A future of the model organism model

Jasper Rine
Department of Molecular and Cell Biology and California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720-3220

ABSTRACT **Changes in technology are fundamentally reframing our concept of what constitutes a model organism. Nevertheless, research advances in the more traditional model organisms have enabled fresh and exciting opportunities for young scientists to establish new careers and offer the hope of comprehensive understanding of fundamental processes in life. New advances in translational research can be expected to heighten the importance of basic research in model organisms and expand opportunities. However, researchers must take special care and implement new resources to enable the newest members of the community to engage fully with the remarkable legacy of information in these fields.**

Monitoring Editor
Keith G. Kozminski
University of Virginia

Received: Aug 26, 2013
Revised: Dec 2, 2013
Accepted: Dec 20, 2013

INTRODUCTION

Technological advances in genome sequencing and editing are reshaping the common conception of a model organism. Even in the recent past, a fairly large and dedicated collection of research teams was needed to whip an organism into shape for the kinds of investigations that interest the readership of this and similar journals. With today's tools in hand, it is only a matter of time before even the Samoan pololo worm becomes the subject of a fascinating article in *Molecular Biology of the Cell*.

For this Perspective, I confine my comments to the more traditional meaning of a model organism, one that has nucleated a committed band of intrepid explorers seeking the answers to questions about biology that are favorably accessible for experimental investigation in that organism. For our tribe of scientists, mouse, maize, guinea pig, *Neurospora*, and *Drosophila* were early members of the pantheon, with T4, *Escherichia coli* and *Salmonella*, *Saccharomyces* and *Schizosaccharomyces*, *Caenorhabditis elegans*, *Arabidopsis*, *Zebrafish*, and others joining later. A critical mass of investigators working on the same organism was necessary to create enough common tools and reagents for them to achieve a satisfying measure of synergism. The most successful models have enjoyed a culture of resource and information sharing, typically inculcated by the senior founders and embraced by others.

DOI:10.1091/mbc.E12-10-0768

Address correspondence to: Jasper Rine (jrine@berkeley.edu).

Abbreviations used: CRISPR, clustered regularly interspaced short palindromic repeat; EMAP, epistatic miniarray profile; SELEX, systematic evolution of ligands by exponential enrichment; TALEN, transcription activator-like effector nuclease; ZFN, zinc-finger nuclease.

The difference between the science infrastructure that a new graduate student begins with today versus what was available when I began in 1975 is staggering. One can easily understand how a new student could get the impression that model organism research is moving to ever-more-refined details and that the greener pastures for establishing a new career are elsewhere. However, there are incredibly exciting emergent possibilities created by the model organism research communities that demand renewed emphasis and offer exceptional opportunities for distinguished and impactful careers. I focus on just a few of these emergent possibilities, with a bias on the organisms I know best.

THE POSSIBILITY OF COMPREHENSIVE BIOLOGY

Biomedical research has many imperfect and biased routes to discovery and no perfect one. Biochemists discover proteins biased by the abundance or activity of a protein in their assays. Geneticists discover genes biased by their size and mutability. Cell biologists often require assays that can work in crude extracts. We have routinely worked with known unknowns, and with the near certainty of unknown unknowns, as first articulated in a different context. Model organisms are our best hope for a comprehensive understanding of biology at any scale. Comprehensive knowledge about any aspect of biology can be as valuable for eliminating the unknowns, which can then be ignored, as it is for the facts that it establishes.

The description of the complete developmental lineage of *C. elegans* (Sulston and Horvitz, 1977; Sulston *et al.*, 1983) is a perfect example of a comprehensive biological understanding, and its elucidation will forever remain a heroic contribution to biology. Knowing what each cell does in development provides tremendous focus by removing doubt about the myriad things that a cell might have done were the lineage not complete. Moreover, knowing the pattern serves as the foundation for recognizing when cells do not follow it because of a mutation. Similarly, knowing the complete wiring

diagram of the *C. elegans* nervous system (White *et al.*, 1986) illuminates which cells can and cannot contribute to any given behavior. These examples also illustrate the impact that a small team of highly motivated individuals can make in producing a comprehensive description of a process, transforming the way that research in an organism is done. Although the value of hypothesis-driven science is undisputed, it is useful to consider that these two "descriptive" studies would have been hard to sell to some study sections.

We must keep in mind that "comprehensive" is defined by the context in which the data were comprehended. Complete genome sequences have been highly visible contributions to comprehensive biology by offering the potential of a complete parts list of proteins and RNAs for an organism. One can simply look at the genome sequence and deduce the number of, for example, protein kinases in an organism as a step toward figuring out which is responsible for phosphorylating a particular target. However, our genome annotation tools remain decidedly imperfect. We routinely fail to consider small open reading frames, not because they are unimportant, but because they are a statistical challenge for the annotation tools. Similarly, we find transcribed sequences only when we evaluate RNA populations under conditions in which the RNAs are expressed. In *Saccharomyces*, 16 years after the genome sequence was complete, a recent study of mRNA profiles in meiotic cells uncovered many previously unrecognized genes, some of which produce noncoding RNAs or use noncanonical start codons and hence were previously overlooked (Brar *et al.*, 2012).

Another premier example of comprehensive biology is the complete set of knockout mutations in *Saccharomyces cerevisiae* (Winzeler *et al.*, 1999). The importance of this collection vastly transcends the ease of having null alleles readily available. It is one thing to screen for a phenotype among 100,000 mutagenized colonies hoping to find a mutant. It is quite different when one needs to screen only a few thousand strains for a phenotype. In that case, the simplified logistics allows use of much more quantitative assays that detect contributions of genes that would be too subtle to detect when screening 100,000 mutagenized colonies, all of which would typically have 50–100 other mutations. As one measure of the difference between screening comprehensive collections of mutants versus the classic approaches, consider the genes involved in galactose metabolism in *Saccharomyces*. In decades of classic studies, approximately eight genes were found affecting galactose metabolism. Comprehensive screening of the knockout collection doubled that number with a single figure in a single paper (Giaever *et al.*, 2002). Highly parallel RNA interference–based assays allow systematic functional screens to be applied at scale to *C. elegans* (Fraser *et al.*, 2000).

Any complete collection of mutants enables more-sophisticated systematic studies. In *Saccharomyces* this resource has fueled systematic double-mutant analyses, first with synthetic genetic arrays (Tong *et al.*, 2001) and later with epistatic miniarray profiles (EMAPs), inspiring similar analyses with small interfering RNAs in other organisms (e.g., Tischler *et al.*, 2006). The interaction maps that result from such systematic double-mutant analyses (e.g., Collins *et al.*, 2007; Costanzo *et al.*, 2010), jokingly referred to as the hairballs of genetics, have had tremendous, and I think unexpected, predictive value for inferring the function of unstudied or understudied proteins by the transitive logic of synthetic interactions. Few of us imagined that the network of interactions between, for example, yeast and human would be sufficiently conserved that modules of interactions among genes involved in cell wall biosynthesis in yeast would predict the function of "phenologues" of angiogenesis in humans (McGary *et al.*, 2010).

SYNTHETIC LETHALITY, THE MYTH OF REDUNDANCY, AND CANCER

Redundancy, the unicorn of genetics, is a common misconception applied to cases in which a gene can be eliminated with no apparent effect on an organism. With the exception of very recent gene duplications and possibly a few theoretical contexts (Nowak *et al.*, 1997), redundancy cannot be selected for. Hence, functional genes all provide some unique benefit to the organism in which they reside. Our perception of redundancy is, most often, a reflection of our limited ability to discern phenotype. Model organisms are the only contexts in which we have any reasonable hope of uncovering the factors that give the appearance of redundancy.

Synthetic genetic interactions are proving to be a powerful way of uncovering phenotype by asking which mutant gene, when paired with another mutant gene, causes a phenotype or, in the limit case, lethality. As described earlier, synthetic-lethal phenotypes have proven to be abundant in the interaction networks of model organisms. Because cancer cells typically contain many mutations, it is possible that they are therefore vulnerable to other mutations in ways that normal cells are not. If so, one could imagine that drugs that target the synthetic-lethal partners of a tumor suppressor gene mutation would have special impact on cancer cells. Indeed, recent work established both the feasibility of this approach and the predictive power of synthetic-lethal interactions from model organisms as applied to human cells (O'Neil *et al.*, 2013; Pourdehnad *et al.*, 2013; van Pel *et al.*, 2013).

THE IMPORTANCE OF ISOGENIC COMPARISONS, AND THE IMPORTANCE OF GETTING OVER ISOGENICITY

One of the strengths of model organism research is the luxury of a defined wild type against which other isolates and mutants can be compared. A defined wild type, combined with the power of isogenic comparisons, allows the genetic foundation for even the subtlest of phenotypes to be identified from the pattern of phenotypic segregation in a cross. However, model organisms have also proven their value in characterizing the patterns and principles of genetic variation in wild species. For example, a study of why some genes are essential in certain strains of *S. cerevisiae* but not in another revealed an unexpectedly complex genetic structure, which casts a sobering light on the limits of genome-wide association studies of humans (Dowell *et al.*, 2010). On the flip side, human genome-wide association studies ran into an apparent roadblock, dubbed the "mystery of the missing heritability," reflecting the tendency of genetic associations to account for only a fraction of the phenotypic variance in a population. Studies in yeast and *C. elegans* largely resolved the mystery by revealing the complexity of the genetic architecture governing complex traits (e.g., Gaertner *et al.*, 2012; Bloom *et al.*, 2013). With denser markers and better phenotyping, the mystery largely disappears.

WHAT ARE WE MISSING?

Given the daunting nature of the literature of model organisms, it is easy to lose sight of how little we know about fundamental aspects of biology, including some of the most thoroughly studied topics in model organisms. An understanding of what we do not know is indispensable to motivating the development of technologies to push the frontiers. The remarkable advances in zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and now clustered regularly interspaced short palindromic repeats (CRISPRs) put targeted genome editing solidly in the genetic tool box of all organism researchers. However, for the foreseeable future, we will be limited by our ability to describe phenotype more

precisely. The lack of precision in phenotyping is important because efforts to find the genes responsible for phenotypic differences are most successful when the phenotype can be narrowly defined or broken down into multiple, well-defined components.

Here are some of my favorite areas in which better phenotyping would make a big impact. I invite readers to add their favorites to the list.

1) The Yeastnet database lists 1394 nonspontaneous reactions in yeast, of which 622 are not yet assigned to an enzyme (Herrgård *et al.*, 2008)—and you thought the gospels of Stryer and Lehninger had complete descriptions of metabolism! As one example, an entire new pathway in central metabolism, riboneogenesis was recently discovered by the use of advanced small-molecule mass spectrometry (Clasquin *et al.*, 2011). There are sure to be many new aspects of metabolism to be discovered in model organisms.
2) We need better ways of recognizing differences between cells that look alike by conventional assays. Immunologists lead this game, and we would be well served to develop markers capable of distinguishing cell types to better resolve developmental processes and signaling networks. I was rather stunned by the recent discovery of a new cell type in yeast—the quiescent cell—which had been overlooked by the thousands of people who have studied yeast over the years (Aragon *et al.*, 2008). How many more such examples are there?
3) The lab is not the environment in which our model organisms evolved. Hence, we are blind to many influences that molded the genome of our models. For example, at the beginning of a wine fermentation, yeast metabolism begins in a concentration of sugar that is 12 times higher than is used in any lab and soon becomes hypoxoic. Yet not a single genome-wide double-mutant analysis has, to my knowledge, been conducted under these conditions. similarly, most of what we know about metabolism in *E. coli* was learned from highly oxygenated, mid log cultures grown on Luria broth. Yet our gut is nearly anaerobic and decidedly lacking in Luria broth. There is much to learn by increasing the range of conditions under which we study our model organisms to get better measures of phenotype.
4) tRNAs enjoy a wide variety of posttranscriptional chemical modifications influencing their function, and, to a lesser extent, rRNA does as well. The recognition of the diversity of RNA species, including the abundance of noncoding RNAs in all species examined, has expanded the field substantially. Yet what do we know about chemical modifications on mRNA or any of the newly discovered noncoding RNAs? Could nature possibly have excluded these RNAs with their capacity to form structures every bit as complex as tRNAs from the benefits of base modifications? I suppose it is possible, but this lacuna in our knowledge more likely stems from the challenge in purifying homogeneous samples of a single mRNA species. With the ability of systematic evolution of ligands by exponential enrichment (SELEX) to allow selection of RNA motifs with tight binding affinities for ligands, it is now possible to imagine "epitope tagging" an mRNA for affinity purification and subsequent analysis of its bases. Next-generation sequencing approaches also look promising for identifying at least some modified bases in mRNA (Edelheit *et al.*, 2013).

TRANSLATIONAL OPPORTUNITIES

The expanding focus on translational research by funders has not escaped the notice of those of us working on model organisms. Anyone with an appreciation of the scientific roots of, for example, the entire biotechnology industry will appreciate that the enabling breakthroughs came through curiosity-driven basic research in model organisms, with no idea of a practical application in mind when the research began. It is absolutely critical that we not lose sight of this indispensable value of curiosity-driven research in model organisms and work hard to protect such opportunities from the vicissitudes of manner in science funding.

However, it is equally important that we continue to make the point that model organism research is a critical and efficient path forward on some of the most critical translational challenges of the age. For the sake of brevity I restrict my translational discussion to three of these challenges: the critical need for global food security, interpreting the impact of human gene sequence variation, and the soon-to-appear rush of new human disease genes.

CLIMATE CHANGE AND GLOBAL FOOD SECURITY

There are ~7 billion people in the world, ~1 billion of whom receive inadequate calories on a regular basis. The United Nations projects that an additional billion people will need be fed by 2050 with little if any increase in farmable land (Food and Agriculture Organization of the United Nations, 2009). Setting aside the debate on the cause of climate change, the need for more food will be an especially difficult struggle when weather extremes threaten the yield of vital crops. Historically, plant breeding has had a primary focus on improved yield. More recently, yield stability—the ability to produce a good yield under a variety of conditions—has been recognized as a more critical trait than yield per se. In the future, robustness to extreme weather fluctuations is likely to be even more important. So how does model organism research fit in to this picture?

At the biological level, climate change will challenge plants with extremes of temperature, dessication, salinity, flooding, and probably of pathogen burden. Genes have been found in *Saccharomyces* that cause a 10^5-fold difference in dessication resistance between wild type and strains lacking these genes (Calahan *et al.*, 2011), illustrating how the fundamental biology of environmental stress is accessible in the lab in model organisms. There is already compelling evidence of the value of genes from model organisms in improving the quality of transgenic crops. For example, cold shock protein B (cspB), an RNA chaperone, of *Bacillus subtilis* substantially improves the yield of transgenic corn carrying that gene when grown under limited water conditions (Castiglioni *et al.*, 2008).

Transcription factor genes may have unusually strong conservation of function among plants and hence potential for quickly affecting crop yields. For example, systematic overexpression of genes for transcription factors in *Arabidopsis* has been done on a scale that allows these many transgenic lines to be screened for traits of agronomic interest. In multiple cases, when the native orthologue of that gene is expressed in corn, rice, or soybean, the transgenic crops enjoy the same quantitative improvement predicted from the *Arabidopsis* model (Nelson *et al.*, 2007; Preuss *et al.*, 2012; Yu *et al.*, 2013). Such approaches can take many years off the development cycle for new crop strains.

Were I able to wave a magic funding wand, I would think hard about a substantial investment in the fundamental biology of stress and mechanisms of stress resistance. The need is obvious, and the foregoing examples convince me that there are ways that discoveries from basic research in this area would find rapid utility.

MODEL ORGANISMS AS INTERPRETERS OF GENE SEQUENCE VARIATION

The speed at which the cost of human exome sequencing has dropped underscores the need for work on model organisms to

interpret the effect of the variants we all carry. In my view, the variants in the coding sequence, although smaller in number than the total variants in a human genome, are the sweet spot for having an effect on human health because medicines target proteins, not noncoding sequences. Among the 3 million or so single-nucleotide polymorphisms that differ between any two people, ~10,000 will cause a change in an amino acid in a protein. Sequence comparisons of human genes with those of model organisms have been used to identify the conserved regions of genes, a technique that offers some guide to predicting the effect of a change. There are, or course limitations, such as the 2% of human disease alleles that are identical to the sequence of the corresponding position of orthologous gene of mouse (cited in Ng and Henikoff, 2006).

Surrogate genetics—the testing of a gene from one organism in the context of a model organism—provides a better, although still imperfect assessment for situations in which the human gene can complement the function of the orthologue in the model organism or cause a distinct phenotype. This approach has proven remarkably effective, even in yeast, on proteins ranging from metabolic enzymes (e.g., Mayfield *et al.*, 2012) to p53 (Kato *et al.*, 2003). Thousands of variants in thousands of human genes could be tested for function in model organisms.

THE FLOOD TIDE OF HUMAN DISEASE GENES

I have been deeply impressed by the ability of whole-exome sequencing of a handful of related individuals to pinpoint the causal mutations in rare medical conditions with Mendelian inheritance (1000 Genomes Project Consortium *et al.*, 2012; Tennessen *et al.*, 2012). For these diseases, there would never be an opportunity to assemble the population sizes needed for conventional gene mapping or association studies, yet exome sequencing has been successful. I foresee enormous progress on this front in a very short time, 2–3 years, as many of the 6000–8000 rare inherited diseases (http://raregenomics.org/) yield information on the gene affected by the causal mutation. The pace of discovery of new rare-disease-causing genes by exome sequencing jumped from ~30 in 2011 to 140 in 2012 and is still accelerating (Boycott *et al.*, 2013) It is reasonable to expect that many of these genes will have obvious orthologues in model organisms, yet with no obvious clue from human biology about the mechanism of disease. Racers, on the mark, get set,….

Will we have the ability to deliver on this challenge with the opportunities available in the model organism community? Here is one way to think about the scale of the challenge. Imagine that the gene behind 3000 of these diseases is largely unstudied, and an appropriate model organism exists for studying the biology of that gene. If there were one modular 4-year R-01 style National Institutes of Health (NIH) grant funded for each gene, we could expect an enormous amount of progress for on the order of the 3 billion dollars, about the price of the human genome project, or, put another way, ~10% of the annual NIH budget. Because these genes will not all be discovered at once, the cost would be a few percent of the budget per year. A broad-based attack on thousands of diseases that leverages the amazing opportunities in model organism research is feasible and affordable. Our community needs to be relentless in pushing this agenda. The alternative approach would be another big-science initiative described with easy-to-understand sound bites. Big-science projects have their place when the path forward is clear, the goal easily defined, and the chief challenge being one of scale. The challenge of understanding the biology behind the thousands of new disease genes soon to be discovered fits the R-01 model best.

CLOSING THOUGHTS

When I was a graduate student in the Herskowitz lab at the University of Oregon, half of the lab worked on aspects of lambda gene regulation, and the rest of us worked on yeast. My experience with lambda made two lasting impressions. First, the science was intensely interesting and had a legacy of literature produced by some of the finest scientists of the day. Second, the level of arcane nomenclature that one had to master as the ante for doing even a simple experiment was so daunting that I chose the yeast option over lambda simply because there was little literature on yeast by today's standards. Hence there was at least the illusion that one could master the literature relevant to a research question. I worry that students entering a model organism lab today face a situation not so different from my experience with the lambdology of yesteryear.

We have done a better job, I hope, in keeping the jargon to a minimum. Nevertheless, all of us who teach beginning graduate students recognize that the language and culture of our fields are as challenging to understand as the science itself. Under the leadership of Mark Johnston, Editor-in-Chief of *Genetics*, and Alan Hinnebush, prominent figures of the yeast field have published chapters in what is known as *The Yeast Book*, appearing regularly in the pages of *Genetics*. These chapters have proven to be a magnificent entrée to the field and a worthy successor to the Cold Spring Harbor monographs with which many of us grew up. I encourage all model organism communities to consider this model for helping the newest members of their fields as they approach a critical size.

Finally, although the modern scientific literature is richer than ever before, I know no one who feels truly on top of the literature in his or her field. Dedicated model organism databases address this problem, but in a limited way. The *Saccharomyces* Genome Database is an indispensable part of everyday activity in my lab. We would be crippled without it. The team that maintains this resource is second to none among those working on model organism databases. However, some of the content in such databases is one or more steps away from the primary data and reflects the informed interpretation of a smart and experienced curator. Yet there is never a substitute for knowing the primary data in one's field, and there is no hope of knowing all of the primary data. This is the existential dilemma of modern model organism research. My way to deal with the dilemma is to acknowledge it and struggle to embrace as much of the primary data as I can. Immersion in the lively and interactive teams that characterize the majority of model organism research labs is one way to approach the struggle, and surely the most fun.

ACKNOWLEDGMENTS

Research in my lab has been supported continuously by the National Institutes of Health for 32 years, with additional support from the Defense Advanced Research Projects Agency, the Army Office of Research, and predoctoral and postdoctoral fellowships from the National Institutes of Health, the National Science Foundation, the American Cancer Society, the Damon Runyon Cancer Fund, the Helen Hay Whitney Foundation, the Jane Coffin Childs Foundation, and the Miller Institute. I thank Cynthia Kenyon and two reviewers for helpful comments on the manuscript.

REFERENCES

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65.

Aragon AD, Rodriguez AL, Meirelles O, Roy S, Davidson GS, Tapia PH, Allen C, Joe R, Benn D, Werner-Washburne M (2008). Characterization of differentiated quiescent and nonquiescent cells in yeast stationary-phase cultures. Mol Biol Cell 19, 1271–1280.

Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L (2013). Finding the sources of missing heritability in a yeast cross. Nature 2494, 234–237.

Boycott KM, Vanstone MR, Bulman DE, Mackenzie AE (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 14, 681–691.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335, 552–557.

Calahan D, Dunham M, DeSevo C, Koshland DE (2011). Genetic analysis of dessication tolerance in *Saccharomyces cerevisiae*. Genetics 189, 507–519.

Castiglioni P *et al*. (2008). Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. Plant Physiol 147, 446–455.

Clasquin MF *et al*. (2011). Riboneogenesis in yeast. Cell 145, 969–980.

Collins SR *et al*. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446, 806–810.

Costanzo M *et al*. (2010). The genetic landscape of a cell. Science 327, 425–431.

Dowell RD *et al*. (2010). Genotype to phenotype: a complex problem. Science 328, 469.

Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, Archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet 9, e1003602.

Food and Agriculture Organization of the United Nations (2009). How to Feed the World in 2050. Available at http://www.fao.org/wsfs/forum2050/en/.

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature 408, 325–330.

Gaertner BE, Parmenter MD, Rockman MV, Kruglyak L, Phillips PC (2012). More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *Caenorhabditis elegans*. Genetics 192, 1533–1542.

Giaever G *et al*. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418, 387–391.

Herrgård MJ *et al*. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26, 1155–1160.

Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C (2003). Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci USA 105, 8055–8060.

Mayfield JA *et al*. (2012). Surrogate genetics and metabolic profiling for characterization of human disease alleles. Genetics 192, 1309–1323.

McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc Natl Acad Sci USA 107, 6544–6549.

Nelson DE *et al*. (2007). Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104, 16450–16455.

Ng PC, Henikoff S (2006). Predicting the effects of amino acid substitutions on protein function. Ann Rev Genomics Hum Genet 7, 61–80.

Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997). Evolution of genetic redundancy. Nature 388, 167–171.

O'Neil NJ, van Pel DM, Hieter P (2013). Synthetic lethality and cancer: cohesin and PARP at the replication fork. Trends Genet 29, 290–297.

Pourdehnad M, Truitt ML, Siddiqi IN, Ducker GS, Shokat KM, Ruggero D (2013). Myc and mTOR converge on a common node in protein synthesis control that confers synthetic lethality in Myc-driven cancers. Proc Natl Acad Sci USA 110, 11988–11993.

Preuss SB *et al*. (2012). Expression of the *Arabidopsis thaliana BBX32* gene in soybean increases grain yield. PLoS One 7, e30717.

Sulston JE, Horvitz HR (1977). Post-embryonic cell lineages of the nematode. *Caenorhabditis elegans*. Dev Biol 56, 110–156.

Sulston JE, Schierenberg E, White JG, Thomson JN (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol 100, 64–119.

Tennessen JA *et al*. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337, 64–69.

Tischler J, Lehner B, Chen N, Fraser AG (2006). Combinatorial RNA interference in *Caenorhabditis elegans* reveal that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol 7, R69.

Tong AH *et al*. (2001). Systematic genetic analysis with an ordered array of yeast deletion mutants. Science 294, 2364–2368.

van Pel DM, Stirling PC, Minaker SW, Sipahimalani P, Hieter P (2013). *Saccharomyces cerevisiae* genetics predicts candidate therapeutic genetic interactions at the mammalian replication fork. G3 (Bethesda) 3, 273–282.

White JG, Southgate E, Thomson JN, Brenner S (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. Philos Trans R Soc Lond B Biol Sci 314, 1–340.

Winzeler EA *et al*. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285, 901–906.

Yu L, Chen X, Wang Z, Wang S, Wang Y, Zhu Q, Li S, Xiang C (2013). *Arabidopsis Enhanced Drought Tolerance1/HOMEODOMAIN GLABROUS11* confers drought tolerance in transgenic rice without yield penalty. Plant Physiol 162, 1378–1391.