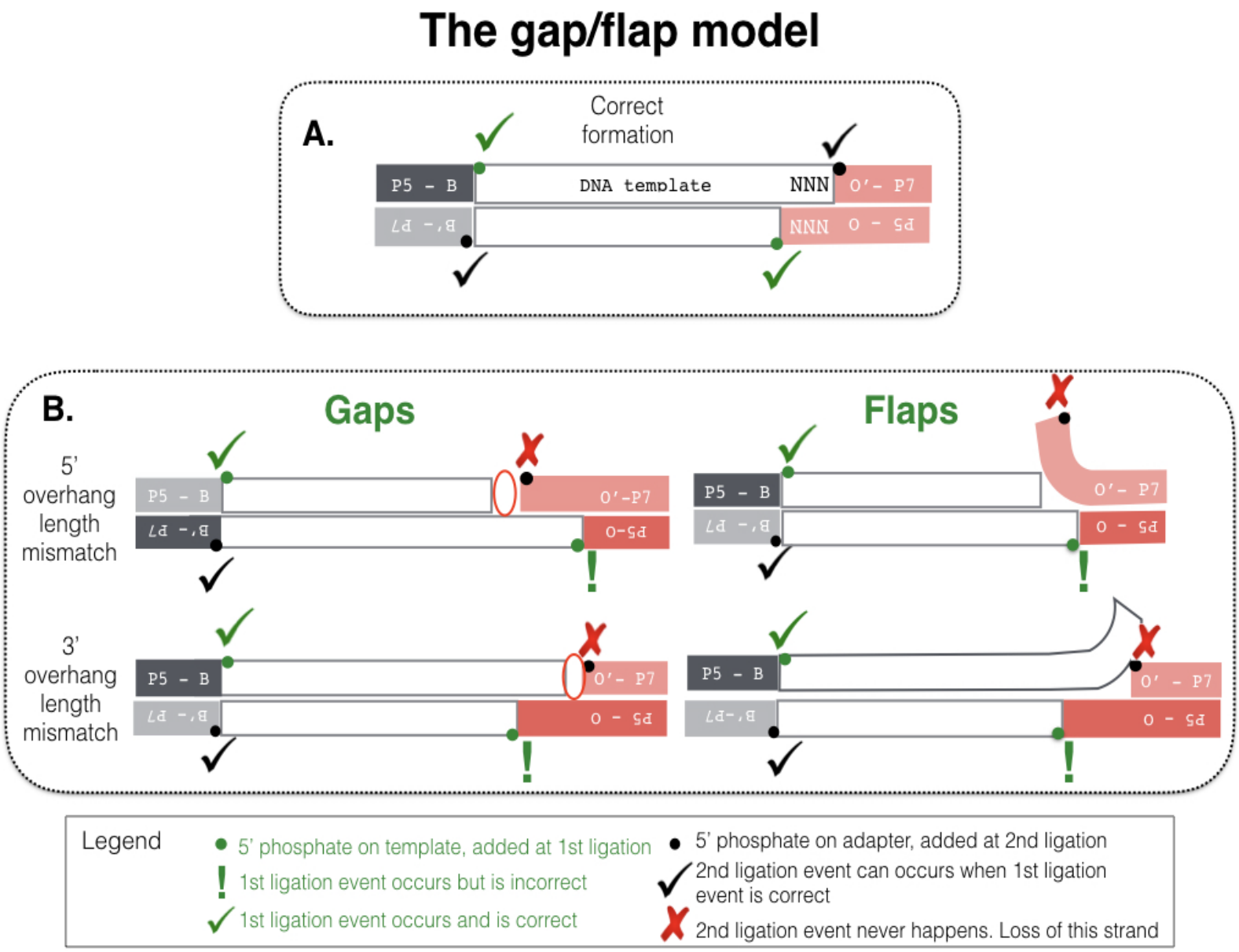
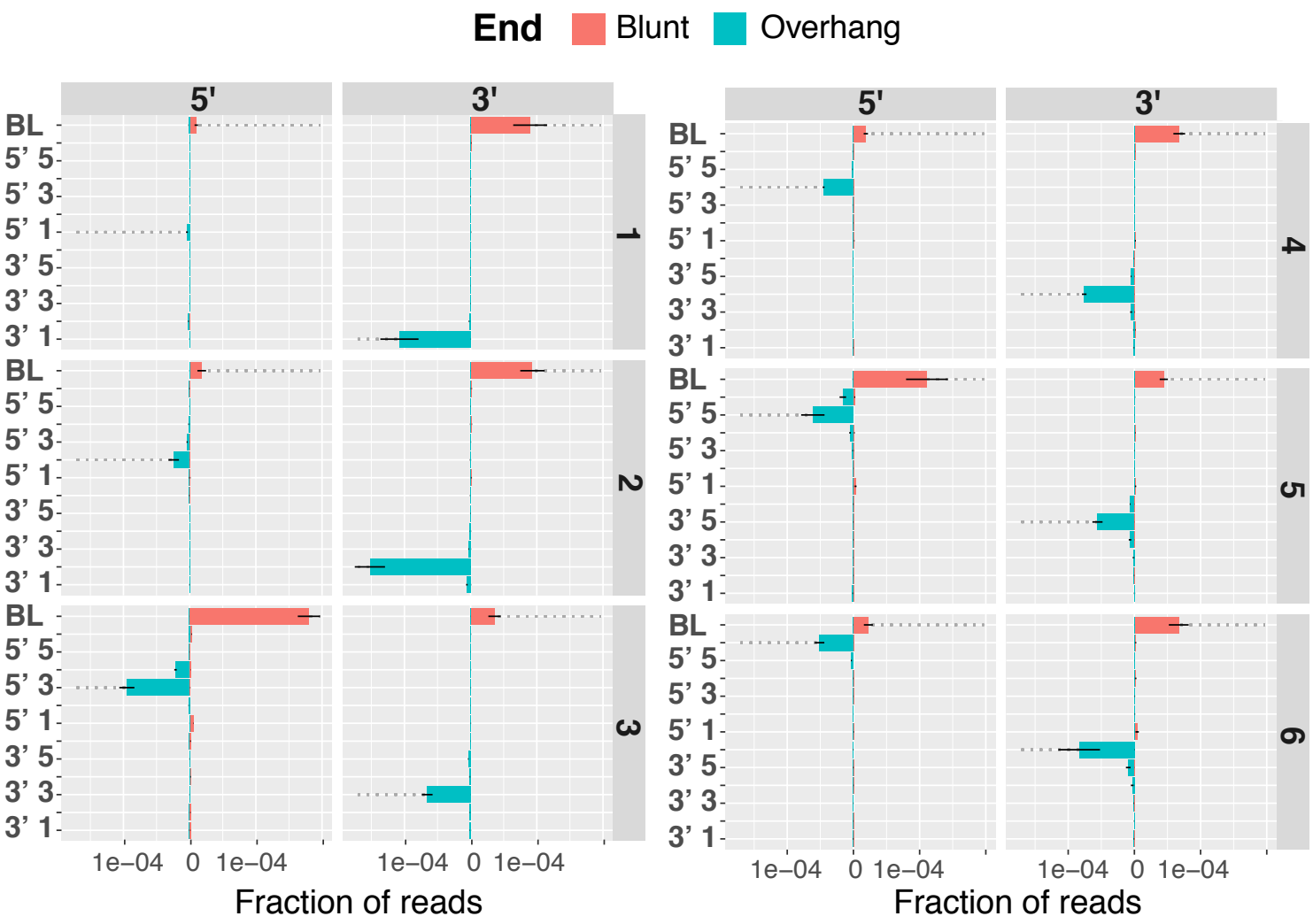


Supplementary Figure 1

Our control oligos with known overhangs are engineered with a blunt and an overhanging terminus, as depicted. In the sequencing data of known control oligos we observed, 1. an excess of only one of the two original strands, and 2. more accuracy for P7 than P5 UEIs. We propose these observations in the sequencing data are caused by errors during ligation resulting from gaps and flaps of mismatching overhang lengths.

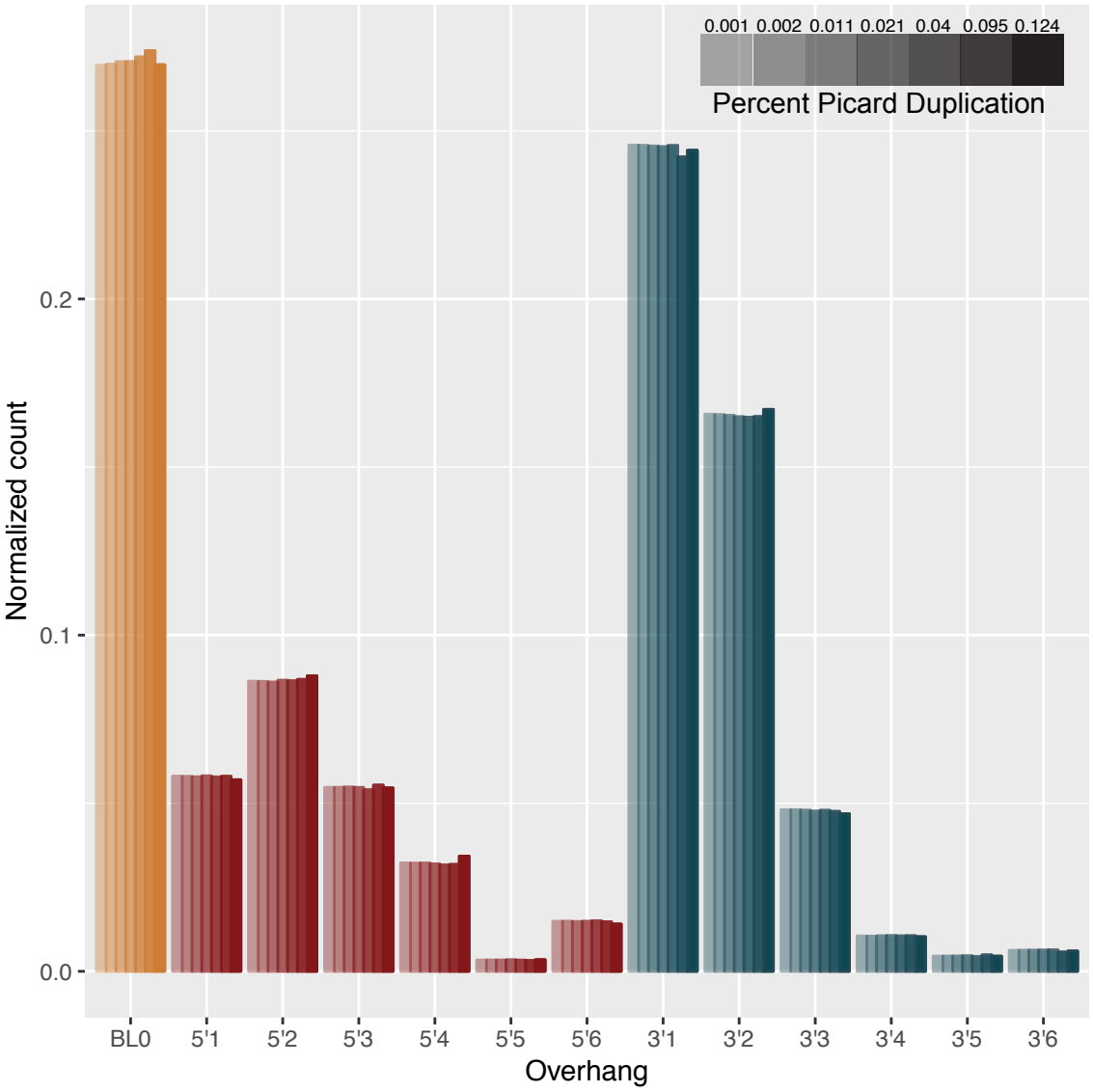
Supplementary Figure 1 depicts the proposed error modes that, if occurring during adapter ligation, may cause the loss of one of the two DNA strands, or misidentification of the true length. Green circles represent the location of the 1st ligation event where the 5' phosphate of the template and 3' hydroxyl of the P5 UEI ligate. Because the UEI adapters lack phosphates, a 2nd ligation event (black circles) is required to phosphorylate the 5' termini of the P7 UEI adapters, enabling ligation with the 3' hydroxyl of template DNA, thus forming a fully double-stranded library molecule.

A. The top panel shows an example of a correctly formed XACTLY molecule given a template with one blunt end and one overhanging terminus. B. Examples of Gaps and Flaps. On the left, a mismatch in the length of a 5' or 3' overhang causes a gap. On the right, a mismatch in the length of a 5' or 3' overhang causes a flap. When the lengths of the template/UEI overhangs are mismatched, an 'incorrect' covalent bond during the 1st ligation (green) can occur, inhibiting the 2nd ligation (black). This leads to conversion of only one strand and the loss of the other strand. Furthermore, in these cases the P5 UEI will report the wrong overhang length while the P7 UEI will be correct. We observe 1.7x increase in accuracy of the P7 UEI over using both UEIs, whether blunt or overhangs; for this reason we only use P7 UEIs during our analyses.



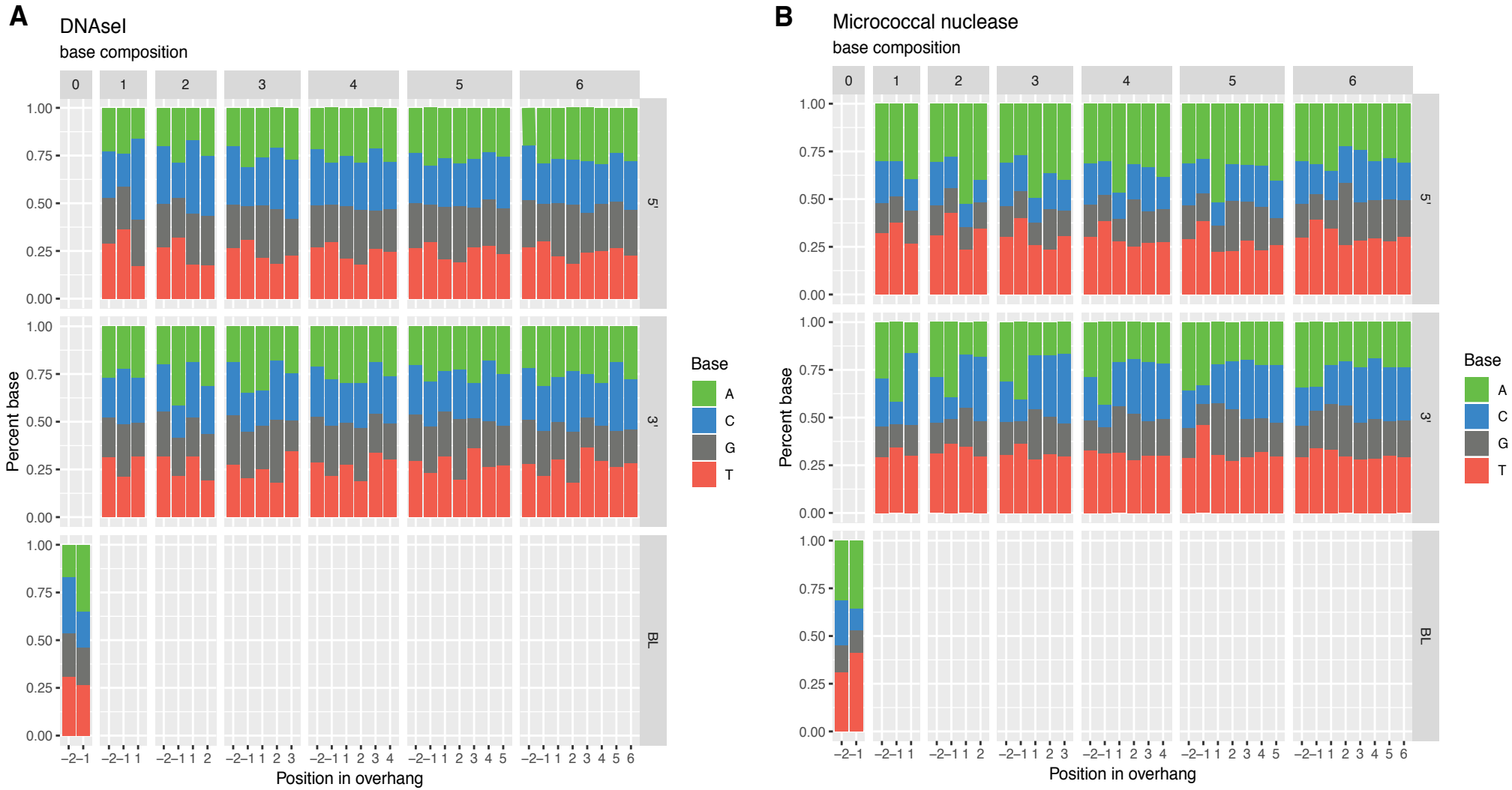
Supplementary Figure 2

Correctness of UEI/control oligo combinations by control oligo type (panels). Each panel represents a specific double-stranded control oligo. Each control oligo is engineered with one blunt end (pink) and at the other, a given length of single-stranded 5' or 3' overhang (blue), from 1 to 6nt in length. The Y-axis represents the UEI adapter observed; X-axis is the observed fraction of reads. Dotted lines show correct combinations; solid lines show 1 standard deviation across 17 replicate libraries.



Supplementary Figure 3

DNA termini profile of an XACTLY library generated from Bioruptor-sheared NA12878 gDNA sequenced PE 150 on a Illumina NextSeq (6.7 million read pairs) and MiSeq (182,299 read pairs). From the NextSeq data, 6,757,000 read pairs, which represents 12% duplication rate, we down-sampled the raw sequencing data using 'seqtk sample' to increasingly shallow sequencing depths: 5 million read pairs (9.5% duplication), 2 million read pairs (4% duplication) 1 million read pairs (2.1% duplication), 0.5 million read pairs (1.1% duplication) and 0.1 million read pairs (0.1% duplication). We also plotted the Illumina Miseq data (0.2% duplication rate). The darker the shade, the higher sequencing depth, reported as % duplication. The DNA termini profiles remain consistent across a range of sequencing depths and two Illumina platforms, suggesting that fewer than 1 million reads are required for fragment end profiling using the XACTLY data type.



Supplementary Figure 4

Position-specific base composition for overhangs produced via different shearing techniques. All values are means across two replicate libraries; Positions -1 and -2 are base composition of the human reference genome (hg19) at one and two bases before first aligned position of the template molecule. A: libraries digested using DNAseI. B: libraries digested using micrococcal nuclease.

Supplemental Table 1. Synthetic oligo design

Overhang Type	Sequence 1	Sequence 2
3' 1bp	CCATACTGTGGTCGTCACCTATTACCCCGCGTA AAGGTAGGCTATGTCATN$_1$	ATGACATAGCCTACCTTTACGCGGGGTAATAGGTGA CGACCACAGTATGG
3' 2bp	GTGAATTGTTGATGTCCTGGGTGCCTCGTCCC AAAAGCTGTCCTCACGACN$_2$	GTCGTGAGGACAGCTTTTGGGACGAGGCACCCAGG ACATCAACAATTCAC
3' 3bp	GCTTCTCGAACCCGCGATCCGGCCGATCCGGC ATAATGGGTTGATTTAGAN$_3$	TCTAAATCAACCCATTATGCCGGATCGGCCGGATCG CGGGTTCGAGAAGC
3' 4bp	CGACACGGATATTCCATCAAGAGACGGGCCTA TGGTCCCTGTGATGATGTN$_4$	ACATCATCACAGGGACCATAGGCCCGTCTCTTGATG GAATATCCGTGTCG
3' 5bp	ACCTTGTGTGTTGCTGAAGCAAAGCCGCGTGA CCGTTTTAACCAGCGAACN$_5$	GTTCGCTGGTTAAAACGGTCACGCGGCTTTGCTTCA GCAACACACAAGGT
3' 6bp	ATTTTACCACGAGTTCCTTACGACGGCTGTGAT GCCACGGTAGGCAGGTAN$_6$	TACCTGCCTACCGTGGCATCACAGCCGTCGTAAGGA ACTCGTGGTAAAAT
5' 1bp	N$_1$CGCTTTACGGGTCCTGGGCCGGGGTGCGAT ACCTTGCAGAAATCGAGGCC	GGCCTCGATTTCTGCAAGGTATCGCACCCCGGCCCA GGACCCGTAAAGCG
5' 2bp	N$_2$AGGACTCTGCCGTCGACGAGTTCGTTAATTC ACGGCATCACGTGCGTAGT	ACTACGCACGTGATGCCGTGAATTAACGAACTCGTC GACGGCAGAGTCCT
5' 3bp	N$_3$ACCTCCGTCGCGCTATGTTCTGTTGCATTCG ACCTTCTCCGTTCTGTGGG	CCCACAGAACGGAGAAGGTCGAATGCAACAGAACA TAGCGCGACGGAGGT
5' 4bp	N$_4$ACAAGAGGAGCATCCGTATTACCGCCTATA TCGCCTACGTTTAGAGCATT	AATGCTCTAAACGTAGGCGATATAGGCGGTAATAC GGATGCTCCTCTTGT
5' 5bp	N$_5$GTAAATCCCACACAGCTGTCGGCTTATATG GTCATTGGACGGCGTAATAG	CTATTACGCCGTCCAATGACCATATAAGCCGACAGC TGTGTGGGATTTAC
5' 6bp	N$_6$CCAGACAGCCATAGAGGTTACAAGCATAGC AATTTGCATCAGTTCGCAGA	TCTGCGAACTGATGCAAATTGCTATGCTTGTAACCTC TATGGCTGTCTGG

Supplemental Table 2. Blood collection tubes used in synthetic spike experiments.

Blood Collection Tube	Anti-coagulant	Nuclease inhibited
Red top tube	None	No - Additional nucleases released during clotting
Yellow top tube	Sodium Citrate	No - Citrate has no nuclease inhibition function
Purple top tube	Potassium EDTA	May be - EDTA can inhibit nuclease function
Streck DNA tube	Potassium EDTA	Yes - contains nuclease and cell lysis inhibitors

Supplemental Table 3. XACTLY Adapters and UEIs (lower case) associated with termini type

Adapter	Sequence 1	Sequence 2
Blunt	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tataccgc	gcggtatAGATCGGAAGAGCACACGTCTGAACTCCAGT CAC/3SPC3/
3' 1bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tgatatcg*N_1	cgatatcAGATCGGAAGAGCACACGTCTGAACTCCAGT CAC/3SPC3/
3' 2bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC TgtctgacN_1*N_1	gtcagacAGATCGGAAGAGCACACGTCTGAACTCCAG TCAC/3SPC3/
3' 3bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC TgagccaaN_2*N_1	ttggctcAGATCGGAAGAGCACACGTCTGAACTCCAGT CAC/3SPC3/
3' 4bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC TcgccataN_3*N_1	tatggcgAGATCGGAAGAGCACACGTCTGAACTCCAGT CAC/3SPC3/
3' 5bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC TcgtatatN_4*N_1	atatacgAGATCGGAAGAGCACACGTCTGAACTCCAG TCAC/3SPC3/
3' 6bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC TgactaagN_5*N_1	cttagtcAGATCGGAAGAGCACACGTCTGAACTCCAGT CAC/3SPC3/
5' 1bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tagtacgg	N_1*ccgtactAGATCGGAAGAGCACACGTCTGAACTCC AGTCAC/3SPC3/
5' 2bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tagcagcg	N_1*N_1cgctgctAGATCGGAAGAGCACACGTCTGAACT CCAGTCAC/3SPC3/
5' 3bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tccatatg	N_1*N_2catatggAGATCGGAAGAGCACACGTCTGAACT CCAGTCAC/3SPC3/
5' 4bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tagcctgg	N_1*N_3ccaggctAGATCGGAAGAGCACACGTCTGAACT CCAGTCAC/3SPC3/
5' 5bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tatacgcg	N_1*N_4cgcgtatAGATCGGAAGAGCACACGTCTGAACT CCAGTCAC/3SPC3/
5' 6bp	/5SpC3/ACACTCTTTCCCTACACGACGCTCTTCCGATC Tgctaggc	N_1*N_5gcctagcAGATCGGAAGAGCACACGTCTGAACT CCAGTCAC/3SPC3/

Supplementary Table 4. Sequencing metrics for a randomly downsampled XACTLY library - Library ID APN702 - generated from Bioruptor-sheared gDNA (relates main figure 2)

Library	Number of reads	% Reads Mapped to hg19	Picard percent duplicates
Hiseq (original)	6756906	82.7	0.124
Downsampling-1	5000000	82.7	0.095
Downsampling-2	2000000	82.7	0.040
Downsampling-3	1000000	82.8	0.021
Downsampling-4	500000	82.8	0.011
Downsampling-5	100000	82.7	0.002
Miseq (original)	182299	89.3	0.018

Supplementary Table 5. Sequencing data for sensitivity experiments (relates to main figure 3)

Sample Type	Library ID	Number of Read Pairs Sequenced	% Reads Mapped to hg19	Number of Overhangs on R2 Observed	Number of 5' AATT Observed in R2 Overhangs	% of 5' AATT in R2 Overhangs
100% Mechanical Shear	APN674	138921	91	91011	41	0.05
1% RE Digest	APN675	132516	90	86052	647	0.75
5% RE Digest	APN676	134791	90	88194	2977	3.38
10% RE Digest	APN677	129736	90	85432	4268	5.00
25% RE Digest	APN678	106698	88	68811	6499	9.44
50% RE Digest	APN679	125340	86	75954	12211	16.08
100% RE Digest	APN680	97724	76	44139	17523	39.70

Supplementary Table 6. Sequencing data for common shearing mechanisms and nucleases (relates to main figure 4)

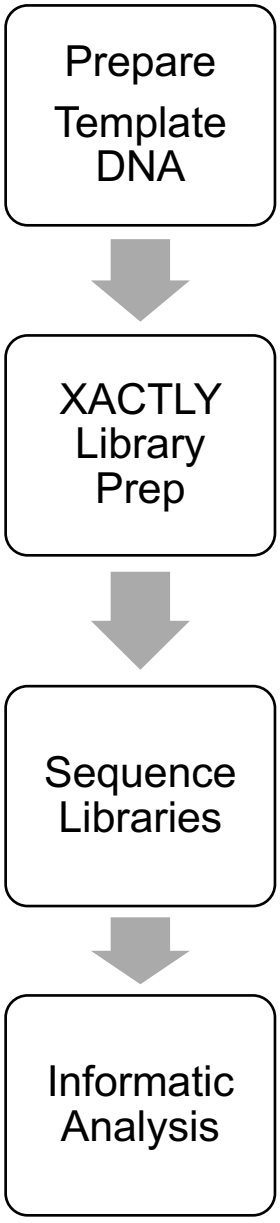
Sample Type	Library IDs	Number of Read Pairs Sequenced	% Molecules that Contain UEIs	% UEI Molecules Mapped to hg19
DNaseI	APN694, APN698	184, 938 ± 27,744	92	66
Micrococcal Nuclease	APN697, APN701	178,653 ± 30,933	95	82

Protocol: XACTLY Library Preparation for Fragmented DNA

Materials

- 0.2 mL PCR tubes
- 0.2 mL Magnet Rack (Diagenode DiaMag02)
- 80% Ethanol
- ddH2O (DNA-free, DNAse/RNase free)
- 10x CutSmart Buffer (New England Biolabs)
- Shrimp Alkaline Phosphatase (rSAP) (1000 units/mL)
- 10x T4 DNA Ligase Buffer
- T4 Polynucleotide Kinase (PNK) (10,000 units/mL)
- T4 DNA Ligase (400,000 units/mL)
- 50% PEG8000
- XACTLY UEI Adapters: Pool of blunt, and with 5' & 3' overhangs 1-6 nucleotide
- 2x PCR Mastermix (e.g. Kapa Biosystem KK2601)
- SPRI purification beads
- EB Buffer (10mM Tris-HCL, pH 8.5)
- TE Buffer (10mM Tris-HCL, 1mM EDTA, pH 8)
- Primers (as published in Meyer et al. 2010):
    - IS4
    - P7 indexing primer
    - IS5
    - IS6
- Thermal cycler

Overview of Workflow



- rSAP template DNA to remove errant phosphates (45min)
- Phosphorylate 5' termini (1hr)
- Optional stopping point (no cleanup)

- Ligate XACTLY Adapters then purify DNA (90min)
- Perform second ligation/nick repair then purify (75min)
- Index PCR and purify (1hr)
- Optional stopping points after any purification

- Sequence libraries on MiSeq or other Illumina platform
- >100,000 read-pairs/library (2x150bp)
- Titrate PhiX as necessary to increase sequencing run complexity

- Download sequencing run data to the server
- Run informatics pipelines

Part 1. Prepare Template DNA

The amount of rSAP used to dephosphorylate the template DNA depends on moles of ends present in the template. Use NEBioCalculator to estimate moles of ends (http://nebiocalculator.neb.com/#!/dsdnaends)

Step 1. rSAP Template DNA

Reagent	Volume
Template DNA	Up to 1pmol of ends
10x Cutsmart buffer	2µl
rSAP	1µl
H2O	To 20 µl
Total Vol	**20 µl***

1. Perform reaction in a 0.2ml PCR tube
2. Incubate sample at 37°C for 30 minutes then 65°C for 10 minutes.
3. Proceed reaction to PNK

*Scale reaction volume appropriately for more pmoles of ends.

Step 2. PNK Template DNA

Reagent	Volume
Prepared Template DNA	20 µl
10x T4 DNA ligase buffer	4 µl
50% PEG 8000	4 µl
H2O	10 µl
T4 PNK	2 µl
Total Vol	**40 µl**

1. Perform reaction in the same tube as the rSAP reaction if in 20 µl.
2. Incubate samples at 37°C for 30 minutes then 65°C for 30 minutes
3. Template DNA is now ready for XACTLY Ligation, no DNA cleanup necessary
4. Calculate the pmoles/µl (uM) of your sample by taking the original pmoles placed into the rSAP reaction/40µl. Currently this is calculated not by pmols of the ends but by total pmols in the sample, get this by dividing the pmols of ends by 2.
   Ex: 1pmol of ends = 500fmols total. 500fmols/40µl = 12.5fmol/µlof template after T4 PNK treatment.

Note: Safe stopping point. Samples can be stored for several weeks at 4°C or -20°C

Part 2. XACTLY Library Preparation

Step 1. Ligate the XACTLY Adapters

Reagent	Volume
Prepared Template DNA	0.05pmol
XACTLY Adapter Pool	1 pmol
10x T4 DNA Ligase Buffer	6 µl
T4 DNA Ligase	2 µl
H2O	To 60 µl
Total Vol	**60 µl**

1. Incubate prepared template DNA in the above reaction at 20°C for 1hr in a 0.2ml PCR tube
2. *Purify*. Perform a 2x SPRI bead cleanup for any substrate DNA under 100bp. Perform a 1.2x SPRI clean for any substrate DNA with an average size range greater than 100bp
    a. For a 2x SPRI bead clean:
        i. Add 120 µl of SPRI beads to the sample tube
        ii. Fully resuspend the beads and incubate for 5-10 minutes at room temperature off magnet.
        iii. Quick spin the tube and place on the magnet for 2-5 minutes. Pipet off and discard supernatant once it is clear.
        iv. Leave tube on the magnet, and wash beads twice with 200 µl 80% EtOH by incubating the EtOH with the beads for 30 seconds with each wash, then aspirate EtOH. There is no need to resuspend the beads for these washes.
        v. Quick spin the tube and place on the magnet for 1 minute. Use a 20 µl pipet tip to remove traces of EtOH.
        vi. Air dry beads for 10 minutes on the magnet.
        vii. Add 20 µl EBT to sample tube.
        viii. Fully resuspend the beads and incubate for 2 min at room temperature off magnet.
        ix. Quick spin the tube and place on the magnet until supernatant is clear. Transfer all the supernatant to a new 0.2ml PCR tube. The tube containing the supernatant is now the sample. Discard beads.
3. Proceed to Nick Repair

Note: Safe stopping point. Samples can be stored overnight at 4°C.

Step 2. Nick Repair

Reagent	Volume
Eluted DNA from Ligation	20 µl
H2O	21.8 µl
10x T4 DNA Ligase Buffer	5 µl
T4 PNK	2 µl
Total Vol	**48.8 µl**

1. Incubate samples at 37°C for 30 minutes
2. Add 1.2 µl of T4 DNA ligase to the nick repair reaction
    a. Incubate samples at 20°C for 15 minutes
3. Purify. Perform a 2x SPRI clean (100 µl SPRI beads) as outlined in Part 2, Step 1.2.
    a. Elute samples in 20 µl EBT

Note: Safe stopping point. Samples can be stored overnight at 4°C.

Part 3. Library Amplification

Step 1: Index PCR

Reagent	Volume
Eluted DNA from Nick Repair	10 µl
H2O	11 µl
2x HiFi HS Readymix (Kapa)*	25 µl
IS4 indexing primer (10uM)	2 µl
Index Primer (10uM)	2 µl
Total Vol	**50 µl**

1. Amplify reaction in a thermal cycler as follows:
    a. 3 min 98°C
    b. 20s at 98°C / 30s at 65°C / 30s at 72°C, repeat for desired cycle #**
    c. 1 min at 72°C
    d. Hold at 12°C

*if other polymerase is desired, please use manufacturer's concentration and cycling recommendations
** cycle number recommendation

Input DNA amount	Initial Index PCRc (Best Guess)
0.05pmol (<100 bp)	15c
0.05pmol (cfDNA)	18c
0.025pmol (cfDNA)	22c

Note: The number of PCR cycles is dependent on the amount, type, and quality of input substrate DNA. The goal is to reach post-index yields of > 4nM. If your sample fails to reach 4nM after your initial best guess for number of PCR cycles you will have to re-amplify your DNA in PCR reaction using IS5 and IS6 primers. See the below tables for details.

Reamp PCR Reagent	Reamp PCR Vol
Eluted DNA from Initial Index PCR	18 µl
H2O	2 µl
2x HiFi HS Readymix (Kapa)	25 µl
IS5 (10uM)	2
IS6 (10uM)	2
Total Vol	50 µl

2. *Purify*. For DNA template < 100 bp (prior to ligation), perform a 1.5x SPRI clean (75 µl SPRI beads). For DNA template > 100 bp (prior to ligation), perform a 1.2x SPRI clean. Follow the protocol in Part 2, Step 1.2.
    a. Elute samples in 20 µl EBT
3. This is the final sequencing library. Save the remaining 10 µl not used for index PCR as a backup stock of the library

Note: For added complexity or added [DNA] perform the above index PCR reaction in duplicate using the additional 10 µl of the sample and the same index primer. Pool the sample into one tube prior to the 1.5x SPRI clean.

Step 2a. QC Indexed XACTLY Library

1. Quantify the sequencing library using the HS dsDNA Qubit
2. Quantify the sequencing library using the D1000 tapestation tape.
3. If library has the expected size distribution (lack of heteroduplexes or high % dimers) use the following equation to quantify the final nM of the sequencing library:
   $(Qubit_{ng/\mu l} \times TapeSation_{nM}) / TapeStation_{ng/\mu l} = Average\ Final_{nM}$ Library

Step 2b. Index PCR (Troubleshooting)

1. **If the sequencing library contains heteroduplexes** either 1 µl can be reconditioned in the PCR machine for 2 additional PCR cycles, quantified using the D1000 tapestation, and then back calculated to determine the Average $Final_{nM}$ Library or the entire library can be reconditioned in the PCR machine for 2 cycles using the same protocol listed above and all 18 µl of the remaining product with only 3 µl of H20 to make up the void volume. If the entire final library is reconditioned it must be 1.5x SPRI cleaned again before quantification in sub-steps 4-6 of step 3a.
2. **If the sequencing library contains a significant fraction of adapter dimers** samples can be size selected using the Pippen prep. Use a 3% gel and manufacturer's instructions to size select a library in the 175-250 bp range. Use a 2% gel and manufacturer's instructions to size select a cfDNA library in the 175-600bp range. 1.5% gel size selection is an option if molecules up to 800bp wish to be retained.

NEXT: Sequencing. Follow manufacturer instructions for pooling and paired-end sequencing on Illumina platforms