bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Precise and Predictable DNA Fragment Editing Reveals Principles of Cas9-mediated Nucleotide Insertion

Jia Shou[1,2,3,#], Jinhuan Li[1,2,3,#], Qiang Wu[1,2,3,*]

[1]Center for Comparative Biomedicine, MOE Key Laboratory of Systems Biomedicine, Institute of Systems Biomedicine, SCSB, Shanghai Jiao Tong University (SJTU), Shanghai, 200240, China

[2]State Key Laboratory of Oncogenes and Related Genes, SJTU Medical School, Shanghai, 200240, China

[3]School of Life Sciences and Biotechnology, SJTU, Shanghai, 200240, China

[#]Co-first author

[*]Correspondence: qwu123@gmail.com (Q.W.)

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Summary

DNA fragment editing (DFE) or chromosomal engineering including inversions, deletions, and duplications by Cas9 with paired sgRNAs are important to investigate structural genome variations and developmental gene regulation, but little is known about the underlying mechanisms. Here we report that debilitating CtIP, which is thought to function in NHEJ, enhances precise DNA fragment deletion. By analyzing the inserted nucleotides at the junctions of DNA fragment inversions, deletions, and duplications, we find that Cas9 cleaves the noncomplementary strand with a flexible profile upstream of the PAM site and rationally-designed Cas9 nucleases have distinct cleavage profiles. Finally, Cas9-mediated nucleotide insertions of DFE are nonrandom and are equal to the combined sequences upstream of both PAM sites with predicted frequencies. Thus, precise and predictable DFEs could be achieved by perturbing DNA repair genes and using appropriate PAM configurations. These findings have important applications regarding 3D chromatin folding and enhancer insulation during gene regulation.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Organisms defend themselves against virus infection with adaptive immunity and maintain their genome integrity against genetic damage with DNA repair machinery. In bacteria and archaea, for example, clustered regularly interspaced short palindromic repeats (CRISPR) are a natural adaptive immune system against virus infection and phage conjugation (Deltcheva et al., 2011; Jiang et al., 2016; Jinek et al., 2012; Makarova et al., 2015; Marraffini and Sontheimer, 2010; Sternberg et al., 2015). For the type II system, a single endonuclease of the CRISPR associated protein (Cas9) is programed to cleave the invading viral genomes by crRNA (CRISPR RNA) and tracrRNA (trans-activating crRNA), generating blunt-ended double-strand breaks (DSBs) 3 base pairs upstream (-3 bp) of PAM (protospace adjacent motif, NGG for Cas9 from *Streptococcus pyogenes*) by the HNH and RuvC nuclease domains for the complementary and noncomplementary strands, respectively (Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012; Makarova et al., 2015). In particular, the R-loop locking of HNH by guide RNA enables the cutting of the complementary strand at the exact -3 position upstream of PAM (Jiang et al., 2016; Sternberg et al., 2015).

In eukaryotes, Cas9 can be reprogramed for DNA-fragment editing (DFE) by paired sgRNAs to generate four DSB ends, which are repaired by several competing end-joining (EJ) pathways including canonical or classic non-homologous EJ (cNHEJ), alternative NHEJ (alt-NHEJ or MMEJ), and

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

homologous recombination (Bétermier et al., 2014; Bhargava et al., 2017; Jasin and Haber, 2016), resulting in large DNA-fragment deletions and inversions as well as duplications (Bhargava et al., 2017; Canver et al., 2014; Huang and Wu, 2016; Kraft et al., 2015; Li et al., 2015a). This DFE method of chromosomal rearrangement or genome engineering is useful to probe 3D chromatin architecture and gene regulation; however, the underlying mechanisms are largely unknown (Dekker and Mirny, 2016; Franke et al., 2016; Guo et al., 2015; Huang and Wu, 2016; Lupiáñez et al., 2015).

Here we report that debilitating CtIP enhances precise DNA fragment deletion (DFD). In addition, in contrast to cleavage by HNH at the exact -3 position of the complementary strand (Jiang et al., 2016; Sternberg et al., 2015), we found that the cleavage by RuvC on the noncomplementary strand is flexible. Specifically, RuvC cleaves the noncomplementary strand at -3 as well as further upstream positions of PAM *in vivo*, resulting in blunt as well as non-blunt DSB ends with 5' overhang. Through rational design of a set of engineered Cas9 nucleases, we achieved predictable DNA fragment inversion (DFI) *in vivo* based on the scissile profiles of RuvC. Finally, with this, we dissected the orientation and function of a composite enhancer containing two CBSs (CTCF-binding sites), revealing that one single CBS at topological domain boundary functions as an insulator to ensure specific long-distance chromatin-looping interactions between the distal enhancer and its target promoter. Thus, our mechanistic insights of Cas9 cleavage and of subsequent DSB-repair events

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

should have broad implications for precise and predictable editing of millions of noncoding elements for 3D genome architecture and chromosomal rearrangement in genetic diseases.

Results

The Role of CtIP in Precise DFD Revealed by Screening DNA Repair Genes

DFD by Cas9 with paired sgRNAs has the advantage of avoiding single-sgRNA-guided Cas9 re-cutting (Paquet et al., 2016). Indeed, we observed many more precise junctions of DFD guided by paired sgRNAs (Guo et al., 2015; Li et al., 2015a). To further increase the efficiency of precise DFD, we screened a panel of seven DNA-repair genes and found that debilitating CtIP (C-terminal binding protein[CTBP]-interacting protein, also known as RBBP8: RB-binding protein 8, or Sae2) results in a significant increase of precise DFD (Figure 1). Specifically, we knocked out each of the seven DNA-repair genes and found, through next generation sequencing (NGS), that precise DFDs at the protocadherin (*Pcdh*) locus are significantly increased with CtIP debilitation (Figure 1A).

To see whether this is true for other loci, we screened the same panel of DNA-repair genes for precise DFD in the *β-globin* locus. Indeed, CtIP deficiency also enhances the efficiency of the *β-globin* precise DFD (Figure 1B). In addition, we found a corresponding decrease of small deletions and no significant

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alteration of insertions at the DFD junctions (Figures 1C-1D and S1A-S1B). We confirmed these NGS results by cloning and Sanger sequencing (Figures 1E-1H). Finally, CtIP debilitation results in no significant changes at the junctions of DNA fragment inversion or duplication (Figures S1C-S1H), suggesting that the mechanism of DFD is different from that of DNA-fragment inversion and duplication (*trans*-allelic translocation) (Ghezraoui et al., 2014).

We next investigated four combinatorial CRISPR PAM configurations with paired sgRNAs for distinct DNA fragments of different sizes in the *β-globin* and *HoxD* loci. We found that CtIP deficiency results in a significant increase of precise ligations and a corresponding decrease of small deletions at the DFD junctions, but no significant changes in the junctions of DNA fragment inversion and duplications in all four PAM configurations (Figures 1I-1J and S1I-S1L).

Confirmation the role of CtIP in DFD by Triapine and CRISPR Single Cell Clones

To confirm the role of CtIP in DFD, we used triapine (3AP) to block CtIP activity and found that 3AP increases precise DFD and decreases the percentages of small deletions but no alteration of insertions for both the *Pcdh* and *β-globin* loci (Figures 2A-2D and S2A-S2B). However, there are no significant alterations in the junctions of DNA fragment inversion and duplication (Figures S2C-S2H).

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, we screened for CtIP-mutant CRISPR cell clones and measured the precise DFD efficiency at *Pcdh*, *β-globin*, and its *LCR* (locus control region) by NGS. We obtained two CtIP-mutant CRISPR cell clones (C14 and C27) and found both of them display a significantly higher efficiency of precise DFD and a corresponding decrease of small deletions at DFD junctions (Figures 2E and 2F). However, these two cell clones display no significant alterations in the junctions of DNA fragment inversion and duplication (Figures S2I-S2N). Together, these data demonstrate that debilitating CtIP enhances precise DFD. Since DFD results from two competing DNA-repair pathways (cNHEJ and alt-NHEJ) (Bhargava et al., 2017) and CtIP facilitates or participates in strand resection in alt-NHEJ (Sartori et al., 2007), this suggests that cNHEJ is an error-free DSB repair pathway (Bétermier et al., 2014).

Non-blunted end cleavage by Cas9

Taking advantage of the absence of Cas9 re-cutting in DFE (Paquet et al., 2016), in conjunction with NGS, we next investigated the EJ-repair junctional patterns of deletion, duplication, and inversion of a *Pcdh* DNA fragment (Figure 3). We first analyzed the pattern of indels at DFD junctions and found a striking pattern of nucleotide insertions (Figure 3A). Specifically, at the first position of insertions, there is a “G” nucleotide with a 99.11% frequency (Figure 3A). By contrast, there is a “T” with a frequency of only 0.89% at this position with no “C” or “A” at all. At the second position of insertions, there is 100% “G” with no

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

insertion of “T”, “C”, or “A” (Figure 3A). Remarkably, there is an “A”, “C”, “T”, “G”, or “C” nucleotide each with a 100% frequency at the 3rd, 4th, 5th, 6th, or 7th position of insertions, respectively, with no other nucleotides at all (Figure 3A). Thus, the nucleotide insertions at each position are nonrandom and may reflect the cleavage pattern of RuvC on the noncomplementary strand. Indeed, the actual inserted nucleotides of “+G”, “+GG”, “+AGG”, “+CAGG”, “+TCAGG”, “+GTCAGG”, and “+CGTCAGG” all match perfectly with the nucleotide sequences of the noncomplementary strand at the -4, -5, -6, -7, -8, -9, and -10 positions upstream of the second PAM site, respectively (Figure 3A). These observations strongly suggest that the inserted nucleotides at the DFD junctions result from the flexible cleavage of RuvC of the noncomplementary strand at the -4, -5, -6, -7, -8, -9, and -10 positions.

We then analyzed nucleotide insertions at the junctions of DNA-fragment duplications and found a pattern similar to that of DFD. Not only is there a striking nucleotide bias of “T”, “C”, or “A” with a frequency of 100%, 59.63%, or 100% at the 1st, 2nd, or 6th position of insertions, respectively; but the inserted nucleotides of “+T”, “+CT”, and “+AAGGCT” also match perfectly with the nucleotide sequences of the noncomplementary strand upstream of the first PAM site (Figure 3B). We concluded that the inserted nucleotides at the junctions of DNA-fragment duplication result from the flexible cleavage of RuvC of the noncomplementary strand at the -4, -5, or -9 position upstream of the first PAM site. Although there is no specific nucleotide-insertion pattern at the

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

upstream DFI junctions, we found a remarkable pattern of nucleotide insertions at the downstream DFI junctions (Figures 3C and 3D). Namely, the inserted nucleotides match the combined sequences of the nucleotide insertions at junctions of both DNA-fragment deletion and duplication (Figure 3D).

Finally, we edited four additional DNA fragments of different sizes ranging from ~700 bp to ~930 kb in distinct loci and found remarkably-similar cleavage patterns (Figures S3 and S4). Taken together, these data strongly suggest that, in addition to cleavage of the noncomplementary strand at the -3 position upstream of the PAM site, RuvC also cleaves of the noncomplementary strand at positions further upstream *in vivo*.

Altered Cleavage Profiles of Engineered Cas9 Nucleases by Rational Design

To confirm the non-blunted cutting pattern of Cas9 *in vivo*, we used a structure-based rational design to alter the cleavage pattern of the RuvC domain. Recent structural and molecular studies revealed that allosteric conformational changes between RuvC and HNH for R-loop formation are essential for Cas9 catalytic activation and that the two linker regions between these domains are mostly disordered or helical (Anders et al., 2014; Jiang et al., 2016; Jinek et al., 2014; Nishimasu et al., 2014; Sternberg et al., 2015). We reasoned that, by perturbing these two linker regions and carefully preserving the catalytic activity

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of RuvC and HNH, we may be able to design an engineered Cas9 with altered scissile patterns of RuvC on the noncomplementary strand.

We screened twenty engineered Cas9 nucleases with mutations of residues in the two linker regions and found seven of them with altered scissile profiles for each of the ten sgRNAs we tested (Figures 4A and 4B). We then calculated the scissile profile for each engineered nuclease with these ten sgRNAs and found that each has a distinct scissile profile (Figure 4C). Finally, we tested these engineered nucleases in our CtIP CRISPR cell clone and found that CtIP deficiency results in a significant increase of the percentage of precise DFD and a concomitant decrease of small deletions for each rationally-designed Cas9 nucleases (Figures 4D-4G). These data suggest that the cleavage profile of Cas9 could be altered by inducing a conformational change through engineering the linker regions between the HNH and RuvC domains.

Predictable DFE

Deep sequencing of duplication junctions revealed that all of them except K918A result in a decreased frequency of cleavage of the noncomplementary strand at the -3 position (precise) and that four of them have an increased frequency of cleavage at the -4 position (insertion of “+C”) upstream of the first PAM site (Figures 5A and 5B). In addition, most of them also have an increased frequency of cleavage at the -5 position (insertion of “+GC”) (Figure 5B).

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Deep sequencing of deletion junctions revealed that all of the seven engineered Cas9 nucleases have a decreased frequency of cleavage at the -3 position and the scissile profiles at -4, -5, and -6 positions upstream of the second PAM site are also altered (Figure 5C).

Remarkably, nucleotide insertions at the downstream junctions of DFI perfectly match the combined insertions at the junctions of both DNA-fragment deletion and duplication with the expected frequency (Figure 5D). These observations further support that Cas9 generates both blunt and non-blunt DSB ends with a 5' overhang *in vivo*. Indeed, we could predict the frequency of inserted nucleotides at downstream inversion junctions for all of the seven engineered Cas9 nucleases (Figure 5D).

Finally, we edited two additional DNA fragments of different sizes ranging from ~700 bp to ~930 kb and found the same remarkable principle of Cas9-mediated nucleotide insertions (Figures S5 and S6). Taken together, we conclude that precise and predictable Cas9-mediated DFEs can be achieved by engineered Cas9 with paired sgRNAs.

Single CBS sites at Topological Domain Boundaries Determine Chromatin Looping Directions

This precise and predictable DFE method allows a detailed investigation of the role of enhancer and insulator orientation in 3D genome architecture and gene

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulation. The location and relative orientation of CBSs genome-wide determine the 3D chromosomal architecture (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015). The human *Pcdh* locus comprise 53 tandem genes organized into three sequentially-linked clusters (*Pcdh*α, β, and γ) (Wu and Maniatis, 1999) and two (*Pcdh*α and *Pcdh*$\beta\gamma$) CTCF/cohesin-mediated chromatin contact domains (CCDs or subTADs) (Guo et al., 2015). These clustered *Pcdh* genes are crucial for dendritic self-avoidance and tiling during brain development (Lefebvre et al., 2012). The orientation and location of a repertoire of *Pcdh* CBSs determine the directional looping between the distal *HS5-1* enhancer and its target promoters in the *Pcdh*α variable region (Guo et al., 2015). This composite enhancer comprises two tandem CBSs (*HS5-1a* and *HS5-1b*) both in reverse orientation and numerous TF binding sites that are clustered in a 1-kb region (Figure 6A) (Guo et al., 2015).

To date, there is no evidence that the relative orientation of single CBS sites can determine chromatin-looping directions between enhancers and promoters (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015). To this end, we designed a series of paired sgRNAs to dissect the *HS5-1* enhancer architecture (Figure 6A). We screened single-cell CRISPR clones for inversion of the single CBS *HS5-1a* or *HS5-1b* site as well as its combination with the middle region of the *HS5-1* enhancer. Remarkably, inversion of *HS5-1b* alone or with the middle region results in a significant decrease of DNA-looping interactions between the *HS5-1* enhancer and the *Pcdh*α promoters, and a corresponding

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

increase of DNA-looping interactions with the *Pcdhβγ* promoters (Figure 6B). However, inversion of *HS5-1a* alone or combined with the middle region results in no significant alteration of directional DNA-looping interactions with either *Pcdhα* or *Pcdhβγ* promoters (Figure 6C). Thus, the relative orientation of the single CBS *HS5-1b* at the *Pcdhα* CCD or subTAD boundary determines the chromatin-looping directions between the distal enhancer and its target promoters.

To assess the functional role of the altered chromatin looping, we performed the RNA-seq experiments and found that each CRISPR DFI cell line with disrupted architecture of the *HS5-1* enhancer displays altered patterns of *Pcdh* gene expression, consistent with previous studies that subtle structural changes could result in alternations of gene expression pattern (Figure 6D) (de Laat and Duboule, 2013). Finally, a series of CRISPR inversions of progressive numbers of CBSs in the *β-globin* locus demonstrate that only those inversions covering the CBS at a TAD boundary (CBS15) switch chromatin-looping directions (Figure S7). We conclude that single CBS sites at either TAD or SubTAD boundaries function as insulators.

Discussion

Chromosomal engineering or large genomic fragment editing, including deletions, inversions, duplications, and translocations, could be achieved by

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using the conventional Cre-LoxP system with low efficiencies (Wu et al., 2007; Zheng et al., 2000). Recent development in genome editing by ZFN, TALEN, and CRISPR enabled rapid technical advance for chromosomal rearrangement or segmental editing (Blasco et al., 2014; Canver et al., 2014; Choi and Meyerson, 2014; Cong et al., 2013; Gupta et al., 2013; Kraft et al., 2015; Lee et al., 2012; Li et al., 2015a; Park et al., 2016; Torres et al., 2014; Xiao et al., 2013; Yang et al., 2013). In particular, DFE has been used to investigate developmental gene regulation and to model human genetic diseases (Franke et al., 2016; Guo et al., 2015; Lupiáñez et al., 2015; Maddalo et al., 2014; Park et al., 2015; Tai et al., 2016); however, the underlying mechanism for CRISPR DFE remains unknown and the nucleotide insertions at the junctions are assumed to be random. We found here that CtIP deficiency increases the frequency of precise DFD by Cas9 with paired sgRNAs *in vivo*. In addition, we found that CRISPR/Cas9-mediated nucleotide insertions are nonrandom and can be predicted based on the scissile profiles of RuvC on the noncomplementary strand. Our rational design and *in vivo* cleavage data are consistent with the allosteric conformation change between the HNH and RuvC domains for catalytic activation of Cas9 nucleases (Jiang et al., 2016; Sternberg et al., 2015). Finally, we found, through dissection of a composite enhancer, that single CBS sites at topological boundaries function as an insulator.

Based on our findings that RuvC cleaves the noncomplementary strand further upstream of PAM (Figures 3-4 and S3-S4) and that Cas9-mediated

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nucleotide insertions at the downstream DFI or duplication junctions are equal to the combined sequences with expected frequencies upstream of both PAM sites for the PAM configurations of NGG-NGG or NGG-CCN, respectively (Figures 5 and S5-S6); we propose that Cas9-mediated nucleotide insertions could be achieved for the other two PAM combinations of CCN-NGG and CCN-CCN (Figure 7). In addition, these inserted nucleotide sequences are generated by filling in of nucleotides through 5'-3' extension by unknown DNA polymerases and subsequent precise ligation by DNA ligases during DNA DSB repair (Figure 7). Through these mechanisms, we could achieve predictable DFE for all four PAM configurations.

Our data suggest that, consistent with previous *in vitro* evidence (Jinek et al., 2012; Li et al., 2015b), Cas9 cleaves the dsDNA target to generate blunt as well as previously unrealized non-blunt ends with 5' overhangs *in vivo*, revealing a striking similarity of staggered cutting patterns by the evolutionarily-related class 2 CRISPR effector proteins of Cas9, Cpf1, and C2c1 (Makarova et al., 2015; Yang et al., 2016; Zetsche et al., 2015). Furthermore, predictable insertions at DFI junctions based on the scissile profiles of RuvC of engineered nucleases on the noncomplementary strand reveal a general principle of Cas9-mediated nucleotide insertions. These findings on Cas9 cleavage and DSB-repair outcomes lay a foundation to improve DFE for probing 3D genome architecture and gene regulation as well as for investigating mechanisms of DNA damage response pathways.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Cell culture. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (HyClone) containing 10% FBS (Gibco) and 1% penicillin–streptomycin (Gibco). HEC-1-B cells were cultured in Eagle's Minimum Essential Medium (Gibco) containing 10% FBS (Gibco), 1% penicillin–streptomycin (Gibco), 2 mM glutamine (Gibco), and 1 mM sodium pyruvate (Sigma). All cells were cultured at 37°C and 5% (v/v) CO_2. Cells were plated at a density of approximately 4×10^5 cells per well in 12-well plates and cultured for 24 hours for transfection.

DNA-repair genes screening for precise DFD. First, plasmids for paired sgRNAs targeting the exons of each repair gene of *CtIP*, *Mre11*, *Rad50*, *BRCA1*, *PARP1*, *Exo1*, and *XRCC1* were designed (Supplemental Table S1) and constructed as described previously (Li et al., 2015a). In addition, a pair of sgRNAs targeting specific DNA fragments of the *Pcdh* as well as the *β-globin* loci (Supplemental Table S1) were also generated as described previously (Li et al., 2015a). HEK293T cells were transfected with plasmids encoding SpCas9 (0.8 μg), two sgRNAs targeting specific DNA fragment (each 0.6 μg) and two sgRNAs targeting the exons of each repair gene (each 0.6 μg) by Lipofectamine 2000 (Life Technology) in 12-well plates. The transfected cells were cultured for 48 hours and then harvested. Genomic DNA was extracted by Genomic DNA Purification kit (Promega) for analyses.

NGS for DFD junctions. The *Pcdh* and *β-globin* DFD junctions were amplified

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the high-fidelity DNA polymerase (KOD-Plus-Neo, TOYOBO) by specific primer pairs with P5 and P7 Illumina adapters (Supplemental Table S1). PCR products were purified by the High-Pure Purification Kit (Roche). Libraries were barcoded and sequenced on the Hiseq X-Ten platform. Reads of each sample were de-multiplexed and adaptor sequences were clipped. Reads were mapped to reference sequences of precise DFD junctions (Precise junctions are defined as direct ligations between two blunt DSB ends generated by Cas9 cutting at -3 bp upstream of the PAM site). Indels (insertions and deletions) of junctions were called by the Varscan2 program (Koboldt et al., 2012). The percentages of precise DFD were calculated. A total of 328 Illumina libraries were sequenced for analyzing the role of CtIP in precise DFD.

TA cloning. TA cloning was used to confirm the results of NGS. Briefly, PCR products with a single adenosine attached to the 3'-ends were amplified by the Taq DNA polymerase and directly ligated into pGEM-T easy vector (Promega) by T4 DNA ligase. The ligated products were transformed with Stbl3 (F^- *mcr*B*mrrhsd*S20(r_B^-,m_B^-)*rec*A13*sup*E44*ara*-14*gal*K2*lac*Y1*pro*A2*rps*L20(Str^R) *xyl*-5λ*leumtl*-1) competent cells. About 30 clones for each DNA-repair gene were analyzed by Sanger sequencing (Figure 1). The results of TA-cloning were consistent with NGS. Thus, all subsequent experiments were analyzed by NGS.

PAM configuration. A pair of plasmids encoding paired sgRNAs for each PAM configuration targeting *HoxD* or *β-globin* DNA fragments (Supplemental Table

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S1) were generated as described previously (Li et al., 2015a). After transfection by Lipofectamine 2000 (Life Technology), genomic DNA was extracted and analyzed by NGS for precise DFD.

Drug inhibition of the CtIP activity. Triapine (3-aminopyridine-2-carboxaldehyde thiosemicarbazone, 3AP) is a small-molecule inhibitor that was used to block CtIP phosphorylation (Lin et al., 2014). HEK293T cells were treated with 3AP at the final concentrations of 0.2 μM, 0.4 μM, and 0.8 μM at the time of plasmid transfection and continued culturing for 24 hours. The DFD junctions were then analyzed by NGS.

Single cell screening of *CtIP* mutant by CRISPR. Generation of CRISPR single-cell clones of *CtIP* mutants were as previously described (Guo et al., 2015; Li et al., 2015a). Briefly, HEK293T cells were transfected with Cas9 and two sgRNAs targeting the exons of *CtIP*. After 2 day of transfection, puromycin (2 mg/ml) was added and cells were continued to culture for additional 4 days. The cells were then maintained in the puromycin-free medium for another 8 days and harvested to seed in 96-well plates at an approximately density of one cell per well. The single cells were cultured for another 10 days to grow to single-cell colonies. Single-cell clones were then screened to obtain the *CtIP* mutant cell lines. Of 96 single-cell clones picked, two CtIP mutant clones (C14 and C27) were identified. Precise DFD was then tested in these CtIP mutant clones.

Rational design of Cas9 nuclease. We reasoned that mutating the two linker

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regions of Cas9 between HNH and RuvC catalytic domains while keeping their catalytic center unchanged may result in alteration of the scissile profile. Engineered Cas9 nucleases with substitution or deletion of specific residues within the two linker regions were rationally designed based on the crystal structures of SpCas9 (PDB accession number: 5F9R). Briefly, primer pairs for cloning engineered Cas9 nucleases (Supplemental Table 1) were designed by either substituting or deleting the codon for a specific residue in the two linker regions. Wild-type SpCas9 plasmids were used as templates to amplify mutant DNA fragments with the Q5 polymerase (NEB). The SpCas9 templates were then digested by DpnI. The amplified PCR products were ligated with T4 DNA ligase (NEB) at room temperature for 5 minutes and were transformed with the Stbl3 competent cells. In total we designed 20 Cas9 mutants and found 7 of them with altered scissile profiles. All cloned plasmids encoding engineered Cas9 nucleases were confirmed by Sanger sequencing.

Single-cell screening of DNA fragment inversions by CRISPR. CRISPR single-cell inversion clones were screened as previously published (Guo et al., 2015; Li et al., 2015a). For the *Pcdh* locus, we obtained 2 (A15 and A23), 1 (E85), 2 (C17 and C20), and 2 (D8 and D67) inversion HEC-1-B clones of CBS *HS5-1b* and *HS5-1a* as well as their combination with the enhancer middle regions from 44, 85, 67, and 77 single-cell clones, respectively. For the *β-globin* locus, we obtained 3 (A3, A29 and A49), 1 (B36), 2 (E19 and E37) inversion HEK293T clones of CBS15, CBS14-15, and CBS13-14 from 49, 40, and 40

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

single-cell clones, respectively.

Circularized chromosome conformation capture (4C). The libraries for 4C-seq were generated as described previously (Guo et al., 2012; Guo et al., 2015; Hagège et al., 2007; Splinter et al., 2012). Briefly, a total of 10^7 cells of HEK293T cells or HEC-1-B cells were cross-linked by formaldehyde and lysed to expose cell nucleus. The cross-linked nuclei were digested with HindIII or EcoRI while rotating at 900 rpm overnight and then ligated with T4 DNA ligase (NEB). After reverse cross-linking, the purified genomic DNA was digested with a 4-bp cutter (DpnII or NlaIII) and were ligated again. The re-ligated products were purified by the High Pure DNA Purification kit (Roche). Finally, 4C-seq libraries were amplified by inverse PCR using the high-fidelity DNA polymerase by a pair of primers each with a P5 or P7 Illumina adapter (Supplemental Table S1). The amplified libraries were sequenced on a Hiseq X-Ten platform. Reads were mapped to the human genome (hg19) using the Bowtie2 program (Langmead and Salzberg, 2012). The r3Cseq program in the R/Bioconductor package (Thongjuea et al., 2013) was used to analyze the long-range chromatin-looping interactions. All 4C-seq experiments were performed with replications.

RNA-seq. RNA-seq experiments were performed as previously described (Guo et al., 2015). Briefly, about 4×10^5 cells were used for each RNA-seq. Total RNA were prepared using an RNeasy kit (Qiagen). Polyadenylated mRNAs were then selected by the oligo dT beads (NEB) for library construction. After reverse transcription of polyA mRNAs with a polyT primer, the cDNA libraries

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were barcoded and sequenced on the Illumina Hiseq X ten platform. Specific index primers (Supplemental Table S1) were used to multiplex RNA-seq samples. Sequenced reads were align to the human genome (GRCh37) using the TopHat software (v2.0.14) (Trapnell et al., 2009) with the parameters of -N 0 –g 1 –x 1. The expression levels of each transcript were calculated using the Cufflinks software (v2.2.1) (Trapnell et al., 2010) with the default parameters. The heatmap was generated using the R program to display expression levels of members of the *Pcdh* gene clusters.

NGS and statistical analyses. All experiments for NGS were performed with biological replicates. In total, 1024 libraries were constructed and sequenced by NGS for analyzing of DFE junctional patterns. For CtIP experiments, *P*-values were calculated using the one-tailed Student's *t*-test. For predictable DFE junctions, we assume that the probability of ligations of any two DSB ends by DNA-repair machineries are equal. Thus, the ratio of the DNA repair events for DFE junctions is equal to the product of the percentages of two DSB ends to be ligated. Pearson's product-moment correlation coefficients (PPMCC) was used to compare sets of nucleotide-insertion frequencies of engineered Cas9 nucleases from NGS with corresponding sets of predictions that are equal to the product of multiplication of the normalized insertion frequencies. The correlation coefficients and *P* values were generated by the R program.

Cas9 structural analysis. The structural representation in Figure 4A was generated by PyMOL (http://www.pymol.org/). The coordinates of the structure

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of SpCas9 with sgRNA and dsDNA is from PDB accession number 5F9R (Jiang et al., 2016). We removed the recognition domain and PAM-interaction domain for clarity. The HNH and RuvC nuclease domains are shown in surface representation. The two linker regions are shown in ribbon. The side chains of mutated or deleted residues are shown and labelled with single-letter amino acid code.

Author Contributions Q.W. conceived the experimental design and wrote the manuscript with the help of S.J and J.L., who also performed all the experiments.

Acknowledgments We thank Profs Dan Czajkowsky and Yanli Wang for critical reading of the manuscript and all members of the Wu lab for discussion. This work was supported by grants from NSFC (31630039, 91640118, and 31470820).

References

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Bétermier, M., Bertrand, P., and Lopez, B.S. (2014). Is non-homologous end-joining really an inherently error-prone process? PLoS Genet. *10*, e1004086.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bhargava, R., Carson, C.R., Lee, G., and Stark, J.M. (2017). Contribution of canonical nonhomologous end joining to chromosomal rearrangements is enhanced by ATM kinase deficiency. Proc. Natl Acad. Sci. USA *114*, 728-733.

Blasco, R.B., Karaca, E., Ambrogio, C., Cheong, T.C., Karayol, E., Minero, V.G., Voena, C., and Chiarle, R. (2014). Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology. Cell Rep. *9*, 1219-1227.

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. *289*, 21312-21324.

Choi, P.S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. Nat. Commun. *5*, 3728.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499-506.

de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. Mol. Cell *60*, 676-684.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dekker, J., and Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. Cell *164*, 1110-1121.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., Kraft, K., Kempfer, R., Jerkovic, I., Chan, W.L.*, et al.* (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265-269.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl Acad. Sci. USA *109*, E2579-2586.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C.*, et al.* (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol. Cell *55*, 829-842.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. Proc. Natl Acad. Sci. USA *109*, 21081-21086.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., *et al.* (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900-910.

Gupta, A., Hall, V.L., Kok, F.O., Shin, M., McNulty, J.C., Lawson, N.D., and Wolfe, S.A. (2013). Targeted chromosomal deletions and inversions in zebrafish. Genome Res. *23*, 1008-1017.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. *2*, 1722-1733.

Huang, H., and Wu, Q. (2016). CRISPR double cutting through the labyrinthine architecture of 3D genomes. J. Genet. Genomics *43*, 273-288.

Jasin, M., and Haber, J.E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA repair *44*, 6-16.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., *et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568-576.

Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., *et al.* (2015). Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. Cell Rep. 10, 833-839.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359.

Lee, H.J., Kweon, J., Kim, E., Kim, S., and Kim, J.S. (2012). Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. Genome Res. *22*, 539-548.

Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T., and Sanes, J.R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. Nature *488*, 517-521.

Li, J., Shou, J., Guo, Y., Tang, Y., Wu, Y., Jia, Z., Zhai, Y., Chen, Z., Xu, Q., and Wu, Q. (2015a). Efficient inversions and duplications of mammalian

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulatory DNA elements and gene clusters by CRISPR/Cas9. J. Mol. Cell Biol. *7*, 284-298.

Li, Y., Park, A.I., Mou, H., Colpan, C., Bizhanova, A., Akama-Garren, E., Joshi, N., Hendrickson, E.A., Feldser, D., Yin, H.*, et al.* (2015b). A versatile reporter system for CRISPR-mediated chromosomal rearrangements. Genome Biol. *16*, 111.

Lin, Z.P., Ratner, E.S., Whicker, M.E., Lee, Y., and Sartorelli, A.C. (2014). Triapine disrupts CtIP-mediated homologous recombination repair and sensitizes ovarian cancer cells to PARP and topoisomerase inhibitors. Molecular cancer research : MCR *12*, 381-393.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E.*, et al.* (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722-736.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181-190.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125-129.

Park, C.Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.H., Kim, D.W., and Kim, J.S. (2015). Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. Cell Stem Cell *17*, 213-220.

Park, C.Y., Sung, J.J., Choi, S.H., Lee, D.R., Park, I.H., and Kim, D.W. (2016). Modeling and correction of structural variations in patient-derived iPSCs using CRISPR/Cas9. Nat. Protoc. *11*, 2154-2169.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., *et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl Acad. Sci. USA *112*, E6456-6465.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. Nature *450*, 509-514.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221-230.

Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.

Tai, D.J., Ragavendran, A., Manavalan, P., Stortchevoi, A., Seabra, C.M., Erdin, S., Collins, R.L., Blumenthal, I., Chen, X., Shen, Y.*, et al.* (2016). Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR. Nat. Neurosci. *doi:10.1038/nn.4235*.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. *41*, e132.

Torres, R., Martin, M.C., Garcia, A., Cigudosa, J.C., Ramirez, J.C., and Rodriguez-Perales, S. (2014). Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. Nat. Commun. *5*, 3964.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511-515.

Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell *97*, 779-790.

Wu, S., Ying, G., Wu, Q., and Capecchi, M.R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. Nat. Genet. *39*, 922-930.

Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., Yang, Z., Zu, Y., Li, W., Huang, P., Tong, X.*, et al.* (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res. *41*, e141.

Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. Cell *167*, 1814-1828 e1812.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370-1379.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A.*, et al.*

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zheng, B., Sage, M., Sheppeard, E.A., Jurecic, V., and Bradley, A. (2000). Engineering mouse chromosomes with Cre-*loxP*: range, efficiency, and somatic applications. Mol. Cell Biol. *20*, 648-655.

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends:

Figure 1. CtIP Debilitation Increases the Frequency of Precise DFD

(A-B) Quantitative analyses by NGS of the deletion junctions of a DNA fragment in the *Pcdh* or *β-globin* locus. (C-D) Quantitative analyses of small deletions at DFD junctions in the *Pcdh* and *β-globin* loci. (E-F) Confirmation of the increase of precise DFDs with CtIP debilitation by TA cloning and Sanger sequencing. (G-H) Confirmation of the decrease of small deletions at the DFD junctions with CtIP debilitation by TA cloning and Sanger sequencing. (I) CtIP deficiency leads to increased precise DFD of distinct DNA fragments of different sizes ranging from 6 kb to 930 kb by paired sgRNAs in the four PAM configurations. (J) CtIP debilitation leads to a significant decrease of small deletions at DFD junctions of distinct DNA fragments with different sizes ranging from 6 kb to 930 kb in all four PAM configurations. Data as mean ± SD, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; one-tailed Student's *t*-test.

Figure 2. Confirmation of the Role of CtIP by Triapine and CRISPR Single Cell Clones

(A-B) Inhibition of CtIP activity by triapine (3AP), a drug that leads to inactivation of CtIP, increases the frequency of precise DFD in the *Pcdh* and *β-globin* loci. (C-D) CtIP inhibition by 3AP results in a significant decrease of small deletions at the DFD junctions in the *Pcdh* and *β-globin* loci. (E) CtIP-deficient cell clones (C14, C27) generated by CRISPR engineering display a significant increase of precise DFD frequency. (F) There is a significant decrease of small deletions in

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CtIP mutant cell clones at the *Pcdh*, *β-globin*, and its LCR. Data as mean ± SD, **P* < 0.05, ***P* < 0.01, ****P* < 0.001; one-tailed Student's *t*-test.

Figure 3. Non-random Nucleotide Insertions at DFE Junctions *in vivo*.

(A) The inserted nucleotides at DFD junctions are resulted from the cleavage by RuvC of the noncomplementary strand upstream of the **second** PAM site. The sequences of a targeted *Pcdh* DNA fragment (1,233 bp) are shown with PAM in red. The inserted nucleotides detected by NGS are highlighted in the yellow background and are shown with their frequencies. The relative ratios of "T", "C", "A", "G" at each position of insertions are also shown. (B) The inserted nucleotides at the junctions of DNA fragment duplications are from the cleavage by RuvC of the noncomplementary strand upstream of the **first** PAM site. The inserted nucleotides are highlighted in the blue background and are shown with their frequencies. (C) Nonspecific insertions at the upstream junctions of DFI. (D) The inserted nucleotides at the downstream junctions of DFI are shown with their frequencies. They are resulted from the combined insertions at junctions of both DNA-fragment deletion (Cas9 with sgRNA2) and duplication (Cas9 with sgRNA1).

Figure 4. Altered Scissile Profiles of Rationally Engineered Cas9 Nucleases

(A) A surface representation of the crystal structure of SpCas9 (PDB 5F9R) (Jiang et al., 2016) showing the two linker regions (depicted in green ribbon) between the HNH (yellow) and RuvC (blue) domains. The noncomplementary

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strand is depicted in red, the complementary strand in cyan, and sgRNA in wheat. (B) Altered scissile profiles at -3 to -6 positions on the noncomplementary strand of engineered Cas9 nucleases. (C). Each engineered Cas9 nuclease display a unique scissile profiles for the ten sgRNAs tested. (D-G) The engineered Cas9 nucleases display a significant increase of precise DFD and a concomitant decrease of small deletions in the *Pcdh* and *β-globin* loci in the *CtIP*-deficient cell clone. Data as mean ± SD, **P* < 0.05, ***P* < 0.01, ****P* < 0.001; one-tailed Student's *t*-test.

Figure 5. Predictable Insertions at the Downstream DFI Junctions

(A) The sequences of a targeted *β-globin* DNA fragment (6,277 bp) are shown with PAM in red. (B-C) Shown are the inserted nucleotides and their frequencies at the duplication or deletion junctions of engineered Cas9 nucleases. (D) Predictable Cas9-mediated nucleotide insertions at the downstream DFI junctions. The observed frequencies of the downstream insertions at the DFI junction are predicted to be equal to the product of multiplication of the normalized insertion frequencies at the both duplication and deletion junctions. Note that the nucleotide insertion of "+GCAT" are from both "X_2Y_4" (the combined insertion of "+**G**" and "+CAT") and "X_3Y_3" (the combined insertion of "+**GC**" and "+AT"). The Pearson's product-moment correlation coefficients and their *P* values for all of the engineered Cas9 nucleases are indicated on the right.

Figure 6. Dissection of the *Pcdh HS5-1* Enhancer Architecture by DFI

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reveals the Insulational Role of the Single Boundary CTCF Site

(A) Schematics of the three human *Pcdh* gene clusters. The composite *HS5-1* enhancer and its various CRISPR inversions are enlarged. Two tandem CBSs in a reverse orientation (*HS5-1a* and *HS5-1b*) are indicated. (B) Shown are chromatin-interaction profiles in the wild-type control (Ctr) and CRISPR cell lines with inversion of the CBS *HS5-1b* alone (Inv I) or the *HS5-1b* combined with the middle region (Inv II) using the *HS5-1* enhancer as an anchor. The significance of interactions (*P* value) is shown under the reads density. The log2 ratios between each inversion cell clone and wild-type cells are also indicated. (C) Shown are chromatin-interaction profiles in wild-type (Ctr) and CRISPR cell lines with inversion of the CBS *HS5-1a* alone (Inv III) or the *HS5-1a* combined with the middle region (Inv IV) using the *HS5-1* enhancer as an anchor. (D) Heatmap shows the alteration of expression patterns of members of the *Pcdh* gene clusters resulting from CRISPR inversion of different enhancer elements by RNA-seq experiments.

Figure 7. A Model of Predictable Cas9-mediated Nucleotide Insertions for the Four PAM Configurations

Shown are schematics of predictable DFEs by Cas9 with paired sgRNAs in four PAM configurations. Through filling in of DSB ends by an unknown polymerase and ligation, Cas9-mediated nucleotide insertions match the combined sequences upstream of both PAM sites for all four PAM configurations.

Figure S1. Characterization of Junctions of DNA Fragment Deletion,

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Inversion, and Duplication with Inactivation of a Set of DNA-repair Genes and CtIP Debilitation in Four PAM Configurations, Related to Figure 1

(A-B) Quantitative analyses of insertions by NGS at DFD junctions in the *Pcdh* and *β-globin* loci. (C-D) Analyses of the frequency of precise ligations and indels by NGS at the upstream junctions of DFI in the *Pcdh* and *β-globin* loci. (E-F) Analyses of the frequency of precise ligations and indels by NGS at the downstream junctions of DFI. (G-H) Analyses of the frequency of precise ligations and indels by NGS at the junctions of DNA fragment duplication. (I) CtIP debilitation results in no significant alteration of insertions at DFD junctions. (J-L) CtIP debilitation results in no significant alteration of precise ligations and indels at the upstream and downstream DFI junctions as well as DNA-fragment-duplication junctions. Data as mean ± SD.

Figure S2. Analyses of DFE Junctions with 3AP Inhibition of CtIP Activity and in CtIP Mutant Cell Clones, Related to Figure 2

(A-B) 3AP results in no significant alteration of insertion at the DFD junctions. (C-H) 3AP does not alter the precise ligations and indels at the DFI junctions as well as DNA fragment duplication junctions. (I-J) Generation of subcloned cell lines of the CtIP mutations using the CRISPR/Cas9 system with a pair of sgRNAs. Shown are Sanger sequencing traces for two CtIP mutant CRISPR cell clones (C14 and C27). Sequencing confirmed the CtIP mutations. (K-N) There is no significant alteration of insertions at DFD junctions as well as of precise ligations and indels at the DFI junctions and DNA-fragment-duplication

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

junctions in CtIP mutant cell clones. Data as mean ± SD.

Figure S3. Analyses of Nucleotide Insertions at the Junctions of DFE of 709-bp or 6,277-bp DNA Fragments in the Human *β-globin* Locus, Related to Figure 3

(A) Shown are the schematics of the DNA-fragment deletion, inversion, and duplication. (B) Analyses of the insertions at DFD junctions of a 709-bp DNA fragment. The inserted nucleotides at DFD junctions are resulted from the cleavages by RuvC of the noncomplementary strand upstream of the **second** PAM site. The sequences of a targeted *β-globin* DNA fragment are shown with PAM in red. The inserted nucleotides detected by NGS are highlighted in the yellow background and are shown with their frequencies. The relative ratio of "T", "C", "A", "G" at each position of insertions is also shown. (C) Analyses of insertions at the junctions of a 709-bp DNA fragment duplication. The inserted nucleotides at the junctions of DNA fragment duplications are resulted from the cleavages by RuvC of the noncomplementary strand upstream of the **first** PAM site. The inserted nucleotides are highlighted in the blue background and are shown with their frequencies. (D) Nonspecific insertions at the upstream junctions of DFI of a 709-bp DNA fragment. (E) The inserted nucleotides at the downstream junctions of DFI of a 709-bp DNA fragment are shown with their frequencies. They are resulted from the combined insertions at junctions of both DNA-fragment deletion (Cas9 with sgRNA2) and duplication (Cas9 with sgRNA1). (F-I) Analyses DFE junctions of a 6,277-bp DNA fragment reveal

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

similar principles of Cas9-mediated nucleotide insertions.

Figure S4. Analyses of Nucleotide Insertions at the Junctions of DFE of 80,732-bp or 930-kb DNA Fragments with Distinct PAM Configurations in the Human *β-globin* Locus, Related to Figure 3

(A) Analyses of insertions at the DFD junctions of an 80,732-bp DNA fragment. The inserted nucleotides are resulted from the cleavages by RuvC of the noncomplementary strand upstream of the **first** PAM site. The inserted nucleotides are highlighted in the yellow background and are shown with their frequencies. The relative ratio of "T", "C", "A", "G" at each position of insertions is also shown. (B) Analyses of insertions at the junctions of an 80,732-bp DNA-fragment duplication. The inserted nucleotides are resulted from the cleavages by RuvC of the noncomplementary strand upstream of the **second** PAM site. The inserted nucleotides are highlighted in the pink background and are shown with their frequencies. (C) Analyses of insertions at the **upstream** junctions of DFI of an 80,732-bp DNA fragment. The inserted nucleotides are shown with their frequencies. They are resulted from the combined insertions at junctions of both DNA-fragment deletion (Cas9 with sgRNA1) and duplication (Cas9 with sgRNA2). (D) Nonspecific insertions at the downstream junctions of DFI of an 80,732-bp DNA fragment. (E) Analyses of nucleotide insertions at the upstream junctions of a 930-kb DNA fragment inversion. The inserted nucleotides are resulted from the cleavages by RuvC of the noncomplementary strand upstream of the **second** PAM site. (F) Analyses of nucleotide insertions at the

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

downstream junctions of DFI of a 930-kb DNA fragment. The inserted nucleotides are from the cleavages by RuvC of the noncomplementary strand upstream of the **first** PAM site. The inserted nucleotides are highlighted in the blue background and are shown with their frequencies. (G) Nonspecific insertions at the junctions of DFD of a 930-kb DNA fragment. (H) Analyses of insertions at the junctions of DNA fragment **duplication of** a 930-kb DNA fragment. The inserted nucleotides are derived from the combined insertions at the both upstream (Cas9 with sgRNA2) and downstream junctions of DFI (Cas9 with sgRNA1).

Figure S5. Altered Scissile Profiles of Rationally Engineered Cas9 Nucleases Reveal Predictable DFEs of a 709-bp Fragment in the Human *β-globin* Locus, Related to Figure 5

(A) The sequences of a targeted DNA fragment are shown with PAM in red. (B-C) Shown are the inserted nucleotides at the duplication or deletion junctions and their frequencies. (D-E) The heatmap of the relative ratio of insertions at the duplication and deletion junctions reveals altered scissile profiles of engineered Cas9 nucleases. (F) Predictable insertions at the downstream DFI junctions. The observed frequencies of the downstream insertions at the DFI junction are predicted to be equal to the product of multiplication of the normalized insertion frequencies at the both duplication and deletion junctions. Note that the insertions of “+TC” and “+TCC” are each from two different combinations. The Pearson’s correlation coefficients and their *P* values for all

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the engineered Cas9 nucleases are indicated on the right.

Figure S6. Altered Scissile Profiles of Rationally Engineered Cas9 Nucleases Reveal Predictable DFEs of a 930-kb Fragment in the Human *β-globin* Locus, Related to Figure 5

(A) The sequences of a targeted *β-globin* DNA fragment are shown with PAM in red. (B-C) Shown are the inserted nucleotides at the downstream and upstream junctions of DFI as well as their frequencies. (D-E) The heatmap of the relative ratio of insertions at the downstream and upstream junctions of DFI reveals altered scissile profiles of engineered Cas9 nucleases. (F) Predictable insertions at the **duplication** junctions. The observed frequencies of the insertions at the duplication junctions are predicted to be equal to the product of multiplication of the normalized insertion frequencies at the both upstream and downstream junctions of DFI. The Pearson's correlation coefficients and their *P* values for all of engineered Cas9 nucleases are indicated on the right.

Figure S7. DFI of Progressive Numbers of Tandem CBS Sites at the Border Region of a *β-globin* Topological Domain Reveals Principles of CBS Insulation, Related to Figure 6

(A) Schematic of the orientated CBS clusters and the various DFIs of CBS combinations containing the inversion of CBS15, CBS14-15, or CBS13-14 induced by paired sgRNAs in the *β-globin* locus. (B) Shown are the schematics of predicted long-distance chromatin-looping interactions by inverting the relative orientation of CBSs in the *β-globin* locus. The chromatin-interaction

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

profiles in control (Ctr) and various CBSs inversion clones using CBS13-15 as anchor are shown in the right panel. The CBS13-15 cell clone (E79) is a positive control (Guo et al., 2015). Note that the inversion of the two internal CTCF sites of CBS13-14 does not switch the chromatin-looping direction. Only inversions covering the boundary CBS15 switch the chromatin-looping direction. The significance of interactions (*P* value) is shown under the reads density. The log2 ratios between each inversion and the control clone are also indicated. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Supplemental Table S1. Oligonucleotides Used in This Study, Related to Experimental Procedures

Fig. 1

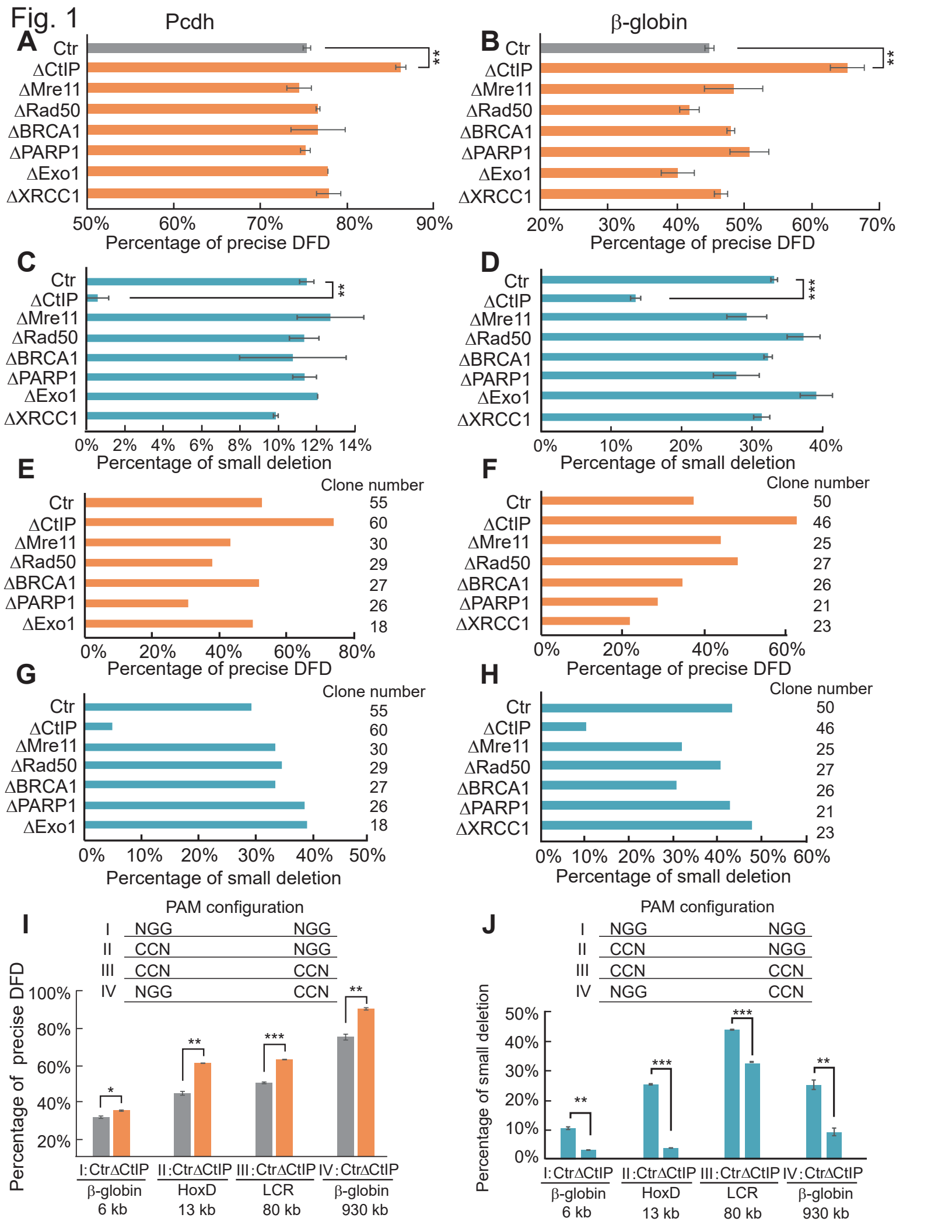

A
Pcdh
Ctr
ΔCtIP
ΔMre11
ΔRad50
ΔBRCA1
ΔPARP1
ΔExo1
ΔXRCC1
**
50%
60%
70%
80%
90%
Percentage of precise DFD
B
β-globin
Ctr
ΔCtIP
ΔMre11
ΔRad50
ΔBRCA1
ΔPARP1
ΔExo1
ΔXRCC1
**
20%
30%
40%
50%
60%
70%
Percentage of precise DFD
C
Ctr
ΔCtIP
ΔMre11
ΔRad50
ΔBRCA1
ΔPARP1
ΔExo1
ΔXRCC1
**
0%
2%
4%
6%
8%
10%
12%
14%
Percentage of small deletion
D
Ctr
ΔCtIP
ΔMre11
ΔRad50
ΔBRCA1
ΔPARP1
ΔExo1
ΔXRCC1
***
0%
10%
20%
30%
40%
Percentage of small deletion
E
Clone number
Ctr 55
ΔCtIP 60
ΔMre11 30
ΔRad50 29
ΔBRCA1 27
ΔPARP1 26
ΔExo1 18
0%
20%
40%
60%
80%
Percentage of precise DFD
F
Clone number
Ctr 50
ΔCtIP 46
ΔMre11 25
ΔRad50 27
ΔBRCA1 26
ΔPARP1 21
ΔXRCC1 23
0%
20%
40%
60%
Percentage of precise DFD
G
Clone number
Ctr 55
ΔCtIP 60
ΔMre11 30
ΔRad50 29
ΔBRCA1 27
ΔPARP1 26
ΔExo1 18
0%
10%
20%
30%
40%
50%
Percentage of small deletion
H
Clone number
Ctr 50
ΔCtIP 46
ΔMre11 25
ΔRad50 27
ΔBRCA1 26
ΔPARP1 21
ΔXRCC1 23
0%
10%
20%
30%
40%
50%
60%
Percentage of small deletion
I
PAM configuration
I NGG NGG
II CCN NGG
III CCN CCN
IV NGG CCN
Percentage of precise DFD
100%
80%
60%
40%
20%
*
**
***
**
I:CtrΔCtIP
β-globin
6 kb
II:CtrΔCtIP
HoxD
13 kb
III:CtrΔCtIP
LCR
80 kb
IV:CtrΔCtIP
β-globin
930 kb
J
PAM configuration
I NGG NGG
II CCN NGG
III CCN CCN
IV NGG CCN
Percentage of small deletion
50%
40%
30%
20%
10%
0%
**
***
***
**
I:CtrΔCtIP
β-globin
6 kb
II:CtrΔCtIP
HoxD
13 kb
III:CtrΔCtIP
LCR
80 kb
IV:CtrΔCtIP
β-globin
930 kb

Fig. 2

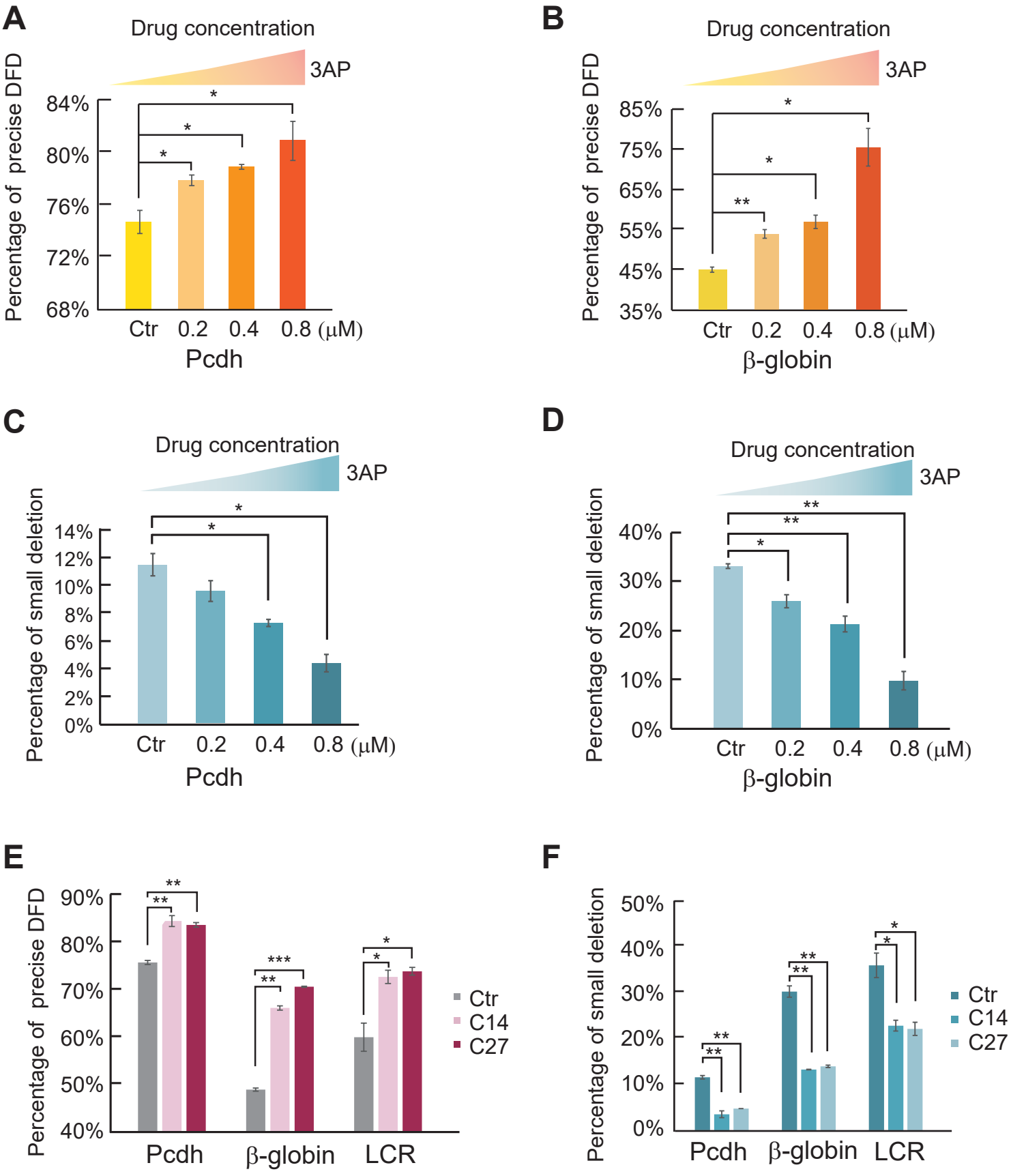

A
Drug concentration
3AP
Percentage of precise DFD
84%
80%
76%
72%
68%
*
*
*
Ctr
0.2
0.4
0.8 (μM)
Pcdh
B
Drug concentration
3AP
Percentage of precise DFD
85%
75%
65%
55%
45%
35%
*
*
**
Ctr
0.2
0.4
0.8 (μM)
β-globin
C
Drug concentration
3AP
Percentage of small deletion
14%
12%
10%
8%
6%
4%
2%
0%
*
*
Ctr
0.2
0.4
0.8 (μM)
Pcdh
D
Drug concentration
3AP
Percentage of small deletion
40%
30%
20%
10%
0%
**
**
*
Ctr
0.2
0.4
0.8 (μM)
β-globin
E
Percentage of precise DFD
90%
80%
70%
60%
50%
40%
**
**
***
**
*
*
Pcdh
β-globin
LCR
Ctr
C14
C27
F
Percentage of small deletion
50%
40%
30%
20%
10%
0%
**
**
**
**
*
*
Pcdh
β-globin
LCR
Ctr
C14
C27

Fig. 3

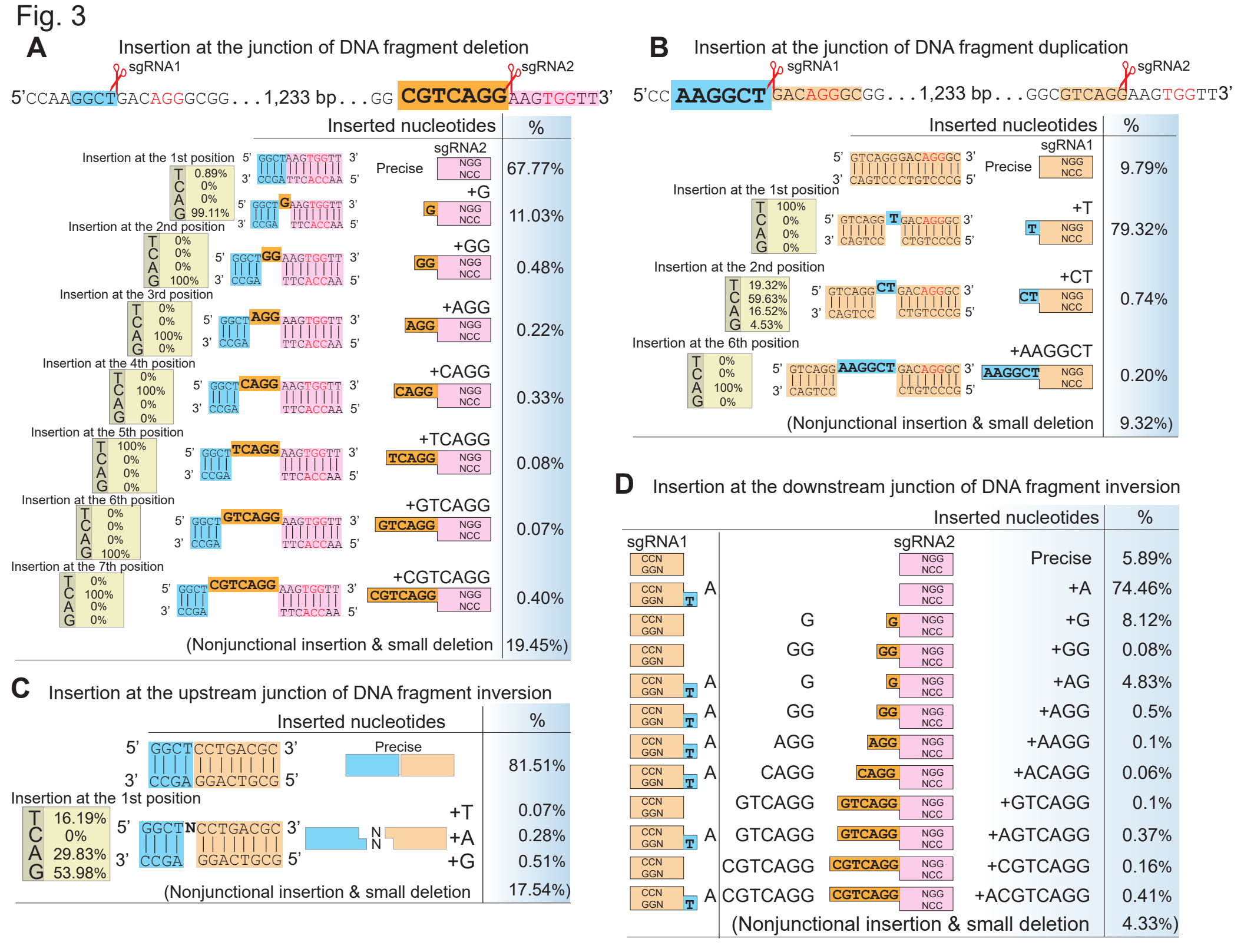

A
Insertion at the junction of DNA fragment deletion
sgRNA1
sgRNA2
5'CCAAGGCTGACAGGGCGG...1,233 bp...GG CGTCAGG AAGTGGTT3'
Inserted nucleotides
%
Insertion at the 1st position
T 0.89% C 0% A 0% G 99.11%
Precise
67.77%
+G
11.03%
Insertion at the 2nd position
T 0% C 0% A 0% G 100%
+GG
0.48%
Insertion at the 3rd position
T 0% C 0% A 100% G 0%
+AGG
0.22%
Insertion at the 4th position
T 0% C 100% A 0% G 0%
+CAGG
0.33%
Insertion at the 5th position
T 100% C 0% A 0% G 0%
+TCAGG
0.08%
Insertion at the 6th position
T 0% C 0% A 0% G 100%
+GTCAGG
0.07%
Insertion at the 7th position
T 0% C 100% A 0% G 0%
+CGTCAGG
0.40%
(Nonjunctional insertion & small deletion
19.45%)
B
Insertion at the junction of DNA fragment duplication
sgRNA1
sgRNA2
5'CC AAGGCT GACAGGGCGG...1,233 bp...GGCGTCAGGAAGTGGTT3'
Inserted nucleotides
%
Precise
9.79%
Insertion at the 1st position
T 100% C 0% A 0% G 0%
+T
79.32%
Insertion at the 2nd position
T 19.32% C 59.63% A 16.52% G 4.53%
+CT
0.74%
Insertion at the 6th position
T 0% C 0% A 100% G 0%
+AAGGCT
0.20%
(Nonjunctional insertion & small deletion
9.32%)
C
Insertion at the upstream junction of DNA fragment inversion
Inserted nucleotides
%
Precise
81.51%
Insertion at the 1st position
T 16.19% C 0% A 29.83% G 53.98%
+T
0.07%
+A
0.28%
+G
0.51%
(Nonjunctional insertion & small deletion
17.54%)
D
Insertion at the downstream junction of DNA fragment inversion
Inserted nucleotides
%
sgRNA1
sgRNA2
Precise
5.89%
+A
74.46%
+G
8.12%
+GG
0.08%
+AG
4.83%
+AGG
0.5%
+AAGG
0.1%
+ACAGG
0.06%
+GTCAGG
0.1%
+AGTCAGG
0.37%
+CGTCAGG
0.16%
+ACGTCAGG
0.41%
(Nonjunctional insertion & small deletion
4.33%)

Fig. 4

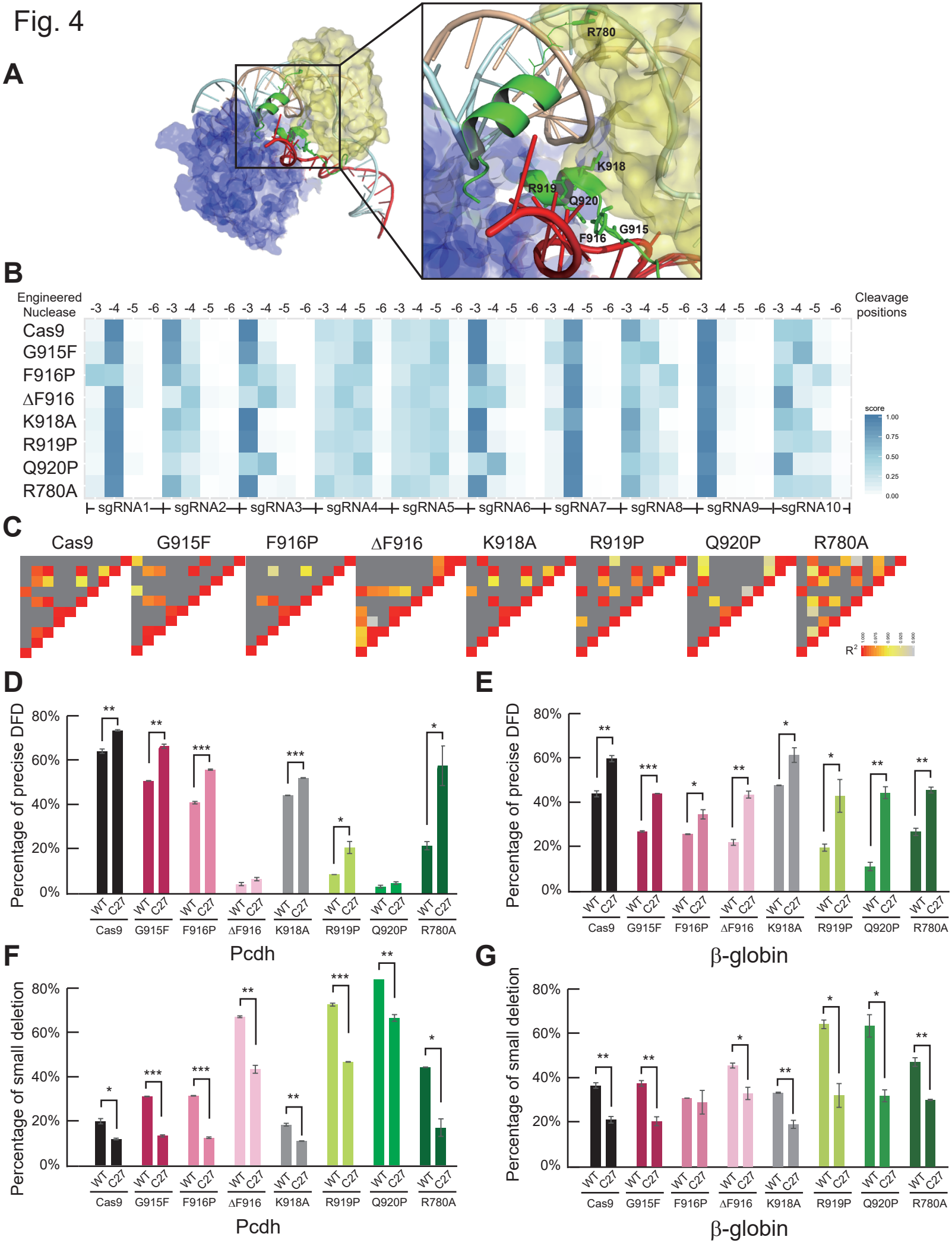
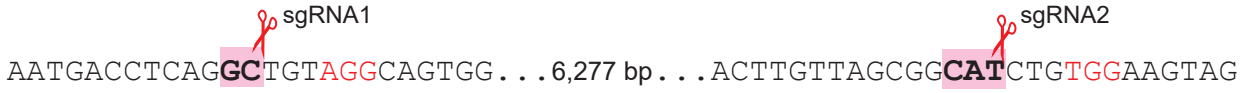

A
R780
K918
R919
Q920
G915
F916
B
Engineered Nuclease
Cleavage positions
Cas9
G915F
F916P
ΔF916
K918A
R919P
Q920P
R780A
score
1.00
0.75
0.50
0.25
0.00
sgRNA1
sgRNA2
sgRNA3
sgRNA4
sgRNA5
sgRNA6
sgRNA7
sgRNA8
sgRNA9
sgRNA10
C
Cas9
G915F
F916P
ΔF916
K918A
R919P
Q920P
R780A
R2
D
Percentage of precise DFD
WT
C27
Pcdh
E
Percentage of precise DFD
β-globin
F
Percentage of small deletion
Pcdh
G
Percentage of small deletion
β-globin

Fig. 5

A



B

Insertion at the duplication junction

SgRNA1 \ Inserted nt (%)	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
NGG NCC	82.76	68.76	38.04	17.25	86.28	64.73	23.47	58.63
+C C NGG NCC	5.00	15.04	22.70	33.46	2.25	3.27	31.17	2.47
+GC GC NGG NCC	0.00	0.20	11.99	11.14	0.00	2.11	0.43	4.03

C

Insertion at the deletion junction

SgRNA2 \ Inserted nt (%)	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
NGG NCC	22.57	14.77	11.07	10.96	17.57	13.56	5.73	18.35
+T T NGG NCC	19.24	17.77	27.18	22.89	27.48	15.31	7.92	11.05
+AT AT NGG NCC	33.43	40.39	28.10	13.34	34.98	12.69	13.48	12.70
+CAT CAT NGG NCC	3.29	2.09	0.57	0.83	2.04	2.50	0.37	1.92

D

Predictable insertion at the downstream DFI junction

Normalized ratio of insertion at the duplication junction

SgRNA1: **X**

	Precise X_1	+C X_2	+GC X_3
Cas9	94.30%	5.70%	0.00%
G915F	81.86%	17.90%	0.24%
F916P	52.30%	31.21%	16.49%
ΔF916	27.89%	54.10%	18.01%
K918A	97.46%	2.54%	0.00%
R919P	92.32%	4.66%	3.01%
Q920P	42.61%	56.60%	0.79%
R780A	90.02%	3.80%	6.19%

$Z =$ *

Normalized ratio of insertion at the deletion junction

SgRNA2: **Y**

	Precise Y_1	+T Y_2	+AT Y_3	+CAT Y_4
Cas9	28.74%	24.50%	42.57%	4.19%
G915F	19.68%	23.69%	53.83%	2.79%
F916P	16.55%	40.62%	41.98%	0.85%
ΔF916	22.82%	47.67%	27.78%	1.73%
K918A	21.41%	33.48%	42.62%	2.48%
R919P	30.78%	34.75%	28.80%	5.67%
Q920P	20.82%	28.80%	49.02%	1.36%
R780A	41.68%	25.10%	28.85%	4.37%

Observed frequency of nucleotide insertion at the downstream DFI junction

	Precise	+T	+AT	+CAT	+G	+GT	+GAT	+GCAT	+GC	+GCT	+GCCAT
Cas9	27.15%	24.07%	36.08%	3.59%	3.32%	2.14%	3.24%	0.40%	0.00%	0.00%	0.00%
G915F	16.40%	22.21%	39.38%	2.05%	5.10%	4.85%	9.37%	0.64%	0.00%	0.00%	0.00%
F916P	4.08%	23.36%	15.01%	1.16%	8.56%	20.33%	17.31%	3.99%	2.35%	3.78%	0.08%
ΔF916	1.74%	17.71%	3.36%	1.07%	10.84%	35.71%	10.32%	8.25%	3.32%	7.35%	0.33%
K918A	21.09%	32.23%	38.60%	2.50%	1.47%	1.76%	2.15%	0.19%	0.00%	0.00%	0.00%
R919P	24.48%	39.52%	24.10%	4.82%	1.24%	1.66%	0.86%	1.94%	0.83%	0.56%	0.00%
Q920P	5.77%	19.19%	11.65%	1.36%	12.21%	21.36%	24.46%	2.61%	0.46%	0.94%	0.00%
R780A	40.59%	24.85%	25.09%	3.95%	1.05%	0.50%	1.01%	1.47%	1.51%	0.00%	0.00%

Predicted frequency of nucleotide insertion at the downstream DFI junction **($Z = X * Y$)**

	Precise X_1Y_1	+T X_1Y_2	+AT X_1Y_3	+CAT X_1Y_4	+G X_2Y_1	+GT X_2Y_2	+GAT X_2Y_3	+GCAT $X_2Y_4+X_3Y_3$	+GC X_3Y_1	+GCT X_3Y_2	+GCCAT X_3Y_4
Cas9	27.10%	23.11%	40.14%	3.95%	1.64%	1.40%	2.43%	0.24%	0.00%	0.00%	0.00%
G915F	16.11%	19.39%	44.07%	2.28%	3.52%	4.24%	9.64%	0.63%	0.05%	0.06%	0.01%
F916P	8.65%	21.24%	21.96%	0.44%	5.16%	12.68%	13.10%	7.19%	2.73%	6.70%	0.14%
ΔF916	6.36%	13.29%	7.75%	0.48%	12.35%	25.79%	15.03%	5.94%	4.11%	8.59%	0.31%
K918A	20.87%	32.63%	41.54%	2.42%	0.54%	0.85%	1.08%	0.06%	0.00%	0.00%	0.00%
R919P	28.41%	32.08%	26.59%	5.23%	1.43%	1.62%	1.34%	1.13%	0.93%	1.05%	0.17%
Q920P	8.87%	12.27%	20.89%	0.58%	11.78%	16.30%	27.75%	1.15%	0.16%	0.23%	0.01%
R780A	37.52%	22.60%	25.97%	3.93%	1.58%	0.95%	1.10%	1.95%	2.58%	1.55%	0.27%

Pearson's correlation coefficients

R^2	*P* value
0.9964	5.686×10^{-11}
0.9929	1.213×10^{-9}
0.8621	6.413×10^{-4}
0.9300	3.358×10^{-5}
0.9989	3.116×10^{-13}
0.9793	1.495×10^{-7}
0.8982	1.732×10^{-4}
0.9978	6.677×10^{-12}

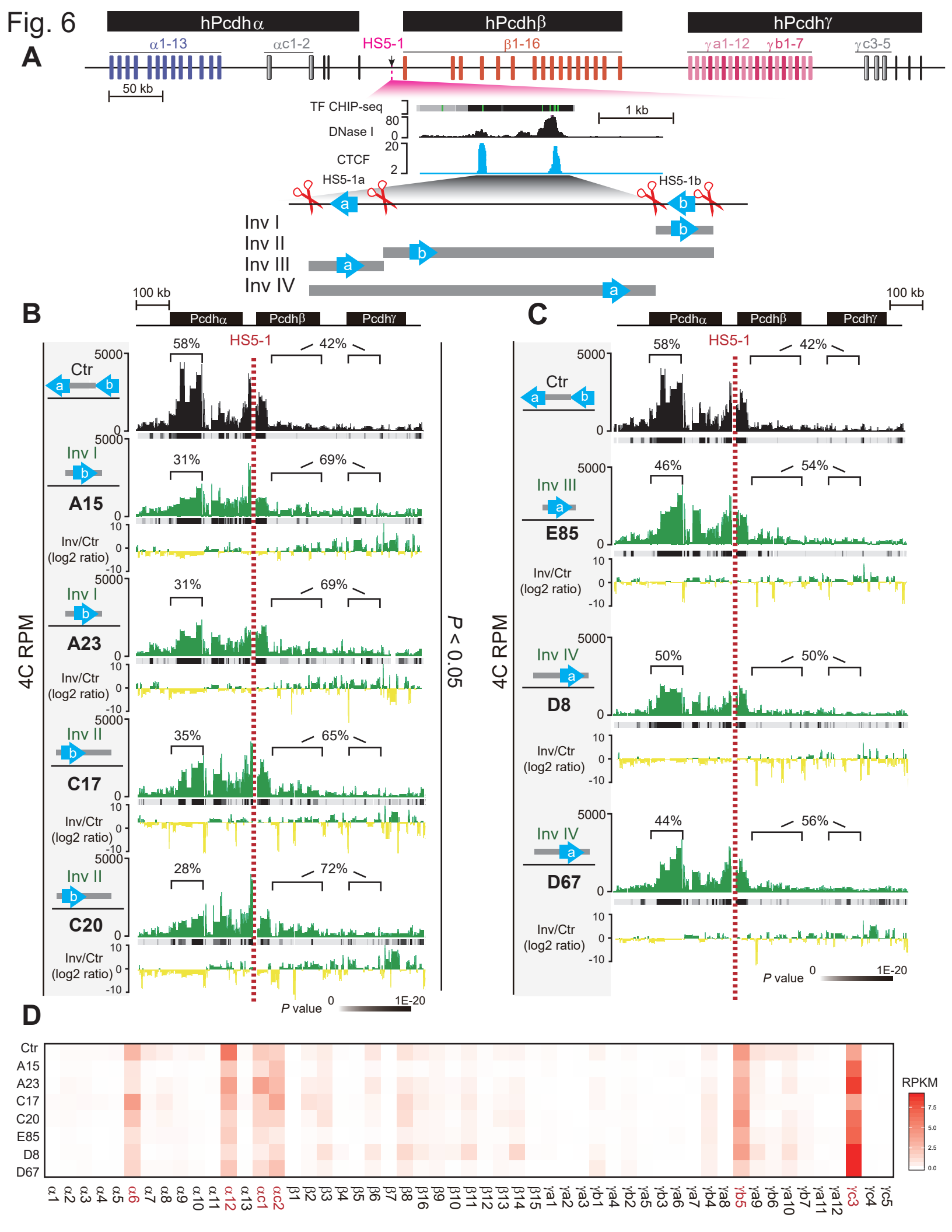


Fig. 6
A
hPcdhα
hPcdhβ
hPcdhγ
α1-13
αc1-2
HS5-1
β1-16
γa1-12
γb1-7
γc3-5
50 kb
TF CHIP-seq
DNase I
CTCF
1 kb
HS5-1a
HS5-1b
Inv I
Inv II
Inv III
Inv IV
B
C
100 kb
Pcdhα
Pcdhβ
Pcdhγ
Ctr
Inv I
A15
A23
Inv II
C17
C20
Inv III
E85
Inv IV
D8
D67
58%
42%
31%
69%
35%
65%
28%
72%
46%
54%
50%
44%
56%
4C RPM
Inv/Ctr (log2 ratio)
P < 0.05
P value
0
1E-20
D
RPKM
Ctr
A15
A23
C17
C20
E85
D8
D67

Fig. 7

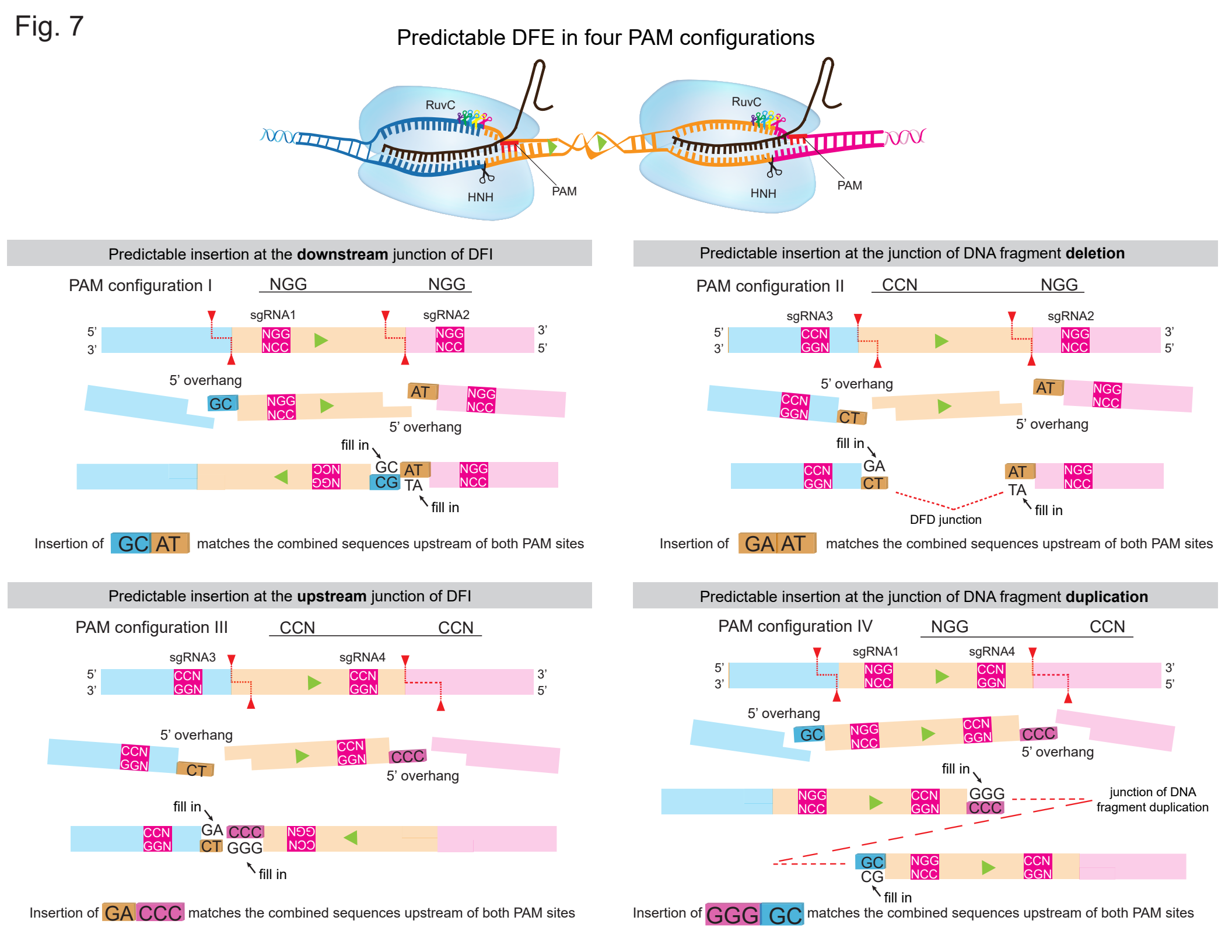

Predictable DFE in four PAM configurations
RuvC
HNH
PAM
RuvC
HNH
PAM
Predictable insertion at the downstream junction of DFI
PAM configuration I
NGG
NGG
sgRNA1
sgRNA2
5'
3'
NGG
NCC
5' overhang
GC
AT
fill in
CG
TA
Insertion of GC AT matches the combined sequences upstream of both PAM sites
Predictable insertion at the junction of DNA fragment deletion
PAM configuration II
CCN
NGG
sgRNA3
sgRNA2
CCN
GGN
CT
DFD junction
GA
Insertion of GA AT matches the combined sequences upstream of both PAM sites
Predictable insertion at the upstream junction of DFI
PAM configuration III
CCN
CCN
sgRNA3
sgRNA4
CCC
GGG
Insertion of GA CCC matches the combined sequences upstream of both PAM sites
Predictable insertion at the junction of DNA fragment duplication
PAM configuration IV
NGG
CCN
sgRNA1
sgRNA4
junction of DNA fragment duplication
Insertion of GGG GC matches the combined sequences upstream of both PAM sites

Fig. S1. Related to Fig. 1

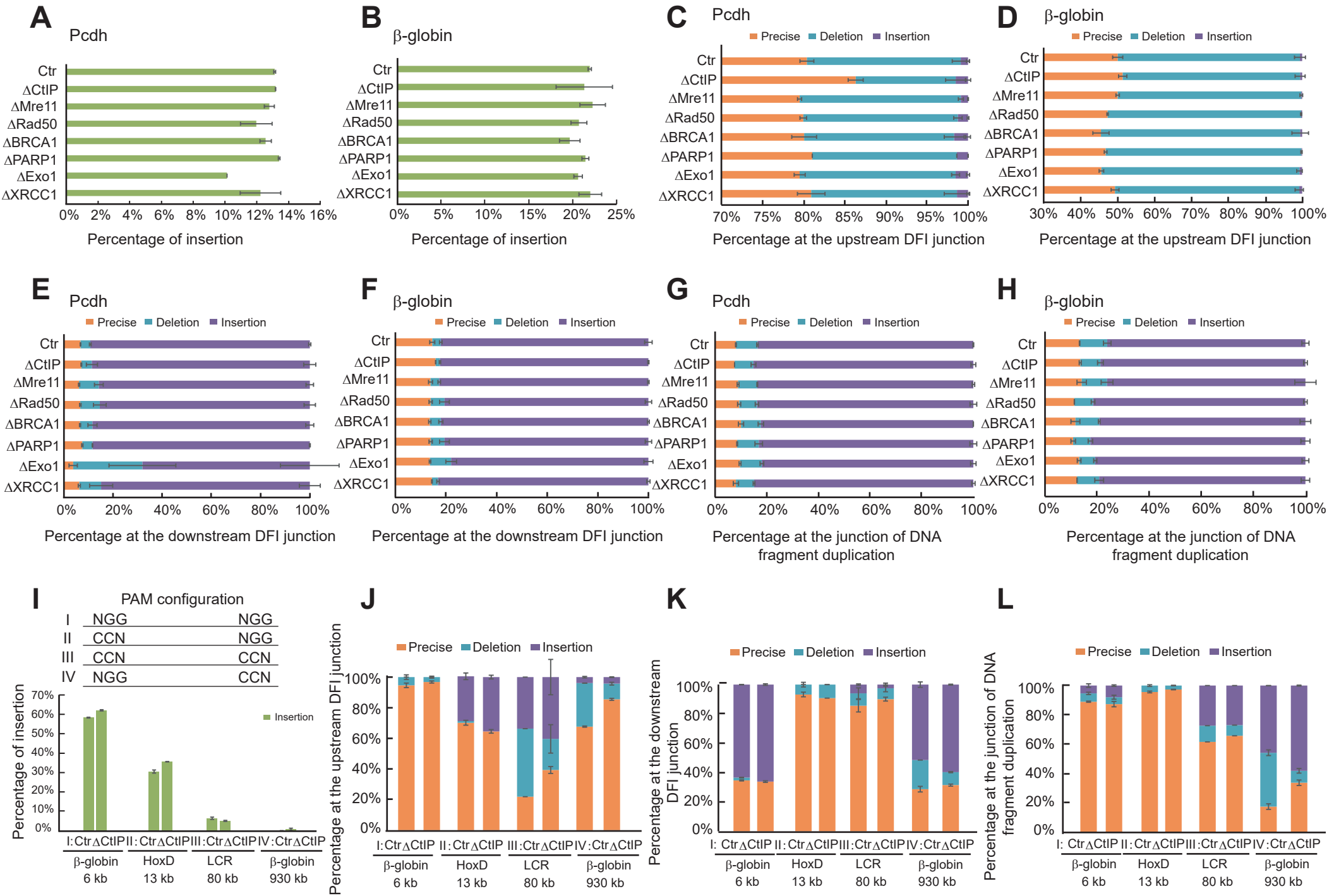


Fig. S2. Related to Fig. 2

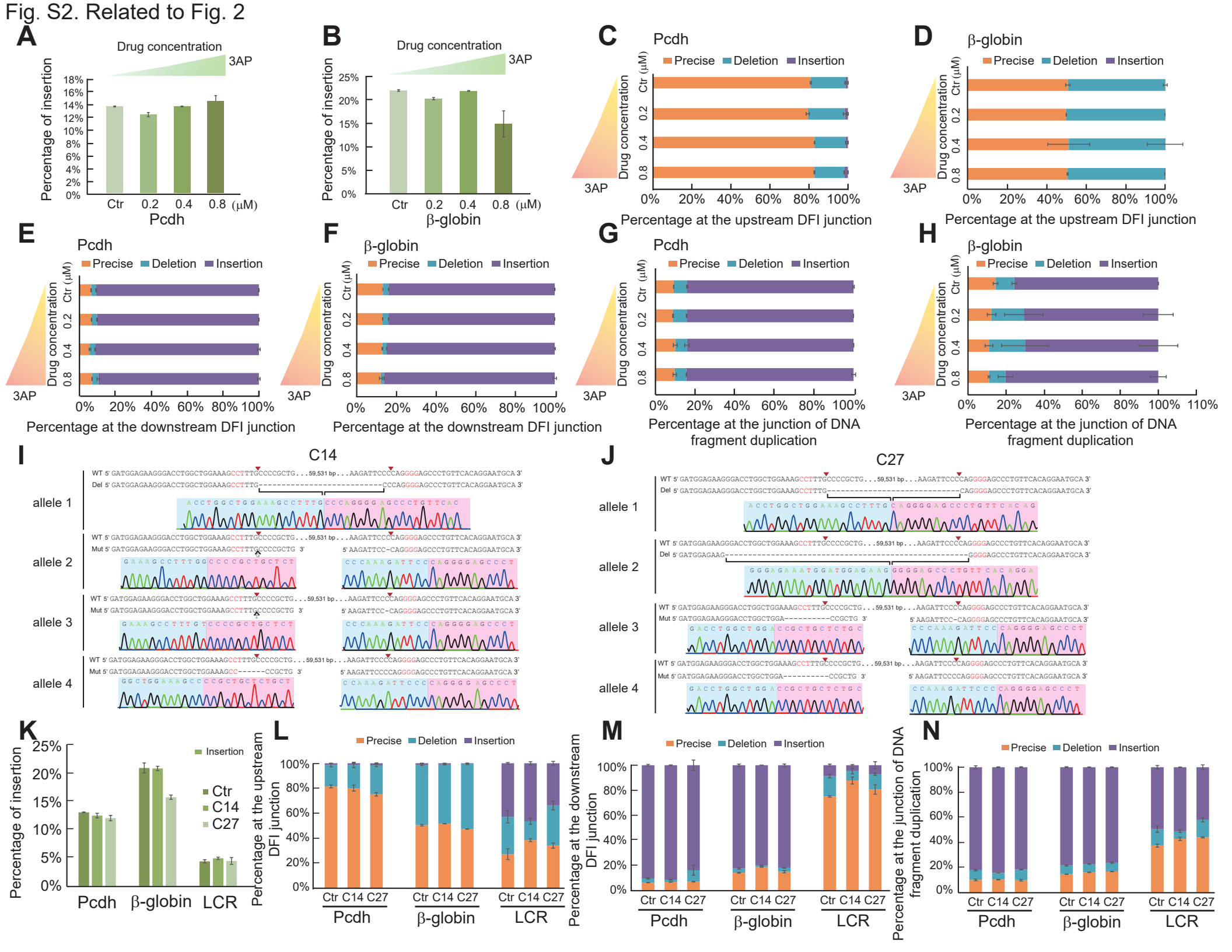


Fig. S3. Related to Fig. 3

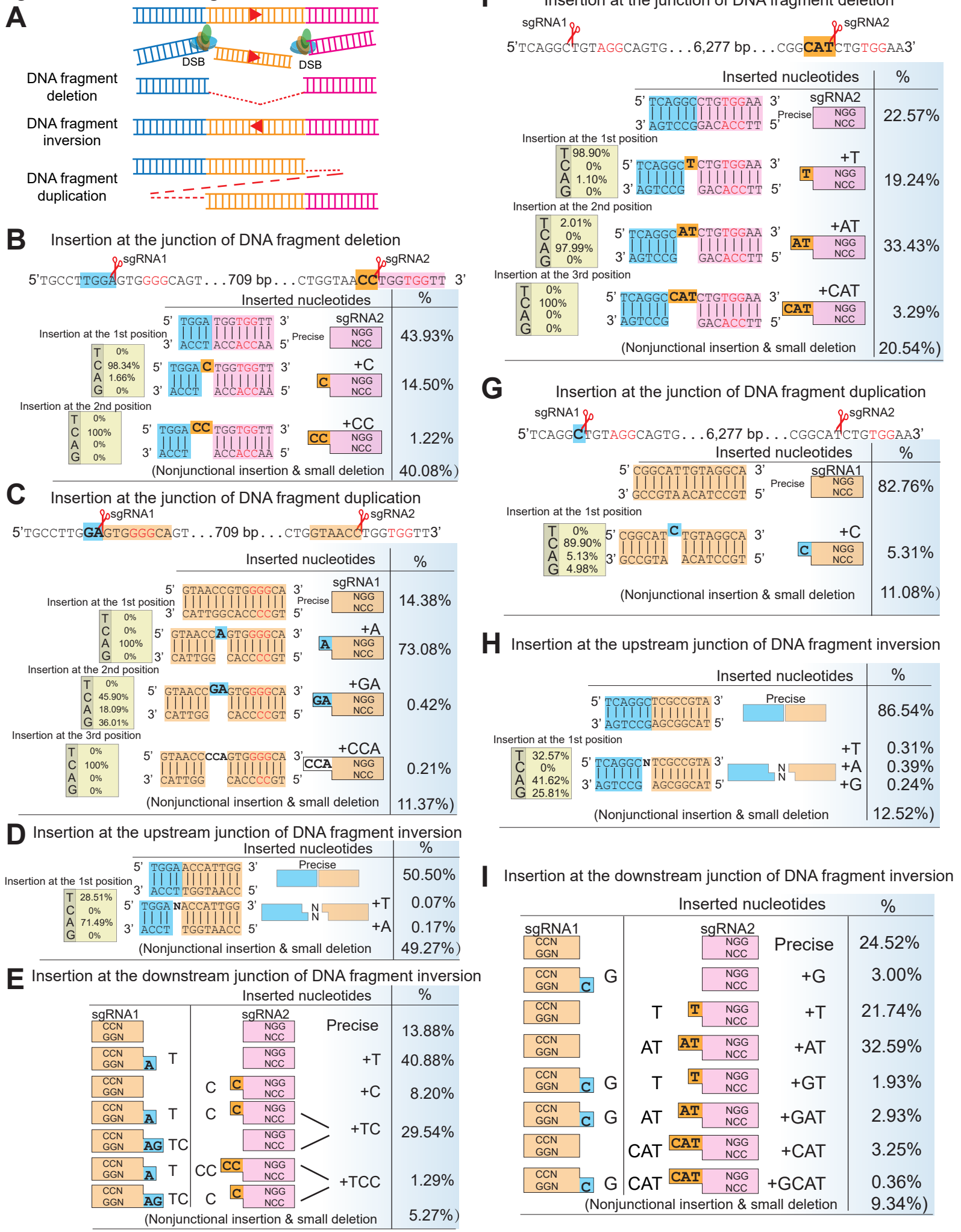


Fig. S4. Related to Fig. 3

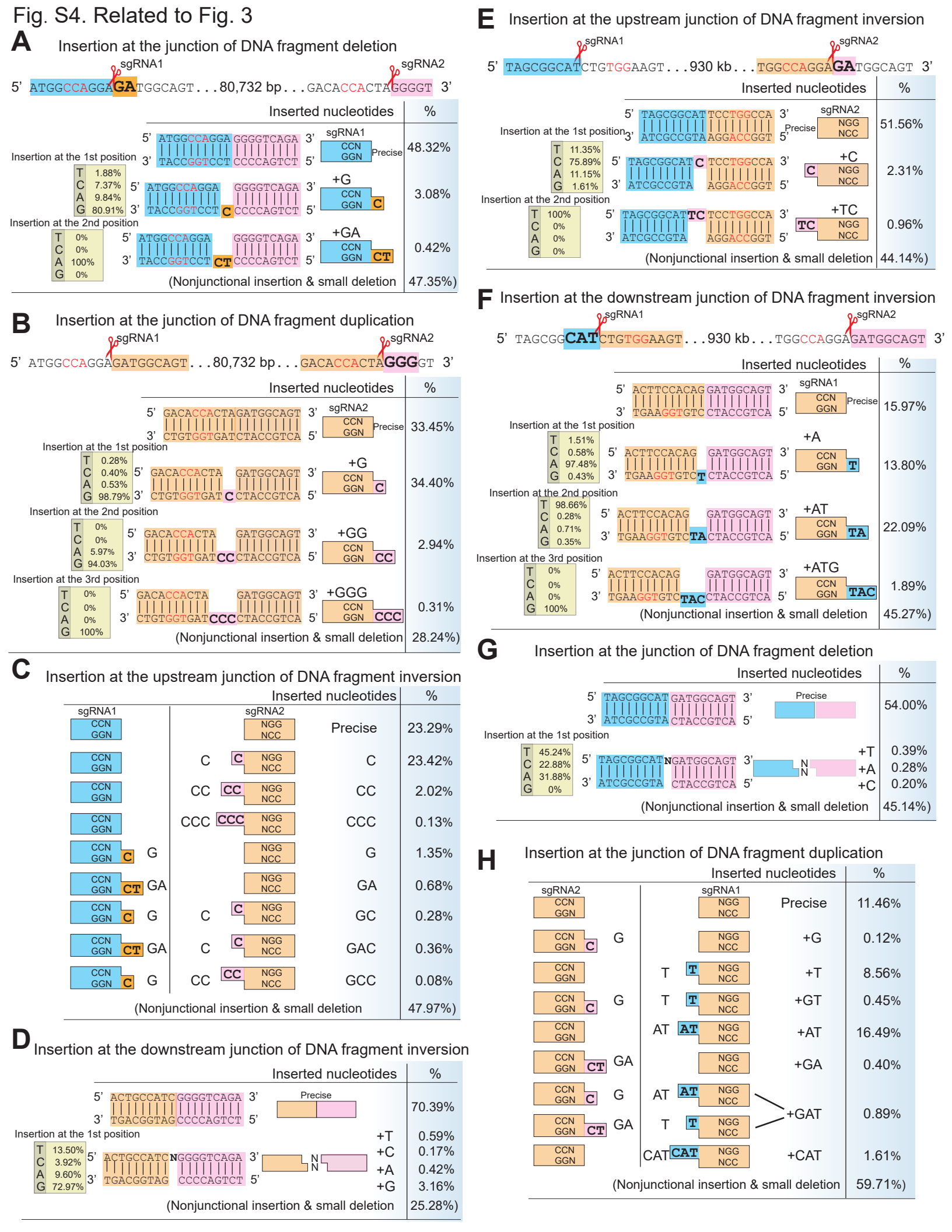
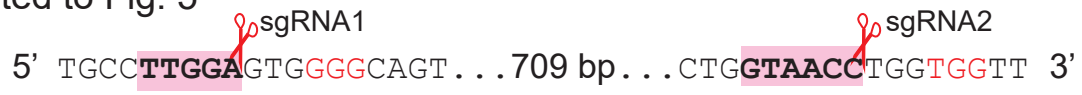


Fig. S5. Related to Fig. 5

A



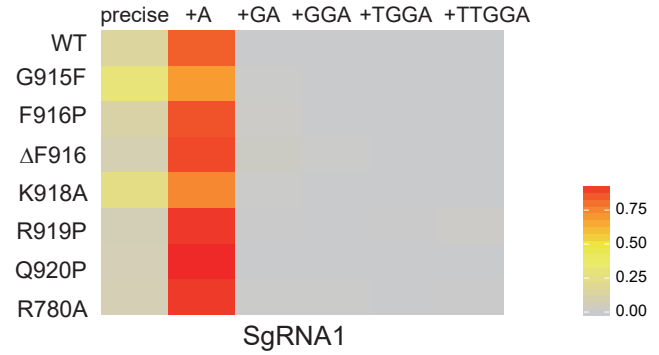
B Insertion at the duplication junction

SgRNA1 / Inserted nt	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
NGG NCC	12.69	24.01	8.28	4.58	20.01	5.23	5.12	5.74
+A (A NGG NCC)	66.92	53.96	56.68	48.06	58.66	67.08	65.44	69.54
+GA (GA NGG NCC)	0.06	0.51	0.66	1.21	0.45	0.26	0.00	0.68

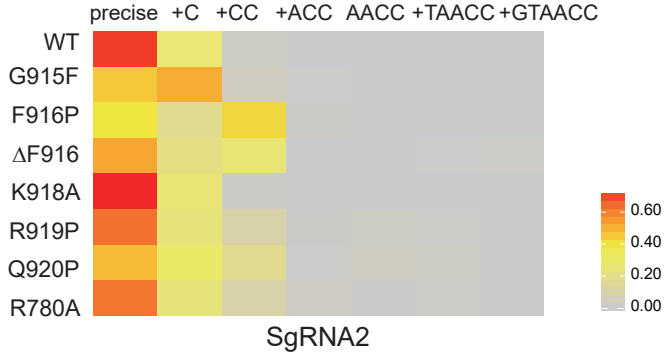
C Insertion at the deletion junction

SgRNA2 / Inserted nt	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
NGG NCC	43.62	26.68	25.44	21.76	47.45	19.38	10.89	26.49
+C (C NGG NCC)	15.83	29.92	11.09	8.61	16.23	7.22	6.84	10.04
+CC (CC NGG NCC)	1.38	1.83	26.95	10.57	0.80	2.78	3.92	4.24

D Insertion profile at the duplication junction



E Insertion profile at the deletion junction



F

Z = X * Y

Normalized ratio of insertion at the duplication junction — SgRNA1: **X**

	Precise X_1	+A X_2	+GA X_3
Cas9	15.93%	83.99%	0.08%
G915F	30.59%	68.76%	0.65%
F916P	12.62%	86.37%	1.01%
ΔF916	8.45%	88.68%	2.23%
K918A	25.29%	74.15%	0.57%
R919P	7.11%	91.11%	0.36%
Q920P	7.26%	92.75%	0.00%
R780A	7.48%	90.68%	0.89%

Normalized ratio of insertion at the deletion junction — SgRNA2: **Y**

	Precise Y_1	+C Y_2	+CC Y_3
Cas9	71.62%	25.99%	2.27%
G915F	45.29%	50.79%	3.11%
F916P	39.60%	17.27%	41.95%
ΔF916	52.40%	20.74%	25.45%
K918A	73.58%	25.17%	1.24%
R919P	63.55%	23.67%	9.13%
Q920P	47.84%	30.04%	17.21%
R780A	62.70%	23.77%	10.03%

Observed frequency of nucleotide insertion at the downstream DFI junction

	Precise	+C	+CC	+T	+TC	+TCC	+TCCC
Cas9	14.08%	9.61%	0.49%	41.60%	32.80%	1.43%	0.00%
G915F	13.47%	30.86%	1.19%	10.71%	42.70%	1.07%	0.00%
F916P	6.63%	5.98%	15.32%	26.75%	18.88%	26.19%	0.24%
ΔF916	6.20%	7.81%	5.87%	31.10%	26.23%	22.62%	0.17%
K918A	19.49%	11.90%	0.24%	43.25%	24.56%	0.57%	0.00%
R919P	15.00%	6.99%	1.34%	38.01%	31.23%	7.42%	0.00%
Q920P	12.45%	12.20%	2.72%	33.49%	26.09%	13.06%	0.00%
R780A	16.26%	9.26%	1.16%	39.66%	30.14%	3.51%	0.00%

Predicted frequency of nucleotide insertion at the downstream DFI junction **(Z = X * Y)**

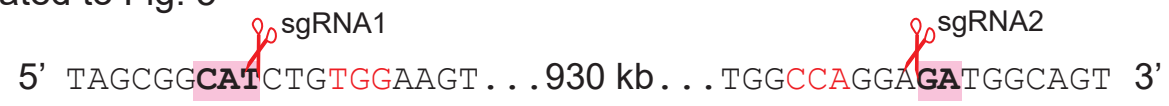
	Precise X_1Y_1	+C X_1Y_2	+CC X_1Y_3	+T X_2Y_1	+TC $X_2Y_2+X_3Y_1$	+TCC $X_2Y_3+X_3Y_2$	+TCCC X_3Y_3
Cas9	11.41%	4.14%	0.36%	60.15%	21.88%	1.93%	0.00%
G915F	13.85%	15.54%	0.95%	31.14%	35.21%	2.47%	0.02%
F916P	5.00%	2.18%	5.29%	34.20%	15.31%	36.41%	0.42%
ΔF916	4.43%	1.75%	2.15%	46.47%	19.56%	23.03%	0.57%
K918A	18.61%	6.37%	0.31%	54.56%	19.08%	1.06%	0.01%
R919P	4.52%	1.68%	0.65%	57.90%	21.80%	8.40%	0.03%
Q920P	3.47%	2.18%	1.25%	44.37%	27.86%	15.96%	0.00%
R780A	4.69%	1.78%	0.75%	56.85%	22.11%	9.31%	0.09%

Pearson's correlation coefficients

R^2	*P* value
0.9193	3.397×10^{-3}
0.7616	4.664×10^{-2}
0.9221	3.121×10^{-3}
0.9231	3.024×10^{-3}
0.9686	3.306×10^{-4}
0.8936	6.687×10^{-3}
0.9349	2.007×10^{-3}
0.8888	7.460×10^{-3}

Fig. S6 related to Fig. 5

A



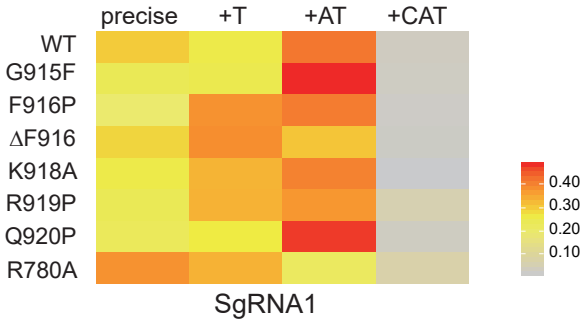
B Insertion at the downstream DFI junction

SgRNA1 Inserted nt \ Cas9	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
NGG NCC	15.97	11.83	9.31	10.87	15.02	10.95	9.13	23.48
+T (T NGG NCC)	13.80	12.14	18.14	14.53	19.39	15.21	10.44	20.79
+AT (AT NGG NCC)	22.09	23.86	19.59	11.73	23.33	16.84	18.74	13.69
+CAT (CAT NGG NCC)	1.89	1.57	1.22	1.07	0.99	2.77	1.34	4.59

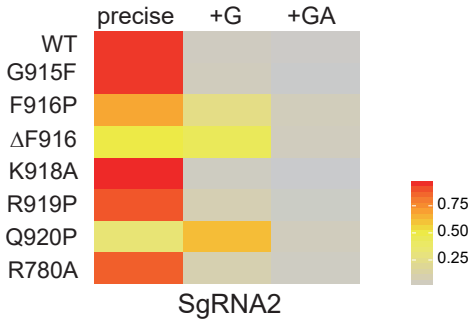
C Insertion at the upstream DFI junction

SgRNA2 Inserted nt \ Cas9	WT 1	G915F 2	F916P 3	ΔF916 4	K918A 5	R919P 6	Q920P 7	R780A 8
CCN GGN	51.56	38.33	28.26	15.05	40.16	30.23	8.98	30.60
+G (CCN GGN C)	2.31	2.16	11.05	13.50	1.62	3.04	16.91	3.42
+GA (CCN GGN CT)	0.96	0.30	2.50	1.59	0.19	0.84	1.47	1.28

D Insertion at the downstream DFI junction



E Insertion at the upstream DFI junction



F

Normalized ratio of insertion at the upstream DFI junction

SgRNA2: **Y**

	Precise Y_1	+G Y_2
Cas9	94.03%	4.21%
G915F	93.97%	5.31%
F916P	67.60%	26.42%
ΔF916	49.93%	44.78%
K918A	95.68%	3.85%
R919P	88.64%	8.90%
Q920P	32.83%	61.79%
R780A	86.70%	9.70%

Z = *

Normalized ratio of insertion at the downstream DFI junction

SgRNA1: **X**

	Precise X_1	+T X_2	+AT X_3	+CAT X_4
Cas9	29.71%	25.68%	41.10%	3.52%
G915F	23.95%	24.58%	48.29%	3.18%
F916P	19.29%	37.58%	40.58%	2.54%
ΔF916	28.45%	38.04%	30.72%	2.79%
K918A	25.58%	33.02%	39.72%	1.68%
R919P	23.93%	33.23%	36.80%	6.04%
Q920P	23.02%	26.33%	47.25%	3.39%
R780A	37.54%	33.24%	21.88%	7.33%

Observed frequency of nucleotide insertion at the duplication junction

	Precise	+G	+T	+GT	+AT	+GAT	+CAT	+GCAT
Cas9	28.96%	0.31%	21.62%	1.14%	41.67%	2.25%	4.07%	0.00%
G915F	20.79%	0.75%	22.28%	0.42%	50.37%	3.32%	2.06%	0.00%
F916P	18.02%	2.52%	26.39%	7.41%	31.71%	11.76%	1.98%	0.22%
ΔF916	16.51%	5.92%	17.93%	21.15%	15.96%	20.20%	1.90%	0.42%
K918A	26.74%	0.45%	27.91%	0.38%	42.09%	0.93%	1.49%	0.00%
R919P	26.97%	0.75%	27.40%	0.78%	37.39%	2.70%	3.15%	0.88%
Q920P	21.93%	6.78%	15.58%	20.56%	8.41%	24.79%	0.00%	1.95%
R780A	21.29%	0.00%	22.92%	1.14%	48.73%	2.06%	3.85%	0.00%

Predicted frequency of nucleotide insertion at the duplication junction **(Z = Y * X)**

	Precise Y_1X_1	+G Y_2X_1	+T Y_1X_2	+GT Y_2X_2	+AT Y_1X_3	+GAT Y_2X_3	+CAT Y_1X_4	+GCAT Y_2X_4
Cas9	27.94%	1.25%	24.14%	1.08%	38.64%	2.18%	3.31%	0.15%
G915F	22.50%	1.27%	23.09%	1.30%	45.38%	2.74%	2.99%	0.17%
F916P	13.04%	5.10%	25.41%	9.93%	27.43%	12.97%	1.71%	0.67%
ΔF916	14.21%	12.74%	18.99%	17.04%	15.34%	15.77%	1.39%	1.25%
K918A	24.47%	0.98%	31.60%	1.27%	38.00%	1.68%	1.61%	0.06%
R919P	21.21%	2.13%	29.45%	2.96%	32.62%	4.09%	5.36%	0.54%
Q920P	7.56%	14.22%	8.64%	16.27%	15.51%	30.62%	1.11%	2.09%
R780A	32.55%	3.64%	28.82%	3.23%	18.97%	3.32%	6.36%	0.71%

Pearson's correlation coefficients

R^2	*P* value
0.9959	1.733×10^{-7}
0.9961	1.446×10^{-7}
0.9792	2.220×10^{-5}
0.9090	1.759×10^{-3}
0.9918	1.357×10^{-6}
0.9847	8.818×10^{-6}
0.6731	6.733×10^{-2}
0.7041	5.125×10^{-2}

bioRxiv preprint doi: https://doi.org/10.1101/135822; this version posted May 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

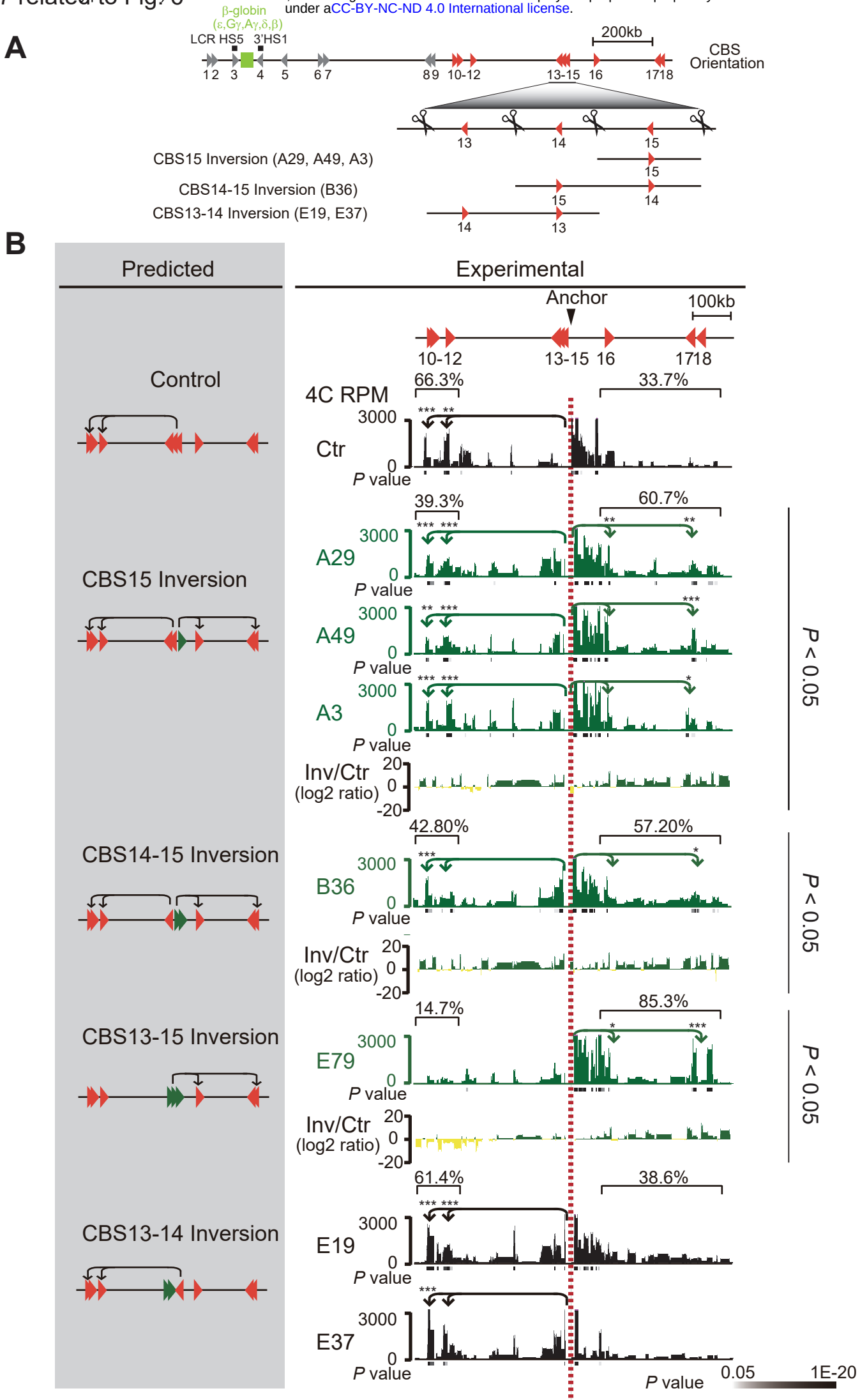



Fig. S7 related to Fig. 6