Genetics and Genomics

Is competition for cellular resources a driver of complex trait heritability?

Olivier Naret, Yuval Simons, Jacques Fellay, Jonathan K Pritchard

School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland • Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland • Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA • Department of Genetics, Stanford University, Stanford, CA 94305, USA • Precision Medicine Unit, Biomedical Data Sciences Center, Lausanne University Hospital and University of Lausanne, Lausanne, 1015, Switzerland • Department of Biology, Stanford University, Stanford, CA 94305, USA

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2
March 14, 2024 (this version)

Reviewed preprint version 1
August 3, 2023

Sent for peer review
May 19, 2023

Posted to preprint server
March 28, 2022

Abstract

Most human complex traits are enormously polygenic, with thousands of contributing variants with small effects, spread across much of the genome. These observations raise questions about why so many variants–and so many genes–impact any given phenotype. Here we consider a possible model in which variant effects are due to competition among genes for pools of shared intracellular resources such as RNA polymerases. To this end, we describe a simple theoretical model of resource competition for polymerases during transcription. We show that as long as a gene uses only a small fraction of the overall supply of polymerases, competition with other genes for this supply will only have a negligible effect on variation in the gene's expression. In particular, although resource competition increases the proportion of heritability explained by trans-eQTLs, this effect is far too small to account for the roughly 70% of expression heritability thought to be due to trans-regulation. Similarly, we find that competition will only have an appreciable effect on complex traits under very limited conditions: that core genes collectively use a large fraction of the cellular pool of polymerases and their overall expression level is strongly correlated (or anti-correlated) with trait values. Our qualitative results should hold for a wide family of models relating to cellular resource limitations. We conclude that, for most traits, resource competition is not a major source of complex trait heritability.

eLife assessment

This **solid** study addresses the unresolved question of why many thousands of small-effect loci contribute more to the heritability of a trait than the large-effect lead variants. The authors explore resource competition within the transcriptional machinery as one possible explanation with a simple theoretical model, concluding that the effects of resource competition would be too small to explain the heritability effects. The topic and approximation of the problem are **important** and offer an intuitive way to think about polygenic variation, but there are concerns on the derivation of the equations with respect to dropping vs. including certain terms that deal inherently with small numbers.

1 Introduction

Since the advent of genome-wide association studies some 15 years ago, there has been huge progress toward determining the genetic basis of many human complex traits (Klein et al. 2005; Wellcome Trust Case Control Consortium 2007; Claussnitzer et al. 2020). However, early studies found something perplexing: namely that the lead variants for any given trait typically explained only a small fraction of the heritability that had been predicted by family studies (Weedon et al. 2008; Manolio et al. 2009). This gap between the heritability accounted for by top variants, and the heritability observed in family studies, caused so much consternation that in 2008 it was referred to as the "mystery of missing heritability" (Maher 2008).

This mystery was largely resolved when it was shown that most of the trait heritability comes from large numbers of common variants with very small effect sizes, whose signals fall far below genome-wide significance (International Schizophrenia Consortium, Purcell, Wray, Stone, Visscher, O'Donovan, Sullivan, and Sklar 2009; Yang et al. 2010). Further work since then has shown that for many complex traits, there are on the order of 10^4 or even 10^5 variants across the genome that affect trait variance (Zhang, Qi, Park, and Chatterjee 2018; Frei et al. 2019; O'Connor, Schoech, Hormozdiari, Gazal, Patterson, and Price 2019; Sinnott-Armstrong, Naqvi, Rivas, and Jonathan K Pritchard 2021). Although there is some contribution from coding variants, most of the heritability comes from non-coding variants impacting gene regulation (Trynka, Sandor, Han, Xu, Stranger, X. S. Liu, and Raychaudhuri 2012; Pickrell 2014; Finucane et al. 2015). These variants are spread surprisingly uniformly across the genome rather than being strongly concentrated near important genes or in particular chromosomal regions (Loh et al. 2015; Shi, Kichaev, and Pasaniuc 2016). Indeed the overall genetic architecture of most complex traits bears a striking resemblance to the classic infinitesimal model of quantitative genetics, introduced in Fisher's 1919 paper (Fisher 1919) and characterized further mathematically by Barton et al and as summarized in Turelli's comment(Turelli 2017; Barton, Etheridge, and Véber 2017).

However, such analyses also indicate another curious feature of the data. While the strongest GWAS signals are usually enriched near trait-relevant genes, in most cases these trait-relevant genes contribute only a small fraction of the heritability (Jostins et al. 2012; Wood et al. 2014; Boyle, Li, and Jonathan K. Pritchard 2017; Zhu and Stephens 2018; Fernández-Tajes, Gaulton, Bunt, Torres, Thurner, Mahajan, Gloyn, Lage, and McCarthy 2019). The observation that the SNPs contributing heritability are spread relatively uniformly across the genome (Loh et al. 2015) implies that a large fraction of genes must be contributing to the trait variance (Boyle, Li, and Jonathan K. Pritchard 2017). For example, a recent paper from our group examined GWAS data for three molecular traits– urate, IGF-1, and testosterone–where a great deal is known about the relevant biological pathways (Sinnott-Armstrong, Naqvi, Rivas, and Jonathan K Pritchard 2021). Aside from one major effect locus for urate, that paper concluded that in aggregate the lead biological pathways for each trait only explain about 10% of the total SNP-based heritability. Instead, for all three traits, the bulk of the heritability comes from a large number of SNPs spread relatively uniformly across the genome: we estimated around 4, 000-12, 000 causal variants for the three molecular traits and 80, 000 causal variants for height. Hence, paradoxically, for typical traits, most of the heritability appears to act mainly through seemingly trait-irrelevant genes.

1.1 Why do so many genes affect trait variance?

Thus, the resolution of the missing-heritability question leads to a second, and more mechanistic question: *Why are complex traits so enormously polygenic, and why do so many different genes affect trait variance?*

In two recent papers, our group proposed a simple quantitative model that we referred to as the "omnigenic" model, to explain this (Boyle, Li, and Jonathan K. Pritchard 2017; X. Liu, Li, and Jonathan K. Pritchard 2019). Summarized very briefly, this model proposes that a modest fraction of all genes have direct effects on a phenotype of interest; these are referred to as "core genes". Meanwhile, all of the other genes expressed in trait-relevant cell types are referred to as "peripheral genes". While the peripheral genes do not exert direct effects on the trait, by definition, the expression levels of peripheral genes can have indirect effects on the trait via gene regulatory networks. Indeed, we proposed that the large majority of the heritability actually flows through indirect trans-regulatory effects from peripheral genes.

This model is currently difficult to test directly due to our limited knowledge of gene networks and core genes. However, our analysis of molecular traits strongly supports the conclusion that core genes typically contribute only small fractions of the heritability (Sinnott-Armstrong, Naqvi, Rivas, and Jonathan K Pritchard 2021). Recent work on correlations between polygenic scores for various traits and whole blood gene expression of likely core genes also supports our model for how trait variation is mediated by a small subset of core genes (Võsa, Claringbould, Westra, Bonder, Deelen, Zeng, Kirsten, Saha, Kreuzhuber, Kasela, et al. 2018).

Furthermore, we showed that there is a natural connection between our model and estimates of cis- and trans-heritability of gene expression (X. Liu, Li, and Jonathan K. Pritchard 2019). Surveying work that measures gene expression heritability in a variety of cell types and species, we estimated that typically $\sim 70\%$ of gene expression heritability is due to trans regulation (X. Liu, Li, and Jonathan K. Pritchard 2019). Since trans-eQTLs have very small effect sizes compared to cis-eQTLs, this implies that a typical gene must be regulated by very large numbers of trans-eQTLs–most of which lie far below the detection threshold for current studies. Based on the 70% estimate for trans heritability of expression, our model implies that peripheral genes can be expected to contribute between $\sim 70\%$ to nearly 100% of the heritability for any given trait, depending on the number of core genes and their relative positions within the regulatory network.

It's worth noting that other types of effects also contribute to the observed architectures of complex traits but do not resolve the paradox of extreme polygenicity, and will not be considered in detail in this paper. First, many disease endpoints are impacted by multiple separate intermediate processes, each of which is, itself, polygenic. For example, diabetes risk is affected by adiposity, lipid levels and distribution, and liver function, each of which has a polygenic basis (Udler 2019). Thus, any variants that affect the intermediate processes can potentially be detected in GWAS of the endpoint trait (Turkheimer 2000; Gottesman and Gould 2003; Pickrell, Berisa, J. Z. Liu, Ségurel, Tung, and Hinds 2016; Udler 2019). While this hierarchical nature of traits certainly contributes to the high polygenicity of some disease endpoints, it seems unlikely to be a complete and general explanation given that virtually all complex traits show high polygenicity. To give just one example, urate, which is controlled mainly by solute channels in the kidneys was estimated to have $\sim 12,000$ causal variants (Sinnott-Armstrong, Naqvi, Rivas, and Jonathan K Pritchard 2021). A second relevant effect is that selective constraint can play a "flattening" role on signals by lowering the allele frequencies of the large-effect variants (O'Connor, Schoech, Hormozdiari, Gazal, Patterson, and Price 2019). This phenomenon likely helps to explain the typically modest contributions of core genes. But at the same time we lack a mechanistic explanation for how it is that so many genes can have nonzero effects. Alternatively, non-biochemical mechanisms can also be a hidden effective trans-acting factor on gene expression. For example, a recent study using computer simulations and a polymer model of chromosomes showed that spatial correlations arising from 3D genome organization lead to stochastic and bursty transcription as well as complex small-world regulatory networks.(Brackley, Gilbert, Michieletto, Papantonis, Pereira, Cook, and Marenduzzo 2021).

1.2 The role of resource competition in trans-regulation and heritability

In this paper, we consider the role of a mechanism for trans regulation that is distinct from the network-based model that we considered previously (Boyle, Li, and Jonathan K. Pritchard 2017; X. Liu, Li, and Jonathan K. Pritchard 2019). In the original phrasing of our model, we assumed implicitly that the effects from peripheral genes are transmitted via specific regulatory interactions in cellular regulatory networks. Examples of specific regulatory interactions include repressors and transcription factors regulating their target genes and protein, but any type of molecular interaction between genes that affects their expression would fit within that framework (Võsa, Claringbould, Westra, Bonder, Deelen, Zeng, Kirsten, Saha, Kreuzhuber, Yazar, et al. 2021; Freimer, Shaked, Naqvi, Sinnott-Armstrong, Kathiria, Chen, Cortez, Greenleaf, Jonathan K. Pritchard, and Marson 2021). In the present paper, we consider a non-specific form of regulation: resource competition.

The foundational premise of the resource competition model is that each cell possesses limited pools of shared molecules crucial for gene expression and regulation. These molecules include RNA Polymerase II, nucleotides, spliceosomes, tRNAs, and ribosomes (Brendler, Godefroy-Colburn, Yu, and Thach 1981; Godefroy-Colburn and Thach 1981; Chu, Barnes, and Haar 2011; Walden, Godefroy-Colburn, and Thach 1981; Brendler, Godefroy-Colburn, Carlill, and Thach 1981; De Vos, Bruggeman, Westerhoff, and Bakker 2011). Recent work underscores competition as the mechanism ensuring that transcription remains consistent irrespective of the genome copy number throughout the cell cycle(Lin and Amir 2018). If an individual carries the high-expressing genotype then we can expect this to very slightly reduce the number of RNA polymerases and other shared resources available to all other genes. Hence, the existence of resource limitations implies that every cis-eQTL must act as a weak trans-eQTL for every other gene.

We should clearly expect the net effect of any single cis-eQTL to be tiny, but what about in aggregate? We know that a large fraction of genes have cis-eQTLs (GTEx Consortium et al. 2018). If there are 10^4 eQTLs in a cell type of interest, then **could these in aggregate drive a meaningful effect on the variance of any given gene, or on the heritability of a trait controlled by that cell type through resource competition?**

In this paper we use a mathematical model to show that the aggregate effects of resource competition are likely to be negligible in practice. Instead, it is more likely that most trans-acting effects on heritability flow through specific molecular interactions in gene regulatory networks.

2 A model for intracellular resource competition

We study these questions using a simple linear model of resource competition in a scenario of complete resource limitation: i.e., where there is a fixed resource pool, and all resources are in full use at all times. When resource competition is partial, when the limited resource is not in full use all of the time, resource competition would have an even smaller effect. To make the model specific, we describe it in terms of competition for RNA Pol II, but competition for other types of resources would be modeled very similarly.

We first examine the effect of the resource competition on the variation in expression level of a single gene and show that it can only account for a tiny fraction of the trans-regulation of gene expression. We then apply this model to a complex trait under the core gene (omnigenic) model of Liu *et al.* (X. Liu, Li, and Jonathan K. Pritchard 2019). We show that, under most plausible conditions, resource competition has only a negligible effect on the proportion of trans-heritability and on the overall trait variance.

2.1 A basic model of competition for polymerases

To focus on the parameters that are directly relevant here, we use a simple model that relates the expression level of a particular gene to the proportion of polymerase bound to and transcribing that gene. Thus, we absorb all the complexities of transcriptional regulation into a single parameter per gene.

We treat the binding of polymerases to promoters as multi-substrate Michaelis-Menten kinetics (Michaelis, Menten, et al. 1913; Schäuble, Stavrum, Puntervoll, Schuster, and Heiland 2013). To emphasize the connection with Michaelis-Menten, we follow the standard notation from chemistry, denoting the number of Pol II molecules bound to gene *i* (either in the promoter or gene body) per cell as the *concentration* of bound polymerase $[PG_i]$.

As we show in the supplement, the concentration of polymerase bound to gene *i* is

$$[PG_i] = g_i \cdot [G] \cdot [P], \tag{1}$$

where [*G*] is the concentration of promoters (assumed constant, i.e., two copies per cell for each gene), [*P*] is the concentration of free polymerase, and g_i is the reciprocal of the Michaelis constant. Note that g_i measures gene *i*'s ability to bind free polymerase and can increase if the polymerase affinity to the gene's promoter increases or the rate of transcription initiation increases. In the genetic context, variants that affect gene expression in cis (i.e., cis eQTLs), would do so by changing g_i.

We denote the overall concentration of polymerase as $[P]_0$, and assume that $[P]_0$ is fixed. Then $[P]_0 = [P] + \sum_{i=1}^{m}[PG_i]$ i.e., the overall concentration of polymerase is the sum of the free polymerase, [*P*], and the polymerase bound to all genes. We are interested in the limit of strong resource competition so we assume there is very little free polymerase, i.e. [*P*] << $[P]_0$, since this is the limit in which resource competition is strongest and would have the largest effect. Therefore,

$$[P]_0 = [P] + \sum_{i=1}^{m}[PG_i] \approx \sum_{i=1}^{m}[PG_i] = \sum_{i=1}^{m} g_i \cdot [G] \cdot [P] = g_{tot} \cdot [G] \cdot [P] \tag{2}$$

with *m* being the number of genes and $g_{tot} \equiv \sum_{i=1}^{m} g_i$ defined as a total free polymerase binding ability. Therefore,

$$[P] \approx \frac{[P]_0}{g_{tot} \cdot [G]} \tag{3}$$

and

$$[PG_i] \approx \frac{g_i}{g_{tot}} \cdot [P]_0. \tag{4}$$

This equation shows that the proportion of time a polymerase is bound to gene *i* is proportional to $\frac{g_i}{g_{tot}}$ or, in other words, a fraction $\frac{g_i}{g_{tot}}$ of the overall polymerase $[P]_0$, is bound to gene *i*.

The rate of transcription of gene *i* is proportional to $[PG_i]$ and if we assume that the over-all polymerase concentration, $[P]_0$, is constant in time and identical between individuals then, at equilibrium, the expression level of gene *i*, x_i, is

$$x_i = \frac{g_i}{g_{tot}} \cdot x_i^{max} \tag{5}$$

with x_i^{max} being the maximal gene expression level of gene *i* (see supplement for full derivation). x_i^{max} is the hypothetical expression level achieved if all the polymerases were actively transcribing gene *i*. Generally, only a small fraction of the polymerase pool is transcribing any single gene, meaning that $x_i << x_i^{max}$. We henceforth measure each gene's expression level in units of x_i^{max}, i.e., we set $x_i^{max} = 1$ for each gene.

Under this model, gene expression of gene *i* is simply proportional to the fraction of the polymerase pool transcribing that gene. Moreover, under this scenario, Equation 5 shows that if the free polymerase binding ability g_i were to increase, it not only increases the number of polymerases transcribing gene *i* but also decreases the number of polymerases transcribing all other genes (by increasing g_{tot}). Crucially, cis-eQTLs do exactly this: they increase (or decrease) g_i, with small opposite-direction effects on all other genes.

Importantly, although this model is motivated by considering transcription, the functional form will hold for any form of extreme competition: gene product levels will be proportional to the gene's share of the limited resource. For example, competition for ribosomes would lead to a similar equation for protein levels with free ribosome binding ability replacing free polymerase binding ability.

Furthermore, although we assume a relatively simple competition model for illustrative purposes, our model's simplicity arises from assuming an extreme form of resource competition with very little free polymerase. Therefore, our key result that resource competition generally has negligible effects should be even stronger for more-complex models with less-stringent resource competition. Moreover, our results stem not from model specifics but from order-of-magnitude arguments and should therefore hold quite generally, as discussed later.

3 Resource competition can only explain a small fraction of trans-regulation

It is estimated that approximately 70% of the heritable variance in gene expression is due to trans-regulation (and the rest from cis-regulation) (X. Liu, Li, and Jonathan K. Pritchard 2019). Resource competition provides a possible mechanism for this, since it implies that every cis-eQTL is a (weak) trans-eQTL for all other genes. Even though for any single QTL this is a small effect, one might conjecture that trans effects could accumulate over all genes and provide a significant fraction of the estimated 70% gene expression trans heritability. However, in this section we will show that, as long as no single gene engages more than a small fraction of the overall pool of polymerases (or another limited resource), resource competition will have only a small effect on variation in gene expression and will account for only a small fraction of trans-regulation.

According to our model (5), the variance in the expression level of gene *i* across individuals is:

$$V[x_i] = V\left[\frac{g_i}{g_{tot}}\right]$$

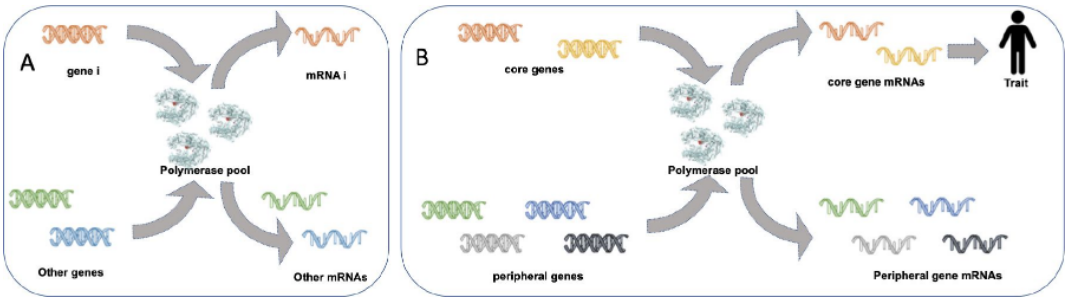

Figure 1.

Illustration of the competition model for polymerases.

(A) Transcription of a single gene may be affected by polymerase depletion due to transcription of other genes. **(B)** This same model extends to complex traits by distinguishing between transcription of core and peripheral genes.

(Note that all expectations and variances henceforth are taken over individuals, i.e. over the variability in g values induced by genetic regulation of polymerase affinity and transcription rate.)

The variance of a ratio $V(a/b)$, when $a \ll b$, can be approximated by a first-order Taylor expansion as follows (Elandt-Johnson and Johnson 1980; Kendall, Stuart, and Ord 1994):

$$V\left[\frac{a}{b}\right] \approx \frac{1}{E[b]^2}\left(V[a] - 2\frac{E[a]}{E[b]}Cov[a,b] + \frac{E[a]^2}{E[b]^2}V[b]\right). \quad (6)$$

Assuming that $g_i \ll g_{tot}$ we can use this approximation to write:

$$V[x_i] \propto \underbrace{V[g_i]}_{basal} \underbrace{- 2\frac{E[g_i]}{E[g_{tot}]}Cov[g_i, g_{tot}] + \frac{E[g_i]^2}{E[g_{tot}]^2}V[g_{tot}]}_{competition}. \quad (7)$$

Without loss of generality, we set the proportionality constant of Equation 7 to be 1, which is equivalent to changing the units in which we measure polymerase affinity.

The first term ($V[g_i]$) on the right side of Equation 7 reflects the various sources of genetic variance that are *not* due to resource competition: the expression variance due to cis genetic effects, environmental variance, and possibly trans effects acting via gene regulatory networks. The second and third terms reflect the effects of resource competition. Notably, the remaining two resource competition terms in this equation represent two opposite effects:

1. $-2\frac{E[g_i]}{E[g_{tot}]}Cov[g_i, g_{tot}]$ This term represents a perhaps unexpected effect. An allele associated with an increase in the free polymerase binding ability of gene i, g_i, also leads to an increase in total free polymerase binding ability, g_{tot}; hence, the impact of any cis-acting change in g_i is slightly counteracted by global depletion of polymerase. Thus, in this term, competition leads to a *reduction* in expression variance as represented by the second term of Equation 7.
2. $\frac{E[g_i]^2}{E[g_{tot}]^2}V[g_{tot}]$ This term represents the intuitive effect that an increase (respectively, decrease) in expression of any gene soaks up some of the free polymerase, thereby slightly reducing (increasing) expression of all other genes. Thus, an increase in $V[g_{tot}]$ will tend to *increase* the variance of gene i, as represented by the third term of Equation 7.

However, critically, the magnitude of both these terms are likely to be small in practice. The two effects of resource competition depend on the ratio of the mean free polymerase binding ability of the gene of interest, $E[g_i]$, and the expected total free polymerase binding ability, $E[g_{tot}]$; i.e. both effects are proportional to the fraction of polymerases bound to the gene of interest. This suggests that, as long as only a small fraction of polymerases are bound to the gene of interest, competition for polymerases would only have a small effect on variation in gene expression.

We show this explicitly using a simple but illustrative example: Assume that for all genes the mean and variance of free polymerase binding ability are identical, i.e., $E[g_i] = E_g$ and $V[g_i] = V_g$ for every i, and that all genes are solely under cis-regulation, i.e. $Cov[g_i, g_j] = 0$ if $i \neq j$. Under these assumptions the overall variance in the expression of gene i is:

$$V[x_i] = \underbrace{V_g}_{basal} \underbrace{- 2\frac{1}{m}V_g}_{cis} + \underbrace{\frac{1}{m^2}mV_g}_{trans}. \quad (8)$$

The first term in this equation is the variation due to cis-regulation of gene *i*. The second term is also a cis term and represents a dampening of cis-regulation due to the limited availability of polymerases. The third term is the trans-regulated variation in gene expression produced by resource competition. In this term, a change in the cis-regulation of any gene changes the proportion of polymerases binding to that gene and thereby mRNA production in the gene of interest.

We see that, in this simple example, resource competition leads to a reduction in the total variance but to an increase in trans regulation. However, this increase in trans regulation is inversely proportional to the number of genes, $m \gg 1$, and is therefore tiny. This tiny effect would clearly fail to explain the roughly 70% of the variation in gene expression that is due to trans-regulation. In fact, with $\sim$10, 000 genes expressed in a typical cell, the impact of resource competition would be 4 orders of magnitude smaller than the observed magnitude of trans regulation.

Relying on this simple order of magnitude argument means that the conclusion that competition has a minute effect on the overall variation and the proportion of trans-regulation is quite general. The effect of resource competition scales like $1/m$ even when genes are not identical and when there are other sources of trans-regulation. This result does not depend on which specific resource is rate limiting for transcription and holds equally for competition during translation, e.g. for ribosomes. The only possible scenarios under which competition is a large effect on gene expression variation are scenarios in which a large fraction of the pool of polymerases (or another limited resource) is bound to the gene of interest. This might plausibly happen for highly transcribed genes, such as rRNA genes; or perhaps in settings like translation that occurs in highly localized subcellular compartments with limited numbers of mRNAs.

We also conducted simulations to validate that the relative effect of resource competition is minor, on the order of $1/m$. Details can be found in the supplementary materials, section 7.2, and are illustrated in **Figures S1A** and **S1B**.

4 Effect of resource competition on the variance of complex traits

Complex traits involve many genes and therefore one might hypothesize that resource competition could have a large effect on complex trait variation even though it has a small effect on the expression of any single gene. In this section, we ask **what is the impact of competition on the phenotypic variance of a complex trait?**

Using the omnigenic model laid out in Liu *et al.* (X. Liu, Li, and Jonathan K. Pritchard 2019), we consider a complex trait whose value is determined by the expression of *c* core genes. We show that there are two conditions for resource limitation to account for a significant fraction of trait variation: First, a large fraction of the pool of polymerases has to be bound to core genes. Second, an increase in the expression of core genes has to systematically increase or decrease the trait.

Following Liu *et al.*, we consider a set of *c* core genes whose expression affects a complex trait, *Y*. We define the effect size per unit gene expression as γ, with γ_i being the effect size for gene *i*. A positive γ_i implies that an increase in the expression of gene *i* increases trait values and a negative γ_i indicates that an increase in the expression of gene *i* decreases trait values. In this model, an individual's phenotype is given by:

$$Y = \bar{Y} + \sum_{i=1}^{c} \gamma_i x_i \tag{9}$$

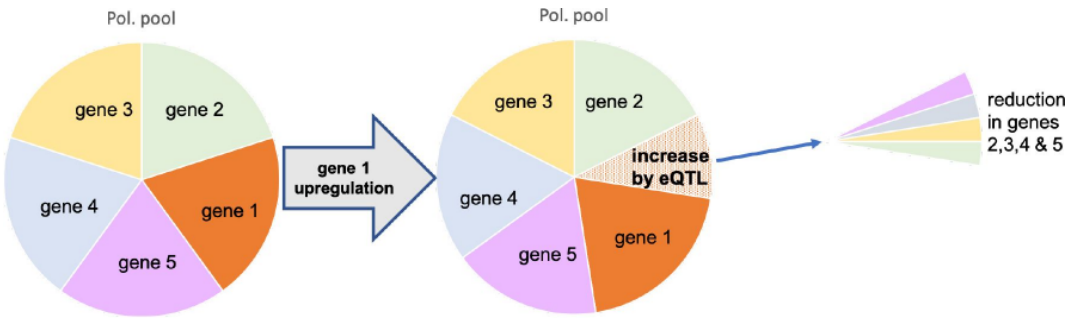

Figure 2.

Illustration of the effect of resource competition on expression.

Upregulation of Gene 1 causes a small downregulation in the expression of all other genes. Since this change is spread between all genes it scales like 1 over the number of genes.

With $\bar{Y}$ being the mean phenotype in the population. We omit, for simplicity, a possible random environmental effect term.

Next, we rewrite Equation 9 in terms of free polymerase binding ability by plugging in Equation 5:

$$\begin{aligned} Y &= \bar{Y} + \sum_{i=1}^{c} \frac{\gamma_i g_i}{g_{tot}} \\ &= \bar{Y} + \frac{g_\gamma}{g_{tot}} \end{aligned} \tag{10}$$

where we define, for convenience, $g_\gamma = \sum_{i=1}^{c} \gamma_i g_i$.

We now use our approximation (Equation 6) to obtain an expression for the phenotypic variance:

$$V[Y] \propto \underbrace{V[g_\gamma]}_{basal} \underbrace{-2\frac{E[g_\gamma]}{E[g_{tot}]} Cov[g_\gamma, g_{tot}] + \frac{E[g_\gamma]^2}{E[g_{tot}]^2} V\left[g_{tot}\right]}_{competition} \tag{11}$$

This expression is very similar in form to Equation 7 except that g_γ replaces g_i. The first term now represents trait variation due to non-competitive effects, including both cis and trans regulation of core genes. The second term now represents the suppression of trait variation due to core genes competing among themselves for the limited pool of polymerases. The third term represents the increase in trait variation due to competition-induced fluctuations in the number of polymerases available for core genes.

As in the previous section (Equation 7), we can see that the effect of resource competition will depend crucially on the ratio between the averages of g_γ and g_{tot}. This ratio is different from the ratio for the variance of a single gene seen in the previous section in two ways: (1) it concerns many core genes and not just one, and (2) each gene is associated with a γ value that can be either positive or negative.

To gain an intuition for the effects of these qualitative differences, we turn, once again, to the very simple model where the expression levels of all genes have identical distributions with variance V_g and resource competition is the only source of trans effects. Assuming that $m >> 1$ and $c >> 1$, equation 11 is then simplified to

$$\begin{aligned} V[Y] &\propto \overline{\gamma^2} c\, V_g - 2\frac{\bar{\gamma} c}{m} \bar{\gamma} c\, V_g + \frac{\bar{\gamma}^2 c^2}{m^2} m\, V_g \\ &= \underbrace{\overline{\gamma^2} c\, V_g}_{basal} - \underbrace{\bar{\gamma}^2 \frac{c}{m} c\, V_g}_{competition} \end{aligned} \tag{12}$$

We immediately see that, in comparison to what we saw in the previous section (Equation 8), there are now two distinct reasons for the competition term to be relatively small. The effect of resource competition will be small relative to the basal term if either of the following conditions holds:

1. $\frac{c}{m} << 1$, i.e. the number of core genes is much smaller than the total number of genes. As with a single gene, this is because only a small fraction of the pool of polymerases is bound to core genes, or:

2. $\overline{\gamma}^2 << \overline{\gamma^2}$, i.e. the mean effect size squared is much smaller than the mean squared effect size. This will generally be true since γ takes both positive and negative values. The only exception would be when trait values are strongly correlated (or strongly anti-correlated) with the overall expression level of core genes such that most γ values have the same sign.

This result, that resource competition is a small effect for these two reasons, should hold quite generally. In particular, it should hold if genes vary in their expression patterns, so long as core genes have comparable expression levels to peripheral genes. However, the direction of the small effect of resource competition is sensitive to such details, since they affect the balance between the two resource competition terms in Equations 7 and 12.

This result is also valid if core genes are co-regulated. As discussed by Liu *et al.* (X. Liu, Li, and Jonathan K. Pritchard 2019), if core genes tend to be co-regulated, trans effects can dominate the variance of a trait, thus inflating $V[g_\gamma]$, which only involves covariances between core genes, considerably compared to $Cov[g_\gamma, g_{tot}]$ and $V[g_{tot}]$. Therefore, in such a case, the relative importance of resource competition will be even smaller.

5 Discussion

In this paper, we explored the possible contribution of resource competition to gene expression and complex trait variance. Since different genes compete for the same cellular resources during transcription and translation, a variant upregulating a single gene may reduce the availability of cellular resources to all other genes. However, it is unclear, *ab initio*, if this could be a substantial effect.

We have presented a simple model of resource competition between genes, at the level of transcription. We have shown that resource competition should only have a minor effect on variation in the expression level of any given gene, as long as a small fraction of the overall pool of polymerases binds to that gene. It can therefore account for only a tiny fraction of the trans-heritability of gene expression. Similarly, only if a large fraction of the overall pool of polymerases binds to the core genes would resource competition have a major effect on trait variation. Even in such a scenario, resource competition would remain a small effect on trait variance unless trait values are strongly correlated (or anti-correlated) with the overall expression level of core genes.

While some traits may meet one of these two conditions, only a few traits should meet both. We do not know much about the expected number of core genes but it could be large for some traits. Even if a trait has a small number of core genes, these genes may bind a large fraction of the pool of polymerases in trait-relevant tissues if the expression is compartmentalized, or during specific periods of development. As for the second condition, we do not know of any category of traits for which trait values correlate with expression levels of core genes, though such traits may exist. Still, we expect that for the vast majority of traits at least one of these conditions is not met, making resource competition a negligible effect.

The work presented here is based on a very simple competition model, but one that assumes very strong competition. A more detailed model may explicitly model regulatory mechanisms, competition for multiple resources, temporal and spatial patterns, gene heterogeneity, or fluctuations. Such a model would result in a similar or weaker form of competition, with gene activity depending only partially and/or stochastically on the limited resource. Therefore, we expect our results to hold, qualitatively, even for more detailed models.

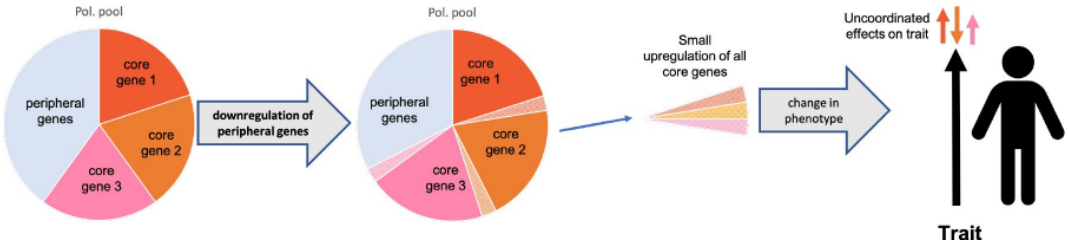

Figure 3.

Illustration of the effect of resource competition on compex trait phenotype.

Downregulation of one or more peripheral genes causes a small upregulation of all core genes. As we show in the text, this will only be an appreciable effect if core genes engage a large fraction of the overall pool of polymerases. However, as can be seen in this illustration, since an increase in core gene expression may both increase and decrease the phenotype, such an increase would only result in a minute phenotypic effect.

Other forms of resource competition are expected to produce near-identical results. Strong competition between genes should give the same equation 5, regardless of what is the limited resource. For example, if we consider competition for transcription factors all that would change is that g_i would parameterize free transcription factor binding. Similarly, if we consider competition at the level of translation instead of transcription, protein level would replace gene expression levels in Equation 5 and g_i could parameterize free ribosome binding (of all of gene *i*'s mRNAs). All such modes of competition, or a combination of them, would result in similar models. However it is worth noting that the conclusions may differ in contexts where a very small number of genes compete for a highly limited resource, such as access to a particular molecular transporter – this would change the relevant number of genes, *m*, such that variation in any single gene could in fact have an appreciable effect.

In summary, we have explored here resource competition as a possible contributing mechanism to expression and complex trait variation. We have laid out a foundation for understanding the effects of resource competition on the architecture of expression, protein, and trait-level variance. Our model suggests that, for most traits, competition will not be a meaningful contributor to phenotypic variance.

Acknowledgements

This work was supported by NIH grants HG011432 and HG008140 to JKP, HG011202 to YS, and by an SNSF Doc.Mobility fellowship to ON. We thank Matthew Aguirre, Hakhamanesh Mostafavi, Roshni Patel, and the entire Pritchard lab, for helpful conversations. The project was prompted in part by helpful conversations with David Botstein, Guy Sella, and Gavin Sherlock. Much of the work was conducted during an extended visit by ON to the Pritchard lab at Stanford in 2019. We would also like to thank the anonymous reviewers for their time and feedback.

7 Supplement: Multi-substrate Michaelis-Menten

7.1 Multi-substrate Michaelis-Menten

We can think of the kinetics of polymerase binding to multiple genes as following multi-substrate Michaelis-Menten kinetics, with the polymerase acting as an enzyme and the different genes as competing substrates. We present here a quick review of such kinetics, following the procedures outlined in Chou and Talalay (Chou and Talaly 1977) and arriving at a similar result to Schauble et al. (Schäuble, Stavrum, Puntervoll, Schuster, and Heiland 2013). We emphasize the assumptions we use and the simplifications arising from them.

We model mRNA production at each gene as an irreversible reaction facilitated by the polymerase, i.e.

$$P + G_i \underset{k_{off,i}}{\overset{k_{on,i}}{\rightleftharpoons}} PG_i \xrightarrow{k_{cat,i}} P + x_i$$

which implies that

$$\frac{d[PG_i]}{dt} = k_{on,i}[P][G_i] - k_{off,i}[PG_i] - k_{cat,i}[PG_i]. \tag{13}$$

and therefore, at equilibrium, when $\frac{d[PG_i]}{dt} = 0$,

$$[PG_i] = \frac{[P][G_i]}{K_{M,i}} \tag{14}$$

with $K_{M,i} = \frac{k_{off,i}+k_{cat,i}}{k_{on,i}}$ being the Michaelis constant.

The overall level of the polymerase, $[P]_0$, is constant (or externally set). Therefore,

$$const = [P]_0 = [P] + \sum_i [PG_i] = [P] + \sum_i \frac{[P][G_i]}{K_{M,i}}. \tag{15}$$

Here comes our first major assumption and simplification - since each gene has exactly two copies per cell then $[G_i]$ is the same for all genes, and we set $[G_i] = [G]$. Therefore,

$$[P]_0 = \left(1 + \sum_i \frac{1}{K_{M,i}}[G]\right)[P] = \left(1 + \sum_i g_i[G]\right)[P]. \tag{16}$$

where, for convenience, we define $g_i \equiv \frac{1}{K_{M,i}}$.

We are interested in scenarios of resource competition, when the proportion of free polymerase is very small. That is, when polymerases spend most of their time bound to and transcribing some gene. Mathematically, this is the limit when $[P] << \sum_i [PG_i]$ or

$$[P]_0 = \left(1 + \sum_i g_i[G]\right)[P] \approx \sum_i g_i \cdot [G] \cdot [P]. \tag{17}$$

From this equation, we arrive at the results that

$$[P] = \frac{[P]_0}{g_{tot} \cdot [G]} \tag{18}$$

with $g_{tot} = \Sigma_i g_i$ and therefore

$$[PG_i] = \frac{g_i}{g_{tot}}[P]_0. \tag{19}$$

The rate of mRNA production of gene i is therefore

$$V_i = k_{cat,i}[PG_i] = V_{max} \cdot \frac{g_i}{g_{tot}}. \tag{20}$$

with $V_{max} = k_{cat,i} \cdot [P]_0$ being the maximum rate of mRNA production. Lastly, the level of mRNA would be

$$x_i = \frac{k_{cat,i}}{k_{decay,i}}[PG_i] = x_i^{max} \cdot \frac{g_i}{g_{tot}} \tag{21}$$

with $x_i^{max} = \frac{k_{cat,i}}{k_{decay,i}} \cdot [P]_0$ being the maximal gene expression level for this concentration of polymerase, $[P]_0$.

7.2 Simulations

We simulated a dataset of 10,000 under our model under two scenarios: First, we simulated m identical genes with polymerase binding activity (g) sampled from a normal distribution with mean μ_g = 1000 and variance 1, and looked at variation in expression of a focal gene, marked gene i. In **figure S1A** you can see that, as predicted by **equation 8**, resource competition reduces the variance, and that this reduction is proportion to relative abundance of gene i. In the second scenario, while for all other genes g had a variance of 1, we varied the variance of g_i, marked as V_{gi}. As you can see in S1B, the change in expression variance is still linear in the relative abundance of gene i (at least for low relative abundance) but the slope depends on V_{gi}. This is because changing V_{gi} changes the relative impact of the two competition terms in **equation 7**.

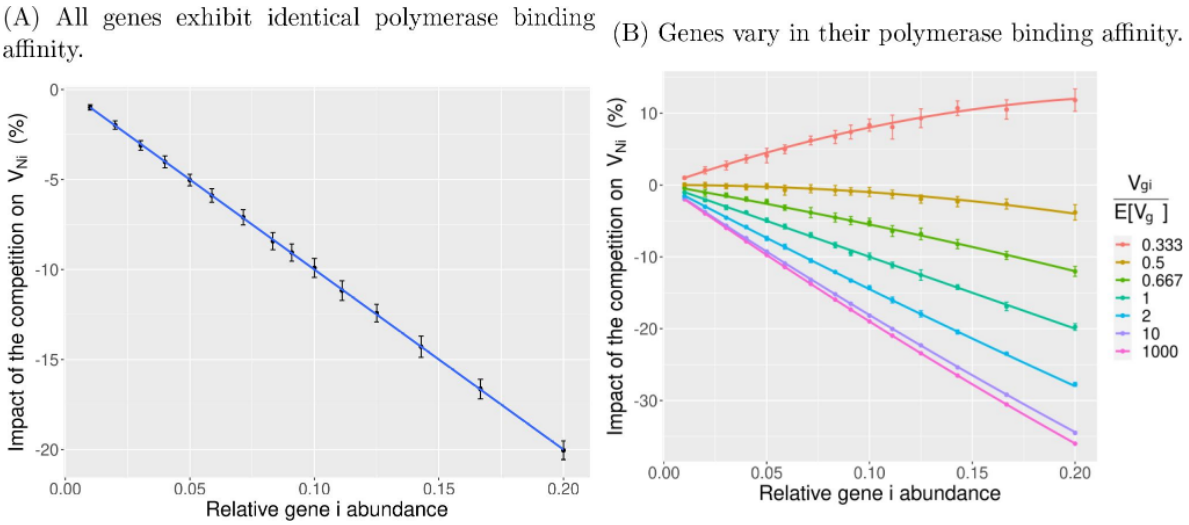

Figure S1

Impact of competition on variance in gene's polymerase binding activity.

(A) Scenario with genes having identical polymerase binding affinity across variable gene numbers *m* ranging from 5 to 100. **(B)** Scenario where genes exhibit heterogeneous polymerase binding affinities, also spanning variable gene numbers *m* from 5 to 100. The simulation data (dots with error bars) are juxtaposed with the theoretical formula (line). Each point represents the average outcome of 100 simulations based on 10,000 samples, encompassing a 90% confidence interval.

References

Barton N. H., Etheridge A. M., Véber A. (2017) **"The infinitesimal model: Definition, derivation, and implications** *In: Theoretical Population Biology* **118**:50–73 https://doi.org/10.1016/j.tpb.2017.06.001

Boyle Evan A., Li Yang I., Pritchard Jonathan K. (2017) **"An Expanded View of Complex Traits: From Polygenic to Omnigenic** *en. In: Cell* **169**:1177–1186 https://doi.org/10.1016/j.cell.2017.05.038

Brackley C. A., Gilbert N., Michieletto D., Papantonis A., Pereira M. C. F., Cook P. R., Marenduzzo D. (2021) **"Complex small-world regulatory networks emerge from the 3D organisation of the human genome** *In: Nature Communications* **12** https://doi.org/10.1038/s41467-021-25875-y

Brendler T., Godefroy-Colburn T., Carlill R. D., Thach R. E. (1981) **"The role of mRNA competition in regulating translation. II. Development of a quantitative in vitro assay** *eng. In: The Journal of Biological Chemistry* **256**:11747–11754

Brendler T., Godefroy-Colburn T., Yu S., Thach R. E. (1981) **"The role of mRNA competition in regulating translation. III. Comparison of in vitro and in vivo results** *eng. In: The Journal of Biological Chemistry* **256**:11755–11761

Chou T C, Talaly P (1977) **"A simple generalized equation for the analysis of multiple inhibitions of Michaelis-Menten kinetic systems** *In: Journal of Biological Chemistry* **252**:6438–6442 https://doi.org/10.1016/S0021-9258(17)39978-7

Chu Dominique, Barnes David J., von der Haar Tobias (2011) **"The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae** *In: Nucleic Acids Research* **39**:6705–6714 https://doi.org/10.1093/nar/gkr300

Claussnitzer Melina, et al. (2020) **"A brief history of human disease genetics** *en. In: Nature* **577**:179–189 https://doi.org/10.1038/s41586-019-1879-7

De Vos Dirk, Bruggeman Frank J., Westerhoff Hans V., Bakker Barbara M. (2011) **"How molecular competition influences fluxes in gene expression networks** *In: PloS One* **6** https://doi.org/10.1371/journal.pone.0028494

Elandt-Johnson Regina C., Johnson Norman L. (1980) **"Survival Models and Data Analysis** *en. In: Survival Models and Data Analysis*

Fernández-Tajes Juan, Gaulton Kyle J., van de Bunt Martijn, Torres Jason, Thurner Matthias, Mahajan Anubha, Gloyn Anna L., Lage Kasper, McCarthy Mark I. (2019) **"Developing a network view of type 2 diabetes risk pathways through integration of genetic, genomic and functional data** *In: Genome Medicine* **11** https://doi.org/10.1186/s13073-019-0628-8

Finucane Hilary K, et al. (2015) **"Partitioning heritability by functional annotation using genome-wide association summary statistics** *In: Nature Genetics* **47**:1228–1235 https://doi.org/10.1038/ng.3404

Fisher R. A. (1919) **"XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance** *en. In: Transactions of the Royal Society of Edinburgh* **52**:399–433 https://doi.org/10.1017/S0080456800012163

Frei Oleksandr, et al. (2019) **"Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation** *en. In: Nature Communications* **10** https://doi.org/10.1038/s41467-019-10310-0

Freimer Jacob W., Shaked Oren, Naqvi Sahin, Sinnott-Armstrong Nasa, Kathiria Arwa, Chen Amy F., Cortez Jessica T., Greenleaf William J., Pritchard Jonathan K., Marson Alexander (2021) **Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. en** https://doi.org/10.1101/2021.04.18.440363

Godefroy-Colburn T., Thach R. E. (1981) **"The Role of mRNA Competition in Regulating Translation. IV. Kinetic model** *eng. In: The Journal of Biological Chemistry* **256**:11762–11773

Gottesman Irving I., Gould Todd D. (2003) **"The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions** *In: American Journal of Psychiatry* **160**:636–645 https://doi.org/10.1176/appi.ajp.160.4.636

GTEx Consortium, et al. (2018) **"Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation** https://doi.org/10.1038/s41588-018-0154-4

International Schizophrenia Consortium, Purcell Shaun M., Wray Naomi R., Stone Jennifer L., Visscher Peter M., O'Donovan Michael C., Sullivan Patrick F., Sklar Pamela (2009) **"Common polygenic variation contributes to risk of schizophrenia and bipolar disorder** *eng. In: Nature* **460**:748–752 https://doi.org/10.1038/nature08185

Jostins Luke, et al. (2012) **"Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease** *en. In: Nature* **491**:119–124 https://doi.org/10.1038/nature11582

Kendall Maurice G, Stuart Alan, Ord John Keith (1994) **"Kendall's Advanced Theory of Statistics** *en. In: The advanced theory of statistics in 3 volumes. 1 1*

Klein Robert J., et al. (2005) **"Complement factor H polymorphism in age-related macular degeneration** *eng. In: Science* :385–389 https://doi.org/10.1126/science.1109557

Lin Jie, Amir Ariel (2018) **"Homeostasis of protein and mRNA concentrations in growing cells** *In: Nature Communications* https://doi.org/10.1038/s41467-018-06714-z

Liu Xuanyao, Li Yang I., Pritchard Jonathan K. (2019) **"Trans Effects on Gene Expression Can Drive Omnigenic Inheritance** *English. In: Cell* **177**:1022–1034 https://doi.org/10.1016/j.cell.2019.04.014

Loh Po-Ru, et al. (2015) **"Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis** *en. In: Nature Genetics* **47**:1385–1392 https://doi.org/10.1038/ng.3431

Maher Brendan (2008) **"Personal genomes: The case of the missing heritability** *eng. In: Nature* **456**:18–21 https://doi.org/10.1038/456018a

Manolio Teri A., et al. (2009) **"Finding the missing heritability of complex diseases** *en. In: Nature* **461**:747–753 https://doi.org/10.1038/nature08494

Michaelis Leonor, Menten Maud L, et al. (1913) **"Die kinetik der invertinwirkung** *In: Biochem. Z* **49**

O'Connor Luke J., Schoech Armin P., Hormozdiari Farhad, Gazal Steven, Patterson Nick, Price Alkes L. (2019) **"Extreme Polygenicity of Complex Traits Is Explained by Negative Selection** https://doi.org/10.1016/j.ajhg.2019.07.003

Pickrell Joseph K. (2014) **"Joint analysis of functional genomic data and genome-wide association studies of 18 human traits** *eng. In: American Journal of Human Genetics* **94**:559–573 https://doi.org/10.1016/j.ajhg.2014.03.004

Pickrell Joseph K., Berisa Tomaz, Liu Jimmy Z., Ségurel Laure, Tung Joyce Y., Hinds David A. (2016) **"Detection and interpretation of shared genetic influences on 42 human traits** *en. In: Nature Genetics* **48**:709–717 https://doi.org/10.1038/ng.3570

Schäuble Sascha, Stavrum Anne Kristin, Puntervoll Pål, Schuster Stefan, Heiland Ines (2013) **"Effect of substrate competition in kinetic models of metabolic networks** *In: FEBS Letters* **587**:2818–2824 https://doi.org/10.1016/j.febslet.2013.06.025

Shi Huwenbo, Kichaev Gleb, Pasaniuc Bogdan (2016) **"Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data** *en. In: The American Journal of Human Genetics* **99**:139–153 https://doi.org/10.1016/j.ajhg.2016.05.013

Sinnott-Armstrong Nasa, Naqvi Sahin, Rivas Manuel, Pritchard Jonathan K, Flint Jonathan, Wittkopp Patricia J, Lynch Vincent J, Wray Naomi, Chakravarti Aravinda (2021) **"GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background** *In: eLife* **10** https://doi.org/10.7554/eLife.58615

Trynka Gosia, Sandor Cynthia, Han Buhm, Xu Han, Stranger Barbara E, Liu X Shirley, Raychaudhuri Soumya (2012) **"Chromatin marks identify critical cell types for fine mapping complex trait variants** *In: Nature Genetics* **45**:124–130 https://doi.org/10.1038/ng.2504

Turelli Michael (2017) **"Commentary: Fisher's infinitesimal model: A story for the ages** *en. In: Theoretical Population Biology* **118**:46–49 https://doi.org/10.1016/j.tpb.2017.09.003

Turkheimer Eric (2000) **"Three Laws of Behavior Genetics and What They Mean** *en. In: Current Directions in Psychological Science* **9**:160–164 https://doi.org/10.1111/1467-8721.00084

Udler Miriam S. (2019) **"Type 2 Diabetes: Multiple Genes, Multiple Diseases** *en. In: Current Diabetes Reports* **19** https://doi.org/10.1007/s11892-019-1169-7

Võsa Urmo *et al.* (2018) **Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. en** https://doi.org/10.1101/447367

Võsa Urmo *et al.* (2021) **"Largescale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression** *en. In: Nature Genetics* **53** https://doi.org/10.1038/s41588-021-00913-z

Walden W. E., Godefroy-Colburn T., Thach R. E. (1981) **"The Role of mRNA Competition in Regulating Translation. I. Demonstration of competition in vivo** *eng. In: The Journal of Biological Chemistry* **256**:11739–11746

Weedon Michael N., et al. (2008) **"Genome-wide association analysis identifies 20 loci that influence adult height** *en. In: Nature Genetics* **40**:575–583 https://doi.org/10.1038/ng.121

Wellcome Trust Case Control Consortium (2007) **"Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls** *eng. In: Nature* **447**:661–678 https://doi.org/10.1038/nature05911

Wood Andrew R, et al. (2014) **"Defining the role of common variation in the genomic and biological architecture of adult human height** *In: Nature Genetics* **46**:1173–1186 https://doi.org/10.1038/ng.3097

Yang Jian, et al. (2010) **"Common SNPs explain a large proportion of the heritability for human height** *In: Nature Genetics* **42**:565–569 https://doi.org/10.1038/ng.608

Zhang Yan, Qi Guanghao, Park Ju-Hyun, Chatterjee Nilanjan (2018) **"Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits** https://doi.org/10.1038/s41588-018-0193-x

Zhu Xiang, Stephens Matthew (2018) **"Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes** *en. In: Nature Communications* **9**:1–14 https://doi.org/10.1038/s41467-018-06805-x

Article and author information

Olivier Naret

School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland, Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland

Yuval Simons

Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, USA, Department of Genetics, Stanford University, Stanford, CA 94305, USA
For correspondence: yuval.simons@uchicago.edu
ORCID iD: 0000-0002-0037-4673

Jacques Fellay

Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland, Precision Medicine Unit, Biomedical Data Sciences Center, Lausanne University Hospital and University of Lausanne, Lausanne, 1015, Switzerland
ORCID iD: 0000-0002-8240-939X

Jonathan K Pritchard

Department of Genetics, Stanford University, Stanford, CA 94305, USA, Department of Biology, Stanford University, Stanford, CA 94305, USA
For correspondence: pritch@stanford.edu

Editors

Reviewing Editor
Diethard Tautz
Max Planck Institute for Evolutionary Biology, Plön, Germany

Senior Editor
Detlef Weigel
Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public Review):

This study explores whether the extreme polygenicity of common traits (the fact that variation in such traits is explained by a very large number of genetic variants) could be explained in part by competition among genes for limiting molecular resources involved in gene regulation, which would cause the expression of most genes to be correlated. While the hypothesis is interesting, I still have some concerns about the analysis and interpretation.

As the authors say in their rebuttal, assuming extreme resource limitation, i.e., going from equation 2 to 5 essentially assumes assuming that 1/(gtot [G]) <<1 and that terms that are order [1/(gtot [G])] can neglected. However, then the authors derive so-called resource competition terms that are order (1/m) where m is the number of genes, so that gtot is proportional to m. My main criticism (which I am not sure was addressed) is thus: can we reliably derive small order (1/m) effects while neglecting order [1/(gtot [G])] terms, when both are presumably similar in order of magnitude? Is this mathematically sound?

I do not think the supplement that the authors have added actually gets to this. For example, section 7.1 just gives the textbook derivation of Michelis-Menten kinetics, and does not address my earlier criticism that the terms neglected in going from eq. 16 to eq. 17 (or from eq. 2 to 3) may be similar in magnitude to the terms being derived and interpreted in eqs. 6 and 7.
Similarly, it is unclear from section 7.2 how the authors are doing the simulations. Are these true Michelis-Menten simulations involving equation 2? If yes, then what is the value of [G] and [P_0] in the simulations? If these are not true Michelis-Menten simulations, but instead something that already uses equation 5, then this still does not address my earlier criticism.

https://doi.org/10.7554/eLife.88770.2.sa2

Reviewer #2 (Public Review):

The question the authors pose is very simple, and yet very important. Does the fact that many genes compete for Pol II to be transcribed explain why so many trans-eQTL contribute to the heritability of complex traits? That is, if a gene uses up a proportion of Pol II, does that in turn affect the transcriptional output of other genes relevant or even irrelevant for the trait in a way that their effect will be captured in a genome-wide association study? If yes, then the

large number of genetic effects associated with variation in complex traits can be explained but such trans-propagating effects on transcriptional output of many genes.

This is a very timely question given that we still don't understand how, mechanistically, so many genes can be involved in complex traits variation. Their approach to this question is very simple and it is framed in classic enzyme-substrate equations. The authors show that the trans-propagating effect is too small to explain the ~70% of heritability of complex traits that is associated with trans-effects. Their conclusion relies on the comparison of the order of magnitude of a) the quantifiable transcriptional effects due to Pol II competition, and b) the observed percentage of variance explained by trans effects (data coming from Liu et al 2019, from the same lab).

The results shown in this manuscript rule out that competition for limiting resources in the cell (not restricted to Pol II, but applicable to any other cellular resource like ribosomes, etc) could explain heritability of complex traits.

https://doi.org/10.7554/eLife.88770.2.sa1

Reviewer #3 (Public Review):

Human complex traits including common diseases are highly polygenic (influenced by thousands of loci). This observation is in need of an explanation. The authors of this manuscript propose a model that a competition for a single global resource (such as RNA polymerase II) may lead to a highly polygenic architecture of traits. Following an analytical examination the authors reject their hypothesis. This work is of clear interest to the field. It remains to be seen if the model covers the variety of possible competition models.

https://doi.org/10.7554/eLife.88770.2.sa0

Author Response

The following is the authors' response to the original reviews.

> ***eLife assessment***
>
> *This important study addresses the fundamentally unresolved question of why many thousands of small-effect loci contribute more to the heritability of a trait than the large-effect lead variants. The authors explore resource competition within the transcriptional machinery as one possible explanation with a simple theoretical model, concluding that the effects of resource competition would be too small to explain the heritability effects. The topic and approximation of the problem are very timely and offer an intuitive way to think about polygenic variation, but the analysis of the simple model appears to be incomplete, leaving the main claims only partially supported.*

We thank eLife for recognizing the importance of our work. We hope the revised manuscript addresses the reviewers' reservations.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *This study explores whether the extreme polygenicity of common traits can be explained in part by competition among genes for limiting molecular resources (such as RNA polymerases) involved in gene regulation. The authors hypothesise that such competition would cause the expression levels of all genes that utilise the same molecular resource to*

> *be correlated and could thus, in principle, partly explain weak trans-regulatory effects and the observation of highly polygenic architectures of gene expression. They study this hypothesis under a very simple model where the same molecule binds to regulatory elements of a large number m of genes, and conclude that this gives rise to trans-regulatory effects that scale as 1/m, and which may thus be negligible for large m.*

We thank the reviewer for their thorough and thoughtful review of our manuscript.

> *The main limitation of this study lies in the details of the mathematical analysis, which does not adequately account for various small effects, whose magnitude scales inversely with the number m of genes that compete for the limiting molecular resource. In particular, the fraction of "free" molecule (which is unbound to any of the genes) also scales as 1/m, but is not accounted for in the analysis, making it difficult to assess whether the quantitative conclusions are indeed correct.*

It is explicitly accounted for in the supplement.

> *Second, the questions raised in this study are better analysed in the framework of a sensitivity or perturbation analysis, i.e., by asking how* changes *in expression level or binding affinity at one gene (rather than the total expression level or total binding affinity) affect expression level at other genes. In the context of complex traits, where an increase in gene expression can either increase or decrease the trait, we believe the most important quantity of interest is variation in expression and, therefore, trait variation. Nevertheless, our results do show that the relative change in expression due to competition is also small.*
>
> *Thus, while the qualitative conclusion that resource competition in itself is unlikely to mediate trans-regulatory effects and explain highly polygenic architectures of gene expression traits probably holds, the mathematical reasoning used to arrive at this conclusion requires more care.*
>
> *In my opinion, the potential impact of this kind of analysis rests at least partly on the plausibility of the initial hypothesis- namely whether most molecular resources involved in gene regulation are indeed "limiting resources". This is not obvious, and may require a careful assessment of existing evidence, e..g., what is the concentration of bound vs. unbound molecular species (such as RNA polymerases) in various cell types?*

We intentionally looked at the most extreme case of extreme resource limitation, and we conclude that since extreme resource limitation is a small effect, the same would be true of weak resource limitation, when unbound molecules play an important role. We put more emphasis on this point in our revised text.

> ***Reviewer #1 (Recommendations For The Authors):***
>
> *While the main conclusion that resource competition in itself is unlikely to mediate trans effects and explain high levels of polygenicity may well be correct, I am not convinced that the mathematical reasoning presented in support of this conclusion is entirely correct. I will attempt to outline my concerns mainly in the context of section 2, since the arguments in sections 3 and 4 build upon this.*
>
> *(a) The key assumption underlying the approximations in equations 3, 4, and 5 is that there is very little free polymerase, in other words /_0 is a small quantity. However, the second and third terms that emerge in equation 7 are also small quantities and (as far as I can see) of the same order as /_0. Thus, one cannot simply use equation 4 or 5 as a starting point to derive eq. 7 and should instead use the exact x_i = (g_i [G])/ (1+g_tot [G]),*

> *in order to make sure that all (and not just some) terms that are similar in order of magnitude are accounted for in the analysis.*

The concentration of free polymerase is marked as [P], and we explicitly assume (just before eq. 2) that [P]<<[P]0 with [P]0 being the overall concentration of polymerase. This is a conservative assumption – we consider extreme resource competition with little free polymerase and since we since only a small effect in this extreme scenario we assume it would be a small effect also for less extreme scenarios. We put more emphasis on this point in our revised text.

> *More concretely, the difference between the exact x_i = (g_i [G])/ (1+g_tot [G]) and the approximate x_i = (g_i / g_tot) is precisely 1/m (for large m) in the example considered line 246 onwards. Thus, I suspect that the conclusion that Var[x_i] = (1-1/m)Var[g_i] in that example is just an artefact of starting with eqs. 4 and 5. As a sanity check, it may be useful to actually simulate resource competition explicitly (maybe using a deterministic simulation) under the explicit model [PG_i] = g_i [G] and _0 = + Sum[[PG]_i , i=1,m] without making any further approximations to see if perturbations in g_i actually produce Order [1/m] effects in the variance of x_i for the example considered line 246 onwards (this would require simulating with a few different m and plotting Var[x_i] vs. m for example).*

The exact equation the reviewer is alluding to describes a scenario of non-extreme resource competition. If g_tot [G]>>1, i.e. if most polymerase is bound to a gene then x_i is equal to g_i/g_tot and this is the scenario we are considering of extreme competition. If g_tot [G]<<1, then x_i=g_i [G] and competition has no effect. While the intermediate case is interesting, we see no reason for the effects to be larger than in the extreme competition case. We have added the results of simulations in the supplement to validate our arguments.

> *Lines 231-239: Because of the concerns highlighted above and questions about the validity of equation 7, I am not convinced that the interpretations given here and also in section 4 are correct.*
>
> *(b) Lines 219-230 (including equations 6 and 7): I think to address the question of whether genetic changes in cis-regulatory elements for a given gene have an effect on other genes (under this model of resource competition), it is better to spell out the argument in terms of Var[dx_i] rather than Var[x_i], where dx_i is the change in expression level at gene i due to changes at all m genes, dg_i is the change in gene activity due to (genetic) changes in the relevant regulatory elements associated with gene i etc. Var[dx_i] can then be expressed as a sum of Var[dg_i], Var[dg_tot] and Cov[d g_i, dg_tot]. However, I suspect that to do this correctly, one should not start with the approximate x_i=g_i/g_tot : see previous comment.*

The variance of the deviation from the mean is mathematically identical to the overall variance, Var[dx_i]= Var[x_i]. Our analysis is therefore equivalent to the suggested analysis.

> *Somewhere in all of this, there is also an implicit assumption that E[dg_i] is zero, i.e, mutations are as likely to increase as to decrease binding affinities so that one needs to only consider Var[dx_i] and not E[dx_i]; this assumption should be spelled out.*

Our results concern the variation around trait means and therefore we have not included a possible mean effect of mutation, which would not affect the results but just shift the mean.

> *Some minor comments (mostly related to the introduction and general context):*

> - *I think it would be worth connecting more with the literature on molecular competition and gene regulation (see e.g., How Molecular Competition Influences Fluxes in Gene Expression Networks, De Vos et al, Plos One 2011). Even though this literature does not frame questions in terms of "polygenicity of traits", these analyses address the same basic questions: to what extent do perturbations in gene expression at one gene affect other genes, or to what extent is there crosstalk between different genes or pathways?*

We have expanded our introduction to refer to De Vos et al, as well as a few other papers we have recently become aware of. (e.g., Jie Lin & Ariel Amir Nature Communications volume 9, Article number: 4496 (2018))

> - *Lines 88-89: "supports the network component of the model" is a vague phrase that does not convey much. It would be useful to clarify and make this more precise.*

We have clarified this phrasing in the text.

> - *Lines 113-114: In the context of "selective constraint", it may also be worth discussing previous work by one of the authors: "A population genetic interpretation of GWAS findings for human quantitative traits". What implications would stabilizing selection on multiple traits (as opposed to simple purifying selection) have for the distribution of variances across trait loci and the extent to which trait architectures appear to be polygenic?*

While most definitely of great interest to some of the authors, the distribution of variance across loci does not affect our results.

> *References: Barton and Etheridge 2018 in line 54 is not the correct reference; it should be Barton et al 2017 (paper with Amandine Veber). Fisher 1919 in line 52 is actually Fisher 1918. The formatting of references in the next paragraph (and in various other places in the paper) is also a bit unusual, with some authors referred to by their full names and others only by their last. I believe that it may be useful to crosscheck references throughout the paper.*

We have crosschecked the references in the paper.

> *Line 164: Some word appears to be missing here. Maybe bound -> bound to ?*

Fixed

> ***Reviewer #2 (Public Review):***
>
> *The question the authors pose is very simple and yet very important. Does the fact that many genes compete for Pol II to be transcribed explain why so many trans-eQTL contribute to the heritability of complex traits? That is, if a gene uses up a proportion of Pol II, does that in turn affect the transcriptional output of other genes relevant or even irrelevant for the trait in a way that their effect will be captured in a genome-wide association study? If yes, then the large number of genetic effects associated with variation in complex traits can be explained but such trans-propagating has effects on the transcriptional output of many genes.*

> *This is a very timely question given that we still don't understand how, mechanistically, so many genes can be involved in complex traits variation. Their approach to this question is very simple and it is framed in classic enzyme-substrate equations. The authors show that the trans-propagating effect is too small to explain the ~70% of heritability of complex traits that are associated with trans-effects. Their conclusion relies on the comparison of the order of magnitude of a) the quantifiable transcriptional effects due to Pol II competition, and b) the observed percentage of variance explained by trans effects (data coming from Liu et al 2019, from the same lab).*
>
> *The results shown in this manuscript rule out that competition for limited resources in the cell (not restricted to Pol II, but applicable to any other cellular resource like ribosomes, etc) could explain the heritability of complex traits.*

We thanked the Reviewer for his resounding support of our paper!

The authors rely on simulated data, and although the conclusions hold in a biologically-realistic scenario given the big difference in effect sizes, I wonder if the authors could provide data from the literature (if available) that give the reader a point of reference for the steady state of cells in terms of free/occupied Pol II molecules and/or free/occupied transcription binding sites. This information won't change the conclusion of the manuscript, but it will put it in the context of real biological data.

We have scoured the literature, but have not found readily available data with which to validate our results (beyond that which is already referenced).

> ***Reviewer #3 (Public Review):***
>
> *Human complex traits including common diseases are highly polygenic (influenced by thousands of loci). This observation is in need of an explanation. The authors of this manuscript propose a model that competition for a single global resource (such as RNA polymerase II) may lead to a highly polygenic architecture of traits. Following an analytical examination, the authors reject their hypothesis. This work is of clear interest to the field. It remains to be seen if the model covers the variety of possible competition models.*

We thank the Reviewer for his assessment, support and comments.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *This manuscript provides a straightforward and elegant quantitative argument that the competition for the RNA polymerase is not a significant source of trans-eQTLs and, more generally, of genetic variance of complex polygenic phenotypes. This is an unusual manuscript because the authors propose a hypothesis that they confidently reject based on a calculation. This negative result is intuitive. Still, the manuscript is of interest. Progress in understanding the highly polygenic architecture of complex traits is welcome, and the resource competition hypothesis is quite natural. I have three specific comments/concerns listed below.*
>
> *(1) The manuscripts states that V(x_i)=V(g_i/g_tot). Unless I am missing something, this seems to result from a very strong implicit assumption that all genetic variance is due to variation in the binding of RNA polymerase, while x_i_max is a constant. I would expect that x_i_max may also be genetically variable due to many effects unrelated to the Pol II*

> ***Reviewer #2 (Recommendations For The Authors):***

The authors rely on simulated data, and although the conclusions hold in a biologically-realistic scenario given the big difference in effect sizes, I wonder if the authors could provide data from the literature (if available) that give the reader a point of reference for the steady state of cells in terms of free/occupied Pol II molecules and/or free/occupied transcription binding sites. This information won't change the conclusion of the manuscript, but it will put it in the context of real biological data.

We have scoured the literature, but have not found readily available data with which to validate our results (beyond that which is already referenced).

> ***Reviewer #3 (Public Review):***
>
> *Human complex traits including common diseases are highly polygenic (influenced by thousands of loci). This observation is in need of an explanation. The authors of this manuscript propose a model that competition for a single global resource (such as RNA polymerase II) may lead to a highly polygenic architecture of traits. Following an analytical examination, the authors reject their hypothesis. This work is of clear interest to the field. It remains to be seen if the model covers the variety of possible competition models.*

We thank the Reviewer for his assessment, support and comments.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *This manuscript provides a straightforward and elegant quantitative argument that the competition for the RNA polymerase is not a significant source of trans-eQTLs and, more generally, of genetic variance of complex polygenic phenotypes. This is an unusual manuscript because the authors propose a hypothesis that they confidently reject based on a calculation. This negative result is intuitive. Still, the manuscript is of interest. Progress in understanding the highly polygenic architecture of complex traits is welcome, and the resource competition hypothesis is quite natural. I have three specific comments/concerns listed below.*
>
> *(1) The manuscripts states that V(x_i)=V(g_i/g_tot). Unless I am missing something, this seems to result from a very strong implicit assumption that all genetic variance is due to variation in the binding of RNA polymerase, while x_i_max is a constant. I would expect that x_i_max may also be genetically variable due to many effects unrelated to the Pol II binding (e.g. transcription rate, bursting, presence of R-loops etc.). I guess that the assumption made by the authors is conservative.*

Indeed. We made conservative assumptions throughout, aiming to consider the most extreme scenario in which resource competition may affect trait variation. Our logic being that if even under the most extreme scenario resource competition is a small effect then it is a small effect in all scenarios. We put more emphasis on this point in our revised text.

> *(2) The manuscript focuses on the competition for RNA polymerase but suggests that the lesson learned is highly generalizable. However, it is an example of a single global limiting resource resulting in first-order kinetics. What happens in a realistic scenario of competition for multiple resources associated with transcription and with downstream processes (free ribonucleotides, spliceosome, polyadenylation machinery, ribosome, post-translational modifications)? It is possible that in most cases a single resource is a limiting factor, but an investigation (or even a brief discussion) of this question would support the claim that the results are generalizable.*

We expect competition for multiple resource to result in similarly weak effects. Since there is not a great number of such resources, we do not expect it to change our qualitative result. We added language to that effect in the main text.

> *(3) Alternatively, what happens in a scenario of competition for multiple local resources shared by a few genes (co-factors, substrates, chaperones, micro-RNAs, post-translational modification factors such as kinases, degradation factors, scaffolding proteins)? In this case, each gene would compete for resources with a few other genes increasing polygenicity without a global competition with all other genes. Intuitively, a large set of such local competitions may lead to a highly polygenic architecture.*

This is indeed a scenario in which competition may be a large effect which we mention in our discussion. “the conclusions may differ in contexts where a very small number of genes compete for a highly limited resource, such as access to a particular molecular transporter”