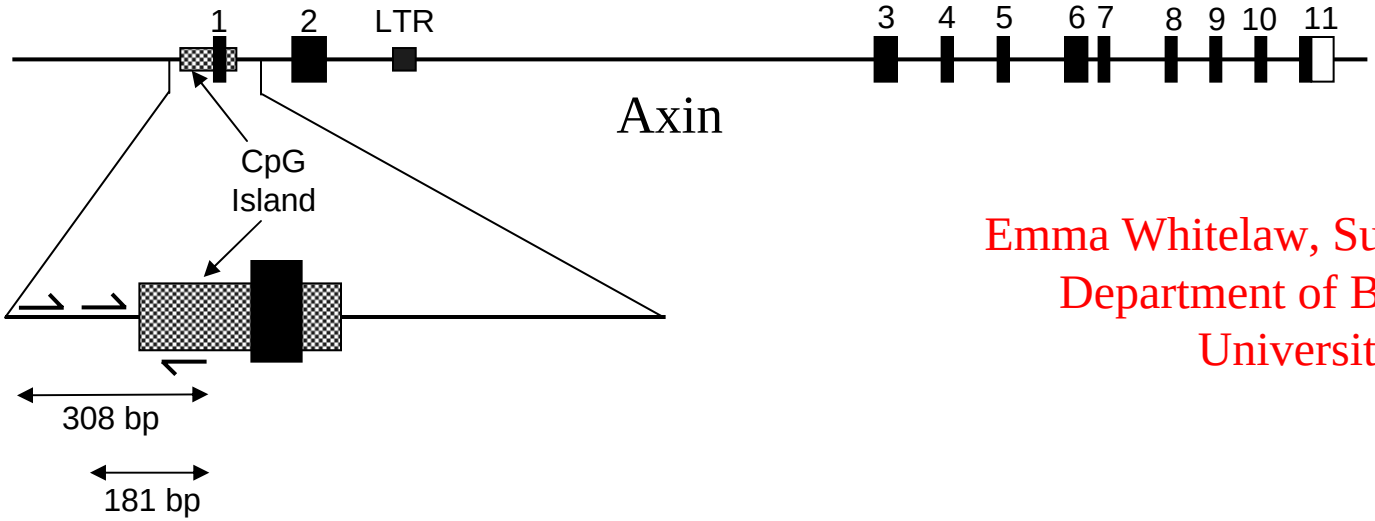


Fig. 2. Patient 1. Radiograph of the vertebral column shows complete duplication of the spine from L4 down.

urethra, a dilated pelvis of the right kidney, bilateral uterus unicornis with normal ovaries, hemivertebrae of thoracic vertebrae 6 and 10, and abnormal curvature of the sacrum. A persistent ductus arteriosus and secundum atrial septum defect was suspected, but results of cardiac investigations at 10 months were normal.

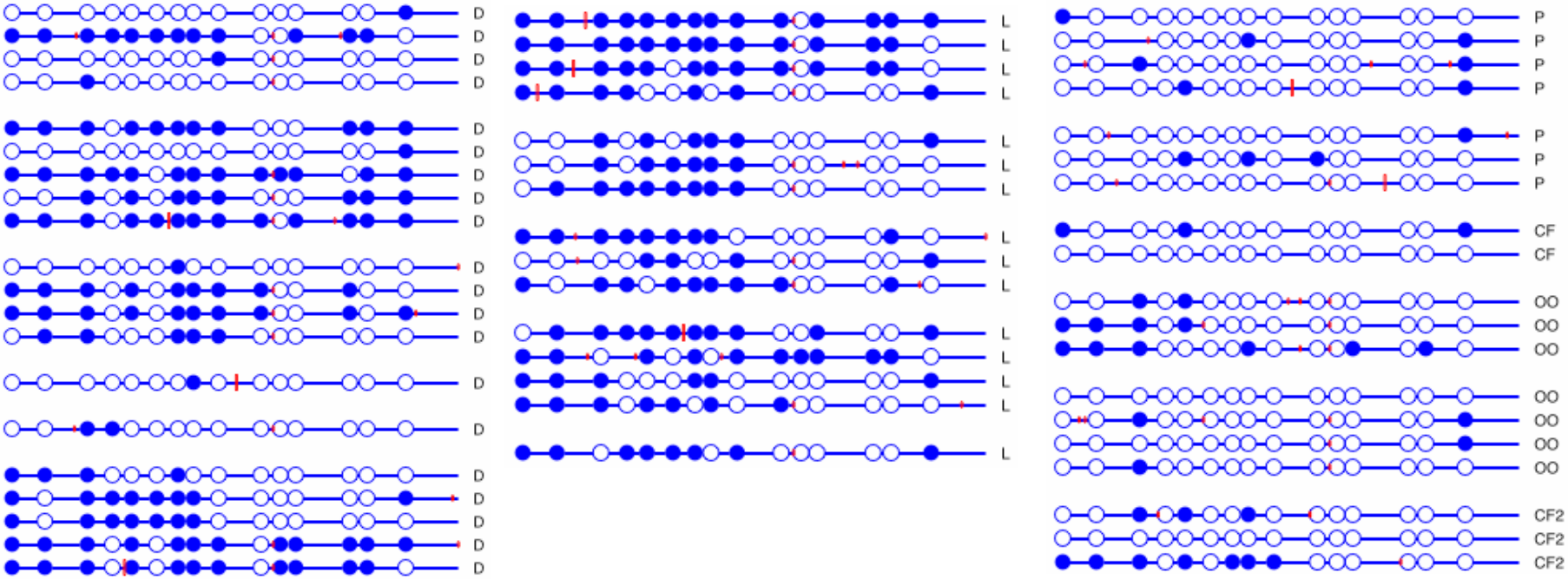
At physical examination for genetic evaluation at 4 months we saw a baby girl with epicanthal folds, but no other minor anomalies. She had a capillary nevus on her left buttock. In the anal region only a dimple was seen. The patient was operated on one day after birth, when a colostomy was made and a fistula connected to the colon

Discordant caudal duplication in MZ twins



Emma Whitelaw, Suyinn Chong
Department of Biochemistry
University of Sydney

Twin 1- unaffected < Twin 2 - affected > Controls [e.g.]



Other studies on MZ discordance

Epilepsy (with S. Berkovic, L. Vadlamudi)

Schizophrenia (with B.Mowry, N.Hayward)

Depression (with A. Petronis, D. Boomsma, P. McGuffin)

Asthma (with M.Ferreira, E.Whitelaw)

Per ardua ad astra

*(Through hard work to the *genes*)*