Slide 2.

Biology is all about phenotypes. But what is a phenotype? Broadly speaking, a phenotype is any measurable trait associated with an organism.

Slide 3.

The history of biology is punctuated by the introduction of new phenotypes. For example, Beadle and Tatum used **auxotrophy** as a phenotype, but to use it effectively, they needed to learn to isolate and maintain auxotrophs. By developing the protocols to do this, they discovered that genes can be sequentially ordered into pathways, which in turn led them to propose that each gene codes for one enzyme.

Slide 4.

The history of biology is punctuated by the introduction of new, VISIBLE, phenotypes. However, under standard laboratory conditions, most genes do not have visible mutant phenotypes when we knock them out—they are invisible. That said, they may have unseen transcriptional consequences. If so, transcriptomes could be a powerful new phenotype with which to study such "invisible" genes and their interactions. However, it was not clear how to perform causal analyses using transcriptomes.

Slide 5.

How do we go about developing causal methodologies? To develop causal methodologies, we did not need to start from scratch—geneticists are experts at using phenotypes to determine causality.

Slide 6.
In fact, geneticists have a suite of methods to dissect genetic interactions and gene structure/function, and we adapted all of these analyses to be applied transcriptome-wide. Today, I will show you how we extended one of these methods to detect genetic interactions using transcriptomes as phenotypes.

Slide 7.
Consider the question: Given a set of genes, can we figure out the set of causal interactions between them? Moreover, if these genes act sequentially, can we order them into a pathway?

Slide 8.
Traditionally, genes are ordered along pathways using a phenomenon called Batesonian epistasis. For example, consider the following example where *C. elegans* has 2 genes, *a* and *b*. When I knock out *a*, the worm turns green. If I knock out *b*, there is no mutant phenotype. But now, suppose I go back to that green worm, and I delete *b* in that green worm. I might expect the worm to stay green, but actually, what happens is that the mutant phenotype goes away.

Instead, notice that the phenotypes of the *b* single and double mutants of are exactly the same. This phenotypic equality is called Batesonian epistasis. Knocking out *b* can suppress the phenotype associated with knocking out *a*.

Slide 9.

Inverting this logic, we can now conclude that *a* inhibits *b*.

Slide 10.

The reason we are allowed to draw this pathway is because epistasis occurs in networks when all paths leading from *a* to *c* go through *b*.

Slide 11.

I just showed you how we can do epistasis using binary phenotypes, but transcriptomes are continuous, not binary. Moreover, it's unclear how you do epistasis in multiple dimensions.

Slide 12.

My solution to this problem was to take my three multi-dimensional phenotypes and summarize their relationships into a single statistic, a single number. We can then use that statistic to check for Batesonian epistasis, and if it is, then we can build a pathway.

Slide 13.

So how do we compute this statistic? First, for each transcript we calculate an expected value for the double mutant by adding the single mutant log Fold Changes. Next, we compute the difference between the observed and expected values for each transcript. Finally, we will make a plot of difference versus expected for all transcripts. The points will form a line, and the slope of that line is our statistic.

Slide 14.

Let me give you an example. Here's some sample data I obtained from some single and double mutants. I have plotted the difference between the observed and expected values against the expected values, and I have calculated my statistic by finding the line of best fit.

Slide 15.

Here, the green curve shows my best guess for the value of the statistic. Next, we need to check whether this statistic is Batesonian.

Slide 16.

To check whether our statistic is Batesonian, we can simulate the values we would expect from our statistic by using only our single mutant data. And now, all we need to do is…

Slide 17.

… overlap our curves. Here, you can see that the observed slope matches exactly with one of the two Batesonian models. With this information, we could now write down a pathway, exactly like we would using classical epistasis analysis.

Slide 18.

With these methods, we blindly reconstructed the hypoxia pathway in *C. elegans* in a pilot project, showing that transcriptomes can serve as phenotypes for causal genetic analysis.

Slide 19.

And using these causal analyses, we discovered a new state in the life-cycle of *C. elegans*. As it happens, *C. elegans* hermaphrodites become female when they lose their endogenous sperm, and when this happens, their transcriptome, metabolome and their behaviors change.

Slide 20.

Moreover, our analyses are now fully automated and hosted on WormBase. Today, any nematode biologist can perform automated epistasis analyses using RNA-seq at the click of a button.

Slide 21.

All this time, we have been talking about epistasis, but I haven't really explained why epistasis happens, or why it works, or how we know what it means. This was the question I set out to answer in a brief postdoctoral stint with Matt Thomson at Caltech. As it turns out, genetics and probability theory are intimately related. In particular, I was able to prove that epistasis analyses are tests of conditional independence in probability theory.

Slide 22.

What does it mean for two things to be conditionally independent given a third variable? Mathematically, conditional independence is a way for us to say that all paths connecting *a* and *c* pass through *b*.

What does this mean genetically? Well, it means that the phenotype of the single mutant of *b* is identical to the phenotype of the double mutant *ab*. Mutants of genes in a pathway exhibit Batesonian epistasis.

Slide 23.

So why does this matter? From a really basic perspective, we now understand what genetics is doing. Epistasis analyses carry out a mathematical operation to reconstruct networks. Moreover, now we can translate between biology and mathematics, and we can use the language of mathematics to develop new algorithms for epistatic analysis. For example, using these discoveries we can develop algorithms that take as input single, double and triple mutants and integrate the data from these mutants to reconstruct networks automatically. By studying the properties of these algorithms, we can also arrive at a simple and powerful conclusion: Epistasis analyses are an extremely powerful way of dissecting complex networks.

Slide 24.

However, perhaps the most exciting aspect about this work is that it explains why any phenotype can be used for genetic analysis. I've shown you that we can think about molecules as phenotypes, but we can also use organismal phenotypes. We can use static phenotypes, such as the pigmentation of a rooster's crest, or we can use dynamic phenotypes such as the action potential of a neuron. And the reason for this is that conditional independence is scale-free. It doesn't care about any of these things, as long as they are causally linked to our pathway.

Slide 25.

The beautiful thing about epistasis, then, is that it can help us reconstruct pathways among individual genes.

Slide 26.

And thanks to my mathematical work, now we can generalize these analyses to study how groups of genes interact with one another.

Slide 27.

And what is a group of genes, if not an organism? Next, I want to study how bacteria interact with one another to form a community using the genetic principles I have shown you thus far.

Slide 28.

I want to know whether bacterial interactions in a community are dense, where every bacterial species interacts with every other bacterial species, or whether they are sparse, with only a few interactions per species.

Slide 29.

To do this, I will generate a platform to build customizable microbiomes. I will use a Bacterial Library that has already been generated by my postdoctoral laboratory to generate complex communities of at least 20 species. An eventual goal would be to generate *in vivo* customizable communities.

Slide 30.

With this platform, we will be able to carry out epistasis analyses to determine bacterial interactions. A particular advantage of these tests is that they circumvent the greatest obstacle in microbiology: The dearth of molecular tools with which to study most bacterial species. Though challenging, my implementations do not require genetically encoded markers or mutations.

Slide 31.

We will carry out these experiments, using 16S abundance as our initial phenotype, measured once the community has saturated.

Slide 32.

I am aware that these experiments are challenging, and to this end I will develop algorithms that can minimize the number of experiments we need to perform, while maximizing the information we gain from each experiment. A simple algorithm that could minimize bacteria is as follows:

Given a wild-type community, sort the names of all bacteria according to phylogenetic distance. Measure the wild-type abundances. Next, remove a subset of the bacteria and figure out if the community composition shifts significantly. If it does, we repeat this experiment, removing two smaller subsets. Finally, we look for epistasis. This top-down approach cuts down the number of experiments because it essentially implements a binary search.

Slide 33.

Biology is all about phenotypes, because phenotypes are a window into causality.