

Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs

Ashley D Sanders¹, Ester Falconer¹, Mark Hills¹, Diana C J Spierings² & Peter M Lansdorp^{1–3}

¹Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada. ²European Research Institute for the Biology of Ageing, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. ³Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. Correspondence should be addressed to P.M.L. (plansdor@bccrc.ca).

Published online 11 May 2017; doi:10.1038/nprot.2017.029

The ability to distinguish between genome sequences of homologous chromosomes in single cells is important for studies of copy-neutral genomic rearrangements (such as inversions and translocations), building chromosome-length haplotypes, refining genome assemblies, mapping sister chromatid exchange events and exploring cellular heterogeneity. Strand-seq is a single-cell sequencing technology that resolves the individual homologs within a cell by restricting sequence analysis to the DNA template strands used during DNA replication. This protocol, which takes up to 4 d to complete, relies on the directionality of DNA, in which each single strand of a DNA molecule is distinguished based on its 5′–3′ orientation. Culturing cells in a thymidine analog for one round of cell division labels nascent DNA strands, allowing for their selective removal during genomic library construction. To preserve directionality of template strands, genomic preamplification is bypassed and labeled nascent strands are nicked and not amplified during library preparation. Each single-cell library is multiplexed for pooling and sequencing, and the resulting sequence data are aligned, mapping to either the minus or plus strand of the reference genome, to assign template strand states for each chromosome in the cell. The major adaptations to conventional single-cell sequencing protocols include harvesting of daughter cells after a single round of BrdU incorporation, bypassing of whole-genome amplification, and removal of the BrdU⁺ strand during Strand-seq library preparation. By sequencing just template strands, the structure and identity of each homolog are preserved.

INTRODUCTION

Advances in high-throughput sequencing (HTS) technologies have allowed many studies on genomic heterogeneity and plasticity in relation to phenotypic variation and diseases. These studies have been fundamental to uncovering complex genetic variation in human populations^{1,2} and understanding how aberrations in genomes drive pathological outcomes^{3,4}. To accomplish this, sequence reads from genome and exome libraries are mapped to reference assemblies, and bioinformatics approaches are used to search for discordant read alignments that predict the location of structural variants in the sequence data⁵. However, even the most sophisticated HTS pipelines face major challenges in detecting copy-neutral variants such as inversions and translocations, which reorganize the homolog structure without altering DNA content^{6,7}. Moreover, detection algorithms typically locate variants only when they are supported by multiple discordant reads^{8–11}. Consequently, most HTS methods require deep (>30×) genomic coverage, which is obtained from micrograms of starting material. Inevitably, such data represent an average derived from millions of input cells.

More recently, single-cell sequencing tools have been developed to better resolve population substructures^{12–17}. These methods have been used to explore cellular heterogeneity in the context of cancer and somatic variation. For instance, clonal expansions of individual tumor cells have been tracked based on variant signatures that were inherited from a founder cell that suffered genome instability^{15,18}, and mosaic aneuploidy in the brain was found to be less prevalent than predicted based on earlier studies using fluorescence *in situ* hybridization techniques¹⁶. Most single-cell sequencing methods begin with a whole-genome amplification step to increase input material, and this can introduce PCR artifacts, amplification biases and allelic dropouts that confound variant calling¹⁹. Furthermore, although conventional single-cell sequencing allows resolution of subpopulations, results still

represent an average of each cell's DNA content, and resolving the individual homologs is near impossible using these types of data. Consequently, the relationship of the variants (both structural and single-nucleotide polymorphisms) to a chromosomal homolog (i.e., the phase) is masked using these approaches.

Strand-seq is a single-cell template strand sequencing technology that generates directional genomic libraries that, when aligned to the reference genome, permits a clear distinction between the individual homologs of a chromosome. Homolog resolution allows diverse types of variants to be located in a single cell that would otherwise be very challenging to detect using conventional approaches. This includes inversions, translocations, copy-number changes, aneuploidy events and complex structural variants, with haplotype awareness. By pooling data from several dozen Strand-seq cells, each homologous chromosome can be uniquely characterized and analyzed. Since our original publication²⁰, the protocol has been scaled and automated for library preparation on a liquid-handling platform, allowing 96 single-cell libraries to be constructed simultaneously. This advance has greatly improved the throughput and efficiency of the Strand-seq method, making it possible to perform more large-scale studies with a fraction of the cost and time. Here, we provide a detailed protocol for the scaled Strand-seq procedure, describe basic data analysis pipelines and illustrate interpretation of results.

Overview of the procedure

In Strand-seq (Fig. 1), parental DNA template strands inherited by daughter cells are sequenced following construction of a modified Illumina genomic library²⁰. The method takes advantage of the directionality of single-stranded DNA molecules, which are distinguished as either Crick (C; forward, or plus, strand of the reference assembly) or Watson (W; reverse, or minus, strand) based

PROTOCOL

on their 5′–3′ orientation (Fig. 1a). Daughter cells, generated following one mitotic division of a parental cell in the presence of BrdU, are isolated and deposited as single cells (or nuclei) into

single wells of a 96-well plate (Fig. 1b). Forgoing any whole-genome preamplification step, the genomic DNA is fragmented using micrococcal nuclease (MNase) to generate mononucleosomal

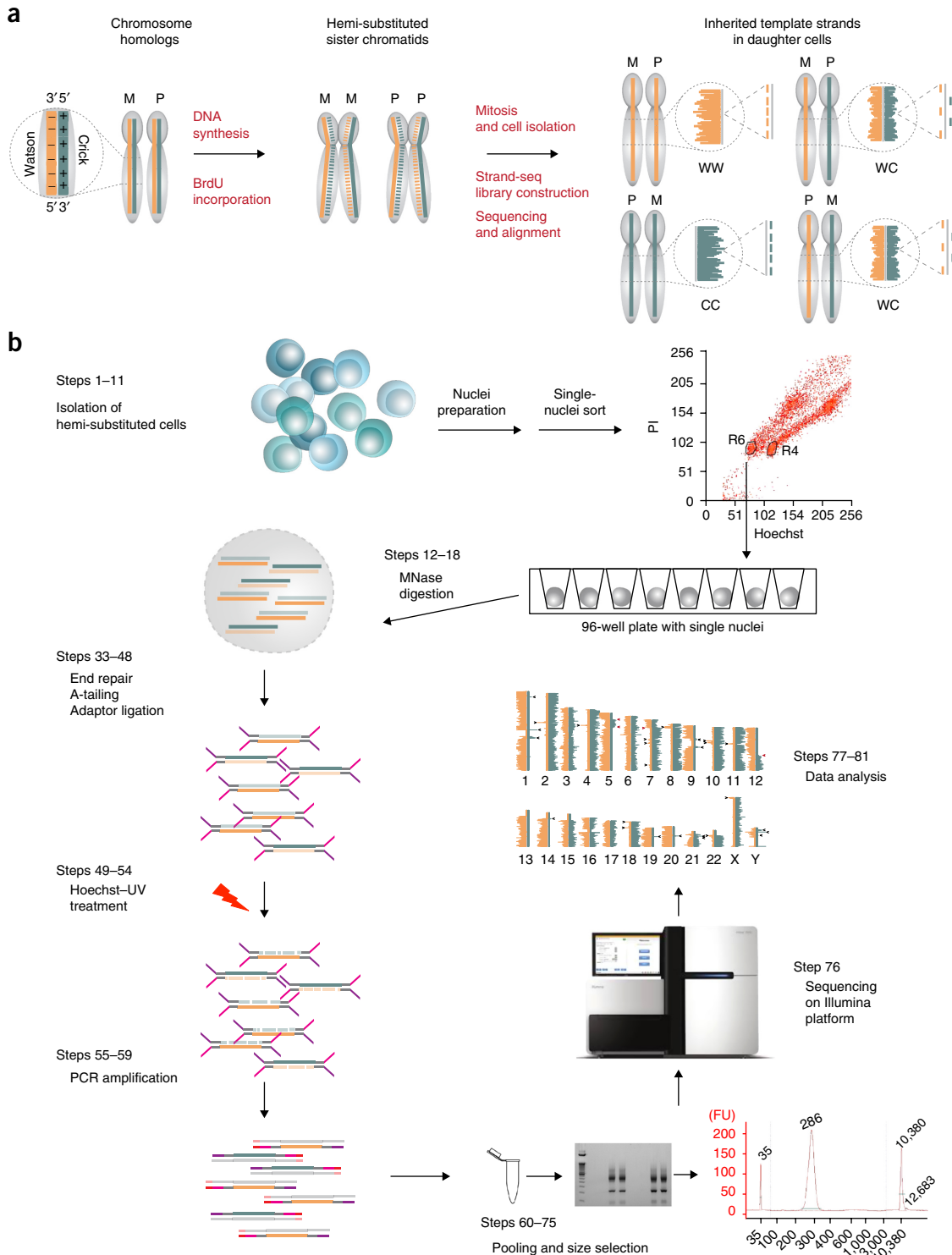


Figure 1 | Overview of Strand-seq protocol. **(a)** Principle of the technology and inheritance of template strands in daughter cells. Maternal (M) and paternal (P) chromosome homologs are composed of a positive (Crick, ‘C’, blue) and a negative (Watson, ‘W’, orange) template strand. During DNA replication, BrdU is incorporated exclusively into nascent DNA strands (dashed lines). Upon cell division, each daughter cell inherits CC, WW or WC template strands of each parental chromosome. Image adapted with permission from Sanders *et al.*²¹, Cold Spring Harbor Laboratory Press. **(b)** During library preparation, the BrdU-substituted strand will be selectively removed by Hoechst-UV treatment after adaptor ligation and subsequently only template strands (orange or blue) are amplified and sequenced. Reads from a Strand-seq library are aligned to the reference genome and binned, and read counts are displayed on ideograms for each chromosome by BAIT. Procedure step numbers are shown on this schematic overview of the experimental workflow for Strand-seq. FU, fluorescence units.

fragments of ~150 bp in length. The MNase-digested chromatin is then blunt-ended and 5'-phosphorylated by an end repair reaction, and a single adenine is attached to the 3' end of the fragment by an A-tailing reaction. This adenine serves as a substrate for standard Illumina forked adaptors, which have a single 3' thymidine overhang (Supplementary Fig. 1a). After adaptor ligation, the BrdU-substituted DNA strands are nicked by photolytic cleavage. This is accomplished by treating samples with Hoechst 33258, followed by UV radiation, which induces a single-stranded nick at sites of BrdU incorporation (Fig. 2). Nicking the BrdU⁺ strand interferes with subsequent PCR amplification, and as a result, only the BrdU⁻ (i.e., the original DNA template) strand is amplified to produce a directional single-cell library (Supplementary Fig. 2). During PCR amplification, a custom multiplexing PCR primer is added to the reaction (Supplementary Fig. 1b), such that each cell receives a unique hexamer barcode. This barcode allows multiple cells to be pooled for gel size selection and sequencing. Following Illumina HTS, the template strands in each library are distinguished by the first sequencing read generated from the A2 Illumina adaptor, and the single-cell barcode is read using a custom sequencing primer generated by the second sequencing read. By demultiplexing and aligning sequencing data to a reference assembly, template strand identity is determined for each chromosome in the cell (Fig. 1b).

The key difference between Strand-seq and other single-cell sequencing protocols is that it aims to sequence only one strand of DNA in each cell, by isolating daughter cells after one cell division. While many steps mirror other single-cell sequencing methodologies, the ability to produce directional single-strand libraries is dependent on nicking the BrdU-incorporated strands through incubation with Hoechst and subsequent UV irradiation. In addition to introducing a photolytic cleavage step, the Strand-seq protocol also bypasses any whole-genome amplification in order to preserve the specific labeling of the DNA strands. This means that genomic libraries are generated from only 6–10 pg of input DNA. Because of this, and the multiple enzymatic cleanup steps that are required during library preparation, we typically obtain genomic coverage in the range of a few percentage points (1–5%)—depending on the number of libraries pooled and the depth of sequencing achieved. The Strand-seq technology sacrifices coverage (depth and breadth) for long-range structural information that offers many new applications for genome biology, as discussed below. Finally, because the resulting sequencing data are directional, bioinformatic analysis must consider whether output reads map to the plus or the minus strand of the reference genome, meaning most standard genomic analysis methods cannot be immediately applied to the data.

Applications, advantages and limitations

The primary advantage of Strand-seq is that it combines the genomic resolution of sequencing technologies with the long-range structural information obtained in traditional cytogenetics. By sequencing only template strands, parental homologs can be directly distinguished and visualized in daughter cells based simply on read alignment. Changes in template strand states can reflect a DNA repair crossover event between chromatids arising in S-phase, or a genomic rearrangement that is either inherited or clonally propagated in many cells. The distinction between sporadic sister chromatid exchange (SCE) events and stable structural rearrangements

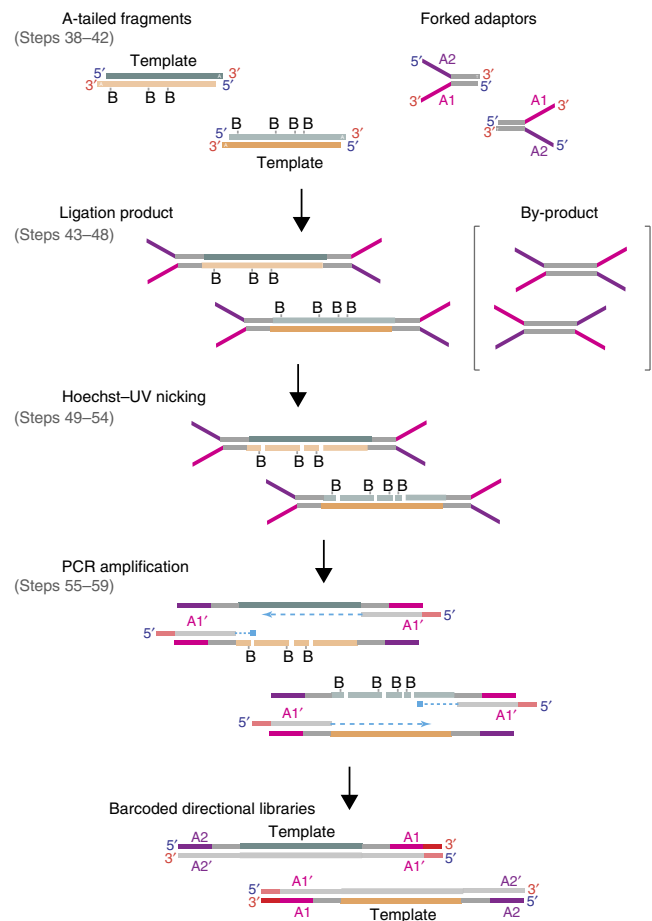


Figure 2 | Generation of directional template strand sequencing libraries. Schematic of the enzymatic steps from adaptor ligation through PCR amplification. MNase-digested chromatin is end-repaired and A-tailed before adding forked Illumina adaptors during the ligation reaction. The reaction also generates adaptor dimer by-products. Hoechst-UV treatment creates single-strand nicks at sites of BrdU incorporation (indicated by 'B'). During PCR amplification, the nicks inhibit elongation of the BrdU⁺ strand, allowing the selective amplification of template strands. The hexamer barcode (red) is introduced by the PE 2.0 primer during PCR amplification (Supplementary Figs. 1 and 2). The directional library contains the A2 adaptor at the 5' end and the A1 adaptor at the 3' end of the template strand. Libraries are pooled and submitted for sequencing on an Illumina platform.

is easily made by analysis of a few dozen cells²¹. Precisely locating the change in strand state allows fine mapping of these rearrangements at the kilobase scale, which is several orders of magnitude better than that for conventional cytogenetic techniques^{20,22}. Preserving the structure of each chromosomal homolog allows copy-neutral rearrangements, such as inversions and translocations, to be confidently localized even in complex and repetitive regions of the genome²¹. The directionality of Strand-seq data also allows direct genotyping of structural variants on a genome-wide scale. In addition, using template strands to distinguish and phase variants found on each parental homolog allows whole chromosomes to be rapidly haplotyped²³. This means haploid genomes can be reconstructed without the need for generational information. Finally, by treating template strand states as biallelic markers, genome reference assemblies can be refined by looking for co-inheritance of unplaced fragments. These applications are described in greater detail in the ANTICIPATED RESULTS section.

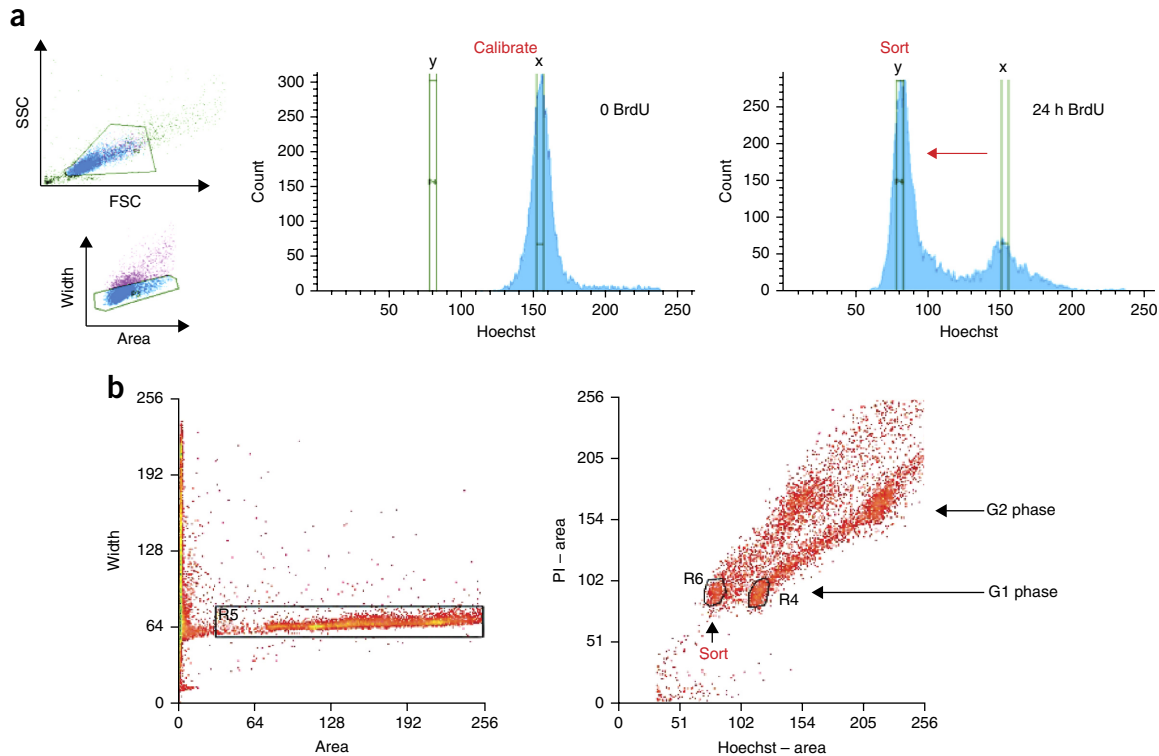


Figure 3 | Cell sorting gates based on Hoechst quenching. **(a,b)** Cells that undergo a single division in the presence of BrdU are sorted based on **(a)** Hoechst fluorescence and, optionally, **(b)** PI fluorescence. **(a)** FACS gates are first set to select nuclei based on forward (FSC) and side (SSC) scatter, and doublets are avoided by size. The negative-control sample (0 BrdU) is used to set up laser voltages, sorting gates and establish the Hoechst profile (visualized on a linear scale) for the undivided population. During calibration, a narrow gate (gate ‘x’) is set on the Hoechst channel to establish the mean fluorescence for this population (table). Notice that during calibration the Hoechst peak is placed in the right-hand half of the plot, to allow room for the quenched population (right-hand panel). The sorting gate (gate ‘y’) is set on the peak showing ~1/2 Hoechst fluorescence, compared with the undivided (0 BrdU) control. The percentage of cells that occupy this peak increases with BrdU exposure, indicating that cells have undergone mitosis and taken up BrdU (compare tables). **(b)** A plot of Hoechst area against PI area is used to distinguish cell cycle stage based on PI fluorescence (which is not affected by BrdU). Bivariate fluorescence plots comparing Hoechst against PI intensity for a mixture of cells cultured with and without BrdU allow identification of cells suitable for Strand-seq library construction. Cells in sorting region ‘R6’ have lower Hoechst fluorescence as compared with cells in region ‘R4’ and show PI fluorescence reflecting 2N DNA content (i.e., G1 phase). Cells in R6 with quenched Hoechst fluorescence signal are sorted as single cells for subsequent library preparation.

The principal limitation of Strand-seq is that the technology centers on hemi-substituted DNA to distinguish template strands. This requirement limits the method to cells that can incorporate BrdU during S-phase. Consequently, the protocol requires viable mitotic cells, and libraries cannot be generated from extracted DNA or fixed tissue samples. Here we focus on *in vitro* conditions to incorporate BrdU into genomic DNA; however, *in vivo* experiments in model organisms are also feasible. These will probably require several BrdU injections to ensure continuous and consistent incorporation of BrdU during S-phase (for which at least one BrdU analog is required for each ~150-bp interval)²⁴. As Strand-seq is a modified Illumina-based protocol, read length is currently limited to shorter fragments and requires sequencing on either a HiSeq or a MiSeq Illumina platform. In addition, the low read coverage obtained from individual libraries limits the size of the variant detected and the resolution of breakpoint mapping within a single cell. Indeed, coverage can be variable between cells from the same experiment, and between experiments. And, because of the multiple enzymatic and cleanup steps required, the protocol is expensive in terms of time (requiring up to 4 d to prepare 96 libraries) and money (costing ~\$1,300 USD per 96-well plate,

excluding sequencing). Future protocol advancements will aim to improve reproducibility and throughput.

A further caveat to this approach is that strand inheritance patterns are confounded in polyploid samples. For example, a trisomic chromosome will have three template strands sequenced that result in complex inheritance patterns (e.g., WWW, WWC, WCW, CWW, CWC, WCC, CCW, CCC), which are marked by differential read doses. As such, chromosomal rearrangement can be more challenging to locate within polyploid loci. For instance, an SCE event on one homolog of a trisomy causes a more ‘subtle’ transition in template strand states, with the ratio of reads changing by 33%, rather than 50%, as it would for a diploid chromosome. Nevertheless, copy-number variation (CNV) analyses can be performed on these data. By measuring shifts in read doses within a chromosome, and between chromosomes, CNV and aneuploidy events can be predicted. In addition, by incorporating the phasing information inherent in Strand-seq, haplotype support for variant predictions can be used to improve sensitivity of conventional single-cell approaches. Currently, bioinformatic analysis tools are under active development to combine strand inheritance patterns with CNV analyses, which we predict will

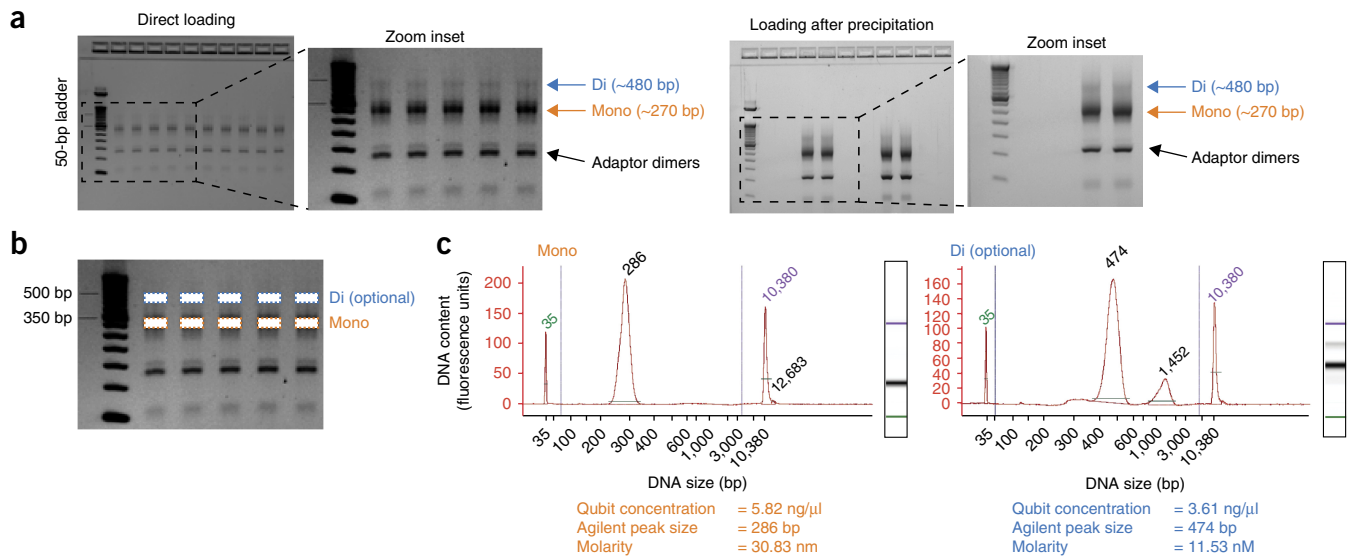


Figure 4 | Size selection of final Strand-seq library pool. Completed Strand-seq libraries are pooled and size-selected before sequencing. Size selection can be performed immediately after post-PCR cleanup ('Direct loading'), or following an ethanol precipitation step ('loading after precipitation'). (a) Gels of 96-pooled libraries are shown. Upon running the library pool on a 2% agarose gel, bands corresponding to dimer by-products, mononucleosomal fragments (Mono) and dinucleosomal fragments (Di) are visible. (b) The nucleosomal fragments are enriched by carefully excising the appropriate band(s) from the gel (dotted boxes). Multiple gel bands corresponding to products of similar size (i.e., those representing the mono- or dinucleosomal DNA insert sizes) can be pooled together for gel purification and sequencing. If the band is very faint (for example, the Di product), use the DNA ladder as a guide for cutting. (c) To calculate yield, the purified library is run on an Agilent Bioanalyzer, to confirm that the major peak is of the expected size. The concentration is also measured using a Qubit Assay. Using these readings, the molarity is calculated to ensure that the appropriate nanomolar (nM) concentration is loaded on the flow cell for sequencing.

be a powerful approach to studying structural variation with haplotype resolution.

We believe that the opportunity to distinguish homologs in single cells has far-reaching applications. We have already shown how Strand-seq can be used to map SCEs with unprecedented resolution²⁰, characterize genome-wide structural rearrangements, such as polymorphic inversions²¹, and build highly accurate chromosome-length haplotypes²³. Further applications include using contig strand states to refine reference genome assemblies, as well as mapping translocations and other chromosomal abnormalities, to study clonal heterogeneity and tumor evolution (data not shown). This single technology offers the potential to haplotype-resolve chromosomal rearrangements in normal and diseased individuals in a rapid and cost-effective manner, which has important implications for many scientific communities, including those studying population and evolutionary genetics, genome instability and integrity, cellular heterogeneity and lineage-tracing, cancer biology and personalized medicine.

Experimental design

BrdU Incorporation for a single-cell division. To generate Strand-seq libraries, cells are cultured in the presence of BrdU. Daughter cells arising from a single-cell division in BrdU contain hemi-substituted DNA, meaning that the nascent DNA strand has incorporated BrdU and the original template DNA strand has not (Fig. 1a). This allows the BrdU⁺ strands to be targeted and selectively ablated during genomic library construction, ensuring that only the BrdU⁻ template strands are sequenced. To accomplish this, cell cultures are initiated and BrdU (recommended 40 μM, final concentration) is added to the culture medium while cells undergo a single mitosis. The timing and dose of BrdU exposure is dependent on the length of the S-phase for the cells of interest

and their sensitivity to BrdU²⁴. Proper control experiments must ensure that the course of BrdU administration is restricted to one DNA synthesis phase (S-phase) of the cell cycle. Capturing daughter cells after a single cell division is critical to the protocol. Cells that have not undergone mitosis will have no, or partial, BrdU labeling (leading to variable 'whole-genome' sequencing data), whereas those that have undergone two cell divisions will contain DNA template strands with BrdU (resulting in chromosomal dropouts in the Strand-seq data). To harvest the desired cell population, single cells can be grown in the presence of BrdU and daughters mechanically separated by micromanipulation after the cell has divided. Although labor-intensive, the resulting daughter cells will have gone through only one DNA replication and will be hemi-substituted. To capture larger numbers of cells, bulk populations can be sorted by FACS; however, this can result in the inclusion of cells that have not undergone a complete cell division or that have undergone two cell divisions in the analysis. Purity is dependent on the quality of the sort, the size of the sorting gates and the duration of BrdU exposure. A difficulty here is that the growth rate of individual cells cannot always be established by measuring doubling time alone because primary human cell cultures can contain increasing numbers of nondividing cells over time, and BrdU exposure can delay the cell cycle²⁵. It is therefore worth first establishing the optimal concentration and length of BrdU pulse for the cells to be assayed. We recommend performing a flow cytometric time-course analysis for each cell type of interest to determine the percentage of hemi-substituted cells in culture, after various periods of BrdU exposure, based on quenching of Hoechst fluorescence intensity^{26,27} (as described in the following section). This will establish the best conditions for selecting hemi-substituted daughter cells from the culture. In addition, cell cultures can be synchronized, for instance, by

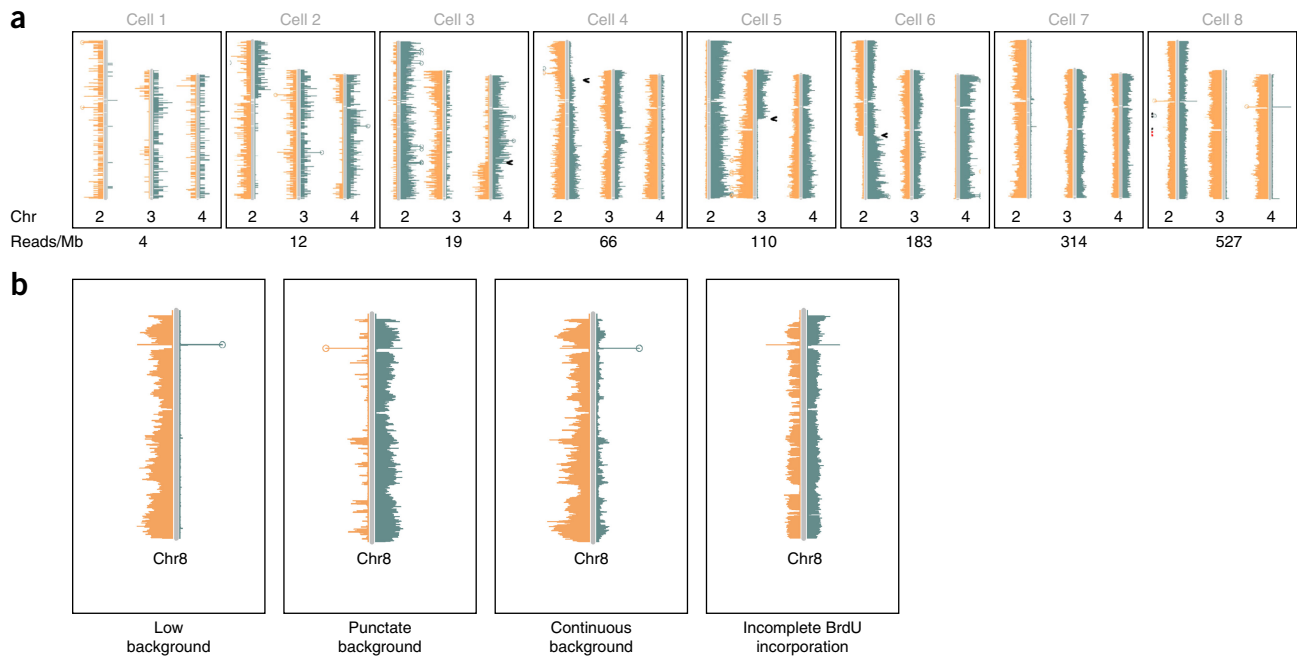


Figure 5 | Important parameters when considering Strand-seq libraries. **(a)** Human data sets from eight libraries with increasing depth. For three example chromosomes, the ability to resolve which template was inherited, and features such as SCE locations, increases at greater read depths. Although overarching trends can be seen even in libraries with less than 20 reads/Mb, the accuracy of strand-state change identification, breakpoint region localization and copy-number change calculation becomes unreliable. Black arrowheads represent BAIT-identified strand-state changes. **(b)** Examples of the different classes of background identified in Strand-seq libraries. Whereas most libraries are considered low background, some can be found with punctate regions of increased reads. These may be erroneously called by analysis software as strand-state changes or CNV. Some libraries display a continuous background, which still allows accurate calling of template strands, but hinders localization of breakpoints and haplotyping. In addition, depending on gating conditions during the FACS sort, some libraries may have gone through more than a single round of BrdU incorporation, leading to libraries with patches of missing reads for one or both templates.

G2 arrest using a CDK1 inhibitor, before BrdU treatment and sorting²⁰. Although this will help normalize the length of BrdU exposure, suspending the cell cycle may cause genetic aberrations that could confound downstream analyses.

Isolation of hemi-substituted daughter cells. To select the desired cell population for Strand-seq, hemi-substituted daughter cells are isolated from culture either by manual micromanipulation (to capture pairs of sister cells) or by sorting nuclei based on Hoechst quenching from bulk cultures (for population studies). For bulk cultures, cells containing hemi-substituted DNA can be captured based on the fluorescence intensity of Hoechst 33258 in the nucleus (Fig. 3). BrdU is incorporated into DNA instead of thymidine (T), and Hoechst is an intercalating fluorescent dye that preferentially binds double-stranded DNA at adenine–thymidine (A–T) base pairs. When incorporated into DNA, the bromine group of BrdU absorbs the emission energy of adjacent Hoechst dye molecules excited by light^{28,29}. Consequently, a cell that has hemi-substituted DNA (i.e., one strand of the DNA molecule contains BrdU) exhibits approximately half the fluorescence of a cell that did not incorporate BrdU into its DNA. We take advantage of this biochemistry to capture cells after one cell division by sorting on Hoechst fluorescence. For this, a BrdU⁻ control sample is included to identify the undivided population and calibrate the FACS system (Fig. 3a). Because these cells have not been exposed to BrdU, they will exhibit ‘full’ Hoechst fluorescence. To achieve the desired fluorescence profile, a gentle cell lysis step is performed that isolates nuclei for staining. Different nuclei extraction buffers can be tested on the

cell type of interest to optimize cell lysis and ensure that intact nuclei are obtained (**Supplementary Methods**). The stained nuclei are then run through a FACS machine, and single cells are gated, using size to discriminate doublets. When isolating cells by FACS, at least 10,000 cells are required to visualize a stable population in the gating plots, and we recommend harvesting a minimum of 5×10^5 cells for all samples to be analyzed on a cell sorter. Hoechst is excited with a UV (350-nm) laser, and a 488-nm detector records the emission. The BrdU⁻ control population visualized on a linear histogram should be placed in the right-hand quadrant of the FACS plot (Fig. 3a). The cell population that underwent one cell division in the presence of BrdU is identified as the peak showing substantially less Hoechst fluorescence as compared with that of control cells that did not incorporate BrdU. Optionally, nuclei can be counterstained with propidium iodide (PI), to identify cell cycle stage and gate on cells in G1 phase³⁰ (Fig. 3b). In either case, a narrow gate should be set to directly deposit the single-sorted nuclei into the wells of a 96-well plate for Strand-seq library preparation. During the sort, library preparation controls can be introduced by sorting no cells (negative control) or multiple cells (e.g., 10, as positive control) into some wells. The 0-cell controls should show no sequencing reads, and the 10-cell controls should show ~10-fold more coverage and exhibit a whole-genome sequence pattern (i.e., every chromosome represented by Watson and Crick reads). The total number of cells required for a Strand-seq experiment depends on the application, and we typically process between 94 and 188 single cells (allowing for one negative and one positive control per plate) per experiment.

Strand-seq library preparation. The single-sorted cells or nuclei are subjected to a modified paired-end sequencing library preparation protocol that (i) fragments chromatin using MNase, (ii) processes the fragment ends as substrates for adaptor ligation, (iii) ligates forked Illumina adaptors onto nucleosomal DNA fragments, (iv) removes the BrdU⁺ strand and (v) introduces a multiplexing barcode during PCR amplification (Figs. 1 and 2). It is critical to remember that the library is constructed from unamplified starting material (which is ~6–10 pg of DNA from a single cell), and therefore special care must be taken at all times to minimize sample loss. After each enzymatic reaction, a cleanup step is required to purify the DNA for subsequent processing. We use solid-phase reversible immobilization paramagnetic beads (Agencourt AMPure XP), which can also be optimized to select the desired product based on fragment size (Supplementary Fig. 3). This is important following adaptor ligation, when the excess of adaptors in the reaction mix (as compared with the availability of nucleosomal fragments) can generate adaptor-dimers that are preferentially amplified in the PCR reaction. The adaptor ligation reaction has been modified to help minimize this, and the subsequent size-selection step helps remove undesired dimers, for which we recommend using a 0.8× bead:DNA ratio to enrich for nucleosomal fragments (>150 bp) and remove smaller dimer-by-products.

Pooling and size selection. Following the PCR reaction, each single-cell library contains a unique hexamer barcode, which is introduced through the multiplexing PE 2.0 primer (Supplementary Fig. 2; Supplementary Table 1). This allows multiple single-cell libraries to be pooled together for size selection and sequencing. We typically combine between 96 and 192 samples (from one or two 96-well plates, respectively) for a single experiment. The entire volume of each uniquely labeled single-cell sample can be combined in a single microcentrifuge tube (Fig. 4). Alternatively, half of the volume of each library can be pooled together, with the remaining material reserved for future resequencing of individual or subsets of libraries. For instance, a MiSeq run can be performed on a complete pool to first determine the best libraries from the experiment, which can then be selectively pooled for targeted resequencing on a HiSeq platform. The library pool is subjected to a final agarose gel purification step, to ensure that a narrow size range of DNA is loaded onto the flow cell used for sequencing. Here, we aim to specifically enrich the mononucleosomal read fragments (~270 bp, following adaptor ligation and PCR amplification (Supplementary Fig. 3)) by excising the material between the 250- and 350-bp range (Fig. 4b). Optionally, the dinucleosomal read fragments (~480 bp, including adaptor sequences) can also be harvested for sequencing. In our experience, these larger fragments often represent more heterochromatic regions of the genome, and the overall genomic coverage of the final sequence data can be improved by including the dinucleosomal fragments with the more abundant mononucleosomal fragments. Often a visible band is seen at ~125 bp (Fig. 4a). This band represents adaptor dimers and should be strictly avoided during final size selection. The final size distribution is confirmed by running some of the product on an Agilent high-sensitivity chip, and the final concentration is determined by running a Quant-iT dsDNA HS assay kit on the Qubit fluorometer. This allows the final molarity of the purified product to be calculated for sequencing on a MiSeq or HiSeq Illumina sequencer (Fig. 4b).

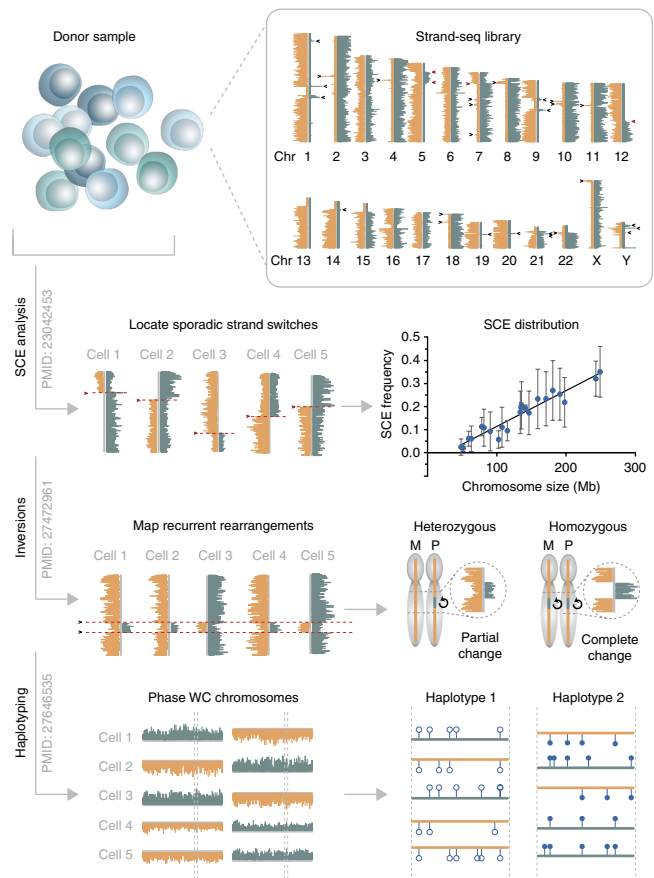


Figure 6 | Overview of genomic features evident from a Strand-seq experiment. From a single Strand-seq experiment consisting of a pool of single cells of interest (e.g., from a single donor sample), template strand inheritance patterns are first examined for each chromosome. By locating strand-state changes independently for each library, patterns can be analyzed across multiple Strand-seq libraries to (i) map sporadic sister chromatid exchanges with unprecedented resolution, (ii) locate and genotype stably inherited inversions genome-wide and (iii) assign phase variant alleles to either haplotype 1 or haplotype 2 across each chromosome of the donor genome.

The type of sequencing depends on the research question: although a 50-bp single-read sequencing protocol is typically sufficient for mapping breakpoints of structural rearrangements, 76-bp paired-end sequencing runs are more suitable for obtaining complete coverage of mononucleosomal fragments for haplotyping purposes. The dinucleosomes can be pooled with the mononucleosomes for sequencing on a single lane (at a 1:1.3 ratio, to reduce sequencing costs), or sequenced on separate lanes (for instance, using a 151-bp paired-end sequencing protocol, to improve overall coverage). Resulting sequencing reads are aligned to a reference assembly in order to assign template strand states.

Strand-seq data analysis. As the protocol centers on template strand information, the primary concerns in analyzing Strand-seq data involve determining (i) where in the reference assembly the sequencing read aligns, and (ii) whether the read is in the Crick (C; aligns to the plus (forward) strand of the reference assembly) or Watson (W; aligns to the minus (reverse) strand) orientation. Using this information, template strand states are assigned and the inheritance pattern is determined for each chromosome of the

PROTOCOL

sequenced cell. In a successful Strand-seq library derived from a diploid cell, expect to see ~50% of the chromosomes exhibiting either a WW or CC state. If all chromosomes in the library have a WC inheritance pattern (i.e., it appears as a ‘whole-genome’ sequencing library), Strand-seq was probably unsuccessful for that cell. By assessing the genomic coverage, level of background and proportion of WW:WC:CC chromosomes for each library

(Fig. 5), the success rate of the experiment can be assessed and successful libraries selected for further analyses. Within each successful Strand-seq library, template strand states can be monitored along individual chromosomes to detect genomic rearrangements, and locate copy-number changes and phase variant alleles (Fig. 6). Further details of these analyses are found in the ANTICIPATED RESULTS section.

MATERIALS

REAGENTS

- Cells of interest (e.g., EBV-transformed B-lymphocyte cell lines, such as GM12878²³, available from Coriell Institute for Medical Research)
! CAUTION The cell lines used in your research should be regularly checked to ensure that they are authentic and that they are not infected with mycoplasma.
- Cell culture medium appropriate for cell type of interest (e.g., RPMI 1640 medium (Gibco, cat. no. 11875101) supplemented with 15% FBS (Sigma-Aldrich, cat. no. F1051))
- PBS (1×; Life Technologies, cat. no. 14190250)
- Cell dissociation solution appropriate for the cell type of interest, if adherent (commonly Trypsin-EDTA, 0.25% (Life Technologies, cat. no. 25200056))
- Trypan blue solution, 0.4% (Sigma-Aldrich, cat. no. 93595)
- 5-Bromo-2'-deoxyuridine (BrdU; Sigma-Aldrich, cat. no. B9285)
! CAUTION BrdU is a mutagen and potential carcinogen that may cause skin irritation and be harmful if absorbed through the skin or inhaled. Handle it with care and use appropriate personal protective equipment.
- Ultrapure Hoechst 33258 (Enzo Life Sciences, cat. no. ENZ-52402)
! CAUTION This compound is a mutagen and is harmful if swallowed. It causes skin irritation, and it may cause respiratory irritation. It is suspected of causing genetic defects; handle it while wearing appropriate personal protective equipment. Keep it protected from light.
- Propidium iodide (PI; Sigma-Aldrich, cat. no. P4170) **! CAUTION** This compound is an irritant and can be harmful if swallowed. It is suspected of causing genetic defects; handle it while wearing appropriate personal protective equipment. Keep it protected from light.
- Nuclei Isolation Kit: Nuclei EZ Prep (Sigma-Aldrich, cat. no. NUC101-1KT)
- ProFreeze-CDM, nonanimal origin, chemically defined freeze medium (2×; Lonza, cat. no. 12-769E)
- DMSO, molecular biology grade (Sigma-Aldrich, cat. no. D8414)
- Ultrapure Tris-HCl buffer, pH 7.5 (1 M; Invitrogen, cat. no. 15567-027)
- NaCl solution, molecular biology grade (5 M; Sigma-Aldrich, cat. no. S5150)
- MgCl₂ solution, molecular biology grade (1 M; Sigma-Aldrich, cat. no. M1028)
- CaCl₂ solution, molecular biology grade (1 M; Sigma-Aldrich, cat. no. 21115)
- Bovine Albumin Fraction V (BSA, 7.5% (vol/vol) solution; Gibco, cat. no. 15260037)
- Nonidet P40 Substitute (NP40; Sigma-Aldrich, cat. no. 74385)
! CAUTION This compound contains materials that may cause respiratory tract, eye and skin irritation. It may be harmful if swallowed; handle it with appropriate care.
- Ultrapure glycerol, molecular biology grade (Invitrogen, cat. no. 15514011)
- EDTA solution, molecular biology grade (0.5 M, pH 8.0; Ambion, cat. no. AM9261)
- Ultrapure DNase/RNase-free dH₂O (Invitrogen, cat. no. 10977035)
- DTT (1 M; Sigma-Aldrich, cat. no. 43816) **! CAUTION** Direct contact can cause irritation.
- PEG 6000, molecular biology grade (Calbiochem, cat. no. 528877)
- Elution buffer (EB; Qiagen, cat. no. 19086)
- MNase (micrococcal nuclease), 200 U/μl (NEB, cat. no. M0247S; includes MNase reaction buffer (cat. no. B0247B)) **▲ CRITICAL** Frequently replace MNase stock. The enzyme is highly diluted in the single-cell reaction (1.25 units/sample), and therefore any drop in activity will be detrimental to library quality.
- T4 DNA polymerase (NEB, cat. no. M0203S; includes NEB buffer 2.1)
- dNTP set, 10 mM (NEB, cat. no. N0447S)

- T4 DNA ligase buffer with 10 mM ATP (NEB, cat. no. B0202S)
 - Klenow DNA polymerase (NEB, cat. no. M0210S; includes NEB buffer 2)
 - T4 polynucleotide kinase (NEB, cat. no. M0201S; includes T4 polynucleotide kinase reaction buffer)
 - dATP solution (100 mM; NEB, cat. no. N0440S)
 - Klenow fragment (3'→5' exo-; NEB, cat. no. M0212S; includes NEB buffer 2)
 - Quick ligation kit (NEB, cat. no. M2200L; includes Quick Ligase and 2× Quick Ligation Reaction Buffer)
 - Phusion HF PCR Master Mix (NEB, cat. no. M0531L; includes a 2× PCR Master Mix containing Phusion High-Fidelity DNA Polymerase and dNTPs)
 - Agencourt AMPure XP (60 ml; Beckman Coulter, cat. no. A63881)
▲ CRITICAL Make sure that AMPure XP beads are resuspended and equilibrated to room temperature (RT; 22–25 °C) for at least 30 min before addition to samples.
 - Ethanol, molecular biology grade (absolute, 200 proof; Fisher Scientific, cat. no. BP2818500) **! CAUTION** This compound is flammable; keep it away from open flame.
 - GenElute-LPA (Sigma-Aldrich, cat. no. 56575)
 - NaOAc solution, molecular biology grade (3M; Sigma-Aldrich, cat. no. 71196)
 - E-Gel EX Agarose Gels, 2% (Invitrogen, cat. no. G401002)
 - E-Gel EX Agarose Gels, 1% (Invitrogen, cat. no. G401001)
 - E-Gel 50 bp DNA Ladder (Invitrogen, cat. no. 10488099)
 - Zymoclean Gel DNA Recovery Kit (Zymo Research, cat. no. D4008)
 - Agilent high-sensitivity DNA kit (Agilent, cat. no. 5067-4626)
 - Qubit dsDNA high-sensitivity assay kit (Invitrogen, cat. no. Q32851)
▲ CRITICAL Qubit DNA binding dye is light-sensitive. Keep it protected from light.
 - DNA Away Surface Decontaminant (Thermo Fisher Scientific, cat. no. 7010)
 - Adaptor and primer oligonucleotides (see REAGENT SETUP section). All oligonucleotides can be ordered from Integrated DNA Technologies (IDT; <https://eu.idtdna.com/site>). Adaptor and index sequencing oligonucleotides are subjected to PAGE purification; the PCR amplification oligonucleotides are standard desalted and lyophilized oligonucleotides for shipping.
- ### EQUIPMENT
- Tissue culture vessels appropriate for cells of interest (e.g., T25 flasks (Thermo Fisher Scientific, cat. no. 169900))
 - Hemocytometer (Sigma-Aldrich, cat. no. Z359629)
 - Polystyrene round-bottom tube with cell-strainer cap (5 ml; Falcon, cat. no. 352235)
 - Cell sorter equipped with UV laser (e.g., FACSJazz or FASCARIA (BD Biosciences) or MoFlo Astrios (Beckman Coulter))
 - UV3 HEPA PCR Workstation (UVP, cat. no. 95-0438-04)
 - RNase- and DNase-free Microfuge tubes (1.5 ml; Ambion, cat. no. AM12400)
 - Amber microcentrifuge tubes, Safe-Lock (0.5 ml; Eppendorf, cat. no. 211-2145)
 - Polypropylene DNA LoBind Polypropylene Microcentrifuge Tubes (1.5 ml; Eppendorf, cat. no. 0030 108.051)
 - Screw-cap tubes with O-ring (VWR, cat. no. 211-0094)
 - Labtop cooler, -20 °C, Nalgene (VWR, cat. no. 471-0001)
 - Cold block: Eppendorf PCR cooler, iceless cold storage system for 96-well plates and PCR tubes (Eppendorf, cat. no. 3881000015)
 - Hard-shell 96-well PCR plates, clear wells (Thermo Scientific, cat. no. AB2396)
 - Hard-shell 384-well PCR plates, clear wells (Thermo Scientific, cat. no. AB2384)
 - Magnetic Bead Separation Block for 384-well PCR plates (V&P Scientific, cat. no. VP 771G-7A)
 - Microseal 'C' optical seals (Bio-Rad, cat. no. MSC1001)
 - SILVERseal, aluminum foil, temperature-resistant from -80 °C to +110 °C (Greiner Bio-One, cat. no. 676090)

- Easy Seals (Greiner Bio-One, cat. no. 676002)
- Pipetman G pipettes (P2G, P20G, P200G and P1000G, equipped with stainless steel ejectors; Gilson, cat. nos. F144054M and F167900)
- Low-retention filtered pipette tips: 0.5–2, 1–30, 1–200 and 100–1,000 μl (Corning, cat. nos. 4801, 4821, 4823, 4809)
- UV cross-linker, equipped with 365-nm Longwave bulbs (UVP, cat. no. CL-1000L) **! CAUTION** UV radiation is harmful to the unprotected skin and eye. Use appropriate personal protective equipment.
- Qubit assay tubes (Invitrogen, cat. no. Q32856)
- E-Gel iBase and E-Gel Safe Imager Combo Kit (Invitrogen, cat. no. G6465-EU)
- Gel Doc EZ System (Bio-Rad, cat. no. 170-8270)
- Blue sample tray (Bio-Rad, cat. no. 170-8273)
- Gel knife (Invitrogen, cat. no. EI9010)
- GeneCatcher disposable gel excision tips, 4.0 mm \times 1.0 mm (GelCompany, cat. no. PKB4.0)
- QBD2 dry block heater (Grant, cat. no. 11470958)
- Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad, cat. no. 185-1384)
- Centrifuge 5810R (Eppendorf, cat. no. 5811000320)
- Sorvall Legend MicroCL 17R (Thermo Scientific, cat. no. 75002440)
- MPS 1,000 Mini PCR Plate Spinner (Labnet International, cat. no. C1000)
- Spectrafuge Mini Laboratory Centrifuge (Labnet International, cat. no. C1301)
- Agilent 2100 Bioanalyzer (Agilent Technologies, cat. no. G2938C)
- Qubit 2.0 fluorometer (Invitrogen, cat. no. Q32866)
- Illumina DNA sequencing instrument (e.g., MiSeq, HiSeq, NextSeq)
- Bowtie2³¹ software package (v2.2.4; available from <https://sourceforge.net/projects/bowtie-bio/files/bowtie2/>)
- SAMtools³² software package (v0.1.18; available from <http://www.htslib.org/>)
- BAIT²² software package (available from <https://sourceforge.net/projects/bait/>)
- contiBAIT software package (version >1.0.0; available from <https://bioconductor.org>)
- BWA (<https://sourceforge.net/projects/bwa/>)
- bedtools (<https://github.com/arq5x/bedtools2/releases>)

REAGENT SETUP

General setup At all times, take care to avoid DNA or cross-well contamination. Clean all work surfaces and pipettes with DNA Away Surface Decontaminant (or 5% (vol/vol) bleach in H_2O) before use. Change gloves regularly and between procedural steps. Never reuse pipette tips.

Create working aliquots of all reagents to avoid any cross-contamination between batches of libraries. We recommend preparing aliquots of sufficient volume for the preparation of 96 libraries at a time (see Master Mix recipes and increase aliquots by ~10% to circumvent pipetting errors). Prepare aliquots in a UV3 HEPA PCR Workstation decontaminated by 30-min shortwave UV exposure and cleaned with DNA Away Surface Decontaminant. Upon thawing stock reagents, mix all buffers very well using a vortexer (except enzymes, which should never be vortexed) before preparing aliquots. Keep enzymes in labtop coolers at all times while preparing aliquots. Also prepare multiple 1.5-ml aliquots of ultrapure H_2O and elution buffer for Master Mixes and cleanup steps. For convenience, a protocol ‘cheat sheet’ is provided in the **Supplementary Methods**, which contains Master Mix recipes, bead cleanup and elution volumes. A 12-sample recipe is included for the case in which fewer samples (e.g., a single row of a 96-well PCR plate) are being prepared. **! CRITICAL** Do not use the same aliquots for different sets of libraries, and discard any unused Master Mixes. Do not vortex Master Mixes containing enzymes, rather pipette up and down several times to mix. Pulse-spin Master Mixes before use.

For the scaled protocol, we recommend using a robotic liquid-handling platform that is equipped to handle 96-well and/or 384-well plates, such as the Bravo Automated Liquid Handling Platform (Agilent Technologies). This will facilitate precise and rapid Master Mix dispensing during each procedural step.

For efficient delivery of reagents to multiple samples simultaneously, Master Mixes can be aliquotted into 96-well PCR plates (e.g., distributed evenly between the 12 wells of row A, or throughout all 96 wells). This should be done immediately before dispensing to samples; the Master Mix plate should be pulse-spun and kept in a 4 °C cold block at all times while dispensing. The liquid handler, or a multichannel pipette, can then be used to dispense Master Mixes.

BrdU solution, 40 mM Using sterile conditions, dissolve 50 mg of BrdU in 4.07 ml of 1 \times PBS to obtain a 40 mM BrdU stock solution. Vortex well and heat the solution at 37 °C until all precipitates are dissolved. Divide the solution into 400- μl aliquots of in 0.5-ml amber microcentrifuge tubes and store them at –80 °C for up to 1 year.

NP40, 10% (vol/vol) To prepare 10 ml of 10% (vol/vol) NP40, add 1 ml of NP40 to 9 ml of water and mix well. Store the solution at RT for up to 6 months.

Hoechst 33258 solution, 10 mg/ml Dissolve 100 mg of Hoechst 33258 in 10 ml of ultrapure water to prepare a 10 mg/ml solution. Prepare 200- μl aliquots in 0.5-ml amber microcentrifuge tubes and store them at –20 °C for up to 1 year. This Hoechst dilution is used for staining nuclei. **! CRITICAL** Keep Hoechst 33258 and Hoechst solution protected from light at all times.

Hoechst 33258 solution, 1 mg/ml Dilute 100 μl of 10 mg/ml Hoechst 33258 solution in 900 μl of ultrapure water to obtain a 1 mg/ml solution. Prepare 90- μl aliquots in 0.5-ml amber microcentrifuge tubes and store them at –20 °C for up to 3 months. This Hoechst dilution is used during the library preparation.

! CRITICAL Keep Hoechst solution protected from light at all times.

PI solution, 2 mg/ml Dissolve 100 mg of PI in 50 ml of ultrapure water to obtain a 2 mg/ml solution. Store the solution at 4 °C for up to 1 year.

! CRITICAL Keep PI protected from light at all times.

Nuclei staining buffer A To prepare 10 ml, add 1.0 ml of 1 M Tris–HCl, pH 7.5, 308 μl of 5 M NaCl, 10 μl of 1 M CaCl_2 , 5 μl of 1 M MgCl_2 , 266.5 μl of 7.5% BSA solution, 100 μl of 10% (vol/vol) NP40 and 10 μl of 10 mg/ml Hoechst 33258 to 8.3 ml of water. Keep the buffer on ice, and use it immediately. **! CRITICAL** Nuclei staining buffer (NSB) A does not contain PI and allows a narrow sorting gate to be set on a histogram of Hoechst fluorescence. **! CRITICAL** Freshly prepare NSB A on the day of sorting and protect it from light at all times. **! CRITICAL** This NSB is optimized for isolating and staining nuclei of human hematopoietic cells (e.g., lymphoblastic cell lines and primary cultures). If different cell types of interest are used, it may be necessary to test alternative NSB recipes (**Supplementary Methods**).

Ultimately, aim to isolate single and intact nuclei stained with Hoechst \pm PI.

NSB B To prepare 10 ml, add 50 μl of 2 mg/ml PI to NSB A. Keep the solution on ice, and use it immediately. **! CRITICAL** NSB B contains PI for optional cell-phase analysis, and requires a 2D sorting gate to be set.

! CRITICAL Freshly prepare NSB B on the day of sorting and protect it from light at all times. **! CRITICAL** This NSB is optimized for isolating and staining nuclei of human hematopoietic cells (e.g., lymphoblastic cell lines and primary cultures). If different cell types of interest are used, it may be necessary to test alternative NSB recipes (**Supplementary Methods**).

Ultimately, aim to isolate single and intact nuclei stained with Hoechst \pm PI.

Freeze buffer Combine 425 μl of 2 \times ProFreeze–CDM with 75 μl of 100% DMSO and 500 μl of 1 \times PBS to obtain 1 ml of freeze buffer. Use it directly to prepare sort plates.

EDTA, 100 mM Dilute 2 ml of 0.5 M EDTA stock with 8 ml of water. Mix it well and divide it into 1-ml aliquots; store it at RT for up to 6 months.

MNase dilution buffer The MNase dilution buffer consists of 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA and 50% (vol/vol) glycerol.

To prepare 1 ml, add 10 μl of 1 M Tris–HCl, pH 7.5, 10 μl of 5 M NaCl, 10 μl of 100 mM EDTA and 500 μl of glycerol to 470 μl of water. Mix well and divide the buffer into 19- μl aliquots; store them at –20 °C for up to 6 months.

! CRITICAL The MNase enzyme will be diluted directly in these aliquots; therefore, precisely 19 μl must be deposited.

MNase, 10 U/ μl Dilute 1 μl of MNase stock solution (200 U/ μl) in 19 μl of MNase dilution buffer. Keep the solution on ice, and use it immediately.

! CRITICAL Keep the enzyme on ice during use. Prepare this diluted MNase solution just before use when the other reagents have already been added to the MNase Master Mix. Add to the Master Mix immediately before dispensing.

PEG 6000, 50% (wt/vol) Dissolve 5 g of PEG 6000 powder in 10 ml of water. Autoclave to sterilize, divide the solution into 300- μl aliquots and store them at RT for up to 1 year.

DTT, 200 mM Dilute 200 μl of 1 M DTT stock with 800 μl of water. Mix well and divide the solution into 18- μl aliquots; store them at –20 °C for up to 6 months. **! CRITICAL** Vortex the DTT stock very well and ensure that no precipitates are visible in solution before diluting to working stock.

dATP, 10 mM Dilute 80 μl of 100 mM dATP in 720 μl of water. Mix well and divide the solution into 36- μl aliquots; store them at –20 °C for up to 6 months.

PROTOCOL

10× Annealing buffer The annealing buffer consists of 100 mM Tris–HCl, pH 7.5, 1 M NaCl and 10 mM EDTA. To prepare 1 ml, combine 100 µl of 1 M Tris–HCl, pH 7.5, 200 µl of 5 M NaCl and 20 µl of 0.5 M EDTA with 680 µl of water. Use freshly prepared buffer for adaptor annealing.

Illumina PE adaptor-1 5′-75Phos/

GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3′: dissolve the lyophilized oligonucleotide in water to a final concentration of 110 µM. Store the solution at –20 °C for up to 3 years. Discard after 10 freeze–thaw cycles.

▲ **CRITICAL** This oligonucleotide contains a 5′ phosphate group.

Illumina PE adaptor-2 5′-ACACTCTTCCCTACACGACGCTCTTCCGATC*T-3′: dissolve the lyophilized oligonucleotide in water to a final concentration of 110 µM. Store the solution at –20 °C for up to 3 years. Discard after 10 freeze–thaw cycles. ▲ **CRITICAL** This oligonucleotide contains a 5′–3′ phosphorothioate linkage between the last two nucleotides (indicated by *).

Annealed illumina PE adaptors To make 50 µM annealed adaptors, mix 45 µl of PE adaptor-1 and 45 µl of PE adaptor-2 (each at 110 µM) with 10 µl of 10× annealing buffer in a screw-cap tube with an O-ring. Place the sealed tube in a heat block at 95 °C for 5 min. Turn off the heat block and let the tube equilibrate to RT. Divide the annealed adaptors into aliquots in multiple tubes (10 µl each) for storage at –20 °C for up to 1 year. Dilute annealed adaptors to a working concentration of 1 µM by adding 10 µl of 50 µM stock to 490 µl of EB. Dispense into 80-µl aliquots and store them at –20 °C for up to 3 months.

Illumina PCR Primer PE 1.0 5′-AATGATACGCGACCACCGAGATCTCACTCTTCCCTACACGACGCTCTTCCGATCT-3′: dissolve the lyophilized PE 1.0 primer oligonucleotide in EB to a 100 µM stock solution. To make 1 ml of 10 µM working PE 1.0 primer solution, dilute 100 µl of the stock in 900 µl of EB. Divide both stock (100 µl each) and working solution (120 µl each) into multiple aliquots for storage at –20 °C. Stock solutions can be stored for up to 3 years. Working solutions can be stored for up to 6 months. Discard after 10 freeze–thaw cycles. ▲ **CRITICAL** The PE 1.0 primer is delivered to all samples, and is added to the PCR Master Mix.

Custom multiplexing PCR Primer PE 2.0 5′-CAAGCAGAAGACGGCATACGAGATNNNNNCGGTCTCGCATTCCTGCTGAACCGCTCT

TCCGATCT-3′: if lyophilized, dissolve the PE 2.0 primer oligonucleotides in EB to make a 100 µM stock solution. Dilute primers to a 10 µM working solution by adding 10 µl of stock solution to 90 µl of EB. Divide the solution in multiple 96-well plates with 10 µl per well (multiplexing primer plate). Store both stock and working solutions at –20 °C. Stock solutions can be stored for up to 3 years. Working solutions can be stored for up to 6 months. Discard after 5 freeze–thaw cycles. ▲ **CRITICAL** The multiplexing primers are custom-designed to contain a unique fault-tolerant hexamer barcode with a minimal hamming distance of 2, at position ‘NNNNNN’ (**Supplementary Table 1**). These primers are used for barcoding single-cell libraries, and each PE 2.0 primer is delivered separately to each sample. ▲ **CRITICAL** Custom primers can be ordered in a 96-well plate format already suspended at a 100 µM stock concentration. If ordered separately, resuspend the individual primers to stock concentration and dispense them into a 96-well plate for efficient dilution and delivery to samples.

Custom index sequencing primer 5′-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG-3′: dissolve the lyophilized oligonucleotide in EB to obtain a 100 µM stock solution. Create 10-µl aliquots and store them at –20 °C for up to 3 years. Discard the stock after 10 freeze–thaw cycles.

▲ **CRITICAL** This custom index sequencing primer is used to sequence the hexamer barcode.

EQUIPMENT SETUP

UV cross-linker 8-W 254-nm Shortwave UV bulbs come standard issue with most UV cross-linker ovens. Be sure to replace these bulbs with 8-W 365-nm longwave UV bulbs. ▲ **CRITICAL** The higher-wavelength bulbs are used to induce debromination of BrdU and subsequent nicking of DNA at sites of BrdU incorporation with minimal formation of pyrimidine dimers and free radicals that are induced by shorter-wavelength UV irradiation.

Magnetic bead separation block Place a 3- to 3.5-mm solid rubber gasket around the magnet to adjust the height of the 384-well PCR plate during AMPure XP bead cleanup steps (**Supplementary Fig. 4**). This ensures that the paramagnetic bead pellet collected at the side of the well will be positioned lower in the well, enabling reduced elution and reaction volumes.

PROCEDURE

Isolation of single cells ● TIMING 3 h

1| Culture cells according to accepted protocols specific to the cell type of interest.

▲ **CRITICAL STEP** If you plan to test different BrdU-exposure periods (described in Step 3), initiate multiple cultures for the experiment. At a minimum, have at least two cultures set up—one sample will be used as a BrdU⁻ control, and the other(s) will receive BrdU (BrdU⁺). The BrdU⁻ sample is used to calibrate the FACS sorting gates (**Fig. 3**, and Step 9), and the BrdU⁺ sample is sorted for library preparation (Steps 10 and 11).

2| To semiconfluent cultures, when the cells are in logarithmic phase and proliferating exponentially, add 40 µM of BrdU (final concentration) to the BrdU⁺ sample(s) by dispensing 40 mM BrdU solution at a 1:1,000 ratio to the culture volume.

! **CAUTION** BrdU may cause skin irritation and may be harmful if absorbed through the skin or inhaled. Work with appropriate personal protective gear.

▲ **CRITICAL STEP** Ensure that you have a BrdU⁻ sample, which is required to calibrate gates for FACS sorting (Step 9).

3| Upon adding BrdU, culture cells of interest for a single round of cell division (e.g., 24 h) to allow BrdU to incorporate into newly synthesized DNA strands.

! **CAUTION** BrdU can slow the cell cycle. Therefore, we advise starting with two time points (e.g., a 24-h and a 36-h culture period) to increase the opportunity to capture cells after a single cell division. The appropriate cell population will be selected during FACS sorting (in Step 10).

▲ **CRITICAL STEP** The total length of BrdU incorporation will depend on the length of S-phase, and must be determined experimentally for each cell type, as described in the Experimental Design section.

- 4| Harvest cells (including the BrdU⁻ and BrdU⁺ samples) either by trypsinization of monolayer cultures (resuspend in complete medium) or by collecting suspension cultures in medium.
- 5| Perform a viable cell count (e.g., using a hemocytometer) and gently centrifuge the cells at 300g for 4 min at RT.
- 6| Resuspend the cell pellet in cold NSB (see REAGENT SETUP section) to a concentration of 1×10^6 cells/ml.
! CAUTION NSB contains Hoechst, and optionally PI (NSB B), which are mutagens. Handle with gloves and dispose of safely in accordance with institutional regulations.
▲ CRITICAL STEP Keep NSB and resuspended cells protected from light at all times.
- 7| Incubate the sample on ice for at least 15 min, no longer than 3 h. Before sorting, filter the sample through a cell strainer into a 5-ml round-bottom Falcon tube. During incubation, prepare sorting plates (see Step 8)
! CAUTION Do not vortex the sample. Vortexing can cause substantial damage to the nuclei.
▲ CRITICAL STEP To ensure that nuclei isolation was successful, stain a small aliquot with trypan blue (1:1) and visualize with a hemocytometer (**Supplementary Methods**). If successful, dark blue and intact nuclei that are free of cell membrane will be visible.
? TROUBLESHOOTING
- 8| Prepare sorting plates by dispensing 5 μ l of freeze buffer (see REAGENT SETUP) per well of a hard-shell 96-well PCR plate. Keep on ice or at 4 °C until you are ready to sort. Alternatively, prepare sorting plates with 5 μ l of Nuclei EZ Prep lysis buffer if proceeding directly to MNase digestion (Step 12).
▲ CRITICAL STEP Plates should be hard-shell to resist warping in the FACS machine. Any distortion of plate shape can result in imprecision during the single-cell sort.
- 9| Calibrate the FACS machine (**Fig. 3a**). Run the BrdU⁻ sample on the sorter to set gates on populations of interest. Single nuclei are gated based on size (using forward scatter (FSC), side (SSC) scatter) and width (to avoid doublets). Using the UV laser in linear scale, place the Hoechst⁺ population on the right-hand side of the plot. A gate can be drawn on this population to calculate the mean Hoechst fluorescence of the BrdU⁻ sample.
▲ CRITICAL STEP For optimal results, set sort precision to single-cell mode with either a 0.5- or 1-drop envelope. To ensure efficient and accurate single nuclei deposition, sort single fluorescent beads into a 96-well plate and check the presence of these beads at the bottom of each well using a fluorescence microscope (optional). Before each sort, check machine alignment by sorting single drops onto a sealed or empty 96-well plate and examining their position. The sorted drops should be centered precisely in the middle of the wells before proceeding (to Step 10).
- 10| Load the BrdU⁺ sample to visualize the divided population, and draw the sorting gate. Cells that divided and took up BrdU should show $\sim 1/2$ the Hoechst fluorescence of the BrdU⁻ sample, appearing in the left-hand side of the UV plot. Any cells that overlap with the BrdU⁻ sample have not incorporated BrdU and/or did not divide in culture. Draw a very narrow sorting gate on the divided population (as illustrated in **Fig. 3**).
! CAUTION A shift in the FSC/SSC profile might be seen for the BrdU⁺ sample, which is probably due to the thymidine analog altering the granularity of the cell. Adjust this gate accordingly to the population.
▲ CRITICAL STEP Exact sorting gates depend on which NSB is used. For NSB B, DNA content can be analyzed. Set parameters for Hoechst and PI area, width and height (**Fig. 3b**).
? TROUBLESHOOTING
- 11| Sort single nuclei from Step 10 into single wells of the sorting plate prepared during Step 8 (i.e., containing 5 μ l of freeze buffer or Nuclei EZ Prep lysis buffer). Immediately after sorting, cover the sample plate with a sealing film (use a SILVERseal if samples will be frozen) and spin in a prechilled (4 °C) centrifuge for 5 min at 500g.
▲ CRITICAL STEP To include experimental controls for library preparation, leave at least one well empty during the sort (0-cell negative control), and sort 10 nuclei into at least one well (10-cell positive control). Be sure to record the well numbers for controls.
▲ CRITICAL STEP If nuclei are sorted into Nuclei EZ lysis buffer, proceed immediately with MNase digestion (Step 12).
? TROUBLESHOOTING
■ PAUSE POINT Sample plates containing freeze buffer can be frozen at -80 °C until further processing, up to 12 months.

PROTOCOL

MNase digestion ● TIMING 30 min

12| Thaw MNase reagents and prepare MNase Master Mix in a microcentrifuge tube using the below recipe (omitting MNase* enzyme at this time). Mix thoroughly, pulse-spin the tube and keep it on ice.

MNase Master Mix			
Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
10× MNase buffer	1.5	30	216
200 mM DTT	0.11	2.2	16
50% PEG 6000	1.51	30.2	216
Ultrapure H ₂ O	6.83	136.6	974
10 U/μl MNase*	0.05	1	7.2
Final volume	10	200	1,429.2

▲ **CRITICAL STEP** Store all Master Mix reagents (here and for all subsequent reactions) as aliquots at $-20\text{ }^{\circ}\text{C}$ (see Reagent Setup, General section) and thaw them on ice immediately before use.

▲ **CRITICAL STEP** Do not dilute or add MNase enzyme to the Master Mix until the pulse-spin is in progress. Do this immediately before dispensing Master Mix to samples (i.e., during Step 14).

13| Thaw frozen sample plate (from Step 11, if frozen) and spin the plate briefly (500g, 30 s, $4\text{ }^{\circ}\text{C}$) to ensure that samples are

at the bottom of wells. Samples should be in 5 μl of lysis or freezing buffer, in a 96-well plate. Once the spin is complete, take the plate containing samples to the bench; do not place them on ice.

▲ **CRITICAL STEP** Before each step, it is good practice to pulse-spin all plates in a Mini PCR Plate Spinner or in a prechilled centrifuge (500g, 5 s at $4\text{ }^{\circ}\text{C}$) to remove any air bubbles from solutions.

14| Dilute MNase enzyme to 10 U/μl in MNase dilution buffer (as described in REAGENT SETUP). Mix gently. Add the diluted MNase to MNase Master Mix (from Step 12) and mix well. Dispense complete Master Mix to samples immediately, as described in Step 15.

15| Add 10 μl of MNase Master Mix (completed in Step 14) to each sample in the plate (from Step 13). Use pipette tips to mix very well (20–30 times). The final reaction volume is 15 μl.

! **CAUTION** Take care while distributing and mixing reaction volumes to avoid cross-contamination of adjacent wells.

▲ **CRITICAL STEP** Store Master Mix on ice or in a $4\text{ }^{\circ}\text{C}$ cold block while distributing to samples.

▲ **CRITICAL STEP** Reaction timing is critical to obtaining mononucleosomal fragments. Dispense Master Mix to all wells simultaneously to ensure equal incubation time.

16| Once all wells receive MNase Master Mix, allow digestion to proceed at RT for 8 min. Cover the plate during incubation with a sealing film (i.e., Easy Seals).

17| After 8 min, stop each reaction by adding 1.66 μl of 100 mM EDTA to each sample (final EDTA is 10 mM). Mix the samples very well. Total reaction volume is 16.66 μl.

▲ **CRITICAL STEP** Reaction timing is critical to obtaining mononucleosomal fragments. Add EDTA to all wells simultaneously to ensure equal incubation time.

18| Proceed immediately to AMPure bead cleanup (following Steps 19–32). During Step 21 (after adding AMPure XP beads to samples), transfer the samples with beads from the 96-well PCR plate to a 384-well PCR plate. This is required to fit the plate on the magnetic block. To transfer, use one well per quadrant of the 384-well plate (such that wells A1, B2 and C3 of the 96-well plate are, respectively, transferred to wells A1, C3 and E5 of the 384-well plate). All subsequent procedural steps can be performed in 384-well PCR plates.

■ **PAUSE POINT** After cleanup, the sample plate can be spun down and frozen overnight at $-20\text{ }^{\circ}\text{C}$, until proceeding to end repair (Step 33).

AMPure bead cleanup ● **TIMING 45 min**

19| Resuspend AMPure XP beads by gentle shaking. Take the aliquot required for the cleanup step (see table in Step 21, and multiply by total number of samples), and equilibrate the beads to RT.

! CAUTION Do not vortex, as this can crush the beads.

20| Prepare 70% (vol/vol) EtOH, for two wash steps, as follows, and keep it on ice:

70% EtOH wash buffer			
Reagent	Per cell (μl)	12 Cells (μl)	96-Well plate (ml)
100% EtOH	42	700	4.9
Ultrapure H ₂ O	18	300	2.1
Final volume	2 × 30	1,000	7

▲ CRITICAL STEP 70% ethanol is hygroscopic. For optimal results, prepare fresh 70% ethanol for each cleanup step and use it immediately.

21| Add AMPure XP to each sample from the specified Procedure Step, at the volume indicated in the following table. Mix very well by pipetting up and down at least 30 times to evenly distribute the beads in the sample.

AMPure XP bead volume	
Procedure Step sample is taken from	Volume AMPure XP to add per cell (μl)
MNase (Step 18)	30
End repair (Step 37)	22.5
A-tailing (Step 42)	18
Adaptor ligation (Step 48)	20
Post-PCR (Step 59)	20

22| Incubate the samples at RT for 5 min.

23| Place the samples on a magnetic block for at least 2–5 min, to clear the sample of beads.

▲ CRITICAL STEP The solution is quite viscous; ensure that it is completely clear before removing the supernatant. Leave the samples longer on a magnet if necessary.

24| With the plate remaining on the magnetic block, carefully remove the supernatant and discard it.

▲ CRITICAL STEP Avoid disrupting the beads while removing the supernatant. This is best achieved by placing the pipette tip on the wall opposite the magnet/bead pellet while aspirating (**Supplementary Fig. 4**).

25| Perform wash 1: add 30 μl of 70% EtOH, running EtOH gently over the side of the well where magnetic particles have adhered. Do not dislodge the pellet. With the plate remaining on the magnetic rack, carefully remove the supernatant and discard it.

26| Perform wash 2: repeat Step 25.

27| When the second wash is complete and the supernatant has been removed, cover the plate with sealing film, remove it from the magnetic block and pulse-spin for 1 s—just enough to force residual EtOH to the bottom of the wells.

! CAUTION Do not over-spin; otherwise, the magnetic beads will clump together and will be harder to resuspend in EB.

28| After the pulse-spin, place the plate back on the magnetic block and remove traces of EtOH.

29| Leave the plate uncovered (while on the magnetic block) and allow the beads to dry for 2–2.5 min.

! CAUTION Avoid overdrying beads, which causes them to dislodge and flake away.

▲ CRITICAL STEP Check inside surface of beads for any sheen, which should disappear when beads are dry. In addition, in areas in which beads are thicker, tiny cracks appear when beads are dry.

PROTOCOL

30| Once the beads are dry, remove the plate from the magnetic rack and elute the DNA with EB. The EB volume required after each Procedure Step is indicated in the following table. Add EB directly over the bead pellet, and ensure that it is covered with the buffer (V). Allow samples to stand at RT for 2 min to rehydrate the beads; you should see the beads start to distribute within the buffer.

Elution buffer (EB) volume	
Procedure Step sample is taken from	Volume of EB to add Per cell (μl)
MNase (Step 18)	10
End repair (Step 37)	8.5
A-tailing (Step 42)	12.5
Adaptor ligation (Step 48)	9.5
Post-PCR (Step 59)	4

31| Pipette up and down many times (20–30) to ensure full resuspension and a homogeneous light brown color.

! CAUTION Pay special attention while resuspending beads to ensure that there are no bead aggregates, which could reduce yield.

? TROUBLESHOOTING

32| Once all wells have been resuspended, place the plate back on the magnetic block to collect beads on the side. Wait until the supernatant is clear (up to 5 min), then remove the eluant (which contains the templates for the next step) and place it into a new 384-well PCR plate.

! CAUTION When recovering eluant, avoid bead carry-over, as beads can interfere with subsequent reaction steps (e.g., PCR amplification).

▲ CRITICAL STEP The full volume of the EB can be recovered. However, the surface tension of the EB will tend to pull beads away from the side of the well and could result in the transfer of some beads.

End repair ● TIMING 45 min + 45 min for AMPure bead cleanup

33| Thaw MNase-digested samples (if frozen), pulse-spin briefly and keep them on ice until the Master Mix is prepared. Each sample should be in 10 μl of EB buffer in a 384-well PCR plate.

34| Thaw end repair reagents and prepare Master Mix in a microcentrifuge tube using the recipe below. Mix thoroughly, pulse-spin and dispense to samples immediately.

End Repair Master Mix			
Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
10 \times T4 ligase buffer	1.25	27.5	145
10 mM dNTP mix	0.5	11	58
T4 DNA polymerase	0.25	5.5	29
Klenow	0.05	1.1	5.8
T4 PNK	0.25	5.5	29
Ultrapure H ₂ O	0.2	4.4	23.2
Final volume	2.5	55	290

▲ CRITICAL STEP Vortex the T4 DNA ligase buffer well to resolubilize any precipitates before adding to Master Mix.

▲ CRITICAL STEP Store Master Mix on ice or in a 4 °C cold block while distributing to samples.

35| Add 2.5 μl of End Repair Master Mix (prepared in Step 34) to each sample (from Step 33). Mix very well. The final reaction volume is 12.5 μl .

36| Once End Repair Master Mix has been added to all wells, allow the reaction to proceed at RT for 30 min. Cover the plate during incubation with sealing film.

37| Proceed to AMPure bead cleanup (by repeating Steps 19–32).

■ **PAUSE POINT** After cleanup, the sample plate can be spun down and frozen overnight at $-20\text{ }^{\circ}\text{C}$, until proceeding to A-tailing (Step 38).

A-tailing ● **TIMING 45 min + 45 min for AMPure bead cleanup**

▲ **CRITICAL** A-tailing and adaptor ligation reactions must be done sequentially.

38| Thaw end-repaired samples (if frozen), pulse-spin briefly and keep them on ice until Master Mix is prepared. Each sample should be in $8.5\text{ }\mu\text{l}$ of EB buffer in a 384-well PCR plate.

39| Thaw A-tailing reagents and prepare Master Mix in a microcentrifuge tube using the recipe below. Mix thoroughly, pulse-spin and immediately proceed to the next step.

A-tailing Master Mix

Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
NEB buffer 2	1	30	132
10 mM dATP	0.25	7.5	33
Klenow (exo-)	0.25	7.5	33
Final volume	1.5	45	198

▲ **CRITICAL STEP** Store Master Mix on ice or in a $4\text{ }^{\circ}\text{C}$ cold block while distributing to samples.

40| Add $1.5\text{ }\mu\text{l}$ of A-Tailing Master Mix (prepared in Step 39) to each sample (from Step 38). Mix the samples very well. The final reaction volume is $10\text{ }\mu\text{l}$.

41| Once A-Tailing Master Mix has been added to all wells, allow the reaction to proceed at $37\text{ }^{\circ}\text{C}$ for 30 min. Use a prewarmed heating block or a PCR machine. Cover the plate during incubation with Microseal 'C' Optical sealing film.

▲ **CRITICAL STEP** If you are using a heating block, be sure to prewarm it by setting it to $37\text{ }^{\circ}\text{C}$ before beginning this procedure (i.e., at the beginning of Step 38)

42| Proceed to AMPure bead cleanup (see Steps 19–32).

▲ **CRITICAL STEP** Do not freeze the plate. Proceed directly to adaptor ligation.

Adaptor ligation ● **TIMING 30 min + 45 min for AMPure bead cleanup**

▲ **CRITICAL** A-tailing and adaptor ligation reactions must be done sequentially.

43| Pulse-spin A-tailed samples and keep them on ice until the Master Mix is prepared. Each sample should be in $12.5\text{ }\mu\text{l}$ of EB buffer in a 384-well PCR plate.

44| Thaw adaptor ligation reagents and prepare Master Mix in a microcentrifuge tube, using the below recipe. Mix Master Mix thoroughly, pulse-spin and immediately proceed to the next step.

Adaptor Ligation Master Mix

Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
$10\times$ T4 ligase buffer	2	28	213.3
$1\text{ }\mu\text{M}$ PE adaptors	0.67	9.4	71.5
Quick Ligase	1.08	15.1	115
Ultrapure H_2O	3.75	52.5	400.2
Final volume	7.5	105	800

▲ **CRITICAL STEP** Store Master Mix on ice or in a $4\text{ }^{\circ}\text{C}$ cold block while distributing to samples.

▲ **CRITICAL STEP** By performing adaptor ligation in the absence of the crowding reagent PEG 6000 (which is contained in the Quick Ligation Reaction Buffer), this recipe aims to reduce the formation of adaptor dimers and promote efficient ligation of adaptors to DNA templates, resulting in increased library yield.

PROTOCOL

45| Add 7.5 μl of adaptor ligation master mix (prepared in Step 44) to each sample (from Step 43). Mix very well. The final reaction volume is 20 μl , and the final concentration of PE adaptors is 33.5 nM.

46| Once adaptor ligation master mix has been added to all wells, allow the reaction to proceed at RT for 15 min. Cover the plate with sealing film during incubation.

47| After 15 min, dilute the samples by adding 5 μl of EB. Mix very well. The total volume should be 25 μl .

48| Proceed to AMPure bead cleanup (see Steps 19–32).

▲ **CRITICAL STEP** Do not increase the volume of beads in this AMPure cleanup step. After adaptor ligation, ensure that a 0.8:1 ratio (bead/sample) ratio is used to minimize capture of adaptor dimer byproducts (Supplementary Fig. 3).

■ **PAUSE POINT** After cleanup, the sample plate can be spun down and frozen overnight at $-20\text{ }^{\circ}\text{C}$, until Hoechst–UV treatment (Step 49).

Hoechst–UV treatment ● **TIMING 40 min**

▲ **CRITICAL** Hoechst–UV treatment and PCR amplification must be done sequentially.

49| Thaw adaptor-ligated samples (if frozen), pulse-spin briefly and keep them on ice until Hoechst is prepared. Each sample should be in 9.5 μl of EB buffer in a 384-well PCR plate.

50| Prepare 100 ng/ μl Hoechst 33258 using the below recipe. Use an amber microcentrifuge tube to keep the Hoechst protected from light.

Hoechst Treatment Master Mix

Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
1 mg/ml Hoechst 33258	0.1	10	80
Ultrapure H_2O	0.9	90	720
Final volume	1	100	800

! **CAUTION** Hoechst is a mutagen and suspected of causing genetic defects; handle it with care and while wearing appropriate personal protective equipment.

▲ **CRITICAL STEP** Keep Hoechst protected from light at all times.

51| With lights switched off, add 1 μl of Hoechst 33258 (10 ng/ μl , final concentration; prepared in Step 50) to each sample (from Step 49). Dispense to all wells simultaneously and mix it very well. The final volume should be 10.5 μl .

! **CAUTION** Dispose of Hoechst plate, tips, tubes and gloves in accordance with local and national regulations in regard to cytotoxic waste.

52| Incubate for 15 min at RT, while protected from light. Cover the plate during incubation with sealing film and a piece of foil.

53| Uncover the plate and place it in a UV cross-linker balanced on top of a glass beaker.

▲ **CRITICAL STEP** Balancing the plate on a glass beaker, at an $\sim 9\text{-cm}$ distance from the bulbs, places samples closer to the UV source and prevents occlusion of UV rays.

▲ **CRITICAL STEP** For peak Hoechst excitation, and to minimize the likelihood of formation of cyclobutane pyrimidine dimers, ensure that the cross-linker is equipped with 365-nm longwave UV bulbs.

54| While uncovered, irradiate with UV radiation for 15 min. The total dose should be $2.7 \times 10^3\text{ J/m}^2$. Meanwhile, prepare PCR Amplification Master Mix (Step 55).

PCR amplification ● TIMING 45 min + 45 min for AMPure bead cleanup

▲ **CRITICAL** Hoechst–UV treatment and PCR amplification must be done sequentially.

55| Thaw PCR reagents and prepare Master Mix using the below recipe (Master Mix can be prepared during the Hoechst–UV treatment; Step 54). Mix thoroughly, pulse-spin and immediately proceed to the next step.

PCR Amplification Master Mix			
Reagent	Per cell (μl)	12 Samples (μl)	96 Samples (μl)
Phusion HF Master Mix	12.5	200	1,300
PE 1.0 primer	1	16	104
Final volume	13.5	216	1,404

56| Thaw multiplexing primer plate (see REAGENT SETUP) containing indexed Primer PE 2.0 primers, and pulse-spin. Dispense 1 μl of a single PE 2.0 primer to each sample (all of which were subjected to UV irradiation in Step 54). Mix the samples very well.

! **CAUTION** The multiplexing primer plate is the only aliquotted reagent that is reused between experiments. Therefore, be extremely careful to avoid cross-contamination, and use it a maximum of five times to avoid too many freeze–thaw cycles.

▲ **CRITICAL STEP** Ensure that each sample that will be pooled for sequencing receives a different indexed primer containing a unique hexamer barcode.

57| Add 13.5 μl of PCR Master Mix (prepared in Step 55) to each sample (from Step 56). Mix well, but avoid introducing bubbles. The final reaction volume is 25 μl.

58| Cover the plate with Microseal ‘C’ Optical sealing film and run the PCR program for 17 rounds of amplification, as follows:

Cycle number	Denature	Anneal	Extend	Hold
1	98 °C, 30 s			
2–18	98 °C, 10 s	65 °C, 30 s	72 °C, 30 s	
19			72 °C, 5 min	
20				4 °C

■ **PAUSE POINT** The sample plate can be spun down and frozen at –20 °C overnight before AMPure bead cleanup.

59| Proceed to AMPure bead cleanup (see Steps 19–32) to enrich the final product, which consists of completed directional single-cell libraries. Transfer completed libraries from the 384-well plate to a 96-well PCR plate.

! **CAUTION** The magnetic beads will tend to cluster in a tighter ‘pellet’ in the PCR reaction buffer than when cleaning up after adaptor ligation. Therefore, check that the pellet is completely dry after the 70% (vol/vol) EtOH wash by observing a loss of sheen on the magnet pellet, and observing visible cracks in the magnetic bead pellet, or the tendency of the pellet to curl away from the side of the tube.

▲ **CRITICAL STEP** If possible, perform all subsequent steps, including the final AMPure bead cleanup, in a post-PCR area with dedicated post-PCR pipettes and post-PCR filter tips.

▲ **CRITICAL STEP** Do not increase the volume of beads in this AMPure cleanup step. The 0.8:1 ratio (bead/sample) helps minimize capture of adaptor and primer dimers to enrich for the larger nucleosomal fragments (**Supplementary Fig. 3**)

■ **PAUSE POINT** After cleanup, the sample plate can be spun down and frozen at –20 °C, for up to one week, until further processing.

Pooling and size selection ● TIMING 5 h

60| Thaw completed libraries (if frozen), and pulse-spin. Each sample should be in EB buffer in a 96-well PCR plate.

PROTOCOL

61 Pool all 96 individual libraries (or desired subsets) together in one LoBind microcentrifuge tube, mix and pulse-spin briefly.

▲ **CRITICAL STEP** Pool the entire volume of each sample, or preserve a fraction of each sample for future resequencing (see Experimental Design section).

▲ **CRITICAL STEP** Use only 1 μ l of each 10-cell control library for pooling.

62 Size selection can be performed using either direct loading (option A) or loading after DNA precipitation (option B) (Fig. 4). Both options work equally well; however, with option B the library products are easier to visualize on a gel.

(A) Direct loading

(i) Load library pool into gel wells (20 μ l/well) of two 2% E-Gel EX agarose gels with SYBR Safe Stain together with a 50-bp DNA ladder, and run the gels using program no. 7 for 10 min.

▲ **CRITICAL STEP** Leave at least one empty well between the first library pool well and the well containing DNA ladder. When loading two different library pools on the same gel, make sure to leave at least one empty well between the two pools to avoid cross-contamination.

(B) Loading after DNA precipitation

(i) Add 3.81 μ l (1/100 volume) of GenElute-LPA to the library pool of 381 μ l total. Add 38.1 μ l (1/10 volume) of NaOAc, and mix it very well by pulse-vortexing followed by a pulse-spin. Add 952.5 μ l of 100% EtOH (2.5 \times volume) and vortex the sample for 10 s to mix. Incubate at -20 $^{\circ}$ C for at least 30 min, or longer (e.g. overnight). Prechill the microcentrifuge to 4 $^{\circ}$ C before this incubation step.

▲ **CRITICAL STEP** Volumes provided are for precipitating a full 96-well plate library pool, for which the final volume will be 381 μ l, accounting for 4 μ l per sample ($n = 95$ single-cell and 0-cell control samples), plus 1 μ l for a 10-cell control. If a smaller volume, or fewer libraries, is pooled, be sure to adjust volumes of LPA, NaOAc and 100% EtOH accordingly.

(ii) Centrifuge the tube at 21,000*g* (i.e., maximum speed) for 30 min at 4 $^{\circ}$ C, in a prechilled centrifuge.

(iii) Carefully remove the supernatant and wash the pellet with 200 μ l of freshly prepared 70% EtOH.

▲ **CRITICAL STEP** The DNA/LPA pellet is sometimes loose, so ensure that you do not lose it during aspiration of the supernatant.

(iv) Spin the tube at 21,000*g* (i.e. maximum speed) for 5 min at 4 $^{\circ}$ C.

(v) Remove the supernatant. Quick-spin for 10 s and remove the remaining EtOH.

(vi) Leave the tube open to air-dry. Once dry, add 40 μ l of EB and let the pellet rehydrate at RT for 5 min.

(vii) Dissolve the DNA by carefully pipetting up and down.

(viii) Repeat Step 60A(i).

63 While running the gel, place the microcentrifuge tube with EB in a heat block to preheat it to 55 $^{\circ}$ C. When the gel run is complete, take a picture of the gel with the Gel Doc EZ system using the Blue Sample Tray (or similar system) for records.

64 Carefully open the gel cassette with a gel knife cleaned with DNA Away Surface Decontaminant and place it on the Safe Imager transilluminator.

65 Excise the mononucleosome band using a disposable gel excision P1000 tip and eject the band into a prelabeled clean 1.5-ml LoBind microcentrifuge tube. To pool multiple lanes containing the same library pool, the same tip can be used to eject all slices into one tube, with a maximum of 10 lanes per tube. Take a picture of the excised gel for your records.

▲ **CRITICAL STEP** To capture the mononucleosomal product, excise the DNA band between 250 and 350 bp. Optionally, the dinucleosome product (~470 bp) can also be excised and further processed (see the Experimental Design section for details).

? TROUBLESHOOTING

66 Briefly spin down the tube(s) to collect all gel slices at the bottom. Add 60 μ l of buffer ADB (Zymo kit) per gel slice and incubate the mixture at 55 $^{\circ}$ C in a heat block for ~ 10 min. Mix during incubation with several gentle pulse-spins.

67 Briefly pulse-spin the tube and transfer melted gel slices to a prelabeled Zymo-spin column in a collection tube.

68 Spin at 21,000*g* (i.e., maximum speed) for 30 s at RT.

69 Wash the column once with 200 μ l of 80% EtOH, and spin it at 21,000*g* (i.e., maximum speed) for 1 min at RT.

70 Transfer the column to a clean LoBind 1.5-ml tube and leave the column to dry for 1 min.

71| Add 13 μl of warmed EB directly to the column matrix to elute the DNA. Let it stand for 1 min and then spin it at 21,000g (i.e., maximum speed) for 1 min at RT to collect the eluate.

■ **PAUSE POINT** The sample can be spun down and frozen at $-20\text{ }^{\circ}\text{C}$ for up to 1 week (longer storage periods may be possible but have not been formally tested).

72| Measure the final concentration of size-selected DNA in the library pool using a Qubit fluorometer according to the manufacturer's instructions. We recommend using 2 μl of the library pool for a sensitive reading.

73| Measure the final fragment length of size-selected DNA in the library pool using an Agilent Bioanalyzer. Run 1 μl of the library pool on a DNA high-sensitivity chip according to the manufacturer's instructions.

▲ **CRITICAL STEP** Avoid overloading the Bioanalyzer Chip, and run a maximum of 1 ng per lane in order to accurately determine molarity. If the library pool is too concentrated, dilute to 1 ng/ μl before running on the Bioanalyzer.

▲ **CRITICAL STEP** If the Bioanalyzer profile shows an incorrect size distribution of library fragments (e.g., an extra peak at 125 bp, corresponding to adaptor dimers), perform a second size selection step by running the library pool on a 1% E-Gel EX gel, and repeat Steps 63–73.

74| Calculate the molarity of the final library pool using the concentration ($\text{ng } \mu\text{l}^{-1}$) obtained with the Qubit measurement and the molarity obtained by region analysis on the Bioanalyzer.

▲ **CRITICAL STEP** To avoid overloading the flow cell used for sequencing, compare the concentrations obtained by Qubit and Bioanalyzer using region analysis. If these concentrations differ considerably, it is advisable to remeasure the sample. If the Bioanalyzer and Qubit concentrations are within range, we recommend calculating a correction factor ($[\text{Qubit}]/[\text{Bioanalyzer}]$). Taking into account that the Qubit is more accurate in measuring DNA concentrations, use this factor to correct the molarity of the library pool by multiplying the molarity obtained on the Bioanalyzer by the correction factor.

■ **PAUSE POINT** The sample can be spun down and frozen at $-20\text{ }^{\circ}\text{C}$ for up to 1 week (longer storage periods may be possible but have not been formally tested).

75| Dilute the library pool with EB to get a 2 nM solution. Measure the diluted library by Qubit and adjust the concentration to 2 nM, if necessary. Store at $-20\text{ }^{\circ}\text{C}$ until sequencing.

▲ **CRITICAL STEP** Prepare 2 nM dilutions shortly before sequencing. Store diluted library pool at $-20\text{ }^{\circ}\text{C}$ for a maximum of 1 week.

ILLUMINA SEQUENCING ● TIMING ~1 d–2 weeks

76| Generate single-end or paired-end sequencing reads from the libraries using an Illumina sequencing platform (as discussed in Experimental Design). We typically load 9–10 pM of a size-selected library pool per lane for a Rapid Run on a HiSeq 2500.

▲ **CRITICAL STEP** Include the custom index sequencing primers to sequence the hexamer barcode. For a HiSeq 2500 Rapid Run, we add 8.8 μl of 100 μM primer to 1.75 ml of Illumina i7 index primer solution in the tube at position no. 17 (final concentration of 0.5 μM).

? TROUBLESHOOTING

Bioinformatic analysis: generation of .bam files from an indexed reference assembly ● TIMING variable

▲ **CRITICAL** If the output files from the sequencer is aligned .bam files, it is not necessary to complete Steps 77–80. Analysis is typically performed on a Linux platform and requires the following freely available software: SAMtools, bedtools, BWA and BAIT. Alternatively, Strand-seq libraries in the form of aligned .bam files may be analyzed in a Linux, Windows or Macintosh environment using the R-based Bioconductor package contiBAIT.

77| Download the reference assembly for the organism corresponding to your cells of interest. Assemblies can be downloaded from the UCSC Genome Browser: <http://genome.ucsc.edu/cgi-bin/hgGateway> (the file with the '.fa.gz' extension from the FTP bigZips folder). They can also be downloaded from GenBank (<ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/>) for the most up-to-date, patched builds.

▲ **CRITICAL STEP** If downstream analysis (e.g., data visualization) will be performed on the UCSC Genome Browser, we recommend downloading the assembly from the UCSC website. Some formatting differences between assembly sources (such as '1' instead of 'chr1' or 'chrMT' instead of 'chrM') will cause an error when uploading data to this website.

78| Unzip the assembly. For .gz files, use the following command:

```
gunzip <genome.fa.gz>
```

where <genome.fa.gz> is the location of the zipped file.

Box 1 | Worked example of analysis pipeline ● TIMING variable

Procedure

The following assumes that either BAIT is installed on a Linux computer or that R, Bioconductor and contiBAIT are installed on a Windows/Macintosh/Linux computer.

1. Strand-seq data can be downloaded from the SRA for all published projects. To avoid the need to convert .sra files to .bam format, download the test data file, called testData.tar.bz, from the BAIT website: <https://sourceforge.net/projects/bait/files/?source=navbar>
2. The test data are contained in a GRCm37-aligned .bam file²⁰ zipped in a tarball. This can be unzipped in a Windows/Macintosh environment with a program such as WinZip, or in the Linux environment with the following command:

```
tar -xvzf testData.tar.bz
```

3. Navigate to the unzipped test/ directory. There will be a file named 'testData.AAAAAA.bam'.
4. To make a Strand-seq ideogram plot in a Windows, Macintosh or Linux environment, open an R terminal and run the following commands. The runtime should be ~4 s:

```
library(contiBAIT)
```

```
BAIT(path='.', splitBy=200000, pairedEnd=TRUE, chroms=c(1:21))
```

This creates an ideogram plot in which the data are binned into 200-kb bins (the splitBy term). 'PairedEnd' should be set to 'TRUE', as these data are paired end. The chroms variable allows selection of particular chromosomes. As GRCm37 contains the 19 autosomes and 2 allosomes, but also 45 scaffolds that have not been placed within the assembly, we take the first 21 elements from the .bam file header (chr1 through to chrY).

5. Alternatively, in a Linux environment, BAIT can be run directly from a terminal window. The runtime should be ~30 s:

```
BAIT -arv -7 Mus_musculus -8 mm9
```

79| After unzipping, index the assembly using the following command:

```
bwa index -a bwtsv -p <genome.fa> <genome.fa>
```

where <genome.fa> is the location of the assembly .fasta file.

80| To process Illumina reads in the .fasta format to the assembly, download the simple BASH script that (i) aligns the data to .sam format, (ii) compresses the data to .bam format and (iii) sorts the reads. This program is available at https://sourceforge.net/projects/whambam/files/wham_bam.sh/download. Typically, Strand-seq data involve a large number of single-cell libraries, and this program will align all files within a directory. There are a number of options that are selected, based on the data types. For example, the -p option is used if these data are single-end reads. By default, the program will search for a 6-character string of A, T, C and G in the file name and will assign this as the index. If 6-nucleotide indexes are not present in the filenames, a prompt will guide the user to select a suitable portion of the filename to distinguish the files. The program can be run with the following command:

```
bash filepath/wham_bam.sh -g <filepath/genome.fa> -v
```

where <filepath/genome.fa> is the location of the indexed assembly.

▲ CRITICAL STEP The alignment of several hundred .fastq files to a genome assembly can take >24 h of computer time, depending on system configuration and the number of reads. For a standard 2.0 GHz i7 processor with 168 GB of RAM, expect ~6 h per 96 samples. Converting files on a computer cluster will considerably speed up this alignment.

Bioinformatic analysis: generation of Strand-seq ideograms and finding of genomic features ● TIMING variable

81| As a first pass, there are currently three options for analyzing Strand-seq data: use of a Linux-based method using BAIT (option A), use of an R-based method using contiBAIT (option B) or design of your own custom scripts that address your specific experimental questions (option C).

▲ CRITICAL STEP It is important to note whether reads are paired-end vs single-end, as this affects the number of bits in the bitwise flag. There are options for analysis of single- vs. paired-end data in BAIT and contiBAIT.

? TROUBLESHOOTING

(A) Linux-based method using BAIT

- (i) The Linux-based method currently offers the most control to the end user. BAIT²² is a combination of BASH and R scripts that perform standard analysis of Strand-seq data. A working example of running BAIT, the custom bioinformatic pipeline used in-house to visualize Strand-seq data and select successful libraries, is provided in **Box 1**.

By default, this program will generate an ideogram for each library, automatically search for SCE locations and call changes in copy number. The program can be downloaded from <https://sourceforge.net/p/bait/wiki/Home/>, and includes an extensive wiki with tutorials. To carry out a standard run, navigate to the folder containing the Strand-seq .bam files and run the following command:

```
BAIT -t -r -a -v -s
```

This executes BAIT and creates a metric table (-t parameter), computationally identifies regions of template strand exchange such as SCEs (-r parameter), draws arrowheads on the ideograms at SCE locations (-a parameter), informs the user in a verbose fashion which functions are being run (-v parameter) and calculates CNV based on a circular binary segmentation algorithm (-s parameter).

▲ **CRITICAL STEP** BAIT can take several hours to run, depending on computer speed, the size/number of .bam files and the parameters used.

(B) R-based method using contiBAIT

- (i) A pure R-based implementation of BAIT, called contiBAIT, was recently released through Bioconductor. This method currently generates only ideograms, but it is more stable, platform-independent, simpler and faster than BAIT. In an R session, install contiBAIT using the following command:

```
source("https://bioconductor.org/biocLite.R")
biocLite("contiBAIT")
```

Then run the following command:

```
BAIT(path="location_of_bams", pairedEnd=TRUE,
plotName="BAIT_plot.pdf")
```

Where `location_of_bams` is the path to the Strand-seq bam files, and the `pairedEnd` parameter is a logical statement as to whether single- or paired-end sequencing was carried out. contiBAIT is substantially faster than BAIT, and is more memory efficient. It also works on the Windows and Mac operating systems. However, it does not identify SCEs or CNVs at this time.

(C) Custom bioinformatic analysis

- (i) All the strand directionality information used in Strand-seq analysis is stored in the second column of the .bam file, the bitwise flag. This column is a single integer that represents 12 binary parameters or 'bits'. Strand directionality is represented by the fifth bit. For paired-end data, the read number (the first read or the mate pair) is represented by the seventh and eighth bits. Using this information, Strand-seq data can be analyzed by custom scripts.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
7	Improper nuclei isolation	Cell lysis was too strong	Reduce the percentage of NP40 in the NSB or use an alternative nuclei isolation protocol. Suggestions are included in the Supplementary Methods
		Nuclei were in the NSB for too long before sorting	Shorten incubation time and prepare nuclei solution just before sorting
		Inefficient cell lysis	Increase the percentage of NP40 or use a stronger detergent in the nuclei isolation protocol. Suggestions are included in Supplementary Methods
10	Control sample does not have defined peak in the UV channel	Hoechst staining of nuclei was not optimized	Confirm that cell lysis was complete (described above). Ensure that UV laser is being used to excite Hoechst

(continued)

PROTOCOL

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
11	BrdU ⁺ sample does not show a shift in UV channel	Cells did not take up BrdU	Perform time course to test optimal duration of BrdU exposure. Titrate the amount of BrdU added to the sample, and determine optimal concentration for cells of interest
31	AMPure XP beads will not resuspend	Beads were dried for too long	Shorten the drying for the beads Rehydrate the beads for a longer time in EB to enable efficient resuspension, and mix by pipetting
65	Incorrect size distribution of DNA fragments	Inefficient MNase treatment, resulting in larger fragments than expected (e.g., dinucleosomes) Many adaptor dimers, low-yield DNA fragments	Replace MNase enzyme Many empty wells: Make sure that during the sort each well contains a nucleus Decrease the amount of adaptors used during the ligation step or use ligation buffer without PEG
76	Overloaded flow cell	Incorrect calculation of molarity	Take area under the curve of the peak, using region analysis to calculate molarity, instead of using peak analysis to calculate average peak size Compare the concentrations obtained by Qubit and Bioanalyzer using region analysis. If these concentrations differ considerably, it is advisable to remeasure the sample. If the concentrations are in the same range, we recommend calculating a correction factor ([Qubit]/[Bioanalyzer]). Taking into account that the Qubit is more accurate in measuring DNA concentration, use this factor to correct the molarity of the library pool by multiplying the molarity obtained with the Bioanalyzer by the correction factor.
81	Empty libraries	No nucleus was deposited into the well Sorted nucleus may not have been deposited into the sorting buffer, but instead ended up on the side of the well	To ensure efficient and accurate single-nuclei sorting, sort single fluorescent beads into a 96-well plate and check the presence of these beads at the bottom of each well using a fluorescence microscope Check sort alignment by sorting drops onto a sealed 96-well plate, and examine the position of the drops. The sorted drops should be centered precisely in the middle of the wells
	High percentage of background	Not enough BrdU was incorporated Incorrect nuclei were sorted	When many cells are present in the culture, it is possible that the total amount of BrdU incorporated per cell is not high enough to ensure that the nascent strand is completely removed during Hoechst/UV treatment. Increase the BrdU concentration during cell culture Incomplete BrdU incorporation occurred because cells were already halfway the S-phase when BrdU was added to cell culture
	Some chromosomes show no reads	Cell was already in second cell division during BrdU incorporation	Shorten BrdU incubation time
	All chromosomes contain WC reads (e.g., whole-genome sequencing library)	Cells have not divided	Check cell cycle timing and ensure proper gate settings during sort

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
81		More than one nucleus is present	Ensure during sort that only one nucleus per well is sorted
	Programs crash or produce unusual results (e.g., only Crick reads present on ideograms)	Single-end data are treated as paired-end or vice versa	Ensure that the correct options are selected with the programs Do not mix single- and paired-end data sets
		Files with different headers are used	Ensure that all .bam files are aligned to the same assembly. Different assembly versions (e.g., GRCh37 and GRCh38) will cause errors, as will different versions of the same assembly (e.g., GRCh38 with chromosomes in the form 'chr1', 'chr2', 'chr3' and so on versus GRCh38 with chromosomes in the form '1', '2', '3', and so on)

● TIMING

Steps 1–11, isolation of single cells: 3 h

Steps 12–18, MNase digestion: 30 min

Steps 19–32, AMPure bead cleanup: 45 min

Steps 33–37, end repair: 45 min (+ 45 min for AMPure bead cleanup)

Steps 38–42, A-tailing: 45 min (+ 45 min for AMPure bead cleanup)

Steps 43–48, adaptor ligation: 30 min (+ 45 min for AMPure bead cleanup)

Steps 49–54, Hoechst–UV treatment: 40 min

Steps 55–59, PCR amplification: 45 min (+ 45 min for AMPure bead cleanup)

Steps 60–75, pooling and size selection, 5 h

Step 76, sequencing: 1 d (MiSeq or HiSeq rapid run)–2 weeks (high-output HiSeq run)

Steps 77–81, data analysis: 1+ days (depending on computing power, library depth and number of libraries)

Box 1, worked example of analysis pipeline: variable (depending on computing power, library depth and number of libraries): A complete set of 96 Strand-seq libraries can be generated in a total of 4 d. Typically, Steps 1–37, corresponding to single-cell isolation, MNase digestion and end repair (and including the two AMPure bead cleanups required after each enzymatic step), can be completed on day 1. A-tailing through to PCR amplification (with two AMPure bead cleanups, corresponding to Steps 38–59) can be completed on day 2. The final post-PCR AMPure bead cleanup, pooling and size selection (Steps 60–75) can be completed on day 3. The pooled libraries can be sequenced on a HiSeq Rapid Run or MiSeq instrument on day 4. The time required for data analysis depends on the computing system configuration and research questions; however, sequencing reads can be demultiplexed, aligned to the reference assembly and analyzed with BAIT within 1 d.

ANTICIPATED RESULTS

Single Strand-seq libraries initially require demultiplexing from the single lane of sequencing through index retrieval. It is recommended that this be done directly on the Illumina platform within the sample sheet. Each indexed Strand-seq library should then be processed by aligning it to a genome assembly and converting it into .bam format. The indexes used in the Strand-seq protocol have a minimum Hamming distance of 2 (mean Hamming distance of 4.5; on average each index will have a distance of 2 with 13 other indexes). Because of this, it is not recommended that any fault-tolerant criterion be used when demultiplexing, as single mismatches within the 6-nucleotide sequence are likely to ambiguously match with more than one index.

The quality of Strand-seq libraries is assessed in a number of ways. Successful Strand-seq libraries are suitably deep, with a minimum of 20 reads/Mb, and have a mapping quality of >q10 (**Fig. 5a**). Libraries with fewer reads are typically not included in the analysis, as the accuracy of calling strand-state breakpoints and CNV diminishes. Although this is a minimum, ideal Strand-seq libraries have a read depth greater than 100 reads/Mb. Furthermore, it is important to distinguish Strand-seq libraries from whole-genome libraries; cells that failed to incorporate BrdU are occasionally erroneously sorted and show no informative strand-state information (reads map to both Watson and Crick strands equally, so every chromosome appears WC). These must also be removed from the analysis.

Strand-seq library metrics

Strand-seq libraries should also ideally be of low background (**Fig. 5b**). The level of background within libraries is difficult to quantify, and whether to include a particular library with high background is dependent on the type of analysis being

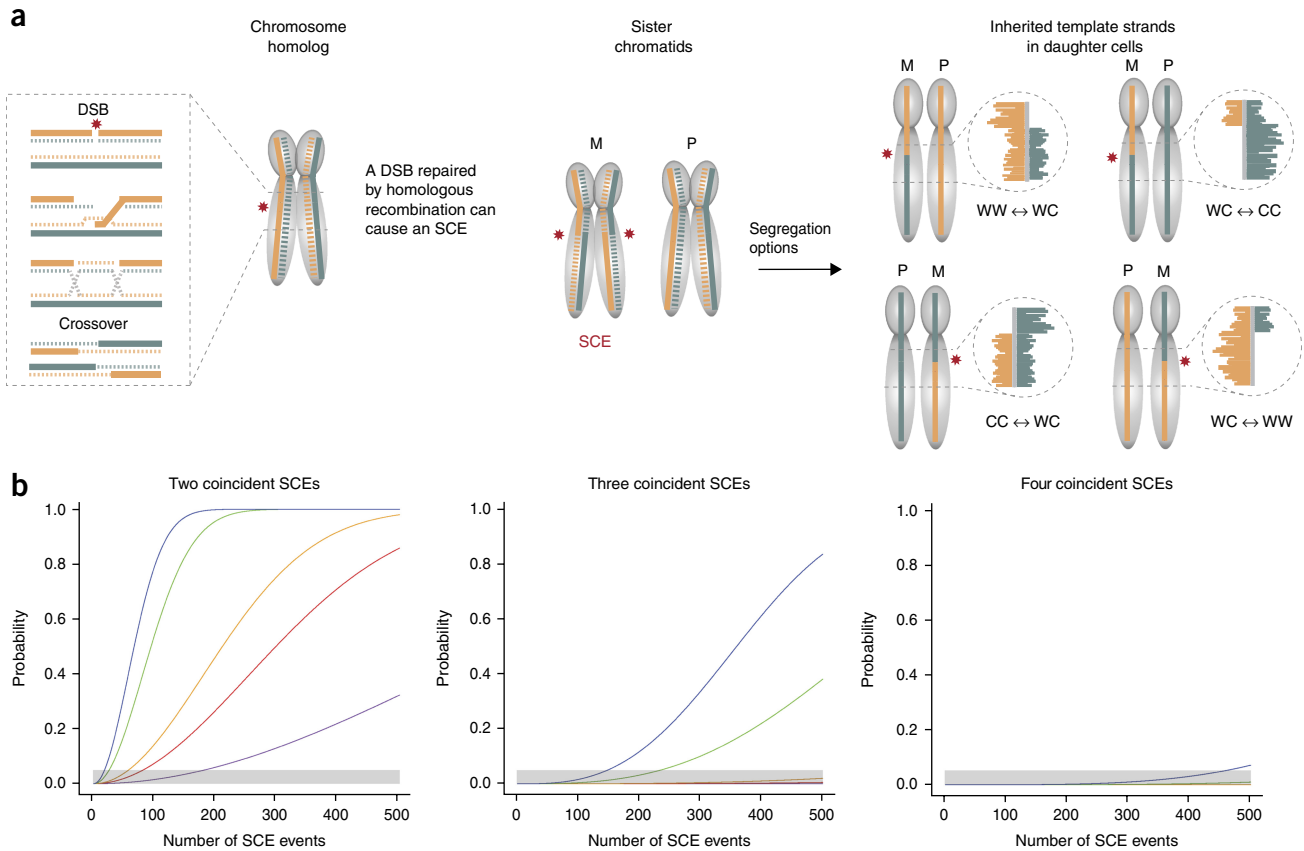


Figure 7 | Probabilities for template strand changes to occur in the same vicinity between libraries. **(a)** Sister chromatid exchanges (SCEs) are somatic recombination events that occur when there is a double-stranded DNA break (DSB; red asterisks) in a chromosomal homolog that is repaired by a homologous recombination (boxed region). The repair can result in a non-crossover or crossover event, and only the latter is shown for simplicity. In Strand-seq libraries (right panel), SCEs appear as a change in template strand state along the affected chromosome (e.g., a chromosome switching between WW and WC states, or WC and CC states). Watson (W) is shown in orange, and Crick (C) is shown in blue. Nascent BrdU⁺ strands are shown as dotted lines. **(b)** As the number of SCE calls increases for an experiment, so too does the probability that two of those calls will be close together. The plots show the probability of two or more coincident SCE calls within 10 (purple), 50 (red), 100 (orange), 500 (green) and 1,000 (blue) kilobases of each other. Note that the probability of three or more SCE events occurring in the same location by chance is far smaller. Gray bars represent significance. Probabilities were calculated based on a variant of the birthday problem.

undertaken. There are cases in which insufficient BrdU leads to a consistent low number of reads mapping to both strands. In these cases, strand information is not lost, but the ability to accurately map locations of strand-state changes is diminished because of the pervasive underlying background. Furthermore, some libraries exhibit punctate regions of high background, which can confuse the bioinformatic algorithms used to identify changes in strand state and lead to false-positive calls. Finally, in unsynchronized cell populations, some cells may incorporate BrdU during more than one DNA replication, manifesting as regions (or whole chromosomes) where no reads map.

BAIT includes a function that generates a basic metric table that outlines the number of high-quality reads that map, and the percentages of duplicates and adaptor dimers. Although the latter two metrics do not necessarily have bearing on the quality of the libraries, they do highlight library complexity and give an indication as to whether deeper sequencing would be appropriate. The BAIT program flags all instances of libraries with less than 20 reads/Mb and does not attempt to run strand-state changes or CNV functions on them. BAIT also attempts to attribute a level of background to each individual library.

Identification of strand inheritance patterns

For experiments designed to test the immortal strand or silent sister hypotheses³³, a representation of the strand inheritance pattern across single cells of a particular type is needed. BAIT will produce a data table highlighting the strand state for each chromosome for each library as WW, WC, or CC, or SCE, for chromosomes that contain SCEs. In the absence of any selective retention of template strands, each chromosome should be WC in ~50% of libraries, WW in ~25% of libraries and CC in ~25% of libraries. It is possible that particular regions rather than whole chromosomes show template strand biases between cell types, and so BAIT also calculates the template strand state in 200-kb intervals across the entire length of each chromosome.

Identification of sister chromatid exchange events

SCEs manifest as changes in template strand states within Strand-seq libraries. Essentially these events occur when the template strands of one chromosome homolog recombine in order to repair DNA damage (Fig. 7a). As the other template strand should remain intact, we anticipate that these changes appear as a transition from a WC state to either a CC or WW state. For libraries that are of high read depth and low background, BAIT will accurately call these changes in template strand states, and map the breakpoints with high accuracy²². Alternatively, .bed files of the data (created using the -b option in BAIT) can be uploaded to the UCSC Genome Browser and breakpoints can be mapped manually by locating the region where the strand state changes. On the basis of cultured cell lines we have studied thus far, between 5 and 8 SCEs are anticipated per cell. This number appears to be reduced to 2–5 SCEs when dealing with primary samples or cells derived from *in vivo* BrdU-injected animal models. SCEs are the end result of repair by the homologous recombination machinery, and are used as a chromosomal stability index in response to environmental and mutagenic stimuli. Therefore, we anticipate that the number (and potentially location) of such events will change depending on the toxicity of the stimuli. Assuming a random distribution of SCEs, the likelihood of two SCEs occurring at the same location in two different cells can be calculated using the birthday problem. The probability of two events occurring in close proximity by chance is dependent on the degree of proximity, the overall number of SCE calls and the number of coincident SCE calls (Fig. 7b). Using this calculation, investigations into the special distribution of SCEs can be undertaken to search for genomic regions prone to recombine.

A template-strand-state transition from WW to CC is unlikely to be the result of an SCE, as it would require both chromosomal homologs to have an exchange event between sister chromatids at the same location. Despite the unexpected nature of this type of strand-state change, this phenomenon has been seen in our data^{20,21}. In cases in which such events were discovered, they were present across multiple single cells and were confirmed as misorientations of the assembly to which the data were aligned²⁰. Identification of events occurring at the same location across multiple libraries will be discussed in the next section.

Identification of inversions/misorientations

For outbred populations such as humans, the number of template-strand-state changes called is often higher because the assembly to which the data are aligned does not account for structural rearrangements such as inversions that are polymorphic within the populations²¹. Similar to misorientations, germline (i.e., inherited) genomic rearrangements will map to the same locations in every informative Strand-seq library that is generated from the same donor cell sample. Both BAIT and the R-based package 'Invert.R'²¹ (available at <https://sourceforge.net/projects/strandseq-invertr/>) determine the proximity of template-strand state changes between libraries to assess whether certain regions are putative sites of structural variation. Although BAIT provides a general overview and is better suited to SCE analyses, the Invert.R package is specifically designed to look for polymorphic structural changes and is recommended for inversion analyses. Inversions can be either homozygous or heterozygous within a particular individual, and Invert.R includes a function to genotype these regions. By analyzing multiple single cells from the same individual, a complete inversion profile can be created for the experiment²¹.

Phasing variant positions on whole chromosomes

When chromosomes are inherited as WC, each template strand represents a single parental homolog. Consequently, all variants (structural variants and single-nucleotide polymorphisms) that map to the Watson strand are present on one homolog (e.g., maternal), and all the variants on the Crick strand are present on the other homolog (e.g., paternal). This means that the variants are phased for the entire length of the chromosome, albeit at very low density for a single cell and only for the WC chromosomes. To build more dense haplotypes that represent the entire genome of an individual, multiple cells can be analyzed. We recently showed how this can be used to build complete and accurate haplotypes without the need for generational (i.e., trio) information²³. The R-based bioinformatic package, called 'Strand-Phase.R' (<https://github.com/daewoooo/StrandPhaseR.git>), is available for haplotyping Strand-seq data. Although it is possible that the UV treatment inherent in the Strand-seq protocol may yield premutagenic lesions that are erroneously called as variants, to limit this (and sequencing errors), Strand-Phase.R makes high-quality calls based on variants present in more than one cell/library.

A schematic summary of the genomic features seen in a Strand-seq experiment is shown in Figure 6.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This work was supported by grants from the Canadian Institutes of Health Research (RMF-92093 and 105265), the US National Institutes of Health (R01GM094146) and the Terry Fox Foundation (018006), as well as an Advanced Grant (Nr 294740) from the European Research Council to P.M.L., and a Vanier Graduate Scholarship from the Natural Science and Engineering Research Council to A.D.S.

AUTHOR CONTRIBUTIONS E.F. and A.D.S. developed and optimized the protocol. M.H. and A.D.S. designed bioinformatic analysis tools. D.C.J.S. established the protocol in a sister laboratory. A.D.S., M.H., D.C.J.S. and P.M.L. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Mills, R.E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
2. Sudmant, P.H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
3. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
4. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
5. Alkan, C., Coe, B.P. & Eichler, E.E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).
6. Feuk, L. Inversion variants in the human genome: role in disease and genome architecture. *Genome Med.* **2**, 11 (2010).
7. Sindi, S.S. & Raphael, B.J. Identification of structural variation. in *Genome Analysis: Current Procedures and Applications* (ed. Poptsova, M.S.), (Caister Academic Press, 2014).
8. Korbel, J.O. *et al.* PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol.* **10**, R23 (2009).
9. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
10. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
11. Xie, C. & Tammi, M.T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10**, 80 (2009).
12. Voet, T. *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res.* **41**, 6119–6138 (2013).
13. Nagano, T. *et al.* Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* **10**, 1986–2003 (2015).
14. Leung, M.L. *et al.* Highly multiplexed targeted DNA sequencing from single nuclei. *Nat. Protoc.* **11**, 214–235 (2016).
15. Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol.* **17**, 115 (2016).
16. van den Bos, H. *et al.* Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol.* **17**, 116 (2016).
17. Navin, N.E. The first five years of single-cell cancer genomics and beyond. *Genome Res.* **25**, 1499–1507 (2015).
18. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
19. Macaulay, I.C. & Voet, T. Single cell genomics: advances and future perspectives. *PLoS Genet.* **10**, e1004126 (2014).
20. Falconer, E. *et al.* DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nat. Methods* **9**, 1107–1112 (2012).
21. Sanders, A.D. *et al.* Characterizing polymorphic inversions in human genomes by single cell sequencing. *Genome Res.* **26**, 1575–1587 (2016).
22. Hills, M., O'Neill, K., Falconer, E., Brinkman, R. & Lansdorp, P.M. BAIT: organizing genomes and mapping rearrangements in single cells. *Genome Med.* **5**, 82 (2013).
23. Porubsky, D. *et al.* Building complete chromosomal haplotypes using single-cell sequencing. *Genome Res.* **26**, 1565–1574 (2016).
24. Falconer, E., Chavez, E., Henderson, A. & Lansdorp, P.M. Chromosome orientation fluorescence *in situ* hybridization to study sister chromatid segregation *in vivo*. *Nat. Protoc.* **5**, 1362–1377 (2010).
25. Smith, J.R. & Pereira-Smith, O.M. Replicative senescence: implications for *in vivo* aging and tumor suppression. *Science* **273**, 63–67 (1996).
26. Rabinovitch, P.S. Regulation of human fibroblast growth rate by both noncycling cell fraction transition probability is shown by growth in 5-bromodeoxyuridine followed by Hoechst 33258 flow cytometry. *Proc. Natl. Acad. Sci. USA* **80**, 2951–2955 (1983).
27. Poot, M., Silber, J.R. & Rabinovitch, P.S. A novel flow cytometric technique for drug cytotoxicity gives results comparable to colony-forming assays. *Cytometry* **48**, 1–5 (2002).
28. Latt, S.A., George, Y.S. & Gray, J.W. Flow cytometric analysis of bromodeoxyuridine-substituted cells stained with 33258 Hoechst. *J. Histochem. Cytochem.* **25**, 927–934 (1977).
29. Kubbies, M. & Rabinovitch, P.S. Flow cytometric analysis of factors which influence the BrdUrd-Hoechst quenching effect in cultivated human fibroblasts and lymphocytes. *Cytometry* **3**, 276–281 (1983).
30. Fogt, F. *et al.* Flow cytometric measurement of cell cycle kinetics in rat Walker-256 carcinoma following *in vivo* and *in vitro* pulse labelling with bromodeoxyuridine. *Cytometry* **12**, 33–41 (1991).
31. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
32. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
33. Lansdorp, P.M., Falconer, E., Tao, J., Brind'Amour, J. & Naumann, U. Epigenetic differences between sister chromatids? *Ann. N. Y. Acad. Sci.* **1266**, 1–6 (2012).