Inferring direction of replication fork and mechanism of DNA damage using sequencing data

Norbert Dojer[1], Abhishek Mitra[1], Ji Li[1], Yea-Lih Lin[2], Anna Kubicka[3], Magdalena Skrzypczak[3], Krzysztof Ginalski[3], Philippe Pasero[3] and **Maga Rowicka**[1],

[1] Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77554
[2] Institut de Genetique Humaine, Centre National de la Recherche Scientifique, UPR 1142, Montpellier, France
[3] Centre of New Technologies, University of Warsaw, Warsaw, Poland

Double-stranded DNA breaks (DSBs) are a genotoxic form of DNA damage. The damage to both DNA strands precludes the straightforward use of the complementary strand as a template for repair, resulting in mutagenic lesions. Despite many studies on the mechanisms of DSB formation, our knowledge of them is very incomplete. A main reason for our limited knowledge is that, to date, DSB formation has been extensively

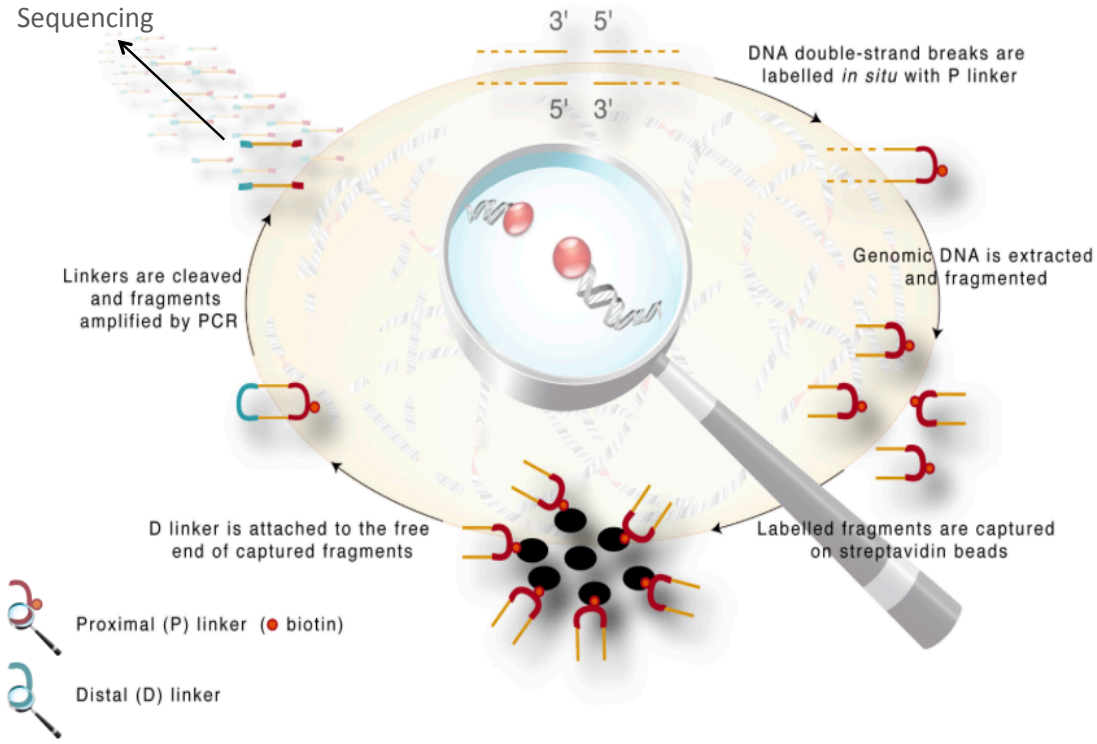

Figure 1. Overview of our DSB labeling method. From the top, clockwise: (1) DSBs are ligated *in situ* to a proximal linker P (red arch) covalently linked to biotin (red oval), (2) genomic DNA is extracted and fragmented and (3) labeled fragments are captured on streptavidin beads (black ovals). (4) A distal linker D (cyan arch) is then ligated to the free extremity of captured fragments. (5) Fragments are released by linker digestion with I-SceI endonuclease and then amplified by PCR using linker-specific primers and (6) sequenced.

studied only at specific loci but remains largely unexplored at the genome-wide level. We

recently developed a method to label DSBs *in situ* followed by deep sequencing (BLESS), and used it to map DSBs in human cells [1] with a resolution 2-3 orders of magnitude better than previously achieved. The method is explained in Fig. 1.

There are many factors inducing DSBs, including replication stress, oxidative stress and irradiation. Most of them cause two-ended DSBs (having two free ends of DNA), the only exception is replication stress which usually induces one-ended DSBs (caused by replication fork stalling and collapse), as shown in Fig. 2 below.

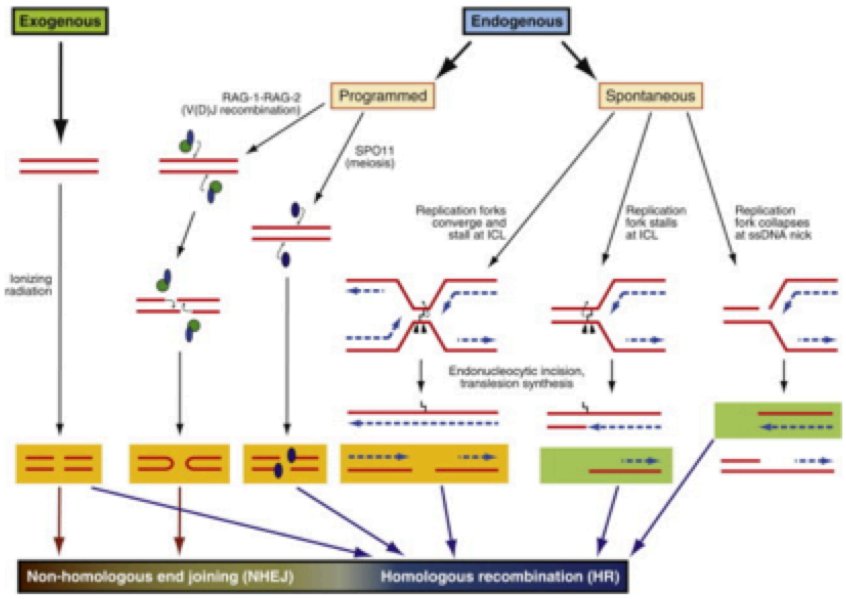

Figure 2. One-ended and two-ended breaks. Different mechanisms leading to DSBs result either in two-ended breaks (yellow) or one-ended breaks (green). Therefore, computationally distinguishing between one-ended and two-ended breaks can be very important tool for clarifying mechanisms of DSB formations in different conditions and in different chromatin context. Figure from Chapman JR, Taylor MR, Boulton SJ, *Mol Cell.* 2012 Aug 24;47(4):497-510.

We use this observation to infer DSBs resulting from replication stress and to analyze chromatin context and sequence features related to replication stress-induced DSBs. Moreover, we show how to reconstruct the direction of replication fork movement from BLESS-Seq read pattern. We apply this concept to infer replication domain boundaries for several cell lines and conditions and to analyze how they change upon treatments and vary between cell lines. We also provide experimental verification for the proposed computational method and show that purely computational methods can predict >80% of experimentally detected DSBs.

References:

[1] Crosetto N, Mitra A, Silva M, Bienko M, Dojer N, Wang Q, et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing, *Nature Methods*, 2013;10:361-5.

Keywords: DNA double-stranded break, DSB, BLESS-Seq, sequencing, replication, replication fork, replication stress