

VIEWPOINT

Exploring the SSBreakome: genome-wide mapping of DNA single-strand breaks by next-generation sequencing

 Nicola Zilio  and Helle D. Ulrich 

Institute of Molecular Biology (IMB) gGmbH, Mainz, Germany

Keywords

DNA damage mapping; DNA single-strand breaks; genome stability; GLOE-Seq; next-generation sequencing; Nick-Seq; SSB-Seq; SSiNGLe

Correspondence

H. D. Ulrich, Institute of Molecular Biology (IMB) gGmbH, Ackermannweg 4, D – 55128 Mainz, Germany
 Tel: +49 6131 3921490
 E-mail: h.ulrich@imb-mainz.de

[Correction added on 17 December 2020, after first online publication: Peer review history is not available for this article, so the peer review history statement has been removed.]

(Received 17 July 2020, revised 7 September 2020, accepted 14 September 2020)

doi:10.1111/febs.15568

Mapping the genome-wide distribution of DNA lesions is key to understanding damage signalling and DNA repair in the context of genome and chromatin structure. Analytical tools based on high-throughput next-generation sequencing have revolutionized our progress with such investigations, and numerous methods are now available for various base lesions and modifications as well as for DNA double-strand breaks. Considering that single-strand breaks are by far the most common type of lesion and arise not only from exposure to exogenous DNA-damaging agents, but also as obligatory intermediates of DNA replication, recombination and repair, it is surprising that our insight into their genome-wide patterns, that is the ‘SSBreakome’, has remained rather obscure until recently, due to a lack of suitable mapping technology. Here we briefly review classical methods for analysing single-strand breaks and discuss and compare in detail a series of recently developed high-resolution approaches for the genome-wide mapping of these lesions, their advantages and limitations and how they have already provided valuable insight into the impact of this type of damage on the genome.

Introduction

A human cell experiences ~ 70 000 DNA lesions per day [1]. Among these, single-strand breaks (SSBs), that is discontinuities in one of the two phosphodiester backbones of the DNA duplex, are by far the most frequent, amounting to ~ 75%. SSBs can assume the

form of nicks or gaps, depending on whether their flanking nucleotides are adjacent or not, and the corresponding 5' and 3' termini may carry hydroxy (OH) groups, phosphate (PO₄) or other structures [2,3]. SSBs most frequently originate from oxidative attack by

Abbreviations

AOA1, ataxia oculomotor apraxia; ChIP, chromatin immunoprecipitation; dA, deoxyadenosine; dATP, deoxyadenosine triphosphate; dC, deoxycytidine; dG, deoxyguanosine; DNA, deoxyribonucleic acid; dNTP, deoxyribonucleotide triphosphate; DSB, double-strand break; dT, deoxythymidine; dUTP, deoxyuridine triphosphate; GLOE-Seq, genome-wide ligation of 3'-OH ends and sequencing; ILM, Illumina; MNase, micrococcal nuclease; NGS, next-generation sequencing; nt, nucleotide; OH, hydroxy; PCR, polymerase chain reaction; PLA, proximity ligation assay; PO₄, phosphate; RNA, ribonucleic acid; RT, reverse transcriptase; SCAN1, spinocerebellar ataxia with axonal neuropathy; Seq, sequencing; SMS, single-molecule sequencing; SSB, single-strand break; SSB-Seq, single-strand break sequencing; SSiNGLe, single-strand break mapping at nucleotide genome level; STRIDE, sensitive recognition of individual DNA ends; TdT, terminal deoxynucleotidyl transferase; TUNEL, TdT-mediated dUTP-biotin nick end labelling; UMI, unique molecular identifier; UV, ultraviolet.

reactive oxygen species, a common by-product of metabolism, or from disintegration of oxidized sugars [2,3]. In addition, they arise indirectly as intermediates of base or nucleotide excision repair pathways, during discontinuous replication of the lagging strand [4,5] or due to abortive activity of topoisomerases [2,3]. If left unrepaired, SSBs can have serious consequences for genome stability and cell survival. By collision with replisomes, SSBs can stall or collapse replication forks or – in the worst-case scenario – be converted into double-strand breaks (DSBs) [6]. Even on their own, SSBs can cause havoc: they interfere with transcription by inhibiting RNA polymerase progression, they can initiate mutagenesis, apoptosis and senescence, and they induce the expression of inflammatory cytokines [7–11]. The importance of dealing with SSBs is underscored by the existence of dedicated repair pathways and of hereditary diseases resulting from misregulation of these pathways. SSB-associated illnesses, such as spinocerebellar ataxia with axonal neuropathy (SCAN1) or ataxia oculomotor apraxia (AOA1), are mostly of a neurodegenerative nature [2,3]. However, both cancer and cardiac disease have also been linked to SSB repair defects [9,10].

Pivotal for understanding the genesis and the repair of any DNA lesions as well as their implications for genome stability are tools for their detection and analysis. Beyond methods for the identification of particular types of lesion, being able to map their positions with high resolution and to accurately quantify their incidence in specific regions of the genome under defined conditions, for example in response to specific environmental stresses or at certain developmental stages, in different cell types or organisms, could provide profound insight into a variety of fundamental biological processes in health and disease. This type of information could also be extremely valuable in other, more applied fields of biomedicine, such as genome engineering, as well as toxicology and biomarker development, where it could help better understand cancer development and treatment.

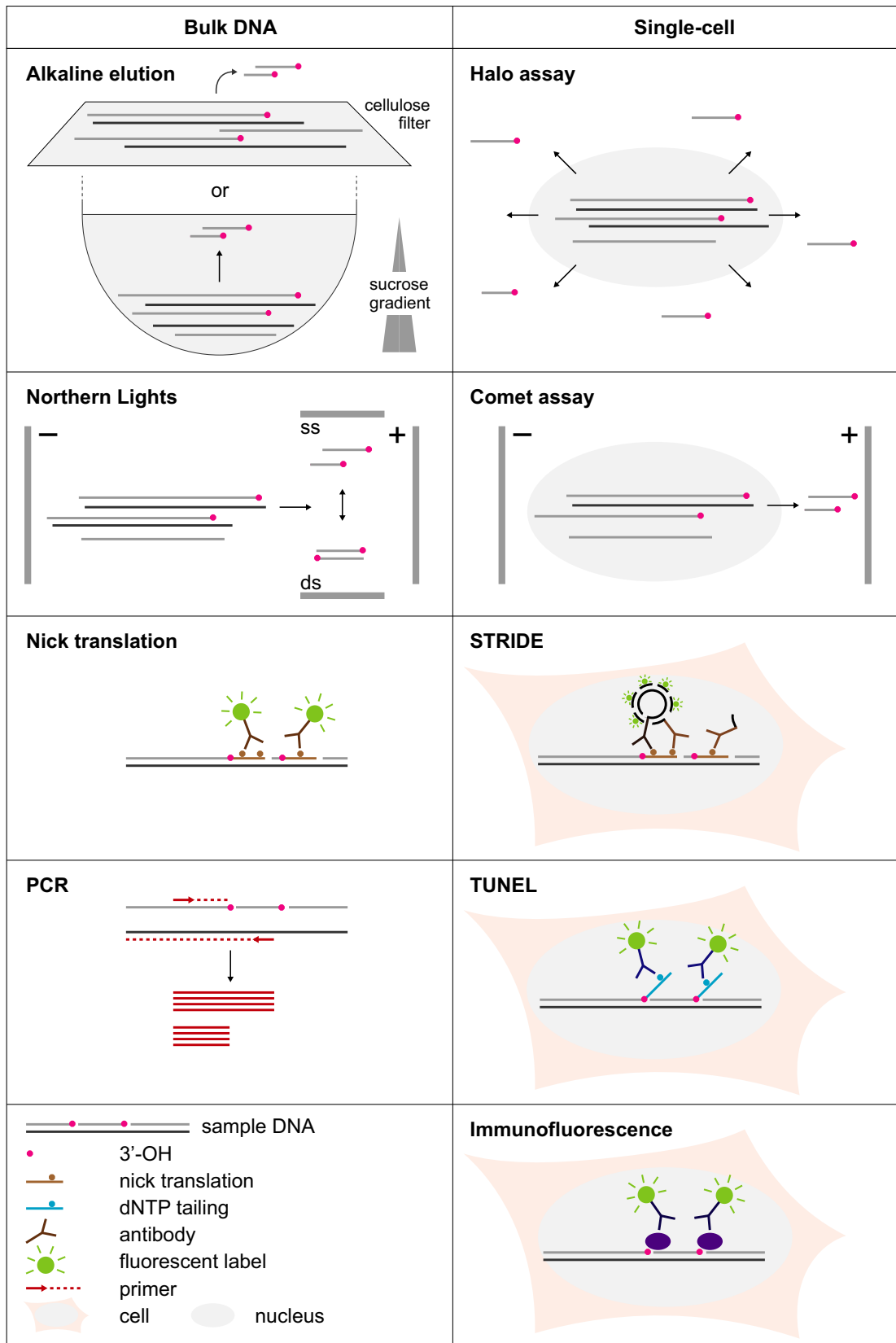
Traditionally, the detection of DNA damage has involved either radioactive labelling of bulk DNA or microscopic visualization in individual cells by means of fluorescent probes (Fig. 1). Although powerful for overall quantification, these methods suffer from several shortcomings, such as low resolution and throughput. The advent of next-generation sequencing (NGS) has revolutionized genome stability research by allowing the genome-wide mapping and quantification of DNA damage at an unprecedented resolution. Although several NGS methods are available to map various kinds of DNA damage, such as abasic sites, oxidative or methylation damage, incorporated ribonucleotides and DSBs,

until very recently methods for accurately mapping SSBs were missing. Here, we briefly review common ‘classical’ methods for detecting SSBs before discussing in detail the development of NGS-based tools for this important lesion, their respective advantages and limitations as well as their potential for future development.

‘Classical’ methods to detect and quantify SSBs

Early methods to detect SSBs involved the elution of radioactively labelled DNA from cellulose membranes or sedimentation through a sucrose gradient under denaturing conditions to liberate small fragments arising from the presence of SSBs [12]. An alternative bulk technique currently in use is nick translation, where a DNA polymerase harbouring 5′–3′-exonucleolytic activity is used to incorporate labelled nucleotides at templated 3′-OH groups, thus ‘translating’ the position of the original break [13]. On the single-cell level, the same principle is exploited by the highly sensitive STRIDE (SensiTive Recognition of Individual DNA Ends) assay [14], which combines *in situ* nick translation with the proximity ligation assay (PLA) to detect either SSBs (sSTRIDE) or DSBs (dSTRIDE). Here, lesions labelled by nick translation with biotinylated nucleotides are detected by a PLA signal originating from two anti-biotin antibodies from different species. The strategy of *in situ* labelling of lesions also applies to the more established TUNEL (Terminal deoxynucleotidyl transferase dUTP Nick End Labelling) assay, originally developed to identify apoptotic cells [15] and capable of detecting both SSBs and DSBs. Here, labelled nucleotides are added enzymatically to 3′-OH termini in a nontemplated manner and subsequently detected by fluorescence microscopy.

Best known among single-cell SSB assays is the comet assay [16,17], which uses an electrical field to extrude damaged DNA out of the nuclei (‘heads’) of agarose-embedded cells, thereby producing a ‘tail’. The length of this tail and the fraction of DNA in it, visualized by a relevant dye, are used to calculate the tail moment [18], which positively correlates with the amount of DNA damage. Like the TUNEL assay, the comet assay detects both DSBs and SSBs. Usually, neutral versus alkaline conditions are used to distinguish DSBs from SSBs; however, SSBs can apparently also be detected under neutral conditions [19,20]. The traditional comet assay has been adapted to two-dimensional (2D) gel electrophoresis in order to distinguish the signals from SSBs and DSBs (two-tailed comet assay, 2T-Comet [21,22]). When used on bulk purified DNA, separation of DNA by length, strandness,



structure and shape via 2D gel electrophoresis is known as the Northern Lights assay [23,24]. Finally, a simplified variation of the comet assay is the halo assay, where damaged DNA is liberated from embedded nuclei by passive diffusion, thus forming a ‘halo’ whose area is proportional to the amount of damage [25,26].

As an alternative to detecting DNA lesions directly, dedicated DNA repair factors or specific post-translational modifications, detected via fluorescence microscopy, are widely used as sensitive damage reporters. Accordingly, localization of a core SSB repair factor, XRCC1, has allowed for monitoring SSBs [27–31]. Obvious limitations to this approach are common problems associated with chromatin accessibility and the staining procedure, a dependence on the repair process itself, a possible accumulation of DNA damage makers at nondamaged sites [32,33], and the uncertainty of how well the position and intensity of a repair focus correlate to the properties and number of the causative lesion(s) [29,34].

Perhaps the most serious drawback of these classical methods is their inability to assign the position of a lesion in the nucleus to a genomic locus with an accuracy that goes beyond a crude differentiation between hetero- and euchromatin or nuclear compartments. Combination with fluorescence in situ hybridization [35–37] or quantitative PCR approaches that exploit the interference of lesions with efficient amplification [38–41] can in principle overcome this limitation; however, both techniques realistically restrict analysis to a small number of loci, and PCR-based methods are insensitive to the type of lesion interfering with amplification.

NGS-based methods for genome-wide mapping of SSBs

The first genome-wide approaches for mapping DNA damage predate NGS technology, as they relied – like RNA analyses and chromatin immunoprecipitation (ChIP) – on the use of microarrays for sequence identification [42–44]. However, the advent of commercial high-throughput sequencers has transformed the ways in which DNA damage can be analysed. On the basis of this

disruptive technology, a multitude of creative new protocols have been developed that now provide the means of identifying various types of lesions with accurate quantification at single nucleotide resolution [45,46].

Double-strand breaks were the first lesions to be targeted for genome-wide mapping [43,44,47], and many NGS-based techniques are available today for this type of damage [48–58]. Most of them rely on ‘polishing’ DSBs to generate blunt ends for ligation to suitable sequencing adaptors. Mis-incorporated ribonucleotides, alkylated bases and UV-induced pyrimidine dimers have been mapped via cleavage by a damage-specific endonuclease and capture of the resulting ends [59–64]. Some base lesions, such as UV damage, cisplatin and benzo[a]pyrene DNA adducts can be affinity-enriched using specific antibodies, lesion-binding proteins or DNA damage markers [42,65–68] and chemical labelling with a specific probe has proven useful for capturing abasic sites [69,70].

To date, four NGS-based methods have been developed for the genome-wide mapping of SSBs: SSB-Seq [71,72], Nick-Seq [73], SSiNGLe [74] and GLOE-Seq [75] (Table 1). Although the molecular steps involved in generating sequencing libraries vary, they all rely on the common principle of detecting SSBs by capturing free 3'-OH termini.

SSB-Seq

An early attempt at genome-wide SSB mapping made use of nick translation in the presence of digoxigenin-labelled dUTP to label 3'-OH termini for subsequent capture with an antibody, followed by standard dsDNA library preparation (Fig. 2A) [71,72]. Omission of the digoxigenin label served as a negative control in this approach. Although in principle feasible, SSB-Seq has not been thoroughly validated; therefore, no information about resolution, sensitivity or sequence bias is presently available. Although inclusion of a small amount of dideoxynucleotides to reduce the size of the labelled patches of DNA is used to increase mapping resolution, the method is unlikely to provide true nucleotide resolution, since the position

Fig. 1. Common methods for the detection and quantification of SSBs. Techniques based on bulk genomic DNA (left) or single cells or nuclei (right) are depicted schematically to illustrate the principles by which SSBs are detected. While alkaline elution and the halo assay exploit the increased mobility of small DNA fragments under denaturing conditions, the Northern Lights and the comet assay separate DNA fragments by means of an electric field. Like nick translation of isolated DNA, the cell-based STRIDE assay marks 3'-OH termini by polymerase-mediated extension with labelled nucleotides for detection by fluorescent antibodies or PLA with fluorescent oligonucleotides. Similarly, the TUNEL assay makes use of TdT for tailing 3'-OH ends with labelled nucleotides, and standard immunofluorescence detects proteins associated with SSBs. PCR can be used to analyse SSBs in defined sequences via the inhibitory effect of a break on the amplification of this template.

Table 1. Comparison of NGS methods for the genome-wide mapping of SSBs.

	SSB-Seq	Nick-Seq	SSINGLE	GLOE-Seq
Sample preparation				
Starting material	Purified DNA	Purified DNA	Isolated nuclei from <i>in situ</i> formaldehyde-crosslinked cells	Purified DNA or agarose-embedded nuclei
3'-labelling principle	Nick translation with digoxigenin-labelled dUTP	a Nick translation with α -thio-dNTPs b Poly(dT) tailing	Poly(dA) tailing	Ligation to biotinylated adaptor
Capturing principle	Capture via labelled nucleotides	a Degradation of unlabelled DNA b None	None	Capture via the ligated adaptor
Sequencing details				
Platform	Illumina	Illumina	SMS: Helicos/SeqLL ILM: Illumina	Illumina
Read length	36 nt (single-end)	75 nt (paired-end)	SMS: > 25 nt ILM: 150 nt (paired-end)	\geq 35 nt (single-end)
Data analysis				
Filtering & quality control	Removal of redundant reads	a Read 1 of the Illumina protocol b Read 2 of the Illumina protocol	SMS: filtering for length (\geq 25 nt) ILM: filtering for correct flanking sequences (C ₁₀ N _x T ₁₂)	Removal of low mapping quality reads
Analysis	Bowtie 2 Extension of reads to average length of library fragments	Galaxy (Trim Galore!, Bowtie 2, Bamtools, bedtools) Custom scripts to identify putative SSBs from the intersection of significant signals from (a) and (b) https://github.com/BoCaolab/Nick-seq	Both: removal of low mapping quality reads, signals near genomic regions with high poly(A) content and in untailed control samples Basic shell pipeline combining custom scripts for filtering, read identification and determination of break positions https://github.com/huifencao/DB-Nature_Communications-2019	Dedicated pipeline (GLOE-Pipe) for identification of break positions and determination of statistical significance https://github.com/helle-ulrich-lab/ngs-gloepipe
Signal definition	Normalized read pileup, based on average track length of sequenced fragments	Normalized read count ratios (to neighbouring nt and control)	Number of reads initiating opposite each genomic position	Normalized number of reads initiating opposite each genomic position
Features and applications				
Validation	None	<i>Escherichia coli</i> DNA digested with nicking endonucleases	Human DNA digested with nicking endonuclease	<i>Saccharomyces cerevisiae</i> DNA digested with nicking and restriction endonucleases
Reported applications	Etoposide-induced breaks (human cells)	Phosphorothioate modification (<i>Salmonella enterica</i> serovar Cerro 87) Peroxide-induced termini (<i>E. coli</i>)	SSBs (human cells) DSBs (human cells) Peroxide-induced termini (human cells) Mitochondrial 7S DNA (human cells)	SSBs (yeast, human cells) DSBs (yeast) Pyrimidine dimers (yeast) m ⁷ dG (yeast) Okazaki fragments (yeast, human cells)

Table 1. (Continued).

	SSB-Seq	Nick-Seq	SSiNGLe	GLOE-Seq
Resolution	Ca. 50–250 nt	Single nt	Single or few nt	Single nt
Strand information	No	Yes	Yes	Yes
Detection of DSBs	No	Yes	Yes	Yes
Advantages	Potential to detect rare events	High confidence signals from combination of (a) and (b)	SMS: no need for amplification	Validated application to many different structures
Limitations	Not fully validated	Not suited for pre-existing SSBs; Requires a certain signal penetrance for detection	Not suited for 3'-PO ₄ termini; Unable to accurately map SSBs in genomic regions with high poly(dA) content	Not suited for SSBs spaced at < 100 nt

of the lesion is lost during the random fragmentation and end-polishing steps involved in library preparation. Nevertheless, SSB-Seq has the potential to detect very rare events, since the degree of labelling could in principle be increased, at the expense of resolution, by extending the nick translation reaction.

Nick-Seq

Nick-Seq is perhaps a misnomer for a method designed to detect base lesions or modifications rather than SSBs (Fig. 2B) [73]. In fact, pre-existing 3'-OH ends are initially blocked by appending dideoxynucleotides. The lesions of interest are then converted enzymatically or chemically into clean 3'-OH termini and further processed in parallel by two complementary strategies. One portion of the DNA is subjected to nick translation using α -thio-dNTPs, which, upon incorporation into DNA, form a nuclease-resistant phosphorothioate bond. Nonlabelled DNA is removed with a cocktail of nucleases and the remaining phosphorothioate-containing polynucleotides are converted into sequencing libraries. The other portion is treated with terminal deoxynucleotidyl transferase (TdT) to add poly(dT) tails, which are then used to anneal oligo(dA) primers for the synthesis of cDNA libraries by reverse transcriptase. High confidence signals are identified by combining the results from both libraries. Thus, while each of the two strategies suffers from limitations such as high background (for nick translation) or poor resolution (for poly(dT) tailing), in combination they are capable of producing high-quality single nucleotide resolution maps of DNA damage. Although certainly powerful, one potential limitation of this two-pronged tactic is that it inherently requires a certain amount of penetrance for a lesion at a particular site to be detected by both procedures. Nick-Seq's biggest current limitation, its inability to detect pre-existing nicks, could potentially be overcome by simply omitting the initial blocking step as well as any further chemical or enzymatic treatment of the DNA, to map naturally occurring strand breaks.

SSiNGLe

SSiNGLe (Single-strand break mapping at Nucleotide Genome Level) is the first NGS-based method purposely developed and properly validated for mapping SSBs (Fig. 2C) [74]. However, like Nick-Seq and GLOE-Seq, it could be easily adapted to any lesion that can be converted, enzymatically or chemically, to a nick with a free 3'-OH group. In SSiNGLe, cells are initially crosslinked *in situ* with formaldehyde and

processed to isolate nuclei. Next, high-molecular-weight DNA is fragmented to a size compatible with NGS using micrococcal nuclease (MNase), which generates 3'-PO₄ termini. After deproteinization and de-crosslinking, TdT in the presence of dATP serves to add poly(dA) tails selectively to all free 3'-OH termini. In this manner, biologically significant SSBs are differentiated from the 3'-PO₄-protected MNase-derived breaks and can be directly subjected to amplification-free sequencing on a Helicos/SeqLL sequencer, which uses flow cells coated with oligo(dT) (SSiNGLe-SMS). Because of the limited availability of this platform, the authors also adapted their core library workflow for the more commonly used Illumina system (SSiNGLe-ILM) by limited linear amplification with an oligo(dT) primer and subsequent poly(dC) tailing of the 3' ends of the newly generated DNA. Unwanted poly(dC) tailing of any remaining oligo(dT) primer is prevented by including three thymidine ribonucleotides at the 3' end of this primer, which cannot be extended by TdT. The resulting oligo(dT)/poly(dC)-tagged molecules are PCR-amplified using primers that include oligo(dT) and oligo(dG) tracts at their 3' ends, respectively, and relevant Illumina-compatible sequences at their 5' extremities. Once mapped, the 3' end of a read marks the original position of break.

Leveraging the unique features of the Helicos/SeqLL platform to directly sequence the captured molecules at their original break sites is an elegant approach because it obviates the need for amplification and therefore avoids the problem of 'PCR duplicates'. These are identical reads derived from amplification of a single molecule, which, in the context of break mapping, can complicate quantification or distort the original representation of sequences [76–78]. Direct comparison of SSiNGLe-SMS and SSiNGLe-ILM on the same input DNA yielded comparable results, indicating that amplification bias is unlikely to be problematic in SSiNGLe-ILM. Nevertheless, accessibility to the Helicos/SeqLL platform is at present very limited; thus, the Illumina-based protocol is expected to become more widely used, despite a few clear limitations.

While introducing unique molecular identifier (UMI) sequences into the oligo(dT) primer would allow the elimination of PCR duplicates [79], the presence of homopolymeric runs of nucleotides at both ends of each read further complicates analysis on the Illumina system, as accurate base calling requires all four bases to be represented approximately evenly, especially during the first few sequencing cycles. Employing patterned flow cells could help; alternatively, cycles without imaging could be used to bypass the homopolymers, or libraries could be diluted by spiking in a whole-genome library, such as one produced from the phi X174 genome, at the expense of relevant sequencing output. Another caveat concerns the use of MNase for fragmentation, which produces unligatable 3'-PO₄ termini and thus cannot be used in situations where this type of modification is of interest. After all, 3'-PO₄, together with 3'-phosphoglycolate, represent the majority of direct SSBs inflicted by reactive oxygen species [2]. Treatment with a phosphatase before poly(dA) tailing of 3' ends and replacing the MNase treatment with a mechanical fragmentation step after the tailing reaction would overcome this problem. Finally, SSiNGLe suffers from the same inherent limitation as other poly(dA) tailing-based methods, that is the inability to accurately map SSBs adjacent to one or more dA residues.

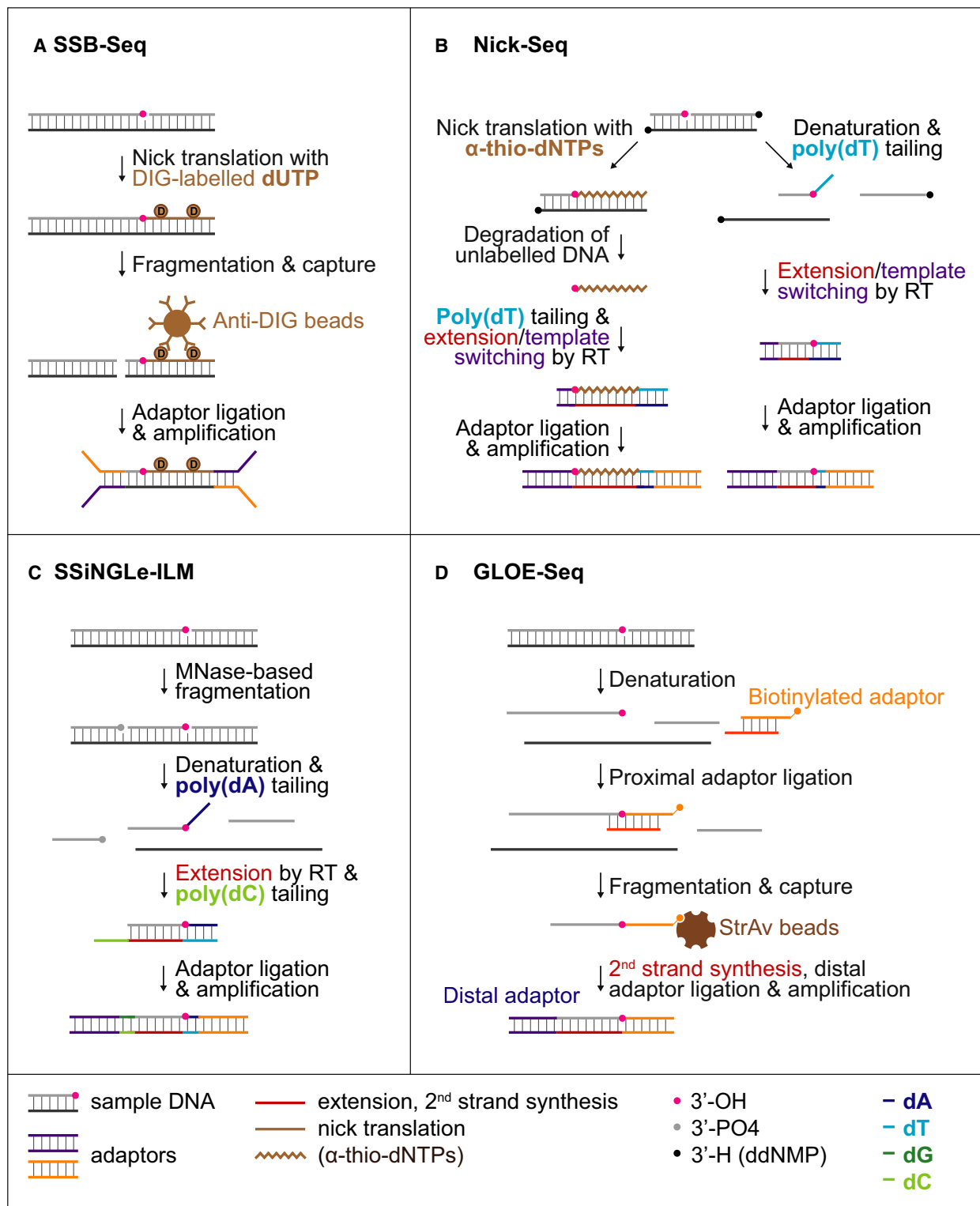
GLOE-Seq

GLOE-Seq (Genome-wide Ligation of 3'-OH Ends) has been developed to specifically map pre-existing 3'-OH groups but has also been validated for a number of base lesions (Fig. 2D) [75]. It employs tried-and-tested steps from protocols originally designed for mapping modified bases [80], but combines them in a unique fashion and optimizes the entire workflow to efficiently capture and recover SSBs. Genomic DNA is initially heat-denatured to make free 3'-OH termini accessible for ligation to a proximal adaptor that bears a random single-stranded hexanucleotide overhang at one end and a biotin moiety at the other. After removal of excess unligated proximal adaptor, the ligated DNA is

Fig. 2. NGS methods for the genome-wide mapping of SSBs. (A) SSB-Seq makes use of nick translation to label and subsequently capture sequences terminating with a 3'-OH group. (B) Nick-Seq exploits a combination of nick translation (left) and poly(dT) tailing (right) to identify 3'-OH termini. The need for capturing the labelled DNA is overcome by using nuclease-resistant nucleotide analogues for nick translation (left) and selective reverse transcriptase (RT)-mediated template switching followed by amplification of the tailed DNA (right). (C) SSiNGLe relies on poly(dA) tailing of 3'-OH termini. Conversion to Illumina (ILM)-compatible libraries involves additional tailing and ligation of specific adaptors. (D) GLOE-Seq (scheme adapted from Ref. [75]) is based on the direct ligation of 3'-OH termini to an adaptor containing an overhang of 6 random nucleotides, followed by capture, second-strand synthesis and amplification of the captured DNA.

fragmented to 200–300 nt and captured on streptavidin beads. The purified ssDNA is converted to dsDNA and then ligated to a second, distal, adaptor, which allows

its amplification by PCR in preparation for Illumina sequencing. After mapping, the 5' end of a read pinpoints the original location of a 3'-OH.



GLOE-Seq's biggest advantage is that 3'-OH ends are directly captured in preparations of total, intact genomic DNA and fragmentation is delayed until after ligation to the biotinylated adaptor. This ensures that very little unwanted 'technical' background is introduced due to handling the DNA during library preparation. Paradoxically, however, this virtue has also revealed a problematic feature inherent in all methods based on 3' end capturing: GLOE-Seq efficiently detects what appears to be a biologically meaningful, and therefore unavoidable, 'background' pattern of SSBs (see below).

Like SSiNGLe, and most other sequencing workflows based on the Illumina platform, GLOE-Seq in its current form cannot determine whether identical reads represent PCR duplicates or originated from two independent breaks at the same location. However, unlike in SSiNGLe, this limitation can be circumvented in GLOE-Seq by introducing UMIs in either the proximal or the distal adaptor.

GLOE-Seq also has a size limitation: the need to remove unincorporated oligonucleotides/adaptors of various stages of library preparation translates into poor recovery of DNA fragments smaller than 100 nt. This means that multiple events in close proximity cannot be efficiently detected. Whether SSiNGLe or Nick-Seq handle closely spaced events any better has not been reported.

Applications and implications

SSiNGLe and GLOE-Seq have provided us with a first glimpse into the previously unknown genome-wide distribution of spontaneous SSBs in both yeast and human cells. Unexpected biases and asymmetries in the cellular SSBreakome revealed in this manner underscore the importance of investigating the underlying mechanisms responsible for SSB formation and processing.

First and foremost, it has become clear that SSBs are not distributed randomly in the genome. In unperturbed yeast cells, SSBs are highly enriched around centromeres and in well-defined peaks within pericentromeric areas [75]. The former may well represent sites of topological stress or topoisomerase activity, but the origin of the latter is unclear and deserves further investigation. In human cultured cells, GLOE-Seq, which detects the 3' ends of both SSBs and DSBs, has revealed a transcription-dependent depletion of breaks around transcription start sites, a pattern contrary to that of DSBs as detected by END-Seq [81] and illustrating the predominance of SSBs over DSBs even at loci where DSBs accumulate. In contrast,

SSiNGLe analysis reports the strongest overlap of SSBs with promoters within coding regions [74], a finding echoed by SSB-Seq data [71]. The reason for this discrepancy is unclear but may be related to the use of different cell types. SSiNGLe also revealed an enrichment of breaks at exons and other transcriptional regulatory elements, active histone marks and CTCF binding sites, but not introns. The difference between introns and exons, identified by GLOE-Seq only for apparently untranscribed genes, is intriguing and should be further pursued.

Other correlations between the SSiNGLe SSBreakome and functionally relevant genomic regions include an enrichment of breaks at early replicating regions. Fork stalling and collapse at poly(dA:dT) loci often found adjacent to replication origins [82] could explain this phenomenon. In this case, SSiNGLe would likely underestimate the incidence of breaks because of its problems with reads within poly(dA) stretches. Breaks were also found to correlate with satellite repeats, topoisomerase IIA cleavage clusters, SNPs and evolutionarily less conserved loci, which would be consistent with the mutagenic nature of DNA damage [83,84]. Remarkably, the association of breaks with sequence variants is not constant throughout the genome. For example, a depletion within exons suggests that different parts of the genome experience different levels of DNA damage and/or repair.

Perhaps the most intriguing finding presented by SSiNGLe is that specific biological states, for example cell type, particular treatments or even age of the organism, appear to be associated with unique, identifiable genome-wide SSB patterns. Not only will this information provide new insight into disease and development, but if correlations between the specific SSBreakome of a diseased state, such as cancer, and its response to a drug can be identified, the use of such patterns as biomarkers is worth exploring further.

GLOE-Seq data have raised the intriguing possibility that nicks or gaps formed during excision of ribonucleotides mis-incorporated in DNA might be a significant source of 'spontaneous' SSBs. In unperturbed yeast and human cells, the incidence of SSBs follows the activity of the leading strand-specific DNA polymerase ϵ . This bias is unexpected given that SSBs should emerge on the lagging strand with a considerably higher frequency than on the leading strand, due to its discontinuous synthesis. Accordingly, GLOE-Seq in combination with depletion of the replicative DNA ligase I has allowed the reliable mapping of replication origins by means of a lagging strand bias of SSBs. However, polymerase ϵ mis-incorporates ribonucleotides about four times as frequently as the lagging strand-

specific DNA polymerase δ [85]. The leading strand bias in ligase-competent cells could therefore suggest that mis-incorporated ribonucleotides are processed significantly more slowly than unligated Okazaki fragments. Intriguingly, an alternative explanation for this bias has been proposed in a very recent report that made use of yet another 3'-end capturing strategy, called TrAEL-seq, designed specifically to map 3'-overhangs of resected DSBs [53]. Here, TdT is used to append a short tail of 1–3 (ribo-)adenosine nucleotides that is ligatable to a suitable adaptor by a specialized RNA ligase. Since TdT requires at least 3 nucleotides of ssDNA [86], TrAEL-seq should capture protruding single-stranded 3'-ends with much better efficiency than SSBs with annealed termini. Nevertheless, application of the procedure to undamaged DNA produced an even stronger lagging strand bias than GLOE-Seq, which led the authors to propose that partially or fully reversed replication forks could be responsible for the pattern.

These findings highlight an inherent, and unfortunately most likely insurmountable, limitation of all methods designed to capture DNA ends. They imply that the mapping of any SSBs, whether arising from the action of genotoxins or generated as intermediates of DNA repair, recombination or replication, has to be performed against a significant, possibly biologically meaningful 'background'.

Future developments

With a handful of techniques released in just over a year, the field of SSB mapping is still in its infancy, leaving significant scope for further improvements.

For example, methods for absolute quantification of breaks will need to be implemented. As a first step, UMIs – although not without caveats [87] – would help eliminate reads derived from PCR duplicates and therefore allow an estimation of the 'true' number of breaks per cell [57]. PCR-derived biases and distortions [78] could be further reduced by linear or quasi-linear amplification methods [88–91]. An interesting alternative to UMIs, developed specifically for DSBs but easily adaptable to SSBs, uses 'spike-in' DSBs, generated by predigesting samples with a rare-cutting restriction enzyme, to quantify biologically relevant DSBs via comparison [87].

Sensitivity of detection and minimization down to the single-cell level are important issues to be tackled, as cell-to-cell variation can have significant consequences for the behaviour of a population [92,93] and should therefore be taken into account when assessing the SSBreakome in the context of diseases such as cancer or of developmental processes.

Complementary to the approach of mapping 3'-OH ends, developing corresponding procedures for capturing 5' termini would be a valuable addition to the current toolbox, particularly on a single-cell scale. For termini on the same strand, simultaneous mapping of both ends would be useful for investigating Okazaki fragments or postreplicative daughter-strand gaps [94–96]. When applied to DSBs, this strategy has the potential to access untapped information about the processing of such lesions. So far, all protocols developed for DSBs suffer from one major drawback: in preparation for DSB capture, the 3'-ssDNA overhangs generated by the cell during the resection process are removed to produce blunt ends compatible with adaptor ligation, thus effectively obfuscating the original position of the break. Individual capture of both ends would overcome this limitation. Even the current GLOE-Seq method has been used to reveal exposed 3' overhangs of genomic DSBs. In combination with DSB-specific protocols or by carefully avoiding denaturation of the genomic DNA, a differentiation between SSBs and (resected) DSBs should thus become possible. Adaptation to the single-cell scale could then provide unprecedented insight into matched pairs of 5' and 3' termini.

Last but not least, advances in third-generation sequencing platforms, such as nanopore-based sequencing (e.g. Oxford Nanopore) [97], promises new opportunities for simplified SSB mapping. Following heat denaturation of genomic DNA and one-step ligation of relevant adaptor–motor enzyme complexes for nanopore sequencing, protocols for the sequencing of ssDNA fragments [98] could potentially reveal the positions of 5' and 3' extremities resulting from SSBs directly and without PCR amplification, and even provide information about their absolute frequency.

Conclusion

The ever-increasing number of NGS-based tools testifies to the notion that applications of high-throughput sequencing have by now well outgrown their original purpose of assembling new genomes or detecting sequence variants. Compared to such staple technology as RNA-Seq and CHIP-Seq, the development of NGS methods designed to detect DNA damage at single nucleotide resolution has somewhat lagged behind. This is particularly true for SSBs, which, unlike DSBs, have only very recently been tackled productively – partly due to technical challenges involved in capturing such lesions, but possibly also because they are often perceived to be less dangerous. SSBs arguably pose less of a risk to a cell's survival than DSBs. However, arising both from the direct action of several genotoxins and as

common intermediates of most DNA transactions, they are no less biologically meaningful and capable of shaping the genome, not only in evolution but also in cancer development. The value of reliable techniques for genome-wide mapping of SSBs is clear: they have revealed intriguing physiological patterns of SSBs along the genome that have already raised a multitude of questions and should pave the way for a new age of studying the underlying molecular processes of genome maintenance.

Acknowledgements

Funding was provided by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation; project number 393547839 – SFB 1361, subproject 07). The funding source had no involvement in the writing of the report or the decision to submit the article for publication. Open access funding enabled and organized by Projekt DEAL

Conflict of interest

The authors declare no conflict of interest.

Author contributions

NZ and HDU jointly conceived, wrote and edited this manuscript.

References

- Lindahl T & Barnes DE (2000) Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127–133.
- Caldecott KW (2008) Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619–631.
- Hossain MA, Lin Y & Yan S (2018) Single-strand break end resection in genome integrity: mechanism and regulation by APE2. *Int J Mol Sci* **19**, 2389.
- Hanzlikova H, Kalasova I, Demin AA, Pennicott LE, Cihlarova Z & Caldecott KW (2018) The importance of poly(ADP-ribose) polymerase as a sensor of unligated Okazaki fragments during DNA replication. *Mol Cell* **71**, 319–331.
- Hanzlikova H & Caldecott KW (2019) Perspectives on PARPs in S phase. *Trends Genet* **35**, 412–422.
- Branzei D & Foiani M (2008) Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**, 297–308.
- Tann AW, Boldogh I, Meiss G, Qian W, Van Houten B, Mitra S & Szczesny B (2011) Apoptosis induced by persistent single-strand breaks in mitochondrial genome: critical role of EXOG (5'-EXO/endonuclease) in their repair. *J Biol Chem* **286**, 31975–31983.
- Schiml S, Fauser F & Puchta H (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *Proc Natl Acad Sci USA* **113**, 7266–7271.
- Nassour J & Abbadie C (2016) A novel role for DNA single-strand breaks in senescence and neoplastic escape of epithelial cells. *Mol Cell Oncol* **3**, e1190885.
- Higo T, Naito AT, Sumida T, Shibamoto M, Okada K, Nomura S, Nakagawa A, Yamaguchi T, Sakai T, Hashimoto A *et al.* (2017) DNA single-strand break-induced DNA damage response causes heart failure. *Nat Commun* **8**, 15104.
- Kathe SD, Shen GP & Wallace SS (2004) Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *J Biol Chem* **279**, 18511–18520.
- Crump PW, Fielden EM, Jenner TJ & O'Neill P (1990) A comparison of the techniques of alkaline filter elution and alkaline sucrose sedimentation used to assess DNA damage induced by 2-nitroimidazoles. *Biochem Pharmacol* **40**, 621–627.
- Rigby PW, Dieckmann M, Rhodes C & Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J Mol Biol* **113**, 237–251.
- Kordon MM, Zarebski M, Solarczyk K, Ma H, Pederson T & Dobrucki JW (2020) STRIDE-a fluorescence method for direct, specific in situ detection of individual single- or double-strand DNA breaks in fixed cells. *Nucleic Acids Res* **48**, e14.
- Gavrieli Y, Sherman Y & Ben-Sasson SA (1992) Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J Cell Biol* **119**, 493–501.
- Langie SA, Azqueta A & Collins AR (2015) The comet assay: past, present, and future. *Front Genet* **6**, 266.
- Ostling O & Johanson KJ (1984) Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem Biophys Res Commun* **123**, 291–298.
- Olive PL, Banath JP & Durand RE (1990) Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells measured using the “comet” assay. *Radiat Res* **122**, 86–94.
- Collins AR, Oscoz AA, Brunborg G, Gaivao I, Giovannelli L, Kruszewski M, Smith CC & Stetina R (2008) The comet assay: topical issues. *Mutagenesis* **23**, 143–151.
- Azqueta A & Collins AR (2013) The essential comet assay: a comprehensive guide to measuring DNA damage and repair. *Arch Toxicol* **87**, 949–968.
- Cortes-Gutierrez EI, Fernandez JL, Davila-Rodriguez MI, Lopez-Fernandez C & Gosalvez J (2017) Two-tailed comet assay (2T-Comet): simultaneous detection

- of DNA single and double strand breaks. *Methods Mol Biol* **1560**, 285–293.
- 22 Enciso M, Sarasa J, Agarwal A, Fernandez JL & Gosalvez J (2009) A two-tailed Comet assay for assessing DNA damage in spermatozoa. *Reprod Biomed Online* **18**, 609–616.
 - 23 Gudmundsson B, Thormar HG, Sigurdsson A, Dankers W, Steinarsdottir M, Hermanowicz S, Sigurdsson S, Olafsson D, Halldorsdottir AM, Meyn S *et al.* (2018) Northern lights assay: a versatile method for comprehensive detection of DNA damage. *Nucleic Acids Res* **46**, e118.
 - 24 Gunnarsson GH, Gudmundsson B, Thormar HG, Alfredsson A & Jonsson JJ (2006) Two-dimensional strandness-dependent electrophoresis: a method to characterize single-stranded DNA, double-stranded DNA, and RNA-DNA hybrids in complex samples. *Anal Biochem* **350**, 120–127.
 - 25 Sestili P, Martinelli C & Stocchi V (2006) The fast halo assay: an improved method to quantify genomic DNA strand breakage at the single-cell level. *Mutat Res* **607**, 205–214.
 - 26 Sestili P & Cantoni O (1999) Osmotically driven radial diffusion of single-stranded DNA fragments on an agarose bed as a convenient measure of DNA strand scission. *Free Radic Biol Med* **26**, 1019–1026.
 - 27 Solarczyk KJ, Kordon M, Berniak K & Dobrucki JW (2016) Two stages of XRCC1 recruitment and two classes of XRCC1 foci formed in response to low level DNA damage induced by visible light, or stress triggered by heat shock. *DNA Repair (Amst)* **37**, 12–21.
 - 28 Campalans A, Kortulewski T, Amouroux R, Menoni H, Vermeulen W & Radicella JP (2013) Distinct spatiotemporal patterns and PARP dependence of XRCC1 recruitment to single-strand break and base excision repair. *Nucleic Acids Res* **41**, 3115–3129.
 - 29 Kordon MM, Szczurek A, Berniak K, Szelest O, Solarczyk K, Tworzydło M, Wachsmann-Hogiu S, Vaahtokari A, Cremer C, Pederson T *et al.* (2019) PML-like subnuclear bodies, containing XRCC1, juxtaposed to DNA replication-based single-strand breaks. *FASEB J* **33**, 2301–2313.
 - 30 Lan L, Nakajima S, Oohata Y, Takao M, Okano S, Masutani M, Wilson SH & Yasui A (2004) In situ analysis of repair processes for oxidative DNA damage in mammalian cells. *Proc Natl Acad Sci USA* **101**, 13738–13743.
 - 31 Abdou I, Poirier GG, Hendzel MJ & Weinfeld M (2015) DNA ligase III acts as a DNA strand break sensor in the cellular orchestration of DNA strand break repair. *Nucleic Acids Res* **43**, 875–892.
 - 32 Ziani S, Nagy Z, Alekseev S, Soutoglou E, Egly JM & Coin F (2014) Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J Cell Biol* **206**, 589–598.
 - 33 Bekker-Jensen S, Lukas C, Melander F, Bartek J & Lukas J (2005) Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *J Cell Biol* **170**, 201–211.
 - 34 Lisby M, Mortensen UH & Rothstein R (2003) Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat Cell Biol* **5**, 572–577.
 - 35 Martin M, Terradas M, Hernandez L & Genesca A (2014) gammaH2AX foci on apparently intact mitotic chromosomes: not signatures of misrejoining events but signals of unresolved DNA damage. *Cell Cycle* **13**, 3026–3036.
 - 36 Shaposhnikov S, Thomsen PD & Collins AR (2011) Combining fluorescent in situ hybridization with the comet assay for targeted examination of DNA damage and repair. *Methods Mol Biol* **682**, 115–132.
 - 37 Spivak G (2010) The Comet-FISH assay for the analysis of DNA damage and repair. *Methods Mol Biol* **659**, 129–145.
 - 38 Zhu S & Coffman JA (2017) Simple and fast quantification of DNA damage by real-time PCR, and its application to nuclear and mitochondrial DNA from multiple tissues of aging zebrafish. *BMC Res Notes* **10**, 269.
 - 39 Furda A, Santos JH, Meyer JN & Van Houten B (2014) Quantitative PCR-based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. *Methods Mol Biol* **1105**, 419–437.
 - 40 Figueroa-Gonzalez G & Perez-Plasencia C (2017) Strategies for the evaluation of DNA damage and repair mechanisms in cancer. *Oncol Lett* **13**, 3982–3988.
 - 41 Lehle S, Hildebrand DG, Merz B, Malak PN, Becker MS, Schmezer P, Essmann F, Schulze-Osthoff K & Rothfuss O (2014) LORD-Q: a long-run real-time PCR-based DNA-damage quantification method for nuclear and mitochondrial genome analysis. *Nucleic Acids Res* **42**, e41.
 - 42 Teng Y, Bennett M, Evans KE, Zhuang-Jackson H, Higgs A, Reed SH & Waters R (2011) A novel method for the genome-wide high resolution analysis of DNA damage. *Nucleic Acids Res* **39**, e10.
 - 43 Tchurikov NA, Kretova OV, Sosin DV, Zykov IA, Zhimulev IF & Kravatsky YV (2011) Genome-wide profiling of forum domains in *Drosophila melanogaster*. *Nucleic Acids Res* **39**, 3667–3685.
 - 44 Feng W, Di Rienzi SC, Raghuraman MK & Brewer BJ (2011) Replication stress-induced chromosome breakage is correlated with replication fork progression and is preceded by single-stranded DNA formation. *G3 (Bethesda)* **1**, 327–335.
 - 45 Pareek CS, Smoczynski R & Tretyn A (2011) Sequencing technologies and genome sequencing. *J Appl Genet* **52**, 413–435.
 - 46 Sloan DB, Broz AK, Sharbrough J & Wu Z (2018) Detecting rare mutations and DNA damage with

- sequencing-based methods. *Trends Biotechnol* **36**, 729–740.
- 47 Leduc F, Faucher D, Bikond Nkoma G, Gregoire MC, Arguin M, Wellinger RJ & Boissonneault G (2011) Genome-wide mapping of DNA strand breaks. *PLoS One* **6**, e17353.
- 48 Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, Gradia S, Vidal B, Garner E, Storach EM *et al.* (2017) Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600–606.
- 49 Canela A, Sridharan S, Sciascia N, Tubbs A, Meltzer P, Sleckman BP & Nussenzweig A (2016) DNA breaks and end resection measured genome-wide by end sequencing. *Mol Cell* **63**, 898–911.
- 50 Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K *et al.* (2013) Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**, 361–365.
- 51 Frock RL, Hu J, Meyers RM, Ho YJ, Kii E & Alt FW (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179–186.
- 52 Hoffman EA, McCulley A, Haarer B, Arnak R & Feng W (2015) Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Res* **25**, 402–412.
- 53 Kara N, Krueger F, Rugg-Gunn P & Houseley J (2020) Genome-wide analysis of DNA replication and DNA double strand breaks by TrAEL-seq. *bioRxiv* <https://doi.org/10.1101/2020.1108.1110.243931> [PREPRINT].
- 54 Lensing SV, Marsico G, Hansel-Hertsch R, Lam EY, Tannahill D & Balasubramanian S (2016) DSBCapture: in situ capture and sequencing of DNA breaks. *Nat Methods* **13**, 855–857.
- 55 Tchurikov NA, Kretova OV, Fedoseeva DM, Chechetkin VR, Gorbacheva MA, Karnaukhov AA, Kravatskaya GI & Kravatsky YV (2015) Mapping of genomic double-strand breaks by ligation of biotinylated oligonucleotides to fork domains: analysis of the data obtained for human rDNA units. *Genom Data* **3**, 15–18.
- 56 Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–197.
- 57 Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, Kallas T, Custodio J, Wernersson E, Li Y, Gao L *et al.* (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* **8**, 15058.
- 58 Zhou ZX, Zhang MJ, Peng X, Takayama Y, Xu XY, Huang LZ & Du LL (2013) Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Res* **23**, 705–715.
- 59 Bryan DS, Ransom M, Adane B, York K & Hesselberth JR (2014) High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res* **24**, 1534–1542.
- 60 Mao P, Brown AJ, Malc EP, Mieczkowski PA, Smerdon MJ, Roberts SA & Wyrick JJ (2017) Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. *Genome Res* **27**, 1674–1684.
- 61 Daigaku Y, Keszthelyi A, Muller CA, Miyabe I, Brooks T, Retkute R, Hubank M, Nieduszynski CA & Carr AM (2015) A global profile of replicative polymerase usage. *Nat Struct Mol Biol* **22**, 192–198.
- 62 Reijns MAM, Kemp H, Ding J, de Proce SM, Jackson AP & Taylor MS (2015) Lagging-strand replication shapes the mutational landscape of the genome. *Nature* **518**, 502–506.
- 63 Clausen AR, Lujan SA, Burkholder AB, Orebaugh CD, Williams JS, Clausen MF, Malc EP, Mieczkowski PA, Fargo DC, Smith DJ *et al.* (2015) Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol* **22**, 185–191.
- 64 Mao P, Smerdon MJ, Roberts SA & Wyrick JJ (2016) Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proc Natl Acad Sci USA* **113**, 9057–9062.
- 65 Hu J, Adar S, Selby CP, Lieb JD & Sancar A (2015) Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev* **29**, 948–960.
- 66 Shu X, Xiong X, Song J, He C & Yi C (2016) Base-resolution analysis of cisplatin-DNA adducts at the genome scale. *Angew Chem Int Ed Engl* **55**, 14246–14249.
- 67 Hu J, Adebali O, Adar S & Sancar A (2017) Dynamic maps of UV damage formation and repair for the human genome. *Proc Natl Acad Sci USA* **114**, 6758–6763.
- 68 Li W, Hu J, Adebali O, Adar S, Yang Y, Chiou YY & Sancar A (2017) Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[a]pyrene. *Proc Natl Acad Sci USA* **114**, 6752–6757.
- 69 Poetsch AR, Boulton SJ & Luscombe NM (2018) Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol* **19**, 215.
- 70 Liu ZJ, Martinez Cuesta S, van Delft P & Balasubramanian S (2019) Sequencing abasic sites in DNA at single-nucleotide resolution. *Nat Chem* **11**, 629–637.
- 71 Baranello L, Kouzine F, Wojtowicz D, Cui K, Przytycka TM, Zhao K & Levens D (2014) DNA break mapping reveals topoisomerase II activity genome-wide. *Int J Mol Sci* **15**, 13111–13122.

- 72 Baranello L, Kouzine F, Wojtowicz D, Cui K, Zhao K, Przytycka TM, Capranico G & Levens D (2018) Mapping DNA breaks by next-generation sequencing. *Methods Mol Biol* **1672**, 155–166.
- 73 Cao B, Wu X, Zhou J, Wu H, Liu L, Zhang Q, DeMott MS, Gu C, Wang L, You D *et al.* (2020) Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res* **48**, 6715–6725.
- 74 Cao H, Salazar-Garcia L, Gao F, Wahlestedt T, Wu CL, Han X, Cai Y, Xu D, Wang F, Tang L *et al.* (2019) Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. *Nat Commun* **10**, 5799.
- 75 Sriramachandran AM, Petrosino G, Mendez-Lago M, Schafer AJ, Batista-Nascimento LS, Zilio N & Ulrich HD (2020) Genome-wide nucleotide-resolution mapping of DNA replication patterns, single-strand breaks, and lesions by GLOE-Seq. *Mol Cell* **78**, 975–985.
- 76 Dabney J & Meyer M (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94.
- 77 Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C & Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12**, R18.
- 78 Kebschull JM & Zador AM (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res* **43**, e143.
- 79 Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, Parekh S, Geuder J, Hellmann I & Enard W (2018) Sensitive and powerful single-cell RNA sequencing using mcSCR-seq. *Nat Commun* **9**, 2937.
- 80 Ding J, Taylor MS, Jackson AP & Reijns MA (2015) Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nat Protoc* **10**, 1433–1444.
- 81 Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, Kieffer-Kwon KR, Pekowska A, Zhang H, Rao SSP *et al.* (2017) Genome organization drives chromosome fragility. *Cell* **170**, 507–521.
- 82 Tubbs A, Sridharan S, van Wietmarschen N, Maman Y, Callen E, Stanlie A, Wu W, Wu X, Day A, Wong N *et al.* (2018) Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142.
- 83 Chen J & Furano AV (2015) Breaking bad: the mutagenic effect of DNA repair. *DNA Repair (Amst)* **32**, 43–51.
- 84 Chatterjee N & Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* **58**, 235–263.
- 85 Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A & Kunkel TA (2010) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci USA* **107**, 4949–4954.
- 86 Bollum FJ (1974) Terminal deoxynucleotidyl transferase. *Enzymes* **10**, 145–171.
- 87 Zhu Y, Biernacka A, Pardo B, Dojer N, Forey R, Skrzypczak M, Fongang B, Nde J, Yousefi R, Pasero P *et al.* (2019) qDSB-Seq is a general method for genome-wide quantification of DNA double-strand breaks using sequencing. *Nat Commun* **10**, 2313.
- 88 Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* **99**, 5261–5266.
- 89 Phillips J & Eberwine JH (1996) Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods* **10**, 283–288.
- 90 Zong C, Lu S, Chapman AR & Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626.
- 91 Hoeijmakers WA, Bartfai R, Francoijs KJ & Stunnenberg HG (2011) Linear amplification for deep sequencing. *Nat Protoc* **6**, 1026–1036.
- 92 Dueck H, Eberwine J & Kim J (2016) Variation is function: are single cell differences functionally important?: testing the hypothesis that single cell variation is required for aggregate function. *BioEssays* **38**, 172–180.
- 93 Saadatpour A, Lai S, Guo G & Yuan GC (2015) Single-cell analysis in cancer genomics. *Trends Genet* **31**, 576–586.
- 94 Smith DJ & Whitehouse I (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–438.
- 95 Wong RP, Garcia-Rodriguez N, Zilio N, Hanulova M & Ulrich HD (2020) Processing of DNA polymerase-blocking lesions during genome replication is spatially and temporally segregated from replication forks. *Mol Cell* **77**, 3–16.
- 96 Lopes M, Foiani M & Sogo JM (2006) Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell* **21**, 15–27.
- 97 Feng Y, Zhang Y, Ying C, Wang D & Du C (2015) Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* **13**, 4–16.
- 98 Radukic MT, Brandt D, Haak M, Müller KM & Kalinowski J (2020) Nanopore sequencing of native adeno-associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol. *bioRxiv* 2019.2012.2027.885319 [PREPRINT].