UNIVERSITY OF CALIFORNIA SAN DIEGO

Engineering the Jerbouse: Genetic Strategies for Modeling Evolutionarily Divergent Phenotypes in Mouse

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Hannah Ariel Grunwald

Committee in charge:

Professor Kimberly Cooper, Chair
Professor Vineet Bafna
Professor Ethan Bier
Professor Prashant Mali
Professor Deborah Yelon

2020

The dissertation of Hannah Ariel Grunwald is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California San Diego

2020

It is a gift always, to be loved.
This dissertation is then dedicated, with my most profound gratitude and devotion to Robin Carpenter, David Grunwald, Ellen Wilson, Sonia Grunwald, Kaisa Hanley, Eileen Moreno, Oryx, Diego, and Pepper.

“from so simple a beginning
endless forms most beautiful
and most wonderful
have been, and are being, evolved.”
-Charles Darwin

“This time it had been magic.
And it didn't stop being magic just because
you found out how it was done.”
-Terry Pratchett

Table of Contents

List of Figures

List of Tables

Acknowledgements

With my deepest gratitude, I acknowledge the incredible mentorship, scientific advice, career guidance, and general brilliance of my mentor and committee chair, Dr. Kimberly Cooper. Dr. Cooper's dedication to me, as a scientist, activist, and individual, has never faltered. Without her incredible strength of mentorship and friendship, this work could not have been completed. I am honored to have had the opportunity to work with Dr. Cooper.

I would like to further acknowledge the incredible work of all the Cooper lab members, past and present, without whose support and advice (technical, personal, and otherwise), this dissertation would not be possible. In particular, I would like to recognize and thank the superb undergraduate students I have had the honor of working with, Fayth Tan, Angela Green, Andrew Chen, and Alfredo Escobedo.

I am deeply indebted to the various scientists who have contributed their scientific expertise and advice over the course of my time at UCSD, including, but not limited to Dr. Chris Organ, Dr. Heidi Cook-Andersen, Dr. Stephen Hedrick, Dr. Richard Behringer, Dr. Pieter de Jong, Dr. Cris Jung, and Kaisa Hanley. I also appreciate and recognize the advice of Anusorn Mudla and Brian Tenner with regards to Gibson cloning. I am deeply appreciative of the help regarding Nanopore sequencing provided by Cong Dinh and I thank everyone at the University of Michigan Transgenic Animal Core, most especially Thom Saunders, Elizabeth Hughes, Corey Ziebell, and Anna LaForest for their work producing the jerbouse. I thank Sonia Grunwald for her exceptional advice and aid in regards to computational analysis, David Grunwald for his sage advice, and in particular

for showing me the Metz paper on chromosome pairing in *Diptera*, and Robin Carpenter for his help with the production of Figure 1.5 and formatting of this thesis.

I feel deeply honored to have had the advice and mentorship of my incredible committee members, Dr. Deborah Yelon, Dr. Ethan Bier, Dr. Prashant Mali, and Dr. Vineet Bafna, who helped shape the content, story, and trajectory of this work.

I am fundamentally grateful to have had the opportunity to participate in workshops and symposia in association with the Genetic Biocontrol of Invasive Rodents project and the TATA Institute for Genetics and Society and in particular want to thank Royden Saah for his generosity of spirit, general enthusiasm, and wealth of knowledge and connections and Sarah Werner for her infinite capacity for patience and organization. I am also grateful for the advice and support I've received from members of the Active Genetics community at UCSD including (but not limited to) Ethan Bier, Omar Akbari, and John Evans.

I am permanently indebted to the staff who helped me navigate the bureaucracy of academia and extend my findings into the real world—their outstanding work and selfless attitude made this work possible. I therefore would like to thank Melody Bazyar, Suzi Harlow, Natalie Noles, Ria Del Rosario, Mario Aguilera, Victoria Berdin, Amielyn Diza, Alex Tea, Donna Aizuss, Lien Ngo, and Jessica Whittier.

Finally, it is with great joy and appreciation that I recognize the contribution of everyone who has contributed to my growth as a human. Without these individuals, this work could not have been completed, but perhaps more importantly, it would not have held meaning. Without a doubt, I will miss many who deserve thanks, but I would like to acknowledge first my brilliant and supportive family, Ellen Wilson, David Grunwald,

Sonia Grunwald, Jack Litle, Julia Heberle, John Carpenter, Eleanor Carpenter, and Jesse Carpenter; my closest small loved ones, who never miss an opportunity to make me feel loved, Oryx, Diego, and Pepper; my incredible friends and heroes Kaisa Hanley, Alberto Carreño, Charles Carreño-Hanley, Eileen Moreno, Jesus Moreno, Andy Ryan, Anna Guzikowski, Laura Chipman, Gina Powers, Jack Olmstead, Elizabeth Bullard, Alyson Abel, Megan Peet, Jamie Ivey, Jill Castellano, Anton Khodakov, Kurt Naro, and Eric Jones.

In particular, I want to thank and recognize the outstanding contributions of Robin Carpenter, without whom, I would, simply put, be less human. His strength of character, easy dedication and persistence in all his endeavors, and gentle kindness lift up everyone around him. I'm honored to have had a partner who cooked me matzah ball soup and homemade croissants, took me camping, and put up with all the terrible TV I insisted on watching while writing this dissertation, who drove me to work when it rained, helped me with mini-preps, and patiently focused (and eventually bought me colored pencils) while I practiced explanations of intensely technical genetic concepts, using up reams of paper and hours of our lives. I have never met anyone as patient, generous, bright, or grounded as Robin Carpenter. This thesis exists in large part because of his efforts.

Chapter 1, in part, is being prepared for submission for publication of the material. Grunwald, Hannah; Chen, Andrew; Organ, Chris; Cooper, Kimberly. The dissertation author was the primary researcher and author of this material.

Chapter 2, in part, is a reprint of the material as it appears in Nature 2019. Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E. & Cooper, K. L. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566,** 105–109 (2019). The dissertation author was the primary researcher and author of this paper.

Chapter 3, in part, has been submitted for publication of the material as it may appear in Nature Protocols, 2020. Grunwald, Hannah; Weitzel, Alexander; Cooper, Kimberly, 2020. The dissertation author was the primary author of this paper.

Vita

2014 Bachelor of Arts, Swarthmore College

2016-2020 Teaching Assistant, University of California San Diego

2014-2020 Research Assistant, University of California San Diego

2020 Doctor of Philosophy, University of California San Diego

Publications

Grunwald, H. A. *et al.* Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566**, 105–109 (2019).

Fields of Study

Major Field: Biology

Genetics
Genetic Engineering
Development
Evolution

Research supervised by Professor Kimberly Cooper

Abstract of the Dissertation

Engineering the Jerbouse: Genetic Strategies for Modeling Evolutionarily Divergent Phenotypes in Mouse

by

Hannah Ariel Grunwald

Doctor of Philosophy in Biology

University of California San Diego, 2020

Professor Kimberly Cooper, Chair

Over the course of millions of years, modifications in mammalian genomes have produced an incredible variety of phenotypes, from the antlers of deer, to the breathing capacity of whales, to the dexterous fingers of a human. This remarkable diversity is perhaps nowhere more evident than in the limb of the mammal, which, despite its underlying structural similarities, exhibits specialized morphologies that look strikingly different in different species. Thus the hands typing this abstract are structurally similar but morphologically and functionally divergent from the fins of a manatee, the wings of a

bat, or the legs of a horse. Comparing the genetic and regulatory underpinnings of this divergence (and indeed, of the similarities between species) allows us to simultaneously reconstruct genomic evolution and delve into the developmental mechanisms that produce limbs. The lesser Egyptian jerboa, *Jaculus jaculus*, is genomically similar to the common laboratory mouse, *Mus musculus*, but has highly divergent derived characteristics in the hindlimb. The hindlimb of the jerboa features a striking elongation of metatarsals and tibia, the loss of two digits, and the fusing of metatarsals into one central bone. By engineering 'jerboanized' mouse models, where regions of the mouse genome are replaced with their jerboa homologues to produce animals with jerboa-esque phenotypes, we can elucidate the mechanisms of divergence and development in the jerboa in a well-established and highly manipulable model system. Here, I present work on two genetic engineering techniques that promise to transform interspecies comparisons in rodent systems. First, I introduce a large interspecies conversion that simultaneously deletes a ~26kb region of the mouse genome and replaces it with ~31kb of homologous sequence from the jerboa. Second, I present the first demonstration of active genetics, a technique that makes use of the CRISPR Cas9 system to convert genotypes, in a rodent model. Active genetics in the rodent may be used to increase the rate of inheritance of an allele, which has profound implications the future of genetic engineering in both laboratory and applied settings. Together these tools may be used to develop complex mouse models of jerboanized phenotypes to probe the specific genomic changes responsible for the evolution of the development of the jerboa morphology.

Introduction

Every living organism starts as a single cell—a single, self-contained creature with all the machinery it needs to propagate its genetic material. Multicellular organisms clone that single cell over and over and over to form a mass of genetically identical cells, each of which will eventually perform a specific function in a complex body system made up of different tissue types. Think, for a moment, of the way the bones in an arm interact with the corresponding muscles, tendons, and ligaments that allow someone to type a word or grasp a fork. Think about the complicated system of vasculature that brings oxygen to each of those muscles by way of trachea, lungs, and heart. Think about the different neurons that snake their way from the spinal cord down the arm so that an organism can detect heat and subsequently jerk its fingers away. Every piece of that system—every piece of *you*—has precisely the same DNA, perfectly identical instructions on how to behave and grow.

How, then, do our bodies build complex interwoven structures, when each player in the system has precisely the same genetic material? DNA is a global set of instructions, a manual describing the construction of every possible protein and RNA the body will ever need. Cells decide which sections to read and which pieces to build based on input from outside sources. They react to signals from their neighbors[1], hormones and chemicals they encounter[2], and even gravity[3,4], to ensure that specific functions are performed at the right time and in the right place: an eye here, a kneecap there, a neuron, an immune cell, a muscle fiber.

Intriguingly, some structures that are defined using essentially the same set of genes manage to look quite different from one another. Think of your arms and legs.

Despite their underlying structural similarities, the human hindlimb is optimized for standing and running, the forelimb for precise manipulation. The differences are both readily apparent and obviously functional. As suggested by the gross morphological similarities between fore- and hindlimb, the genes essential for their patterning are largely overlapping. This suggests that when it comes to some critical distinctions between structures it is not sufficient to think of gene expression as varying at the tissue or organ level; it matters not only *which* genes are expressed in a tissue but how many cells they are expressed in, and at what levels, and for how long.

A superb example of this phenomenon is the variation in digit number in the Australian skink. Different species of skink within the genus *Hemiergis* have different numbers of digits, and some species differ in the number of digits between fore- and hindlimbs within a single animal. Regardless of the specific species or number of digits on a limb, each skink expresses the gene *Sonic hedgehog* (*Shh*) in the posterior part of the most distal section of each limb. *Shh* has been well described in the limb as a signal that patterns the anterior-posterior axis of the autopod and promotes cell survival. Surprisingly, however, the duration of *Shh* expression varies between *Hemiergis* limbs and species. Researchers observed a striking correlation between the number of digits on a limb and the duration of *Shh* expression. When *Shh* expression was diminished earlier than expected, limbs produced fewer digits. This difference was apparent even within single animals that had different numbers of digits on fore- and hindlimbs. In this case, the fine details of expression, specifically the duration of expression, were responsible for the observed variation in phenotype between similar structures within a single organism[5].

The *Hemiergis* example also brings to light a second intriguing window into how variation in gene expression patterns can contribute to functional diversity of structures. Much of the information for patterning a body has been highly conserved across the animal kingdom. The basic tetrapod body plan, for example, is so well conserved that it is easy to identify any given homologous bone in a frog, a giraffe, and a hawk, animals that likely shared their last common ancestor more than 350 million years ago (mya)[6].

Indeed, this conservation extends to the cellular, molecular, and genetic levels. Many genes are highly conserved between taxa and can be identified by sequence homology alone. Perhaps more compelling, many genes have stereotypical patterns of expression and similar or identical function in multiple animals. Returning to the example of the limb, *Shh*, *Fgfs*, and *Hox* genes have all been shown to be essential for limb patterning in chick[7,8], axolotl[9], and mouse[10]. The evidence for a tight conservation of gene products extends even beyond sequence and function similarity. Many proteins from one species are functional in others. *Sox9* homologues, for example, from cuttlefish and horseshoe crab were able to successfully recognize downstream chondrogenic gene targets in a reporter assay in rat cells[11]. Together, this suggests that many protein-coding regions are well conserved and functionally similar across animal species that are phenotypically distinct.

In some ways, this phenomenon of close conservation is quite surprising, given the depth of phenotypic variation we see in animals. Evolution has spent hundreds of millions of years shaping organisms into dramatic and chaotic 'endless forms most beautiful'. Using the same set of DNA in the ancestor of vertebrates, changes here and there have resulted in the snake's tongue, the hummingbird's wings, the camel's hump.

Even within a closely related clade of animals—mammals, for example—the variation is readily apparent. This combination of divergent morphology and relatively conserved gene sequences and expression patterns suggests, once again, that changes in the *regulation* of genes, rather than gene products themselves, may be in large part responsible for the diversity of multicellular forms.

As technologies to manipulate and study non-model organisms improve, researchers are increasingly taking advantage of the diversity seen in nature. By comparing two animals with dramatically distinct phenotypes, we are offered insight into the genetic control of development to a depth not possible using any single organism. Comparing a goat to a giraffe, for example, may lend us insight not only into the common mechanisms of neck specification and growth, which is likely to be reliant on genes and expression patterns common between the two animals, but also the genetic possibilities for manipulating gene expression to produce different neck morphologies. Comparing species allows us to simultaneously ask questions about development and about evolution in a context where dramatic phenotypes are plentiful.

An animal model of the development and evolution of the vertebrate limb

When it comes to choosing models of variation, limbs in mammals showcase some of the most obvious and extreme diversity apparent in the phyla. Whereas human hands are capable of performing careful and precise manipulations, horse forelimbs are simple single-toed levers used for running, bat forelimbs have elongated webbed digits

for powered flight, and dolphins and whales have evolved paddle-like forelimbs for propulsion underwater.

Decades of work illuminating the generalized mechanism of limb outgrowth and patterning have provided developmental biologists with a detailed model of the general development of a limb (reviewed here[12,13] and elsewhere). In each limb bud, early expression of *Wnt3a* in the ectoderm[14] and *Fgf10*[15] in the mesenchyme promote the development of the apical ectodermal ridge (AER), a thickened epithelium overlying the limb mesenchyme at the dorsal-ventral boundary. *Fgf10* induces strong *Fgf8* expression in the AER[15], which promotes cell proliferation and limb outgrowth[16,17]. Signals from the AER, including *Fgf8,* work in opposition to signals from the flank, which likely include *RA*, to define the proximodistal axis[18,19]. Proximo-distal limb segments are further characterized by differential *Hox* expression, with the most proximal region, the stylopod, expressing *Hox9/10* and *Meis1/2*, the intermediate region, the zeugopod, expressing *Hox11*, and the most distal region, the autopod, expressing *Hox13*[20].

Establishment of an anterior-posterior axis is facilitated by localized expression of *Hand2* and *Shh* in the zone of polarizing activity (ZPA), located in the posterior region of the autopod. This localized posterior expression establishes a posterior to anterior gradient of *Hedgehog* signaling required for cell survival, digit formation, and the establishment of the anterior-posterior axis[21,22]. *Wnt7a*[23] and *Bmp*s[24,25] act in opposition to define the dorsal-ventral axis.

The autopod is further patterned by a Turing reaction-diffusion network wherein *Bmps, Wnts,* and the chondrogenic signal *Sox9* define the locations of the chondrogenic condensations that will become the digits[26]. In many animals, the interdigital space

between these condensations is carved away by apoptotic cell death controlled by *Bmp4* and *Msx2*[27,28].

Despite the wealth of knowledge about the signals required to pattern and maintain limb outgrowth, the specifics of how the limb has naturally diversified remain unknown. For all our work and detailed understanding of how limbs are formed, we have yet to understand how signals are modulated to make human tibias longer than metacarpals or human fingers more dexterous than toes. It may be noted that the gene networks involved in limb growth are largely overlapping between fore- and hindlimbs, suggesting that variation between these structures is indeed primarily regulatory in origin. As previously noted, these networks are also largely conserved across vertebrates, raising questions of how they can be manipulated to produce the apparent species diversity.

One way to study regulatory changes with phenotypic consequences in the limb is to identify a model organism that provides a platform to investigate limb variation both within and between species. Such a model system would need to have substantive differences between hind- and forelimb, to facilitate studies on intra-animal variation, as well as marked divergence from existing animal models, to allow for the study of inter-animal variation. In addition, to maximize its utility, it would need to be small and easily raised in a laboratory setting. Finally, it would benefit from having a well-studied relative against which comparisons could be performed.

The lesser Egyptian jerboa, *Jaculus jaculus*, satisfies all of these criteria, and as such, provides an unparalleled model for the study of mammalian limbs. *J. jaculus* is a member of the rodent family *Dipodidae*, whose 33 extant members, all of which are obligate bipedal rodents, have a variety of limb-specific traits that are thought to be

adaptive for cursorial and saltatorial locomotion (running and jumping) in desert landscapes[29,30]. Together the changes in the jerboa leg have allowed it to survive the harsh conditions of an open expanse, where cover and food are both scarce[31]. Importantly, the clade displays dramatic inter- and intra-species limb diversity, with variations in morphotypes between species as well as obvious differences between fore- and hindlimbs[30].

The jerboa hindlimb adaptations have long been described as similar to those of more easily recognizable cursorial animals, including the horse and deer[32]. The convergence of these limb traits, including elongated distal limb elements and the reduction of distal weight through loss of excess digits and muscle makes the jerboa particularly appealing as a model of convergent cursorial adaptation as well as general rodent limb development.

Access to jerboa specimens for research is relatively easy. Embryos can be gathered from jerboas in the deserts of China and Mongolia, where both five- (*Allactaga elater*) and three-toed (*Dipus sagitta*) jerboas are prevalent. Though these specimens are ideal for embryological studies, these animals hibernate yearly, making it difficult to maintain lab colonies. Unlike *D. sagitta* and *A. elater*, another three-toed species jerboa species, *Jaculus jaculus* (Figure 0.1), native to North Africa and the Middle East, does not hibernate and can be reared in captivity. The establishment of a colony of *J. jaculus*[29,33] at Harvard in 2010, now housed in our laboratory at UC San Diego, and the subsequent sequencing of the *J. jaculus* genome facilitate laboratory study of these animals.

As a member of the most derived clade of *Dipodidae*, *J. jaculus* also has one of the most extreme limb morphologies documented within rodents, with elongated hindlimbs, exaggerated proportion within the hindlimb, loss of intrinsic muscle, only three digits on the hindfeet, and fused metatarsals. While the hindlimb features myriad derived traits, the forelimb of the jerboa is more 'mouse-like', with five toes, retained muscle, and diminutive metacarpals[29,30]. The substantial differences between hind- and forelimbs allow for complex comparisons between modified limb structures within a single animal where the primary source of variation is likely expression differences rather than alterations to the coding sequences of genes used in both pairs of limbs.

In addition to the ability to investigate intra-animal differences, jerboas are well positioned for inter-species comparisons as near relatives of the laboratory mouse, *Mus musculus*. *Dipodidae* are close relatives of *Muroidea*, a superfamily including gerbils, hamsters, rats, and mice, which diverged from the *Dipodoids* only 55 mya[6]. Thus jerboas are an accessible source of natural variation and have highly derived hindlimbs that can be compared both to their more evolutionarily conserved forelimbs, and also to the fore- and hindlimbs of their close relative, the common lab mouse.

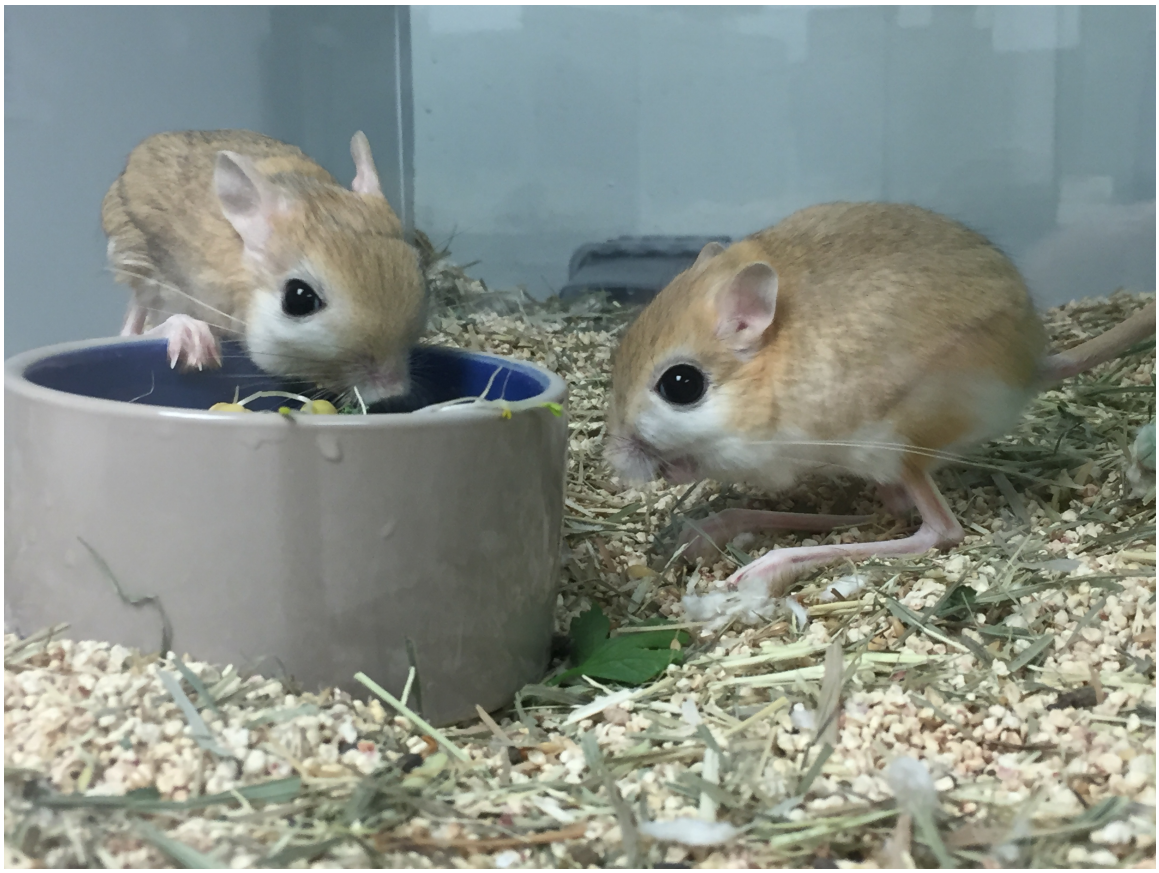


Figure 0.1 | The lesser Egyptian Jerboa

Methods to link genotype and phenotype across species

Limbs are complex structures controlled by number of genes responsible for patterning and modulation of growth. As with all polygenic traits, it is difficult to reduce the complexity of a genome's worth of variation down to the individual genetic modules responsible for specific phenotypes. A common method for inferring the import of given genomic regions on a phenotype is the use of quantitative trait loci (QTL) mapping, where organisms with differing phenotypes (sometimes from different species) are crossed, producing hybrid offspring. These hybrids, when further crossed, produce offspring with a range of intermediate phenotypes that can be mapped back to specific stretches of DNA inherited from one of the two original lineages[34].

When the organisms being compared are not sufficiently close relatives to produce viable hybrid offspring, some researchers have employed an alternative method where genetic sequences from one species are artificially introduced into another. A key feature of this technique is the simplification of millions of years of accumulated genomic change into a single region of interest. Rather than investigate the phenotypic output of a given region in the context of a divergent genome, where variation in other genomic regions may muddy interpretations, these studies test the sufficiency of a single genetic region to produce a phenotype. Furthermore, by transferring genetic regions from non-model organisms into common model systems amenable to manipulation, researchers are able to investigate complex evolved morphological variation within a commonly used species with well-established protocols.

Previous work along these lines has focused on the introduction of species-specific enhancers into mice to test both their expression patterns and biological function. The introduction of a transgenic reporter under the control of a *Tbx4* enhancer demonstrated that the Anolis enhancer was sufficient to direct expression to the hindlimbs and genital tubercule in mouse, while snake enhancers could only direct expression to the genital tubercule[35]. Another study examining the degradation of limb enhancers in snake used reporters and functional replacements to study the ZRS, a well described limb enhancer for *Shh*[36]. These researchers removed a core section of the ZRS in mouse, replacing it with the homologous region from humans, coelacanths, and snakes. Though the human and coelacanth ZRS were sufficient to produce limbs indistinguishable from those of a wildtype mouse, animals that had the endogenous ZRS replaced with the snake homologue had severely truncated limbs. The re-introduction of a 17 bp sequence, which is conserved across tetrapods and fish but lost in the snake, was sufficient to rescue this phenotype[37].

Similar studies have been performed to investigate the evolution of the bat wing. Focusing on *Prx1*, a gene known to be essential for limb outgrowth, researchers replaced a previously identified mouse enhancer with the homologous enhancer from bats. The resulting animals had significantly higher expression of *Prx1* and also had significantly longer forelimbs than their wildtype littermates. Unfortunately, this change was only significant at E18.5, before feeding differences could cause variability in growth between littermates[38].

As technologies for targeted replacements have improved, corresponding studies have emerged in the biomedical field with the goal of producing mice with human-like or

'humanized' phenotypes. Humanized mice provide biomedical researchers with a more accurate platform to study human diseases and more reliably test putative therapeutics. Traditionally, humanized mice have been produced by xenograft transplantation of human tissues into mice[39], but as genetic engineering has grown in ease and popularity, new genetically humanized mouse models have become common. The transgenic introduction of human proteins into mouse has allowed researchers to model the Hepatitis C virus, which is otherwise unable to enter mouse cells[40]. The introduction of a mutant allele of the *F9* gene, which causes hemophilia in humans, into the *Rosa26* safe harbor locus in mouse, allowed researchers to test gene therapies to correct the mutation in vivo[41]. The humanization of genes involved in the metabolism of drugs, including *Cyp3a4*, *PXR*, and *CAR*, has allowed for more precise testing of human drug metabolism and has the potential to aid in therapeutic development and dosage estimations[42,43].

Whereas evolutionary studies have so far focused on individual genes or enhancers, several humanized mouse models have replaced tens or hundreds of kb of mouse genome with human sequence in an effort to capture many genes or regulatory regions simultaneously. By replacing 120kb spanning the α globin domain, including its regulatory region, mice were produced that expressed only human α globin chains[44]. The replacement of the mouse Ig heavy and light chain variable regions with human sequences (a process in which a total of 6Mb of mouse genome was exchanged for 1.5Mb of human sequence) resulted in a mouse with a functionally humanized immune system[45,46].

We reasoned that similarly large replacements between jerboa and mouse could provide a novel method to study the variations in regulatory sequence that contribute to

change over evolutionary time. The advantages of such a strategy are multifold. First, the replacement of a large region, rather than a single enhancer, allows for the potential capture of multiple regulatory regions simultaneously. Genomic regions with 'enhancer-like' chromatin states outnumber regions with 'promoter-like' chromatin states in the human genome by at least a factor of five[47]. This number of enhancers corresponds to the wealth of evidence that suggests many developmental enhancers have overlapping, redundant, or compensatory functions[48–50]. In the limb, knockouts of 10 conserved enhancers with strong limb expression in transgenic reporter mice resulted in no detectable phenotype. It was only when researchers knocked out pairs of enhancers controlling single genes that limb defects were apparent, suggesting substantial functional redundancy[48].

This redundancy was also apparent in the *Prx1* bat enhancer study. Whereas knock outs of *Prx1* gene function result in mice with shortened limbs, knocking out the validated *Prx1* limb enhancer in mouse produced no phenotype[38]. Indeed the ZRS, the enhancer modified in the previously mentioned snake studies, is unusual in that it is indispensible for limb function. Knock-outs of the ZRS alone result in mice with severely truncated limbs[37]. Intriguingly, multiple conserved subregions of the ZRS have proven to be dispensable for proper limb development, suggesting that functional redundancy exists between subregions of the ZRS[51]. Together, the evidence for redundancy suggests that animals have systems to buffer modification of important developmental enhancers.

Redundancy has been primarily investigated in the context of loss of function mutations, where a gene or enhancer is rendered non-functional. It remains to be seen what effect, if any, redundant regulation may have on a replacement or gain of function.

Notably, when the mouse *Prx1* enhancer was replaced with the bat *Prx1* enhancer, the resulting phenotype was mild, though significant. Embryonic forelimbs of animals with the bat enhancer were only 6% longer than their wild type littermates[38]. Of note, limb length has consistently been demonstrated to be a highly polygenic trait, suggesting that any single gene may be insufficient to produce a detectable phenotype[52–54]. Therefore the mild phenotype could be due to the minimal additive effect of *Prx1* amidst a large number of genes promoting limb growth. Perhaps more importantly, the difference in the length of forelimbs was only significant at a single time point, and adult animals were indistinguishable from wild type individuals[38]. The reversal of this lengthened forelimb phenotype may suggest a mechanism of altered limb gene expression to correct for the bat enhancer, either through differential regulation of other limb genes or through cooperation and signaling between multiple regulators of *Prx1* itself.

The use of large regions that encapsulate coding sequences in concert with *cis* regulatory elements has the potential to circumvent redundancy between enhancers for a single gene while maintaining the local genomic structure. Such a replacement could allow for the detection and study of complex regulatory interactions between multiple sequences within a locus. As some regulatory elements for developmental genes can be located at a great distance from their target (the *Bmp4* mesoderm enhancer is located 46kb distal from its promoter[55] while the ZRS is located nearly 1Mb from *Shh*[36]), large replacements may, in some cases, be the only possible method by which an enhancer and coding sequence can be transferred concordantly. Furthermore, the inclusion of multiple regulatory regions in a single genetic replacement has the potential to simultaneously produce more than one phenotype in a single animal as many developmental genes are

pleiotropic. In all cases, subsequent analysis by chromatin profiling or chromosome capture methods could be used to identify subregions responsible for particular phenotypes.

Importantly, the use of a large replacement, rather than a targeted enhancer replacement, is essentially agnostic to the specific sequence that may produce a phenotype. Rather than relying on previously reported regulatory elements, this technique has the potential to link phenotypes to genetic sequences that would not have been obviously selected based on previous research.

The prospect of performing these large replacements between jerboa and mouse is particularly appealing because of the close phylogenetic relationship shared between the two rodents. Whereas snakes diverged from mice more than 350 mya and bats diverged from mice more than 90 mya, the jerboa's more recent common ancestor with the mouse (55 mya) may increase the likelihood that mouse transcription factors and repressors will appropriately bind to sites in a jerboa sequence. As such, a 'jerbouse' might act as an interesting system to test whether diverged regulatory sequences are capable of being recognized across evolutionary time. If mouse cells are competent to recognize and interact with jerboa sequences, we can identify not only the jerboa sequences responsible for the jerboa's derived limb morphology, but also the limits of interspecies recognition and divergence.

Technical advances and limitations to large genomic replacements

The production of transgenic animals has now been common practice for decades, but the technology has continued to grow and change. The introduction of exogenous DNA into an organism can be divided broadly into random or targeted insertion techniques. The introduction of double stranded DNA into cultured cells[56] or zygotes[57] by microinjection, can result in the random integration of the exogenous DNA into loci throughout a cell's genome. This integration appears to happen at random locations within the genome and multiple insertions can occur within a single cell or lineage[58]. Because of the inability to control where an integration may occur, these 'transgenic' lines may exhibit variability due to position effect[59,60] or copy number differences[61,62], or lethality due to the interruption of essential genes (reviewed here[63]).

Targeted genetic modifications, unlike random transgenic insertions, result in the modification of a specific locus. Early work showed that flanking a selection cassette with sequence homology to the intended locus of integration could result in the targeted insertion of the cassette by homologous recombination (HR)[64,65]. Subsequent work revealed that HR in both plants and animals was dramatically more efficient when a double stranded DNA break (DSB) was generated by a rare-cutting site-specific endonuclease[66,67]. Utilization of such DSBs for the purposes of genome editing was limited for a time by the need for an endonuclease recognition site that was present at the locus of interest but absent from the rest of the genome.

The subsequent development of programmable zinc-finger nucleases (ZFNs) and TAL-effector nucleases (TALENs) (the history of which is reviewed beautifully by

Chandrasegaran and Carroll[68]) and the discovery[69] and implementation[70–73] of the more flexible CRISPR/Cas9 system, allowed researchers to direct the generation of DSBs to particular loci, opening up a new era of genome editing potential.

CRISPR/Cas9, the most flexible and most widely used of these technologies, was initially discovered in bacteria, where it is used to combat invading viruses by selectively cleaving viral DNA. In brief, short regions of viral DNA are incorporated into the bacterial host chromosome. Subsequent transcription of these sequences produces RNA molecules complementary to the virus. These CRISPR RNAs (crRNAs) interact with a second trans-activating crRNA (tracrRNA), which in turn forms a complex with the Cas9 endonuclease. This complex is guided to the viral target by the crRNA, where Cas9 cleaves the viral DNA. Cleavage is dependent both upon complementarity to the target site and also upon the presence of a protospacer adjacent motif (PAM), typically NGG, immediately adjacent to the target[69]. The bacterial targeting of viral DNA can be recapitulated in other cells by the production of a single molecule guide RNA (sgRNA or gRNA) that fulfills the roles of crRNA and tracrRNA together[69,70]. When combined with Cas9, this system can be used to generate targeted DSBs in vitro and in vivo.

Regardless of the method of DSB generation, it is the subsequent repair of the break that will define the ultimate genotype of any engineered animal. Upon recognition of a DSB, cells deploy a variety of endogenous DNA repair machinery. End-joining (EJ) repair pathways, including the non-homologous end-joining (NHEJ)[74,75] and micro-homology mediated end-joining (MMEJ, also termed alt-NHEJ)[76] pathways predominate during most of the cell cycle in somatic cells[75,77]. These pathways frequently result in small insertion or deletion mutations (indels), which can disrupt the function of a coding

sequence. As a result, the most common use for gene editing technology in the lab is the generation of mutant alleles. It is important to note that a secondary consequence of such indel generation is the disruption of the targeted sequence such that it can no longer be recognized by the same gRNA or programmable endonuclease.

An alternative DNA repair pathway, the homology directed repair (HDR) pathway, can result in the seamless repair of a DSB through recognition of homology between sequences flanking a DSB and sequences in a corresponding template[75,77]. By introducing a template with a desired insert flanked by homologous sequences, researchers can incorporate their desired sequence directly into the targeted location after DSB generation. While such 'knock-ins' have been shown to be effective in multiple animal systems, they have highly variable and typically low efficiency. A 2015 review of genetic engineering in mice reported HDR efficiencies from 8.1%-66%[78]. Furthermore, efficiency of integration seems to decrease as a function of the size of the insertion and it is rare to see published insertions larger than 3kb.

While CRISPR/Cas9 modification of ESCs is possible, many researchers have opted to directly edit zygotes to speed up the process of producing a modified animal. When working with ESCs, several rounds of screening are required to produce a live animal. ESCs must be modified and then injected into blastocysts to produce mosaic mice. Those mice must subsequently be mated to isolate the products of germ cells carrying the desired modification. In contrast, zygotic injection directly produces mosaic mice, eliminating one round of modification and screening. Nevertheless, the engineering of ESCs has a potent advantage over genetic modifications made by zygotic injections. ESCs, unlike zygotes, can be grown in antibiotic containing medium. The incorporation

of antibiotic resistance genes into an HDR template allows researchers to selectively enrich for clones that have incorporated the desired insertion.

Such selection in a system where thousands of cells can be assessed improves identification of low efficiency HDR events, which is presumably why the largest HDR insertions made in mammalian cells to date have been made in cell culture. Whereas, to my knowledge, the largest insertion by HDR in zygotes is 5.8kb with a reported 32% efficiency, the same study successfully introduced 7.4kb into the same locus with a reported 67% efficiency in ESCs[79]. To my knowledge, the largest insertion to date in mouse ESCs, published in May, 2020, reports the deletion of an 18.1kb genomic region and replacement with 45.5kb of recombinant sequence using CRISPR-mediated HDR, albeit with the frequent generation of undesirable indels on at least one chromosome[80]. Interestingly, both of these studies emphasize the use of long homology arms to maximize efficiency.

Even larger scale replacements have been performed in ESCs, using non-CRISPR based techniques (beautifully reviewed here[81,82]). Humanization efforts have taken advantage of Recombinase Mediated Cassette Exchange (RMCE) and HR in order to produce mice with 120kb and 200kb replacements respectively. In brief, in an RMCE strategy, site-specific recombination sites (SSRs) are introduced at either end of the region to be replaced. For example, a region in the genome can be modified so that it carries a loxP site at the 5' end and a lox511 site at the 3' end. A targeting vector carrying the desired replacement is modified so that it, too, carries these SSRs. Finally, a recombinase capable of recognizing the SSRs (Cre, in our example) is introduced. This recombinase facilitates recombination between each pair of SSRs (loxP with loxP;

lox511 with lox511) resulting in a direct exchange of the sequence lying between the two sites in the donor and recipient. Thus the genomic region between the two sites is exchanged for the sequence in the targeting vector[44].

This method, while effective, requires several rounds of targeted modification in ESCs, as SSRs must be introduced before an exchange can be performed. Such manipulation of ESCs can limit the success of germ cell contribution after blastocyst injections as ESCs are known to lose pluripotency with successive passaging and manipulations[83]. In contrast, a method dependent on HR can be used for the targeted insertion of large regions in a single step. It has been shown that efficiency of HR and HDR seems to improve with increased length and specificity of the homology between donor and recipient[84–86]. By using a donor vector with massive amounts of homology flanking the desired insertion, HR can be used to produce targeted insertions of more than 200kb in a single step[45].

We opted to replace a mouse locus with the homologous jerboa region using this latter method, based on the observed efficiency and low number of manipulations required in ESCs. Although it was anticipated that the addition of CRISPR/Cas9 could improve the rate of HR, we determined that the production of unwanted indels either at our locus of interest or at off-target sites could substantially confound our findings.

Combining modified regions of interest

Modeling of multigenic traits, including limb length and morphology, may require the assembly of more than one engineered locus. While we anticipate that large

replacements might contain multiple regulatory elements of interest, many of the classic limb patterning genes that may contribute to a phenotype fall on independently assorting chromosomes. Indeed, efforts to humanize mice have frequently required multiple engineered loci to be combined in a single mouse to produce a suitably humanized phenotype[43,45].

In order to combine engineered loci, multiple rounds of breeding may be required, making the assembly of many homozygous loci in a single animal extremely inefficient. If, for example, a researcher crossed mice that were heterozygous at each of three loci, the rules of Mendelian inheritance predict that less than 2% of all offspring would be homozygous at all three loci. This problem is compounded as further loci are added. Addition of just a single locus to our hypothetical cross means that a vanishingly small 0.4% of offspring are expected to be homozygous at all loci.

Development of self-propagating alleles, which are inherited at greater-than-Mendelian proportions may rectify this problem. Naturally occurring 'selfish' alleles, which are inherited at higher rates than other genomic regions, have been identified in both plants and animals. Transposable elements, first discovered in maize, can insert themselves to novel locations in a genome, multiplying even within a single cell[87]. Naturally occurring 'meiotic drivers', including the *SD* locus in *Drosophila*[88] and the t-haplotype in mice[89] are able to skew germline transmission such that they are inherited at higher rates than their wildtype counterparts.

Perhaps most interesting of these selfish alleles are the Homing Endonuclease Genes (HEGs), found in fungi, plants, algae, and bacteria. HEGs encode sequence-specific endonucleases capable of cutting precisely their own location of insertion. When

a HEG+ chromosome encounters a HEG- chromosome, the encoded endonuclease cleaves the HEG- site. The resulting DSB is repaired by HDR using the HEG+ chromosome as a template, allowing the HEG+ allele to propagate by being copied from one chromosome to another[90].

In order to produce self-propagating alleles that could be manipulated in the lab, researchers theorized that synthetic HEGs could be developed[91]. These artificial homing cassettes, now used successfully in flies[92,93] and mosquitoes[94], are inserted into the genome at a gRNA target sequence, and consist of the coding sequence for Cas9 as well as the gRNA targeting the precise location of insertion on the opposite chromosome. When the Cas9 and gRNA are expressed, the gRNA recognizes the opposite chromosome and Cas9 generates a DSB, which can be repaired by HDR from the homologous chromosome. The coding sequences for Cas9 and gRNA are copied to the opposite chromosome producing a cell homozygous for the Cas9/gRNA allele. By duplicating the allele, this system ensures that any offspring of this cell will inherit the Cas9/gRNA allele. Importantly, this system takes advantage of the high flexibility of CRISPR-directed endonuclease systems, and so can be encoded almost anywhere in the genome.

Reasoning that such a system could be used to propagate multiple alleles to successive generations at rates higher than those possible by Mendelian segregation, we endeavored to test the feasibility of such a system in a vertebrate by developing a modified ‘split-drive’ system in mouse.

Here I present two systems for the genetic manipulation of mouse to produce models of ‘jerboanized’ phenotypes. First, I elaborate on the direct replacement of the mouse *Msx2* locus with the jerboa’s homologous region utilizing HR as previously used in humanized mouse models. Then I describe a method for the super-Mendelian inheritance of alleles in mouse, which may be used to combine multiple alleles contributing to a jerboa limb phenotype.

Chapter 1: Large interspecies genomic replacement to investigate evolution of the developmental role of *Msx2* in the lesser Egyptian jerboa

Abstract

Naturalists dating back to 1903 have remarked on the converged morphology of highly divergent species that navigate their environments through cursorial locomotion. In particular, many cursorial species have lost digits, presumably to conserve energy by decreasing the weight of the distal limb. Of the animals that have exhibited a digit loss phenotype, horses, camels, and jerboas, a family of obligate bipedal rodents, have expanded expression of a highly conserved apoptotic pathway controlled by *Bmp4* and *Msx2*. Here we examine the *Msx2* locus, highlighting significant differences seen between the lesser Egyptian jerboa and other rodents, including accelerated evolution at this locus. Finally, we demonstrate a method to replace 25.8kb of the mouse *Msx2* locus with the corresponding 31.4kb of jerboa homologous sequence in embryonic stem cells to produce a 'jerboanized' mouse model of *Msx2*.

Main Text

Despite their diversity, mammals have remarkably conserved underlying skeletal morphology, particularly recognizable in the long bones of the limb. While proportions may differ, the identity of individual limb bones has remained essentially static across the phylum so that nearly every mammal species has a readily identifiable humerus, radius,

and metacarpal. An interesting exception to this pattern is the number of toes found in different species. Though the vast majority of mammalian species retain the ancestral five digits on both fore and hindlimbs, the loss of one or more digits has arisen many distinct times in the mammalian lineage. Digits have been lost multiple times in the Artiodactyls, which include cows, pigs, and deer, as well as in the Perissodactyls, which include horses, rhinos, and tapirs, and in some rodents, including the jerboas[30,95]. Intriguingly, zoologists dating back to the early 1900s have linked the morphological adaptations observed in deer, horses, and jerboas to their shared cursorial locomotion[32] and it has been hypothesized that reduction of weight by digit loss in the distal limb may provide an energy saving adaptation for cursorial locomotion[96–98].

While the loss of digits may be a common adaptation, little is known about the mechanism by which the autopod may be modified to carry fewer digits. Research into *Hemiergis* skinks[5], cows[99], pigs, camels, horses, and jerboas[95] has illuminated some of the shifting patterns of gene expression that may contribute to a digit loss phenotype in these animals, but the underlying genetic basis of these expression changes remains unknown.

Whereas acquisition of embryonic Artiodactyl and Perissodactyl specimens can prove challenging, and rearing these animals in a lab is impractical, the jerboa is small, easy to work with, and can be reared in the lab. As a model system for digit loss, the jerboa has two additional advantages over species like the horse or camel. The jerboa is closely related to the common laboratory mouse, *Mus musculus*, with whom it shared a common ancestor only ~55 million years ago (mya). Furthermore, unlike the horse or camel, the jerboa has lost the first and fifth digits exclusively on its hindlimb; the more

'mouse-like' forelimb of the jerboa retained all five digits. Thus the jerboa can be easily studied in the lab and provides a superb model to understand divergent morphologies both between species, by comparing jerboa to mouse, and within an organism, by comparing forelimb to hindlimb.

Previous work in *Dipus sagitta*, a three-toed jerboa for which embryos could be readily obtained, examined the expression domains of classical limb-patterning genes in the hindlimb and forelimb of jerboa and mouse. This work established that while most limb patterning genes have stereotypical patterns of expression in both rodents, *Msx2*, a homeobox gene transcription factor, and *Bmp4*, a secreted ligand from the TGFβ superfamily, have expanded domains of expression specifically in the developing jerboa hindlimb[95]. *Msx2* and *Bmp4* are expressed between the forming digits of most tetrapods in a well described pattern associated with interdigital cell death[27,28,100], but in the jerboa, both *Msx2* and *Bmp4* are expressed in an additional expanded domain at the anterior and posterior margins of the distal hindlimb. This expression overlaps closely with the region that is expected, in mouse, to produce the first and fifth digit and corresponds to a jerboa hindlimb-specific domain of apoptotic cells (Figure 1.1a-c[95]).

It is well established that *Bmps* and *Msx2* cooperate in a positive feedback loop that results in programmed cell death between developing digits[27,28,101] and it has been suggested that inter-species differences in digital webbing may be the result of modulation of this pathway[27,102,103]. *Bmp* expression induces *Msx2* expression[28], which, in turn, binds to its own enhancers[104] and *Bmp* enhancers[101], simultaneously upregulating its own expression and that of *Bmp4*[105], and ultimately stimulating apoptosis. Overexpression of either *Bmp4*[106], *Bmp5*[107], or *Msx2*[105] is sufficient to cause aberrant cell

death and viral overexpression of *Msx2* has been shown to cause aberrant sculpting of the limb and diminished chondrogenesis[105]. It seems possible then that this pathway could be co-opted to sculpt the limb of the jerboa by carving away the first and fifth digit.

Intriguingly, despite the canonical description of this pathway as a direct result of *Bmp* signaling, the expansion of *Msx2* expression in the jerboa is detectable at E11, a full day before the expansion of *Bmp4* expression can be detected at E12 (Figure 1.1 b,c[95]). This early expansion of the *Msx2* domain is also coincident with the first detectable changes in gross limb morphology, with the jerboa hindlimb becoming steadily more distinct from the mouse hindlimb after E11.5 (Figure 1.1d[95]).

The initial investigation of *Bmp4* and *Msx2* expression was performed on wild-caught *D. sagitta* embryos because these animals, prevalent in the deserts of China, have a narrow window of reproduction, maximizing the likelihood of capturing pregnant females for embryological studies[29]. To further investigate the genetic mechanism by which *Msx2* and *Bmp4* acquire their expanded expression domains in the jerboa hindlimb, we turned to our lab model system, the lesser Egyptian jerboa, *Jaculus jaculus*, which has a sequenced genome and can be lab reared[33]. *D. sagitta* and *J. jaculus* both belong to a derived clade of three-toed jerboas in which all species have lost digits I and V. Since parsimony infers that their most recent common ancestor, from which they diverged only ~12 mya[6] had three toes, it is likely that the mechanism of digit loss is shared in *D. sagitta* and *J. jaculus*[30].

This shared mechanism for the early expansion of the *Msx2* expression domain in the jerboa hindlimb may be the result of either *cis* or *trans* regulatory changes. Changes to local genomic structure could cause the modification, gain, or loss of regulatory

elements that control *Msx2* expression. Alternatively, changes in the expression pattern of transcription factors that target conserved *cis* elements that govern *Msx2* could result in an altered domain of expression. It is further possible that undetectable alteration of *Bmp4* expression in the early limb bud is responsible for the changes in *Msx2* expression.

We chose to focus on *Msx2*, rather than *Bmp4*, because the *Msx2* domain of expression was expanded earlier, hinting that *Msx2* may be acting upstream of *Bmp4* in the apoptotic pathway that sculpts digits in the jerboa. Interestingly, *Msx2* lies within a gene desert, with the nearest coding region more than 90 kb distal to the end of the *Msx2* transcript. Because regulatory information is largely coded in intergenic regions, we reasoned that the region surrounding *Msx2* might be enriched for regulatory domains. As the coding sequence (Figure 1.2) and known limb enhancer are largely preserved between the two species, we chose to expand our region of focus past the boundaries of the *Msx2* gene.

First, to determine whether there had been large-scale organizational changes in the locus, I searched for regions of homology in the intergenic region surrounding *Msx2* and identified 13 distinct conserved regions (Cons1-Cons13) that span more than 30 kb in jerboa (Figure 1.3a). These conserved regions include the predicted transcript of *Msx2* (largely represented by Cons6-10) and the previously validated limb enhancer (largely represented by Cons4). To determine the depth of homology and synteny of these regions within the mammalian phylum, we searched for Cons1-Cons13 in an additional 21 rodents and 48 other mammals with sequenced genomes. Eight of these regions were identifiable in all mammals examined, and more than half of species examined had all 13 conserved regions (Table 1.1). In all cases, the synteny of the locus was maintained,

suggesting that no major rearrangements have occurred. We therefore identified the *Msx2* region of interest (*Msx2* ROI) as the genomic stretch between and including Cons1 and Cons13 (Figure 1.3a).

Intriguingly, despite the absence of obvious rearrangements, an obvious difference was apparent between the mouse and jerboa. The full length of the jerboa *Msx2* ROI is 31.4 kb, whereas the equivalent region in the mouse is only 25.8 kb (Table 1.2). It is possible that the length of the region is highly variable across mammals, that the jerboa has gained approximately 5.6 kb of sequence, or that the mouse has lost approximately 5.6 kb of sequence.

To differentiate between these possibilities, I compared the length of the *Msx2* ROI in 71 mammalian species. Across all mammals sampled in this study, the average length is 26.1 kb ± 2.0kb (standard deviation). Within rodents sampled, the average length is 25.4 kb ± 1.8 kb (standard deviation), with jerboa the only significant outlier (Grubbs' test, two-sided, $p<0.01$). This suggests that it is the jerboa, and not the mouse, that is anomalous (Table 1.3, Figure 1.4).

To determine whether the increased sequence length is due to a localized accumulation of sequence in one area or spread equally along the full length of the region, I divided the *Msx2* ROI into four distinct subregions, based on conserved regions present in all species assayed. Because there exists no single nucleotide boundary adjacent to islands of high sequence conservation, I included the full length of the conserved region at either end of each subregion. Subregion A stretches from the beginning of the *Msx2* ROI to the known limb enhancer (Cons1-Cons4); Subregion B from the enhancer to the approximate beginning *Msx2* transcription start site (Cons4-Cons6); Subregion C

encompasses the annotated transcript of *Msx2* (Cons6-Cons10); and Subregion D includes the remainder of the *Msx2* ROI (Cons10-Cons13) (Figure 1.3a).

Measuring the length of each subregion revealed that while each subregion was expanded compared to mouse, Subregion A was disproportionately expanded, and accounted for 80% of the total difference in length between jerboa and mouse (Table 1.2). Interestingly, the size of Subregion A correlates strongly with the full length of the *Msx2* ROI across all 71 mammals assayed (slope=0.8605, R^2=0.9345, p<0.0001), whereas the adjacent Subregion B has no significant correlation (slope=0.04821, R^2=0.04263, p=0.0841) (Figure 1.5a). This suggests that while the length of Subregion B has stayed fairly static across all mammals in this analysis, the expansion and contraction of the full *Msx2* ROI is tightly linked to the expansion and contraction of Subregion A.

Again, it is possible that changes to the length of Subregion A are localized to one particular location or spread across the subregion, so we further characterized this region in mouse and jerboa by measuring Subregion A-1 (Cons1-Cons3) and A-2 (Cons3-Cons4). Variance in Subregion A-1 accounts for 73% of the total difference in Subregion A and Subregion A-2 accounts for 29% (Table 1.2). Therefore, while the bulk of the change is localized in Subregion A-1, there are at least two domains of expansion even within Subregion A. The extension of this analysis to all mammals was not possible because Cons3 could not be identified in all species analyzed.

Although the length of both *Msx2* ROI and Subregion A in jerboa were significant outliers compared to other rodents (Figure 1.4), it is possible that this is due to biases in species sampling. Only one jerboa species has a published genome sequence to date, whereas 8/23 rodent sequences used in this study are from species within the Muridae

family, which includes *Mus musculus*. Thus a trait change in the ancestor of Muridae may skew our comparison. We corrected for this by applying a phylogenetic correction to weight the character state of 'region length' and used the resulting inferred trees to estimate rates of evolution for this trait on both the length of Subregion A and B (Figure 1.5c,d). Interestingly, the lengths of both Subregion A and B have evolved faster in the jerboa than in the mouse compared to the neutral rate of nucleotide substitution, though the difference was more profound in Subregion A (neutral J:M= 0.683, Subregion A J:M=1.13, Subregion B J:M=0.99).

Although the difference in length between mouse and jerboa *Msx2* ROI is both obvious and phylogenetically interesting, it remains unclear what specific sequence differences exist between mouse and jerboa, let alone the degree to which any change has biological significance. While direct comparisons between mouse and jerboa can reveal profound insights into the mechanism of development and evolution[54,95,108,109], interpretations of such comparisons can be rendered difficult due to the accumulated genetic change of 55 million years of evolution. To simplify the investigation of the genetic changes responsible for digit-loss in the jerboa, we endeavored to produce a mouse model in which the mouse *Msx2* ROI was replaced with the jerboa's homologous region (Figure 1.3c) to probe the sufficiency of the jerboa *Msx2* ROI to produce the molecular, cellular, and organismal phenotypes seen in the jerboa. By isolating the jerboa *Msx2* ROI in a mouse model, any resulting phenotypes can be confidently linked to this individual genomic region.

Due to the large-scale augmentation of the region in jerboa across all subregions, we opted to make a mouse model of the entire *Msx2* ROI to preserve any and all local

regulatory and coding sequence differences in the native genomic context. To produce this model, we aimed to fully replace the mouse *Msx2* ROI with the jerboa *Msx2* ROI. A direct replacement avoids the potential pitfalls associated with randomly inserted transgenic alleles by mitigating the risk of disrupting important genomic regions or varying expression levels according to copy number[61,62] and position effect[60]. Such a large replacement further allows for the preservation of local genomic architecture, including potential interactions between sites. It is worth noting that reports indicate that expression may be partially dependent on the genomic distance between enhancer and promoter elements[110]. Consistent with this, the length of Subregion B, which spans the distance from the known limb enhancer to the beginning of the *Msx2* coding sequence, is highly invariate across all mammals in this study. Therefore it may be important to preserve the precise distances between elements in the *Msx2* ROI for an accurate reproduction of a 'jerboanized' phenotype.

The replacement of 25.8 kb of mouse sequence with 31.4 kb of homologous jerboa sequence is not currently feasible using common CRISPR/Cas9-based knock-in techniques in zygotes. To our knowledge, the largest reported CRISPR knock-in by zygotic injection to date was only 5.8 kb[79]. Instead, we employed a method, reliant on homologous recombination (HR) in mouse embryonic stem cells (ESCs) which has previously been used to successfully humanize massive stretches of the mouse genome[45].

To perform this interspecies genomic replacement, I first constructed a chimeric bacterial artificial chromosome (BAC), detailed further below, which contains the desired jerboa genomic replacement, a selection marker flanked by lox sites, and tens of kb of flanking mouse homology. This BAC was electroporated into ESCs where endogenous

machinery incorporates the replacement by HR between the massive homology arms and the mouse genome (Figure 1.3b,c). The efficiency, even of this previously published technique, is extremely low, with correct incorporation estimated to occur in 0.1-0.5% under stringent selection[45].

Owing to the need for a selection cassette to facilitate identification of successful HR in cell culture, one end of the replacement must contain a 'scar', a difference in the mouse genome that is not solely due to homologous replacement with jerboa genomic DNA. We chose to embed the selection cassette, with its inevitable loxP scar, just upstream of the alignment between mouse and jerboa in Cons1. To ensure the least disruption to the integrity of the locus, we chose to make the downstream transition from jerboa to mouse entirely seamless by targeting the downstream transition point to a region of perfect homology between mouse and jerboa within Cons13 (Figure 1.6b).

The chimeric BAC was constructed using viral recombineering[111,112] in three phases using a jerboa BAC (jBAC) and a mouse BAC (mBAC), each containing the *Msx2* ROI. First, I added cassettes to the jBAC at the upstream and downstream transition points by electroporating linearized donor cassettes and inducing expression of viral proteins that recognize and facilitate recombination between sites of homology. The upstream cassette contains a jerboa homology arm (177 bp), followed by a unique I-CeuI cut site, 154 bp of mouse homology, a Kanamycin resistance (Kan^R) cassette bounded by loxP sites, and finally a second jerboa homology arm (206 bp). The floxed Kan^R cassette contains a mammalian PGK promoter as well as the synthetic EM7 bacterial promoter, each of which can drive expression of an aminoglycoside phosphotransferase that provides resistance to Kanamycin (in bacteria) and G418 (in mouse ESCs). This dual

promoter setup ensures that the same resistance gene can be used to screen bacterial clones during recombineering and mouse ESC clones after HR. (Figure 1.7a).

The process was repeated at the downstream transition point using a cassette containing a jerboa homology arm (208 bp), a second I-CeuI cut site, a bacterial promoter driving a Zeocin resistance gene (Zeo^R), and a second jerboa homology arm (180 bp). It is essential to note that the first jerboa homology arm adjacent to the I-CeuI site begins within the Jerboa Cons13 region and encompasses 85 bp of sequence with perfect conservation in mouse and jerboa (Figure 1.7a).

The resulting doubly modified jBAC was digested using I-CeuI to release a linear fragment containing the floxed Kan^R cassette and 31.4 kb of jerboa sequence bounded by small regions of mouse homology. This enormous construct was incorporated into the mBAC by recombineering to produce a chimeric BAC in which the full mouse *Msx2* ROI was replaced with the jerboa *Msx2* ROI (Figure 1.7b). This chimeric BAC was then linearized and electroporated into mouse ESCs by the University of Michigan Transgenic Animal core. 102 clones were selected on G418 and shipped to UCSD to screen for the deletion and replacement of the mouse *Msx2* ROI.

While the acquisition of G418 resistance does implicate the integration of at least the Kan^R cassette, it is possible for the desired jerboa insert to be truncated or missing altogether. Therefore, in order to identify clones that acquired the entire jerboa locus, I amplified 6 regions sampling the jerboa *Msx2* ROI by PCR. This technique measures only presence or absence of an insert but does not rule out the possibility of a random insertion, so additional screening was performed on clones containing the jerboa *Msx2* ROI to assay correct targeting.

In a typical knock-in strategy, either PCR or Southern blot is employed to identify correct junctions between the genome and insertion. Unfortunately, as the length of homology arms on either end of the desired replacement is upwards of 50kb, it is impossible to use either of these methods to detect the entire junction, which by definition must span endogenous mouse sequence, 50kb of mouse homology arm, and inserted jerboa sequence.

Instead, I employed the Loss Of Native Allele (LONA) method[45,113], which uses qPCR to assay the zygosity of the region of interest. In the case of a single allele insertion, we expect the mouse *Msx2* ROI to be deleted from one chromosome and replaced with the jerboa locus. LONA assumes that detection of both the deletion of the "native allele" and the insertion of the new allele in a single clone indicates a successful exchange of the two.

To detect deletion of the native allele, I performed quantitative amplification of the mouse *Msx2* ROI from clones that had acquired the jerboa insertion. In successfully targeted clones, we expect to detect approximately half the amount of mouse *Msx2* ROI detected in a wildtype clone or a clone with a random insertion when normalized to an unrelated homozygous locus. To ensure deletion of this extensive region, zygosity was assessed at five locations spanning the mouse *Msx2* ROI.

To minimize the possibility of erroneously identifying a clone where the deletion of the mouse region and insertion of the jerboa region are unlinked, I also assayed the zygosity of a homology arm near the terminus of the jerboa *Msx2* ROI. In the event of a random insertion encompassing the full jerboa region, it is likely that this section of the homology arm would be included. Therefore, in the case of a random insertion, we

expect to detect 1.5 times the amount of this homology arm compared to either wild type clones or successfully modified clones.

Finally, to measure copy number of the jerboa *Msx2* ROI insertion, I used qPCR to compare relative amounts of jerboa-specific amplification between clones.

I identified a single clone (of 102 G418 resistant candidates) that appeared to have lost one copy of the full mouse *Msx2* ROI and also acquired a single copy of the entire jerboa *Msx2* ROI, while maintaining the expected level of homology arms (Table 1.4). When taken together, the quantitative analysis of this region suggests that the mouse *Msx2* ROI was correctly targeted in this clone, and was likely deleted and replaced with the jerboa *Msx2* ROI in a single insertion event. This clone was injected into blastocysts to produce chimeric mice by the University of Michigan Transgenic Animal Core. Of seven animals born after blastocyst injections, one male and one female have visible mosaicism indicative of high levels of chimerism (Figure 1.8).

Although the phenotypes of the chimeras and their offspring have not yet been analyzed, the successful integration of 31.4 kb of jerboa genomic DNA into a mouse ESC to replace the homologous endogenous mouse locus, represents a significant step towards the production of a jerboanized mouse. Though the phenotypes have not been measured, it is interesting to speculate about the possible outcomes of such an interspecies replacement.

If the *Msx2* ROI does indeed contain the information responsible for the developmental and evolutionary expansion of *Msx2* expression in the jerboa hindlimb, we anticipate detection of a similarly expanded domain in the jerboanized mice. In order to differentiate between expression deriving from the endogenous mouse *Msx2* and the

introduced jerboa *Msx2*, we have already produced in situ hybridization probes that detect the species specific 3' UTR of the *Msx2* transcript. Using two-color in situ hybridization, we can visualize an expanded expression domain from a single allele, and also potentially determine whether an expanded domain of expression of jerboa *Msx2* is sufficient to upregulate ectopic expression of the endogenous mouse *Msx2*. If an expansion of *Msx2* is sufficient to replicate aspects of digits loss, we may see an increase in the domain of apoptotic cells and ultimately observe mice with fewer, smaller, or malformed digits.

It is worth noting, of course, that our approach is intentionally agnostic to any particular phenotype. We identified a region from which gene expression differs in an evolutionarily meaningful context and for which there is statistically and phylogenetically significant sequence divergence from other mammals. We suspect interesting phenotype(s) may arise from this exchange of information, but in truth, there is no direct evidence that these sequence differences are directly linked to divergent *Msx2* expression or the three-toed phenotype. However, *Msx2* is a highly pleiotropic gene, with roles in cell survival, cell proliferation, and cell death in diverse tissues including limb outgrowth[114], spinal cord[115], eye development[116], and facial structure[116]. Therefore it is entirely possible that we will observe an *Msx2*-linked phenotype elsewhere in the body in addition to or instead of the limb. Intriguingly, the same positive feedback loop wherein *Bmp4* and *Msx2* drive cell death in the interdigital region is also seen in neural crest, where apoptosis selectively ablates rhombomeres, which are segmented structures in the neural tube that give rise to specific populations of neural crest derivatives[117]. Neural crest cells diversify in the early embryo to produce a variety of tissue types including

neurons, pigment, connective tissue and muscle, and craniofacial structure[118]. *Msx2* loss has been linked to craniofacial malformations in humans[119], so it is particularly interesting to note that the jerboa has a shorter and broader facial structure compared to mouse as well as modified dentition[29].

It is further possible that any phenotype seen in the resulting mouse model will be unrelated to *Msx2* expression entirely. The genomic location of *Msx2* lies in a gene desert that may coincidentally place it nearby enhancers or other regulatory elements responsible for the control of entirely unrelated genes. Regulatory elements, particularly for developmental genes, have been reported at great distances from their corresponding coding regions. *Bmp4*, itself, has a prominent limb enhancer more than 46 kb from its transcription start site (TSS)[55,101]. Perhaps the most famous limb enhancer, the ZRS, which controls *Sonic Hedgehog* expression in the limb, is located an incredible 1Mb away from the TSS of *Shh*[36].

Any observed phenotype can be further investigated and linked to more specific genomic sequences by using chromatin interrogation methods to identify small regions active within the large replacement. ATAC-Seq can be used to identify regions of open chromatin with a precision of 100-200bp[120], while HINT-ATAC can be used to subsequently predict which of those open regions are occupied by transcription factors and repressors[121]. Circular chromosome conformation capture, also termed 4-C, can reveal physical interactions between the region of replacement and any other genomic regions, allowing for the identification of genes controlled by long-range enhancers located in the *Msx2* ROI[122].

Ultimately, small regions specific to the jerboa that are predicted to have regulatory significance can be individually replaced in mice to produce precise targeted replacement models. As these regions are likely to be substantially smaller than the full *Msx2* ROI, small targeted replacements can be engineered by more conventional means, including zygotic injection with CRISPR/Cas9 to induce HDR, saving substantial time and money and facilitating the simultaneous production of multiple small replacements.

Finally, we would be remiss if we did not consider the possibility that the jerbouse will have no obvious phenotype. In itself, this would be intriguing. Despite the homology of coding sequence and conserved regions, the jerboa and mouse genomes have undergone about 55 million years of evolutionary divergence. As previously noted, evidence suggests that at least the length of this region has evolved more quickly in the jerboa than the basal rate of nucleotide substitution. An absence of phenotype would, therefore, be not only surprising, but also informative, as it could indicate mechanisms to buffer variation in the mouse.

Regardless of phenotypic outcome, we suggest that large replacements have potential to reveal the mechanisms of evolutionary change and species diversity. By isolating specific genomic regions of interest, such interspecies exchanges may allow for the study of dramatically divergent phenotypes in an identical genetic background, reducing the potential noise from millions of years of genomic divergence. Observation of a phenotype resulting from such a replacement provides a perfect proof of sufficiency that a given genetic sequence is responsible for that phenotype, needing no additional evolutionary modifications.

Whereas previous studies have used interspecies enhancer replacements to investigate the modification of gene expression over evolutionary time[35,37,38], direct enhancer replacements rely on the knowledge of specific and distinct enhancers in different species. Furthermore, as enhancers are frequently highly redundant[48–50], the modification of a single enhancer may be insufficient to produce a phenotype. By capturing a large region, we hope to encapsulate a greater portion of the local regulatory landscape, preserving complex interactions and maximizing the likelihood of including relevant regulatory information to ultimately produce a model of a jerboa phenotype in a mouse.

Figures and Tables

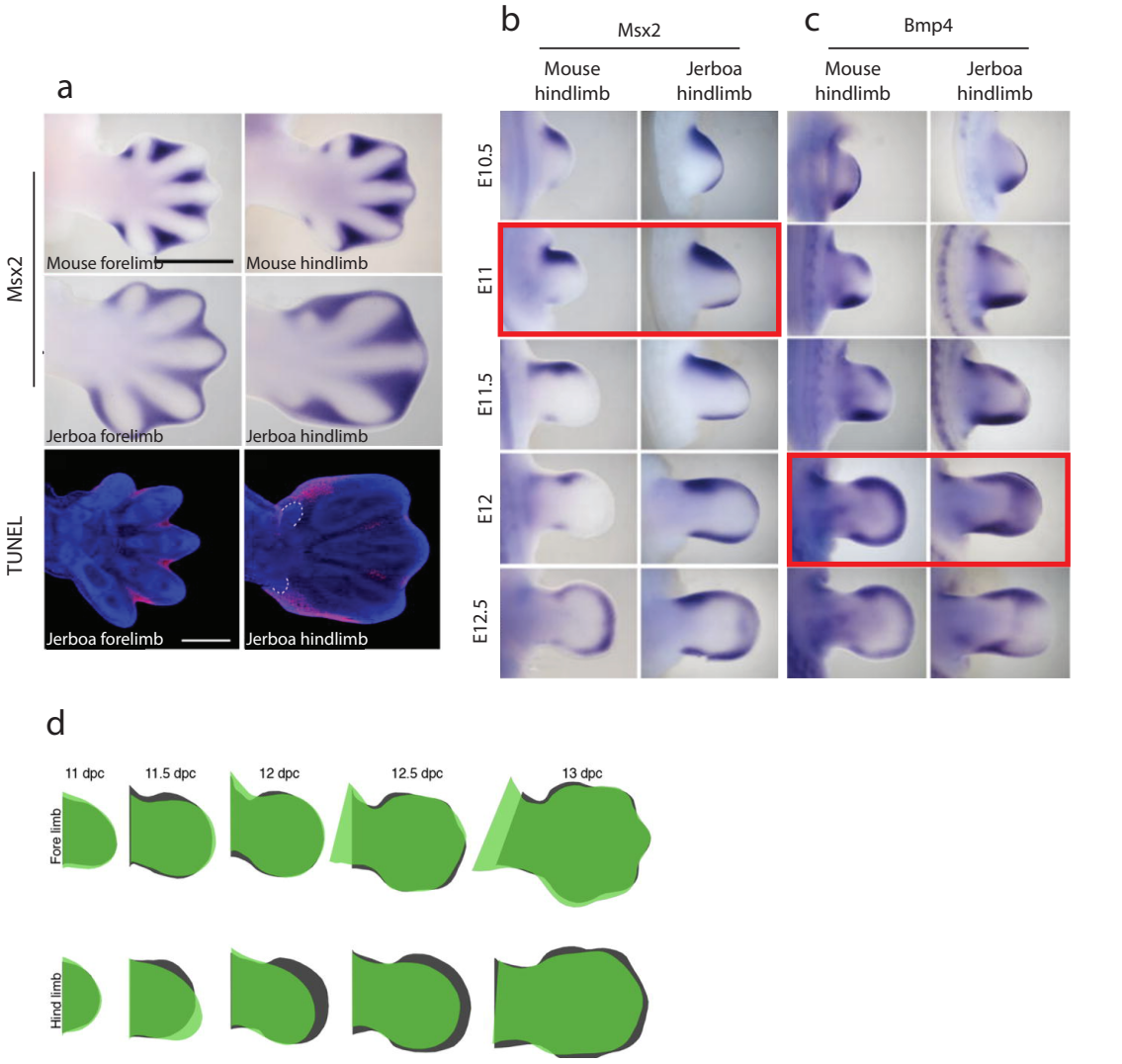


Figure 1.1 | Expansion of *Msx2*, *Bmp4*, and apoptosis in the developing jerboa hindlimb. a. In situ hybridization reveals expanded domain of *Msx2* expression in jerboa hindlimb at E13.5; TUNEL staining reveals expanded domain of apoptosis in the jerboa hindlimb at E13.5. **b,c.** *Msx2* has visibly expanded expression in the jerboa hindlimb a day before the expansion of *Bmp4* is detectable (red box- first stage with expanded domain of expression). **d.** The jerboa hindlimb changes shape compared to the mouse hindlimb (green- silhouette of mouse limb; grey- silhouette of jerboa hindlimb). This figure, in its entirety was adapted from Cooper, K. L., Sears, K. E., Uygur, A., Maier, J., Baczkowski, K.-S., Brosnahan, M., Antczak, D., Skidmore, J. A. & Tabin, C. J. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511,** 41–45 (2014).

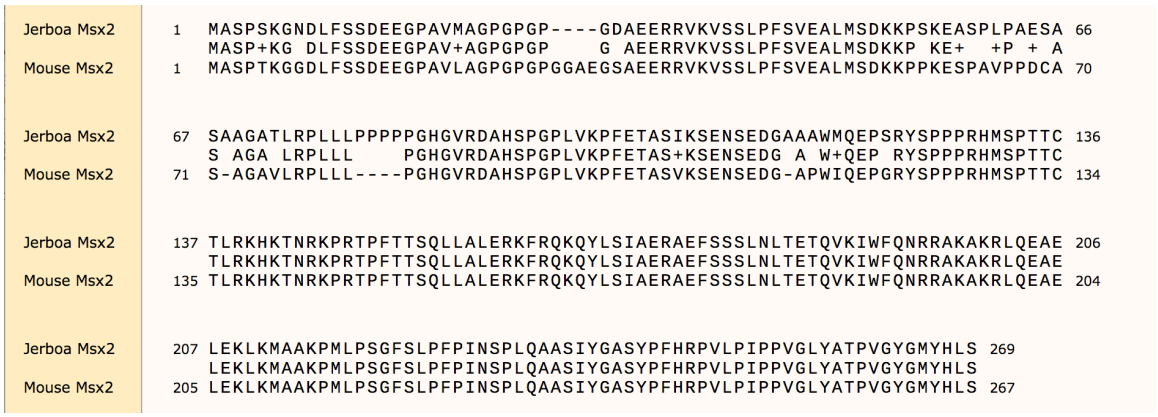


Figure 1.2 | Protein alignment of Mouse and Jerboa *Msx2*.

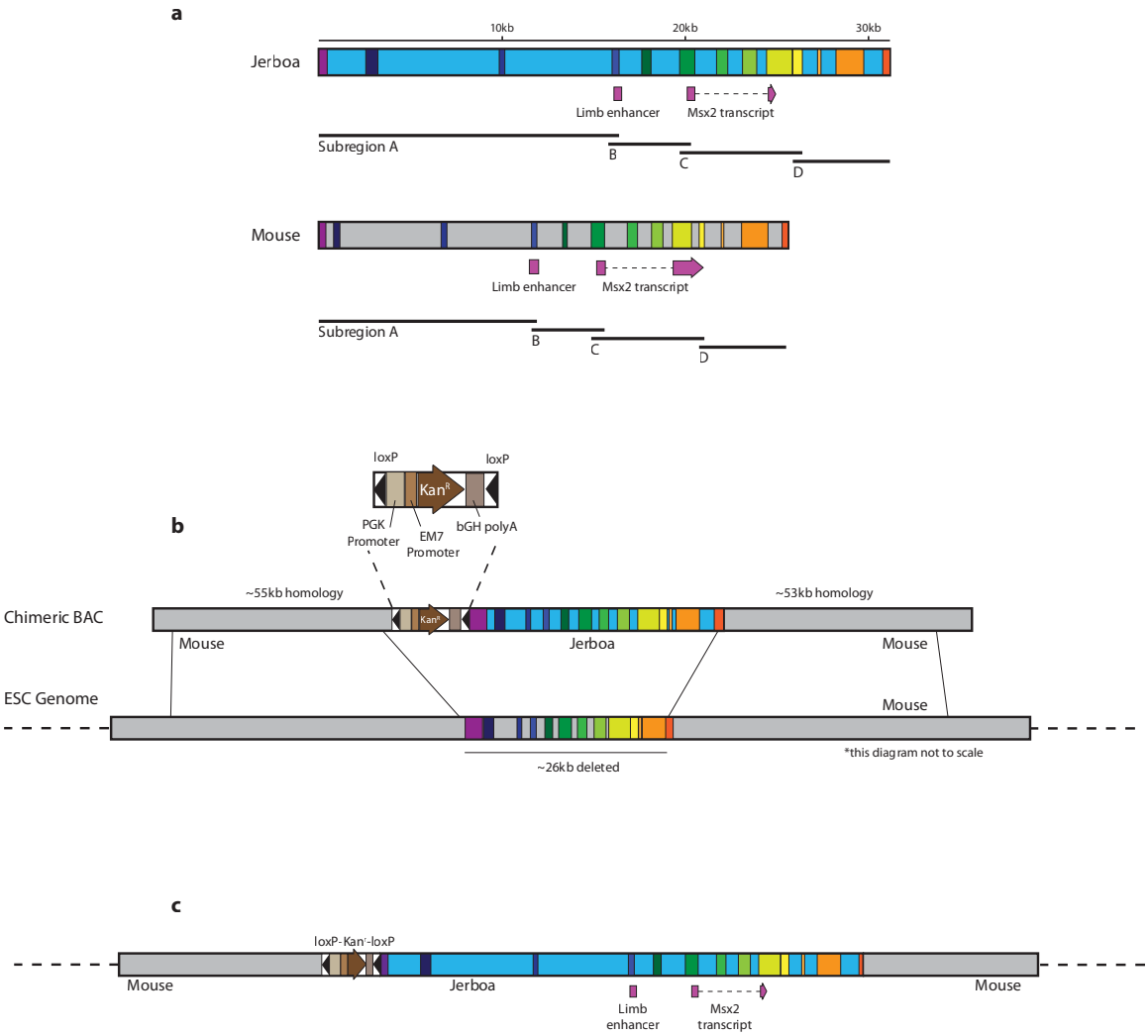


Figure 1.3 | Strategy for exchanging *Msx2* Region Of Interest (ROI). a. The *Msx2* ROI. The jerboa and mouse *Msx2* ROI share 13 conserved regions (rainbow boxes) with maintained synteny (light blue- jerboa sequence; grey- mouse sequence). The *Msx2* ROI is divided into Subregions A-D. **b.** Replacement strategy. A chimeric BAC is produced containing the jerboa *Msx2* ROI, a floxed Kanamycin resistance cassette (Kan^R), and more than 50kb of mouse homology on either end. When electroporated into embryonic stem cells (ESCs), the jerboa ROI is incorporated into the ESC genome by homologous recombination, simultaneously deleting the mouse ROI and replacing it with the jerboa sequence. Diagram b is not to scale. **c.** Mouse genome with successful jerboa *Msx2* ROI replacement.

Figure 1.4 | Schematic of *Msx2* ROI length in rodents. Molecular phylogeny (see Figure 1.5) of rodents with lagomorphs as an out group. Bars depict the length of the full *Msx2* ROI (black), Subregion A (blue), or Subregion B (red) for each species. The lesser Egyptian jerboa is the only outlier for the full length region or Subregion A. Marmot is the only outlier for Subregion B (Grubbs' test, two tailed, $p<0.01$).

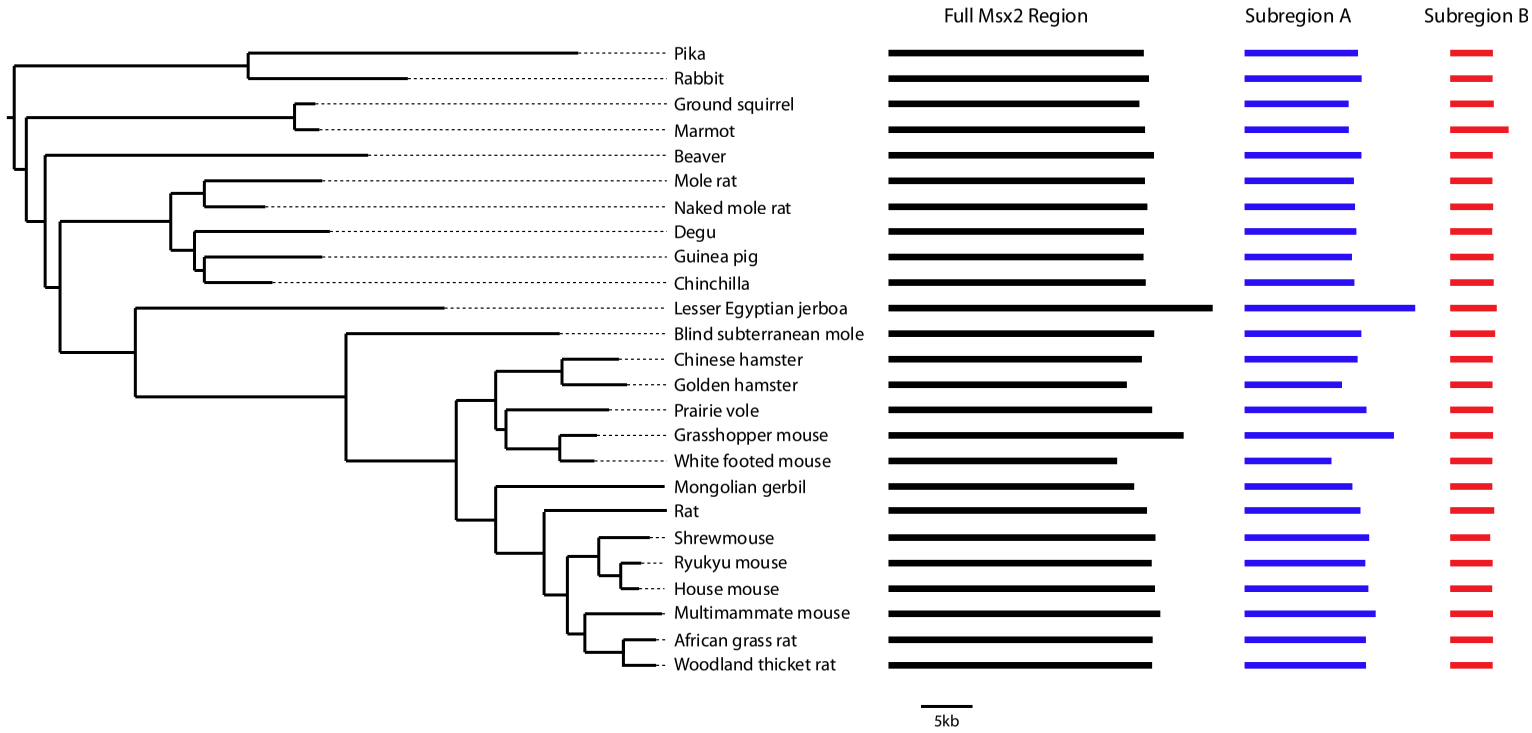
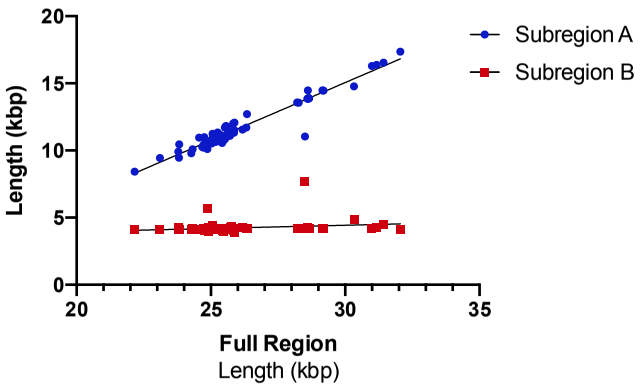
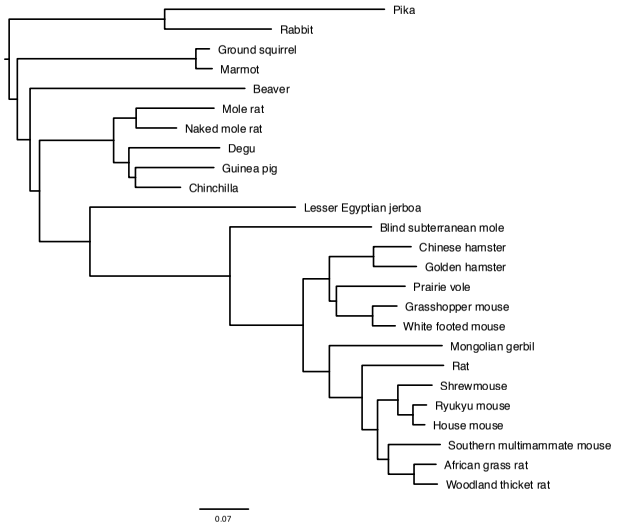

Full Msx2 Region
Subregion A
Subregion B
Pika
Rabbit
Ground squirrel
Marmot
Beaver
Mole rat
Naked mole rat
Degu
Guinea pig
Chinchilla
Lesser Egyptian jerboa
Blind subterranean mole
Chinese hamster
Golden hamster
Prairie vole
Grasshopper mouse
White footed mouse
Mongolian gerbil
Rat
Shrewmouse
Ryukyu mouse
House mouse
Multimammate mouse
African grass rat
Woodland thicket rat
5kb

Figure 1.5 | Phylogenetic characterization of variation in the lengths of Subregions A and B. a. The length of Subregion A, but not Subregion B, strongly correlates with the length of the *Msx2* ROI across 71 mammals. **b.** A molecular phylogeny of rodents based on the nucleotide sequence of the *Msx2* gene. Lagomorphs (rabbit and pika) are included as an outgroup. Branch lengths represent relative rate of evolution. Scale bar- 0.07 substitutions per position. **c.** Phylogeny from b, weighted with the length of Subregion A for each species. **d.** Phylogeny from b, weighted with the length of Subregion B for each species.

a


Subregion A
Subregion B
Length (kbp)
Full Region
Length (kbp)
0
5
10
15
20
20
25
30
35

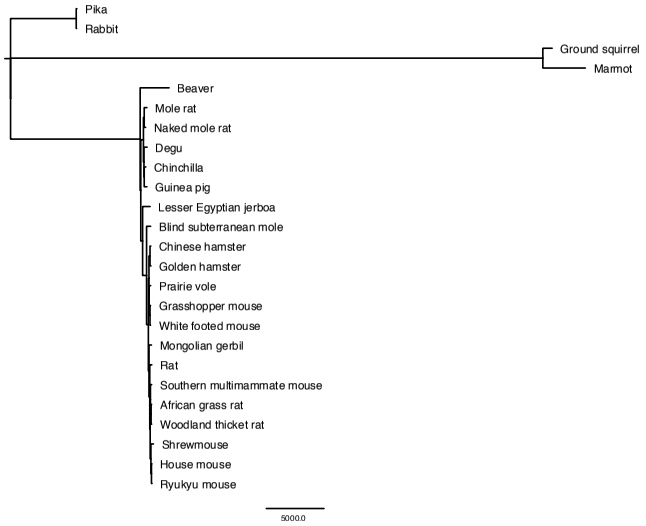
b


Pika
Rabbit
Ground squirrel
Marmot
Beaver
Mole rat
Naked mole rat
Degu
Guinea pig
Chinchilla
Lesser Egyptian jerboa
Blind subterranean mole
Chinese hamster
Golden hamster
Prairie vole
Grasshopper mouse
White footed mouse
Mongolian gerbil
Rat
Shrewmouse
Ryukyu mouse
House mouse
Southern multimammate mouse
African grass rat
Woodland thicket rat
0.07

c


Pika
Rabbit
Ground squirrel
Marmot
Beaver
Mole rat
Naked mole rat
Degu
Chinchilla
Guinea pig
Lesser Egyptian jerboa
Blind subterranean mole
Chinese hamster
Golden hamster
Prairie vole
Grasshopper mouse
White footed mouse
Mongolian gerbil
Rat
Southern multimammate mouse
African grass rat
Woodland thicket rat
Shrewmouse
House mouse
Ryukyu mouse
50000.0

d


Pika
Rabbit
Ground squirrel
Marmot
Beaver
Mole rat
Naked mole rat
Degu
Chinchilla
Guinea pig
Lesser Egyptian jerboa
Blind subterranean mole
Chinese hamster
Golden hamster
Prairie vole
Grasshopper mouse
White footed mouse
Mongolian gerbil
Rat
Southern multimammate mouse
African grass rat
Woodland thicket rat
Shrewmouse
House mouse
Ryukyu mouse
5000.0

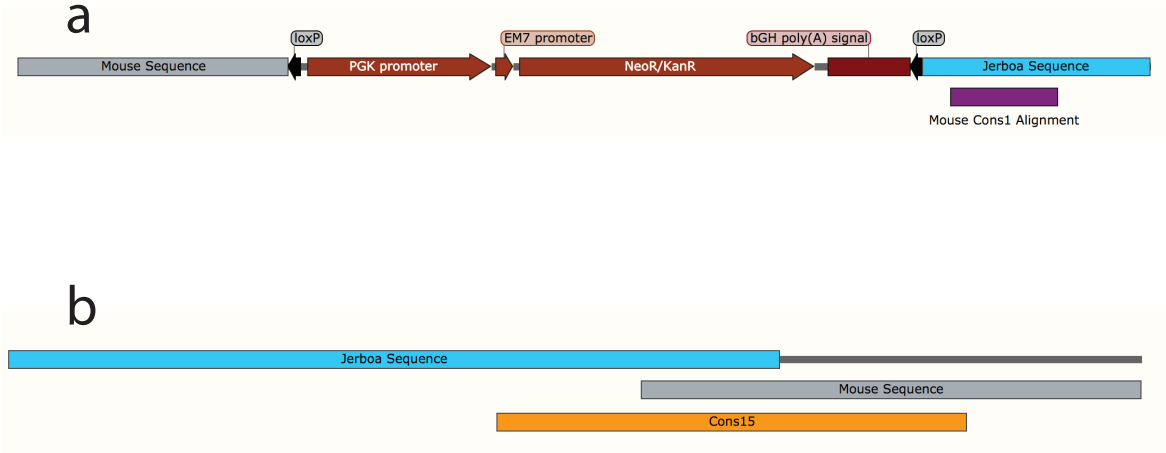


Figure 1.6 | Replacement transition points. **a.** The upstream transition point with floxed Kanamycin resistance (Kan^R) cassette is located 79 bp upstream of the end of Mouse Cons1 alignment with jerboa. **b.** The downstream transition point is seamless as it falls within a region of perfectly conserved sequence in Cons15.

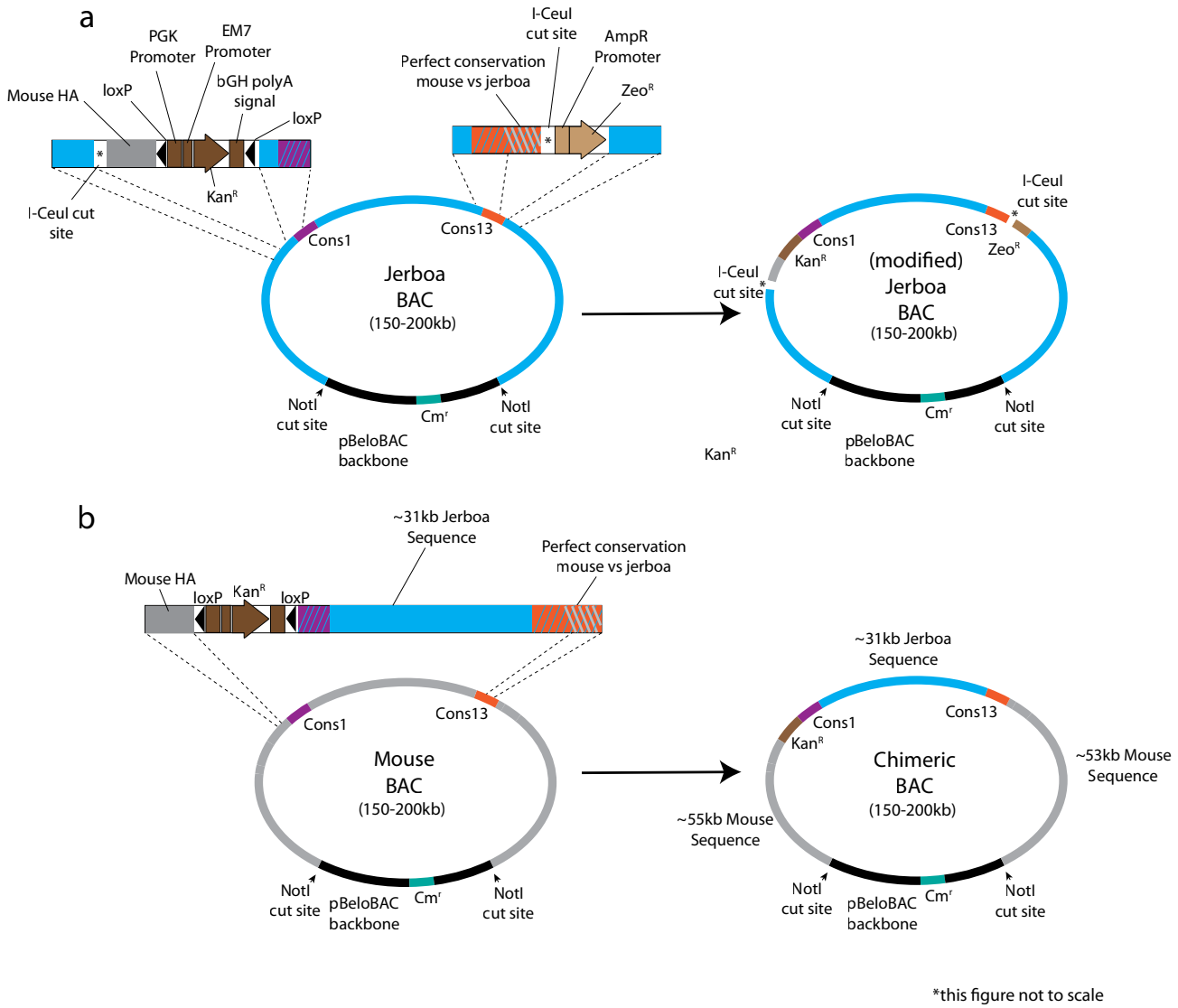


Figure 1.7 | Construction of the chimeric BAC. a. Addition of upstream and downstream cassettes to the jerboa BAC (light blue). The upstream transition point is immediately upstream of Conserved 1 (Cons1, purple). The downstream transition point falls in a region of perfect conservation within Conserved 13 (Cons13, orange). **b.** Deletion and replacement of the mouse *Msx2* ROI in a mouse BAC. The modified jerboa region is linearized using I-CeuI. Dotted lines indicate homologous recombination. Hash marks on conserved regions indicate whether the sequence perfectly matches mouse (grey) or jerboa (blue). (light blue- jerboa sequence; grey- mouse sequence; Kan^R- kanamycin resistance; Zeo^R-Zeocin resistance; Cm^R-chloramphenicol resistance; asterisks- I-CeuI cut sites; arrowheads-NotI cut sites)

Figure 1.8 | Chimeric mice born after blastocyst injection. Modified ESCs carrying the jerboa *Msx2* ROI replacement in a C57B6 background were injected into blastocysts with an albino background. Patches of pigmented skin are presumed to be donated from the modified ESCs.

Table 1.1 | Presence or Absence of Conserved Regions 1-13 in Mammals

1 indicates presence, 0 indicates absence, asterisk indicates all regions present

		Conserved Regions												
Common name	**Species name**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
African elephant	Loxodonta	1	0	1	1	1	1	1	1	1	1	1	1	1
Sheep	Ovis aries	1	0	1	1	1	1	1	1	1	1	1	1	1
Cow	Bos taurus	1	0	1	1	1	1	1	1	1	1	0	1	1
White tailed deer*	Odocoileus virginianus	1	1	1	1	1	1	1	1	1	1	1	1	1
Water buffalo	Bubalus bubalis	1	0	1	1	1	1	1	1	1	1	1	1	1
Pilot whale	Globicephala melas	1	0	1	1	1	1	1	1	1	1	1	1	1
Orca*	Orcinus orca	1	1	1	1	1	1	1	1	1	1	1	1	1
Bottlenose dolphin*	Tursiops truncatus	1	1	1	1	1	1	1	1	1	1	1	1	1
Narwhal*	Monodon monocerus	1	1	1	1	1	1	1	1	1	1	1	1	1
Dromedary camel	Camelus dromedarius	1	0	0	1	1	1	1	1	1	1	1	1	1
Wild boar	Sus scrofa	1	0	1	1	1	1	1	1	1	1	1	1	1
Hyena	Crocuta crocuta	1	0	1	1	1	1	1	1	1	1	0	1	1
Grey seal*	Halichoerus grypus	1	1	1	1	1	1	1	1	1	1	1	1	1
Elephant seal	Mirounga leonina	1	1	0	1	1	1	1	1	1	1	1	1	1
Walrus*	Odobenus rosmarus	1	1	1	1	1	1	1	1	1	1	1	1	1
Grizzly bear*	Ursus arctos horribilis	1	1	1	1	1	1	1	1	1	1	1	1	1
Meerkat	Suricata suricatta	1	0	1	1	1	1	1	1	1	1	0	1	1
Giant panda*	Ailuropoda melanoleuca	1	1	1	1	1	1	1	1	1	1	1	1	1
Sea lion*	Zalophus californianus	1	1	1	1	1	1	1	1	1	1	1	1	1
Stoat*	Mustela erminea	1	1	1	1	1	1	1	1	1	1	1	1	1
River otter*	Lontra canadensis	1	1	1	1	1	1	1	1	1	1	1	1	1
Cheetah	Acinonyx jubatus	1	0	0	1	1	1	1	1	1	1	1	1	1
Red fox	Vulpes vulpes	1	1	0	1	1	1	1	1	1	1	1	1	1
Cat	Felis catus	1	0	0	1	1	1	1	1	1	1	1	1	1
Dingo	Canis lupis dingo	1	0	0	1	1	1	1	1	1	1	1	1	1
Canada lynx	Lynx canadensis	1	0	0	1	1	1	1	1	1	1	1	1	1
Mouse eared bat	Myotis myotis	1	0	0	1	1	1	1	1	1	1	1	1	1
Egyptian fruit bat*	Rousettus aegyptiacus	1	1	1	1	1	1	1	1	1	1	1	1	1
Greater horseshoe bat	Rhinolophus ferrumequinum	1	1	0	1	0	1	1	1	1	1	0	1	1
Big brown bat	Eptesicus fuscus	1	1	0	1	1	1	1	1	1	1	0	1	1
Star-nosed mole	Condylura cristata	1	1	0	1	0	1	1	1	1	1	1	1	1
Pika	Ochotona princeps	1	1	0	1	1	1	1	1	1	1	1	1	1
Rabbit	Oryctolagus cuniculus	1	1	0	1	1	1	1	1	1	1	1	1	1
Horse*	Equus caballus	1	1	1	1	1	1	1	1	1	1	1	1	1
Pangolin*	Manis javanica	1	1	1	1	1	1	1	1	1	1	1	1	1
Pig tailed macaque*	Macaca nemestrina	1	1	1	1	1	1	1	1	1	1	1	1	1
Olive baboon*	Papio anubis	1	1	1	1	1	1	1	1	1	1	1	1	1
Orangutan*	Pongo abelii	1	1	1	1	1	1	1	1	1	1	1	1	1
Red colobus monkey*	Piliocolobus tephrosceles	1	1	1	1	1	1	1	1	1	1	1	1	1
Bonobo*	Pan paniscus	1	1	1	1	1	1	1	1	1	1	1	1	1
Gorilla*	Gorilla gorilla	1	1	1	1	1	1	1	1	1	1	1	1	1
Human*	Homo sapiens	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 1.1, Continued | Presence or Absence of Conserved Regions 1-13 in Mammals

1 indicates presence, 0 indicates absence, asterisk indicates all regions present

Marmoset*	Callithrix jacchus	1	1	1	1	1	1	1	1	1	1	1	1	1
Francois langur monkey*	Trachypithecus francoisi	1	1	1	1	1	1	1	1	1	1	1	1	1
Silvery gibbon*	Hylobates moloch	1	1	1	1	1	1	1	1	1	1	1	1	1
White cheeked gibbon*	Nomascus leucogenys	1	1	1	1	1	1	1	1	1	1	1	1	1
Tufted capuchin*	Sapajus apella	1	1	1	1	1	1	1	1	1	1	1	1	1
White footed mouse	Peromyscus leucopus	1	1	1	1	0	1	1	1	1	1	1	1	1
Golden hamster*	Mesocricetus auratus	1	1	1	1	1	1	1	1	1	1	1	1	1
Mongolian gerbil*	Meriones unguiculatus	1	1	1	1	1	1	1	1	1	1	1	1	1
Ground squirrel	Ictidomys tridecemlineatus	1	0	1	1	1	1	1	1	1	1	1	1	1
Chinese Hamster	Cricetulus griseus	1	0	0	1	0	1	1	1	1	1	1	1	1
Guinea pig	Cavia porcellus	1	1	1	1	1	1	1	1	1	1	0	1	1
Degu	Octodon degus	1	1	1	1	0	1	1	1	1	1	1	1	1
Mole rat*	Fukomys damarensis	1	1	1	1	1	1	1	1	1	1	1	1	1
Marmot	marmota marmota	1	0	0	1	1	1	1	0	1	1	1	1	1
Chinchilla*	Chinchilla lanigera	1	1	1	1	1	1	1	1	1	1	1	1	1
Rat*	Rattus rattus	1	1	1	1	1	1	1	1	1	1	1	1	1
Naked mole rat*	Heterocephalus glaber	1	1	1	1	1	1	1	1	1	1	1	1	1
Mus caroli*	Mus caroli	1	1	1	1	1	1	1	1	1	1	1	1	1
Woodland thicket rat*	Grammomys surdaster	1	1	1	1	1	1	1	1	1	1	1	1	1
Prairie vole	Microtus ochrogaster	1	1	1	1	1	1	1	1	1	1	0	1	1
African grass rat*	Arvicanthis niloticus	1	1	1	1	1	1	1	1	1	1	1	1	1
Beaver*	Castor canadensis	1	1	1	1	1	1	1	1	1	1	1	1	1
Blind subterarnnean mole	Nannospalax galili	1	1	1	1	1	1	1	1	1	1	0	1	1
House Mouse*	Mus musculus	1	1	1	1	1	1	1	1	1	1	1	1	1
Mus pahari*	Mus pahari	1	1	1	1	1	1	1	1	1	1	1	1	1
Multimammate mouse*	Mastomys coucha	1	1	1	1	1	1	1	1	1	1	1	1	1
Grasshopper mouse	Onychomys torridus	1	1	1	1	0	1	1	1	1	1	1	1	1
Lesser Egyptian Jerboa*	Jaculus jaculus	1	1	1	1	1	1	1	1	1	1	1	1	1
Chinese tree shrew	Tupaia chinensis	1	0	1	1	1	1	1	1	1	1	0	1	1
	Portion of species with this region	1	0.75	0.79	1	0.92	1	1	0.99	1	1	0.87	1	1
	Number of species with this region	71	53	56	71	65	71	71	70	71	71	62	71	71

Table 1.2 | Comparison of *Msx2* ROI Subregions in mouse and jerboa

	Full Region	Subregion A	Subregion A-1	Subregion A-2	Subregion B	Subregion C	Subregion D
House mouse length (kbp)	25.8	12.0	7.0	5.3	4.1	6.3	4.8
Lesser Egyptian jerboa length (kbp)	31.4	16.5	10.3	6.6	4.5	6.7	5.3
Difference (kbp)	5.6	4.5	3.3	1.3	0.4	0.4	0.5
Percent change	+21.7%	+37.5%	+47.1%	+24.5%	+9.8%	+6.3%	+10.4%
% of total difference		80.4%	58.9%	23.2%	7.1%	7.1%	8.9%

Table 1.3 | Length of *Msx2* ROI and Subregions in Mammals

Common name	Scientific name	Full Region (bp)	Sub-Region A (bp)	Sub-Region B (bp)	Sub-Region C (bp)	Sub-Region D (bp)	Clade
African elephant	Loxodonta africana	30325	14781	4875	6837	5392	Afrotheria
Sheep	Ovis aries	25178	10657	4155	6510	5430	Artiodactyla
Cow	Bos taurus	25440	11001	4147	6429	5425	Artiodactyla
White tailed deer	Odocoileus virginianus	25489	11051	3990	6493	5263	Artiodactyla
Water buffalo	Bubalus bubalis	25520	11014	4149	6509	5413	Artiodactyla
Pilot whale	Globicephala melas	25645	11245	4161	6389	5311	Artiodactyla
Orca	Orcinus orca	25697	11274	4165	6403	5299	Artiodactyla
Bottlenose dolphin	Tursiops truncatus	25729	11283	4164	6407	5319	Artiodactyla
Narwhal	Monodon monocerus	25784	11300	4157	6550	5338	Artiodactyla
Dromedary camel	Camelus dromedarius	26176	11551	4279	6394	5425	Artiodactyla
Wild boar	Sus scrofa	26325	11713	4219	6442	5435	Artiodactyla
Hyena	Crocuta crocuta	24917	10690	3951	6370	5181	Carnivora
Grey seal	Halichoerus grypus	25020	10531	4134	6453	5276	Carnivora
Elephant seal	Mirounga leonina	25036	10545	4139	6456	5271	Carnivora
Walrus	Odobenus rosmarus	25183	10653	4124	6464	5346	Carnivora
Grizzly bear	Ursus arctos horribilis	25269	10687	4138	6459	5361	Carnivora
Meerkat	Suricata suricatta	25366	11155	4225	6279	5168	Carnivora
Giant panda	Ailuropoda melanoleuca	25417	10566	4055	6740	5343	Carnivora
Sea lion	Zalophus californianus	25456	10900	4123	6464	5372	Carnivora
Stoat	Mustela erminea	25476	10760	4156	6453	5483	Carnivora
River otter	Lontra canadensis	25537	10870	4137	6466	5444	Carnivora
Cheetah	Acinonyx jubatus	25609	11219	4193	6377	5263	Carnivora
Red fox	Vulpes vulpes	25714	11086	4294	6414	5369	Carnivora
Cat	Felis catus	25781	11400	4089	6395	5272	Carnivora
Dingo	Canis lupis dingo	25860	11321	4178	6365	5360	Carnivora
Canada lynx	Lynx canadensis	25868	11523	4189	6367	5258	Carnivora
Mouse eared bat	Myotis myotis	23788	9916	4118	6398	4697	Chiroptera
Egyptian fruit bat	Rousettus aegyptiacus	23812	9453	4249	6362	5066	Chiroptera
Greater horseshoe bat	Rhinolophus ferrumequinum	24274	9784	4110	6440	5340	Chiroptera
Big brown bat	Eptesicus fuscus	25082	11031	4406	6409	4679	Chiroptera
Star nosed mole	Condylura cristata	24893	10487	4281	6493	5177	Eulipotyphla
Pika	Ochotona princeps	24752	10998	4140	6138	4767	Lagomorpha
Rabbit	Oryctolagus cuniculus	25252	11349	4114	6228	4930	Lagomorpha
Horse	Equus caballus	24728	10215	4167	6443	5270	Perissodactyla
Pangolin	Manis javanica	24676	10258	4119	6580	5212	Pholidota
Pig tailed macaque	Macaca nemestrina	28209	13571	4192	6600	5221	Primates
Olive baboon	Papio anubis	28268	13563	4190	6622	5265	Primates
Orangutan	Pongo abelii	28594	13865	4198	6642	5262	Primates
Red colobus monkey	Piliocolobus tephrosceles	28605	13862	4273	6607	5184	Primates
Gorilla	Gorilla gorilla	28649	13887	4202	6674	5220	Primates
Bonobo	Pan paniscus	28649	13891	4198	6668	5259	Primates
Human	Homo sapiens	28657	13883	4204	6658	5270	Primates
Marmoset	Callithri jacchus	29151	14482	4155	6615	5283	Primates
Francois langur monkey	Trachypithecus francoisi	29193	14462	4203	6614	5157	Primates
Silvery gibbon	Hylobates moloch	30994	16291	4184	6619	5259	Primates
White cheeked gibbon	Nomascus leucogenys	31174	16373	4270	6637	5350	Primates
Tufted capuchin	Sapajus apella	32056	17354	4136	6641	5301	Primates
White footed mouse	Peromyscus leucopus	22164	8425	4090	6400	4800	Rodentia
Golden hamster	Mesocricetus auratus	23108	9443	4106	6340	4607	Rodentia
Mongolian gerbil	Meriones unguiculatus	23818	10454	4086	6092	4504	Rodentia
Ground squirrel	Ictidomys tridecemlineatus	24321	10088	4229	6432	5007	Rodentia
Chinese hamster	Cricetulus griseus	24564	10959	4127	6262	4562	Rodentia
Guinea pig	Cavia porcellus	24724	10410	4210	6425	5132	Rodentia
Degu	Octodon degus	24769	10838	4074	6333	4898	Rodentia
Mole rat	Fukomys damarensis	24860	10597	4096	6375	5179	Rodentia
Marmot	Marmota marmota	24868	10104	5667	5439	5015	Rodentia
Chinchilla	Chinchilla lanigera	24937	10647	4219	6396	5122	Rodentia
Rat	Rattus rattus	25061	11235	4268	6066	4707	Rodentia
Naked mole rat	Heterocephalus glaber	25096	10706	4165	6490	5221	Rodentia
Ryukyu mouse	Mus caroli	25518	11703	4116	6224	4824	Rodentia
Woodland thicket rat	Grammomys surdaster	25554	11774	4120	6223	4738	Rodentia
Prairie vole	Microtus ochrogaster	25559	11818	4170	6213	4835	Rodentia

Table 1.3, Continued | Length of *Msx2* ROI and Subregions in Mammals

African grass rat	Arvicanthis niloticus	25608	11762	4144	6225	4768	Rodentia
Beaver	Castor canadensis	25736	11328	4131	6760	5082	Rodentia
Blind subterranean mole	Nannospalax galili	25756	11319	4359	6429	5014	Rodentia
House mouse	Mus musculus	25831	12001	4073	6268	4814	Rodentia
Shrewmouse	Mus pahari	25880	12077	3882	6254	4681	Rodentia
Multimammate mouse	Mastomys coucha	26349	12709	4151	6243	5356	Rodentia
Grasshopper mouse	Onychomys torridus	28611	14479	4158	6204	5212	Rodentia
Lesser Egyptian jerboa	Jaculus jaculus	31425	16540	4515	6712	5330	Rodentia
Chinese tree shrew	Tupaia chinensis	28507	11047	7720	6350	4856	Scandentia

Table 1.4 | Representative results from screening of ESCs.
All values are relative to genomic sequence of *Bmp4*, expected to have two copies. Yellow- Areas that appear to be hetero- or hemizygous. Mouse *Msx2* ROI 1, 2, and 3 are distinct primer pairs at distinct locations in the mouse *Msx2* ROI.

	Relative Amount		
	Mouse *Msx2* ROI-1	**Homology Arm**	**Jerboa *Msx2* ROI**
Clone 1	0.98	0.92	0.86
Clone 2	1.76	0.90	1.21
Clone 3	1.18	1.22	1.01
Clone 4	0.65	0.87	0.97
Clone 4 (duplicate)	0.66	0.90	0.97

	Relative Amount	
	Mouse *Msx2* ROI-2	**Mouse *Msx2* ROI-3**
Clone 6	1.25	1.29
Clone 7	1.34	1.49
Clone 8	1.29	1.21
Clone 9	0.92	1.08
Clone 10	0.93	0.83
Clone 11	0.94	0.86
Clone 4	0.41	0.40

Methods

Analysis of the *Msx2* ROI

I used NCBI BLAST (Blastn, align two sequences)[123], to compare a large region surrounding the *Msx2* coding sequence (approximately 20kb upstream of the transcription start site and 6kb downstream of the final codon) between jerboa and mouse. I pursued regions with E values less than 1×10^{-25}, identifying 15 regions of high homology between jerboa and mouse. Alignment was confirmed using the SnapGene alignment tool. Regions with a SnapGene identified alignment of fewer than 20bp were removed, leaving 13 regions.

SnapGene was used to align jerboa regions to 21 rodents and 48 other mammals (Table 1.3). Grubbs' test for outliers was performed using GraphPad, alpha=.01. Regression analysis was performed using GraphPad Prism Version 9.0.0.

A molecular phylogeny was constructed using the *Msx2* ROI Subregion 3 nucleotide sequence (Cons6-10), which approximates the annotated *Msx2* gene. This sequence was aligned for rodents and lagomorphs in Table X using the online Clustal Omega tool provided by EMBL-EBI[124,125]. The alignment was imported into SeaView[126] to produce a maximum likelihood phylogenetic tree (PhyML using BioNJ starting tree, default settings), which was rerooted at the base of the Lagomorphs (Figure 1.4).

Bayestraits (http://www.evolution.rdg.ac.uk/SoftwareMain.html) was used to phylogenetically weight the trait of 'region length' using this molecular phylogeny (Independent Contrast, MCMC, 10,000,000 iterations, sampling every 1000th iteration)[127]. A consensus tree was derived using BayesTrees

(http://www.evolution.rdg.ac.uk/SoftwareMain.html) (Figure 1.5). Comparing branch lengths of lesser Egyptian jerboa and house mouse was used to estimate relative evolution across trees.

Producing the chimeric BAC

Jerboa BAC (jBAC)

Clemson University produced a BAC library based on the jerboa genome using HindIII. BACs were estimated to contain 150-200kb of genomic sequence. At our request, our colleagues at Clemson screened the BACs used pooled overgo probes[128]. Probes used for Msx2 are in Table 1.5.

Table 1.5 | Overgo Probes for Msx2

Msx2 Upstream	GAGGAATTTGAGCTTTTCGCCCCAAATCAGCTCTTTAAT TTTTGTTTTTTTCTTGCAGGGTCATTTTTATGACACTTGA GTTTTCTTCACAATGAAAATA
Msx2 Downstream	CAATATATGCGCGCTGACACCGGGTCCAGCCAGCACCA CTGGCCCGGGACATCGCAGGCGACCAGGTGCACTCAAC CGCCCCCCCCTCCGCACCCCCATC

The resulting positives were screened with PCR to identify a clone containing the jerboa *Msx2* ROI. This clone, 0123O11, was used for the remainder of cloning.

Primers are listed in Table 1.6. In general, endpoint PCR was performed using Bioline MyTaq Red MasterMix and 500nM primers with the following conditions: 95° C for 1'; 32 cycles of 95° C for 15", 60° C for 15", 72° C for 45"/kb; 72° C for 5'. Products were run on an agarose gel (1-2%) to check amplicon size.

Table 1.6 | BAC Screening Primers

Name	Sequence
Msx2 20kb Up L1	TTTTTCTTGCAGGGTCATTTTT
Msx2 20kb Up R1	GACTCCAGGTGTTCTCCGATAC
Msx2 gene L2	ACAGCTGTGTGGTTGTTTATGG
Msx2 gene R2	TATCTTCTCCAGGGTGACCTGT
Msx2 10kb Down L1	CGGGTTATCAATCACCCAGC
Msx2 10kb Down R1	GCCTGCGATGTCCCGGGC

Mouse BAC (mBAC)

The B6nNg01-248G4 BAC was identified using the UCSC genome browser. This BAC contains a portion of *Mus musculus* C57BL-6J chromosome 13 (accession number GRCm38.p6) including the mouse *Msx2* ROI, ~55kb of upstream region and ~53kb of downstream region. It should be noted that Nanopore sequencing indicates that this BAC contains a region slightly shifted compared to that reported.

Recombineering

Recombineering was performed using the pSIJ8 plasmid[111], purchased from Addgene (#68122). This plasmid contains L-arabinose inducible viral recombinase proteins that recognize sequences of homology and facilitate recombination. In general, recombineering steps were performed using the following protocol, adapted from a recombineering protocol generously provided by our colleagues Pieter de Jong and Cris Jung at CHORI. Cm- Chloramphenicol, Amp-Ampicillin, Kan-Kanamycin, Zeo-Zeocin.

Step overview:

1. The jBAC was made electrocompetent and then electroporated with pSIJ8.

2. The resulting Cm^R; Amp^R clones were recombineered with the upstream targeting cassette (UTC) and selected on Kan.

3. Kan resistant clones that had sequence confirmed by PCR were made electrocompetent and electroporated with pSIJ8.

4. The resulting Kan^R; Cm^R; Amp^R clones were recombineered with the DTC targeting cassette (DTC) and selected on Zeo and Kan.

5. The doubly modified jBAC was linearized by I-CeuI digestion to release the mBAC targeting cassette (MTC- mouse homology, Kan^R selection, 31kb jerboa sequence, mouse homology) and purified by membrane dialysis for 2hr on a nitrocellulose membrane in TE.

6. The mBAC was made electrocompetent and then electroporated with pSIJ8.

7. The resulting Cm^R; Amp^R clones were recombineered with the full MTC and selected on Kan.

Electroporating with pSIJ8

Grow 5mL overnight liquid culture (target plasmid, Cm) at 32° C, shaking, 16-20hr. Transfer 1mL overnight culture to 100mL LB with Cm. Grow at 32° C, shaking until OD600 = 0.350-0.400. Incubate on ice 30'. Centrifuge at 1000g for 20' at 4° C. Discard supernatant. Wash twice with ice cold sterile diH2O and once with ice cold sterile 10% glycerol, centrifuging as before. Resuspend in 1mL 10% sterile glycerol. Aliquots of 20-50uL can be stored for electroporation.

Add 1uL (50ng) pSIJ8. Electroporate (exponential decay, 25uF, 200ohm, 1800V, 0.1cm cuvette). Add 1mL LB, incubate at 37° C, shaking, for 1.5hr. Plate on LB with Amp and Cm and incubate at 37° C overnight.

Recombineering

Grow 5mL overnight liquid culture (target plasmid, pSIJ8, appropriate antibiotics) at 32° C, shaking, 16-20hr. Transfer 30uL of culture to 1.4mL LB with antibiotics. Shake at 32° C until OD600 = 0.300. Add 20uL 10% L-arabinose in sterile water. Shake 37° C for 40'. Incubate on ice 5'. Centrifuge at 13,000rpm for 1' at 4° C. Remove supernatant. Wash three times with ice-cold sterile 10% glycerol, centrifuging each time at 13,000rpm for 1' at 4° C and removing the supernatant. In the last wash, resuspend in 20-50uL of remaining supernatant Add 200-800ng recombineering cassette. Electroporate (exponential decay, 25uF, 200ohm, 1800V, 0.1cm cuvette). Add 1mL LB, incubate at 37° C, shaking, for 1.5hr. Plate on LB with 25ug/mL Kan and/or Zeo and incubate at 37° C overnight. After overnight growth, clones should be subsequently plated or grown with 50ug/mL Kan.

Cloning Recombineering Casettes

PCR for all cloning was performed using NEB Q5 2X MasterMix with annealing temperatures and extension times as recommended by NEB. Gibson cloning was performed using the NEB Hifi DNA Assembly MasterMix with incubation at 50degrees C for 1hr.

The downstream targetting cassette (DTC) was synthesized as a gene block from IDT. The DTC contains 202 basepairs of jerboa sequence upstream of and including Cons13 (caps). This sequence contains 79 bp of perfect sequence homology between mouse and jerboa (pink). It also contains an I-CeuI recognition site (turquoise), the commonly used Ampicilillin resistance promoter (underlined), a bleomycin resistance gene (yellow), and 180bp of jerboa sequence continuing within and downstream of Cons13 (caps). The full sequence is below.

CCCTTCTAGAAGCGATGGGCGTCGGGAGCTAGAAAGAAGGACACCGTCCTGG
CGCGCGGCCCCTTCCGCCCCGCCTGCCGCCAGCCGCCCGTTTGATGTCGCGGG
CCCTGCAGGAATGCGAGGCGGGTTATCAATCACCCAGCTGGATCCCGAAGGT
CTCAGCCTAATCACATTTAATTGCTCGTGGAGGCCCATTCTCGCCcccggctcgctac
cttaggaccgttatagttacgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctga
taaatgcttcaataatattgaaaaaggaagagtatggccaagttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgcc
ggagcggtcgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttcgccggtgtggtccgggacg
acgtgaccctgttcatcagcgcagtacaggaccaggtagtaccagataacacccttgcttgggtttgggtgagaggtcttgacga
gctttacgctgagtggtcggaggttgtttccacgaacttcagagacgcttccggtcctgctatgaccgagatcggtgagcagcctt
ggggtcgtgagttcgctcttcgtgaccctgctggtaactgcgttcacttcgttgctgaggagcaggactgaCCGCCGCCT
CGCTCAATTACTCCCCAAATATCTGCCATCAATATATGCGCGCTGACACCGGG
TCCAGCCAGCACCACTGGCCCGGGACATCGCAGGCGACCAGGTGCACTCAAC
CGCCCCCCCCTCCGCACCCCCATCATTAAACTTACACCTCGACGCCCAGGGG
AGTAAGCCAAGCAG

The DTC gene block was A-tailed using dATP and Taq Polymerase for 15' at 72degrees C. It was then incorporated into pGEMT-easy using a Quick Ligation Kit from NEB. It was subsequently amplified using Primer 1 and 2.

The Upstream targeting cassette (UTC) was adapted from constructs built for an earlier strategy (see Appendix A) and therefore the cloning process follows a circuitous route. pNeoXTR[129] was acquired from Addgene (69157). Primer 3 and Primer 4-loxP were used to amplify a product containing the PGK promoter, the EM7 promoter, the Neo/Kan resistance gene, and the bGH poly A signal. Primer 4 adds a loxP site upstream

of the PGK promoter to produce pLPN (lox-PGK-Neo). A PiggyBac Transposase 3' Recognition sequence (not relevant to this study) was added by amplification from pScarlessHD-DSRed (Addgene 64703, generously made public by Kate O'Connor-Giles) with Primer 5 and 6, which added MluI and NsiI sites, respectively. Restriction cloning was used to incorporate this fragment to produce pLPNP (lox-PGK-Neo-PB3').

A region surrounding Cons1 was amplified from *J. jaculus* gDNA using primers 7 and 8 and ligated into pGEMT-easy by TA cloning with a Quick Ligation Kit. The resulting plasmid was linearized by PCR using primers 9 and 10, which added overhangs appropriate for Gibson cloning. The loxP-PGK-EM7-Neo-PiggyBac3' cassette was amplified from pLPNP using primers 11 and 12. Gibson cloning was used to combine these fragments to produce a plasmid with jerboa sequence followed by the LPNP cassette (pJLPNP). Sequencing determined several errors in this plasmid, so Gibson cloning was simultaneously used to remove the PiggyBac arm, fix these errors, and add a second loxP site immediately following the bGH polyA signal. Four fragments were amplified from pJLPNP using primers 13 and 14, 15 and 16, 17 and 18, and 19 and 20. These were then combined by Gibson assembly to produce a plasmid with jerboa sequence surrounding Cons1 followed by loxP-PGK-EM7-Neo-bGHpolyA-loxP (pJerboa-floxedPGK-Neo).

Primers 21 and 22 were used to amplify jerboa sequence from *J. jaculus* gDNA, adding an overlap so that this amplicon could be combined with that produced by Primers 23 and 24 which amplified mouse sequence from *M. musculus* gDNA. Overlap extension PCR was subsequently cloned into the pGEMT-easy backbone. This plasmid was linearized by PCR using Primers 25 and 26 and combined with the floxedPGK-Neo

fragment by amplifying from pJerboa-floxedPGK-Neo with Primers 20 and 28 to produce pJMfloxPN.

Finally, to add a second jerboa arm and to add the required I-CeuI recognition site, jerboa sequence was amplified from *J. jaculus* gDNA with Primers 29 and 30, while pJMfloxPN was linearized with Primers 31 and 32. These two fragments were combined by overlap extension PCR to produce UTC, which contains 178bp of Jerboa sequence just upstream of the mouse alignment of Cons1 (pink), an I-CeuI recognition site (turquoise), 154bp of mouse sequence just upstream of the mouse alignment of Cons1 (pink), the loxP-PGK-EM7-Neo/Kan-bGHpolyA-loxP casette (yellow) and 206bp of jerboa sequence including parts of Cons1.

TATGCCAAGTTTTCGCCTGAAGCACATTTTTACAGCTGAGTAATAAAT
CCCAGAAAAACATGCCTTATGATGGAAACATTGAACTCGTCTGTTCCAACAG
GACCACTGTGGCTGCGGAGATGAGGAACATTCTTATCAAAGAAGGAGGGAA
AAAGGAAGAAGAAGGGACTGGGTAAATCGCTACCTTAGGACCGTTATAGTTA
GAGGTGCTTGTGACAGAAACAGTAAACTTCGTCTGTCCCTAAAGAGCCATTG
TGGCCATGGGGATGGATGGAAGAACATTTTAATCAAAGAAAGTTGGGAAAG
AGAGACTTGCTAGGTGAATGCCCCAGAGGCAGTGCTTTGGGAGTGACGATTT
ATAACTTCGTATAATGTATGCTATACGAAGTTATCAGGTCGTCGAAATTCTAC
CGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAG
CCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCC
ACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCC
ACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGT
CGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGAT
GGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCA
ATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGT
CCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGT
CCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCT
GTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTGCAGCCTGTTGACAATTAA
TCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAA
CCATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGC
TTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGC
TCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGT
CAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCG
GCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTG
TCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGG
ATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGAT

GCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACC
AAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGT
CGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACT
GTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGACC
CATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGG
ATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCG
TTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCT
TCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTAT
CGCCTTCTTGACGAGTTCTTCTGAGGGGATCAATTCTCTAGAGCTCGCTGATC
AGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATG
AGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCT
GGGGATGCGGTGGGCTCTATGGCATAACTTCGTATAATGTATGCTATACGAA
GTTATTATTTACCCCTGGAGGCAGCgttgTTGATAATGAGGAATTTGAGCTTTTC
GCCCCAAATCAGCTCTTTAATTTTTGTTTTTTTCTTGCAGGGTCATTTTTATGA
CACTTGAGTTTTCTTCACAATGAAAATACACGCTTGACAAGGGGGACGTGAG
AGTGATGGAGAGGCATGGAGCTATTTTTATGCCCGTCTTCTTTACAAGCCG

Table 1.7 | Cloning Primers

Name	Sequence
1	CCCTTCTAGAAGCGATGGGCGTC
2	CTGCTTGGCTTACTCCCCTGG
3	CTGGTTCTTTCCGCCTCAGAAG
4	ATAACTTCGTATAATGTATGCTATACGAAGTTATCAGGTCGTCGAAATTCTACC
5	TTACGCGTGATCCTAAAAGTTTTGTTACTTTATAGAAG
6	TTATGCATTTAACCCTAGAAAGATAATCATATTGTGAC
7	GACACTTGAGTTTTCTTCACAATG
8	AAACTGAGGCATAGAGTGAG
9	ctagggttaaTGTCTGGGAGACTCTAGAAG
10	gcggccgccaTTATGAGGATCAAGATAGCAG
11	atcctcataaTGGCGGCCGCGGGAATTC
12	ctcccagacaTTAACCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCG
13	GCCGCCATTATGAGGATCAAGATAGCAGT
14	ATAACTTCGTATAATGTATGCTATACGAAGTTATGCGTTGGATGCATAGCTTGAGTA
15	actgctatCTTGATCCTCATAATGGCGGC
16	AAGGAGCAAAGCTGCTATTGGCCGC
17	GCGGCCAATAGCAGCTTTGCTCCTT
18	TGCAATCCATCTTGTTCAATGGCCG
19	CGGCCATTGAACAAGATGGATTGCA
20	ATAACTTCGTATAGCATACATTATACGAAGTTATGCCATAGAGCCCACCGCATCCCCAG
25	GCATAACTTCGTATAATGTATGCTATACGAAGTTATTATTTACCCCTGGAGGCAGCgttg
26	GCATACATTATACGAAGTTATAAATCGTCACTCCCAAAGCACTGC
28	TTGGGAGTGACGATTTATAACTTCGTATAATGTATGCTATACGAAGTTATCAGGTCGTCG
29	TATGCCAAGTTTTCGCCTGAAGCAC
30	TAACTATAACGGTCCTAAGGTAGCGATTTACCCAGTCCCTTCTTCTTCCTTTTTCC
31	CGGCTTGTAAAGAAGACGGGC
32	CGCTACCTTAGGACCGTTATAGTTAGAGGTGCTTGTGACAGAAACAGTAAAC

Nanopore sequencing and subsequent consensus sequence analysis to confirm the identity of the full chimeric BAC was graciously performed by Cong Dinh of Dr. Rachel Dutton's lab.

Mouse ESC work

All ESC work was performed by the Transgenic Animal Core at University of Michigan. The chimeric BAC was linearized by NotI digestion and purified. ESCs were electroporated with the chimeric BAC and selected on G418. Clones were expanded and DNA was extracted and sent to UCSD for screening. The correctly identified clone was further expanded and a chromosome spread was performed to characterize proportion of euploidy. Clones were injected into blastocysts and implanted in pseudopregnant females.

Screening

G418 resistant clones were screened by PCR for the presence of jerboa sequences within the *Msx2* ROI, using primers in Table 1.8. PCR was performed using Bioline MyTaq Red MasterMix as described above.

Table 1.8 | Endpoint Screening Primers for Jerboa *Msx2* ROI Presence

Name	Sequence
Jac Insert1 F	CTTCTTCCATAACCCAGGTTGGGG
Jac Insert1 R	TTCTAGTCCCCAGTACCCATGTAAAACC
Jac Insert2 F	CCTCGGAAGTACCTTCCAGTGCT
Jac Insert2 R	GGGAAGACGTCGGGATTCAGAGAAG
Jac Insert3 F	CTTTGGGGACATCCTTTCGTAGACCG
Jac Insert3 R	cCGGAACTTTAGGCAAGCAGAGT
Jac DS F	GGGAAAGGGCGACAAAAACCC
Jac DS R	AGCTCCCGACGCCCATC
Jac US F	GAGGAATTTGAGCTTTTCGCCCCAA
Jac US R	GTCCCCCTTGTCAAGCGTGTATT
Msx2 gene L2	ACAGCTGTGTGGTTGTTTATGG
Msx2 gene R2	TATCTTCTCCAGGGTGACCTGT

Clones were further screened by qPCR in technical triplicate to assay zygosity of mouse and jerboa *Msx2* ROI at multiple locations as described. Primers are listed in Table 1.9. As qPCR is highly sensitive to changes in salt and protein concentration and therefore susceptible to variation according DNA extraction protocols, it was not possible to use an outside control with known zygosity of the *Msx2* ROI. Furthermore, all available samples were G418 resistant and therefore presumed to be modified in some way. It was therefore impractical to use any single clone, even one lacking jerboa regions, as a control. For this reason, we established a protocol whereby each individual clone is compared to the average of all other clones included in an experiment. This method reliably identified differences between DNA concentration in blinded tests using diluted mouse and jerboa DNA.

The comparison is laid out below:

1. Ct values are averaged among triplicates to produce AvgCt
2. Control AvgCt (using primers at an unrelated location) is removed from Experimental AvgCt to reveal DeltaCt
3. The DeltaCq values are averaged across all individuals within an experiment for a given Experimental primer set to give the AvgDeltaCt

4. The DeltaCt for each sample is compared to the AvgDeltaCt to produce the Delta-deltaCt

5. 2^Delta-deltaCt is an approximation of the relative amount of the experimental amplicon compared to the average amount (between all samples in an experiment) as normalized to the control amplicon

Table 1.9 | qPCR Primers for Loss of Native Allele Assay

Name	Sequence	Test for?
Control F1	AACACCGATCAATAGCGAGAAA	Control Seq (Bmp4)
Control R1	CAAACTTGATCTTTCGGACCTG	Control Seq (Bmp4)
Control F2	CCCAAATCAGATAGCCTCCA	Control Seq2 (Bmp4)
Control R2	GCTGAAGGTCCGAAGTGAAG	Control Seq2 (Bmp4)
Mouse F1	CCAGCAAAGGAAAGCCTCATTGAC	Mouse *Msx2* ROI
Mouse R1	GGGAGCTAACTTGCTTTGCTGCTG	Mouse *Msx2* ROI
Mouse F2	GTGCATGGTTGAGTAAGAGAATGTGAGGG	Mouse *Msx2* ROI
Mouse R2	CCATCTCCAACATGAGACCAGGC	Mouse *Msx2* ROI
Mouse F3	TTGTTCCTCCCCCAATTCTTTACCC	Mouse *Msx2* ROI
Mouse R3	GGGGTGTGGGAGATTGAGAGAAGG	Mouse *Msx2* ROI
Mouse F4	TGCCTAGGGAGGCTAGAACAG	Mouse *Msx2* ROI
Mouse R4	TGTGCATCTGGGTATCAGGAG	Mouse *Msx2* ROI
Mouse F5	GCAAGGGATGATTTTCCTGATGGG	Mouse *Msx2* ROI
Mouse R5	GAGGCTCTGGGTTCCATCTGTA	Mouse *Msx2* ROI
Jerboa F1	CTTCTTCCATAACCCAGGTTGGGG	Jerboa *Msx2* ROI
Jerboa R1	TTCTAGTCCCCAGTACCCATGTAAAACC	Jerboa *Msx2* ROI
Mouse HA F1	AGAGACAGTCCCTGCTCCAA	Mouse Homology Arm
Mouse HA R1	GAGCGGTGGGTAGAAGTGAG	Mouse Homology Arm

Appendix A

Our initial strategy to produce a large homologous replacement of the *Msx2* ROI relied on Recombinase Mediated Cassette Exchange (RMCE). In brief, heterologous lox sites are introduced into the mouse genome at either end of the region to be replaced and also into the jBAC at either end of the region to be replaced. Addition of Cre recombinase results in recombination between matching lox sites, exchanging the mouse genomic region for the homologous region from the jBAC.

While this strategy has been demonstrated to be effective, replacing up to 120kb at a time in humanized mouse models, it requires extensive manipulation and selection in ESCs[44]. Typically, each lox site is introduced independently into the genome in successive rounds of targeting and selection before the replacement is introduced, and the final rounds of selection are performed. Extensive manipulation and passaging of ESCs has been shown to cause alterations to epigenetic landscape, increase aneuploidy, and decrease germ cell contribution to ESC-derived chimeras[83].

A second deficit of this strategy is the retention of unwanted insertions at either end of the final replacement. Each end will retain a 34 bp lox 'scar' and at least one end will require the presence of a selection cassette. This selection cassette could theoretically be bounded by an additional set of recombinase sites, for instance FRT sites recognized by Flippase, in which case one end of the replacement would ultimately have a 68bp scar.

Our modified RMCE strategy therefore aimed to improve upon existing methods by eliminating a step in ESCs and by designing the exchange such that one end of the replacement would be seamless and scarless. To eliminate one round of ESC targeting, we designed our strategy to delete the full mouse *Msx2* ROI and simultaneously replace it

with a single small cassette containing positive and negative selection markers flanked by the appropriate heterologous lox sites. By integrating the lox sites in a single step, rather than introducing each lox site independently, we could reduce the degree of ESC manipulations, passaging, and selection.

In order to eliminate one lox scar, we devised a configuration that would result in a seamless transition between mouse and jerboa at the upstream boundary of the replacement, leaving only a 34 bp *lox2272* scar at the downstream boundary. In order to achieve this seamless transition, we employed the PiggyBac Transposase system, which can facilitate scarless removal of transgenes at TTAA sites[130]. By engineering our strategy such that the transition between mouse and jerboa happened at a naturally occurring TTAA site in mouse, we anticipated being able to remove all selection markers and one of the two lox sites.

The strategy in full is detailed below.

Theoretical RMCE Strategy for *Msx2* ROI Replacement

First, the mouse genome is targeted by dual gRNAs, each of which recognizes one end of the *Msx2* ROI. We identified three putative gRNA target sites with appropriate PAM near the upstream boundary of the replacement (U1, U2, U3) and two putative gRNA sites with appropriate PAM near the downstream boundary (D1, D2). These predicted target sequences (gRNA spacers) were generated by IDT and cloned into the pX330 vector, which contains a U6 promoter driving transcription of the full gRNA scaffold and the strong mammalian CAG promoter driving transcription of *SpCas9*[71].

After several attempts to test cutting efficiency in vitro, we tested the capability of predicted gRNAs to generate a deletion of the full *Msx2* ROI in cell culture. MEFs were transfected with two modified pX330 plasmids, each containing one gRNA target sequence. The capability of pairs of gRNAs to generate the full deletion were validated by PCR amplification using primers external to the expected deletion. As the full mouse *Msx2* ROI is 25.8kb, a length impossible to amplify in the lab with a typical polymerase, only the deletion results in the production of an amplicon of known size. Any pair of an upstream gRNA with a downstream gRNA was sufficient to produce a deletion of expected size, whereas amplification from cells transfected with a gRNA targeting EGFP failed. Interestingly, cutting with the U3 gRNA routinely produced the brightest bands, regardless of which downstream gRNA it was paired with. Additionally, the U3 gRNA fortuitously overlay a TTAA site in mouse, which could be used for seamless PiggyBac excision. We therefore developed a targeting strategy using the U3 upstream and D1 downstream gRNAs.

We designed a selection cassette (Figure 1.9a) that contains a 2058bp homology arm that immediately abuts the predicted cut site of the U3 gRNA. External to the upstream homology arm, the mammalian PGK promoter drives transcription of diphtheria toxin A (DTA) to reduce the likelihood of random integration. Immediately following the homology arm is the 5' recognition sequence for the PiggyBac excisionase (PB 5'), such that the TTAA left behind by the PiggyBac system falls precisely at the location of the endogenous mouse TTAA found in the U3 target site. Following the PB 5' sequence is a *loxP* site, and the mammalian CMV promoter driving transcription of a fusion product between the hygromycin-B resistance gene and a truncated version of the

herpes simplex virus' thymidine kinase (Hyg-TK). The Hyg-TK fusion product provides both resistance to hygromycin and also sensitivity to ganciclovir, making this single fusion gene a positive and negative selection marker. The Hyg-TK gene is followed by a *lox2272* site, which has been shown to recombine efficiently with other *lox2272* sites, but does not appear to recombine with *loxP*[131], and a 2630 bp downstream homology arm that begins 16 bp away from the predicted cut site of the D1 gRNA. This selection cassette is designed to fully eliminate the target sequence of the D1 gRNA and truncate the target sequence of the U3 gRNA such that it is missing 4bp adjacent to the PAM.

When gRNAs, Cas9, and selection cassette are added to ESCs, we anticipated that the selection cassette would be incorporated by HDR, and the resulting cells could be selected on hygromycin (Figure 1.9a).

After selection and screening, ESCs wherein the Hyg-TK cassette replaced the endogenous *Msx2* ROI locus would be electroporated with the Cre recombinase and a modified jBAC carrying appropriate lox sites and selection (Figure 1.9b). We designed the modifications to the jBAC such that at the downstream boundary of the jerboa *Msx2* ROI has a *lox2272* site in the same orientation as the *lox2272* site in the modified mouse ESCs. In order to introduce this *lox2272* site, we would necessarily need to include a selection marker suitable for cloning, such as an ampicillin resistance gene. By placing the ampicillin resistance marker downstream of the *lox2272* site, we ensure that ampicillin resistance will not be transferred in the Cre-recombination step. At the upstream boundary of the jerboa *Msx2* ROI, the modified jBAC has a *loxP* site (in the same orientation as the *loxP* site in the mouse ESCs) followed by a neomycin resistance gene under the control of a dual promoter system. In this dual promoter system,

neomycin can be driven by either the synthetic bacterial EM7 promoter, to facilitate cloning in bacterial cells, or by the mammalian PGK promoter to facilitate screening in ESCs. To ensure successful resistance in mammalian cells, the neomycin resistance coding sequence is followed by the bovine growth hormone polyA signal. Finally, the neomycin resistance gene is followed by the 3' recognition sequence of the PiggyBac excisionase (PB 3') such that the TTAA sequence replaces an ATAA sequence in jerboa that is homologous to the TTAA identified in mouse (Figure 1.9b).

We anticipated that when the modified jBAC was added to hygromycin-resistant ESCs with a Cre recombinase, the Hyg-TK selection cassette would be exchanged for the full jerboa *Msx2* ROI with neomycin resistance (Figure 1.9c). Resulting cells could be selected on G418 for acquisition of neomycin resistance, indicating successful insertion of the jerboa sequence, and on ganciclovir for loss of the thymidine kinase. This negative selection allows for differentiation between targeted exchange and random insertion, as cells with random insertions of the neomycin resistance would retain sensitivity to ganciclovir.

The resulting ESCs are expected to have the jerboa *Msx2* ROI in place of the mouse *Msx2* ROI, a *lox2272* 'scar', and a neomycin resistance cassette with adjacent *loxP* site flanked by PiggyBac recognition sequences (Figure 1.9c). Application of the PiggyBac excisionase removes all selection markers, leaving cells with a single 34 bp *lox2272* 'scar' at the downstream end of the *Msx2* ROI (Figure 1.9d).

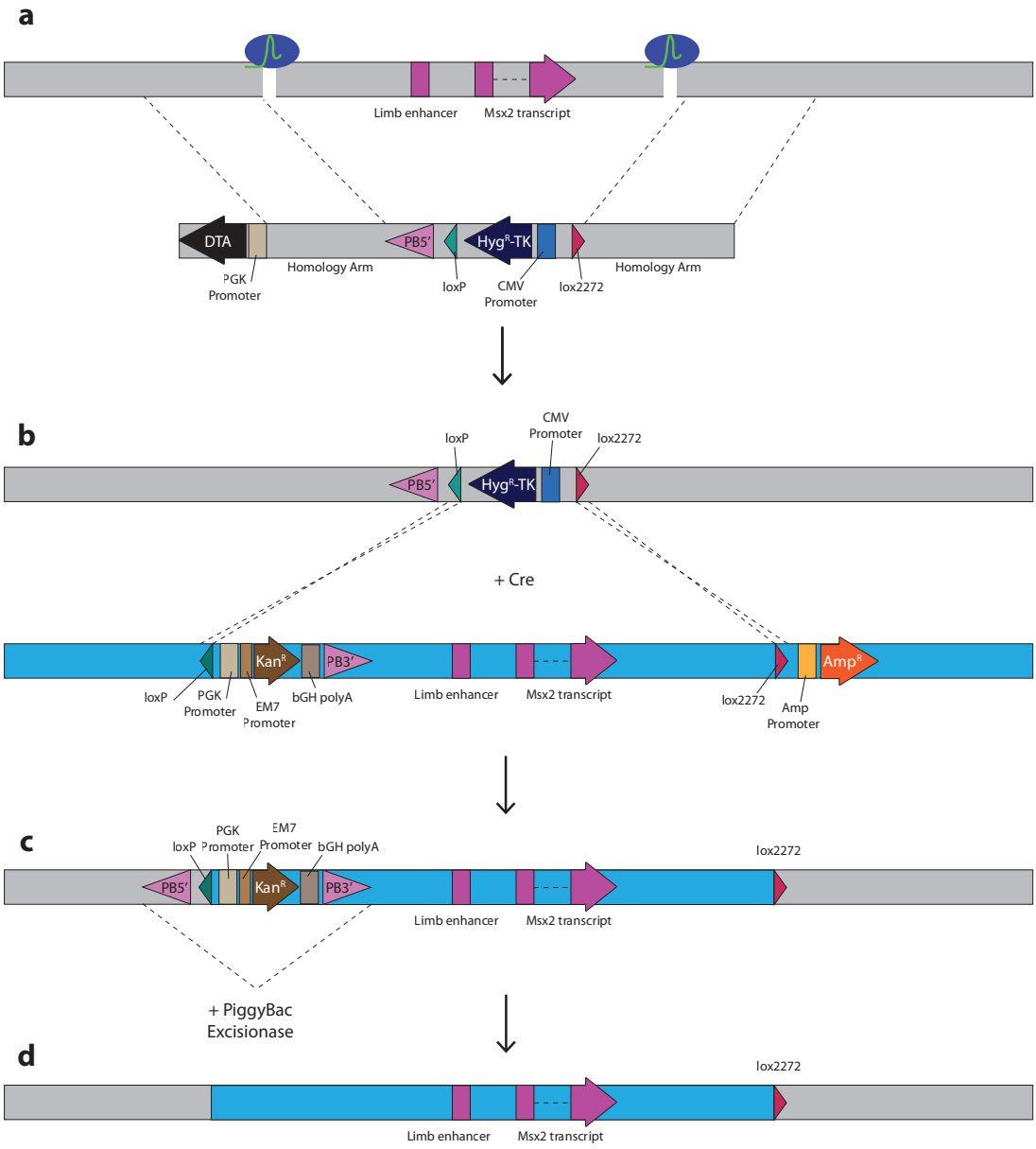


Figure 1.9 | RMCE Strategy for *Msx2* ROI Replacement. a. The mouse genome (grey) is cut with two gRNAs, allowing for simultaneous deletion of the mouse *Msx2* ROI and replacement with a selection cassette bounded by two heterologous *lox* sites. **b.** The resulting cells are electroporated with a Cre recombinase and a jBAC (blue) modified such that it contains the appropriate *lox* sites and selection. Cre facilitates recombination between paired *lox* sites. **c.** The resulting cells contain the jerboa *Msx2* ROI, an upstream selection cassette, and a downstream *lox2272* site. The selection cassette can be seamlessly excised using the PiggyBac system. **d.** The final clones have the full mouse *Msx2* ROI replaced with the jerboa *Msx2* ROI and retain a single *lox* scar. (Grey- mouse sequence; blue- jerboa sequence; PB5' and PB3'- PiggyBac 5' and 3' recognition sequences, respectively; Hyg^R-TK- hygromycin resistance gene fused to thymidine kinase; Kan^R- kanamycin resistance; bGH – bovine growth hormone; DTA- diphtheria toxin A)

Unsuccessful integration of the Hyg-TK selection cassette

The U3 and D1 pX330 plasmids were electroporated into ESCs by the UCSD Transgenic Mouse Core in concert with the Hyg-TK selection construct and cells were selected on hygromycin. PCR to assay for the presence of the Hyg-TK cassette revealed that the full cassette (including lox sites, PB 5', and Hyg-TK fusion gene, but not homology arms) was detectable in at least two thirds of all clones assayed.

PCR across the upstream junction of the insertion routinely failed, whereas primers that spanned the downstream junction were able, in some cases, to amplify bands of the correct size. Because of the difficulty in amplifying across junctions, I tried to determine whether the full insertion, complete with homology arms, was present in clones that were hygromycin resistant. To do this, I designed primers at progressively more distal locations in the homology arms. Intriguingly, of 60 clones assayed, only 19 had any regions extending into the downstream homology arm. More surprisingly, only one had detectable amplification from the upstream homology arm. Even the single clone that preserved a portion of the upstream homology arm did not preserve the full homology arm.

There are several factors that could potentially account for the failure to fully integrate the Hyg-TK selection cassette into the ESC genome. First, it should be noted that the diphtheria toxin was present at the end of the upstream homology arm. It seems plausible that the presence of this toxin did successfully select against random integration of the upstream portion of the construct, resulting in fewer clones that had the upstream homology arm.

Interestingly, a subset of clones did have the correct downstream junction, suggesting targeted, rather than random integration. These clones nevertheless did not contain the correct upstream junction (or indeed the full upstream homology arm), suggesting that in some cases, targeting occurred only at the downstream boundary of the *Msx2* ROI. It is possible that differences in the efficiency of DSB generation between the U3 and D1 gRNAs is partially responsible for this phenomenon. If the D1 gRNA generated DSBs earlier or with higher frequency than the U3 gRNA, then it is possible that integration at the downstream end of the *Msx2* ROI would be favored, eliminating the D1 target site before the U3 site was cut. To avoid this possibility, future attempts to perform RMCE for this purpose should quantitatively assay gRNAs using Inference of CRISPR Edits (ICE)[132] to determine the relative efficiency of DSB generation using each gRNA.

It is also possible that HDR was improved at the downstream junction because of the relative length of the downstream homology arm. Whereas the upstream homology arm was 2058 bp, the downstream homology arm was 2630 bp. Studies have consistently shown that larger homology arms improve the rate of HDR[84–86]. In the future, attempts to perform large replacements using this technology could make use of more extensive homology arms.

While we ultimately chose a different direction to pursue the goal of creating large interspecies genomic replacements, this technique may be of future use, particularly in studies that aim to make mouse models exploring the phenotypic consequences of the same genomic region from more than one animal. The replacement of the endogenous mouse locus with a *loxP-lox2272* cassette results in the production of an ESC line that

can be used to introduce any other DNA bounded by the same Cre-recognition sites. By producing modified BACs from multiple animals such that they carry the *loxP* and *lox2272* sites, one could theoretically streamline the process of producing many mouse models concurrently.

Acknowledgements

We would like to thank Pieter de Jong and Cris Jung at CHORI for sharing their extensive knowledge regarding recombineering, Michael Atkins at Clemson University for producing our Jerboa BAC library, Cong Dinh and Rachel Dutton for assistance with Nanopore sequencing, and Elizabeth Hughes and Thom Saunders at University of Michigan for their assistance with ESCs. In addition, we would like to thank Sonia Grunwald for her technical assistance.

Chapter 1, in part, is being prepared for submission for publication of the material. Grunwald, Hannah; Chen, Andrew; Organ, Chris; Cooper, Kimberly. The dissertation author was the primary researcher and author of this material.

Chapter 2: Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline

Abstract

A gene drive biases the transmission of one of the two copies of a gene such that it is inherited more frequently than by random segregation. Highly efficient gene drive systems have recently been developed in insects, which leverage the sequence-targeted DNA cleavage activity of CRISPR–Cas9 and endogenous homology-directed repair mechanisms to convert heterozygous genotypes to homozygosity[92,94,133,134]. If implemented in laboratory rodents, similar systems would enable the rapid assembly of currently impractical genotypes that involve multiple homozygous genes (for example, to model multigenic human diseases). To our knowledge, however, such a system has not yet been demonstrated in mammals. Here we use an active genetic element that encodes a guide RNA, which is embedded in the mouse tyrosinase (*Tyr*) gene, to evaluate whether targeted gene conversion can occur when CRISPR–Cas9 is active in the early embryo or in the developing germline. Although Cas9 efficiently induces double-stranded DNA breaks in the early embryo and male germline, these breaks are not corrected by homology- directed repair. By contrast, Cas9 expression limited to the female germline induces double-stranded breaks that are corrected by homology-directed repair, which copies the active genetic element from the donor to the receiver chromosome and increases its rate of inheritance in the next generation. These results demonstrate the feasibility of CRISPR–Cas9-mediated systems that bias inheritance of desired alleles in mice and that have the potential to transform the use of rodent models in basic and biomedical research.

Main Text

A cross between mice that are heterozygous for each of three unlinked genes must produce 146 offspring for a 90% probability that one will be a triple homozygous mutant. The likelihood decreases further if any of the three mutations are genetically linked but on opposite homologous chromosomes, because recombination events that combine alleles onto the same chromosome would be very infrequent. The cost, time and requirement for a large number of mice to obtain a few individuals of the desired genotype are therefore prohibitive for certain complex models of multigenic evolutionary traits or human diseases, such as arthritis and cancer.

Recently, CRISPR–Cas9-mediated gene drive systems were developed in *Drosophila* and anopheline mosquitoes that increase the frequency of inheritance of desired alleles[92,94,133,134]. These used genetic elements, which we refer to broadly as active genetic elements, that can carry transgenes or orthologous sequences from other species[135]. Notably, an active genetic element includes a guide RNA (gRNA) and is inserted into the genome at the precise location that is targeted for cleavage by the encoded gRNA. In a heterozygous animal that also expresses the Cas9 nuclease, the gRNA targets cleavage of the wild-type homologous chromosome. Genomic sequences that flank the active genetic element then correct the double-stranded break (DSB) by homology-directed repair (HDR), which copies the active genetic element from the donor to the receiver chromosome and converts the heterozygous genotype to homozygosity. The frequency of transmitting the active genetic element to the next generation is therefore greater than expected by random segregation of heterozygous alleles and is referred to as ‘super-Mendelian’. In addition to the potential to overcome the obstacles of

assembling complex genotypes in laboratory rodents, variations of a CRISPR–Cas9-mediated system have been proposed that might help to suppress invasive rodent populations and/or reduce the prevalence of rodent-borne disease[91,136].

Despite the high efficiency observed in insects, the approximately 790 million years of divergence since their last common ancestor with mammals presents two potential obstacles to the implementation of active genetics in mice; the frequency of DSB formation using a genetically encoded Cas9 and gRNA and/or the frequency of HDR may prevent efficient gene conversion. The alternative DSB repair pathway, non-homologous end joining (NHEJ), frequently generates small insertions and deletions (indels) that make CRISPR–Cas9 an effective means of mutating specific sites in the genome. Although HDR of CRISPR–Cas9-induced DSBs does occur in vitro and in vivo in mammalian cells and embryos, usually from a plasmid or single-stranded DNA template, NHEJ is the predominant mechanism of DSB repair in somatic cells[77,137].

To assess the feasibility of active genetic systems in mice, we designed a 'CopyCat' element[138] that differs from the genetic element used initially in insects in that it cannot self-propagate, because it encodes a gRNA but not the Cas9 protein (Fig. 2.1a). We designed our strategy to disrupt the *Tyrosinase* gene (*Tyr*), because of the obvious albino phenotype of homozygous loss-of-function mice[139] and to make use of a previously characterized *Tyr* gRNA with high activity[140]. The precise insertion of this element into the gRNA cut site in exon 4 of *Tyr* to obtain the $Tyr^{CopyCat}$ knock-in allele is shown in Fig 2.1a. In brief, the *Tyr* gRNA is transcribed from a constitutive human RNA polymerase III U6 promoter[141]. On the reverse strand of the DNA, to minimize possible transcriptional conflict, mCherry is ubiquitously expressed from the human

cytomegalovirus (CMV) immediate-early promoter and enhancer[142]. As the 2.8-kb insert disrupts the *Tyr* open reading frame, $Tyr^{CopyCat}$ is a functionally null (albino) allele that results in a red fluorescent phenotype (Fig. 2.2) and that is propagated by Mendelian inheritance in the absence of Cas9. Crossing mice that carry the $Tyr^{CopyCat}$ element to transgenic mice that express Cas9 enabled us to test whether it is possible to observe super-Mendelian inheritance of the $Tyr^{CopyCat}$ allele. For this analysis, we assessed eight different genetic strategies that use existing tools to provide spatial and temporal control of Cas9 expression in the early embryo and in the male and female germlines.

We used two 'constitutive' *Cas9* transgenic lines, *Rosa26-Cas9*[143] and *H11-Cas9*[144], which express Cas9 in all tissues that have been assessed. Each is driven by a ubiquitous and highly efficient CAG promoter and is placed in the respective *Rosa26* or *H11* 'safe harbour' locus. To track the inheritance of the chromosome that is targeted for gene conversion, we bred the *chinchilla* allele of tyrosinase ($Tyr^{c\text{-}ch}$, here simplified to Tyr^{ch}) into each *Cas9* transgenic line. Tyr^{ch} encodes a hypomorphic point mutation in exon 5 that is tightly linked to the gRNA target site in exon 4. Tyr^{ch} homozygotes or heterozygotes that also have a null allele have a grey coat colour, and the G to A single-nucleotide polymorphism can be scored with certainty by PCR followed by DNA sequencing[145] (Figure 2.3).

Female *Rosa26-Cas9;*$Tyr^{ch/ch}$ and *H11-Cas9;*$Tyr^{ch/ch}$ mice were each crossed to $Tyr^{CopyCat/+}$males to combine the gRNA and Cas9 protein in the early embryo (Fig. 2.1c). In absence of a loss-of-function mutation in exon 4 of the receiver chromosome, $Tyr^{CopyCat/ch}$ mice should appear grey (see $Cas9^{-}$;$Tyr^{CopyCat/ch}$ mice in Fig. 2.1e). However, we did not observe any grey $Cas9^{+}$;$Tyr^{CopyCat/ch}$ mice in the F_2 offspring

of either cross. Instead, all 17 of the *Rosa26-Cas9;Tyr*$^{CopyCat/ch}$ mice were entirely white. Among *H11-Cas9;Tyr*$^{CopyCat/ch}$ mice, 21 of the F_2 progeny (87.5%) were a mosaic mixture of grey and white fur, and three mice (12.5%) were entirely white (Fig. 2.1e, 2.1f and Table 2.1). The prevalence of mosaicism in the *H11-Cas9* mice, compared to the all-white mice produced by *Rosa26-Cas9*, suggests that there may be a difference in the level and/or timing of Cas9 expression driven by these two transgenes.

Our next goal was to determine what type of repair events (NHEJ mutations or gene conversions by HDR) were transmitted to the next generation. To assess inheritance in many offspring, we crossed each F_2 male *Rosa26-Cas9;Tyr*$^{CopyCat/ch}$ and *H11-Cas9;Tyr*$^{CopyCat/ch}$ mouse to multiple albino CD-1 females (Fig. 2.1d), which carry a loss-of-function point mutation in the *Tyr* exon 1 (*Tyr*^{c}, here designated as *Tyr*^{null})[139,145]. We then genotyped F_3 offspring of this cross by PCR and DNA sequencing of exon 5 to identify those that inherited the *Tyr*^{ch}-marked receiver chromosome (Fig 2.3). In the absence of gene conversion, effectively none of these chromosomes would be predicted to also carry the *Tyr*^{CopyCat} allele, because *Tyr* exons 4 and 5 are separated by only approximately 9.1 kb, and therefore have a very low estimated rate of recombination.

Tyr$^{ch/null}$ mice should appear grey because of the partial activity of the hypomorphic *Tyr*^{ch} allele. However, among F_3 offspring with this genotype, 100% in the *Rosa26-Cas9* lineage and 90.4% in the *H11-Cas9* lineage were completely white, indicating frequent transmission of a CRISPR–Cas9-induced loss-of-function mutation on the receiver chromosome and consistent with the primarily albino coat colour of the F_2 parents (Fig. 2.1d and Table 2.2). If the induced null alleles resulted from inter-homologue HDR copying the *Tyr*^{CopyCat} allele from the donor to the *Tyr*^{ch}-marked

receiver chromosome, these white animals should also express the fluorescent mCherry marker. However, in these experiments none of the F_3 offspring that inherited the receiver chromosome in either the *Rosa26-Cas9* or *H11-Cas9* lineages expressed mCherry. PCR amplification of *Tyr* exon 4 confirmed that the $Tyr^{CopyCat}$ element was not present in white $Tyr^{ch/null}$ F_3 progeny (Table 2.2).

The different propensities to yield a full albino or mosaic coat colour pattern in the F_2 generation of *Rosa26-Cas9* and *H11-Cas9* lineages were consistent with differences in the number of unique NHEJ mutations that we identified on receiver chromosomes in individuals of each genotype. Sequenced PCR products from *Rosa26-Cas9;*$Tyr^{CopyCat}$ F_2 tails—which are somatic tissues that consist of both ectodermal and mesodermal derivatives—and from individual F_3 outcross offspring (representing the germline) typically exhibited only two unique NHEJ mutations, suggesting that many of these Cas9-induced mutations may have been generated in embryos at the 2–4-cell stage (average 2.4 alleles among offspring of five families). By contrast, *H11-Cas9;*$Tyr^{CopyCat}$ F_2 tails and F_3 offspring had significantly more unique NHEJ mutations (average 4.6 alleles in five families; two-tailed Student's *t*-test, $P = 0.041$), consistent with the hypothesis that Cas9 is expressed at a later embryonic stage and/or at lower levels in this lineage (Fig. 2.4 and Table 2.3).

We considered two explanations for the observation that $Tyr^{CopyCat}$ was not copied to the receiver chromosome in the early embryo. The first possibility is that homologous chromosomes are not aligned for inter-homologue HDR to repair DSBs. Second, the DNA repair machinery in somatic cells typically favours NHEJ over HDR[77,137]. A possible solution to overcome both potential obstacles is to restrict CRISPR–Cas9

activity to coincide with meiosis in the developing germline. During meiosis, recombination of the maternal and paternal genomes is initiated by the formation of DSBs that are repaired by exchanging regions of homologous chromosomes that are physically paired during meiosis I[146]. Indeed, the molecular mechanisms of NHEJ are actively repressed to favour HDR during meiosis in many species, including mice[147].

To test the hypothesis that CRISPR–Cas9 activity will convert a heterozygous active genetic element to homozygosity during meiosis, we designed a crossing scheme to initiate Cas9 expression during germline development in *Tyr*$^{CopyCat/ch}$ mice. As no currently available transgenic mice express Cas9 under direct control of a germline-specific promoter, we crossed mice with a conditional *Rosa26*- or *H11-LSL-Cas9* transgene, each with a *loxP*-Stop-*loxP* (LSL) site preceding the Cas9 translation start site[143,144], to *Vasa-Cre* (also known as *Ddx4-Cre*) or *Stra8-Cre* germline transgenic mice. *Vasa-Cre* is expressed later than the endogenous *Vasa* transcript in both male and female germ cells[148], whereas *Stra8-Cre* is limited to the male germline and is initiated in early-stage spermatogonia[149]. Although oogonia and spermatogonia are pre-meiotic, and spermatogonia are in fact mitotic, we reasoned that Cre protein must first accumulate before *Cas9* can be expressed from the *LSL-Cas9* conditional allele. The possible time delay may require initiation of *Cre* expression before the onset of meiosis so that Cas9-induced DSBs can be resolved by inter-homologue HDR before segregation of homologous chromosomes at the end of meiosis I. We generated each combination of these *Cre* and conditional *Cas9* lines in case the timing or levels of Cas9 expression were critical variables in these crosses. We also assessed males and females of

the *Vasa* strategies in case there were sex-dependent differences in animals that inherit the same genotype.

Males heterozygous for $Tyr^{CopyCat}$ and the *Vasa-Cre* transgene were crossed to females homozygous for both the Tyr^{ch} allele and one of the two conditional *Cas9* transgenes (Fig. 2.5a). We avoided the reverse cross using female *VasaCre* mice, because Cre protein maternally deposited in the egg[148] might induce recombination of the conditional *Cas9* allele and induce mutations in the early embryo similar to what we observed in the experiments using constitutive *Cas9* transgenes. Introducing the *Vasa-Cre* transgene by inheritance from the male instead resulted in most offspring that were entirely grey, owing to the $Tyr^{CopyCat/ch}$ genotype, and a few mosaic animals (Table 2.4). The presence of mosaicism suggests that this conditional approach to restrict *Cas9* expression to the germline resulted in some spurious cleavage of the *Tyr* locus in somatic tissues.

We first tested whether Cas9 in the female germline could promote copying of the $Tyr^{CopyCat}$ element onto the receiver chromosome by crossing F_3 female mice of each *Vasa-Cre* lineage to CD-1 (Tyr^{null}) males. In each cross, we identified F_4 offspring that inherited the Tyr^{ch}-marked chromosome (Fig. 2.5b). As in the crosses to assess the effects of embryonic *Cas9* expression, we expected that $Tyr^{ch/null}$ mice without a loss-of-function mutation in exon 4 of the receiver chromosome would be grey. Mice with a CRISPR–Cas9-induced NHEJ mutation in exon 4 should be white. Mice carrying a CRISPR–Cas9-induced mutation that was repaired by inter-homologue HDR should not only be white, but also show expression of mCherry (red fluorescence) owing to transmission of the mCherry-marked $Tyr^{CopyCat}$ active genetic element.

Figure 2.5b summarizes the results of these crosses that demonstrate gene conversion upon Cas9 expression in the female germline. In contrast to early embryonic expression of *Cas9*, we observed that the $Tyr^{CopyCat}$ transgene was copied to the Tyr^{ch}-marked receiver chromosome in both *Vasa-Cre;Rosa26-LSL-Cas9* and *Vasa-Cre;H11-LSL-Cas9* lineages. However, the observed efficiency differed between genotypes; three out of five females of the *Vasa-Cre;Rosa26-LSL-Cas9* lineage and all five females of the *Vasa-Cre;H11-LSL-Cas9* lineage transmitted a Tyr^{ch}-marked chromosome that contained a $Tyr^{CopyCat}$ insertion to at least one offspring. Although there was considerable variation among females with the same genotype, the highest observed frequency of gene conversion (72.2%) within a germline produced 13 out of 18 Tyr^{ch} offspring with a $Tyr^{CopyCat}$ insertion in the *Vasa-Cre;H11-LSL-Cas9* lineage (Fig. 2.5b, Table 2.5). The probability of obtaining an animal with this genotype by natural meiotic recombination mechanisms is very low (4.7×10^{-5}) owing to ultra-tight linkage between the $Tyr^{CopyCat}$ and Tyr^{ch} alleles. Although it seems probable that inter-homologue HDR of Cas9-induced DSBs uses the same DNA-repair machinery that is active during meiotic recombination, these copying events cannot be explained by an increased incidence of chromosomal crossover, because all animals that inherited the donor chromosome lacking the Tyr^{ch} marker expressed mCherry (Table 2.5).

In contrast to the 41 copying events that we observed out of a total of 132 $Tyr^{ch/null}$ offspring of female mice, we observed no copying in a total of 113 offspring of males in which conditional *Cas9* expression was induced by either *Vasa-Cre* or *Stra8-Cre* (Fig. 2.5b, Table 2.5). It is possible, however, that the number of families and of offspring in each family—which was limited by unexplained low male fertility—was

insufficient to detect low-efficiency copying in each of four genetic strategies. If there is indeed a difference between males and females in the efficiency of $Tyr^{CopyCat}$ copying, we can consider two potential explanations. First, despite equivalent genotypes in the male and female *Vasa-Cre* lineages, Cre, Cas9 and/or gRNA may not be well-expressed in the male germline. However, the high frequency of white $Tyr^{ch/null}$ offspring suggests that DSB formation is very efficient in males. Second, spermatogonia continually undergo mitosis and produce new primary spermatocytes throughout the life of a male in mammals[150]. By contrast, oogonia directly enlarge without further mitosis to form all of the primary oocytes in the embryo[151]. The difference in the observed efficiency of inter-homologue HDR between females and males at this locus may therefore reflect a requirement for the precise timing of CRISPR–Cas9 activity to coincide with meiosis (Fig. 2.6). NHEJ indels in males may result from DSB repair that occurs before the alignment of homologous chromosomes in meiosis I. Similarly, the higher observed efficiency of inter-homologue HDR in females of the *H11-LSL-Cas9* conditional strategy may relate to the lower or delayed *Cas9* expression from the *H11* locus, compared to *Rosa26*, which was evident in the constitutive crosses (Fig. 2.1e, f and Tables 2.1, 2.3). Thus, in the *Vasa-Cre;H11-LSL-Cas9* mice, DSB formation may have been fortuitously delayed to fall within a more optimal window during female meiosis.

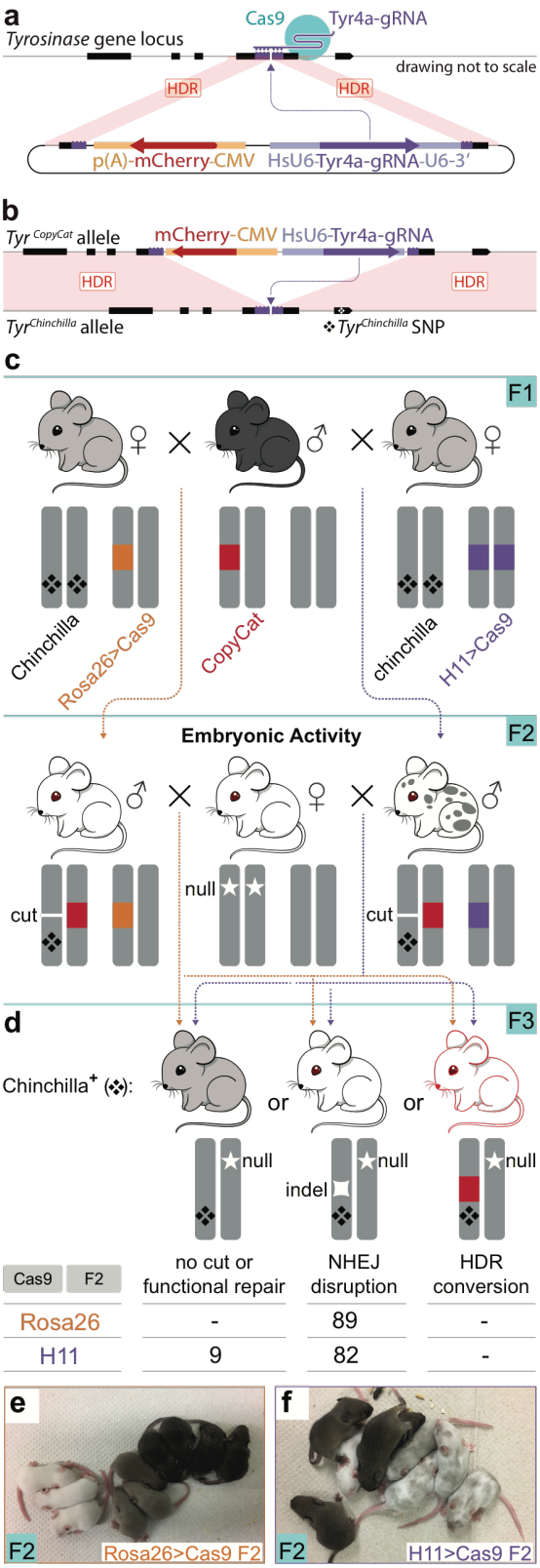
In summary, we demonstrate that the fundamental mechanism of a CRISPR–Cas9-mediated gene drive is feasible in mice. However, our comparison of eight different genetic strategies indicates that the precise timing of Cas9 expression may present a greater challenge in rodents than in insects to restrict DSB formation to a window when breaks can be efficiently repaired by the endogenous meiotic recombination machinery.

These data are therefore critical to the ongoing discussion about whether CRISPR–Cas9-mediated gene drives could be used to reduce invasive rodent populations, because it appears that both the optimism and concerns are likely to be premature. Further optimization to increase the frequency of gene conversion in both males and females and to reduce the prevalence of drive-resistant alleles (NHEJ indels that alter the gRNA target site) would be necessary to achieve rapid and sustained suppression of wild populations[152–157].

Nevertheless, the copying efficiencies that we observed here would be more than sufficient for a broad range of laboratory applications. For example, the average observed copying rate of 44% using the most efficient genetic strategy in females (*Vasa-Cre;H11-LSL-Cas9*) combined ultra-tightly linked tyrosinase mutations such that 22.5% of all offspring inherited a chromosome with both alleles, which would not be possible through Mendelian inheritance. This observed average copying rate would also be expected to increase the inheritance of a single desired allele from 50% to 72%, and the highest rate of gene conversion that we observed (72.2%) would result in an 86% frequency of transmitting a desired allele. If multiple genes could be simultaneously converted to homozygosity, such high transmission frequencies that bypass the onerous constraint of genetic linkage stand to greatly accelerate the production of rodent models for a variety of complex genetic traits.

Figures and Tables

Figure 2.1 | Embryonic Cas9 activity does not copy the *Tyr*^{CopyCat} allele from the donor to the receiver chromosome. a. Knock-in strategy using the *Tyr*^{CopyCat} targeting vector to produce the genetically encoded *Tyr*^{CopyCat} element. The U6-Tyr4a gRNA (*Tyrosinase* exon 4 gRNA a) and CMV-mCherry were inserted into the cut site of the *Tyr4a* gRNA by HDR after CRISPR–Cas9 DSB formation targeted by the *Tyr4a* gRNA. **b.** The genetically encoded *Tyr*^{CopyCat} element, when combined with a transgenic source of Cas9, is expected to induce a DSB in the *Tyr*^{ch}-marked receiver chromosome, which could be repaired by inter-homologue HDR. **c**. Breeding strategy to combine *Tyr*^{CopyCat} with a constitutive *Cas9* transgene followed by a cross between *Tyr*^{CopyCat} and *Tyr*^{null} mice. **d**. Summary of the F_3 cross offspring of five independent families for each *Rosa26-Cas9* and *H11-Cas9* genotype. **e**. A representative litter of six *Rosa26-Cas9* F_2 litters. Black mice did not inherit *Tyr*^{CopyCat}. Grey mice inherited *Tyr*^{CopyCat} but not *Cas9*. White mice inherited both transgenes. **f**. A representative litter of five F_2 litters in which all offspring inherited *H11-Cas9*. The mosaic mice also inherited *Tyr*^{CopyCat}.


a
Tyrosinase gene locus
Cas9
Tyr4a-gRNA
drawing not to scale
HDR
HDR
p(A)-mCherry-CMV
HsU6-Tyr4a-gRNA-U6-3'
b
Tyr CopyCat allele
mCherry-CMV HsU6-Tyr4a-gRNA
HDR
HDR
Tyr Chinchilla allele
Tyr Chinchilla SNP
c
F1
Chinchilla
Rosa26>Cas9
CopyCat
chinchilla
H11>Cas9
Embryonic Activity
F2
cut
null
cut
d
F3
Chinchilla+ (❖):
or
or
null
indel
null
null
Cas9
F2
no cut or functional repair
NHEJ disruption
HDR conversion
Rosa26
-
89
-
H11
9
82
-
e
F2
Rosa26>Cas9 F2
f
F2
H11>Cas9 F2

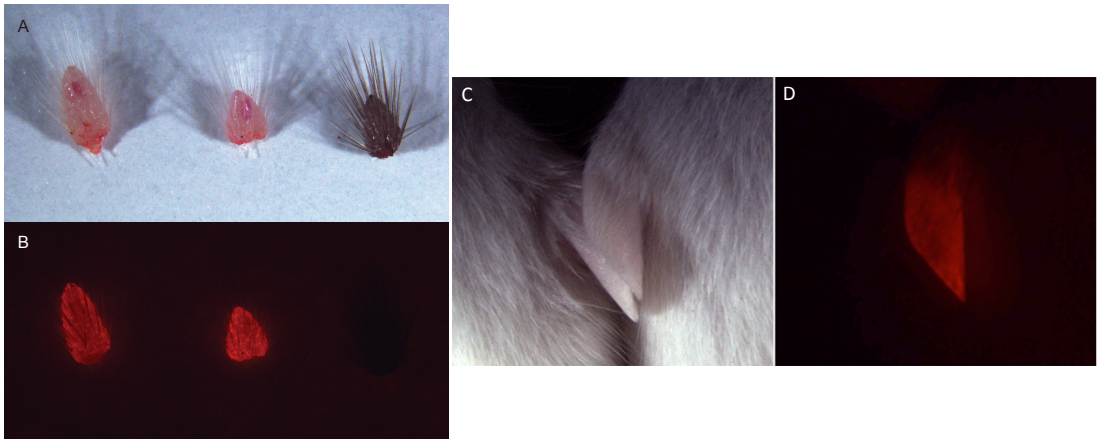


Figure 2.2 | mCherry fluorescence marks *Tyrosinase*^{CopyCat} tails and ears. a and b. Two tail tips from F_2 mice of the *Rosa26:Cas9* lineage with *Tyr*^{CopyCat} (left and middle) and one from a mouse that did not inherit the *Tyr*^{CopyCat} transgene (right; *Tyr*$^{WT/ch}$). mCherry is visible only in tails with an allele of the *Tyr*^{CopyCat} transgene. **c and d.** F_3 offspring of the constitutive *Rosa26:Cas9* lineage. The right mouse inherited the original *Tyr*^{CopyCat} transgene with mCherry fluorescence in an outcross to CD-1 *Tyrosinase*^{null}. The left mouse inherited the *Tyr*^{ch}-marked target chromosome with an NHEJ mutation and no mCherry fluorescence.

Figure 2.3 | Sanger sequencing traces of *Tyrosinase* exon 5 differentiate individuals that are wild type, heterozygous, and homozygous for the *Tyrosinase*^{chinchilla} SNP.

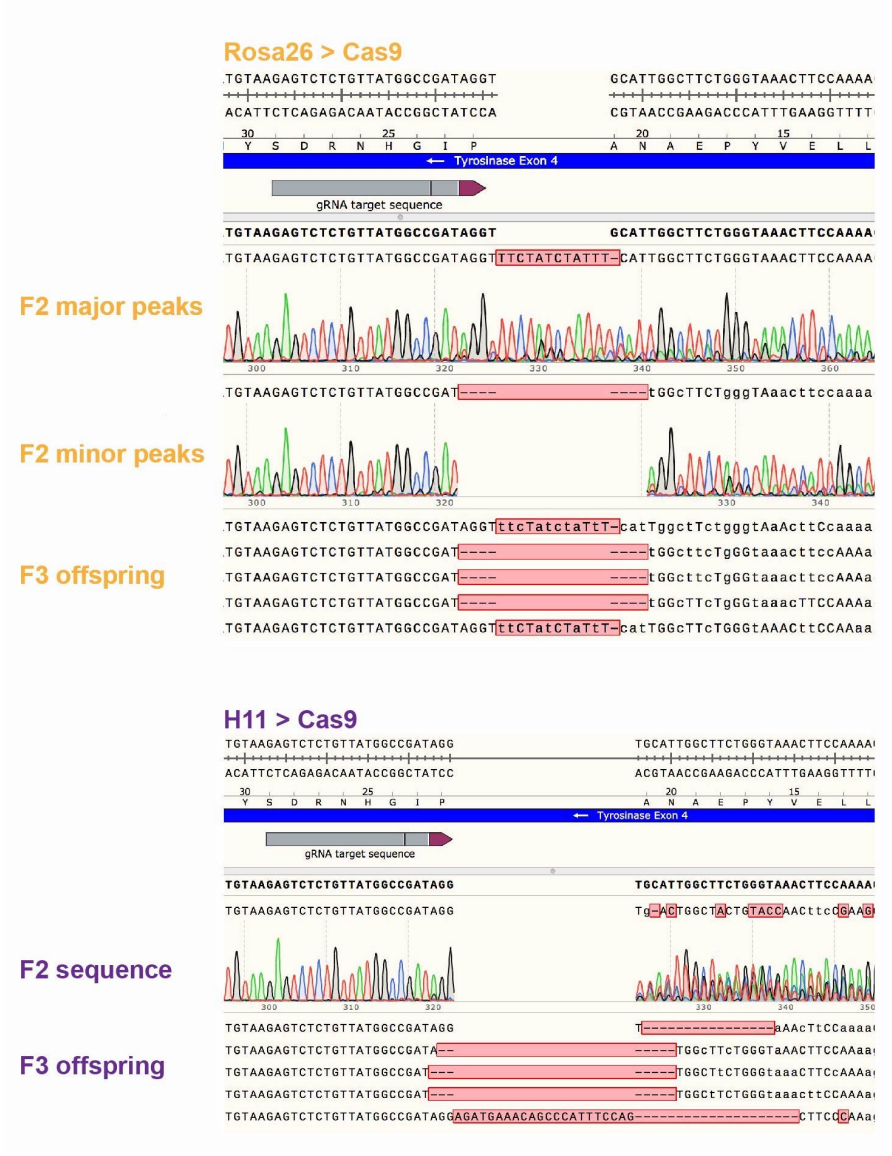
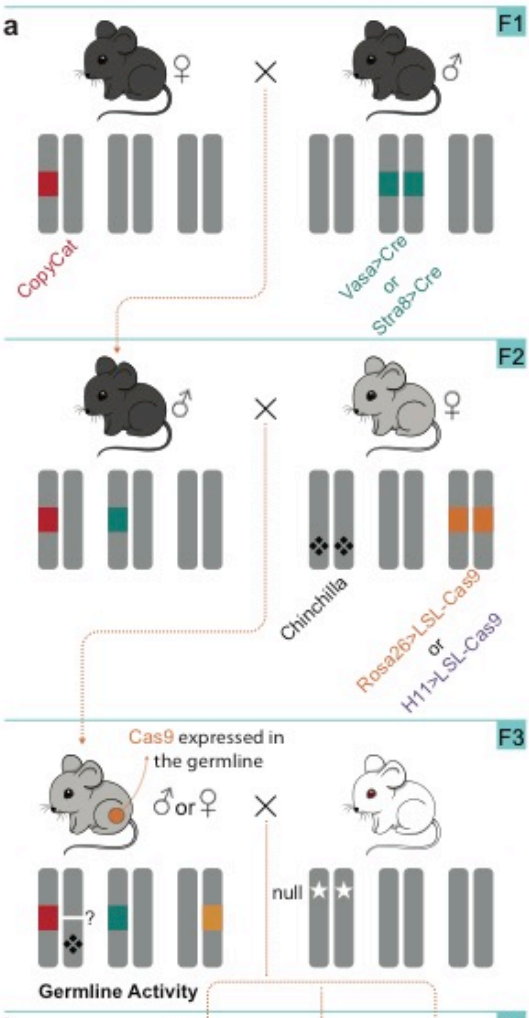


Figure 2.4 | *Rosa26-Cas9* and *H11-Cas9* constitutive lineages have different numbers of unique NHEJ indels. Top, a single representative Sanger sequence trace of the bulk PCR product amplified from a *Rosa26-Cas9;Tyr*^{CopyCat}-positive F_2 mouse (*Rosa26* family 1 in Table 2.3). Highlighting either major or minor peaks reveals two distinct alleles. *Tyr*^{ch}-positive offspring of this F_2 individual each match one of the two alleles. Bottom, a single representative Sanger sequence trace of the bulk PCR product amplified from an *H11-Cas9*;*Tyr*^{CopyCat}-positive F_2 mouse (*H11* family 1 in Table 2.3). Alternate alleles cannot be highlighted because of the complexity of overlapping peaks. *Tyr*^{ch}-positive offspring each have one of four unique alleles.

Fig. 2.5 | Breeding strategy to produce *Tyr*$^{CopyCat/ch}$ mice with a conditional *Cas9* transgene and a germline restricted *Cre* transgene. a. Breeding strategy to unite *Cre*, *Cas9*, and *Tyr*^{CopyCat} transgenes. F_3 offspring were crossed to *Tyr*^{null} mice. **b.** F_4 *Tyr*^{ch}+ phenotypes and genotypes reveal genotype conversion in the female germline.



Cre	Cas9	F3	no cut or functional repair	NHEJ disruption	HDR conversion	% HDR conversion
Vasa-Cre	Rosa26 ♀	1	-	15	-	-
		2	1	2	-	-
		3	-	3	1	25%
		4	4	5	3	25%
		5	6	6	1	8%
	Rosa26 ♂	1	-	25	-	-
		2	-	17	-	-
	H11 ♀	1	4	1	13	72%
		2	10	7	4	19%
		3	4	-	5	56%
		4	8	4	4	25%
		5	8	3	10	48%
	H11 ♂	1	-	15	-	-
		2	-	5	-	-
		3	-	2	-	-
		4	-	3	-	-
Stra8	R26♂	1	-	19	-	-
		2	-	3	-	-
	H11♂	1	3	21	-	-

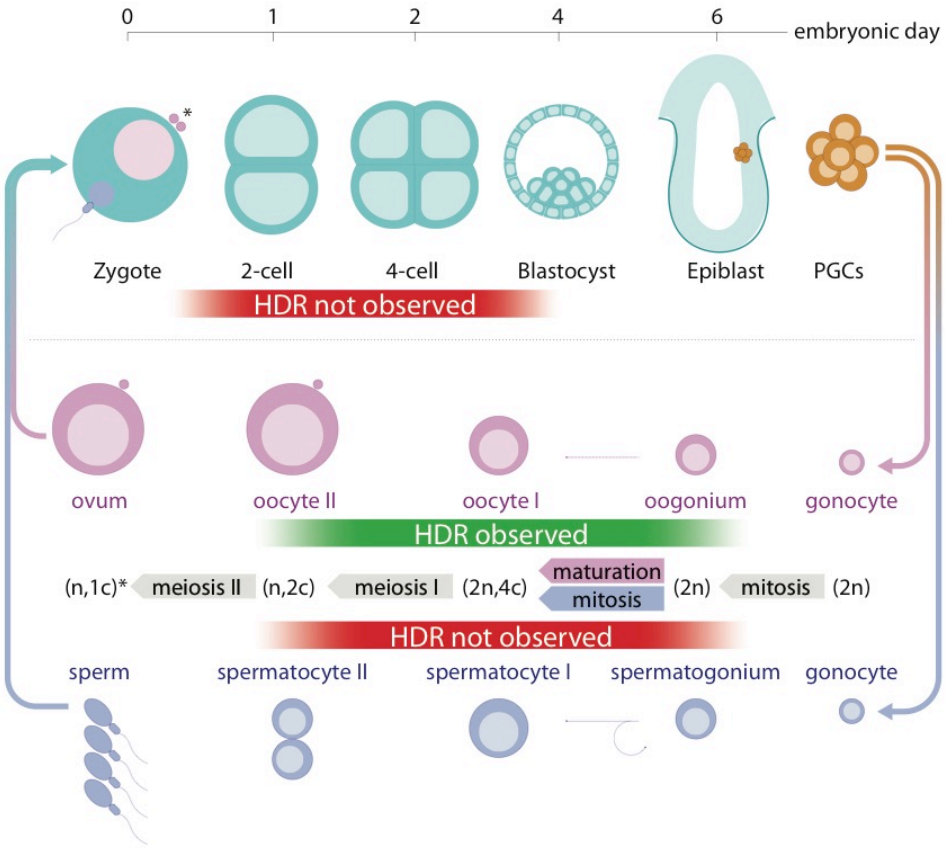


Fig. 2.6 | Genotype conversion by an active genetic element was observed in the female germline and not in the male germline or in the early embryo. Schematic of early embryonic and male and female germline development overlaid with the presence or absence of observed HDR. PGCs, primordial germ cells; *n*, number of homologous chromosomes; *c*, chromosome copy number. The asterisk indicates the difference between male sperm (*n*, 1*c*) and female ovum, which remains (*n*, 2*c*) until second polar body extrusion after fertilization.

Table 2.1 | Coat colour of F_2 individuals that were constitutive *Cas9*-positive and *Tyr*$^{CopyCat/ch}$

	Rosa26>Cas9	H11>Cas9
White	17	3
Mosaic	0	21

Table 2.2 | Analyses of phenotypes and genotypes of all F_3 progeny of the constitutive *Cas9* crosses Genotypes and phenotypes of all F_3 progeny in Fig. 2.1d-f that are the offspring of a subset of F_2 individuals listed in Table 2.1. Prim., primarily; referring to the coat colour that covers the greatest total area when there was obviously not an equal proportion of white and grey. *A family with offspring that were Tyr^{ch}-negative and mCherry-negative, suggesting a large deletion in the recipient chromosome that may encompass the Tyr^{ch} SNP.

			Chinchilla + F3 offspring			
	F2 parent color	Total F3 Chinchilla-	Total F3 Chinchilla+	White (NHEJ mutation)	Grey (no cut or functional repair)	mCherry+ (HDR gene conversion)
Rosa26 Family 1	white	30	20	20	0	0
Rosa26 Family 2	white	11	15	15	0	0
Rosa26 Family 3*	white	28*	16	16	0	0
Rosa26 Family 4	white	22	22	22	0	0
Rosa26 Family 5	white	16	16	16	0	0
H11 Family 1	prim. white mosaic	14	15	15	0	0
H11 Family 2	prim. grey mosaic	23	28	25	3	0
H11 Family 3	prim. white mosaic	13	9	9	0	0
H11 Family 4	prim. white mosaic	12	15	15	0	0
H11 Family 5	mosaic	23	24	18	6	0

Table 2. 3 | **Allelic complexity of the constitutive *Rosa26*- and *H11-Cas9* families.** *A family with offspring that were Tyr^{ch}-negative and mCherry-negative, suggesting a large deletion in the recipient chromosome that may encompass the Tyr^{ch} SNP. This was counted as one of the two unique NHEJ indels.

	Number of distinct NHEJ alleles in 'n' $Tyr^{chinchilla}$+ offspring
Rosa26 Family 1	2 (n=9)
Rosa26 Family 2	3 (n=10)
Rosa26 Family 3*	2 (n=9)
Rosa26 Family 4	2 (n=7)
Rosa26 Family 5	3 (n=13)
H11 Family 1	4 (n=10)
H11 Family 2	6 (n=9)
H11 Family 3	4 (n=6)
H11 Family 4	2 (n=8)
H11 Family 5	7 (n=9)

Table 2.4 | Coat color of *Tyr*$^{CopyCat/ch}$ F_3 individuals that inherited a germline *Cre* transgene and a *loxSTOPlox:Cas9* conditional allele.

	Vasa>Cre				Stra8>Cre	
	Rosa26>LSLCas9		H11>LSLCas9		Rosa26>LSLCas9	H11>LSLCas9
	Female	Male	Female	Male	Male	Male
White	0	0	0	0	0	0
Grey	3	4	5	4	2	1
Mosaic	2	1	0	1	0	0

Table 2.5 | Analyses of phenotypes and genotypes of all F_4 progeny of the germline Cas9 crosses Phenotypes and genotypes of all F_4 progeny that are the offspring of a subset of F_3 individuals listed in Table 2.4.

Vasa>Cre
Rosa26>LSL Cas9

Female

	F3 parent color	Chinchilla - mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
Rosa26 Family 1	mosaic	10	15	14	1	0	0	0.0
Rosa26 Family 2	grey	8	3	1	1	1	0	0.0
Rosa26 Family 3	mosaic	6	4	3	0	0	1	25.0
Rosa26 Family 4	grey	9	12	2	3	4	3	25.0
Rosa26 Family 5	grey	14	13	6	0	6	1	7.7

Male

	F3 parent color	Chinchilla – mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
Rosa26 Family 1	grey	30	25	25	0	0	0	0.0
Rosa26 Family 2	grey	19	17	17	0	0	0	0.0

Vasa>Cre
H11>LSL Cas9

Female

	F3 parent color	Chinchilla – mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
H11 Family 1	grey	17	18	0	1	4	13	72.2
H11 Family 2	grey	18	21	5	2	10	4	19.0
H11 Family 3	grey	13	9	0	0	4	5	55.6
H11 Family 4	grey	14	16	2	2	8	4	25.0
H11 Family 5	grey	13	21	2	1	8	10	47.6

Male

	F3 parent color	Chinchilla – mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
H11 Family 1	grey	33	15	15	0	0	0	0.0
H11 Family 2	grey	9	5	5	0	0	0	0.0
H11 Family 3	grey	2	2	2	0	0	0	0.0
H11 Family 4	grey	2	3	3	0	0	0	0.0

Stra8>Cre

Rosa26>LSL Cas9
Male

	F3 parent color	Chinchilla – mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
Rosa26 Family 1	grey	23	19	19	0	0	0	0.0
Rosa26 Family 2	grey	3	3	3	0	0	0	0.0

H11>LSL Cas9
Male

	F3 parent color	Chinchilla – mCherry + F4 Total	Chinchilla + F4 offspring: Chinchilla + F4 Total	White, mCherry - (NHEJ mutation)	Mosaic (variable NHEJ mutations)	Grey (no cut or functional repair)	White, mCherry + (HDR conversion)	% Copying in F3 germline
H11 Family 1	grey	22	24	21	0	3	0	0.0

Methods

Statistics and reproducibility

The five families in each of the constitutive crosses (Table 2.2) are considered five independent experiments with each F_3 offspring representing an early embryonic DSB repair event in the F_2 parent. Each F_4 offspring in Table 2.5 represents a germline DSB repair event, an independent data point and each family is considered an independent trial. No statistical method was used to predetermine sample size. Given breeding limitations, we assessed as many offspring as possible in each family and aimed to assess up to five families for each strategy. To detect gene conversion events, we used genetic linkage rather than a statistical test of inheritance greater than 50% expected by Mendelian segregation. Specifically, the receiver chromosome was marked with a SNP (Tyr^{ch}) located approximately 9.1 kb from the target site for gene conversion. The probability of a naturally occurring recombination event that would unite these ultra-tightly linked loci on the same chromosome is 4.7×10^{-5}, because the average genetic distance[158] for mouse chromosome 7 is 0.52 cM Mb^{-1}.

Cloning of the *Tyrosinase*^{CopyCat} transgene

All primers for cloning are listed in Table 2.6. Using primers v851 and v852, we amplified a backbone for bacterial propagation that also contained a Human U6 promoter and gRNA scaffold. We amplified a second fragment of DNA that contained the CMV enhancer and promoter driving expression of the mCherry fluorophore from plasmid #548 (provided by Dr. Mark Tuszynski), using the primers v853 and v854. The two fragments were joined using the Gibson Assembly technique with reagents from New England Biolabs (NEB) (Cat.# E5520S) to obtain the plasmid pVG211, which carried all

the components of the CopyCat except for the gRNA target sequence. To obtain the final transgene sequence, the *Tyrosinase* Exon 4 gRNA target (Tyr4a-gRNA) sequence was inserted by performing a plasmid primer mutagenesis using the primers v878 and v875 and the NEB Q5 Site-Directed Mutagenesis Kit (Cat.# E0554S) to obtain the pVG242 plasmid.

The Jackson Laboratory modified this plasmid to include homology arms for homologous recombination into the *Tyrosinase* locus (*Tyr*), precisely at the Tyr4a-gRNA target cut site. This donor plasmid was then used to introduce the targeting vector into the *Tyr* locus by pronuclear injection into zygotes of the C57BL/6J strain. Briefly, the Jackson Laboratory purchased capped Cas9 mRNA from Trilink for co-injection at 60 ng/ul together with 25-50 ng/ul of guide RNA (Tyr4a-gRNA target sequence: 5'-GTTATGGCCGATAGGTGCAT-3') and 10-20 ng/ul of the donor plasmid. The resulting founders were backcrossed, and offspring were screened for germline transmission.

Table 2.6 | Primers that were used for cloning the *Tyrosinase*^{CopyCat} transgene.

V851	CCAGCTAGCAGAGGGCCTATTTCCC
V852	GAGCTCGAATTCACTGGCCGTC
V853	TAGGCCCTCTGCTAGCTGGGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACG
V854	TGTAAAACGACGGCCAGTGAATTCGAGCTCCCATAGAGCCCACCGCAT
V875	GTTATGGCCGATAGGTGCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
V878	GGTGTTTCGTCCTTTCCACAAG

Mouse care

Mouse stocks used in this study are listed in Table 2.7. All mice were housed in accordance with UCSD Institutional Animal Care and Use Committee protocols and fed on a standard breeders diet. Adult males and females were used for breeding.

Table 2.7 | Mouse stocks that were used in this study.

Jackson Labs Stock Number	Jackson Labs Stock Name	Notes
26175	B6J.129(B6N)-Gt(ROSA)26Sortm1(CAG-cas9*,-EGFP)Fezh/J	Rosa:lox-STOP-lox Cas9
26179	B6J.129(Cg)-Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J	Rosa:constitutive Cas9
26816	B6;129-Igs2tm1(CAG-cas9*)Mmw/J	H11:lox-STOP-lox Cas9
27650	STOCK Igs2tm1.1(CAG-cas9*)Mmw/J	H11:constitutive Cas9
17490	B6.FVB-Tg(Stra8-icre)1Reb/LguJ	Stra8:Cre
6954	FVB-Tg(Ddx4-cre)1Dcas/J	Vasa:Cre
4828	FVB.129P2-Pde6b+ Tyrc-ch/AntJ	Tyrosinase^{chinchilla}

DNA extraction

We obtained <5 mm of tail tissue from each mouse between birth and postnatal day 21 for genotyping. We sealed tail wounds with KwikStop Stypic Powder. We then screened tails for expression of mCherry using a fluorescent dissecting microscope. We submerged tails in 500 uL of TNES buffer (10mM Tris, pH 7.5; 400mM NaCl; 100 mM EDTA; 0.6% SDS) with 3 uL of 10 mg/mL Proteinase K and digested overnight (8-20 hr) in a 56°C water bath. We then added 139 uL of 6 M NaCl to each sample, vortexed, and centrifuged for 10 minutes at 14,000g at room temperature. We transferred supernatant to a clean tube and precipitated DNA by adding 700 uL ice-cold 95% EtOH and placing samples overnight at -20°C. We pelleted the precipitated DNA by centrifugation at

14,000g for 10 minutes at 4°C. We washed the pelleted DNA with ice- cold 70% EtOH and allowed it to air-dry before resuspension in TE.

PCR reactions

We performed PCR using either Bioline Red MyTaq MasterMix or NEB Q5 2X MasterMix (if the product was to be submitted for Sanger sequencing) with the following recipes and cycling parameters.

Bioline Red MyTaq:
1X MasterMix, 0.5 uM primers, 1 uL DNA (between 10-200 ng DNA) in 20 uL with the following cycle parameters. “n” represents the annealing temperature, and “q” represents the elongation time; each is designated in Table 2.8.
95°C for 3’
30 repeats
    95°C for 15”
    n°C for 15”
    72°C for q”
72°C for 5’
10°C for ∞

NEB Q5:
1X MasterMix, 0.5 uM primers, 1 uL DNA (between 10-200 ng DNA) in 50 uL
98°C for 30”
35 repeats
    98°C for 30”
    64°C for 30”
    72°C for 3’
72°C for 5’
10°C for ∞

Table 2.8 | PCR primers that were used for genotyping in this study. *indicates primers that were designed at the Jackson Laboratory or in citations referenced in the manuscript.
IPC- internal positive control, amplifies *interleukin2* from chromosome 3.

Primer Name	Primer Sequence	Amplicon	Amplicon Length	Polymerase	Annealing Temp	Elongation Time	Sequencing Primer	Notes
Chinchilla L3	GGGAGGAAAGGGTGCTTGAG							
Chinchilla R1	CAGCAAGCTGTGGTAGTCGT	Chinchilla SNP	392bp	MyTaq	60	1'	Chinchilla R1	SNP G>A at position 252
EGFP L1	ACATGAAGCAGCACGACTTCT							Indicates presence of Rosa26-Cas9
EGFP R1	ACGTTGTGGCTGTTGTAGTTGT	EGFP	220bp	MyTaq	60	45"		and Rosa26-loxSTOPloxCas9
RLC-WT F*	GTTCGTGCAAGTTGAGTCCATC							Indicates presence of Rosa26 wild
RLC-WT R*	GGACTGAGAATAGGCCCAAATG	RLC-WT	~1200bp	MyTaq	60	45"		type allele
HCC-Tg-F*	GGGCAACGTGCTGGTTATTG							Indicates presence of H11-Cas9
HCC-Tg-R*	CCAGGCCGATGCTGTACTTC	HCC-Tg	425bp	MyTaq	60	10"		allele
HC-WT-F*	GGGGCCTCCAAGTCTTGACAGTAGAT							Indicates presence of H11 wild type
HC-Common-R*	CTGACCAGTGGGACTGCTTTTTCCAG	HLC-WT	200bp	MyTaq	60	10"		allele
HLC-Cas9-F*	CGGCCGCCACTCGACGATGTA							Indicates presence of H11-
HC-Common-R*	CTGACCAGTGGGACTGCTTTTTCCAG	HLC-Tg	350bp	MyTaq	60	10"		loxSTOPloxCas9 allele
HC-WT-F*	GGGGCCTCCAAGTCTTGACAGTAGAT							Indicates presence of H11 wild type
HLC-Common-R*	CTGACCAGTGGGACTGCTTTTTCCAG	HLC-WT	200bp	MyTaq	60	10"		allele
Stra8:Cre F*	AGATGCCAGGACATCAGGAACCTG							
Stra8:Cre R*	ATCAGCCACACCAGACACAGAGATC	Stra8:Cre	236bp	MyTaq	60	10"		Indicates presence of Stra8:Cre
IPC F*	CTAGGCCACAGAATTGAAAGATCT	Internal Positive						Internal Positive Control confirms
IPC R*	GTAGGTGGAAATTCTAGCATCATCC	Control	324bp	MyTaq	60	10"		PCR success
Vasa:Cre F*	CACGTGCAGCCGTTTAAGCCGCGT							
Vasa:Cre R*	TTCCCATTCTAAACAACACCCTGAA	Vasa:Cre	240bp	MyTaq	59	10"		Indicates presence of Vasa:Cre
IPC F*	CTAGGCCACAGAATTGAAAGATCT	Internal Positive						Internal Positive Control confirms
IPC R*	GTAGGTGGAAATTCTAGCATCATCC	Control	324bp	MyTaq	59	10"		PCR success
Tyr HAL F2	AATGGCTGCGAAGGCAC							CopyCat band indicates presence of
Tyr HAR R2	GGTTCAAAAGCTTCCCAATCCT	CopyCat/NHEJ	CopyCat: 2606bp NHEJ/WT: ~400bp	Q5	64	3'	Tyr HAR R2	CopyCat. NHEJ/WT band can be sequenced to identify NHEJ
cc F1	TCAATGTCCCAGCTAGCAGAGGG							
Tyr HAR R2	GGTTCAAAAGCTTCCCAATCCT	CopyCat	838bp	MyTaq	60	1'		Indicates presence of CopyCat allele
IPC F*	CTAGGCCACAGAATTGAAAGATCT	Internal Positive						Internal Positive Control confirms
IPC R*	GTAGGTGGAAATTCTAGCATCATCC	Control	324bp	MyTaq	60	1'		PCR success

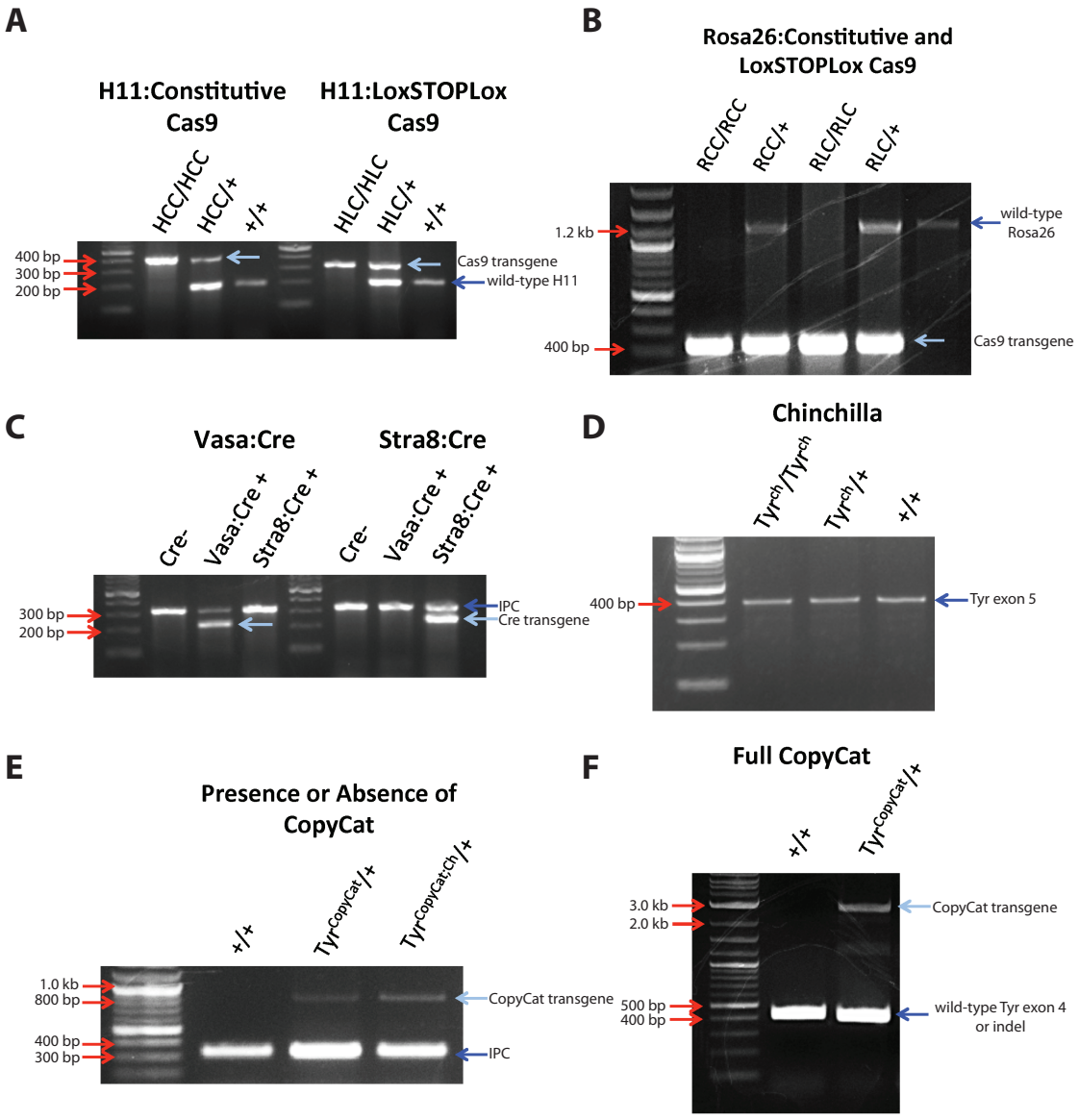


Figure 2.7 | Sample genotype results for each allele using primers as indicated in Table 2.8. For all: dark blue arrows indicate the wild type alleles or internal positive controls (IPC, amplifies *interleukin2* on chromosome 3), light blue arrows indicate transgenes. Red arrows denote relevant size markers in the DNA ladder for comparison.

Gel purification

We ran samples on 1-2% agarose gels to separate bands. Representatives of each genotyping reaction are shown in Fig. 2.7. Non-homologous end joining (NHEJ) amplicons or *Tyr*^{ch} amplicons were gel extracted using a QiaQuick Gel Extraction Kit as instructed. We submitted purified DNA for Sanger sequencing using the amplification primers noted in Table 2.8.

Blinded genotyping of F3 constitutive and F4 germline conditional offspring

The researcher did not have information about the status of mCherry fluorescence or coat color when PCR identification of the transgene and TyrChinchilla SNP was performed from each tail tip DNA sample. Presence or absence of the *Tyr*^{CopyCat} allele (visualized as a band on agarose gel) and presence or absence of the *Tyr*^{ch} SNP (determined from Sanger sequence trace) was documented separately for each individual and then merged to annotate the genotype at both sites. Randomization was not appropriate for this work.

Acknowledgements

We thank K. Hanley for the DNA extraction protocol; A. Green and A.-C. Chen for genotyping assistance; M. Tran for laser-capture microdissection in an effort to genotype spermatogonia; P. Jain for assistance with fibroblast transfection; H. Cook-Andersen and M. Wilkinson for conversations about mouse germline development; L. Montoliu for discussion of the tyrosinase locus; M. Tuszynski for plasmids and for early support of the project. This work was funded by a Searle Scholar Award from the

Kinship Foundation, a Pew Biomedical Scholar Award from the Pew Charitable Trusts, a Packard Fellowship in Science and Engineering from the David and Lucile Packard Foundation, and NIH grant R21GM129448 awarded to K.L.C. E.B. was supported by NIH grant R01GM117321, a Paul G. Allen Frontiers Group Distinguished Investigators Award and a gift from the Tata Trusts in India to TIGS-UCSD and TIGS-India. H.A.G. was supported by a Ruth Stern Graduate Fellowship and by the NIH Cell and Molecular Genetics training grant T32GM724039; V.M.G. was supported by NIH grant DP5OD023098. We thank B. Conklin, S. Qi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Chapter 2, in part, is a reprint of the material as it appears in Nature 2019. Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E. & Cooper, K. L. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566,** 105–109 (2019). The dissertation author was the primary researcher and author of this paper.

Conclusion: Beyond Gene Drive: Applications and considerations for genetically encoded CRISPR/Cas9 systems in laboratory rodents

Abstract

Self-propagating genetic elements that transmit at super-Mendelian frequencies have gained recent attention, as they could be used to drive desired alleles through a population with the goal of eliminating invasive species or disease vectors. We recently demonstrated that the genotype conversion mechanism underlying a CRISPR/Cas9-mediated gene drive is feasible in mice. Though substantial technical hurdles remain, overcoming these could lead to strategies that might decrease the spread of rodent-Bourne Lyme disease or eliminate invasive rat populations that devastate island ecology. Perhaps more immediately achievable, applications of CRISPR/Cas9-mediated genotype conversion in a laboratory setting could produce complex genotypes that reduce the time and cost in dollars and animal lives compared to Mendelian inheritance strategies. Here, we discuss what we have learned from early efforts to achieve CRISPR/Cas9-mediated genotype conversion, current limitations, potential for broader applications in the laboratory, and plans for optimizing this potentially powerful technology.

Main Text

For over a century, mouse models have provided insights into human development and disease[159]. Nevertheless, when compared to animals that produce hundreds of offspring at each mating event, mouse genetic approaches are an exercise in patient persistence. Researchers can spend years managing mouse colonies that seem to expand endlessly as they search for the exceptionally rare combination of homozygous

alleles at three or more loci. This goal is effectively impossible to achieve if two alleles are closely linked in *trans* on opposite chromosomes due to infrequency of natural meiotic recombination.

We recently demonstrated that a genetically encoded CRISPR/Cas9 system, termed an 'active genetic' system, can increase the transmission of a desired allele to 'super-Mendelian' frequencies. Whereas Mendelian inheritance transmits each of the two alleles to half of all offspring, an active genetic system favors transmission of one allele over the other by converting the parental genotype for a particular allele from heterozygosity to homozygosity in the germline. This is achieved by genotype conversion, which copies genetic information from 'donor' to 'recipient' locus on homologous chromosomes, and is therefore also a highly-effective approach to unite ultra-tightly linked alleles of two loci by converting them from a *trans* to a *cis* configuration.

Applications of these systems would improve the efficiency and ethics of laboratory mouse genetics by decreasing the time, cost, and number of animals needed to produce complex models of development, disease, and evolution. Outside of the laboratory, it has been suggested that active genetic systems could be used in a 'gene drive' approach to spread desired alleles over multiple generations through a wild population with the goal to modify animals that act as disease vectors or to eliminate invasive populations that devastate local ecosystems[91,160,161], but as of yet no vertebrate gene drive has been tested.

Here, we expand on the practical considerations and further research needed to fulfill the potential of active genetics in rodents.

An introduction to active genetics in mice

An active genetic system favors transmission of a 'preferred' donor allele by using genetically encoded gRNA and Cas9 to generate a double strand DNA break (DSB) in the 'non-preferred' recipient allele on the homologous chromosome. The resulting DSB can be resolved by interchromosomal homology directed repair (HDR) seeded by sequences on the donor chromosome that flank the preferred allele, copying the preferred allele to the recipient homologous chromosome where it replaces the non-preferred allele. The result is 'genotype conversion' at this locus from heterozygous to homozygous for the preferred allele.

Cas9 and gRNA can be encoded together in *cis* in the same genetic element[92,94] or they can be encoded in *trans* at different locations in the genome[138,162,163]. As demonstrated in *Drosophila* and *Anopheles* mosquitoes, a *cis*-encoded system that targets its own site of insertion has all the necessary machinery in a single transgene to copy itself onto the recipient chromosome and can therefore 'self-propagate,' such that every gamete at every generation could inherit the transgenic allele (Figure 3.1A,C). These 'gene drives' form the foundation of proposed wild release strategies. Because of their potential to transform populations, stringent biocontainment protocols are essential when working with a *cis*-encoded drive.

The *trans* configuration, referred to as a 'split-drive' system, used successfully in *Drosophila*[138,163] and in mouse[162], does not self-propagate. Cas9 and gRNA are encoded on distinct chromosomes and will independently assort, as would any traditional genetic modification (Figure 3.1B,C). We therefore recommend the *trans* configuration for

laboratory applications, because these animals require no exceptional biocontainment precautions.

In the first successful active genetic experiments in mouse, we inserted a 'CopyCat' transgene encoding a ubiquitously-expressed gRNA targeting *Tyrosinase* into the exact cut site of the encoded gRNA. The insertion of the CopyCat transgene results in a null allele of *Tyrosinase* and disrupts the gRNA recognition sequence such that the transgenic allele cannot be cut again by its encoded gRNA. When inherited together with a Cas9 transgene, the gRNA directs cleavage of the wild-type allele, allowing interchromosomal HDR to copy the $Tyrosinase^{CopyCat}$ transgene to the homologous chromosome[162]. We initially tested eight genetic strategies - two in the early embryo, two in the female germline, and four in the male germline. In our most successful strategy in the female germline, we observed an average 44% genotype conversion, bringing transmission of the transgene to 72% from the expected 50% by Mendelian inheritance[162]. Whereas previous experiments in flies and mosquitoes showed genotype conversion at high efficiencies in all cases[92,94], we never detected genotype conversion in the embryo nor initially in the male germline[162].

We detected genotype conversion by inheritance of the $Tyrosinase^{CopyCat}$ transgene together with a single nucleotide polymorphism (SNP) located exclusively on the recipient chromosome. The SNP is ultra-tightly linked to the gRNA target site such that reciprocal homologous recombination (HR) during meiosis would be exceedingly rare. This strategy allowed us to detect single instances of genotype conversion that copied the $Tyrosinase^{CopyCat}$ transgene onto the SNP-marked recipient chromosome, while differentiating true genotype conversion from reciprocal HR; reciprocal HR would

produce chromosomes with neither *Tyrosinase*^{CopyCat} nor the SNP allele, which were not observed.

Combining the transgene and SNP on a single chromosome also directly demonstrates the most immediate practical breakthrough of our pilot study. The estimated probability of recombination between the transgene and SNP loci is 4.7×10^{-5}, making it nearly impossible for a researcher to produce an animal that is homozygous at both sites by traditional breeding. By promoting genotype conversion of the transgenic locus via HDR, more than one-fifth of offspring in this study inherited the two closely-linked alleles on the same chromosome[162].

The meiotic window is critical for interchromosomal HDR

Mouse zygotic injection of gRNA and Cas9 with a DNA template is now standard practice to generate knock-in alleles by HDR[78]. Why then did we not observe interchromosomal HDR in the early embryo, despite efficient generation of DSBs as indicated by frequent insertion-deletion mutations (indels)? This may be at least partially explained by observations that end joining DNA repair pathways, which can cause indels, predominate over HDR in somatic cells and are the primary mechanism of DSB repair during G1 and M phase. Methods to chemically or genetically suppress end joining[164,165] or to restrict Cas9 expression to particular phases of the cell cycle[166,167] result in a moderate increase the relative frequency of HDR after zygotic CRISPR/Cas9 and template injection.

Despite the natural bias toward end joining, HDR is observed in 8-60% of zygotic CRISPR/Cas9 injections[78], in stark contrast to our observed 0% HDR in the early

embryo[162]. When injecting zygotes, however, thousands of copies of template are introduced into the cell[168]. In contrast, interchromosomal genotype conversion must use a chromosomal template, which limits the likelihood of interaction between donor and recipient. Methods to force a physical interaction between DSB and donor template in cultured somatic cells increased efficiency of HDR by up to 30-fold[169,170]. Further supporting a role for donor-recipient pairing, active genetics is highly efficient in *Dipteran* insects[92,94] where homologous chromosomes are aligned in somatic cells[171], and Cas9-dependent genotype conversion decreases precipitously in *Drosophila* when chromosomes have an inversion that would disrupt homologous alignment[93]. Thus physical proximity and, more specifically, appropriate alignment of chromosomes may be required for CRISPR-mediated genotype conversion.

Unlike *Diptera*, mammalian homologues are aligned only during early meiosis when DSBs are induced by the endogenous *Spo11* endonuclease and are subsequently resolved by reciprocal HR, also termed 'crossing over'[172]. Thus, meiotic cells naturally exchange information between paired homologous chromosomes, and indeed some portion of natural recombination events occur by allelic duplication, as in our system, rather than reciprocal HR[173–176]. In mice, recombination is dependent upon maintenance of the synaptonemal complex, which physically binds homologous chromosomes during meiosis I[177]. If alignment is indeed essential for interchromosomal HDR, active genetics in most non-*Dipteran* species may require precise meiotic timing.

A requirement for DSB formation coincident with chromosome alignment may also explain differences in the efficiency of genotype conversion in male and female germlines in our initial study. In female mice, oogonia initiate meiosis at embryonic day

13.5 (E13.5) and arrest with homologous chromosomes aligned at the end of prophase until ovulation cycles begin in the adult[178,179]. In contrast, male spermatogonia remain mitotic for the life of the animal. Beginning at around post-natal day 7 (P7), groups of spermatagonia periodically initiate a differentiation program, to become meiotic primary spermatocytes and ultimately haploid spermatozoa[180–182] (Figure 3.2). In our initial study, using a *Vasa:Cre* transgene first expressed at E15 to activate conditional Cas9 expression, DSBs may have been generated in meiotic oocytes, thus allowing genotype conversion, and in mitotic pre-spermatogonia, causing indel mutations that would be resistant to further cutting. Recent work in our laboratory has demonstrated that when Cas9 expression is more tightly restricted to early meiosis, using *Spo11* regulatory sequences, genotype conversion can be achieved in male mice, albeit at low efficiency (unpublished).

Even with Cas9 restriction to early meiosis, it is possible that other reproductive differences contribute to the efficiency of interchromosomal HDR. In both males and females, homologous chromosomes are unpaired up until early prophase of meiosis and remain paired until shortly before the first division, when they are segregated into individual cells, but the total duration of chromosome alignment differs between male and female germ cells. During spermatogenesis, the first cohort of maturing spermatocytes enters meiosis between P7-P9 and divides at around P18, allowing for approximately 10 days of chromosomal alignment[180] when genotype conversion could be possible. During oogenesis, homologous chromosomes are held together by the synaptonemal complex until oocytes arrest around birth and remain quiescent until ovulation, which will occur, at minimum, several weeks later. During this arrest,

homologues are associated by chiasmata at cross-over points, but are not fully synapsed along the length of the chromosome[178]. Thus during egg development, HDR can occur during the 6 day window when chromosomes are fully synapsed or possibly during the weeks- to months-long period of arrest where homologous alignment is maintained by chiasmata (Figure 3.2).

Population modification and wild release

Active genetic 'gene drives' have successfully increased transmission of alleles that make mosquitoes resistant to the malaria parasite, *Plasmodium falciparum*[94], that impair mosquito fertility[133], and that bias sex ratios in mosquito populations[183]. Applications of these findings have potential to make profound impacts on human health worldwide by, for example, making progress toward a malaria-free future. It has been similarly proposed that a gene drive could reduce transmission of rodent-borne Lyme disease or curb invasive rodent populations that decimate island biodiversity[160,161] by reducing the number of offspring in each successive generation and ultimately resulting in population collapse. However, strong selection against such a maladaptive allele will likely eliminate this 'preferred' allele from the population unless nearly every animal inherits the allele at every generation.

Our most effective genetic strategy transmitted an introduced transgene to 72% of offspring on average and only through the maternal lineage, an improvement over Mendelian ratios but almost certainly insufficient for rapid 'gene drive' to high frequency in a population. Of equal or perhaps greater importance is the ratio of HDR events to indel mutations, as each indel is resistant to future genotype conversion. The total

frequency of indel alleles can surpass the preferred allele frequency in a population and block a gene drive. One way to overcome this is to aggressively select against NHEJ events. This may be achieved by targeting genotype conversion at an essential gene such that faithful HDR preserves gene function, but offspring that inherit an indel are not viable. Multiple strategies with varying efficiencies of HDR, allele fitness, and prevalence of resistant alleles have been mathematically modeled to predict the resulting effective rise in preferred allele frequency over time[154,155].

While the efficiency of current systems in rodents is likely not sufficient for wild-release applications, encoded CRISPR/Cas9 strategies have immediate potential to transform the development of mouse models in the laboratory. Here, the ability to genotype and hand-select animals of interest at each generation tolerates both lower rates of HDR and a higher proportion of indels. The remainder of this article will highlight practical considerations for implementation and areas of research to improve laboratory applications of these technologies.

Multigenic and humanized rodent models

Single gene knockout models commonly used in laboratories are often insufficient to completely elucidate the mechanism of a complex process or disease or to test candidate therapeutics. Phenotypic analysis of knock-outs is complicated by the fact that 10-20% of single gene knock-out mice exhibit no detectable phenotype[184], even among genes characterized as "essential" in humans[185]. An absence of phenotype may be attributed to exon skipping[186,187], functional redundancy[184], or compensation[188],

suggesting that revealing the role of some genes may require the simultaneous inactivation of multiple genes in a pathway.

Regardless of the number of genes assessed, many mouse models fail to recapitulate human disease and development simply because mice are not humans. In clinical cancer trials, fewer than 10% of drugs that are effective in animal models have been successful with human patients[189], suggesting that cancer, immune function, and/or drug metabolism in mice is not sufficiently similar to humans to mimic human outcomes. 'Humanized' mouse models with regions of the mouse genome replaced by homologous human sequences promise to improve the value of the laboratory mouse as a model system in biomedical research[81,82]. By humanizing mice at multiple loci simultaneously, researchers can gain unprecedented insight into development and disease and better characterize and predict outcomes of novel drug therapies.

Both basic research, particularly in the fields of development and evolution where many genes orchestrate a given process, and biomedical research where multiple loci have been implicated in a variety of diseases, including heart disease[190], Alzheimer's disease[191], and diabetes[192], would benefit substantially from genetically encoded systems that can rapidly homozygose multiple loci of interest by genotype conversion and/or produce homozygous multi-gene knock-outs by sequence disruption.

A strategy to produce bi-allelic, tissue specific, multi-locus loss-of-function mice

While some gene knockouts have no phenotype, some pleiotropic genes have complex phenotypes that affect multiple tissues and may cause early lethality. There is, therefore, great value in the development of a system that allows for multiple genes to be

inactivated in a tissue-specific manner to simultaneously avoid pleiotropy and redundancy.

Cre-mediated recombination allows for conditional knockouts in a tissue-specific manner but requires modification of *both* alleles of *each* gene of interest so that each otherwise functional allele is flanked by lox sites (floxed). To abrogate function of multiple genes simultaneously in a tissue, a breeding scheme must assure assembly of 2 floxed alleles for each locus plus a copy of the Cre allele so that seven alleles must be combined in one animal to conditionally knock out three genes (Figure 3.3A).

We propose an alternative strategy, not yet tested, termed BATMN (Bi-Allelic,Tissue-specific, Multiplex kNock-out) to simultaneously knockout multiple genes in a tissue-specific manner by combining just three alleles. Our initial experiment in *Tyrosinase* used a publicly available Cre-inducible Rosa26:loxSTOPloxCas9 element (RosaLSLCas9)[143], which generated a null allele of *Tyrosinase* by NHEJ in 100% of males using either Vasa-Cre[148]or Stra8-Cre[149]. By combining this highly efficient RosaLSLCas9 allele with an efficient tissue-restricted Cre and encoded gRNAs, one could generate conditional DSBs in many loci simultaneously. Rather than encode each gRNA of interest separately, we propose the use of Polycistronic tRNA-gRNA (PTG) arrays, comprised of alternating tRNAs and gRNAs that can be transcribed and processed by endogenous cellular machinery into multiple functional gRNAs to target up to nine loci simultaneously[193–196]. Regardless of the number of loci to be targeted, there need only be one copy each of the Cre, RosaLSLCas9, and PTG alleles, and with the breadth of available Cre transgenes, only the PTG array need be engineered de novo (Figure 3.3B).

Though this technology could be powerful, it will be necessarily mosaic, as in each cell where Cas9 is expressed, an individual NHEJ repair event will occur. Nevertheless, with a high rate of true null allele generation, a genetically heterogeneous tissue could still be phenotypically uniform. Of note, when gRNA and Cas9 were injected into the zebrafish zygote, the resulting mosaic fish fully phenocopied established null mutant lines[197]. In some cases, mosaicism may actually be an asset when studying autonomous/non-autonomous signaling in a tissue. Furthermore, mosaicism may allow for the dilution of rare off target mutations across a tissue, such that if only a subset of cells carry an off target mutation, the tissue sum phenotype will be reduced. Nevertheless, in order to minimize phenotypic mosaicism, one must maximize the likelihood of fully abrogating gene function in every cell.

gRNAs should be designed to efficiently target essential domains of each gene. Tools like Inference of CRISPR Edits (ICE)[132] quantify overall indel generation and also the frequency of generating *specific* indels. gRNAs have been reported to preferentially generate specific indels[198,199], so one can identify highly efficient gRNAs while avoiding those that preferentially return indels of +/- three base pairs. gRNAs should be targeted to essential domains to produce a true phenotypic knock-out, rather than relying on the generation of early stop codons, which can result in exon skipping to produce functional, albeit truncated proteins[186,187]. Finally, we recommend using two gRNAs for any gene of interest, each targeting important or conserved domains, to increase the chances of producing at least one non-functional domain and/or deleting the region flanked by the gRNAs entirely.

Strategies to drive alleles to homozygosity

While our proof-of-principle study mobilized an introduced transgene, a preferred allele might be an existing or engineered sequence variant of a locus sometimes differentiated by only a few nucleotides or a single SNP. Recently in *Drosophila*, 'allelic drive' was demonstrated to convert a genotype at an existing sequence polymorphism when the non-preferred allele contains an appropriate gRNA target site[93] (Figure 3.4B). In this strategy, the gRNA and Cas9 are encoded at an unrelated genomic location, so that when the gRNA is expressed, the non-preferred allele is cut, allowing for HDR-mediated copying of the preferred pre-existing allele without any additional cargo. This principle can be extended to gRNA target sites within close proximity of, but not directly overlapping, the allele to be replaced because upon cleavage, the region surrounding the DSB may be copied, including the desired allele[93]. After cleavage, endogenous machinery resects the broken DNA ends at a rate of approximately 4 kb/hr. As strand invasion can take 30-45 minutes, we estimate that a region extending up to 3 kb in either direction from the DSB may be copied in a genotype conversion event[200] (Figure 3.4B').

In some cases, two preferred alleles exist but are tightly linked in *trans*, such that it is impossible to produce animals that are homozygous for both alleles. As we showed in our pilot study, targeting one of two tightly linked alleles for genotype conversion can bring the two alleles together in a *cis* configuration at a high rate[162], provided the allele of interest on the recipient chromosome is outside the window of resection (Figure 3.4C). As with other active genetic approaches, all these strategies require a sequence difference between homologous chromosomes at the gRNA target site, discussed in greater detail below. Such a sequence difference ensures that only the recipient chromosome is cut and

thus allows the preferred allele to directionally replace the linked non-preferred allele (Figure 3.4B).

Genetically encoded CRISPR/Cas9 systems could also assist in rapidly homozygosing multiple large genomic regions that have been humanized or replaced with homologous sequence of another species. Combining gRNAs can efficiently produce large-scale deletion of tens to hundreds of kb[201]. In cultured cells, two gRNAs that flank a region of interest can initiate a simultaneous deletion and replacement with exogenously provided DNA template. In human cells, this strategy has been used to delete up to 10 kb of sequence and simultaneously replace it with ~1 kb of donor sequence[202]. To our knowledge, no one has attempted a larger replacement using this strategy. In a mouse with one copy of a humanized locus, for example, gRNAs could be encoded to target each end of the homologous sequence in the mouse genome. This might facilitate removal of the mouse sequence, and interchromosomal HDR might replace the deletion with humanized sequence from the donor chromosome (Figure 3.4D). If implemented simultaneously at multiple loci, such an approach would accelerate production of genetically complex humanized models.

There are caveats to this untested approach. First, two DSBs on a chromosome can instead delete the region between them[197,202,203] potentially removing the mouse allele without converting it to a humanized state. Nevertheless, such a deletion could be a desirable alternative outcome, as the resulting animal would express only humanized transcripts from this locus (Figure 3.4E). Second, homology between mouse and human sequence within the humanized region could result in partial genotype conversion such that the humanized locus is not contiguously homozygous (Figure 3.4F). For studies of

the functional consequence of evolutionary sequence divergence, this alternative outcome may help to assign genotype-phenotype relationships to subregions of larger sequences without the need to engineer new lines. Last, and perhaps the least desirable outcome, it is also possible that DSBs are resolved by end joining at one or both ends of the humanized region, leaving an indel that is resistant to subsequent gRNA targeting but without altering gene function (Figure 3.4G).

Considerations when designing an active genetic system in rodents

Despite the potentially powerful applications of these emerging technologies, there are hurdles yet to be overcome and important technical details that must be taken into account for the successful design of any genotype conversion strategy. In particular, genotype conversion relies on careful analysis of gRNA target sequences, the timing of Cas9 expression, and the size and sequence of the preferred allele to be copied. Here, we expand on these technical considerations.

Control over the directionality of genotype conversion is at the core of an active genetic system, since the non-preferred allele needs to be replaced by the homologous preferred allele. To ensure unidirectional genotype conversion, gRNAs must exclusively target the recipient allele and never the preferred donor allele. In human cells, gRNAs will tolerate up to three nucleotide mismatches to cut target sequences[204]. The location of these mismatches seems to directly correlate with the efficiency of cutting such that a mismatch close to the PAM site prohibits cutting more effectively than a mismatch located more distally in the target sequence[205,206]. Furthermore, *SpCas9* is capable of targeting sites with a 5'-NAG PAM albeit with a lower efficiency than with a 5'-NGG

PAM[207]. Thus single changes in either the target site or the PAM site are insufficient to ensure directional genotype conversion.

Active genetic strategies should therefore be designed to target a recipient allele where the homologous sequence of the donor chromosome contains a fully disrupted PAM site or four or more nucleotide mismatches, preferably located near the PAM site. Both the efficiency and exclusivity of DSB generation in the recipient but not the donor allele should also be empirically determined using *in vitro* methods prior to engineering animals with the encoded gRNA[208]. If no unique gRNA target sequence can be found on the recipient chromosome near the locus of interest, researchers can first use CRISPR/Cas9 to preemptively mutate the target site on the donor chromosome prior to initiating active genetic breeding strategies.

In addition to the importance of ensuring directionality, we think that genotype conversion by interchromosomal HDR requires that the timing of DSB formation *in vivo* is limited to early meiosis. The strategy in mice that yielded the highest rate of genotype conversion to date required two transgenes for Cas9 expression: germline-restricted Cre[148,149] and conditional Cas9[143,144], which is not optimal if the goal is to simplify complex genetic strategies. While studies in fly and mosquito achieved genotype conversion using a *Vasa* promoter to drive Cas9 expression, the *Vasa* homolog in mouse (*Vasa*/*Ddx4*) is first expressed early in primordial germ cell (PGC) specification, at E6.5 in males and females[209]. Since females don't enter meiosis until E13.5, and males don't enter meiosis until well after birth, *Vasa* is likely expressed too early to capture the window of chromosome alignment in rodents.

It is worth noting that the Vasa:Cre transgene successfully used in our initial studies of genotype conversion uses a regulatory sequence that initiates germline-specific expression at E15, more than a week after the onset of endogenous *Vasa* expression. However, a recent study in mouse using this same *Vasa* regulatory sequence to directly drive Cas9 failed to achieve statistically significant evidence of genotype conversion, though their detection methods relied on observed allele frequencies, rather than genetic association with a linked marker. It is therefore possible that some level of genotype conversion did occur in their experiments at incredibly low levels[210].

In an effort to improve genotype conversion efficiencies in the male and female mouse by more precisely limiting DSB formation, we engineered a Cas9 knock in allele at the *Spo11* locus (*Spo11:Spo11-Cas9*). *Spo11*, the endonuclease that creates DSBs to initiate crossing over[172], is expressed during the meiotic window of chromosome alignment. This strategy did produce detectable rates of HDR in the male germline, suggesting that restricting expression of Cas9 to more precisely fall within the window of chromosome alignment may indeed allow for genotype conversion in both males and females. However, the overall efficiency of DSB generation, and therefore of genotype conversion, was very low. Intriguingly, however, the ratio of genotype conversion to indel mutation increased in both male and female germlines, consistent with the hypothesis that HDR is preferentially used for DNA repair during the meiotic window (unpublished).

The fact that Cas9 under direct *Vasa* or *Spo11* regulatory control produced lower degrees of DSB generation may speak to requirement for a sufficient level of Cas9 protein expression. In the most successful genotype conversion experiments to date,

germline-restricted Cre expression allowed irreversible Cas9 expression under the very strong CAG promoter. In contrast, when Cas9 is under control of the *Spo11* promoter, for example, it is only expressed during the time and to the level that *Spo11* is expressed. Thus, it is possible that sufficient DSB generation and subsequent genotype conversion relies on an extended duration and/or high level of Cas9 expression.

Other possible candidates to more precisely initiate Cas9 expression during chromosome alignment include regulatory sequences associated with *Dmc1*, a homolog of the *E. coli* recombinase gene, *RecA,* required for crossing over[211]; *Msh4*[212]/*Msh5*[213], DNA repair proteins essential for chromosome pairing; and *Scp3*[214] and other synaptonemal complex proteins directly responsible for synapsis of homologous chromosomes. Alternatively, it may be most effective to express Cas9 *before* chromosomes are paired to avoid missing the window of chromosome alignment. Although the Stra8-Cre transgene we used previously is not expressed in female mice[149], endogenous *Stra8* is first expressed at E12.5 in female mice, a day before meiosis begins[215]. *Stra8* is, however, expressed too early prior to meiosis in males to be useful[216] (Figure 3.2). Strong promoter control of chemically inducible Cas9 fusion proteins that are restricted to the cytoplasm or are rapidly degraded without a provided ligand may provide both high levels of Cas9 and tighter control over the timing of DSB generation[217].

It is not known whether there is an upper limit to the size of a region that can undergo genotype conversion. The *Tyrosinase*^{CopyCat} allele used for the initial proof-of-feasibility experiments in mice, the only active genetic element successfully converted to homozygosity in a vertebrate to date, is approximately 2.8 kb[162]. The largest active

genetic element assessed so far in any species is a ~17 kb transgene in mosquitoes, which achieved over 95% genotype conversion[94]. Since HDR tends to scale inversely with the length of a bounded sequence[84–86], it is possible that the efficiency of genotype conversion for larger elements may be lower. Indeed, one study in *Drosophila* showed that genotype conversion of a few base pairs exceeded the efficiency of copying a larger transgene at a closely-linked locus, suggesting that further increasing the cargo size in rodents might further decrease the efficiency of genotype conversion[93]. However, the efficiency of HDR using exogenous DNA sequences as template also tends to increase as a function of flanking homology arm length[84–86]. It is therefore possible that decreased genotype conversion efficiency will be at least partially mitigated by the extent of homology surrounding a genetically encoded donor allele in the context of a whole chromosome.

The unknown mechanics of mobilizing multiple elements at once

Implementation of active genetic approaches in the laboratory mouse promises to have the most transformative impact if multiple alleles can be readily combined to produce models of genetically complex phenotypes. To date, however, CRISPR/Cas9-mediated interchromosomal HDR has been tested only at the *Tyrosinase* locus in mouse[162,210]. The targeting efficiencies of individual gRNAs are variable with respect to one another, and both Cas9 occupancy and relative rate of HDR to end joining repair have been shown to depend on chromatin state[69,70,166]. This implies that different loci may have different rates of DSB generation and HDR, and these may vary according to cell type. It is also possible that genetically encoded and highly transcribed Cas9 and gRNAs

may counter these phenomena to produce higher rates of DSB generation and HDR than systems where there are finite amounts of Cas9 and gRNA that subsequently degrade. However, these advantages may be offset by the necessity to restrict DSBs to the germline, which narrows the time in which HDR must occur.

An unknown that affects multi-locus genotype conversion efficiency is whether the combined probability at multiple loci is multiplicative or coordinated. If the rate of interchromosomal HDR is independent at each locus in a cell, then the combined probability should be multiplicative. If interchromosomal HDR is a function of cellular environment, including chromosome alignment and predominant repair machinery, then genotype conversion of multiple loci would be coordinated, and thus the combined probability is the same as the probability of a single genotype conversion event. Interestingly, a high correlation of genotype conversion at two distinct closely-linked loci was observed in *Drosophila*[93], lending support to a coordinated model for multiple loci. Calculations in Figure 3.5 of the hypothetical rate of inheritance according to each of these models use a 44% probability of genotype conversion, the average observed at the *Tyrosinase* locus using our most-optimal genetic strategy in female mice. We consider theoretical loci A, B, and C, in scenarios where a researcher breeds two mice heterozygous at each locus.

In a multiplicative model, each parent has 72% probability of transmitting each desired allele to an offspring (50% probability of transmitting the donor chromosome plus 44% conversion of the recipient chromosomes). If both parents perform genotype conversion at this rate, the likelihood of obtaining offspring homozygous at all three loci is 13.9%, an eight-fold improvement over Mendelian inheritance (Figure 3.5B). Even if

only one parent is capable of genotype conversion at this rate, the likelihood of obtaining a triple homozygote is 4.7%, a 2.9-fold improvement (Figure 3.5C).

In a coordinated model, 44% of germ cells have 'conversion-permissive' cellular conditions. An optimal 'conversion-permissive' cell would convert all loci to homozygosity, always producing a gamete with all three preferred alleles. The remaining 56% without permissive conditions would retain heterozygous alleles that segregate normally, and therefore have a Mendelian probability that 12.5% of these gametes carry all preferred alleles. Thus each animal has 51% chance of producing a gamete with all three preferred alleles and two such animals bred together have a 26% chance of producing a triple homozygous offspring, a 16-fold increase over Mendelian probability (Figure 3.5D).

Since biology rarely follows tidy models, the actual efficiency of genotype conversion at multiple loci will likely fall somewhere between a multiplicative and a coordinated probability. Targeting a cell at exactly the right time when HDR is favored over end joining and when homologous chromosomes are aligned might increase genotype conversion at all loci in a cell. Nevertheless, it is likely that even "perfect" cellular conditions will produce different outcomes at loci with different chromatin structure and rates of DSB formation. The lowest efficiency scenario, in which genotype conversion at different loci is purely multiplicative, is still substantially better than Mendelian inheritance. It is worth noting that the conceptual models here consider only three loci, but these methods could be used with many more. Regardless of which model is correct in practice, gains over Mendelian ratios expand as a function of the number of targeted loci (Table 3.1).

Conclusion

While active genetic technology has not yet achieved genotype conversion efficiencies in rodents that would make it practical for wild release to modify populations, it stands to transform laboratory science. At moderate efficiencies, genetically encoded CRISPR/Cas9 machinery could rapidly produce homozygous animals with multiple humanized regions, introduced transgenes, and/or polymorphic alleles. The ability to combine ultra-tightly linked alleles on a single chromosome can produce animals with genotypes that are too difficult to obtain by natural breeding. Importantly, all of these applications are theoretically scalable; individual elements can be designed that carry multiple gRNAs to target multiple sites at once, which further opens possibilities to simplify the study of complex genetic traits using encoded CRISPR/Cas9 systems. If successfully implemented, this technology promises to save money, time, and animal lives, while simultaneously expanding our ability to investigate some of the most complex developmental questions and most prevalent human diseases.

Figures and Tables

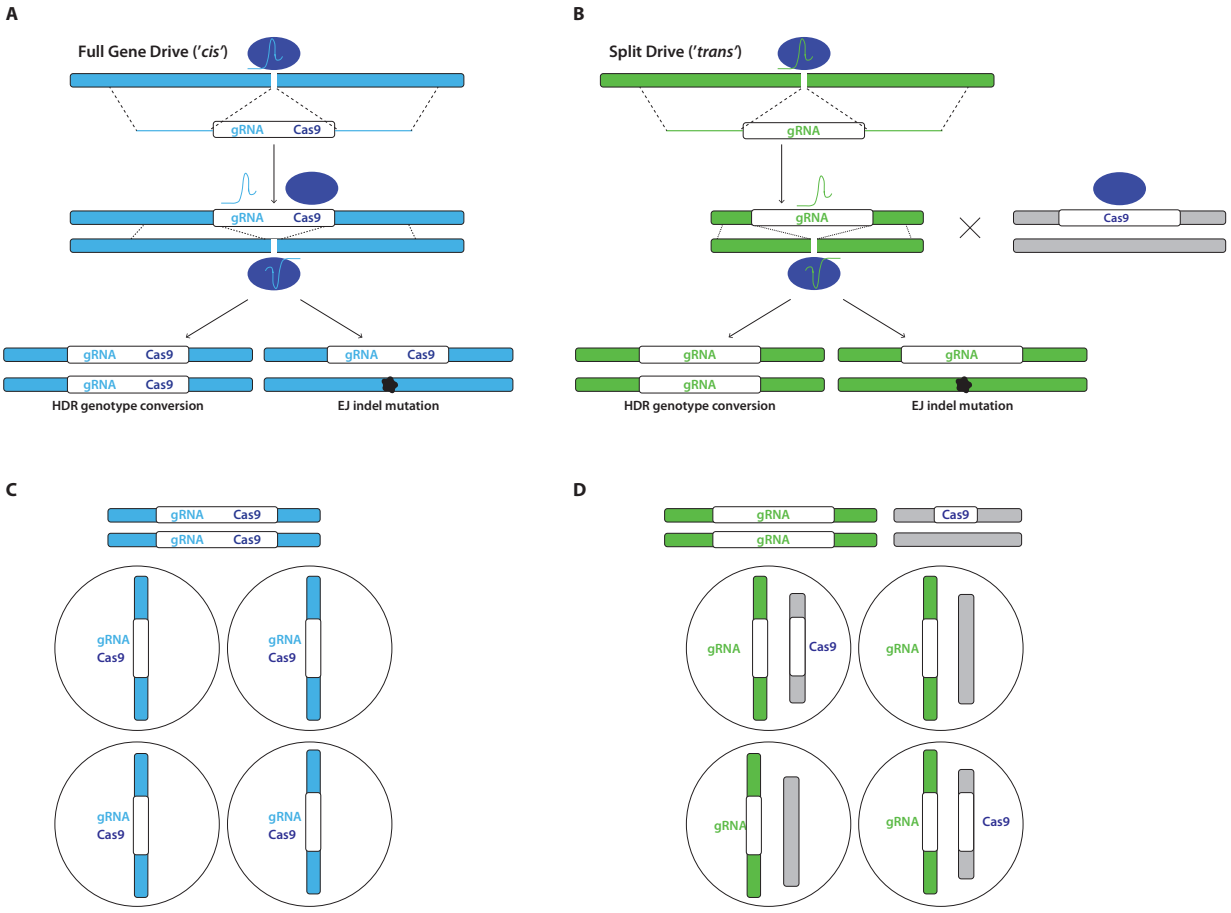


Figure 3.1| Self-Propagating Gene Drive vs. Split Drive. A. The genetic configuration of a 'self-propagating gene drive' system. A transgene containing both the gRNA and the coding sequence for Cas9 is knocked into a locus at the precise location of the gRNA's own cut site. Cas9 and gRNA form a DSB in the homologous wild-type chromosome. The DSB will be repaired by homology directed repair (HDR) using the homologous chromosome as a template or by end joining (EJ), leaving an indel mutation. **B**. The genetic configuration of a 'split drive' system. A transgene containing only the gRNA is knocked into the locus at the gRNA's own cut site. Cas9 is encoded by an unlinked transgene that segregates independently. **C** and **D**. Segregation of components in gametes after successful genotype conversion in a self-propagating gene drive (**C**) compared to a split drive (**D**).

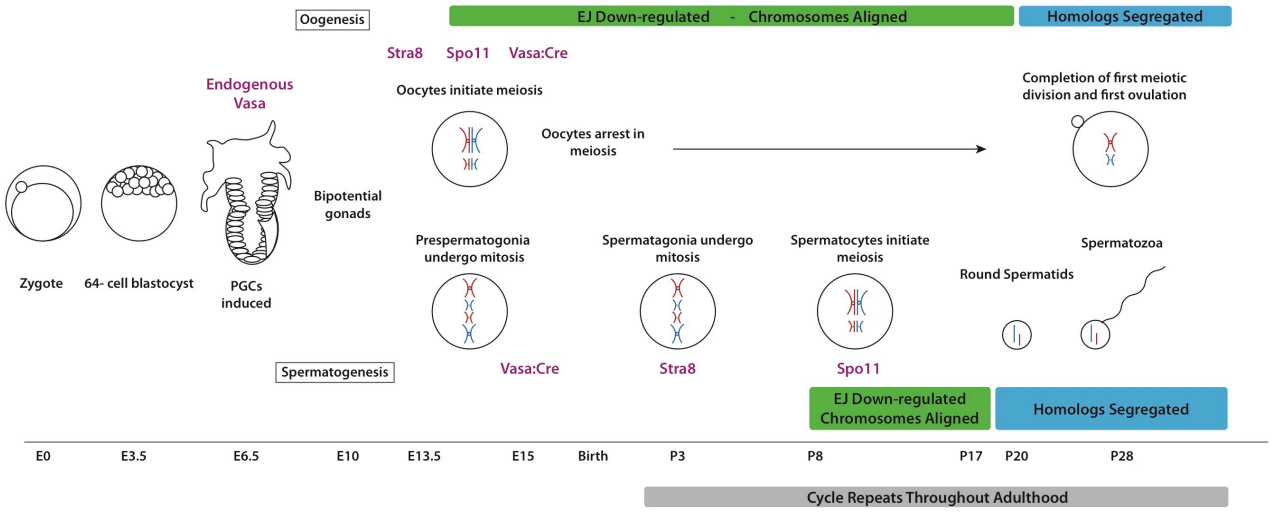


Figure 3.2 | Meiotic timelines differ between male and female mice. Male and female mice specify primordial germ cells (PGCs) at E6.5 and develop bipotential gonads by E10. **Oogenesis (top)**: Oocytes initiate meiosis between E13.5-E15 and arrest in metaphase of meiosis I by birth. The first primary oocytes to mature complete their first meiotic division near the time of the first ovulation at around P28. Throughout adulthood, sets of primary oocytes mature and divide with every estrus cycle. **Spermatogenesis (bottom)**: Prespermatogonia undergo mitosis between E13.5-E15 to produce self-renewing spermatogonia. Starting around P3 and continuing for the life of the animal, spermatogonia divide asymmetrically with one daughter undergoing four rounds of trans-amplifying mitoses to produce meiotic primary spermatocytes. The first round of these mitotic divisions begins at P3, with the first round of primary spermatocytes initiating meiosis between P7-P10. The first round of meiotic divisions are complete by P20. Purple text: Approximate onset of expression for *Spo11*, *Stra8*, *Vasa*, and *Vasa:Cre* in the germ line. Green bars indicate windows of development when homologous chromosomes are aligned and end joining (EJ) is downregulated. Blue bars indicate windows of development after the first meiotic division, when homologous chromosomes have segregated. Grey bar at bottom indicates stages of germ cell progression, aligned in both male and female above, that are repeated in cycles throughout adulthood.

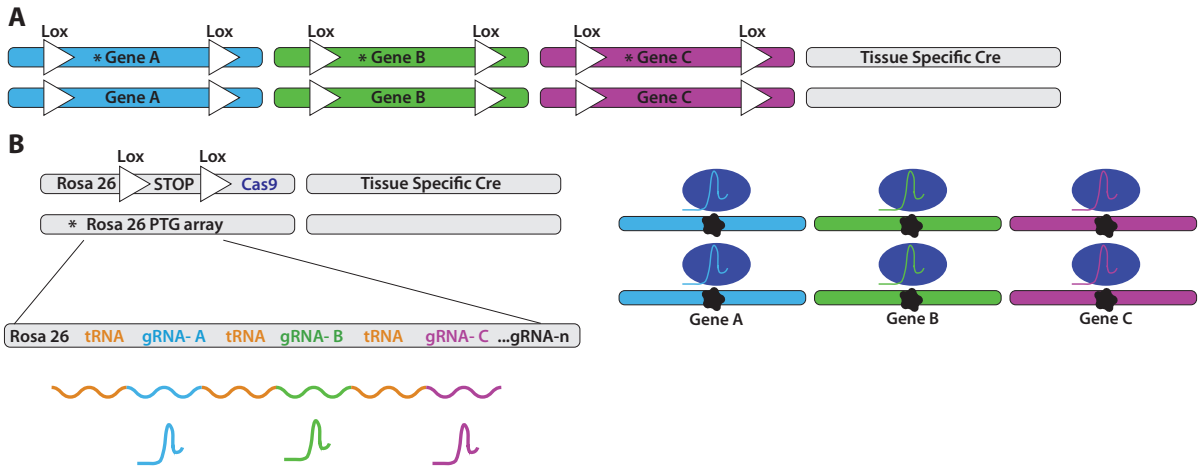


Figure 3.3 | Bi-Allelic Tissue-specific Multiplex kNock-out (BATMN). A. A traditional conditional knock-out consists of two floxed alleles for each gene to be knocked out and at least one copy of a tissue-specific Cre. Although many tissue-specific Cre lines exist, researchers must generate new floxed genes for each new gene being studied. **B.** BATMN knocks out multiple genes simultaneously while requiring assembly of only three alleles. A tissue-specific Cre excises a STOP signal to allow for expression of Cas9 in cells where Cre has been active. A polycistronic tRNA gRNA array (PTG) is knocked into the Rosa 26 locus on the opposite chromosome. The PTG array produces a polycistronic RNA molecule that is cleaved by endogenous tRNA processing machinery within the cell to free gRNAs, which direct Cas9 to genes of interest and generate DSBs which are subsequently resolved by end joining, generating indels. Only the PTG array must be engineered de novo. The example here depicts three gRNAs targeting three loci, but there is no known upper limit on the number of gRNAs that could be encoded in a single knock-in to the Rosa locus. Asterisks demarcate new alleles that must be engineered for each new experiment.

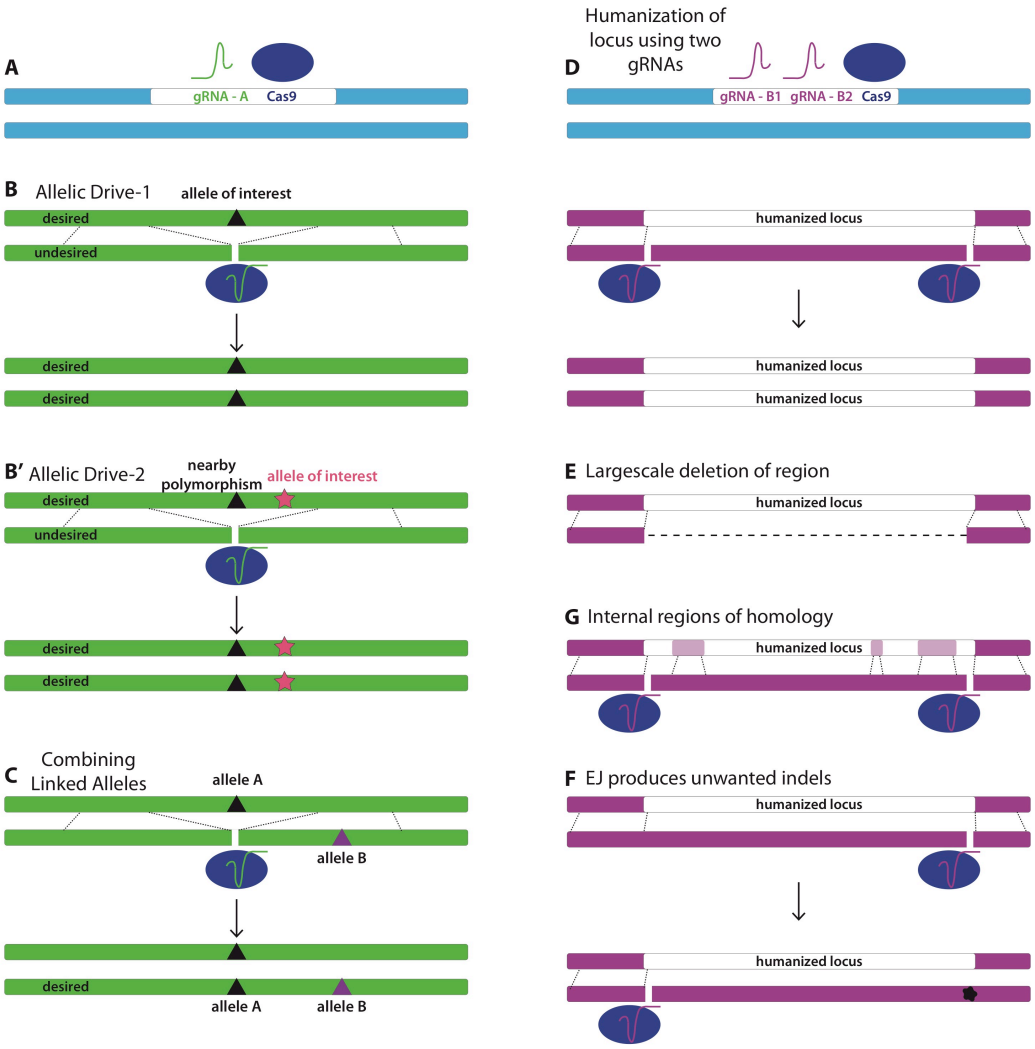


Figure 3.4 | Alternative active genetic strategies for the production of mouse models. A-C. *Strategies to make use of existing alleles*. **A**. Cas9 and gRNA are encoded together in a genomic location distinct from the target site of the gRNA (blue chromosome). This configuration is not self-propagating, because sequences flanking the gRNA/Cas9 transgene are not used as template to repair the DSB. **B**. *Allelic Drive-1*. The preferred allele interrupts the gRNA site so that only the non-preferred allele can be cut. The DSB is resolved by HDR from the preferred allele on the homologous chromosome. **B'**. *Allelic Drive-2*. The preferred allele sits nearby a polymorphism that disrupts a gRNA target site present in *cis* with the non-preferred allele. A DSB on the chromosome carrying the non-preferred allele is resolved by HDR, which converts the genotype of the entire region, copying the preferred allele to the homologous chromosome. **C**. *Combining Linked Alleles*. Allele A and B are closely linked in *trans*. Targeting Allele A for genotype conversion copies this allele to the chromosome that encodes Allele B so both preferred alleles are united on a single chromosome. **D-G**. *Humanization of a mouse locus*. **D**. gRNAs targeting either end of the locus are encoded with Cas9 at a genomic location distant from the locus of interest (blue chromosome). This configuration is not self-propagating. Cas9 and gRNAs target either end of the locus and generate DSBs, resulting in (**D**) genotype conversion, (**E**) large-scale deletion of the intervening region, (**F**) recombination at sites of internal regions of homology between the humanized region and the mouse chromosome, resulting in a discontiguous humanized haplotype, or (**G**) production of one or more indels, resulting in a resistant allele.

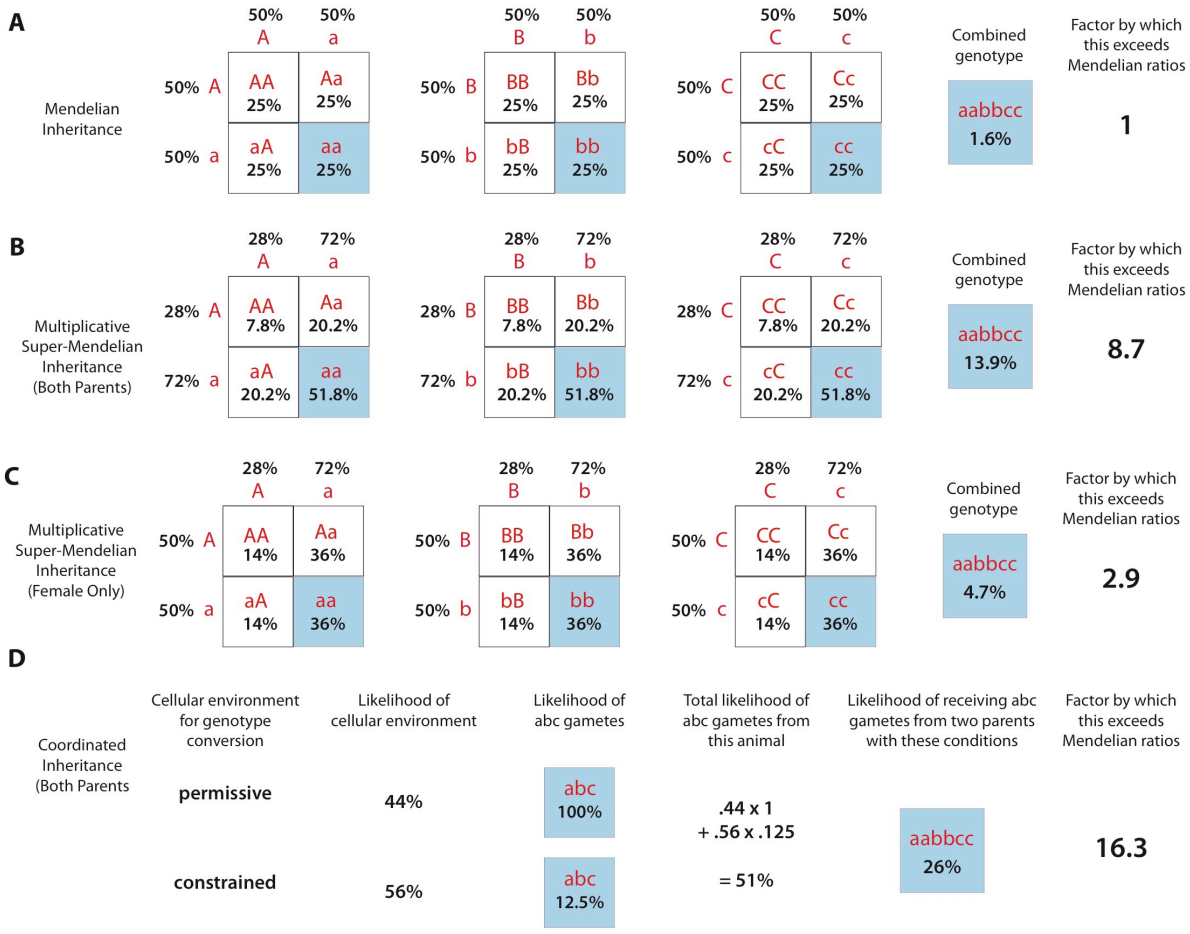


Figure 3.5 | Models of multiplex inheritance to obtain a triple homozygous genotype (aabbcc) from a cross of two animals heterozygous at each locus (AaBbCc x AaBbCc). Estimated probabilities according to principles of (**A**) normal Mendelian segregation; (**B**) a "Multiplicative" model where both parents can perform genotype conversion; (**C**) a "Multiplicative" model, where only the female parent is capable of genotype conversion; (**D**) a "Coordinated" model where genotype conversion can occur in both parents.

Table 3.1 | Likelihood of obtaining a gamete with preferred alleles at multiple unlinked loci by Mendelian ratios compared to predicted ratios from a Multiplicative Model and a Coordinated Model. Multiplicative Model assumes a 44% genotype conversion rate at each locus by each parent. Coordinated Model assumes 44% permissive cellular conditions for genotype conversion in each parent.

# loci of interest	Mendelian ratio	Estimated Multiplicative Ratio	Factor by which Multiplicative exceeds Mendelian	Estimated Coordinated Ratio	Factor by which Coordinated exceeds Mendelian
1	50.0%	72.0%	1.4	72.0%	1.4
2	25.0%	51.8%	2.1	58.0%	2.3
3	12.5%	37.3%	3.0	51.0%	4.1
4	6.3%	26.9%	4.3	47.5%	7.5
5	3.1%	19.3%	6.2	45.7%	14.7
6	1.6%	13.9%	8.7	44.9%	28.1
7	0.8%	10.0%	12.5	44.4%	55.5

Acknowledgements

Chapter 3, in part, has been submitted for publication of the material as it may appear in Nature Protocols, 2020. Grunwald, Hannah; Weitzel, Alexander; Cooper, Kimberly, 2020. The dissertation author was the primary author of this paper.

References

1. Basson, M. A. Signaling in cell differentiation and morphogenesis. *Cold Spring Harb. Perspect. Biol.* **4,** a008151 (2012).

2. Mirth, C. K. & Shingleton, A. W. Coordinating Development: How Do Animals Integrate Plastic and Robust Developmental Processes? *Front. Cell Dev. Biol.* **7,** 8 (2019).

3. Duprat, A.-M., Husson, D. & Gualandris-Parisot, L. Does gravity influence the early stages of the development of the nervous system in an amphibian? *Brain Res. Rev.* **28,** 19–24 (1998).

4. Hoson, T. Plant Growth and Morphogenesis under Different Gravity Conditions: Relevance to Plant Life in Space. *Life Basel Switz.* **4,** 205–216 (2014).

5. Shapiro, M. D., Hanken, J. & Rosenthal, N. Developmental basis of evolutionary digit loss in the Australian lizardHemiergis. *J. Exp. Zool.* **297B,** 48–56 (2003).

6. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34,** 1812–1819 (2017).

7. Rodrigues, A. R., Yakushiji-Kaminatsui, N., Atsuta, Y., Andrey, G., Schorderet, P., Duboule, D. & Tabin, C. J. Integration of Shh and Fgf signaling in controlling *Hox* gene expression in cultured limb cells. *Proc. Natl. Acad. Sci.* **114,** 3139 (2017).

8. Sheth, R., Grégoire, D., Dumouchel, A., Scotti, M., Pham, J. M. T., Nemec, S., Bastida, M. F., Ros, M. A. & Kmita, M. Decoupling the function of Hox and Shh in developing limb reveals multiple inputs of Hox genes on limb growth. *Development* **140,** 2130 (2013).

9. Purushothaman, S., Elewa, A. & Seifert, A. W. Fgf-signaling is compartmentalized within the mesenchyme and controls proliferation during salamander limb development. *eLife* **8,** e48507 (2019).

10. Panman, L., Galli, A., Lagarde, N., Michos, O., Soete, G., Zuniga, A. & Zeller, R. Differential regulation of gene expression in the digit forming area of the mouse limb bud by SHH and gremlin 1/FGF-mediated epithelial-mesenchymal signalling. *Development* **133,** 3419 (2006).

11. Tarazona, O. A., Slota, L. A., Lopez, D. H., Zhang, G. & Cohn, M. J. The genetic program for cartilage development has deep homology within Bilateria. *Nature* **533,** 86–89 (2016).

12. Hill, R. E. & Lettice, L. A. in *Kaufmans Atlas Mouse Dev. Suppl.* (eds. Baldock, R., Bard, J., Davidson, D. R. & Morriss-Kay, G.) 193–205 (Academic Press, 2016). doi:10.1016/B978-0-12-800043-4.00015-4

13. McQueen, C. & Towers, M. Establishing the pattern of the vertebrate limb. *Development* **147,** dev177956 (2020).

14. Barrow, J. R., Thomas, K. R., Boussadia-Zahui, O., Moore, R., Kemler, R., Capecchi, M. R. & McMahon, A. P. Ectodermal Wnt3/beta-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev.* **17,** 394–409 (2003).

15. Min, H., Danilenko, D. M., Scully, S. A., Bolon, B., Ring, B. D., Tarpley, J. E., DeRose, M. & Simonet, W. S. Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless. *Genes Dev.* **12,** 3156–3161 (1998).

16. Boulet, A. M., Moon, A. M., Arenkiel, B. R. & Capecchi, M. R. The roles of Fgf4 and Fgf8 in limb bud initiation and outgrowth. *Dev. Biol.* **273,** 361–372 (2004).

17. Crossley, P. H., Minowada, G., MacArthur, C. A. & Martin, G. R. Roles for FGF8 in the Induction, Initiation, and Maintenance of Chick Limb Development. *Cell* **84,** 127–136 (1996).

18. Cunningham, T. J., Zhao, X., Sandell, L. L., Evans, S. M., Trainor, P. A. & Duester, G. Antagonism between retinoic acid and fibroblast growth factor signaling during limb development. *Cell Rep.* **3,** 1503–1511 (2013).

19. Cooper, K. L., Hu, J. K.-H., ten Berge, D., Fernandez-Teran, M., Ros, M. A. & Tabin, C. J. Initiation of Proximal-Distal Patterning in the Vertebrate Limb by Signals and Growth. *Science* **332,** 1083 (2011).

20. Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capecchi, M. R. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. *Nature* **375,** 791–795 (1995).

21. Osterwalder, M., Speziale, D., Shoukry, M., Mohan, R., Ivanek, R., Kohler, M., Beisel, C., Wen, X., Scales, S. J., Christoffels, V. M., Visel, A., Lopez-Rios, J. & Zeller, R. HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Dev. Cell* **31,** 345–357 (2014).

22. Lettice, L. A., Horikoshi, T., Heaney, S. J. H., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., Hill, R. E. & Noji, S. Disruption of a long-range cis-

acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7548–7553 (2002).

23. Parr, B. A. & McMahon, A. P. Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. *Nature* **374,** 350–353 (1995).

24. Ahn, K., Mishina, Y., Hanks, M. C., Behringer, R. R. & Crenshaw, E. B. BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. *Dev. Camb. Engl.* **128,** 4449–4461 (2001).

25. Pizette, S., Abate-Shen, C. & Niswander, L. BMP controls proximodistal outgrowth, via induction of the apical ectodermal ridge, and dorsoventral patterning in the vertebrate limb. *Dev. Camb. Engl.* **128,** 4463–4474 (2001).

26. Raspopovic, J., Marcon, L., Russo, L. & Sharpe, J. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science* **345,** 566 (2014).

27. Zuzarte-Luis, V. & Hurle, J. M. Programmed cell death in the embryonic vertebrate limb. *Semin. Cell Dev. Biol.* **16,** 261–269 (2005).

28. Marazzi, G., Wang, Y. & Sassoon, D. Msx2 is a transcriptional regulator in the BMP4-mediated programmed cell death pathway. *Dev. Biol.* **186,** 127–138 (1997).

29. Cooper, K. L. The Lesser Egyptian Jerboa, Jaculus jaculus: A Unique Rodent Model for Evolution and Development. *Cold Spring Harb. Protoc.* **2011,** pdb.emo066704 (2011).

30. Moore, T. Y., Organ, C. L., Edwards, S. V., Biewener, A. A., Tabin, C. J., Jenkins Jr., F. A. & Cooper, K. L. Multiple Phylogenetically Distinct Events Shaped the Evolution of Limb Skeletal Morphologies Associated with Bipedalism in the Jerboas. *Curr. Biol.* **25,** 2785–2794 (2015).

31. Moore, T. Y., Cooper, K. L., Biewener, A. A. & Vasudevan, R. Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents. *Nat. Commun.* **8,** 440 (2017).

32. Shimer, H. W. Adaptations to Aquatic, Arboreal, Fossorial and Cursorial Habits in Mammals. III. Fossorial Adaptations. *Am. Nat.* **37,** 819–825 (1903).

33. Jordan, B., Vercammen, P. & Cooper, K. L. Husbandry and Breeding of the Lesser Egyptian Jerboa, *Jaculus jaculus*. *Cold Spring Harb. Protoc.* **2011,** pdb.prot066712 (2011).

34. Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. & Pang, E. C. K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142,** 169–196 (2005).

35. Infante, C. R., Mihala, A. G., Park, S., Wang, J. S., Johnson, K. K., Lauderdale, J. D. & Menke, D. B. Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-Genital cis-Regulatory Element in Snakes. *Dev. Cell* **35,** 107–119 (2015).

36. Lettice, L. A. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12,** 1725–1735 (2003).

37. Kvon, E. Z., Kamneva, O. K., Melo, U. S., Barozzi, I., Osterwalder, M., Mannion, B. J., Tissières, V., Pickle, C. S., Plajzer-Frick, I., Lee, E. A., Kato, M., Garvin, T. H., Akiyama, J. A., Afzal, V., Lopez-Rios, J., Rubin, E. M., Dickel, D. E., Pennacchio, L. A. & Visel, A. Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167,** 633-642.e11 (2016).

38. Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J. & Behringer, R. R. Regulatory divergence modifies limb length between mammals. *Genes Dev.* **22,** 141–151 (2008).

39. Shultz, L. D., Ishikawa, F. & Greiner, D. L. Humanized mice in translational biomedical research. *Nat. Rev. Immunol.* **7,** 118–130 (2007).

40. Dorner, M., Horwitz, J. A., Donovan, B. M., Labitt, R. N., Budell, W. C., Friling, T., Vogt, A., Catanese, M. T., Satoh, T., Kawai, T., Akira, S., Law, M., Rice, C. M. & Ploss, A. Completion of the entire hepatitis C virus life cycle in genetically humanized mice. *Nature* **501,** 237–241 (2013).

41. Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S. Y., Bhagwat, A. S., Malani, N., Anguela, X. M., Sharma, R., Ivanciu, L., Murphy, S. L., Finn, J. D., Khazi, F. R., Zhou, S., Paschon, D. E., Rebar, E. J., Bushman, F. D., Gregory, P. D., Holmes, M. C. & High, K. A. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475,** 217–221 (2011).

42. Scheer, N., Ross, J., Rode, A., Zevnik, B., Niehaves, S., Faust, N. & Wolf, C. R. A novel panel of mouse models to evaluate the role of human pregnane X receptor and constitutive androstane receptor in drug response. *J. Clin. Invest.* **118,** 3228–3239 (2008).

43. Hasegawa, M., Kapelyukh, Y., Tahara, H., Seibler, J., Rode, A., Krueger, S., Lee, D. N., Wolf, C. R. & Scheer, N. Quantitative Prediction of Human Pregnane X Receptor and Cytochrome P450 3A4 Mediated Drug-Drug Interaction in a Novel Multiple Humanized Mouse Line. *Mol. Pharmacol.* **80,** 518–528 (2011).

44. Wallace, H. A. C., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J. A., Hughes, J., Wood, W. G., Higgs, D. R. & Smith, A. J. H. Manipulating the Mouse Genome to Engineer Precise Functional Syntenic Replacements with Human Sequence. *Cell* **128,** 197–209 (2007).

45. Macdonald, L. E., Karow, M., Stevens, S., Auerbach, W., Poueymirou, W. T., Yasenchak, J., Frendewey, D., Valenzuela, D. M., Giallourakis, C. C., Alt, F. W., Yancopoulos, G. D. & Murphy, A. J. Precise and in situ genetic humanization of 6 Mb of mouse immunoglobulin genes. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 5147–5152 (2014).

46. Murphy, A. J., Macdonald, L. E., Stevens, S., Karow, M., Dore, A. T., Pobursky, K., Huang, T. T., Poueymirou, W. T., Esau, L., Meola, M., Mikulka, W., Krueger, P., Fairhurst, J., Valenzuela, D. M., Papadopoulos, N. & Yancopoulos, G. D. Mice with megabase humanization of their immunoglobulin genes generate antibodies as efficiently as normal mice. *Proc. Natl. Acad. Sci.* **111,** 5153–5158 (2014).

47. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

48. Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., Kato, M., Garvin, T. H., Pham, Q. T., Harrington, A. N., Akiyama, J. A., Afzal, V., Lopez-Rios, J., Dickel, D. E., Visel, A. & Pennacchio, L. A. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554,** 239–243 (2018).

49. Song, W. & Ovcharenko, I. Dichotomy in redundant enhancers points to presence of initiators of gene regulation. *BMC Genomics* **19,** 947 (2018).

50. Fujiwara, S. & Cañestro, C. Reporter Analyses Reveal Redundant Enhancers that Confer Robustness on Cis-Regulatory Mechanisms. *Adv. Exp. Med. Biol.* **1029,** 69–79 (2018).

51. Lettice, L. A., Devenney, P., De Angelis, C. & Hill, R. E. The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. *Cell Rep.* **20,** 1396–1408 (2017).

52. Guo, Y., Hou, L., Zhang, X., Huang, M., Mao, H., Chen, H., Ma, J., Chen, C., Ai, H., Ren, J. & Huang, L. A meta analysis of genome-wide association studies for limb bone lengths in four pig populations. *BMC Genet.* **16,** 95–95 (2015).

53. Chan, Y., Salem, R. M., Hsu, Y.-H. H., McMahon, G., Pers, T. H., Vedantam, S., Esko, T., Guo, M. H., Lim, E. T., GIANT Consortium, Franke, L., Smith, G. D., Strachan, D. P. & Hirschhorn, J. N. Genome-wide Analysis of Body Proportion

Classifies Height-Associated Variants by Mechanism of Action and Implicates Genes Important for Skeletal Development. *Am. J. Hum. Genet.* **96,** 695–708 (2015).

54. Saxena, A., Sharma, V., Neufeld, S. J., Tran, M. P., Gutierrez, H. L., Erberich, J. M., Birmingham, A., Cobb, J., Hiller, M. & Cooper, K. L. Interspecies transcriptome analyses identify genes that control the development and evolution of limb skeletal proportion. *bioRxiv* 754002 (2019). doi:10.1101/754002

55. Chandler, K. J., Chandler, R. L. & Mortlock, D. P. Identification of an ancient Bmp4 mesoderm enhancer located 46 kb from the promoter. *Dev. Biol.* **327,** 590–602 (2009).

56. Capecchi, M. R. High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell* **22,** 479–488 (1980).

57. Rülicke, T. & Hübscher, U. Germ line transformation of mammals by pronuclear microinjection. *Exp. Physiol.* **85,** 589–601 (2000).

58. Michalova, K., Bucchini, D., Ripoche, M.-A., Pictet, R. & Jami, J. Chromosome localization of the human insulin gene in transgenic mouse lines. *Hum. Genet.* **80,** 247–252 (1988).

59. Bestor, T. H. Gene silencing as a threat to the success of gene therapy. *J. Clin. Invest.* **105,** 409–411 (2000).

60. Feng, Y. Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. *Mol. Cell. Biol.* **21,** 298–309 (2001).

61. Hobbs, S. L. A., Warkentin, T. D. & DeLong, C. M. O. Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* **21,** 17–26 (1993).

62. Alexander, G. M., Erwin, K. L., Byers, N., Deitch, J. S., Augelli, B. J., Blankenhorn, E. P. & Heiman-Patterson, T. D. Effect of transgene copy number on survival in the G93A SOD1 transgenic mouse model of ALS. *Brain Res. Mol. Brain Res.* **130,** 7–15 (2004).

63. Haruyama, N., Cho, A. & Kulkarni, A. B. Overview: engineering transgenic constructs and mice. *Curr. Protoc. Cell Biol.* **Chapter 19,** Unit-19.10 (2009).

64. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51,** 503–512 (1987).

65. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336,** 348–352 (1988).

66. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci.* **91,** 6064 (1994).

67. Puchta, H., Dujon, B. & Hohn, B. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.* **21,** 5034–5040 (1993).

68. Chandrasegaran, S. & Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J. Mol. Biol.* **428,** 963–989 (2016).

69. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337,** 816 (2012).

70. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. & Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **339,** 823 (2013).

71. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. & Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819 (2013).

72. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 230–232 (2013).

73. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. & Doudna, J. RNA-programmed genome editing in human cells. *eLife* **2,** e00471 (2013).

74. Roth, D. B. & Wilson, J. H. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell. Biol.* **6,** 4295–4304 (1986).

75. Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. & Takeda, S. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17,** 5497–5508 (1998).

76. McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24,** 529–538 (2008).

77. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **7,** 1765–1771 (2008).

78. Singh, P., Schimenti, J. C. & Bolcun-Filas, E. A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics* **199,** 1–15 (2015).

79. Wang, B., Li, K., Wang, A., Reiser, M., Saunders, T., Lockey, R. F. & Wang, J.-W. Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. *BioTechniques* **59,** 201–208 (2015).

80. Zhang, F., Cheng, D., Wang, S. & Zhu, J. Crispr/Cas9‐mediated cleavages facilitate homologous recombination during genetic engineering of a large chromosomal region. *Biotechnol. Bioeng.* **117,** 2816–2826 (2020).

81. Devoy, A., Bunton-Stasyshyn, R. K. A., Tybulewicz, V. L. J., Smith, A. J. H. & Fisher, E. M. C. Genomically humanized mice: technologies and promises. *Nat. Rev. Genet.* **13,** 14–20 (2011).

82. Zhu, F., Nair, R. R., Fisher, E. M. C. & Cunningham, T. J. Humanising the mouse genome piece by piece. *Nat. Commun.* **10,** 1845 (2019).

83. Carstea, A. C., Pirity, M. K. & Dinnyes, A. Germline competence of mouse ES and iPS cell lines: Chimera technologies and genetic background. *World J. Stem Cells* **1,** 22–29 (2009).

84. Deng, C. & Capecchi, M. R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12,** 3365–3371 (1992).

85. Beumer, K. J., Trautman, J. K., Mukherjee, K. & Carroll, D. Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. *G3 Bethesda Md* **3,** 657–664 (2013).

86. Byrne, S. M., Ortiz, L., Mali, P., Aach, J. & Church, G. M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* **43,** e21–e21 (2015).

87. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* **36,** 344 (1950).

88. Sandler, L. & Golic, K. Segregation distortion in drosophila. *Trends Genet.* **1,** 181–185 (1985).

89. Schimenti, J. Segregation distortion of mouse t haplotypes. *Trends Genet.* **16,** 240–243 (2000).

90. Burt, A. & Koufopanou, V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14,** 609–615 (2004).

91. Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* **3,** e03401 (2014).

92. Gantz, V. M. & Bier, E. The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* **348,** 442–444 (2015).

93. Guichard, A., Haque, T., Bobik, M., Xu, X.-R. S., Klanseck, C., Kushwah, R. B. S., Berni, M., Kaduskar, B., Gantz, V. M. & Bier, E. Efficient allelic-drive in Drosophila. *Nat. Commun.* **10,** 1640 (2019).

94. Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. & James, A. A. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci.* **112,** E6736–E6743 (2015).

95. Cooper, K. L., Sears, K. E., Uygur, A., Maier, J., Baczkowski, K.-S., Brosnahan, M., Antczak, D., Skidmore, J. A. & Tabin, C. J. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511,** 41–45 (2014).

96. Myers, M. J. & Steudel, K. Effect of limb mass and its distribution on the energetic cost of running. *J. Exp. Biol.* **116,** 363 (1985).

97. BROWNING, R. C., MODICA, J. R., KRAM, R. & GOSWAMI, A. The Effects of Adding Mass to the Legs on the Energetics and Biomechanics of Walking. *Med. Sci. Sports Exerc.* **39,** (2007).

98. McHorse, B. K., Biewener, A. A. & Pierce, S. E. Mechanics of evolutionary digit reduction in fossil horses (Equidae). *Proc. R. Soc. B Biol. Sci.* **284,** 20171174 (2017).

99. Lopez-Rios, J., Duchesne, A., Speziale, D., Andrey, G., Peterson, K. A., Germann, P., Unal, E., Liu, J., Floriot, S., Barbey, S., Gallard, Y., Muller-Gerbl, M., Courtney, A. D., Klopp, C., Rodriguez, S., Ivanek, R., Beisel, C., Wicking, C., Iber, D., Robert, B., McMahon, A. P., Duboule, D. & Zeller, R. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511,** 46–51 (2014).

100. Guha, U., Gomes, W. A., Kobayashi, T., Pestell, R. G. & Kessler, J. A. In vivo evidence that BMP signaling is necessary for apoptosis in the mouse limb. *Dev. Biol.* **249,** 108–120 (2002).

101. Jumlongras, D., Lachke, S. A., O'Connell, D. J., Aboukhalil, A., Li, X., Choe, S. E., Ho, J. W. K., Turbe-Doan, A., Robertson, E. A., Olsen, B. R., Bulyk, M. L., Amendt, B. A. & Maas, R. L. An Evolutionarily Conserved Enhancer Regulates

Bmp4 Expression in Developing Incisor and Limb Bud. *PLoS ONE* **7,** e38568 (2012).

102. Gañan, Y., Macias, D., Basco, R. D., Merino, R. & Hurle, J. M. Morphological Diversity of the Avian Foot Is Related with the Pattern ofmsxGene Expression in the Developing Autopod. *Dev. Biol.* **196,** 33–41 (1998).

103. Fallon, J. F. & Cameron, J. A. Interdigital cell death during limb development of the turtle and lizard with an interpretation of evolutionary significance. *J. Embryol. Exp. Morphol.* **40,** 285 (1977).

104. Pan, Z.-Z., Kronenberg, M. S., Huang, D.-Y., Sumoy, L., Rogina, B., Lichtler, A. C. & Upholt, W. B. Msx2 Expression in the Apical Ectoderm Ridge Is Regulated by an Msx2 and Dlx5 Binding Site. *Biochem. Biophys. Res. Commun.* **290,** 955–961 (2002).

105. Ferrari, D., Lichtler, A. C., Pan, Z. Z., Dealy, C. N., Upholt, W. B. & Kosher, R. A. Ectopic expression of Msx-2 in posterior limb bud mesoderm impairs limb morphogenesis while inducing BMP-4 expression, inhibiting cell proliferation, and promoting apoptosis. *Dev. Biol.* **197,** 12–24 (1998).

106. Gañan, Y., Macias, D., Duterque-Coquillaud, M., Ros, M. A. & Hurle, J. M. Role of TGFβs and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* **122,** 2349–2357 (1996).

107. Zuzarte-Luís, V., Montero, J. A., Rodriguez-León, J., Merino, R., Rodríguez-Rey, J. C. & Hurlé, J. M. A new role for BMP5 during limb development acting through the synergic activation of Smad and MAPK pathways. *Dev. Biol.* **272,** 39–52 (2004).

108. Gutierrez, H. L., Tsutsumi, R., Moore, T. Y. & Cooper, K. L. Convergent metatarsal fusion in jerboas and chickens is mediated by similarities and differences in the patterns of osteoblast and osteoclast activities. *Evol. Dev.* **21,** 320–329 (2019).

109. Tran, M. P., Tsutsumi, R., Erberich, J. M., Chen, K. D., Flores, M. D. & Cooper, K. L. Evolutionary loss of foot muscle during development with characteristics of atrophy and no evidence of cell death. *eLife* **8,** e50645 (2019).

110. Quintero-Cadena, P. & Sternberg, P. W. Enhancer Sharing Promotes Neighborhoods of Transcriptional Regulation Across Eukaryotes. *G3amp58 GenesGenomesGenetics* **6,** 4167–4174 (2016).

111. Jensen, S. I., Lennen, R. M., Herrgård, M. J. & Nielsen, A. T. Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. *Sci. Rep.* **5,** 17874–17874 (2015).

112. G., R., Blank, K. & Wille, T. in *Genet. Manip. DNA Protein - Ex. Curr. Res.* (ed. Figurski, D.) (InTech, 2013). at <http://www.intechopen.com/books/genetic-manipulation-of-dna-and-protein-examples-from-current-research/site-directed-mutagenesis-using-oligonucleotide-based-recombineering>

113. Valenzuela, D. M., Murphy, A. J., Frendewey, D., Gale, N. W., Economides, A. N., Auerbach, W., Poueymirou, W. T., Adams, N. C., Rojas, J., Yasenchak, J., Chernomorsky, R., Boucher, M., Elsasser, A. L., Esau, L., Zheng, J., Griffiths, J. A., Wang, X., Su, H., Xue, Y., Dominguez, M. G., Noguera, I., Torres, R., Macdonald, L. E., Stewart, A. F., DeChiara, T. M. & Yancopoulos, G. D. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotech* **21,** 652–659 (2003).

114. Lallemand, Y., Nicola, M.-A., Ramos, C., Bach, A., Clement, C. S. & Robert, B. Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb. *Development* **132,** 3003–3014 (2005).

115. Duval, N., Daubas, P., Carbon, C. B. de, Clement, C. S., Tinevez, J.-Y., Lopes, M., Ribes, V. & Robert, B. Msx1 and Msx2 act as essential activators of Atoh1 expression in the murine spinal cord. *Development* **141,** 1726–1736 (2014).
116. Foerst-Potts, L. & Sadler, T. w. Disruption of Msx-1 and Msx-2 reveals roles for these genes in craniofacial, eye, and axial development. *Dev. Dyn.* **209,** 70–84 (1997).

117. Graham, A., Koentges, G. & Lumsden, A. Neural crest apoptosis and the establishment of craniofacial pattern: an honorable death. *Mol. Cell. Neurosci.* **8,** 76–83 (1996).

118. Achilleos, A. & Trainor, P. A. Neural crest stem cells: discovery, properties and potential for therapy. *Cell Res.* **22,** 288–304 (2012).

119. ALAPPAT, S., ZHANG, Z. Y. & CHEN, Y. P. Msx homeobox gene family and craniofacial development. *Cell Res.* **13,** 429–442 (2003).

120. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109,** 21.29.1-21.29.9 (2015).

121. Li, Z., Schulz, M. H., Look, T., Begemann, M., Zenke, M. & Costa, I. G. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20,** 45 (2019).

122. Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh Sandhu, K., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. & Ohlsson, R. Circular chromosome conformation capture (4C)

uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38,** 1341–1347 (2006).

123. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

124. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38,** W695–W699 (2010).

125. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. & Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** W597–W600 (2013).

126. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* **27,** 221–224 (2010).

127. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401,** 877–884 (1999).

128. Madishetty, K., Condamine, P., Svensson, J. T., Rodriguez, E. & Close, T. J. An improved method to identify BAC clones using pooled overgos. *Nucleic Acids Res.* **35,** e5–e5 (2007).

129. Robles-Oteiza, C., Taylor, S., Yates, T., Cicchini, M., Lauderback, B., Cashman, C. R., Burds, A. A., Winslow, M. M., Jacks, T. & Feldser, D. M. Recombinase-based conditional and reversible gene regulation via XTR alleles. *Nat. Commun.* **6,** 8783–8783 (2015).

130. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat. Methods* **6,** 363–369 (2009).

131. Lee, G. & Saito, I. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* **216,** 55–65 (1998).

132. Hsiau, T., Maures, T., Waite, K., Yang, J., Kelso, R., Holden, K. & Stoner, R. Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv* 251082 (2018). doi:10.1101/251082

133. Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A. & Nolan, T. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34,** 78–83 (2016).

134. Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T. & Crisanti, A. A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36,** 1062–1066 (2018).

135. Gantz, V. M. & Bier, E. The dawn of active genetics. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **38,** 50–63 (2016).

136. Gould, F. BROADENING THE APPLICATION OF EVOLUTIONARILY BASED GENETIC PEST MANAGEMENT. *Evolution* **62,** 500–510 (2008).

137. Miyaoka, Y., Berman, J. R., Cooper, S. B., Mayerl, S. J., Chan, A. H., Zhang, B., Karlin-Neumann, G. A. & Conklin, B. R. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6,** 23549 (2016).

138. Xu, X.-R. S., Gantz, V. M., Siomava, N. & Bier, E. CRISPR/Cas9 and active genetics-based trans-species replacement of the endogenous Drosophila kni-L2 CRM reveals unexpected complexity. *eLife* **6,** e30281 (2017).

139. Yokoyama, T., Silversides, D. W., Waymire, K. G., Kwon, B. S., Takeuchi, T. & Overbeek, P. A. Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. *Nucleic Acids Res.* **18,** 7293–7298 (1990).

140. Yen, S.-T., Zhang, M., Deng, J. M., Usman, S. J., Smith, C. N., Parker-Thornburg, J., Swinton, P. G., Martin, J. F. & Behringer, R. R. Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. *Dev. Biol.* **393,** 3–9 (2014).

141. Miyagishi, M. & Taira, K. U6 promoter–driven siRNAs with four uridine 3′ overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* **20,** 497–500 (2002).

142. Boshart, M., Weber, F., Jahn, G., Dorschhler, K., Fleckenstein, B. & Schaffner, W. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* **41,** 521–530 (1985).

143. Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., Dahlman, J. E., Parnas, O., Eisenhaure, T. M., Jovanovic, M., Graham, D. B., Jhunjhunwala, S., Heidenreich, M., Xavier, R. J., Langer, R., Anderson, D. G., Hacohen, N., Regev, A., Feng, G., Sharp, P. A. & Zhang, F. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159,** 440–455

144. Chiou, S.-H., Winters, I. P., Wang, J., Naranjo, S., Dudgeon, C., Tamburini, F. B., Brady, J. J., Yang, D., Grüner, B. M., Chuang, C.-H., Caswell, D. R., Zeng, H., Chu,

P., Kim, G. E., Carpizo, D. R., Kim, S. K. & Winslow, M. M. Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes Dev.* (2015).

145. Beermann, F., Orlow, S. J. & Lamoreux, M. L. The Tyr (albino) locus of the laboratory mouse. *Mamm. Genome* **15,** 749–758 (2004).

146. Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. *Cell* **88,** 375–384 (1997).

147. Goedecke, W., Eijpe, M., Offenberg, H. H., Aalderen, M. van & Heyting, C. Mre11 and Ku70 interact in somatic cells, but are differentially expressed in early meiosis. *Nat. Genet.* **23,** 194–198 (1999).

148. Gallardo, T., Shirley, L., John, G. & Castrillon, D. H. Generation of a germ cell-specific mouse transgenic Cre line, Vasa-Cre. *Genes. N. Y. N 2000* **45,** 413–417 (2007).

149. Sadate-Ngatchou, P. I., Payne, C. J., Dearth, A. T. & Braun, R. E. Cre Recombinase Activity Specific to Postnatal, Premeiotic Male Germ Cells in Transgenic Mice. *Genes. N. Y. N 2000* **46,** 738–742 (2008).

150. de Rooij, D. G. & Grootegoed, J. A. Spermatogonial stem cells. *Curr. Opin. Cell Biol.* **10,** 694–701 (1998).

151. Pepling, M. E. From primordial germ cell to primordial follicle: mammalian female germ cell development. *genesis* **44,** 622–632 (2006).

152. Burt, A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B Biol. Sci.* **270,** 921–928 (2003).

153. Deredec, A., Burt, A. & Godfray, H. C. J. The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics* **179,** 2013–2026 (2008).

154. Unckless, R. L., Clark, A. G. & Messer, P. W. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. *Genetics* **205,** 827–841 (2017).

155. Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M. & Nowak, M. A. Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3,** e1601964 (2017).

156. Marshall, J. M., Buchman, A., Sánchez C., H. M. & Akbari, O. S. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci. Rep.* **7,** 3776 (2017).

157. Noble, C., Adlam, B., Church, G. M., Esvelt, K. M. & Nowak, M. A. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* **7,** e33423 (2018).

158. Jensen-Seaman, M. I. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. *Genome Res.* **14,** 528–538 (2004).

159. Ericsson, A. C., Crim, M. J. & Franklin, C. L. A brief history of animal modeling. *Mo. Med.* **110,** 201–205 (2013).

160. Callaway, E. Controversial CRISPR 'gene drives' tested in mammals for the first time. *Nature* **559,** 164–164 (2018).

161. Scudellari, M. Self-destructing mosquitoes and sterilized rodents: the promise of gene drives. *Nature* **571,** 160–162 (2019).

162. Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E. & Cooper, K. L. Super-Mendelian inheritance mediated by CRISPR–Cas9 in the female mouse germline. *Nature* **566,** 105–109 (2019).

163. López Del Amo, V., Bishop, A. L., Sánchez C., H. M., Bennett, J. B., Feng, X., Marshall, J. M., Bier, E. & Gantz, V. M. A transcomplementing gene drive provides a flexible platform for laboratory investigation and potential field deployment. *Nat. Commun.* **11,** 352 (2020).

164. Li, G., Zhang, X., Zhong, C., Mo, J., Quan, R., Yang, J., Liu, D., Li, Z., Yang, H. & Wu, Z. Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Sci. Rep.* **7,** 8943 (2017).

165. Hu, Z., Shi, Z., Guo, X., Jiang, B., Wang, G., Luo, D., Chen, Y. & Zhu, Y.-S. Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR-Cas9 and ssODN in human cancer cells. *Cell Biosci.* **8,** 12–12 (2018).

166. Janssen, J. M., Chen, X., Liu, J. & Gonçalves, M. A. F. V. The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. *Mol. Ther. - Nucleic Acids* **16,** 141–154 (2019).

167. Gu, B., Posfai, E. & Rossant, J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat. Biotechnol.* **36,** 632–637 (2018).

168. Raveux, A., Vandormael-Pournin, S. & Cohen-Tannoudji, M. Optimization of the production of knock-in alleles by CRISPR/Cas9 microinjection into the mouse zygote. *Sci. Rep.* **7,** 42661 (2017).

169. Carlson-Stevermer, J., Abdeen, A. A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M. & Saha, K. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8,** 1711 (2017).

170. Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* **1,** 54 (2018).

171. Metz, C. W. Chromosome studies on the Diptera. II. The paired association of chromosomes in the Diptera, and its significance. *J. Exp. Zool.* **21,** 213–279 (1916).

172. Romanienko, P. J. & Camerini-Otero, R. D. The Mouse Spo11 Gene Is Required for Meiotic Chromosome Synapsis. *Mol. Cell* **6,** 975–987 (2000).

173. Comeron, J. M., Ratnappan, R. & Bailin, S. The Many Landscapes of Recombination in Drosophila melanogaster. *PLOS Genet.* **8,** e1002905 (2012).

174. Duret, L. & Galtier, N. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu. Rev. Genomics Hum. Genet.* **10,** 285–311 (2009).

175. Osada, N. & Innan, H. Duplication and Gene Conversion in the Drosophila melanogaster Genome. *PLOS Genet.* **4,** e1000305 (2008).

176. Williams, A. L., Genovese, G., Dyer, T., Altemose, N., Truax, K., Jun, G., Patterson, N., Myers, S. R., Curran, J. E., Duggirala, R., Blangero, J., Reich, D. & Przeworski, M. Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *eLife* **4,** e04637 (2015).

177. de Vries, F. A. T., de Boer, E., van den Bosch, M., Baarends, W. M., Ooms, M., Yuan, L., Liu, J.-G., van Zeeland, A. A., Heyting, C. & Pastink, A. Mouse Sycp1 functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. *Genes Dev.* **19,** 1376–1389 (2005).

178. MacLennan, M., Crichton, J. H., Playfoot, C. J. & Adams, I. R. Oocyte development, meiosis and aneuploidy. *Plasma Membr. Repair Dev. Pathol. Gonad* **45,** 68–76 (2015).

179. Pan, B. & Li, J. The art of oocyte meiotic arrest regulation. *Reprod. Biol. Endocrinol.* **17,** 8 (2019).

180. Goetz, P., Chandley, A. C. & Speed, R. M. Morphological and temporal sequence of meiotic prophase development at puberty in the male mouse. *J. Cell Sci.* **65,** 249 (1984).

181. Oakberg, E. F. A description of spermiogenesis in the mouse and its use in analysis of the cycle of the seminiferous epithelium and germ cell renewal. *Am. J. Anat.* **99,** 391–413 (1956).

182. de Kretser, D. M., Loveland, K. L., Meinhardt, A., Simorangkir, D. & Wreford, N. Spermatogenesis. *Hum. Reprod.* **13,** 1–8 (1998).

183. Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G., Burt, A., Nolan, T. & Crisanti, A. A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nat. Biotechnol.* (2020). doi:10.1038/s41587-020-0508-1

184. Barbaric, I., Miller, G. & Dear, T. N. Appearances can be deceiving: phenotypes of knockout mice. *Brief. Funct. Genomics* **6,** 91–103 (2007).

185. Liao, B.-Y. & Zhang, J. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci.* **105,** 6987 (2008).

186. Mou, H., Smith, J. L., Peng, L., Yin, H., Moore, J., Zhang, X.-O., Song, C.-Q., Sheel, A., Wu, Q., Ozata, D. M., Li, Y., Anderson, D. G., Emerson, C. P., Sontheimer, E. J., Moore, M. J., Weng, Z. & Xue, W. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* **18,** 108–108 (2017).

187. Sui, T., Song, Y., Liu, Z., Chen, M., Deng, J., Xu, Y., Lai, L. & Li, Z. CRISPR-induced exon skipping is dependent on premature termination codon mutations. *Genome Biol.* **19,** 164–164 (2018).

188. El-Brolosy, M. A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G. L. M., Takacs, C. M., Lai, S.-L., Fukuda, R., Gerri, C., Giraldez, A. J. & Stainier, D. Y. R. Genetic compensation triggered by mutant mRNA degradation. *Nature* **568,** 193–197 (2019).

189. Mak, I. W., Evaniew, N. & Ghert, M. Lost in translation: animal models and clinical trials in cancer treatment. *Am. J. Transl. Res.* **6,** 114–118 (2014).

190. Cooke, G. E. Pharmacogenetics of multigenic disease: Heart disease as an example. *Pharmacogenetics Heart Dis.* **44,** 66–74 (2006).

191. Bertram, L. & Tanzi, R. E. in *Prog. Mol. Biol. Transl. Sci.* (ed. Teplow, D. B.) **107,** 79–100 (Academic Press, 2012).

192. Bottino, R. & Trucco, M. Multifaceted Therapeutic Approaches for a Multigenic Disease. *Diabetes* **54,** S79–S86 (2005).

193. Dong, F., Xie, K., Chen, Y., Yang, Y. & Mao, Y. Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. *Biochem. Biophys. Res. Commun.* **482,** 889–895 (2017).

194. Port, F. & Bullock, S. L. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat. Methods* **13,** 852–854 (2016).

195. Shiraki, T. & Kawakami, K. A tRNA-based multiplex sgRNA expression system in zebrafish and its application to generation of transgenic albino fish. *Sci. Rep.* **8,** 13366 (2018).

196. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3570–3575 (2015).

197. Hoshijima, K., Jurynec, M. J., Klatt Shaw, D., Jacobi, A. M., Behlke, M. A. & Grunwald, D. J. Highly Efficient CRISPR-Cas9-Based Methods for Generating Deletion Mutations and F0 Embryos that Lack Gene Function in Zebrafish. *Dev. Cell* **51,** 645-657.e4 (2019).

198. van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., Reece-Hoyes, J. S., Nye, C., Gradia, S., Vidal, B., Zheng, J., Hoffman, G. R., Fuller, C. K. & May, A. P. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63,** 633–646 (2016).

199. Lemos, B. R., Kaplan, A. C., Bae, J. E., Ferrazzoli, A. E., Kuo, J., Anand, R. P., Waterman, D. P. & Haber, J. E. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl. Acad. Sci. U. S. A.* **115,** E2040–E2047 (2018).

200. Mehta, A. & Haber, J. E. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harb. Perspect. Biol.* **6,** a016428–a016428 (2014).

201. Song, Y., Lai, L. & Li, Z. Large-scale genomic deletions mediated by CRISPR/Cas9 system. *Oncotarget* **8,** 5647–5647 (2017).

202. Zheng, Q., Cai, X., Tan, M. H., Schaffert, S., Arnold, C. P., Gong, X., Chen, C.-Z. & Huang, S. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *BioTechniques* **57,** 115–124 (2014).

203. Chen, X., Xu, F., Zhu, C., Ji, J., Zhou, X., Feng, X. & Guang, S. Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in Caenorhabditis elegans. *Sci. Rep.* **4,** 7581 (2014).

204. Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. & Church, G. M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31,** 833–838 (2013).

205. Anderson, E. M., Haupt, A., Schiel, J. A., Chou, E., Machado, H. B., Strezoska, Ž., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A. & Smith, A. van B. Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* **211,** 56–65 (2015).

206. Zheng, T., Hou, Y., Zhang, P., Zhang, Z., Xu, Y., Zhang, L., Niu, L., Yang, Y., Liang, D., Yi, F., Peng, W., Feng, W., Yang, Y., Chen, J., Zhu, Y. Y., Zhang, L.-H. & Du, Q. Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci. Rep.* **7,** 40638 (2017).

207. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. & Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

208. Grainger, S., Lonquich, B., Oon, C. H., Nguyen, N., Willert, K. & Traver, D. CRISPR Guide RNA Validation *In Vitro*. *Zebrafish* **14,** 383–386 (2017).

209. Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M. & Noce, T. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech. Dev.* **93,** 139–149 (2000).

210. Pfitzner, C., Hughes, J., White, M., Scherer, M., Piltz, S. & Thomas, P. *Development of zygotic and germline gene drives in mice*. (Genetics, 2020).

211. Yoshida, K., Kondoh, G., Matsuda, Y., Habu, T., Nishimune, Y. & Morita, T. The Mouse RecA-like Gene Dmc1 Is Required for Homologous Chromosome Synapsis during Meiosis. *Mol. Cell* **1,** 707–718 (1998).

212. Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Jr, Kolodner, R. D., Kucherlapati, R., Pollard, J. W. & Edelmann, W. MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes Dev.* **14,** 1085–1097 (2000).

213. de Vries, S. S., Baart, E. B., Dekker, M., Siezen, A., de Rooij, D. G., de Boer, P. & te Riele, H. Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev.* **13,** 523–531 (1999).

214. Yuan, L., Liu, J.-G., Zhao, J., Brundell, E., Daneholt, B. & Höög, C. The Murine SCP3 Gene Is Required for Synaptonemal Complex Assembly, Chromosome Synapsis, and Male Fertility. *Mol. Cell* **5,** 73–83 (2000).

215. Menke, D. B., Koubova, J. & Page, D. C. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Dev. Biol.* **262,** 303–312 (2003).

216. Zhou, Q., Li, Y., Nie, R., Friel, P., Mitchell, D., Evanoff, R. M., Pouchnik, D., Banasik, B., McCarrey, J. R., Small, C. & Griswold, M. D. Expression of stimulated by retinoic acid gene 8 (Stra8) and maturation of murine gonocytes and spermatogonia induced by retinoic acid in vitro. *Biol. Reprod.* **78,** 537–545 (2008).

217. Zhang, J., Chen, L., Zhang, J. & Wang, Y. Drug Inducible CRISPR/Cas Systems. *Comput. Struct. Biotechnol. J.* **17,** 1171–1177 (2019).