Genetic Epidemiology in the Genomic Age: The Role of Twin Studies in the Genomic Era & Missing Heritability

Nick Martin

Queensland Institute of Medical Research

Brisbane

Intro Workshop

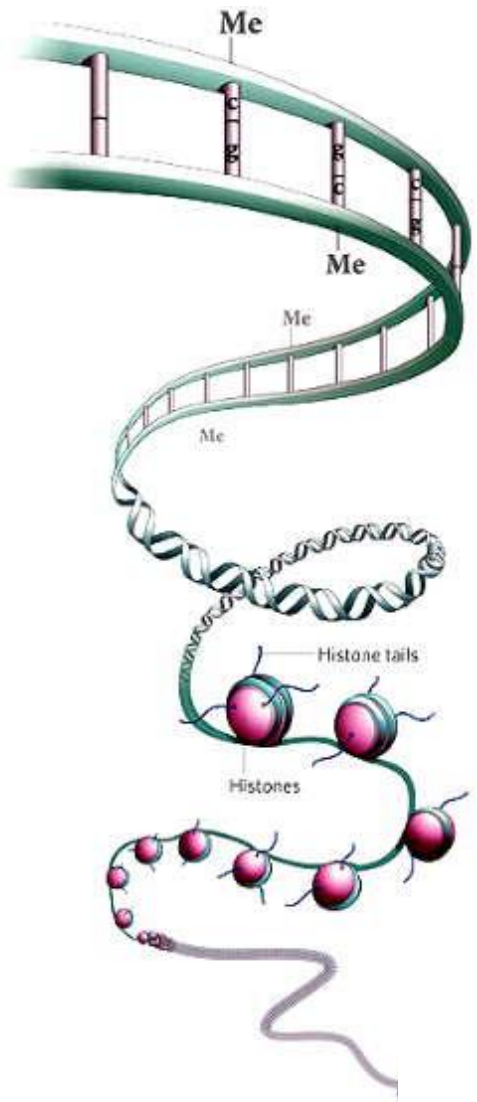
Boulder

March 7, 2014

Genetic Epidemiology: Stages of Genetic Mapping

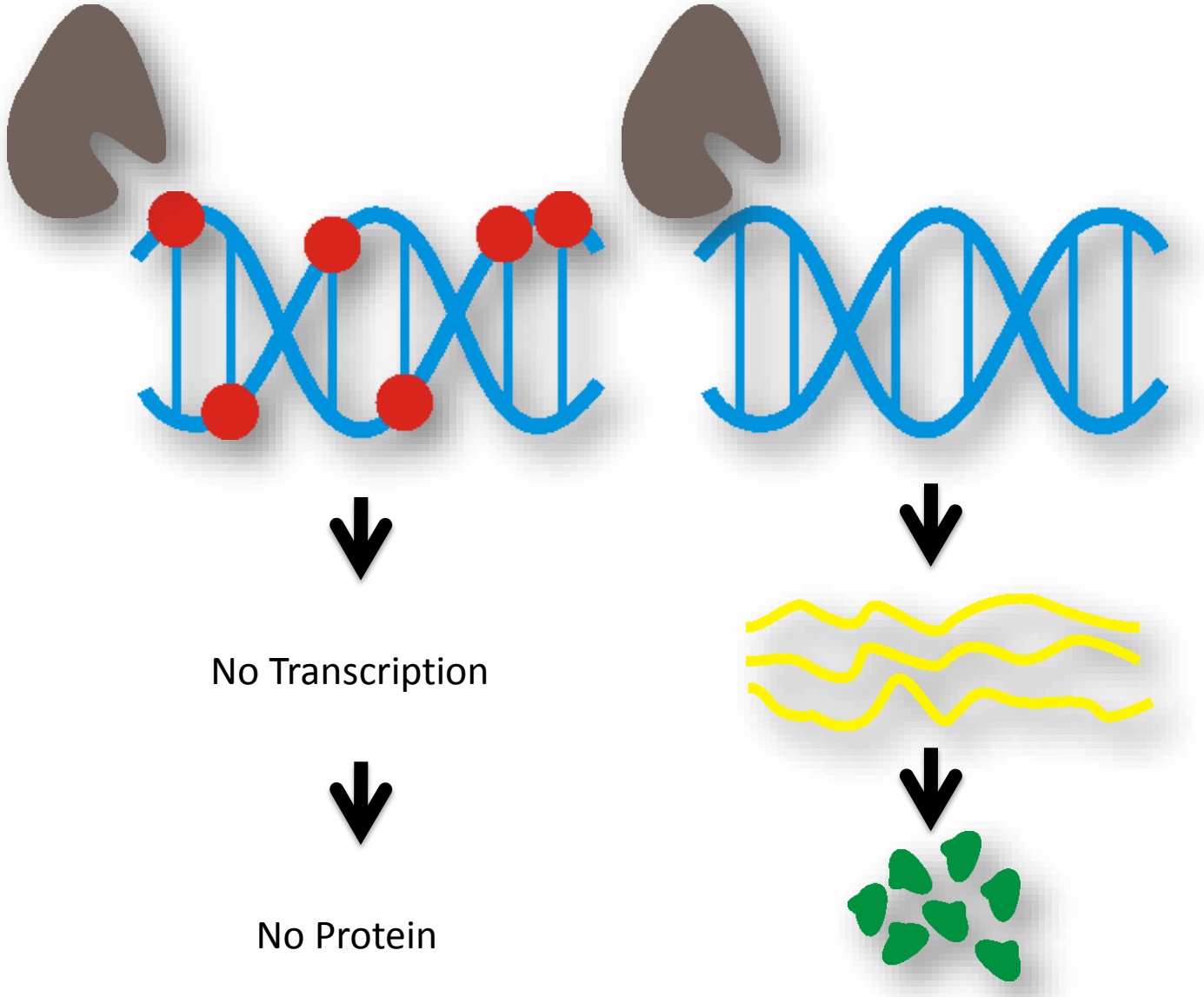
- Are there genes influencing this trait?
  - Twin family studies of some genomic phenotypes
- Where are those genes?
  - Linkage analysis
- What are those genes?
  - Association analysis
- How do they work beyond the sequence?
  - Epigenetics, transcriptomics, proteomics
- What can we do with them ?
  - Translational medicine

Epigenetic mechanisms
- DNA methylation
- Histone binding



- Modifications of genome other than nucleotide changes that regulate gene expression (e.g. methylation of cytosines, histone modifications, microRNAs, …)

How DNA methylation affects gene transcription (gene expression)

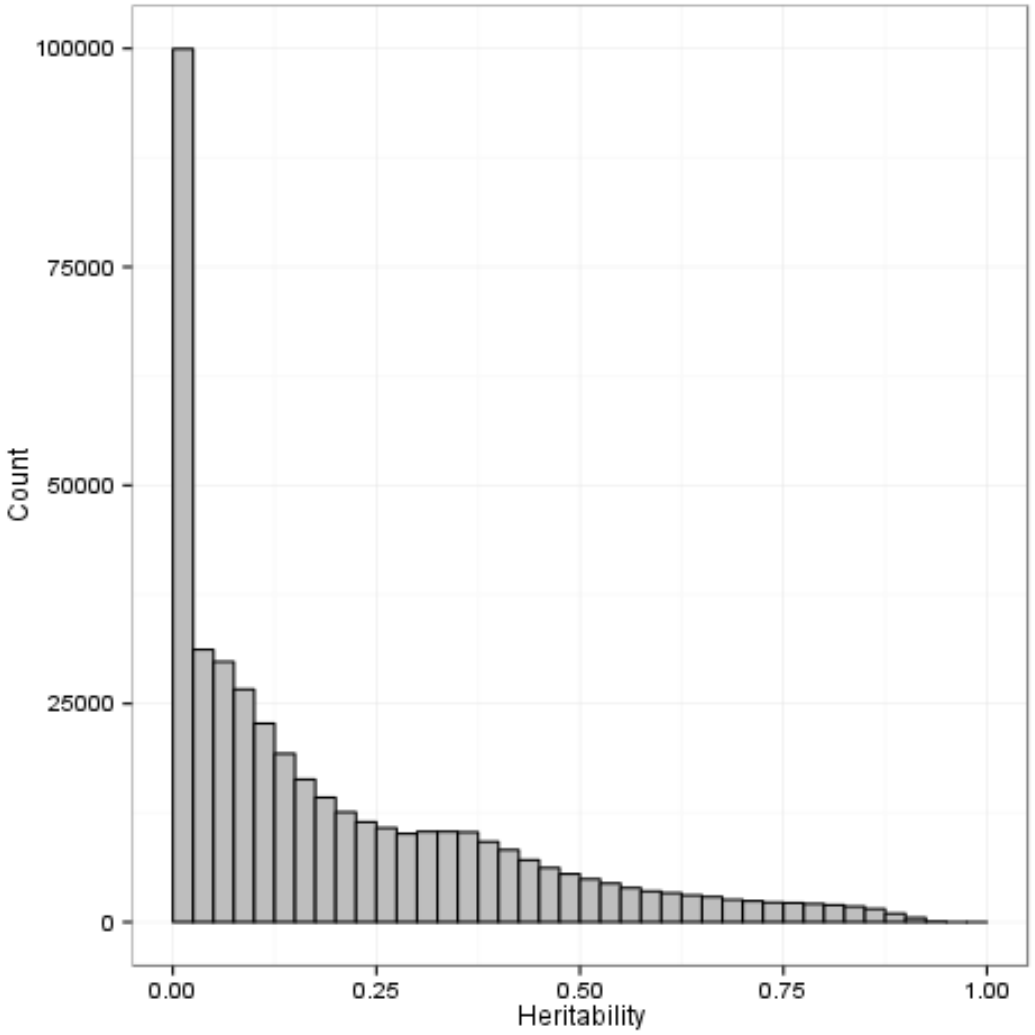


Average correlation across all probes of normalised methylation measurements between relative pairs

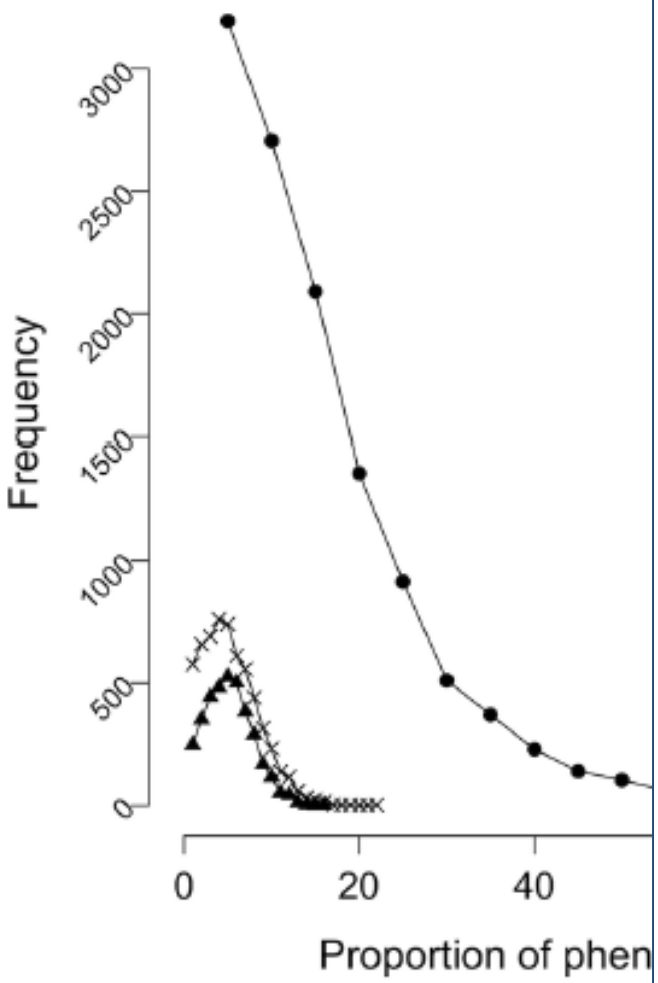
Relationship	# Pairs	Correlation	Expected
MZ twin	67	0.200	h^2
DZ twin	111	0.109	$h^2/2$
Sibling	262	0.090	$h^2/2$
Parent – Offspring	362	0.089	$h^2/2$
Parent – Parent	58	0.023	0
Unrelated	187331	-0.002	0

Allan McRae

Distribution of heritability estimates for DNA methylation levels



Allan McRae



Distribution heritability estimates across 47,585 transcripts (~700 MZ & 600 DZ Dutch pairs)

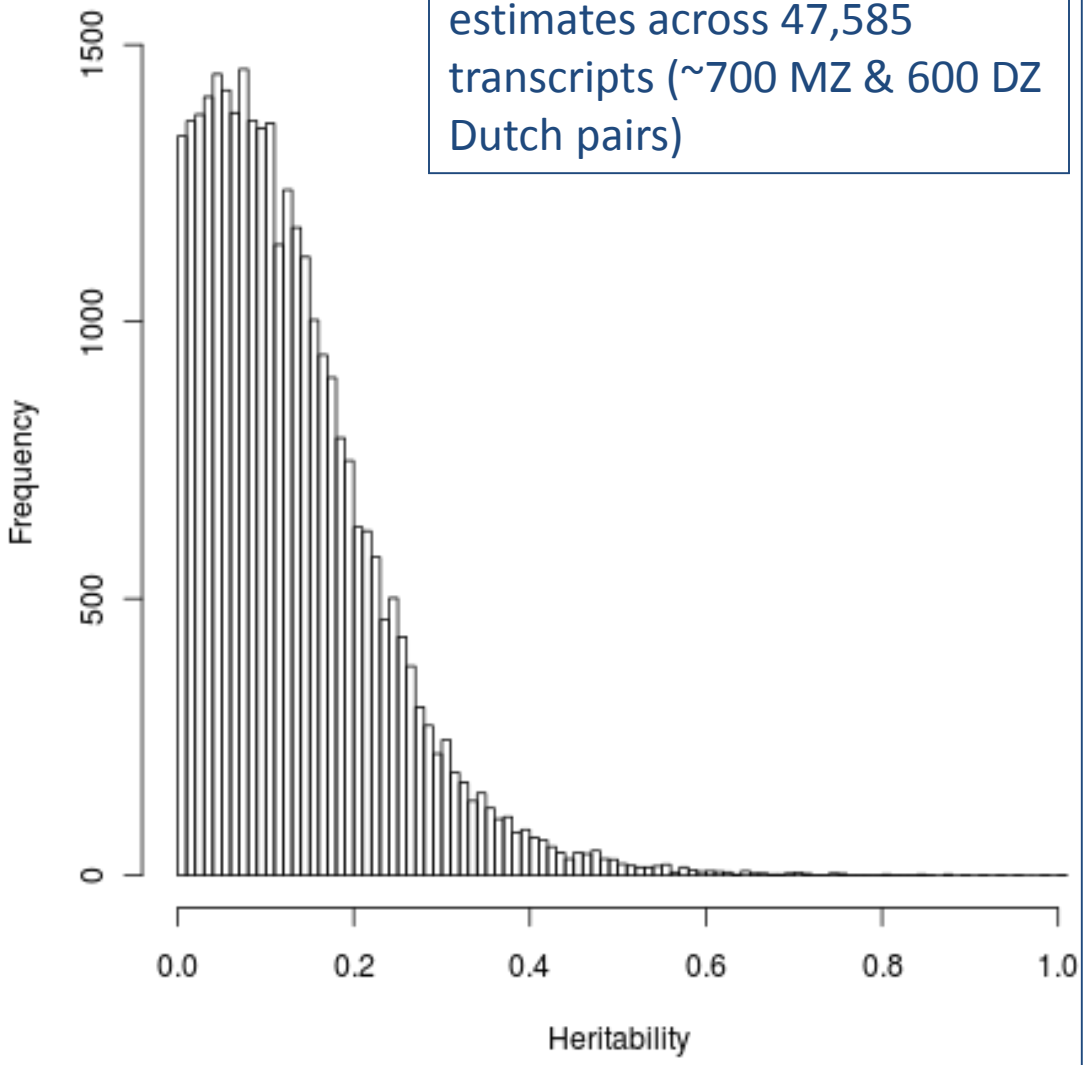


Figure 1. Components of variation
proportion of phenotypic variance attribut
non-additive genetic (d^2) and common fa
whose estimates are greater than zero are included. The distributions
for all probes are given in Figure S1. Estimates of h^2 and d^2 were
obtained by fitting an model [1], whist f^2 estimates were obtained from
model [3]

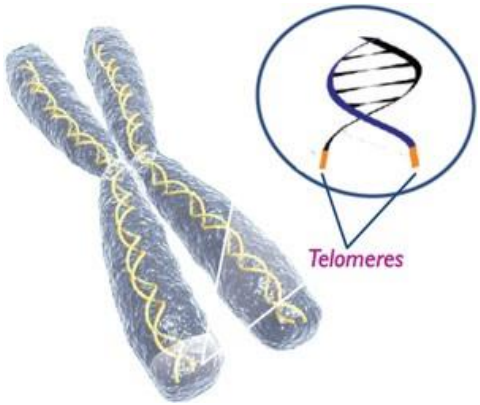
Meta-analysis of telomere length in 19,713 subjects

Linda Broer et al. (ENGAGE consortium)

	n	r	p-value
Siblings	1,553	**0.49**	$3.46*10^{-96}$
Monozygotic twins	2,534	**0.69**	0*
Dizygotic twins	1,940	**0.25**	$2.82*10^{-30}$
Spouses (<55)	962	**0.20**	$3.24*10^{-10}$
Spouses (>55)	977	**0.31**	$4.27*10^{-23}$

Parent offspring	n	r	p-value
Father-son	791	**0.34**	$2.57*10^{-23}$
Father-daughter	882	**0.33**	$3.99*10^{-24}$
Mother-son	850	**0.42**	$5.06*10^{-37}$
Mother-daughter	1,005	**0.42**	$2.99*10^{-45}$

Heritability ~70%



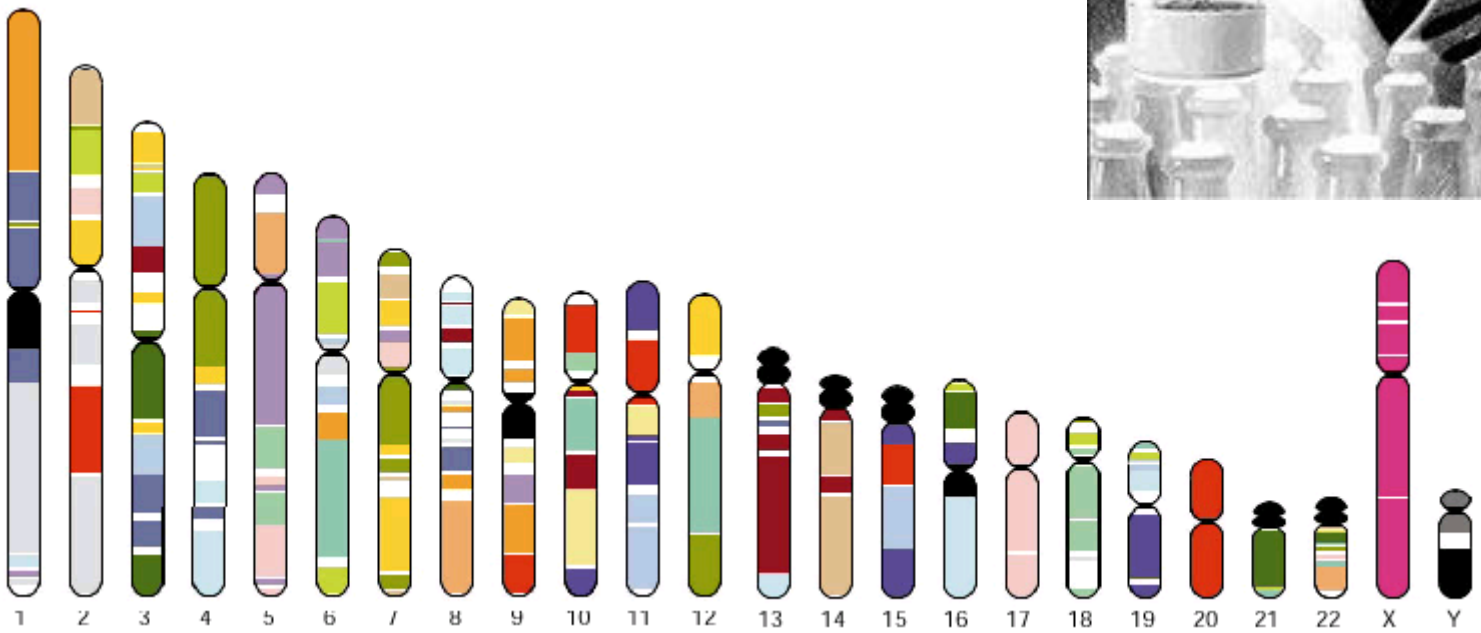
Eur J Hum Genet. 2013 Oct;21(10):1163-8.

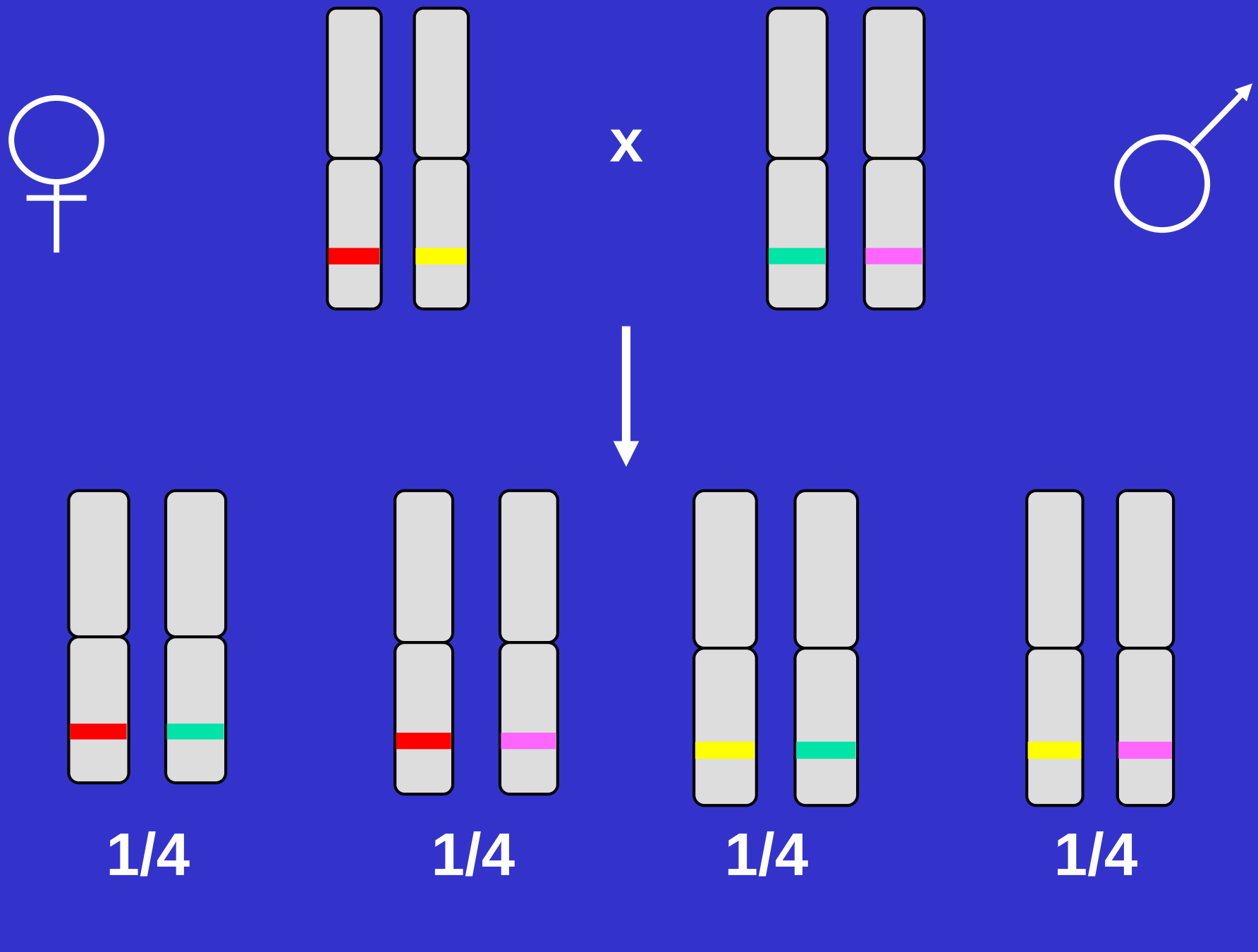
Genetic Epidemiology: Stages of Genetic Mapping

- Are there genes influencing this trait?
  - Genetic epidemiological studies
- Where are those genes?
  - Linkage analysis
- What are those genes?
  - Association analysis
- How do they work beyond the sequence?
  - Epigenetics, transcriptomics, proteomics
- What can we do with them ?
  - Translational medicine

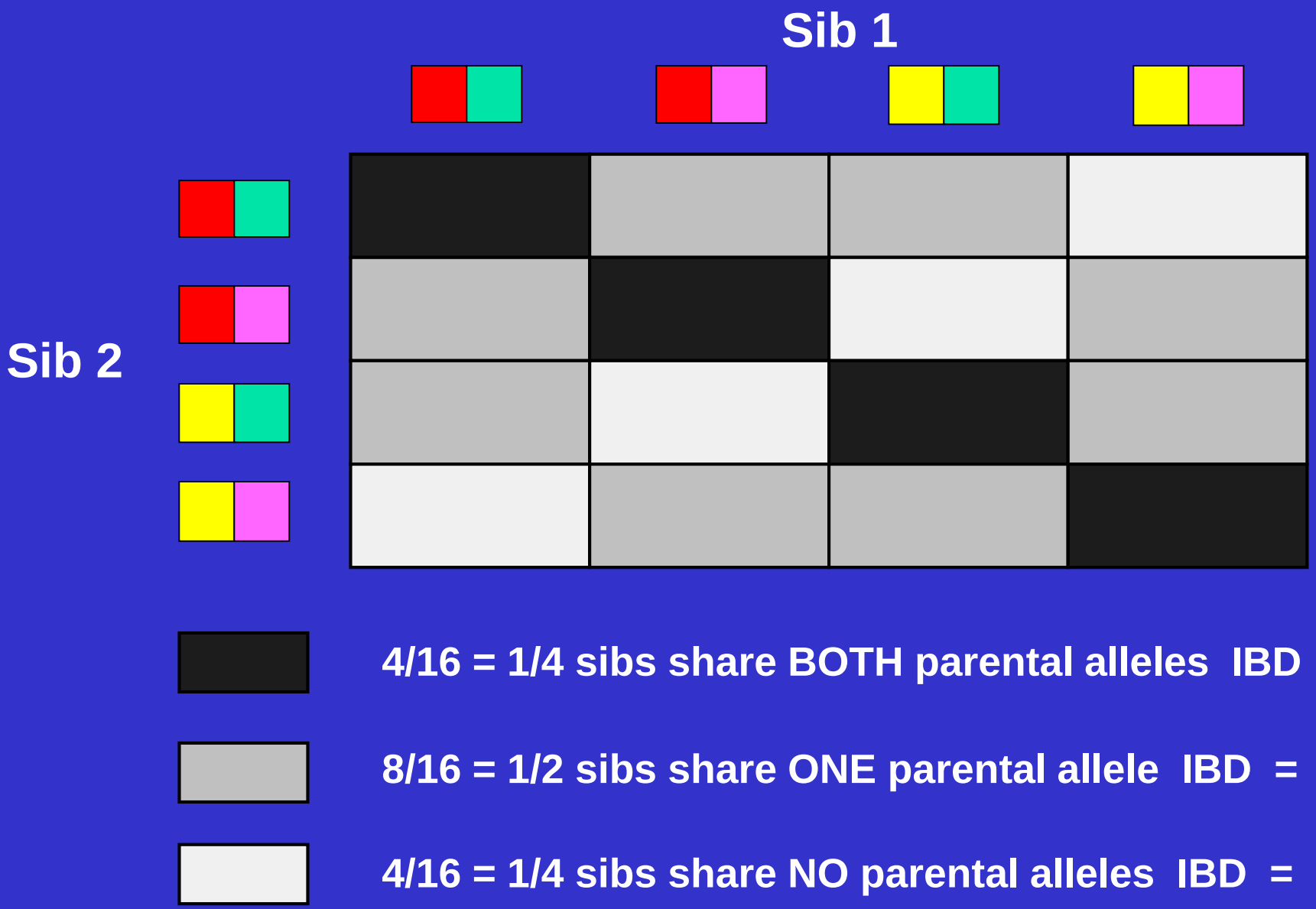
Thomas Hunt Morgan – discoverer of linkage

Linkage analysis

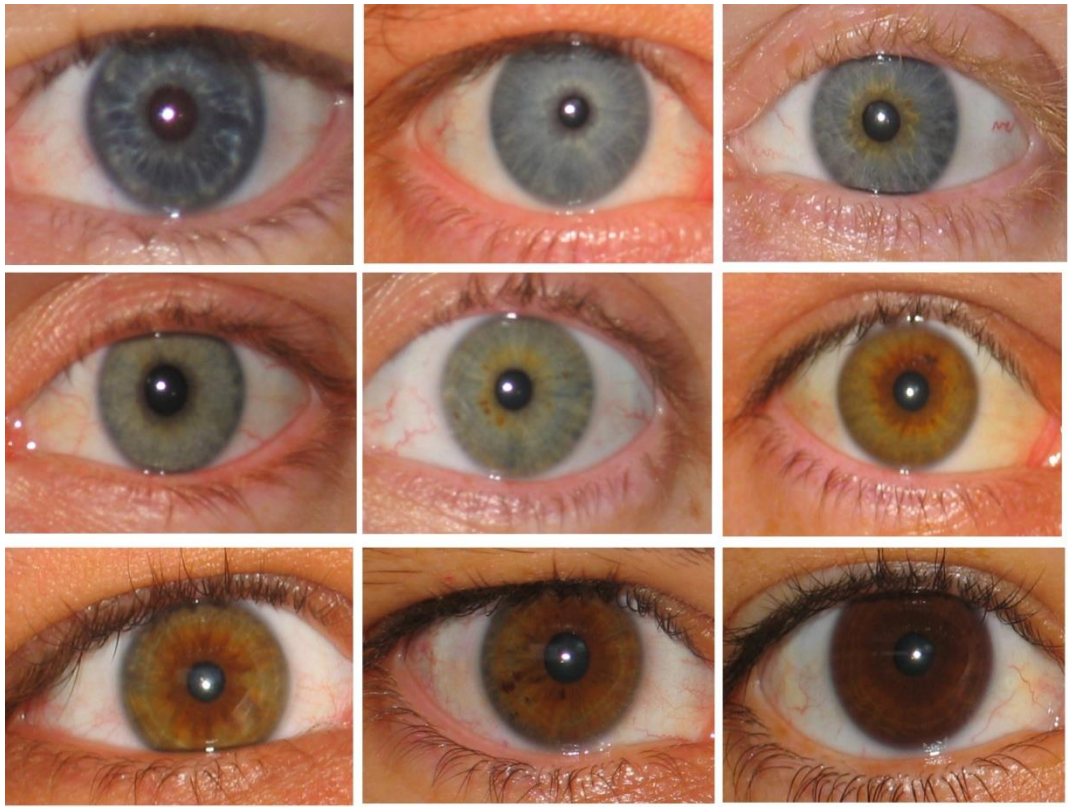


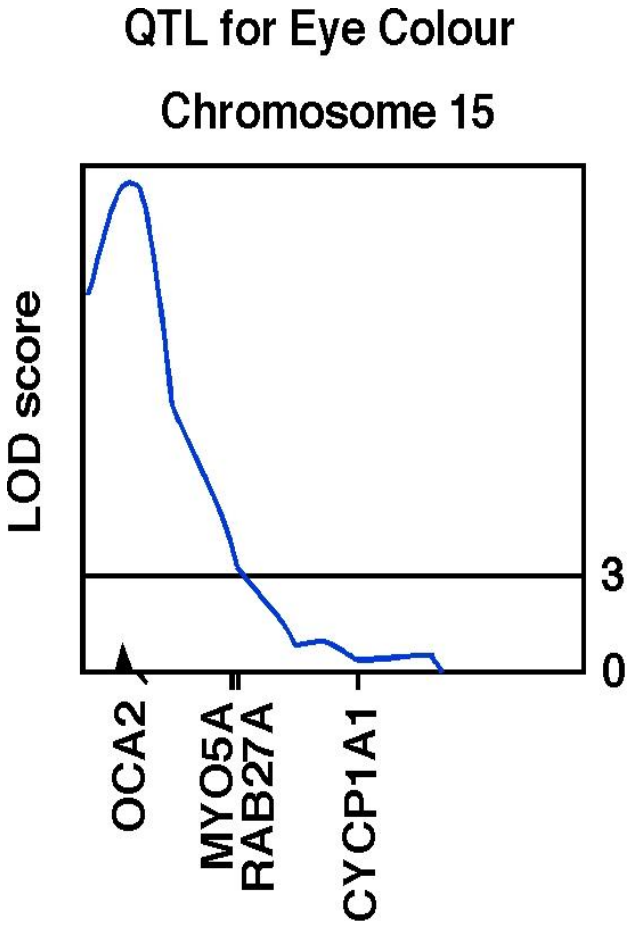

x
1/4
1/4
1/4
1/4

IDENTITY BY DESCENT



Human OCA2 and eye colour



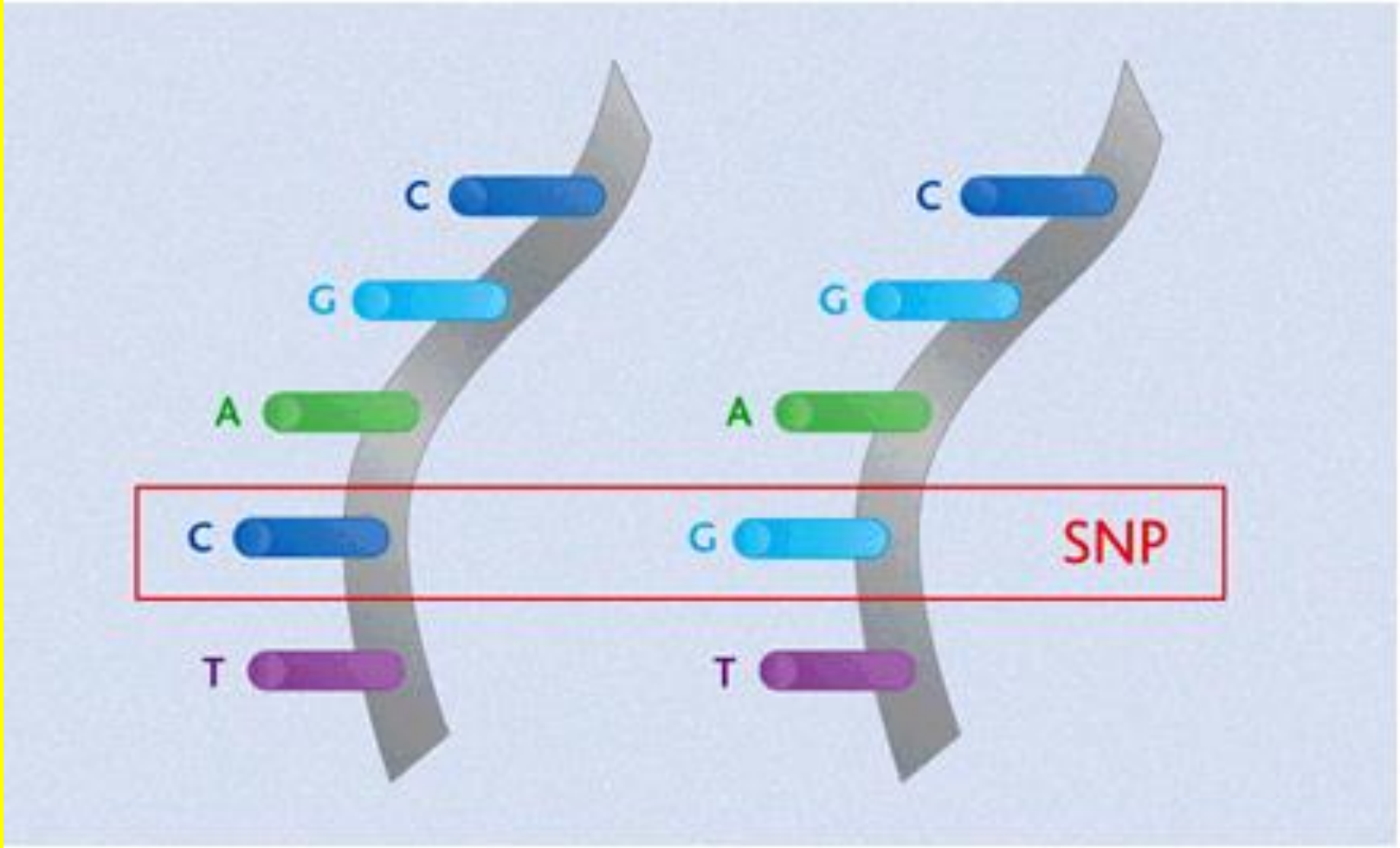


Zhu et al., *Twin Research* 7:197-210 (2004)

Finding the genes - association

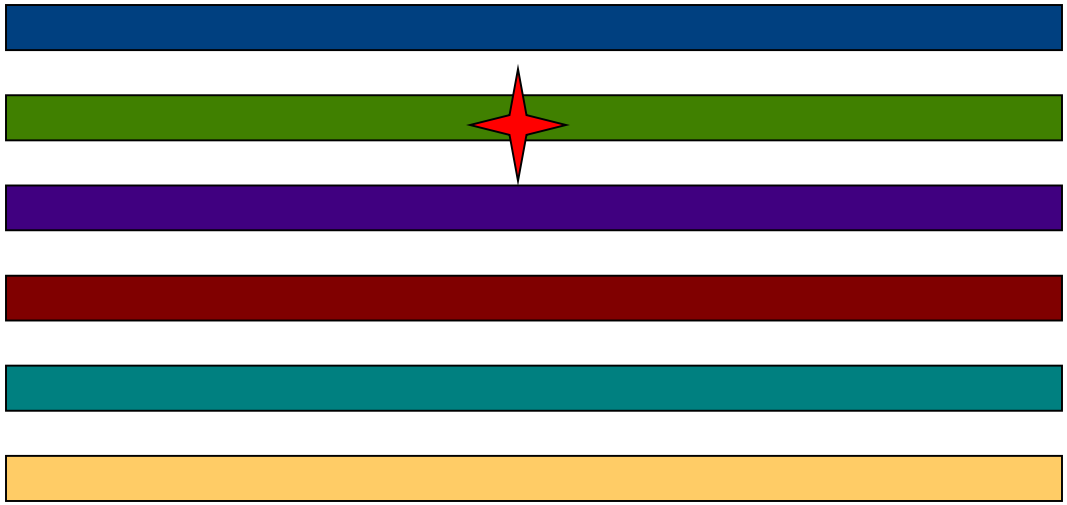
- Looks for correlation between specific alleles and phenotype (trait value, disease risk)

Variation: Single Nucleotide Polymorphisms

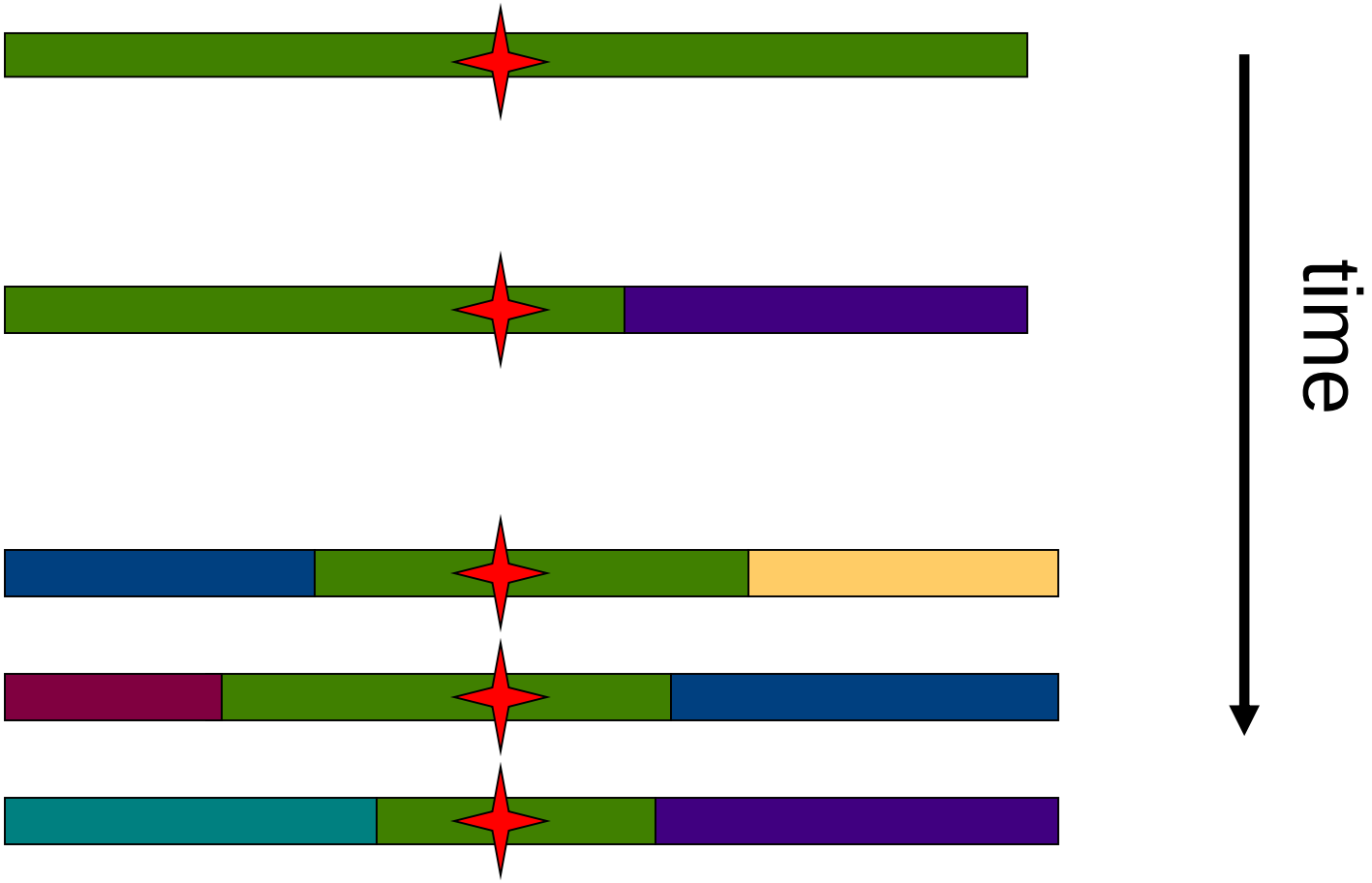


Complex disease marker? SNPs are single-base differences in DNA.

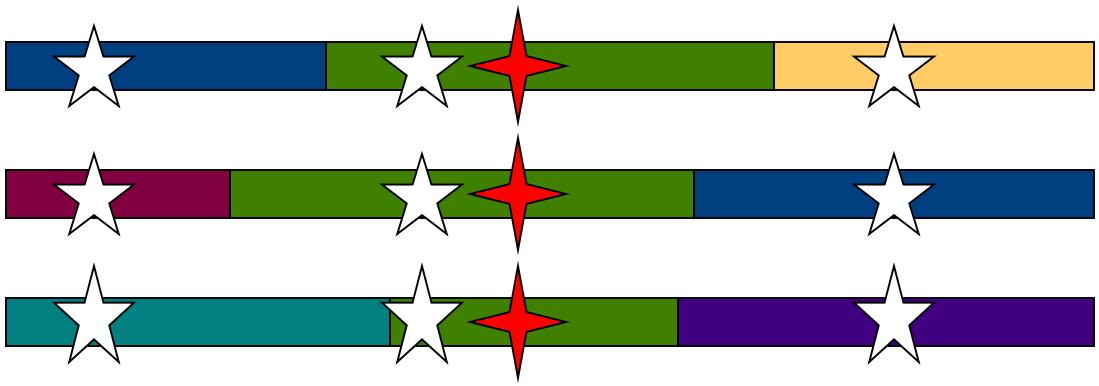
Linkage disequilibrium



Linkage disequilibrium

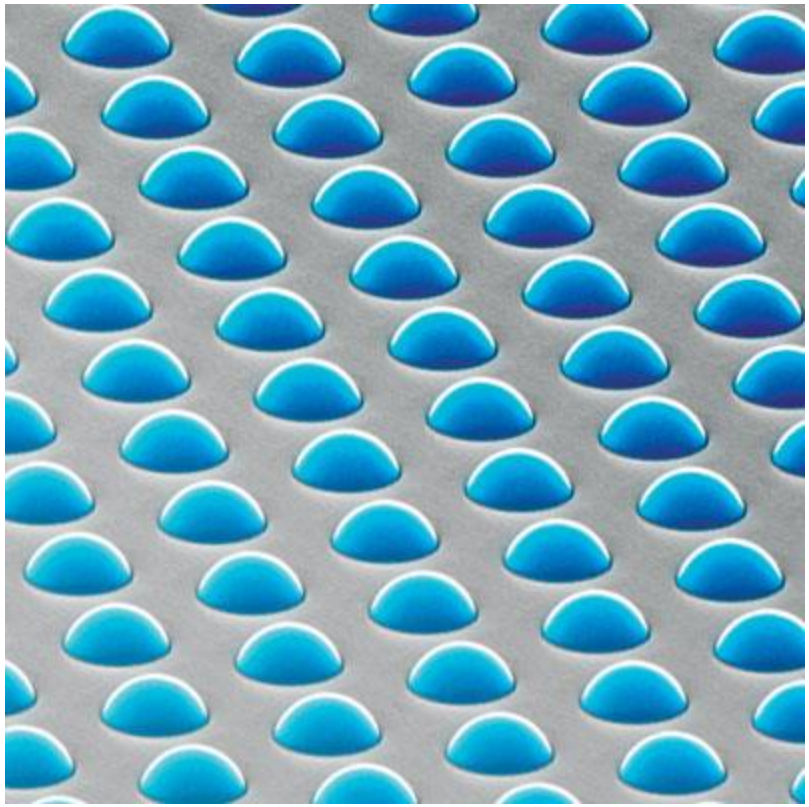


Indirect association

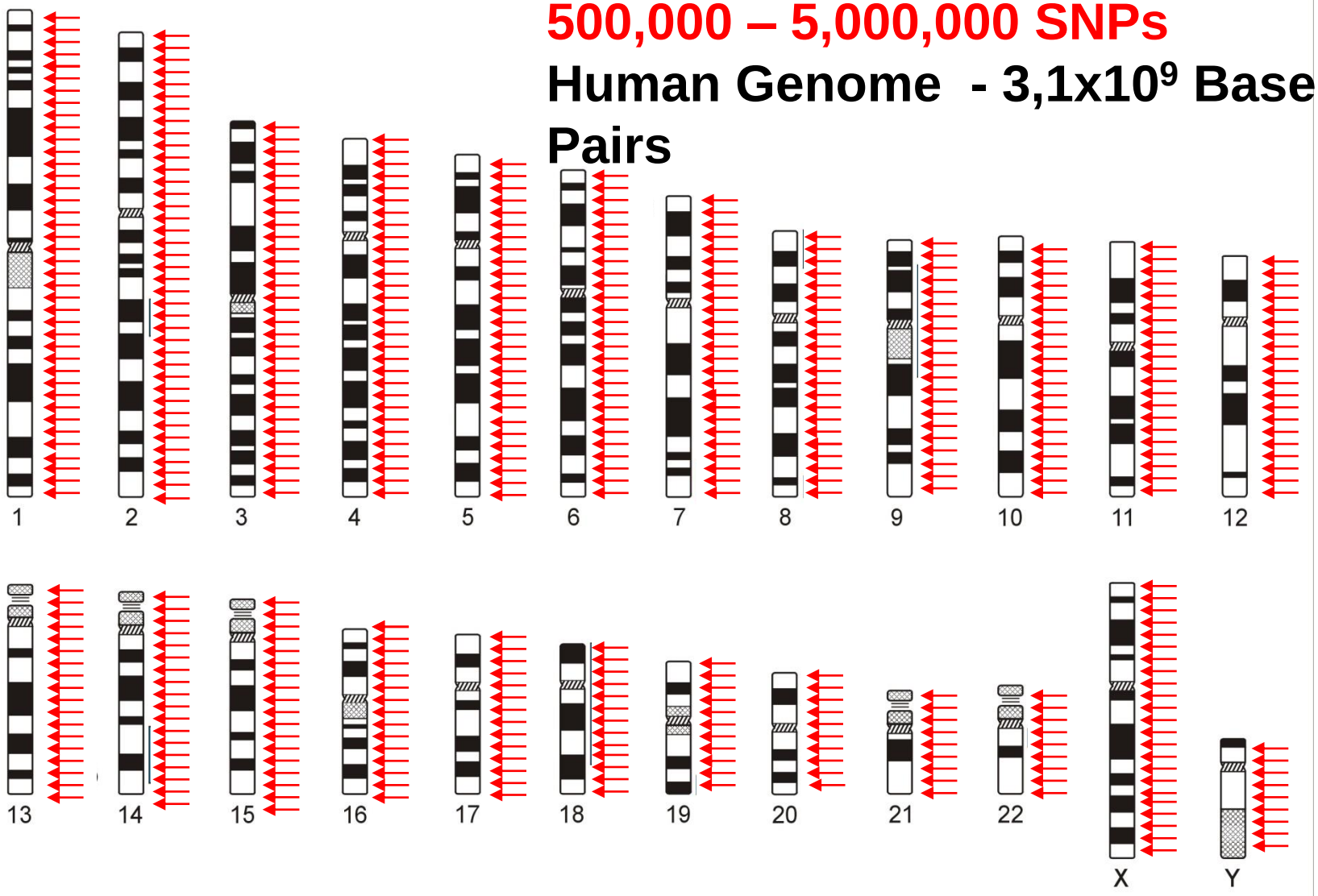


this SNP will be associated with disease

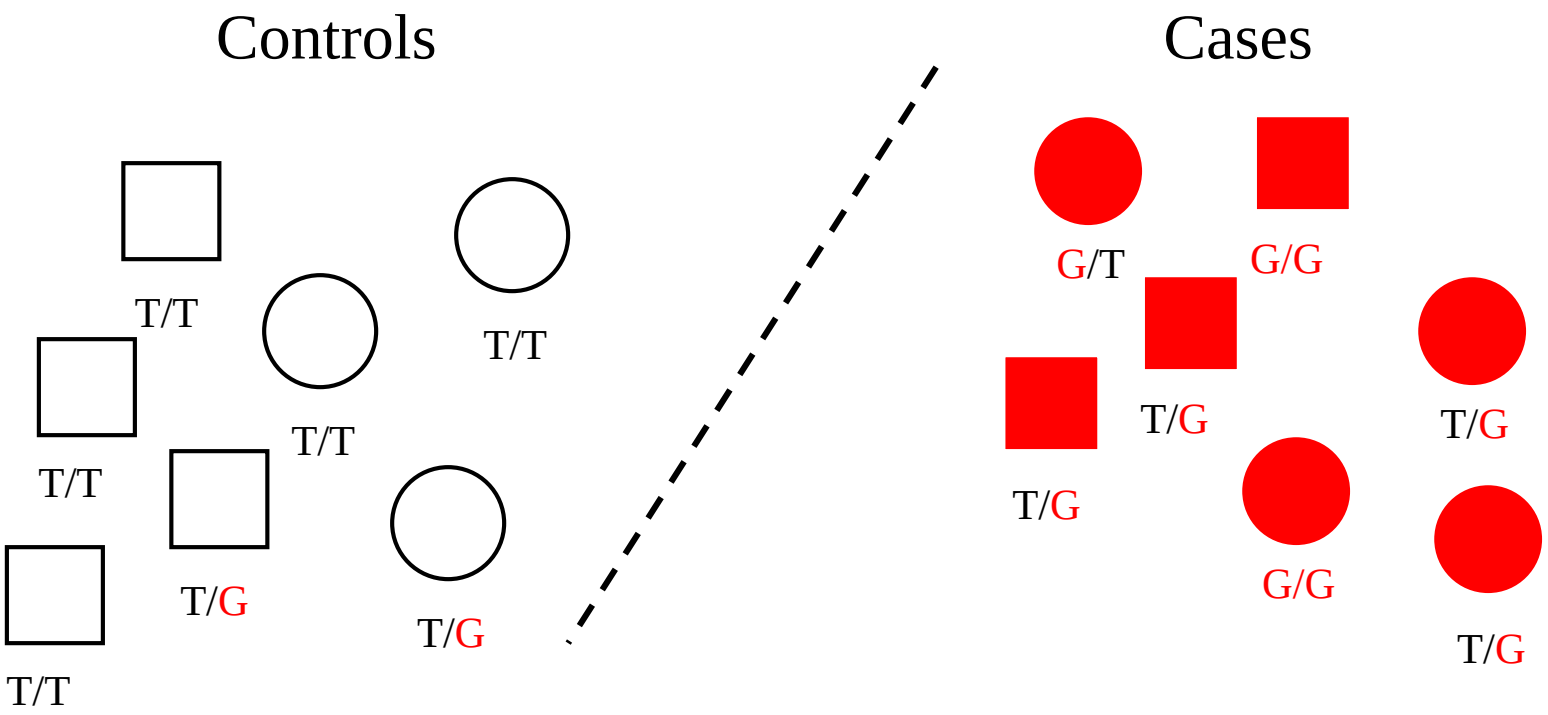
High density SNP arrays – up to 1 million SNPs



Genome-Wide Association Studies



Genetic Case Control Study



Allele G is ‘associated’ with disease

Allele-based tests (case-control)

- Each individual contributes two counts to 2x2 table.
- Test of association

$$X^2 = \sum_{i=0,1} \sum_{j=A,U} \frac{\left(n_{ij} - E\left[n_{ij}\right]\right)^2}{E\left[n_{ij}\right]}$$

where

$$E\left[n_{ij}\right] = \frac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$$

	Cases	Controls	Total
G	n_{1A}	n_{1U}	$n_{1\cdot}$
T	n_{0A}	n_{0U}	$n_{0\cdot}$
Total	$n_{\cdot A}$	$n_{\cdot U}$	$n_{\cdot\cdot}$

- X^2 has χ^2 distribution with 1 degrees of freedom under null hypothesis.

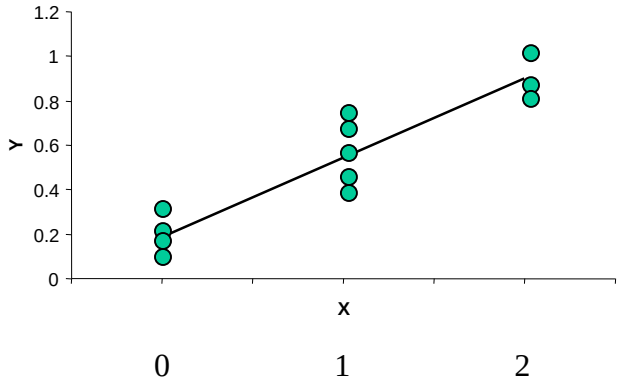
Simple Regression Model of Association (continuous trait)

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

where

Y_i = trait value for individual i

X_i = number of 'A' alleles an individual has



Association test is whether $\beta > 0$

We define genome-wide significance as .05/1 million effective tests = 5 x 10^{-8}

Figure 3. Genome-wide Association Findings in Rheumatoid Arthritis

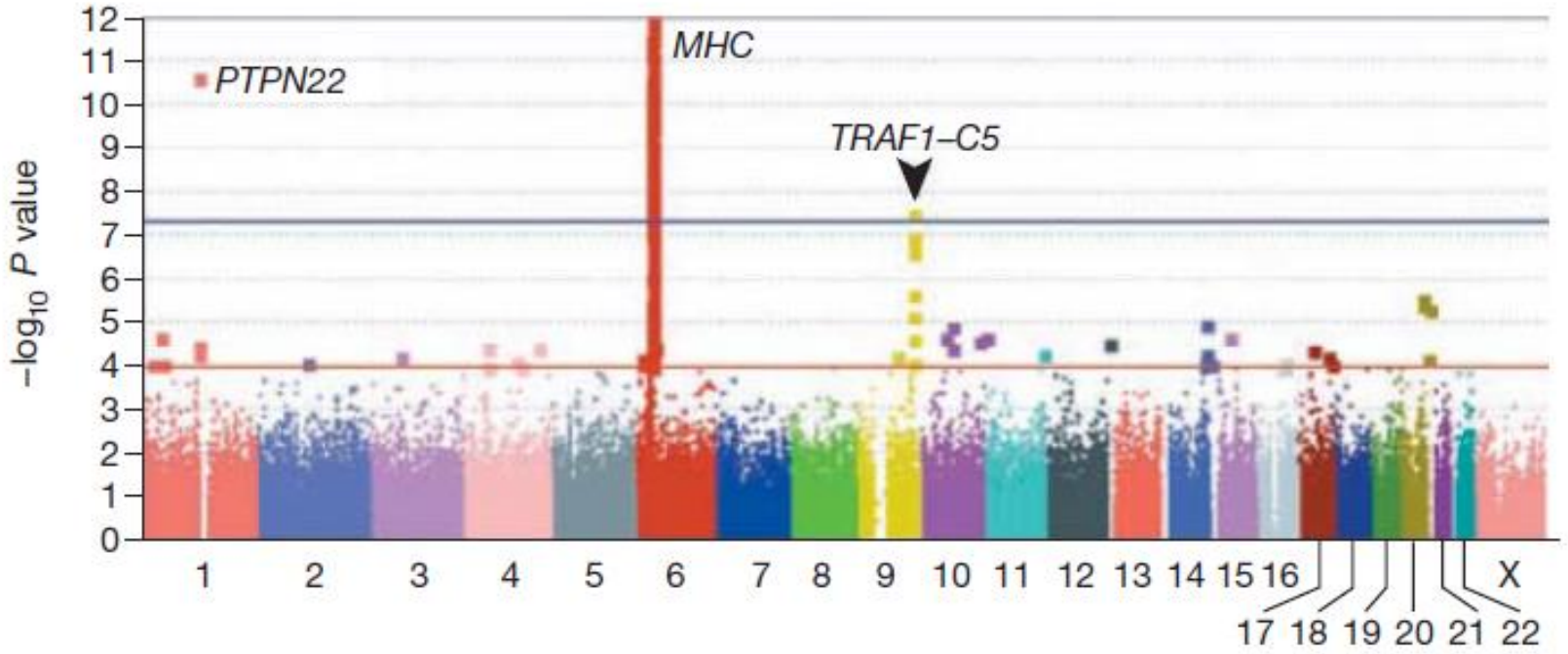
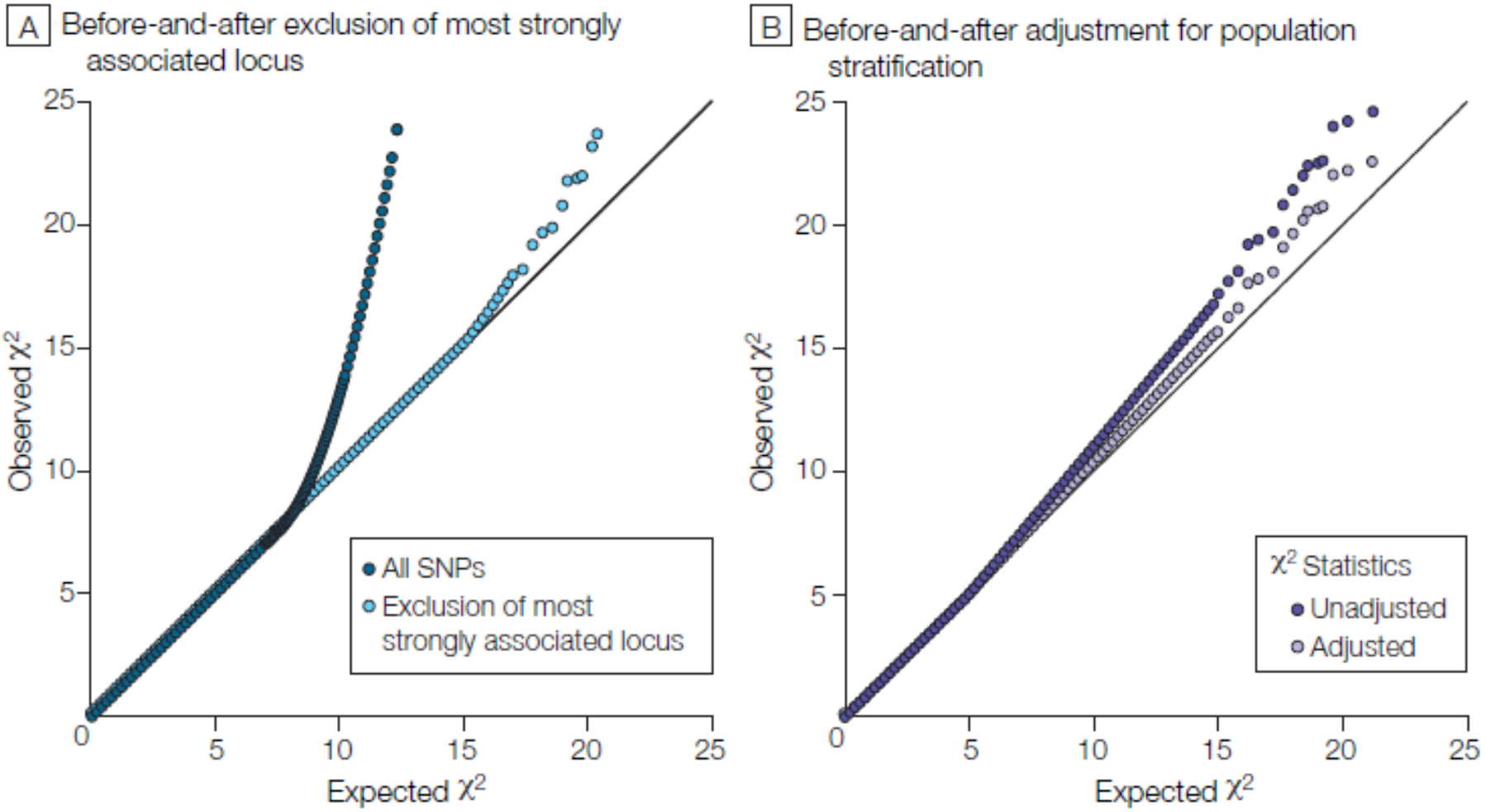


Figure 1. Hypothetical Quantile-Quantile Plots in Genome-wide Association Studies



Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways

To newly identify loci for age at natural menopause, we carried out a meta-analysis of 22 genome-wide association studies (GWAS) in 38,968 women of European descent, with replication in up to 14,435 women. In addition to four known loci, we identified 13 loci newly associated with age at natural menopause (at $P < 5 \times 10^{-8}$). Candidate genes located at these newly associated loci include genes implicated in DNA repair (*EXO1*, *HELQ*, *UIMC1*, *FAM175A*, *FANCI*, *TLK1*, *POLG* and *PRIM1*) and immune function (*IL11*, *NLRP11* and *PRRC2A* (also known as *BAT2*)). Gene-set enrichment pathway analyses using the full GWAS data set identified exoDNase, NF-κB signaling and mitochondrial dysfunction as biological processes related to timing of menopause.

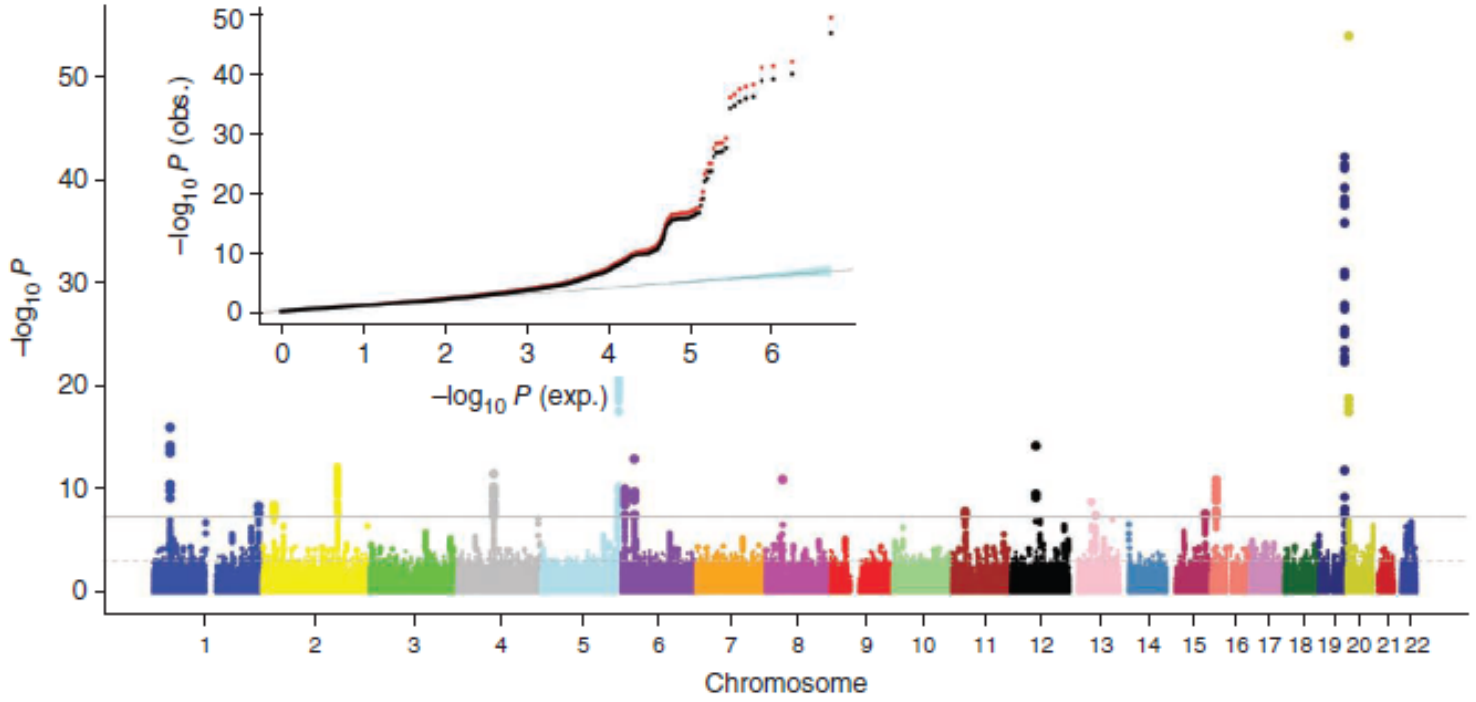


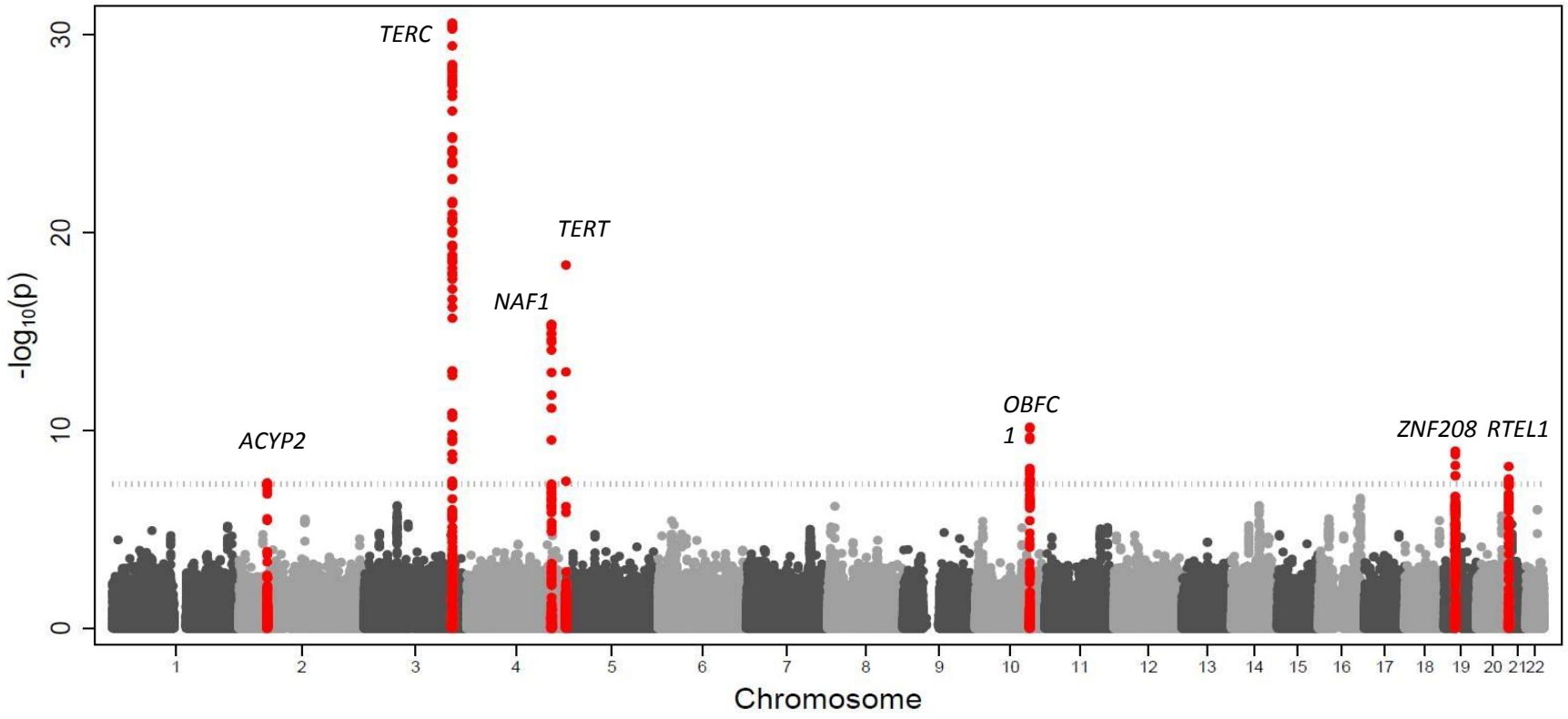
Figure 1 Discovery GWAS results. Manhattan plot of discovery meta-analysis. Inset, quantile-quantile plot of discovery primary analysis (red) and double genomic control–adjusted primary analysis (black). Obs., observed; exp., expected.

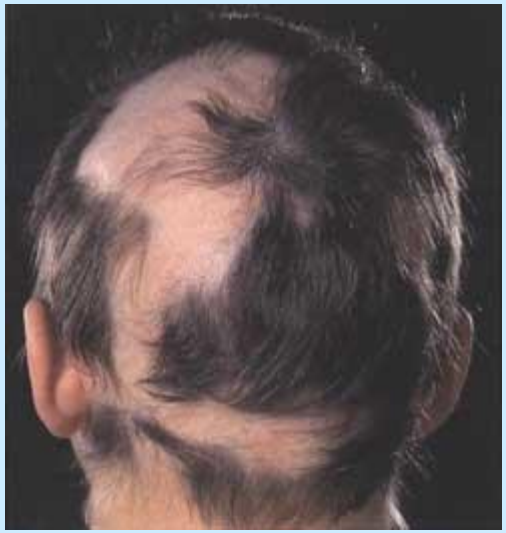
NATURE GENETICS VOLUME 44 | NUMBER 3 | MARCH 2012

Identification of seven loci affecting mean telomere length and their association with disease

Veryan Codd et al. (ENGAGE consortium) *NG*, 2013

Twin registries supplied 34% of samples





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

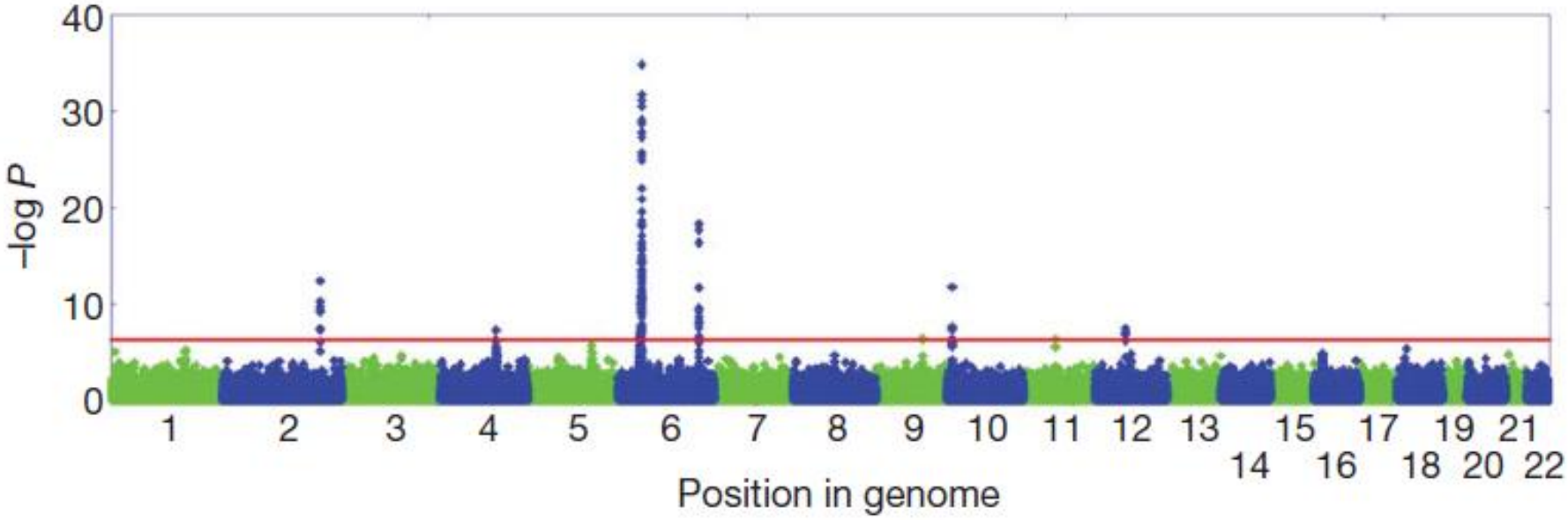
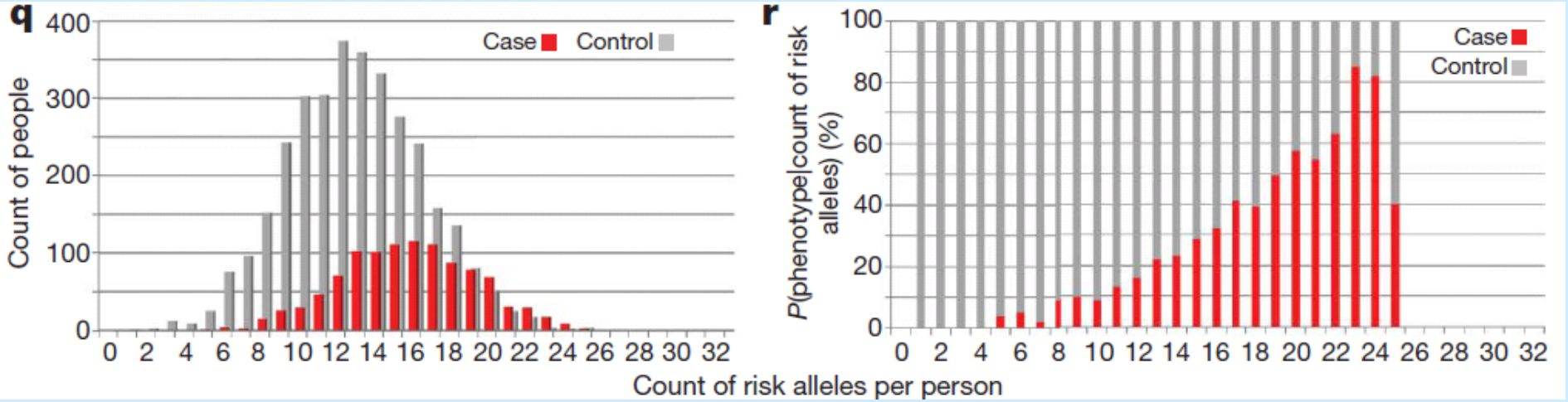


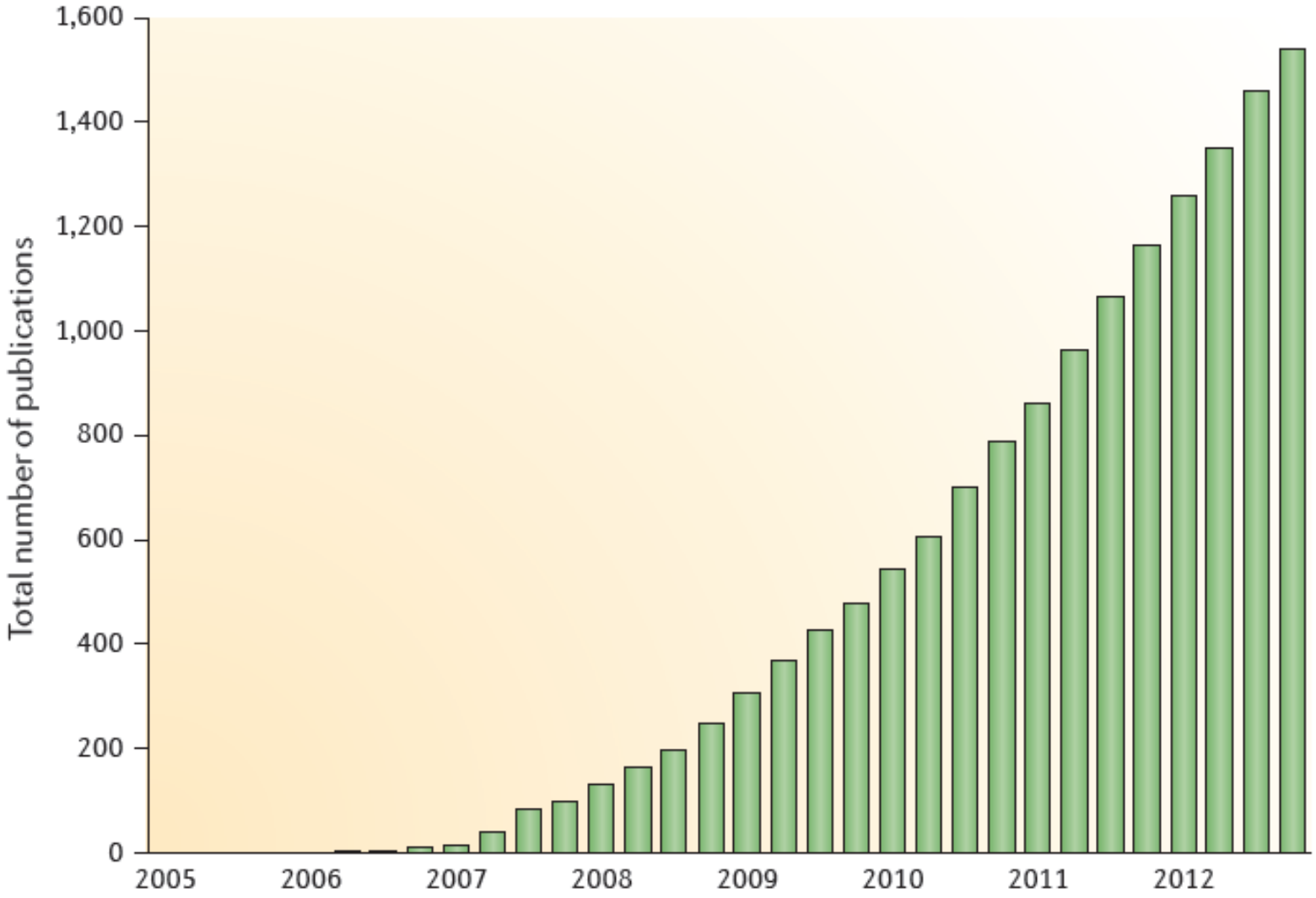
Table 1 | Genes with significant association to AA

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).

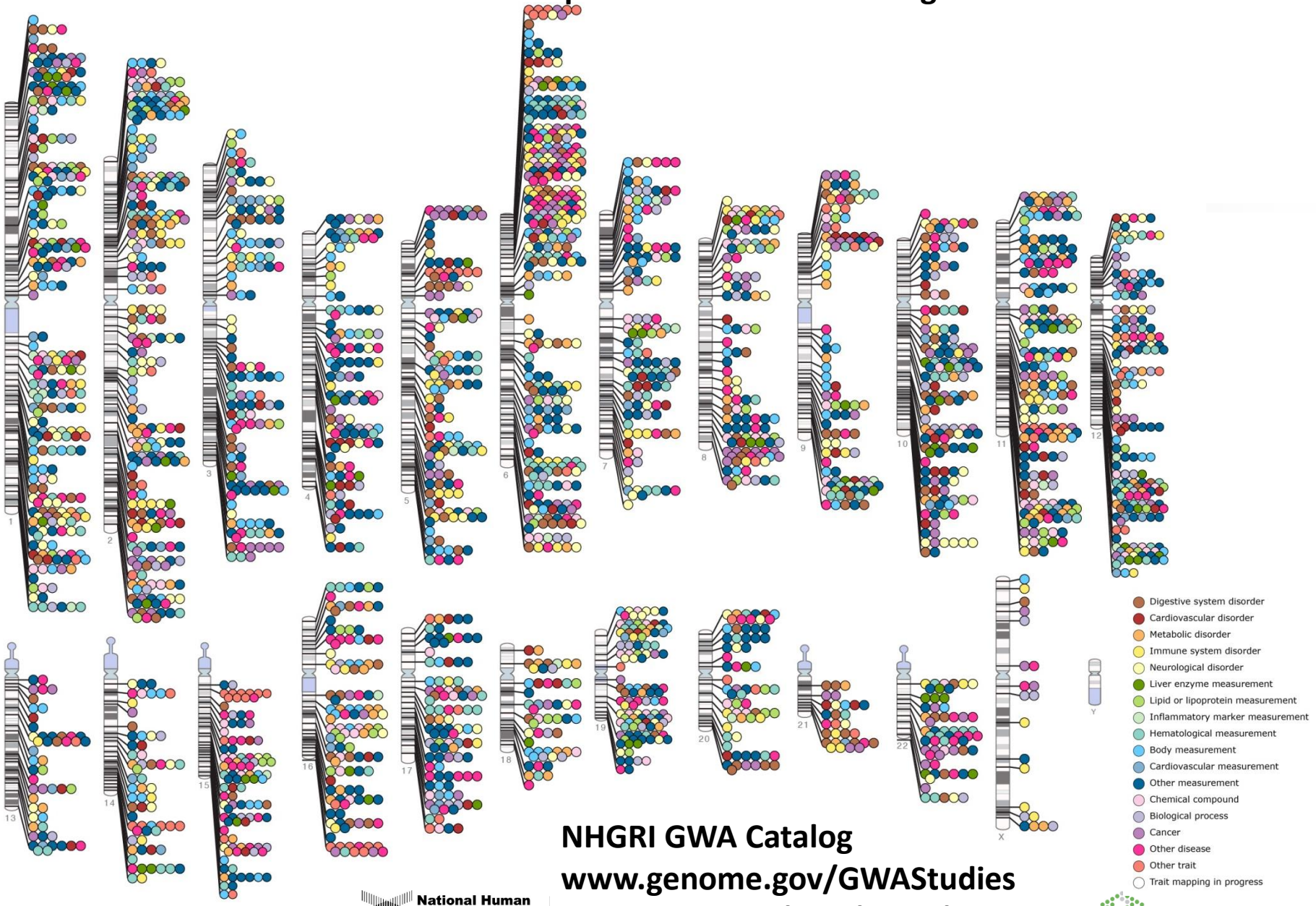


GWAS publications since 2005



Manolio, *Nature Reviews Genetics*, August 2013

Published Genome-Wide Associations through 07/2012
Published GWA at $p \leq 5X10^{-8}$ for 18 trait categories



NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

EMBL-EBI

Examples of Previously Unsuspected Associations between Certain Conditions and Genes and the Related Metabolic Function or Pathway, According to Genomewide Association Studies

Table 1. Examples of Previously Unsuspected Associations between Certain Conditions and Genes and the Related Metabolic Function or Pathway, According to Genomewide Association Studies.

Condition	Gene	Function or Pathway	Source of Data
Age-related macular degeneration	*CFH*	Complement-mediated inflammation	Klein et al.[25]
Coronary disease	*CDKN2A, CDKN2B*	Cell-cycle regulator	Helgadottir et al.[36]
Childhood asthma	*ORMDL3*	Unknown	Moffatt et al.[37]
Type 2 diabetes	*CDKAL1*	Cell-cycle regulator	Scott et al.[3]
Crohn's disease	*ATG16L1*	Autophagy	Rioux et al.[38]

Manolio T. N Engl J Med 2010;363:166-176

The NEW ENGLAND JOURNAL of MEDICINE

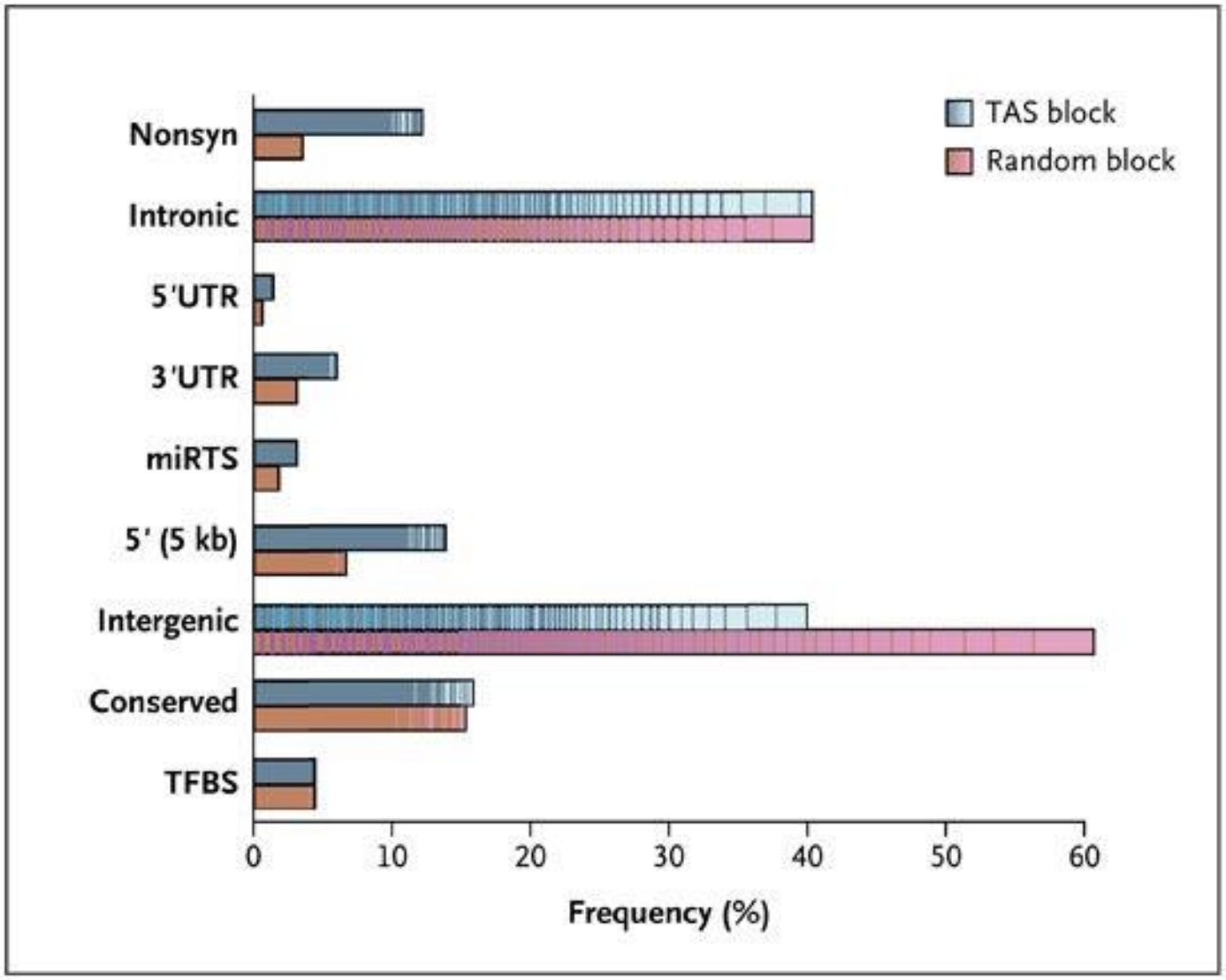
Examples of loci shared by conditions or traits previously thought to be unrelated, according to Genomewide Association Studies

Table 2. Examples of Loci Shared by Conditions or Traits Previously Thought to Be Unrelated, According to Genomewide Association Studies.

Gene	Conditions Sharing Associations	Source of Data
CDKN2A, CDKN2B	Coronary disease	Helgadottir et al.[36]
	Type 2 diabetes	Scott et al.[3]
	Invasive melanoma	Kamb et al.[43]
ORMDL3	Childhood asthma	Moffatt et al.[37]
	Crohn's disease	Barrett et al.[27]
CDKAL1	Type 2 diabetes	Scott et al.[3]
	Prostate cancer	Steinthorsdottir et al.[44]
LRRK2	Parkinson's disease	Paisán-Ruíz et al.[45]
	Crohn's disease	Barrett et al.[27]
KITLG	Testicular carcinoma	Rapley et al.[46]
	Blond or brown hair	Sulem et al.[47]
C10orf67	Sarcoidosis	Franke et al.[48]
	Celiac disease	Franke et al.[48]
JAZF1	Height	Johansson et al.[49]
	Type 2 diabetes	Zeggini et al.[50]
	Prostate cancer	Thomas et al.[17]

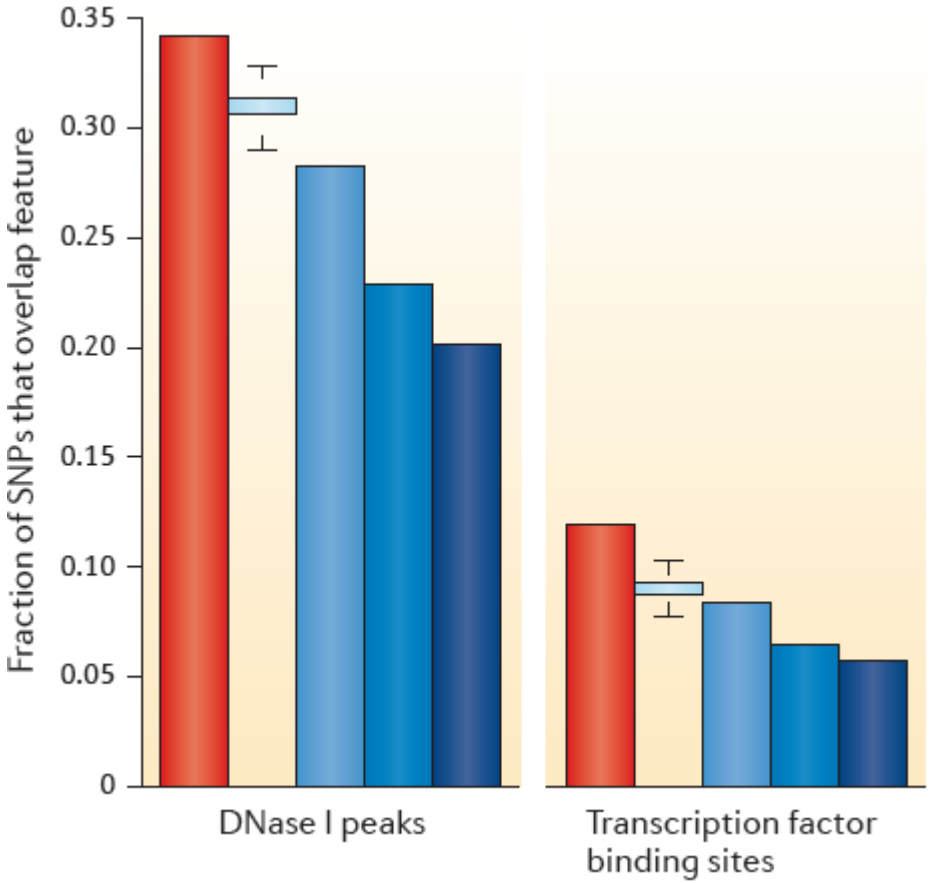
Manolio T. N Engl J Med 2010;363:166-176

Functional classifications of 465 Trait-Associated SNPs and the SNPs in Linkage Disequilibrium with them



Manolio T. N Engl J Med 2010;363:166-176

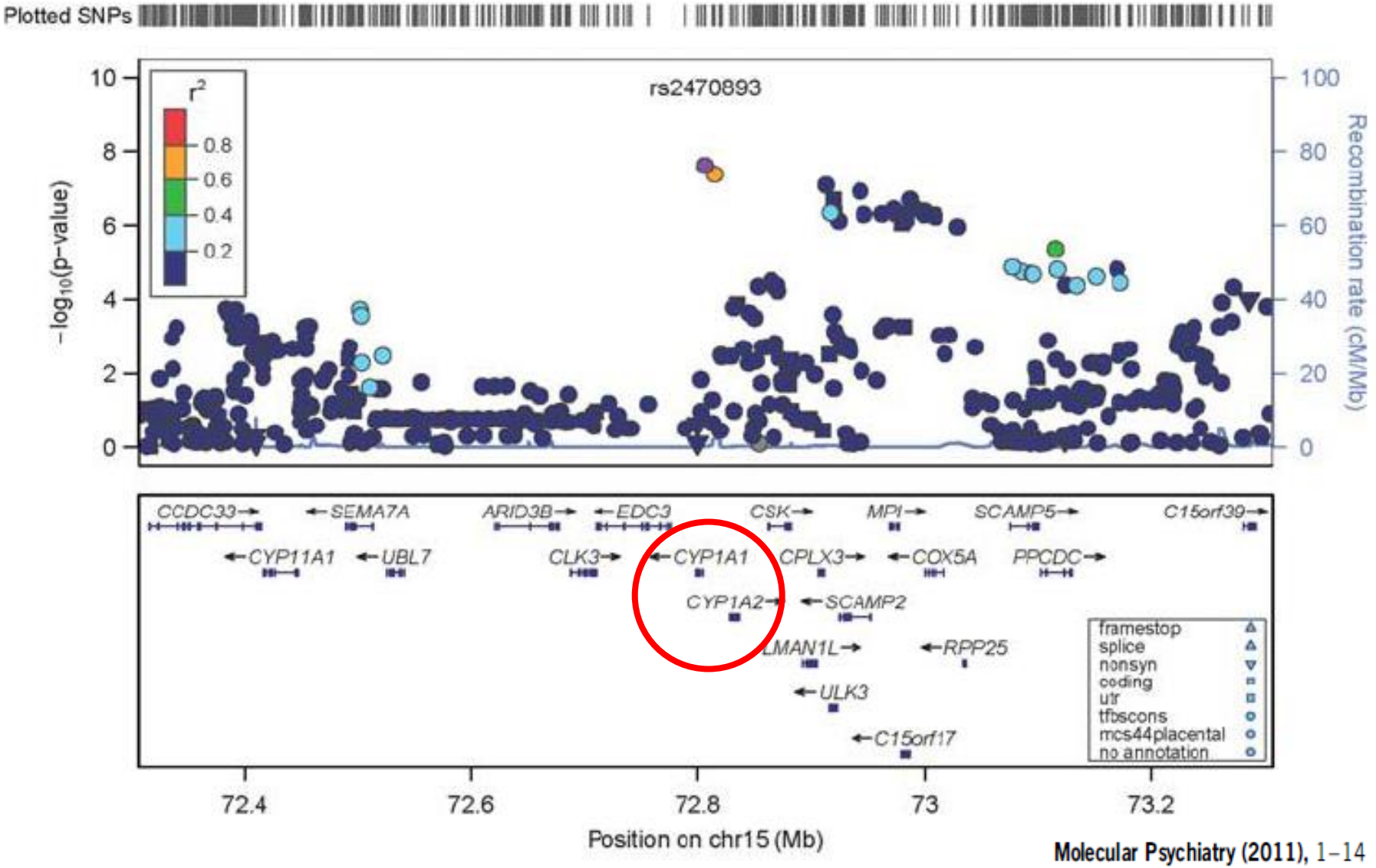
Correlations of presumed regulatory regions defined from GWAS



DNaseI peaks indicate regions of open chromatin accessible to the transcription apparatus and transcription factor binding sites where this this apparatus attached to the DNA

Manolio, Nature Reviews Genetics, August 2013

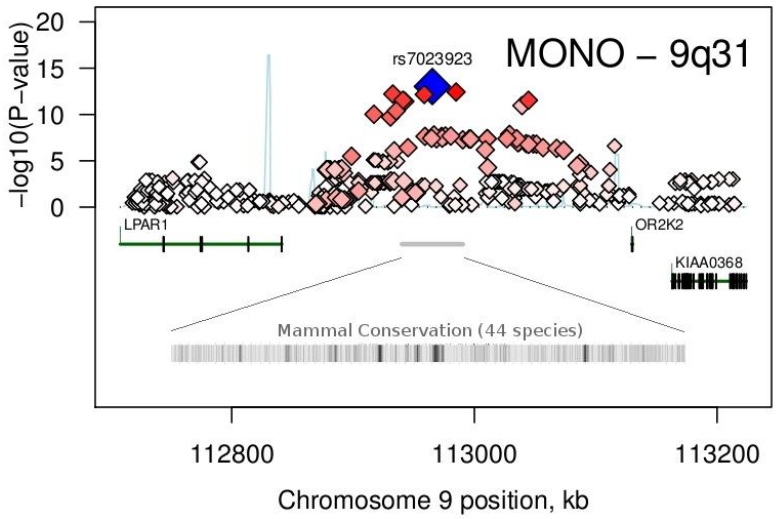
Genome-wide association analysis of coffee drinking suggests association with *CYP1A1/CYP1A2* and *NRCAM*

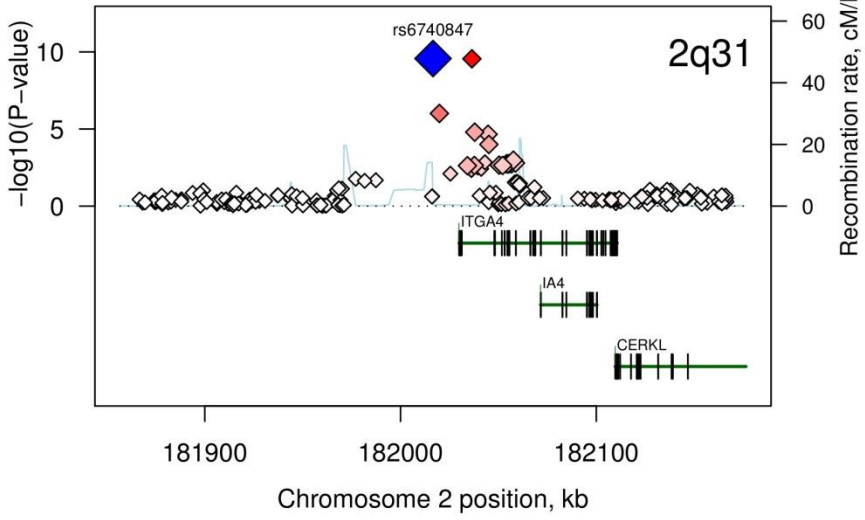


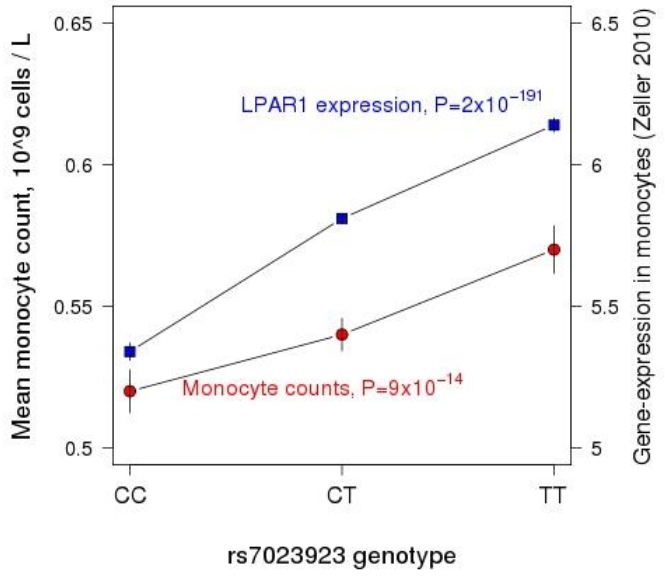
Molecular Psychiatry (2011), 1–14

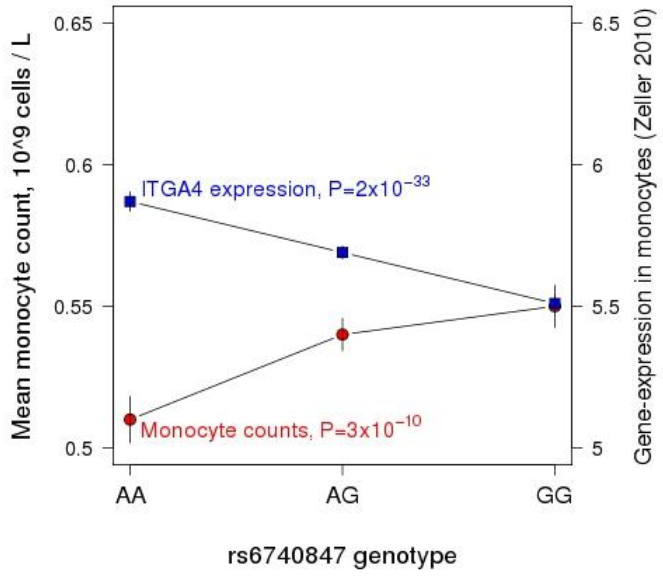
GWAS of monocyte counts – **help from expression data**

- Discovery N=4,225 (QIMR+NTR), replication N=1,517 (Busselton, GenomEUtwin)

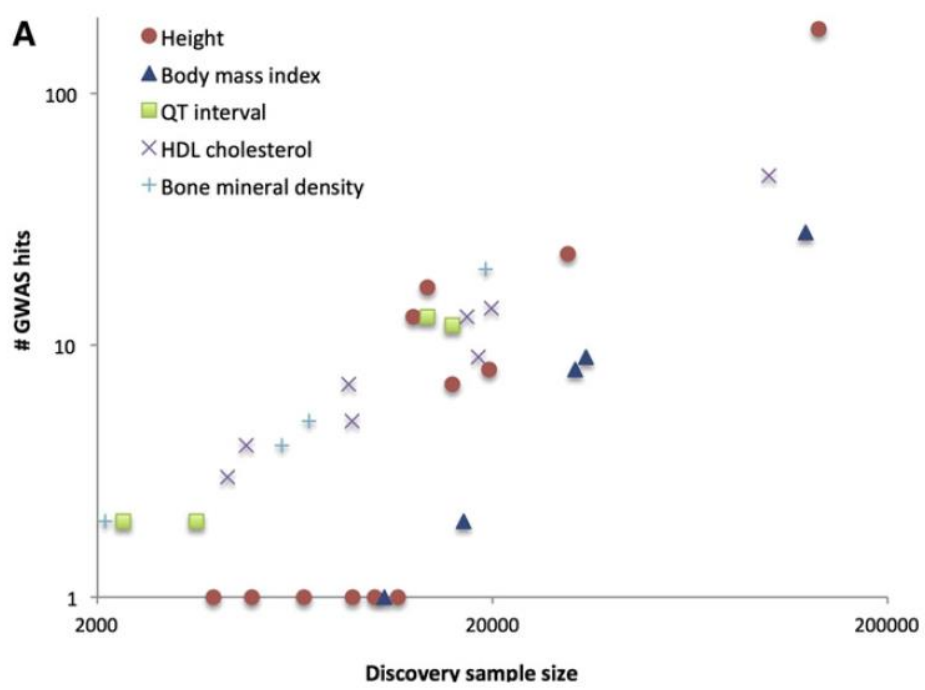


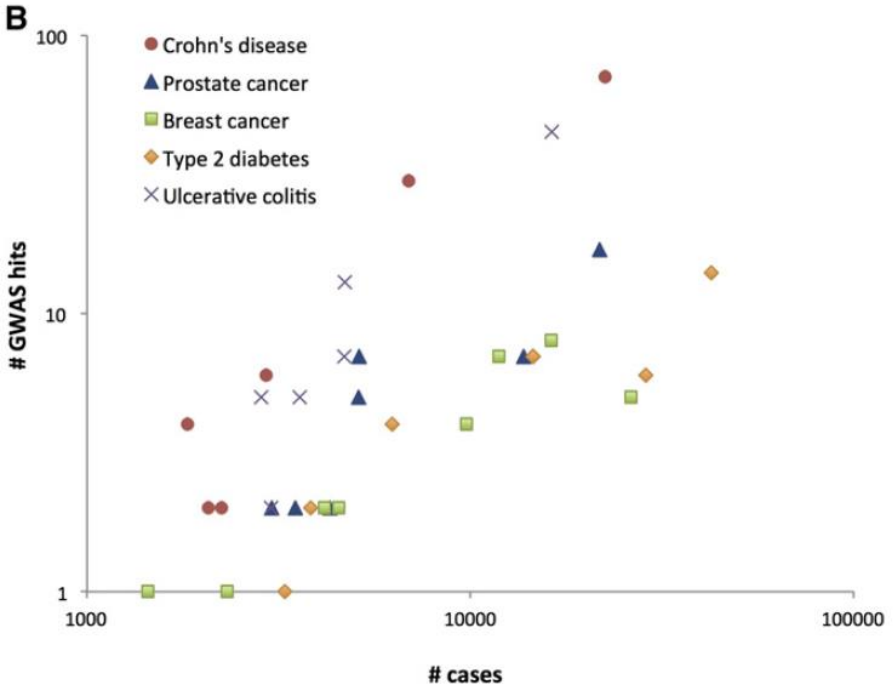






Ferreira et al. (2009) AJHG 85: 745; Zeller et al. (2010) PLoS One 5: e10693.

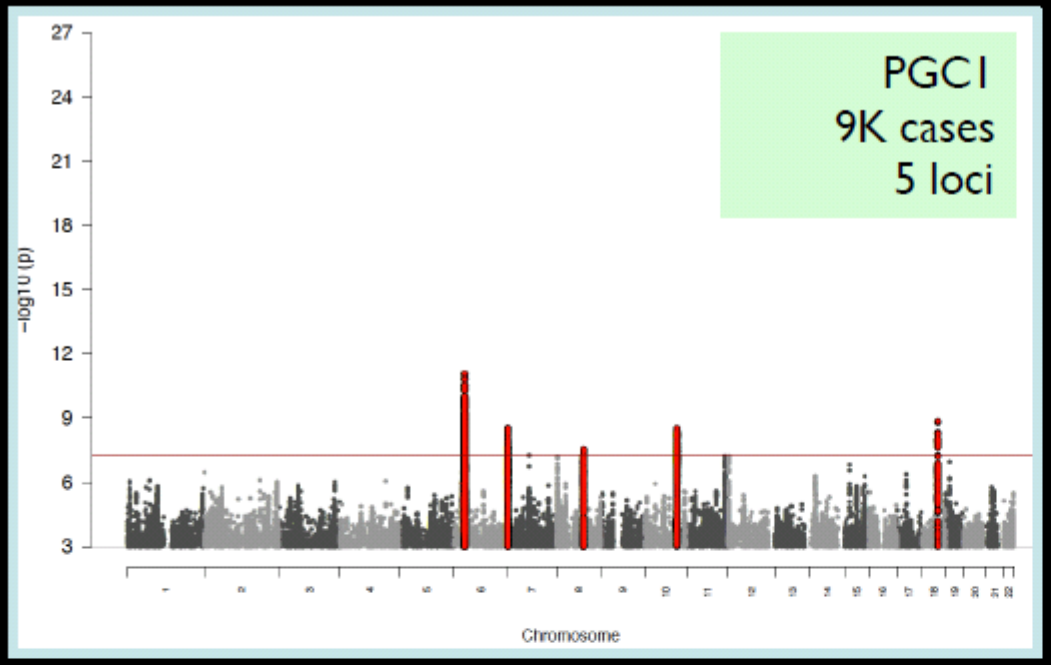




Selected quantitative traits **Selected diseases**

Number of Loci Identified is a Function of Sample Size

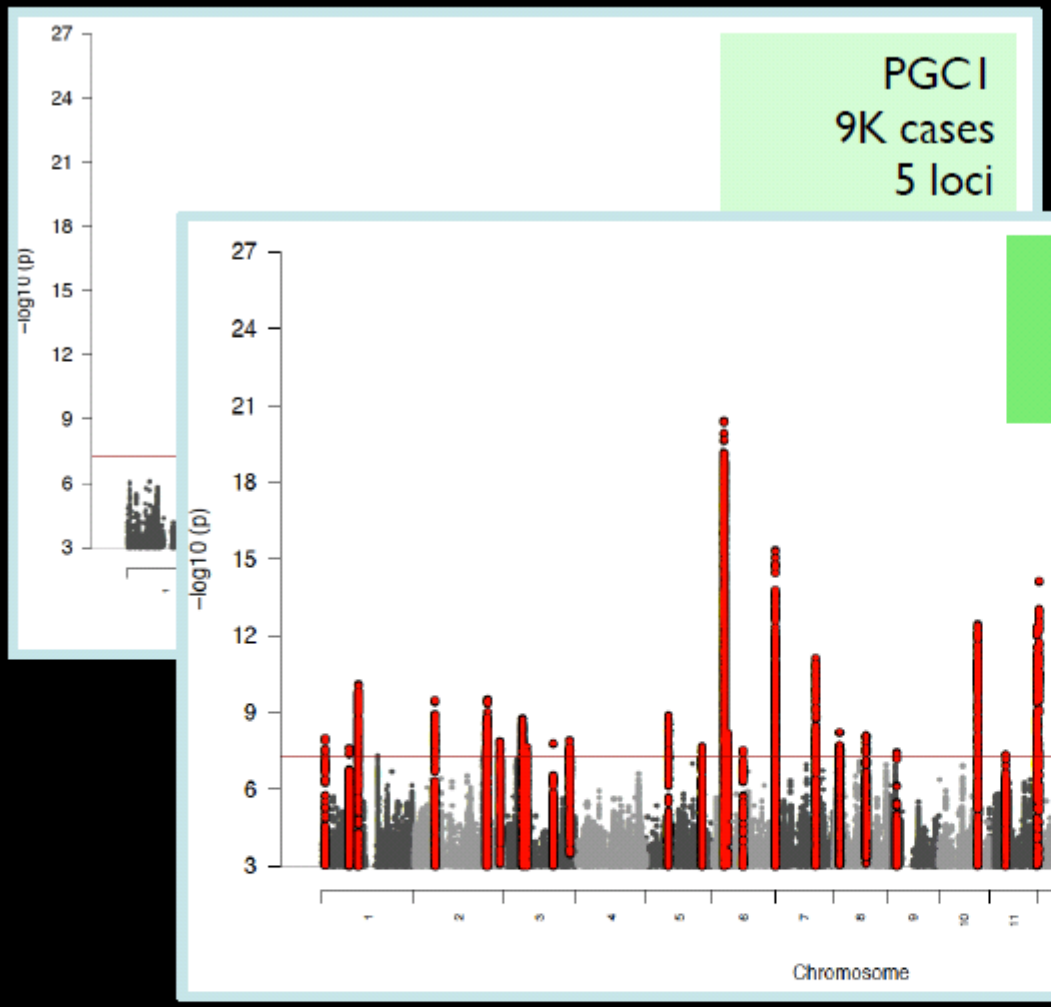
Visscher PM, et.al. (2012) Am J Hum Genetics



October 2011

Dramatic progress in GWAS for Schizophrenia

July 2012

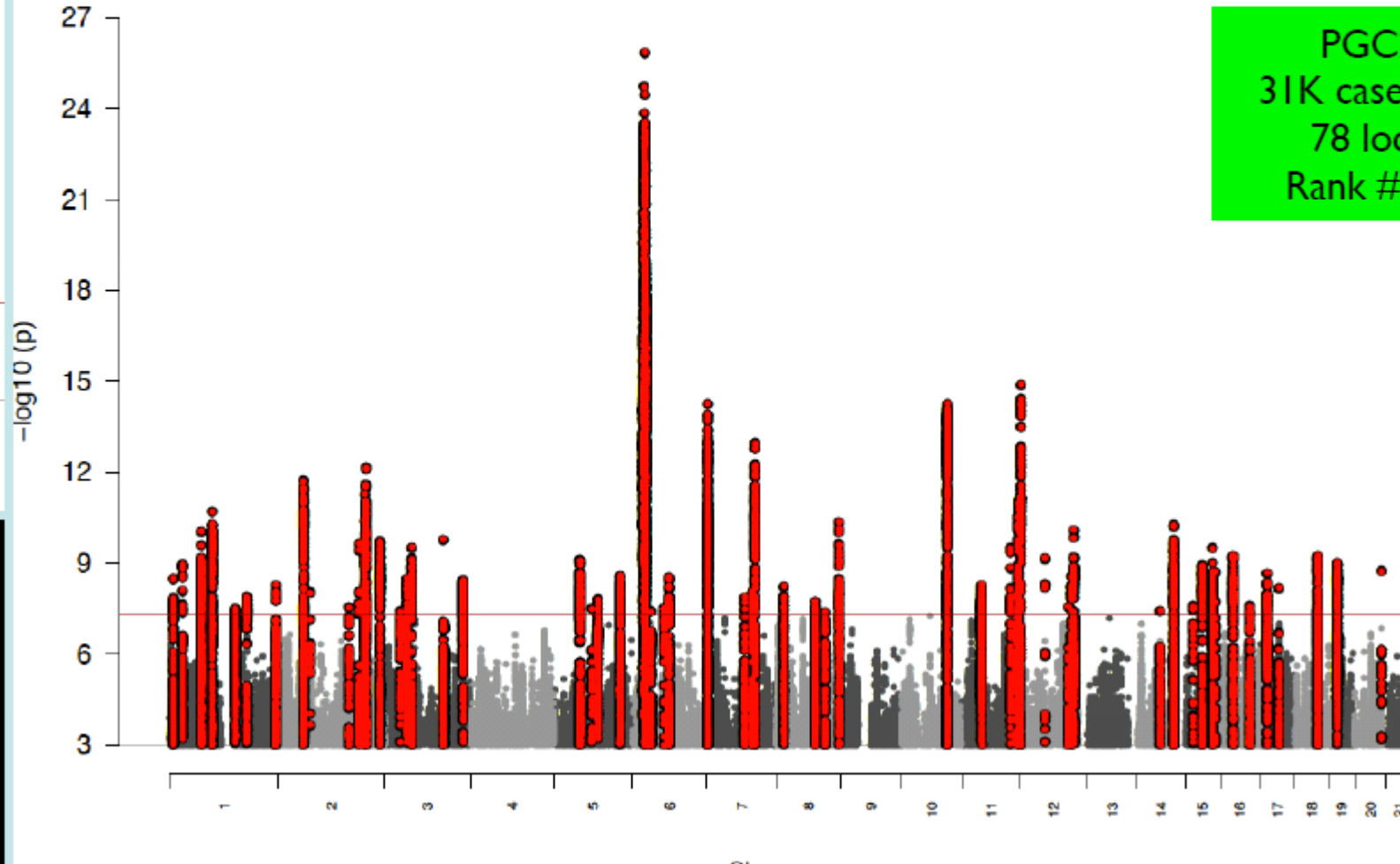

PGC1
9K cases
5 loci
PGC1 + Sweden
14K cases
22 loci
-log10 (p)
Chromosome

April 2013

PGC1
9K cases
5 loci

PGC1 + Sweden
14K cases
22 loci

PGC
31K case
78 loc
Rank #



Molecular Psychiatry (2012), 1–15

www.nature.com/mp

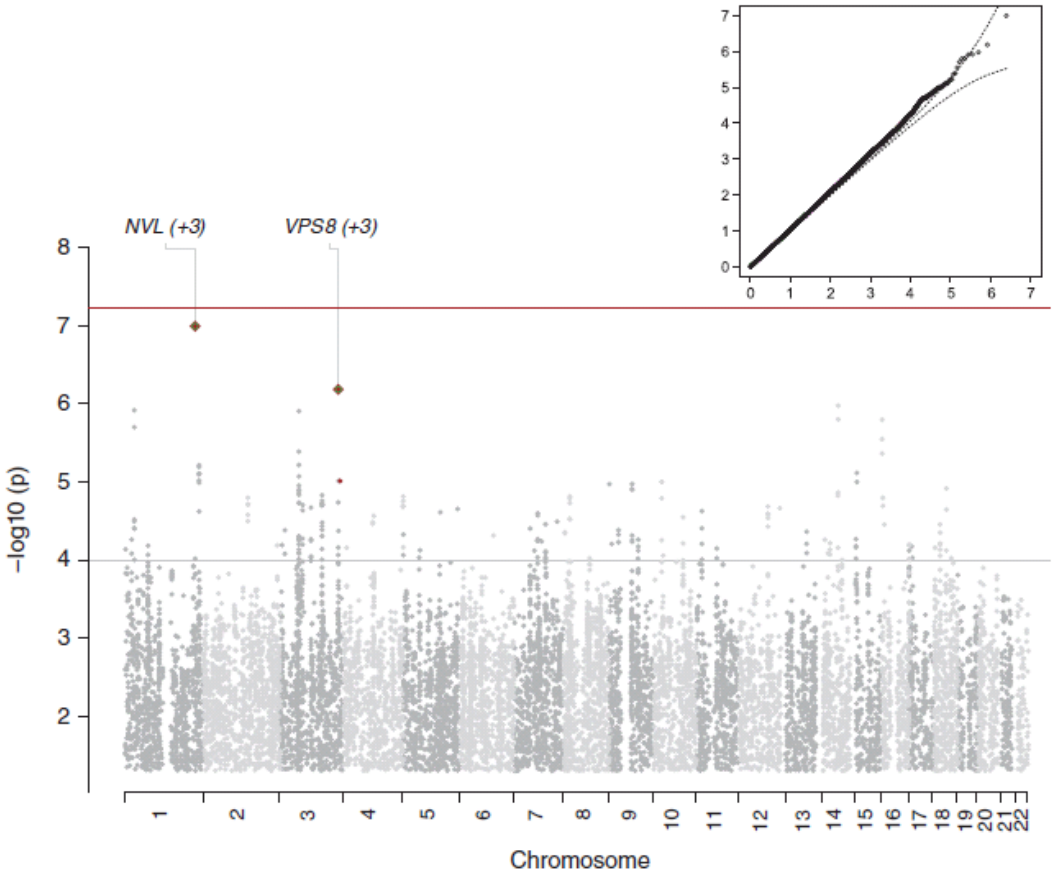
ORIGINAL ARTICLE

A mega-analysis of genome-wide association studies for major depressive disorder

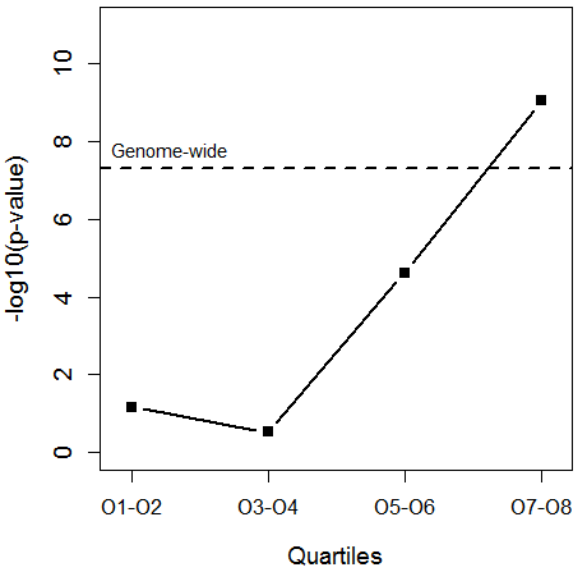
Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium[1]

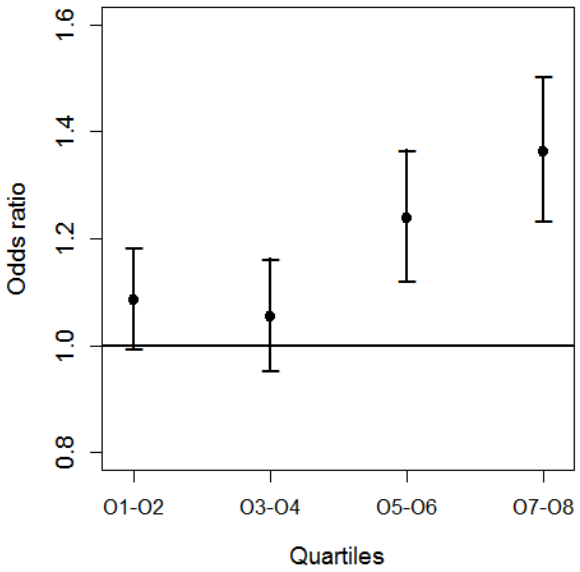
9240 MDD cases
9519 controls
….Nothing ☹

In the MDD-bipolar cross-disorder analysis, 15 SNPs exceeded GWS, and all were in a 248 kb interval of high LD on 3p21.1(rs2535629)

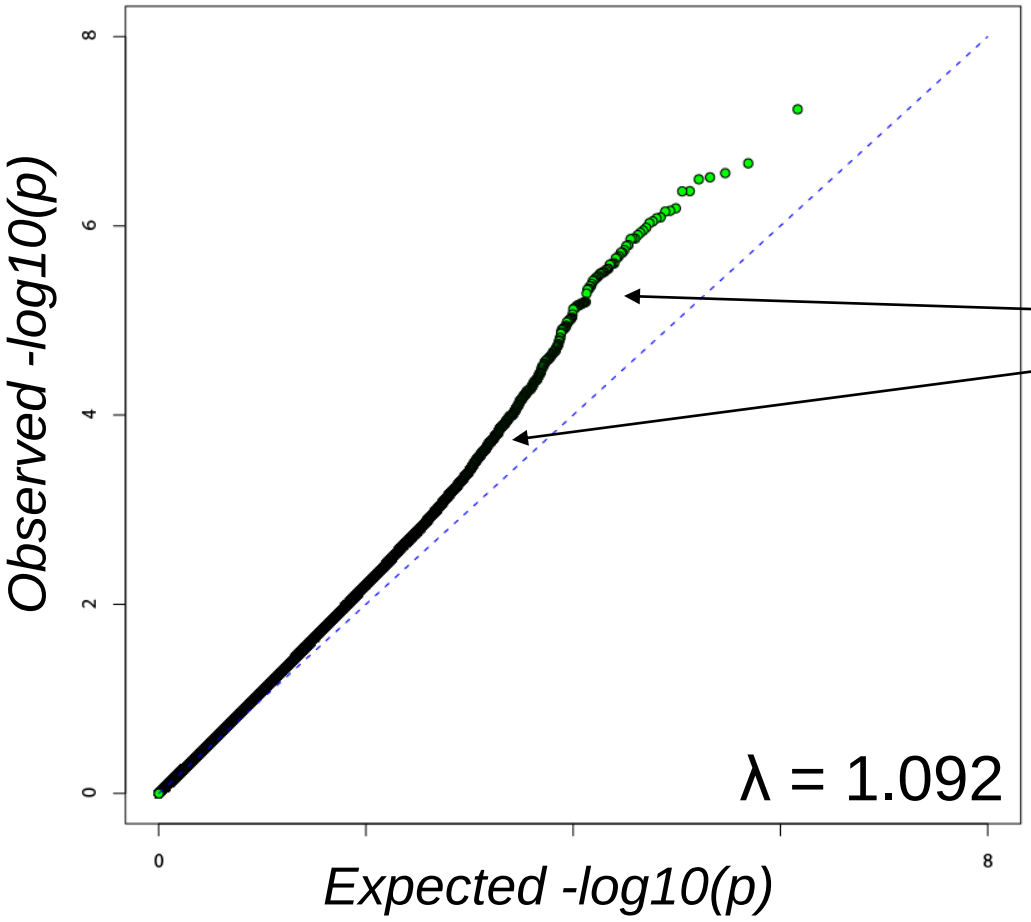


Significance and effect size for the top hit with cases split into non-overlapping quartiles by age-at-onset within their study





Schizophrenia (ISC) Q-Q plot



Consistent with:

Stratification?

Genotyping bias?

Distribution of true polygenic effects?

Genomic inflation factors under polygenic inheritance

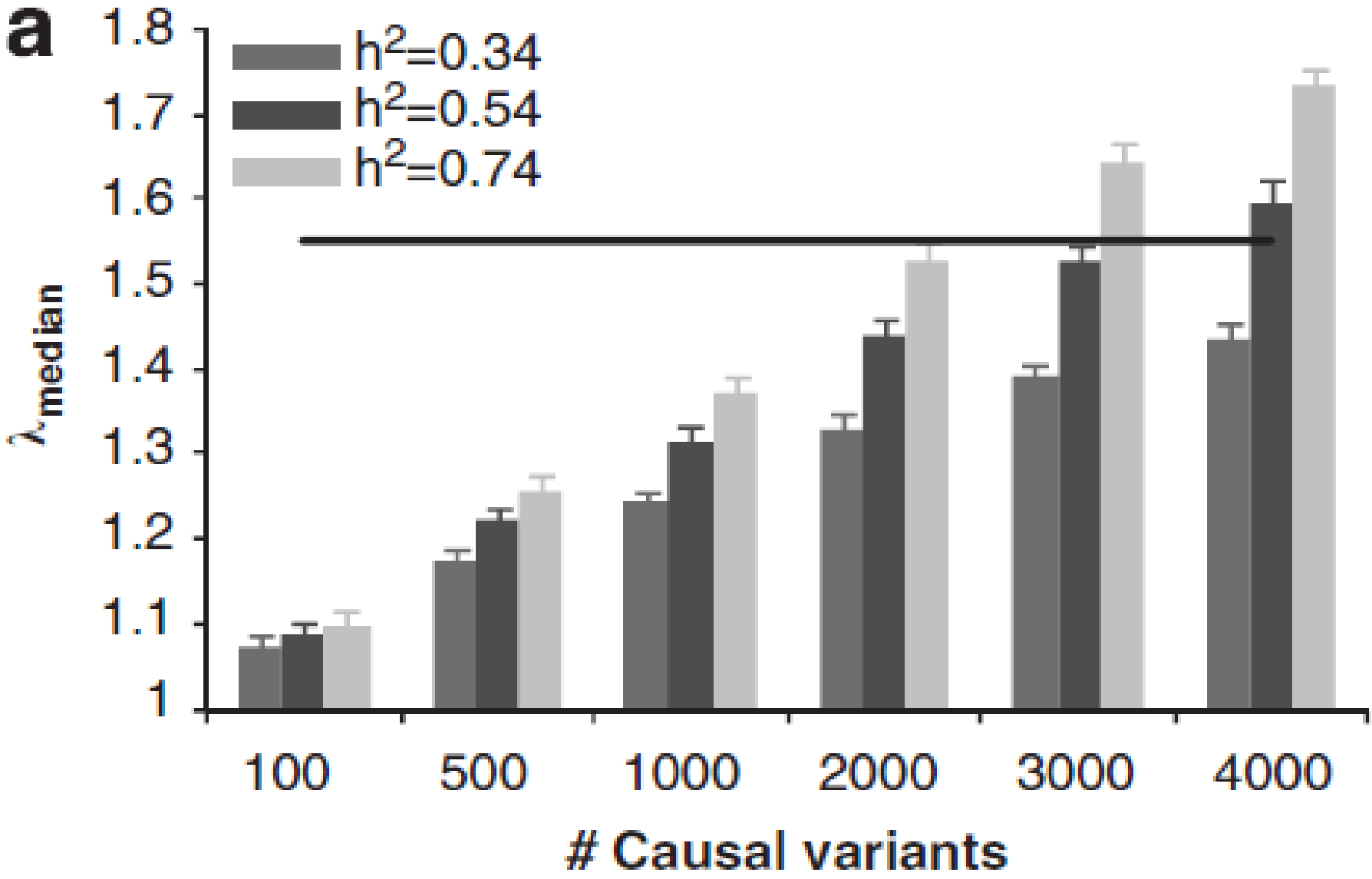
Jian Yang*[,1], Michael N Weedon[2], Shaun Purcell[3,4], Guillaume Lettre[5], Karol Estrada[6], Cristen J Willer[7], Albert V Smith[8], Erik Ingelsson[9], Jeffrey R O'Connell[10], Massimo Mangino[11], Reedik Mägi[12], Pamela A Madden[13], Andrew C Heath[13], Dale R Nyholt[1], Nicholas G Martin[1], Grant W Montgomery[1], Timothy M Frayling[2], Joel N Hirschhorn[3,14,15], Mark I McCarthy[12,16], Michael E Goddard[17], Peter M Visscher[1] and the GIANT Consortium

Population structure, including population stratification and cryptic relatedness, can cause spurious associations in genome-wide association studies (GWAS). Usually, the scaled median or mean test statistic for association calculated from multiple single-nucleotide-polymorphisms across the genome is used to assess such effects, and 'genomic control' can be applied subsequently to adjust test statistics at individual loci by a genomic inflation factor. Published GWAS have clearly shown that there are many loci underlying genetic variation for a wide range of complex diseases and traits, implying that a substantial proportion of the genome should show inflation of the test statistic. Here, we show by theory, simulation and analysis of data that in the absence of population structure and other technical artefacts, but in the presence of polygenic inheritance, substantial genomic inflation is expected. Its magnitude depends on sample size, heritability, linkage disequilibrium structure and the number of causal variants. Our predictions are consistent with empirical observations on height in independent samples of ~4000 and ~133 000 individuals.

$$\lambda_{\text{mean}}^{\text{QT}} \approx 1 + \frac{Nh^2\overline{r^2}\bar{s}}{n}$$

$$\lambda_{\text{mean}}^{\text{CC}} \approx 1 + \frac{Nh^2\overline{r^2}\bar{s}i^2v(1-v)}{n(1-K)^2}$$

European Journal of Human Genetics (2011) 19, 807–812



European Journal of Human Genetics (2011) 19, 807–812

Sciencexpress Reports

GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment

All authors with their affiliations appear at the end of this paper.

ured at an age at which subjects were very likely to have completed their education [over 95% of the sample was at least 30; (*5*)]. On average, subjects have 13.3 years of schooling, and 23.1% have a college degree. To enable pooling of GWAS results, all studies conducted analyses with data imputed to the HapMap 2 CEU (r22.b36) reference set. To guard against population stratification, the first four principal were nort- cific lled, us- sam- hree little f *p*- es of

30, 2013

- Finnish twin cohort
- Netherlands twin register
- QIMR (Australian twin register)
- Swedish twin register
- TwinsUK
- Minnesota Twin – family study
- **Twin registers supply 44,751 Ss (i.e. >35% of total sample size)**
- **There are 6 twin cohorts and total of 52 cohorts (11%)**

The value of DZ twins for within-pair association tests for ruling out population stratification

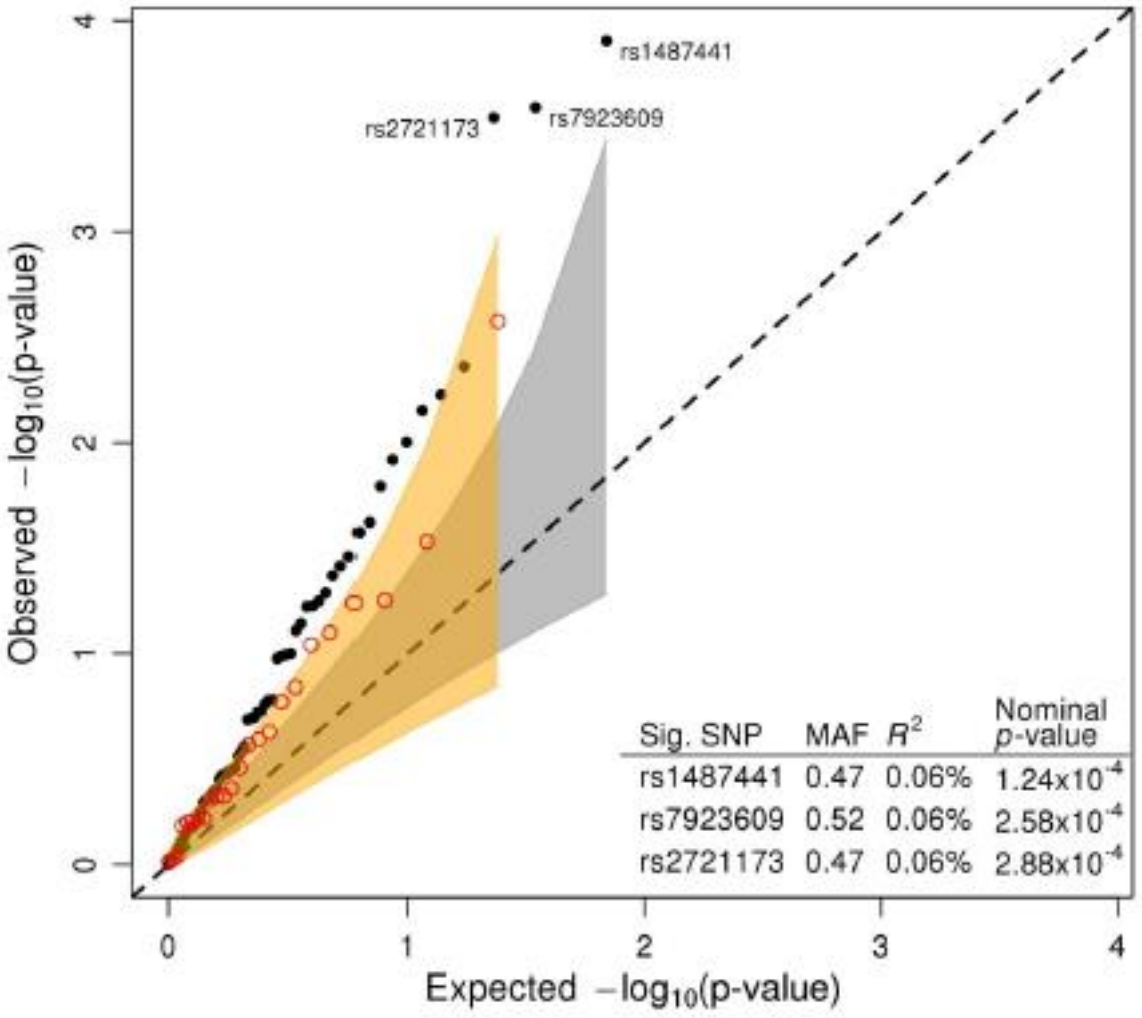
Within-family regression results of the polygenic scores on *College* and *EduYears* in the QIMR and Swedish Twin Registry cohorts using SNPs selected from the meta-analysis excluding the QIMR and STR cohorts.

		Prediction in QIMR + STR			
Phenotype (PGS)		$p_{SNPs} < 5\times10^{-8}$	$p_{SNPs} < 5\times10^{-5}$	$p_{SNPs} < 5\times10^{-3}$	All SNPs
EduYears (*College*)	R^2 (%)	0.017	0.003	0.220	0.310
	P	0.455	0.739	0.006	0.001
EduYears (*EduYears*)	R^2 (%)	0.002	0.001	0.110	0.190
	P	0.791	0.846	0.065	0.011

Analyses for QIMR are based on 572 full-sib pairs from independent 572 families. Analyses for STR are based on 2,774 DZ twins from 2,774 independent families.

Science. 2013 Jun 21;340:1467-71

Education SNPs predict IQ



Koellinger, submitted

Genetic variants associated with breast size also influence breast cancer risk

GWAS of Bra cup size on 16,000 women (23andMe)

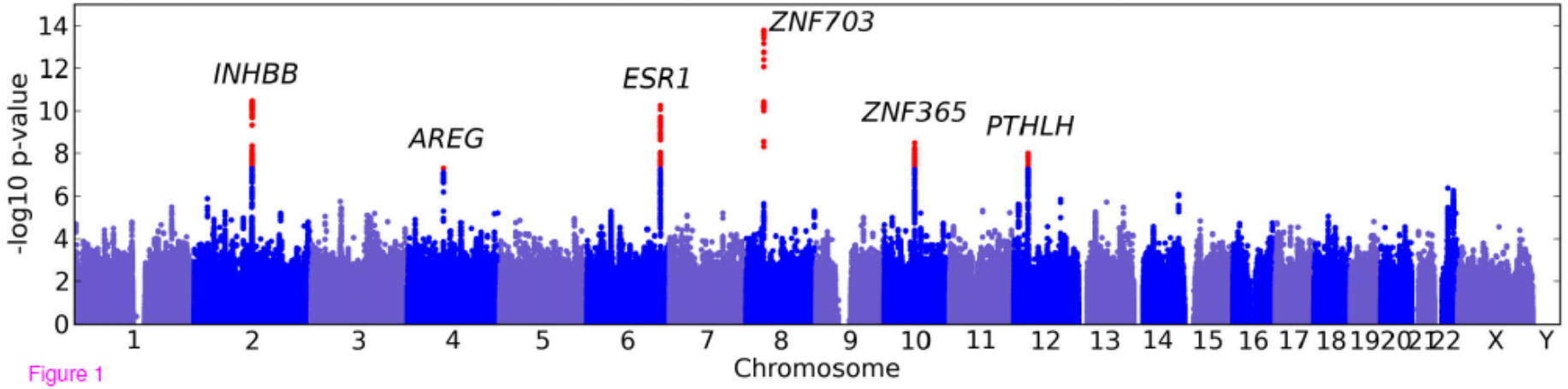
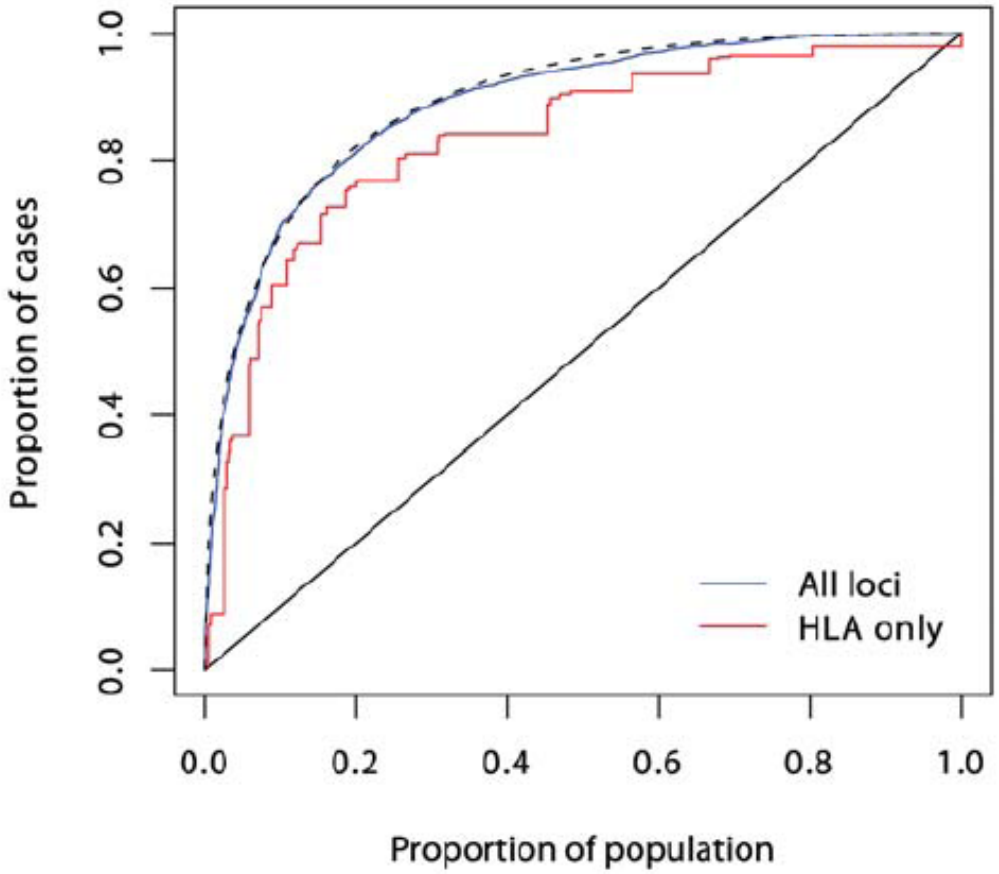


Figure 1

BMC Medical Genetics 2012, 13:53

How much variance have GWAS studies explained?

GWAS' greatest success: T1D



Current known loci explain a λ_s of just under five, as compared with the value of 15 often quoted. However, it is likely that the latter figure is exaggerated, and the λ_s attributable to inheritance is likely to be less than ten. The heritability explained will be increased to some degree when the known regions are more fully studied, but the bulk of the remaining heritability is likely to be attributable to many small (or rare) effects, most of which are unlikely to be mapped. Thus, even for this highly heritable disease, the prediction achievable could fall some way short of that required for a targeted prevention strategy.

Figure 5. ROC curve prediction from all the SNPs listed in Supplementary Table 1 in Text S1 (in blue). The prediction curve using the six MHC SNPs alone is shown in red, and the dashed curve corresponds to a polygenic multiplicative model with $\lambda_s = 4.75$.

PLoS GENETICS
OPEN ACCESS Freely available online
Viewpoints

Prediction and Interaction in Complex Disease Genetics: Experience in Type 1 Diabetes

David G. Clayton*

Variance explained by GWAS for selected complex traits

Trait or Disease	h^2 Pedigree Studies	h^2 GWAS Hits[a]	h^2 All GWAS SNPs[b]
Type 1 diabetes	0.9[98]	0.6[99,c]	0.3[12]
Type 2 diabetes	0.3–0.6[100]	0.05-0.10[34]	
Obesity (BMI)	0.4–0.6[101,102]	0.01-0.02[36]	0.2[14]
Crohn's disease	0.6–0.8[103]	0.1[11]	0.4[12]
Ulcerative colitis	0.5[103]	0.05[12]	
Multiple sclerosis	0.3–0.8[104]	0.1[45]	
Ankylosing spondylitis	>0.90[105]	0.2[106]	
Rheumatoid arthritis	0.6[107]		
Schizophrenia	0.7–0.8[108]	0.01[79]	0.3[109]
Bipolar disorder	0.6–0.7[108]	0.02[79]	0.4[12]
Breast cancer	0.3[110]	0.08[111]	
Von Willebrand factor	0.66–0.75[112,113]	0.13[114]	0.25[14]
Height	0.8[115,116]	0.1[13]	0.5[13,14]
Bone mineral density	0.6-0.8[117]	0.05[118]	
QT interval	0.37–0.60[119,120]	0.07[121]	0.2[14]
HDL cholesterol	0.5[122]	0.1[57]	
Platelet count	0.8[123]	0.05–0.1[58]	

Visscher PM, et.al. (2012) Am J Hum Genetics

Possible explanations for missing heritability

(not mutually exclusive, but in order of increasing plausibility ?)

- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

Possible explanations for missing heritability

(in order of increasing plausibility ?)

- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

Non-additive variance?

OPEN ACCESS Freely available online

PLoS GENETICS

Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits

William G. Hill[1]*, Michael E. Goddard[2,3], Peter M. Visscher[4]

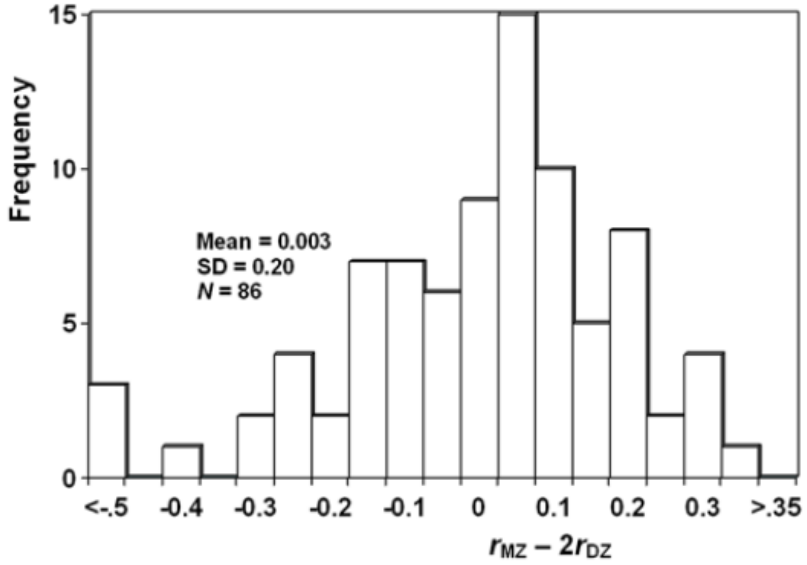
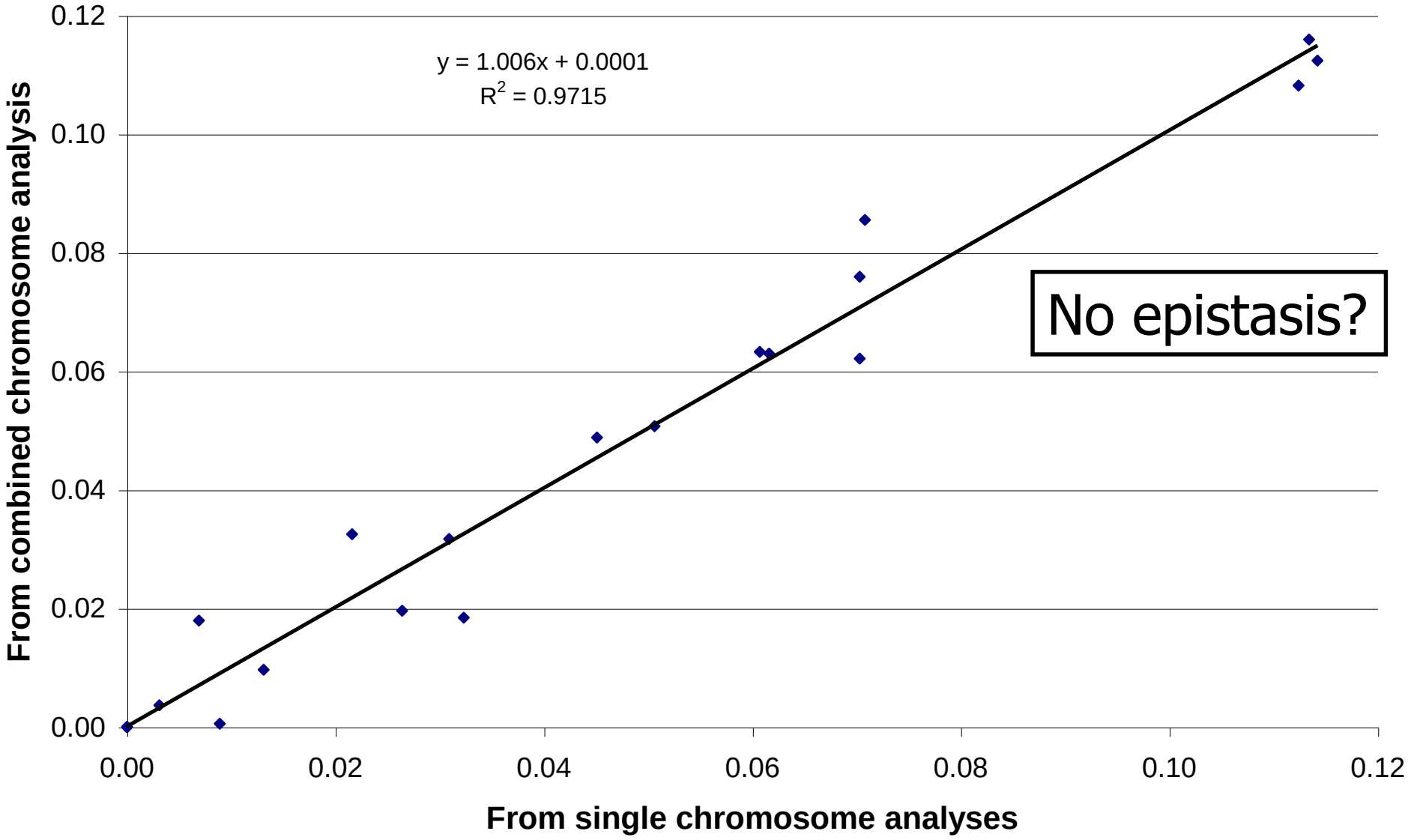


Figure 1. Distribution of $r_{MZ} - 2r_{DZ}$ for all traits on human twins.

Estimates of chromosomal heritabilities for height



EVIDENCE FOR POLYGENIC EPISTATIC INTERACTIONS IN MAN?

A. C. HEATH, N. G. MARTIN, L. J. EAVES AND D. LOESCH

Observed familial correlations for finger pattern intensity and the expected contributions of the main components of gene action

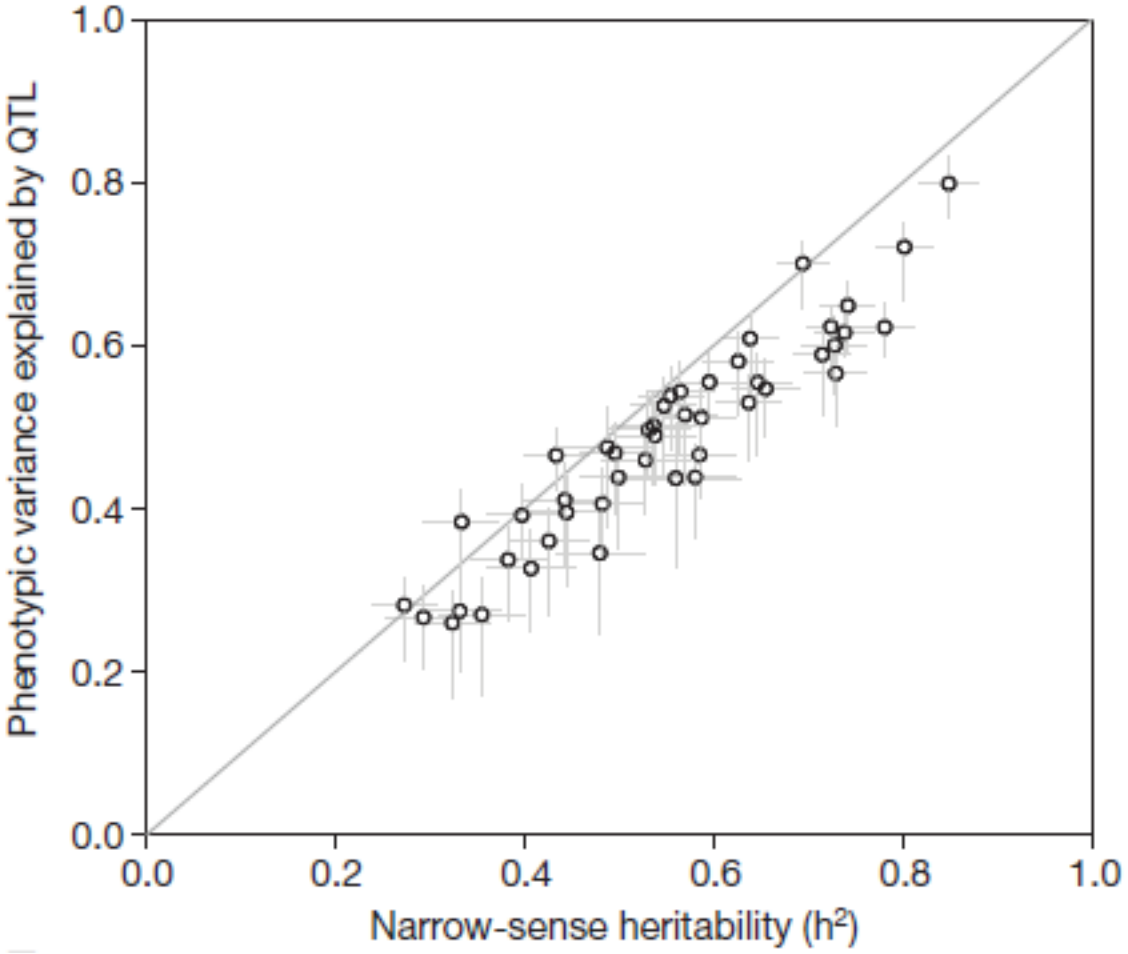
Relationship	*N*	*r*	Genetic contribution		
			VA	VD	VAA
MZ male twins	60	0.91	1	1	1
MZ female twins	50	0.90	1	1	1
DZ male twins	62	0.24	½	¼	¼
DZ female twins	49	0.36	½	¼	¼
Male siblings	461	0.40	½	¼	¼
Female siblings	309	0.33	½	¼	¼
Opposite-sex siblings	857	0.33	½	¼	¼
Father-son	469	0.33	½	0	¼
Father-daughter	547	0.40	½	0	¼
Mother-son	460	0.41	½	0	¼
Mother-daughter	540	0.31	½	0	¼
Spouses	281	0.04	0	0	0

Abbreviations used are: VA, additive genetic variance; VD, dominance variance; VAA, epistatic variance arising from interactions of additive effects of genes.

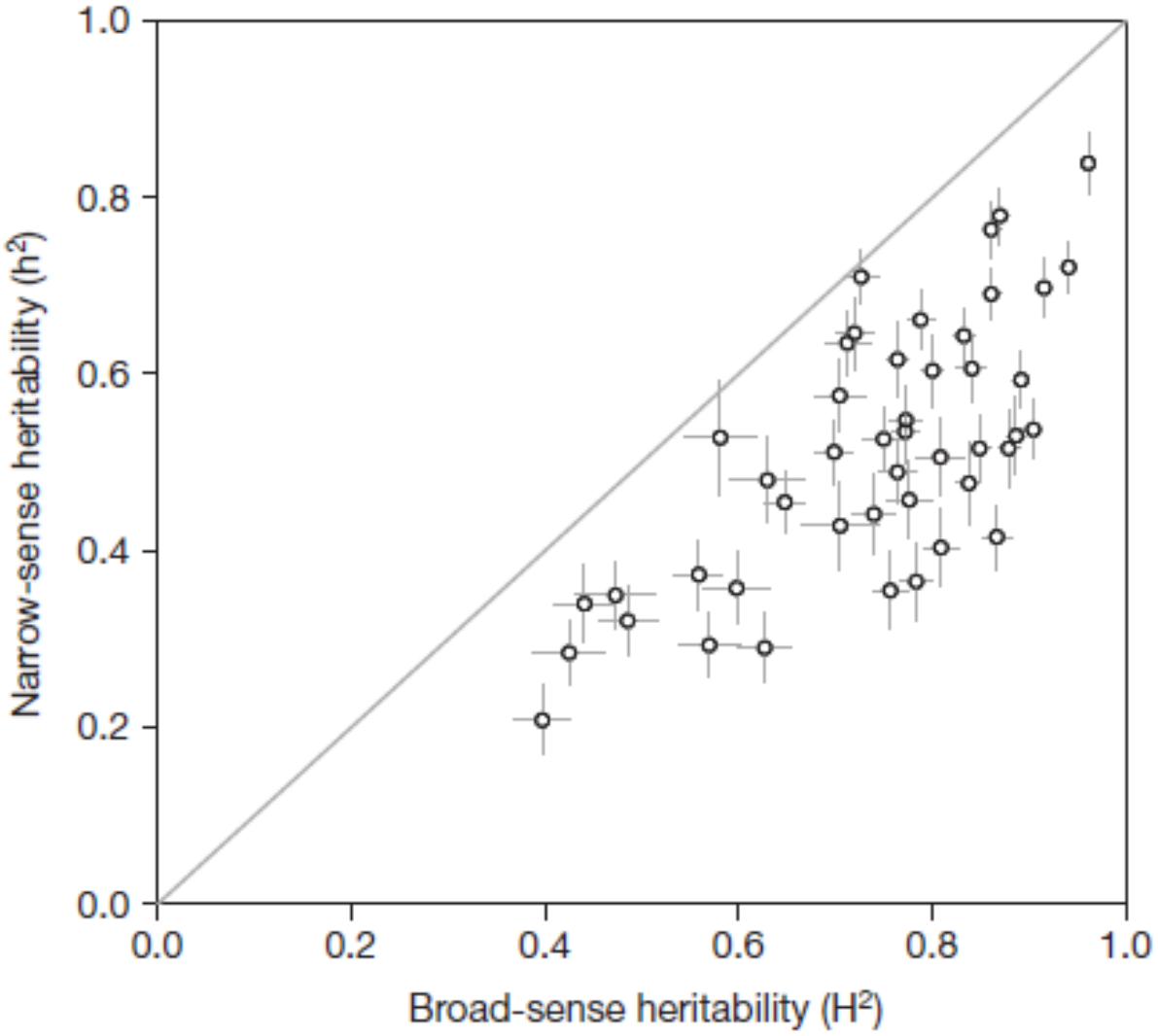
Genetics 106: 719-727,1984

Finding the sources of missing heritability in a yeast cross

Joshua S. Bloom[1,2], Ian M. Ehrenreich[1,3], Wesley T. Loo[1,2], Thúy-Lan Võ Lite[1,2] & Leonid Kruglyak[1,4,5]



Contribution to heritability of gene–gene interactions varies among traits, from ~0 to ~50%

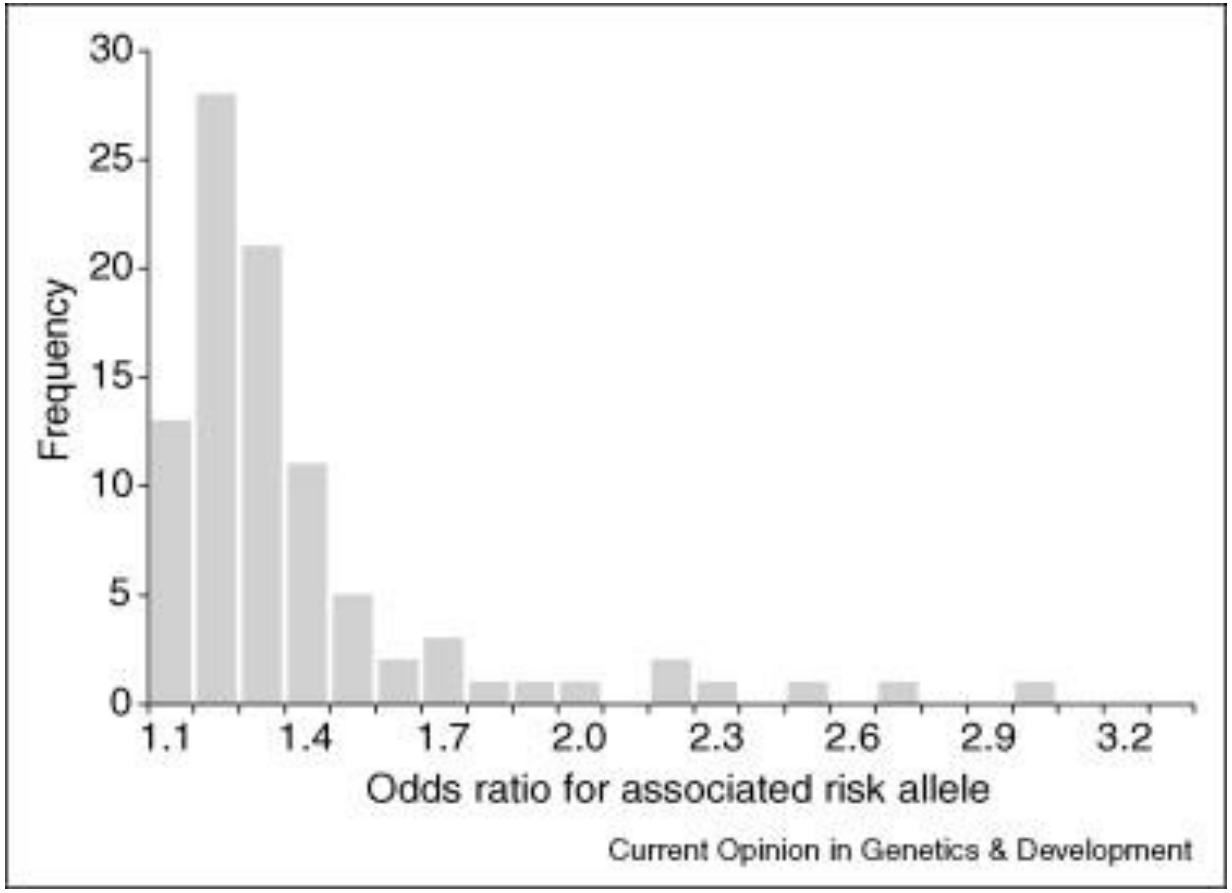


Possible explanations for missing heritability

(in order of increasing plausibility ?)

- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

Effects sizes of validated variants from 1st 16 GWAS studies



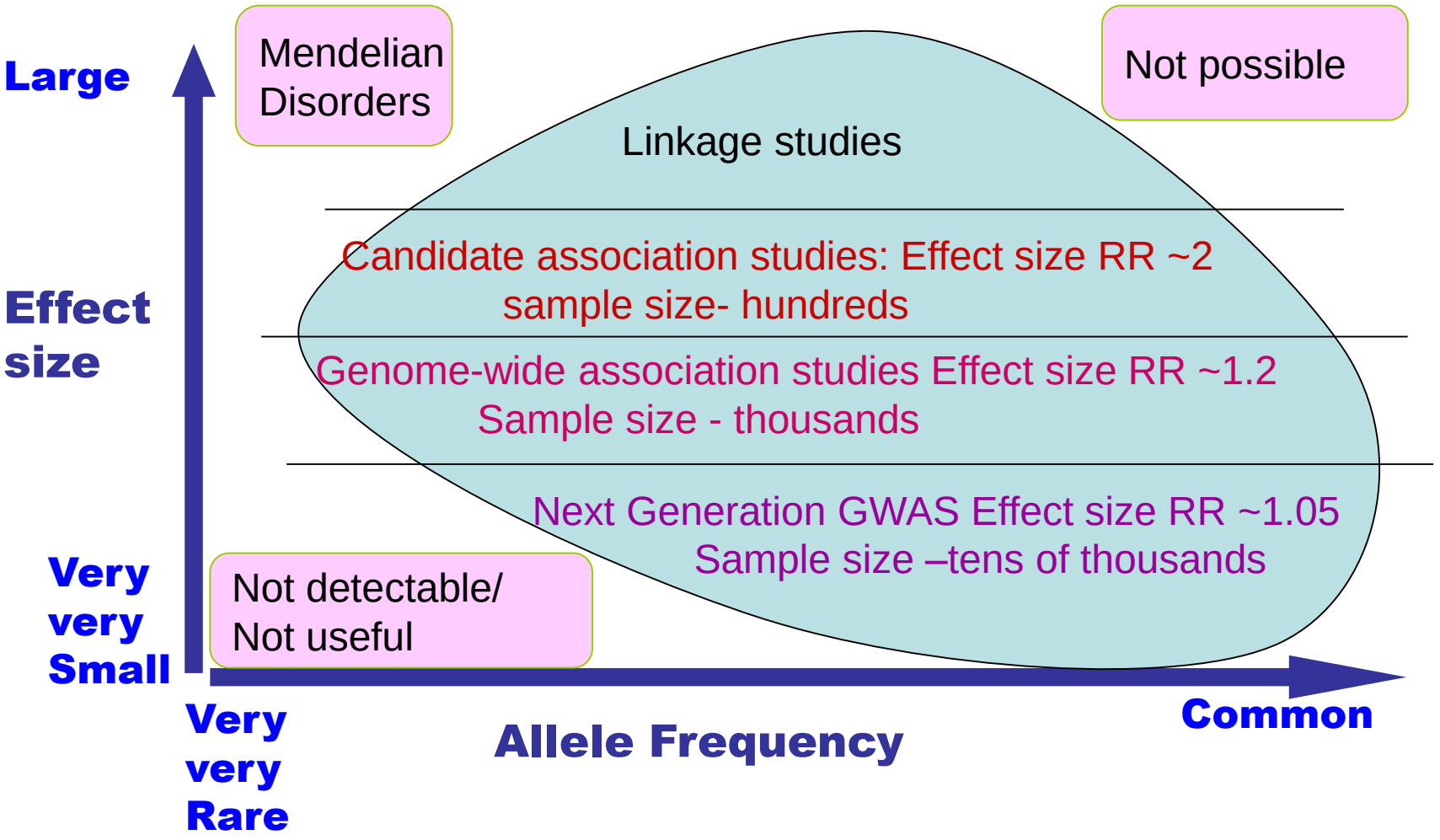
Most effect sizes are very small <1.1

Prediction of individual genetic risk of complex disease

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

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

…and will need huge sample sizes to detect



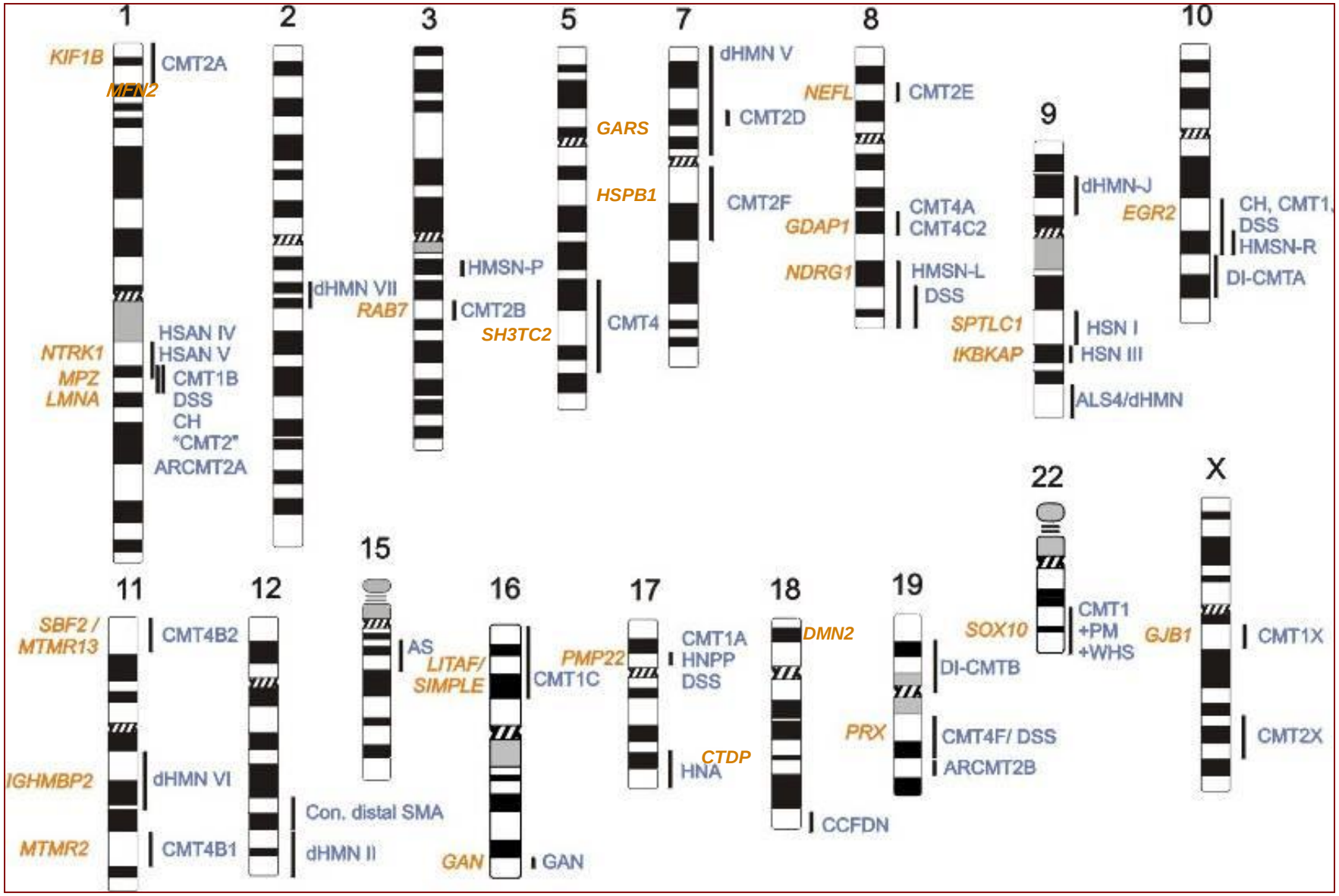
Possible explanations for missing heritability

(in order of increasing plausibility ?)

- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

What if our “disease” is actually dozens (hundreds, thousands) of different diseases that all look the same?

Loci for Inherited Peripheral Neuropathies
Multiple causal loci for Charcot Marie Tooth disease (CMT)



Possible explanations for missing heritability

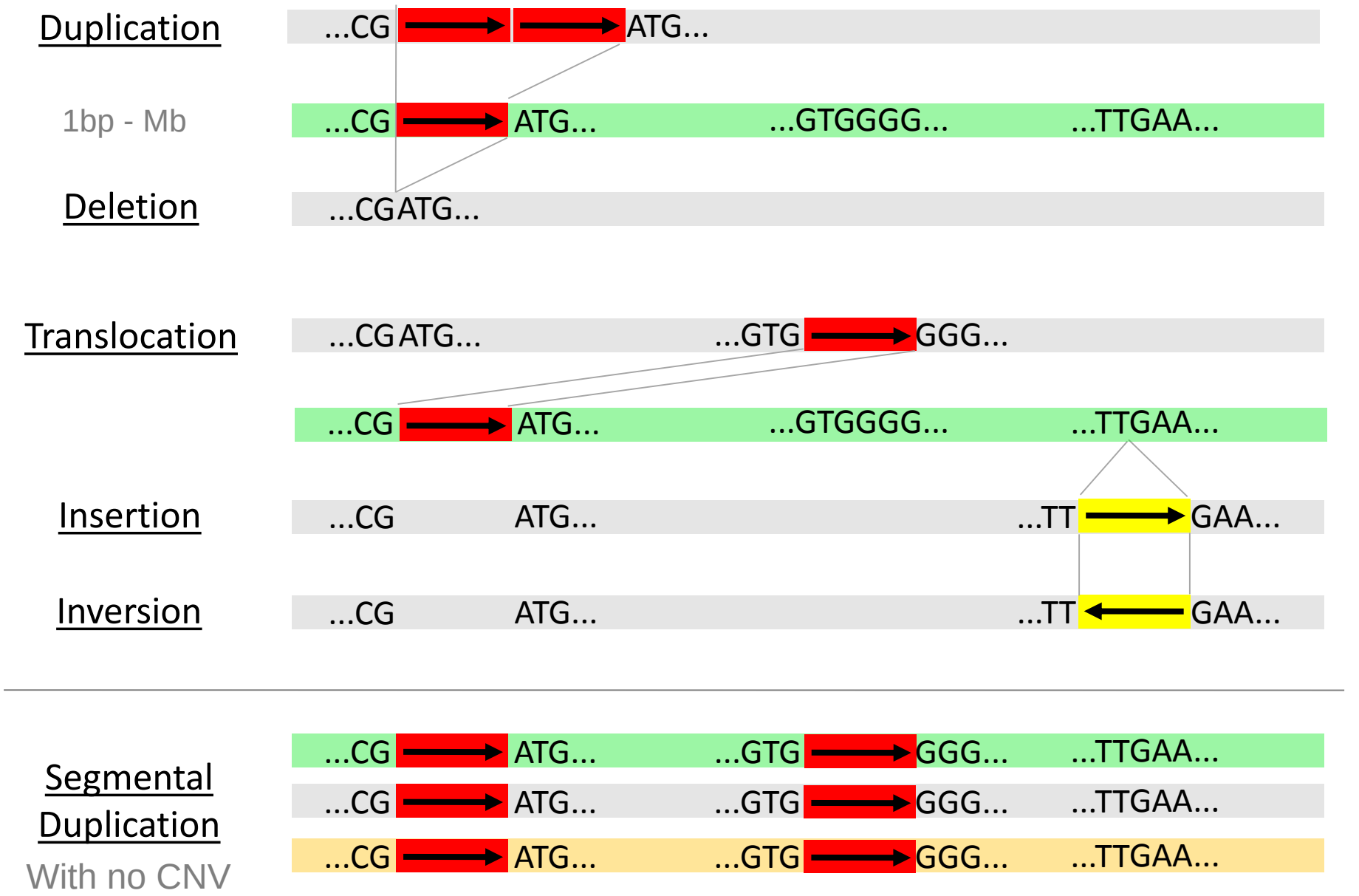
(in order of increasing plausibility ?)

- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

Genetic diversity is larger than differences in DNA sequence

When we take into account:

- Structural variation [e.g. copy number variants (CNV)]
- Epigenetic differences (DNA methylation status)


Duplication
...CG
ATG...
1bp - Mb
...CG
ATG...
...GTGGGG...
...TTGAA...
Deletion
...CGATG...
Translocation
...CG ATG...
...GTG
GGG...
...CG
ATG...
...GTGGGG...
...TTGAA...
Insertion
...CG
ATG...
...TT
GAA...
Inversion
...CG
ATG...
...TT
GAA...
Segmental Duplication
With no CNV
...CG
ATG...
...GTG
GGG...
...TTGAA...
...CG
ATG...
...GTG
GGG...
...TTGAA...
...CG
ATG...
...GTG
GGG...
...TTGAA...

For example: Bipolar disorder

Molecular Psychiatry (2009) 14, 376–380

www.nature.com/mp

IMMEDIATE COMMUNICATION

Singleton deletions throughout the genome increase risk of bipolar disorder

D Zhang[1], L Cheng[1], Y Qian[1], N Alliey-Rodriguez[1], JR Kelsoe[2], T Greenwood[2], C Nievergelt[2], TB Barrett[2], R McKinney[2], N Schork[3,4], EN Smith[3,4], C Bloss[3,4], J Nurnberger[5], HJ Edenberg[6,7], T Foroud[7], W Sheftner[8], WB Lawson[9], EA Nwulia[9], M Hipolito[9], W Coryell[10], J Rice[11], W Byerley[12], F McMahon[13], TG Schulze[13], W Berrettini[14], JB Potash[15], PL Belmonte[15], PP Zandi[15], MG McInnis[16], S Zöllner[16], D Craig[17], S Szelinger[17], D Koller[5], SL Christian[18], C Liu[1*] and ES Gershon[1,18*]

… we present a genome-wide copy number variant (CNV) survey of 1001 cases and 1034 controls ... Singleton deletions (deletions that appear only once in the dataset) more than 100 kb in length are present in 16.2% of BD cases and in 12.3% of controls (permutation P = 0.007).
Our results strongly suggest that BD can result from the effects of multiple rare structural variants.

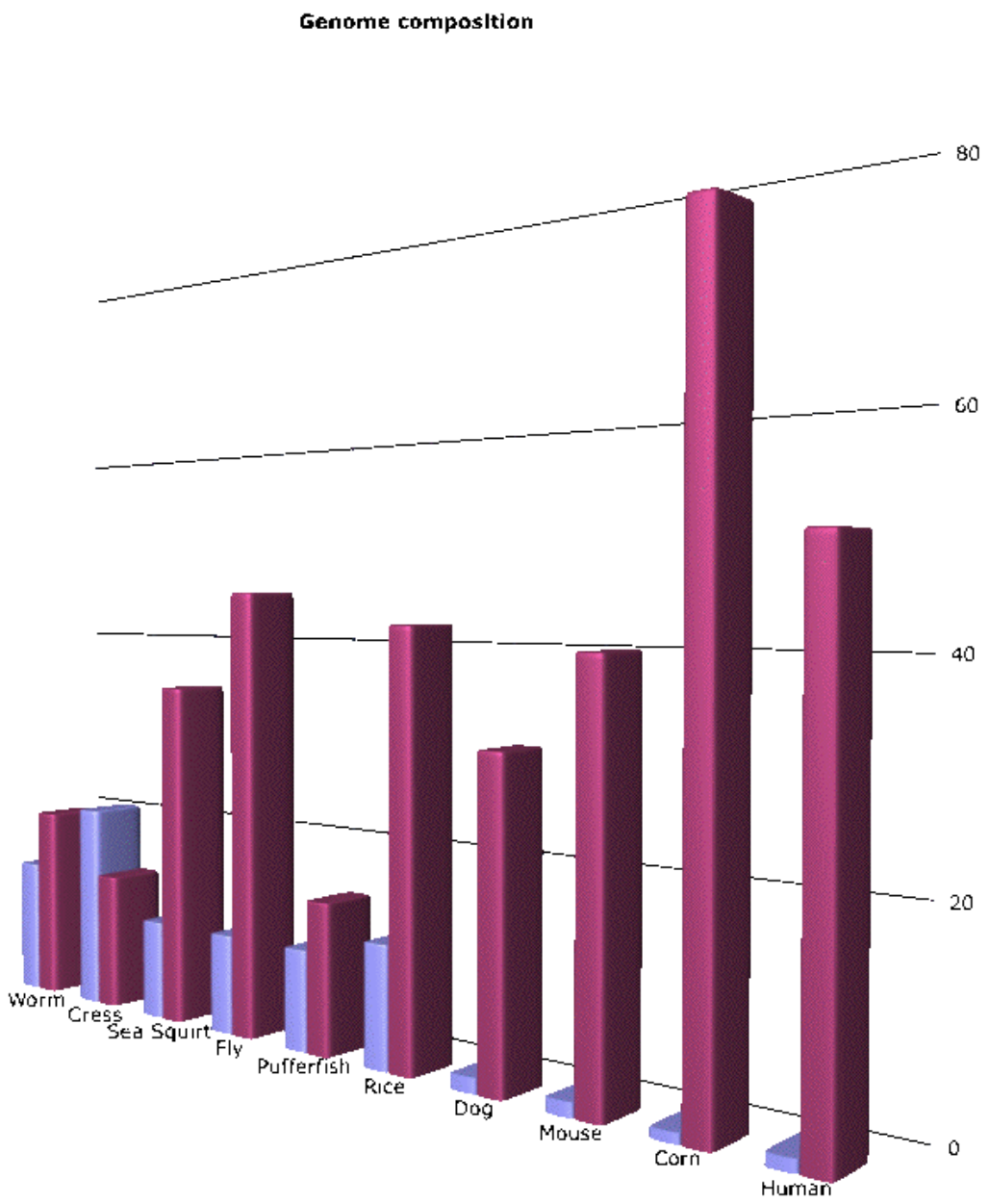
Possible explanations for missing heritability

(in order of increasing plausibility ?)

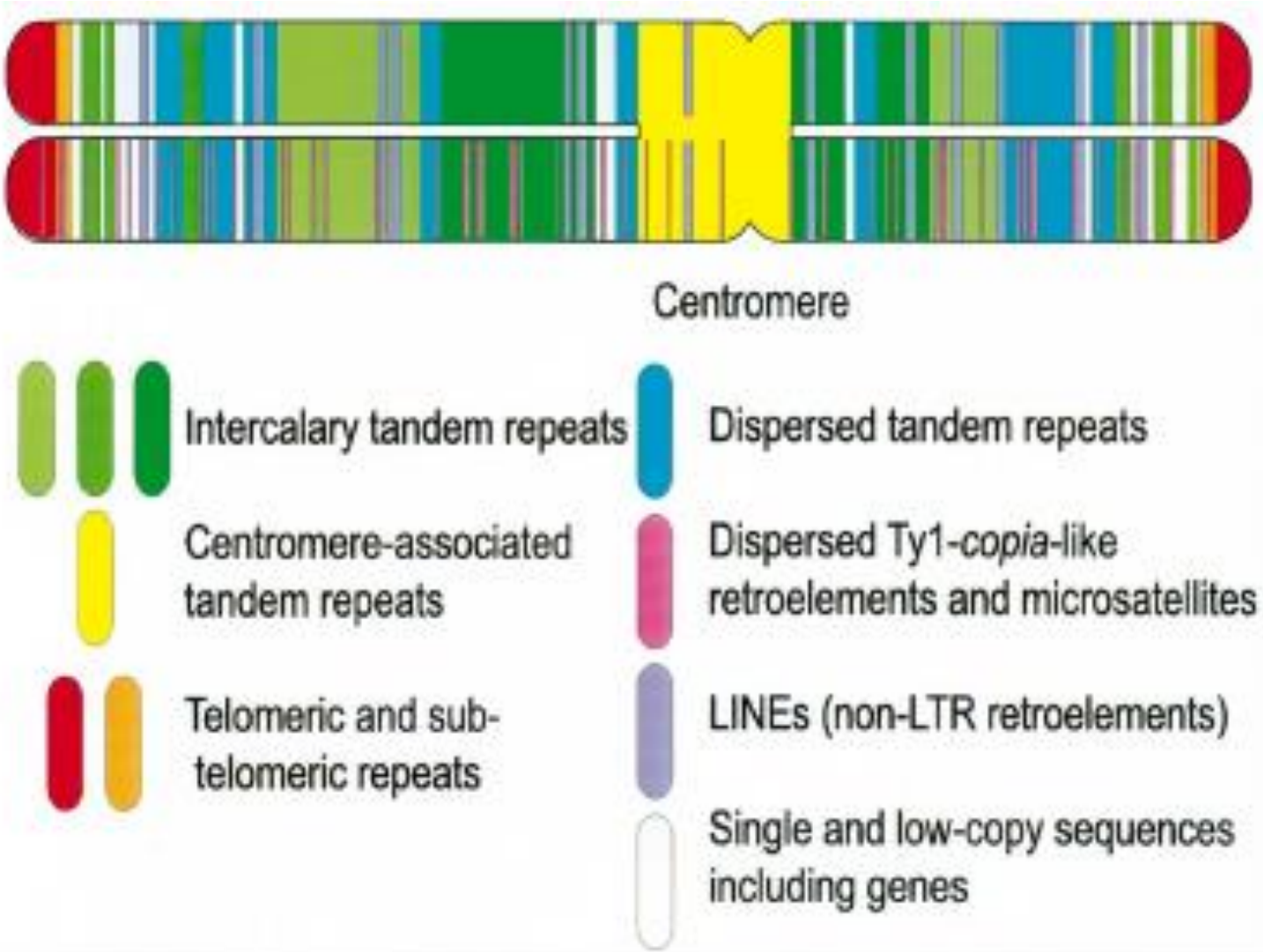
- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect (including CNVs)
- Poor tagging (2)
  - common variants in problematic genomic regions

50% of human genome is repetitive DNA.

Only 1.2% is coding



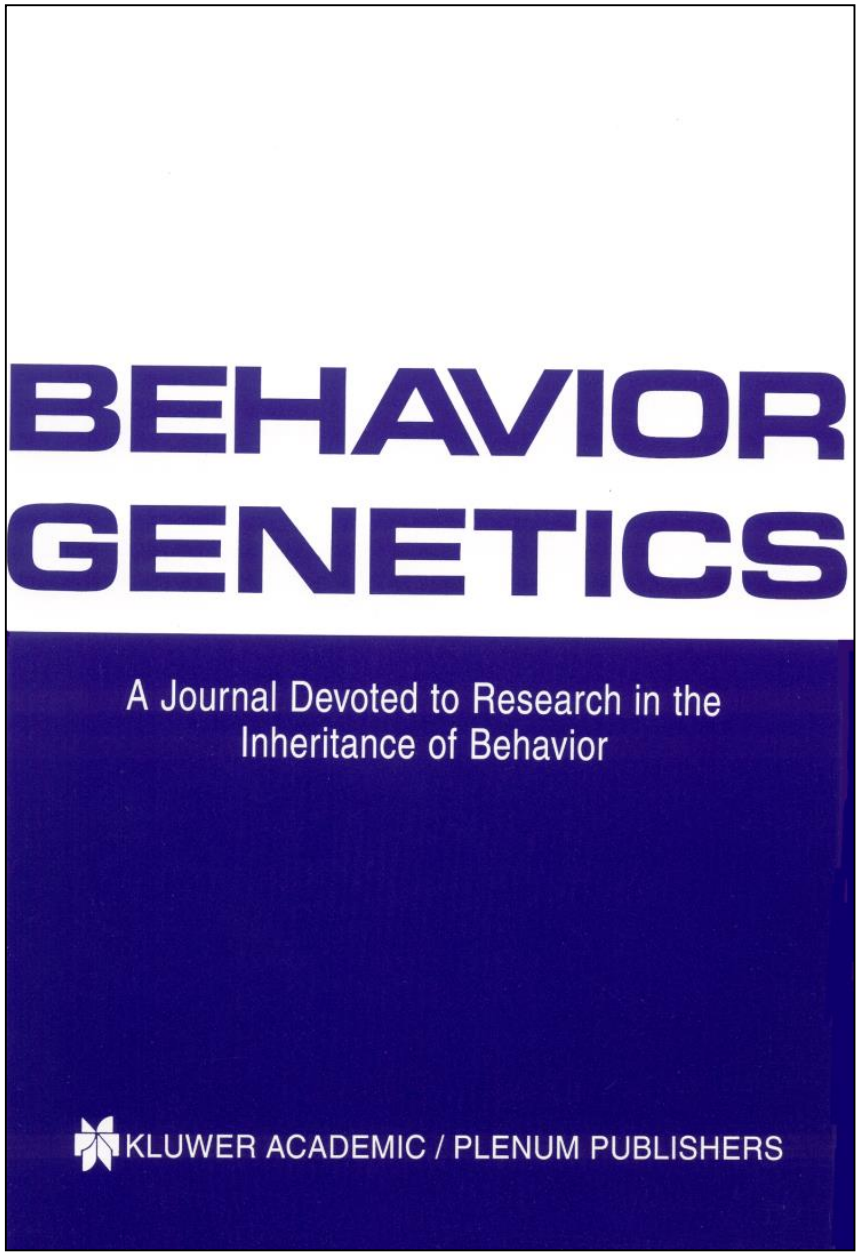
Types of repetitive elements and their chromosomal locations



Summary

- Huge amount of repetitive sequence
- Highly polymorphic
- Some evidence that it has functional significance
- Earlier studies too small (100s) to detect effect sizes now known to be realistic
- Much (most?) such variation poorly tagged with current chips
- Current CNV arrays only detect large variants; no systematic coverage of the vast number of small CNVs (including microsatellites)

- Editor: Nick Martin
- Publisher: Cambridge University Press
- Fully online
- Fast turnaround
- **First submission free to workshop participants!!!!!☺**



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