State of the Science of Genomic Associations: Current and Future Directions

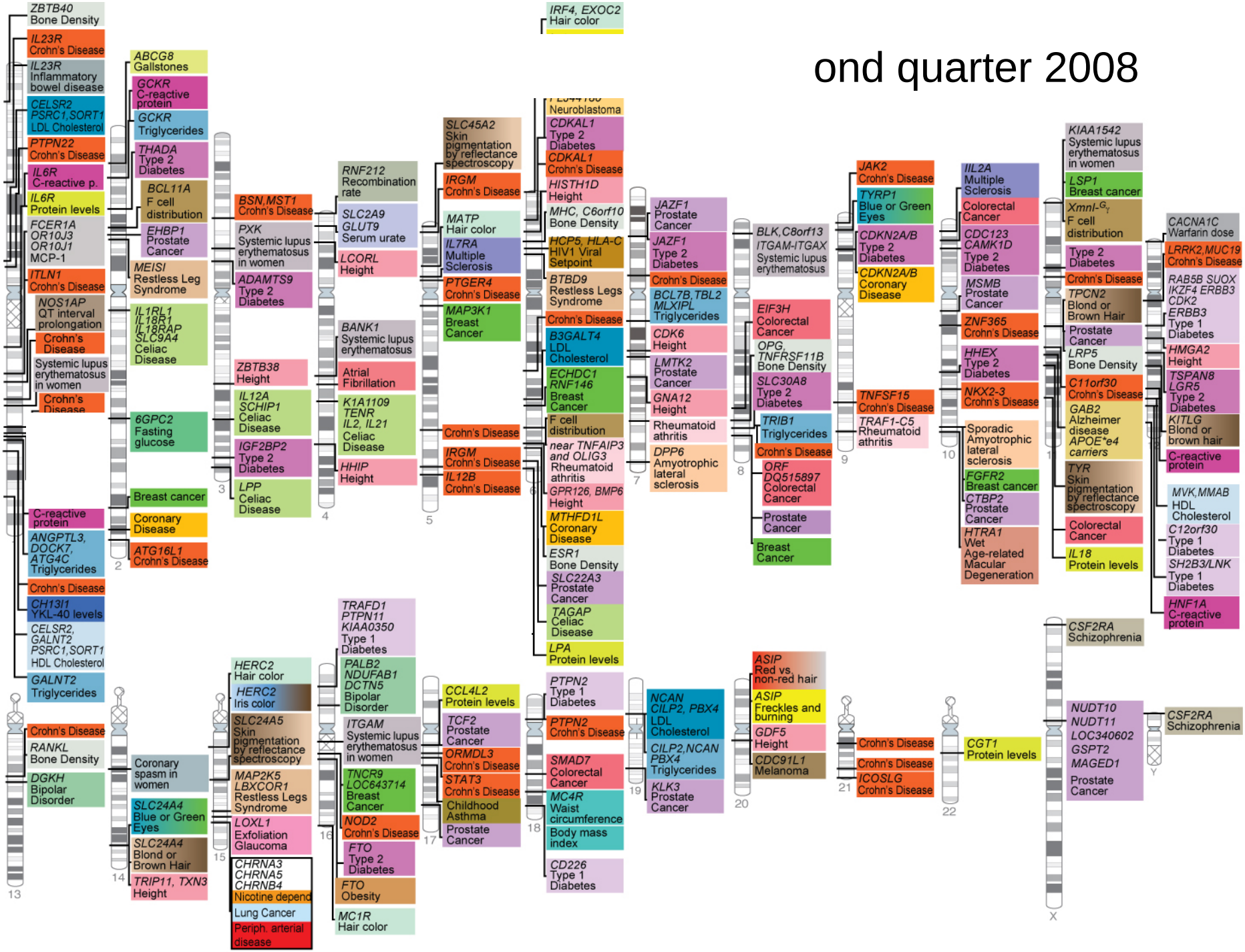
U.S. Department of Health and Human Services
National Institutes of Health
National Human Genome Research Institute

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

July 8, 2008

State of the Science of Genomic Associations

- What are the recent advances in genomics research? How have these advances facilitated the emergence of personal genome services?
- For which diseases are strong genetic associations and/or markers established?
- What criteria should be considered in determining whether the association between a particular genetic marker and phenotype is strong enough for that marker to be included in genetic testing?
- What are the limitations of genetic markers in risk assessment for disease?


ond quarter 2008
ZBTB40 Bone Density
IL23R Crohn's Disease
IL23R Inflammatory bowel disease
CELSR2 PSRC1,SORT1 LDL Cholesterol
PTPN22 Crohn's Disease
IL6R C-reactive p.
IL6R Protein levels
FCER1A OR10J3 OR10J1 MCP-1
ITLN1 Crohn's Disease
NOS1AP QT interval prolongation
Crohn's Disease
Systemic lupus erythematosus in women
Crohn's Disease
C-reactive protein
ANGPTL3, DOCK7, ATG4C Triglycerides
Crohn's Disease
CH13I1 YKL-40 levels
CELSR2, GALNT2 PSRC1,SORT1 HDL Cholesterol
GALNT2 Triglycerides
Crohn's Disease
RANKL Bone Density
DGKH Bipolar Disorder
13
ABCG8 Gallstones
GCKR C-reactive protein
GCKR Triglycerides
THADA Type 2 Diabetes
BCL11A F cell distribution
EHBP1 Prostate Cancer
MEISI Restless Leg Syndrome
IL1RL1 IL18R1 IL18RAP SLC9A4 Celiac Disease
6GPC2 Fasting glucose
Breast cancer
Coronary Disease
ATG16L1 Crohn's Disease
2
Coronary spasm in women
SLC24A4 Blue or Green Eyes
SLC24A4 Blond or Brown Hair
TRIP11, TXN3 Height
14
BSN,MST1 Crohn's Disease
PXK Systemic lupus erythematosus in women
ADAMTS9 Type 2 Diabetes
ZBTB38 Height
IL12A SCHIP1 Celiac Disease
IGF2BP2 Type 2 Diabetes
LPP Celiac Disease
3
HERC2 Hair color
HERC2 Iris color
SLC24A5 Skin pigmentation by reflectance spectroscopy
MAP2K5 LBXCOR1 Restless Legs Syndrome
LOXL1 Exfoliation Glaucoma
CHRNA3 CHRNA5 CHRNB4 Nicotine depend
Lung Cancer
Periph. arterial disease
15
RNF212 Recombination rate
SLC2A9 GLUT9 Serum urate
LCORL Height
BANK1 Systemic lupus erythematosus
Atrial Fibrillation
K1A1109 TENR IL2, IL21 Celiac Disease
HHIP Height
4
TRAFD1 PTN11 KIAA0350 Type 1 Diabetes
PALB2 NDUFAB1 DCTN5 Bipolar Disorder
ITGAM Systemic lupus erythematosus in women
TNCR9 LOC643714 Breast Cancer
NOD2 Crohn's Disease
FTO Type 2 Diabetes
FTO Obesity
MC1R Hair color
16
SLC45A2 Skin pigmentation by reflectance spectroscopy
IRGM Crohn's Disease
MATP Hair color
IL7RA Multiple Sclerosis
PTGER4 Crohn's Disease
MAP3K1 Breast Cancer
Crohn's Disease
IRGM Crohn's Disease
IL12B Crohn's Disease
5
CCL4L2 Protein levels
TCF2 Prostate Cancer
ORMDL3 Crohn's Disease
STAT3 Crohn's Disease
Childhood Asthma
Prostate Cancer
17
IRF4, EXOC2 Hair color
Neuroblastoma
CDKAL1 Type 2 Diabetes
CDKAL1 Crohn's Disease
HISTH1D Height
MHC, C6orf10 Bone Density
HCP5, HLA-C HIV1 Viral Setpoint
BTBD9 Restless Legs Syndrome
Crohn's Disease
B3GALT4 LDL Cholesterol
ECHDC1 RNF146 Breast Cancer
F cell distribution
near TNFAIP3 and OLIG3 Rheumatoid athritis
GPR126, BMP6 Height
MTHFD1L Coronary Disease
ESR1 Bone Density
SLC22A3 Prostate Cancer
TAGAP Celiac Disease
LPA Protein levels
6
PTPN2 Type 1 Diabetes
PTPN2 Crohn's Disease
SMAD7 Colorectal Cancer
MC4R Waist circumference
Body mass index
CD226 Type 1 Diabetes
18
JAZF1 Prostate Cancer
JAZF1 Type 2 Diabetes
Crohn's Disease
BCL7B, TBL2 MLXIPL Triglycerides
CDK6 Height
LMTK2 Prostate Cancer
GNA12 Height
Rheumatoid athritis
DPP6 Amyotrophic lateral sclerosis
7
NCAN CILP2, PBX4 LDL Cholesterol
CILP2, NCAN PBX4 Triglycerides
KLK3 Prostate Cancer
19
BLK, C8orf13 ITGAM-ITGAX Systemic lupus erythematosus
EIF3H Colorectal Cancer
OPG, TNFRSF11B Bone Density
SLC30A8 Type 2 Diabetes
TRIB1 Triglycerides
Crohn's Disease
ORF DQ515897 Colorectal Cancer
Prostate Cancer
Breast Cancer
8
ASIP Red vs. non-red hair
ASIP Freckles and burning
GDF5 Height
CDC91L1 Melanoma
20
JAK2 Crohn's Disease
TYRP1 Blue or Green Eyes
CDKN2A/B Type 2 Diabetes
CDKN2A/B Coronary Disease
TNFSF15 Crohn's Disease
TRAF1-C5 Rheumatoid athritis
9
Crohn's Disease
Crohn's Disease
ICOSLG Crohn's Disease
21
IIL2A Multiple Sclerosis
Colorectal Cancer
CDC123 CAMK1D Type 2 Diabetes
MSMB Prostate Cancer
ZNF365 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
CGT1 Protein levels
22
KIAA1542 Systemic lupus erythematosus in women
LSP1 Breast cancer
Xmnl-G$_\gamma$ F cell distribution
Type 2 Diabetes
Crohn's Disease
TPCN2 Blond or Brown Hair
Prostate Cancer
LRP5 Bone Density
C11orf30 Crohn's Disease
GAB2 Alzheimer disease APOE*e4 carriers
TYR Skin pigmentation by reflectance spectroscopy
Colorectal Cancer
IL18 Protein levels
1
CSF2RA Schizophrenia
NUDT10 NUDT11 LOC340602 GSPT2 MAGED1
Prostate Cancer
X
CACNA1C Warfarin dose
LRRK2,MUC19 Crohn's Disease
RAB5B SUOX IKZF4 ERBB3 CDK2 ERBB3 Type 1 Diabetes
HMGA2 Height
TSPAN8 LGR5 Type 2 Diabetes
KITLG Blond or brown hair
C-reactive protein
MVK,MMAB HDL Cholesterol
C12orf30 Type 1 Diabetes
SH2B3/LNK Type 1 Diabetes
HNF1A C-reactive protein
CSF2RA Schizophrenia
Y

2007: The Year of GWA Studies

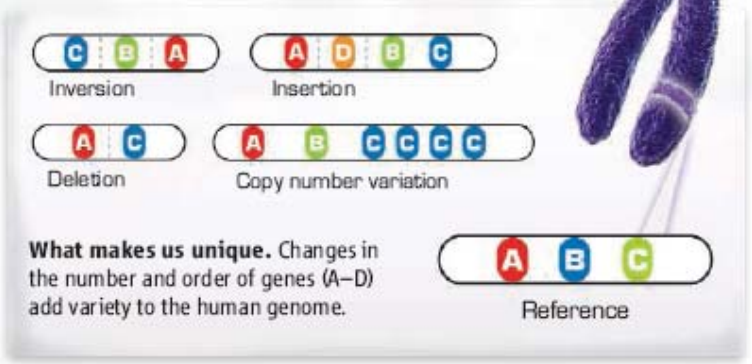
BREAKTHROUGH OF THE YEAR

Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another

THE UNVEILING OF THE HUMAN GENOME ALMOST 7 YEARS AGO cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.



What makes us unique. Changes in the number and order of genes (A–D) add variety to the human genome.

Pennisi E, *Science* 2007; 318:1842-43.

www.hapmap.org

Vol 437|27 October 2005|doi:10.1038/nature04226 nature

Vol 449|18 October 2007|doi:10.1038/nature06258 nature

Nature 2007; 449:851-61.

ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

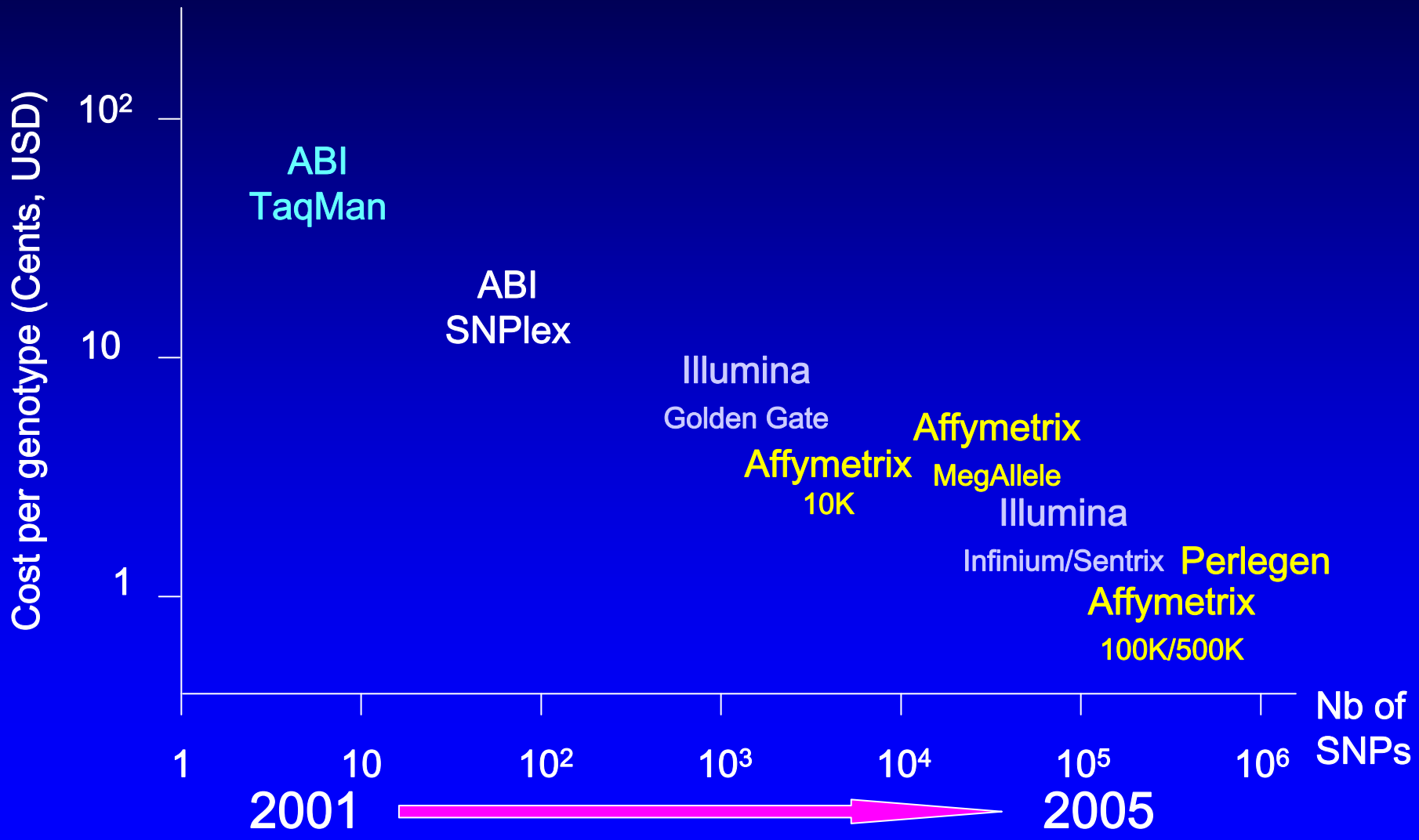
The International HapMap Consortium*

We describe the Phase II HapMap, which characterizes over 3.1 million human single nucleotide polymorphisms (SNPs) genotyped in 270 individuals from four geographically diverse populations and includes 25–35% of common SNP variation in the populations surveyed. The map is estimated to capture untyped common variation with an average maximum r^2 of between 0.9 and 0.96 depending on population. We demonstrate that the current generation of commercial genome-wide genotyping products captures common Phase II SNPs with an average maximum r^2 of up to 0.8 in African and up to 0.95 in non-African populations, and that potential gains in power in association studies can be obtained through imputation. These data also reveal novel aspects of the structure of linkage disequilibrium. We show that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry and that up to 1% of all common variants are untaggable, primarily because they lie within recombination hotspots. We show that recombination

A HapMap for More Efficient Association Studies: Goals

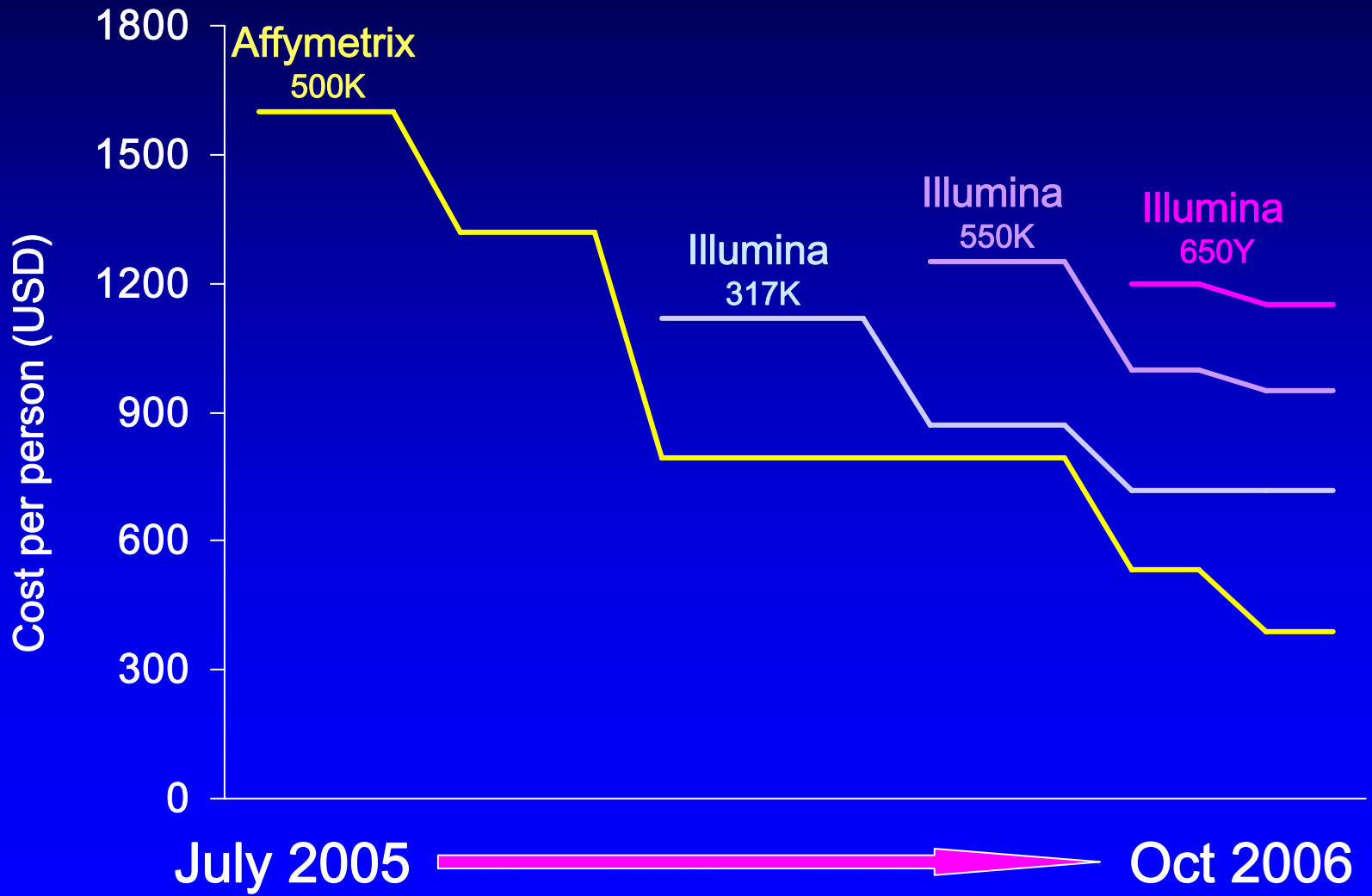
- Use just the density of SNPs needed to find associations between SNPs and diseases
- Do not miss chromosomal regions with disease association
- Produce a tool to assist in finding genes affecting health and disease
- Use more SNPs for complete genome coverage of populations of recent African ancestry populations due to shorter LD

Progress in Genotyping Technology



Courtesy S. Chanock, NCI

Continued Progress in Genotyping Technology



Courtesy S. Gabriel, Broad/MIT

Diseases and Traits with Published GWA Studies (n = 58, 7/7/08)

- Macular Degeneration
- Exfoliation Glaucoma

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

- Inflamm. Bowel Disease
- Celiac Disease
- Gallstones
- Irritable Bowel Syndrome

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

- Warfarin Dosing
- Ximelegatran Adv. Resp.

- 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-St. 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

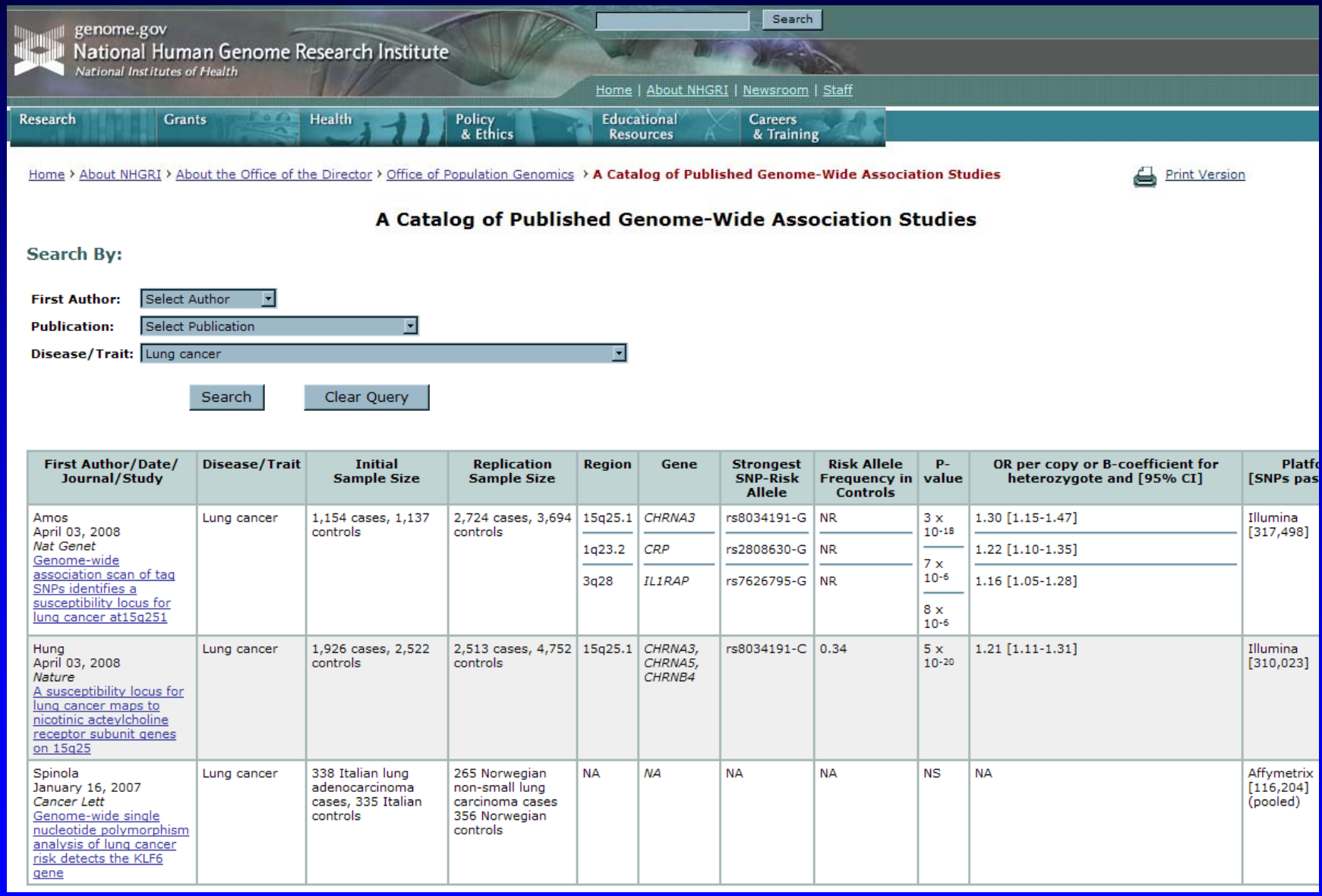
and in this issue of the *Journal*, coronary artery disease (reported by Samani et al., pages 443–453). These genomewide association studies have been able to examine interpatient differences in inherited genetic variability at an unprecedented level of resolution, thanks to the development of microarrays,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

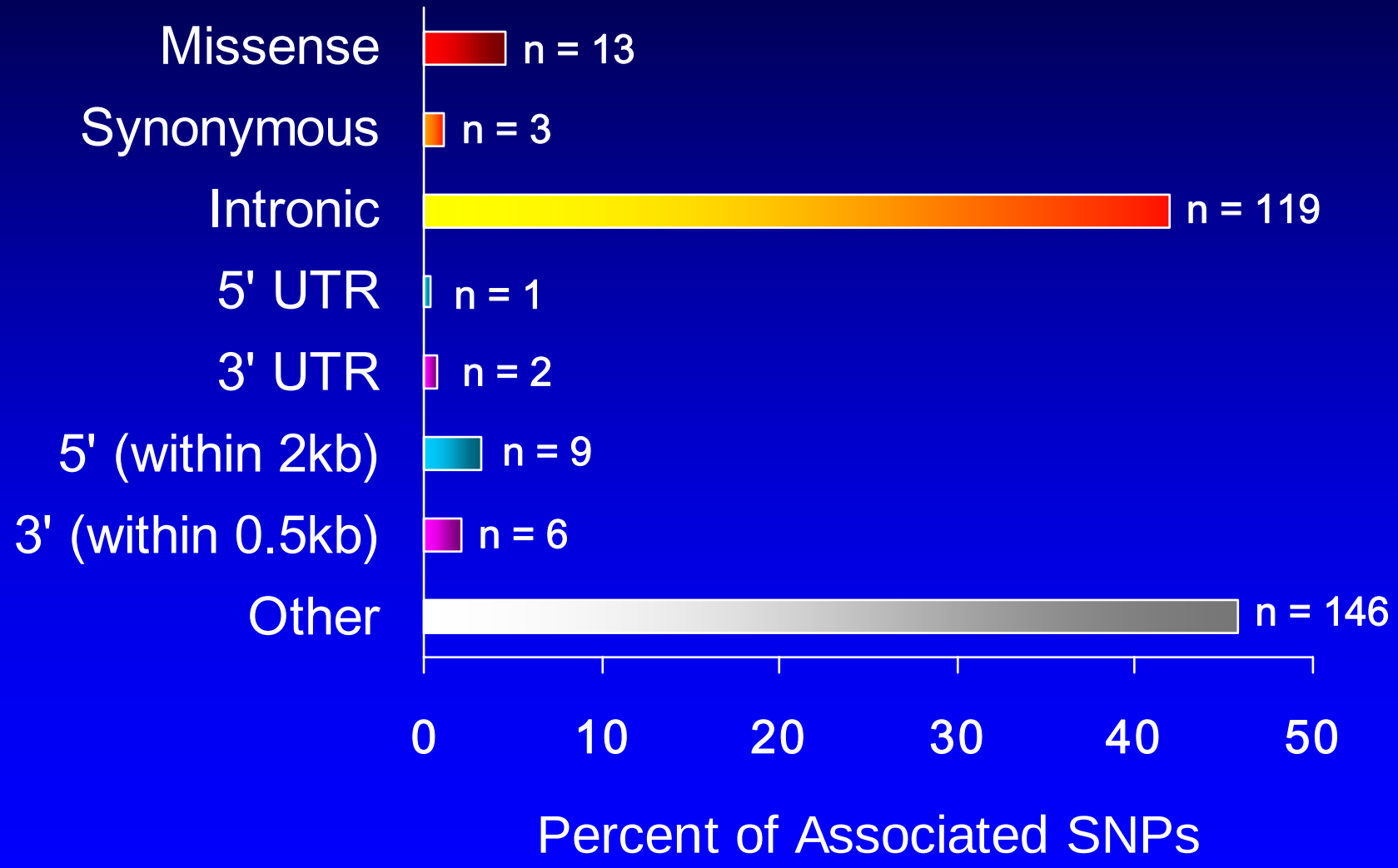
Publication: Select Publication

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)

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*

What are the recent advances in genomics research? How have these advances facilitated the emergence of personal genome services?

- Low-cost high-throughput genotyping within reach of large-scale population studies since 2006
- Over 150 such studies completed with over 180 well-replicated loci in nearly 60 diseases/traits
- Genotyping costs now within reach of (well-to-do) consumers

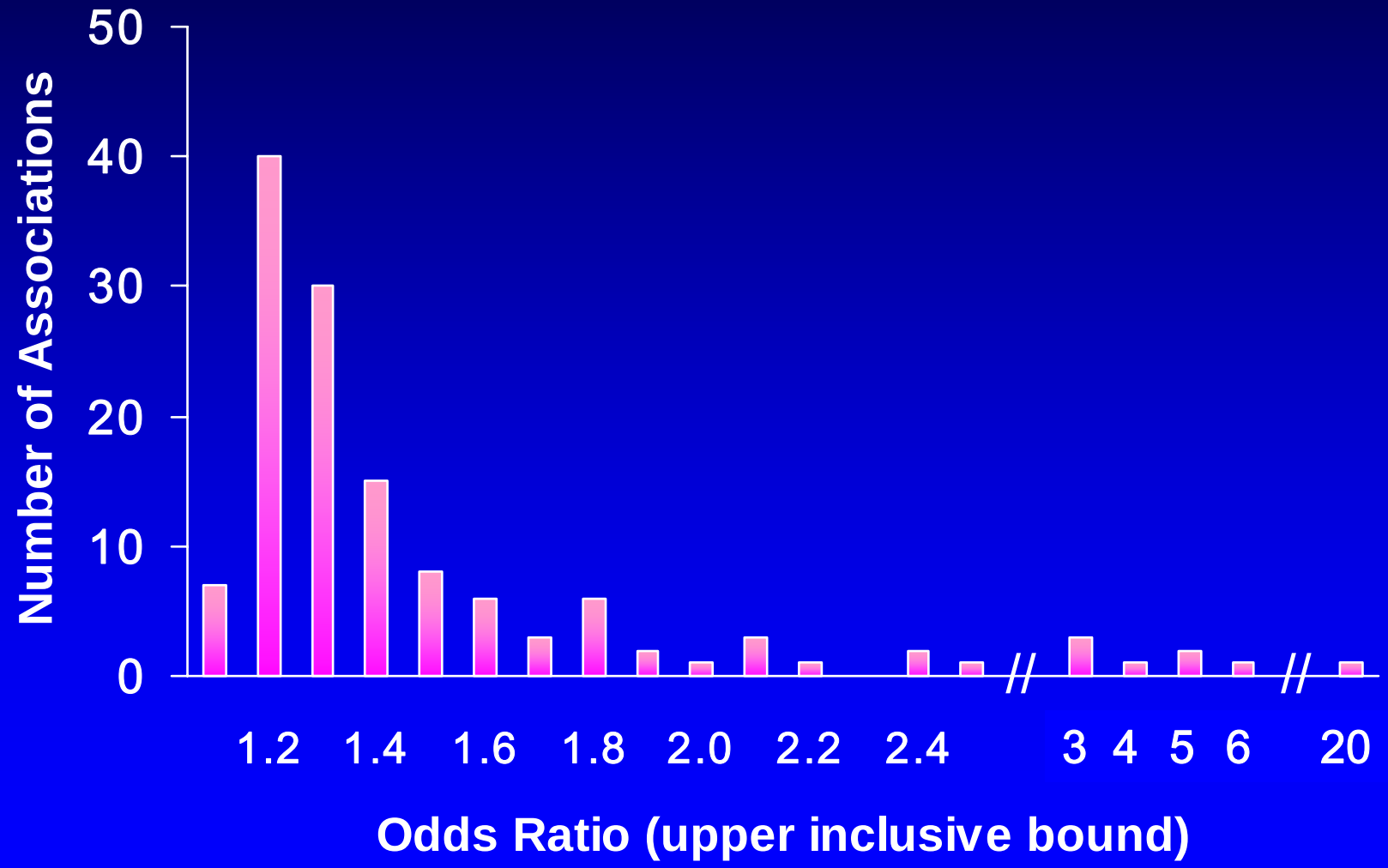
On the horizon…

- Copy number variants
- Next generation sequencing; 1,000 Genomes
- DNA methylation and gene expression

For which diseases are strong genetic associations and/or markers established?

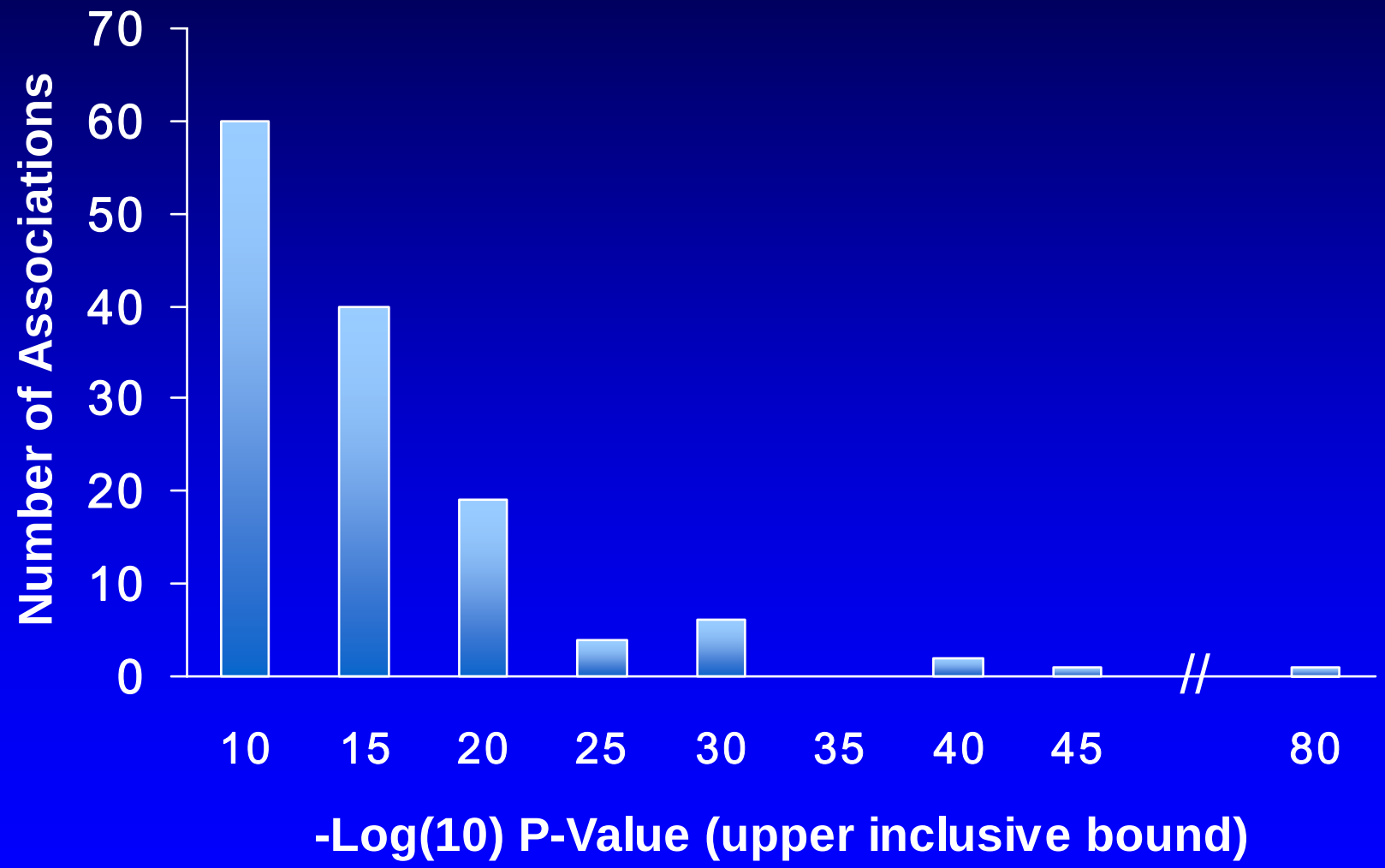
- Define “strong”!
- Large odds ratio?
- Very small p-value?
- Very frequently occurring risk allele?
- Large proportion of disease attributable to risk allele?
- Explaining large proportion of genetic variance?

Odds Ratios of Discrete Associations



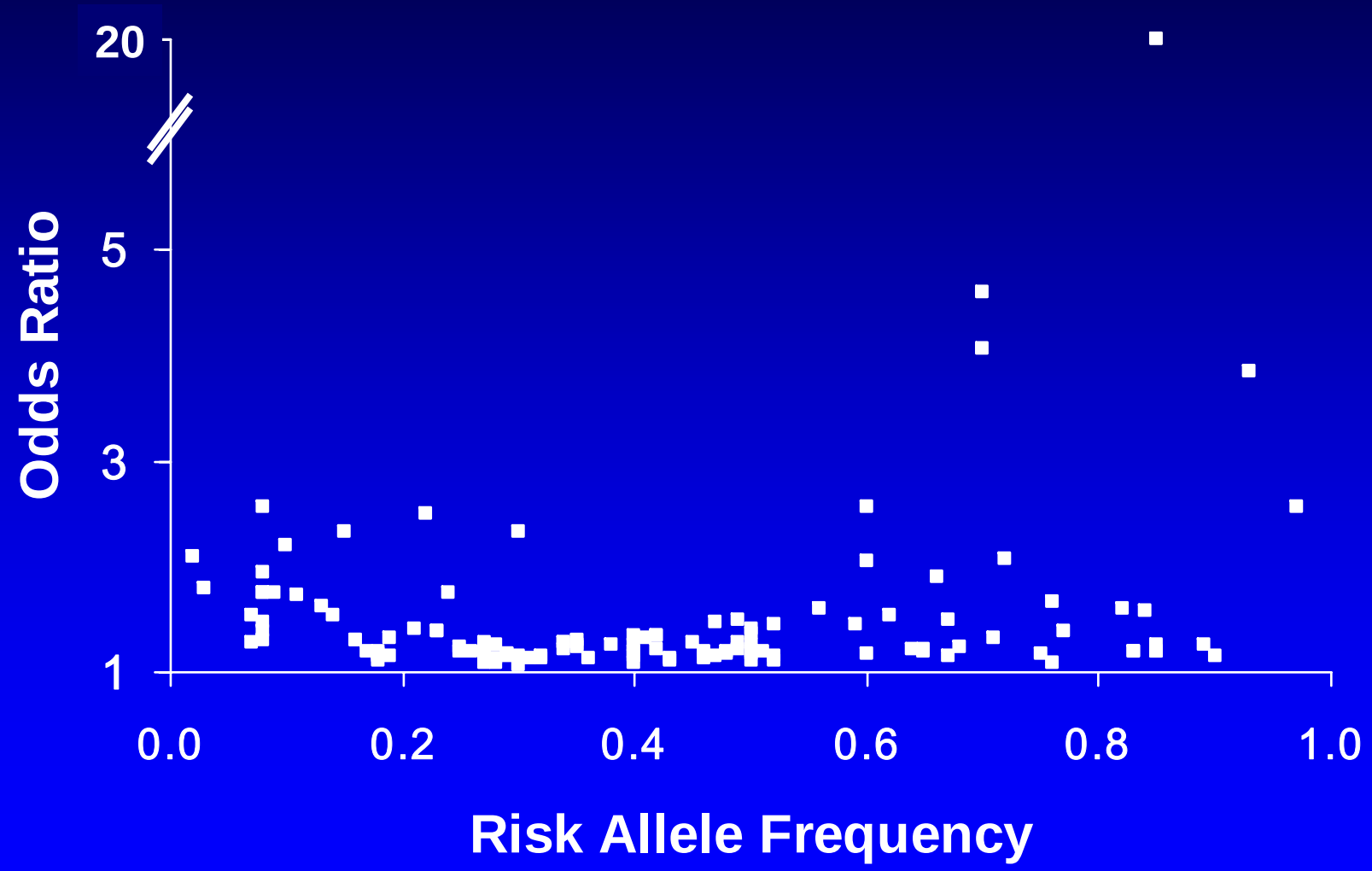
http://www.genome.gov/gwastudies/

- Log_{10} P-Values of Discrete Associations



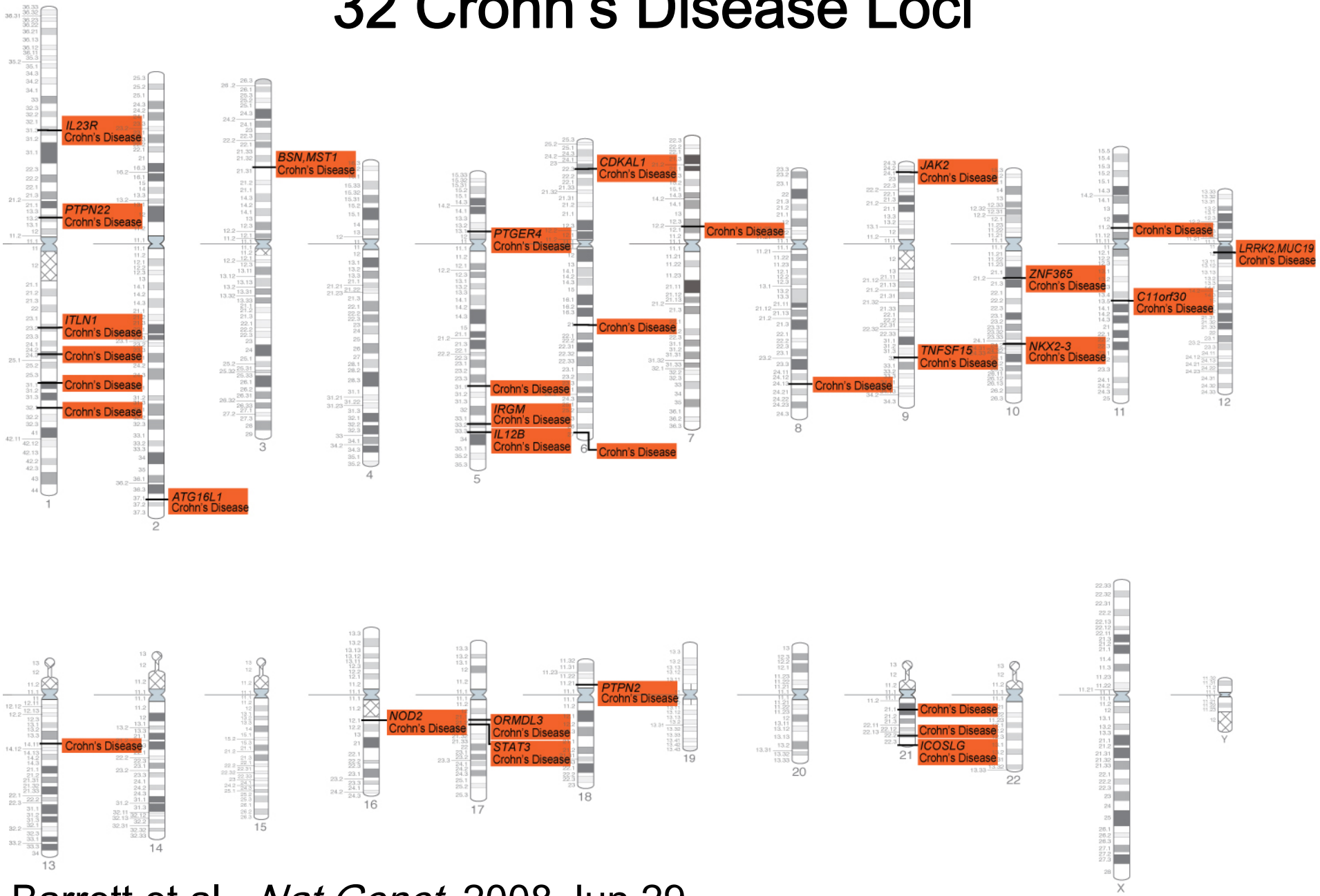
http://www.genome.gov/gwastudies/

Odds Ratio by Risk Allele Frequency for Discrete Associations



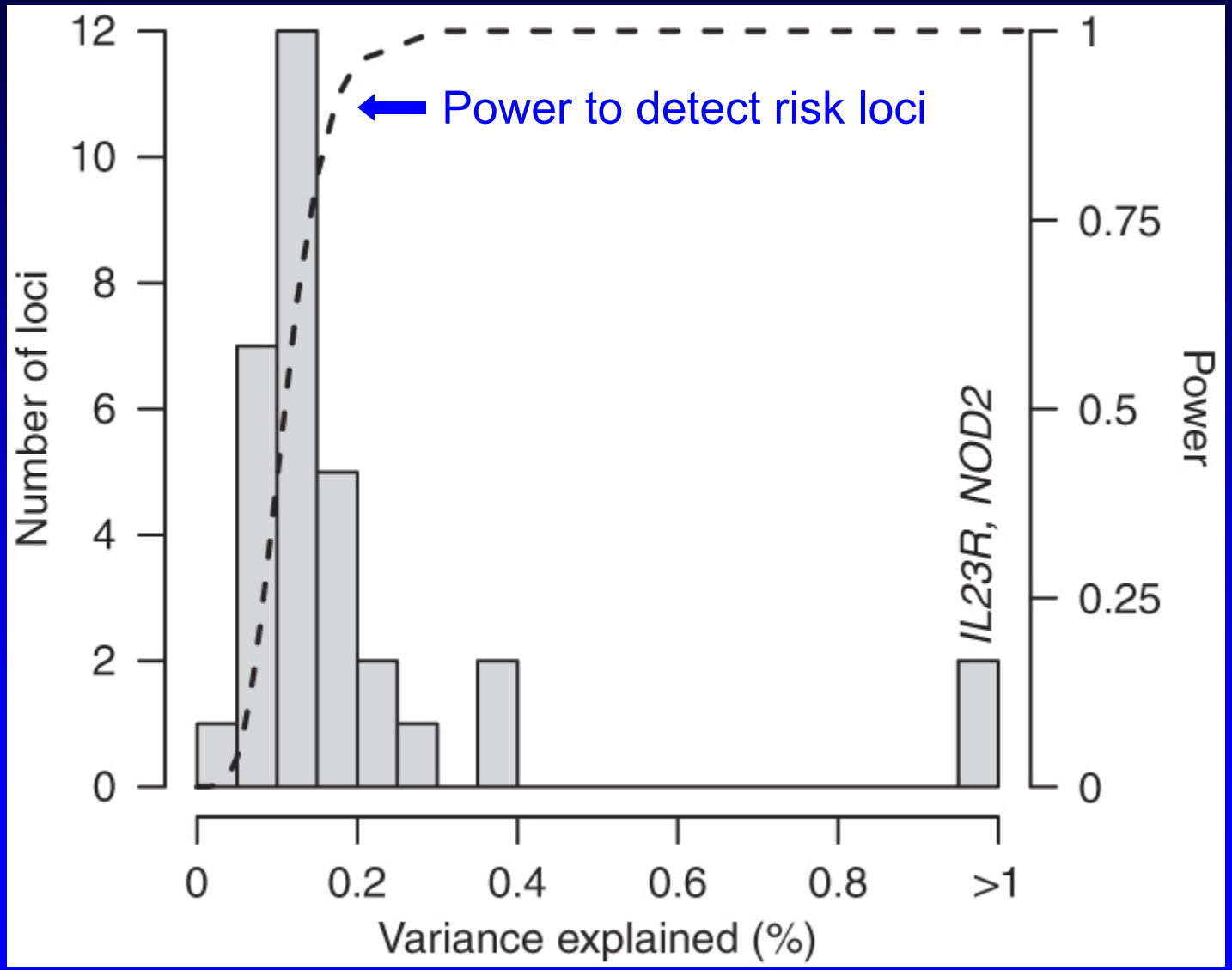
http://www.genome.gov/gwastudies/

32 Crohn's Disease Loci



Barrett et al., *Nat Genet* 2008 Jun 29.

Percent of Variance in Disease Risk Explained by 32 Established CD Risk Loci



Barrett et al., *Nat Genet* 2008 Jun 29.

What criteria should be considered in determining whether the association between a particular genetic marker and phenotype is strong enough for that marker to be included in genetic testing?

What criteria should be considered in determining whether a particular genetic marker should be included in genetic testing?

Depends to very large degree on purpose of testing…

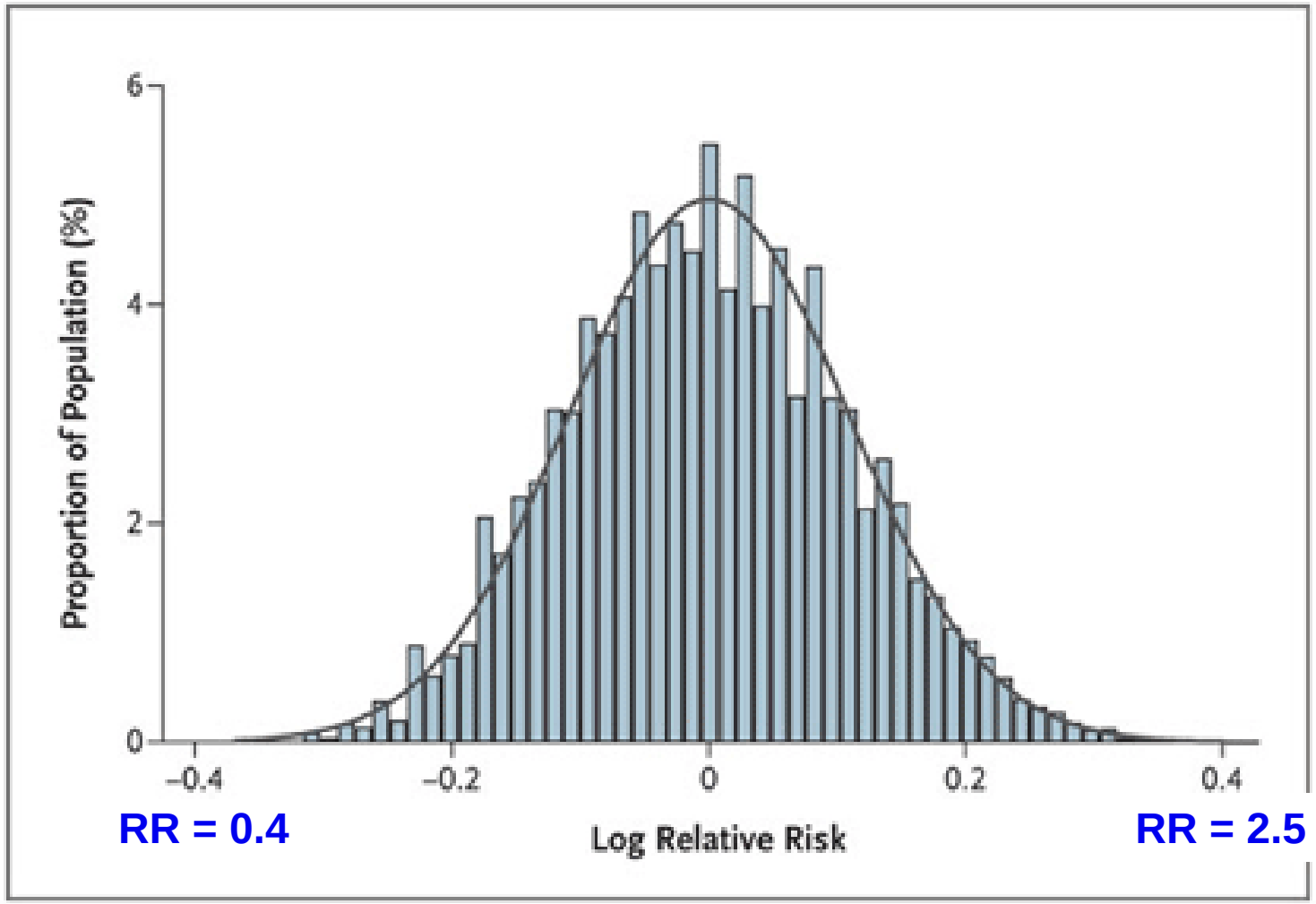
Possible Purposes of Genetic Testing

- To improve health and prevent disease
- To provide targeted, proven risk reduction strategies to those at greatest risk
- To identify persons at high risk, for later rapid implementation of newly-proven interventions
- To improve cost efficiency of non-genetic risk reduction strategies
- To facilitate reproductive choices
- To provide information of personal value to individuals, regardless of whether “actionable”

Criteria to be Considered in Selecting Genetic Variants for Testing

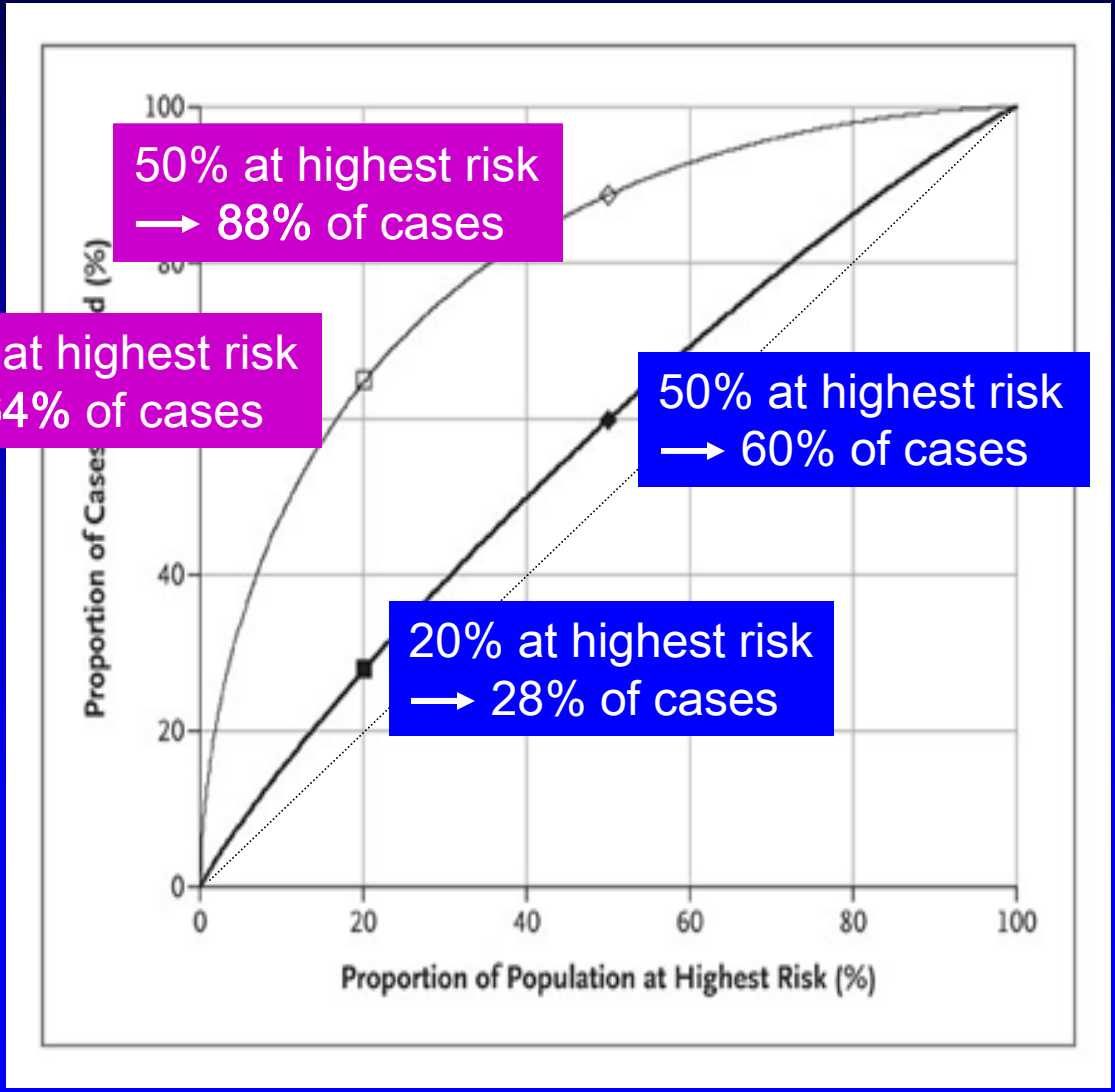
- Strength of evidence for risk association
- Availability and acceptability of proven risk-reducing interventions
- Validity, availability, cost of the test
- Potential anxiety, stigma, cost, additional testing, or other harms from receiving results
- Trade-offs in other testing or care that cannot be paid for within fixed budget

Distribution of Genetic Risk in the Population



Pharoah et al., *N Engl J Med* 2008; 358:2796-803.

Proportion of Breast Cancer Cases Explained by Proportion of Population at Highest Risk



Pharoah et al., *N Engl J Med* 2008; 358:2796-803.

Improving Efficiency of Screening Program for Breast Cancer

- 50-year-old woman in UK has 2.3% risk of breast cancer in next ten years; mammography currently offered to all women over 50
- Offer screening to all women at that risk level?
- Do not offer screening to women not at that risk level?
- Women in 40th percentile of population risk have 10-year risk at age 50 of 2.1%
- Women in 5th percentile of population risk have 10-year risk at age 50 of 1.5% and *never reach* a 2.3% risk

Pharoah et al., *N Engl J Med* 2008; 358:2796-803.

What are the limitations of genetic markers in risk assessment for disease?

- Most markers are not deterministic– many people without the markers will develop disease, and many people with the markers will not
- Most of the genetic risk remains unexplained
- Little or no evidence to date that interventions based on genotype improve outcome
- Genetic markers may provide additional risk information for more aggressive risk management in carriers, but again little evidence

Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S)

PREVENTION OF CARDIOVASCULAR EVENTS AND DEATH WITH PRAVASTATIN IN PATIENTS WITH CORONARY HEART DISEASE AND A BROAD RANGE OF INITIAL CHOLESTEROL LEVELS

Primary Prevention of Acute Coronary Events With Lovastatin in Men and Women With Average Cholesterol Levels

MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial

Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) JAMA 2001

Multiplex Genetic Susceptibility Testing:

A prototype for applied research to inform personalized medicine

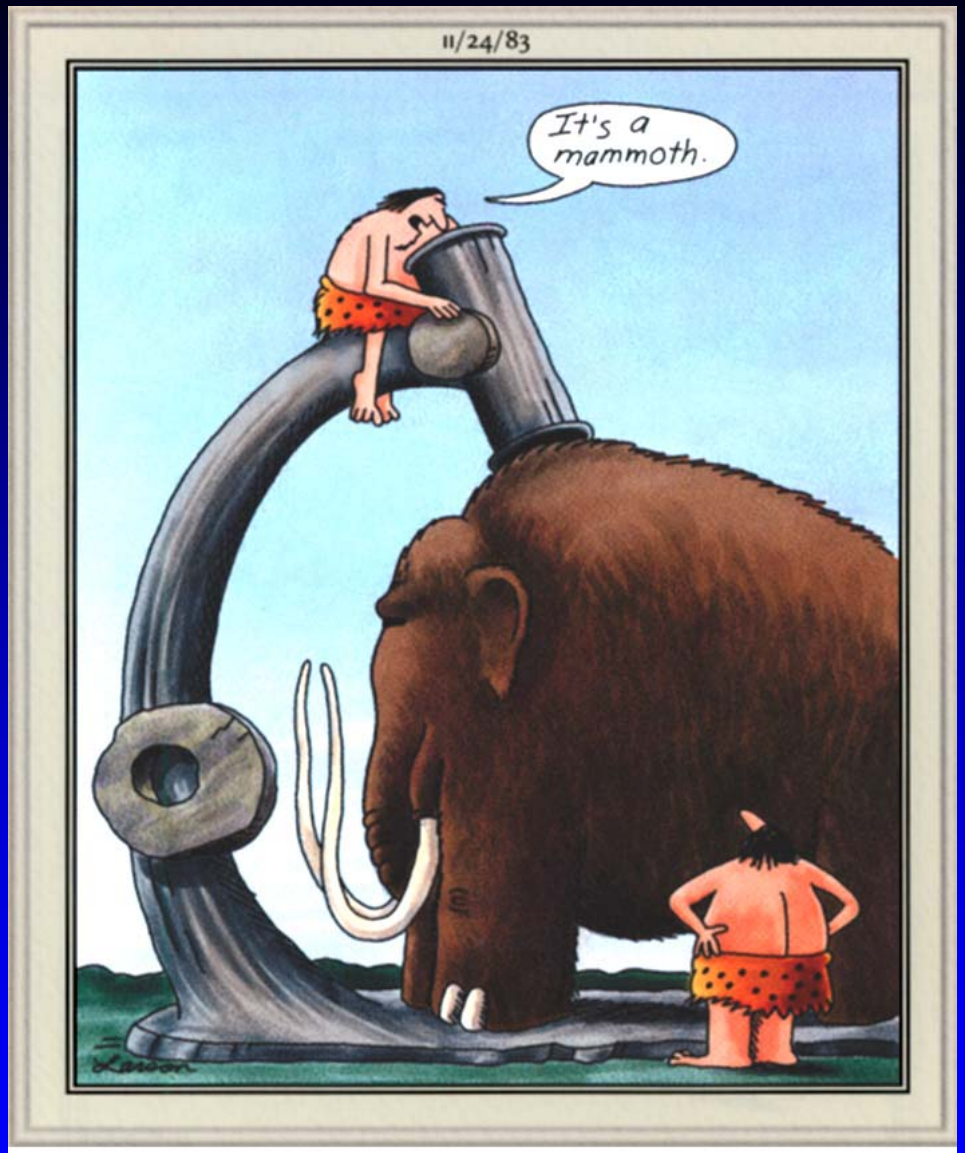
Colleen M. McBride, PhD. and Larry Brody, Ph.D.

Research Partners:

National Human Genome Research Institute
Henry Ford Health System
Group Health Cooperative
Cancer Research Network (NCI)

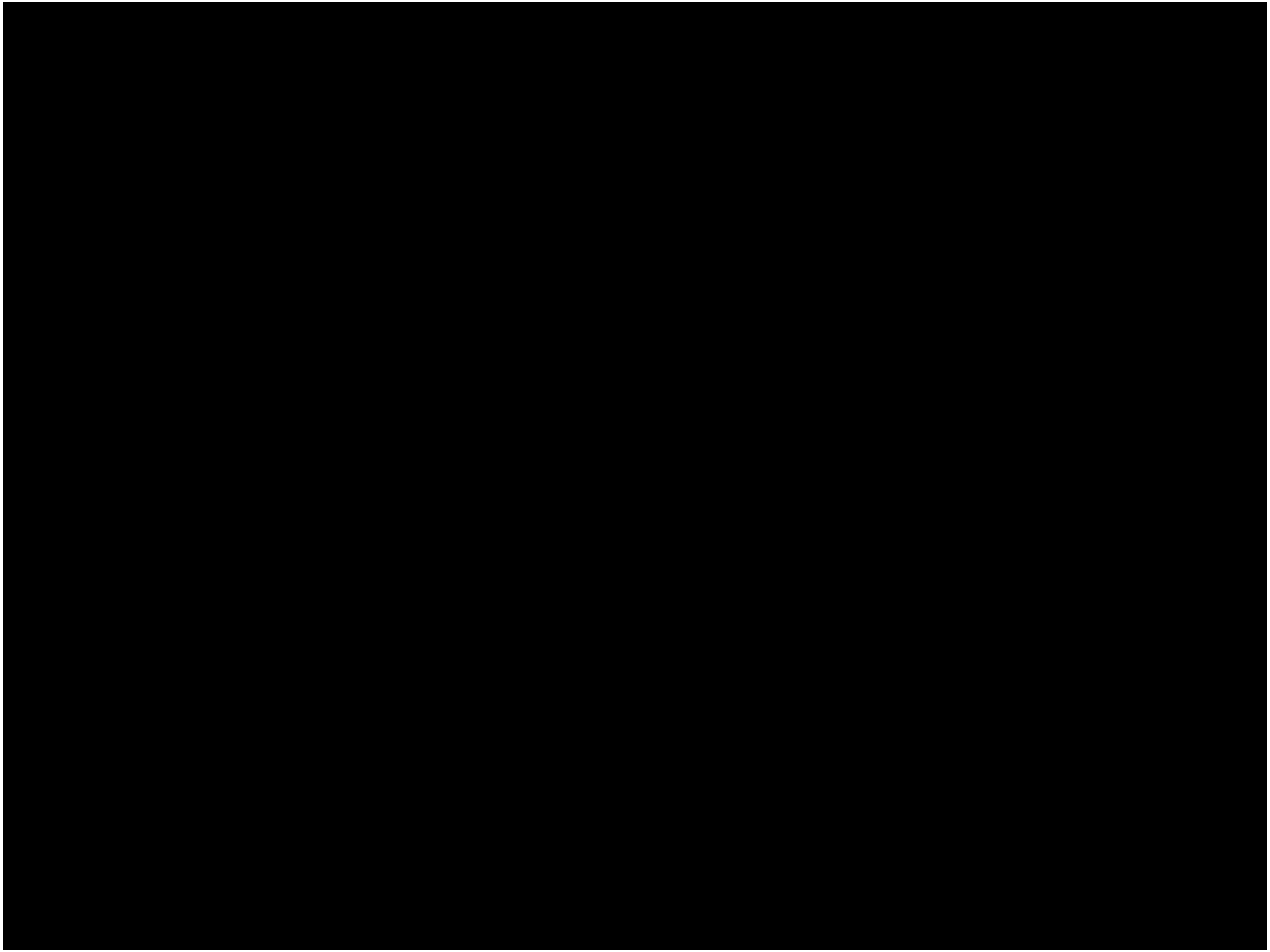
Multiplex Project Aims

- **To develop a prototype for multiplex genetic susceptibility testing**
  - Multiple markers of susceptibility for multiple diseases
  - Provide risk feedback to target populations
- **To create an infrastructure to facilitate public health research**
  - Decide upon “standard of care” for consent, feedback & support services
  - Identify optimal study population(s) and recruitment approach



Early microscope

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



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*