Genome-wide profiling of DNA repair identifies higher-order coordination in single cells

Kim L. de Luca[1,2], Pim M. J. Rullens[1,2,3], Gaëlle Legube[4], Jop Kind[1,2,3]*

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) & University Medical Center Utrecht; Utrecht, the Netherlands.

[2]Oncode Institute; the Netherlands.

[3]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, the Netherlands.

[4]MCD, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse; Toulouse, France.

*Corresponding author. Email: j.kind@hubrecht.eu

Abstract

Accurate repair of DNA damage is critical for maintenance of genomic integrity and cellular survival. Because damage occurs non-uniformly across the genome, single-cell resolution is required for proper interrogation, but sensitive detection has remained challenging. Here, we present genome-wide binding profiles of DNA double-strand break repair proteins in single cells, allowing for the study of heterogeneity in genomic damage locations and associated repair features. By unbiasedly detecting repair-enriched segments, we find that repair proteins often occupy entire topologically associating domains and mimic variability in chromatin loop anchoring. Genomic loci that form damage-specific 3D contacts show multi-way repair coordination in individual cells that becomes stronger according to the number of interaction partners. These findings advance our understanding of genome stability in the context of nuclear organization.

Introduction

The eukaryotic nucleus is constantly exposed to endogenous and exogenous sources of damage to the genome. Double-stranded breaks (DSBs) in the DNA are exceptionally toxic lesions because they completely sever the DNA fiber, leaving the genome at risk of small nucleotide changes and larger structural aberrations such as translocations and deletions. These types of genomic instability are associated with tumorigenesis as well as aging-related diseases. In addition to the molecular pathways involved in the DNA damage response (DDR), repair is dependent on the chromatin environment of a DSB and its location within the highly compartmentalized nucleus (*1, 2*).

Because the occurrence and repair of DNA damage are variable across cells, information obtained from individual cells is required to accurately profile their distribution. A recent study reported single-cell detection of spontaneous DNA damage by using whole-genome amplification (*3*), but insight into DNA repair remains elusive. Here, we present single-cell profiles of DNA repair proteins involved in the main DSB repair pathways: non-homologous end joining (NHEJ) and homology-directed repair (HDR). Using various sequencing-based approaches, we validate that the obtained signals represent *bona fide* repair events, and perform comparisons with the state of the art, showing superior sensitivity of our approach. Our data reveal that DSB loci processed by HDR proteins are repaired in a cell cycle-dependent manner that correlates well with replication timing. Further, we investigate repaired regions in the context of genome organization. Even at the resolution of individual cells, repair proteins spread along the genome and inside the nucleus according to underlying genome topology, following the three-dimensional structure that exists prior to damage. We detect heterogeneity in the modes of spreading that display features of loop extrusion and may therefore reflect interaction domains. In addition, newly formed, damage-specific, genomic contacts exhibit multi-way coordination of repair activity, shedding new light on DSB clustering events in single cells.

Detecting double-strand break repair proteins genome-wide with DamID and ChIC

We established an experimental and computational workflow to unbiasedly identify DSB repair profiles in single cells (Fig. 1A). Using DamID (*4-7*) and ChIC (*8-10*), we measured genomic contacts of proteins involved in NHEJ and HDR, distinguishing between pathways by choosing different Dam-fusion proteins (for DamID) or antibodies (for ChIC). After filtering on quality criteria (see Methods), we obtained a collective total of ~12,000 single-cell profiles of DSB repair proteins 53BP1, MDC1, and γH2AX, and HDR-specific proteins RAD51 and BRCA1 (Fig. 1B, S1A-B). We used the previously established human DIvA cell line, which generates DSBs at sequence-specific positions in the genome by the endonuclease AsiSI-ER under control of 4-hydroxytamoxifen (4OHT) (*11, 12*). To determine significant enrichments of repair occupancy on the genome, we developed a computational method for single-cell domain calling. We modified a multi-scale representation of genomic signals (MSR) approach (*13*) that unbiasedly identifies enrichments of variable sizes (Fig. S2, Methods). Upon damage induction, specific repair signal accumulates at DSB sites (Fig. 1B, S1C-E), which compares well with publicly available population-based ChIP-seq datasets (Fig. S1E-G). Measured across all cells and DSB sites, we calculated the proportion of cells in which each DSB was repaired by the different target proteins (referred to as "repair frequency"). Our collective dataset indicates superior sensitivity compared to previous reports (Fig. 1C), allowing for the study of a broader range of infrequently captured DSB sites. In accordance with the DDR signaling cascade, nearly all DSB sites are most frequently occupied by mediator of DNA damage checkpoint protein 1 (MDC1) and ubiquitous DSB marker γH2AX (the direct chromatin substrate for MDC1 binding), followed by NHEJ-promoting 53BP1, and HDR-specific RAD51 and BRCA1 (Fig. 1C). Nonetheless, repair signatures clearly vary across sites, distinguishing between NHEJ and HDR preference. To further quantify the repair domains per cell, we defined a set of single-cell profiles in which we could most accurately determine the number of segments (see Methods). At the median sequencing depth (3×10^3 UMIs), DSB-treated and control cells respectively contained 28 and 3 repair-enriched segments, reaching a plateau at 50 and 7 segments (Fig. S1H), in agreement with the number of AsiSI-induced 53BP1 foci detected by imaging (*14*).

DSB repair has been widely studied with imaging-based methods that detect proteins, while DamID relies on detection of genomic DNA that has been marked by the Dam-fusion protein. We previously developed the ^{m6}A-Tracer system for visualization and tracking of Dam-methylated contacts (*15*), and apply it here to validate that HDR-specific Dam-RAD51 generates foci that visually overlap with γH2AX, a universal DSB marker (Fig. 1D). Of the DamID RAD51 signal, ~65% colocalizes with γH2AX; vice versa, a smaller ~30% of γH2AX signal corresponds to Dam-RAD51 (Fig. 1E). The latter is anticipated to be lower since not all γH2AX-marked DSBs are repaired by RAD51, as can also be observed from the smaller number of Dam-RAD51 foci.

Together, these results show that implementation of DamID and ChIC can be used for sensitive and specific single-cell genomics as well as quantitative imaging analyses of DSB repair.

Homology-directed repair mediated by RAD51 correlates with replication timing and transcriptional activity

Repair pathway choice between NHEJ and HDR is highly regulated at multiple levels, including nuclear structure, global spatial genome organization, local chromatin context, and sequence specificity (*16*). In addition to such regulatory processes, the cell cycle state of individual cells has been linked to heterogeneity in repair pathway usage (*17, 18*). Our DamID experimental setup includes recording of live-cell DNA content during fluorescence-activated cell sorting (FACS), thereby establishing a procedure to address DSB repair protein occupancy in relation to the cell cycle at unprecedented resolution (Fig. 2A). We observed that some DSB sites exhibit differences in repair frequency according to cell cycle phase, particularly when undergoing HDR (Fig. S3A). To further explore this relationship, we ordered all Dam-RAD51 cells on their cell cycle stage, and noticed differences in repair enrichment at DSB sites during S phase (Fig. 2B-C), a genome-wide trend that is not present for NHEJ protein 53BP1 (Fig. S3B). Because HDR requires a sister chromatid as its repair template, we annotated the genomic regions by their replication timing (RT) using publicly available Repli-Seq data (*19, 20*). This showed considerable

concordance between RT of a site and the relative frequency with which it is repaired by Dam-RAD51 along S and G2 (Fig. 2D), in agreement with imaging data suggesting that active replication influences HDR employment (*17*).

RT is significantly correlated with chromatin state and genome organization (*21, 22*); moreover, RT-driven remodeling of the epigenome has been linked to cancer-specific chromosomal rearrangements (*23*). Several studies have further highlighted the role of chromatin in DNA repair (*24-33*); notably, RAD51 is preferentially recruited to transcribed loci enriched in active histone modification H3K36me3 (*24*). We sought to directly measure the relationship between DNA repair and chromatin, by jointly profiling RAD51 occupancy (with DamID) and either H3K36me3 or repressive mark H3K9me3 (with ChIC) in the same cell (see Methods and (*34*)). Our combined single-cell Dam&ChIC repair profiles were of similar quality to DamID as the only genomic readout (Fig. S3C-D), and ChIC histone modification signal was in excellent concordance with published ChIP-seq (Fig. S3E-F). We show that H3K36me3 enrichment is significantly higher in cells that have RAD51 occupancy than in cells without, for a given AsiSI site (Fig. 2E) and across all sites (Fig. S3G), while H3K9me3 is expectedly low and similar between repaired states (Fig. 2E, S3G). This confirms a mechanistic link between homology-directed repair and transcriptionally active chromatin at the single-cell level.

Repair protein occupancy in single cells follows pre-existing genome topology

Next, we set out to address chromatin conformation of DSB repair sites at single-cell resolution. Recent population-based work has reported local chromatin reorganization upon DNA damage (*35, 36*), and proposed cohesin-mediated loop extrusion as the mechanism by which DNA repair domains are established (*35*). This further substantiates the thought that chromatin is compacted during the DDR, forming incontiguous domains according to topological features (*37-41*).

We projected our DamID 53BP1-enriched segments onto Hi-C maps to examine the interplay between genome topology and DSB repair in single cells (Fig. S4, 3A). Repair domain enrichment occurs in different scenarios, varying mostly according to extent, directionality, and “modality” of spreading. While domains of repair occupancy seem generally constrained by borders of chromosomal domains, modality is interpreted as an indication of the spectrum of observed domain sizes, reflecting cell-to-cell variability (Fig. S4, 3A). In highly insulated chromatin, 53BP1 segments clearly overlap with the local genome structure, while DSB sites devoid of detectable structure show repair occupancy in a scattered fashion, without topological constraints. Notably, in an example of multimodal enrichment, 53BP1 spreading is halted at a moderate insulation border in only a fraction of cells (Fig. 3A-B blue line), whereas the majority of repair segments extend to the much stronger flanking borders (orange lines).

We wondered whether this apparent relationship between repair protein spreading in single cells and border strength of topologically associating domains (TADs) could be generalized to all DSB sites. Indeed, genome-wide, there is a strong dip in the insulation score at repair segment boundaries, indicating overlap with TAD borders (Fig. 3C). Repair segments also more frequently end at stronger TAD borders (i.e., with lower insulation scores) (Fig. 3D). We note that these measurements of repair protein occupancy are a direct result of our newly implemented multi-scale domain calling approach and its ability to detect enrichment without any *a priori* information of segment size. Our observation of heterogeneity within the distributions of repair domains thereby highlights the importance of profiling repair events at single-cell resolution.

Coordinated repair protein signal at loop extrusion borders

In the comparison with Hi-C above, we defined repair protein occupancy as the enrichment from start to end of the MSR segments. To further strengthen the notion that such occupancy occurs in 3D, we compared repair spreading at a given DSB site to the chromosomal conformation of that locus (“virtual 4C” from Hi-C data). The quantitative repair counts—i.e., sequencing read output, not domain calls—captured by both DamID and ChIC strongly resemble topological contacts along the linear genome (Fig. 3E). Hence, the quantitative *in silico*

population signal, single-cell profiles, and binarized MSR segments all show that DNA repair occurs in the context of genome organization.

Importantly, single-cell Hi-C and super-resolution chromatin tracing methods have indicated that boundaries of TAD-like domains vary across cells, despite preferential anchoring at population-based boundaries (*42-45*). We reasoned that, if repair proteins spread according to these topology-driven rules of variability and boundary anchoring, loci with stronger Hi-C contacts should more frequently show coordinated occupancy of repair proteins across single cells. Conversely, repair signal should be independent for loci with weak (or few) Hi-C contacts. Coordination is thus a measure of how frequently a repair protein occupies a DSB locus, measured across all cells: it should reflect the diversity of topological configurations commonly observable in a given Hi-C window. Indeed, repair protein coordination corresponds to features in Hi-C data that are related to cohesin-mediated loop extrusion, suggesting that the repair machinery spreads within an ample variety of preferentially anchored single-cell loops (Fig. 3F).

Altogether, we interpret these results to mean that the spreading of repair proteins on the genome follows underlying topology, explaining differences in repaired genome segment size as well as quantitative repair signal across DSB loci.

Multi-way coordination of DNA repair activity at interchromosomal contacts

Besides cohesin-mediated reorganization of the genome in response to damage, microscopy- and sequencing-based approaches have shown that DSBs exhibit “clustering”, i.e., intra-nuclear motion and coalescence of genomic loci on different chromosomes (*14, 46-49*). DSBs can further trigger genome-wide compartment switching (*36*), with recent findings suggesting the formation of a novel, damage-specific sub-compartment (*50*).

Long-range coordination of single-cell repair activity corresponds to Hi-C contact frequency

Our analysis of recent Hi-C data (*35*) confirms that some DSB sites form contacts upon damage (Fig. 4A, Fig. S5A), predominantly inter-chromosomally (Fig. S5B). We defined pairs of sites as “contacting” by setting a threshold on the differential Hi-C matrix, and validate enriched interactions up to +/- 0.5 Mb surrounding the DSB sites (Fig. S5C top). The single-cell nature of our data uniquely positions us to examine coordinated repair protein binding events within the same cell. We reasoned that, in order for DSB pairs to physically interact upon DNA damage in a given cell, both sites should be occupied by the repair machinery. Indeed, we find that coordination of 53BP1 repair events is considerably higher for contacting DSB pairs (Fig. S5C bottom, S5D), to a degree that repair coordination at individual break sites strongly overlaps the enriched Hi-C damage contacts (Fig. 4B). In light of the relationship between HDR and RT (replication timing, Figure 2), we categorized the DSB pairs by their replication score. Repair coordination of RAD51 and BRCA1 was significantly higher for the matching categories (early / early; late / late) than the early / late pairs.

Higher-order repair coordination in single cells increases with hub size

In the analyses described above, both the Hi-C contacts and repair coordination measurements describe pairwise events, and are thus limited to two DSB sites at a time. However, population-based Hi-C data indicates that many DSBs physically interact with more than one other DSB, and some do so with >15 other sites (Fig. S5F). Consequently, it is currently an open question whether contacting sites form higher-order, or multi-way, DSB repair “hubs” within the same cell, or mutually exclusive interactions that occur independently across cells (Fig. 4Di). Although the scDamID and scChIC repair protein profiles do not measure spatial localization of DSBs, they offer a distinct ability to analyze (unlimited) multi-way coordination of repair activity as a proxy for repair hub formation.

To systematically evaluate higher-order repair coordination, we selected DSB sites that form 3 or more proximal contacts (Fig. S5F-G, 4C). Hierarchical clustering identified subsets of sites that frequently interact (Fig. S5G, 4C), which are remarkably mirrored by pairwise repair coordination across single cells (Fig. 4C). Based on

the DSB Hi-C matrix, we identify a few hundred hubs of DSB sites that have pairwise interactions between all participating loci. These interaction hubs vary in size (i.e., number of contacts), with some extremes consisting of up to 6 DSBs (Fig. S5H). To directly test higher-order coordination in those hubs, we applied a multiple-site similarity measure that is based on the pairwise Sørensen-Dice coefficient (see Methods and (*51, 52*)). Contacting hubs showed considerably more multi-way repair coordination than randomly selected control hubs of the same size (Fig. 4E, left). Moreover, coordination consistently increased according to hub size, implying that cooperative repair activity (rather than multiple separate hubs) preferentially occurs within one cell.

Intercellular variability in higher-order repair coordination

Interestingly, in bulk Hi-C data, we noticed sets of sites that independently contact shared other sites and display increased repair coordination (Fig. 4C and 4E, e.g., teal and purple hubs), while lacking considerable contacts and showing decreased coordinated repair amongst themselves (orange hub). We hypothesize that such instances likely reflect intercellular variability, where DNA repair hubs exist mutually exclusively in different cells (Fig. 4Di). To test the premise of heterogeneity, we examined “triplets” of contacting DSB sites identified from Hi-C data (Fig. 4Dii). Triplets were classified as either “complete”, in which contacts were found between all three sites (forming AB, BC, and AC), or “incomplete”, which lacked one contact (AC). Using these categories based on Hi-C data, we found that complete triplets are more likely to show cooperative repair between AB and BC, while incomplete triplets do not (Fig. 4F). Further, the strength of a triplet is a measure of the fraction of cells that show repair enrichment at all three DSB sites in the single-cell repair data. By this definition, we observe that the strongest and weakest triplets are generally complete and incomplete, respectively (Fig. S5I). Our analysis thus shows evidence of both higher-order and mutually exclusive DSB repair contacts, which are indistinguishable in Hi-C population data where both appear as pairwise interactions. Collectively, these data illustrate a range of repair coordination, with multi-way architectural repair hubs that may heterogeneously exist across single cells in various conformations.

Discussion

Above, we established the framework for sensitive, genome-wide, single-cell DNA repair profiling. Our experimental and computational approaches enable detection of repair enrichment at kilobase-scale resolution in thousands of individual cells. The sequence-specific DSB induction system allowed us to examine coordinated repair activity at known genomic positions. We show that repair proteins spread according to topological features but exhibit intercellular variability. To reveal the underlying mechanism whereby DDR signaling meets genome control, we envision that future work will combine genome-wide mapping of repair and chromosomal contacts or structural proteins in single cells. Our analysis of multiple repaired loci across the genome showed that DNA repair occurs in a cooperative fashion. Loci known to form contacts more frequently showed pairwise as well as higher-order repair coordination. This observation provides unambiguous evidence of DSBs undergoing simultaneous repair in single cells, information not currently attainable from spatial contact approaches.

The genomics toolbox now available enables study of other DNA damage and repair mechanisms, including stochastic systems and *in vivo* models. Stratification of single-cell DNA repair profiles according to an additional layer of information is amenable to many different questions of interest. We demonstrated the applicability by using live DNA content staining prior to index sorting to explore the cell cycle. A similar approach can be envisioned for any cellular observation that is compatible with FACS, e.g., antibody staining for cell type annotation, classification of apoptosis or other stress responses, and mitochondrial labeling. We foresee that (multimodal) single-cell measurements will disentangle the role of chromatin, transcription, and other factors involved in DNA repair at high resolution, with implications for our understanding of cellular response and fate after damage.

Acknowledgments

We thank members of the Kind and Legube groups for useful feedback. Specifically, we acknowledge Koos Rooijers for developing computational methods and performing bioinformatic analyses. Purified pA-MNase and ^{m6}A-Tracer protein were respective gifts from Alexander van Oudenaarden and Bas van Steensel. We thank Reinier van der Linden at the Hubrecht FACS Facility and Anko de Graaff at the Hubrecht Imaging Center for excellent technical support and training. High-throughput sequencing and computational resources were provided by the Utrecht Sequencing Facility (USEQ) and High-Performance Computing Facility.

Funding

The Oncode Institute is supported by the KWF Dutch Cancer Society. This work was further funded by: Dutch Research Council (NWO-ALW) Vidi grant 161.339 (JK), Fulbright Commission the Netherlands promovendus grant (KLdL), European Molecular Biology Organization grant 9178 (KLdL).

Author contributions

Conceptualization: KLdL and JK. Methodology, Investigation, Data curation & Validation: KLdL. Formal analysis & Visualization: KLdL and PMJR. Software: PMJR. Project administration & Supervision: JK. Resources: GL and JK. Funding acquisition: KLdL and JK. Writing – original draft: KLdL with input from PMJR and JK. Writing – review & editing: all authors.

Competing interests

The authors declare no competing interests.

Data, code, and materials availability

Newly generated sequencing data will be made publicly available on the NCBI Gene Expression Omnibus under accession code GSE229874; use token wjizommazbgnrsd while under private status. See Methods for accession numbers of publicly available sequencing data analyzed in this manuscript. Imaging data will be made publicly available on Mendeley Data. Key scripts will be made publicly available on GitHub / Zenodo. Any additional material and/or information required to reproduce this work is available from the corresponding author upon request and may be subject to an MTA.

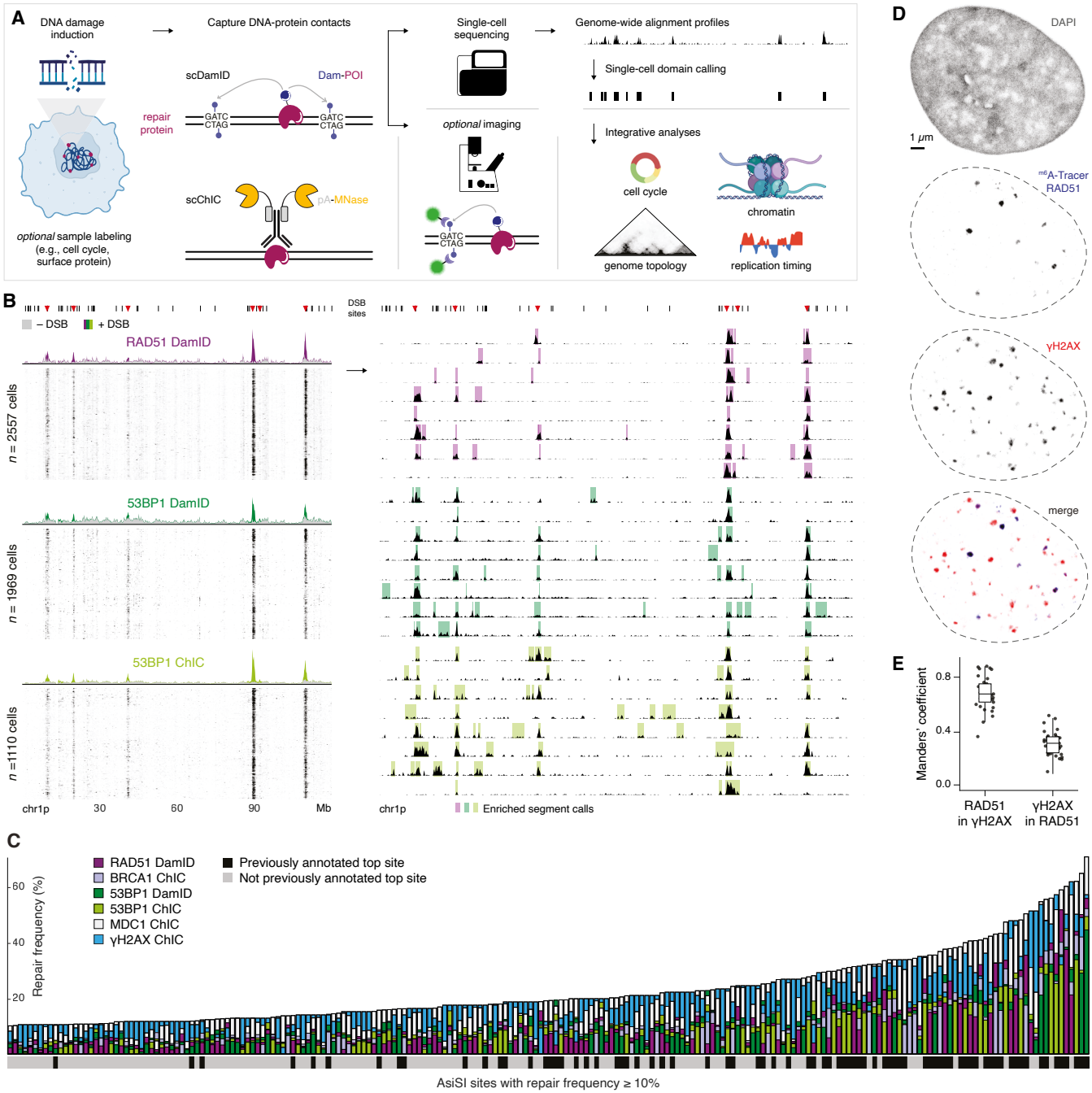
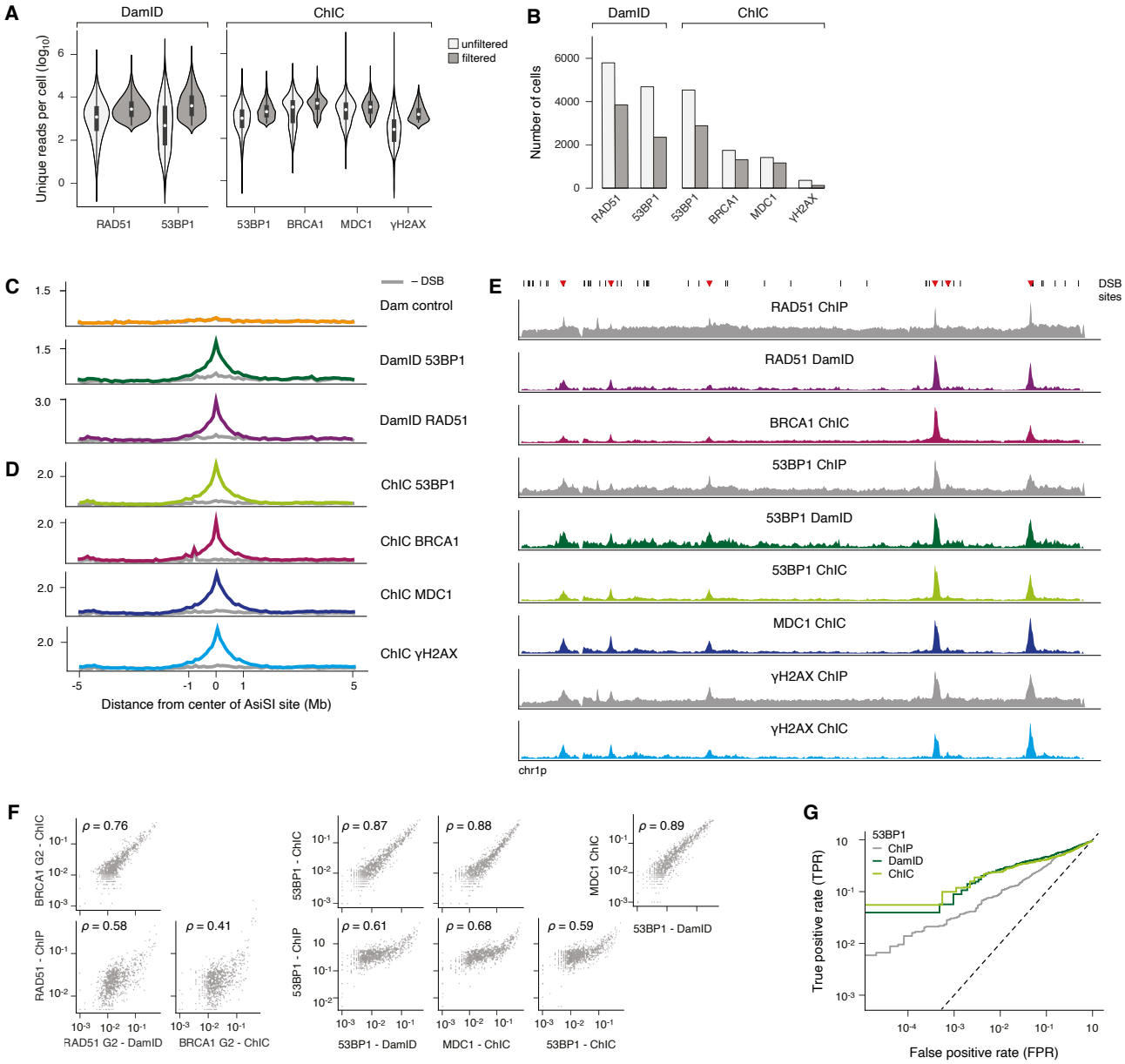


Fig. 1. Detecting double-strand break repair proteins with DamID and ChIC.

(**A**) Overview of experimental setup and integrations. (**B**) Signal on chromosome 1p. Left: single-cell heatmaps (RPKM) in +DSB condition, with single-cell aggregates of +DSB (colored) and –DSB (gray) conditions on top. Right: single-cell line plots with overlaid MSR calls. AsiSI motifs annotated with black lines, and red triangles indicating frequently cleaved (or "top") sites. (**C**) Bar plot of all AsiSI sites with repair frequency ≥ 10%, colored by repair frequencies per dataset (target protein & method) per site. Sites are ordered (on the x-axis) by increasing absolute repair frequency (i.e., highest frequency in any dataset). Per site, bars are ordered by increasing repair frequency per dataset (front to back; i.e., bars are not stacked). Bottom horizontal bar indicates previous (lack of) annotation as top site. (**D**) Confocal images of one representative nucleus showing DAPI, RAD51 DamID ^{m6}A-Tracer, and endogenous γH2AX immunofluorescent staining. (**E**) Quantification of signal colocalization (Manders' A and B per nucleus), n = 33 nuclei.



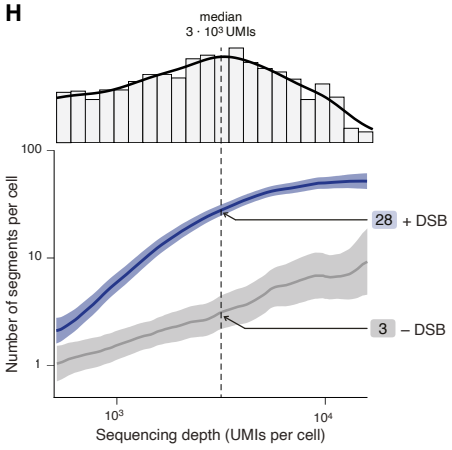


Fig. S1. Technical validation of DamID and ChIC repair protein data.

(**A**) Distribution of sequencing depth per protein target per method, unfiltered and filtered (retaining samples passing threshold). (**B**) Number of cells per protein target per method, unfiltered and filtered. (**C**) DamID alignment plots (averaged across single cells) over top AsiSI sites, in +DSB (colored) and –DSB (gray) conditions. (**D**) ChIC alignment plots (averaged across 100-cell samples) as in (A). (**E**) Line plots of chromosome 1p as in Figure 1B, including external ChIP-seq data. (**F**) Scatter plots of ChIP RPKM versus DamID and ChIC repair frequency, per protein and relevant cell cycle phase. Each dot is an AsiSI site. Correlation values indicate Spearman's *rho*. (**G**) Receiver-Operator Characteristic (ROC) curves showing TPR versus FPR for 53BP1 ChIP, ChIC and DamID, testing presence of a called repair segment and whether it falls on an AsiSI motif. Dashed line: x = y. (**H**) Line graphs showing the number of enriched segments as a function of sequencing depth in both DSB induction conditions. Bold line indicates mean, shaded area indicates 95% CI. Histogram of sequencing depth is annotated on top, dashed vertical line indicating the median.

Fig. S2

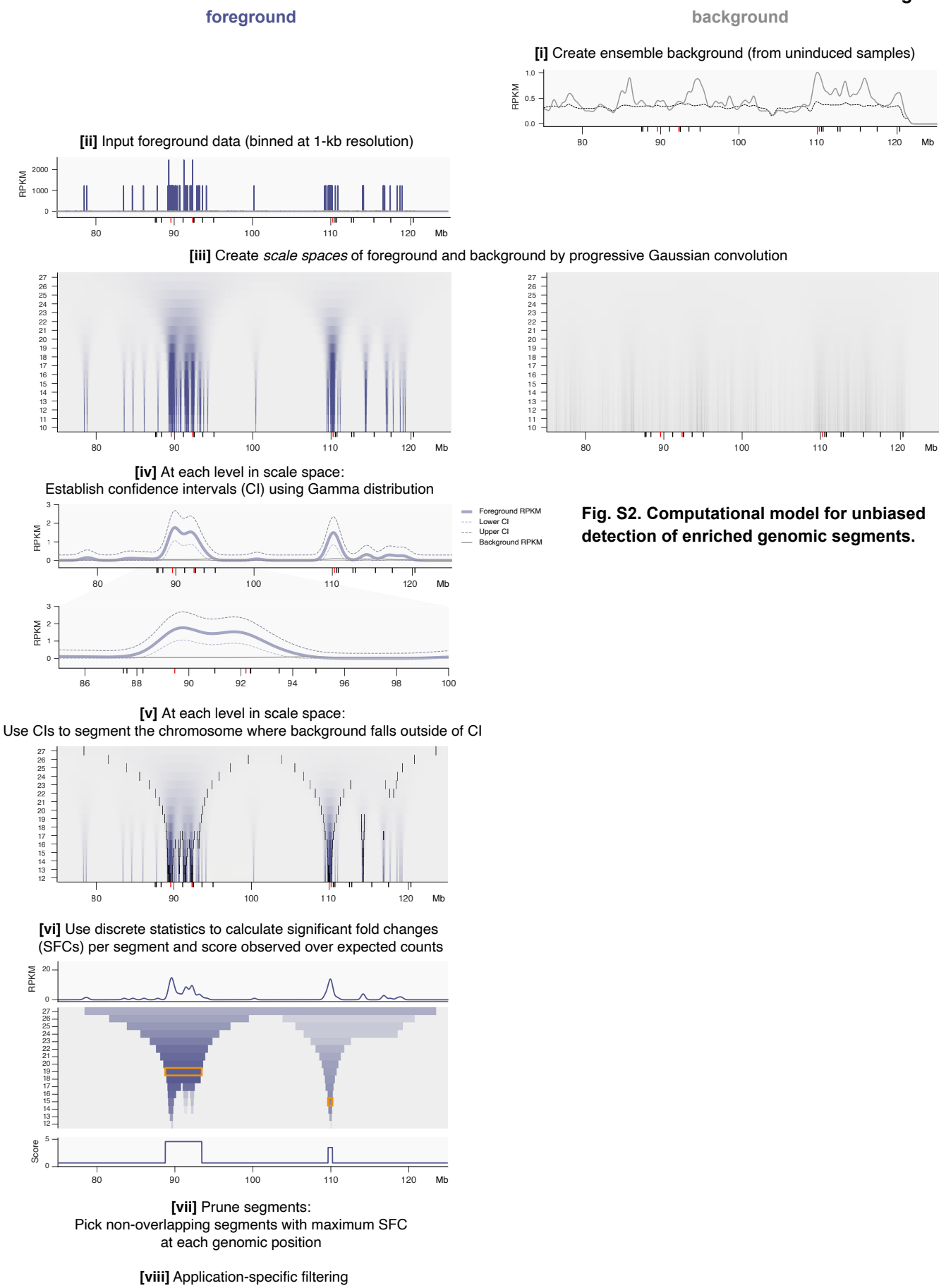


Fig. S2. Computational model for unbiased detection of enriched genomic segments.

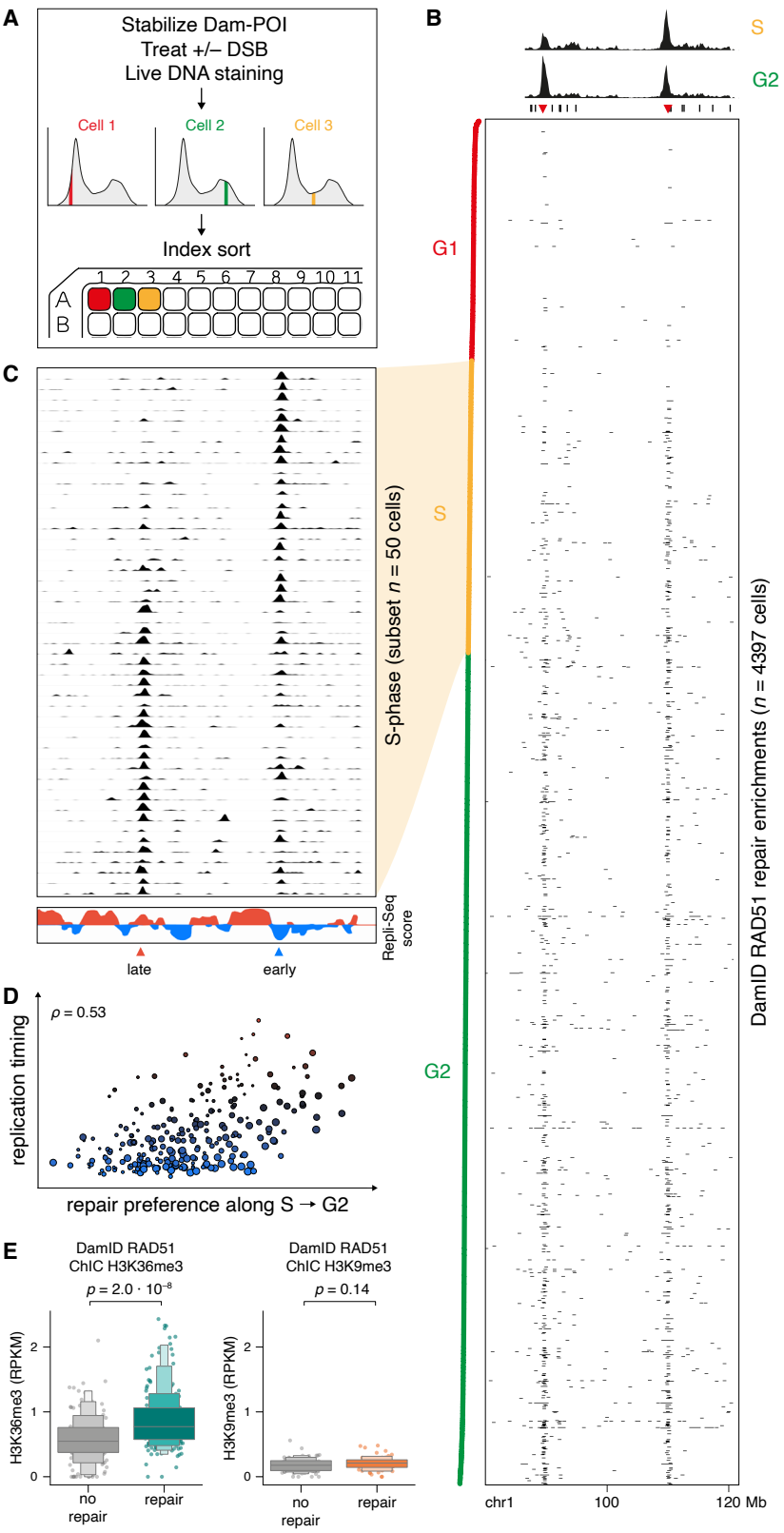
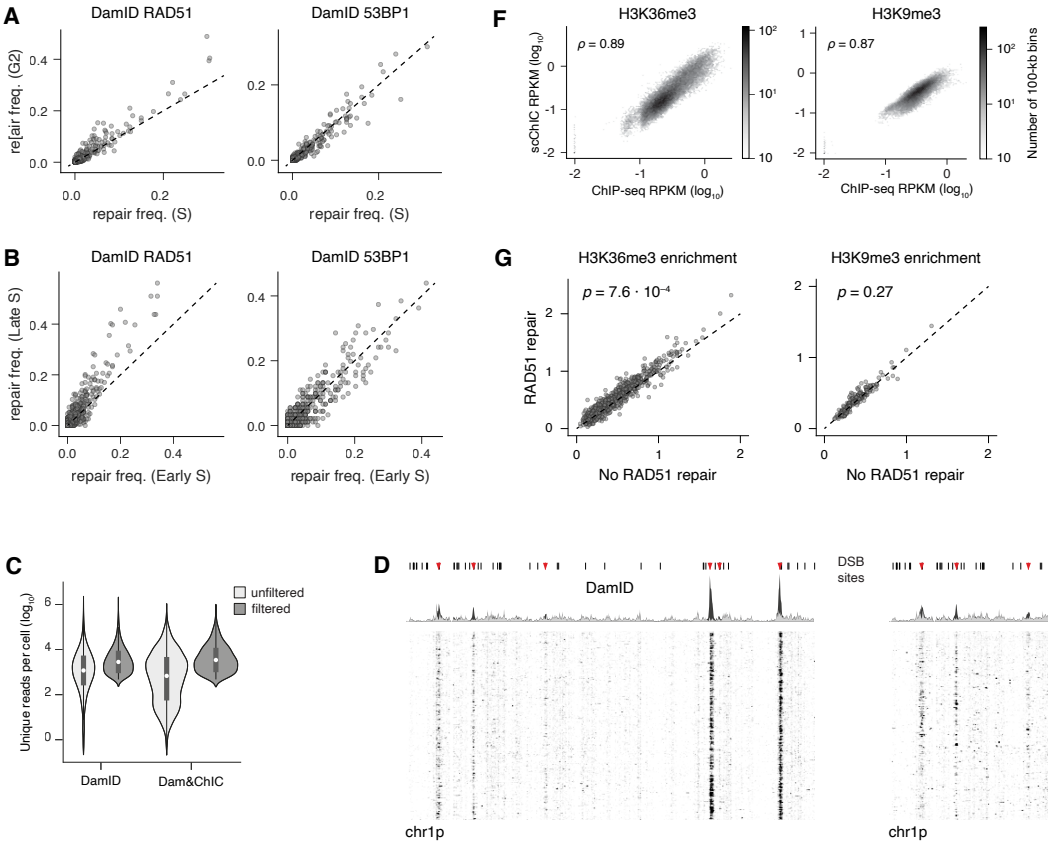


Fig. 2. RAD51 correlates with cell cycle phase, replication timing, and transcriptionally active chromatin in single cells.

(**A**) Overview of DamID experimental setup up to and including FACS. (**B**) Heatmap showing binarized RAD51 DamID enrichment (MSR calls) of 4397 single cells, at a selected region on chromosome 1. DNA content as measured by live Hoechst staining during FACS is annotated along the vertical axis (left side): G1 in red, S in yellow, G2 in green. Single-cell aggregates of +DSB condition for S and G2 phases on top, with AsiSI motifs annotated as in (1B). (**C**) Line plots showing RAD51 signal (RPKM) for a subset of cells in S phase, region as in (B). (**D**) Quantification of correlation between cell cycle preference and Repli-Seq scores for AsiSI clusters with repair frequency ≥ 5%. Size of circles indicates repair frequency. (**E**) Boxen plots showing H3K36me3 (left) and H3K9me3 (right) signal (RPKM) at AsiSI site 505, for cells with (right) and without (left) RAD51 enrichment. Statistical significance was calculated by two-sample Kolmogorov-Smirnov test.



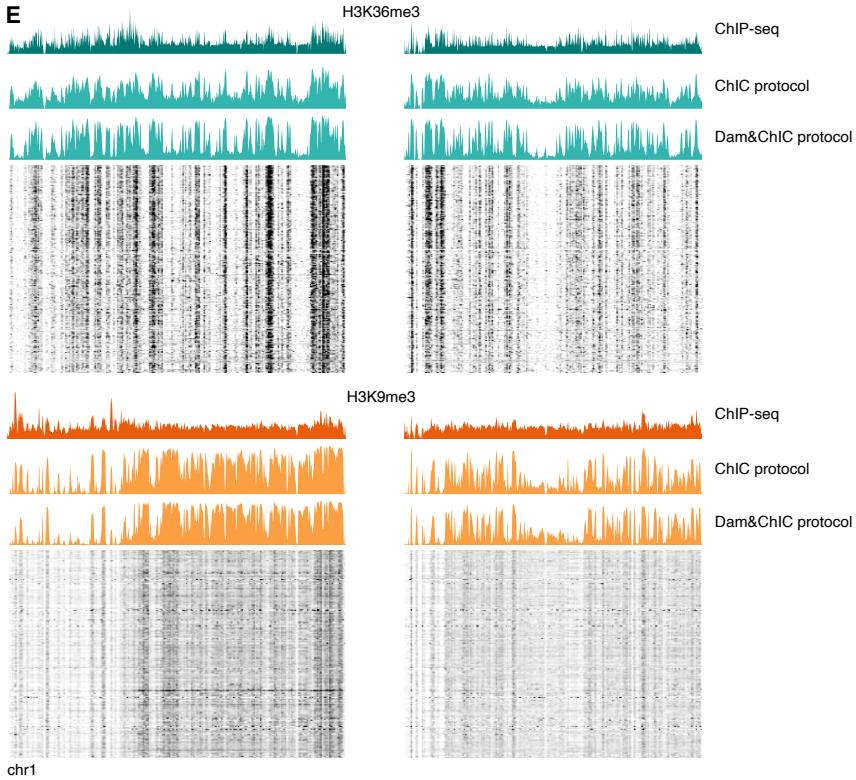


Fig. S3. Validation and characterization of pathway-specific repair frequencies.

(**A**) Scatter plots of RAD51 and 53BP1 DamID repair frequency in G2 vs S phase. Each dot is an AsiSI cluster. Dashed line: x = y. (**B**) Scatter plots as in (A), showing Late vs Early S phase. (**C**) Sequencing depth (UMIs per cell) for samples prepared with ChIC or Dam&ChIC protocols. (**D**) DamID RAD51 signal on chromosome 1p, generated with DamID (left) or Dam&ChIC (right) protocols. (**E**) Signal on chr1, for H3K36me3 (top) and H3K9me3 (bottom). Line plots show ChIP-seq (RPKM) and single-cell aggregate ChIC (contact frequency). Heatmaps show scChIC (RPKM) of 1000 cells. Aggregate ChIC signal is split on sample preparation by ChIC or Dam&ChIC. (**F**) Hexbin density plots showing the bin-based correlations between ChIP-seq and scChIC sequencing depth. Hexbins are colored by the number of 100-kb genomic bins. (**G**) Genome-wide quantification of H3K36me3 (left) or H3K9me3 (right) signal based on presence or absence of RAD51 enrichment. For each AsiSI site, scatter plots show the mean hPTM ChIC signal in cells that are repaired by RAD51 (y) or not (x). Statistical significance was calculated by one-sample Kolmogorov-Smirnov test.

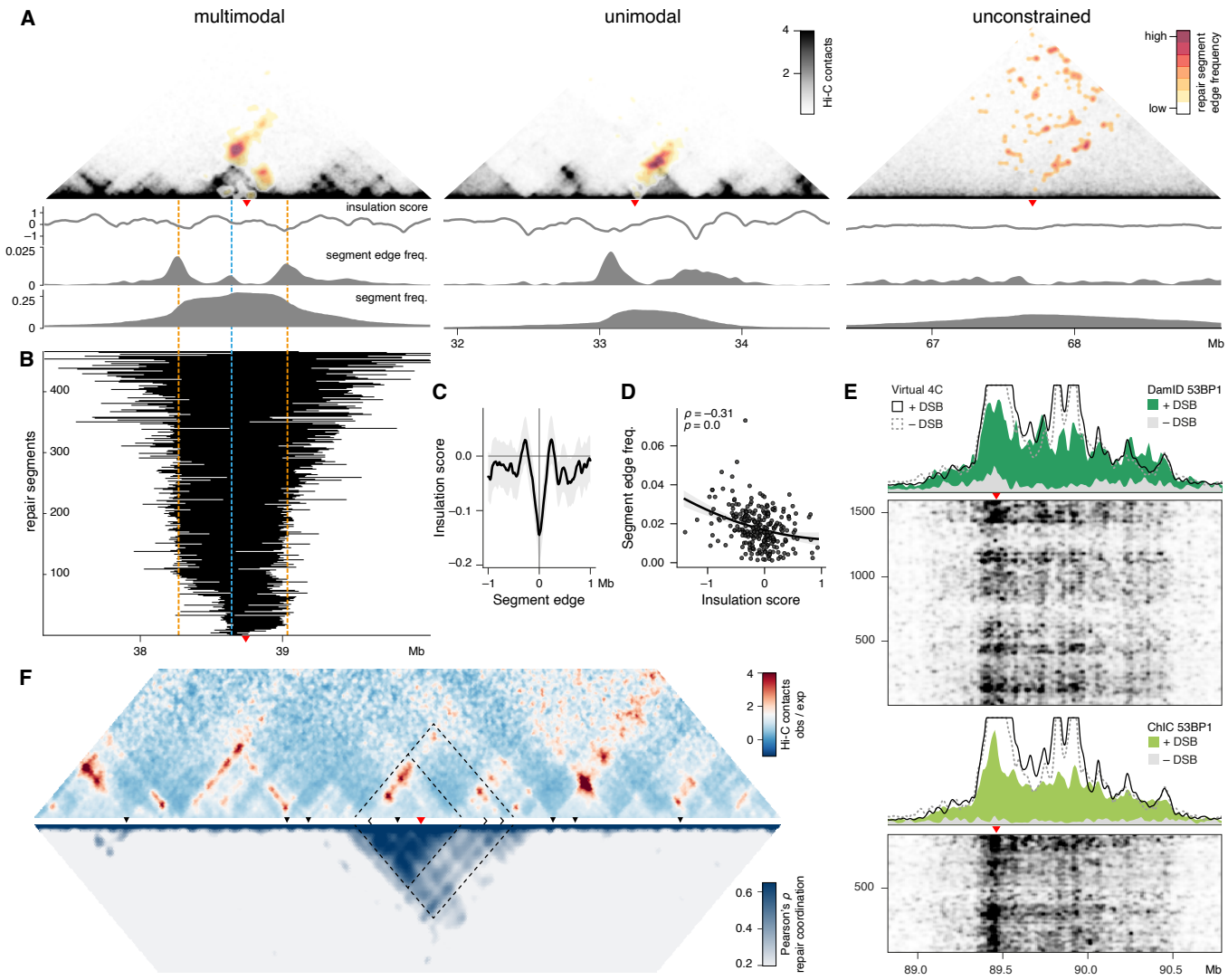


Fig. 3. Repair protein spreading in single cells mimics underlying genome topology.

(**A**) Overlay of 53BP1 DamID repair segment edges onto the Hi-C contact matrix for regions harboring a DSB site. Repair signal is colored by the frequencies with which those edges are observed across single cells. One example region is shown per different scenario of repair signal spreading: multimodal (left), unimodal (middle), and unconstrained (right). (**B**) Individual repair segments as called per single cell, sorted on segment size. Dashed lines indicate local maxima of edge frequency: outermost (orange) and middle (blue) segment edges. (**C**) Quantification of insulation score across all segment edges (in a surrounding window of 2 Mb). Shaded area indicates 95% CI. (**D**) Quantification of the relationship (Pearson's *r*) between repair-based segment edge frequency and topology-based insulation score at that edge. (**E**) Comparison of virtual 4C (representation of Hi-C contacts formed along the linear genome) and 53BP1 signal (DamID top, ChIC bottom). Single-cell aggregates for both conditions are plotted above single-cell heatmaps +DSB. (**F**) Comparison of the normalized Hi-C contact matrix (top) and single-cell repair coordination (bottom) for a 5-Mb region. Repair coordination across cells is measured per pair of genomic bins.

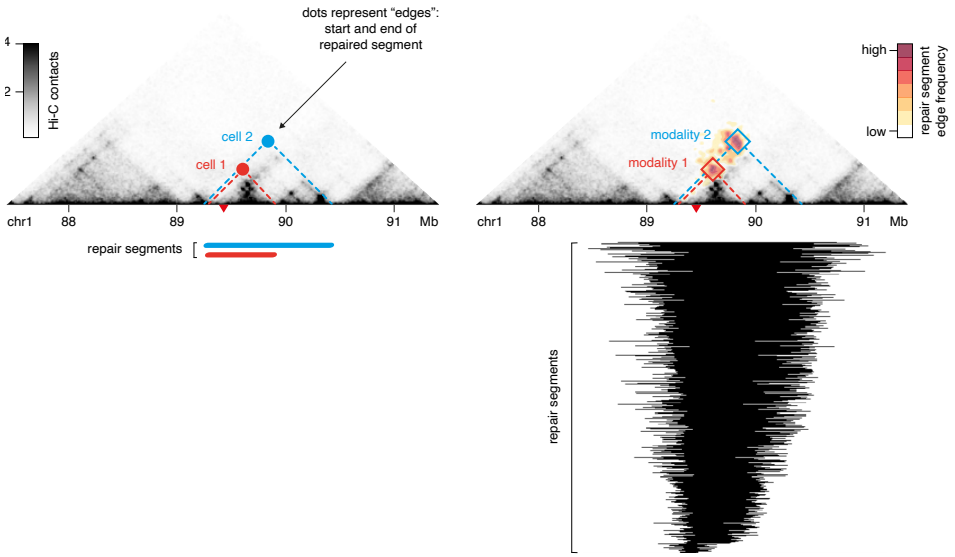


Fig. S4. Intercellular variability visualized by projection of repair segment edges onto Hi-C contact profiles.

Overlay of repair segment edges onto the Hi-C contact matrix for regions harboring a DSB site. Repair signal is colored by the frequencies with which those edges are observed across single cells. Left: cartoon indicates repair segments of different sizes, originating from two individual cells. Each repair data point projected onto the Hi-C map represents the genomic coordinates of segment start and end. Repair segments (as called on the linear genome) are annotated underneath, sorted on segment size (large to small, top to bottom).

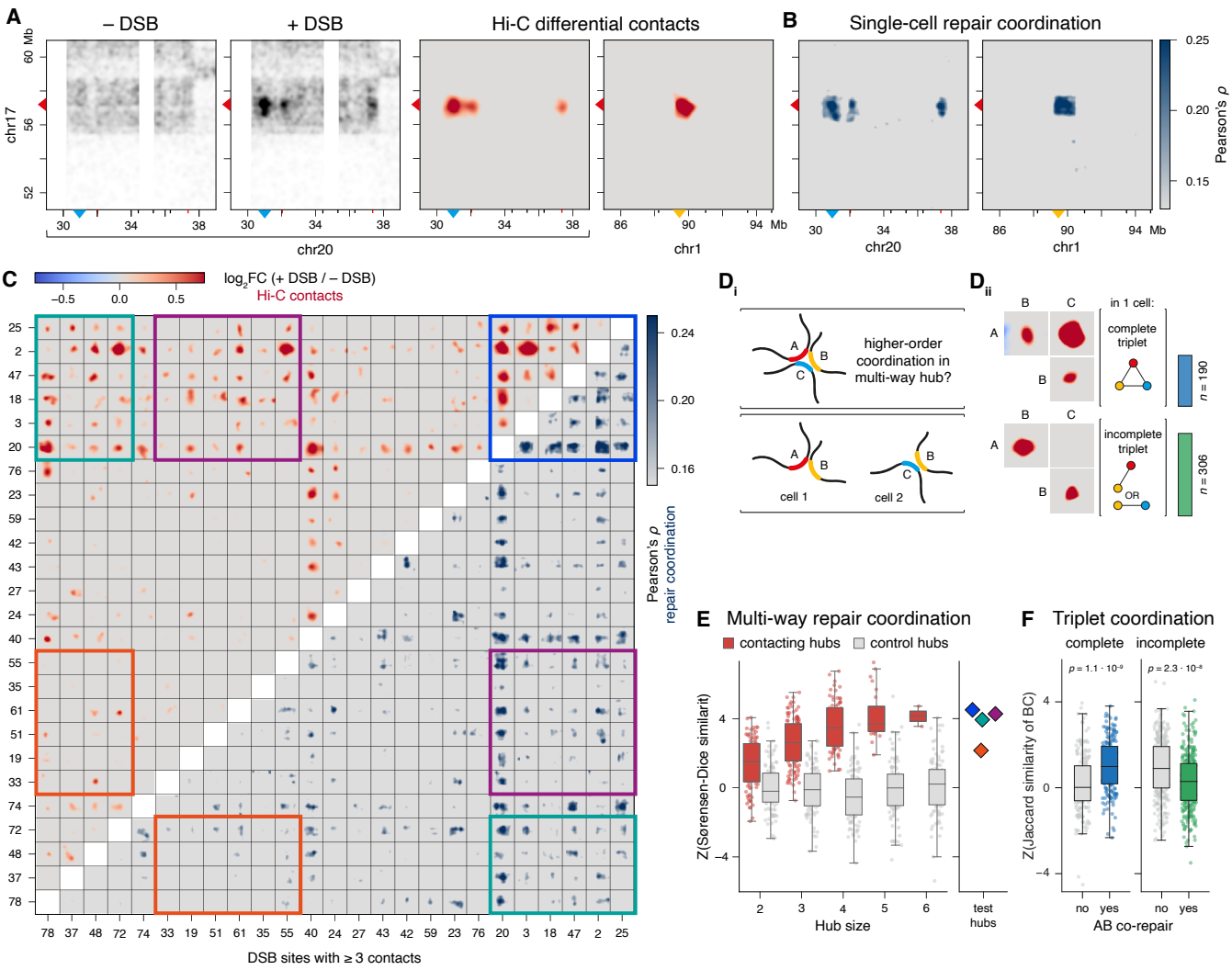


Fig. 4. Higher-order repair coordination of long-range contacts in single cells.

(**A**) Hi-C signal in 10-Mb regions harboring DSB sites. Left two panels show Hi-C contacts of chr17 (y) and chr20 (x) in –DSB and +DSB conditions. Right two panels show differential Hi-C contacts ($(\log_2 FC$ (+DSB / –DSB)) of chr17 with chr20 and chr1, respectively. (**B**) Single-cell repair coordination of all pairwise combinations of bins in selected regions as in (A). Coordination is measured as Pearson's correlation between 40-kb bins across cells. (**C**) Clustered heatmap and hub annotation as in (S4G), displaying pairwise damage-specific Hi-C contacts (red, top left) and single-cell 53BP1 repair coordination (dark blue, bottom right), in a region of 1 Mb surrounding each AsiSI site. (**D**) (**i**) Schematic representation of higher-order (top) and mutually exclusive (bottom) contacts. (**ii**) Categorization into triplets of DSB sites [A, B, C]. Incomplete triplets lack contact AC. (**E**) Higher-order repair coordination measured across hub sizes, for hubs in contact (red) and control hubs of the same size (gray). Similarity is measured with the normalized multi-way Sørensen-Dice coefficient. (**F**) Triplet-based coordination (or co-repair) measured for complete and incomplete triplets. Co-repair of BC is quantified in cells showing absence or presence of AB co-repair.

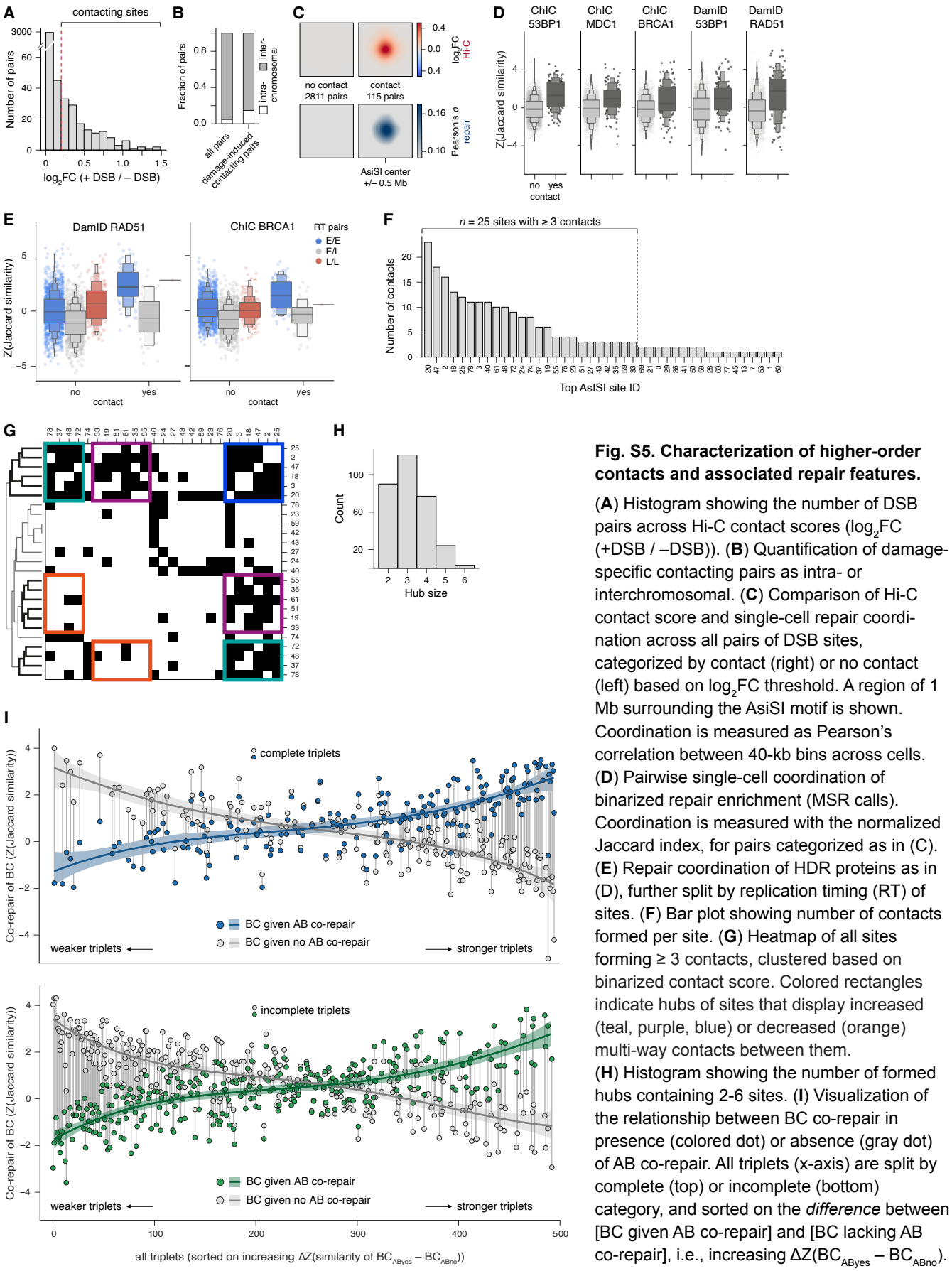


Fig. S5. Characterization of higher-order contacts and associated repair features.

(**A**) Histogram showing the number of DSB pairs across Hi-C contact scores ($\log_2$FC (+DSB / –DSB)). (**B**) Quantification of damage-specific contacting pairs as intra- or interchromosomal. (**C**) Comparison of Hi-C contact score and single-cell repair coordination across all pairs of DSB sites, categorized by contact (right) or no contact (left) based on $\log_2$FC threshold. A region of 1 Mb surrounding the AsiSI motif is shown. Coordination is measured as Pearson's correlation between 40-kb bins across cells. (**D**) Pairwise single-cell coordination of binarized repair enrichment (MSR calls). Coordination is measured with the normalized Jaccard index, for pairs categorized as in (C). (**E**) Repair coordination of HDR proteins as in (D), further split by replication timing (RT) of sites. (**F**) Bar plot showing number of contacts formed per site. (**G**) Heatmap of all sites forming ≥ 3 contacts, clustered based on binarized contact score. Colored rectangles indicate hubs of sites that display increased (teal, purple, blue) or decreased (orange) multi-way contacts between them. (**H**) Histogram showing the number of formed hubs containing 2-6 sites. (**I**) Visualization of the relationship between BC co-repair in presence (colored dot) or absence (gray dot) of AB co-repair. All triplets (x-axis) are split by complete (top) or incomplete (bottom) category, and sorted on the *difference* between [BC given AB co-repair] and [BC lacking AB co-repair], i.e., increasing $\Delta Z(BC_{AByes} - BC_{ABno})$.

Methods

Experimental methods

DamID construct design

Shield1-inducible Dam lines were created by cloning into the multiple cloning site (MCS) downstream of the double degron (DD) of ProteoTuner vector pPTuner IRES3 (PT4040-5). This generates an in-frame construct of DD-POI-V5-Dam. POI 53BP1 contains the minimal focus-forming region (amino acids 1221-1711). POI RAD51 contains the full endogenous protein (amino acids 1-339).

Generation of cell lines

Stable, clonal cell lines containing Dam-POI constructs were established by transfection and antibiotic selection. DIvA cells were grown in 24-well plates and transfected with 500 ng DamID plasmid and 1.5 µL Lipofectamine 2000 per well. Each well was passaged to a 15-cm dish and subjected to antibiotic resistance selection with 500 µg/mL G418 (Gibco) for 10 days (at complete death of untransfected control dishes). Monoclonal cell populations were hand-picked, expanded, and characterized by performing bulk DamID. Dam methylation levels were checked by evaluating methylation-specific amplification on agarose gel (as previously described in (*5,*53)) and on-target methylation signal was evaluated with high-throughput sequencing. One clone per construct with the highest signal-to-noise ratio (SNR) was chosen for single-cell experiments.

Cell culture and experimental treatment conditions

Cell lines were grown in DMEM containing high glucose, GlutaMAX supplement, and sodium pyruvate (Gibco) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich) and 1X Penicillin Streptomycin (Gibco), at 37 °C with 5% CO_2. For maintenance, cells were split 1/10 every 3 days and routinely tested for *Mycoplasma*. For experimental procedures, cells were plated the day prior, to achieve 60-70% confluency during induction treatment. Dam-POIs were stabilized by addition of 500 nM Shield1 ligand (AOBIOUS) for 4 or 8 h. Nuclear translocation of AsiSI-ER was simultaneously induced by addition of 300 nM 4-hydroxytamoxifen (4OHT; Sigma). Cells were harvested by washing twice with PBS0 followed by trypsinization with 1X TrypLE (Gibco), inactivation with DMEM, pipetting to yield a single-cell suspension, and centrifugation at 300 *g*. For FACS, cell pellets were resuspended in growth medium containing 10 µg/mL Hoechst 34580 (Invitrogen) per $1x10^6$ cells and incubated for 45-60 minutes at 37 °C, for live DNA content staining. Prior to sorting, cell suspension was passed through a 20-µm mesh. For bulk DamID, genomic DNA was isolated from cell pellets using commercial reagents (e.g., Promega Wizard). During each genomics experiment, cells were concurrently plated on glass and treated to verify proper induction and cleavage activity of AsiSI-ER by immunofluorescent staining and imaging.

Immunofluorescent staining

Cells were grown as described above, with the exception of plating on glass coverslips the day prior to experimental treatment. At the end of induction, cells were washed twice with PBS and chemically crosslinked with fresh formaldehyde solution (2% in PBS) for 10 minutes at RT, then permeabilized (with 0.5% IGEPAL® CA-630 in PBS) for 20 minutes and blocked (with 1% bovine serum albumin (BSA) in PBS) for 30 minutes. All antibody incubations were performed in final 1% BSA in PBS followed by three PBS washes at RT. Incubation with primary antibody against the endogenous protein as well as purified ^{m6}A-Tracer protein (recognizing methylated DNA) was performed at 4 °C for 16 hours (overnight), followed by anti-GFP (against ^{m6}A-Tracer protein) incubation at RT for 1 hour, and secondary antibody incubations at RT for 1 hour. The final PBS wash was simultaneously an incubation with DAPI (Invitrogen) at 0.5 µg/mL for 2 min, followed by a wash in MilliQ and sample mounting on glass slides.

Primary antibodies: anti-53BP1 Santa Cruz (rabbit) at 1/500, anti-γH2AX BioLegend (mouse) at 1/1000, anti-MDC1 Bethyl Laboratories (rabbit) at 1/500, anti-GFP Aves GFP-1020 (chicken) at 1/1000.
Secondary antibodies: all AlexaFluor at 1/500. Anti-chicken 488, anti-mouse 555 or 647, anti-rabbit 555 or 647.

Confocal imaging
Imaging (12-bit) was performed on an inverted scanning confocal microscope (Leica TCS SP8) with a HC PL APO CS3 63X (NA 1.40) oil-immersion objective and HyD detectors. Pinhole was set to 1 Airy Unit. Scanning zoom was set to 4-5X at a speed of 400 Hz. Full-nucleus images were acquired as *Z*-stacks at 0.2-µm intervals. Multi-color images were acquired sequentially (by frame).

Image analysis
Images were processed in Imaris 9.3 (Bitplane) by baseline subtraction and background correction with a 3x3(x3) median filter. Colocalization was calculated by Manders' coefficients M1 and M2 between channel 2 (DamID ^{m6}A-Tracer) and channel 3 (endogenous DSB repair marker). Pixels were retained that contained signal (intensity >10) in channel 1 (DAPI). Pixel intensity thresholds for the colocalization analysis were determined using Costes' method with default settings.

FACS
FACS was performed on BD FACSJazz or BD FACSInflux Cell Sorter instruments with BD Sortware. Index information was recorded for all sorts. Single cells were gated on forward and side scatters, trigger pulse width, and Hoechst cell cycle profiles. One cell per well was sorted into 384-well hard-shell plates containing 5 µL of filtered mineral oil and protocol-specific reagent.

DamID and derivative methods
<u>High-throughput sequencing</u>
Libraries were sequenced on the Illumina NextSeq 500 (75-bp single-end reads) or NextSeq 2000 (100-bp paired-end reads) platform.

<u>Bulk DamID</u>
DamID on populations was performed as previously described (*53*). Briefly, genomic DNA (gDNA) was isolated from cell pellets, digested with DpnI to enrich for Dam-methylated GATCs, and ligated to universal (not barcoded) double-stranded DamID adapter molecules. Methylation-specific PCR was performed with barcoded primer (unique per sample). Samples were pooled per clone, further processed to construct Illumina-compatible libraries, and sequenced to approximately 10 M raw reads per sample.

<u>Automated liquid handling</u>
Liquid reagent dispensing steps for single-cell protocols in microwell plates were performed on a Nanodrop II robot (Innovadyne Technologies / BioNex). Addition of barcoded adapters was done with a mosquito LV (SPT Labtech).

<u>Single-cell DamID2</u>
DamID on single cells was performed as previously described in detail *(54)*. Briefly, after FACS, cells were lysed and treated with proteinase K, after which methylated GATCs resulting from Dam enzyme activity were specifically digested with DpnI. Double-stranded adapters containing cell-specific barcodes and a T7 promoter were ligated to the blunt (DpnI-cleaved) DNA ends. Cells with non-overlapping barcodes were pooled together to undergo *in vitro* transcription (IVT), amplifying the genomic DamID-specific product in a linear manner. Library preparation was then performed on the amplified RNA, to generate molecules compatible with Illumina sequencing.

Single-cell Dam&T-seq

One single-cell DamID experiment was performed using the combinatorial scDam&T-seq approach capturing DamID and transcriptome *(7)*, as previously described in detail *(54)*, with the exception that all volumes were halved (to reduce costs). Briefly, after FACS, cells were lysed, followed by reverse transcription and second-strand synthesis in order to convert cellular mRNA into cDNA. Subsequent steps were followed according to the scDamID2 protocol.

(Single-cell) ChIC

ChIC was performed as described in detail in *(10)*, with adaptations as follows. After experimental treatment of cell cultures as described above, nuclei were isolated and permeabilized, incubated with primary antibody, then incubated with pA-MNase (Protein A IgG-binding domain fused to micrococcal nuclease, for antibody-specific binding) and Hoechst (for DNA content staining). If the primary antibody was raised in mouse, nuclei were incubated with secondary antibody (rabbit anti-mouse) before incubation with pA-MNase. After FACS, proximity-based cleaving by pA-MNase was activated (for exactly 30 min on ice), followed by inactivation and proteinase K treatment. MNase-cleaved ends were then blunted and phosphorylated, and double-stranded adapters were ligated. For one experiment, A-tailing was performed after end repair of MNase-cleaved ends, followed by ligation of T-tailed adapters. DNA molecules were then further processed for sequencing as in scDamID2.

Antibodies: anti-53BP1 Santa Cruz [H-300] sc-22760 (rabbit) at 1/500, anti-γH2AX BioLegend [2F3] 61340x (mouse) at 1/500, anti-MDC1 Bethyl Laboratories A300-053A (rabbit) at 1/500, anti-BRCA1 Santa Cruz [D-9] sc-6954 (mouse) at 1/500, anti-H3K9me3 Abcam ab8898 (rabbit) at 1/1000, anti-H3K36me3 Active Motif 6110x (rabbit) at 1/500, anti-mouse IgG Abcam ab6709 (rabbit) at 1/500.

Single-cell Dam&ChIC

Combinatorial profiling of DamID and ChIC was performed as described in *(34)*, with adaptations as follows. The same procedures of nuclei isolation, antibody treatment, FACS, and molecular preparation were followed as for scChIC. After end repair of MNase-cleaved ends, the scDamID procedure was followed, namely DpnI digestion to enrich for Dam-methylated GATCs, adapter ligation, and subsequent library preparation steps.

Computational methods

Raw data processing

Data generated by the scDamID and scDam&T-seq protocols was largely processed with the workflow and scripts described in (Markodimitraki et al., 2020) (see also www.github.com/KindLab/scDamAndTools). The procedure is described below, in brief. For detailed parameters and exact software versions, see www.github.com/KindLab/scRepair.

DamID and Dam&T samples

DamID data contains reads that result from DpnI-restriction activity. Briefly, raw sequencing data was processed by removing reads that contained contaminants (resulting from library preparation), using `cutadapt` (v2.0). Samples were then demultiplexed (using an in-house python script), and for every single sample one (or two, in the case of paired-end data) files were obtained. Demultiplexed data were aligned using `hisat2` (v2.1.0) to human reference genome hg19/GRCh37, supplemented with the ERCC spike-in sequences. Since DpnI cleaves GAˇTC, we prefixed each read *in silico* with a 'GA' dimer to improve alignment rates. Aligned reads were counted per genomic GATC position and a vector of counts per chromosome was stored. For DamID, reads not aligning at GATC positions were discarded. Reads were then binned in either 1-kb or 100-kb bins for further analysis and plotting.

ChIC and Dam&ChIC samples

Contaminant removal, demultiplexing and alignment was done as for DamID samples. Reads were counted per GATC position (for both ChIC-only and Dam&ChIC samples), but reads not aligning on GATC position, as well as unaligned reads were stored in an auxiliary BAM file. These non-GATC reads were processed by removal of the (*in silico* prefixed) 'GA' dimer, and realignment to the human reference genome. Realigned non-GATC reads were then counted UMI-unique per genomic position and counts were binned in either 1-kb or 100-kb bins for further analysis and plotting.

RPKMs and scaling factors

RPKMs were only used for inspection of raw read densities and were calculated using the common definition: $\mathrm{RPKM} = \frac{b}{\frac{T}{10^6} \times \frac{S}{10^3}}$ where b is the number of reads per bin, T is the sample total and S is the size of bins (in kb).

The nature of DSB binding patterns make the signal generally "inherently peaky" (high variance in read density along linear chromosomes). However, the number of DSBs and the ratio of signal (signal from DSB loci) and background (signal from inter-DSB loci) varies substantially between samples (i.e., SNR, signal-to-noise ratios). Naively scaling the read density as with RPKMs led to false enrichments/depletions, e.g., comparing a high SNR sample to an uninduced (no DSBs) background would register depletions across most inter-DSB regions. To overcome the inherent variation in SNR between the single-cell samples we employed the "scaling factor" normalization from PoissonSeq (*55*), which weakens the assumptions implicit of RPKM scaling by assuming only that a fraction (here: 50%) of bins are not differentially enriched between two conditions.

Enriched segment calling

To identify regions of significant enrichment and depletion we adapted the workflow of multi-scale representation of genomic signals as described in (*13*). A diagram of the workflow can be found in Figure S2. Each induced sample is compared to a suitable background. For DamID, this is an average of single-cell samples that are not induced with 4OHT (one background per clone). For ChIC, this is an average of single-cell samples that are not induced with 4OHT (one per antibody target). Starting from 1-kb bins, both foreground and background signals are convolved with a Gaussian kernel which standard deviation increases with $\sqrt{2}$ at each level, up to a standard deviation of 10 Mb. This creates the scale-space. Each level of the scale-space is segmented along the linear chromosome by comparing where background signal is either above or below the foreground signal (in practice a small confidence interval is built around the foreground signal to deal with zero data in the foreground signal and numerical rounding issues after Gaussian convolution, raising the possibility that background signal is *within* the confidence interval of the foreground signal; those segments are ignored in further steps). Each segment, where background signal is either below or above the foreground signal, at each level in the scale space is then tested for significant enrichment or depletion, respectively. We use the Gamma distribution to create a confidence interval around the background, using the observed signal density and the scaling factor (see above) of the foreground versus the background signal, and a *P*-value of 10^{-9}. A segment is considered significantly enriched or depleted when the foreground signal is either above or below this confidence interval. Similar to the original MSR, we calculate the "significant fold change" (SFC) as the observed foreground signal over the confidence interval boundary. In addition, we also record the true observed fold change of observed versus expected densities. Finally, we prune the enriched/depleted segments across all levels of the scale-space by selecting the level that yields the highest (absolute) SFC score.

Post-hoc filtering

We focus on "high-fidelity" enrichments in downstream analyses, satisfying the following conditions: observed logFC ≥1.25, size ≤10 Mb, segments encompass at least 10 UMI-unique reads in foreground sample.

Normalization of Hoechst measurements across batches

A Gaussian mixture model with 2 or 3 components, depending on fit, was applied to Hoechst intensity values, after which G1 and G2 peaks were assigned. Script is available on GitHub / Zenodo.

Defining top AsiSI sites

Top AsiSI site annotations were taken from (*28*).

External data processing

ChIP-seq

Raw ChIP-seq data was obtained from GSE48423, GSE97589 and E-MTAB-5817. Reads were aligned using HISAT2 (same parameters as our DamID and ChIC samples). Aligned reads with MAPQ ≥10 were counted in genomic bins of either 1 kb or 100 kb and used for downstream analyses.

Repli-seq

Repli-seq data was obtained from the 4DN project, SRP126407 and SRP197558.

Samples used:

`4DNFI6KIPWXQ - 4DNFID41JKT6 - 4DNFIFEZTAI1 - 4DNFIZDPE9T6.`

`SRR6363337 - SRR6363338 - SRR6363341 - SRR6363342 - SRR6363345 - SRR6363348 - SRR6363350 - SRR9040713 - SRR9040714.`

Sequencing data was aligned using HISAT2 using the same parameters as DamID and ChIC samples. Reads were counted in bins of 1 kb. Using sample annotations of early and late replication (and for one sample set, mid-replication) we used glmPCA (*56*) to obtain a principal component per chromosome indicative of replication timing in U2OS cells.

Hi-C

Files in `.hic` format were obtained from E-MTAB-8851.

Samples used:

`HiC_mOHT_rep1 - HiC_mOHT_rep2 - HiC_pOHT_rep1 - HiC_pOHT_rep2.`

The observed over expected Hi-C matrix presented in Figure 3F was calculated by dividing the normalized Hi-C matrix (binned at 25-kb resolution) by the average intrachrotmosomal contact probability between increasingly distant pairs of loci (i.e., distance-dependent contact decay) using `cooltools.lib.numutils.observed_over_expected`.

For Figure (S)4, the normalized Hi-C matrices (binned at 25-kb resolution) were smoothed using a Gaussian kernel (σ = 200 kb). The fold change was calculated (by dividing the smoothed induced matrix over the uninduced) and subsequently $\log_2$-transformed, after which insignificant logFCs were masked from the interaction maps.

hPTMs in cells with and without RAD51 repaired sites

The analysis presented Figure 2E and S2F, that demonstrates enriched H3K36me3 on AsiSI sites in cells where sites are repaired by RAD51, was performed as follows. Per site, cells were stratified by the presence or absence of an MSR segment call. Only sites that overlap with an MSR segment in at least 5 cells were included. In the resulting two groups of cells (i.e., with and without repair), the mean enrichment of either H3K36me3 or H3K9me3 ChIC signal was calculated in the consecutive non-overlapping 100-kb bin wherein the geometric mean of the AsiSI cluster was located.

MSR segment edge peak calling

For systematic comparison between MSR segment edge frequency and Hi-C insulation score (computed using `cooltools.insulation` on 25-kb binned normalized Hi-C matrix) presented in Figure 3C-D, segment edge peaks were identified as follows. The segment edge frequency was calculated by averaging the binary MSR calls over the single cells in 25-kb bins. The edge frequency vector was smoothed by fitting a Gaussian kernel (SD = 2.5×10^4). Segment edge frequency peaks were subsequently detected by `scipy.signal.find_peaks` (with prominence = P_i where the i-th Percentile is 97.5). Hi-C insulation score and segment edge frequency at called peak positions were used.

Identification of pairs of contacting repair sites

For the identification of contacting pairs of repair sites, a total of 79 AsiSI clusters that contain a previously annotated top site was included. The theoretical number of pairs is defined as $P = \frac{n!}{k!(n-k)!}$, where n is the number of AsiSI clusters (79) and $k = 2$, was tested for the presence of DSB-induced (+4OHT) 3D contacts in the logFC Hi-C matrix described above. The mean logFC was calculated within a window of 50-kb surrounding the pixel where both cluster geometric means of the pair meet in the Hi-C matrix. Based on the distribution of mean logFC values across the 3081 pairs, a threshold was set at ≥0.2 to discriminate contacting from not-contacting pairs (Figure S4A). Pairs of sites less than 2 Mb apart in linear genomic distance were excluded from the analysis.

Analysis of repair coordination

Analysis of repair coordination across single cells is largely categorized in two: 1) based on the binary MSR segment calls, that were used for both pairwise and multi-way similarity metrics (Figure 4E-G, S4D-E) and 2) based on quantitative depth-normalized repair signal, the Pearson's coefficient was calculated to evaluate pairwise coordination. The details of both approaches are described below.

<u>Coordination of binary MSR repair segments</u>

To measure coordination of binary repair signal, MSR segments overlapping the AsiSI cluster geometric mean were used to determine whether a cluster is repaired in a given cell. A total of 79 AsiSI clusters that contain a previously annotated top site was included. The resulting $m \times n$ matrix is defined as $M = (a_{ij})$, where i is a cell and j is an AsiSI cluster. Pairwise distances were calculated between all pairs of clusters using the Jaccard similarity index, defined as the size of the intersection over the size of the union:

$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ where A and B represent any pair of clusters in M.

For multi-way coordination analysis of repair hubs (i.e., ≥3 contacting AsiSI clusters), we applied a multiple-site extension of the Sørensen-Dice similarity index on binary matrix M (*51*). For a detailed justification and explanation of the Sørensen-Dice multi-way similarity metric, see (*51, 57, 58*). We briefly describe the way the metric was used here, below. The multiple-site similarity index for any number of T AsiSI clusters can be formulated as

$$C_S^T = \frac{T}{T-1}\left(\frac{\sum_{i<j} a_{ij} - \sum_{i<j<k} a_{ijk} + \sum_{i<j<k<l} a_{ijkl} - \dots}{\sum_i a_i}\right)$$

where a_i is the number of cells in which cluster A_i is repaired, a_{ij} the number of cells that share repair of cluster A_i and A_j, and a_{ijk} the number of cells that share repair of cluster A_i, A_j and A_k, and so on. In case of $T = 2$ the outcome would reflect the definition of the original pairwise Sørensen-Dice similarity index. We wrote a Python implementation of the `betapart` R package (*52*) to compute the multiple-site Sørensen similarity index on a subset of matrix M defined as: $N = (M_{ij})_{\substack{1 \le i < a \\ j \in c}}$ where c is a vector of AsiSI cluster indices of any size in M. Our `pyBetapart` function is available on GitHub / Zenodo.

Coordination of quantitative repair signal

Coordinated behavior of quantitative repair signal for the analysis presented in Figure 3F and (S)4 was computed as follows. For intra-chromosomal coordination maps (Figure 3F), the single-cell data was first binned at 20-kb resolution, RPKM-normalized and smoothed using a Gaussian kernel (SD = 4 × 10^4). For inter-chromosomal coordination maps (Figure (S)4), data was binned at 40 kb, RPKM-normalized and smoothed (SD = 4 × 10^5). For all possible pairs of bins surrounding the AsiSI cluster geometric mean i and j, repair coordination was calculated as the Pearson's correlation coefficient between the vectors b_i and b_j that contain the RPKM values of bin i and j across the single cells.

Identifying contacting hubs

The identification of hubs of contacting AsiSI clusters builds on the pair identification described above, but is elaborated by walking through Hi-C matrix M^2 (Figure S4G), testing all theoretical combinations of hubs with defined sizes for Hi-C contacts. The hub identification function for any number of k AsiSI clusters included in the hub can be written as $H = \frac{n!}{k!(n-k)!}$ where n is the total number of AsiSI clusters included in M^2. All theoretical combinations of hubs were called as contacting hub if all AsiSI clusters within H contact each other (Hi-C logFC ≥0.2). Our hub identification function written in Python is available on GitHub/ Zenodo.

Triplet categorization

For the analysis of (in)complete triplets (Figure 4), we built on our pairwise identification described above. We first determined all combinations of three among a total of 79 AsiSI clusters that contain a previously annotated top site. Theoretical triplets T were tested for the presence of Hi-C based contacts (logFC ≥ 0.2) and called "complete" in case all three sites contact each other (AB, AC, BC) or "incomplete" if one pair within the triplet does not contact while both other pairs do $(AB, \cancel{AC}, BC)$.

Conditioning of cells for triplet analysis

For triplet analysis presented in Figure 4F-G, cells were stratified by the presence of a binary MSR segment on *both* AsiSI cluster A *and* B (i.e., "AB co-repair" or "synchronous AB"). Cells with an MSR segment on *either* A *or* B were used for "no AB co-repair" or "asynchronous AB". In these two conditions of cells, the pairwise coordination of BC was measured using the Jaccard similarity index as described above.

Coordination Z-score normalization by matrix permutation

Two potential problems are posed on our coordination analysis that might introduce undesirable bias: 1) typical sparsity of single-cell data results in non-uniformly distributed signal dropout and 2) binary similarity metrics can be sensitive to site prominence that differs between sites, but which does not reflect coordination. To solve both problems, we applied a previously described algorithm (*59*), to randomize the abovementioned presence-absence matrix M n-times (n = 100), without altering row and column totals. The resulting randomized matrices were used to Z-score normalize binary coordination metrics, which can be written as $Z = \frac{x-\mu}{\sigma}$, where x is the similarity score of the observed, μ the mean similarity, and σ the standard deviation of the random controls.

References

1. I. Mitrentsi, D. Yilmaz, E. Soutoglou, How to maintain the genome in nuclear space. *Curr Opin Cell Biol* **64**, 58-66 (2020).
2. A. Carre-Simon, E. Fabre, 3D Genome Organization: Causes and Consequences for DNA Damage and Repair. *Genes (Basel)* **13**, (2021).
3. Q. Zhu, Y. Niu, M. Gundry, C. Zong, Single-cell damagenome profiling unveils vulnerable genes and functional pathways in human genome toward DNA damage. *Sci Adv* **7**, (2021).
4. B. van Steensel, S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-428 (2000).
5. M. J. Vogel, D. Peric-Hupkes, B. van Steensel, Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467-1478 (2007).
6. J. Kind, L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey, M. Bienko, Y. Zhan, B. Lajoie, C. A. de Graaf, M. Amendola, G. Fudenberg, M. Imakaev, L. A. Mirny, K. Jalink, J. Dekker, A. van Oudenaarden, B. van Steensel, Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-147 (2015).
7. K. Rooijers, C. M. Markodimitraki, F. J. Rang, S. S. de Vries, A. Chialastri, K. L. de Luca, D. Mooijman, S. S. Dey, J. Kind, Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat Biotechnol* **37**, 766-772 (2019).
8. M. Schmid, T. Durussel, U. K. Laemmli, ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**, 147-157 (2004).
9. W. L. Ku, K. Nakamura, W. Gao, K. Cui, G. Hu, Q. Tang, B. Ni, K. Zhao, Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
10. P. Zeller, J. Yeung, H. Viñas Gaza, B. A. de Barbanson, V. Bhardwaj, M. Florescu, R. van der Linden, A. van Oudenaarden, Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat Genet* **55**, 333-345 (2023).
11. J. S. Iacovoni, P. Caron, I. Lassadi, E. Nicolas, L. Massip, D. Trouche, G. Legube, High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* **29**, 1446-1457 (2010).
12. L. Massip, P. Caron, J. S. Iacovoni, D. Trouche, G. Legube, Deciphering the chromatin landscape induced around DNA double strand breaks. *Cell Cycle* **9**, 2963-2972 (2010).
13. T. A. Knijnenburg, S. A. Ramsey, B. P. Berman, K. A. Kennedy, A. F. Smit, L. F. Wessels, P. W. Laird, A. Aderem, I. Shmulevich, Multiscale representation of genomic signals. *Nat Methods* **11**, 689-694 (2014).
14. P. Caron, J. Choudjaye, T. Clouaire, B. Bugler, V. Daburon, M. Aguirrebengoa, T. Mangeat, J. S. Iacovoni, A. Alvarez-Quilon, F. Cortes-Ledesma, G. Legube, Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep* **13**, 1598-1609 (2015).
15. J. Kind, L. Pagie, H. Ortabozkoyun, S. Boyle, S. S. de Vries, H. Janssen, M. Amendola, L. D. Nolen, W. A. Bickmore, B. van Steensel, Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-192 (2013).
16. R. Scully, A. Panday, R. Elango, N. A. Willis, DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**, 698-714 (2019).
17. K. Karanam, R. Kafri, A. Loewer, G. Lahav, Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol Cell* **47**, 320-329 (2012).
18. N. Hustedt, D. Durocher, The control of DNA repair by the cell cycle. *Nat Cell Biol* **19**, 1-9 (2016).
19. M. Macheret, T. D. Halazonetis, Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112-116 (2018).
20. T. N. Moiseeva, Y. Yin, M. J. Calderon, C. Qian, S. Schamus-Haynes, N. Sugitani, H. U. Osmanbeyoglu, E. Rothenberg, S. C. Watkins, C. J. Bakkenist, An ATR and CHK1 kinase signaling mechanism that limits origin firing during unperturbed DNA replication. *Proc Natl Acad Sci U S A* **116**, 13374-13383 (2019).
21. K. N. Klein, P. A. Zhao, X. Lyu, T. Sasaki, D. A. Bartlett, A. M. Singh, I. Tasan, M. Zhang, L. P. Watts, S. I. Hiraga, T. Natsume, X. Zhou, T. Baslan, D. Leung, M. T. Kanemaki, A. D. Donaldson, H. Zhao, S. Dalton, V. G. Corces, D. M. Gilbert, Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371-378 (2021).
22. C. Marchal, J. Sima, D. M. Gilbert, Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* **20**, 721-737 (2019).
23. Q. Du, S. A. Bert, N. J. Armstrong, C. E. Caldon, J. Z. Song, S. S. Nair, C. M. Gould, P. L. Luu, T. Peters, A. Khoury, W. Qu, E. Zotenko, C. Stirzaker, S. J. Clark, Replication timing and epigenome remodelling are associated with the

nature of chromosomal rearrangements in cancer. *Nat Commun* **10**, 416 (2019).

24. F. Aymard, B. Bugler, C. K. Schmidt, E. Guillou, P. Caron, S. Briois, J. S. Iacovoni, V. Daburon, K. M. Miller, S. P. Jackson, G. Legube, Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366-374 (2014).

25. A. Janssen, G. A. Breuer, E. K. Brinkman, A. I. van der Meulen, S. V. Borden, B. van Steensel, R. S. Bindra, J. R. LaRocque, G. H. Karpen, A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* **30**, 1645-1657 (2016).

26. K. Tsouroula, A. Furst, M. Rogier, V. Heyer, A. Maglott-Roth, A. Ferrand, B. Reina-San-Martin, E. Soutoglou, Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* **63**, 293-305 (2016).

27. S. Pellegrino, J. Michelena, F. Teloni, R. Imhof, M. Altmeyer, Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. *Cell Rep* **19**, 1819-1831 (2017).

28. T. Clouaire, V. Rocher, A. Lashgari, C. Arnould, M. Aguirrebengoa, A. Biernacka, M. Skrzypczak, F. Aymard, B. Fongang, N. Dojer, J. S. Iacovoni, M. Rowicka, K. Ginalski, J. Cote, G. Legube, Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell* **72**, 250-262 e256 (2018).

29. K. Nakamura, G. Saredi, J. R. Becker, B. M. Foster, N. V. Nguyen, T. E. Beyer, L. C. Cesa, P. A. Faull, S. Lukauskas, T. Frimurer, J. R. Chapman, T. Bartke, A. Groth, H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat Cell Biol* **21**, 311-318 (2019).

30. J. R. Becker, G. Clifford, C. Bonnet, A. Groth, M. D. Wilson, J. R. Chapman, BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature* **596**, 433-437 (2021).

31. R. Schep, E. K. Brinkman, C. Leemans, X. Vergara, R. H. van der Weide, B. Morris, T. van Schaik, S. G. Manzo, D. Peric-Hupkes, J. van den Berg, R. L. Beijersbergen, R. H. Medema, B. van Steensel, Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* **81**, 2216-2230 e2210 (2021).

32. T. Clouaire, G. Legube, A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks. *Trends Genet* **35**, 330-345 (2019).

33. J. Ferrand, A. Plessier, S. E. Polo, Control of the chromatin response to DNA damage: Histone proteins pull the strings. *Semin Cell Dev Biol* **113**, 75-87 (2021).

34. S. Kefalopoulou, P. M. J. Rullens, K. L. de Luca, A. van Oudenaarden, P. Zeller, J. Kind, Combined scDamID and sortChIC enables multifactorial chromatin profiling at single-cell resolution. Unpublished manuscript, *bioRxiv* pending (2023).

35. C. Arnould, V. Rocher, A. L. Finoux, T. Clouaire, K. Li, F. Zhou, P. Caron, P. E. Mangeot, E. P. Ricci, R. Mourad, J. E. Haber, D. Noordermeer, G. Legube, Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660-665 (2021).

36. J. Zagelbaum, A. Schooley, J. Zhao, B. R. Schrank, E. Callen, S. Zha, M. E. Gottesman, A. Nussenzweig, R. Rabadan, J. Dekker, J. Gautier, Multiscale reorganization of the genome following DNA damage facilitates chromosome translocations via nuclear actin polymerization. *Nat Struct Mol Biol* **30**, 99-106 (2023).

37. F. Natale, A. Rapp, W. Yu, A. Maiser, H. Harz, A. Scholl, S. Grulich, T. Anton, D. Horl, W. Chen, M. Durante, G. Taucher-Scholz, H. Leonhardt, M. C. Cardoso, Identification of the elementary structural units of the DNA damage response. *Nat Commun* **8**, 15760 (2017).

38. F. Ochs, G. Karemore, E. Miron, J. Brown, H. Sedlackova, M. B. Rask, M. Lampe, V. Buckle, L. Schermelleh, J. Lukas, C. Lukas, Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**, 571-574 (2019).

39. J. T. Sanders, T. F. Freeman, Y. Xu, R. Golloshi, M. A. Stallard, A. M. Hill, R. San Martin, A. S. Balajee, R. P. McCord, Radiation-induced DNA damage and repair effects on 3D genome organization. *Nat Commun* **11**, 6178 (2020).

40. P. L. Collins, C. Purman, S. I. Porter, V. Nganga, A. Saini, K. E. Hayer, G. L. Gurewitz, B. P. Sleckman, J. J. Bednarski, C. H. Bassing, E. M. Oltz, DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nat Commun* **11**, 3158 (2020).

41. J. Huang, C. Wu, J. A. Kloeber, H. Gao, M. Gao, Q. Zhu, Y. Chang, F. Zhao, G. Guo, K. Luo, H. Dai, S. Liu, Q. Huang, W. Kim, Q. Zhou, S. Zhu, Z. Wu, X. Tu, P. Yin, M. Deng, L. Wang, J. Yuan, Z. Lou, SLFN5-mediated chromatin dynamics sculpt higher-order DNA repair topology. *Mol Cell*, (2023).

42. T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, P. Fraser, Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).

43. T. Nagano, Y. Lubling, C. Varnai, C. Dudley, W. Leung, Y. Baran, N. Mendelson Cohen, S. Wingett, P. Fraser, A. Tanay, Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61-67 (2017).

44. T. J. Stevens, D. Lando, S. Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O'Shaughnessy-Kirwan, J. Cramard, A. J. Faure, M. Ralser, E. Blanco, L. Morey, M. Sanso, M. G. S. Palayret, B. Lehner, L. Di Croce, A. Wutz, B. Hendrich, D. Klenerman, E. D. Laue, 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59-64 (2017).

45. B. Bintu, L. J. Mateo, J. H. Su, N. A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A. N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).

46. T. Neumaier, J. Swenson, C. Pham, A. Polyzos, A. T. Lo, P. Yang, J. Dyball, A. Asaithamby, D. J. Chen, M. J. Bissell, S. Thalhammer, S. V. Costes, Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells. *Proc Natl Acad Sci U S A* **109**, 443-448 (2012).

47. F. Aymard, M. Aguirrebengoa, E. Guillou, B. M. Javierre, B. Bugler, C. Arnould, V. Rocher, J. S. Iacovoni, A. Biernacka, M. Skrzypczak, K. Ginalski, M. Rowicka, P. Fraser, G. Legube, Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* **24**, 353-361 (2017).

48. C. P. Caridi, C. D'Agostino, T. Ryu, G. Zapotoczny, L. Delabaere, X. Li, V. Y. Khodaverdian, N. Amaral, E. Lin, A. R. Rau, I. Chiolo, Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* **559**, 54-60 (2018).

49. B. R. Schrank, T. Aparicio, Y. Li, W. Chang, B. T. Chait, G. G. Gundersen, M. E. Gottesman, J. Gautier, Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61-66 (2018).

50. C. Arnould, V. Rocher, A. S. Bader, E. Lesage, N. Puget, T. Clouaire, R. Mourad, D. Noordermeer, M. Bushell, G. Legube, ATM-dependent formation of a novel chromatin compartment regulates the Response to DNA Double Strand Breaks and the biogenesis of translocations. *bioRxiv*, 2021.2011.2007.467654 (2021).

51. O. H. Diserud, F. Odegaard, A multiple-site similarity measure. *Biol Lett* **3**, 20-22 (2007).

52. A. Baselga, C. D. L. Orme, Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**, 808-812 (2012).

53. K. L. de Luca, J. Kind, Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells. *Methods Mol Biol* **2157**, 159-172 (2021).

54. C. M. Markodimitraki, F. J. Rang, K. Rooijers, S. S. de Vries, A. Chialastri, K. L. de Luca, S. J. A. Lochs, D. Mooijman, S. S. Dey, J. Kind, Simultaneous quantification of protein-DNA interactions and transcriptomes in single cells with scDam&T-seq. *Nat Protoc* **15**, 1922-1953 (2020).

55. J. Li, D. M. Witten, I. M. Johnstone, R. Tibshirani, Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **13**, 523-538 (2012).

56. F. W. Townes, S. C. Hicks, M. J. Aryee, R. A. Irizarry, Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol* **20**, 295 (2019).

57. A. Baselga, A. Jimenez-Valverde, G. Niccolini, A multiple-site similarity measure independent of richness. *Biol Lett* **3**, 642-645 (2007).

58. A. Baselga, Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution* **4**, 552-557 (2013).

59. G. Strona, D. Nappo, F. Boccacci, S. Fattorini, J. San-Miguel-Ayanz, A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat Commun* **5**, 4114 (2014).