



Rodent Genetics

Fernando Benavides and Jean-Louis Guénet

Abstract

This chapter is an overview of the current and growing knowledge of the genetics of laboratory rodents, specifically the mouse (*Mus musculus*) and the rat (*Rattus norvegicus*), the two main species used in biomedical research. We present basic information about Mendelian genetics and on the structure of the mouse and rat genomes, including the protein-coding DNA and the more intriguing non-coding DNA sequences, abundant in repetitive DNA, transposable elements and different types of genetic polymorphisms. Experiments should be performed with carefully designed and approved protocols, including the use of genetically defined animals. Thus, in this chapter we discussed the different types of genetically standardized laboratory strains and the aspects related to their genetic quality control. We also present the different types of genetically altered mice and rats, including spontaneous and chemically induced mutations, random transgenesis, targeted mutagenesis using embry-

onic stem cells and the novel genome editing techniques. It is very important for the veterinarians and technicians in charge of animal facilities, as well as for researchers and students using mouse and rat models that they have an available up-to-date information devoted to the genetics of these species.

Keywords

laboratory rodents · Mice · Rats · Inbred strains · Genetic polymorphisms · Quality control · Transgenesis · Genome edition

1 Introduction to Mendelian Genetics and Genomics

1.1 Genes, Alleles and Their Interactions

Several years after Gregor Mendel's seminal publication *Experiments on Plant Hybridization* was published (1866), Hugo de Vries published *Intracellular Pangenesis* (1889) in which he recommended the word *pangens* be used to specify Mendel's 'hereditary particles'. The Danish biologist Wilhelm Johannsen proposed in 1909 that the (Danish) word *gen* be used to describe the units of heredity. Almost at the same time,

F. Benavides (✉)
Department of Epigenetics and Molecular
Carcinogenesis, The University of Texas MD Anderson
Cancer Center, Smithville, TX, USA
e-mail: fbenavid@mdanderson.org

J.-L. Guénet
Institut Pasteur (Emeritus), Paris, France

© Springer Nature Switzerland AG 2021

J. M. Sánchez Morgado, A. Brønstad (eds.), *Experimental Design and Reproducibility in Preclinical Animal Studies*, Laboratory Animal Science and Medicine 1,
https://doi.org/10.1007/978-3-030-66147-2_2

Johannsen introduced the terms *phenotype* and *genotype*. William Bateson proposed the term *genetics* to describe the science dealing with *gens* (*genes*). Shortly after the confirmation that DNA was the molecular basis of inheritance (seminal work published by Avery, McCarty and MacLeod in 1944), the gene was defined in molecular terms as ‘a segment of DNA of variable size encoding an enzyme’. This definition was revised to ‘one gene, one polypeptide’ when it was recognized that some proteins are not enzymes. With the completion of the first (draft) sequences of the human (2001), mouse (2002) and rat (2004) genomes and the confirmation that many genes are not translated into polypeptides, the definition of the gene changed again.

Today, a gene corresponds to a segment of DNA that is transcribed into RNA. Some RNA molecules, like messenger RNAs (mRNAs), are translated into polypeptides, whereas many others are not translated but nevertheless have important functions. Recently, information collected from the systematic analysis of a single transcriptome revealed that mammalian DNA is pervasively transcribed from both strands and that the proportion of DNA transcribed into RNA is much greater than expected. The same analysis also revealed that not all mammalian genes are easily identified in DNA; on the contrary, their limits are often difficult to delineate, with some small genes being nested inside the larger ones (e.g. inserted in the introns). Thus, it seems clear that the concept of the gene must be reconsidered and its definition reformulated. Nonetheless, we will work with the idea that a gene is a functional unit contained in a short DNA segment that is transcribed into RNA and whose inheritance can be followed experimentally generation after generation. Genes can be precisely localized on a specific chromosome using a variety of techniques, and this position defines its *locus* (plural *loci*), the Latin word for ‘place’.

For decades, *genome* is referred to the collection of genes in a given species. Now, the concept includes both the genes (i.e. the coding sequences) and the sum of heterogeneous DNA intermingled with the genes. Thus, when we refer to the genome sequence, we are referring to the

sequence of all nuclear DNA. The number of protein-coding genes in the mammalian genome is predicted to be 22,000 to 24,000 genes, on par with the 22,628 currently listed in the mouse GRCm38 assembly and the 22,250 in the rat Rnor_6.0 assembly. However, some genes vary in copy number across different strains, and even between individuals, with many being non-functional, whereas others are present in only some strains (or species) and absent in others. Such gene variations complicate accurately evaluating organismal gene number. Predicting gene number becomes even more difficult given that multiple RNAs (coding and non-coding) can be transcribed from the same gene via *alternative splicing*, tremendously increasing the number and diversity of molecules potentially encoded in the genome. Obviously, it is the sum of these transcripts, not the raw number of genes that is important for defining the genome.

Most genes exist in alternative forms (variants) called *alleles*. The word ‘allele’ is an abbreviation of the ancient word *allelomorph*, which described the different forms of a gene. Formerly, the concept of alleles was tightly associated with mutations that produce phenotypes different from wild type (i.e. the version most commonly found in wild animals), for example, a different coat colour, a heritable skeletal defect or a debilitating neurological disease. The new version of the gene was called a *mutant allele*. The concept of the allele, like the gene, has changed over time so that now any alteration of DNA sequence within a gene is defined as a new allele, regardless of whether the change produces a phenotype.

The term *polymorphism* can refer to many things, including the alleles present at a specific locus or to all loci of a strain or species. The whole collection of alleles segregating in a given population represents what geneticists call the *genetic polymorphism*. In the mouse, the gene encoding tyrosinase (*Tyr*), an enzyme that is instrumental for the synthesis of the pigment melanin, was one of the first (if not the first) genes to be identified based on a variation in coat colour. At the *Tyr* locus, the wild-type allele encodes a functional tyrosinase, but many mutant alleles encode non-functional enzymes resulting in albinism. Over

120 different mutations have been identified at the *Tyr* locus, some of them affecting coat colour (e.g. chinchilla, *Tyr^{c-ch}*; extreme dilution, *Tyr^{c-e}*; and Himalayan, *Tyr^{c-h}*).

1.1.1 Dominance, Recessivity and Co-dominance

When the alleles at a given locus are identical on both chromosomes, the animal is *homozygous* for that allele. When the two alleles are different, the animal is *heterozygous*, and the phenotype will depend upon the interactions between the two alleles. To illustrate, we will again consider the *Tyr* gene in the mouse. *Tyr* has several alleles, some of which are non-functional, like *Tyr^c*. *Tyr^c/Tyr^c* mice are albino, but *Tyr^c/Tyr⁺* heterozygotes are pigmented like wild mice because the mutant *Tyr^c* allele is *recessive* to the *dominant* wild-type allele (*Tyr⁺* or sometimes only +). In this case, the lack of functional tyrosinase due the presence of the *Tyr^c* allele is completely compensated for by a single copy of the wild-type allele.

Other *Tyr* alleles have less dramatic effects than *Tyr^c* on the synthesis of melanin. In many cases the mice are pigmented, although always less than or differently from the wild type. Mice homozygous for the chinchilla allele *Tyr^{c-ch}* have a diluted coat colour, but mice homozygous for the Himalayan allele *Tyr^{c-h}* have a remarkable pattern of pigmentation. They have light-ruby eyes and a coat that is mainly white with only the tip of the nose, tip of the ears and the tail pigmented normally, like Siamese cats. This pattern results from the *Tyr^{c-h}* allele-encoded, thermo-labile tyrosinase being active only in the colder parts of the body, where the temperature is below 35 °C. With so many *Tyr* alleles available, one could breed a wide variety of mice heterozygous or homozygous for the different alleles to find that the normal allele (*Tyr⁺*) is dominant over all other alleles. However, if the mice were graded based on decreasing coat colour intensity for all possible combinations of the *Tyr⁺*, *Tyr^{c-ch}*, *Tyr^{c-e}* and *Tyr^c* alleles, we would observe an almost continuous gradient of pigmentation from wild type to albino. Therefore, dominance and recessivity must be considered only in the context of a specific allele pair.

Semi-dominance (sometimes referred to as *incomplete dominance*) describes mutant alleles that produce heterozygotes with a phenotype that is different from and often intermediate to both kinds of homozygotes. A typical example is the *Kit^{W-f}* allele. *Kit^{W-f}/+* heterozygous mice have a light grey coat with a white spot on the belly and on the forehead, whereas *Kit^{W-f}/Kit^{W-f}* homozygous mice are extensively spotted. Amazingly, the tails of these mice perfectly characterize the situation; the tail is completely pigmented from the base to the tip in wild-type mice, half-pigmented in heterozygotes and unpigmented in homozygotes.

Another type of allelic interaction common in mammals is *co-dominance*. Co-dominance occurs when the two alleles at a given locus are both expressed in the heterozygote to create a unique phenotype. Most genetics textbooks illustrate the concept of co-dominance using the AB blood groups in humans, where AB heterozygotes have a phenotype in which both the A and B antigens are expressed on red blood cells. Blood groups homologous to the human AB system do not exist in the mouse or rat, but nearly all alleles that encode forms of the same protein that vary by charge are co-dominantly expressed.

Other allelic interactions have been discovered by studying the process such as sex determination. In mammalian species, males have only one X-chromosome and therefore are *hemizygous* for all genes carried by this chromosome, and all are fully expressed. In females, X-inactivation, a mechanism of dosage compensation causes most X-linked genes to be functionally haploid; only one copy of each gene is transcribed, and the other copy is switched off. The choice of which allele to inactive is usually a random process. In mammals, a few genes in the so-called pseudo-autosomal region of the X-chromosome are not inactivated and behave as autosomal genes [1]. Notably, certain autosomal regions, sometimes reduced to one or a few genes, are also functionally haploid, expressing the allele(s) inherited from only one of the two parents, a phenomenon called genomic imprinting, also resulting from epigenetic mechanisms [2, 3].

1.1.2 Epistasis and Pleiotropy

Many phenotypic traits are controlled by more than one gene, and a single gene can contribute to the phenotypic expression of one or several other genes. *Epistasis* occurs when the phenotypic expression of gene (or allele) *A* depends on the presence of one or more specific alleles (*B*, *C*, *D*) at other loci to modify or suppress the classical phenotype of gene *A*. In other words, epistasis is an interaction between nonallelic genes in which one gene suppresses or enhances the expression of another. The gene that is expressed is *epistatic* over the other genes, which are themselves *hypostatic*. The genes that determine coat colour offer simple, didactic examples. Exploiting the variety of alleles at the five major loci governing mouse coat colour (agouti, *A*; tyrosinase, *Tyr*; brown, *Tyrp1*; dilute, *Myo5a*; and pink-eyed dilution, *Oca2*), one can generate a large collection of mice with a wide array of coat colours. However, sometimes the effects of a given mutant allele cannot be observed in the presence of another particular allele. For example, a mouse with a non-agouti brown coat colour (genotype *a/a*; *Tyrp1^b/Tyrp1^b*) would appear ‘chocolate’, except in the presence of two copies of the *Tyr^c* mutant allele (homozygous) that causes the mouse to be albino. In this case, the *Tyr^c* allele exhibits an epistatic interaction with all other coat colour genes because without tyrosinase there is no pigment.

Pleiotropy describes a common genetic phenomenon in which a mutant allele influences multiple phenotypic traits. In fact, if we carefully analyse mutants with deleterious phenotypes, we would discover that almost all of them exhibit a range of altered phenotypes. The yellow allele (*A^y*) was identified because of its beautiful yellow coat colour, but these mutants are also slightly diabetic, exhibit liver hypertrophy and often become obese and sterile following the first few months of life [4]. Compared to normal mice, these mice are also more susceptible to several kinds of tumours and are more aggressive. Given that the products of most genes have multiple functions, pleiotropy is more a rule than an exception. It simply means that the gene in question codes for a product that is used by various cell types, signals to multiple targets or regulates more than one pathway, as a transcription factor might.

1.1.3 Penetrance and Expressivity

Penetrance is a term used to express the fraction (percentage) of individuals of a given genotype that effectively exhibits the expected phenotype. For example, if a particular dominant mutation has 80% penetrance, then 80% of the mice carrying the mutant allele will develop the phenotype, and 20% will look normal. A genotype exhibits variable *expressivity* when individuals with that genotype differ in the extent to which they express the phenotype. One example illustrating the concept of expressivity and differentiating it from the concept of penetrance (which is not always easy) is the case of spotting in cattle. When observing an herd of Holstein Friesian cattle, one may notice that, although all the cows are spotted (penetrance is 100%), the ratio of black/white is highly variable from one animal to the next. The spotting is highly variable in shape (no surprise) and extent (which is more surprising). Similarly, rodents can also display a large amount of phenotypic variations among individuals with the same genotype, for example, the case of a mutation in the brachyury gene (*T*) which encodes a transcription factor important for proper formation of the tail and the *Ednrb^s* spotting mouse mutation (Fig. 1).

The causes of variable penetrance and expressivity are not well understood. In the mouse and rat, one can study the phenotypic expression of the same mutation in different genetic backgrounds and note more or less consistent differences, indicating the influence of a genetic component (modifier genes). However, one can also observe phenotypic variations in animals having exactly the same mutation in exactly the same genetic background – meaning that nongenetic factors, such as epigenetic and environmental factors, also influence penetrance and expressivity.

1.2 Genomes and Genetic Variation

The sizes of the laboratory mouse (*Mus musculus* strain C57BL/6J) and rat (*Rattus norvegicus* mixed female BN and male SHR) genomes are 2.7 Gbp and 2.8 Gbp, respectively [5, 6]. Both genomes are 14% smaller than the human genome

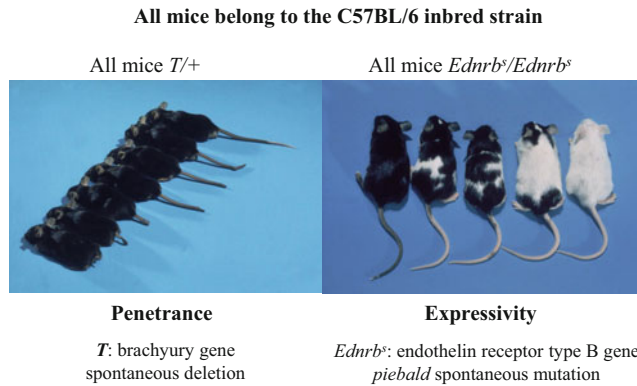


Fig. 1 Penetrance and expressivity. The picture illustrates two major characteristics of the phenotypic expression of mutant alleles in mammalian species. In the present case, all seven mice on the left panel are affected by the same mutation (*brachyury* (T), with 100% penetrance) affecting the length of the tail, but they exhibit great variations in the phenotype, with some mice (top of the picture) with a normal-looking tail. On the right-hand side, all mice exhibit a spotted coat with wide variations in expressivity

(mutation $Ednrb^s$). The penetrance characterizes the fraction of individuals of a given genotype that actually shows a particular phenotype irrespective of the degree of its expression. The expressivity characterizes the phenotypic variation among individuals having the same genotype. It is now well established that modifier genes influence the phenotypic expression, but these genes cannot explain all the variations, since these deviations are also observed in inbred strains

(3.1 Gbp) likely due to a higher rate of deletions in the mouse lineage [5]. Such sequence loss indicates that the mammalian genome is a mosaic of sequences of dissimilar importance. This suggestion is supported by the decades-old observations of cytogeneticists who found that certain large chromosomal deletions (i.e. visible through the optical microscope) did not affect the phenotype of mice homozygous for the deletion. Below, we will briefly review the different kinds of DNA sequences within the mammalian genome. Besides the genome sequence of C57BL/6J, deep genome sequencing and variation analysis has been now finalized for new mouse inbred strains, including wild-derived strains [7, 8]. These new sequences show that, remarkably, genetically similar inbred strains can sometimes show divergent phenotypes and that extensive strain-specific haplotype variation still exists in these supposedly completely inbred genomes. These new genomes not only improve the mouse reference genome but also help in the discovery of unknown genes.

Approximately 5% of mammalian genome contains highly conserved sequences, of which no more than 1.5% encode proteins (one estimate is 1.27% for the mouse genome and 1.0% for the

human genome) [9] (Fig. 2). The remaining 3.5% consists of sequences whose functions are only partially known but includes sequences important for regulating gene expression (e.g. DNA-binding sites), chromosome architecture and folding and binding to the mitotic spindle. Interestingly, some of these conserved non-coding sequences have been completely eliminated in mice without substantially affecting phenotype [10].

Annotation of the mouse and rat genomes (the process of identifying functional elements along the DNA sequence) is progressing thanks, in part, to the thousands of spontaneous and induced mutations. Yet, only 14,700 mouse genes have been functionally annotated based on the existence of one or more mutant alleles or through expression assays (MGI, October 2018). Because many genes are conserved in both sequence and function, genes identified in any one of the human, mouse or rat genomes may also aid in the annotation of related genes in the other species. For example, approximately 99% of mouse genes have a human orthologue. This and other examples clearly justify the ‘comparative genomics’ approach [11–13].

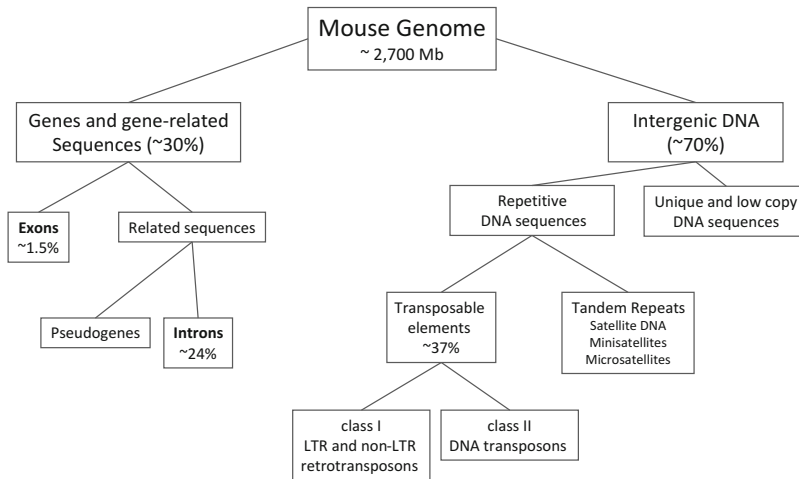


Fig. 2 Types of DNA in the mammalian genome. The graphic shows the different types of DNA sequences present in mammalian genomes, including rodents. It is estimated that only around 30% of the genome is represented by genes (protein-coding sequences) and gene-related sequences (e.g. introns, regulatory sequences and pseudogenes). On the other hand, the so-called intergenic DNA constitutes up to 70% of the genome. This non-coding DNA corresponds to different categories of repeti-

tive and transposable sequences, together with single copy and low copy number sequences (see text for details). This DNA (inaccurately referred to as ‘junk’ DNA) is poorly known; however, many non-coding DNA sequences are highly conserved between mammals, most likely because they have important biological functions. At the same time, genetic variations in non-coding sequences have been widely used as tools in rodent genetics, including quality controls

1.2.1 Genes, Gene Families and Pseudogenes

Mouse and rat genes have an architecture similar to other mammalian genes, typically composed of coding exonic and non-coding intronic sequences flanked by additional canonical upstream and downstream sequences. The smallest gene known is 0.1 kbp and encodes t-RNA^{Tyr}. The biggest gene is 2.3 Mbp in mouse, rat and humans and encodes dystrophin (*Dmd*). Gene introns also vary in size, ranging from 0.5 kbp for the shortest intron to 30 kbp for the biggest *Dmd*, with an average intron size of 4.7 kbp. For exons, the average exon size is 300 bp with the shortest being only 9 bp (exon 2 of *MyoVla*) and the longest being 7.6 kbp (exon 26 of *Apob*). The number of exons per gene varies from 1 to 314 with an average of 7.5 [14], and about 4000 genes have only one exon.

As in other species, mouse and rat genes are alternatively spliced, meaning that not all exons of a given gene are represented in all transcripts (mRNA) from that gene. Alternative splicing is a clever, evolutionarily conserved mechanism that

allows more than one protein to be encoded by a single gene, based on the exons present in a particular transcript. This also means that the number of genes in an organism does not reflect the degree of genetic complexity of that organism. Instead, the total number of exons may provide a better estimate of complexity. Interestingly, interspecific comparisons indicate that although most exons in the mouse, rat and human genomes are strongly conserved, exons present only in alternatively spliced forms are less conserved and likely represent recent exon creation or loss events [15].

Interspecific comparisons of mouse and other mammalian genomes indicate that the mouse genes are *syntenic* with those of humans and rats. That is, most mouse genes are conserved in blocks, with the same linear arrangement as in the human or rat genomes. For example, when a hypothetical gene G_2 is flanked by genes G_1 and G_3 in mouse, there is a very high probability that the same linear order (G_1 - G_2 - G_3) is preserved in the other two species. This conservation of *synteny* (from the Greek, meaning ‘on the same ribbon’) helps validate candidate genes. It also

aids in identifying duplications and/or deletions among species. For example, about 90% of the mouse and human genomes can be partitioned into regions of conserved synteny, reflecting the structural organization of the chromosome in the common ancestor. These genomes share about 350 segments of conserved synteny, with sizes ranging from 300 kbp to 65 Mbp.

In contrast to genes conserved across species, the mouse and rat genomes also contain rodent-specific genes. The majority of these belong to gene families associated with reproductive functions, exhibiting spermatid- or oocyte-specific expression, or with vomeronasal receptors [9, 16]. Some of these new genes originate from relatively recent duplication events in the mouse lineage subsequent to its divergence from the rat, around 20 million years ago (<http://www.timetree.org/>). In comparison, the human genome (the primate lineage) has lost genes coding for olfactory and vomeronasal receptors [17].

The mammalian genome contains a great number of sequences that resemble protein-coding genes but are not. These *pseudogenes* may be processed or unprocessed. Processed pseudogenes originate from the retro-transcription of messenger RNAs back into the genomic DNA in more or less random locations. They lack introns and contain mutations, including frameshift mutations and premature stop codons, indicating that they are not transcribed. Unprocessed pseudogenes arise from either the tandem duplication of a gene during DNA replication or are degenerated genes that become inactive and are no longer under selection. There are roughly 12,000 pseudogenes in the mouse genome assembly (Mouse Reference GRCh38), but identifying them is often difficult. Synonymous mutations, those that will not modify the amino acid sequence, occur at the same frequency in genes and pseudogenes, whereas non-synonymous mutations are rare in functional genes. The ratio of the number of non-synonymous substitutions to the number of synonymous substitutions in orthologous genes is a strong evidence for deciding whether a 'gene' is a true gene or a pseudogene.

As mentioned, the majority of the mammalian genome consists of non-coding sequences. How-

ever, even some non-coding sequences are highly conserved between humans and rodents, likely because they have important biological functions [18]. The function of these conserved non-coding sequences is the subject of intense research, and it has been suggested that these sequences may be associated with certain diseases [19]. However, a significant portion of non-coding DNA is not conserved and therefore exhibits a higher degree of genetic variation (polymorphism) than conserved non-coding DNA.

1.2.2 Repetitive DNA Sequences

Repetitive DNA sequences are non-coding sequences present in multiple copies within mammalian genomes. Depending on the number of repeats, they are classified as moderately or highly repetitive DNA sequences. The latter include tandem and interspersed repeats. Interspersed repeats are derived from transposable elements, as explained in Sect. 1.2.3. Tandem repeats form when multiple copies of a motif are adjacent to each other in the genome. Depending on the number of nucleotides in the motif, these repeats are categorized as *satellite* DNA (between 120 and 250 nucleotides), *minisatellites* (between 10 and 60 nucleotides) and *microsatellites* (between 2 and 6 nucleotides). Polymorphisms result from variations in the number of tandem repeats within a locus and allow different alleles to be distinguished. In the mouse, satellite DNA comprises about 5% of the genome with major satellite repeats being 6 Mb long and located pericentrically and minor satellite repeats being from 500 kb to 1.2 Mb long and located in the centromere [20]. Minisatellite loci, also known as variable number tandem repeats (VNTRs), are 5–10 kb in size, extremely abundant and distributed throughout the mammalian genome [21]. These highly polymorphic loci were used as genetic markers in the late 1980s, particularly in human studies. They were also the basis for the famed DNA fingerprinting that revolutionized forensic science [22]. However, even though minisatellites were used in a few mouse linkage studies and for the genetic monitoring of inbred strains (isogenic individuals within an inbred strain share the same band pattern) [23–25], the use of DNA fingerprinting in

genetic monitoring was quickly surpassed by the use of microsatellite makers. Microsatellites are very abundant (hundreds of thousands of copies per genome), extremely polymorphic and widely distributed throughout the genomes of animals and plants. Since the early 1990s, microsatellites have been ideal genetic markers because their analysis is simple, affordable and highly reliable [2]. Microsatellites are valuable for genome scans in linkage studies and background characterization of mouse and rat inbred strains [26, 27]. The use of microsatellites for genetic quality control is described in Sect. 4.

1.2.3 Copy Number Variations, Indels, Transposable Elements and SNPs

Although deletions, insertions and other large genomic rearrangements have been known since the 1980s, over the last decades, there has been an increasing interest in the study of segmental duplications and copy number variations (CNVs). CNVs are structural variants that result in copy number changes in a specific chromosomal region. As a consequence, certain large DNA segments (from 1 kb to several Mb and with more than 90% sequence conservation) can vary in copy number when compared with a reference genome, with other individuals of the same species or between inbred strains. Most importantly, CNVs are thought to affect gene expression (altering transcript dosage) and phenotypic variability in genetic diseases (e.g. affecting the penetrance of the trait) [28]. This can be particularly relevant given that the genomes of two randomly selected individuals may differ by at least 1%, mainly due to CNVs and SNPs. In the mouse, approximately 100 genomic regions harbour CNVs across the 19 autosomes, ranging in size from 20 kb to 2 Mb [29–31]. The change in gene dosage associated with these CNVs could easily explain their involvement in phenotypic variation in the mouse [32].

Transposable elements (TEs), found in virtually all eukaryotes, are genomic DNA sequences that move from location to location and exist as interspersed, repetitive DNA sequences. TEs can be inserted into different locations through

DNA recombination, and after many generations, the repeated sequence can spread over various regions. There are two classes of TEs: class I, composed of long terminal repeats (LTRs) and non-LTR retrotransposons, which transpose via an RNA intermediate in a ‘copy and paste’ fashion, and class II, composed of DNA transposons, further divided into subclasses 1 and 2, which use a ‘cut and paste’ mechanism that does not involve an RNA intermediate [33, 34]. LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) are among the most studied class I non-LTR retrotransposons.

LINEs are autonomous retrotransposons and include the family of LINE-1 (L1) sequences, the most active non-LTR element identified in mammalian genomes, with 100,000 copies per haploid genome. SINEs are non-autonomous retrotransposons with repeated motifs of approximately a few hundreds of base pairs. Common examples are the Alu sequences in humans and the B1 and B2 sequences in mice, rats and other rodents [35]. In evolutionary terms, these interspersed sequences are classified as lineage-specific (added to the mouse or rat genomes after the divergence from a common ancestor with other rodents) or ancestral (before the divergence). It is estimated that lineage-specific sequences make up 32% of the mouse genome, compared with 24% in the human genome. In contrast, ancestral sequences represent only 5% of the mouse genome, compared with 22% of the human genome [36].

The nature of the TE-host relationship (e.g. parasitism, symbiosis or commensalism) and the role of TEs in disease and evolution have been debated extensively. There are several reports of human diseases caused by L1-driven insertional mutagenesis [35], but compared to endogenous retrovirus insertions, LINE- and SINE-related pathologies are less common in mice [37]. Even though the role of TEs in the evolution of vertebrate genomes remains controversial, these mobile elements can facilitate sequence-mediated chromosomal rearrangements that can potentially generate new gene regulatory sites [38]. Finally, these transposable elements have made pathways to new germline mutagenesis systems, such as Sleeping Beauty and PiggyBac,

in the mouse and other mammals [39, 40]. This section would not be complete without mentioning endogenous retroviruses. Retroviral infections have also shaped the rodent genome. Endogenous retrovirus expression has been associated with both physiological function and disease [41]. In the mouse, a classic example of an endogenous retrovirus acting as a mutagen is the insertion into the hairless (*Hr*) gene creating the hairless (*hr*) allele [42]. Here, the insertion affects a gene splicing event and results in a hairless phenotype.

Although single nucleotide polymorphisms (SNPs) have been known for many years, their use in linkage and genome-wide association studies has rapidly expanded more recently. A SNP (pronounced ‘snip’) is a single nucleotide change identified by comparing the genomes of individuals of the same species or inbred strains (Fig. 3). SNPs are the most abundant genetic variation and are present in both coding and non-coding sequences. In coding sequences, non-synonymous SNPs create an amino acid change, whereas synonymous SNPs do not. Nonsense SNPs introduce a premature **stop codon**. Almost all SNPs are bi-allelic; only two variants segregate in a population (e.g. homozygous G/G or T/T or heterozygous G/T). In humans, the frequency of certain SNPs varies between populations, that is, a SNP allele can be common in one geographical or ethnic group and atypical in another [43]. Inbred mouse and rat strains possess long segments of DNA with either extremely high (40 SNPs per 10 kb) or extremely low (0.5 SNPs per 10 kb) levels of polymorphism, creating SNP-poor and SNP-rich genomic segments [36, 44]. Nonetheless, several SNP panels, with markers evenly distributed across the mouse and rat genome, have been developed [45–47]. The use of SNPs for genetic quality control will be presented in Sect. 4.

1.2.4 Functional Annotation of the Mouse Genome

As discussed earlier, the massive size and heterogeneous sequence structure of the mammalian genome makes it difficult to analyse. Some elements are repeated, some are unique and some

are present but not essential. To make sense of the bulk of available sequence data, creating and improving the current reference gene annotation that identifies and describes gene structures are essential.

Gene annotation procedures are largely computational but are continually refined manually. We believe that annotation efforts should concentrate on the myriad of genomic transcripts (tRNA, rRNA, shRNA, miRNAs, snoRNAs, lncRNA, etc.) rather than genomic sequence per se. Both the GENCODE and FANTOM projects are essential to the process. The GENCODE project (<https://www.genecodegenes.org/mouse/>) produces comprehensive gene annotation for the reference mouse genome [48]. The FANTOM consortium (Functional Annotation of the Mammalian Genome), at RIKEN in Yokohama, has collected and sequenced 103,000 full-length mouse cDNAs [49]. The FANTOM project has been fundamental; it improved estimates of the total number of genes (and their alternative transcript isoforms) in the mouse, expanded our knowledge of gene families and revealed that a large fraction of the *transcriptome* is non-coding. Currently, tissue-specific expression of genes is being catalogued; consequently, it is already possible, for example, to make an exhaustive inventory of those genes that are expressed in the brain at a particular embryonic day [50] (see the Eur-express Atlas at <http://www.eurexpress.org/ee/>).

Readers seeking more detailed genomic information can consult the Mouse Genome Informatics (MGI) resource [51], an international database that provides integrated genetic, genomic and biological data. The MGI consortium (<http://www.informatics.jax.org>) coordinates several databases and resources, including the Mouse Phenome Database (MPD), the Mouse Genome Database (MGD), the Gene Expression Database (GXD), the Mouse Tumor Biology Database (MTB), the Gene Ontology Project (GO), MouseMine, the International Mouse Strain Resource (IMSR), Cre recombinase activity data, on-line books and information regarding standard nomenclature. The MGI-LIST is a forum for topics in mouse genetics and MGI news updates. It is an active, moderated, email-

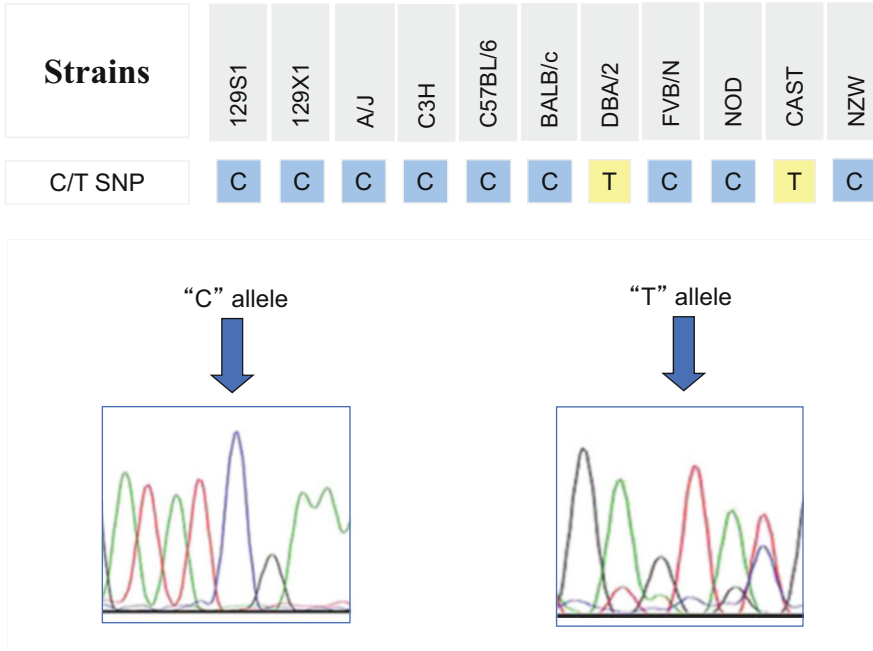


Fig. 3 Single nucleotide polymorphisms (SNPs). SNPs are discrete DNA sequence variations occurring when a single nucleotide in the genome differs between members of the same species. These SNPs are common and they are scattered throughout the genome of all species. They result from random point mutations occurring at a constant rate during evolution, either in the coding regions or in between genes, and they are inherited like a Mendelian trait. In the mouse genome, they are very unevenly distributed along the chromosomes with ‘SNP-rich’ and ‘SNP-poor’ regions

depending on the phylogenetic origin of the chromosomal segment. This allows the determination of a SNP pattern, which is unique to a given strain and accordingly can be used for assessing strain purity. The upper panel represents a C/T SNP that is polymorphic between strains DBA/2 and CAST (homozygous for the ‘T’ allele) and other common inbred strains (homozygous for the ‘C’ allele). The lower panel presents DNA sequencing electropherograms showing the SNP (arrow)

based bulletin board for the scientific community supported by the MGD User Support group.

The Rat Genome Database (RGD, <http://rgd.mcw.edu>) provides the most comprehensive data repository and informatics platform related to the laboratory rat, one of the most important model organisms for disease studies. It includes (i) genomic variation, (ii) phenotypes and diseases, (iii) data related to the environment and experimental conditions and (iv) datasets and software tools that allow the user to explore and analyse the interactions among these and their impact on disease [52, 53].

2 Standardized Strains of Laboratory Rodents

Clarence C. Little, while at Harvard University, was the first to try to develop ‘pure’ mouse

lines by inbreeding. Simultaneously, Helen D. King worked towards developing inbred rat lines at The Wistar Institute, eventually creating the WKA and PA inbred rat strains. The first mouse inbred strain, *dba*, was started in 1909 by Little through inbreeding mice homozygous for three recessive coat colour alleles (*d*, dilute; *b*, brown; and *a*, non-agouti). Similarly, Little established strain C57BL/6 in 1921 via a cross between two ‘black’ mice, female 57 and male 52, obtained from Miss Abbie Lathrop, a retired teacher and a mouse supplier from Massachusetts. A few other mouse strains were developed concurrently by other scientists, in particular Leonell C. Strong (C3H strain) at Cold Spring Harbor and Nadine Dobrovolskaia-Zavadskaia in Paris [54, 55]. In addition to these North American and European researchers, Japanese scientists established a number of colonies from fancy mice [56].

2.1 Inbred Strains and Substrains

According to the definition of the International Committee on Standardized Genetic Nomenclature for Mice, ‘Strains can be termed *inbred* if they have been mated, brother \times sister (sib-mating), for 20 or more consecutive generations, and individuals of the strain can be traced to a single ancestral pair at the 20th or subsequent generation’. However, it has been estimated that 24 generations of sib-mating are needed to reach a heterozygosity rate $< 1\%$ and 36 generations to reach complete fixation [57] and be regarded, for most purposes, as genetically identical (Fig. 4a). In practice, most of the mouse strains commonly used in research laboratories have undergone several tens of generations of brother \times sister matings (indicated with an ‘F’, for filial), with some of the oldest lines surpassing 200 generations (e.g. in 2018 DBA/2 J reached F224). The definition of an inbred strain calls for some explanation. Individuals of the same inbred strain are genetically identical except for the sex-linked characters, and because of strict inbreeding, all of the individuals of a given strain have become *homozygous* at all loci that were segregating in the founder ancestors (the original or ancestral breeding pair). Each mouse is homozygous for the same allele, meaning that the maternal and paternal chromosomes are identical. This is also known as *autozygosity* because the two alleles are copies of the same ancestral allele. To describe this important characteristic, geneticists refer to the animals as being genetically identical or *isogenic*. The process leading to homozygosity by progressive allele loss (or fixation) is simply that, if an allele that was present at generation F_n is not transmitted to at least one member of the breeding pair at generation F_{n+1} , then it is permanently lost. In other words, as inbreeding progresses, alleles are constantly lost but never introduced (with the exception of de novo mutations), leading to both homozygosity and isogenicity (Fig. 4b) [2].

During inbreeding, the progression towards homozygosity is not linear. During the first few generations, many genes become homozygous, but fewer genes become homozygous in subse-

quent generations. Still, after 20 generations of inbreeding, no more than 2% of the loci that were heterozygous in the ancestors will still be segregating. This is because the genes becoming homozygous are linked and arranged linearly on chromosomes and the evolution towards homozygosity involves variable-sized blocks of DNA, not individual genes. This also explains why independent inbred strains carrying the same allele at a given locus have a greater chance of sharing the same short segment of neighbouring DNA (haplotype) flanking the allele in question. For example, if we analyse four classical albino strains (A, AKR, BALB/c and SJL), they are likely to be homozygous for the same short segment of chromosome 7 that flanks the albino mutation (Tyr^c), because the mutation shared by these strains results from an event that occurred well before the creation of these strains (i.e. identical by descent). In fact, all of the common albino rat strains share the same *Tyr* missense mutation, suggesting that they also share a common ancestor [58].

In most mammalian species, inbreeding of a natural population often has deleterious effects of variable intensity. These adverse manifestations are commonly referred to as *inbreeding depression*. Recent genetic studies suggest that inbreeding depression is caused predominantly by the presence of recessive deleterious mutations in natural populations that are progressively fixed in the homozygous state while inbreeding progresses. Alternative explanations, such as epistatic interactions, are also possible. Surprisingly, inbreeding depression is not a serious issue in some rodent species if the breeders stem from the same natural population of closely related individuals. Besides mice and rats, there are a few inbred strains from other rodents, like the Syrian hamster (*Mesocricetus auratus*) LSH/N strain, the guinea pig (*Cavia porcellus*) classical 2/N and 13/N strains and the gerbil (*Meriones unguiculatus*) MON/Tum strain.

The fact that all members of the same inbred strain are nearly genetically identical is the major reason why they have become so prevalent in biomedical research. Scientists working with the same inbred strain, but in different laboratories or at different time periods, can perform

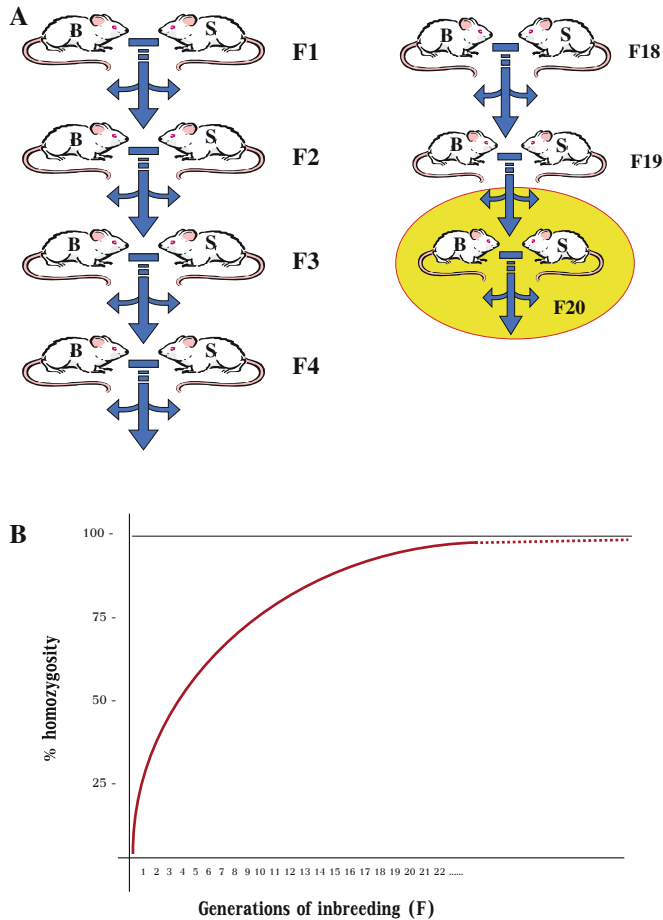


Fig. 4 Inbred strains. (a) This drawing represents schematically the breeding system that is commonly used to produce an inbred strain: mating a male and a female from the same litter (brother x sister) in successive generations. The uppercase letter F, followed by the number of generations, symbolizes each generation of inbreeding. When this number is not known, a question mark is often used; F? + 27, for example, would indicate that the number of brother x sister matings was not known when the strain was acquired, but 27 generations of

unrelaxed inbreeding have been added since this time. According to the definition of the International Committee on Standardized Genetic Nomenclature for Mice, strains can be termed inbred if they have been mated (sib-mating) for 20 or more consecutive generations. (b) The curve was drawn based on the Fibonacci series and represents relatively faithfully the cumulated percentage of genes that have become fixed in the homozygous state as inbreeding progresses. From generation F5 onwards, this percentage is incremented by 19.6% at each generation

experiments where, by definition, variations in experimental results will not be due to differences in the genetic constitution of the animals. Finally, being isogenic, mice and rats of the same inbred strain are also *histocompatible* (or syngeneic). This means that they permanently accept tissue transplantations from any individual of the same strain (and sex). Researchers have used this peculiarity extensively, since it allows studying the fate of cells with an immunological function in

different contexts (cellular cooperation), especially for the serial transplantation of cancer cell lines.

While inbreeding effectively eliminates a proportion of new mutant alleles, another fraction may become progressively fixed in the homozygous state (estimated between 10 and 30 mutations per generation) and replace the original allele, a process known as *genetic drift*. Genetic drift, a slow but unavoidable natural process, con-

tributes inexorably to strain divergence and the generation of *substrains* when the same strain is propagated independently in different places [59]. Examples of mouse substrains are abundant, for example, there are 10 BALB/c substrains and 15 C57BL/6 substrains including the J and N substrains from The Jackson Laboratory and the National Institutes of Health, respectively. Some spontaneous mutations differentially segregate in these common substrains of C57BL/6, first separated in 1951. These include a retinal degeneration mutation in the *Crb1* gene (*Crb1^{rd8}*) and a non-synonymous SNP in the *Cyfp2* gene, present only in the N substrain, and a deletion in the *Nnt* gene, present only in the J substrain [60–62] (Table 1). The most comprehensive comparative phenotypic and genomic analysis of these popular substrains was recently published [63]. Notably, we can take advantage of genetic drift to accelerate the identification of causative mutations resulting in phenotypic differences between closely related substrains [62]. Considered as substrains (although we could argue that they are just related strains), the 129 family of strains is unusual for its high level of divergence, including different coat colours. For example, 129X1/SvJ and 129P3 strains are albino (or chinchilla), whereas 129S1, 129S4, 129S6 and 129S7 (*Still* group) are agouti [64] (Table 2) (for more information see http://www.informatics.jax.org/mgihome/nomen/strain_129.shtml). In the same way, many rat inbred strains present at least two substrains, for example, SHR has four substrains, including SHR/Ola and SHR/NCrI, and WKY and F344 have three substrains each. Substrain variability has been confirmed by sequencing for these rat substrains, with WKY showing the highest degree of substrain variation [65].

The insidious and unavoidable occurrence of new mutations in strains justifies the recommendation in the Guidelines for Nomenclature of Mouse and Rat Strains that inbreeding should never be relaxed. Inbreeding is inefficient in preventing mutations but helps eliminate a substantial proportion of new mutant alleles, thus preserving the genetic profile of a given strain. Similarly, the same international committee on nomenclature has stated that two strains

with the same origin, but separated in different colonies for 20 or more generations (e.g. 12 generations in laboratory A and 10 in laboratory B), should be considered two different substrains and designated appropriately. The Institute of Laboratory Animal Resources (ILAR) maintains the International Laboratory Code Registry (<https://www.nationalacademies.org/ilar/lab-code-database>). Each lab code contains one to five letters and identifies the institute, laboratory or investigator that produced and/or maintains a particular strain [66].

Inbred strains are often described as artificial populations because their genetic constitution (isogenicity and homozygosity) has no natural equivalent. This description is supported by historical records indicating that modern mouse lines do not stem from a single subspecies of the *Mus* genus. Indeed, the polyphyletic origin (i.e. from different subspecies) of modern inbred strains has been substantiated by the complete high-resolution sequencing of the genomes of a large panel of inbred strains [45, 67]. Overall, the genomes of inbred laboratory mice are a mosaic of chromosomal regions with distinct subspecific origins. Recent estimates indicate classical inbred strains were predominantly derived from *M. m. domesticus* (94%), with variable contributions from *M. m. musculus* (5%) and *M. m. castaneus* (<1%) subspecies [68].

Over the last 30 years, a variety of strains derived from small groups of wild specimens trapped in well-defined geographical regions and belonging to well-characterized taxonomic groups, have been established in various laboratories [69]. With the increasing use of PCR amplification for the detection of genetic polymorphisms, the inbred strains derived from these wild populations have become valuable for gene mapping. Examples of these strains are PWK/PhJ (*Mus m. musculus*), MOLD/RkJ (*Mus m. molossinus*) and CAST/EiJ (*Mus m. castaneus*). Special mention must be made of those derived from *Mus spretus* (SEG/Pas, SPRET/Ei and STF/Pas) because this species is one of the most distantly related to the laboratory strains that can still produce fertile hybrids with them. In contrast to laboratory

Table 1 Mutations present in the different C57BL/6 substrains

Substrain	Vendor	Affected Gene (mutant allele)	<i>Snca</i> (<i>Snca^{del,1}</i>)	<i>Mmrrn1</i> (<i>Mmrrn1^{del,ya}</i>)	<i>Nlrp12</i> (<i>Nlrp12^{B6J}</i>)	<i>Dock2</i> (<i>Dock2^{Hsd}</i>)	<i>Cyfpip2</i> (<i>Cyfpip2^{B6N}</i>)
C57BL/6J	JAX (C57BL/6J)	<i>Crb1</i> (<i>Crb1^{rd8}</i>) Mutated	Wild type (WT)	Wild type (WT)	Mutated	Wild type (WT)	Wild type (WT)
	CRL (Europe) (C57BL/6JCrI)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Supposedly WT	Wild type (WT)
	Janvier (C57BL/6JRj)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Supposedly WT	Wild type (WT)
C57BL/6N	JAX (C57BL/6NJ)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Mutated
	CRL (C57BL/6NCRl)	Wild type (WT)	Wild type (WT)	Wild type (WT)	Supposedly WT	Wild type (WT)	Mutated
	Envigo (C57BL/6NHsd)	Mutated	Wild type (WT)	Wild type (WT)	Supposedly WT	Mutated	Mutated
	Taconic (C57BL/6NTac)	Mutated	Wild type (WT)	Wild type (WT)	Supposedly WT	Wild type (WT)	Mutated
	Janvier (C57BL/6NRj)	Mutated	Wild type (WT)	Wild type (WT)	Supposedly WT	Supposedly WT	Mutated
C57BL/6ByJ	JAX	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Not tested	Wild type (WT)
C57BL/6JEJ	JAX	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Not tested	Wild type (WT)
C57BL/6JOlaHsd	Envigo	Wild type (WT)	Mutated	Mutated	Not tested	Not tested	Not tested
C57BL/6JRccHsd	Envigo	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Not tested	Not tested
C57BL/6JBomTac	Taconic ^b	Wild type (WT)	Wild type (WT)	Wild type (WT)	Not tested	Not tested	Not tested

^aDel(6)*Srca1*Slab^bY-chromosome mutation recently detected in C57BL/6JBomTac

Table 2 Current nomenclature and coat colour for the 129 families of strains

	Old nomenclature	New nomenclature	Abbreviation	Coat colour	ES cells
P (parent) group	129/ReJ	129P1/ReJ	129P1	White-bellied, pink-eyed, light chinchilla (light tan)	
	129/OlaHsd	129P2/OlaHsd	129P2	White-bellied, pink-eyed, light chinchilla (light tan)	E14, E14.1
	129/J	129P3/J	129P3	White-bellied, pink-eyed, light chinchilla or albino	EMS32
	129/SvImJ	129S1/SvImJ	129S1	White (or light)-bellied agouti	W9.5, CJ7
S (steel) group	129/SvPas	129S2/SvPas	129S2		D3, D3H
	129/SvJae	129S4/SvJae	129S4		AK7, RF8
	129/SvEvBrd	129S5/SvEvBrd	129S5		Lex-1, Lex-2
	129/SvEvTac	129S6/SvEvTac	129S6		IT2, KG1
	129/SvEvBrd	129S7/SvEvBrd	129S7		AB1, AB2.1
	129/SvEv	129S8/SvEv	129S8		
	129/Sv	129T1/Sv	129T1	White (or light)-bellied chinchilla	CI368
	129/SvEms	129T2/SvEms	129T2		
T (teratoma) group	129/SvJ	129X1/SvJ	129X1	White-bellied, pink-eyed, light chinchilla or albino	CI, CI3

mice, all laboratory rat strains have been derived exclusively from *Rattus norvegicus* (no subspecies are recognized).

2.2 F1 Hybrids

F1 hybrids result from the cross of two inbred strains and are heterozygous at all loci for which the parental strains have different alleles but, like inbred strains, are genetically uniform (Fig. 5). They are also histocompatible and permanently accept tissue transplantations from either parental strain, from their littermates and from all their offspring; however, the parental strains will not

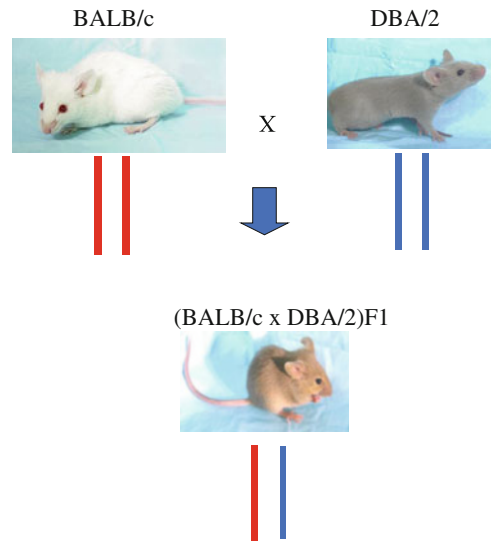


Fig. 5 Hybrid F1. This figure depicts the creation of hybrid F1 mice by intercrossing parental inbred strains BALB/c (albino, with coat colour loci AA;bb;cc;DD) and DBA/2 (diluted brown, with loci aa;bb;CC;dd). Below the mouse pictures, only one pair of chromosomes is shown as an example, with different colours representing the different backgrounds (although not all alleles will be polymorphic between the parental strains). Note that the hybrid F1 mouse obtained has the characteristic brown agouti ‘cinnamon’ coat colour (Aa;bb;Cc;Dd genotype). The standard nomenclature is (BALB/c x DBA/2)F1 (maternal strain listed first). Also acceptable is the abbreviated version CD2F1. Note that hybrid F1 mice are isogenic because they all receive the same maternal and paternal chromosomes. However, crossing F1 mice will generate hybrid F2 mice that are not isogenic because they will have recombinant chromosomes showing different patterns of BALB/c and DBA/2 alleles

accept a graft from the F1 hybrids. F1 mice and rats also exhibit *hybrid vigour* (heterosis), the opposite of inbreeding depression, making them the material of choice in many experimental protocols, e.g. in the protocols aimed at the production of genetically engineered animals. In this case, F1 hybrids are used because of their robust production of preimplantation embryos that are highly resistant to manipulation (e.g. DNA pronuclear microinjection). However, a major drawback is that their progeny (F2) is genetically heterogeneous when intercrossed, since the alleles at all polymorphic loci start segregating, due to meiotic recombination events, in the F1 gametes. Interstrain hybrids can also be used to generate genetically heterogeneous populations. For example, F1 hybrids between strain A and strain B (abbreviated ABF1 or AXBF1) can be crossed with F1 hybrids between strain C and strain D (CDF1 or CXDF1) to generate a four-way heterogeneous stock. In this case, the basic ingredients of the genetically heterogeneous stock (i.e. the original inbred strains A, B, C and D) are perfectly identified, and similar, but not identical stocks can be produced.

2.3 Co-isogenic, Congenic and Consomic Strains

When a mutation occurs in the breeding nucleus of an inbred strain, and the new mutant allele has replaced the original one (probability = 0.25), the new inbred strain differs from the original at only that one specific locus. If the new mutant is viable and the mutation does not impair fertility, the new strain can be propagated by mating brother to sister mutant mice or, preferably, by mating, at each generation, to a nonmutant mouse of the original inbred strain. The original strain and new mutant strain are *co-isogenic*. Co-isogenic strains are extremely useful for gene annotation because they allow a comparison of the phenotypes associated with the original and mutant alleles without the influence of genetic background. A large number of co-isogenic strains are held in several mouse and rat repositories worldwide. Some common mouse strains, like C57BL/6, have several co-

isogenic ‘companion’ strains segregating for a variety of allelic forms controlling, for example, coat colour. Co-isogenic C57BL/6-*Tyr^c* (albino) mice are commonly used to create easily recognizable chimeric mice derived from C57BL/6 ES cells injected into albino C57BL/6-*Tyr^c/Tyr^c* blastocysts [70]. In addition to coat colour, other mutations in co-isogenic strains may cause detrimental effects on development or metabolism. These strains have aided the analysis of developmental and metabolic pathophysiology by providing both the experimental animal and its control. However, co-isogenic strains have two major drawbacks inherent to their origin: (i) they arise mainly as a consequence of a rare mutation, and (ii), although they can emerge in any inbred strain, they generally emerge in a strain other than the one of primary interest.

Congenic strains are an alternative to co-isogenic strains with the advantage that any allele of interest may be moved (i.e. *introgressed*) into any inbred background. The *donor strain* carries the allele or chromosome region of interest (i.e. spontaneous, induced or targeted mutations, as well as transgenes) and is crossed to the *recipient* or *background strain*. The F1 offspring generated by crossing the donor and recipient strains are again backcrossed to the background strain, and the offspring that carry the allele of interest (i.e. the one originating from the donor strain) are repeatedly backcrossed to the background strain, typically for ten or more successive generations (Fig. 6), unless marker-assisted breeding is used (see Sect. 4.4). Ideally, the crosses initiate with a donor female and a recipient male. Then, the F1 mutant males will carry the correct Y-chromosome, and after mating to a recipient female, males of the N2 generation will carry the correct X- and Y-chromosomes of the recipient strain.

During the successive backcrosses, the chromosomes of the background strain progressively replace those of the donor, except for the one that carries the allele of interest. For this chromosome, the segment containing the selected allele is reduced in size only when a recombination event occurs that replaces a piece of chromosome of the donor for the homologous segment of the

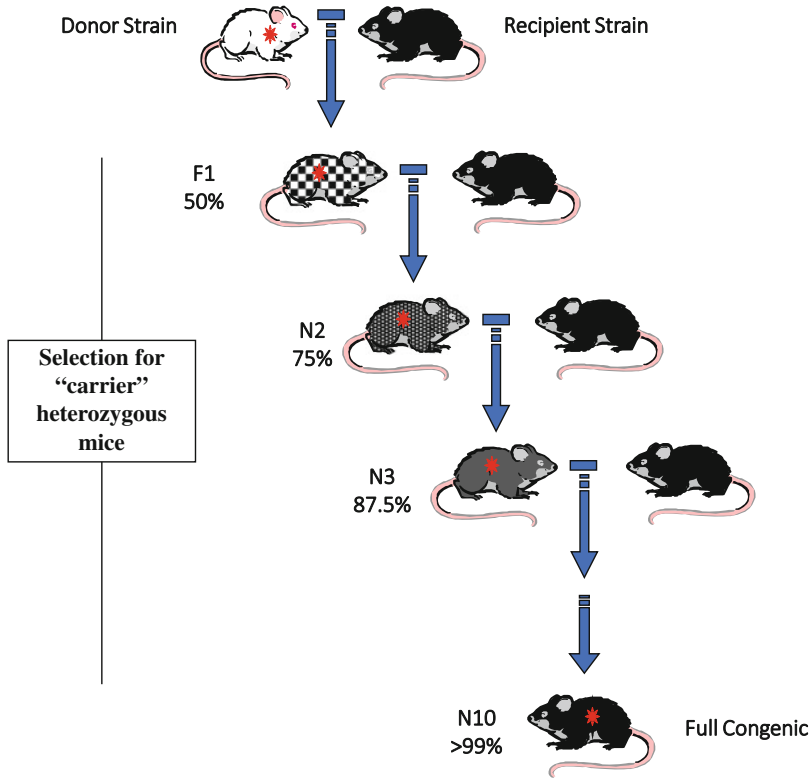


Fig. 6 Congenic strains. This scheme represents the successive steps in the establishment of a congenic strains. The first step is to cross a mouse from the donor strain (albino in the example) carrying the gene of interest (e.g. a transgene or a targeted null allele) with a mouse from the recipient inbred strain. At each generation a breeder carrying the gene of interest (*) is backcrossed to a partner of the recipient strain (black in this example). The letter

'N' is used to indicate the generation of backcross, starting with N2. The degree of grey colour is only to show how, after each backcross generation, the offspring have increasing amounts of the recipient genome. After each backcross generation, on the average, 50% of the genomic DNA of the donor strain is replaced by the equivalent proportion of genomic DNA of the (recipient) background strain

background strain. Over generations, such replacement events cause the chromosome carrying the targeted allele to gradually be 'eroded' on both sides of the allele in a nonlinear manner. Ultimately, the chromosomal segments flanking the selected locus generally remain associated with it, thus marking the basic difference between congenic and co-isogenic strains. In other words, while co-isogenic strains differ from the background strain at a single locus, congenic strains differ not only at the locus but also by a short chromosomal segment flanking the targeted locus, with the size of the flanking region being progressively reduced during backcrossing.

On average, at each generation, an equivalent proportion of the background strain replaces

one half of the genome of the donor strain; thus the progression of genome substitution is given by the formula $1/2^N$, where N is the number of backcross generations. Theoretically, after ten backcross generations, only $1/2^{10}$ (1/1000) of the donor genome will remain in the congenic strain; however this is only an approximation. The actual percentage of donor genome replaced at each generation will vary. In addition, and as previously discussed, this estimate is valid only for those chromosomes lacking the allele of interest. For the chromosome bearing the allele of interest, the reduction in size is a much slower process. It is estimated that there is only a 10% chance that the segment carrying the introgressed gene will be smaller than 1 cM after a series of ten

backcrosses. This is not negligible: on average, 1 cM (1.8 Mbp) of the mouse genome will contain dozens of genes, depending on the region. Congenic strains have been used extensively since the early days of mouse genetics and are still used as tools for the analysis of quantitative (complex) traits. It is precisely by developing such strains that George D. Snell and his colleagues from The Jackson Laboratory were able to elucidate the genetic determinism of histocompatibility resulting in a Nobel Prize in 1980 to G.D. Snell, J. Dausset and B. Benacerraf.

Consonic strains, also called chromosome substitution strains (CSSs), are a variation on the congenic strain concept, but the introgressed DNA is a complete chromosome, rather than a piece of chromosome flanking a given gene [71]. These strains are useful for rapidly mapping phenotypic traits to a specific chromosome and for QTL analysis. QTLs, or quantitative trait loci, are chromosomal regions that influence a particular complex, multigenic/multifactorial phenotype (e.g. resistance or susceptibility to carcinogenesis). However, in consonic strains, small fragments of donor strain chromosomes might escape the selection process.

2.4 Recombinant Inbred Strains and Recombinant Congenic Strains

Recombinant inbred strains (RISs) are developed by crossing two parental inbred strains to generate F1 hybrids followed by intercrossing these F1 to generate F2s. Then, randomly chosen F2 animals are brother-sister mated over 20 or more generations to develop a group of related inbred strains (Fig. 7) [72]. A collection of RISs derived from the same parental strains form a set (also referred to as a panel). For example, the largest RIS mouse panel is currently C57BL/6 × DBA/2 (BXD) with more than 100 strains and thousands of measured phenotypes and typed genetic markers (see GeneNetwork at <http://www.genenetwork.org/webqtl/main.py>). RISs are true inbred strains (an ‘immortal’ resource), homozygous at all loci but with a unique, fixed combination of parental

alleles in a 50:50 ratio (on average). For example, each strain of the set of 33 AXB-BXA strains, derived from the initial cross of a C57BL/6 mouse with a A/J mouse, carries either the B6 allele or the A allele at each genetic locus. By typing all of these allelic forms, one can establish a strain distribution pattern (SDP) for each strain, listing the collection of alleles inherited from either parental strain A or parental strain B6. High-resolution maps of some mouse RISs and CSSs are also available [57]. Sets of rat RISs have also been created between the LE/Stm and F344 inbred strains (LEXF) [73]. Overall, RISs have proven very helpful for gene mapping, particularly for the rapid regional assignment of microsatellites on a given chromosome. They have also been used to map QTLs involved in controlling behaviour (e.g. alcohol intake, etc.) and certain immunological responses.

Recombinant congenic strains (RCSs) resemble RISs in their genomic structure except that the proportion of the parental alleles in a given strain is not 50:50 but 75:25 or 87.5:12.5, depending on the set. RCSs are established by inbreeding mice of the first or second backcross generation onto the background strain. RCSs are helpful for identifying genes associated with polygenic inheritance, especially when the number of genes is high. For example, RCSs have been very helpful for unravelling the genetic determinism of colon cancer in the mouse [74]. Interspecific recombinant congenic strains (IRCSs) have also been developed from the parental strain C57BL/6JPas and SEG/Pas (*Mus spretus*) [75]. This set of strains has proven particularly useful for describing the genetic basis of some anatomical traits [76].

2.5 The Mouse Collaborative Cross

The *Collaborative Cross* (CC) is a variation on the RIS concept but with a much higher power of resolution and level of genetic diversity segregating in the panel [77, 78]. The CC is a randomized cross of eight inbred mouse strains that have been carefully selected by a panel of mouse geneticists

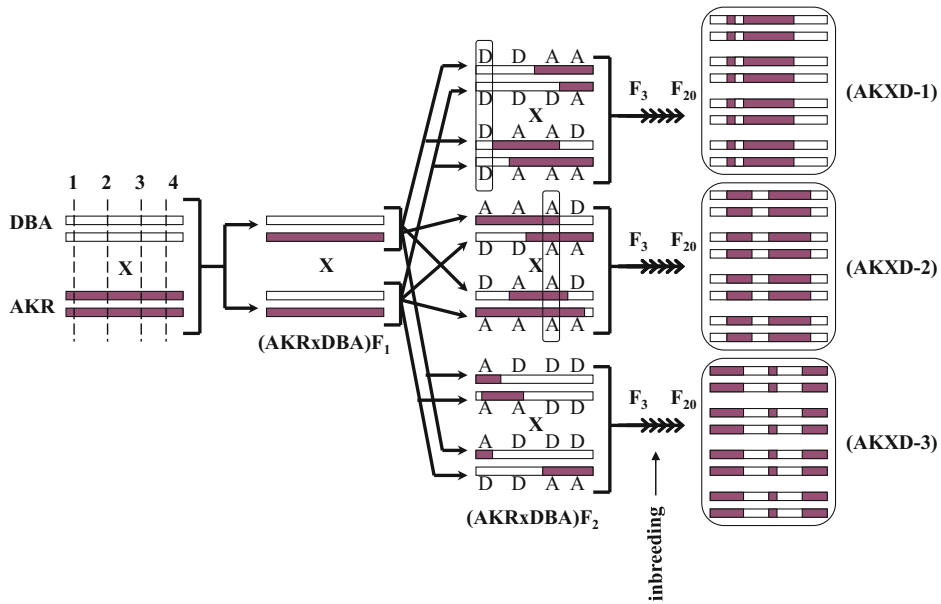


Fig. 7 Recombinant inbred strains. This diagram represents the creation of a set of three recombinant inbred strains (RIS) originated by intercrossing parental inbred strains DBA (D) and AKR (A) (only one pair of chromosomes is shown as an example). The positions of four hypothetical loci are indicated with dotted lines in the parental chromosomes (numbers 1–4). The rectangles show alleles that are already fixed (D or A) in some breeders at the F2 generation. After >20 generations of inbreeding, we obtained truly inbred strains that carry,

on average, 50% of alleles from each parental strain. The boxes on the right represent the same chromosome pair showing identical patterns in four random mice from three different RIS (AKXD-1, AKXD-2 and AKXD-3). Individual RISs have a unique combination of loci derived by recombination of the alleles present in the original parental strains. Since RISs are inbred and each strain has a unique genotype, RISs have a number of advantages over F2 or backcross mouse populations as tools for mapping genes or quantitative trait loci (QTL)

(the Complex Trait Consortium). These strains consist of (i) three classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ), (ii) two inbred strains afflicted by diabetes or obesity (NOD/LtJ and NZO) and (iii) three strains recently derived from wild progenitors (CAST/Ei, PWK and WSB/Ei). The eight strains are first crossed pairwise to make all ($8 \times 7 = 56$) possible G1 parents; then all eight genomes are brought together in a series of crosses, and the offspring of these crosses are inbred for several generations (Fig. 8). Several hundreds of new inbred strains (recombinant for variable proportions of the original eight parental strains) are progressively becoming available. These strains can be used to make biologically relevant correlations among thousands of measured traits providing an unprecedented power of resolution [79, 80]. To

increase mapping resolution power, investigators may also use the first-generation (F1) progeny from crosses of CC strains (designated CC-recombinant intercross or CC-RIX).

2.6 Outbred Stocks

Outbred stocks are populations of laboratory animals that are genetically heterogeneous and therefore radically different from those already discussed. Outbred stocks are ‘closed populations (for at least four generations) of genetically variable animals that are bred to maintain maximum heterozygosity’. Compared with inbred strains, F1 hybrids and congenic strains, the genetic constitution of a given animal taken randomly from an outbred stock is not known a priori. Outbred

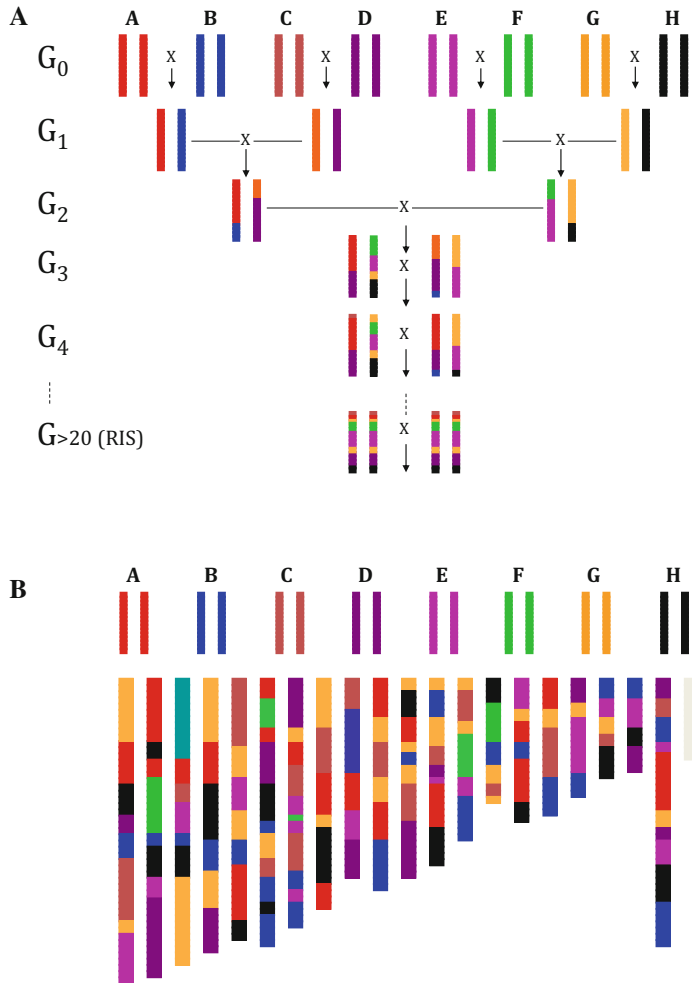


Fig. 8 The Collaborative Cross (CC). **(a)** This is a randomized cross of eight unrelated mouse inbred strains designed by members of the Complex Trait Consortium. The lines are first crossed pairwise to make all 56 possible G₁ parents. A set of possible four-way crosses is performed, keeping Y-chromosome and mitochondrial balance. Finally, all eight genomes are brought together in G₂:F₁, and the offspring of this cross are inbred. The Collaborative Cross is a community resource that was initially designed

for the purpose of mapping complex traits. **(b)** The initial previsions were to breed around 1000 inbred strains where all the alleles of the initial inbred strains would be associated in a wide and unique variety of combinations. Only one strain is represented in this illustration; other strains would be similar but with a different pattern of parental strain distribution. The pool of strains selected for the CC is constituted by five classical unrelated inbred strains (A/J, C57BL/6J, 129S1, NOD and NZO) and three wild-derived strains (CAST/Ei, PWK/PhJ and WSB/Ei)

stocks are normally bred according to a system that minimizes inbreeding and maintains a certain amount of heterozygosity in the population [81]. One frequently used outbreeding system is the ‘rotational breeding’ system described by Poiley [82]. Software for generating random mating schemes is freely available [83].

The degree of genetic heterogeneity in outbred colonies depends on colony history [84]. Heterogeneity can be very low, for example, as a consequence of genetic drift (or the bottleneck effect) or when the pool of breeders has been accidentally or intentionally reduced to a few individuals, as is common when starting a new

breeding program with a small group of imported breeders. In contrast, genetic heterogeneity can be very high when the stock has been recently outcrossed. Although the methodology and results are not always made public, it is likely that reputable commercial breeders regularly monitor the polymorphisms segregating in their outbred stocks. Examples of outbred stocks of mice are ICR (CD-1), CFW and NMRI (all derived from the original ‘Swiss’ mice imported to the USA by Clara J. Lynch in 1926) and the non-Swiss CF-1 mice [84]. Examples of outbred rat stocks are Sprague Dawley (SD), Wistar (WI) and Long-Evans (LE). Outbred stocks of other laboratory rodents, including guinea pig, Syrian hamster, Chinese hamster (*Cricetulus griseus*), gerbil, cotton rat (*Sigmodon hispidus*) and sand rat (*Psamomys obesus*), are also available.

Because outbred colonies, like human populations, are heterogeneous, they are often considered the most appropriate category of laboratory animals for toxicology and pharmacology research. However, several geneticists have disputed this point and have even suggested that in many studies, outbred mice were used inappropriately, wasting animals’ lives and research resources on suboptimal experiments [85]. In fact, any outbred stock can be replaced with a ‘synthetic’ population obtained by intercrossing classical inbred strains. As mentioned, crossing two inbred strains to produce F1 progeny followed by crossing two independent F1 individuals generates a four-way polymorphic population. This population is heterogenic, in the sense that individuals are genetically different. In addition, the population often carries a greater number of allelic forms, which is generally considered an advantage compared to a classical outbred population. Recently, however, researchers have realized that outbred stocks might be useful for refining QTL mapping experiments, because these heterogeneous stocks accumulate many recombination breakpoints that over time split their chromosomes into ‘fine-grained mosaics’, facilitating high-resolution mapping of complex traits [86, 87]. Other investigators recently claimed that contrary to conventional understanding, outbred

mice might be better subjects for some biomedical research [88].

3 Genetically Altered (GA) Rodents

There are numerous terms used to describe genetic changes in rodents. In mice, the terms genetically engineered mice (GEM) and genetically modified mice (GMM) typically describe any genetically modified mouse. Here, we use the term genetically altered (GA) rodent to also include animals carrying spontaneous and/or chemically induced mutations and refer to ‘lines’ rather than ‘strains’ for GA rodents. GA lines are created using various genetic manipulation technologies that are summarized in several popular books and articles [89–91]. We also recommend visiting the webpage of the International Society for Transgenic Technologies (ISTT) at <http://www.transtechsociety.org/>.

3.1 Spontaneous and Chemically Induced Mutants

Every scientist in charge of a colony of inbred mice or rats, even if only for a few years, has almost certainly discovered a mutation segregating in a breeding nucleus. For example, dominant spotting (*Kit^W*), a mutant allele of the oncogene *Kit*, is very common and easy to identify on a C57BL/6, C3H or CBA background because it lightens coat colour, particularly in the tail, and often induces a white belly spot. In fact, 74 spontaneous mutations have been identified for *Kit*, with similar but not completely identical phenotypes. Other mutations are also quite common, especially those with an obvious viable phenotype (e.g. skeletal anomalies, cerebellar defects, neuromuscular syndromes, anaemia, skin defects and inner ear defects), and are generally either recessive or dominant. Since inbreeding increases the level of homozygosity in populations, it also enhances the probability of discovering recessive mutant phenotypes; however, inbreed-

ing does not primarily increase the frequency of mutations.

It is also important to classify mutations based on their effect on the activity of their gene products. For example, an *amorphic* allele (null or loss of function) will eliminate activity completely, whereas a *hypomorphic* allele will produce a gene product with less activity than the wild-type gene product. In the same way, a *hypermorphic* allele will have increased activity, a *neomorphic* allele will have a new function and an *antimorphic* allele will have a dominant negative function.

Spontaneous mutations typically occur at low frequency, but frequency varies among loci. Some advantages of working with spontaneous mutations are that they are produced at virtually no cost and are usually freely available. In addition, they generally have an obvious phenotype, given that they are identified based on observation. Collectively, spontaneous mutations represent a great variety of molecular events, including deletions, insertions and point mutations. Such mutations generate not only loss-of-function alleles but also hypomorphs and hypermorphs. In many cases, spontaneous mutations can help establish better animal models than those produced by KO models [92–94]. Unfortunately, spontaneous mutations also have drawbacks. One major disadvantage is that the mutation's primary molecular defect is almost always unknown and therefore has unpredictable utility for gene annotation. Nonetheless, documenting spontaneous mutations is important; the Mouse Mutant Resource (MMR) at The Jackson Laboratory has been characterising (genetically and phenotypically) mice carrying spontaneous mutations for decades.

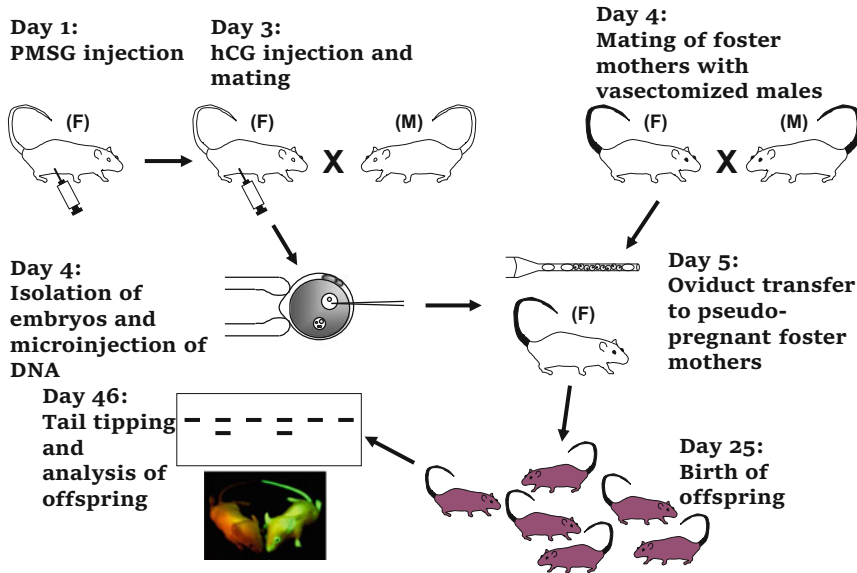
Ever since William Russell, of Oak Ridge National Laboratory, USA [95], reported that N-ethyl-N-nitrosourea (ENU) was 'the most potent mutagen in the mouse', ENU and other chemical mutagens have been used to generate mutations. ENU has numerous advantages as a mutagen, and its mode of action has been studied extensively [96, 97]. ENU is an alkylating agent producing mostly base pair changes (point mutations). In optimal conditions, ENU induces an average of 0.7–1.9 nucleotide substitutions per Mbp of DNA or one mutation at a specific locus in every 670–

1000 mice of a G3 generation. Several collaborative projects aimed at the mass production of new mutant alleles were launched in the late 1990s, particularly in Europe, Japan and North America [98, 99]. In most instances, these projects were associated with downstream phenotypic screens designed to recover specific types of mutations (e.g. mutations leading to neuromuscular defects or to deafness). Interestingly, data contained in the Mutagenetix database (<https://mutagenetix.utsouthwestern.edu>) of mouse phenotypes and mutations induced with ENU indicates, based on over 100,000 mutations, that putative null mutations have a 61% probability of causing (phenotypically) detectable damage in the homozygous state [100].

Forward genetics is one genetic strategy used to identify the gene(s) responsible for a particular phenotype or biological process. It is a bottom-up approach that proceeds from the phenotype to the genotype. In this strategy individuals with spontaneous or induced mutations causing a phenotype of interest supply the raw material. Mapping the mutation requires subsequent breeding and a genetic map with as many informative genetic markers as possible [101]. *Positional cloning* is the process of identifying a gene based on its position in the genome, without any prior idea of its function. A good historical example of positional cloning is the identification of the gene responsible for the obese mutation (*ob*, later renamed *Lep^{ob}*) [102].

3.2 Classical Transgenesis by Pronuclear Microinjection (Random Insertion)

Transgenic rodents are created by the microinjection of foreign DNA fragments directly into one of the two pronuclei of one-celled embryos (zygotes), a technique widely used in the mouse and to a lesser extent in the rat [103–105]. In this process of *additive transgenesis*, the microinjected transgene randomly integrates into the genome as a concatemer with variable copy number (Fig. 9). The mouse and rat models created with this system typically overexpress a transgene placed un-



Flowchart for the generation of transgenic mice

Fig. 9 Producing transgenic mice by pronuclear injection. The flowchart represents the different steps for the production of transgenic mice by pronuclear injection. One-cell embryos are flushed out of the oviduct immediately after fertilization, and then the transgene is microinjected *in vitro* with a glass micropipette into one of the pronuclei (typically the male pronucleus). Once injected, the embryos are kept *in vitro* for a few hours

and then transplanted into pseudo-pregnant females (previously mated with vasectomized males). Genotyping of the G0 (presumptive) transgenic mice can be achieved at any time from birth onwards. Every pup genotyped as positive by PCR (i.e. hemizygous Tg/0 carrier) or expressing a reporter protein (e.g. GFP) should be considered a 'founder', and independent lines should be developed from each founder

der the control of a tissue-specific, developmental stage-specific or ubiquitous promoter (along with other regulatory elements), all contained in the transgene DNA construct.

The number of copies of the transgene that integrates into the host genome is not controlled and ranges from one to several tens or hundreds. DNA copies are generally arranged in head-to-tail arrays in the transgenic insertion with potential rearrangements in the flanking regions. In addition, the site of integration is random and can seriously influence transgene expression due to position effects. Position effects cause unpredictable, unexpected and somewhat erratic variations in transgene expression. For example, when an insertion occurs in a hyper-methylated region of the genome, the transgene will be weakly or not expressed. Position effects are one of the main weaknesses of pronuclear transgenesis. As it is impossible to predict either the integration site

or the number of copies that will integrate, it is impossible to know how well a transgene introduced by this method will be expressed. Therefore, when developing a transgenic line, it is highly recommended to compare the offspring of several different founder mice. Likewise, it is important to avoid intercrossing mice originating from different founders; independent transgenic lines should be developed from each founder.

The recommended generic symbol for a transgenic insertion is Tg. Founder transgenic animals are hemizygous for the newly introduced DNA segment and are designated Tg/0. Establishing a transgenic line, in which the transgene is propagated by sexual reproduction, requires genotyping each generation to which the transgene was transmitted, unless the carriers have an obvious phenotype [106]. Lines are normally kept by backcrossing transgenic carriers (hemizygous Tg/0) with wild-type animals from

the inbred background strain and by selecting carriers at each generation. When viability and fertility are unaffected, a transgene may be maintained by keeping transgenic lines in the homozygous state. Traditionally, to distinguish between homozygous (Tg/Tg) and hemizygous (Tg/0) mice, the mouse of interest was crossed to a non-transgenic partner, and the progeny was statistically analysed for Mendelian segregation of the transgene. Today, quantitative real-time PCR (qPCR) can be used to distinguish hemizygous from homozygous transgenic mice [107]. In order to achieve a pure genetic background, it is recommended to inject the transgene into embryos derived from an inbred strain, such as FVB/N, which is widely used because its zygotes possess large and prominent male pronuclei and the females are excellent breeders that produce large litters [108].

A later improvement on the original constructs used for transgenesis was the introduction of inducible systems allowing transgene expression to be turned on and off. Currently, the most common strategies are the Tet-on and Tet-off expression systems. In these systems, transcription of a given transgene is placed under the control of a tetracycline-controlled transactivator protein, which can be regulated, both reversibly and quantitatively, by exposing the transgenic mice to either *tetracycline* (Tc) or one of its derivatives, such as *doxycycline* (Dox). Both Tet-on and Tet-off are binary systems that require the generation of double transgenic (*bigenic*) mice. These mice carry both a responder construct, consisting of a tetracycline response element (TRE)-regulated transgene, and an effector construct (tTA or rtTA), containing a tetracycline-controlled transactivator [109].

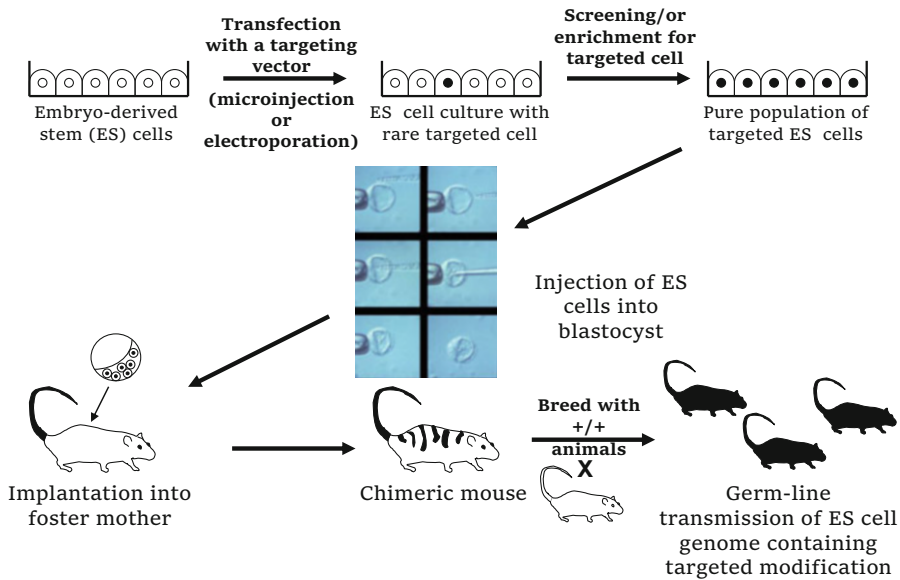
3.3 Targeted Mutagenesis Using ES Cells

Another mouse genetic engineering technology uses pluripotent embryonic stem (ES) cell lines. ES cells are undifferentiated pluripotent embryonic cells derived from the inner cell mass of preimplantation blastocysts that can

participate in the formation of the germ cell lineage of chimeric mice, an indispensable step in generating founder mice carrying the targeted mutation (Fig. 10). Most early ES cell lines were derived from embryos of the 129 families of inbred strains (129S2, 129P3, etc.). Today, ES cell lines come from a variety of strains. For example, the ES cell lines derived from C57BL/6N have become widespread and are often selected for many international projects (e.g. EUCOMM). In contrast to mice, the development of germline-competent ES cells in rats has only recently become possible [110], and their use remains limited.

Chimeras resulting from the admixture of engineered ES cells (carrying the targeted mutation in the gene of interest) with cells of the inner cell mass of a recipient blastocyst can be identified as soon as a few days after birth based on their dappled coat colour. The dappled coat is obvious when the ES cells are derived from C57BL/6N (which is non-agouti *a/a* – i.e. solid back) and the recipient blastocyst is from either a wild-type (agouti *A/A*) or albino (*Tyr^f/Tyr^f*) strain. In these conditions, the chimeras exhibit a mixture of black and agouti (or albino) spots. Using coat colour as a reference, one can estimate the degree of chimerism, but a high level of chimerism does not necessarily parallel with a high rate of germline transmission. Although chimeras can be from either sex, males are generally the only sex with germline transmission because the majority of ES cell lines are XY. To avoid mixed background lines down the road, it is recommended to generate co-isogenic KO/KI lines by crossing the chimeras with wild-type mice from the same inbred background as the ES cells. For example, when C57BL/6-derived ES cells are injected into albino C57BL/6 blastocysts, the chimeric mice are easily identified because their coats exhibit white and black patches. These chimeras can then be crossed with albino C57BL/6 mice to test for germline transmission, validated by the appearance of ES cell-derived black offspring [70].

Other gene-targeting strategies have been developed to create conditional rather than constitutive KO mutations. Conditional mutations bypass



Generation of germ line chimeras from embryo-derived stem (ES) cells

Fig. 10 Targeted mutagenesis in the mouse using engineered ES cells. The flowchart represents the different steps for the production of targeted mutants (KO and KI) using genetically modified ES cells. Pluripotent ES cells can be cultured *in vitro*, for several generations, remaining in an undifferentiated state. While *in vitro*, the ES cells can be manipulated like ordinary somatic cell lines and selected on the basis of specific criteria. ES cells are then typically injected into blastocysts (less commonly into eight-cell or morula stage) where they spontaneously

merge with the inner cell mass. After embryo transfer into the uterus of a pseudo-pregnant female, and provided that the ES cells are still pluripotent, fertile chimeric mice can result from these reconstructed blastocysts. The chimeras with the best level of chimerism are then crossed with wild-type mice in order to confirm germ line transmission, basically the production of genotypically heterozygous mice carrying the targeted allele. One extra generation is necessary to observe the alteration in the homozygous state

some of the drawbacks of using constitutive null alleles of endogenous genes (e.g. pre- and post-natal lethality, fertility and welfare problems). With conditional mutations, the time and tissue in which the gene is inactivated can be controlled. Conditional KO production requires a cross between two independent lines to generate bigenic mice. The most popular conditional KO strategy is based on the Cre-*loxP* system, although a FLP-*FRT* system also exists. In the Cre-*loxP* strategy, Cre recombinase, derived from bacteriophage P1, cuts and recombines the DNA strand at specific sites called *loxP* sites (short for locus of X-ing over P1). These *loxP* sites consist of two 13-bp inverted (palindromic) repeats separated by an 8-bp asymmetric spacer region that define the orientation of the site. When the *loxP* sites are in the same orientation and on the same strand (*in cis*), the intervening stretch of DNA is excised as a

circular loop. When two *loxP* sites are in opposite orientations and on the same chromosome, the intervening DNA segment is inverted. Finally, when the *loxP* sites are on two different chromosomes (*in trans*), the recombinase generates a reciprocal translocation [111].

The Cre transgene can be made inducible, adding more sophistication to the system, for example, by using Cre^{ERT2}, which can be induced by administration of tamoxifen [112]. Nowadays, many Cre-expressing lines are produced as KI mice with the Cre sequence incorporated directly into the gene of interest (rather than creating transgenic lines using pronuclear microinjection). The Cre-*loxP* strategy can also be used to control the expression of reporter genes. For example, the *lacZ* gene can be driven by a ubiquitous promoter (e.g. *Rosa 26*) with a floxed 'stop' sequence consisting of a short segment of DNA

with several termination codons inserted between the promoter and the *lacZ* coding sequence, thus preventing translation of the *lacZ* gene product beta-galactosidase. When the Cre activity causes deletion of the floxed ‘stop’ sequence in specific cells or tissues, beta-galactosidase is produced in those cells or tissues. [113]. Because of the widespread use of this conditional targeting approach, databases cataloguing strains that synthesize Cre (designated Cre-deleters), either ubiquitously or in specific tissues, have been developed (see, for example, The Jackson Laboratory Cre Portal at <https://www.jax.org/research-and-faculty/resources/cre-repository> or the MGI Mouse Recombinase at <http://www.informatics.jax.org/home/recombinase>).

When using the Cre-*loxP* system, keep in mind the following: (i) Results may vary depending upon whether Cre is transmitted from the female or the male parent (e.g. Cre is significantly more efficient when transmitted maternally in the EIIa-Cre line). (ii) The presence of Cre alone might produce a phenotype (always include a Cre + control mouse without floxed sequences). (iii) The Cre-*loxP* system can be combined with the Tet-on or Tet-off inducible system. (iv) Cre mosaicism has been reported in some strains, resulting in variable expression. (v) Some floxed alleles are more easily recombined than others. (vi) Tamoxifen-inducible Cre lines can be leaky, that is, Cre can sometimes be active in the absence of tamoxifen.

3.4 Gene Editing Using Nucleases

Over the last 10 years, several new techniques have been developed using engineered nucleases to create targeted mutations. These techniques provide ES cell-independent approaches for the production of targeted mutations in mice, rats and other species [114].

3.4.1 Zinc-Finger Nucleases and TALEN

The production of mutations using zinc-finger nucleases (ZFNs) relies on the precise design of a chimeric protein containing a specifically

designed zinc-finger DNA-binding domain and a *FokI* endonuclease domain. Two complementary and sequence-specific multifinger peptides are designed to recognize a specific DNA sequence spanning 9–18 bp on either side of a 5–6 bp sequence, which defines the targeted region. When injected into the pronucleus or the cytoplasm of zygotes, the ZFNs bind tightly on both sides of the targeted site, one on each strand, allowing dimerization of *FokI* which then makes double-strand breaks (DSBs) at the selected site. Once cleaved by *FokI*, the cellular mechanisms controlling DNA integrity (DNA repair pathways) are triggered to repair the damage by either homology-dependent repair (HDR) or nonhomologous end joining (NHEJ). HDR requires a homologous sequence as a template to direct repair and accurately re-establish the original sequence. NHEJ is a much less precise mechanism that restores damaged strands incompletely, leaving behind deletions, thus creating frameshifts that commonly result in loss-of-function mutations. ZFN technology can be used to create a homozygous KO mutation faster than traditional KO strategies using ES cells and is applicable to all strains of mice and rats, allowing for the production of mutations in different inbred backgrounds. Mice and rats carrying null alleles or sequence-specific modifications have already been produced using ZFN technology [115, 116].

Like ZFN technology, transcription activator-like effector nuclease (TALEN) technology combines a nonspecific DNA endonuclease having robust cleavage activity with a DNA-binding domain that can be easily engineered to target a particular DNA sequence. In recent years, several groups have used TALENs (originally described in bacterial pathogens of crop plants) to modify endogenous genes in a wide variety of species, including zebrafish, rat, mouse, pig and cow [117, 118]. The advantages of TALENs over ZFNs are easier design and assembly, higher specificity and lower cost.

3.4.2 CRISPR-Cas System

This newly developed technology depends on small RNAs for RNA-guided cleavage of specific

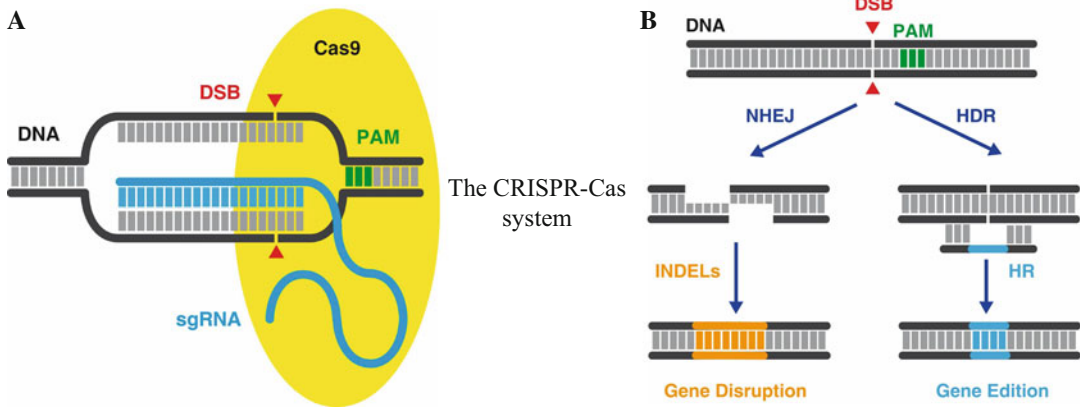


Fig. 11 Genome editing using site-specific RNA-guided DNA endonuclease (CRISPR/Cas system). (a) With the CRISPR strategy, Cas9 unwinds the DNA duplex and performs a double-strand break (DSB) after recognition of a specific (20 bp) target by the gRNA, provided that the correct protospacer adjacent motif (PAM) is present. (b) DSBs are repaired through nonhomologous end joining (NHEJ) or through homology-directed repair (HDR). In the case of DSBs repaired by NHEJ, the mechanism will induce indels and potentially produce KO alleles.

For HDR to occur requires that a DNA molecule or a single-stranded synthetic DNA be added as a template. If the sequence of the template differs from the endogenous sequence by the addition or substitution of some nucleotides (light blue colour), this results in a KI allele. These methods for producing mutations at specifically targeted sites are very efficient. Figures kindly provided by Dr. Lluís Montoliu, CNB-CSIC, CIBERER-ISCI, Centro Nacional de Biotecnología, Campus de Cantoblanco, Madrid, Spain

DNA sequences by a Cas endonuclease. The strategy was developed after the identification and characterization of a primitive bacterial/archaeal defence mechanism called CRISPR-Cas that allows these organisms to fight against infections from viruses, plasmids and phages [119, 120]. Engineered modifications to CRISPR (clusters of regularly interspaced short palindromic repeats) and the Cas enzyme (Cas9 is the most commonly used RNA-guided DNA nuclease) have led to an efficient system to produce DSBs at will. The guide RNA (gRNA or sgRNA) binds to the target DNA sequence and directs the Cas9 nuclease to create precise DSBs at the location of interest (Fig. 11).

RNA-guided endonucleases can be engineered to cleave virtually any DNA sequence by appropriately designing the gRNA, for example, to generate KO mice and rats [121–123]. CRISPR-Cas9 has several advantages over ZFNs and TALENs. It can be used to create mutations in multiple genes across the genome in a single step, by injecting multiple gRNAs targeting different sequences simultaneously. Such multiplex gene

editing has proven successful not only to modify cells in vitro but also to modify mouse and rat embryos [124]. This saves substantial breeding time when several specific mutations are required in the same genome. Given the ease and speed of this method, it is clear why it is revolutionizing mammalian genetic engineering [125–128]. CRISPR-Cas also confers the possibility of producing KO lines on any inbred background because constructs are introduced either by injection into the cytoplasm or pronuclei of one-cell or two-cell stage embryos [129] or by electroporation [130, 131], thus avoiding ES cells and chimera production. However, as each indel mutation generated is unique, CRISPR-Cas-based genetic engineering requires extensive sequencing and bioinformatic analyses to characterize multiple founders (G0) to ensure against mosaicism and off-target mutations while also verifying the presence of the expected genetic change. The selected founder should then be bred with wild-type animals to evaluate transmission of the mutation to their offspring.

4 Genetic Quality Control for Mice and Rats

Genetic markers are specific DNA sequences with a known chromosomal location. The current gold standard for genetic quality control of laboratory rodents requires the analysis of polymorphic genetic markers that can distinguish between different genetic backgrounds. Historically, many of the techniques used to detect and analyse these markers have been shared with forensic DNA profiling.

4.1 Current Tools for Genetic Quality Control

Although many polymorphisms have been described in the mouse and rat, only two types are widely used in modern QA programmes: microsatellites (also known as simple sequence length polymorphisms (SSLPs) or short tandem repeats (STRs)) and/or SNPs. It is still too early to determine whether high-throughput, whole-exome sequencing (sequencing the exons of all protein-coding genes in a genome) will be useful for QA purposes, but it does provide both a robust method to discover hereditary factors contributing to rare Mendelian disorders in humans and a means to identify the precise molecular aberration underlying mutations mapped through positional cloning in mice and rats [132]. Whole-exome sequencing could also be very useful for the characterization of substrains.

4.1.1 Microsatellite (SSLP) Markers

Microsatellite markers are still used in genetic quality control programmes because they are extremely easy to type at a very low cost. Microsatellite analysis requires PCR amplification of the short, tandemly arranged, repeating DNA sequences, typically di- and tri-nucleotides (Fig. 12). The PCR products, 100–300 bp in size, are analysed on agarose or polyacrylamide gels. There are enormous numbers of microsatellite loci in the mouse and rat genomes (10^5), and

identifying a set of markers whose amplification products will create a strain-specific pattern is not generally problematic. Routine analysis of DNA samples with microsatellite markers will confirm isogenicity (in the case of inbred strains) and provided the markers have been carefully selected, strain authenticity. One advantage of microsatellites is that they are multiallelic markers, meaning that, when tested in different inbred strains, a single marker can identify multiple alleles, distinguished by PCR products of different sizes. Microsatellite technology has been enhanced through the introduction of fluorescently labelled primers combined with capillary electrophoresis to provide a fast, automated system for genetic monitoring [27]. Here, PCR products are distinguished from each other by both their size and the fluorescent dye associated with them. The availability of different dyes allows multiplexing the PCR reaction (i.e. combining multiple primer sets to simultaneously amplify multiple loci in one reaction) and/or pooling several PCR reactions/products into one capillary [46]. Well-defined panels of SSLPs for mouse and rat inbred strains are available [27, 133, 134].

The MGI [51] presents comprehensive SSLP data, including primer sequences and the expected sizes of their amplified products for several mouse inbred strains (<http://www.informatics.jax.org/marker>). A collection of mapped SSLP markers for inbred strains of rats is available at the Rat Genome Database (RGD).

4.1.2 Single Nucleotide Polymorphisms (SNPs)

SNP genotyping is inexpensive and can be performed in most research institutions or outsourced to providers. Petkov and co-workers from The Jackson Laboratory have described the allelic distribution of 235 SNPs in 48 mouse strains and selected a panel of 28 such SNPs, enough to characterize most of the approximately 300 inbred, recombinant inbred, wild-derived, congenic and consomic strains maintained at The Jackson Laboratory [135]. This set of markers, encompassing all mouse chromosomes, is an

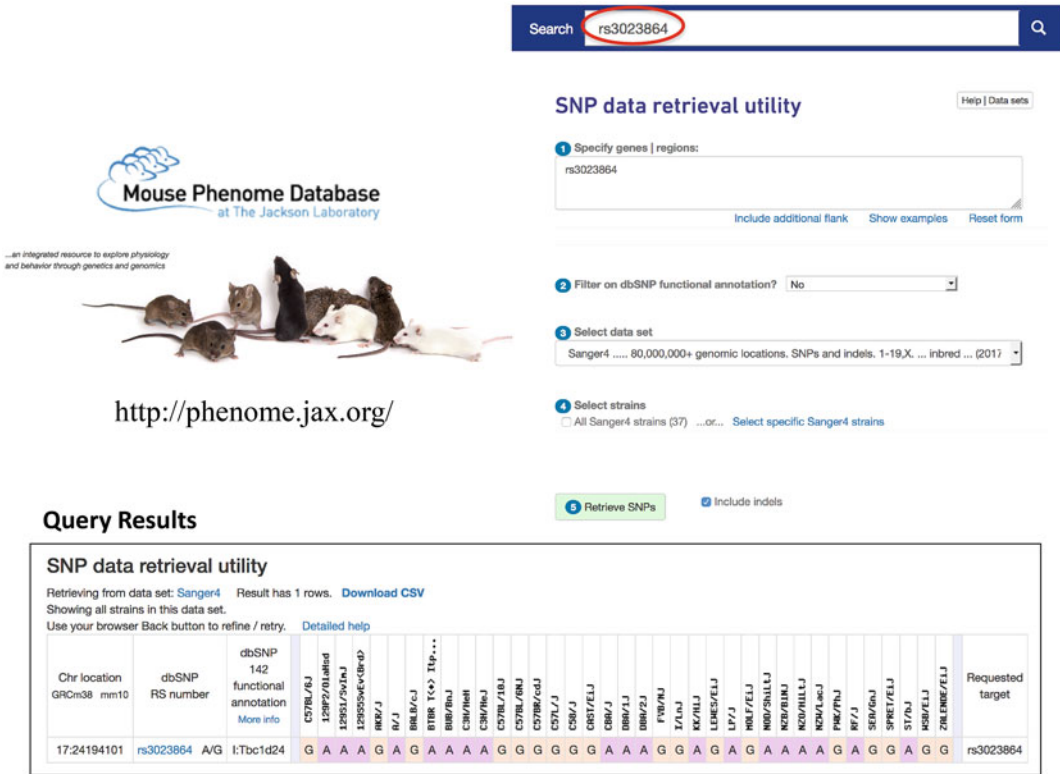


Fig. 13 SNP databases. This figure shows the results of a search using the Mouse Phenome Database seeking for polymorphism between inbred mouse strains for a SNP (ID rs3023864) located on chromosome 17. In this case, the Sanger4 set of 37 inbred strains was selected as dataset.

The bottom of the figure is a screen capture of the results of the query showing the alleles (G or A) present in each of the strains (A = A/A; G = G/G). There are several other options for searching SNP data online (see text)

SNP genotyping is the current method chosen for genetic monitoring by most commercial suppliers of laboratory mice and rats. SNP genotyping assays are currently based on allele-specific PCR (including KASPar fluorescent technology) [47], real-time PCR (TaqMan®), direct sequencing and DNA arrays [101]. Another clever option is to exploit those SNPs that create a restriction fragment length polymorphism (RFLP) [143], making them easy to identify using simple technology. Databases, including the Mouse Phenome Database (MPD), the Mouse Genome Informatics (MGI), the Sanger Institute’s Mouse Genomes Project and the Rat Genome Database (RGD), contain information for hundreds of thousands of SNPs for common mouse and rat inbred strains regarding their genomic locations and

which alleles (C, G, A or T) to expect for a particular SNP/strain combination (Fig. 13).

4.2 Genetic Quality Control of Inbred Strains and Outbred Stocks

Historically, most techniques used to assay the genetic quality of inbred strains were based on the postulate that each inbred strain is expected a priori to be homozygous at almost all loci [144]. These techniques were designed based on the genetic tools available contemporaneously and generally consisted of analysing a few traits, controlled by a set of specific alleles, to define a unique pattern for each strain. Analysis of

biochemical markers, mainly enzymatic proteins (isozymes), by electrophoresis became popular in the mid-1970s; however, this technique was expensive because each test required specific and costly reagents. Other techniques used for genetic monitoring have included immunological markers (particularly H2 haplotype), *osteometry* (mandible) traits and coat colour testcrosses [144–146].

Although genetic monitoring now relies on molecular techniques, the genetic purity of rodent populations must also be considered in a broader context that includes monitoring nonmolecular parameters, such as coat colour, behaviour, characteristics of genetic predispositions, breeding performance and/or other unique strain features [145]. For example, a sudden increase in litter sizes or elevation of the breeding index in an inbred strain is a strong indicator of possible genetic contamination. Likewise, monitoring for strain-specific pathologies is also important for quickly discovering possible genetic contamination and genetic drift.

Commercial breeders are extremely sensitized to the risk linked with genetic contamination and perform regular monitoring of their strains to detect such contamination. Most breeders monitor their nucleus colonies using SNPs, and larger vendors typically establish special programmes to tackle the issue of genetic drift. For example, The Jackson Laboratory has developed the patented Genetic Stability Program, initiated in 2003 [147]. This programme effectively limits cumulative genetic drift by rebuilding foundation stocks from cryopreserved (pedigreed) embryos every five generations. Starting in 2005, The Jackson Laboratory began selling only C57BL/6J mice descended from two chosen mice (Adam and Eve mice) through hundreds of frozen embryos of the duo's grandchildren, enough to last for 25–30 years [148]. For academic institutions, The International Council for Laboratory Animal Science (ICLAS) is promoting and helping develop genetic monitoring programmes to improve the level of QA for academically held mouse and rat models. Current ICLAS recommendations were recently reviewed by Fahey et al. [149].

4.2.1 Genetic Monitoring to Confirm Strain Identity

When inbred mice and rats are kept in-house, it is best to purchase animals from reliable vendors and refresh the colony with mice from the same vendor every 3–5 years rather than maintain independent colonies of classical inbred strains. Established vendors have excellent genetic quality programmes that allow smaller facilities to circumvent genetic monitoring altogether. As an additional benefit, acquiring animals from the same vendor prevents the formation of substrains harbouring potential mutations. Nonetheless, it is the best practice to use a small panel of SSLPs for strain authentication in those facilities that lack sophisticated equipment but wish to authenticate strains in-house. The number of markers to use has not been standardized because each situation and facility is different. However, a panel of 30–40 SSLPs, evenly distributed across the autosomal chromosomes, is generally considered adequate to rule out (recent) genetic contamination, typically resulting from accidental crosses with animals of a different inbred strain or outbred stock. Accidental crosses are more common when a facility maintains strains with the same coat colour in the same room, a particularly dangerous practice if not using individually ventilated cage (IVC) systems. The key characteristic of the SSLP panel used to detect contamination is that the markers must be polymorphic between the suspected strains.

An alternative to authenticating strains maintained in-house is to request SNP genotyping services from a commercial laboratory. Most commercial services are based on fixed DNA microarrays, so it is important to consider that only a fraction of the SNPs on any one array will be polymorphic between the strains under analysis (e.g. 40% for some classical inbred strain combinations). In addition to small-scale SNP genotyping (100–400 SNPs), there are high-density microarrays available. Although high-density arrays were designed primarily for gene mapping purposes, they may also be used to perform a complete SNP profile characterization for new or non-characterized inbred strains and substrains. For example, the Mouse Universal Genotyping Array (MUGA) in its MiniMUGA format has

11,000 SNPs, and the MegaMUGA format has 78,000 SNPs with both being built on the Illumina Infinium platform.

4.2.2 Discrimination of Substrains

The consensus is that if an inbred colony has been isolated for more than 20 generations, it should be considered a substrain, regardless of whether genetic differences between it and the parental strain have been confirmed. Opposed to standard genetic monitoring, the use of SSLPs is not recommended for identification of substrains because there are insufficient numbers of informative markers to distinguish between most of the common substrains. Instead, SNPs should be used, but the initial characterization of a substrain that has been isolated from the parent for several years requires a large set of SNPs. As an example, a pairwise comparison of sister strains using the MegaMUGA array showed that the number of polymorphic SNPs is 154 between C57BL/6J and C57BL/6N, 134 between BALB/cJ and BALB/cByJ and 827 between C3H/HeJ and C3H/HeN [150]. However, only complete exome sequencing can provide exhaustive information regarding specific mutations accumulated in protein-coding genes. Nevertheless, if the goal is only to identify to which classical substrain a colony (or an animal) is associated with, then a small number of SNPs, based on the information available in the SNP databases, can be selected for comparison. This is particularly easy for common substrains such as C57BL/6J and C57BL/6N, where small sets of markers have already been published [137–139].

4.2.3 Genetic Monitoring for Outbred Colonies

Genetic monitoring of outbred stocks is much more complex, because the essential nature of these mice and rats is that they are not genetically uniform. Outbred colonies are groups of closely related animals with common ancestors and group identity (e.g. tame, albino, prolific, etc.), but that still exhibits some level of genomic heterozygosity [81]. Outbred colonies should be treated as a population, making it difficult to establish a standard genetic monitoring programme

with just a few genetic markers. However, monitoring the frequencies of different alleles present in the population with an adequate number of SNPs or SSLPs could reveal stock identity and help preserve the genetic heterogeneity (and allele pool) of a colony. This complex process requires analysing a large number of animals and access to historical allelic frequency (and level of heterozygosity) data for that particular colony.

One of the main issues with maintaining small colonies of outbred rodents with a very small number of breeders is that it reduces the number of alleles in the population and increases the inbreeding coefficient. Therefore, these colonies are neither truly outbred nor inbred. In any case, if it is not possible to keep a large number of breeders, it is better to purchase outbred rodents from vendors that maintain a very large colony and use special breeding schemes that reduce inbreeding.

4.3 Background Characterization for GA Rodents

The recent enormous increase in the number of GA lines will likely exacerbate the problem of undefined ‘mixed backgrounds’ in experimental rodents. This is particularly worrying in the case of inducible and conditional models that require the cross of two independent lines (e.g. Cre-expressing lines crossed with ‘floxed’ lines). It is well recognized that the genetic background (i.e. all genomic sequences other than the gene of interest) can influence the phenotype of an animal model. Spontaneous and induced mutations, transgenes and targeted alleles that are introgressed into a different background have been reported to exhibit altered phenotypes [151, 152]. These changes are mainly due to the influence of modifier genes in the genetic background.

One of the first cases documenting the influence of modifier genes involved the classical diabetes mutation *Lepr^{db}* that presented transient diabetes in the C57BL/6 background but overt diabetes in the C57BLKS background [153]. Later, the dominant *Apc^{Min}* (adenomatosis polyposis coli) mutation presented with an

increased frequency of intestinal tumours in C57BL/6 mice but not in an AKR background. In this case, the responsible genetic modifier is an amorphic allele of *Pla2g2a* fixed in C57BL/6 [154]. Other examples include background effects on survival rate in *Egfr* (epidermal growth factor receptor) KO mice [155], effects on tumour incidence and spectrum in *Trp53* and *Pten* KO mice [156, 157] and milder phenotypes in the *Dmd^{mdx}* mouse model for Duchenne muscular dystrophy when moved to 129X1 [158]. There are also examples from rat models, like the influence of genetic background on prostate tumorigenesis in Pb-SV40 transgenic rats [159] and changes in phenotype severity in *Ednrb^{sl}* mutant rats [160].

On the other hand, mutations hidden in the genomes of introgressed strains or substrains (congenic lines) that can affect the outcome of an experiment are sometimes referred to as ‘passenger mutations’ [161]. There are many examples in the literature where substrains, although stemming from the same original inbred strains, have acquired new and unique phenotypes as a consequence of genetic drift [61, 162]. Mice of the C57BL/6JOLA^{Hsd} substrain, for example, are homozygous for a deletion of the *Snca* locus (encoding for α -synuclein) on chromosome 6 [163]. Alone, this deletion has modest phenotypic effects, but it could interfere unpredictably with other mutations if used as a background strain for making a knockout. Another interesting example stems from using different substrains of C57BL/6 mice as controls in acetaminophen-induced liver injury studies of *Jnk2* KOs. Researchers reported exactly opposite conclusions regarding JNK2 in helping or hurting liver health [164]. Similarly, due to the presence of a spontaneous mutation at the *Tlr4* locus (encoding for a *Toll*-like receptor) in substrain C3H/HeJ, where all mice are homozygous for the defective allele *Tlr4^{Lps-d}*, when C3H/HeJ mice are experimentally infected with Gram-negative bacteria, they may react very differently from mice of substrain C3H/HeN that lacks this mutation [165]. Berghe and colleagues recently reported that passenger mutations are common in most GA lines derived from 129 ES cells and that these mutations persist even after the creation of

fully congenic strains [161]. This is not trivial; Berghe et al. estimate that close to 1000 protein-coding genes might be aberrantly expressed in the 129-derived chromosomal segments that are still segregating in these congenic lines. This finding emphasizes the need for proper controls to identify phenotypes due to background mutations or the combination of background mutations and the genetic modification of interest, rather than the modification itself.

Genome scans can be performed on a GA line with a mixed background to estimate the percentages of the genome contributed by different inbred origins. This process is referred to as a *background characterization* and is a service offered by some commercial enterprises and institutional core facilities. A typical background characterization requires genetic markers that are polymorphic between the most likely involved inbred strains and evenly distributed across the genome. In most mouse cases, these are C57BL/6 (the most common background strain for GA lines) and 129 substrains. The reason for the prevalence of 129 substrains is that, historically, the ES cells needed for the development of KO and KI were derived exclusively from 129 substrains [64]. The dominance of 129 substrains is now slowly changing with the availability of ES cell lines derived from other strains, particularly C57BL/6, and the arrival of genome editing techniques that create targeted alterations in any mouse or rat strain.

In any case, it is recommended to circumvent the problem of mixed background altogether by (i) injecting transgenes or nucleases (Cas9-sgRNA) into inbred embryos from the strain of choice, (ii) modifying the gene of interest in ES cells from the preferred background strain (e.g. using C57BL/6 ES cells) and (iii) crossing chimeras and KO/KI founders with mice of the same strain as the ES cells used for the targeting. Finally, if the GA is already developed (acquired from a collaborator or repository), a background characterization should be performed, and if needed, a fully congenic strain should be established through either classical backcrossing protocols or speed congenics.

4.4 Marker-Assisted Backcrossing (Speed Congenics)

Compared to traditional backcrossing schemes, *marker-assisted backcrossing*, or *speed congenics*, is a rapid and rigorous method that accelerates congenic strain development through the use of DNA markers [166, 167]. The principle that underlies the speed congenic process is based on the selection of breeders, at each generation of backcrossing, based on their percentage of donor genome as determined by analysing the presence of polymorphic genetic markers covering the whole genome. The animal with the lowest percentage of donor DNA is then selected as a breeder for setting the next backcross (Fig. 14). This process greatly reduces the number of generations necessary to reach full congenicity. Using marker-assisted crosses, we can obtain 80% recipient background at N2, 94% at N3 and 99% at N4 (instead of the classical mean values of 75.0%, 87.5% and 93.7%, respectively). It is important to note that once a marker is typed ‘homozygous’ for the allelic form of the background strain, it is no longer necessary to genotype the offspring of the future N generations for this marker because

it is permanently fixed. Using additional markers also assists in the selection of breeders with the smallest amount of flanking DNA, helping to alleviate the ‘flanking gene’ concern [168, 169].

5 Mouse and Rat Phenomics

5.1 Standardized Phenotyping Protocols

Researchers now have all the means and tools to create a great variety of alterations in the mouse and rat genomes. Many of these alterations are expected to result in changes in phenotype, and the careful analysis of these phenotypic changes is fundamental for the process of genome annotation. However, even if it is relatively easy to characterize a DNA sequence, it remains difficult to unambiguously establish the link between a DNA alteration and an abnormal phenotype. The collection of physical and biochemical traits of an animal is known as the *phenome*, and *phenomics* is the discipline that deals with the measurement of these traits.

Speed Congenic Timeline (~18 months)

- Start crossing donor (carrier) female with recipient strain male in order to generate F1 carrier males (PI).
- Backcross F1 males to generate ~20 N2 carrier males (~25%) (PI)
- Scan N2 carriers with SNPs and select the best breeders (Service)
- Cross best N2 males with several recipient strain females (PI)
- Generate ~20 N3 carrier males (~25%) (PI)
- Genotype N3 mice for heterozygous SNPs in N2 analysis (Service)
- Cross best N3 males with several recipient strain females (PI)
- Repeat same scheme at N4 (and if necessary at N5)

Fig. 14 Speed Congenics Timeline. Selecting at each backcross generation, the breeder with the lowest percentage of introgressed (donor) DNA greatly accelerates the establishment of a congenic strain. It is important to note that genotyping requires many polymorphic DNA markers only for the first backcross progeny (N2). Once a marker is characterized as homozygous, it is no longer necessary to type it in the forthcoming generations. Although carrier males (heterozygous for the gene of interest) are typically

recommended as ‘best breeders’, females can also be used, as long as they have high percentages of the recipient genome. The prediction of >98% recipient genome at N5 is based on the use of 20 best breeders (carriers) at each generation (Markel et al. [167]); however this number is not always available, and fewer breeders can be used, with disparate results, depending also on chance. *PI*, Principal Investigator

Phenotyping of rodent models has become a main concern over the last decades. Therefore, many laboratories and institutions have developed highly standardized phenotyping protocols. The range of phenotyping platforms, including dual-energy X-ray absorptiometry, electrocardiography, high-resolution imaging and FACS, ensures the recovery of phenotype data across multiple systems and disease states. In most cases the basic protocols include behaviour, neurology, clinical chemistry, development, immunology, energy metabolism, vision and hearing, pain perception and cardiovascular and gross pathology assessments. The use of standard procedures and defined protocols allows data to be comparable and shareable, even across species, which may help identify mouse and rat models of human diseases [170]. However, phenotyping loss-of-function mutations cannot predict the relevance of these alleles (and their phenotypes) to complex human diseases that are likely driven by several alleles of modest effect [171].

5.2 International Mice Phenotyping Consortia

One of the earliest collaborative projects using standard phenotyping procedures was the European Eumorphia project. This programme also developed the Europhenome data repository and the European Mouse Phenotyping Resource for Standardized Screens [172]. The European Mouse Disease Clinical programme, together with the Sanger Mouse Genetics Program (MGP), continued the collaborative work of Eumorphia, developing protocols and phenotyping mutant mouse lines (mostly from the IKMC mutant ES cell lines) [173]. The Jackson Laboratory has developed a programme to collect baseline phenotypic data on the most commonly used inbred strains of mice through a coordinated international effort. Information collected through this programme (*The Mouse Phenome Database*) is freely available to the community through the Internet (<http://phenome.jax.org/>) [174]. The establishment (and updating) of this database is possible only because inbred mice are isogenic and genetically stable in the long term.

The International Mouse Phenotyping Consortium (IMPC) was established in 2011 with several goals: (i) to maintain and expand a worldwide consortium of institutions with capacity and expertise to produce germ line transmission of targeted KO mutations in ES cells, (ii) to test each mutant mouse line through a broad-based primary phenotyping pipeline, (iii) to systematically aim to discover and ascribe biological function to each gene, (iv) to maintain and expand collaborative ‘networks’ with specialist phenotyping consortia or laboratories and (v) to provide a centralized data centre and portal for free, unrestricted access to primary and secondary data from the scientific community [175]. The current European members of IMPC are the Medical Research Council (Harwell), the Wellcome Trust Sanger Institute (Cambridge) and the European Bioinformatics Institute (Hinxton) in the UK; the Helmholtz-Zentrum Muenchen in Germany; the PHENOMIN (Strasbourg) in France; the CNR (Monterotondo) in Italy; the Czech Centre for Phenogenomics in the Czech Republic; and the Universitat Autònoma de Barcelona in Spain [176]. Phenotyping data are accessible on the IMPC website (<http://www.mousephenotype.org/>). Using both gene trapping and gene targeting approaches, the IMPC has developed mutant ES cells (many with conditional mutations) for more than 18,000 genes representing more than 90% of the mouse protein-coding genes [171]. The ultimate goal is to produce a comprehensive catalogue of mouse gene functions by generating and characterizing null mutations for every mouse gene.

So far, the IMPC has used ES cells [177] to generate the mouse mutants, all on a C57BL/6N background (National Institutes of Health substrain). For example, EUCOMM and KOMP-CSD (CHORI, Sanger Institute and UC Davis) use promoter-less and promoter-driven targeting cassettes for the generation of the KO alleles [178]. This strategy relies on the identification of a critical exon common to all transcript variants that, when deleted, creates a frameshift mutation. The KO-first (*Tmla*) allele is flexible and can produce reporter knockouts, conditional knockouts and null alleles following exposure to site-specific recombinases. For example, excising

the *Tm1a* allele with Cre creates the *Tm1b* (*lacZ* tagged) allele that is a true KO because skipping over the *lacZ* cassette will no longer restore gene expression. The cassette expresses *lacZ* in tissues where the gene of interest is knocked out. Beta-galactosidase staining can be used to follow the tissue expression of the gene of interest. Finally, the *Tm1c* (conditional ready) allele has a phenotypically wild-type state where the exons are spliced together normally. However, the critical exon(s) are still flanked by *loxP* sites. Crosses with tissue-specific Cre-deleter mice can be used to create a tissue-specific KO line. Nowadays, the IMPC is starting to use CRISPR/Cas9 technology to generate the KO mutants by deleting an early critical exon.

The IMPC uses the International Mouse Phenotyping Resource of Standardised Screens (IMPreSS) phenotyping protocols, which are essential for the characterization of mouse phenotypes (see <https://www.mousephenotype.org/impress>). In this case, homozygous (or heterozygous in the case of embryonic lethal mutations) adult mutant mice enter a standardized pipeline [179] where cohorts of males and females undergo a wide range of phenotyping tests from 9 to 16 weeks, followed by a variety of terminal tests. The phenotyping of both male and female cohorts has allowed an in-depth analysis of the extent of sexual dimorphism. To date, the IMPC has generated over 7000 mutant lines, and phenotype data have been collected on over 5000 lines with a large number of novel phenotypes revealed [179]. More importantly, approximately 90% of the gene-phenotype annotations described by the IMPC have not been previously reported [171]. Data from the IMPC shows that around 24% of genes will not produce homozygous KO (null allele) offspring because they are homozygous lethal.

References

1. Guenet JL, Benavides F, Panthier J, Montagutelli X. Genetics of the mouse. Berlin: Springer; 2015.
2. Silver L. Mouse genetics. Concepts and applications. Oxford: Oxford University Press; 1995.
3. MacDonald WA, Mann MR. Epigenetic regulation of genomic imprinting from germ line to preimplantation. *Mol Reprod Dev.* 2014;81(2):126–40.
4. Michaud EJ, Bultman SJ, Klebig ML, van Vugt MJ, Stubbs LJ, Russell LB, et al. A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (*Ay*) mutation. *Proc Natl Acad Sci USA.* 1994;91(7):2562–6.
5. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002;420(6915):520–62.
6. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature.* 2004;428(6982):493–521.
7. Doran AG, Wong K, Flint J, Adams DJ, Hunter KW, Keane TM. Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. *Genome Biol.* 2016;17(1):167.
8. Lilue J, Doran AG, Fiddes IT, Abrudan M, Armstrong J, Bennett R, et al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat Genet.* 2018;50(11):1574–83.
9. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, et al. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* 2009;7(5):e1000112.
10. Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM. Megabase deletions of gene deserts result in viable mice. *Nature.* 2004;431(7011):988–93.
11. Windsor AJ, Mitchell-Olds T. Comparative genomics as a tool for gene discovery. *Curr Opin Biotechnol.* 2006;17(2):161–7.
12. Liao BY, Zhang J. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA.* 2008;105(19):6987–92.
13. Tollis M, Schiffman JD, Boddy AM. Evolution of cancer suppression as revealed by mammalian comparative genomics. *Curr Opin Genet Dev.* 2017;42:40–7.
14. Sakharkar MK, Perumal BS, Sakharkar KR, Kanguane P. An analysis on gene architecture in human and mouse genomes. *In Silico Biol.* 2005;5(4):347–65.
15. Modrek B, Lee CJ. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet.* 2003;34(2):177–80.
16. Choi E, Lee J, Oh J, Park I, Han C, Yi C, et al. Integrative characterization of germ cell-specific genes from mouse spermatocyte UniGene library. *BMC Genomics.* 2007;8:256.

17. Rouquier S, Blancher A, Giorgi D. The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. *Proc Natl Acad Sci USA*. 2000;97(6):2870–4.
18. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 2006;444(7118):499–502.
19. Cobb J, Busst C, Petrou S, Harrap S, Ellis J. Searching for functional genetic variants in non-coding DNA. *Clin Exp Pharmacol Physiol*. 2008;35(4):372–5.
20. Kuznetsova IS, Prusov AN, Erukashvily NI, Podgornaya OI. New types of mouse centromeric satellite DNAs. *Chromosom Res*. 2005;13(1):9–25.
21. Bois PR. Hypermutable minisatellites, a human affair? *Genomics*. 2003;81(4):349–55.
22. Jeffreys AJ, Wilson V, Thein SL. Individual-specific ‘fingerprints’ of human DNA. *Nature*. 1985;316(6023):76–9.
23. Jeffreys AJ, Wilson V, Kelly R, Taylor BA, Bulfield G. Mouse DNA ‘fingerprints’: analysis of chromosome localization and germ-line stability of hyper-variable loci in recombinant inbred strains. *Nucleic Acids Res*. 1987;15(7):2823–36.
24. Kurtz TW, Montano M, Chan L, Kabra P. Molecular evidence of genetic heterogeneity in Wistar-Kyoto rats: implications for research with the spontaneously hypertensive rat. *Hypertension*. 1989;13(2):188–92.
25. Benavides F, Cazalla D, Pereira C, Fontanals A, Salaverri M, Goldman A, et al. Evidence of genetic heterogeneity in a BALB/c mouse colony as determined by DNA fingerprinting. *Lab Anim*. 1998;32(1):80–5.
26. Benavides F, Glasscock E, Coghlan LG, Stern MC, Weiss DA, Conti CJ. PCR-based microsatellite analysis for differentiation and genetic monitoring of nine inbred SENCAR mouse strains. *Lab Anim*. 2001;35(2):157–62.
27. Mashimo T, Voigt B, Tsurumi T, Naoi K, Nakanishi S, Yamasaki K, et al. A set of highly informative rat simple sequence length polymorphism (SSLP) markers and genetically defined rat strains. *BMC Genet*. 2006;7:19.
28. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315(5813):848–53.
29. Adams DJ, Dermitzakis ET, Cox T, Smith J, Davies R, Banerjee R, et al. Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nat Genet*. 2005;37(5):532–6.
30. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, et al. A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet*. 2007;3(1):e3.
31. She X, Cheng Z, Zollner S, Church DM, Eichler EE. Mouse segmental duplication and copy number variation. *Nat Genet*. 2008;40(7):909–14.
32. Watkins-Chow DE, Pavan WJ. Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res*. 2008;18(1):60–6.
33. Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet*. 2007;8:241–59.
34. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8(12):973–82.
35. Belancio VP, Hedges DJ, Deininger P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res*. 2008;18(3):343–58.
36. Wade CM, Kulbokas EJ 3rd, Kirby AW, Zody MC, Mullikin JC, Lander ES, et al. The mosaic structure of variation in the laboratory mouse genome. *Nature*. 2002;420(6915):574–8.
37. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet*. 2001;35:501–38.
38. Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, et al. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci USA*. 2008;105(11):4220–5.
39. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature*. 2005;436(7048):221–6.
40. Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, et al. piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci USA*. 2006;103(41):15008–13.
41. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet*. 2008;42:709–32.
42. Stoye JP, Fenner S, Greenoak GE, Moran C, Coffin JM. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell*. 1988;54(3):383–91.
43. Hughes AL, Welch R, Puri V, Matthews C, Haque K, Chanock SJ, et al. Genome-wide SNP typing reveals signatures of population history. *Genomics*. 2008;92(1):1–8.
44. Zhang J, Hunter KW, Gandolph M, Rowe WL, Finney RP, Kelley JM, et al. A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. *Genome Res*. 2005;15(2):241–9.
45. 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.

46. Bryda EC, Riley LK. Multiplex microsatellite marker panels for genetic monitoring of common rat strains. *J Am Assoc Lab Anim Sci*. 2008;47(3):37–41.
47. Nijman IJ, Kuipers S, Verheul M, Guryev V, Cuppen E. A genome-wide SNP panel for mapping and association studies in the rat. *BMC Genomics*. 2008;9:95.
48. Mudge JM, Harrow J. Creating reference gene annotation for the mouse C57BL/6J genome assembly. *Mamm Genome*. 2015;26(9–10):366–78.
49. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559–63.
50. Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol*. 2011;9(1):e1000582.
51. Eppig JT. Mouse Genome Informatics (MGI) resource: genetic, genomic, and biological knowledgebase for the laboratory mouse. *ILAR J*. 2017;58(1):17–41.
52. Shimoyama M, De Pons J, Hayman GT, Lauderkind SJ, Liu W, Nigam R, et al. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res*. 2015;43(Database issue):D743–50.
53. Lauderkind SJ, Hayman GT, Wang SJ, Smith JR, Petri V, Hoffman MJ, et al. A primer for the Rat Genome Database (RGD). *Methods Mol Biol*. 1757;2018:163–209.
54. Morse HC 3rd. *Origins of inbred mice*. New York: Academic Press; 1978.
55. Rader K. *Making mice: standardizing animals for American Biomedical Research, 1900–1955*. Princeton: Princeton University Press; 2004.
56. Moriwaki K, Shiroishi T, Yonekawa H. *Genetics in wild mice: its application to biomedical research*. Tokyo: Japan Scientific Societies Press; 1994.
57. Simecek P, Forejt J, Williams RW, Shiroishi T, Takada T, Lu L, et al. High-resolution maps of mouse reference populations. *G3 (Bethesda)*. 2017;7(10):3427–34.
58. Kuramoto T, Nakanishi S, Ochiai M, Nakagama H, Voigt B, Serikawa T. Origins of albino and hooded rats: implications from molecular genetic analysis across modern laboratory rat strains. *PLoS One*. 2012;7(8):e43059.
59. Peters H, Reifenberg K, Wedekind, D. Substrains of inbred strains. *GV-SOLAS*. 2013; Specialist Information.
60. Freeman HC, Hugill A, Dear NT, Ashcroft FM, Cox RD. Deletion of nicotinamide nucleotide transhydrogenase: a new quantitative trait locus accounting for glucose intolerance in C57BL/6J mice. *Diabetes*. 2006;55(7):2153–6.
61. Mattapallil MJ, Wawrousek EF, Chan CC, Zhao H, Roychoudhury J, Ferguson TA, et al. The Rd8 mutation of the *Crb1* gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes. *Invest Ophthalmol Vis Sci*. 2012;53(6):2921–7.
62. Kumar V, Kim K, Joseph C, Kourrich S, Yoo SH, Huang HC, et al. C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science*. 2013;342(6165):1508–12.
63. Simon MM, Greenaway S, White JK, Fuchs H, Gailus-Durner V, Wells S, et al. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol*. 2013;14(7):R82.
64. Simpson EM, Linder CC, Sargent EE, Davisson MT, Mobraaten LE, Sharp JJ. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat Genet*. 1997;16(1):19–27.
65. Hermsen R, de Ligt J, Spee W, Blokzijl F, Schafer S, Adami E, et al. Genomic landscape of rat strain and substrain variation. *BMC Genomics*. 2015;16:357.
66. Guenet JL, Benavides FJ. Mouse strains and genetic nomenclature. *Curr Protoc Mouse Biol*. 2011;1(1):213–38.
67. Yalcin B, Fullerton J, Miller S, Keays DA, Brady S, Bhomra A, et al. Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. *Proc Natl Acad Sci USA*. 2004;101(26):9734–9.
68. Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, et al. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet*. 2011;43(7):648–55.
69. Guenet JL, Bonhomme F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet*. 2003;19(1):24–31.
70. Schuster-Gossler K, Lee AW, Lerner CP, Parker HJ, Dyer VW, Scott VE, et al. Use of coisogenic host blastocysts for efficient establishment of germline chimeras with C57BL/6J ES cell lines. *BioTechniques*. 2001;31(5):1022–4. 6
71. Nadeau JH, Singer JB, Matin A, Lander ES. Analysing complex genetic traits with chromosome substitution strains. *Nat Genet*. 2000;24(3):221–5.
72. Bailey DW. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. *Transplantation*. 1971;11(3):325–7.
73. Shisa H, Lu L, Katoh H, Kawarai A, Tanuma J, Matsushima Y, et al. The LEXF: a new set of rat recombinant inbred strains between LE/Stm and F344. *Mamm Genome*. 1997;8(5):324–7.
74. Demant P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat Rev Genet*. 2003;4(9):721–34.
75. Burgio G, Szatanik M, Guenet JL, Arnau MR, Panthier JJ, Montagutelli X. Interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species: a powerful tool to dissect genetic control of complex traits. *Genetics*. 2007;177(4):2321–33.

76. Burgio G, Baylac M, Heyer E, Montgutelli X. Genetic analysis of skull shape variation and morphological integration in the mouse using interspecific recombinant congenic strains between C57BL/6 and mice of the *mus spretus* species. *Evolution*. 2009;63(10):2668–86.
77. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*. 2004;36(11):1133–7.
78. Chesler EJ, Miller DR, Branstetter LR, Galloway LD, Jackson BL, Philip VM, et al. The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm Genome*. 2008;19(6):382–9.
79. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, et al. Genetic analysis of complex traits in the emerging collaborative cross. *Genome Res*. 2011;21:1213–22.
80. Srivastava A, Morgan AP, Najarian ML, Sarsani VK, Sigmon JS, Shorter JR, et al. Genomes of the mouse Collaborative Cross. *Genetics*. 2017;206(2):537–56.
81. Hartl DL. Genetic management of outbred laboratory rodent populations. *Charles River Genetic Literature*. 2001.
82. Poiley SM. A systematic method of breeder rotation for non-inbred laboratory animals colonies. *Proc Anim Care Panel*. 1960;10:159.
83. Schmitt AO, Bortfeldt R, Neuschl C, Brockmann GA. RandoMate: a program for the generation of random mating schemes for small laboratory animals. *Mamm Genome*. 2009;20(5):321–5.
84. Chia R, Achilli F, Festing MF, Fisher EM. The origins and uses of mouse outbred stocks. *Nat Genet*. 2005;37(11):1181–6.
85. Festing MF. Inbred strains should replace outbred stocks in toxicology, safety testing, and drug development. *Toxicol Pathol*. 2010;38(5):681–90.
86. Yalcin B, Nicod J, Bhomra A, Davidson S, Cleak J, Farinelli L, et al. Commercially available outbred mice for genome-wide association studies. *PLoS Genet*. 2010;6(9):e1001085.
87. Churchill GA, Gatti DM, Munger SC, Svenson KL. The diversity outbred mouse population. *Mamm Genome*. 2012;23(9–10):713–8.
88. Tuttle AH, Philip VM, Chesler EJ, Mogil JS. Comparing phenotypic variation between inbred and outbred mice. *Nat Methods*. 2018;15(12):994–6.
89. Jackson IJ, Abbott CM. *Mouse genetics and transgenics: a practical approach*. Oxford: Oxford University Press; 2000.
90. Nagy A, Gertsenstein M, Vintersten K, Behringer R. *Manipulating the mouse embryo, a laboratory manual*. 3rd ed. New York: Cold Spring Harbor Press; 2003.
91. Koentgen F, Suess G, Naf D. Engineering the mouse genome to model human disease for drug discovery. *Methods Mol Biol*. 2010;602:55–77.
92. Guenet JL. Animal models of human genetic diseases: do they need to be faithful to be useful? *Mol Gen Genomics*. 2011;286(1):1–20.
93. Perez CJ, Jaubert J, Guenet JL, Barnhart KF, Ross-Inta CM, Quintanilla VC, et al. Two hypomorphic alleles of mouse *Ass1* as a new animal model of citrullinemia type I and other hyperammonemic syndromes. *Am J Pathol*. 2010;177(4):1958–68.
94. Bao J, Perez CJ, Kim J, Zhang H, Murphy CJ, Hamidi T, et al. Deficient LRRC8A-dependent volume-regulated anion channel activity is associated with male infertility in mice. *JCI Insight*. 2018;3(16):e99767.
95. 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 USA*. 1979;76(11):5818–9.
96. Guenet JL. Chemical mutagenesis of the mouse genome: an overview. *Genetica*. 2004;122(1):9–24.
97. Gondo Y. Trends in large-scale mouse mutagenesis: from genetics to functional genomics. *Nat Rev Genet*. 2008;9(10):803–10.
98. Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, et al. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet*. 2000;25(4):440–3.
99. Hrabe 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.
100. Wang T, Bu CH, Hildebrand S, Jia G, Siggs OM, Lyon S, et al. Probability of phenotypically detectable protein damage by ENU-induced mutations in the Mutagenetix database. *Nat Commun*. 2018;9(1):441.
101. Moran JL, Bolton AD, Tran PV, Brown A, Dwyer ND, Manning DK, et al. Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res*. 2006;16(3):436–40.
102. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature*. 1994;372(6505):425–32.
103. 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.
104. Costantini F, Lacy E. Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature*. 1981;294(5836):92–4.
105. Gordon JW, Ruddle FH. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science*. 1981;214(4526):1244–6.
106. Bonaparte D, Cinelli P, Douni E, Herault Y, Maas M, Pakarinen P, et al. FELASA guidelines for the

- refinement of methods for genotyping genetically-modified rodents: a report of the Federation of European Laboratory Animal Science Associations Working Group. *Lab Anim.* 2013;47(3):134–45.
107. Ballester M, Castello A, Ibanez E, Sanchez A, Folch JM. Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals. *BioTechniques.* 2004;37(4):610–3.
 108. Taketo M, Schroeder AC, Mobraaten LE, Gunning KB, Hanten G, Fox RR, et al. FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci USA.* 1991;88(6):2065–9.
 109. Furth PA, St Onge L, Boger H, Gruss P, Gossen M, Kistner A, et al. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proc Natl Acad Sci USA.* 1994;91(20):9302–6.
 110. Li P, Tong C, Mehrian-Shai R, Jia L, Wu N, Yan Y, et al. Germline competent embryonic stem cells derived from rat blastocysts. *Cell.* 2008;135(7):1299–310.
 111. McLellan MA, Rosenthal NA, Pinto AR. Cre-loxP-mediated recombination: general principles and experimental considerations. *Curr Protoc Mouse Biol.* 2017;7(1):1–12.
 112. Feil S, Valtcheva N, Feil R. Inducible Cre mice. *Methods Mol Biol.* 2009;530:343–63.
 113. West DB, Pasumarthi RK, Baridon B, Djan E, Trainor A, Griffey SM, et al. A lacZ reporter gene expression atlas for 313 adult KOMP mutant mouse lines. *Genome Res.* 2015;25(4):598–607.
 114. Kaneko T, Mashimo T. Creating knockout and knockin rodents using engineered endonucleases via direct embryo injection. *Methods Mol Biol.* 2015;1239:307–15.
 115. Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science.* 2009;325(5939):433.
 116. Mashimo T. Gene targeting technologies in rats: zinc finger nucleases, transcription activator-like effector nucleases, and clustered regularly interspaced short palindromic repeats. *Develop Growth Differ.* 2014;56(1):46–52.
 117. Sung YH, Baek IJ, Kim DH, Jeon J, Lee J, Lee K, et al. Knockout mice created by TALEN-mediated gene targeting. *Nat Biotechnol.* 2013;31(1):23–4.
 118. Tesson L, Remy S, Menoret S, Usal C, Thinnard R, Savignard C, et al. Genome editing in rats using TALE nucleases. *Methods Mol Biol.* 2016;1338:245–59.
 119. Pennisi E. The CRISPR craze. *Science.* 2013;341(6148):833–6.
 120. Fernandez A, Josa S, Montoliu L. A history of genome editing in mammals. *Mamm Genome.* 2017;28(7–8):237–46.
 121. Horii T, Arai Y, Yamazaki M, Morita S, Kimura M, Itoh M, et al. Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Sci Rep.* 2014;4:4513.
 122. Guan Y, Shao Y, Li D, Liu M. Generation of site-specific mutations in the rat genome via CRISPR/Cas9. *Methods Enzymol.* 2014;546:297–317.
 123. Shao Y, Guan Y, Wang L, Qiu Z, Liu M, Chen Y, et al. CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. *Nat Protoc.* 2014;9(10):2493–512.
 124. 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.
 125. Yoshimi K, Kaneko T, Voigt B, Mashimo T. Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform. *Nat Commun.* 2014;5:4240.
 126. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* 2014;157(6):1262–78.
 127. Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet.* 2014;23(R1):R40–6.
 128. Seruggia D, Fernandez A, Cantero M, Pelczar P, Montoliu L. Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis. *Nucleic Acids Res.* 2015;43(10):4855–67.
 129. Gu B, Posfai E, Rossant J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat Biotechnol.* 2018;36(7):632–7.
 130. Chen S, Lee B, Lee AY, Modzelewski AJ, He L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J Biol Chem.* 2016;291(28):14457–67.
 131. Kobayashi T, Namba M, Koyano T, Fukushima M, Sato M, Ohtsuka M, et al. Successful production of genome-edited rats by the rGONAD method. *BMC Biotechnol.* 2018;18(1):19.
 132. Fairfield H, Srivastava A, Ananda G, Liu R, Kircher M, Lakshminarayana A, et al. Exome sequencing reveals pathogenic mutations in 91 strains of mice with Mendelian disorders. *Genome Res.* 2015;25(7):948–57.
 133. Otsen M, Den Bieman M, Winer ES, Jacob HJ, Szpirer J, Szpirer C, et al. Use of simple sequence length polymorphisms for genetic characterization of rat inbred strains. *Mamm Genome.* 1995;6(9):595–601.
 134. Gurumurthy CB, Joshi PS, Kurz SG, Ohtsuka M, Quadros RM, Harms DW, et al. Validation of simple sequence length polymorphism regions of commonly used mouse strains for marker assisted speed congenics screening. *Int J Genomics.* 2015;2015:735845.
 135. Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, et al. An efficient SNP system

- for mouse genome scanning and elucidating strain relationships. *Genome Res.* 2004;14(9):1806–11.
136. Myakishev MV, Khripin Y, Hu S, Hamer DH. High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Res.* 2001;11(1):163–9.
137. Zurita E, Chagoyen M, Cantero M, Alonso R, Gonzalez-Neira A, Lopez-Jimenez A, et al. Genetic polymorphisms among C57BL/6 mouse inbred strains. *Transgenic Res.* 2011;20(3):481–9.
138. Mekada K, Abe K, Murakami A, Nakamura S, Nakata H, Moriwaki K, et al. Genetic differences among C57BL/6 substrains. *Exp Anim.* 2009;58(2):141–9.
139. Mekada K, Hirose M, Murakami A, Yoshiki A. Development of SNP markers for C57BL/6N-derived mouse inbred strains. *Exp Anim.* 2015;64(1):91–100.
140. Zimdahl H, Nyakatura G, Brandt P, Schulz H, Hummel O, Fartmann B, et al. A SNP map of the rat genome generated from cDNA sequences. *Science.* 2004;303(5659):807.
141. Smits BM, Guryev V, Zeegers D, Wedekind D, Hedrich HJ, Cuppen E. Efficient single nucleotide polymorphism discovery in laboratory rat strains using wild rat-derived SNP candidates. *BMC Genomics.* 2005;6:170.
142. Consortium S, Saar K, Beck A, Bihoreau MT, Birney E, Brocklebank D, et al. SNP and haplotype mapping for genetic analysis in the rat. *Nat Genet.* 2008;40(5):560–6.
143. Beckstead WA, Bjork BC, Stottmann RW, Sunyaev S, Beier DR. SNP2RFLP: a computational tool to facilitate genetic mapping using benchtop analysis of SNPs. *Mamm Genome.* 2008;19(10–12):687–90.
144. Wedekind D, Reifenberg K, Hedrich HJ. Genetic monitoring of inbred strains. In: Hedrich HJ, editor. *The laboratory mouse.* Boston: Elsevier; 2012. p. 621–37.
145. Guenet JL, Benavides F. Genetic monitoring of laboratory rodents. In: Patrinos GP, Ansoorge W, editors. *Molecular diagnostics.* 2nd ed. Oxford: Oxford Academic Press; 2010.
146. Reifenberg K, Hedrich H, Wedekind D, Howells N. Objective and methods of genetic monitoring of isogenic mouse and rat strains. *GV-SOLAS Specialist Information.* 2014.
147. Taft RA, Davisson M, Wiles MV. Know thy mouse. *Trends Genet.* 2006;22(12):649–53.
148. Reardon S. Lab mice's ancestral 'Eve' gets her genome sequenced. *Nature.* 2017;551(7680):281.
149. Fahey JR, Katoh H, Malcolm R, Perez AV. The case for genetic monitoring of mice and rats used in biomedical research. *Mamm Genome.* 2013;24(3–4):89–94.
150. Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse HC 3rd, de Villena FP. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. *BMC Genomics.* 2014;15:847.
151. Linder CC. The influence of genetic background on spontaneous and genetically engineered mouse models of complex diseases. *Lab Anim (NY).* 2001;30(5):34–9.
152. Doetschman T. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol.* 2009;530:423–33.
153. Hummel KP, Coleman DL, Lane PW. The influence of genetic background on expression of mutations at the diabetes locus in the mouse. I. C57BL-KsJ and C57BL-6J strains. *Biochem Genet.* 1972;7(1):1–13.
154. Dietrich WF, Lander ES, Smith JS, Moser AR, Gould KA, Luongo C, et al. Genetic identification of Mom-1, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. *Cell.* 1993;75(4):631–9.
155. Threadgill DW, Dlugosz AA, Hansen LA, Tennenbaum T, Lichti U, Yee D, et al. Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science.* 1995;269(5221):230–4.
156. Kuperwasser C, Hurlbut GD, Kittrell FS, Dickinson ES, Laucirica R, Medina D, et al. Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice. A model for Li-Fraumeni syndrome. *Am J Pathol.* 2000;157(6):2151–9.
157. Freeman D, Lesche R, Kertesz N, Wang S, Li G, Gao J, et al. Genetic background controls tumor development in PTEN-deficient mice. *Cancer Res.* 2006;66(13):6492–6.
158. Calyjur PC, Almeida Cde F, Ayub-Guerrieri D, Ribeiro AF Jr, Fernandes Sde A, Ishiba R, et al. The mdx mutation in the 129/Sv background results in a milder phenotype: transcriptome comparative analysis searching for the protective factors. *PLoS One.* 2016;11(3):e0150748.
159. Asamoto M, Hokaiwado N, Cho YM, Shirai T. Effects of genetic background on prostate and taste bud carcinogenesis due to SV40 T antigen expression under probasin gene promoter control. *Carcinogenesis.* 2002;23(3):463–7.
160. Dang R, Torigoe D, Suzuki S, Kikkawa Y, Moritoh K, Sasaki N, et al. Genetic background strongly modifies the severity of symptoms of Hirschsprung disease, but not hearing loss in rats carrying Ednrn(sl) mutations. *PLoS One.* 2011;6(9):e24086.
161. Vanden Berghe T, Hulpiau P, Martens L, Vandenbroucke RE, Van Wonterghem E, Perry SW, et al. Passenger mutations confound interpretation of all genetically modified congenic mice. *Immunity.* 2015;43(1):200–9.
162. Stevens JC, Banks GT, Festing MF, Fisher EM. Quiet mutations in inbred strains of mice. *Trends Mol Med.* 2007;13(12):512–9.
163. Specht CG, Schoepfer R. Deletion of the alpha-synuclein locus in a subpopulation of C57BL/6J inbred mice. *BMC Neurosci.* 2001;2:11.

164. Bourdi M, Davies JS, Pohl LR. Mispairing C57BL/6 substrains of genetically engineered mice and wild-type controls can lead to confounding results as it did in studies of JNK2 in acetaminophen and concanavalin A liver injury. *Chem Res Toxicol*. 2011;24(6):794–6.
165. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science*. 1998;282(5396):2085–8.
166. Wakeland E, Morel L, Achey K, Yui M, Longmate J. Speed congenics: a classic technique in the fast lane (relatively speaking). *Immunol Today*. 1997;18(10):472–7.
167. Markel P, Shu P, Ebeling C, Carlson GA, Nagle DL, Smutko JS, et al. Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet*. 1997;17(3):280–4.
168. Wolfer DP, Crusio WE, Lipp HP. Knockout mice: simple solutions to the problems of genetic background and flanking genes. *Trends Neurosci*. 2002;25(7):336–40.
169. Chen S, Kadomatsu K, Kondo M, Toyama Y, Toshimori K, Ueno S, et al. Effects of flanking genes on the phenotypes of mice deficient in basigin/CD147. *Biochem Biophys Res Commun*. 2004;324(1):147–53.
170. Brommage R, Liu J, Hansen GM, Kirkpatrick LL, Potter DG, Sands AT, et al. High-throughput screening of mouse gene knockouts identifies established and novel skeletal phenotypes. *Bone Res*. 2014;2:14034.
171. Brown SDM, Holmes CC, Mallon AM, Meehan TF, Smedley D, Wells S. High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet*. 2018;19(6):357–70.
172. Mallon AM, Blake A, Hancock JM. EuroPhenome and EMPReSS: online mouse phenotyping resource. *Nucleic Acids Res*. 2008;36(Database issue):D715–8.
173. 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. *Nat Genet*. 2015;47(9):969–78.
174. Paigen K, Eppig JT. A mouse phenome project. *Mamm Genome*. 2000;11(9):715–7.
175. Brown SD, Moore MW. The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. *Mamm Genome*. 2012;23(9–10):632–40.
176. Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res*. 2014;42(Database issue):D802–9.
177. Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling MC, et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm Genome*. 2012;23(9–10):580–6.
178. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*. 2011;474(7351):337–42.
179. Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, Warren J, et al. Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat Genet*. 2017;49(8):1231–8.