

Optimize the selection of guide RNA by ChIP to keep CRISPR on-target

Claude Van Campenhout¹, Anne-Clémence Veillard¹, Agnieszka Zelisko-Schmidt¹, Cyril Gueydan², Véronique Kruys², Céline Sabatel¹

¹ Diagenode SA, Liège Science Park, Rue du Bois Saint-Jean 3, 4102 Seraing, Belgium

² Laboratoire de Biologie Moléculaire du Gène, IBMM, Université Libre de Bruxelles, Rue des Professeurs Jeener et Brachet 12, 6041 Gosselies, Belgium

INTRODUCTION

In recent years, CRISPR has evolved from “the curious sequence of unknown biological function” into a highly functional genome editing tool. The technique has recently become extremely popular as it can rapidly modify the genome of a large variety of organisms with unprecedented ease. The CRISPR/Cas9 genome editing system can be used to introduce non-specific insertions and/or deletions (indels) but also specific sequence replacements at any genomic location. In addition, the CRISPR/Cas9 technology is also used as a platform to recruit a variety of functionalities to a genomic region of interest. Introducing two amino acid changes (D10A and H840A) in Cas9 coding sequence result in a nuclease-inactive DNA binding protein named “dead Cas9” (dCas9). Recently, dCas9 variants have been generated that allow repurposing of the system to a variety of applications. Fusing dCas9 to a transcriptional activator or transcriptional repressor, for instance, proved to be a potent way of regulating gene expression. Moreover, dCas9 could be fused to domains that regulate the epigenetic landscape at endogenous loci. Dead Cas9 can also be used to label endogenous loci for live visualization or to edit a single base in the genome (Dominguez et al., 2016).

The mechanisms of target recognition and target specificity of the Cas9 protein is still not completely understood. A major hurdle of this technology is the introduction of double-strand breaks (DSBs) at sites other than the intended on-target site (off-target effects). Several strategies have been developed to mitigate off-target effects including sgRNA truncation, nickases, enzyme engineering or vector-free delivery methods. The binding of dCas9 fusion protein to an off-target site might also lead to unwanted modifications of the cellular physiology. To overcome this specificity issue, an approach is to improve single guide RNAs design (sgRNA). Numerous online tools are available to assist in sgRNA design but the correlation between the predictions and the actual measurements vary considerably since sequence homology alone is not fully predictive of off-target sites (Haeussler et al., 2016).

Together, all CRISPR/Cas9 applications require the verification of the specific binding of the sgRNA at the locus of interest. Chromatin immunoprecipitation followed by real-time PCR (ChIP-qPCR) is a technique of choice for studying protein-DNA interactions. In this study, we show a successful ChIP-qPCR method to verify the binding efficiency of the dCas9/sgRNA complex in the targeted region; and ChIP-seq – to monitor off-target bindings of the dCas9/sgRNA complex in the genome.

METHODS

Cell lines and plasmids

HEK293T cells were cultured in Dulbecco's modified Eagles' medium supplemented with 10% Fetal Bovine Serum and 1x penicillin/streptomycin. The cells were stably transfected with a dCas9 expression plasmid based on the pSpCas9n(BB)-2A-Puro (PX462) V2.0 backbone. The sgRNAs were cloned in the dCas9 expression vector as described in Ran and colleagues (2013). The target sequences are: H19-sgRNA-1 GTCTATCTCTGACAACCCTC, GAPDH-sgRNA-1 GTCTGGCGCCCTCTGGTGGC and GAPDH-sgRNA-2 AAAGACTCGGTCTGGTCT. Non transfected cells were used as negative controls.

Immunolabelling

For imaging, transfected cells (48 hours post-transfection), cells were washed once with phosphate-buffer saline (PBS), fixed with 4% PFA for 5 minutes at room temperature (RT) and washed three times in PBS for 5 minutes each. For immunolabelling, cells were permeabilized with PBS-0.5% Triton X-100 for 15 minutes at RT. Blocking was performed in PBS-5% Calf serum-1% bovine serum albumin for 1 hour at RT. The cells were incubated overnight at 4°C with the anti-Cas9 monoclonal antibody 4G10 (Diagenode Cat. No. C15200216) diluted 1/400 in blocking solution. The next day, cells were washed 3 times 10 minutes with PBS and incubated with secondary anti-mouse antibody coupled to Alexa Fluor 596 diluted 1/800 in blocking solution for 2 hours at RT in the dark. Next, cells were washed 3 times in PBS for 5 minutes each at RT. The nuclei were counterstained with DAPI.

Western blot analyses

HEK293T cells were transfected with plasmids encoding for dCas9-vectors. Cells were harvested and lysed in RIPA Buffer supplemented with Protease Inhibitor Cocktail. Samples were sonicated using a Diagenode Bioruptor® Pico (Cat. No. B01060010) and protein concentration was measured using the Biorad Protein Assay Reagent and an Eppendorf BioPhotometer. The cell lysate was fractionated using 8% SDS polyacrylamide gel. Proteins were transferred to a PVDF membrane using a Biorad system. TBS-T (0.1% tween) with 5% BSA was used for all the blocking and antibody dilutions. Membranes were incubated with the following primary antibodies: mouse monoclonal 4G10 anti-Cas9 antibody (1:5000) or mouse monoclonal anti-GAPDH antibody for 16h at 4°C. After washing, membranes were incubated with an anti-mouse IgG, HRP-conjugated antibody for 2h at RT, then washed three times with TBS-T for 15 min each and finally visualized by ECL Western Blotting Detection Kit (Amersham). Images were captured using a ChemiDoc Touch Imaging System (Bio-Rad).

Chromatin immunoprecipitation (ChIP) and qPCR

The HEK293T cells expressing dCas9 and the different sgRNAs were cultured in the presence of Puromycin (3µg/ml). ChIP assays were performed using the iDeal ChIP-seq Kit for Transcription Factors (Diagenode Cat. No. C01010055). In brief, freshly prepared formaldehyde solution was added

into the cell culture medium (1% final concentration). The culture flasks were subsequently agitated for exactly 10 minutes at RT. Next, glycine was applied to stop the fixation process. After a 5 minute incubation period at RT, the cells were scraped and transferred to a 50-ml tube. The collected cells were subjected to lysis and centrifugation as described in the iDeal ChIP-seq Kit for Transcription Factors manual. The Bioruptor Pico, thermo-controlled sonication device, was used for the shearing of chromatin. Samples were sonicated for 10 cycles of 30'' ON/30'' OFF at 4°C. A total of 4 million cells were used per immunoprecipitation. The ChIP-validated anti-SpCas9 polyclonal antibody (Diagenode Cat. No. C15310258) was used for the ChIP. The negative Ctrl IgG from rabbit was used in parallel (Diagenode Cat. No. C15410206). After washes and reversal of DNA-protein cross-links, the ChIP samples were diluted 10 times and analysed by quantitative real-time PCR (qPCR). The qPCR amplifications were carried out in triplicates for each sample and primer pair. Primers to the GAPDH intron 8 or H19 target sites were used as positive control primers. Human Myt1 and SLC17A4 were used as negative control regions. The qPCR amplifications were performed with the aid of a Roche Light Cycler 96 system. ChIP dCas9 was also performed using the Auto iDeal ChIP-Seq Kit for Transcription Factors (Diagenode, Cat. No. C01010172) and the IP-Star® Compact Automated System (Diagenode Cat. No. B03000002) as described in the manual.

Library preparation, sequencing and data analysis

Illumina compatible libraries were prepared on immunoprecipitated DNA using the MicroPlex v2 library preparation kit according to the manual instructions (Diagenode Cat. No. C05010012). One nanogram of DNA was used per library preparation. The libraries were sequenced on HiSeq 3000 using SE 50pb reads. Cluster generation and sequencing were performed according to the manufacturer's instructions. After trimming using Cutadapt, the reads were mapped to the human genome with the BWA aligner. Subsequent peak calling was performed using MACS.

RESULTS

Locus H19

sgRNA#1
PAM

GAGT**GTCTATCTCTGACAACCTC**GGGAACCAATGCGGCACCTGGCC**TGGGAGACCTGGGACGTTTCTGTGGGTGAACCCCA**
 CTCACAGATAGAGACTGTTGGGAGCCCTTGGTTACGCCGTGGACCGACCCTCTGGACCCTGCAAAGACACCCACTTGGGGT

Locus GAPDH

sgRNA#1
PAM

GGTATGTGGCTGGGGCCAGAGACTGGCTCTTAAAAAGTGCAGG**GTCTGGCGCCCTCTGGTGGCTGG**CTCAGAAAAAGGGCC
 CCATACACCGACCCCGGCTCTGACCGAGAATTTTTCACGTCCAGACCG**GGGAGACCACCGACCGAGTCTTT**TTCCCGG

PAM
sgRNA#2

Figure 1. sgRNA sequences from human H19 and GAPDH loci. sgRNAs sequences are marked in red and PAM are indicated in blue.

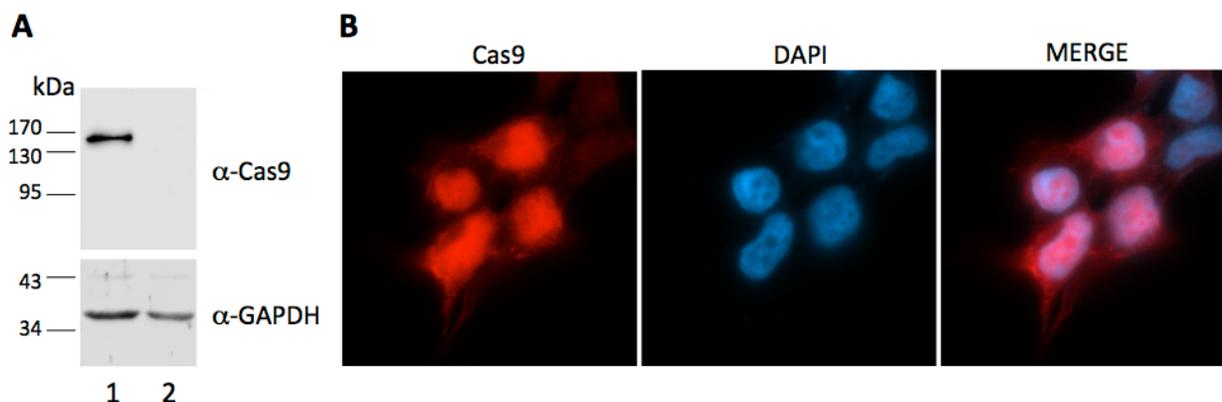


Figure 2. *dCas9* expression can be detected by Western blot and immunofluorescence in transfected HEK293T cells.

- A.** WB was performed on whole protein extracts from HEK293T cells transfected with a *Cas9* expression vector (lane 1) or untransfected cells (lane 2) using Diagenode antibody against *Cas9* (Cat. No. C15200216).
- B.** HEK293T cells were transfected with a *dCas9* expression vector. The cells were stained with the *Cas9* antibody (left) and with DAPI (middle). A merge of the two stainings is shown on the right.

We selected one sgRNA targeting the human H19 locus and two sgRNAs targeting the human GAPDH locus (#1 and #2) with the help of the CRISPR Design tool (<http://crispr.mit.edu/>) (Figure 1). All sgRNAs were scored “high quality” by the algorithm. The different sgRNAs were cloned in the pSpdeadCas9-2A-Puro plasmid and the resulting vectors were used to stably transfect HEK293T cells.

After puromycin selection, we verified the expression of *dCas9* in the cells by Western blot (Figure 2A). Immunofluorescence shows that *dCas9* is detected in the nucleus of HEK293T stable transformants (Figure 2B). These results show that *dCas9* is correctly expressed in the transformed cells and that the protein is found in the nucleus, as expected.

Next, we used the iDeal ChIP-seq Kit for Transcription Factors to validate that the selected guide RNAs could successfully guide *Cas9* to the target site. This kit is ideal for use with low abundance target proteins and provides high yields with excellent specificity and sensitivity. One of the first steps in ChIP is to shear the chromatin, which is one of the most critical steps for a successful ChIP experiment. Ideally, chromatin fragments should range between 100-600 bp. The optimal time of sonication depends on cell type, cell density, sample volume, fixation time, lysis buffer etc. We performed an initial time-course experiment of 8, 10 and 12 sonication cycles with Bioruptor® Pico. We chose 10 sonication cycles as the shortest sonication time resulting in efficient chromatin shearing (data not shown). Over-sonication should be avoided as it may lead to a drop in efficiency in ChIP experiments, especially when non-histone proteins are to be evaluated by ChIP. After the sonication, the fragmented chromatin was subjected to immunoprecipitation with anti-*Cas9* antibody. Using high-quality antibodies is also a critical aspect for successful chromatin immunoprecipitation. Here we used a Sp*Cas9* polyclonal ChIP-grade antibody (Diagenode Cat. No. C15310258) to immunoprecipitate the *dCas9* protein together with the associated chromatin. Subsequently, cross-linking was reversed and DNA was purified from the isolated chromatin to use in qPCR analysis.

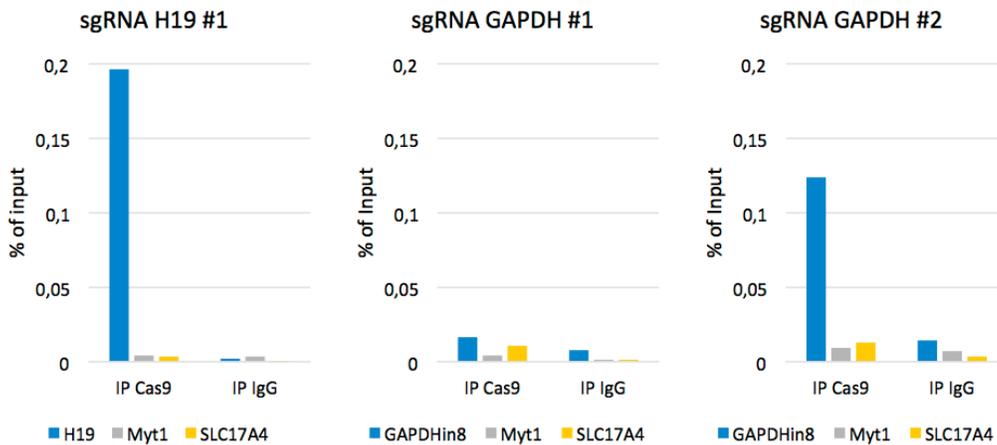


Figure 3. dCas9 ChIP can be used to determine the quality of the sgRNA design

ChIP was performed on sheared chromatin from 4,000,000 HEK293T cells using the iDeal ChIP-seq Kit for Transcription Factors, 5µl of the polyclonal Cas9 antibody and 1µg of the negative control IgG. Primers specific for the human H19, GAPDH, Myt1 and SLC17A4 were used for the qPCR. The figure shows the recovery, expressed as a % of input (the relative amount of immunoprecipitated DNA compared to input DNA after qPCR analysis).

As shown in Figure 3, dCas9 ChIP using sgRNA H19 could successfully isolate the H19 locus (0.2% of the input), whereas the irrelevant MYT1 and SLC7A4 loci were not enriched. One of the two tested sgRNA directed against the GAPDH locus (sgRNA GAPDH #2) was also specifically bound to the GAPDH locus (Figure 3). However, the second sgRNA GAPDH #1 did not allow a specific enrichment of the target region (Figure 3). These results underline the importance of testing each sgRNA for its efficiency. Computational methods alone cannot fully predict whether a sgRNA will bind efficiently and specifically the locus of interest. Interestingly, the same results were obtained with the fully automated IP-Star Compact Automated System combined with the Auto iDeal ChIP-seq kit for Transcription Factors (data not shown). Use of the IP-Star Compact significantly reduce the hands-on time for ChIP-qPCR. The data exhibited high technical reproducibility compared to data from ChIP-qPCR performed manually.

In order to monitor the potential binding of dCas9 to aspecific regions genome-wide, the DNA immunoprecipitated with Cas9 antibody was further analysed in sequencing. As shown in figure 4, H19 sgRNA specifically targets dCas9 to this genomic location without any off-target binding. By contrast, GAPDH sgRNA induces the binding of dCas9 to the GAPDH gene and to a region surrounding the promoter of YIPF4 gene. These results demonstrated the importance of monitoring the binding of dCas9 genome-wide to avoid off-target effects. As depicted on figure 4, similar results were obtained on DNA immunoprecipitated using the IP-Star Compact.

These results indicate that the iDeal ChIP-seq Kit for Transcription Factors combined with the polyclonal anti-Cas9 antibody is sensitive enough to isolate a single genomic locus. In addition, ChIP dCas9 can be used to verify the binding efficiency and specificity of a sgRNA.

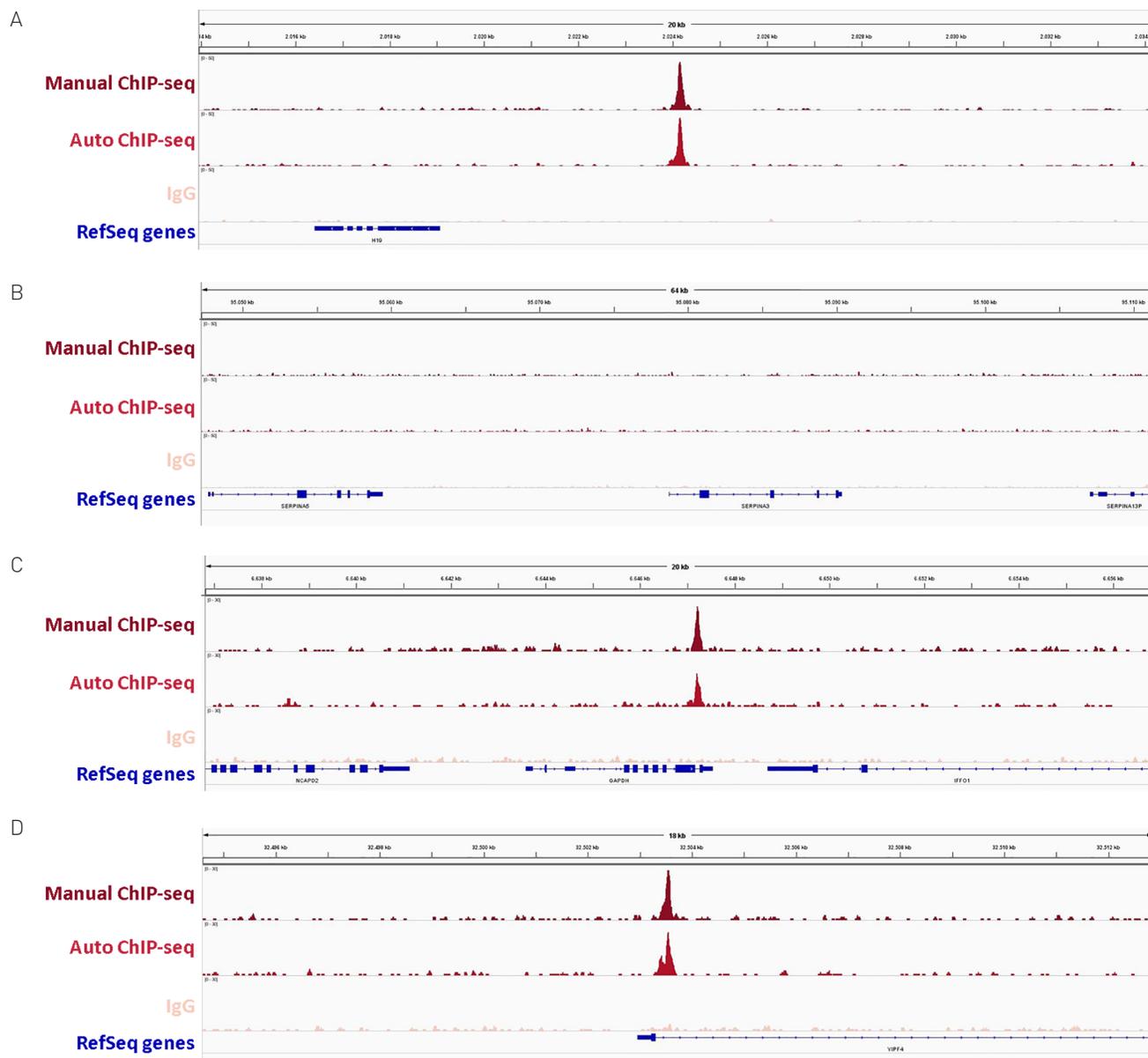


Figure 4. dCas9 ChIP can be used to determine the specificity of the sgRNA design

ChIP was performed on sheared chromatin from 4,000,000 HEK293T cells using the iDeal ChIP-seq Kit for Transcription Factors, 5µl of the polyclonal Cas9 antibody and 1µg of the negative control IgG. The libraries were sequenced on Illumina HiSeq 3000 in 1X50 pb. The picture shows the read distribution for the manual IP (top), the automated IP (middle) and IgG (bottom) samples. A-B. Peaks distribution of the three datasets in the region surrounding H19 (A) and a representative region of the genome (B) for HEK293T cells expressing a sgRNA for H19. C-D. Peaks distribution of the three datasets in the region surrounding GAPDH (C) and YIPF4 (D) for HEK293T cells expressing a sgRNA for GAPDH (sgRNA GAPDH #2).

DISCUSSION

Therapeutic, industrial, and research applications will place high demand on improving the specificity and efficiency of the CRISPR/Cas9 system. Multiple computational tools are available to aid in the prediction and design of sgRNA to target specific loci. Nevertheless, there is still limited predictability of whether the Cas9 and sgRNA of interest will be able to bind specifically the genome. The underlying reasons for this variation in efficiency are still limited but some best-practices may increase the chances of a successful targeting. While the activity of Cas9 can be studied using various techniques based on indels detection, these methods are not suitable for dCas9 applications. We showed here that the iDeal ChIP-seq Kit for Transcription Factors provides a robust ChIP protocol suitable for the investigation of dCas9 binding in the genome. A specific enrichment of dCas9 in the on-target region can be verified by ChIP followed by real time PCR assay. This information is essential to verify the correct binding of dCas9 fused to an effector (Neguembor et al., 2017).

To detect sgRNA-mediated binding of the Cas9 protein at genomic regions, DNA fragments isolated by ChIP can be sequenced and mapped to their originating locations in the genome. The number of off-target sites bound by dCas9 varies from ~10 to >1000, depending on the sgRNA (Kuscu et al., 2014; Wu et al., 2014). ChIP-seq is a powerful approach to survey off-target binding for a multitude of sgRNAs due to its rapidity, reduced sequencing cost and high coverage. As no method can guarantee the complete detection of off-target cleavages, it is ideal to combine multiple approaches. Using both genome-wide ChIP-seq dCas9 binding analysis and the sequence-base *in silico* prediction can be combined to efficiently identify off-target sites (O'Geen et al., 2015). Here, we show also that the ChIP dCas9 can be automated using the auto iDeal ChIP-Seq Kit for Transcription Factors and the automated IP-Star Compact Automated System. Overall, the ChIP dCas9 application supports the discovery of the mechanisms governing dCas9 binding specificity and the optimization of CRISPR experiments.

REFERENCES

1. Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol.* 2016 Jan;17(1):5-15.
2. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, Joly JS, Concordet JP. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 2016 Jul 5;17(1):148.
3. Kescu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol.* 2014 Jul;32(7):677-83.
4. Neguembor MV, Sebastian-Perez R, Aulicino F, Gomez-Garcia PA, Cosma MP, Lakadamyali M. (Po) STAC (Polycistronic SunTA modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. *Nucleic Acids Res.* 2017 Dec 23.
5. O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* 2015 Mar 31;43(6):3389-404.
6. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 2013 Nov;8(11):2281-2308.
7. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, Jaenisch R, Zhang F, Sharp PA. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol.* 2014 Jul;32(7):670-6.