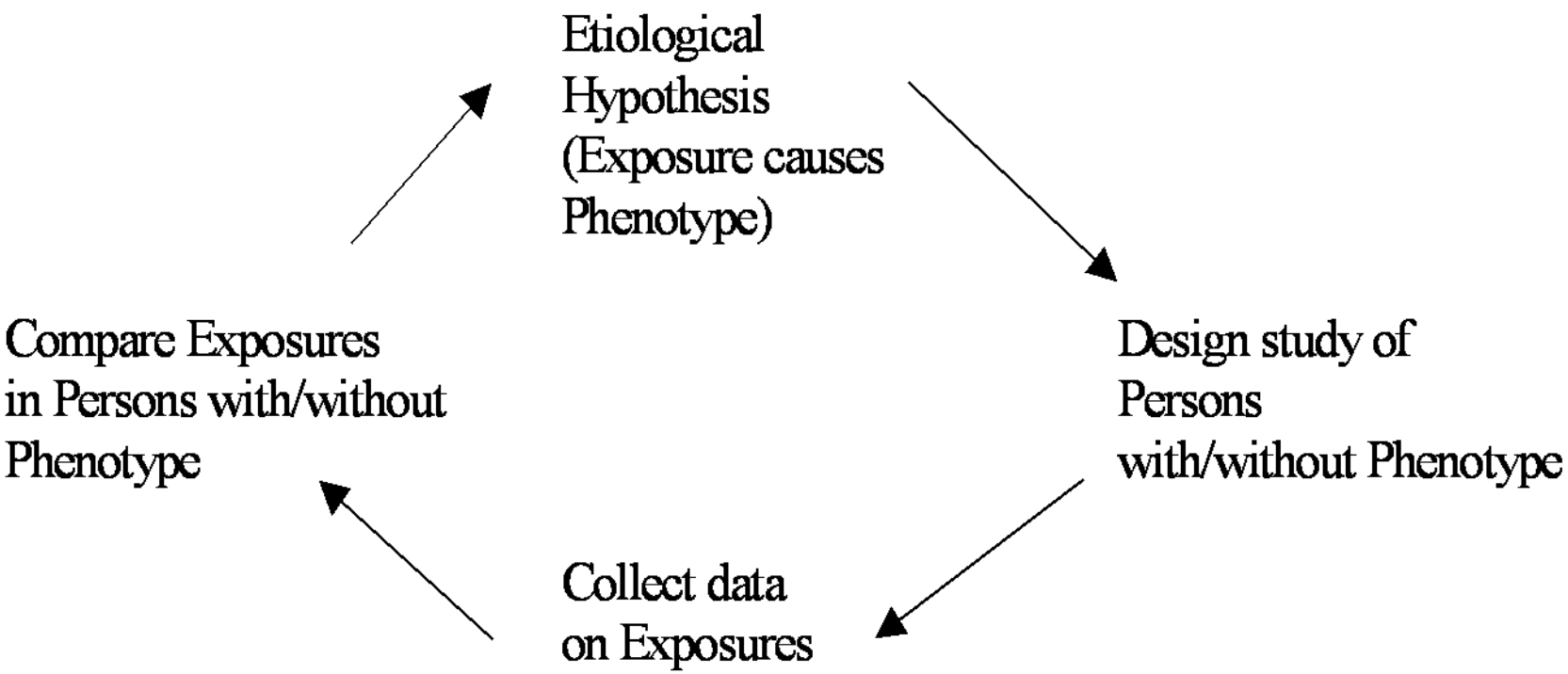
Lecture 8: Practical Applications of Epidemiologic Methods to Human Genome Research

Thomas A. Pearson, MD, PhD
University of Rochester School of Medicine
Visiting Scientist, NHGRI (9/1/07-5/30/08)

Learning Objectives

- To appreciate the use of epidemiologic methods in the practical applications of human genomic data.
  - To review the types and breadth of data required to infer that a gene variant is causal of a phenotype or disease.
  - To consider the role of basic science in establishing the functional role of a gene variant in the causation of disease.
  - To describe the role of gene variants in the efficacy and safety of pharmacotherapies.
- To be aware of current guidelines for the access to and sharing of data generated by gene association studies.

The Epidemiologic Method



U.S. Surgeon General's Criteria for Causal Association

1. Temporal relationship
2. Strength of the association
3. Dose-response relationship
4. Replication of findings
5. Biologic plausibility
6. Consideration of alternate explanations
7. Cessation of exposure
8. Consistency with other knowledge
9. Specificity of the association

* Report of the Advisory Committee to the Surgeon General, 1964

GWAS and the U.S. Surgeon General's Criteria for Causal Association

Criteria	GWAS Evidence
1. Temporal Relationship	Genome precedes disease ?Expression of gene
2. Strength of association	Multiple SNP's and other gene variants ?Composite risk of all variants known or unknown
3. Dose-response relationship	Number of alleles Recessive vs. Dominant

Case-Control Study of Any Smoking vs. Camel Smoking

	Dis	No Dis	Total	
Smoke Cigarettes	1200	800	2000	OR=2.25
Do Not Smoke	800	1200	2000	
Assume 10% Smoke Camel Cigarettes				
Smoke Camels	120	80	200	OR=1.53
Do Not Smoke Camels	1880	1920	3800	

GWAS Demonstrating Risk Per Allele for Breast Cancer*

	OR per allele	Heterozygous OR	Homozygous OR	P
FGFR2	1.26	1.23	1.63	10^{-16}
TNRC9	1.11	1.14	1.23	10^{-7}
MAP3KI	1.13	1.13	1.27	10^{-6}
LSP1	1.07	1.06	1.17	10^{-6}
H19	.96	.94	.95	10^{-6}

*Easton, DF, et al. Nature 2997; 447: 1087-1093

GWAS and the U.S. Surgeon General's Criteria for Causal Association (Cont.)

Criteria	GWAS Evidence
4. Replication of findings	Required ?Heterogeneity real or due to bias
5. Biologic plausibility	Functional studies ?Invivo studies required
6. Consideration of alternate explanations	Complex models of genetic etiology ?Attribution of all genetic risk

Possible Explanations of Heterogeneity of Results in Genetic Association Studies

- Biologic mechanisms
  - Genetic heterogeneity
  - Gene-gene interactions
  - Gene-environment interactions
- Spurious mechanisms
  - Selection bias
  - Information bias
  - Publication bias
    Confounding (population stratification)
  - Cohort, age, period (secular effects
  - Type I error

Structure of Human Genes: Potential Sites of Gene Variation

- Exons
- Introns
- Regulatory Elements
  - Promoters
  - PolyA Tail
  - Enhancers
  - Silencers
  - Locus Control Regions

GWAS to Identify Novel Breast Cancer Susceptibility Loci*

- Known breast cancer loci explain <25% of familial risk.
- Two stage study of 4398 cases and 4316 controls with replication of 30 SNP's in 21,860 cases and 22,578 controls.
- 227,876 SNP's genotyped.
- 5 novel loci related to breast cancer at $P<I0^{-7}$ explain an additional 3.6% of familial risk.
- 1792 additional SNP's associated at P<.05 with 1343 expected, suggesting many additional susceptibility alleles exist.

Easton DF, et al. Nature 2007; 447: 1087-1093

GWAS and the U.S. Surgeon General's Criteria for Causal Association (Cont.)

Criteria	GWAS Evidence
7. Cessation of exposure	Currently not possible in humans ?Intervene to reduce substrate of defective gene action or replace defective gene product
8. Consistency with other knowledge	Functional evidence Animal models including knock-outs
9. Specificity of association	One gene-one protein ?Shared association diseases with gene variants

Intervention in Children with Hutchinson-Gilford Progeria Syndrome*

- Rare disorder of accelerated aging with death from cardiovascular disease by age 13 years.
- Defect is a glycine GGC to glycine GGT in codon 608 of exon 11 of *lamin A* gene.
  - Activates a cryptic splice donor to produce an abnormal protein, Lamin A.
  - *Lamin A* or progerin cannot release from farnesyl-cysteine tether site on the nuclear membrane and alters transcription.
- Farnesyl transferase inhibition prevents anchoring of progerin in fibroblasts and in transgenic mouse models.
- Open label clinical trial of inhibition of farnesyl transferase with ABT 100 is underway.

*Merideth MA, et. al. NEJM 2008; 358: 592-604

Diseases with Common Genetic Associations Identified in GWAS

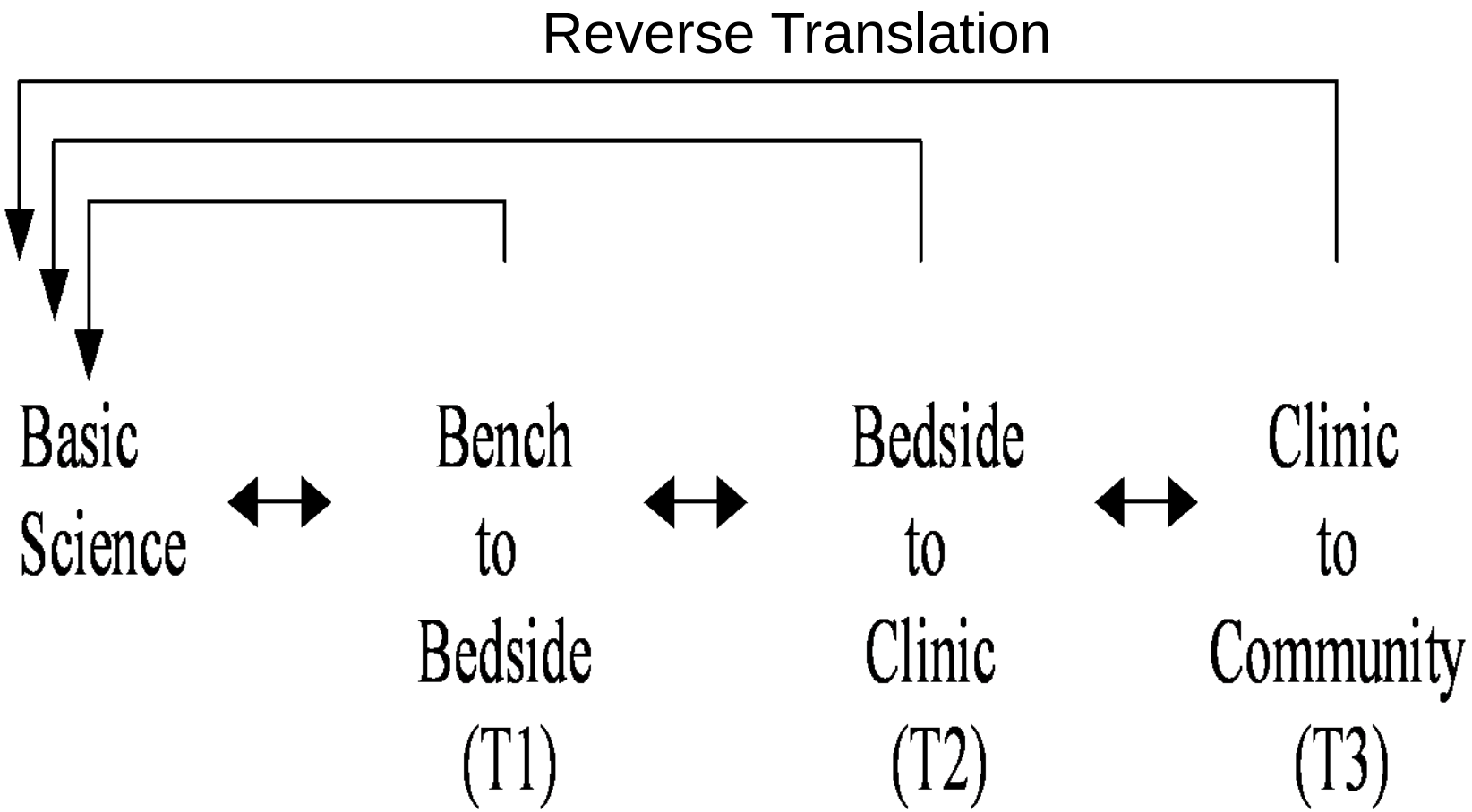
Diseases	Genes
Diabetes, CHD, Melanoma, Frailty	CDKN2A/2B
Prostate, Breast, Colorectal Cancers	8q24 region
Crohn's Dis., Psoriasis	IL23R
Crohn's Dis., T1DM	PTPN2
Rheumatoid Arthritis, T1DM	PTPN22

GWAS Identifies Gene Variant rs4430796 Which Confers Risk for Prostate Cancer and Protection from Type 2 Diabetes*

	Iceland	All Groups in Replication Study
Prostate Cancer		
Cases/Controls	1501/11289	3490/14345
OR	1.20	1.22
95% CI	1.11-1.31	1.15-1.30
P	1.4×10^{-5}	1.4×10^{-11}
Type 2 Diabetes		
Cases/Controls	1380/9840	9936/23087
OR	.86	.91
95% CI	.78-.95	.87-.94
P	.0021	2.7×10^{-7}

*Gundmundsson J, et al. Nat Gen 7/1/07

Translational Research
Reverse Translation



Sample Collection and Processing

- Obtaining samples for DNA preparation
  - Blood
  - Buccal cells
  - Serum
  - Pathology specimens
  - Other?
- Purifying and quantifying DNA
- Whole genome amplification (WGA)
- Trace individual DNAs (QC)

Courtesy S. Chanock, NCi

Contributions of GWAS to Basic Science

Genome structure and function

- Exons, introns
- Regulatory elements

Novel mechanisms of disease

Proteins as therapeutics

Drug targets

- Mass screening of small molecule inhibitors

The Genetic Etiology of Disease

Gene Variant

↓

Gene Expression

↓

Gene Product

↓

Altered Physiology

↓

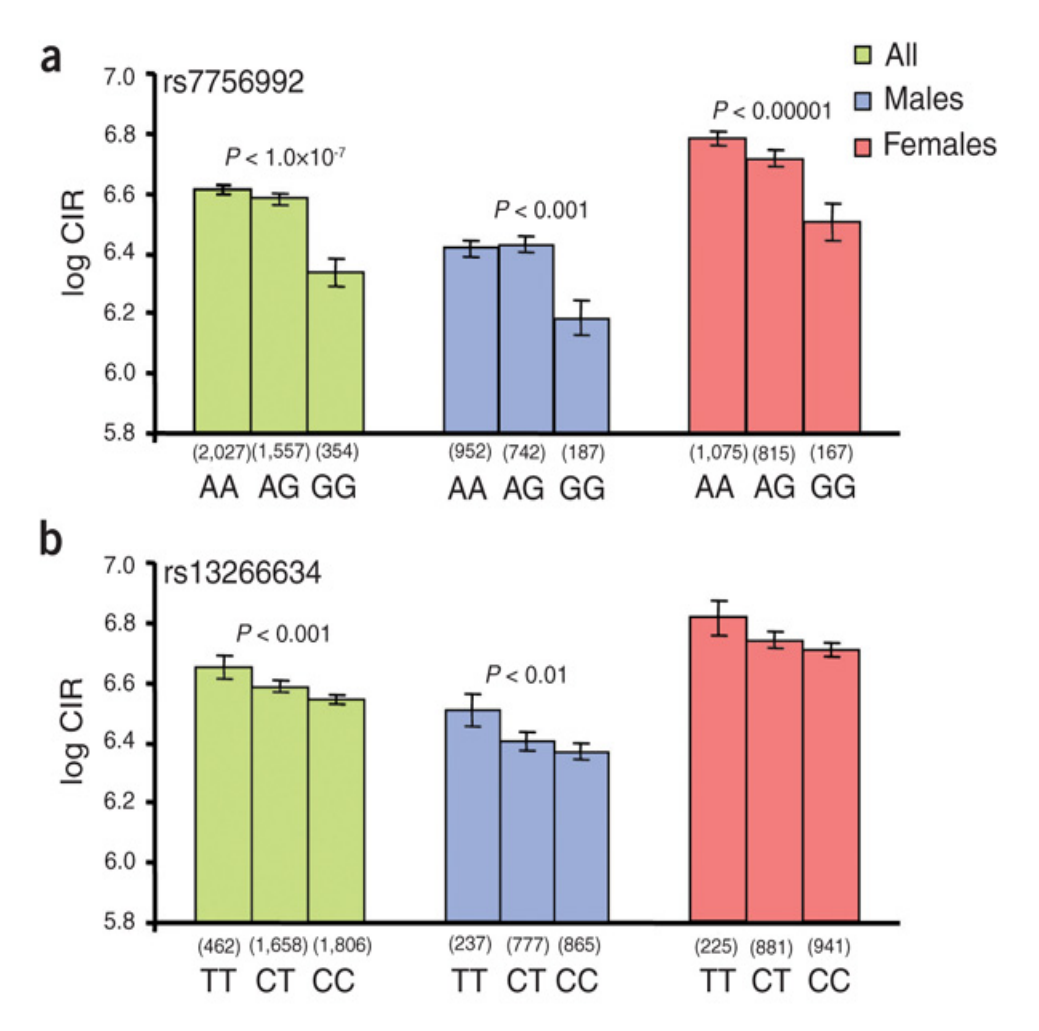
Phenotype (Disease)

Correlation of SNPs with Intermediate Phenotypes

- rs7756992 on 6p22.3 associated with type 2 diabetes (OR 1.20, $p < 8 \times 10^{-8}$), resides in intron 5 of CDK5 regulatory subunit associated protein 1-like1 (*CDKAL1*)
- rs13244434 on 8q24 also associated with T2DM: OR 1.15, $p < 4 \times 10^{-6}$
  - Nonsynonymous arginine to tryptophan change in last exon of solute carrier family 30 (zinc transporter), member 8 (*SLC30A8*)
  - Specific to pancreas and expressed in beta cells

Steinthorsdottir et al, *Nat Genet* 2007; 39:770-75.

Relationship of Diabetes-Associated SNPs with Insulin Secretion



Steinthorsdottir et al, *Nat Genet* 2007; 39:770-75.

Co-Localization of Gene Product with Histopathologic Changes

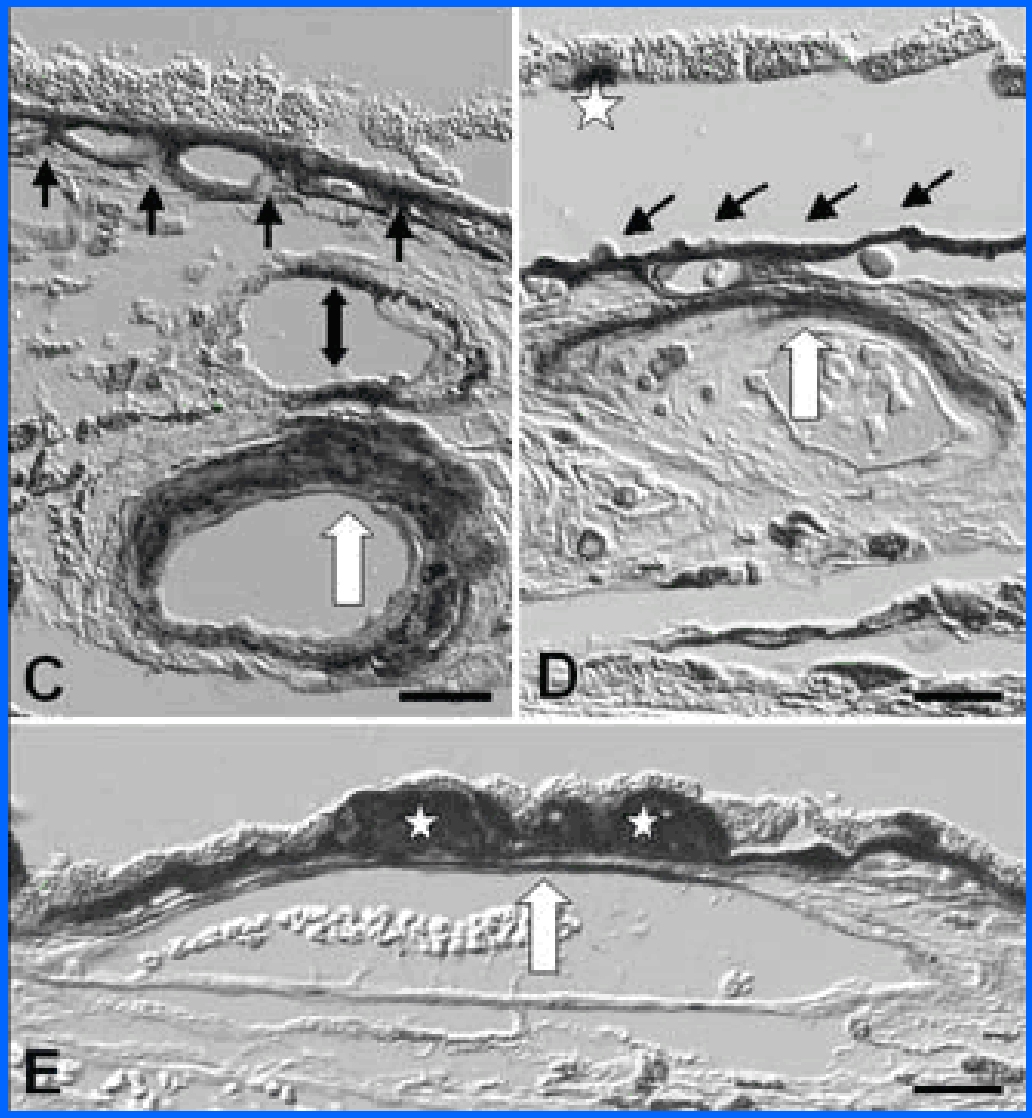
- *CFH* in retina and drusen (macular degeneration)
- *GAB2* in dystrophic neurons (Alzheimers disease)

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



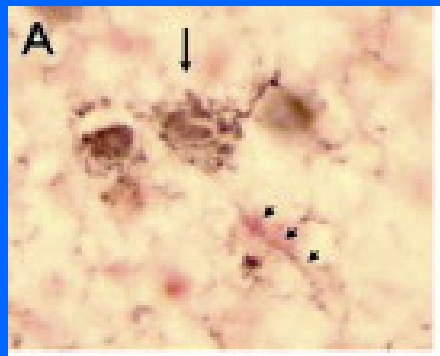
Gab2 Colocalizes with Dystrophic Neurons in LOAD Brain

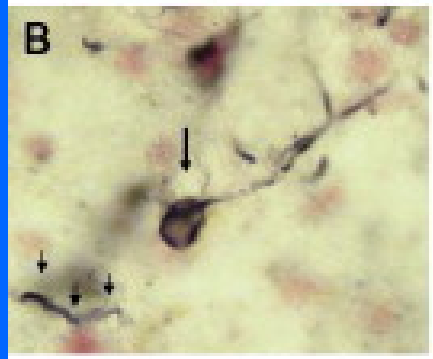
Dystrophic neuron (arrow) and neurites (arrowheads)

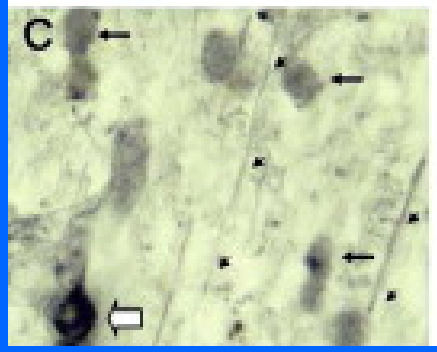
Tangle-containing neuron (arrow), dystrophic neurites (arrowheads)

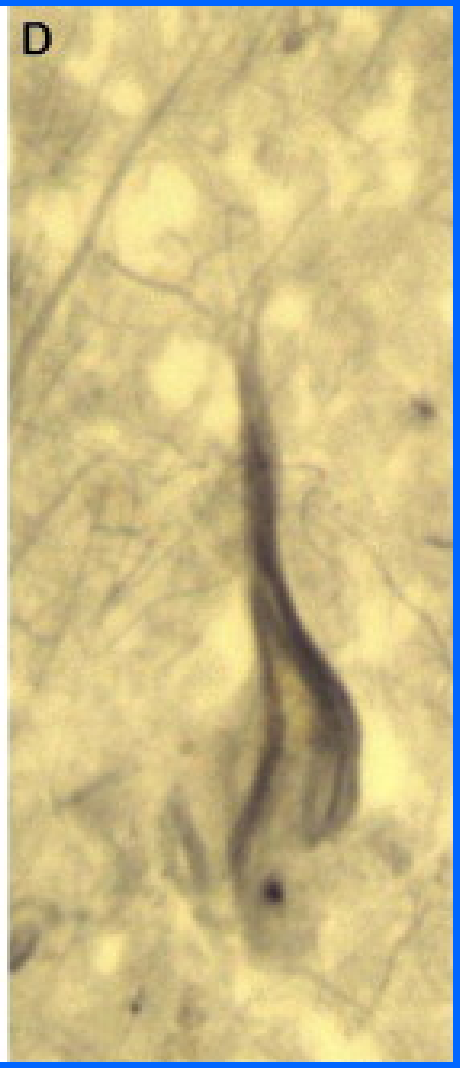
Tangle-bearing neuron (open arrow), immuno-reactive structures resembling dendrites (arrowheads)

Gab2 immunoreactive cell with flame-shaped tangle-like inclusion



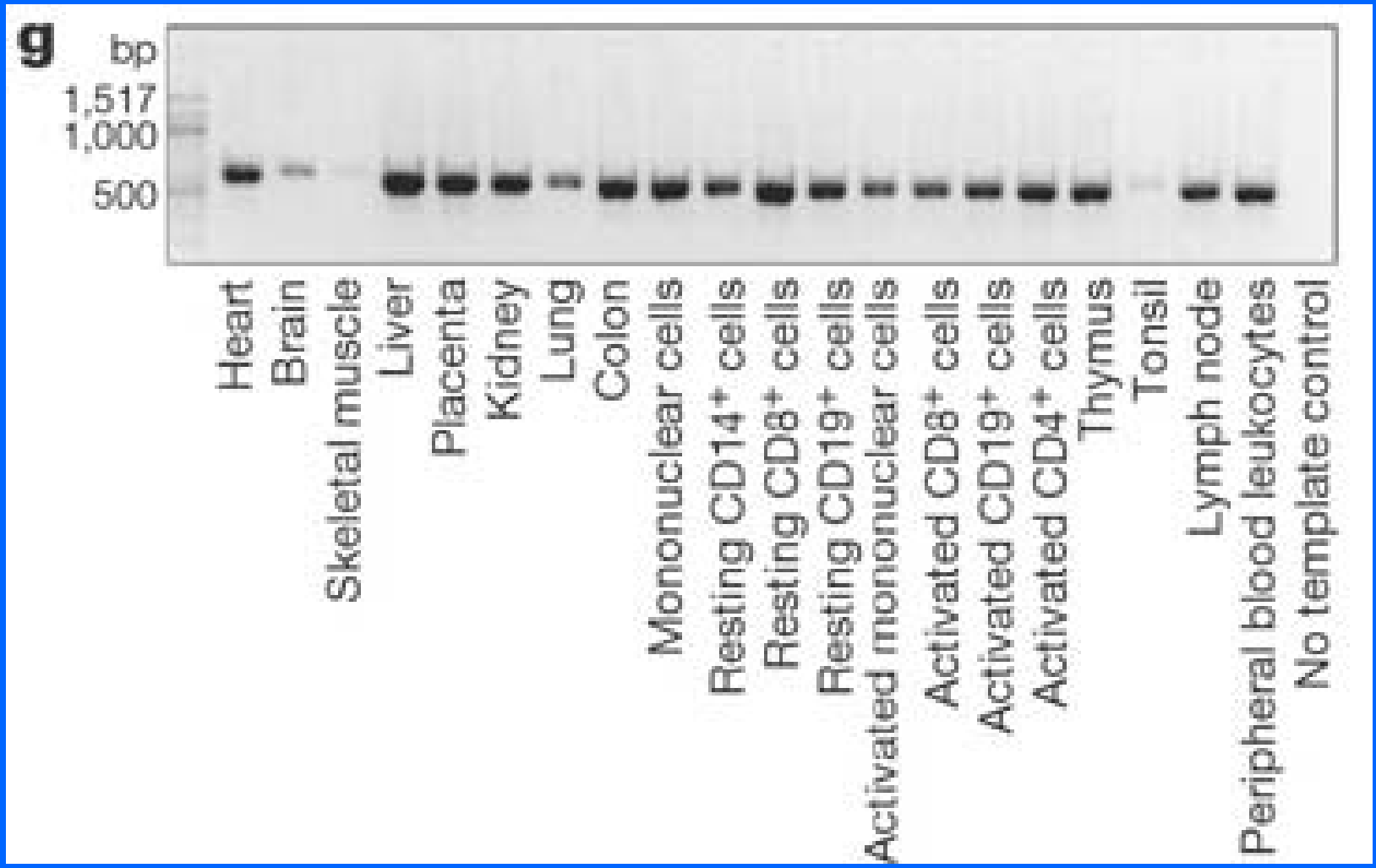






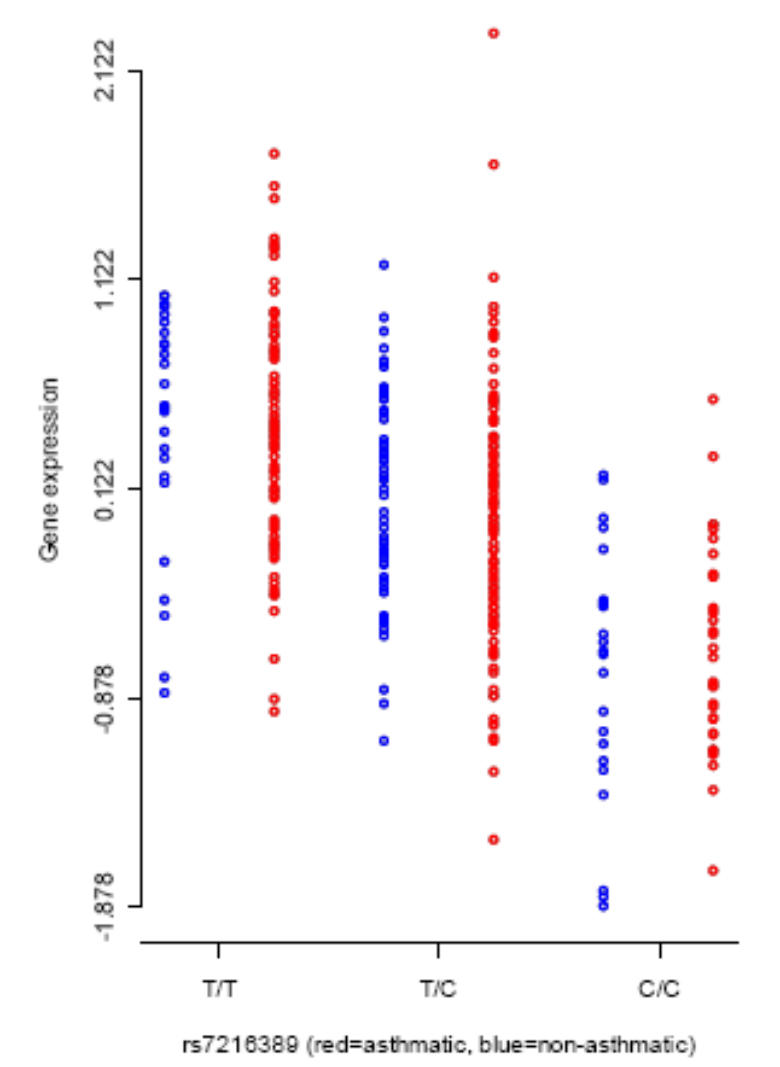
Reiman et al, *Neuron* 2007; 54:713-20.

Conservation and Expression Studies: Asthma and *ORMDL3*



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

Conservation and Expression Studies: Asthma and *ORMDL3*

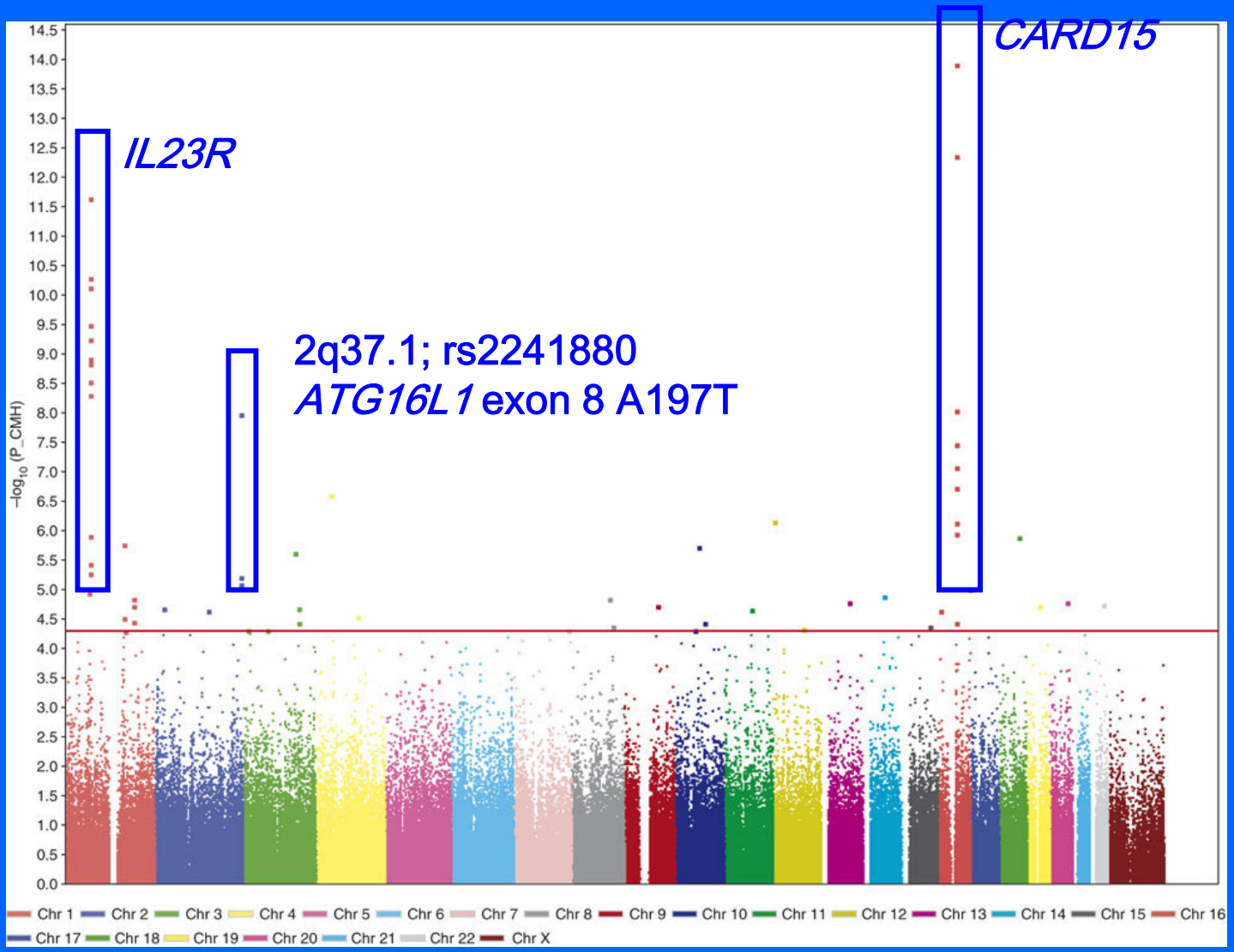


Moffatt et al, *Nature* 2007; 448:470-73.

Knockdown and Knockout Studies

- Knockdown of *ATG16L1*
  - Associated with Crohn's disease
  - Reduces phagocytosis of *S. typhimurium* in HeLa cells
- Knockdown of *GAB2*
  - Associated with Azheimer's disease
  - Increases tau phosphorylation
- Knockout of *MLXIPL*
  - Associated with lower triglyceride levels
  - Knockout shows lower triglyceride levels
  - Transgenic (knockin) shows higher levels

Genome-Wide Associations in Crohn's Disease



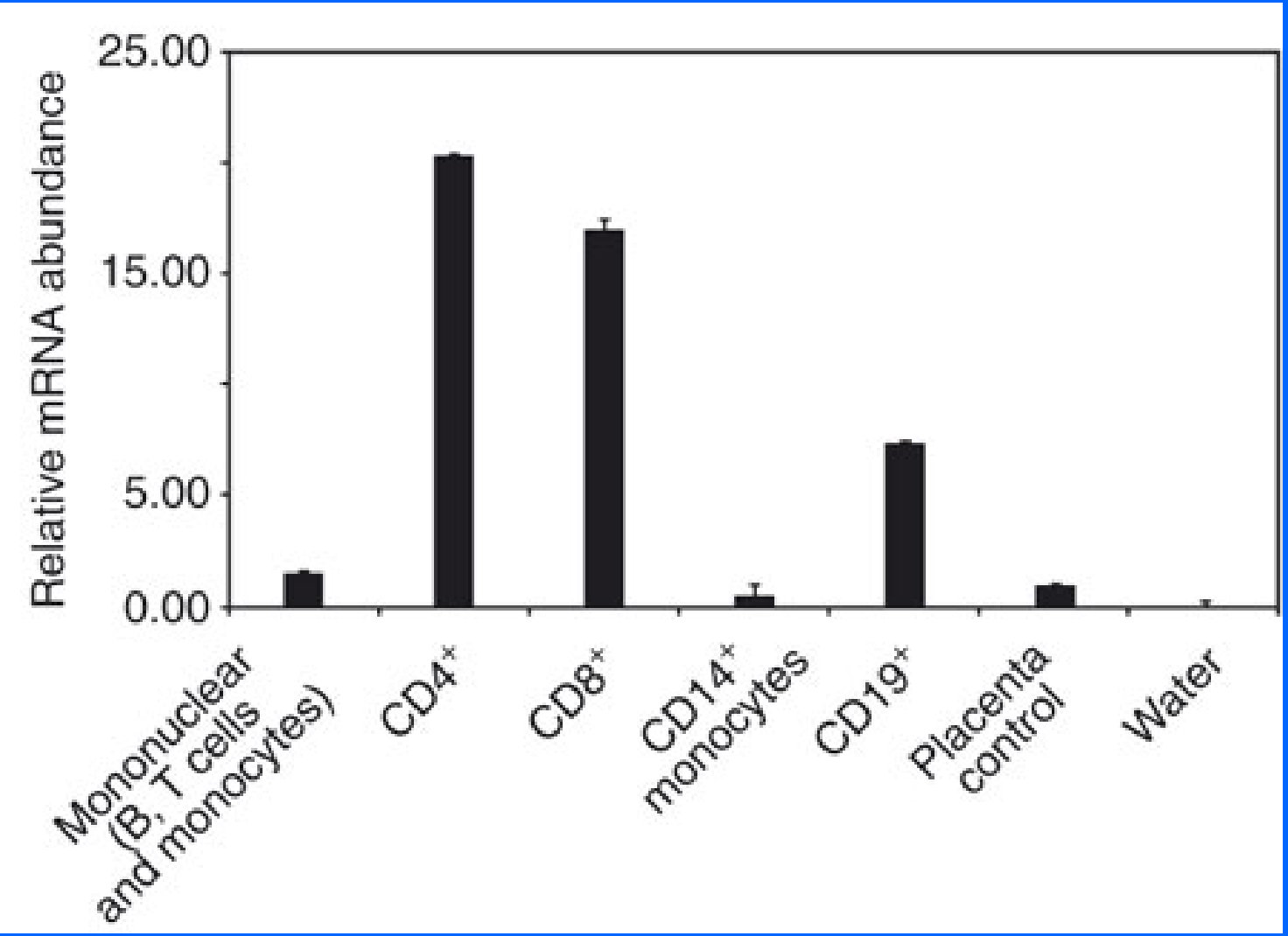
Rioux et al, *Nat Genet* 2007; 39:596-604.

Gene Expression in Crohn's Disease Study

- 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

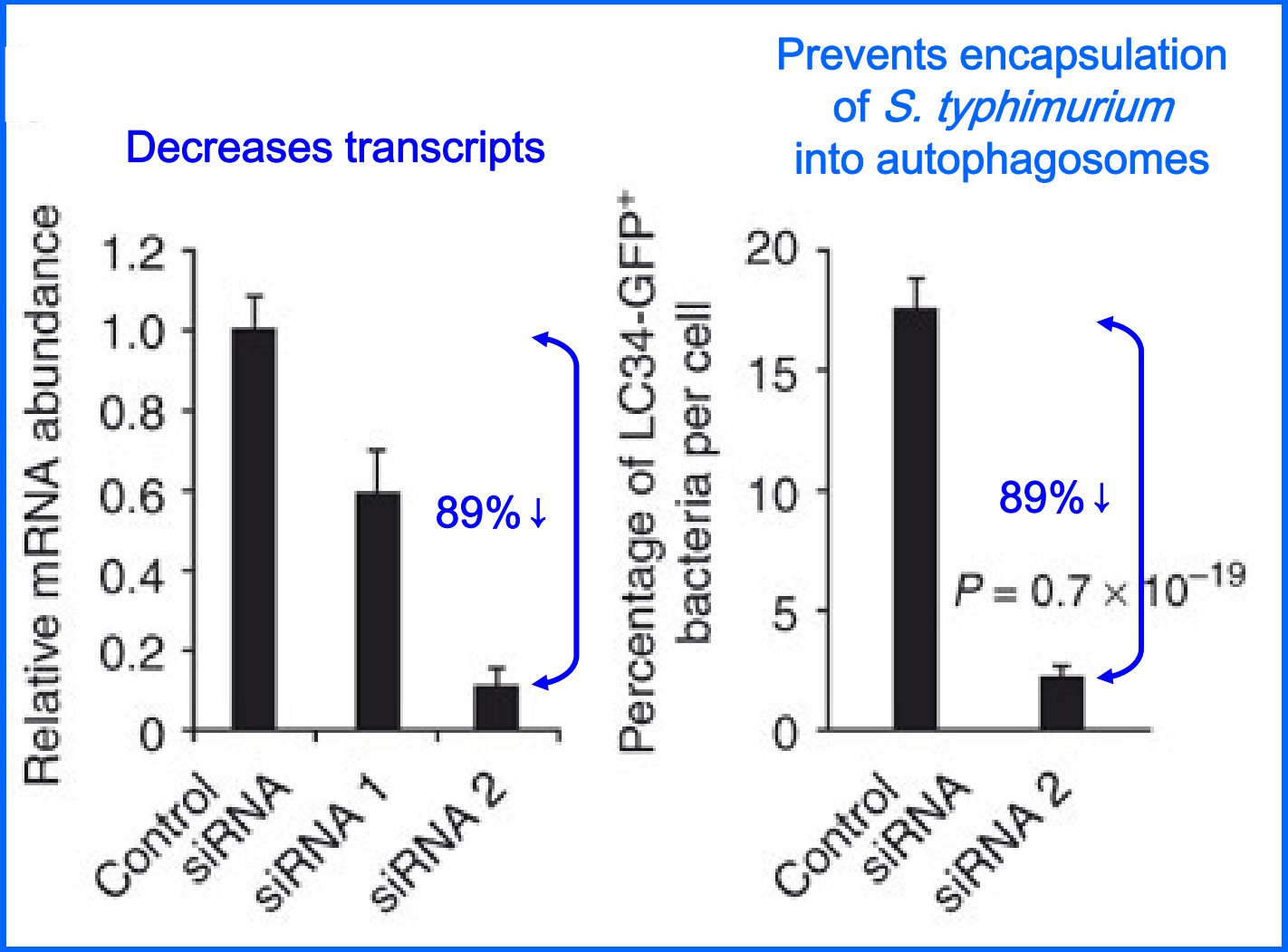
Rioux et al, *Nat Genet* 2007; 39:596-604.

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.

Finding (Putative) Causal Variants Post GWA

- 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 animal models

Pharmacogenetics: The study of differences in drug response due to allelic variation in genes affecting drug metabolism, efficacy, and toxicity.

- Drug metabolism under genetic control
  - Hydroxylation
  - Conjugation
    - Glucuronidation
    - Acetylation
    - Methylation
- Phenotypes of drug metabolism

  Normal metabolizers

  Poor metabolizers

  Ultrafast metabolizers

Frequency of Slow-Acelylator Phenotype Affecting Isoniazid Metabolism

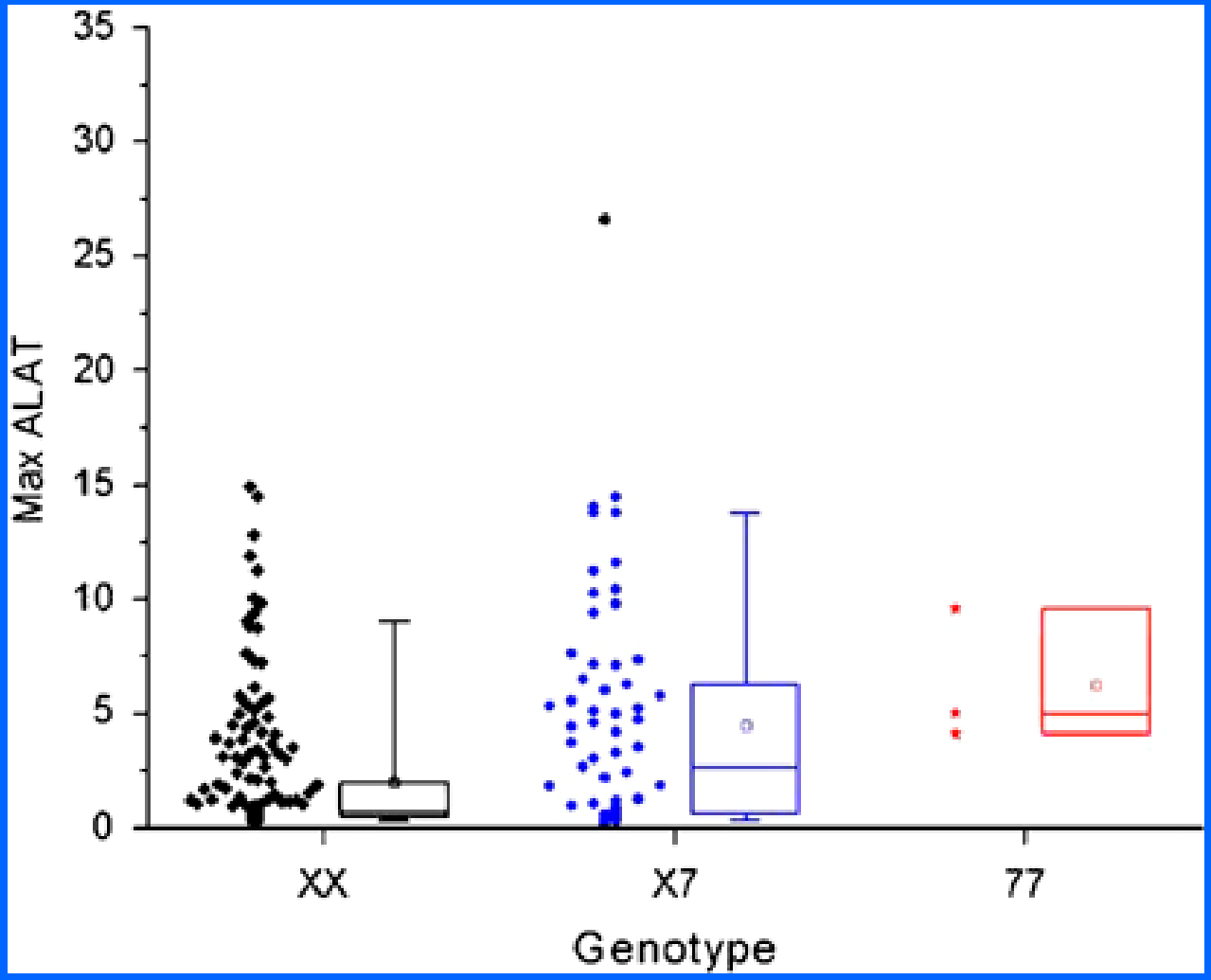
Population	Frequency (%)
African and African-American	51
White	58
Chinese	22
Japanese	10
Inuit	6

Burroughs VJ, et al., cited in Nussbaum R, Thompson and Thompson's Genetic Medicine, 2007

GWAS Involving Pharmacogenetics

Drug	Phenotypes Studies	Reference
Nicotine	Dependent vs. not Dependent by Fagerstrom Score	Bierut LJ, et al Hum Molec Gen 2007 16:24-35
Beta interferon	Responses vs. no response in multiple sclerosis patients	Byum E, et al. Arch Neurol 2008. 65 (3)
Direct Thombin Inhibitor, ximelagatran	Elevation vs. no elevation of serum transaminase levels	Kindmark A, et al, Pharmocogenomics 2007; 5/15/07
Methamphetamines	Dependence vs. controls	Uhl G, et al. Arch Gen Psych 2008; 65: 345-355
Nicotine	Inability to quit vs. able to quit smoking	Uhl G, et al. BMC Genetics 2007; 8:10

HLA *DRB1**0701 and Transaminase Elevations Following Ximelegatran Treatment



Kindmark et al, *Pharmacogen J* 2007; May 15 (on-line)

Policy for Sharing of Data Obtained in NIH Supported and Conducted GWAS (NOT-OD-07-088)

Goal:

To make available the genotype and phenotype datasets as rapidly as possible to a wide range of scientific investigators.

Components:

Data repository (NCBI, dbGAP)

Data submission and protection

Data access

Publication

Intellectual property

Investigators Requesting and Receiving GWAS Data

- Submit a description of proposed research project
- Submit a data access request, co-signed by Institutional Official
- Protect data confidentiality
- Ensure data security measures are in place
- Notify appropriate Data Access Committee of policy violations, if any
- Submit annual reports on research findings

Practical Application Problem

Investigator X identifies several SNPs to be associated with a dread disease that you are studying in your laboratory. The problem is that the SNPs are not in or near any known genes associated with this disease. Dr. X asks for your assistance in establishing the biologic plausibility of the gene-disease association. How could you help Dr. X?

Practical Application Problem: Possible Functional Studies

- Fine mapping, sequencing of region.
- Identification of nearby genes and conservation in association region.
- Measurement of levels of protein product of putative genes.
- Histopathologic studies co-localizing gene expression and pathologic changes.
- Measurement of expression levels in gene variant groups.
- Knockdown, knockout animal models

Summary Points

- The inference that a gene variant causes disease is rarely established by a single study, and usually requires numerous studies of different types.
- Gene association studies not only offer basic laboratory insights into novel disease mechanisms, but relies on them to elucidate these mechanisms to support a causal inference.
- Pharmacologic agents are a type of environmental exposure for which genetic mechanisms can affect efficacy and safety.
- Open access and sharing of gene association study results offer the best opportunity to learn the full importance of genetic diversity.