A History of Mouse Genetics: From Fancy Mice to Mutations in Every Gene

María J García-García

1.1 Overview

The laboratory mouse has become the model organism of choice in numerous areas of biological and biomedical research, including the study of congenital birth defects. The appeal of mice for these experimental studies stems from the similarities between the physiology, anatomy, and reproduction of these small mammals with our own, but it is also based on a number of practical reasons: mice are easy to maintain in a laboratory environment, are incredibly prolific, and have a relatively short reproductive cycle. Another compelling reason for choosing mice as research subjects is the number of tools and resources that have been developed after more than a century of working with these small rodents in laboratory environments. As will become obvious from the reading of the different chapters in this book, research in mice has already helped uncover many of the genes and processes responsible for congenital birth malformations and human diseases. In this chapter, we will provide an overview of the methods, scientific advances, and serendipitous circumstances that

have made these discoveries possible, with a special emphasis on how the use of genetics has propelled scientific progress in mouse research and paved the way for future discoveries.

1.2 Establishing the Mouse as a Mammalian Model for Research

Mice have accompanied humans since the early days of agriculture. Therefore, it is no surprise that people developed curiosity about these small mammals and even fancied them as pets [1]. Ancient Chinese and Japanese records report the domestication and breeding of many varieties of mice with different coat colors, like *albino* or *yellow*, and peculiar behaviors, such as those of "waltzing mice," which tend to run around in circles due to mutations that affect the inner ear [2]. During the 1800s and well into the early twentieth century, these "fancy" mice gained popularity among Europeans and Americans, who imported them and set up breeding programs, showing their most fancy specimens at mouse shows and clubs. As fortune would have it, one of these "mouse fanciers," Miss Abbie E. C. Lathrop, a retired school teacher who set up a mouse pet farm around 1900, played an important role in the establishment of mice as a model organism for research experimentation [1].

M. J. García-García (✉)
Molecular Biology and Genetics Department, Cornell University, Ithaca, NY, USA
e-mail: garciamj@cornell.edu

1.2.1 Mendelian Genetics in Mice

The first reports of mice being used for research purposes date back to the sixteenth century, when Robert Hooke analyzed the effects of increased air pressure on mice [3]. However, it would not be until the beginning of the twentieth century that scientists unleashed the power of mouse genetics by demonstrating that Mendel's laws of inheritance are also applicable to these small mammals. In 1902, French biologist Lucien Cuénot was the first to report the use of *albino* coat-color mice to confirm Mendel's laws of inheritance. This report was quickly followed by work from additional scientists, who confirmed and extended these findings to other genetic mouse traits [2]. It would be one of these scientists, American William E. Castle, who would become recognized as the father of mammalian genetics, a merit based on his multiple research contributions, as well as his influential role as the director of the Bussey Institute of Experimental Biology at Harvard from 1909 until 1937, where many prominent scientists trained and worked under his supervision, including Clarence C. Little, Sewal Wright, Leonel Strong, George D. Snell, and Leslie C. Dunn, to name a few [4].

Miss Abbie Lathrop's mice farm, located in Granby, Massachusetts, played a critical role during these first years of research on mouse genetics. At her farm, Miss Lathrop bred several colonies of mice, either collected from the wild or imported from European "mouse fanciers," with the intention of selling them as pets. However, she unexpectedly became the supplier of mice for the Bussey Institute, as well as a few other research institutions. Many of the mice currently used in laboratories worldwide can be traced to the colonies initially established by Abbie Lathrop. However, Lathrop's contributions were not limited to being a mouse provider. She was a meticulous breeder and a perceptive observer of her mice, as attested by her careful breeding records and the multiple research papers she contributed to.

1.2.2 Inbred Mouse Strains

While the birth of mouse genetics was linked to the rediscovery of Mendel's laws, it was research on cancer that dominated during the following five decades. Through her careful breeding, Abbie Lathrop had noticed that some of her mice colonies had a propensity to develop skin lesions. In an effort to diagnose these, she sent mice to Dr. Leo Loeb, an experimental pathologist at the University of Pennsylvania, who concluded that Lathrop's mice were developing cancers [5]. This interaction marked the beginning of their collaboration, which rendered several important publications on cancer susceptibility of different mouse strains. Meanwhile, other investigators were experimenting with transplanting tumors in mice and grappling with the idea of whether cancer susceptibility was a heritable Mendelian trait. Support for this hypothesis came from early observations that tumors could be transplanted among waltzing mice, but failed to grow if transplanted onto mice of a different colony. Starting in 1909 and all the way into the 1920s, critical papers from Ernest E. Tyzzer, Leo Loeb, Maude Slye, and Halsey Bagg supported the heritability of cancer susceptibility. However, these investigators found so much variability in their data that they had problems verifying their own observations or concluding whether cancer susceptibility was a dominant or a recessive trait. Around 1909, Clarence C. Little and Leonell C. Strong postulated that the culprit of such variability was the inherent genetic heterogeneity of the mouse strains that were being used for experimentation. To solve this problem, they launched intensive breeding programs to achieve isogenic mouse strains [5]. Their thought was that by systematically performing brother-to-sister matings for more than 20 generations, the genetic constitution of the resulting mice will become homogeneous and stable (isogenic), making them ideal for research subjects (Fig. 1.1). This idea was received with great skepticism in the scientific community since inbreeding was known to be
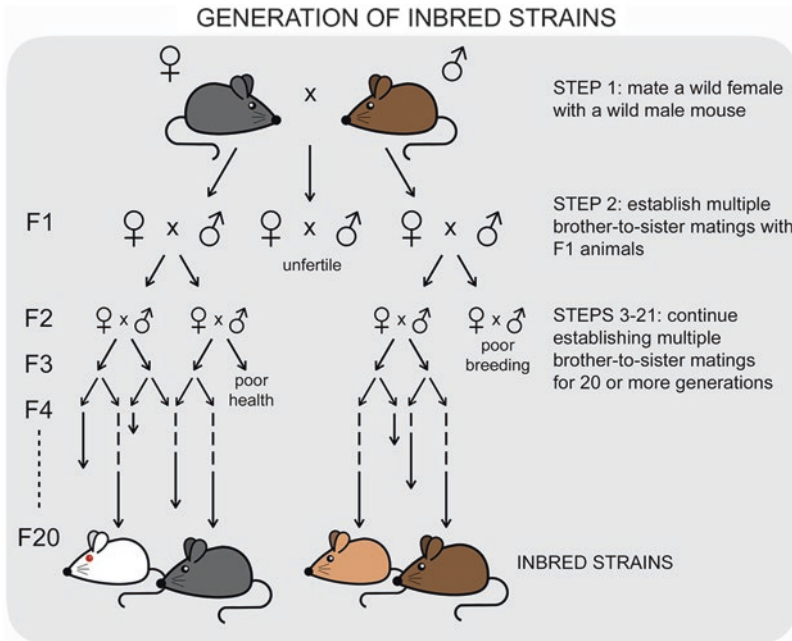
Fig. 1.1 *Inbred mouse strains*. Inbred mice are generated by crossing two wild mice, then systematically performing sister-to-brother matings for more than 20 generations. As breeding proceeds and alleles segregate, individual traits eventually reach homozygosity. Homozygosity for some traits can affect the fertility or viability of mice, compromising further breeding. In other cases, homozygosity pro-duces distinct visible phenotypes, such as different coat colors. Selection of healthy breeders with specific characteristics is performed in each generation in order to render different isogenic mouse strains that can be easily maintained in research facilities. The genome of inbred mice is 98% identical to that of their siblings by 20 generations and 99.5% identical at 40 generations

evolutionarily discouraged, and it was feared that, as recessive factors present in wild mice populations reached homozygosity, a "sterility barrier" would be encountered. However, Little and Strong thought that it would be possible to bypass this "sterility barrier" by keeping multiple independent crosses for each generation and selecting those that did not carry factors detrimental to vigor, reproduction, or susceptibility to diseases. Years of breeding would be needed for these and a few other investigators to reach their goals and establish several viable lines of isogenic mice. Their journeys were not exempt from unexpected challenges, such as disease outbreaks and accidents, including one that decimated 80% of the ongoing crosses due to the escape of stove gases into one of the "mouse coops" [5]. The resulting colonies became known as **inbred mouse strains** and constitute the first innovation in the field of mouse genetic research. The first inbred strain, called DBA (which carries three color-coat alleles for *dilute*, *brown*, and *non-agouti*), was established by Clarence C. Little, but many others followed. Today, there are more than 450 available inbred strains, whose genealogies can be found in the following review [6].

1.3   Getting to Know the Mouse Genome: From Inbred Lines to Genetic Maps

The establishment of inbred strains provided standardized, genetically uniform strains of mice to be used in the study of cancer. Additionally, the large breeding programs required for their generation had important ripples in the field of mouse genetics. As different inbred strains became available during the 1920s and 1930s, it was obvious that they differed in a variety of characteristics, not only cancer susceptibility, but also coat color, behavior, longevity, and many

others. At that time, it was clear that these differences were likely due to either the separation of different allele variants present in the original populations of wild mice used to generate inbred strains or to spontaneous mutations arising during breeding. The **chromosome theory of inheritance**, proposed by Walter Sutton in 1902, and the term **gene**, coined by Wilhelm Johannsen in 1909, provided a framework for understanding that the different alleles of inbred strains correlated with physical entities in chromosomes. However, it would not be until the mid-1940s that the nature of nucleic acids as carriers of genetic information would be recognized. As a consequence, during most of the first half of the twentieth century, genes were just viewed as alleles that segregated in specific ways during breeding, causing dominant or recessive phenotypes.

1.3.1 Inbred and Congenic Strains

At a time when tools to analyze the mouse genome were scarce, early mouse geneticists focused on using inbred strains to identify different alleles responsible for particular traits and follow their segregation through breeding. With the objective of applying this genetics methodology to the study of cancer in mice, Clarence C. Little founded in 1929 the Roscoe B. Jackson Memorial Laboratories in Bar Harbor, Maine. This institution would become an important center for mouse genetics, both as a research organization and, later, as a supply center for mice strains to other institutions [7]. During the first few years, research at the Jackson Labs focused on the identification of alleles that could explain the ability of inbred strains to accept or reject transplanted tumors. For this, animals from two inbred strains with different ability for accepting transplanted tumors were crossed with each other, then the resulting progeny (from F1, F2, F3, and subsequent generations) were analyzed for the inheritance of resistance to tumor transplant. By applying this, so-called **outcross-intercross method** at Jackson Labs, Clarence Little and Leonel Strong were able to deduce that transplant rejection was controlled by multiple loci, which were called *histocompatibility* (*H*)

loci. However, it was not until 1948 that George Snell could isolate independent alleles responsible for tumor rejection. To isolate different *histocompatibility* loci, Snell applied a new breeding scheme known as the "**outcross-backcross-intercross method**," which entailed breeding mice from two inbred strains, one of which (recipient strain) rejected tumors from the other one (donor strain), followed by mating the F1 hybrid progeny to animals from the parental inbred donor strain, then continue backcrossing to the donor strain individuals selected from the G2, G3, G4, and subsequent generations for their ability to carry the allele causing tumor rejection (Fig. 1.2). This selection of carriers often required brother–sister intercrosses since many of the alleles for cancer rejection behaved in a recessive fashion. Snell calculated that, by backcrossing selected carriers for more than ten generations, the genome of the resulting mice will mostly originate from the donor strain, except for a small chromosomal segment containing the loci responsible for the tumor rejection phenotype [8]. Strains produced through this method, later called **congenic strains**, represent an important method for the identification of specific genetic loci [1]. Snell's congenic strains carrying alleles for tumor rejection turned out to be critical for the analysis of the *H2 histocompatibility complex*, a work that granted him the Nobel Prize in 1980.

1.3.2 The Origins of Developmental Genetics

In 1927, Nelly Dobrovolskaia-Zavadskaia, a cancer research scientist working at the Pasteur Institute in Paris, discovered a dominant mutation in the course of an X-ray mutagenesis screen that caused animals to develop a short tail [9]. This mutation, called *Brachyury* or *T*, became one of the first developmental mutations studied in mammals. Initially, however, the interest in this mutation focused not on understanding embryology, but rather on unraveling the puzzling genetic behavior of the *T* locus, which presented several violations of Mendel's laws. The first of these violations was an abnormal proportion of mice with short tails in the progeny of heterozygote *T*
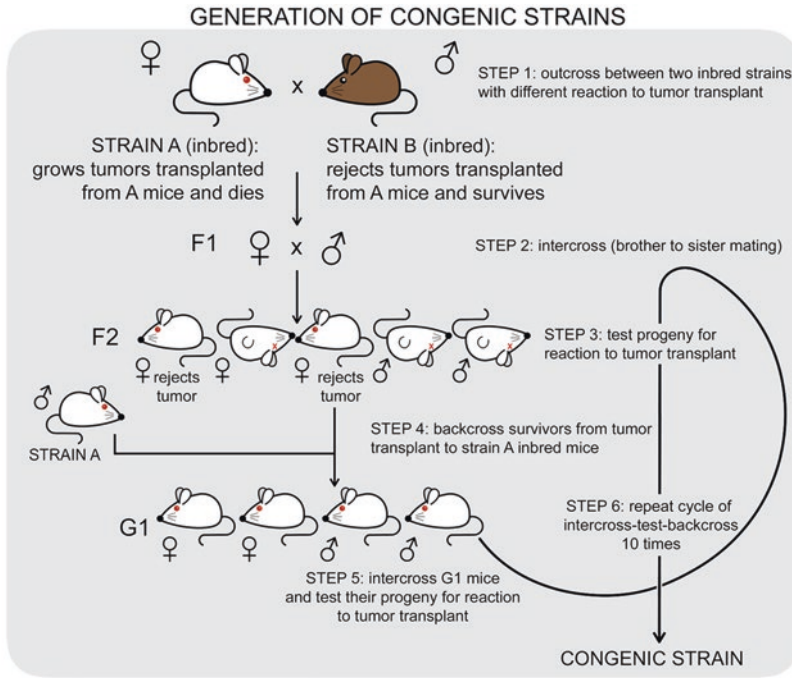
Fig. 1.2 *Congenic mouse strains*. The outcross-intercross-backcross method allows the genetic isolation and propagation of an individual genetic element responsible for a selectable trait. This method was first used by George Snell to isolate loci responsible for the rejection to tumor transplant. By backcrossing selected mice carrying the allele for tumor rejection for ten or more generations to inbred mice that lack this allele (strain A), the genome of the resulting congenic strain originates mostly from strain A, except for a small chromosomal segment that contains the locus responsible for tumor rejection (which originated from strain B). Note that the breeding scheme in the illustration applies to the generation of congenic strains for recessive traits. For dominant traits, intercrosses are not required since selection for tumor rejection can be done directly in the progeny of each backcross

animals, a phenomenon due to the fact that homozygote *T/T* animals died in utero shortly after gastrulation. However, many other mysteries surrounded the *T* locus, including the findings that some *T* chromosomes showed puzzling genetic interactions with alleles from wild mice, could suppress recombination, and caused transmission ratio distortion in males (reviewed in [10]). The study of these anomalies revealed that the genetic behaviors of the *T* locus were in fact due to several linked loci that became known as the *t-complex*, an area later discovered to span the third distal part of chromosome 17, containing more than 500 genes. Additionally, it was found that certain allelic combinations (haplotypes) of the *t-complex* contained embryonic lethal mutations, small inversions (which were responsible for the suppressed recombination), and alleles causing male sterility (which explained the transmission ratio distortion). Sorting out these mysteries took

more than 70 years of research and the work of numerous investigators, including Leslie C. Dunn, Salome Glueckshon-Schoenheimer, Mary Lyon, Dorothea Bennett, Lee Silver and Karen Artz, to name a few. Because of the numerous embryonic lethal mutations at the *t-complex*, understanding the intricacies of this locus inspired the study of embryonic development, contributing to the identification and characterization of many mutations that disrupted development at different embryonic stages [11].

1.3.3 Linkage Analysis, Complementation Tests, and Recombination Maps

Through experiments with flies, Thomas Morgan had shown that the segregation of certain alleles violated Mendel's laws of independent assortment

and that the basis for this phenomenon was the location of cosegregating alleles in the same chromosome [12]. This principle, called **genetic linkage**, was first demonstrated in mice by John B. S. Haldane using *albino* and *pink-eyed dilution* fancy mice to show that alleles for these two loci segregated together [13]. Early mouse geneticists soon adopted linkage as a convenient tool for tracking loci of interest in inbred and congenic strains. Linkage was useful because, if dominant alleles with visible phenotypes were found linked to alleles that would otherwise only be detectable with the help of time-consuming tests, linked alleles could be used as visible markers in breeding schemes, enormously facilitating the maintenance and analysis of "invisible" interesting alleles (Fig. 1.3). The convenience of using linkage as a tool prompted the generation of inbred strains that contained different "marker" traits, such as different coat colors (albino, brown, pink-eyed dilution) or other morphological characters (i.e., the short tail of *T* mice and the kinked tail of *Fused* mice). Inbred strains simultaneously containing several of these markers, called "**linkage testing stocks**," were especially useful, since they allowed to establish whether or not a new phenotype was linked to one of different markers in the course of a single cross strategy [14].

As more allele variants were discovered in different inbred strains, it became important to discern whether some of the observed phenotypes were controlled by the same or through different loci. For recessive alleles, this was done by crossing two mice, each heterozygote for one of the alleles to be tested, then inquiring whether the progeny showed the recessive phenotype, a breeding strategy known as **complementation test** (Fig. 1.4).

Although early mouse geneticists could not pinpoint where their alleles were exactly located within chromosomes in physical or molecular terms, linkage analysis allowed them to map their position in relationship with other known alleles. This strategy was previously exploited by fly geneticists in the early twentieth century for the generation of what became known as **recombination maps or linkage maps**. Linkage maps relied on the facts that any two loci in close proximity within a chromosome will have a tendency to segregate together and that recombination between these loci, due to crossovers during the generation of gametes, can be used as an index of the distance between them (Fig. 1.3; [15]). In mice, recombination mapping efforts were initially limited to alleles that were interesting as based on their relevance to human disease. As a consequence, linkage maps grew very slowly. By 1941, the first edition of the Biology of the Laboratory Mouse, a text of reference for mouse investigators at the time [16], listed 24 independent loci, 15 of which were mapped to 7 different linkage groups. The progress of linkage maps has been captured in the regular publication of the Mouse News Letter (MNL), a free biannual bulletin that ran between 1949 and 1991 and was used by geneticists to report new mutants, inbred strains, as well as updates of the "Mouse Linkage Map." Leslie C. Dunn, Salome Gluecksohn-Waelsch, Margaret Green, and Mary Lyon were among the first editors of the newsletter, which constituted the first "journal" on mammalian genetics until *Mouse Genome* was created [5, 10, 17]. A historical event marking the progress of linkage analysis took place in 1958 at the Tenth Congress of Genetics in Montreal, where the staff of the Jackson Laboratory put together a Live Linkage Map of the Mouse, with live mice from about 60 different strains, each in a small cage, showcased onto 18 lines that represented different linkage groups. While the exhibit proudly displayed the achievements of the scientific community at the time, it is worth mentioning that linkage groups were listed in the order in which they were discovered, since it was not yet possible to assign these groups to any chromosomal location.

1.3.4 Cytogenetics: Chromosomal Maps and Rearrangements

Around the 1920s, the use of dyes such as orcein, Giemsa, or Feulgen was used to karyotype animals of different species and determine differences in their genome organization. Using these
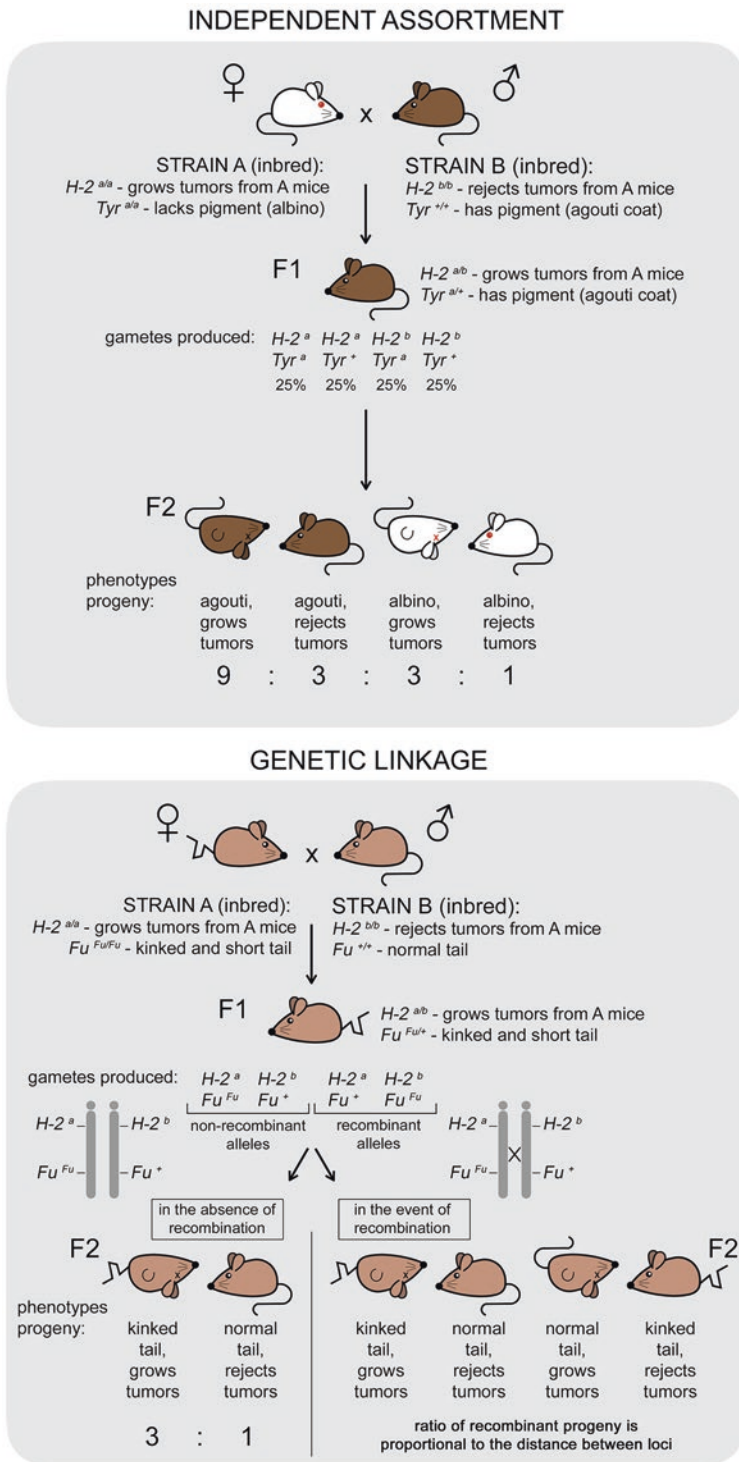
Fig. 1.3 *Independent assortment and genetic linkage.* According to Mendel's law of independent assortment (upper panel), the alleles for different genes segregate independently during gamete formation. In the illustrated example, the alleles for albinism (*Tyr*^a) and the ability to grow/reject tumors (alleles *H-2*^a & *H-2*^b) segregate

dyes, Theophilus S. Painter was the first one to determine that house mice contain 20 chromosome pairs (19 autosomes, plus the X and Y sex chromosomes) [18]. Proper identification of mouse chromosomes was initially challenging due to the facts that early staining protocols revealed uniformly stained chromosomes, and that all mouse chromosomes were found to be telocentric. However, advances in cytogenetics during the 1960s and 1970s led to the development of alternative staining protocols called Q, G, R, or C **banding methods**. In these new protocols, samples were subject to chromatin denaturation and/or a mild enzymatic digestion prior to staining with a DNA-binding dye. These treatments affected chromatin differently, depending on its composition and/or structure and, as a result, the dyes revealed reproducible patterns of high- and low-intensity bands that were unique to each chromosome. In this way, banding methods allowed the identification of individual chromosomes and the generation of detailed **cytogenetic or chromosomal maps** of the mouse genome [1, 5].

As different laboratory and wild mouse strains were analyzed with banding methods, differences among strains were detected in the form of chromosomal translocations, deletions, duplications, and inversions. By analyzing the banding patterns of these chromosomal rearrangements, and especially those that disrupted known loci and/or linkage relationships, investigators could determine the chromosomal location of genes. For instance, by using Q and G banding on a deletion involving the *albino* locus, it was possible to

determine its location to chromosome 7 [19]. Using this strategy, linkage groups previously identified through recombination mapping could finally be assigned to specific chromosomes, an achievement that was reflected for the first time in the 1975 issue of the Mouse News Letter [20]. By 1980, all linkage groups had been assigned to physical chromosomes (reviewed in [21]).

1.3.5 Improving Linkage Maps: New Markers, Recombinant Inbred Lines, and Interspecific Backcrosses

Because linkage maps depend on the recombination between alleles that can serve as markers, the resolution of these maps depends on two factors: the number of markers available and the number of crossover events that can be analyzed. Efforts to address these limiting factors and produce a comprehensive map of the mouse genome spanned most of the second half of the twentieth century.

Initially, linkage analysis could only be performed using a limited number of **morphological markers**, allele variants with phenotypes that could be directly observed in animals, such as coat-color variants (Fig. 1.5, top-left panel). However, advances in molecular biology allowed the development of two additional types of markers: biochemical polymorphisms and DNA polymorphisms. **Biochemical markers** became available during the 1940s and 1950s, when it was discovered that protein extracts from different inbred strains sometimes showed differences

---

Fig. 1.3 (continued) independent of each other, generating four types of gametes that, when randomly combined during fertilization, give rise to four different phenotypes in the F2 progeny at the indicated 9:3:3:1 ratios. Genes located on the same chromosome do not obey Mendel's law of independent assortment and, instead, segregate together in gametes (lower panel, left). In the example, the genes *H-2* and *Fused (Fu)* are genetically linked, and as a consequence, heterozygote animals at these loci only produce two types of gametes that, when randomly combined, give rise to two phenotypes in the F2 progeny at 3:1 ratios. Linkage between alleles can be used for tracking the inheritance of "invisible" traits. In this example, the morphology of the tail can be used to track the inheritance

of the ability of grow/reject tumors. During gametogenesis, the "linkage" between alleles located on the same chromosome can be disrupted in the event of chromosome recombination (lower panel, right). In this case, recombinant allelic combinations can be found in gametes, and four phenotypes can be observed in F2 progeny. While these four different phenotypes are similar to the ones expected if the genes had undergone independent assortment, their observed ratios are not 9:3:3:1. The ratio of progeny from recombinant gametes is proportional to the physical distance between the genes on the chromosomes. This principle can be used to infer the relative location of genetic elements in the genome and is the basis for the generation of genetic linkage maps
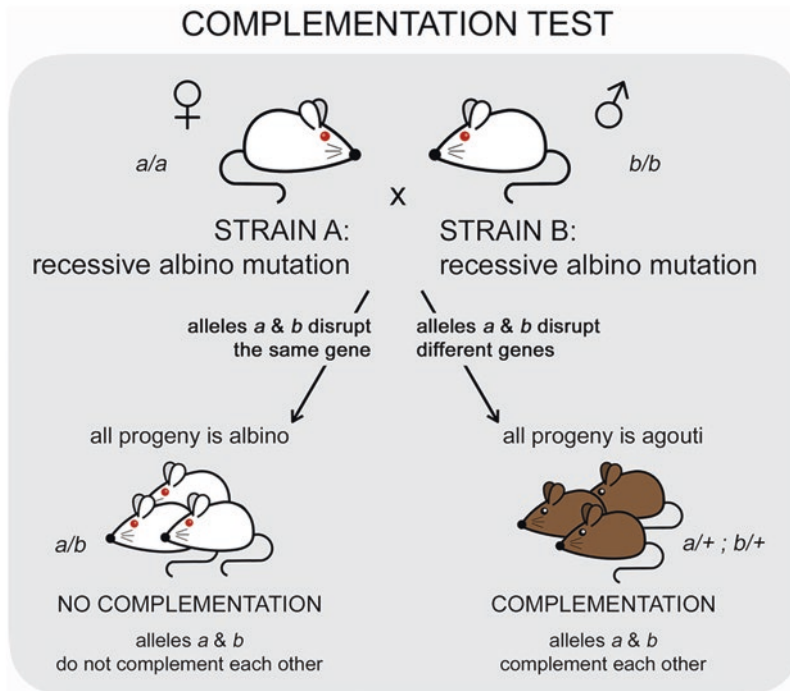
Fig. 1.4 *Complementation test.* By analyzing the F1 progeny from two animals carrying recessive alleles that cause the same phenotype, it can be determined whether the two alleles disrupt the same or different genes. In the illustrated example, if the *a* and *b* alleles causing albinism correspond to the same gene (left), albinism will be observed in the progeny and the alleles are said to not complement. If, on the contrary, the alleles correspond to different genes (right), mice with normal coat-color pigmentation will be observed in the progeny and the alleles are said to "complement" each other

in their biochemical properties (Fig. 1.5, bottom-left panel). Differences included changes in the electrophoretic mobility of proteins, in their enzymatic activity, their solubility in certain buffers, their thermal inactivation profile, their distribution in organelles, or their immunoreactivity [1, 5]. While the basis for these protein polymorphisms was thought to reside in allele variants for the genes encoding them, the actual genes and/or nucleotide changes were in many cases unknown.

Biochemical markers contributed to improving recombination maps by providing additional anchor points in the genome for linkage analysis. However, finding novel morphological or biochemical markers for linkage studies depended on serendipitous discoveries. As a consequence, the number of available markers remained an important limiting factor toward obtaining detailed linkage maps for many years. This situation changed dramatically in the 1980s and 1990s with the development of recombinant DNA technologies and improvements in DNA sequencing (reviewed in [1, Chapters 7 and 8]). By enabling the cloning, sequencing, and analysis of genomic sequences, these techniques led to the discovery of sequence differences between the DNA from different inbred strains (Fig. 1.5, right panel). These sequence differences, known as **DNA polymorphisms**, had two advantages over morphological and biochemical markers: they could be actively identified by comparing sequencing data between inbred strains, and they seemed to be distributed randomly throughout the genome, therefore providing a wide source of additional anchor points for linkage analysis. While DNA polymorphisms could be detected by sequencing, this approach was not practical for linkage analysis at the time, since sequencing methods were laborious and linkage analysis required testing hundreds of recombinant samples.
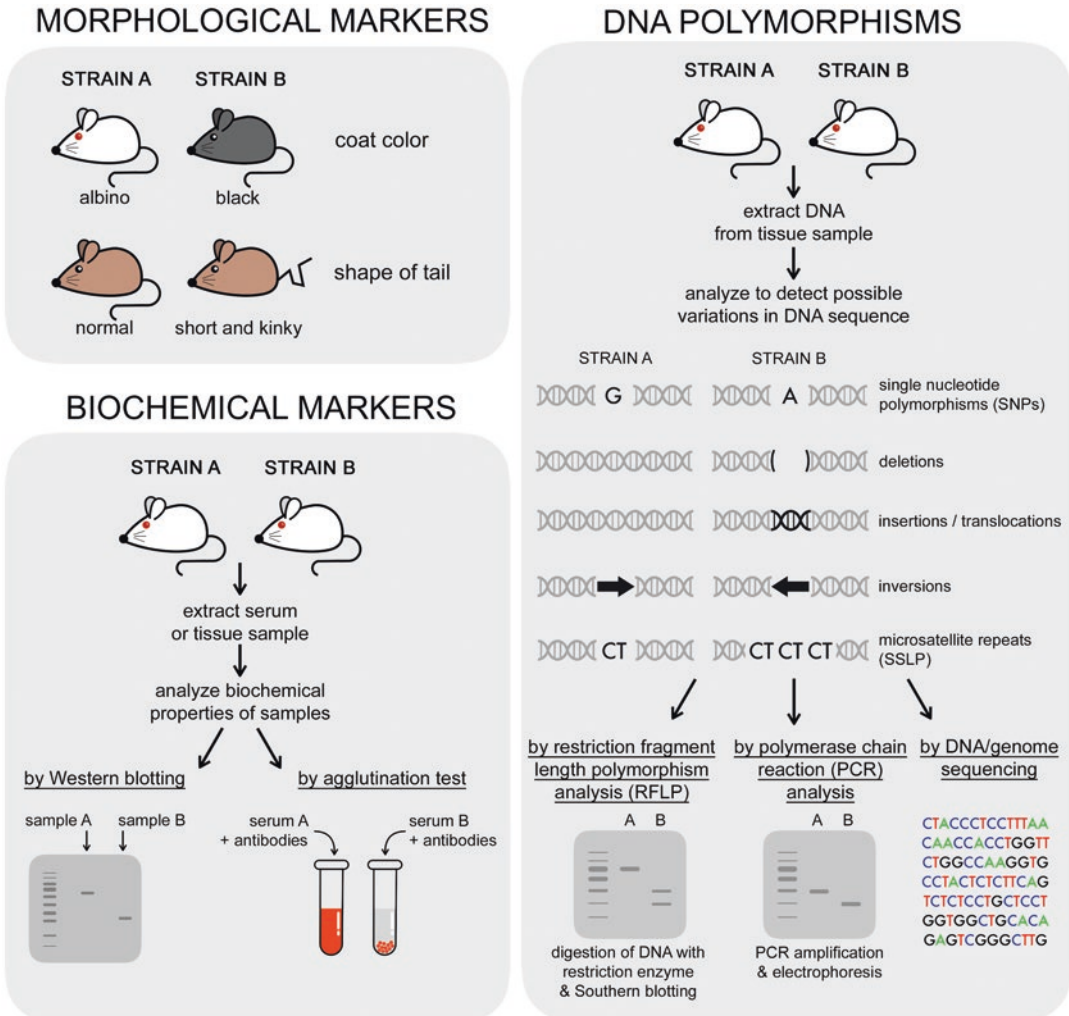
Fig. 1.5 *Markers for linkage analysis*. Linkage analysis necessitates detectable markers to establish the relative chromosomal location of genetic elements. Morphological markers (top-left panel) rely on phenotypes that can be directly observed in animals, such as coat color or the shape of the tail. Biochemical markers (bottom-left panel) are based on differences in the biochemical properties of tissue samples obtained from animals. The detection of these biochemical differences generally requires labora-tory tests such as Western blotting or agglutination assays. DNA polymorphisms (right panel) are based on differences in the DNA sequence, such as deletions, insertions, translocations, inversions, nucleotide changes (also referred to as single-nucleotide polymorphisms (SNPs)), and variations in the number of microsatellite repeats. Detection of DNA polymorphisms can be done by direct sequencing, RFLP analysis, or PCR-based methods

One of the first practical methods developed for the detection of DNA polymorphisms made use of the ability of restriction enzymes to cut DNA at specific sequences. The principle behind this method relies on the fact that sequence differences among strains might disrupt recognition sites for certain restriction enzymes. As a consequence, DNA polymorphisms can be visualized as **restriction fragment length polymorphisms (RFLP)**, differences in the size of the fragments that resulted from digesting genomic DNA from different strains with restriction enzymes. This approach was laborious, since the detection of restriction fragments required the use of Southern blotting with a probe located near the position of each known DNA polymorphism, but allowed

testing multiple samples in a single experiment. Another advantage of this method is that it allowed the detection of many types of DNA polymorphisms, not only single nucleotide changes, but also a variety of chromosomal rearrangements such as deletions, insertions, or translocations. As a consequence, the use of RFLP contributed significantly to improve the resolution of linkage maps.

With the popularization of polymerase chain reaction (PCR) methods in the 1990s, the detection of RFLP was greatly facilitated by eliminating the need to use Southern blotting to identify a particular genomic region. However, around this time, the use of RFLP as markers was relegated by the discovery of a new type of DNA polymorphisms involving repetitive genome sequences, which demonstrated unmatched benefits as markers for linkage analysis [22]. The most useful of these repetitive elements were **microsatellites**, genomic elements that contain mono-, di-, tri-, or tetrameric sequences repeated in tandem multiple times at specific locations in the genome. Microsatellite repeats do not have any known function and are thought to generate from recombination or replication errors at genome areas that are not critical for gene function. As a consequence, the number of tandem repeats at a given loci tends to vary among different laboratory mouse strains, making them ideal markers for linkage analysis. Also, microsatellites seemed widely distributed across the mouse genome and therefore could provide a widespread coverage of anchor points for linkage. On the practical side, it was easy to design PCR-based methods to detect microsatellite polymorphisms, also called **simple sequence length polymorphisms (SSLP)**. Similar to RFLPs, microsatellites were also easy to identify in the data that started outpouring from the sequencing of cDNA, bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs), and cosmid libraries. As a consequence, the number of available SSLP increased rapidly in just a few years. The Center for Genome Research at the Whitehead Institute/MIT led a systematic search for polymorphic microsatellite loci that could be used for linkage analysis [23]. The completion of this ambitious

project identified more than 6000 SSLPs and mapped them with respect to each other and with existing RFLP linkage maps, providing the first comprehensive linkage map of the mouse genome [24].

In parallel to the development of polymorphic markers, mouse geneticists worked toward finding efficient ways to test linkage and establish detailed maps. In its early days, linkage analysis entailed setting up breeding crosses between mouse strains carrying different morphological markers, then scoring the progeny for recombination events. However, a methodological breakthrough came in the 1970s, when Donald Bailey and Benjamin Taylor established **recombinant inbred (RI) strains** and conceptualized its use for linkage analysis [5, 25]. Recombinant inbred strains are obtained by crossing two known inbred strains and then establishing inbred colonies from the progeny. The resulting set of RI strains provide a collection of samples in which recombination events are preserved for future analysis through inbreeding (Fig. 1.6). Many RI strains, as well as genomic DNA samples of mice from these colonies, were maintained at Jackson Labs and were available to investigators for a small fee. Consequently, new markers could be mapped with respect to existing ones without the need to perform any breeding. Despite the convenience of RI strains, the scarcity of polymorphic markers at the time remained an important limitation to increase the resolution of existing linkage maps. In fact, almost two decades had to pass before a substantial number of DNA polymorphisms became available and RI strains could show its full potential for recombination mapping.

Meanwhile, investigators realized that the convenience of RI strains was tainted by the fact that most of the laboratory inbred strains used to generate them originated from just a few animals captured in the same geographical area and, as a consequence, their genomes were not very polymorphic. The discovery that fertile progeny could be obtained from **interspecific crosses** between laboratory strains (*Mus musculus*) and the distantly related species *M. spretus* [26] opened the possibility of using the genetic diversity between
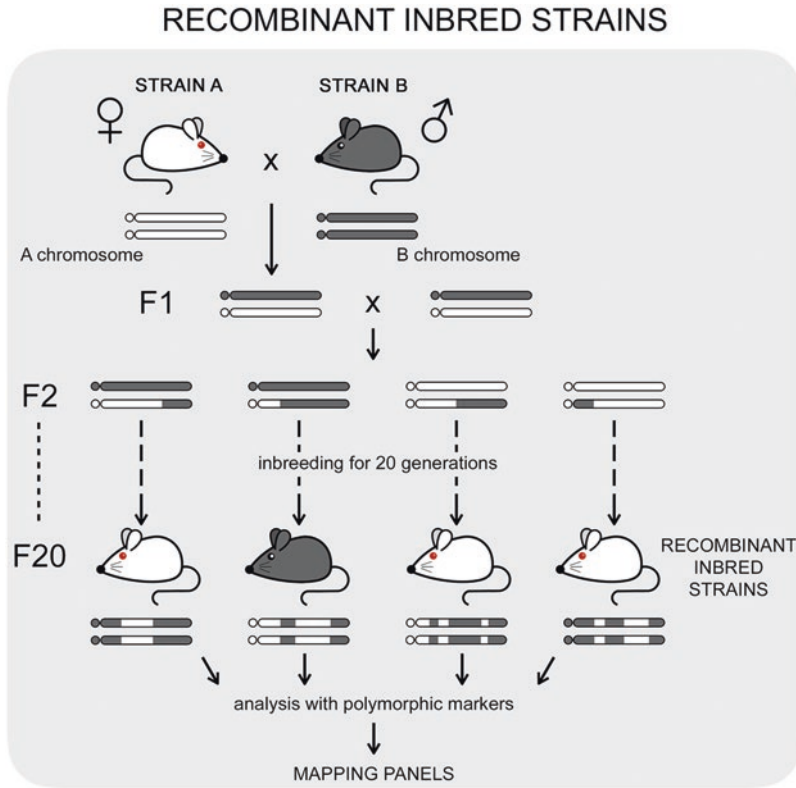
RECOMBINANT INBRED STRAINS

Fig. 1.6 *Recombinant inbred strains*. Recombinant inbred strains are generated by crossing mice from two previously established inbred strains, then performing brother-to-sister matings for 20 or more generations. Each of the resulting RI strains is genetically homogeneous and contains a mix of chromosome segments from the two original inbred strains. However, different RI strains differ in their genetic composition depending on the recombina-tion history of alleles in each of the breeding lines. The collection of DNA samples from the resulting RI strains constitutes a mapping panel. These samples can be tested with markers that are polymorphic between the original inbred strains, and the linkage relationships among the alleles from the A strain and the B strain can be used to construct linkage maps

species as a source of polymorphisms for linkage analysis [27]. Unfortunately, only the F1 hybrid females from interspecific crosses were fertile, but these could be backcrossed to males from laboratory strains, and DNA from the progeny could be preserved for analysis. Using this approach, several initiatives, including one at the Jackson Labs and another one in Europe (EUCIB—European Collaborative Backcross) performed interspecific backcrosses (with *M. spretus*, *M. castaneous*, and *M. domesticus*) and generated collections of DNA samples that, together with those obtained from RI strains, became known as **mapping panels** (reviewed in [28] and [1, Chapter 9]).

Mapping panels from interspecific crosses and RI strains played an important role in achieving a high-resolution linkage map of the mouse genome. By the 1980s, the amount of linkage information grew to a point where the "index card" system initially established by Margaret Green, and periodically published in the Mouse News Letter, became impractical [17]. To adjust to the demands of this research progress, Muriel Davisson and Thomas Roderick, at Jackson Labs, compiled all the existing information by 1990 and created one of the first computer-based mouse databases, the Genomic Database of the Mouse (Gbase). In 1992, the information from Gbase and other useful databases was compiled into a single

online portal, the precursor of today's **Mouse Genome Informatics** (MGI) (http://www.informatics.jax.org/). Since its inception, MGI has remained the most comprehensive database serving the international research community on mouse genetics, incorporating links to many useful Internet resources [29].

1.3.6 From Linkage Maps to Physical Maps

Genetic distances in linkage maps are measured in **centimorgans** (cM), an arbitrary unit that corresponds to the distance between two loci that segregate separately in the progeny at a frequency of 1 in every 100 individuals (which represents a crossover rate of 1%). While the frequency of recombination between two loci is roughly proportional to the length of DNA that separates them, numerous factors affect the frequencies at which recombination is observed and the interpretation of the results (reviewed in [1, Chapter 7]). For instance, loci separated by 50 cM or more have recombination frequencies similar to those of loci located in different chromosomes, making them appear as unlinked. Additionally, loci located far away from each other can undergo multiple crossovers, which skew the observed ratios of recombination in the progeny (i.e., an even number of crossover events between two loci produces the same allele combination as in the parental line and is therefore undetected). Another consideration is that recombination events within a chromosome are not independent of each other since the formation of a crossover site inhibits the initiation of additional recombination events nearby, a phenomenon known as **genetic interference**. As investigators became aware of these limitations, mathematical mapping functions were developed to correct for the effects of multiple crossovers and genetic interference ([14] and references therein). Nonetheless, as more linkage, cytogenetic, and sequence data became available, additional factors influencing recombination mapping were recognized. Among these, it was found that recombination sites are not randomly distributed across the genome: telomeric regions are more recombinogenic than are centromeric regions [30], and certain regions within chromosomes, known as **recombination hotspots**, have a higher incidence of recombination [31]. Recombination frequencies were also found to differ depending on the sex of the hybrid analyzed (recombination is higher in females than in males) and among different mouse strains [32, 33]. In recognizing these factors, it became clear that linkage and cytogenetic maps provided a comprehensive look at the mouse genome, but there were limits to their resolution, and therefore they could not substitute for a detailed physical map, where genes could be accurately placed in order onto chromosomes. As we will describe below, this accomplishment was made possible with the advent of molecular biology techniques, but would not be fully materialized until 2002, when the first draft of the mouse genome sequence was published [34].

1.4 The Molecular Biology Revolution and Mouse Genetics

The events that led to the birth of molecular biology and the publication of its central dogma in 1958 transformed the scope of genetic research [35, 36]. During the 1970s and 1980s, DNA cloning, DNA sequencing, nucleic acid hybridization, and the polymerase chain reaction made it possible to analyze the genome of any species with an unprecedented level of detail. Genes were no longer just alleles that manifested in different phenotypes; they could be identified as DNA sequences that were transcribed in specific tissues to produce proteins with specific cellular functions. As a result, natural alleles and induced mutations could now be analyzed at the molecular level as variations in the nucleotide sequence of genes that caused alterations in protein functions.

During the early years of molecular biology, libraries containing DNA fragments and complementary DNA (cDNA, DNA complementary to gene transcripts) were created in a variety of

vectors (bacterial plasmids, BACs, YACs, and cosmids), and sequences from these cloned DNAs were published in public repositories, including the European Molecular Biology Laboratory database (founded in 1980, currently part of EMBL-EBI), GenBank (founded in 1982, currently part of NCBI), and the DNA Data Bank of Japan (DDBJ) (founded in 1986), among others (reviewed in [37]). An important aspect of how these methods contributed to revolutionizing scientific research was that these repositories were all public: everyone could contribute their results to the databases, and archived sequences were available to anyone in the scientific community (although perhaps not as easily as we are used to nowadays since e-mail, the Internet, and the World Wide Web were not yet publicly available then). Also critical during these early years was the publication of the practical handbook *Molecular Cloning: A Laboratory Manual*, which, by offering detailed protocols, democratized the use of recombinant DNA techniques worldwide [38].

Sequencing information and molecular biology techniques had such a transformative impact on the field of mouse genetics that it is impossible to provide here a detailed account of the many techniques and approaches that contributed to this revolution. Nonetheless, we will mention a few highlights in the areas of linkage analysis, gene expression, and gene function. As discussed above, sequencing data provided a source of novel RFLP and SSLP polymorphisms that could be used to increase the resolution of linkage maps. Additionally, in situ hybridization techniques allowed the visualization of DNA in cytological preparations, enabling the mapping of genes and DNA sequences directly onto chromosomes [39]. As a result, these techniques made it possible to reconcile existing linkage, cytogenetic, and physical maps. Beyond linkage maps, molecular cloning and in situ hybridization techniques allowed investigators to determine that genes were transcribed in specific tissues and organs (reviewed in [40]), providing clues about their possible functions. Meanwhile, sequence comparison among different organisms revealed that many sequences and genes were evolution-

arily conserved across species, suggesting that research findings in a given organism could provide valuable information to determine gene function in another, an approach that later solidified in the creation of gene ontology databases [41]. In turn, these and other molecular biology contributions enabled the implementation of additional approaches toward the study of gene function. For instance, linkage maps became critical for the positional cloning of spontaneous and induced mutations. Additionally, the development of transgenesis and gene targeting approaches in the 1980s (see below) hinged on the ability of investigators to obtain and manipulate genomic sequences.

1.5   Manipulating the Mouse Genome: Making Mutants

Understanding the functional elements of mammalian genomes requires mechanisms to study how changes in DNA sequence and organization affect the physiology and/or reproduction of organisms. In the early years of mouse genetics, allele variants within natural populations and inbred strains were the only way to study the relationship between genes and phenotypes. Later on, the intense breeding programs carried out by mouse fanciers and research labs uncovered spontaneous mutations, providing additional genetic variants that could be correlated with disease outcomes and/or morphological differences [5]. However, investigators soon found more efficient ways to manipulate the genome by either using mutagenic agents, introducing exogenous pieces of DNA, or engineering customized changes in the genome's DNA sequence.

From a methodological perspective, two fundamental strategies have been historically used to study the effects of mutations (Fig. 1.7). **Forward genetics** is a phenotype-driven approach where naturally occurring or induced mutations are selected based on their phenotype, then further studied to determine the genetic and molecular causes for the morphological or physiological defects observed. Conversely, **reverse genetics** is
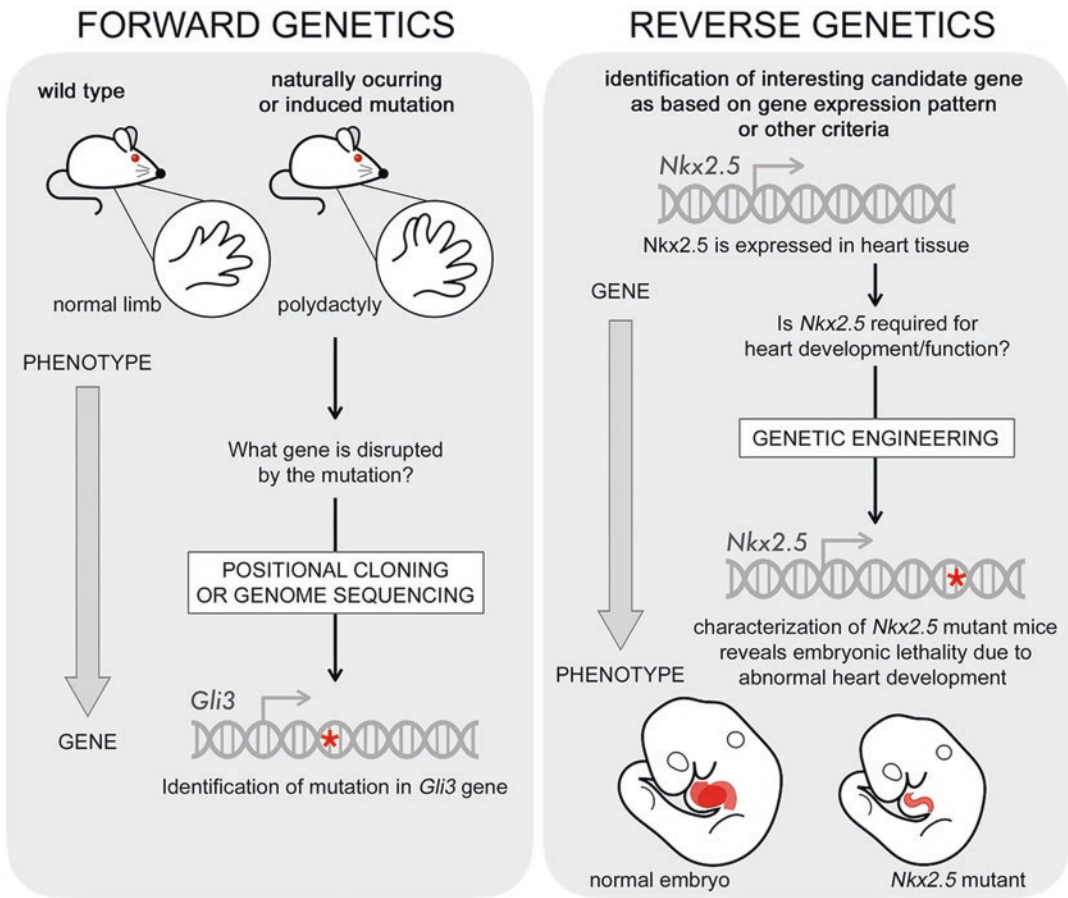
Fig. 1.7 *Genetic approaches to study gene function.* Genetic studies rely on the ability to link the genetic makeup of an organism (genotype) to its morphological or physiological constitution (phenotype). Forward genetic approaches (left panel) start with the analysis of naturally occurring or induced mutations that cause interesting phenotypes such as polydactyly, a condition that causes the appearance of extra digits in the extremities of mammals. Positional cloning or genome sequencing can be later used to identify the gene/mutation linked to the phenotype observed. Reverse genetic approaches (right panel) start with a known element of the genome, such as the gene *Nkx2.5*, and use genetic engineering methods to establish the effect of mutations disrupting that genetic element. Reverse genetic approaches are used by investigators to determine the function of genes that are interesting as based on previous research results or to establish mouse models of genetic human diseases

a DNA-driven approach that starts with the targeted disruption of a specific genetic element, then follows up on the study of the effects of this mutation on the *phenotype* of an organism. Forward genetic approaches do not require any previous knowledge of what sequences in the genome might be functional. Therefore, they constitute an unbiased strategy toward the discovery of novel elements in the genome. On the other side, reverse genetic approaches are ideal for studying previously identified elements of the genome whose functions are unknown. The development of techniques to support forward and reverse genetic approaches evolved in parallel since the 1950s, supported by discoveries in research areas as disparate as developmental biology, teratogenesis, and bacterial genomics, providing another example of how serendipity in research often promotes scientific progress in unanticipated ways.

1.5.1   The Power of Mutagens

The ability of radiation and certain chemicals to induce mutations was well known before the end of World War II from work on *Drosophila* and maize. However, bombings in Hiroshima and Nagasaki raised an interest in understanding the impact that atomic warfare and nuclear power plants could have on exposed individuals and their descendants. In the US, a big project toward this goal was initiated at the Biology Division of the Oak Ridge National Laboratories (ORNL) (reviewed in [42]). Mice offered an ideal model system for investigating the effects of radiation on mammals. Consequently, Bill Russell, who had trained with Sewall Wright studying phenotypic variability in inbred strains, was recruited to lead the operations in 1947. To test the rates of mutagenesis elicited by different mutagenic agents in germ cells, Russell introduced a methodological innovation known as the **specific locus test (SLT)**. This test involved crosses between mutagenized animals and "tester stocks" carrying alleles for seven different recessive markers with morphological phenotypes easy to distinguish by visual inspection. By scoring for the appearance of the recessive phenotypes in the F1 progeny, these crosses provided a standardized way to evaluate and compare the mutagenic rates of different types of mutagens and mutagenic regimes. In the early years, studies at ORNL centered on the effects of both external radiation sources (including X-rays, gamma rays, neutrons) and internal emitters (animals treated with radioactive isotopes such as tritium and plutonium). Later on, the successful platforms established at ORNL were also used to test the mutagenesis rates of chemicals. Studies with chemicals were initiated in the early 1960s and were greatly expanded in the 1980s, covering a wide spectrum of substances.

While the main focus of the ORNL programs was to study the effects of mutagens in female and male germ cells by using the SLT, work at ORNL spawned research in a variety of areas. Efforts to understand the effects of mutagens on testes and ovaries resulted in basic research on gametogenesis. Also, studies with embryos revealed that the early stages of embryogenesis were especially sensitive to the effects of radiation, a result that led to clinical recommendations for the practice of radiology on women of childbearing age. Perhaps the most influential contribution of the mutagenesis program at ORNL was that, as expected from such an intense use of mutagenic agents, lots of mutations and chromosome aberrations were obtained. From the onset, ORNL was committed to keeping mutants for their use in basic research projects. As a consequence, mouse genetics was no longer limited to the study of inbred strains and/or spontaneous mutations. Studies on some of the mutants obtained at ORNL contributed to important scientific discoveries, including the mechanism of sex determination in mice and the phenomenon of X-chromosome inactivation. Mutations obtained at ORNL were distributed to investigators worldwide for analysis and, as techniques were developed for freezing embryos and sperm [43–45], the ORNL devoted resources toward cryopreserving the entire ORNL stock collection for future investigation on the molecular effects of mutagens.

While the big genetic programs at ORNL were the first ones to be established, they were not the only ones. The United Kingdom initiated a similar mutagenesis program in the early 1950s, first located at the University of Edinburgh, then at Harwell. Focused on the analysis of chromosomal rearrangements, research at Harwell provided critical materials for cytogenetic analysis and genetic mapping [17]. Additionally, the Federal Republic of Germany recruited Udo Ehling to carry a chemical mutagenesis program at Neuherberg, near Munich in the mid-1960s [42]. Taken together, the use of mutagens represented the first methodology for investigators to manipulate the genome and generate mutations. The study of the resulting mutants highlighted the power of this approach to uncover the roles of the genome in regulating biological processes. Unfortunately, tools were not yet in place for investigators to be able to identify the genes disrupted by the mutations induced. However, once these tools became available during the 1980s and 1990s, the use of chemical mutagens for the

functional analysis of the mouse genome resurrected in the form of forward mutagenesis screens (see below).

1.5.2   Transgenesis: Introducing Exogenous DNA in the Mouse Genome

Starting in the early 1970s, a variety of methods for introducing foreign DNA into somatic and germ cell lineages of mice were developed. The first attempts used **viral DNA or retroviral particles** to infect early embryos [46], but shortly afterward techniques became available for introducing recombinant DNA into fertilized oocytes [47] and zygotic pronuclei [48]. Fundamental for the success of these techniques was the previous establishment of strategies to extract and manipulate embryos from pregnant mice, then reintroduce them into surrogate females. These strategies were developed during the 1950s and 1960s under the auspices of experimental embryologists who, motivated by their interest in understanding mammalian reproduction, required techniques to observe embryos outside of the uterus without disrupting their normal development. Thus, several developmental biologists contributed to optimizing protocols for growing two to eight cell embryos to the blastocyst stage in culture, aggregating cultured embryonic cells into chimeric embryos, and transferring cultured embryos into the oviduct of females (reviewed in [49]). Another embryonic manipulation that would become widely used for the generation of transgenic animals was **pronuclear injection**, a procedure that involved the injection of foreign genetic material directly into the pronucleus of fertilized mouse oocytes (Fig. 1.8; [50]). The success of pronuclear injection, and the exciting possibilities that transgenesis brought for genetic research, are exemplified in the fact that numerous groups adopted this technique just a year after it was first published [51–54]. Mice were the first organisms in which transgenesis was accomplished. Therefore, the establishment of these techniques constituted an important landmark that opened the door for genetic manipula-

tions in other organisms, including the generation of transgenic farm animals that could be used to produce large quantities of pharmacological compounds or that could be modified for improved agricultural productivity [55].

In mice, transgenesis provided a new tool for the analysis of gene function by allowing investigators to analyze the effects of **ectopic expression of genes** in a tissue and/or specific developmental stage. Spatial and/or temporal expression was usually accomplished by placing known enhancers or inducible promoters in transgenes [56]. Some applications of this strategy include studies on the oncogenic activity of certain genes (*Myc*, *Ras*), the analysis of immune responses to self-antigens, and the effects of developmental regulators (reviewed in [57]). In general terms, the ectopic expression of transgenic genes represents a **gain-of-function mutation**. However, transgenesis has also been used to generate **dominant negative conditions** by introducing mutated versions of genes (such as truncations or point mutations) that can sequester wild-type products in an inactive conformation (e.g., as inactive dimers; [58]). The ectopic expression of transgenes can also be used in the context of **complementation tests**, to evaluate whether candidate genes can rescue loss of function mutations [59]. Transgenesis was also used by developmental biologists in the context of **cell lineage analysis**, either by expression of reporter genes under the control of cell-/tissue-specific regulatory sequences [60]—such as the bacterial *lacZ* gene [61] or the gene encoding green fluorescent protein (GFP) [62]—or by ablation of certain cells through the controlled expression of toxic genes (e.g., diphtheria toxin; [63]).

While retroviral vectors for transgenesis have a limit to the length of the fragments that can be cloned into them, large fragments cloned in bacterial and yeast artificial chromosomes (BAC and YAC vectors) can successfully be integrated into the mouse genome through pronuclear injection (reviewed in [64, 65]). By allowing the integration of large genomic regions, transgenesis through pronuclear injection brought investigators a novel tool for the **identification of regulatory sequences**. The logic of this type of
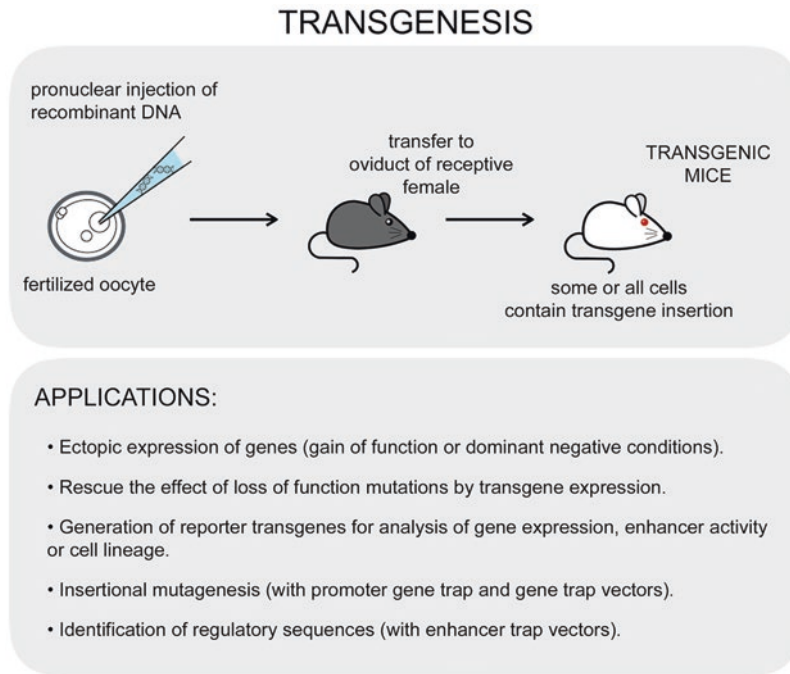
Fig. 1.8 *Transgenesis*. Efficient introduction of exoge-
nous genetic material into the mouse genome can be
accomplished by pronuclear injection of recombinant
DNA into fertilized oocytes. Injected oocytes are then
briefly cultured in vitro, then transferred to the oviduct of
females that are hormonally receptive to these embryos
(pseudopregnant females). The pups born from the pseu-
dopregnant female/s will be transgenic if the injected
recombinant DNA integrated into the oocyte genome. If
the integration takes place before the first mitotic division
of the oocyte, then all the cells of transgenic animals
would contain the transgene. If integration takes place
later during embryogenesis, then the transgene might only
integrate in some of the cells from the resulting transgenic
animals

experiments is that if a large transgene containing
a given gene of interest is expressed correctly
after integrating into a random location in the
genome, it can be inferred that the regulatory ele-
ments required for its expression were also pres-
ent in the transgene. If so, these elements can be
later localized by either deleting candidate
sequences from the original BAC/YAC clone or
testing the function of these candidate sequences
in the context of reporter transgenes. This strat-
egy was used extensively to identify regulatory
elements conferring spatial and/or temporal tran-
scriptional control in a variety of genes (reviewed
in [65]). This type of information contributed to a
better understanding of the mechanisms that reg-
ulate gene expression and identified regulatory
sequences that could be used to drive ectopic
gene expression in particular tissues or develop-
mental stages.

Another important application of transgenesis
related to the fact that the insertion of foreign
genetic material can disrupt genes or functional
elements of the genome located at the integration
site. As a consequence of this effect, called **inser-
tional mutagenesis**, many transgenesis
experiments unexpectedly led to abnormal phe-
notypes in transgenic embryos/animals [66]. In
these cases, since the mutagenic agent (the for-
eign DNA) remained integrated in the genome, it
could be used as a tag from which to clone the
genes disrupted. The overall frequency of inser-
tional mutagenesis was found to be relatively low
(7%, [65]). Nonetheless, at a time when there
were few mechanisms to identify the genes dis-
rupted by other mutagenic agents (such as radia-
tion or chemicals), insertional mutagenesis
provided investigators with a useful strategy to
characterize the functional elements of the

genome. This approach was later employed as the basis for large-scale gene-trap mutagenesis screens (see below).

1.5.3   Targeted Mutagenesis Through Homologous Recombination

While mutagens, including transgenes, provided mechanisms to manipulate the genome, the location where mutations were introduced was out of the control of the investigator. This situation changed in the mid-1980s with the development of techniques that allowed the modification of a target sequence of interest in a controlled fashion. Several lines of experimentation had to merge for the development of these techniques. The first critical step was the discovery that **pluripotent embryonic stem cells (ES cells)** could be isolated from early mouse embryos [67, 68] and that, when injected into blastocysts, these cells could contribute to any cell lineage in the resulting embryos, including the germline [69]. These findings inspired experiments to generate genetically modified animals by using ES cells that had been previously manipulated in culture, either by exposure to retroviruses or by transfection of DNA [70, 71]. Meanwhile, Mario Capecchi and Oliver Smithies were experimenting with the idea of whether **homologous recombination**, a process known to promote the exchange of DNA between DNA fragments with similar sequence, could be used to modify genes in mammalian cells. In 1987, both groups reported the successful use of homologous recombination to modify genes in ES cells [72, 73]. Shortly afterward, the first gene-targeted mice were born [74–78]. At this point, advances in recombinant DNA and molecular biology techniques had provided investigators with a wealth of information about cloned mammalian genes and, in some cases, their association to human diseases. Therefore, the possibility of using targeted mutagenesis to introduce mutations in any locus of interest opened the door to interrogating the function of any known sequence in the genome and/or to generate mouse mutants

that could serve as models to study human disease (reviewed in [79, 80]). Gene targeting was soon adopted by many investigators, and the number of mouse mutants obtained through this technology, which became known as **knockout (KO) mice**, grew exponentially during the last decade of the twentieth century. In recognition of the transformative impact that gene targeting had on the field of mouse genetics, Mario R. Capecchi, Martin J. Evans, and Oliver Smithies received the Nobel Prize in Physiology or Medicine in 2007.

The generation of knockout mice starts with the design of an appropriate targeting vector containing the desired gene modifications (Fig. 1.9). The vector is then electroporated into ES cells for homologous recombination to take place. Because the efficiency of homologous recombination is very low compared to the rate of transgenesis, careful selection of ES cell clones is required to identify those in which the desired locus has been modified. This step was initially very laborious. However, smart improvements in vector design made the selection process less cumbersome by introducing sequences that allow the positive selection of the cells that have incorporated the vector through homologous recombination and the negative selection of cells in which the vector has randomly integrated in the genome. Over the years, vector design incorporated additional modifications to facilitate the selection process and eliminate undesired effects at the targeted locus (reviewed in [81]).

Targeting vectors can be designed to introduce a variety of modifications at the target locus of interest, including deletions, point mutations, insertions, and sequence substitutions. This versatility enabled the use of targeted mutagenesis for a variety of applications, including the possibility of rescuing a mutant allele by replacing the mutated gene for a wild-type copy, the generation of mouse models of human disease through the introduction of point mutations identified in humans, and the generation of knock-in mice containing reporter alleles, to name a few [56, 81]. Despite this versatility, when inquiring about the roles of a previously uncharacterized gene, investigators usually chose to design targeting vectors toward the generation of null mutations,
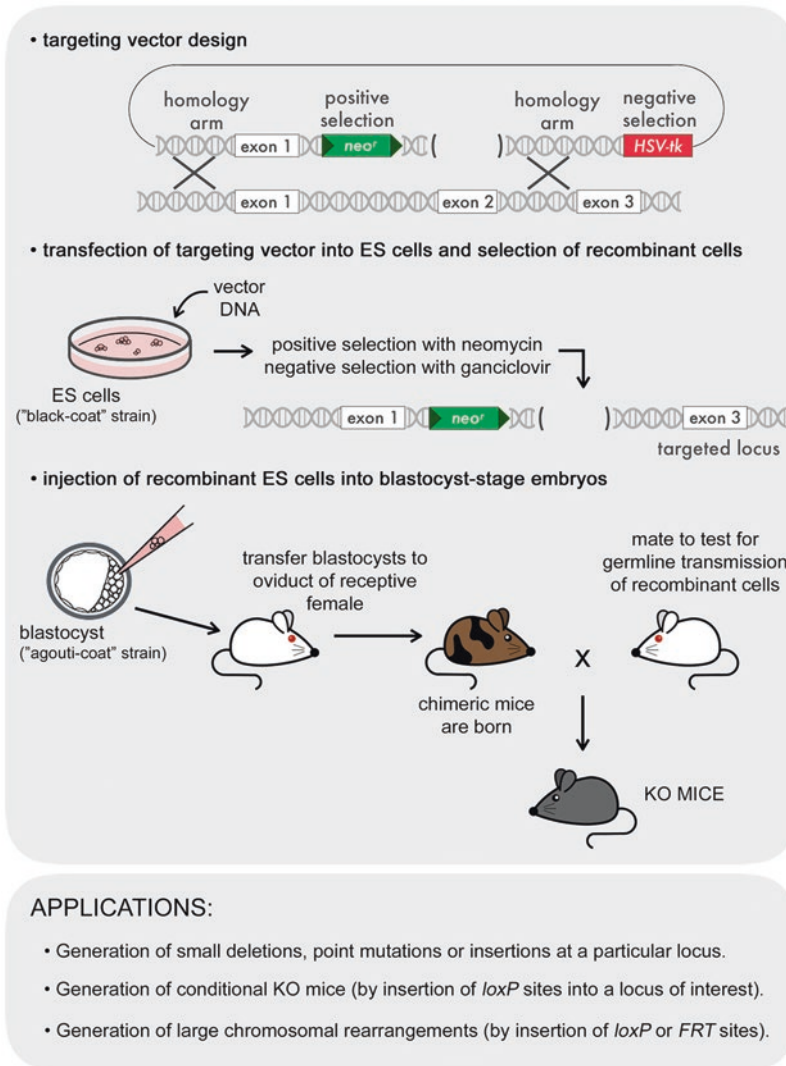
TARGETED MUTAGENESIS BY HOMOLOGOUS RECOMBINATION

• targeting vector design

homology arm — positive selection — homology arm — negative selection

exon 1 | neo^r | HSV-tk

exon 1 | exon 2 | exon 3

• transfection of targeting vector into ES cells and selection of recombinant cells

vector DNA

ES cells ("black-coat" strain)

positive selection with neomycin
negative selection with ganciclovir

exon 1 | neo^r | exon 3

targeted locus

• injection of recombinant ES cells into blastocyst-stage embryos

blastocyst ("agouti-coat" strain)

transfer blastocysts to oviduct of receptive female

chimeric mice are born

mate to test for germline transmission of recombinant cells

X

KO MICE

APPLICATIONS:

• Generation of small deletions, point mutations or insertions at a particular locus.

• Generation of conditional KO mice (by insertion of *loxP* sites into a locus of interest).

• Generation of large chromosomal rearrangements (by insertion of *loxP* or *FRT* sites).

Fig. 1.9 *Targeted mutagenesis by homologous recombination*. The generation of KO mice starts with the design and engineering of a plasmid targeting vector containing the DNA sequence with the desired modification/mutation to be introduced and a positive selection cassette (generally *neo^r*, which confers resistance to neomycin). These elements need to be flanked by two regions with complete homology to the locus to be genetically modified (homology arms). Also, a negative selection cassette located after one of the homology arms (generally *HSV-tk*, which confers resistance to ganciclovir) is needed to select against cases where the targeting vector does not undergo homologous recombination but, instead, integrates randomly in the genome. The targeting vector is electroporated into ES cells and, after positive and negative selection, recombi-nant ES cell clones carrying the targeting vector into the locus of interest are injected into blastocyst-stage embryos. These embryos are then transferred to pseudo-pregnant females. The pups born from these females are chimeric, bearing cells from the blastocyst that was injected and from the ES cells that were introduced. The genetic makeup of ES cells and blastocysts can be chosen to provide coat-color markers that can facilitate the assessment of chimerism and selection of KO mice. In the illustration, "black" coat color marks cells of ES cell origin, while agouti coat color marks blastocysts cells. To further study the effects of the mutation, chimeric mice must be able to transmit the modifications to their progeny (germline transmission)

generally by eliminating one or more exons of the targeted gene. The phenotypes of null mutations were often quite unexpected, offering lessons of humility to investigators who, eager to find phenotypes in tissues where the targeted gene was known to be expressed, were confronted with finding no phenotype at all or unpredicted phenotypes, such as embryonic defects and/or lethality. These unforeseen effects highlighted the fact that the function of mammalian genes is sometimes redundant with closely related genes, and therefore phenotypes are not obvious unless two or more genes are knocked out simultaneously. In other cases, mammalian genes have pleiotropic functions at different developmental stages, and therefore early phenotypes preclude the analysis of later functions.

To address the analysis of pleiotropic gene functions, the design of targeting vectors was refined such that gene function would only be altered in specific tissues and/or at precise developmental timepoints. One of these refinements was the use of the *Cre/loxP* recombinase system to generate **conditional knockout mice**. This system is based on the ability of the *Cre* recombinase from the P1 bacteriophage to excise any region of DNA placed between two recognition motifs called *loxP* sites. Consequently, the two elements of this system, *Cre* recombinase and *loxP* sites, need to be introduced into mice for the generation of conditional knockouts. On one side, homologous recombination is used to place two *loxP* sites flanking an essential exon of the gene to be knocked out. If done properly, the resulting mice would contain a **"floxed" allele** in which the *loxP* sites do not interfere with the normal transcription or splicing of the gene. Mice with a floxed allele are then mated to transgenic animals in which the *Cre* recombinase is expressed under the control of tissue-specific enhancers. As a result, excision of the floxed allele will only happen in specific tissues of the progeny. As more labs adopted this strategy to analyze the function of genes expressed in a particular tissue of interest, a variety of *Cre* lines became available. Many of these lines can now be obtained through public repositories [82]. Variations of the *Cre/loxP* approach employing other recombinases (*FLP-*

FRT system) or incorporating inducible gene expression systems (tamoxifen or tetracycline-dependent expression) provided alternative methods and additional versatility for the conditional inactivation of genes. Another interesting application of the *Cre/loxP* system was the **engineering of chromosomal rearrangements** such as large deletions, duplications, inversions, and translocations, some of which could be used as mouse balancer chromosomes [83].

1.5.4 Genome Engineering with Endonucleases: CRISPR/Cas9 Engineering

For more than 20 years, homologous recombination remained the only reverse genetic approach to purposely target a known element of the mouse genome. However, in the early years of the twenty-first century, endonucleases emerged as powerful tools for gene editing. Endonucleases work by generating double-strand breaks (DSBs) in the DNA, thereby triggering one of several DNA repair mechanisms that are endogenous to cells. Nonhomologous end joining is an error-prone repair mechanism that frequently leads to the production of small insertions or deletions (indels), which can potentially disrupt genes or other functional elements of the genome. DSBs can also be repaired through high-fidelity homology-directed repair mechanisms. In normal conditions, homology-directed repair uses a sister chromatid as template, but this repair system can be deceived to use a single-stranded or double-stranded DNA cointroduced into the cell, as long as it bears homology to the locus being repaired. Therefore, by delivering simultaneously a nuclease with an alternative repair template containing mutations, any desired sequence change, such as nucleotide substitutions, deletions, or insertions, can be introduced at or near the induced DSB. Key to the use of endonucleases for targeted mutagenesis was the development of methods to direct these enzymes to introduce DSBs exclusively at a desired locus in the genome. This has been accomplished through different strategies.

The first endonuclease systems used for gene editing were **zinc-finger nucleases (ZFNs)** and **transcription activator-like effector nucleases (TALENs)**. Both ZFNs and TALENs are modular enzymes that contain the bacterial FokI endonuclease domain fused with a DNA recognition motif that can be engineered to recognize any known sequence in the genome (reviewed in [84]). ZFNs and TALENs were successfully applied for gene editing in a variety of experimental systems. However, their use for gene editing was eclipsed by the difficulties associated with the design of specific DNA recognition motifs and the advent of a novel and more versatile gene editing system. This new gene editing method was adapted from a bacterial locus that confers adaptive immunity against bacteriophages and comprises three different elements: (1) an array of **clustered regularly interspaced short palindromic repeats (CRISPR)**, which contains sequences derived from bacteriophages or other invading genetic elements and is transcribed to produce a CRISPR RNA (crRNA); (2) a nuclease, encoded by nearby CRISPR-associated genes (**Cas**); and (3) a trans-activating crRNA sequence (tracrRNA), which is transcribed into an RNA complementary to parts of the crRNA and can recruit the Cas nuclease. These three elements form a functional Cas-crRNA-tracrRNA complex able to recognize and digest exogenous DNA with sequences complementary to those in the crRNA sequence, thereby protecting bacteria from the harmful effects of a phage infection. While three different types of CRISPR/Cas systems have been described [85], the type II **CRISPR/Cas9 system** has been adopted for gene editing due to its high efficiency and adaptability to a variety of organisms (reviewed in [86]). In 2013, the CRISPR/Cas9 system was successfully used for the first time to edit the genome of mouse and human cells [87, 88]. This was accomplished by transfecting cells with plasmids encoding the Cas9 and an engineered guiding RNA (gRNA), which contained a reprogrammed crRNA with sequence complementary to a 30 bp unique target site in the genome and an 89 bp tracrRNA (Fig. 1.10). Since then, the technique has been further developed

for its application to a wide spectrum of model organisms. Additionally, protocol improvements have been introduced to provide better efficiency, target specificity, and to favor homology-directed repair [89].

In mice, it was found that co-injecting the components of the CRISPR/Cas9 system directly into one-cell embryos can result in gene editing, either through nonhomologous end joining or through homology-dependent repair mechanisms (Fig. 1.10; [90, 91]). This finding revolutionized gene targeting in mice since it allowed the generation of CRISPR-edited animals, referred to as **CRISPRed mice**, as early as 6 weeks after embryo injections. Consequently, CRISPR/Cas9 editing offers a faster and simpler one-step protocol as compared to the process of obtaining KO mice, which takes several months of cumbersome vector design, ES cell selection and injection into blastocysts, followed by selection of animals with germ-line contribution. Also, the CRISPR system offers additional advantages over targeted mutagenesis by homologous recombination in ES cells. First, the high targeting efficiency of CRISPR/Cas9 makes it possible to edit the two homologous chromosomes in a single procedure, facilitating the analysis of recessive traits. Also, by injecting multiple gRNA constructs, CRISPR/Cas9 can be multiplexed to accomplish the simultaneous targeting of several loci, making it possible to generate double or triple mutant mice directly. While CRISPR-induced editing is not yet exempt from a few experimental pitfalls [89], its application in one-cell mouse embryos was found to be highly specific, alleviating the concerns about possible off-target effects that had been raised in other experimental settings [91, 92]. Moreover, the last few years have seen the publication of protocol variations, such as CRISPR-EZ, that continue improving the fidelity, efficiency, and versatility of this system [93]. At present, CRISPR/Cas9 has been successfully used to generate indels that disrupt gene function, to introduce subtle genomic modifications such as point mutations, to insert exogenous sequences such as those that allow to generate conditional floxed alleles or epitope tags [92, 94], and to generate relatively large deletions (up to
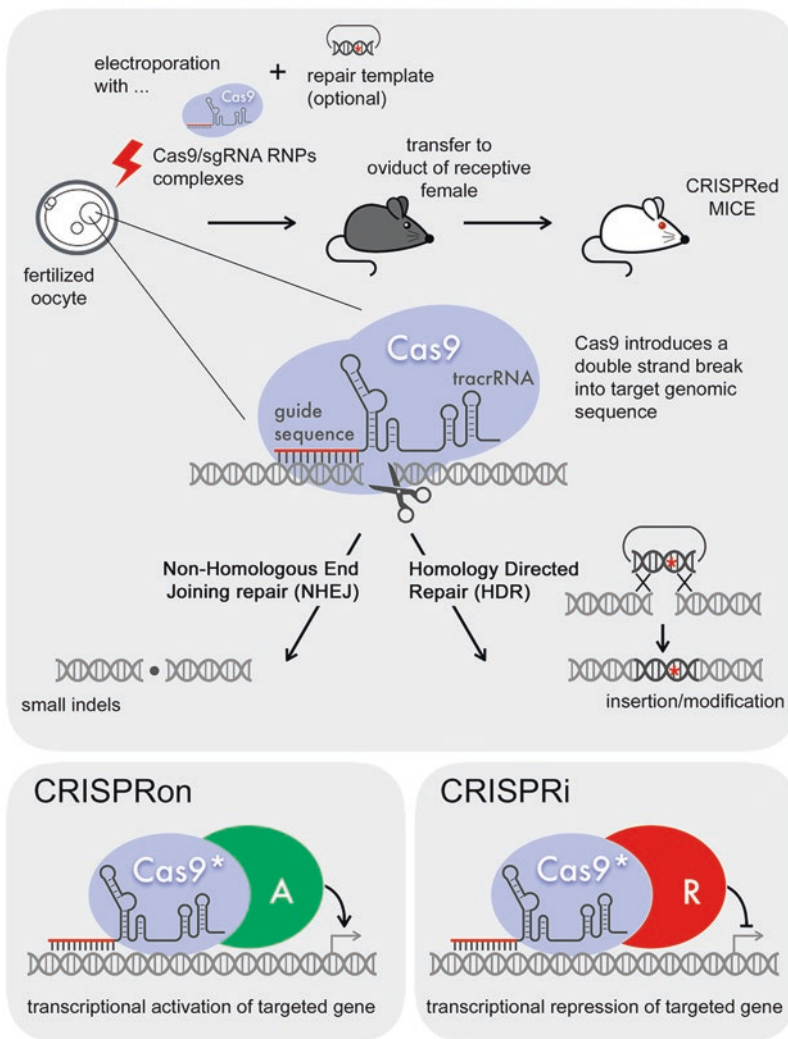
Fig. 1.10 *CRISPR/Cas9 engineering*. Adapted from a bacterial immunity system against bacteriophages, CRISPR/Cas9 genetic engineering is based on the ability of the Cas9 endonuclease to introduce double-strand breaks (DSB) into the DNA (upper panel). Cas9 (light blue) can be targeted to particular genomic loci when cotransfected with RNA molecules that contain an area of homology to the target locus (guide RNA sequence) and a specific RNA sequence able to interact with Cas9 (tracrRNA). DSB introduced by Cas9 can be repaired through one of two available repair mechanisms endogenous to cells: the nonhomologous end-joining repair (NHEJ) pathway, which frequently causes small deletions or insertions (indels) at the repaired site, or the homology-directed repair (HDR) pathway, which uses sequences with homology to the affected locus to repair the damage. Normally, homologous chromosomes serve as templates for HDR, but alternative repair templates (plasmids containing mutations or modifications) can be provided experimentally. In mice, electroporation of Cas9/sgRNA ribonucleoprotein complexes into fertilized oocytes (with or without repair templates) can efficiently promote genome editing. Electroporated oocytes can be cultured in vitro to the blastocyst stage, then transferred to pseudopregnant females. Following this procedure, CRISPR-edited (CRISPRed) mice can be obtained in about 6 weeks. CRISPRon (lower left panel) and CRISPRi (lower right panel) are alternative applications of the CRISPR/Cas9 system. By using versions of Cas9 that lack endonuclease activity and are fused to a transcriptional activator (CRISPRon) or a transcriptional repressor (CRISPRon), these systems can respectively produce transcriptional activation or repression at the locus targeted by the cotransfected gRNA sequence

10 kb, [95]). Additionally, modified versions of the CRISPR system have been recently developed for applications beyond genome editing. Examples are the CRISPR-on and CRISPRi (Fig. 1.10, lower panels), which constitute tools for targeted regulation of gene expression by using Cas9 variants that lack nuclease activity and are tethered to transcriptional activator or repressor proteins [96–98].

1.6 The Mouse Genome Sequence

In 1985, a group of scientists, summoned by Robert Sinsheimer, met at the University of California Santa Cruz to discuss the feasibility of sequencing the whole human genome. This meeting was the first effort toward what ended up crystalizing in 1990 as the **International Human Genome Project**. While the ultimate goal of the project was to produce a complete assembly of the human genome, it was decided from the outset that the project should also include the analysis of other species, including mice [99]. The benefits of including other species were seen as double: on one side, sequencing smaller genomes, such as those of bacteria, yeast, *Drosophila melanogaster* (a fruit fly) and *Caenorhabditis elegans* (a nematode worm), would serve as a proof of principle that the DNA sequencing technology, as well as the computational methods required for the alignment and assembly of the resulting sequence, were ready to handle more complex genomes. On the other side, understanding the functional roles of the human genome would benefit from comparative studies among different organisms. The Human Genome Sequencing Project, which published its first complete draft in 2001, has arguably been one of the most ambitious scientific endeavors undertaken by humankind and one of exemplary international cooperation [100, 101]. Nonetheless, this ambitious project was not exempt from many scientific and political issues, which have been the topic of multiple divulgation books [102, 103].

The **Mouse Genome Sequencing Consortium** was created in 1999 and ran in par-

allel to the sequencing of other genomes. Initially, the strategy chosen for accomplishing a high-quality sequence of the mouse genome was similar to the one devised for the human genome and included two phases: in the first phase, efforts focused on improving the resolution of linkage maps with additional DNA polymorphisms and on using this information as a blueprint to map the chromosomal location of DNA fragments cloned into a variety of vectors (including expression libraries, BACs, YACs, and cosmids). A second phase comprised the sequencing of the DNA fragments in these libraries, its assembly into contigs (sets of overlapping sequences corresponding to a large genomic region), and the filling of the gaps between contigs to accomplish a contiguous sequence for each chromosome. While these were the initial plans, lessons learned from the "public" *Human Genome Project* and the "private" sequencing ventures initiated by the company *Celera Genomics* suggested that an alternative sequencing strategy known as **shotgun sequencing** could significantly accelerate genome sequencing. Shotgun sequencing relies on the use of computational approaches to align sequencing results from random clones and DNA fragments, whose location in the genome is initially unknown. By eliminating the need for mapping the location of each DNA clone within the genome, this strategy proved to be a useful method to assemble a first draft of the genome quickly. Nonetheless, it was found that the presence of highly repetitive sequences in complex genomes complicates the computational alignment of shotgun sequences. As a consequence, the mouse genome sequence was obtained through a diversified strategy that involved both shotgun sequencing and the sequencing of DNA fragments previously mapped to existing linkage maps. By combining the benefits of these two types of approaches, the first high-quality assembly of the mouse genome was accomplished in 2002 [34]. This achievement marked a new era in mouse research, facilitating the application of both forward and reverse genetic approaches toward the functional characterization of all the genes in the genome.

Following the completion of the mouse genome project, the increased availability of DNA sequence from different mouse strains led to the discovery of a new type of sequence variation among inbred strains, termed **single-nucleotide polymorphisms (SNPs)**. At present, SNPs constitute the best type of DNA polymorphic markers for linkage analysis due to their dense distribution across the genome, their high variability among mouse strains, and the availability of multiplexed genotyping platforms that simplify their detection in mouse samples [104, 105]. Because the discovery and detection of SNPs was only possible after technological advances allowed cheap and reliable genome sequencing, SNPs did not play a significant role in the initial development of accurate linkage maps. Nonetheless, the unsurpassed density of SNPs between different inbred strains made this type of polymorphisms stand out as powerful tools for the positional cloning of mouse mutations, as we will discuss below. To fully characterize molecular variations between the most common inbred strains, the **Mouse Genomes Project** was launched in 2009. Since then, sequence from more than 35 different inbred strains has provided a catalog of SNPs, as well as other genetic variants such as short indels and transposable elements [106].

1.7   A Mutant in Every Gene: Large-Scale Approaches to Study Gene Function

The use of mutagens and, later, the development of transgenesis and targeted mutagenesis accelerated the pace of mouse genetic research by giving investigators the tools required to obtain mutations that could inform about the roles of the different functional elements in mammalian genomes. However, as indicated above, each of these methods to manipulate the genome presented a different set of strengths and limitations, making it clear that understanding the functions of every gene in the genome would require complementary strategies. Up to this day, investigators have grappled with choosing the mutagenesis

method most appropriate to address their particular research goals. In the process, technical improvements in targeted mutagenesis and transgenesis, as well as progress from the mouse genome sequencing project, opened venues for large-scale mutagenesis efforts. As a consequence, the ambitious goal of obtaining mutations for each of the genes in the mouse genome became feasible. Below we review how large-scale chemical mutagenesis, transgenesis, and genetic engineering have contributed toward this goal, which is expected to be accomplished in 2020.

1.7.1   Forward Mutagenesis Screens and Positional Cloning

Results from research at ORNL identified N-ethyl-N-nitrosourea (ENU) as one of the most powerful mutagens in mice [107]. This finding raised the possibility of using ENU to perform genome-wide mutagenesis screens akin to those performed in other organisms, such as the Nüsslein-Volhard and Wieschaus screens in fruit flies, which would be later recognized with a Nobel Prize in 1995 [108]. The promise of using ENU for forward genetic approaches was initially diminished by the fact that linkage maps during the 1980s were still rudimentary, and many investigators feared that the identification of the point mutations responsible for the phenotypes recovered would be extremely difficult. Despite this, a few early investigators pioneered the use of ENU toward the identification of alleles at the *t complex* [109, 110], a genomic region that had long been subject to intense investigation due to its importance in the control of histocompatibility and embryonic development [10]. The success of these projects, together with improvements in linkage maps and the discovery of SSLP polymorphisms in the early 1990s, resurrected the enthusiasm in using **large-scale ENU-based forward genetic approaches** to uncover the functional elements of the mouse genome. This enthusiasm reinvigorated even further with the publication of a few additional successful ENU mutagenesis projects, which identified dominant-

effect point mutations causing tumorigenesis and circadian clock phenotypes [111–114]. As a consequence, several large-scale mutagenesis screens were initiated worldwide (reviewed in [115–117]).

The first large-scale ENU-based screens focused on the identification of **dominant phenotypes**, given the simpler breeding scheme required for the detection and perpetuation of dominant mutations. However, **recessive screens**, which require three generations of crosses before animals can be screened, were launched shortly afterward. Because forward genetics is a phenotype-driven approach, the establishment of reliable **phenotyping platforms** for the analysis of mutants is a must. The SHIRPA platform set the groundwork for the systematic assessment of a wide gamut of physiological and behavioral parameters [118, 119], but each project introduced its own screening protocol as based on its particular interest. In fact, individual forward mutagenesis projects focused on the identification of mutations causing particular phenotypes ranging from the identification of neurological defects to hematological conditions, behavioral anomalies, or developmental malformations (reviewed in [115, 116]). Projects also differed in their scope. Some projects run **genome-wide screens,** while others performed **focused screens** on specific areas of the genome. Focused screens were accomplished by mating mutagenized animals to mice carrying chromosomal rearrangements (deficiencies or balancer chromosomes), a strategy that allows the fast identification of mutations affecting a particular chromosome or genomic regions [120–126]. A few labs carried out **small-scale ENU screens** that, while not achieving a full saturation of the genome, proved really successful in identifying novel genes and pathways involved in specific biological processes, including embryogenesis, immunity, and neuronal development [127–135]. Other labs embarked on **sensitized screens**, mating mutagenized animals to known mutants, then screening for mutations that could either enhance or suppress their phenotypes [136, 137]. As a whole, ENU-based screens demonstrated that forward genetic approaches, regardless of their scale, are really valuable for the unbiased discovery of genes and pathways involved in any biological process for which a reliable phenotyping method can be established.

The identification of the point mutations responsible for the phenotypes obtained in ENU screens was initially accomplished through a process known as **positional cloning** (Fig. 1.11). For this, screens used a breeding strategy that involved two different inbred strains: first, ENU was injected in mice of one strain (strain A), then these animals were mated to a different strain (strain B), and the progeny was screened for interesting phenotypes. Selected carriers for these phenotypes were then systematically outcrossed to the second strain until establishing congenic mutant lines. As a consequence of the outbreeding process, the content of DNA from strain B increases throughout the genome, except for a small chromosomal interval from the mutagenized strain (strain A) selected to carry the mutation of interest. In this way, linkage analysis to the mutagenized strain can be used to identify the chromosomal region containing the mutation. For the early pioneers of ENU mutagenesis, mapping mutations to small chromosomal intervals was extremely laborious, given the inaccuracies in linkage maps and the scarcity of polymorphic markers. Even more challenging was to identify the genes mapping to the particular chromosomal interval and to sequence candidate genes in search of ENU-induced mutations [138–140]. However, these challenges disappeared as improvements in linkage maps and in the availability of polymorphic markers, including SNPs, streamlined positional cloning. More recently, the development of next-generation sequencing methods has significantly reduced the cost of sequencing whole genomes, making it possible to sequence all transcribed sequences in samples from ENU-induced mutant animals and directly identify the causative mutations, even without the need for positional cloning [141–144].

An attractive part of ENU mutagenesis is that it introduces point mutations into DNA, and therefore, as opposed to homologous recombination or transgenesis, it can uncover hypomorphic
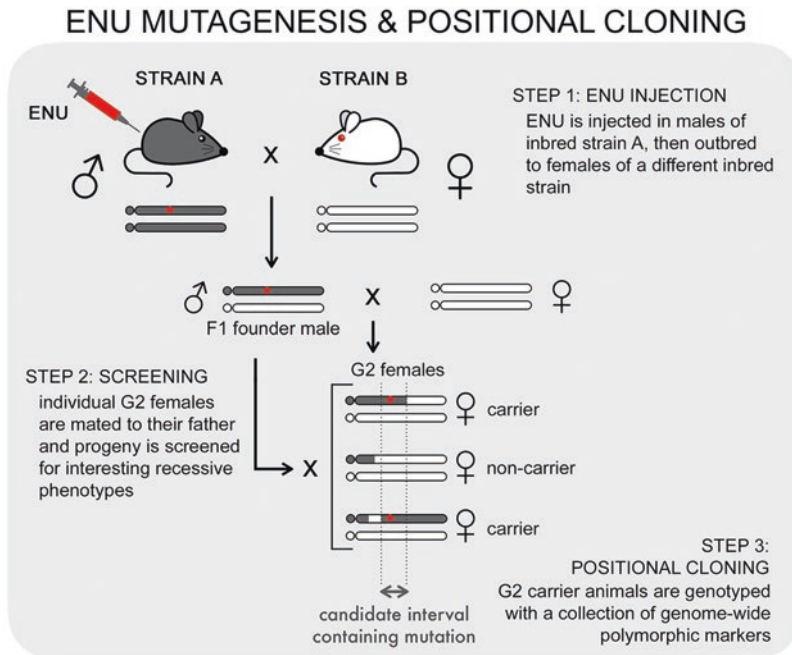
Fig. 1.11 *Large-scale ENU mutagenesis and positional cloning*. The different steps of a forward genetic chemical mutagenesis approach are illustrated for a screen aimed at identifying recessive mutations. First, a chemical mutagen, such as *N*-ethyl-*N*-nitrosourea (ENU), is injected into male mice of a particular inbred strain (strain A). Injected males are then mated to females of a different inbred strain (strain B) to propagate the germline mutations caused by ENU (red asterisk). Polymorphisms between the DNA of these inbred strains (illustrated as black and white chromosomes) will later facilitate the identification of the mutations. Males from the progeny (F1 founder males) are used to establish independent colonies and screen for interesting mutations. For this, F1 founder males are crossed to inbred females of strain B, then females from their G2 progeny are mated back to their father. The progeny from these G2 females is then screened for interesting phenotypes in embryos or adults. The logic of this breeding scheme is that if the F1 founder male is a carrier of an interesting mutation, 50% of his progeny (G2 animals) will also carry that mutation, and when they are mated back to their father, they will produce progeny-carrying recessive traits in homozygosis. Therefore, by establishing random crosses between G2 females with their father and selecting those G2 females that produce interesting and reproducible phenotypes, a collection of G2 animals that are heterozygote carriers for the mutation can be identified. The use of different inbred strains in the breeding scheme implies that recombination in the germ line of F1 founder males (and subsequent generations) produces recombinant chromosomes that contain DNA from strains A and B. Positional cloning is based on the fact that ENU mutations will be linked to DNA originating from strain A. Therefore, DNA from selected carriers can be genotyped with a collection of genome-wide DNA polymorphic markers, and the genotype of these animals can help identify linkage of the mutation to a particular chromosomal region. Genes within this interval are candidates to contain ENU-induced mutations and can be sequenced to identify the point mutation responsible for the phenotype identified through screening

and dominant alleles, mutations more similar to those arising spontaneously in humans. ENU also has a wide spectrum of genomic targets, and although biases have been observed, including mutagenic hot spots and a preference for transcribed genomic regions [145], it can lead to the identification of genes located anywhere in the genome without prior knowledge of their existence. Consequently, the greatest advantage of ENU mutagenesis is that it provides an unbiased strategy to identify essential genes whose functions would be difficult to uncover using hypothesis-driven approaches. An example of this was the discovery of the cilium as a cellular structure required for signal transduction in mammalian cells (reviewed in [146]).

1.7.2 Gene-Trap Mutagenesis Screens

The finding that the random insertion of transgenes in the genome could disrupt gene function [66] and that the transgene insertion could facilitate the identification of the integration site [147, 148] motivated the use of transgenesis for what became known as **insertional mutagenesis**. However, as explained above, the most efficient transgenesis method initially involved pronuclear injection of exogenous DNA into fertilized oocytes, a demanding and time-consuming process that was not optimal for scaling up the generation and screening of transgene insertions. As a consequence, large-scale insertional mutagenesis was not possible until a few technical improvements came into place.

The first breakthrough that made large-scale insertional mutagenesis screens possible was the establishment of ES cells as useful platforms for transgenesis [149]. ES cells offered several advantages: first, exogenous DNA could be introduced efficiently through electroporation or retroviral infection. Second, ES cells could be grown in multiplexed platforms, facilitating the generation, screening, and characterization of new insertions. Third, transgenic ES cell lines could be kept frozen until the insertion sites could be characterized. Fourth, the identification of the insertion site for each ES cell line could be easily accomplished through procedures such as **plasmid rescue** [147] or **rapid amplification of cDNA ends** (RACE) [148], both of which used vector sequences as entry points to identify the adjacent genomic sequences. Last but not least, ES cells selected to contain interesting insertions could be injected into early blastocyst-stage embryos, allowing the analysis of the resulting chimeric embryos and/or the selection of chimeras with germ-line transmission of the transgene, ultimately making it possible to test whether the transgene insertion caused any abnormal phenotype in animals. Because of these numerous advantages, ES cells were soon adopted as platforms for large-scale insertional mutagenesis screens [150], establishing this technique as a powerful method to systematically obtain and catalog mutations in each of the genes in the mouse genome (Fig. 1.12, upper panel).

Also critical for the success of large-scale insertional mutagenesis screens were improvements in the design of DNA vectors that could disrupt gene function with high efficiency (Fig. 1.12, lower panel; reviewed in [151]). The first vectors used for mouse insertional mutagenesis derived from plasmids originally employed for **enhancer-trap screens** in flies [152]. These enhancer-trap plasmids contained the bacterial *lacZ* reporter gene under the control of a weak promoter, plus a marker that allowed the selection of animals in which the transgene had successfully integrated into the genome. The weak promoter was insufficient for the detection of reporter gene expression, unless the plasmid integrated in the vicinity of a transcriptional regulatory element, in which case *lacZ* would be expressed with the temporal and/or spatial expression pattern dictated by the "trapped" enhancer. The use of these vectors in mice led to the identification of transgenic animals that expressed *lacZ* in a variety of tissue-specific patterns [150, 153, 154]. Similar transgenic experiments were later performed with promoter-less reporter vectors that, when inserted in frame within the exon of a gene, could simultaneously report the expression pattern of the gene and disrupt its function, either totally or partially [155]. These later **promoter-trap vectors** demonstrated a higher mutagenicity rate than enhancer-trap vectors. However, it would be a third type of vectors, called **gene-trap vectors**, that became more widely used for large-scale insertional mutagenesis screens due to their high mutagenic rate. The increased mutagenicity of gene-trap vectors relied on the presence of a splicing acceptor site in front of a promoterless *lacZ* reporter gene, followed by a strong polyadenylation signal such that, upon integration in any of the intronic sequences of a gene, the reporter would divert its normal splicing, causing protein truncations or missense transcripts, while also reporting the areas where the gene was expressed. Gene-trap vectors were not exempt from certain pitfalls, including their preference for inserting in certain genome
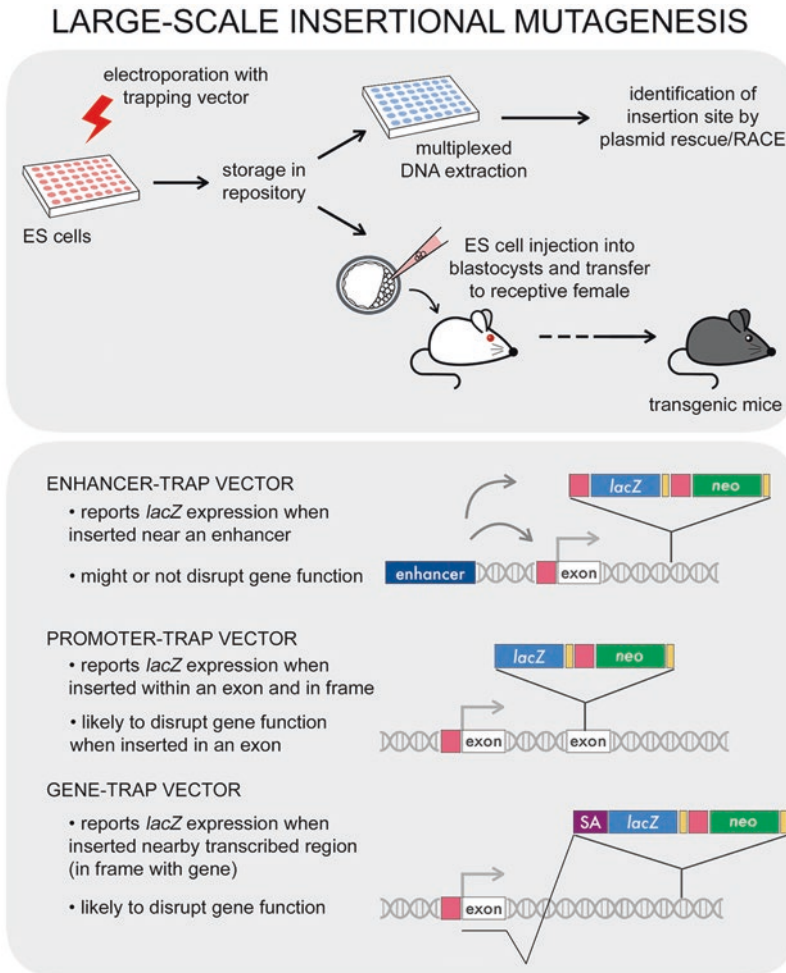
Fig. 1.12 *Large-scale insertional mutagenesis*. The possibility of electroporating vectors into ES cells opened the door to large-scale insertional mutagenesis (upper panel). After selecting for transgene insertion, ES cells can be frozen and stored until the molecular or phenotyping analysis of the insertions can be performed. The molecular characterization of the insertion site can be accomplished through plasmid rescue or RACE techniques, both of which make use of the known vector sequences to isolate the genomic areas flanking the insertion site. ES cell lines with interesting insertions can be injected into blastocysts to generate transgenic mice. All vectors contain a selection cassette that confers neomycin resistance (*neo^r*, green box) to the ES cells that have incorporated the transgene. Other features vary among trapping vectors (lower panel). In enhancer-trap vectors, the *lacZ* reporter gene is placed downstream of a basic promoter (pink box) and upstream of a polyadenylation site (yellow box), enabling these vectors to report the expression pattern dictated by enhancers nearby the insertion site. In promoter-trap vectors, the *lacZ* reporter gene lacks a promoter and therefore can only report expression when inserted in frame within a coding sequence. In gene-trap vectors, the *lacZ* reporter is preceded by a splicing acceptor site (SA, purple box), which functions to divert the normal splicing of genes in the vicinity of the insertion site

regions [156]. However, several generations of gene-trap vectors, with increasing degrees of sophistication, were developed to bypass some of these drawbacks and to facilitate the selection of different types of insertions (reviewed in [151]).

Compared to other contemporary methods, gene-trap mutagenesis stood out as one of the most practical approaches to generating mutations in mouse genes: large-scale mutagenesis screens using chemicals were still impractical due to the hardship of identifying the genes

disrupted by the mutations, and targeted mutagenesis through homologous recombination was not yet amenable to large-scale pipelines. In contrast, gene-trap mutagenesis could be performed on ES cells in a multiplexed format, and the resulting lines could be kept frozen until the insertion sites could be further characterized. As a consequence, the use of large-scale gene-trap mutagenesis spread quickly during the mid-1980s, with multiple academic groups throughout the globe, as well as the private company Lexicon Genetics, performing screen with different vectors and diverse overall goals [147, 150, 155–166]. Together, by the early years of the twenty-first century, these initiatives contributed to trapping nearly two thirds of all genes in mice. Importantly, the public gene-trap mutagenesis efforts, united under the operational umbrella of the **International Gene Trap Consortium** (IGTC) (https://igtc.org), provided annotated information about each transgenic line through online databases and made frozen ES cell stocks available without restriction to investigators worldwide [167, 168].

1.7.3 The International Knockout and Phenotyping Consortia

The first drafts of the human and mouse genome sequence revealed that the total number of genes in mammalian organisms would not be as high as the 150,000 that had been initially predicted but rather be between 25,000 and 30,000 genes. This lower number raised optimism that, given the genetic tools available in mice, it would be feasible to undertake a systematic functional characterization of all the genes in the mouse genome. At the time, it was estimated that the combined efforts of the scientific community, including large-scale ENU mutagenesis screens, gene-trap insertional mutagenesis, and mouse knockouts generated by individual investigators, already accounted for functional annotations in about 5000 genes [167]. Therefore, generating mutations in an additional 20,000–25,000 genes seemed attainable. Contributing to this optimism was the fact that the scientific community was

still under the spell of the recent successes of the international genome sequencing projects and the International Gene Trap Consortium, both of which left clear that international public investments in "big science" projects can enormously facilitate scientific exploration and spark new research venues. Inspired by this positive climate, scientists worldwide initiated discussions to endorse the systematic mutagenesis of all mouse genes and to devise the best approaches to reach this goal. Pan-European discussions, sponsored by the European Commission (EC Frame Program 6), started as early as 2002 [169], and a historical international meeting, held at the Banbury Center of the Cold Spring Harbor Laboratory in September of 2003, solidified the proposal for an international resource that could generate mutations in all mouse genes and make them available to the scientific community [170]. This proposal became a reality in 2007 with the creation of the **International Knockout Mouse Consortium (IKMC)**, a partnership of three different initiatives led and financed by the EU (European Conditional Mouse Mutagenesis (EUCOMM) Program, https://www.eucomm.org), the US (Knock Out Mouse Project (KOMP), https://www.komp.org), and Canada (North American Conditional Mouse Mutagenesis (NorCOMM) project, http://www.norcomm2.org/) [171].

From the outset, it was recognized that this ambitious enterprise will require complementary mutagenesis approaches and a coordination of all parties involved. **Large-scale gene-trap mutagenesis in ES cells** was considered the fastest and most cost-effective method to obtain gene mutations. Hence, additional gene-trap screens were launched using newest and more powerful vectors that, by including target sites for FLP and Cre recombinases, made it possible to generate conditional alleles [172]. Progress reports demonstrated the success of this strategy to obtain mutant ES cell lines [173, 174]. However, because some genes are recalcitrant to insertional mutagenesis and gene-trap strategies cannot guarantee null mutations, it was clear that targeted mutagenesis would also be needed to deliver a comprehensive catalog of mutations in

all genes [175]. Fortunately, technological innovations in bacterial recombineering [176–178], together with the introduction of robotics and computerized vector design [179], transformed the originally laborious homologous recombination protocols into streamlined high-throughput and automated processes. As a consequence, the IKMC phased out gene-trap mutagenesis and substituted this approach with automated homologous recombination ES cell pipelines [174]. These pipelines benefited from versatile new targeting vectors that, by borrowing some "tricks" from gene-trap vectors, allowed the generation of **"KO first, conditional ready" alleles** that could also report the expression pattern of the targeted genes. As a result of these combined efforts, thousands of mutant ES cell clones have become available through repositories worldwide, making it easier for individual investigators to obtain and analyze mouse mutants in their favorite genes. The number of ES cell clone requests processed by IKMC repositories attests to the impact that these resources have had in the scientific community [173].

While the first goal of the IKMC was to generate mutant ES cell lines for every gene in the genome, original discussions also recognized that the functional annotation of all of the genes in the mouse genome would also benefit from the systematic generation of live mice carrying the resulting mutations and their phenotyping through standardized tests. These second phase goals solidified in the creation of the **International Mouse Phenotyping Consortium (IMPC)** in 2010 [180]. This initiative benefited from the centers, infrastructure, and resources of the IKMC, which first used the available mutant ES cell lines to generate live mice colonies that would then be subject to standardized phenotyping. However, as of 2013, IMPC centers adopted CRISPR/Cas9 genetic engineering methods for gene targeting since this approach can be directly applied to embryos with high efficiency and specificity, bypassing the need to generate ES cell-line intermediates and, therefore, facilitating the workflow required to analyze gene function [181]. Regarding the phenotypic analysis of

mouse mutants, the IMPC benefited from the accrued experience of the European Mouse Clinics, which had been developing standardized phenotyping tests for the systematic analysis of ENU-induced mouse mutants for about a decade. Thus, initial IMPC efforts used the standardized high-throughput phenotyping pipelines defined by the European Mouse Phenotyping Resource of Standardized Screens (EMPReSS) as part of the EUMORPHIA (European Union Mouse Research for Public Health and Industrial Applications) program [182]. These pipelines include about 20 different platforms for the systematic analysis and statistical analysis of more than 400 variables relating to lethality, morphology, metabolism, skeletal and cardiovascular systems, neurobehavioral and sensory systems, hematology, biochemistry, and immunity [183, 184]. Since IMPC's inception, additional platforms have been developed for the evaluation of additional phenotypes, such as auditory dysfunction, ophthalmic diseases, congenital disorders, and complex traits, as well as for the identification of disease susceptibility under different environmental conditions, such as diet variations or infection [185–187].

All mouse models generated by the IKMC and the IMPC are available from worldwide repositories, either as live mice or as frozen sperm or embryos. Additionally, data and conclusions from the phenotypic analysis of mouse mutants are publicly available and regularly updated at the IMPC online portal (https://www.mousephenotype.org/). As of the writing of this book, the last IMPC update reported that 5861 mouse genes have already been completely or partially phenotyped, resulting in 69,982 phenotype calls reported and millions of data points produced [187]. Even though the international community will still need a few more years to complete the ambitious goals established in 2007, the data so far indicate that 30% of the mutations analyzed cause embryonic lethality and, therefore, are essential for life [186, 187]. Moreover, the analysis of these data through computerized algorithms has revealed that IMPC efforts have produced mouse models for about a third of all known

human Mendelian conditions, making the IMPC catalog a critical resource for understanding the molecular and genetic basis of human diseases.

1.8   Future Perspectives

A century of research on mouse genetics has transformed fancy mice into a powerful model system for understanding human biology. From the availability of inbred strains to the sequence and functional annotation of the mouse genome, the tools and resources currently available constitute invaluable assets to the scientific community. While the research accomplishments to date are countless, we are still far from understanding how our genomes make us who we are and how mutations cause disease. Some of the mysteries still lurking in our genomes include the inheritance of complex traits, the identification of regulatory elements, as well as the mechanisms responsible for epigenetic inheritance and cellular reprogramming. Scientific advances in the areas of genomics and computational biology are already increasing the research toolbox to dissect these fascinating phenomena [188]. These and future innovations, combined with the power of mouse genetics for uncovering the functional elements of our genomes, make the future ahead nothing but exciting.

References

 1. Silver LM. Mouse genetics. Oxford: Oxford University Press; 1995. 1 p.
 2. Paigen K. One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). Genetics. 2003;163(1):1–7.
 3. Guenet J-L, Bonhomme F. Origin of the laboratory mouse and related subspecies. The laboratory mouse: Elsevier Ltd.; 2004. 9 p.
 4. Russell ES. One man's influence: a tribute to William Ernest Castle. J Hered Narnia. 1954;45(5):211–4.
 5. Morse HCI. Origins of inbred mice [Internet]. Academic Press; 1978. 1 p. http://www.informatics.jax.org/morsebook/frames/frame45.shtml.
 6. Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MFW, et al. Genealogies of mouse inbred strains. Nat Genet. 2000;24(1):23.
 7. Crow JF. C. C. Little, cancer and inbred mice. Genetics. 2002;161(4):1357–61.
 8. Klein J. George Snell's first foray into the unexplored territory of the major histocompatibility complex. Genetics. 2001;159:435–9.
 9. Dobrovolskaia-Zavadskaia N. Sur la mortification spontanée de la queue chez la souris nouveau-née et sur l'existence d'un caractère (facteur) héréditaire "non-viable". Contes Rendues Soc Biol Paris. 1927;97:114–6.
10. Artzt K. Mammalian developmental genetics in the twentieth century. Genetics. 2012;192(4):1151–63.
11. Bennett D. Abnormalities associated with a chromosome region in the mouse. II. Embryological effects of lethal alleles in the T-region. Science. 1964;144(3616):263–7.
12. Morgan TH. Data for the study of sex-linked inheritance in Drosophila. J Exp Zool. 1912;13:79–101.
13. Haldane JBS, Sprunt AD, Haldane NM. Reduplication in mice (preliminary communication). J Genet. 1915;5(2):133–5.
14. Carter TC, Falconer DS. Stocks for detecting linkage in the mouse, and the theory of their design. J Genet. 1951;50(2):307–23.
15. Sturtevant AH. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. J Exp Zool. 1913;14(1):43–59.
16. Snell GD. Biology of lab mouse; 1941. pp. 1–520.
17. Lyon MF. A personal history of the mouse genome. Annu Rev Genome Human Genet. 2002;3(1):1–16.
18. Painter TS. The chromosomes of rodents. Science. 1926;64(1657):336.
19. Miller DA, Dev VG, Tantravahi R, Miller OJ, Schiffman MB, Yates RA, et al. Cytological detection of the c-25H deletion involving the albino (c) locus on chromosome 7 in the mouse. Genetics. 1974;78(3):905–10.
20. Miller OJ, Miller DA. Cytogenetics of the mouse. Annu Rev Genet. 1975;9:285–303.
21. Eicher EM. Foundation for the future: formal genetics of the mouse. Prog Clin Biol Res. 1981;45:1–49.
22. Love JM, Knight AM, McAleer MA, Todd JA. Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. Nucleic Acids Res. 1990;18(14):4123–30.
23. Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli NC, et al. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics. 1992;131(2):423–47.
24. Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, et al. A comprehensive genetic map of the mouse genome. Nature. 1996;380(6570):149–52.
25. Bailey DW. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation. 1971;11(3):325–7.
26. Bonhomme F, Benmehdi F, Britton-Davidian J, Martin S. [Genetic analysis of interspecific crosses Mus musculus L. x Mus spretus Lataste: linkage of Adh-1 with Amy-1 on chromosome 3 and Es-14

with Mod-1 on chromosome 9]. C R Acad Sci D. 1979;289(6):545–8.

27. Avner P, Amar L, Dandolo L, Guenet JL. Genetic analysis of the mouse using interspecific crosses. Trends Genet. 1988;4(1):18–23.

28. Copeland NG, Jenkins NA, Gilbert DJ, Eppig JT, Maltais LJ, Miller JC, et al. A genetic linkage map of the mouse: current applications and future prospects. Science. 1993;262(5130):57–66.

29. Saunders TL. A survey of internet resources for mouse development. Methods Enzymol. 2010;476:3–21.

30. de Boer P, Groen A. Fertility and meiotic behavior of male T70H tertiary trisomics of the mouse (Mus musculus). A case of preferential telomeric meiotic pairing in a mammal. Cytogenet Cell Genet. 1974;13(6):489–510.

31. Steinmetz M, Uematsu Y, Lindahl KF. Hotspots of homologous recombination in mammalian genome. Trends Genet. 1987;3:7–10.

32. Seldin MF, Howard TA, D'Eustachio P. Comparison of linkage maps of mouse chromosome 12 derived from laboratory strain intraspecific and Mus spretus interspecific backcrosses. Genomics. 1989;5(1):24–8.

33. Reeves RH, Crowley MR, Moseley WS, Seldin MF. Comparison of interspecific to intersubspecific backcrosses demonstrates species and sex differences in recombination frequency on mouse chromosome 16. Mamm Genome. 1991;1(3):158–64.

34. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520–62.

35. Morange M, Cobb M. A history of molecular biology. Cambridge: Harvard University Press; 2000. 1 p.

36. Judson HF. The eighth day of creation; 2004. 1 p.

37. Attwood TK, Gisel A, Bongcam-Rudloff E, Eriksson N-E. Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective; 2011.

38. Maniatis T, Fritsch EF, Sambrook J. Molecular cloning; 1982. 1 p.

39. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A. 1969;63(2):378–83.

40. Gall JG. The origin of in situ hybridization—a personal history. Methods. 2016;98:4–9.

41. Gene Ontology Consortium. The gene ontology project in 2008. Nucleic Acids Res. 2008;36(Database issue):D440–4.

42. Russell LB. The mouse house: a brief history of the ORNL mouse-genetics program, 1947–2009. Mutat Res. 2013;753(2):69–90.

43. Nakagata N. Cryopreservation of mouse spermatozoa. Mamm Genome. 2000;11(7):572–6.

44. Sztein JM, Takeo T, Nakagata N. History of cryobiology, with special emphasis in evolution of mouse sperm cryopreservation. Cryobiology. 2018;82:57–63.

45. Whittingham DG, Leibo SP, Mazur P. Survival of mouse embryos frozen to −196 degrees and −269 degrees C. Science. 1972;178(4059):411–4.

46. Jaenisch R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc Natl Acad Sci U S A. 1976;73(4):1260–4.

47. Brinster RL, Chen HY, Trumbauer ME, Avarbock MR. Translation of globin messenger RNA by the mouse ovum. Nature. 1980;283(5746):499–501.

48. Gordon JW, Ruddle FH. Integration and stable germ line transmission of genes injected into mouse pronuclei. Science. 1981;214(4526):1244–6.

49. Nagy A, Behringer RR, Gertsenstein M, Vintersten K. Manipulating the mouse embryo: a laboratory manual. 3rd ed. New York: Cold Spring Harbor Laboratory Press; 2003.

50. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A. 1980;77(12):7380–4.

51. Brinster RL, Chen HY, Trumbauer M, Senear AW, Warren R, Palmiter RD. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell. 1981;27(1 Pt 2):223–31.

52. Costantini F, Lacy E. Introduction of a rabbit beta-globin gene into the mouse germ line. Nature. 1981;294(5836):92–4.

53. Wagner TE, Hoppe PC, Jollick JD, Scholl DR, Hodinka RL, Gault JB. Microinjection of a rabbit beta-globin gene into zygotes and its subsequent expression in adult mice and their offspring. Proc Natl Acad Sci U S A. 1981;78(10):6376–80.

54. Wagner EF, Stewart TA, Mintz B. The human beta-globin gene and a functional viral thymidine kinase gene in developing mice. Proc Natl Acad Sci U S A. 1981;78(8):5016–20.

55. Niemann H, Kues W, Carnwath JW. Transgenic farm animals: present and future. Rev Sci Tech. 2005;24(1):285–98.

56. Lewandoski M. Conditional control of gene expression in the mouse. Nat Rev Genet. 2001;2(10):743–55.

57. Hanahan D. Transgenic mice as probes into complex systems. Science. 1989;246(4935):1265–75.

58. Stacey A, Bateman J, Choi T, Mascara T, Cole W, Jaenisch R. Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered mutant pro-alpha 1(I) collagen gene. Nature. 1988;332(6160):131–6.

59. Readhead C, Popko B, Takahashi N, Shine HD, Saavedra RA, Sidman RL, et al. Expression of a myelin basic protein gene in transgenic shiverer mice: correction of the dysmyelinating phenotype. Cell. 1987;48(4):703–12.

60. Lo CW, Coulling M, Kirby C. Tracking of mouse cell lineage using microinjected DNA sequences: analyses using genomic Southern blotting and

tissue-section in situ hybridizations. Differentiation. 1987;35(1):37–44.

61. Goring DR, Rossant J, Clapoff S, Breitman ML, Tsui LC. In situ detection of beta-galactosidase in lenses of transgenic mice with a gamma-crystallin/lacZ gene. Science. 1987;235(4787):456–8.

62. Okabe M, Ikawa M, Kominami K, Nakanishi T, Nishimune Y. "Green mice" as a source of ubiquitous green cells. FEBS Lett. 1997;407(3):313–9.

63. Palmiter RD, Behringer RR, Quaife CJ, Maxwell F, Maxwell IH, Brinster RL. Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. Cell. 1987;50(3):435–43.

64. Jaenisch R. Transgenic animals. Science. 1988;240(4858):1468–74.

65. Palmiter RD, Brinster RL. Germ-line transformation of mice. Annu Rev Genet. 1986;20(1):465–99.

66. Soriano P, Gridley T, Jaenisch R. Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. Genes Dev. 1987;1(4):366–75.

67. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981;292(5819):154–6.

68. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A. 1981;78(12):7634–8.

69. Bradley A, Evans M, Kaufman MH, Robertson E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature. 1984;309(5965):255–6.

70. Robertson E, Bradley A, Kuehn M, Evans M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature. 1986;323(6087):445–8.

71. Gossler A, Doetschman T, Korn R, Serfling E, Kemler R. Transgenesis by means of blastocyst-derived embryonic stem cell lines. Proc Natl Acad Sci U S A. 1986;83(23):9065–9.

72. Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S, et al. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature. 1987;330(6148):576–8.

73. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell. 1987;51(3):503–12.

74. Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell. 1989;56(2):313–21.

75. Schwartzberg PL, Goff SP, Robertson EJ. Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. Science. 1989;246(4931):799–803.

76. Thomas KR, Capecchi MR. Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature. 1990;346(6287):847–50.

77. Zijlstra M, Li E, Sajjadi F, Subramani S, Jaenisch R. Germ-line transmission of a disrupted beta 2-microglobulin gene produced by homologous recombination in embryonic stem cells. Nature. 1989;342(6248):435–8.

78. Koller BH, Hagemann LJ, Doetschman T, Hagaman JR, Huang S, Williams PJ, et al. Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. Proc Natl Acad Sci U S A. 1989;86(22):8927–31.

79. Babinet C, Cohen-Tannoudji M. Genome engineering via homologous recombination in mouse embryonic stem (ES) cells: an amazingly versatile tool for the study of mammalian biology. An Acad Bras Cienc. 2001;73(3):365–83.

80. Paigen K. One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981–2002). Genetics. Genet Soc Am. 2003;163(4):1227–35.

81. Hall B, Limaye A, Kulkarni AB. Overview: generation of gene knockout mice. Curr Protoc Cell Biol. 2009; Chapter 19(1):Unit19.1219.12.1–17.

82. Nagy A. Cre recombinase: the universal reagent for genome tailoring. Genesis. 2000;26(2):99–109.

83. Yu Y, Bradley A. Engineering chromosomal rearrangements in mice. Nat Rev Genet. 2001;2(10):780–90.

84. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 2013;31(7):397–405.

85. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Micro. 2011;9(6):467–77.

86. Lander ES. The heroes of CRISPR. Cell. 2016;164(1–2):18–28.

87. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819–23.

88. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013;2:e00471.

89. Singh P, Schimenti JC, Bolcun-Filas E. A mouse geneticist's practical guide to CRISPR applications. Genetics. 2015;199(1):1–15.

90. Shen B, Zhang J, Wu H, Wang J, Ma K, Li Z, et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res. 2013;23(5):720–3.

91. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013;153(4):910–8.

92. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell. 2013;154(6):1370–9.

93. Modzelewski AJ, Chen S, Willis BJ, Lloyd KCK, Wood JA, He L. Efficient mouse genome engineering by CRISPR-EZ technology. Nat Protoc. 2018;13(6):1253–74.

94. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat Protoc. 2014;9(8):1956–68.

95. Fujii W, Kawasaki K, Sugiura K, Naito K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. Nucleic Acids Res. 2013;41:e187–e187.

96. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013; 152(5):1173–83.

97. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442–51.

98. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013;23(10):1163–71.

99. National Research Council, Studies DOEAL, Sciences COL, Genome COMASTH. Mapping and sequencing the human genome. National Academies Press; 1988. 1 p.

100. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.

101. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304–51.

102. Ferry G, Sulston J. The common thread. Washington, DC: Joseph Henry Press; 2002. 1 p

103. Cook-Deegan RM. The gene wars. New York: W. W. Norton & Company; 1996. 1 p

104. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, et al. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet. 2000;24(4):381–6.

105. Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature. 2007;448(7157):1050–3.

106. Adams DJ, Doran AG, Lilue J, Keane TM. The Mouse Genomes Project: a repository of inbred laboratory mouse strain genomes. Mamm Genome. 2015;26(9–10):403–12.

107. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A. 1979;76(11):5818–9.

108. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila. Nature. 1980;287(5785):795–801.

109. Bode VC. Ethylnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the T region of mouse chromosome 17. Genetics. 1984;108(2):457–70.

110. Shedlovsky A, Guenet JL, Johnson LL, Dove WF. Induction of recessive lethal mutations in the T/t-H-2 region of the mouse genome by a point mutagen. Genet Res. 1986;47(2):135–42.

111. Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, et al. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science. 1994;264(5159):719–25.

112. Moser AR, Pitot HC, Dove WF. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science. 1990;247(4940):322–4.

113. King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, et al. Positional cloning of the mouse circadian clock gene. Cell. 1997;89(4):641–53.

114. Su LK, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, et al. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science. 1992;256(5057):668–70.

115. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. Mouse ENU mutagenesis. Hum Mol Genet. 1999;8(10):1955–63.

116. Clark AT, Goldowitz D, Takahashi JS, Vitaterna MH, Siepka SM, Peters LL, et al. Implementing large-scale ENU mutagenesis screens in North America. Genetica. 2004;122(1):51–64.

117. Gondo Y. Trends in large-scale mouse mutagenesis: from genetics to functional genomics. Nat Rev Genet. 2008;9(10):803–10.

118. Rogers DC, Fisher EM, Brown SD, Peters J, Hunter AJ, Martin JE. Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. Mamm Genome. 1997;8(10):711–3.

119. Hrabé de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, Marschall S, et al. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet. 2000;25(4):444–7.

120. Rinchik EM. Carpenter DA. N-ethyl-N-nitrosourea mutagenesis of a 6- to 11-cM subregion of the Fah-Hbb interval of mouse chromosome 7: completed testing of 4557 gametes and deletion mapping and complementation analysis of 31 mutations. Genetics. 1999;152(1):373–83.

121. Schimenti J, Bucan M. Functional genomics in the mouse: phenotype-based mutagenesis screens. Genome Res. 1998;8(7):698–710.

122. Justice MJ, Zheng B, Woychik RP, Bradley A. Using targeted large deletions and high-efficiency N-ethyl-N-nitrosourea mutagenesis for functional analyses of the mammalian genome. Methods. 1997;13(4):423–36.

123. Kile BT, Hentges KE, Clark AT, Nakamura H, Salinger AP, Liu B, et al. Functional genetic analysis of mouse chromosome 11. Nature. 2003;425(6953):81–6.

124. Hagge-Greenberg A, Snow P, O'Brien TP. Establishing an ENU mutagenesis screen for the piebald region of mouse chromosome 14. Mamm Genome. 2001;12(12):938–41.

125. Boles MK, Wilkinson BM, Maxwell A, Lai L, Mills AA, Nishijima I, et al. A mouse chromosome 4 balancer ENU-mutagenesis screen isolates eleven lethal lines. BMC Genet. 2009;10(1):12.

126. Ching Y-H, Munroe RJ, Moran JL, Barker AK, Mauceli E, Fennell T, et al. High resolution mapping and positional cloning of ENU-induced mutations in the Rw region of mouse chromosome 5. BMC Genet. 2010;11(1):106.

127. Kasarskis A, Manova K, Anderson KV. A phenotype-based screen for embryonic lethal mutations in the mouse. Proc Natl Acad Sci U S A. 1998;95(13):7485–90.

128. Herron BJ, Lu W, Rao C, Liu S, Peters H, Bronson RT, et al. Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. Nat Genet. 2002;30(2):185–9.

129. Zarbalis K, May SR, Shen Y, Ekker M, Rubenstein JLR, Peterson AS. A focused and efficient genetic screening strategy in the mouse: identification of mutations that disrupt cortical development. PLoS Biol. 2004;2(8):E219. Joshua R Sanes, editor. Public Library of Science.

130. Hoebe K, Beutler B. Unraveling innate immunity using large scale N-ethyl-N-nitrosourea mutagenesis. Tissue Antigens. 2005;65(5):395–401.

131. Garcia-Garcia MJ, Eggenschwiler JT, Caspary T, Alcorn HL, Wyler MR, Huangfu D, et al. Analysis of mouse embryonic patterning and morphogenesis by forward genetics. Proc Natl Acad Sci U S A. 2005;102(17):5913–9.

132. Papathanasiou P, Goodnow CC. Connecting mammalian genome with phenome by ENU mouse mutagenesis: gene combinations specifying the immune system. Annu Rev Genet. 2005;39(1):241–62.

133. Stottmann RW, Moran JL, Turbe-Doan A, Driver E, Kelley M, Beier DR. Focusing forward genetics: a tripartite ENU screen for neurodevelopmental mutations in the mouse. Genetics. 2011;188(3):615–24.

134. Ha S, Stottmann RW, Furley AJ, Beier DR. A forward genetic screen in mice identifies mutants with abnormal cortical patterning. Cereb Cortex. 2015;25(1):167–79.

135. Hentges K, Thompson K, Peterson A. The flat-top gene is required for the expansion and regionalization of the telencephalic primordium. Development. 1999;126(8):1601–9.

136. Rubio-Aliaga I, Soewarto D, Wagner S, Klaften M, Fuchs H, Kalaydjiev S, et al. A genetic screen for modifiers of the delta1-dependent notch signaling function in the mouse. Genetics. 2007;175(3):1451–63.

137. Matera I, Watkins-Chow DE, Loftus SK, Hou L, Incao A, Silver DL, et al. A sensitized mutagenesis screen identifies Gli3 as a modifier of Sox10 neurocristopathy. Hum Mol Genet. 2008;17(14):2118–31.

138. Horner VL, Caspary T. Creating a "hopeful monster": mouse forward genetic screens. Methods Mol Biol. 2011;770:313–36.

139. Caspary T. Chapter 16—phenotype-driven mouse ENU mutagenesis screen. In: Guide to techniques in mouse development, part B: mouse molecular genetics, 2nd ed., vol. 477. Elsevier Inc.; 2010. 15 p.

140. Stottmann RW, Beier DR. Using ENU mutagenesis for phenotype-driven analysis of the mouse. Meth Enzymol. 2010;477:329–48.

141. Moresco EMY, Li X, Beutler B. Going forward with genetics. Am J Pathol. 2013;182(5):1462–73.

142. Sun M, Mondal K, Patel V, Horner VL, Long AB, Cutler DJ, et al. Multiplex chromosomal exome sequencing accelerates identification of ENU-induced mutations in the mouse. G3 (Bethesda). 2012;2(1):143–50.

143. Gallego-Llamas J, Timms AE, Geister KA, Lindsay A, Beier DR. Variant mapping and mutation discovery in inbred mice using next-generation sequencing. BMC Genomics. 2015;16(1):913.

144. Geister KA, Timms AE, Beier DR. Optimizing genomic methods for mapping and identification of candidate variants in ENU mutagenesis screens using inbred mice. G3 (Bethesda). 2018;8(2):401–9.

145. Takahasi KR, Sakuraba Y, Gondo Y. Mutational pattern and frequency of induced nucleotide changes in mouse ENU mutagenesis. BMC Mol Biol. 2007;8(1):52.

146. Eggenschwiler JT, Anderson KV. Cilia and developmental signaling. Annu Rev Cell Dev Biol. 2007;23(1):345–73.

147. Hicks GG, Shi EG, Li XM, Li CH, Pawlak M, Ruley HE. Functional genomics in mice by tagged sequence mutagenesis. Nat Genet. 1997;16(4):338–44.

148. Townley DJ, Avery BJ, Rosen B, Skarnes WC. Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. Genome Res. 1997;7(3):293–8.

149. Lovell-Badge RH, Bygrave AE, Bradley A, Robertson E, Evans MJ, Cheah KS. Transformation of embryonic stem cells with the human type-II collagen gene and its expression in chimeric mice. Cold Spring Harb Symp Quant Biol. 1985;50:707–11.

150. Gossler A, Joyner AL, Rossant J, Skarnes WC. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science. 1989;244(4903):463–5.

151. Stanford WL, Cohn JB, Cordes SP. Gene-trap mutagenesis: past, present and beyond. Nat Rev Genet. 2001;2(10):756–68.

152. O'Kane CJ, Gehring WJ. Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A. 1987;84(24):9123–7.

153. Allen ND, Cran DG, Barton SC, Hettle S, Reik W, Surani MA. Transgenes as probes for active chromosomal domains in mouse development. Nature. 1988;333(6176):852–5.

154. Kothary R, Clapoff S, Brown A, Campbell R, Peterson A, Rossant J. A transgene containing lacZ

inserted into the dystonia locus is expressed in neural tube. Nature. 1988;335(6189):435–7.

155. Friedrich G, Soriano P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. 1991;5(9):1513–23.

156. Hansen J, Floss T, Van Sloun P, Füchtbauer E-M, Vauti F, Arnold H-H, et al. A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A. 2003;100(17):9918–22.

157. Skarnes WC, Auerbach BA, Joyner AL. A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes Dev. 1992;6(6):903–18.

158. Niwa H, Araki K, Kimura S, Taniguchi S, Wakasugi S, Yamamura K. An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. J Biochem. 1993;113(3):343–9.

159. Skarnes WC, Moss JE, Hurtley SM, Beddington RS. Capturing genes encoding membrane and secreted proteins important for mouse development. Proc Natl Acad Sci U S A. 1995;92(14):6592–6.

160. Wurst W, Rossant J, Prideaux V, Kownacka M, Joyner A, Hill DP, et al. A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. Genetics. 1995;139(2):889–99.

161. Salminen M, Meyer BI, Gruss P. Efficient poly A trap approach allows the capture of genes specifically active in differentiated embryonic stem cells and in mouse embryos. Dev Dyn. 1998;212(2):326–33.

162. Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature. 1998;392(6676):608–11.

163. Araki K, Imaizumi T, Sekimoto T, Yoshinobu K, Yoshimuta J, Akizuki M, et al. Exchangeable gene trap using the Cre/mutated lox system. Cell Mol Biol (Noisy-le-grand). 1999;45(5):737–50.

164. Wiles MV, Vauti F, Otte J, Füchtbauer EM, Ruiz P, Füchtbauer A, et al. Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat Genet. 2000;24(1):13–4.

165. Hardouin N, Nagy A. Gene-trap-based target site for cre-mediated transgenic insertion. Genesis. 2000;26(4):245–52.

166. Stryke D. BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. Nucleic Acids Res. 2003;31(1):278–81.

167. The International Mouse Mutagenesis Consortium, Nadeau JH, Balling R, Barsh G, Beier D, SDM B, et al. Functional annotation of mouse genome sequences. Science. 2001;291(5507):1251–5.

168. Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, Cox T, et al. A public gene trap resource for mouse functional genomics. Nat Genet. 2004;36(6):543–4.

169. Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, et al. The European dimension for the mouse genome mutagenesis program. Nat Genet. 2004;36(9):925–7.

170. Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, et al. The knockout mouse project. Nat Genet. 2004;36(9):921–4.

171. International Mouse Knockout Consortium, Collins FS, Rossant J, Wurst W. A mouse for all reasons. Cell. 2007;128(1):9–13.

172. Schnütgen F, De-Zolt S, Van Sloun P, Hollatz M, Floss T, Hansen J, et al. Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A. 2005;102(20):7221–6.

173. Lloyd KCK. A knockout mouse resource for the biomedical research community. Ann N Y Acad Sci. 2011;1245(1):24–6.

174. Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling M-C, et al. The mammalian gene function resource: the International Knockout Mouse Consortium. Mamm Genome. 2012;23(9–10):580–6.

175. Friedel RH, Seisenberger C, Kaloff C, Wurst W. EUCOMM the European conditional mouse mutagenesis program. Brief Funct Genomic Proteomic. 2007;6(3):180–5.

176. Angrand PO, Daigle N, van der Hoeven F, Schöler HR, Stewart AF. Simplified generation of targeting constructs using ET recombination. Nucleic Acids Res. 1999;27(17):e16.

177. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 1998;20(2):123–8.

178. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet. 2001;2(10):769–79.

179. Valenzuela DM, Murphy AJ, Frendewey D, Gale NW, Economides AN, Auerbach W, et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat Biotechnol. 2003;21(6):652–9.

180. Abbott A. Mouse project to find each gene's role. Nature. 2010;465(7297):410.

181. Lanza DG, Gaspero A, Lorenzo I, Liao L, Zheng P, Wang Y, et al. Comparative analysis of single-stranded DNA donors to generate conditional null mouse alleles. BMC Biol. 2018;16(1):69.

182. Brown SDM, Chambon P, de Angelis MH, Eumorphia Consortium. EMPReSS: standardized phenotype screens for functional annotation of the mouse genome. Nat Genet. 2005;37(11):1155.

183. Brown SDM, Moore MW. The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. Mamm Genome. 2012;23(9–10):632–40.

184. de Angelis MH, Nicholson G, Selloum M, White J, Morgan H, Ramirez-Solis R, et al. Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. Nature. 2015;47(9):969–78.

185. Ayadi A, Birling M-C, Bottomley J, Bussell J, Fuchs H, Fray M, et al. Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mamm Genome. 2012;23(9–10):600–10.

186. Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, Warren J, et al. Disease model discovery from 3328 gene knockouts by The International Mouse Phenotyping Consortium. Nature. 2017;49(8):1231–8.

187. Cacheiro P, Haendel MA, Smedley D, Meehan T, Mason J, Mashhadi HH, et al. New models for human disease from the International Mouse Phenotyping Consortium. Mamm Genome. 2019;30(5–6):143–50.

188. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636–40.