

**DSBCapture: *in situ* capture and direct sequencing of dsDNA breaks**

Stefanie V. Lensing<sup>1</sup>, Giovanni Marsico<sup>1</sup>, Robert Hänsel-Hertsch<sup>1</sup>, Enid Y. Lam<sup>1</sup>, David Tannahill<sup>1</sup> & Shankar Balasubramanian<sup>1,2,3</sup>

1. Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK.
  2. Department of Chemistry, University of Cambridge, Cambridge, UK.
  3. School of Clinical Medicine, University of Cambridge, Cambridge, UK.
  4. Present address: Cancer Research Division, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.
- Correspondence should be addressed to S.B.: sb10031@cam.ac.uk

**Supplementary Table 1.** EcoRV (HeLa) DSBCapture sequencing data. Alignment statistics for a paired-end MiSeq sequencing run of the EcoRV DSBCapture library.

EcoRV (HeLa) DSBCapture experiment	Number
Reads sequenced	43,794,614
Clean reads aligned	31,267,705
Clean reads from read 1	16,366,684
Unduplicated clean reads from read 1	13,449,911
Unduplicated clean reads from read 1 (forward strand)	6,718,566
Unduplicated clean reads from read 1 (reverse strand)	6,731,345

**Supplementary Table 2.** AID-DIVa U2OS DSBCapture sequencing data. Alignment statistics for a paired-end NextSeq sequencing run of the AID-DIVa U2OS DSBCapture library. The % of proximal reads used was calculated as the number of unduplicated proximal reads divided by the total number of proximal reads sequenced (total number of proximal reads = # Reads/2).

Library	# Reads	# Clean Aligned Reads	# Unduplicated Reads	% Duplication	# Unduplicated Proximal Reads	% Proximal Reads Used
U2OS DSB-Capture	267,536,292	230,074,135	83,734,433	63.6	37,347,231	27.9

**Supplementary Table 3.** NHEK DSBCapture and BLESS sequencing data. Combined data from multiple paired-end NextSeq sequencing runs of the NHEK BLESS and DSBCapture libraries, where two independent biological replicates were performed (N1 and N2). The % of proximal reads used was calculated as the number of unduplicated proximal reads divided by the total number of proximal reads sequenced (total number of proximal reads = # Reads/2).

Library	# Reads	# Clean Aligned Reads	% Clean Reads	# Unduplicated Reads	% Duplication	# Unduplicated Proximal Reads	% Proximal Reads Used
DSB-Capture N1	353,056,932	314,223,106	89.0	256,152,484	18.5	124,192,457	70.4
DSB-Capture N2	229,658,364	204,047,616	88.8	160,803,701	21.2	77,523,430	67.5
BLESS N1	324,872,812	228,953,681	70.5	147,477,098	35.6	71,305,629	43.9
BLESS N2	585,674,924	317,449,269	54.2	51,725,762	83.7	24,766,932	8.5

**Supplementary Table 4.** Overlap of high confidence DSBCapture peaks in NHEKs with various marks.

Mark	# DSBs overlapping mark	Fold enrichment of DSB overlapping mark/random	% DSBs overlapping mark
DNase	64,763	33.3	76.2
H3K4me1	58,844	10.7	69.3
H3K4me2	58,842	14.5	69.3
H2AZ	48,959	11.4	57.6
H3K27ac	45,991	15.2	54.1
POL2B	32,172	12.2	37.9
CTCF	25,347	25.0	29.8
H3K4me3	20,171	17.4	23.7
H3K79me2	18,242	3.3	21.5
P63	6,131	39.1	7.2
EZH2	4,884	3.2	5.7
H3K27me3	3,766	1.2	4.4
H3K36me3	3,044	0.8	3.6
H3K9me3	1,908	0.7	2.2