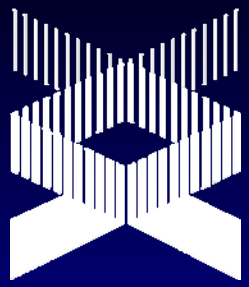


National Human Genome Research Institute

National Institutes of Health

U.S. Department of Health and Human Services

State of the Science in Genomics: Genome-Wide Association and Complex Diseases

U.S. Department of Health and Human Services

National Institutes of Health

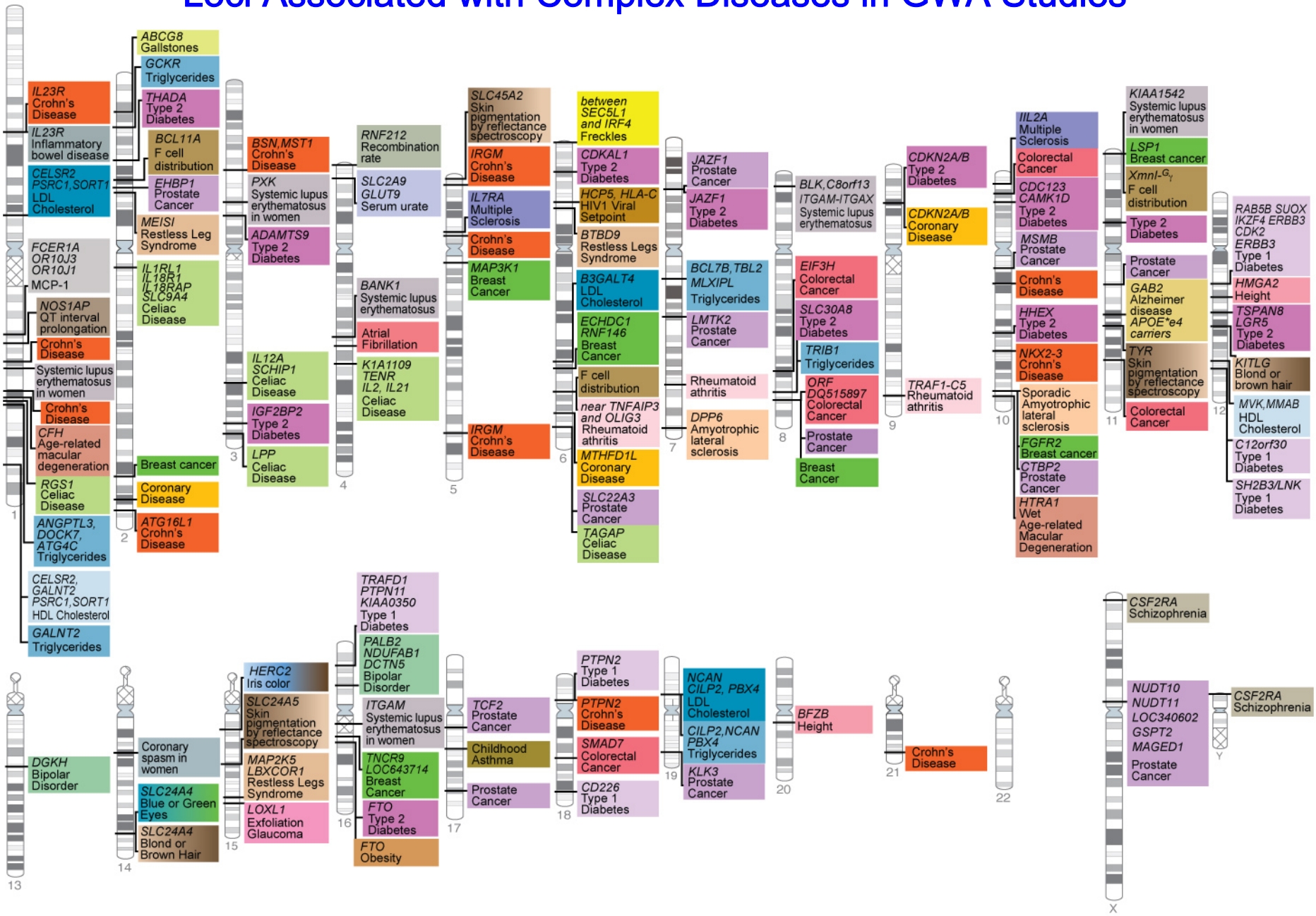
National Human Genome Research Institute

June 4, 2008

New Insights into Complex Diseases from Genome-Wide Association

- Early output of genome-wide association studies
- Initial lessons learned
- Need for large samples sizes and collaboration among population studies
- Data sharing through the Database of Genotype and Phenotype (dbGaP)
- Integration of basic science into population-based GWA studies

Loci Associated with Complex Diseases in GWA Studies



Manolio, Brooks, Collins, *J Clin Invest* 2008; 118:1590-605.

Vol 447|7 June 2007|doi:10.1038/nature05911 nature

ARTICLES

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

The Wellcome Trust Case Control Consortium*

There is increasing evidence that genome-wide association (GWA) studies represent a powerful approach to the identification of genes involved in common human diseases. We describe a joint GWA study (using the Affymetrix GeneChip 500K Mapping Array Set) undertaken in the British population, which has examined ~2,000 individuals for each of 7 major diseases and a shared set of ~3,000 controls. Case-control comparisons identified 24 independent association signals at $P < 5 \times 10^{-7}$: 1 in bipolar disorder, 1 in coronary artery disease, 9 in Crohn's disease, 3 in rheumatoid arthritis, 7 in type 1 diabetes and 3 in type 2 diabetes. On the basis of prior findings and replication studies thus-far completed, almost all of these signals reflect genuine susceptibility effects. We observed association at many previously identified loci, and found compelling evidence that some loci confer risk for more than one of the diseases studied. Across all diseases, we identified a large number of further signals (including 58 loci with single-point *P* values between 10^{-5} and 5×10^{-7}) likely to yield

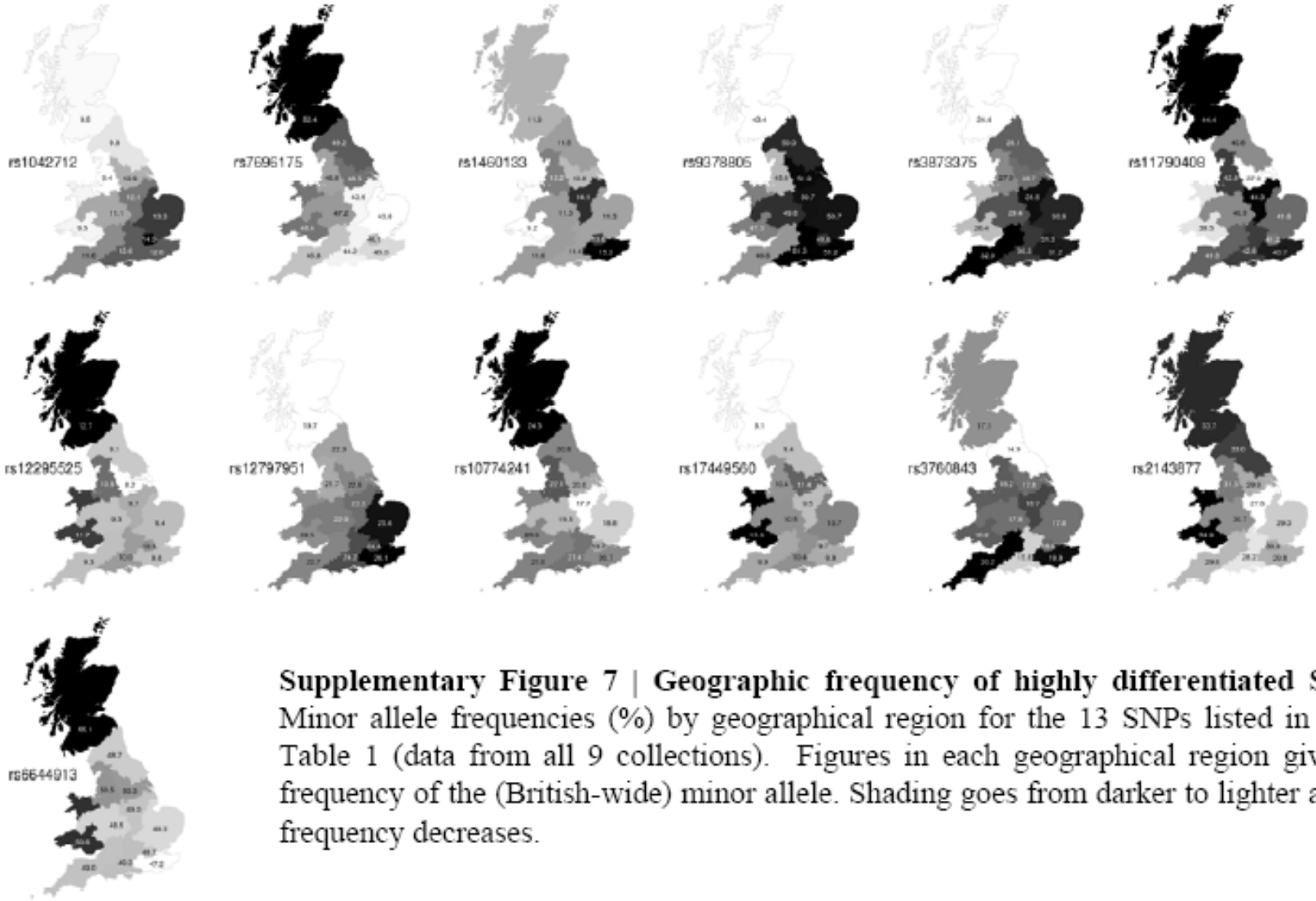
Nature and *Nature Genetics*, 7Jun2007

Contributions of Wellcome Trust Case-Control Consortium to GWA Studies

- 24 independent association signals in six diseases
- Value of shared controls
- Improved methods for genotype calling
- Importance of quality control and review of unprocessed genotyping data
- New methods for imputing genotypes across platforms
- Improved power estimates

Nature and *Nature Genetics*, 7Jun2007

Allele Frequency Differences Across Britain



Supplementary Figure 7 | Geographic frequency of highly differentiated SNPs. Minor allele frequencies (%) by geographical region for the 13 SNPs listed in Main Table 1 (data from all 9 collections). Figures in each geographical region give the frequency of the (British-wide) minor allele. Shading goes from darker to lighter as this frequency decreases.

WTCCC, Nature 2007, Supplementary Fig 7

Diseases and Traits with Published GWA Studies (n = 55, 6/3/08)

- Macular Degeneration
- Exfoliation Glaucoma

- Lung Cancer
- Prostate Cancer
- Breast Cancer
- Colorectal Cancer
- Neuroblastoma
- Melanoma

- Crohn's Disease
- Celiac Disease
- Gallstones
- Irritable Bowel Syndrome

- QT Prolongation
- Coronary Disease
- Stroke
- Hypertension
- Atrial Fibrillation/Flutter
- Coronary Spasm
- Lipids and Lipoproteins

- Parkinson Disease
- Amyotrophic Lat. Sclerosis
- Multiple Sclerosis
- Prog. Supranuclear Palsy
- MS Interferon-β Response
- Alzheimer's Disease
- Cognitive Ability
- Memory
- Restless Legs Syndrome
- Nicotine Dependence
- Methamphetamine Depend.
- Neuroticism
- Schizophrenia
- Bipolar Disorder
- Family Chaos

- Rheumatoid Arthritis
- Systemic Lupus Erythematosus
- Psoriasis

- HIV Viral Setpoint
- Childhood Asthma

- Type 1 Diabetes
- Type 2 Diabetes
- Diabetic Nephropathy
- End-Stage Renal Disease
- Obesity, BMI, Waist, IR
- Height
- Osteoporosis
- Osteoarthritis

- F-Cell Distribution
- Fetal Hgb Levels
- C-Reactive Protein
- 18 groups of Framingham Traits
- Pigmentation
- Uric Acid Levels
- Recombination Rate
- Protein Levels

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 studained in samples ower to

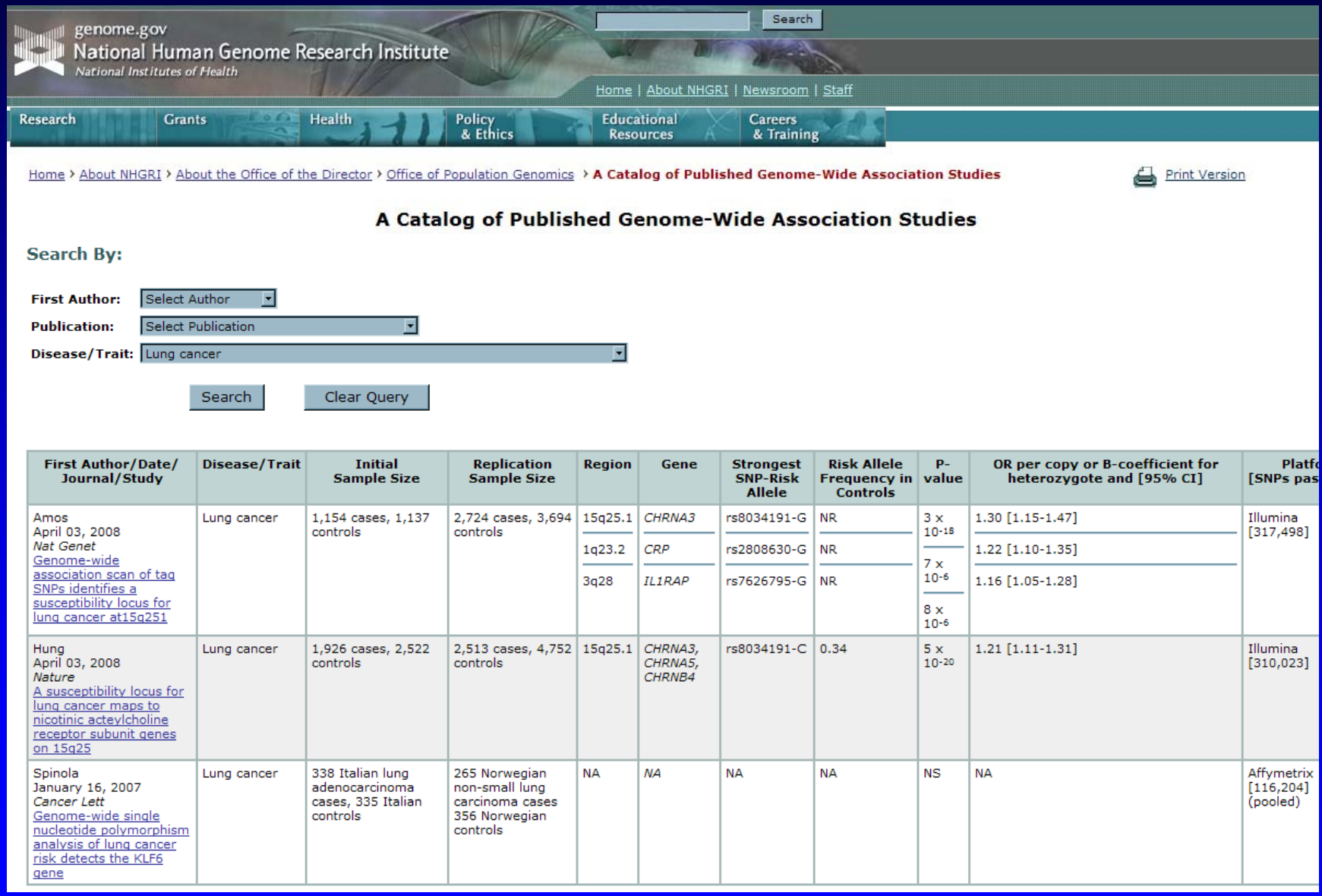
and in this issue of the *Journal*, coronary artery disease (reported by Samani et al., pages 443–453). These genomewide association studies have been able to examine interpatient differences in inherited genetic variability at an unprecedented level of resolution, thanks to the development of microarrays,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

NHGRI Catalog of GWA Studies:
http://www.genome.gov/gwastudies/



genome.gov
National Human Genome Research Institute
National Institutes of Health

Search

Home | About NHGRI | Newsroom | Staff

Research | Grants | Health | Policy & Ethics | Educational Resources | Careers & Training

Home › About NHGRI › About the Office of the Director › Office of Population Genomics › **A Catalog of Published Genome-Wide Association Studies**

Print Version

A Catalog of Published Genome-Wide Association Studies

Search By:

First Author: Select Author

Publication: Select Publication

Disease/Trait: Lung cancer

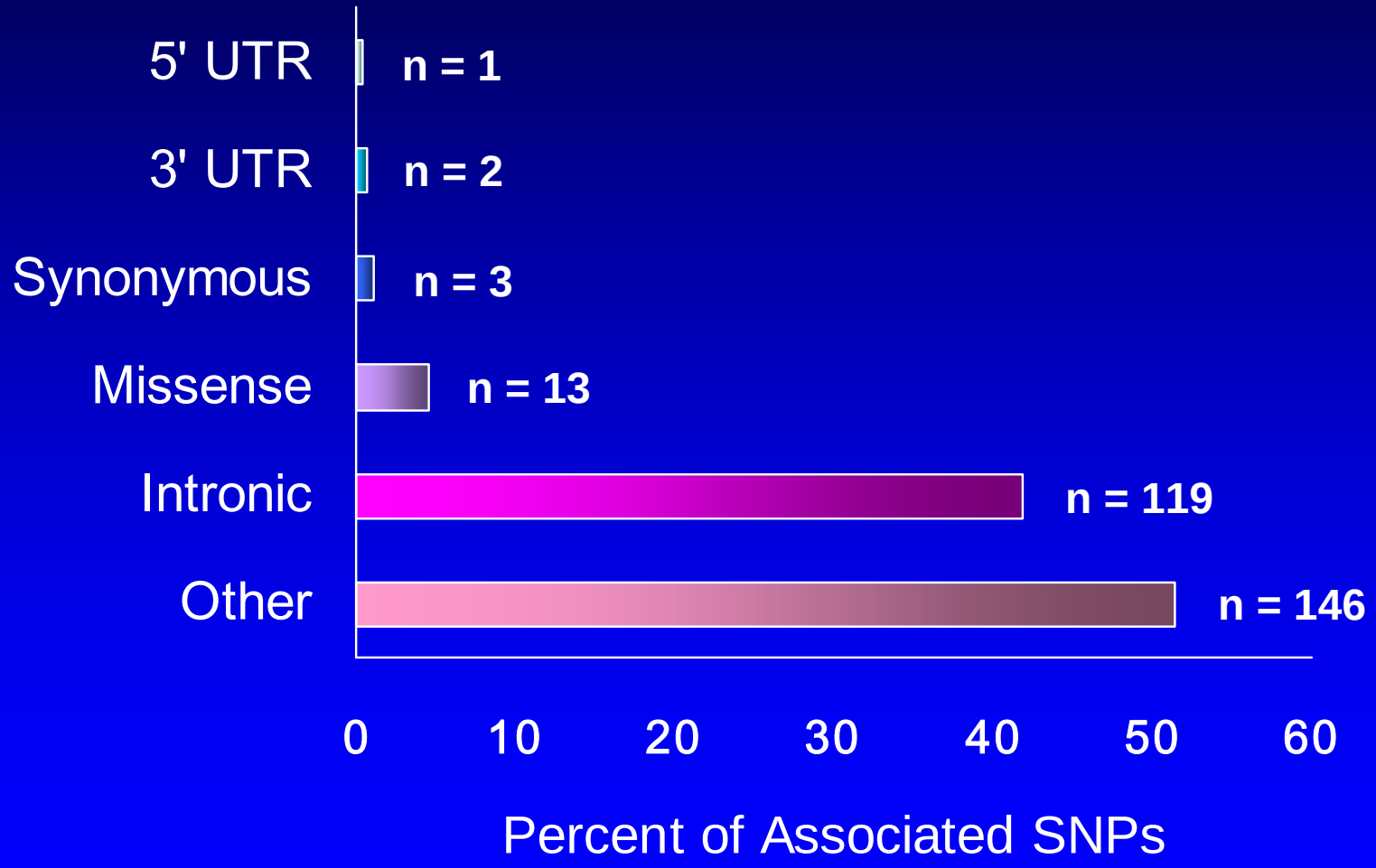
Search | Clear Query

First Author/Date/ Journal/Study	Disease/Trait	Initial Sample Size	Replication Sample Size	Region	Gene	Strongest SNP-Risk Allele	Risk Allele Frequency in Controls	P-value	OR per copy or B-coefficient for heterozygote and [95% CI]	Platfo [SNPs pas
Amos April 03, 2008 *Nat Genet* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at15q251	Lung cancer	1,154 cases, 1,137 controls	2,724 cases, 3,694 controls	15q25.1 1q23.2 3q28	*CHRNA3* *CRP* *IL1RAP*	rs8034191-G rs2808630-G rs7626795-G	NR NR NR	3 x 10^{-18} 7 x 10^{-6} 8 x 10^{-6}	1.30 [1.15-1.47] 1.22 [1.10-1.35] 1.16 [1.05-1.28]	Illumina [317,498]
Hung April 03, 2008 *Nature* A susceptibility locus for lung cancer maps to nicotinic acteylcholine receptor subunit genes on 15q25	Lung cancer	1,926 cases, 2,522 controls	2,513 cases, 4,752 controls	15q25.1	*CHRNA3, CHRNA5, CHRNB4*	rs8034191-C	0.34	5 x 10^{-20}	1.21 [1.11-1.31]	Illumina [310,023]
Spinola January 16, 2007 *Cancer Lett* Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene	Lung cancer	338 Italian lung adenocarcinoma cases, 335 Italian controls	265 Norwegian non-small lung carcinoma cases 356 Norwegian controls	NA	*NA*	NA	NA	NS	NA	Affymetrix [116,204] (pooled)

NHGRI Catalog of GWA Studies:
http://www.genome.gov/gwastudies/

- First author/Data/Journal/Study
- Disease/Trait
- Initial Sample Size
- Replication Sample Size
- Region
- Gene
- Strongest SNP – Risk Allele
- Risk Allele Frequency in Controls
- P-value
- OR per copy [95% CI]
- Platform and SNPs passing QC

Functional Classification of 284 SNPs Associated with Complex Traits



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

Lessons Learned from Initial GWA Studies

Signals in Previously Unsuspected Genes	
Macular Degeneration	*CFH*
Coronary Disease	*CDKN2A/2B*
Childhood Asthma	*ORMDL3*
Type II Diabetes	*CDKAL1*
QT interval prolongation	*NOS1AP*

Lessons Learned from Initial GWA Studies

Signals in Previously Unsuspected Genes	
Macular Degeneration	*CFH*
Coronary Disease	*CDKN2A/2B*
Childhood Asthma	*ORMDL3*
Type II Diabetes	*CDKAL1*
QT interval prolongation	*NOS1AP*
Signals in Gene "Deserts"	
Prostate Cancer	8q24
Crohn's Disease	5p13.1, 1q31.2, 10p21

Lessons Learned from Initial GWA Studies

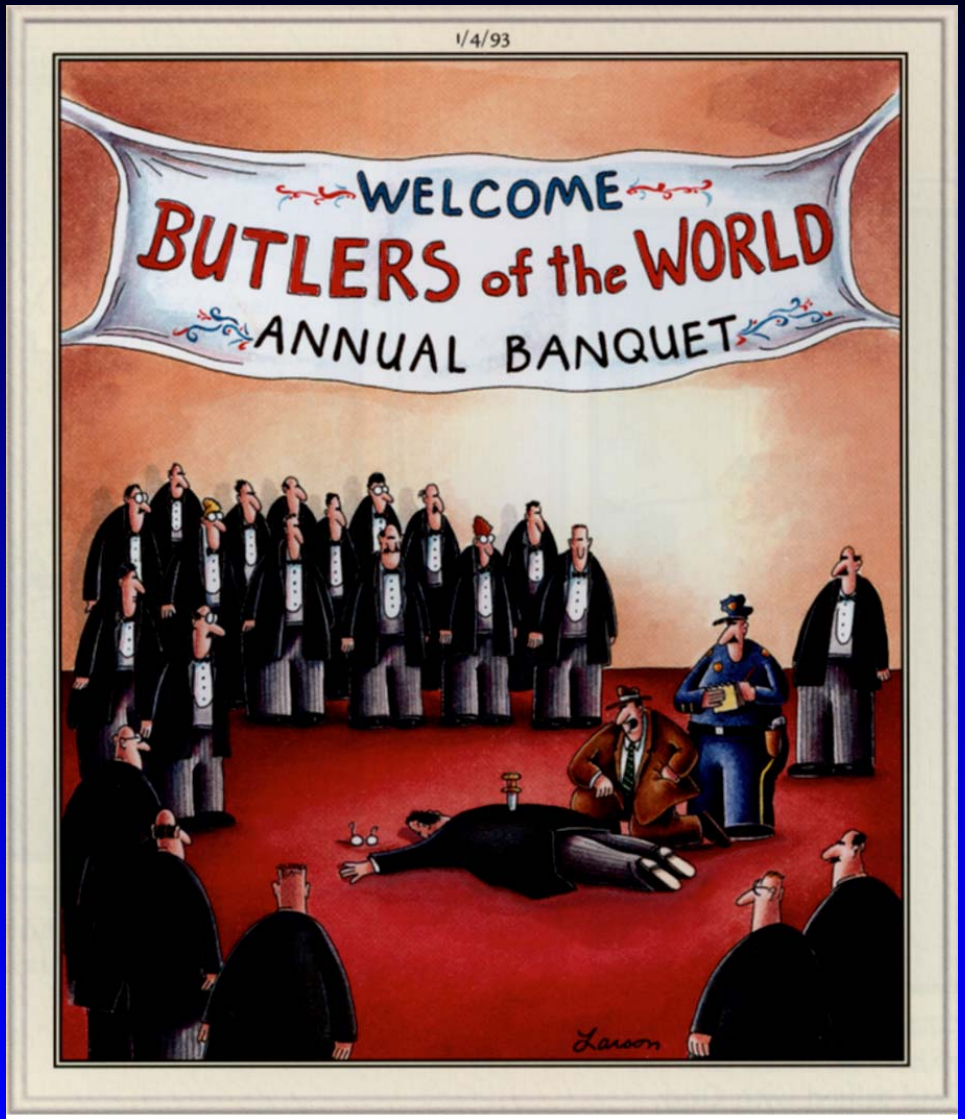
Signals in Previously Unsuspected Genes	
Macular Degeneration	*CFH*
Coronary Disease	*CDKN2A/2B*
Childhood Asthma	*ORMDL3*
Type II Diabetes	*CDKAL1*
QT interval prolongation	*NOS1AP*
Signals in Gene "Deserts"	
Prostate Cancer	8q24
Crohn's Disease	5p13.1, 1q31.2, 10p21
Signals in Common	
Diabetes, CHD, Melanoma, Frailty	*CDKN2A/2B*
Prostate, Breast, Colorectal Cancer	8q24 region
Crohn's Disease, Psoriasis	*IL23R*
Crohn's Disease, T1DM	*PTPN2*
Rheumatoid Arthritis, T1DM	*PTPN22*

Unique Aspects of GWA Studies

- Permit examination of inherited genetic variability at unprecedented level of resolution
- Permit "agnostic" genome-wide evaluation
- Once genome measured, can be related to any trait
- Most robust associations in GWA studies have not been with genes previously suspected of association with 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."

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



"God, Collings, I hate to start a Monday with a case like this."

Larson, G. *The Complete Far Side*. 2003.

FEATURE

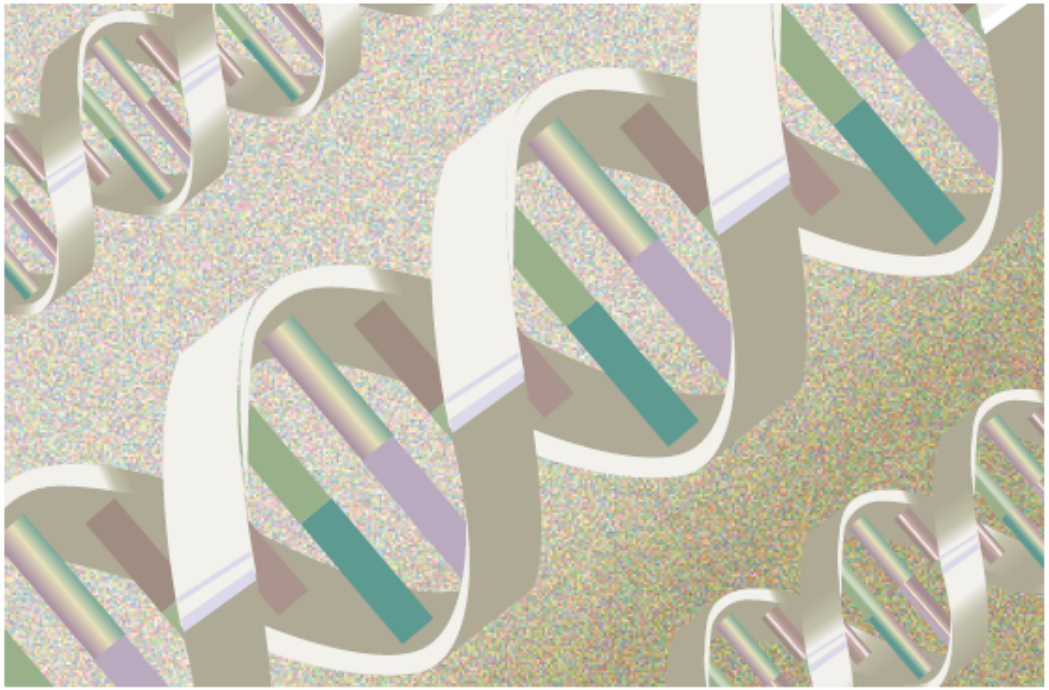
Replicating genotype–phenotype associations

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 S, Manolio T, et al, *Nature* 2007; 447:655-660.

Replication, Replication, Replication

Initial study: Sufficient description to permit replication

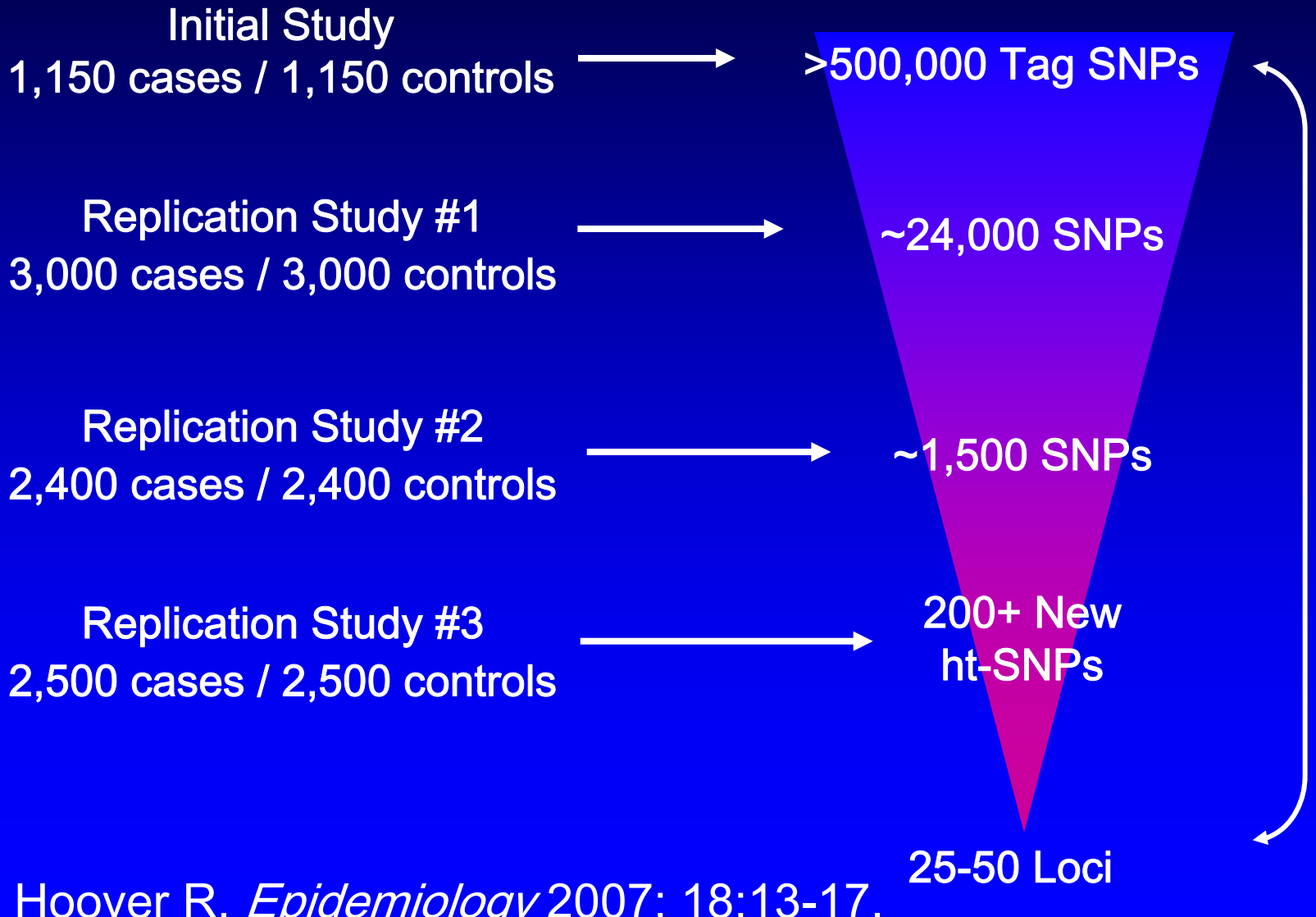
- Sources of cases and controls
- Participation rates and flow chart of selection
- Methods for assessing affected status
- Standard “Table 1” including rates of missing data
- Assessment of population heterogeneity
- Genotyping methods and QC metrics

Replication study:

- Similar population, similar phenotype
- Same genetic model, same SNP, same direction
- Adequately powered to detect postulated effect

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

Replication Strategy for Prostate Cancer Study in CGEMS



Hoover R, *Epidemiology* 2007; 18:13-17.

Replication Strategy in Easton Breast Cancer Study

Stage	Cases	Controls	SNPs
1	408	400	266,722

Easton et al, *Nature* 2007; 447:1087-93.

Replication Strategy in Easton Breast Cancer Study

Stage	Cases	Controls	SNPs
1	408	400	266,722
2	3,990	3,916	13,023

Easton et al, *Nature* 2007; 447:1087-93.

Replication Strategy in Easton Breast Cancer Study

Stage	Cases	Controls	SNPs
1	408	400	266,722
2	3,990	3,916	13,023
3	23,734	23,639	31

Easton et al, *Nature* 2007; 447:1087-93.

Replication Strategy in Easton Breast Cancer Study

Stage	Cases	Controls	SNPs
1	408	400	266,722
2	3,990	3,916	13,023
3	23,734	23,639	31
Final			6

- ABCFS
- BCST
- COPS
- GENICA
- HBCS
- HBCP
- TBCS
- KConFab/AOCS
- KBCP
- LUMCBCS
- MCBCS
- MCCS
- MEC-W
- MEC-J
- NHS
- PBCS
- RBCS
- SASBAC
- SEARCH2
- SEARCH3
- SBCP
- SBCS
- CNIOBCS
- USRT

Easton et al, *Nature* 2007; 447:1087-93.

Larson, G. *The Complete Far Side*. 2003.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869

Thomas et al, *Nat Genet* 2008; 40:310-15.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

Thomas et al, *Nat Genet* 2008; 40:310-15.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

* Selected for $p < 0.068$

Thomas et al, *Nat Genet* 2008; 40:310-15.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

* Selected for $p < 0.068$

SNP	Gene	Stage 1+2 P-value
rs4962416	*MSMB*	7×10^{-13}
rs10896449	11q13	2×10^{-9}
rs10993994	*CTBP2*	2×10^{-7}
rs10486567	*JAZF1*	2×10^{-6}

Thomas et al, *Nat Genet* 2008; 40:310-15.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

* Selected for p < 0.068

SNP	Gene	Stage 1+2 P-value	Initial Rank
rs4962416	*MSMB*	7×10^{-13}	24,223
rs10896449	11q13	2×10^{-9}	
rs10993994	*CTBP2*	2×10^{-7}	
rs10486567	*JAZF1*	2×10^{-6}	

Thomas et al, *Nat Genet* 2008; 40:310-15.

Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

* Selected for p < 0.068

SNP	Gene	Stage 1+2 P-value	Initial Rank
rs4962416	*MSMB*	7×10^{-13}	24,223
rs10896449	11q13	2×10^{-9}	2,439
rs10993994	*CTBP2*	2×10^{-7}	319
rs10486567	*JAZF1*	2×10^{-6}	24,407

Thomas et al, *Nat Genet* 2008; 40:310-15.

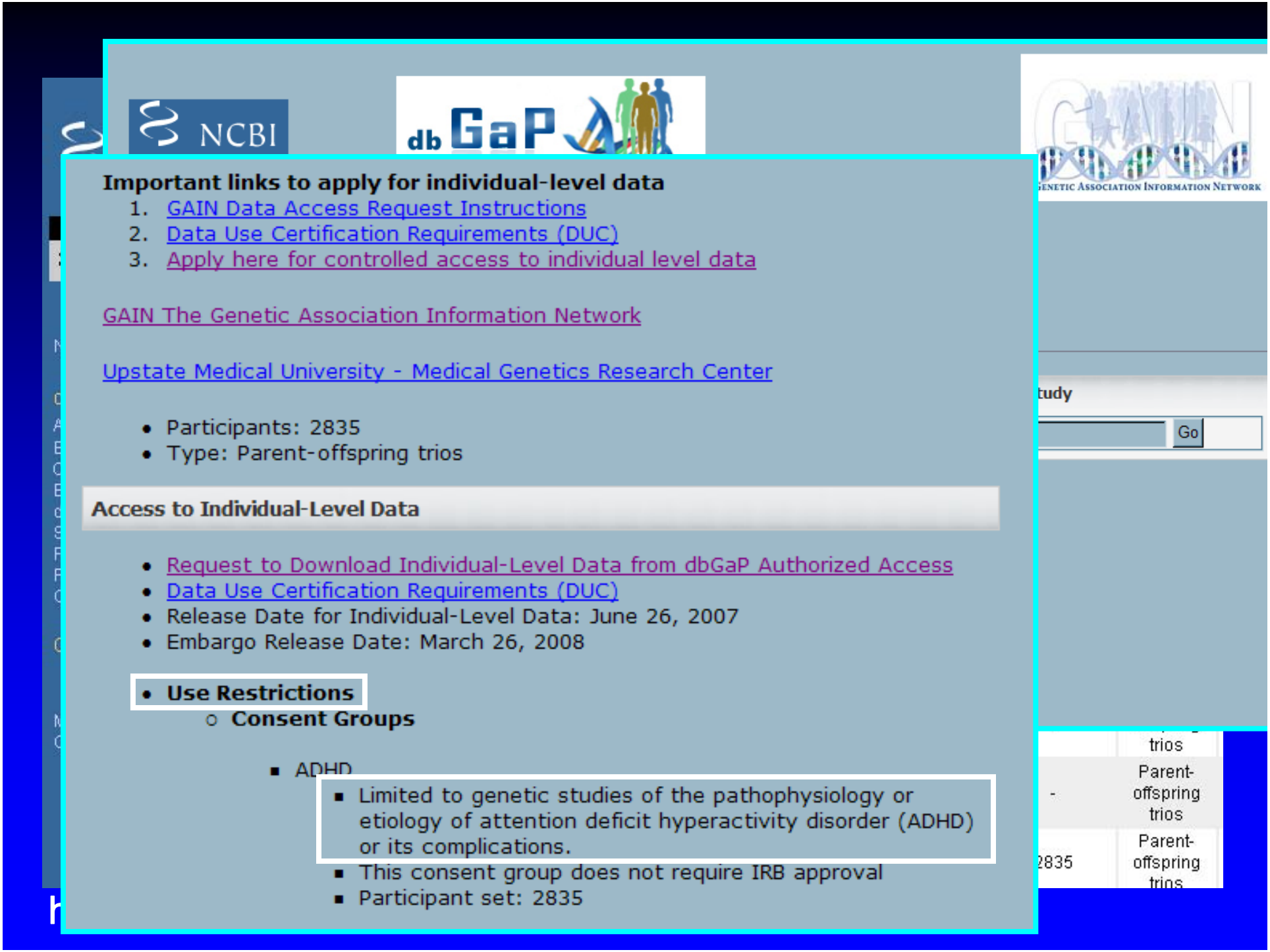
Replication Strategy in CGEMS Prostate Cancer Study

Stage	Cases	Controls	SNPs
1	1,172	1,157	527,869
2	3,941	3,964	26,958*

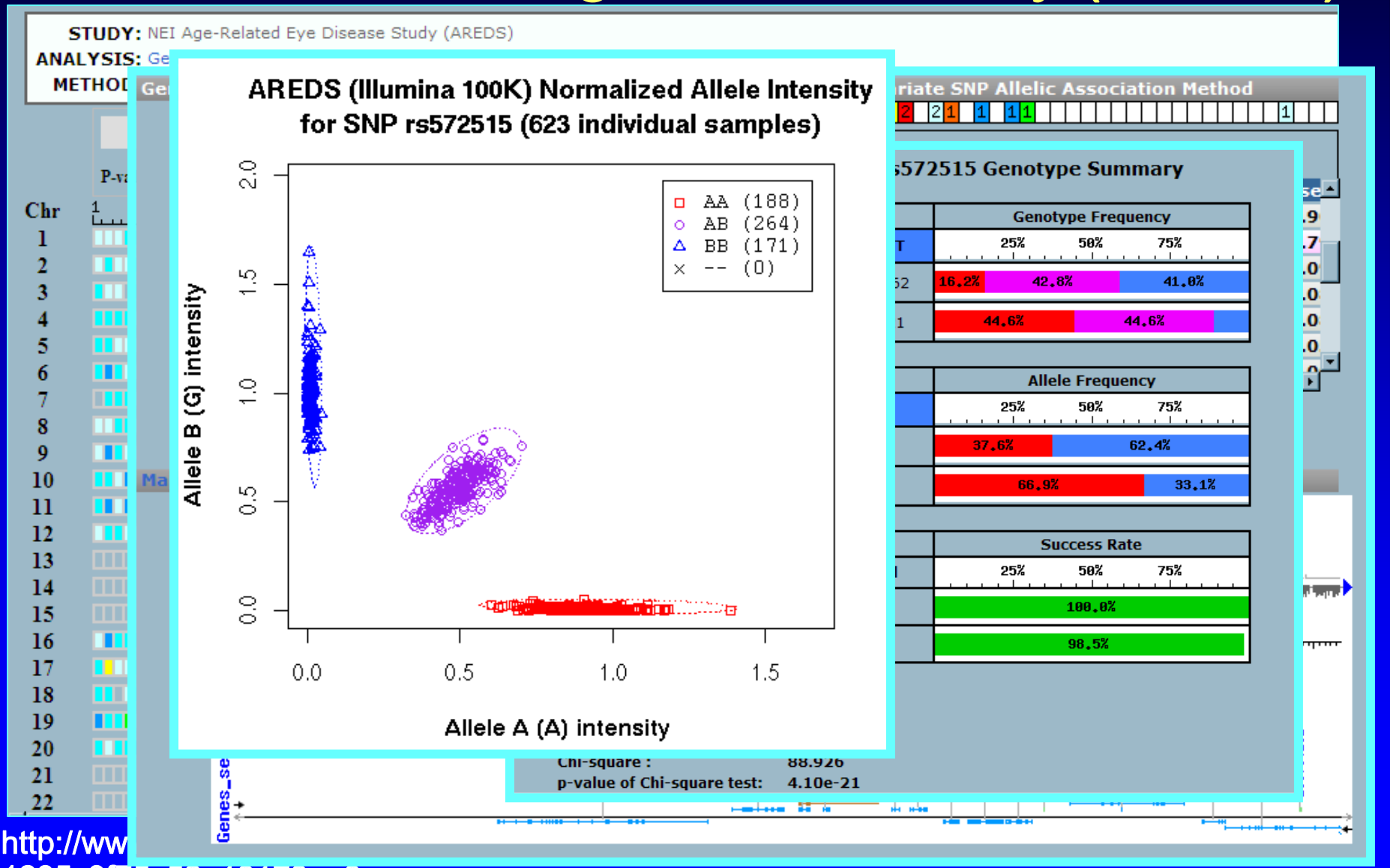
* Selected for p < 0.068

SNP	Gene	Stage 1+2 P-value	Initial Rank	Initial P-value
rs4962416	*MSMB*	7×10^{-13}	24,223	0.042
rs10896449	11q13	2×10^{-9}	2,439	0.004
rs10993994	*CTBP2*	2×10^{-7}	319	4×10^{-4}
rs10486567	*JAZF1*	2×10^{-6}	24,407	0.042

Thomas et al, *Nat Genet* 2008; 40:310-15.


NCBI
dbGaP
GAIN
Genetic Association Information Network
Important links to apply for individual-level data
1. GAIN Data Access Request Instructions
2. Data Use Certification Requirements (DUC)
3. Apply here for controlled access to individual level data
GAIN The Genetic Association Information Network
Upstate Medical University - Medical Genetics Research Center
• Participants: 2835
• Type: Parent-offspring trios
Access to Individual-Level Data
• Request to Download Individual-Level Data from dbGaP Authorized Access
• Data Use Certification Requirements (DUC)
• Release Date for Individual-Level Data: June 26, 2007
• Embargo Release Date: March 26, 2008
• Use Restrictions
○ Consent Groups
▪ ADHD
▪ Limited to genetic studies of the pathophysiology or etiology of attention deficit hyperactivity disorder (ADHD) or its complications.
▪ This consent group does not require IRB approval
▪ Participant set: 2835
tudy
Go
trios
Parent-offspring trios
2835
Parent-offspring trios

Genome-Wide Allelic Association Results, Age-Related Macular Degeneration Study (AREDS)



http://ww
1285c9f75c52e19173ae9

From Bench to Bedside and Back

- GWA studies provide unprecedented opportunity for linking laboratory and population science
- Several GWA studies have incorporated histopathologic, gene expression, or knockdown findings directly into initial report

- Knockdown of *ATG16L1*
  - Associated with Crohn's disease
  - Reduces phagocytosis of *S. typhimurium* in HeLa cells

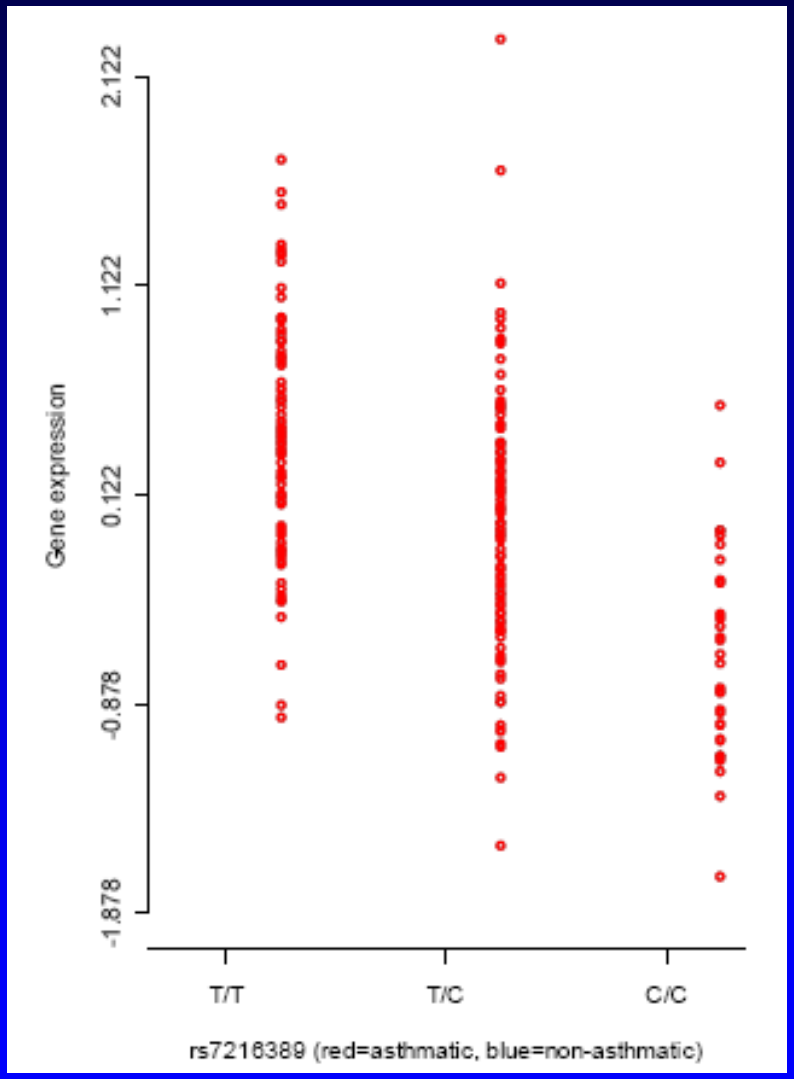
Conservation and Expression Studies: Asthma and *ORMDL3*

Moffatt et al, *Nature* 2007; 448:470-73.

Conservation and Expression Studies: Asthma and *ORMDL3*

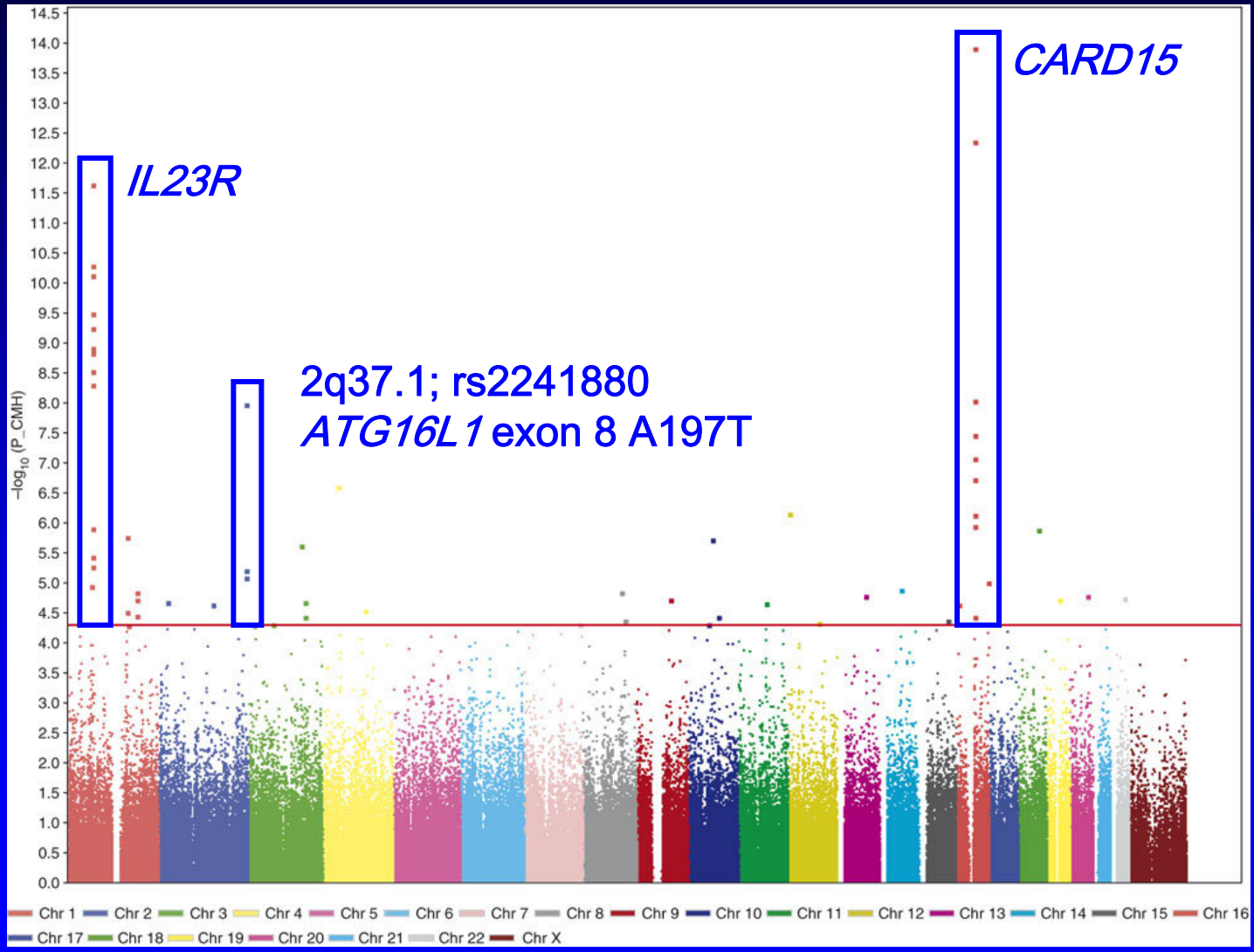
Moffatt et al, *Nature* 2007; 448:470-73.

Conservation and Expression Studies: Asthma and *ORMDL3*



Moffatt et al, *Nature* 2007; 448:470-73.

Genome-Wide Associations in Crohn's Disease



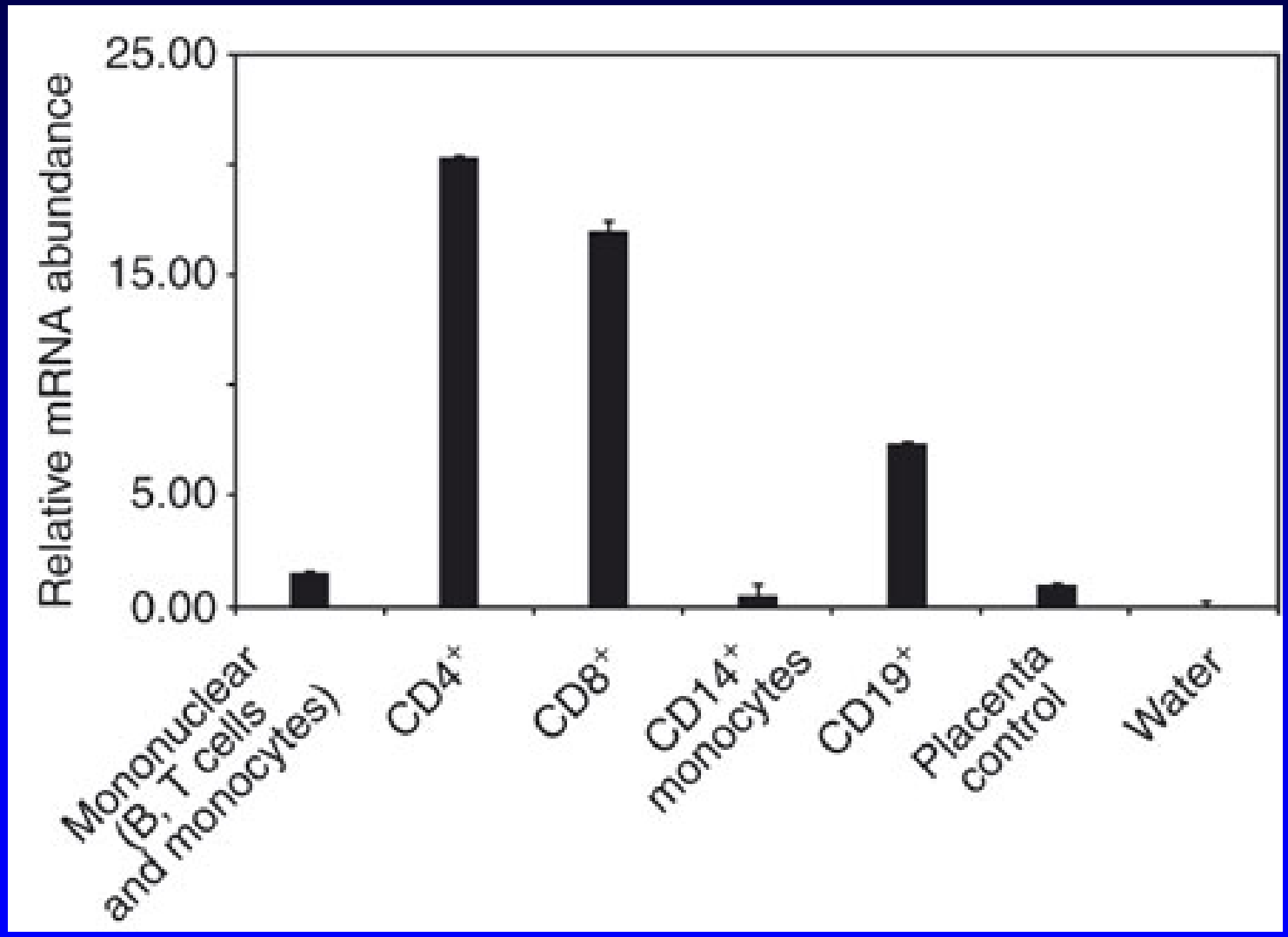
Rioux et al, *Nat Genet* 2007; 39:596-604.

Gene Expression in Crohn's Disease

- rs2241880 associated at $p < 10^{-8}$
- Nonsynonymous amino acid change in exon 8 of autophagy-related 16-like 1 (*ATG16L1*)
- Autophagy is biologic process involved in protein degradation, antigen processing, absorption of cellular organelles, initiation and regulation of inflammatory response

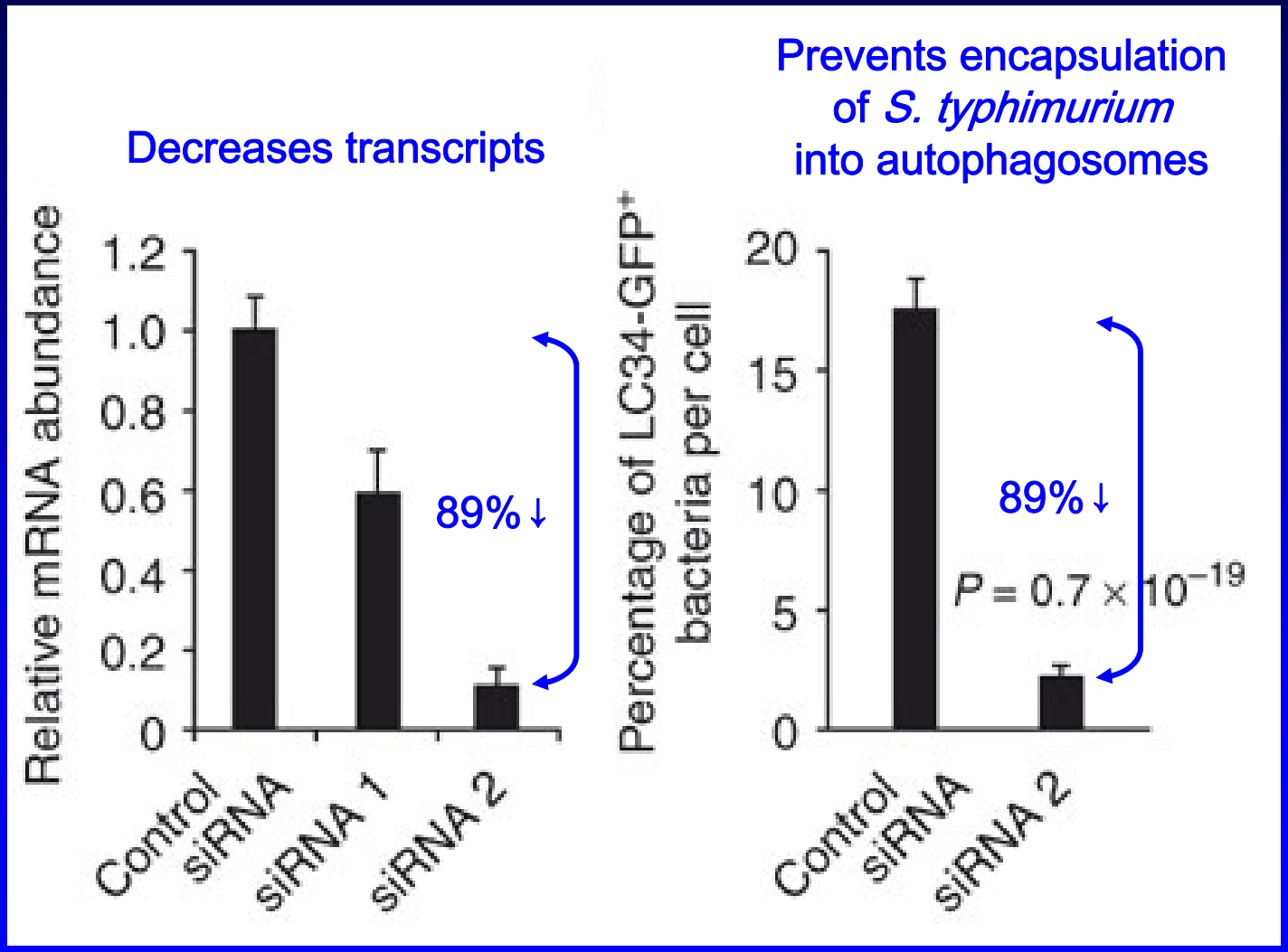
Rioux et al, *Nat Genet* 2007; 39:596-604.

Expression of *ATG16L1* in Human Primary Immune Cells



Rioux et al, *Nat Genet* 2007; 39:596-604.

Knockdown of Endogenous *ATG16L1* by siRNA 2 in HeLa Cells



Rioux et al, *Nat Genet* 2007; 39:596-604.

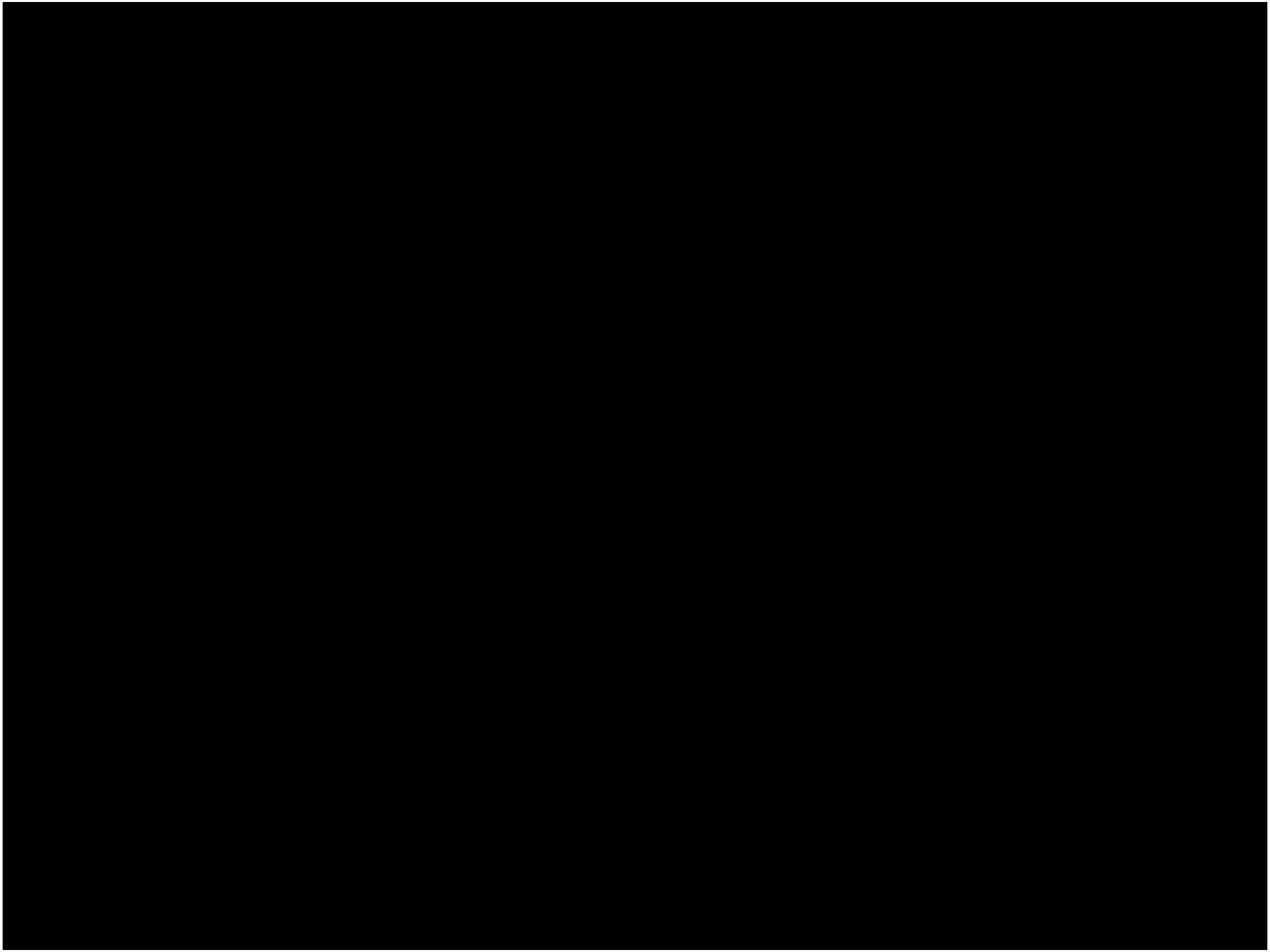
Post GWA: Finding (Putative) Causal Variants

- Narrowing region with fine mapping, sequencing
- Structure of association region: nearby genes, conservation
- Association with levels of protein product
- Co-localization with histopathologic changes
- Association with expression levels
- Knockdown, knockout studies

“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



Co-Localization of Gene Product with Histopathologic Changes

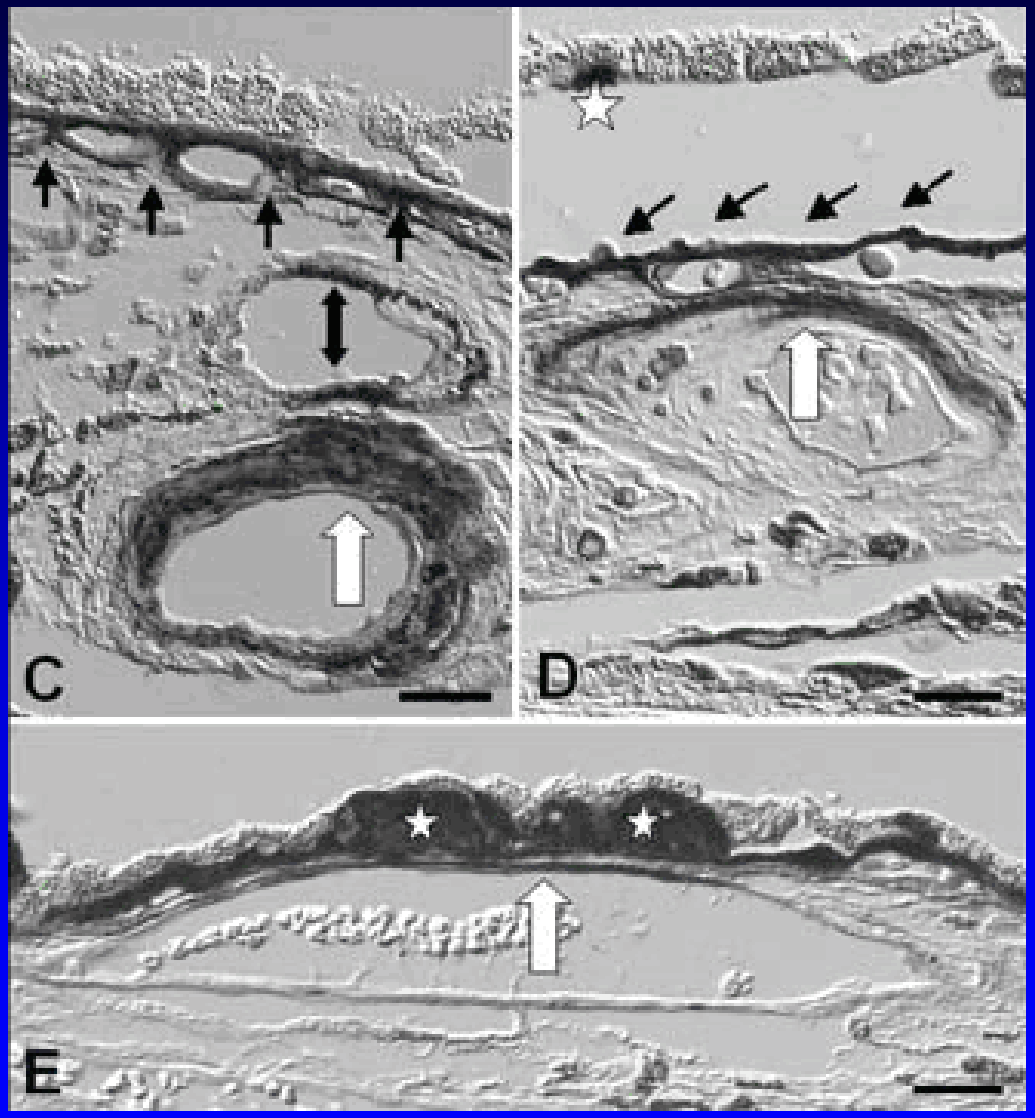
- *CFH* in retina and drusen (macular degeneration)
- *GAB2* in dystrophic neurons (Alzheimers disease)

Complement Deposition in Affected Retina

Complement deposition in Bruch's membrane (thin black arrows)

Deposition also in choroidal artery (double headed arrow, pt C) and choroidal vein (white arrow, both)

Deposition in drusen (*) as well as Bruch's membrane and choroidal vein



Klein et al, *Science* 2005; 308:385-89.

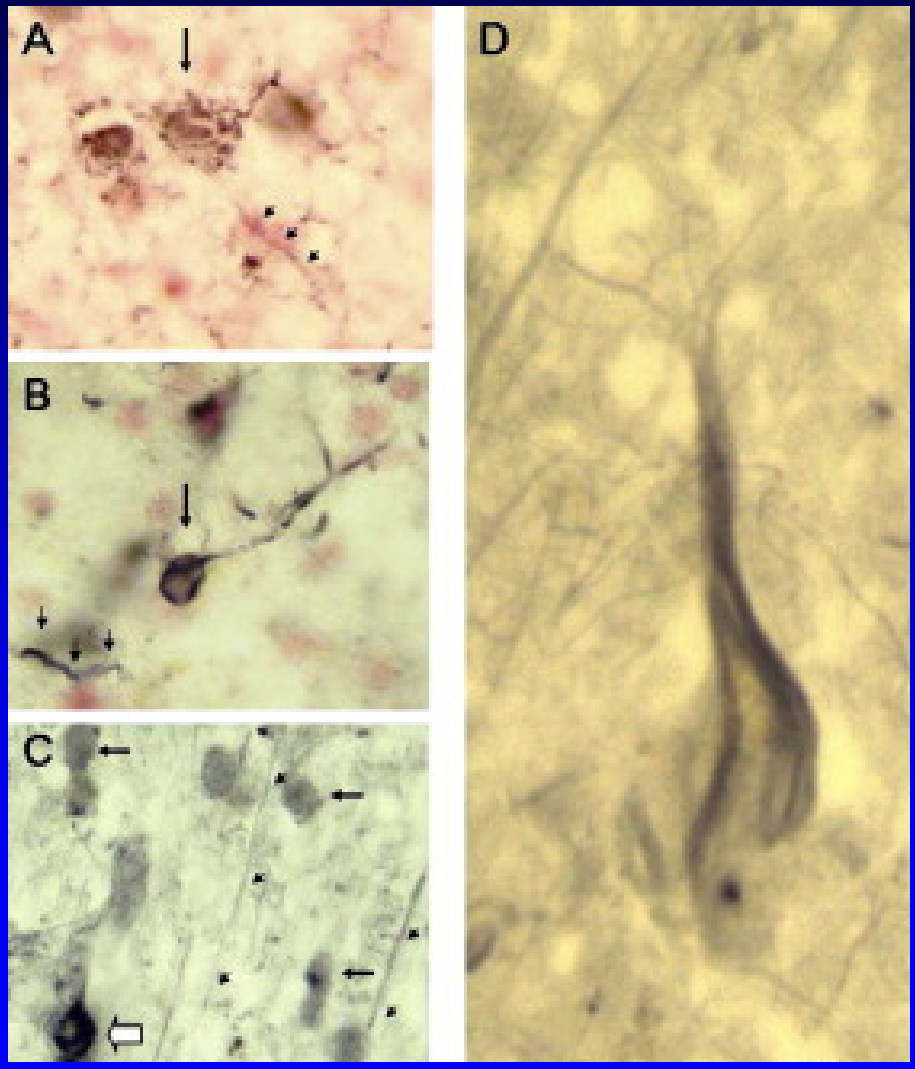
Gab2 Colocalizes with Dystrophic Neurons in LOAD Brain

Dystrophic neuron (arrow) and neurites (arrowheads)

Tangle-containing neuron (arrow), dystrophic neurites (arrowheads)

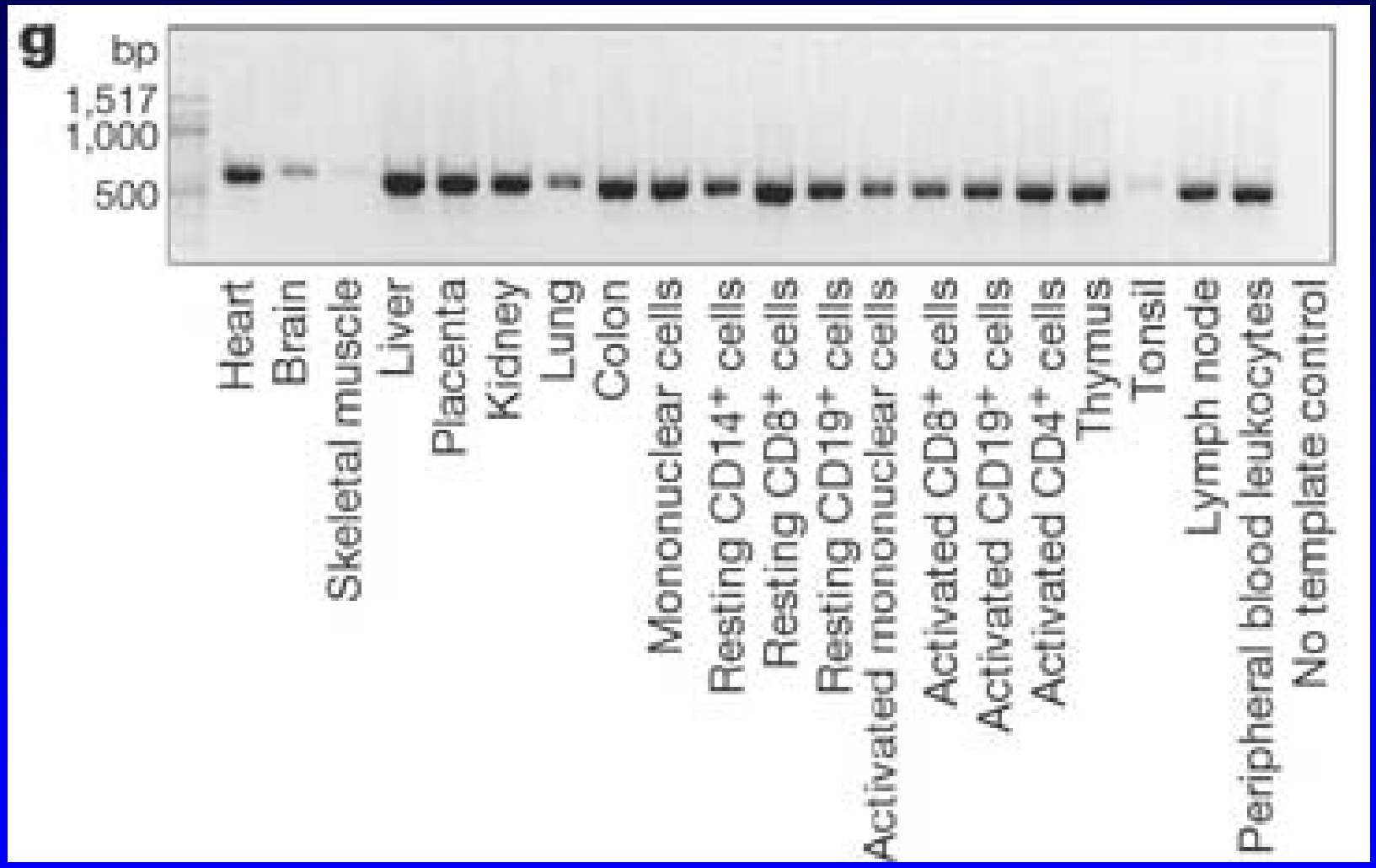
Tangle-bearing neuron (open arrow), immuno-reactive structures resembling dendrites (arrowheads)

Gab2 immunoreactive cell with flame-shaped tangle-like inclusion



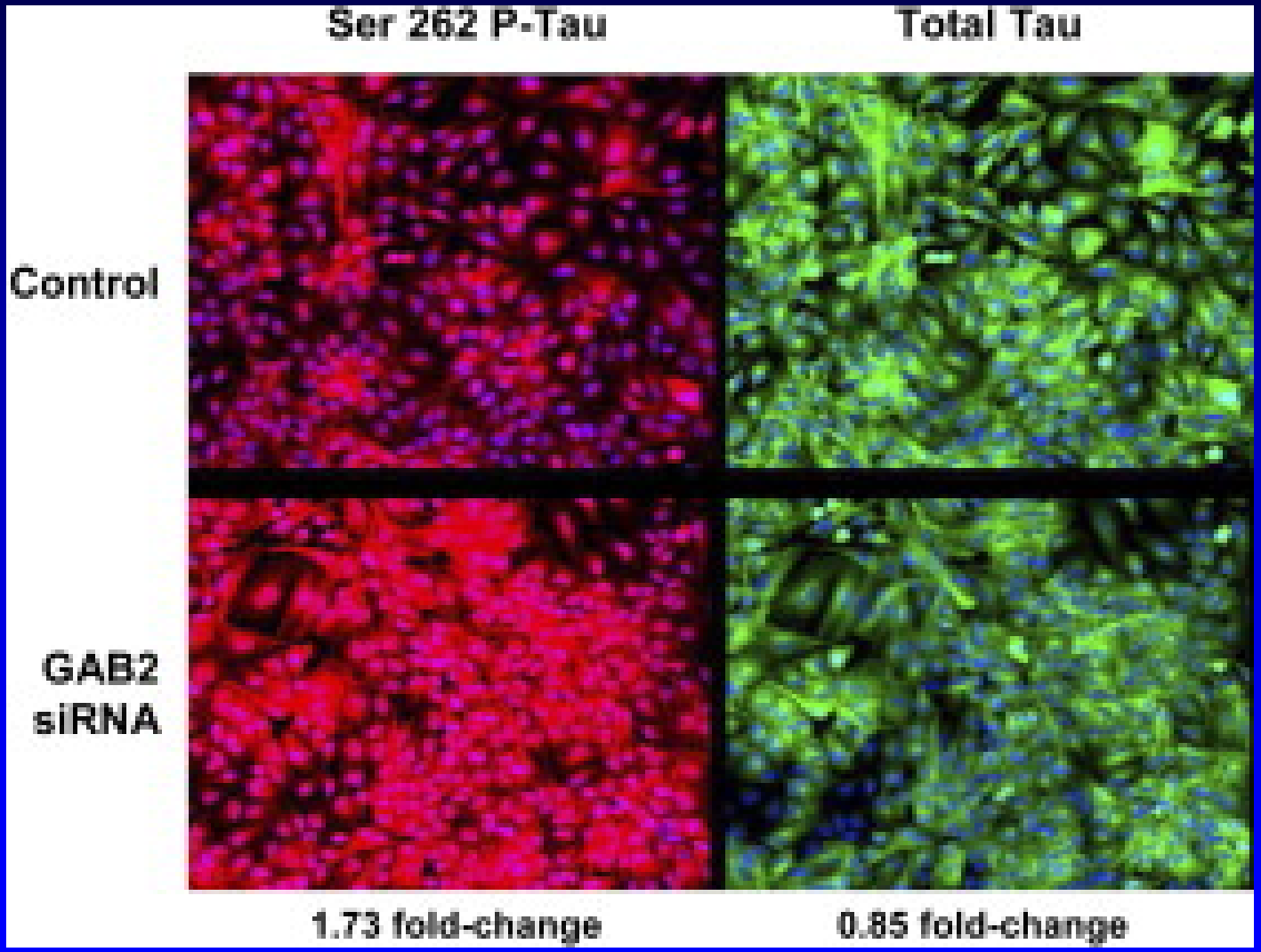
Reiman et al, *Neuron* 2007; 54:713-20.

Conservation and Expression Studies: Asthma and *ORMDL3*



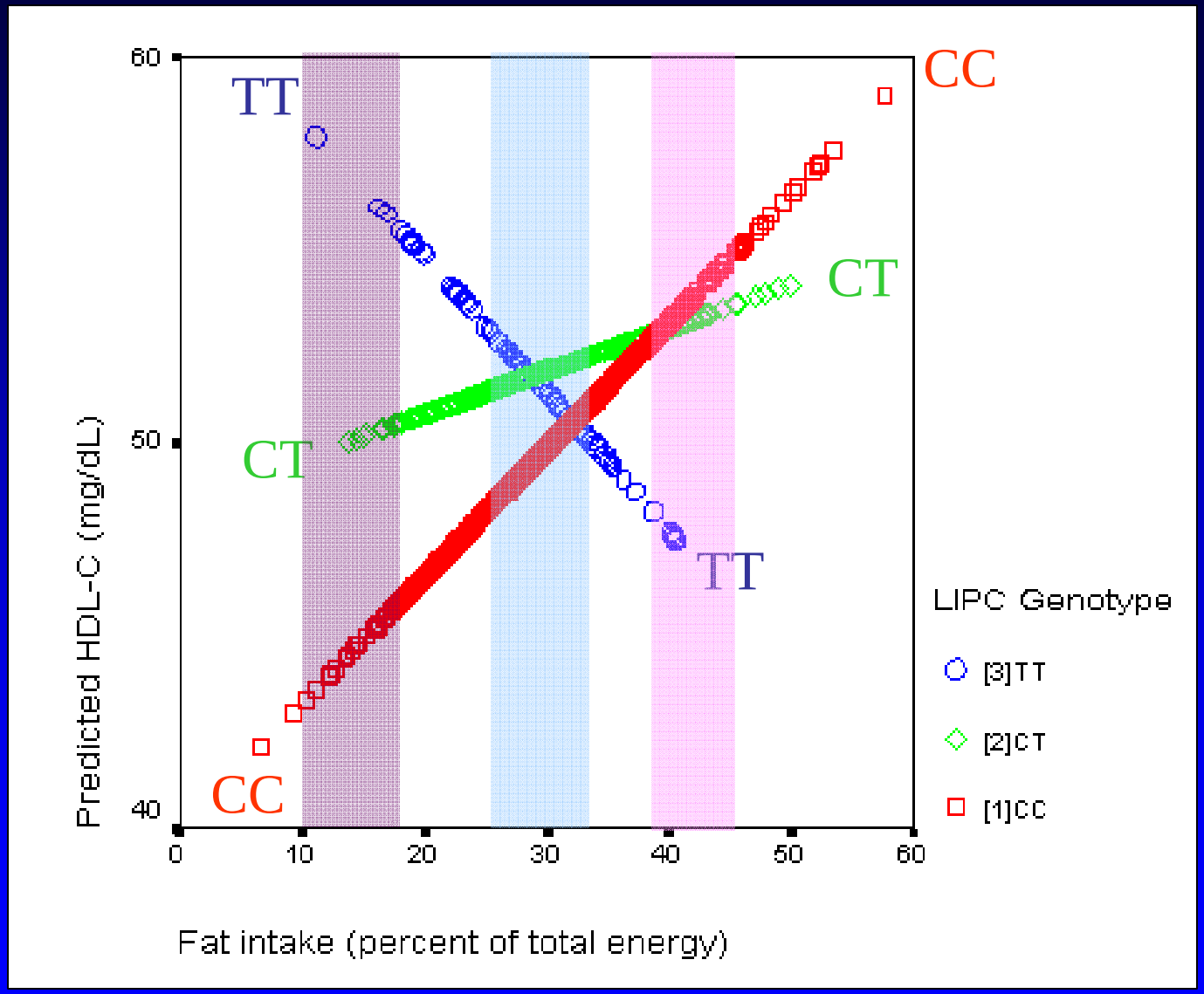
Moffatt et al, *Nature* 2007; 448:470-73.

siRNA Knockdown of *GAB2* Increases Tau Phosphylation without Increasing Total Tau



Reiman et al, *Neuron* 2007; 54:713-20.

Interaction: Is *LIPC* Genotype Related to HDL-C?



Ordovas et al, *Circulation* 2002; 106:2315-2321.

Challenges in Studying Gene-Environment Interactions

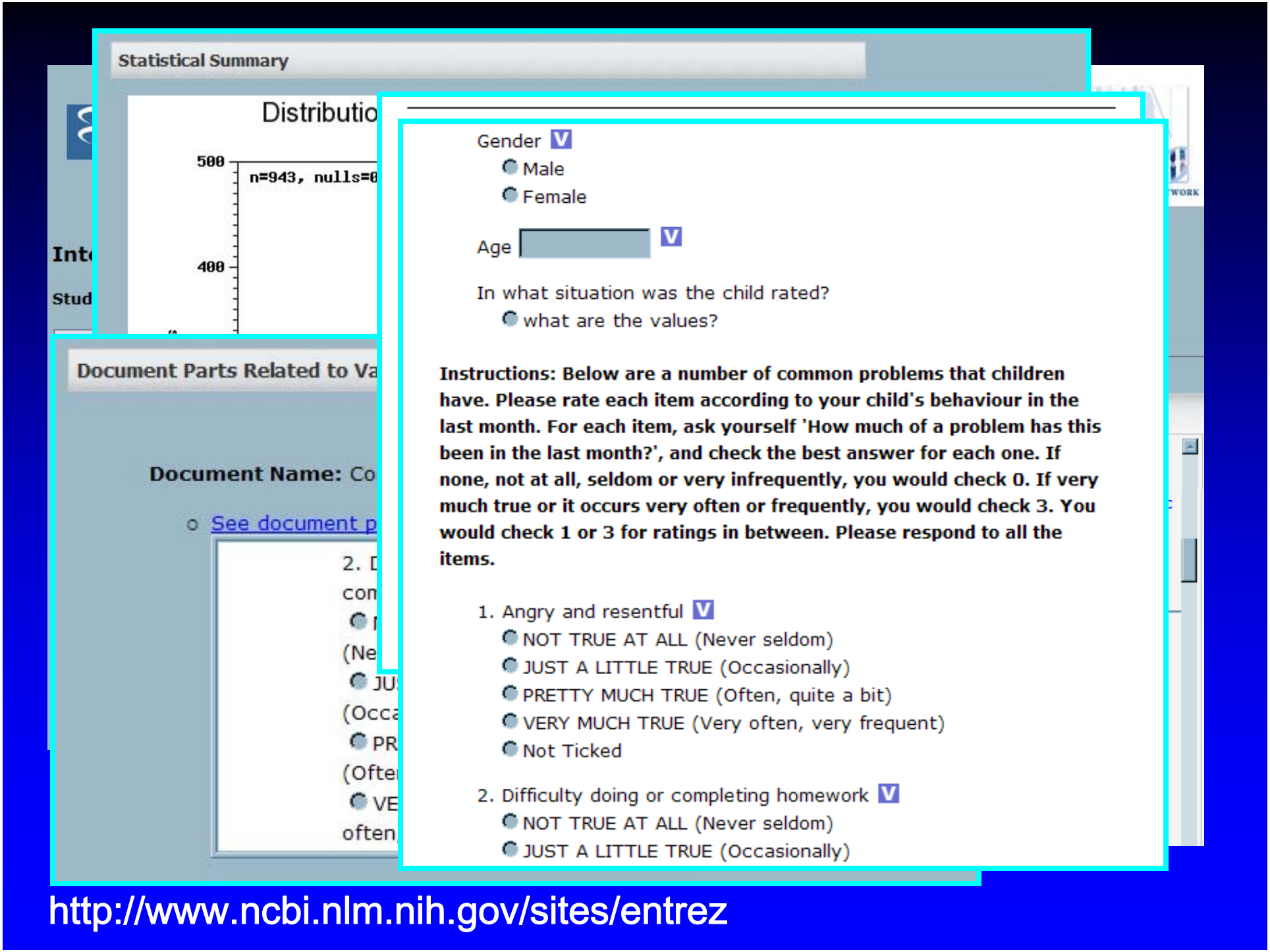
Challenge	Genes	Environment
Ease of measure	Pretty easy	Often hard
Variability over time	Low/none	High
Recall bias	None	Possible
Temporal relation to disease	Easy	Hard

Key Points: Genetic Association Studies

- Candidate gene studies enormously prone to spurious associations
- GWA presents new paradigm, is unconstrained by current imperfect understanding of genome structure and function
- Initial findings astoundingly positive
- Most are skimming surface of what could be learned
- GWA beginning to be applied to cohort studies
- Very little work in genetic association in clinical trials and treatment response

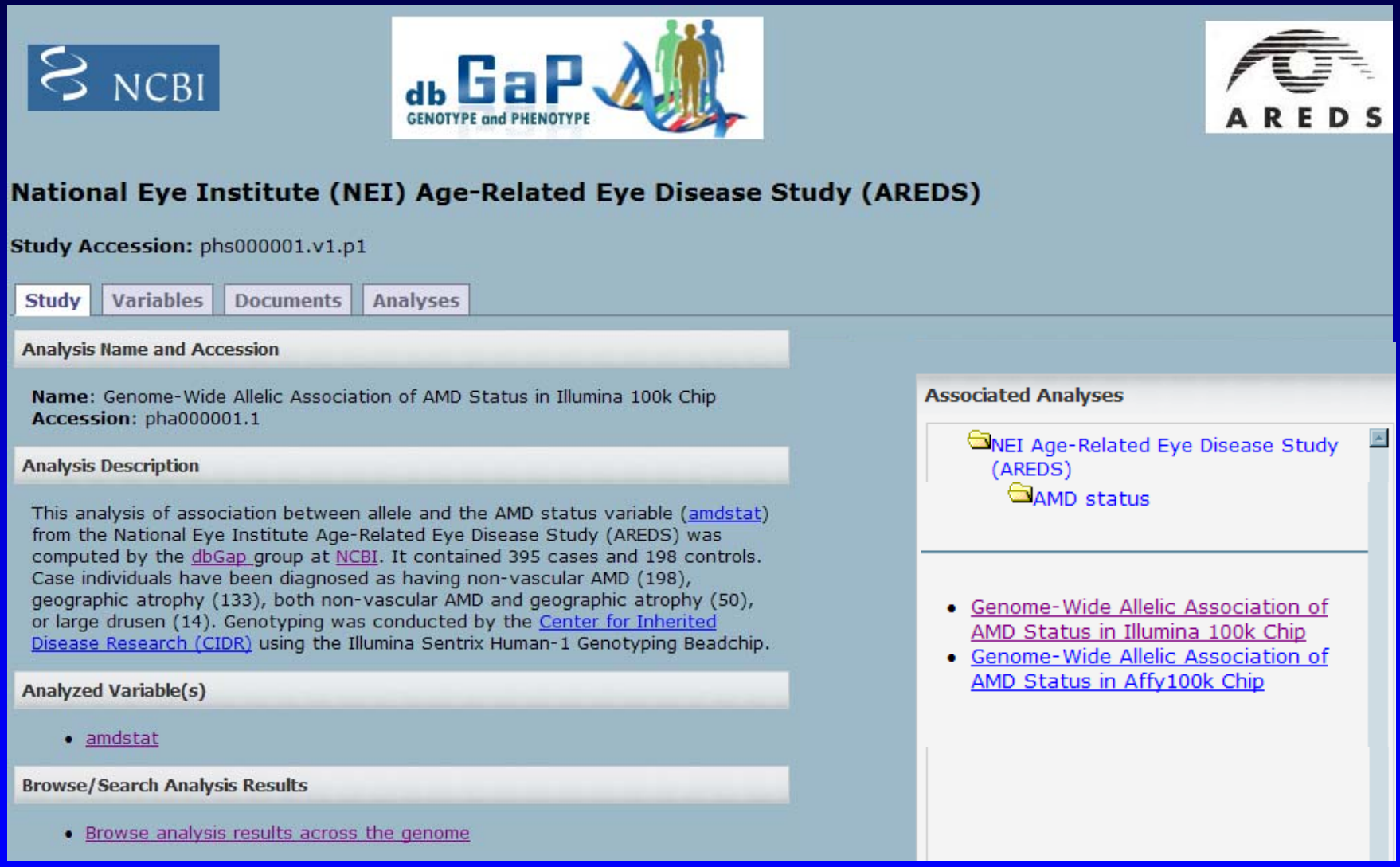
Ways of Dealing with Multiple Testing

- Bonferroni correction: most common, typically $p < 10^{-7}$ or 10^{-8}
- False discovery rate: proportion of significant associations that are actually false positives
- False positive report probability: probability that the null hypothesis is true, given a statistically significant finding
- Replication, replication, replication



http://www.ncbi.nlm.nih.gov/sites/entrez

Genome-Wide Allelic Association Results, Age-Related Macular Degeneration Study (AREDS)



http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/analysis.cgi?study_id=phs000001.v1.p1&phv=&phd=&pha=1&phsf=&phvf=&phdf=&phaf=1