Bringing genetic background into focus

Clement Y. Chow

Researchers should embrace differences in genetic background to build richer disease models that more accurately reflect the level of variation in the human population, posits Clement Chow.

> "the effect of genetic background is still treated as a nuisance to be avoided"

Department of Human Genetics, University of Utah School of Medicine, Room 5150, 15 North 2030 East, Salt Lake City, Utah 84112, USA.
cchow@genetics.utah.edu

doi:10.1038/nrg.2015.9
Published online 14 Dec 2015

At the heart of the U.S. National Institutes of Health (NIH) Precision Medicine Initiative (PMI) is the understanding that each patient is an individual and disease outcomes vary greatly between patients. Most Mendelian and complex diseases are influenced by a variety of factors, not the least of which is genetic background. To fulfil the goals of the PMI, we need to understand how genetic background interacts with other factors to produce differential susceptibility to disease. However, most model organism studies are conducted with a reductionist approach, on single-strain backgrounds, to simplify analysis. Recent advances in model organism biology now make it feasible to ask questions and test hypotheses regarding the role of genetic background on disease outcomes. We need to embrace background genetic variation as an opportunity to understand the phenotypic spectrum and genetic architecture of disease.

Genetic background and modifier genes

Many genetic diseases are caused by one (or a few) primary genetic variants, but the background genetic milieu of each individual can have profound effects on that primary disease. One of the best examples of this is the cystic fibrosis transmembrane conductance regulator (CFTR) variant p.Phe508del (also known as F508del in legacy nomenclature), the most common cause of cystic fibrosis[1]. Individuals homozygous for F508del present with a wide range of phenotypes and can differ in disease severity. Recent studies have identified modifier genes in these patients that drive some of this phenotypic variability[1].

Identification of modifier genes in humans is a difficult process that is often hampered by vast differences in environmental factors. Model organisms, however, provide a controlled, systematic approach to studying variation. In fact, genetic modifiers have a strong history in model organisms, especially in mouse genetics. Some of the most seminal work in modifier genes comes from the recognition that a particular mouse mutant, when crossed onto two different inbred strains, can have drastically different phenotypic outcomes. For example, discovery of one of the first neurological disease modifier genes came from the recognition that, on the C3H mouse strain, the medJ splicing mutation in the sodium channel gene *Scn8a* results in a severe movement disorder but normal lifespan[2]. However, on the C57BL/6 strain, the mice display progressive paralysis and lethality by 4 weeks of age. Positional cloning analysis identified the modifier as a premature stop codon in the splicing factor SCNM1 (REF. 2).

Both of these examples nicely illustrate that a 'simple' Mendelian disease is quite complex when genetic background is considered. However, these two examples are also relatively straightforward, whereas most genetic variation is complex and involves many modifier genes. Knowledge of this phenomenon is widespread, but the effect of genetic background is still treated as a nuisance to be avoided.

Model organisms and genetic background

Our disease models need to more closely reflect what is found in the human population — including the wide range of genetic variation. We need tractable systems to not only document genetic variation but also determine the functional consequences of such variation. There are two main ways to study genetic variation in model organisms. The classic approach is mutagenesis, wherein a mutagen is used to semi-randomly generate a known frequency of mutations throughout the genome. This approach is often done on a sensitized background that carries a primary mutation of interest, and subsequent mapping or sequencing enables the identification of mutations that segregate with a modified phenotype. This method, which has been commonly used in *Drosophila melanogaster*, *Caenorhabditis elegans* and mice, has been invaluable in identifying new pathway elements and interacting partners for a particular disease gene. However, these mutagenesis studies often produce very specific kinds of mutations, and commonly used methods tend to enrich for mutations with large effects. It is unclear whether these mutations reflect the standing variation found in populations. For example, many mutations identified by mutagenesis have detrimental phenotypes that would preclude their persistence in a wild population.

> every phenotype, be it simple or complex, is highly variable and strongly influenced by genetic background

Another way to study genetic variation is to examine natural variation within a population. Genetic variation found in 'wild-type' strains of model organisms is not necessarily meant to reflect the human population, but it provides the opportunity to model the effects of genetic variation in a controlled manner. Because this variation is generally compatible with life, we expect each variant to be of smaller effect size, which, in turn, may reflect differences in genetic network architecture. Variants of small effect can produce genetic networks that are very different from those of large-effect mutations generated by mutagenesis.

In *D. melanogaster*, natural variation can be captured in collections of hundreds of inbred lines derived from individuals in a natural population, such that each homozygous strain contains a different genome sampled from that population. Some of these panels include, but are not limited to, the *Drosophila* Genetic Reference Panel (DGRP) and the *Drosophila* Population Genomics Project (DPGP)[3]. Each resource has different advantages, but they all capture genetic variation that was previously unavailable to the *Drosophila* research community.

The study of genetic variation in mice is a larger, more challenging prospect, owing to the complicated relationship between wild mouse populations and the laboratory strains established over the past century. The mouse Collaborative Cross (CC)[4], for example, starting with eight diverse mouse strains, has generated ~300 new recombinant inbred strains that shuffle the original genomes. Another related resource, the Diversity Outbred stock[4], was generated from the CC as an advanced mapping, outbred population. Again, these two resources are unique in different ways, but they provide the mouse community with powerful new tools to probe genetic variation.

These resources have been employed in many different ways. We are using natural genetic variation to identify novel Mendelian disease modifiers. In a recently completed study, we crossed an autosomal-dominant fly model of retinal degeneration onto ~200 strains of the DGRP. We found that the disease phenotype was essentially a quantitative trait: nearly every background conferred a slightly different severity[5], similar to what might be expected in the human population. The quantitative nature of the phenotype also allowed us to nominate new candidate modifier genes. In a separate study, we challenged the parental strains of the mouse CC with endoplasmic reticulum stress, which is a very basic disease-related cellular stress, and observed that the response in each strain differs: some genetic backgrounds use certain aspects of the response network more than others[6]. In a third study by Rasmussen *et al.*, the mouse CC was even used to demonstrate the extensive phenotypic and genetic variation in Ebola infection outcomes[7]. Many other studies have been conducted in the DGRP and CC resources, examining the genetic variation of phenotypes ranging from metabolism to cancer to behaviour. The general lesson from all of these studies is that every phenotype, be it simple or complex, is highly variable and strongly influenced by genetic background.

Moving forward

To study a particular mutation or pathway on many (hundreds) of genetic backgrounds is not a trivial endeavour. This type of analysis takes much effort and statistical consideration. Both the DGRP and CC communities have made analytical tools available to simplify some of the downstream analysis. This type of thinking, that genetic background is as important as the primary mutation, is not new, but we are in an era of biology where computational, sequencing and analytical tools have converged to make these types of studies possible.

While documenting genetic and phenotypic diversity can reveal novel insights into biological pathways, the larger goal is to bring this knowledge back to the clinic. Studies utilizing single genetic backgrounds are incredibly informative, but we risk missing the nuances that make treating patients challenging. Incorporating genetic variation into a model organism study might help to prioritize findings that are relevant to differences observed in patients. This is especially true with the increasing availability of thousands of whole-genome and whole-exome sequences. We need better ways of prioritizing potential disease-causing variants. Studying genetic variation in model organisms might be one tool to help us to better predict pathogenic variants. Thus, we need to embrace genetic background differences and use them to build richer disease models that more accurately reflect the level of variation in the human population.

1. Cutting, G. R. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat. Rev. Genet.* **16**, 45–56 (2015).
2. Buchner, D. A., Trudeau, M. & Meisler, M. H. SCNM1, a putative RNA splicing factor that modifies disease severity in mice. *Science* **301**, 967–969 (2003).
3. Lack, J. B.*et al.* The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241 (2015).
4. Threadgill, D. W. & Churchill, G. A. Ten years of the Collaborative Cross. *Genetics* **190**, 291–294 (2012).
5. Chow, C. Y., Kelsey, K., Wolfner, M. F. & Clark, A. G. Candidate genetic modifiers of retinitis pigmentosa identified by exploiting natural variation in *Drosophila*. *Hum. Mol. Gen.* (in press)
6. Chow, C. Y. *et al.* The genetic architecture of the genome-wide transcriptional response to ER stress in the mouse. *PLoS Genet.* **11** (2015).
7. Rasmussen, A. L. *et al.* Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science* **346**, 987–991 (2014).

Acknowledgements
The author would like to thank Charles Murtaugh, Carl Thummel and Brooke LaFlamme for helpful comments.

Competing interests statement
The author declares no competing interests.

FURTHER INFORMATION
Drosophila Genetic Reference Panel (DGRP): http://dgrp2.gnets.ncsu.edu/
Drosophila Population Genomics Project (DPGP): http://www.dpgp.org/
Collaborative Cross (CC): http://compgen.unc.edu/wp/?page_id=99
The Diversity Outbred (DO): http://churchill.jax.org/research/cc/doresources.shtml

ALL LINKS ARE ACTIVE IN THE ONLINE PDF