Journal club Boyle *et al*, Cell 2017

Nathalie Vialaneix, INRA/MIAT

Biopuces – December 12, 2019

Purpose of the talk

Leading Edge
Perspective

Cell

An Expanded View of Complex Traits: From Polygenic to Omnigenic

Evan A. Boyle,[1,*] Yang I. Li,[1,*] and Jonathan K. Pritchard[1,2,3,*]
[1]Department of Genetics
[2]Department of Biology
[3]Howard Hughes Medical Institute
Stanford University, Stanford, CA 94305, USA
*Correspondence: eaboyle@stanford.edu (E.A.B.), yangili@stanford.edu (Y.I.L.), pritch@stanford.edu (J.K.P.)
http://dx.doi.org/10.1016/j.cell.2017.05.038

A central goal of genetics is to understand the links between genetic variation and disease. Intuitively, one might expect disease-causing variants to cluster into key pathways that drive disease etiology. But for complex traits, association signals tend to be spread across most of the genome—including near many genes without an obvious connection to disease. We propose that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most heritability can be explained by effects on genes outside core pathways. We refer to this hypothesis as an "omnigenic" model.

The longest-standing question in genetics is to understand how genetic variation contributes to phenotypic variation. In the early

typical traits, even the most important loci in the genome have small effect sizes and that, together, the significant hits only

Main messages

Usually/initially thought: complex traits driven by a few variants with moderate effects

But:

- effects are weak and causal variants might be rare
- variants are mainly **non coding**

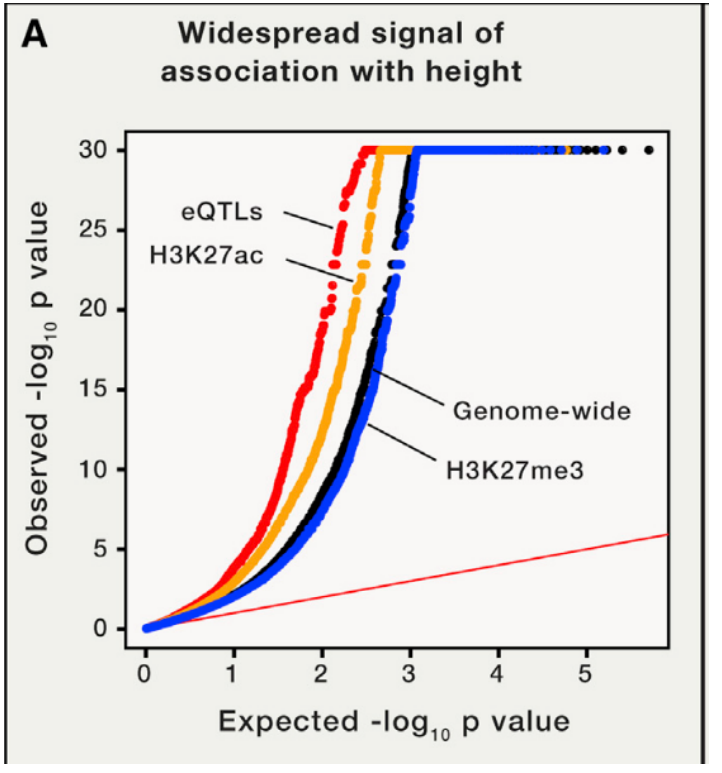
The paper explains that:

- heritability is spread all over genome (nearly all genes)
- complex traits are explained by an accumulation of weak effects on key genes and regulatory pathways
- propogation of small effects through networks (transcriptional, post-transcriptional, PPI, ...)

Careful study of variants associated to height (GIANT)

Distribution of GWAS signals accross genome I/III

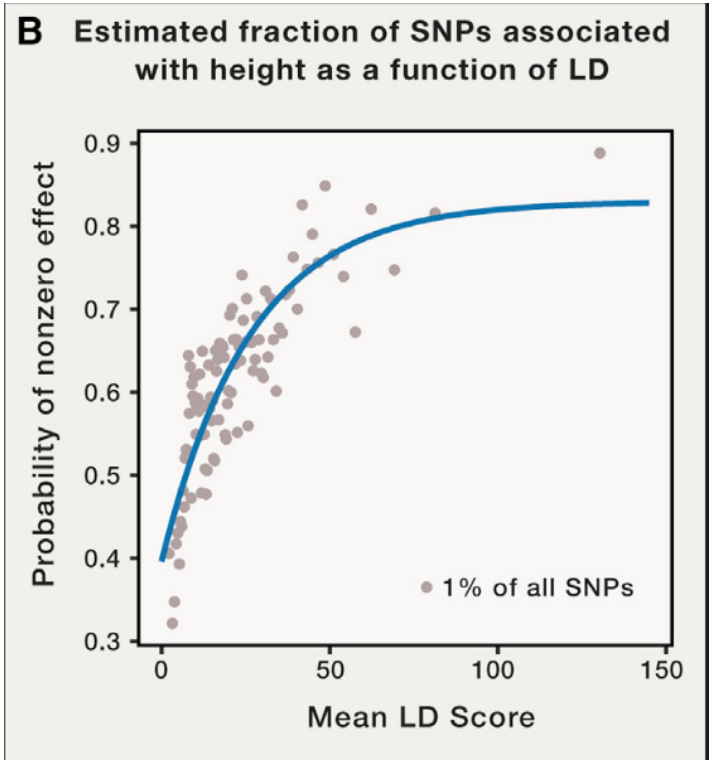
- use of the GIANT study
- 697 significant loci explaining 16% of the height variance
- observed p-values in this study vs expected (under the null hypothesis)



$\Rightarrow$ p-values are **smaller than expected**, especially for eQTL and in active chromatin (enrichment of signal in gene-regulatory regions)

Distribution of GWAS signals accross genome II/III

- for the 697 significant loci, check the % of loci with a non zero effect as a function of the LD

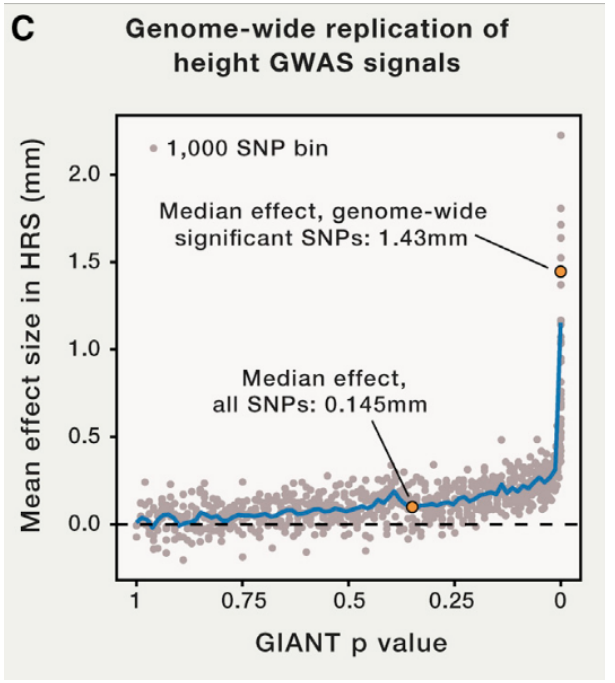


$\Rightarrow$ most 100kb windows in the genome include variants with a non zero effect and SNPs with more LD partners are more likely be associated with height

Overall 62% of the common SNPs have a non zero effect.

Distribution of GWAS signals accross genome III/III

- using a replication independant cohort, computation of SNP effect in the cohort
- results displayed as a function of the p-value in GIANT



$\Rightarrow$ distribution is not centered around zero, even for extremely large p-values (*e.g.*, 0.5) which indicates that observed effect size is a **lower bound** of true effect size

More than 100,000 SNPs have causal effect on height

Main conclusions

- extremely large number of causal variants
- … with tiny effect size
- most genome contributes to height variance

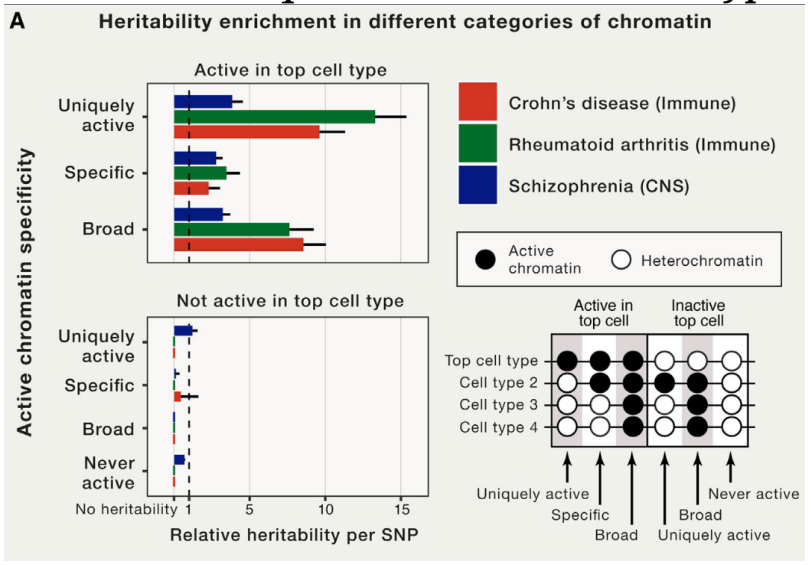
$\Rightarrow$ these conclusions are inconsistent with the assumption that complex trait variants are specific relevant genes and pathway (hence GO analyses of causal variants is maybe not relevant)

What about tissue specific and GO enrichment (complex diseases)?

Are the previous conclusions in contradiction with other analyses? I/III (ATACseq)

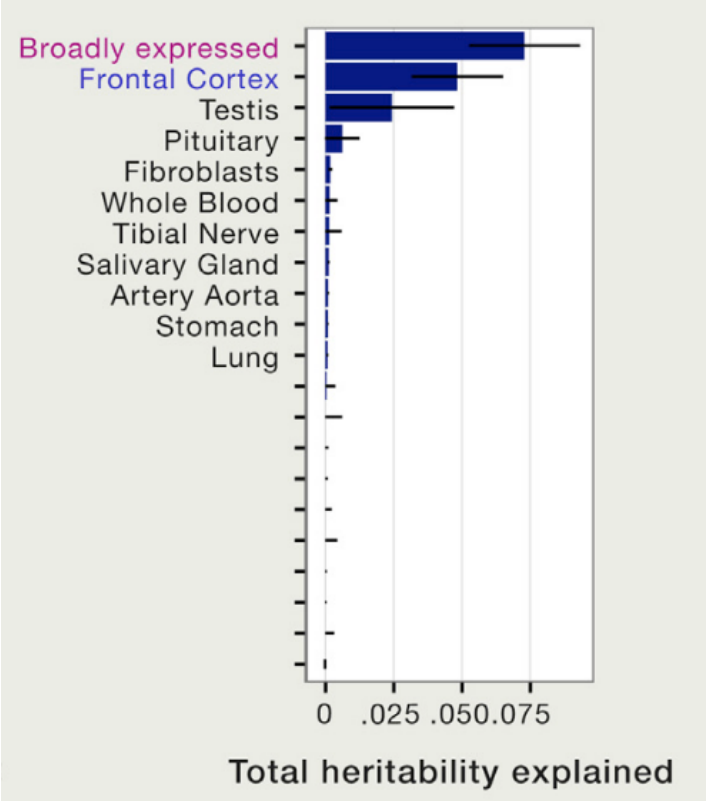
Starting point: most studies in complex diseases (Crohn, rheumatoid arthritis and schizophrenia) show an enrichment in chromatine active in the cell type relevant to the disease (immune system and central nervous system).

Use of ATACseq data on different cell types to check the specificity:



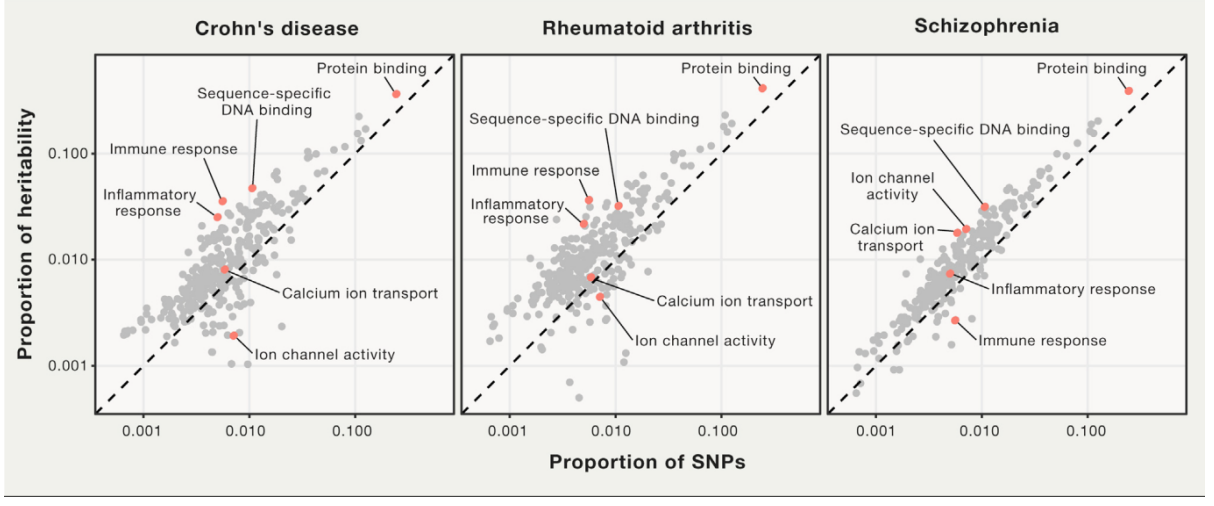
⇒ active chromatine shows an enriched heritability larger than specifically active chromatine & not active chromatine and chromatine only active in irrelevant cell types contribute very lowly to heritability

Are the previous conclusions in contradiction with other analyses? II/III (gene expression)



- SNP near genes that are broadly expressed contribute more to heritability than SNP near genes that are specifically expressed in brain
- part of this result comes from the fact that the number of genes specifically expressed in brain is very low

Are the previous conclusions in contradiction with other analyses? III/III (GO)



- linear relation between number of SNP implicated in a given function and explained proportion of heritability
- broad categories (protein binding) explain more heritability than specific ones
- the only exception holds for studies on rare variants

Main conclusions

- genetic contributions to disease is concentrated in active regions
- enrichment for regions specifically active in relevant tissues is very low
- enrichment is mainly a function of the number of SNP in a given category

$\Rightarrow$ these conclusions are inconsistent with the assumption that complex trait variants are specific relevant genes and pathway (hence GO analyses of causal variants is maybe not relevant)

A new proposition: omnigenic model

Omnigenic model

- traits are directly affected by a few **core** genes/pathways
- nearly all genes affect core genes through networks (effects of individual genes are weighted by these networks)
- the relative effect sizes are such that, since core genes are hugely outnumbered by peripheral genes, a large fraction of the total genetic contributions to disease comes from peripheral genes that do not play direct roles in disease