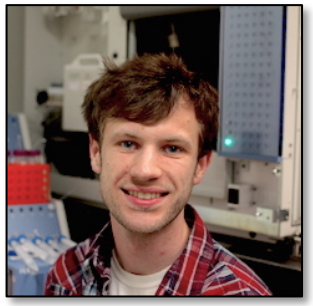
BIRS 2017

An expanded view of complex traits: from polygenic to omnigenic

How does human genetic variation drive variation in complex traits/disease risk?

Yang I Li
Stanford University

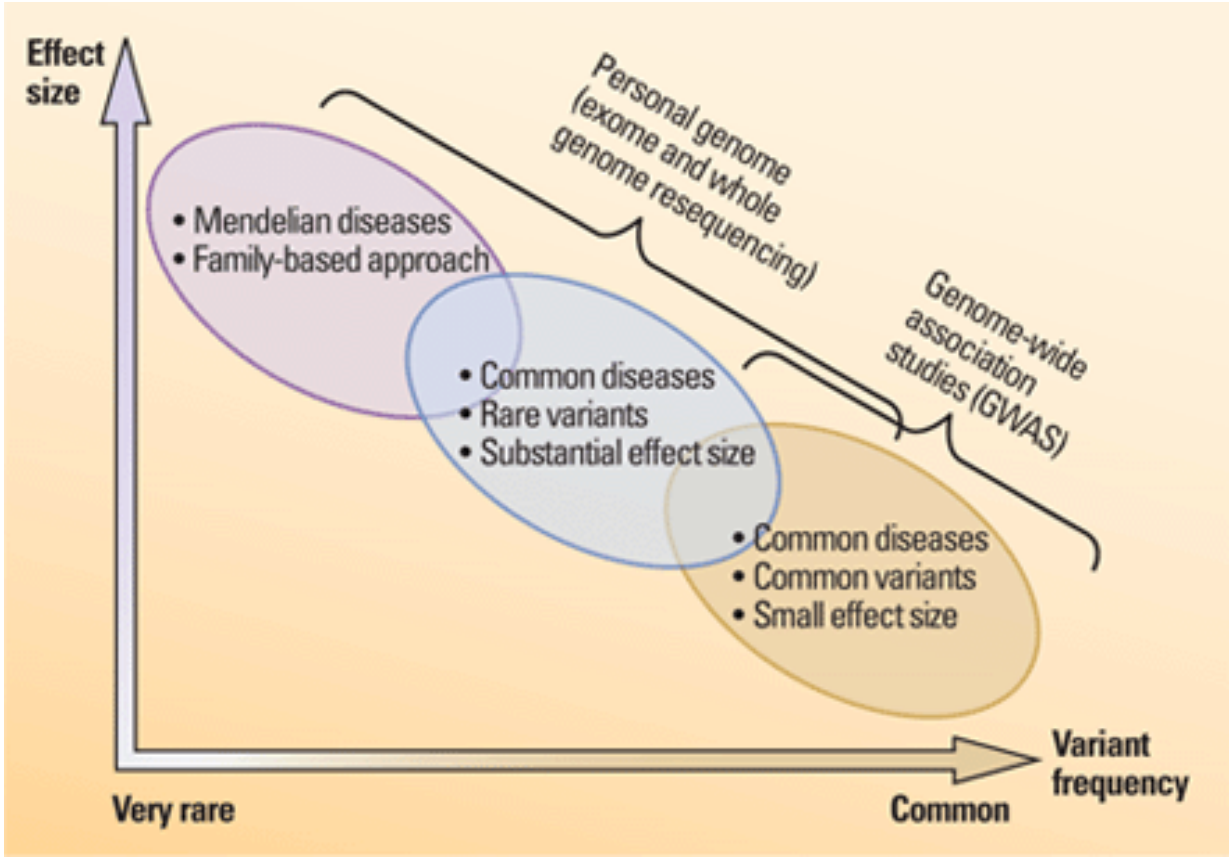


Evan Boyle

Jonathan Pritchard

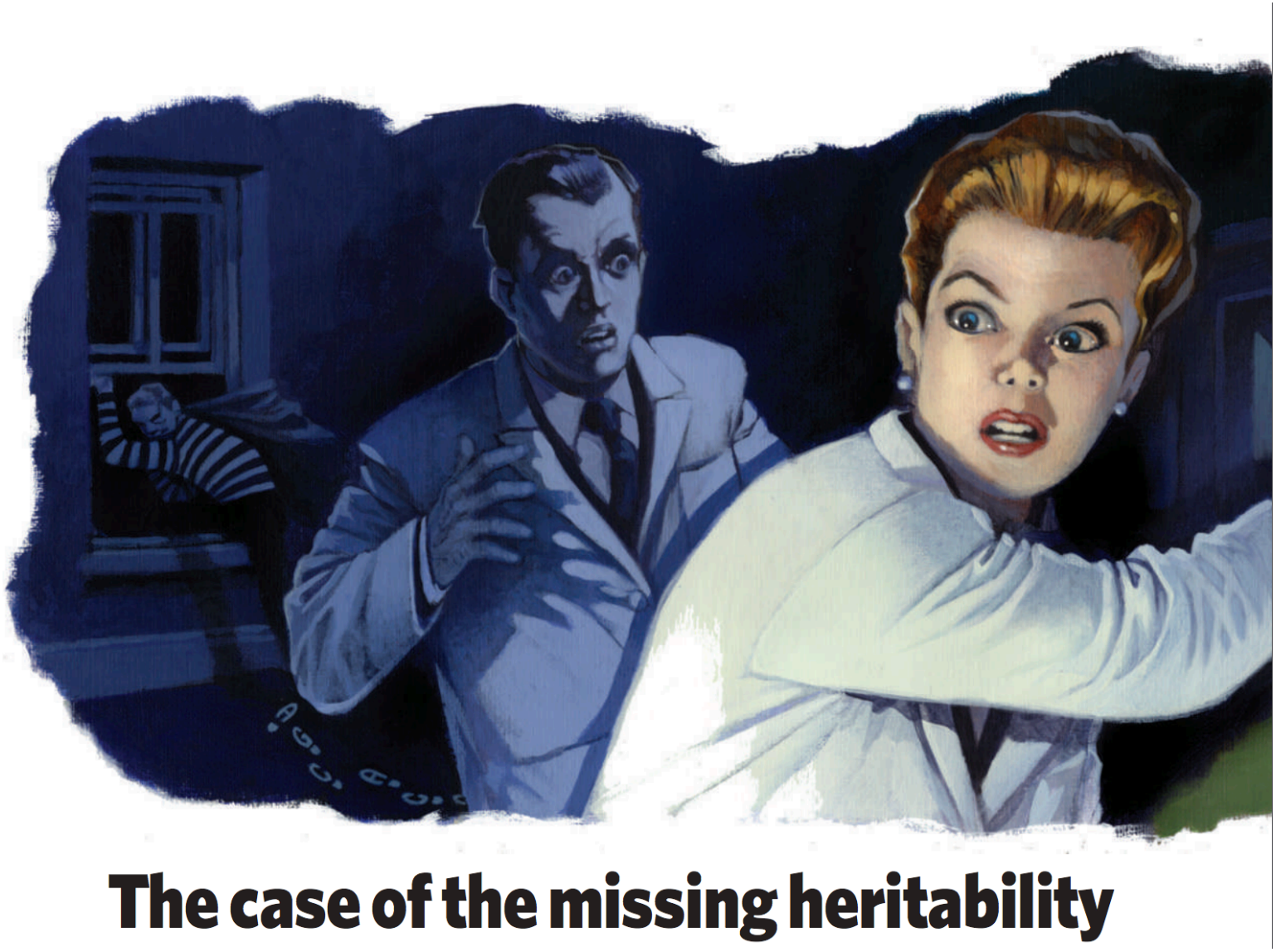
Stanford University

What did we learn from genome-wide association studies?



- Very polygenic architecture.
- Effect sizes are generally very small.

(Kaiser 2012, *Science*)



- GWAS significant hits explain only a small % of heritability.

(Maher 2008, *Nature*)

Complex traits variation mostly explained by common variants

Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index

Jian Yang[1,2,24], Andrew Bakshi[1], Zhihong Zhu[1], Gibran Hemani[1,3], Anna A E Vinkhuyzen[1], Sang Hong Lee[1,4], Matthew R Robinson[1], John R B Perry[5], Ilja M Nolte[6], Jana V van Vliet-Ostaptchouk[6,7], Harold Snieder[6], The LifeLines Cohort Study[8], Tonu Esko[9–12], Lili Milani[9], Reedik Mägi[9], Andres Metspalu[9,13], Anders Hamsten[14], Patrik K E Magnusson[15], Nancy L Pedersen[15], Erik Ingelsson[16,17], Nicole Soranzo[18,19], Matthew C Keller[20,21], Naomi R Wray[1], Michael E Goddard[22,23] & Peter M Visscher[1,2,24]

- GWAS significant hits explain only a small % of heritability.
- But considering all common SNPs explains most of heritability.

Of course there may be some exceptions...

PGC group	N_{case}	Hits	Twin-h^2	SNP-h^2	Strongest genetic correlation
Schizophrenia	60,995	155	81%	45%	Bipolar disorder
MDD	130,664	20+	30-40%	14%	Neuroticism
Bipolar disorder	20,352	19	80%	21%	
ADHD	20,183	12	70-80%	22%	Educational attainment
Autism spectrum disorder	18,381	3	75%	12%	Subjective well-being
Anorexia nervosa	3,495	1	56%	~20%	Metabolic traits
Substance use disorders	3,772	1	50%	~10%	Smoking
Tourette syndrome	4,232	1	60-80%	58%	OCD
OCD	2,688	0	45-65%	37%	Tourette syndrome
PTSD	3,749	0	30-40%	5-35%	Schizophrenia

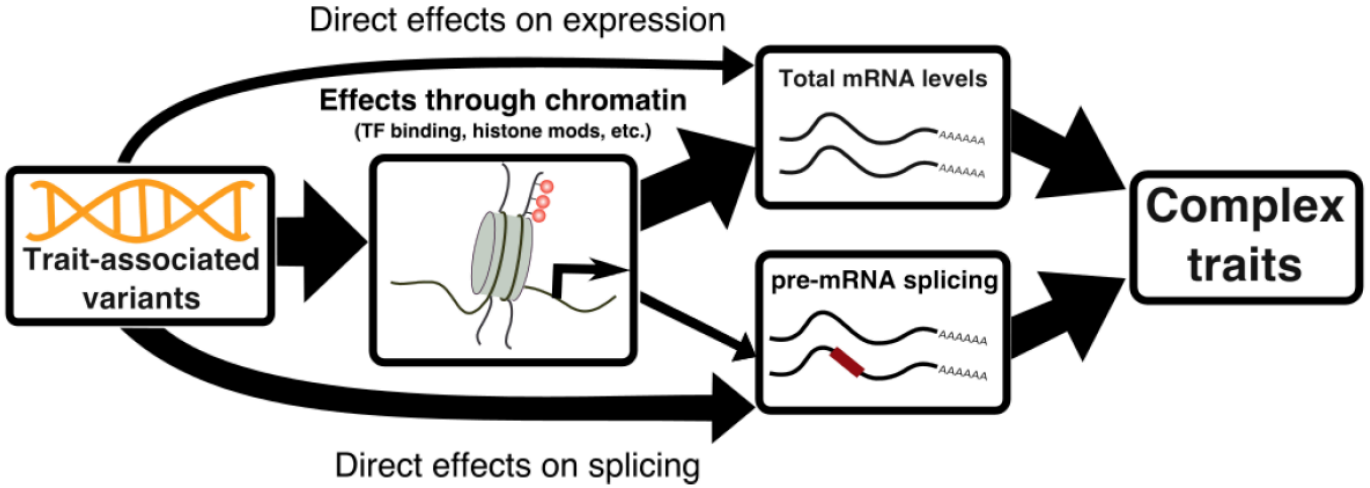
- Rare variants and de novo mutations play larger role in e.g. Autism.

(Sullivan et al., *bioRxiv, 2017*)

Regulatory variation, not coding variation, drive variation in complex traits

- Over 90% of GWAS SNPs are noncoding
- Enrichment in chromatin marks, eQTLs, sQTLs in relevant cell-types (**Farh et al., 2015**; **Trynka et al., 2013**; **Finucane et al., 2015**; and many others…)

Three primary regulatory mechanisms link common genetic variants to complex traits



(Li et al., 2016, *Science*)

What does this tell us about why people get disease?

The classic view would be that **causal variants** are **concentrated in core genes, pathways that drive disease/traits**.

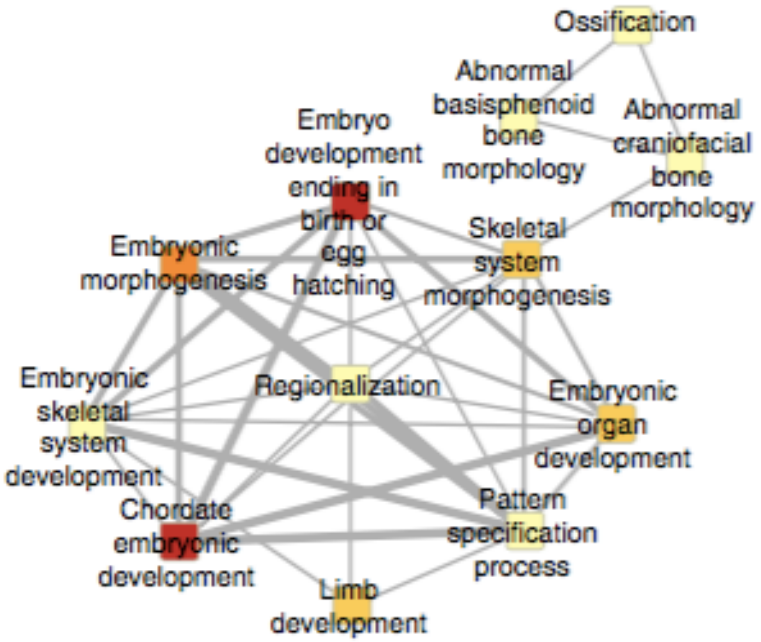
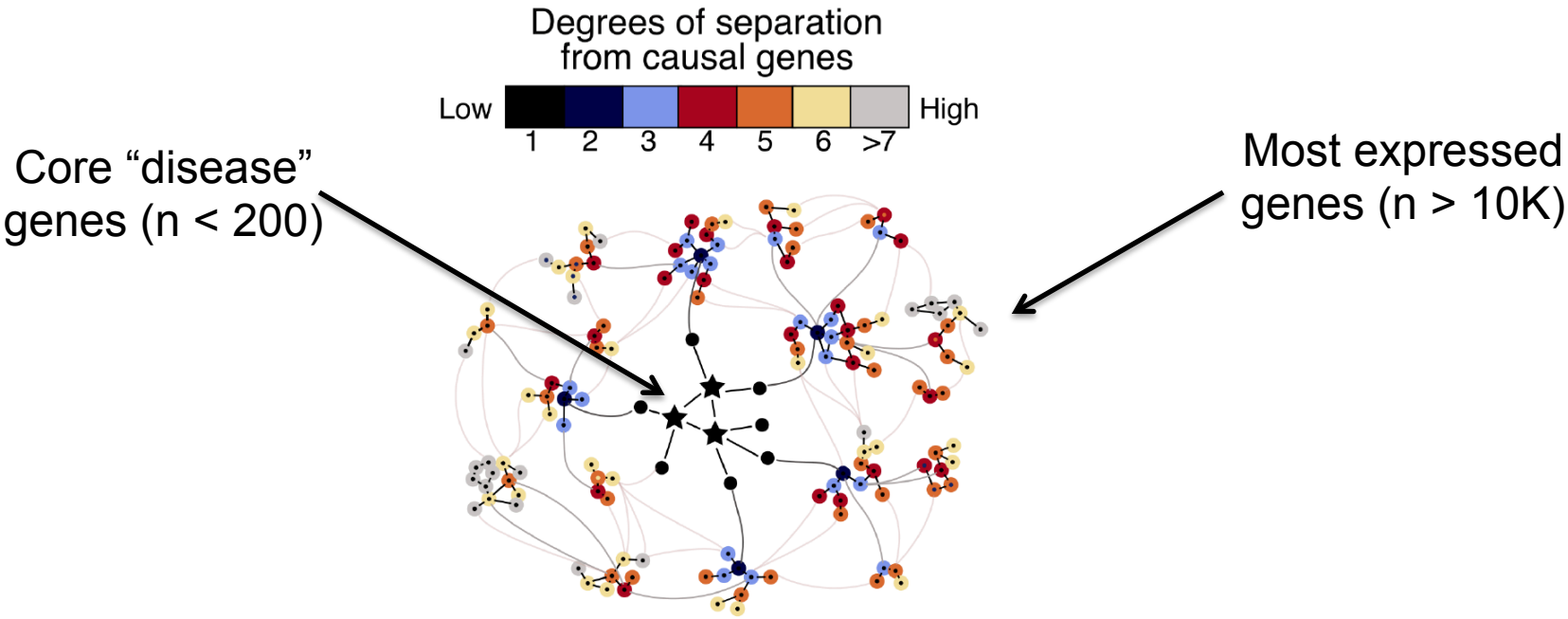


Figure from Wood et al 2014 NG

- Synaptic pruning in Schizophrenia (Sekar et al., 2016)
- Adipocyte differentiation in obesity (Smemo et al., 2014, Claussnitzer et al., 2015)

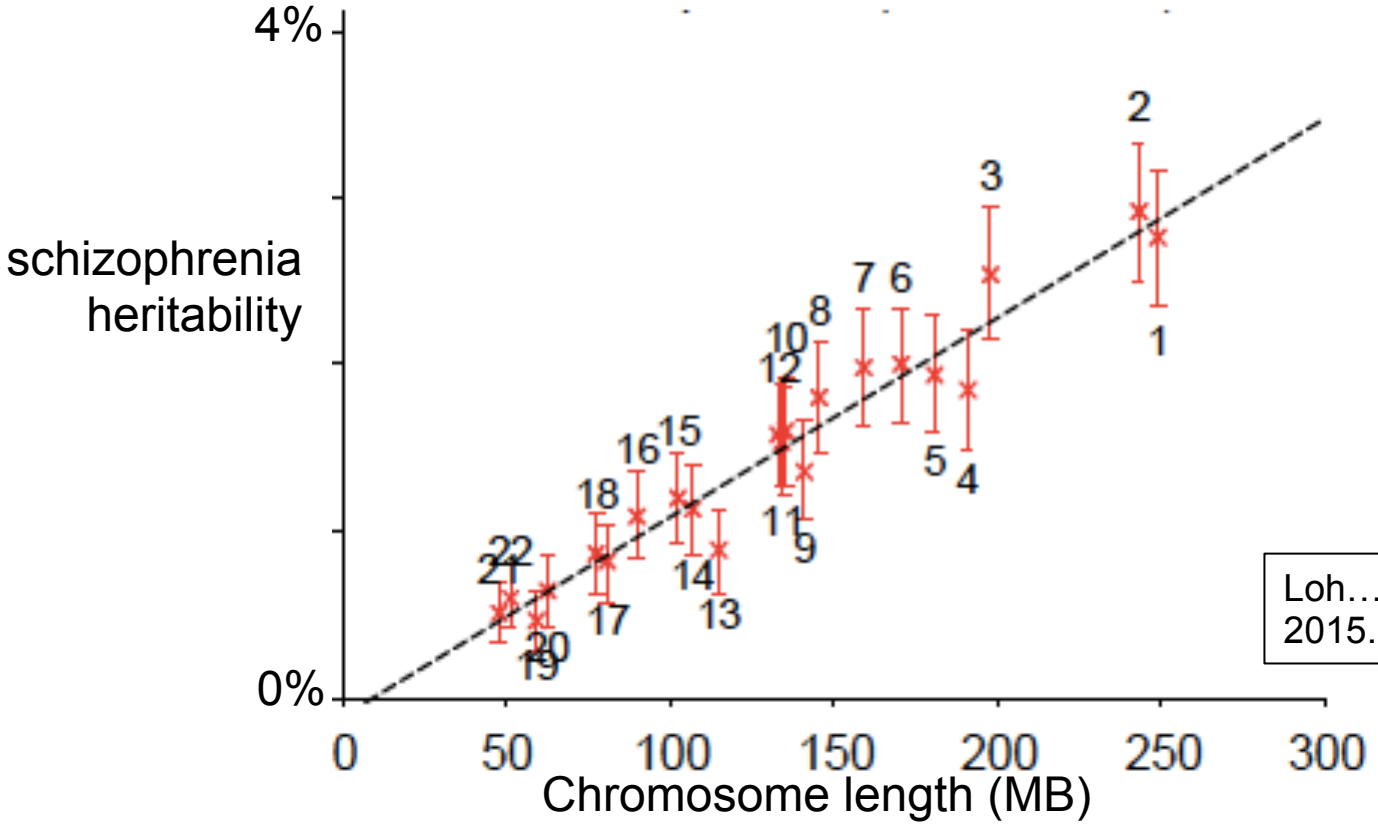
We argue that data from GWAS do not support this model: Heritability of many complex traits (e.g. schizophrenia) is (1) **spread very widely across the genome** and (2) **shows limited pathway enrichment**.



Instead we propose an updated model that hypothesizes that **most genes expressed in relevant tissues affect disease risk** through **highly-connected tissue-specific interaction networks**. (We refer to this hypothesis as the "**Omnigenic**" model.)

Observation #1: For many traits, causal loci are spread nearly uniformly across the genome

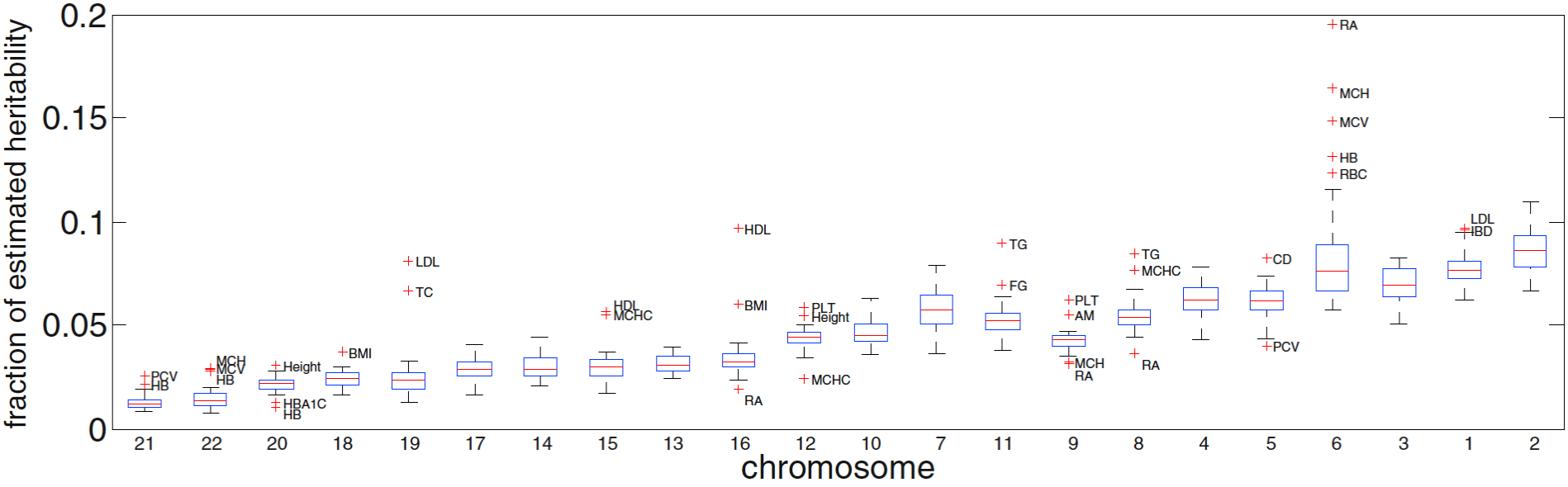
Amount of schizophrenia heritability explained by each chromosome is highly correlated with its length



Loh…Price, Nature Genetics 2015.

- At a broad scale, causal SNPs are spread widely across the genome
- Loh et al: >70% of MB windows in the genome contribute to schizophrenia heritability

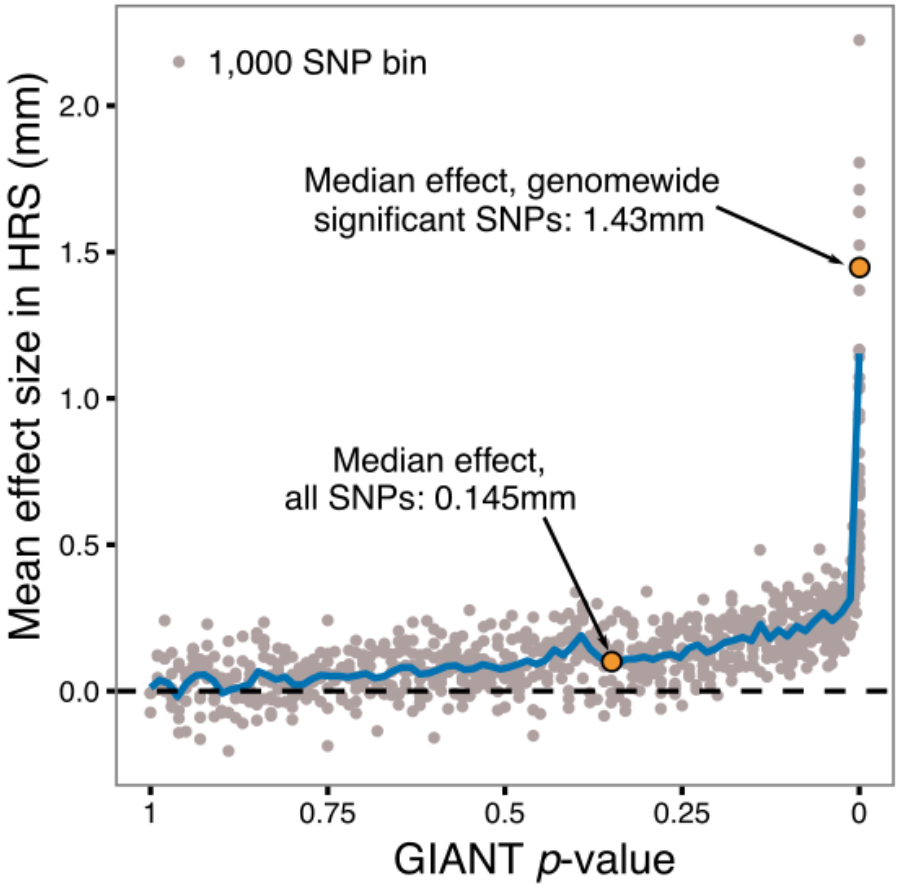
Nearly all complex traits show a strong polygenic signature at a broad scale



- 30 traits were considered ranging from autoimmune diseases, to anthropomorphic traits, to metabolic traits, etc...

Shi et al. AJHG 2016

Replication of height signal throughout p-value range



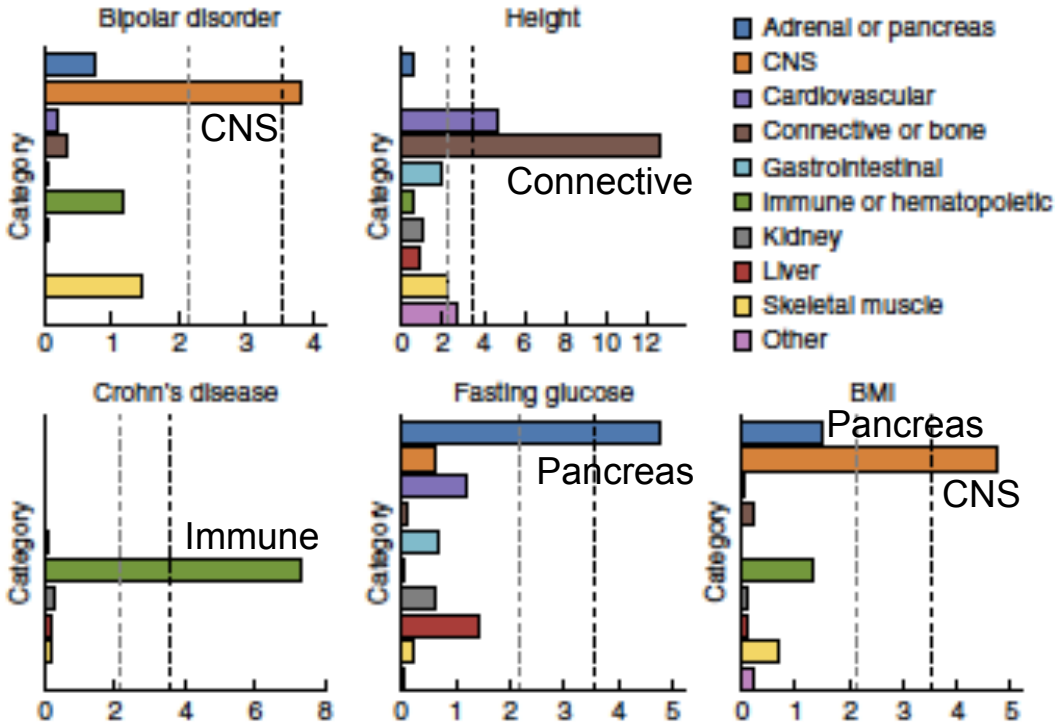
- Median effect of the 697 significant hits on height is 1.43mm, median effect of all SNPs is quite large: 1/10th of that, i.e. 0.145mm.
- Conservative back-of-the-envelope calculation: >23K independent genetic loci affect height (need >150K SNPs to explain height variance in human).

Evan Boyle

For many traits, causal SNPs are spread nearly uniformly across the genome

But "disease genes" are not, so causal SNPs might often target other genes?

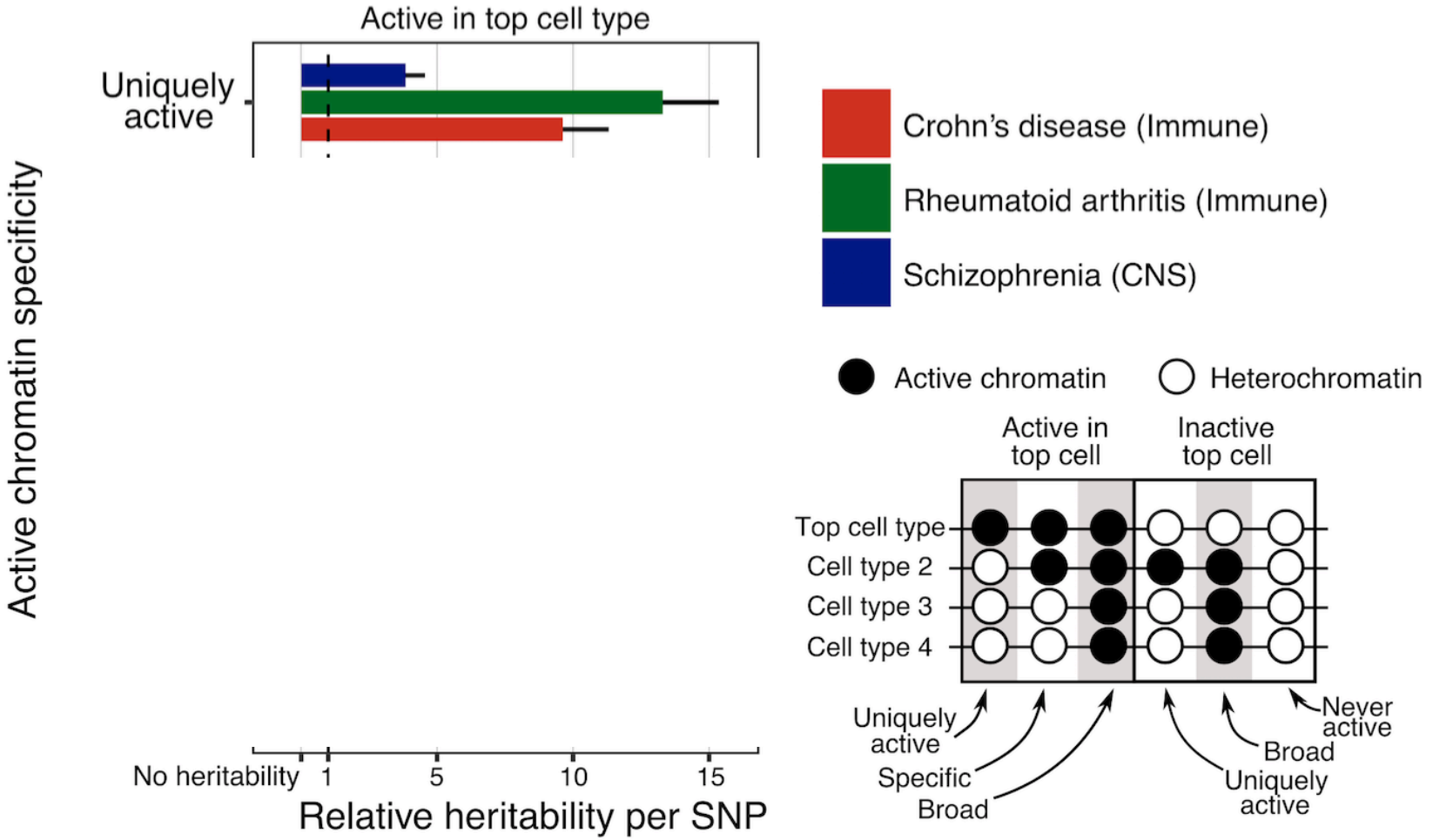
Observation #2: GWAS signals are enriched in chromatin that is active in cell-types that “make sense”.



Analysis by Stratified LD Score regression
[Finucane…Alkes Price (2015) NG]

- GWAS SNPs generally affect cell-type-specific processes (classic view).
- GWAS SNPs affect non-specific processes as long as they are “active” (omnigenic view).

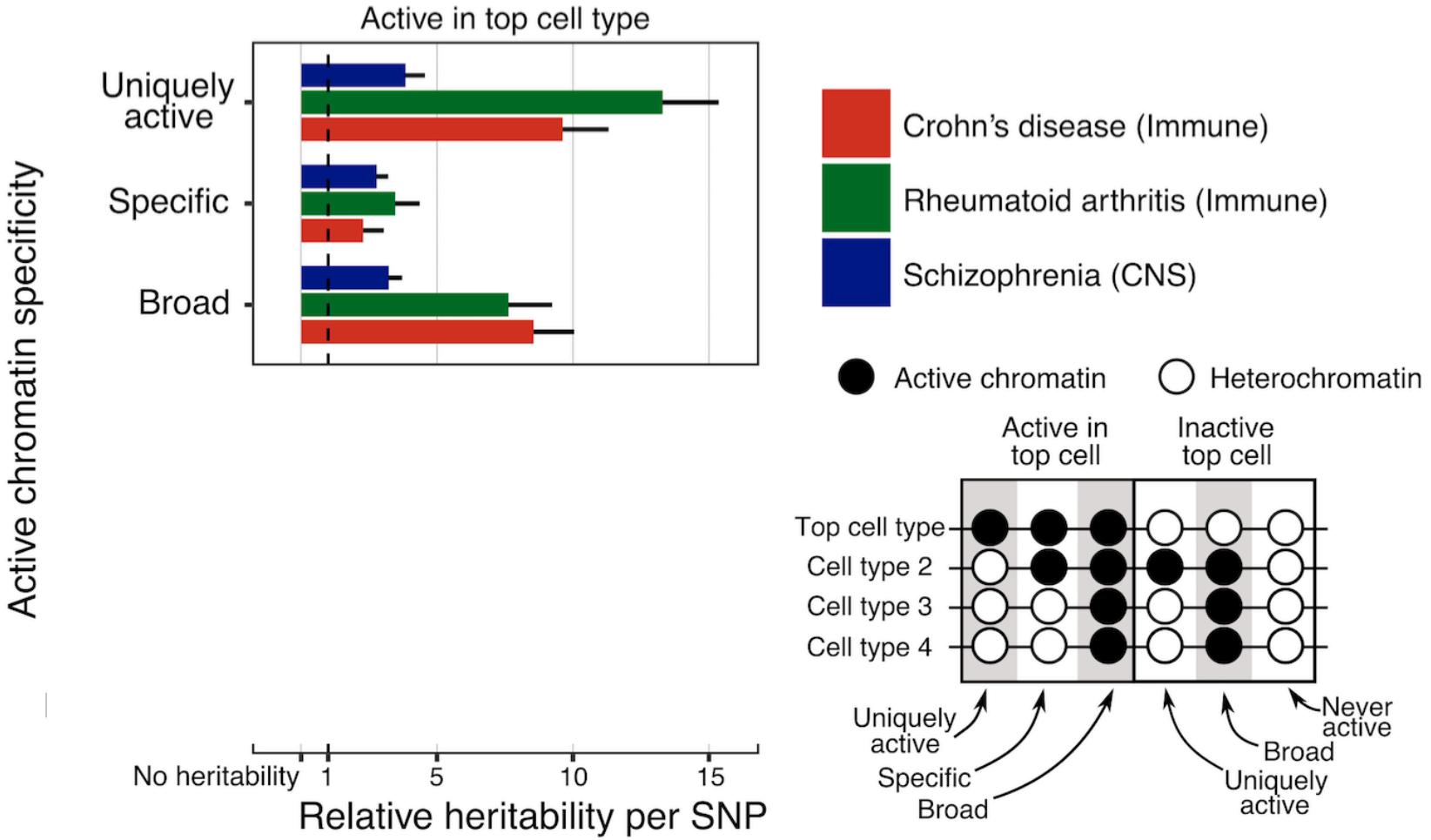
Disease heritability is enriched in chromatin uniquely active in relevant cell types



Analysis using Stratified LD Score regression and annotations from Finucane et al. 2015

Evan Boyle

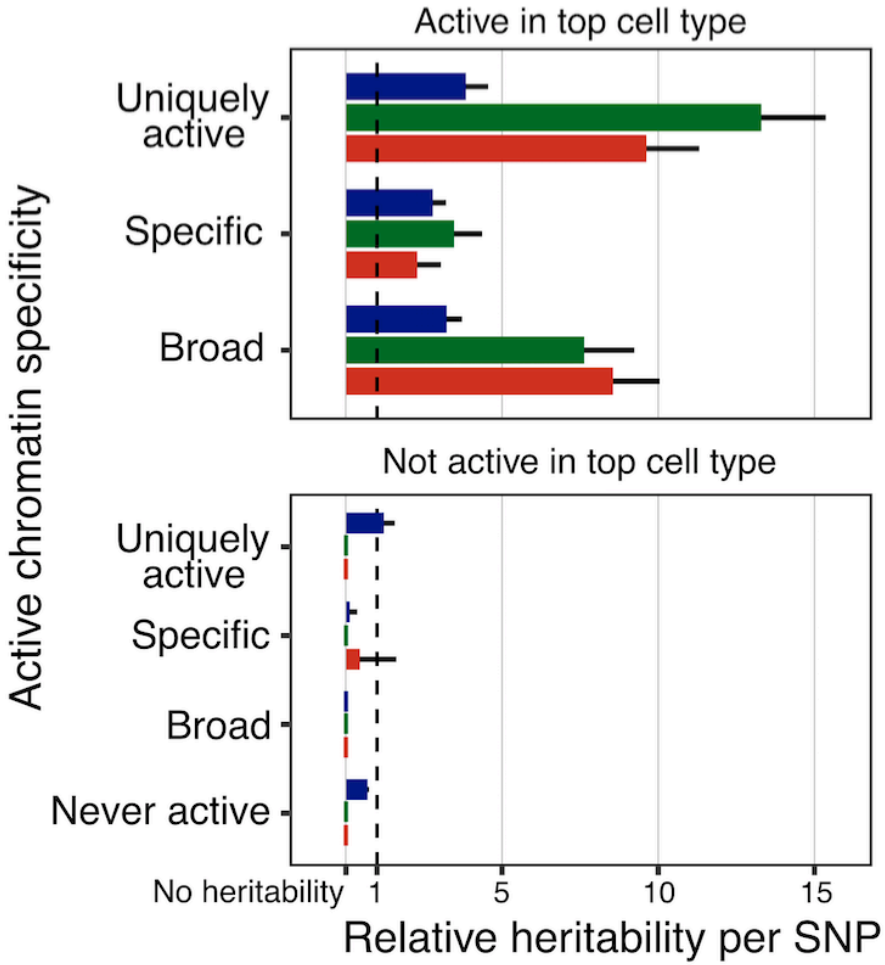
Observation #2A: It doesn't matter much whether the chromatin is broadly active, or uniquely active

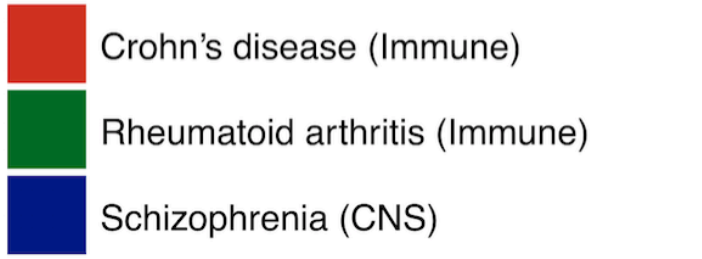


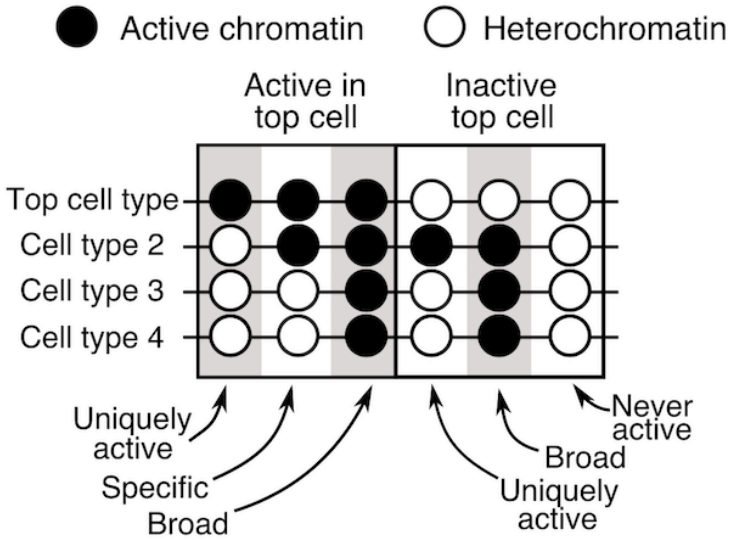
Analysis using Stratified LD Score regression and annotations from Finucane et al. 2015

Evan Boyle

What matters is that they are active in the relevant cell type



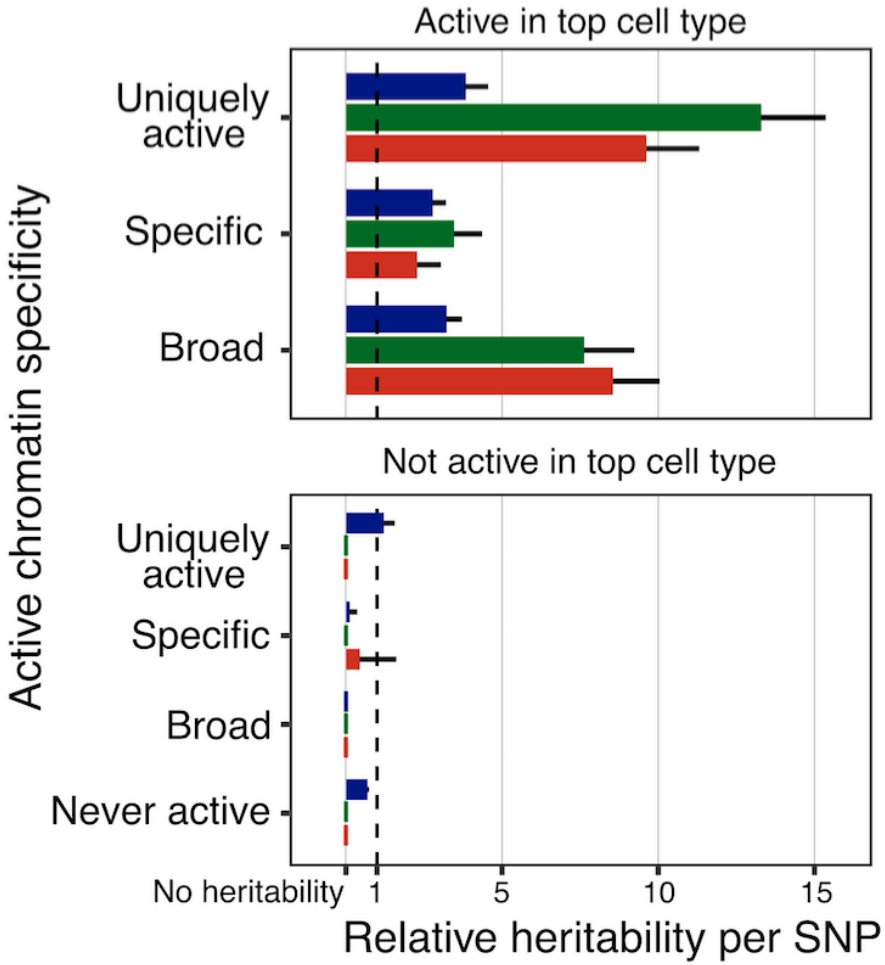


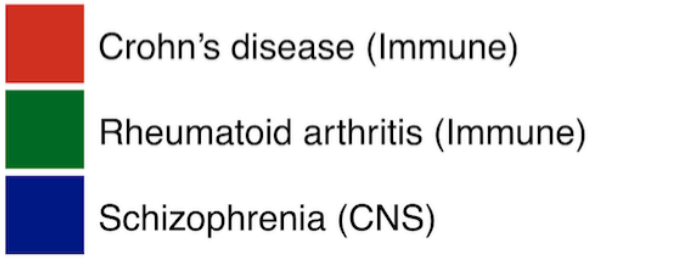


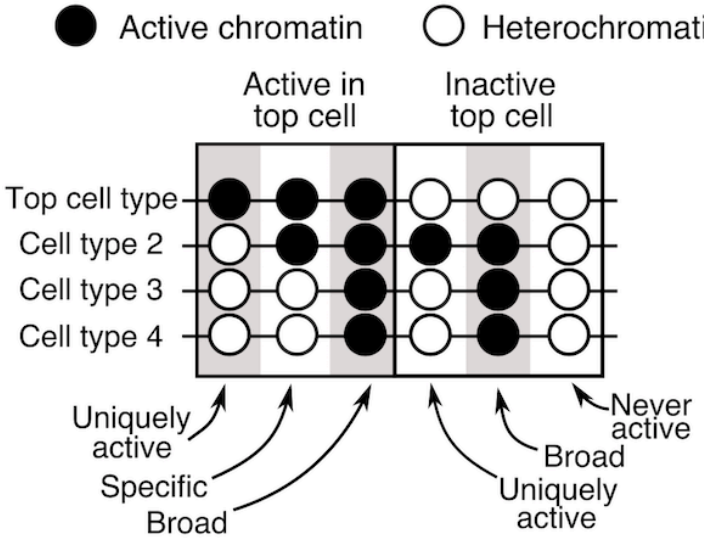
Analysis using Stratified LD Score regression and annotations from Finucane et al. 2015

Evan Boyle

What matters is that they are active in the relevant cell type





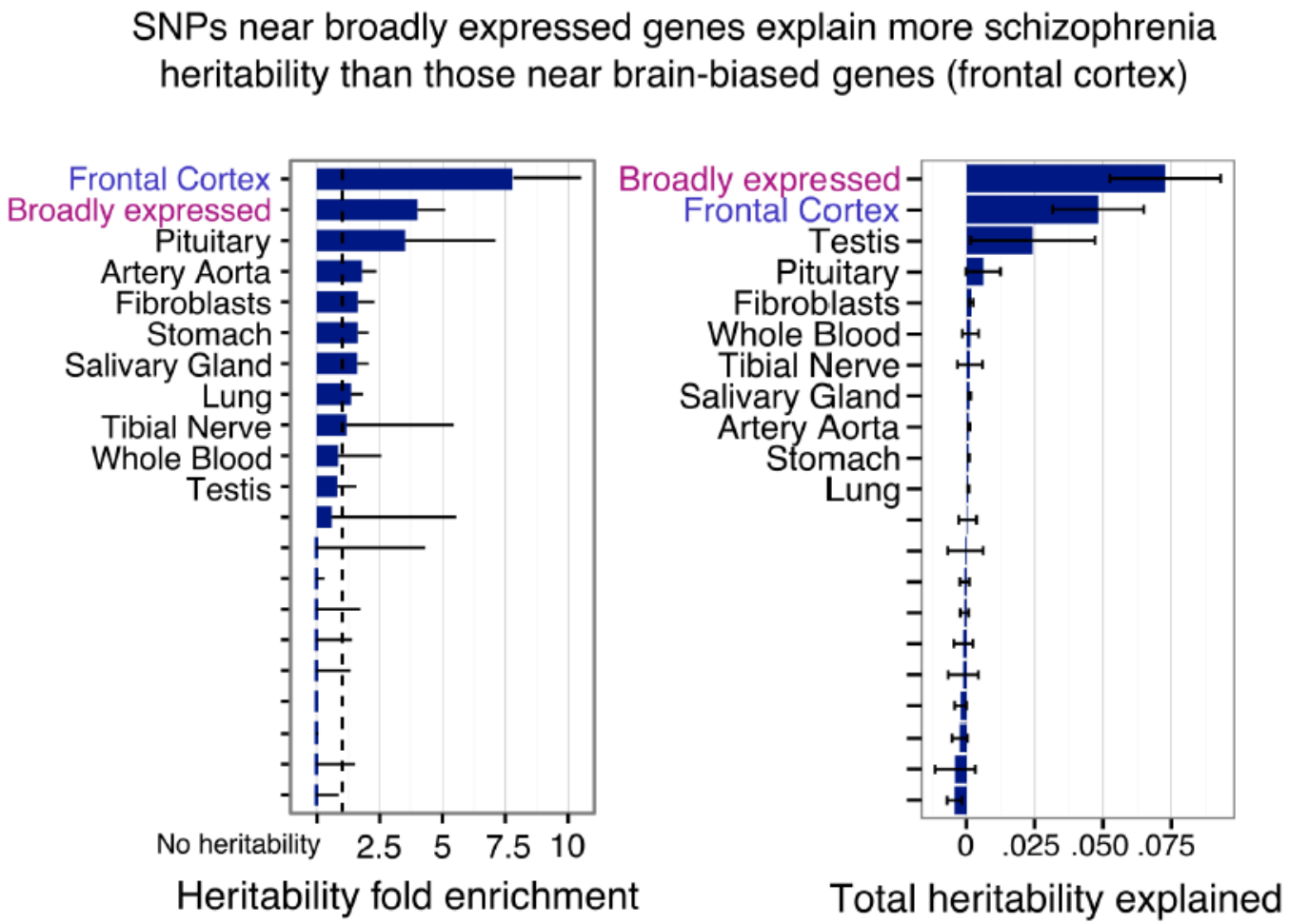


Genetic effects are not mediated through cell-type-specific function?

Analysis using Stratified LD Score regression and annotations from Finucane et al. 2015

Evan Boyle

Heritability near genes with preferential tissue expression

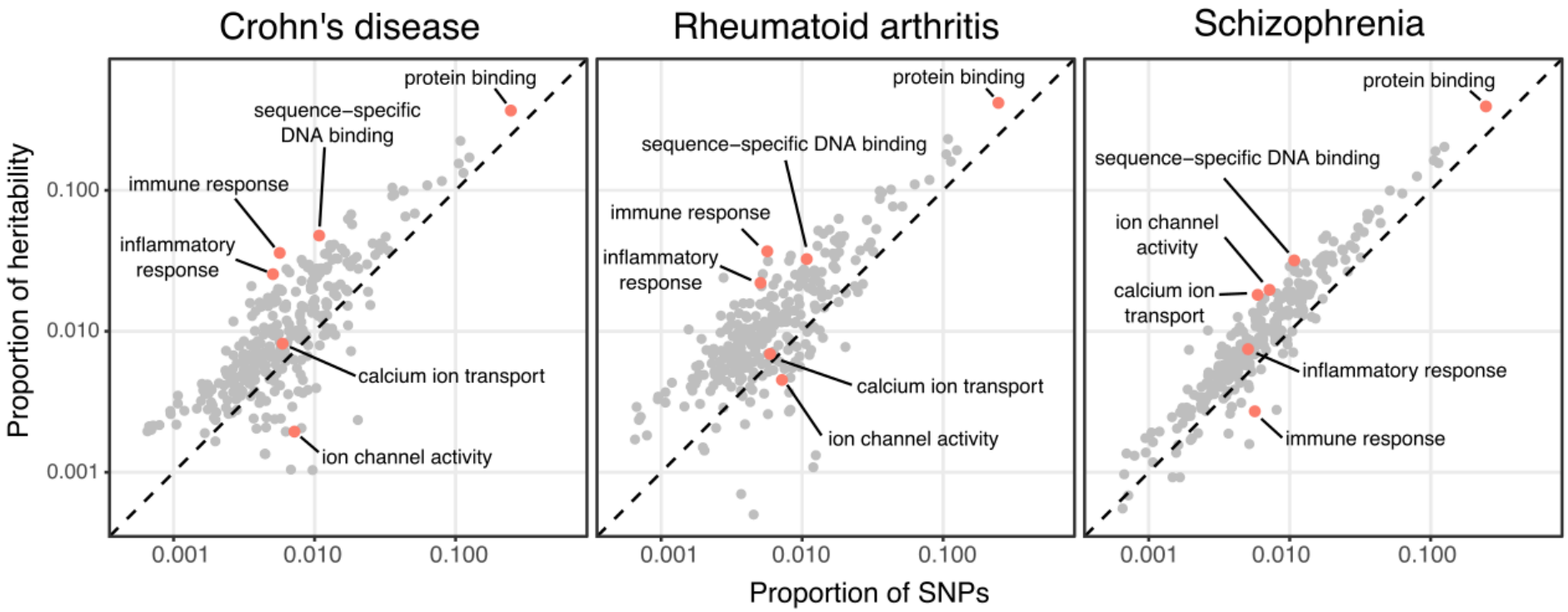


Using stratified LD score regression

- SCZ heritability more enriched near genes preferentially expressed in frontal cortex
- SNPs near genes expressed broadly explained more total heritability (because they are more numerous)

- *Hypothesis: Genes that do not have a direct function in disease pathways might play, in aggregate, a larger role in disease so long as they have a function in the relevant tissue*

Observation #3: SNPs near genes with relevant functional ontologies explain only a small fraction of disease heritability



- Relevant functional categories are enriched in heritability.
- For all three diseases the category that explained the most heritability was simply the largest category, i.e. “protein binding”.

Evan Boyle

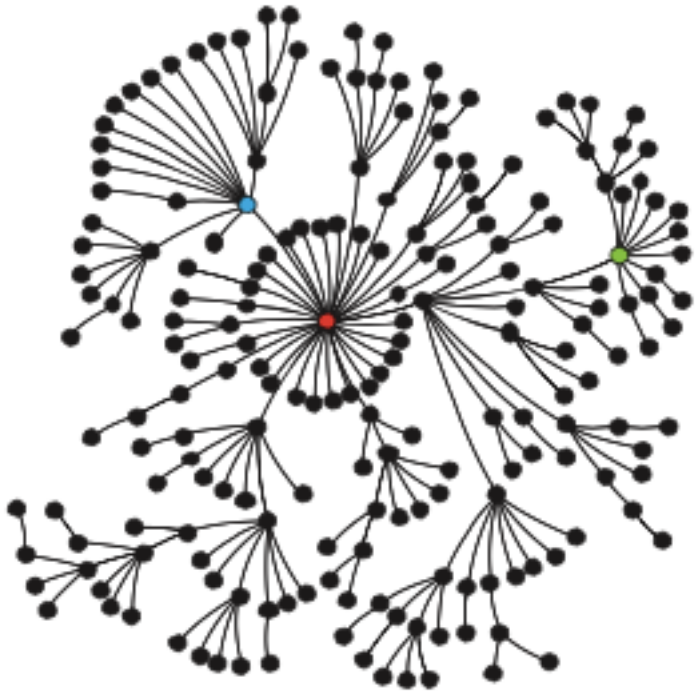
Summary: Variants in most of the genome contributes to heritability

The main shared feature of contributing variants is that they **are in regions that are active in relevant tissues**, but **not necessarily in pathways directly relevant to disease**.

Model: 3 types of genes

- **Core genes**: direct roles in disease (genes that make sense: eg synaptic genes for schizophrenia)
- **Peripheral genes**: Any gene expressed in the “right” cell types can affect regulation/function of core genes, and most of them do
- **Inactive genes**: Genes not expressed in the “right” cell types do not contribute to heritability

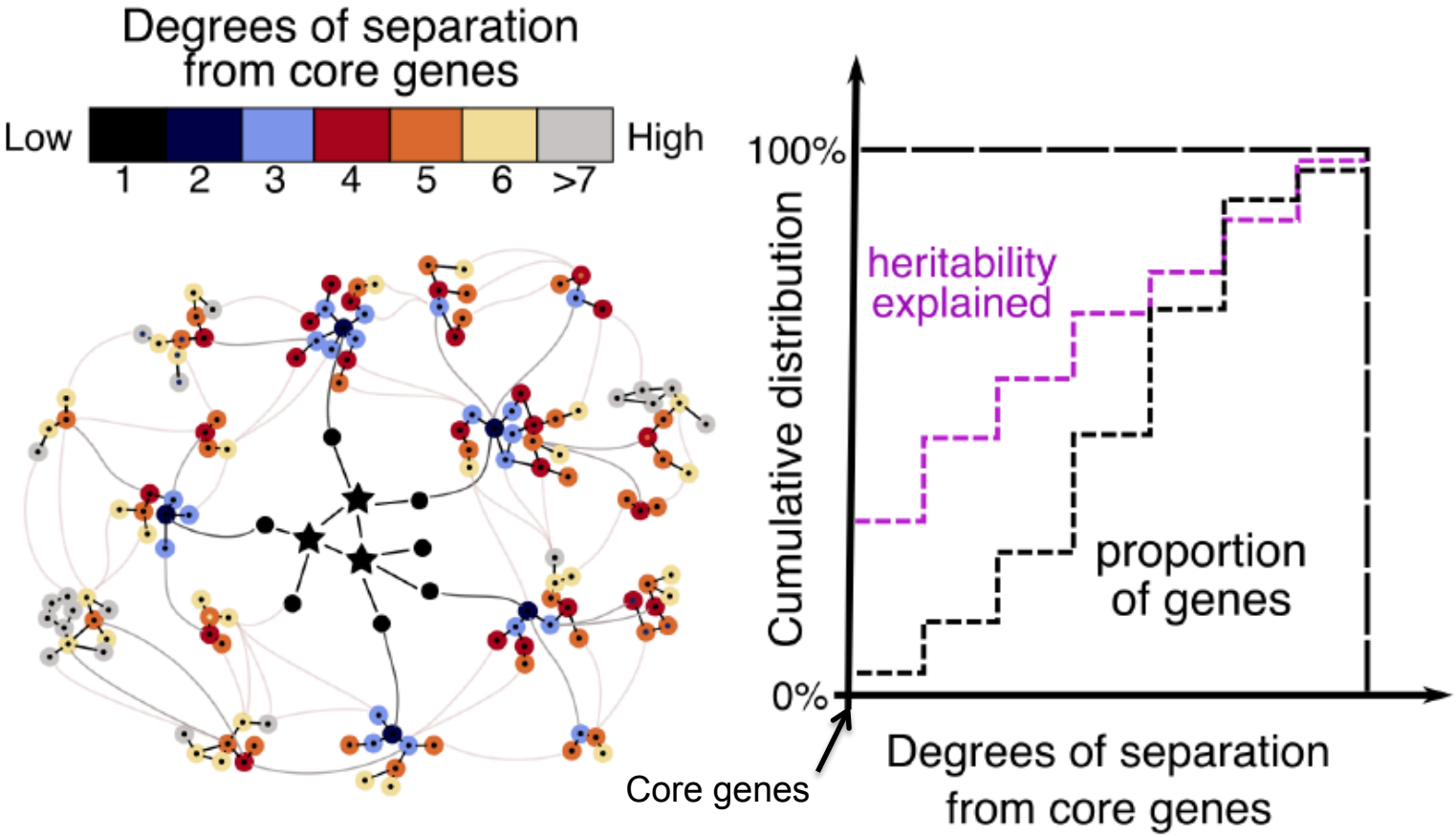
Small world property of networks: most nodes can be reached from every other node by a small number of steps



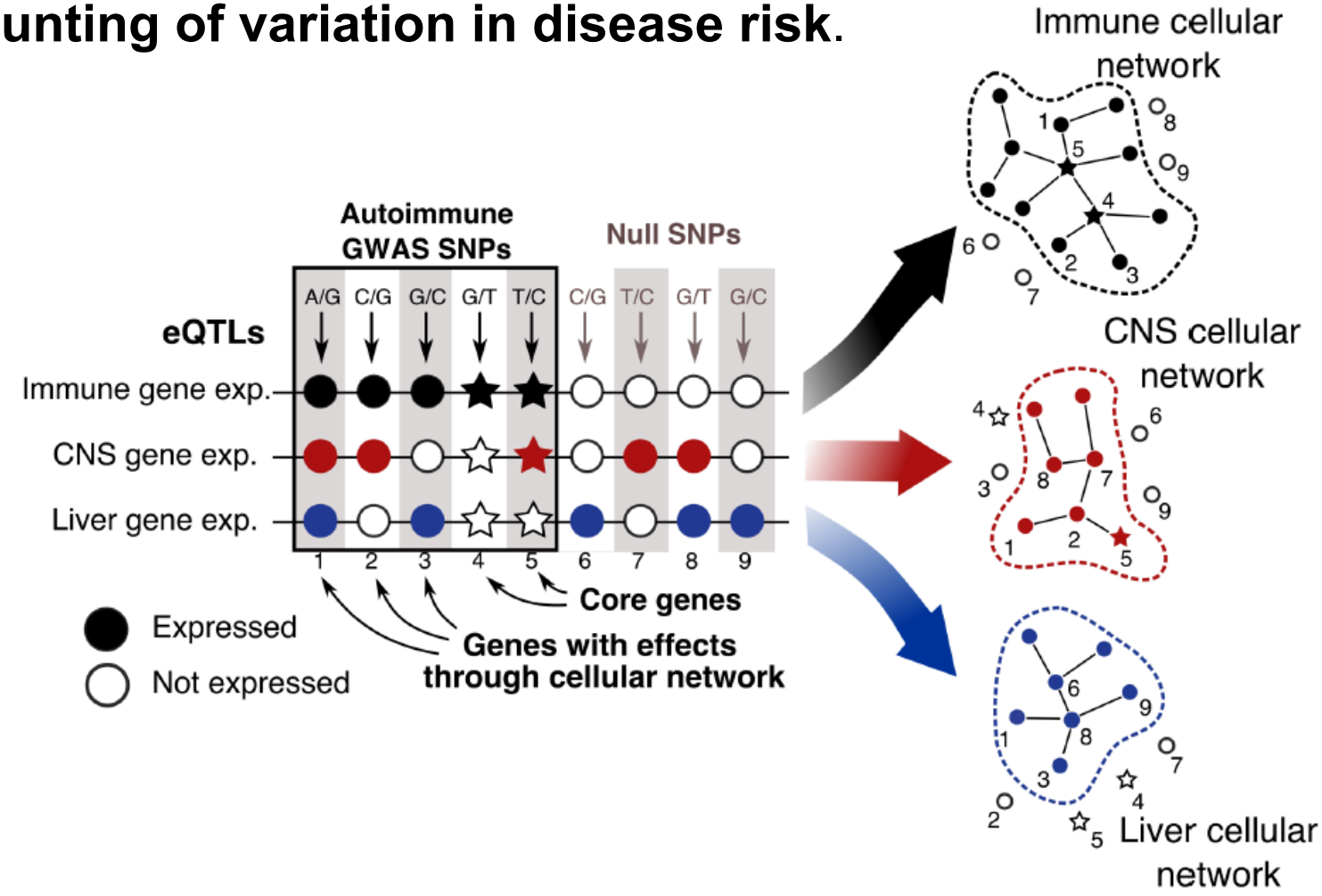
Papers on “small world” property: applies to most real-life networks
Watts & Strogatz 1998
Strogatz 2001

- Suggests that most expressed genes may be “close” to core genes.
- Network could be at the transcriptional, post-transcriptional level, and/or an emergent property (e.g. cell function).

If the *Small World* property applies here then genes outside core pathways might cumulatively contribute more heritability than the much smaller number of genes inside core pathways



Summary: The contributions of thousands (tens of) of regulatory QTLs in peripheral genes might (paradoxically) drive most of the disease heritability. **Therefore focusing only on core genes/pathways might never provide us with a full accounting of variation in disease risk**.



Final remarks

More implications/future directions:

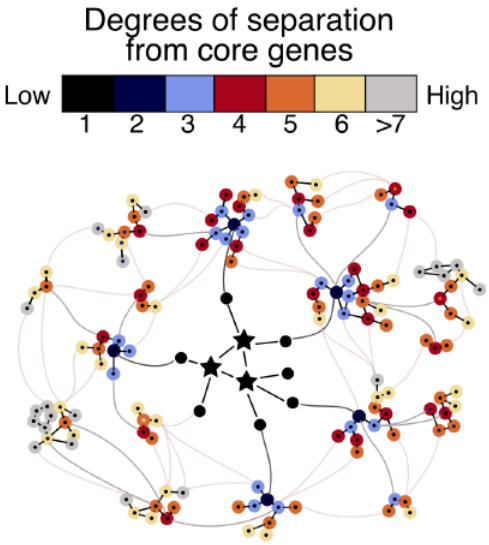
- Mouse models for complex traits have limited use (dysregulated pathway only account for a small fraction of disease cases)
- Drugs may need to focus on altering system states rather than individual genes/pathways.
- **We need a better understanding of how genes/ proteins interact together in a cell-type-specific fashion (could be PPI, transcriptional, etc..)**

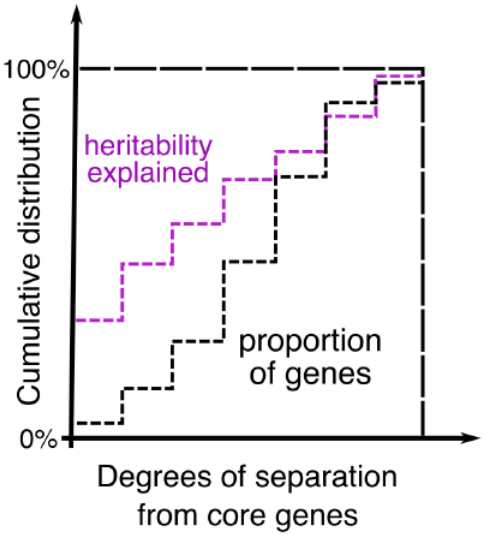
More implications in our perspective:
Boyle*, Li*, and Pritchard* (under review)

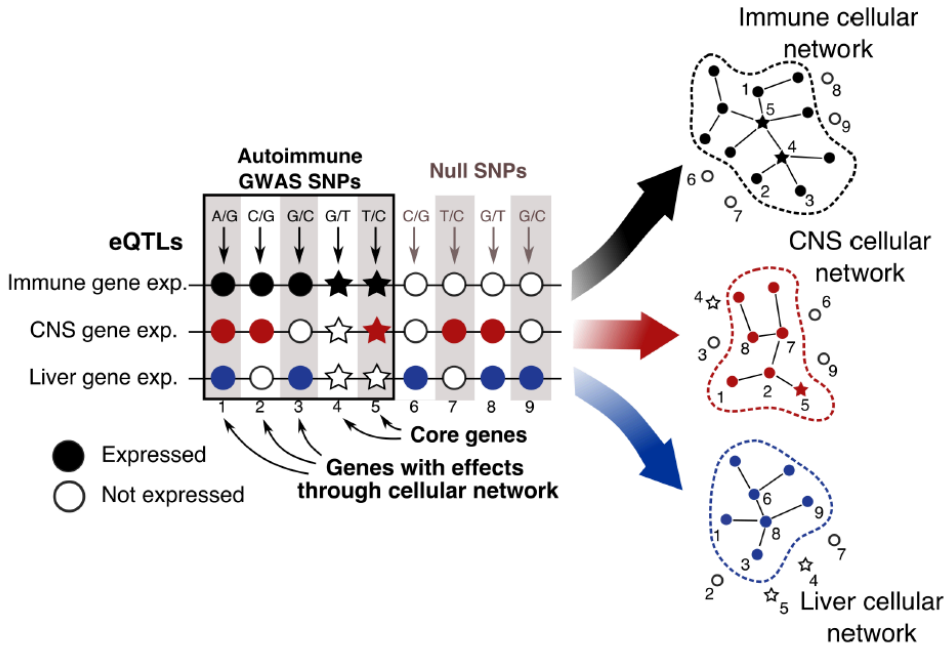
Acknowledgments

Arbel Harpak (Stanford)
William Greenleaf (Stanford)
Alex Marson (UCSF)
Joe Pickrell (NYGC/Columbia)
Guy Sella (Columbia)
Molly Przeworski (Columbia)

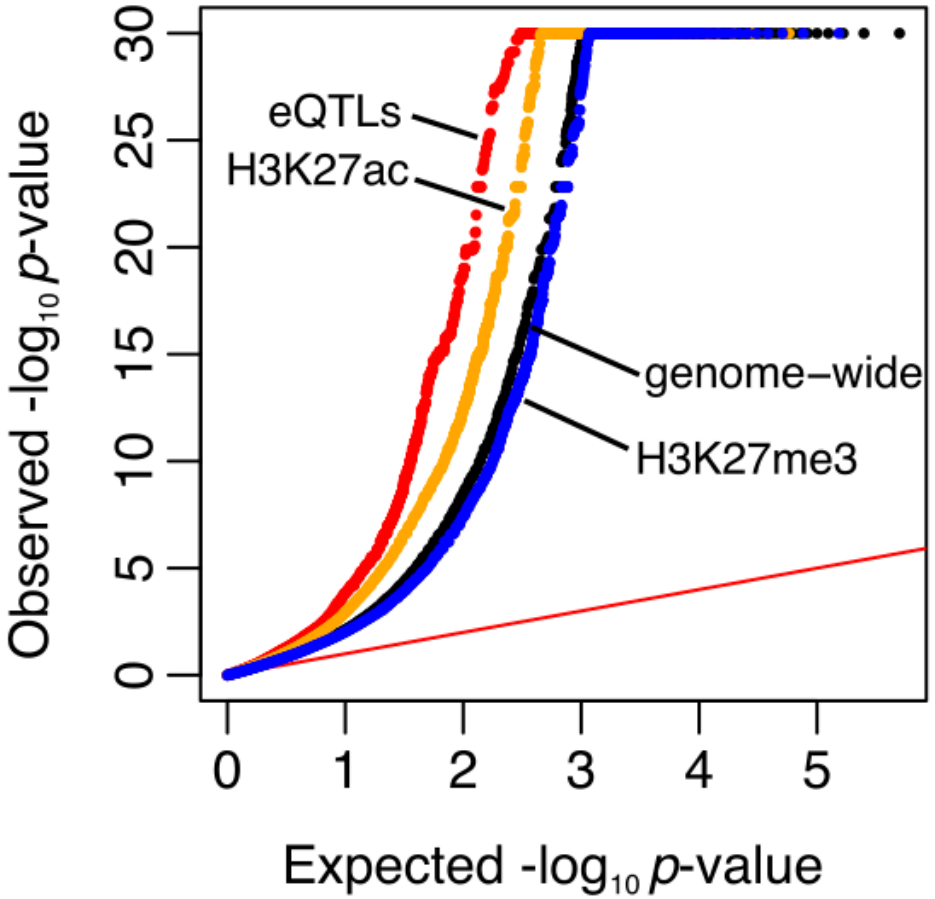
Discussion/criticism welcome!





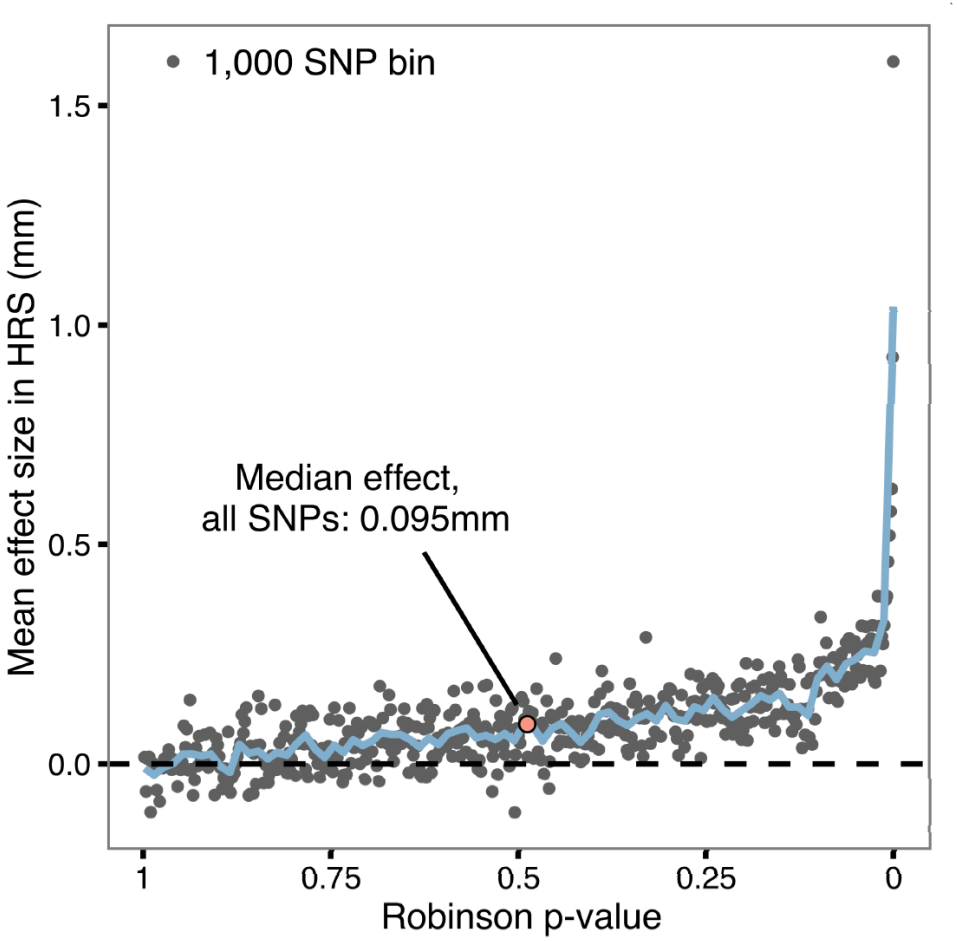


Height GWAS SNPs are enriched in functional elements



- Although GWAS SNPs appear to be spread uniformly across the genome, they often fall within functional regions that suggests regulatory function.

Family-based GWAS confirms the signals are not driven by confounding from population structure



- Median effect smaller likely due to higher rate of sign errors (sample size is 10% of GIANT).

Evan Boyle

Variant type	Gene Set/Ontology	Enrichment p-value	Reference
Rare	ARC voltage-gated calcium channel	$p = 1.6 \times 10^{-3}$ $p = 1.9 \times 10^{-3}$	Purcell et al. (2014)
de novo	ARC N-methyl-D-aspartate receptor (NMDAR)	$p = 4.8 \times 10^{-4}$ $p = 2.5 \times 10^{-2}$	Fromer et al. (2014)
CNV	ARC Synaptic gene	$p = 1.8 \times 10^{-4}$ $p = 2.8 \times 10^{-11}$	PGC (2016)
GWAS	glutamatergic neurotransmission synaptic plasticity	not significant*	Ripke et al. (2014)

Table 1: *Summary of gene sets that show functional enrichment in recent large-scale papers on schizophrenia. Rare variant studies show clearer evidence of enrichment than seen in GWAS. All p-values are nominal, but with corrected* $p < 0.05$. *ARC: activity-regulated cytoskeleton-associated scaffold protein nominally significant prior to multiple testing correction. *Consistent with studies of rare variants, Ripke et al. (2014) identified associated loci near several genes involved in glutamatergic neurotransmission and synaptic plasticity, but these categories did not show a statistically significant enrichment for GWAS hits.*