SUPPLEMENTAL INFORMATION

SUPPLEMENTAL METHODS

smFRET Trajectory Selection
We selected smFRET trajectories that displayed persistent PEC formation (30+ seconds) to generate histograms, autocorrelation values, and transition energies for our analysis. Some of the substrates used also have transient PECs included in the analysis - defined as those that lasted <30 seconds, as both associated and dissociated during our observation (Reid et al., 2017).

smFRET NHEJ Pairing Efficiency
Pairing efficiency is calculated by using a positive control NHE reaction on each PEG-slide (4nt complementary overhang with 5' phosphate and 3' hydroxyl) for normalization. Multiple images from each flow lane are used to calculate the average number of observed molecules in the control and variable. The resulting normalized pairing efficiency from three independent experiments is used to determine the mean and SEM efficiency.

Histograms and Gaussian Histogram Fitting
Histograms were composed of at least 200 molecules for persistent PECs and 50 molecules for transient PECs. SmFRET histograms were then fit with either single or multiple Gaussian peak fitting in OriginLab software. To compare the FWHM of the various histograms, we compared the primary peaks (>80% of the molecules).

Autocorrelation and Energy Calculation of smFRET Trajectories.
Autocorrelation correlates a signal with itself as a function of time, allowing us to extract the underlying transition frequency in our FRET trajectories. This method was used to compare transition frequencies in persistent PECs, and use Boltzmann inversion to calculate the relative energetic stability of the states occupied (Reid et al., 2017; Rothenberg et al., 2008).

Autocorrelation of FRET efficiency was defined as:

$$AC(\tau) = \frac{\langle E_{FRET}(t) E_{FRET}(t+\tau) \rangle_t}{\langle E_{FRET}(t) \rangle_t^{\,2}}$$

Where $AC(\tau)$ is the autocorrelation function of the lagging time τ; $E_{FRET}(t)$ represents the time-trajectory of the obtained FRET efficiency. To calculate the correlation function, we applied the Fourier Transfer algorithm as:

$$AC(\tau) = \frac{1}{\langle E_{FRET}(t) \rangle_t^{\,2}} \mathrm{iFFT}\left(\left| \mathrm{FFT}\left(E_{FRET}(t) \right) \right|^2 \right)$$

Where FFT and iFFT denotes the Fast Fourier Transfer and inverse Fast Fourier Transfer (MathWorks), respectively.

The generated autocorrelation curve can then be fit with either an exponential, multi-exponential, or stretch exponential function in OriginLab software to measure the average transition time for the PEC.

The obtained autocorrelation time constants and dwell times were calculated as previously described (Rothenberg et al., 2008). The energies for the complexes were obtained from the parametric solution

$$E_{tot} = -RT(\ln(1/[t_{corr}]*a])) \text{ or } E_{tot} = -RT(\ln(1/[t_{dwell}]*a]))$$

Where a is the preexponential factor, assumed to be constant for all configurations and $T = 25°C$ (room temperature), and t_{corr}/t_{dwell} are the autocorrelation and dwell times calculated. The error associated with the reported energy calculations are based on the errors from the curve fitting of the autocorrelation and dwell times.

SUPPLEMENTAL FIGURES

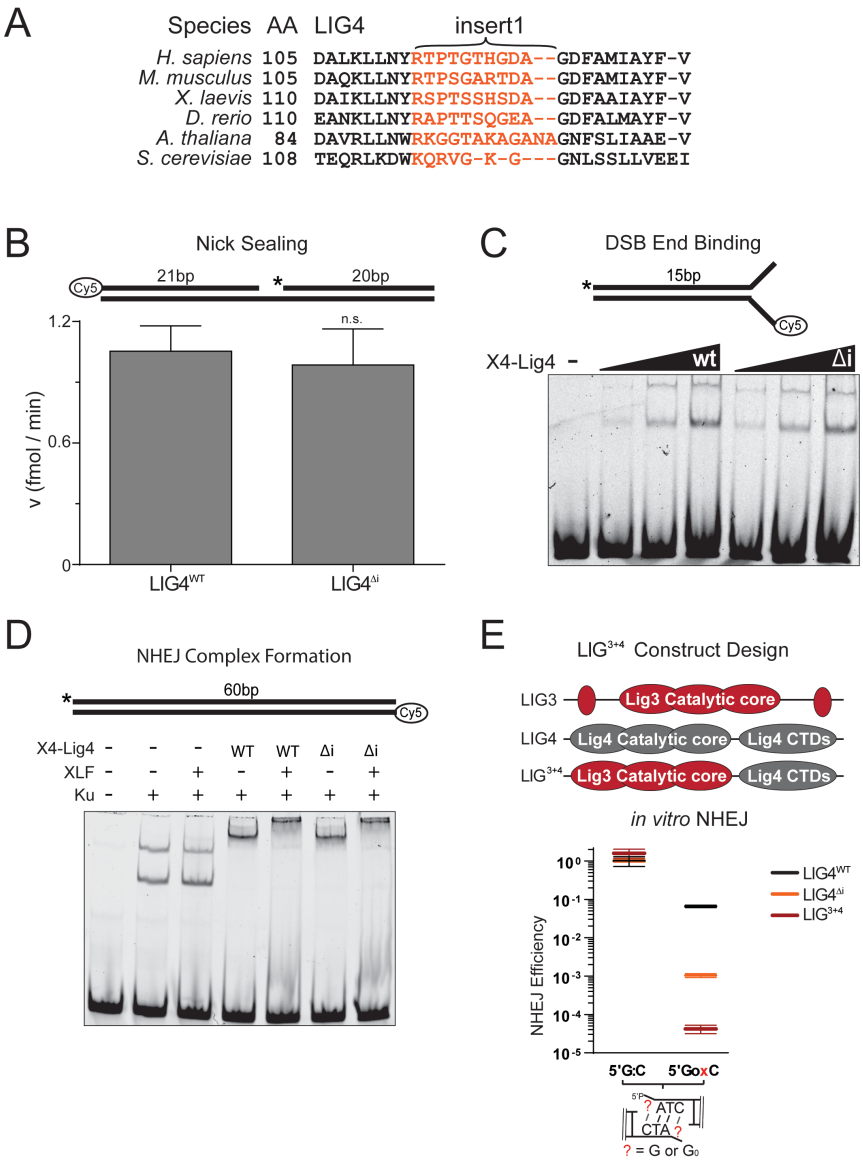

Figure S1. Biochemical characterization of LIG4 variants, Related to Figure 1. **(A)** Amino acid sequence alignment of LIG4 orthologs, with insert1 in orange. **(B)** A nicked, Cy5-labeled 41 bp substrate was incubated with XRCC4-LIG4^{WT} or XRCC4-LIG4$^{\Delta i}$ in triplicate and joining was assessed by stand-denaturing gel electrophoresis. Error bars represent standard error of the mean for 3 experiments. The mean for LIG4$^{\Delta i}$ was assessed by two-tailed t-test as not statistically significantly different (ns) from control (LIG4^{WT}). **(C)** A 15 bp Cy5-labeled substrate was incubated with XRCC4-LIG4^{WT} or XRCC4-LIG4$^{\Delta i}$ and substrate binding was assessed by native gel electrophoresis. **(D)** A 60 bp Cy5 labeled substrate was incubated with indicated NHEJ factors and NHEJ complex formation was assessed by native gel electrophoresis. **(E)** LIG^{3+4} chimera was generated by fusing catalytic domains of LIG3 with C-terminal domains of LIG4 and purified after co-expression with XRCC4. NHEJ reactions were performed *in vitro* as in Figure 1B using undamaged (5' G:C) and damaged (5' GoxC) substrates. Joining efficiency is expressed as a fraction of the total junctions recovered using the 5' G:C substrate with LIG4^{WT}. Data represent the means of 3 ligation reactions, and error bars represent the range of observed data points.

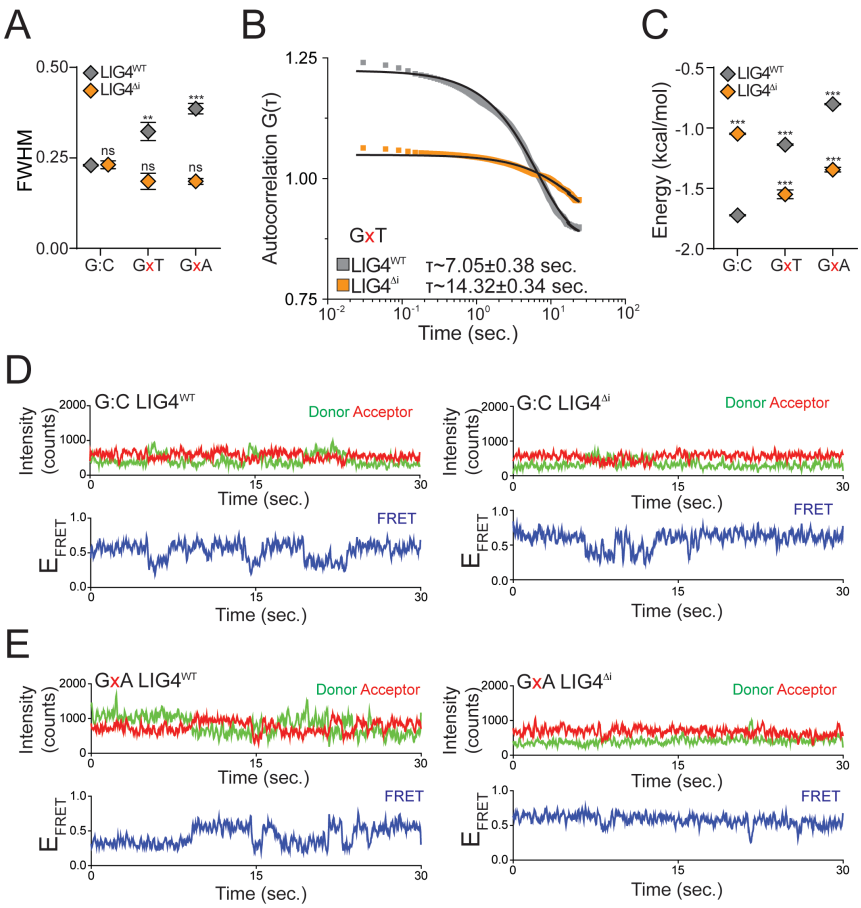

Figure S2. Effect of distorted ends on pairing dynamics of single molecule complexes with $LIG4^{WT}$ or $LIG4^{\Delta i}$, Related to Figure 2. **(A)** Full width at half maximum (FWHM) of peaks was calculated from E_{FRET} histograms for G:C, GxT, and GxA substrates. For **(A)** and **(C)**, error bars represent standard error of the mean for 3 experiments, and means were assessed for significance as in Figure 2B with confidence $p<0.01$ (**), $p<0.001$ (***) or not significantly different (ns). **(B)** Autocorrelation of individual FRET trajectories was used to calculate average transition times (τ) between FRET states of PECs formed on the GxT substrate with $LIG4^{WT}$ or $LIG4^{\Delta i}$. **(C)** Transition energy between FRET states calculated from autocorrelation. **(D)** Representative smFRET trajectory for $LIG4^{WT}$ and $LIG4^{\Delta i}$ PECs formed with G:C complementary ends **(E)** Representative smFRET trajectories of $LIG4^{WT}$ and $LIG4^{\Delta i}$ PECs formed with GxA ends.

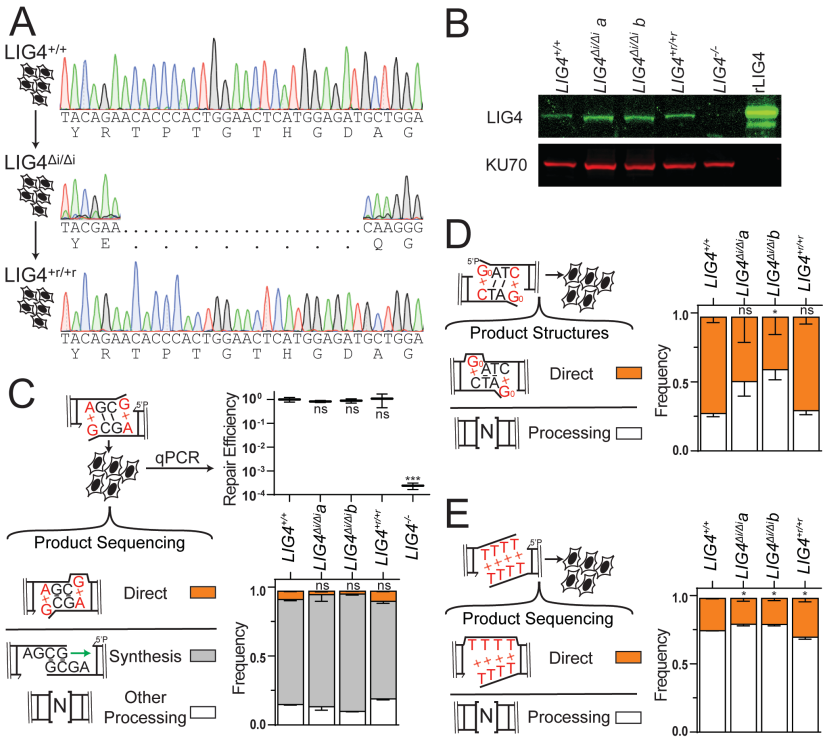

Figure S3. Effect of LIG4 insert1 on cellular joining of complex ends, Related to Figure 3. **(A)** Sequences of targeted region of genomic DNA harvested from *LIG4*$^{+/+}$, *LIG4*$^{\Delta i/\Delta i}$, and *LIG4*$^{+r/+r}$ cells. **(B)** Western blot was performed to validate similar LIG4 expression in the indicated cell lines **(C-E)** Substrates with varied end structures were introduced in the noted cell types. **(C)** G:A mispaired overhangs was electroporated into cells. Repair efficiency was quantified by qPCR and repair product structures were determined by sequencing. Product structures were classified as directly repaired (orange), gap fill-in synthesis (gray), or other processing (white) **(D)** A substrate with radiomimetic terminal 8-oxoguanine damage was electroporated into cells. Repair product structures were chacterized by diagnostic restriction digests and classified as either directly ligated (orange) or ligated after end processing (white) **(E)** A substrate with fully mispaired TTTT overhangs was electroporated into cells. Repair product structures were characterized by sequencing and classified as either directly repaired (orange) or processed (white). Error bars for **(C-E)** represent standard error of the mean for 3 experiments. Means were assessed for significance as in Figure 3B-E with confidence $p<0.05$ (*), $p<0.001$ (***) or not significantly different (ns).

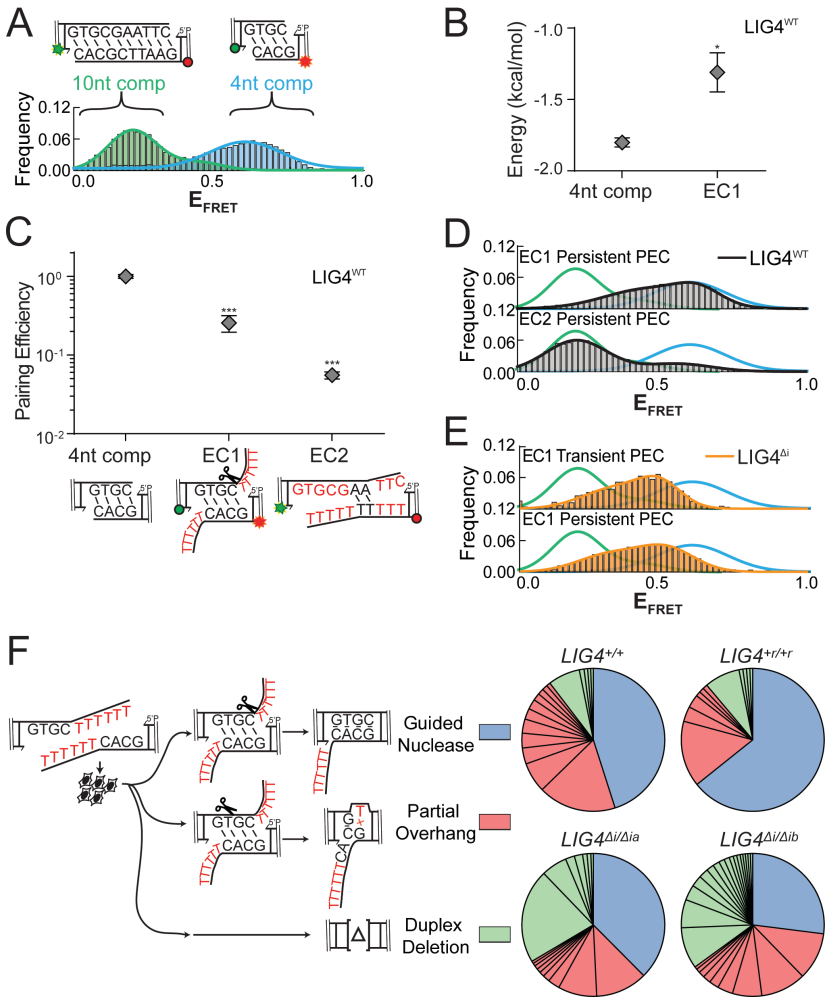

Figure S4. Effect of PEC flexibility on nucleolytic end processing, Related to Figure 4. **(A)** Histogram of E_{FRET} of PECs formed with FRET standards containing fully complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(B)** Transition energies calculated for 4nt complementary and EC1 substrates with $LIG4^{WT}$. Error bars represent standard error of the mean for 3 experiments. The mean energy for the EC1 substrate was assessed by t-test as significantly different from control (4nt complementary) with confidence $p<0.05$ (*). **(C)** Quantitation of pairing efficiency of a substrate where embedded complementarity was reduced and relocated (EC2), relative to 4 nt complementary overhangs. Error bars represent standard error of the mean for 3 experiments. Means were assessed by one-way ANOVA as significantly different from control (4nt complementary) with confidence $p<0.001$ (***). **(D)** Histograms of E_{FRET} of PECs formed on the EC1 (top) and EC2 (bottom) substrates (black), compared to FRET standards with complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(E)** Histograms of E_{FRET} of PECs formed with $LIG4^{\Delta i}$ on the EC1 substrate (orange), compared to FRET standards with complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(F)** The EC1 substrate was electroporated into cells. Repair product structures were determined by sequencing and classified as either deletion guided by embedded complementarity (blue), other deletions limited to single stranded overhang (red), or deletions that extended into double stranded flanking DNA (green).

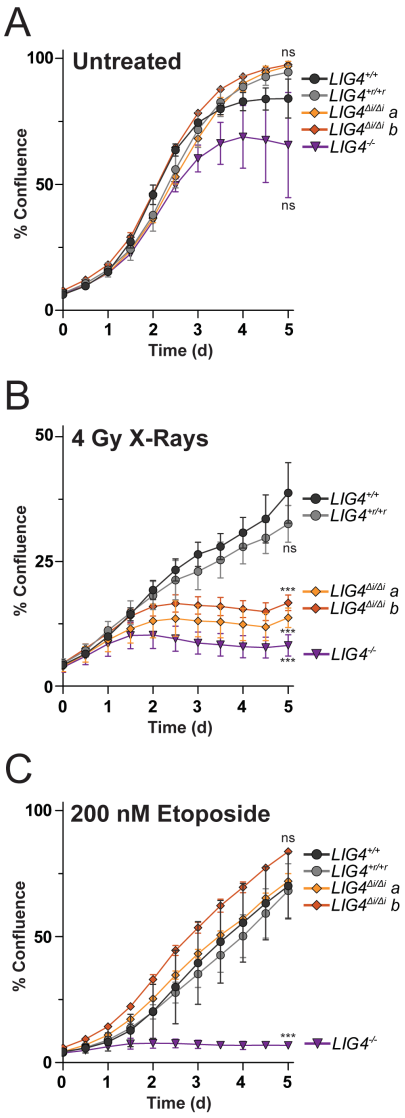

Figure S5. Effect of LIG4 insert1 on cell growth following treatment with ionizing radiation, Realted to Figure 5. **(A-C)** Cell growth was assessed by live cell imaging every 4 hours for 5 days after seeding cultures. Cells were either **(A)** left untreated, **(B)** irradiated with 4 Gy X-rays, or **(C)** treated with 200 nM etoposide. Mean growth after 5 days was assessed by one-way ANOVA as significantly different from control (*LIG4*$^{+/+}$ cells) with confidence $p<0.001$ (***) or not significantly different (ns). For all live cell imaging experiments, error bars represent the standard deviation of a triplicate.

Supplemental Table S1. Sequences of DNA Reagents, Related to Figure 1-4

Substrate Construction Oligonucleotides

Core	
CAAGTGGTCTCAGACTGGCTACCCTGCTTCTTTGAGCATTTCTGAAACTATCACTTGTGTTTAT TATTACACTGGCATTCATTCTCCAGAGAACATGTCTAGCCTATTCCCAGCTTTGCTTACGGAGT TACTCTGTATCTTTGCCTTGGAGAGTGCCAGAATCTGGTTTCAGAGTAAGATTTTATACATCAT TTTTAGACATAGAAGCCACAGACATAGACAACGGAAGAAAGAGACTTTGGATTCTACTTACG TTTGATTTCCCTGACGGAGACCTCGGC	
5’ G:C Left Cap Top Strand	GATCCTCACACCCATCTCA
5’ G:C Left Cap Bottom Strand	/5PHOS/AGTCTGAGATGGGTGTGAG
5’ G:C Right Cap Top Strand	GATCCTCGCTTAGCTGTATA
5’ G:C Right Cap Bottom Strand	/5PHOS/TGACTATACAGCTAAGCGAC
5’ GoxC Left Cap Top Strand	G_OATCCTCACACCCATCTCA
5’ GoxC Left Cap Bottom Strand	/5PHOS/AGTCTGAGATGGGTGTGAG
5’ GoxC Right Cap Top Strand	G_OATCCTCGCTTAGCTGTATA
5’ GoxC Right Cap Bottom Strand	/5PHOS/TGACTATACAGCTAAGCGAC
3’ G:C Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCGGCGCG
3’ G:C Left Cap Bottom Strand	CCGCACACCCATCTCA
3’ G:C Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCGGCGCG
3’ G:C Right Cap Bottom Strand	CCGCTTAGCTGTATA
3’ GxT Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCTGTGCG
3’ GxT Left Cap Bottom Strand	CAGCACACCCATCTCA
3’ GxT Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCTGTGCG
3’ GxT Right Cap Bottom Strand	CAGCTTAGCTGTATA
3’ GxA Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCAGAGCG
3’ GxA Left Cap Bottom Strand	CTGCACACCCATCTCA
3’ GxA Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCAGAGCG
3’ GxA Right Cap Bottom Strand	CTGCTTAGCTGTATA
3’ TTTT Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCTGTTTT
3’ TTTT Left Cap Bottom Strand	CAGCACACCCATCTCA
3’ TTTT Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCTGTTTT
3’ TTTT Right Cap Bottom Strand	CAGCTTAGCTGTATA
EC1 Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGTTTGTGCTTTTTT
EC1 Left Cap Bottom Strand	AAACACACCCATCTCA
EC1 Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCGTGCACTTTTTT
EC1 Right Cap Bottom Strand	ACGCTTAGCTGTATA
Nicked Substrate Top Strand Left Fragment	/5CY5/AGAAAACTGGCCCTTGCCATT
Nicked Substrate Top Strand Right Fragment	/5PHOS/CTCGGTGAGAGCATCGCTTA
Nicked Substrate Bottom Strand	TAAGCGATGCTCTCACCGAGAATGGCAAGGGCCAGT TTTCT
DSB End Substrate Top Strand	/5PHOS/TCACACACGCACGCATTTTT
DSB End Substrate Bottom Strand	/5CY5/TTTTTTGCGTGCGTGTGTGA
Complex Formation Substrate Top Strand	/5PHOS/CTCAGCTGGGAATTCCATATGAGTACTGCAG ATGCACTTGCTCGATAGATCTAACATGAG
Complex Formation Substrate Bottom Strand	/5CY5/GTAGGGCTCATGTTAGATCTATCGAGCAAGTG CATCTGCAGTACTCATATGGAATTCCCAGCTGAG
FRET Acceptor Top Strand	/5PHOS/CGTG/ICY5/AGAGGAGACAGAGTGCGGGCGA ACAACATAAATCGTACCCTCGTATGTATCGTATGGCT CATGCTTATCAGATGCT/3BIO/
FRET Acceptor Bottom Strand (G:C)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC

	TCTCACGCGCG
FRET Acceptor Bottom Strand (GxT)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC TCTCACGTGCG
FRET Acceptor Bottom Strand (GxA)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC TCTCACGAGCG
FRET Acceptor Bottom Strand (4 nt comp)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC TCTCACGGCAC
FRET Acceptor Bottom Strand (EC1)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC TCTCACGGCACTTTTTT
FRET Acceptor Bottom Strand (EC2)	AGCATCTGATAAGCATGAGCCATACGATACATACGA GGGTACGATTTATGTTGTTCGCCCGCACTCTGTCTCC TCTCACGGAATTCGCAC
FRET Donor Bottom Strand	/5PHOS/TCTG/ICY3/ATAAGCATGAGCCATACGATACA TACGAGGGTACGATTTATGTTGTTCGCCCGCACTCTG TCTCCTCTCACGTTTTCGTGAGAGGAGACAGAGTGC
FRET Donor Top Strand (G:C)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGACGCG
FRET Donor Top Strand (GxT)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGATGCG
FRET Donor Top Strand (GxA)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGAAGCG
FRET Donor Top Strand (4 nt comp)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGAGTGC
FRET Donor Top Strand (10 nt comp)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGAGTGCGAATTC
FRET Donor Top Strand (EC1)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGAGTGCTTTTTT
FRET Donor Top Strand (EC2)	GGGCGAACAACATAAATCGTACCCTCGTATGTATCG TATGGCTCATGCTTATCAGATTTTTTTTTT

PCR Primer Oligonucleotides

qPCR NHEJ Assay Forward	CTTACGTTTGATTTCCCTGACTATACAG
qPCR NHEJ Assay Reverse	GCAGGGTAGCCAGTCTGAGATG
Illumina Amplification Forward	NNNNNNCTTACGTTTGATTTCCCTGACTATACAG
Illumina Amplification Reverse	NNNNNNGCAGGGTAGCCAGTCTGAGATG
Illumina Adapter Top Strand	/5PHOS/GATCGGAAGAGCGGTTCAGCAGGAATGCCG AG
Illumina Adapter Bottom Strand	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Illumina Enrichment Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT
Illumina Enrichment Reverse	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCAT TCCTGCTGAACCGCTCTTCCGATCT
CRISPR Screening Forward	TGAGTTGGAATCGAATGCTG
CRISPR Screening Reverse	GAGGGGGCTTCTCTGCTACT
GoxC Amplification Forward	/5CY5/CCTTGGAGAGTGCCAGAATC
GoxC Amplification Reverse	CTGGAGAATGAATGCCAGTG

Oligonucleotides to generate sgRNAs (guide italicized and underlined)

Target *LIG4*^{WT} Top Strand	ccgg*GCATCTCCATGAGTTCCAGT*
Target *LIG4*^{WT} Bottom Strand	aaac*ACTGGAACTCATGGAGATGC*
Target *LIG4*$^{\Delta i}$ Top Strand	ccgg*ACTTTTAAACTACGAACAAG*
Target *LIG4*$^{\Delta i}$ Bottom Strand	aaac*CTTGTTCGTAGTTTAAAAGT*

Gene Targeting Regions

LIG4^{WT} and *LIG4*$^{+r/+r}$ sequence (guide italicized and underlined; BpmI site bolded)
AAACTTTTAAACTACAGAACA*CCCACTGGAACTCATGGAGATG**C***TGGAGACTTTGC
LIG4$^{\Delta i}$ sequence (guide italicized and underlined; BsmFI site bolded)
AA*ACTTTTAAACTACGAACAA----------G**GG***GAC TTTGC

SUPPLEMENTARY REFERENCES

Reid, D.A., Conlin, M.P., Yin, Y., Chang, H.H., Watanabe, G., Lieber, M.R., Ramsden, D.A., and Rothenberg, E. (2017). Bridging of double-stranded breaks by the nonhomologous end-joining ligation complex is modulated by DNA end chemistry. Nucleic acids research *45*, 1872-1878.

Rothenberg, E., Grimme, J.M., Spies, M., and Ha, T. (2008). Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. Proceedings of the National Academy of Sciences of the United States of America *105*, 20274-20279.