Journal of Mammary Gland Biology and Neoplasia (2019) 24:47–59
https://doi.org/10.1007/s10911-018-9417-z

Dissecting Tissue-Specific Super-Enhancers by Integrating Genome-Wide Analyses and CRISPR/Cas9 Genome Editing

Kyung Hyun Yoo[1] · Lothar Hennighausen[2,3] · Ha Youn Shin[4]

Received: 29 June 2018 / Accepted: 1 October 2018 / Published online: 6 October 2018

Abstract

Recent advances in genome-wide sequencing technologies have provided researchers with unprecedented opportunities to discover the genomic structures of gene regulatory units in living organisms. In particular, the integration of ChIP-seq, RNA-seq, and DNase-seq techniques has facilitated the mapping of a new class of regulatory elements. These elements, called super-enhancers, can regulate cell-type-specific gene sets and even fine-tune gene expression regulation in response to external stimuli, and have become a hot topic in genome biology. However, there is scant genetic evidence demonstrating their unique biological relevance and the mechanisms underlying these biological functions. In this review, we describe a robust genome-wide strategy for mapping cell-type-specific enhancers or super-enhancers in the mammary genome. In this strategy, genome-wide screening of active enhancer clusters that are co-occupied by mammary-enriched transcription factors, co-factors, and active enhancer marks is used to identify bona fide mammary tissue-specific super-enhancers. The in vivo function of these super-enhancers and their associated regulatory elements may then be investigated in various ways using the advanced CRISPR/Cas9 genome-editing technology. Based on our experience targeting various mammary genomic sites using CRISPR/Cas9 in mice, we comprehensively discuss the molecular consequences of the different targeting methods, such as the number of gRNAs and the dependence on their simultaneous or sequential injections. We also mention the considerations that are essential for obtaining accurate results and shed light on recent progress that has been made in developing modified CRISPR/Cas9 genome-editing techniques. In the future, the coupling of advanced genome-wide sequencing and genome-editing technologies could provide new insights into the complex genetic regulatory networks involved in mammary-gland development.

Keywords Super-enhancer · Genome-wide analysis · CRISPR/Cas9 · Cell type-specific gene regulation · Mammary gland development

Abbreviations

SE Super-enhancer

Introduction

During animal development, different gene sets are precisely activated in distinct organs at specific stages. This tightly regulated developmental process is mainly controlled by regulatory units in genomic DNA. Although extensive studies have examined the roles of these regulatory elements (e.g., promoters and enhancers), recent progress in next-generation-sequencing technologies has given researchers new opportunities to map the genome-wide chromatin landscape of regulatory units [1]. In such strategies, DNase-seq is commonly used to predict the chromatin accessibility of the genome [2]. Active regulatory elements are most often located in nucleosome-depleted regions, and are thus predominantly found in DNase I hypersensitive sites (DHS) that are easily cleaved by DNase I. An open (i.e., nucleosome-depleted)

✉ Ha Youn Shin
hayounshin@konkuk.ac.kr

[1] Department of Biological Sciences, Sookmyung Women's University, Seoul 04310, Republic of Korea

[2] Laboratory of Genetics and Physiology, National Institutes of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[3] BK21 Biological Science Visiting Professor, Sookmyung Women's University, Seoul 04310, Republic of Korea

[4] Department of Biomedical Science and Engineering, Konkuk University, Seoul 05029, Republic of Korea

chromatin region serves as a platform for the transcriptional machinery responsible for gene activation, which includes transcription factors, co-factors, and RNA polymerase II. Genome-wide transcription factor binding patterns and histone modifications can be detected by chromatin immunoprecipitation followed by sequencing (ChIP-seq) [3, 4]. ChIP-seq profiling has revealed that active promoters are post-translationally modified by the tri-methylation of lysine 4 at histone 3 (H3K4me3), whereas active enhancers are enriched for the acetylation of H3K27. In addition to histone modifications, the co-occupancy of lineage-specific transcription factors and co-factors is often found in cell-type-specific enhancers [5, 6]. The extent of the gene activation controlled by the nearest regulatory elements can be further examined by RNA sequencing (RNA-seq) [7, 8]. When integrated together, these next-generation-sequencing technologies can enable researchers to predict putative regulatory elements throughout the genome and facilitate the characterization of their unique chromatin structures across a wide range of living organisms.

Although next-generation-sequencing techniques are useful for examining the genomic structures of regulatory elements, they do not reflect biological functions. The recent emergence of nuclease-mediated genome-editing technologies opened a new era in functional mutagenesis studies and has enabled researchers to determine the in vivo biological functions of putative regulatory elements. The first endonuclease utilized for genome engineering was the zinc-finger nuclease (ZFN) [9, 10], which was introduced for this purpose in 2005. The zinc finger DNA-binding domains of ZFN recognize specific target sequences, whereupon the associated endonuclease, FoKI, cleaves target DNA sequences. Subsequently, a synthetic nuclease, called transcription activator-like effector nuclease (TAELN), was developed to provide an alternative platform for genome engineering. Similar to ZFN, TALEN consists of a TALE effector DNA-binding domain and a DNA-cleavage domain [11–14]. Although both ZFN and TALEN have the capacity to alter specific DNA sequences, it remains technically challenging and time-consuming to engineer DNA-binding modules. More recently, researchers developed a clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system that has been widely used for genome engineering in various organisms. CRISPR/Cas9 was originally adapted from the bacterial immune system [15–17]. Prokaryotes use CRISPR/Cas9 to degrade invading nucleic acids (e.g., viral and plasmid DNAs) for adaptive immunity. In 2012 and 2013, three different research groups simultaneously succeeded in applying a CRISPR/Cas9-based system for genome engineering in eukaryotic cells [18–23]. CRISPR/Cas9-based genome engineering is simpler, faster, and more cost effective than the previous strategies and has been extensively used to establish mouse models.

In this review, we first describe a strategy that may be used to obtain a reliable set of cell-type-specific enhancers or super-enhancers, and how their biological functions may be examined using the CRISPR/Cas9 system. We then comprehensively illustrate the preferential deletion patterns that may be obtained from different CRISPR/Cas9 targeting methods and address recent updates in genome-engineering technologies. We further provide practical guidance and new insights on the application of genome-wide analyses and genome-editing technologies to study mammary gland biology.

Identification of Super-Enhancers Using Genome-Wide Analyses

Although enhancers have been widely studied for more than several decades, their importance was recently re-emphasized by the discovery that they play a unique role in cell-type-specific gene control. Richard Young and his colleagues were the first to identify enhancers essential for lineage-specific gene regulation and designate them as “super-enhancers” [24, 25]. Unlike typical enhancers, super-enhancers are comprised of multiple enhancers and cover long regions of genomic DNA, generally over 12.5 kb. The constituent enhancers of a super-enhancer are densely occupied by lineage-specific or master transcription factors, the mediator complex subunit MED1 [26], and the active enhancer mark H3K27ac. Importantly, super-enhancers are highly associated with genes involved in cell identity. Further studies demonstrated that super-enhancers have unique chromatin features in various cell types of mouse and human tissues [27–29]. Super-enhancers are commonly identified by the Rank Ordering of Super-Enhancers (ROSE) algorithm, using the ChIP-seq profiles of lineage-specific transcription factors, MED1, or H3K27ac [24]. This program identifies stitched enhancer regions and ranks them by their ChIP-seq signal intensity. However, most studies have applied only one or two ChIP-seq profiles to the ROSE algorithm. For example, MED1 ChIP-seq profiles were used to identify super-enhancers in embryonic stem cells, B cells [24] and adipocytes [28], and H3K27ac ChIP-seq signals were used to identify super-enhancers in hair follicle stem cells [29] and erythrocytes [30]. Therefore, there is some question as to whether this is sufficient to obtain genuine cell-type-specific enhancers.

Hennighausen and co-workers used a more robust strategy to identify super-enhancers in mammary tissue [31]. ChIP-seq experiments were performed initially to examine the genome-wide binding patterns of the transcription factor, STAT5, which is a master regulator of mammary gland development (Fig. 1). These experiments identified regions in which STAT5 binding and H3K27ac coincided, followed by the use of non-promoter regions to identify super-enhancers. Additional ChIP-seq experiments were performed with mammary-enriched transcription factors, such as the glucocorticoid receptor (GR) [32] and MED1, on day 1 of lactation (L1), when mammary-specific genes are highly expressed. GR has been found to synergize with STAT5 to induce

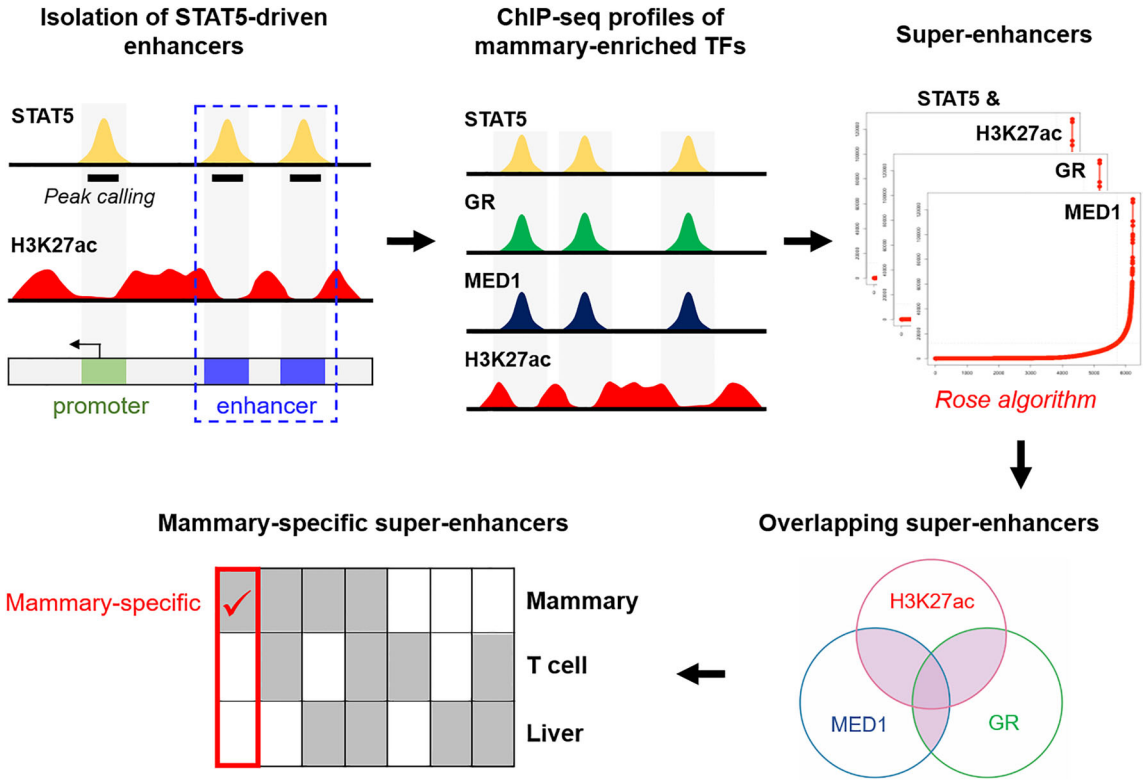

Fig. 1 Identification of mammary-specific super-enhancers. ChIP-seq experiments conducted with the master transcription factor, STAT5, in mammary tissue identified genomic regions coinciding with the active enhancer mark, H3K27ac. Non-promoter regions were excluded, and the individual ChIP-seq profiles of STAT5, the mammary-enriched transcription factor, GR, the mediator, MED1, and H3K27ac were subjected to the ROSE algorithm for identification of putative super-enhancers. Overlapping super-enhancers were isolated, and those common to mammary tissue, liver or T cells were further removed to obtain a high-confidence set of mammary-specific super-enhancers

expression of genes encoding milk proteins. Subsequently, the individual ChIP-seq profiles were subjected to the ROSE algorithm, and overlapping super-enhancers were isolated. After removing the STAT5-driven super-enhancers found in both mammary and non-mammary tissues (e.g., liver and T cells), 440 mammary-specific super-enhancers were obtained. DNase-seq data confirmed that these super-enhancer regions were located in open chromatin regions, and additional ChIP-seq profiles revealed the co-localization of mammary-enriched transcription factors, such as nuclear factor 1-B type (NFIB) [33] and E74-like factor 5 (ELF5) [34], at constituent enhancers. NFIB functions as a co-activator of STAT5 in mammary gland development, whereas ELF5 is essential for mammary gland development and differentiation.

The high-confidence mammary-specific super-enhancers were further annotated to the nearest genes, and their expression levels were examined by RNA-seq in mammary tissue. RNA-seq data revealed that the genes associated with STAT5-driven super-enhancers were more highly expressed than genes associated with STAT5-driven lone enhancers. Moreover, genes associated with mammary-specific super-enhancers were highly expressed in mammary tissue, but were not expressed or expressed only at basal levels in non-mammary tissues, including liver and T cells. Further gene-set-enrichment analysis (GSEA) showed that the levels of expression of super-enhancer-associated genes were highly dependent on the expression level of STAT5 and that more than half of these genes were significantly induced during lactation. For example, the mammary-specific super-enhancer, *Wap*, comprises three individual enhancers that are densely occupied by H3K27ac, MED1 and mammary enriched transcription factors such as STAT5, GR, NFIB, and ELF5 [31]. More complex mammary-specific gene loci comprised of four *Csn* genes and also contain four super-enhancers, two between *Csn2* and *Csn1s2a*, one upstream of *Odam*, and one in front of *Csn3*, and that these super-enhancers are densely occupied by mammary-enriched transcription factors and H3K27ac marks [35]. However, STAT5-driven lone enhancers did not bind to mammary-enriched transcription factors. Additional ChIP-seq experiments revealed that both *Wap* and *Csn* super-enhancers are demarcated by the zinc finger protein, CTCF, which is involved in insulating neighboring genes [35, 36]. Chromosome conformation capture-on-ChIP (4C) experiments also showed interactions among CTCF sites within *Csn* super-enhancers. Subsequently, RNA-seq data demonstrated that both *Wap* and *Csn*, which are major milk protein genes in mice, are highly induced up to 10,000-fold during lactation [31], whereas genes-associated with STAT5-driven lone enhancers were only

expressed at basal levels in mammary tissue. These findings clearly demonstrate that the sequential and comprehensive analysis of results generated by complementary next-generation-sequencing tools can allow researchers to identify bona fide super-enhancers specific to mammary tissue.

Recently, super-enhancers have also been identified in human breast cancers. In addition to the association between super-enhancers and cell identity genes, aberrant forms of super-enhancers have been frequently found in various diseases including leukemia [37, 38], solid tumors [25, 39], diabetes [25, 40], auto-immune diseases [25, 27] and neurodegenerative diseases such as Alzheimer's disease [25] and Parkinson's disease [41, 42]. Several features of disease-associated super-enhancers have been characterized. Single nucleotide polymorphisms (SNPs), small insertions, DNA translocation, focal amplification and overexpression of transcription factors such as bromodomain-containing protein-4 (BRD4) or oncogenes have often been found within super-enhancers [43]. BRD4 is a member of the BET family that binds to mediator complexes and participates in transcriptional elongation [44]. To identify super-enhancers in breast cancer, Zuber and colleagues used H3K27ac and BRD4 ChIP-seq profiles for the MCF-7 breast cancer cell line [45] and Glodzik and co-workers used H3K27ac ChIP-seq profiles for human breast cancer tissues [46]. Although these studies identified super-enhancers in breast tissue, they only used an active enhancer marker or a marker preferentially found in disease-associated super-enhancers (e.g., BRD4) that is not specific to mammary tissue.

In Vitro Approaches to Investigate the Functional Role of Super-Enhancers

Although cell-type-specific chromatin features of super-enhancers have been well characterized, their biological relevance has not been fully determined. Initial attempts to investigate the functional activities of super-enhancers involved the linking of super-enhancer fragments to the reporter gene and assessing enhancement by in vitro reporter assays [40, 47–49]. Whyte et al. induced the overexpression of super-enhancer fragments, which were linked to the luciferase gene in embryonic stem cells, finding that super-enhancers have higher activity than lone enhancers [24]. Adam et al. also showed that ectopic expression of the super-enhancer specific to hair follicle stem cells linked to the GFP reporter gene restricts enhancer activity in hair follicle stem cells, but not other cell types [29].

Additional studies examined the functional roles of super-enhancers by inhibiting key transcription factors within super-enhancers using siRNA [50, 51], shRNA [24, 28], miRNA [52] or small molecule inhibitors [47, 53–61]. shRNA inhibition of the key transcription factor Oct4 within super-enhancers specific to embryonic stem cells significantly reduced the expression of genes associated with super-enhancers, but not lone enhancers [24]. Treatment with BRD4 inhibitors such as JQ1 has also been used to test the activity of super-enhancers found in various cell types, specifically in diseases [47, 53–58, 60, 61]. JQ1 preferentially reduced expression levels of disease-associated genes by selectively preventing BRD4 binding at super-enhancers. Following the development of various genome engineering tools, many research groups have tried to delete entire super-enhancers or induce mutations in individual enhancers at transcription factor binding sites in various cell lines [30, 62–68]. However, the in vivo biological function of super-enhancers was not demonstrated in animals until 2016.

Using CRISPR/Cas9 Genome Editing to Investigate the In Vivo Functional Role of Cell-Type-Specific Regulatory Elements

The first demonstration of the in vivo biological function of super-enhancers was reported by Shin et al. in mammary tissue [31]. CRISPR/Cas9 genome editing was used to target STAT5 binding sites at three constituent *Wap* super-enhancers, E1, E2 and E3, both individually and in various combinations. Experiments with mice having individual enhancer mutations, ΔE1, ΔE2 or ΔE3, revealed that the most distal E3 enhancer plays a critical role in mammary-specific *Wap* gene expression during lactation and was largely dependent on STAT5. Experiments with mice having combined mutations, ΔE1/2, ΔE2/3, or ΔE1/2/3, showed that all three individual enhancers are necessary for full induction of mammary-specific *Wap* gene during lactation. These results indicate that the mammary-specific *Wap* super-enhancer, identified by complementary genome-wide analysis tools, activates a mammary-specific gene. Moreover, these findings demonstrate that CRISPR/Cas9 genome editing is a useful tool to examine in vivo biological functions of super-enhancers. In a study investigating the role of CTCF as an insulator, five CTCF binding sites within the *Wap* super-enhancer were also mutated by CRISPR/Cas9 [36]. Although no CTCF binding site mutation was found to affect *Wap* gene expression, the deletion of specific CTCF sites significantly induced the expression of the neighboring gene, *Ramp3*, and established de novo chromatin interactions between two distinct regulatory regions. These results indicate that CTCF may not completely block the super-enhancer unit; instead, it may silence neighboring gene activation.

Although CRISPR/Cas9 has been used successfully to mutate specific DNA sequences in genes or regulatory elements of eukaryotic cells [31, 35, 36, 69–71], only limited studies have described strategies to introduce mutations in the mouse germline, and the subsequent molecular consequences had not been systemically explored prior to a study in 2017 [72]. To investigate the global molecular consequences driven by CRISPR/Cas9 editing of the mouse genome, 17 genomic sites within mammary-specific enhancers were systematically

targeted, and individual mutations in 630 founder mice were analyzed. Specifically, the sequences of the STAT5- and CTCF-binding sites within mammary-specific super-enhancers were mutated, either individually or in combination, followed by a comprehensive investigation of the preferential deletion patterns and consequences of the different targeting strategies obtained by CRISPR/Cas9 genome editing.

The CRISPR/Cas9 system consists of two essential components: a single guide RNA (gRNA) that recognizes target DNA sequences and a Cas9 nuclease that originated from the bacterium, *Streptococcus pyogenes* [19]. When a gRNA binds to target sequences, Cas9 induces a double strand break (DSB), which is repaired by either non-homologous end joining (NHEJ) or homology-direct repair (HDR). NHEJ dominates when the CRISPR/Cas9 system is used, and it efficiently generates deletions at target sites. When the donor DNA template is introduced into cells along with a gRNA and Cas9, HDR alternatively occurs and the donor sequence is inserted into the target region. Thus, NHEJ is widely used to generate knockout mice, whereas HDR is commonly used to establish knockin mouse models. These unique features allow researchers to directly examine the in vivo function of putative regulatory elements in the mouse genome, especially in mammary organs that may not be represented by an in vitro system.

Preferential Deletion Patterns Associated with CRISPR/Cas9 Genome Editing

Cas9 recognizes a protospacer-adjacent motif (PAM), typically an NGG sequence adjacent to a gRNA within the target region, and cleaves double-stranded DNA between the third and fourth nucleotides from the PAM. To investigate whether there is any preferential orientation of the deletions generated by CRISPR/Cas9, mutational analyses were performed at nine different genomic sites in 139 founder mice [72]. Notably, the deletions were prevalently generated toward the 5′- or 3′-ends from the Cas9 cutting site (Fig. 2a); more than 70% of the deletions extended to 1.5-fold of either end, and over 80% exceeded a 2-fold difference. Prevalent asymmetric deletions were also found in data obtained from the study of Kim et al. [73]. Thus, although there does not appear to be any distinct preference for a deletion to be oriented toward the 5′- or 3′-end (Fig. 2b), it is clear that asymmetric deletions prevail at the Cas9 cutting site.

Deletions with repeat sequences are also frequently seen with the CRISPR/Cas9 system [72]. Analysis of seven different target regions in 122 founder mice showed that a single or duplicate unit of a repeat sequence was often preferentially retained at a deletion site (Fig. 2c). Although Kim et al. reported that microhomology-mediated deletion can produce a single unit of a repeat sequence at a deletion site [74], the frequency of this deletion pattern had not been systemically investigated in such a large cohort, and no previous report had described deletions in which repeat sequences align at both ends. We found that in specific mutant lines, more than 65% of mice carried deletions with a single copy of a repeat sequence, while 35% carried duplicate copies [72]. Although future studies are needed to unravel the mechanisms underlying the preferential asymmetric deletions and deletions with repeat sequences obtained using CRISPR/Cas9, recognition of these unique deletion patterns may help researchers predict prevalent deletion sites and/or design an efficient deletion strategy.

Molecular Consequences of Different CRISPR/Cas9 Targeting Methods

Zhou et al. reported that the co-injection of multiple gRNAs at a single site frequently generates large deletions [75]. We further assessed whether large deletions are generated by a single gRNA injection at a single genomic site [72]. Analysis of deletion sizes in 122 founder mice disclosed that the median deletion size generated by a single gRNA injection (9 bp) was obviously smaller than that gained with multiple gRNAs (84 bp). In a specific mutant line generated by injection of multiple gRNAs, more than 40% of mice carried deletions of over 200 bp in size. However, surprisingly, large deletions of up to 600 bp were also created with a single gRNA (Fig. 3a).

Several studies have reported that the simultaneous injection of more than one gRNA at multiple sites often induces large deletions that can reach mega bases in size [22, 76–78]. With a view to developing a more efficient targeting method for juxtaposed genomic sites, we compared the deletion patterns obtained by simultaneous or sequential injection of more than one gRNA at multiple genomic sites [72]. While the smaller deletions obtained by the two targeting methods were not significantly different (8 bp vs. 34 bp on average), large deletions of 400 bp to 24 kb were generated only by the simultaneous targeting method. Remarkably, more than half of the simultaneously targeted mutant lines carried these large deletions. These results suggest that different targeting methods should be chosen for certain purposes: the sequential targeting method is more appropriate for generating specific short deletions at individual sites, whereas simultaneous targeting is more efficient at creating large deletions that span adjacent sites.

Misleading Genotypes Caused by Large Deletions

We further demonstrated that large deletions generated using the CRISPR/Cas9 system often yield misleading results when genotyped [72]. In the typical genotyping method, target site fragments of approximately 400 bp are amplified by polymerase chain reaction (PCR), and their sequences are analyzed by Sanger sequencing (Fig. 3b, top panel). The PCR primers will

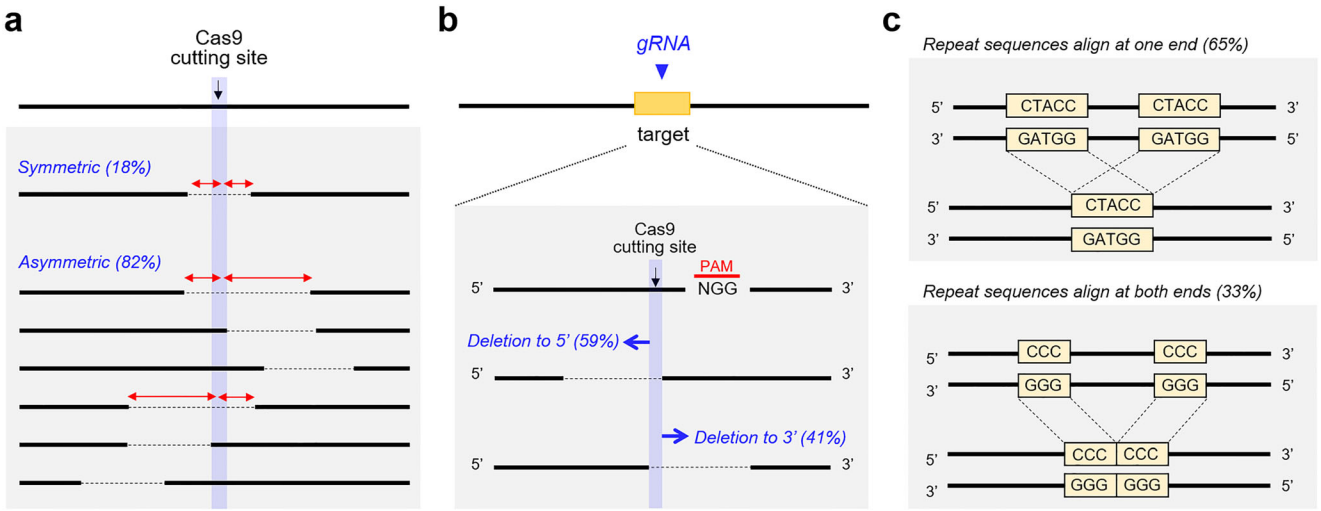

Fig. 2 Prevalent deletion patterns obtained by CRISPR/Cas9 genome editing. **a** Asymmetric deletions are preferentially introduced by CRISPR/Cas9. More than 80% of mutant mice carry deletions with one arm exceeding the other by 2-fold or more. **b** Orientation of asymmetric deletions. There is no distinct preference for the specific orientation of the asymmetric deletions. **c** Deletions preferentially occur at repeat sequences. Approximately 65% of the deletions retain a single copy of the repeat sequence, while 33% retain duplicate copies of the repeat sequence

fail to attach to the site of a larger (over 400 bp) deletion, which will not be visualized by the typical PCR-based genotyping method. Therefore, if a large deletion is created at one allele to yield a heterozygous mutant, the genotyping result would exhibit only the wild-type (WT) allele (Fig. 3b, middle panel). Similarly, if deletions of different sizes are generated at the two alleles, and one is large, the genotyping results may only reflect the smaller of the two. Finally, if large deletions exist at both alleles, the PCR may fail.

When biallelic deletions of different sizes are present at multiple sites, the true genotype is even harder to decode. Compound heterozygous mutations are frequently generated by the use of simultaneous targeting in founder mice [72]. If one allele contains a continuous or stitched large deletion, the conventional genotyping method may not be able to distinguish two alleles (Fig. 3b, bottom panel). If a large deletion contains a combination of continuous and stitched deletions, the true genotype may not be revealed unless whole-genome sequencing is conducted. If a large deletion is generated at the intervening region between two target sites, it may be overlooked if only the target sites are sequenced [72]. Moreover, mosaic mutations can occur in founder mice, further complicating efforts to distinguish individual genotypes. Evidence of complex large deletions generated by CRISPR/Cas9 was further supported by two recent studies [79, 80].

Thus, researchers should thoroughly analyze their founder genotyping results to avoid unexpected genotyping pitfalls and only trust genotyping results of F1 heterozygotes. When generating F1 mice, we suggest mating pure WT mice with founder mutants. In the absence of a determination that homozygous or heterozygous mutants have large hidden deletions prior to their breeding with each other, PCR genotyping often failed in some F2 mice due to large biallelic deletions. This can result in an inability to identify true genotypes, further delaying the establishment of homozygous mutant lines. We also suggest screening more than ten founder mice, maintaining at least three different mutant lines, and keeping track of their genotyping results until F2 homozygous mice are obtained. Along with these mating strategies, various genotyping methods can be used to further screen out unwanted large deletions. Once the genotype of founder mutants has been identified by PCR-based genotyping methods, the presence of large deletions within target regions can be assessed by quantitative PCR genotyping. If the mutant has a large deletion on one allele, only one-half levels of WT DNA will be amplified. Further PCR genotyping with serial primers spanning extended target sites can investigate whether there is any off-target deletion near or between target regions. If necessary, whole genome sequencing (WGS) should be performed to obtain accurate genotypes. Although WGS has been used to screen off-target mutations in single cell-derived clones and animals, it is very exhaustive and not sufficiently sensitive to detect indels in bulk heterogeneous populations of cells. To solve these issues, various methods have been developed that enable detection of regions of unrepaired DSBs or of DSBs repaired by erroneous NHEJ [81, 82]. The regions of unrepaired DSBs can be directly screened by in vitro nuclease-digested sequencing (Digenome-seq) [83] or in situ break labeling, enrichments on streptavidin and next-generation sequencing (BLESS) [23, 84]. However, because BLESS only labels DSBs present at the time of cell fixation, it shows limited sensitivity. Methods of detecting regions of erroneous NHEJ-mediated DSB repair include integrase-deficient lentivirus (IDLV)-capture [85, 86], high-throughput genomic translocation sequencing (HTGTS) [87], and

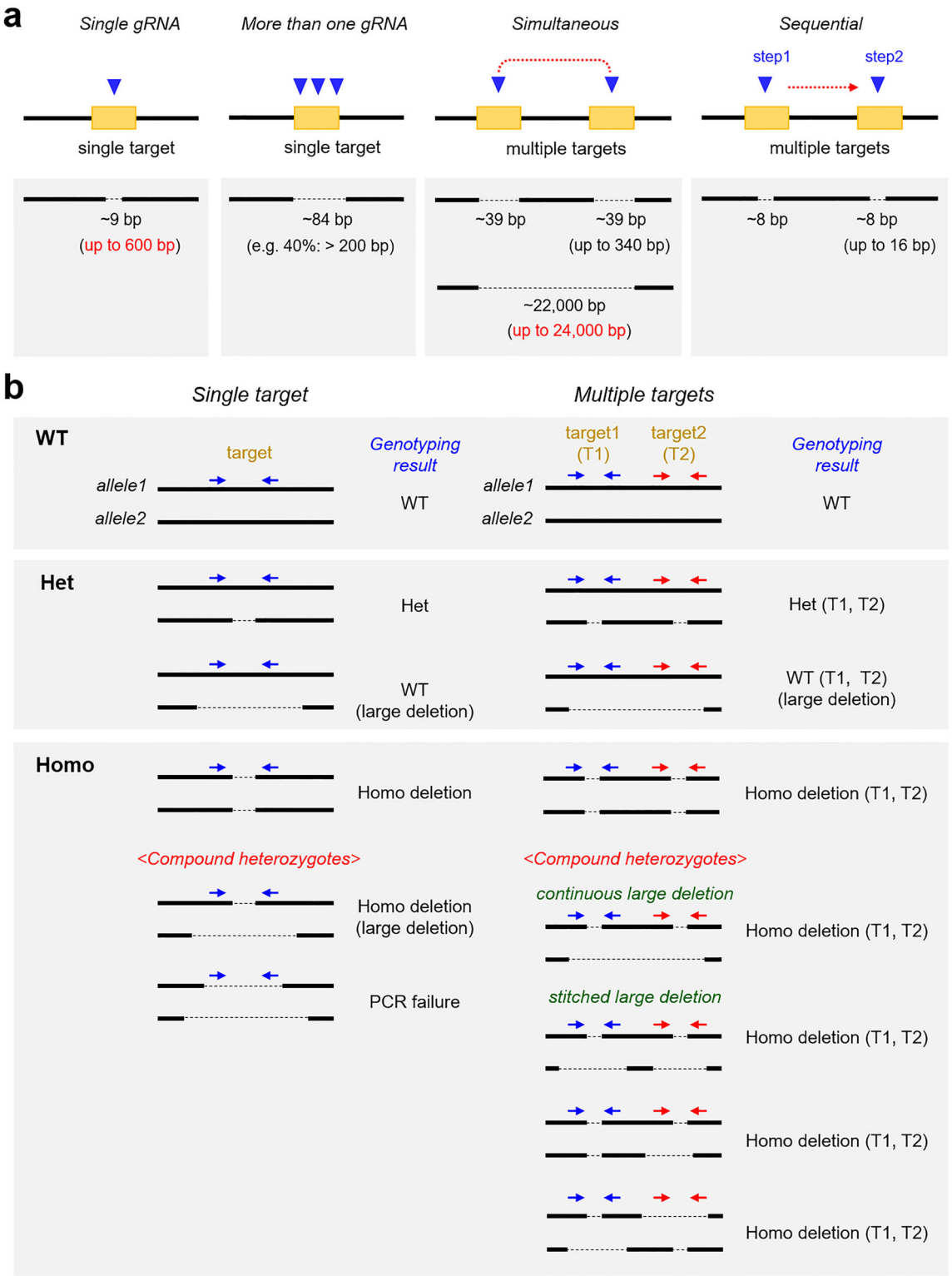

Fig. 3 Deletion sizes obtained from different gRNA injection strategies. **a** Comparison of deletion sizes obtained from different numbers of gRNAs and targeting methods. Although the median deletion size obtained with a single gRNA is smaller than that obtained with multiple gRNAs, large deletions (up to 600 bp) may be obtained with a single gRNA at a single site. Large deletions (up to 24 kb) are frequently generated when multiple sites are simultaneously targeted by more than one gRNA, but not when the sequential targeting method is used. **b** Various types of "hidden" large deletions (i.e., those that are not visualized by standard PCR-based genotyping methods) may be generated by CRISPR/Cas9. The panels illustrate the potential genotyping results of wild-type (WT, top), heterozygous mutants (middle), and homozygous mutants (bottom)

genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq) [88]. The IDLV-capture method, which utilizes IDLV DNA to tag DSBs followed by PCR amplification coupled with deep sequencing, shows poor sensitivity. GUIDE-seq, which utilizes phosphorothioate-protected double-stranded oligonucleotides to label DSB regions, shows higher sensitivity than IDLV-capture and its identification of off-target sites correlates well with that of HTGTS. HTGTS detects improper genomic rearrangements by employing a target site as 'bait' to capture off-target 'prey' sequences. Although GUIDE-seq seems to be generally more advantageous than other methods, it is less sensitive when used with Cas9 variants such as Cas9-nickase. HTGTS can be applied to Cas9 variants but requires two concurrent DSBs in a single cell. Whereas Digenome-seq is a cell-free in vitro system, GUIDE-seq and HTGTS are cell-based systems that are more advantageous in detecting cell-type specific off-target mutations. Therefore, selection of the most suitable method is dependent on each case.

Recent Progress in Genome-Engineering Technologies and Applications to Studies of Mammary Gland Biology

Although the introduction of the CRISPR/Cas9 system revolutionized the field of genome engineering, several concerns still exist, including the low efficiency by which its generates knockins, the inefficient cellular delivery of the relatively large Cas9, and the potential for off-target activity. To overcome these issues, many attempts have been made to improve CRISPR/Cas9 system (Fig. 4). *Streptococcus pyogenes*-derived Cas9 (SpCas9) is fairly large (4.1 kb), and is thus difficult to pack into certain viral delivery vectors (e.g., adeno-associated virus; AAV) for in vivo genome engineering. To improve the cellular delivery of the utilized endonuclease, researchers have used smaller nuclease variants, including Cpf1 [89], SaCas9 [23], and CjCas9 [90]. Cpf1, which was derived from *Francisella novicida*, is slightly smaller (3.9 kb) than SpCas9, and CRISPR/Cpf1 offers several advantages over CRISPR/Cas9. For example, Cpf1 recognizes different PAMs, which allows new targeting possibilities. It also creates sticky ends rather than the blunt ends generated by Cas9, which enhances target specificity. Cpf1 cleaves target DNA 18–23 nucleotides downstream of the PAM, which is farther away than the Cas9 cleavage site (3 nt) and enables multiple rounds of DNA cleavage. SaCas9, which is derived from *Staphylococcus aureus*, is even smaller (3.16 kb), but still resists being packed with some reporter genes in the AAV vector. CjCas, which is derived from *Campylobacter jejuni*, is the smallest Cas9 ortholog that has been reported to date (2.95 kb) and has been shown to drive highly specific mutagenesis in human and mouse genomes.

Beyond the genomic deletions that may be engineered by Cas9, its modification has enabled the system to be used for the transcriptional activation or repression of specific genes. One such example, CRISPRi, is the fusion protein of catalytically inactive "dead" Cas9 (dCas9) linked to transcriptional repressors, such as KRAB or SID [91–93]. dCas9 does not induce DNA double-strand breaks but retains the DNA-binding feature, and thus facilitates the action of various effectors linked to dCas9. CRISPRa, in contrast, is the fusion protein of dCas9 and transcriptional activators such as VP64 or p300 [94–97]. These modified fusion proteins enable researchers to reversibly control the transcriptional regulation of targeted genes.

Although the efficiency of CRISPR/Cas9-directed gene knockout is very high (> 80%), given the dominance of NHEJ, knockin occurs with an efficiency of only ~0.5~5% due to the low frequency of HDR. The recent development of a CRISPR/Cas9-derived base-editing technology has provided researchers with a powerful tool to introduce specific point mutations at desired genomic sites. The fusion of dCas9 with an APOBEC1 cytidine deaminase generated a base editor (BE1) that can introduce specific C to T or G to A nucleotide changes without requiring any DNA double-strand break or donor DNA template [98]. Further optimization of targeting efficiency led to the establishment of various base-editing platforms, including BE2, BE3, and BE4, which are based on the use of Cas9 nickase (Cas9 D10A) instead of dCas9 [98–100]. Cytidine deaminase strictly converts C to T or G to A, but the recently discovered adenine base editors (ABEs) can also covert A to G or T to C [101]. Moreover, the recent development of Cpf1 coupled with APOBEC1 cytidine deaminase has broadened the targeting range by adding the ability to use different PAMs [102]. Thus, the base editors have been successfully used to switch a single nucleotide and have been found to induce less off-target genome modification. However, they are still large fusion proteins that cannot be easily packed within viral vectors for in vivo delivery. Further improvements might be necessary to reduce the molecular sizes of base editors and enable their efficient viral gene delivery, or to explore the use of an alternative gene delivery method, such as nanotechnology.

Mammary glands are unique organs specialized for milk secretion. Although mammary-like organoids have been successfully established, these methods are technically challenging and the resulting organoids do not fully recapitulate the secretory function of mammary glands, their major and foremost activity [103–106]. Thus, mammalian models, such as mouse models, would be the most reliable experimental models to study mammary gland biology. Although conventional knockout mouse models have been extensively used to study normal mammary development and to investigate the molecular mechanisms underlying breast cancer formation, these methods are time-consuming and laborious [107–109].

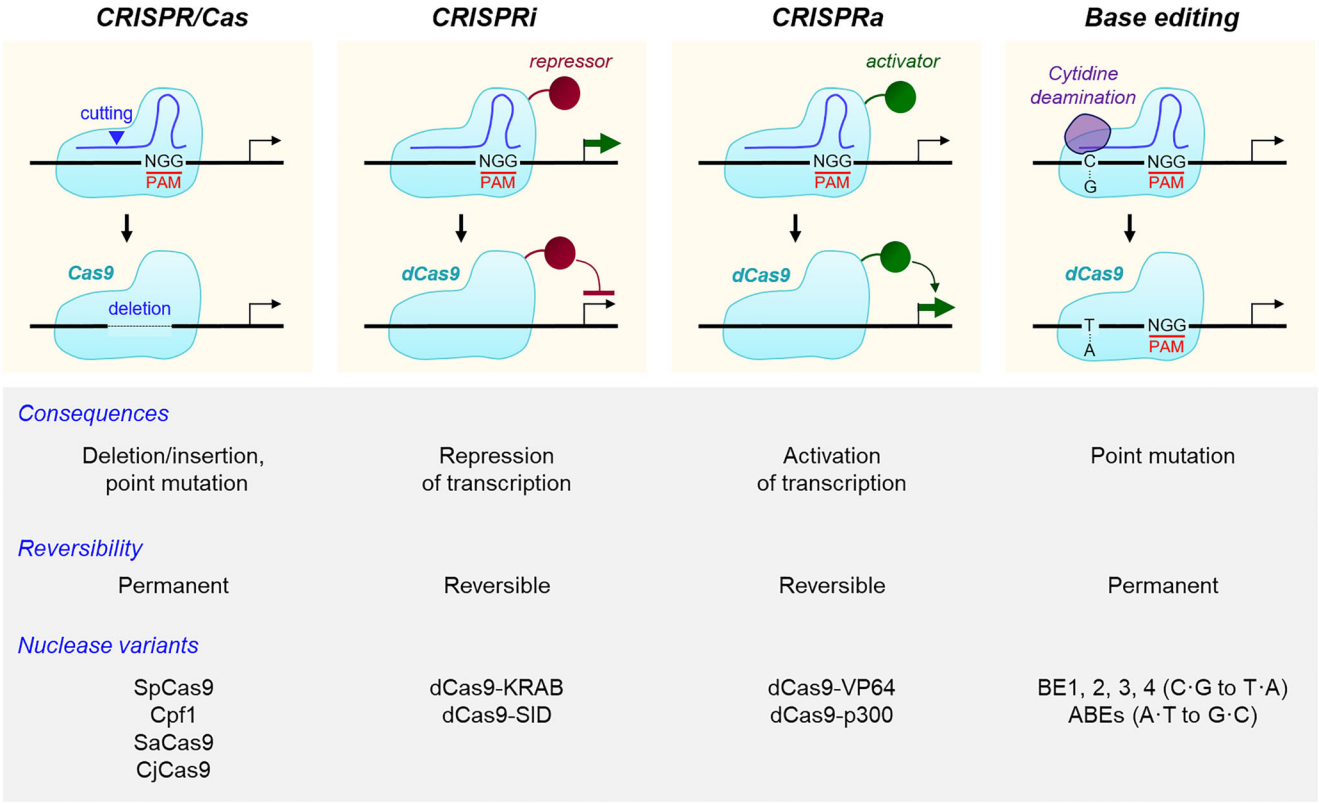

Fig. 4 Molecular consequences of distinct RNA-guided endonucleases. While Cas9 generates permanent deletions, insertions, and point mutations in the genomic DNA, the use of CRISPRi or CRISPRa induces the reversible transcriptional repression or activation, respectively, of specific genes. Base editing specifically introduces permanent point mutations without inducing DNA double-strand breaks. Representative examples of nuclease variants are presented

In addition, because super-enhancers, including those specific to mammary glands, are large (over 12.5 kb) genomic fragments, it is quite exhaustive and technically challenging to delete entire super-enhancer regions or mutate specific constituent enhancers in mice by conventional knockout or knock-in methods. In contrast, use of advanced CRISPR/Cas9 genome editing techniques can easily create germline mutant mice or conditional mutant mice having the desired mutations within a short period of time at relatively low cost. Although improvements in targeting accuracy are required, CRISPR/Cas9 is more advantageous in studying the detailed underlying mechanisms of mammary-specific super-enhancers. It remains unclear whether a super-enhancer unit is sufficient to convert cell type specificity in vivo. It would be intriguing to examine whether translocation of a mammary-specific super-enhancer into the universal genomic region confers mammary-specific gene controls. This genomic rearrangement can be achieved by the CRISPR/Cas9 system. Although STAT5 has been found essential for the activity of mammary-specific super-enhancers, the sequential process of super-enhancer establishment is largely unknown. Using advanced base-editing techniques, the DNA binding sites of key transcription factors, other than STAT5, can be specifically mutated within constituent super-enhancers and their sequential occupancy can be thoroughly investigated. Moreover, the detailed molecular mechanisms underlying the establishment of mammary-specific super-enhancers can be examined using Cas9 fusion proteins linked to activators or repressors by specifically controlling the transcriptional activity of individual enhancers without disrupting any enhancer sequences. Indeed, mammary tissue contains a heterogeneous population of cells, including basal and luminal cells with different functions. Therefore, single cell sequencing may be required to investigate the distinct features of super-enhancers specific to each cell subtype. However, because transcription factors easily dissociate from chromatin during isolation of single cells, it is hard to obtain reliable ChIP-seq results. Further studies are required to improve these technical issues.

Closing Remarks

Recent advances in next-generation-sequencing and genome-engineering technologies have provided researchers with unprecedented opportunities to understand the genomic features of the regulatory units that are responsible for controlling mammalian development, including the differentiation of the

mammary epithelium. Compared to the use of genome-wide analysis with a single tool or single dataset, the integration of results obtained from several relevant genome-wide analysis tools, such as DNase-seq, ChIP-seq, and RNase-seq, may yield more robust research conclusions. The recently introduced CRISPR/Cas9 genome-editing technique has enabled researchers to determine the biological relevance of putative regulatory elements by directly mutating any given sequence, including specific transcription factor binding sites. Although CRISPR/Cas9 shows great promise for editing specific DNA sequences, several issues remain unresolved. In addition to the generation of unexpectedly large deletions, inversion of large fragments or duplications are observed when two adjacent sites are co-cut [110]. Unwanted random insertions are also frequent, and off-target effects (i.e., mutations at non-target sites) are a serious concern for the clinical use of CRISPR/Cas9 in genome engineering. Finally, recent reports have suggested that the CRISPR/Cas9 may increase the cancer risk by inactivating p53 [111, 112]. Further improvements in the techniques for genome-wide analysis and genome engineering are needed if we hope to improve our fundamental understanding of the complex genomic circuits essential for mammalian development and move toward the clinical application of genome-editing tools for disease treatment.

Acknowledgements This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2018R1C1B6001117) and the Collaborative Genome Program for Fostering New Post-Genome Industry of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (MSIT) (NRF-2017M3C9A6044519).

References

1. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272–86. https://doi.org/10.1038/nrg3682.
2. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311–22. https://doi.org/10.1016/j.cell.2007.12.014.
3. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316(5830):1497–502. https://doi.org/10.1126/science.1141319.
4. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. 2007;4(8):651–7. https://doi.org/10.1038/nmeth1068.
5. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 2011;12(4):283–93. https://doi.org/10.1038/nrg2957.
6. Ong CT, Corces VG. Enhancers: emerging roles in cell fate specification. EMBO Rep. 2012;13(5):423–30. https://doi.org/10.1038/embor.2012.52.
7. Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques. 2008;45(1):81–94. https://doi.org/10.2144/000112900.
8. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63. https://doi.org/10.1038/nrg2484.
9. Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature. 2005;435(7042):646–51. https://doi.org/10.1038/nature03556.
10. Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. 2007;25(7):778–85. https://doi.org/10.1038/nbt1319.
11. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326(5959):1509–12. https://doi.org/10.1126/science.1178811.
12. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. 2009;326(5959):1501. https://doi.org/10.1126/science.1178817.
13. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 2010;186(2):757–61. https://doi.org/10.1534/genetics.110.120717.
14. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011;29(2):143–8. https://doi.org/10.1038/nbt.1755.
15. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60(2):174–82. https://doi.org/10.1007/s00239-004-0046-3.
16. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology. 2005;151(Pt 3):653–63. https://doi.org/10.1099/mic.0.27437-0.
17. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(Pt 8):2551–61. https://doi.org/10.1099/mic.0.28048-0.
18. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012;482(7385):331–8. https://doi.org/10.1038/nature10886.
19. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816–21. https://doi.org/10.1126/science.1225829.
20. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819–23. https://doi.org/10.1126/science.1231143.
21. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823–6. https://doi.org/10.1126/science.1232033.
22. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013;153(4):910–8. https://doi.org/10.1016/j.cell.2013.04.025.
23. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520(7546):186–91. https://doi.org/10.1038/nature14299.

24. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307–19. https://doi.org/10.1016/j.cell.2013.03.035.
25. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934–47. https://doi.org/10.1016/j.cell.2013.09.053.
26. Yin JW, Wang G. The mediator complex: a master coordinator of transcription and cell lineage development. Development. 2014;141(5):977–87. https://doi.org/10.1242/dev.098392.
27. Vahedi G, Kanno Y, Furumoto Y, Jiang K, Parker SC, Erdos MR, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature. 2015;520(7548):558–62. https://doi.org/10.1038/nature14154.
28. Siersbaek R, Rabiee A, Nielsen R, Sidoli S, Traynor S, Loft A, et al. Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. Cell Rep. 2014;7(5):1443–55. https://doi.org/10.1016/j.celrep.2014.04.042.
29. Adam RC, Yang H, Rockowitz S, Larsen SB, Nikolova M, Oristian DS, et al. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature. 2015;521(7552):366–70. https://doi.org/10.1038/nature14289.
30. Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y, et al. Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. Dev Cell. 2016;36(1):9–23. https://doi.org/10.1016/j.devcel.2015.12.014.
31. Shin HY, Willi M, HyunYoo K, Zeng X, Wang C, Metser G, et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat Genet. 2016;48(8):904–11. https://doi.org/10.1038/ng.3606.
32. Pittius CW, Sankaran L, Topper YJ, Hennighausen L. Comparison of the regulation of the whey acidic protein gene with that of a hybrid gene containing the whey acidic protein gene promoter in transgenic mice. Mol Endocrinol. 1988;2(11):1027–32. https://doi.org/10.1210/mend-2-11-1027.
33. Robinson GW, Kang K, Yoo KH, Tang Y, Zhu BM, Yamaji D, et al. Coregulation of genetic programs by the transcription factors NFIB and STAT5. Mol Endocrinol. 2014;28(5):758–67. https://doi.org/10.1210/me.2012-1387.
34. Zhou J, Chehab R, Tkalcevic J, Naylor MJ, Harris J, Wilson TJ, et al. Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. EMBO J. 2005;24(3):635–44. https://doi.org/10.1038/sj.emboj.7600538.
35. Lee HK, Willi M, Wang C, Yang CM, Smith HE, Liu C, et al. Functional assessment of CTCF sites at cytokine-sensing mammary enhancers using CRISPR/Cas9 gene editing in mice. Nucleic Acids Res. 2017;45(8):4606–18. https://doi.org/10.1093/nar/gkx185.
36. Willi M, Yoo KH, Reinisch F, Kuhns TM, Lee HK, Wang C, et al. Facultative CTCF sites moderate mammary super-enhancer activity and regulate juxtaposed gene in non-mammary cells. Nat Commun. 2017;8:16069. https://doi.org/10.1038/ncomms16069.
37. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science. 2014;346(6215):1373–7. https://doi.org/10.1126/science.1259037.
38. Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe JS, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev. 2013;27(24):2648–62. https://doi.org/10.1101/gad.232710.113.
39. Drier Y, Cotton MJ, Williamson KE, Gillespie SM, Ryan RJ, Kluk MJ, et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. Nat Genet. 2016;48(3):265–72. https://doi.org/10.1038/ng.3502.
40. Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A. 2013;110(44):17921–6. https://doi.org/10.1073/pnas.1317023110.
41. Achour M, Le Gras S, Keime C, Parmentier F, Lejeune FX, Boutillier AL, et al. Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of Huntington's disease mice. Hum Mol Genet. 2015;24(12):3481–96. https://doi.org/10.1093/hmg/ddv099.
42. Le Gras S, Keime C, Anthony A, Lotz C, De Longprez L, Brouillet E, et al. Altered enhancer transcription underlies Huntington's disease striatal transcriptional signature. Sci Rep. 2017;7:42875. https://doi.org/10.1038/srep42875.
43. Shin HY. Targeting super-enhancers for disease treatment and diagnosis. Mol Cells. 2018;41(6):506–14. https://doi.org/10.14348/molcells.2018.2297.
44. Zeng L, Zhou MM. Bromodomain: an acetyl-lysine binding domain. FEBS Lett. 2002;513(1):124–8.
45. Zuber V, Bettella F, Witoelar A, Consortium P, Cruk G, Consortium B, et al. Bromodomain protein 4 discriminates tissue-specific super-enhancers containing disease-specific susceptibility loci in prostate and breast cancer. BMC Genomics. 2017;18(1):270. https://doi.org/10.1186/s12864-017-3620-y.
46. Glodzik D, Morganella S, Davies H, Simpson PT, Li Y, Zou X, et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. Nat Genet. 2017;49(3):341–8. https://doi.org/10.1038/ng.3771.
47. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320–34. https://doi.org/10.1016/j.cell.2013.03.036.
48. Ohba S, He X, Hojo H, McMahon AP. Distinct transcriptional programs underlie Sox9 regulation of the mammalian chondrocyte. Cell Rep. 2015;12(2):229–43. https://doi.org/10.1016/j.celrep.2015.06.013.
49. Liu CF, Lefebvre V. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. Nucleic Acids Res. 2015;43(17):8183–203. https://doi.org/10.1093/nar/gkv688.
50. Wang AH, Juan AH, Ko KD, Tsai PF, Zare H, Dell'Orso S, et al. The elongation factor Spt6 maintains ESC pluripotency by controlling super-enhancers and counteracting Polycomb proteins. Mol Cell. 2017;68(2):398–413 e6. https://doi.org/10.1016/j.molcel.2017.09.016.
51. Lomberk G, Blum Y, Nicolle R, Nair A, Gaonkar KS, Marisa L, et al. Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. Nat Commun. 2018;9(1):1978. https://doi.org/10.1038/s41467-018-04383-6.
52. Suzuki HI, Young RA, Sharp PA. Super-enhancer-mediated RNA processing revealed by integrative MicroRNA network analysis. Cell. 2017;168(6):1000–14 e15. https://doi.org/10.1016/j.cell.2017.02.015.
53. Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MG, Qi J, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell. 2013;24(6):777–90. https://doi.org/10.1016/j.ccr.2013.11.003.
54. Pelish HE, Liau BB, Nitulescu II, Tangpeerachaikul A, Poss ZC, Da Silva DH, et al. Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature. 2015;526(7572):273–6. https://doi.org/10.1038/nature14904.
55. Das S, Senapati P, Chen Z, Reddy MA, Ganguly R, Lanting L, et al. Regulation of angiotensin II actions by enhancers and super-enhancers in vascular smooth muscle cells. Nat Commun. 2017;8(1):1467. https://doi.org/10.1038/s41467-017-01629-7.

56. Nakamura Y, Hattori N, Iida N, Yamashita S, Mori A, Kimura K, et al. Targeting of super-enhancers and mutant BRAF can suppress growth of BRAF-mutant colon cancer cells via repression of MAPK signaling pathway. Cancer Lett. 2017;402:100–9. https://doi.org/10.1016/j.canlet.2017.05.017.
57. Gelato KA, Schockel L, Klingbeil O, Ruckert T, Lesche R, Toedling J, et al. Super-enhancers define a proliferative PGC-1alpha-expressing melanoma subgroup sensitive to BET inhibition. Oncogene. 2018;37(4):512–21. https://doi.org/10.1038/onc.2017.325.
58. Andricovich J, Perkail S, Kai Y, Casasanta N, Peng W, Tzatsos A. Loss of KDM6A activates super-enhancers to induce gender-specific squamous-like pancreatic Cancer and confers sensitivity to BET inhibitors. Cancer Cell. 2018;33(3):512–26 e8. https://doi.org/10.1016/j.ccell.2018.02.003.
59. Gerlach D, Tontsch-Grunt U, Baum A, Popow J, Scharn D, Hofmann MH, et al. The novel BET bromodomain inhibitor BI 894999 represses super-enhancer-associated transcription and synergizes with CDK9 inhibition in AML. Oncogene. 2018;37(20):2687–701. https://doi.org/10.1038/s41388-018-0150-2.
60. Nakagawa M, Shaffer AL 3rd, Ceribelli M, Zhang M, Wright G, Huang DW, et al. Targeting the HTLV-I-regulated BATF3/IRF4 transcriptional network in adult T cell leukemia/lymphoma. Cancer Cell. 2018;34(2):286–97e10. https://doi.org/10.1016/j.ccell.2018.06.014.
61. Chen D, Zhao Z, Huang Z, Chen DC, Zhu XX, Wang YZ, et al. Super enhancer inhibitors suppress MYC driven transcriptional amplification and tumor progression in osteosarcoma. Bone Res. 2018;6:11. https://doi.org/10.1038/s41413-018-0009-8.
62. Li Y, Rivera CM, Ishii H, Jin F, Selvaraj S, Lee AY, et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. PLoS One. 2014;9(12):e114485. https://doi.org/10.1371/journal.pone.0114485.
63. Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol Cell. 2015;58(2):362–70. https://doi.org/10.1016/j.molcel.2015.02.014.
64. Moorthy SD, Davidson S, Shchuka VM, Singh G, Malek-Gilani N, Langroudi L, et al. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res. 2017;27(2):246–58. https://doi.org/10.1101/gr.210930.116.
65. Dave K, Sur I, Yan J, Zhang J, Kaasinen E, Zhong F, et al. Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth. Elife. 2017;6. https://doi.org/10.7554/eLife.23382.
66. Huang J, Li K, Cai W, Liu X, Zhang Y, Orkin SH, et al. Dissecting super-enhancer hierarchy based on chromatin interactions. Nat Commun. 2018;9(1):943. https://doi.org/10.1038/s41467-018-03279-9.
67. Schuijers J, Manteiga JC, Weintraub AS, Day DS, Zamudio AV, Hnisz D, et al. Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. Cell Rep. 2018;23(2):349–60. https://doi.org/10.1016/j.celrep.2018.03.056.
68. Pantera H, Moran JJ, Hung HA, Pak E, Dutra A, Svaren J. Regulation of the neuropathy-associated Pmp22 gene by a distal super-enhancer. Hum Mol Genet. 2018;27(16):2830–9. https://doi.org/10.1093/hmg/ddy191.
69. Metser G, Shin HY, Wang C, Yoo KH, Oh S, Villarino AV, et al. An autoregulatory enhancer controls mammary-specific STAT5 functions. Nucleic Acids Res. 2016;44(3):1052–63. https://doi.org/10.1093/nar/gkv999.
70. Zeng X, Willi M, Shin HY, Hennighausen L, Wang C. Lineage-specific and non-specific cytokine-sensing genes respond differentially to the master regulator STAT5. Cell Rep. 2016;17(12):3333–46. https://doi.org/10.1016/j.celrep.2016.11.079.
71. Willi M, Yoo KH, Wang C, Trajanoski Z, Hennighausen L. Differential cytokine sensitivities of STAT5-dependent enhancers rely on Stat5 autoregulation. Nucleic Acids Res. 2016;44(21):10277–91. https://doi.org/10.1093/nar/gkw844.
72. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat Commun. 2017;8:15464. https://doi.org/10.1038/ncomms15464.
73. Sung YH, Kim JM, Kim HT, Lee J, Jeon J, Jin Y, et al. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res. 2014;24(1):125–31. https://doi.org/10.1101/gr.163394.113.
74. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008;24(11):529–38. https://doi.org/10.1016/j.tig.2008.08.007.
75. Zhou J, Wang J, Shen B, Chen L, Su Y, Yang J, et al. Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. FEBS J. 2014;281(7):1717–25. https://doi.org/10.1111/febs.12735.
76. Fujii W, Kawasaki K, Sugiura K, Naito K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. Nucleic Acids Res. 2013;41(20):e187. https://doi.org/10.1093/nar/gkt772.
77. Wang L, Shao Y, Guan Y, Li L, Wu L, Chen F, et al. Large genomic fragment deletion and functional gene cassette knock-in via Cas9 protein mediated genome editing in one-cell rodent embryos. Sci Rep. 2015;5:17517. https://doi.org/10.1038/srep17517.
78. Hara S, Kato T, Goto Y, Kubota S, Tamano M, Terao M, et al. Microinjection-based generation of mutant mice with a double mutation and a 0.5 Mb deletion in their genome by the CRISPR/Cas9 system. J Reprod Dev. 2016;62(5):531–6. https://doi.org/10.1262/jrd.2016-058.
79. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018;36(8):765–71. https://doi.org/10.1038/nbt.4192.
80. Lee H, Kim JS. Unexpected CRISPR on-target effects. Nat Biotechnol. 2018;36(8):703–4. https://doi.org/10.1038/nbt.4207.
81. Bolukbasi MF, Gupta A, Wolfe SA. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. Nat Methods. 2016;13(1):41–50. https://doi.org/10.1038/nmeth.3684.
82. Kanchiswamy CN, Maffei M, Malnoy M, Velasco R, Kim JS. Fine-tuning next-generation genome editing tools. Trends Biotechnol. 2016;34(7):562–74. https://doi.org/10.1016/j.tibtech.2016.03.007.
83. Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods. 2015;12(3):237–43, 1 p following 43. https://doi.org/10.1038/nmeth.3284.
84. Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods. 2013;10(4):361–5. https://doi.org/10.1038/nmeth.2408.
85. Wang X, Wang Y, Wu X, Wang J, Wang Y, Qiu Z, et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat Biotechnol. 2015;33(2):175–8. https://doi.org/10.1038/nbt.3127.
86. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, Kaeppel C, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol. 2011;29(9):816–23. https://doi.org/10.1038/nbt.1948.

87. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol. 2015;33(2):179–86. https://doi.org/10.1038/nbt.3101.
88. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015;33(2):187–97. https://doi.org/10.1038/nbt.3117.
89. Kim HK, Song M, Lee J, Menon AV, Jung S, Kang YM, et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. Nat Methods. 2017;14(2):153–9. https://doi.org/10.1038/nmeth.4104.
90. Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, et al. In vivo genome editing with a small Cas9 orthologue derived from campylobacter jejuni. Nat Commun. 2017;8:14500. https://doi.org/10.1038/ncomms14500.
91. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods. 2015;12(12):1143–9. https://doi.org/10.1038/nmeth.3630.
92. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442–51. https://doi.org/10.1016/j.cell.2013.06.044.
93. Konermann S, Brigham MD, Trevino A, Hsu PD, Heidenreich M, Cong L, et al. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013;500(7463):472–6. https://doi.org/10.1038/nature12466.
94. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol. 2015;33(5):510–7. https://doi.org/10.1038/nbt.3199.
95. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973–6. https://doi.org/10.1038/nmeth.2600.
96. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977–9. https://doi.org/10.1038/nmeth.2598.
97. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol. 2013;31(9):833–8. https://doi.org/10.1038/nbt.2675.
98. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533(7603):420–4. https://doi.org/10.1038/nature17946.
99. Rees HA, Komor AC, Yeh WH, Caetano-Lopes J, Warman M, Edge ASB, et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun. 2017;8:15790. https://doi.org/10.1038/ncomms15790.
100. Liang P, Sun H, Sun Y, Zhang X, Xie X, Zhang J, et al. Effective gene editing by high-fidelity base editor 2 in mouse zygotes. Protein Cell. 2017;8(8):601–11. https://doi.org/10.1007/s13238-017-0418-2.
101. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Publisher correction: programmable base editing of a*T to G*C in genomic DNA without DNA cleavage. Nature. 2018;559:E8. https://doi.org/10.1038/s41586-018-0070-x.
102. Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, et al. Base editing with a Cpf1-cytidine deaminase fusion. Nat Biotechnol. 2018;36(4):324–7. https://doi.org/10.1038/nbt.4102.
103. Qu Y, Han B, Gao B, Bose S, Gong Y, Wawrowsky K, et al. Differentiation of human induced pluripotent stem cells to mammary-like organoids. Stem Cell Reports. 2017;8(2):205–15. https://doi.org/10.1016/j.stemcr.2016.12.023.
104. Jamieson PR, Dekkers JF, Rios AC, Fu NY, Lindeman GJ, Visvader JE. Derivation of a robust mouse mammary organoid system for studying tissue dynamics. Development. 2017;144(6):1065–71. https://doi.org/10.1242/dev.145045.
105. Laperrousaz B, Porte S, Gerbaud S, Harma V, Kermarrec F, Hourtane V, et al. Direct transfection of clonal organoids in Matrigel microbeads: a promising approach toward organoid-based genetic screens. Nucleic Acids Res. 2018;46(12):e70. https://doi.org/10.1093/nar/gky030.
106. Hennighausen L. Mouse models for breast cancer. Breast Cancer Res. 2000;2(1):2–7.
107. Pittius CW, Hennighausen L, Lee E, Westphal H, Nicols E, Vitale J, et al. A milk protein gene promoter directs the expression of human tissue plasminogen activator cDNA to the mammary gland in transgenic mice. Proc Natl Acad Sci U S A. 1988;85(16):5874–8.
108. Liu X, Robinson GW, Wagner KU, Garrett L, Wynshaw-Boris A, Hennighausen L. Stat5a is mandatory for adult mammary gland development and lactogenesis. Genes Dev. 1997;11(2):179–86.
109. Wagner KU, Wall RJ, St-Onge L, Gruss P, Wynshaw-Boris A, Garrett L, et al. Cre-mediated gene deletion in the mammary gland. Nucleic Acids Res. 1997;25(21):4323–30.
110. Boroviak K, Doe B, Banerjee R, Yang F, Bradley A. Chromosome engineering in zygotes with CRISPR/Cas9. Genesis. 2016;54(2):78–85. https://doi.org/10.1002/dvg.22915.
111. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat Med. 2018;24:927–30. https://doi.org/10.1038/s41591-018-0049-z.
112. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat Med. 2018. https://doi.org/10.1038/s41591-018-0050-6.