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

Yuval Simons[*1], Olivier Naret[*2,3], Jacques Fellay[2,3,4], Jonathan K Pritchard[1,5]

[1] Department of Genetics, Stanford University, Stanford, CA 94305, USA
[2] School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland
[3] Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
[4] Precision Medicine Unit, Biomedical Data Sciences Center, Lausanne University Hospital and University of Lausanne, Lausanne, 1015, Switzerland
[5] Department of Biology, Stanford University, Stanford, CA 94305, USA
*Co-first authors
Correspondence: pritch@stanford.edu

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.

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 [1, 2, 3]. 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 [4, 5]. 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" [6].

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 [7, 8]. 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 [9, 10, 11, 12]. Although there is some contribution from coding variants, most of the heritability comes from non-coding variants impacting gene regulation [13, 14, 15]. These variants are spread surprisingly uniformly across the genome rather than being strongly concentrated near important genes or in particular chromosomal regions [16, 17]. 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 [18].

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 [19, 20, 21, 22, 23]. The observation that the SNPs contributing heritability are spread relatively uniformly across the genome [16] implies that a large fraction of genes must be contributing to the trait variance [21]. 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 [12]. 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.

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 [21, 24]. 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 [12]. Recent work on correlations between polygenic scores for various traits and whole blood gene expression of likely core genes also supports the network component of the model [25].

Furthermore, we showed that there is a natural connection between our model and estimates of cis- and trans-heritability of gene expression [24]. 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 [24]. 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 [26]. Thus, any variants that affect the intermediate processes can potentially be detected in GWAS of the endpoint trait [27, 28, 29, 30, 26]. 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 [12]. 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 [11]. 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.

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 [21, 24]. 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 [31, 32]. In the present paper, we consider a non-specific form of regulation: resource competition.

The fundamental notion of the resource competition model is that each cell contains finite pools of shared molecules that are important for gene expression and regulation, including RNA Polymerase II, nucleotides, spliceosomes, tRNAs and ribosomes [33, 34, 35, 36, 37]. Consider a cis-eQTL for a particular gene. 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 [38]. 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. 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.* [24]. 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.

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 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 [39, 40]. 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 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. $[P] << [P]_0$. 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 overall 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.

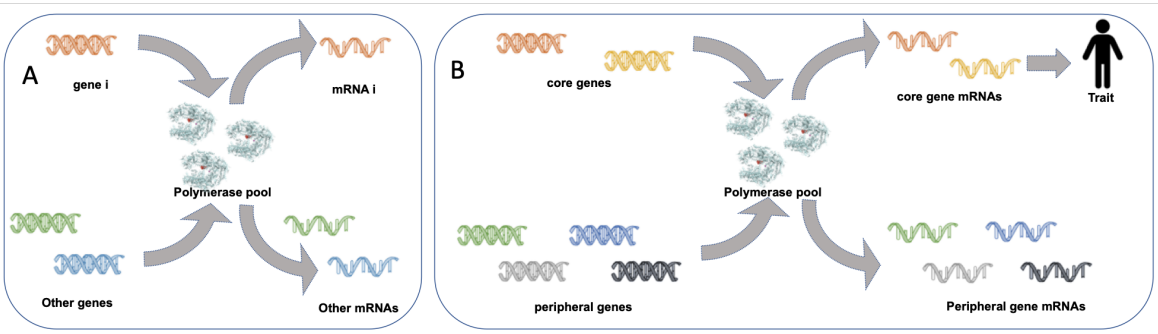

Figure 1: **Illustration of the competition model for polymerases. (A)** *Transcription of a single gene may be affected by polymerase depletion due to transcription of other genes.* **(B)** *This same model extends to complex traits by distinguishing between transcription of core and peripheral genes.*

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) [24]. 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:

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

(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 $a << b$, can be approximated by a first-order Taylor expansion as follows [41, 42]:

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

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

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

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) $\frac{E[g_i]^2}{E[g_{tot}]^2}V\left[g_{tot}\right]$: 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:

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

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 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) are 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.

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.* [24], 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

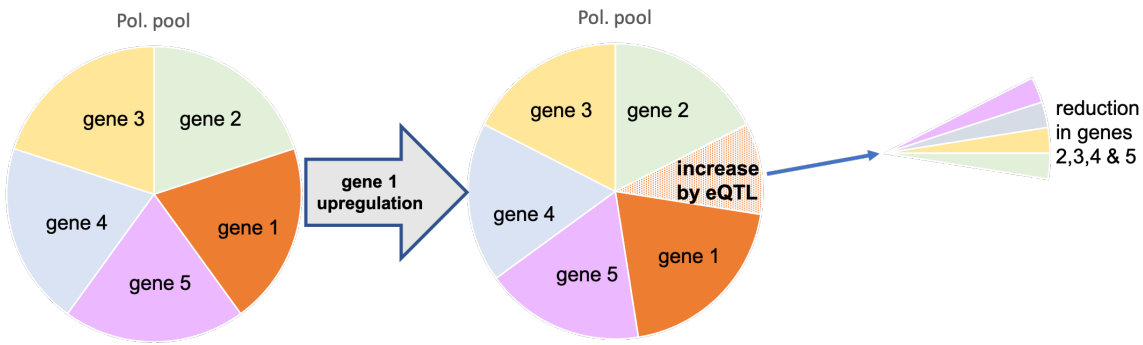

Figure 2: **Illustration of the effect of resource competition on expression.** *Upregulation of Gene* 1 *causes a small downregulation in the expression of all other genes. Since this change is spread between all gene it scales like* 1 *over the number of genes.*

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 affect 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 expression of gene i increases trait values and a negative γ_i indicates that an increase in expression of gene i decreases trait values. In this model, an individual's phenotype is given by:

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

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 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 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

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

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.* [24], 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

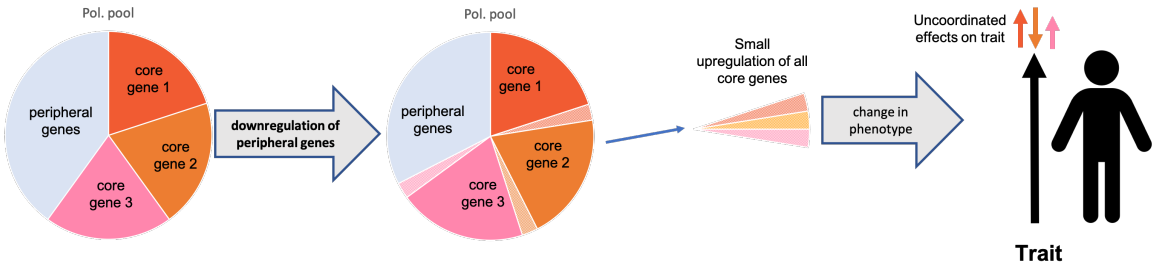

Figure 3: **Illustration of the effect of resource competition on compex trait phenotype.** *Downregulation of one or more peripheral genes causes a small upregulation of all core genes. As we show in the text, this will only be an appreciable effect if core genes engage a large fraction of the overall 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.*

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 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 condition 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, temporal and spatial patters, 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.

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.

6 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.

7 Supplement: 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 [43] and arriving at a similar result to Schauble et al. [40]. 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 $g_{tot} = \sum_i g_i$ 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, $[P]_0$.

References

[1] Robert J. Klein, Caroline Zeiss, Emily Y. Chew, et al. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.)*, 308(5720):385–389, April 2005.

[2] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, June 2007.

[3] Melina Claussnitzer, Judy H. Cho, Rory Collins, et al. A brief history of human disease genetics. *Nature*, 577(7789):179–189, January 2020. Number: 7789 Publisher: Nature Publishing Group.

[4] Michael N. Weedon, Hana Lango, Cecilia M. Lindgren, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, 40(5):575–583, May 2008. Number: 5 Publisher: Nature Publishing Group.

[5] Teri A. Manolio, Francis S. Collins, Nancy J. Cox, et al. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753, October 2009.

[6] Brendan Maher. Personal genomes: The case of the missing heritability. *Nature*, 456(7218):18–21, November 2008.

[7] International Schizophrenia Consortium, Shaun M. Purcell, Naomi R. Wray, Jennifer L. Stone, Peter M. Visscher, Michael C. O'Donovan, Patrick F. Sullivan, and Pamela Sklar. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256):748–752, August 2009.

[8] Jian Yang, Beben Benyamin, Brian P McEvoy, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, July 2010.

[9] Yan Zhang, Guanghao Qi, Ju-Hyun Park, and Nilanjan Chatterjee. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics*, page 1, August 2018.

[10] Oleksandr Frei, Dominic Holland, Olav B. Smeland, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications*, 10(1):2417, June 2019. Number: 1 Publisher: Nature Publishing Group.

[11] Luke J. O'Connor, Armin P. Schoech, Farhad Hormozdiari, Steven Gazal, Nick Patterson, and Alkes L. Price. Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *The American Journal of Human Genetics*, August 2019.

[12] Nasa Sinnott-Armstrong, Sahin Naqvi, Manuel Rivas, and Jonathan K Pritchard. GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *eLife*, 10:e58615, February 2021. Publisher: eLife Sciences Publications, Ltd.

[13] Gosia Trynka, Cynthia Sandor, Buhm Han, Han Xu, Barbara E Stranger, X Shirley Liu, and Soumya Raychaudhuri. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genetics*, 45(2):124–130, December 2012.

[14] Joseph K. Pickrell. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American Journal of Human Genetics*, 94(4):559–573, April 2014.

[15] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11):1228–1235, September 2015.

[16] Po-Ru Loh, Gaurav Bhatia, Alexander Gusev, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, 47(12):1385–1392, December 2015.

[17] Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *The American Journal of Human Genetics*, 99(1):139–153, July 2016.

[18] R. A. Fisher. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 52(02):399–433, 1919.

[19] Luke Jostins, Stephan Ripke, Rinse K. Weersma, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 491(7422):119–124, November 2012.

[20] Andrew R Wood, Tonu Esko, Jian Yang, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46(11):1173–1186, October 2014.

[21] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186, June 2017.

[22] Xiang Zhu and Matthew Stephens. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nature Communications*, 9(1):1–14, October 2018.

[23] Juan Fernández-Tajes, Kyle J. Gaulton, Martijn van de Bunt, Jason Torres, Matthias Thurner, Anubha Mahajan, Anna L. Gloyn, Kasper Lage, and Mark I. McCarthy. Developing a network view of type 2 diabetes risk pathways through integration of genetic, genomic and functional data. *Genome Medicine*, 11(1):19, March 2019.

[24] Xuanyao Liu, Yang I. Li, and Jonathan K. Pritchard. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell*, 177(4):1022–1034.e6, May 2019.

[25] Urmo Võsa, Annique Claringbould, Harm-Jan Westra, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. preprint, Genomics, October 2018.

[26] Miriam S. Udler. Type 2 Diabetes: Multiple Genes, Multiple Diseases. *Current Diabetes Reports*, 19(8):55, July 2019.

[27] Eric Turkheimer. Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*, 9(5):160–164, October 2000. Publisher: SAGE Publications Inc.

[28] Irving I. Gottesman and Todd D. Gould. The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *American Journal of Psychiatry*, 160(4):636–645, April 2003. Publisher: American Psychiatric Publishing.

[29] G. Bittante, M. Penasa, and A. Cecchinato. Invited review: Genetics and modeling of milk coagulation properties. *Journal of Dairy Science*, 95(12):6843–6870, December 2012.

[30] Joseph K. Pickrell, Tomaz Berisa, Jimmy Z. Liu, Laure Ségurel, Joyce Y. Tung, and David A. Hinds. Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, 48(7):709–717, July 2016.

[31] Urmo Võsa, Annique Claringbould, Harm-Jan Westra, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics*, 53(9):1300–1310, September 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Gene expression;Gene regulation;Genome-wide association studies Subject_term_id: gene-expression;gene-regulation;genome-wide-association-studies.

[32] Jacob W. Freimer, Oren Shaked, Sahin Naqvi, Nasa Sinnott-Armstrong, Arwa Kathiria, Amy F. Chen, Jessica T. Cortez, William J. Greenleaf, Jonathan K. Pritchard, and Alexander Marson. Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. Technical report, April 2021. Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article.

[33] T. Brendler, T. Godefroy-Colburn, S. Yu, and R. E. Thach. The role of mRNA competition in regulating translation. III. Comparison of in vitro and in vivo results. *The Journal of Biological Chemistry*, 256(22):11755–11761, November 1981.

[34] T. Godefroy-Colburn and R. E. Thach. The Role of mRNA Competition in Regulating Translation. IV. Kinetic model. *The Journal of Biological Chemistry*, 256(22):11762–11773, November 1981.

[35] Dominique Chu, David J. Barnes, and Tobias von der Haar. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic Acids Research*, 39(15):6705–6714, August 2011.

[36] W. E. Walden, T. Godefroy-Colburn, and R. E. Thach. The Role of mRNA Competition in Regulating Translation. I. Demonstration of competition in vivo. *The Journal of Biological Chemistry*, 256(22):11739–11746, November 1981.

[37] T. Brendler, T. Godefroy-Colburn, R. D. Carlill, and R. E. Thach. The role of mRNA competition in regulating translation. II. Development of a quantitative in vitro assay. *The Journal of Biological Chemistry*, 256(22):11747–11754, November 1981.

[38] GTEx Consortium, Eric R. Gamazon, Ayellet V. Segrè, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics*, June 2018.

[39] Leonor Michaelis, Maud L Menten, et al. Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352, 1913.

[40] Sascha Schäuble, Anne Kristin Stavrum, Pål Puntervoll, Stefan Schuster, and Ines Heiland. Effect of substrate competition in kinetic models of metabolic networks. *FEBS Letters*, 587(17):2818–2824, 2013. A century of Michaelis - Menten kinetics.

[41] Regina C. Elandt-Johnson and Norman L. Johnson. Survival Models and Data Analysis. In *Survival Models and Data Analysis*, page 69. John Wiley & Sons, September 1980.

[42] Maurice G Kendall, Alan Stuart, and John Keith Ord. Kendall's Advanced Theory of Statistics. In *The advanced theory of statistics in 3 volumes. 1 1*, page 351. Griffin, London, 1994. OCLC: 1071028235.

[43] T C Chou and P Talaly. A simple generalized equation for the analysis of multiple inhibitions of michaelis-menten kinetic systems. *Journal of Biological Chemistry*, 252(18):6438–6442, 1977.