UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Development and Application of Genome Editing Approaches to Investigate Endogenous Retroviruses

Permalink

https://escholarship.org/uc/item/2vs2c834

Author

Chen, Sean

Publication Date

2017

Peer reviewed|Thesis/dissertation

Development and Application of Genome Editing Approaches to Investigate
Endogenous Retroviruses

by

Sean Chen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Lin He, Chair
Professor Henk Roelink
Professor Dirk Hockemeyer
Professor Michael W. Nachman

Fall 2017

Abstract

Development and Application of Genome Editing Approaches to Investigate
Endogenous Retroviruses

by

Sean Chen

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Lin He, Chair

Endogenous retroviruses (ERVs) constitute a significant fraction of mammalian genomes, but their impact on host biology remains poorly understood. One group of ERVs, murine ERV with leucine tRNA primer (MERVL), is highly expressed during the 2-cell (2C) stage of mouse preimplantation development but is silenced thereafter. While active, MERVL-derived *cis*-regulatory elements drive expression of hundreds of host genes, including "chimeric" isoforms with exonized MERVL sequences. Remarkably, loss of a single miRNA *miR-34a* in pluripotent stem cells is sufficient to derepress MERVL and imbue expanded fate potential reminiscent of totipotent 2C blastomeres. Using bioinformatic prediction and reporter assays, I identified *gata2* as the primary target of miR-34a that mediates MERVL derepression in pluripotent stem cells. While miR-34a is required for MERVL silencing in pluripotent stem cells, it is dispensable for repressing MERVL during preimplantation development. To evaluate the role of MERVL *in vivo,* I applied a candidate approach to assess the role of one MERVL-driven chimeric gene *tead4:MT2B1*. Loss of *tead4:MT2B1* results in compensatory upregulation of canonical *tead4* transcripts, suggesting that the chimeric isoform functions redundantly in development. In order to facilitate efficient *in vivo* screening of additional candidates, I developed a high-throughput electroporation-based genome editing technique called CRISPR RNP Electroporation of Zygotes (CRISR-EZ). Compared to previous methodologies, CRISPR-EZ offers significant advantages in throughput, cost, and simplicity. Altogether, I have elucidated a molecular axis involved in the regulation of MERVL and fate potency, setting the stage for further *in vivo* characterization using improved genome editing tools.

1

Table of Contents

List of Tables and Figures

Chapter 1
Introduction

Overview of mammalian endogenous retroviruses

In 1950, Barbara McClintock published a groundbreaking study examining mosaic kernel color variation in maize, wherein she proposed that mobile "controlling elements" could alter expression of adjacent genes.[1] Widely dismissed by her contemporaries, the significance of her findings was not appreciated until transposition was later observed in bacteria, leading to the discovery of transposable elements (TEs).[2] It is now recognized that TEs occupy a significant fraction of all eukaryotic genomes—12% in *C. elegans,* ~40% in mammals, and up to 90% in some plants.[3–5] Arguments to rationalize the pervasiveness of TEs have fallen under two doctrines. The writings of Doolittle and Sapienza, along with those of Orgel and Crick, depicted TEs as selfish genomic parasites whose propagation comes at the expense of host fitness.[6,7] In contrast, McClintock and others believed that these "controlling elements" confer regulatory complexity to host genomes.[8] In more recent years, these conflicting views have converged on a paradigm in which TEs, while capable of causing genomic instability, have nonetheless become domesticated into host gene regulatory networks, and may act as powerful drivers of genome evolution.[2]

TEs are broadly divided into DNA transposons and retrotransposons. DNA transposons undergo excision and reintegration through a self-encoded transposase enzyme, a "cut-and-paste" mechanism that does not result in increased copy number.[9] In contrast, retrotransposons propagate by a "copy-and-paste" mechanism, utilizing an RNA intermediate that is reverse-transcribed into DNA before integration into a new genomic locus.[10] Thus, retrotransposons can accumulate within host genomes, and in mammals retrotransposons vastly outnumber other TEs.[11] Retrotransposons are further classified as long terminal repeat (LTR) retrotransposons, also called endogenous retroviruses (ERVs) in mammals, and Non-LTR retrotransposons, which are mainly comprised of Long Interspersed Elements (LINEs), and Short Interspersed Elements (SINEs).

ERVs account for approximately 10% of the human genome.[4] Remnants of ancient retroviral infection of the host germline, the vast majority of ERVs are heavily mutated and/or truncated; nevertheless, "complete" elements resemble exogenous retroviruses in structure and retrotransposition mechanism. Their 5' and 3' LTRs harbor all required signals for gene expression, including enhancer, promoter, and polyadenylation signal. The LTRs flank internal sequences containing regions called *gag*, *pro*, *pol*, and *env*, which encode viral core structural proteins, protease, reverse-transcriptase/integrase/RNAse H, and envelop protein respectively. Following transcription by RNA polymerase II, the binding of host tRNA to a region called the primer binding site (PBS) initiates reverse-transcription, a multi-step reaction that ultimately generates double-stranded DNA that may be reinserted into the host genome by viral integrase.

The most widely accepted method of classifying mammalian ERVs is based on homology of *pol* sequences to exogenous retroviruses: Class I, which resemble gamma- and epsilon-retroviruses, Class II, which resemble alpha- and beta-retroviruses, and Class III, which resemble spumaviruses.[12] However, many ambiguities remain—for example MaLR elements lack *pol* sequences but are often considered Class III on the basis of slight *gag* sequence similarity.[12] Moreover, re-classification based on *env* sequences frequently contradicts that based on *pol,* likely due to extensive recombination between ERV species.[13] Within each class, individual ERV insertions are further subdivided into groups believed to be descendants of a single retroviral infection. Some examples include the Class I MLV, VL30, and RLTR4; the Class II MMTV, intracisternal A-type particle (IAP) and MusD; and Class III MERVL and MaLR. Nomenclature for ERV groups remains confounding due to a lack of universally accepted naming criteria. Existing naming schemes are often based on homology to exogenous viruses, host species, tRNA used to prime reverse-transcription, and neighboring genes.

Although Class I, II, and III ERVs have been found in all tested mammals, their abundance both in terms of group diversity (from 1 to 20 groups) and number of individual insertions (from 1 to several thousand loci) can vary drastically between species.[14] For example, Class II ERVs have undergone dramatic amplification in the mouse, while primate genomes have a larger fraction of Class I ERVs.[12] The chronology and frequency of ERV expansion within their host genomes remains somewhat contested, but estimates have been made based on insertion synteny across different species, as well as the degree of mutation between the two LTRs of an intact ERV (which are identical upon initial insertion). From such analyses, for example, it was determined that some of the oldest Class III ERVs predate the divergence of rodents and primates, while the youngest HERVK insertions are polymorphic in humans.[15–17] It has become clear that mammalian ERVs have expanded within their hosts gradually over tens of millions of years, sometimes punctuated with bursts of amplification.[18] In mice, novel ERVs actively colonize the germline even in the present day, while in other species including human this is seldom observed.[15,19] Although this difference in activity remains poorly understood, it is likely due to the combined effects of divergent silencing mechanisms, selection, and genetic drift.

Pathological effects of ERVs

Because insertion into exons or enhancers can disrupt gene expression, transposition is generally considered deleterious to host genome integrity. Indeed, in invertebrate model organisms such as *C. elegans* and *D. melanogaster,* germline transposon reactivation results in genome instability and sterility.[20,21] Although such dramatic phenotypes are rarely observed in mammalians models, a few cases of mutagenic TE insertions underlying human disease have been reported.[22–24] Still, insertions that disrupt host exons are exceedingly rare in humans, owing to a low ERV retrotransposition rate.[25] This stands in stark contrast to the mouse, wherein at least 10% of germline mutations are thought to arise from novel ERV insertions.[26]

Perhaps the most easily understood pathology associated with ERV insertion is cancer, as exogenous retroviruses have long been recognized for their carcinogenic effects through insertional activation of proto-oncogenes. In AKR mice that develop spontaneous leukemia, endogenous MLVs generate infectious retroviruses that cause malignancy,[27] while in mice deficient for the DNA methyltransferase Dnmt1, IAP retrotransposition can activate proto-oncogenes and induce lymphoma.[28] In humans, a handful of examples of so-called "onco-exaptation" have been reported.[24,29–32] In some cases, promoter activity of LTRs drives aberrant expression of host genes—notable examples include a Class III THE1B insertion upstream of the receptor-tyrosine kinase CSF1R, and a Class I LOR1a insertion driving ectopic expression of the transcription factor IRF5, both identified in cases of Hodgkin's Lymphoma.[24,29] Alternatively, insertional mutagenesis can lead to protein truncation and loss of regulatory domains. Such was reported for a melanoma case in which insertion of the Class III LTR16B created an alternative short isoform of the receptor-tyrosine kinase ALK.[30] A similar case was shown for anaplastic large cell lymphoma in which insertion of the Class III MLT1C truncated the receptor-tyrosine kinase ERBB4.[33] It is important to qualify that the few examples highlighted here constitute almost the entire body of known carcinogenic ERV insertions. Whether such reports represent an underappreciated widespread phenomenon, or are merely exceptional cases, remains unclear.

ERVs have also become increasingly implicated in auto-immune disorders. Just as hosts mount an immune response against exogenous retroviruses, it seems plausible that ERVs could illicit similar effects under specific conditions. Indeed, mice with systemic lupus erythematosus (SLE) produce antibodies against the envelop glycoprotein of GP70 ERVs.[34] Even more

astoundingly, immunodeficient B6 mice lacking antibodies produce ecotropic MLV that can infect other mice and induce lymphomas.[35] In humans, HERVE are specifically derepressed in kidney tumors, and antigens derived from these ERVs could stimulate cytotoxic T-cells *in vitro* and *in vivo*.[36] Similarly, HERVW has been described as a biomarker for neuroinflammatory disorders including multiple sclerosis (MS), and overexpression of the HERVW-encoded protein SYNCYTIN-1 in glia phenocopies some aspects of MS including endoplasmic reticulum stress and production of free radicals.[37] While no ERVs have been directly implicated in causing human autoimmune diseases, these provocative observations merit further study into this relationship.

Mechanisms of ERV silencing

To prevent mutagenic transposition, host organisms utilize a plethora of tools to suppress ERV activity both at the transcriptional and post-transcriptional level. Foremost among these silencing mechanisms are the epigenetic machineries that facilitate heterochromatin formation through DNA and histone methylation. It has long been appreciated that DNA methylation plays a major role in ERV silencing. Injection of mice with 5-azacytidin activates ERV expression in a variety of somatic tissues including thymus, spleen, and liver.[38] Dnmt1, which is responsible for maintenance of CpG methylation on newly synthesized DNA, was shown to be essential for silencing a range of ERVs both in somatic and germ cells—mice carrying a hypomorphic mutant of Dnmt1 develop thymic lymphomas as a result of oncogenic IAP insertion into the *Notch1* locus, and loss of Dnmt1 leads to reactivation of IAP in germ cells.[28,39] During preimplantation development, global demethylation leads to rampant reactivation of several ERV groups, including MERVL, MusD, and IAP.[11,40]

Histone methylation, particularly tri and di-methylation of histone 3 lysine 9, has also been established as a universal mechanism of TE silencing in diverse organisms.[41–43] In mouse embryonic stem cells (ESCs), depletion of the H3K9me3 methyltransferase Setdb1 results in derepression of Class I and Class II ERVs,[43] while depletion of the H3K9me2 methyltransferase G9a causes derepression of Class III ERVs.[44] Although the determinants of this specificity are not completely understood, it is thought that initiating factors may recognize ERVs and subsequently recruit histone modifying proteins. For example, Kap1 recruits Setdb1 to ERVs in mouse embryonic stem cells, and *kap1* deletion results in loss of H3K9me3 as well as derepression of a range of ERVs, particularly IAP.[45] Liver-specific *kap1* knockout mice exhibit derepression of VL30 ERVs in hepatocytes and mount an antiviral interferon response mimicking that of exogenous viral infection.[46] The interaction of Kap1 with kruppel-associated box zinc finger proteins (KRAB-ZFPs) provides an attractive explanation for the specificity of ERV silencing, as mammals possess hundreds of KRAB-ZFPs that bind unique sequences.[47] It has even been proposed that the rapid diversification of KRAB-ZFPs in mammals constitutes an evolutionary "arms race" against invading exogenous retroviruses, as the number of KRAB-ZFP genes appears closely correlated with the number of ERVs in a species.[47] One notable example is the KRAB-ZFP Znf809, which is required for silencing MLV through binding its PBS in mouse ESCs.[48] Other zinc finger proteins have been shown to play similar roles. In mouse erythroid cells, Kruppel-like factor 3 (Klf3) was shown to silence ORR1AO LTRs, as loss of Klf3 resulted in widespread ERV derepression.[49] Ying Yang 1 (Yy1), a zinc finger protein that interacts with Kap1, can bind the LTRs of many endogenous and exogenous viruses, and deletion of the Yy1 binding site leads to derepression and reduction of H3K9me3 over proviral loci.[50] A zinc finger protein related to Yy1 called Rex1 was shown to regulate MERVL in both mouse ESCs and preimplantation embryos, potentially through interaction with the lysine demethylase Lsd1.[51,52] Thus, the utilization of zinc

finger proteins as sequence-specific adapters for global epigenetic machineries appears to be a common theme in ERV suppression.

In addition to these epigenetic mechanisms, several host proteins play roles in post-transcriptional ERV silencing. Many of these were first identified as restriction factors against exogenous retroviruses. Prime examples include the APOBEC3 family of cytidine deaminases, of which APOBEC3G and APOBEC3F were identified as potent inhibitors of HIV-1 by viral genome hypermutation and reverse-transcriptional interference.[53,54] Many ERVs harbor signatures of APOBEC3-mediated cytidine deamination, with some insertions carrying hundreds of mutations, hinting at a substantial role in ERV silencing.[55–57] Another major restriction factor is TRIM5alpha, which binds to viral capsid protein and disrupts core stability, reverse-transcription, and nuclear entry.[58,59] Additionally, Tetherin is an extracellular restriction factor that prevents release of enveloped viruses by crosslinking budded virions to the cell membrane.[60,61] It has been proposed that sequence differences between ERVs and their ancestral retroviruses may actually reflect adaptations to evade restriction factors—the lack of *env* genes and a mutation in *gag* genes of IAP and MusD elements allows intracellular assembly and budding without being targeted by extracellular restriction factors.[62]

Small RNAs are also recognized as major regulators of TEs, particularly in the germline. Such mechanisms have been extensively characterized in invertebrate models. In *C. elegans*, natural dsRNAs generated from read-through transcription of Tc1 elements trigger RNAi–dependent silencing of transposition.[3] Loss of Argonaute (Ago) or Dicer causes derepression of TEs in many organisms,[63] but the significance of these endogenous siRNA (endo-siRNA) pathways in mammalian species is less understood. Very recent work has shown that, upon loss of Dnmt1 in mouse ESCs, sense and anti-sense transcripts derived from TEs can form dsRNA, which are processed by Dicer to generate endo-siRNAs.[64] Immunoprecipitation of Ago2 followed by small RNA-Seq revealed that these endo-siRNAs are engaged by the RNA-induced silencing complex (RISC), and knockdown of Ago2 led to increased expression of TEs, including many ERVs.[65] These findings suggest that endo-siRNAs may act as a secondary defense against ERV activation under hypomethylated conditions.

Another distinct class of small RNAs involved in TE silencing are the Piwi-interacting RNAs (piRNAs). Called repeat-associated small interfering RNAs (rasiRNAs) in *Drosophila*, they comprise the largest class of small RNAs expressed in the germline, and play a crucial role in silencing transposition.[66] Mature piRNAs (~26-31 nt in length) are generally processed via a Dicer-independent pathway from single-stranded precursors, which are transcribed from TE-rich piRNA clusters.[67,68] Interacting with members of the Piwi clade of Argonaute proteins, including Piwi, Aubergine (Aub), and Argonaute 3 (Ago3)*,* piRNAs function in both transcriptional and post-transcriptional repression of TEs. Disruption of this pathway in *Drosophila* leads to mobilization of gypsy class ERVs and germline deterioration.[69] As with invertebrates, the mammalian Piwi proteins Miwi and Mili also play integral roles in piRNA-mediated transposon silencing. Analysis of pre-pachytene piRNAs in mouse spermatocytes revealed that more than a third were complementary to transposons, and *Mili* knockout mutants derepressed LINE-1 and IAP expression by 5-10 fold.[70] Furthermore, null mutants of *Miwi* and *Mili* show defects in spermatogenesis, indicating the importance of this pathway in germ cell maturation. It is important to note that epigenetic and small RNA-dependent silencing mechanisms are closely intertwined in many organisms. For example, small RNA-dependent recruitment of H3K9me3 machinery has been demonstrated in fission yeast, flies, and nematodes.[71–73] Interestingly, loss of *Mili* in mouse

testes led to substantial demethylation of TEs, suggesting that piRNAs may guide methylation in the mouse germline.[70]

Along with siRNA and piRNA, microRNAs (miRNAs) play an increasingly recognized role in the regulation of TEs. miRNAs are a class of small RNAs that mediate posttranscriptional gene silencing through the combined mechanisms of translational repression and mRNA degradation,[67,68] and are involved in a broad range of biological processes.[74–76] In mice, specific classes of TEs become derepressed upon loss of Dgcr8 or Dicer.[77,78] Intriguingly, many miRNA are generated from TE-derived transcripts, reminiscent of siRNA biogenesis in *C. elegans* and *Drosphila*,[79] but functional studies remain scarce. It was recently shown that miR-155 expression could suppress *env* genes of the avian ERV ALVE in a tissue-specific manner.[80] In both mouse ESCs and iPSCs, loss of *miR-34a* leads to highly specific derepression of MERVL as well as its neighboring genes. The mechanism of this regulation appears to be indirect—*miR-34a* targets the transcription factor Gata2, which in turn is required for MERVL induction.[81] It remains to be determined whether the regulation of ERVs by miRNAs extends beyond these few reported cases.

Lastly, accumulating evidence has begun to implicate transfer RNA-derived fragments (tRFs) in ERV suppression. tRFs are a structurally diverse class of small RNAs generated from endonucleolytic cleavage of both mature and precursor tRNAs.[82,83] Although tRF biogenesis remains an area of active investigation, in many cases it appears to share much of the same core machinery as canonical miRNA processing (Dicer, Drosha, and Dgcr8).[84–86] In one study, a species of tRF found in sperm, designated as 5' tRNA-Gly-CCC, was shown to downregulate ERVs *in vitro* and *in vivo*—antisense oligos that interfered with tRNA-Gly-CCC caused upregulation of the Class III ERV MERVL in both ESCs and preimplantation embryos.[87] Astonishingly, tRNA-Gly-CCC was upregulated in the sperm of males fed a low-protein diet, and oocytes fertilized with such sperm generated embryos with reduced levels of MERVL.[87] These findings appear to suggest that parental diet can influence ERV expression in offspring through sperm-born tRFs. Another study identified abundant tRFs produced in ESCs and TSCs lacking the ERV-silencing histone methyltransferase Setdb1. Two classes of 3' tRF were detected: 22 nt tRFs, which induce RNAi against IAP and MusD elements; and 18 nt tRFs, which bind to the PBS of these elements and strongly inhibit their reverse-transcription.[88] Thus, in the absence of H3K9me3 silencing marks, tRFs may confer an additional layer of regulation against ERV activity.

Host co-option of ERV proteins

Given the prevalence of ERVs in mammalian genomes and their myriad gene regulatory activities, it is impossible to overlook the possibility that a subset ERVs may become domesticated to enhance host fitness. Examples of ERV exaptation fall into two main categories: ERV-derived protein co-option, and host gene regulatory co-option. A smaller number of cases fall into a third category, in which ERV transcripts act as functional long non-coding RNAs (lncRNAs), which will be discussed later in more detail.

In no tissue is ERV protein domestication more evident than in the placenta. During embryonic development, the developing placenta undergoes continuous syncytial fusion to form the syncytiotrophoblast layer that invades the uterine wall. Molecular players in this process remained elusive until, after extensive searching, two ERV-derived proteins were discovered with fusogenic activity in human trophoblast cells.[89–91] Sequence analysis revealed that these membrane glycoproteins, designated SYNCYTIN-1 and SYNCITIN-2, were derived from the *env* genes of HERVW and HERV-FRD, respectively, and siRNA knockdown of *Syncytin-2* abrogated cell fusion in a trophoblast cell line.[92] In mice, a distinct set of *Syncytin* genes derived from independent

co-option events, called *Syncytin A* and *Syncytin B,* also play crucial roles in placental trophoblast fusion, with *Syncytin A* null mutants exhibiting embryonic lethal placental defects.[93,94] It is now clear that ERV-derived envelop proteins have been co-opted independently at least seven times throughout the evolution of placental mammals.[89–99] Surprisingly, a *Syncytin* gene was even recently identified in a non-mammalian species, the placental *Mabuya* lizard, where it is expressed in the placenta and induces cell fusion.[100] It is thus likely that the striking diversity of placental morphology, function, and development can be largely attributed to ERV domestication.[101,102]

Somewhat paradoxically, several ERV-derived proteins have become repurposed by their hosts to defend against exogenous viral infection. The best known example of these restriction factors is the murine *Fv1* gene. Sharing approximately 40% sequence similarity to MERVL *gag*, Fv1 binds the capsid of retroviruses and is thought to block proviral integration.[103,104] Another *gag*-related protein, derived from Jaagsiekte sheep retrovirus (enJSRV), functions as a dominant negative to inhibit viral trafficking and virion release.[57] Aside from *gag,* restriction factors derived from *env* genes have also been reported, including enFeLV in cats, as well as *Fv4* and *Rmcf2* in mice.[105–107] In humans, it has been reported that HERVK is actively transcribed during normal preimplantation development, and that overexpression of its Rec protein in a pluripotent cell line increases levels of the antiviral protein IFITM1. It was thus proposed that HERVK serves a protective function against viral infection during early human development.[108]

Regulatory exaptation of ERVs

Far outnumbering known protein domestications are cases of ERVs contributing novel regulatory modalities to their surrounding host genes.[109] As the LTRs of ERVs have evolved to utilize host transcription factors to facilitate viral replication, is it easy to see how these signals can become co-opted as *cis*-regulatory elements in normal host physiology. While the majority of ERV insertions have neutral or deleterious effects on host fitness, rare integrations can become fixed in a population through evolutionary selection.[10] Some of the first discovered regulatory ERVs include an HERVE LTR insertion that allows a pancreatic amylase gene paralog to become expressed in human saliva,[110] as well as an LTR dubbed *Imp1* that confers androgen regulation to the mouse sex-limited protein (*Slp*) gene.[111,112] Another well-studied example is an ERV9 LTR integration near the locus control region upstream of the human beta-globin gene. This LTR possesses multiple binding cites for ubiquitous and hematopoietic transcription factors including NF-Y, MZF1, and GATA2, and exhibits potent long-range enhancer activity in erythroid cells.[113–115] Interestingly, it was shown that the ERV9 LTR drives expression of intergenic lncRNAs in this region, though the functional contribution of these transcripts to its enhancer activity remains unclear.[115]

Given the high copy numbers of certain ERV groups, it is immediately tempting to envision that ERVs within the same group can exert concerted regulation across large swathes of the genome, simultaneously activating hundreds or even thousands of genes. In this manner, a specific tissue or cell type can gain access to entirely new gene regulatory networks through LTR co-option. The use of high-throughput genomic approaches has provided strong evidence of this type of genome rewiring.[116–119] Like ERV protein domestication, one tissue in which this has repeatedly occurred is the placenta. ChIP-Seq studies comparing mouse and rat trophoblast stem cells (TSCs) revealed that species-specific enhancers are highly enriched for ERVs, and it was shown that one such ERV group, RLTR13D5, contributes hundreds of mouse-specific enhancers that bind core TSC transcription factors including Cdx2, Eomes, and Elf5.[116] Numerous instances of human placenta-specific gene expression driven by ERVs have been documented, including leptin via a

MER11 insertion, pleiotrophin via a Class I HERV insertion, and interleukin 2 (IL-2) receptor beta subunit via a THE1D insertion.[120–122] In chicken ESCs, the Ens-1 LTR provides binding sites for Gata4, Nanog, and Ets1 that may prime the specification of extraembryonic lineages.[123] It has been proposed that trophoblast-specific tolerance of ERVs is a direct outcome of the integral role played by ERVs in placenta evolution.[102]

Evidence of LTR-regulated gene networks have been described in other contexts as well, including innate immunity, pluripotency, oogenesis, early development, and tumor suppression.[11,117,124,125] It was shown through ChIP-Seq analysis that ERV LTRs may be integral components of the human interferon (IFN) transcriptional network, providing thousands of IRF1 and STAT1 binding sites near IFN-inducible genes. CRISPR-mediated ablation of one such element, MER41, reduced the interferon response in HeLa cells.[124] In human ESCs, around 20% of OCT4 and 15% of NANOG binding sites can be attributed to TEs, with Class I ERVs comprising the largest fraction, and RNAi knockdown of *OCT4* led to reduced expression of genes near these TEs.[117] The same study also showed that comparison between human and mouse ESCs showed less than 5% conservation of binding sites, suggesting that species-specific TEs substantially influenced their divergent pluripotency transcriptional circuitries.[117] Indeed, it appears that the majority of primate-specific enhancers may be derived from TEs.[118] In the oocyte and in early embryogenesis, many ERVs are expressed in great abundance, and it has been estimated that 5% of host genes expressed during these stages are driven by LTRs.[11,126] Overall, it is clear that ERVs have significantly impacted the regulatory landscape of mammalian genomes.

In addition to *cis*-regulatory effects, ERVs also frequently undergo a variety of exonization schemes, whereby they contribute alternative promoters, splice sites, and polyadenylation signals to neighboring host genes. Such "chimeric" gene isoforms are generally more easily detected than ERV enhancer effects, owing to the presence of unique sequenced junctions between the ERV insertion and host exon. Perhaps the most convincing case of a developmentally functional chimeric gene is an oocyte-specific isoform of Dicer *(*Dicer^O*)*. In mice, an MTC element insertion between exons 6 and 7 acts as an alternative promoter exclusively in oocytes, resulting in a truncated isoform lacking the N-terminal DExD helicase domain. Unlike full-length somatic Dicer (Dicer^S), Dicer^O efficiently cleaves dsRNA into endo-siRNAs. Genetic ablation of the MTC element leads to increased levels of siRNA targets, meiotic spindle defects, and female sterility.[64] Another interesting example is an ERV9-driven isoform of p63, the only isoform expressed in the human male germline. This chimeric p63 isoform was shown to be responsive to DNA damage, inducing apoptosis upon caspase cleavage.[125] In some instances, LTR-derived promoters have entirely replaced the putative ancestral host promoter. For example, in the rat, an IAP LTR constitutes the sole promoter of *Ocm2*.[127] In rodents, an ORR1E LTR acts as the major promoter for the anti-apoptotic gene *Naip*, in contrast to humans, where an LTR acts as a testes-specific promoter.[128] Owing to high-throughput sequencing technologies and bioinformatic prediction, hundreds of mammalian chimeric isoforms have now been identified, including genes involved in germline and preimplantation development, metabolism, and hematopoiesis.[11,49,109,129–133] However, direct proof of function remains scarce, possibly due to inadequate *in vivo* genetic tools. In the absence of functional characterization in animal models, it remains challenging to unambiguously distinguish spurious transcriptional noise from *bona fide* promoter co-option.

An accumulating body of work is beginning to implicate ERV-derived lncRNAs in gene regulation. Genome-scale cap analysis of gene expression (CAGE) analysis revealed that tens of thousands of anti-sense transcripts initiate from TEs in human,[134] and depending on tissue 30-80% of human lncRNAs contain TE sequences.[135–137] Although it is unclear what fraction of these

transcripts serve important functions, some interesting examples have been reported. As previously mentioned, the ERV9 LTR upstream of the human globin locus generates intergenic transcripts, and it has been proposed that this unidirectional transcription guides the long-range looping interactions mediating its enhancer activity.[115] In human ESCs, HERVH expression is both abundant and highly specific,[138,139] and it has been reported that HERVH acts as an enhancer lncRNA to recruit coactivators and pluripotency factors such as OCT4. siRNA knockdown of HERVH led to reduced levels of pluripotency markers and differentiation.[140] A caveat of many of these studies is the difficulty of uncoupling enhancer RNA (eRNA) activity from ERV *cis*-regulatory effects. More rigorous investigation is needed to conclusively assess the role of ERV-derived lncRNAs.

ERVs: markers or regulators of fate potency?

A relatively new area of interest is the burgeoning field of ERVs involved in developmental fate potential. A flurry of publications over the last few years have mainly implicated two ERVs, the previously mentioned HERVH in human and MERVL in the mouse, as markers and/or regulators of unique developmental states, predominantly in pluripotent stem cells but also in preimplantation stage embryos. What follows is a review of the current literature and a critical examination of the presented evidence.

HERVH is a primate-specific Class I ERV, named for its use of a histidine tRNA to prime reverse transcription.[141,142] Although HERVH elements have been detected in all primates, significant expansion occurred in the Old World Monkey lineage, with 80% of human HERVH integrations occurring within the last 30 million years.[143–145] In terms of sheer abundance, HERVH is among the most highly expressed genes in human ESCs, constituting an estimated 2% of all polyadenylated RNAs.[138] Its exclusive expression in human ESCs led to the designation of HERVH as pluripotency marker.[138] HERVH elements are closely associated with open chromatin marks in ESCs but not in any other analyzed cell type, and levels of HERVH RNA drop precipitously upon differentiation.[138] It was later shown that HERVH is not ubiquitously expressed in human ESCs, but marks a subpopulation that manifests properties of "naïve" stem cells.[139] Compared to the mouse, traditional human ESC derivation methods generate ESCs of limited developmental potential. These "primed" ESCs are more akin to mouse epiblast stem cells (EpiSCs), being capable of unlimited self-renewal and differentiation into the three germ layers, but being incapable of forming germline chimeras.[146] In contrast, naïve human stem cells can recapitulate all aspects of pluripotency and were long considered the "holy grail" of human ESCs.[147–149] Thus the discovery of HERVH as a molecular marker of naïve human ESCs generated considerable enthusiasm. It was reported that the flanking LTR7 elements of HERV provide binding sites for naïve ESC-specific transcription factors including LBP9, as well as pluripotency factors OCT4 and NANOG, which drive expression of chimeric transcripts including many lncRNA.[119] Depletion of LBP9 or HERVH resulted in markedly altered cell morphology and loss of self-renewel.[119,140] While exciting, these findings have since been contradicted by transcriptome analyses comparing naïve and primed human ESCs with early human embryos—in fact, these investigators found that HERVH/LTR7 is more strongly associated with the primed state.[150] Such conflicting results may reflect differences in HERVH mapping and annotation, naïve ESC derivation methodologies, or bioinformatic analysis pipeline.

In parallel to these observations, a few studies had also identified HERVH and LTR7 as players in somatic reprogramming. LTR7 expression increases throughout the course of reprogramming, becoming transiently hyperactivated above ESC levels immediately prior to the

iPSC stage, before dropping back down upon completion of reprogramming.[151] However, as global TE deregulation appears to be a general feature of iPSC generation, more investigation was needed to determine whether HERVH was required for this process. Additional evidence would come from experimental manipulation of HERVH levels in the context of somatic reprogramming— overexpression of LBP9 or a subset of HERVH-driven lncRNAs accelerated iPSC maturation, while depletion of a HERVH-driven lncRNA reduced the efficiency of iPSC generation.[139,152]

While much *in vitro* characterization of HERVH has been performed, *in vivo* investigation remains unsurprisingly scant. In a screen to identify functional TE-derived lncRNAs, one study identified three candidate human pluripotency-associated transcripts (HPATs) as possible regulators of cell fate. Of the three candidates (designated HPAT2, HPAT3, and HPAT5), HPAT2 and HPAT3 were derived from HERVH, while HPAT5 was derived from the unrelated LTR8. HPAT3 and HPAT5 were both detected by RNA FISH in human blastocysts. Strikingly, when siRNAs targeting all three candidates were injected into a single blastomere of 2-cell embryos, siRNA-injected cells failed to contribute to the inner cell mass (ICM) of the resulting blastocysts, while control injected cells contributed to both ICM and trophectoderm (TE) lineages. While these remarkable findings suggest that HERVH-derived lncRNAs could function in maintaining pluripotency *in vivo,* this study has some notable caveats. Since siRNAs targeting all three candidates were co-injected simultaneously, it is impossible to ascertain which of the three candidates were responsible for resulting phenotype, and the small sample size of this experiment (n=3 blastocysts) precludes making definitive conclusions.

Nearly concurrent with the commotion surrounding HERVH in human ESCs, a similar trend was unfolding around MERVL in mouse ESCs. ERVL (ERV with leucine tRNA primer) is an ancient (70 MYA) Class III family of mammalian ERVs with approximately 200 intact copies in human (HERVL) and 700 copies in mice (MERVL).[10,153] However, the number of incomplete and solo LTRs greatly outnumber the intact elements, with nearly 1800 such loci in the mouse genome.[81,153] Since most placental mammals retain ERVL at much lower copy number, HERVL and MERVL are believed to have undergone independent amplification in their respective host lineages.[10] The expansion of MERVL in mice appears to have occurred recently, as rats maintain ancestral copy numbers of ERVL.[153]

In mice, families of ERVs are highly expressed in distinct waves during early development.[11] MERVL is among the first zygotically expressed genes, reaching maximal activity at the 2-cell (2C) stage where it constitutes roughly 3% of mRNAs.[11,154] After the 2C stage, MERVL is rapidly silenced, and these elements were thought to remain inactive henceforth until a landmark study by Macfarlan *et. al.* demonstrated that a subpopulation of mouse ESCs retains high MERVL activity.[129] These MERVL-positive cells not only possess a molecular profile similar to 2C embryos, but also functionally recapitulate the fate potential of 2C blastomeres in chimeric embryos—while MERVL-negative mouse ESCs contributed only to the ICM of reconstituted blastocysts, MERVL-positive embryos colonized both the ICM and the TE.[129] As pluripotent ESCs were thought to solely give rise to ICM derivatives, these findings suggest that MERVL marks a rare population of ESCs with expanded fate potential resembling totipotency.

The significance of this work cannot be understated, as it provided the first hints that totipotent stem cells could be obtained *in vitro,* a prospect that had eluded investigators for decades. Within a short time, several other mouse ESC populations with similar molecular and functional characteristics were discovered.[81,155–159] Genetic ablation of epigenetic regulators, including the histone demethylase Lsd1 and chromatin assembly factor Caf1, led to upregulation of MERVL and adoption of 2C-like properties.[158,160] While Macfarlan *et. al*. refrained from calling

MERVL-positive ESCs "totipotent," other investigators were less hesitant to brandish this designation. Morgani *et. al*. enriched for a subpopulation of ESCs using the extraembryonic endoderm marker *Hex*. These cells had high levels of MERVL and contributed to both ICM and TE derivatives in chimeric embryos even at the single-cell level, leading the authors to designate these cells as *bona fide* totipotent stem cells.[155] Choi *et. al*. identified the first miRNA involved in this pathway, showing that *miR-34a* knockout ESCs and iPSCs derepressed MERVL and contributed to all cell lineages in chimeric embryos.[81] While most studies were performed using cell lines, attempts were also made to address the role of MERVL *in vivo*. These studies utilized antisense oligos or siRNA to knock down MERVL in developing embryos, resulting in mild developmental delay at low concentrations and preimplantation arrest at higher dosages.[154,161] In spite of limited *in vivo* data, the centrality of MERVL to the 2C state, either as a marker or a regulator, is rapidly solidifying.

Two questions immediately follow from these studies. First, what factor(s) induce MERVL expression in preimplantation development? And second, by what mechanism(s) does MERVL regulate fate potency? In addressing the former, a handful of candidate transcription factors have been proposed as activators of MERVL in stem cells and embryos.[81,159,161,162] *gata2* was identified in a screen to uncover downstream mediators of the MERVL derepression phenotype in *miR-34a* knockout cells. Highly expressed MERVL LTRs contain conserved Gata2 binding sites whose deletion reduced MERVL activity in luciferase assays. While *gata2* knockdown abrogated both MERVL derepression and expanded fate potential in *miR-34a* knockout cells, overexpression in wild-type cells did not induce MERVL expression, suggesting that *gata2* is required but not sufficient for MERVL activation.[81] *stella* is a maternally inherited factor whose absence leads to failure to activate MERVL and chimeric genes *in vivo,* resulting in reduced blastocyst viability.[161] The sole candidate shown to be both necessary and sufficient to induce MERVL is the double-homeodomain transcription factor Dux*,* whose role was elucidated in two back-to-back publications.[159,162] siRNA knockdown of *dux* significantly compromised Caf1-dependent induction of the 2C state in mouse ESCs, and more remarkably, *dux* overexpression was sufficient to convert up to 74% of ESCs into 2C-like cells.[159] Moreover, CRISPR knockout of *dux* in mouse zygotes severely compromised embryonic development, indicating that *dux* functions both in cultured ESCs and in preimplantation development.[162] These findings have led to the designation of *dux* as a potential master regulator of MERVL and the 2C state.

How does MERVL regulate fate potency? Despite continued efforts, a mechanistic understanding remains elusive. Theories typically fall under two categories: *cis*, whereby MERVL-regulated neighboring genes specify/maintain the 2C state; or *trans*, whereby MERVL RNA or protein products facilitate this role. Several studies have favored *cis*-based explanations, as hundreds of MERVL-containing chimeric transcripts are generated during preimplantation development.[11,81,129] These include chimeric isoforms of well-established specifiers of the TE (*tead4)* and primitive endoderm (*gata4*).[129] Non-chimeric host genes driven by MERVL enhancer activity may also be involved—one such gene, *zscan4*, has been implicated in telomere maintenance in ESCs.[163] A cohort of MERVL chimeric genes may thus constitute a transcriptional program underlying cell fate decisions in the early embryo.[11] While this model is certainly attractive, almost no direct evidence exists—isoform-specific genetic knockouts with preimplantation cell fate defects are sorely needed.

In contrast, siRNA knockdown experiments targeting the *gag* region of MERVL do elicit preimplantation defects, albeit mild ones, suggesting that this activity may be mediated at least in part by MERVL RNA or proteins. As previously mentioned, ERV protein co-option in mammalian

11

development has been well documented, including several examples of *gag*-derived host proteins.[57,103,104] For example, the murine restriction factor Fv1 almost certainly arose from MERVL *gag* sequences.[103,104] It is plausible that MERVL proteins serve a yet undefined role in fate potency. Alternatively, MERVL mRNA might function as a lncRNA, similar to HERVH lncRNAs in human cells.[140,164] In one study, disruption of an MERVL-containing lncRNA called *LincGET* resulted in 100% penetrant 2C stage arrest, but had no major effects on zygotic gene expression.[165] Of course, these possibilities are not mutually exclusive—the respective roles of MERVL proteins and RNA certainly merit further investigation.

Closing remarks

Much like ERVs themselves, public perception of these genetic elements has evolved considerably over the years. What was once solely regarded as parasitic, mutagenic proviral remnants has become increasingly recognized as a source of genomic innovation and as key actors in mammalian evolution. Examples of ERV-derived host proteins, regulatory elements, and exonized sequences number in the hundreds, with new cases uncovered on a regular basis.[109] Yet in spite of their apparently frequent evolutionary co-option, the question of essentiality remains largely unsolved. While numerous ERV-regulated host genes have been documented, to date only one lethal phenotype in a ERV-specific genetic knockout animal has been observed.[64] Given the growing availability of recently developed high-throughput genome editing technologies, *in vivo* evidence remains disappointingly lacking.[166–169]

Does this spell the end of Barbara McClintock's original vision? It may be premature to dismiss the importance of ERV co-option merely on the basis of animal phenotypes. Here, one might draw parallels between the ERV and miRNA fields. Individual knockout experiments performed in *C. elegans* disseminated the popular view that miRNAs were largely dispensable for development.[170,171] It was not until investigators completely knocked out redundant miRNA families did strong, fully-penetrant defects become apparent.[172,173] Similarly, the huge abundance of ERVs in mammalian genomes renders it unlikely that removal of any single locus will be of major consequence to development. It can perhaps be argued that ERVs confer an additional level of robustness to gene regulation, and only upon global disruption or under stress conditions are their effects apparent. Indeed, siRNA disruption of MERVL and HERVL at the family level does appear to perturb embryonic development.[154,161,164] The field may benefit greatly from multiplex CRISPR or CRISPRi-based approaches; indeed, such methods have been utilized to genetically inactivate all porcine endogenous retroviruses (PERVs) in live pigs.[174] These new technologies offer a chance to definitively show the essentiality of ERVs in development, and may represent the ultimate test for mammalian ERV co-option.

Chapter 2
Dissecting the role of MERVL in regulating cell fate potency

Background

Major strides have been made over the last three decades in methods to derive and maintain embryonic stem cells (ESCs).[175–177] Pluripotent ESCs are capable of generating the three embryonic germ layers and possess indefinite self-renewal, making them attractive substrates for regenerative medicine, disease modeling, and developmental biology. Significant effort has been poured into investigating the molecular basis of pluripotency, leading to breakthroughs such as the advent of induced pluripotent stem cells (iPSCs), wherein somatic cells can be reprogrammed by the introduction of core pluripotency transcription factors.[178] In contrast, far less is understood about the mechanisms underlying totipotency, the ability of a cell to give rise to all tissues of an organism, including extraembryonic lineages. Currently, the only means to experimentally induce totipotency is by somatic cell nuclear transfer, a technically challenging and inefficient procedure that involves implanting a somatic nucleus into an enucleated oocyte.[179,180] A comprehensive understanding of the molecular players involved in totipotency is essential for addressing these shortcomings, and would afford broad insight into the fundamental basis of cell fate plasticity.

In the developing mouse embryo, blastomeres up until the 4 to 8-cell stage are totipotent, after which this potential is restricted upon commitment to the ICM and TE lineages of the blastocyst.[181–183] Although it was previously believed that this fate decision signified an irreversible loss of totipotency, the identification of rare cells with expanded fate potential among cultured ESCs suggests that conditions in which totipotent-like cells arise can be recapitulated *in vitro*.[81,129,155–159] These ESC subpopulations, isolated on the basis of specific molecular markers, culture conditions, or genetic manipulations, highly resemble 2-cell (2C) stage blastomeres in molecular profile and developmental potential.

Our lab demonstrated that deletion of a single miRNA, *miR-34a*, leads to acquisition of 2C-like properties in both ESCs and iPSCs. *miR-34a* belongs to family of miRNAs encoded by two genomic loci consisting of *miR-34a* and *miR-34b/c*, a family previously reported to inhibit somatic reprogramming by targeting core pluripotency transcription factors.[184] *miR-34a* expression begins around the 8-cell stage and increases through the blastocyst stage, consistent with a potential role in regulating totipotency.[185] *miR-34a*-/- pluripotent stem cells exhibit expanded fate potential reminiscent of 2C blastomeres in multiple functional assays, giving rise to both embryonic and extraembryonic tissues in chimeric embryos, forming embryonic and placental derivatives in teratomas, and expressing elevated levels of extraembryonic lineage markers upon embryoid body differentiation.[81] Moreover, *miR-34a*-/- ESCs share molecular features with totipotent blastomeres, include elevated expression of the 2C-specific ERV MERVL, whose activity has been consistently correlated with totipotent-like fate potency.[81,129,155–159]

While the 2C-like features of *miR-34a*-/- ESCs are striking, the molecular mechanisms underlying this phenotype were completely unknown. Using a luciferase reporter system, I determined the minimal sequence within the MERVL LTR required for its activation in *miR-34a*-/- cells. Based on this minimal fragment, I identified *gata2* as a key transcription factor downstream of *miR-34a* that activates MERVL. *gata2* is a direct target of *miR-34a*, being required but not sufficient to induce MERVL in *miR-34a*-/- cells. I then evaluated the role of the *miR-34a*/Gata2/MERVL axis *in vivo,* demonstrating that all three components are expressed during preimplantation development but may not functionally recapitulate phenotypes observed *in vitro.* Finally, I applied a candidate approach to assess whether MERVL LTR regulatory co-option contributes to cell fate decisions *in vivo,* investigating the role of a MERVL-driven chimeric gene *tead4* using genome editing.

14

Results

A minimal LTR fragment recapitulates MERVL induction in *miR-34a*^{-/-} ESCs

As no *miR-34a* seed matches were detected in the MERVL consensus sequence, I hypothesized that *miR-34a*-dependent MERVL suppression was mediated by a downstream effector, most likely a transcription factor that could bind to the MERVL LTR. To screen for this factor in a highly tractable system, I employed a luciferase reporter containing the entire 5' MERVL LTR as well as a portion of the *gag* sequence as its promoter (Fig. 1A). When this full length reporter construct (MERVL$_{1-1000}$) was transfected into wild-type and *miR-34a*^{-/-} ESCs, substantially greater luciferase signal was detected in *miR-34a*^{-/-} ESCs, indicating that the reporter faithfully recapitulates *miR-34a*-dependent MERVL activity (Fig. 1B). I then tested a series of derivative constructs containing different portions of the full length MERVL$_{1-1000}$ reporter (Fig. 1A). A reporter containing only the LTR (MERVL$_{1-493}$) retained high luciferase activity, while this induction was completely abolished when only the *gag* portion (MERVL$_{500-1000}$) was included, suggesting that the LTR is necessary and sufficient to activate MERVL in *miR-34a*^{-/-} ESCs (Fig. 1B). Furthermore, a minimal 250-bp fragment within the LTR (MERVL$_{125-375}$) could recapitulate MERVL derepression in *miR-34a*^{-/-} ESCs (Fig. 1B). These results implicate a transcriptional mechanism involving the MERVL$_{125-375}$ minimal sequence that mediates MERVL derepression in *miR-34a*^{-/-} ESCs.

Gata2 is required for MERVL induction in *miR-34a*^{-/-} iPSCs

Having determined the minimal sequence required for MERVL induction, I leveraged bioinformatic prediction to identify candidate transcription factors that could bind this sequence. As over 70 candidates were initially produced, the following additional criteria were applied to further narrow down candidates: 1) presence of *miR-34a* binding sites; 2) overexpression in *miR-34a*^{-/-} cells; 3) co-expression with MERVL in preimplantation development. Among all candidates, only *gata2* fulfilled these criteria—it contains three predicted *miR-34a* binding sites (Fig. 3A), is overexpressed in *miR-34a*^{-/-} ESCs as measured by qPCR (Fig. 2C, 3C), and its expression pattern closely mimics that of MERVL in preimplantation development, as determined by published preimplantation RNA-Seq data (Fig. 4A).[186]

Sequence alignment of the 18 most highly expressed MERVL loci in *miR-34a*^{-/-} iPSCs reveals the presence of two fully conserved and one partially conserved Gata1/2/3 binding sites within the minimal LTR fragment (Fig. 2A). Mutating either of the fully conserved binding sites reduces luciferase activity in *miR-34a*^{-/-} pluripotent stem cells, while mutating both binding sites further reduces activity in a synergistic manner (Fig. 2B)**.** Thus, conserved Gata1/2/3 binding sites on the MERVL LTR are required for its induction in *miR-34a*^{-/-} pluripotent stem cells.

To determine whether Gata2 protein is required for MERVL activation, two independent shRNAs were employed to knock down *gata2* in two wild-type and two *miR-34a*^{-/-} iPSC lines. Both shRNA designs reduced *gata2* expression to WT levels (Fig. 2C). Strikingly, *gata2* knockdown completely abolished the induction of MERVL in *miR-34a*^{-/-} iPSCs (Fig. 2C). Notably, MERVL levels were effectively restored to wild-type levels, suggesting that *Gata2* is the primary effector responsible for MERVL induction. Taken together, these findings suggest that *gata2* mediates the induction MERVL in *miR-34a*^{-/-} pluripotent cells by binding to conserved Gata1/2/3 sites within the minimal LTR fragment.

gata2 is a *miR-34a* target

g*ata2* contains three predicted *miR-34a* binding sites, two located on the 3'UTR, and one on the 3' end of the open reading frame (Fig. 3A). Two of these sites (sites 1 and 3) were predicted to be strong *miR-34a* targets, with perfect seed matches or strong 3' complementarity.[187] In order to determine whether these sites are required for *gata2* regulation by *miR-34a*, I cloned the entire *gata2* 3'UTR plus the 3' end of the *gata2* open reading frame into a 3'UTR luciferase reporter construct (Fig. 3B). This reporter construct was transfected into ESCs in the presence or absence of *miR-34a* overexpression. *miR-34a* overexpression resulted in attenuation of luciferase activity, indicating that sequences within the *gata2* cloned region confer *miR-34a*-dependent repression (Fig. 3B). Furthermore, mutating all three predicted *miR-34a* target sites abrogated this regulation, demonstrating that these sites are required for repression by *miR-34a* (Fig. 3B). Finally, Gata2 is upregulated both at the protein and mRNA level in *miR-34a*[-/-] cells, and overexpression of *miR-34a* rescues Gata2 repression (Fig. 3C, 3D). These results indicate that *gata2* is a direct target of *miR-34a* in pluripotent stem cells.

Expression profiling of *gata2, miR-34a,* and MERVL during preimplantation development

The role of the *miR-34a*/Gata2/MERVL axis in pluripotent stem cells raises the intriguing possibility that the same pathway functions in preimplantation development. MERVL is highly expressed in 2C embryos and has been repeatedly associated with the totipotent state.[81,129,155–159] It is plausible that the onset of *miR-34a* expression in preimplantation development contributes to the restriction of cell fate by repressing Gata2/MERVL. Indeed, this model is consistent with previously published preimplantation RNA-Seq and miRNA profiling data (Fig. 4A). To further validate this expression data, I performed single-embryo qPCR to measure levels of *Gata2,* MERVL, and primary *miR-34a (pri-miR-34a)* in wild-type oocytes and preimplantation embryos. These results confirm that Gata2 and MERVL are co-expressed at the 2C stage (Fig. 4B). *Pri-miR-34a* is also expressed from the 2C stage onwards (Fig. 4B); however, it has been previously reported that miRNA function is globally suppressed during early preimplantation development.[188] Thus these expression data are consistent with a model wherein *miR-34a* suppresses Gata2/MERVL *in vivo.*

miR-34a is dispensable for MERVL silencing *in vivo*

To further test the hypothesis that *miR-34a* regulates Gata2/MERVL *in vivo,* I immunostained preimplantation embryos using antibodies against the MERVL *gag* protein. In wild-type embryos, strong MERVL signal is detected in 2C embryos, but drops to background levels by the morula stage (Fig. 5A). If *miR-34a* functions to repress MERVL during preimplantation development, one would expect sustained MERVL expression past the 2C stage in the absence of *miR-34a*. However, *miR-34a*[-/-] morula and blastocysts do not have discernably altered MERVL *gag* levels (Fig. 5A). Furthermore, single-embryo qPCR measurements of MERVL and its associated chimeric gene *zfp352* detected no significant differences in expression between wild-type and *miR-34a*[-/-] preimplantation stage embryos (Fig. 5B). Therefore, *miR-34a* appears to be dispensable for silencing MERVL in preimplantation embryos.

MERVL drives expression of the essential TE gene *tead4*

In order to investigate potential *cis*-regulatory roles of MERVL in developing embryos, I adopted a candidate approach to identify MERVL chimeric genes implicated in preimplantation cell fate specification. *tead4* was previously identified as a chimeric gene driven by an MERVL-

16

related LTR MT2B1 (*tead4:MT2B1*).[129] *tead4* encodes a transcription factor required for specification of the TE—*tead4* null mutants are unable to form blastocysts, exhibit total loss of TE lineage markers, and fail to implant.[189,190] *tead4:MT2B1* is derived from an MT2B1 alternative promoter that splices into exon 2, skipping the canonical promoter, and is not predicted to alter the open reading frame (Fig. 6A).

Using isoform specific primers, I performed single-embryo qPCR to measure levels of both canonical and MT2B1-driven isoforms of *tead4*. Strikingly, the vast majority of *tead4* transcripts are derived from the MT2B1 isoform (Fig. 6B). At the 8-cell stage, during which TE specification is thought to initiate, *tead4:MT2B1* transcripts are hundreds of times more abundant than canonical *tead4*. At the blastocyst stage after the TE is fully specified, *tead4:MT2B1* expression drops to levels comparable with canonical *tead4*. This expression pattern shows that an ERV-driven chimeric transcript is the predominant isoform of *tead4* during preimplantation development.

Generation of *tead4^{ΔMT2B1}* mice

To functionally characterize the role of *tead4:MT2B1* in development, I employed a pair of single-guide RNAs (sgRNAs) flanking the MT2B1 element to delete this sequence using CRISPR/Cas9 microinjection (Fig. 7A). 7 heterozygous founders were obtained, and one male was backcrossed to a C57BL/6J wild-type female to establish an isogenic line. The resulting F2 heterozygous offspring were intercrossed to obtain homozygous *tead4^{ΔMT2B1}* mice. 31 pups from 6 litters were born, comprising 15 wild-type, 9 heterozygous, and 7 homozygous knockout mice. Although this non-Mendelian ratio may be indicative of low-penetrance embryonic lethality, all born mice appeared morphologically normal and had no discernable breeding defects.

Molecular characterization of *tead4^{ΔMT2B1}* embryos

To determine whether deletion of *tead4^{ΔMT2B1}* alters levels of Tead4 protein in preimplantation stage embryos, I immunostained wild-type and *tead4^{ΔMT2B1}* morula using a Tead4-specific antibody. *tead4^{ΔMT2B}* morula display markedly decreased, but not abolished, nuclear Tead4 signal (Fig. 7C). Interestingly, this loss of nuclear signal was accompanied by an increase in cytoplasmic localization. These results suggest that ablation of *tead4:MT2B1* may alter subcellular localization of Tead4 protein. Notably, it has been reported that loss of nuclear localization of Tead4 impairs specification of the TE.[191]

While nuclear Tead4 localization was reduced, it was not completely absent. Given that *tead4:MT2B1* constitutes nearly the entire pool of *tead4* mRNA in preimplantation embryos, this result was difficult to reconcile. In order to measure isoform-specific and total expression of Tead4 in mutant embryos, I performed single embryo qPCR using both isoform-specific and universal primers on 8-cell stage embryos. Strikingly, while *tead4:MT2B1* expression was completely lost in *tead4^{ΔMT2B1}* embryos, total *tead4* as measured by universal primers was unaltered (Fig. 7D). Furthermore, a compensatory increase in canonical *tead4* levels was also observed, suggesting that in the absence of *tead4:MT2B1*, redundant mechanisms may act to maintain total levels of *tead4* mRNA.

tead4^{ΔMT2B1} embryos do not exhibit altered fate potency

To assess the fate potency of *tead4^{ΔMT2B1}* embryos, I performed an *in vitro* blastocoel formation assay. Morulae were harvested from both *tead4^{ΔMT2B1}* and wild-type matings and cultured for 24 hours. 22/26 (84.6%) of wild-type embryos formed blastocysts, while 16/19 (84.2%) of *tead4^{ΔMT2B1}* formed blastocysts, such that blastocoel formation rate was essentially

identical. I then sought to assay fate potency using a more sensitive assay by aggregating mutant morula with wild-type morula constitutively expressing GFP.[192] In this competition assay, subtle differences in fate potential between the two cell populations can manifest as lineage biases in the resulting chimeric blastocysts. However, equal contribution of both $tead4^{\Delta MT2B1}$ and wild-type morulae to ICM and TE lineages was observed (Fig. 7E). Overall, these results suggest that $tead4^{\Delta MT2B1}$ embryos do not possess altered fate potency compared to wild-type embryos.

Discussion

In nature, totipotency is phenomenon only observed in zygotes and early blastomeres, as these cells have the capacity to generate all cell types in the developing organism.[181–183] With each cell division thereafter, this fate potency is further restricted. By the blastocyst stage, the first cell fate decision has partitioned the embryo into the ICM, which forms the embryo proper, and the TE which forms the placenta. Cells from the ICM can be expanded *in vitro* to generate pluripotent ESCs, which recapitulate the entire gamut of pluripotent characteristics exhibited by the ICM, including indefinite self-renewal, differentiation into the three germ layers, and ability to form germline chimeras. Conventional ESCs seldom give rise to extraembryonic lineages, owing to the presence of molecular barriers that limit cell fate.

A growing body of work suggests that such barriers can be experimentally overcome. Such methods include somatic cell nuclear transfer, genetic manipulation, enrichment using specific markers, or special culture conditions.[81,129,155–159] These unique ESCs with expanded fate potential, often described as "2C-like", "totipotent-like," or "bi-potential," can give rise to both embryonic and extraembryonic cell types, and share many molecular characteristics with 2C stage blastomeres, including derepression of the ERV MERVL. *miR-34a* is the first non-coding RNA shown to be involved in the 2C state, whose deficiency leads to induction of 2C-like cells in ESCs and iPSCs. *miR-34a* appears to have multiple functions in pluripotent stem cells, as it was previously demonstrated to inhibit somatic reprogramming by targeting core pluripotency factors.[184] As a whole, these findings demonstrate the centrality of *miR-34a* in the regulation of cell fate in pluripotent stem cells.

In contrast to the profound effects of *miR-34a* deficiency in pluripotent stem cells, *miR-34a^{-/-}* preimplantation embryos show no development, morphological, or molecular abnormalities. No obvious difference in MERVL mRNA or protein levels were detected between *miR-34a^{-/-}* and wild-type embryos. It is conceivable that other regulatory mechanism act redundantly *in vivo* to maintain repression of MERVL in the absence of *miR-34a*. *miR-34a* is one of six miRNAs in the miR-34/449 family that could function redundantly in preimplantation development. Moreover, early mouse embryogenesis is highly robust and able to tolerate considerable perturbation.[193,194]

I have demonstrated that *gata2* is the primary target of *miR-34a* that mediates the derepression of MERVL in *miR-34a^{-/-}* pluripotent stem cells. This activity requires conserved Gata1/2/3 binding sites within the MERVL LTR, and *gata2* knockdown completely abolishes induction of MERVL. However, *gata2* overexpression in wild-type pluripotent stem cells was not sufficient to induce MERVL (data not shown), suggesting that additional factors are required. Identifying other activators of MERVL would be of great value to elucidating this pathway.

In *miR-34a^{-/-}* cells, loss of repression on *gata2* leads to aberrant activation of MERVL and expanded fate potential, but it remains undetermined whether MERVL functionally contributes to this fate potency. To address this question, I adopted a candidate approach to identify early cell fate specifiers regulated by MERVL. *tead4* was an attractive candidate, owing to its requirement

for TE specification, its abundant MERVL-driven chimeric isoform *tead4:MT2B1*, and its striking expression pattern in preimplantation development. However, genetic ablation of the chimeric isoform did not result in a discernable loss of fate potency, as measured by *ex vivo* assays. Furthermore, loss of *tead4:MT2B1* resulted in compensatory upregulation of the canonical isoform. These results argue that chimeric transcripts may confer robustness to essential regulatory networks by providing a redundant source of alternative transcripts. It does appear that *tead4^{ΔMT2B1}*embryos have reduced levels of nuclear Tead4 protein at the morula stage, but given the inherent plasticity of mouse preimplantation embryos,[193,194] this reduction in Tead4 may not be sufficient to illicit obvious developmental defects.

Taken together, these findings implicate a complex network of miRNA, protein-coding genes, and ERVs in the control of cell fate potency. This work may help lay the foundation for fully recapitulating conditions to derive totipotent stem cells *in vitro.* Still, the functional role of MERVL in this pathway remains unknown. MERVL can activate a wide range of genes during preimplantation development,[11,129] the vast majority of which has not been characterized *in vivo*. As the generation of knockout mice has traditionally been a costly and time-consuming endeavor, the use of high-throughput *in vivo* genome editing tools would greatly aid in these efforts.

Methods

Animal generation, breeding, and genotyping

The generation of *miR-34a^{-/-}* mice was described previously.[184] *tead4^{ΔMT2B1}* mice were generated by CRISPR mRNA and sgRNA microinjection. sgRNAs were synthesized by T7 *in vitro* transcription as previous described (oligos listed in Table 1).[195] An injection mix of 60μl total volume was prepared at 100 ng/μl Cas9 mRNA (Life Technologies, Cat. # A25640) and 50 ng/μl sgRNA final concentration. The mix was filtered by centrifugation through a Spin-X Centrifuge Tube Filter (Sigma-Aldrich, cat. # CLS8160-24EA) for 1 min at 10,000 x g. The filter was then removed and the solution was centrifuged for 10 mins at 17,000 x g. 45μl was aspirated for the final injection mix. Microinjection was performed using continuous flow settings.

Animal genotyping primers are listed in Table 1. All mice were maintained on an isogenic C57BL/6J background and housed in a non-barrier animal facility at UC-Berkeley. All animal work was performed with the approval of the UC-Berkley's Animal Care and Use Committee. UC-Berkeley's assurance number is A3084-01, and is on file at the National Institutes of Health Office of Laboratory Animal Welfare.

Luciferase Assays

For MERVL-luciferase reporter assays, pGL3 luciferase reporter vectors (Promega, Cat. # E1751) harboring MERVL$_{1-1000}$, MERVL$_{1-493}$, and MERVL$_{500-1000}$ fragments, were used as described by Macfarlan *et. al.*[129] The MERVL$_{125-375}$ reporter was constructed by truncating the MERVL$_{1-493}$ fragment using a QuikChange Site-Directed Mutagenesis Kit (Strategene, Cat. # 200518). The two fully conserved Gata2 binding sites (BS1 and BS3) were ablated in the MERVL$_{125-375}$ reporter construct, either individually or in combination, using a QuikChange Site-Directed Mutagenesis Kit.

MERVL reporters and control Renilla luciferase reporter pRL-TK (Promega, Cat. # E2241) were co-transfected (600 ng and 150 ng per well of a 12-well plate, respectively) using Lipofectamine 2000 reagent (Life Technologies, Cat. # 11668027) into ESCs. Transfection complexes containing the reporter constructs were prepared in Opti-MEM Reduced-Serum

Medium (Life Technologies, Cat. # 31985062) according to manufacturer's instructions. After trypsinization with 0.25% Trypsin + EDTA (Life Technologies, Cat. # 25200-056), 100,000 cells were resuspended in ES media lacking Pen Strep, incubated with transfection complexes for 10 minutes at 37° C, and then transferred to one well of a 12-well plate containing feeders. After 48 hours, transfected ESCs were trypsinized, plated onto gelatin-coated plates for 1 hour to remove feeders, and then assayed for luciferase activity by Dual-Luciferase® Reporter Assay System (Promega, Cat. # E1910) using a Glomax 20/20 Luminometer (Promega).

For *gata2* 3'UTR luciferase assays, a fragment that includes the 3' portion of the ORF and the entire Gata2 3'UTR was amplified by PCR. The fragment was then cloned into a psiCheck-2 vector (Promega, Cat. # C8021) to generate the *gata2* 3'UTR-Luc reporter. *miR-34a* binding site mutants were generated using a QuikChange Site-Directed Mutagenesis Kit. The *gata2* 3'UTR-Luc reporters (2 ng per well of 12-well plate) were co-transfected with 100 nM siGFP or mature *miR-34a* RNA mimics using Lipofectamine 2000 reagent (Life Technologies, Cat. #11668027) into a feeder-free mouse ESC line. After 48 hours, cells were lysed and assayed for luciferase activity by Dual-Luciferase® Reporter Assay System (Promega, Cat. # E1910) using a Glomax 20/20 Luminometer (Promega).

All luciferase assay primers are listed in Table 1.

Prediction of transcription factor binding sites

MATCH (version 8.6) was used to search for transcription factor binding motifs of the TRANSFAC database within the 250-bp central portion of the MERVL LTR (MERVL$_{125-375}$). MinSUM profile was used as cutoff to balance both the false positives and false negatives in detecting binding sites. For highlighting the GATA binding sites in all MERVL LTR sequences, CLUSTAL W (version 1.83) was used to perform multiple alignment with default settings.

Real-time qPCR

RNA was isolated by Trizol extraction following manufacturer's instruction (Life Technologies, Cat. # 15596). cDNA was reverse-transcribed using iScript Advanced Reverse-Transcriptase (Bio-Rad, Cat. # 1725037). All real-time qPCR analyses were performed using SYBR FAST qPCR Master Mix (Kapa Biosystems, Cat. # KK4604). To detect MERVL expression, four pairs of primers were designed to amplify specific regions of MERVL (Fig. 5C) and yielded similar results. One pair of primers detecting the MERVL *pol* region was used for all other MERVL real-time qPCR analyses.

For single-embryo real time PCR, embryos were collected at different developmental stages by ampulla puncture (oocyte), oviduct flushing (2-cell through 8-cell), or uterus flushing (morula and blastocyst) with KSOM Medium (Millipore, Cat. # MR-121-D), and then were subject to real-time PCR analyses using a Single Cell-to-Ct qRT-PCR kit (Life Technologies, Cat. #4458236).

All qPCR primers are listed in Table 1.

Transfection and retrovirus/lentivirus transduction

To overexpress *miR-34a* in wild-type and *miR-34a*$^{-/-}$ iPSCs, cells were infected with MSCV (murine stem cell virus) retrovirus that encoded a LTR-*miR-34a* and a PGK-puromycin-IRES-GFP cassette. MSCV and MSCV-miR-34a transduced iPSCs were selected with 3 µg/ml puromycin for two days before being collected for real-time qPCR analyses and western blotting.

To knock down *gata2* by RNAi, two Gata2 shRNAs were cloned into pLKO.1 lentiviral vector (Addgene, #10878). All shRNA sequences are listed in Table 1. The corresponding lentiviruses were produced by co-transfecting pLKO.1 shRNA vectors with pMD2.G and psPAX2 to HEK293T cells. After infection, iPSCs were selected in 3 µg/ml puromycin for two days and expanded for *in vitro* and *in vivo* analyses.

Western blot

For ESC or iPSC collection, trypsinized cells were plated on a gelatin-coated plate for 1 hour to remove feeders. Cells separated from the feeders were then lysed in Laemmli sample buffer (60 mM Tris-Cl pH 6.8, 2% SDS, 100 mM DTT, 10% glycerol, 0.02% bromophenol blue) and subjected to western analyses. Antibodies against mouse Gata2 (Santa Cruz Biotechnology, Cat # CG2-96) was used at 1:500 dilution, and α-tubulin (Sigma, clone B-5-1-2) was used at a 1:4,000 dilution as a loading control. The quantitation of all western analyses was carried out with ImageJ (NIH).

Immunofluorescence staining

Preimplantation embryos were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, Cat # 19202) for 15 min at room temperature and then transferred to PBS containing 0.1% bovine serum albumin (BSA, Sigma, Cat # A3311). Permeabilization was performed by incubating the embryos with PBS containing 0.1% Triton X-100 and 0.1% BSA for 5 minutes. Subsequently, embryos were blocked for 1 hour at room temperature in blocking solution (PBS containing 10% goat serum Fisher Cat#31872, and 0.1% BSA). Embryos were then incubated with MERVL-Gag antibody (1:2000, a gift from T. Heidmann laboratory) in blocking solution at 4°C overnight, and subsequently in Alexa Fluor 594 Goat anti mouse IgG (1:400, ThermoFisher, Cat # A21125) in blocking solution at 4°C overnight. Embryos were imaged using a Zeiss Observer A1 fluorescent microscope.

Chimeric blastocyst reconstitution assay

To generate chimeric blastocysts by morula aggregation, *tead4^ΔMT2B1^* or WT (expressing GFP under the control of an Ubiquitin promoter)[196] morulae were collected from superovulated matings at day 3.0 postcoitum (dpc), treated with acid Tyrode's solution (Sigma, Cat. # T1788) to remove the zona pellucida, and cultured as 1-to-1 pairs in concave depressions formed on cell culture plates with EmbryoMax KSOM Medium (Millipore, Cat. # MR-121-D). After 24 hours, resulting chimeric blastocysts were imaged using a Zeiss Observer A1 fluorescent microscope.

Purpose	Name	Sequence
qPCR primers	β-actin F	GATCTGGCACCACACCTTCT
	β-actin R	GGGGTGTGTTGAAGGTCTCAAA
	Mervl F	GGTGGTCGAGATGGAGGTTA
	Mervl R	CGGATTGCGGGTTTGATCTC
	Iap F	GCTCCTGAAGATGTAAGCAATAAAG
	Iap R	CTTCCTTGCGCCAGTCCCGAG
	Zfp352 F	AAGGTCCCACATCTGAAGAA
	Zfp352 R	GGGTATGAGGATTCACCCA
	Gata2 F	CACCCCGCCGTATTGAATG
	Gata2 R	CCTGCGAGTCGAGATGGTTG
	Tead4 canon. F	TCCTCTGCAAACTCCAGTCC
	Tead4 canon. R	AGCTCCACTCGTTGGAGGTA
	Tead4 MT2B1 F	GGCAAGCCTACTTCTTCAGG
	Tead4 MT2B1 R	AGCTCCACTCGTTGGAGGTA
	Tead4 total F	GCCAGCAAGATCCCGACAC
	Tead4 total R	TTTCGAGGTAGGGGTCACTGT
Genotyping Primers	*miR-34a*-Common-R	ACTGCTGTACCCTGCTGCTT
	miR-34a-WT-F	GTACCCCGACATGCAAACTT
	miR-34a-KO-F	GCAGGACCACTGGATCATTT
	Tead4 MT2B1 F1	ATCACATGGTCCCTGCCTTG
	Tead4 MT2B1 R1	CCCACATCAAACCCAGCTCT
	Tead4 MT2B1 R2	CCTCGGCCTCTCAAACTCTAG
tead4:MT2B1 sgRNA oligos	T7 Tead4 MT2B1 left F	GGATCCTAATACGACTCACTATAGCCCATCCAACAAGCGTGTGGGTTTTAGAGCTAGAA
	T7 Tead4 MT2B1 right F	GGATCCTAATACGACTCACTATAGATAGCTACAGTTCCATCGATGTTTTAGAGCTAGAA
	T7 sgRNA R	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
	T7 Amp F	GGATCCTAATACGACTCACTATAG
	T7 Amp R	AAAAAAGCACCGACTCGG
Luciferase assay primers	MERVL 1-375 F	TGGACTTCCATTCACCTCGAGATCTGCGATCTAAGTAAGC
	MERVL 1-375 R	ATCGCAGATCTCGAGGTGAATGGAAGTCCAAGGATCTAGC
	MERVL 125-375 F	CTTACGCGTGCTAGCGATCTTGAGCCATAGTGGCTATGGA
	MERVL 125-375 R	CTATGGCTCAAGATCGCTAGCACGCGTAAGAGCTCGGTAC
	Gata2ΔBS1 F	TCTCCGAGTTTAAGGAACACACCTTTGGGCTACGCCTTTC
	Gata2ΔBS1 R	AATCCCAGATGAAAGGCGTAGCCCAAAGGTGTGTTCCTTA
	Gata2ΔBS3 F	TTAAAGGTGTGGTGGAACACACCTTTGGGCTACACCTTCT
	Gata2ΔBS3 R	TGTCTCCAGCAGAAGGTGTAGCCCAAAGGTGTGTTCCACC
	Gata2 3'UTR F XhoI	CTCGAGAGTCTCTCTTTTGGCCACCC
	Gata2 3'UTR R NotI	GCGGCCGCCAAGGCCACCTGACAGCTTA
	Gata2 3'UTRΔ34aBS F	CCGTCCAGCATGGTGATGGGCTAGGCAAGCCTCCCACTGG
	Gata2 3'UTRΔ34aBS R	GCTTGCCTAGCCCATCACCATGCTGGACGGGTGGGGGTGG
	Gata2 3'UTRΔ34aBS2 F	AGAGACCCACTTCCTGCCTAGCCTGGCCGAAGCCACCTCT
	Gata2 3'UTRΔ34aBS2 R	TCGGCCAGGCTAGGCAGGAAGTGGGTCTCTTGGGATGGGC
	Gata2 3'UTRΔ34aBS3 F	CTTCTTTGGGACCTCCCAGTCAGGGCTCTCGGGGGCAGAC
	Gata2 3'UTRΔ34aBS3 R	GAGAGCCCTGACTGGGAGGTCCCAAAGAAGGACCCCAAGA
shRNA oligos	shgata2#1 sense	CCGGGAGGTGGATGTCTTCTTCAACCACTCGAGTGGTTGAAGAAGACATCCACCTCTTTTTG
	shgata2#1 antisense	AATTCAAAAAGAGGTGGATGTCTTCTTCAACCACTCGAGTGGTTGAAGAAGACATCCACCTC
	shgata2#2 sense	CCGGGGACGAGGTGGATGTCTTCTTCAACTCGAGTTGAAGAAGACATCCACCTCGTCCTTTTTG
	shgata2#2 antisense	AATTCAAAAAGGACGAGGTGGATGTCTTCTTCAACTCGAGTTGAAGAAGACATCCACCTCGTCC

Table 1: Primer and oligo sequences used in Chapter 2

A

B
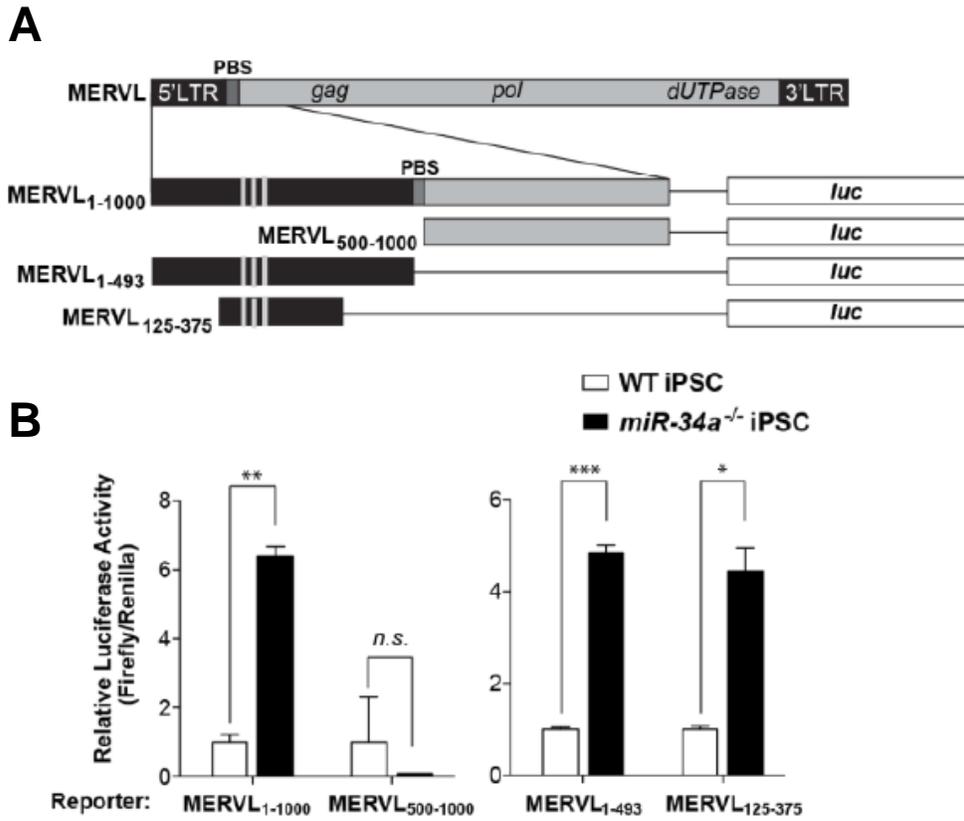
Figure 1. A minimal LTR fragment recapitulates MERVL derepression in *miR-34a$^{-/-}$* pluripotent stem cells

A. Schematic diagram of MERVL fragments that were tested for promoter activity using luciferase assays. **B**. The luciferase reporters driven by MERVL fragments containing the full length LTR (MERVL$_{1-1000}$ and MERVL$_{1-493}$) are highly active in *miR-34a$^{-/-}$* ESCs, but not in wild-type ESCs. A 250-bp MERVL$_{125-375}$ fragment recapitulates this differential reporter activity in wild-type and *miR-34a$^{-/-}$* ESCs. Error bars: *s.d.*, n=2. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, *n.s.*, not significant.

A
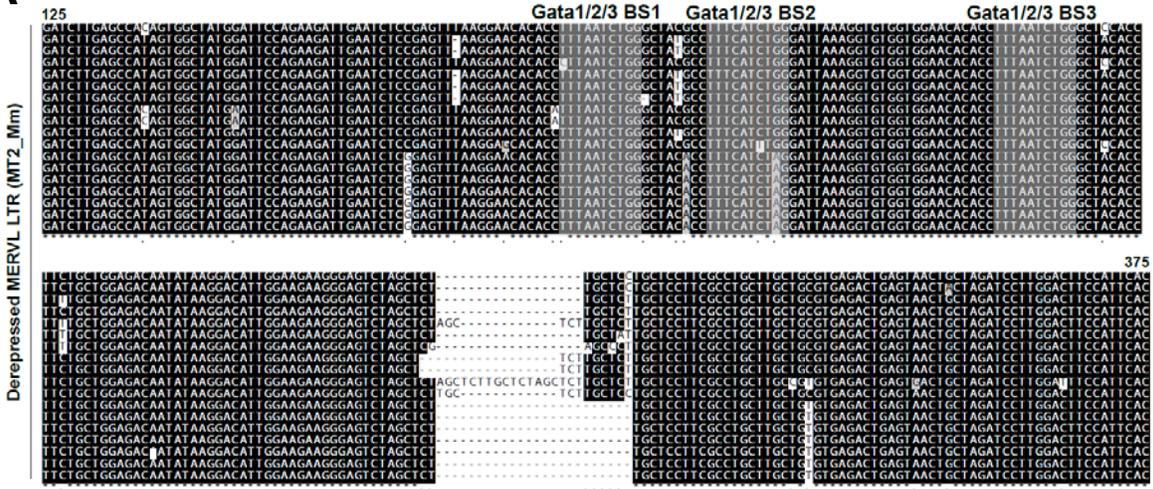
B
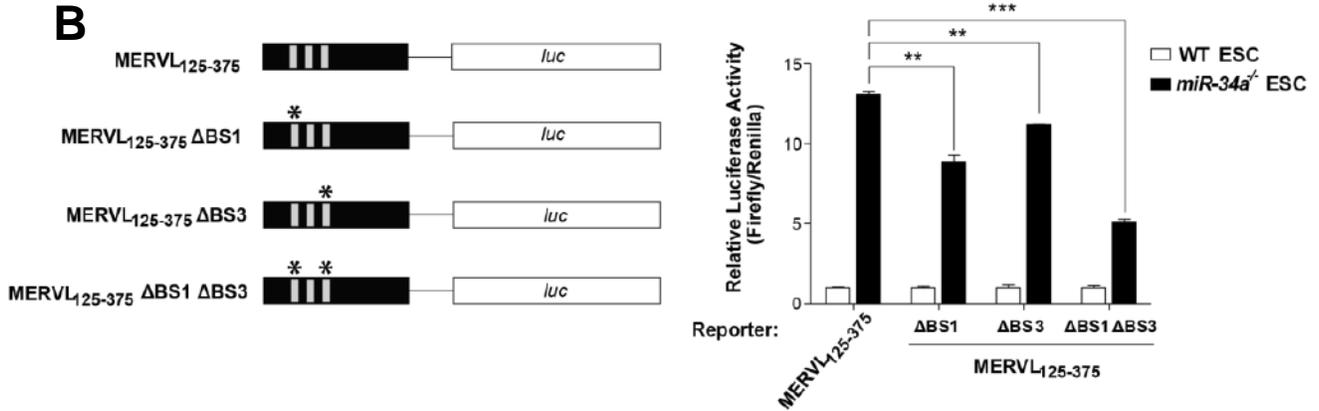
C
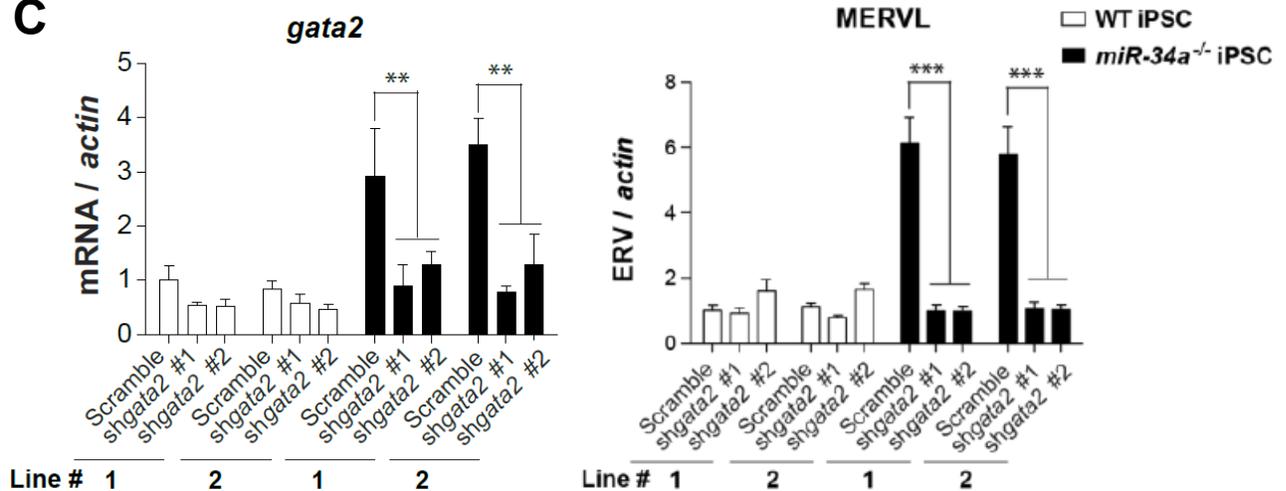
Figure 2. *gata2* is required for MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells

A. Clustal-W LTR sequence alignment of the MERVL$_{125-375}$ fragment from 18 highly expressed MERVL loci in *miR-34a-/-* iPSCs reveals three conserved predicted Gata1/2/3 binding sites. Among the three predicted binding sites (designated as BS1, BS2, and BS3 and highlighted in grey), BS1 and BS3 are fully conserved across all 18 MERVL elements, while BS2 is partially conserved. **B.** (Left) Schematic diagram of the MERVL$_{125-375}$ reporter and derivatives harboring mutations of predicted Gata2 binding sites. (Right) Mutation of both BS1 and BS3 in the MERVL$_{125-375}$ reporter synergistically impairs promoter reporter activity in *miR-34a*$^{-/-}$ ESCs. Error bars: *s.d.*, n=2. ** $P < 0.01$, *** $P < 0.001$, *n.s.*, not significant. **C.** *gata2* knockdown in *miR-34a*$^{-/-}$ iPSCs significantly decreases the expression of MERVL (performed by Chao-Po Lin). Error bars, *s.e.m.*, n = 2. ** $P < 0.01$, *** $P < 0.001$, *n.s.*, not significant.
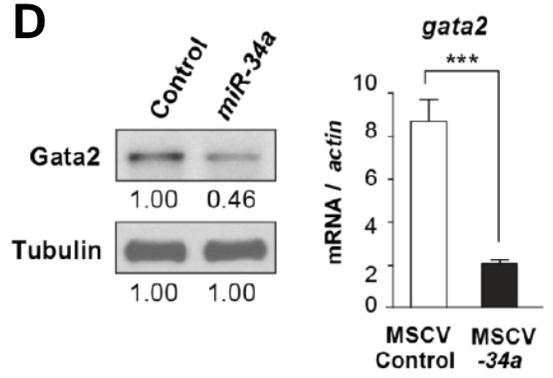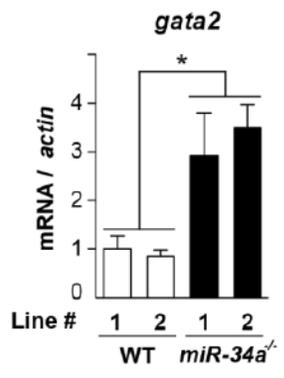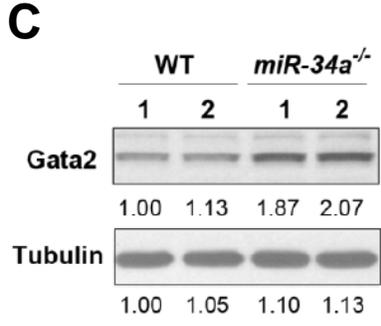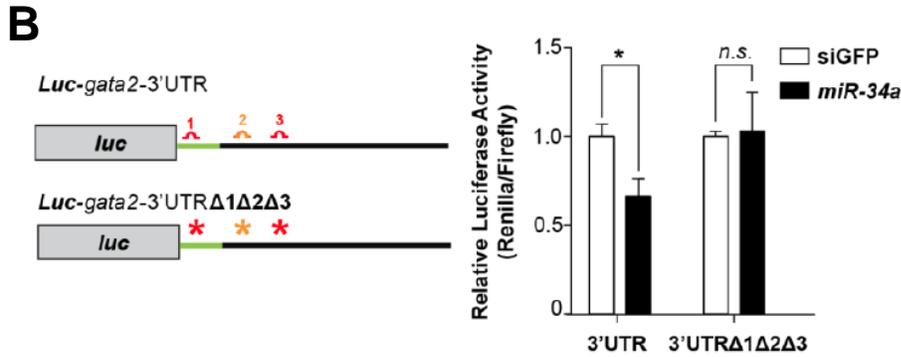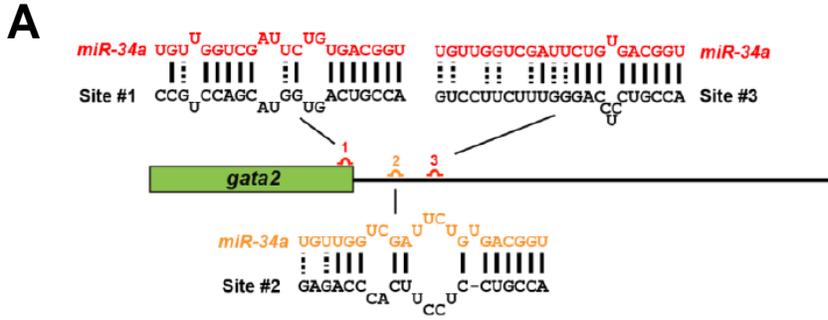
Figure 3. *gata2* is a *miR-34a* target in *miR-34a^{-/-}* pluripotent stem cells

A. Schematic representation of three predicted *miR-34a* binding sites in the *gata2* mRNA, with one site (1) located at the 3' end of the open reading frame and two sites (2 and 3) located within the 3'-UTR. Site 1 (red) is predicted as a strong *miR-34a* binding site by both the 7mer-A1 seed-match rule and duplex folding energy.[197] Site 3 (red) does not have a perfect seed match, but contains compensatory 3' base-pairing and exhibits strong folding energy. In comparison, site 2 (orange) represents a weaker prediction but with a reasonable folding energy.[197] **B**. Mutating all three predicted *miR-34a* binding sites within the *gata2*-3'UTR luciferase reporter completely abolishes *miR-34a*-dependent repression. Error bars: *s.d.*, n=2. * $P < 0.05$, *n.s.*, not significant. **C.** Gata2 protein (left) and *gata2* mRNA (right) levels are elevated in *miR-34a^{-/-}* iPSCs compared to wild-type iPSCs. Two independent pairs of passage- and littermate-controlled wild-type and *miR-34a^{-/-}* iPSC lines were measured by western blotting and real-time PCR analyses. Error bars: *s.e.m.*, n=3. * $P < 0.05$. **D**. Overexpression of *miR-34a* in *miR-34a^{-/-}* iPSCs using MSCV retroviral vectors represses Gata2 protein and mRNA levels (performed by Chao-Po Lin). Error bars: *s.e.m.*, n=3. *** $P < 0.001$.

Figure 4. *gata2*, MERVL, and *miR-34a* expression in preimplantation development
A. Expression patterns of MERVL, *gata2* and *miR-34a* during mouse preimplantation development. In published datasets, levels of MERVL and *gata2* both peak in 2C embryos. Mature *miR-34a* is induced at the 4-cell stage, and remains highly expressed until the blastocyst stage. **B.** Using single-embryo real-time PCR analyses, we validated the expression patterns of MERVL, *gata2* and *pri-miR-34a* in mouse preimplantation embryos. N=16 per stage.

(Adapted from Choi et. al., 2017)

29
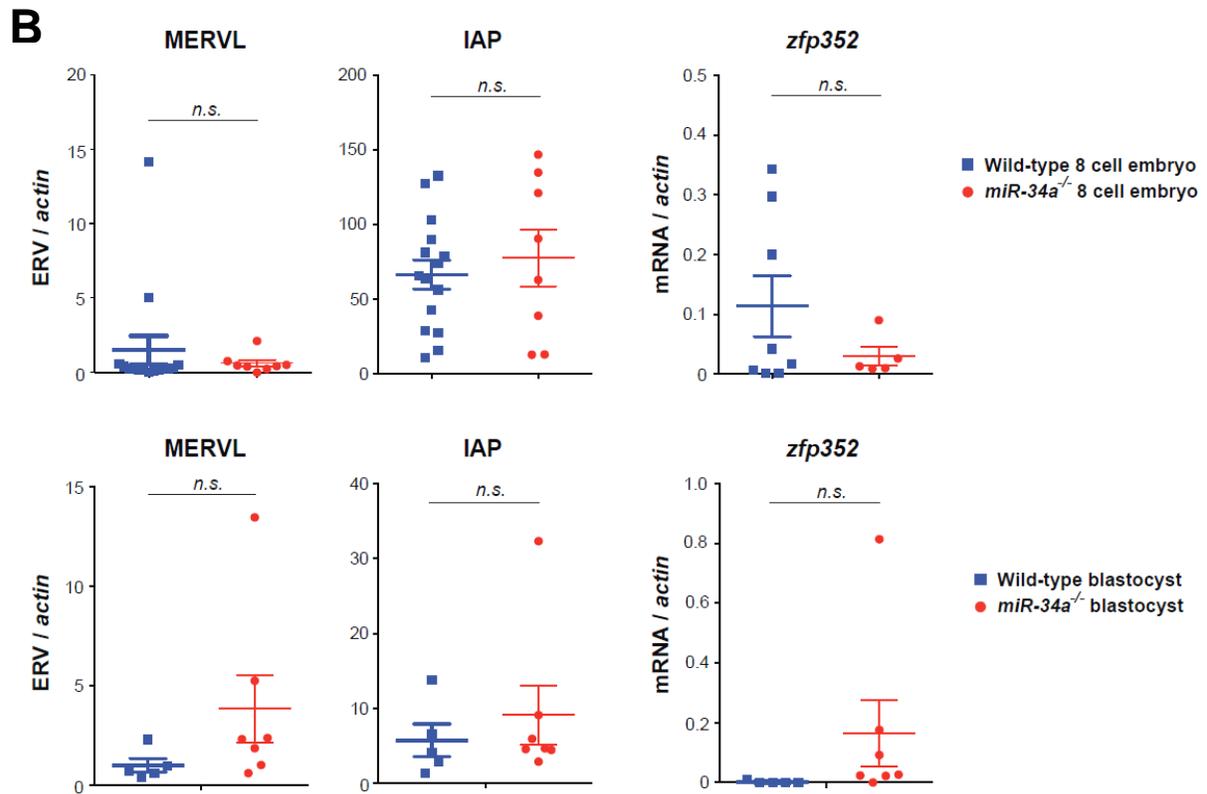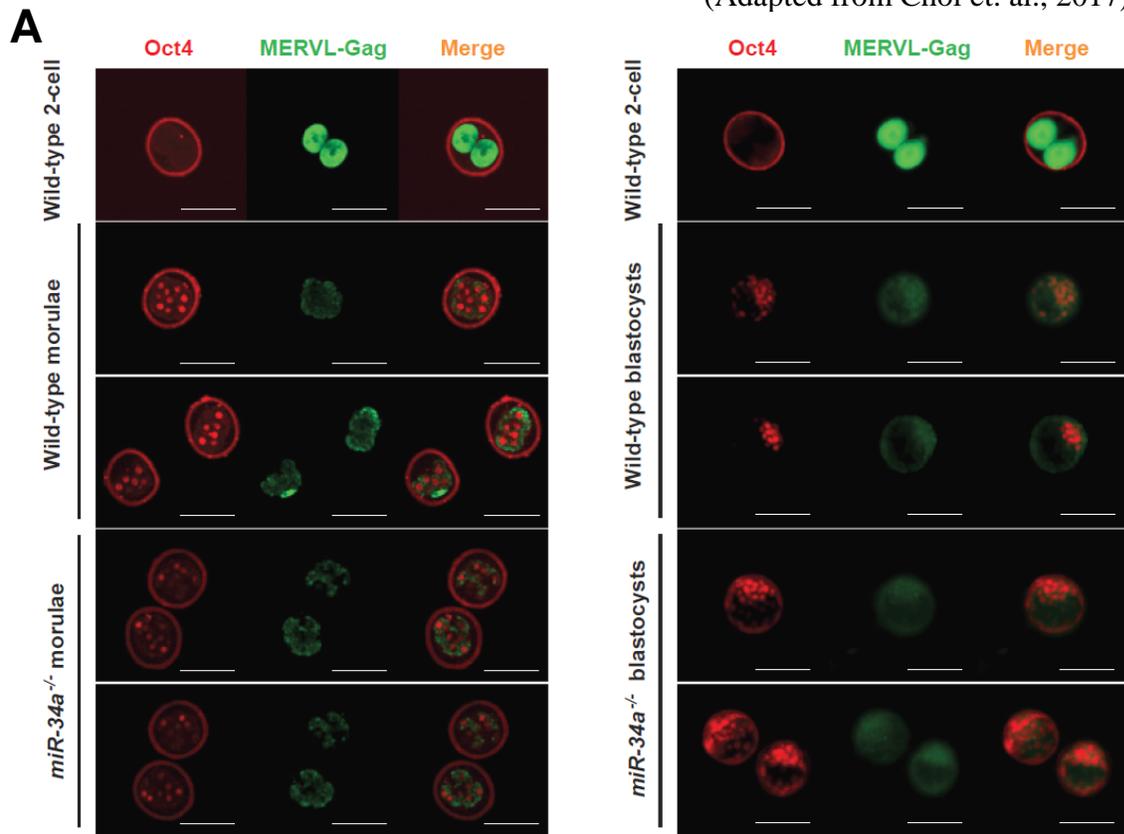
Figure 5. *miR-34a* **is dispensable for MERVL silencing** *in vivo*

A. Representative immunofluorescence staining of MERVL-Gag (green) and Oct4 (red) in wild-type and *miR-34a*[-/-] embryos at the morula and blastocyst stages showed similar levels of MERVL-Gag. Wild-type 2C blastomeres were included to demonstrate the specificity of MERVL-Gag IF staining. Wild-type morula, n=7; *miR-34a*[-/-] morula, n=14; wild-type blastocysts, n=7; *miR-34a*[-/-] blastocysts, n=15. **B.** Expression level of MERVL, IAP and MERVL-proximal gene *zfp352* in wild-type and *miR-34a*[-/-] embryos at the 8-cell (top) and blastocyst (bottom) stage, as measured by single-embryo real-time PCR analyses. Error bars: *s.e.m. n.s.*, not significant.

A. *tead4*

CCCGTTGGAC AGGCCGCAAA

Ex12 Ex11 // Ex2 Ex1 MT2B1

B. Relative expression

- *tead4* (canon.)
- *tead4:MT2B1*

oocyte 2-cell 8-cell blastocyst
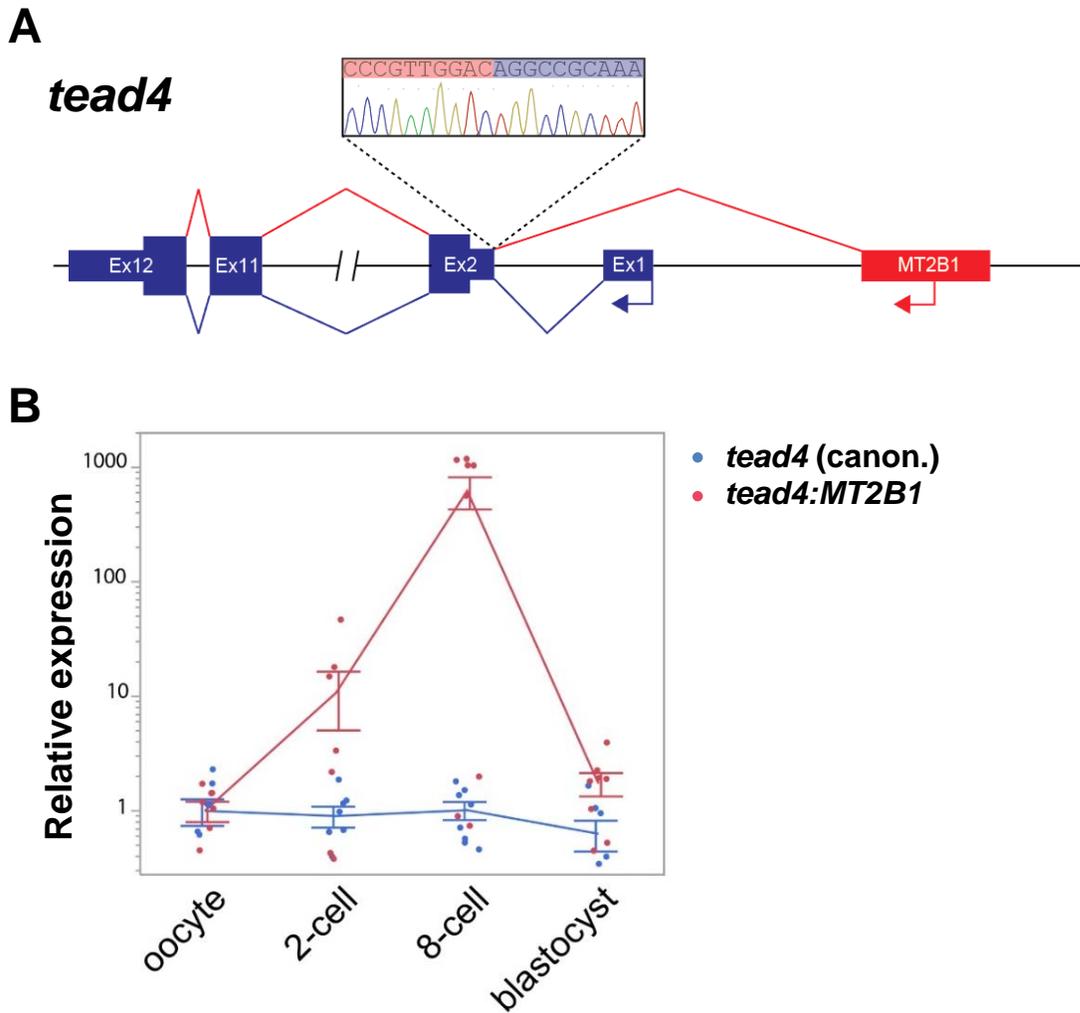
Figure 6. *tead4:MT2B1* **is the predominant isoform of** *tead4* **in preimplantation development**
A. Schematic diagram of *tead4* gene structure. An MT2B1 element upstream of exon 1 acts as an alternative promoter and forms a splice junction with exon 2 (sequence validation in box insert).
B. Single-embryo real time qPCR analysis using primers of equal efficiency reveals that *tead4:MT2B1* is the expressed at substantially higher levels than the canonical isoform during preimplantation development, particularly during the 8-cells stage. Error bars: *s.d.*

A *tead4*

B

F + R1

F + R2

C DAPI   Tead4   Merge

WT

tead4^{ΔMT2B1}

D
● Wild-type
○ *tead4^{ΔMT2B1}*

tead4 (canon.)   *tead4:MT2B1*   *tead4 (total)*

E
ICM + TE
ICM only
TE only

Number of blastocysts
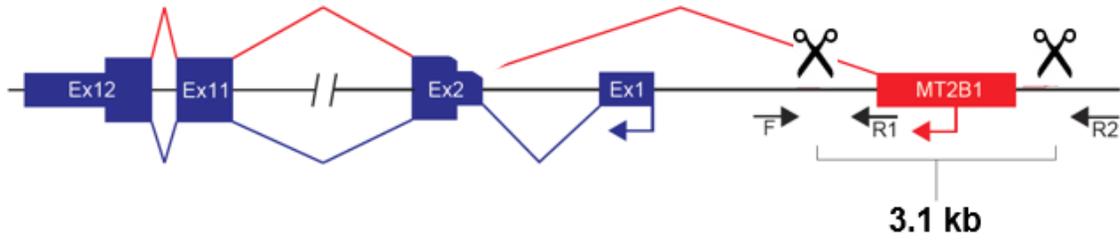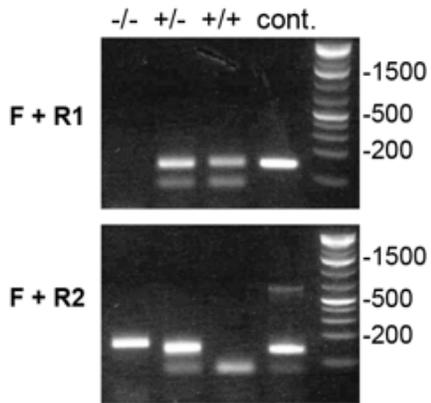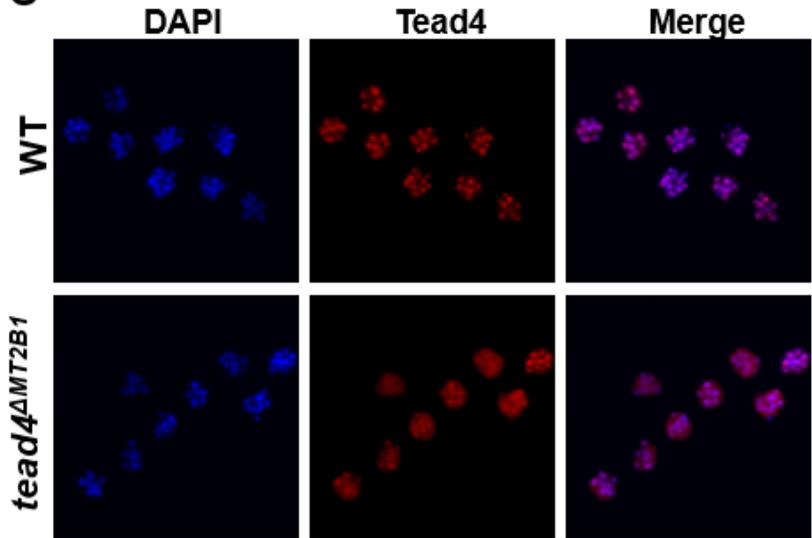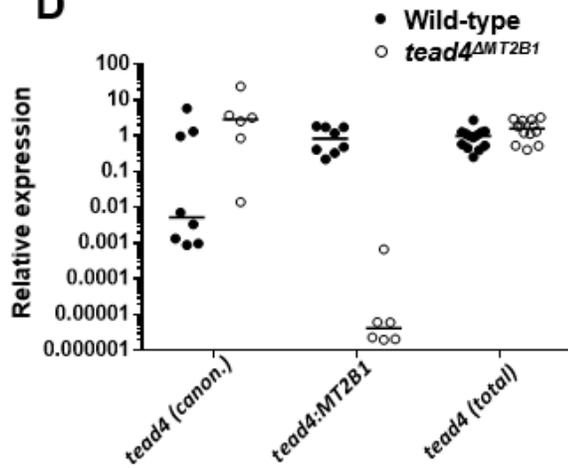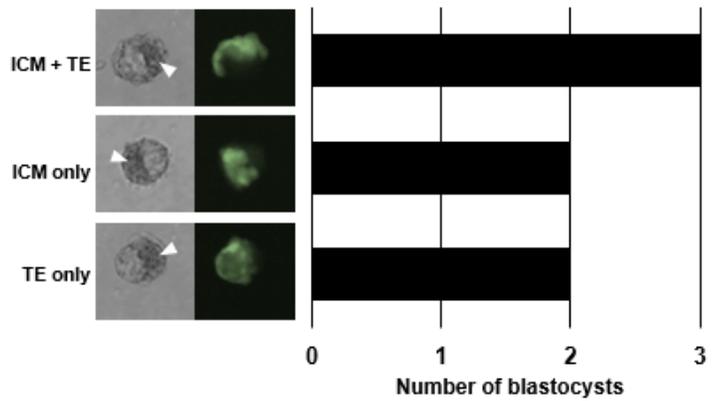
Figure 7. Generation and characterization of *tead4^ΔMT2B1^* mice

A. Schematic diagram of gene editing strategy. A pair of sgRNAs flanking the MT2B1 element mediate 3.1 kb deletion upon successful editing. Genotyping primers F, R1, and R2 are shown as arrows. **B.** Representative genotyping PCR. F + R1 produces an ~150 bp band in the presence of the unedited allele, while F + R2 produces an ~150 bp in the presence of the edited allele. Pooled edited and unedited DNA was used as a control. **C.** Immunofluorescence staining was performed using Tead4 antibodies on wild-type and *tead4^ΔMT2B1^* morulae. *tead4^ΔMT2B1^* morula display reduced nuclear signal and increased cytoplasmic signal. n=8 morulae per genotype. **D**. Single-embryo real time qPCR was performed on wild-type and *tead4^ΔMT2B1^* morulae using isoform-specific and total *tead4* primers. n=12 morulae per genotype. **E.** *tead4^ΔMT2B1^* morulae were aggregated with wild-type morulae expressing GFP under an Ubiquitin promoter. (cite). After 24 hours, resulting chimeric blastocysts were scored. n=10 blastocysts.

Chapter 3
Development of a high-throughput mouse genome editing technique

Background

While cell culture systems can recapitulate many biological processes, animal models have long served as the gold standard for demonstrating gene function *in vivo*. Genetically modified mouse models are invaluable tools for probing genes function within the context of a living animal, but their production has been historically tedious, costly, and time-consuming. Classic mouse genome engineering is accomplished by injection of clonally edited ESCs into recipient blastocysts, which are implanted into pseudopregnant females. The resulting chimeric mice are then backcrossed to isogeny and intercrossed before homozygous edited animals are obtained.[175,198,199]

The emergence of programmable nucleases enabled a major shift in genome editing strategies. Through the formation and subsequent repair of double-strand breaks (DSBs) at desired genomic loci, investigators could obtain random insertion/deletion mutations (indels) through non-homologous end joining (NHEJ), or precise sequence replacement through homology-directed repair.[200–205] Early iterations of this technology, which utilized zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), improved editing efficiency in cell lines by orders of magnitude over traditional homologous recombination-based methods; however, their complex design and synthesis workflows have limited their adoption for *in vivo* editing.[201,203,206]

These caveats were largely overcome by the advent CRISPR/Cas9 technology. Clustered regularly-interspaced short palindromic repeats (CRISPR) RNAs and CRISPR-associated (Cas) proteins were initially discovered as part of a bacterial adaptive immune system used to defend against viral infection.[207,208] The most popular form of this system, derived from *Streptococcus pyogenes,* has been reconstituted *in vitro* and adapted for efficient genome editing in almost all metazoans.[209–215] Cas9 nuclease activity is directed by a single-guide RNA (sgRNA) that is complementary to its target locus. The simplicity and efficiency of this system has propelled CRISPR/Cas9 to the forefront of the genome editing field, with applications including multiplex editing *in vivo*, large deletions, engineering of reporters and conditional alleles, and inducing genomic rearrangements.[209,216–220]

Using CRISPR/Cas9 technology, it is possible to completely bypass ESC manipulation by microinjecting Cas9 DNA or mRNA together with sgRNAs into fertilized zygotes.[218,221–224] While a major advancement compared to traditional ESC methods, microinjection is a technically challenging procedure, requiring highly trained personnel and sophisticated equipment. Cas9 mRNA electroporation has been demonstrated as an alternative to microinjection; however, such methods often require custom instrumentation and exhibit reduced efficiency compared to microinjection.[168,225,226]

Structural studies reveal that Cas9 forms a stable ribonucleoprotein (RNP) complex with sgRNA in its active conformation,[227–229] and delivery of pre-assembled Cas9/sgRNA ribonucleoproteins (RNPs) facilitates highly efficient editing in cell lines and in many organisms.[230–234] Based on these principles, we developed an electroporation-based method to deliver Cas9/sgRNA RNPs into zygotes for high-throughput genome editing, designated as **CRISP**R **R**NP **E**lectroporation of **Z**ygotes (CRISPR-EZ) (Fig. 8A). Using CRISPR-EZ, we achieved high efficiency editing for multiple genes, and generated live animals and mouse embryos with a variety of editing schemes, including indel mutations, point mutations, genomic deletions and small precise insertions. Taken together, CRISPR-EZ is a simple, economic, high-throughput, and highly efficient technique for genome editing *in vivo*, with the potential to replace microinjection as the *de facto* standard method for mammalian genome editing.

35

Results

Optimization of conditions

To optimize experimental conditions for efficient genome editing by CRISPR-EZ, we targeted exon 1 of the *Tyr* gene (Fig. 8B), an essential enzyme for pigment synthesis[235]. A cloning-free *in vitro* transcription method was used to generate the sgRNA,[233,236] which was then combined with purified Cas9 protein to form Cas9/sgRNA RNPs. Prior to electroporation, we collected C57B6/J mouse zygotes from superovulated females, and subjected them to a brief treatment with Acid Tyrode's solution to weaken the zona pellucida and facilitate Cas9 RNP delivery. We then combined 30-40 pre-treated mouse zygotes with assembled Cas9/sgRNA RNPs for each electroporation, and cultured the electroporated zygotes to the morula stage for genotyping analysis, or to the 2-cell stage for oviduct transfer to pseudopregnant mothers (Fig. 8A). Successful NHEJ-mediated *Tyr* editing was predicted to ablate a HinfI restriction site 1 nt upstream of the Protospacer Adjacent Motif (PAM) (Fig. 8B), allowing us to design a restriction fragment length polymorphism (RFLP) assay to distinguish among bi-allelic editing, partial editing, and unedited mouse embryos. Partially edited embryos could contain either mono-allelic or mosaic editing.

To optimize experimental conditions for CRISPR-EZ, we determined the NHEJ editing efficiency and embryo survival rate at different RNP concentrations (8 uM or 16 uM) using multiple electroporation conditions (2 pulses of electroporation at 1 msec, 3 msec and 10 msec pulse length) (Fig. 8C). Electroporated embryos were cultured to the compacted morula stage and subjected to RFLP and sequencing analyses for genotyping (Fig. 8C, 8D). From these experiments, we determined that electroporation pulse length was the critical parameter that impacted the efficiency of genome editing in mouse embryos (Fig. 8E). CRISPR-EZ at 1 msec pulse length yielded mostly partially edited embryos, but the 3 msec and 10 msec pulse length conditions mostly generated bi-allelic editing (83% and 100%, respectively, Fig. 8C, 8E, Table 3), indicating that a longer electroporation pulse facilitated greater CRISPR editing in mouse embryos. It is possible that a longer pulse permits greater Cas9/sgRNA RNP delivery, hence improving NHEJ editing efficiency. Notably, all embryos exhibited partial or complete editing at 3 or 10 msec pulse length conditions, demonstrating 100% delivery of Cas9/sgRNA RNPs into mouse zygotes at these conditions (Fig. 8E). Not surprisingly, a higher concentration of Cas9/sgRNA RNPs in CRISPR-EZ experiments also yielded improved NHEJ editing efficiency (Fig. 8E), possibly due to increased Cas9/sgRNA RNP delivery. While a longer pulse length and/or a greater RNP concentration significantly enhanced the editing efficiency, they compromised embryo viability as fewer embryos developed to the morula stage post electroporation (Fig. 8F, Table 3). This decreased embryo survival rate is possibility due to off-target Cas9 cleavage when excessive Cas9/sgRNA RNPs were delivered to mouse zygotes. Based on our optimizations, CRISPR-EZ experiments using 8 uM Cas9/sgRNA RNP with two pulses of 3 msec electroporation at 30V achieved an ideal balance between efficient editing and optimal embryo survival (67% bi-allelic editing and 60% survival to morula embryo). However, longer pulse conditions or increased pulse numbers can be used to enable better CRISPR editing efficiency at the cost of embryo viability, when the sgRNA design is suboptimal.

CRISPR-EZ facilitates efficient NHEJ and HDR-based editing

To evaluate the robustness of the CRISPR-EZ technology, we targeted three additional genes, *Cdh1, Cdk8,* and *Kif11,* for NHEJ-mediated editing in mouse embryos. In each case, sgRNAs were designed to target a restriction site 3-4 nt upstream of the PAM, thus allowing us to diagnose NHEJ editing efficiency by RFLP analyses. While CRISPR-EZ editing efficiency varied

across different sgRNA designs, at least 50% of mouse embryos exhibited editing for each experiment, as demonstrated by RFLP analyses and sequencing validation (Fig. 9A, 9B, Table 3). Thus, CRISPR-EZ efficiently delivers Cas9/sgRNA RNPs to broadly induce indel mutations through the NHEJ pathway in mouse embryos. Since Cas9/sgRNA RNPs can be delivered to mouse zygotes with 100% delivery efficiency in CRISPR-EZ experiments (Fig. 8E), varying editing efficiencies can likely be attributed to sgRNA design.

While NHEJ-mediated CRISPR editing is a powerful means to engineer loss of function mutations, HDR-mediated CRISPR editing enables more precise and sophisticated editing schemes. To determine the efficiency of HDR-mediated editing in CRISPR-EZ experiments, we designed a 92 nt ssDNA donor oligo that substitutes the endogenous HinfI site in *Tyr* exon 1 for an EcoRI site, causing a frameshift mutation and early termination of the open reading frame (Fig. 9C). Using CRISPR-EZ, we co-electroporated Cas9/sgRNA RNPs with the ssDNA donor oligo into mouse zygotes, and obtained 46% morula embryos that harbored the precise sequence modification we engineered (Fig. 9D). Thus, Cas9/sgRNA RNPs, together with an ssDNA HDR donor, can be delivered into mouse zygotes by CRISPR-EZ to achieve highly efficient HDR-mediated editing.

CRISPR-EZ generates HDR-mediated genome modifications in mice

As Tyr is the rate-limiting enzyme in pigment synthesis[237,238], the extent of albino coat color in edited mice can be used as a proxy for bi-allelic *Tyr* inactivation. Mosaicism in bi-allelic editing, which was otherwise difficult to determine by RFLP, can be estimated by the extent of coat color mosaicism in the edited animals. To assess *Tyr* editing in live mice, we performed CRISPR-EZ using both 1 msec and 3 msec pulse length conditions to deliver Cas9/sgRNA RNPs and the ssDNA HDR donor into mouse zygotes, cultured electroporated zygotes to the 2-cell stage, and transferred viable 2-cell embryos to the oviducts of pseudopregnant mothers. The 3 msec pulse length condition resulted in 100% efficiency in Cas9 RNP delivery, as all live pups exhibited evidence of HDR and/or NHEJ editing in RFLP analyses (Fig. 10D, 10E, Table 5). Notably, 88% (29/33) of animals appeared completely albino, indicating bi-allelic *Tyr* editing; 9% (3/33) were mosaic with a ~50% albino coat, suggesting that bi-allelic *Tyr* editing occurred after S-phase initiation in 1-cell zygotes; and 3% (1/33) had a black coat yet carried an edited *Tyr* allele (Fig. 10B, Table 4), suggesting complete or partial mono-allelic *Tyr* editing. We speculate that the timing of Cas9/sgRNA delivery is crucial for efficient gene editing in mice, as any editing events occurring after S-phase initiation in 1-cell zygotes could lead to partial editing and less efficient germline transmission of the edited allele(s). While slightly increasing the live birth rate (Figure 10C), the 1 msec pulse length condition, yielded less efficient editing, with 27% (12/44) albino mice and 27% (12/44) efficiency in HDR-mediated editing (Fig. 10B, Table 5). Remarkably, when compared to a standard microinjection-based CRISPR experiment using the same *Tyr* sgRNA, both CRISPR-EZ pulse conditions offered a significant improvement on the live birth rate of edited mice (Fig. 10C, Table 4). To better appreciate the exact molecular outcome of these editing events, we then performed RLFP analysis on DNA isolated from tail samples of the mice. For the 3 msec pulse condition, 42% (13/31) of the assayed animals harbored the HDR-mediated precise modification (Fig. 10D, Table 5), one of which was homozygous HDR-edited. Using sequencing confirmation, all tested albino mice harbored NHEJ and/or HDR editing as determined by RLFP (Fig. 10E), including some that appeared to be bi-allelic HDR edited.

CRISPR-EZ mediates large deletions and small insertions

In addition to indel mutations and small sequence modifications, we also successfully applied CRISPR-EZ to generate a genomic deletion (~720 bp in length), and a precise sequence insertion (42bp in length) in mouse embryos. By co-electroporating RNPs containing two sgRNAs flanking exon 3 of the *Mecp2* gene[218] (Fig. 11A)*,* we generated 71% (17/24) morula embryos harboring a ~720 bp deletion of the intervening sequence, as determined by PCR analyses and validated by sequencing (Fig. 11B, 11C). Additionally, using CRISPR-EZ, we introduced Cas9/sgRNA RNPs and a 162 nt ssDNA donor oligo that directed the insertion of a V5 tag sequence (42 bp) to the 3' end of the *Sox2* ORF[218] (Fig. 12A), achieving 31% insertion efficiency (5/16) (Fig 12B). Successful integration of the V5 epitope tag to the endogenous *Sox2* ORF was demonstrated by PCR and by immunofluorescence staining in blastocyst mouse embryos (Fig. 12B, 12C). Thus, CRISPR-EZ allows efficient delivery of Cas9/sgRNA RNPs and ssDNA donors into mouse zygotes to mediate a variety of editing schemes. Taken together, the CRISPR-EZ technology generates both NHEJ and HDR-edited mice with unprecedented ease, speed, throughput, and efficiency, while significantly improving animal viability.

Discussion

A significant advantage of the CRISPR/Cas9 system is the possibility to bypass ESC manipulation by directly engineering the genome of zygotes. The current standard practice requires microinjection of Cas9 DNA or mRNA together with sgRNAs into zygotes; yet this procedure remains rate-limiting due to its laborious and costly nature and high technical barriers. In contrast, CRISPR-EZ utilizes commonly available reagents and equipment, and can be performed by laboratory personnel with basic training in embryo manipulation. With our optimized conditions, CRISPR-EZ delivers Cas9/sgRNA RNPs into mouse zygotes with 100% efficiency, allowing for a variety of genome editing schemes with unprecedented efficiency, while significantly improving embryo viability compared to microinjection. Finally, electroporation is a rapid procedure that can be applied to many zygotes simultaneously, hence enabling high-throughput production. Given the lengthy, costly, and laborious nature of the microinjection procedure, and its negative impact on embryo/pup viability, CRISPR-EZ is superior to microinjection, and can become the primary methodology for *in vivo* CRISPR editing in mice, and possibly other mammals.

Methods

in vitro synthesis of sgRNAs
Candidate sgRNA designs were selected from a number of algorithms, including Sequence Scan for CRISPR (SSC)[239], the Gene Perturbation Platform (GPP)[240], Chop-Chop[241] and CRISPR Design[242]. For most experiments, we selected three to four candidate sgRNAs based on the predicted scores from multiple sgRNA design algorithms and the proximity to desired target sites. We then experimentally determined the best sgRNA design by measuring targeted DNA cleavage efficiency using the Surveyor assay[243] in a Cas9-overexpressing 368T1 mouse lung cancer cell line that harbors a Kras[G12D] mutation and a p53 deletion.

To *in vitro* synthesize sgRNAs, a DNA oligonucleotide template that contained a T7 promoter, a 20 nucleotide (nt) guide sequence and a sgRNA scaffold[236], was generated by overlapping polymerase chain reaction (PCR)[233]. Specifically, we performed PCR reactions using Phusion® High-Fidelity DNA Polymerase (NEB, Cat# M0530), with the annealed product from a uniquely designed oligo (5'-GGA TCC TAA TAC GAC TCA CTA TAG---guide-sequence---GTT

TTA GAG CTA GAA-3', 0.02 µM) and a common oligo T7RevLong (5'AAA AAA GCA CCG ACT CGG TGC CAC TTT TTC AAG TTG ATA ACG GAC TAG CCT TAT TTT AAC TTG CTA TTT CTA GCT CTA AAA C-3', 0.02 µM) as the template, and T7FwdAmp (5'-GGA TCC TAA TAC GAC TCA CTA TAG-3', 1 µM) and T7RevAmp (5'-AAA AAA GCA CCG ACT CGG-3', 1 µM) as two common primers (Supplementary Fig. S1A). All sgRNA sequences used in this analysis are listed in Table 4. The thermocycler setting consisted of 30 cycles of 95°C for 10 s, 57°C for 10 s and 72°C for 10 s. A 20 µl *in vitro* transcription reaction consisting of 25ng/µl of PCR amplified DNA template, 10mM NTPs and 1 unit T7 RNA polymerase (NEB, Cat# E2040S) was incubated at 37ºC for more than 18hrs, followed by a brief treatment of RNase-Free DNase I (NEB, Cat# M0303S, 2 units) at room temperature for 20 min. The *in vitro* synthesized sgRNAs were cleaned up by magnetic beads that allowed solid-phase reversible immobilization (SPRI) of RNAs[244]. The *in vitro* transcription reaction was first brought to 150µl in volume with 100% ethanol, followed by gentle mixing of 100µl of SeraMeg Speedbeads magnetic carboxylate modified particles (GE Healthcare, Cat # 65152105050250) for 10 times before a 5 min room temperature incubation. The reaction was subsequently placed on a magnetic stand (Invitrogen, Cat# 12321D) for 5 minutes under room temperature to allow the formation of compact RNA/bead pellets. After the supernatant was carefully aspirated by pipet, we washed the pellets gently with 80% ethanol three times (2 minute wash each time, without pipetting), and air dried the pellets for ten minutes. sgRNAs bound to the beads were eluted by incubating with 20µl of RNase-Free H_2O (Ambion, Cat# AM9937) and stored at -80ºC.

Assembly of Cas9/sgRNA RNPs

To assemble the Cas9/sgRNA RNPs, we incubated purified Cas9 Protein (QB3 Macrolab, UC Berkeley) in a 1:1.5 molar ratio with sgRNAs to obtain a final concentration of 8 or 16 uM Cas9/sgRNA RNPs in a 10ul solution containing 20mM HEPES pH7.5 (Sigma, Cat# H3375), 150mM KCl (Sigma, Cat# P9333), 1mM $MgCl_2$ (Sigma, Cat# M8266), 10% glycerol (Fisher, Cat# BP229) and 1mM reducing agent TCEP (*tris*(2-carboxyethyl)phosphine, Sigma, Cat# C4706). a final concentration of 8 or 16 uM Cas9/sgRNA RNPs. Whenever appropriate, 200 pmol of HDR single stranded DNA (ssDNA) oligo donor (up to 162 nt in length) was also included in the 10ul reaction by diluting a concentrated ssDNA stock solution in distilled water. All donor oligo sequences used in this analysis are listed in Table 4. The Cas9/sgRNA RNP complex was prepared by incubating the mixture at 37ºC for 10 min immediately before electroporation.

Delivery of the Cas9/sgRNA RNPs to mouse zygotes by CRISPR-EZ

Three-to-five week old female C57BL/6J mice (Jackson Laboratory, Cat# 000664) were superovulated by intraperitoneal (i.p.) administration of 5IU of Pregnant Mare Serum Gonadotropin (PMSG, Calbiochem, Millipore, Cat# 367222), and 46-48 hours later, 5IU Human Chorion Gonadotropin (hCG, Calbiochem (Millipore, Cat# 230734). Superovulated females were mated at a 1:1 ratio with 3-8 month old C57BL/6J males to generate 1-cell zygotes at 0.5 days post coitum (0.5 dpc). Under a stereomicroscope (Nikon SMZ-U or equivalent), the ampulla of oviduct was nicked, releasing fertilized zygotes and oocytes associated with surrounding cumulus cells into 50ul M2+BSA media, consisting of M2 media (Millipore, Cat# MR-015-D) supplemented with 4mg/ml bovine serum albumin (BSA, Sigma, Cat# A3311). Using a handheld pipette set to 50uL, zygotes were subsequently dissociated from cumulus cells after the cell clumps were transferred into a 200µl droplet of Hyaluronidase/M2 solution (Millipore, Cat# MR-051-F), incubated for 1 minute, and passed through five washes in the M2+BSA media to remove cumulus

cells. All embryos from this point on were manipulated by mouth-pipetting with the use of a 15-inch aspirator tube (Sigma, Cat# A5177), and a hand-made glass needle fashioned by glass pulling of capillary tubes (Sigma P0674) over an open flame. Embryo are passed through five washes of M2+BSA to remove cumulus cells. With as little additional volume as is reasonable, embryos are transferred to a 200uL droplet of Acid Tyrode's Solution (Sigma T1788). Subsequently, zygotes were transferred to a 200µl droplet of acidic Tyrode's Solution (Sigma, Cat# T1788) to weaken the zona pellucida in preparation for the electoroporation. Due to batch-to-batch variation of the acidic Tyrode's solution, the exact timing of the acidic Tyrode's treatment needs to be determined empirically by observing the thinning of zona pellucida under the stereomicroscope. Typically, we incubated the zygotes in the acidic Tyrode's solution until ~15-20% zona pellucida was digested by visual inspection, which typically took 30-60 seconds. Untreated embryos served as a useful control for determining the appropriate timing for acidic Tyrode's treatment. The proper weakening of the zona is critical for the efficient electroporation of the Cas9/sgRNA RNPs into zygotes, yet prolonged zygote exposure to the acidic Tyrode's solution can lead to reduced embryo viability. Treated zygotes were subsequently washed four times in M2+BSA droplets to remove acidic Tyrode's solution. Subsequently, zygotes can be temporarily cultured in M2+BSA in a water jacketed CO_2 incubator, (5% CO_2, 37°C and 95% humidity) until electroporation.

During electroporation, ~30-40 zygotes were pooled and washed once with Opti-MEM reduced serum media (Thermo Fisher Scientific, Cat# 31985062) to remove the M2+BSA media. Subsequently, 30-40 zygotes in 10ul of Opti-MEM reduced serum media were combined with 10ul freshly made Cas9/sgRNA RNP solution with or without the corresponding HDR oligos. Using a standard handheld pipette, the 20µl embryo and RNP mixture was pipetted into a 1mm electroporation cuvette (Biorad, Cat# 1652089) and loaded into the Biorad Gene Pulser Xcell electroporator. A standard square wave electroporation was performed using two pulses at 30V for 3msec, which were separated by a 100msec interval. Immediately following electroporation, zygotes were recovered from the cuvette by flushing with 100µl of KSOM+AA media (KCl-enriched simplex optimization medium with amino acid supplement, Zenith Biotech, Cat #ZEKS-050) once or twice. Embryos were then washed once in KSOM+BSA media that was equilibrated at least 3-4hrs prior to the start of the CRISPR-EZ experiment, and then cultured in 20ul droplets of KSOM+BSA in 35x10mm culture dishes (CellStar Greiner Bio-One 627160) in a water jacketed CO_2 incubator, (5% CO_2 37°C and 95% humidity). The embryos that successfully developed into 2-cell embryos were transferred into the oviduct of CD1 pseudopregnant females, with ~10 embryos per oviduct. The viable pups were then subjected to genotyping and phenotyping analyses.

Delivery of Cas9 mRNA and sgRNA to mouse zygotes by microinjection

UC-Berkeley transgenic facility performed all microinjection experiments under the standard protocol. Pronucleus embryos were pre-selected from collected superovulated embryos by visual inspection for the presence of the second pronuclei. Microinjection was performed in M2 media (Sigma, Cat# M7176), using an inverted microscope (Nikon Corporation, Tokyo, Japan) and micromanipulators (Narishige, Tokyo, Japan). A solution containing 100ng/µl Cas9 mRNA (Life Technologies, Cat# A25640) and 50ng/µl *in vitro* transcribed sgRNA were injected into pronucleus embryos by microinjection. After microinjection, the embryos were cultured in KSOM in a CO_2 incubator (5.0% CO_2 at 37°C) overnight, surviving 2-cell stage embryos were transferred to 0.5dpc CD1 pseudopregnant mothers via oviduct transfer.

RFLP and genotyping analyses

  To extract DNA from cultured morula embryos, embryos were washed twice with PBS, and 1ul of PBS solution containing a single embryo was transferred into 10ul of embryo lysis buffer containing 50mM KCl (Fisher, Cat# P217-3), 10mM Tris-HCl PH=8.5 (Fisher, Cat # BP1531), 2.5mM MgCl2 (Fisher, Cat# M33-500), 0.1mg/mL Gelatin (Fisher, Cat# G7-500), 0.45% NP-40 (Fluka, Cat# 74385), 0.45% Tween-20 (Sigma-Aldrich, Cat# P7949-500) and 0.2mg/mL Proteinase K (Fisher, Cat# BP1700-100). Lysis was performed in a thermocycler with the following conditions: 55°C for 4 hours, 95°C for 10 minutes, and 10°C hold. To extract DNA from mouse tails, we used a standard chloroform extraction protocol.

  Following DNA isolation, PCR was performed using GoTaq (Promega: M712). 3ul of the embryo lysis solution and 20ng of tail DNA were used as the PCR templates for embryo and mouse genotyping, respectively, to generate an amplicon containing the edited region. Nested PCR reactions were then performed using 1ul of 1:10 diluted primary PCR product. For *Tyr, Cdh1, Cdk8,* and *Kif11* editing experiments, nested PCR products were further subject to RLFP analyses for genotyping (see below); for the *Mecp2* deletion, PCR genotyping was performed using nested primers flanking the deleted sequences; for the V5 insertion into *Sox2*, PCR genotyping was performed using nested primers, with two primers residing outside the V5 sequence sequence and one primer residing within the V5 sequence.

  Using the nested PCR products amplified from embryo or mouse DNA, we designed a restriction fragment length polymorphism (RFLP) based strategy to distinguish unedited, NHEJ, or HDR events for the *Tyr* editing experiments. All primer sequences used in this analysis are listed in Table 4. We first identified a unique restriction site within the unedited amplicon near the PAM, and designed sgRNAs that were predicted to disrupt the endogenous restriction site by a successful NHEJ or HDR editing. For HDR editing of the *Tyr* gene, we designed the donor oligo to replace the endogenous HinfI restriction site with a new EcoRI restriction site while simultaneously disrupting the target sequence to prevent secondary editing. For restriction digest, 3 uL of PCR product was added to a 10 uL reaction containing 1 U of the restriction enzyme and incubated at 37° for 4 hours. Digested PCR products were electrophoresed on a 2% agarose gel and imaged on a Bio-Rad Gel Doc XR+. The following restriction enzymes were used: HinfI (NEB, Cat# R0155S), EcoRI (NEB, Cat# R0101S), XmaI (NEB, Cat# R0180S), EcoNI (NEB, Cat# R0521S), BslI (NEB, Cat# R0555S).

  To validate the CRISPR editing by Sanger sequencing, nested PCR products were gel purified using GeneJET Gel Extraction Kit (Thermo Scientific, Cat # K0691) and cloned into pGEM-T Easy vector (Promega, Cat# A1360) by TA cloning. Plasmid DNA were submitted for Sanger sequencing to determine the DNA sequences in edited embryo or mouse.

Embryo Immunofluorescence (IF)

  Blastocyst embryos were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, Cat # 19202) for 15 min at room temperature and then transferred to PBS containing 0.1% bovine serum albumin (BSA, Sigma, Cat # A3311). Permeabilization was performed by incubating the embryos with PBS containing 0.1% Triton X-100 and 0.1% BSA for 5 minutes. Subsequently, embryos were blocked for 1 hour at room temperature in blocking solution (PBS containing 10% goat serum Fisher Cat#31872, and 0.1% BSA). Embryos were then incubated with anti-V5 primary antibody (1:100, ThermoFisher, Cat # R960-25) in blocking solution at 4°C overnight, and subsequently in Alexa Fluor 594 Goat anti mouse IgG (1:400, ThermoFisher, Cat # A21125) in

blocking solution at 4°C overnight. Embryos were imaged using a Zeiss Observer A1 fluorescent microscope.

Purpose	Name	Sequence
sgRNA sequences	sgTyr	GGGTGGATGACCGTGAGTCC
	sgCdh1	TATGACTGGAGTCCCGGGCG
	sgCdk8	AGACAGAAACACCTTCAGAA
	sgKif11	CGTGGAATTATACCAGCCAG
	sgMecp2 L	CCCAAGGATACAGTATCCTA
	sgMecp2 R	AGGAGTGAGGTCTAGTACTT
	sgSox2	TGCCCCTGTCGCACATGTGA
Nested Primers for RLFP and genotyping	sgTyr F1	TCTTTTCGGAGACACTCAAATCA
	sgTyr F2	TCTGTACAATTTGGGCCCCC
	sgTyr R1	GCTTTCAGGCAGAGGTTCCT
	sgCdh1 F1	TCTCCGGGTAGGGTTGTTCA
	sgCdh1 F2	CCTGTCTGTGATCTGTCCACTT
	sgCdh1 R1	CCAACAAGTCCCCAGTGCTA
	sgCdk8 F1	CACTTCCAAGCAGCCAGGTA
	sgCdk8 F2	GGCCGTGGCATATCCTTGTA
	sgCdk8 R1	GGTGACTCCTAGTGCAGTGG
	sgKif11 F1	GGATGGGAGGTGTAGCTGAG
	sgKif11 F2	GGATCAGTCCTCAGTGTTGCA
	sgKif11 R1	CCTTGTTCCGGGGATCATCAA
	sgMecp2 F1	GGCCAGATGCATGGGTAGAA
	sgMecp2 F2	TGAAAACAGAGGACCTGCCG
	sgMecp2 WT R1	CCTTGCCTGAAGGTTGGACA
	sgMecp2 KO R1	TTGTCATGTGGCAAGCCCA
	sgSox2 F1	ACATGATCAGCATGTACCTCC
	sgSox2 V5 F2	ACATGGGCAAGCCCATCC
	sgSox2 R1	TAATTTGGATGGGATTGGTGG
ssDNA HDR Donor oligos	Tyr ssDNA	GTGCACCATCTGGACCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGAATTCCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGTGCTC
	Sox2 V5 ssDNA	TACCAGAGCGGCCCGGTGCCCGGCACGGCCATTAACGGCACACTGCCCCTGTCGCACATGGGCAAGCCCATCCCCAACCCCCTGCTGGGCCTGGACAGCACCTGAGGGCTGGACTGCGAACTGGAGAAGGGGAGAGATTTTCAAAGAGATACAAGGGAATTG

Table 2: Primer and oligo sequences used in Chapter 3

Gene	CRISPR-EZ conditions	Zygotes treated	Embryos survived to morula	Embryos assayed by RFLP	Editing		
					Bi-allelic	Partial	Unedited
Tyr	16 uM Cas9 RNPs, 1 msec	29	8 (28%)	8	1 (13%)	7 (87%)	0
	16 uM Cas9 RNPs, 3 msec	24	6 (25%)	6	5 (83%)	1 (17%)	0
	16 uM Cas9 RNPs, 10 msec	28	5 (18%)	5	5 (100%)	0	0
	8 uM Cas9 RNPs, 1 msec	30	19 (63%)	12	2 (16%)	9 (75%)	1 (8%)
	8 uM Cas9 RNPs, 3 msec	30	18 (60%)	12	8 (67%)	4 (33%)	0
	8 uM Cas9 RNPs, 10 msec	30	12 (40%)	12	11 (92%)	1 (8%)	0
Cdh1	8 uM Cas9 RNPs, 3 msec	35	n.d.	25	0	14 (56%)	11 (44%)
Cdk8	8 uM Cas9 RNPs, 3 msec	35	n.d.	22	3 (14%)	19 (86%)	0
Kif11	8 uM Cas9 RNPs, 3 msec	35	n.d.	24	1 (4%)	12 (50%)	11 (46%)

Table 3. NHEJ editing efficiency and embryo viability under multiple CRISRP-EZ conditions

Genome editing method	Embryos treated	Embryos transferred	Mice born	Coat Color		
				Black	Albino	Mosaic
mRNA injection	136 (93)[a]	60	5	1 (20%)	3 (60%)	1 (20%)
CRISPR-EZ, 1 msec	140	85	44	23 (53%)	12 (27%)	9 (20%)
CRISPR-EZ, 3 msec	120	90	33	1 (3%)	29 (88%)	3 (9%)

Table 4. Phenotype and viability of *Tyr* edited embryos and mice by CRISPR-EZ
[a]93 of 136 collected embryos were injected after screening for presence of pronuclei. In contrast, CRISPR-EZ was performed on all collected embryos.

CRISPR-EZ pulse length	Mice assayed by RFLP	Type of editing events detected by RFLP analyses					
		Unedited	Unedited; NHEJ	Unedited; HDR	NHEJ; HDR	Bi-allelic NHEJ	Bi-allelic HDR
1 msec	44	18 (41%)	6 (14%)	2 (5%)	8 (18%)	8 (18%)	2 (5%)
3 msec	31	0	2 (6%)	1 (3%)	11 (35%)	16 (52%)	1 (3%)

Table 5. Efficiency of NHEJ and HDR editing in CRISPR-EZ edited mice

A CRISPR-EZ

superovulation → embryo harvest → electroporation → oviduct transfer → CRISPR edited mice

sgRNA synthesis → RNP assembly → culture to morula → RFLP genotyping

B *Tyr*

200 bp

Ex 1   Ex 2   Ex 4   Ex 5

HinfI PAM

5′ GTGCACCATCTGGACCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGAGTCCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGTGCTC 3′

NHEJ

PAM

5′ GTGCACCATCTGGACCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGA CCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGTGCTC 3′

HinfI

C

1 msec PCR / HinfI ◄edited ◄unedited

3 msec PCR / HinfI ◄edited ◄unedited

10 msec PCR / HinfI ◄edited ◄unedited

D

HinfI

WT CCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGAG TCCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGT

#1 CCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGAGC TCCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGT

#2 CCTCAGTTCCCCTTCAAAGGGGTGGATGACCGTGAGCC TCCTGGCCCTCTGTGTTTTATAATAGGACCTGCCAGT

E NHEJ Editing

% of embryos

n=12 n=12 n=12   n=8 n=6 n=5

bi-allelic / partial / unedited

Pulse length: 1 ms 3 ms 10 ms   1 ms 3 ms 10 ms

RNP concentration: 8 uM   16 uM

F Survival

% of embryos

n=30 n=30 n=30   n=29 n=24 n=28

Pulse length: 1 ms 3 ms 10 ms   1 ms 3 ms 10 ms

RNP concentration: 8 uM   16 uM

44

Figure 8. Optimization of CRISPR-EZ conditions
A. Diagram illustrating the workflow of the CRISPR-EZ technology in mouse genome editing. Fertilized embryos are combined with pre-assembled Cas9/sgRNA RNPs for electroporation, and then transferred to pseudopregnant mothers to generate edited mice. **B**. A gene schematic illustrating the NHEJ-mediated editing design that targets *Tyr* exon1. The HinfI restriction site within *Tyr* exon 1, located 2 nt upstream of the PAM, will be disrupted upon successful NHEJ editing. Arrows indicate the positions of primers that amplify a DNA fragment for RFLP genotyping analyses. **C, D.** Representative RFLP genotyping analyses (**C**) and sequencing confirmation (**D**) of *Tyr* NHEJ editing in mouse morula embryos using 8uM Cas9/sgRNA RNPs under different electroporation conditions. **C**. The presence of an undigested PCR product (200 bp) indicates successful NHEJ editing. Top: nested PCR amplicons from morula embryos following CRISPR-EZ; bottom: HinfI digestion using nested PCR amplicons in RFLP analyses. **D.** Chromatograms and alignment of sequences from two edited mouse morula embryos compared to the wildtype *Tyr* sequence. Red boxes indicate edited sequences. **E, F.** Key experimental conditions in CRISPR-EZ were optimized to achieve high editing efficiency on *Tyr* and favorable embryo viability in culture. Three electroporation conditions and two Cas9 RNP concentrations were compared for *Tyr* NHEJ editing efficiency (**E**) and for embryo viability (**F**). **F.** Percent survival was calculated as the ratio between the number of embryos that developed to the morula stage and the total number of zygotes electroporated.
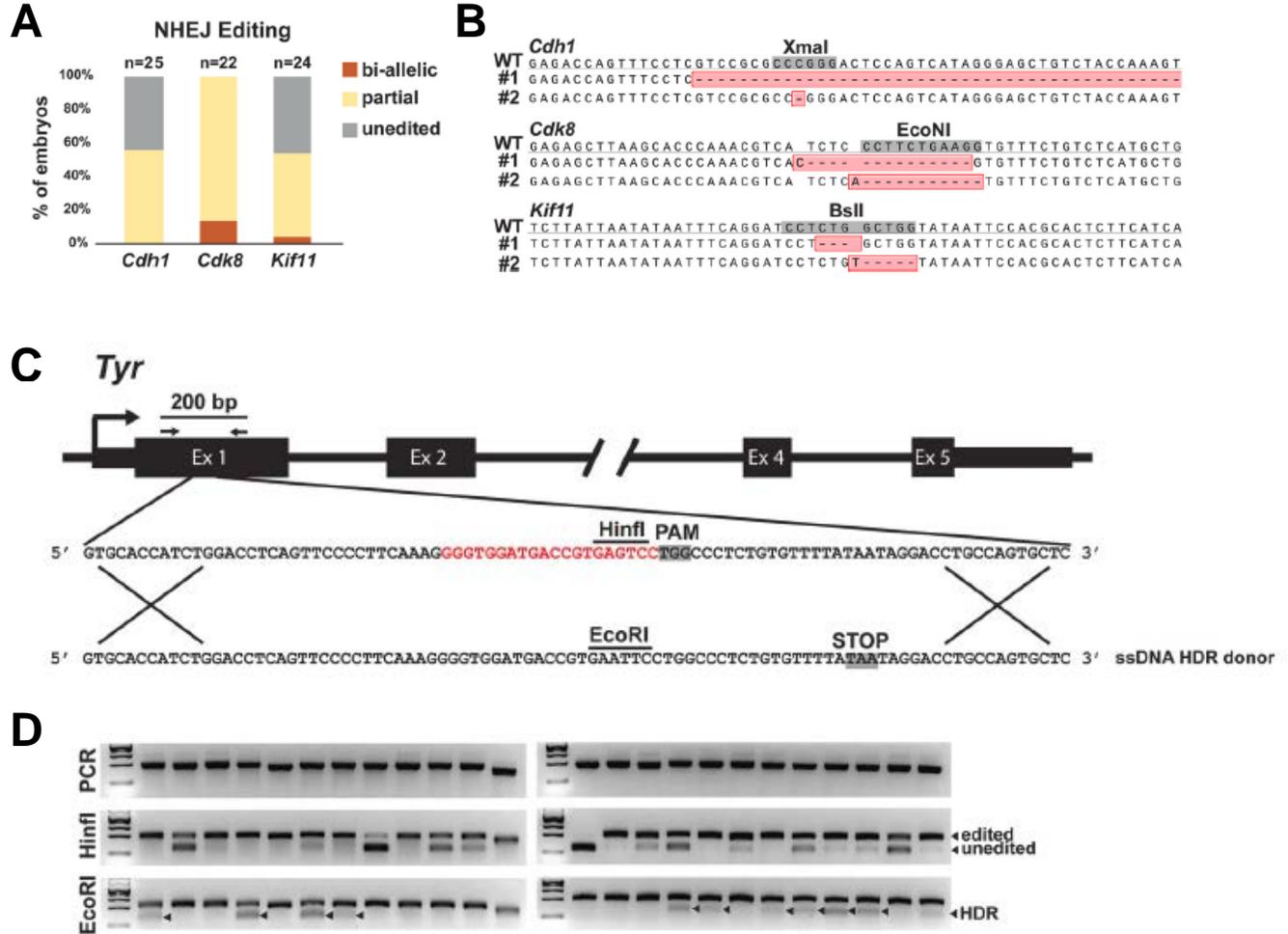
Figure 9. CRISPR-EZ efficiently induces NHEJ and HDR-mediated edits
A. CRISPR-EZ achieved efficient NHEJ editing of *Cdh1*, *Cdk8* and *Kif11*. Following CRISPR-EZ using optimized conditions (8uM Cas9/sgRNA RNPs, 2 pulses of electroporation at 3 msec pulse length), the efficiency of NHEJ editing on *Cdh1*, *Cdk8* and *Kif11* were measured by RFLP analyses. **B**. Sequence validation is shown for representative *Cdh1*, *Cdk8* and *Kif11* editing events. **C**. Diagram illustrating the HDR editing scheme that targets *Tyr* exon 1. A synthesized 92nt ssDNA donor oligo directs HDR-mediated editing, which replaces the endogenous HinfI restriction site with an EcoRI site, and causes a frameshift mutation to introduce a premature stop codon in *Tyr* exon 1. Arrows indicate the positions of primers that amplify a 200 bp DNA fragment for RFLP genotyping analyses. **D.** RFLP genotyping analyses revealed the successful NHEJ and HDR editing in morula embryos. Nested PCR amplicons (top row) were digested with both HinfI (middle row) and EcoRI (bottom row) to assay for NHEJ editing and HDR editing, respectively. HDR-specific digestion products (~100 bp, migrate as one band) are marked with black arrowheads.

A mRNA injection   CRISPR-EZ (1 msec)   CRISPR-EZ (3 msec)

B Coat Color

C Viability

- mRNA injection (*Tyr*)
- mRNA injection (UC-Berkeley transgenic facility avg)
- 1 msec CRISPR-EZ (*Tyr*)
- 3 msec CRISPR-EZ (*Tyr*)

D albino mice (1 msec)   albino mice (3 msec)

PCR   HinfI   EcoRI   ◄edited   ◄unedited   ◄HDR

E

Figure 10. CRISPR-EZ generates live edited mice with precise edits
A. Representative *Tyr* edited mouse litters for a microinjection experiment (left), a CRISPR-EZ experiment at 1 msec pulse length (middle) and a CRISPR-EZ experiment at 3 msec pulse length (right). **B.** Quantification of coat color phenotypes of live, *Tyr* edited mice generated by microinjection and CRISPR-EZ experiments. **C.** CRISPR-EZ significantly improves mouse viability after genome editing compared to microinjection-based experiments. UC Berkeley transgenic facility averages were calculated based on data collected across recent 5 CRISPR experiments that inject *cas9* mRNA and sgRNAs for genome editing. **D, E.** Albino mice obtained from CRISPR-EZ experiments were subjected to RFLP genotyping analyses (**D**) to demonstrate NHEJ and/or HDR editing; and select albino mice were sequence confirmed (**E**). HDR-specific digestion products (~100 bp, migrate as one band) are marked with black arrowheads (**D**). Red boxes indicate edited sequences, while red letters indicate the HDR-mediated precise modification (**E**).
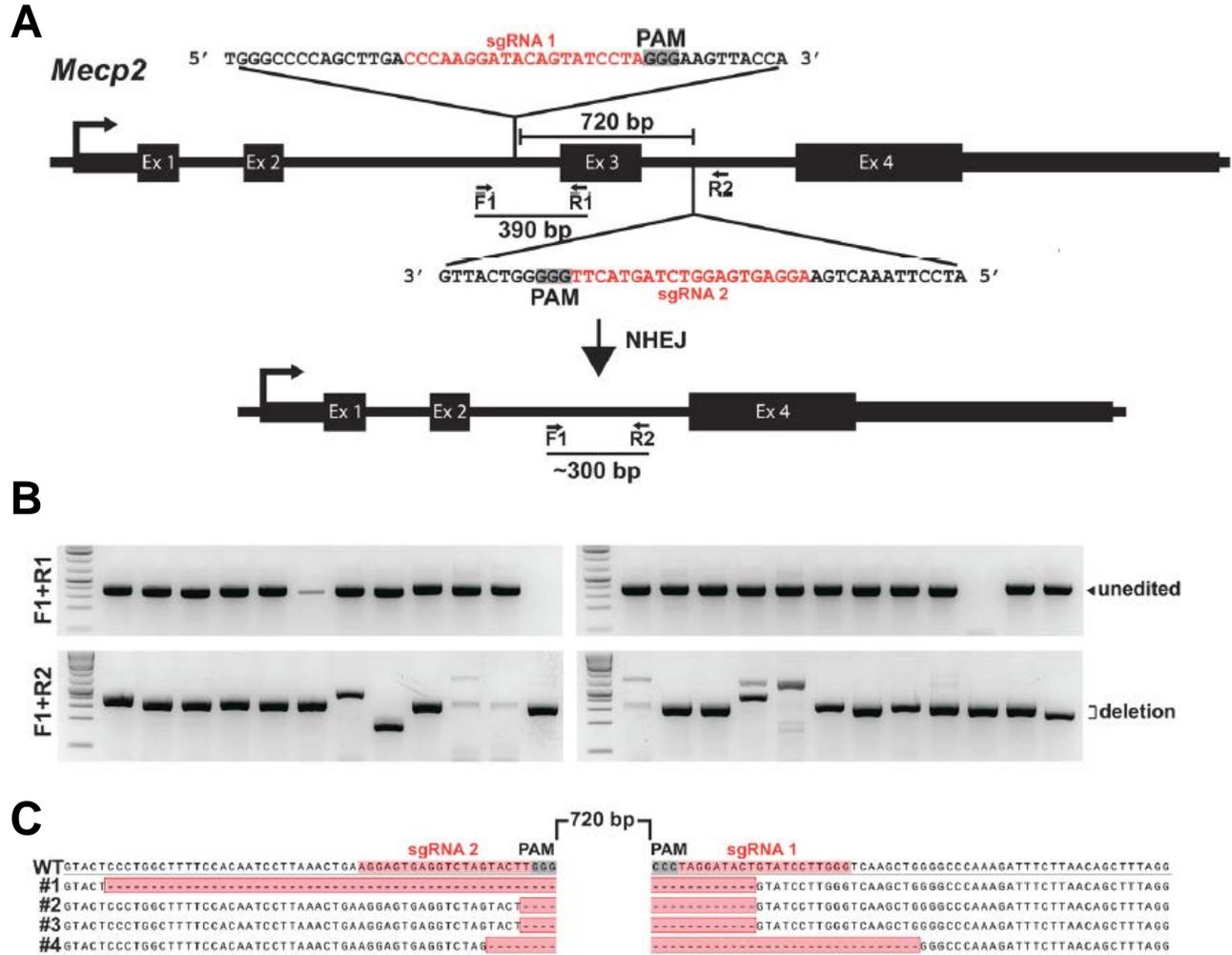
Figure 11. CRISPR-EZ efficiently facilitates large deletions.
A. Diagram illustrating the editing scheme to delete exon 3 of the *Mecp2* gene by CRISPR-EZ. Two sgRNAs were designed to direct Cas9 cleavage in *Mecp2* intron 2 and intron 3 to generate the ~720 bp deletion of exon 3. Arrows indicate the positions of primers used for PCR genotyping that amplifies across the deleted region. **B, C.** Representative PCR genotyping analyses (**B**) and sequencing confirmation (**C**) for assessing the editing efficiency of *Mecp2* in mouse morula embryos. Red boxes indicate deleted sequences.
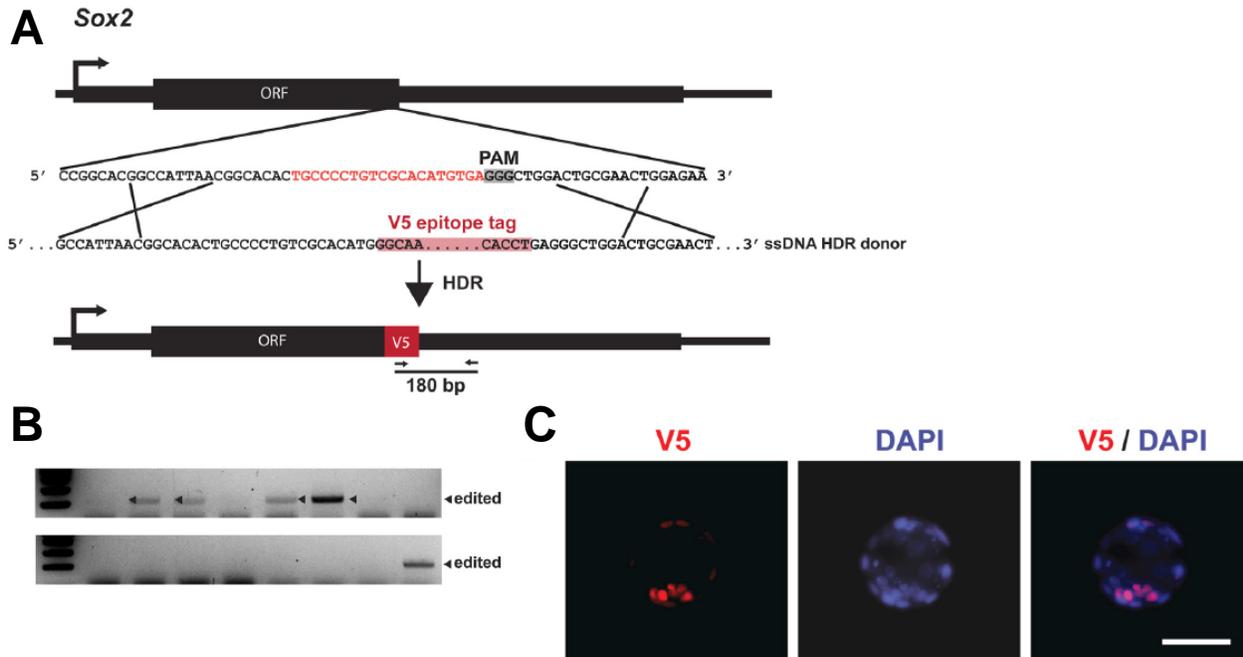
Figure 12. CRISPR-EZ facilitates small insertions
A. Diagram illustrating HDR editing scheme that inserts a V5 epitope tag (42bp in length) at the 3' of the *Sox2* open reading frame via a 162nt ssDNA donor oligo using CRISPR-EZ. Arrows indicate the positions of primers that amplify a 180 bp DNA fragment across the edited genomic region. **B, C.** PCR genotyping analysis (**B**) and immunofluorescence staining (**C**) for assessing the editing efficiency of the *sox2* gene in mouse morula and blastocyst embryos, respectively. Scale bar = 20 µM

References

1. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* **36,** 344–355 (1950).
2. Rebollo, R., Romanish, M. T. & Mager, D. L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46,** 21–42 (2012).
3. Sijen, T. & Plasterk, R. H. A. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature* **426,** 310–4 (2003).
4. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–62 (2002).
5. Vicient, C. M. Transcriptional activity of transposable elements in maize. *BMC Genomics* **11,** 601 (2010).
6. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284,** 601–3 (1980).
7. Orgel, L. E. & Crick, F. H. C. Selfish DNA: the ultimate parasite. *Nature* **284,** 604–607 (1980).
8. McClintock, B. CHROMOSOME ORGANIZATION AND GENIC EXPRESSION. *Cold Spring Harb. Symp. Quant. Biol.* **16,** 13–47 (1951).
9. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41,** 331–68 (2007).
10. Gifford, W. D., Pfaff, S. L. & Macfarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23,** 218–26 (2013).
11. Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D. & Knowles, B. B. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Dev. Cell* **7,** 597–606 (2004).
12. Mager, D. L. & Stoye, J. P. Mammalian Endogenous Retroviruses. *Microbiol. Spectr.* **3,** (2015).
13. Henzy, J. E. & Johnson, W. E. Pushing the endogenous envelope. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120506 (2013).
14. Gifford, R. & Tristem, M. The Evolution, Distribution and Diversity of Endogenous Retroviruses. *Virus Genes* **26,** 291–315 (2003).
15. Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65,** 3383–98 (2008).
16. Wildschutte, J. H., Williams, Z. H., Montesion, M., Subramanian, R. P., Kidd, J. M. & Coffin, J. M. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci.* 201602336 (2016). doi:10.1073/pnas.1602336113
17. Belshaw, R., Dawson, A. L. A., Woolven-Allen, J., Redding, J., Burt, A. & Tristem, M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J. Virol.* **79,** 12507–14 (2005).
18. Katzourakis, A., Rambaut, A. & Pybus, O. G. The evolutionary dynamics of endogenous retroviruses. doi:10.1016/j.tim.2005.08.004
19. Subramanian, R. P., Wildschutte, J. H., Russo, C. & Coffin, J. M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **8,** 90 (2011).
20. Ketting, R. F., Haverkamp, T. H. ., van Luenen, H. G. A. . & Plasterk, R. H. . mut-7 of C. elegans, Required for Transposon Silencing and RNA Interference, Is a Homolog of

Werner Syndrome Helicase and RNaseD. *Cell* **99,** 133–141 (1999).

21. Castro, J. P. & Carareto, C. M. A. Drosophila melanogaster P Transposable Elements: Mechanisms of Transposition and Regulation. *Genetica* **121,** 107–118 (2004).

22. Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. & Antonarakis, S. E. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332,** 164–6 (1988).

23. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B. & Nakamura, Y. Disruption of the APC Gene by a Retrotransposal Insertion of L1 Sequence in a Colon Cancer. *Cancer Res.* **52,** 643–645 (1992).

24. Lamprecht, B. et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16,** 571–9, 1p following 579 (2010).

25. Bannert, N. & Kurth, R. The Evolutionary Dynamics of Human Endogenous Retroviral Families. *Annu. Rev. Genomics Hum. Genet* **7,** 149–73 (2006).

26. Maksakova, I. A., Romanish, M. T., Gagnier, L., Dunn, C. A., van de Lagemaat, L. N. & Mager, D. L. Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. *PLoS Genet.* **2,** e2 (2006).

27. Stoye, J. P., Moroni, C. & Coffin ', J. M. Virological Events Leading to Spontaneous AKR Thymomas. *J. Virol.* 1273–1285 (1991). at <http://jvi.asm.org/content/65/3/1273.full.pdf>

28. Howard, G., Eiges, R., Gaudet, F., Jaenisch, R. & Eden, A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27,** 404–408 (2008).

29. Babaian, A., Romanish, M. T., Gagnier, L., Kuo, L. Y., Karimi, M. M., Steidl, C. & Mager, D. L. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **35,** 2542–2546 (2016).

30. Wiesner, T. et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **526,** 453–7 (2015).

31. Nagai, M., Furihata, T., Matsumoto, S., Ishii, S., Motohashi, S., Yoshino, I., Ugajin, M., Miyajima, A., Matsumoto, S. & Chiba, K. Identification of a new organic anion transporting polypeptide 1B3 mRNA isoform primarily expressed in human cancerous tissues and cells. *Biochem. Biophys. Res. Commun.* **418,** 818–823 (2012).

32. Lock, F. E., Rebollo, R., Miceli-Royer, K., Gagnier, L., Kuah, S., Babaian, A., Sistiaga-Poveda, M., Lai, C. B., Nemirovsky, O., Serrano, I., Steidl, C., Karimi, M. M. & Mager, D. L. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E3534-43 (2014).

33. Scarfò, I. et al. Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. *Blood* **127,** 221–32 (2016).

34. Baudino, L., Yoshinobu, K., Morito, N., Santiago-Raber, M.-L. & Izui, S. Role of endogenous retroviruses in murine SLE. *Autoimmun. Rev.* **10,** 27–34 (2010).

35. Young, G. R., Eksmond, U., Salcedo, R., Alexopoulou, L., Stoye, J. P. & Kassiotis, G. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* **491,** nature11599 (2012).

36. Cherkasova, E., Weisman, Q. & Childs, R. W. Endogenous Retroviruses as Targets for Antitumor Immunity in Renal Cell Cancer and Other Tumors. *Front. Oncol.* **3,** 243 (2013).

37. Antony, J. M., Deslauriers, A. M., Bhat, R. K., Ellestad, K. K. & Power, C. Human endogenous retroviruses and multiple sclerosis: Innocent bystanders or disease determinants? (2011). doi:10.1016/j.bbadis.2010.07.016

38. Jaenisch, R., Schnieket, A., Harbers, K. & Vogt, P. K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues (toxicity of 5-azacytidine/in vivo activation of transcription/recombination of defective proviral genomes). *Cell Biol.* **82,** 1451–1455 (1985).

39. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20,** 116–7 (1998).

40. Brolet, P., Condamine, H. & Jacob, F. Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis (endogenous retrovirus/RNA DNA in situ hybridization/long terminal repeat/early transposon). *Dev. Biol.* **82,** 2054–2058 (1985).

41. Gu, S. G., Pak, J., Guang, S., Maniar, J. M., Kennedy, S. & Fire, A. Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat. Genet.* **44,** 157–64 (2012).

42. Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. a & Tóth, K. F. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27,** 390–9 (2013).

43. Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C. & Shinkai, Y. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464,** 927–31 (2010).

44. Maksakova, I. A., Thompson, P. J., Goyal, P., Jones, S. J., Singh, P. B., Karimi, M. M. & Lorincz, M. C. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6,** 15 (2013).

45. Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B. & Trono, D. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463,** 237–240 (2010).

46. Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R. & Davidson, I. Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. *Nat. Struct. Mol. Biol.* **20,** 339–346 (2013).

47. Thomas, J. H. & Schneider, S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* **21,** 1800–12 (2011).

48. Wolf, G., Yang, P., Füchtbauer, A. C., Füchtbauer, E.-M., Silva, A. M., Park, C., Wu, W., Nielsen, A. L., Pedersen, F. S. & Macfarlan, T. S. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* **29,** 538–54 (2015).

49. Mak, K. S. et al. Repression of chimeric transcripts emanating from endogenous retrotransposons by a sequence-specific transcription factor. *Genome Biol.* **15,** 3252 (2014).

50. Schlesinger, S., Lee, A. H., Wang, G. Z., Green, L. & Goff, S. P. Proviral silencing in embryonic cells is regulated by yin yang 1. *Cell Rep.* **4,** 50–8 (2013).

51. Guallar, D., Pérez-Palacios, R., Climent, M., Martínez-Abadía, I., Larraga, A., Fernández-Juan, M., Vallejo, C., Muniesa, P. & Schoorlemmer, J. Expression of endogenous

retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic Acids Res.* **40,** 8993–9007 (2012).

52.   Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front. Oncol.* **4,** 14 (2014).

53.   Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).

54.   Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364,** 675–87 (2009).

55.   Knisbacher, B. A. & Levanon, E. Y. DNA Editing of LTR Retrotransposons Reveals the Impact of APOBECs on Vertebrate Genomes. *Mol. Biol. Evol.* **33,** 554–567 (2016).

56.   Jern, P., Stoye, J. P. & Coffin, J. M. Role of APOBEC3 in Genetic Diversity among Endogenous Murine Leukemia Viruses. *PLoS Genet.* **3,** e183 (2007).

57.   Arnaud, F., Murcia, P. R. & Palmarini, M. Mechanisms of late restriction induced by an endogenous retrovirus. *J. Virol.* **81,** 11441–51 (2007).

58.   Stoye, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbiol.* **10,** 395 (2012).

59.   Sanz-Ramos, M. & Stoye, J. P. Capsid-binding retrovirus restriction factors: discovery, restriction specificity and implications for the development of novel therapeutics. *J. Gen. Virol.* **94,** 2587–2598 (2013).

60.   Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).

61.   Martin-Serrano, J. & Neil, S. J. D. Host factors involved in retroviral budding and release. *Nat. Rev. Microbiol.* **9,** 519–531 (2011).

62.   Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G. & Heidmann, T. An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* **18,** 597–609 (2008).

63.   Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8,** 272–85 (2007).

64.   Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K. & Svoboda, P. A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. *Cell* **155,** 807–816 (2013).

65.   Berrens, R. V., Andrews, S., Spensberger, D., Santos, F., Dean, W., Gould, P., Sharif, J., Olova, N., Chandra, T., Koseki, H., von Meyenn, F. & Reik, W. An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell* **21,** 694–703.e7 (2017).

66.   Vagin, V. V, Sigova, A., Li, C., Seitz, H., Gvozdev, V. & Zamore, P. D. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–4 (2006).

67.   Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. & Zamore, P. D. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science (80-.).* **313,** 320–324 (2006).

68.   Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. & Hannon, G. J. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128,** 1089–1103 (2007).

69. Sarot, E. Evidence for a piwi-Dependent RNA Silencing of the gypsy Endogenous Retrovirus by the Drosophila melanogaster flamenco Gene. *Genetics* **166,** 1313–1321 (2004).

70. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–7 (2007).

71. Irvine, D. V, Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L. & Martienssen, R. A. Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313,** 1134–7 (2006).

72. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151,** 964–80 (2012).

73. Guang, S., Bochner, A. F., Burkhart, K. B., Burton, N., Pavelec, D. M. & Kennedy, S. Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465,** 1097–101 (2010).

74. Shenoy, A. & Blelloch, R. H. Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat. Rev. Mol. Cell Biol.* **15,** 565–576 (2014).

75. Ambros, V. MicroRNAs and developmental timing. *Curr. Opin. Genet. Dev.* **21,** 511–517 (2011).

76. Bushati, N. & Cohen, S. M. microRNA Functions. *Annu. Rev. Cell Dev. Biol.* **23,** 175–205 (2007).

77. Heras, S. R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyras, E., Garcia-Perez, J. L. & Cáceres, J. F. The Microprocessor controls the activity of mammalian retrotransposons. *Nat. Struct. Mol. Biol.* **advance on,** (2013).

78. Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J. & Schultz, R. M. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269,** 276–85 (2004).

79. Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18097–102 (2007).

80. Hu, X., Zhu, W., Chen, S., Liu, Y., Sun, Z., Geng, T., Wang, X., Gao, B., Song, C., Qin, A. & Cui, H. Expression of the env gene from the avian endogenous retrovirus ALVE and regulation by miR-155. *Arch. Virol.* **161,** 1623–1632 (2016).

81. Choi, Y. J., Lin, C.-P., Risso, D., Chen, S., Kim, T. A., Tan, M. H., Li, J. B., Wu, Y., Chen, C., Xuan, Z., Macfarlan, T., Peng, W., Lloyd, K. C. K., Kim, S. Y., Speed, T. P. & He, L. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science (80-.).* **355,** (2017).

82. Gebetsberger, J. & Polacek, N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* **10,** 1798–1806 (2013).

83. Thompson, D. M. & Parker, R. The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae. *J. Cell Biol.* **185,** 43–50 (2009).

84. Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A. Z. & Kay, M. A. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16,** 673–95 (2010).

85. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent

small RNAs. *Genes Dev.* **22,** 2773–85 (2008).

86.   Cole, C., Sobala, A., Lu, C., Thatcher, S. R., Bowman, A., Brown, J. W. S., Green, P. J., Barton, G. J. & Hutvagner, G. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* **15,** 2147–60 (2009).

87.   Sharma, U. et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science (80-.).* (2015).

88.   Schorn, A. J., Gutbrod, M. J., LeBlanc, C. & Martienssen, R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170,** 61–71.e11 (2017).

89.   Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.-Y., Edouard, P., Howes, S., Keith, J. C. & McCoy, J. M. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–789 (2000).

90.   Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F. & Cosset, F. L. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* **74,** 3321–9 (2000).

91.   Blaise, S., de Parseval, N., Bénit, L. & Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 13013–8 (2003).

92.   Vargas, A., Moreau, J., Landry, S., Lebellego, F., Toufaily, C., Rassart, É., Lafond, J. & Barbeau∗, B. Syncytin-2 Plays an Important Role in the Fusion of Human Trophoblast Cells. *J. Mol. Biol.* **392,** 301–318 (2009).

93.   Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P. & Heidmann, T. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 12127–32 (2009).

94.   Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G. & Heidmann, T. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc. Natl. Acad. Sci. U. S. A.* **108,** E1164-73 (2011).

95.   Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B. C., Catzeflis, F., Mulot, B., Heidmann, O., Heidmann, T. & Dupressoir, A. Capture of syncytin-Mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the Rodentia squirrel-related clade. *J. Virol.* **88,** 7915–28 (2014).

96.   Cornelis, G., Heidmann, O., Bernard-Stoecklin, S., Reynaud, K., Véron, G., Mulot, B., Dupressoir, A. & Heidmann, T. Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E432-41 (2012).

97.   Nakaya, Y., Koshi, K., Nakagawa, S., Hashizume, K. & Miyazawa, T. Fematrin-1 is involved in fetomaternal cell-to-cell fusion in Bovinae placenta and has contributed to diversity of ruminant placentation. *J. Virol.* **87,** 10563–72 (2013).

98.   Cornelis, G., Heidmann, O., Degrelle, S. A., Vernochet, C., Lavialle, C., Letzelter, C., Bernard-Stoecklin, S., Hassanin, A., Mulot, B., Guillomot, M., Hue, I., Heidmann, T. & Dupressoir, A. Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E828-37 (2013).

99. Heidmann, O., Vernochet, C., Dupressoir, A. & Heidmann, T. Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals. *Retrovirology* **6,** 107 (2009).

100. Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O. A., Meurice, G., Heidmann, O., Dupressoir, A., Miralles, A., Ramirez-Pinilla, M. P. & Heidmann, T. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proc. Natl. Acad. Sci. U. S. A.* 201714590 (2017). doi:10.1073/pnas.1714590114

101. Chuong, E. B. Retroviruses facilitate the rapid evolution of the mammalian placenta. *Bioessays* **35,** 853–61 (2013).

102. Dupressoir, A., Lavialle, C. & Heidmann, T. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta* **33,** 663–671 (2012).

103. Best, S., Tissier, P. Le, Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene Fvl. *Nature* **382,** 826–829 (1996).

104. Lilly, F. Fv-2: identification and location of a second gene governing the spleen focus response to Friend leukemia virus in mice. *J. Natl. Cancer Inst.* **45,** 163–9 (1970).

105. Wu, T., Yan, Y. & Kozak, C. A. Rmcf2, a xenotropic provirus in the Asian mouse species Mus castaneus, blocks infection by polytropic mouse gammaretroviruses. *J. Virol.* **79,** 9677–84 (2005).

106. Odaka, T., Ikeda, H. & Akatsuka, T. Restricted expression of endogenous N-Tropic XC-positive leukemia virus in hybrids between G and AKR mice: An effect of the Fv-4r gene. *Int. J. Cancer* **25,** 757–762 (1980).

107. Mcdougall, A. S., Terry, A., Tzavaras, T., Cheney, C., Rojko, J. & Neili, J. C. Defective Endogenous Proviruses Are Expressed in Feline Lymphoid Cells: Evidence for a Role in Natural Resistance to Subgroup B Feline Leukemia Viruses. *J. Virol.* **68,** 2151–2160 (1994).

108. Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., Martin, L., Ware, C. B., Blish, C. A., Chang, H. Y., Reijo Pera, R. A. & Wysocka, J. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522,** 221–225 (2015).

109. Rebollo, R., Farivar, S. & Mager, D. L. C-GATE - catalogue of genes affected by transposable elements. *Mob. DNA* **3,** 9 (2012).

110. Samuelson, L. C., Phillips, R. S. & Swanberg, L. J. Amylase gene structures in primates: retroposon insertions and promoter evolution. *Mol. Biol. Evol.* **13,** 767–779 (1996).

111. Stavenhagen, J. B. & Robins, D. M. An Ancient Provirus Has Imposed Androgen Regulation on the Adjacent Mouse Sex-Limited Protein Gene. *Cell* **55,** 247–254 (1988).

112. Ramakrishnan, C. & Robins, D. M. Steroid hormone responsiveness of a family of closely related mouse proviral elements. *Mamm. Genome* **8,** 811–817 (1997).

113. Ling, J., Pi, W., Bollag, R., Zeng, S., Keskintepe, M., Saliman, H., Krantz, S., Whitney, B. & Tuan, D. The Solitary Long Terminal Repeats of ERV-9 Endogenous Retrovirus Are Conserved during Primate Evolution and Possess Enhancer Activities in Embryonic and Hematopoietic Cells. *J. Virol.* **76,** 2410–2423 (2002).

114. Yu, X., Zhu, X., Pi, W., Ling, J., Ko, L., Takeda, Y. & Tuan, D. The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active

LTR enhancer complex NF-Y/MZF1/GATA-2. *J. Biol. Chem.* **280,** 35184–94 (2005).

115. Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J. & Tuan, D. Long-range function of an intergenic retrotransposon. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12992–7 (2010).

116. Chuong, E. B., Rumi, M. a K., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45,** 325–9 (2013).

117. Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H. & Bourque, G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42,** 631–4 (2010).

118. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9,** e1003504 (2013).

119. Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., Chen, W., Lorincz, M. C., Ivics, Z., Hurst, L. D. & Izsvák, Z. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516,** 405–409 (2014).

120. Bi, S., Gavrilova, O., Gong, D. W., Mason, M. M. & Reitman, M. Identification of a placental enhancer for the human leptin gene. *J. Biol. Chem.* **272,** 30583–8 (1997).

121. Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T. & Wellstein, A. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Med. Sci.* **93,** 14759–14764 (1996).

122. Cohen, C. J., Rebollo, R., Babovic, S., Dai, E. L., Robinson, W. P. & Mager, D. L. Placenta-specific expression of the interleukin-2 (IL-2) receptor β subunit from an endogenous retroviral promoter. *J. Biol. Chem.* **286,** 35543–52 (2011).

123. Mey, A., Acloque, H., Lerat, E., Gounel, S., Tribollet, V., Blanc, S., Curton, D., Birot, A.-M., Nieto, M. A. & Samarut, J. The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue. *Retrovirology* **9,** 21 (2012).

124. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (80-.).* **351,** 1083–1087 (2016).

125. Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S. M., Brachmann, R. K. & Haussler, D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18613–8 (2007).

126. Evsikov, A. V & Marín de Evsikova, C. Friend or Foe: Epigenetic Regulation of Retrotransposons in Mammalian Oogenesis and Early Development. *Yale J. Biol. Med.* **89,** 487–497 (2016).

127. Banvillet, D. & Boie, Y. Retroviral Long Terminal Repeat is the Promoter of the Gene Encoding the Tumor-associated Calcium-binding Protein Oncomodulin in , the Rat. *J. Mol. Biol* **207,** (1989).

128. Romanish, M. T. et al. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. *PLoS Genet.* **3,** e10 (2007).

129. Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D. & Pfaff, S. L. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).

130. Dunn, C. A., Medstrand, P. & Mager, D. L. An endogenous retroviral long terminal repeat

LTR enhancer complex NF-Y/MZF1/GATA-2. *J. Biol. Chem.* **280,** 35184–94 (2005).

115. Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J. & Tuan, D. Long-range function of an intergenic retrotransposon. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 12992–7 (2010).

116. Chuong, E. B., Rumi, M. a K., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45,** 325–9 (2013).

117. Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H. & Bourque, G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42,** 631–4 (2010).

118. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9,** e1003504 (2013).

119. Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., Chen, W., Lorincz, M. C., Ivics, Z., Hurst, L. D. & Izsvák, Z. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516,** 405–409 (2014).

120. Bi, S., Gavrilova, O., Gong, D. W., Mason, M. M. & Reitman, M. Identification of a placental enhancer for the human leptin gene. *J. Biol. Chem.* **272,** 30583–8 (1997).

121. Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T. & Wellstein, A. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Med. Sci.* **93,** 14759–14764 (1996).

122. Cohen, C. J., Rebollo, R., Babovic, S., Dai, E. L., Robinson, W. P. & Mager, D. L. Placenta-specific expression of the interleukin-2 (IL-2) receptor β subunit from an endogenous retroviral promoter. *J. Biol. Chem.* **286,** 35543–52 (2011).

123. Mey, A., Acloque, H., Lerat, E., Gounel, S., Tribollet, V., Blanc, S., Curton, D., Birot, A.-M., Nieto, M. A. & Samarut, J. The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extra embryonic tissue. *Retrovirology* **9,** 21 (2012).

124. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (80-.).* **351,** 1083–1087 (2016).

125. Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S. M., Brachmann, R. K. & Haussler, D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 18613–8 (2007).

126. Evsikov, A. V & Marín de Evsikova, C. Friend or Foe: Epigenetic Regulation of Retrotransposons in Mammalian Oogenesis and Early Development. *Yale J. Biol. Med.* **89,** 487–497 (2016).

127. Banvillet, D. & Boie, Y. Retroviral Long Terminal Repeat is the Promoter of the Gene Encoding the Tumor-associated Calcium-binding Protein Oncomodulin in , the Rat. *J. Mol. Biol* **207,** (1989).

128. Romanish, M. T. et al. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. *PLoS Genet.* **3,** e10 (2007).

129. Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D. & Pfaff, S. L. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).

130. Dunn, C. A., Medstrand, P. & Mager, D. L. An endogenous retroviral long terminal repeat

is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12841–6 (2003).

131. Dunn, C. A. & Mager, D. L. Transcription of the human and rodent SPAM1 / PH-20 genes initiates within an ancient endogenous retrovirus. *BMC Genomics* **6,** 47 (2005).

132. van de Lagemaat, L. N., Landry, J.-R., Mager, D. L. & Medstrand, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. at <https://ac.els-cdn.com/S0168952503002312/1-s2.0-S0168952503002312-main.pdf?_tid=71ec4966-d61e-11e7-81db-00000aab0f6c&acdnat=1512081371_d721796a4fe2af55a444d39c17053e57>

133. Conley, A. B., Piriyapongsa, J. & Jordan, I. K. Retroviral promoters in the human genome. *Bioinformatics* **24,** 1563–7 (2008).

134. Conley, A. B., Miller, W. J. & Jordan, I. K. Human cis natural antisense transcripts initiated by transposable elements. doi:10.1016/j.tig.2007.11.005

135. Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M. & Feschotte, C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9,** e1003470 (2013).

136. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13,** R107 (2012).

137. St Laurent, G., Shtokalo, D., Dong, B., Tackett, M. R., Fan, X., Lazorthes, S., Nicolas, E., Sang, N., Triche, T. J., McCaffrey, T. A., Xiao, W. & Kapranov, P. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol.* **14,** R73 (2013).

138. Santoni, F. A., Guerra, J. & Luban, J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9,** 111 (2012).

139. Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., Chen, W., Lorincz, M. C., Ivics, Z., Hurst, L. D. & Izsvák, Z. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516,** 405–409 (2014).

140. Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G. & Ng, H.-H. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21,** 423–425 (2014).

141. Jern, P., Sperber, G. O. & Blomberg, J. Definition and variation of human endogenous retrovirus H. (2004). doi:10.1016/j.virol.2004.06.023

142. Mager, D. L. & Freeman, J. D. Human Endogenous Retroviruslike Genome with Type C pol Sequences and gag Sequences Related to Human T-Cell Lymphotropic Viruses. **61,**

143. Goodchild, N. L., Wilkinson, D. A. & Mager, D. L. Recent Evolutionary Expansion of a Subfamily of RTVL-H Human Endogenous Retrovirus-like Elements. *Virology* **196,** 778–788 (1993).

144. Mager, D. L. & Freeman, J. D. HERV-H Endogenous Retroviruses: Presence in the New World Branch but Amplification in the Old World Primate Lineage. *Virology* **213,** 395–404 (1995).

145. Magiorkinis, G., Blanco-Melo, D. & Belshaw, R. The decline of human endogenous retroviruses: extinction and survival. *Retrovirology* **12,** 8 (2015).

146. Huang, Y., Osorno, R., Tsakiridis, A. & Wilson, V. In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. (2012).

doi:10.1016/j.celrep.2012.10.022

147. Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. & Smith, A. The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–23 (2008).

148. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4,** 487–92 (2009).

149. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **advance on,** (2013).

150. Theunissen, T. W. et al. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19,** 502–515 (2016).

151. Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Díaz, N., Rowe, H. M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., Mangeat, B., Wildhaber, B. E., Naldini, L. & Trono, D. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* **24,** 1251–9 (2014).

152. Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura, M., Tokunaga, Y., Nakamura, M., Watanabe, A., Yamanaka, S. & Takahashi, K. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 12426–31 (2014).

153. Benit, L., Lallemand, J.-B., Casella, J.-F., Philippe, H. & Heidmann, T. ERV-L Elements: a Family of Endogenous Retrovirus-Like Elements Active throughout the Evolution of Mammals. *J. Virol.* **73,** 3301–3308 (1999).

154. Kigami, D. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biol. Reprod.* **68,** 651–654 (2002).

155. Morgani, S. M., Canham, M. A., Nichols, J., Sharov, A. A., Migueles, R. P., Ko, M. S. H. & Brickman, J. M. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3,** 1945–57 (2013).

156. Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J. M. & Torres-Padilla, M.-E. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22,** 662–671 (2015).

157. Abad, M., Mosteiro, L., Pantoja, C., Cañamero, M., Rayon, T., Ors, I., Graña, O., Megías, D., Domínguez, O., Martínez, D., Manzanares, M., Ortega, S. & Serrano, M. Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature* **502,** 340–345 (2013).

158. Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S. E., Franco, L., Rosenfeld, M. G., Ren, B. & Pfaff, S. L. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25,** 594–607 (2011).

159. Hendrickson, P. G., Doráis, J. A., Grow, E. J., Whiddon, J. L., Lim, J.-W., Wike, C. L., Weaver, B. D., Pflueger, C., Emery, B. R., Wilcox, A. L., Nix, D. A., Matthew Peterson, C., Tapscott, S. J., Carrell, D. T. & Cairns, B. R. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Publ. Gr.* (2017). doi:10.1038/ng.3844

160. Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J. M. & Torres-Padilla, M.-E. Early embryonic-like cells are induced by downregulating replication-dependent chromatin

assembly. *Nat. Struct. Mol. Biol.* **22,** 662–671 (2015).

161. Huang, Y. et al. Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *Elife* **6,** 1131–1139 (2017).

162. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J. & Trono, D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49,** 941–945 (2017).

163. Zalzman, M., Falco, G., Sharova, L. V, Nishiyama, A., Thomas, M., Lee, S.-L., Stagg, C. A., Hoang, H. G., Yang, H.-T., Indig, F. E., Wersto, R. P. & Ko, M. S. H. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464,** 858–63 (2010).

164. Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E. J., Davila, J., Mall, M., Wong, W. H., Wysocka, J., Au, K. F. & Reijo Pera, R. A. The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat. Genet.* **48,** 44–52 (2015).

165. Wang, J. et al. A novel long intergenic noncoding RNA indispensable for the cleavage of mouse two-cell embryos. *EMBO Rep.* **17,** 1452–1470 (2016).

166. Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J. & He, L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J. Biol. Chem.* **291,** 14457–67 (2016).

167. Qin, W., Dion, S. L., Kutny, P. M., Zhang, Y., Cheng, A. W., Jillette, N. L., Malhotra, A., Geurts, A. M., Chen, Y.-G. & Wang, H. Efficient CRISPR/Cas9-Mediated Genome Editing in Mice by Zygote Electroporation of Nuclease. *Genetics* **200,** 423–430 (2015).

168. Hashimoto, M. & Takemoto, T. Electroporation enables the efficient mRNA delivery into the mouse zygotes and facilitates CRISPR/Cas9-based genome editing. *Sci. Rep.* **5,** 11315 (2015).

169. Willis, B. J., Lloyd, K. C. K. & Wood, J. A. High throughput electroporation of zygotes increases efficiency in the generation of genetically edited rodent models using CRISPR-Cas9. in *Transgenic Animal Research Conference XI* (2017).

170. Miska, E. A., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R. & Horvitz, H. R. Most Caenorhabditis elegans microRNAs Are Individually Not Essential for Development or Viability. *PLoS Genet.* **3,** e215 (2007).

171. Alvarez-Saavedra, E. & Horvitz, H. R. Many families of C. elegans microRNAs are not essential for development or viability. *Curr. Biol.* **20,** 367–73 (2010).

172. Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J. S., Dixon, G., Harland, R., Wan, Y., Lishko, P., Lize, M., Kessel, M. & He, L. miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature* **510,** 115–120 (2014).

173. Wu, J., Bao, J., Kim, M., Yuan, S., Tang, C., Zheng, H., Mastick, G. S., Xu, C. & Yan, W. Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E2851-7 (2014).

174. Niu, D. et al. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* **357,** 1303–1307 (2017).

175. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).

176. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in

medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 7634–8 (1981).

177. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. & Jones, J. M. Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–7 (1998).

178. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–76 (2006).

179. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394,** 369–74 (1998).

180. Tachibana, M. et al. Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153,** 1228–38 (2013).

181. Kelly, S. J. Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J. Exp. Zool.* **200,** 365–76 (1977).

182. Ziomek, C. A., Johnson, M. H. & Handyside, A. H. The developmental potential of mouse 16-cell blastomeres. *J. Exp. Zool.* **221,** 345–55 (1982).

183. Suwińska, A., Czołowska, R., Ozdzeński, W. & Tarkowski, A. K. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. *Dev. Biol.* **322,** 133–44 (2008).

184. Choi, Y. J., Lin, C.-P., Ho, J. J., He, X., Okada, N., Bu, P., Zhong, Y., Kim, S. Y., Bennett, M. J., Chen, C., Ozturk, A., Hicks, G. G., Hannon, G. J. & He, L. miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nat. Cell Biol.* **13,** 1353–60 (2011).

185. Liu, W.-M., Pang, R. T. K., Chiu, P. C. N., Wong, B. P. C., Lao, K., Lee, K.-F. & Yeung, W. S. B. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 490–4 (2012).

186. Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J., Horvath, S. & Fan, G. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500,** 593–7 (2013).

187. Miranda, K. C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A. M., Lim, B. & Rigoutsos, I. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126,** 1203–17 (2006).

188. Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J. & Blelloch, R. MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr. Biol.* **20,** 271–7 (2010).

189. Yagi, R., Kohn, M. J., Karavanova, I., Kaneko, K. J., Vullhorst, D., DePamphilis, M. L. & Buonanno, A. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134,** 3827–36 (2007).

190. Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K. & Sasaki, H. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125,** 270–83 (2008).

191. Home, P., Saha, B., Ray, S., Dutta, D., Gunewardena, S., Yoo, B., Pal, A., Vivian, J. L., Larson, M., Petroff, M., Gallagher, P. G., Schulz, V. P., White, K. L., Golos, T. G., Behr, B. & Paul, S. Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 7362–7 (2012).

192. Schaefer, B. C., Schaefer, M. L., Kappler, J. W., Marrack, P. & Kedl, R. M. Observation of Antigen-Dependent CD8+ T-Cell/ Dendritic Cell Interactions in Vivo. *Cell. Immunol.* **214,** 110–122 (2001).
193. TARKOWSKI, A. K. Experiments on the Development of Isolated Blastomeres of Mouse Eggs. *Nature* **184,** 1286–1287 (1959).
194. Rossant, J. Postimplantation development of blastomeres isolated from 4-and 8-cell mouse eggs. *Embryol. exp. Morph* **36,** 283–290 (1976).
195. Dewitt, M. A., Corn, J. E. & Carroll, D. Genome editing via delivery of Cas9 ribonucleoprotein. (2017). doi:10.1016/j.ymeth.2017.04.003
196. Schaefer, B. C., Schaefer, M. L., Kappler, J. W., Marrack, P. & Kedl, R. M. Observation of Antigen-Dependent CD8+ T-Cell/ Dendritic Cell Interactions in Vivo. *Cell. Immunol.* **214,** 110–122 (2001).
197. Miranda, K. C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A. M., Lim, B. & Rigoutsos, I. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126,** 1203–17 (2006).
198. 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,** 348–52 (1988).
199. Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet.* **6,** 507–12 (2005).
200. Geurts, A. M. et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325,** 433 (2009).
201. Carbery, I. D., Ji, D., Harrington, A., Brown, V., Weinstein, E. J., Liaw, L. & Cui, X. Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186,** 451–9 (2010).
202. Tesson, L., Usal, C., Ménoret, S., Leung, E., Niles, B. J., Remy, S., Santiago, Y., Vincent, A. I., Meng, X., Zhang, L., Gregory, P. D., Anegon, I. & Cost, G. J. Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.* **29,** 695–6 (2011).
203. Sung, Y. H., Baek, I.-J., Kim, D. H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.-S. & Lee, H.-W. Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* **31,** 23–4 (2013).
204. Meyer, M., de Angelis, M. H., Wurst, W. & Kühn, R. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15022–6 (2010).
205. Cui, X., Ji, D., Fisher, D. A., Wu, Y., Briner, D. M. & Weinstein, E. J. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat. Biotechnol.* **29,** 64–7 (2011).
206. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188,** 773–82 (2011).
207. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–12 (2007).
208. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V & van der Oost, J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–4 (2008).
209. Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., Yang, Z., Zu, Y., Li, W., Huang, P., Tong, X., Zhu, Z., Lin, S. & Zhang, B. Chromosomal deletions and inversions mediated by

TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* **41,** e141 (2013).

210. Niu, Y. et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156,** 836–43 (2014).

211. Guo, X., Zhang, T., Hu, Z., Zhang, Y., Shi, Z., Wang, Q., Cui, Y., Wang, F., Zhao, H. & Chen, Y. Efficient RNA/Cas9-mediated genome editing in Xenopus tropicalis. *Development* **141,** 707–14 (2014).

212. Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B. B. L. G., Tijsterman, M., van den Heuvel, S. & Boxem, M. CRISPR/Cas9-targeted mutagenesis in Caenorhabditis elegans. *Genetics* **195,** 1187–91 (2013).

213. Gokcezade, J., Sienski, G. & Duchek, P. Efficient CRISPR/Cas9 plasmids for rapid and versatile genome editing in Drosophila. *G3 (Bethesda).* **4,** 2279–82 (2014).

214. 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,** 816–21 (2012).

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

216. Maddalo, D., Manchado, E., Concepcion, C. P., Bonetti, C., Vidigal, J. A., Han, Y.-C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., Lowe, S. W. & Ventura, A. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516,** 423–427 (2014).

217. Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., Maeda, T., Paw, B. H. & Orkin, S. H. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289,** 21312–24 (2014).

218. Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L. & Jaenisch, R. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154,** 1370–9 (2013).

219. 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,** 819–23 (2013).

220. Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. & Jaenisch, R. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell* **153,** 910–918 (2013).

221. Irion, U., Krauss, J. & Nüsslein-Volhard, C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* **141,** 4827–30 (2014).

222. Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. & O'Connor-Giles, K. M. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194,** 1029–35 (2013).

223. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Rep.* **6,** 1178–1179 (2014).

224. Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. & Jaenisch, R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–8 (2013).

225. Qin, W., Dion, S. L., Kutny, P. M., Zhang, Y., Cheng, A., Jillette, N. L., Malhotra, A., Geurts, A. M., Chen, Y.-G. & Wang, H. Efficient CRISPR/Cas9-Mediated Genome

Editing in Mice by Zygote Electroporation of Nuclease. *Genetics* (2015). doi:10.1534/genetics.115.176594

226. Takahashi, G., Gurumurthy, C. B., Wada, K., Miura, H., Sato, M. & Ohtsuka, M. GONAD: Genome-editing via Oviductal Nucleic Acids Delivery system: a novel microinjection independent genome engineering method in mice. *Sci. Rep.* **5,** 11406 (2015).

227. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E2579-86 (2012).

228. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F. & Nureki, O. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–49 (2014).

229. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E. & Doudna, J. A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343,** 1247997 (2014).

230. Cho, S. W., Lee, J., Carroll, D., Kim, J.-S. & Lee, J. Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* **195,** 1177–80 (2013).

231. Lee, J.-S., Kwak, S.-J., Kim, J., Kim, A.-K., Noh, H. M., Kim, J.-S. & Yu, K. RNA-guided genome editing in Drosophila with the purified Cas9 protein. *G3 (Bethesda).* **4,** 1291–5 (2014).

232. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–9 (2014).

233. Lin, S., Staahl, B., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3,** e04766 (2014).

234. Schumann, K., Lin, S., Boyer, E., Simeonov, D. R., Subramaniam, M., Gate, R. E., Haliburton, G. E., Ye, C. J., Bluestone, J. A., Doudna, J. A. & Marson, A. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl. Acad. Sci.* **112,** 201512503 (2015).

235. Mizuno, S., Dinh, T. T. H., Kato, K., Mizuno-Iijima, S., Tanimoto, Y., Daitoku, Y., Hoshino, Y., Ikawa, M., Takahashi, S., Sugiyama, F. & Yagami, K. Simple generation of albino C57BL/6J mice with G291T mutation in the tyrosinase gene by the CRISPR/Cas9 system. *Mamm. Genome* **25,** 327–34 (2014).

236. Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S. & Huang, B. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–91 (2013).

237. Basrur, V., Yang, F., Kushimoto, T., Higashimoto, Y., Yasumoto, K., Valencia, J., Muller, J., Vieira, W. D., Watabe, H., Shabanowitz, J., Hearing, V. J., Hunt, D. F. & Appella, E. Proteomic Analysis of Early Melanosomes: Identification of Novel Melanosomal Proteins. *J. Proteome Res.* **2,** 69–79 (2003).

238. Paterson, E. K., Fielder, T. J., MacGregor, G. R., Ito, S., Wakamatsu, K., Gillen, D. L., Eby, V., Boissy, R. E. & Ganesan, A. K. Tyrosinase Depletion Prevents the Maturation of

Melanosomes in the Mouse Hair Follicle. *PLoS One* **10,** e0143702 (2015).

239. Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J. S., Brown, M. & Liu, S. X. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* gr.191452.115 (2015). at <http://genome.cshlp.org/content/early/2015/06/10/gr.191452.115>

240. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J. & Root, D. E. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–191 (2016).

241. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42,** W401-7 (2014).

242. 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,** 827–32 (2013).

243. Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J. & Gerard, G. F. Mutation detection using Surveyor nuclease. *Biotechniques* **36,** 702–7 (2004).

244. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–46 (2012).