Morgan's Legacy: Fruit Flies and the Functional Annotation of Conserved Genes

Hugo J. Bellen[1,2,3,4,5,*] and Shinya Yamamoto[1,2,4]
[1]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas 77030, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA
[3]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA
[4]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA
[5]Howard Hughes Medical Institute (HHMI), Houston, TX, 77030, USA
*Correspondence: hbellen@bcm.edu
http://dx.doi.org/10.1016/j.cell.2015.09.009

In 1915, "The Mechanism of Mendelian Heredity" was published by four prominent *Drosophila* geneticists. They discovered that genes form linkage groups on chromosomes inherited in a Mendelian fashion and laid the genetic foundation that promoted *Drosophila* as a model organism. Flies continue to offer great opportunities, including studies in the field of functional genomics.

This year we celebrate the 100th anniversary of the publication of the book "The Mechanism of Mendelian Heredity" by Thomas H. Morgan, Alfred H. Sturtevant, Hermann J. Muller, and Calvin B. Bridges (Morgan et al., 1915). The work published by these four giants in the *Drosophila* field was the most influential scientific work in the field of genetics since Gregor Mendel's work in 1866. Although the achievements of Mendel were ignored in the 19th century, the rediscovery of Mendel's law in 1900 led to the foundation of the field of genetics. Morgan, who initiated his work on *Drosophila* in 1909, was an embryologist who became attracted to flies because of the discovery of genetic variants. Interestingly, in his early career (1900–1910), Morgan was critical of the Mendelian theory of heredity and skeptical of the fact that species arise by natural selection as postulated by Charles Darwin. Moreover, in his acceptance speech for the Nobel Prize of 1933, he downplayed the contribution of *Drosophila* research to human biology and medicine with one exception: genetic counseling. Morgan quickly changed his mind and became an advocate of Mendel's and Darwin's work, while researchers later showed that he was overly modest about the implications of *Drosophila* research on human biology.

Drosophila Research in the 20th Century

Morgan initiated his work on *Drosophila* in 1909 at Columbia University. He quickly attracted a set of superb scientists, and together, they elegantly documented many of the basic tenets of genetics, discovering that factors (now known as alleles of genes) form linkage groups, and that these linkage groups exhibited the same inheritance pattern as the chromosomes to which they mapped. Experimental data with mutants that map to sex chromosomes in *Drosophila* provided the central support for their hypothesis that genes are independent physical entities present in a linear array on chromosomes that follow Mendel's law of independent segregation. They concluded their book by stating that: "Although Mendel's law does not explain the phenomena of development, and does not pretend to explain them, it stands as a scientific explanation of heredity, because it fulfills all the requirements of any causal explanation" (Morgan et al., 1915). Despite the criticism toward Mendel's work—that he had ignored or failed to report data that did not support his hypothesis—Morgan and colleagues gave Mendel the proper credit for discovering the principles of heredity, as is obvious from this statement as well as from the title of their book.

Muller, Sturtevant, and Bridges as well as other fly geneticists continued to perform experiments that laid the basis of much of eukaryotic genetics between 1910 and 1940. Muller developed the first balancer chromosomes which allowed him to discover that X-rays are mutagenic (Muller, 1927), for which he was awarded the Nobel Prize in 1946. Balancer chromosomes are still the most elegant means of preventing the exchange of genetic information between two homologous chromosomes, thereby giving researchers an efficient method to maintain thousands of recessive lethal and sterile stocks without the need of molecular genotyping. Sturtevant demonstrated that the Bar eye phenotype is caused by unequal crossover, a phenomenon which plays an important role in the generation of small chromosomal duplications and deletions linked to human diseases (Lupski et al., 1996). Bridges constructed the first physical map of chromosomes for any organism by describing the banding pattern of the polytene chromosomes in the salivary gland of flies and provided a physical map of genes on the banded chromosomes (Bridges, 1935). Bridges' work demonstrated the correlation between the physical structure of chromosomes and genetically defined linkage groups.

Drosophila research lost prominence in the 1940s as phages and bacteria dominated the field of genetics. However, a rebirth occurred in the early 1970s as two fields, neuroscience and developmental biology, converged onto *Drosophila* research. This resurgence was in part because of the reagents created by the founders, the availability of many mutations affecting numerous traits, and the ability to efficiently create new mutations (Lewis and Bacher, 1968). Indeed, no higher eukaryotic model

organism in the seventies had the tools that allowed the manipulation of genes as elegantly and probingly as in *Drosophila*.

The use of *Drosophila* as a model organism for neuroscience and developmental biology led to discoveries that provided a lasting impact. Seymour Benzer and colleagues studied genes affecting visual behavior, olfaction, sexual behavior, learning and memory, diurnal rhythms, aging, and neurodegeneration (Jan and Jan, 2008). Their work led to the discovery of numerous important genes and proteins such as the first potassium and transient receptor potential (TRP) channels, key circadian clock genes, and genes required for learning and memory. Similarly, in 1978 Christiane Nüsslein-Volhard and Eric Wieschaus decided to pursue a systematic genetic strategy to screen for mutants that affect the development of the embryo pattern and discovered many of the genes that are key players of developmental signaling pathways such as Notch, Wnt, Hedgehog, TGF-β/BMP, and Toll/TLR (Nüsslein-Volhard and Wieschaus, 1980). The impact of these discoveries have permeated almost every area of biology, including medical genetics and cancer biology (Wangler et al., 2015).

The ability to manipulate the *Drosophila* genome was bolstered tremendously by the technology to introduce any type of DNA into the fly genome using *P*-element-mediated transposition (Rubin and Spradling, 1982). Since then numerous technologies have been developed that allow extensive biological and genetic manipulation (Perrimon, 2014). The ability to manipulate the fly genome has enabled numerous scientists to contribute significantly to almost all areas of biology, including genetics, developmental biology, cell biology, neuroscience, physiology and metabolism, disease mechanisms, population genetics, and evolution.

Drosophila as a Model System for In Vivo Functional Genomics

The breadth of tools that have been developed and that are shared among the members of the fly community, in the tradition of the founders, permits sophisticated experiments that can be performed in very few model organisms. For example, these tools are being used to tease apart neuronal networks, assess and control specific behaviors, determine gene function in specific cells, and study physiological functions of proteins and metabolites. An area that has expanded significantly in the past 10 years is the study of fly genes whose human homologs cause genetic disorders. These studies attempt to better understand the basic biology of these genes and products, and attempt to probe the mechanism by which specific mutations cause pathological phenomena such as neurodegeneration (Jaiswal et al., 2012). Approximately 60% of the ~13,000 protein coding fly genes are evolutionarily conserved in human, yet, a functional annotation of most of these genes is still lacking (Yamamoto et al., 2014). Better and more detailed annotations of function and expression of thousands of *Drosophila* genes would help not only to better understand fly biology, but also to functionally annotate the human genome. Here, we will expand on some recently developed strategies that aim at providing functional data on fly genes and their expression patterns. These strategies also attempt to assess the function of human genes and provide data about the pathogenic impact of human mutations or variants.

In his 2015 State of the Union Address, President Obama announced the launch of the "Precision Medicine Initiative," with the ultimate goal of improving medical care by providing individuals with tailor-made prevention and treatment strategies. Due to the resources generated through the human genome project and the recent advances in sequencing technology and bioinformatics, human geneticists can quickly identify the majority of the polymorphisms and variants in a personal genome. The real challenge in precision medicine, however, is the interpretation of such genomic data. Our ability to extract meaningful data from whole-exome sequencing data is dampened by the existence of numerous rare variants of uncertain/unknown significance and, more importantly, by the lack of in vivo functional information of the majority of human genes. Hence, high-throughput strategies to quickly assess whether or not a variant of interest have functional effects is in high demand.

Although functional information can be obtained using cultured human cells, such as iPSCs, these experiments do not provide in vivo information. *Drosophila* is an ideal model organism to fill this niche, thanks to its short-life cycle, low maintenance costs, conserved biology, and powerful genetic toolbox.

Functional annotation of genes is typically done one by one, with individual laboratories devoting years to study the role of one or a few genes in a specific biological process or pathway. As most genes are also pleiotropic, different labs often study the same genes in different processes. This level of annotation has been the mainstay and the foundation of success of *Drosophila* research. In addition to this detailed level of gene characterization, cursory but rapid function examination of conserved genes in *Drosophila* can also provide important data to fill the gap between genetic and phenotypic information.

A cursory functional annotation of genes should start with the generation of null alleles or strong loss-of-function (LOF) mutations since this will provide a reference point and a context to study the in vivo function of a gene. Once a phenotype is identified, integration and expression of human cDNA homologous to the fly gene can be tested for its rescuing ability. An example of a simple strategy is shown in Figure 1. Integration of the yeast GAL4 transcription factor with a ribosome skipping peptide (2A) in a gene of interest will create a severe LOF allele (Diao et al., 2015). Upon identification of the phenotype in the fly, rescue experiments by the UAS-human cDNA transgene that is expressed in the proper spatial and temporal domain permit testing the conservation of gene function between fly and human. Comparing the rescue efficiency of human cDNAs with reference (wild-type) versus variant (mutant) sequences is a rapid method of assessing whether a particular variant found in a human patient might be affecting the normal function of this gene. Finally, overexpression of reference and variant human cDNA sequences in wild-type flies can also lead to detection of dominant phenotypes associated with variants found in human patients.

Another key step in the functional annotation of genes is to determine the
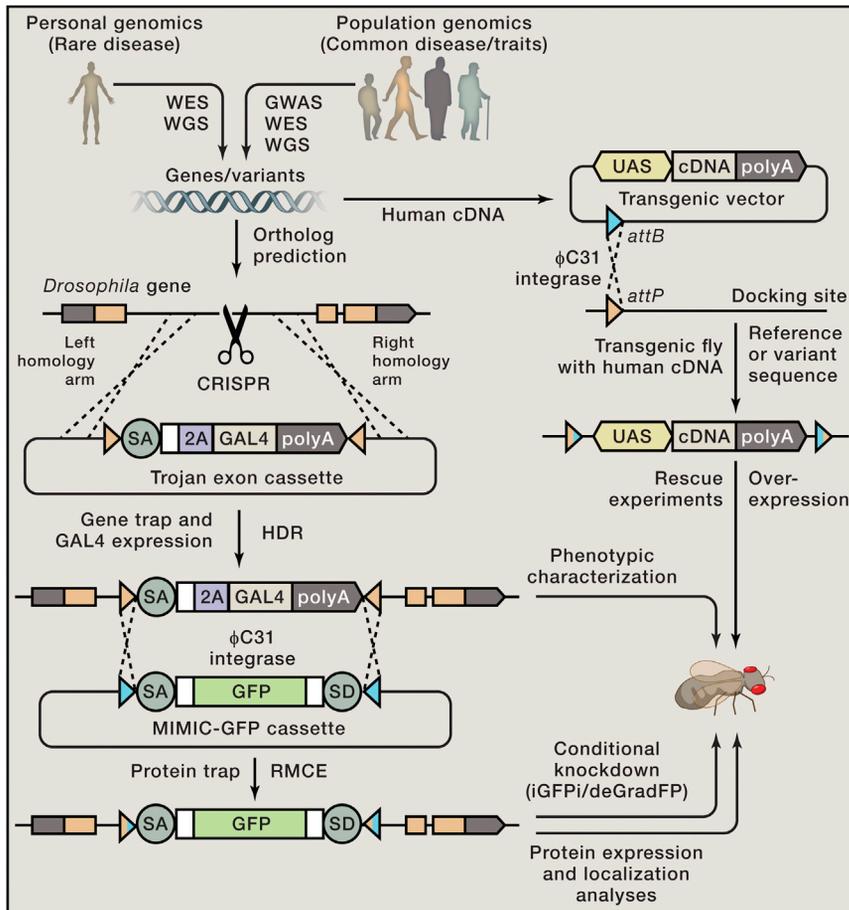
Figure 1. Functional Annotation of Conserved Genes using *Drosophila*
Rapid functional annotation of conserved genes is possible in *Drosophila* by combining a number of technologies and resources. First, the potential fly ortholog of a human gene of interest is identified. An insertion of an artificial exon that functions as a gene trap and allows expression of GAL4 (Trojan exon cassette (Diao et al., 2015)) can be introduced in an intron between two coding exons via Recombination Mediated Cassette Exchange (RMCE) of available MiMIC (*Minos* Mediated Integration Cassette) insertions (Venken et al., 2011). Alternatively, this can be achieved via Homology Directed Repair (HDR) using CRISPR. This Trojan exon consist of splice acceptor (SA) followed by a ribosomal skipping peptide (2A), the *GAL4* gene, and a polyadenylation (polyA) sequence, allowing the expression of GAL4 in the pattern of the gene of interest in loss-of-function (LOF) mutants. By crossing these lines with flies that carry a transgene of the human cDNA under the control of UAS (DNA sequence recognized by GAL4), it can be determined if a human cDNA is able to rescue the fly mutant phenotype. If rescue is achieved with the wild-type (reference sequence) protein, one can further assess the function of variants found in human patients. UAS-human cDNA lines can also be used to assess dominant phenotypes (antimorphic, hypermorphic, or neomorphic) by overexpressing the human gene in a wild-type fly. MiMIC or Trojan gene-traps can be converted into protein-traps via RMCE, allowing intronic tagging of the gene of interest. GFP-tagged genes/proteins can be further knocked down using strategies to degrade the transcript (iGFPi) or protein (deGradFP) in a conditional and tissue specific manner (Nagarkar-Jaiswal et al., 2015), providing stage and tissue specific gene function information.

temporal, cellular, and subcellular distribution of the protein of interest. The simplest strategy is to tag genes in genomic constructs (plasmids, fosmids, or BAC clones), generate transgenic strains, and monitor the tag (e.g., GFP) in vivo. Alternatively, the above mentioned GAL4 cassette can be modified to be replaced with an artificial exon that contains a protein tag (Venken et al., 2011). These tagged proteins are expressed under the control of endogenous regulatory elements, allowing documentation of protein expression patterns and subcellular localization without overexpression. Although the tag is internal to the protein, 75% of the proteins tagged with GFP tested so far have been shown to be functional in vivo (Nagarkar-Jaiswal et al., 2015). In summary, by combining genomic technologies, one should be able to quickly assess the LOF phenotypes and expression pattern of a yet uncharacterized gene, identify the human ortholog, and assess the function of human variants.

Morgan may have been modest about the impact of *Drosophila* research in human physiology and medicine but the long-term impact is obvious: he selected a cost-effective model organism that has provided countless insights into biology, many of which have been directly applicable to human biology and medicine. Going forward, *Drosophila* has the potential to keep on making great contributions, and the era of functional genomics is no exception.

REFERENCES

Bridges, C.B. (1935). J. Hered. *26*, 60–64.

Diao, F., Ironfield, H., Luan, H., Diao, F., Shropshire, W.C., Ewer, J., Marr, E., Potter, C.J., Landgraf, M., and White, B.H. (2015). Cell Rep. *10*, 1410–1421.

Jaiswal, M., Sandoval, H., Zhang, K., Bayat, V., and Bellen, H.J. (2012). Annu. Rev. Genet. *46*, 371–396.

Jan, Y.N., and Jan, L. (2008). Science *319*, 45.

Lewis, E.B., and Bacher, F. (1968). Drosoph. Inf. Serv. *43*, 193.

Lupski, J.R., Roth, J.R., and Weinstock, G.M. (1996). Am. J. Hum. Genet. *58*, 21–27.

Morgan, T.H., Sturtevant, A.H., Muller, H.J., and Bridges, C.B. (1915). The Mechanism of Mendelian Heredity (New York, NY: Henry Holt and Company).

Muller, H.J. (1927). Science *66*, 84–87.

Nagarkar-Jaiswal, S., Lee, P.T., Campbell, M.E., Chen, K., Anguiano-Zarate, S., Gutierrez, M.C., Busby, T., Lin, W.W., He, Y., Schulze, K.L., et al. (2015). eLife *4*, e05338.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Nature *287*, 795–801.

Perrimon, N. (2014). Methods: Drosophila developmental biology methods, *Volume 68* (Amsterdam: Elsevier).

Rubin, G.M., and Spradling, A.C. (1982). Science *218*, 348–353.

Venken, K.J., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A., and Bellen, H.J. (2011). Nat. Methods *8*, 737–743.

Wangler, M.F., Yamamoto, S., and Bellen, H.J. (2015). Genetics *199*, 639–653.

Yamamoto, S., Jaiswal, M., Charng, W.L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). Cell *159*, 200–214.