Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Modeling Genetic Susceptibility to Disease

AC Veith, C Chu, and B Moorthy, Baylor College of Medicine, Houston, TX, United States

© 2018 Elsevier Ltd. All rights reserved.

8.25.1 Genomics and Toxicology
8.25.2 Animal Models for Toxicological Research
8.25.3 Genome Engineering Techniques
  8.25.3.1 Gene Targeting Vector Design
  8.25.3.2 Gene Targeting in Embryonic Stem Cells
  8.25.3.3 Genome Editing via Targeted Endonucleases
8.25.4 Gene Targeting Approaches
  8.25.4.1 Null Alleles
  8.25.4.2 Conditional Alleles
    • Recombinases and recognition sites
    • Spatially controlled conditional alleles
    • Temporally controlled conditional alleles
  8.25.4.3 Point Mutations
  8.25.4.4 Transgenic Rescue
  8.25.4.5 Reporter Genes
  8.25.4.6 Humanized Mouse Models
    • Transgenic humanized mouse models
    • Humanized chimeric mouse models
8.25.5 Screening Methods to Identify Animal Models of Disease
  8.25.5.1 The Collaborative Mouse Models
  8.25.5.2 Large-Scale Mutagenesis-Induced Phenotypic Screens
  8.25.5.3 Gene Trapping and Transposon-Based Screening
  8.25.5.4 Commercially Available Mouse Stocks and Publically Accessible Databases
8.25.6 Alternatives to Mouse Models
  8.25.6.1 Fruit Fly
  8.25.6.2 Zebrafish
  8.25.6.3 Rat
8.25.7 Future Directions
References
Relevant Websites

Abbreviations
GWAS Genome Wide Association Study
SNPs Single nucleotide polymorphisms
ES cell Embryonic stem cell
DSB Double-stranded DNA break
NHEJ Nonhomologous end joining
HDR Homology directed repair
ZFN Zinc finger nuclease
TALEN Transcription activator-like effect nuclease
CRISPR Clustered regularly-interspaced short palindromic repeats
Cas CRISPR associated genes/proteins
PAM Protoscaler adjacent motif
eQTL Expressed qualitative trait loci
ENU N-ethyl-N-nitrosourea

Change History: May 2017. A Veith, C Chu, B Moorthy updated all sections and included new sections and figures from R.G. Gregg and T.E. Geoghegan, Modeling Genetic Susceptibility to Disease, Comprehensive Toxicology, Second Edition. 2010. This is an update of RG Gregg and TE Geoghegan, Modeling Genetic Susceptibility to Disease, In Comprehensive Toxicology (Second Edition), edited by Charlene A. McQueen, Elsevier, Oxford, 2010, Pages 325–333.
8.25.1 Genomics and Toxicology

Genomics has become one of the most important facets of personalized medicine. From a toxicological perspective, an individual’s genetic makeup has been shown to influence how they respond to a wide range of compounds, including pharmaceuticals and xenobiotics. Pharmaceutical and toxicological research today must incorporate the effect of genetic variation on drug dosage, efficacy, bioavailability and toxicity (Blais et al., 2017; Davis et al., 2017).

Next generation sequencing technologies have fostered a genomic revolution where large scale sequencing and genome-wide association studies (GWAS) have facilitated understanding of how genetic variation can impact a wide range of complex diseases, such as cancer or diabetes (Koboldt et al., 2013). GWAS studies have identified a large number of single nucleotide polymorphisms (SNPs) and a number of resources have become available to aid investigators in determining the biological function and impact of SNPs. For example, PharmGKB is an online resource that has been developed to identify the impact of polymorphisms in pharmacology (Whirl-Carrillo et al., 2012). Other resources have also been developed to provide more information surrounding the impact of SNPs in biology and medicine, including dbGAP, F-SNP, SNPinfo Web Server, and SNP Function Portal (Lee and Shatkay, 2008; Wang et al., 2006; Xu and Taylor, 2009).

Additionally, novel gene editing technologies have been developed that allow for an unprecedented ability to generate animal models harboring these types of genetic variation. These animals serve as preclinical platforms for a wide range of toxicology applications, such as drug toxicity testing to understanding how exposure to xenobiotics impact different individuals.

8.25.2 Animal Models for Toxicological Research

Animal models, including transgenic animal models, have been at the heart of many toxicology programs, including drug development and risk assessment (Boverhof et al., 2011). Animal models also play a role in the bench-to-bedside spectrum, from driving basic science discoveries to serving as preclinical models. Animal models provide an unparalleled insight into understanding complex physiological systems and modeling human disease. The development of genetically modified animal models has facilitated the ability to interrogate gene expression and function in both biological and pathological processes. Recent technological advances have enabled the rapid generation of genetically engineered animal models at a rate that not previously possible. Coupled with the increased availability of large-scale genomic datasets and bioinformatics tools, toxicologists now have the unparalleled ability to model genetic susceptibility to disease.

This chapter focuses on advances in genetic engineering that have provided an unprecedented ability to generate animals that can model genetic susceptibility to disease. Novel methods have been developed to allow investigators to generate genetic animal models for their disease or gene of interest at an incredible pace. Moreover, these advances have allowed for the generation of genetically engineered animal models of other species. There are a growing number of databases and resources available to investigators to help identify candidate genes or previously generated genetic animal models for their research. This chapter also highlights novel gene targeting methods and approaches as well as screening methods. The goal is to provide the reader an understanding of what is currently being done in animals to model genetic susceptibility to disease.

8.25.3 Genome Engineering Techniques

Gene targeting, and more recently gene editing, have been used to create targeted mutations to study gene function and to replicate mutations observed in humans. The importance of gene targeting in biology and medicine was recognized in 2007 with the Nobel Prize in Physiology and Medicine to Mario Capecchi, Martin Evans, and Oliver Smithies. This methodology harnessed the cell’s endogenous DNA repair mechanisms to incorporate exogenous DNA target vectors with desired mutations. The pioneering approach utilized homologous recombination in embryonic stem cells, however novel endonuclease-based gene editing technologies are quickly supplanting gene targeting. This discussion primarily features mice, in part due their wide popularity among
biomedical investigators. However, today’s ability to generate genetically engineered animal models extends to other organisms that may be more ideally suited to address a specific research hypothesis or experimental condition.

### 8.25.3.1 Gene Targeting Vector Design

The crucial first step in gene targeting is to design the correct targeting vector and therefore, this initial discussion focuses on some of the key principles of vector design. A variety of mutations can be introduced via a gene-targeting vector, with simplified examples presented in Fig. 1. Subsequent gene targeting approaches may also highlight different aspects of vector design. Gene targeting vectors can be designed to incorporate different genetic mutations that allow investigators to answer a specific research hypothesis, whether to study a gene’s function or create mutations similar to those identified in a specific patient population. Original efforts to create transgenic mice were done with viruses, such as Simian Virus 40, and while the transgene was able to pass on to subsequent generations, the integration site was random (Costantini and Lacy, 1981; Gordon and Ruddle, 1981; Jaenisch and Mintz, 1974). In contrast, gene targeting relies on the ability of a cell to recombine different DNA molecules sharing the same sequence through homologous recombination (Capecchi, 1989). A simplified example of gene targeting vector incorporation is shown in Fig. 2.

**Fig. 1** Mutations that are commonly introduced by gene targeting vectors. **Black** indicates nucleotide sequence, **blue** indicates amino acid sequence, **red** indicates mutations. (A) Missense mutation—a point mutation that results in a change of an amino acid residue. (B) Nonsense mutation—a point mutation that results in the change of an amino acid residue into a nonsense, or stop, codon. (C) Insertion (Frameshift)—an insertion of nucleotide(s) that results in a change of the open reading frame. (D) Deletion (Frameshift)—a deletion of nucleotide(s) that results in a change of the open reading frame. (E) Duplication—duplication of a segment of DNA. (E) Trinucleotide expansion—expansion of trinucleotide repeats.

**Fig. 2** Example gene targeting schematic. In this example, endogenous exon 2 is replaced with a mutant exon 2 containing a point mutation. (A) Murine gene locus containing wild type exons (gray). (B) Linearized targeting vector containing mutant exon 2 (red), positive selection marker neo (orange) flanked by loxP sites (yellow), and negative selection marker HSV-TK (blue). Recombination between A and B at regions of homology produce (C) successful incorporation of mutant exon 2 targeting vector. (D) Addition of Cre recombinase excises neo creating mutant allele. Note, one loxP site remains in the modified allele.
Homologous recombination requires regions of homology between different DNA molecules. These regions of homology are often >1 kilobase pair (kb) in length and flank the 5′ and 3′ end of the desired mutation. As seen in Fig. 2, regions of homology are the sites of recombination between 2A and 2B. It is important that the homologous sequences between the endogenous genome and the gene-targeting vector be isogenic, and able to match the strain of mouse being used. Linearization of the targeting vector increases the low frequency of homologous recombination (Brenner et al., 1986; Hasty et al., 1992). Linearization is commonly achieved through restriction enzyme digest of the targeting vector that is then transfected into isolated embryonic stem cells (ES cells).

Despite linearization of the targeting vector and efficient transfection methods, the frequency of homologous recombination remains low. Therefore, positive and negative selection criteria are often utilized to enrich for cells that incorporate the transgene successfully. The positive selection marker is used to identify ES cells that have incorporated the transgene at the correct site. A common positive selection marker is a neomycin-resistance gene (neo) which provides cells resistance against the antibiotic G418 (Sedivy and Sharp, 1989). Addition of the antibiotic G418 allows only cells which express neo to survive, therefore, the neo cassette or other positive selection markers, are positioned between the regions of homology. The neo cassette is flanked by loxP sites that allow Cre recombinase to excise the neo gene, as shown in Fig. 2. It is also possible for the gene-targeting vector to randomly insert into the genome. Negative selection markers, such as herpes simplex virus thymidine kinase (HSV-TK) marker or diphtheria toxin A (DT-A), can be used to select against these events (Czakó and Márton, 1994; Yagi et al., 1993). Cells expressing DT-A and HSV-TK produce toxic byproducts that render them unable to survive and thus help to select against random insertions. Negative selection markers are positioned outside of the homology arms to allow selection of undesired incorporations of the gene-targeting vector (Fig. 2B). Subsequent PCR, genomic southern, or sequencing can also be performed to ensure the correct insertion of the gene targeting vector. There are a number of other reviews available to highlight the basic principles of gene targeting (Hall et al., 2009).

### 8.25.3.2 Gene Targeting in Embryonic Stem Cells

A tremendous amount of pioneering work has gone into developing gene targeting methods that have been used as the basis for generating animal models with mutant alleles (Muller, 1999). One of the foundations for this method is the ability to integrate foreign DNA into the genome in a targeted manner. This was first shown by introduction of a modified b-globin gene into the endogenous human b-globin locus (Smithies et al., 1985). The compilation of several studies culminated with the ability to successfully transfec exogenous DNA into murine ES cells and that these modified ES cells could be re-introduced into a recipient blastocyst, creating chimeric progeny (Bradley et al., 1984; Evans and Kaufman, 1981; Thomas and Capecchi, 1987). In this method, ES cells are isolated from donor blastocysts, cultured, transfected with the gene-targeting vector, and selected for expansion. ES cells that have correctly incorporated the gene-targeting vector are then reintroduced into a recipient blastocyst by microinjection, and the recipient blastocyst is implanted into a pseudopregnant female mouse (see extensive review here Adams and van der Weyden, 2008). A common scheme involves modifying the genome of 129S1/SvImJ (129/Sv) mice because of their ease in culturing and ability to be reintroduced into C57BL/6J blastocystcs. Chimeric progeny are generally identified by coat color; in this case, consisting of the agouti coat color of the 129/Sv strain and black coat color of the C57BL/6J strain. Mice with the highest percentage agouti coat color are then bred with C57BL/6J mice to verify germ line incorporation of gene targeting vector in subsequent progeny. It is often necessary to further backcross these mice for 10 generations to obtain a pure C57BL/6J background, traditionally one of the major disadvantages of this method. A strategy to overcome this has been to create protocols and conditions by which ES cells from other strains of mice can be modified and used in this scheme (Czechanski et al., 2014). In this approach, a minimum of two generations is needed to obtain mice with ubiquitous incorporation of the transgene and can take several months. A number of other reviews on this topic exist (Bouabe and Okkenhaug, 2013).

### 8.25.3.3 Genome Editing via Targeted Endonucleases

The majority of previously generated mutant animal strains have been created by gene targeting in ES cells. However, recent advances have allowed for the development of endonuclease-based methods to precisely and rapidly create mutant animals. These targetable endonucleases cleave DNA and create double-stranded DNA breaks (DSBs) which can dramatically increase homologous recombination rates (Bibikova et al., 2001). These DSBs are then repaired either by error prone nonhomologous end joining (NHEJ) or homology directed repair (HDR) (Carbery et al., 2010; Kandavelou et al., 2009). NHEJ is an error prone repair mechanism that can produce small insertions or deletions (indels) that result in a frameshift.

A number of nuclease-based gene editing technologies have been developed, including meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effect nucleases (TALENs), and clustered regularly-interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes. Meganucleases are a family of DNA nucleases found in a variety of organisms (Belfort and Roberts, 1997). These nucleases possess a DNA binding domain that recognizes specific nucleotide sequences (Epinat et al., 2003). Both ZFNs and TALENs have emerged as promising technologies to make genomic modification in different model organisms (Cui et al., 2011; Sommer et al., 2015; Welers et al., 2013). Nucleases, such as ZFNs and TALENs have programmable DNA recognition domains that can be engineered to recognize desired DNA sequences (Kim and Kim, 2014). ZFNs and TALENs are programmed by DNA-protein interactions and therefore, a new protein domain needs to be engineered, which can be laborious.
Meganucleases, ZFNs, and TALENs have limited throughput because of the dependence of these systems on protein-DNA interactions, as a new nuclease must be engineered to target different DNA sequences.

A novel methodology has been developed from a prokaryotic immune defense system, designed to destroy foreign DNA, termed clustered regularly-interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes (Sorek et al., 2008). Transformative work has shown how the CRISPR/Cas system (hereafter referred to as CRISPR) functioned as a DNA endonuclease that could be targeted by a single RNA molecule (Jinek et al., 2012). The utility of CRISPR in gene targeting quickly emerged, as a PubMed search for CRISPR yields 1340 results in 2015 alone, highlighting how CRISPR has become the new frontier for the creation of genetically modified organisms. A major advantage of this system is the ease of use, a knockout mouse model can be achieved by inserting Cas9 (Cas nuclease from *Streptococcus pyogenes*) mRNA or protein with a guide RNA (gRNA) into the zygote (Singh et al., 2015). The revolutionary aspect of the CRISPR system lies in the ability to target the nuclease via a gRNA to multiple locations using a single enzyme. For example, Wang et al. (2013) achieved biallelic knockout of two genes, *Tet1* and *Tet2*, simultaneously in a single blastocyst. With previous methods, this would have taken multiple generations of breeding to obtain homozygous knockouts for multiple genes. The same group was then able to show the CRISPR system could be used to tag genes with fluorescent markers or create a conditional allele in one generation (Yang et al., 2013). Using this approach, the authors directly injected the Cas9 mRNA, gRNAs, and DNA constructs to create reporter and conditional alleles by HDR. CRISPR has revolutionized the pace at which we can create transgenic animals. While this technology has shown amazing promise, CRISPR systems have to recognize protospacer adjacent motifs (PAMs) adjacent to the complementary DNA sequence of the gRNA for successful targeting (Mojica et al., 2009). For example, the Cas9 protein has to recognize a PAM sequence of NGG, however current research efforts are underway to ease this requirement (Kleinstiver et al., 2015). While many of these examples use mice as a model organism, CRISPR has become increasingly popular to edit genomes of other model organisms. A more thorough review detailing the protocol for using CRISPR to edit genomes is provided here (Ran et al., 2013). There are also a number of reviews highlighting the future of this system in genome engineering that may be of interest (Doudna and Charpentier, 2014; Sander and Joung, 2014). The CRISPR-Cas system is changing our ability to create transgenic animal models and enabling our ability to model genetic susceptibility to disease in way that was not previously possible. A simplified CRISPR schematic is illustrated in Fig. 3, including the types of alleles that can be readily generated through NHEJ and HDR repair mechanisms.

### 8.25.4 Gene Targeting Approaches

A number of alleles can be generated to best model genetic susceptibility to disease in an animal model. Different mutations and alleles can have different effects on a given gene expression and function. Here we highlight several different types of alleles that are...
commonly used for toxicological research applications. As previously mentioned, the examples focus on murine models, but these types of alleles are applicable in other animal models as well. Table 1 contains an abbreviated summary of the gene targeting approaches, impact on gene expression, and possible research questions that allele can address.

### 8.25.4.1 Null Alleles

The null or knockout (KO) allele is among the most commonly used gene targeting approaches. Early work focused on the hypoxanthine-guanine phosphoribosyl transferase (HPRT) gene, due to its clinical relevance to Lesch-Nyhan syndrome as well as the ease of selection of HPRT-deficient cells [Kuehn et al., 1987; Seegmiller et al., 1967]. Early strategies often used selection markers, such as *neo*, to disrupt the gene of interest, creating null alleles (Doetschman et al., 1987). However, as sequencing of model organisms has grown, more gene targeting vector schemes can be used to create null alleles. These types of alleles traditionally provide investigators insight into the relevance and function of the gene target in biological and pathological processes. For instance, our group was interested in understanding what role, if any, cytochrome P450 (CYP) 1A1 played in hyperoxic lung injury as a model of acute respiratory distress syndrome (Lingappan et al., 2014). The results from this study identified Cyp1a1 as playing a protective role against hyperoxic toxicity. Despite the wealth of information that null alleles have provided basic science researchers, these types of alleles generally do not represent patient populations. Consequently, other gene targeting approaches have been developed to more accurately replicate the alleles found in patients.

As mentioned previously, several targeting vectors have been employed to generate null alleles. A common strategy is to create frameshift mutations, possibly NHEJ indel mutations, that disrupt the correct reading frame of the gene of interest. This has become technically easy due to the use of targeted endonucleases. Positive selection markers, such as *neo*, have been used to disrupt or replace a coding exon in the gene of interest, producing a null allele (Buters et al., 1999). Many commercially available alleles, such as those from the International Mouse Knockout Consortium (IMKC) involve deletion of a coding exon, creating a frameshift mutation (Rosen et al., 2013). In conventional null alleles, the gene of interest is deleted in all cell types and tissues as well as all developmental stages.

Due to the wealth of knowledge provided by KO mice, several collaborative efforts are underway to delete every protein-coding gene in the mouse genome and make the mice, vectors, or ES cells available commercially to investigators. The Knockout Mouse Project (KOMP), International Mouse Phenotype Consortium (IMPC), and Mouse Genome Informatics (MGI) websites are excellent resources for investigators who wish to work with an available KO mouse strain without the need to generate it themselves.

KO mice have helped to clarify the involvement of many genes in various biological and pathological processes and these works span multiple fields including medicine and toxicology, to list a few. These types of alleles are commonly used in a wide range of toxicological research areas. However, these alleles are generally not representative of the mutations seen in human or patient population, thus diminishing their translational potential. Moreover, many genes are also needed for various developmental processes and removal of these genes may lead to embryonic lethality. Fortunately, other gene targeting approaches can generate alleles to overcome these limitations.

### 8.25.4.2 Conditional Alleles

Conditional alleles have been developed for several reasons, including the need to overcome embryonic lethality, as well as to provide investigators a more precise method to interrogate gene function. Conditional alleles selectively delete the gene of interest,
generally spatially or temporally. The ability to generate these types of alleles relies on the power of recombinase enzymes that can be transcriptionally activated by specific promoters to achieve the desired deletion. The mutations that generate knockout alleles in conditional approaches are often similar to those in null alleles, i.e., through frameshift mutations or deletion of crucial coding exons. The following sections will expand on the tools and strategies generally used to generate conditional alleles.

### 8.25.4.2.1 Recombinases and recognition sites

The ability to generate conditional alleles relies on recombination enzymes to drive selective gene deletion. The vast majority of strategies utilize Cre recombinase, a bacteriophage recombinase enzyme that recognizes palindromic sequences, loxP sites, to drive recombination (Abremski and Hoess, 1984; Sternberg and Hamilton, 1981). This system was quickly adopted to manipulate eukaryotic genomes, often to excise DNA between two loxP sites (Sauer and Henderson, 1988). This is often referred to as being ‘floxed’. Another recombination enzyme system that is used is the Flp-FRT recombinase system. In this system, Flp recombinase, isolated from *Saccharomyces cerevisiae*, promotes site-specific recombination at FRT sites (Maiti and Sinha, 1992). The directionality of the sites is important, loxP sites must be in the same direction to excise DNA, or the floxed DNA sequence will be inverted. The IMPC project currently uses these systems in combination to generate alleles that can be combined with Flp and Cre mice to generate different alleles (International Mouse Knockout Consortium et al., 2007). The power of these recombinase enzymes lies in the ability to knock-in the cre and fl genes downstream of a desired promoter to achieve spatial or temporal transcriptional control of the recombinase enzyme. The Jackson Laboratory Cre Repository is an excellent resource to search for previously generated Cre strains to match specific research needs.

### 8.25.4.2.2 Spatially controlled conditional alleles

Specifying Cre expression to a subset of cells can create powerful studies that aim to identify the cell type responsible for a given phenotype. For example, spatially regulated Cre expression can allow an investigator to understand how signaling pathways acted in different cell types within the liver to drive liver pathology. The authors used several spatially controlled Cre lines to selectively delete *Pten* and *Tgfb2* in the liver (*Albumin-Cre*), as well as AAV8-TBG-Cre to limit Cre expression to hepatocytes (Mu et al., 2016).

These types of strategies can help provide more informed conclusions regarding how gene expression in a specific cell type may be contributing to disease pathogenesis. This approach has also proven to be effective to study genes for which the null allele results in embryonic lethality. For example, Kobayashi et al. conditionally knocked out *Nrf1* in the central nervous system with Cre driven by the *Nestin* promoter, allowing for the study of in motor neuron dysfunction (Kobayashi et al., 2011). This approach allowed for interrogation of gene function in a specific biological or pathological context, while maintaining endogenous function in other tissues. These types of alleles are generally best used to answer specific research questions.

### 8.25.4.2.3 Temporally controlled conditional alleles

Cre expression can be limited temporally, with the majority of temporally-controlled conditional alleles are driven by inducible promoters. There is a growing number of inducible Cre lines to generate temporally controlled conditional alleles (Feil et al., 2009). One of the more common strategies is to use a Cre recombinase fused to the hormone-binding domain of a mutant murine estrogen receptor that only binds the synthetic analog of estradiol, 4-hydroxytamoxifen (Danielian et al., 1998). In this strategy, Cre is localized in the cytoplasm until binding of tamoxifen, which allows translocation to the nucleus and facilitates recombination.

One of the powers of inducible strategies is that they afford investigators the ability to delete a gene during a specific treatment or time, allowing that gene to fulfill its developmental and homeostatic functions until the experiment is initiated. Another strategy that can be used is to drive Cre expression by a promoter that is only expressed in certain developmental stages. For example, the promoter for the keratin 19 gene (*K19*) is expressed in early embryonic stages and can drive Cre expression (Means et al., 2005).

### 8.25.4.3 Point Mutations

Previously discussed gene targeting approaches have focused on deleting a gene of interest, completely inactivating a gene’s function and expression. However, many mutations and SNPs that occur in human and patient populations do not have such a dramatic effect. Rather, point mutations can create hypo- or hypermorphic alleles that subtly alter a gene’s function or expression. Large-scale genomic and sequencing efforts have identified numerous SNPs and as their effect on disease or phenotype risk grows, modeling these types of alleles in animal models has become increasingly important. For example, missense mutations in B-cardiac myosin heavy chain (B-MHC) account for a significant proportion of hypertrophic cardiomyopathy (HCM) cases (Blankenburg et al., 2014). In the study conducted by Blankenburg et al. (2014), mice carrying a Val606Met substitution had a similar HCM phenotype compared to wild type mice under homeostatic conditions, however, their phenotype was remarkably worsened when given cyclosporine. This is an excellent example of the benefit to toxicology-based research applications, and how incorporating genomic data in animal models to identify risk factors or off target drug effects can be effective.

These types of mutations are generally achieved through replacement gene targeting vectors. These vectors carry the desired mutation and replace the endogenous allele after positive and negative selection criteria. While modeling individual SNPs would be ideal, a lack of homology between human and mouse genes may not allow for similar mutations in both organisms. However, understanding the impact of a SNP or mutation on the amino acid sequence, for example, may allow for similar mutations to be replicated in animal models.
8.25.4.4 Transgenic Rescue

Selectively deleting a gene by targeting one promoter at a time may be highly inefficient to answer the question of interest. An alternate approach is to transgenically rescue expression of a gene in a subset of cells. This may be helpful in cases of embryonic lethality or cases in which there is a need to address the confounding of target phenotypes. This may require extensive knowledge of the gene’s expression profile and identification of critical tissue or cell types. For example, to study the effect of Na+/H+ exchanger 3 (Nhe3) in hypertension, Li et al. (2015) reintroduced NHE3 expression in the small intestine to protect against off-target intestinal dysfunction during hypertension. The authors describe how Nhe3+/−− mice showed phenotypes in the small intestine that may confound their results and prevent correct conclusions. In this study, the authors only wished to restore NHE3 gene expression to the small intestine and maintain the knockout in the kidney. It may be more efficient to rescue gene expression in a single organ or cell type rather than knock out gene expression in multiple organs or cell types at the same time, highlighting the advantage of using a transgenic rescue.

8.25.4.5 Reporter Genes

There is growing interest in having the ability to track gene expression, both temporally and spatially. To this end, several reporter genes are commonly used including luciferase, green fluorescent protein (GFP), and β-galactosidase (lacZ). Some strategies create null alleles with reporter genes in replacement vectors. Our laboratory, for example, has been interested in studying differences in gene expression CYP1A1 and CYP1A2 expression in response to 3-methylcholanthrene (3-MC) (Jiang et al., 2009) and has used transgenic knock-in of humanized CYP1A1 and murine CYP1A2 promoters driving luciferase expression to study differences in 3-MC induction of CYP1A enzymes in humans and mice. The importance of tracking gene expression via reporter genes is well recognized and the IMPC gene construct includes the ability to express a lacZ reporter that can be used to stain for gene expression (Rosen et al., 2015).

There are also exciting strategies that can couple reporter gene expression with the endogenous gene so that the endogenous gene is not knocked out. This is important given that tagging a protein with a reporter gene, such as GFP, can change a protein’s localization or possibly impair its activity. To overcome this limitation, transcripts are generated that contain two proteins that are translated independently, referred to as bicistronic. The first method described was to include an internal ribosome entry site (IRES). This allows for simultaneous expression of the endogenous gene and reporter gene (Yu et al., 2000). More recently, 2A peptide cleavage sites can be included. In this strategy, a 2A peptide sequence is inserted between the gene of interest and the reporter gene so that they are transcribed together, but translated separately (Kim et al., 2011). These approaches may also prove useful in that there are limited reagents available to track expression of your gene of interest and can use reporter gene expression as a surrogate.

8.25.4.6 Humanized Mouse Models

Despite technical and strategic improvements in genome engineering, there may be diseases or experimental designs where use of animal model’s genes and physiological systems is unable to accurately replicate the human condition. For example, there are diseases that mice cannot replicate because they do not encode the functions necessary for disease development. This has led to the development of humanized murine models of disease. These models may be best adept to model genetic susceptibility to disease. These models have become increasingly popular for engineering mice a range of applications, from human infectious diseases to determining metabolic profiles of human enzymes. Several examples are provided below.

8.25.4.6.1 Transgenic humanized mouse models

There are different strategies that can be taken in creating humanized mouse models. One method is to insert the human gene into the murine locus of the same gene, placing the human gene under the control of the murine promoter. Another method is to insert the human gene, including its promoter at another locus in the murine genome. A common problem that toxicologists face is the metabolism of critical tissue or cell types. For example, to study the effect of Na+/H+ exchanger 3 (Nhe3) in hypertension, Li et al. (2015) reintroduced NHE3 expression in the small intestine to protect against off-target intestinal dysfunction during hypertension. The authors describe how Nhe3+/−− mice showed phenotypes in the small intestine that may confound their results and prevent correct conclusions. In this study, the authors only wished to restore NHE3 gene expression to the small intestine and maintain the knockout in the kidney. It may be more efficient to rescue gene expression in a single organ or cell type rather than knock out gene expression in multiple organs or cell types at the same time, highlighting the advantage of using a transgenic rescue.

8.25.4.6.2 Humanized chimeric mouse models

Humanizing mice by introduction of a single gene may be insufficient or inaccurate to model some human diseases. Major methodological advances have been made that support the development of chimeric mice with human immune and liver cells, allowing for the study of a multitude of human diseases in animal models (Brehm et al., 2014; Grompe and Strom, 2013). For toxicological based applications, this allows for an unprecedented ability to identify drug-drug interactions, predict drug metabolites, and
8.25.5 Screening Methods to Identify Animal Models of Disease

Many of the diseases that currently impact human health are complex diseases such as cancer, heart disease, and diabetes. These diseases involve several genes as well as environmental interactions that contribute to disease pathogenesis. While some of the genes that contribute to these conditions are known, there are unidentified SNPs and expressed qualitative trait loci (eQTL) that contribute to disease. Phenotype-driven screening programs to create random mutations and screen for desired phenotypes may be an optimal method for gene discovery-based applications. By screening for desired phenotypes, one may be able to make novel genetic associations with the screened phenotype.

8.25.5.1 The Collaborative Mouse Models

Historically, investigators have used the DBA, Dahl, and obese models in numerous collaborative projects. Nebert and Felton (1976) showed that genetic differences in toxicity are caused by various environmental pollutants in several inbred strains of mice and in siblings of the (C57BL/6N)(DBA/2N)F1 × DBA/2N backcross, in which the phenotypes “aromatic hydrocarbon responsiveness” or “nonresponsiveness” had been predetermined. Similarly, the obese (ob/ob) model has been extensively used in diabetes (Jeanrenaud, 1979; Islam, 2013) and other diseases such as hypertension (Imperatore et al., 2017). Another model of historical importance is the Dahl strain rat, which provides a model for examining mechanisms involved in the genetic sensitivity or resistance to salt-induced hypertension. Dahl salt-sensitive rats develop hypertension when fed a high salt diet; Dahl salt-resistant rats remain normotensive (Mark, 1991).

More recently, the laborious nature of creating knockout mouse models has led to the development of many collaborative efforts to catalog different alleles and their phenotypes to aid investigators in identifying a candidate model ideally suited to test their hypothesis. The KOMP is currently working to mutate all protein-encoding genes in the mouse and make them available to investigators while the International Mouse Phenotyping Consortium (IMPC) is working to provide publicly basic phenotypic data for those knockout mice (International Mouse Knockout Consortium et al., 2007; Rosen et al., 2015). Large-scale forward genetic screens may be used to identify candidate genes that are positive for a given phenotype. Several repositories and consortiums are working toward knocking out every protein-coding gene and providing basic phenotype data that can be made available to the scientific community to allow for reverse genetic screens. Finally, many large scale genomic studies generally make their datasets available online and with the aid of various bioinformatics tools, investigators can tailor their mouse model to most accurately reflect the genetic changes observed in patients in their disease of interest. Several examples are presented below.

8.25.5.2 Large-Scale Mutagenesis-Induced Phenotypic Screens

Large homozygous deletions do not often replicate the genomic lesions seen in human populations and therefore, conventional null alleles generated in animal models may not accurately replicate the expected phenotype. N-ethyl-N-nitrosourea (ENU) is a chemical agent that causes point mutations to occur randomly in the germ line of male mice (Oliver and Davies, 2012). Breeding these mice with wild type female mice can yield many progeny with random mutations that can then be screened for the phenotype of interest (Nolan et al., 2000). As expected with all screening programs, a large number of mice are needed, and an efficient method for detecting the desired phenotype is needed, which may compromise the value of this approach for an individual investigator. However, these types of large-scale screens are readily done in other model organisms, such as zebrafish. For example, Just et al. found that T-box 20 (Tbx20) is involved in embryonic cardiomyocyte development in an ENU mutagenesis screen (Just et al., 2016).

8.25.5.3 Gene Trapping and Transposon-Based Screening

A more complex method for generating a large number of mutations to screen for desired phenotypes is called gene trapping. Gene trapping involves use of a vector that contains a reporter tag that randomly inserts into the genome. This may generate null alleles that can be identified by sequencing, while expression of the target is tracked by the reporter tag. Many previously generated gene trap vectors have been generated by the International Gene Trap Consortium and are available at the Mutant Mouse Resource and Research Centers (MMRRC). Other gene trap vectors have been used to interrogate the function of those genes, such as the use of a USF1 gene trap vector which helped identify sex-specific regulation of angiotensinogen expression (Park et al., 2012). An excellent review on gene trapping, with vector diagrams, in zebrafish is provided here (Trinh le and Fraser, 2013).

Another methodology for generating a large number of mutations is via transposon based screening. Transposons are vectors within the DNA that are excised by transposase and that reintegrate into the genome (Moriarity and Largaespada, 2015).
8.25.5.4 Commercially Available Mouse Stocks and Publically Accessible Databases

Recent years have seen several organizations and funding agencies work toward creating consortia and repositories for embryonic stem cells, plasmids, and live mice that are made available to the scientific community. These types of resources allow investigators to accomplish reverse genetic screens, screening known mutations for a phenotype of interest. One resource that is currently available is the International Mouse Phenotyping Consortium Web Portal. This allows you to browse phenotype data generated by the IMKC. A description of this online resource is reviewed in Koscielny et al. (2014), and details their protocols, quality control measures, and data access available to the scientific community.

There are also resources of publicly available data from human studies that can be used to identify mutations or SNPs of interest to model in transgenic animals. The National Human Genome Research Institute and European Bioinformatics Institute have a GWAS Catalog that allows users to search publicly available GWAS data. eQTL may also identify polymorphisms associated with the expression of other genes. The Genotype-Tissue Expression (GTEx) project provides a resource to study the impact of genetic variation on gene expression and may help design novel mouse models of complex diseases.

8.25.6 Alternatives to Mouse Models

This review has focused on the use of mice to model genetic susceptibility to disease. While mice are arguably the most commonly used animal model in toxicology-based research, other animal models are viable alternatives that can provide their own advantages to toxicology-based research applications. Of relevance in this regard are studies focusing on Drosophila melanogaster, zebrafish, and rats as alternative models to identify genetic susceptibility to disease. It should be noted that gene targeting experiments using pigs as surrogate models have also been reported. While pigs are excellent models for toxicological based research due to numerous physiological and anatomical similarities with humans, gene targeting experiments using the CRISPR-Cas system have only recently been described (Swindle et al., 2012; Tan et al., 2013; Zhou et al., 2016). It is important to also stress that the brevity in this review is not sufficient to cover the utility of these organisms and other more comprehensive reviews will be included to aid investigators in choosing a model to address their specific research hypothesis.

8.25.6.1 Fruit Fly

Drosophila melanogaster has been used extensively in biomedical and toxicological research for years because they are easy to grow, have small size, reproduce quickly, and can easily be genetically modified. The importance of flies in biology has been well established, as Drosophila was the organism used by Thomas Morgan Hunt to understand chromosomal heredity, whose work in this field was acknowledged with a Nobel Prize in Physiology and Medicine in 1933. These advantages have made Drosophila a favored model organism for forward genetic screens and discovery of candidate genes in a disease model or phenotype (Wolf and Rockman, 2011; Yamamoto et al., 2014). Drosophila is invertebrates that may not replicate human phenotypes accurately and yet, there are phenotypic outputs, in drug discovery applications for example, that can be used to identify candidate genes for testing in other organisms (Pandey and Nichols, 2011). The addition of nuclease-based gene editing technologies in creating precise mutations in a gene of interest is also available in Drosophila (Beumer and Carroll, 2014). Due to the longstanding use of Drosophila in genetics and biomedical research, there is a wealth of online databases, such as Flybase and Flymine available to aid investigators in identifying previously generated fly lines as well as other useful bioinformatics resources (Mohr et al., 2014).

8.25.6.2 Zebrafish

Danio rerio, or zebrafish, is another model commonly used in toxicology and forward genetics-based screening platforms. There are a number of unique advantages that zebrafish can provide, especially in studying early development, as their embryos are transparent (Pickart and Klee, 2014). These vertebrate animals are small and produce large numbers of offspring, also highlighting their convenience in genetic screens (Patton and Zon, 2001). The ability of zebrafish to model human diseases has been reviewed extensively (Lieschke and Currie, 2007). As with other model organisms, targeted nucleases have increased in use in zebrafish as well, enabling rapid generation of new mutant alleles (Hwang et al., 2013). For example, a protocol was developed which CRISPR was used in zebrafish to convert eGFP-transgenic zebrafish into with a Gal4/UAS system (Auer et al., 2014). There are a variety of toxicological research fields, such as chemical and xenobiotic exposure, that have taken advantage of zebrafish models. An example of this is a screening of toxic compounds for estrogenic activity under control by the cyp19a1b gene (Brion et al., 2012). Studies like this with other genes of interest or other toxic effects can be easily completed in zebrafish to quickly adapted with the aid of nuclease-based gene targeting approaches. As with many model organisms, several online resources are available to help investigators learn about zebrafish and identify necessary reagents are available at the zebrafish international resource center (ZIRC) and zebrafish model organism database (ZFIN).
8.25.6.3 Rat

Rats are one of the most extensively used animal models in biomedical research (Jacob et al., 2010). While rats have similar advantages to mice, such as small size and genetic homology, rats may be more suitable to replicate certain human pathologies, such as cognition and memory (Iannaccone and Jacob, 2009; Jacob, 1999). Many previous studies that utilized rat models focused on how different rat strains responded to a given experimental approach or treatment and studying these differences (Mark, 1991; Huang et al., 2016; Oltra-Noguera et al., 2015). While technologies to edit the rat genome have existed for some time, only recently with the increasing use of CRISPR technologies and genome sequencing has the ability to edit the rat genome approached the ability of the mouse (Shao et al., 2014). CRISPR has been harnessed to create eGRF and Cre knocking rats, allowing for development conditional and reporter alleles (Ma et al., 2014a). Other groups have successfully reporter simultaneous deletion of multiple genes in one generation in rat embryos with CRISPR (Ma et al., 2014b). The rat has been a preferred animal for pharmacological and toxicological studies in drug discovery because of their physiological similarities. As a means to study calcium signaling and cardiotoxicity in drug discovery applications, Szebényi et al. (2015) created a transgenic rat model that would express a fluorescent GCaMP2 calcium sensor. In this study, the authors were interested in understanding cardiotoxicity of various antimalarial agents. This is an outstanding example of harnessing newer gene targeting methods and tailoring them to the model organism best suited to address a specific hypothesis.

8.25.7 Future Directions

Animal models have proven to be tremendously valuable in providing insight into the fundamentals of biology and medicine in an effort to model genetic susceptibility to disease. Gene targeting and editing methods and approaches have enabled an unprecedented ability to alter the murine genome and create mutant alleles. While mice have traditionally been at the forefront of modeling genetic diseases, advances in genomics and sequencing are quickly allowing other animal models to be used in this capacity just as easily. This allows other animal models that may be preferred for a given research hypothesis to be used, ideally increasing our ability to translate findings in animal models to humans. With the advent of gene-editing endonuclease technology coupled with large-scale human genomic datasets, investigators are now able to identify candidate mutations in animal models, and create necessary alleles at a pace that was not possible 5 years ago. We are now able to identify genetic markers associated with xenobiotic toxicity or chemotherapy resistance that can be generated in mice in months as opposed to years. Other tools, such as conditional knockouts and humanized models are allowing for more precise interrogation of gene function and more accurate translation of data from animal models to humans. These improvements in gene editing technology and approaches are allowing unprecedented ability to characterize the toxicity of chemical compounds on physiology and model genetic susceptibility to disease.

See also: 8.34. Molecular Biomarkers.

References

Abremski, K., & Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *The Journal of Biological Chemistry, 259*, 1509–1514.

Adams, D. J., & van der Weyden, L. (2008). Contemporary approaches for modifying the mouse genome. *Physiological Genomics, 34*, 225–238.

Agrawal, A. S., Garron, T., Tao, X., Peng, B.-H., Wakamiya, M., Chan, T.-S., Couch, R. B., & Tseng, C.-T. K. (2015). Generation of a transgenic mouse model of Middle East respiratory syndrome coronavirus infection and disease. *Journal of Virology, 89*, 3659–3670.

Auer, T. G., Duroure, K., Concordet, J.-P., & Del Bene, F. (2014). CRISPR/Cas9-mediated conversion of eGFP- into Gal4-transgenic lines in zebra fish. *Nature Protocols, 9*, 2823–2840.

Beemer, K. J., & Carroll, D. (2014). Targeted genome engineering techniques in *Drosophila*. *Methods, 68*, 29–37.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and Cellular Biology, 21*, 289–297.

Blais, E. M., Rawls, K. D., Dougherty, B. V., Li, Z. I., Kolling, G. L., Ye, P., Waltz, A., & Papin, J. A. (2017). Reconciled rat and human metabolic networks for comparative toxigenomics and biomarker predictions. *Nature Communications, 8*, 14250.

Blankenburg, R., Hackert, K., Wurster, S., Deenm, R., Seilman, J. G., Seilman, C. E., Lohe, M. J., & Schmitt, J. P. (2014). β-Mysin heavy chain variant Val606Met causes very mild hypertrophic cardiomyopathy in mice, but exacerbates HCM phenotypes in mice carrying other HCM mutations. *Circulation Research, 115*, 3388–3398.

Bouab, H., & Okkenhaug, K. (2013). Gene targeting in mice: A review. *Methods in Molecular Biology (Clifton, N.J.), 1064*, 315–336.

Boverhof, D. R., Chamberlain, M. P., Elcombe, C. R., Gonzalez, F. J., Heffich, R. H., Hernández, L. G., Jacobs, A. C., Jacobson-Kram, D., Luijten, M., Maggi, A., Manjanatha, M. G., van Berthem, J., & Gollapudi, B. B. (2011). Transgenic animal models in toxicology: Historical perspectives and future outlook. *Toxicological Sciences: An Official Journal of the Society of Toxicology, 121*, 207–233.

Bradley, A., Evans, M., Kaufman, M. H., & Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature, 309*, 255–256.

Brehm, M. A., Wiles, M. V., Greiner, D. L., & Shultz, L. D. (2014). Generation of improved humanized mouse models for human infectious diseases. *Journal of Immunological Methods, 410*, 3–17.

Brenner, D. A., Smigocki, A. C., & Camerini-Dero, R. D. (1986). Double-strand gap repair results in homologous recombination in mouse L cells. *Proceedings of the National Academy of Sciences of the United States of America, 83*, 1762–1766.
Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thom, C. F., Altman, R. B., & Klein, T. E. (2012). Pharmacogenomics knowledge for personalized medicine. Clinical Pharmacology and Therapeutics, 92, 414–417.

Wolf, M. J., & Rockman, H. A. (2011). Drosophila, genetic screens, and cardiac function. Circulation Research, 109, 794–806.

Xu, Z., & Taylor, J. A. (2009). SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Research, 37, NaN–NaN.

Yagi, T., Nada, S., Watanabe, N., Tamemoto, H., Kohnmura, N., Ikawa, Y., & Aizawa, S. (1993). A novel negative selection for homologous recombinants using diphtheria toxin A fragment gene. Analytical Biochemistry, 214, 77–86.

Yamamoto, S., Jaiswal, M., Chang, W.-L., Gambin, T., Karaca, E., Mirzazai, G., Wiszniekski, W., Sandoval, H., Haeltzerman, N. A., Xiong, B., Zhang, K., Bayat, V., David, G., Li, T., Chen, K., Galla, U., Hanel, T., Pehtivan, D., Penney, S., Visiers, L., E. L. M., de Ligt, J., Jhangiani, S. N., Xie, Y., Tsang, S. H., Parmar, Y., Sivaci, M., Battaoglu, E., Muzny, D., Wan, Y.-W., Liu, Z., Lin-Moore, A. T., Clark, R. D., Curry, C. J., Link, N., Schulze, K. L., Boenwinkle, E., Dobyns, W. B., Ailskovets, R., Gibbs, R. A., Chen, R., Lupski, J. R., Wangler, M. F., & Bellen, H. J. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell, 159, 200–214.

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., & Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell, 154, 200–214.

Yu, Y., Annala, A. J., Banito, J. R., Toyokuni, T., Satyamurthy, N., Namavari, M., Cherry, S. R., Phelps, E. H., Herschman, H. R., & Gambhir, S. S. (2000). Quantification of target gene expression by imaging reporter gene expression in living animals. Nature Medicine, 6, 933–937.

Zhou, X., Wang, L., Du, Y., Xie, F., Li, L., Liu, Y., Liu, C., Wang, S., Zhang, S., Huang, X., Wang, Y., & Wei, H. (2016). Efficient generation of gene-modified pigs harboring precise orthologous human mutation via CRISPR/Cas9-induced homology-directed repair in zygotes. Human Mutation, 37, 110–118.

**Relevant Websites**

- [http://ctdbase.org](http://ctdbase.org)—Comparative Toxicogenomics Database.
- [http://flybase.org](http://flybase.org)—A database of Drosophila Genes & Genomes.
- [http://flymine.org](http://flymine.org)—FlyMine Database.
- [http://ghr.nlm.nih.gov](http://ghr.nlm.nih.gov)—Genetics Home Reference.
- [http://informatics.jax.org](http://informatics.jax.org)—Mouse Genome Informatics.
- [http://jax.org](http://jax.org)—Jackson Laboratories.
- [http://komp.org](http://komp.org)—Knockout Mouse Project.
- [http://mmmr.org](http://mmmr.org)—Mutant Mouse Resource & Research Centers.
- [http://mousephenotype.org](http://mousephenotype.org)—International Mouse Phenotype Consortium.
- [http://pharmgkb.org](http://pharmgkb.org)—The Pharmacogenomics Knowledgebase.
- [http://rgd.mcw.edu](http://rgd.mcw.edu)—Rat Genome Database.
- [http://zebrafish.org](http://zebrafish.org)—Zebrafish International Resource Center.
- [http://zfin.org](http://zfin.org)—Zebrafish Model Organism Database.