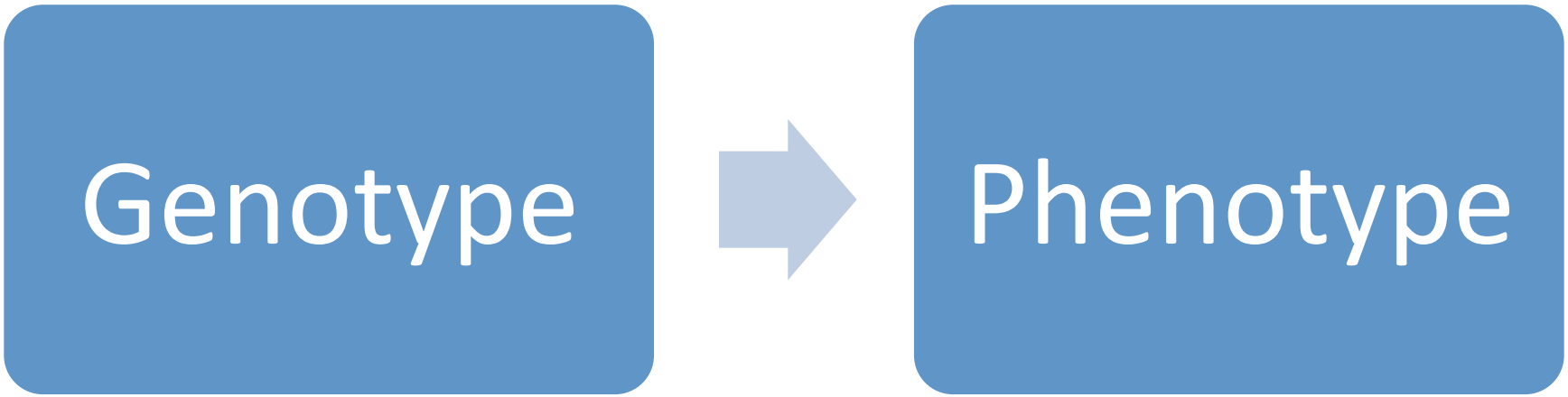
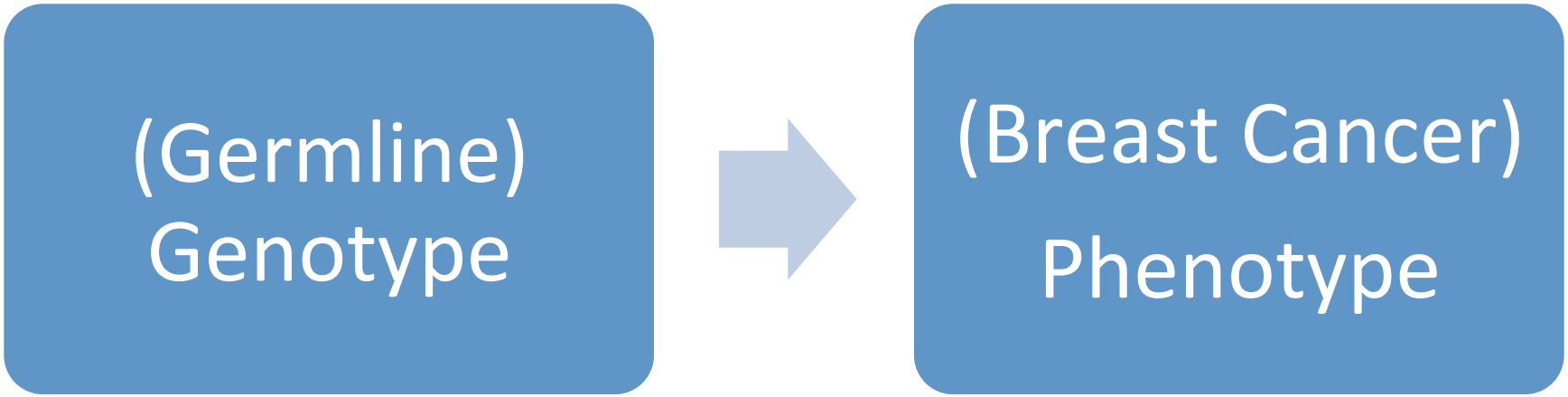
Genome-wide analysis of expression quantitative trait loci in breast cancer

Nicholas Knoblauch
Research Assistant II
Beck Lab
Department of Pathology, Beth Israel Deaconess Medical Center & Harvard Medical School

Introduction

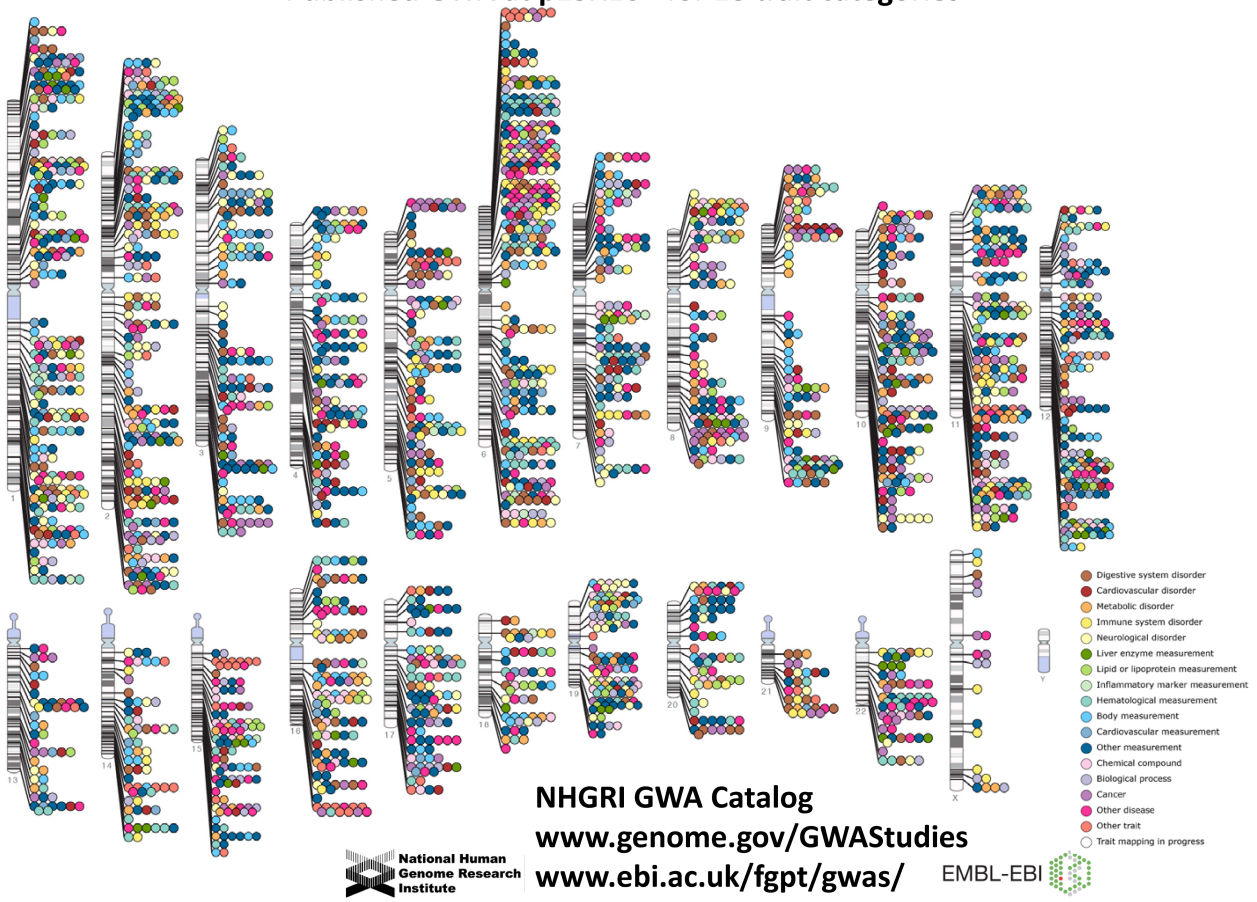


Introduction



Genotype → Phenotype

Published Genome-Wide Associations through 07/2012
Published GWA at p≤5X10^{-8} for 18 trait categories



NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

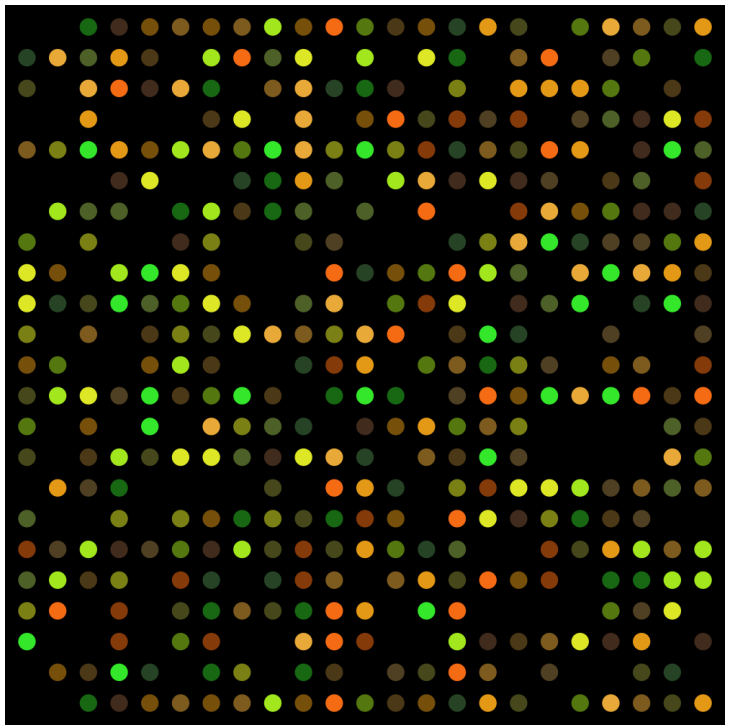
National Human Genome Research Institute

EMBL-EBI

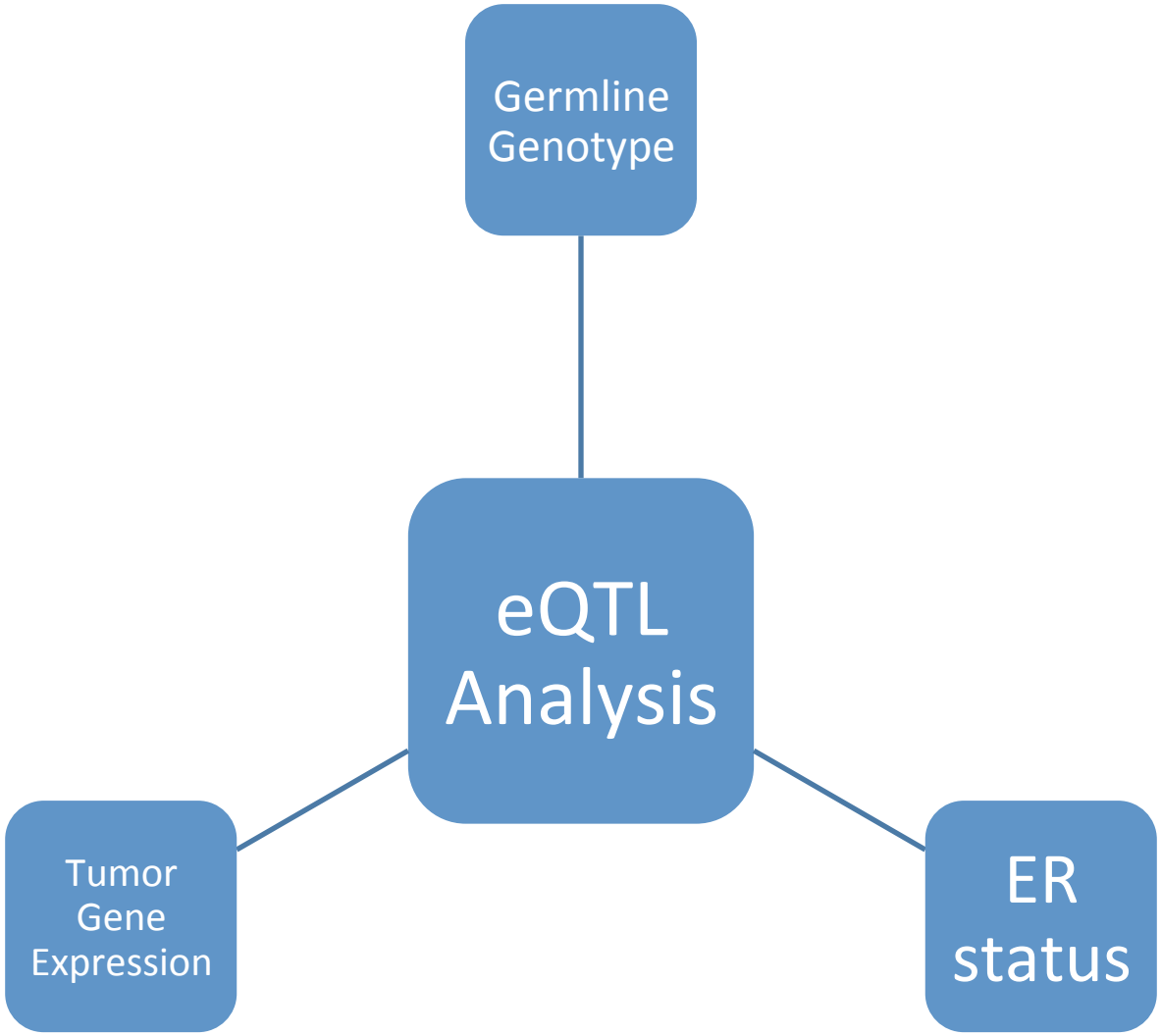
Breast Cancer GWAS

- A number of SNPs are associated with increased risk of Breast cancer
  - ~50 in GWAS catalog
- How do we infer the mechanism of these risk alleles?
  - How do we study the functional consequences of variation at these loci?

Gene Expression as a phenotype

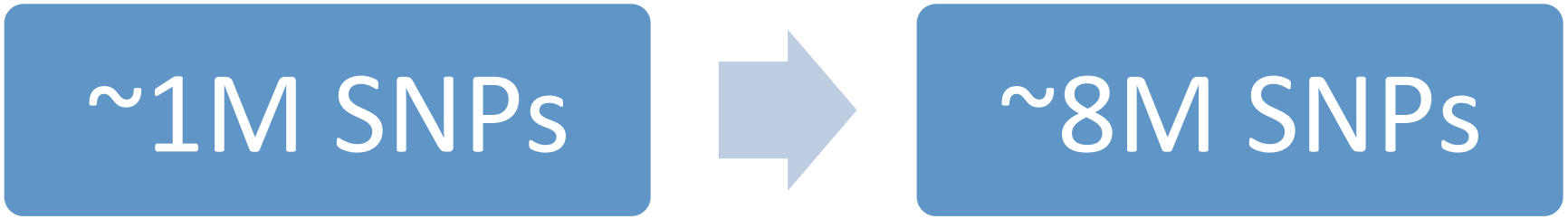


- Easy to measure tens of thousands of features simultaneously
- Facilitates investigation of functional the consequences of genetic variance


Germline Genotype
eQTL Analysis
Tumor Gene Expression
ER status

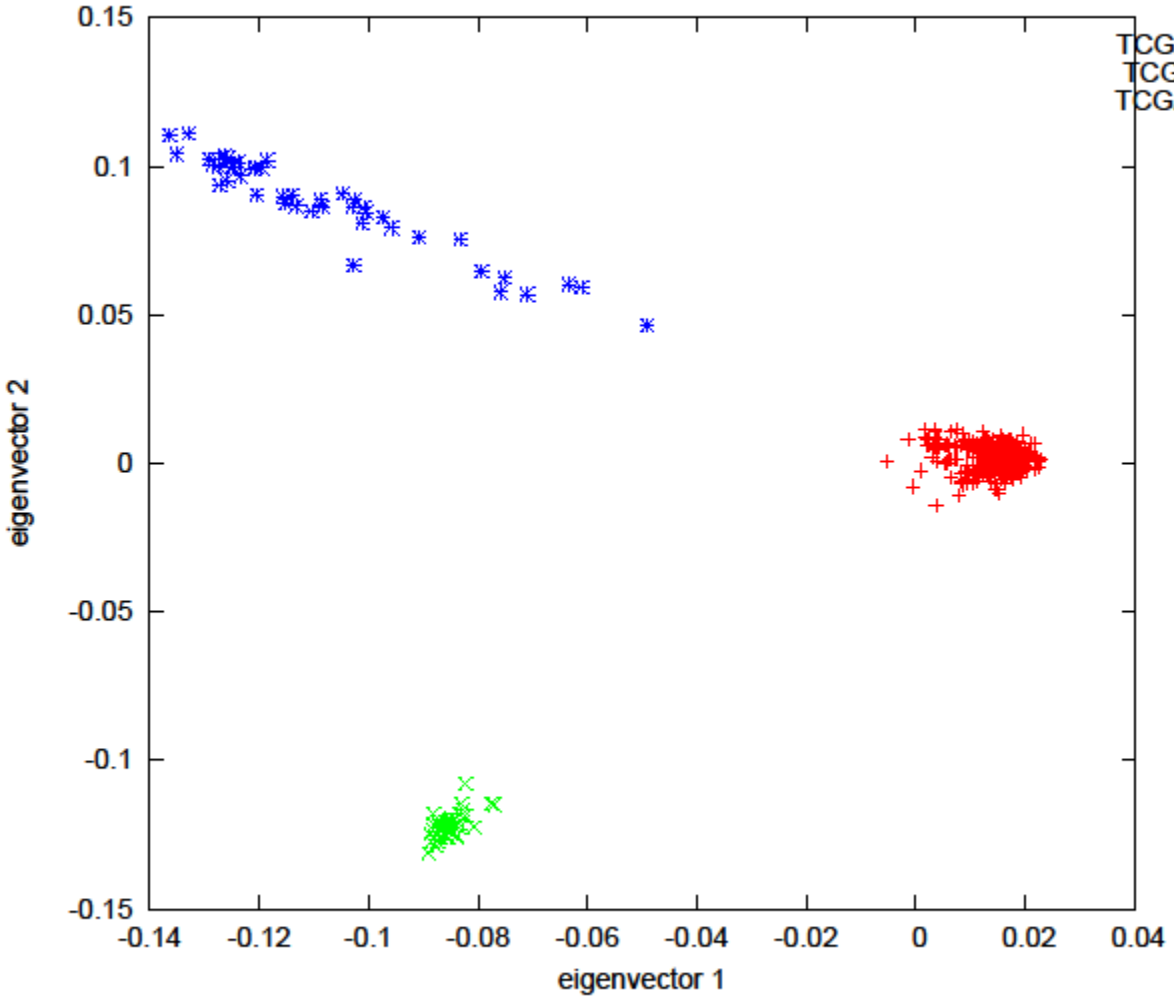
Data

- 382 TCGA invasive breast cancer cases
- Germline SNP data from Affymetrix 6.0 SNP
- Expression from Agilent G4502A 244k Array
- Imputed SNP data
  - ~8 million SNPs



- Imputation
  - Estimate genotype for ungenotyped markers using a genotyped reference panel
  - BEAGLE
    - Infers haplotypes for unrelated individuals
  - minimac
    - Low memory footprint
    - Implements MaCH algorithm
  - 906600 SNPs
  - 1000G (Nov 2010) for imputation

EIGENSTRAT Analysis (PC1 vs. PC2)



Genetically estimated

White	577
Asian	42
Black	46
Total	665

Linear Model

$$g = \alpha + \gamma x + \beta s + \varepsilon$$

- Parameters for intercept,genotype and covariate (ER status)
- MatrixEQTL
  - Tests each SNP-transcript pair
  - “Ultra-fast”
    - ~2 days with 70GB RAM

Results

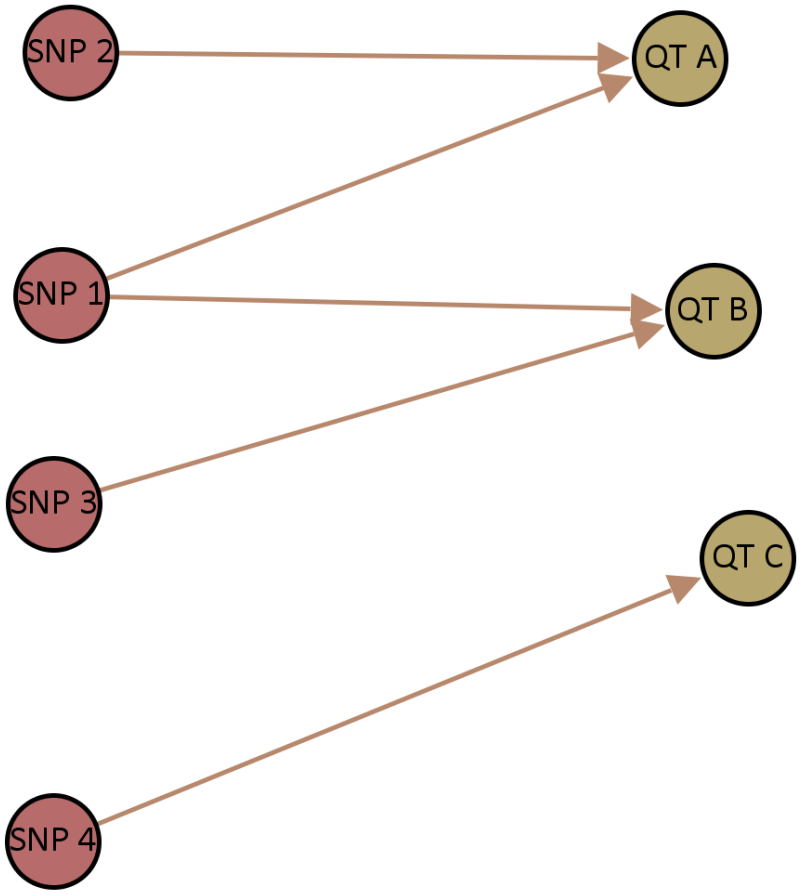
- Of ~8 million SNPs:
  - ~140,000 SNP are eQTL
  - None of 51 known breast cancer risk alleles were detected as eQTL at the given significance threshold

Risk Alleles and eQTL

	Not Risk Allele	Is a Risk Allele
Not eQTL	7880425	53
Is an eQTL	138138	0

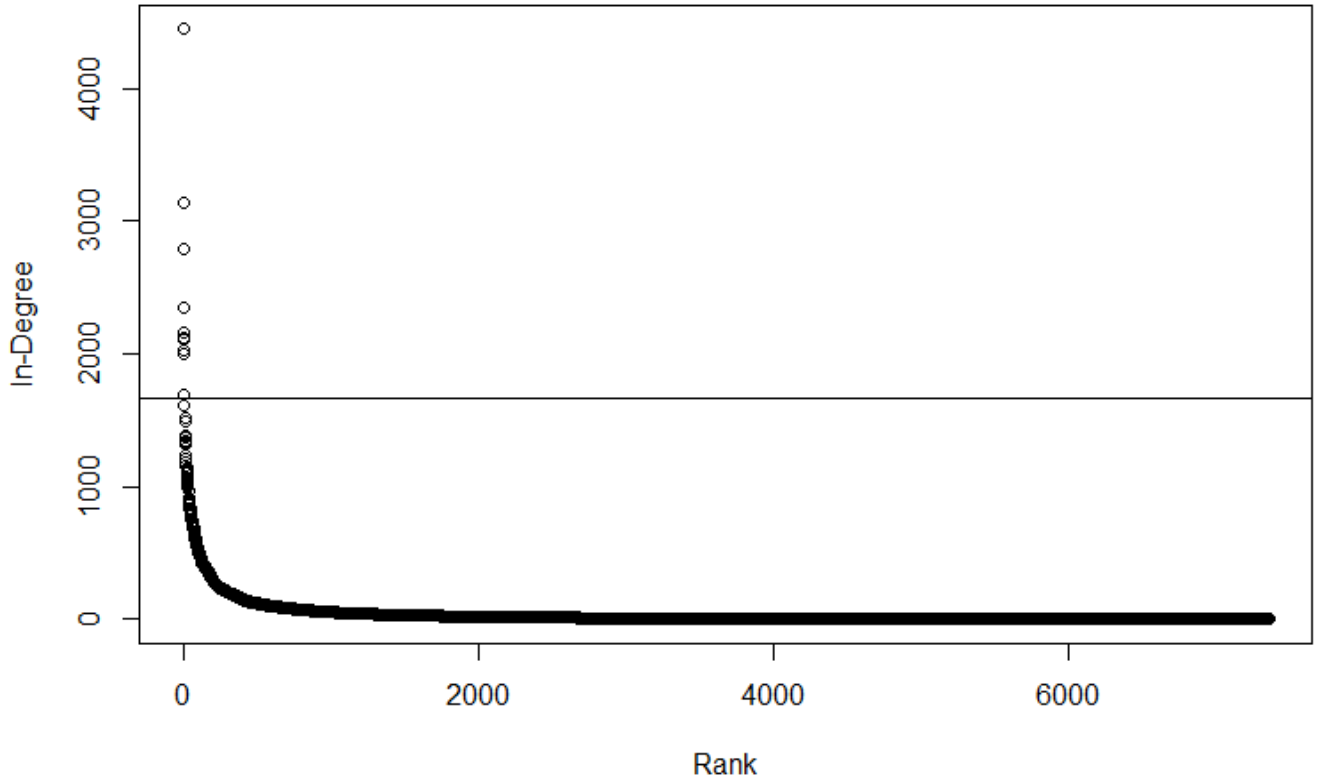
X^2=0.1901, p=0.6628

eQTL Network

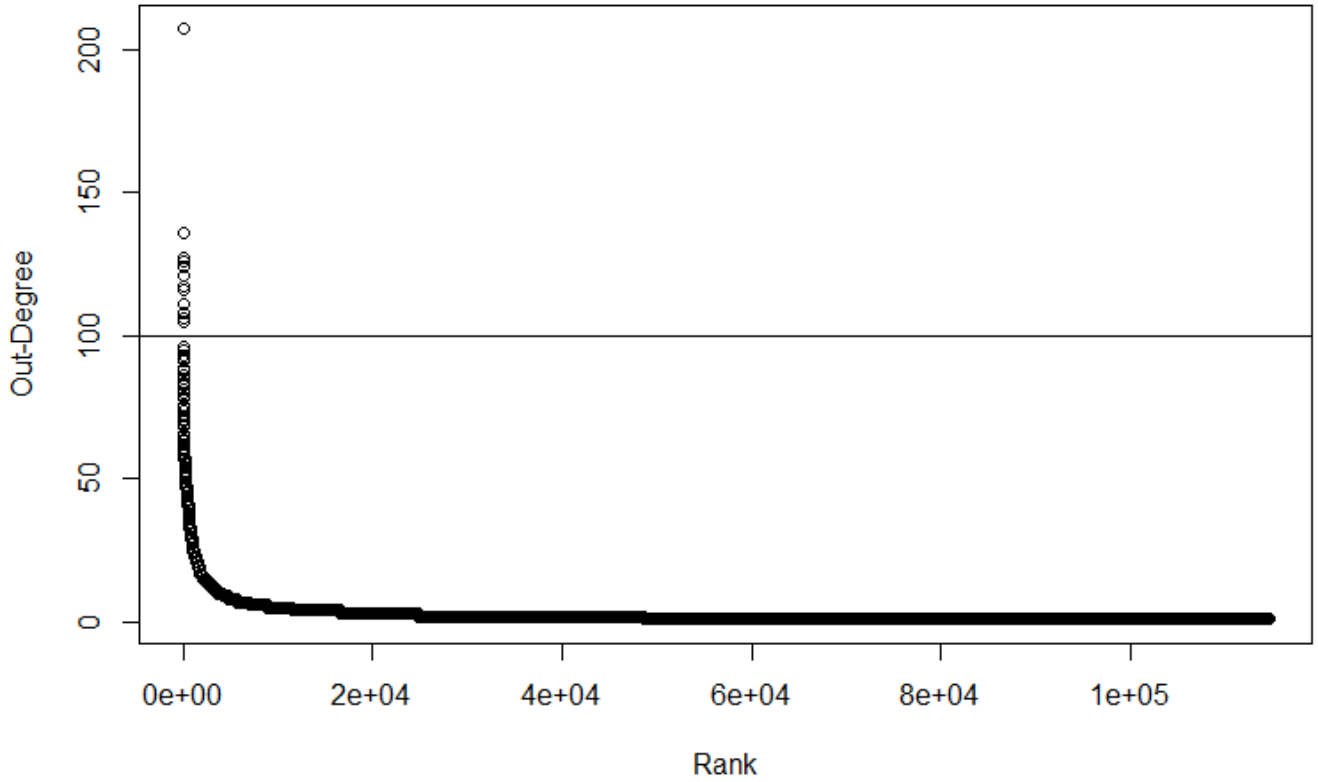


- Loci -> Transcripts
- In-degree
- Out-degree
- Connectivity

In-Degree Distribution

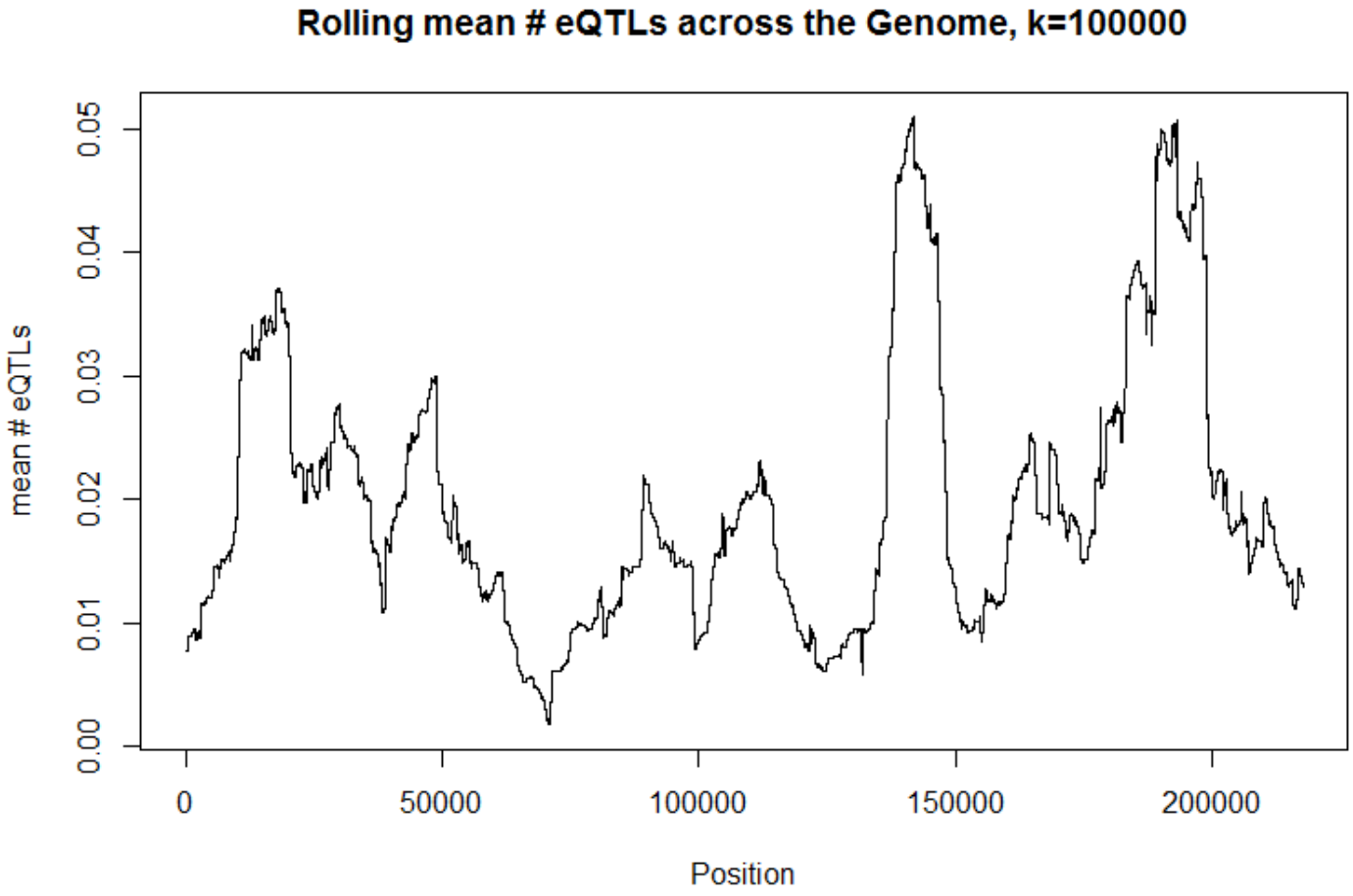


Out-Degree Distribution



QTs with highest in-degree

Gene	In-degree	Description
TSSK1B	4446	testis-specific serine kinase
CYB5A	3142	Cytochrome b5 type A
DPF3	2794	Zinc Finger Protein
PRL	2354	Prolactin
MEN1	2156	multiple endocrine neoplasia
CSH1	2120	chorionic somatomammotropin hormone 1 (placental lactogen)


Rolling mean # eQTLs across the Genome, k=100000
mean # eQTLs
0.00
0.01
0.02
0.03
0.04
0.05
0
50000
100000
150000
200000
Position

“ER-dimorphic” eQTL

- 32 eQTL with opposite t-statistic in ER+ and ER- cases.

Gene	Description
CD5L	inhibitor of apoptosis
C19orf6(membralin)	Tumor-associated protein
MUC4	inhibitor of apoptosis
SPATA19	Spermiogenesis
IGF1R	anti-apoptotic agent by enhancing cell survival
H2AFB3	Atypical histone H2A which can replace conventional H2A in some nucleosomes

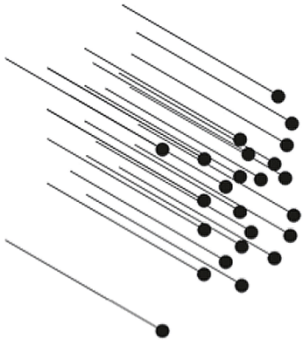
Conclusions

- Of 1,253,331,753,741 SNP-transcript interactions, 375,127 eQTL were found
- Risk allele status does not predict eQTL status
- ER status can interact with direction of eQTL
- Germline genotype can lend insight into Breast cancer phenotype

Acknowledgements

Beth Israel Deaconess Medical Center

A teaching hospital of Harvard Medical School



BECK LAB

- Andrew Beck

HARVARD SCHOOL OF PUBLIC HEALTH

- Aditi Hazra
- Pete Kraft
- John Quackenbush
- Constance Chen

THE CANCER GENOME ATLAS