Update on Genome-Wide Association Studies: We Live in Interesting Times

U.S. Department of Health and Human Services
National Institutes of Health
National Human Genome Research Institute

Teri A. Manolio, M.D., Ph.D.
Director, Office of Population Genomics
Senior Advisor to the Director, NHGRI,
for Population Genomics

September 19, 2007

We Live in Interesting Times…

"'May he live in interesting times.' Like it or not we live in interesting times."

--Robert Kennedy, June 7, 1966

May you come to the attention of those in authority.

May you find what you are looking for.

Wikipedia, accessed 11Sep07

May You Live in Interesting Times…

Since 2005, over 30 genome-wide association studies have identified robust associations with genetic variants for nearly 20 common, complex diseases and traits:

- 10 Mar 2005: Age-related macular degeneration
- 30 Apr 2006: QT interval prolongation
- 19 Oct 2006: Neovascular AMD
- 26 Oct 2006: Inflammatory bowel disease
- 11 Feb 2007: Type 2 diabetes
- 5 Mar 2007: Crohn's disease
- 12 Apr 2007: Obesity

Genome-wide association study of prostate cancer identifies a second risk locus at 8q24

Meredith Yeager[1,2], Nick Orr[3], Richard B Hayes[2], Kevin B Jacobs[4], Peter Kraft[5], Sholom Wacholder[2], Mark J Minichiello[6], Paul Fearnhead[7], Kai Yu[2], Nilanjan Chatterjee[2], Zhaoming Wang[1,2], Robert Welch[1,2], Brian J Staats[1,2], Eugenia E Calle[8], Heather Spencer Feigelson[8], Michael J Thun[8], Carmen Rodriguez[8], Demetri
Walter
Edward
David J

Multiple regions within 8q24 independently affect risk for prostate cancer

Christopher A Haiman[1], Nick Patterson[2], Matthew L Freedman[2,3], Simon R Myers[2], Malcolm C Pike[1], Alicja Waliszewska[2,4,5], Julie Neubauer[2,4], Arti Tandon[2,4], Christine Schirmer[2,4], Gavin J McDonald[2,4], Steven
David
Kathlee
Brian E

Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24

Julius Gudmundsson[1,17], Patrick Sulem[1,17], Andrei Manolescu[1,17], Laufey T Amundadottir[1,17], Daniel Gudbjartsson[1], Agnar Helgason[1], Thorunn Rafnar[1], Jon T Bergthorsson[1], Bjarni A Agnarsson[2], Adam Baker[1], Asgeir Sigurdsson[1], Kristrun R Benediktsdottir[2], Margret Jakobsdottir[1], Jianfeng Xu[3], Thorarinn Blondal[1], Jelena Kostic[1], Jielin Sun[3], Shyamali Ghosh[1], Simon N Stacey[1], Magali Mouy[1], Jona Saemundsdottir[1], Valgerdur M Backman[1], Kristleifur Kristjansson[1], Alejandro Tres[4,7], Alan W Partin[5], Marjo T Albers-Akkers[6], Javier Godino-Ivan Marcos[7], Patrick C Walsh[5], Dorine W Swinkels[8], Sebastian Navarrete[9], Sarah D Isaacs[5], Katja K Aben[10], Theresa Graif[11], John Cashy[11], Manuel Ruiz-Echarri[4],

Nature Genetics, 1Apr2007

A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura
Willi
Anne
Tianl
Andr
Craig
Thom
Gonç
Jaakk

Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels

Diab
and

Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes

Eleftheria Zeggini,[1,2*] Michael N. Weedon,[3,4*] Cecilia M. Lindgren,[1,2*] Timothy M. Frayling,[3,4*] Katherine S. Elliott,[2] Hana Lango,[3,4] Nicholas J. Timpson,[2,5] John R. B. Perry,[3,4]

Sciencexpress, 26Apr2007

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

Ruth McPherson,[1*†] Alexander Pertsemlidis,[2*] Nihan Kavaslar,[1] Alexandre Stewart,[1] Robert Roberts,[1] David R. Cox,[3] David A. Hinds,[3] Len A. Pennacchio,[4,5] Anne Tybjaerg-Hansen,[6] Aaron R. Folsom,[7] Eric Boerwinkle,[8] Helen H. Hobbs,[2,9] Jonathan C. Cohen[2,10†]

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,[1*] Gudmar Thorleifsson,[1*] Andrei Manolescu,[1*] Solveig Gretarsdottir,[1] Thorarinn Blondal,[1] Aslaug Jonasdottir,[1] Adalbjorg Jonasdottir,[1] Asgeir Sigurdsson,[1] Adam Baker,[1] Arnar Palsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1] Kristinn P. Magnusson,[1] Karl Andersen,[2] Allan I. Levey,[3] Valgerdur M. Backman,[1] Sigurborg Matthiasdottir,[1] Thorbjorg Jonsdottir,[1] Stefan Palsson,[1] Helga Einarsdottir,[1] Steinunn Gunnarsdottir,[1]

Sciencexpress, 3May2007

Genome-wide association study identifies novel breast cancer susceptibility loci

Dougla
Dennis
Shahar
Christo
Suleep
Hui-Ch
Sheila
Børge
Jolanta
Daehe

A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer

David J
Sholom
Nilanja
Saundra
Richard
Gilles T

Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor–positive breast cancer

Simon N Stacey[1], Andrei Manolescu[1], Patrick Sulem[1], Thorunn Rafnar[1], Julius Gudmundsson[1], Sigurjon A Gudjonsson[1], Gisli Masson[1], Margret Jakobsdottir[1], Steinunn Thorlacius[1], Agnar Helgason[1], Katja K Aben[2,3], Luc J Strobbe[4], Marjo T Albers-Akkers[5], Dorine W Swinkels[3], Brian E Henderson[6], Laurence N Kolonel[7], Loic Le Marchand[7], Esther Millastre[8], Raquel Andres[8], Javier Godino[9], Maria Dolores Garcia-Prats[10], Eduardo Polo[11], Alejandro Tres[8], Magali Mouy[1], Jona Saemundsdottir[1], Valgerdur M Backman[1], Larus Gudmundsson[1], Kristleifur Kristjansson[1], Jon T Bergthorsson[1], Jelena Kostic[1], Michael L Frigge[1], Frank Geller[1], Daniel Gudbjartsson[1], Helgi Sigurdsson[12], Thora Jonsdottir[12],

Nature and *Nature Genetics*, 27May2007

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The W

Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci cont

susc

Miles Pa
Mark Tr
Roland

We followed up on 37 SNPs from 31 distinct loci, associated at $P < 10^{-5}$ on initial analysis of the WTCCC data set. Support for some of these markers diminished in the final WTCCC analysis after extensive data filtering[5]. We selected two markers for each locus where low linkage disequilibrium (LD) between associated SNPs in

Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes

John A Todd[1], Neil M Walker[1,9], Jason D Cooper[1,9], Deborah J Smyth[1,9], Kate Downes[1], Vincent Plagnol[1], Rebecca Bailey[1], Sergey Nejentsev[1], Sarah F Field[1], Felicity Payne[1], Christopher E Lowe[1], Jeffrey S Szeszko[1], Jason P Hafler[1], Lauren Zeitels[1], Jennie H M Yang[1], Adrian Vella[1,8], Sarah Nutland[1], Helen E Stevens[1], Helen Schuilenburg[1], Gillian Coleman[1], Meeta Maisuria[1], William Meadows[1], Luc J Smink[1], Barry Healy[1], Oliver S Burren[1], Alex A C Lam[1], Nigel R Ovington[1], James Allen[1], Ellen Adlem[1], Hin-Tak Leung[1], Chris Wallace[2], Joanna M M Howson[1], Cristian Guja[3], Constantin Ionescu-Tîrgovişte[3], Genetics of Type 1 Diabetes in Finland[4], Matthew J Simmonds[5], Joanne M Heward[5], Stephen C L Gough[5], The Wellcome Trust Case Control Consortium[6], David B Dunger[7], Linda S Wicker[1] & David G Clayton[1]

Nature and *Nature Genetics*, 7Jun2007

2007: The Year of GWA Studies?

Consistently replicated associations found for:

- 10 Jun 2007: Celiac disease
- 1 Jul 2007: Atrial fibrillation
- 8 Jul 2007 : Colorectal cancer
- 15 Jul 2007: Gallstones
- 18 Jul 2007: Periodic limb movements in sleep
- 19 Jul 2007: HIV viral setpoint
- 26 Jul 2007: Childhood asthma
- 29 Jul 2007: Multiple sclerosis
- 1 Aug 2007: Amyotrophic Lateral Sclerosis
- 9 Aug 2007: Exfoliation glaucoma
- 2 Sep 2007: Height
- 5 Sep 2007: Rheumatoid arthritis
- 18 Sep 2007: ??

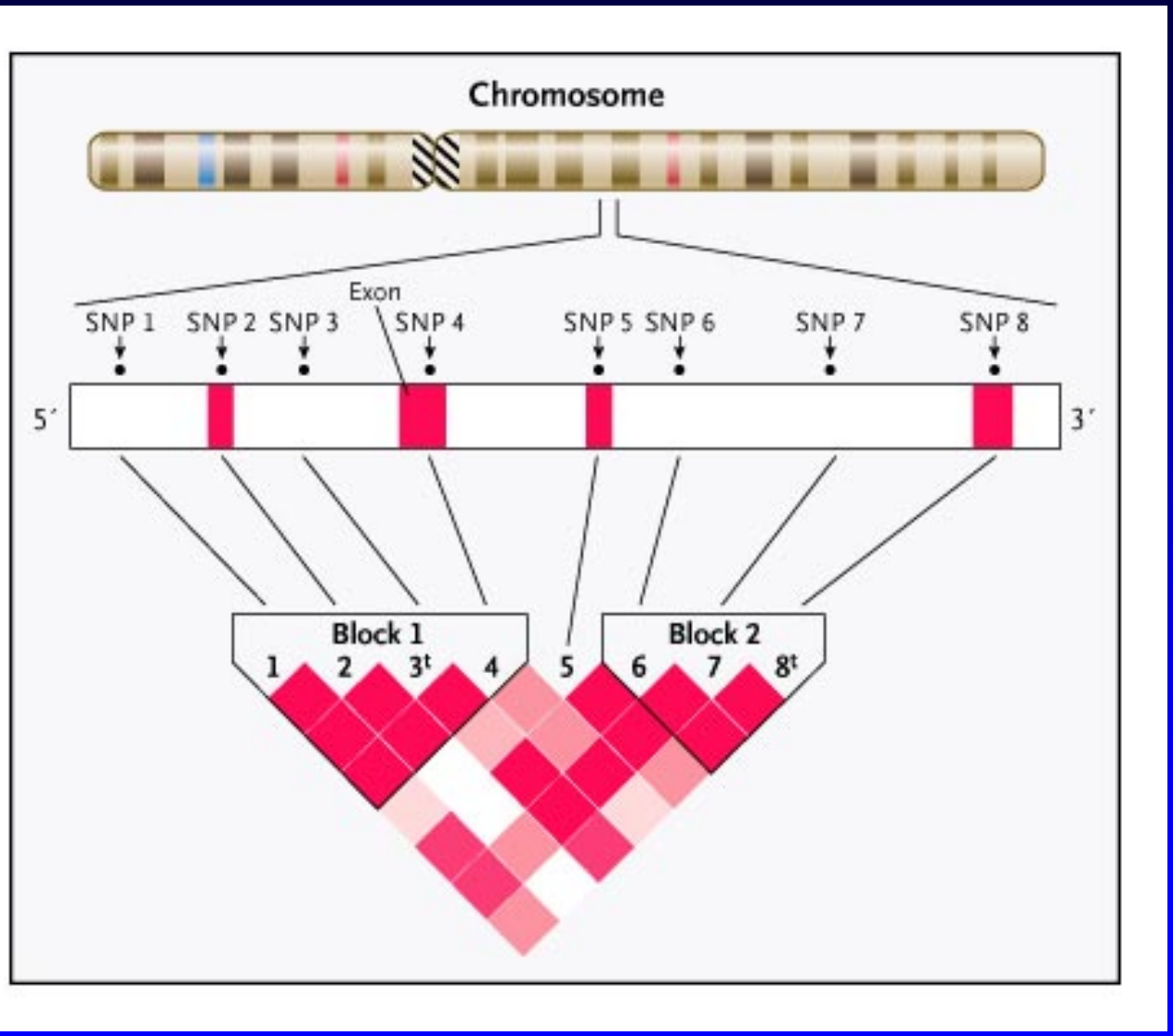
What is a GWA Study?

- Method for interrogating all 10 million variable points across human genome
- Variation inherited in groups, or blocks, so not all 10 million points have to be tested
- Blocks are shorter (so need to test more points) the less closely people are related
- Technology now allows studies in unrelated persons, assuming ~10,000 base pair lengths in common (300,000 - 500,000 markers)

DNA on Chromosome 7

SNPs 1 / 300 bases

Mapping the Relationships Among SNPs



Christensen and Murray, *N Engl J Med* 2007; 356:1094-1097.

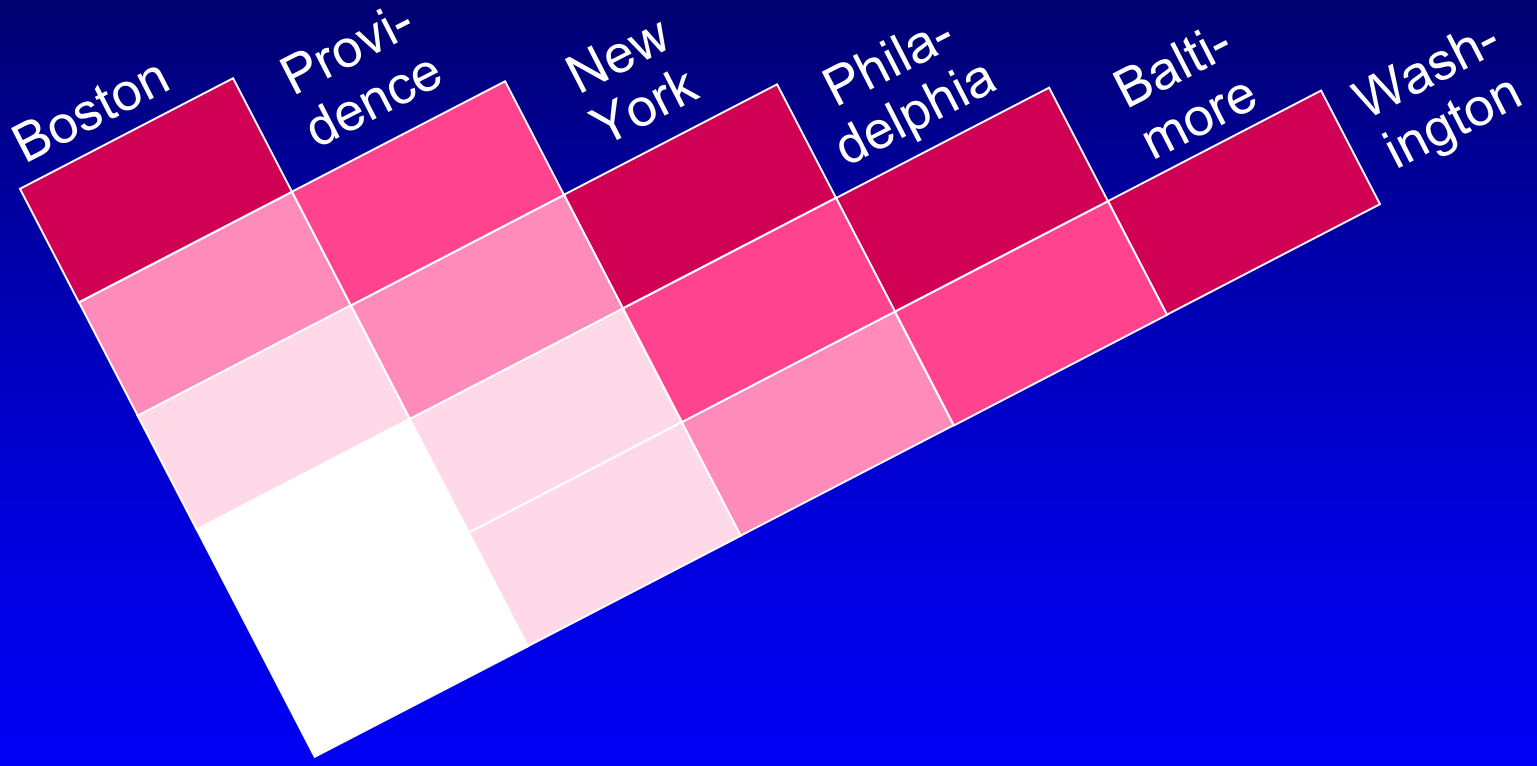
Distances Among East Coast Cities

	Boston	Provi-dence	New York	Phila-delphia	Balti-more
Providence	59				
New York	210	152			
Philadelphia	320	237	86		
Baltimore	430	325	173	87	
Washington	450	358	206	120	34

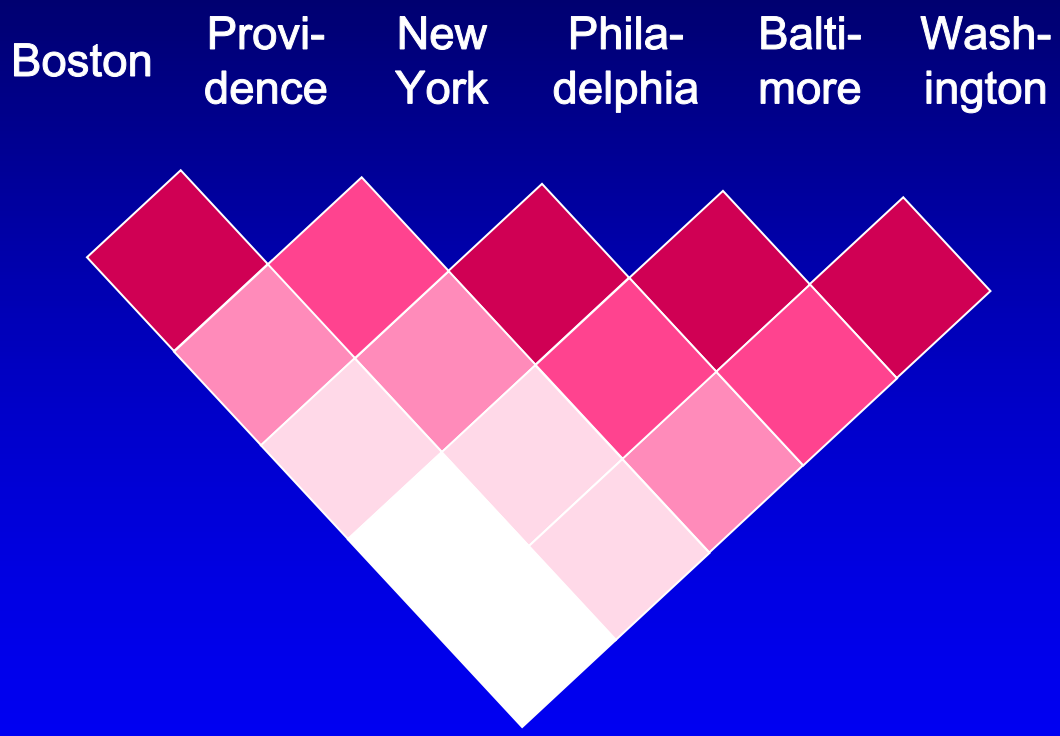
Distances Among East Coast Cities

	Boston	Provi-dence	New York	Phila-delphia	Balti-more
Providence	59				
New York	210	152			
Philadelphia	320	237	86		
Baltimore		325	173	87	
Washington		358	206	120	34

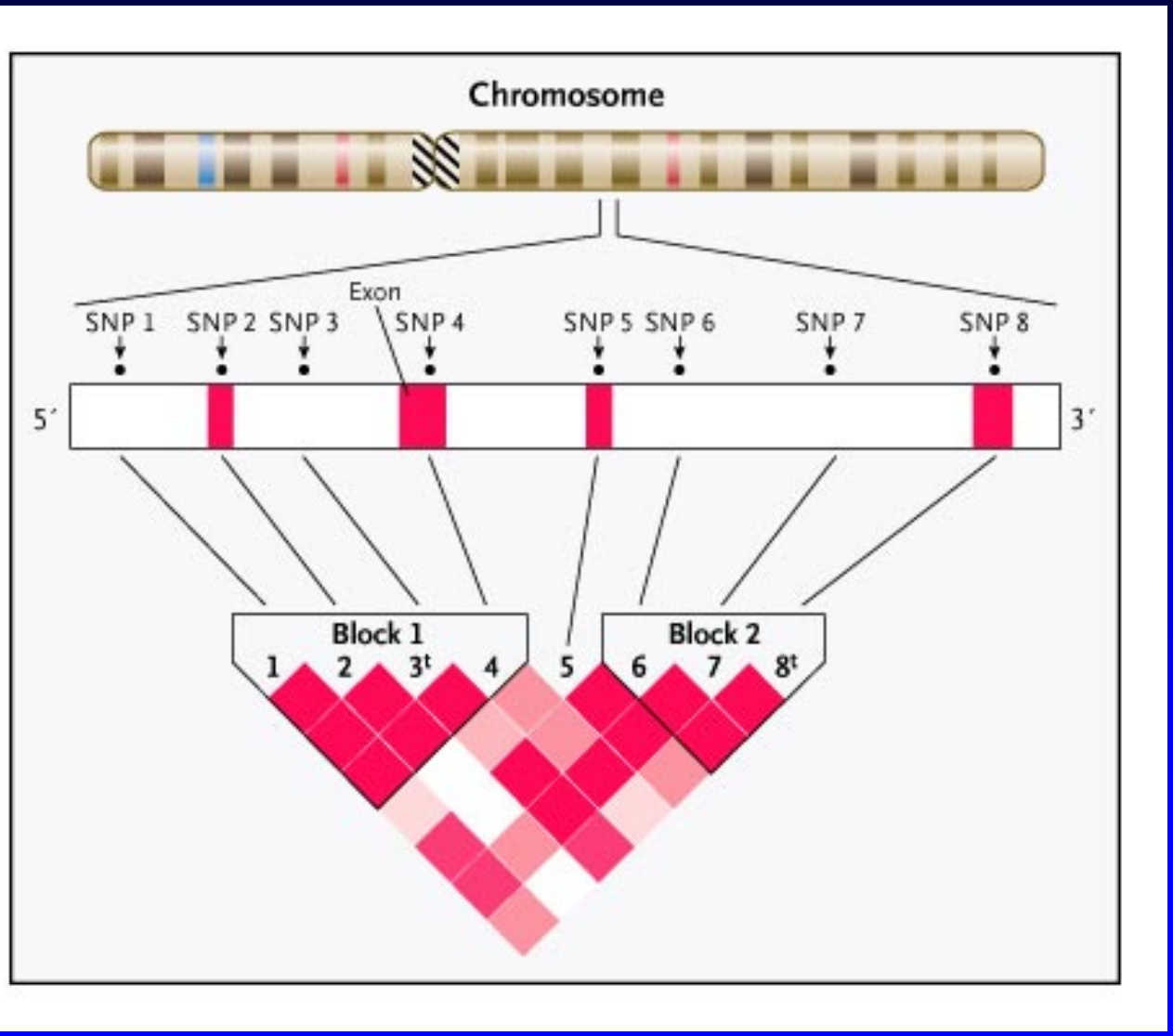
Distances Among East Coast Cities



Distances Among East Coast Cities

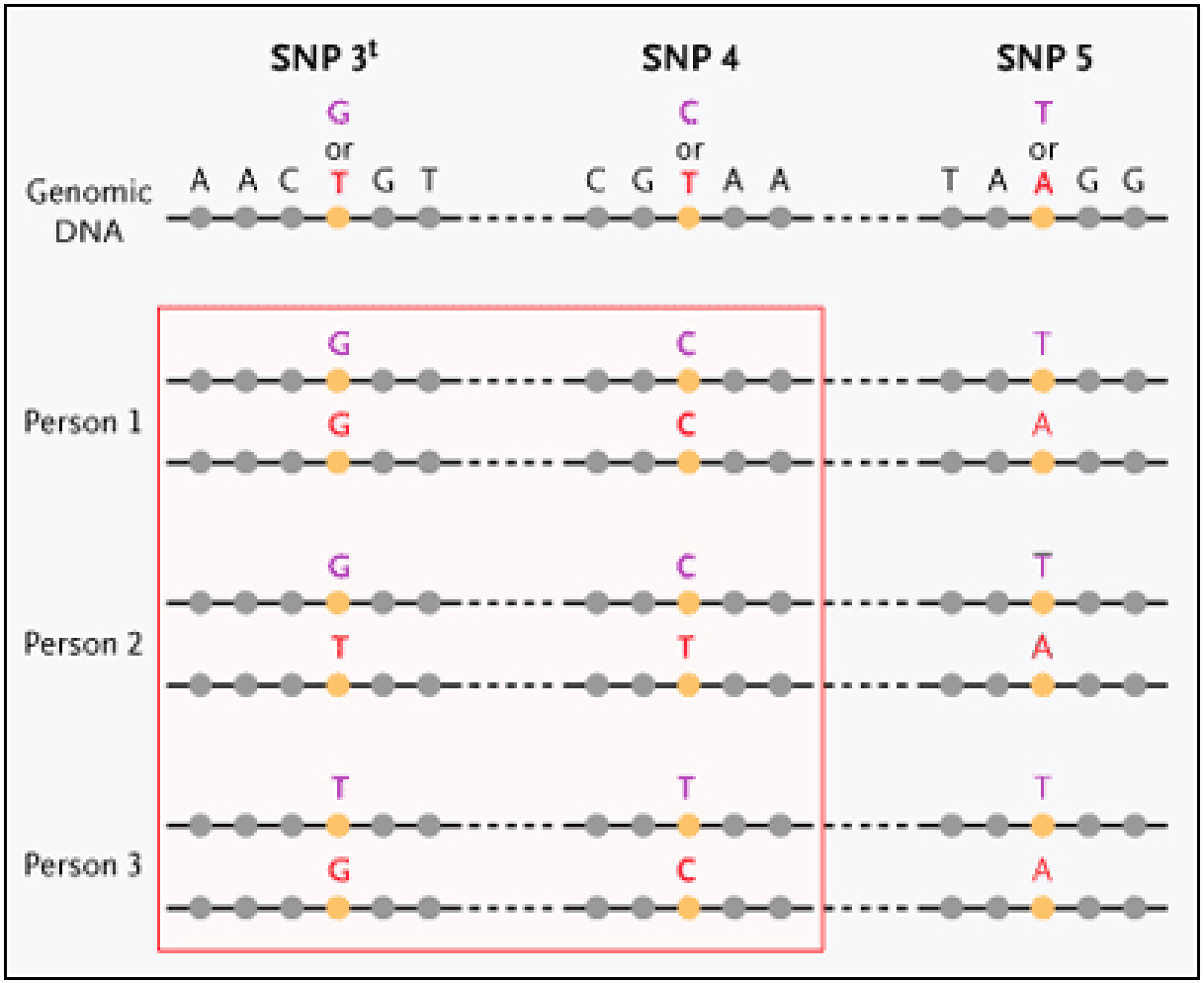


Mapping the Relationships Among SNPs



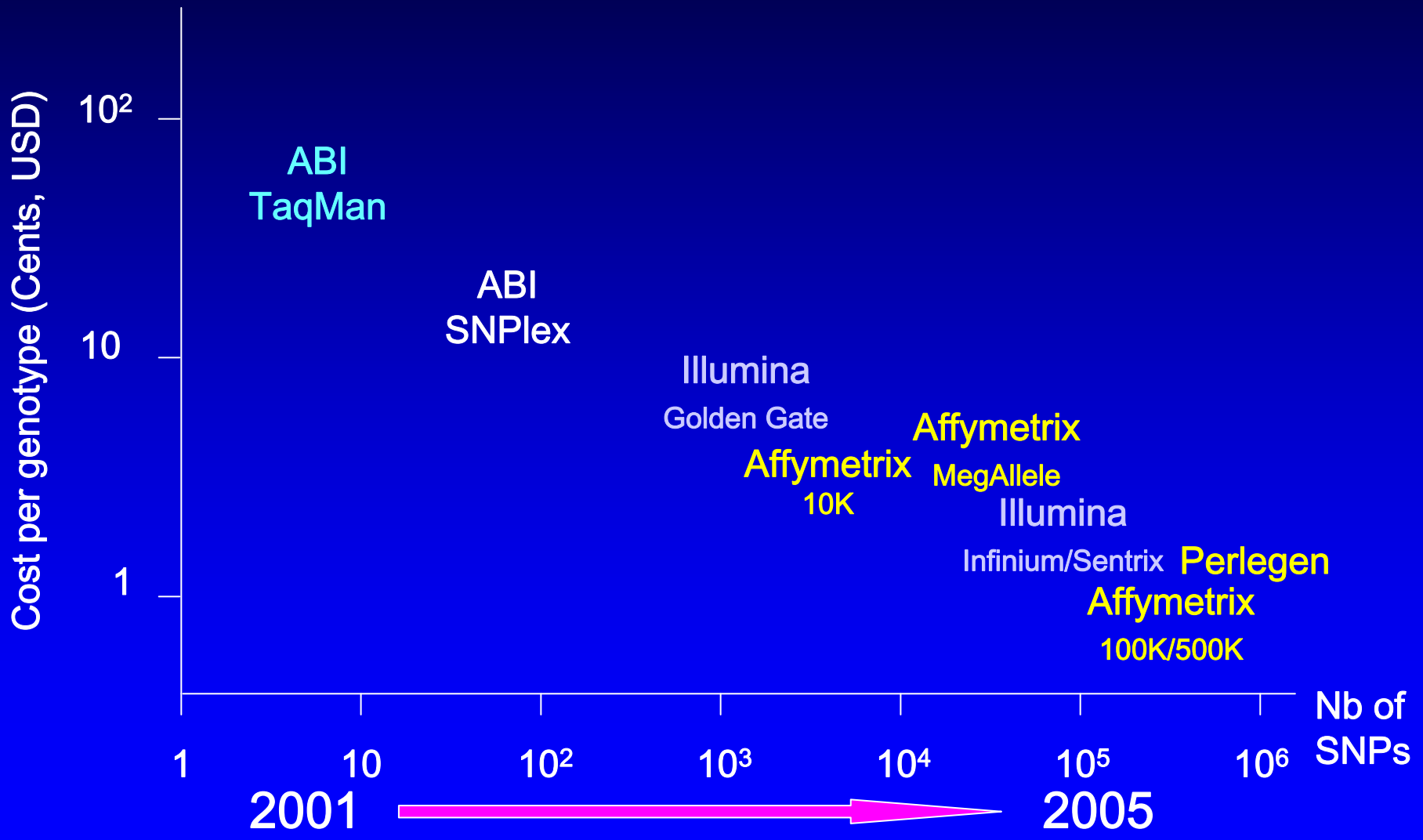
Christensen and Murray, *N Engl J Med* 2007; 356:1094-1097.

One SNP May Serve as Proxy for Many



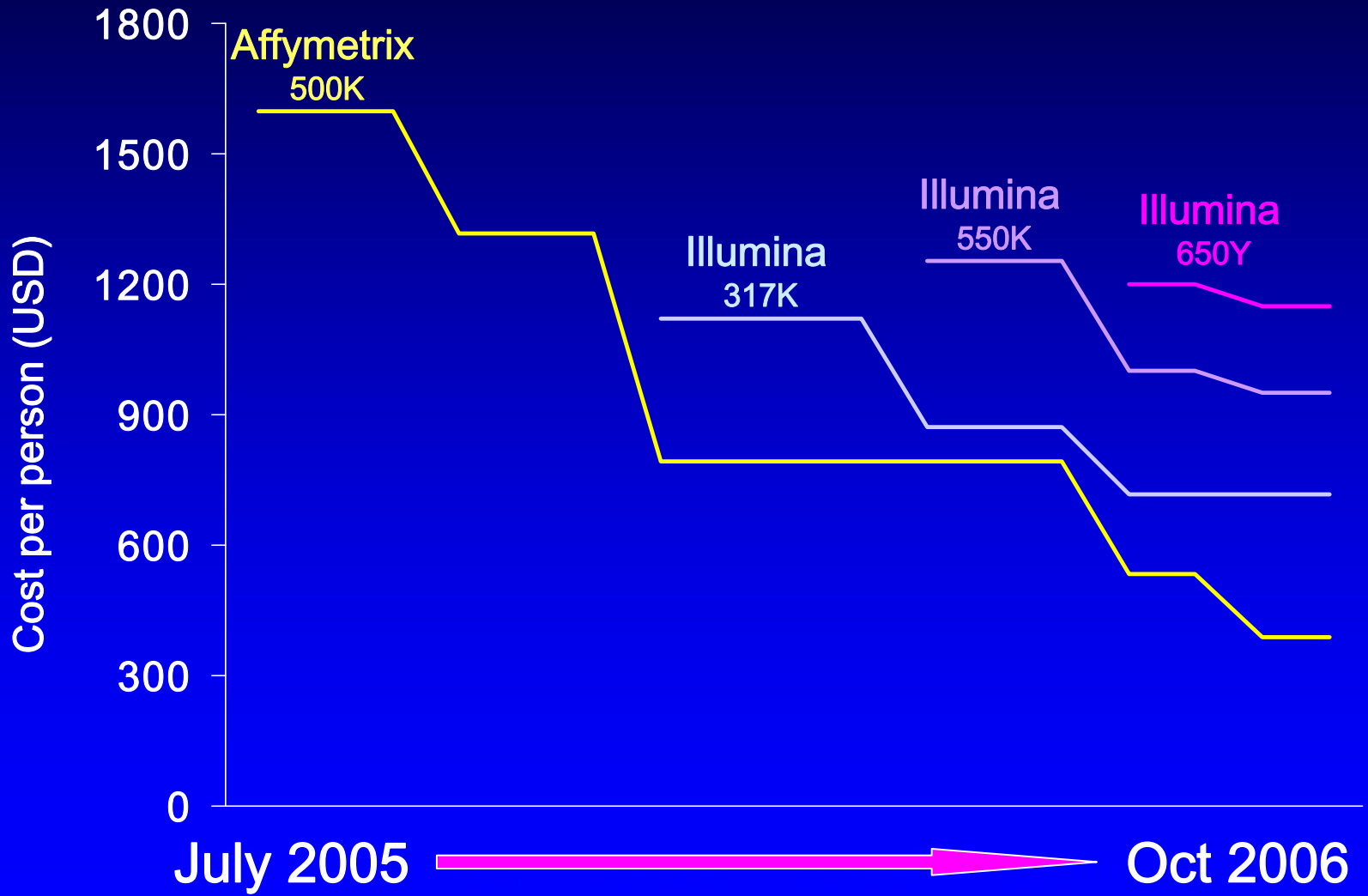
Christensen and Murray, *N Engl J Med* 2007; 356:1094-1097.

Progress in Genotyping Technology



Courtesy S. Chanock, NCI

Continued Progress in Genotyping Technology



Courtesy S. Gabriel, Broad/MIT

Cost of a Genome-Wide Association Study in 2,000 People

Year	Number of SNPs	Cost/SNP	Cost/Study
2001	10,000,000	$1.00	$20 billion
2007	500,000	0.1¢	$1 million

GWA Genotyping Data, Chromosome 22, Parkinson's Study

Study ID	Case/ Control Status	rs5747620		rs2236639	
		Allele 1	Allele 2	Allele 1	Allele 2
14	Case	T	T	G	G
20	Case	T	C	G	G
41	Case	T	C	G	G
412	Control	T	C	G	G
592	Control	C	C	G	G
665	Control	T	C	A	G

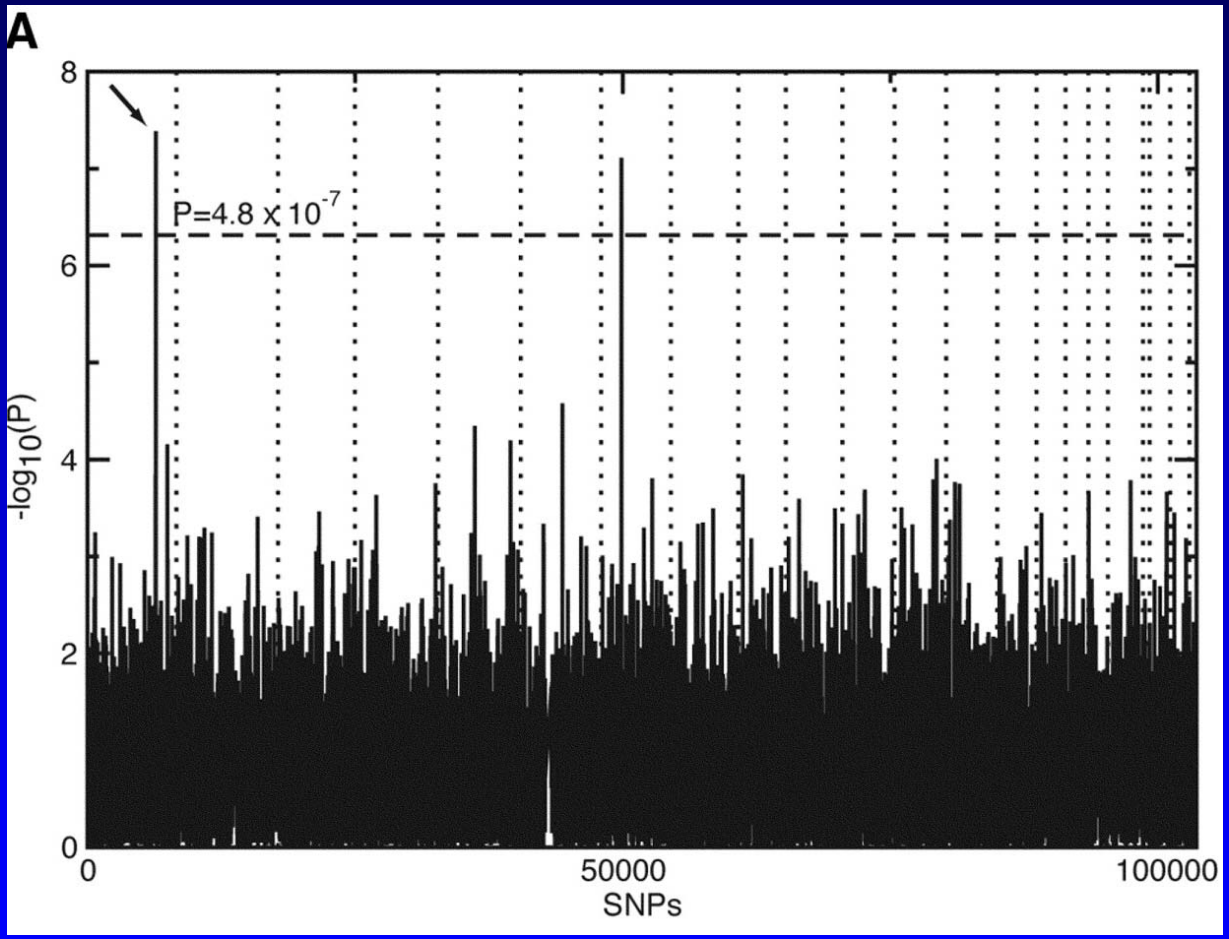
http://ccr.coriell.org/ninds/

Association of rs2236639 Alleles with Development of Parkinson Disease (Made Up!)

Variant Allele (A)	Development of Disease: Develop Disease	Development of Disease: Do Not Develop Disease	Total
Present	10	70	80
Absent	40	880	920
Total	50	950	1,000

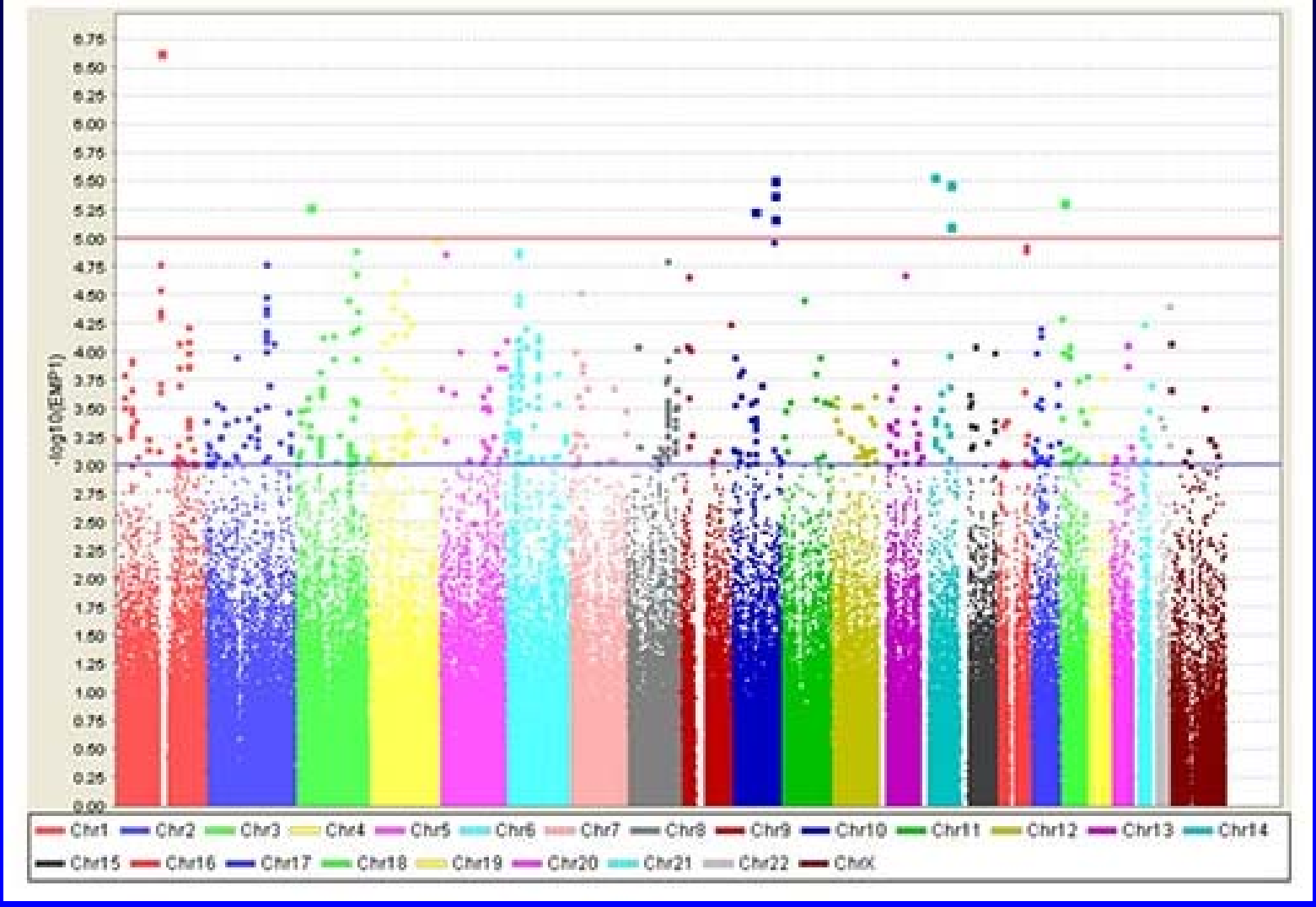
$$\text{Relative Risk} = \frac{\text{Risk in Exposed}}{\text{Risk in Unexposed}} = \frac{10/80}{40/920} = \frac{12.5\%}{4.3\%} = 2.9$$

P Values of GWA Scan for Age-Related Macular Degeneration



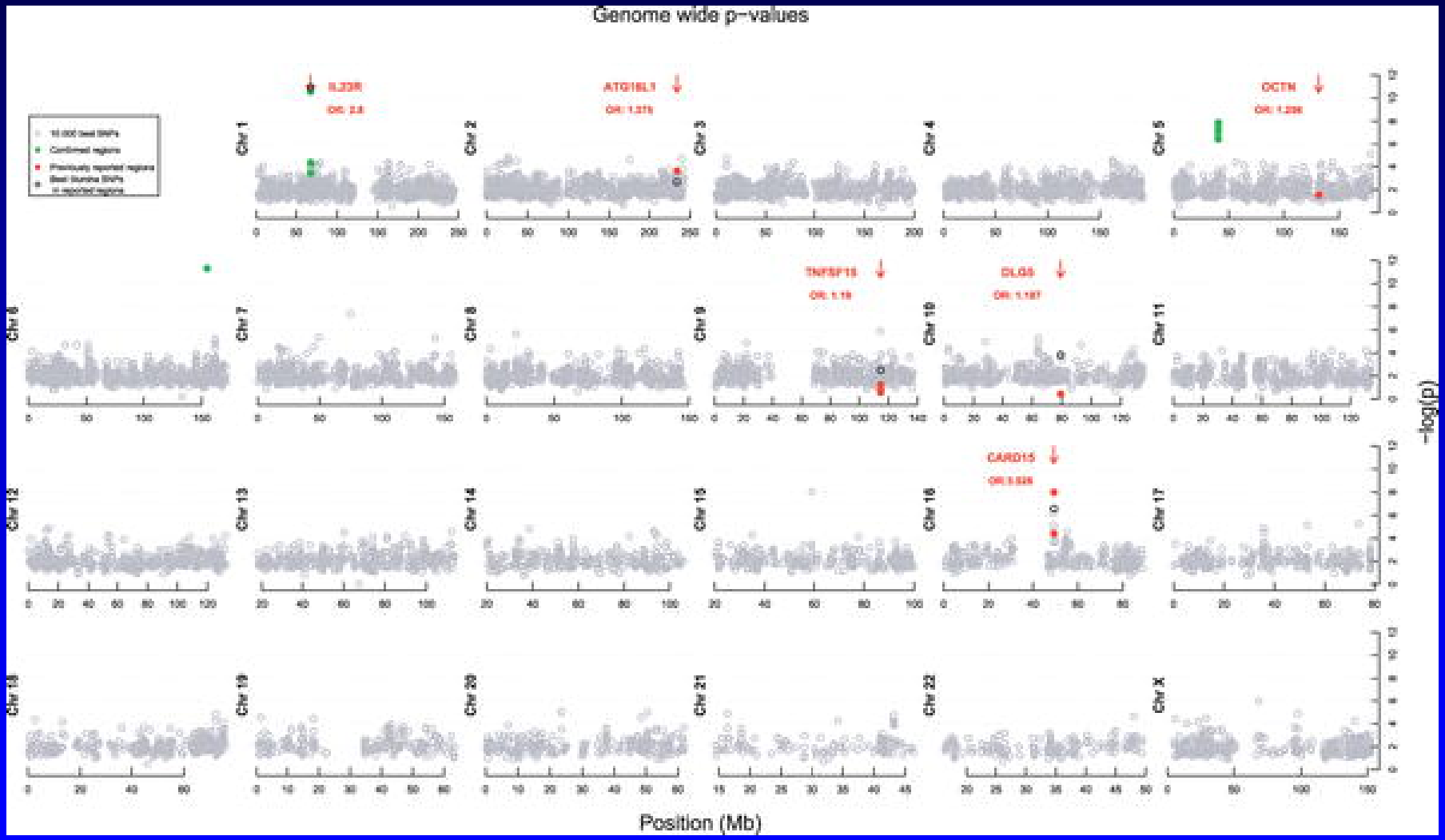
Klein et al, *Science* 2005; 308:385-389.

Genome-Wide Scan for Type 2 Diabetes in a Scandinavian Cohort



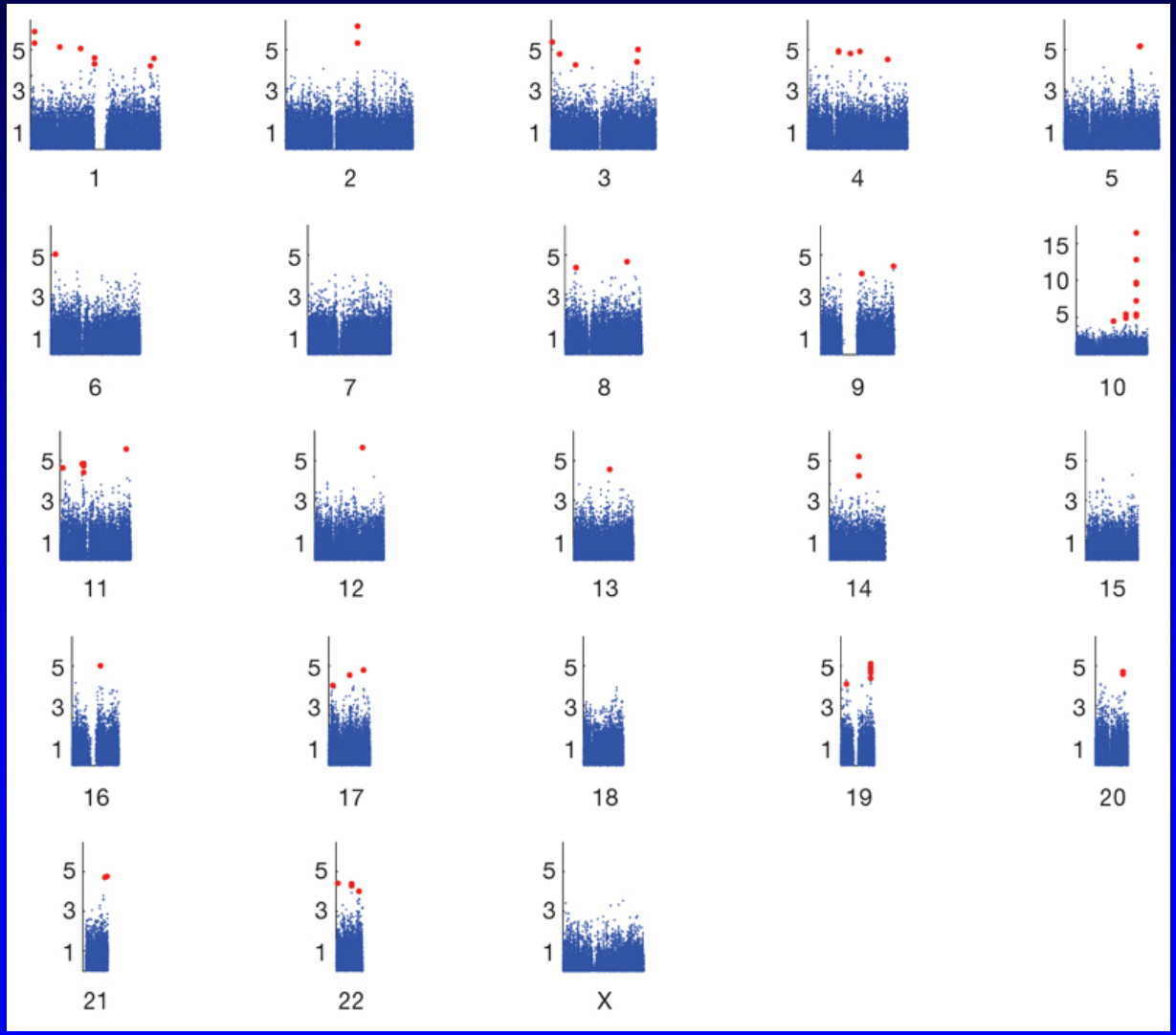
http://www.broad.mit.edu/diabetes/scandinavs/type2.html

Genome-Wide Scan for Crohn Disease in Belgian Cases and Controls



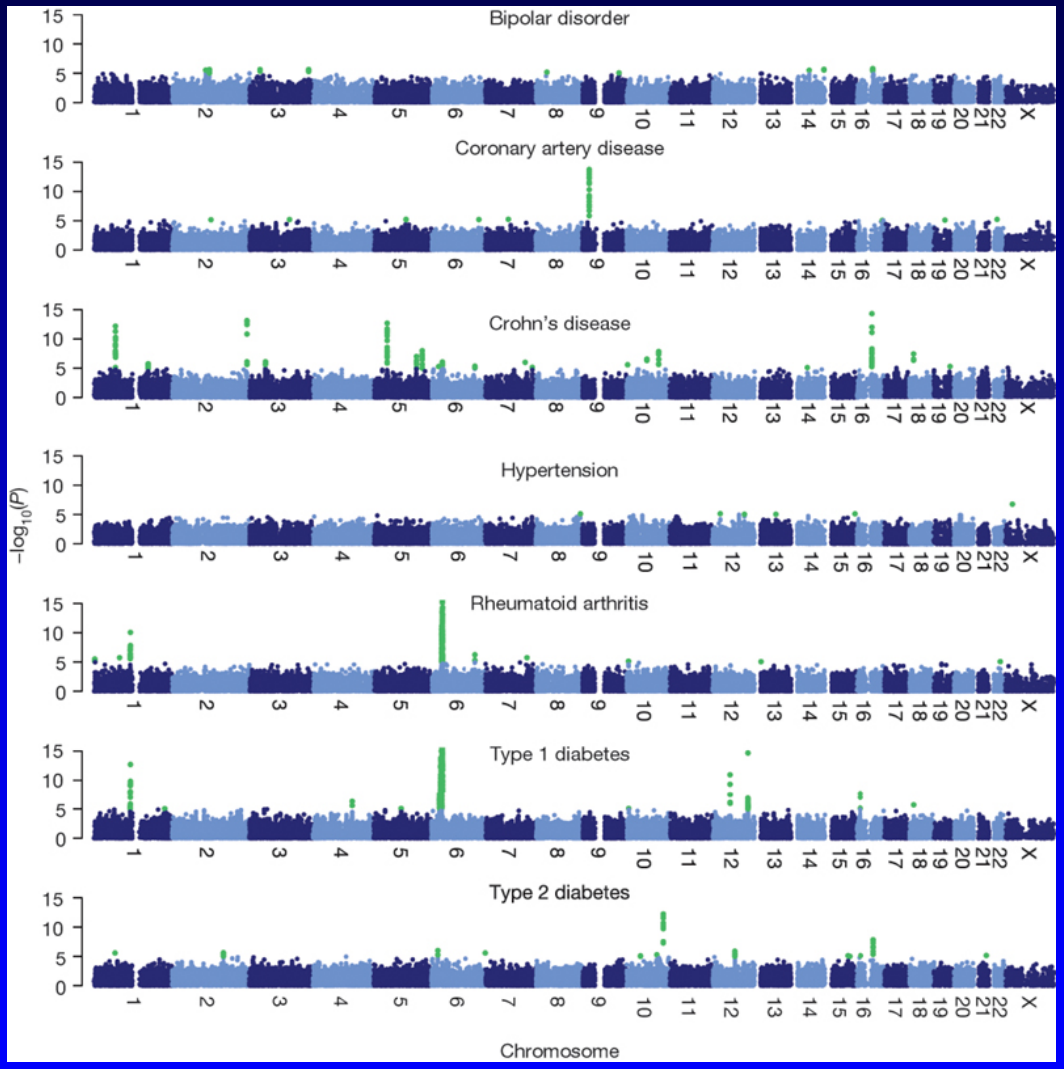
Libioulle C et al, *PLoS Genet*; 2007 Apr 20;3(4):e58.

Genome-Wide Scan for Type 2 Diabetes in French Case-Control Study



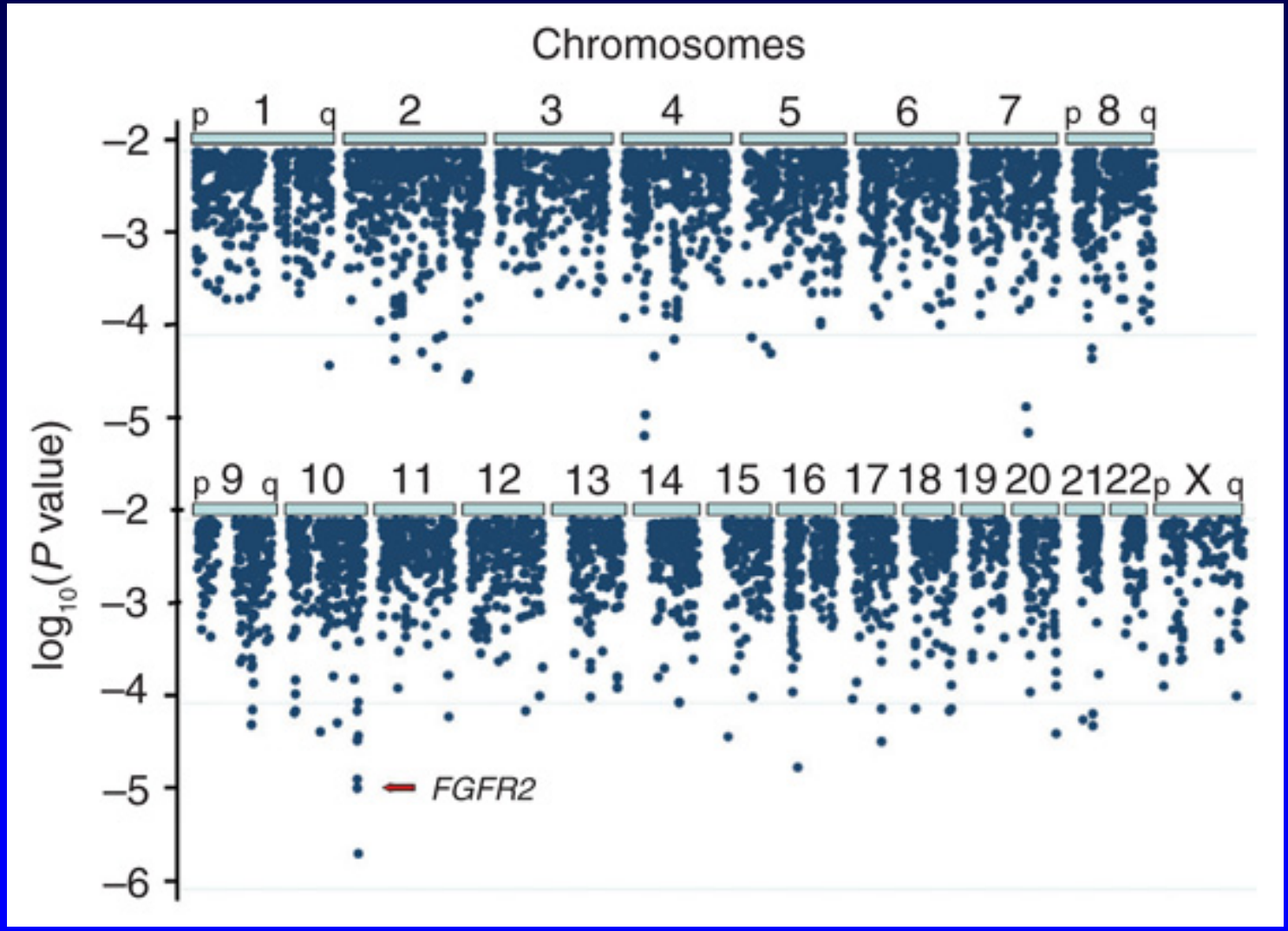
Sladek R et al, *Nature* 2007; 445, 881-885.

Wellcome Trust Genome-Wide Association Study of Seven Common Diseases



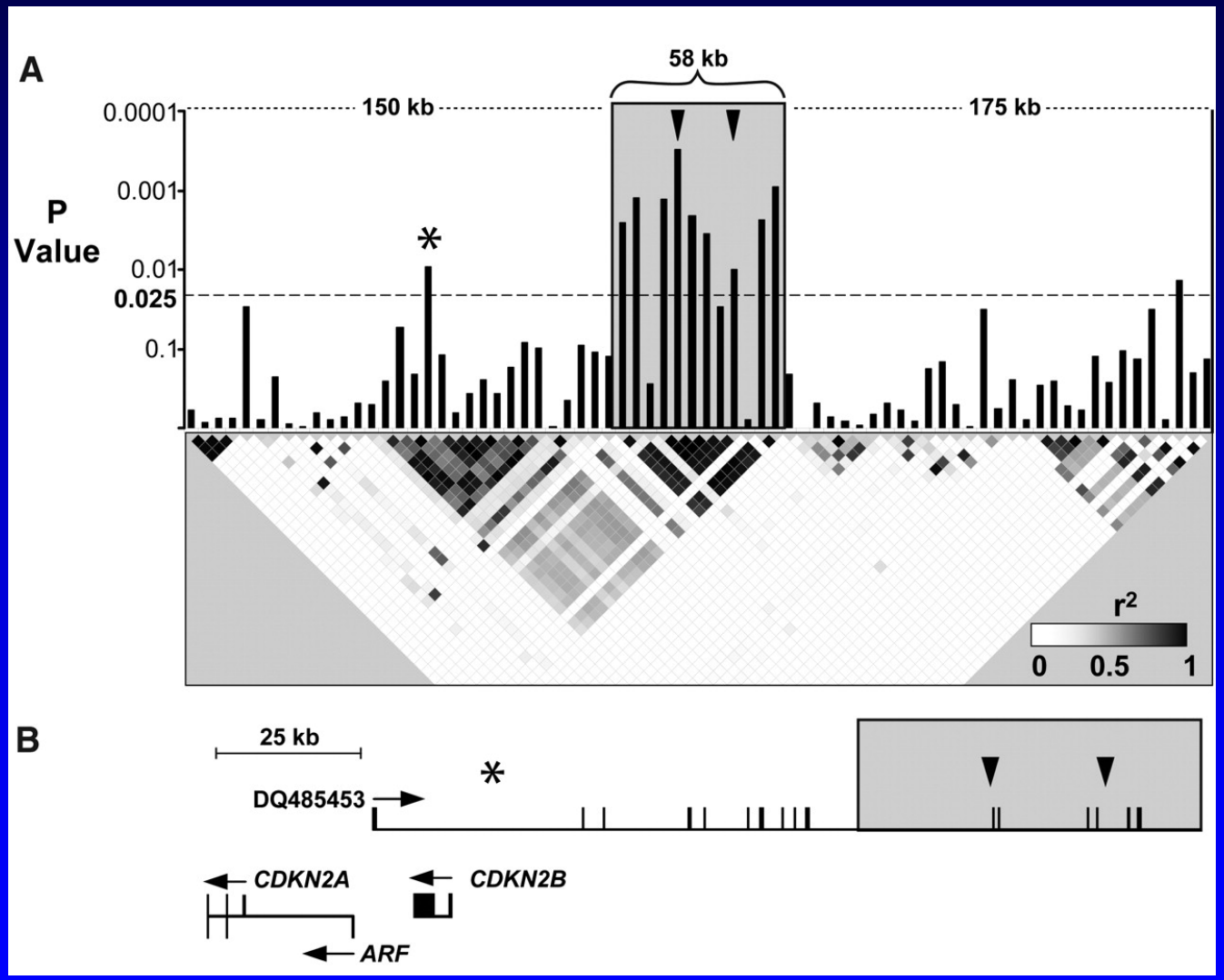
WTCCC, *Nature* 2007; 447:661-678.

Genome-Wide Scan for Breast Cancer in Postmenopausal Women



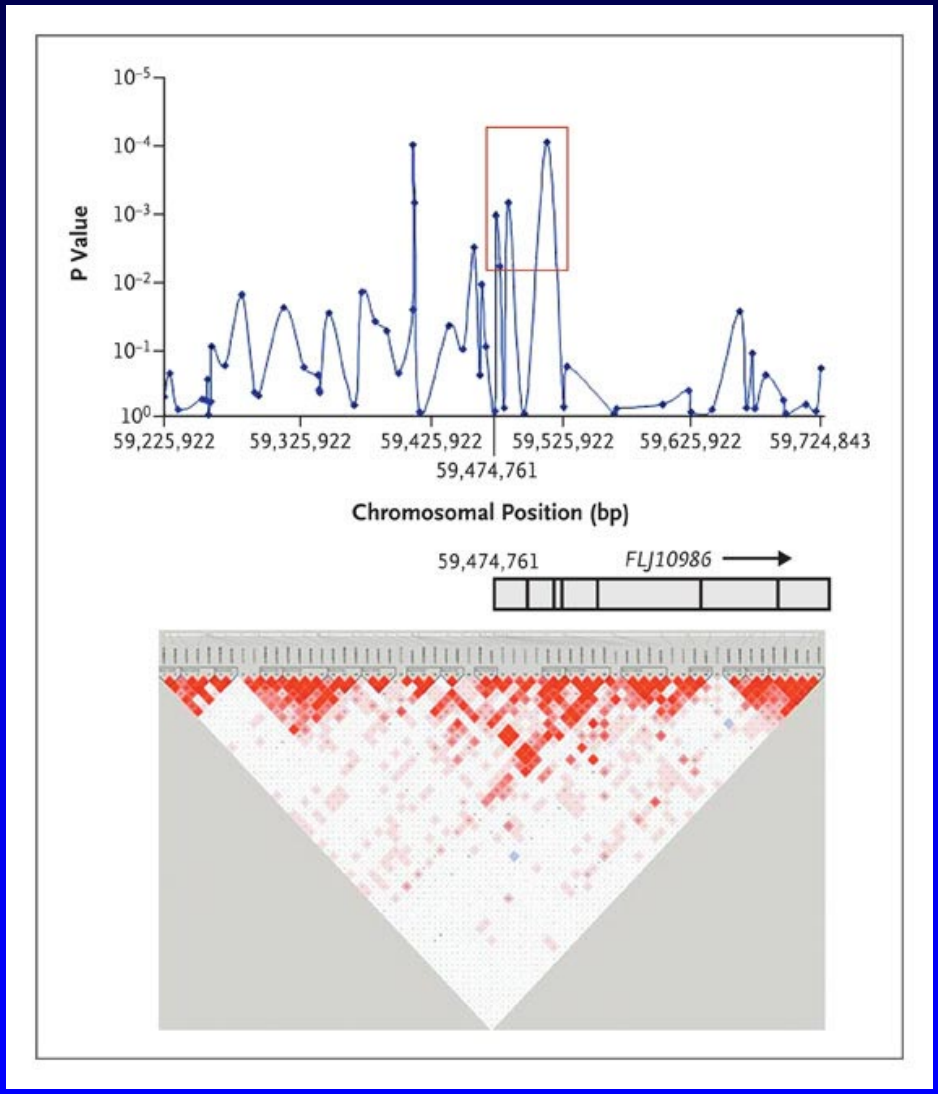
Hunter DJ et al, *Nat Genet* 2007; 39:870-874.

Genome-Wide Scan for Coronary Heart Disease in Ottawa Case-Control Study



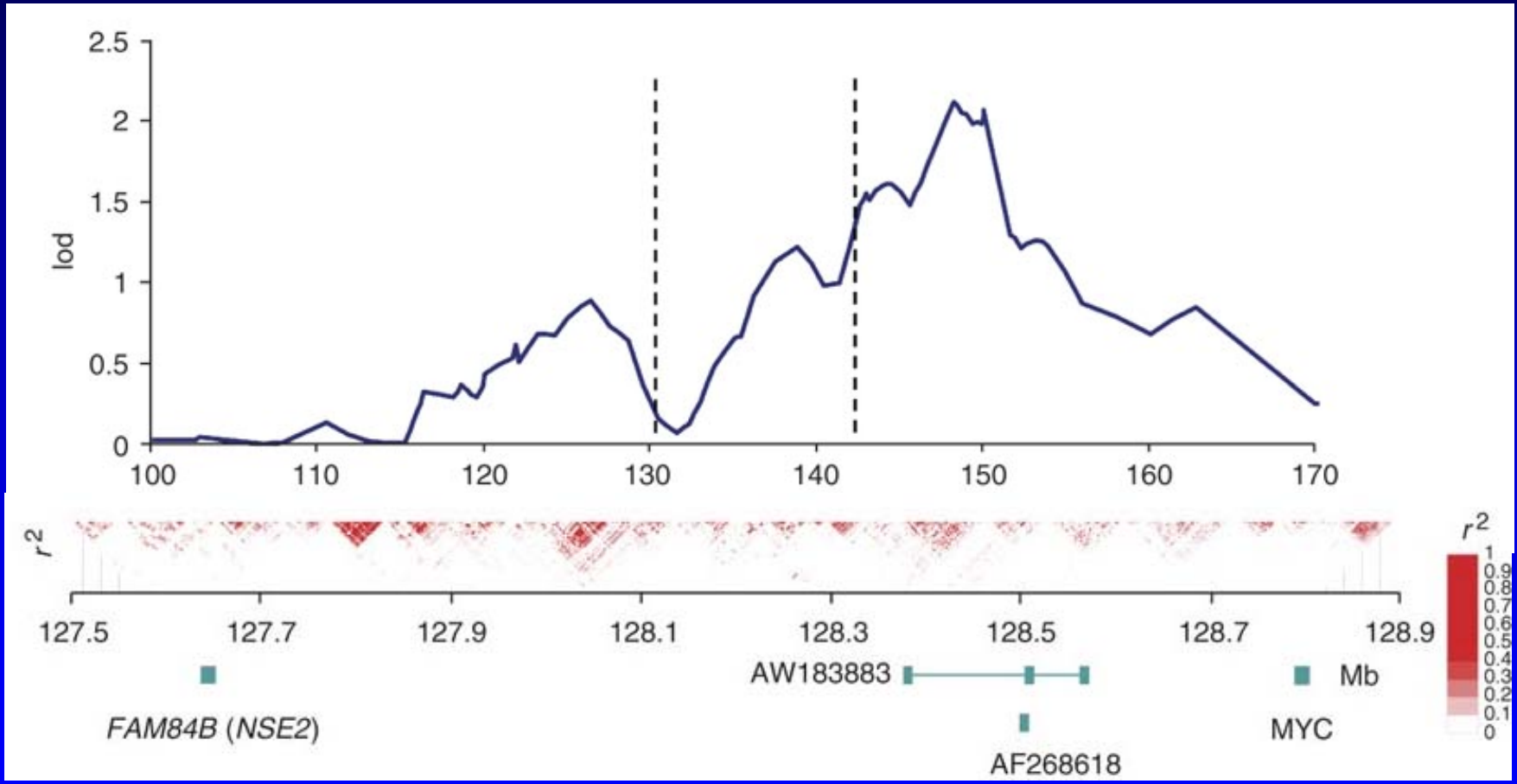
McPherson R et al, *Nature* 2007; 316:1488-1491.

Genome-Wide Scan for Sporadic Amyotrophic Lateral Sclerosis



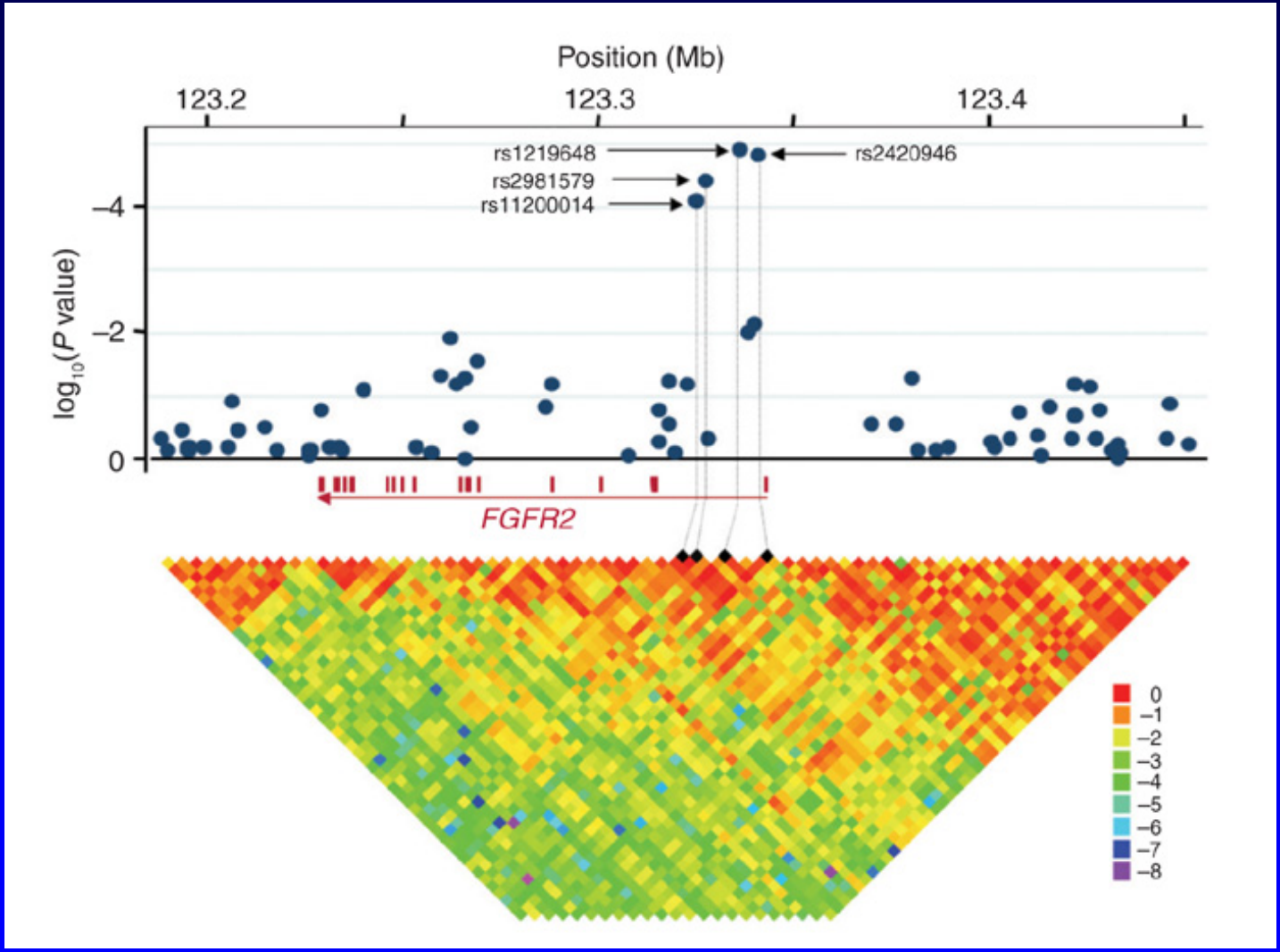
Dunckley T et al, *N Engl J Med* 2007; 357:775-788.

Genome-Wide Scan for Prostate Cancer

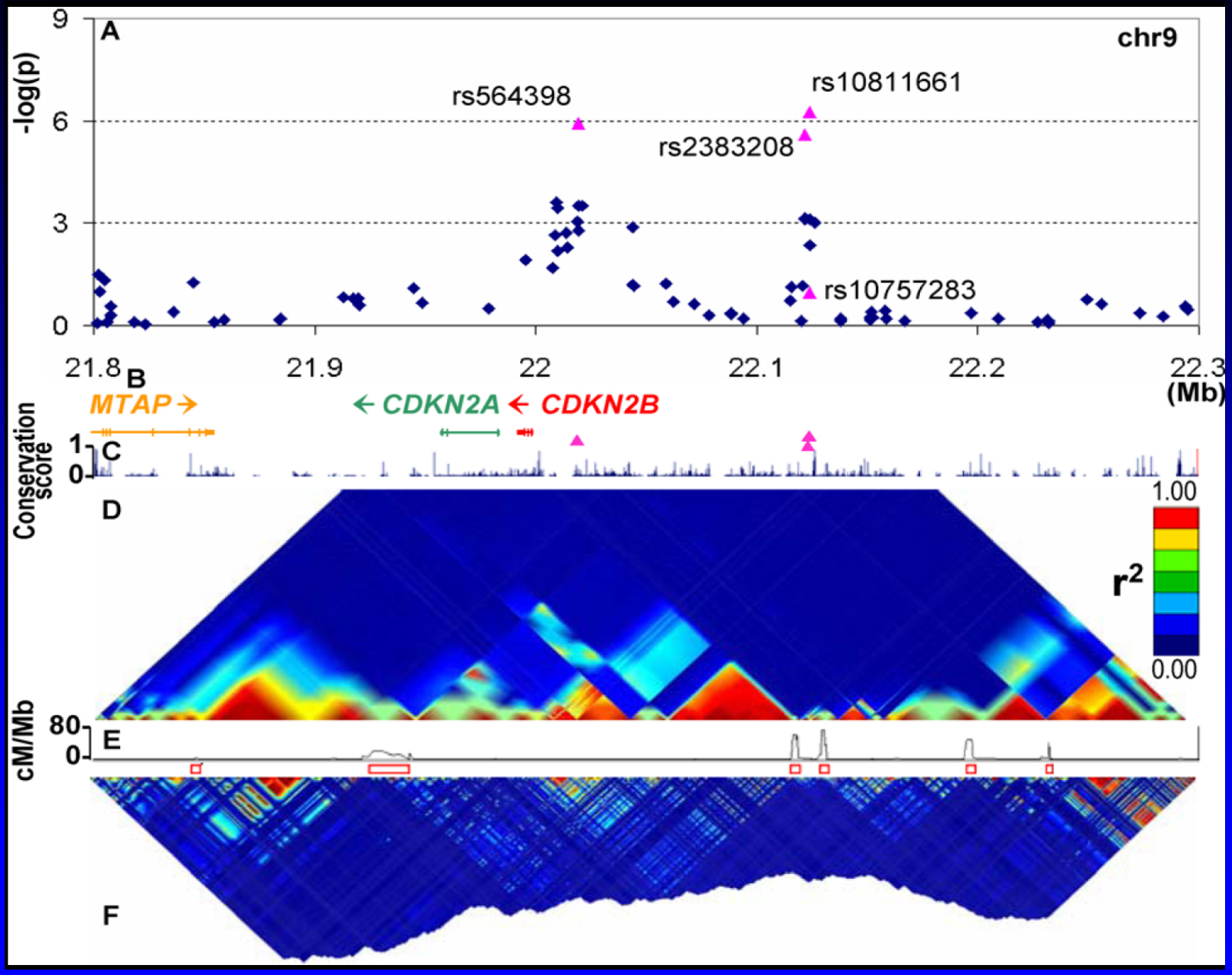


Gudmundsson J et al, *Nat Genet* 2007; 39:631-637.

Association Analysis of SNPs across FGFR2



Hunter DJ et al, *Nat Genet* 2007; 39:870-874.



Courtesy, F. Collins

Lessons Learned from Initial GWA Studies

Signals in Gene "Deserts"	
Prostate Cancer	8q24
Crohn's Disease	5p13.1, 1q31.2, 10p21
Signals in Common	
Diabetes, CHD, Melanoma	*CDKN2A/2B*
Prostate, Breast, Colon Cancer	8q24 region
Crohn's Disease, Psoriasis	*IL23R*
Crohn's Disease, T1DM	*PTPN2*

STATISTICS AND MEDICINE

Drinking from the Fire Hose — Statistical Issues in Genomewide Association Studies

David J. Hunter, M.B., B.S., and Peter Kraft, Ph.D.

Related article, page 443

"There have been few, if any, similar bursts of discovery in the history of medical research…"

The past 3 months have seen the publication of a series of

ating the need for guessing which genes are likely to harbor variants

The main problem with this strategy is that because of the

ost stud-
ained in
samples
ower to

and in this issue of the *Journal*, coronary artery disease (reported by Samani et al., pages 443–453). These genomewide association studies have been able to examine interpatient differences in inherited genetic variability at an unprecedented level of resolution, thanks to the development of microarrays, or chips, capable of as-

lated to the disease. Some of these associations have been found in regions not even known to harbor genes, such as the 8q24 region, in which multiple variants have been found to be associated with prostate cancer.[2] Such findings promise to open up new avenues of research, through both the discovery of new genes rele-

generate P values as small as 10^{-7}. In addition, most variants identified recently have been associated with modest relative risks (e.g., 1.3 for heterozygotes and 1.6 for homozygotes), and many true associations are not likely to exceed P values as extreme as 10^{-7} in an initial study. On the other hand, a "statistically significant" finding

Hunter DJ and Kraft P, *N Engl J Med* 2007; 357:436-439.

Unique Aspects of GWA Studies

- Permits examination of inherited genetic variability at unprecedented level of resolution
- Permits "agnostic" genomewide comparison
- Most robust associations in GWA studies have not been with genes previously suspected of being related to the disease
- Some associations in regions not even known to harbor genes

"The chief strength of the new approach also contains its chief problem: with more than 500,000 comparisons per study, the potential for false positive results is unprecedented."

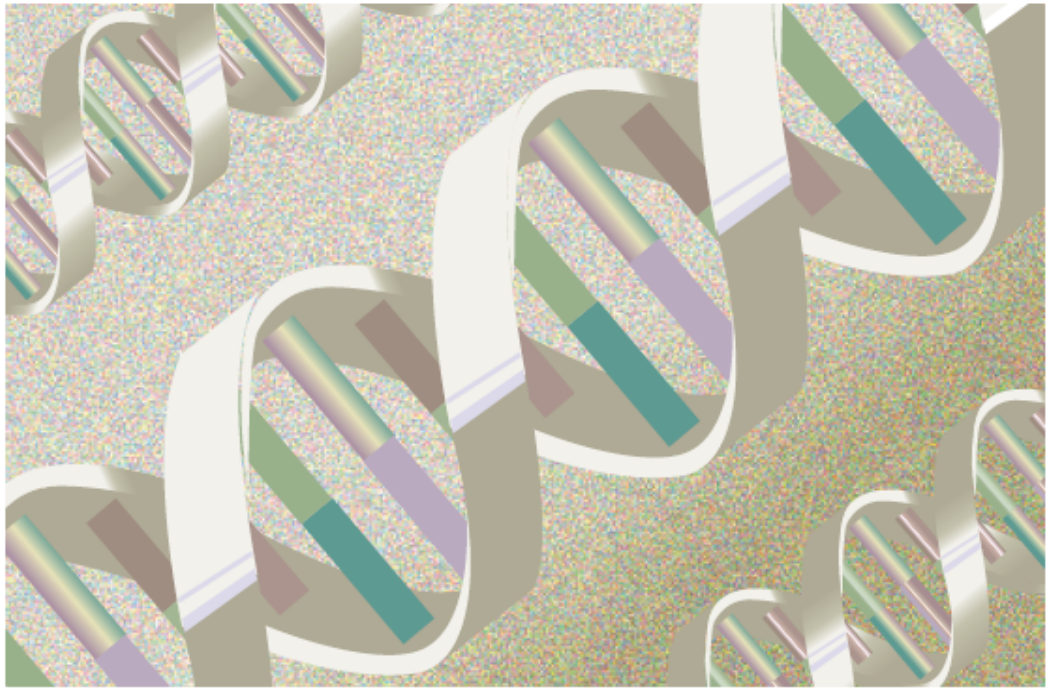
N Engl J Med 2007; 357:436-439.

FEATURE

Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies



The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)[1–3]. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even 'agnostic', approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

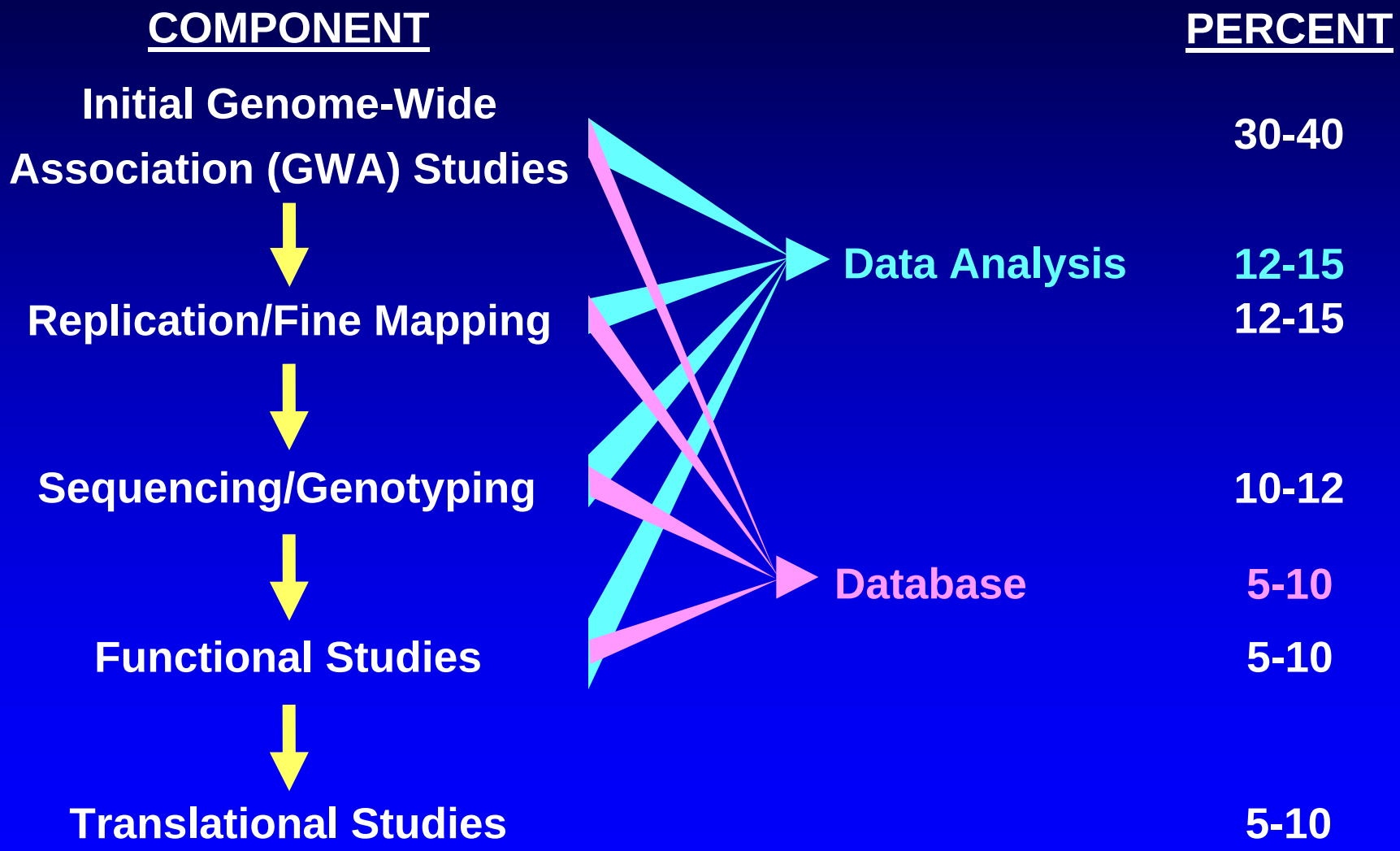
So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies[4–7]. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of false positives attained through attempts to rep-

studies because of issues in either the initial study or the attempted replication[4–6,32,33]. Small sample size is a frequent problem and can result

conclusion from the literature because follow-up studies have not consistently analysed the same markers or those in perfect linkage dis-

Chanock et al, *Nature* 2007; 447:655-660.

Flow of Investigation: From Genome-Wide Association to Clinical Translation



Availability of GWA Data in NIH Databases: Current

- Database of Genotype and Phenotype (dbGaP): http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap
- Cancer Biomedical Information Grid (caBIG) and Cancer Genetic Markers of Susceptibility (CGEMS): https://caintegrator.nci.nih.gov/cgems/

Possible Implications of Many Variants of Small Effect

- Need not carry all of them to develop disease
- Probably need to carry more than one, unless very strong environmental interaction
- Some may affect same pathways and be duplicative
- Others may affect different pathways, so some key combination(s) needed
- Should be possible to identify “clusters” of variants carried by different groups of cases
- May be possible to classify on molecular basis

“The more we find, the more we see, the more we come to learn.

The more that we explore, the more we shall return.”

Sir Tim Rice, *Aida*, 2000