National Human Genome Research Institute

National Institutes of Health

U.S. Department of Health and Human Services

Multi-IC Symposium on Application of Genomic Technologies to Population-Based Studies

James Battey, NIDCR
Stephen Chanock, NCI
Katrina Gwinn-Hardy, NINDS
Teri Manolio, NHGRI
Rebekah Rasooly, NIDDK
Winifred Rossi, NIA
Gerald Sharp, NIAID

Opportunities and Challenges in Applying Genomic Techniques to Population Studies

Teri Manolio, M.D., Ph.D.

Senior Advisor to the Director for Population Genomics

June 5, 2006

Background and Opportunities

- Understanding of genome structure provides unprecedented opportunities to define genetic contributions to health and disease, particularly in relation to environmental effects
- Progress in application of genomic knowledge to health hampered by two separate worlds:
  - Genomics
  - Population-based research

anonymized

anonymized

coding SNP

anonymized

indel

coding SNP

anonymized

indel

coding SNP

inversion

silencer

whole genome amplification

phenocopy

linkage disequilibrium


anonymized
indel
shotgun sequencing
coding SNP
contig
5' UTR
inversion
silencer
whole genome amplification
phenocopy
ancestral markers
linkage disequilibrium
lambda
admixture


anonymized
indel
shotgun sequencing
coding SNP
contig
5' UTR
enhancer
epigenetics
inversion
silencer
epistasis
minor allele frequency
population stratification
whole genome amplification
selective sweep
phenocopy
promoter
ancestral markers
genome wide association
linkage disequilibrium
lambda
gene-environment interaction
admixture

The Far Side

MAY

22

Thursday

Understanding only German, Fritz was unaware that the clouds were becoming threatening.

Gary Larson, *The Far Side.*

Basic Definitions

Locus: Place on a chromosome where a specific gene or set of markers reside

Gene: Contiguous piece of DNA that can contain information to make or modify 'expression' of specific protein(s)

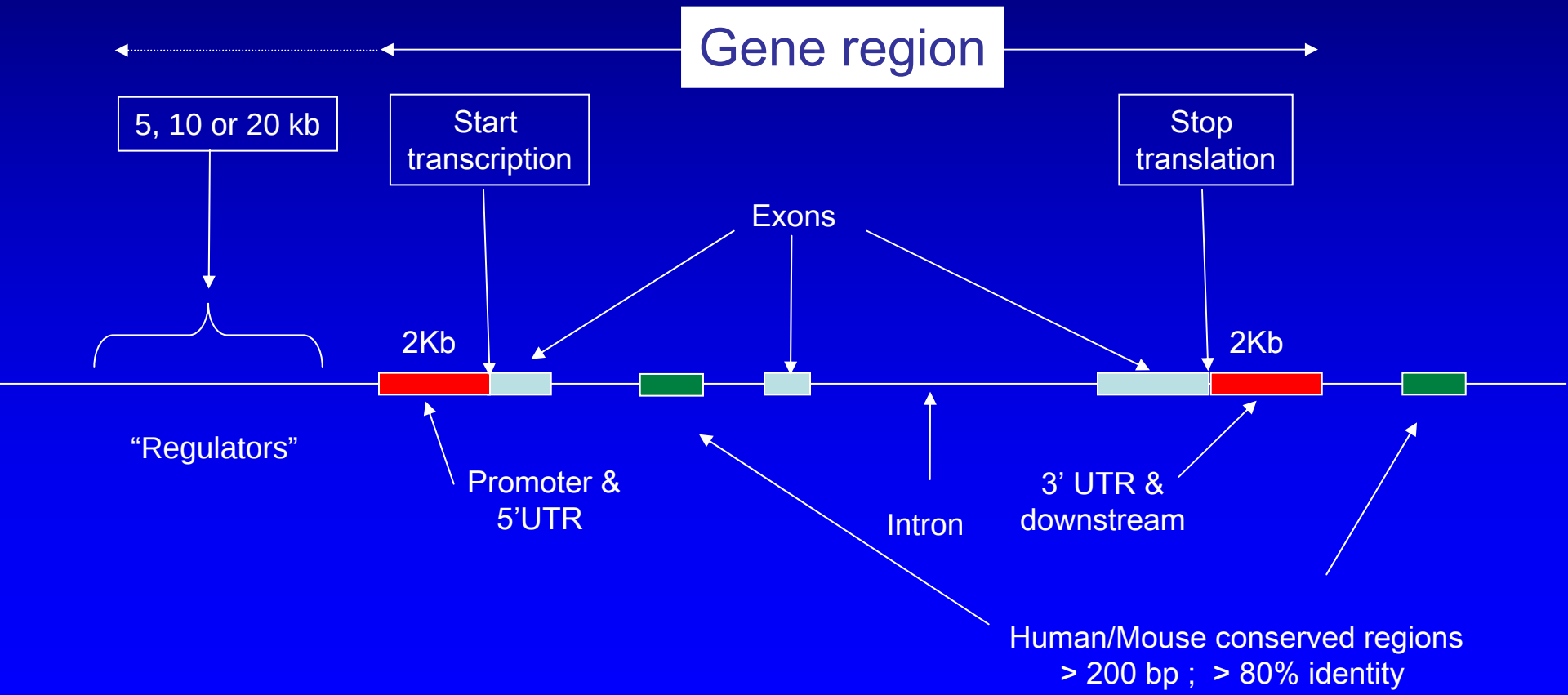
Polymorphism: Variation in the sequence of DNA among individuals

Allele: A variant form of a DNA sequence at a particular locus on a chromosome

Note: these terms were defined when we did not have access to complete DNA sequence

Courtesy S. Chanock, NCI

What is a Classical Gene?



Courtesy S. Chanock, NCI

Mutations and Single Nucleotide Polymorphisms (SNPs)

- Mutation = change in bp sequence
- Point substitution: most frequent in genome
- Point mutations occur roughly once every 10^8 replication events
- "Common" SNPs are defined as >1% minor allele frequency (MAF) in at least one population
- Rare SNPs are hard to identify and validate (probably requires sequencing), but there are likely a large number per individual

After S. Chanock, NCI

SNPs and Function

- Majority are “silent” (excellent as markers)
  - No known functional change
- Alter gene expression/regulation
  - Promoter/enhancer/silencer
  - mRNA stability, splice site variants
  - Small RNAs
- Alter function of gene product
  - Change sequence of protein

After S. Chanock, NCI

Coding SNPs (cSNPs)

- **Synonymous**: no change in amino acid previously termed "silent" but…..
  Can alter mRNA stability
  DRD2 (Duan et al 2002)
- **Nonsynonymous**: changes amino acid, conservative and radical
- **Nonsense**: change to stop codon
- **Insertion/deletion (indel)**: disrupts codon sequence, rare but disruptive

After S. Chanock, NCI

Other Types of Polymorphisms (Structural Variation)

- Duplication/expansion/repeat:
  - Minisatellites (VNTRs), 25 – 100s of bp
  - Microsatellites (STRs), 2 – 7 bp
  - Copy number variants: gains and losses of large chunks of DNA sequence, 10K – 5M bp
- Inversion: segment of DNA reversed end to end, recombination suppressed across that segment
- Translocation: segment of DNA moved to another position, especially another chromosome

After S. Chanock, NCI

Sidney Harris, http://www.sciencecartoonsplus.com/gallery.htm.

Public Health and Genomics

- Primary goal:
  Improve health and prevent disease
- Progress in:
  Defining genome structure and function
- But:
  Not ready for direct clinical or public health application in detection, treatment, and prevention of disease
- Need:
  Concerted effort to stimulate steps from gene identification to public health implementation

Identifying Risk-Related Genetic Variants on a Population Basis

"Genes are merely risk factors passed on from parents to children...."

- Determine prevalence of variants in diverse groups
- Examine associations identified in family studies, assess magnitude and independence
  - common risk factors are not strong
  - strong risk factors are not common
- Define associations with variety of phenotypes
- Identify factors modifying genotype-phenotype relationships (gene-environment interactions)

Genetic Studies in Unrelated Individuals *(pre-2004)*

- Goal: to characterize candidate genes and variants identified as related to disease
- Most effective if begun *after* disease-related genes and variants identified
- Not typically intended to “find genes”

Genetic Studies in Unrelated Individuals *(pre-2004)*

- Assess generalizability of family-based observations in population-based studies (genetic heterogeneity)
- Assess importance of allelic variation at population level (population attributable risk)
  - Allele frequency
  - Size of effect
- Identify modification of genetic association by environmental factors (GxE interaction)

Age-Adjusted Odds on Hypertension by ACE ID/DD Genotype and Sex

	DD	ID	II	P-value
Men: % HTN	53.1	45.8	44.4	
Men: OR	1.67	1.19	1.00	0.004
Women: % HTN	43.3	41.8	44.4	
Women: OR	1.01	0.80	1.00	0.15

after O'Donnell C et al, *Circulation* 1998; 97:1766-1772.

www.hapmap.org

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

ARTICLES

A haplotype map of the human genome

The International HapMap Consortium*

Inherited genetic variation has a critical but as yet largely uncharacterized role in human disease. Here we report a public database of common variation in the human genome: more than one million single nucleotide polymorphisms (SNPs) for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations, including ten 500-kilobase regions in which essentially all information about common DNA variation has been extracted. These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. We show how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution.

International HapMap Consortium, *Nature* 2005; 437:1299-1320.

A HapMap for More Efficient Association Studies

- 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.

Courtesy L. Brooks, NHGRI

Genetic Studies in Unrelated Individuals *post-2004* : Whole Genome Association Studies (WGAS, WGS, GWAS…)

- Identify genes related to complex diseases
- Complex diseases: caused by multiple genes of small effect, not amenable to family studies
- Whole genome: interrogate all variation throughout genome, two main approaches
  1. Family linkage study with 400 microsatellite markers, assumes ~10mb regions of LD
  2. Unrelated case-control study with 300-500K SNPs, assumes ~10kb regions of LD

Value of WGA Studies in Unrelated Individuals

- Unrelated individuals tend to be easier to study
- Many existing collections of population samples
  - Often extremely well-characterized
  - Often followed for long periods
  - Often diverse in origin, exposures
- Large families with common diseases remain very valuable for gene-finding
  - Not so common anymore
  - Families tend to share environmental factors more than do unrelated individuals

Whole Genome Scans (AKA Association Studies)

- Public Health Impact
- Primary Goal(s)
  - Etiology
  - Survival
  - Pharmacogenomics
- Value-added Analyses
  - Covariates
  - Biomarkers
  - Gene-environment interactions

What WGA Studies Will and Won't Do

WILL:

- Identify lots of common SNPs with statistical association to disease or trait
- Identify a blizzard of spurious associations
- Provide clues to causative genetic variants
- Provide clues to environmental modifiers
- Generate terabytes of data

WON'T:

- Identify rare SNPs associated with disease/trait
- Identify causative genetic variants
- Explain biology or function

What Sequencing Studies Will and Won't Do

WILL:

- Identify rare SNPs in persons with disease/trait
- Suggest clues to causative genetic variants
- Generate terabytes of data

WON'T:

- Identify causative genetic variants
- Explain biology or function
- Define importance of variants in population

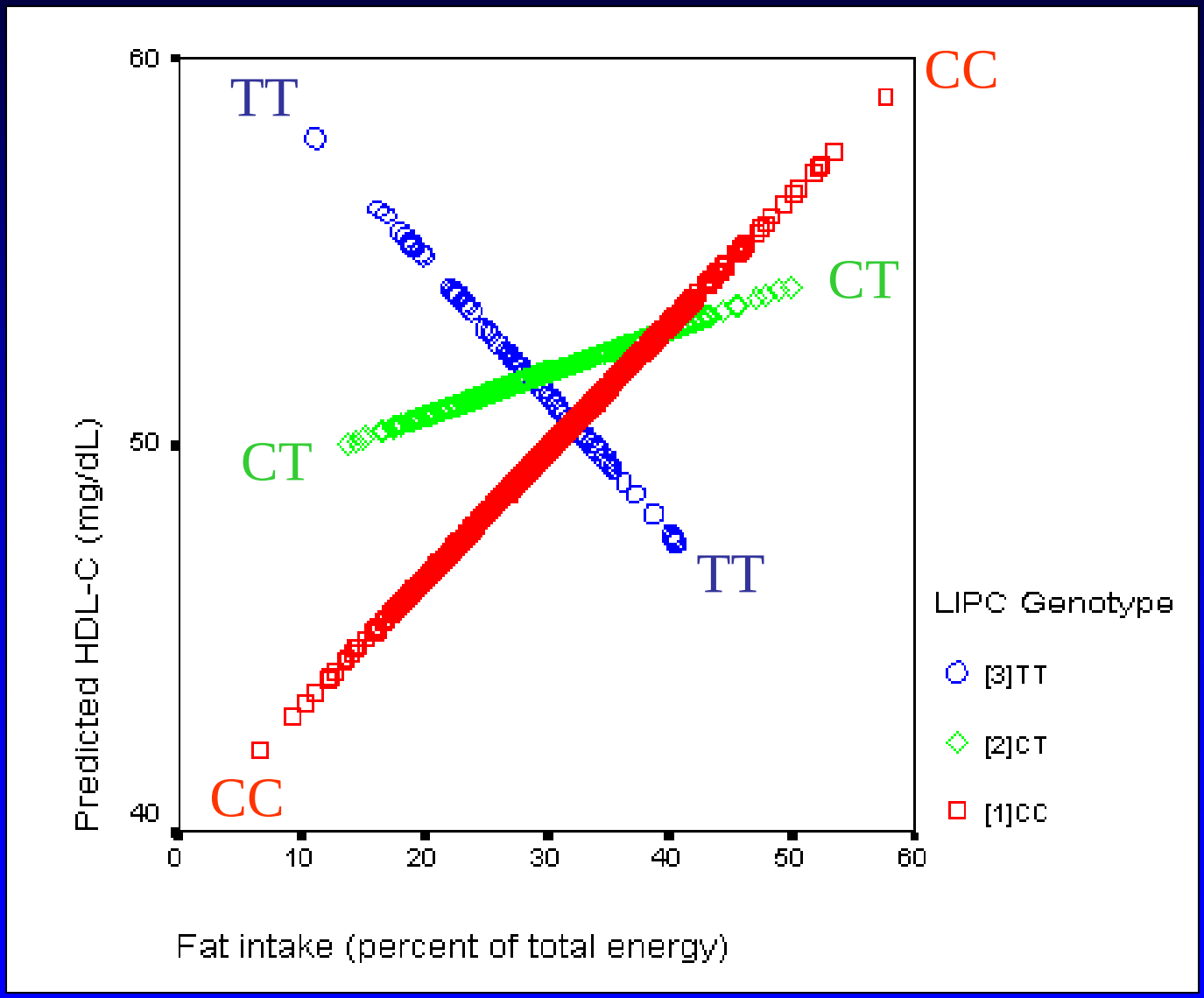
Why are Gene-Environment Interactions so Important to Public Health?

- Environmental and behavioral changes interacting with genetic predisposition have likely produced most of the recent epidemics of chronic diseases
- GxE may be key in reversing their course, by suggesting approaches for modifying effects of deleterious genes
- Future public health measures may focus on avoiding deleterious environmental exposure, especially in genetically susceptible persons

Why are Gene-Environment Interactions so Important to Research?

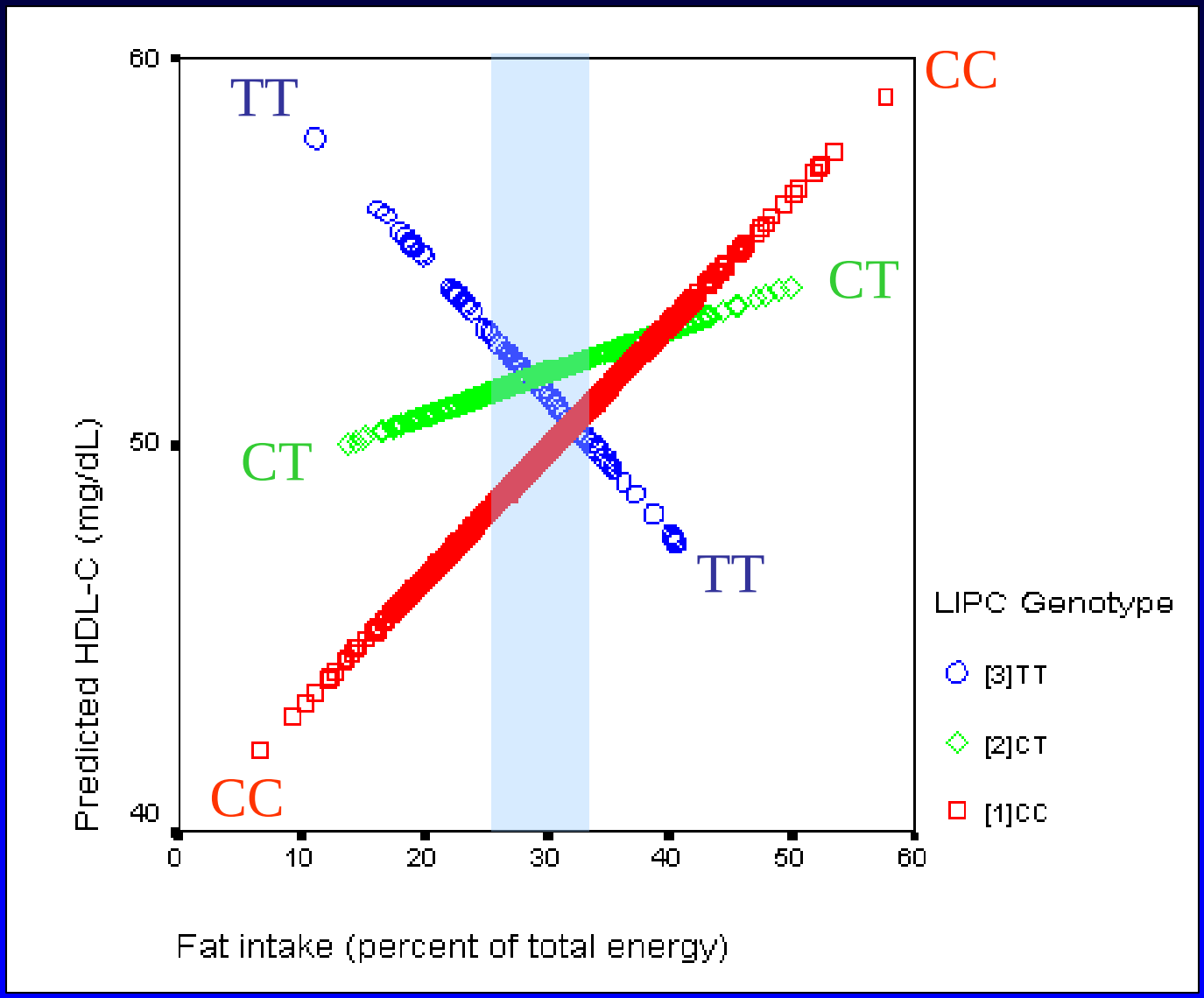
- Can mask detection of genetic (or environmental) effect if they are not identified and controlled for
- Can lead to inconsistencies in disease associations in different populations with:
  - Different environmental exposures that modify the effect of a genetic variant
  - Different prevalences of genetic variants that modify the effect of an environmental exposure

Is LIPC Genotype Related to HDL-C?



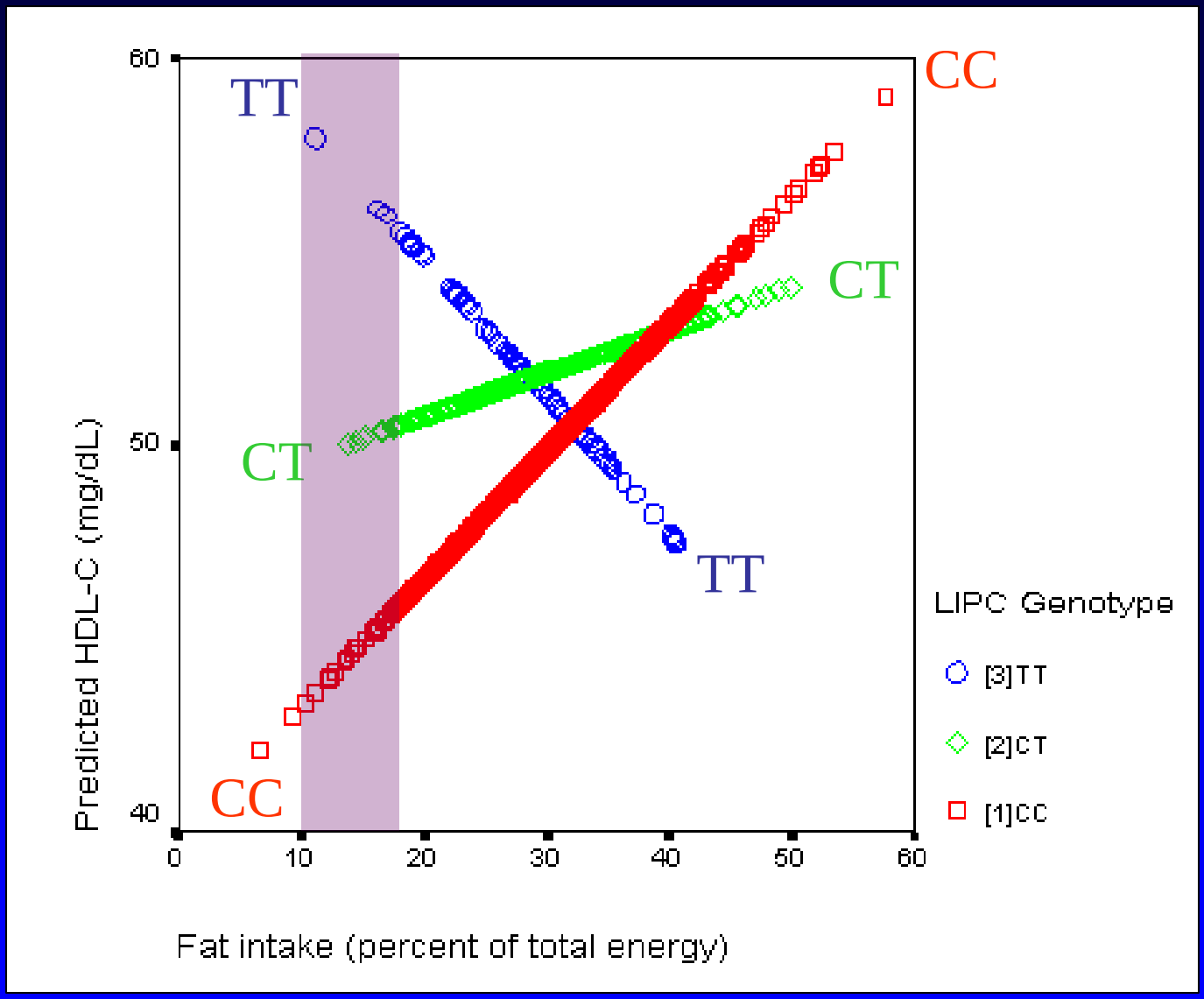
Ordovas et al, *Circulation* 2002; 106:2315-2321.

Is LIPC Genotype Related to HDL-C?



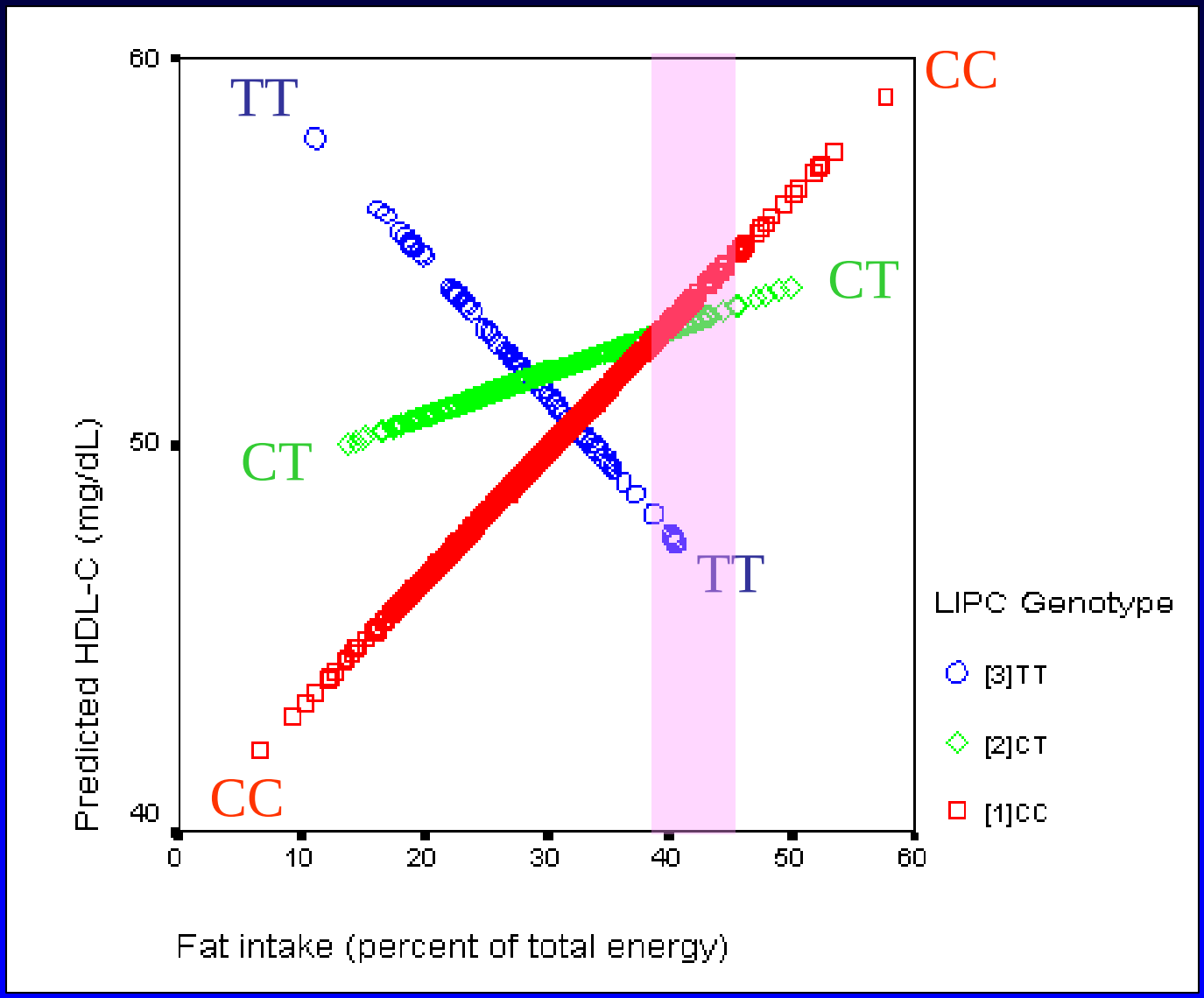
Ordovas et al, *Circulation* 2002; 106:2315-2321.

Is LIPC Genotype Related to HDL-C?



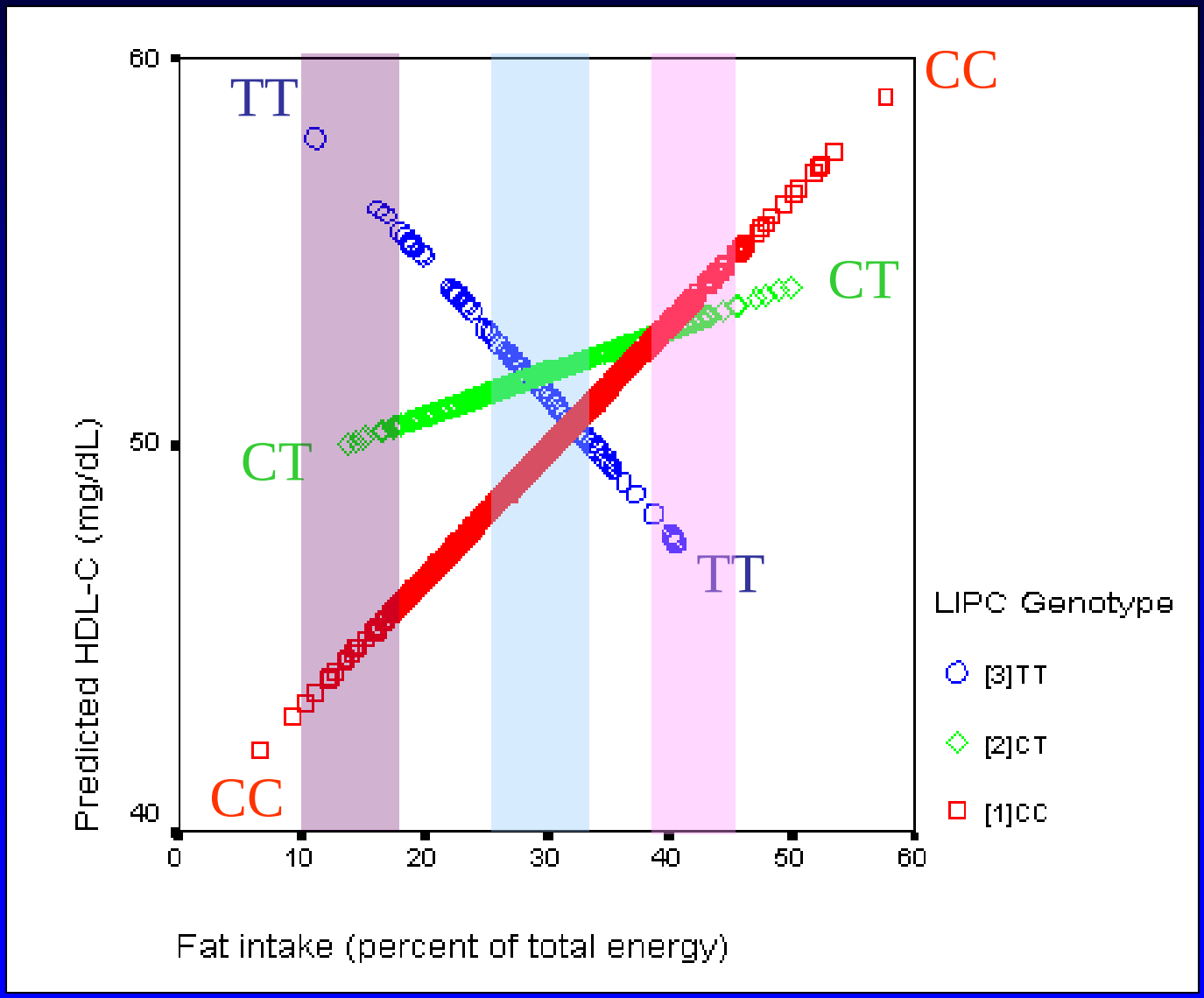
Ordovas et al, *Circulation* 2002; 106:2315-2321.

Is LIPC Genotype Related to HDL-C?

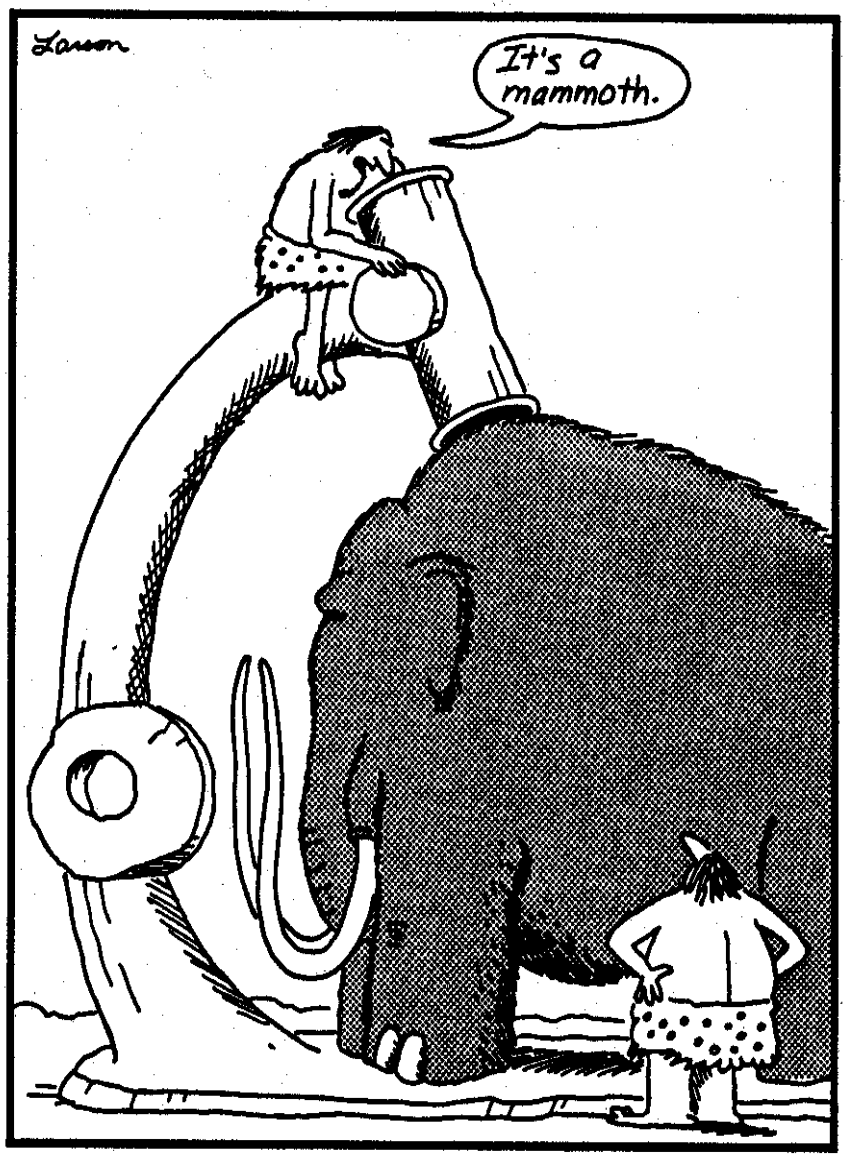


Ordovas et al, *Circulation* 2002; 106:2315-2321.

Is LIPC Genotype Related to HDL-C?



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



Early microscope

Gary Larson, *The Far Side.*

Challenges in Applying Genomic Technologies to Population Studies

- Which technologies? How to evaluate ever-changing technologies and ensure most reliable and cost-effective applied to your IC's studies?
- How to manage data? How to increase access and usefulness to outside groups, and enhance comparability across studies?
- How to ensure adequate consent and human subjects approval for future studies? How to protect participant confidentiality?

Challenges in Applying Genomic Technologies to Population Studies (2)

- Which population studies? How to facilitate use of large-scale studies for multiple ICs' needs?
- Which WGA or sequencing studies? How to select and prioritize appropriate population samples and follow-up studies?
- How to fund and coordinate addition of genomic technologies to ongoing studies? What issues related to consent, confidentiality, and IP need to be addressed in existing vs *de novo* studies?

Symposium Website

http://genome.gov/pages/extranets/PopulationGenomicsTraining/

Username: training

Password: summer06

Stay tuned for migration, updates…