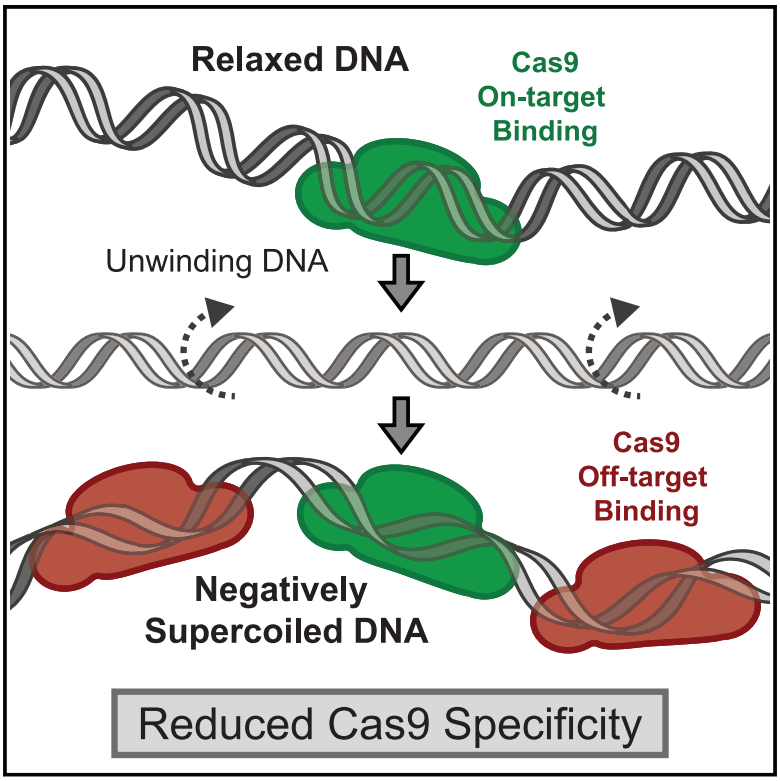
Molecular Cell

Article

Negative DNA supercoiling induces genome-wide Cas9 off-target activity

Graphical abstract



Highlights

- Cas9 shows increased off-target activity on negatively supercoiled DNA
- Negative supercoiling results in Cas9 cleavage at thousands of genomic sites
- Local DNA distortion increases Cas9 mismatch tolerance
- Supercoiling-dependent off-target cleavage can be detected during cell editing

Authors

Matthew D. Newton, Marialucrezia Losito, Quentin M. Smith, ..., Graeme A. King, Maria Emanuela Cuomo, David S. Rueda

Correspondence

g.king@ucl.ac.uk (G.A.K.), emanuela.cuomo@astrazeneca.com (M.E.C.), david.rueda@imperial.ac.uk (D.S.R.)

In brief

Newton et al. demonstrate that changes in DNA structure dramatically reduce the specificity of the gene editing technique CRISPR-Cas9. Underwound (negatively supercoiled) DNA results in off-target activity at thousands of sites across the human genome. This is important to understand for the development of safe and efficient gene editing therapeutics.

Newton et al., 2023, Molecular Cell *83*, 3533–3545
October 5, 2023
https://doi.org/10.1016/j.molcel.2023.09.008

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Article

Negative DNA supercoiling induces genome-wide Cas9 off-target activity

Matthew D. Newton,[1,2,3] Marialucrezia Losito,[1,2,4] Quentin M. Smith,[1,2] Nishita Parnandi,[3] Benjamin J. Taylor,[4] Pinar Akcakaya,[5] Marcello Maresca,[5] Patrick van Eijk,[6] Simon H. Reed,[6] Simon J. Boulton,[3] Graeme A. King,[7,*] Maria Emanuela Cuomo,[8,*] and David S. Rueda[1,2,9,*]

[1]Department of Infectious Disease, Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0HS, UK
[2]Single Molecule Imaging, MRC-London Institute of Medical Sciences, Du Cane Road, London W12 0HS, UK
[3]DSB Repair Metabolism Laboratory, The Francis Crick Institute, London NW1 1AT, UK
[4]Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge CB2 0AA, UK
[5]Genome Engineering, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, 43183 Gothenburg, Sweden
[6]Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4AW, UK
[7]Institute of Structural and Molecular Biology, University College London, London WC1E 6BT, UK
[8]Early Oncology, Oncology R&D, AstraZeneca, Cambridge CB2 0AA, UK
[9]Lead contact
*Correspondence: g.king@ucl.ac.uk (G.A.K.), emanuela.cuomo@astrazeneca.com (M.E.C.), david.rueda@imperial.ac.uk (D.S.R.)
https://doi.org/10.1016/j.molcel.2023.09.008

SUMMARY

CRISPR-Cas9 is a powerful gene-editing technology; however, off-target activity remains an important consideration for therapeutic applications. We have previously shown that force-stretching DNA induces off-target activity and hypothesized that distortions of the DNA topology *in vivo*, such as negative DNA supercoiling, could reduce Cas9 specificity. Using single-molecule optical-tweezers, we demonstrate that negative supercoiling λ-DNA induces sequence-specific Cas9 off-target binding at multiple sites, even at low forces. Using an adapted CIRCLE-seq approach, we detect over 10,000 negative-supercoiling-induced Cas9 off-target double-strand breaks genome-wide caused by increased mismatch tolerance. We further demonstrate *in vivo* that directed local DNA distortion increases off-target activity in cells and that induced off-target events can be detected during Cas9 genome editing. These data demonstrate that Cas9 off-target activity is regulated by DNA topology *in vitro* and *in vivo*, suggesting that cellular processes, such as transcription and replication, could induce off-target activity at previously overlooked sites.

INTRODUCTION

Since the discovery of the CRISPR-Cas9 system and its application as a genome-editing tool, off-target activity of Cas9 has presented a barrier to therapeutic applications. There has been a concerted effort to reduce off-target activity through the use of bioinformatic tools for intelligent guide design,[1–3] engineering of higher-fidelity CRISPR systems,[4–13] development of sensitive methods for off-target detection,[14–22] and the application of machine learning approaches to better predict editing outcomes *in silico*.[3,23–28]

The *S. pyogenes*-derived *Sp*Cas9 binds the CRISPR-RNA (crRNA) and *trans*-activating CRISPR-RNA (tracrRNA) to form the active endonuclease complex with DNA target specificity governed by a 20 nt sequence within the crRNA (Figure 1).[29] Single-molecule and structural studies have provided crucial details on the mechanisms of Cas9 target search, on-target binding, and off-target discrimination.[7,28,30–38] These demonstrate a target recognition mechanism of initial engagement with the three-nucleotide protospacer-adjacent motif (PAM) site (5′-NGG-3′) followed by directional R-loop formation with target DNA melting and annealing of the guide crRNA across the target site. Off-target discrimination occurs at the level of target binding and target cleavage. Firstly, the presence of mismatches that disrupt R-loop formation results in an increased rate of dissociation at off-target sites. Secondly, even at off-target sites where stable Cas9 binding can occur, there is a proposed "conformational checkpoint" where the conformational change of the HNH domain required for DNA cleavage cannot occur in the presence of excessive mismatches between the target DNA and crRNA.[7,33,39]

Work from ourselves and others have questioned the model of a fixed conformational checkpoint.[41–44] We have previously demonstrated that DNA stretching induces Cas9 off-target activity due to destabilization of the target DNA, permitting cleavage at highly mismatched sites even in the absence of full R-loop formation.[43] Such findings are consistent with kinetic models describing Cas9 off-target discrimination through the energy landscape of target unwinding and RNA:DNA hybridization

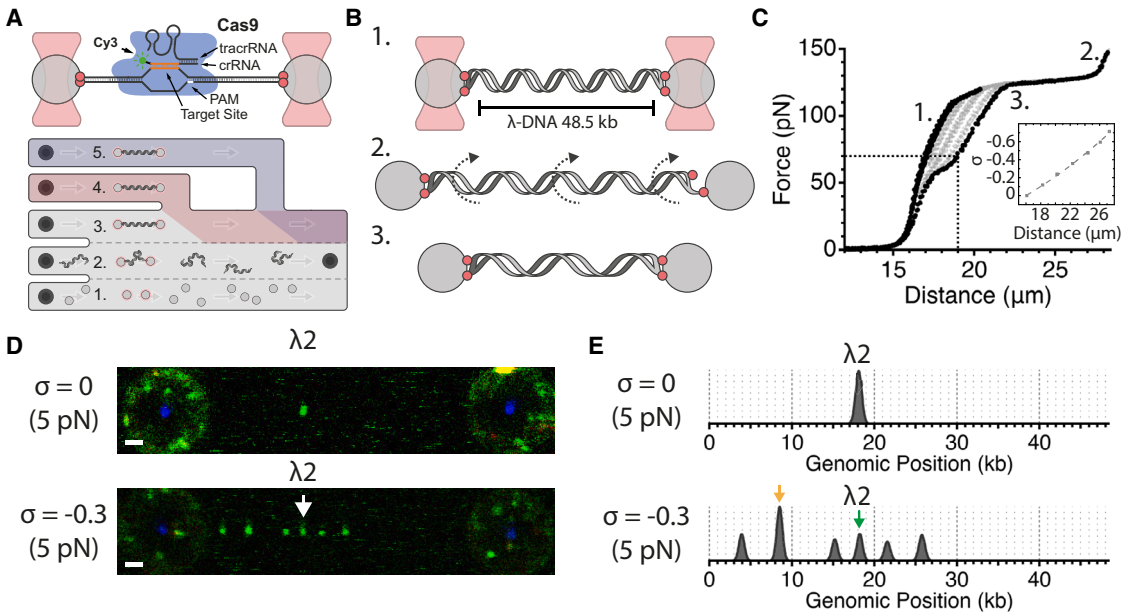


Figure 1. Negative supercoiling induces Cas9 off-target binding

(A) Experimental design. Top: A torsionally constrained λ-DNA is tethered between two optically trapped beads. Cy3-labeled Cas9 complex is targeted to the λ2 site. Bottom: Schematic of the five-channel microfluidic flow cell: (1) Bead Channel, (2) DNA Channel, (3) Buffer Channel, and (4 and 5) Protein/Imaging Channels.

(B) Negative supercoiling of DNA is generated within the optical-tweezers assay using Optical DNA Supercoiling (ODS).[40] A torsionally constrained λ-DNA molecule is tethered between optically trapped beads (1). DNA is stretched to >150 pN, and transient rupture of biotin-streptavidin interactions results in underwinding of the DNA (2). Biotin-streptavidin interactions are reformed as tension is reduced, yielding stable negatively supercoiled λ-DNA (3). The supercoiling density can be controlled through repeated stretch cycles.

(C) Example force-distance plot recorded during the generation of negatively supercoiled DNA, numbered as in (B). Insert shows a calibration plot for σ values as a function of DNA extension at 70 pN, adapted from King et al.[40]

(D) Representative 2D confocal scan of λ2-crRNA:tracrRNA dCas9 complex binding to λ-DNA at low force (5 pN) when DNA is torsionally constrained but non-supercoiled (top, $\sigma = 0$) and negatively supercoiled (bottom, $\sigma = -0.3$). The white arrow shows the on-target λ2 binding site. Scale bars, 1 μm.

(E) Intensity peaks from confocal scans in (D), mapped to the λ genome. The on-target binding site λ2 and the predicted off-target site at 8.4 kb (orange arrow) are indicated.

during R-loop formation.[26,27] Although the conformational checkpoint prevents cleavage in the presence of an excessive number of mismatches, other external factors, such as the local distortion of the DNA target, could favor the active conformation, overcoming the sequence-specificity requirement. Consistent with this, it has been demonstrated at individual sites that negative supercoiling modulates the energy landscape of R-loop formation.[44]

Based on these findings, we hypothesized that destabilization of the target DNA *in vivo* as a result of negative supercoiling during processes such as transcription and DNA replication could alter Cas9 specificity and induce off-target activity at previously overlooked sites. To test this hypothesis, we have used a combination of single-molecule optical-tweezers, next-generation sequencing (NGS)-based CIRCLE-seq and INDUCE-seq off-target detection, and cellular assays. Our data show that negative supercoiling induces Cas9 off-target binding and cleavage, reducing specificity and resulting in the potential for cleavage at over 10,000 sites across the human genome. We further demonstrate *in vivo* that localized DNA deformation in cells can increase Cas9 activity at off-target sites. This work highlights the requirement to carefully consider the context of potential off-target sites during the development of safe Cas9 therapeutics.

RESULTS

Negatively supercoiled DNA induces Cas9 off-target binding

To detect the on- and off-target activity of Cas9 on relaxed and negatively supercoiled DNA in real time at single-molecule resolution, we used optical-tweezers combined with confocal microscopy (Figure 1A). Using a multi-channel microfluidic flow cell, a torsionally constrained λ-DNA molecule (~48.5 kb) was tethered between two optically trapped streptavidin-coated polystyrene beads through biotin-streptavidin interactions (Figures 1A and 1B). The presence of a single torsionally constrained DNA molecule was confirmed by recording force-extension curves from 0 to 150 pN (Figure 1C). Torsionally constrained DNA displays a characteristic overstretching plateau at ~120 pN, compared to ~60 pN for torsionally unconstrained DNA (Figure 1C).[40] Negatively supercoiled DNA was generated by repeated cycles of high force extension (~150 pN) and retraction in an approach termed Optical DNA Supercoiling (ODS), as previously described (Figure 1C).[40] The supercoiling density (σ) can be directly calculated by measuring the extension at 70 pN and comparing this to literature values (Figure 1C, insert).[40]

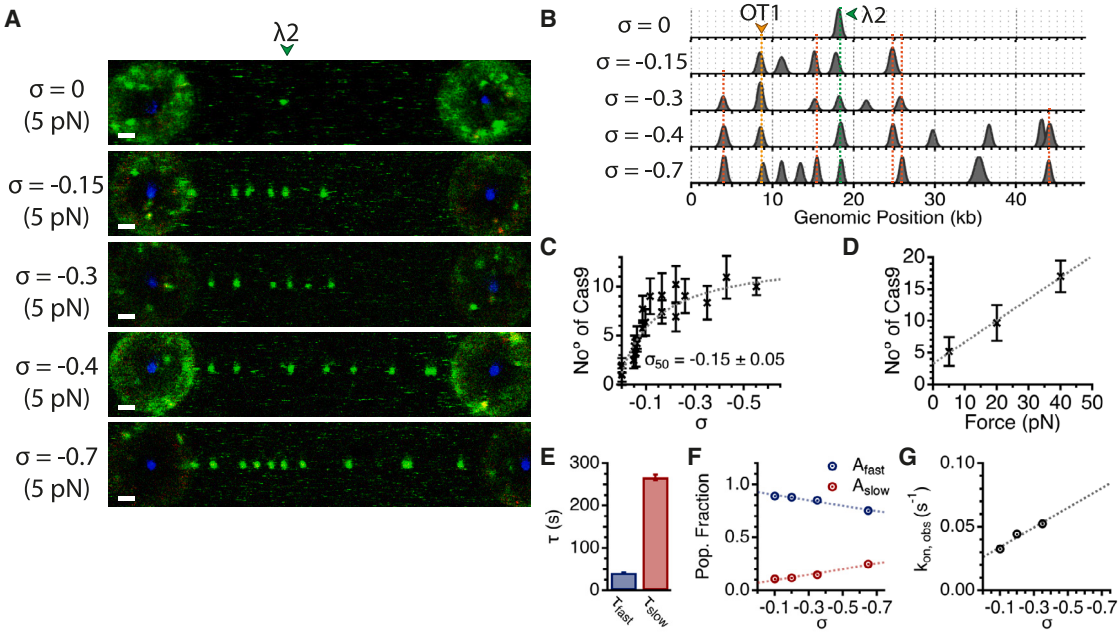


Figure 2. Off-target binding increases with supercoiling density and DNA tension

(A) Representative 2D confocal scans of λ-DNA with supercoiling densities from σ = 0 to σ = −0.7 held at low force (5 pN) in the presence of λ2-crRNA:tracrRNA:dCas9. Scale bars, 1 μm.

(B) Intensity profiles from (A) mapped to the λ genome. The expected on-target site is present at all supercoiling densities (λ2, green arrow). Off-target events are often observed at the same locations at different supercoiling densities (colored vertical lines), including the predicted off-target (OT1, orange arrow).

(C) Plot of the number of dCas9 complexes bound at 5 pN at different supercoiling densities. A binding isotherm is fitted to the data (gray dashed, N_{max} = 13 ± 2, σ_{50} = −0.16 ± 0.05). Each point is the average dCas9 bound across all frames for an individual DNA molecule. n = 10 frames per DNA molecule. Error bars represent the standard deviation.

(D) Plot of the number of Cas9 complexes bound as a function of force for negatively supercoiled DNA. σ = −0.15, n ≥ 3 measurements at each force. Error bars represent the standard deviation.

(E) Time constants for the fast- and slow-dissociating populations, as derived from the fits in Figure S2. Error bars correspond to the errors from global fits.

(F) Plot of the fraction of bound Cas9 complexes in the fast-dissociating and slow-dissociating population as a function of supercoiling density.

(G) Plot of the binding frequency of Cas9 at off-target sites as a function of supercoiling density.

We first investigated the binding of Cas9 to torsionally constrained but relaxed non-supercoiled DNA (σ = 0). The DNA was held at a constant low force of 5 pN in the presence of the catalytically dead Cas9 complex (dCas9, 1 nM), assembled with tracrRNA and a Cy3-labeled crRNA targeting a single site on the λ-DNA, λ2 (Table S1), and imaged by 2D confocal microscopy to detect sites of Cas9 binding. In the presence of Cy3-λ2-crRNA:tracrRNA:dCas9, a single binding event was observed at the expected λ2 binding site, 18.2 kb, one-third of the way along the DNA molecule (Figures 1D and 1E, top; Video S1). This is consistent with previous work demonstrating high specificity of Cas9 on torsionally unconstrained DNA at the same low force.[43]

To investigate whether Cas9 specificity is altered in the presence of negative supercoiling, we used ODS to generate a DNA molecule with a supercoiling density (σ) of −0.3, corresponding to a 30% reduction in the number of helical turns relative to non-supercoiled DNA (Figures 1B and 1C).[40] For comparison, *E. coli* is estimated to maintain an average supercoiling state of σ ∼ −0.07.[45,46] The DNA was then held at 5 pN, and upon incubation with Cy3-λ2-crRNA:tracrRNA:dCas9, we observed binding at the expected on-target site but also additional binding at multiple off-target locations (Figures 1D and 1E, bottom; Video S2).

For all non-supercoiled molecules (n = 3), each imaged over ∼4 min, binding of Cas9 at any off-target DNA site was rare (Figure S1A). When the DNA was negatively supercoiled (σ = −0.3) we observed hundreds of off-target binding events across multiple molecules (n = 3) (Figure S1B; Video S2). Interestingly, we observed an off-target binding event at 8.5 kb (orange arrow, Figures 1E and S1B), which is the only off-target site of the λ2-crRNA within the λ genome predicted by Cas-OFFinder[3] and previously observed as a force-induced off-target binding site on non-supercoiled DNA.[43] Based on these data, we conclude that negative supercoiling induces Cas9 off-target binding.

To further understand how supercoiling alters Cas9 specificity, we monitored Cas9 binding to low-force-stretched λ-DNA with different degrees of negative supercoiling, from σ = 0 to σ = −0.7. Even at a low level of negative supercoiling (σ = −0.15), multiple off-target binding events were observed (Figure 2A). We next mapped and compared the locations of off-target binding events from all different supercoiling states to the λ-DNA genomic position. As expected, the on-target binding location and the previously predicted off-target site at 8.5 kb were observed at all supercoiling states tested. Interestingly, some of the other off-target sites were also repeatedly observed

between different supercoiling states (Figure 2B). These events occurred at the same position as those previously detected for force-induced Cas9 off-target binding,[43] further supporting the idea that these are sequence-driven off-target sites.

To explore the relationship between negative supercoiling and the reduction in Cas9 specificity, we quantified the number of off-target binding events as a function of supercoiling state (Figure 2C). We observed an increase in the number of induced off-target events as negative supercoiling increased, which could be fit to a simple binding isotherm, giving a σ_{50} value of -0.16 ± 0.05 (Figure 2C). This demonstrates that the majority of the reduction in Cas9 specificity occurs at low supercoiling states.

To understand whether the increase in off-target bound Cas9 is driven by increased off-target binding frequency (k_{on}) or increased off-target dwell time (τ), we quantified individual off-target binding events for $\sigma = -0.1$ to -0.65 and constructed cumulative probability distributions of the dwell times (Figures S2C–S2E). A double exponential model was required to fit the data, suggesting the presence of a short-lived population (τ_{fast}) and a longer-lived population (τ_{slow}) (Figure S2E). Fitting with global values for the time constants τ_{fast} (41 ± 1 s) and τ_{slow} (268 ± 7 s) (Figure 2E) revealed that as negative supercoiling increases, there is an increase in the proportion of molecules in the long-lived population (A_{slow}) compared to the short-lived population (A_{fast}) (Figures 2F and S2E). Furthermore, as negative supercoiling is increased from $\sigma = -0.1$ to $\sigma = -0.3$, there is an increase in the apparent binding frequency ($k_{on,obs}$) (Figure 2G).

To compare this to the previously demonstrated loss of Cas9 specificity on force-stretched DNA,[43] we next investigated how force-stretching negatively supercoiled DNA affects Cas9 off-target binding. First, we generated a DNA molecule with a low level of negative supercoiling ($\sigma = -0.15$) and held it at a force of 5 pN. Upon incubation with Cas9, on- and off-target binding events were observed as expected (Figures 2D and S2). Next, we increased the force on the DNA molecule to 20 pN and 40 pN; upon stretching the DNA, we observed an increase in the number of off-target sites (Figures 2D and S2). A linear relationship was observed between DNA tension and the number of induced off-target binding events, demonstrating a combinatory effect of force and negative supercoiling on the loss of Cas9 specificity.

Together, these data demonstrate that increasing levels of negative supercoiling reduces the specificity of Cas9 binding by both increasing the frequency of off-target binding events and also increasing the proportion of long-lived off-target binding events. A dramatic reduction in Cas9 binding specificity occurs even at modest levels of supercoiling with the DNA held at low force; however, this can be further exacerbated by increasing the DNA tension. We hypothesize that this supercoiling-induced loss of Cas9 binding specificity could result in increased off-target cleavage events.

Negative supercoiling induces genome-wide Cas9 off-target cleavage

Based on these findings, we next asked whether negative supercoiling in DNA could induce Cas9 off-target cleavage at sites across the human genome. To test this, we used a modified next-generation sequencing (NGS) approach based on the previously described CIRCLE-seq method for detection of genome-wide Cas9 off-target cleavage sites.[15] First, we purified and fragmented genomic DNA from HEK293T cells and ligated the fragments into ~1,000 bp circles. Any remaining linear DNA was degraded by exonucleases to generate a pool of circular DNA molecules with sequences covering the entire human genome.[15] Next, half of the sample was treated with DNA gyrase to negatively supercoil the DNA circles, estimated to generate a negative supercoiling state of $\sigma = -0.09$ (Figure 3A).[47] The two pools of circular DNA, relaxed (no gyrase) and negatively supercoiled (with gyrase), were then incubated with Cas9 complex. Any site of Cas9 cleavage resulted in linearization of the DNA circles, to which sequencing adapters are ligated, allowing the sites of Cas9 cleavage to be determined by NGS (Figure 3A).

We first used a Cas9 complex targeting the previously characterized EMX1-1 locus (EXM1-1:Cas9; Table S1).[15,48–50] In the presence of relaxed circular DNA, 638 off-target cleavage sites were detected across the human genome compared to untreated controls (see dataset in Key resources table). The detected off-target sites were consistent with previous CIRCLE-seq experiments performed with this guide sequence.[15] Strikingly, when EMX1-1:Cas9 was incubated with gyrase-treated negatively supercoiled genomic DNA, over 13,000 off-target sites were detected (Figure 3B; see dataset in Key resources table). The majority of off-target sites detected in the absence of supercoiling were also detected when the DNA was negatively supercoiled (n = 460/638) (Figure 3B). Importantly, to exclude the possibility that some of the double-strand breaks (DSBs) could be generated by gyrase treatment, we compared a gyrase-treated no-Cas9 control with an untreated no-Cas9 control. Only 101 cleaved sites were detected in the presence of gyrase alone, of which only 6 contained $\leq$7 mismatches. This demonstrates that the additional ~12,500 sites observed for gyrase-treated DNA in the presence of Cas9 are bona fide Cas9-induced cleavage sites (see dataset in Key resources table).

To further verify these results, we repeated the experiment with a second previously characterized and less promiscuous guide (EMX1-2; Table S1), which targets an alternative site in the EMX1 locus.[49] In the presence of relaxed genomic DNA, only 48 off-target cleavage sites were detected across the human genome (Figure 3B; see dataset in Key resources table). However, upon incubation of gyrase-treated negatively supercoiled DNA with Cas9, over 14,000 off-target cleavage sites were detected (Figure 3B; see dataset in Key resources table). Again, we compared the DSBs present in the gyrase-treated no-Cas9 control with the untreated no-Cas9 control. Only 12 sites were detected, of which only 1 contained $\leq$7 mismatches, confirming that the 14,000 sites observed for gyrase-treated DNA in the presence of Cas9 are bona fide Cas9-induced cleavage sites (see dataset in Key resources table). These striking results demonstrate how the reduction in Cas9 specificity induced by negative supercoiling has the potential to allow off-target cleavage at thousands of off-target locations across the human genome.

Genome-wide off-target sites are induced by increased mismatch tolerance

To further understand the nature of the loss of Cas9 specificity, we interrogated the sequences of detected off-target sites for

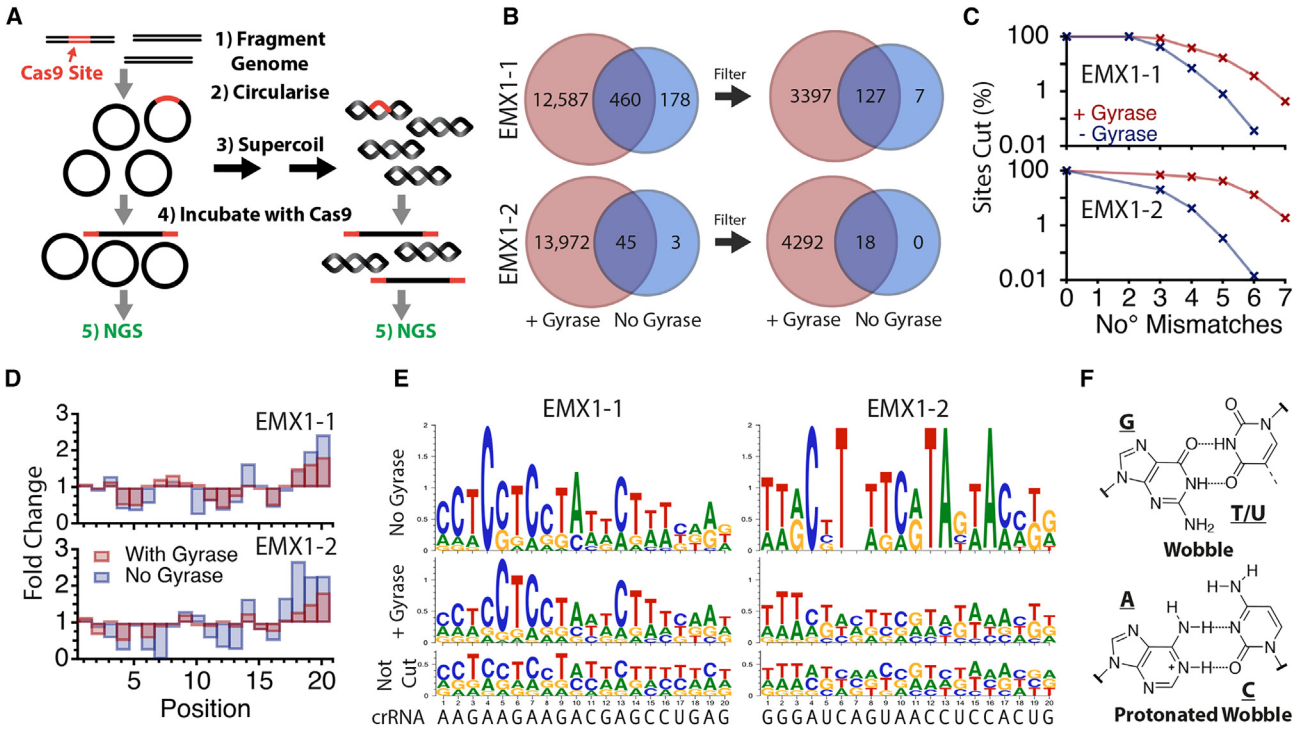


Figure 3. Negative supercoiling induces genome-wide off-targets

(A) Overview of the method for CIRCLE-seq analysis with negative supercoiling.

(B) Number of off-target Cas9 cleavage sites identified in the non-supercoiled (no gyrase) and negatively supercoiled (+gyrase) case for the EMX1-1 and EMX1-2 crRNA guide sequence across the human genome before and after filtering.

(C) The percentage of all potential genome-wide off-target sites with ≤7 mismatches cut by Cas9 targeted with EMX1-1 and EMX1-2 in non-supercoiled DNA (no gyrase) and negatively supercoiled DNA (+gyrase).

(D) The fold change in mismatch frequencies at each position along the target sequence for genome-wide Cas9 off-target cleavage events compared to the uncut sites for EMX1-1 and EMX1-2 in the presence of relaxed DNA (no gyrase, blue), and negatively supercoiled DNA (with gyrase, red).

(E) Mismatch sequence motifs of the mismatching bases present in the Cas9 off-target cleavage sites for non-supercoiled DNA (no gyrase) and negatively supercoiled DNA (+gyrase) alongside the uncut sites (not cut) for EMX1-1 and EMX1-2 crRNA guided complexes.

(F) Base pairing configuration of G-T/U wobble base pair and A-C protonated wobble base pair.

both EMX1-1 and EMX1-2 guides in the context of all the possible off-target sites present in the human genome. First, we performed an exhaustive genome-wide search of all potential Cas9 off-target sites. Using CAS-OFFinder,[3] we generated a comparative list of all possible sites with ≤7 mismatches utilizing the 5'-NRG-3' (where R is G or A) PAM sequence motif and not requiring accommodation of bulges in the target DNA of crRNA. Although there is evidence of Cas9 being able to accommodate bulges by base skipping,[38] DNA and RNA bulges were necessarily excluded from this analysis to avoid identifying an unmanageable number of sites. However, even in their absence, over 250,000 sites were identified for the EMX1-1 guide sequence and 100,000 sites for EMX1-2. As the number of mismatches is increased, the number of off-targets detected increases by orders of magnitude (Figure S3A).

Next, we filtered the off-target cleavage sites detected by CIRCLE-seq in the presence and absence of supercoiling using the same parameters as the CAS-OFFinder search, retaining only the detected off-target sites containing ≤7 mismatches with NRG PAM motifs and removing any sites requiring accommodation of DNA or RNA bulges (Figure S3B).

We first asked what percentage of all possible genome-wide sites were cut by Cas9 in the presence and absence of supercoiling for a given number of mismatches between the guide and off-target site (Figure 3C). With the EMX1-1 guide, 100% of sites with <3 mismatches are cut under both conditions. However, at sites with ≥3 mismatches, negative supercoiling increases the percentage of sites cut by Cas9. For example, in the presence of relaxed genomic DNA, EMX1-1:Cas9 cleaves only 44% of all possible off-target sites with 3 mismatches, but when the DNA is negatively supercoiled, 85% of these sites are cleaved (Figure 3C). A substantial increase in the percentage of cleaved off-target sites was observed when the DNA is negatively supercoiled for all sites with ≥3 mismatches for both the EMX1-1- and EMX1-2-guided Cas9 (Figure 3C). As the number of mismatches increases, the percentage of the total off-target sites that are cut by Cas9 sharply decreases for both relaxed and negatively supercoiled DNA (Figure 3C). However, as the number of mismatches increases, the number of potential off-target sites genome-wide increases exponentially for both EMX1-1 and EMX1-2 (Figure S3A). As a result, when DNA is negatively supercoiled, over 2,000 genomic off-target sites

Table 1. EMX1-1 target site sequences for plasmid cleavage experiments

Name	Sequence	Mismatches
ON	*CCC* TTCTTCTTCTGCTCGGACTC	0
OT1	*CCA* TTCTTC**CTT**TGCTCGG**G**CT**T**	4
OT2	*CCC* TTCTTCT**C**CTGCTC**CA**ACTT	3
OT3	*CCC* TT**A**TTCT**C**CTGCT**TA**GACTC	4
OT4	*CCC* TTCTT**T**TT**G**TGCT**T**GG**G**CTC	4
OT5	*CTC* TTCT**C**CT**C**CTGCT**G**GGA**GAA**	6
OT6	*CCC* TT**A**TTCT**C**CTGC**CT**GGA**GAA**	7
OT7	*CCT* **A**T**AC**TCTTCTGCTC**T**GA**TT**G	6
OT8	*CCC* TT**A**TT**G**TT**T**TGCT**T**GGA**AT**G	6
OT9	*CTC* TT**T**TTCT**CT**TGCT**T**GGA**AT**A	6

The target strand is shown. ON represents on-target; OT denotes off-target; PAM site in italics; mismatches bolded.

with as many as 6 or 7 mismatches are cut by Cas9 (Figures S3B and S3C; see dataset in Key resources table). Together, these data demonstrate that negative supercoiling dramatically increases the number of genome-wide off-target sites cut by Cas9 due to an increase in mismatch tolerance.

Negative supercoiling increases mismatch tolerance through reduced stringency of mismatch distribution and type

To understand the factors that determine which sites can or cannot be cut by Cas9 in the presence and absence of supercoiling, we further analyzed the sequence features of the detected cleaved off-targets in comparison to the uncleaved off-targets present in the genome. We first compared the distribution of mismatches across the cut and uncut genome-wide off-target sites. For the off-target sites cut in the absence and presence of negative supercoiling, we calculated the probability of each position along the 20 nt target site being mismatched. Using the calculated probability of mismatches at each position in the list of all potential off-target sequences, we determined the fold-change in mismatch frequency between the uncut and cut off-targets in the absence and presence of negative supercoiling (Figure 3D). For both EMX1-1 and EMX1-2 off-target sites, where cleavage was detected in the presence of relaxed DNA (no gyrase), we observed an increased probability of mismatches occurring at the PAM distal end of the target site compared to the uncut sites (Figure 3D, blue), consistent with previous work.[14,49,51] Interestingly, when we considered the off-target sites cleaved in the presence of negatively supercoiled DNA (with gyrase), we observed a reduction in the prevalence of PAM distal mismatches (Figure 3D, red). This suggests that when DNA is negatively supercoiled, Cas9 is more tolerant of mismatches spread throughout the target site.

To further understand the changes in mismatch tolerance, we next considered the identities of mismatches found in the sites where cleavage was observed in the absence and presence of negative supercoiling and calculated sequence motifs of the mismatch identities (Figure 3E). As a control, we first calculated the mismatch sequence motif for the uncut genome-wide potential off-targets. As expected, no strong enrichment of particular mismatch types was observed among the uncut off-target sequences for EMX1-1 or EMX1-2 (Figure 3E, bottom). In contrast, when we considered the off-target sequences cleaved by Cas9 in the absence of supercoiling, we observed a strong enrichment of particular types of mismatch (Figure 3E, top). However, in the presence of negative supercoiling, we observed a reduction in the enrichment of particular types of mismatches (Figure 3E, middle). This further demonstrates that the presence of negative supercoiling reduces the stringency of requirements for specific off-target mismatch identities.

We next considered the types of mismatches that preferentially allow Cas9 cleavage at off-target sites in the absence of supercoiling. Interestingly, we observed a preference for mismatches with the potential to result in wobble base pairing. For instance, at those positions in the crRNA where guanine is present, thymine is observed as the favored mismatch, and at positions where adenine is present, mismatches with cytosine are highly favored (Figure 3E, EMX1-1 and EMX1-2 top), both resulting in the potential for formation of a wobble base pair (Figure 3F). The formation of wobble base pairing during off-target Cas9 binding is consistent with off-targets detected by CHANGE-seq[52] and recent high-resolution crystal structures of Cas9 bound to off-target sites where wobble base pairing was observed to be accommodated within the bound complex.[38] We next asked how supercoiling affects the mismatch preferences. When we analyzed the mismatch identities in the negatively supercoiled off-targets, we detected a reduction in the enrichment of the wobble base pairing identities. This demonstrates an increased tolerance of mismatches unable to form wobble base pairing.

Taken together, these data demonstrate that negative supercoiling increases the number of Cas9 off-target sites through reduced stringency in the positions of mismatches across the target sequence and through accommodation of more unfavored mismatches.

Off-target cleavage kinetics with unfavored mismatches

To further understand how supercoiling affects the ability of Cas9 to tolerate unfavored mismatches at off-target sites, we performed detailed kinetic analysis of a number of off-target sites identified by CIRCLE-seq (Table 1). Using purified supercoiled plasmids containing the EMX1-1 on-target and four off-target sites (OT1–OT4), each containing 4 mismatches (Figures 4 and S4; Table 1), we performed bulk kinetic cleavage time course experiments from 20 s to 30 min. The generation of nicked and fully cleaved plasmid products was analyzed by agarose gel electrophoresis (Figure 4). First, we tested the cleavage of the on-target EMX1-1 site. Before the first time point (20 s), we observed full double-strand cleavage with no detectable nicked intermediate (Figure 4A). This defines the lower limit of the on-target cleavage rate constant of Cas9 under our experimental conditions ($k \gg 3\ min^{-1}$). Next, we tested the four EMX1-1 off-target sites. With all four off-targets tested, accumulation of fully linearized plasmids could be observed over multiple time points, as well as the formation of nicked intermediates (Figures 4B, 4C, and S4A–S4D). As such, we were able to directly measure the rate of first strand cleavage (k_1) and second strand

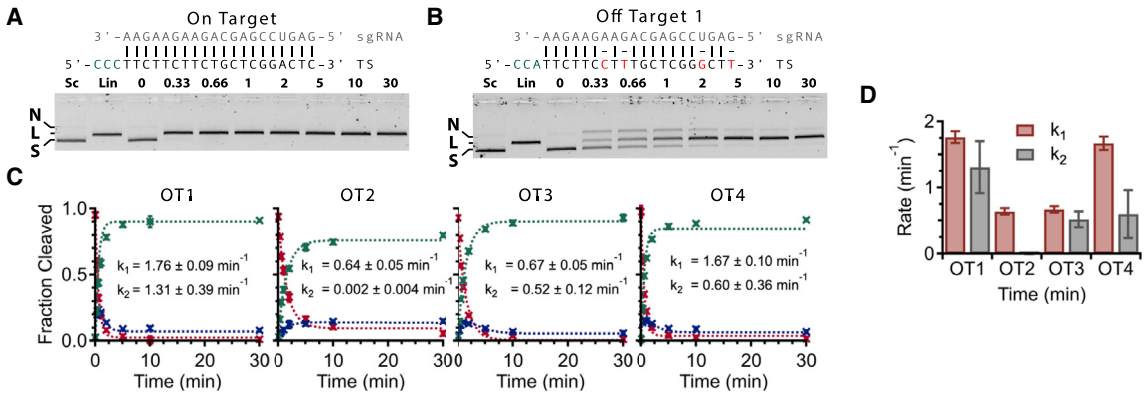


Figure 4. Negative supercoiling induces off-target cleavage *in vitro*

(A) EMX1-1 on-target representative time course agarose gel showing supercoiled (S), nicked (N), and linear (L) products over a 30 min period for supercoiled plasmid cleavage.

(B) EMX1-1 Off-target 1 (OT1) representative time course agarose gel for supercoiled plasmid cleavage.

(C) Quantification of cleavage products: supercoiled (red), nicked (blue), and linear (green) over 30 min for Off-targets 1–4 (OT1–OT4). Error bars = SEM. k_1 and k_2 are the rates of the first and second strand cleavage, respectively.

(D) Cleavage rates k_1 (red) and k_2 (gray) for OT1–OT4. Error bars represent error from fits in (C).

cleavage (k_2) for the four off-target sequences (Figures 4C and 4D). Control experiments with linearized plasmids show no off-target cutting (Figure S4E).

This demonstrates that although Cas9 cleaves supercoiled off-targets with high efficiency, the rate of cleavage is slower than at on-target sites. Furthermore, the presence of mismatches in the supercoiled off-targets can differentially modulate both first and second strand cleavage. For example, when comparing Off-target 1 (OT1) and Off-target 4 (OT4), the first strand cleavage rates are comparable; however, the rate of second strand cleavage for OT4 is significantly lower (Figure 4D). Conversely, although the rate of first strand cleavage for OT3 is significantly lower than for OT4, the second strand cleave rates are comparable.

We also generated plasmids containing five of the off-targets with 6–7 mismatches that had been detected by CIRCLE-seq (Table 1, OT5–OT9) and tested *in vitro* cleavage of relaxed and negatively supercoiled plasmids. Consistent with the CIRCLE-seq data, cleavage was only observed on negatively supercoiled DNA (Figure S4F). Despite similar read counts from the CIRCLE-seq data, cleavage efficiencies ranged from low efficiency (OT5, OT6, OT8, and OT9) to relatively high efficiency (OT7) (Figure S4F).

Together, these data suggest that Cas9 retains a degree of kinetic discrimination between negatively supercoiled off-targets, which can differentially alter first and second strand cleavage rates, and validates the ability of Cas9 to cleave off-target sites with high numbers of mismatches when the DNA is negatively supercoiled.

DNA structural distortion enhances Cas9 off-targeting in cells

We have demonstrated through multiple complementary approaches that negative supercoiling in DNA dramatically reduces Cas9 specificity *in vitro*. Next, we asked whether changes in Cas9 specificity could be observed *in vivo*, in cells, upon perturbation of the local DNA topology. To address this question, we used an adapted proxy-Cas9 system.[53] Proxy-Cas9 approaches involve targeting non-cleavage-competent Cas9 complexes to sites proximal to the intended Cas9 cleavage site (Figure 5A). It has been demonstrated that in some instances, nearby binding of such Cas9 complexes (proxy-CRISPR) can increase the cleavage efficiency of Cas9 at the on-target site.[53] However, this is not universally observed at all sites, and the mechanism underlying this increased efficiency is still not well understood.[53]

We hypothesized that the presence of such proximal Cas9 binding causes local DNA distortion and alters the tolerance of mismatches at neighboring Cas9 target sites. To test this, we utilized the Traffic-Light DSB repair reporter (TLR) system, which enables single-cell quantitative detection of Cas9 cutting activity through a fluorescent readout (Figure 5A).[55] We first validated the TLR system in U2OS TLR cells; cells were transfected by electroporation with a single guide RNA (sgRNA):Cas9 complex containing a fully matching sgRNA targeting the TLR cassette. This fully matching sgRNA resulted in an observed editing efficiency of 35% (Figure 5B). Next, we measured the editing efficiency with guides containing each of the three possible single nucleotide mismatches at positions 1, 5, 7, 8, or 9 along the target sequence. The observed editing efficiencies were diminished overall, ranged from 1% to 30% of the population, and were highly dependent on both mismatch identity and position along the guide (Figure 5B).

We then asked whether deforming the local DNA topology alters the editing efficiency of mismatched Cas9 complexes. We adapted the proxy-CRISPR system using the cleavage-competent SpCas9 with short sgRNA (14 nt proxy-guide RNA [gRNA]), which can still guide Cas9 to the specific target but prevents the nuclease cleavage activity.[12] We first tested a number of different proxy-gRNAs positioned at different positions relative to the main site. Cas9 targeting the TLR but with a mismatch at position 5 was co-transfected with a proxy-CRISPR targeting one of seven sites positioned 18 to 63 bp from the main site

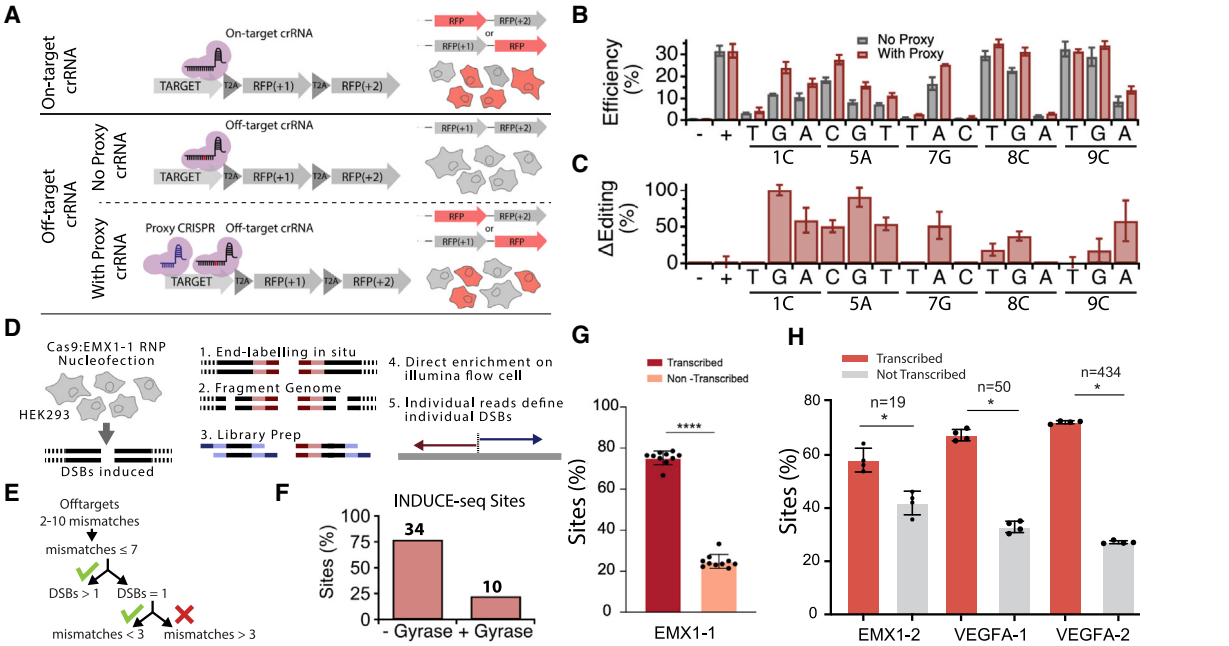


Figure 5. DNA distortion induces off-target activity in cells

(A) Schematic of Traffic-Light Reporter (TLR) system and proxy-CRISPR. Cas9-induced DSB is repaired by non-homologous end-joining, generating a frameshift that places the red fluorescent protein (RFP) coding sequence in-frame, providing a redout for cleavage efficiency.

(B) Cas9 off-target editing efficiency in TLR U2OS cells. Cas9 only (−) shows no cutting, and proxy-CRISPR has no effect. Cas9 complexed with fully matching guide (+) shows 35% editing efficiency with and without proxy-CRISPR co-transfection. Cas9 complexed with mismatched guides (1 mismatch at positions 1, 5, 7, 8, and 9 from the PAM) show enhanced editing efficiency with proxy-CRISPR co-transfection (red bars). Error bars show standard deviation.

(C) Percentage change in off-target editing efficiency with proxy-CRISPR co-transfection. Error bars show standard deviation.

(D) Overview of the INDUCE-seq method for detection of Cas9-induced DSBs.[54] A non-transfected control (NTC) is used to quantify endogenous breaks.

(E) Identified sites of DSBs are filtered based on the number of mismatches and the number of times DSBs are detected.

(F) Percentage of unique sites detected by INDUCE-seq that were detected by CIRCLE-seq without supercoiling (−gyrase) or with supercoiling (+gyrase). Absolute numbers indicated.

(G) Percentage of sites detected by INDUCE-seq with EMX1-1 guide that are located in transcribed or non-transcribed regions. Error bars show standard deviation; $p \leq 0.00001$.

(H) Percentage of sites detected by INDUCE-seq with EMX1-2, VEGFA site 1, and VEGFA site 2 guides located in transcribed or non-transcribed regions. Error bars show standard deviation; $p \leq 0.05$.

(Table 2). In the absence of the proxy-CRISPR, an editing efficiency of <20% was observed (Figure S5A); however, the presence of the proxy-CRISPR significantly increased editing efficiency for 5/7 of the proxy-gRNAs tested (Figure S5A). Based on the level of editing enhancement, we selected proxy-gRNA P1, 37 bp away from the main site, for further experiments.

We next characterized the effect of this proxy-CRISPR on editing efficiency in the presence and absence of different mismatches (Figure 5B). In the absence of mismatches, no change in editing efficiency was observed on co-transfection with proxy-CRISPR P1 (Figures 5B and 5C). However, on co-transfection of Cas9 complexes containing single mismatches with proxy-CRISPR P1, a significant increase in editing efficiency was observed for 10 of the 15 single mismatch positions and identities tested (Figures 5B and 5C). In many instances, the proxy-CRISPR restored editing efficiencies to, or close to, the levels observed in the absence of mismatches, increasing editing efficiencies with mismatched guides by as much as 100% (Figure 5C). Together, these data suggest that *in vivo*, local DNA deformation can reduce Cas9 specificity by enabling an increase in mismatch tolerance.

Supercoiling-induced Cas9 off-targets can be directly detected in cellular gene editing

We next asked whether it was possible to directly detect the presence of supercoiling-induced off-targets in cellular gene editing experiments. Recently, INDUCE-seq has been developed as a highly sensitive method for *de novo* DSB detection in cells (Figure 5D).[54] This PCR-free method is highly sensitive for detection and quantification of DSBs present in a population of cells at a specific time point. To validate this method, the authors nucleofected cells with EMX1 sgRNA/Cas9 ribonucleoprotein (RNP) complexes and performed INDUCE-seq to detect Cas9-induced on- and off-target DSBs. The EMX1-1 guide used in our CIRCLE-seq experiments was also used in the INDUCE-seq cellular experiments. We therefore asked whether cleavage was observed in cells at any of the supercoiling-induced off-target sites we have identified.

Table 2. Table of proxy-gRNAs used in proxy-CRISPR experiments

GUIDE	Sequence (5′ to 3′)	PAM	Distance (bp)	Guide orientation	Result
P1	CACGGGCAGCTTGC	NGG	37	Upstream	Optimal
P2	GACGAGGGTGGGCC	NGG	19	Upstream	Too close
P3	TGACGAGGGTGGGCC	NGG	18	Upstream	Too close
P4	TCCGCCATGCCCGA	NGG	63	Downstream	Too far
P5	CTGGACGTAGCCTT	NGG	59	Downstream	Too far
P6	TCTGATAAAAGCAGC	ACG	46	Downstream	Near optimal
P7	AGCACGACTTCTTC	AAG	34	Downstream	A bit too close

Includes targeted sequence, PAM site sequence, distance from main site, orientation, and result in editing enhancement experiments.

INDUCE-seq is sensitive enough to detect individual endogenous DSBs[54]; however, to avoid incorrect assignment of endogenous breaks as Cas9-induced, and therefore avoid false positive detection, we performed stringent filtering steps (Figure 5E). Firstly, detected DSBs were only classified as Cas9-induced if they contained up to 7 mismatches. Next, sites where only a single DSB was detected were only kept if they contained less than 3 mismatches (Figure 5E). This allows us to confidently differentiate between endogenous and Cas9-induced DSBs (Figures S5B and S5C).

Using this pre-existing INDUCE-seq dataset, we identified Cas9-induced DSBs in cells at 44 off-target sites that were also identified in our *in vitro* CIRCLE-seq off-target detection experiments (Figure S5D). Strikingly, ~23% of these sites (10/44) were only detected *in vitro* in the presence of negative supercoiling (Figure 5F). This suggests that negative supercoiling, or alternative DNA structural distortions, is occurring in the cellular context and driving Cas9 off-target cleavage at these sites. As with the off-target sites detected by CIRCLE-seq, the EMX1-1 off-targets detected by INDUCE-seq showed a strong preference for mismatches resulting in wobble base pairing (Figure 5E).

Next, to understand the potential cellular source of DNA structural distortion, we asked whether these Cas9-induced off-target DSBs fall into transcribed or non-transcribed genomic regions. Previous genome-wide studies of supercoiling in mammalian cells have demonstrated increased negative supercoiling around transcription start sites at distances as far as ±10 kb.[56,57] Over 70% of the EMX1-1 off-target sites detected by INDUCE-seq were found to occur in transcribed regions (Figures 5G and S5F).

To ensure this was not a guide-specific effect, we performed additional INDUCE-seq experiments with three distinct guides, EMX1-2, VEGFA site 1, and VEGFA site 2. With these new guides, at a single time point 7 h after Cas9 nucleofection, we detected 19, 50, and 434 Cas9 off-target sites, respectively. As observed for the EMX1-1 dataset, a significant preference was observed for off-target cleavage in transcribed versus non-transcribed regions (Figure 5F). An increased propensity for off-target activity in highly transcribed regions has previously been suggested; however, this has been largely attributed to the presence of open chromatin.[52]

While chromatin state is likely to also be an important factor, our detection of off-targets *in vivo* that were only found by CIRCLE-seq when the DNA is negatively supercoiled points to a role in transcription-driven DNA distortion or supercoiling in the induction of off-target activity in actively transcribed regions.

DISCUSSION

Our single-molecule optical-tweezers and biochemical cleavage data demonstrate that negative supercoiling increases both Cas9 off-target binding (Figures 1 and 2) and cutting (Figure 4). Furthermore, our CIRCLE-seq experiments show that negative supercoiling induces tens of thousands of genome-wide off-target sites, due to a tolerance for higher numbers of mismatches and less energetically favorable mismatches (Figure 3). Ensemble cleavage assays with supercoiled plasmids show that despite high cleavage efficiency, off-target cleavage kinetics are slower than on-target (Figure 4). Interestingly, the presence of mismatches at these off-targets can alter both first and second strand cleavage kinetics. Finally, using two distinct approaches, we demonstrate that DNA distortion can modulate Cas9 specificity in the environment of the cell. Firstly, we show evidence with the TLR system that specifically distorting the DNA topology in the vicinity of a target site modulates off-target activity (Figure 5). Secondly, using INDUCE-seq data, we demonstrate that off-target cleavage is observed in cellular editing experiments at sites that require negative supercoiling for cleavage *in vitro* (Figure 5).

This demonstration that the local supercoiling density of the DNA modulates off-target activity predicts that regions of the genome that are predominantly negatively supercoiled will result in higher off-target rates in cells. Our findings suggest that the threshold of the conformational checkpoint required for Cas9 cleavage can be altered by the target DNA topology. This might be achieved by modulating the allosteric networks that regulate target binding, R-loop formation, and cleavage.[33] Our kinetic data suggest that the reduction in Cas9 specificity involves an increase in off-target binding frequency and increased population of long-lived off-target binding events (Figure 2). However, supercoiling has also been shown to alter cleavage efficiency itself, and this could additionally contribute to increased off-target cleavage.[44]

In the context of previously described kinetic models for Cas9 specificity,[26,27] this can be described as a tilting of the energy landscape for binding and cleavage. As such, a kinetic model whose energy landscape can be modulated to take into account the tilting caused by supercoiling or mechanical stretching might be more accurate in predicting off-targets. In genome-wide studies of supercoiling in mammalian and other eukaryotic genomes, it has been shown that there are increased levels of negative supercoiling around transcription start sites and that highly expressed genes are more negatively supercoiled than other regions.[56–58] This correlates well with genome-wide

analyses that show Cas9 binding specifically at gene proximal regions (including gene promoters up to 1 kb upstream of the transcription start site).[59] There have been previous reports of increased Cas9 off-target cleavage at sites with high levels of transcription[52]; however, this effect has been largely attributed to changes in chromatin state.

The highly variable specificity of Cas9 depending on DNA topology and the genomic context is particularly important to understand when considering therapeutic application at the whole-organism level. While guides may be screened for potential off-target activity in cellular assays, this may not adequately predict the potential for off-target activity across all the cell types and genomic states present in an entire organism.

The use of non-canonical wobble base pairing for Cas9 off-target interactions is observed in both the CIRCLE-seq and TLR experiments. While this is consistent with recent Cas9 structures[38] and was incorporated into our previous thermodynamic model to identify force-induced off-target sites,[43] these have not yet been incorporated into algorithms for the prediction of potential off-target sites *in vivo*. More surprisingly, purine-purine mismatches are observed to be accommodated in these structures[38] and were also observed to be particularly well-tolerated in our TLR experiments (Figure 5). Future prediction algorithms should incorporate these findings to better predict potential off-target sites.

Finally, it is interesting to consider the fact that CRISPR-Cas9 systems evolved in the context of highly negatively supercoiled bacterial genomes. As such, what are the implications of these findings for their endogenous activity in bacterial innate immunity? In early work to characterize CRISPR-based immunity, it was observed that transfected circular DNA molecules were more vulnerable to degradation than linear molecules, even in the presence of mismatches in the spacer and PAM sequences.[60,61] While the authors offered no explanation at the time, one possible explanation based on our results is that the negative supercoiling present in a plasmid increases the tolerance for mismatches and allows degradation of the plasmid even with imperfect matching between the spacer and the plasmid or non-canonical PAM sequences. It has been argued that CRISPR systems have evolved with an intrinsic tolerance for mismatches between the guide and target to prevent invasive viruses bypassing them through individual base mutations. In summary, our work raises the intriguing possibility that in the endogenous context, when targeting negatively supercoiled DNA, CRISPR systems may be able to accommodate many more mismatches than previously considered.

Limitations of the study

In the optical-tweezers experiments, we use dCas9 to prevent loss of the tethered DNA and therefore cannot determine whether the induced off-target binding sites would actually result in DNA cleavage. To address this, we demonstrate supercoiling-dependent cleavage in the *in vitro* cleavage assays and CIRCLE-seq assays.

A limitation of the CIRCLE-seq experiments is that we use only a single supercoiling state of approximately $\sigma = -0.1$ as induced by treatment with DNA gyrase, and this is not likely to be representative of the average state of the human genome. It would be interesting to look further at how the genome-wide off-targets change across a range of different supercoiling states. Although it is challenging to determine the absolute levels of supercoiling present in mammalian genomes, it can be estimated from the measured relative levels of psoralen binding from *in vitro*[62] and *in vivo*[56] studies that many loci do display a supercoiling density of approximately $\sigma = -0.06$. This shows that our data at $\sigma = -0.1$ approaches supercoiling levels that could be present physiologically.

Furthermore, while we observe tens of thousands of off-targets in our CIRCLE-seq experiments, the vast majority of these sites may not be observable in a population-averaged cellular experiment. This is particularly true if the accessibility of Cas9 to these off-targets is regulated by cellular functions, as different off-targets may be targeted by Cas9 in different cells in a population. We have demonstrated that transcription likely modulates the efficiency of activity at these sites (Figure 5). Nonetheless, it is also likely that other factors, such as the local chromatin state (euchromatin vs. heterochromatin) and the cell-cycle stage, may also have an effect. It will be important to investigate the influence of other cellular processes, and in particular to consider whether regulation of these processes, e.g., transcriptional inhibition, could improve Cas9 specificity.

Future studies could be performed using transcription and topoisomerase inhibitors or though chromatin modulation to further understand the effect of specific cellular processes on Cas9 specificity. However, such experiments are challenging for multiple reasons. Firstly, chemical inhibitors of transcription and topoisomerases are highly toxic and even at low doses can cause DSBs, which increase the background and reduce the sensitivity of off-target detection.[63,64] Secondly, as demonstrated by the CIRCLE-seq data, many of the potential sites contain up to 7 mismatches. When we consider sites with this many mismatches, detection becomes extremely challenging, as events can be spread across such a large number of potential sites and detection of these sites is well beyond the sensitivity of most of the currently used off-target detection methods.[20–22]

We have demonstrated that PCR-free DSB detection methods such as INDUCE-seq may be able to overcome this issue; however, much larger-scale experiments would be required to draw conclusions on how individual off-target sites are affected by perturbing cellular functions. One existing solution to this is through the use of multi-targeting guide RNAs deliberately designed to target a large number of genomic sites.[65] Interestingly, using this approach, it was observed that Cas9 cleavage was more efficient at transcribed regions, consistent with our INDUCE-seq data.[65]

In the TLR experiments, editing efficiency was highly dependent on both the mismatch identity and also the position of the mismatch within the guide (Figure 5). This highlights the complexity of the rules governing Cas9 specificity and the need for additional work to define exactly how the sequence requirements are altered in the context of DNA distortion. Furthermore, the use of proxy-CRISPR is not directly analogous to altering supercoiling state; the fact that we see an increased mismatch tolerance supports the idea that it is causing local DNA distortion, but the exact nature of this in unclear.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.09.008.

ACKNOWLEDGMENTS

We would like to thank Dr. Yi-Fang Wang (MRC-LMS) for discussions and initial help with bioinformatics analysis. We would like to thank Dr. Euan Gordon for provision of purified Cas9 proteins and the cell biology team at AZ for provision of U2OS-TLR cells. We would like to thank Julia Lindgren (AZ) and the MRC-LMS genomics facility for running the sequencing and Mike Firth for his help processing the CIRCLE-seq data.

The Rueda lab is supported by a core grant of the MRC-London Institute of Medical Sciences (UKRIMC-A658-5TY10), a Wellcome Trust Collaborative Grant (206292/Z/17/Z), and BBSRC CASE-studentship (to M.D.N.). M.L. is funded by the LightDyNamics consortium (https://www.lightdynamics.eu/). The King lab is supported by a Wellcome ISSF grant (204841/Z/16/Z) and a Royal Society Research grant (RGS\R2\202043). The Boulton lab is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC0010048), the UK Medical Research Council (FC0010048), and the Wellcome Trust (FC0010048). S.J.B. is also funded by a European Research Council (ERC) Advanced Investigator Grant (TelMetab) and Wellcome Trust Senior Investigator and Collaborative Grants.

AUTHOR CONTRIBUTIONS

Studies were initially conceived by M.D.N., B.J.T., M.E.C., and D.S.R. M.D.N., B.J.T., M.E.C., and D.S.R. were involved in the design of all the experiments. M.D.N. carried out most of the experiments. G.A.K. helped with the supercoiled DNA tweezers experiments. M.L. performed the traffic light experiments *in vivo*. Q.S. performed the plasmid cleavage assays. P.A. and M.M. helped with the CIRCLE-seq experiments, and Y.F.W. helped with the corresponding bioinformatics analysis. N.P. and S.J.B. performed the DISCOVER-seq data analysis. M.D.N. and D.S.R. analyzed the single molecule data. M.D.N. and D.S.R. wrote the initial manuscript draft with input from all the authors.

DECLARATION OF INTERESTS

S.J.B. is co-founder and VP Science Strategy at Artios Pharma Ltd and is on the *Molecular Cell* advisory board. G.A.K. is lead inventor on a patent (EP3737756B1) for the ODS method used to generate negatively supercoiled DNA, which is licensed to LUMICKS B.V.

Received: August 19, 2022
Revised: May 30, 2023
Accepted: September 7, 2023
Published: October 5, 2023

REFERENCES

1. Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633. https://doi.org/10.1371/journal.pone.0124633.
2. Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., and Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics *30*, 1180–1182. https://doi.org/10.1093/bioinformatics/btt764.
3. Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475. https://doi.org/10.1093/bioinformatics/btu048.
4. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495. https://doi.org/10.1038/nature16526.
5. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88. 1979. https://doi.org/10.1126/science.aad5227.
6. Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J.S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun. *9*, 3048. https://doi.org/10.1038/s41467-018-05477-x.
7. Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. Nature *550*, 407–410. https://doi.org/10.1038/nature24268.
8. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., et al. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat. Biotechnol. *36*, 265–271. https://doi.org/10.1038/nbt.4066.
9. Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., and Liu, D.R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature *556*, 57–63. https://doi.org/10.1038/nature26155.
10. Kim, H.Y., Kang, S.J., Jeon, Y., An, J., Park, J., Lee, H.J., Jang, J.-E., Ahn, J., Bang, D., Chung, H.S., et al. (2019). Chimeric crRNAs with 19 DNA residues in the guide region show the retained DNA cleavage activity of Cas9 with potential to improve the specificity. Chem. Commun. *55*, 3552–3555. https://doi.org/10.1039/C8CC08468H.
11. Yin, H., Song, C.-Q., Suresh, S., Kwan, S.-Y., Wu, Q., Walsh, S., Ding, J., Bogorad, R.L., Zhu, L.J., Wolfe, S.A., et al. (2018). Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. Nat. Chem. Biol. *14*, 311–316. https://doi.org/10.1038/nchembio.2559.
12. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol *32*, 279–284. https://doi.org/10.1038/nbt.2808.
13. Rueda, F.O., Bista, M., Newton, M.D., Goeppert, A.U., Cuomo, M.E., Gordon, E., Kröner, F., Read, J.A., Wrigley, J.D., Rueda, D., and Taylor, B.J.M. (2017). Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. Nat. Commun. *8*, 1610. https://doi.org/10.1038/s41467-017-01732-9.
14. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals

RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843. https://doi.org/10.1038/nbt.2673.

15. Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. Nat. Methods *14*, 607–614. https://doi.org/10.1038/nmeth.4278.
16. Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. *33*, 179–186. https://doi.org/10.1038/nbt.3101.
17. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191. https://doi.org/10.1038/nature14299.
18. Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., et al. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat. Methods *10*, 361–365. https://doi.org/10.1038/nmeth.2408.
19. Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L., et al. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat. Commun. *8*, 15058. https://doi.org/10.1038/ncomms15058.
20. Akcakaya, P., Bobbin, M.L., Guo, J.A., Malagon-Lopez, J., Clement, K., Garcia, S.P., Fellows, M.D., Porritt, M.J., Firth, M.A., Carreras, A., et al. (2018). In vivo CRISPR editing with no detectable genome-wide off-target mutations. Nature *561*, 416–419. https://doi.org/10.1038/s41586-018-0500-9.
21. Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., Gradia, S., Vidal, B., Garner, E., Slorach, E.M., et al. (2017). Mapping the genomic landscape of CRISPR–Cas9 cleavage. Nat. Methods *14*, 600–606. https://doi.org/10.1038/nmeth.4284.
22. Wienert, B., Wyman, S.K., Richardson, C.D., Yeh, C.D., Akcakaya, P., Porritt, M.J., Morlock, M., Vu, J.T., Kazane, K.R., Watry, H.L., et al. (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. Science *364*, 286–289. https://doi.org/10.1126/science.aav9023.
23. Abadi, S., Yan, W.X., Amar, D., and Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. PLoS Comput. Biol. *13*, e1005807. https://doi.org/10.1371/journal.pcbi.1005807.
24. Listgarten, J., Weinstein, M., Kleinstiver, B.P., Sousa, A.A., Joung, J.K., Crawford, J., Gao, K., Hoang, L., Elibol, M., Doench, J.G., and Fusi, N. (2018). Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. Nat. Biomed. Eng. *2*, 38–47. https://doi.org/10.1038/s41551-017-0178-6.
25. Lin, J., and Wong, K.C. (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. Bioinformatics *34*, i656–i663. https://doi.org/10.1093/bioinformatics/bty554.
26. Klein, M., Eslami-Mossallam, B., Arroyo, D.G., and Depken, M. (2018). Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. Cell Rep. *22*, 1413–1423. https://doi.org/10.1016/j.celrep.2018.01.045.
27. Eslami-Mossallam, B., Klein, M., Smagt, C.V.D., Sanden, K.V.D., Jones, S.K., Hawkins, J.A., Finkelstein, I.J., and Depken, M. (2022). A kinetic model predicts SpCas9 activity, improves off-target classification, and reveals the physical basis of targeting fidelity. Nat. Commun. *13*, 1367. https://doi.org/10.1038/s41467-022-28994-2.
28. Bravo, J.P.K., Liu, M.S., Hibshman, G.N., Dangerfield, T.L., Jung, K., McCool, R.S., Johnson, K.A., and Taylor, D.W. (2022). Structural basis for mismatch surveillance by CRISPR–Cas9. Nature *603*, 343–347. https://doi.org/10.1038/s41586-022-04470-1.
29. 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. https://doi.org/10.1126/science.1225829.
30. Lim, Y., Bak, S.Y., Sung, K., Jeong, E., Lee, S.H., Kim, J.-S., Bae, S., and Kim, S.K. (2016). Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. Nat. Commun. *7*, 13350. https://doi.org/10.1038/ncomms13350.
31. Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat. Commun. *7*, 12778. https://doi.org/10.1038/ncomms12778.
32. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67. https://doi.org/10.1038/nature13011.
33. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A., and Yildiz, A. (2017). A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci. Adv. *3*, eaao0027. https://doi.org/10.1126/sciadv.aao0027.
34. 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. https://doi.org/10.1038/nature13579.
35. 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. https://doi.org/10.1126/science.1247997.
36. 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. https://doi.org/10.1016/j.cell.2014.02.001.
37. Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477–1481. https://doi.org/10.1126/science.aab1452.
38. Pacesa, M., Lin, C.H., Cléry, A., Saha, A., Arantes, P.R., Bargsten, K., Irby, M.J., Allain, F.H.T., Palermo, G., Cameron, P., et al. (2022). Structural basis for Cas9 off-target activity. Cell *185*, 4067–4081. https://doi.org/10.1016/J.CELL.2022.09.026.
39. 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. https://doi.org/10.1038/nature15544.
40. King, G.A., Burla, F., Peterman, E.J.G., and Wuite, G.J.L. (2019). Supercoiling DNA optically. Proc. Natl. Acad. Sci. USA. *116*, 26534–26539. https://doi.org/10.1073/pnas.1908826116.
41. Singh, D., Wang, Y., Mallon, J., Yang, O., Fei, J., Poddar, A., Ceylan, D., Bailey, S., and Ha, T. (2018). Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. Nat. Struct. Mol. Biol. *25*, 347–354. https://doi.org/10.1038/s41594-018-0051-7.
42. Okafor, I.C., Singh, D., Wang, Y., Jung, M., Wang, H., Mallon, J., Bailey, S., Lee, J.K., and Ha, T. (2019). Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. Nucleic Acids Res. *47*, 11880–11888. https://doi.org/10.1093/nar/gkz1058.
43. Newton, M.D., Taylor, B.J., Driessen, R.P.C., Roos, L., Cvetesic, N., Allyjaun, S., Lenhard, B., Cuomo, M.E., and Rueda, D.S. (2019). DNA stretching induces Cas9 off-target activity. Nat. Struct. Mol. Biol. *26*, 185–192. https://doi.org/10.1038/s41594-019-0188-z.
44. Ivanov, I.E., Wright, A.V., Cofsky, J.C., Aris, K.D.P., Doudna, J.A., and Bryant, Z. (2020). Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. Proc. Natl. Acad. Sci. USA. *117*, 5853–5860. https://doi.org/10.1073/pnas.1913445117.
45. Mayán-Santos, M.D., Martínez-Robles, M.L., Hernández, P., Krimer, D., and Schvartzman, J.B. (2007). DNA is more negatively supercoiled in bacterial plasmids than in minichromosomes isolated from budding yeast. Electrophoresis *28*, 3845–3853. https://doi.org/10.1002/ELPS.200700294.

46. Giaever, G.N., and Wang, J.C. (1988). Supercoiling of intracellular DNA can occur in eukaryotic cells. Cell *55*, 849–856. https://doi.org/10.1016/0092-8674(88)90140-7.

47. Ferrándiz, M.J., Martín-Galiano, A.J., Arnanz, C., Camacho-Soguero, I., Tirado-Vélez, J.M., and De La Campa, A.G. (2016). An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene. Nucleic Acids Res. *44*, 7292–7303. https://doi.org/10.1093/nar/gkw602.

48. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197. https://doi.org/10.1038/nbt.3117.

49. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832. https://doi.org/10.1038/nbt.2647.

50. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826. https://doi.org/10.1038/nbt.2623.

51. Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Res. *42*, 7473–7485. https://doi.org/10.1093/nar/gku402.

52. Lazzarotto, C.R., Malinin, N.L., Li, Y., Zhang, R., Yang, Y., Lee, G., Cowley, E., He, Y., Lan, X., Jividen, K., et al. (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. Nat. Biotechnol. *38*, 1317–1327. 2020. https://doi.org/10.1038/s41587-020-0555-7.

53. Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., and Davis, G.D. (2017). Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat. Commun. *8*, 14958. https://doi.org/10.1038/ncomms14958.

54. Dobbs, F.M., van Eijk, P., Fellows, M.D., Loiacono, L., Nitsch, R., and Reed, S.H. (2022). Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq. Nat. Commun. *13*, 3989. https://doi.org/10.1038/s41467-022-31702-9.

55. Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., and Scharenberg, A.M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. Nat. Methods *8*, 671–676. https://doi.org/10.1038/nmeth.1648.

56. Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat. Struct. Mol. Biol. *20*, 387–395. https://doi.org/10.1038/nsmb.2509.

57. Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat. Struct. Mol. Biol. *20*, 396–403. https://doi.org/10.1038/nsmb.2517.

58. Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. Nature *577*, 701–705. https://doi.org/10.1038/s41586-020-1934-4.

59. O'Geen, H., Henry, I.M., Bhakta, M.S., Meckler, J.F., and Segal, D.J. (2015). A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. *43*, 3389–3404. https://doi.org/10.1093/NAR/GKV137.

60. Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71. https://doi.org/10.1038/nature09523.

61. Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. *190*, 1390–1400. https://doi.org/10.1128/JB.01412-07.

62. Kramer, P.R., Bat, O., and Sinden, R.R. (1999). Measurement of localized DNA supercoiling and topological domain size in eukaryotic cells. Methods Enzymol. *304*, 639–650. https://doi.org/10.1016/S0076-6879(99)04038-0.

63. Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. Chem. Biol. *17*, 421–433. https://doi.org/10.1016/J.CHEMBIOL.2010.04.012.

64. Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? Transcription *2*, 103–108. https://doi.org/10.4161/trns.2.3.16172.

65. Zou, R.S., Marin-Gonzalez, A., Liu, Y., Liu, H.B., Shen, L., Dveirin, R.K., Luo, J.X.J., Kalhor, R., and Ha, T. (2022). Massively parallel genomic perturbations with multi-target CRISPR interrogates Cas9 activity and DNA repair at endogenous sites. Nat. Cell Biol. *24*, 1433–1444. https://doi.org/10.1038/s41556-022-00975-z.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Urea	Sigma-Aldrich	U5378-1KG
10x TBE Buffer (Tris–borate–EDTA)	Thermo Scientific	B52
Acrylamide–bis-acrylamide, 40% solution	Sigma-Aldrich	A9926-100ML
N,N,N′,N′-tetramethylethylenediamine (TEMED)	Sigma-Aldrich	T9281-25ML
10% ammonium persulfate (APS) made up to 10% w/v in ddH_2O.	Sigma-Aldrich	A3678
Formamide	Thermo Scientific	17899
0.5 M EDTA	Millipore	324506-100ML
RNA Elution Buffer: 0.1 mM EDTA, pH 8.0, 0.1% (w/v) sodium dodecyl sulfate (SDS), and 500 mM ammonium acetate.	In House	N/A
λ-DNA	Thermo Fisher	SD0011
T7 RNA Polymerase 50 U/μL	NEB	M0251S
T7 Buffer (10×)	NEB	B9012SVIAL
E. coli IPP 1 U/μL	NEB	M0361S
1 M DTT, 1,4-Dithiothreitol	Roche	10197777001
3,4-Dihydroxybenzoic acid	Merck	99-50-3
Protocatechuate 3,4-Dioxygenase	Merck	P8279-25UN
Bleach, Sodium Hypochlorite NaClO	Honeywell	15685490
0.5% Pluronic F127 made up in PBS and filtered	Sigma-Aldrich	P2443-250G
Streptavidin Coated Polystyrene Particles 0.5% w/v	Spherotech	SVP-40-5
QIAquick PCR Cleanup Kit	Qiagen	28,104
5 mL HisTrap column	GE Healthcare	10571680
5 mL Heparin HiTrap column	GE Healthcare	10298944
Superdex 26/600 column	GE Healthcare	11360342
microTUBE AFA Fiber Snap-Cap	Covaris	PN 520045
Lambda Exonuclease	NEB	M0262S
E. coli Exonuclease I	NEB	M0293S
oSQT1288/5Phos/CGGTGGACCGATGATC/ideoxyU/ ATCGGTCCACCG*T (* indicates phosphorothioate linkage)	IDT	N/A
E. coli DNA Gyrase	Inspiralis	G1001
AMPure XP beads	Beckman Coulter	A63881
KAPA Hyper Prep Kit (Illumina, 96 rxns)	Kapa Biosystems	07962363001
NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1)	NEB	E7600S
Qubit BR dsDNA kit	ThermoFisher	Q32851
Amaxa™ 96-well Shuttle™	Lonza	V4SC-9096
Qiagen Gentra Puregene Cell Kit	Qiagen	158388
Plasmid-Safe ATP-dependent DNase	Cambio Ltd	E3101K
NEBNext Illumina adaptor	NEB	E7601A
Phenol:Chloroform:Isoamyl Alcohol (25:24:1) (Sigma)	Sigma-Aldrich	P2069-100ML
Bacterial and virus strains		
DH5α	In House	N/A
BL21DE3	NEB	C2527H

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Original Confocal Images and Gel Images	This Study	Mendeley Data: https://doi.org/10.17632/4r55hvz87r.1
Data used to plot figures	This Study	Mendeley Data: https://doi.org/10.17632/4r55hvz87r.1
Output from CIRCLE-seq analysis	This Study	Mendeley Data: https://doi.org/10.17632/4r55hvz87r.1
CIRCLE-seq sequencing reads	This Study	NCBI SRA: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1001791
INDUCE-seq sequencing reads	This Study	NCBI SRA: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1001791
Experimental models: cell lines		
HEK293T	MRC LMB Cell Repository	N/A
HEK293	Crick Cell services	N/A
U2OS-TLR	In House, AstraZeneca	N/A
Oligonucleotides		
crRNA RNA Oligonucleotides	IDT	Table S1
DNA Oligos for tracrRNA *in vitro* transcription	IDT	Table S1
Recombinant DNA		
pET24a-Cas9	Previous study[43]	N/A
Software and algorithms		
FIJI	Open source	https://imagej.net/software/fiji/
Igor Pro 9	Wave Metrics	www.wavemetrics.com
Lumicks Pylake	Lumicks Python package	https://lumicks-pylake.readthedocs.io/en/stable/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David Rueda (david.rueda@imperial.ac.uk).

Materials availability

Plasmids, recombinant proteins, DNA substrates and cell lines are available without restriction upon requests, which should be directed to the lead contact.

Data and code availability

- The unprocessed confocal images, gel images and output data from CIRCLE-seq experiments and data used to plot the figures have been deposited to Mendeley and are publicly available as of the date of publication. The DOI is listed in the Key resources table.
- Sequencing read datasets from the CIRCLE-seq sequencing and INDUCE-seq sequencing have been deposited to the SRA database and are publicly available as of the date of publication. The accession number is listed in the Key resources table.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell Culture: The HEK293T cells were obtained from the MRC LMS Cell Repository and were cultured at 37C 5% CO2 in DMEM (11965092, Gibco) and 10% FBS. The HEK293 cells were obtained from the Francis Crick Cell Repository and were cultured at 37C 5% CO2 in DMEM (11965092, Gibco) and 10% FBS. The U2OS TLR cells were obtained from the AstraZeneca Cell Repository and cultured at 37C 5% CO2 in McCoy's Modified 5A, 10% FCS, 1% Glutamax.

METHOD DETAILS

RNA and DNA purification, labeling and annealing

tracrRNA was synthesised by T7 transcription (Table S1). Labeled crRNA guides were synthesized, labeled and HPLC purified by IDT (Table S1), unlabelled crRNA guides were synthesized by IDT (Table S1) and were further purified by 18% denaturing polyacrylamide gel electrophoresis. Equimolar concentrations of crRNA and tracrRNA were pre-annealed by heating to 90°C for 1 min in annealing buffer (100 mN NaCl, 50 mM Tris-HCl pH8 and 1 mM $MgCl_2$) followed by ~10 min cooling at room temperature.

Protein purification, storage and complexing

Cas9 and catalytically dead Cas9 (dCas9; D10A and H840A) were cloned and purified as described previously.[13] Briefly, a synthetic gene coding for Cas9 or dCas9 with an N-terminal 6xHN tag and a C-terminal nucleoplasmin NLS sequence was synthesised and subcloned into pET24a to generate pET24a-Cas9. *E. coli* BL21DE3* transformants of pET24a-Cas9 were selected on LB plates containing 100 μg/mL kanamycin. Large scale cultures were performed by inoculating 750 mL of TB media at 37°C, and the cultures grown at 37°C until OD_{600}~0.6, after which the culture temperature was lowered to 20°C and protein production was induced by the addition of 100 μM IPTG. Incubation was continued overnight. Cas9 was purified from cell lysate via immobilised metal affinity chromatography. Briefly, cells were lysed using an Avestin Emulsiflex C5 in a buffer consisting of 20 mM Tris HCl (pH 7.5), 300 mM NaCl, 10% glycerol, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and 10 mM imidazole, and the lysate clarified by centrifugation at 20,000 × *g* for 20 min. The clarified lysate was loaded onto a 5 mL HisTrap column (GE Healthcare) and, after washing with the same buffer, eluted with a gradient to 300 mM imidazole. After dilution, the Cas9 containing fractions were loaded onto a 5 mL Heparin HiTrap column (GE Healthcare) and eluted with a linear gradient of NaCl from 0.1 to 1 M, in a buffer containing 20 mM Tris HCl pH 7.5, and 10% glycerol. The protein was polished by SEC, on a Superdex 26/600 column (GE Healthcare) in a buffer containing 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol and 1 mM TCEP. Fractions containing Cas9 were pooled and concentrated to 10–20 mg/mL and aliquots flash frozen in liquid nitrogen. For working stocks dCas9 was diluted to 1.6 μm g/μL in 300 mM NaCl, 10 mM Tris HCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol and stored at −20°C.

Single-molecule experiments

For single-molecule experiments, Cas9 was first complexed at 1 μM concentration with a 1:1 ratio of protein to pre-annealed crRNA: tracrRNA at room temperature for 10 min and subsequently diluted to required working concentration with imaging buffer (100 mM NaCl, 50 mM Tris-HCl pH 8, 1 mM $MgCl_2$, 0.2 mg/mL BSA and the PCA/PCD oxygen scavenger system with 5 mM PCA, 100 nM PCD).

Optical-tweezers with confocal microscopy

Optical-tweezer confocal microscopy experiments were performed on the commercially available Lumicks C-trap with integrated confocal microscopy and microfluidics. Protein channels of the microfluidics chip were first passivated with BSA (0.1% w/v in PBS) and Pluronics F128 (0.5% w/v in PBS), both flowed through over a period of 30 min.

Torsionally constrained λ-DNA was generated by ligating biotinylated DNA end-caps (Table S1) to each end of the λ-DNA substrate, as previously described[40] and then tethering the end-capped substrate between two optically-trapped 4.5 μm SPHERO streptavidin coated polystyrene particles at 0.005% w/v using the laminar flow cell. Negatively supercoiled DNA was generated using the Optical DNA Supercoiling (ODS) method, as previously described.[40] In brief, the DNA integrity was first verified by recording force extension curves from 0 to 80 pN: torsionally constrained molecules do not exhibit an overstretching plateau at ~60 pN, in contrast to torsionally unconstrained DNA. Next, the DNA was stretched for a few seconds to high force (~150 pN) before returning to low force (~20 pN). Transient disruption of the biotin-streptavidin interactions occurs at high force, which induces the formation of negative supercoiling. As the DNA becomes supercoiled, the force extension curve changes, displaying an increase in extension at 70 pN, from which the supercoiling density can be determined.[40] The supercoiling density can typically be tuned from $\sigma = 0$ to -0.7 through repeated stretch and retract cycles. Once the desired supercoiling density is obtained, the DNA molecule is held at forces <50 pN to prevent any further supercoiling induction.

For confocal imaging of Cy3 labeled crRNA, an excitation wavelength of 532 nm was used and emission detected in three channels with blue filter 512/25 nm, green filter 585/75 nm and red filter 640 LP.

Mapping kymograph peaks onto the λ-DNA genome

Analysis of the location of binding events was performed as described previously.[43] The fluorescence intensity signal between the two beads was mapped across the length of the known λ-DNA sequence. DNA orientation was determined based on the location of the specific on-target binding. Mapped binned intensities were fit by multipeak fitting using Igor8, and resulting Gaussian fits were plotted for direct comparison.

Target plasmid preparation

Target plasmids generated in this study were based on the pUC19 (Addgene) or pBluescript (Addgene). Complementary single-stranded DNA oligonucleotides containing the target site of interest flanked by restriction sites were ordered from IDT and annealed and inserted into the multiple cloning sites by restriction cloning. Sequences were confirmed by sanger sequencing (Genewiz).

In vitro cleavage assays

tracrRNA and crRNA were heat-annealed in a 1:1 ratio in a 1X annealing buffer (100 mM NaCl, 50 mM Tris-HCl pH 8). The mixture was heated at 95°C for 2 min, and subsequently cooled to room temperature for 10 min before 1:1 incubation with *Sp*Cas9 at room temperature for an additional 10 min to form the Cas9 RNP at 1 μM. The Cas9 RNP was then placed on ice before use in cleavage assays. The Cas9 RNP was mixed with plasmid in a 1:10 ratio, 2 nM and 20 nM of target plasmid and Cas9 RNP respectively. Incubation of plasmid and the Cas9 RNP were done at 37°C in a cleavage buffer containing 100 mM NaCl, 50 mM Tris HCl pH 8, 10 mM $MgCl_2$ over the course of 1 h. Timepoints were taken at 0 s, 20 s, 40 s, 1 min, 2 min, 5 min, 10 min and 30 min intervals and the reactions were stopped by taking equal amounts (10 μL each) of the cleavage reaction, stopping buffer (100 mM Tris-HCl pH 8, 100 mM EDTA, 40% w/v sucrose) and Phenol:Chloroform:Isoamyl Alcohol (25:24:1). The stopped reactions were subsequently vortexed for 5 s and centrifuged at 13,000 rpm for 1 min to separate the aqueous, interface and solvent layers. 10 μL of the reaction from the (top) aqueous layer was taken, loaded onto 1% Agarose gels containing 1X TAE and 1X SybrSafe and ran at 150 V for 30 min. Finally, the gels were imaged with the Cy2 filter and a 5 s exposure time using the Amersham Imager 680.

In vitro cleavage assay kinetics analysis

Imaged gels were visualized using Fiji and four band intensities quantified: supercoiled (S), nicked (N), linear (L) and background. Plasmid cleavage reactions are first order irreversible and consecutive reactions, where:

$$Supercoiled \xrightarrow{k_1} Nicked \xrightarrow{k_2} Linear$$

In order to determine the fraction of each component at each timepoint, the fraction cleaved can be calculated with the following equations:

$$fSupercoiled = \frac{Supercoiled - Background}{Supercoiled+Linear+Nicked - (3 \times Background)}$$

$$fNicked = \frac{Nicked - Background}{Supercoiled+Linear+Nicked - (3 \times Background)}$$

$$fLinear = \frac{Linear - Background}{Supercoiled+Linear+Nicked - (3 \times Background)}$$

The transition between supercoiled starting material and nicked intermediate depends on the first rate constant (k_1) and the subsequent transition between nicked intermediate to final linear product depends on the second rate constant (k_2). The rate constants are calculated using the following equations:

$$S = S_0 e^{-k_1 t} + y0$$

$$N = N_0 \left(\frac{k_1}{k_2 - k_1}\right) \left\{e^{-k_1 t} - e^{-1_2 t}\right\} + y0$$

$$L = L_0 \left(1 - \left(\frac{1}{k_2 - k_1}\right)\right) \left(k_2 e^{-k_1 t} - k_1 e^{-k_2 t}\right) + y0$$

where *S* is supercoiled, *N* is nicked, and *L* is linear plasmid.

CIRCLE-seq with negative supercoiling

Preparation of DNA circles was performed as previously described with a few minor alterations described below.[48] Genomic DNA was extracted from HEK293T cells and mechanically sheared using a Covaris to an average length of 1000 bp. Size distribution was analyzed on a bioanalyser to confirm the correct distribution of fragment sizes of around 1000 bp. The use of 1000 bp fragments ensured the circles were fully relaxed and avoided any potential mechanical strain on the DNA due to the small circle sizes (300 bp) typically used in the CIRCLE-seq protocol.[15] For 600 μg input DNA before fragmentation, approximately 240 μg was recovered. For each condition, 25 μg of DNA is required.

End-repair was then performed using KAPA End Repair Enzyme Mix, followed by A-tailing using KAPA A-tailing Enzyme, ligation of the CIRCLE-seq hairpin adapter (oSQT1288), exonuclease treatment and circularisation as previously described.[48]

After preparation of DNA circles, the sample was split in half and one-half was negatively supercoiled by treatment with DNA gyrase (1 μg DNA circles, 80 μL gyrase reaction buffer, 2 μL DNA gyrase at 5 U/μL, made up to a final volume of 400 μL; 37°C for 1 h). For the

negative control of non-supercoiled circles, the same reaction was performed with the gyrase omitted. Clean-up was performed with 2x AMPure XP beads eluted in 165 μL in 1x TE with beads removed. For each sample, 1 μg of input DNA circles were used in one 400 μL reaction.

Supercoiled and non-supercoiled prepared DNA circles were treated with Cas9 complexed with the guide of interest. First, the tracrRNA and crRNA were annealed (5 μL tracrRNA at 10 μM, 5 μL crRNA at 10 μM, 5 μL Cas9 annealing buffer; 95°C for 2 min, room temperature for >10 min and then placed on ice). Next, the Cas9 nuclease was complexed with the annealed guides (10 μL 10x Cas9 reaction buffer, 9 μL SpCas9 Nuclease at 1 μM, 3 μL annealed crRNA:tracrRNA at 3.3 μM in 45 μL; held at room temperature for 10 min) before addition of the DNA circles and the initiation of the cleavage reaction (45 μL complexed Cas9 mix, 250 ng DNA circles in a total volume of 100 μL; 37°C for 1 h). For the negative controls without Cas9, the same protocol was followed with the Cas9 and annealed guide replaced by ddH2O. Clean-up was performed with 1x AMPure XP beads eluted in 42 μL 1x TE with beads removed.

The samples were then prepared by standard KAPA library preparation for NEBNext sequencing. Samples were indexed with compatible combinations of NEBNext i5 and i7 indexing primers. The concentration of indexed libraries were measured with Qubit BR dsDNA kit and samples run on a bioanalyser to check fragment size. Paired-end 150 bp read length sequencing was performed on an Illumina MiSeq with all samples from one experiment multiplexed. The sample was spiked with 10% PhiX due to potential low sequence diversity of the sample. De-multiplexed FastQC files were then analyzed using the CIRCLE-seq analysis software.[15] To confirm that the detected off-targets were not due to gyrase induced cleavage events, CIRCLE-seq analysis was also performed between the Gyrase, No Cas9 and No Gyrase, No Cas9 samples.

Binding isotherm

The equation of the binding isotherm used in Figure 2C is as follows, where n = number of binding events and σ = supercoiling density:

$$f(\sigma) = n_0 + (n_{max} + n) * \left(\frac{\sigma}{\sigma+\sigma_{50}}\right)$$

Kinetic analysis

To determine off-target binding time constants (τ), the durations of individual off-target binding events were measured and plotted as a normalised cumulative distribution histogram (bin size = 5s). A normalised double exponential with the following equation was used to these data using global parameters for τ_{slow} and τ_{fast} (Figure 2E):

$$f(t) = 1 - \left(A_{slow} \times e^{\left(-\frac{t}{\tau_{slow}}\right)} + \left(1 - A_{slow}\right) \times e^{\left(-\frac{t}{\tau_{fast}}\right)}\right)$$

where A_{slow} and A_{fast} ($A_{fast} = 1 - A_{slow}$), correspond to the respective normalised amplitudes, or population fractions (Figure 2F). To determine binding frequency ($k_{on,obs}$), the time between individual binding events was measured and plotted as a normalised cumulative distribution histogram (bin size = 5s). These were fit with a normalised single exponential with the equation:

$$f(t) = 1 - \left(A \times e^{\left(-\frac{t}{k_{on,obs}}\right)}\right)$$

Traffic light system and PROXY- CRISPR in cells

U2OS-TLR cells were electroporated with Cas9 RNP with on target or mismatched gRNA together targeting the TLR cassette with the PROXY gRNA with 96-well Nucleofector Kit using the Amaxa 96-well Shuttle. Following cleavage by Cas9 at the target site in the TLR, repair by non-homologous end-joining (NHEJ) can result in a frameshift in the cassette leading to expression of red fluorescent protein (RFP).[55] 72 h after electroporation, cells were collected and the percentage of Cas9 edited cells quantified by fluorescence-activated cell sorting (FACS), by counting the number of RFP expressing cells.

INDUCE-seq

The INDUCE-seq experimental and analysis pipelines were previously described.[54] Briefly, HEK293 cells were nucleofected with Cas9 RNP and extracted for INDUCE-seq experiments at 0h, 7h, 12h, 24h, and 30h timepoints. FASTQ (SRA) files were downloaded from NCBI SRA (PRJNA636949), and mapped to the human genome (GRCh38/hg38 assembly) using the BWA MEM algorithm. Low quality alignments (MAPQ scores <30) and alignments mapping to the ends of chromosomes were filtered. To specifically distinguish off-target Cas9 induced DSBs from endogenous DSBs, we only considered off-target sites with ≤7 mismatches. Furthermore at sites where only one DSB was detected, we only classified them as Cas9 induced when they contained ≤3 mismatches (Figure 5E and S5A). p values to compare the number of unique off-target sites in RNP treated vs. non-treated control (NTC) conditions were calculated using Mann-Whitney t-tests. Genomic regions containing DSB sites were classified as transcribed or non-transcribed using the Genomic Features Package in Bioconductor(R).