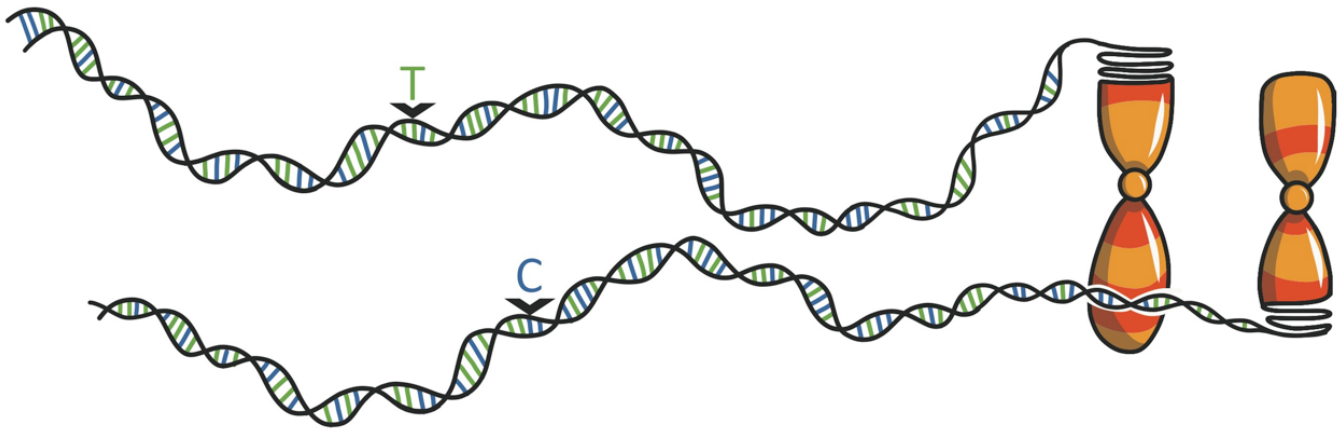
Omnigenic architecture of human complex traits

Jonathan Pritchard

Departments of Genetics & Biology, & HHMI

Stanford University

Joint work with Evan Boyle, Yang Li, Xuanyao Liu



Missing Heritability Workshop,
NIH, May 2018

Questions:

1. Why do the lead hits for any given trait contribute so little heritability?

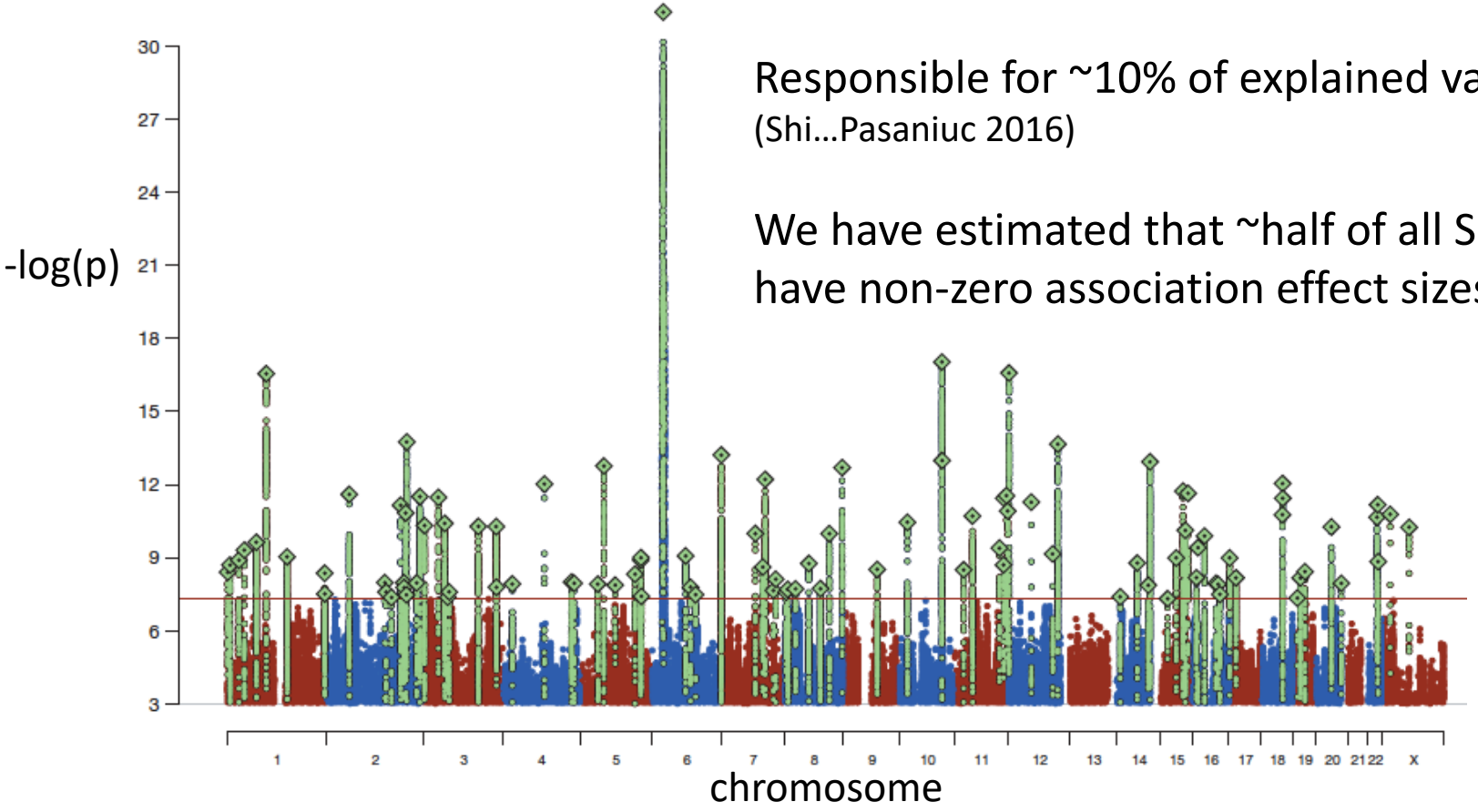
2. Why does so much of the genome contribute to heritability?

Example #1: Schizophrenia

108 genome-wide significant loci so far (Ripke 2014)

Responsible for ~10% of explained variance (Shi…Pasaniuc 2016)

We have estimated that ~half of all SNPs have non-zero association effect sizes (unpub)



See key work on polygenic models and heritability by Visscher, Yang, Pasaniuc, Price, and many others

Example #2:
What about a potentially simpler trait: lipid levels?
(LDL, HDL and triglycerides)

Monogenic lipid disorders
~2 dozen major effect loci

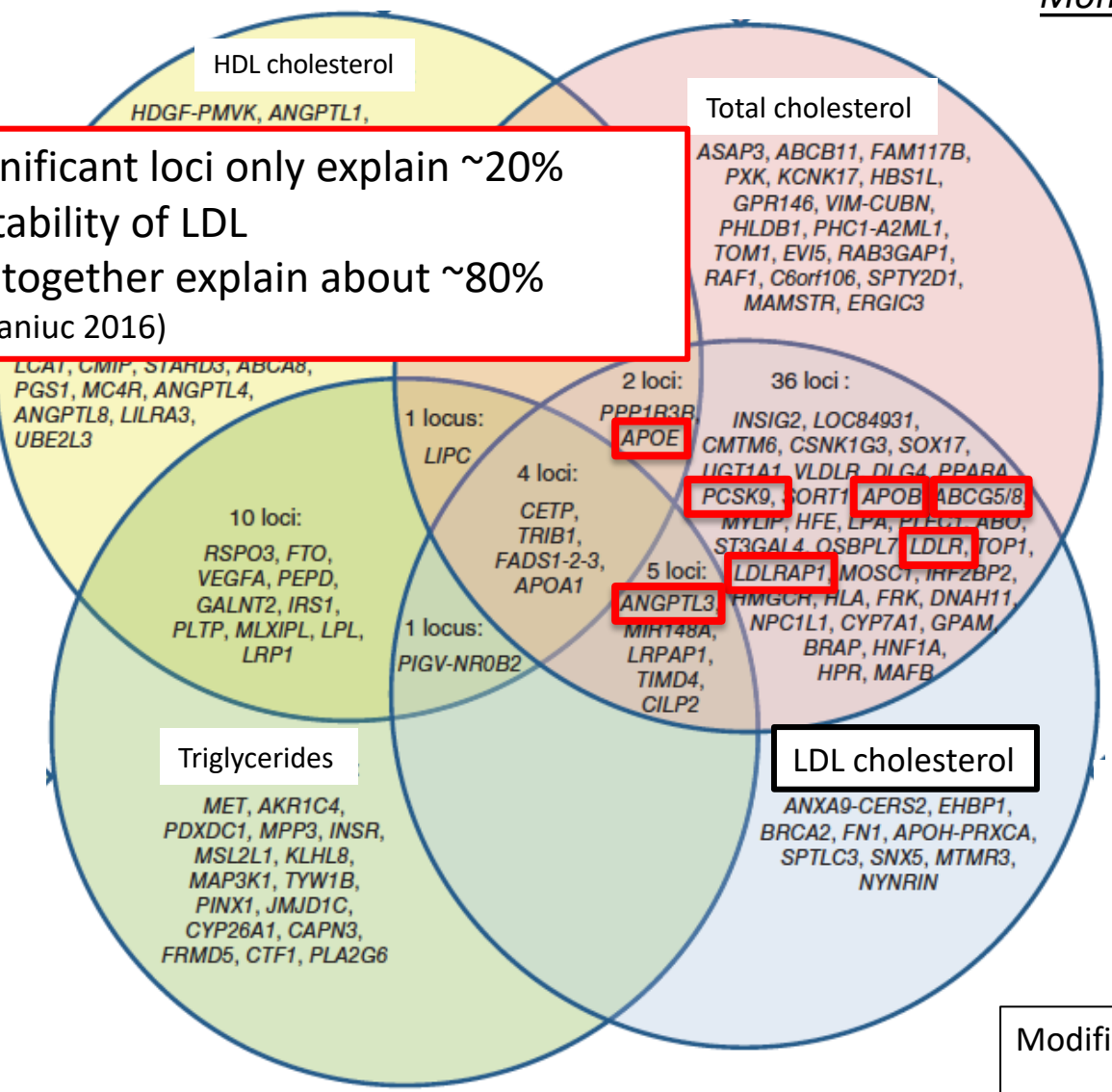
High LDL-C	Familial hypercholesterolemia	*LDLR*
	Familial defective apolipoprotein B	*APOB*
	Autosomal dominant hypercholesterolemia	*PCSK9*
	Autosomal dominant hypercholesterolemia	*STAP1*
	Autosomal dominant hypercholesterolemia	*APOE*
	Autosomal recessive hypercholesterolemia	*LDLRAP1 (ARH)*
	Cholesterol ester storage disease	*LIPA*
	Sitosterolemia	*ABCG5/ ABCG8*
Low LDL-C	Abetalipoproteinemia	*MTTP*
	Hypobetalipoproteinemia	*APOB*
	PCSK9 deficiency with low LDL-C	*PCSK9*
	Familial combined hypolipidemia	*ANGPTL3*
		SAR1B

Modified from Dron et al 2016

Common Variation: GWAS of Lipid Levels

57 genome-wide significant loci (Willer et al 2013)

The significant loci only explain ~20% of heritability of LDL
All loci together explain about ~80%
(Shi…Pasaniuc 2016)



Monogenic genes for LDL

- *LDLR*
- *APOB*
- *PCSK9*
- *STAP1*
- *APOE*
- *LDLRAP1 (ARH)*
- *LIPA*
- *ABCG5/ ABCG8*
- *MTTP*
- *ANGPTL3*
- *SAR1B*

Modified from: Willer et al 2013, Dron et al 2016

For a wide variety of traits and diseases:

- Heritability is spread extremely widely across the genome
- Genes with trait-relevant functions only contribute a small fraction of the total disease risk
- Low frequency-large effect variants often have clearer enrichment in relevant gene sets
- Contributing variants are highly concentrated in regions that are active chromatin in relevant tissues

  (Implies that most effects mediated through gene regulation)

So how should we conceptualize the molecular links from genetic variation to complex traits?

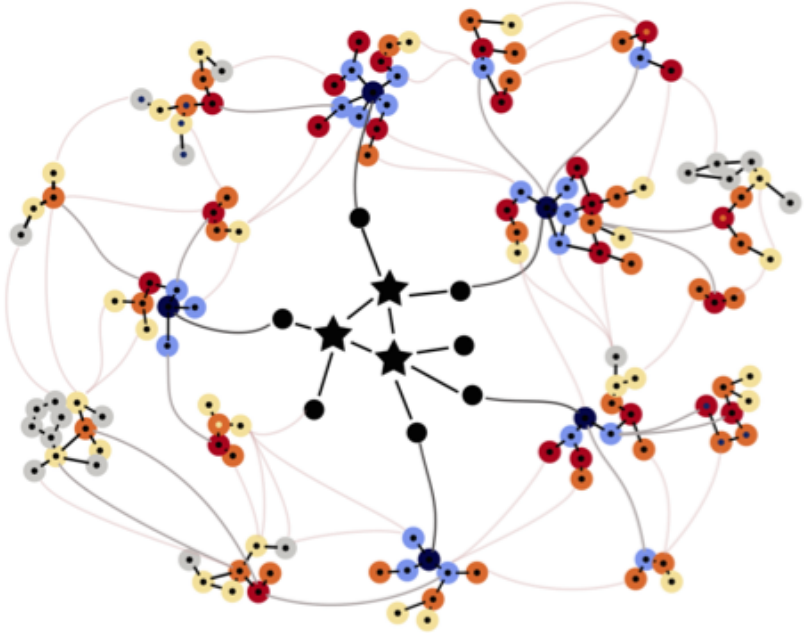
Our model to describe the data: The "*omnigenic*" model

3 types of genes:

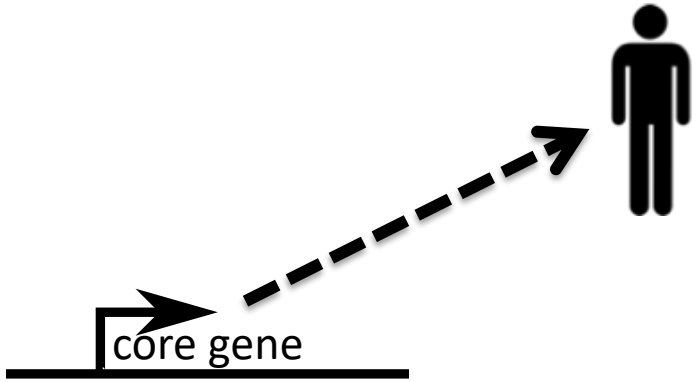
- Tier 1: Core genes: direct roles in disease
- Tier 2: Peripheral genes: essentially all other expressed genes can trans-regulate core genes
- Tier 3: Genes not expressed in the "right" cell types do not contribute to heritability

Most phenotypic variance is due to regulatory variation in peripheral genes

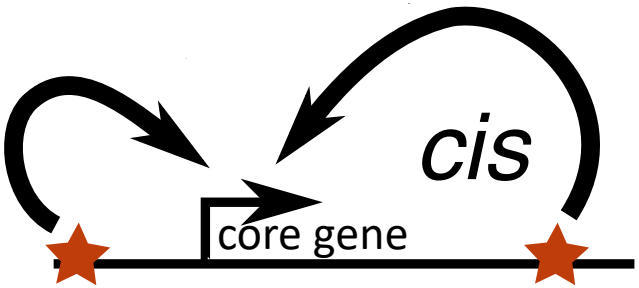
Hypothesis: Peripheral genes outnumber core genes by ~100:1, and likely dominate the phenotypic variance through weak effects rippling through gene networks



cis and *trans* regulation of core genes

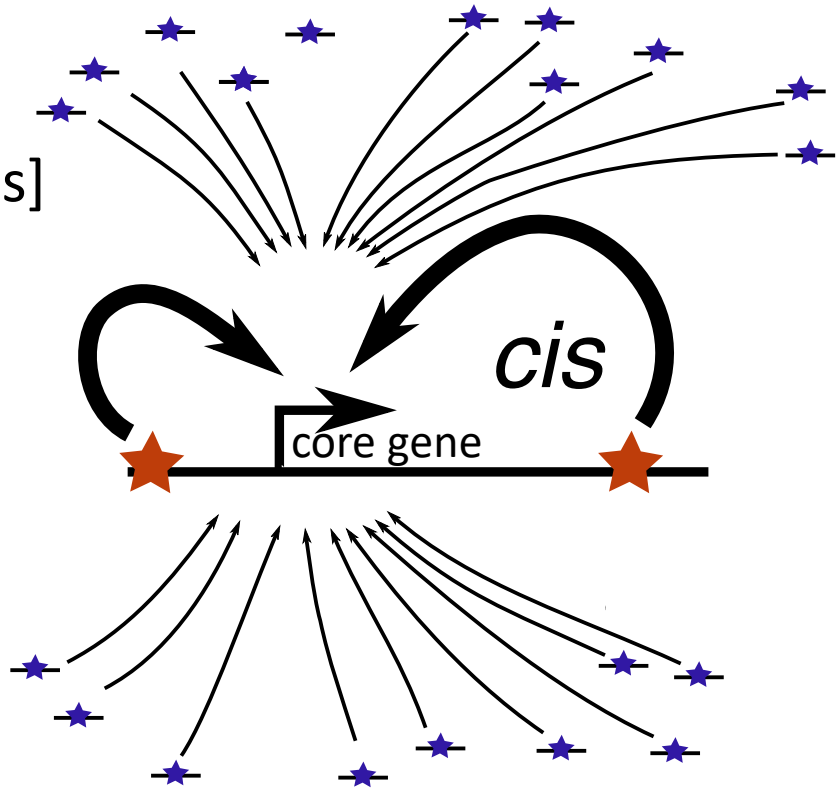


cis and *trans* regulation of core genes

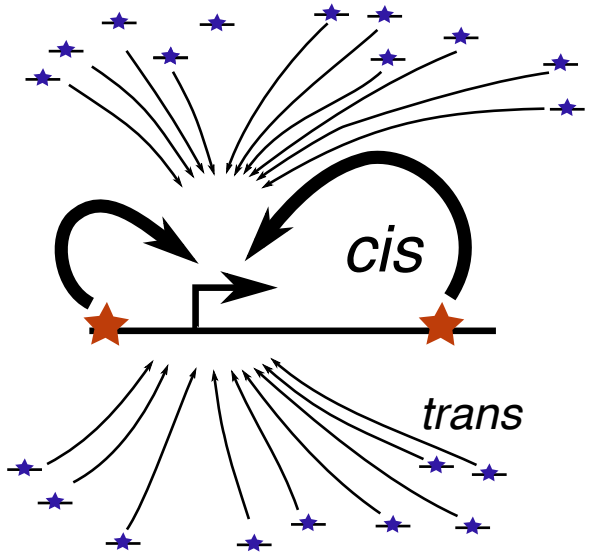


cis and *trans* regulation of core genes

Trans effects
[peripheral genes]



How much of expression variance is due to *cis* vs *trans* effects?



Literature review: genetic variance in gene expression

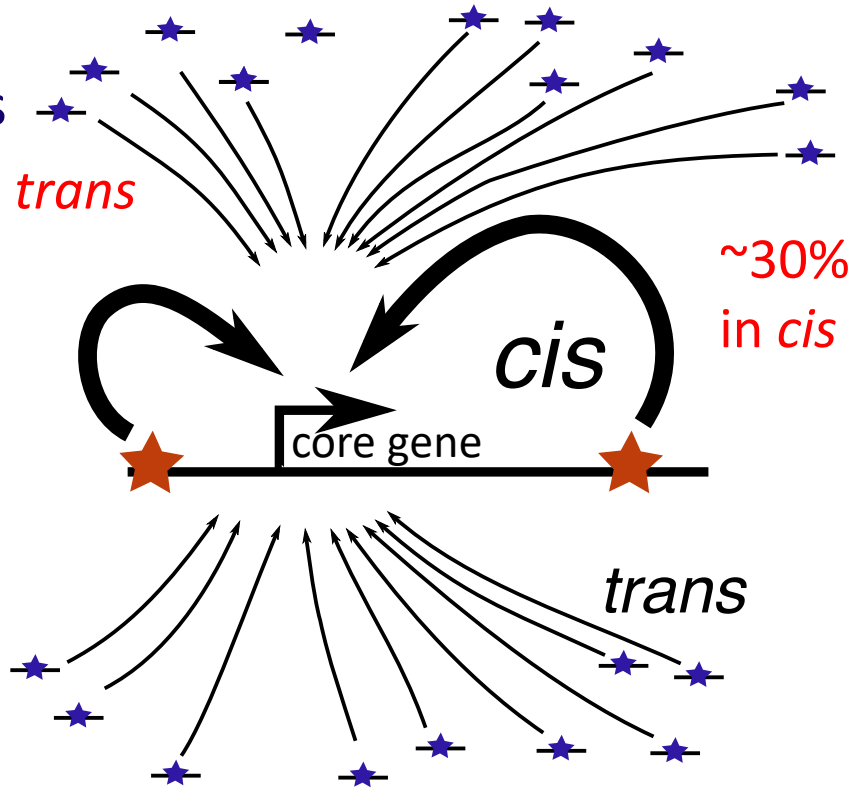
Percent h^2 in trans	Tissue/organism	Method	Reference
88%	lymphoblastoid	African-European ancestry	Price 2008
76%, 61%	Drosophila, whole body	fly hybrids	McManus 2010
76%, 63%	adipose, blood	cis/trans IBD in families	Price 2011
70%, 65%, 64%	adipose, LCL, skin	twin design	Grundberg 2012
77%, 69%	peripheral blood	twin design, LD Score	Wright 2014
72%	yeast segregants	cis vs. trans eQTLs	Albert 2017
62%	mouse liver	GCTA	Our group, In Prep
72%	mouse liver (protein)	GCTA	Our group, In Prep

~70% in trans

cis and *trans* regulation of core genes

peripheral genes

~70% of heritability in *trans*

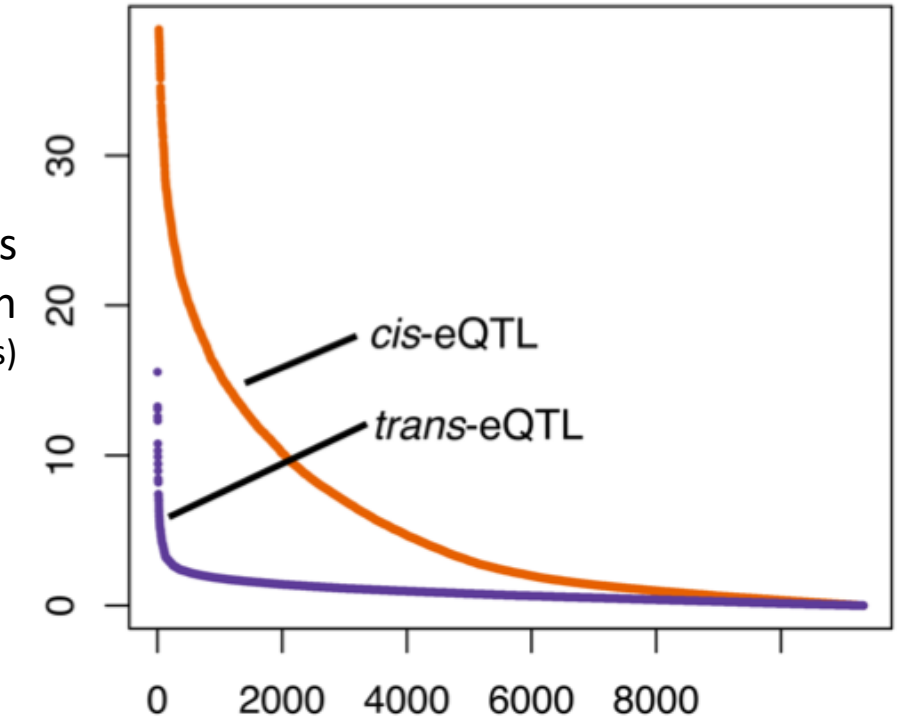


~30% of mRNA heritability in *cis*

But trans eQTLs have very small effect sizes compared to cis

Distribution of *cis* vs *trans* effect sizes

Effect sizes of SNPs on expression (|Z| scores)



Distribution of effect sizes for top hits, cis and trans

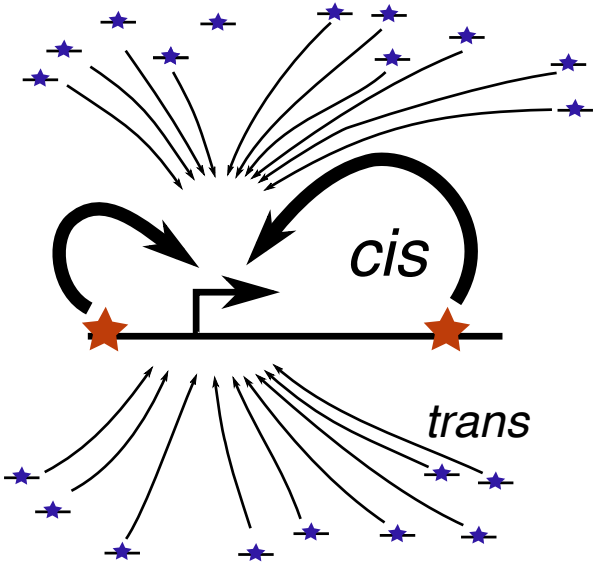
This difference is even more dramatic for (effect size)^2

Xuanyao Liu, unpub'd.

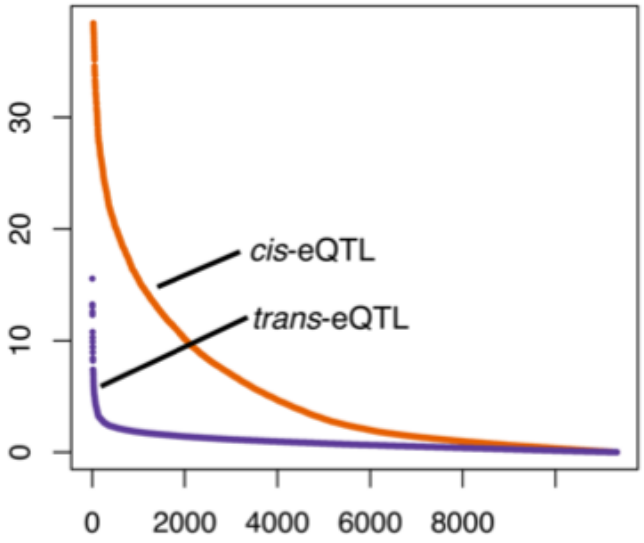
Plot shows replication effect sizes of strongest cis and trans signals from NTR into DGN

Together these observations imply that a typical gene must have huge numbers of weak *trans*-regulators

~70% of variance in *trans*



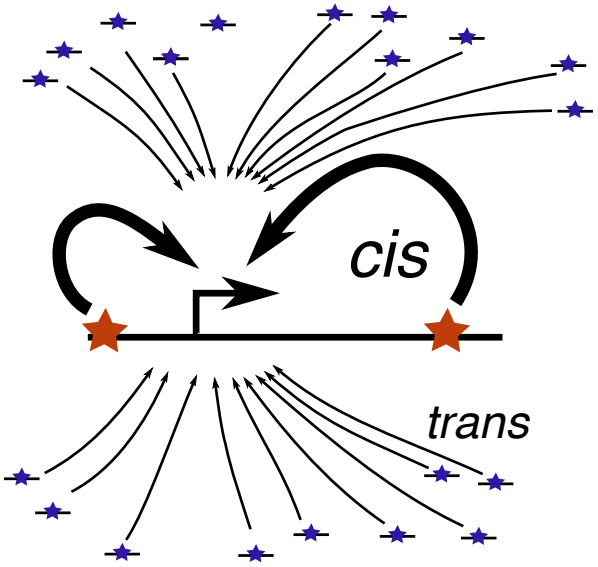
Cis associations much bigger than *trans*



Xuanyao Liu, unpub'd

Together these observations imply that a typical gene must have huge numbers of weak *trans*-regulators

~70% of variance in *trans*

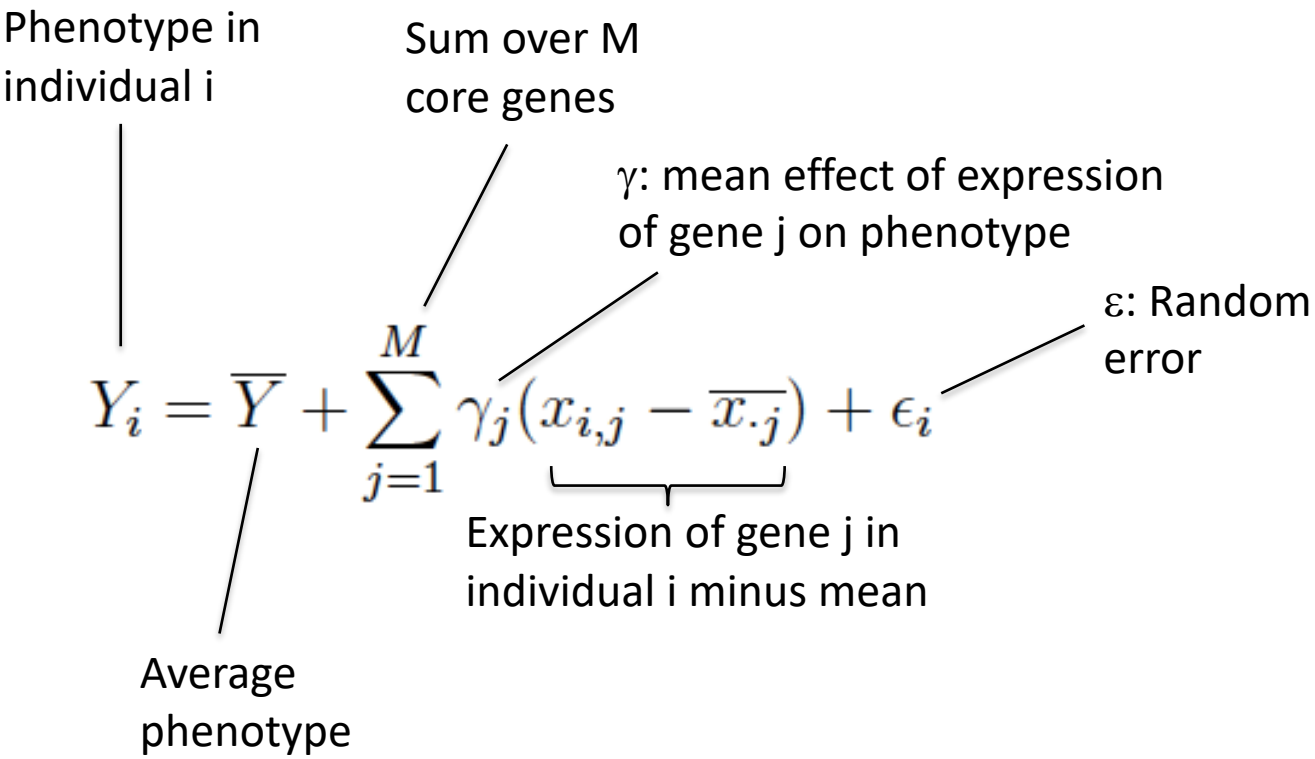


So assuming > tens of core genes, this model explains why such a large fraction of the genome can contribute to any given complex trait

Xuanyao Liu, unpub'd

One last question: why do core genes contribute so little heritability to any given trait?

A simple phenotype model based on expression of core genes



$$Y_i = \overline{Y} + \sum_{j=1}^{M} \gamma_j (x_{i,j} - \overline{x_{\cdot j}}) + \epsilon_i$$

Expression variance:
~1/3 cis, 2/3 trans.
M of these terms

Phenotypic variance

$$\mathrm{Var}(\mathrm{Y_i}) = \sum_{\mathrm{j=1}}^{\mathrm{M}} \gamma_{\mathrm{j}}^2 \mathrm{Var}(\mathrm{x_{\cdot j}}) + \sum_{\mathrm{j \neq k}}^{\mathrm{M}} \gamma_{\mathrm{j}} \gamma_{\mathrm{k}} \mathrm{Cov}(\mathrm{x_{\cdot j}}, \mathrm{x_{\cdot k}}) + \mathrm{Var}(\epsilon)$$

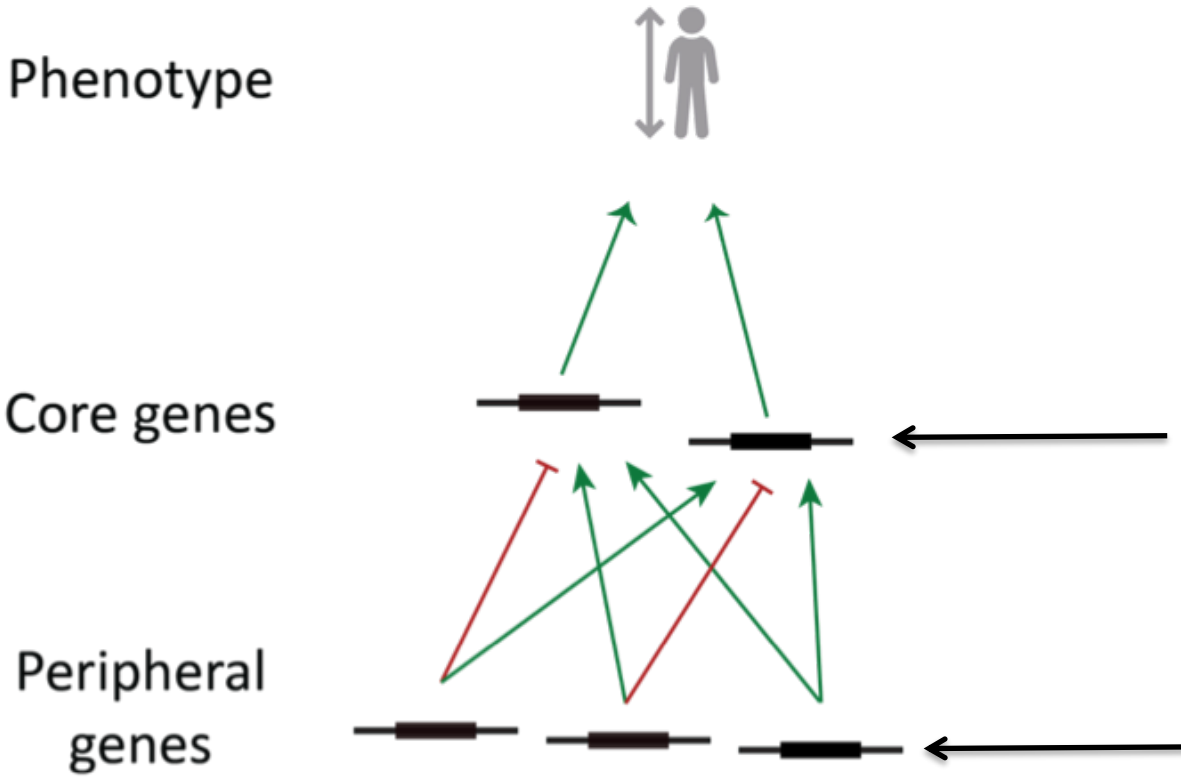
Expression covariance:
Dominated by trans effects
(peripheral genes)
Nearly M^2 of these terms

Two versions of core gene model yield divergent predictions

Model 1:

Expression covariances of core genes low

$$\mathrm{Var}(Y) = \sum_{j=1}^{M} \gamma_j^2 \mathrm{Var}(x_j) + \sum_{j \neq k}^{M} \gamma_j \gamma_k \mathrm{Cov}(x_j, x_k)$$



~30% of expression variance in *cis*

~70% of expression variance in *trans*

~30% of heritability *cis* to core genes

Two versions of core gene model yield divergent predictions

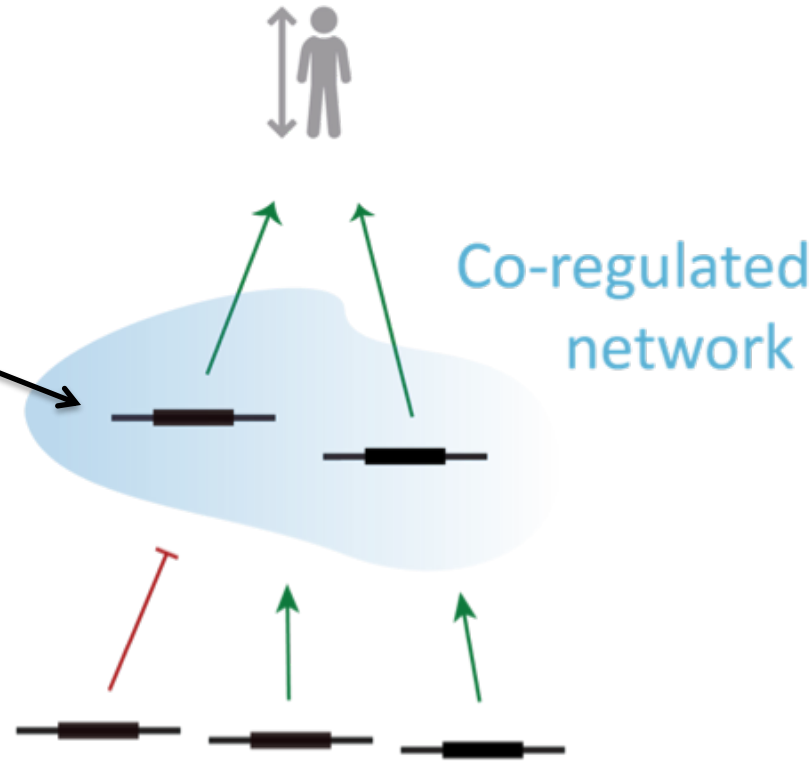
$$\mathrm{Var}(Y) = \sum_{j=1}^{M} \gamma_j^2 \mathrm{Var}(x_j) + \sum_{j \neq k}^{M} \gamma_j \gamma_k \mathrm{Cov}(x_j, x_k)$$

Model 2:

Expression covariances of core genes high

cis effects independent for each core gene

Trans effects often shared across core genes



Most of the heritability transferred to peripheral genes

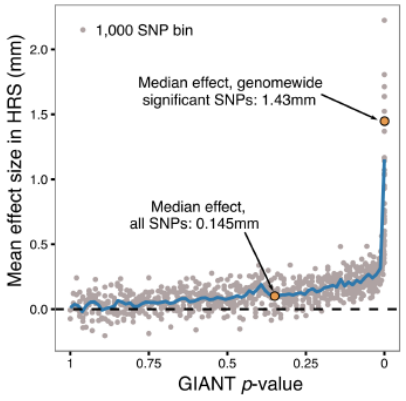
Boyle, Li & Pritchard
Cell 2017

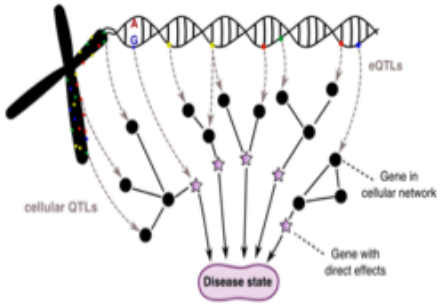
Conclusions (1)

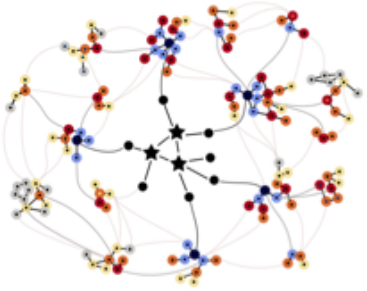
We propose that gene regulatory networks are sufficiently interconnected that

- all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes
- most heritability is due to SNPs outside core pathways.

We refer to this hypothesis as an ‘‘omnigenic’’ model.







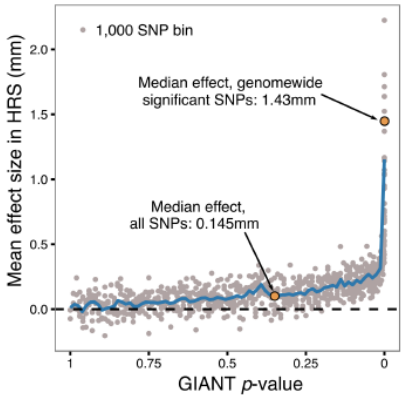
Boyle, Li & Pritchard
Cell 2017

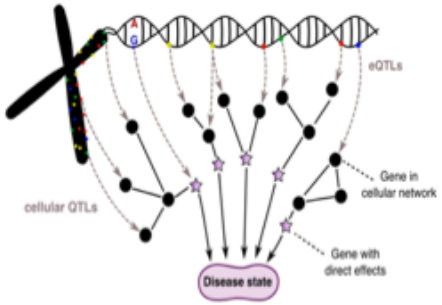
Conclusions (2)

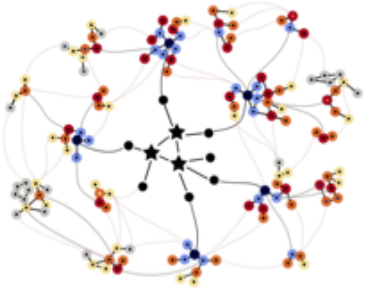
This model is consistent with known properties of *cis*- and *trans*-eQTLs

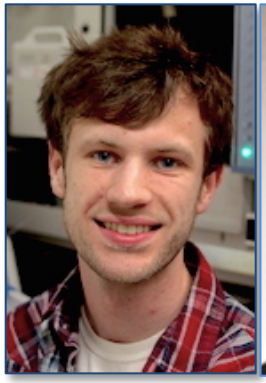
- *trans*-variation is responsible for ~70% of expression heritability
- But effect sizes are nearly uniformly tiny

- Co-regulated gene networks act as amplifiers for peripheral variation








Evan Boyle


Yang Li


Xuanyao Liu

Thanks to many colleagues for great discussions; NIH & HHMI for funding.

We have a draft in prep on the new work (goal: end of May). Please email me if you would like a pre-preprint

pritch@stanford.edu

Lab Reunion 2016

Boyle, Li & Pritchard
Cell 2017

Conclusions (3)

Gene-mapping serves two main goals

- Genetic prediction

  For this, GWAS is essential

- Identification of core genes and pathways

  Some combination of deep exome sequencing to find rare variants with large effects with more GWAS + methods for network inference

Importance of studying long-range network effects of variation