Dear Editors,

I would like to submit the accompanying manuscript, entitled “**Deciphering Complex Traits with High-Order Genetic Analysis**”, for your consideration as a Resource in *Cell*.

A central question in genetics is how the products encoded by genes are organised into systems that give rise to traits. This is often stated as the genotype-to-phenotype problem. One way to address the genotype-to-phenotype problem is to perturb genes and observe changes in phenotype. This way, we can derive an understanding that accounts for the observed perturbations. For example, perturbing two genes at a time can give rise to genetic interactions – e.g. two-gene perturbation phenotypes which can’t be simply explained by just combining the effects of each perturbation alone. These genetic interactions have been widely studied by approaches such as synthetic genetic array analysis (SGA), and have been useful in creating models that can point to functional dependencies between genes to explain the observed phenotypes.

However, a genotype-to-phenotype understanding derived from genetic perturbations is only as good as the perturbation phenotypes it has to account for. Similar to how perturbations of two genes can give rise to phenotypes that can’t be accounted for by one-gene perturbation phenotypes, there has been increasing interest in what happens when multiple genes are perturbed together. It is known that more complex ‘high-order’ genetic interactions can occur, that is, phenotypes arise which can’t be simply explained by just combining the effects of lower-order perturbations. For example, studies in yeast have underscored the prevalence of three-gene interactions, and it is becoming increasingly appreciated from studies of multiple variants at one gene that these higher-order effects can be used to derive functional understanding.

Here, we describe an approach for mapping genotype-to-phenotype that makes use of these high-order interactions to derive understanding of multi-gene systems. This is fundamentally a generalization to approaches such as SGA in order to consider the effects of perturbing any number of (‘*X*’) genes at a time, hence an ‘*X* -gene’ genetic analysis (XGA). Key elements of XGA are the engineering and profiling of multi-gene variants within a set of chosen genes, and the subsequent discovery of high-order genetic effects which can be used to derive more complete genotype-to-phenotype models. This study addresses both the experimental difficulties in generating many combinatorial multi-variant strains, and the analytic challenges in deriving functional models from these data.

We demonstrate our XGA approach on 16 ABC transporters. These ABC transporters are a highly-conserved gene family that contains several well-known multidrug efflux pumps. We were motivated by previous studies which had found surprising knockout effects in this family, and generated 5,352 strains which we profiled for resistance to 16 bioactive compounds (‘drugs’). Indeed, we discovered many high-order interactions between these genes that mediated resistance to these drugs.

and indeed discovered many high-order interactions between these genes.

discovered many surprising functional dependencies which would not have been revealed by knocking out one or two genes.

---

engineering approach that allows us to efficiently create many multi-gene variants

this approach can only account

This idea is the basis for

A central question in genetics is how the products of genes are organised into systems that give rise to traits. This is often stated as the genotype-to-phenotype problem. One way to address the genotype-to-phenotype problem is to perturb genes and observe changes in phenotype. This is a powerful approach because it can be used to learn genotype-to-phenotype models that directly account for the effects of perturbation. However, it isn’t straightforward to use changes in phenotype to model genetic systems directly. Fundamentally, any genotype-to-phenotype understanding is only as good as the observations it accounts for. For example, if we try to model genotype-to-phenotype by only observing what happens when we perturb genes one at a time, then we exclude all observations that point to functional dependencies. To begin to study functional dependencies, we must perturb at least two genes at a time. This way, we may observe pairwise genetic interactions – or two-gene perturbation phenotypes which can’t be simply explained by just combining the effects of each perturbation alone. These genetic interactions have been widely studied by approaches such as synthetic genetic array analysis (SGA), and have been useful in pointing to relatedness and functional dependencies between genes.

However, just as a genotype-to-phenotype understanding learned by only observing single-gene perturbations does not account for all phenotypes revealed by two-gene perturbations, we currently do not have many genotype-to-phenotype models that generally account for the effects of multiple genetic perturbations. This makes any derived genotype-to-phenotype understanding incomplete. It is both experimentally challenging to map out the vast space of possible combinatorial perturbations, and it becomes difficult to manually reason over these complex effects to derive understanding. These are the problem which our manuscript addresses. To tackle the experimental problem, we describe a population engineering approach that allows us to efficiently create many multi-gene variants that we can then profile for phenotypes which are informative for the functions we want to study. The analytic problem is two-fold. The simpler problem is to determine the higher-order genetic interactions that are evident from the combinatorial perturbations. This describes sets of genes which have yielded interesting effects, but does not give a satisfying explanation of why these effects are occurring.

The more difficult problem is to figure out a way to derive genotype-to-phenotype models that more intuitively explain what is going on. That is, we want to replace the laborious, error-prone, and potentially subjective reasoning steps that are usually performed by the geneticist. To address these issues, we developed a neural network model that can learn biologically plausible genotype-to-phenotype models directly from complex genetic data. It may be surprising to see a neural network used in context, as they are often thought of as uninterpretable black boxes. What may be less appreciated is that these networks are a flexible way to use functional knowledge to state a ‘schematic’ of a system that uses genetic data being to fill in the details. Furthermore, these models are quantitative, so that the schematic itself can be evaluated and expanded as needed. We believe that this is a much more powerful approach compared to manual analysis, and that it eliminates many confusions in interpreting complex genetic data.

By addressing both major experimental and analytic challenges, we believe that we have made a fundamental step forward in the ability to learn more complete genotype-to-phenotype models from complex genetic relationships. We demonstrate our approach on ABC transporters, which are a highly-conserved gene family that contains several well-known multidrug efflux pumps. We were motivated by previous studies which had found surprising knockout effects in this family, and discovered many surprising functional dependencies which would not have been revealed by knocking out one or two genes. We termed our study an ‘*X*-gene’ genetic analysis, or XGA. Fundamentally, this is because we are extending approaches such as SGA to consider the effects of perturbing any number of (‘*X*’) genes at a time.

Given the present interest in both combinatorial variant studies, and in more complex genetic interactions, we believe that our manuscript will be of value to the broad readership of *Cell*. Given the advances required to enable XGA, we demonstrate this approach in the yeast model. However, the key ingredients – engineering many ‘*X*-gene’ knockouts and interpreting the effect – are widely applicable to many experimental organisms such as human cell lines. Several proposals for engineering combinatorial genetic perturbations have been already described, and it is even possible to couple these perturbations which single-cell phenotypes. In this regard, our demonstrations are modest in comparison to what is possible, but coupled with other technologies we think our study provides a stepping stone for future XGA.