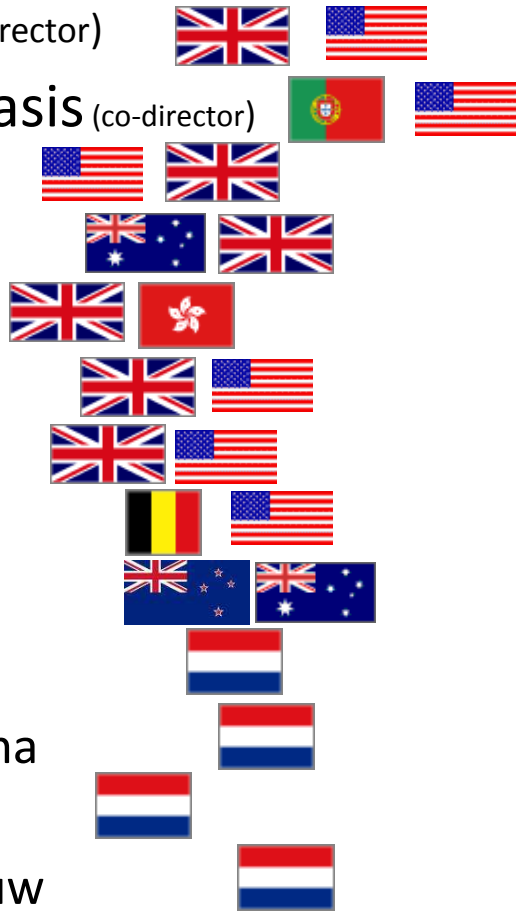
26th International Workshop on Methodology for Human Genomic Studies: “the Advanced course”

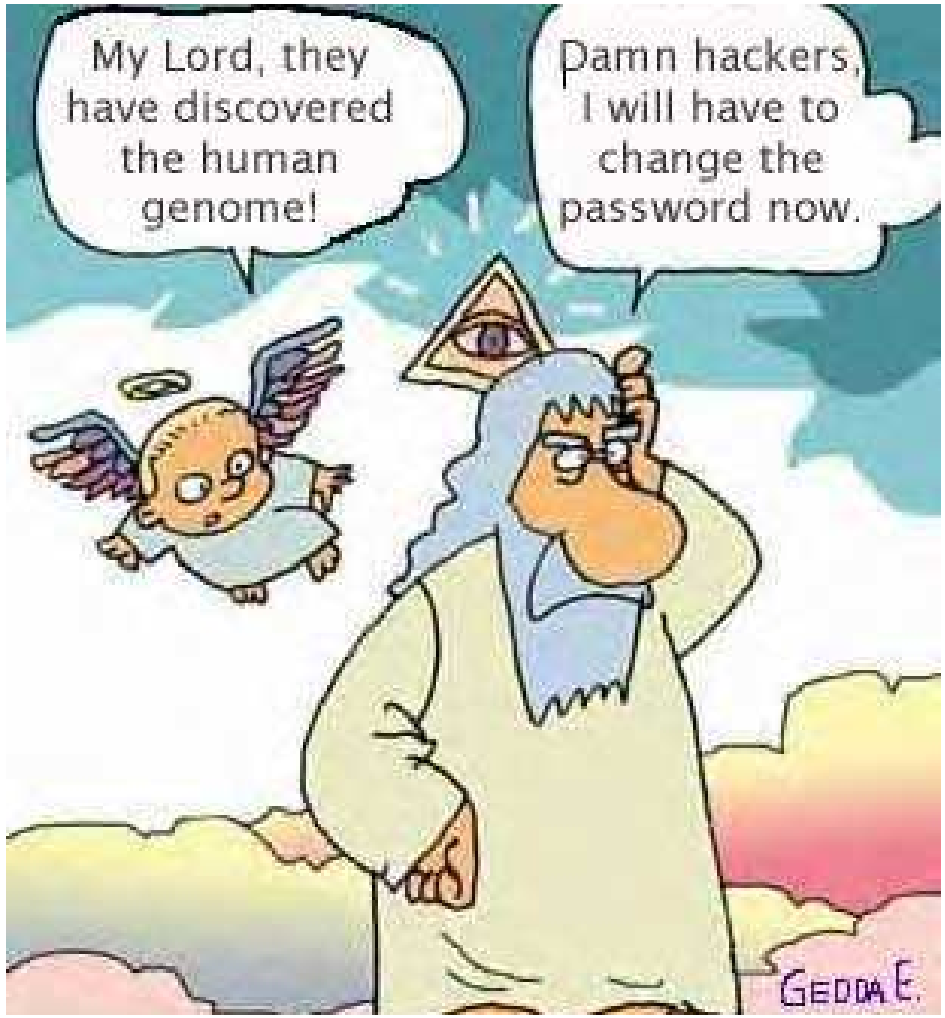
- Ben Neale (co-director)
- Goncalo Abecasis (co-director)
- Jeff Barrett
- David Evans
- Pak Sham
- Lindon Eaves
- Mike Neale
- Hermine Maes
- Sarah Medland
- Dorret Boomsma
- Danielle Posthuma
- Meike Bartels
- Christian de Leeuw



- John Hewitt (host)
- Jeff Lessem
- Matt Keller
- Clara Tang + Emily Wong
- Stacey Cherny
- Miaoxin Li
- Shaun Purcell
- Manuel Ferreira
- Nick Martin
- Abdel Abdellaoui
- Teresa de Candia
- Merry Kate Wing
- Goo Jun

The genetics of complex traits:

historical context and current challenges



- Nick Martin

Queensland Institute of Medical Research, Brisbane

- Dorret Boomsma

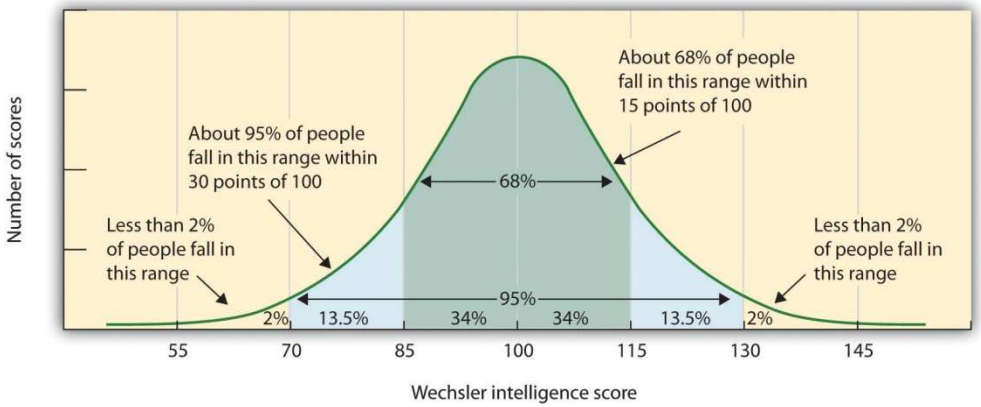
Dept Biological Psychology, VU Univ, Amsterdam

-

Boulder workshop, 2013

Human variation: Height

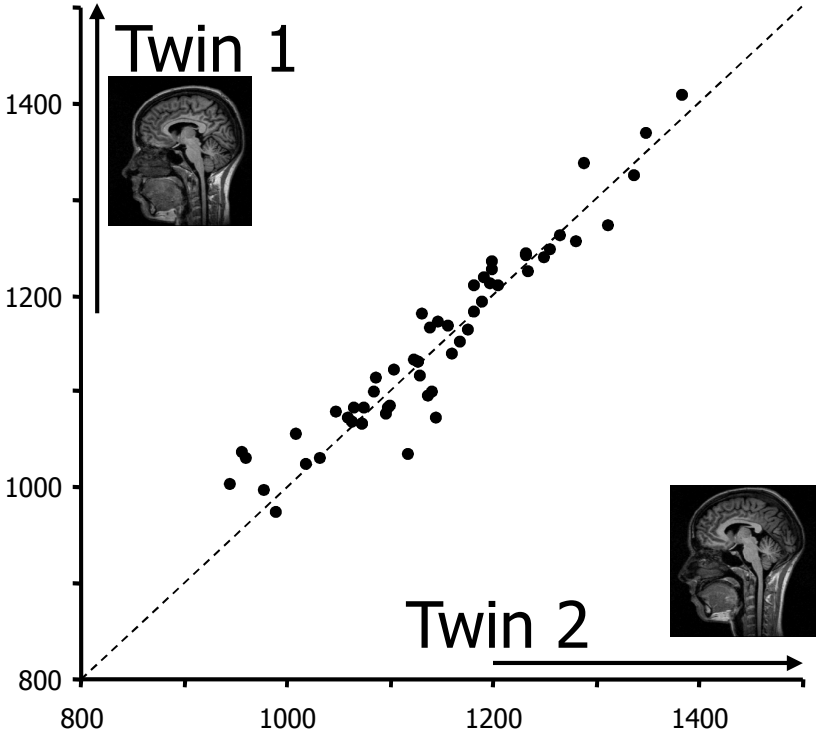
Human variation: IQ



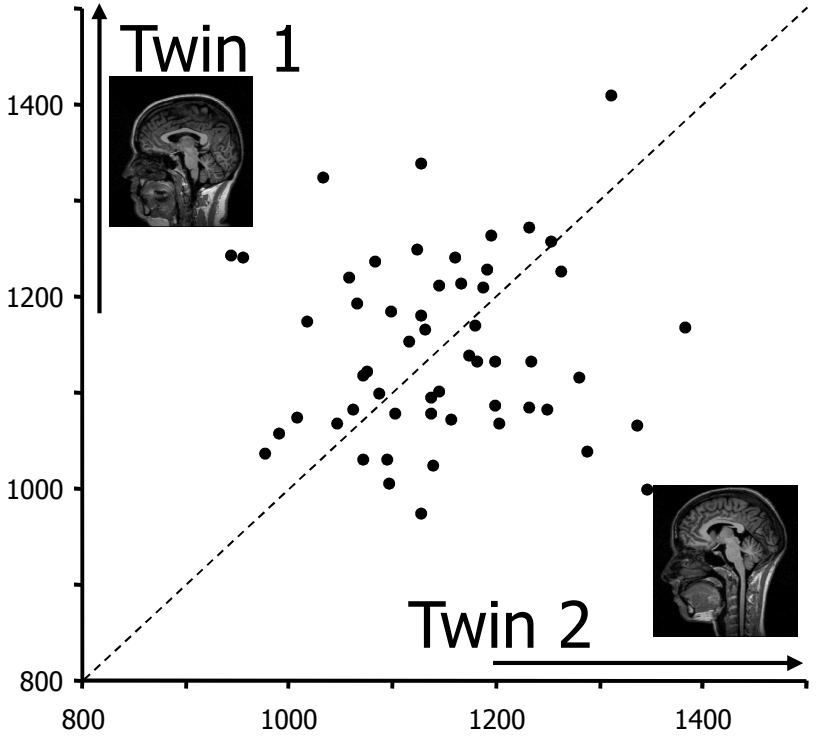
Genetic Epidemiology: Stages of Genetic Mapping

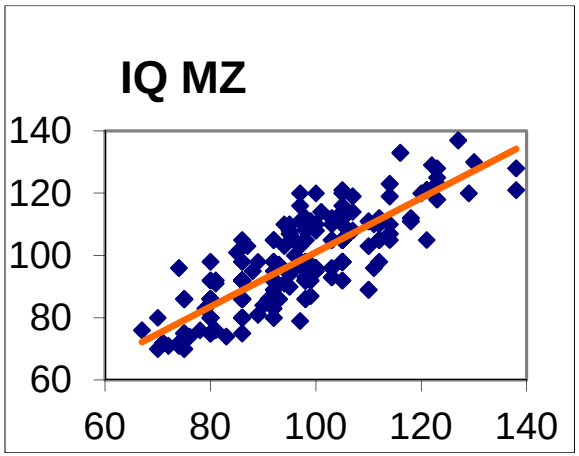
- Are there genes influencing this trait?
  - Genetic epidemiological (twin / family) studies OR heritability based on measured genetic variants
- Where are those genes?
  - Linkage analysis
- What are those genes?
  - Association analysis (meta-analysis / pathway)
- How do they work beyond the sequence?
  - Epigenetics, transcriptomics, proteomics
- What can we do with them ?
  - Translational medicine

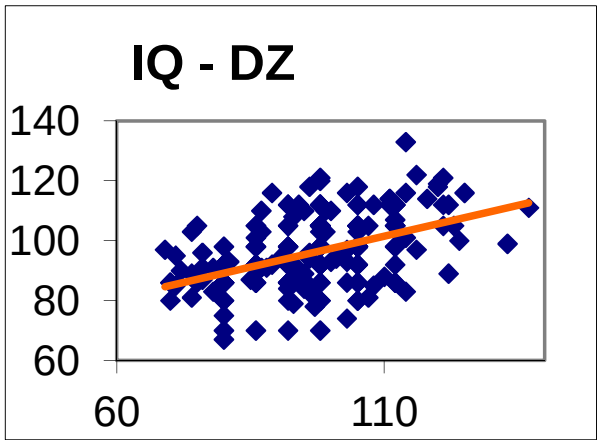
Brain volume MZ twin pairs (milliliter) in twin and co-twin



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

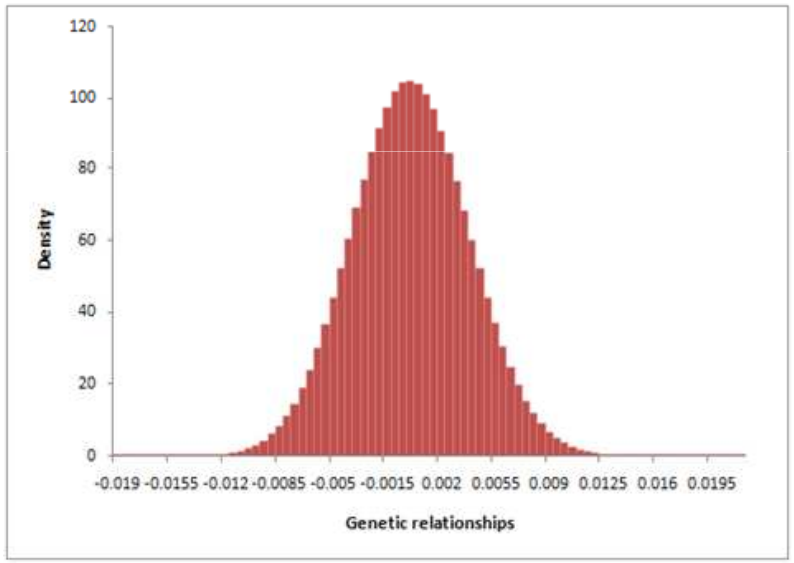


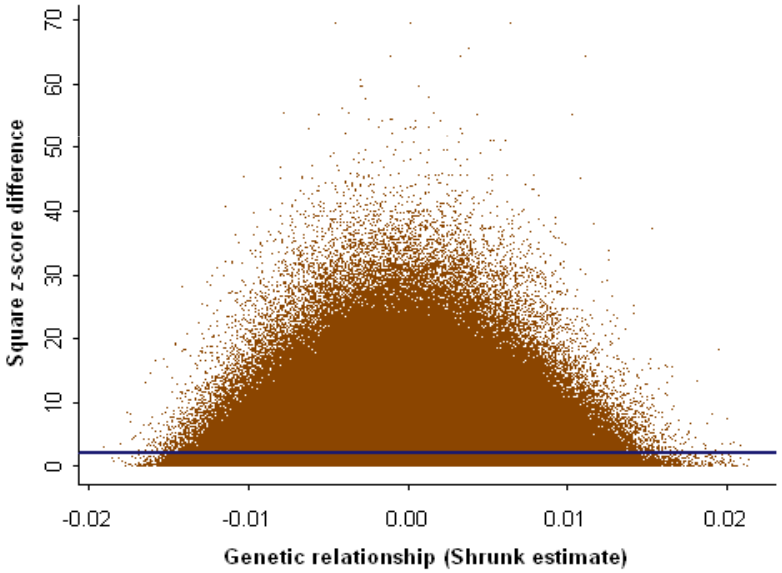




Estimating heritability from ‘unrelated’ individuals

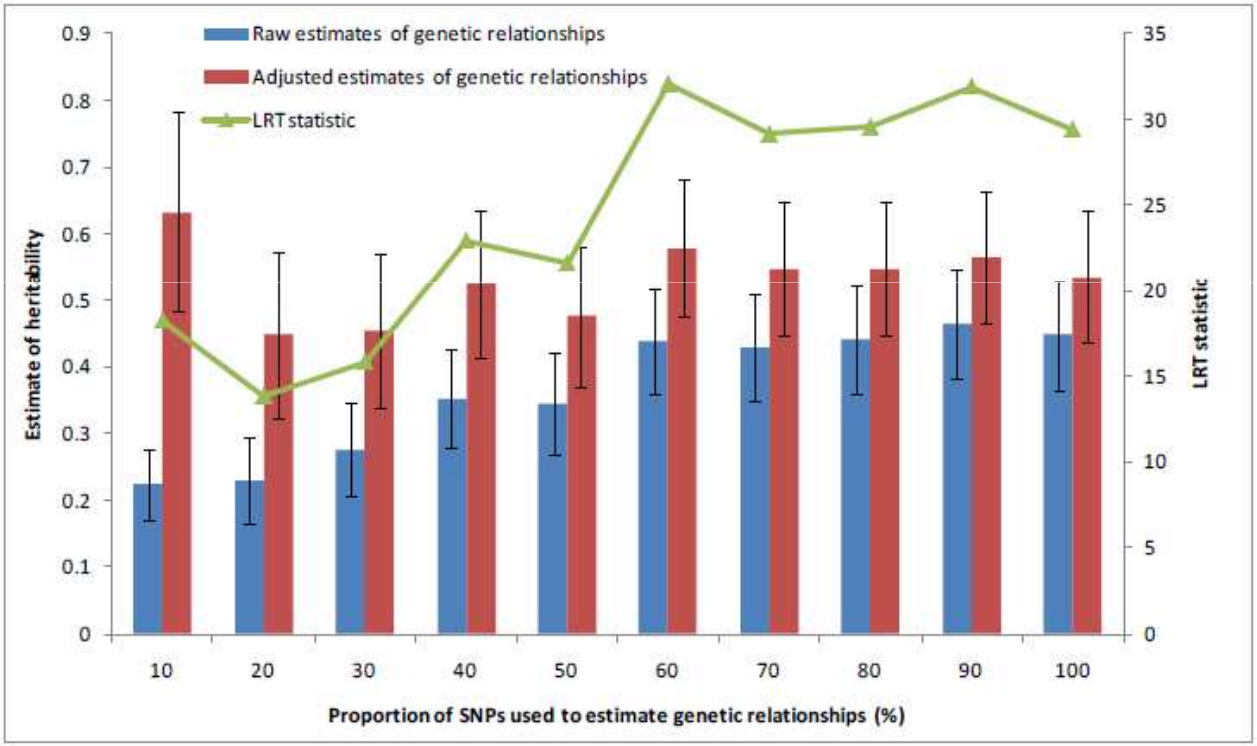
Very distant relatives that share more of their genome by descent are phenotypically more similar than those that share less





Yang et al. Nature Genetics 2010

Proportion of variance in height tagged by SNPs ~ 0.55 (SE 0.1)



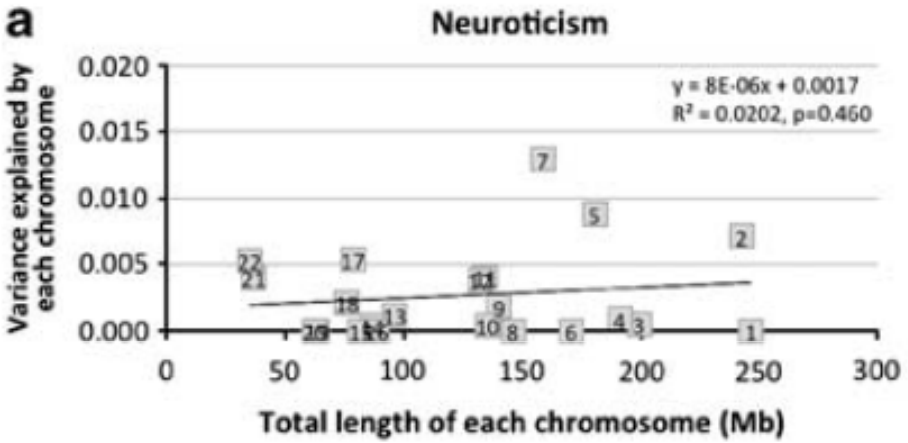
[Yang et al. Nature Genetics 2010]

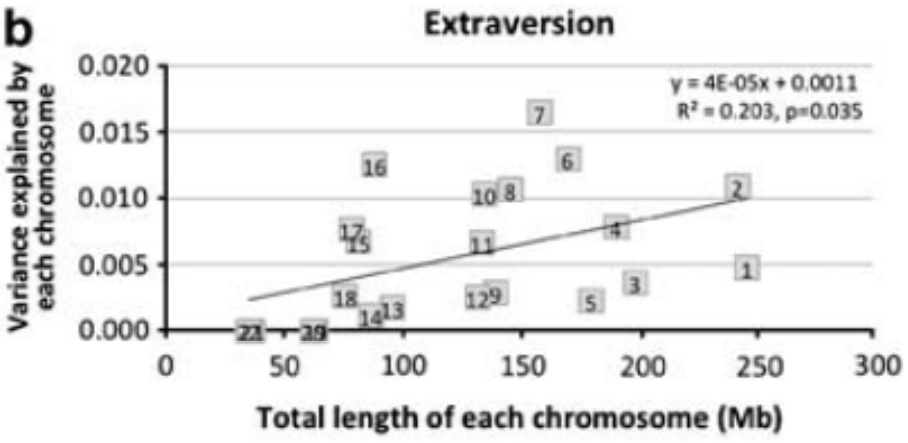
Genome-wide Complex Trait Analysis:
Heritability on measured SNPs

	Major Depression	Smoking Initiation	Current Smoking	Fasting Glucose	Height
Number of Ss	N=3245 N_{case}=1620 $N_{control}$=1625	N=4181 N_{case}=2602 $N_{control}$=1579	N=4181 N_{case}=1189 $N_{control}$=2992	N=3723	N=4199
Method Yang et al.	**.32 (.086)** p= 1.071×10^{-4}	**.19 (.087)** p= 0.024	**.24 (.096)** p= 0.011	**.22 (.059)** p=5.41×10^{-5}	**.42 (.052)** p=0
Method So et al.	**.28 (.058)**	**.28 (.084)**	**.44 (.063)**	**.19 (.036)**	**.29 (.035)**
Heritability twin studies	**.36**	**.44**	**.79**	**.53**	**.90**

Lubke et al. Biological Psychiatry, 2012

Neuroticism - Extraversion heritability by chromosome





Citation: Transl Psychiatry (2012) 2, e102, doi:10.1038/tp.2012.27

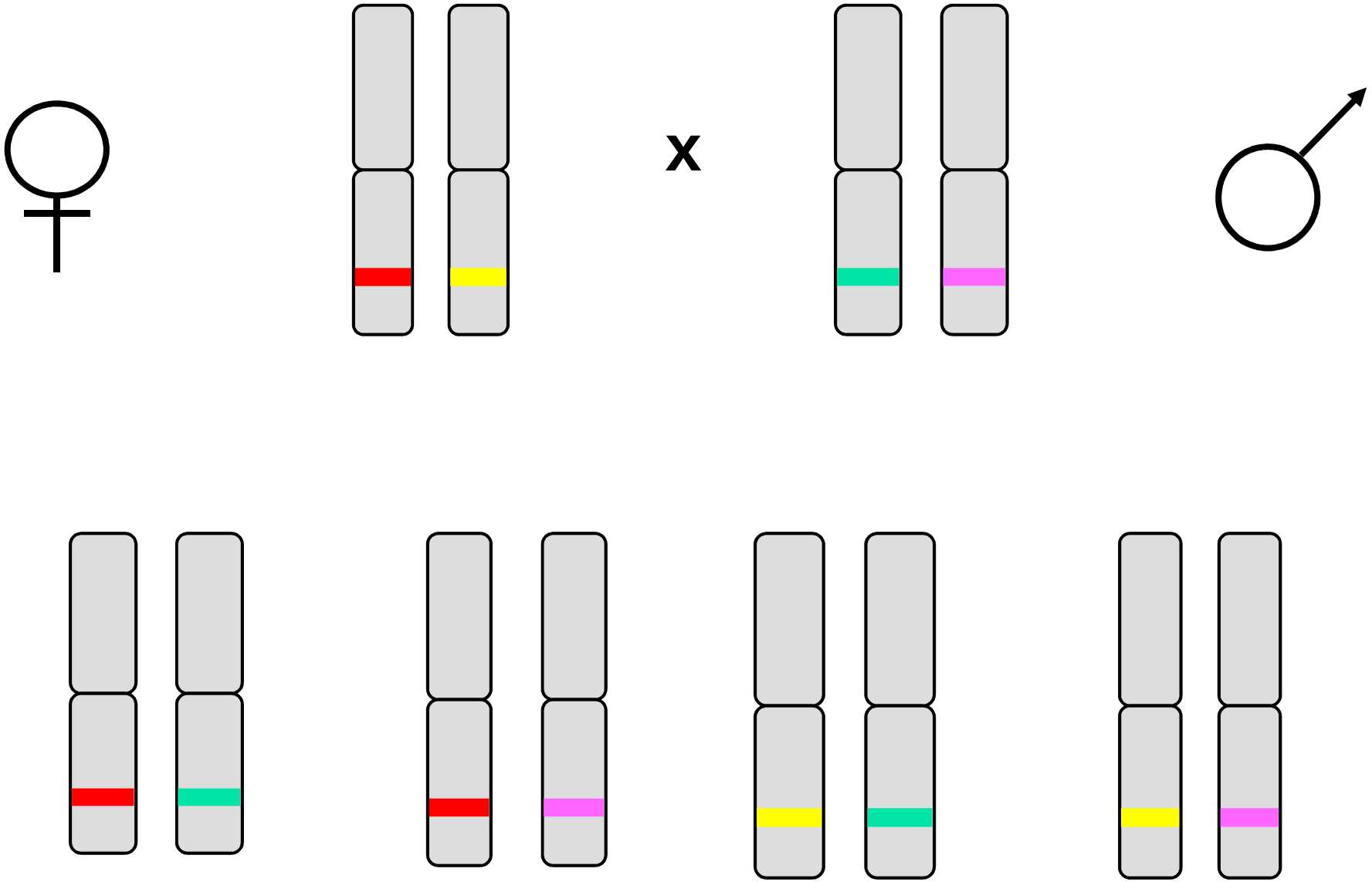
www.nature.com/tp

Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion

AAE Vinkhuyzen[1,2], NL Pedersen[3], J Yang[1], SH Lee[1,2], PKE Magnusson[3], WG Iacono[4], M McGue[4], PAF Madden[5], AC Heath[5], M Luciano[6], A Payton[7], M Horan[8], W Ollier[7], N Pendleton[8], IJ Deary[6], GW Montgomery[1], NG Martin[1], PM Visscher[1] and NR Wray[1,2]

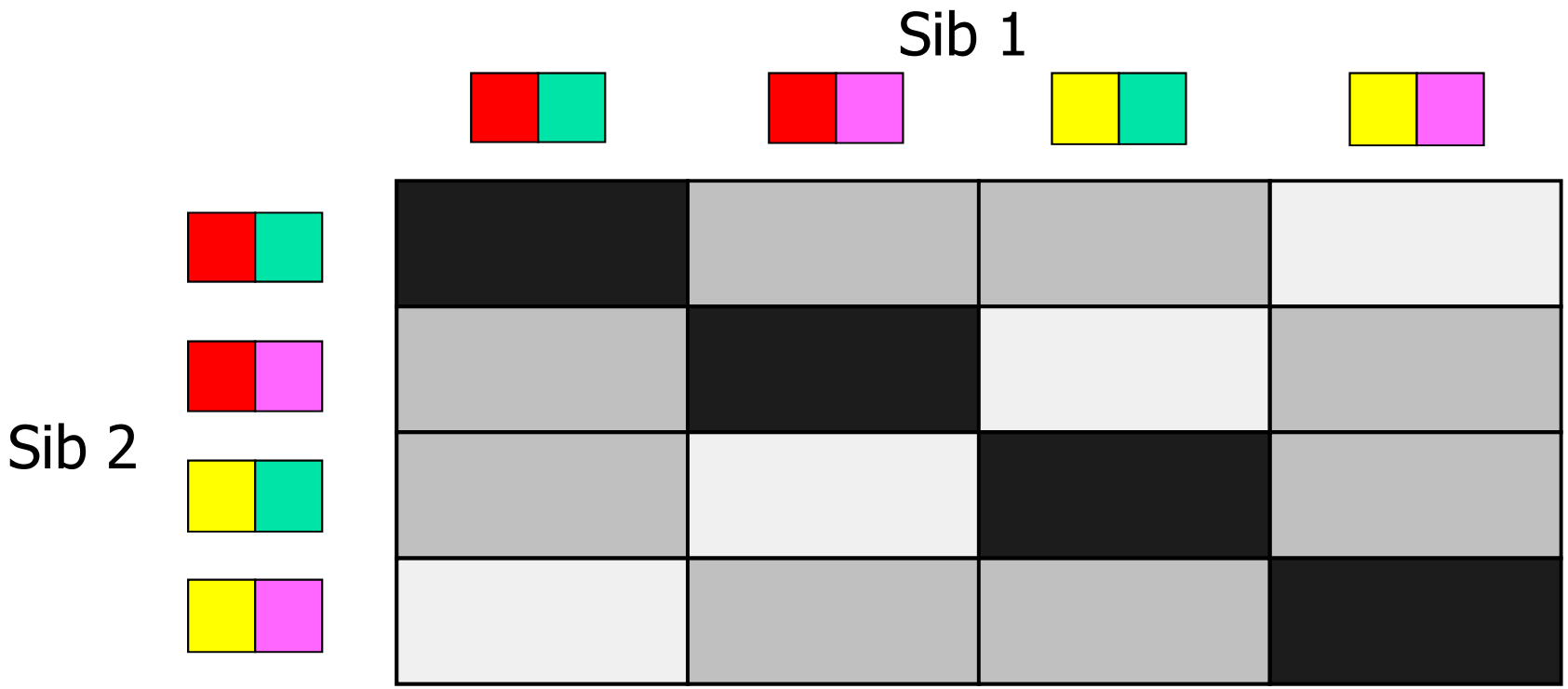
4 Stages of Genetic Mapping

- Are there genes influencing this trait?
  - Genetic epidemiological studies
- Where are those genes?
  - Linkage analysis
- What are those genes?
  - Association analysis
- What can we do with them ?
  - Translational medicine



IBD = IDENTITY BY DESCENT: does IBD sharing correspond with phenotype sharing?

IDENTITY BY DESCENT

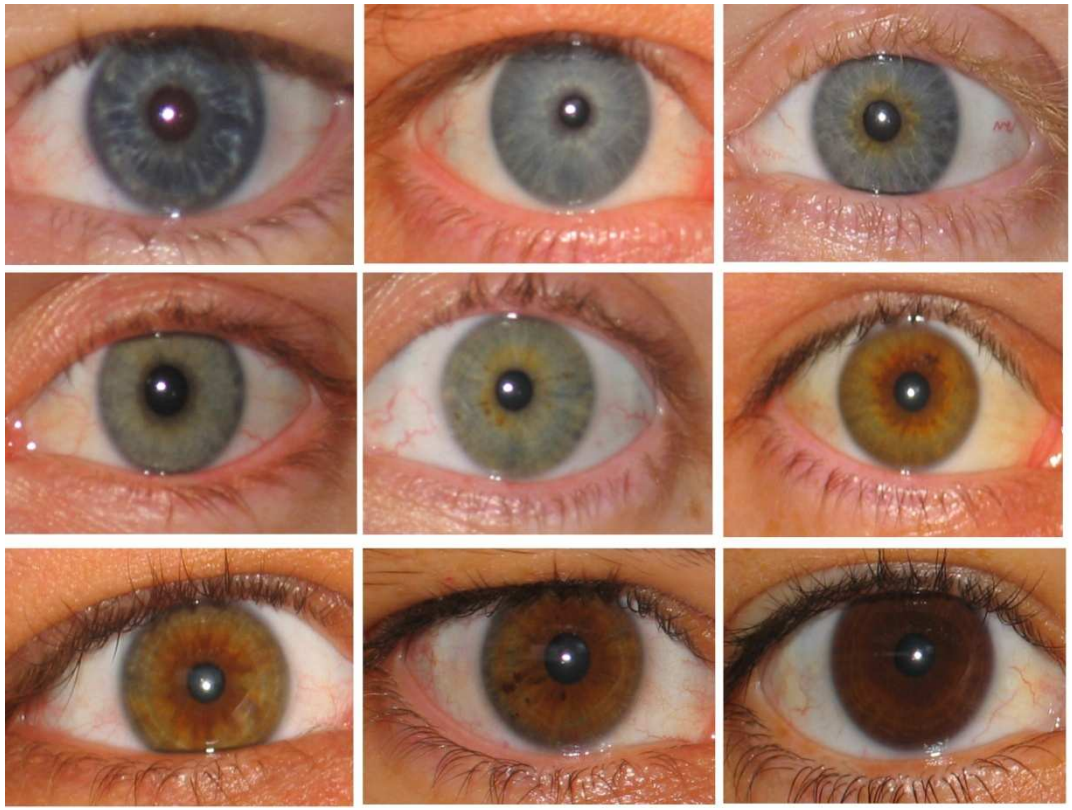


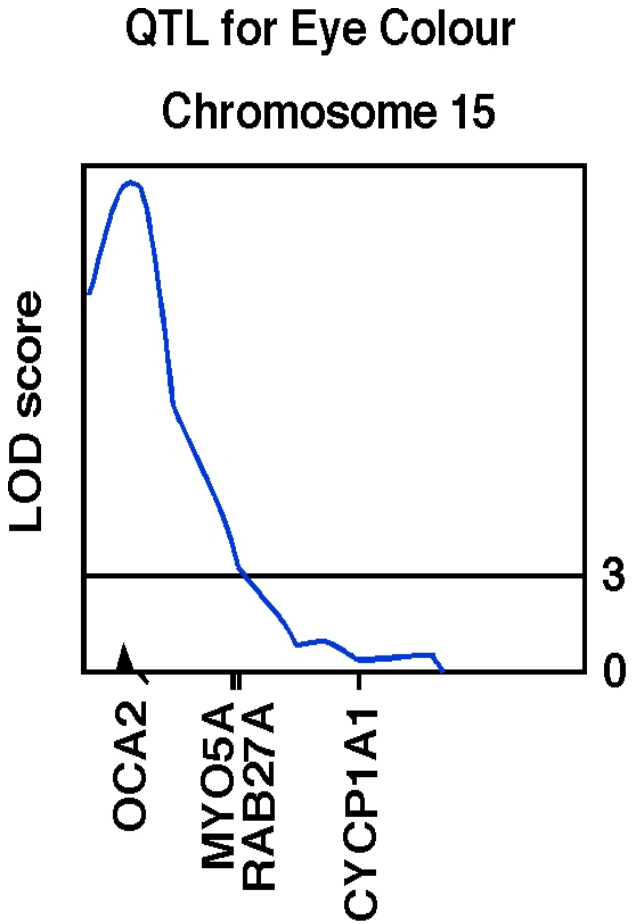
4/16 = 1/4 sibs share BOTH parental alleles IBD = 2

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

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

Example: Human OCA2 and eye colour





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

Finding the genes – association

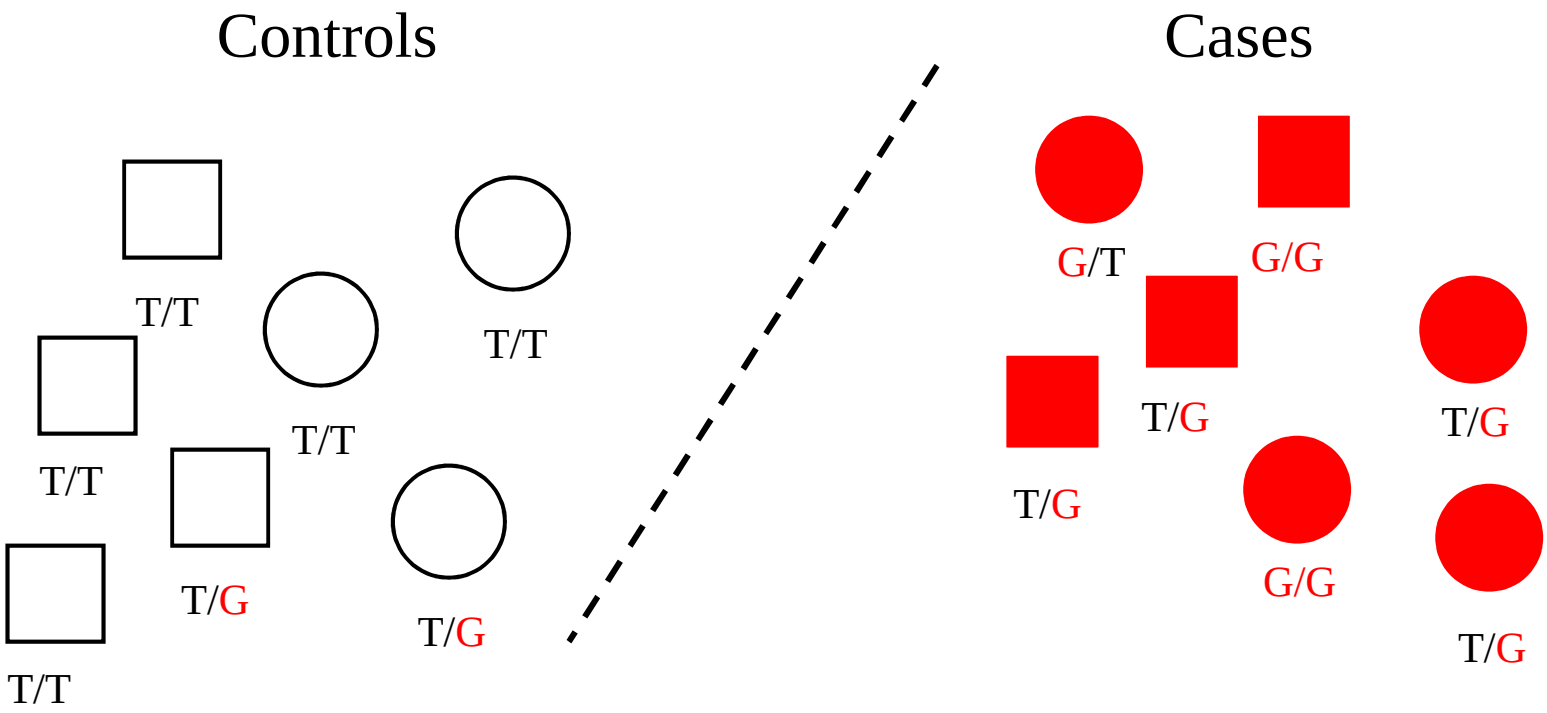
Looks for correlation between specific alleles and phenotype (trait value, disease risk)

How do we test for association?

*with measured (“tag”) SNPs that are correlated (in linkage disequilibrium) with causal variants

*with imputed genotype data

Genetic Case Control Study



Allele G is ‘associated’ with disease

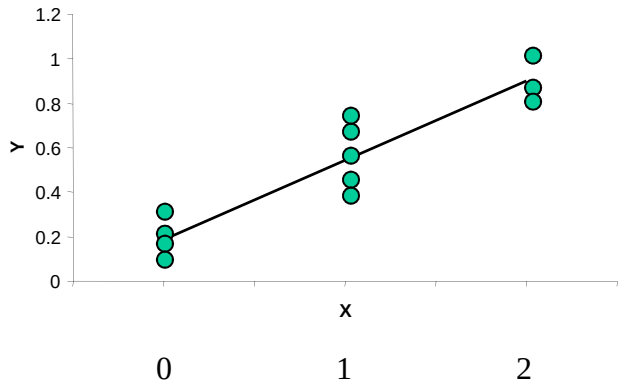
Simple Regression Model of Association (continuous trait)

$$Y_i = \alpha + \beta X_i + e_i$$

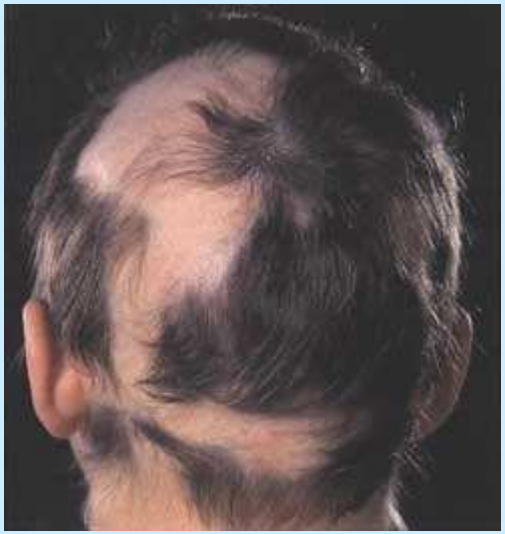
where

Y_i = trait value for individual i

X_i = number of 'MAF' alleles an individual has



Association test is whether $\beta > 0$



Genome-wide association study in alopecia areata implicates both innate and adaptive immunity

Lynn Petukhova[1], Madeleine Duvic[2], Maria Hordinsky[3], David Norris[4], Vera Price[5], Yutaka Shimomura[1], Hyunmi Kim[1], Pallavi Singh[1], Annette Lee[6], Wei V. Chen[7], Katja C. Meyer[8], Ralf Paus[8,9], Colin A. B. Jahoda[10], Christopher I. Amos[7], Peter K. Gregersen[6] & Angela M. Christiano[1,11]

NATURE | Vol 466 | 1 July 2010

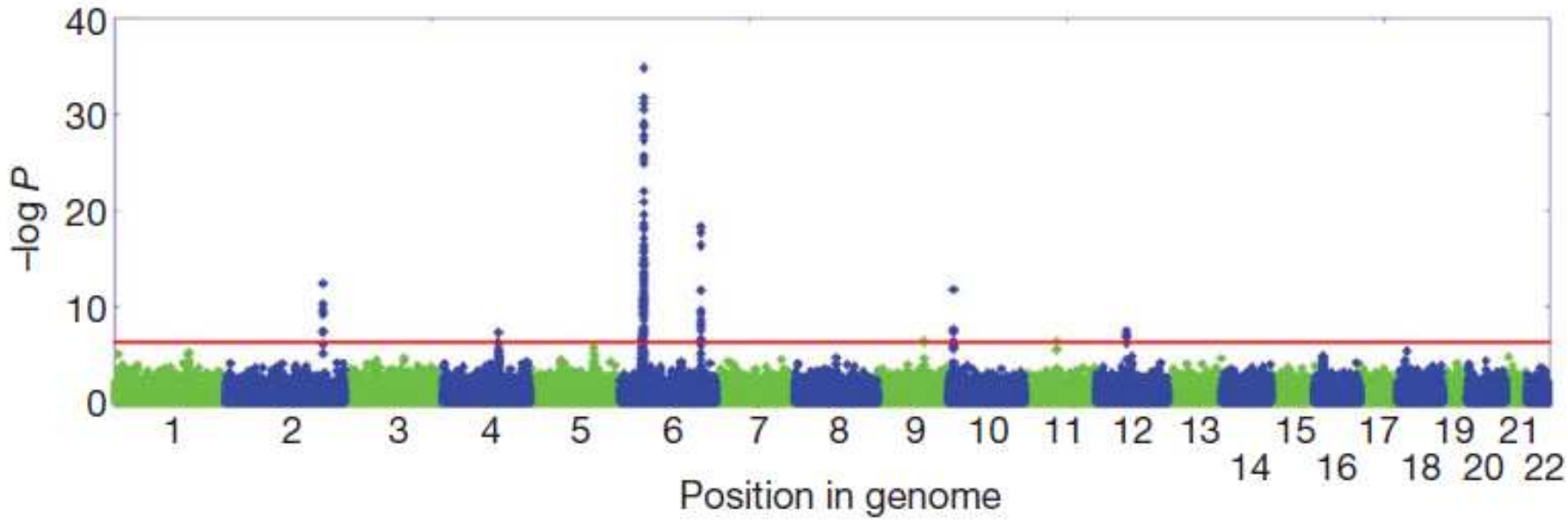
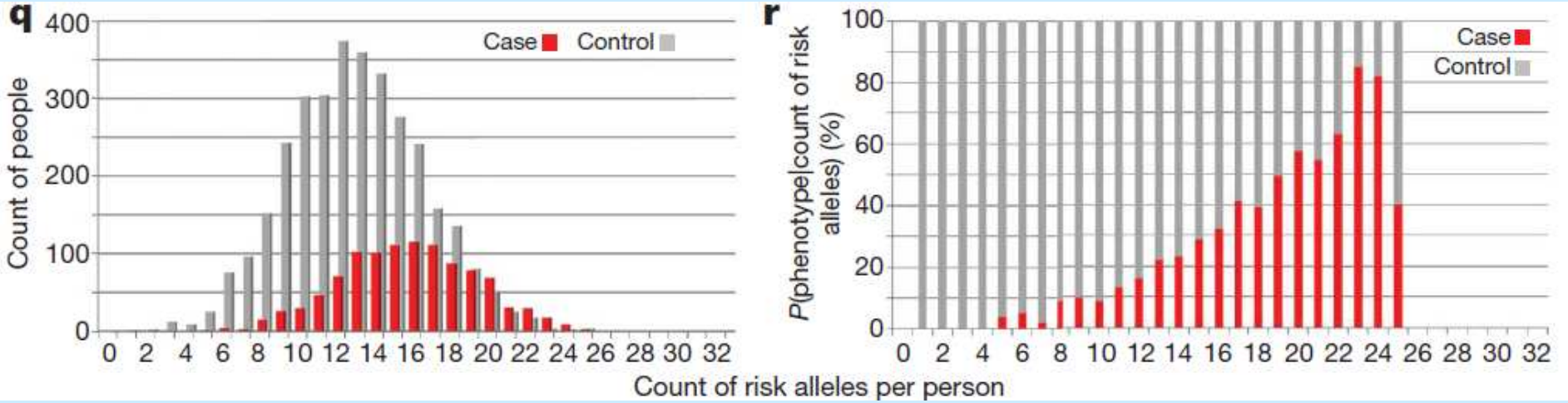


Table 1 | Genes with significant association to AA

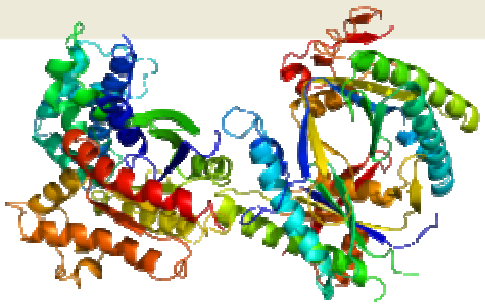
Region	Gene	Function	Strongest association (*P* value)	Maximum odds ratio	Involved in other autoimmune disease
2q33.2	*CTLA4*	Co-stimulatory family	3.55×10^{-13}	1.44	T1D, RA, CeD, MS, SLE, GD
	ICOS	Co-stimulatory family	4.33×10^{-8}	1.32	
4q27	*IL-21/IL-2*	T-, B- and NK-cell proliferation	4.27×10^{-8}	1.34	T1D, RA, CeD, PS
6q25.1	*ULBP6*	NKG2D activating ligand	4.49×10^{-19}	1.65	None
	ULBP3	NKG2D activating ligand	4.43×10^{-17}	1.52	None
9q31.1	*STX17*	Premature hair greying	3.60×10^{-7}	1.33	None
10p15.1	*IL-2RA*	T-cell proliferation	1.74×10^{-12}	1.41	T1D, MS, GD, GV
11q13	*PRDX5*	Antioxidant enzyme	4.14×10^{-7}	1.33	MS
12q13	*Eos* (*IKZF4*)	T_{reg} transcription factor	3.21×10^{-8}	1.34	T1D, SLE
	ERBB3	Epidermal growth factor receptor	1.27×10^{-7}	1.34	T1D, SLE
6p21.32	*MICA*	NKG2D activating ligand	1.19×10^{-7}	1.44	T1D, RA, CeD, UC, PS, SLE
(HLA)	*NOTCH4*	Haematopoietic differentiation	1.03×10^{-8}	1.61	T1D, RA, MS
	C6orf10	Unknown	1.45×10^{-16}	2.36	T1D, RA, PS, GV
	BTNL2	Co-stimulatory family	2.11×10^{-26}	2.70	T1D, RA, UC, CD, SLE, MS, GV
	HLA-DRA	Antigen presentation	2.93×10^{-31}	2.62	T1D, RA, CeD, MS, GV
	HLA-DQA1	Antigen presentation	3.60×10^{-17}	2.15	T1D, RA, CeD, MS, SLE, PS, CD, UC, GD
	HLA-DQA2	Antigen presentation	1.38×10^{-35}	5.43	T1D, RA
	HLA-DQB2	Antigen presentation	1.73×10^{-13}	1.60	RA

Each of the eight regions implicated in our study contains multiple significant SNPs, which are detailed in Supplementary Tables 1 and 2. Here we display candidate genes within the implicated regions, and include the *P* value of the most significant SNP, and the odds ratio for the SNP with the largest effect estimate. Diseases are listed for which a GWAS or previous candidate gene study identified the same region (http://www.genome.gov/gwastudies, http://www.cdc.gov/genomics/hugenet): Crohn's disease (CD), celiac disease (CeD), Graves disease (GD), generalized vitiligo (GV), multiple sclerosis (MS), psoriasis (PS), rheumatoid arthritis (RA), system lupus erythematosus (SLE), type I diabetes (T1D), and ulcerative colitis (UC).



LETTERS

nature genetics



Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight

ARTICLES

nature genetics

New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk

Levels of circulating glucose are tightly regulated. To identify new loci influencing glycemic traits, we performed meta-analyses of 21 genome-wide association studies informative for fasting glucose, fasting insulin and indices of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) in up to 46,186 nondiabetic participants. Follow-up of 25 loci in up to 76,558 additional subjects identified 16 loci associated with fasting glucose and HOMA-B and two loci associated with fasting insulin and HOMA-IR. These include nine loci newly associated with fasting glucose (in or near *ADCY5*, *MADD*, *ADRA2A*, *CRY2*, *FADS1*, *GLIS3*, *SLC2A2*, *PROX1* and *C2CD4B*) and one influencing fasting insulin and HOMA-IR (near *IGF1*). We also demonstrated association of *ADCY5*, *PROX1*, *GCK*, *GCKR* and *DGKB-TMEM195* with type 2 diabetes. Within these loci, likely biological candidate genes

GWAS studies: What did they find? How much variance have they explained?

REVIEW

Five Years of GWAS Discovery

Peter M. Visscher,[1,2,*] Matthew A. Brown,[1] Mark I. McCarthy,[3,4] and Jian Yang[5]

The past five years have seen many scientific and biological discoveries made through the experimental design of genome-wide association studies (GWASs). These studies were aimed at detecting variants at genomic loci that are associated with complex traits in the population and, in particular, at detecting associations between common single-nucleotide polymorphisms (SNPs) and common diseases such as heart disease, diabetes, auto-immune diseases, and psychiatric disorders. We start by giving a number of quotes from scientists and journalists about perceived problems with GWASs. We will then briefly give the history of GWASs and focus on the discoveries made through this experimental design, what those discoveries tell us and do not tell us about the genetics and biology of complex traits, and what immediate utility has come out of these studies. Rather than giving an exhaustive review of all reported findings for all diseases and other complex traits, we focus on the results for auto-immune diseases and metabolic diseases. We return to the perceived failure or disappointment about GWASs in the concluding section.

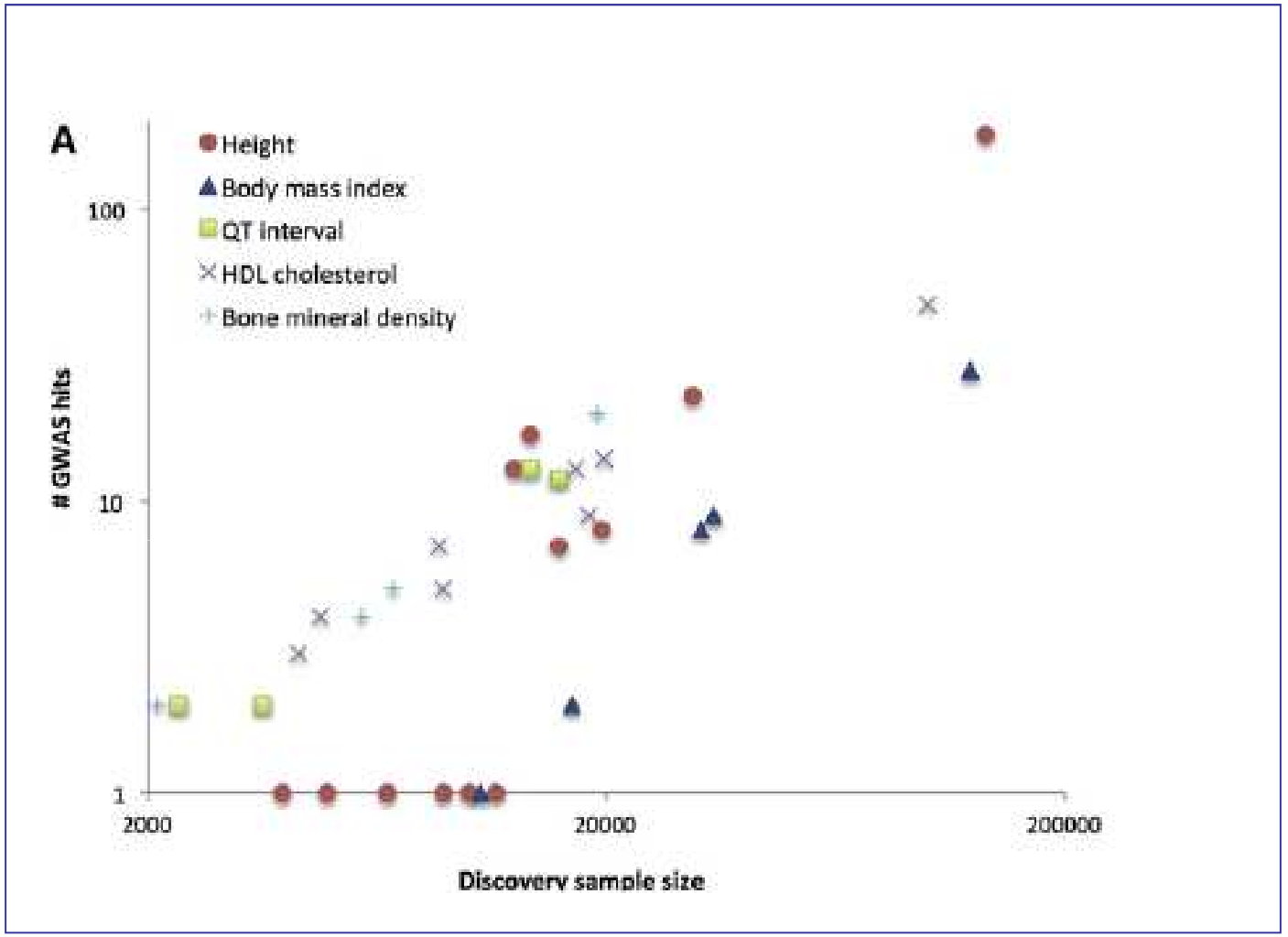
emerged from GWAS, which clearly isn't going to be the answer to everything."

From McCLellan and King, *Cell* 2010[1]:

"To date, genome-wide association studies (GWAS) have published hundreds of common variants whose allele frequencies are statistically correlated with various illnesses and traits. However, the vast majority of such variants have no established biological relevance to disease or clinical utility for prognosis or treatment."

"An odds ratio of 3.0, or even of 2.0 depending on population allele frequencies, would be robust to such population stratification. However, odds ratios of the magnitude generally detected by GWAS (<1.5) can frequently be explained by cryptic population stratification, regardless of the p value associ-

AJHG 2012


A
Height
Body mass index
QT interval
HDL cholesterol
Bone mineral density
100
10
1
GWAS hits
2000
20000
200000
Discovery sample size

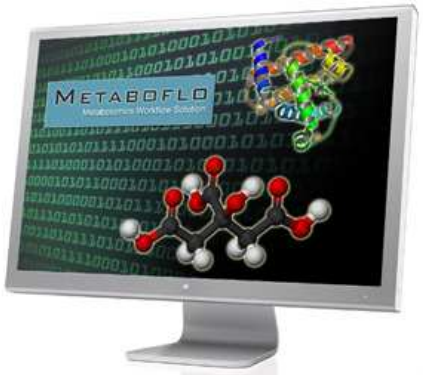
Some hits are ‘easy’: metabolomics

- Metabolites: small molecules (blood, urine)
- Metabolomics: comprehensive measurement

‘Classical’:
One or ‘bulk’

http://test1-img.ehowcdn.com/article-new/ehow/images/a04/ra/b1/normal-levels-ldl-hdl-800x800.jpg

Metabolomics:
MANY



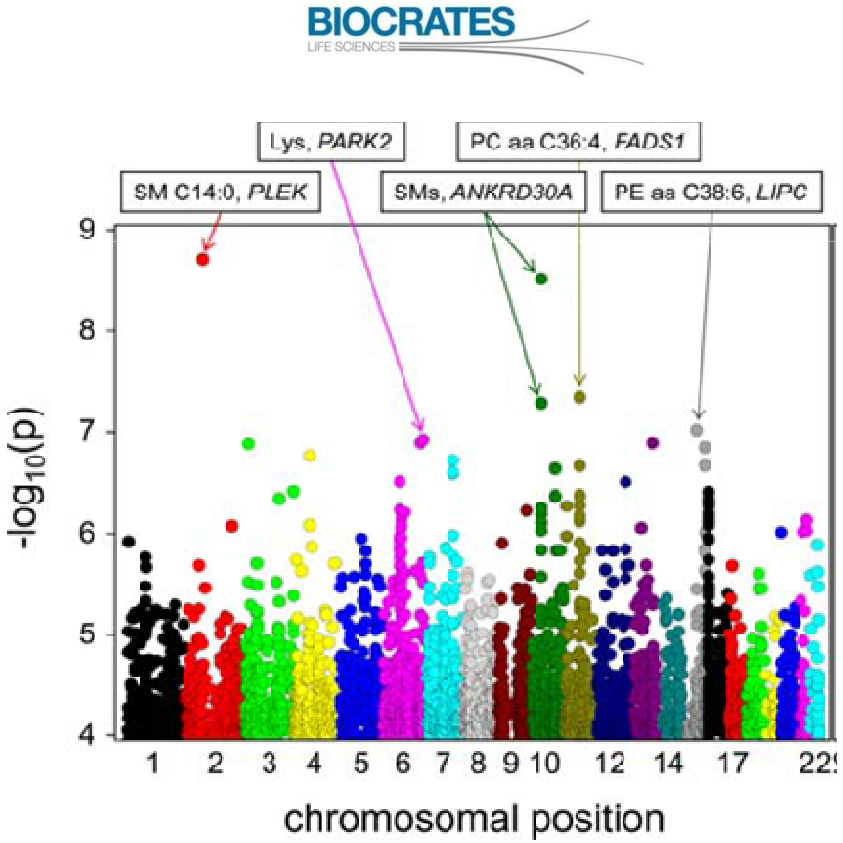
http://insiliflo.com/images/metaboflo/display.png?1349969067

- Techniques (>1970):
  - ^{1}H NMR – nuclear magnetic resonance spectroscopy
  - Mass spectrometry, e.g. LC – MS

API 4000 Triple Quadrupole mass spectrometer (Helmholtz Zentrum, Munich, Germany)

ENGAGE Metabolomics project

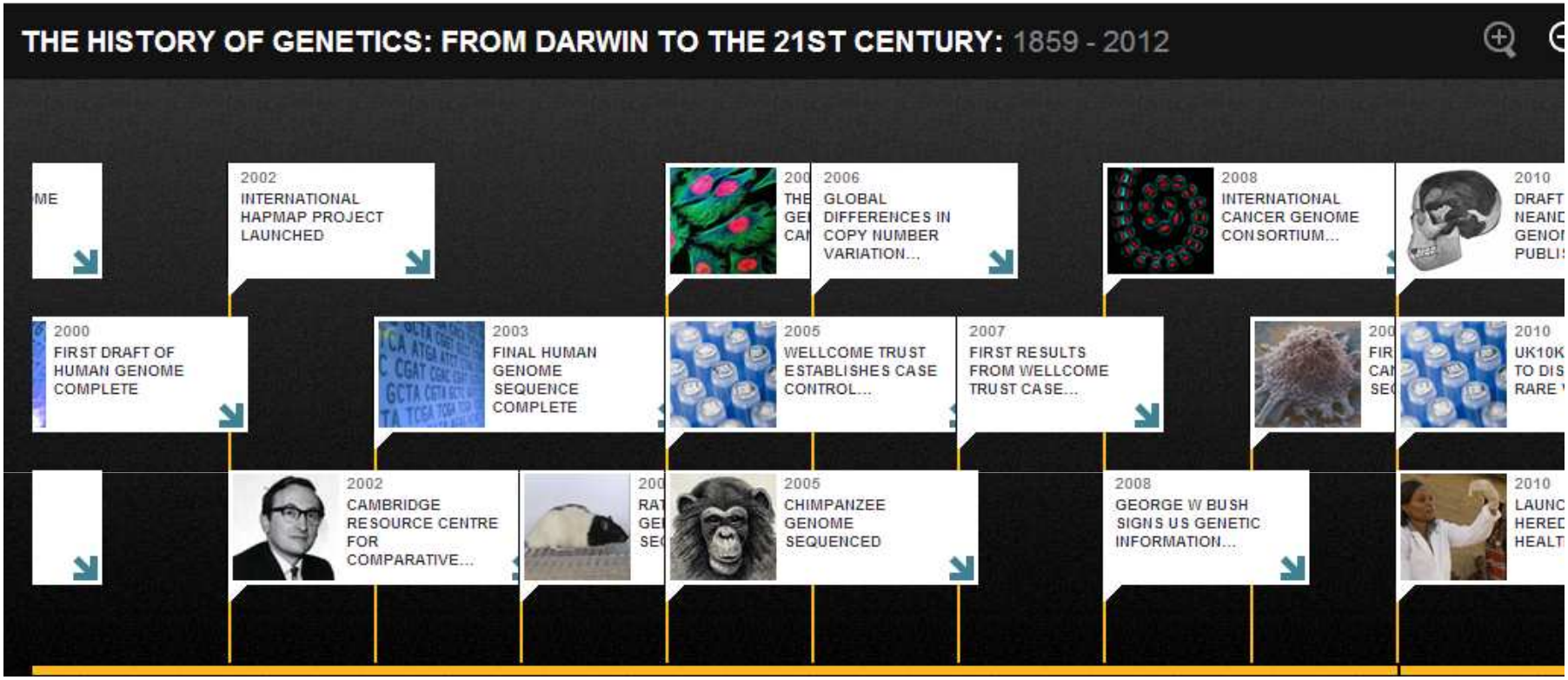
- ‘Follow up’ of Gieger – 2008 & Illig – 2010 GWA studies for fasting serum
- Metabolomics: Biocrates AbsoluteIDQ p150 kit
- Aims:
  - Larger sample sizes → enhanced GWA power
  - Combination with other “omics” data → understanding



Gieger *et al*, PLoS Genet 2008: 4(11): e1000282

The White House - June 26, 2000

Venter
Clinton
Collins


THE HISTORY OF GENETICS: FROM DARWIN TO THE 21ST CENTURY: 1859 - 2012
ME
2002
INTERNATIONAL HAPMAP PROJECT LAUNCHED
2006
GLOBAL DIFFERENCES IN COPY NUMBER VARIATION...
2008
INTERNATIONAL CANCER GENOME CONSORTIUM...
2010
2000
FIRST DRAFT OF HUMAN GENOME COMPLETE
2003
FINAL HUMAN GENOME SEQUENCE COMPLETE
2005
WELLCOME TRUST ESTABLISHES CASE CONTROL...
2007
FIRST RESULTS FROM WELLCOME TRUST CASE...
2010
2002
CAMBRIDGE RESOURCE CENTRE FOR COMPARATIVE...
2005
CHIMPANZEE GENOME SEQUENCED
2008
GEORGE W BUSH SIGNS US GENETIC INFORMATION...
2010

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on

This figure is purely

is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[3,4] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

22 December 2000

Science

Vol. 290 No. 5500
Pages 2201–2372 $8

SEQUENCED GENOMES

Breakthrough of the Year

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

A New Human Genome Sequence Paves the Way for Individualized Genomics

2007

Liza Gross | doi:10.1371/journal.pbio.0050266

Just six years ago, two draft versions of the human genome were published, an achievement widely hailed as one of the most audacious scientific undertakings in history. Both of these versions are composite sequences derived from the haploid genomes—the single set of 23 chromosomes packaged into the sperm or egg of each parent—of (mostly) anonymous donors. But now, one of the principals behind the private human genome initiative has

the draft human genome. Briefly, the shotgun sequencing approach randomly shreds genetic material into millions of fragments, called "reads," each of which is sequenced and then reassembled using a computer (based on sequence similarity), which matches up overlapping reads and merges them into longer sequences. By refining the software algorithms of the computer assembler (to respect the distinct paternal allelic contributions) and

Mary Shago[2], Timothy B. Stockwell[1], Alexia Tsiamouri[1], Vineet Bafna[3], Vikas Bansal[3], Saul A. Kravitz[1], Dana A. Busam[1], Karen Y. Beeson[1], Tina C. McIntosh[1], Karin A. Remington[1], Josep F. Abril[4], John Gill[1], Jon Borman[1], Yu-Hui Rogers[1], Marvin E. Frazier[1], Stephen W. Scherer[2], Robert L. Strausberg[1], J. Craig Venter[1]

1 J. Craig Venter Institute, Rockville, Maryland, United States of America, **2** Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada, **3** Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, United States of America, **4** Genetics Department, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

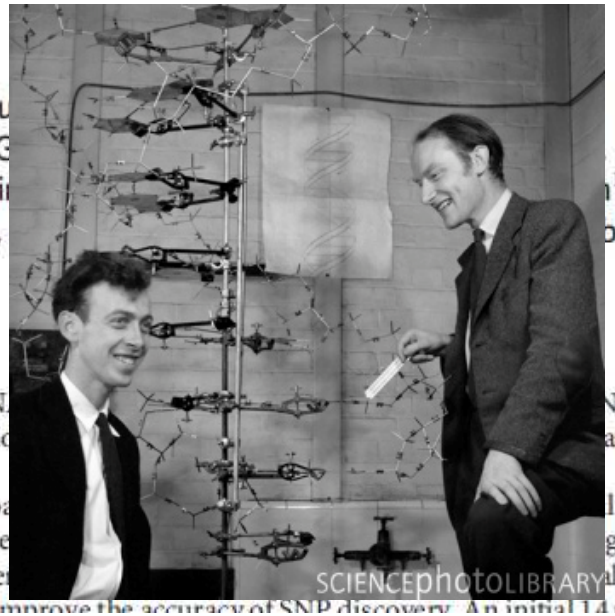
It took 4 months, a handful of scientists and less than US$1.5 mil to sequence the genome of James Watson

nature Vol 452|17 April 2008|doi:10.1038/nature06884

LETTERS

The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yu
Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier G
Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chi
Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2],
& Jonathan M. Rothberg[2]†

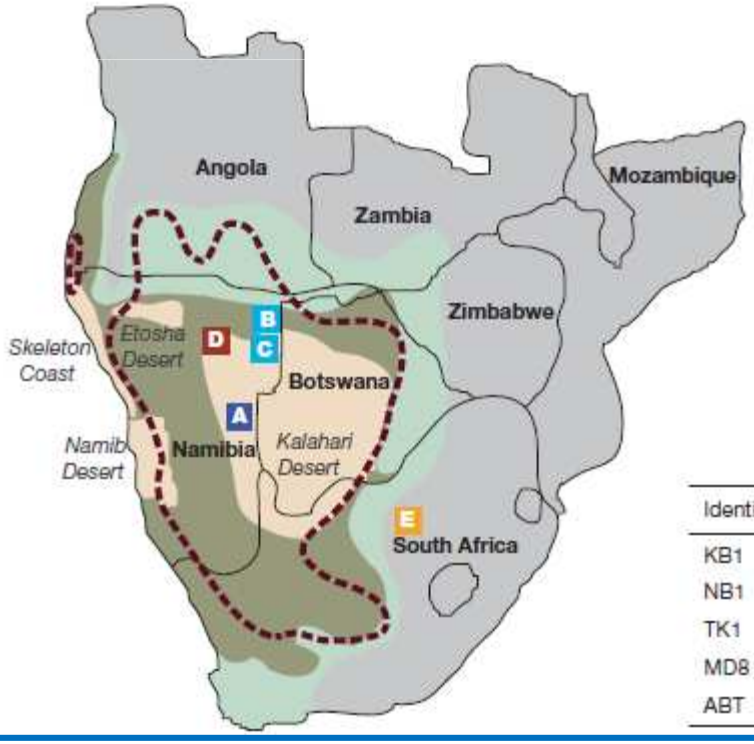


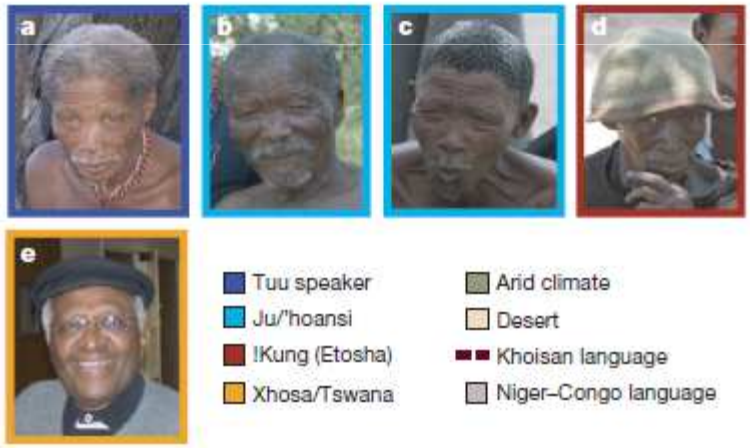
The association of genetic variation with disease and drug response, and improvements in nucleic acid technologies, have given great optimism for the impact of 'genomic medicine'. However, the formidable size of the diploid human genome[1], approximately 6 gigabases, has prevented the routine application of sequencing methods to deciphering complete individual human genomes. To realize the full potential of genomics for human health, this

Vol 463|18 February 2010|doi:10.1038/nature08795 nature

LETTERS

Complete Khoisan and Bantu genomes from southern Africa





Tuu speaker
Ju/'hoansi
!Kung (Etosha)
Xhosa/Tswana
Arid climate
Desert
Khoisan language
Niger–Congo language

Identifier	Name	Location of origin	Linguistic group	Y chromosome
KB1	!Gubi	Southern Kalahari	Tuu-speaker	B2b
NB1	G/aq'o	Northern Kalahari	Juu (Ju/'hoansi)	A3b1
TK1	D#kgao	Northern Kalahari	Juu (Ju/'hoansi)	A2
MD8	!Ai	Northern Kalahari	Juu (!Kung)	E1b1b1
ABT	Tutu	South Africa	Bantu (Xhosa/Tswana)	E1b1a8a *

A map of human genome variation from population-scale sequencing

The 1000 Genomes Project Consortium*

The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for investigating the relationship between genotype and phenotype. Here we present results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms. We undertook three projects: low-coverage whole-genome sequencing of 179 individuals from four populations; high-coverage sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. We describe the location, allele frequency and local haplotype structure of approximately 15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously undescribed. We show that, because we have catalogued the vast majority of common variation, over 95% of the currently accessible variants found in any individual are present in this data set. On average, each person is found to carry approximately 250 to 300 loss-of-function variants in annotated genes and 50 to 100 variants previously implicated in inherited disorders. We demonstrate how these results can be used to inform association and functional studies. From the two trios, we directly estimate the rate of ***de novo*** germline base substitution mutations to be approximately 10^{-8} per base pair per generation. We explore the data with regard to signatures of natural selection, and identify a marked reduction of genetic variation in the neighbourhood of genes, due to selection at linked sites. These methods and public data will support the next phase of human genetic research.

UK10K

Rare Genetic Variants in Health and Disease

Goals and Aims

Many hundreds of genes involved in the disease process have been identified by human genetics studies. However, the discovery of disease-causing variants has been hampered by the resolution of investigation available. This has meant that identified genes have been restricted to those with strong distinctive effects or those with a weak effect that have a

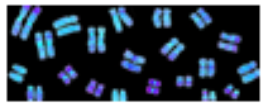
Genome

Aims

Through the genome-wide sequencing of deeply phenotyped cohorts, and exome (protein-coding regions) analysis of selected extreme phenotypes, the UK10K project aims to:

Elucidate singleton variants by maximising variation detected

By utilising pre-existing cohorts of related phenotypes, the UK10K project will explore the DNA sequence at an order of magnitude deeper than the **1000 Genomes Project** for Europe. By carrying out genome-wide sequencing of 4,000 samples from the **TwinsUK** and **ALSPAC** cohorts to 6x sequencing depth, researchers will maximise the amount of variation detected and detect singleton variants.



Adapted from PLoS, doi: 10.1371/ journal.pbio.0030157

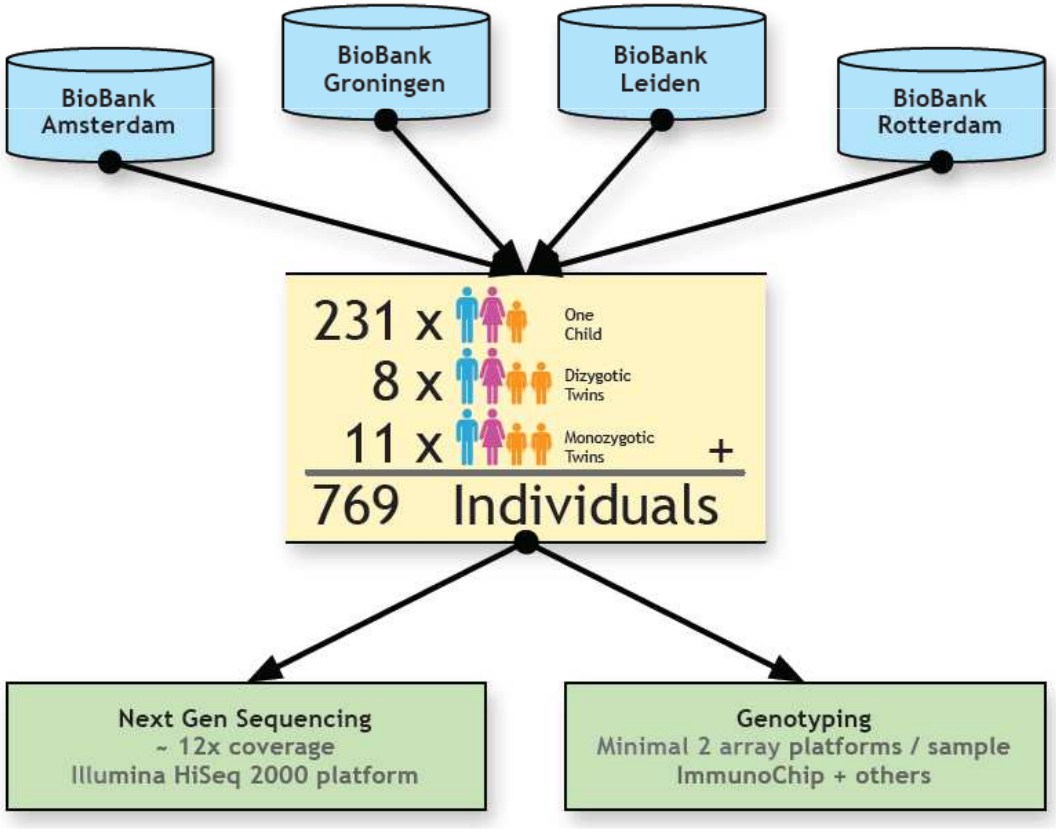
GoNL: Genome of the Netherlands

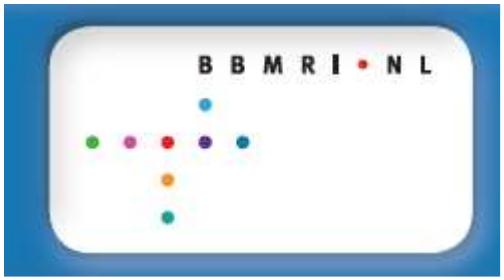
NL Twin Register

Lifelines

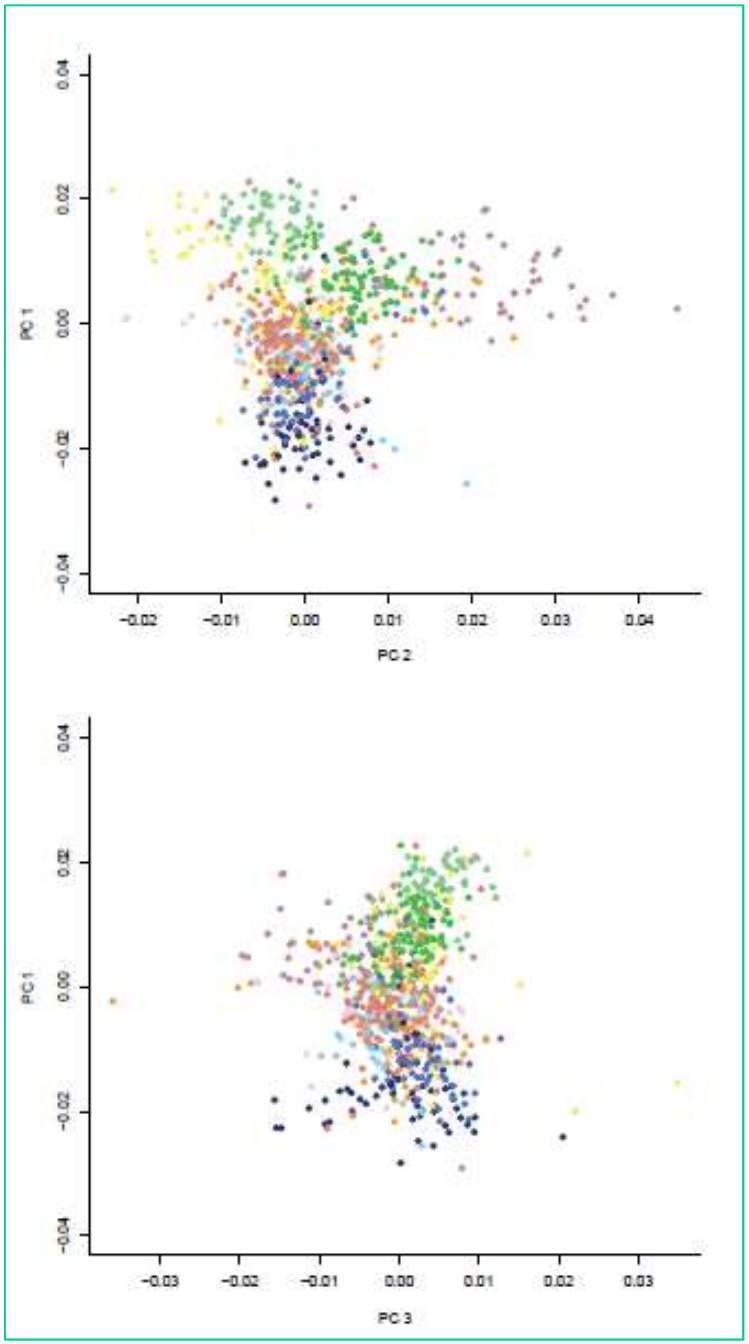
Leiden Longevity

Rotterdam Study





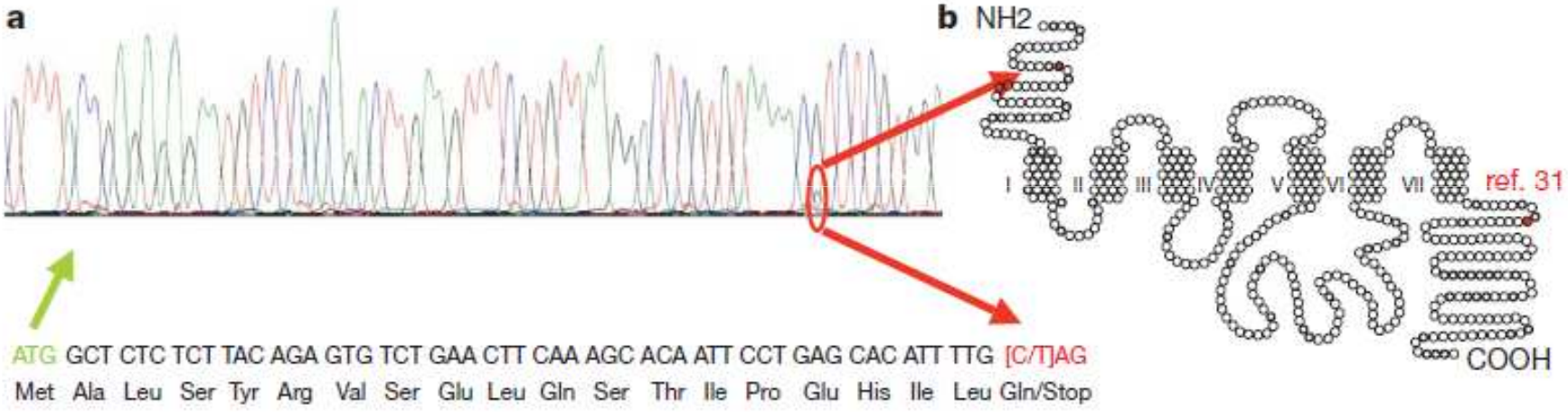
华大基因 BGI



PC1 versus PC2 and PC1 versus PC 3 in 769 GoNL samples . PC (1) North - South of the Netherlands; (2) East and West; and (3) between the middle-band of the Netherlands and the rest of the country.

A population-specific *HTR2B* stop codon predisposes to severe impulsivity

Laura Bevilacqua[1], Stéphane Doly[2], Jaakko Kaprio[3,4,5], Qiaoping Yuan[1], Roope Tikkanen[6], Tiina Paunio[7], Zhifeng Zhou[1], Juho Wedenoja[8,9], Luc Maroteaux[2], Silvina Diaz[2], Arnaud Belmer[2], Colin A. Hodgkinson[1], Liliana Dell'Osso[10], Jaana Suvisaari[7], Emil Coccaro[11], Richard J. Rose[12], Leena Peltonen‡, Matti Virkkunen[6,13] & David Goldman[1]



Exon-focused sequencing of impulsive individuals in a founder population, targeting fourteen genes belonging to the serotonin and dopamine domain. A stop codon in HTR2B was identified that is common (minor allele frequency > 1%) but exclusive to Finnish people (Nature 2010)

HTR2B Q20* is apparently exclusive to Finns. In >3,100 individuals representative of worldwide diversity, including the Human Genome Diversity Panel (Supplementary Table 8), one additional Q20* carrier was observed: a female with a Finnish surname and with alcoholism. Indicative of a common origin and founder population

zygous. However, although few Q20* carriers are criminals, violent criminals with Q20* seem to represent some of the most impulsive individuals within our violent offender cohort. Among 100–155 homicides annually in the Finnish population of 5.3 million, there are few instances of multiple homicide. In our sample, only three individuals were convicted of multiple homicide, and all three carried the Q20* allele.

Sequence differences between MZ twins?

-at candidate loci (e.g. van der Woude syndrome)

-CNV studies (e.g. Forsberg et al: function of age; Ehli et al: ADHD discordance; Veenma et al: Congenital Diaphragmatic Hernia and Esophageal Atresia (EA)

-whole genome: Baranzini (3 pairs discordant for MS), Ye et al. (2 pairs), GoNL (11 pairs)

Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes

Shinji Kondo[1*], Brian C. Schutte[1,2*], Rebecca J. Richardson[3†], Bryan C. Bjork[4†], Alexandra S. Knight[3], Yoriko Watanabe[1], Emma Howard[3], Renata L.L. Ferreira de Lima[5], Sandra Daack-Hirsch[1], Achim Sander[6], Donna M. McDonald-McGinn[7], Elaine H. Zackai[7], Edward J. Lammer[8], Arthur S. Aylsworth[9], Holly H. Ardinger[10], Andrew C. Lidral[11], Barbara R. Pober[12], Lina Moreno[13], Mauricio Arcos-Burgos[14], Consuelo Valencia[14], Claude Houdayer[15], Michel Bahuau[15,16], Danilo Moretti-Ferreira[5], Antonio Richieri-Costa[17], Michael J. Dixon[3] & Jeffrey C. Murray[1,2,18]

nature genetics • volume 32 • october 2002

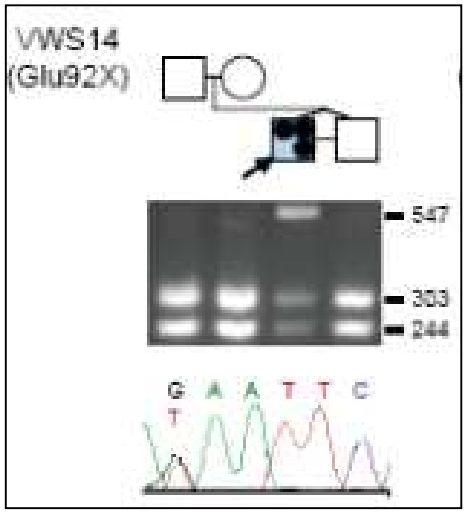
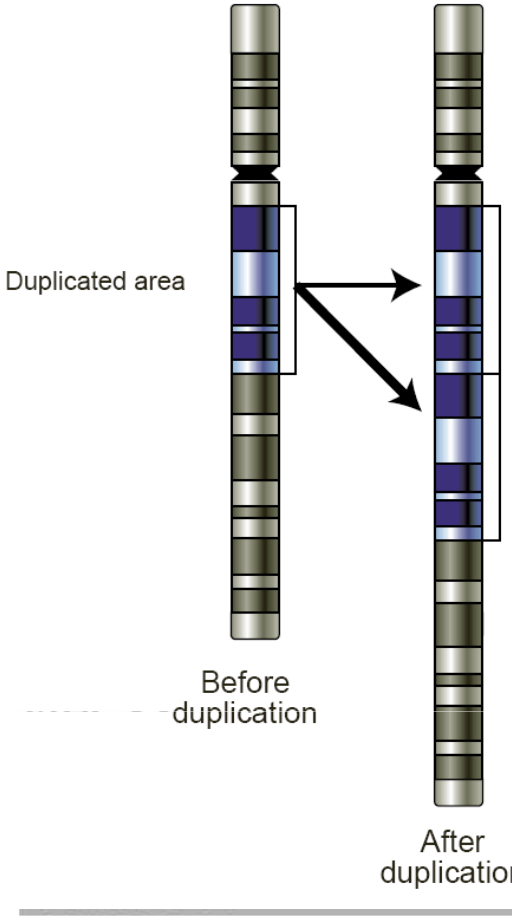


Fig. 1 Mutations in *IRF6* cause VWS and PPS. a, Family number and mutation found for two VWS pedigrees and one PPS pedigree. The gender of each individual was randomly assigned to preserve the anonymity of the pedigrees; the actual pedigrees are available on request. Unaffected individuals (open), probands (arrow) and individuals with VWS (blue) or PPS (red) are indicated. Symbols representing specific phenotypes are shown below the pedigree for family VWS25. The sequence chromatogram derived from the affected proband is shown below the pedigrees for families VWS14 and PPS6. Above is an image of an agarose gel that shows the restriction-fragment length polymorphism (RFLP) assay used to confirm these mutations. Numbers on the side of each gel represent the size of the RFLP products. The mutation in family VWS14 abolishes an *Eco*RI restriction site, whereas the mutation in family PPS6 abolishes an *Hha*I site. Consequently,



Sequence differences between MZ twins?

European Journal of Human Genetics (2012), 1–7

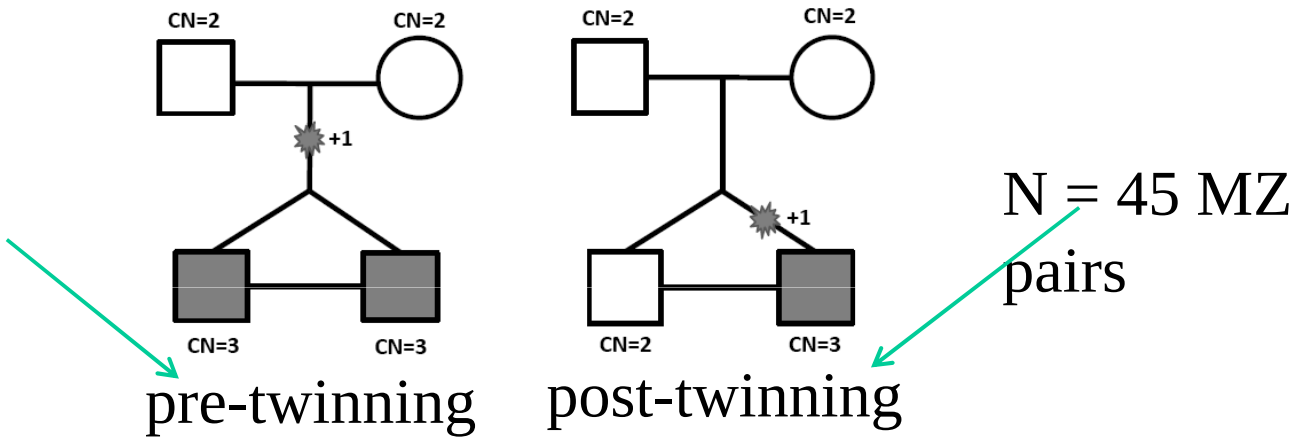
www.nature.com/ejhg

De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems

Erik A Ehli[*,1,2,6], Abdel Abdellaoui[*,3,6], Yueshan Hu[1], Jouke Jan Hottenga[3], Mathijs Kattenberg[3], Toos van Beijsterveldt[3], Meike Bartels[3], Robert R Althoff[4], Xiangjun Xiao[5], Paul Scheet[5], Eco J de Geus[3], James J Hudziak[4], Dorret I Boomsma[3,6] and Gareth E Davies[1,2,6]

CNVs in MZ twin pairs discordant or concordant for AP

N = 25 MZ pairs & their parents.



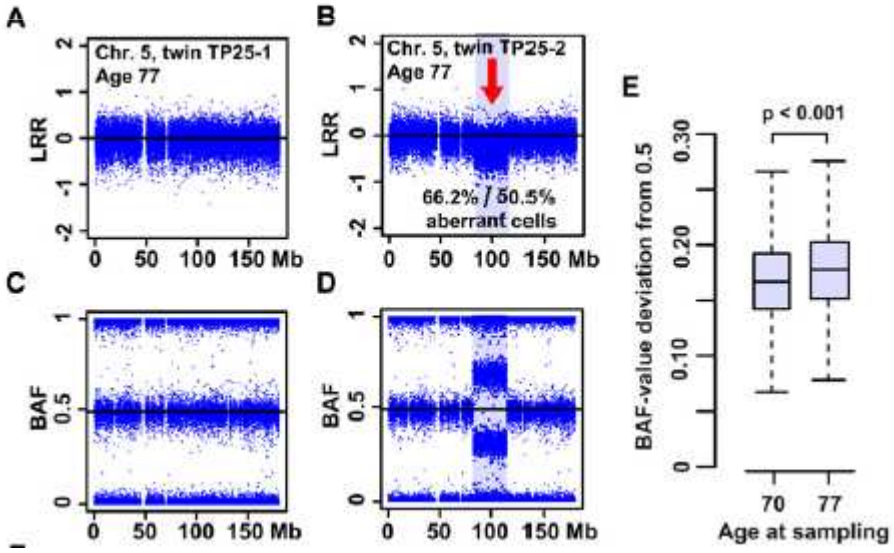
N = 45 MZ pairs

2 *de novo* CNVs were validated through qPCR (1 pre-twinning and 1 post-twinning) & 1 possible *de novo* from a somatic mutation that resulted in mosaicism in the affected twin of a discordant pair

ARTICLE

Age-Related Somatic Structural Changes in the Nuclear Genome of Human Blood Cells

Lars A. Forsberg,[1] Chiara Rasi,[1] Hamid R. Razzaghian,[1] Geeta Pakalapati,[1] Lindsay Waite,[2] Krista Stanton Thilbeault,[2] Anna Ronowicz,[3] Nathan E. Wineinger,[4] Hemant K. Tiwari,[4] Dorret Boomsma,[5] Maxwell P. Westerman,[6] Jennifer R. Harris,[7] Robert Lyle,[8] Magnus Essand,[1] Fredrik Eriksson,[1] Themistocles L. Assimes,[9] Carlos Iribarren,[10] Eric Strachan,[11] Terrance P. O'Hanlon,[12] Lisa G. Rider,[12] Frederick W. Miller,[12] Vilmantas Giedraitis,[13] Lars Lannfelt,[13] Martin Ingelsson,[13] Arkadiusz Piotrowski,[3] Nancy L. Pedersen,[14] Devin Absher,[2] and Jan P. Dumanski[1,*]



(A) A normal profile of MZ twin TP25-1. (B) A 32.5 Mb deletion on 5q of co-twin TP25-2 (deletion uncovered with LRR data from Illumina SNP array)
(C and D) The BAF profiles of twins. The qPCR experiments showed that 66.2% of nucleated blood cells in TP25-2 had the 5q deletion ; 50.5% of the cells had the 5q deletion when twins were 77 years old.
(E) The deviation of BAF values from 0.5 (the allelic fraction of intensity at each heterozygous SNP) Percentage of cells with the 5q deletion was higher when the subjects were 77 years old than when they were 70 years old (p < .001)

Vol 464|29 April 2010|doi:10.1038/nature08990 nature

LETTERS

Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis

Sergio E. Baranzini[1], Joann Mudge[2], Jennifer C. van Velkinburgh[2], Pouya Khankhanian[1], Irina Khrebtukova[3], Neil A. Miller[2], Lu Zhang[3], Andrew D. Farmer[2], Callum J. Bell[2], Ryan W. Kim[2], Gregory D. May[2], Jimmy E. Woodward[2], Stacy J. Caillier[1], Joseph P. McElroy[1], Refujia Gomez[1], Marcelo J. Pando[4], Leonda E. Clendenen[2], Elena E. Ganusova[2], Faye D. Schilkey[2], Thiruvarangan Ramaraj[2], Omar A. Khan[5], Jim J. Huntley[3], Shujun Luo[3], Pui-yan Kwok[6,7], Thomas D. Wu[8], Gary P. Schroth[3], Jorge R. Oksenberg[1,7], Stephen L. Hauser[1,7] & Stephen F. Kingsmore[2]

No reproducible differences were detected between co-twins among 3.6 million single nucleotide polymorphisms (SNPs) or 0.2 million insertion-deletion polymorphisms.

Genetic diagnosis by whole exome capture and massively parallel DNA sequencing

PNAS | November 10, 2009

Murim Choi[a], Ute I. Scholl[a], Weizhen Ji[a], Tiewen Liu[a], Irina R. Tikhonova[b], Paul Zumbo[b], Ahmet Nayir[c], Ayşin Bakkaloğlu[d], Seza Özen[d], Sami Sanjad[e], Carol Nelson-Williams[a], Anita Farhi[a], Shrikant Mane[b], and Richard P. Lifton[a,1]

[a]Depar
Haven,
Rheum

Contrib

Protei
genon
diseas
seque
poten
humar
ing co
DNA s
appro
sensiti
gous v
unant
patien

REPORT

Exome Sequencing Identifies *WDR35* Variants Involved in Sensenbrenner Syndrome

Christian Gilissen,[1,3] Heleen H. Arts,[1,3] Alexander Hois
Peer Arts,[1] Bart van Lier,[1] Marloes Steehouwer,[1] Jeroen
Ronald Roepman,[1] Nine V.A.M. Knoers,[1] Joris A. Veltm

Sensenbrenner syndrome/cranioectodermal dysplasia (CED) is an autoso
and ectodermal and skeletal abnormalities. We sequenced the exomes of
zygous mutations in *WDR35* as the cause of the disease in each of the two
causative gene by sequencing the exome of a single sporadic patient. With
of *WDR35* alters splicing of RNA on the affected allele, introducing a pren
Tubby superfamily) and has previously been characterized as an intrafla
syndrome is a ciliary disorder.

American Journal of Human Genetics *87*, 418–423, September 10, 2010

NATURE GENETICS VOLUME 42 | NUMBER 6 | JUNE 2010

De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome

Alexander Hoischen[1,14], Bregje W M van Bon[1,14], Christian Gilissen[1,14], Peer Arts[1], Bart van Lier[1], Marloes Steehouwer[1], Petra de Vries[1], Rick de Reuver[1], Nienke Wieskamp[1], Geert Mortier[2], Koen Devriendt[3], Marta Z Amorim[4], Nicole Revencu[5], Alexa Kidd[6], Mafalda Barbosa[7], Anne Turner[8], Janine Smith[9], Christina Oley[10], Alex Henderson[11], Ian M Hayes[12], Elizabeth M Thompson[13], Han G Brunner[1], Bert B A de Vries[1] & Joris A Veltman[1]

Schinzel-Giedion syndrome is characterized by severe mental retardation, distinctive facial features and multiple congenital malformations; most affected individuals die before the age of ten. We sequenced the exomes of four affected individuals (cases) and found heterozygous *de novo* variants in *SETBP1* in all four. We also identified *SETBP1* mutations in eight additional cases using Sanger sequencing. All mutations clustered to a highly conserved 11-bp exonic region, suggesting a dominant-negative or gain-of-function effect.

Large twin registers with collections of DNA, GWA, sequence data

- How to deal with data from relatives
- Special value in MZ twins?
- Often longitudinal phenotype data

Heritability studies of

- MtDNA content
- Telomere length
- Expression profiles
- Metabolomics
- Epigenetics

Increased DNA Methylation at the *AXIN1* Gene in a Monozygotic Twin from a Pair Discordant for a Caudal Duplication Anomaly

N. A. Oates, J. van Vliet, D. L. Duffy, H. Y. Kroes, N. G. Martin, D. I. Boomsma, M. Campbell, M. G. Coulthard, E. Whitelaw, and S. Chong

The *AXIN1* gene has been implicated in caudal duplication anomalies. Its coding region was sequenced in both members of a monozygotic (MZ) twin pair discordant for a caudal duplication anomaly, but no mutation was found. Using bisulfite sequencing, we examined methylation at the promoter region of the *AXIN1* gene in these twins and in twin and age-matched singleton controls. Methylation of the promoter region in peripheral blood mononucleated cells was variable among individuals, including MZ pairs. In the MZ pair discordant for the caudal duplication, this region of the affected twin was significantly more methylated than that of the unaffected twin ($P < .0001$), which was significantly more methylated than those of the controls ($P = .02$). We have confirmed that this CpG island does function as a promoter in vitro and that its activity is inversely proportional to the extent of methylation. This finding raises the possibility that hypermethylation of the *AXIN1* promoter, by mechanisms as yet undetermined, is associated with the malformation. This case may be paradigmatic for some cases of MZ discordance.

DISCORDANT MZ TWINS

Epigenetics?

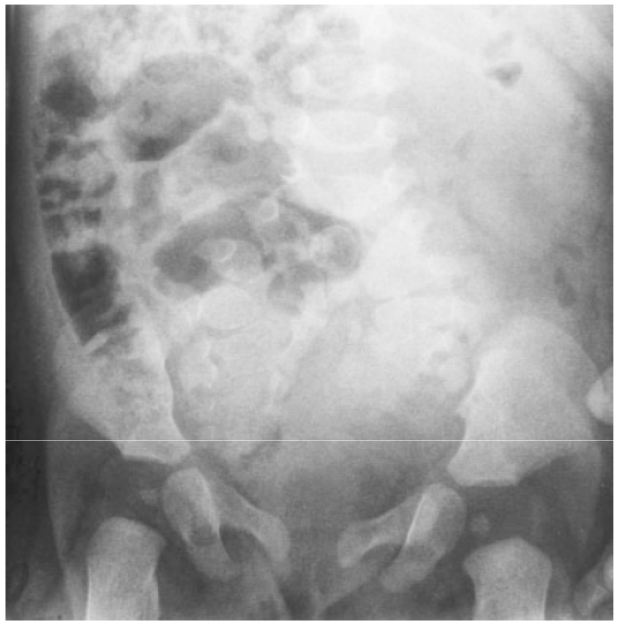
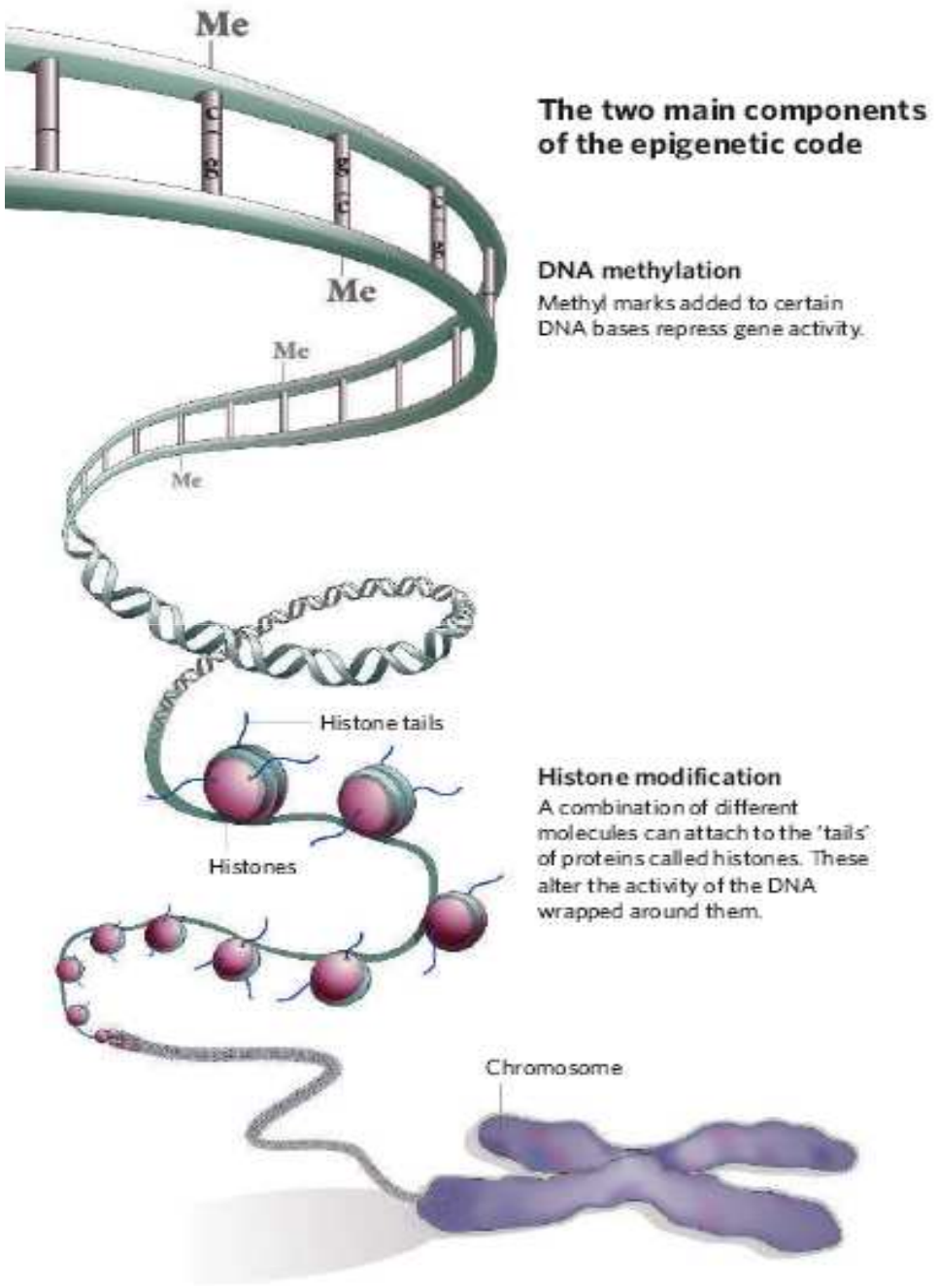


Fig. 2. Patient 1. Radiograph of the vertebral column shows complete duplication of the spine from L4 down.

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



Heritabilities from twin studies in humans

Anthropometric	Heritability	N pairs
Height	**0.68-0.90**	30111
Body Mass Index	**0.64-0.79**	37000
Birth weight	**0.42**	2009
Metabolic and cardiovascular		
Diabetes, Type 1	**0.88**	22650
Diabetes, Type 2	**0.64**	13888
CHD	**M: 0.57; F: 0.38**	10483
SBP	**0.42**	1617
DBP	**0.40**	1617
Markers for cardiovascular disease in blood		
High density lipoprotein	**0.66**	6000
Low density lipoprotein	**0.53**	
Triglyceride level		
Glucose level		
C-reactive protein		

Brain,CNS, psychiatric disorders		
Alzheimer	**0.48**	662
Parkinson	**0.34**	46436 twins
Migraine	**0.34-0.57**	29717
Multiple Sclerosis	**0.25-0.76**	review
ADHD kids	**0.76**	review
Autism Spectrum	**0.71**	11535 twins
Schizophrenia	**0.81**	meta-A
Major Depression	**0.37**	meta-A
EEG measures of brain activity		
Alpha power	**0.79**	meta-A
P300 amplitude	**0.60**	
Skeletal features and disorders		

Discordant MZ twin design

MZ concordance

MZ and DZ twin concordance for complex disease	Probandwise concordance (%)	
	MZ twins	DZ twins
Diabetes Type 1	**42.9**	7.4
Diabetes Type 2	**34**	16
Multiple Sclerosis	**25.3**	5.4
Crohn Disease	**38**	2
Ulcerative Colitis	**15**	8
Alzheimer's Disease	**32.2**	8.7
Parkinson Disease	**15.5**	11.1
Schizophrenia	**40.8**	5.3
Major Depresssion	**31.1**	25.1

Two sides of the coin

Personalized medicine?

Incomplete concordance of MZ twins indicates that a genome cannot predict individual disease outcome.

The fact that MZ twin concordance for common disorders is generally not high has important implications for genomic risk prediction and the ethical concerns that have been raised in this light.