

Locating Polymorphic Determinants of Gene Expression

Richard Spielman

Department of Genetics

**University of Pennsylvania
School of Medicine**

Outline

- **Model for complex traits/diseases**
- **Quantitative trait:**
 - **Recall: Red cell acid phosphatase**
- **Quantitative trait: gene expression phenotype**
 - **Linkage – Genome-wide**
 - **Association / QTDT at target genes**

SIMPLE GENETIC TRAIT OR DISEASE

Gene

Trait / Disease



COMPLEX GENETIC TRAIT

Gene(s)

Trait / Disease

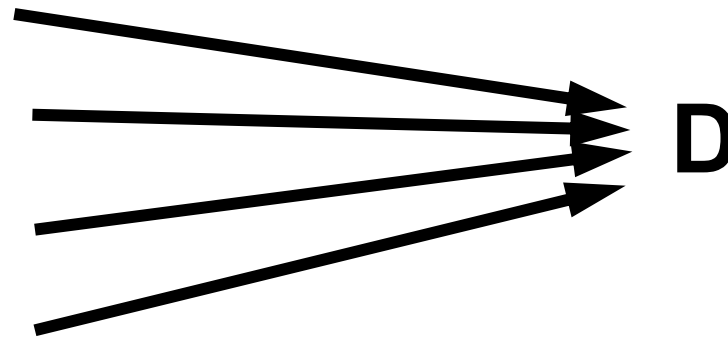
Gene 1

Gene 2

Gene 3

Gene 4

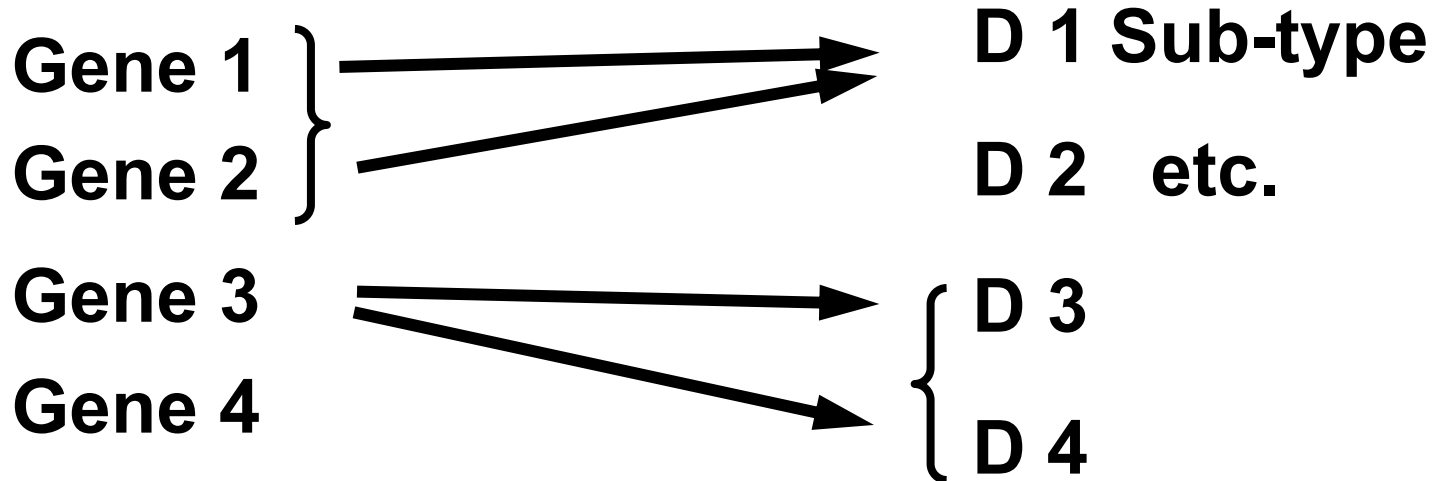
Etc.



MORE COMPLEX: Many to 1 and 1 to Many

Genes

Trait / Disease



COMPLEX DISEASE: FANCIFUL SIMPLIFIED MODEL

Genes

Diabetes

Gene 1



D 1 (Sub-type)

Gene 2 (PPAR γ)



D 2: Ins resistance

Gene 3 (Calpain-10)



Gene 4



D 3: Ins secretion

Overview

Variation – the subject matter

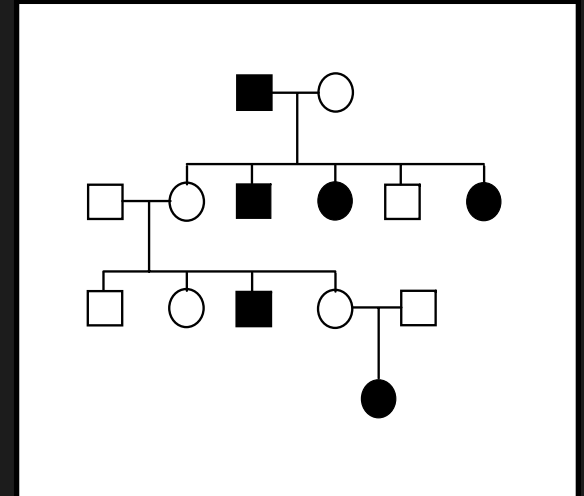
- **Classic example of quantitative variation**
 - **Genetics/genomics of gene expression level**
 - **At level of linkage**
 - **And association**
 - **Ultimately: traits like blood pressure, cholesterol level**
 - **Expression a “simpler” quantitative trait**
 - **Prototype for other complex traits**
-

Outline

- Model for complex traits/diseases
- **Quantitative trait:**
 - **Recall: Red cell acid phosphatase**
- Quantitative trait: Gene Expression
 - Linkage – Genome-wide
 - Association / QTDT at target genes

Phenotypic Variation

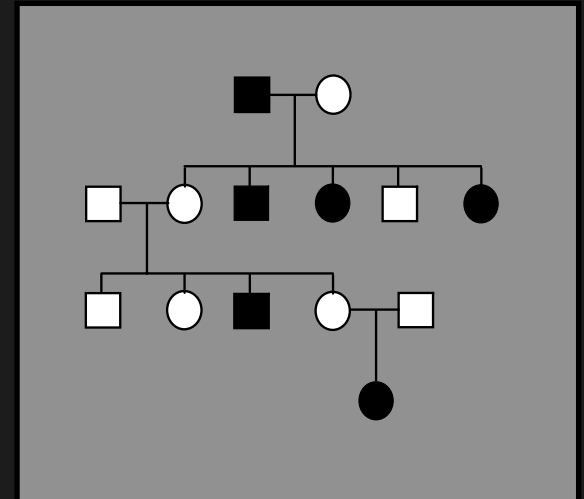
- Qualitative trait –
“Dichotomous”
 - Disease
 - Normal variation (e.g., eye color)



Phenotypic Variation

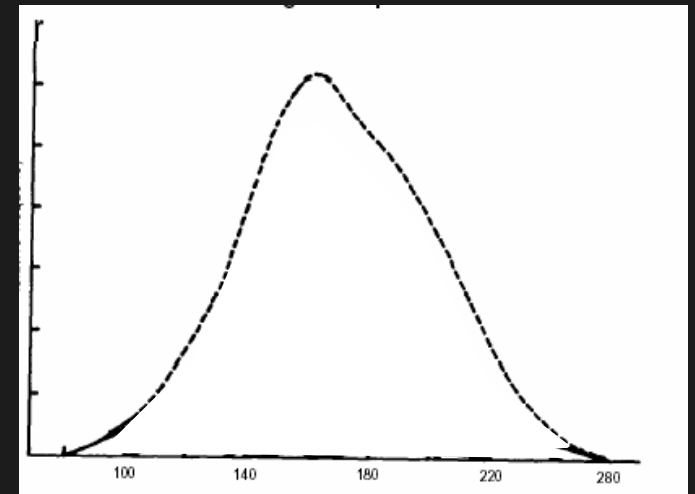
- Qualitative trait – “Dichotomous”

- Disease
- Normal variation (e.g., eye color)



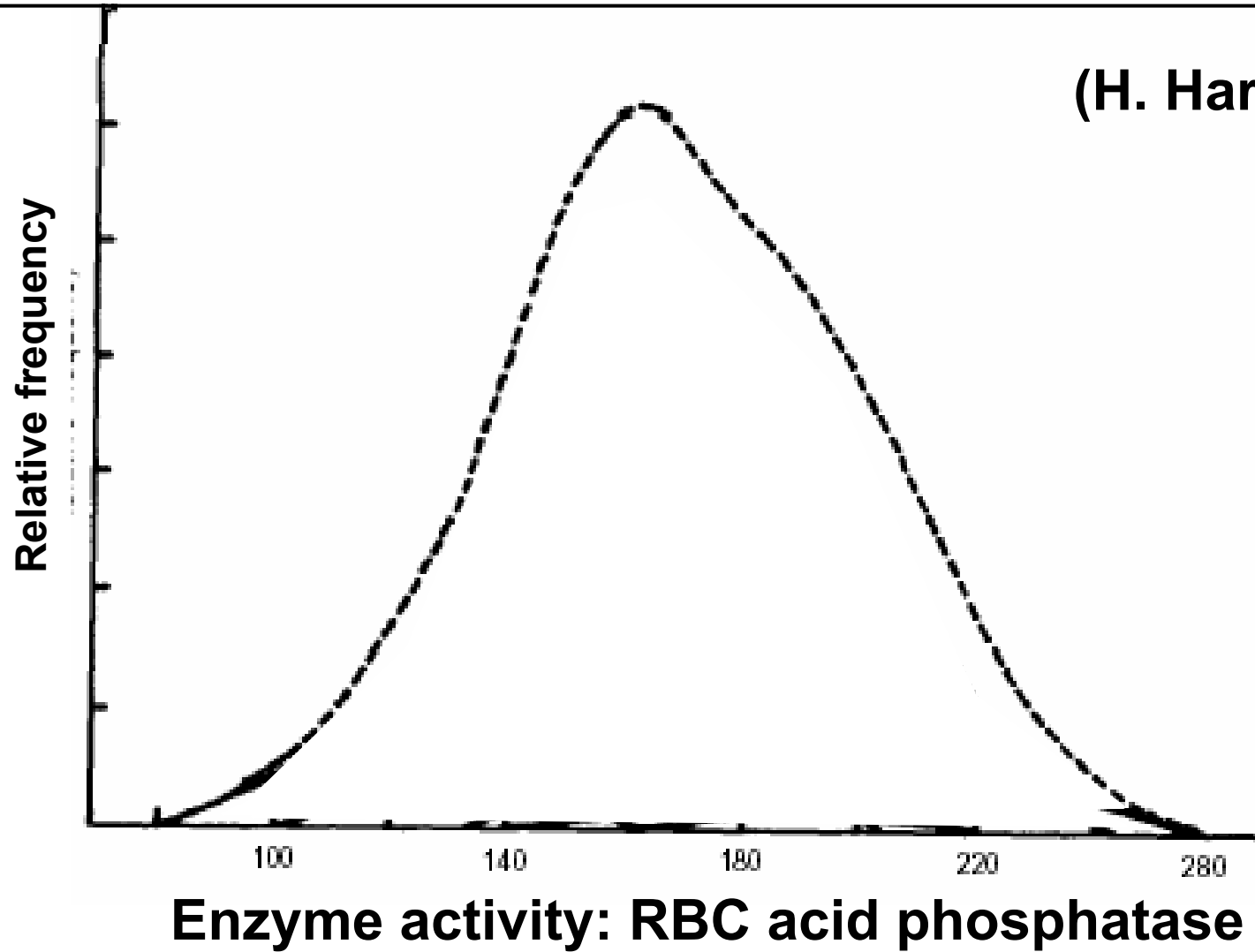
- Quantitative trait - “continuous”

- Normal variation (e.g., height, enzyme level, serum cholesterol)

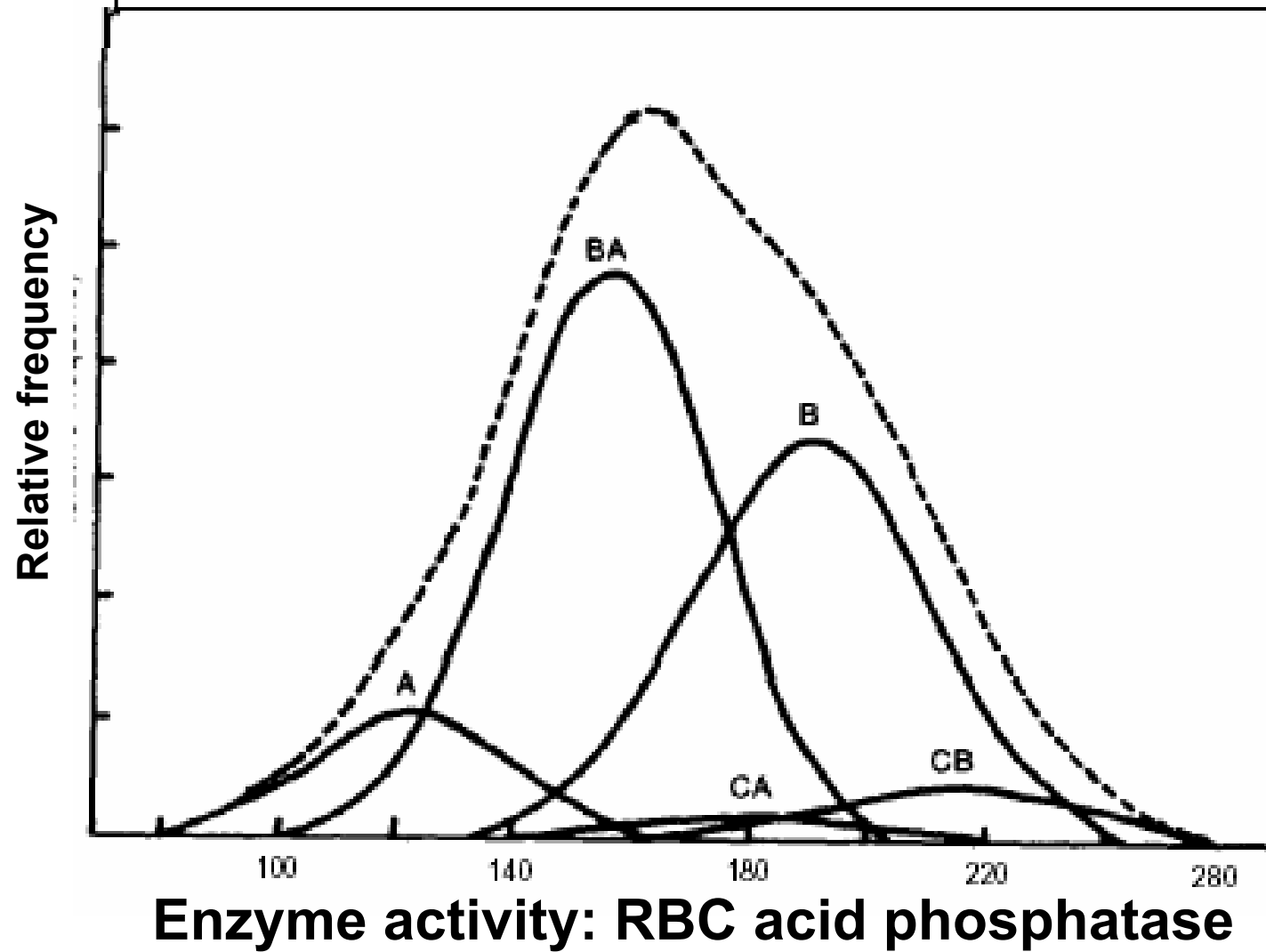


Activity of Red Cell Acid Phosphatase in General Population

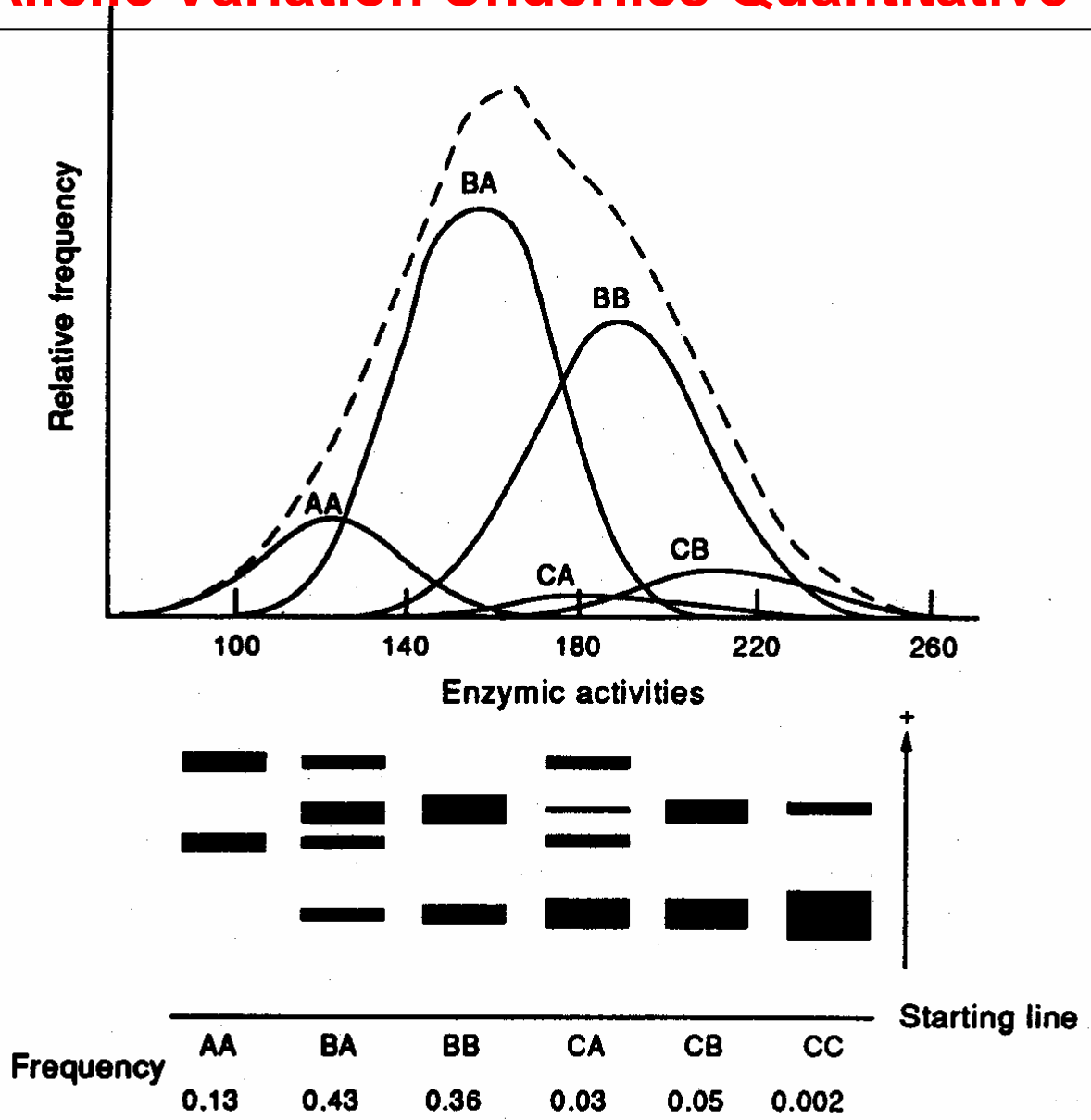
(H. Harris, 1966)



Population Distribution: Component Genotypes



Allelic Variation Underlies Quantitative Variation



Genomic View of Genetic Variation in Gene Expression

A “multi-trait” genome scan

- Genome scan for **differential gene expression at baseline)**
 - Prospect: assess variation at many genes simultaneously
 - Therefore allow: global approach to combined effects
-

Genetics of Variation in Gene Expression: Collaborators

CHEUNG LAB

Michael Morley
Joshua Burdick
Teresa Weber

SPIELMAN LAB

Cliona Molony
Kathy Ewens
Jim Devlin

The Children's Hospital of
Philadelphia

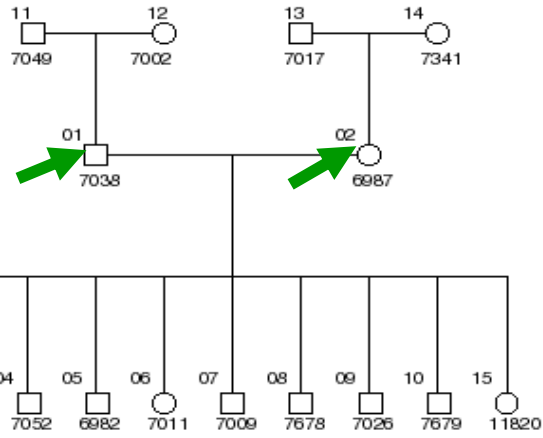
University of
Pennsylvania

Genome-Wide Genetic Analysis of Variation in Gene Expression

- **Desired linkage analysis dictates**
 - **CEPH families – 3 generations, very large sibships**
 - **LCLs – immortalized cell lines**
 - **Analysis: Linkage search for “QTL”**
 - **Analogy: Serum glucose**
 - **For many traits at once**
-

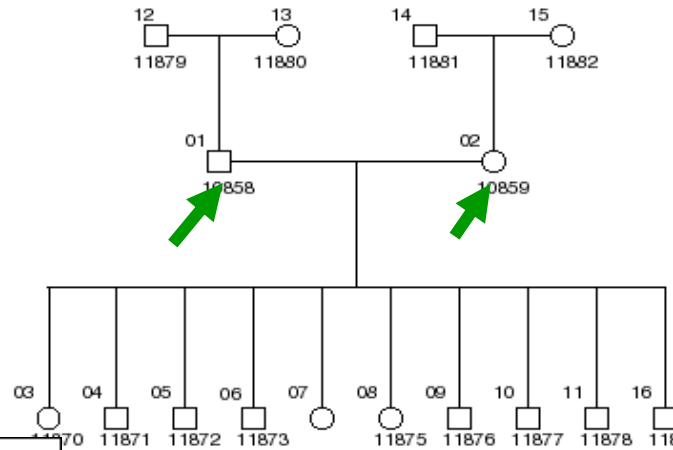
CEPH/Utah Pedigree 1333

9 SIBS



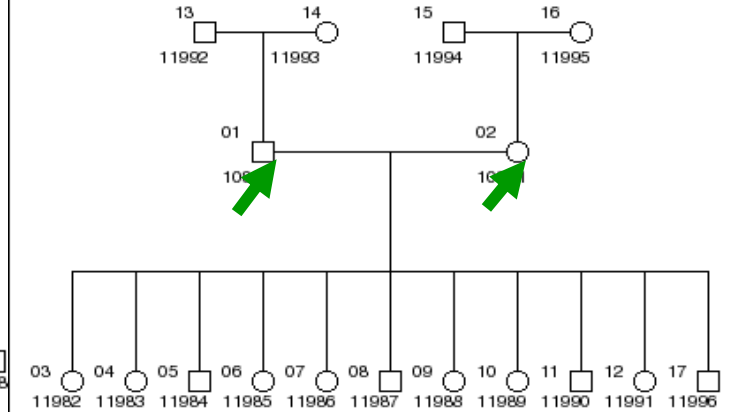
CEPH/Utah Pedigree 1347

9 SIBS



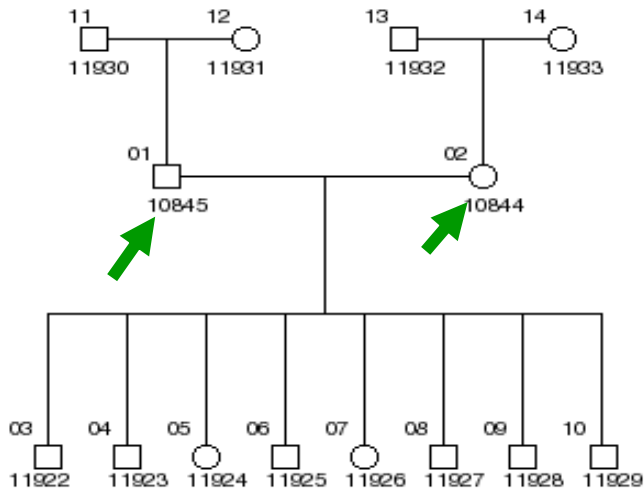
CEPH/Utah Pedigree 1362

SIBS = 11



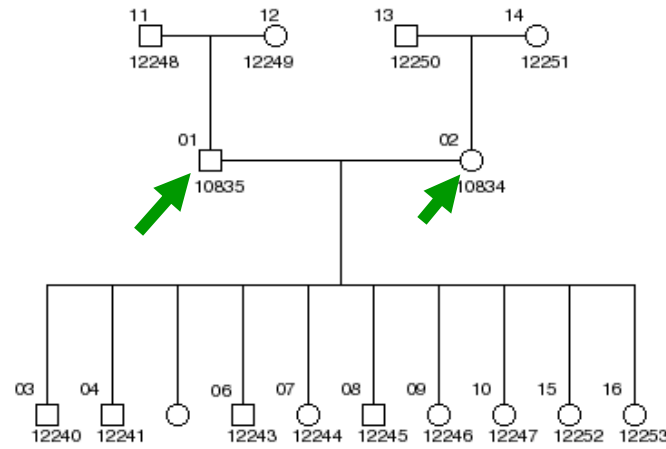
CEPH/Utah Pedigree 1424

SIBS = 8



CEPH/Utah Pedigree 1416

SIBS = 9



(Jean Dausset)

QUANTITATIVE PHENOTYPES = GENE EXPRESSION LEVELS

Affymetrix Genome Focus arrays (~8,500 genes) in 94 unrelated individuals (grandparents) in duplicate



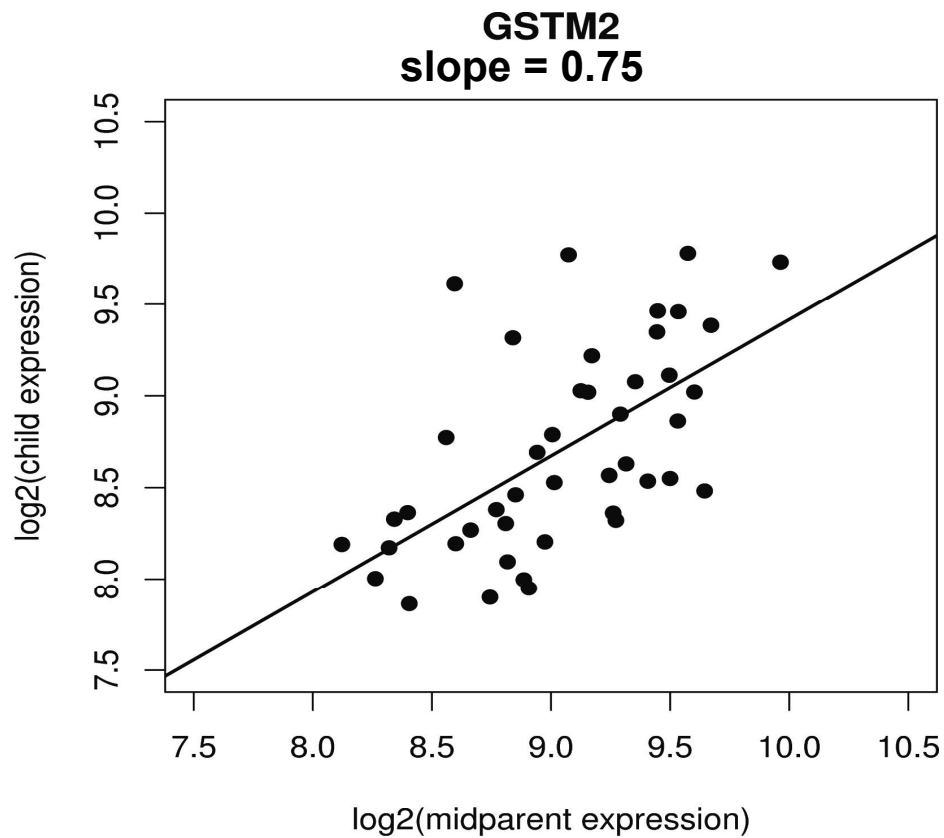
Focused on **3,554 genes – variable expression levels:**

$$\frac{\text{Variance between individuals}}{\text{Variance "within" individuals (replicates)}} > 1$$



Analyzed levels of these 3,554 genes in members of 14 CEPH Utah families

Resemblance between Offspring and Parents (46 Parent-Offspring Trios)



GENE	Narrow-sense heritability
<i>HSD17B12</i>	1.0
<i>LOC64167</i>	1.0
<i>ATP2A3</i>	0.9
<i>ICAP-1A</i>	0.8
<i>UQCRC2</i>	0.8
<i>CLSTN2</i>	0.8
<i>KLF5</i>	0.8
<i>CTSH</i>	0.8
<i>LSM7</i>	0.8

Outline

- **Model for complex traits/diseases**
- **Quantitative trait:**
 - **Recall: Red cell acid phosphatase**
- **Quantitative trait: Gene expression phenotype**
 - **Linkage – Genome-wide**
 - Association / QTDT at target genes

MARKER GENOTYPES, MAPPING

Marker genotypes

- **SNP genotypes from TSC: 2,756 SNP markers covering the autosomes**

Linkage mapping

- **Genome scans to determine the chromosomal locations linked to the expression phenotypes**
- **Haseman-Elston analysis of sib-pair differences (by S.A.G.E., using “sibpal” module)**

Mapping Genetic Elements Contributing to Expression Variation

- **Determinants of expression level**
 - Closely linked to target gene
 - Distantly linked to target gene
 - Unlinked to target
 - **Framework for phenotypic variation (including disease susceptibility)**
 - Focus on expression level, not sequence variant
 - Ultimately: variant sequences are responsible (incl SNPs)
-

LINKAGE APPROACH: **Recap**

- Consider expression level of each gene as a quantitative trait.
- To determine the chromosomal locations linked to the expression levels.
- ~3,500 genes (phenotypes, genome scans)
- 14 CEPH families
- ~3,000 SNP markers genome-wide (TSC)
- S.A.G.E. programs: SIBPAL

GENOME SCAN RESULTS

Among 3,554 expression phenotypes: Most significant evidence for linkage (genome-wide $P < 0.001$, or $P < 0.05$)

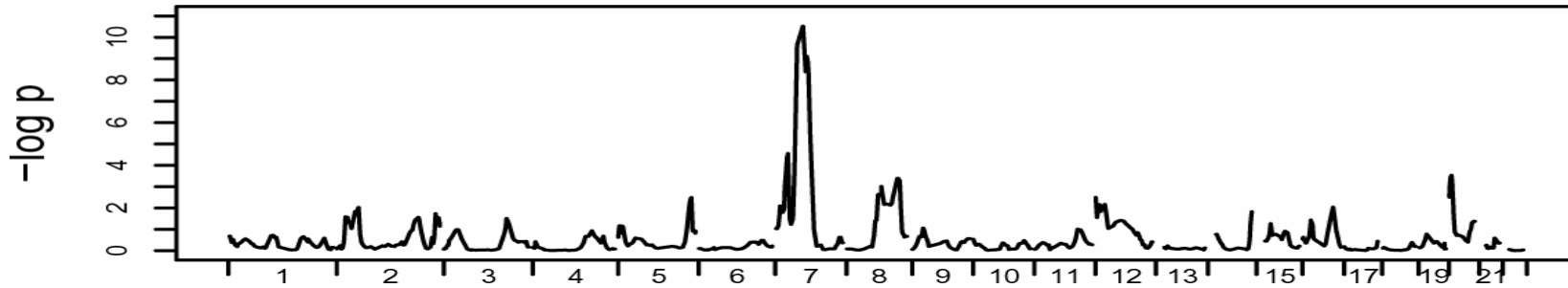
# Phenotypes	P (pointwise)	LOD	P (genome-wide)
142	4.3×10^{-7}	~ 5.3	0.001
984	3.7×10^{-5}	~3.4	0.05

142 Phenotypes with $P = 4.3 \times 10^{-7}$ and beyond

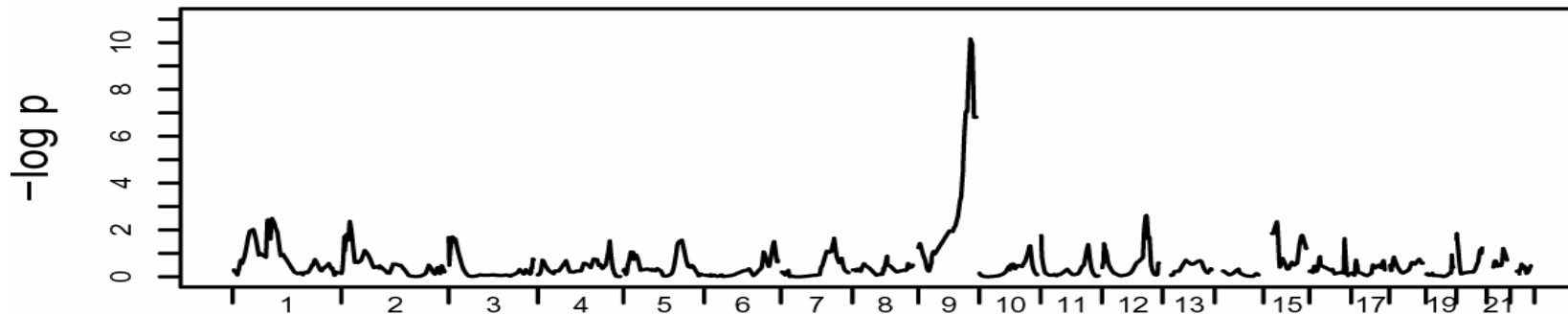
P	Phenotypes with at least one marker beyond P
< 10^{-11}	2
< 10^{-10}	6
< 10^{-9}	5
< $4.3 \text{ E-}7$	129

27 (19%): cis-acting elements, linkage \pm 5 Mb gene
110 (78%): trans-acting
5 (3.5%): multiple regulators

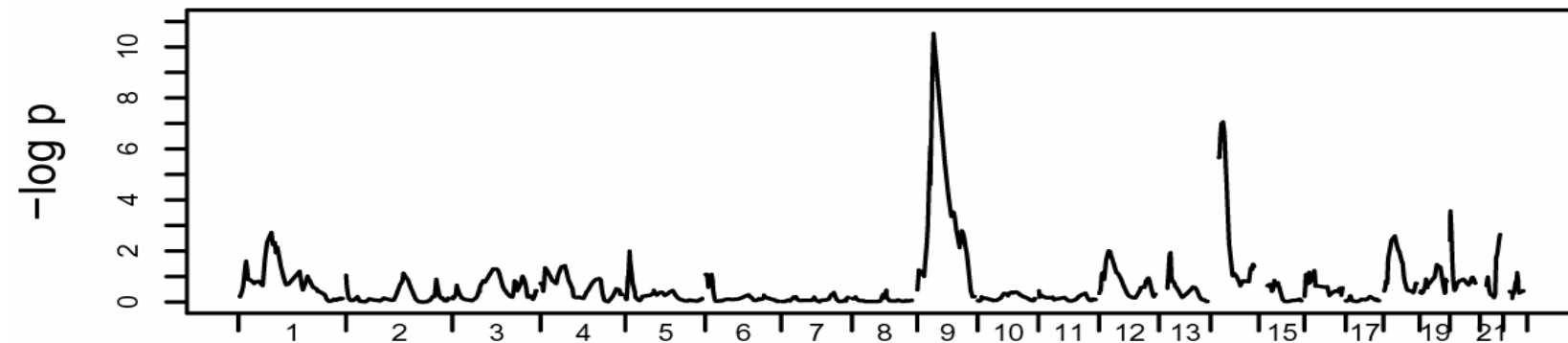
***PSPHL* (Chr 7) “cis”**

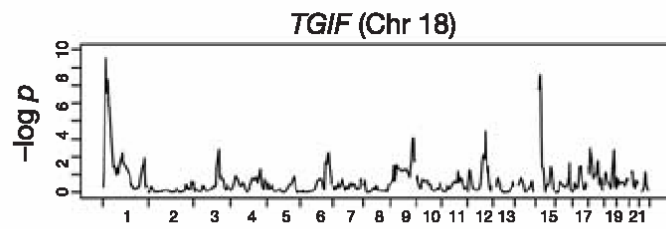
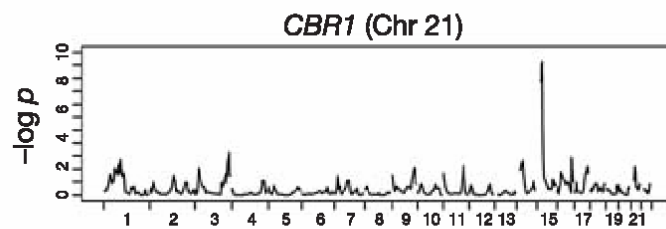
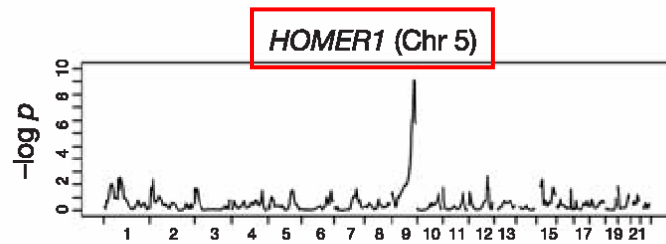
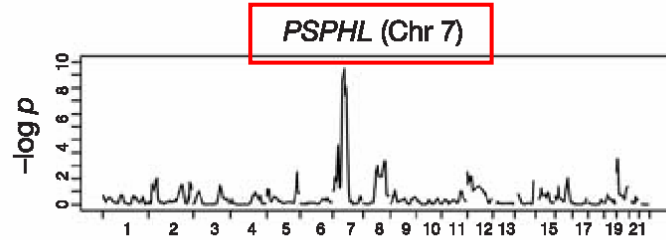
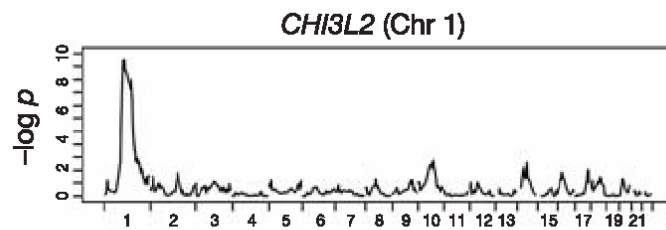


***HOMER1* (Chr 5) “trans”**



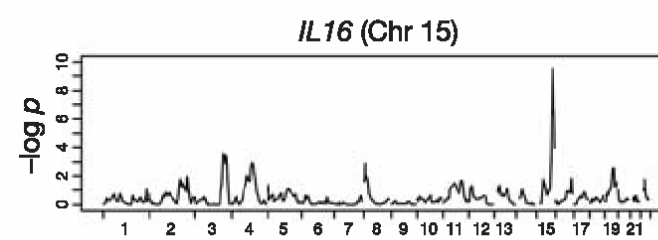
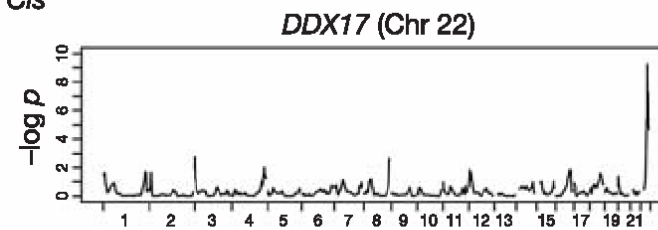
DSCR2* (Chr 21) **multiple*



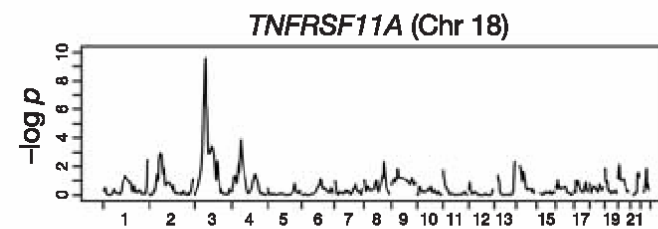
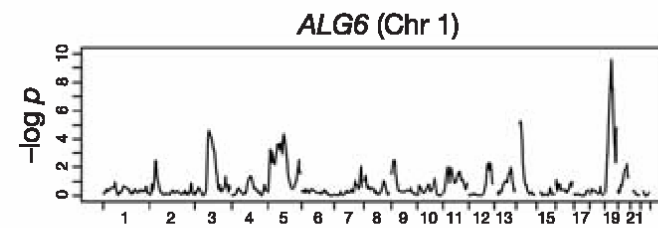


Chromosome

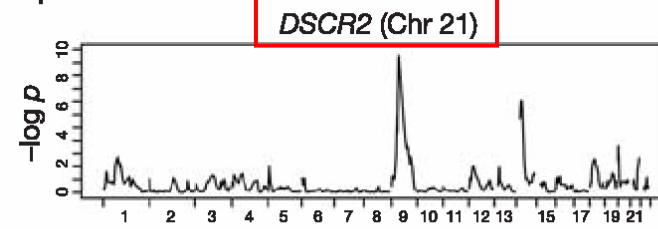
Cis



Trans



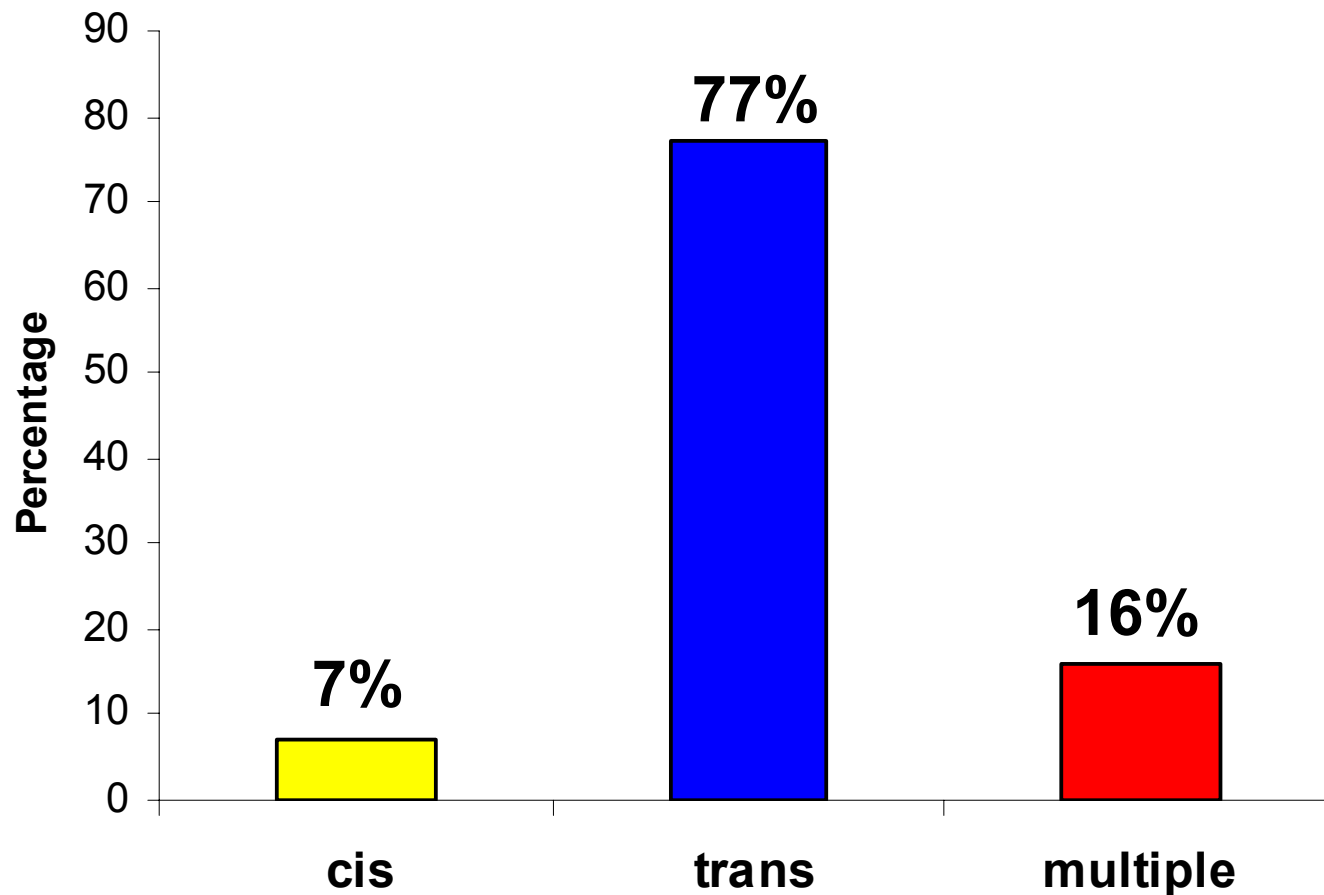
Multiple



Chromosome

CIS AND TRANS REGULATORS

Among the 984 expression phenotypes $P < 3.7 \times 10^{-5}$:



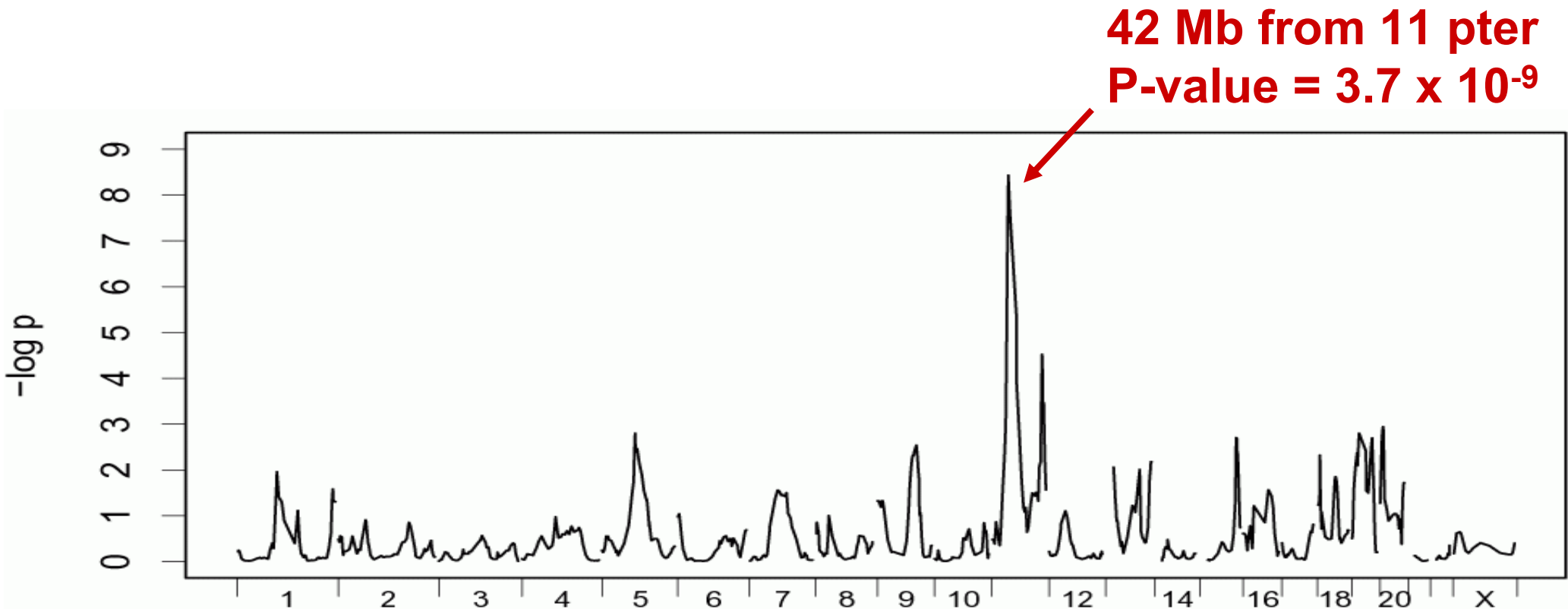
Outline

- **Model for complex traits/diseases**
- **Quantitative trait:**
 - **Recall: Red cell acid phosphatase**
- **Quantitative trait: Gene expression phenotype**
 - **Linkage – Genome-wide**
 - **Association / QTDT at target genes**

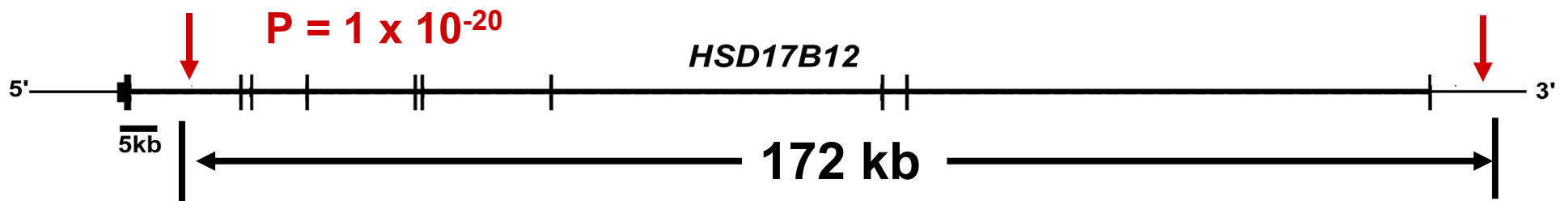
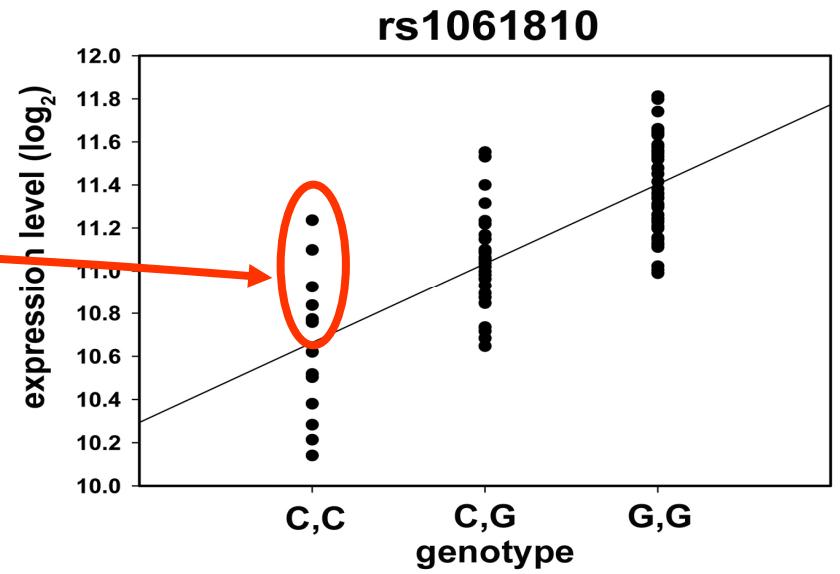
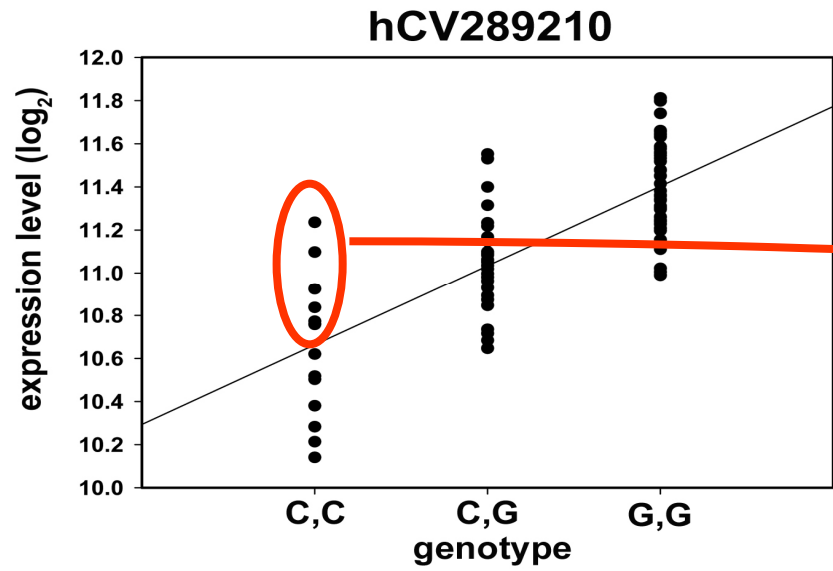
CONFIRMATION OF LINKAGE RESULTS WITH ASSOCIATION ANALYSIS

- **Followed up 17 phenotypes that show “cis”-linkage**
- **Typed new SNPs near or within the target gene**
- **Population and family-based association analysis (QTDT)**

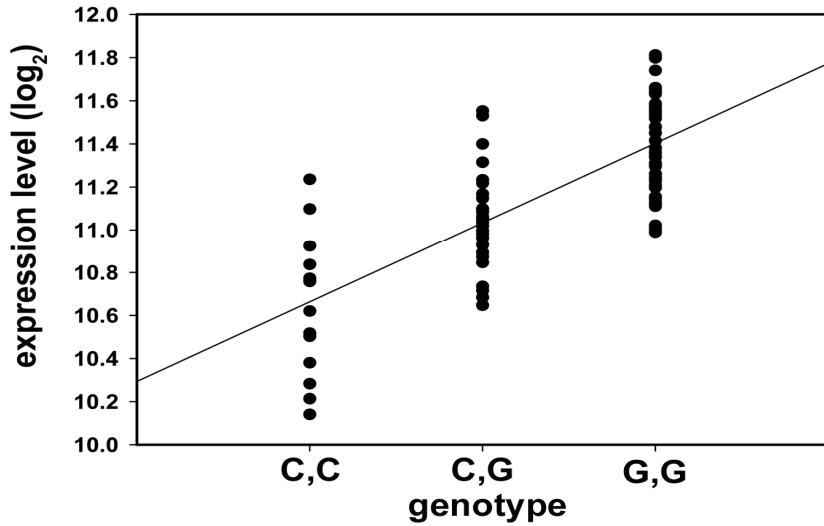
Linkage with HSD17B12 Expression Level



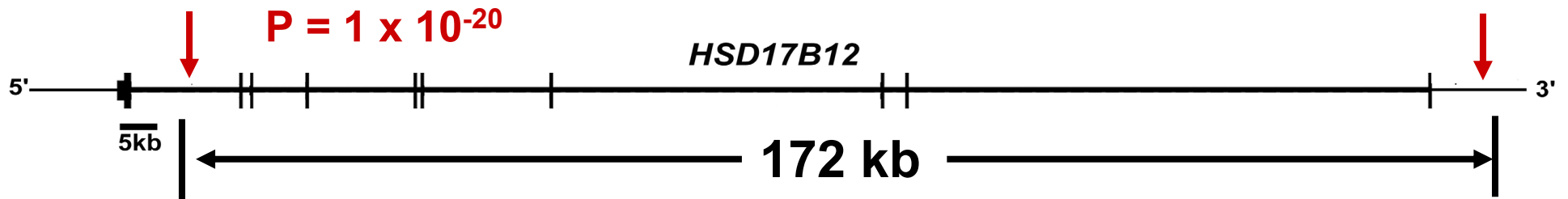
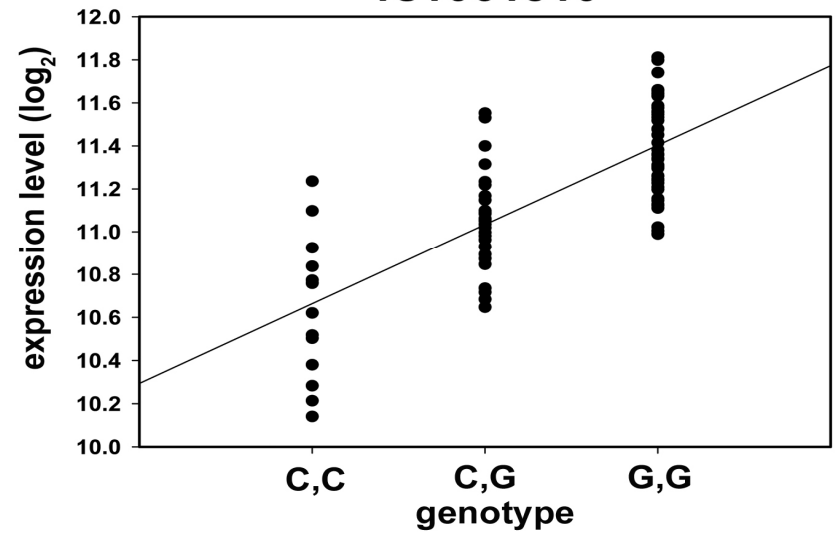
**HSD17B12 target gene: at 44 Mb from 11 pter
Implies cis-regulation**



hCV289210



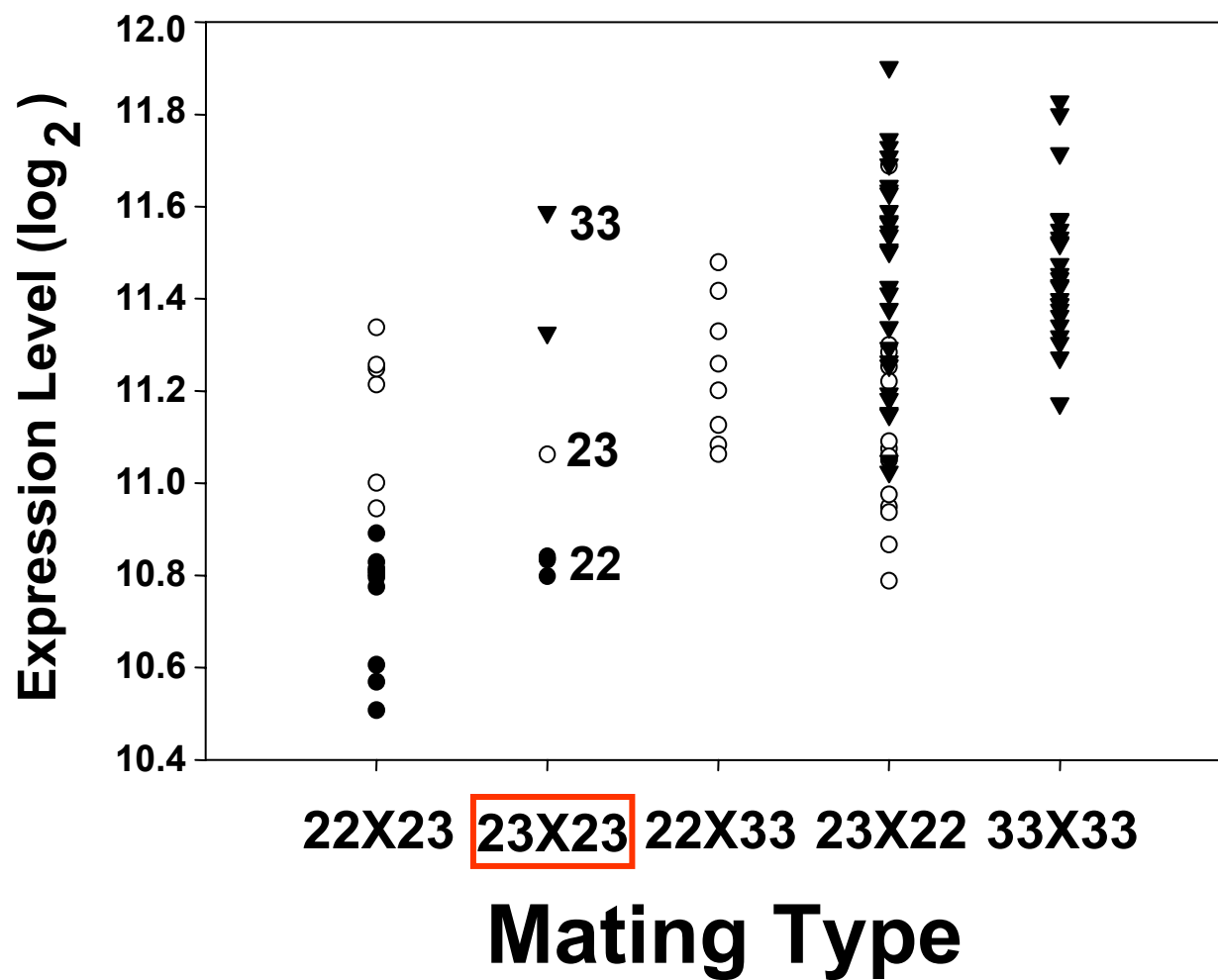
rs1061810



Family-based association by QTDT

$P = 1 \times 10^{-32}$

Parents and Offspring

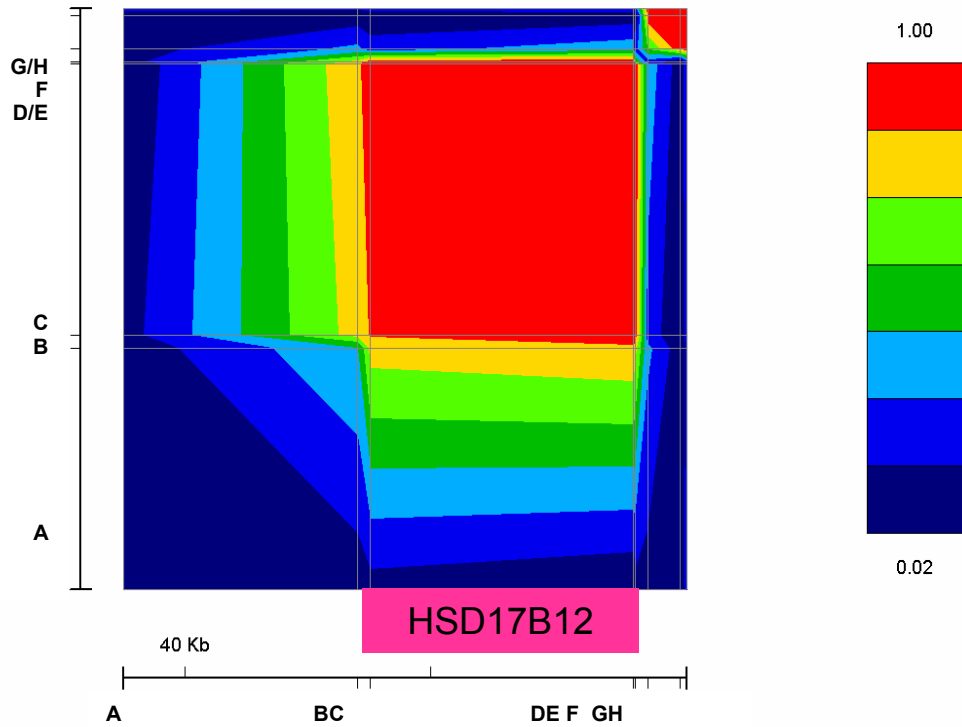


CONFIRMATION OF LINKAGE RESULTS WITH ASSOCIATION ANALYSIS

- Followed-up 17 phenotypes that show “cis”-linkage
- Typed new SNPs near or within the “target” gene
- Population and family-based association analysis (QTDT)
- **Results: 14 (82%) genes showed significant evidence ($P < 0.01$) of allelic association by QTDT:**
 - **Confirmation of linkage**
 - **Differential allelic gene expression**

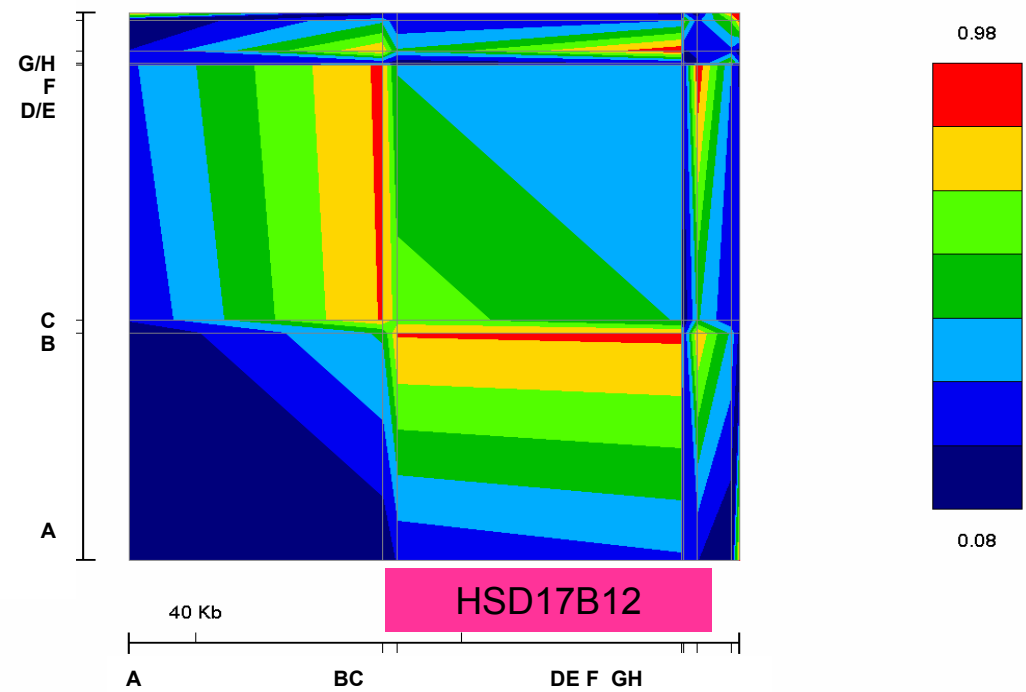
EXTENT OF LD – *HSD17B12*

CEPH



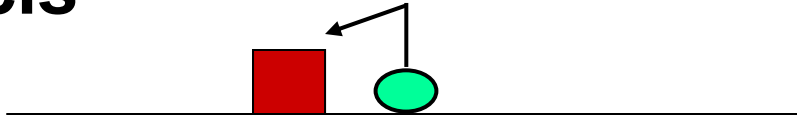
$D' = 1$

African-Americans

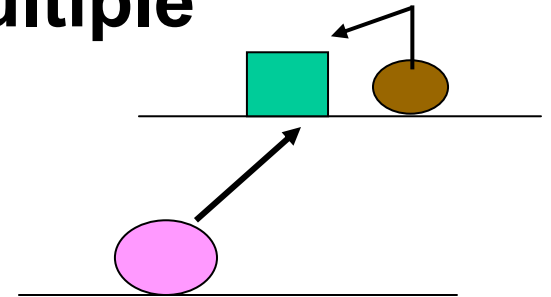


$D' = 0.116$

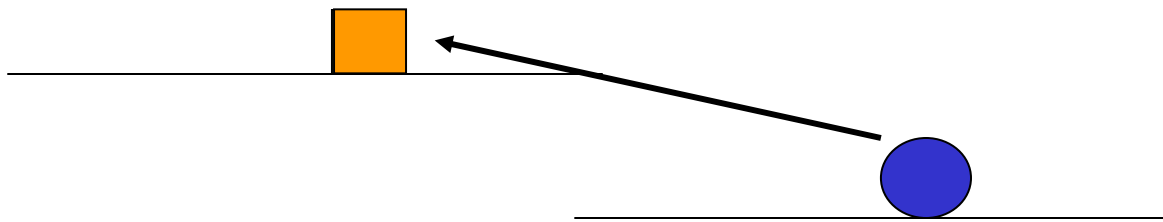
cis



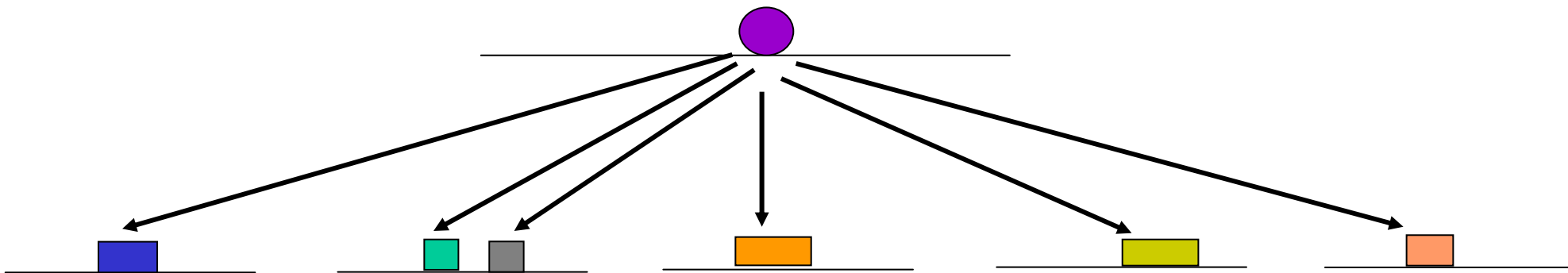
multiple



trans



Master regulators?



Key Features

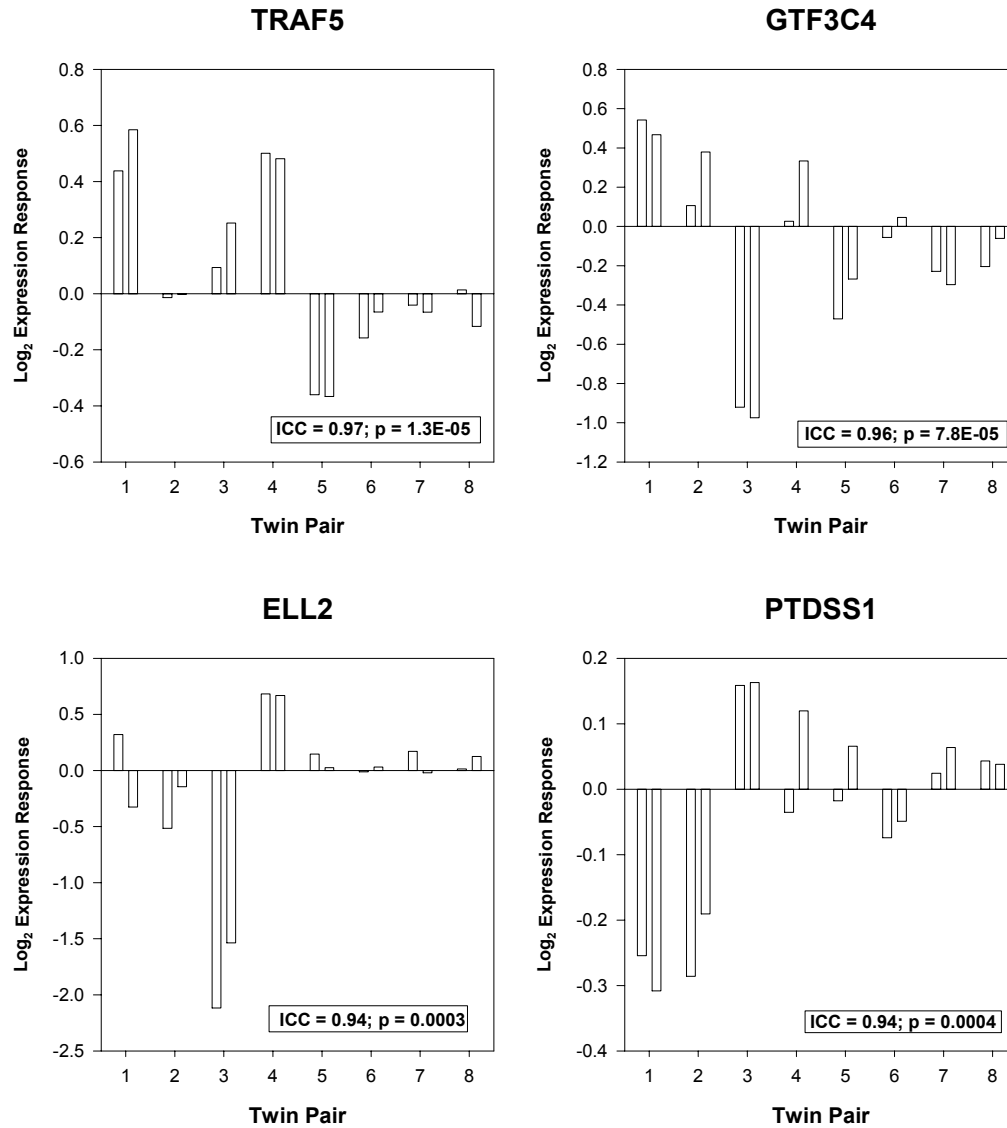
Genetic Variation in Expression Phenotypes

- Germ-line, not somatic
 - Genome-wide linkage analysis
 - Cis- or trans-acting control
 - Other patterns – many-to-one, one-to-many
-

Extensions – Beyond Baseline Variation

- **LCLs respond if tickled**
 - **Ionizing radiation (Vivian Cheung)**
 - **Drugs - (“Pharmaco-”)**
 - **Natural “signals” – glucose, cholesterol, insulin, etc.**
- **Have begun with glucose at “high” and “low” levels – genetic variation in response?**

Expression Response to Glucose: Expression (hi) – Expression (lo)



Expression response to high glucose in 8 MZ twin pairs, for TRAF5 (TNF receptor-associated factor 5), GTF3C4 (general transcription factor IIC, polypeptide 4), ELL2 (elongation factor, RNA polymerase II, 2) and PTDSS1 (phosphatidylserine synthase 1).

Summary

- **Model for complex traits**
 - **By extension, for diseases**
- **Quantitative trait:**
 - **Recall: Red cell acid phosphatase**
- **Quantitative trait: Gene expression phenotype**
 - **Linkage – Genome-wide**
 - **Association / QTDT at target genes**

Genetics of Variation in Gene Expression: Collaborators

CHEUNG LAB

Michael Morley
Joshua Burdick
Teresa Weber

SPIELMAN LAB

Cliona Molony
Kathy Ewens
Jim Devlin

The Children's Hospital of
Philadelphia

University of
Pennsylvania

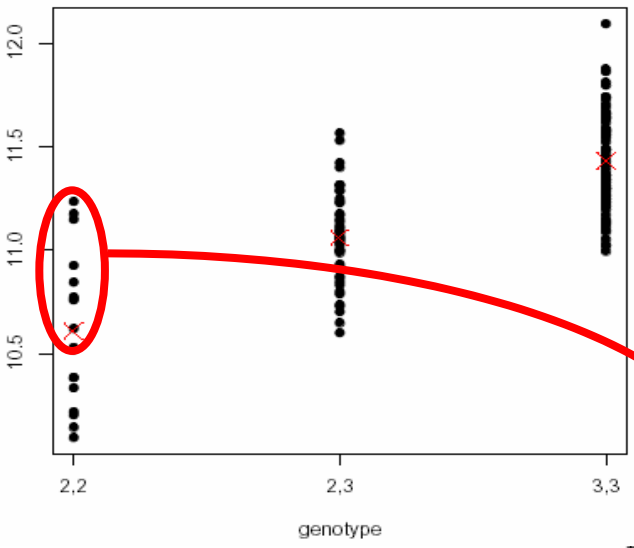
SUMMARY

SUMMARY

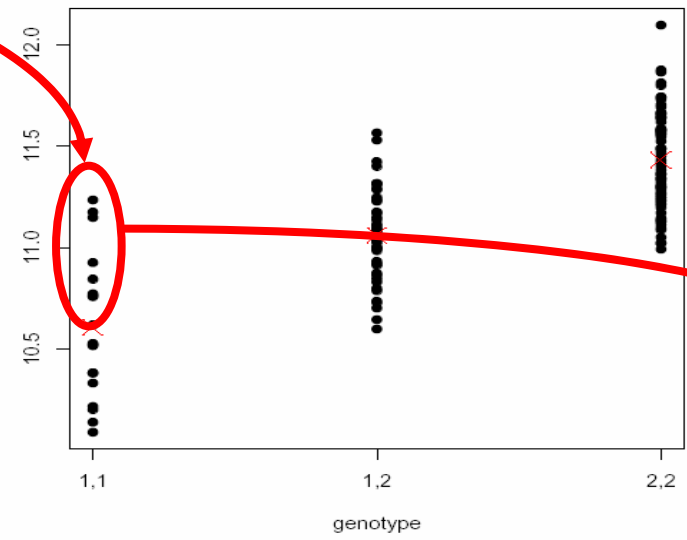
175 kb



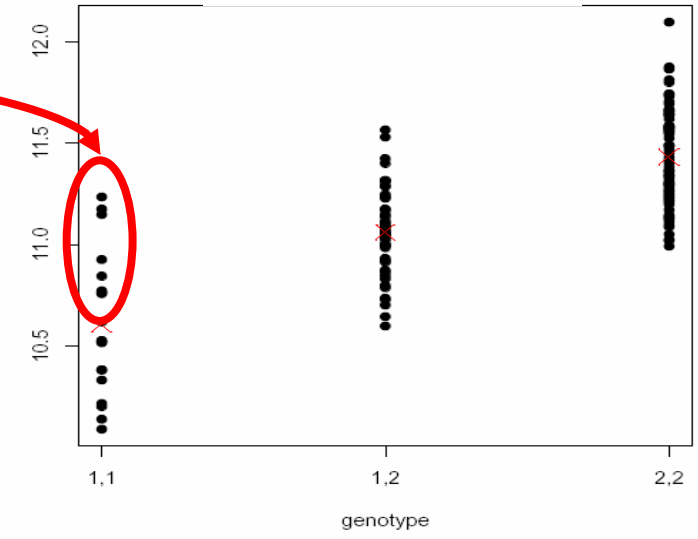
QTDT: $p < 2 \times 10^{-5}$



Marker 2



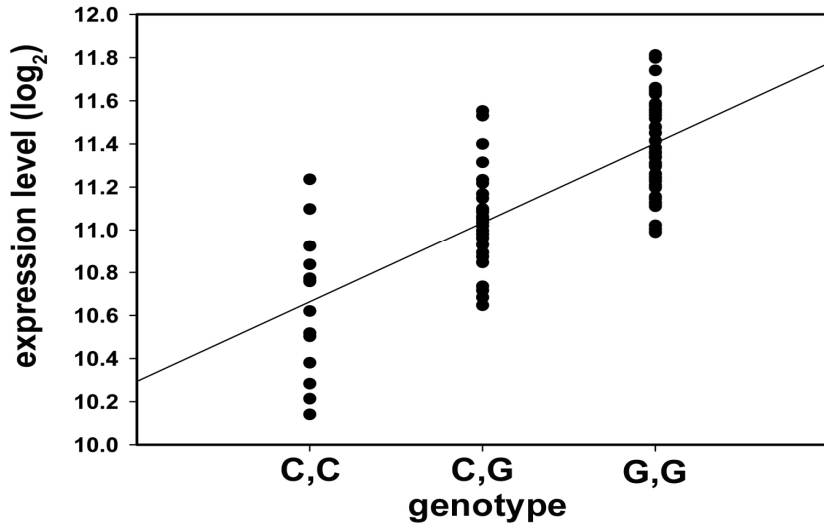
Marker 3



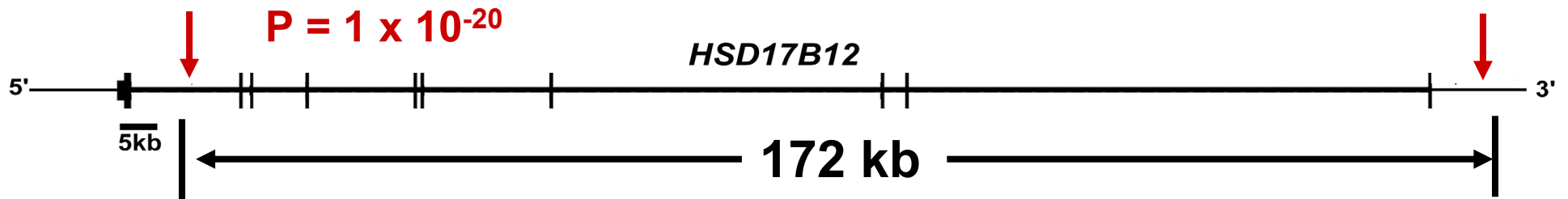
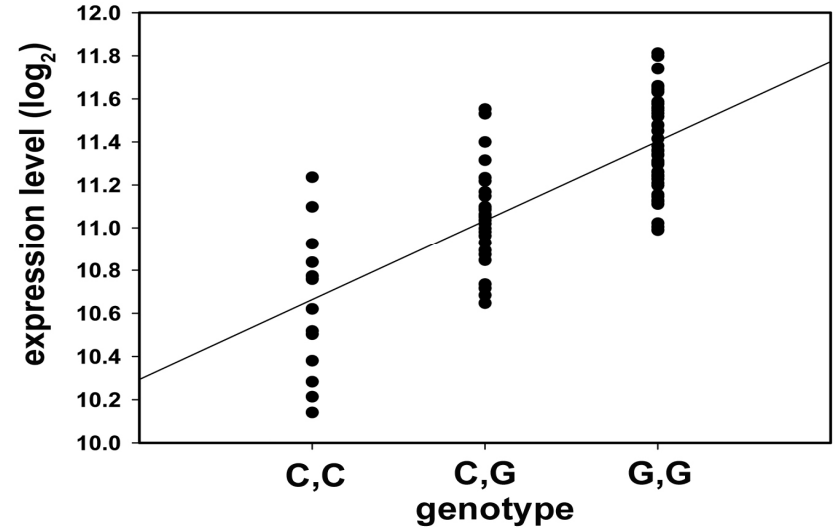
Make LD point clearer

The three markers are in strong LD.

hCV289210



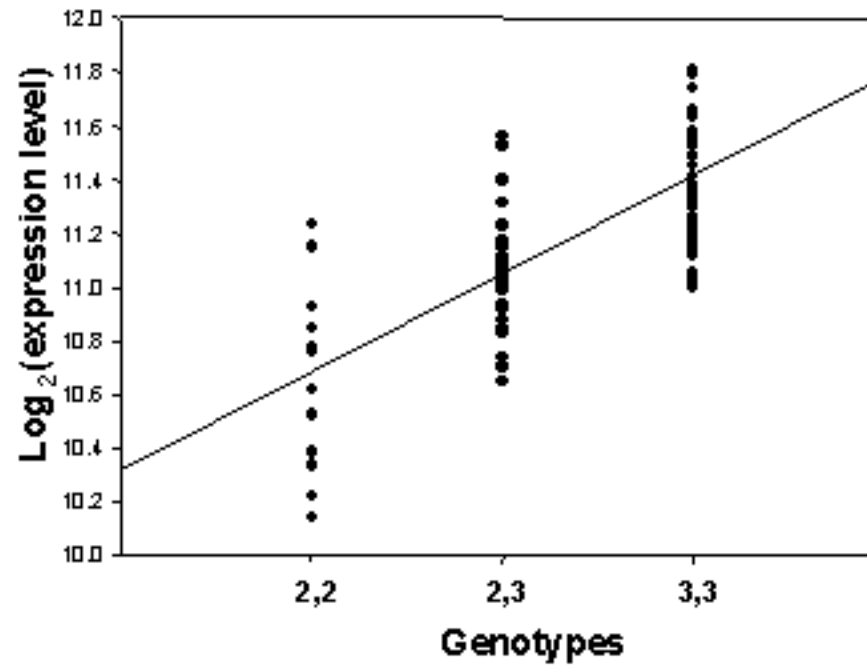
rs1061810



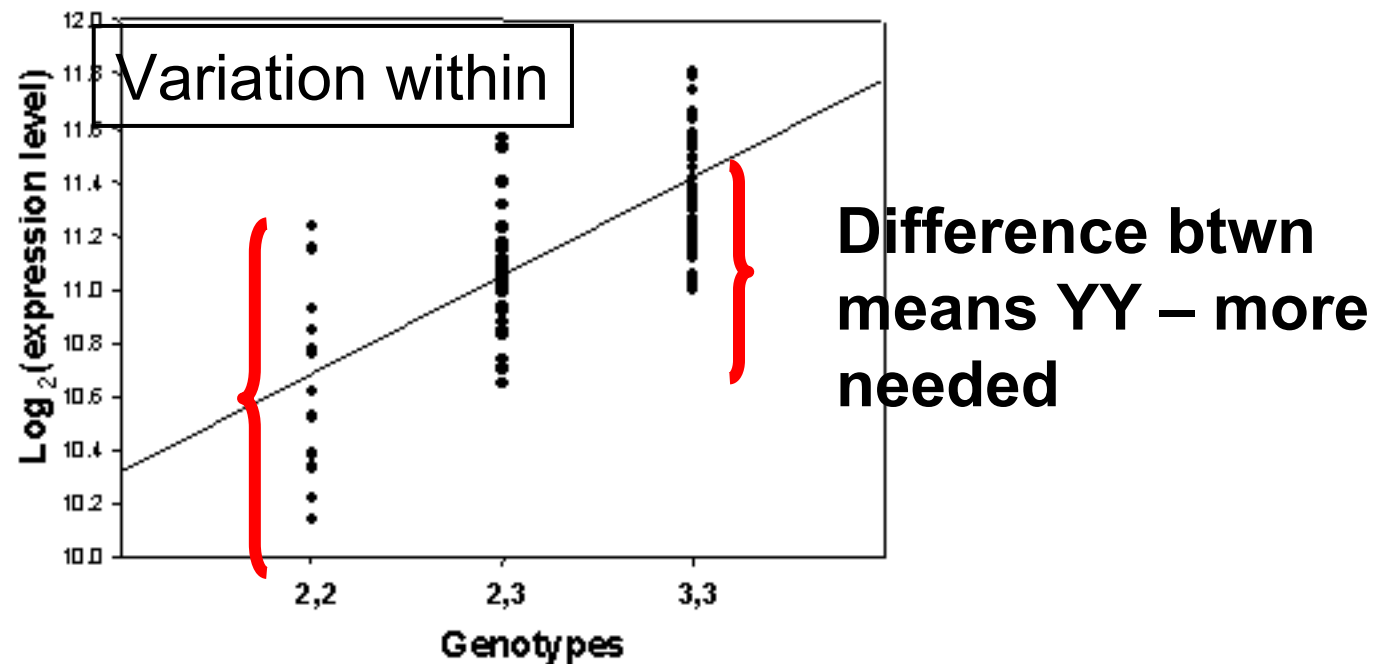
Family-based association with QTDT

$P = 1 \times 10^{-32}$

**Expression of HSD17B12 –
associated with genotype of
SNP marker(s) in the gene.**

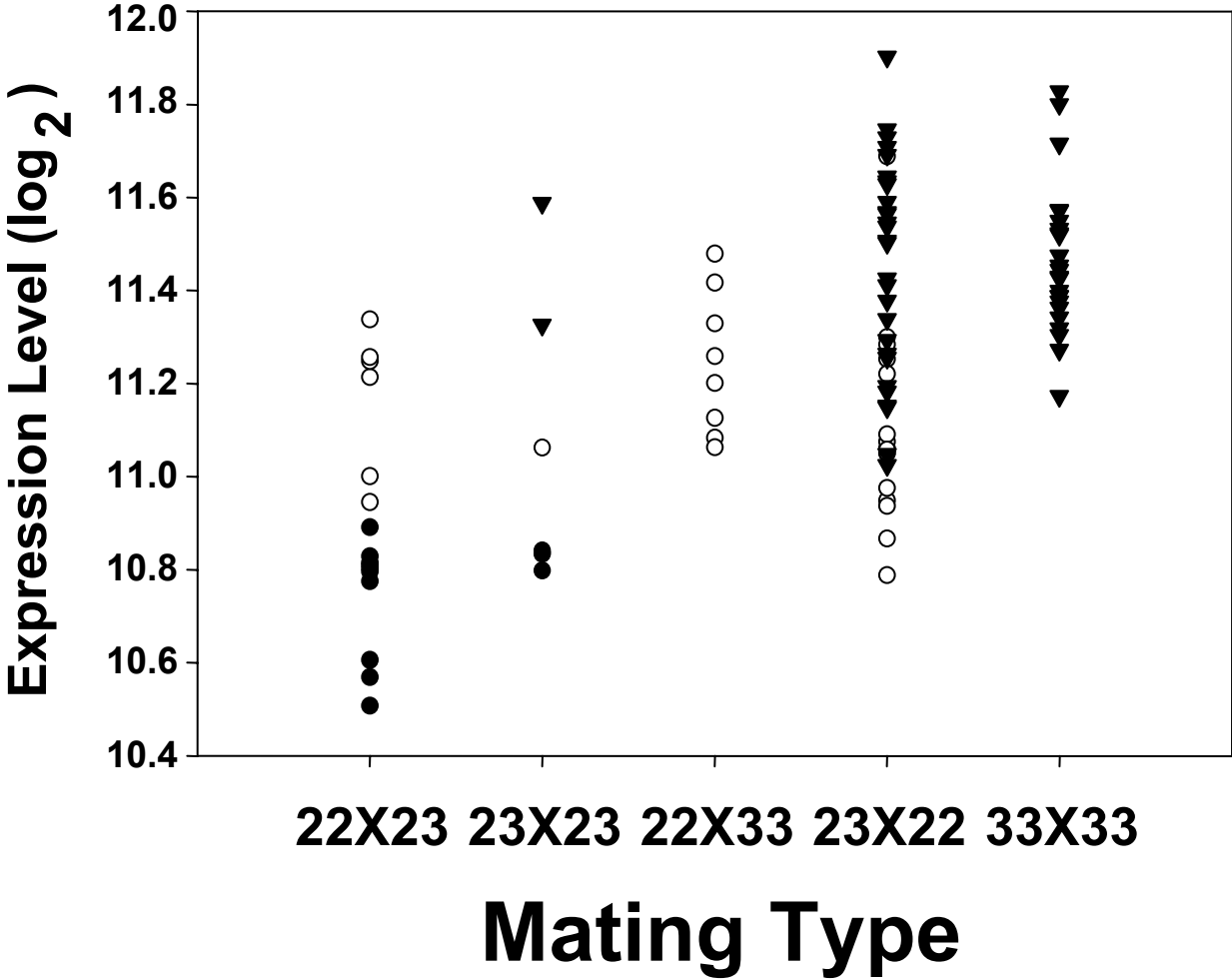


**Expression of HSD17B12 –
associated with genotype of
SNP marker(s) in the gene.**

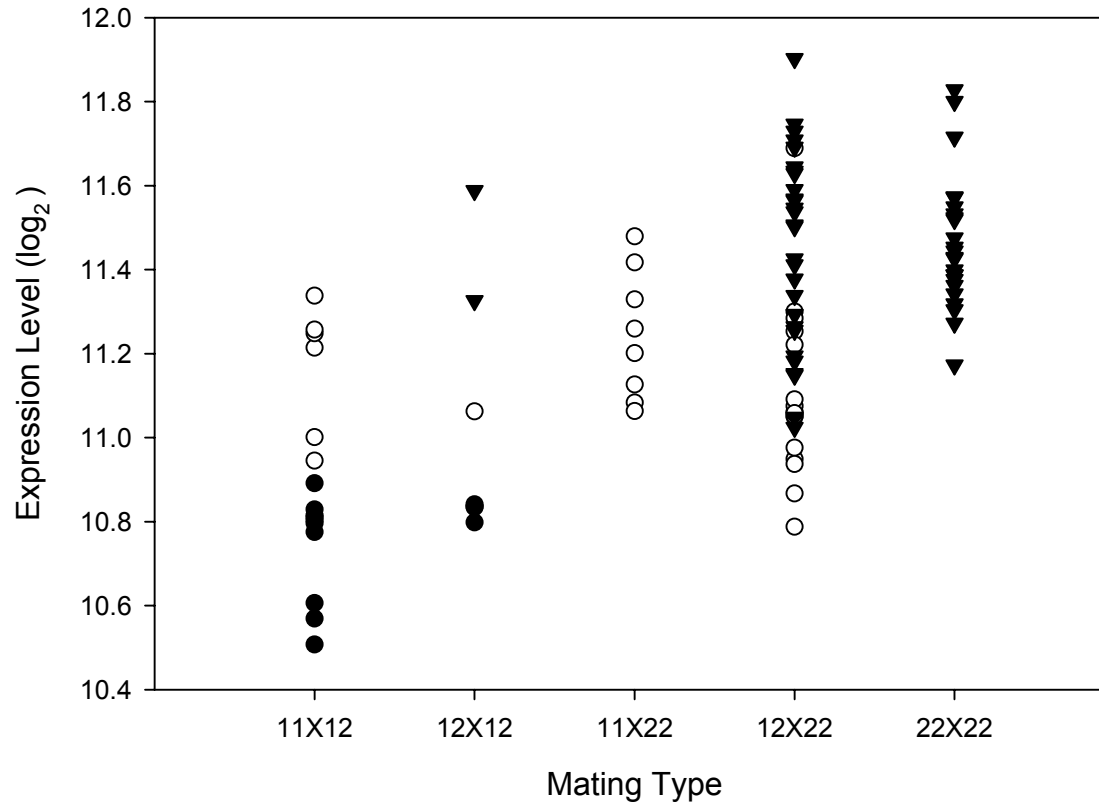


$R^2 = 0.57$ (variation explained)

Offspring in Sibships



Mating Type (Parents + Offspring)



GENETICS OF VARIATION IN GENE EXPRESSION

- **Have mapped individual genetic determinants that contribute to variation in gene expression**
- **Will now identify connections (“networks”) for combined action of determinants**

ASSOCIATION ANALYSIS

Goals:

- Validate the linkage results
- Narrow the candidate regulatory regions

International HapMap Project:

- genotypes for ~ 1 million SNP markers

Whole genome association:

- identify the candidate regulatory regions despite the problem of multiple testing?

WHOLE GENOME ASSOCIATION

57 unrelated CEPH grandparents

- **Phenotype: 27 genes, selected based on highly significant linkage results**
- **Genotype: 770,394 SNP from HapMap (excluded missing data, monomorphic markers)**
- **Expression level (dependent variable) was regressed on SNP genotype (coded 0, 1, 2)**
- **Bonferroni correction for multiple testing**

LINKAGE & ASSOCIATION

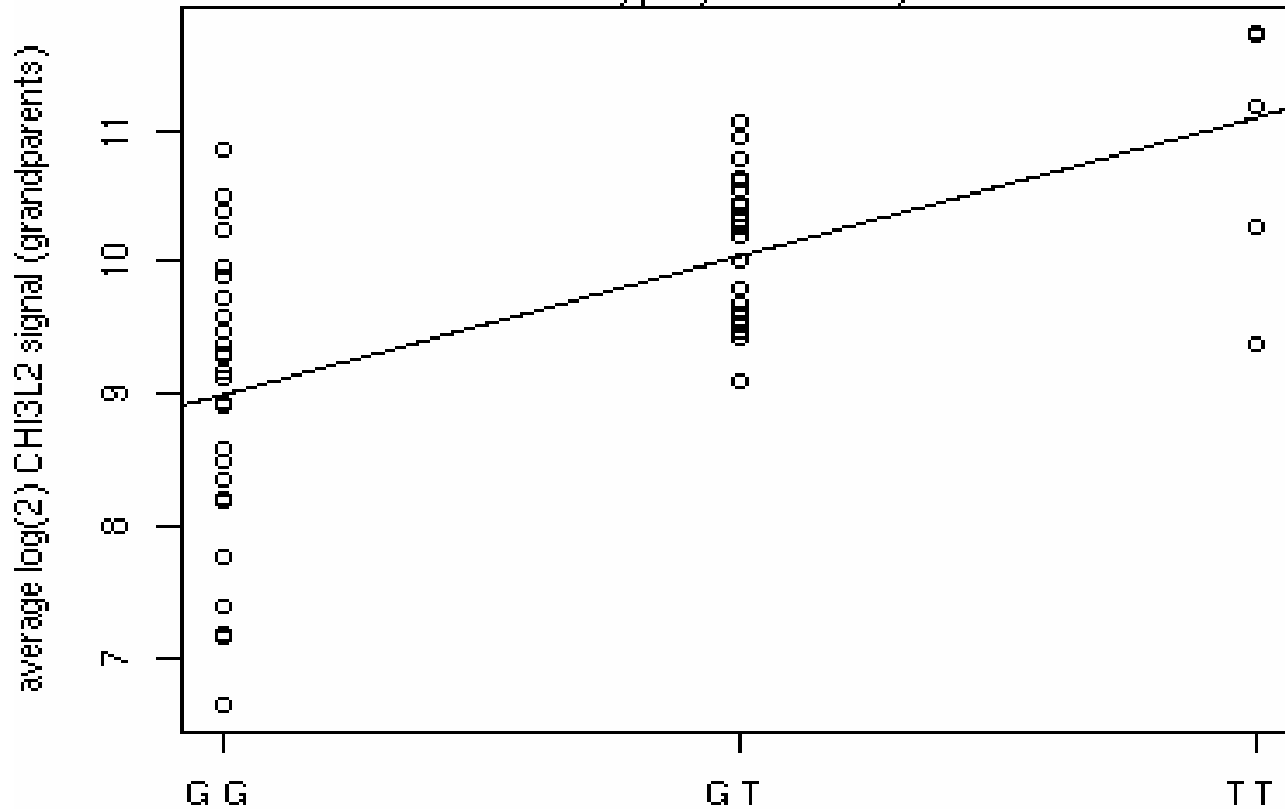
For 15 phenotypes (out of 27): linkage and association pointed to the same regulators

Results from WGA narrowed the candidate regions for all the 15 phenotypes

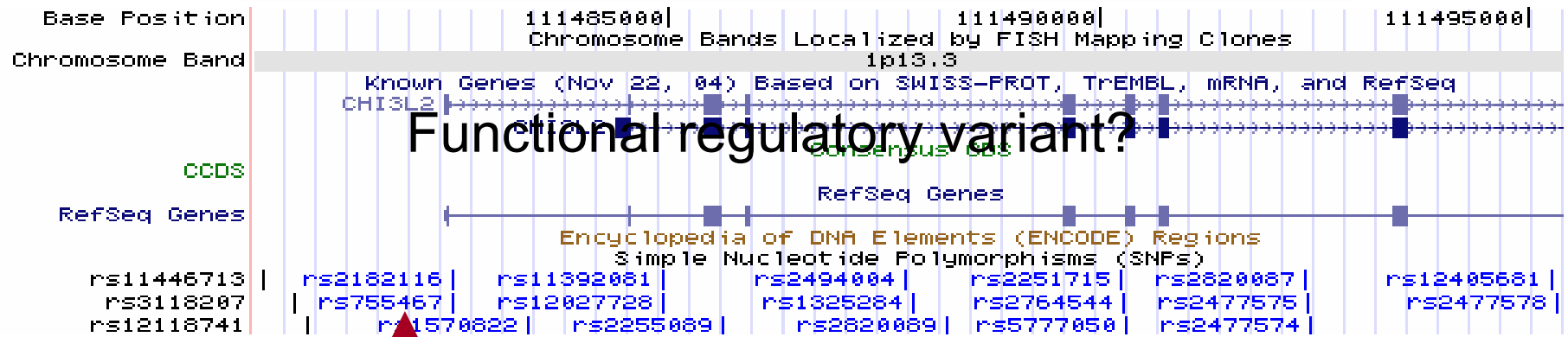
Example with CHI3L2

CHI3L2 association at association peak

F=34.498, p=0, R²=0.385, N=57



rs755467 genotype at 111.482464 Mb from chr 1 pter



LUCIFERASE ASSAY

1. Promoter activity?
2. Allele-specific promoter activity?
3. Preliminary results:

Promoter activity in the region including the SNP, rs755467

Promoter strength T > G allele

LUCIFERASE REPORTER ASSAY

- 1. Promoter activity in the region including the SNP, rs755467**
- 2. Promoter strength T > G allele**
- 3. Why?**
 - HaploChIP assay to examine allele-specific RNA Polymerase II binding**

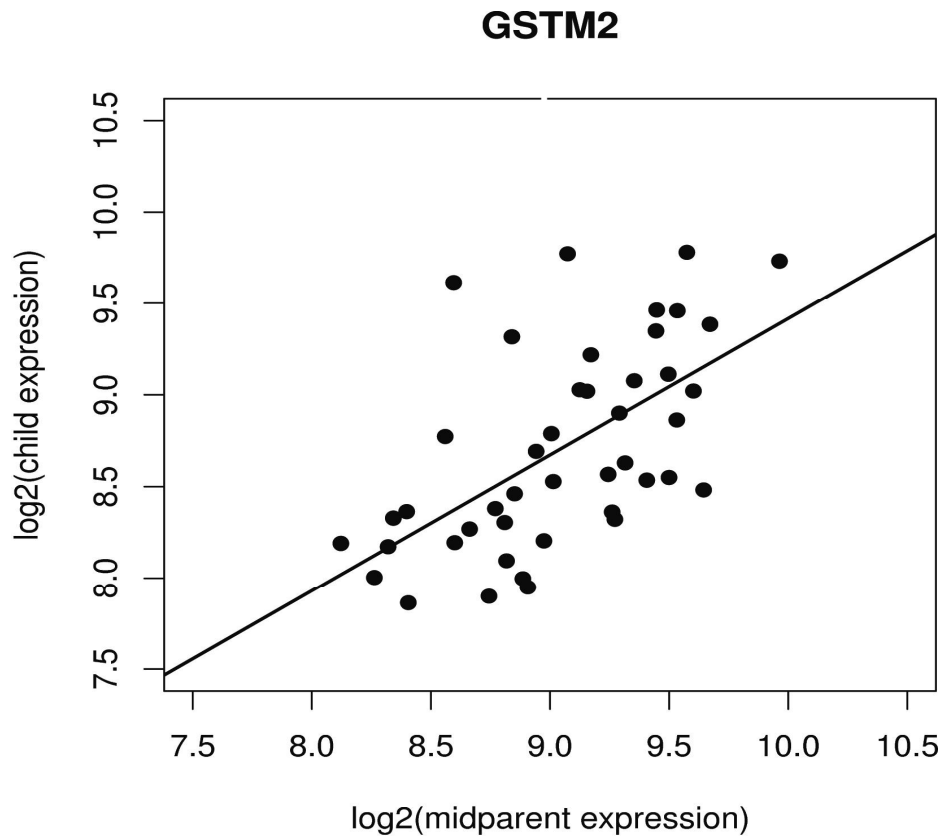
SUMMARY

- **Goal: find all genes (locations) where genetic variation contributes to complex phenotypes, qualitative or quantitative** **MODEL**
- **Genomic approach to genetics: Genetic determinants of baseline level of gene expression (MODEL and REALITY)**

VARIATION IN EXPRESSION LEVELS OF GENES

- **Assess the extent of variation in baseline gene expression**
- Test for familial aggregation
- Identify genetic determinants that contribute to variation in gene expression

Resemblance between offspring and parents



GENE	Narrow-sense heritability
<i>HSD17B12</i>	1.0
<i>LOC64167</i>	1.0
<i>ATP2A3</i>	0.9
<i>ICAP-1A</i>	0.8
<i>UQCRC2</i>	0.8
<i>CLSTN2</i>	0.8
<i>KLF5</i>	0.8
<i>CTSH</i>	0.8
<i>LSM7</i>	0.8

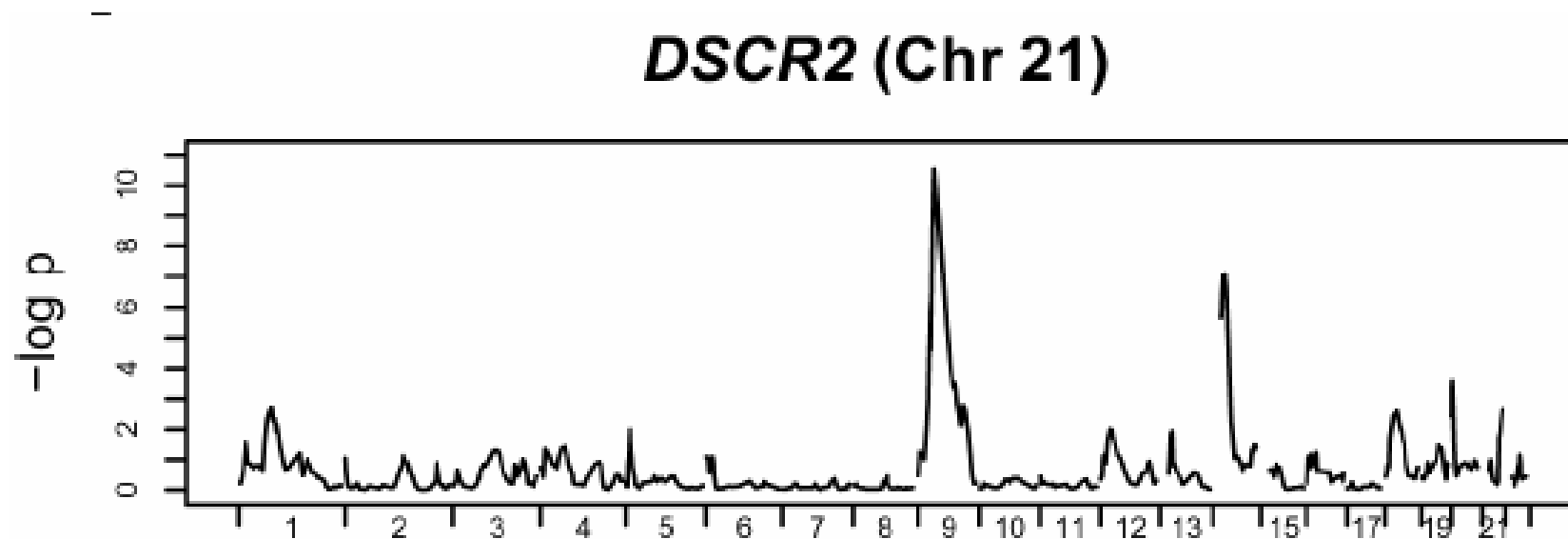
VARIATION IN EXPRESSION LEVELS OF GENES

- Assess the extent of variation in baseline gene expression
- Test for familial aggregation
- **Identify genetic determinants that contribute to variation in gene expression**

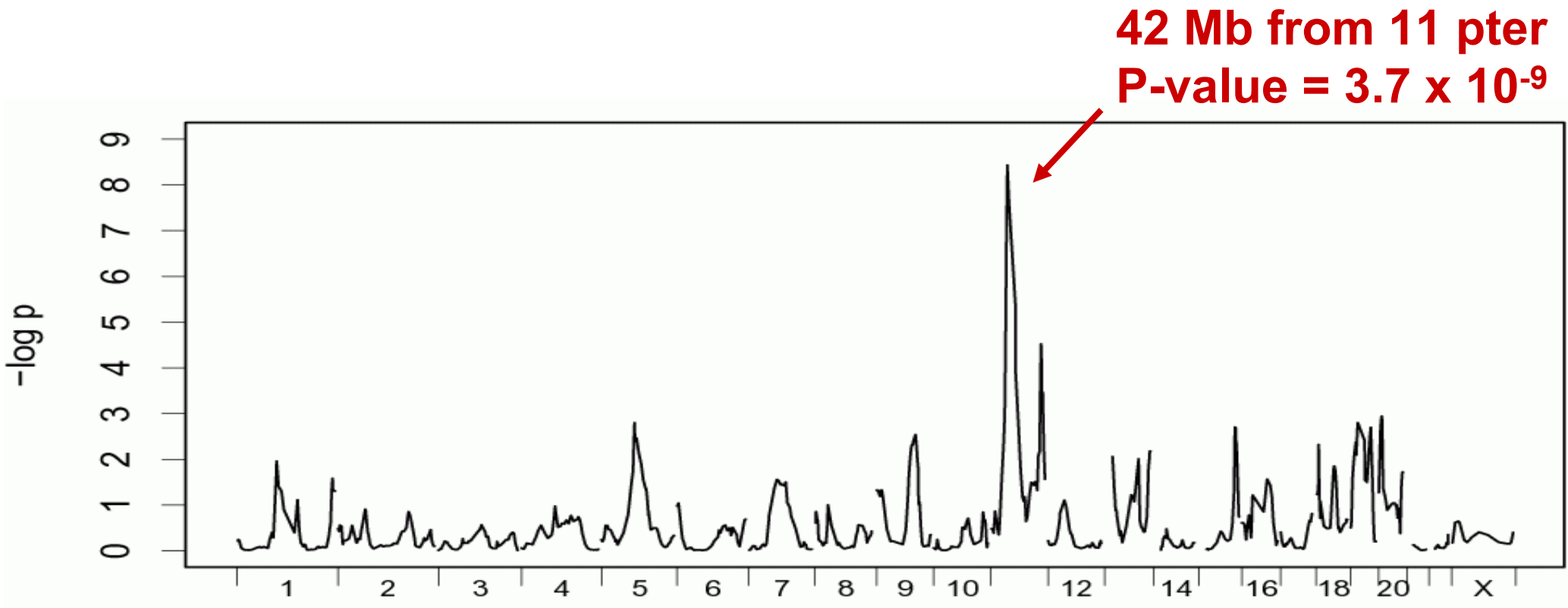
GENOME-WIDE LINKAGE ANALYSES

- **Consider expression level of each gene as a quantitative trait.**
- **To determine the chromosomal locations linked to the expression levels.**
- **3,500 genes - 3,500 genome scans**
- **14 CEPH families**
- **~3,000 SNP markers genome-wide (TSC)**
- **S.A.G.E. programs: SIBPAL (like ASP)**

Genome Scan for One Phenotype



CIS-ACTING REGULATORS



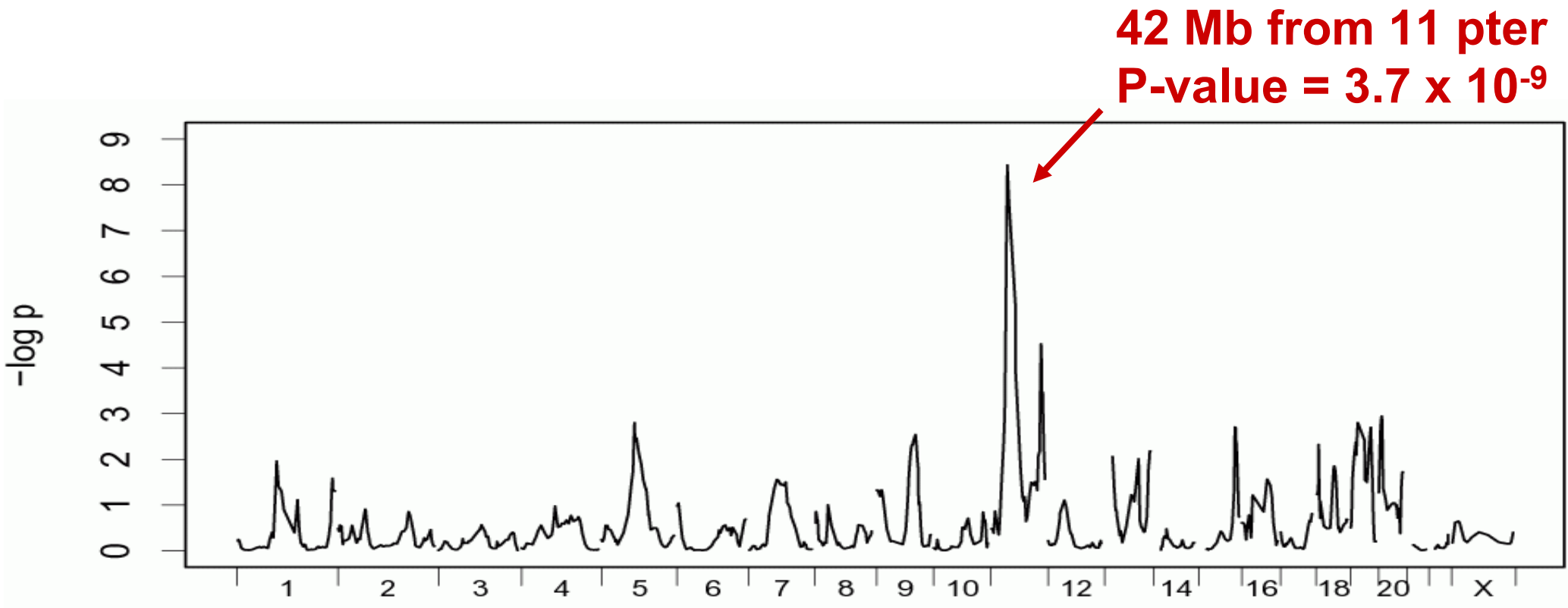
HSD17B12 maps to 44 Mb from 11 pter

Number of genes and level of significance for linkage

p	Genes with at least one marker beyond p-value
< 10⁻¹¹	2
< 10⁻¹⁰	4
< 10⁻⁹	6
<4.3 E-7	130

27 (19%): cis-acting elements, linkage \pm 5 Mb gene
110 (78%): trans-acting
5 (3.5%): multiple regulators

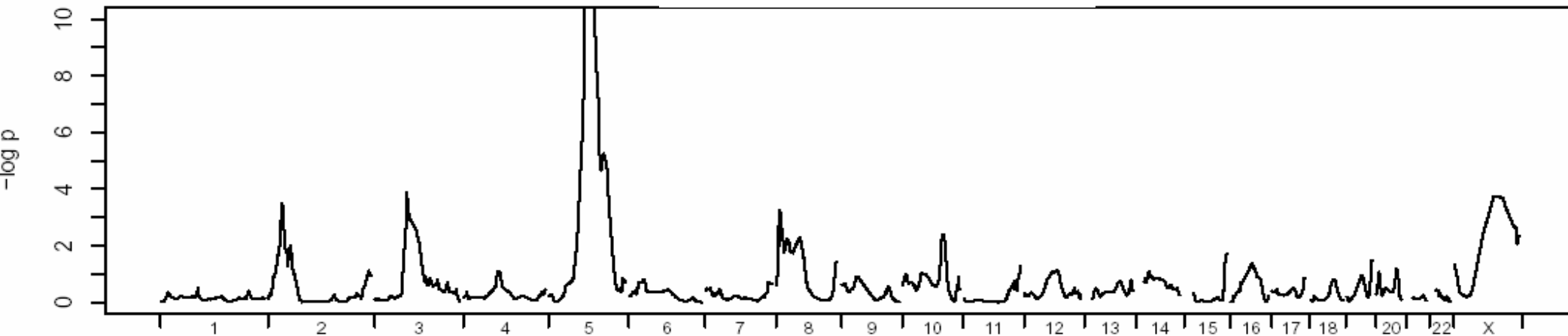
CIS-ACTING REGULATOR



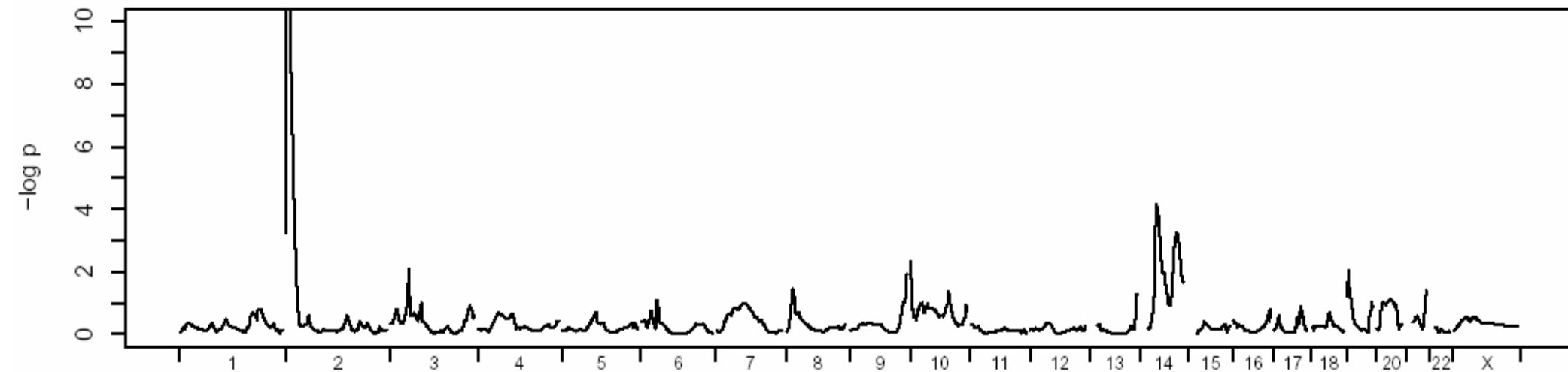
HSD17B12 maps to 44 Mb from 11 pter

OTHER GENES: Cis-acting regulators

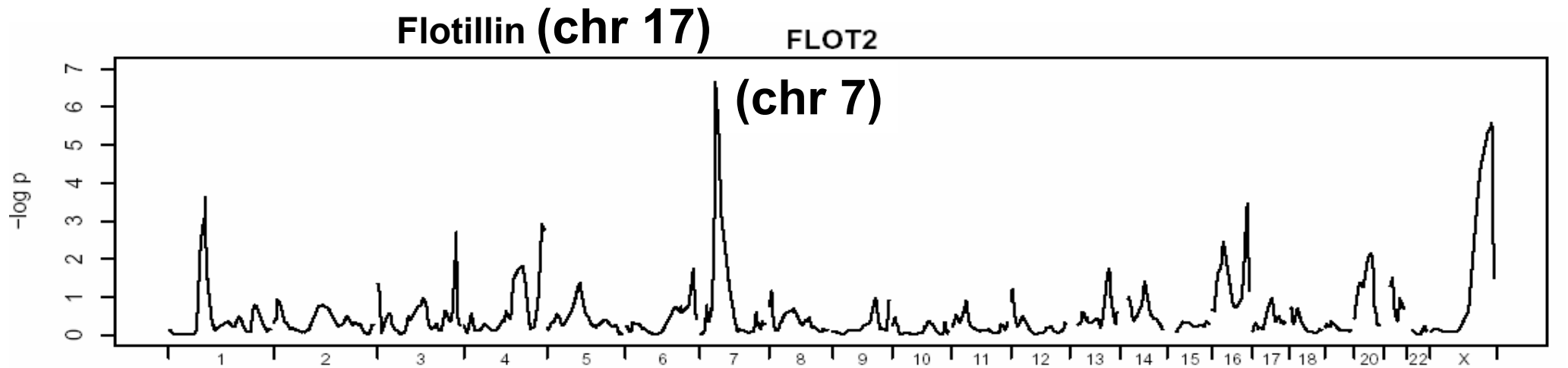
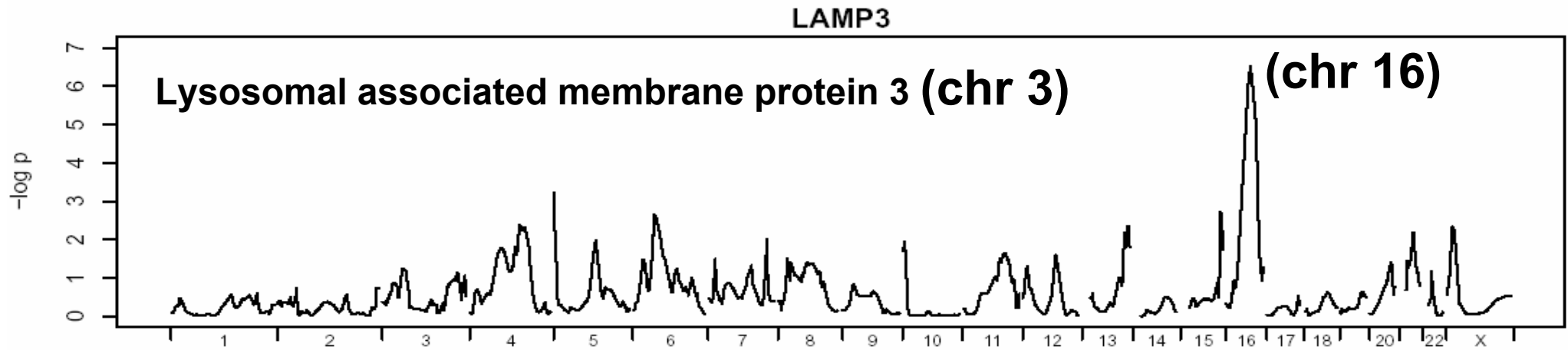
Aminopeptidase – chr 5



Integrin cytoplasmic domain assoc protein 1A – chr 2



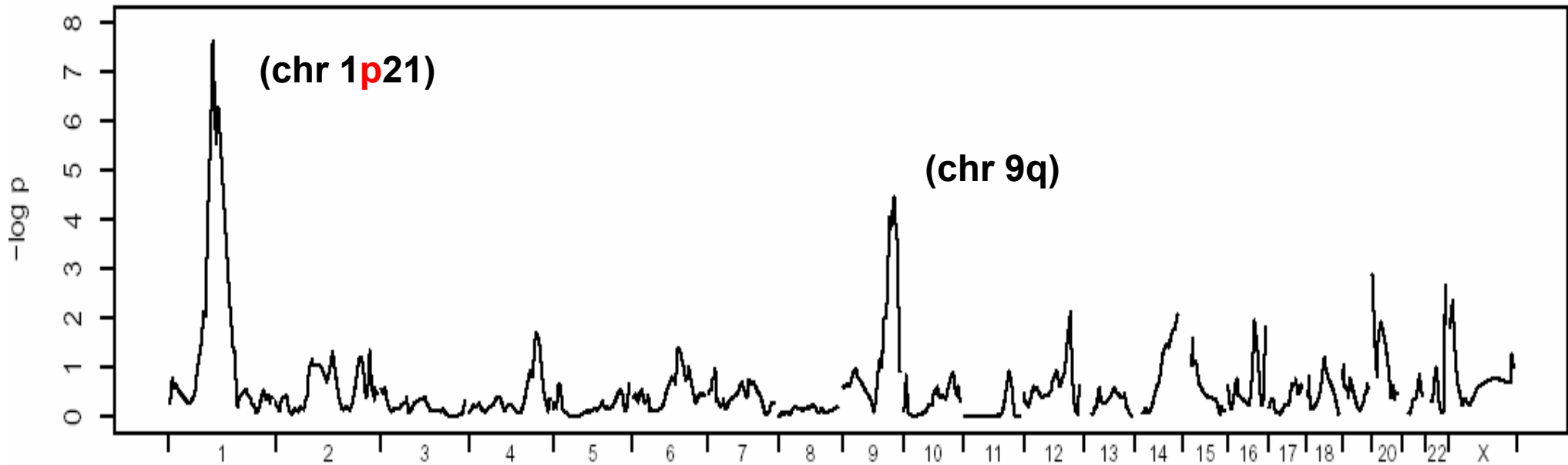
TRANS-ACTING REGULATORS



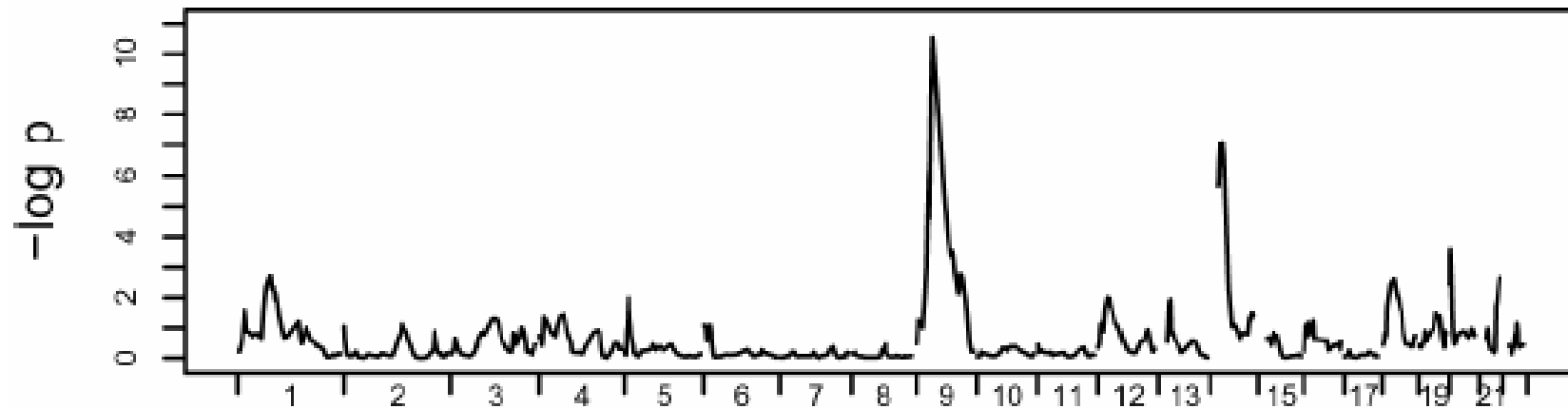
MULTIPLE REGULATORS (for one gene)

Lysophosphatidic acid phosphatase (chr 1q21)

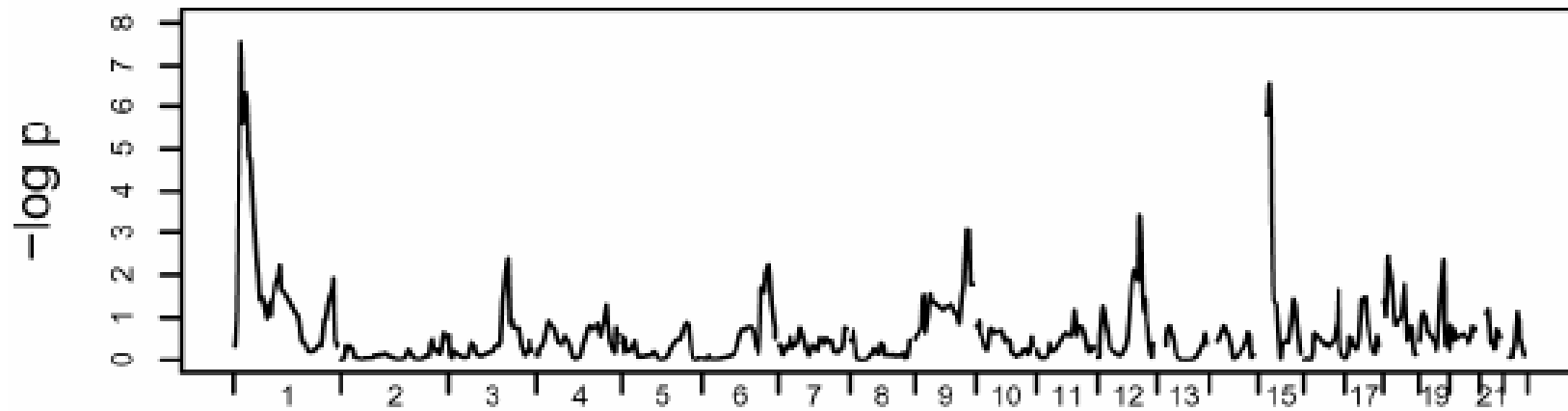
ACP6



DSCR2 (Chr 21)



TGIF (Chr 18)

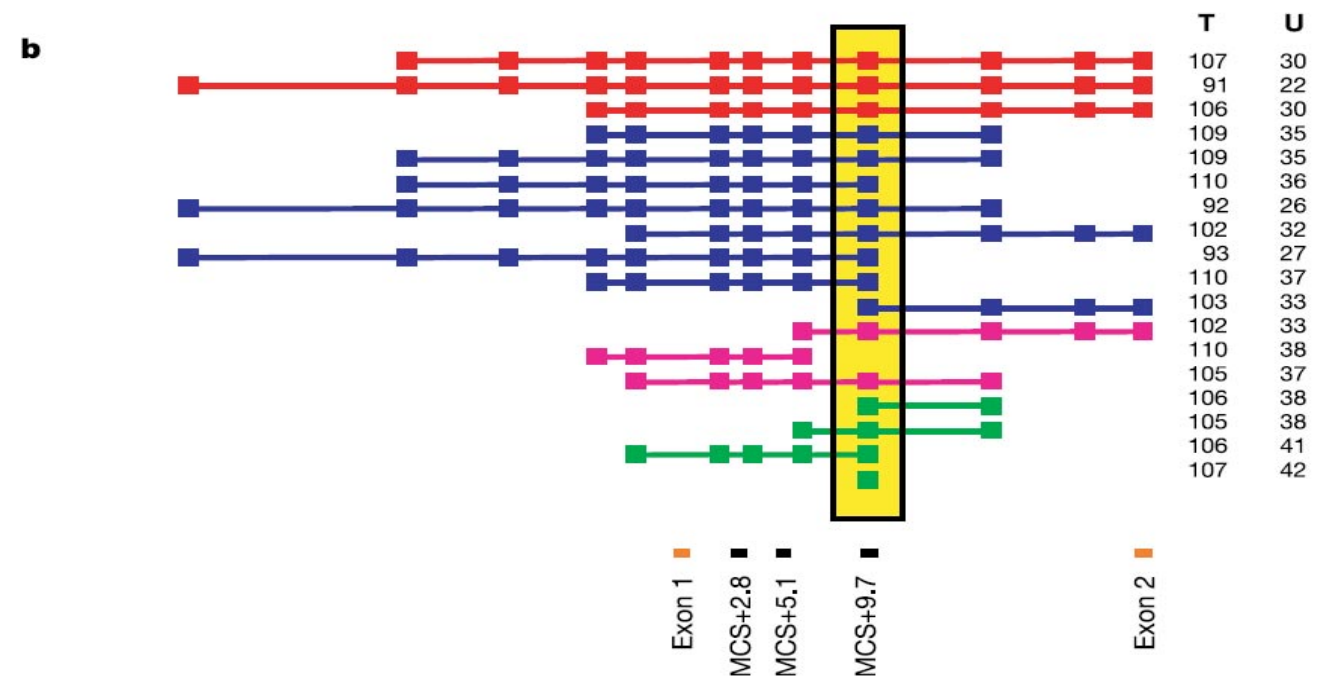
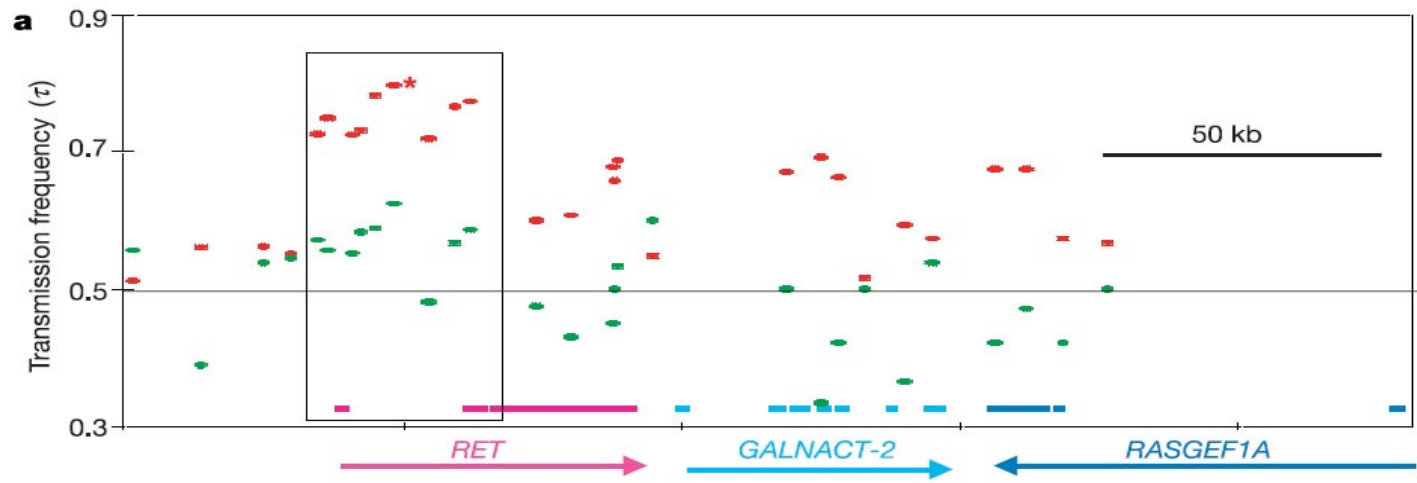


GENETICS OF VARIATION IN GENE EXPRESSION

- **Have mapped genetic determinants that contribute to variation in gene expression**
- **Will now identify connections (“networks”) for combined action of determinants**

Summary

- **Variation – the subject matter**
 - **Linkage and association**
 - **Research stories**
 - **Type 1 diabetes mellitus – fine mapping**
 - **Genetics/genomics of gene expression level**
 - **“Genetic Models for Complex Traits”**
 - **Mathematical / statistical**
 - **Biological / experimental**
-



“EXPRESSION CONTROL ELEMENTS”

- **Molecular biology / Bioinformatics**
 - **Identify network of co-regulated genes, find the regulators**
 - **Genetic approach: map the sites that affect expression levels of variable genes**
 - **Quantitative traits**
 - **Linkage, association methods**
-

GENETICS OF VARIATION IN GENE EXPRESSION

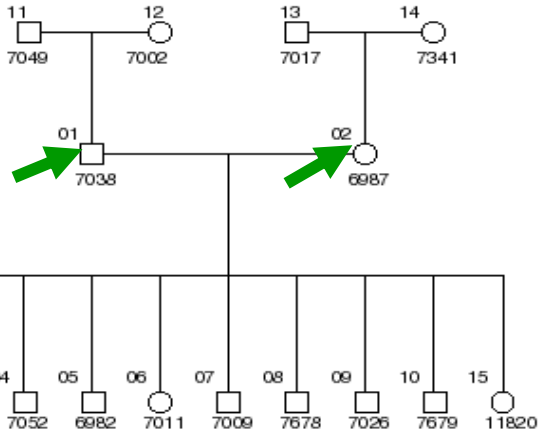
- **Have mapped genetic determinants that contribute to variation in gene expression**
- **Will now identify connections (“networks”) for combined action of determinants**

STUDY MATERIALS

- **Family materials**
 - 3-generation families
 - Marker maps for linkage
 - Lymphoblastoid cell lines (LCLs)
- **Phenotypes: Affymetrix Genome Focus arrays**
 - 8,500 genes; ~3,500 expressed in LCLs
 - Hybridizations done in duplicate

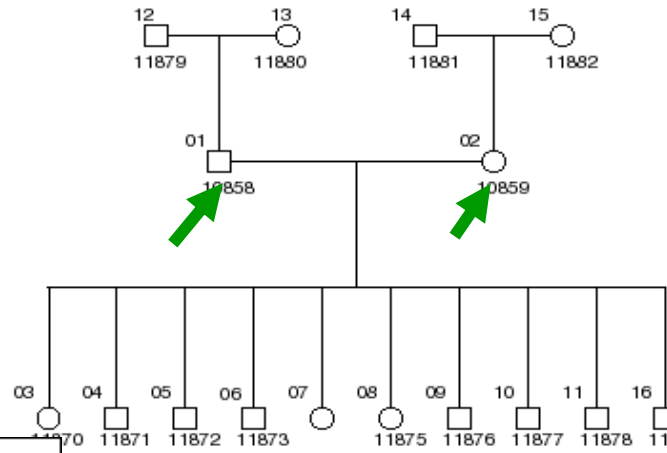
CEPH/Utah Pedigree 1333

SIBS = 9



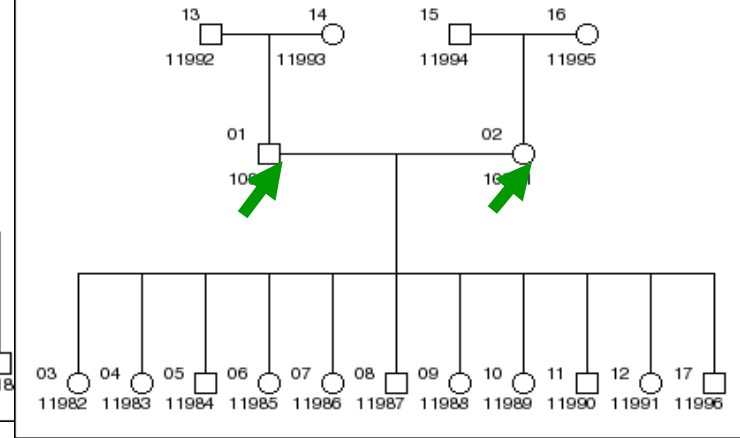
CEPH/Utah Pedigree 1347

SIBS = 9



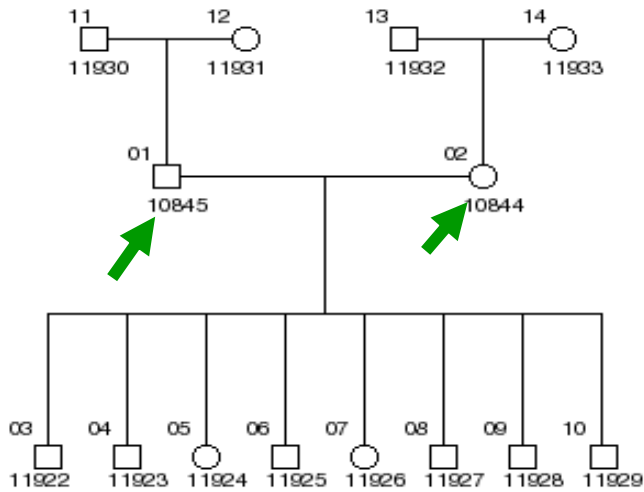
CEPH/Utah Pedigree 1362

SIBS = 11



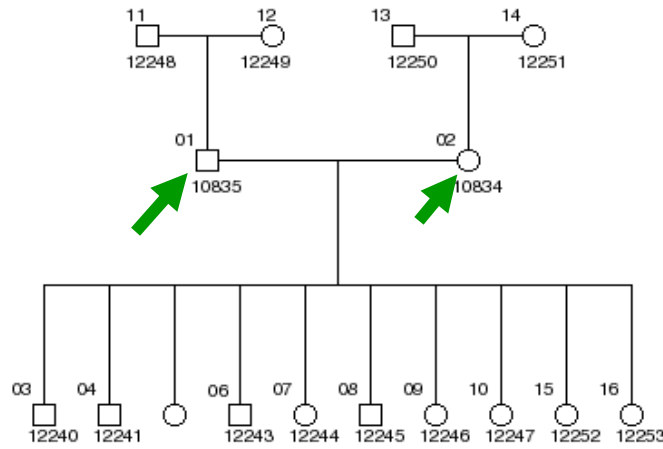
CEPH/Utah Pedigree 1424

SIBS = 8



CEPH/Utah Pedigree 1416

SIBS = 9



XXX Population Geneticist's View of Gene Express

Variation

**Polymorphisms
in DNA**

- STRPs
- SNPs, etc.

Variation

**Intermediate
Phenotypes:**

- aa Sequence
- Structure
- **Expression**

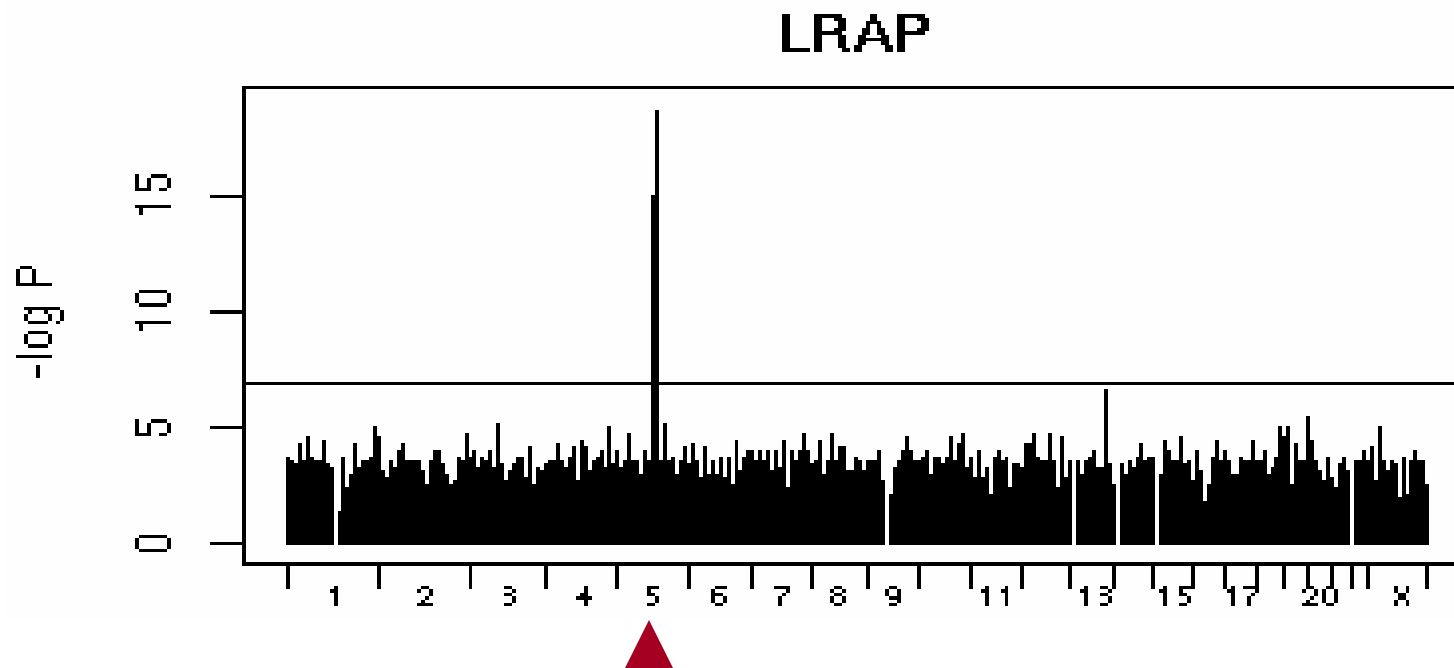
Variation

**Normal
Variation, and
Disease
Susceptibility**

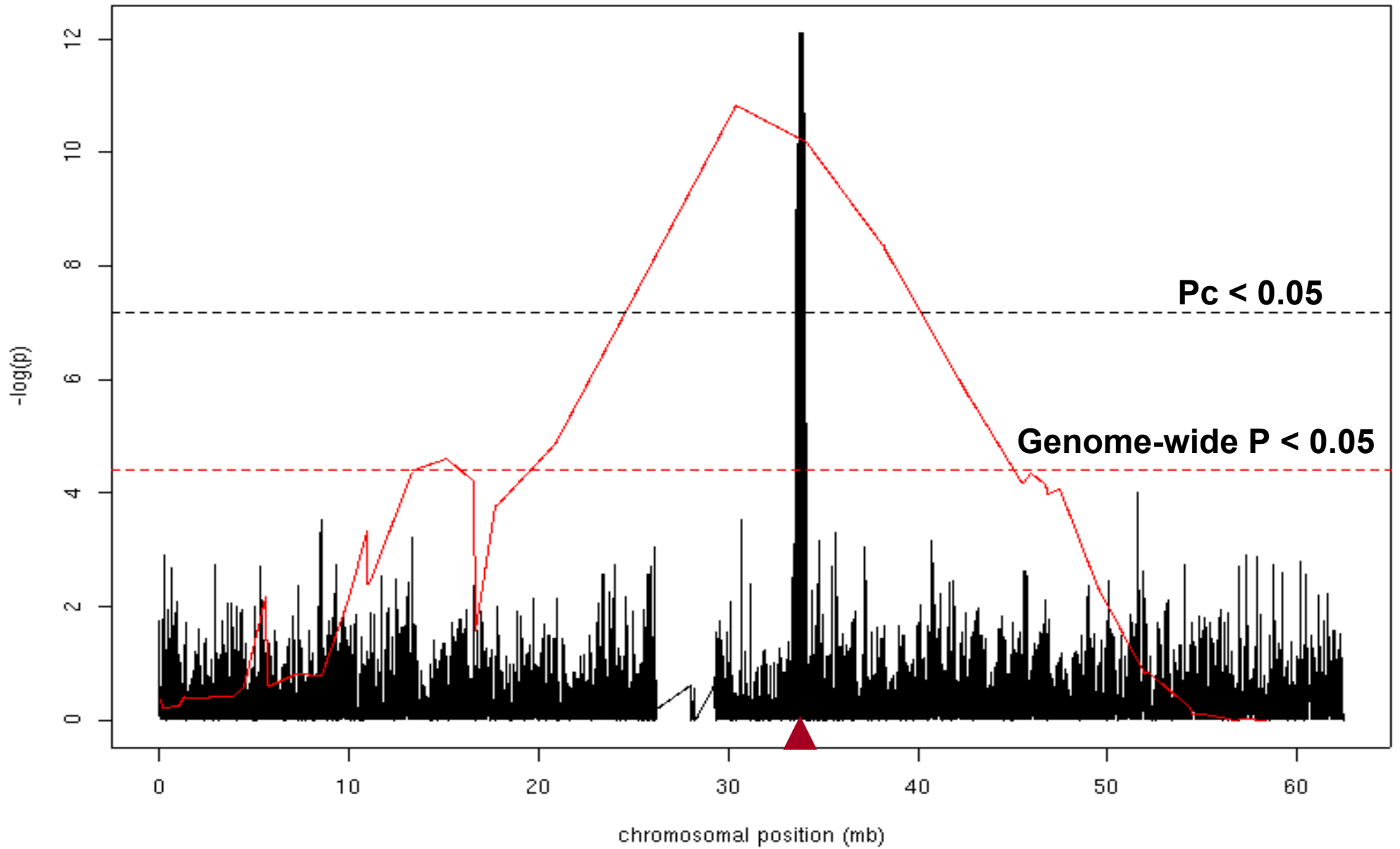


RESULTS

- 14 of the 27 phenotypes: at least one significant marker ($P < 0.05$) after correction



CPNE1 Chromosome 20



Overview

- **Variation – the subject matter**
 - **Simple methods for linkage and association**
 - **Research stories**
 - **Type 1 diabetes mellitus – T1DM**
 - **Genetics/genomics of gene expression level**
 - **“Simpler” quantitative trait**
 - **Prototype for other complex traits**
-