QTL studies: past, present and future

Nick Martin

Dorret Boomsma

Ben Neale

David Evans

and other faculty

Boulder workshop: March 5, 2010

R.A. Fisher, 1918

The explanation of quantitative inheritance in Mendelian terms

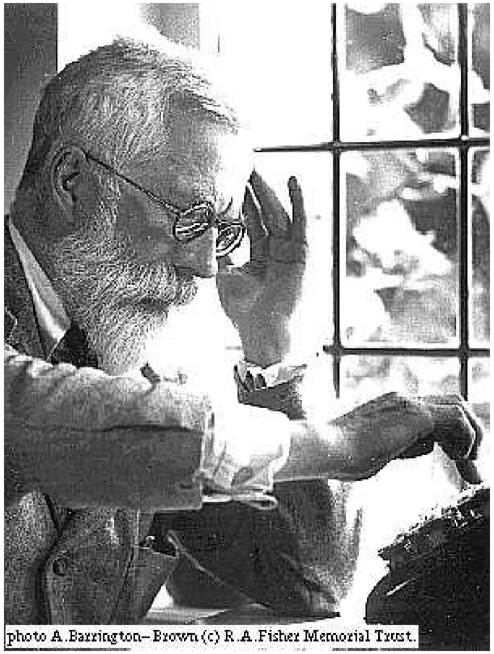


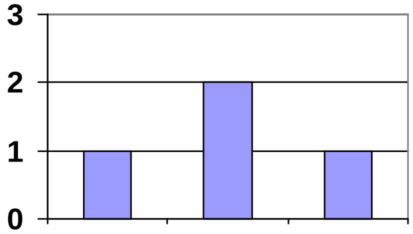
photo A.Barrington–Brown (c) R.A.Fisher Memorial Trust.

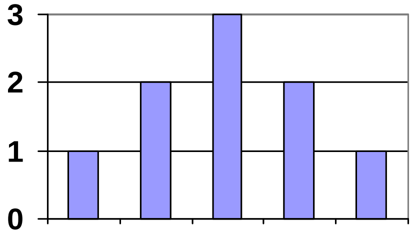
1 Gene
→ 3 Genotypes
→ 3 Phenotypes

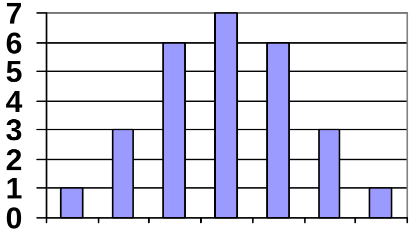
2 Genes
→ 9 Genotypes
→ 5 Phenotypes

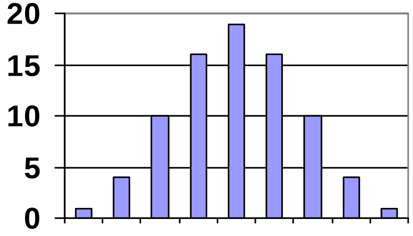
3 Genes
→ 27 Genotypes
→ 7 Phenotypes

4 Genes
→ 81 Genotypes
→ 9 Phenotypes







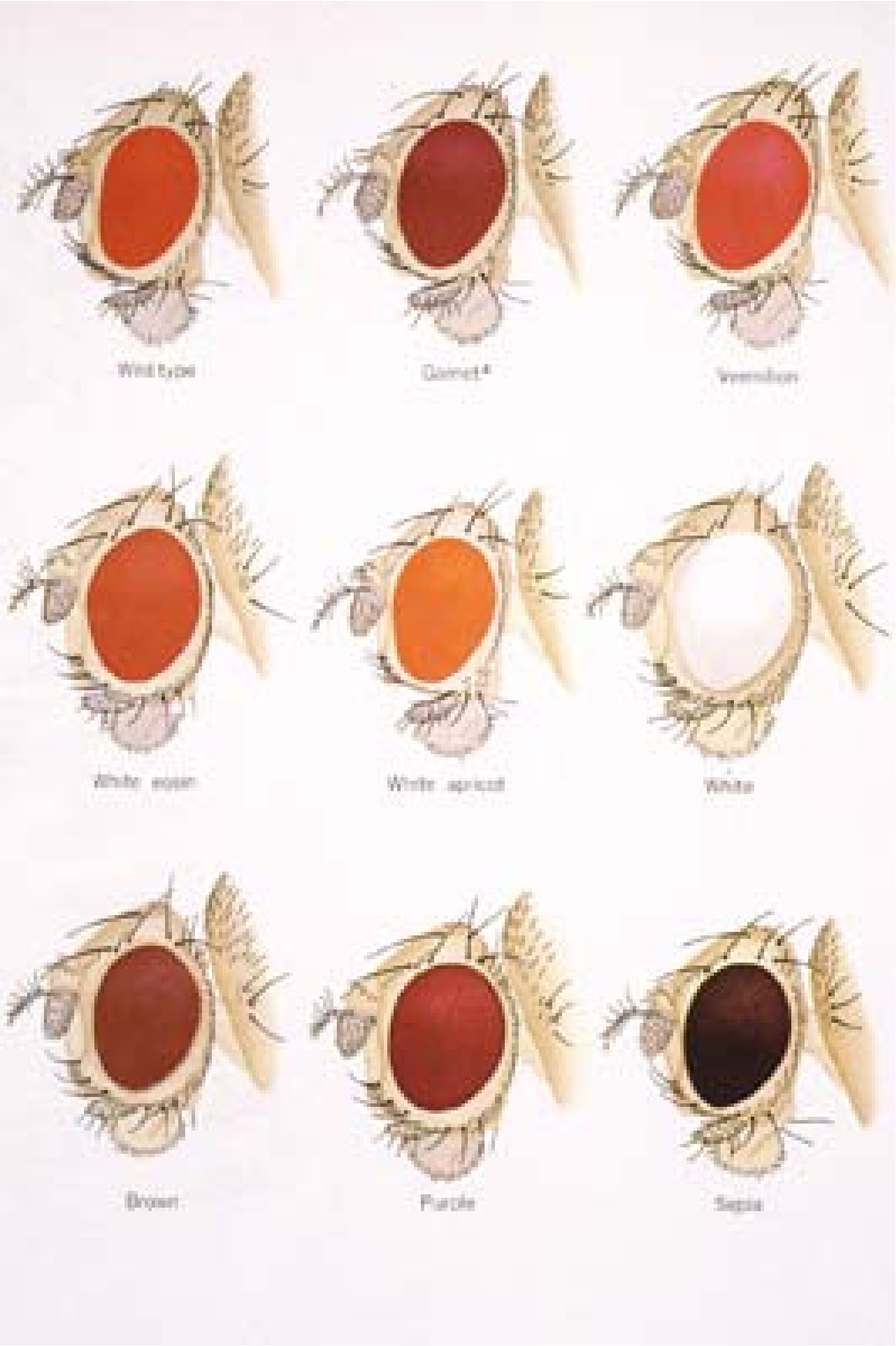


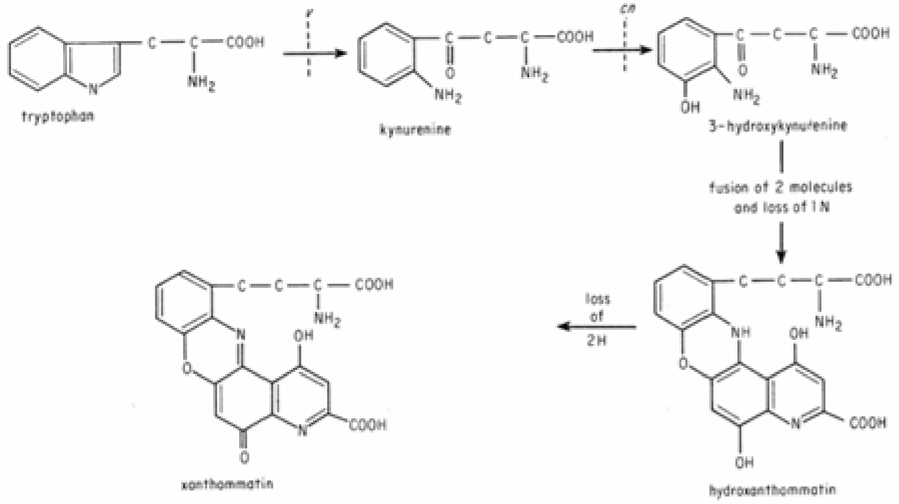
Finding QTLs

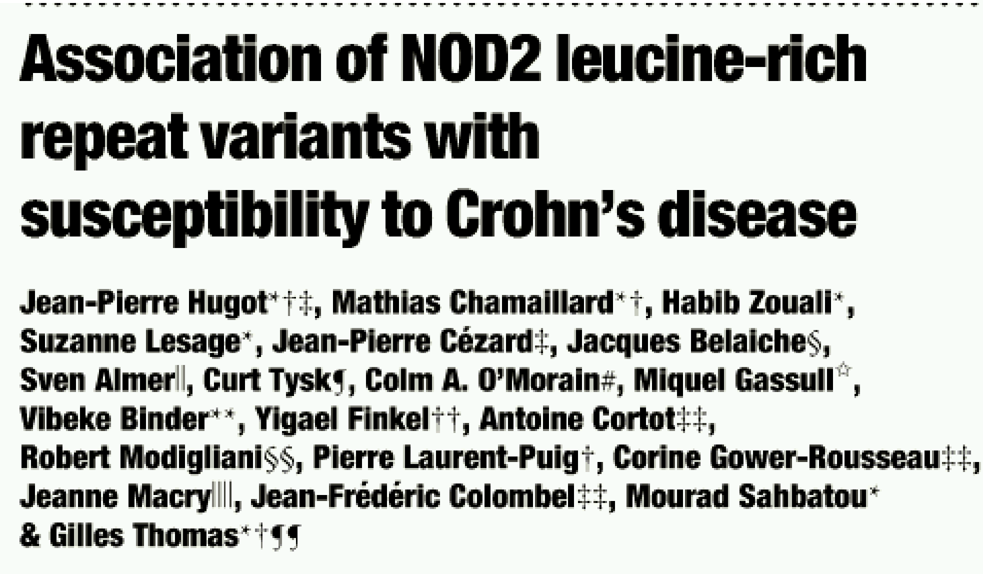
- Linkage
- Association

Using genetics to dissect metabolic pathways: Drosophila eye color

Beadle & Ephrussi, 1936

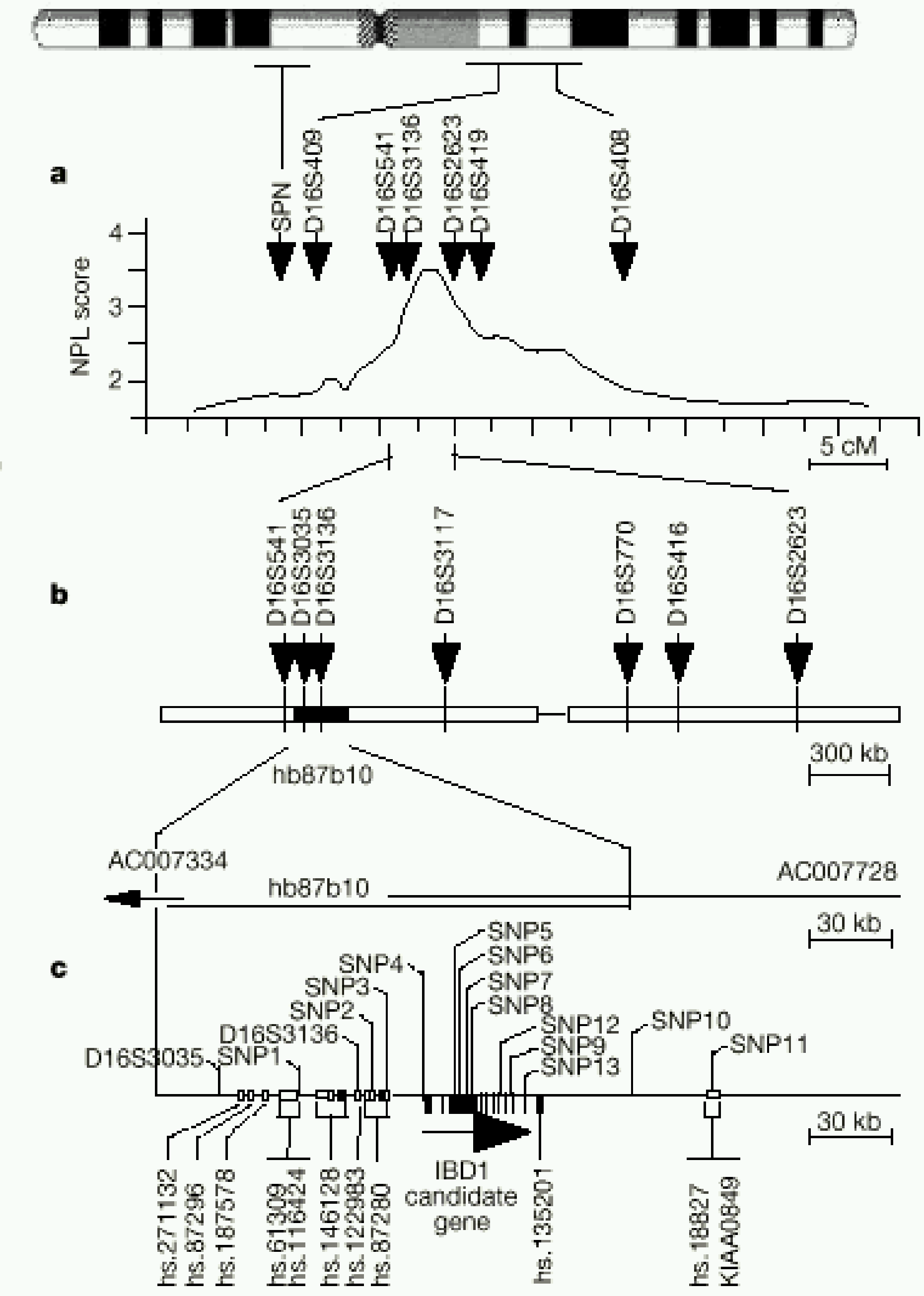


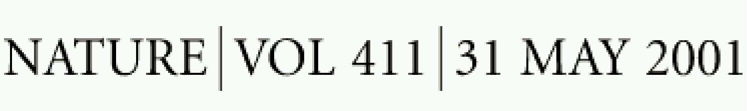




Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

Jean-Pierre Hugot[*†‡], Mathias Chamaillard[*†], Habib Zouali[*], Suzanne Lesage[*], Jean-Pierre Cézard[‡], Jacques Belaiche[§], Sven Almer[||], Curt Tysk[¶], Colm A. O'Morain[#], Miquel Gassull[☆], Vibeke Binder[**], Yigael Finkel[††], Antoine Cortot[‡‡], Robert Modigliani[§§], Pierre Laurent-Puig[†], Corine Gower-Rousseau[‡‡], Jeanne Macry[||||], Jean-Frédéric Colombel[‡‡], Mourad Sahbatou[*] & Gilles Thomas[*†¶¶]

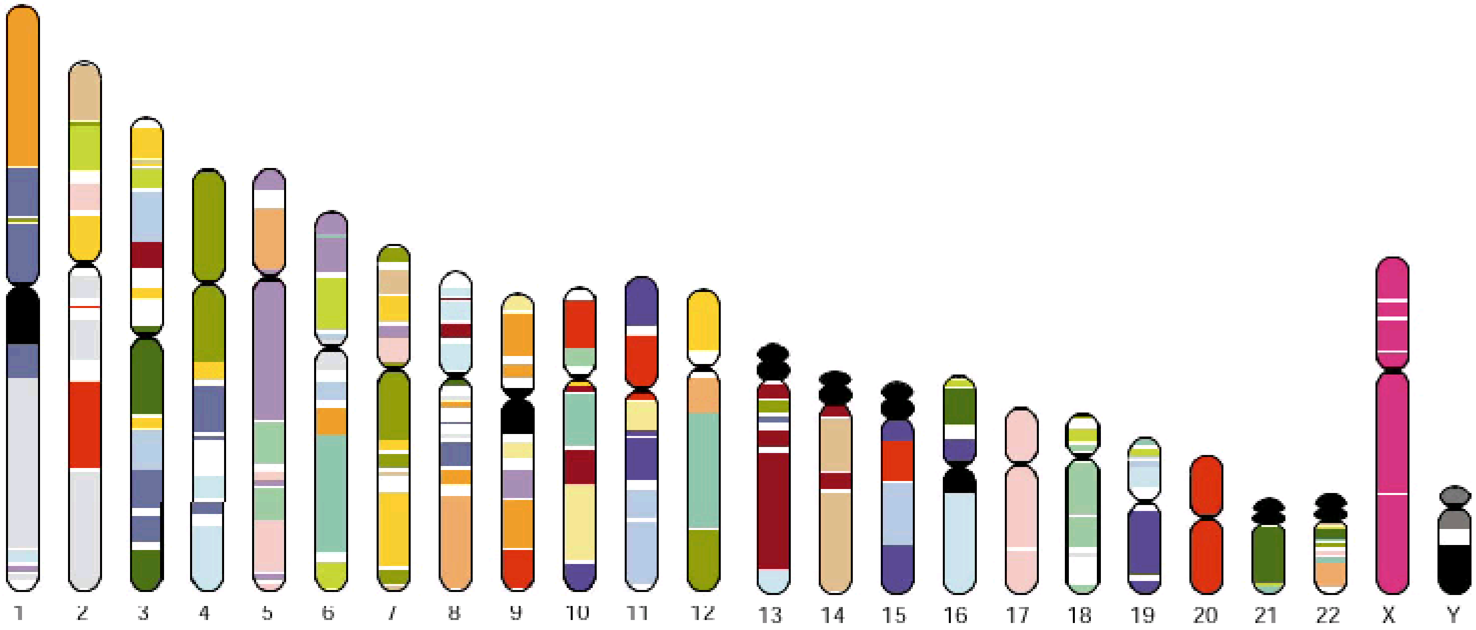




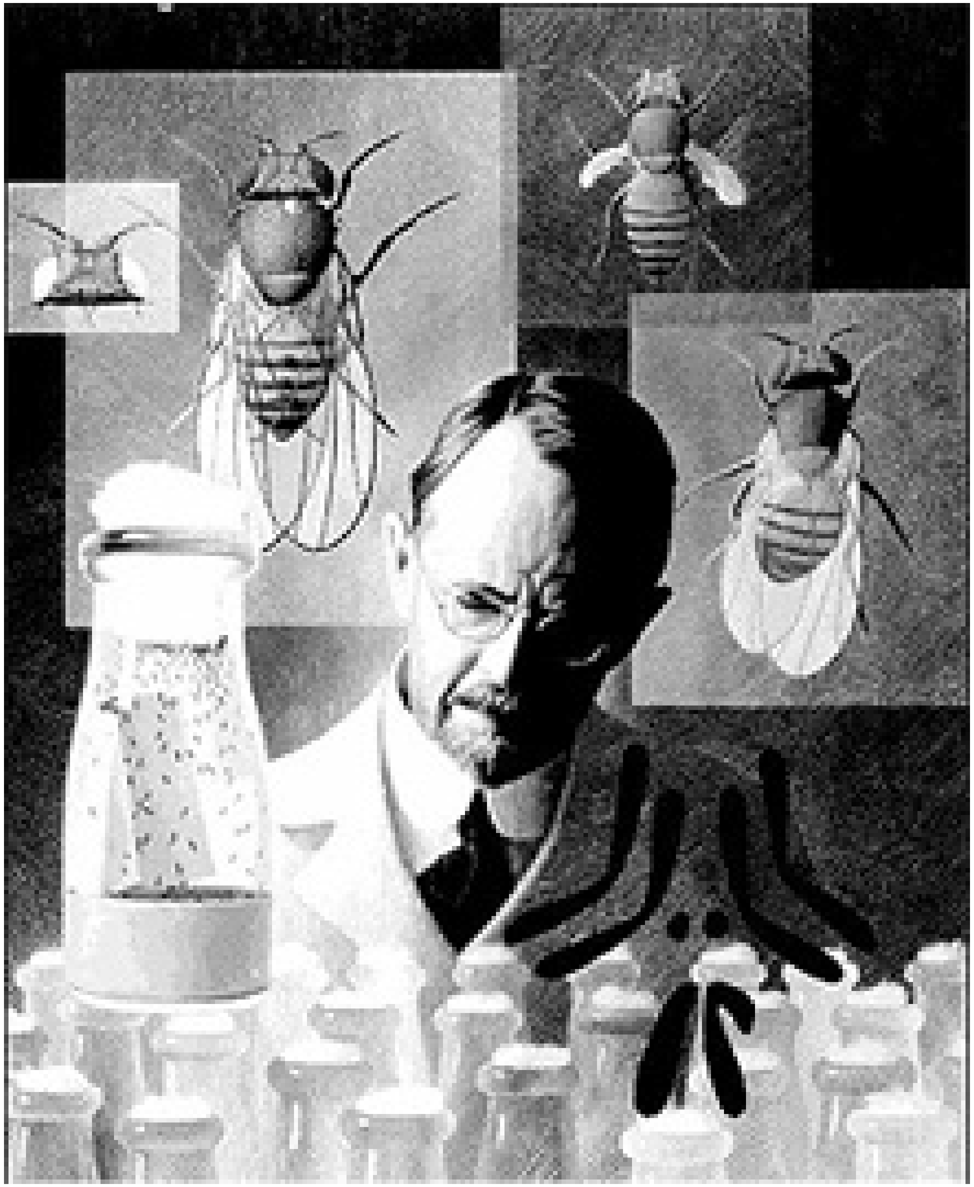
NATURE | VOL 411 | 31 MAY 2001

First (unequivocal) positional cloning of a complex disease QTL !

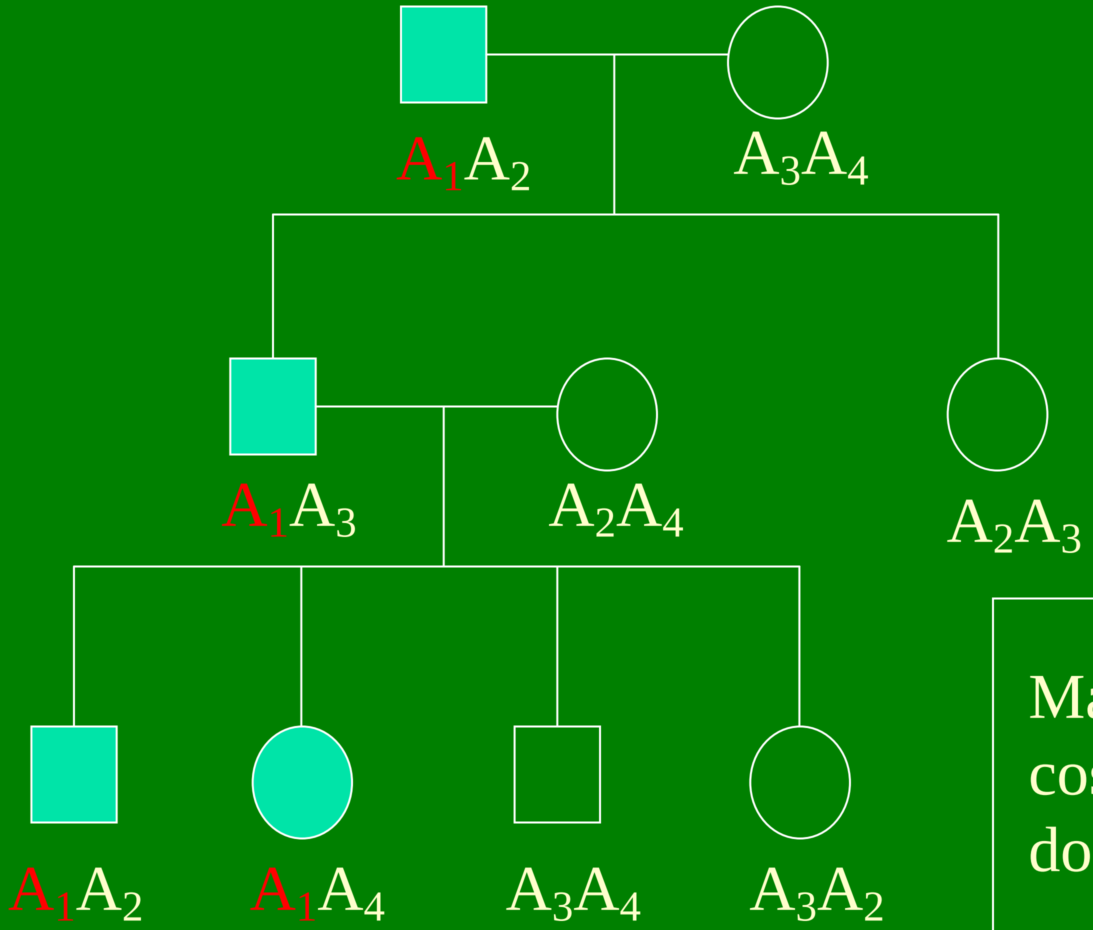
Linkage analysis



Thomas Hunt Morgan – discoverer of linkage

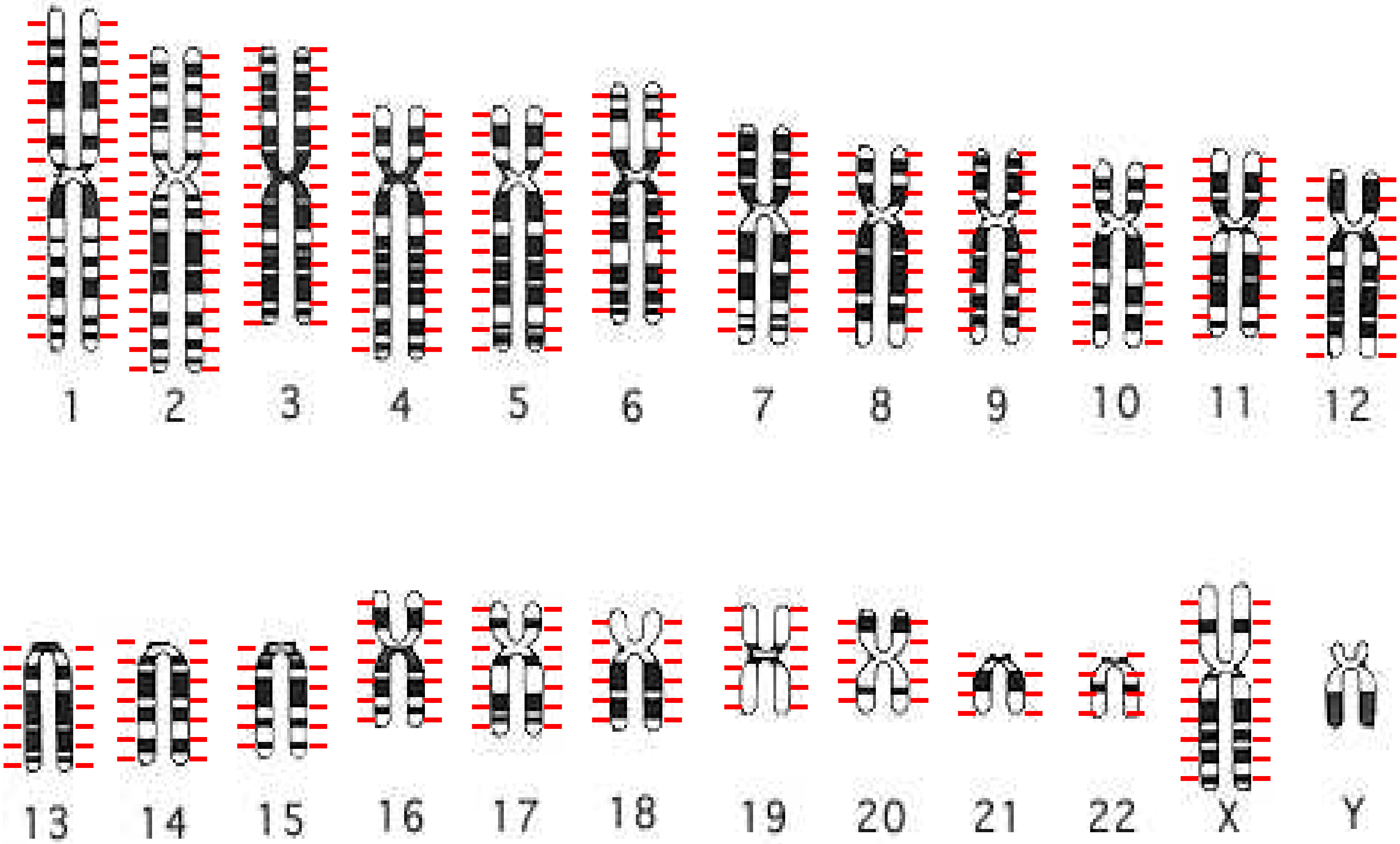


Linkage = Co-segregation

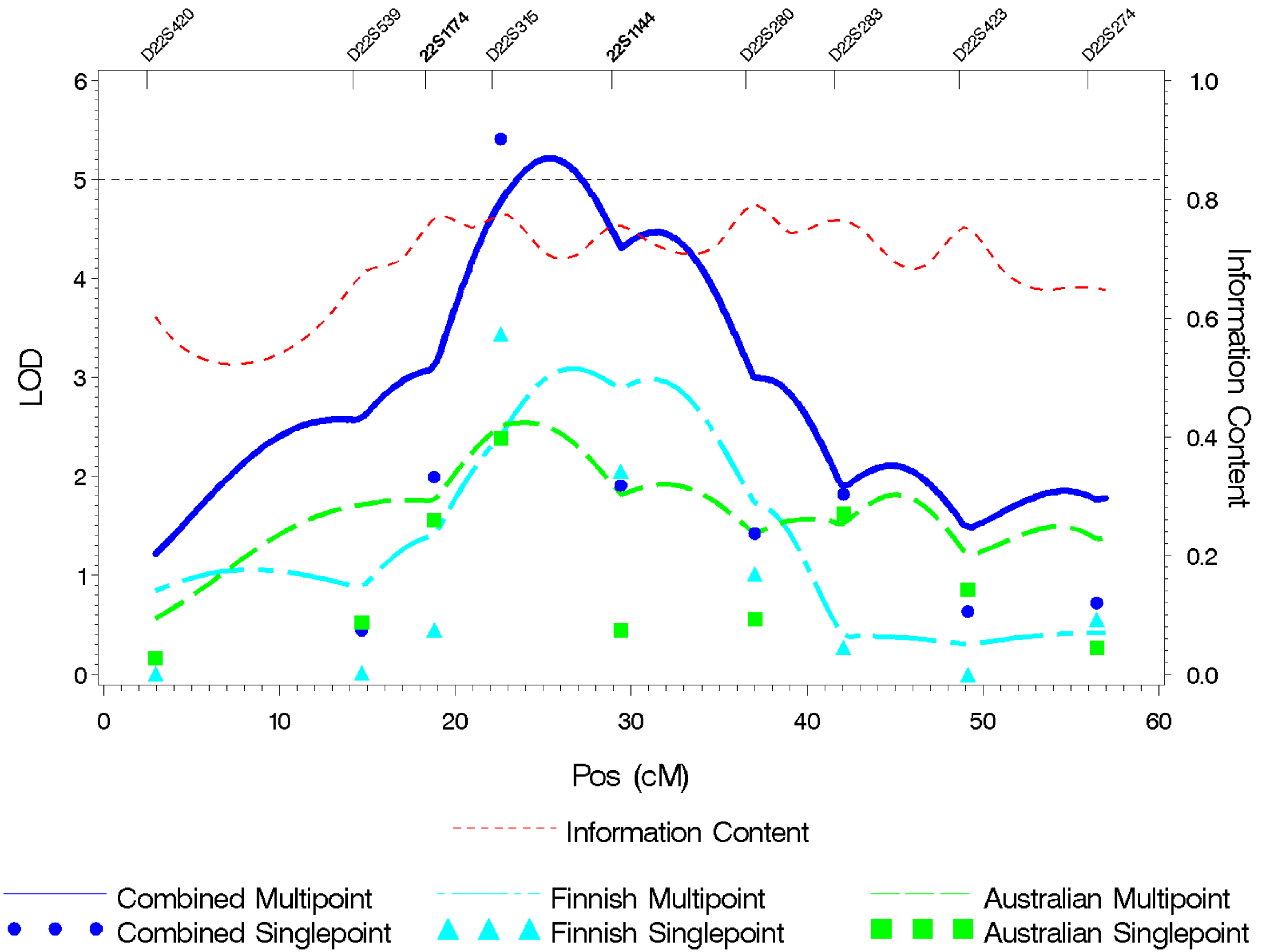


Marker allele A_1 cosegregates with dominant disease 🟩

Linkage Markers…



Linkage for MaxCigs24 in Australia and Finland



AJHG, in press

Linkage

- Doesn’t depend on “guessing gene”
- Works over broad regions
- Only detects large effects (>10%)
- Requires large samples (10,000’s?)
- Can’t guarantee close to gene
- For complex traits results have been disappointing…………

Association

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

Association

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

No. 4356 April 25, 1953 NATURE 737

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

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

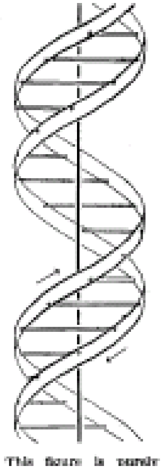
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

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

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

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



This figure is purely

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

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

The novel feature of the structure
in which the two chains are held
purine and pyrimidine bases. The pl
are perpendicular to the fibre axis.
together in pairs, a single base from
hydrogen-bonded to a single base
chain, so that the two lie side by sid
z-co-ordinates. One of the pair must
the other a pyrimidine for bonding
hydrogen bonds are made as follows ;
1 to pyrimidine position 1 ; purin
pyrimidine position 6.

If it is assumed that the bases o
structure in the most plausible ta
(that is, with the keto rather tha
figurations) it is found that only
bases can bond together. These pai
(purine) with thymine (pyrimidine
(purine) with cytosine (pyrimidine).

In other words, if an adenine form
a pair, on either chain, then on th
the other member must be thymin
guanine and cytosine. The sequenc
single chain does not appear to be
way. However, if only specific pairs
formed, it follows that if the sequ
one chain is given, then the sequen
chain is automatically determined.

It has been found experimentally
of the amounts of adenine to thymin
of guanine to cytosine, are always ve
for deoxyribose nucleic acid.

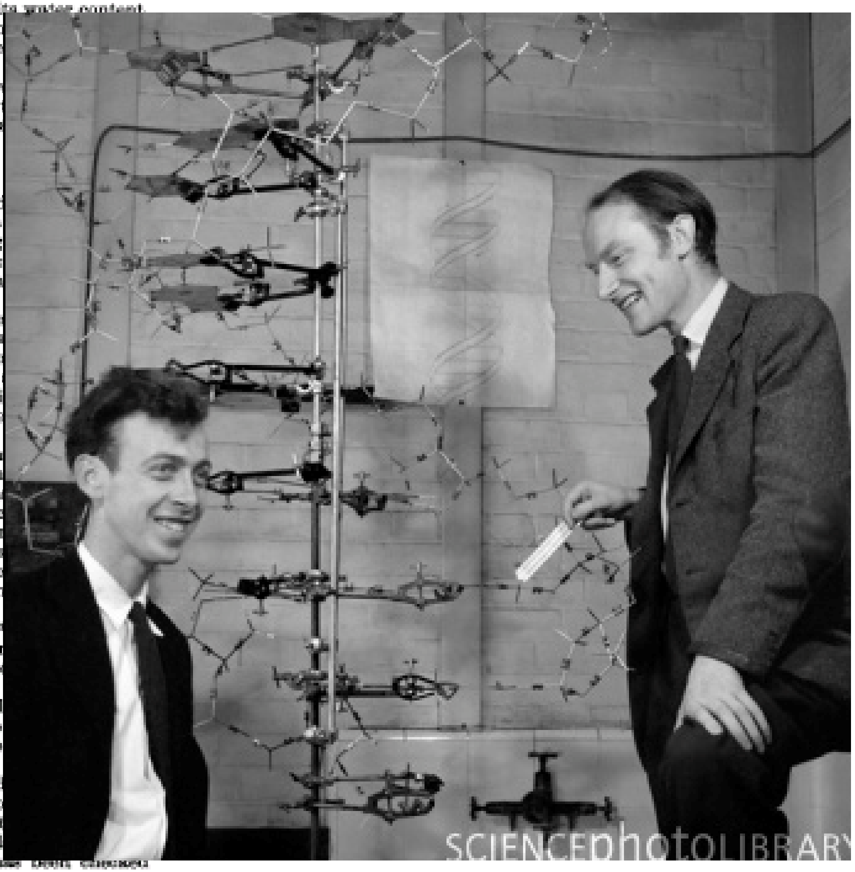
It is probably impossible to buil
with a ribose sugar in place of the
the extra oxygen atom would make
der Waals contact.

The previously published X-ray d
ribose nucleic acid are insufficient fo
of our structure. So far as we can
compatible with the experimental d
be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

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

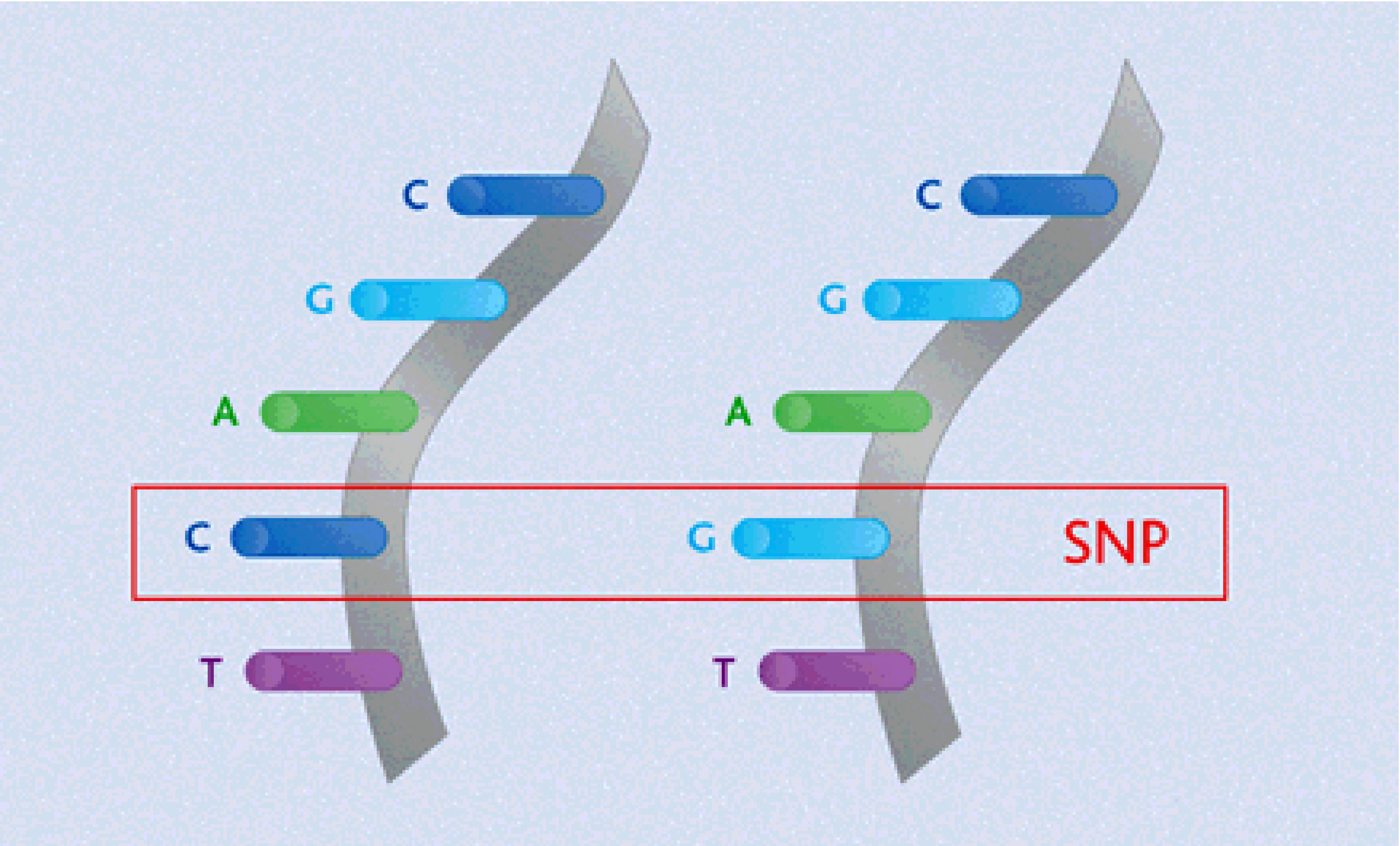
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for

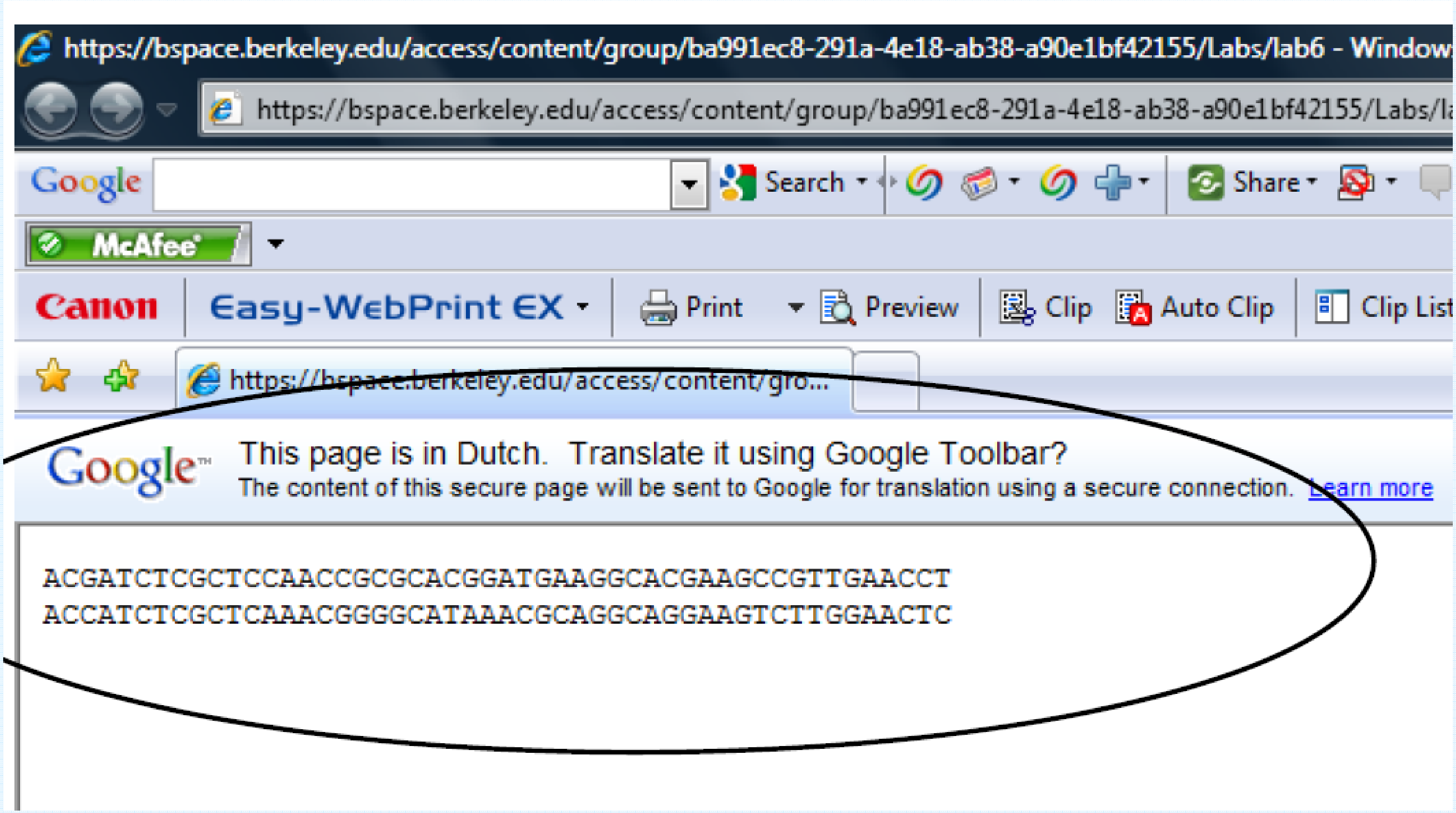


Watson & Crick (1953)

Variation: Single Nucleotide Polymorphisms



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



https://bspace.berkeley.edu/access/content/group/ba991ec8-291a-4e18-ab38-a90e1bf42155/Labs/lab6 - Window

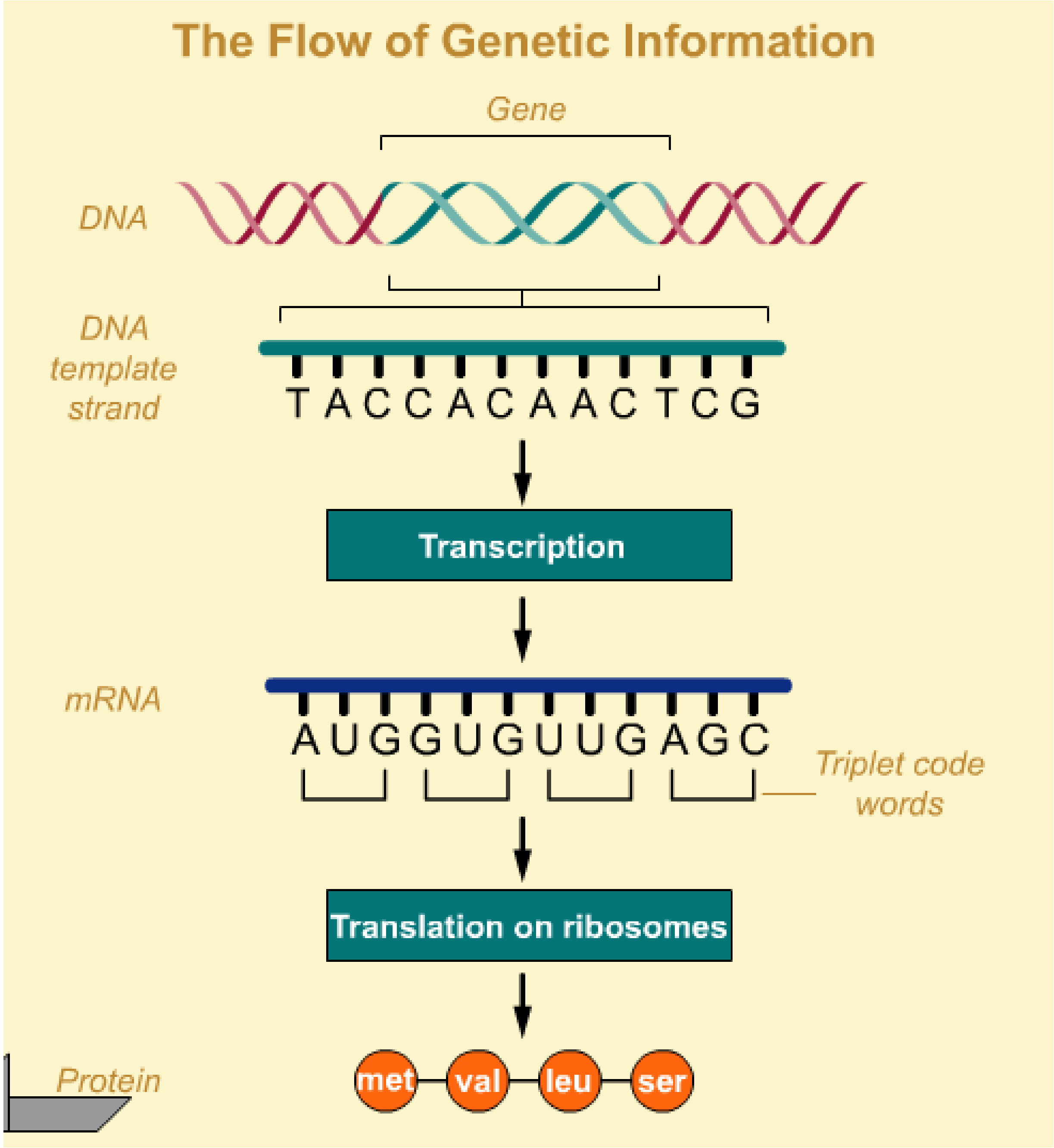
https://bspace.berkeley.edu/access/content/group/ba991ec8-291a-4e18-ab38-a90e1bf42155/Labs/la

https://bspace.berkeley.edu/access/content/gro...

This page is in Dutch. Translate it using Google Toolbar?

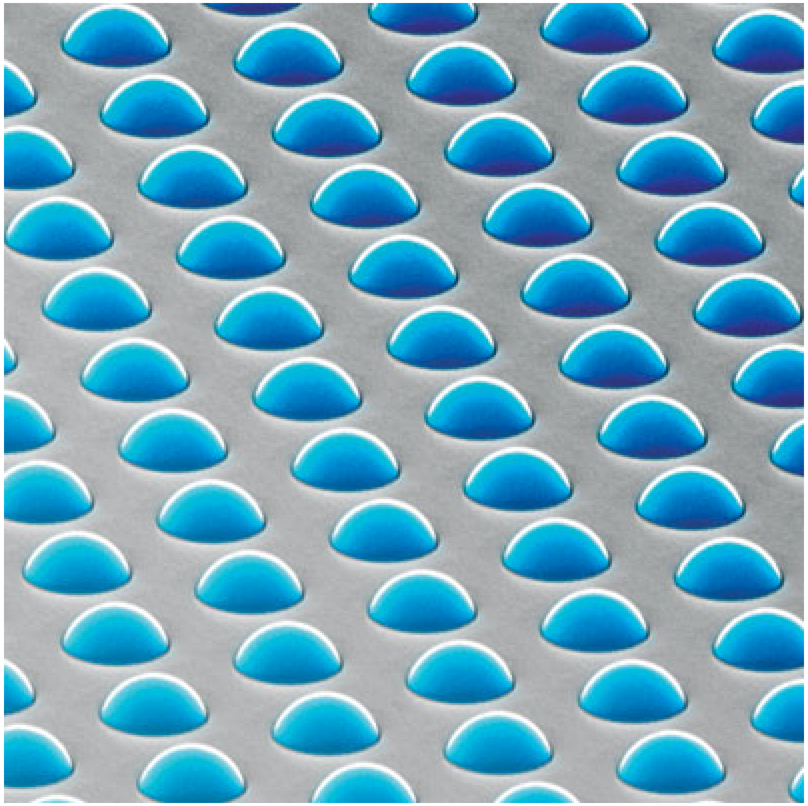
The content of this secure page will be sent to Google for translation using a secure connection. Learn more

```
ACGATCTCGCTCCAACCGCGCACGGATGAAGGCACGAAGCCGTTGAACCT
ACCATCTCGCTCAAACGGGGCATAAACGCAGGCAGGAAGTCTTGGAACTC
```

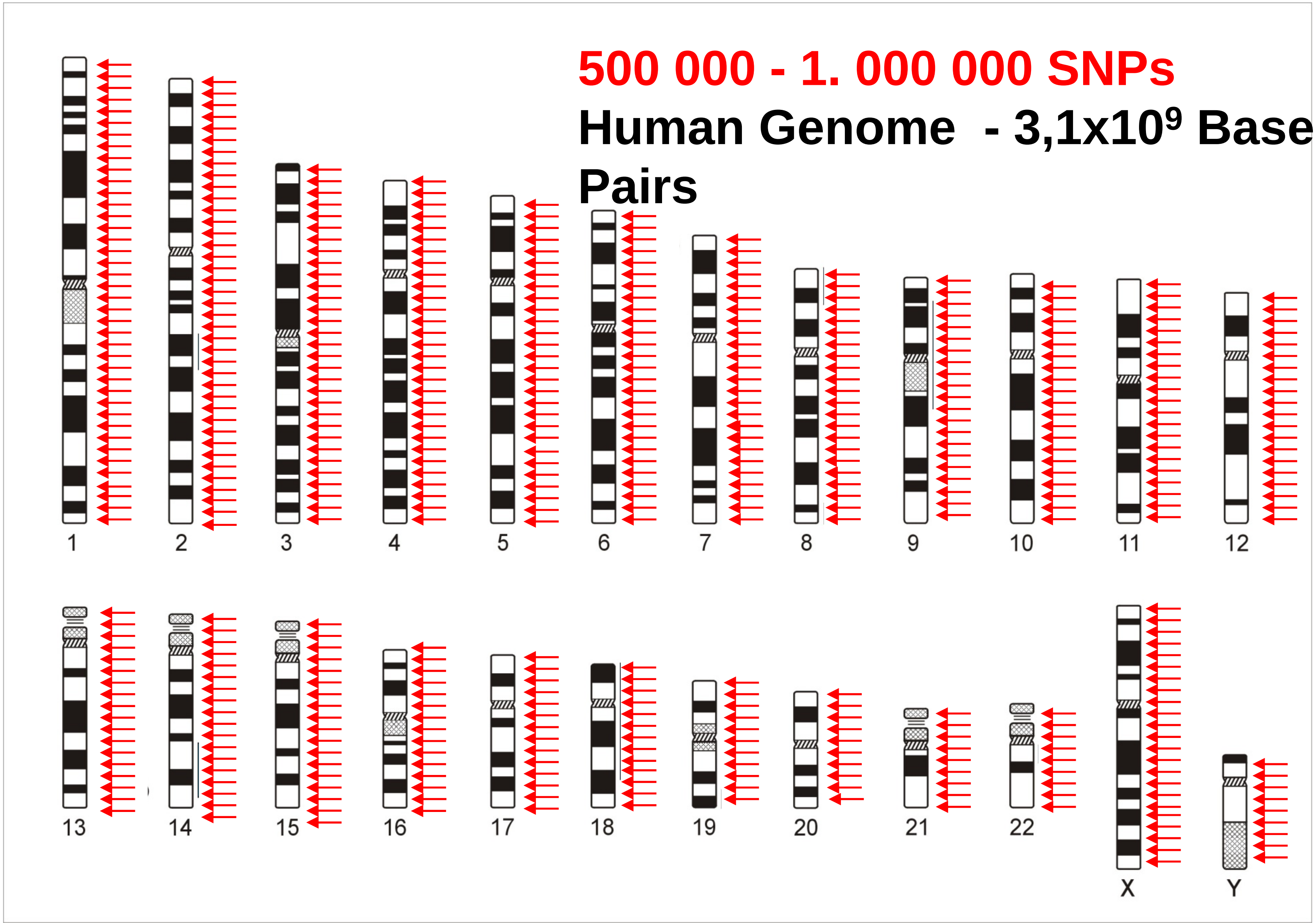


Differences (between subjects) in DNA sequence are responsible for (structural) differences in proteins.

High density SNP arrays – up to 1 million SNPs



Genome-Wide Association Studies

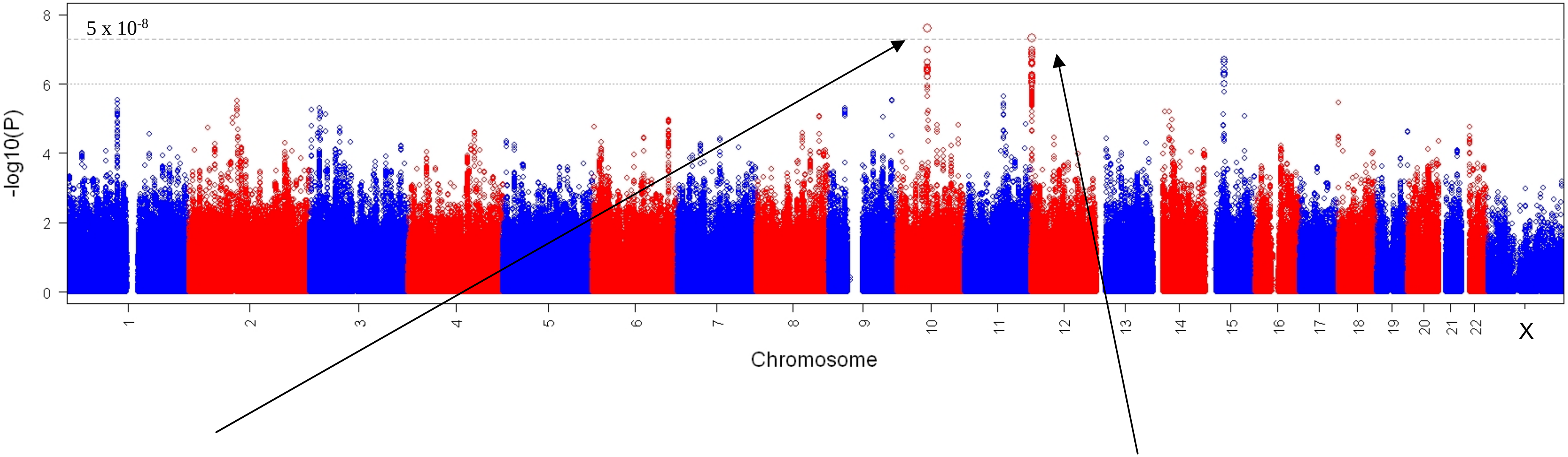


500 000 - 1. 000 000 SNPs

Human Genome - 3,1x10^9 Base Pairs

Bipolar GWAS of 10,648 samples

>1.7 million genotyped and (high confidence) imputed SNPs



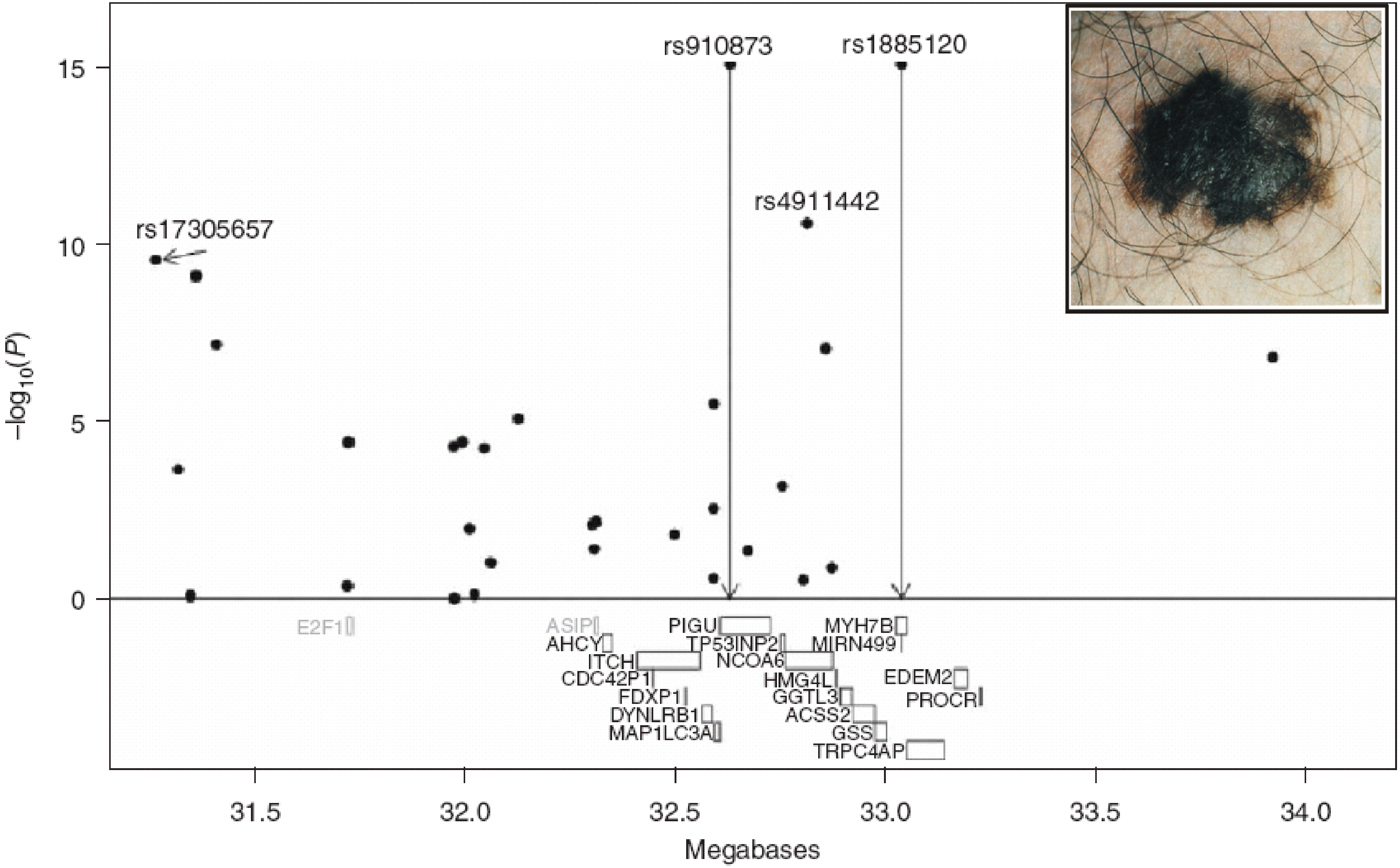
Ankryin-G (ANK3)

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

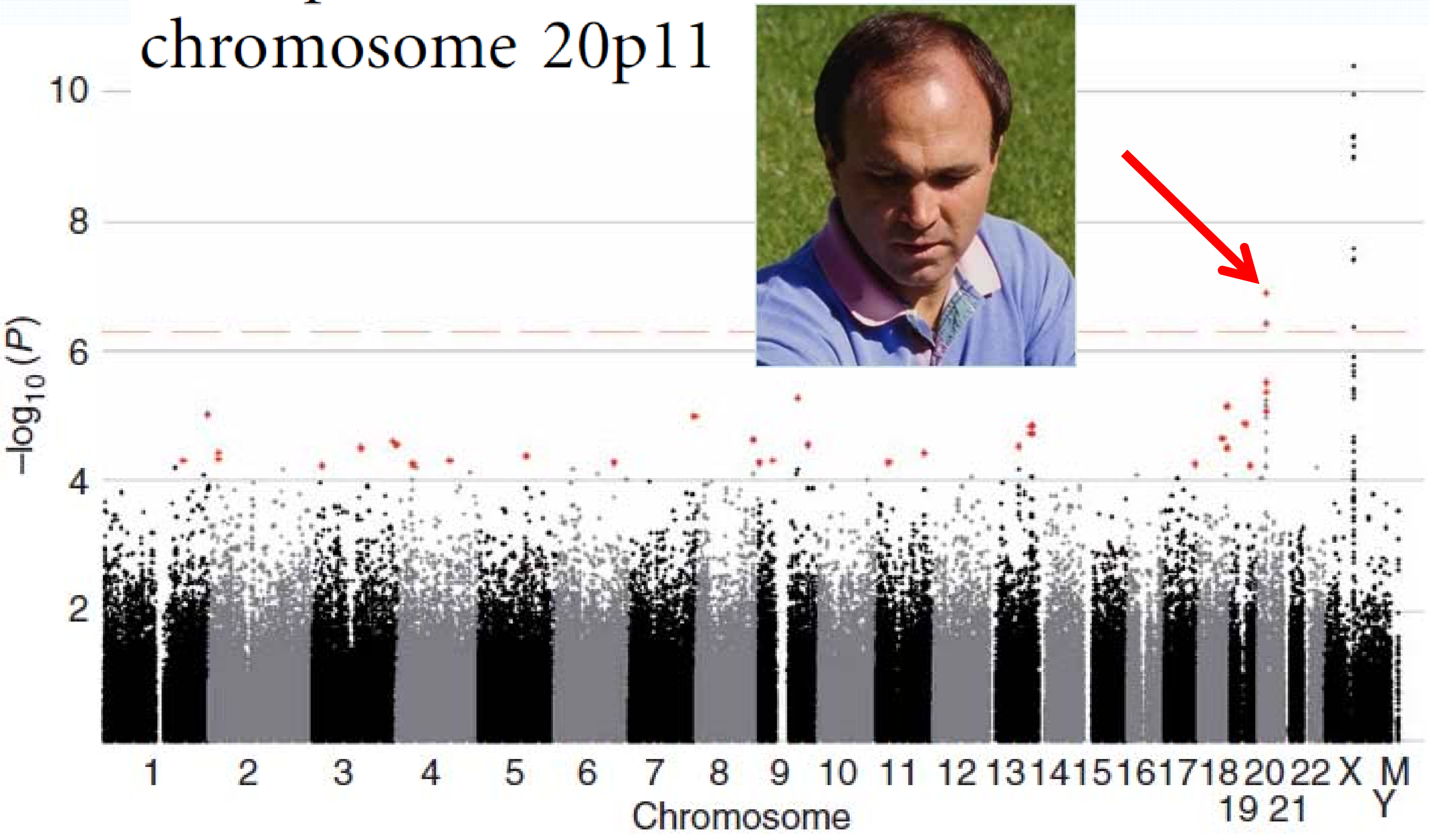
Sample	**Case**	**Controls**	***P*-value**
STEP	35.7%	32.4%	0.0015
WTCCC	35.7%	31.5%	0.0003
EXT	35.3%	33.7%	0.0108
Total	**35.6%**	**32.4%**	**7×10^{-8}**

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

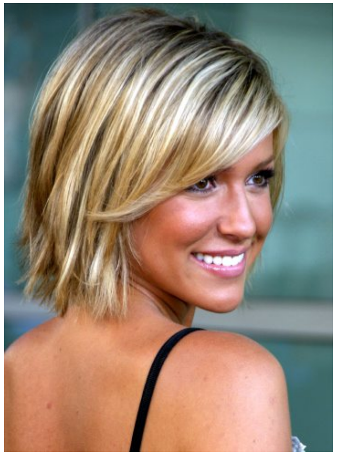


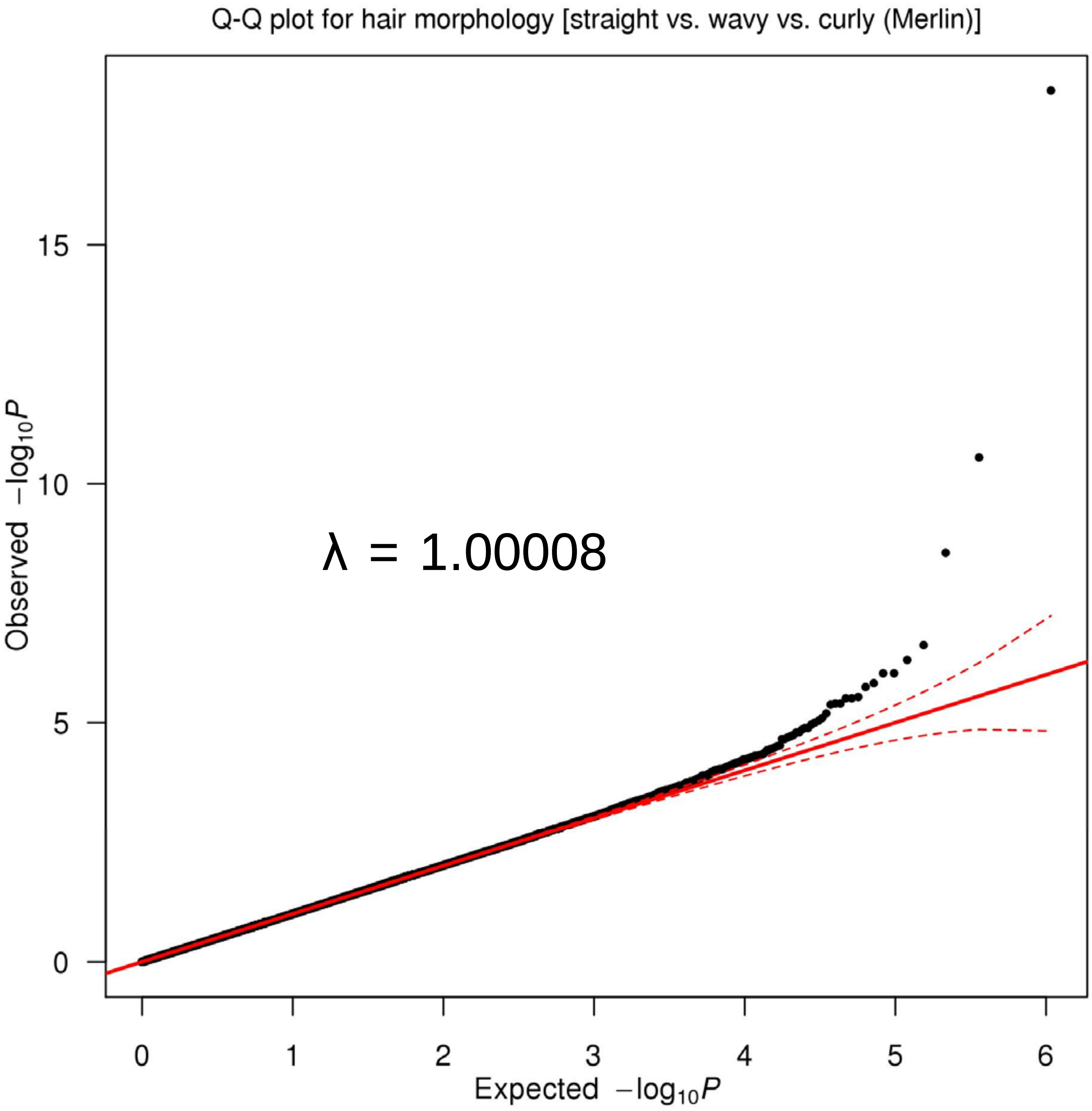
GWAS for Melanoma Association analysis of SNPs across a region of chromosome 20q11.22 for the combined sample. The x-axis is chromosomal position, the left y-axis $-\log_{10}(p)$ for genotyped SNPs. *Nature Genetics* 2008 Jul;40(7):838-40.

Susceptibility variants for male-pattern baldness on chromosome 20p11

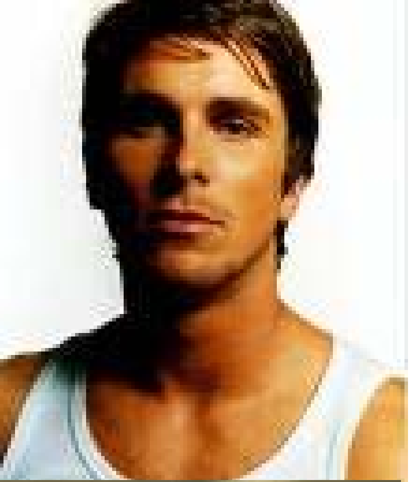


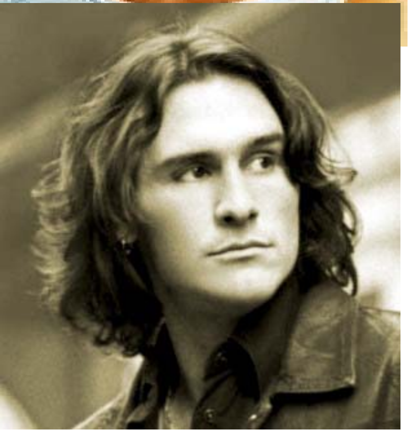
NATURE GENETICS VOLUME 40 | NUMBER 11 | NOVEMBER 2008

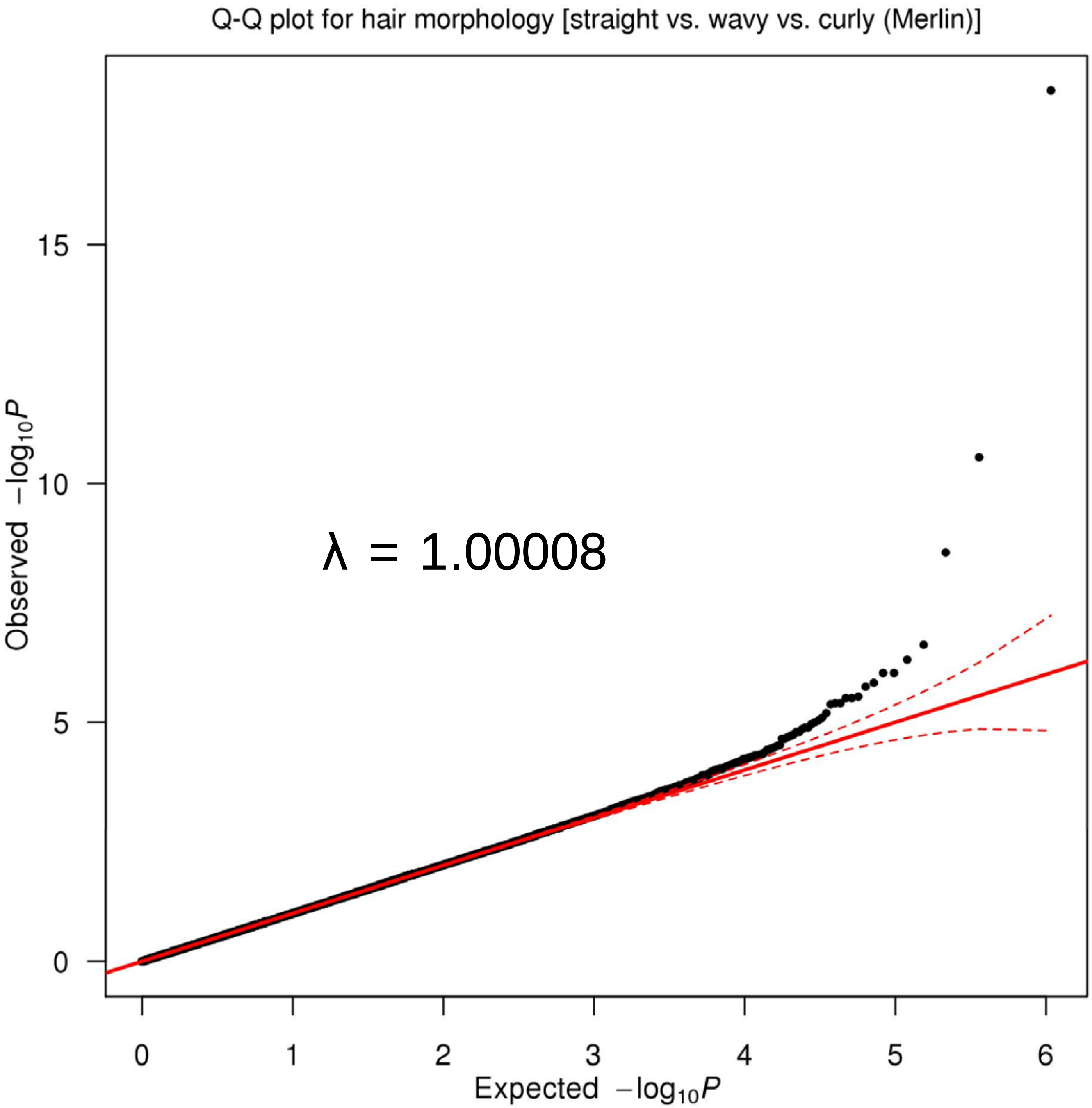



Q-Q plot for hair morphology [straight vs. wavy vs. curly (Merlin)]
Observed $-\log_{10}P$
15
10
5
0
Expected $-\log_{10}P$
0
1
2
3
4
5
6

λ = 1.00008







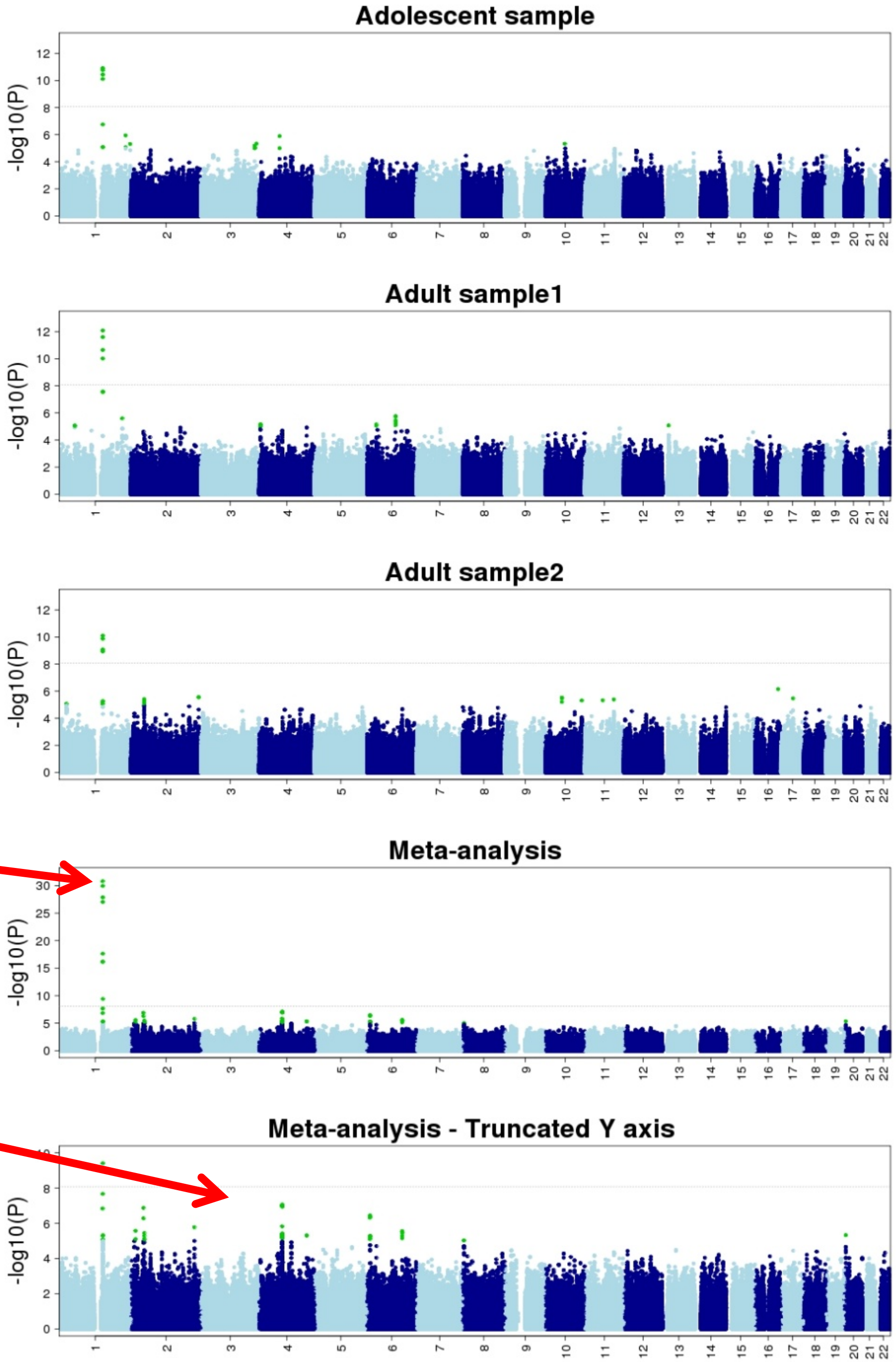
Q-Q plot for hair morphology [straight vs. wavy vs. curly (Merlin)]

λ = 1.00008

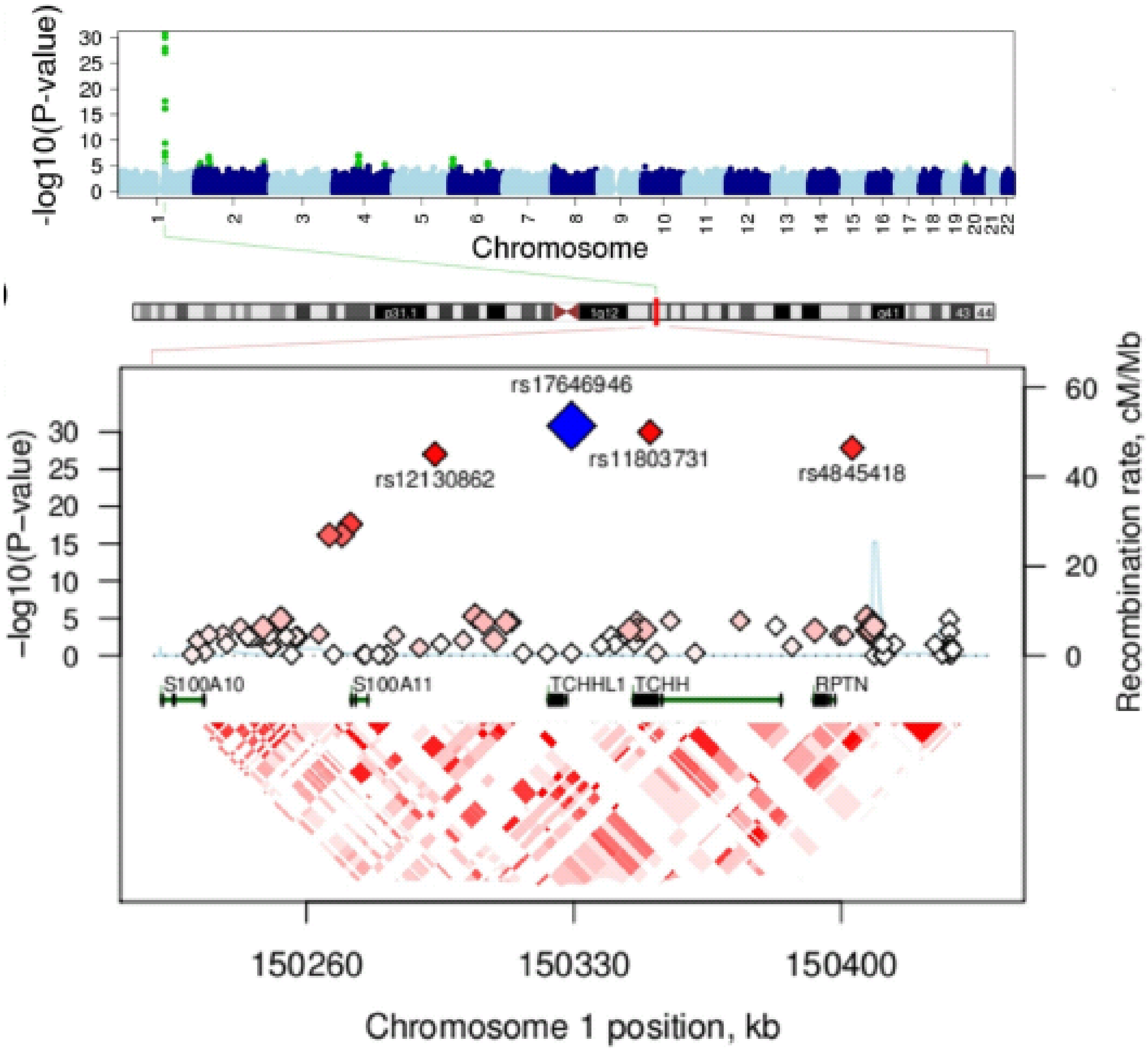
GWAS for curliness in three independent cohorts

$P = 10^{-31}$

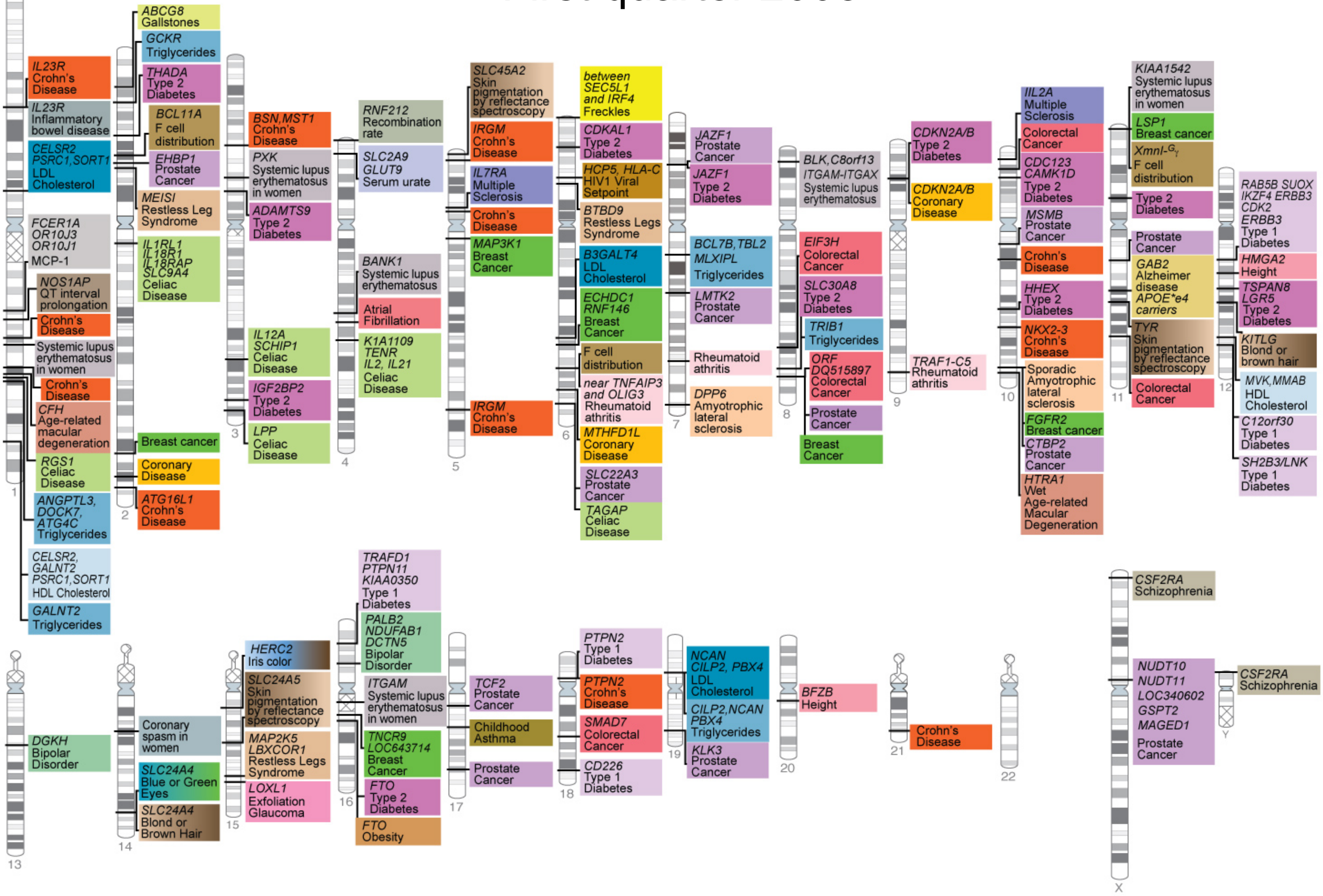
Other peaks



GWAS for hair curliness



First quarter 2008

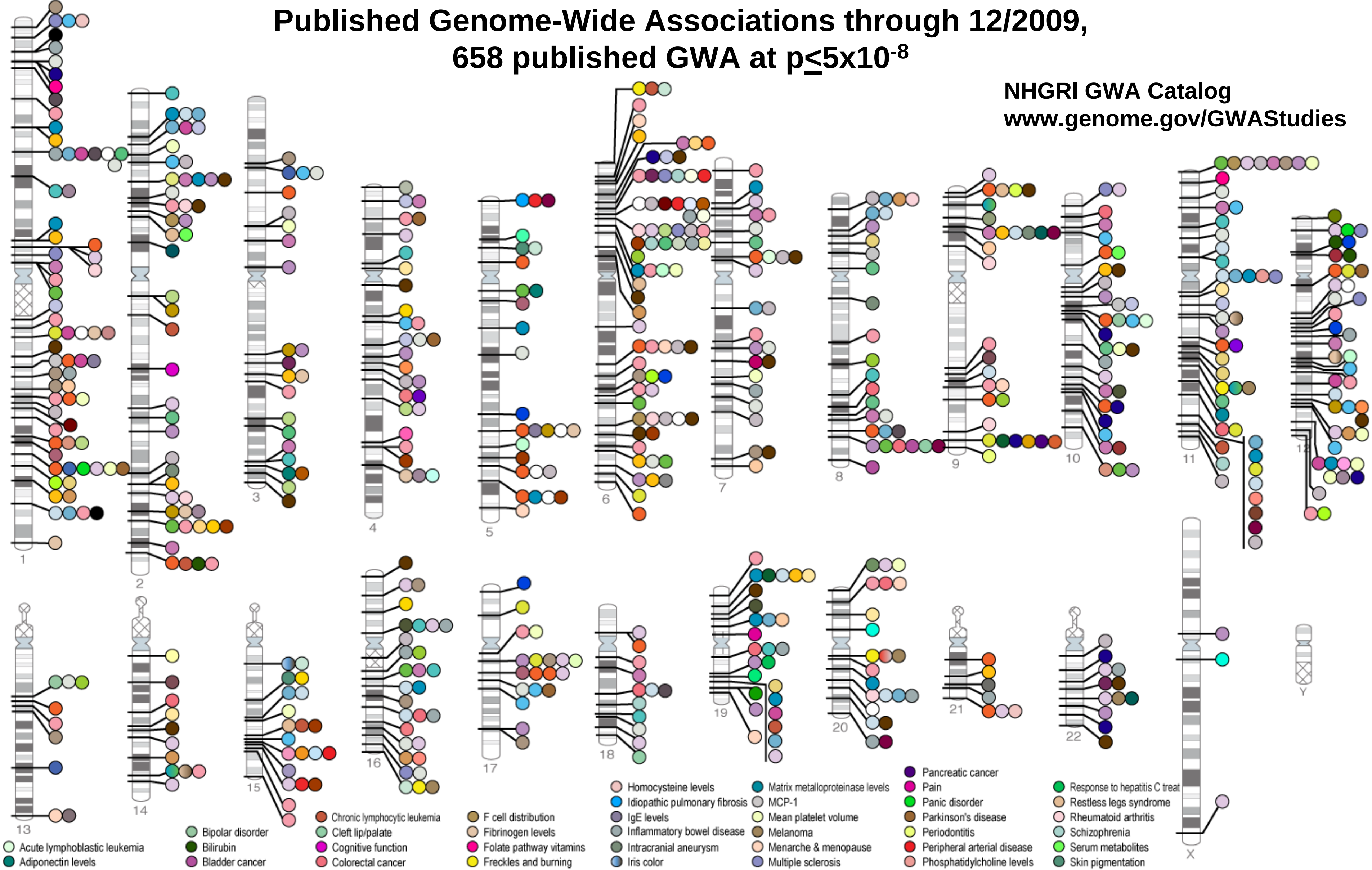

IL23R Crohn's Disease
IL23R Inflammatory bowel disease
CELSR2 PSRC1,SORT1 LDL Cholesterol
FCER1A OR10J3 OR10J1 MCP-1
NOS1AP QT interval prolongation
Crohn's Disease
Systemic lupus erythematosus in women
Crohn's Disease
CFH Age-related macular degeneration
RGS1 Celiac Disease
ANGPTL3, DOCK7, ATG4C Triglycerides
CELSR2, GALNT2 PSRC1,SORT1 HDL Cholesterol
GALNT2 Triglycerides
1
ABCG8 Gallstones
GCKR Triglycerides
THADA Type 2 Diabetes
BCL11A F cell distribution
EHBP1 Prostate Cancer
MEISI Restless Leg Syndrome
IL1RL1 IL18R1 IL18RAP SLC9A4 Celiac Disease
Breast cancer
Coronary Disease
ATG16L1 Crohn's Disease
2
BSN,MST1 Crohn's Disease
PXK Systemic lupus erythematosus in women
ADAMTS9 Type 2 Diabetes
IL12A SCHIP1 Celiac Disease
IGF2BP2 Type 2 Diabetes
LPP Celiac Disease
3
RNF212 Recombination rate
SLC2A9 GLUT9 Serum urate
BANK1 Systemic lupus erythematosus
Atrial Fibrillation
K1A1109 TENR IL2, IL21 Celiac Disease
4
SLC45A2 Skin pigmentation by reflectance spectroscopy
IRGM Crohn's Disease
IL7RA Multiple Sclerosis
Crohn's Disease
MAP3K1 Breast Cancer
IRGM Crohn's Disease
5
between SEC5L1 and IRF4 Freckles
CDKAL1 Type 2 Diabetes
HCP5, HLA-C HIV1 Viral Setpoint
BTBD9 Restless Legs Syndrome
B3GALT4 LDL Cholesterol
ECHDC1 RNF146 Breast Cancer
F cell distribution
near TNFAIP3 and OLIG3 Rheumatoid athritis
MTHFD1L Coronary Disease
SLC22A3 Prostate Cancer
TAGAP Celiac Disease
6
JAZF1 Prostate Cancer
JAZF1 Type 2 Diabetes
BCL7B,TBL2 MLXIPL Triglycerides
LMTK2 Prostate Cancer
Rheumatoid athritis
DPP6 Amyotrophic lateral sclerosis
7
BLK,C8orf13 ITGAM-ITGAX Systemic lupus erythematosus
EIF3H Colorectal Cancer
SLC30A8 Type 2 Diabetes
TRIB1 Triglycerides
ORF DQ515897 Colorectal Cancer
Prostate Cancer
Breast Cancer
8
CDKN2A/B Type 2 Diabetes
CDKN2A/B Coronary Disease
TRAF1-C5 Rheumatoid athritis
9
IIL2A Multiple Sclerosis
Colorectal Cancer
CDC123 CAMK1D Type 2 Diabetes
MSMB Prostate Cancer
Crohn's Disease
HHEX Type 2 Diabetes
NKX2-3 Crohn's Disease
Sporadic Amyotrophic lateral sclerosis
FGFR2 Breast cancer
CTBP2 Prostate Cancer
HTRA1 Wet Age-related Macular Degeneration
10
KIAA1542 Systemic lupus erythematosus in women
LSP1 Breast cancer
XmnI-$^{G}\gamma$ F cell distribution
Type 2 Diabetes
Prostate Cancer
GAB2 Alzheimer disease APOE*e4 carriers
TYR Skin pigmentation by reflectance spectroscopy
Colorectal Cancer
11
RAB5B SUOX IKZF4 ERBB3 CDK2 ERBB3 Type 1 Diabetes
HMGA2 Height
TSPAN8 LGR5 Type 2 Diabetes
KITLG Blond or brown hair
MVK,MMAB HDL Cholesterol
C12orf30 Type 1 Diabetes
SH2B3/LNK Type 1 Diabetes
12
DGKH Bipolar Disorder
13
Coronary spasm in women
SLC24A4 Blue or Green Eyes
SLC24A4 Blond or Brown Hair
14
HERC2 Iris color
SLC24A5 Skin pigmentation by reflectance spectroscopy
MAP2K5 LBXCOR1 Restless Legs Syndrome
LOXL1 Exfoliation Glaucoma
15
TRAFD1 PTPN11 KIAA0350 Type 1 Diabetes
PALB2 NDUFAB1 DCTN5 Bipolar Disorder
ITGAM Systemic lupus erythematosus in women
TNCR9 LOC643714 Breast Cancer
FTO Type 2 Diabetes
FTO Obesity
16
TCF2 Prostate Cancer
Childhood Asthma
Prostate Cancer
17
PTPN2 Type 1 Diabetes
PTPN2 Crohn's Disease
SMAD7 Colorectal Cancer
CD226 Type 1 Diabetes
18
NCAN CILP2, PBX4 LDL Cholesterol
CILP2,NCAN PBX4 Triglycerides
KLK3 Prostate Cancer
19
BFZB Height
20
Crohn's Disease
21
22
CSF2RA Schizophrenia
NUDT10 NUDT11 LOC340602 GSPT2 MAGED1 Prostate Cancer
X
CSF2RA Schizophrenia
Y

Manolio, Brooks, Collins, J. Clin. Invest., May 2008

Stephen Channock

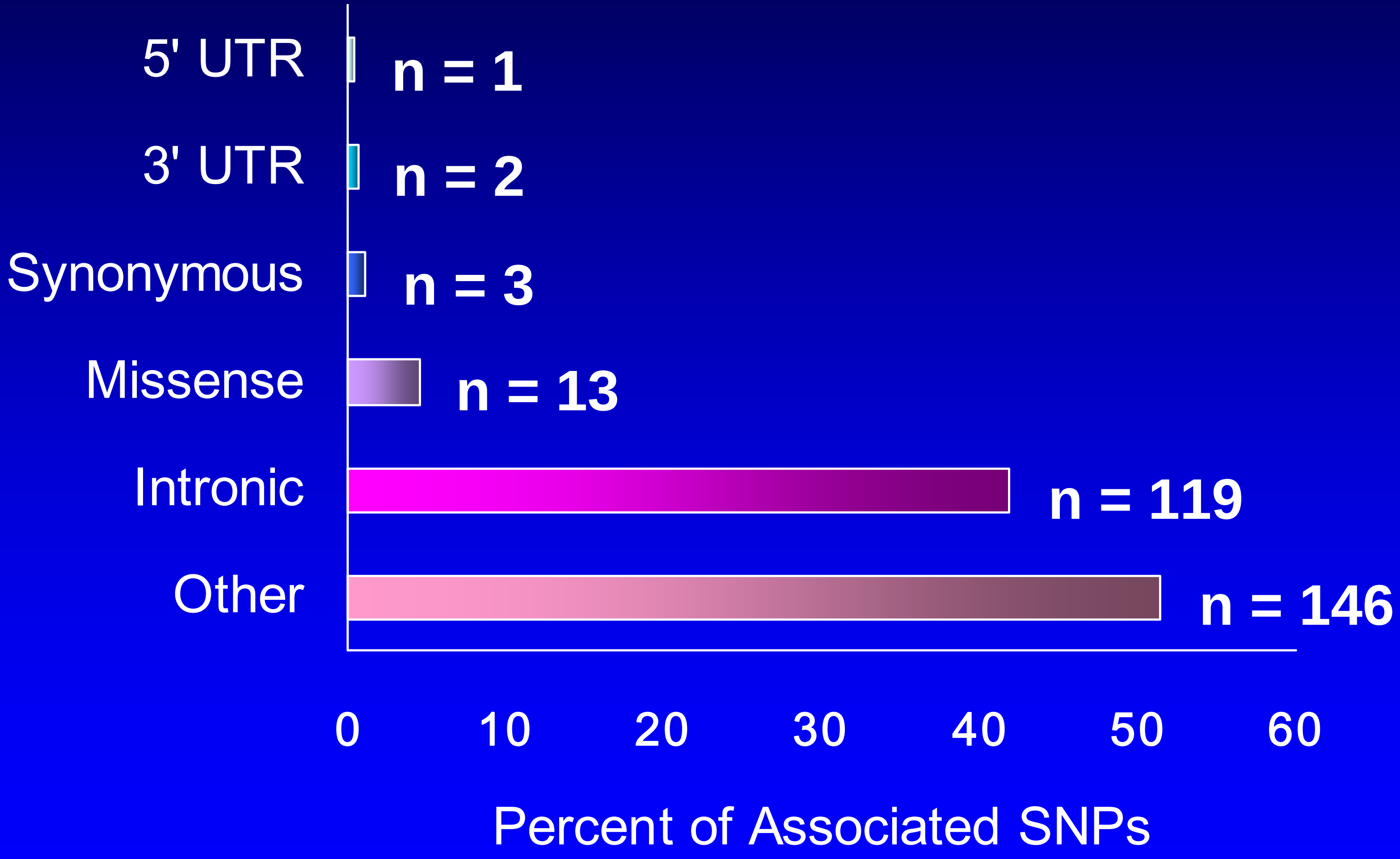
Published Genome-Wide Associations through 12/2009, 658 published GWA at $p \leq 5x10^{-8}$

NHGRI GWA Catalog
www.genome.gov/GWAStudies



- ● Acute lymphoblastic leukemia
- ● Adiponectin levels
- ● Age-related macular degeneration
- ● AIDS progression
- ● Alcohol dependence
- ● Alzheimer disease
- ● Amyotrophic lateral sclerosis
- ● Arterial stiffness
- ● Asthma
- ● Atrial fibrillation
- ● Attention deficit hyperactivity disorder
- ● Autism
- ● Basal cell cancer
- ● Bipolar disorder
- ● Bilirubin
- ● Bladder cancer
- ● Blond or brown hair
- ● Blood pressure
- ● Blue or green eyes
- ● BMI, waist circumference
- ● Bone density
- ● Breast cancer
- ● C-reactive protein
- ● Cardiac structure/function
- ● Carnitine levels
- ● Carotenoid/tocopherol levels
- ● Celiac disease
- ● Chronic lymphocytic leukemia
- ● Cleft lip/palate
- ● Cognitive function
- ● Colorectal cancer
- ● Coronary disease
- ● Creutzfeldt-Jakob disease
- ● Crohn's disease
- ● Cutaneous nevi
- ● Dermatitis
- ● Drug-induced liver injury
- ● Eosinophil count
- ● Erythrocyte parameters
- ● Esophageal cancer
- ● Essential tremor
- ● Exfoliation glaucoma
- ● F cell distribution
- ● Fibrinogen levels
- ● Folate pathway vitamins
- ● Freckles and burning
- ● Gallstones
- ● Glioma
- ● Glycemic traits
- ● Hair color
- ● Hair morphology
- ● Heart rate
- ● Height
- ● Hepatitis
- ● Hirschsprung's disease
- ● HIV-1 control
- ● HDL cholesterol
- ● Homocysteine levels
- ● Idiopathic pulmonary fibrosis
- ● IgE levels
- ● Inflammatory bowel disease
- ● Intracranial aneurysm
- ● Iris color
- ● Iron status markers
- ● Ischemic stroke
- ● Juvenile idiopathic arthritis
- ● Kidney stones
- ● Leprosy
- ● LDL cholesterol
- ● Liver enzymes
- ● LP (a) levels
- ● Lung cancer
- ● Malaria
- ● Male pattern baldness
- ● Matrix metalloproteinase levels
- ● MCP-1
- ● Mean platelet volume
- ● Melanoma
- ● Menarche & menopause
- ● Multiple sclerosis
- ● Myeloproliferative neoplasms
- ● Narcolepsy
- ● Nasopharyngeal cancer
- ● Neuroblastoma
- ● Nicotine dependence
- ● Nonsyndromic cleft lip w/wo cleft palate
- ● Obesity
- ● Open personality
- ● Otosclerosis
- ● Other metabolic traits
- ● Ovarian cancer
- ● Pancreatic cancer
- ● Pain
- ● Panic disorder
- ● Parkinson's disease
- ● Periodontitis
- ● Peripheral arterial disease
- ● Phosphatidylcholine levels
- ● Primary biliary cirrhosis
- ● Prostate cancer
- ● Protein levels
- ● Psoriasis
- ● Pulmonary funct. COPD
- ● QT interval
- ● Quantitative traits
- ● Recombination rate
- ● Red vs.non-red hair
- ● Renal function
- ● Response to antipsychotic therapy
- ● Response to hepatitis C treat
- ● Restless legs syndrome
- ● Rheumatoid arthritis
- ● Schizophrenia
- ● Serum metabolites
- ● Skin pigmentation
- ● Soluble E-selectin
- ● Soluble ICAM-1
- ● Speech perception
- ● Sphingolipid levels
- ● Statin-induced myopathy
- ● Stroke
- ● Systemic lupus erythematosus
- ● Testicular germ cell tumor
- ● Thyroid cancer
- ● Total cholesterol
- ● Triglycerides
- ● Type 1 diabetes
- ● Type 2 diabetes
- ● Urate
- ● Venous thromboembolism
- ● Vitamin B12 levels
- ● Warfarin dose
- ● Weight
- ● White cell count
- ● YKL-40 levels

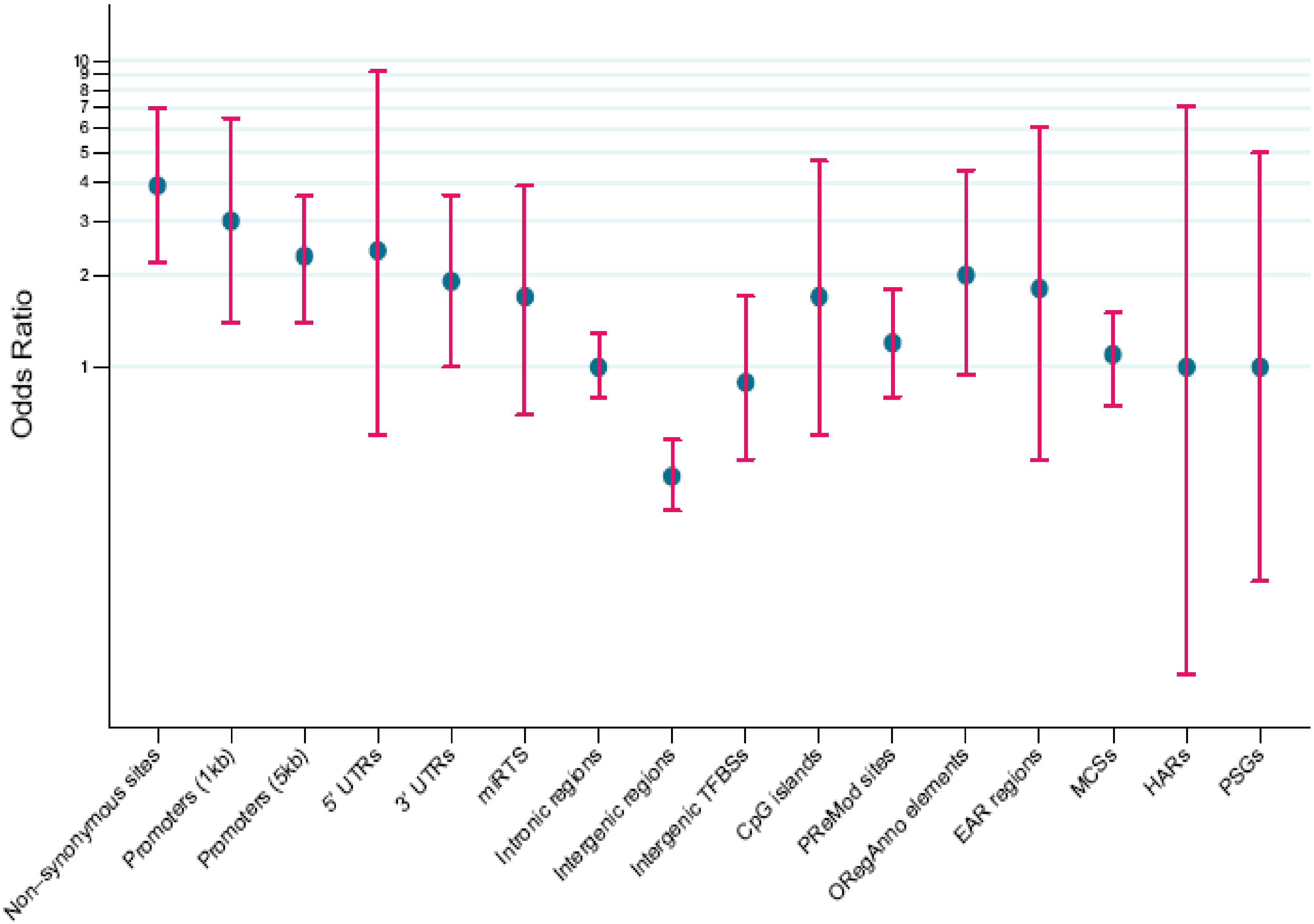
Functional Classification of 284 SNPs Associated with Complex Traits



http://www.genome.gov/gwastudies/

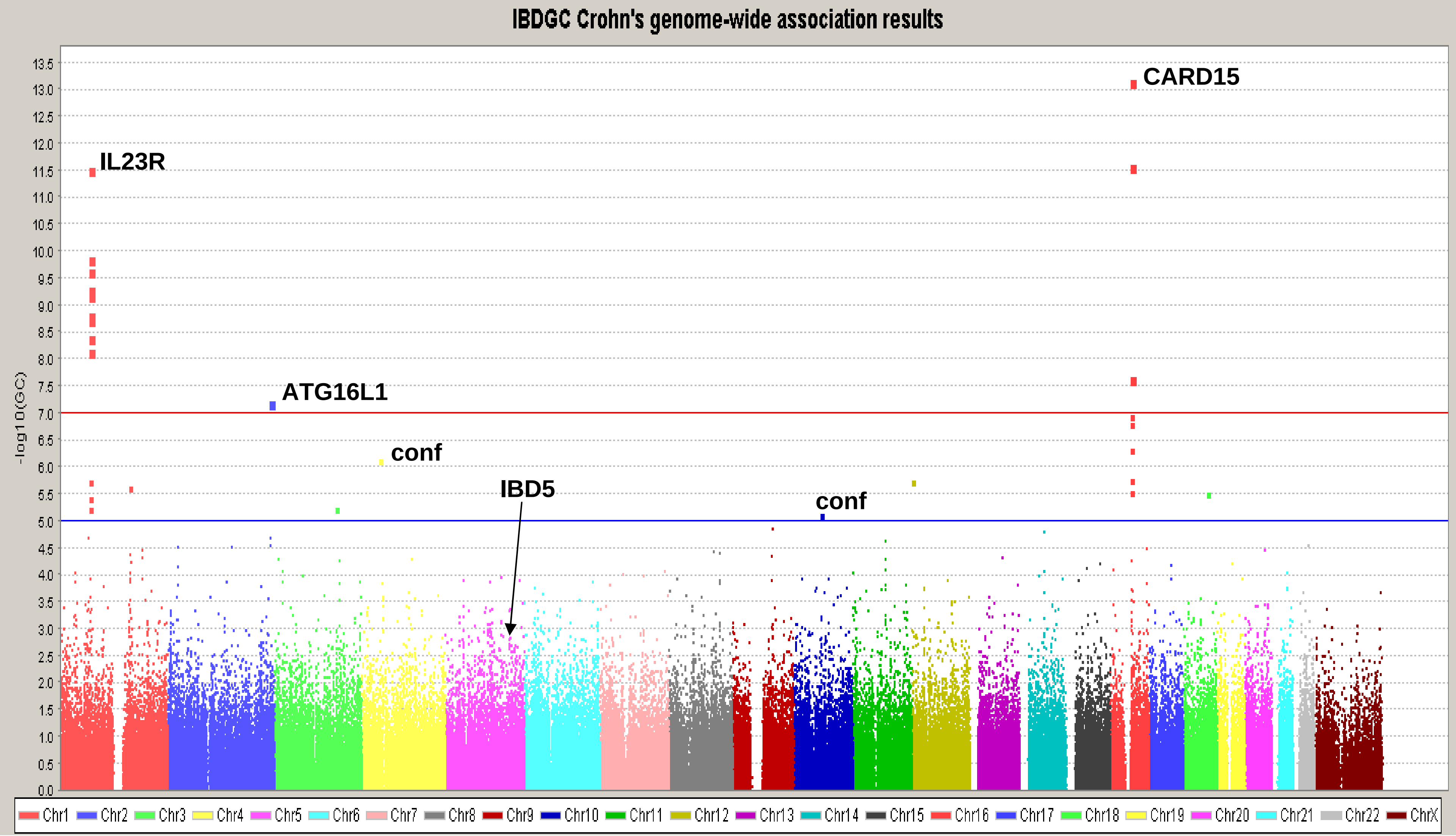
Stephen Channock

Enrichment/depletion analysis after adjusting for 'hitchhiking' effects from non–synonymous sites

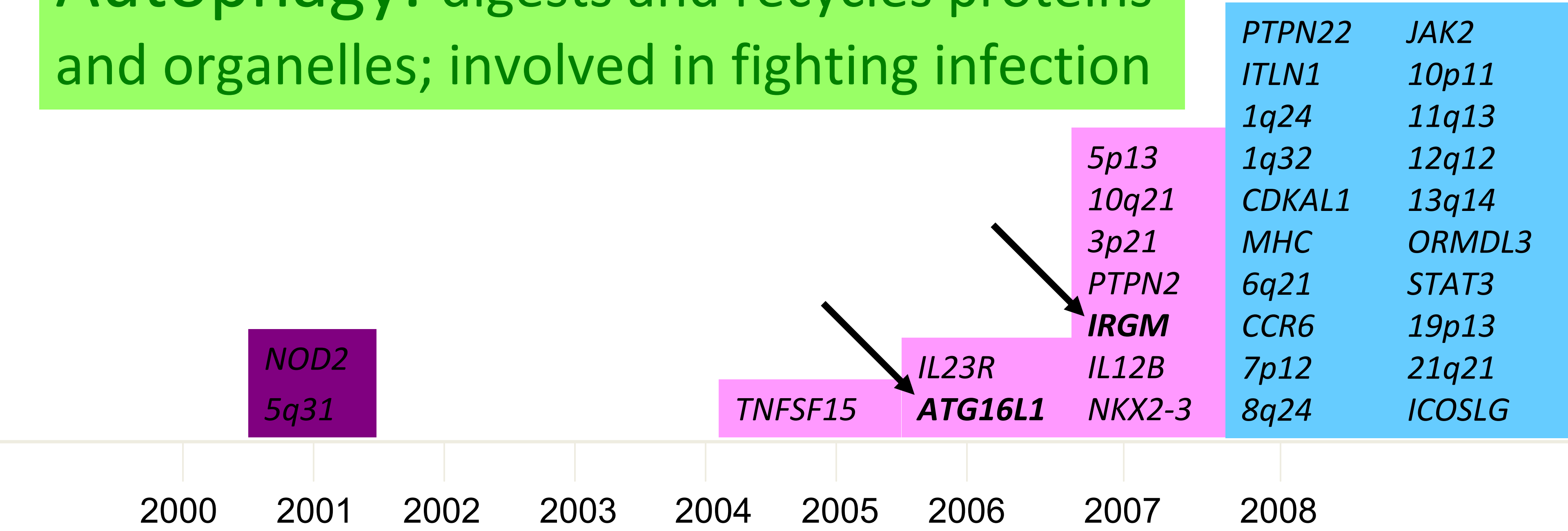


Proc Natl Acad Sci U S A. 2009 Jun 9;106(23):9362-7.

How GWAS can change the research paradigm example: Crohn's Disease (inflammatory bowel)



Unexpected pathway for Crohn's: Autophagy: digests and recycles proteins and organelles; involved in fighting infection



NOD2
5q31

TNFSF15

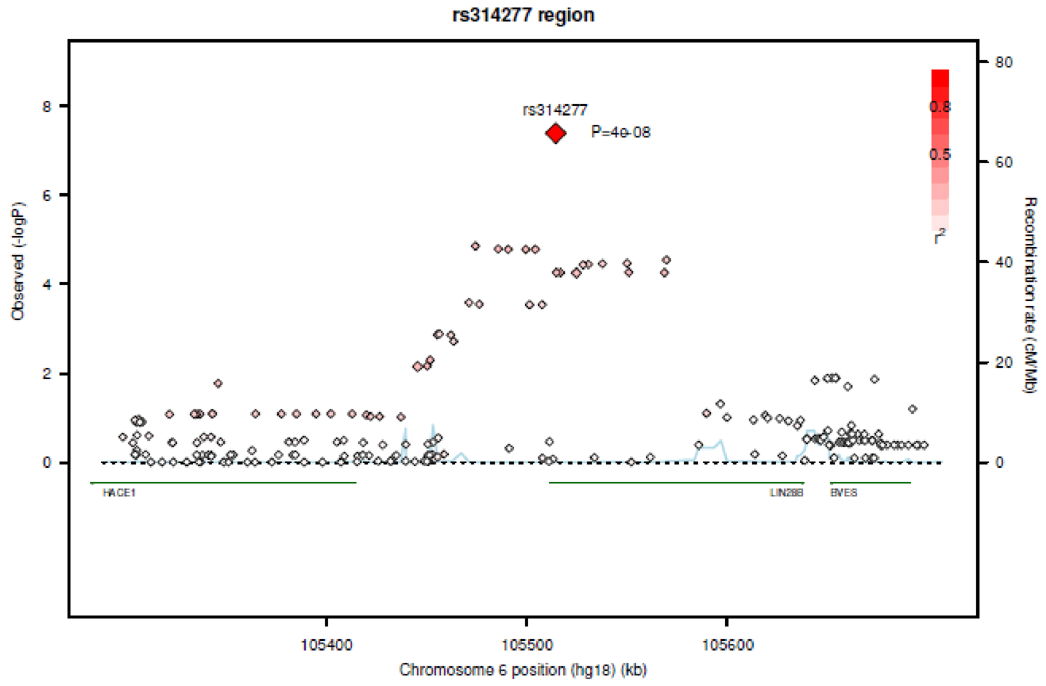
IL23R
ATG16L1

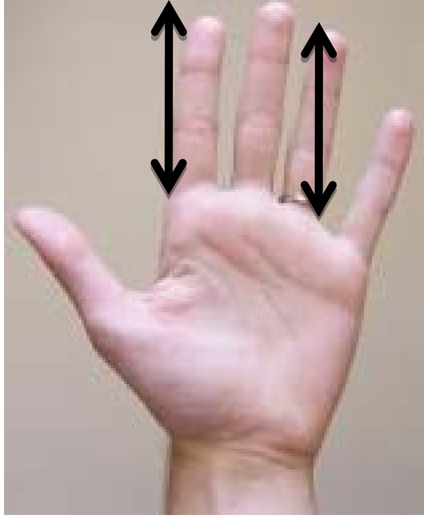
5p13
10q21
3p21
PTPN2
IRGM
IL12B
NKX2-3

PTPN22
ITLN1
1q24
1q32
CDKAL1
MHC
6q21
CCR6
7p12
8q24

JAK2
10p11
11q13
12q12
13q14
ORMDL3
STAT3
19p13
21q21
ICOSLG

Now ~65 genes contributing 12.5% variance in liability

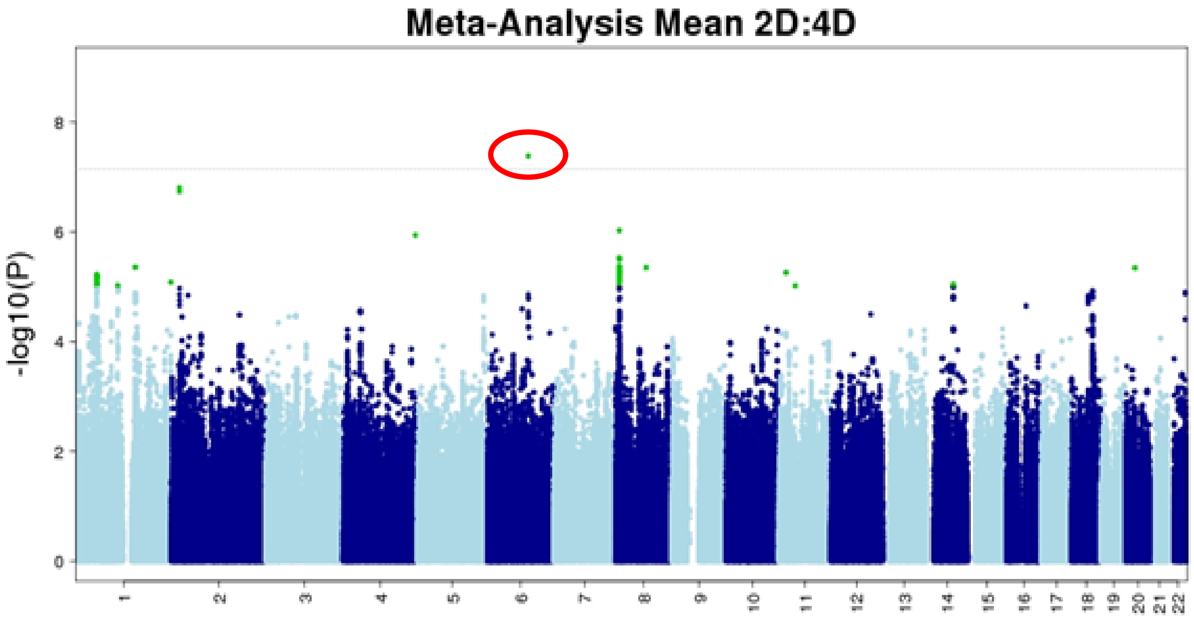




Ratio of 2nd to 4th finger length

Associated with:

- testosterone exposure
- aggression
- ADHD
- homosexuality
- fertility
- others

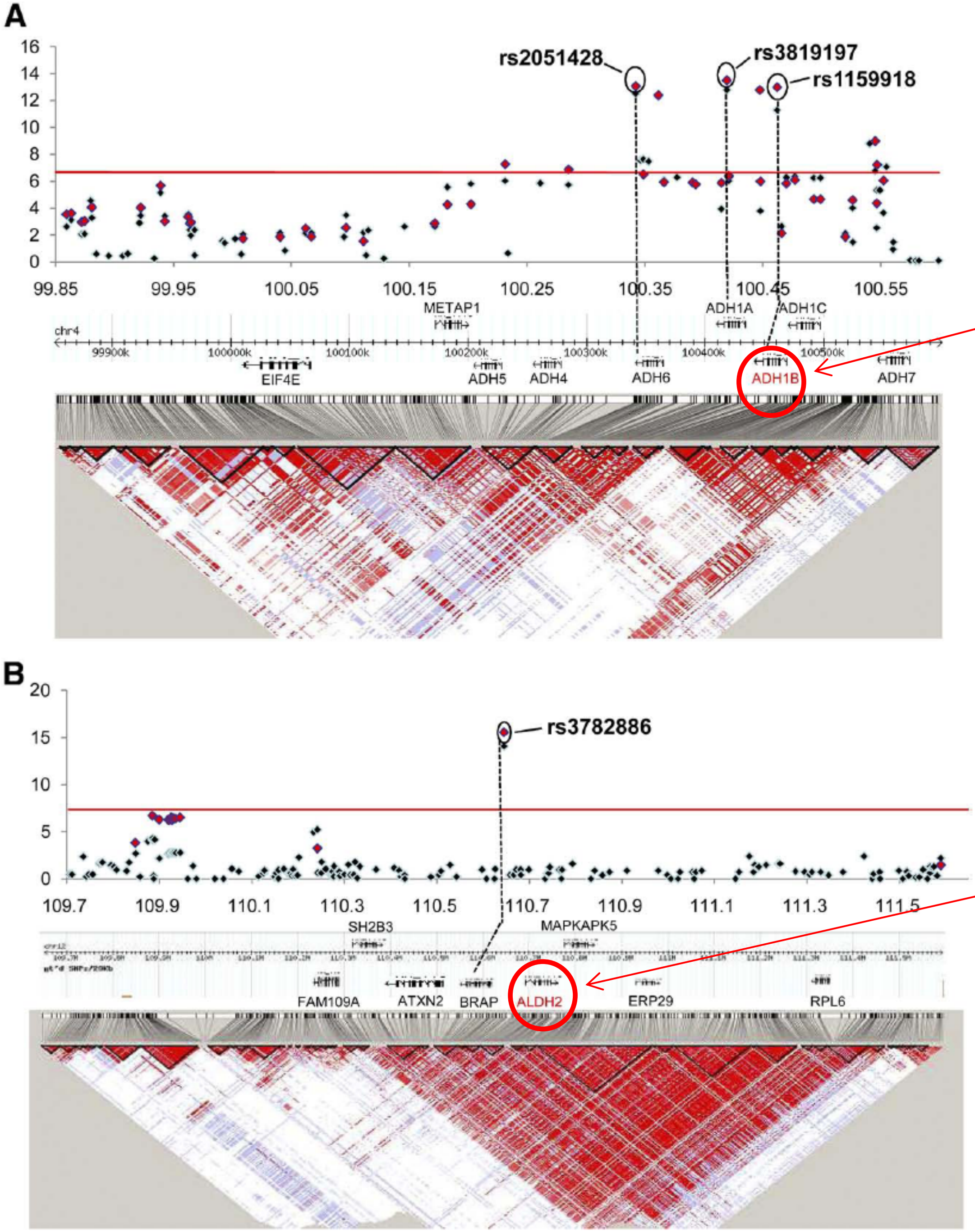


Medland, Martin, Evans (in press) *AJHG*

LIN28B variant associated with:

- 2D:4D ratio
- Age of menarche
- Menopause
- Height

GWAS for esophageal ca



ADH1B

ALDH2

GASTROENTEROLOGY 2009;137:1768–1775

Functional Variants in *ADH1B* and *ALDH2* Coupled With Alcohol and Smoking Synergistically Enhance Esophageal Cancer Risk

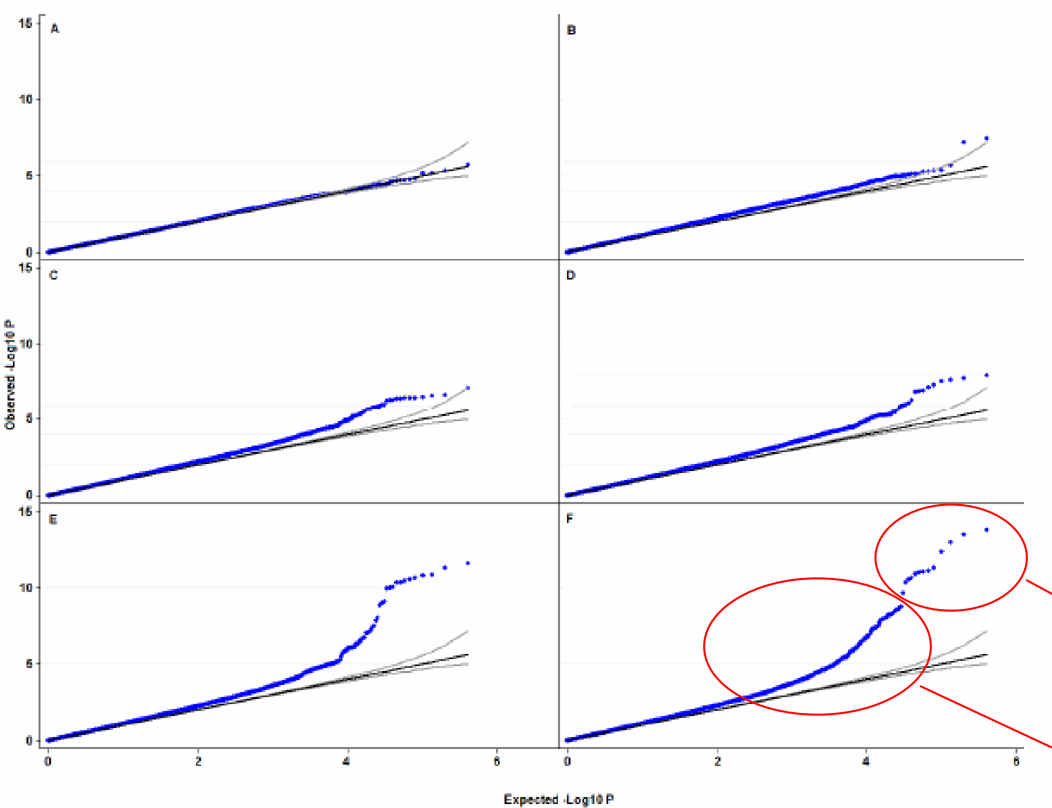
RI CUI,* YOICHIRO KAMATANI,* ATSUSHI TAKAHASHI,‡ MASAYUKI USAMI,* NAOYA HOSONO,§ TAKAHISA KAWAGUCHI,|| TATSUHIKO TSUNODA,|| NAOYUKI KAMATANI,‡ MICHIAKI KUBO,§ YUSUKE NAKAMURA,*,¶ and KOICHI MATSUDA*

- Nature. 2009 Dec 17;462(7275):868-74.
- **Parental origin of sequence variants associated with complex diseases.**
- Kong A, …………., Stefansson K, Altshuler D, Boehnke M, McCarthy MI.
- deCODE genetics, Sturlugata 8, 101 Reykjavík, Iceland. kong@decode.is

- Effects of susceptibility variants may depend on from which parent they are inherited. Although many associations between sequence variants and human traits have been discovered through genome-wide associations, the impact of parental origin has largely been ignored. Here we show that for 38,167 Icelanders genotyped using single nucleotide polymorphism (SNP) chips, the parental origin of most alleles can be determined. We focused on SNPs that associate with diseases and are within 500 kilobases of known imprinted genes. Five SNPs - one with breast cancer, one with basal-cell carcinoma and three with type 2 diabetes-have parental-origin-specific associations. These variants are located in two genomic regions, 11p15 and 7q32, each harbouring a cluster of imprinted genes. Furthermore, we observed a novel association between the SNP rs2334499 at 11p15 and type 2 diabetes. Here the allele that confers risk when paternally inherited is protective when maternally transmitted.

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?

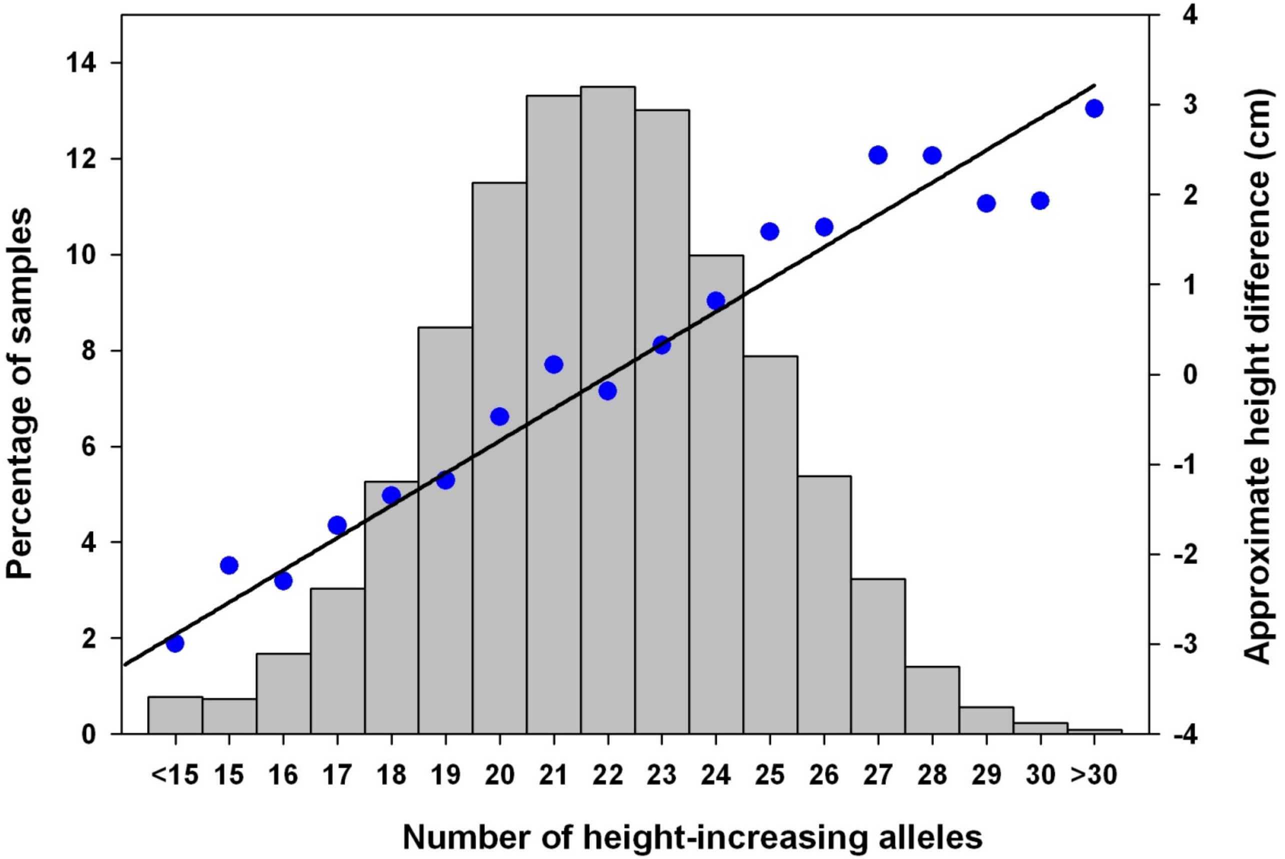
- Large numbers are needed to detect QTLs !!!
- Collaboration is the name of the game !!!

Hedgehog signaling, cell cycle, and extra-cellular matrix genes over-represented

Candidate gene	Monogenic	Knockout mouse	Details*
ZBTB38	-	-	Transcription factor.
CDK6	-	Yes	Involved in the control of the cell cycle.
HMGA2	Yes	Yes	Chromatin architectural factors
GDF5	Yes	Yes	Involved in bone formation
LCORL	-	-	May act as transcription activator
LOC387103	-	-	Not known
EFEMP1	Yes	-	Extra-cellular matrix
C6orf106	-	-	Not known
PTCH1	Yes	Yes	Hedgehog signalling
SPAG17	-	-	Not known
SOCS2	-	Yes	Regulates cytokine signal transduction
HHIP	-	-	Hedgehog signaling
ZNF678	-	-	Transcription factor
DLEU7	-	-	Not known
SCMH1	-	Yes	Polycomb protein
ADAMTSL3	-	-	Extra-cellular matrix
IHH	Yes	Yes	Hedgehog signaling
ANAPC13	-	-	Cell cycle
ACAN	Yes	Yes	Extra-cellular matrix
DYM	Yes	-	Not known

Weedon et al. (i2008) *Nat Genet*

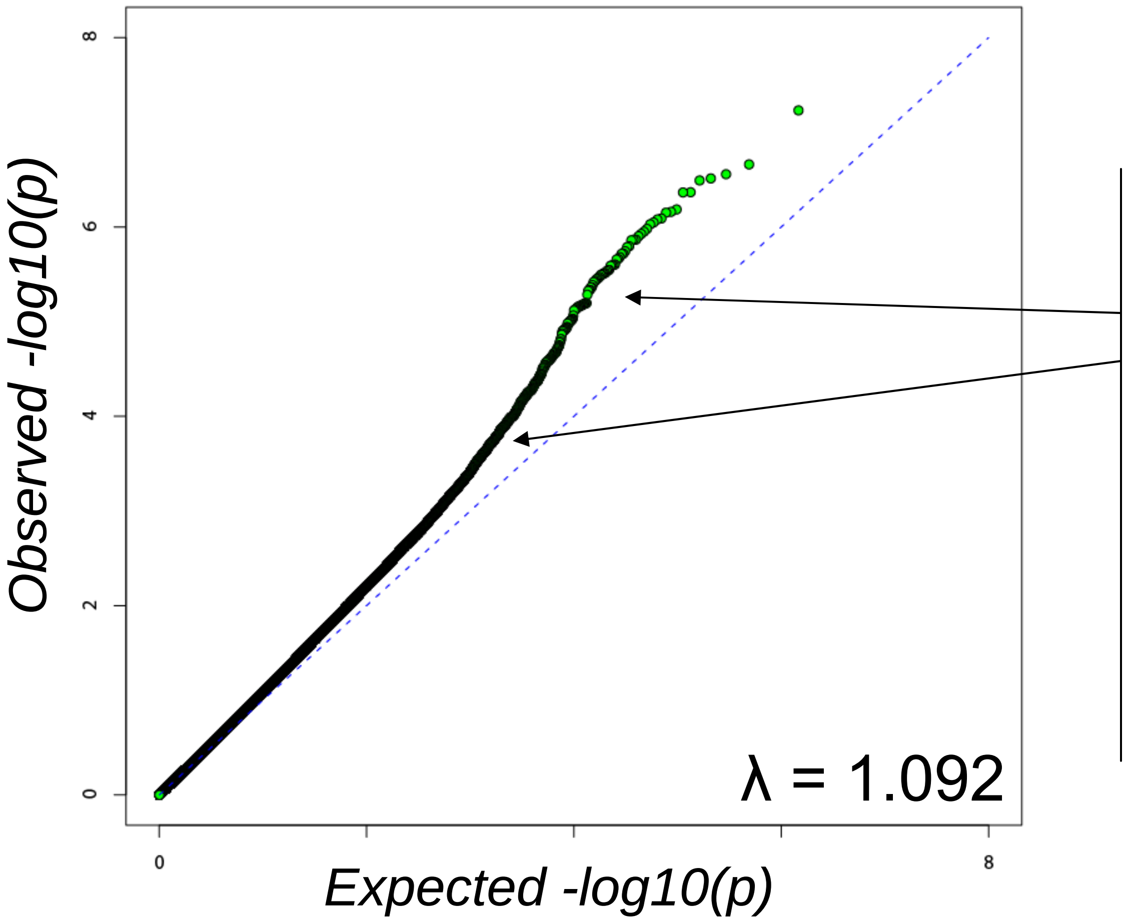
The combined impact of the 20 SNPS with a P < 5 x 10^{-7}



- The 20 SNPs explain only ~3% of the variation of height
- Lots more genes to find – but extremely large numbers needed

Weedon et al. (i2008) *Nat Genet*

Schizophrenia (ISC) Q-Q plot



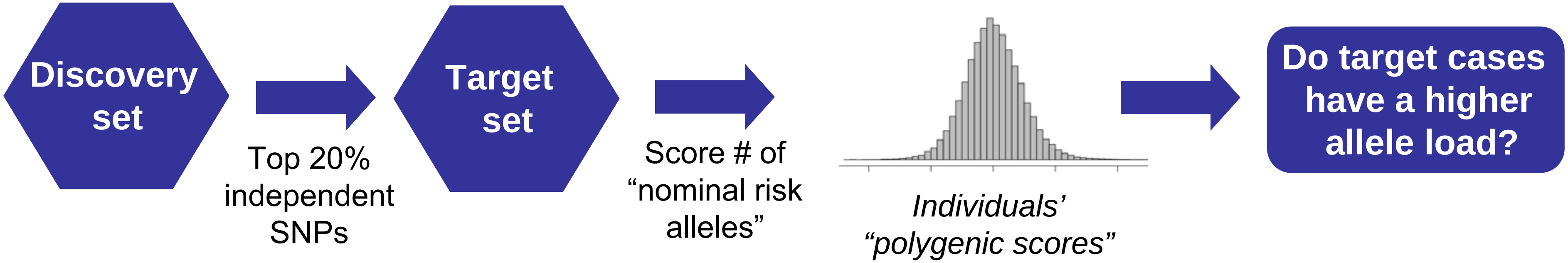
Consistent with:

Stratification?

Genotyping bias?

Distribution of true polygenic effects?

Indexing polygenic variance with large sets of weakly associated alleles



Discovery set

Top 20% independent SNPs

Target set

Score # of "nominal risk alleles"

Individuals' "polygenic scores"

Do target cases have a higher allele load?

IS

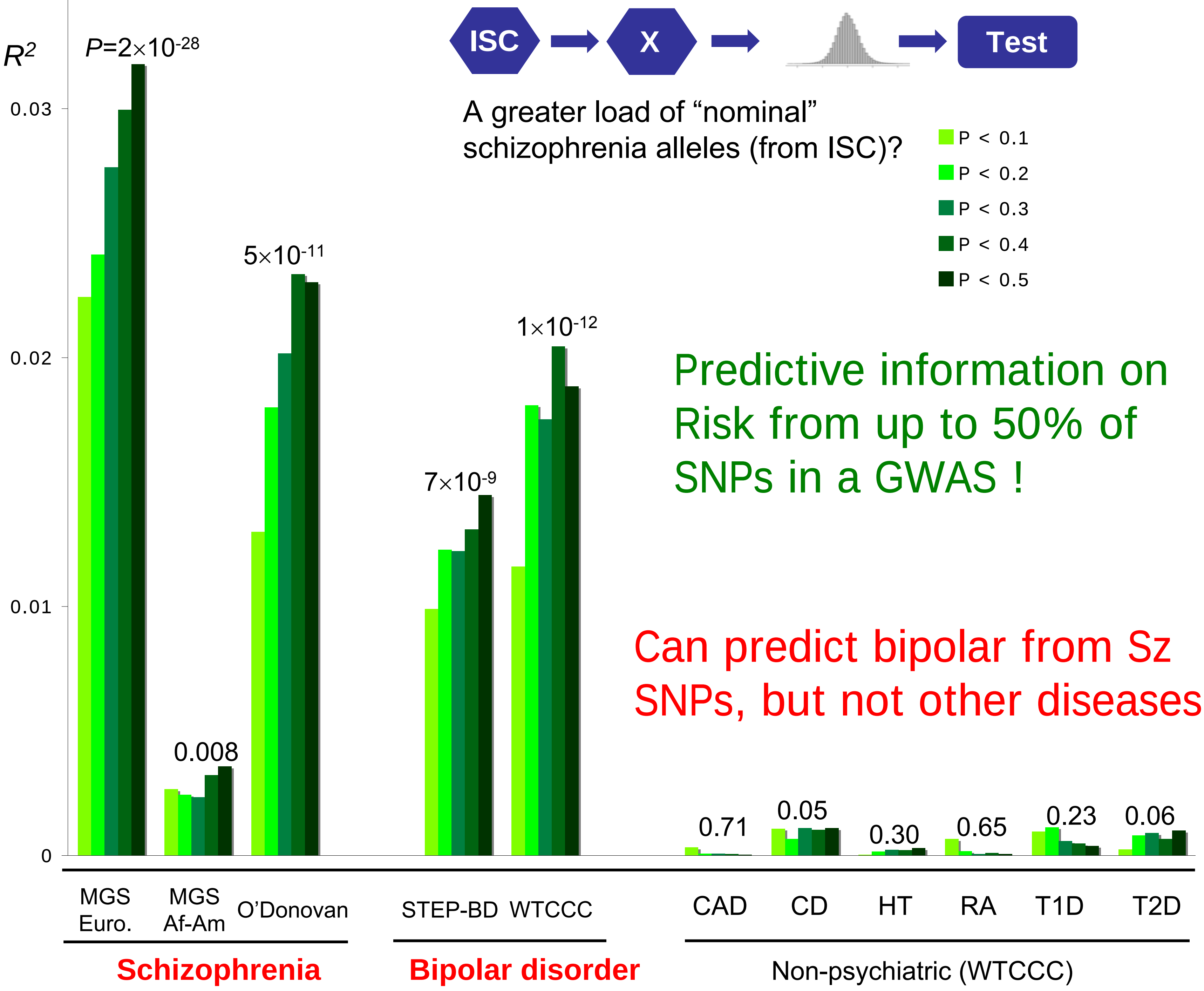
→

→ Independent SCZ studies (MGS,

→ Bipolar disorder (STEP-BD,

→ Non-psychiatric disease

Douglas Levinson, Pablo Gejman,
Jianxin Shi and colleagues



A greater load of "nominal" schizophrenia alleles (from ISC)?

Predictive information on Risk from up to 50% of SNPs in a GWAS !

Can predict bipolar from Sz SNPs, but not other diseases

GWAS' greatest success: T1D

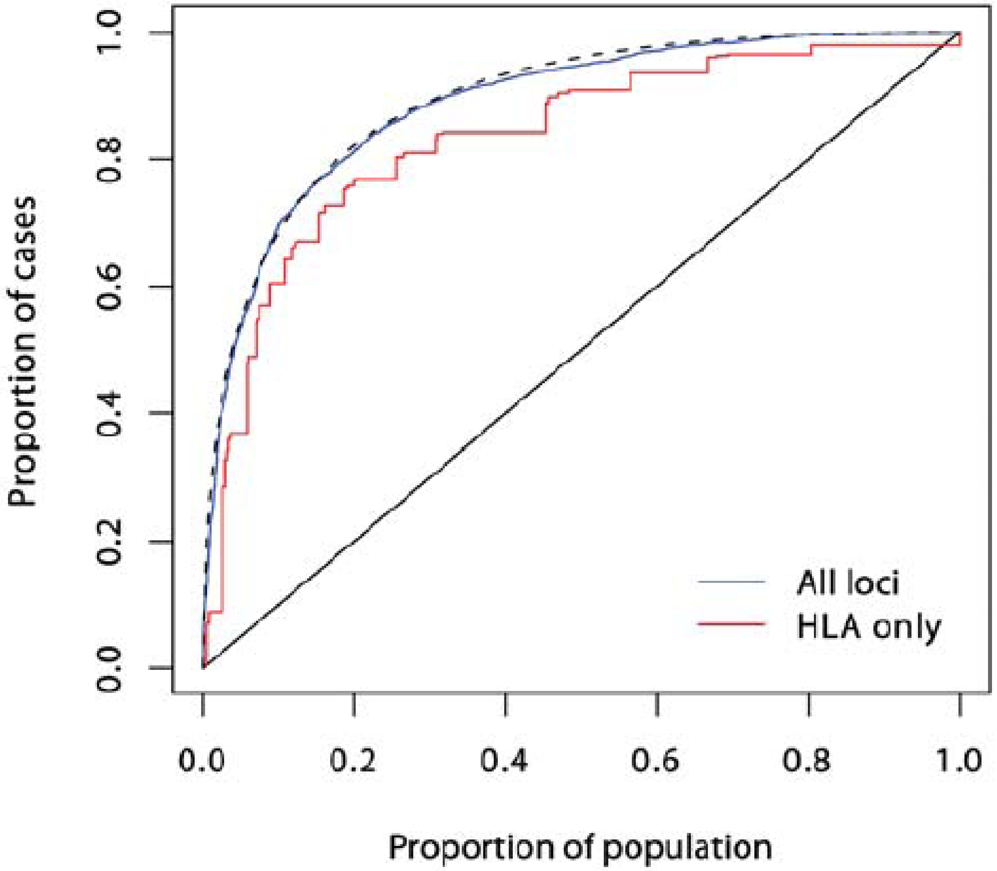


Figure 5. ROC curve prediction from all the SNPs listed in Supplementary Table 1 in Text S1 (in blue). The prediction curve using the six MHC SNPs alone is shown in red, and the dashed curve corresponds to a polygenic multiplicative model with $\lambda_s = 4.75$.

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.

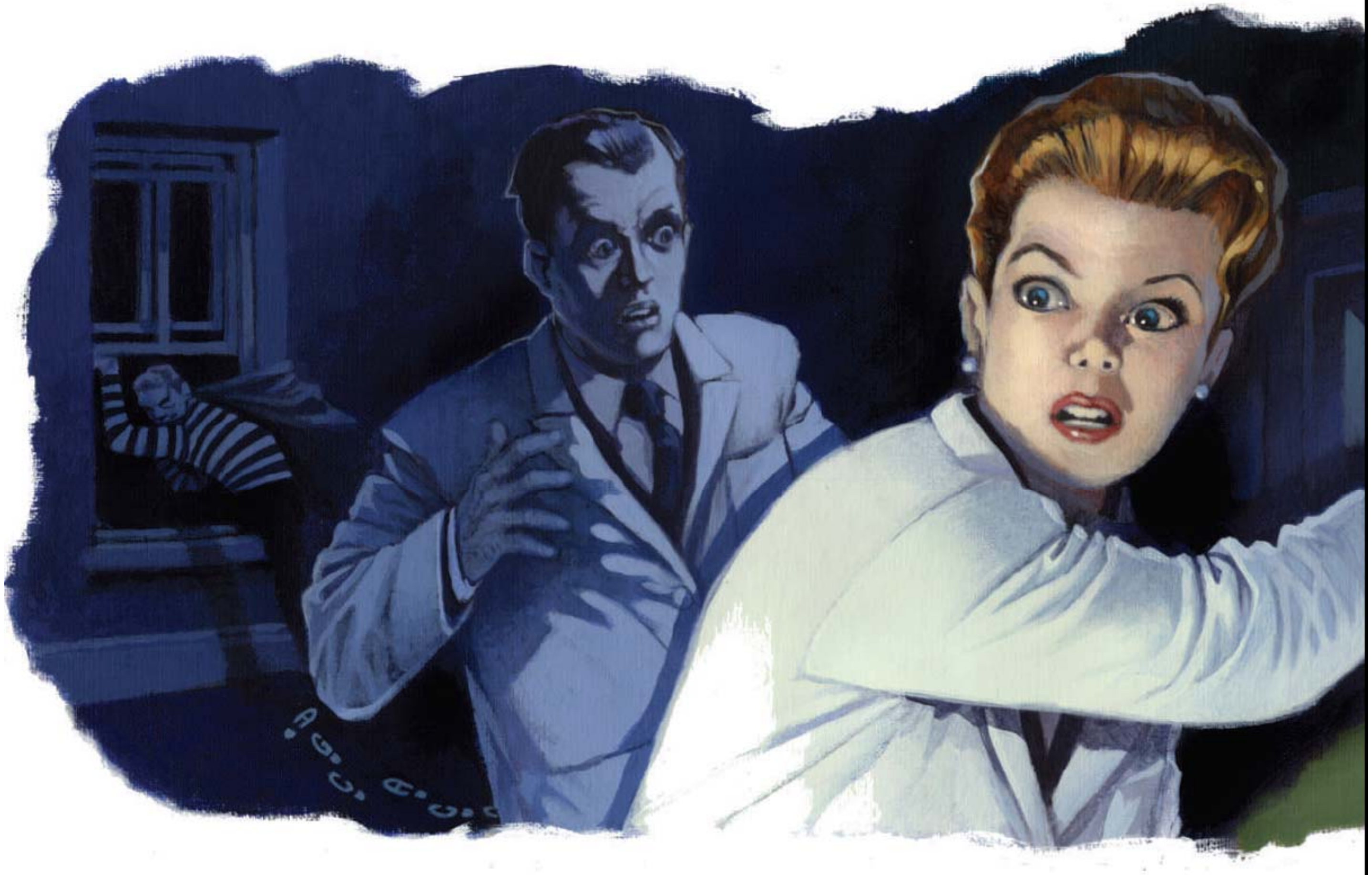
PLOS GENETICS

OPEN ACCESS Freely available online

Viewpoints

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

David G. Clayton*



The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Possible explanations for missing heritability

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

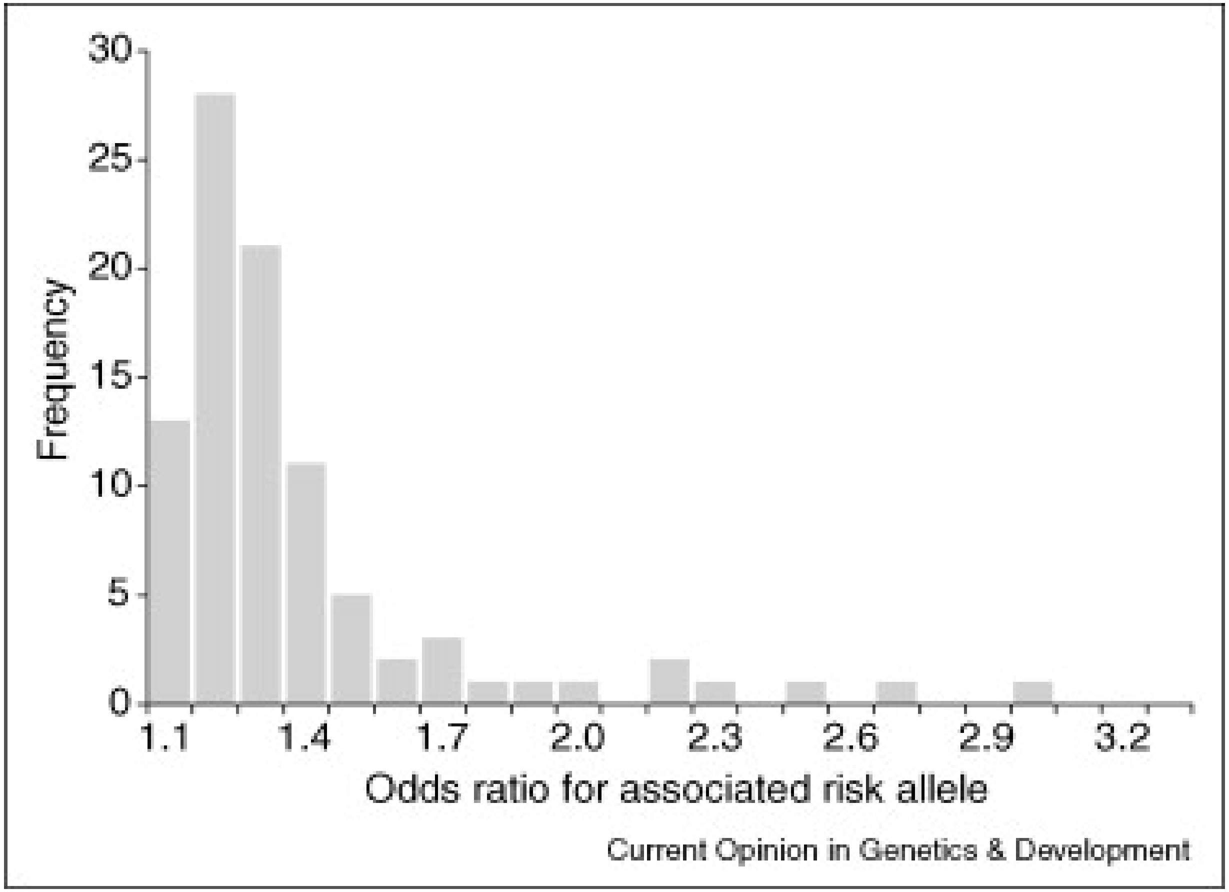
- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect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

Effects sizes of validated variants from 1st 16 GWAS studies



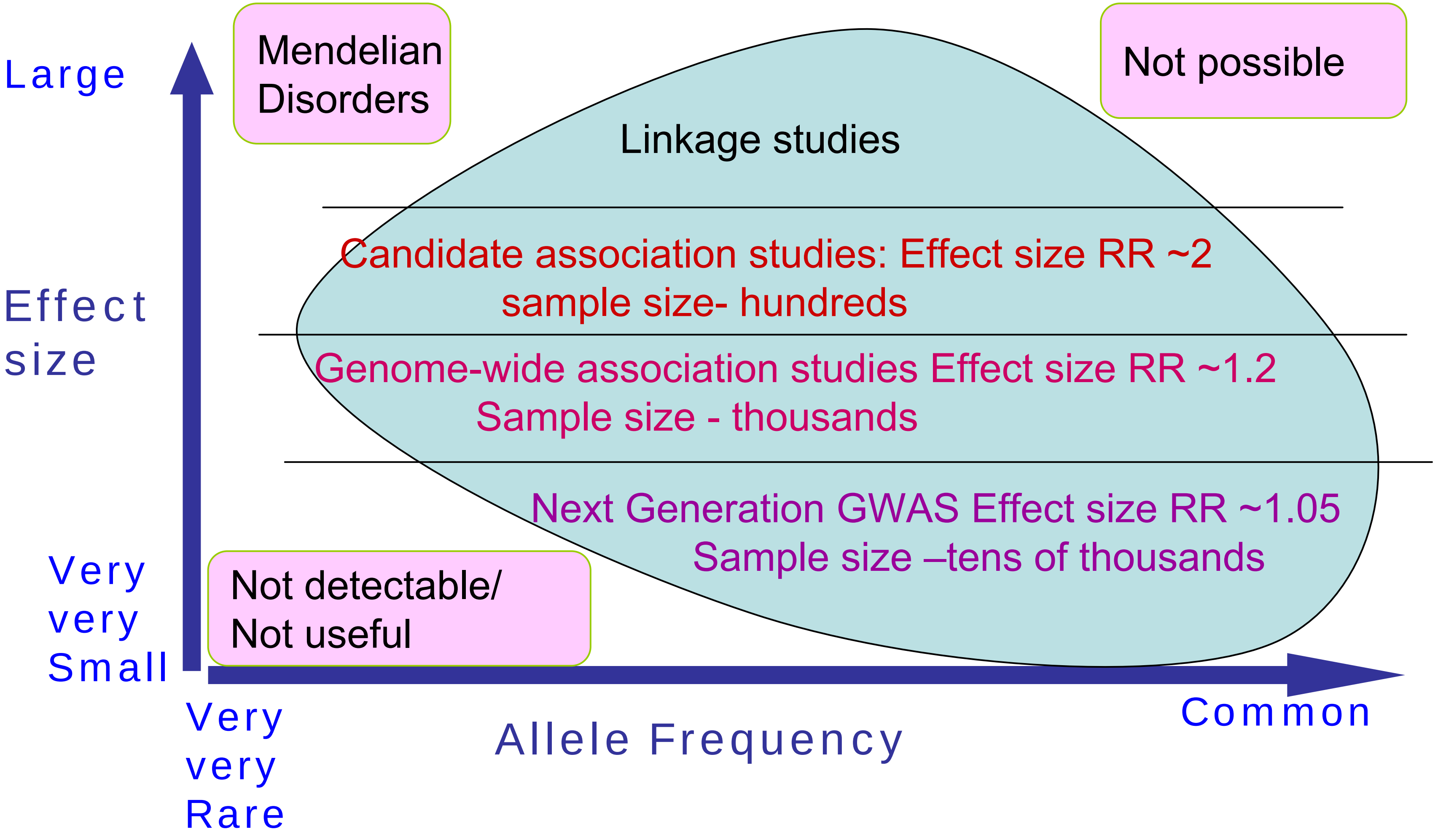
Most effect sizes are very small <1.1

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



Large

Mendelian Disorders

Not possible

Linkage studies

Effect size

Candidate association studies: Effect size RR ~2
sample size- hundreds

Genome-wide association studies Effect size RR ~1.2
Sample size - thousands

Next Generation GWAS Effect size RR ~1.05
Sample size –tens of thousands

Very very Small

Not detectable/
Not useful

Very very Rare

Allele Frequency

Common

GIANT consortium

For those interested in numbers, there are currently 418 authors, from 86 cohorts, affiliated to 240 institutions contributing to three papers combined, with the largest number contributing to the BMI paper. Total N ~100,000 cases !

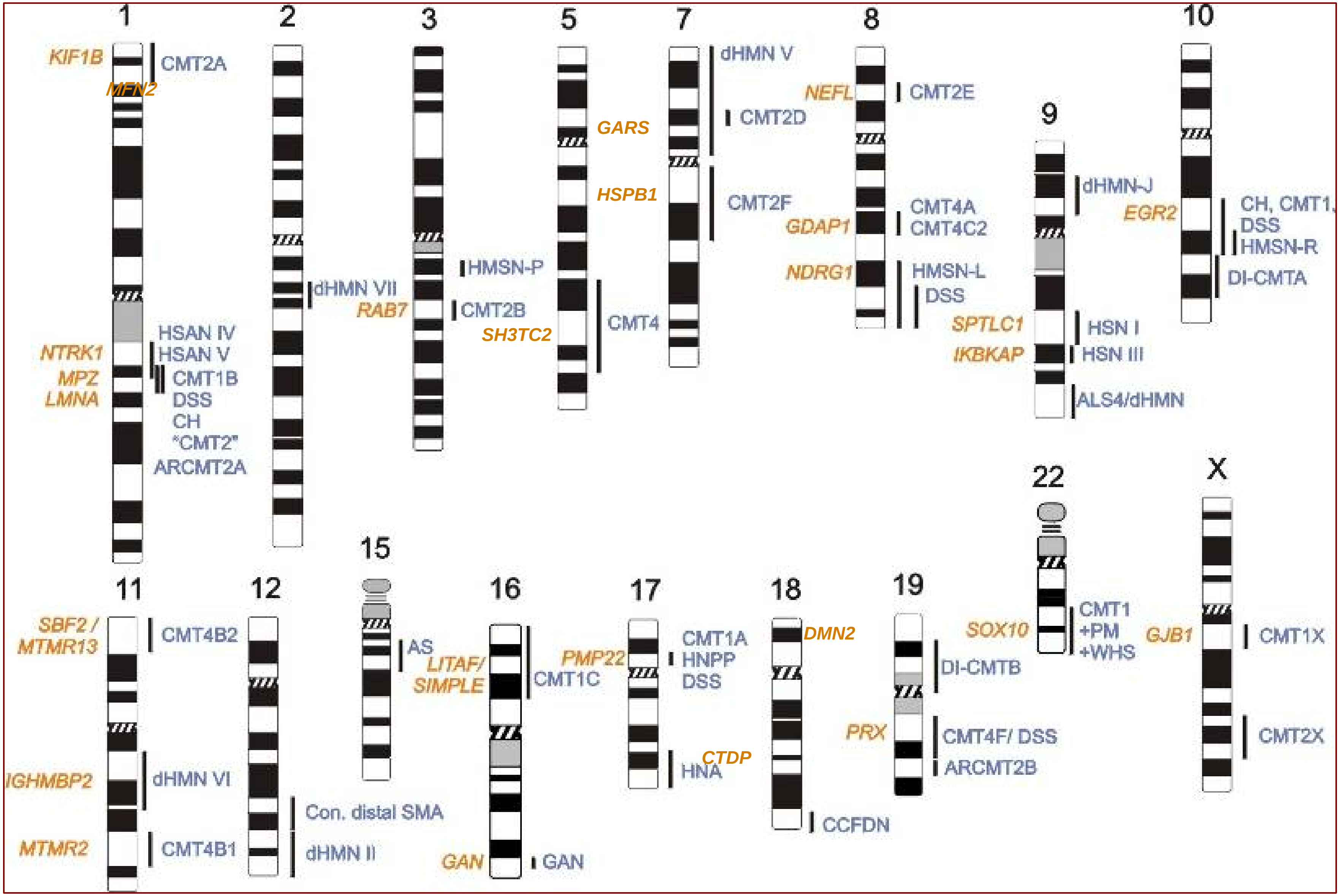
Possible explanations for missing heritability

(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

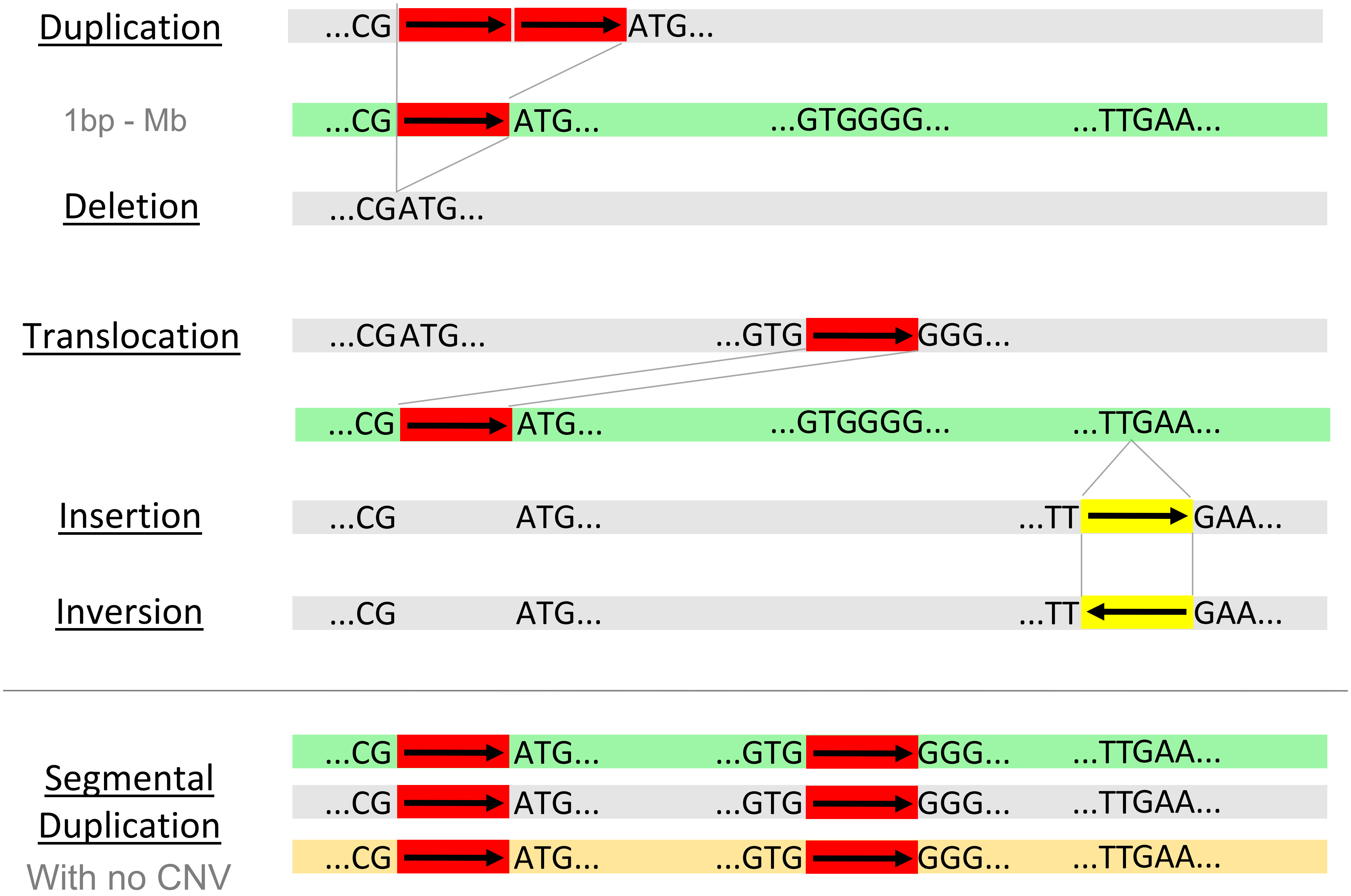
(in order of increasing plausibility ?)

- 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

1bp - Mb

Deletion

Translocation

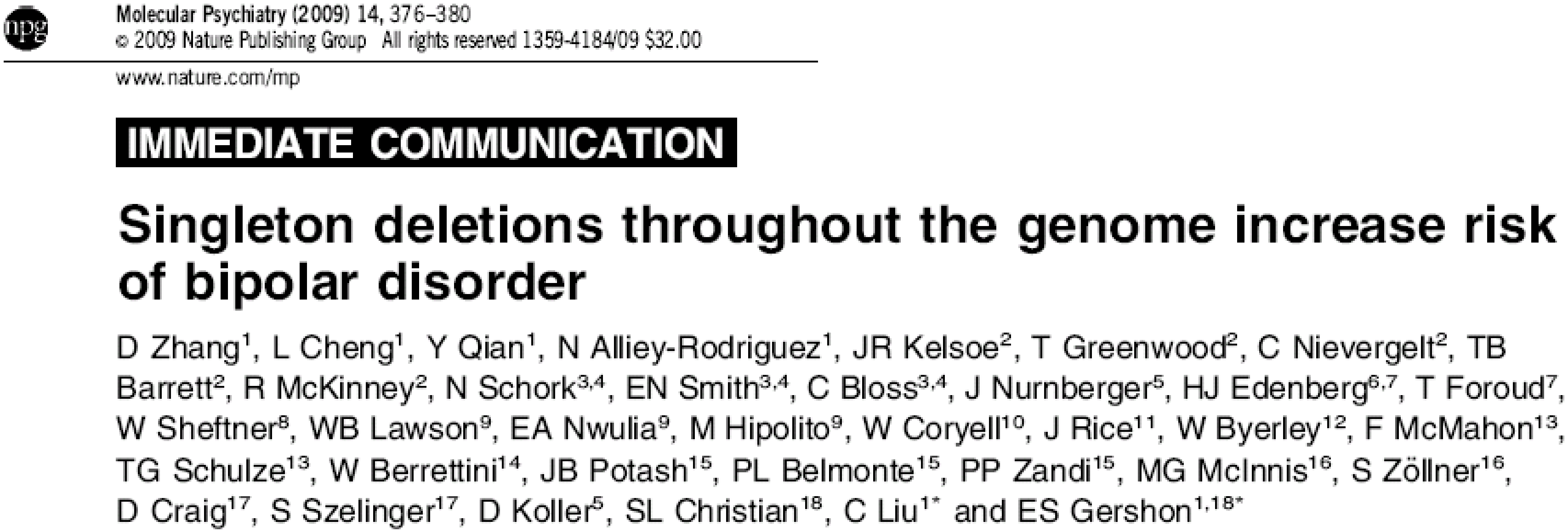
Insertion

Inversion

Segmental Duplication

With no CNV

For example: Bipolar disorder



Molecular Psychiatry (2009) 14, 376–380
© 2009 Nature Publishing Group All rights reserved 1359-4184/09 $32.00
www.nature.com/mp

IMMEDIATE COMMUNICATION

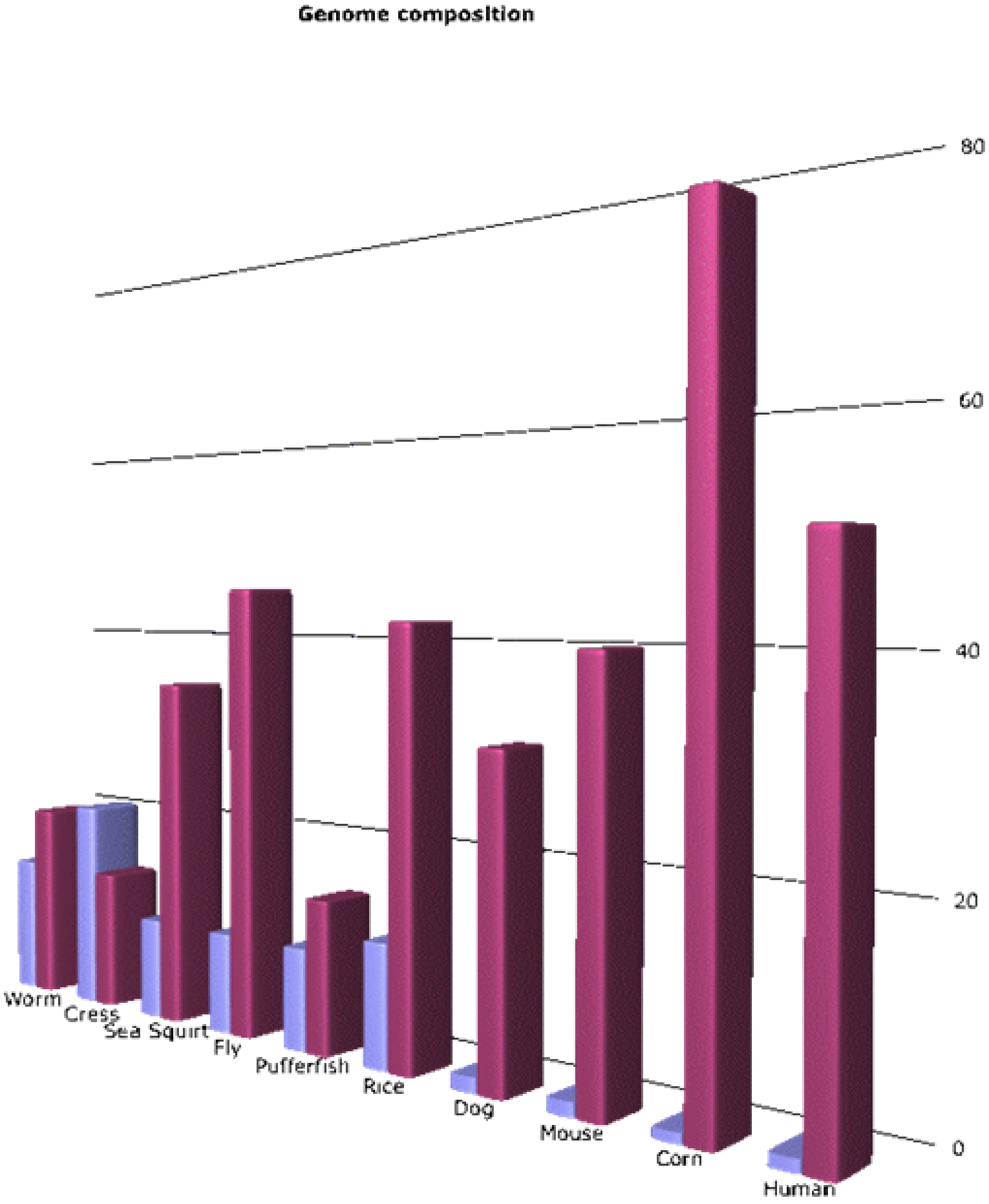
Singleton deletions throughout the genome increase risk of bipolar disorder

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*]

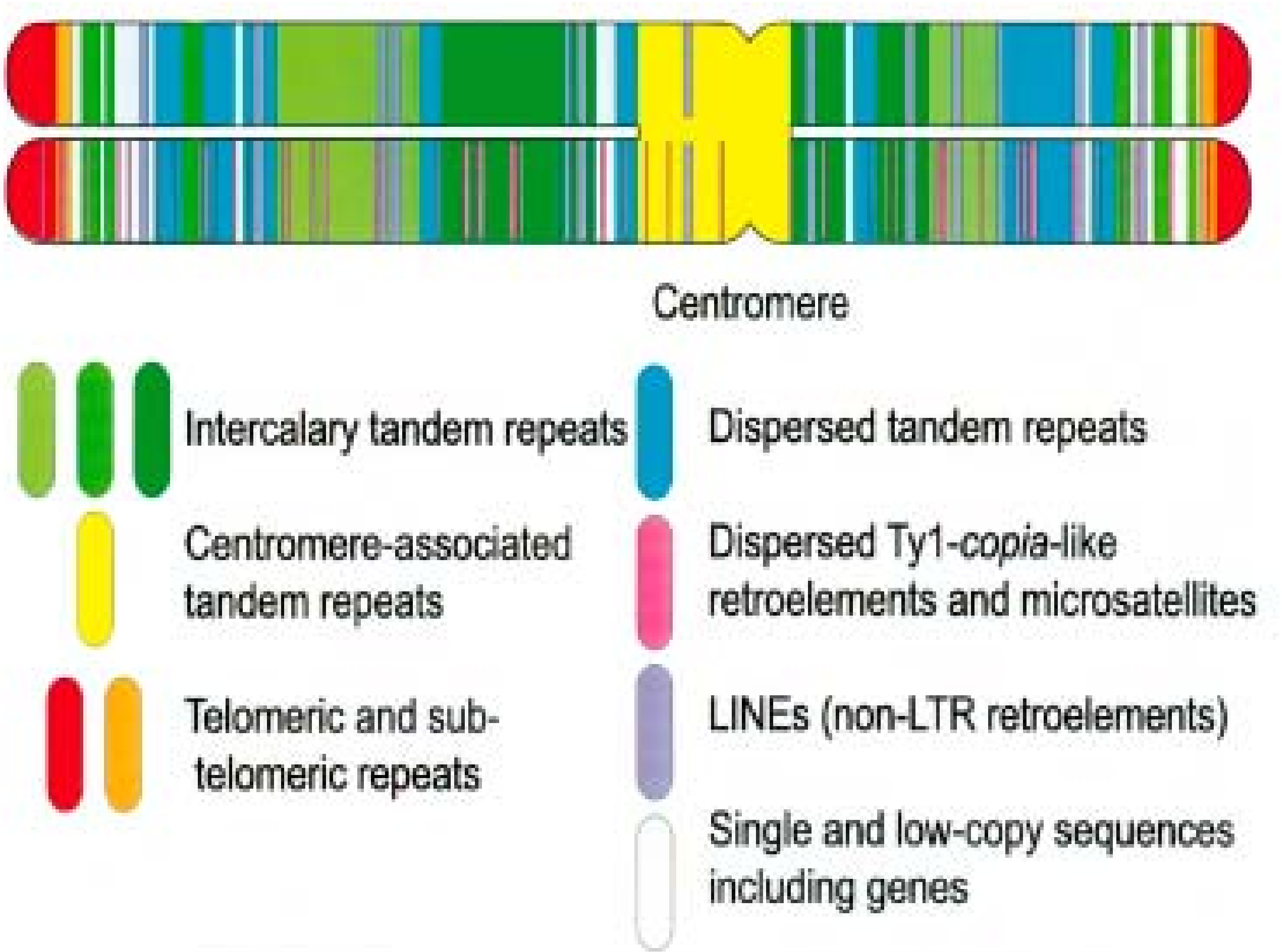
… we present a genome-wide copy number variant (CNV) survey of 1001 cases and 1034 controls ... Singleton deletions (deletions that appear only once in the dataset) more than 100 kb in length are present in 16.2% of BD cases and in 12.3% of controls (permutation P = 0.007).
Our results strongly suggest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

50% of human genome is repetitive DNA.

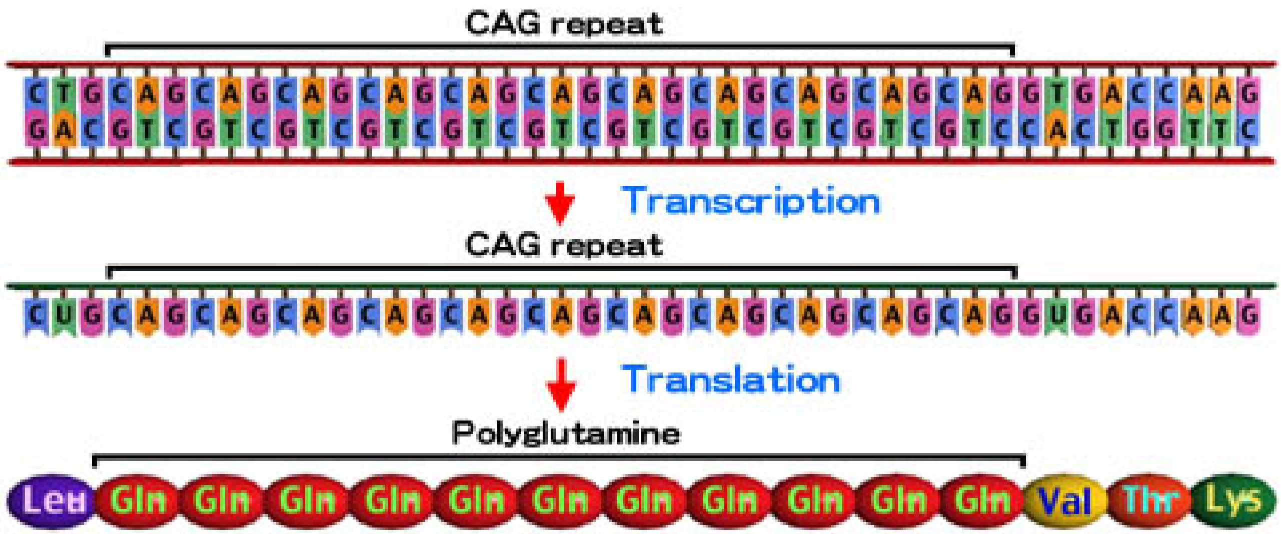
Only 1.2% is coding

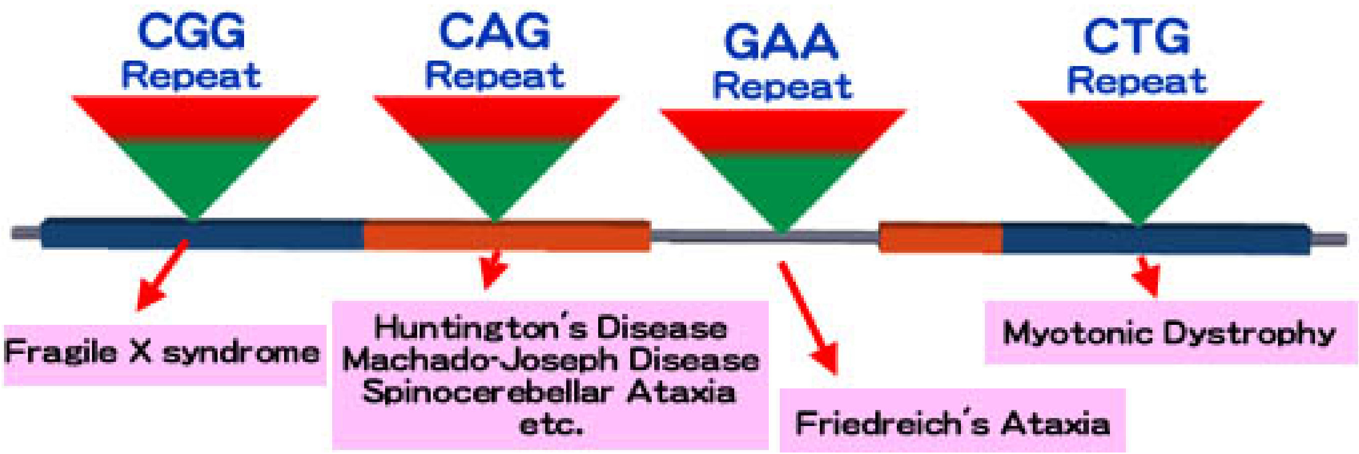


Types of repetitive elements and their chromosomal locations


Centromere
Intercalary tandem repeats
Centromere-associated tandem repeats
Telomeric and sub-telomeric repeats
Dispersed tandem repeats
Dispersed Ty1-copia-like retroelements and microsatellites
LINEs (non-LTR retroelements)
Single and low-copy sequences including genes

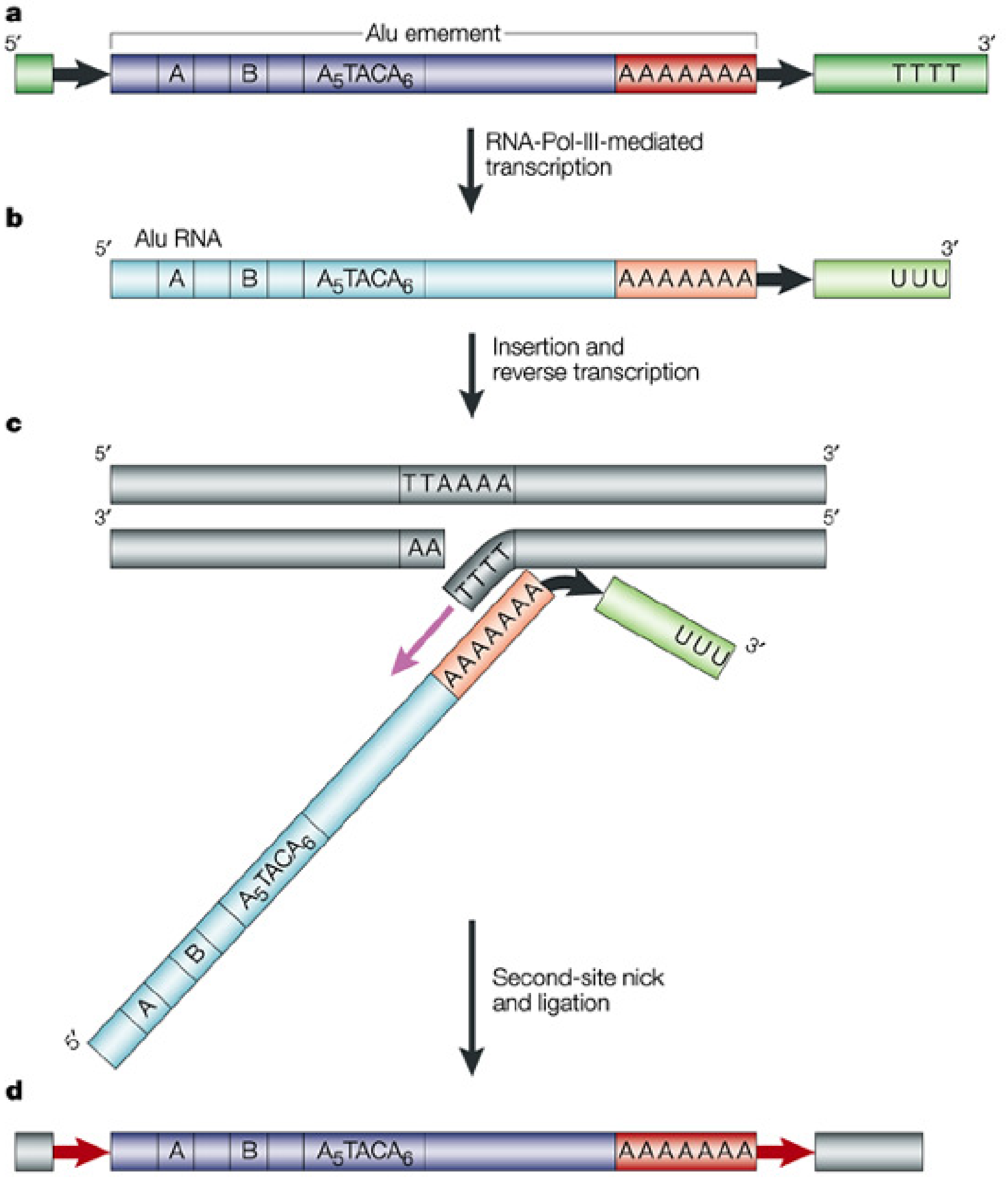
Triplet repeat diseases


CAG repeat
C T G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G G T G A C C A A G
G A C G T C G T C G T C G T C G T C G T C G T C G T C G T C G T C G T C C A C T G G T T C
Transcription
CAG repeat
C U G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G G U G A C C A A G
Translation
Polyglutamine
Leu
Gln
Gln
Gln
Gln
Gln
Gln
Gln
Gln
Gln
Gln
Gln
Val
Thr
Lys



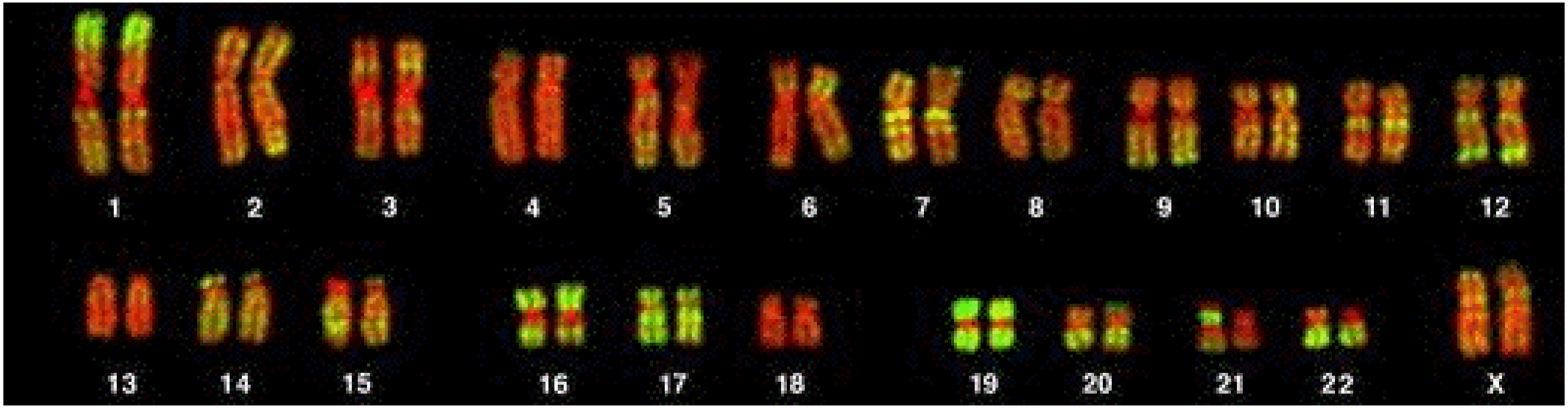
Alu elements

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)

Possible explanations for missing heritability

(in order of increasing plausibility ?)

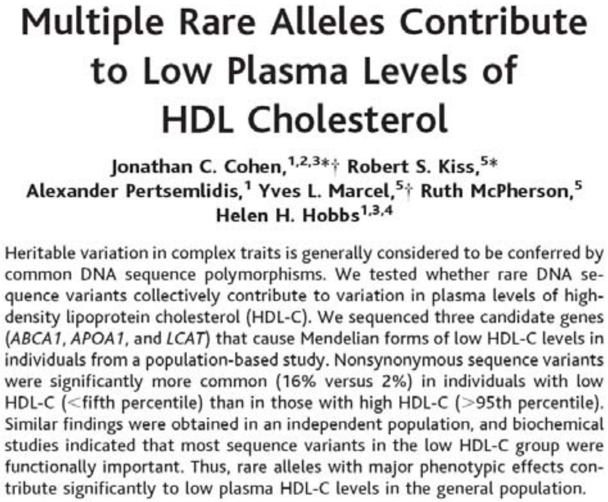
- Heritability estimates are wrong
- Nonadditivity of gene effects – epistasis, GxE
- Epigenetics – including parent-of-origin effects
- Low power for common small effects
- Disease heterogeneity – lots of different diseases with the same phenotype
- Poor tagging (1)
  - rare mutations of large effect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

Even for “simple” diseases the number of alleles is large

- Ischaemic heart disease (LDR) >190
- Breast cancer (BRCA1) >1000
- Colorectal cancer (MLN1) >140

Complex disease: common or rare alleles?

Increasing evidence for Common Disease – Rare Variant hypothesis (CDRV) A paradigm for future sequencing studies ?



Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol

Jonathan C. Cohen,[1,2,3]*† Robert S. Kiss,[5]*
Alexander Pertsemlidis,[1] Yves L. Marcel,[5]† Ruth McPherson,[5]
Helen H. Hobbs[1,3,4]

Heritable variation in complex traits is generally considered to be conferred by common DNA sequence polymorphisms. We tested whether rare DNA sequence variants collectively contribute to variation in plasma levels of high-density lipoprotein cholesterol (HDL-C). We sequenced three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) that cause Mendelian forms of low HDL-C levels in individuals from a population-based study. Nonsynonymous sequence variants were significantly more common (16% versus 2%) in individuals with low HDL-C (<fifth percentile) than in those with high HDL-C (>95th percentile). Similar findings were obtained in an independent population, and biochemical studies indicated that most sequence variants in the low HDL-C group were functionally important. Thus, rare alleles with major phenotypic effects contribute significantly to low plasma HDL-C levels in the general population.

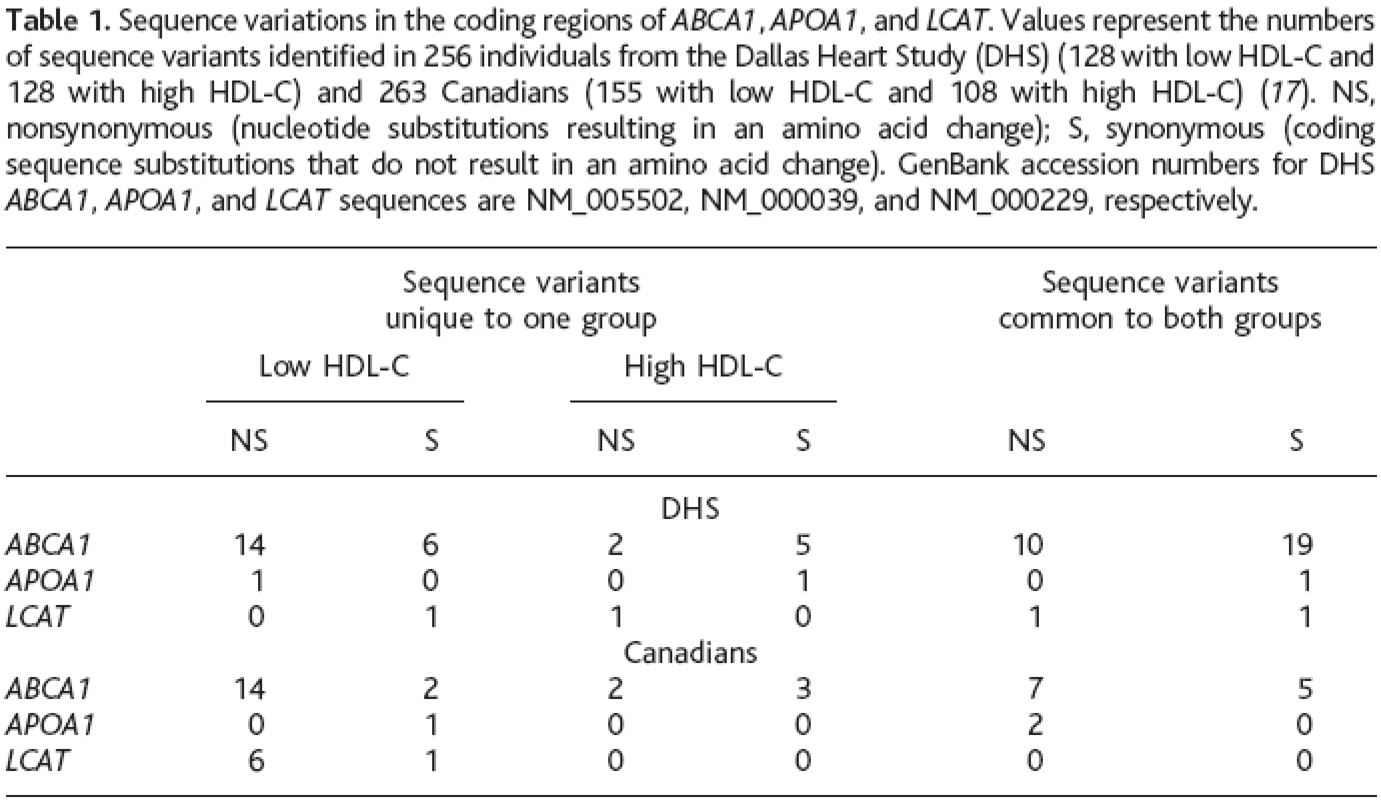
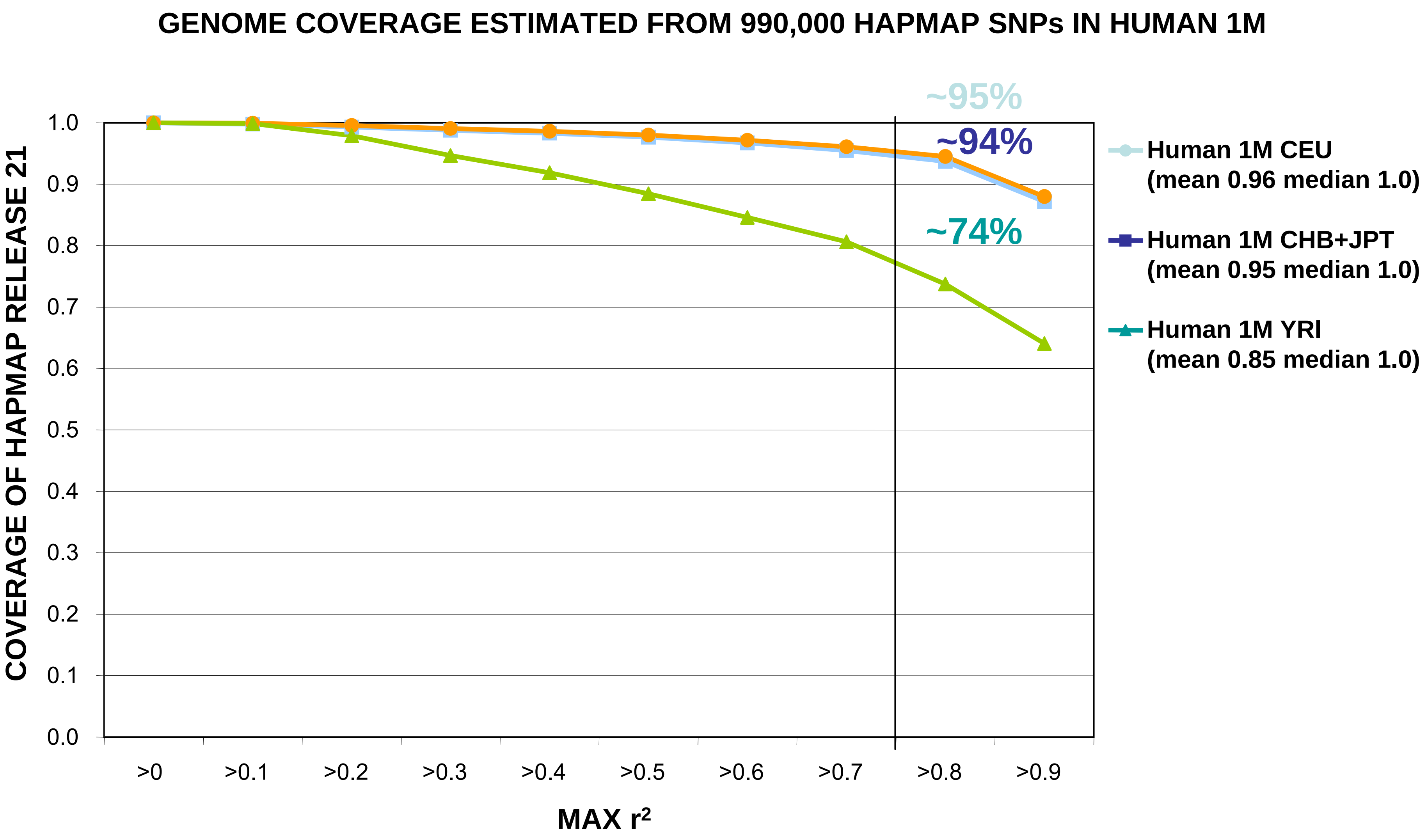


Table 1. Sequence variations in the coding regions of *ABCA1*, *APOA1*, and *LCAT*. Values represent the numbers of sequence variants identified in 256 individuals from the Dallas Heart Study (DHS) (128 with low HDL-C and 128 with high HDL-C) and 263 Canadians (155 with low HDL-C and 108 with high HDL-C) (*17*). NS, nonsynonymous (nucleotide substitutions resulting in an amino acid change); S, synonymous (coding sequence substitutions that do not result in an amino acid change). GenBank accession numbers for DHS *ABCA1*, *APOA1*, and *LCAT* sequences are NM_005502, NM_000039, and NM_000229, respectively.

	Sequence variants unique to one group				Sequence variants common to both groups	
	Low HDL-C		High HDL-C			
	NS	S	NS	S	NS	S
			DHS			
ABCA1	14	6	2	5	10	19
APOA1	1	0	0	1	0	1
LCAT	0	1	1	0	1	1
			Canadians			
ABCA1	14	2	2	3	7	5
APOA1	0	1	0	0	2	0
LCAT	6	1	0	0	0	0

[Science 2004]

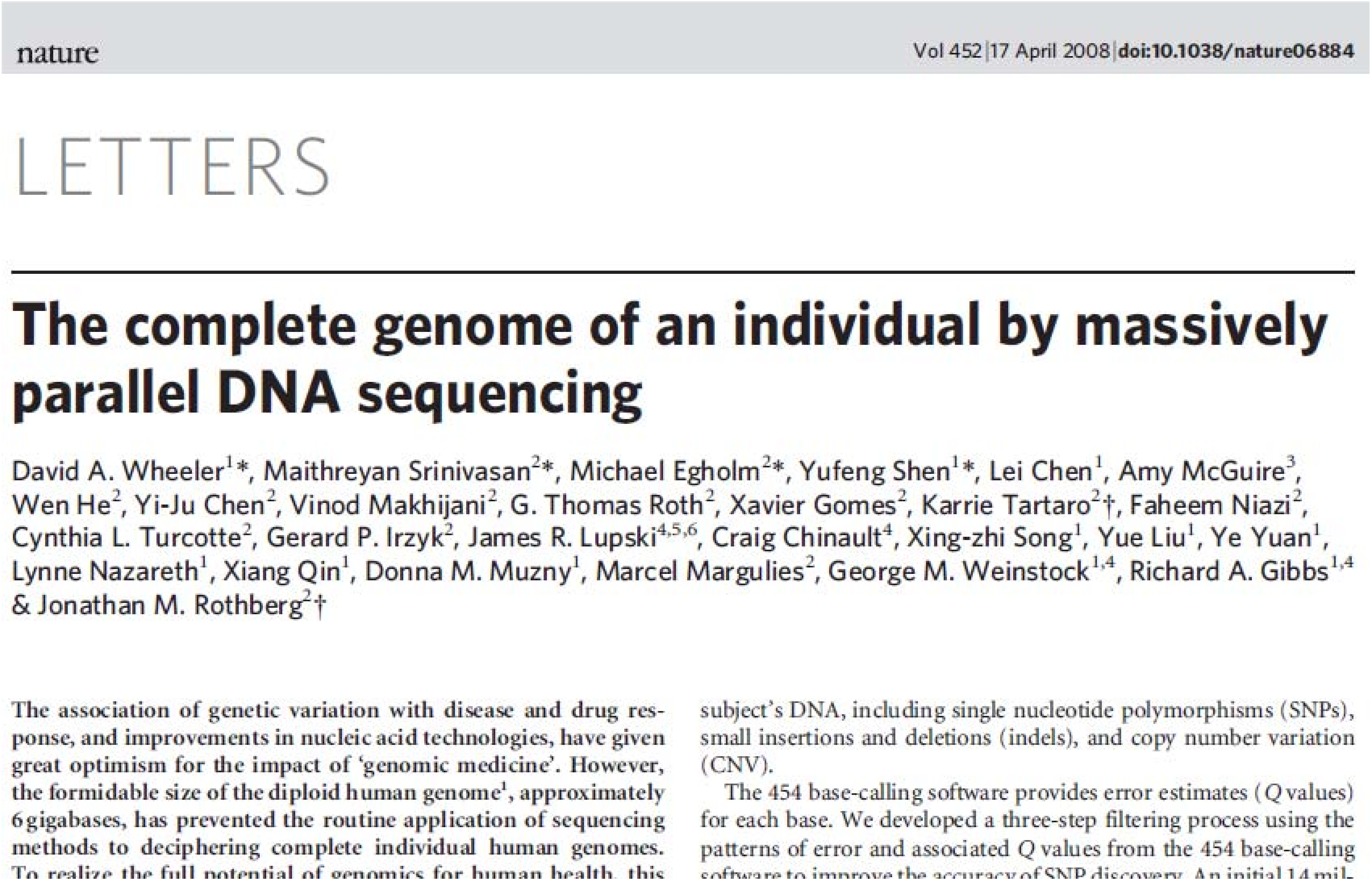
Human 1M HapMap Coverage by Population



The White House - June 26, 2000

Venter
Clinton
Collins

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



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

LETTERS

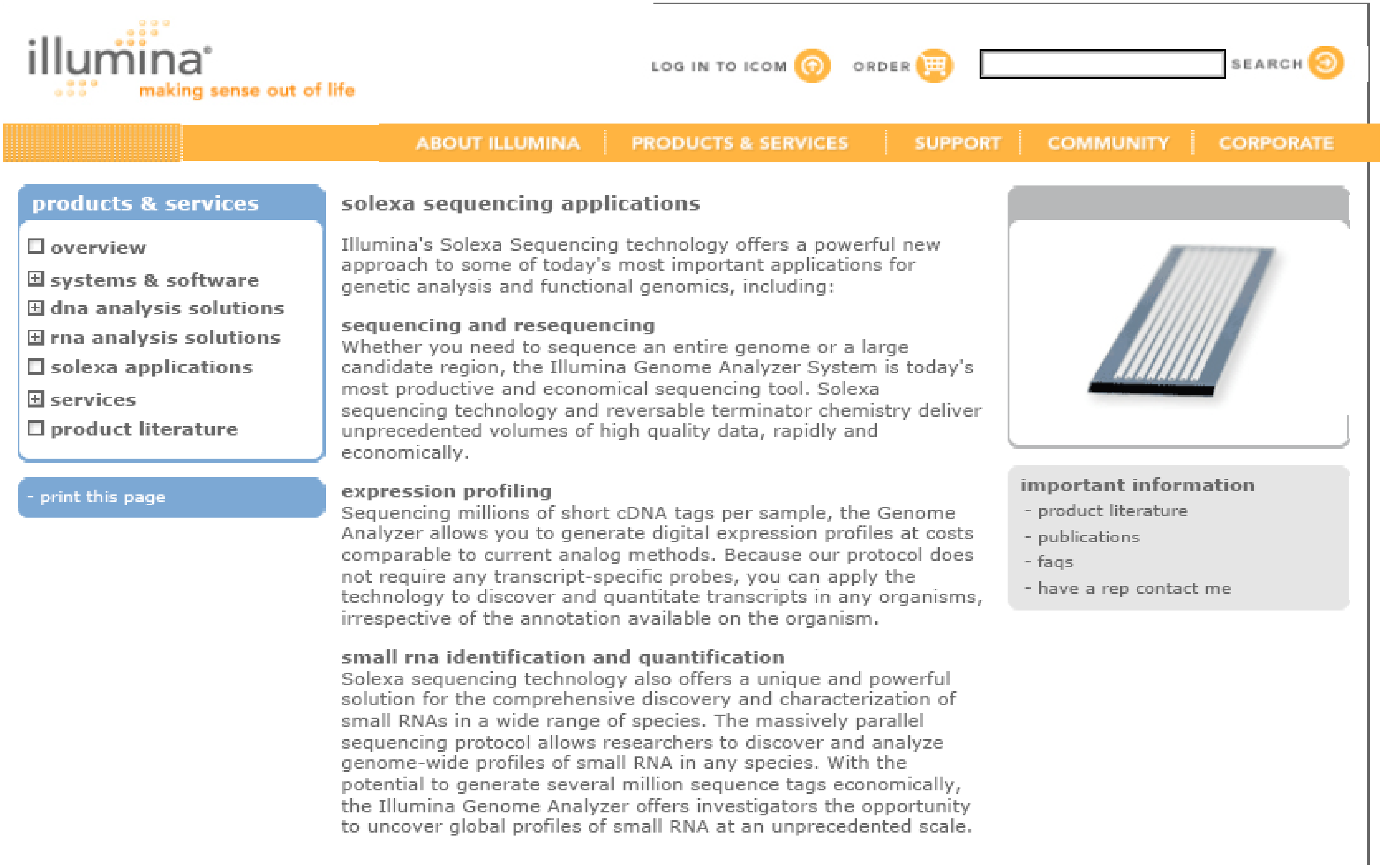
The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yufeng Shen[1]*, Lei Chen[1], Amy McGuire[3], Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier Gomes[2], Karrie Tartaro[2]†, Faheem Niazi[2], Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chinault[4], Xing-zhi Song[1], Yue Liu[1], Ye Yuan[1], Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2], George M. Weinstock[1,4], Richard A. Gibbs[1,4] & Jonathan M. Rothberg[2]†

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

subject's DNA, including single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), and copy number variation (CNV).

The 454 base-calling software provides error estimates (*Q* values) for each base. We developed a three-step filtering process using the patterns of error and associated *Q* values from the 454 base-calling software to improve the accuracy of SNP discovery. An initial 1.4 mil-



solexa sequencing applications

Illumina's Solexa Sequencing technology offers a powerful new approach to some of today's most important applications for genetic analysis and functional genomics, including:

sequencing and resequencing

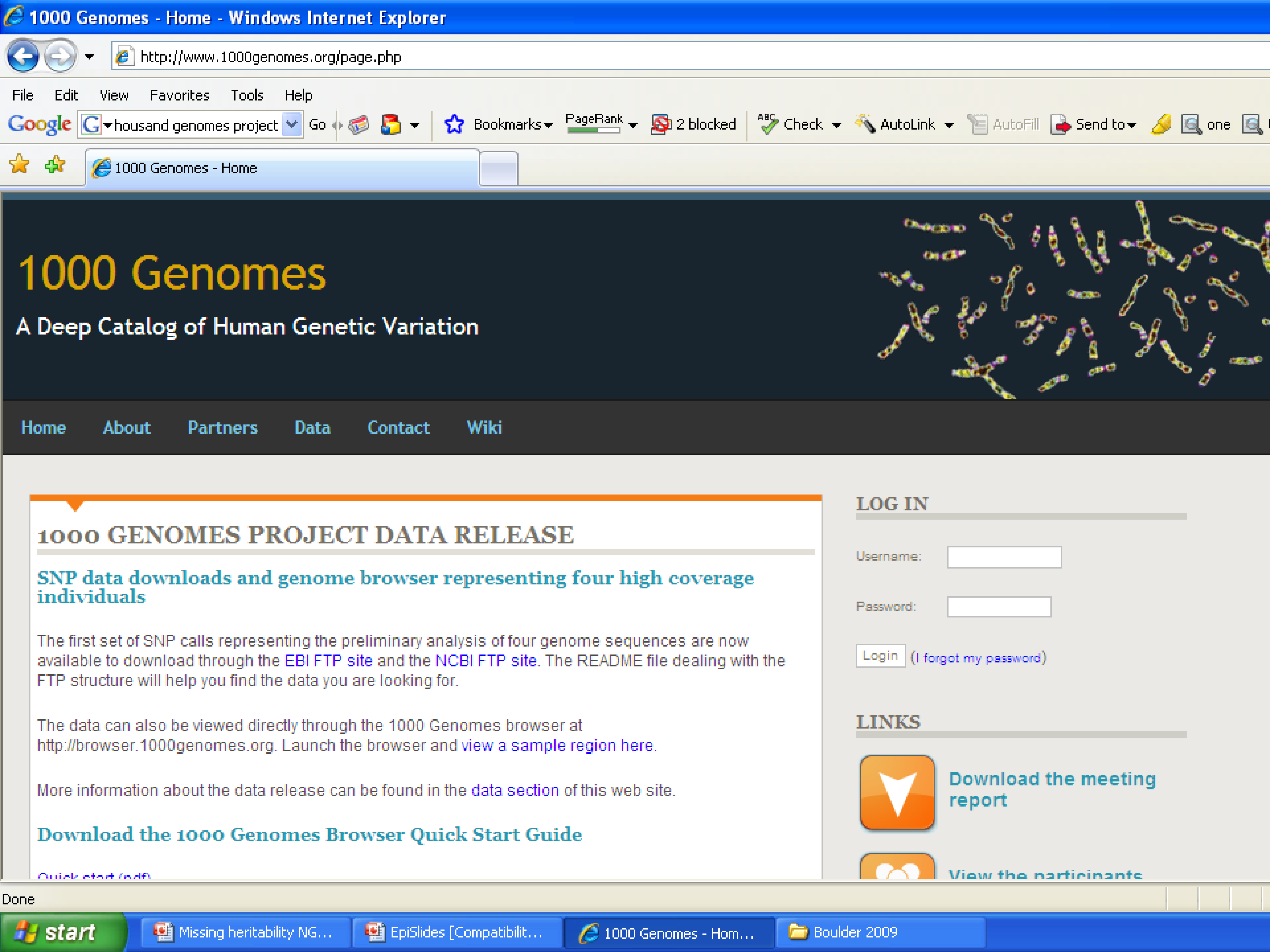
Whether you need to sequence an entire genome or a large candidate region, the Illumina Genome Analyzer System is today's most productive and economical sequencing tool. Solexa sequencing technology and reversable terminator chemistry deliver unprecedented volumes of high quality data, rapidly and economically.

expression profiling

Sequencing millions of short cDNA tags per sample, the Genome Analyzer allows you to generate digital expression profiles at costs comparable to current analog methods. Because our protocol does not require any transcript-specific probes, you can apply the technology to discover and quantitate transcripts in any organisms, irrespective of the annotation available on the organism.

small rna identification and quantification

Solexa sequencing technology also offers a unique and powerful solution for the comprehensive discovery and characterization of small RNAs in a wide range of species. The massively parallel sequencing protocol allows researchers to discover and analyze genome-wide profiles of small RNA in any species. With the potential to generate several million sequence tags economically, the Illumina Genome Analyzer offers investigators the opportunity to uncover global profiles of small RNA at an unprecedented scale.



1000 Genomes

A Deep Catalog of Human Genetic Variation

Home About Partners Data Contact Wiki

1000 GENOMES PROJECT DATA RELEASE

SNP data downloads and genome browser representing four high coverage individuals

The first set of SNP calls representing the preliminary analysis of four genome sequences are now available to download through the EBI FTP site and the NCBI FTP site. The README file dealing with the FTP structure will help you find the data you are looking for.

The data can also be viewed directly through the 1000 Genomes browser at http://browser.1000genomes.org. Launch the browser and view a sample region here.

More information about the data release can be found in the data section of this web site.

Download the 1000 Genomes Browser Quick Start Guide

Quick start (pdf)

LOG IN

Username:

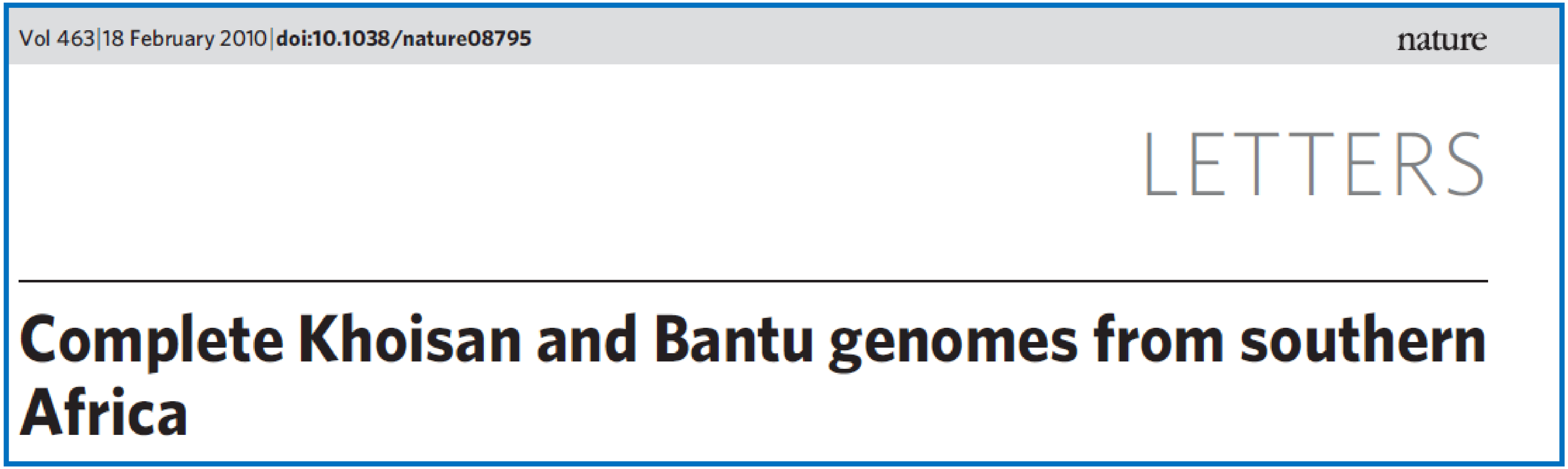
Password:

Login (I forgot my password)

LINKS

Download the meeting report

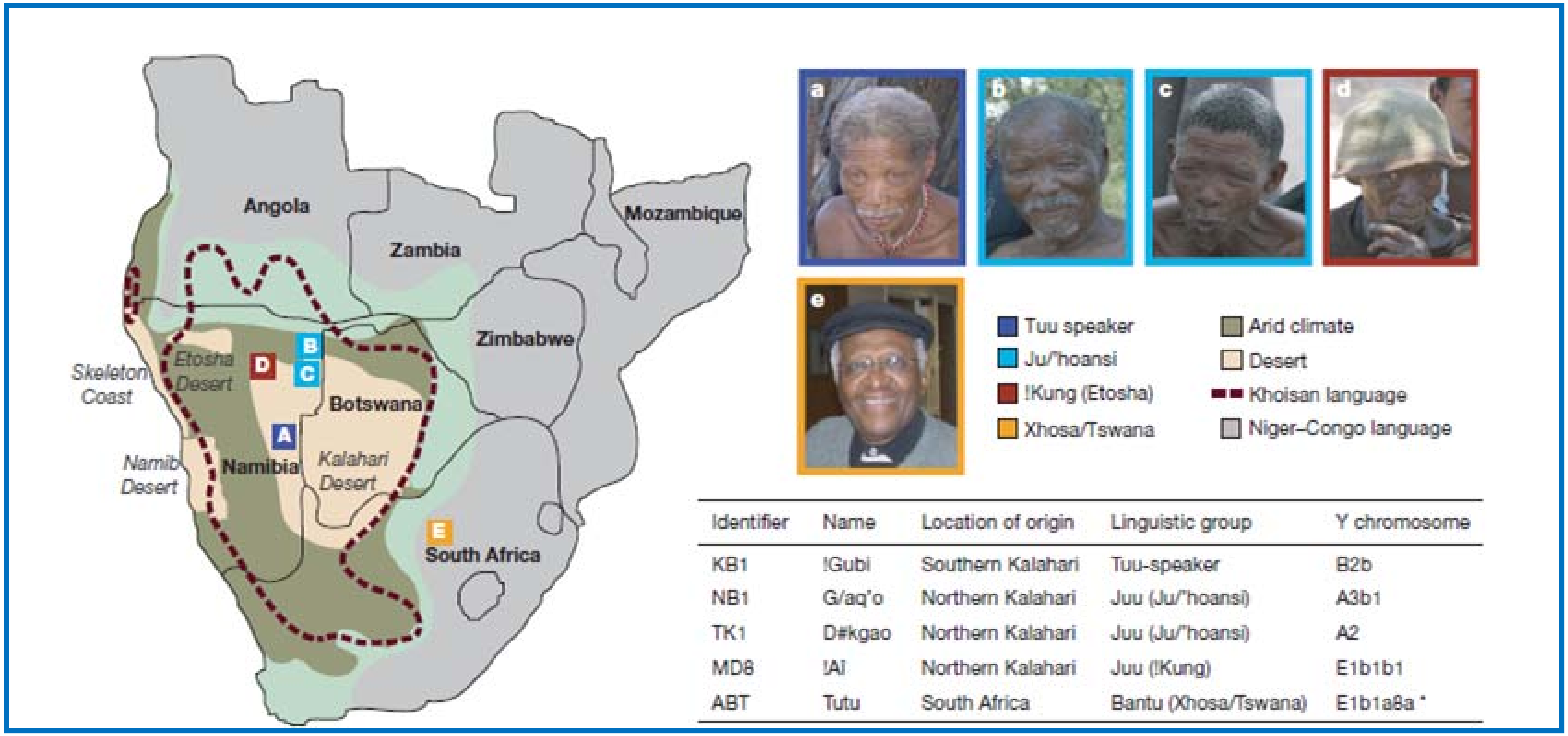
View the participants



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

LETTERS

Complete Khoisan and Bantu genomes from southern Africa



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

The genomes of Archbishop Tutu and one bushman were fully sequenced, and the other three partially (exones).

The bushmen were found to lack genes for digesting milk and malaria resistance, but most had genes linked to enhanced physical prowess. One had a gene linked to increased retention of salt and water, an advantage for a desert dweller.

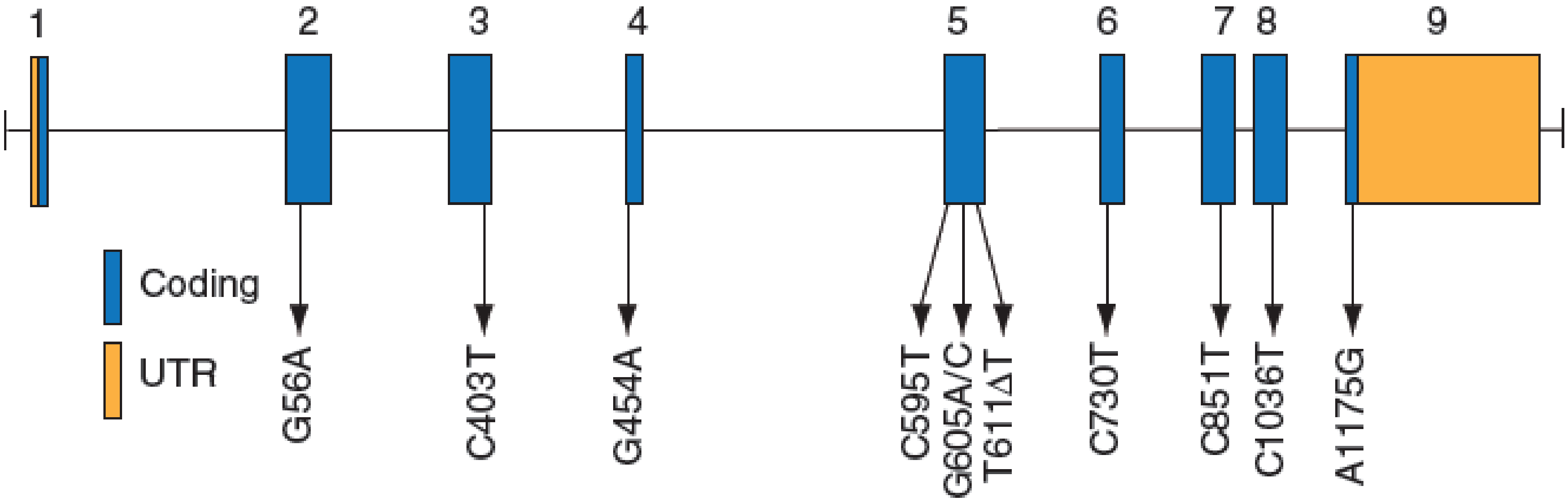
On average there are more genetic differences between any two bushmen in the study than between a European and an Asian

.

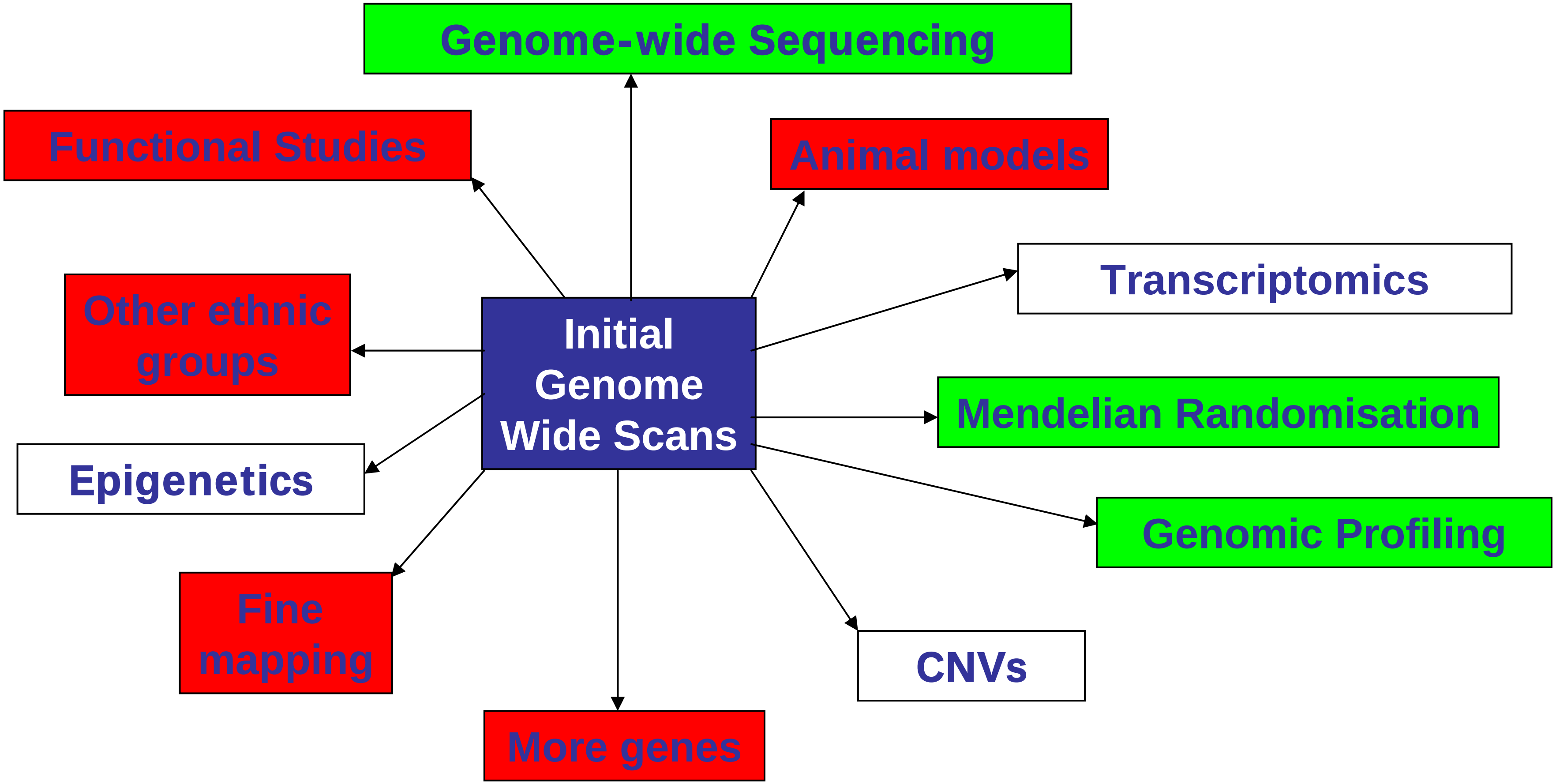
Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[2,10], Choli Lee[1], Abigail W Bigham[2], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

We demonstrate the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome (MIM%263750). For four affected individuals in three independent kindreds, we captured and sequenced coding regions to a mean coverage of 40× and sufficient depth to call variants at ~97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the four individuals identified a single candidate gene, *DHODH*, which encodes a key enzyme in the pyrimidine *de novo* biosynthesis pathway. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome. Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes underlying rare mendelian disorders and will likely transform the genetic analysis of monogenic traits.



What next?


Initial Genome Wide Scans

Genome-wide Sequencing

Functional Studies

Animal models

Transcriptomics

Other ethnic groups

Mendelian Randomisation

Epigenetics

Genomic Profiling

Fine mapping

CNVs

More genes

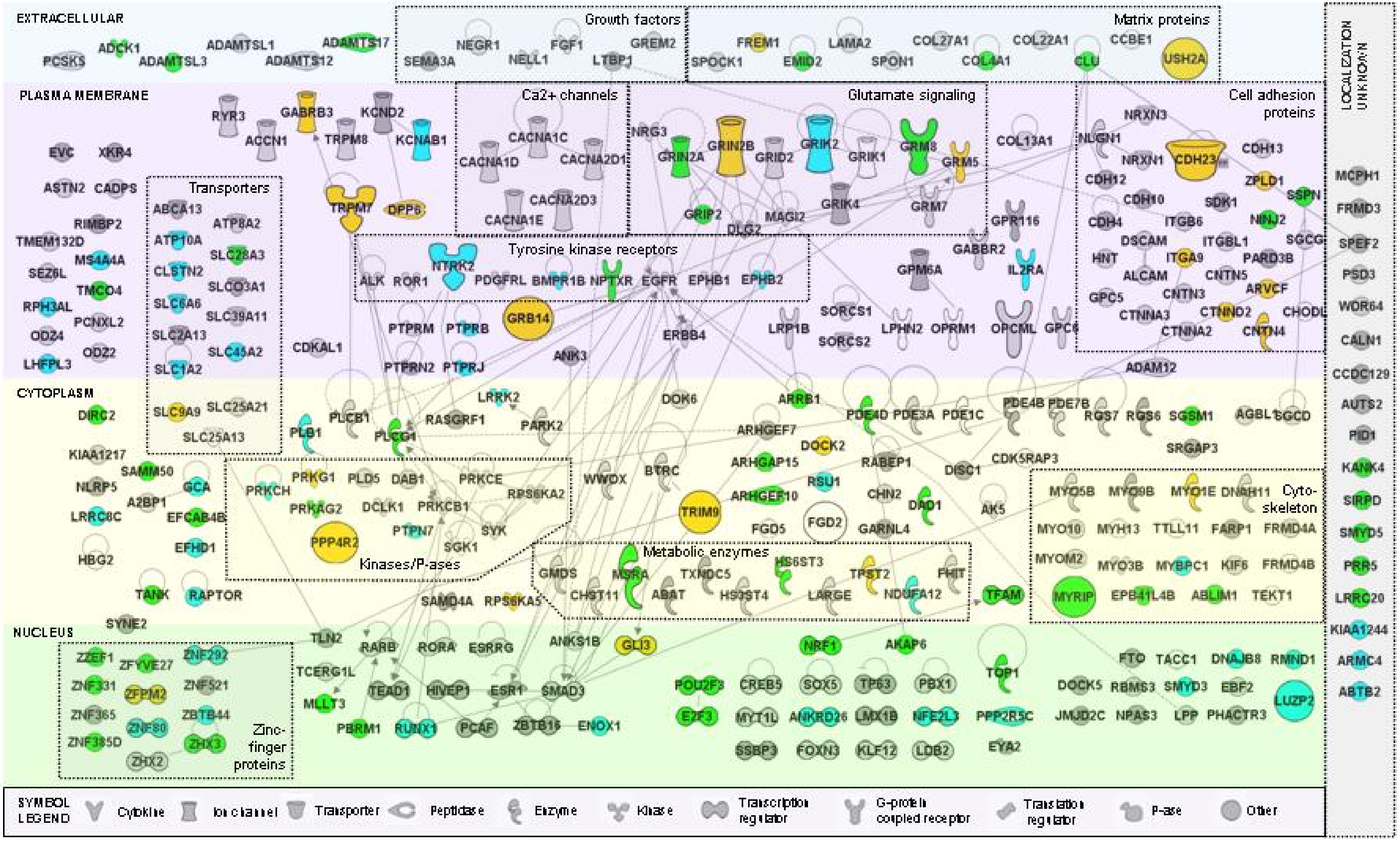
David Evans

Evaluating combined effects of genes

- Select genes that are biologically ‘related’. i.e. they share a pathway or common function

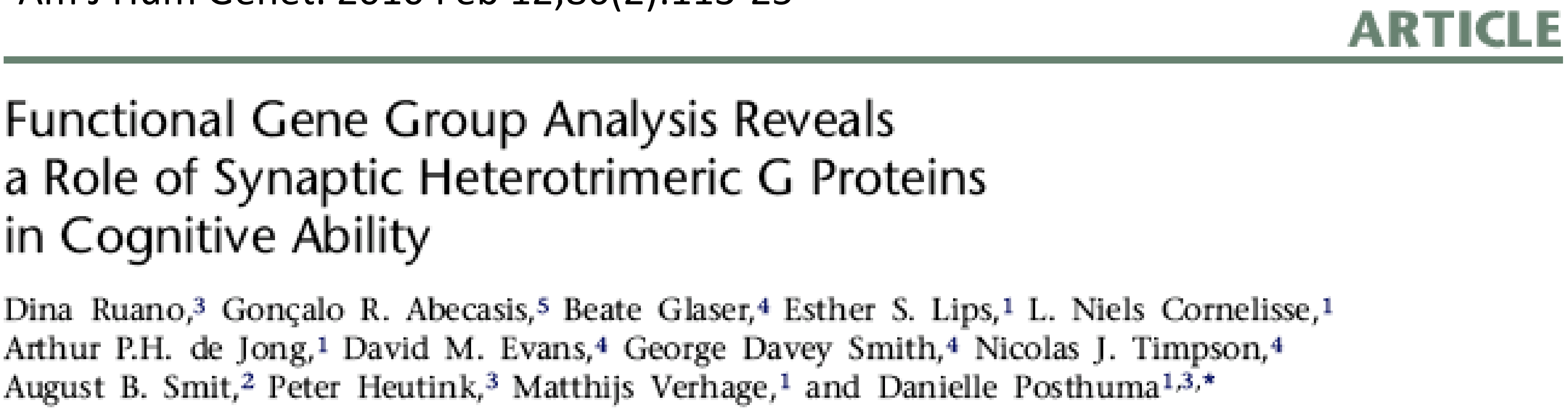
- Networks of genes underlying biological pathways are more likely to be the crucial unit of functioning in the biological system than single SNPs or genes

Pathway (Ingenuity) analysis of GWAS for smoking



Vink et al. 2009 *AJHG* in pres

Am J Hum Genet. 2010 Feb 12;86(2):113-25

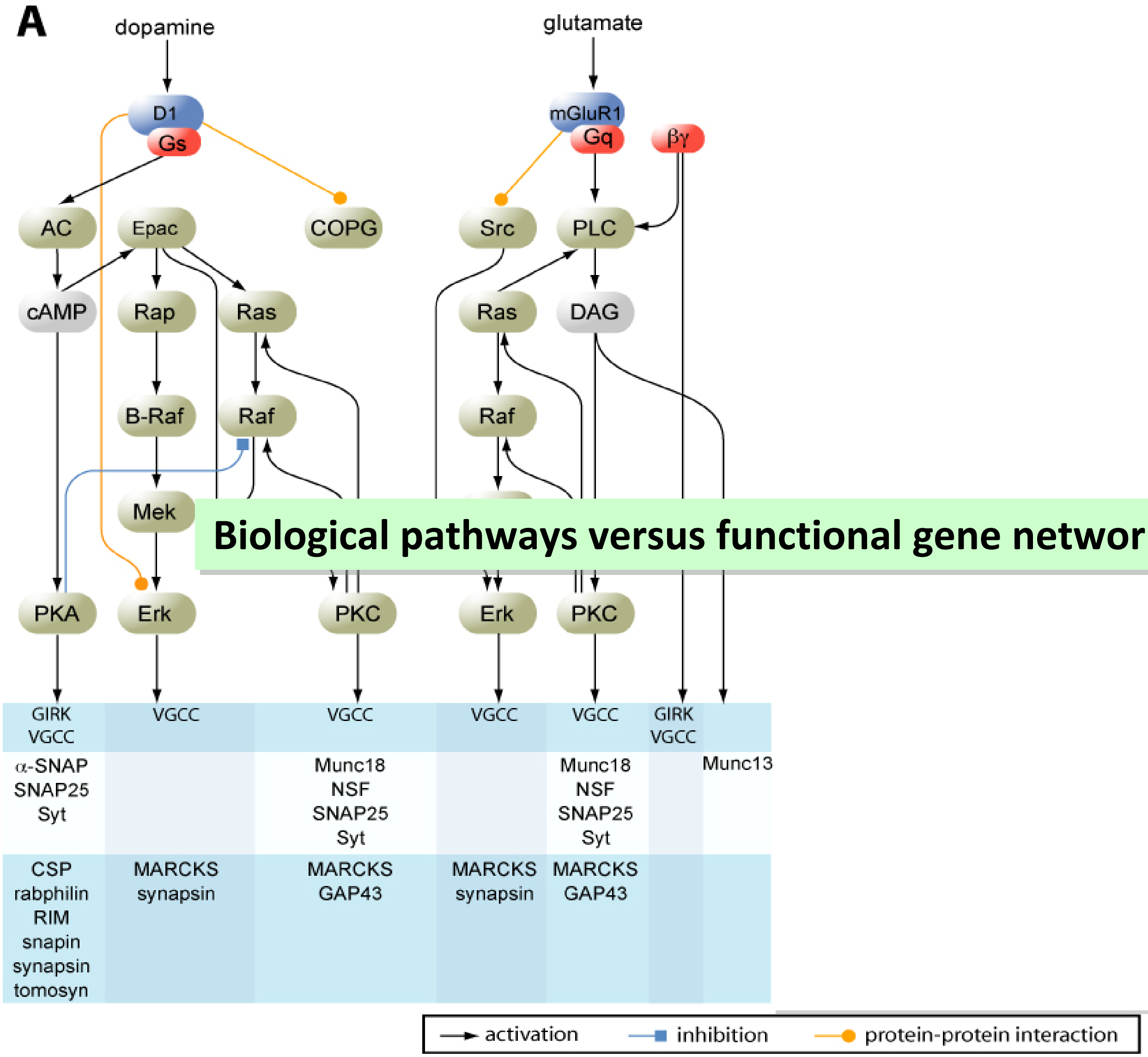


ARTICLE

Functional Gene Group Analysis Reveals a Role of Synaptic Heterotrimeric G Proteins in Cognitive Ability

Dina Ruano,[3] Gonçalo R. Abecasis,[5] Beate Glaser,[4] Esther S. Lips,[1] L. Niels Cornelisse,[1] Arthur P.H. de Jong,[1] David M. Evans,[4] George Davey Smith,[4] Nicolas J. Timpson,[4] August B. Smit,[2] Peter Heutink,[3] Matthijs Verhage,[1] and Danielle Posthuma[1,3,*]

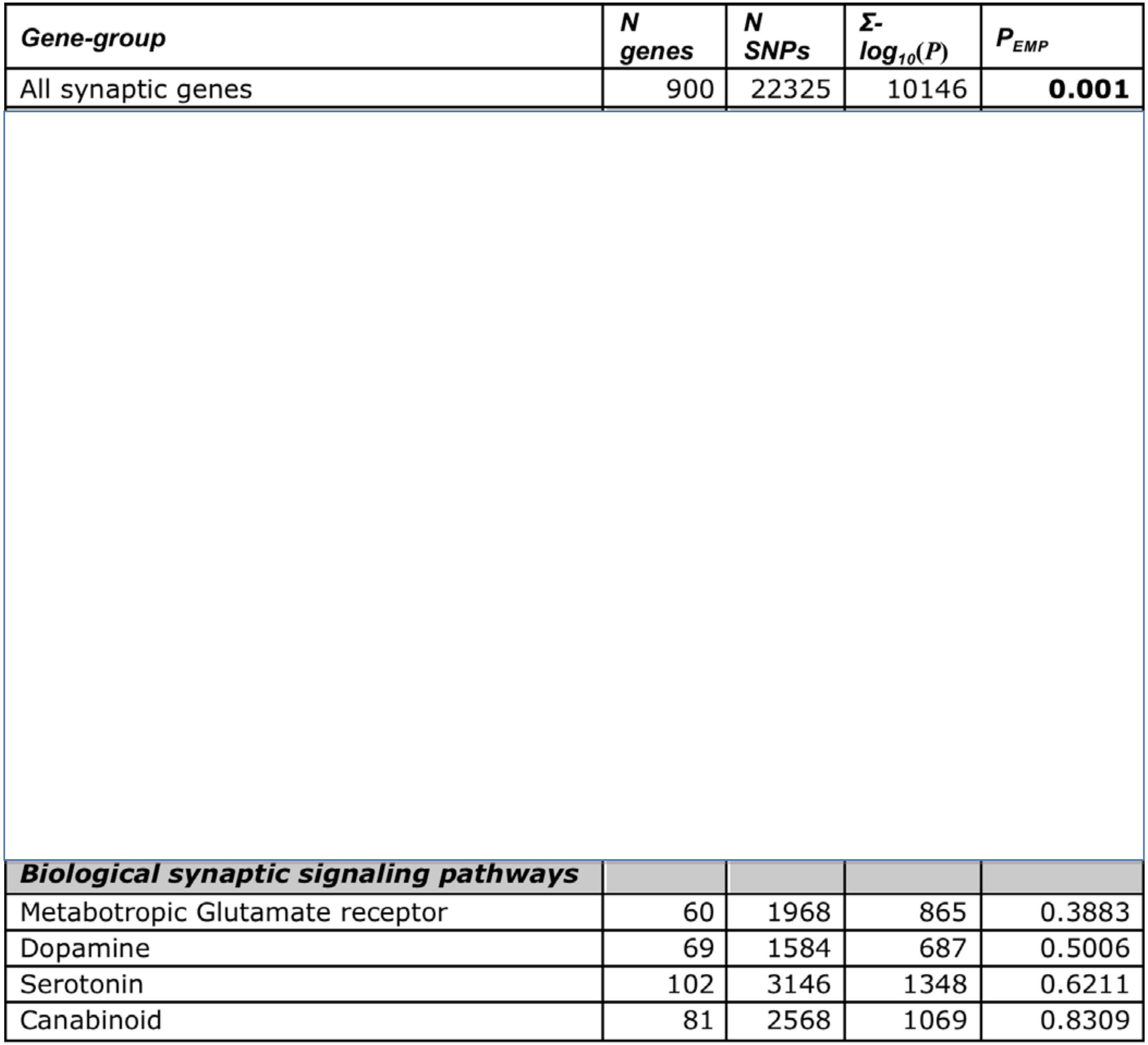
Vertical vs. Horizontal Grouping



Biological pathways versus functional gene networks

Ruano et al., 2010 AJHG

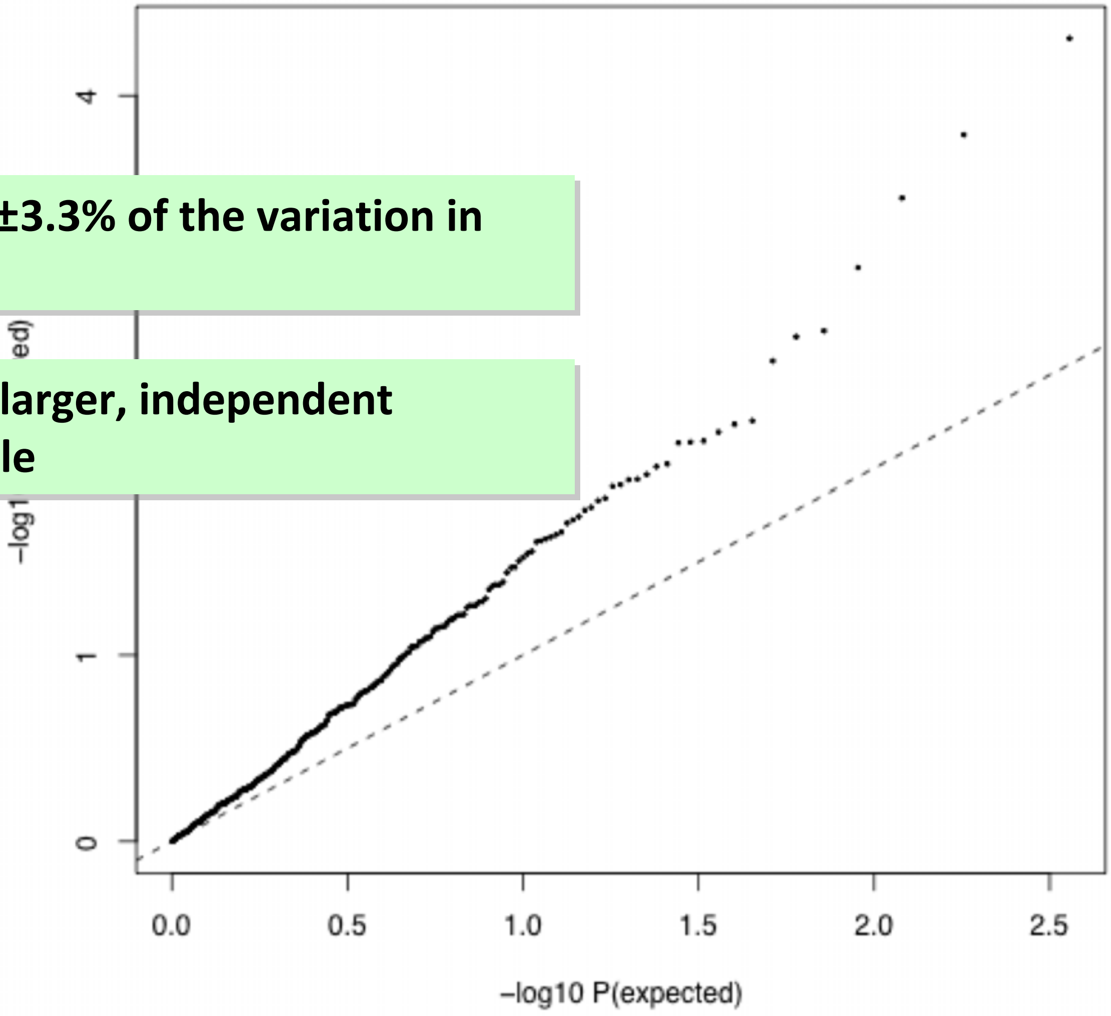
Functional gene networks for intelligence



Gene-group	*N genes*	*N SNPs*	*Σ-log$_{10}$(P)*	*P_{EMP}*
All synaptic genes	900	22325	10146	**0.001**
Biological synaptic signaling pathways				
Metabotropic Glutamate receptor	60	1968	865	0.3883
Dopamine	69	1584	687	0.5006
Serotonin	102	3146	1348	0.6211
Canabinoid	81	2568	1069	0.8309

Ruano et al,AJHG 2010.

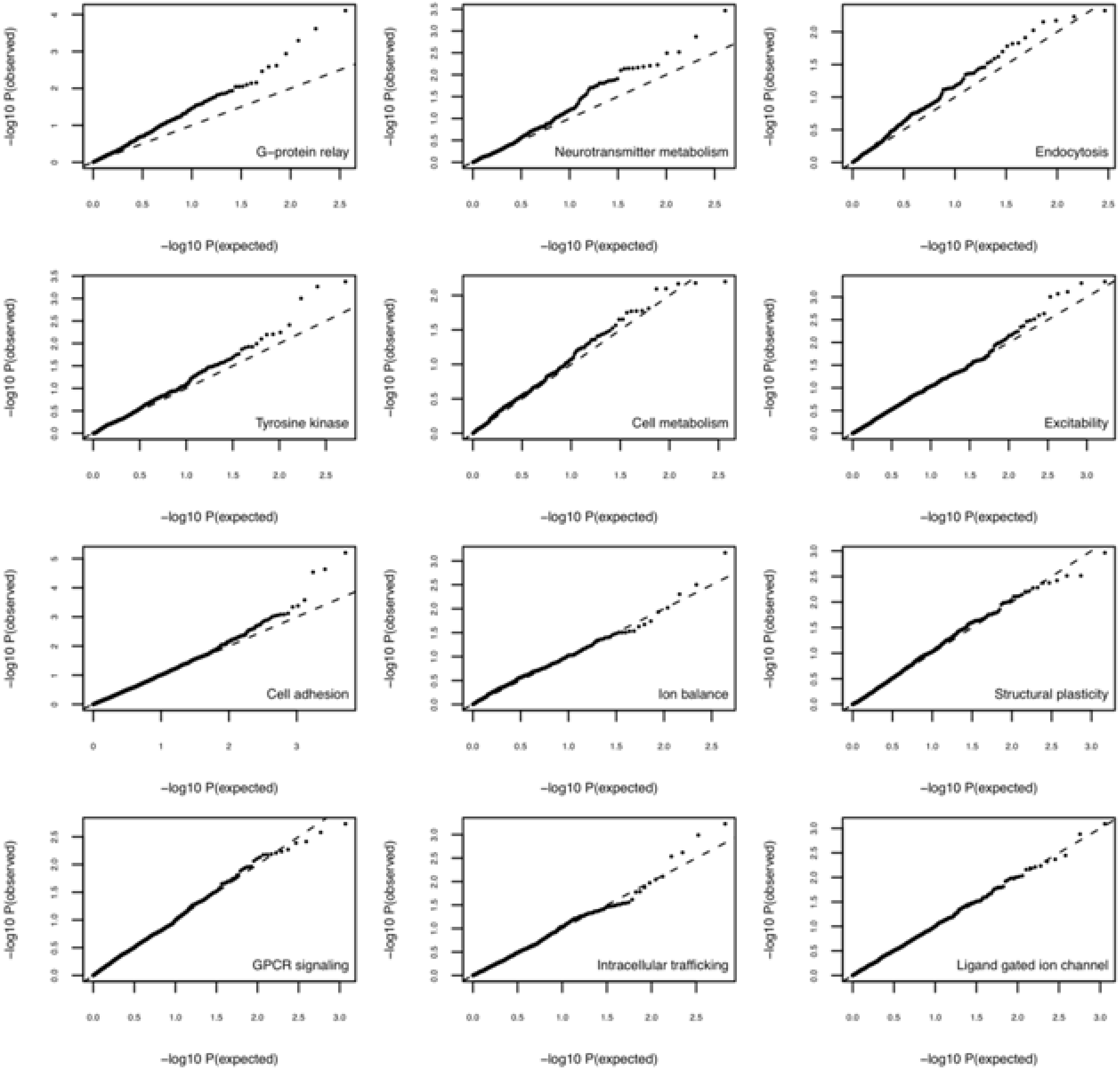
‘QQ-plot’ of p-values of genetic variants in heterotrimeric G proteins



Accounts for ±3.3% of the variation in intelligence

Replicated in larger, independent ALSPAC sample

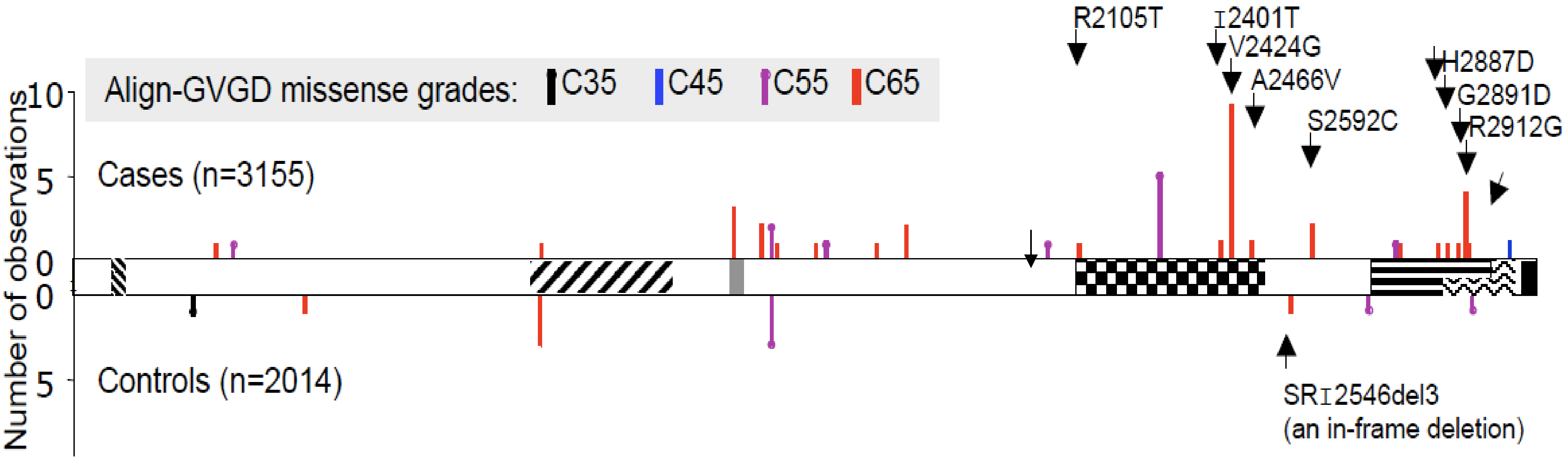
A. FUNCTIONAL GENE GROUPS



Once we have all the rare sequence variants, how do we decide if they are causal / harmful ?

- Too rare to use standard Ca-Co statistical tests
- Can group variants (but heterogeneous?)
- Use DNA/protein functional analysis
- Use evolutionary criteria (sequence conservation across species)

Domain organization of ATM and case-control distribution of rare missense substitutions



▧ p53 & BRCA1 interaction (91-97)

▨ beta-adaptin interaction (950-1233)

■ c-Abl interaction (1373-1382)

↓ Serine 1981

▩ FAT 2096-2491 (PFAM 02259)

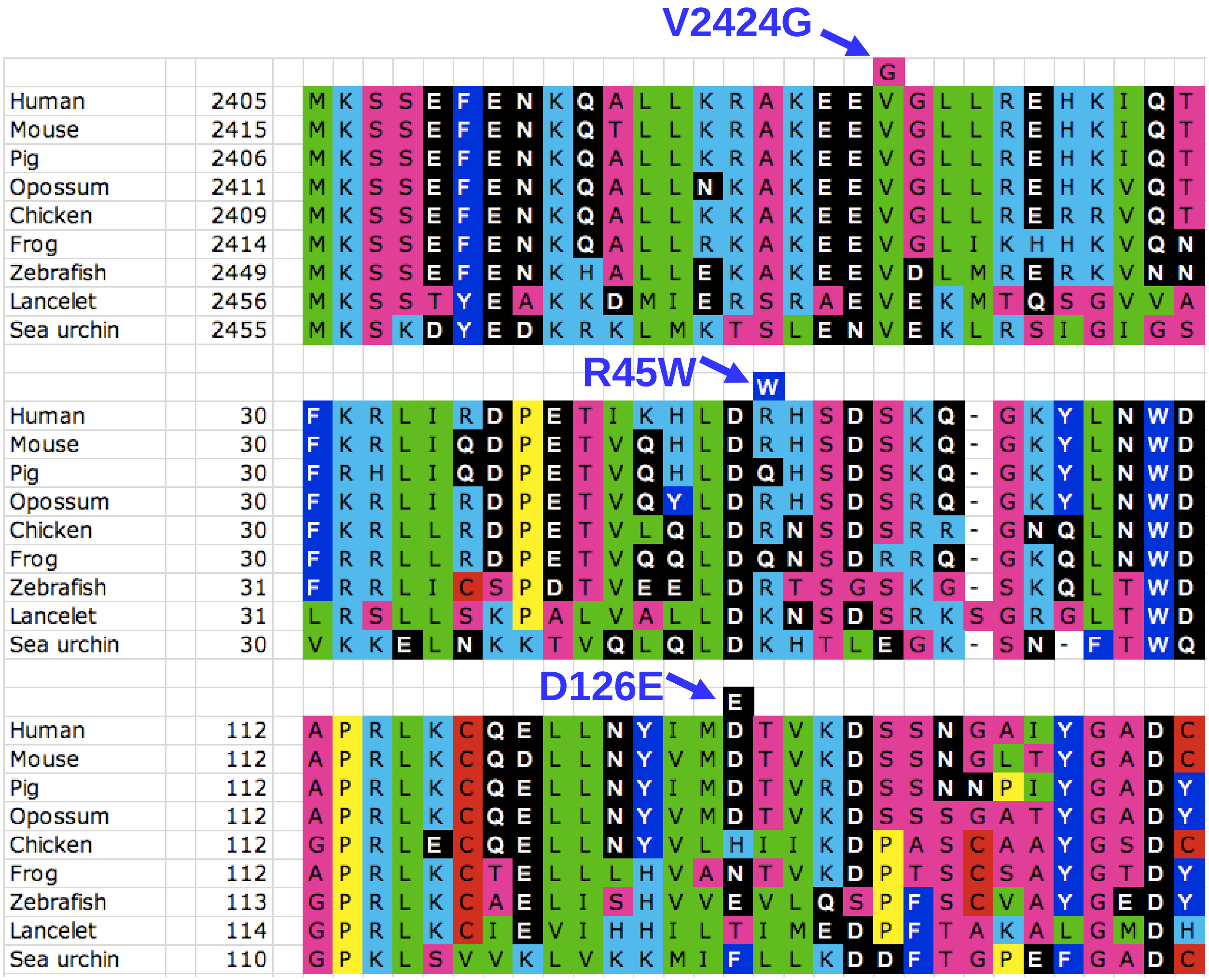
☰ PI3-PI4 kinase 2711-2962 (PFAM 00454)

≋ p53 interaction (2862-3012)

■ FATC 3024-3056 (PFAM 02260)

Am J Hum Genet. 2009 Oct;85(4):427-46.
Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer.
Tavtigian SV, ...Chenevix-Trench G.

Using species comparisons to decide if a mutation is harmful



V2424G																					G										
Human	2405	M	K	S	S	E	F	E	N	K	Q	A	L	L	K	R	A	K	E	E	V	G	L	L	R	E	H	K	I	Q	T
Mouse	2415	M	K	S	S	E	F	E	N	K	Q	T	L	L	K	R	A	K	E	E	V	G	L	L	R	E	H	K	I	Q	T
Pig	2406	M	K	S	S	E	F	E	N	K	Q	A	L	L	K	R	A	K	E	E	V	G	L	L	R	E	H	K	I	Q	T
Opossum	2411	M	K	S	S	E	F	E	N	K	Q	A	L	L	N	K	A	K	E	E	V	G	L	L	R	E	H	K	V	Q	T
Chicken	2409	M	K	S	S	E	F	E	N	K	Q	A	L	L	K	K	A	K	E	E	V	G	L	L	R	E	R	R	V	Q	T
Frog	2414	M	K	S	S	E	F	E	N	K	Q	A	L	L	R	K	A	K	E	E	V	G	L	I	K	H	H	K	V	Q	N
Zebrafish	2449	M	K	S	S	E	F	E	N	K	H	A	L	L	E	K	A	K	E	E	V	D	L	M	R	E	R	K	V	N	N
Lancelet	2456	M	K	S	S	T	Y	E	A	K	K	D	M	I	E	R	S	R	A	E	V	E	K	M	T	Q	S	G	V	V	A
Sea urchin	2455	M	K	S	K	D	Y	E	D	K	R	K	L	M	K	T	S	L	E	N	V	E	K	L	R	S	I	G	I	G	S

V2424G
GV = 0.0
GD = 109.6
Grade: C65
Highest probability to be pathogenic

R45W																	W														
Human	30	F	K	R	L	I	R	D	P	E	T	I	K	H	L	D	R	H	S	D	S	K	Q	-	G	K	Y	L	N	W	D
Mouse	30	F	K	R	L	I	Q	D	P	E	T	V	Q	H	L	D	R	H	S	D	S	K	Q	-	G	K	Y	L	N	W	D
Pig	30	F	R	H	L	I	Q	D	P	E	T	V	Q	H	L	D	Q	H	S	D	S	K	Q	-	G	K	Y	L	N	W	D
Opossum	30	F	K	R	L	I	R	D	P	E	T	V	Q	Y	L	D	R	H	S	D	S	R	Q	-	G	K	Y	L	N	W	D
Chicken	30	F	K	R	L	L	R	D	P	E	T	V	L	Q	L	D	R	N	S	D	S	R	R	-	G	N	Q	L	N	W	D
Frog	30	F	R	R	L	L	R	D	P	E	T	V	Q	Q	L	D	Q	N	S	D	R	R	Q	-	G	K	Q	L	N	W	D
Zebrafish	31	F	R	R	L	I	C	S	P	D	T	V	E	E	L	D	R	T	S	G	S	K	G	-	S	K	Q	L	T	W	D
Lancelet	31	L	R	S	L	L	S	K	P	A	L	V	A	L	L	D	K	N	S	D	S	R	K	S	G	R	G	L	T	W	D
Sea urchin	30	V	K	K	E	L	N	K	K	T	V	Q	L	Q	L	D	K	H	T	L	E	G	K	-	S	N	-	F	T	W	Q

R45W
GV = 56.6
GD = 95.8
Grade: C25
Intermediate probability to be pathogenic

D126E																E															
Human	112	A	P	R	L	K	C	Q	E	L	L	N	Y	I	M	D	T	V	K	D	S	S	N	G	A	I	Y	G	A	D	C
Mouse	112	A	P	R	L	K	C	Q	D	L	L	N	Y	V	M	D	T	V	K	D	S	S	N	G	L	T	Y	G	A	D	C
Pig	112	A	P	R	L	K	C	Q	E	L	L	N	Y	I	M	D	T	V	R	D	S	S	N	N	P	I	Y	G	A	D	Y
Opossum	112	A	P	R	L	K	C	Q	E	L	L	N	Y	V	M	D	T	V	K	D	S	S	S	G	A	T	Y	G	A	D	Y
Chicken	112	G	P	R	L	E	C	Q	E	L	L	N	Y	V	L	H	I	I	K	D	P	A	S	C	A	A	Y	G	S	D	C
Frog	112	A	P	R	L	K	C	T	E	L	L	L	H	V	A	N	T	V	K	D	P	T	S	C	S	A	Y	G	T	D	Y
Zebrafish	113	G	P	R	L	K	C	A	E	L	I	S	H	V	V	E	V	L	Q	S	P	F	S	C	V	A	Y	G	E	D	Y
Lancelet	114	G	P	R	L	K	C	I	E	V	I	H	H	I	L	T	I	M	E	D	P	F	T	A	K	A	L	G	M	D	H
Sea urchin	110	G	P	K	L	S	V	V	K	L	V	K	K	M	I	F	L	L	K	D	D	F	T	G	P	E	F	G	A	D	C

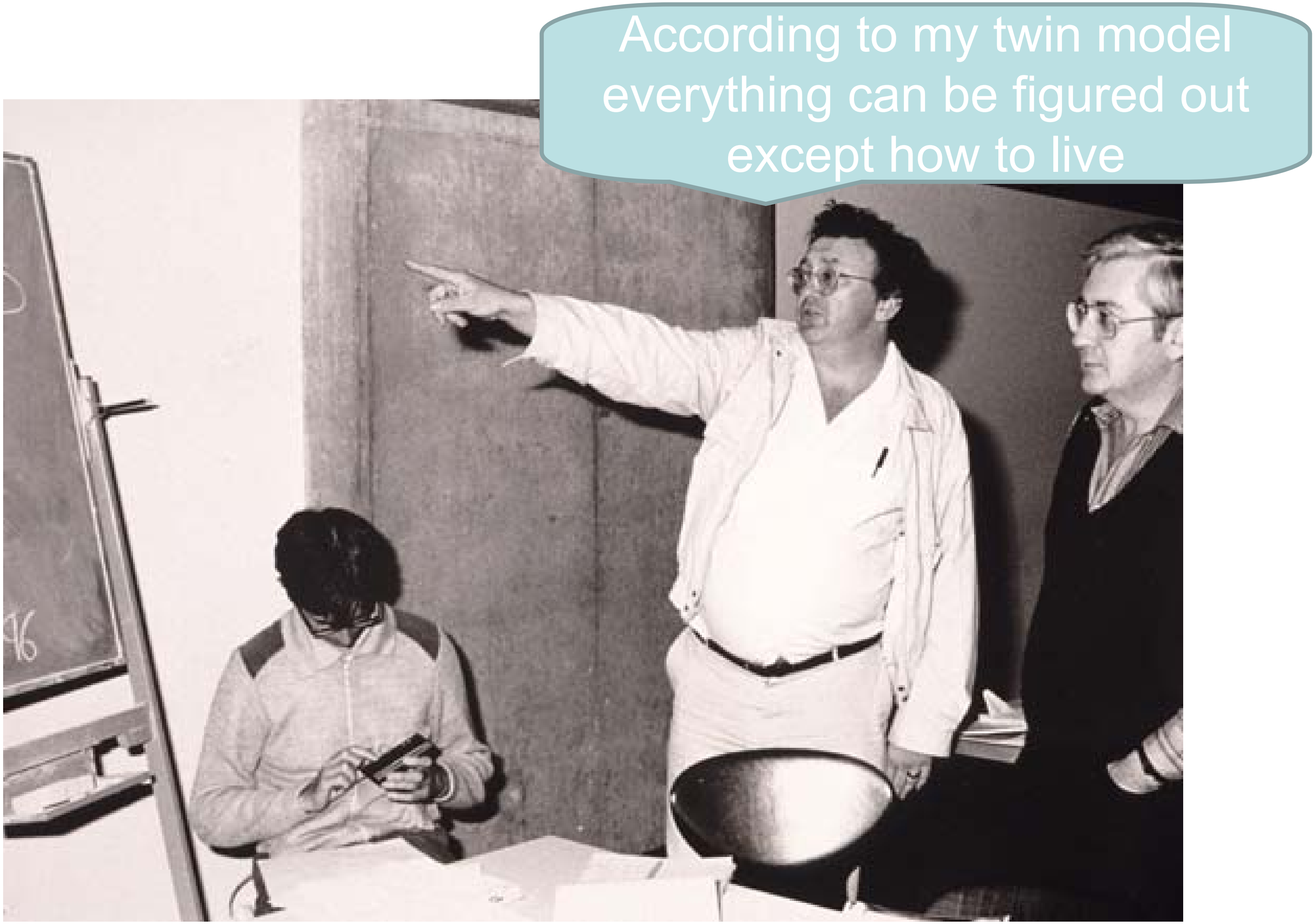
D126E
GV = 176.5
GD = 102.9
Grade: C15
Least probability to be pathogenic

- 🟩 Big Hydrophobic AA
- 🟪 Small Hydrophilic AA
- ▨ Basic AA
- 🟦 Aromatic AA
- ⬛ Acidic and Amid AA
- 🟨 Prolin
- 🟥 Cystein

Parting thought....

“One of the relevant, and scary things, about the Tavtigian paper (and its follow on, not yet written) is that when we tested the 1/1000 ‘pathogenic mutations’ in 5000 more cases, we never saw them again so I suspect there are heaps of them that are super rare, and if we sequenced another 1000 cases, we’d find a different lot”

Georgia Chenevix Trench, March 3 2010


According to my twin model everything can be figured out except how to live