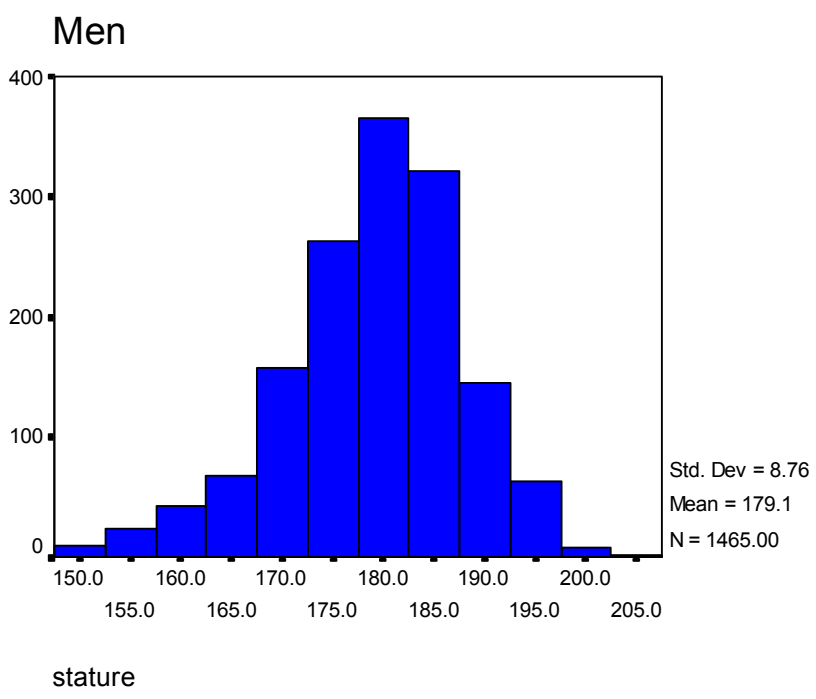
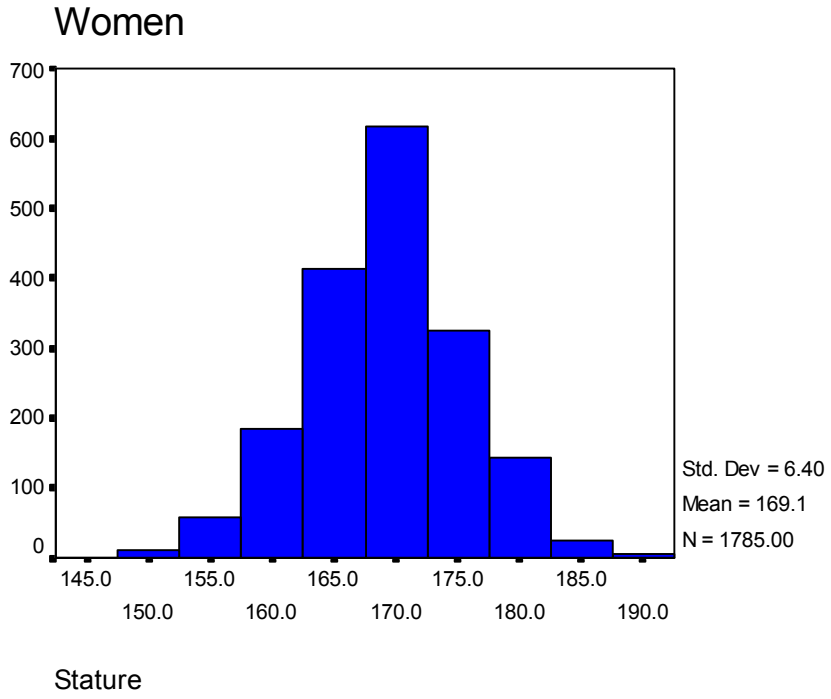
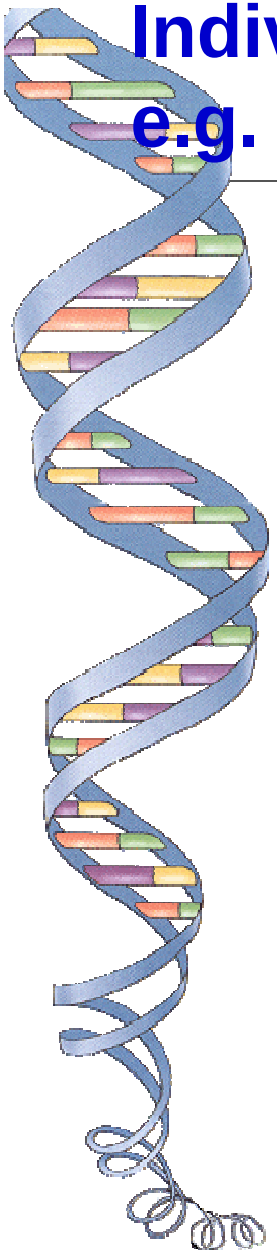
Variation (individual differences): Stature (in cm) in Dutch adolescent twins





Individual differences in human characteristics, e.g. normal and abnormal behavior

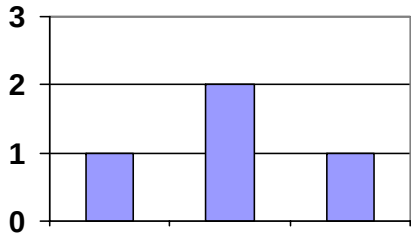


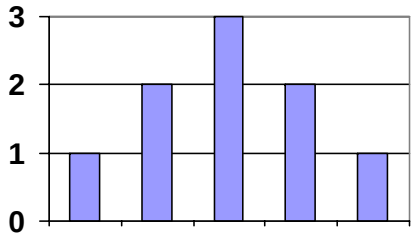
Caused by:

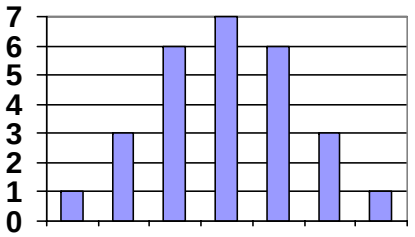
- differences in genotype (G)?
- differences in environment (E)?
- interaction between G and E?

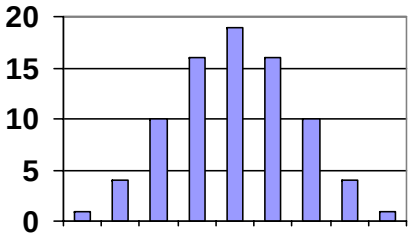
Complex: Polygenic Traits

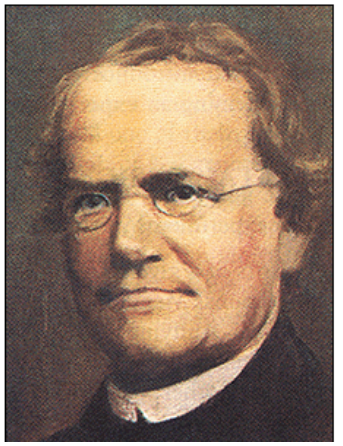
1 Gene	2 Genes	3 Genes	4 Genes
→ 3 Genotypes	→ 9 Genotypes	→ 27 Genotypes	→ 81 Genotypes
→ 3 Phenotypes	→ 5 Phenotypes	→ 7 Phenotypes	→ 9 Phenotypes





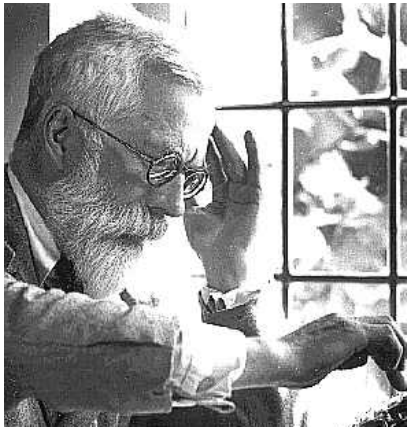






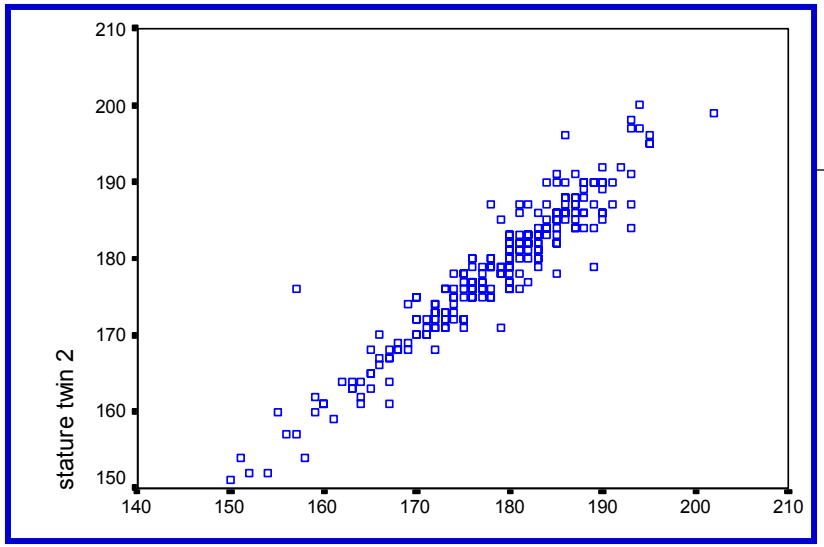
Mendel: Laws of inheritance for monogenic traits:
1 Segregation
2 Independent Assortment

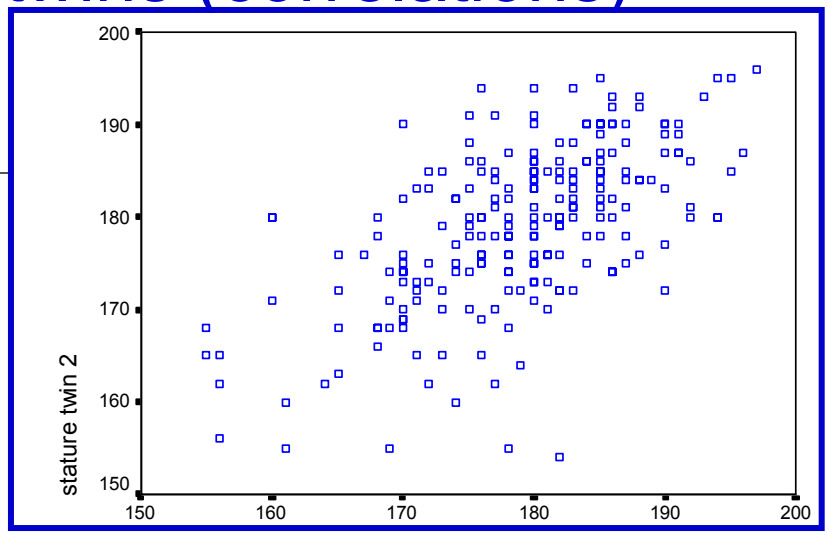
Galton: correlations between family members for continuous traits: Family & Twin Resemblance.



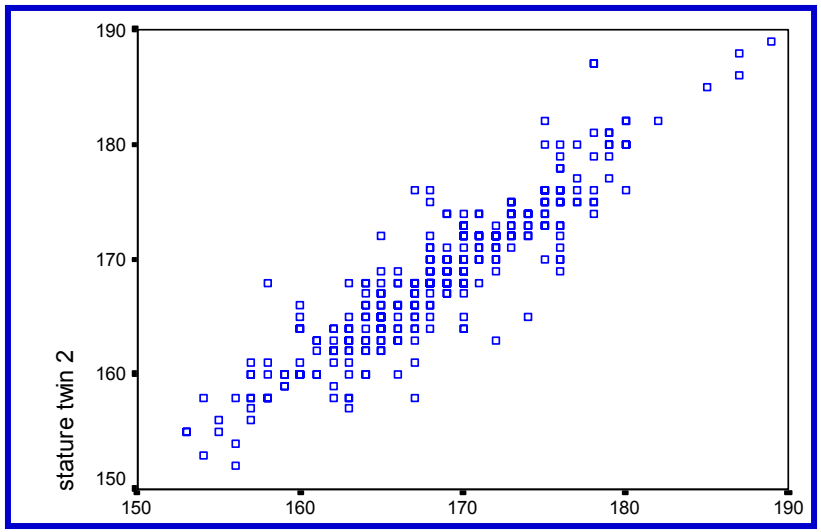
Fisher: traits can be influenced by more than one gene (which each can have small effects). Effects of genes add up and lead to a normal distribution in the population.

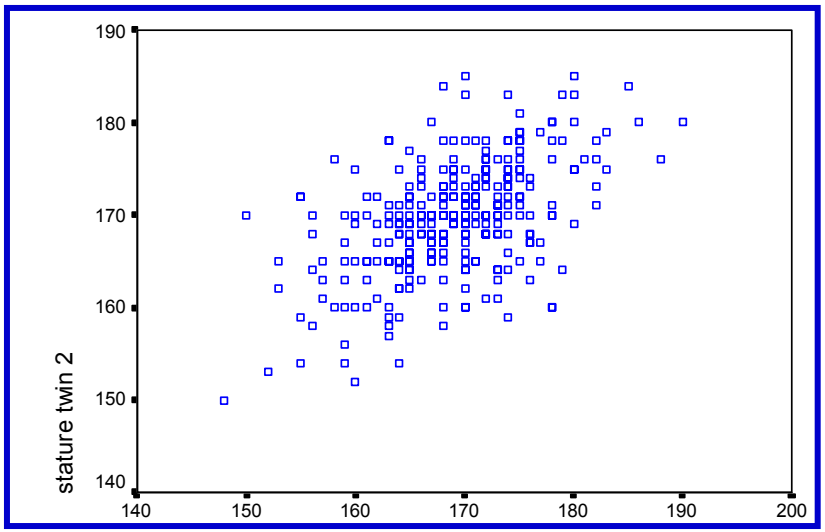
Stature in male and female twins (correlations)



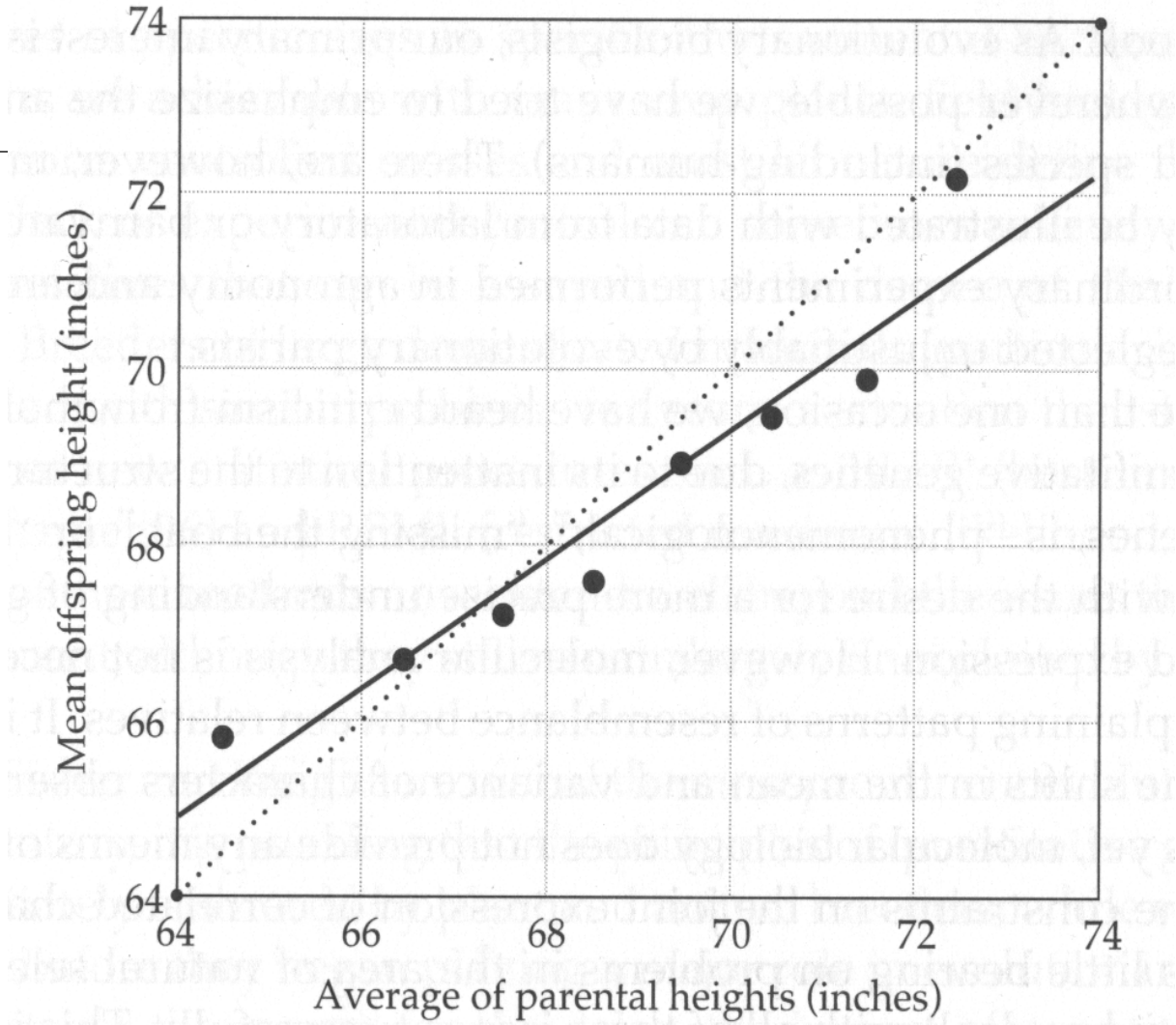


Twin Correlations MZM: 0.95, MZF: 0.92, DZM: 0.60, DZF: 0.52



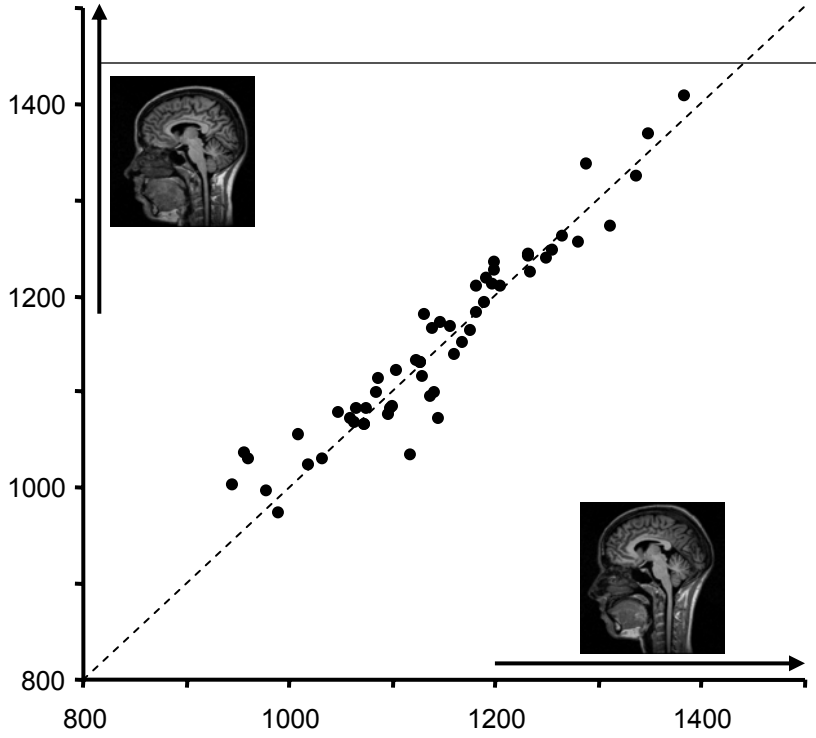


[Galton, 1889]

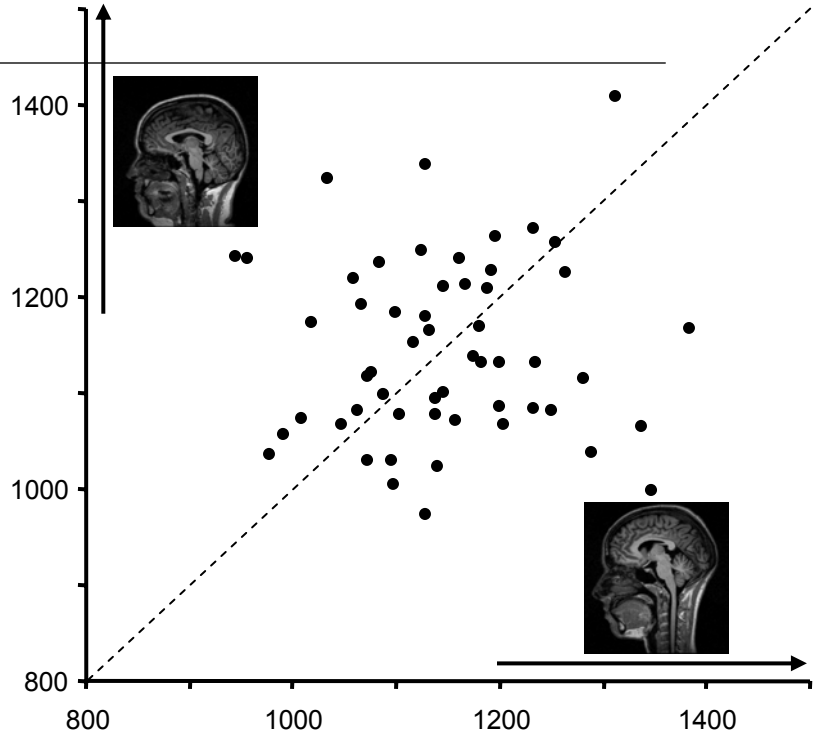


Traits influenced by genes will be correlated among biological relatives

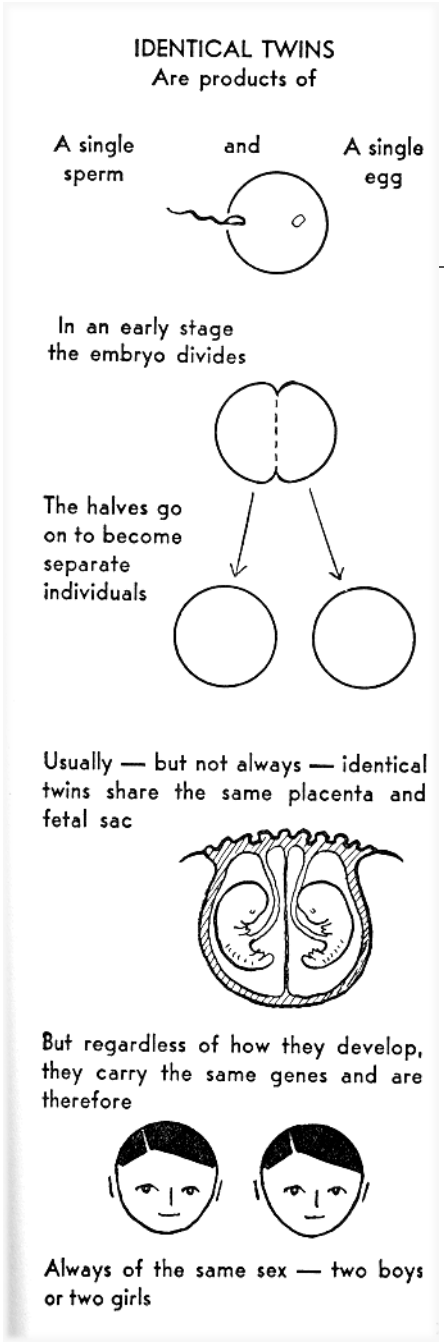
Brain volumes: resemblance of MZ and DZ twins



Brain volume MZ twin pairs (milliliter) in twin and co-twin

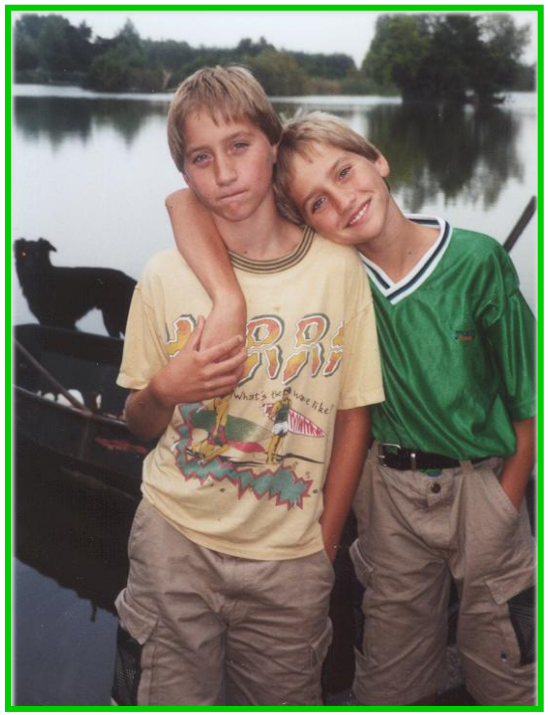


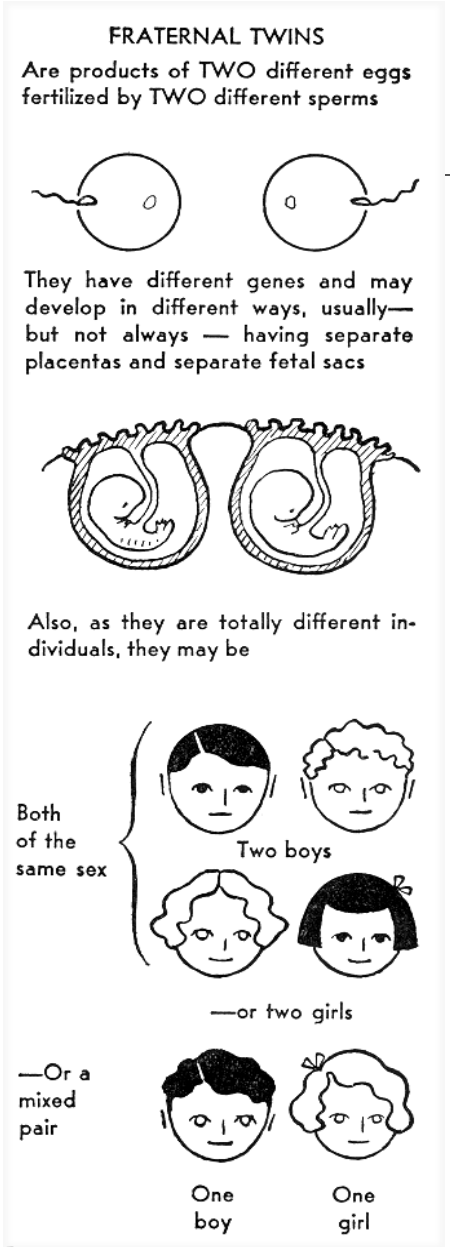
Brain volume DZ twin pairs (milliliter) in twin and co-twin



‘Identical’ twins

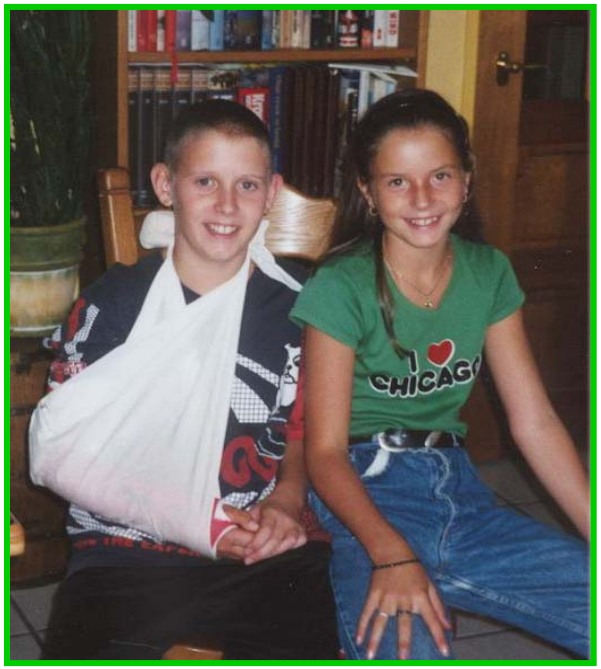
Monozygotic (MZ) twins:
~100% genetically identical





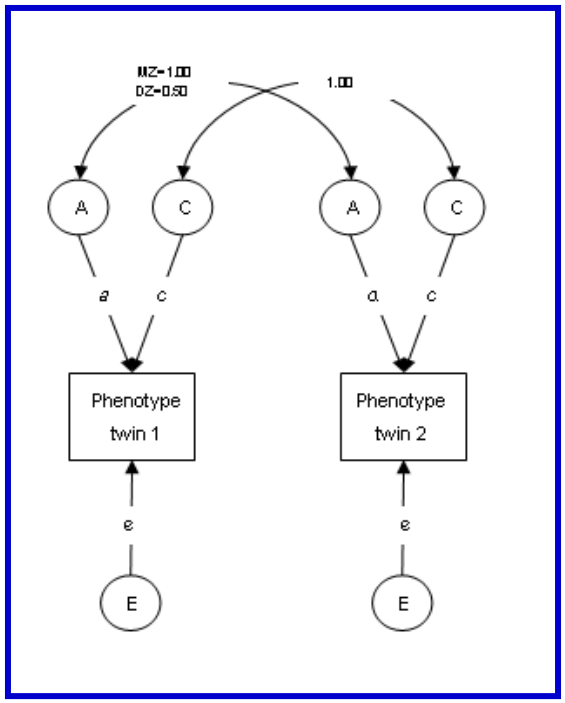
Fraternal twins

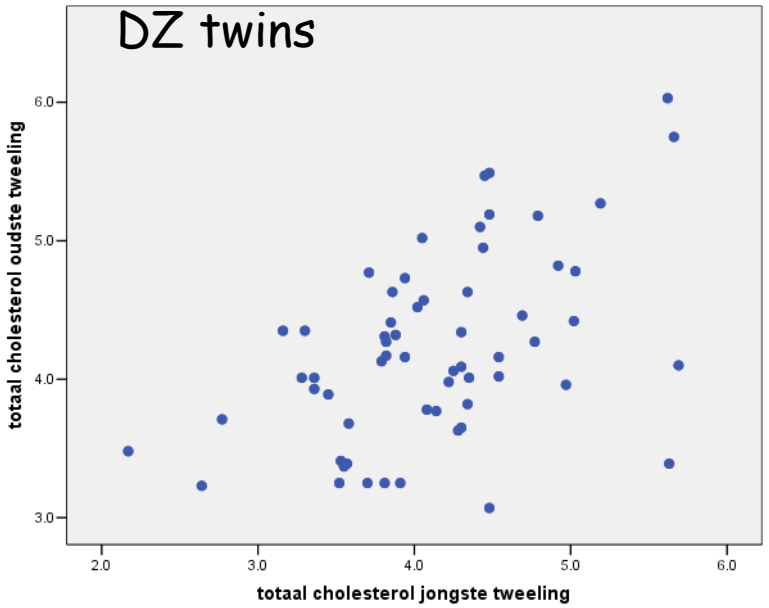
Dizygotic (DZ) twins share ~50% of their segregating genes



Twin Model

- Twin correlations for cholesterol levels (17-yr old twins)
- rMZ = 0.86 & rDZ = 0.46
- Heritability = 80% (=2(.86-.46))





Designs to disentangle G + E

- Family studies – G + C confounded
- MZ twins alone – G + C confounded
- MZ twins reared apart – rare, atypical, selective placement ?
- Adoptions – increasingly rare, atypical, selective placement ?
- MZ and DZ twins reared together
- Extended twin design

Bouchard & McGue: Genetic and environmental influences on human psychological differences (2003)

	Intraclass correlations	
	MZT (626 pairs)	MZA (74 pairs)
Positive emotionality	.55	.43
Negative emotionality	.44	.47
Constraint	.56	.58

Classical twin design: Assumptions

* known zygosity
* EEA: equal environment (including prenatal)
* representative

Zygosity

DZ = DOS

DZ = very unlike in appearance

DZ = different at marker loci

(except for measurement error)

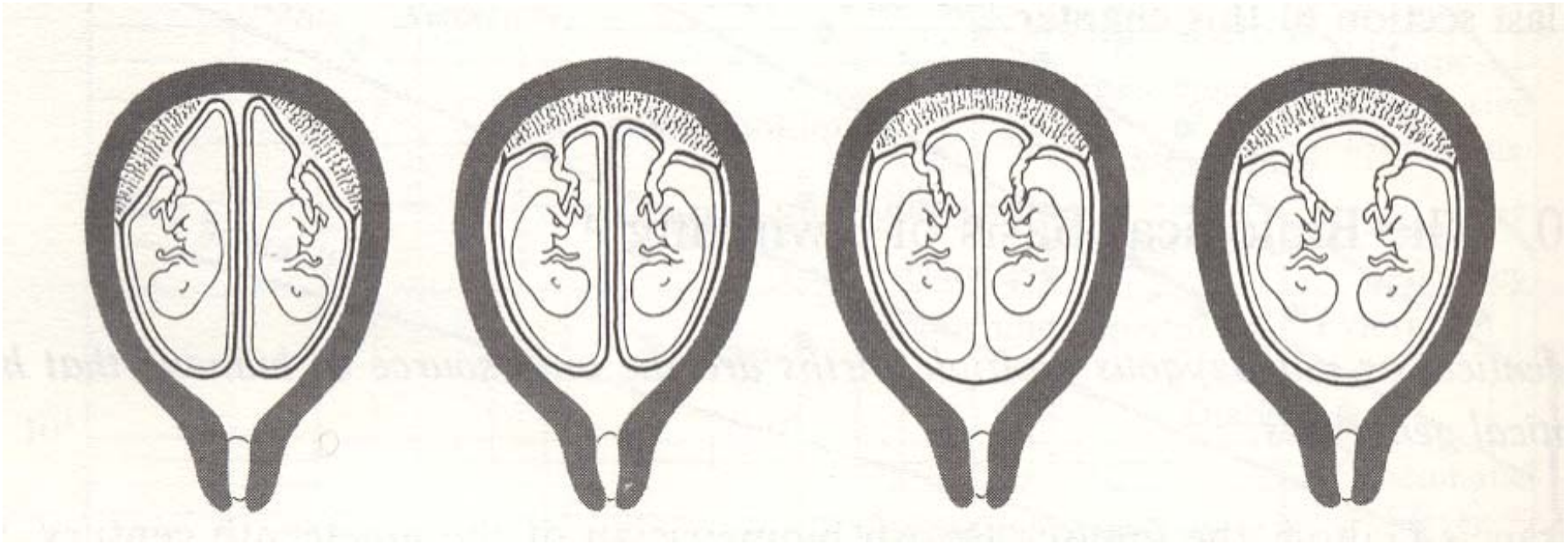
MZ = mono-chorionic

MZ = identical at marker loci

(except for rare mutations)

MZ and DZ twins: determining zygosity using ABI Profiler™ genotyping (9 STR markers + sex)

EEA: Placentation and zygosity

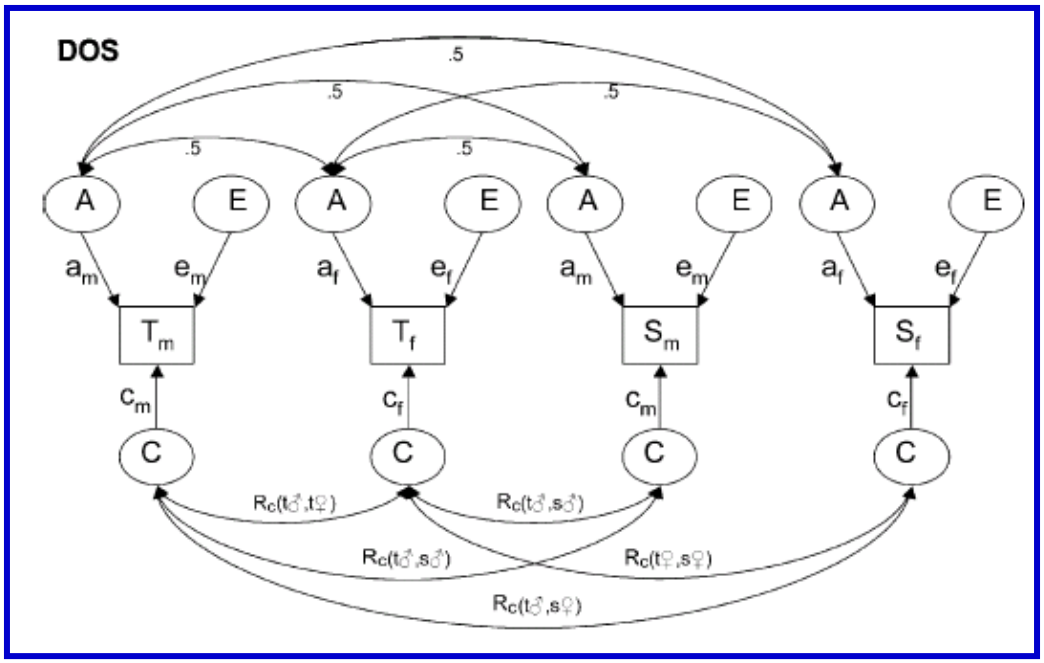


Dichorionic Two placentas	Dichorionic Fused placentas	Monochorionic Diamniotic	Monochorionic Monoamniotic
MZ 19% **DZ 58%**	**MZ 14%** **DZ 42%**	**MZ 63%** **DZ 0%**	**MZ 4%** **DZ 0%**

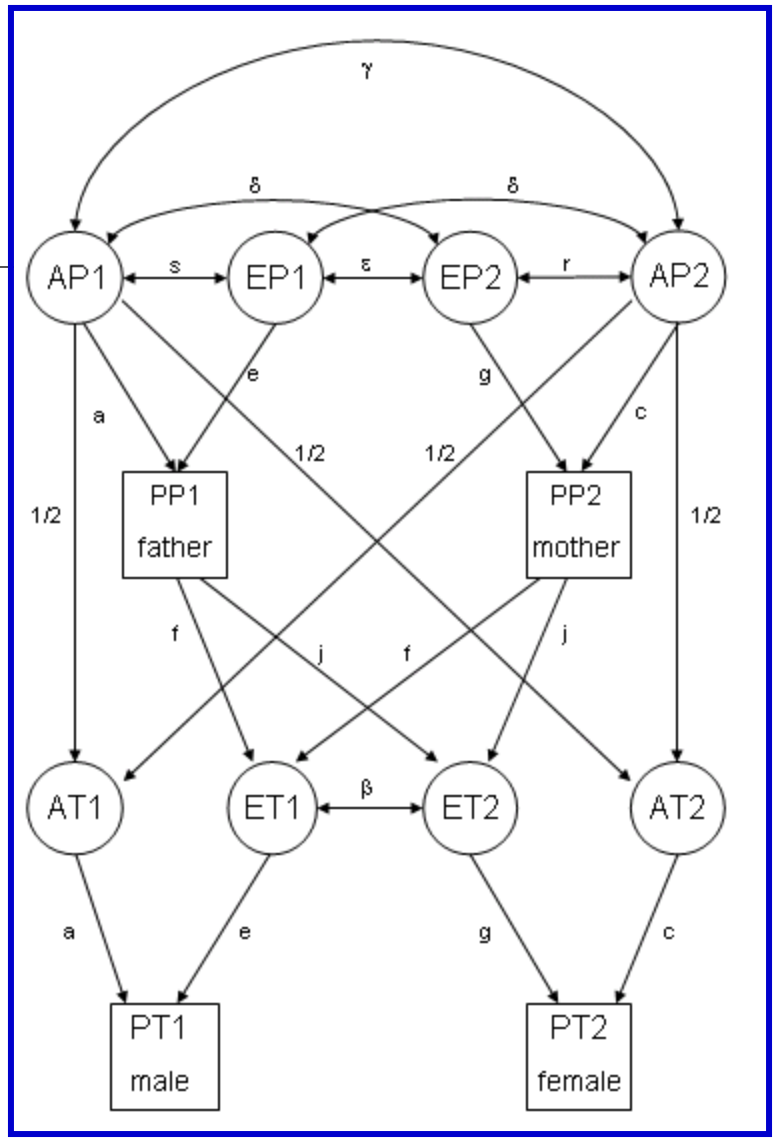
Representative?

- Test for “twin effects”: Include other family members (e.g. siblings of twins)
- Look at resemblance in twins of mistaken zygosity (parents say DZ, testing says MZ)

Extended twin designs

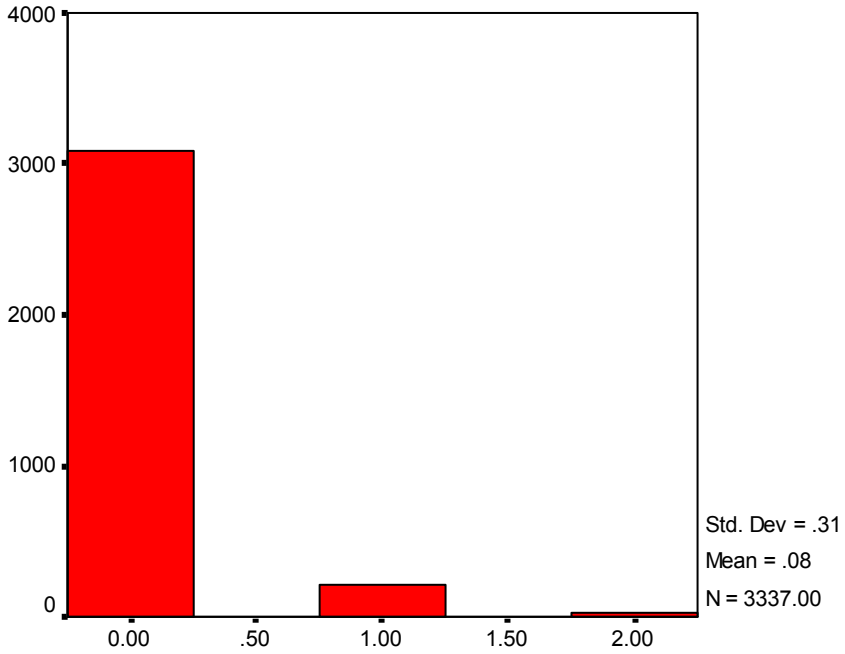


Twin and sibs: tests of special twin effects;
increased power to detect Common environment, Non-additive genetic effects



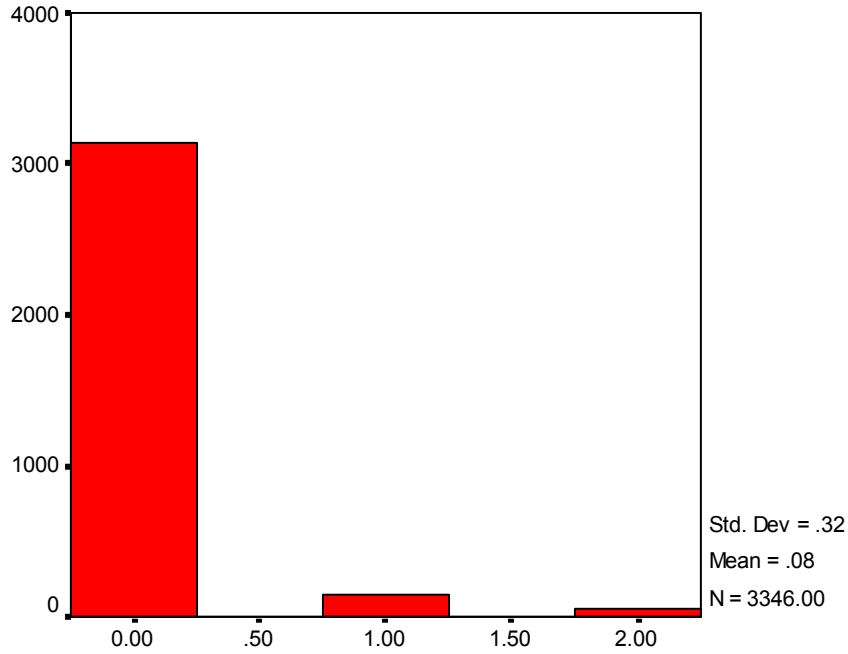
Twin and parents: genetic and cultural transmission, GE correlation, assortment

Individual differences in response to CBCL items on gender identity (3 point scale)



110 - liever van het andere geslacht

Rather be of the other sex

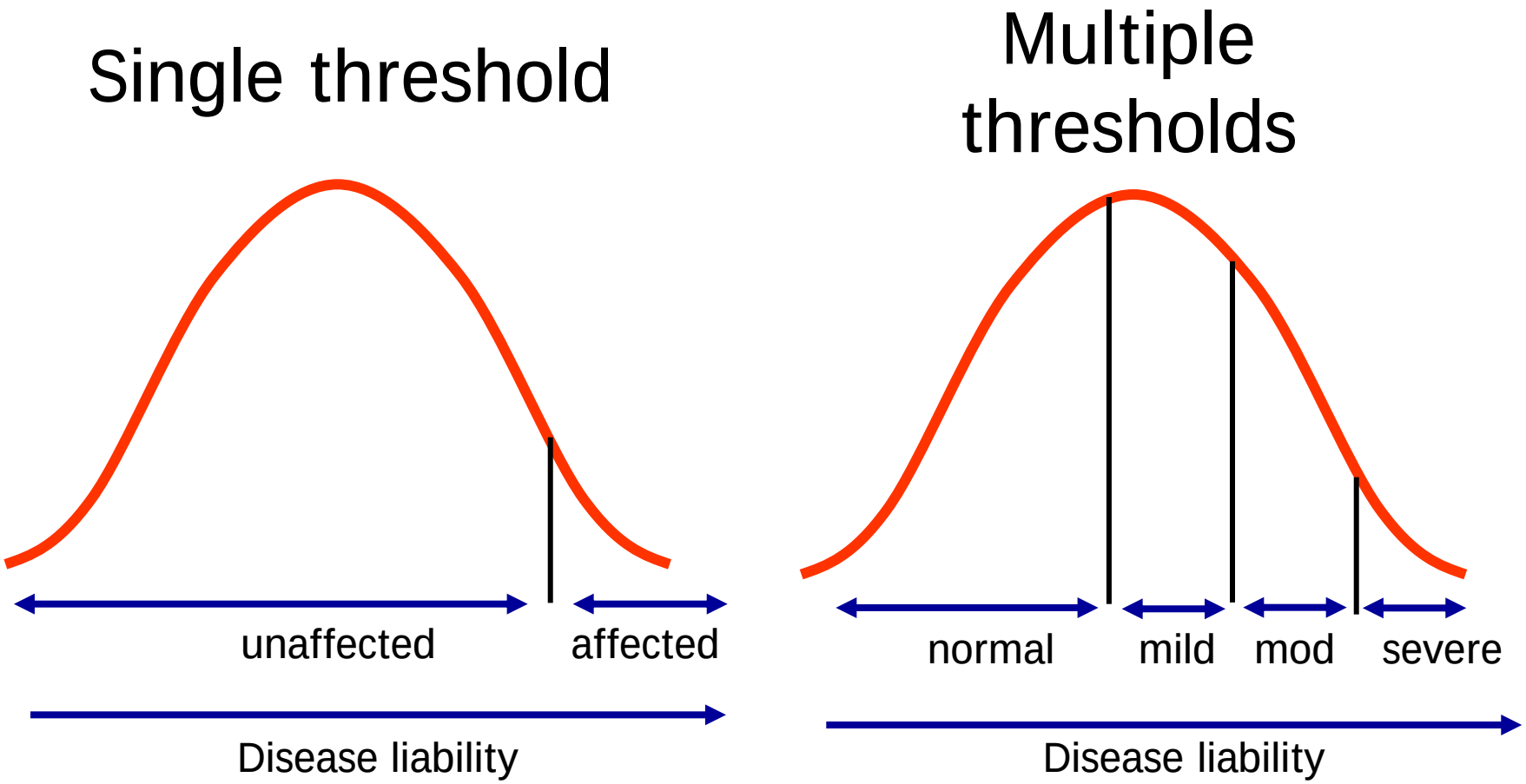


5 - andere geslacht

I am of the other sex

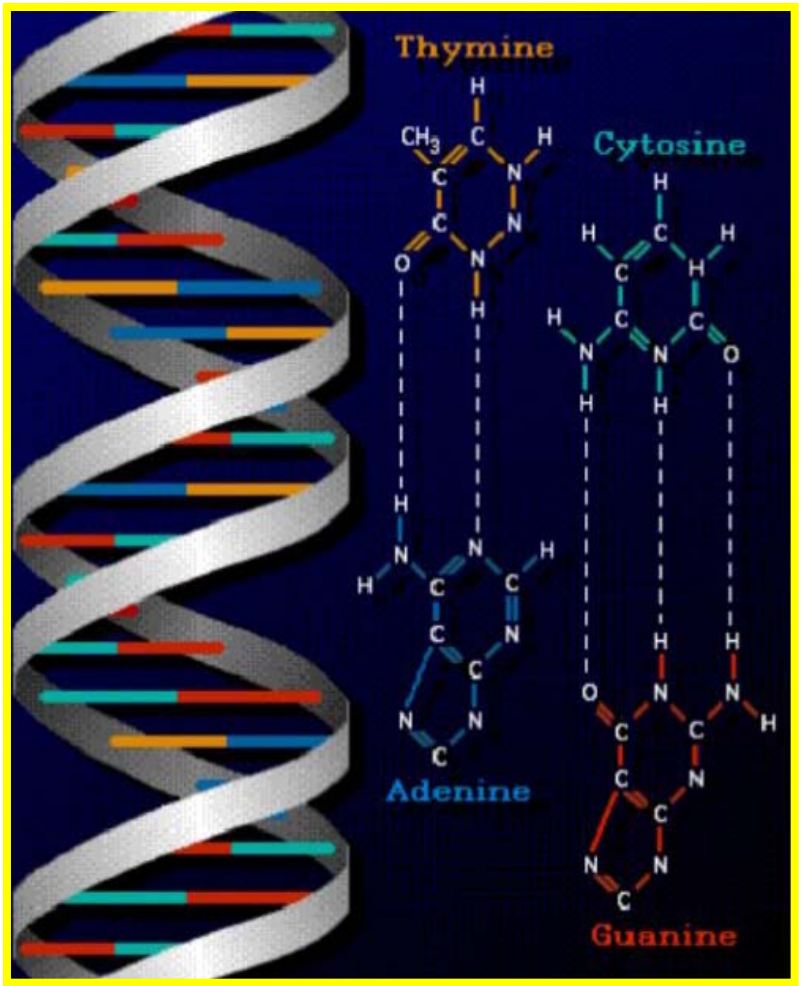
van Beijsterveldt et al. Genetic and environmental influences on cross-gender behavior and relation to behavior problems: a study of Dutch twins at ages 7 and 10 years. Arch Sex Behav. 2006, 35(6):647-58

Multifactorial Threshold Model of Disease



Genetic differences
= differences in DNA sequence

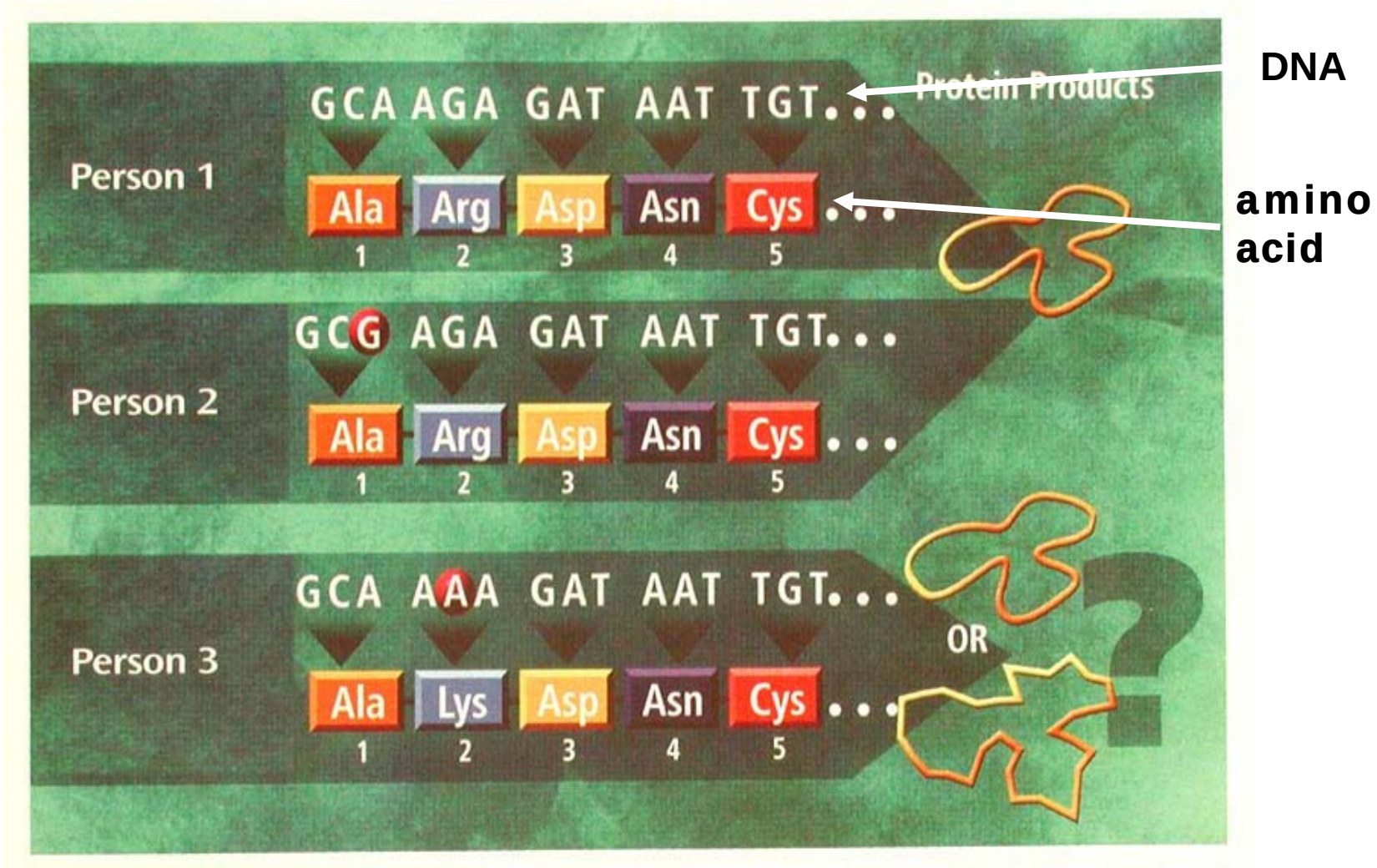
Human-Human 1:1000 = 0.1%



Human-Chimp 1:100 = 1%

Human-Mouse 1:8 = 15%

Sequence differences between individuals

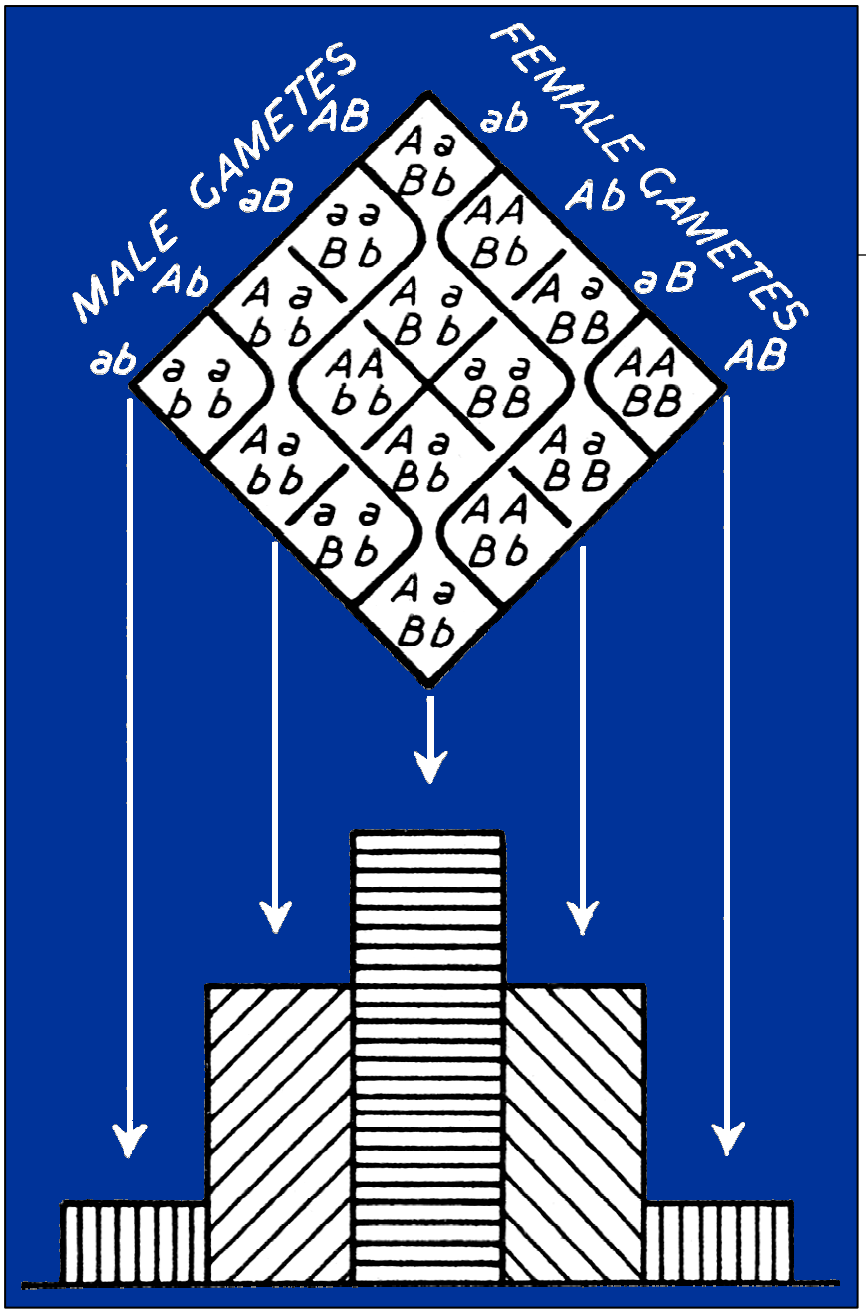


Resemblance between relatives caused by:

- shared Genes (G = A + D)
- environment Common to family members (C)

Differences between relatives caused by:

- non-shared Genes
- Unique environment (U or E)



Punnett square

Genetics explains both the *resemblances* and the *differences* of family members (e.g. sibs).

Distribution of phenotypes in offspring of two heterozygous parents (AaBb).
(2 genes (A & B) with additive allelic effects).

K Mather, Biometrical Genetics, Dover Publ, 1949

what is a gene?

In 2003, estimates from gene-prediction programs suggested there are 24,500 or fewer protein-coding genes.

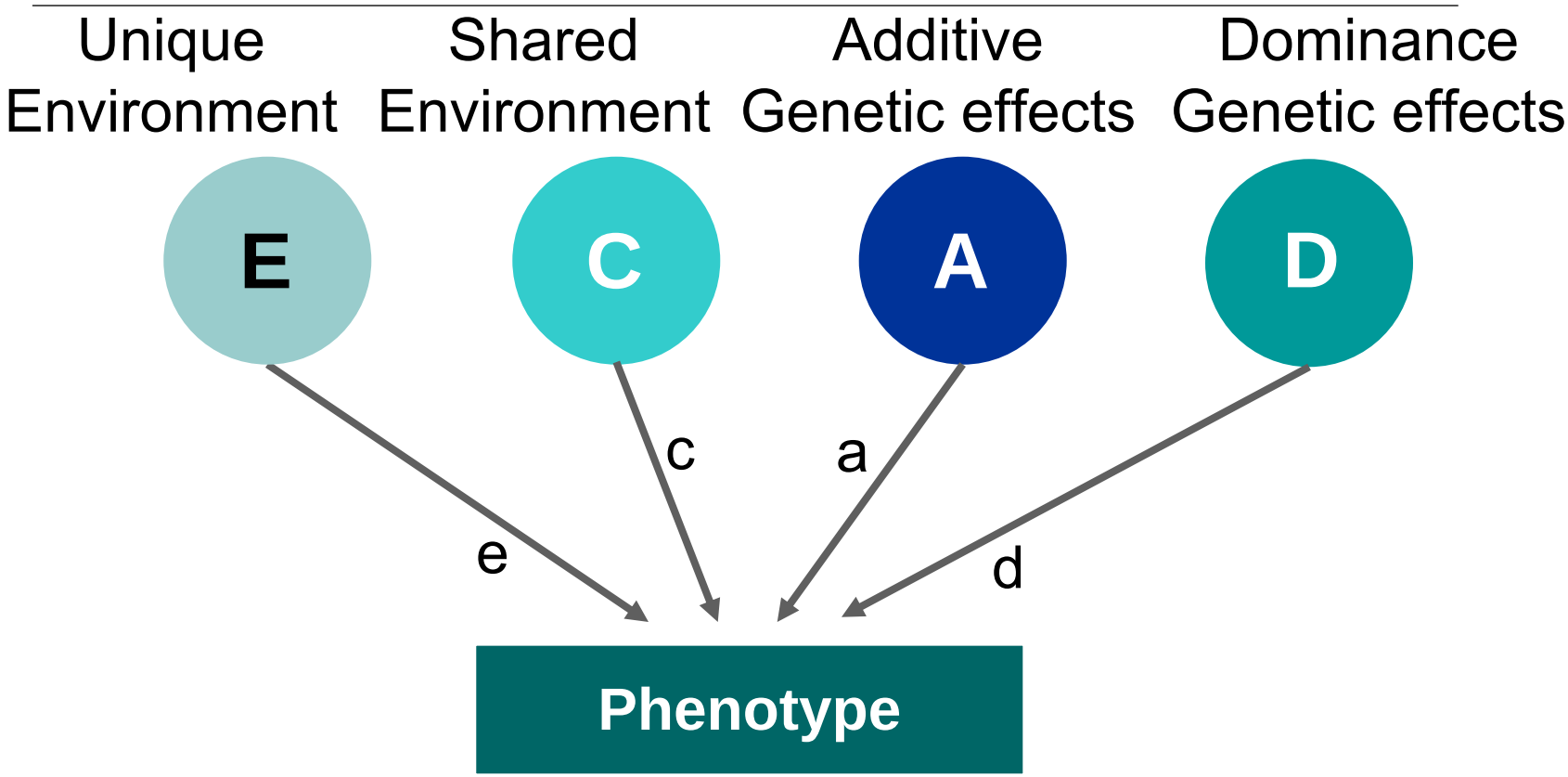
The Ensembl genome-annotation system estimates them at 23,299. Perhaps the biggest obstacle to gene counting is that the definition of a gene is unclear.

Is a gene:

- a heritable unit corresponding to an observable phenotype
- a packet of genetic information that encodes a protein
- a packet of genetic information that encodes RNA
- must it be translated ?
- are genes genes if they are not expressed ?

TK Attwood: The Babel of Bioinformatics, Science, 290:471, 2000

A gene is a latent factor



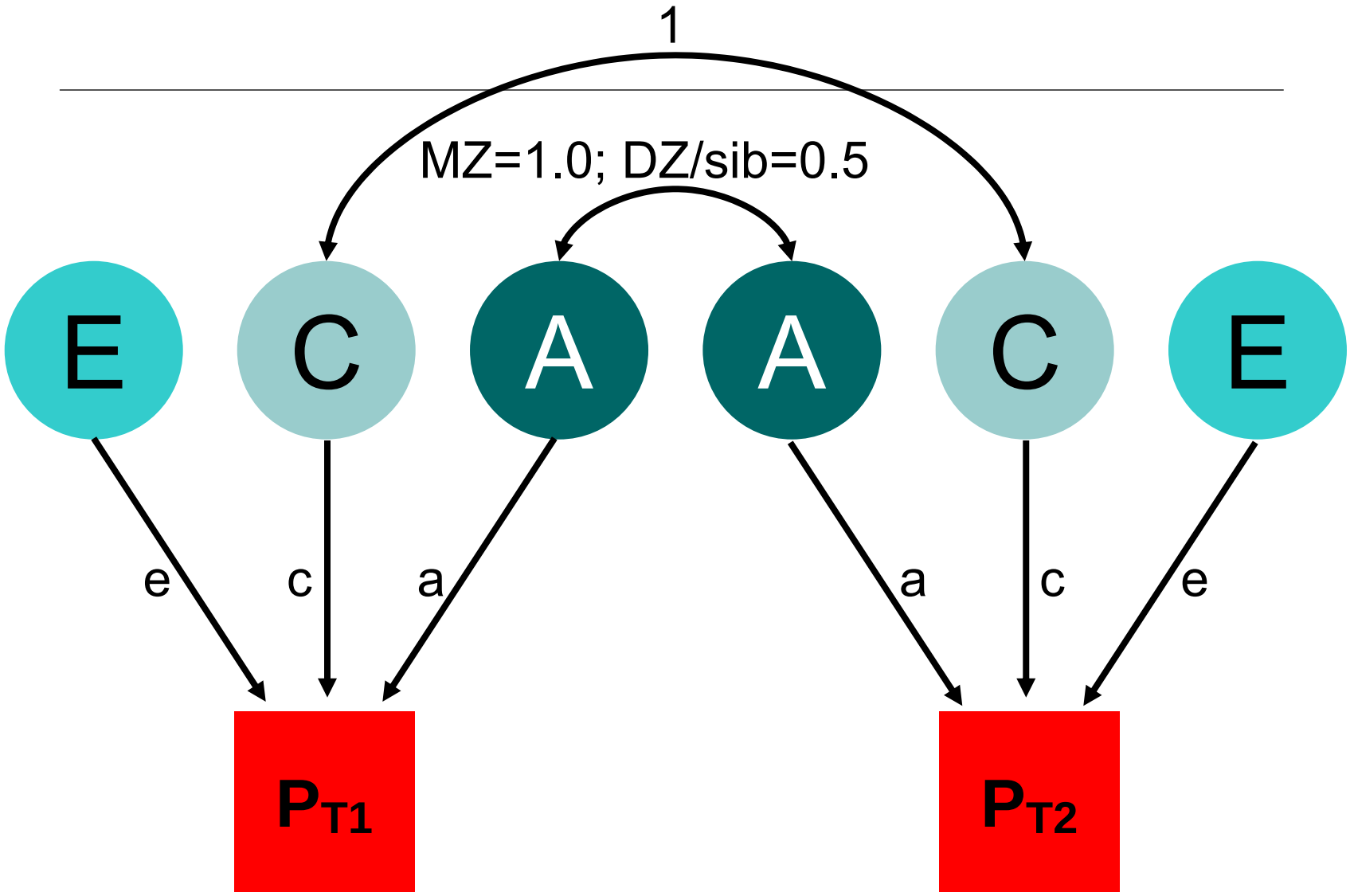
$$P = eE + aA + cC + dD$$

(plus epistasis, assortment, GE interaction, ….)

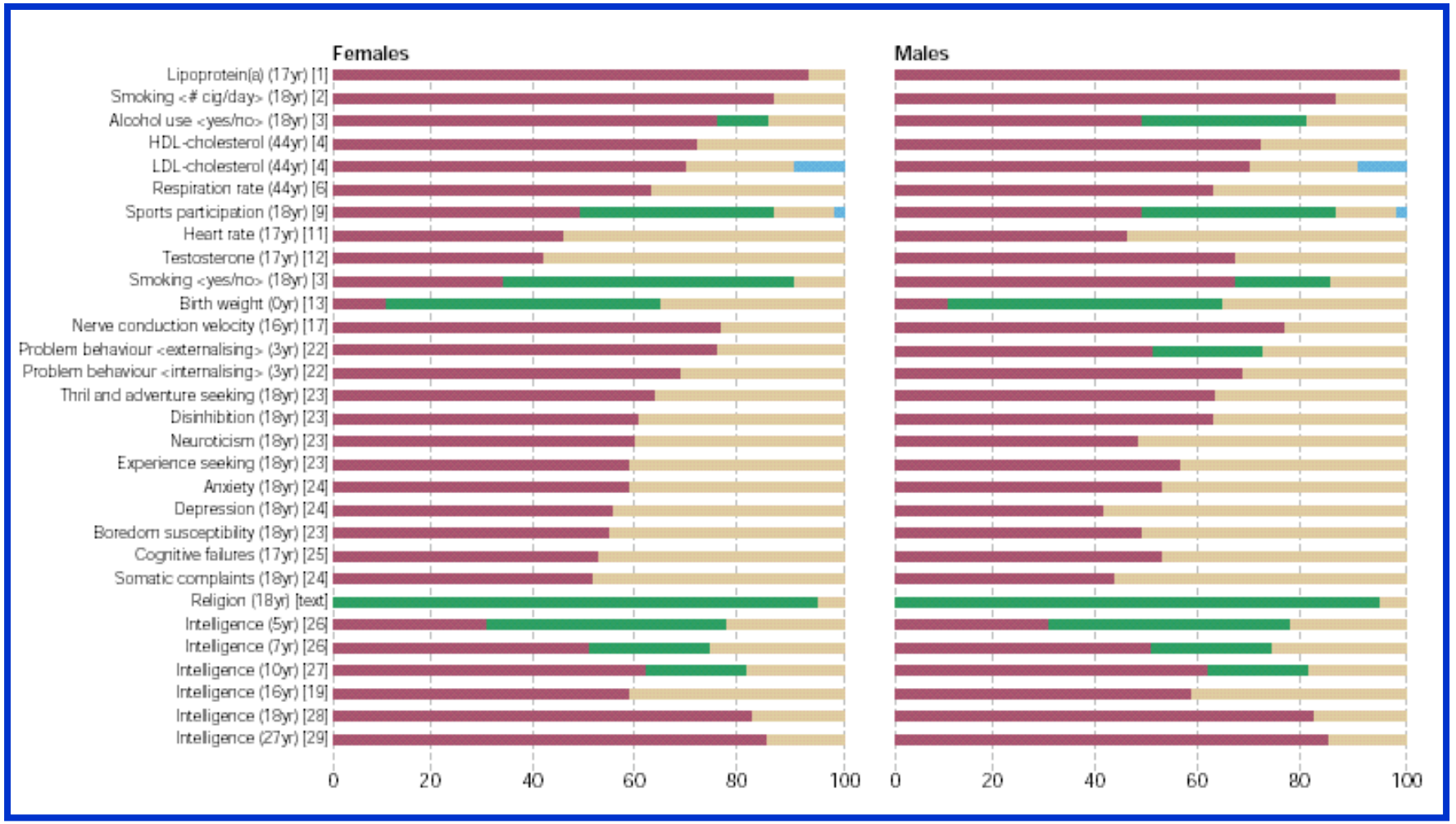
Structural equation modeling

- Both continuous and categorical variables
- Systematic approach to hypothesis testing
- Tests of significance (for effects of G, D, C)
- Can be extended to:
  - More complex questions
  - Multiple variables
  - Other relatives

ACE Model for univariate twin / sib data



Heritability estimates in males and females (ANTR twin data)



Genes

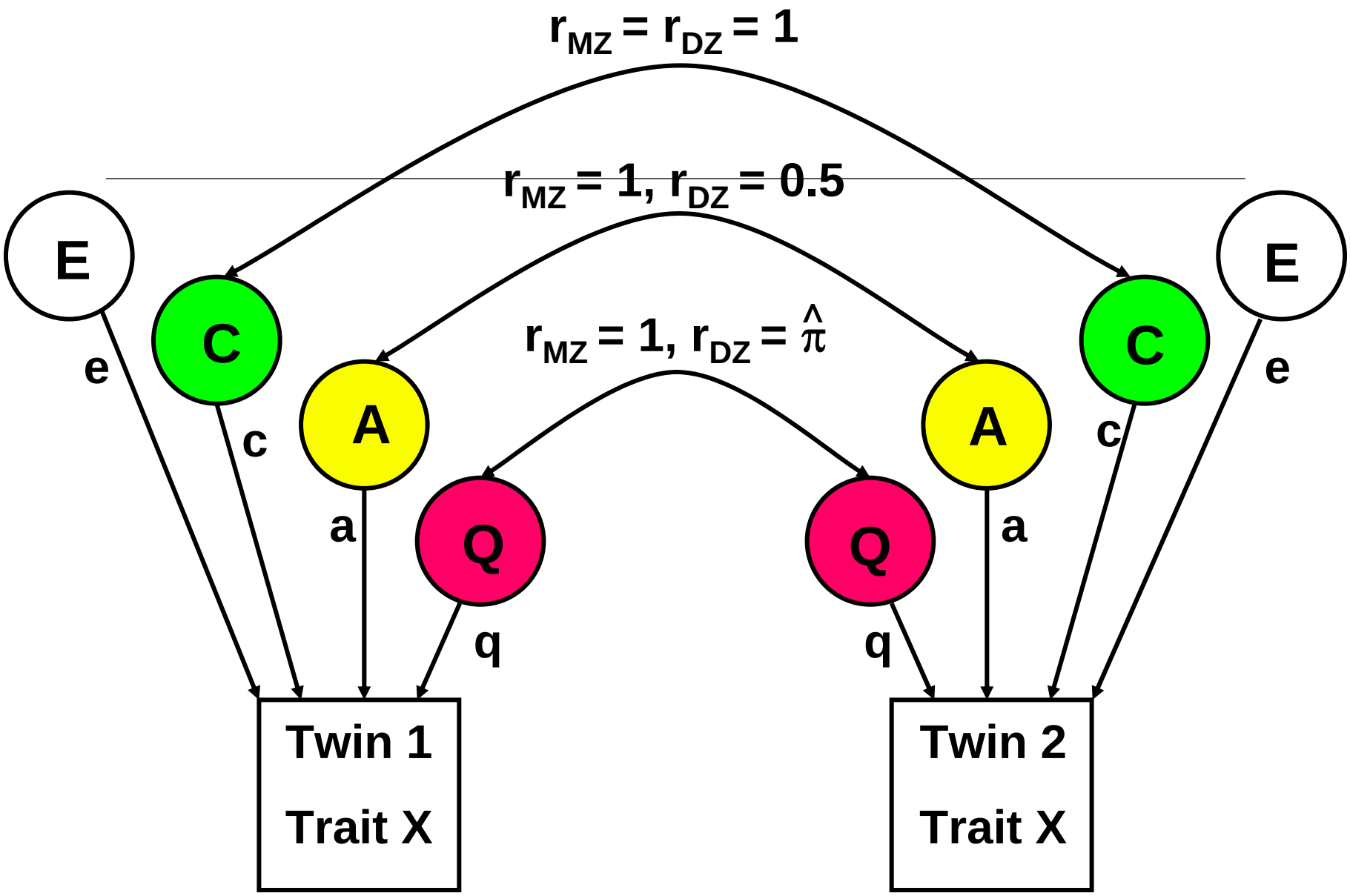
Shared environment

Unique environment

Boomsma et al., 2002, Nat Review Genet

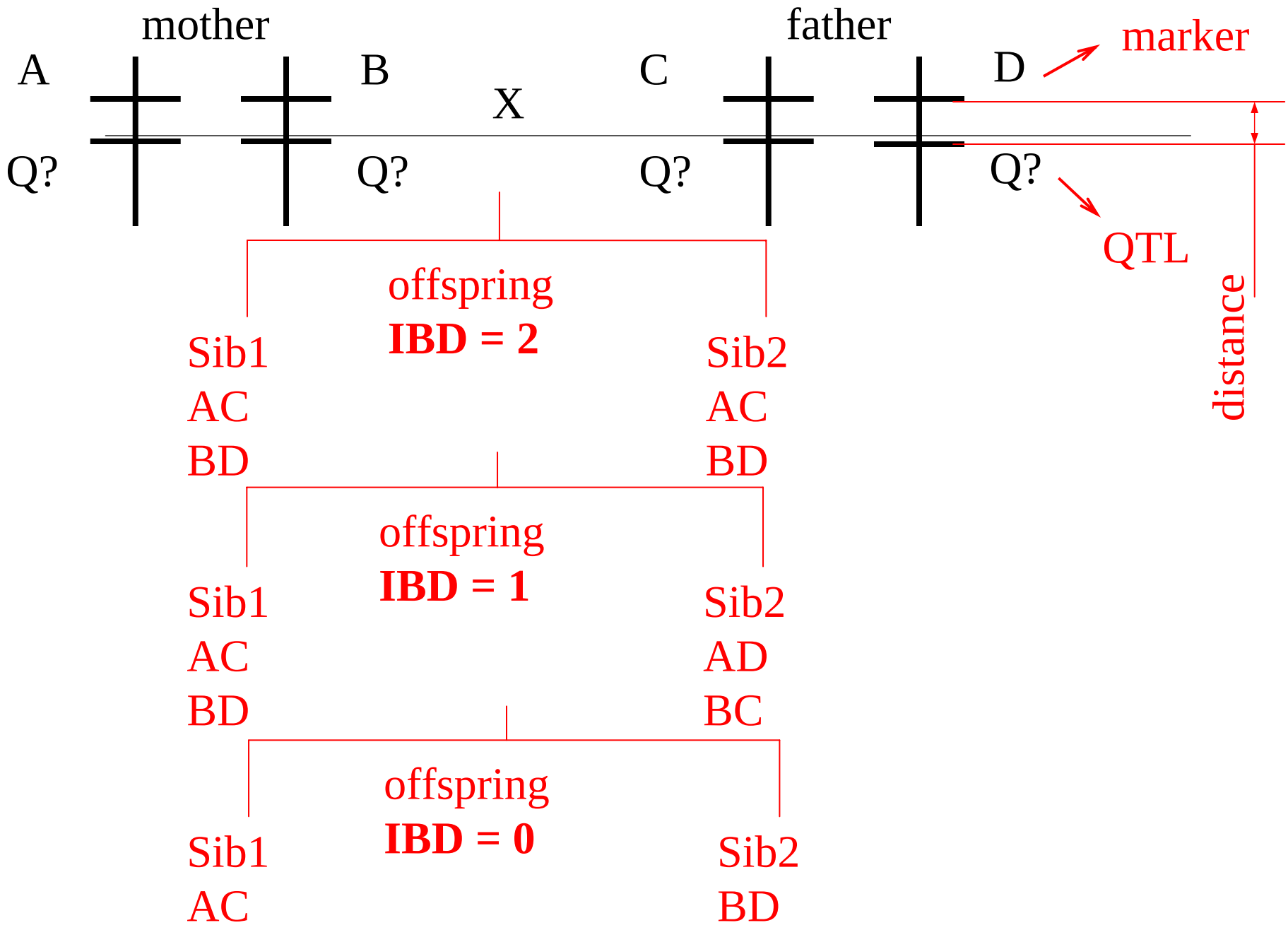
3 Stages of Genetic Mapping

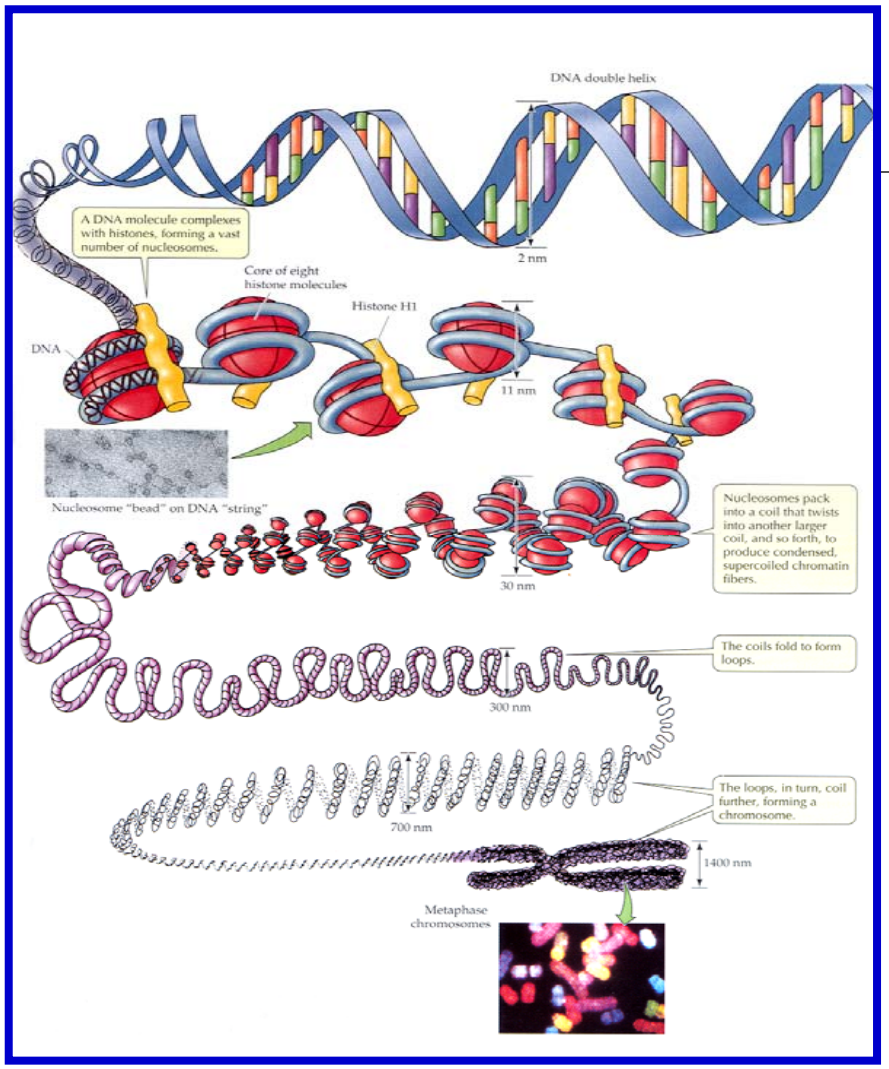
- Are there genes influencing this trait?
  - Genetic epidemiological studies
- Where are those genes?
  - Linkage analysis
  - (look for quantitatve trait loci: QTL)
- What are those genes?
  - Association analysis



π (QTL correlation) is estimated from IBD (identity by descent) data

IBD data: A fully informative mating





Linkage: tracking anonymous DNA markers close to genes of interest in families / sibling pairs.

- "blind" search, low power
- new genes, new mechanisms

Genetic association (based on linkage disequilibrium): direct comparison of regulatory and coding sequences in candidate genes (or markers close to candidate genes).

- high power, high type I/II error rate
- which candidates ?
- Genome wide (GWA)

Anxiety (NL; longitudinal survey data)

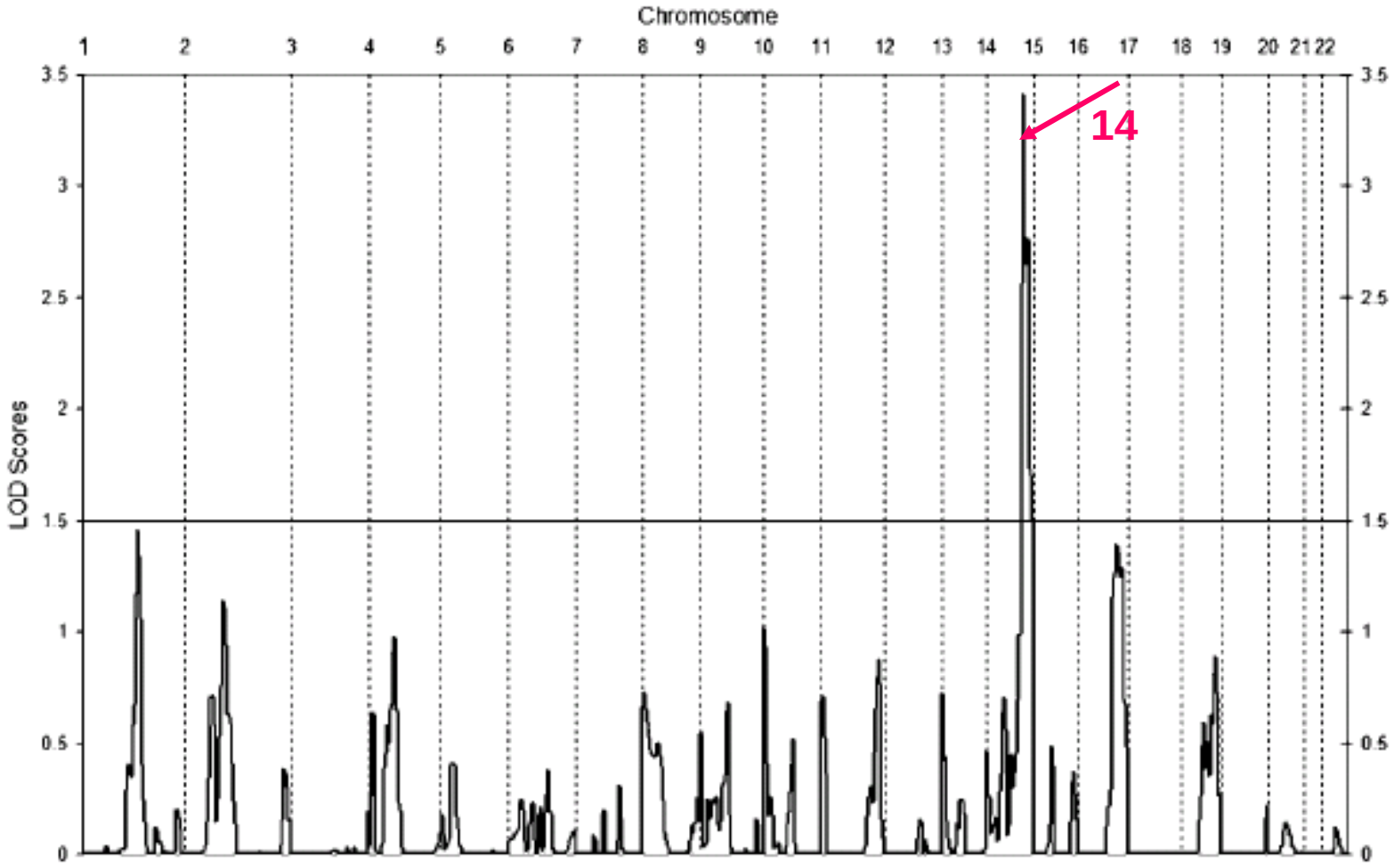


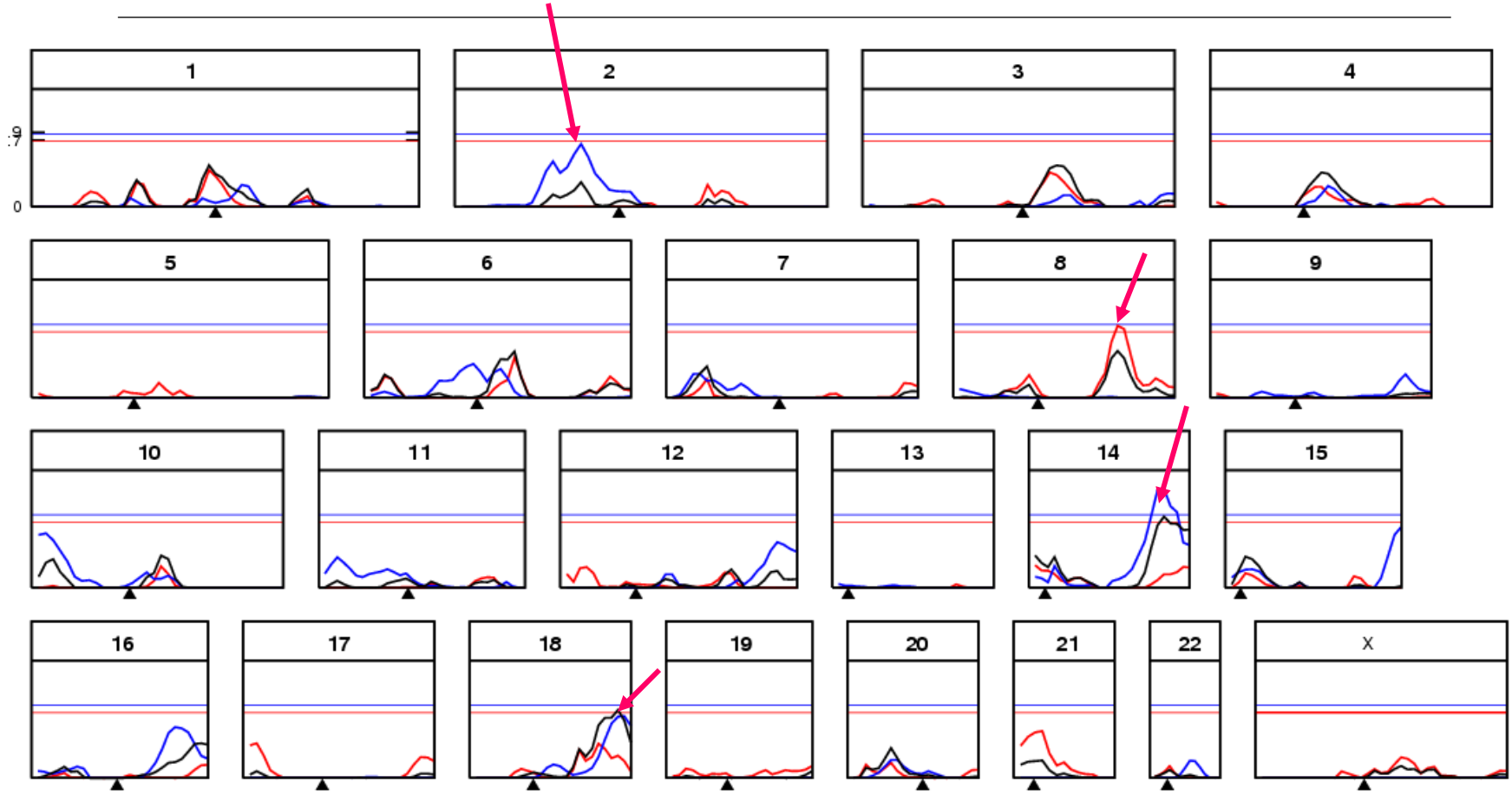
Figure 1 Results of the genome-wide linkage analysis of the anxiety scores averaged across the five occasions.

Middeldorp et al, Molecular Psychiatry, 2008

Neuroticism (endophenotype for depression and anxiety) Data from the Netherlands and Australia (Wray et al. (Arch General Psychiatry, in press))

- 19,635 sibling pairs with data for neuroticism up to five times over a period of up to 22 years.
- **5,424 sib pairs genotyped** with microsatellite markers; pairs concordant or discordant with respect to extreme neuroticism scores were genotyped preferentially.
- 38% (AU) and 51% (NL) of parents were genotyped.
- The average distance between markers was 8.2 cM (Australia) and 11 cM (Netherlands).
- Non-parametric linkage analysis in Merlin-Regress for mean neuroticism score across time.
- Empirical LOD thresholds for suggestive linkage derived from Merlin – simulate.

Neuroticism Netherlands and Australia



90 cM on chr 2
105 cM on chr 14

130 cM on chr 8
115 cM on chr 18

Linkage Analysis

- Models the **covariance structure** among family members

- Marker sharing between relatives
  - Identifies large regions
    - Include several candidates
- Complex disease
  - Scans on sets of small families popular
  - No strong assumptions about disease alleles
  - Low power
  - Limited resolution

Association

- Models **“mean”** values
- Looks for correlation between *specific alleles* and a phenotype (quantitative trait value, disease risk)
- E.g. cases and controls (affected / unaffected)
- Or high and low scoring Ss

Association

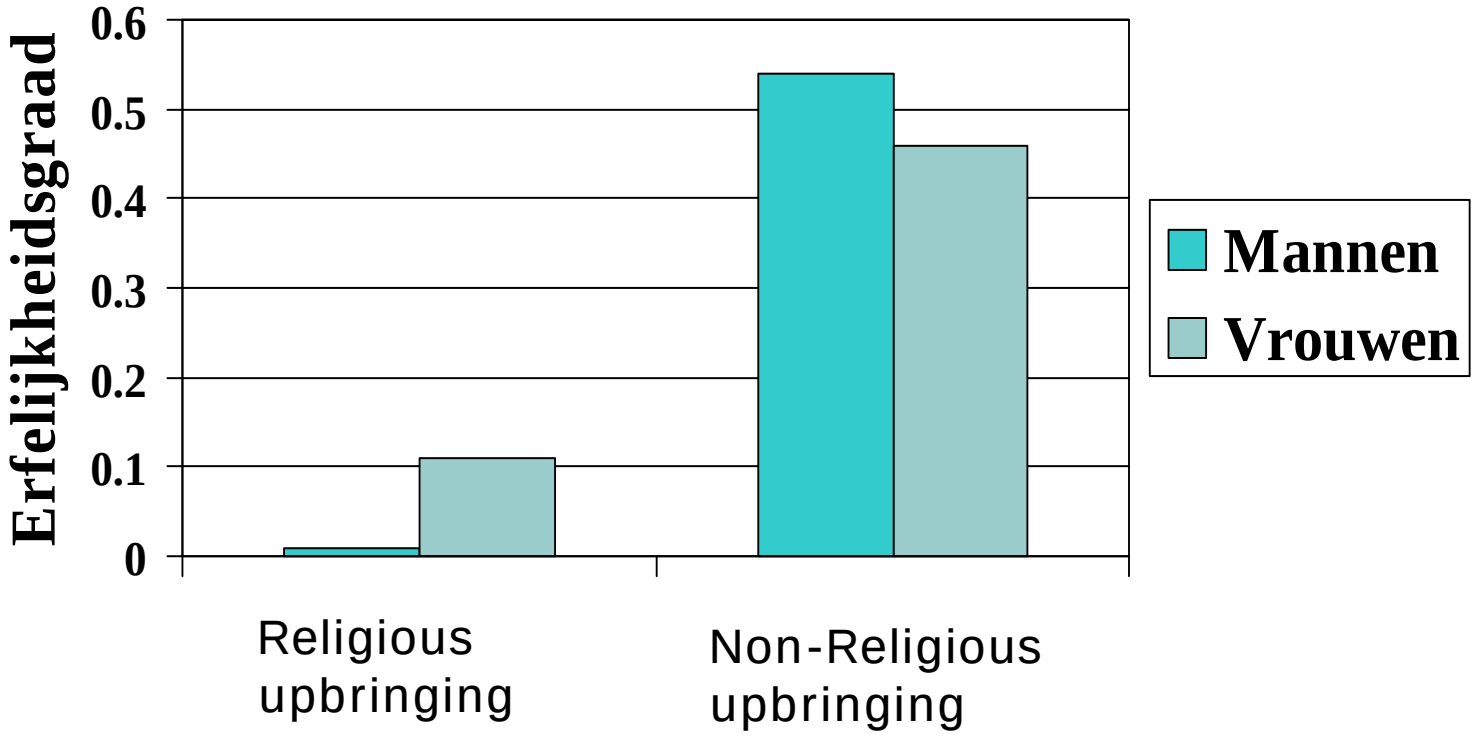
- More sensitive to small effects
- Need to “guess” gene/alleles (“candidate gene”) or be close enough for linkage disequilibrium with nearby loci (GWA: Genome Wide Association)
- May get spurious association (“stratification”) – need to have genetic controls to be convinced
- May get too many “positive” results (if the number of tests is large)

Types of Twin Studies I

Classical MZ -DZ comparison:

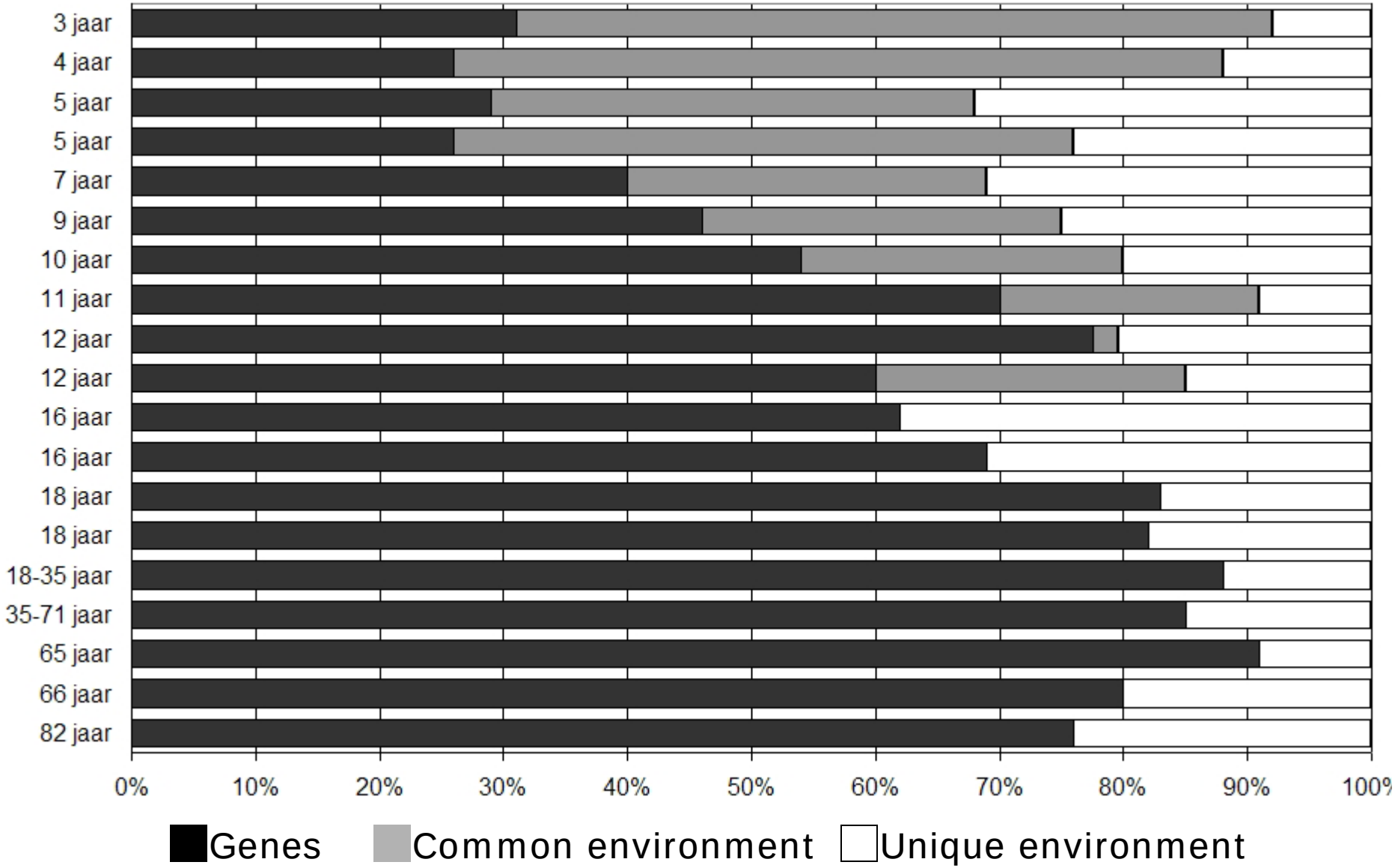
- age differences in heritability
- sex differences in heritability
- genotype x environment interaction
- causal models
- multivariate genetic analyses

Genotype x Environment interaction: Heritability of Disinhibition as a function of religious upbringing

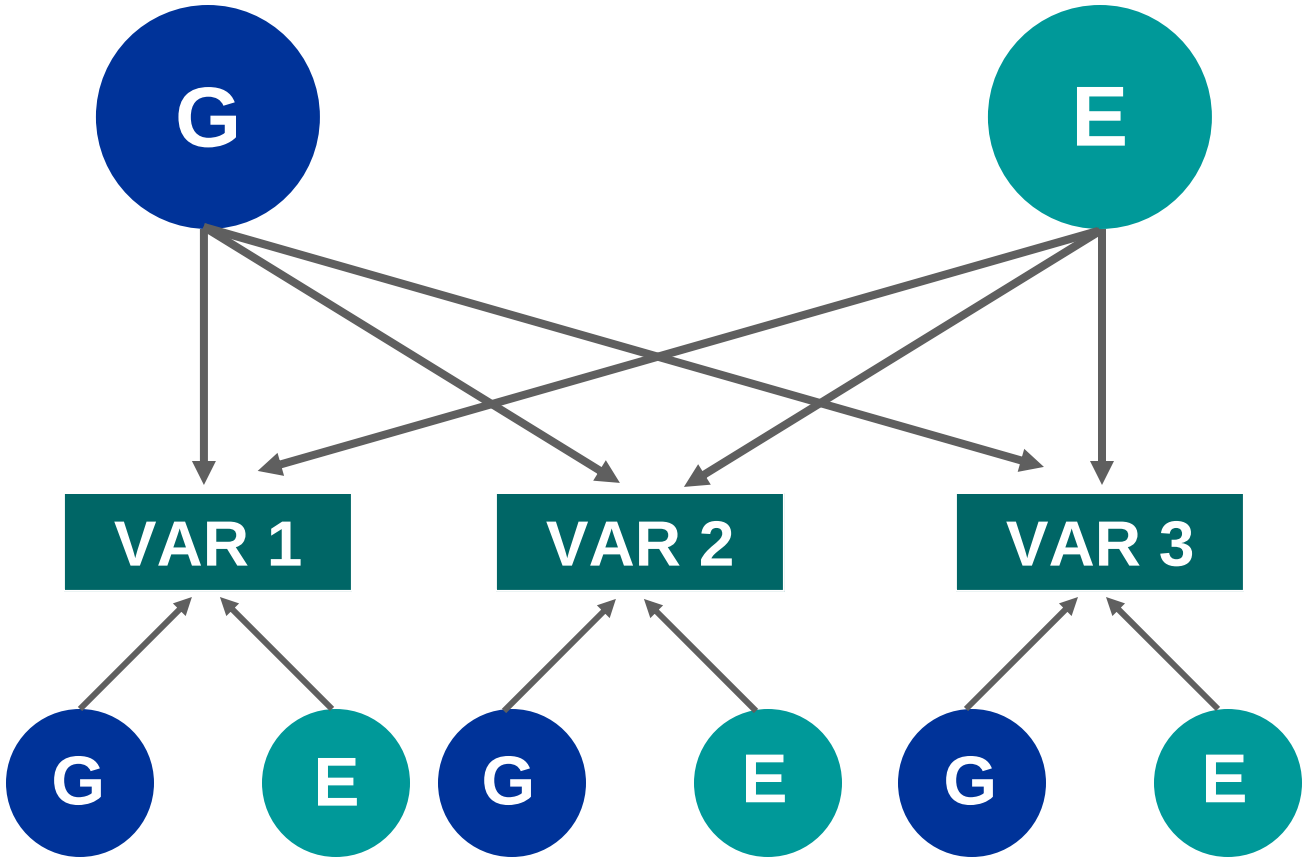


D.I. Boomsma et al. (1999) Twin Research 2, 115-125

IQ heritability (gene x age interaction)



Multivariate analysis: Genetic factor model: do the same latent factors influence multiple traits ?



Classical twin design revisited: Heritability estimation without MZ twins

Why do we use the average sib values of

$r_a = 0.5$ and $r_d = 0.25$

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

OPEN ACCESS Freely available online PLoS GENETICS

Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings

Peter M. Visscher*, Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, Nicholas G. Martin

Genetic Epidemiology Group, Queensland Institute of Medical Research, Brisbane, Australia

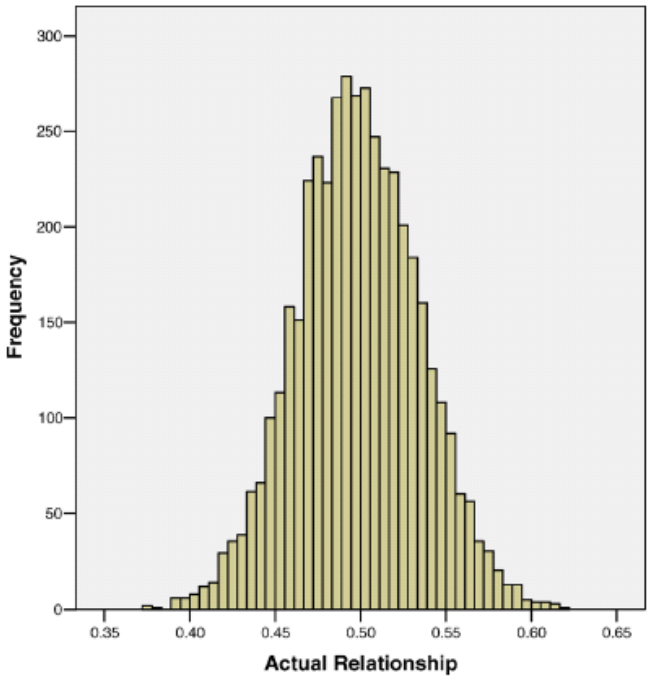


Figure 1. Empirical Distribution of Actual Additive Genetic Relationships of 4,401 Quasi-Independent Pairs of Full Sibs

Histogram of the genome-wide additive genetic relationships of full-sib pairs estimated from genetic markers.

DOI: 10.1371/journal.pgen.0020041.g001

Table 2. ML Estimates of Heritability of Height from Genome-Wide IBD Sharing between Sib Pairs

Data	Model	Estimates (95% CI)	
		f^2	h^2
Adolescents (n = 931)	FAE	0.00 (0.00–0.43)	0.80 (0.00–0.90)
	FE	0.40 (0.34–0.45)	
Adults (n = 2,444)	FAE	0.00 (0.00–0.18)	0.80 (0.43–0.86)
	FE	0.39 (0.36–0.43)	
Combined (n = 3,375)	FAE	0.00 (0.00–0.17)	0.80 (0.46–0.85)
	FE	0.39 (0.36–0.42)	

Types of Twin Studies II

- Co-twin control study
- Extended twin study including:
  - parents: assortative mating
    cultural transmission
  - siblings: social interaction
  - MZ offspring: maternal effects

Monozygotic Twins Discordant for a trait: Identical genomes; differences caused by Environment?

- Different chromosome constitutions because of post-zygotic non-disjunction: e.g. MZ male-female 46,XY - 45,XO
- Differential *methylation* (imprinted genes)
- CNV (copy number variation)
- Skewed X chromosome inactivation in female MZ twins
- Differential trinucleotide repeat expansion
- Post-zygotic mutation
- *Prenatal* differences
- *Postnatal* environmental differences

Martin N, Boomsma DI, Machin G. (1997) *Nature Genetics*

"environmental" factors in MZ twins discordant for Attention problems

Smoking mother during pregnancy

discordant:	**38%**	**(11/29)**	
concordant affected:	**38%**	**(8/21)**	***n.s.***
control:	**14%**	**(10/73)**	***sign.***

Placentation: % of pairs with 2 placenta's in this study:

discordant:	**38%**	**(10/26)**	
concordant affected:	**15%**	**(3/20)**	***sign.***
control:	**13%**	**(13/68)**	***sign.***

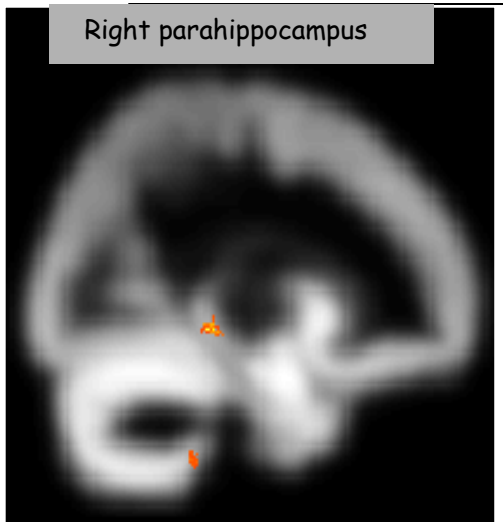
Birth weight

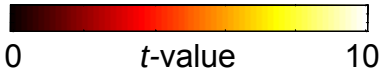
affected twin:	**2425 g**	
unaffected co-twin:	**2580 g**	***sign.***

Time in incubator

affected twin:	**11 days**	
unaffected co-twin:	**7 days**	***sign.***

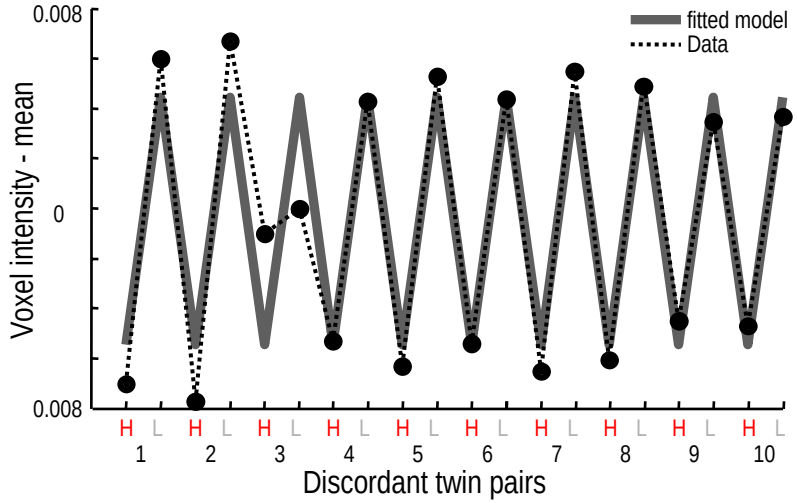
MZ twins discordant for depression risk:
Gray Matter high risk twin < GM low risk twin





(maximum t =8.08,
p < 0.0001 at
x=24,y=-34, z=-6
in MNI space)

p<0.001, min 50 voxels



Right parahippocampus is smaller in the high risk twin from discordant MZ pairs (De Geus et al., 2007)

Types of Twin Studies III

- Genotyping of MZ twins:
  - to detect variability genes
  - to estimate penetrance

- Genotyping of DZ twins to detect linkage and association

Gene – environment interaction in GWA

- Differences within MZ pairs: (mainly) function of Environmental exposure
- Are differences within pairs a function of genotype?
- i.e. is sensitivity to the environment a function of genotype?

New trends

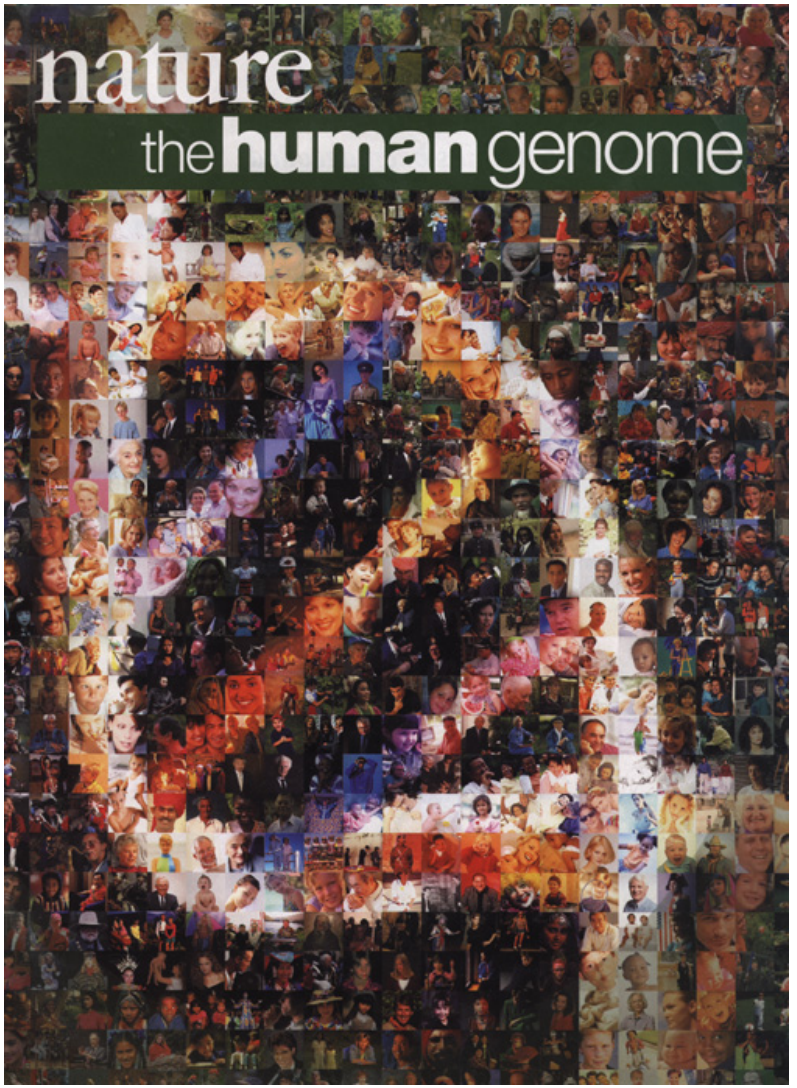
Human Genome Project: **Sequence** of the genome (base sequence)

Variation in the genome (e.g. microsatellites, SNPs, duplicons, copy number variation) related to variation in phenotype?

DNA methylation

Expression of the genome (RNA)

Metabolomics

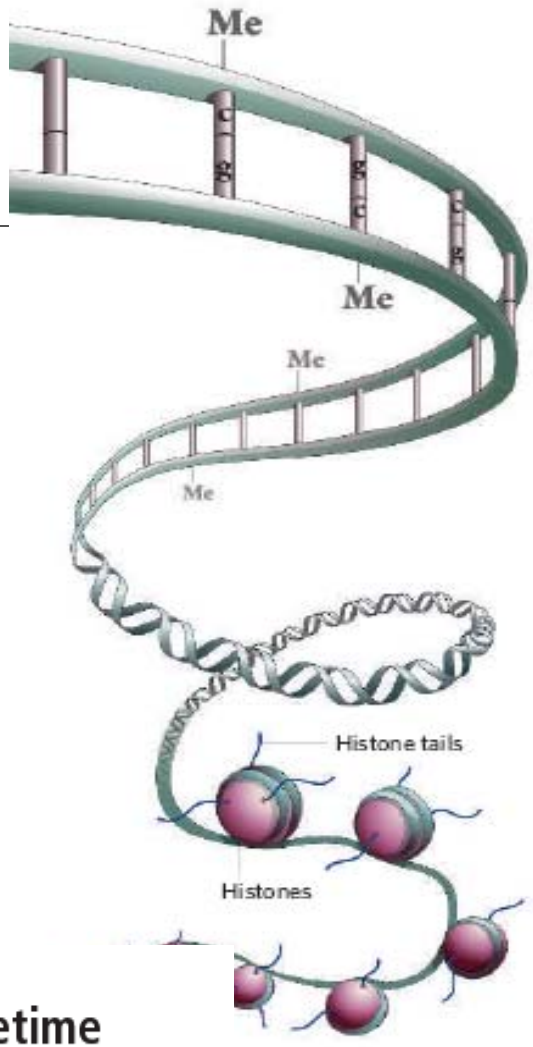


Co-twin control design

DISCORDANCE IN IDENTICAL TWINS

A role for Epigenetics?

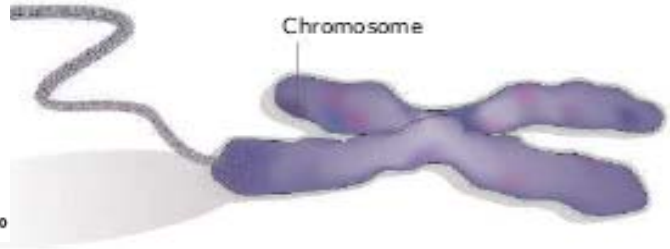
Does epigenetics depend on age?



The two main components of the epigenetic code

DNA methylation
Methyl marks added to certain DNA bases repress gene activity.

Histone modification
A combination of different molecules can attach to the 'tails' of proteins called histones. These alter the activity of the DNA wrapped around them.



PNAS

Epigenetic differences arise during the lifetime of monozygotic twins

Mario F. Fraga*, Esteban Ballestar*, Maria F. Paz*, Santiago Ropero*, Fernando Setien*, Maria L. Ballestar†, Damia Heine-Suñer‡, Juan C. Cigudosa§, Miguel Urioste¶, Javier Benitez¶, Manuel Boix-Chornet†, Abel Sanchez-Aguilera†, Charlotte Lingl, Emma Carlssonl, Pernille Poulsen**, Allan Vaag**, Zarko Stephan††, Tim D. Spector††, Yue-Zhong Wu‡‡, Christoph Plass‡‡, and Manel Esteller*§§

*Epigenetics, §Cytogenetics, and ¶Genetic Laboratories, Spanish National Cancer Centre (CNIO), Melchor Fernandez Almagro 3, 28029 Madrid, Spain; †Department of Behavioral Science, University of Valencia, 46010 Valencia, Spain; ‡Molecular Genetics Laboratory, Genetics Department, Son Dureta Hospital, 07014 Palma de Mallorca, Spain; lDepartment of Clinical Sciences, University Hospital Malmö, Lund University, S-205 02 Malmö, Sweden; **Steno

Discordant Dutch MZ pair: One of the girls has complete duplication of the spine from L4 down

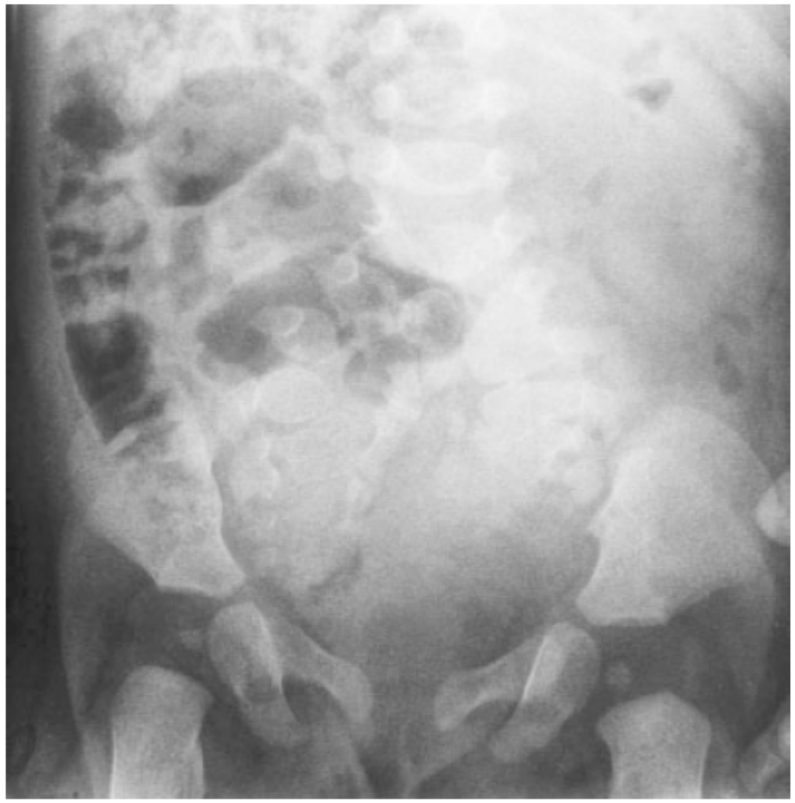


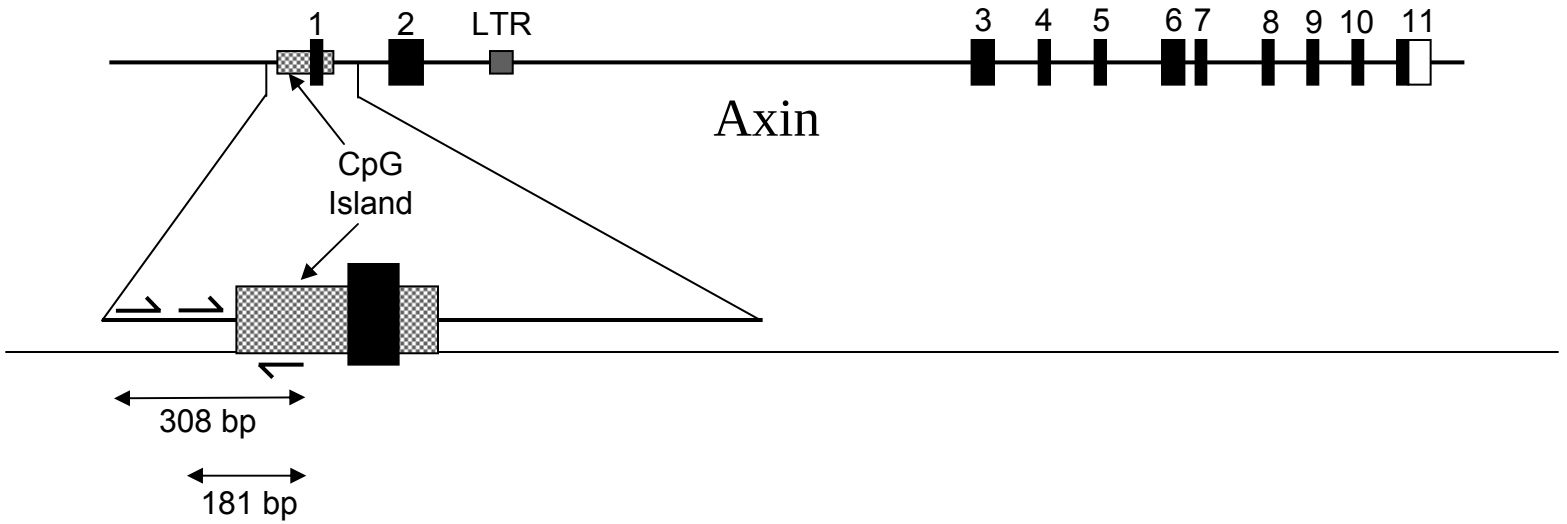
Fig. 2. Patient 1. Radiograph of the vertebral column shows complete duplication of the spine from L4 down.

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.

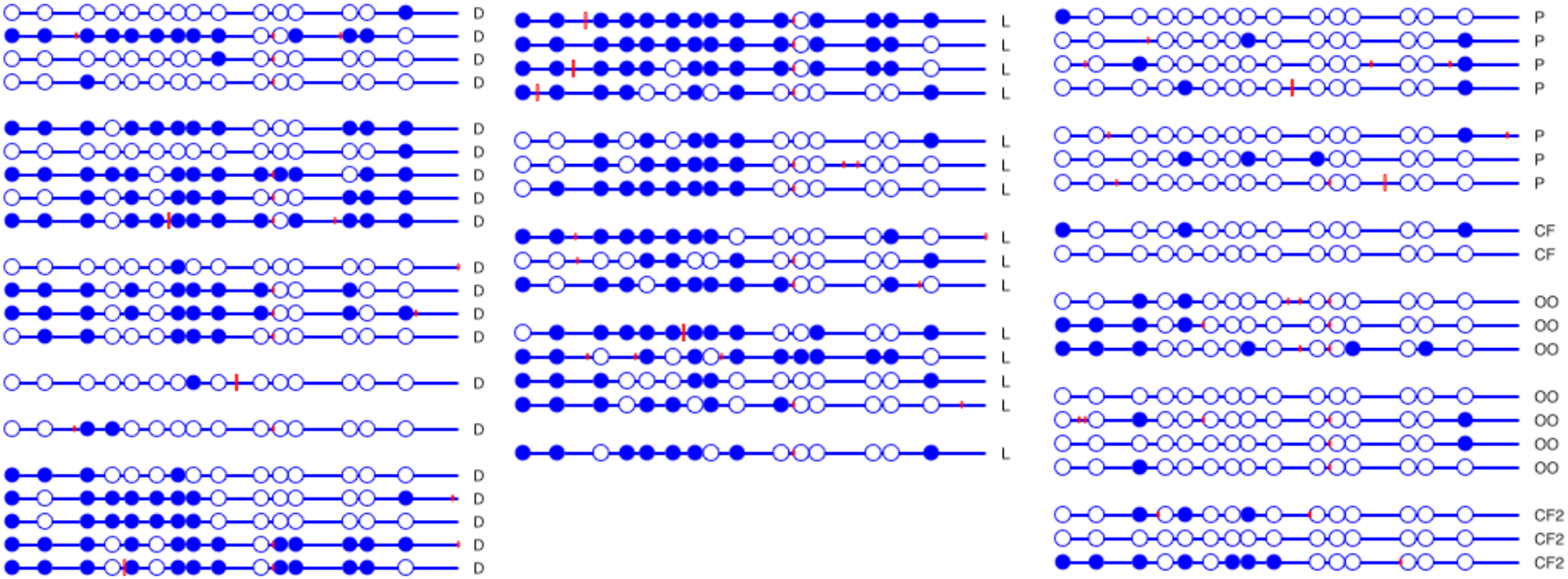
At physical examination for genetic evaluation at 4 months we saw a baby girl with epicanthal folds, but no other minor anomalies. She had a capillary nevus on her left buttock. In the anal region only a dimple was seen. The patient was operated on one day after birth, when a colostomy was made and a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

Oates et al. Increased DNA methylation at the *AXIN1* gene in an MZ twin from a pair discordant for a caudal duplication anomaly. Am J Hum Genet, 2006

Discordant caudal duplication in MZ twins



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



Human Molecular Genetics, 2007, Vol. 16, No. 5 **547–554**
doi:10.1093/hmg/ddm010
Advance Access published on March 5, 2007

Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human *IGF2/H19* locus

Bastiaan T. Heijmans[1,*], Dennis Kremer[1], Elmar W. Tobi[1], Dorret I. Boomsma[2] and P. Eline Slagboom[1]

[1]Molecular Epidemiology Section, Leiden University Medical Centre, Leiden 2333 ZC, The Netherlands and [2]Biological Psychology, Vrije Universiteit, Amsterdam 1081 BT, The Netherlands

Epigenetic variation may significantly contribute to the risk of common disease. Currently, little is known about the extent and causes of epigenetic variation. Here, we investigated the contribution of heritable influences and the combined effect of environmental and stochastic factors to variation in DNA methylation of the *IGF2/H19* locus. Moreover, we tested whether this locus was subject to age-related degeneration of epigenetic patterns as was previously suggested for global methylation. We measured methylation of the *H19* and *IGF2* differentially methylated regions (DMRs) in 196 adolescent and 176 middle-aged twins using a recently developed mass spectrometry-based method. We observed substantial variation in DNA methylation across individuals, underscoring that DNA methylation is a quantitative trait. Analysis of data in monozygotic

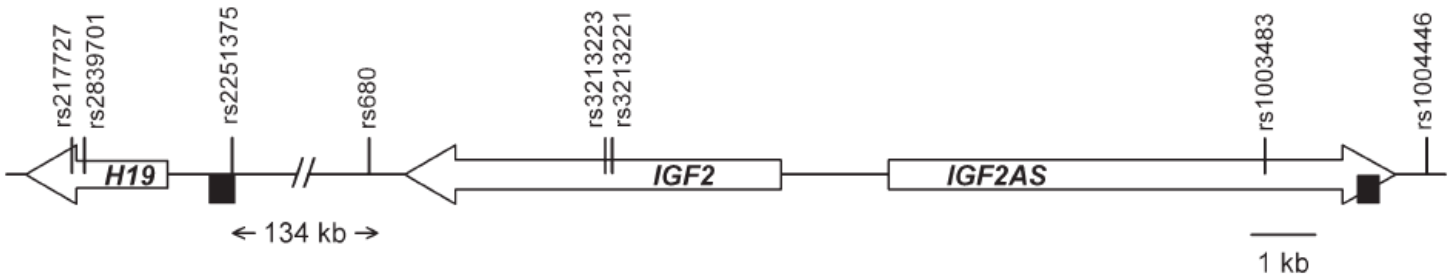


Figure 1. Overview of the *H19* and *IGF2*/*IGF2AS* genomic regions with, in black, the *H19* and *IGF2* DMRs assayed for methylation and the SNPs measured.

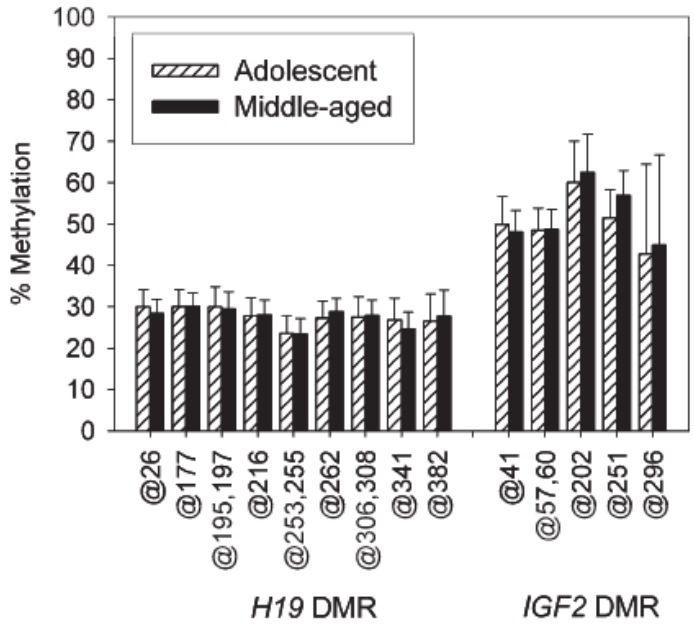


Figure 2. Mean DNA methylation for individual CpG sites in the *H19* and *IGF2* DMRs as observed in 196 adolescent (mean age 16.7) and 176 middle-aged twins (mean age 44.8).

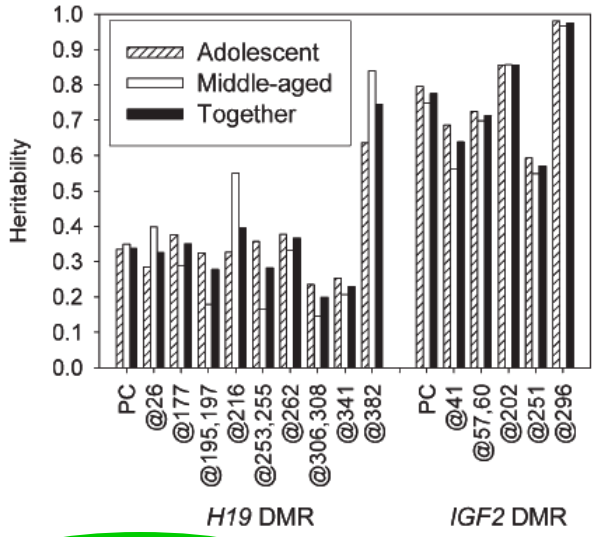
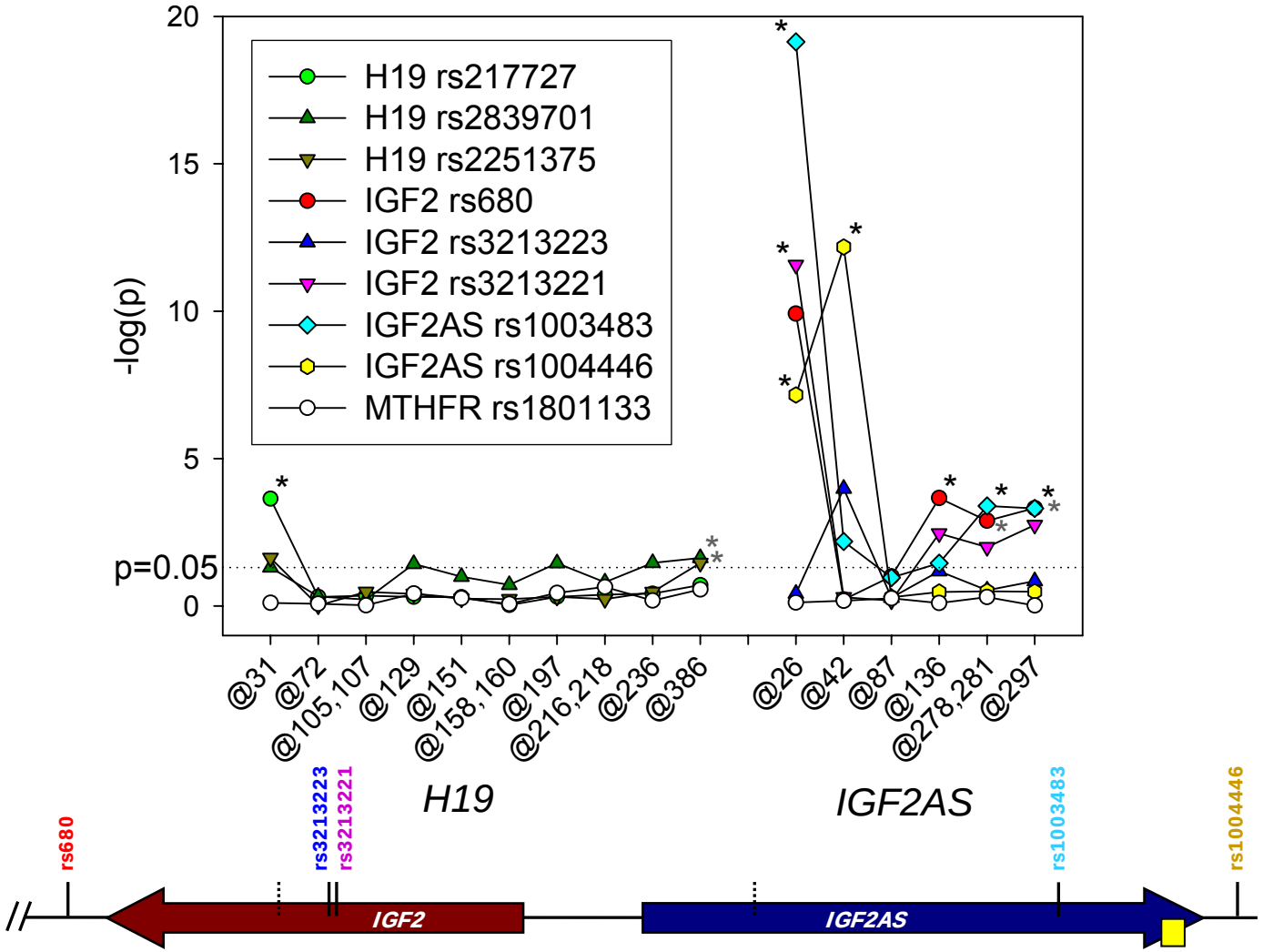


Figure 5. Heritability estimates for CpG sites and component scores based on PCA separately for adolescent and middle-aged twins as well as combined. For one CpG site (*H19* at 382 bp), the heritabilities between the two groups were significantly different ($P = 0.01$) and for another CpG site (*IGF2* DMR at 202 bp), there was evidence for a significant contribution of common environment ($P = 0.02$). PC denotes the methylation score describing the methylation per DMR obtained using PCA.

Association of SNPs in the H19 and IGF2/IGF2AS regions and the MTHFR gene with methylation of individual CpGs. Symbols denote –log(p) for the association of individual SNPs with methylation).



Please cite this article in press as: Bruder et al., Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2007.12.011

REPORT

Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles

Carl E.G. Bruder,[1,*] Arkadiusz Piotrowski,[1] Antoinet A.C.J. Gijsbers,[2,3] Robin Andersson,[4] Stephen Erickson,[5] Teresita Diaz de Ståhl,[6] Uwe Menzel,[6] Johanna Sandgren,[7] Desiree von Tell,[1] Andrzej Poplawski,[1] Michael Crowley,[1] Chiquito Crasto,[1] E. Christopher Partridge,[1] Hemant Tiwari,[5] David B. Allison,[1,5] Jan Komorowski,[4] Gert-Jan B. van Ommen,[2,3] Dorret I. Boomsma,[8] Nancy L. Pedersen,[9] Johan T. den Dunnen,[2,3] Karin Wirdefeldt,[9] and Jan P. Dumanski[1,6]

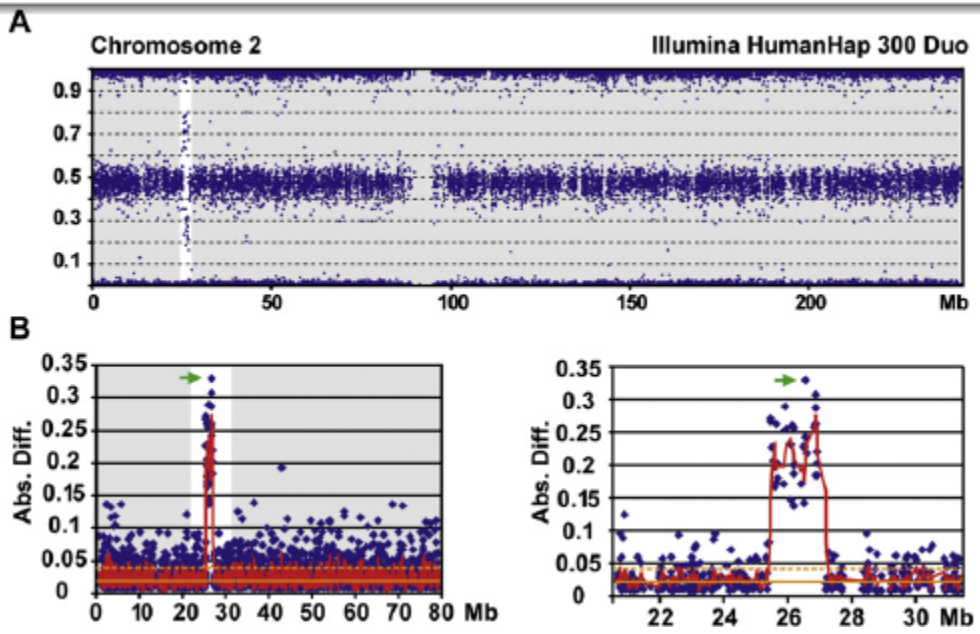


Figure 3. CNV Analysis of Twin D8 Showing the 1.6 Mb Deletion on Chromosome 2

(A) Profile of the entire chromosome 2 from Illumina HumanHap 300 Duo beadchip showing the values of SNP allele ratios. True heterozygous SNPs are expected to be distributed around a value of 0.5. In the highlighted region (white box), the allele ratios differ significantly from 0.5, indicating an imbalance in the allele signals caused by a 1.6 Mb deletion.

(B) Two enlarged views of the deleted region, plotted as values of absolute difference between the heterozygous SNP allele frequencies in twin D8 versus twin D7, calculated in a similar way as shown in Figures 1C and 1E. The red line in both graphs displays the moving average, with a period of

Unselected NTR twins (10 MZ pairs)

- CNV: gains and losses of large chunks of DNA sequence consisting of between ten thousand and five million letters (known as Copy Number Variation).

- Based on shared CNVs patterns twin pairs were easily recognized.
- However, we also detected an unexpected number of unique differences within the monozygotic twin pairs.
- The number of CNVs identified depends mainly on the settings of the scoring algorithms; in the size range of 0.3-1.2 Mb we detect 1-2 per pair.
- CNVs are not present in 100% of the cells. This suggests somatic mosaicism, i.e. a post-meiotic emergence.

OMICS A Journal of Integrative Biology
Volume 12, Number 1, 2008

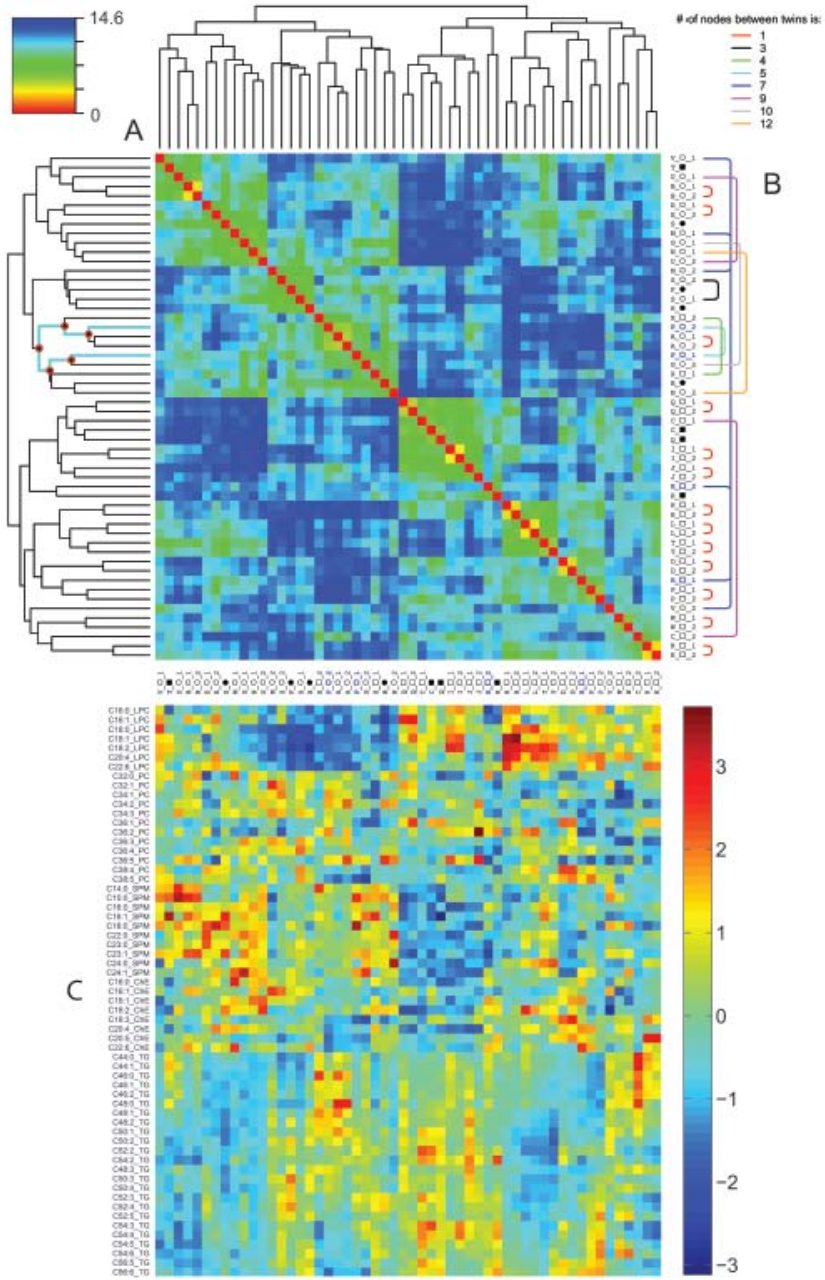
DOI: 10.1089/omi.2007.0048

Similarities and Differences in Lipidomics Profiles among Healthy Monozygotic Twin Pairs

HARMEN H.M. DRAISMA,[1] THEO H. REIJMERS,[1]
IVANA BOBELDIJK-PASTOROVA,[2] JACQUELINE J. MEULMAN,[3]
G. FREDERIEK ESTOURGIE-VAN BURK,[4,5] MEIKE BARTELS,[5]
RAYMOND RAMAKER,[2] JAN VAN DER GREEF,[1] DORRET I. BOOMSMA,[5]
and THOMAS HANKEMEIER[1]

metabolomes. Here we present the results of hierarchical clustering of blood plasma lipid profile data obtained by liquid chromatography–mass spectrometry from 23 healthy, 18-year-old twin pairs, of which 21 pairs were monozygotic, and 8 of their siblings. For 13

Metabolomic data characterized by large number of dependent variables



Euclidean distances among objects and corresponding dendrogram (A); scaled data for each participant (C). In Panel B co-twins are connected by colored lines. In the dendrogram of Panel A an example is drawn of our approach to characterize co-clustering of twins. The keys to Panels A, B and C are given in the upper left, upper right, and lower right corners of the figure. In Panel C lipids are labeled by their class abbreviation (LPC, PC,…) followed by the number of carbon atoms and the number of double bonds (separated by a colon) in the fatty acid.

Boulder 2008

- Dorret Boomsma, NL
- Stacey Cherny, Hong Kong
- Danielle Dick, USA
- David Evans, UK
- Manuel Ferreira, USA
- Nathan Gillespie, USA
- John Hewitt, USA
- Matthew Keller, USA
- Jeff Lessem, USA
- Gitta Lubke, USA
- Hermine Maes, USA
- Nick Martin, OZ
- Sarah Medland, USA
- Katherine Morley, OZ
- Benjamin Neale, UK
- Michael Neale, USA
- Irene Rebollo, NL
- Fruhling Rijsdijk, UK
- William Valdar, UK