

REVIEW

Purification of specific DNA species using the CRISPR system

Toshitsugu Fujita and Hodaka Fujii*

Department of Biochemistry and Genome Biology, Hirosaki University Graduate School of Medicine, Zaifu-cho 5, Hirosaki, Aomori 036-8562, Japan

*Correspondence address. Department of Biochemistry and Genome Biology, Hirosaki University Graduate School of Medicine, Zaifu-cho 5, Hirosaki, Aomori 036-8562, Japan. Tel: +81-(0)172-39-5018; Fax: +81-(0)172-39-5020; E-mail: hodaka@hirosaki-u.ac.jp

Abstract

In 2013, we developed a new method of engineered DNA-binding molecule-mediated chromatin immunoprecipitation that incorporates the clustered regularly interspaced short palindromic repeats (CRISPR) system to purify specific DNA species. This CRISPR-mediated purification can be performed *in-cell* or *in vitro*; CRISPR complexes can be expressed to tag target DNA sequences in the cells to be analyzed, or a CRISPR ribonucleoprotein complex consisting of recombinant nuclease-dead Cas9 (dCas9) and synthetic guide RNA can be used to tag target DNA sequences *in vitro*. Both methods enable purification of specific DNA sequences in chromatin structures for subsequent identification of molecules (proteins, RNAs, and other genomic regions) associated with the target sequences. The *in vitro* method also enables enrichment of purified DNA sequences from a pool of heterogeneous sequences for next-generation sequencing or other applications. In this review, we outline the principle of CRISPR-mediated purification of specific DNA species and discuss recent advances in the technology.

Keywords: CRISPR; enChIP; dCas9; NGS; DNA purification

Introduction

The advent of clustered regularly interspaced short palindromic repeats (CRISPR) technology for genome editing in 2012 [1, 2] opened a wide range of potential applications [3], among them is the purification of specific genomic regions from cells for subsequent biochemical analysis [4]. Several methods had already been developed for this purpose, as discussed in our previous review [5], but all of these methods have drawbacks. For example, tagged oligonucleotide probes can be used to purify specific genomic regions [6], but this requires partial denaturing of chromatin, making it difficult to achieve high yields for analysis of a single copy gene. It is also possible to insert the recognition site of an exogenous DNA-binding molecule such as the bacterial protein LexA into a genomic region of interest and use it as a tag for affinity purification [7], but this requires the laborious and time-consuming step of generating cells harboring the

tagged locus. In addition, insertion of exogenous DNA may disturb the physiological conditions around the insertion site.

To overcome these problems, we developed a general strategy to purify specific DNA species using engineered DNA-binding molecules such as zinc-finger proteins (ZFPs), transcription activator-like (TAL) proteins, and the CRISPR system [4], and called this technology engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). CRISPR technology provides an ideal system for flexible tagging of DNA to be purified. For this purpose, mutant forms of Cas9 lacking nuclease activity but retaining DNA-binding activity (nuclease-dead Cas9, dCas9) are used. Because the specificity of binding to target DNA can be directed by guide RNAs (gRNAs), generation of sequence-specific binding molecules is easier, cheaper, and requires less time than protein-based systems using ZFPs and TALs for which target-specific DNA-binding proteins must be generated each time. These advantages make our CRISPR-based

Received: 1 May 2019; Revised: 2 June 2019; Editorial decision: 3 June 2019; Accepted: 7 June 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

system the method of choice for various applications, including DNA purification.

In addition to purification of specific genomic regions from cells, CRISPR-mediated purification of specific DNA species also enables enrichment of purified DNA sequences from a large number of heterogeneous sequences [8] and may be useful for various applications including clinical diagnosis.

In this review, we outline the principle of CRISPR-mediated purification of specific DNA species and discuss recent advances in the technology.

Principles of CRISPR-mediated purification of specific DNA species

The two methods for CRISPR-mediated purification of specific DNA species differ in whether target DNA is tagged in the cell or *in vitro*.

In-cell tagging

In this method, target DNA is tagged in the cell, typically in the nucleus, by expression of the CRISPR complex (Fig. 1). Expression of the CRISPR complex can be achieved by (i) expression from DNA via transient or stable transfection of expression plasmids, establishment of stable transformants using retroviral or lentiviral vectors, or long-term continuous expression by adenovirus systems; (ii) expression from RNA after transfection of mRNA encoding dCas9 and gRNA (synthetic or *in vitro* transcribed); or (iii) transduction of CRISPR ribonucleoprotein (RNP) complexes consisting of recombinant dCas9 and synthetic gRNA. DNA-mediated expression, particularly sustained expression by retroviral, lentiviral, or adenoviral vectors, may be better for generating sufficient numbers of cells for analysis than RNA transfection or RNP transduction.

The in-cell approach permits targeting of cellular genomic DNA as well as parasitic DNA, which may include viral DNA such as hepatitis B virus and episomal DNA such as human papilloma virus. DNA reverse transcribed from an RNA virus and integrated into the host genome can also be targeted. Usually, these DNAs are associated with histones and other DNA-binding proteins as well as RNA and other genomic regions to form chromatin structures. After DNA purification, mass spectrometry (MS) can be used to identify proteins associated with the target DNA and next-generation sequencing (NGS) can be used to identify RNA and other genomic regions.

In vitro tagging

In this method, target DNA is tagged *in vitro* using CRISPR complexes consisting of recombinant dCas9 and synthetic gRNA (Fig. 2). This method is applicable for targets in purified DNA and in chromatin DNA. One advantage of *in vitro* tagging is that expression of CRISPR complexes in target cells is not necessary, so less time and effort is required for DNA purification. This is especially important when purifying a target from primary cells such as clinical samples. *In vitro* tagging also avoids potentially hazardous effects caused by expression and binding of CRISPR complexes to genomic DNA, which may include inhibition of gene expression (CRISPR interference effect) and changes in chromatin accessibility [9]. One drawback of *in vitro* tagging, however, is that the yield of affinity purification is significantly lower than that for in-cell tagging [8]. It is possible that during *in vitro* tagging, only a fraction of target sites are accessible to CRISPR complexes, whereas in-cell tagging permits binding during the entire cell cycle. The feasibility of expressing CRISPR

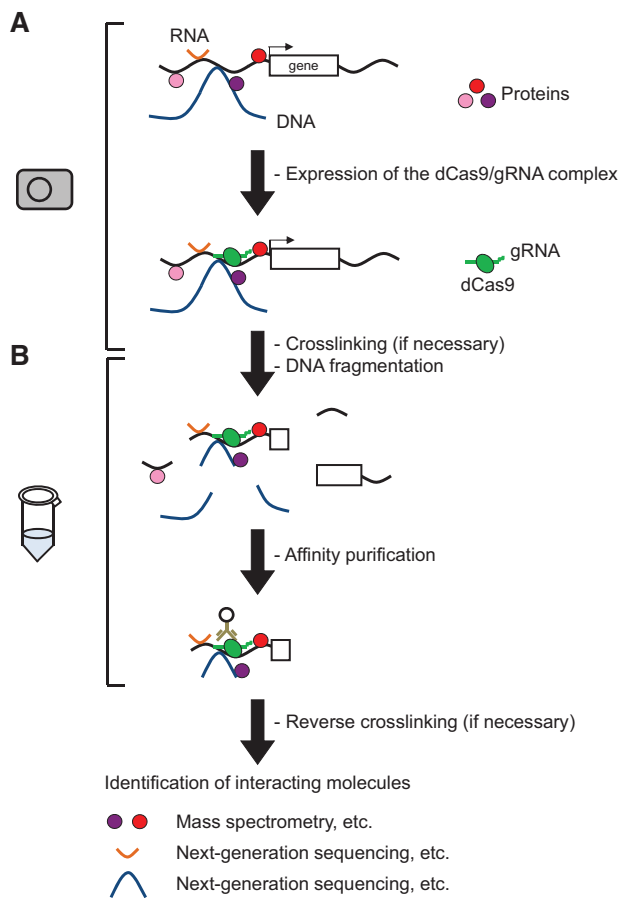


Figure 1: In-cell tagging of specific genomic regions using the CRISPR complex. (A) A locus of interest is tagged with CRISPR complex, consisting of a catalytically inactive form of Cas9 (dCas9) and gRNA that recognizes the target sequence, that is expressed in the cells to be analyzed. (B) Intermolecular interactions can be stabilized with a cross-linker such as formaldehyde, if necessary. Subsequently, chromatin is fragmented by sonication or enzymatic reactions. The target genomic region is isolated by affinity purification. After reverse cross-linking if necessary, molecules interacting with the target genomic region are identified by MS (proteins) and NGS (RNAs and other genomic regions).

complexes and the number of obtainable cells should be considered when deciding which method of tagging to use.

In addition to purification of DNA from chromatin, *in vitro* tagging enables isolation of specific DNA species from a heterogeneous mixture of purified DNA molecules (Fig. 3) [8]. Potential applications include target enrichment for NGS analysis, enrichment of pathogenic DNA from clinical specimens to enable more efficient diagnosis, and enrichment of exogenous DNA integrated during gene therapy and insertional mutagenesis to identify integration sites. *In vitro* tagging can also be used to remove undesirable DNA species, such as reverse-transcribed ribosomal RNA and mitochondrial DNA before transcriptome analysis and 16S rDNA species from mitochondria, chloroplasts, or dominant bacteria during 16S rDNA-based microbiome analysis, among others. Such use decreases the number of reads necessary to achieve sufficient quality of NGS analysis and reduces cost, although reports of such applications are lacking.

Downstream applications of CRISPR-mediated DNA purification

Since our first report of CRISPR-mediated purification of specific DNA species in 2013 [4], an increasing number of applications of

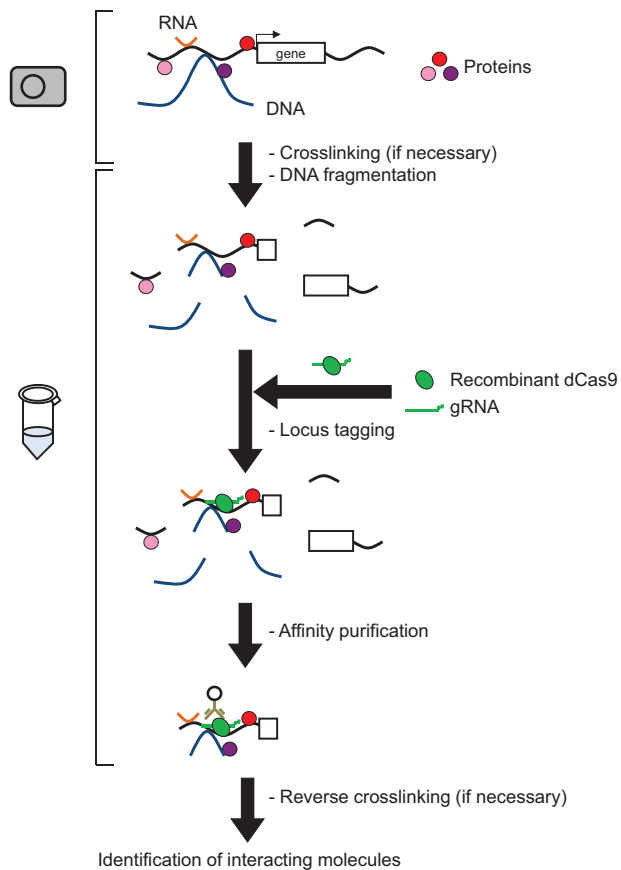


Figure 2: *In vitro* tagging of specific genomic regions using the CRISPR complex. Recombinant dCas9 protein is produced in an appropriate expression system such as *Escherichia coli* or silkworm, and gRNA is chemically synthesized, and then the two are mixed in a test tube to form functional CRISPR complexes. Subsequently, fragmented chromatin (native or fixed) is incubated with the CRISPR complexes for tagging. After affinity purification, molecules associated with the target DNA are identified by MS (proteins) or NGS (RNAs and other genomic regions).

our strategy have been reported (Fig. 4). We discuss these applications below.

Identification of proteins associated with a specific genomic region

CRISPR-mediated purification of specific DNA species can be combined with MS analysis to identify proteins associated with target DNA sequences. In our original report, we targeted the promoter of the *interferon regulatory factor-1* (IRF-1) gene [4, 10] and used a shotgun MS approach and stable isotope labeling using amino acids in cell culture (SILAC), a quantitative form of MS [11]. Many studies have used in-cell locus tagging with the CRISPR complex followed by MS to identify proteins associated with specific genomic regions (Fig. 4). *In vitro* tagging has also successfully been used for MS-mediated identification of proteins associated with a target genomic region [12].

Identification of RNAs associated with a specific genomic region

We have shown that TAL protein-mediated purification of specific DNA species (enChIP using TAL proteins) combined with RT-PCR or RNA sequencing enables detection and identification

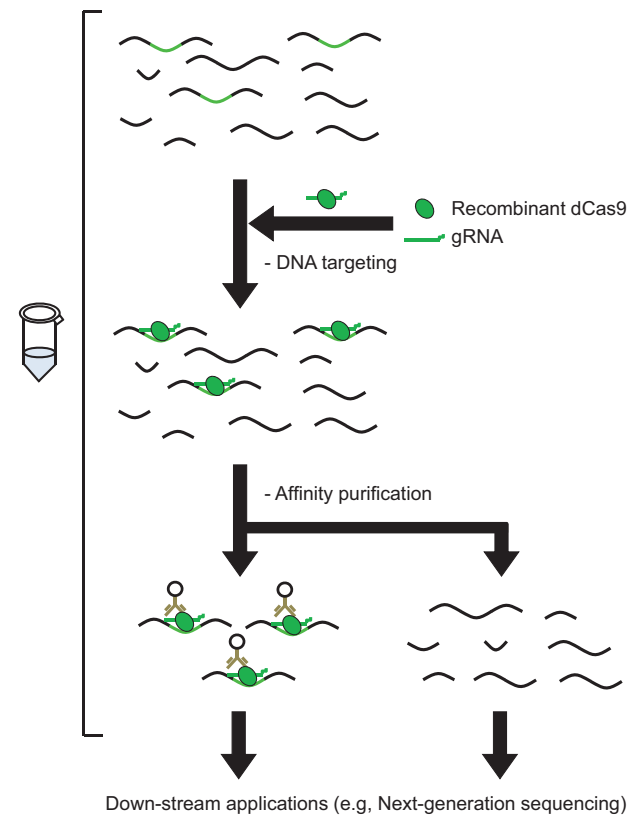


Figure 3: *In vitro* tagging of purified DNA using the CRISPR complex. DNA of interest in a pool of purified DNA species can be tagged using functional CRISPR complexes consisting of recombinant dCas9 protein and synthesized gRNA. Selected DNA species can then be analyzed by downstream applications such as NGS.

of RNAs associated with specific genomic regions [13, 14]. CRISPR-mediated purification of specific genomic regions can also be combined with RNA sequencing to identify RNAs associated with target genomic regions [15, 16].

Identification of intragenomic interactions with a specific genomic region

Investigation of regulation of nuclear organization has revealed that physical interactions between different genomic regions can play critical roles in genome functions such as transcription. Techniques to detect such physical interactions between genomic regions have been developed and reviewed previously [17]. We showed that CRISPR-mediated purification of specific DNA species combined with NGS, a process we call enChIP-Seq, allows identification of physical interactions between genomic regions [18, 19]. We used both in-cell [18] and *in vitro* tagging [19] with the CRISPR complex, and by combining enChIP-Seq, epigenetic profiling of the genomic regions interacting with a target gene promoter, and CRISPR-mediated locus deletion, we successfully identified a potential enhancer located in a different chromosome from the target gene promoter [19]. Subsequently, the same approach was used to identify a functional distant enhancer of the *Sox9* gene [20]. These results demonstrate that CRISPR-mediated locus purification combined with NGS is a powerful tool to identify physical interactions between genomic regions.

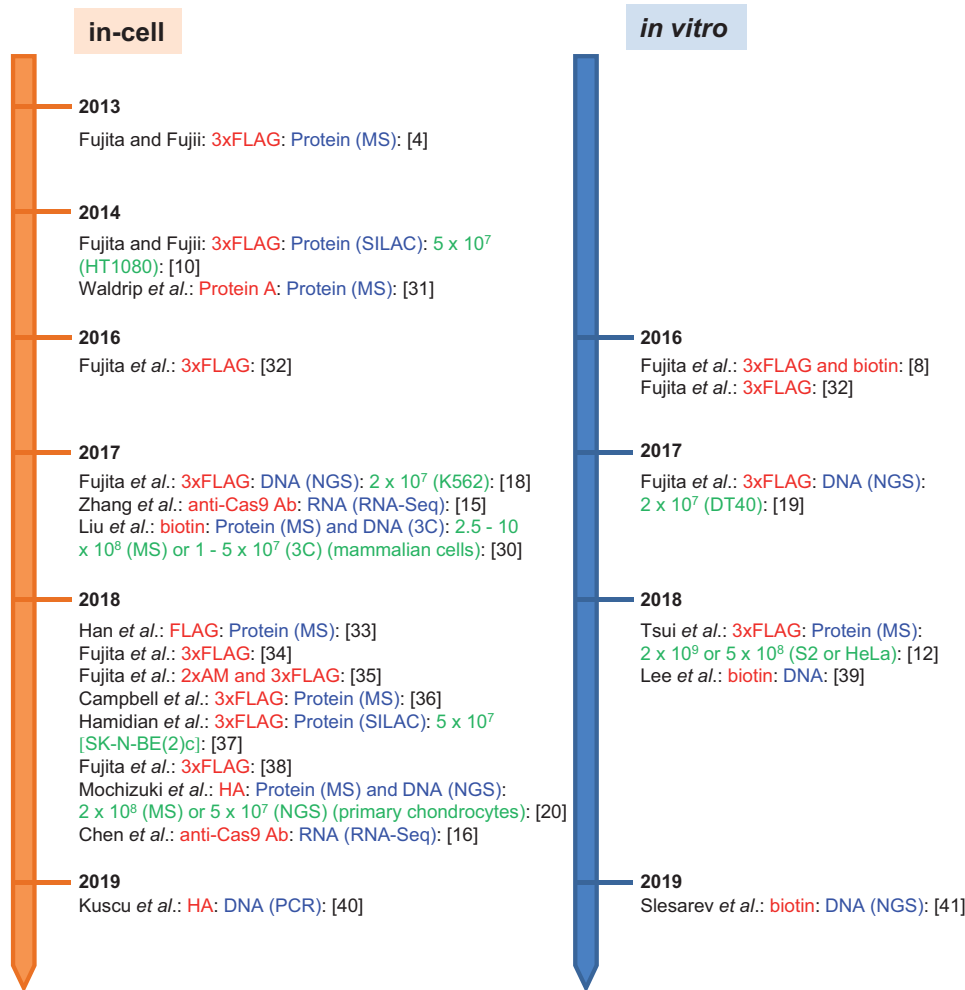


Figure 4: Publications reporting CRISPR-mediated purification of specific DNA species. Epitope tags used for affinity purification and target molecules (and downstream applications) are shown in red and blue [31–34, 36–38, 40, 41]. The CRISPR system of *Streptococcus pyogenes* (*Sp*) was used in all studies. In addition, the CRISPR system of *Staphylococcus aureus* (*Sa*) was used in one study [35]. An updated list of publications, including those using other engineered DNA-binding molecules, can be found at: http://www.med.hirosaki-u.ac.jp/~sim;bgbb/iChIP_related_papers/enChIP/index_e.html.

There are several different approaches to detect physical interactions between genomic regions. The development of chromosome conformation capture (3C) [21] spurred innovation of biochemical detection methods, and many derivatives of 3C have been reported [22]. 3C and related methods depend on ligation of DNA fragments in interacting chromatin complexes, but the ligation step can be a source of noise in the detection of interactions. Therefore, an increasing number of “ligation-free” biochemical methods have been developed, including enChIP-Seq, the in-gel replication of interacting DNA segments (INGRID) assay [23], the split-pool recognition of interactions by tag extension (SPRITE) approach [24], and chromatin interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop) [25] (Table 1). INGRID, SPRITE, and ChIA-Drop all require isolation of a single chromatin complex consisting of different genomic regions. INGRID isolates the complex by its position in a layer of polyacrylamide gel, SPRITE uses a repeat of terminal barcoding of different genomic regions and shuffling, and ChIA-Drop depends on barcoding of different genomic regions in gel-bead-in-emulsion (GEM) droplets.

Of these methods, enChIP-Seq, SPRITE, and ChIA-Drop are non-biased approaches. INGRID is a “one-to-one” approach, whereas enChIP-Seq is a “one-to-all” approach, and SPRITE and

Table 1: Ligation-free biochemical methods to detect interactions between genomic regions

Method	Interactions (n-to-n)	Ref #
INGRID	One-to-one	[23]
enChIP-Seq	One-to-all	[18, 19]
SPRITE	All-to-all	[24]
ChIA-Drop	All-to-all	[25]

ChIA-Drop are “all-to-all” approaches (Table 1). The enChIP-Seq procedure is much simpler than the procedures of SPRITE and ChIA-Drop; because enChIP-Seq using CRISPR is essentially ChIP-Seq of the CRISPR complex, it would not be difficult for anyone experienced with ChIP-Seq to perform. In addition, if the goal is to identify “one-to-all” chromosomal interactions such as identification of enhancers interacting with a target promoter region, enChIP-Seq would be more straightforward and cost effective than SPRITE and ChIA-Drop.

In addition to the biochemical methods discussed, imaging approaches such as fluorescence in situ hybridization [26] and live imaging of specific genomic loci [27–29] have also been

reported. Although these imaging techniques do not provide non-biased detection of interactions between genomic regions, they are valuable for confirming results obtained by 3C-derived methods as well as other ligation-free biochemical methods.

Tag systems, cell numbers, and other experimental conditions used for affinity purification

As shown in Fig. 4, many tag systems are compatible with CRISPR-mediated purification of DNA. Although one report claimed that a biotin tag led to a higher yield than a 1xFLAG tag and anti-Cas9 antibody [30], the authors did not compare the biotin tag with other high-affinity tags such as a 3xFLAG tag. We obtained very high yields using CRISPR-mediated purification of DNA with 3xFLAG-fused dCas9, suggesting that other high-affinity tag systems would also be suitable.

Cell numbers used for CRISPR-mediated purification of DNA vary between experiments. In general, however, identification of proteins associated with a specific genomic region requires more cells than identification of nucleic acids. In total, 5×10^7 to 2×10^9 cells were used for protein identification, whereas 1×10^7 to 5×10^7 cells were used for identification of associated genomic regions (Fig. 4).

Few studies have described the amounts of proteins and nucleic acids used for MS and NGS analyses, respectively. We used 1–5 ng of purified DNA for NGS analysis of genomic regions associated with a specific genomic region [19]. Other reports used 5 ng of DNA for 3C analysis [30] and 10 ng of DNA for NGS [20].

Concluding remarks

In this review, we summarized recent advancements in CRISPR-mediated purification of specific DNA sequences. When this technology is applied to cellular DNA, molecules binding to the target DNA can be identified by MS and NGS. When it is used with purified DNA, target DNA species can be enriched or removed for downstream analyses. This technology will be useful in chromatin and epigenetics research as well as targeted genomics and transcriptomics.

Acknowledgements

We thank members of the Fujii laboratory for development of enChIP and its applications.

Funding

This work was supported by the Takeda Science Foundation (T.F.), Grant-in-Aid for Scientific Research (C) (#15K06895) (T.F.), and Grant-in-Aid for Scientific Research (B) (#15H04329) (T.F., H.F.), ‘Transcription Cycle’ (#15H01354) (H.F.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Conflict of interest statement: T.F. and H.F. have a patent for enChIP (Patent name: Method for isolating specific genomic region using molecule binding specifically to endogenous DNA sequence; Patent number: Japan 5,954,808; Patent application number: WO2014/125668). T.F. and H.F. are founders and directors of Epigeneron, Inc.

References

- Jinek M, Chylinski K, Fonfara I et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–21.
- Gasiunas G, Barrangou R, Horvath P et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 2012;109:E2579–86.
- Xu X, Qi LS. A CRISPR-dCas toolbox for genetic engineering and synthetic biology. *J Mol Biol* 2019;431:34–47.
- Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun* 2013;439:132–36.
- Fujita T, Fujii H. Biochemical analysis of genome functions using locus-specific chromatin immunoprecipitation technologies. *Gene Regul Syst Biol* 2016;10(Suppl. 1):1–9.
- Déjardin J, Kingston RE. Purification of proteins associated with specific genomic loci. *Cell* 2009;136:175–86.
- Hoshino A, Fujii H. Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. *J Biosci Bioeng* 2009;108:446–49.
- Fujita T, Yuno M, Fujii H. Efficient sequence-specific isolation of DNA fragments and chromatin by in vitro enChIP technology using recombinant CRISPR ribonucleoproteins. *Genes Cells* 2016;21:370–77.
- Barkal AA, Srinivasan S, Hashimoto T et al. Cas9 functionally opens chromatin. *PLoS One* 2016;11:e0152683.
- Fujita T, Fujii H. Identification of proteins associated with an IFN γ -responsive promoter by a retroviral expression system for enChIP using CRISPR. *PLoS One* 2014;9:e103084.
- Ong S-E, Blagoev B, Kratchmarova I et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1:376–86.
- Tsui C, Inouye C, Levy M et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc Natl Acad Sci USA* 2018;115:E2734–41.
- Fujita T, Asano Y, Ohtsuka J et al. Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci Rep* 2013;3:3171.
- Fujita T, Yuno M, Okuzaki D et al. Identification of non-coding RNAs associated with telomeres using a combination of enChIP and RNA sequencing. *PLoS One* 2015;10:e0123387.
- Zhang Y, Hu JF, Wang H et al. CRISPR Cas9-guided chromatin immunoprecipitation identifies miR483 as an epigenetic modulator of IGF2 imprinting in tumors. *Oncotarget* 2017;8:34177–90.
- Chen N, Zhao G, Yan X et al. A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biol* 2018;19:218.
- Maass PG, Barutcu AR, Rinn JL. Interchromosomal interactions: a genomic love story of kissing chromosomes. *J Cell Biol* 2019;218:27–38.
- Fujita T, Yuno M, Suzuki Y et al. Identification of physical interactions between genomic regions by enChIP-Seq. *Genes Cells* 2017;22:506–20.
- Fujita T, Kitaura F, Yuno M et al. Locus-specific ChIP combined with NGS analysis reveals genomic regulatory regions that physically interact with the Pax5 promoter in a chicken B cell line. *DNA Res* 2017;24:537–48.

20. Mochizuki Y, Chiba T, Kataoka K et al. Combinatorial CRISPR/Cas9 approach to elucidate a far-upstream enhancer complex for tissue-specific Sox9 expression. *Dev Cell* 2018;**46**: 794–806.e6.
21. Dekker J, Rippe K, Dekker M et al. Capturing chromosome conformation. *Science* 2002;**295**:1306–11.
22. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 2012;**26**:11–24.
23. Gavrilov AA, Chetverina HV, Chermnykh ES et al. Quantitative analysis of genomic element interactions by molecular colony technique. *Nucleic Acids Res* 2014;**42**:e36.
24. Quinodoz SA, Ollikainen N, Tabak B et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* 2018;**174**:744–57.
25. Zheng M, Tian SZ, Capurso D et al. Multiplex chromatin interactions with single-molecule precision. *Nature* 2019;**566**: 558–62.
26. Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. *Nature* 2007;**447**:413–17.
27. Lindhout BI, Franz P, Tessadori F et al. Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res* 2007;**35**:e107.
28. Miyanari Y, Ziegler-Birling C, Torres-Padilla ME. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* 2013;**20**:1321–24.
29. Chen B, Gilbert LA, Cimini BA et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013;**155**:1479–91.
30. Liu X, Zhang Y, Chen Y et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* 2017;**170**: 1028–43.e19.
31. Waldrip ZJ, Byrum SD, Storey AJ et al. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* 2014;**9**:1207–11.
32. Fujita T, Yuno M, Fujii H. Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus. *Sci. Rep* 2016;**6**: 30485.
33. Han B, Zhou B, Qu Y et al. FOXC1-induced non-canonical WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer. *Oncogene* 2018;**37**:1399–408.
34. Fujita T, Kitaura F, Oji A et al. Transgenic mouse lines expressing the 3xFLAG-dCas9 protein for enChIP analysis. *Genes Cells* 2018;**23**:318–25.
35. Fujita T, Yuno M, Fujii H. enChIP systems using different CRISPR orthologues and epitope tags. *BMC Res Notes* 2018;**11**:154.
36. Campbell AE, Shadle SC, Jagannathan S et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *eLife* 2018;**7**:e31023.
37. Hamidian A, Vaapil M, von Stedingk K et al. Promoter-associated proteins of EPAS1 identified by enChIP-MS – a putative role of HDX as a negative regulator. *Biochem Biophys Res Commun* 2018;**499**:291–98.
38. Fujita T, Yuno M, Fujii H. An enChIP system for the analysis of bacterial genome functions. *BMC Res Notes* 2018;**11**:387.
39. Lee J, Lim H, Jang H et al. CRISPR-Cap: multiplexed double-stranded DNA enrichment based on the CRISPR system. *Nucleic Acids Res* 2019;**47**:e1.
40. Kuscic C, Mammadov R, Czikota A et al. Temporal and spatial epigenome editing allows precise gene regulation in mammalian cells. *J Mol Biol* 2019;**431**:111–21.
41. Slesarev A, Viswanathan L, Tang Y et al. CRISPR/CAS9 targeted CAPTURE of mammalian genomic regions for characterization by NGS. *Sci Rep* 2019;**9**:3587.