Science triumphant! the GWAS revolution in complex trait genetics

Nick Martin
Queensland Institute
of Medical Research
Brisbane

Boulder workshop
March 9, 2012

Five Years of GWAS Discovery

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

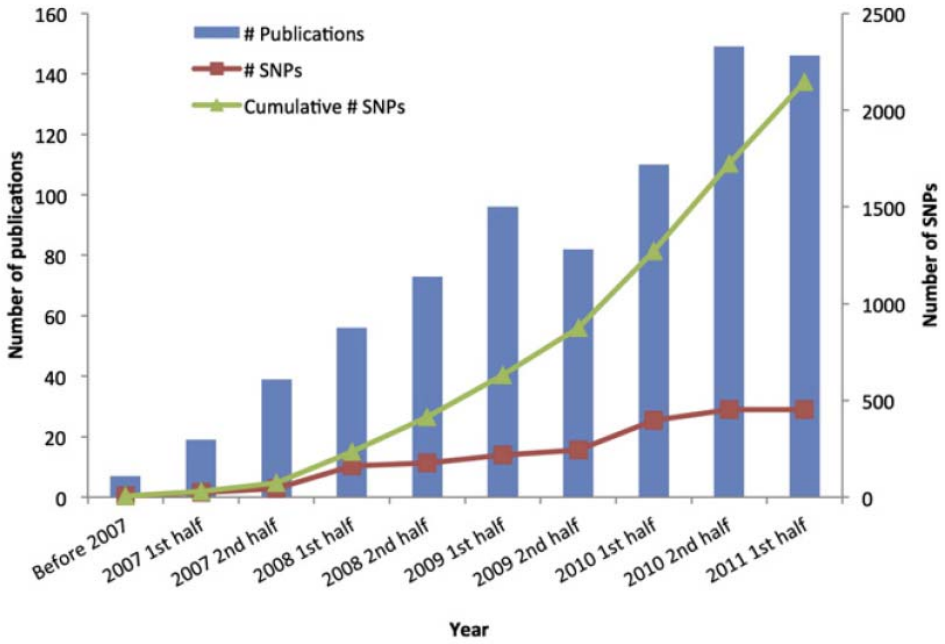


Figure 1. GWAS Discoveries over Time

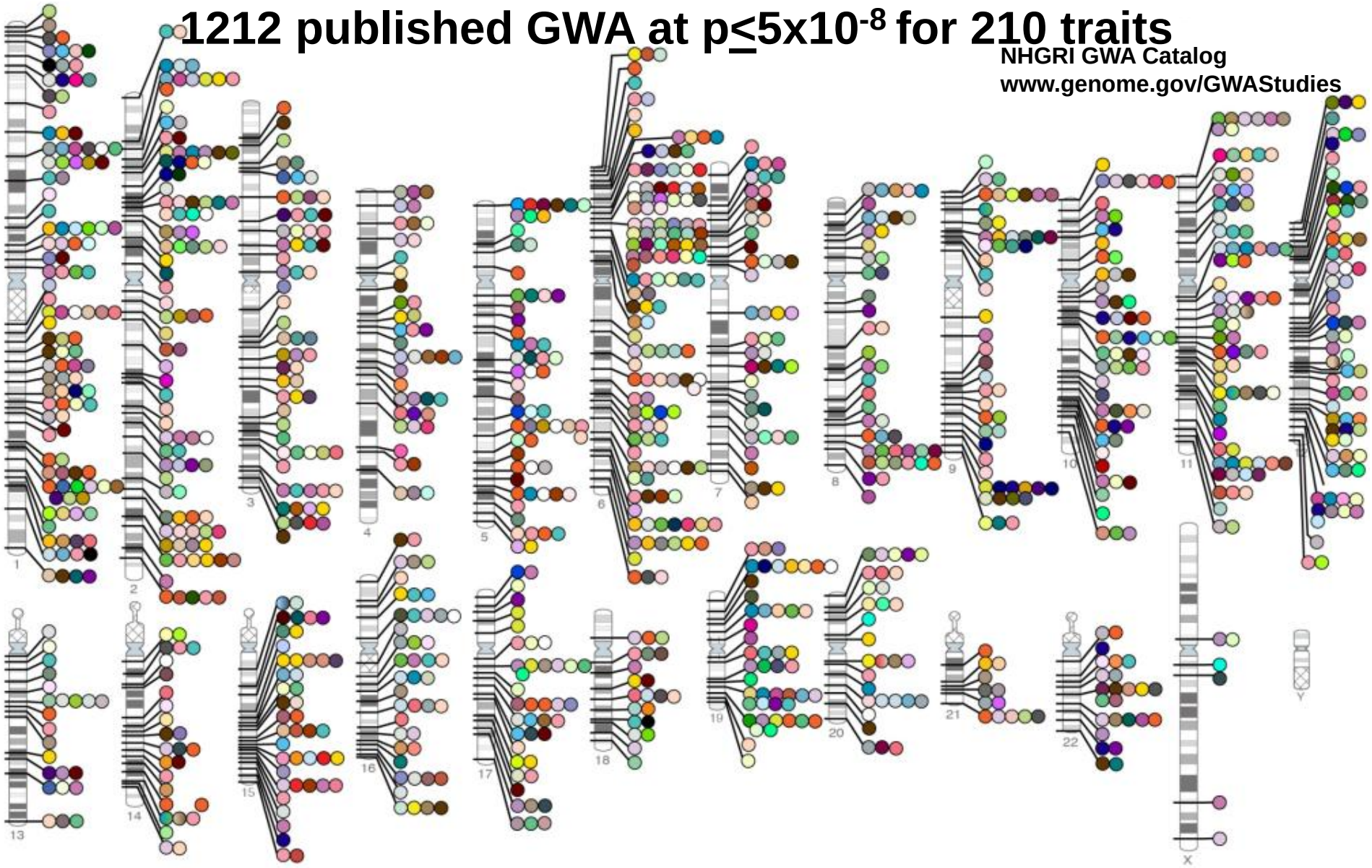
Data obtained from the Published GWAS Catalog (see Web Resources). Only the top SNPs representing loci with association p values $< 5 \times 10^{-8}$ are included, and so that multiple counting is avoided, SNPs identified for the same traits with LD $r^2 > 0.8$ estimated from the entire HapMap samples are excluded.

Summarise criticisms of GWAS
And responses!

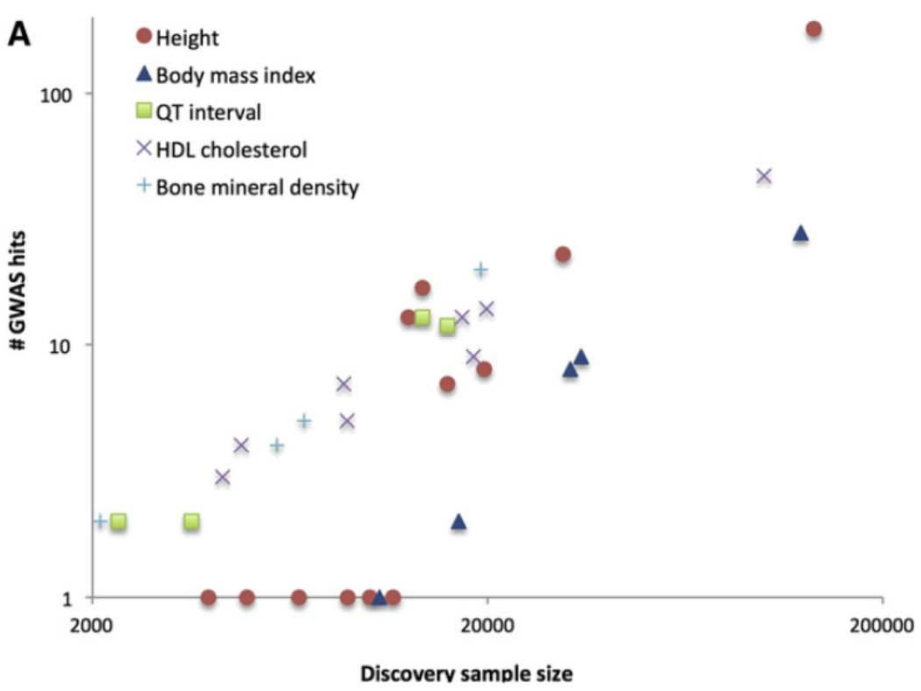
American Journal of Human Genetics *90*, 7–24, January 13, 2012

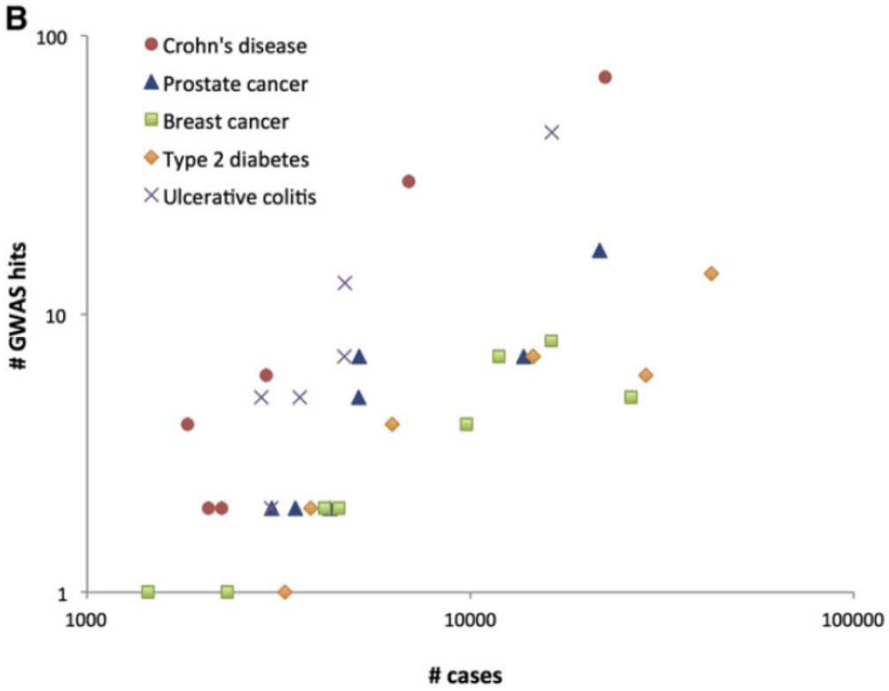
Published Genome-Wide Associations through 12/2010, 1212 published GWA at $p \leq 5 \times 10^{-8}$ for 210 traits

NHGRI GWA Catalog
www.genome.gov/GWAStudies



We are on 5% of these !





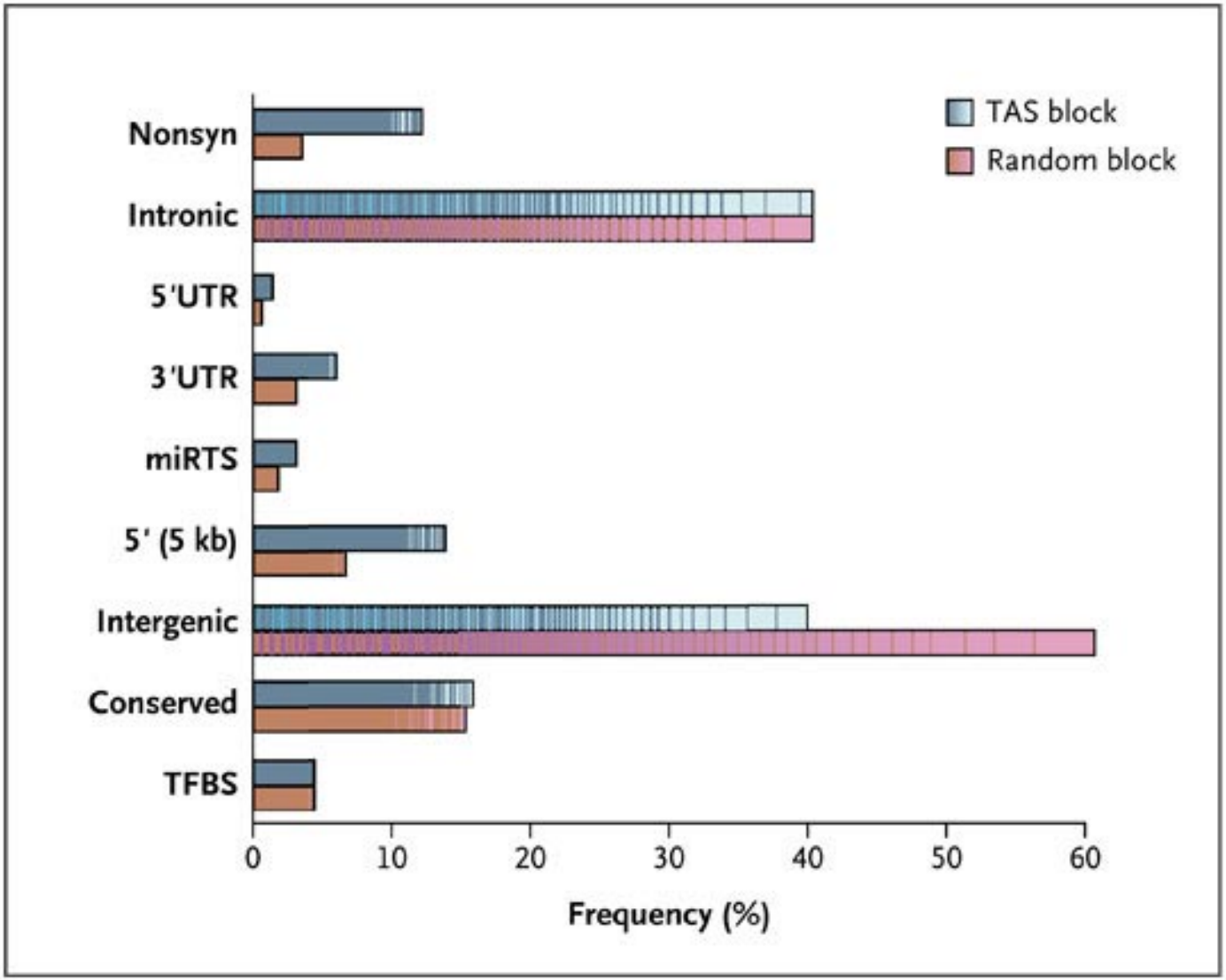
Selected quantitative traits

Selected diseases

Number of Loci Identified is a Function of Sample Size

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

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

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

Sciencexpress Research Article

Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,[1] Caroline Zeiss,[2*] Emily Y. Chew,[3*] Jen-Yue Tsai,[4*] Richard S. Sackler,[1] Chad Haynes,[1] Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6] Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3] Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7†]

[1]Laboratory of Statistical Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA. [2]Department of Ophthalmology and Visual Science, Yale University School of Medicine, 330 Cedar Street, New Haven, CT 06520, USA. [3]National Eye Institute, Building #10, CRC, 10 Center Drive, Bethesda, MD 20892–1204, USA. [4]Biological Imaging Core, National Eye Institute, 9000 Rockville Pike, Bethesda, MD 20892, USA. [5]The EMMES Corporation, 401 North Washington Street, Suite 700, Rockville MD 20850, USA. [6]W.M. Keck Facility, Yale University, 300 George Street, Suite 201, New Haven, CT 06511, USA. [7]Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven CT 06520, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: josephine.hoh@yale.edu

Age-related macular degeneration (AMD) is a major cause of blindness in the elderly. We report a genome-wide screen of 96 cases and 50 controls for polymorphisms associated with AMD. Among 116,204 SNPs genotyped, an intronic and common variant in the complement factor H gene (*CFH*) is strongly associated with AMD (nominal *P* value $<10^{-7}$). Individuals homozygous for the risk alleles have a 7.4-fold increased likelihood of AMD (95% CI 2.9 to 19). Resequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. This polymorphism is in a region of CFH that binds heparin and C-reactive protein. The *CFH* gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

that show evidence of linkage to AMD (*4–8*), but the linkage areas have not been resolved to any causative mutations. Like many other chronic diseases, AMD is caused by a combination of genetic and environmental risk factors. Linkage studies are not as powerful as association studies for the identification of genes contributing to the risk for common, complex diseases (*9*). However, linkage studies have the advantage of searching the whole genome in an unbiased manner without presupposing the involvement of particular genes. Searching the whole genome in an association study requires typing 100,000 or more single nucleotide polymorphisms (SNPs) (*10*). Because of these technical demands, only one whole-genome association study, on susceptibility to myocardial infarction, has been published to date (*11*).

Sciencexpress Research Article

Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,[1] Caroline Zeiss,[2*] Emily Y. Chew,[3*] Jen-Yue Tsai,[4*] Richard S. Sackler,[1] Chad Haynes,[1] Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6] Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3] Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7†]

[1]Laboratory of Statistical Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA. [2]Department of Ophthalmology and Visual Science, Yale University School of Medicine, 330 Cedar Street, New Haven, CT 06520, USA. [3]National Eye Institute, Building #10, CRC, 10 Center Drive, Bethesda, MD 20892–1204, USA. [4]Biological Imaging Core, National Eye Institute, 9000 Rockville Pike, Bethesda, MD 20892, USA. [5]The EMMES Corporation, 401 North Washington Street, Suite 700, Rockville MD 20850, USA. [6]W.M. Keck Facility, Yale University, 300 George Street, Suite 201, New Haven, CT 06511, USA. [7]Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven CT 06520, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: josephine.hoh@yale.edu

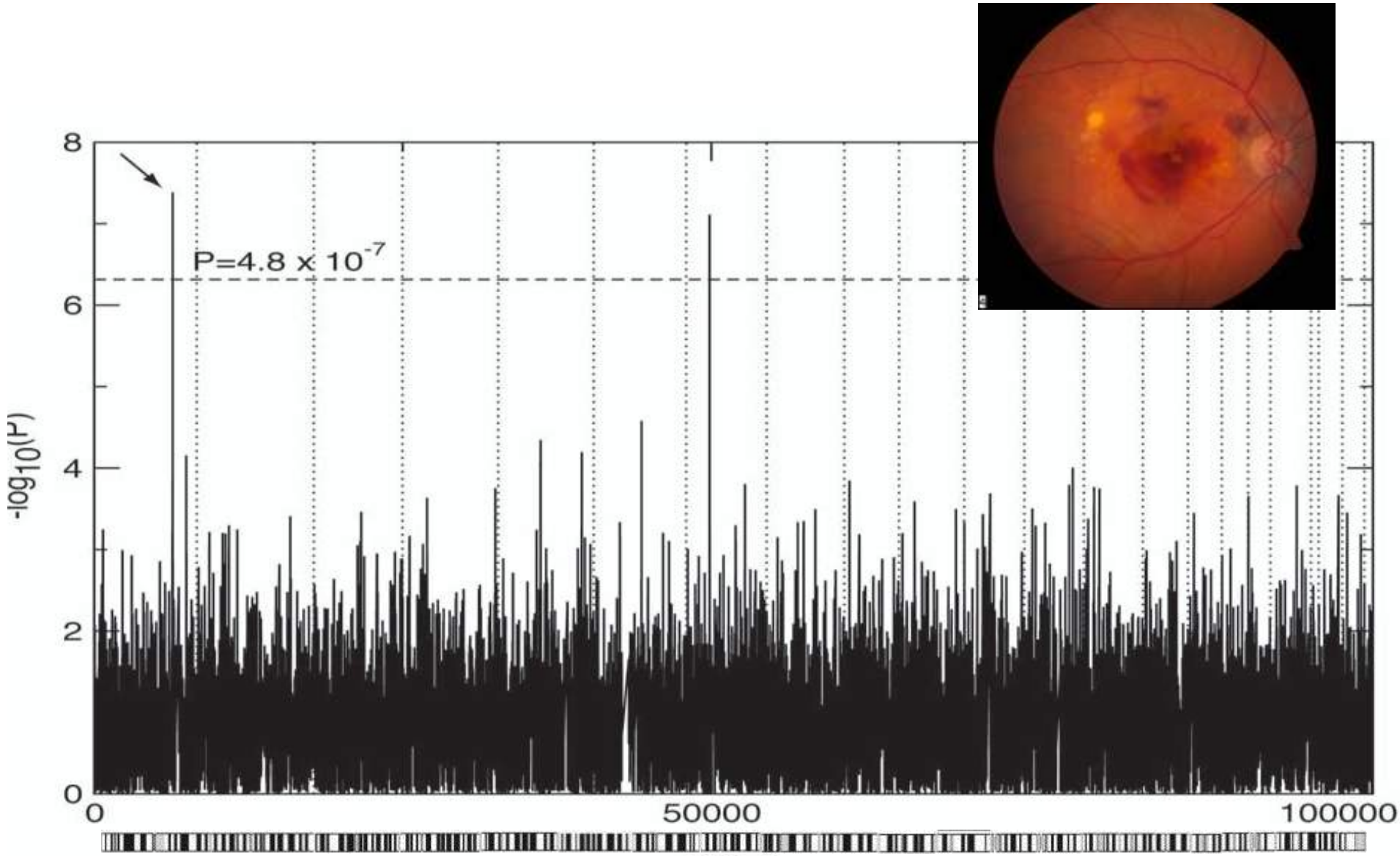
Age-related macular degeneration (AMD) is a major ca
w
p
Sl
co
w
h
likelihood of AMD (95% CI 2.9 to 19). Resequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. This polymorphism is in a region of CFH that binds heparin and C-reactive protein. The *CFH* gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

that show evidence of linkage to AMD (*4–8*), but the linkage

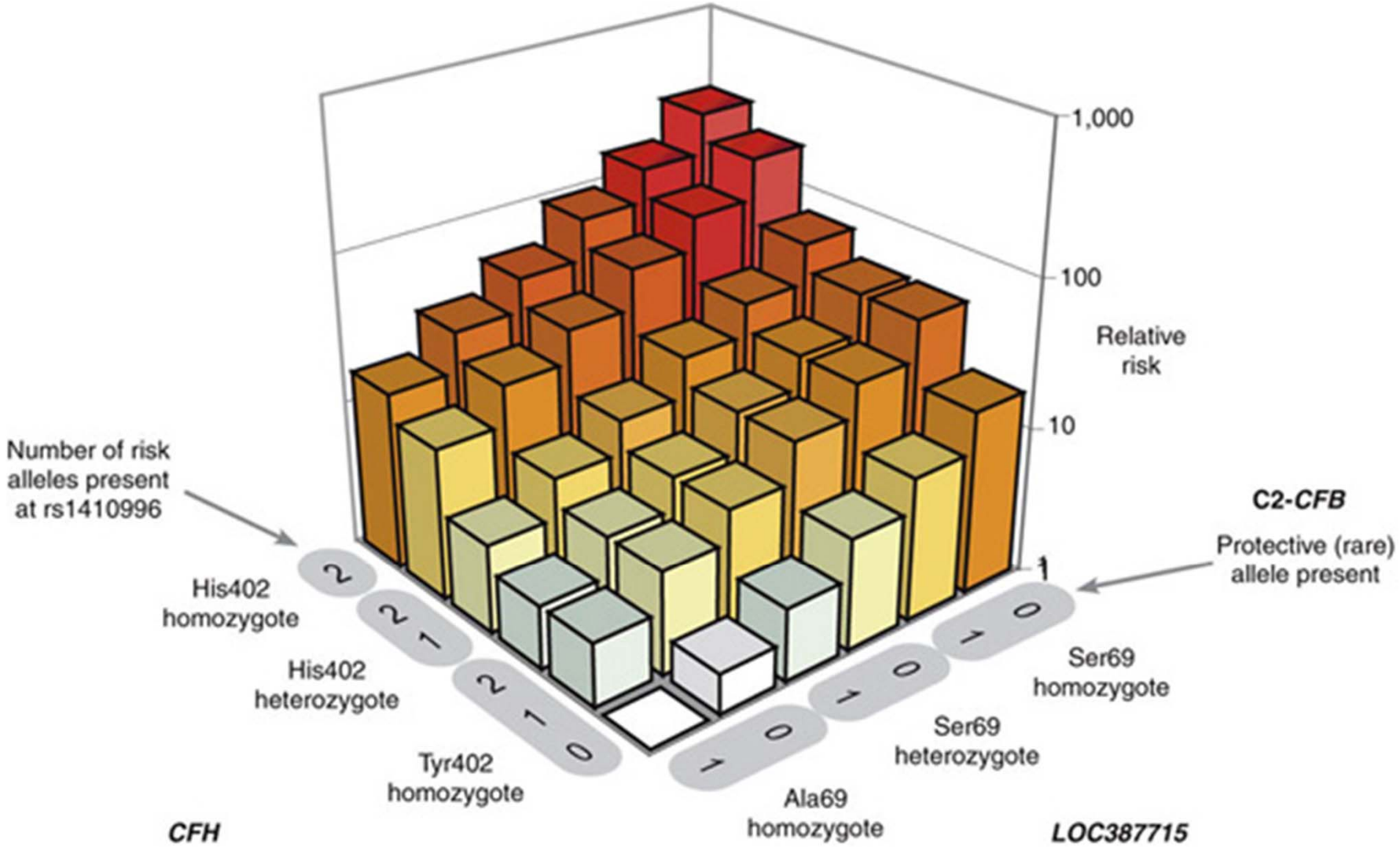
96 cases and 50 controls for

unbiased manner without presupposing the involvement of particular genes. Searching the whole genome in an association study requires typing 100,000 or more single nucleotide polymorphisms (SNPs) (*10*). Because of these technical demands, only one whole-genome association study, on susceptibility to myocardial infarction, has been published to date (*11*).

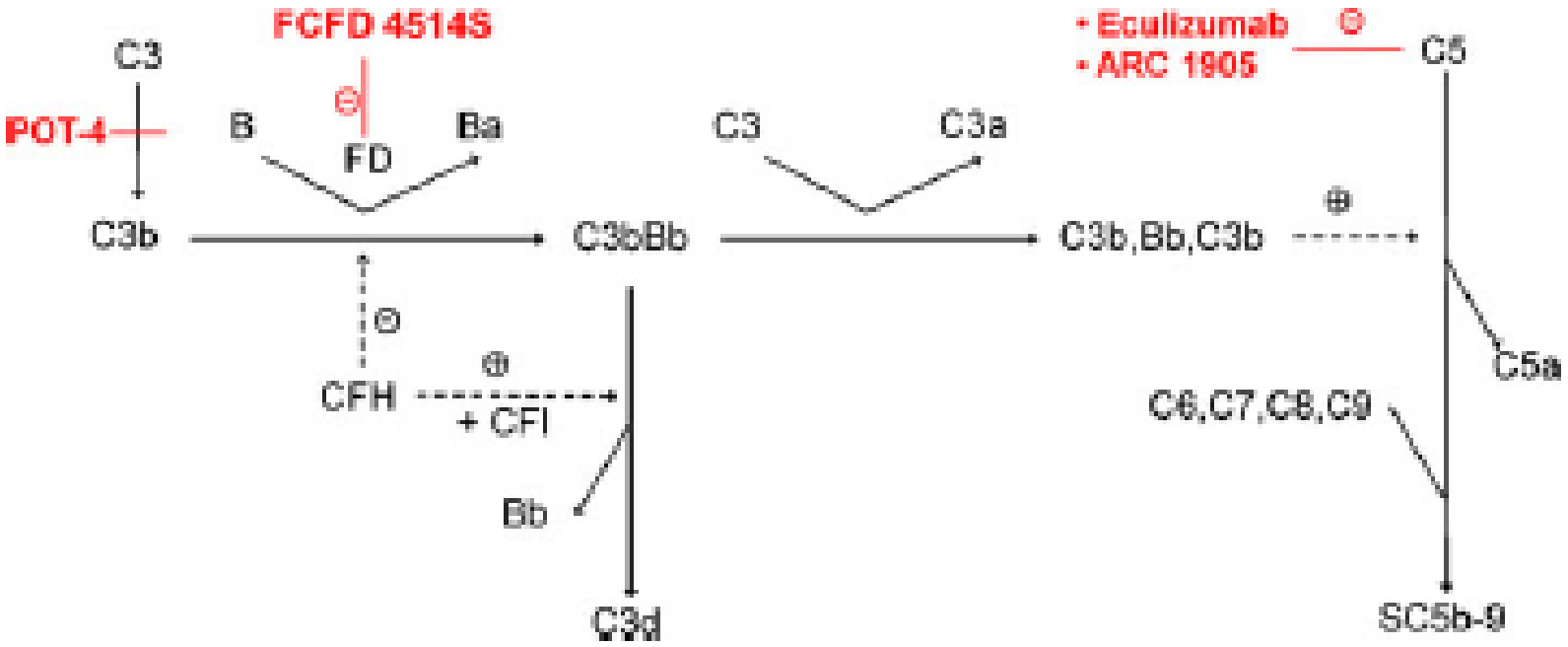
Age-related Macular Degeneration (AMD) – the first GWAS success (2005)



Relative risk plotted as a function of the genetic load of the five variants that influence risk of AMD



Complement Pathway & New Therapeutic Agents Currently in Clinical Trials for AMD



NATURE | Vol 452 | 3 April 2008

A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25

Rayjean J. Hung[1,2]*, James D. McKay[1]*, Valerie Gaborieau[1], Paolo Boffetta[1], Mia Hashibe[1], David Zaridze[3],

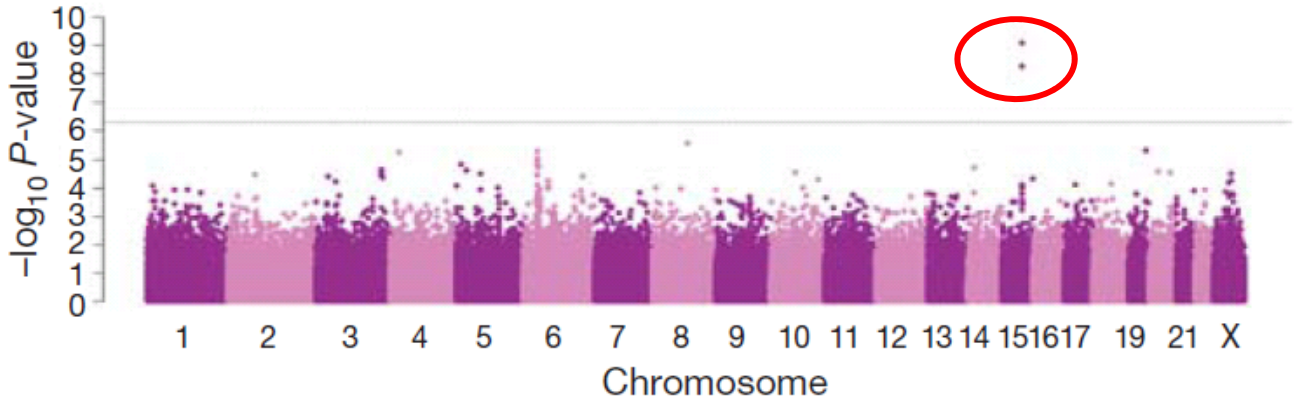
A variant associated with nicotine dependence, lung cancer and peripheral arterial disease

Thorgeir E. Thorgeirsson[1]*, Frank Geller[1]*, Patrick Sulem[1]*, Thorunn Rafnar[1]*, Anna Wiste[1,2],

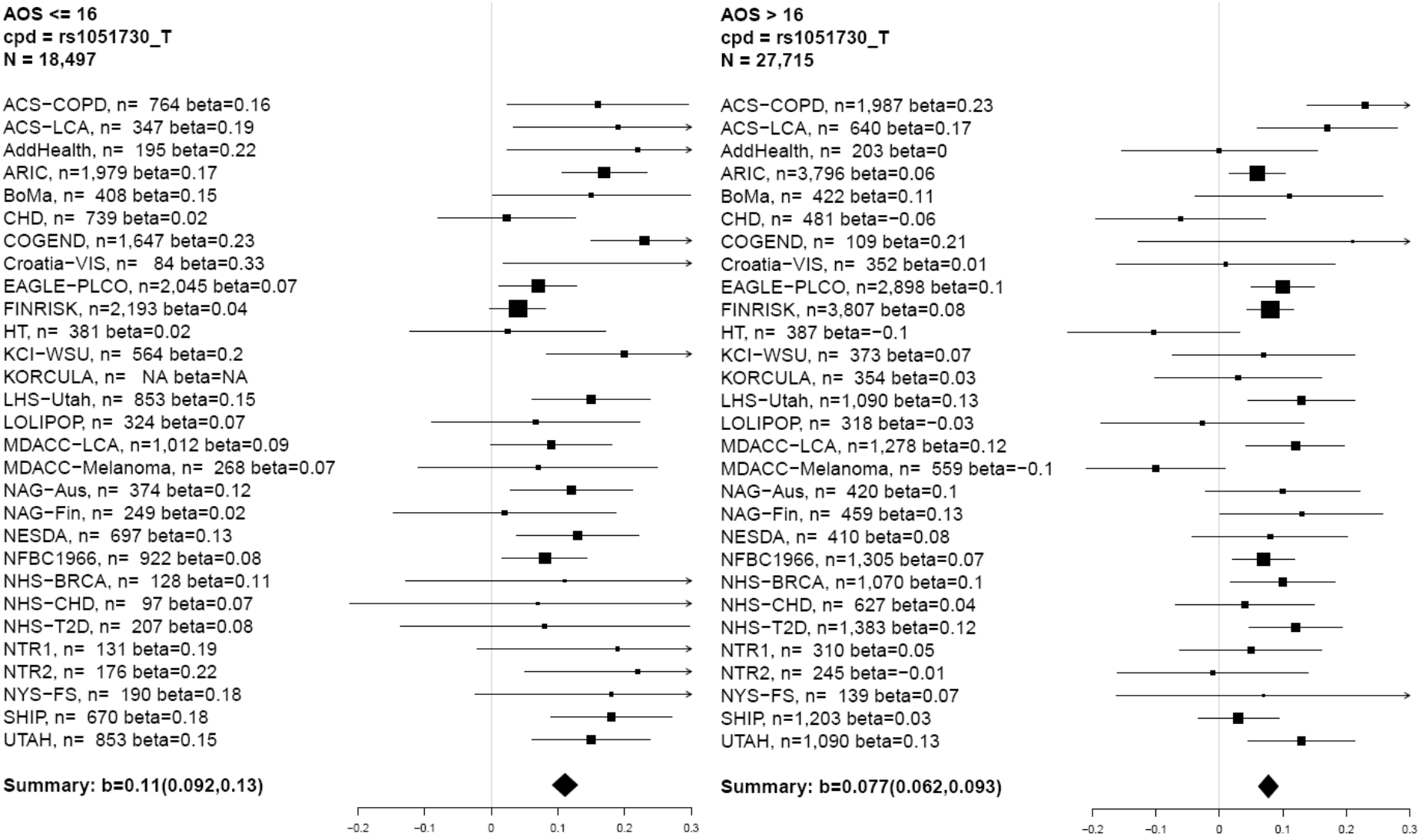
Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1

MAY 2008 NATURE GENETICS

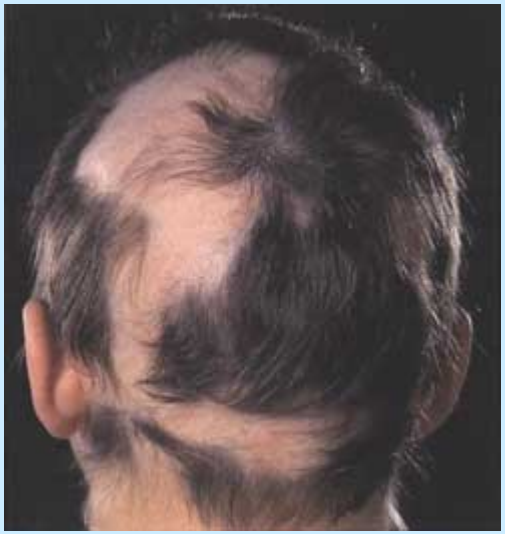
Christopher I Amos[1], Xifeng Wu[1], Peter Broderick[2], Ivan P Gorlov[1], Jian Gu[1], Timothy Eisen[3], Qiong Dong[1],



Association between CPD and rs1051730 across studies where subjects are stratified by age of onset of regular smoking (AOS) at or before age 16 versus after age 16. P-value for difference between meta-analysis betas for the two AOS strata is 0.006.



Sarah Hartz



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

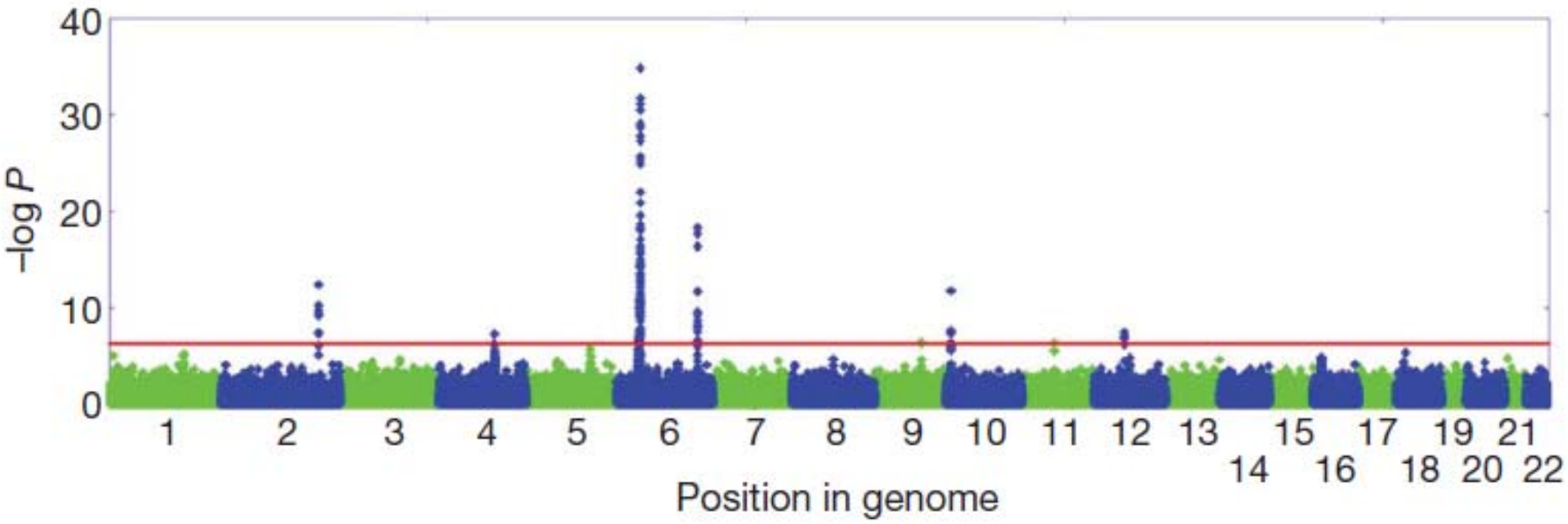
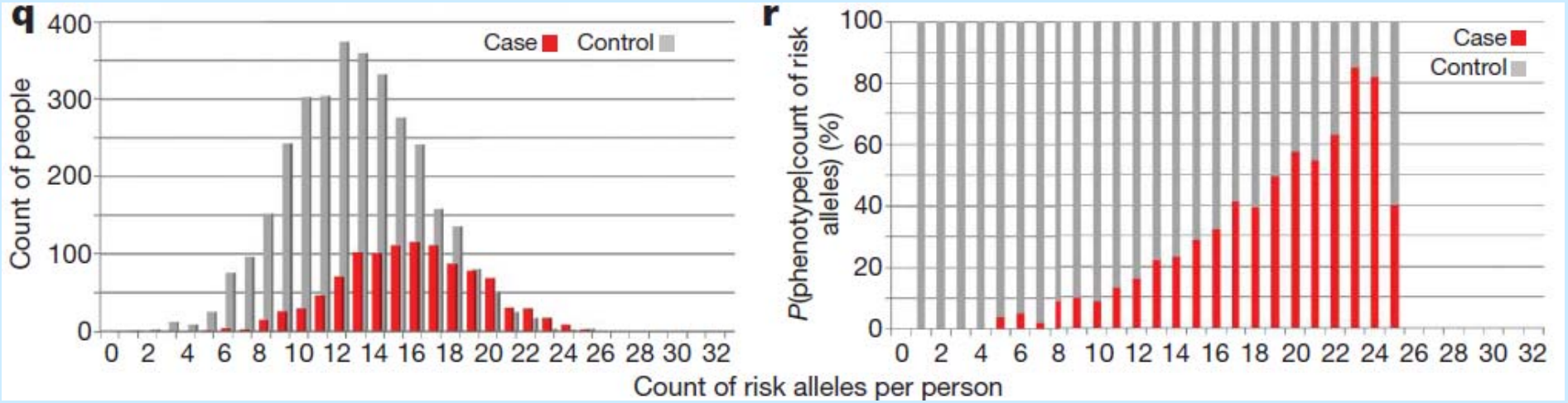


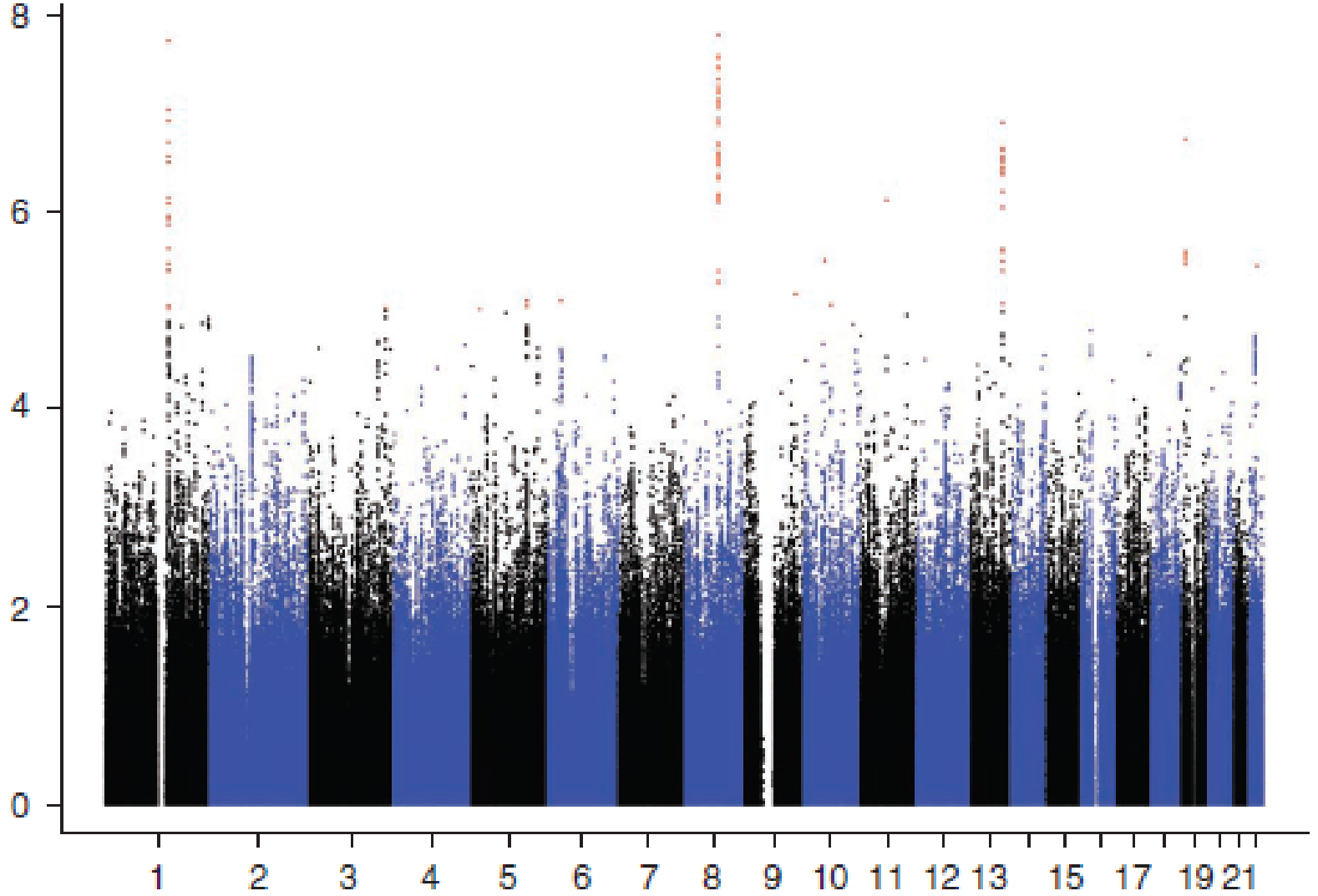
Table 1 | Genes with significant association to AA

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

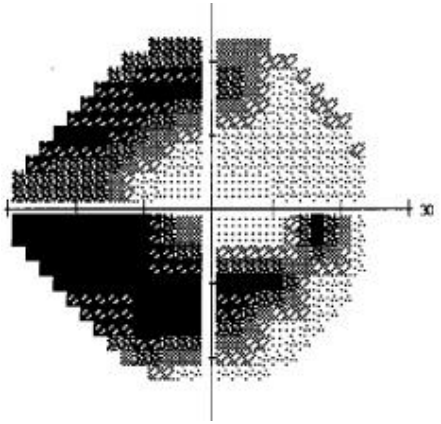
Each of the eight regions implicated in our study contains multiple significant SNPs, which are detailed in Supplementary Tables 1 and 2. Here we display candidate genes within the implicated regions, and include the *P* value of the most significant SNP, and the odds ratio for the SNP with the largest 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).

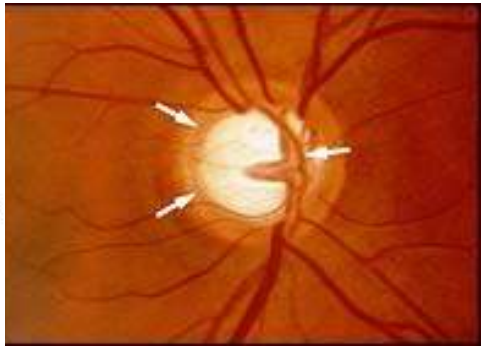


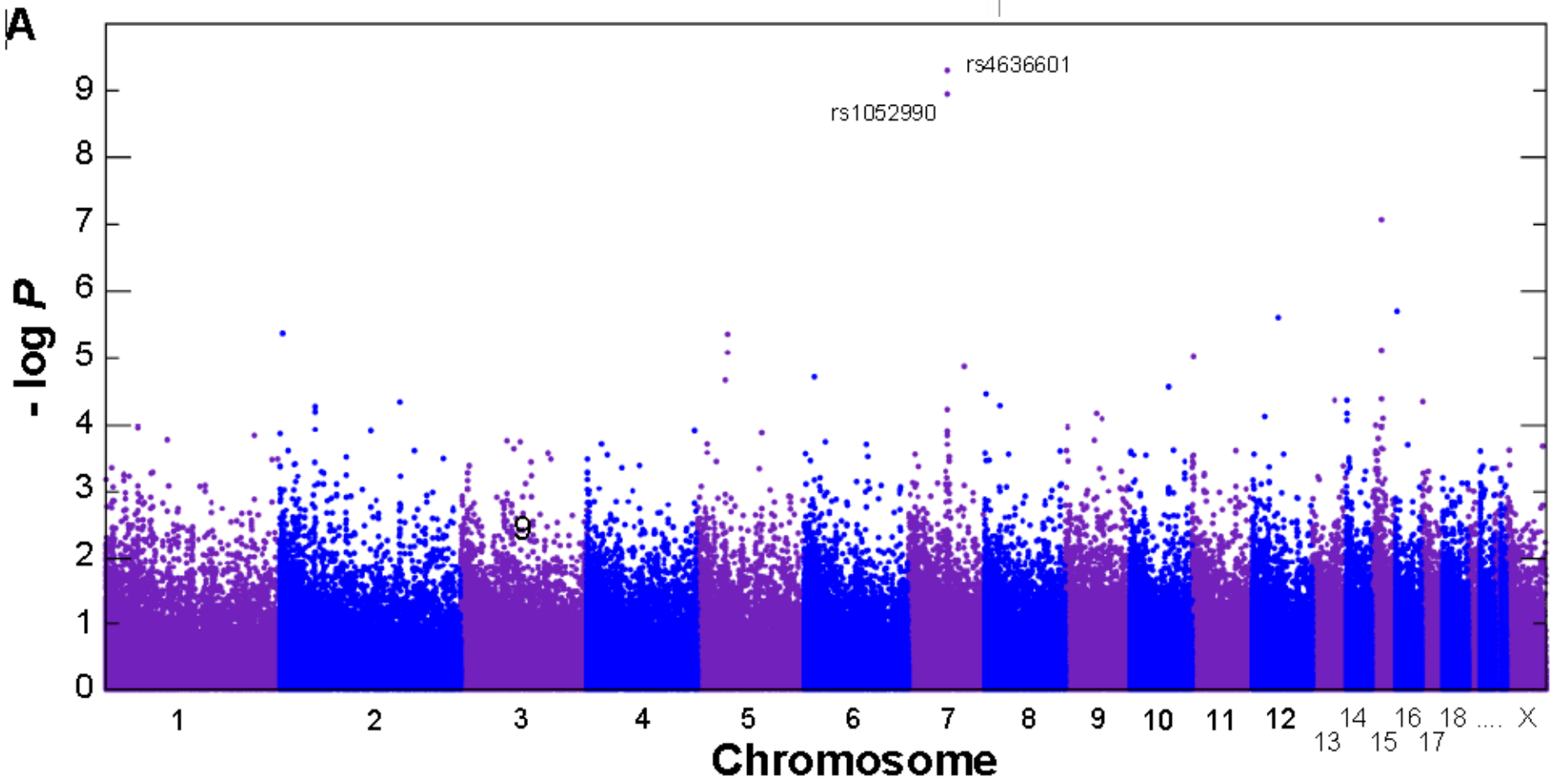
Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis



Manhattan Plot for Glaucoma 2010







Nat Genet. 2010 Oct;42(10):906-9.

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.

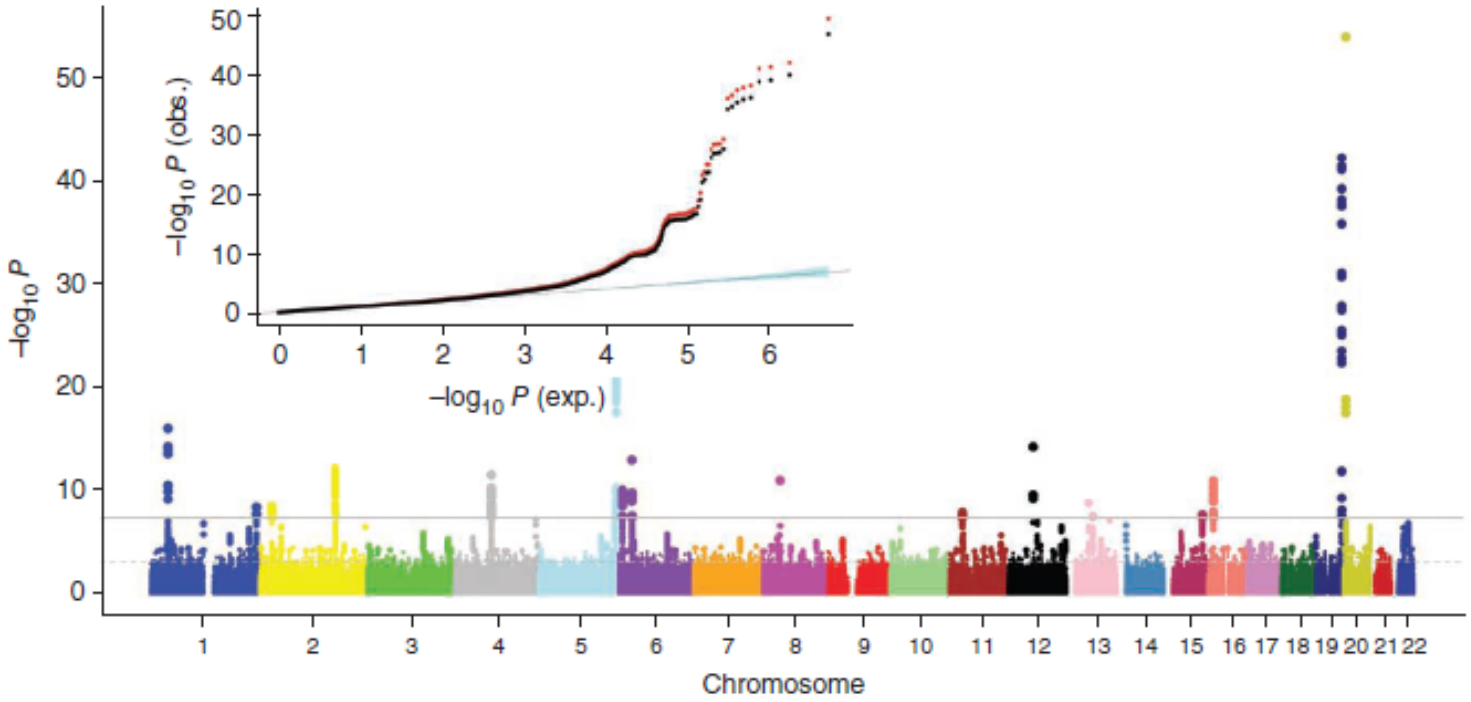


Figure 1 Discovery GWAS results. Manhattan plot of discovery meta-analysis. Inset, quantile-quantile plot of discovery primary analysis (red) and double genomic control–adjusted primary analysis (black). Obs., observed; exp., expected.

New gene functions in megakaryopoiesis and platelet formation

A full list of authors and their affiliations appears at the end of paper.

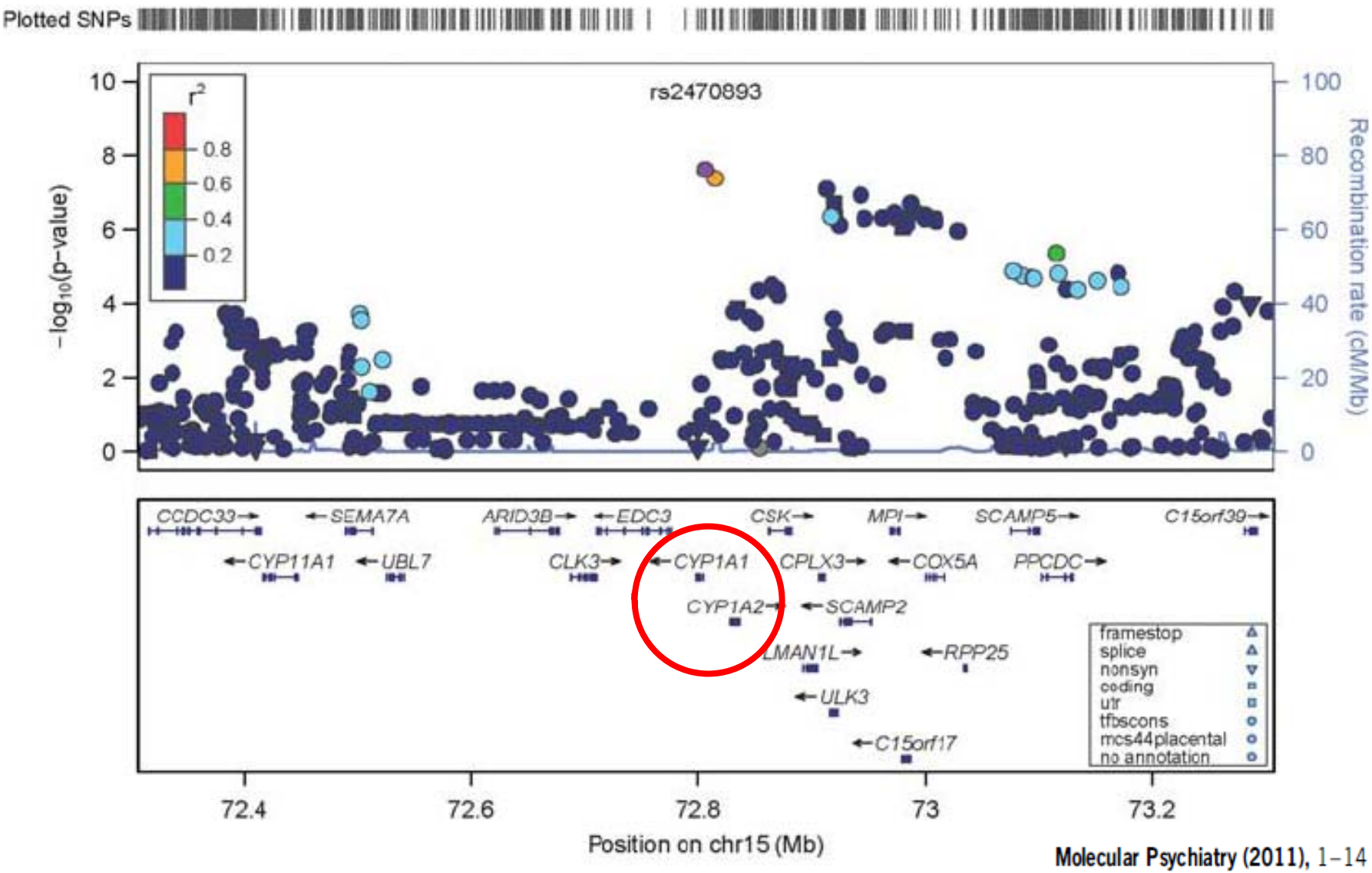
Platelets are the second most abundant cell type in blood and are essential for maintaining haemostasis. Their count and volume are tightly controlled within narrow physiological ranges, but there is only limited understanding of the molecular processes controlling both traits. Here we carried out a high-powered meta-analysis of genome-wide association studies (GWAS) in up to 66,867 individuals of European ancestry, followed by extensive biological and functional assessment. We identified 68 genomic loci reliably associated with platelet count and volume mapping to established and putative novel regulators of megakaryopoiesis and platelet formation. These genes show megakaryocyte-specific gene expression patterns and extensive network connectivity. Using gene silencing in *Danio rerio* and *Drosophila melanogaster*, we identified 11 of the genes as novel regulators of blood cell formation. Taken together, our findings advance understanding of novel gene functions controlling fate-determining events during megakaryopoiesis and platelet formation, providing a new example of successful translation of GWAS to function.

- **66,867 individuals**

- **68 new loci controlling platelet count and volume**

Nature, December 2011

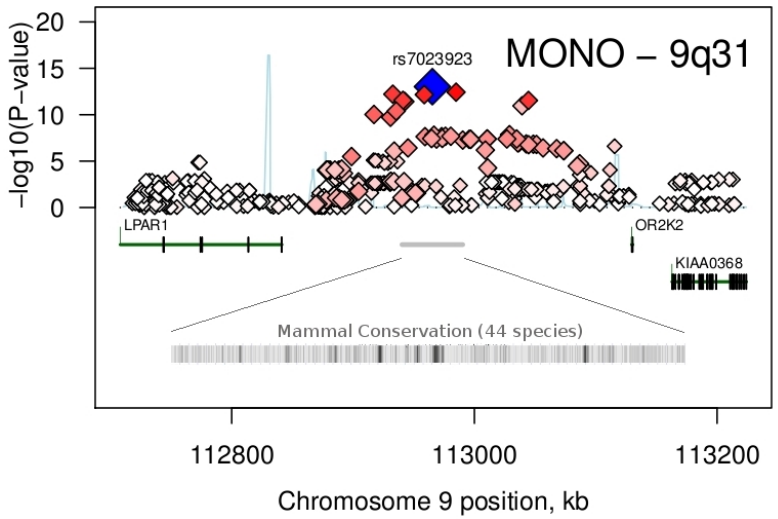
Genome-wide association analysis of coffee drinking suggests association with *CYP1A1/CYP1A2* and *NRCAM*

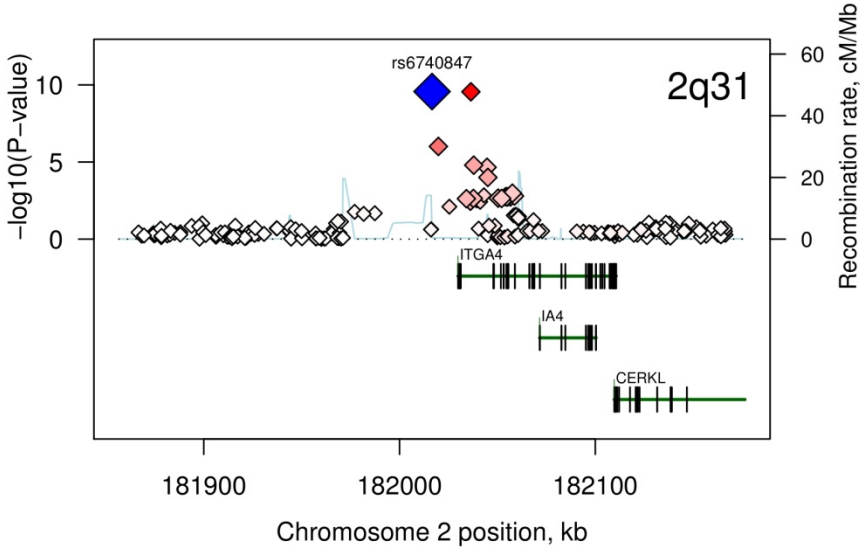


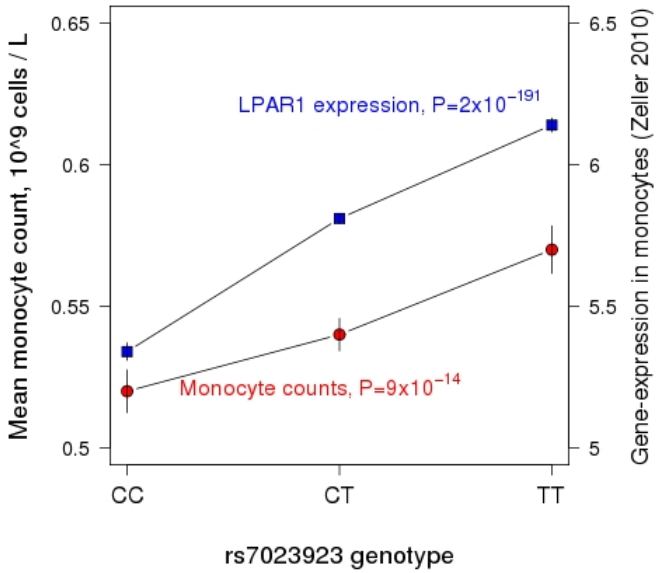
Molecular Psychiatry (2011), 1–14

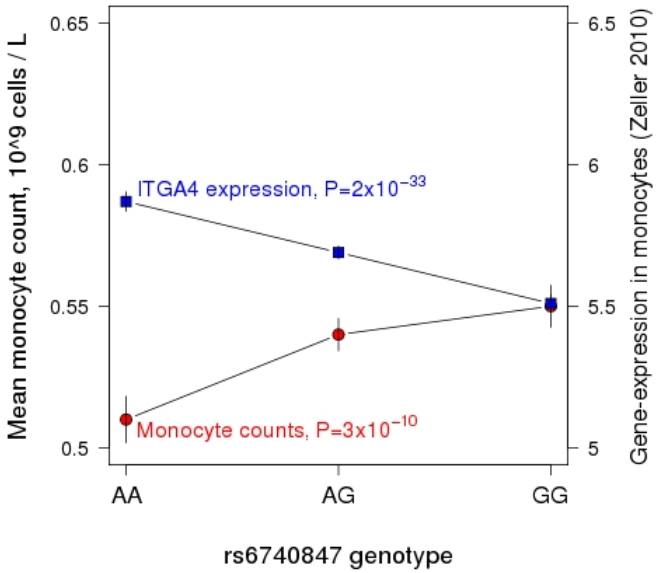
GWAS of monocyte counts – **help from expression data**

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





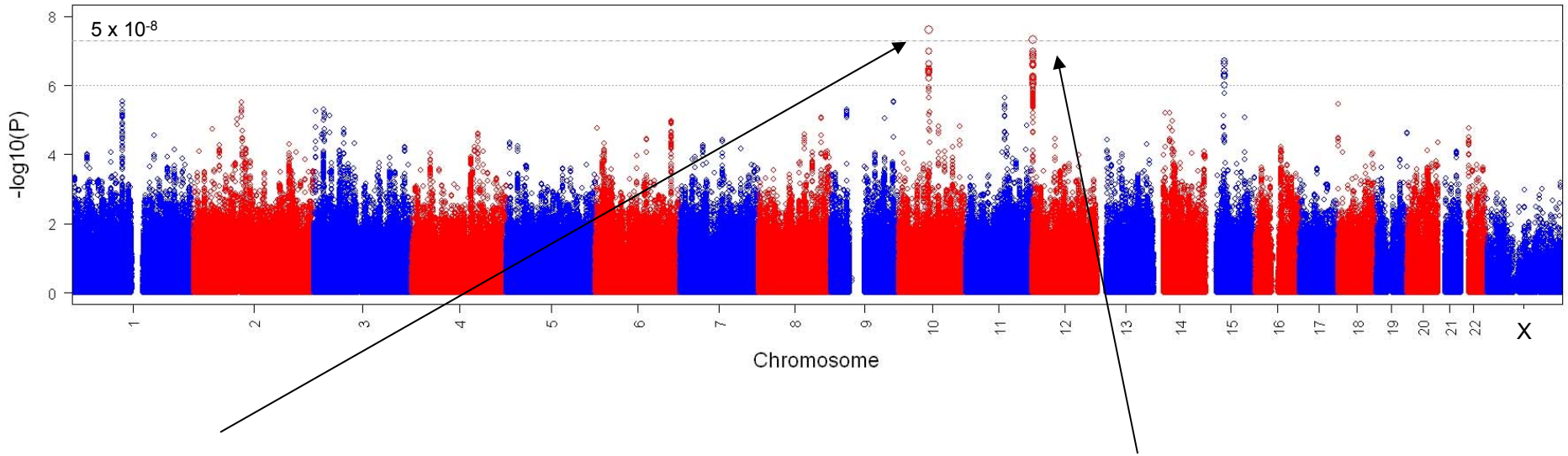




Ferreira et al. (2009) AJHG 85: 745; Zeller et al. (2010) PLoS One 5: e10693.

Bipolar GWAS of 10,648 samples

>1.7 million genotyped and (high confidence) imputed SNPs



Ankryin-G (ANK3)

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

CACNA1C

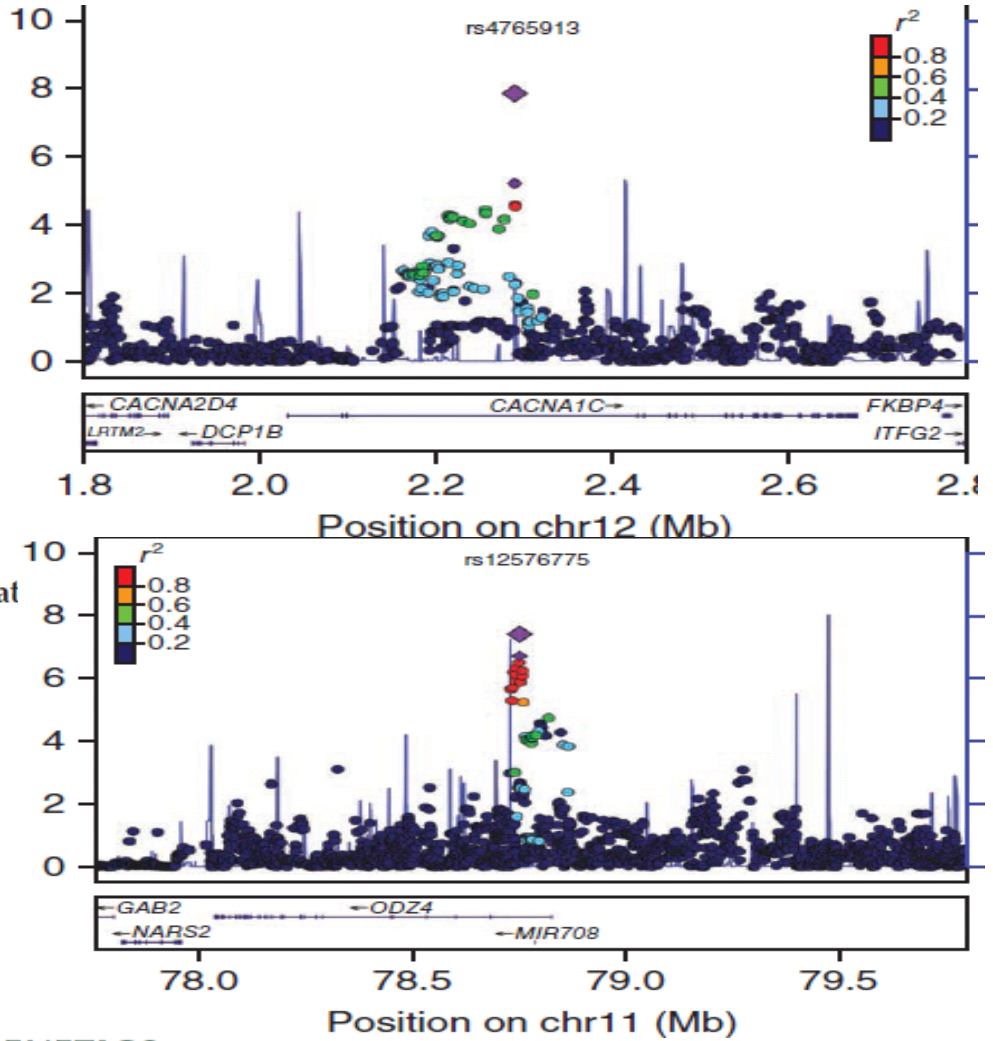
Sample	Case	Controls	*P*-value
STEP	35.7%	32.4%	0.0015
WTCCC	35.7%	31.5%	0.0003
EXT	35.3%	33.7%	0.0108
Total	**35.6%**	**32.4%**	**7×10^{-8}**

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

Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*

Psychiatric GWAS Consortium Bipolar Disorder Working Group[1]

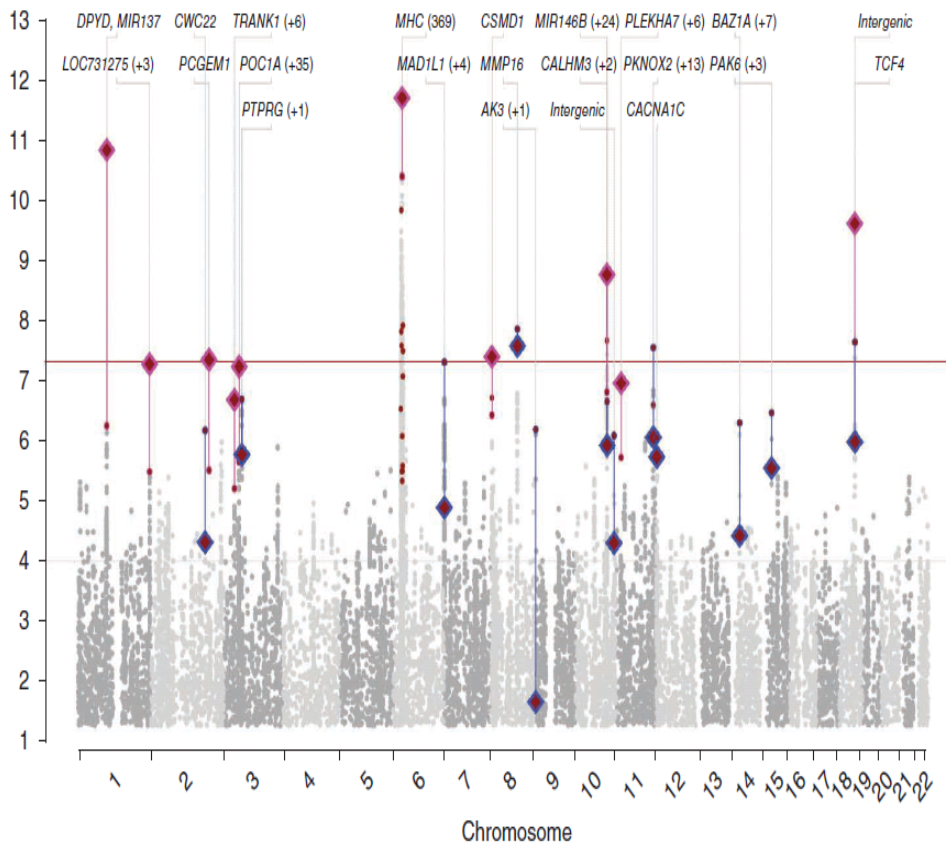
We conducted a combined genome-wide association study (GWAS) of 7,481 individuals with bipolar disorder (cases) and 9,250 controls as part of the Psychiatric GWAS Consortium. Our replication study tested 34 SNPs in 4,496 independent cases with bipolar disorder and 42,422 independent controls and found that 18 of 34 SNPs had $P < 0.05$, with 31 of 34 SNPs having signals with the same direction of effect ($P = 3.8 \times 10^{-7}$). An analysis of all 11,974 bipolar disorder cases and 51,792 controls confirmed genome-wide significant evidence of association for *CACNA1C* and identified a new intronic variant in *ODZ4*. We identified a pathway comprised of subunits of calcium channels enriched in bipolar disorder association intervals. Finally, a combined GWAS analysis of schizophrenia and bipolar disorder yielded strong association evidence for SNPs in *CACNA1C* and in the region of *NEK4-ITIH1-ITIH3-ITIH4*. Our replication results imply that increasing sample sizes in bipolar disorder will confirm many additional loci.



Genome-wide association study identifies five new schizophrenia loci

The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium[1]

We examined the role of common genetic variation in schizophrenia in a genome-wide association study of substantial size: a stage 1 discovery sample of 21,856 individuals of European ancestry and a stage 2 replication sample of 29,839 independent subjects. The combined stage 1 and 2 analysis yielded genome-wide significant associations with schizophrenia for seven loci, five of which are new (1p21.3, 2q32.3, 8p23.2, 8q21.3 and 10q24.32-q24.33) and two of which have been previously implicated (6p21.32-p22.1 and 18q21.2). The strongest new finding ($P = 1.6 \times 10^{-11}$) was with rs1625579 within an intron of a putative primary transcript for *MIR137* (microRNA 137), a known regulator of neuronal development. Four other schizophrenia loci achieving genome-wide significance contain predicted targets of *MIR137*, suggesting *MIR137*-mediated dysregulation as a previously unknown etiologic mechanism in schizophrenia. In a joint analysis with a bipolar disorder sample (16,374 affected individuals and 14,044 controls), three loci reached genome-wide significance: *CACNA1C* (rs4765905, $P = 7.0 \times 10^{-9}$), *ANK3* (rs10994359, $P = 2.5 \times 10^{-8}$) and the *ITIH3-ITIH4* region (rs2239547, $P = 7.8 \times 10^{-9}$).



Open

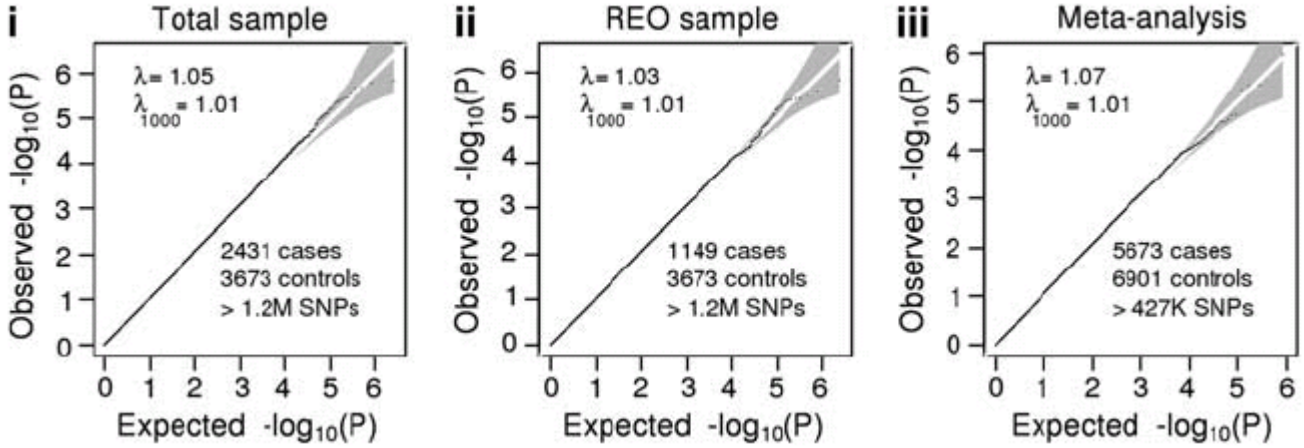
Molecular Psychiatry (2010), 1–13

www.nature.com/mp

ORIGINAL ARTICLE

Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned

NR Wray[1], ML Pergadia[2], DHR Blackwood[3], BWJH Penninx[4], SD Gordon[1], DR Nyholt[1], S Ripke[5,6], DJ MacIntyre[3], KA McGhee[3], AW Maclean[3], JH Smit[4], JJ Hottenga[4], G Willemsen[4], CM Middeldorp[4], EJC de Geus[4], CM Lewis[7], P McGuffin[7], IB Hickie[8], EJCG van den Oord[9], JZ Liu[1], S Macgregor[1], BP McEvoy[1], EM Byrne[1], SE Medland[1], DJ Statham[1,11], AK Henders[1], AC Heath[2], GW Montgomery[1], NG Martin[1], DI Boomsma[4], PAF Madden[2] and PF Sullivan[10]



proportion of total variance in liability. Larger study cohorts characterized for genetic and environmental risk factors accumulated prospectively are likely to be needed to dissect more fully the etiology of MDD.

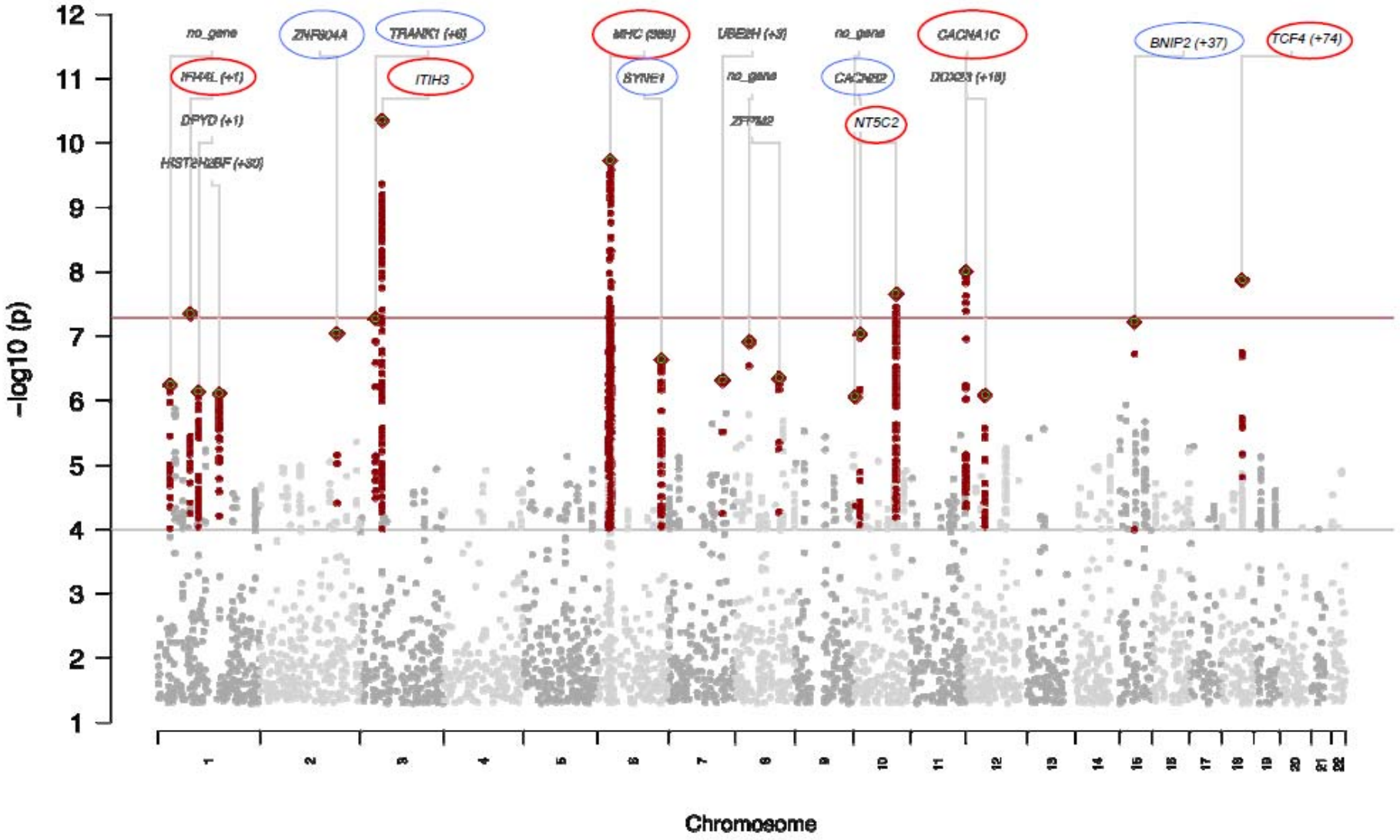
Cross-disorder mega-analysis

CDG: Jordan Smoller

First Analysis: Simple Pooled Analysis of Adult Disorders

Disorder	N Studies	N cases	N controls	Total N (%)
SCZ	17	9,372	7,815	17,817 (38%)
BPD	11	6,988	4,859	11,847 (26%)
MDD	9	9,229	7,347	16,576 (36%)
Total	37	25,589	20,021	45,610

Cross-disorder mega-analysis



GWAS of brain volumes (ADNI sample)

Alzheimer's Disease Neuroimaging Iniative (ADNI)
- mixed sample of healthy controls, MCI, AD

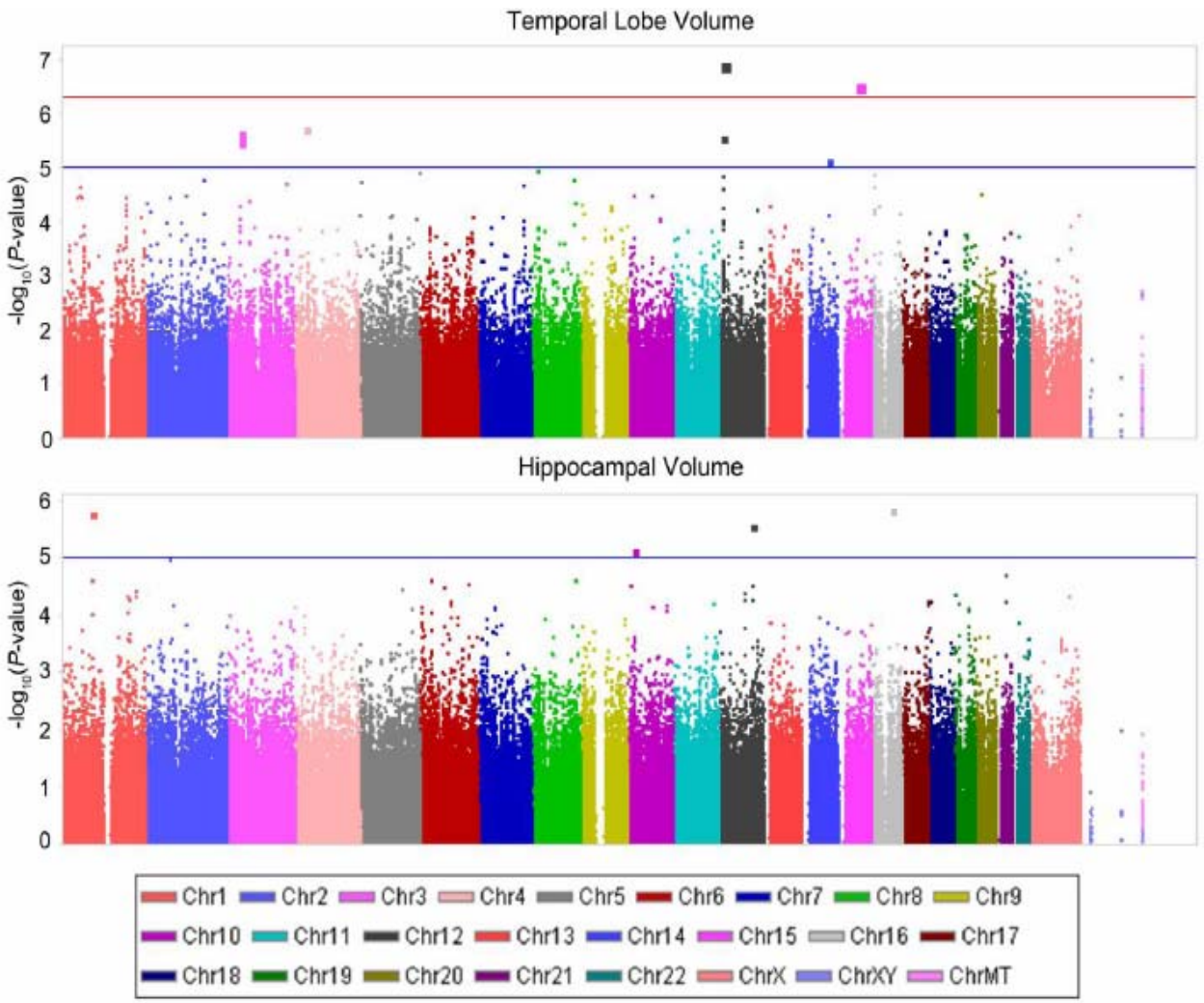
N = 742 (temporal)
N = 698 (hippo)

610K Illumina SNP

Genome –wide evidence or support - chrm. 12

Lower temporal lobe vols were most assoc. with a common variant in GRIN2B .

Risk allele over-represented in AD and MCI vs elderly controls



Stein et al. Neuroimage, 2010

ENIGMA (Enhancing Neuroimaging Genetics through Meta-Analysis)

first meta-analysis on the hippocampus has been completed.

17 cohorts of European ancestry from whom genome-wide single nucleotide polymorphisms (SNPs) and structural MRI data were collected.

Unselected population samples and case-control studies were included, with cases ascertained for neuropsychiatric disorders including depression, anxiety, Alzheimer's disease and schizophrenia.

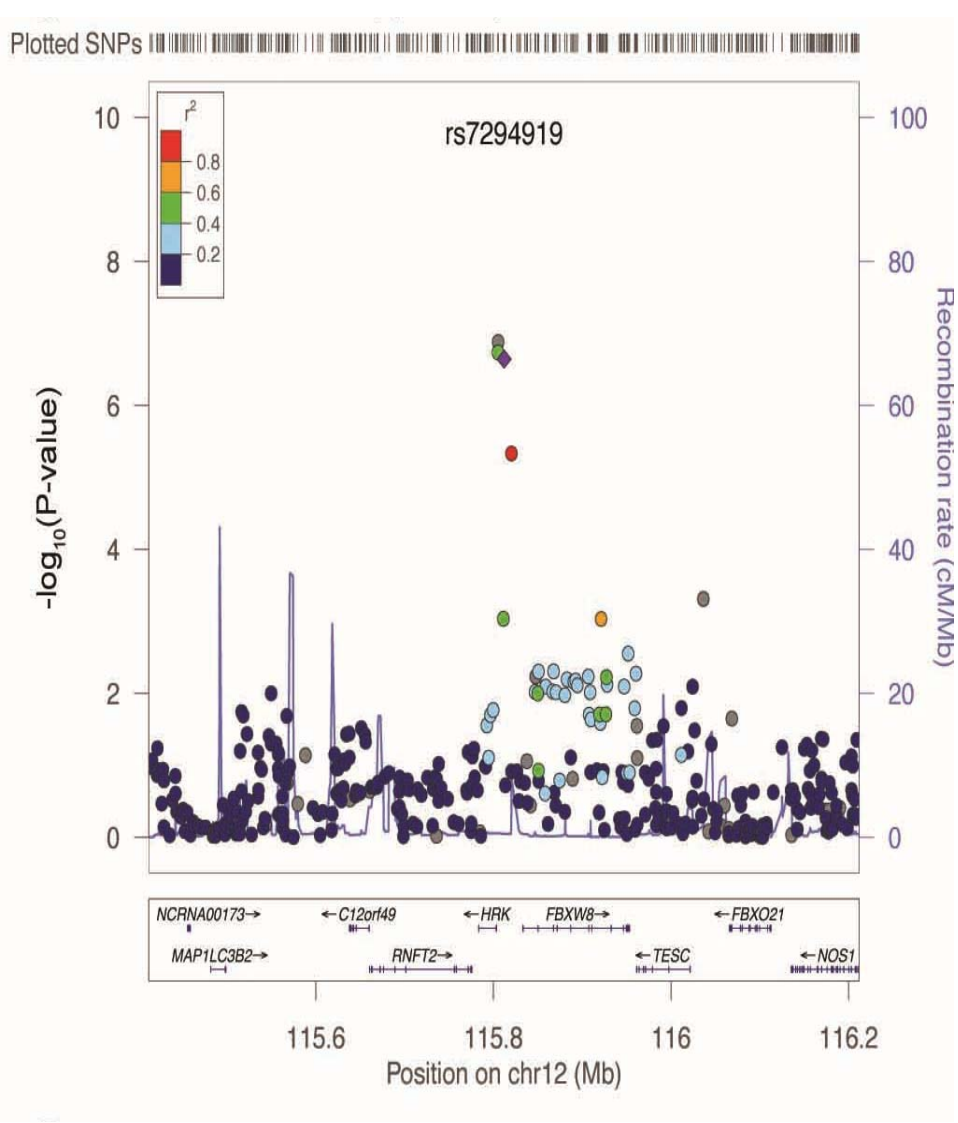
To distinguish whether putative effects at loci varied with disease status, analyses were run in the full sample (*N*=7,795) and in a healthy subsample (*N*=5,776).

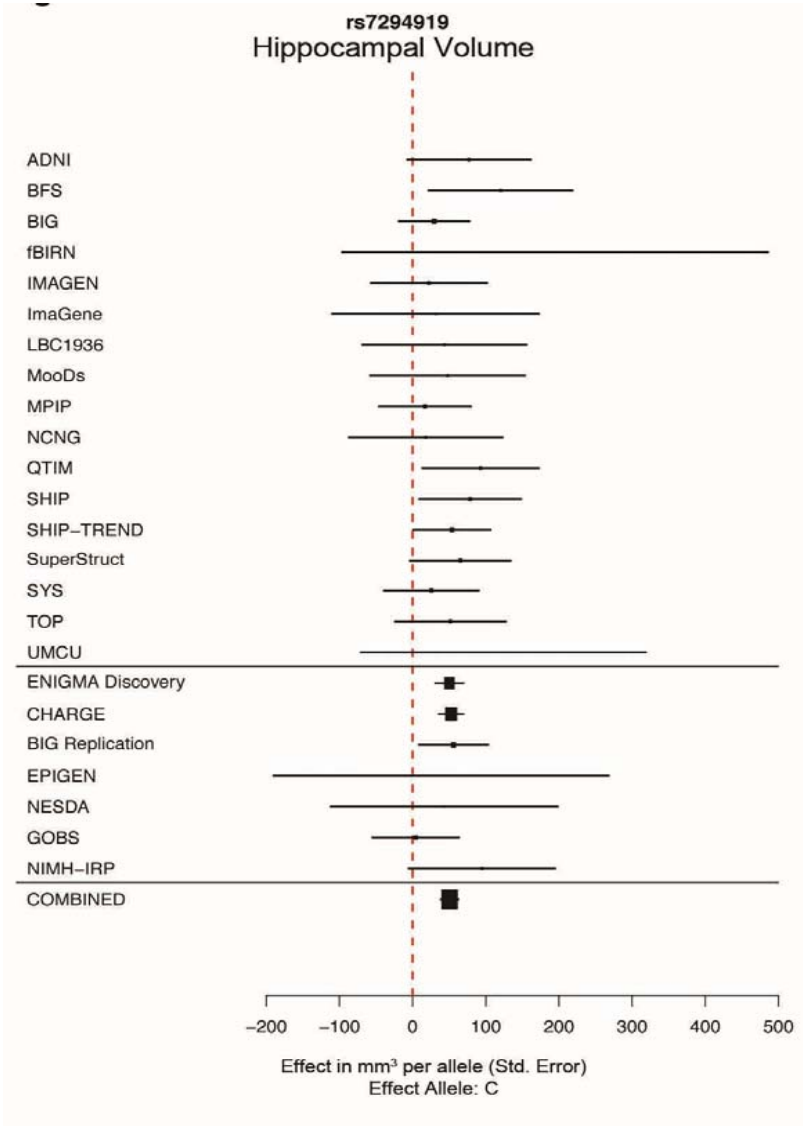
Replicated in CHARGE sample (N~9000)

Next project is 7 subcortical structures:
caudate, putamen, pallidum, thalamus, accumbens, amygdala (and hippocampus)

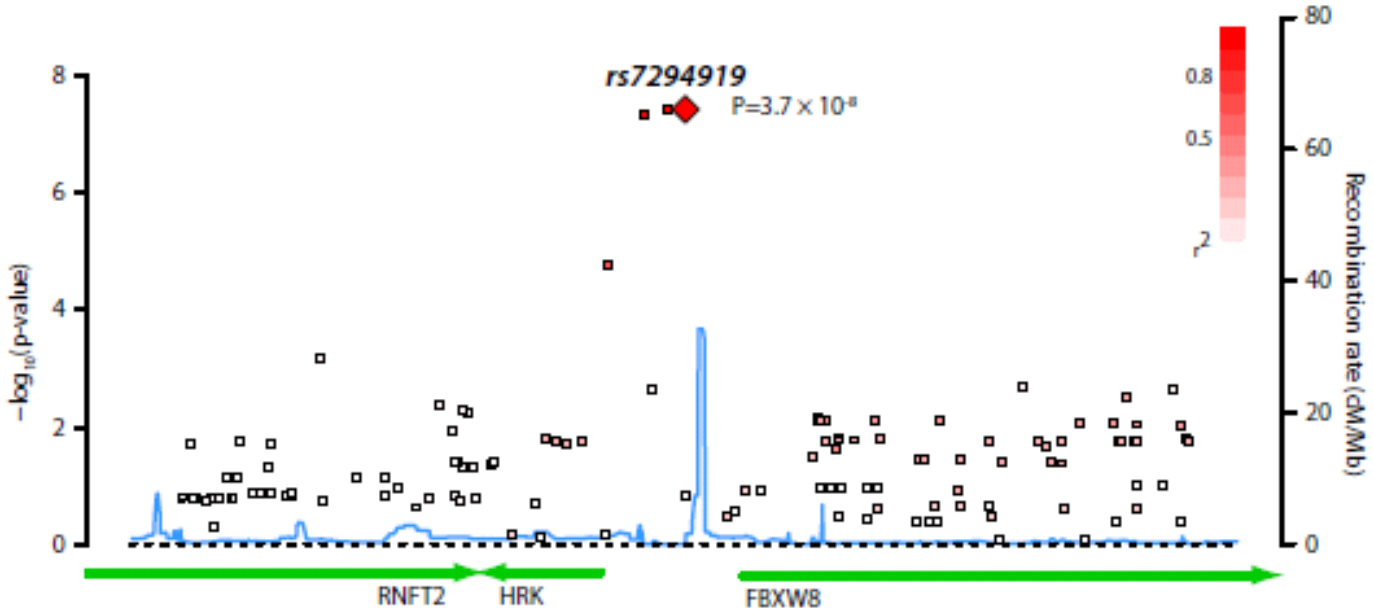
http://enigma.loni.ucla.edu

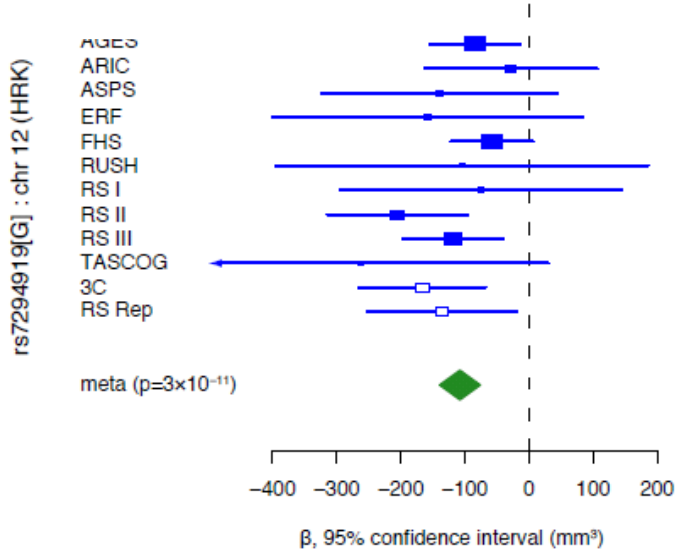
ENIGMA GWAS meta-analysis for hippocampal volume (*N*=7,795)





Chromosome 12 (HRK)



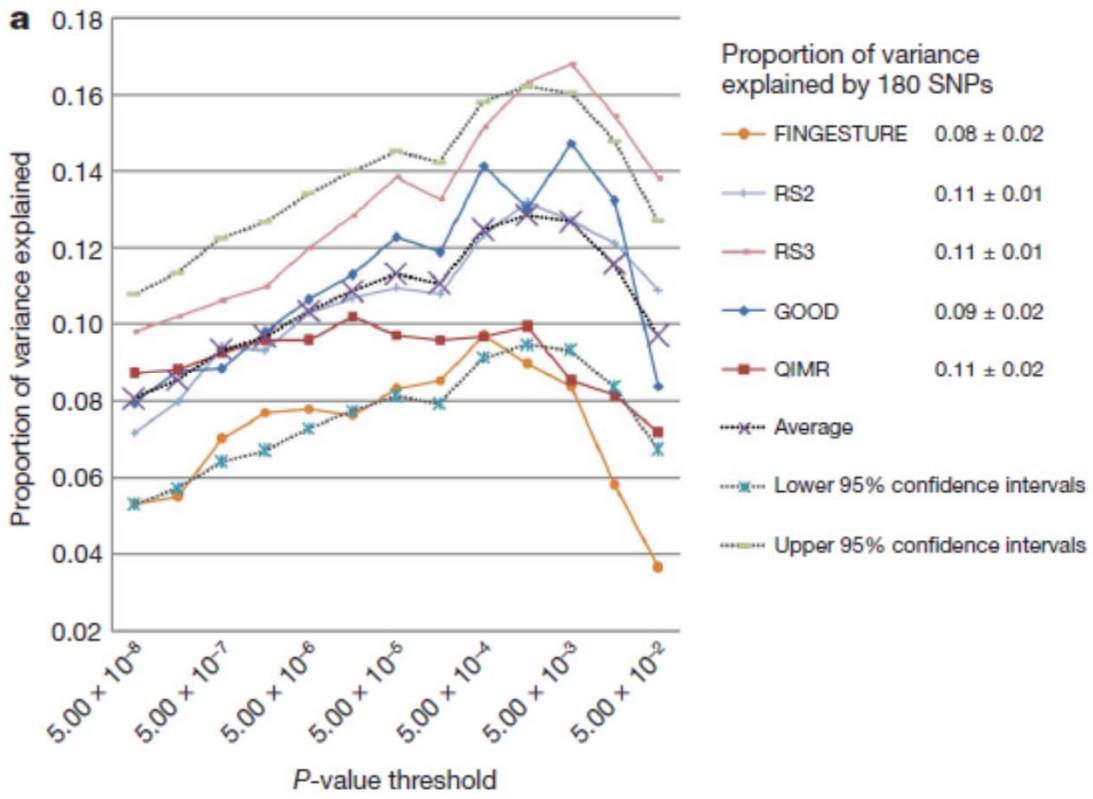


Top hit for hippocampal volume replicated in CHARGE

LETTER

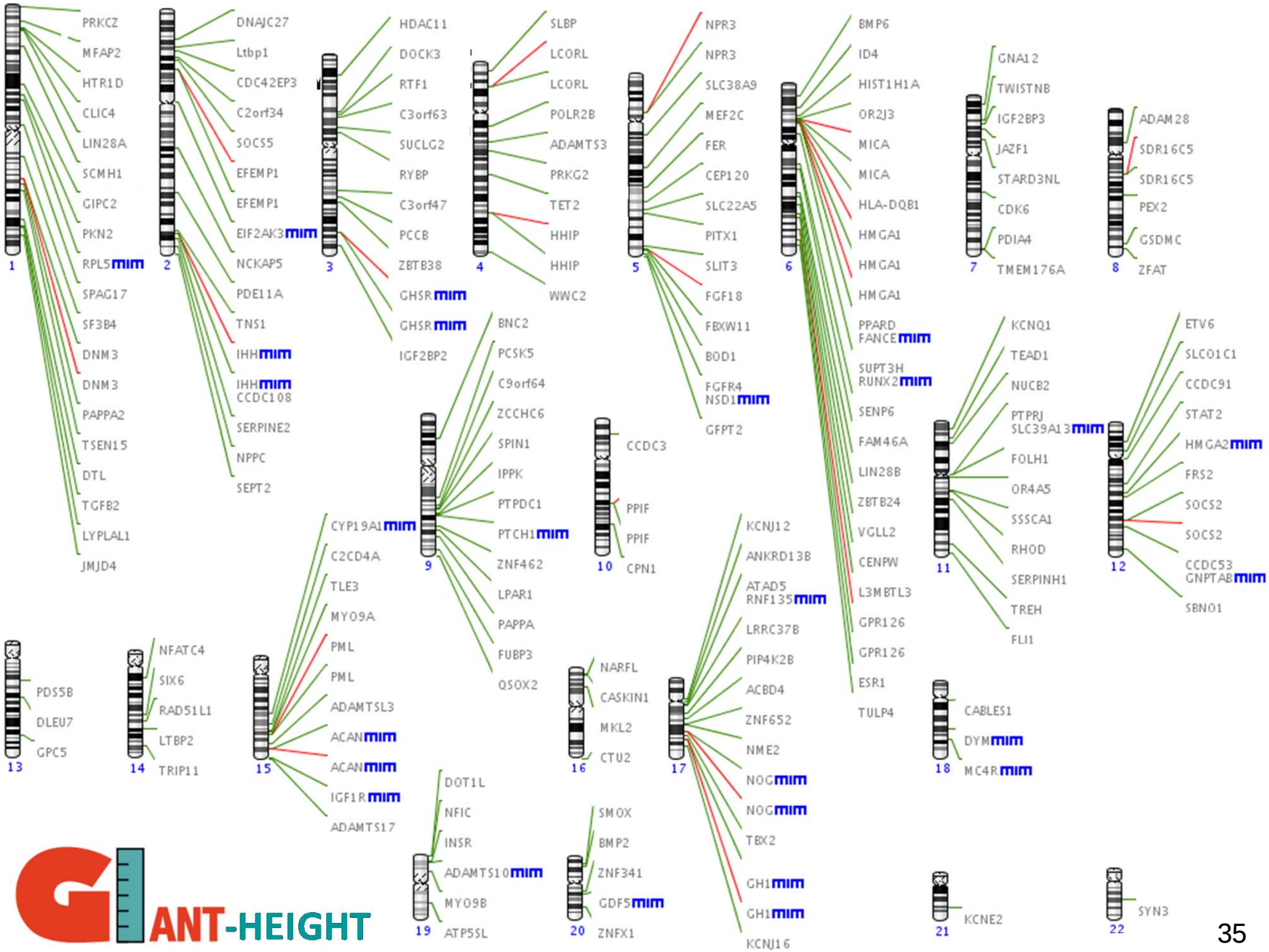
doi:10.1038/nature09410

Hundreds of variants clustered in genomic loci and biological pathways affect human height



GIANT Consortium - Height

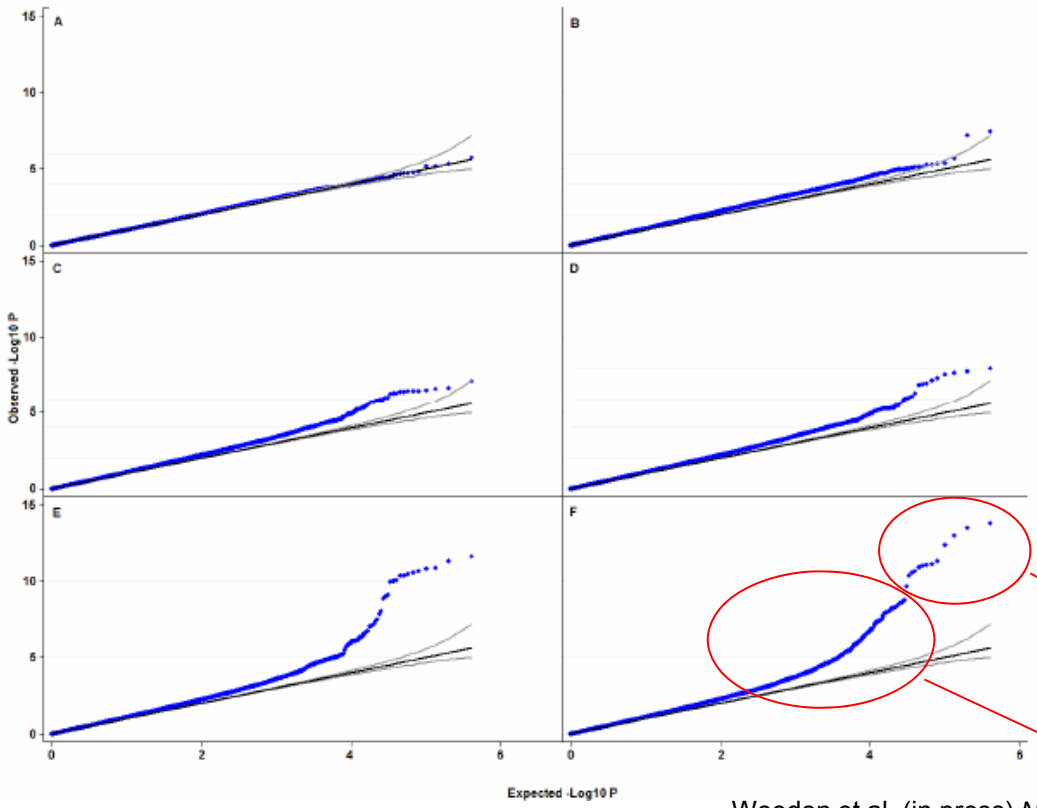
- 180,000 individuals
- 180 loci identified
- Allelic effect sizes 1 to 4 mm
- Enriched for genes that are connected in biological pathways that underlie skeletal growth
- BUT only ~12% of heritability explained !



[Lango Allen *et al*. Nature 2010]

GWAS of Height

Nat Genet. 2008 May;40(5):575-83.
Genome-wide association analysis identifies 20 loci that influence adult height.
Weedon MN, ...**Evans DM,**, , Frayling TM.



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

A- 1914 Cases (WTCCC T2D)

B- 4892 Cases (DGI)

C- 6788 Cases (WTCCC HT)

D- 8668 Cases (WTCCC CAD)

E- 12228 Cases (EPIC)

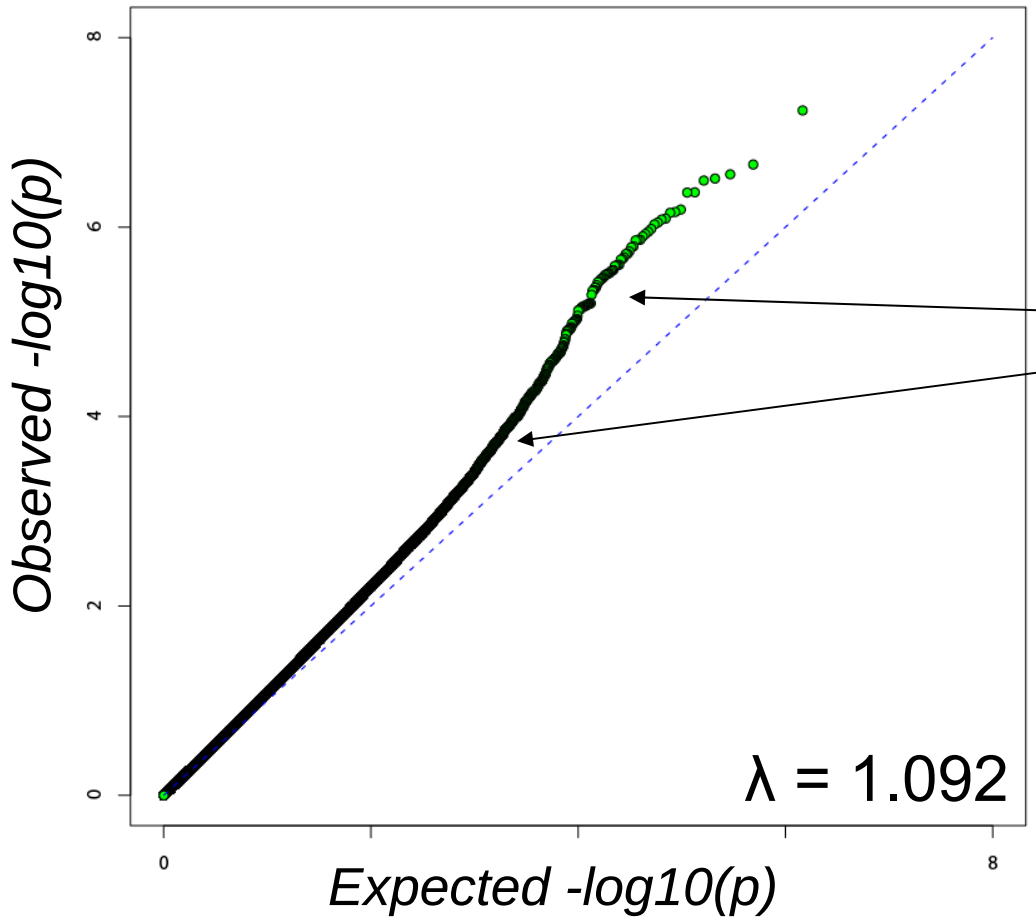
F- 13665 Cases (WTCCC UKBS)

Significant results

Other loci?

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

Schizophrenia (ISC) Q-Q plot



Consistent with:

Stratification?

Genotyping bias?

Distribution of true polygenic effects?

Genomic inflation factors under polygenic inheritance

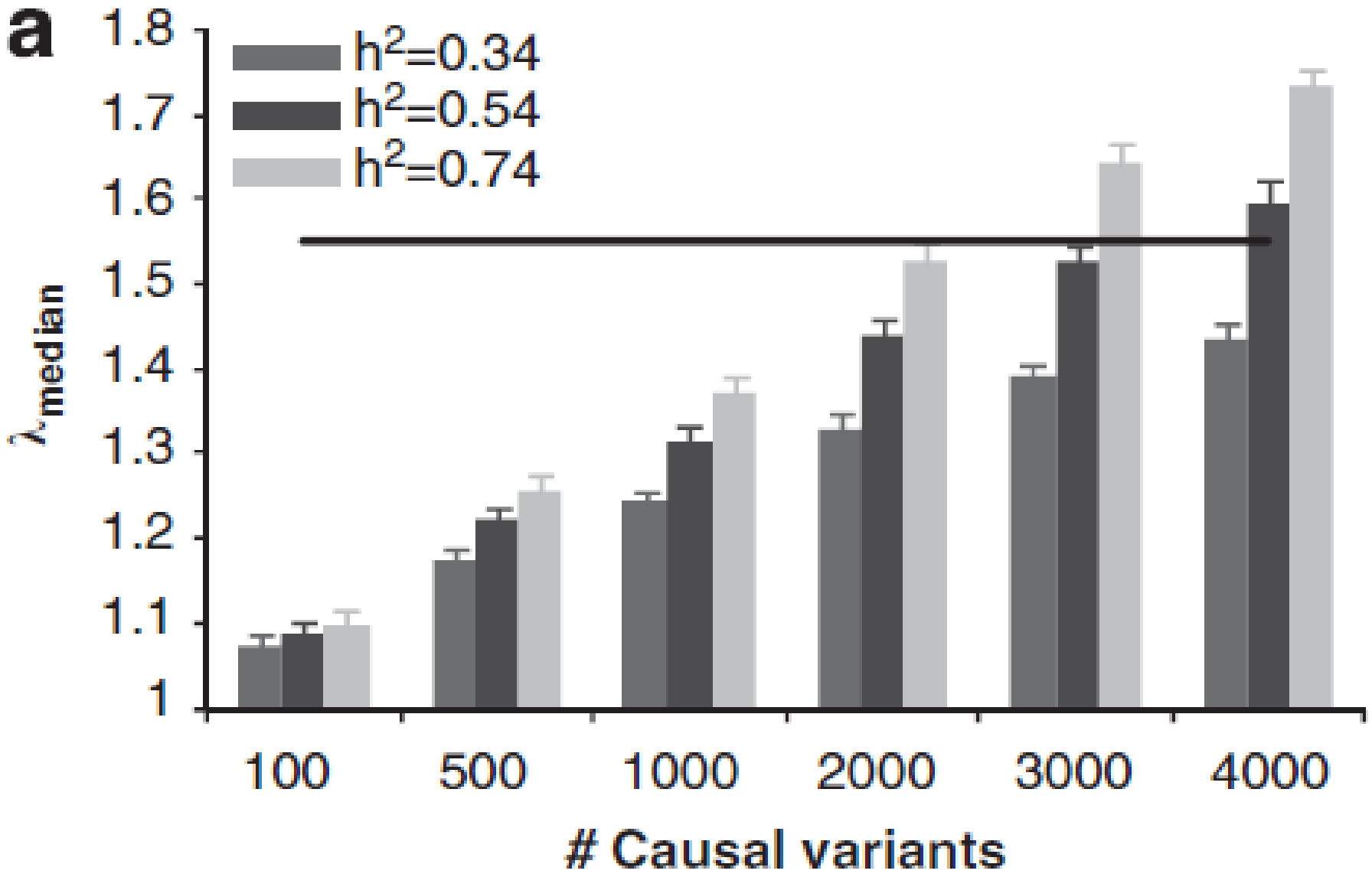
Jian Yang*[,1], Michael N Weedon[2], Shaun Purcell[3,4], Guillaume Lettre[5], Karol Estrada[6], Cristen J Willer[7], Albert V Smith[8], Erik Ingelsson[9], Jeffrey R O'Connell[10], Massimo Mangino[11], Reedik Mägi[12], Pamela A Madden[13], Andrew C Heath[13], Dale R Nyholt[1], Nicholas G Martin[1], Grant 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

Molecular Psychiatry (2012) 17, 337–349

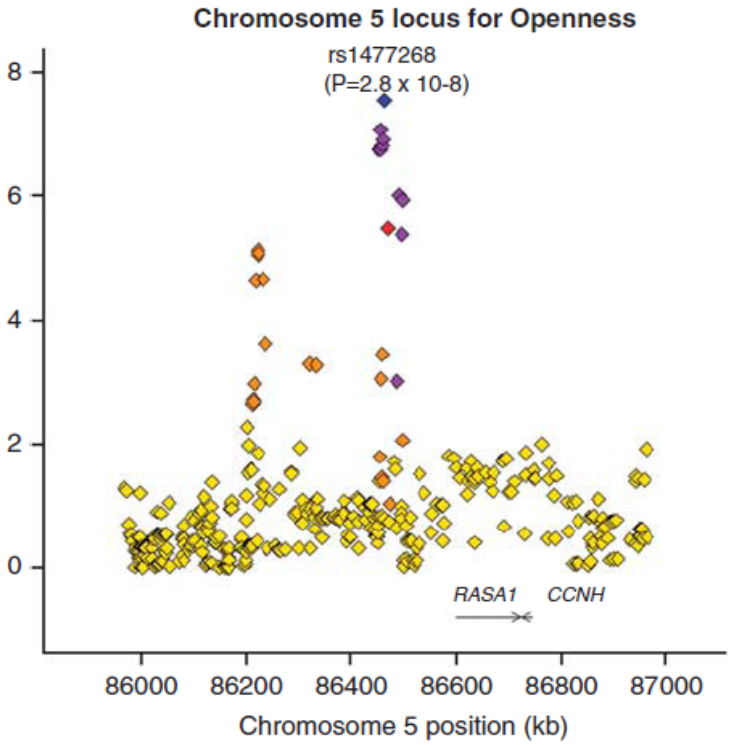
www.nature.com/mp

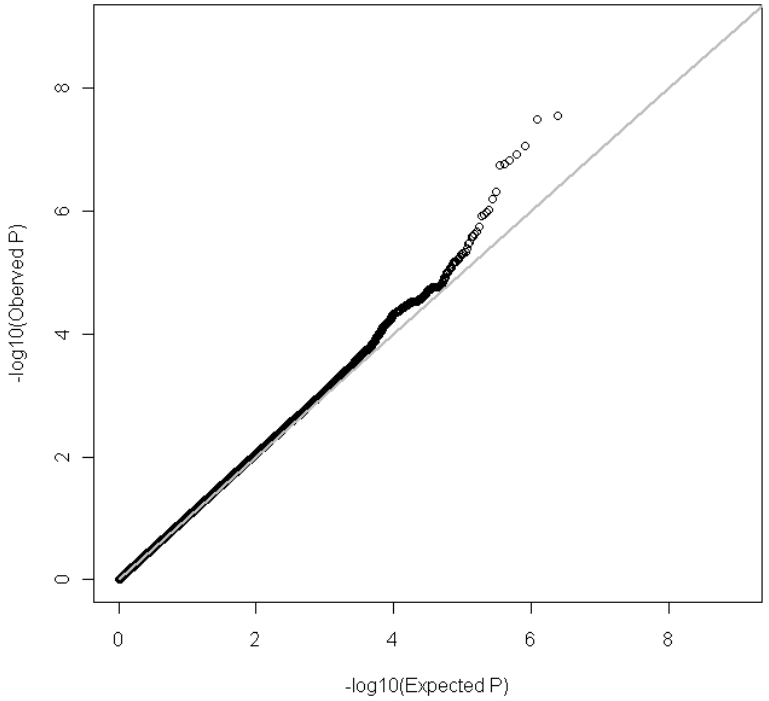
ORIGINAL ARTICLE

Meta-analysis of genome-wide association studies for personality

N = 17,000

MHM de Moor[1], PT Costa[2], A Terracciano[2], RF Krueger[3], EJC de Geus[1], T Toshiko[2],

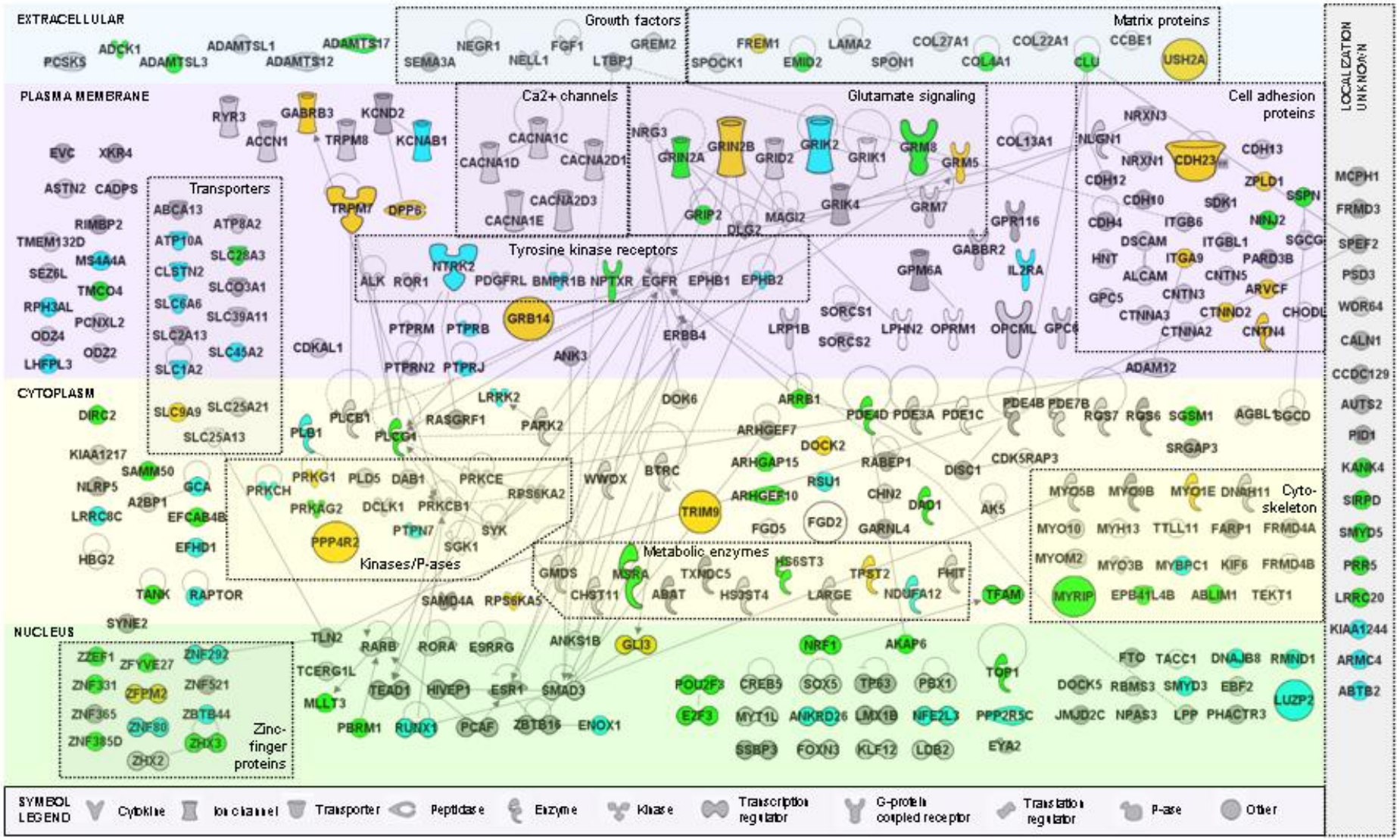




GWILL studies (pace MCN)

- **Personality** – using Item Response Modeling to map different personality scales (EPQ, TPQ, NEO, MPQ) on to each other so samples can be combined – N~60,000 - and more wanted! (Marleen de Moor, Stephanie Vandenberg, Dorret Boomsma)
- **Educational Attainment** – CHARGE Social Science Consortium: N~100,000 (Sarah Medland, Jaime Derringer, Niels Rietvelt, Philip Koellinger, David Caesarini)
- **Need more samples for both !!!**

Pathway (Ingenuity) analysis of GWAS for smoking



Vink et al, Am J Hum Genet 84:367-79,2009

How much variance have GWAS studies explained?

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]	

Variance explained by GWAS for selected complex traits

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

GWAS' greatest success: T1D

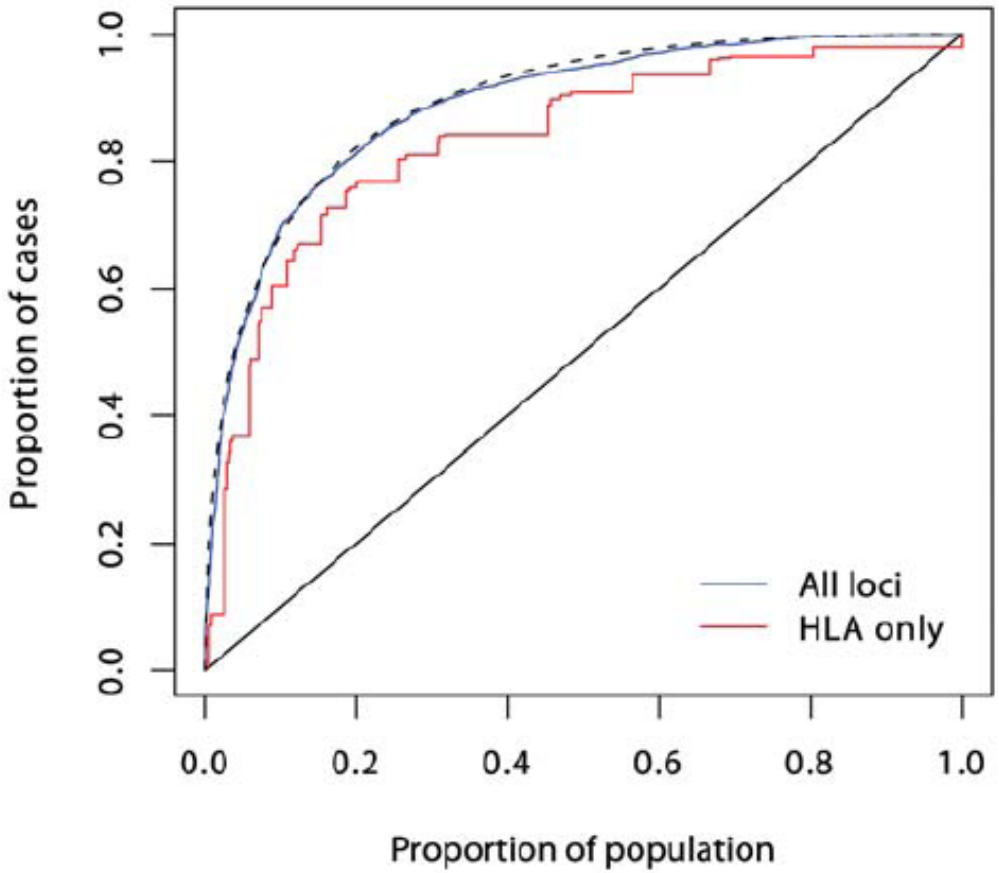


Figure 5. ROC curve prediction from all the SNPs listed in Supplementary Table 1 in Text S1 (in blue). The prediction curve using the six MHC SNPs alone is shown in red, and the dashed curve corresponds to a polygenic multiplicative model with $\lambda_s = 4.75$.

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.

PLOS GENETICS

OPEN ACCESS Freely available online

Viewpoints

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

David G. Clayton*

nature genetics

VOLUME 42 | NUMBER 7 | JULY 2010

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

A Commentary on 'Common SNPs Explain a Large Proportion of the Heritability for Human Height' by Yang et al. (2010)

Peter M. Visscher,[1] Jian Yang[1] and Michael E. Goddard[2,3]

Twin Research and Human Genetics Volume 13 Number 6 pp. 517–524

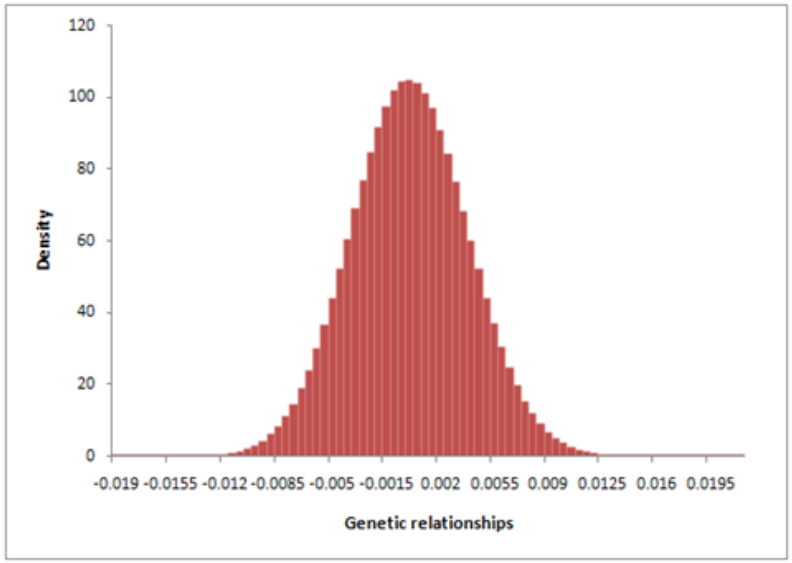
REPORT

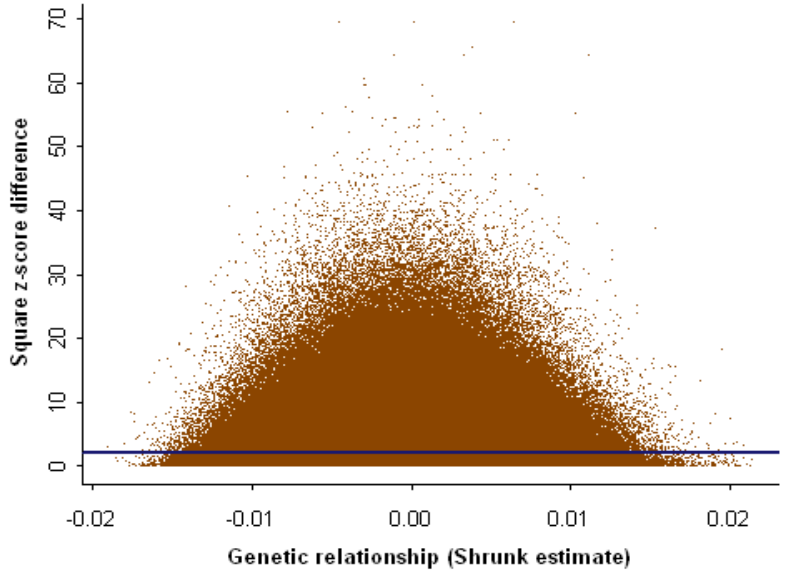
GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

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

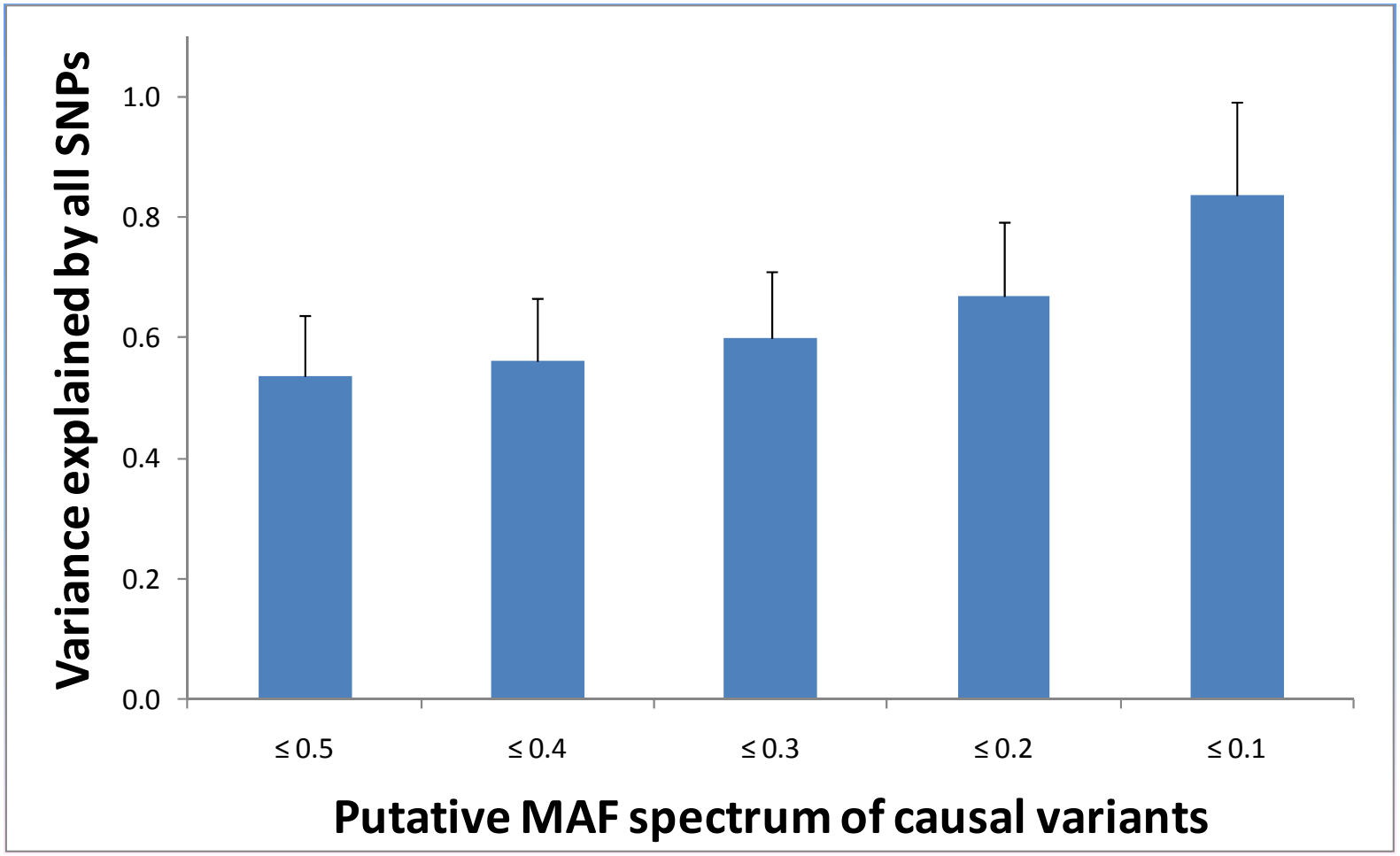




Proportion of height variance tagged by SNPs ~ 0.55 (SE 0.1)

Yang et al. Nature Genetics 2010

More than 50% of genetic variation is tagged by common SNPs



Evidence that causal variants have lower MAF than genotyped SNPs

Molecular Psychiatry (2011), 1–10

www.nature.com/mp

IMMEDIATE COMMUNICATION

Genome-wide association studies establish that human intelligence is highly heritable and polygenic

G Davies[1], A Tenesa[2,3], A Payton[4], J Yang[5], SE Harris[6,7], D Liewald[1,7], X Ke[8], S Le Hellard[9], A Christoforou[9], M Luciano[1,7], K McGhee[1], L Lopez[1,7], AJ Gow[1,7], J Corley[1], P Redmond[1], HC Fox[10], P Haggarty[11], LJ Whalley[10], G McNeill[10], ME Goddard[12,13], T Espeseth[14], AJ Lundervold[15], I Reinvang[14], A Pickles[16], VM Steen[9,17], W Ollier[4], DJ Porteous[6,7], M Horan[18], JM Starr[7,19], N Pendleton[18], PM Visscher[5,7,20] and IJ Deary[1,7,20]

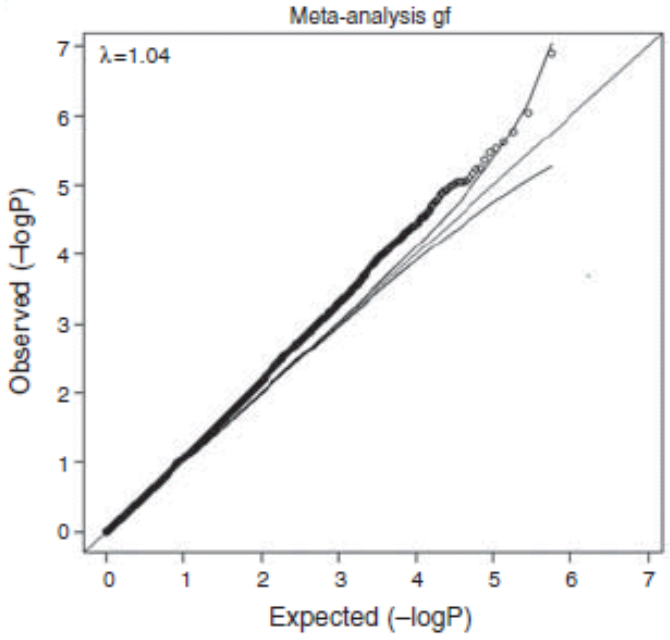


Table 1 Estimates of variance explained by all SNPs

	g_c	g_f
N	3254	3181
h^2 (s.e.)	0.40 (0.11)	0.51 (0.11)
P-value	5.7×10^{-5}	1.2×10^{-7}

Correction for imperfect LD and absence of rare SNPs will push this even higher (for height -> ~0.8)

Estimates of GW SNP-associated variance ~half that estimated from twin studies ?!

Phenotype	Variance explained by typed SNPs	Additive heritability
MDD	**21.1%**	**36%**
Smoking initiation	**12.2%**	**44%**
Current smoking	**42.1%**	**79%**
Fasting glucose	**25.4%**	**53%**
Height	**48.2%**	**90%**

Lubke & Boomsma, submitted , 2011

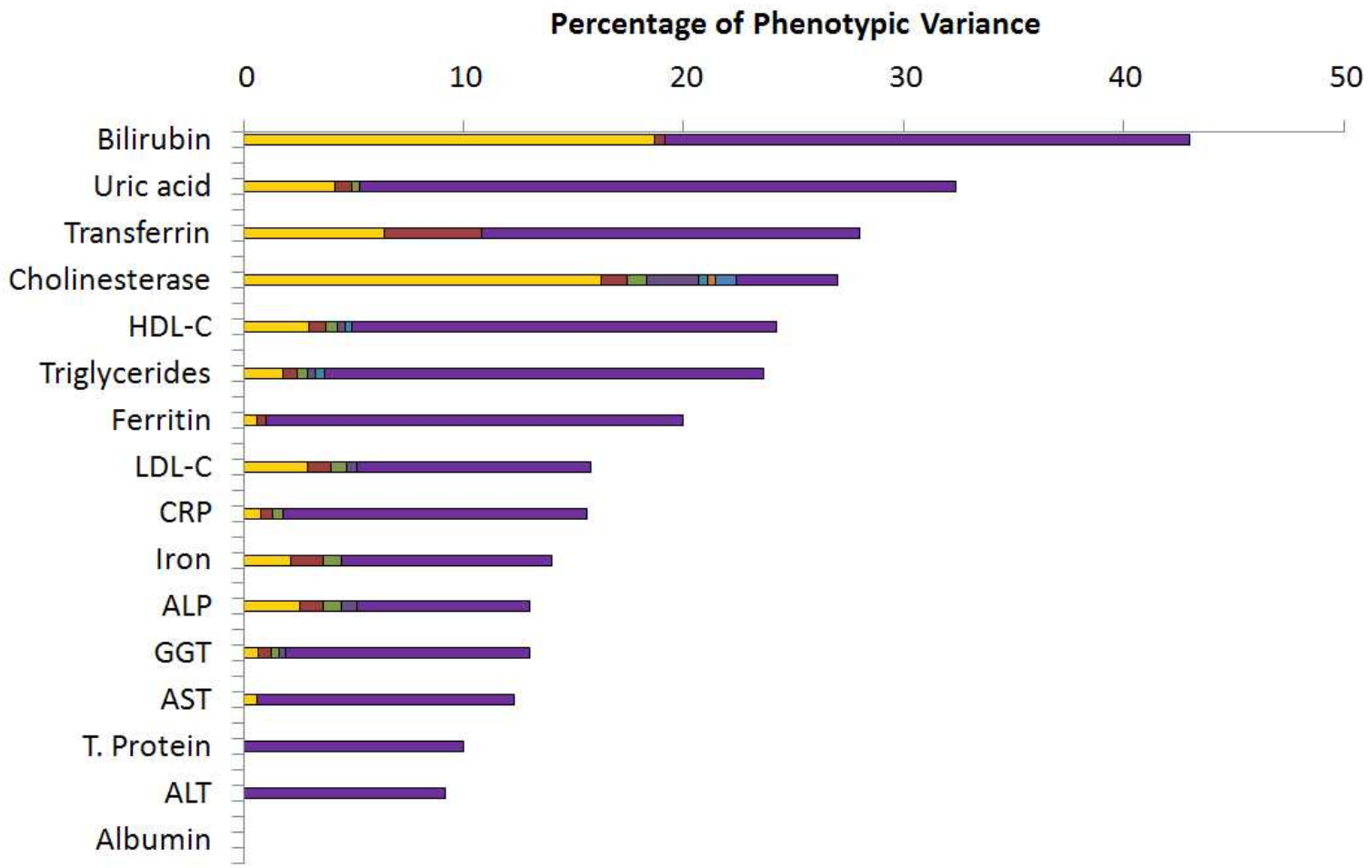
Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

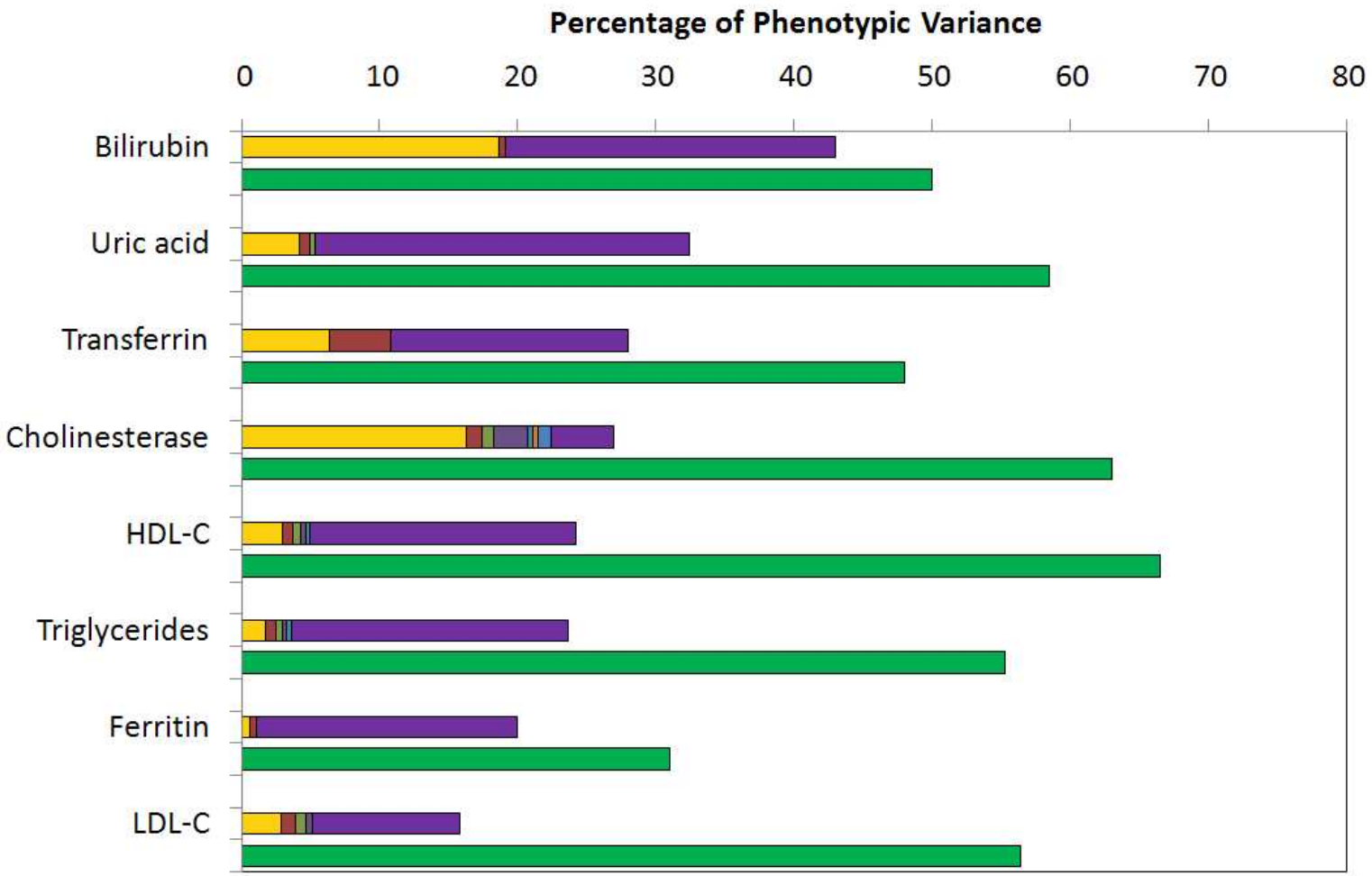
Phenotype	Variance explained by typed SNPs	Additive heritability
Crohn's disease	22-24%	50-60%
Bipolar Disorder	37-41%	60-85%
Type 2 diabetes	28-32%	30-70%

American Journal of Human Genetics *88*, 1–12, March 11, 2011

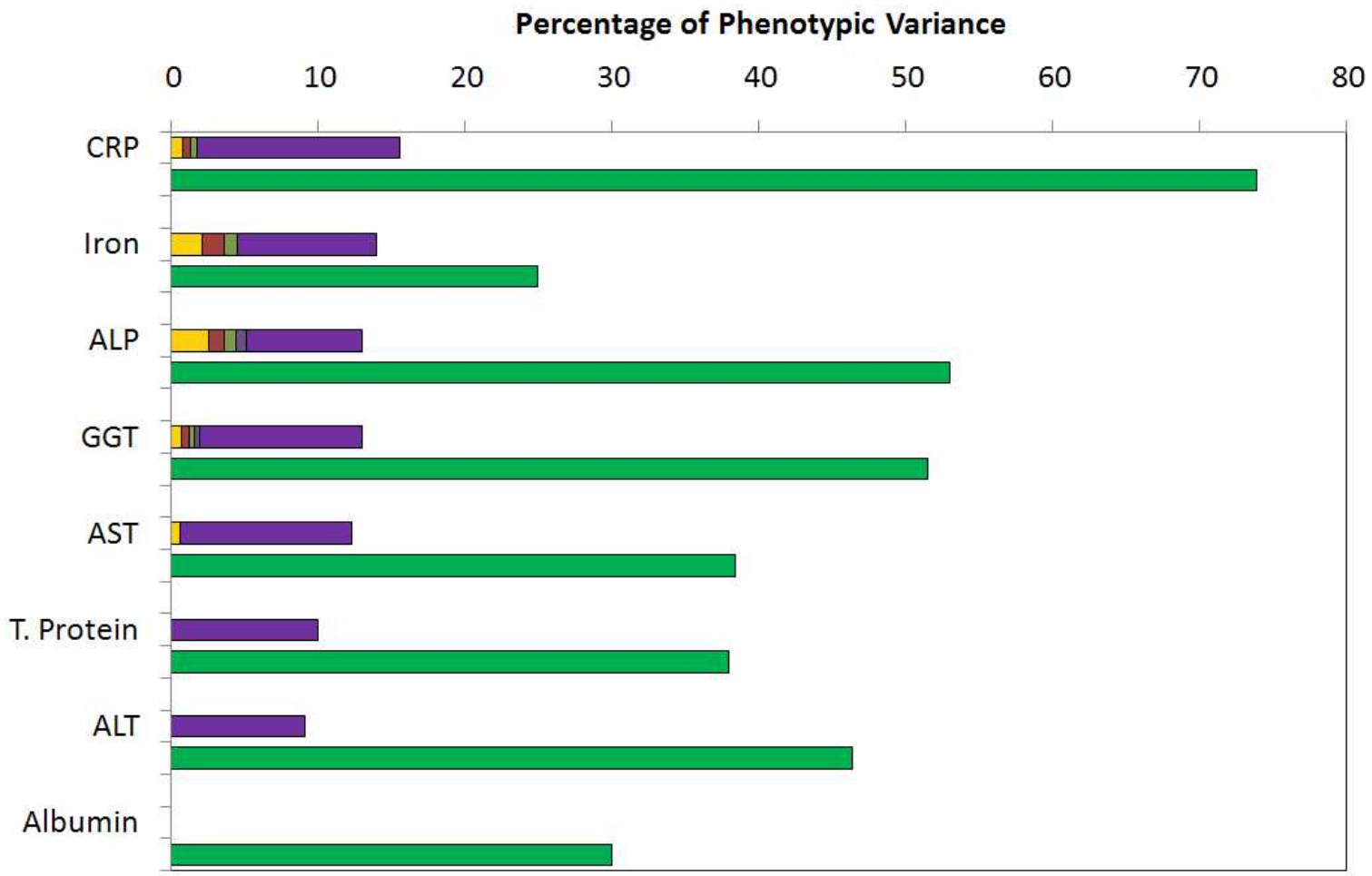
Total SNP-Associated Variance



SNP Variance *versus* Heritability (1)

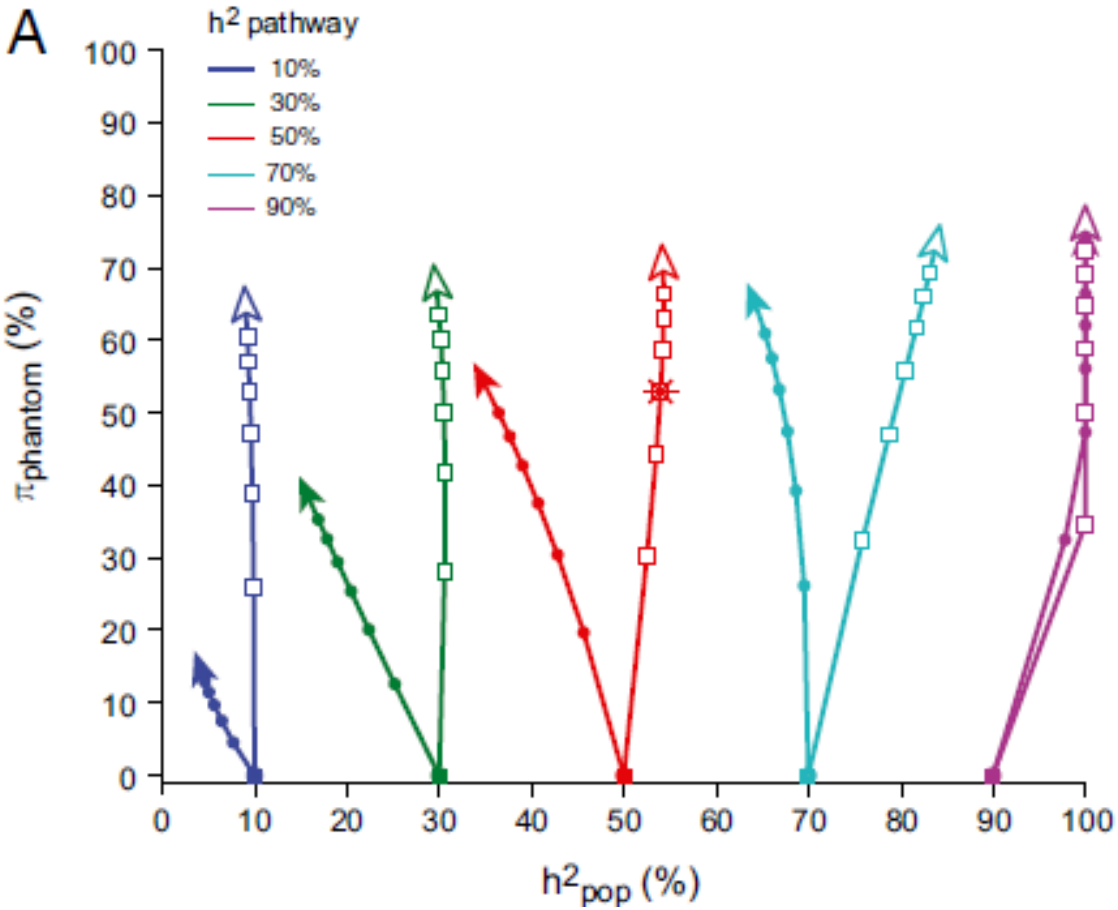


SNP Variance *versus* Heritability (2)



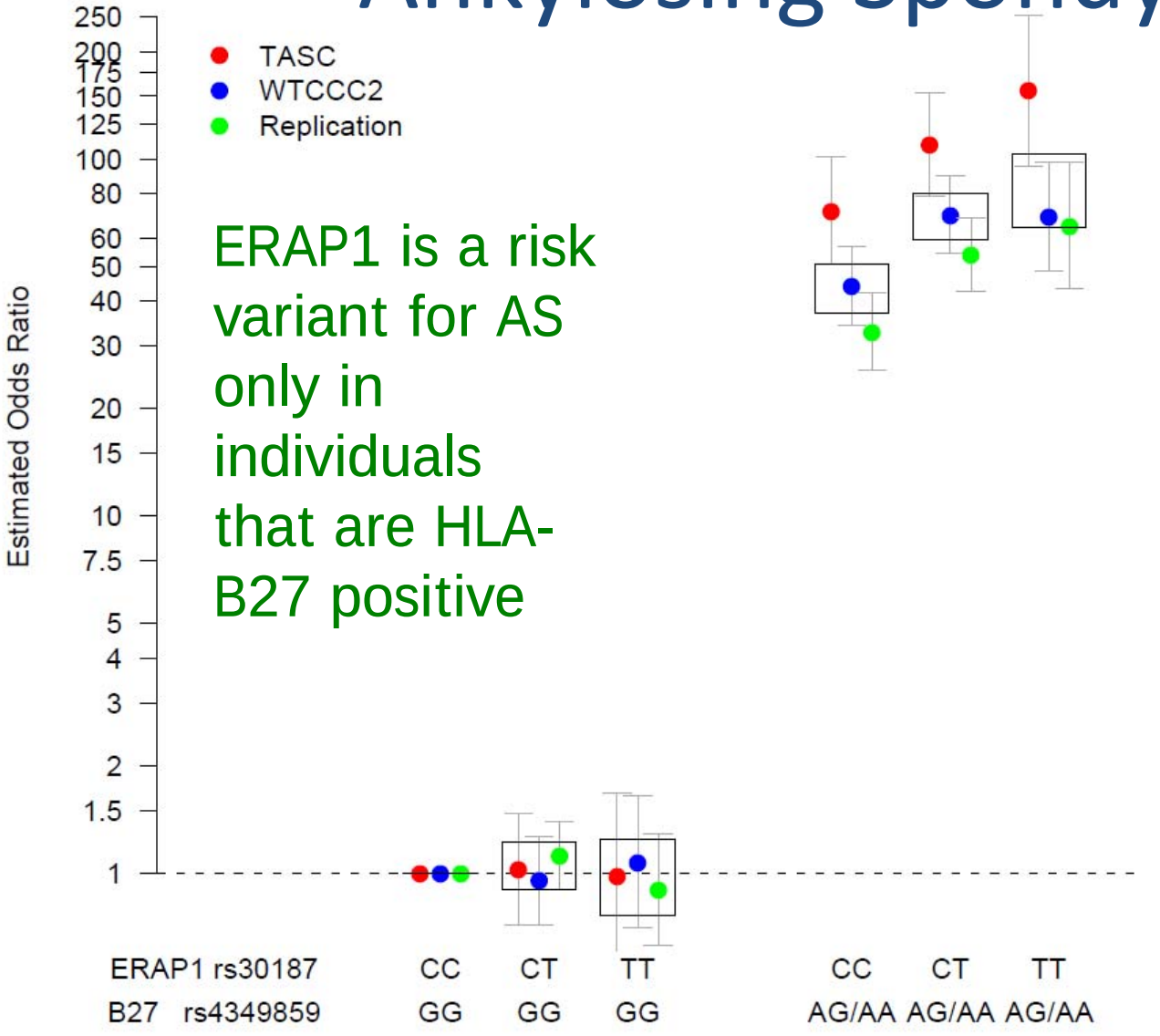
The mystery of missing heritability: Genetic interactions create phantom heritability

Or Zuk[a], Eliana Hechter[a], Shamil R. Sunyaev[a,b], and Eric S. Lander[a,1]

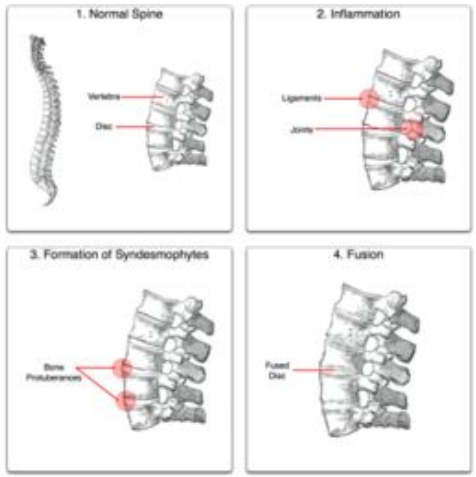


PNAS | January 24, 2012 | vol. 109 | no. 4 | 1193–1198

Gene-Gene Interaction and Ankylosing Spondylitis



ERAP1 is a risk variant for AS only in individuals that are HLA-B27 positive



Evans et al. (2012) Nat Genet

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]

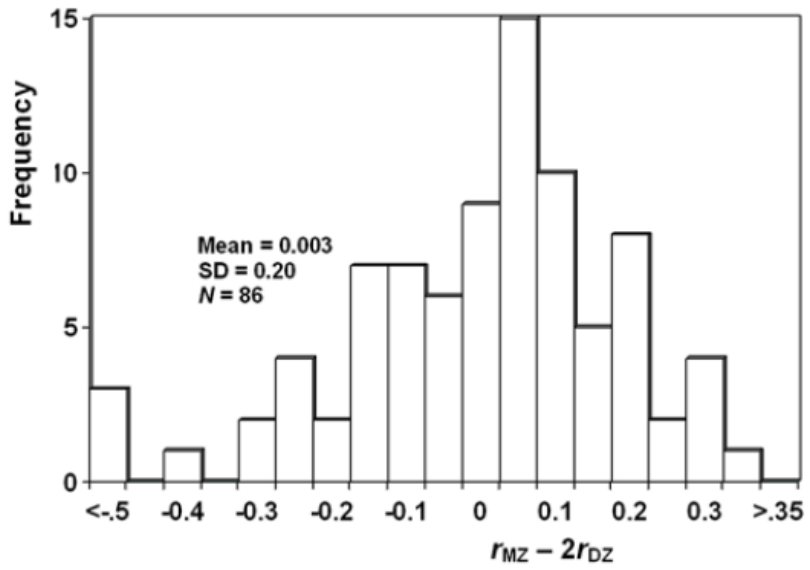
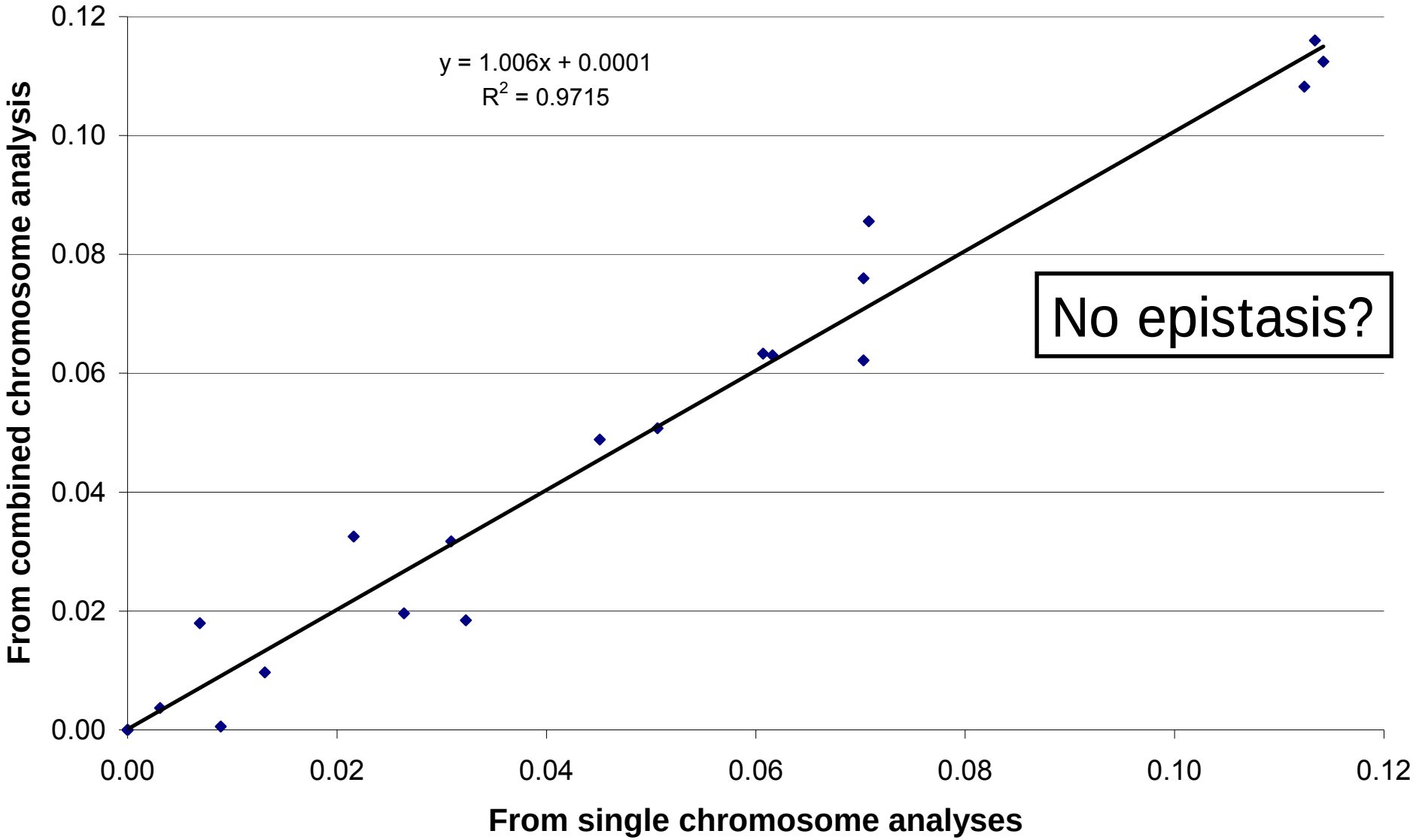


Figure 1. Distribution of $r_{MZ} - 2r_{DZ}$ for all traits on human twins.

Estimates of chromosomal heritabilities for height



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

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

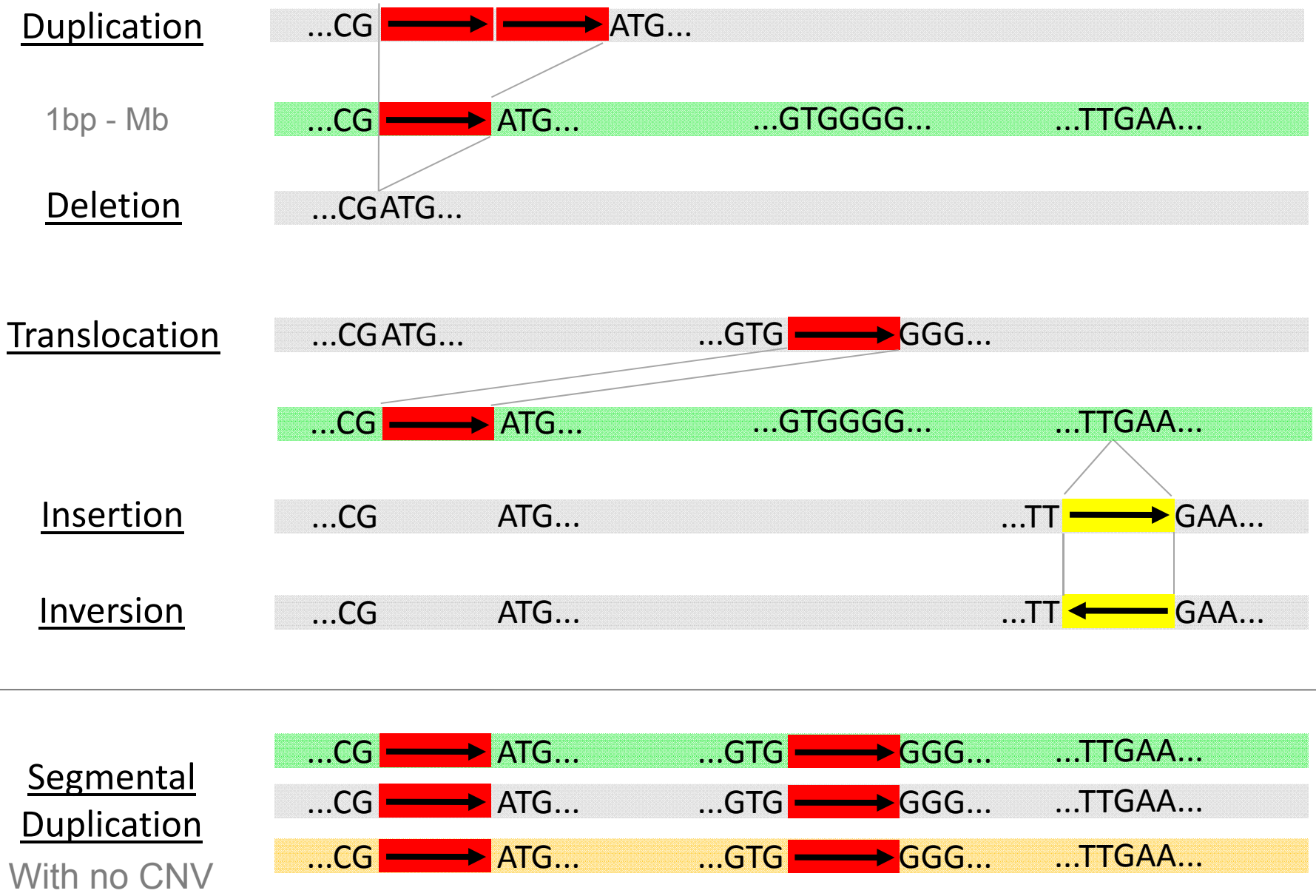
D Zhang[1], L Cheng[1], Y Qian[1], N Alliey-Rodriguez[1], JR Kelsoe[2], T Greenwood[2], C Nievergelt[2], TB Barrett[2], R McKinney[2], N Schork[3,4], EN Smith[3,4], C Bloss[3,4], J Nurnberger[5], HJ Edenberg[6,7], T Foroud[7], W Sheftner[8], WB Lawson[9], EA Nwulia[9], M Hipolito[9], W Coryell[10], J 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

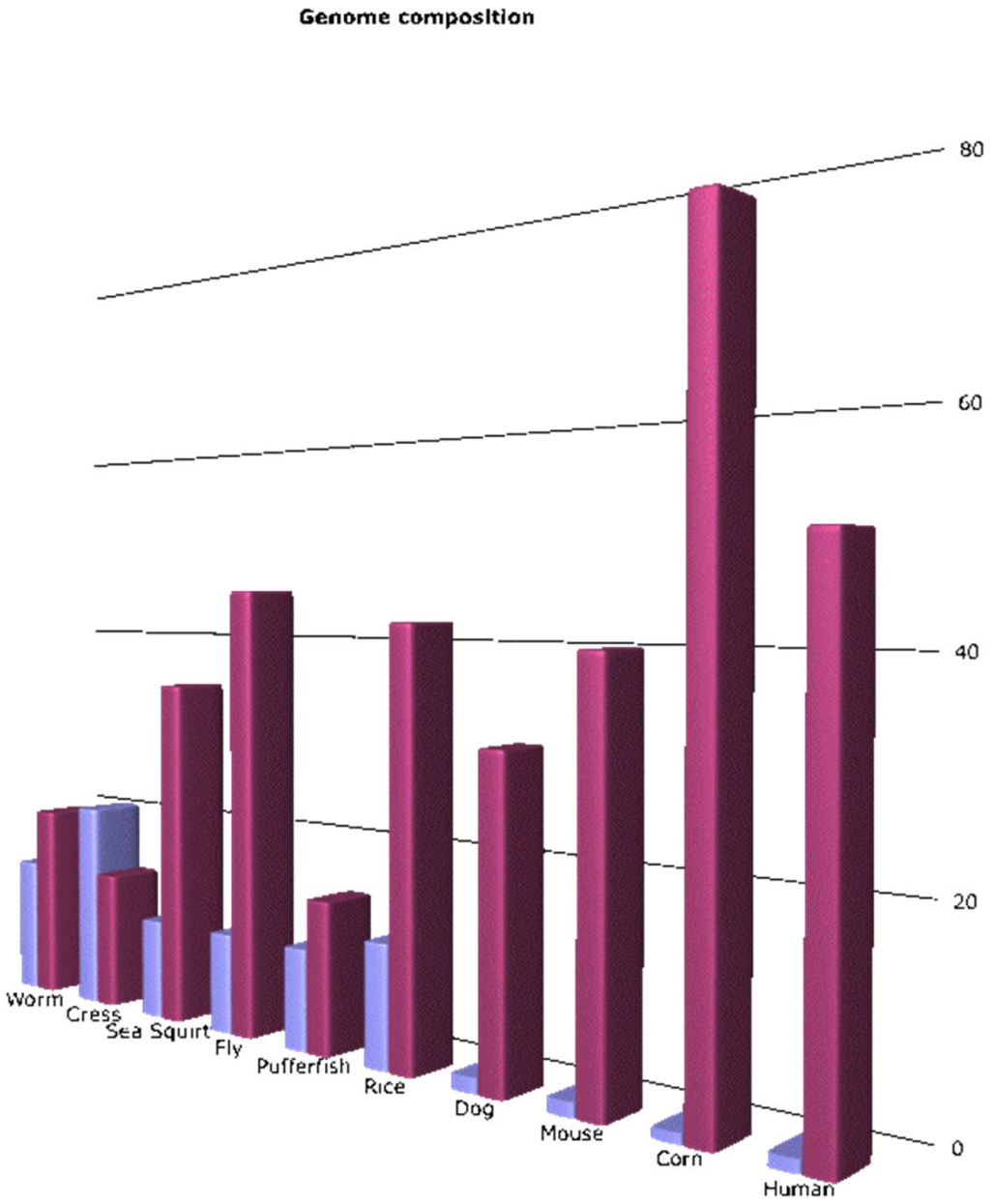
(in order of increasing plausibility ?)

- 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

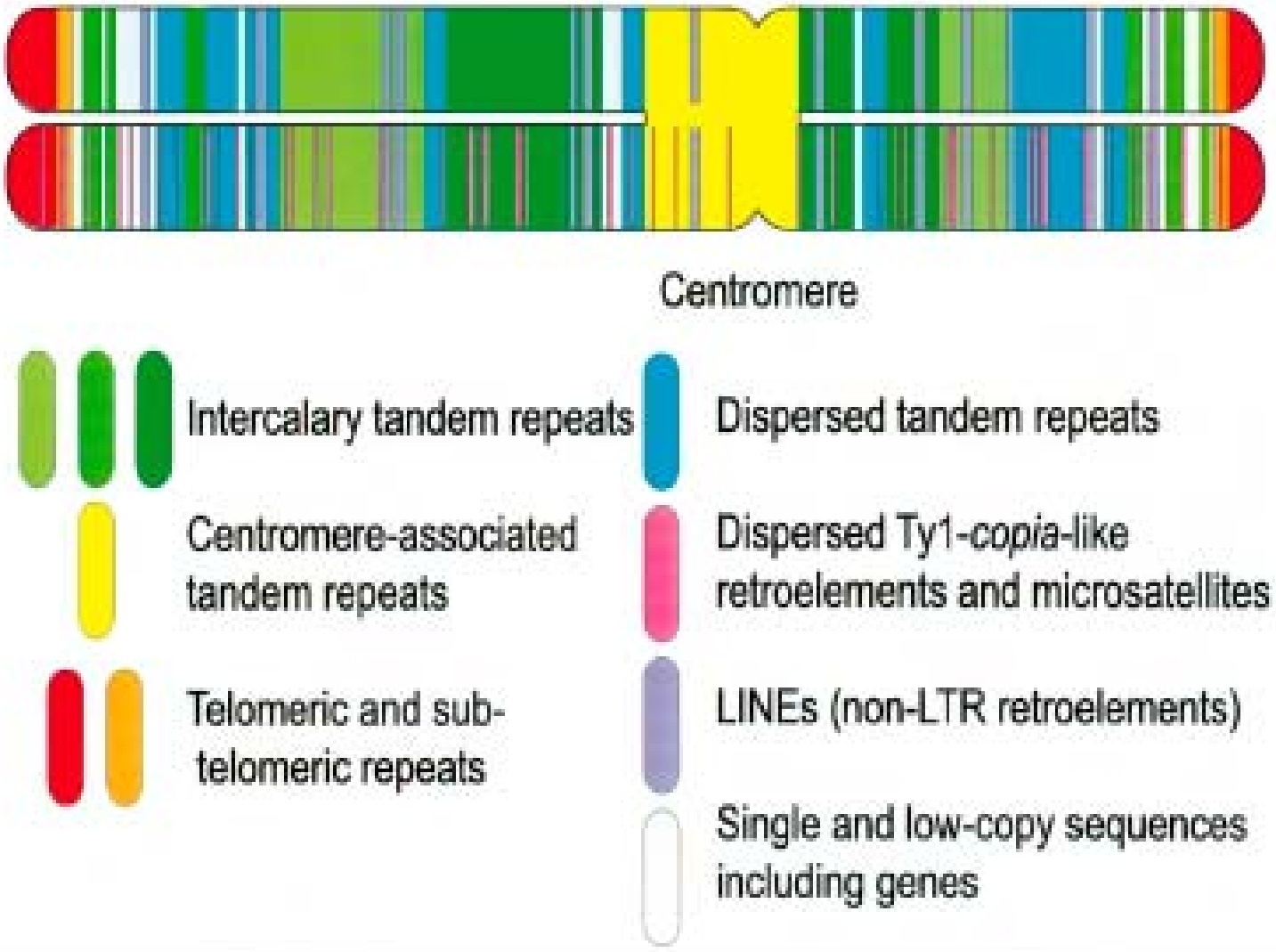

Duplication
...CG
ATG...
1bp - Mb
...CG
ATG...
...GTGGGG...
...TTGAA...
Deletion
...CG ATG...
Translocation
...CG ATG...
...GTG
GGG...
...CG
ATG...
...GTGGGG...
...TTGAA...
Insertion
...CG
ATG...
...TT
GAA...
Inversion
...CG
ATG...
...TT
GAA...
Segmental Duplication
With no CNV
...CG
ATG...
...GTG
GGG...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

50% of the human genome is repetitive DNA.

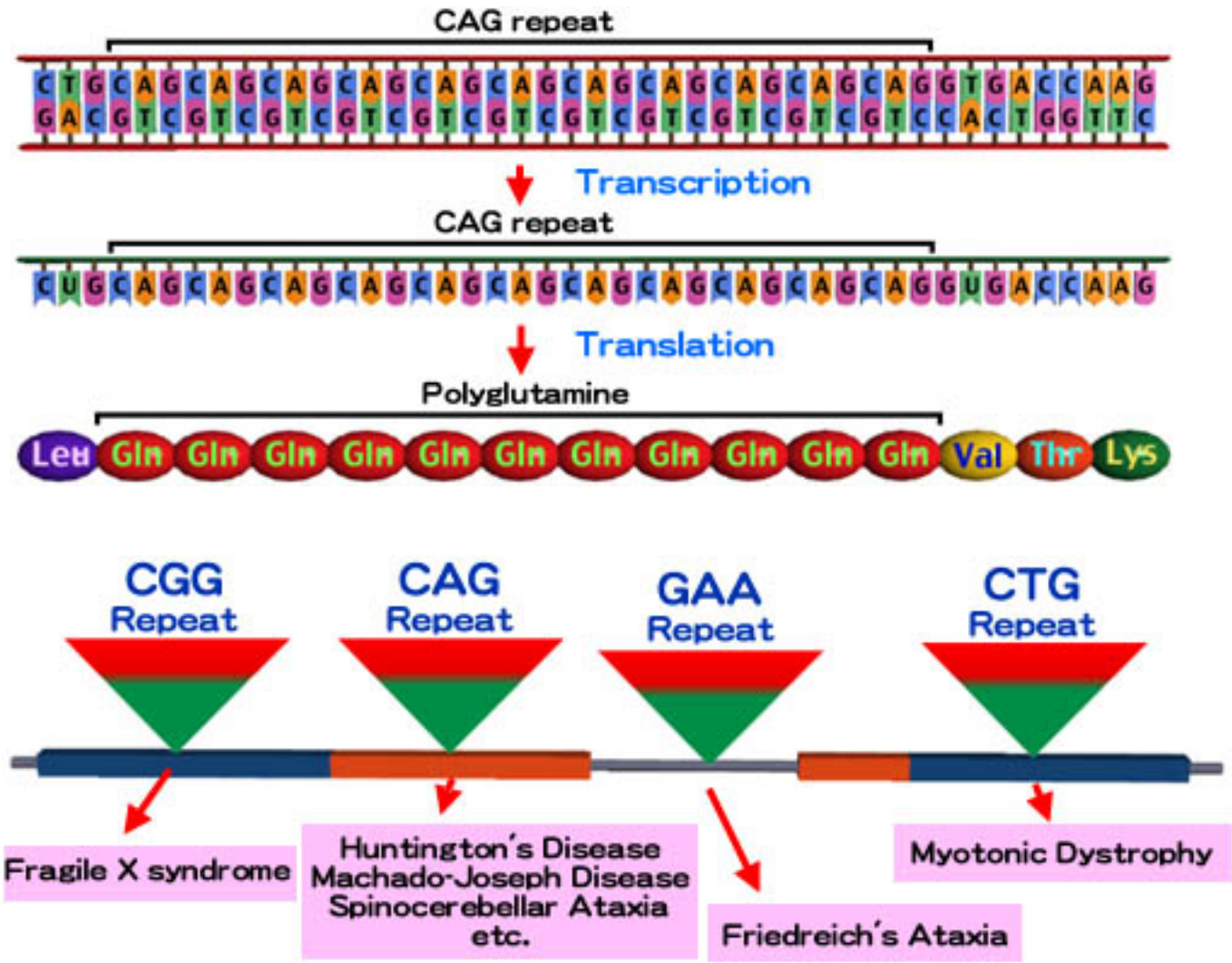
Only 1.2% is coding



Types of repetitive elements and their chromosomal locations

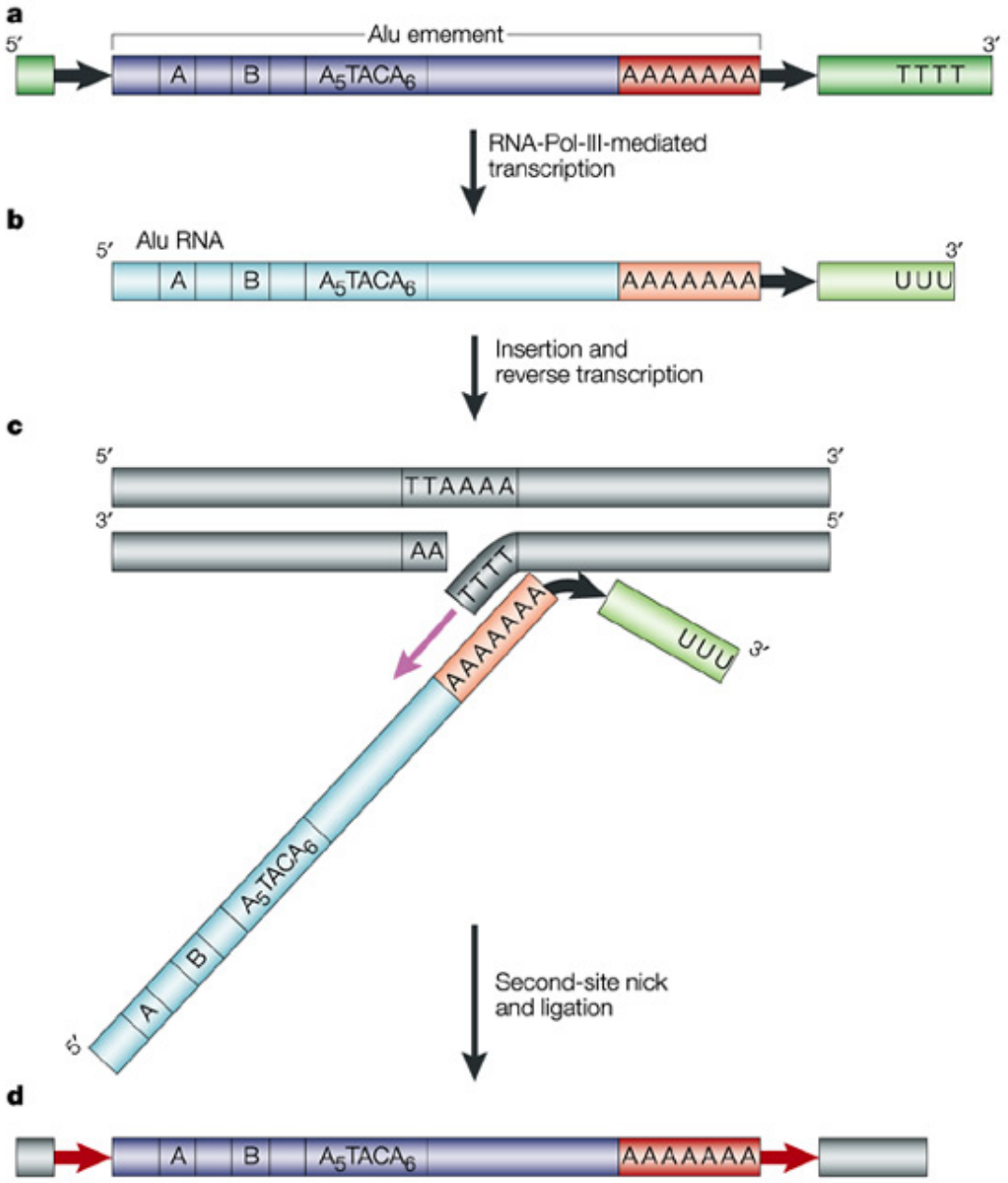


Triplet repeat diseases



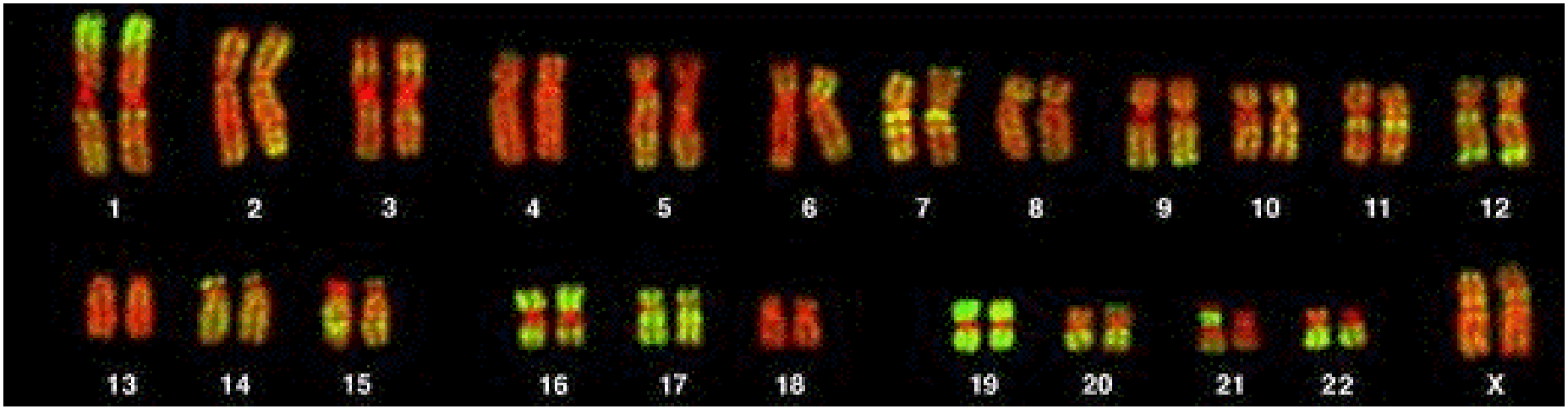
Alu elements

The structure of each Alu element is bi-partite, with the 3' half containing an additional 31-bp insertion (not shown) relative to the 5' half. The total length of each Alu sequence is 300 bp, depending on the length of the 3' oligo(dA)-rich tail. The elements also contain a central A-rich region and are flanked by short intact direct repeats that are derived from the site of insertion (black arrows). The 5' half of each sequence contains an RNA-polymerase-III promoter (A and B boxes). The 3' terminus of the Alu element almost always consists of a run of As that is only occasionally interspersed with other bases (**a**).



Nature Reviews | Genetics

The abundant Alu transposable element, a member of the middle repetitive DNA sequences, is present in all human chromosomes (the Alu element is stained green, while the remainder of the DNA in the chromosomes is stained red).



- > 1 million in genome – unique to humans
- Involved in RNA editing – functional ?
- How well are they tagged ??????

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)

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

Contrit

Protei
genon
diseas
seque
poten
humai
ing co
DNA
appro
sensiti
gous
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 prem
Tubby superfamily) and has previously been characterized as an intrafl
syndrome is a ciliary disorder.

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

NATURE GENETICS VOLUME 42 | NUMBER 6 | JUNE 2010

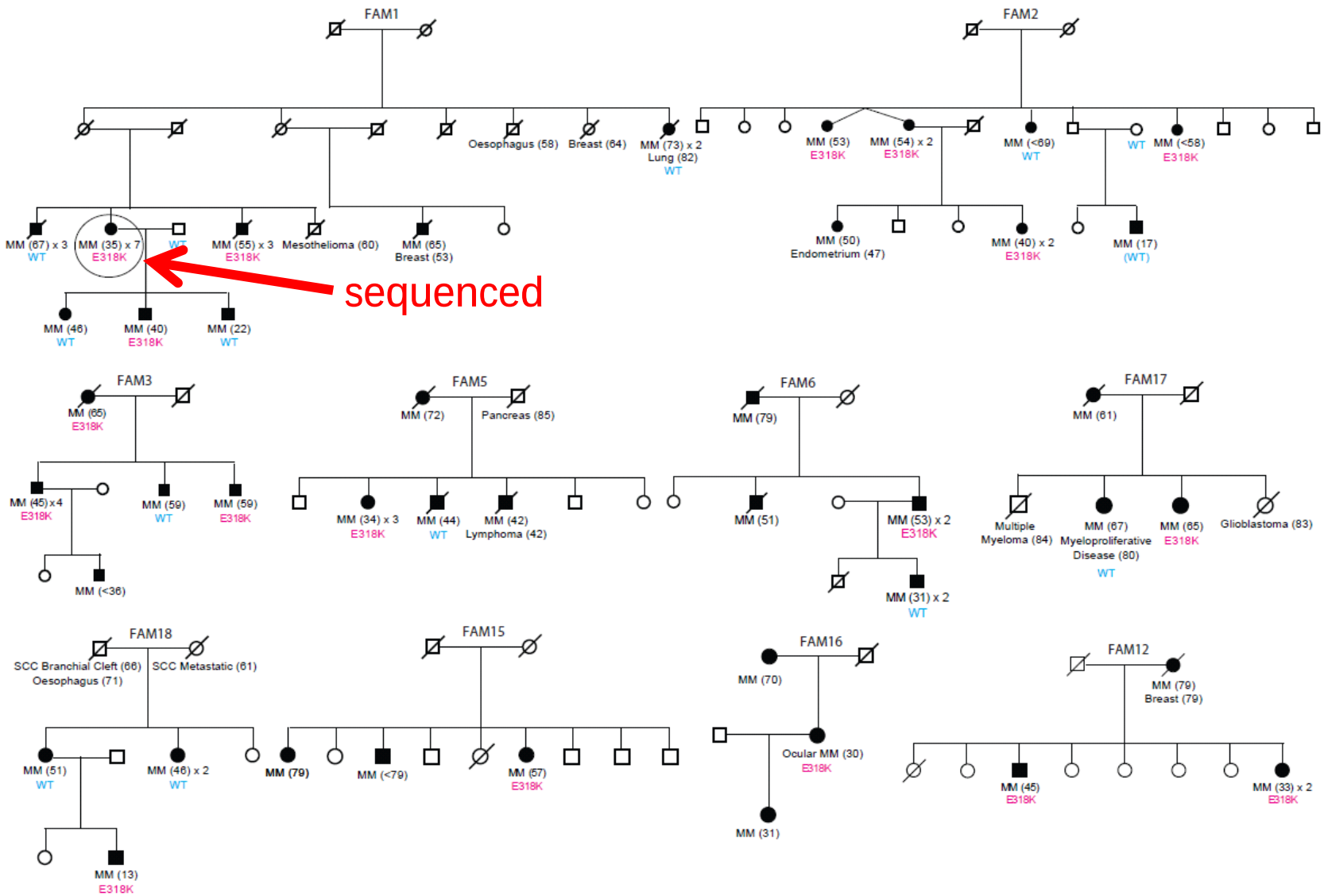
De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome

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.

Can sequencing contribute to the genetics of complex traits ?

MITF (E318K) cosegregates with melanoma in at least 28% of families carrying the variant



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

Reengineering Translational Science: The Time Is Right

Francis S. Collins

Table 1. The GWAS potential. GWAS* can reveal new therapeutic targets for complex diseases (*8*, *56*, *57*).

Disease	Total GWAS hits†	GWAS hits associated with marketed drugs‡	GWAS hits associated with drug effects§
Type 2 diabetes	44	6	8
Hyperlipidemia	39	2	10
Multiple sclerosis	36	5	2
Psoriasis	24	4	1

*Genome-wide association studies (GWAS) assume no knowledge of disease pathogenesis and provide a comprehensive approach to the discovery of common genetic risk factors. Many known drug targets and associated pathways appear on the list of GWAS hits for common diseases, suggesting that other GWAS hits likely represent "druggable" targets worthy of further investigation. †Genetic variants strongly linked to disease susceptibility. ‡Genetic variants that are primary targets of drugs currently marketed for the listed indication. §Genetic variants associated with cellular, pharmacokinetic, pharmacodynamic, or clinical variations in response to one or more drugs currently marketed for the listed indication.

Sci Transl Med. 2011 Jun 15;3(87):87re3.

Whole-genome sequencing for optimized patient management.

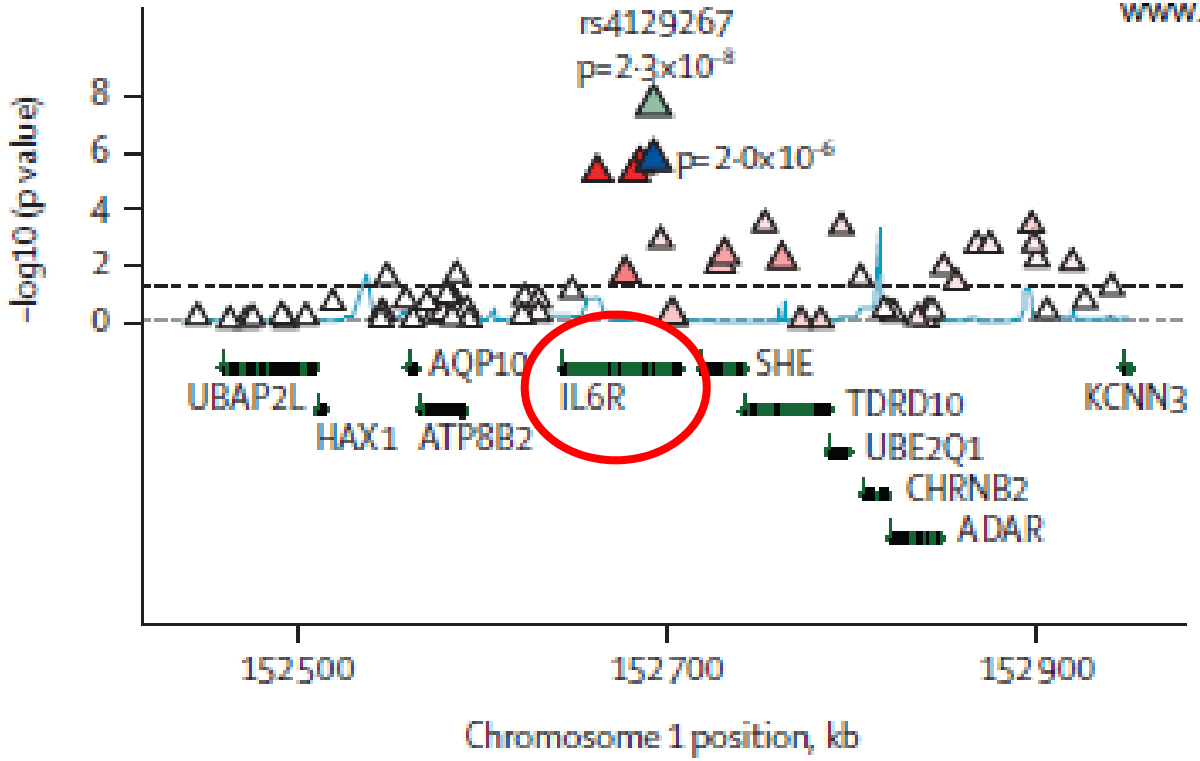
Bainbridge MN, ... Gibbs RA.

Whole-genome sequencing of patient DNA can facilitate diagnosis of a disease, but its potential for guiding treatment has been under-realized. We interrogated the complete genome sequences of a 14-year-old fraternal twin pair diagnosed with dopa (3,4-dihydroxyphenylalanine)-responsive dystonia. Whole-genome sequencing identified compound heterozygous mutations in the SPR gene encoding sepiapterin reductase. Disruption of SPR causes a decrease in tetrahydrobiopterin, a cofactor required for the hydroxylase enzymes that synthesize the neurotransmitters dopamine and serotonin. Supplementation of l-dopa therapy with 5-hydroxytryptophan, a serotonin precursor, resulted in clinical improvements in both twins.

Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma

Manuel A R Ferreira, Melanie C Matheson, David L Duffy, Guy B Marks, Jennie Hui, Peter Le Souëf, Patrick Danoy, Svetlana Baltic, Dale R Nyholt,

www.thelancet.com Vol 378 September 10, 2011



Interpretation The *IL6R* association further supports the hypothesis that cytokine signalling dysregulation affects asthma risk, and raises the possibility that an IL6R antagonist (tocilizumab) may be effective to treat the disease, perhaps in a genotype-dependent manner. Results for the 11q13.5 locus suggest that it directly increases the risk of

Drug Design, Development and Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

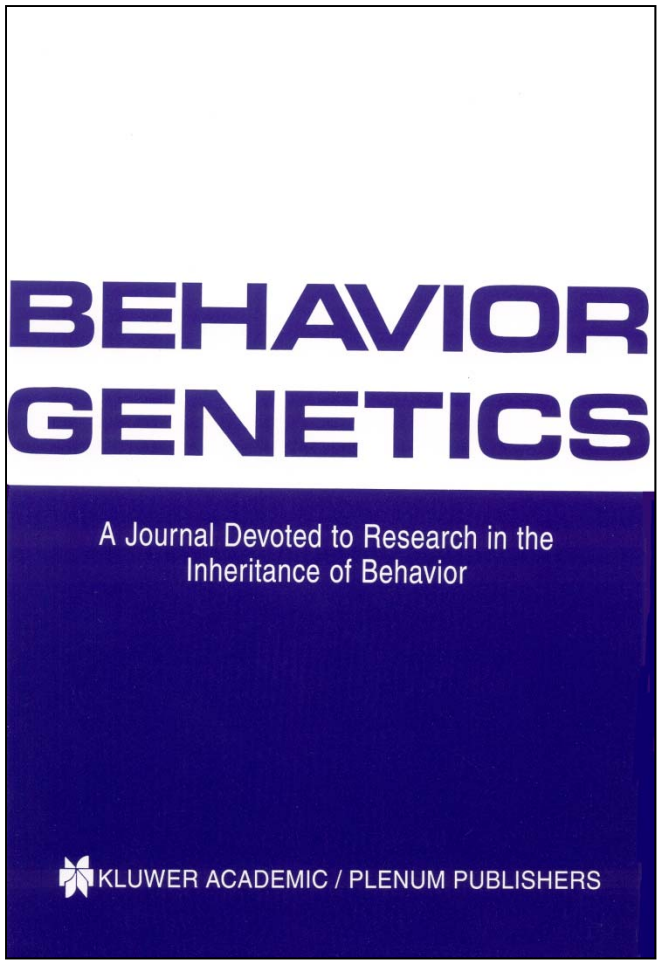
Interleukin-6 inhibition for treatment of rheumatoid arthritis: A review of tocilizumab therapy

Interrupting IL-6–receptor signaling improves atopic dermatitis but associates with bacterial superinfection

The future

- GWAS works!
  - collect larger samples
  - collaborate
- Sequencing may work...
  - collect multiplex families
- Function
  - Correlate GWAS/NGS hits with other –omics
  - take GWAS/NGS hits into animals – not *vice versa*
- Translation
  - new genes/pathways provide targets for intervention
    e.g. *IL6R* and tocilumizab for asthma

We also run two journals (1)



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

We also run two journals (2)

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