

Review

Mapping DNA Topoisomerase Binding and Cleavage Genome Wide Using Next-Generation Sequencing Techniques

Shannon J. McKie ^{1,2} , Anthony Maxwell ²  and Keir C. Neuman ^{1,*} ¹ Laboratory of Single Molecule Biophysics, NHLBI, Bethesda, MD 20892, USA; Shannon.McKie@jic.ac.uk² Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK; tony.maxwell@jic.ac.uk

* Correspondence: neumank@mail.nih.gov

Received: 10 December 2019; Accepted: 10 January 2020; Published: 13 January 2020



Abstract: Next-generation sequencing (NGS) platforms have been adapted to generate genome-wide maps and sequence context of binding and cleavage of DNA topoisomerases (topos). Continuous refinements of these techniques have resulted in the acquisition of data with unprecedented depth and resolution, which has shed new light on *in vivo* topo behavior. Topos regulate DNA topology through the formation of reversible single- or double-stranded DNA breaks. Topo activity is critical for DNA metabolism in general, and in particular to support transcription and replication. However, the binding and activity of topos over the genome *in vivo* was difficult to study until the advent of NGS. Over and above traditional chromatin immunoprecipitation (ChIP)-seq approaches that probe protein binding, the unique formation of covalent protein–DNA linkages associated with DNA cleavage by topos affords the ability to probe cleavage and, by extension, activity over the genome. NGS platforms have facilitated genome-wide studies mapping the behavior of topos *in vivo*, how the behavior varies among species and how inhibitors affect cleavage. Many NGS approaches achieve nucleotide resolution of topo binding and cleavage sites, imparting an extent of information not previously attainable. We review the development of NGS approaches to probe topo interactions over the genome *in vivo* and highlight general conclusions and quandaries that have arisen from this rapidly advancing field of topoisomerase research.

Keywords: DNA topoisomerase; next-generation sequencing; genome wide; topoisomerase cleavage; topoisomerase binding; antibiotics; anticancer drugs

1. Introduction

DNA topoisomerases (topos), pervasive across all domains of life and indispensable for cellular survival, alter DNA topology through the formation of transient single- or double-stranded DNA breaks (DSBs); for comprehensive reviews, see [1,2]. Generally speaking, topos can relax both positive and negative supercoils, as well as unknot and decatenate interlinked DNA. Performing these reactions is vital to maintaining genomic integrity, particularly during DNA transcription, replication, and segregation. A common feature among all topos is the formation of a reversible covalent link between the protein and the DNA backbone, known as a cleavage complex. Whilst this is an integral step in the topo mechanism, it is also a highly vulnerable state for the duplex. Numerous topo poisoning agents exploit this vulnerability by binding at the cleaved DNA site and preventing re-ligation [3,4]. Other endogenous factors such as colliding replication/transcription forks or DNA lesions in close proximity, can also contribute to the cleavage complex becoming a stabilized DSB [5]. Using topo II inhibitors to treat cancer has implicated topos directly in the development of secondary cancers such as acute myeloid leukemia, which arises through genomic translocations caused by the attempted repair of inhibitor-stabilized topo-dependent DSBs [6,7].

The cleavage complex formed when topoisomerases actively engage with the duplex differs amongst the different types (Figure 1). The type I topoisomerases, which form single-stranded breaks (SSBs), are subcategorized as type IA and type IB. Type IA topoisomerases form 5' covalent bonds between the active site tyrosine and the 5' phosphate group of the DNA backbone. The type IB topoisomerases differ in that the active site tyrosine attacks the 3' phosphate group instead. The type II topoisomerases, which form double-stranded breaks (DSBs), are also subcategorized as type IIA and type IIB. The type IIA topoisomerases form two 5' covalent linkages between the active site tyrosines and the 5' phosphate, one on each strand, offset by 4 base-pairs (bp), and therefore generating 4 bp 5' overhangs. The type IIB topoisomerases behave similarly, except they are thought to produce 2 bp 5' overhangs (see Figure 1 for details) [2].

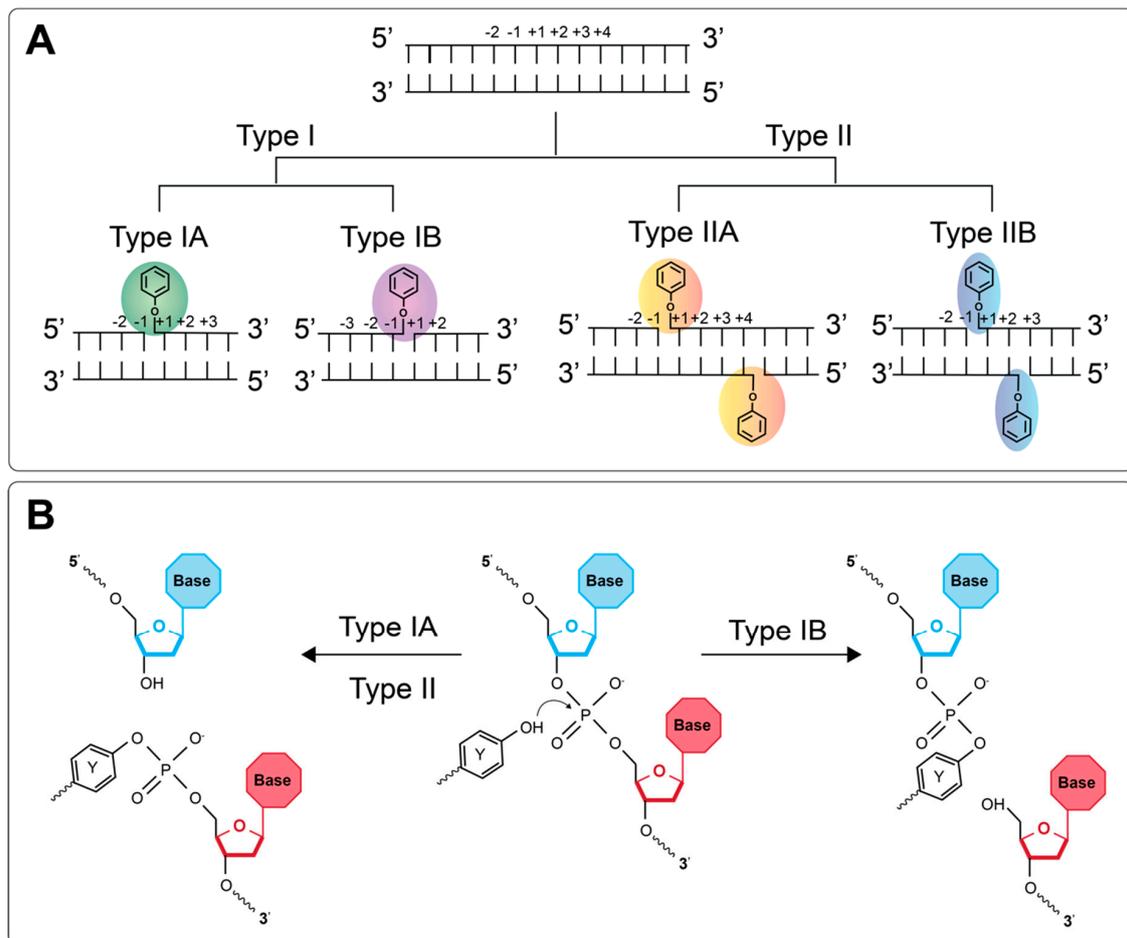


Figure 1. DNA cleavage by the DNA topoisomerases. **(A)** Representation of where the DNA is cleaved and the phosphotyrosyl bond that is formed by each type of topo. **(B)** Chemistry of the covalent attachment between the active site tyrosine of the topo and the DNA backbone. For type IA and II topoisomerases, the 5' phosphate group undergoes nucleophilic attack by the tyrosine OH group, whereas, for type IB, it is the 3' phosphate group.

As they are vital to cellular survival and constitute effective antibiotic and anti-cancer targets, understanding the DNA binding and cleavage profiles of topoisomerases in a genome-wide context is of great interest to the field. Whereas the fundamental biochemistry of the topoisomerase reactions has been established, how their activity is regulated across the genome and through time *in vivo* has only recently begun to be addressed in a systematic manner with the utilization of next-generation sequencing (NGS) technologies. In this review, we discuss the genome-wide mapping of topoisomerases that have been characterized using NGS approaches and the specific nuances of the protocols developed to do so (summarized in Table 1).

Table 1. A summary of the techniques discussed in this review, which were developed to find ways to explore topo binding and cleavage activity genome wide.

| Method | Protein | NGS Platform | Brief Description |
|-------------------------|------------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Type I/II topo ChIP-seq | Type I/II topois | Non-specific | To determine the location and sequence context of topoisomerase binding to DNA using ChIP-seq. Crosslinking agent is applied to cells, before lysis, sonication, and immunoprecipitation of DNA bound to protein of interest. Crosslinking is reversed and DNA is ligated to adapters and sequenced [8]. |
| BLESS | Non-specific | Roche 454 or Illumina Hiseq | To map the location of double strand breaks over the genome using direct in situ Breaks Labelling, Enrichment on Streptavidin and next-generation Sequencing (BLESS). Using human and mouse cells, a fixing agent is used before the cells are lysed and purified nuclei extracted. Biotinylated adapters are used to label DSBs in situ before the DNA is extracted, sonicated and enriched using streptavidin. After a second adapter ligation phase, the DNA is sequenced [9]. |
| END-seq | Non-specific | Illumina Hiseq2500 or Illumina Nextseq500 | A more sensitive and robust approach to map DSBs. Mouse lymphocytes or thymocytes are fixed in agarose before being treated with a proteinase K solution, followed by an RNase solution. The biotinylated adapters are then ligated before the DNA is extracted, sonicated and enriched using streptavidin. The DNA then undergoes a second adapter ligation before being sequenced [10]. This protocol was also adapted for use in human and mouse cells looking at topo II cleavage complexes [11]. |
| Spo11-oligo-seq | Spo11 | Roche 454 | To map the location and sequence context of Spo11 mediated DNA cleavage. Nuclei from meiotic <i>Saccharomyces cerevisiae</i> cultures is extracted, sonicated and ssDNA oligos bound to Spo11 are immunoprecipitated. Proteinase K removes Spo11 and DNA is extended at the 3' end by TdT before a dsDNA adapter is ligated, and the complementary strand synthesized. Adapter 1 contains an inverted dT so the 3' end is unmodifiable, as is the 5' end due to the phosphotyrosyl adduct. The DNA then undergoes another 3' extension and adapter ligation which only affects the newly synthesized strand. The dsDNA fragments are then sequenced [12]. |
| SSB-seq and DSB-seq | Topo II | Illumina GA | To map the location and sequence context of single- and double-stranded DNA breaks induced by Topo II. Human colon cancer cells (HCT116) are treated with etoposide and the genomic DNA is extracted. For single-stranded break (SSB)-seq, SSBs are labelled using nick translation [13] involving digoxigenin labelled dUTP so that fragments can be immunoprecipitated using anti-digoxigenin. For single-stranded break (DSB)-seq, the DNA is 3'-end tailed using TdT in the presence of biotinylated dNTPs and is enriched using streptavidin coated beads. The S1-nuclease is used to cleave the 3'-end tailing and then both the SSB- and DSB-seq DNA populations are adapter ligated and sequenced [14]. |
| CC-seq | Topo II | Illumina Miseq or Nextseq | To map the location and sequence context of Topo II cleavage complexes using cleavage complex (CC)-seq. Human or <i>S. cerevisiae</i> cells are treated with etoposide before lysis. Bulk cellular proteins are removed using a phenol-chloroform extraction and the DNA retained in the aqueous fraction is sonicated. Protein bound DNA is enriched using a silica membrane. The first adapter is ligated to the sonicated DNA-end before Topo II is removed using TDP2, followed by the second adapter ligation and sequencing [15]. |

Table 1. Cont.

| Method | Protein | NGS Platform | Brief Description |
|----------|------------|----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| NorfIIP | Topo IV | Illumina GA | To map the location and sequence context of Topo IV cleavage using norfloxacin immunoprecipitation (NorfIIP). <i>Escherichia coli</i> (<i>E. coli</i>) carrying C-terminal FLAG fusions of ParE and ParC is treated with norfloxacin before lysis, sonication and immune-precipitation using anti-FLAG. Adapters are ligated and fragments sequenced [16,17]. |
| Topo-seq | DNA gyrase | Illumina Nextseq | To map the location and sequence context of DNA gyrase cleavage. <i>E. coli</i> cells are treated with ciprofloxacin, oxolinic acid or microcin B17, before lysis and sonication. DNA bound to gyrase is enriched using immunoprecipitation and proteinase K is used to remove gyrase. The DNA is then treated using the Accel 1S NGS kit, which ligates adapters to the 3' strand from the gyrase cleavage site (5' strand is unmodifiable due to the phosphotyrosyl adduct), and synthesizes the complementary strand. The resultant dsDNA is then sequenced [18]. |
| Top1-seq | Topo 1B | Illumina GAII and SoLid (applied biosystems) | To map the location and sequence context of Topo 1B cleavage. Human colon cancer cells (HCT116) are briefly treated with camptothecin before being lysed, sonicated and immunoprecipitated. Topo 1B is proteolyzed and the DNA ligated to adapters and sequenced [19]. |

1.1. Next-Generation Sequencing

The first widely used sequencing technique was developed in 1977 by Frederick Sanger and colleagues, which employed the use of chain terminating dideoxynucleotides [20]. It was later automated by Applied Biosystems with the development of the AB370 in 1987, which accelerated sequencing time and increased accuracy by using capillary electrophoresis [21]. Whilst this was a revolutionary technique for the time, and facilitated the completion of the first human genome in 2004 [22], this ‘first generation’ sequencing method had its limitations. Sanger sequencing is relatively low throughput, time intensive, and expensive. Efforts to ameliorate sequencing resulted in the development of affordable, high throughput technologies. This led to a number of innovations in the DNA sequencing field and the birth of next-generation sequencing (NGS) platforms [23]. NGS platforms allow cell-free library preparation and are able to perform millions of parallel sequencing reactions, detecting the output directly without electrophoresis. This meant that genomes could be fully sequenced in an extraordinarily short timeframe. The main drawback, particularly for the first NGS platforms, was the relatively short read length. However, this was circumvented through the development of new algorithmic tools for aligning short reads.

In 2005 the first NGS platform was released by Roche, the 454 genome sequencer [24]. The detection method used, called pyrosequencing, relied on the activity of pyrophosphate, which would emit light when a base was incorporated, and allowed the sequencing of ~200,000 reads, ~110 bp long per run, generating 20 Mb of data. More recent advances made reads of 700 bp possible, generating 0.7 GB of data in less than 24 h with the 454 GS FLX Titanium system. In 2006, Sequencing by Oligo Ligation Detection (SOLiD) by Agencourt, was bought by Applied Biosystems [21]. This NGS platform uses octamer oligos, each with a unique fluorescent label, that interrogate the first two bases in the sequence adjacent to the sequencing primer. Once the label color is detected, it is released by cleavage between bases 5 and 6 of the octamer, and another probe will hybridize to the next two free bases and so on. The process is then repeated using a sequencing primer one nucleotide shorter than the first. In this manner, the first-round sequences bases in positions 4, 5, 9, 10, 14, 15, etc., whilst the second round sequences 3, 4, 8, 9, 13, 14, etc., until sequencing primer position 0 is reached [25]. Due to each base being interrogated twice, early SOLiD systems benefitted from high accuracy, with more recent versions able to sequence 85 bp fragments with 99.99% accuracy [21]. Also, in 2006, the Illumina platform was released, which facilitates detection through the use of fluorescently labelled reversible-terminator nucleotides. Each nucleotide is uniquely colored allowing its incorporation to be detected, before the subsequent removal of the 3' terminator group along with the fluorescent probe. Numerous Illumina platforms have since been manufactured, employing this technique, such as MiSeq, HiSeq and NextSeq, with the main difference between these systems being the depth of parallel sequencing power and data output. The most recently released Illumina technology, Novaseq, is capable of performing up to 40 billion sequencing reactions in parallel, of up to 250 bp in length with an output maximum of 6000 Gb. This sequencing scale has opened the door to a wide range of innovative approaches that leverage these platforms to develop unique methodologies to probe field-specific questions. In this review, the research cited relies on using one of the three aforementioned platforms. In most instances, the choice of platform is somewhat arbitrary, largely dictated by the best available sequencing depth and length. A comprehensive overview of sequencing platforms is provided in [21].

1.2. Non-Specific DSB Mapping

Genome metabolism is highly coordinated, dynamic and staggeringly complex, with numerous proteins interacting with both the duplex and other proteins, to maintain DNA integrity while facilitating DNA processing. DNA within each cell is subject to an assault from both endogenous and exogenous factors, resulting in various forms of damage that require either repair or tolerance [26]. Arguably one of the more toxic forms of damage is the induction of DNA breaks, particularly DSBs, which can lead to carcinogenic translocations or cell death [27]. One early use of NGS was to explore the DSB landscape of the genome, to better understand when and where the majority of DSBs occur.

One such study developed a generalized technique called BLESS (direct in situ Breaks Labelling, Enrichment on Streptavidin and next-generation Sequencing) [9]. By labelling DSBs in situ, the level of false positives caused by genome shearing during extraction was minimized, allowing the resolution of the “breakome” in both human and mouse cells. The cells were fixed and lysed to attain purified nuclei, which were then subject to biotinylated adapter ligation. The genomic DNA could then be fragmented and enriched for endogenously labelled DSBs using streptavidin. Exposing cells to aphidicolin, an inhibitor of DNA polymerase, it was found that replication stress is a major driver of DSB generation, particularly within genes and satellite regions. BLESS was also able to accurately identify telomere ends as well as Sce-1 endonuclease sites.

A subsequent study developed an alternative technique, designed to overcome noise and background complications associated with BLESS, as well as to gain information on DSB end structure [10]. Termed END-seq, this protocol involves embedding mouse thymocytes or lymphocytes within agarose before soaking the agarose in proteinase K solution, followed by an RNase solution. The agarose embedded cells are then end repaired and adapter ligated, using a specifically indexed and biotinylated adapter. The DNA is then extracted, sheared, and immunoprecipitated using streptavidin, before ligation of the second adapter. The DNA library was single-end sequenced using Illumina HiSeq2500 or Nextseq500. This technique proved to be highly sensitive, with nucleotide resolution at break sites, and led to a number of new insights in the field of T- and B-cell research. It also has potential as a tissue specific CRISPR-Cas9 off target cleavage detection protocol. A limitation, however, is evident in the end repair stage in which 3′ overhangs are resected and 5′ overhangs are filled-in to blunt the DNA for adapter ligation. This means that any information concerning the length of the overhang is lost.

1.3. Spo11 DSB Mapping

One of the first studies to implement the use of NGS in topoisomerase-specific genome-wide DSB mapping, explored the generation of meiosis specific DNA cleavage by the topoisomerase-related protein, Spo11, in *Saccharomyces cerevisiae* [12]. Spo11 acts as a homodimer and is highly homologous to the topoisomerase VIA subunit [28]. It is responsible for DSB generation during meiosis, which ultimately leads to chromosomal recombination, making Spo11 an important factor in driving evolution through the generation of genome diversity. Spo11 homodimers form covalent 5′ phosphotyrosine linkages on both strands of a DNA duplex, generating a DSB with a 2 bp overhang, a process highly reminiscent of the type II topoisomerases. The main difference being that Spo11 DSBs are irreversible and followed by endonucleolytic cleavage adjacent to the bound Spo11. This releases the Spo11 monomer covalently attached to an oligonucleotide (oligo) and generates 3′ overhangs, which facilitate strand invasion and recombination. The released Spo11-bound oligos are separated into two distinct subpopulations of equal quantity, longer (21–37 nt) and shorter (<12 nt). Pan and colleagues used these oligos to develop an NGS-based approach, termed Spo11-oligo-seq, designed to overcome limitations encountered by previous studies employing DNA microarray analysis, to attain a comprehensive genome-wide Spo11-mediated DSB map with nucleotide resolution at the cleavage sites.

Nuclear extracts were obtained from synchronous meiotic yeast cells and the Spo11-bound oligos captured via immunoprecipitation and then released by proteolytic digestion (Figure 2A). The oligos were then 3′-GTP tailed using terminal deoxynucleotidyl transferase (TdT) and ligated to a double-stranded DNA adapter using T4 DNA ligase before the complementary strand was synthesized using Klenow polymerase. The DNA was purified using denaturing polyacrylamide gel electrophoresis (PAGE), before a second round of 3′-GTP end tailing, adapter ligation and strand synthesis. The original oligo cannot be further modified as the 3′ adapter has an inverted dT and the 5′ DNA end is blocked by the phosphotyrosine residue. Sequencing adapters were then added via PCR before the oligos were sequenced using the 454 platform (Roche). The fragments could then be aligned and bioinformatically processed, generating an in-depth representation of Spo11-DSB activity genome wide. This revealed a complex picture in which the DSB landscape in *S. cerevisiae* is influenced by

numerous factors in a hierarchical manner, and that the concept of DSB ‘hotspots’ [29] is misleading. Spo11 demonstrates an opportunistic cutting activity where most sites can constitute a cleavage site but with variable probability that is modulated by numerous other factors.

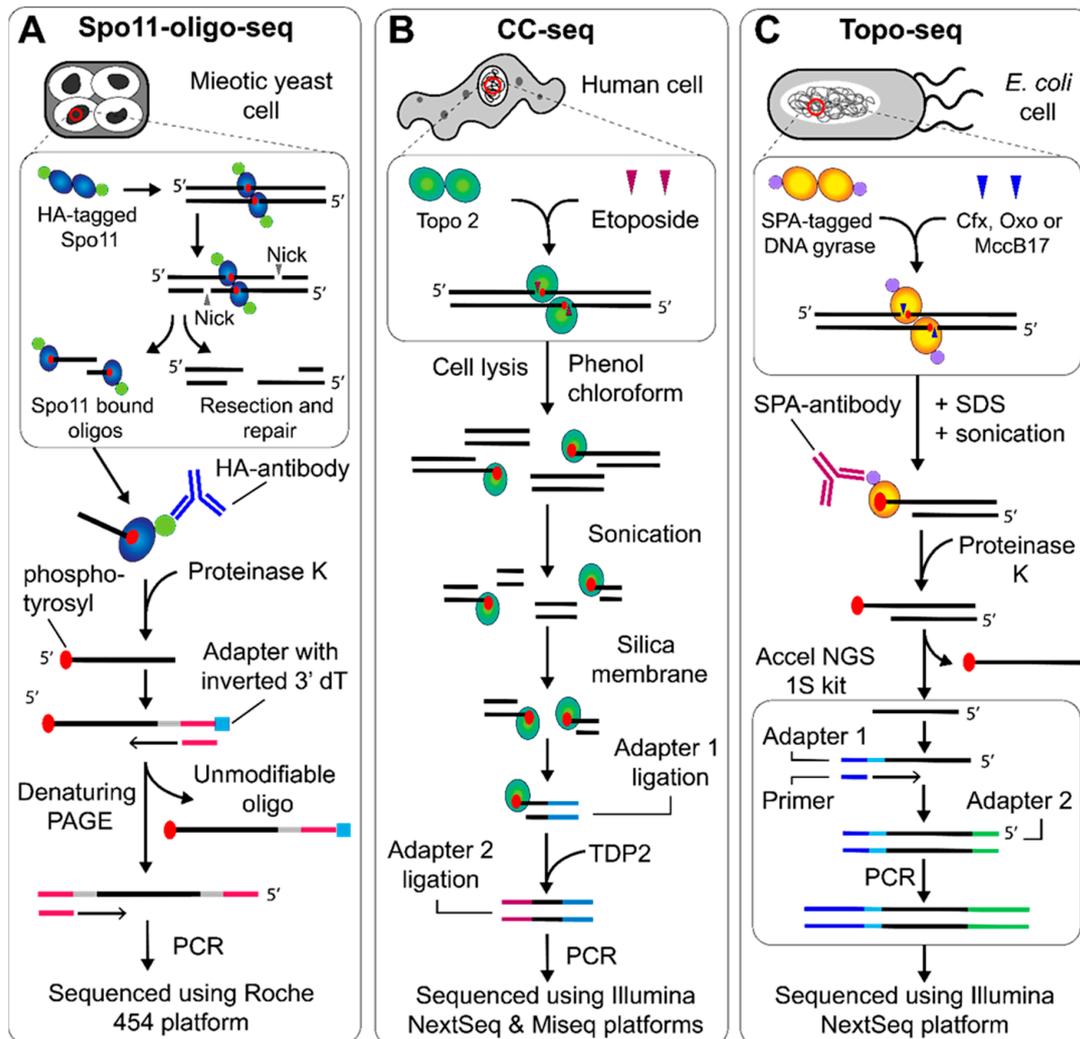


Figure 2. NGS protocols to map topo cleavage sites genome wide. **(A)** Spo11-oligo-seq began with DNA extraction from meiotic *S. cerevisiae* cells. HA-tagged Spo11-bound DNA oligos are enriched with the HA-antibody, before the protein is removed using proteinase K. The 3' end is extended using TdT and an adapter is ligated that contains an inverted dT residue. The complementary strand is synthesized, followed by separation using denaturing PAGE, 3' end tailing and adapter ligation on the newly synthesized strand, which can then be sequenced using the Roche 454 platform [12]. **(B)** CC-seq, used to map topo II cleavage, began with DNA extraction from etoposide treated human cells. The bulk protein is removed using phenol-chloroform extraction, whilst the protein bound to DNA remains in the aqueous phase. The DNA can then be sonicated and enriched using a silica membrane. The first ligation attached adapters to the sonicated end, which was then followed by TDP2-dependent removal of the topo II, before a second adapter ligation followed by sequencing using the Illumina NextSeq and Miseq platforms [15]. **(C)** Topo-seq, used to map gyrase cleavage, began with DNA extraction from *E. coli* cells treated with ciprofloxacin (cfx), oxolinic acid (oxo) or microcin B-17 (MccB17), followed by sonication. Gyrase cleavage complexes were then enriched using the SPA-antibody before proteinase K treatment and use in the Accel NGS 1S kit to generate fragments that can be sequenced using the Illumina Nextseq platform [18].

1.4. DNA Topoisomerase II

In 2014, human topoisomerase II (topo II), both the α - and β -isoforms collectively became the first true topoisomerases explored at the genome-wide scale [14]. Human colon cancer cells (HCT116) were treated with etoposide (a topo II cleavage-complex stabilizing agent) and the genomic DNA extracted. Here, two different methods were implemented to capture both SSBs and DSBs, named SSB-seq and DSB-seq, respectively. In order to label SSBs, nick-translation was used, a technique first described in 1977, which enables labelling of DNA at a nick site [13]. In this case, digoxigenin-labelled dUTP was used so that the nicked DNA fragments, once sonicated, could be immunoprecipitated using anti-digoxigenin. For DSB-seq, the DNA was 3'-end tailed using TdT in the presence of biotinylated dNTPs so that after sonication, the DNA could be enriched using streptavidin coated beads. S1-nuclease was used to cleave the 3'-end tailing and then both the SSB- and DSB-seq DNA populations were adapter ligated and sequenced using Illumina.

This study highlighted the tendency of low-dose etoposide to produce SSBs, as opposed to DSBs, indicating that etoposide predominantly acts on a single monomer of topo II, suggesting in turn that only one cleavage site is poisoned, leaving the other to be freely re-ligated. This is in agreement with a previous study that found only 3% of etoposide-stabilized cleavages are DSBs [30]. It is worth noting, that whilst topo I (type IB) can form single-stranded nicks, these nicks do not have the free 3'-OH group, so are not substrates for nick-translation and won't be labelled, making SSB-seq topo II specific. Whole-genome analysis revealed that topo II SSBs and DSBs were enriched at transcription start sites and more prevalent in highly expressed genes, further demonstrating the cellular necessity of topo II in regulating topology during transcription.

In 2016, came a comprehensive study looking specifically at murine liver topo II β activity, using ChIP-seq, ChIP-exo and Hi-C [31]. For an in-depth ChIP-seq protocol see [8], however, in short, DNA protein interactions were stabilized using a crosslinking agent and topo II β -bound DNA was enriched using anti-topo II β antibody. The crosslinking was reversed, and the DNA underwent library preparation in which the ends were repaired, A-tailed, and adapters were ligated before sequencing using the Illumina HiSeq2500 platform.

Topo II β and topo II α , whilst being structurally and catalytically alike, are functionally distinct, performing different cellular functions. Topo II β knockouts are lethal owing to disruption of neuronal differentiation [32]. Conditional knockouts implicated topo II β activity in retinal development [33] and ovulation [34], and the use of poisons showed it played a role in spermatogenesis [35] and lymphocyte activation [36]. This has led to a picture of topo II β being an important factor in tissue-specific development and cellular differentiation, a concept elaborated upon by Uusküla-Reimand and colleagues with the help of ChIP-seq [31]. They found topo II β interacts with the transcriptional regulator CTCF and the cohesin complex and acts on DNA located at the borders of chromosomal domains as well as conserved transcription factor binding sites.

Topo II α activity genome wide was also explored in depth in 2018 [37], using human leukemia K562 cells treated with etoposide, mitoxantrone, genistein or p-benzoquinone, all of which are known topo II inhibitors. DNA was extracted and sonicated before bead-based immunocapture using rabbit anti-topo II α polyclonal antibody. The DNA was released from the topo II α using calf alkaline phosphatase (CIP), an enzyme which is capable of phosphodiesterase activity at 3% of its canonical monophosphatase activity. CIP is therefore able to partially remove the phosphotyrosine residue, allowing for end repair and adapter ligation. The generated libraries were then sequenced using the illumina HiSeq2500 platform. In agreement with past studies, they found that the majority of topo II α cleavage sites were SSBs rather than DSBs, the latter only making up 1.3–3.4% of cleavage events. Reads clustered in regions of compact genome, reminiscent of the cleavage profile of yeast Spo11, and with a preference for regions with high transcriptional activity. They also identified chromosomes that consistently contained more topo II α cleavage sites than others. In particular chromosome 11 that includes the *KMT2A* gene, which houses a drug-induced leukemia translocation break point. The *KMT2A* gene, along with others, forms the *KMT2A* recombinome, and a proportion of these genes

demonstrated significantly higher levels of topo II α cleavage. They also concluded that topo II α cleavage sites were enriched in genes necessary for DNA metabolism and transcriptional regulation.

As protocols are refined, an increasingly more precise DSB landscape for topo II is emerging. In 2019, DSB maps were generated with cleavage-site nucleotide resolution for human and yeast topo II [15]. The protocol, termed cleavage complex (CC)-seq (Figure 2B), begins with genomic DNA extraction from human or yeast cells treated with etoposide, followed by a phenol-chloroform extraction to remove bulk proteins, leaving protein-bound DNA in the aqueous fraction. The DNA is then sonicated and enriched using a silica membrane, which will only bind DNA that is protein bound. The library preparation then occurs in two phases, with the P7 adapter being ligated to the protein-free DNA end. The topo II is then cleaved using TDP2, an enzyme evolved to rescue stalled topo II cleavage complexes in vivo [38], before the P5 adapter is ligated and the library is sequenced. In this study the NGS platforms used were the Illumina MiSeq and NextSeq 500.

As this study was conducted on both human and yeast cell lines, species-specific conclusions about topo II activity could be made. For both human and yeast, topo II activity was increased in gene dense regions of the genome. However, only a correlation with transcription rate and gene length was observed in human cells. For yeast, topo II activity was concentrated in intergenic regions, which are numerous and evenly spaced. For human cells, topo II activity was found in the gene bodies just downstream of the transcription start site and was maximized for genes that are longer and more actively transcribed. This strengthens the hypothesis that topo II in human cells plays an integral role in topology maintenance during transcription. This gene-local activity pattern may not be necessary in yeast as the genome is both far smaller and less complex. There is also the potential alternative that in yeast cells topo I may be more active in relieving transcriptional stress on the duplex than topo II, whilst in human cells this role is more evenly shared. It is worth noting that for topo II α specifically [37], activity in the absence of cleavage-complex stabilizing drugs was found to be located at the distal ends of genes, with the addition of drugs causing a proximal shift.

An analysis of etoposide-induced topo II cleavage complexes (topoIIcc) in mouse and human cell lines used the aforementioned END-seq protocol, which was adapted to explore both processed and unprocessed topoIIccs [11]. Mouse embryonic fibroblasts (MEFs) depleted of the cohesin subunit, SMC3, exhibited a dramatic decrease in topoIIcc, whereas Smc3^{+/+} control cells exhibited topoIIcc enrichment at cohesin sites. Not only was this decrease in topoIIcc proportional to cohesin levels in MEFs, similar results were obtained with a HCT-116 human cell line depleted of RAD21, another member of the cohesin complex. Moreover, depleting MEFs for the WAPL protein, responsible for cohesin removal from DNA, also increased the levels of etoposide-induced DNA damage. Interestingly, whilst topoII activity as well as binding (also measured here using ChIP-seq) was dependent on cohesin occupancy; the converse relationship was not apparent, with RAD21 and CTCF levels remaining unchanged in topoII β ^{-/-} B cells.

Whilst the activity and binding of topo II was found to be dependent on cohesin, the processing of these sites was sensitive to transcriptional activity. Etoposide-induced topoIIcc occurring within transcription start sites (TSS) or gene bodies were more likely to produce chromosomal translocations than those occurring in intergenic regions, with higher levels of translocations in regions of high transcription and decreased levels of translocations on addition of the transcriptional inhibitor 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB). In eukaryotic cells, the conversion of a topoIIcc to a protein-free DSB is mediated by TDP2. The processing of topoIIcc in the presence of DRB was decreased, with more lesions remaining protein bound, whilst transcriptional activation by INF β increased levels of protein-free DSBs. Hence, the processing and repair of topoIIcc is stimulated by transcriptional activity, and this in turn leads to an increase in translocation rate. This effect seemed to be dependent on proteasome activity, rather than TDP2, with proteasome inhibition by epoxomicin increasing the number of topoIIcc that are reversible (protein-bound) and decreasing the amount of protein-free DSBs 5.5-fold.

1.5. DNA Topoisomerase IV

In 2018, NGS was used to explore DNA topoisomerase IV (topo IV) binding and cleavage sites genome wide in *Escherichia coli* [16,17]. To explore binding in the absence of cleavage, a traditional ChIP-seq method was employed in which topo IV was allowed to bind DNA and then crosslinked, followed by antibody-based fragment capture. To explore topo IV cleavage complexes, a novel approach was developed, called NorfIIP, relying on norfloxacin, a quinolone that traps cleavage complexes produced by prokaryotic type II topoisomerases.

Four highly enriched topo IV binding sites were found, one of which corresponded to the *dif* site, a previously confirmed hotspot of topo IV cleavage. The *dif* (deletion induced filamentation) site is located at the replication terminus and is important for chromosome dimer resolution [39]. To identify additional binding sites, the raw sequencing data were filtered for sites showing the highest Pearson correlation with the sequencing pattern at the *dif* site. This approach led to the identification of 19 additional topo IV binding sites throughout the chromosome. These sites were found frequently within intergenic regions and spanned 200 bp with no significant consensus sequence. Topo IV binding enrichment was also detected in GC rich regions. By combining this binding data with specific topo IV cleavage sites, enhanced with norfloxacin, it was found that only certain binding sites correspond to topo IV cleavage sites, indicating that not all DNA binding results in topo IV cleavage of the DNA. The *dif* site remained strong for both binding and cleavage by topo IV, as well as a site at 2.56 Mb. However, the other strong binding sites were not present in the norfloxacin-dependent cleavage data. A characteristic topo IV-dependent peak profile was found in which two 170 bp peaks were present, separated by a 130 bp reduction in sequencing coverage. This was hypothesized to be due to the presence of topo IV residues remaining covalently bound to the 5' ends of the cut site and preventing adapter ligation and sequencing. This characteristic signal was used to algorithmically detect genome-wide topo IV cleavage sites. Using this approach 88 sites, common across three different experiments, were identified. As expected, topo IV cleavage at the *dif* site was the most enriched. Interestingly, most NorfIIP sites were not observed in the ChIP-seq binding experiments, indicating that not only does binding not strongly correlate with cleavage, but that cleavage does not strongly correlate with binding. No strong consensus sequence preference for topo IV cleavage was detected, beyond a slight bias for GC dinucleotides and an increased spacing between GATC motifs around the cleavage sites.

1.6. DNA Gyrase

In 2018, an NGS approach was developed that allowed the precise identification of gyrase cleavage sites genome-wide in *E. coli*, facilitating in depth analysis of drug driven cleavage in vivo [18]. The method, named Topo-seq (Figure 2C), involved trapping gyrase cleavage complexes using either oxolinic acid (oxo), ciprofloxacin (cfx), or microcin B-17 (MccB17) and immunoprecipitating the resulting fragments. They then underwent a single-strand library preparation protocol that only ligates adapters to the 3' free strand from the cut site. This method was developed to avoid issues caused by the 5' covalent linkage to the active site tyrosine of gyrase. Proteinase K treatment leaves the phosphotyrosyl residue at the 5' cut site, which drastically reduces ligation efficiency of the sequencing adapters. In previous studies, the use of CIP or TDP2 offered an enzymatic solution. However, the efficiency of these enzymes is not always optimal, leading to the introduction of potential artifacts. In topo-seq, this effect is mitigated entirely by solely focusing on the unbound 3' single strand. By sequencing only these strands from either side of the cut site, the expected characteristic signal of gyrase cleavage involves a bimodal peak with a sharp 4 bp sequence loss in the center, indicative of gyrase cleavage. This unique signal allowed gyrase cleavage sites to be accurately called and subsequently analyzed.

Using this approach, Sutormin and colleagues identified 4635 gyrase cleavage sites (GCSs) when cells were treated with cfx, 5478 when treated with oxo and 732 when treated with MccB17. An in-depth comparison among these inhibitors and their cleavage sites revealed not only inhibitor-specific cleavage

patterns, but also inhibitor concentration-dependent effects. Having identified a wealth of cleavage sites, robust analysis of the gyrase sequence preference was facilitated, finding the gyrase binding motif to be both extensive and degenerate. This motif is around 130 bp in length, symmetrical on either side of the cleavage site and present in all cleavage sites, regardless of the inhibitor used. It exhibits a periodic GC content fluctuation, reminiscent of the binding pattern seen for eukaryotic nucleosomes. The only inhibitor-specific differences identified were associated with the precise cleavage site. The tendency for cfx and oxo to intercalate and drive cleavage at guanine nucleotides was confirmed. However, the sequence preference for MccB17 exhibited a variable pattern and was unlike oxo and cfx. This likely reflects the fact that oxo and cfx have a different mode of action to MccB17.

Gyrase showed a strong cleavage preference for highly transcribed loci, producing the most distinct enrichment of GCSs around the rRNA operons, known to have high rates of transcription. Furthermore, the use of rifampicin, a transcriptional inhibitor, not only decreased the number of cfx-dependent cleavage sites by half, but also produced significant relocation of GCSs away from rRNA operons and decreased gyrase avoidance of poorly transcribed regions. Together, these data implicate *E. coli* gyrase directly in topology manipulation during transcription.

1.7. DNA Topoisomerase IA

The first type IA topoisomerase to be explored using NGS was done in concert with gyrase and RNA polymerase (RNAP) from *Mycobacterium tuberculosis* [40]. Whilst *M. tuberculosis* is a well-known and devastating human pathogen, with resistant strains beginning to dominate in some areas of the world [41], it is also a curious bacterial species in regard to topoisomerase encoding. Whilst the majority of bacterial species encode 4 topoisomerases: two type IA topoisomerases (topo I and topo III) and two type IIA topoisomerases (DNA gyrase and topo IV), *M. tuberculosis* has only one type IA (topo I) and one type IIA (DNA gyrase) [42]. The rationale behind this study was to obtain direct evidence of topo I and DNA gyrase interacting at sites of transcription, to further support the twin-supercoiled domain model of transcription and the role that topoisomerases play in removing accumulated torsional stress in the DNA caused by the transcriptional machinery. Hypothesized by Liu and Wang in 1987, the twin-supercoiled domain model describes how the separation of the DNA strands during transcription and replication would lead to positive supercoiling ahead of the protein machinery and negative supercoiling behind [43]. Indeed, numerous in vitro studies support this hypothesis, demonstrating that transcription alters DNA topology [44]; inhibition of DNA gyrase leads to positive supercoil accumulation [45] and mutation of topo I causes negative supercoil accumulation [46]. However, direct evidence of the association of bacterial topoisomerases and the transcriptional machinery was not obtained until a ChIP-seq-based protocol was implemented genome wide using *M. tuberculosis* [40]. ChIP-seq involves the crosslinking of protein to DNA, the DNA of which is enriched for fragments bound to a protein of interest, usually using a protein-specific antibody. So, the data set produced by the subsequent sequencing of this DNA library generates a binding profile for the protein, rather than a direct read out of cleavage activity.

In this study, the binding of *M. tuberculosis* gyrase and topo I were compared to the binding of RNAP. In agreement with the twin-supercoiled domain model, both topo I and gyrase were found to be specifically enriched in regions of transcriptional activity, with a high correlation with RNAP binding. As expected, RNAP was enriched at promoter sequences with topo I enrichment proximally located and gyrase enrichment both overlapping with RNAP and downstream. In addition, both gyrase and topo I were highly enriched at the origin of replication, directly implicating them in replication, whilst only gyrase displayed a binding peak at the termination site. This is likely due to *M. tuberculosis* gyrase having a dual activity of both negative supercoiling and efficient decatenation [47], the latter being a role usually fulfilled by topo IV in other bacteria.

In 2019, another study from the same group, explored topo IA activity in more depth, contrasting binding and cleavage in *Mycobacterium smegmatis*, a close relative of *M. tuberculosis* [48]. In this study, four libraries were produced, three to probe topo IA cleavage, and another for topo IA binding. The first

of the topo IA cleavage protocols involved the use of a poisonous topo IA mutant, with a substitution in the topoisomerase-primase (TOPRIM) domain (D108A), previously shown to cause accumulation of cleavage complexes [49]. Once transformed into competent *M. smegmatis* cells, an increase in cleavage complexes and stimulation of the SOS response were detected, demonstrating that the mutant protein was being expressed and interacting with the genome. The second technique designed to probe topo IA cleavage implemented the use of a newly discovered topo IA inhibitor, imipramine [50]. Imipramine is a tricyclic antidepressant that has been shown to act specifically on topo IA from mycobacteria and prevent re-ligation, thereby increasing DNA cleavage complexes. The third technique, designed to probe binding, followed a traditional ChIP-seq type method. They also managed to capture wild-type topo IA cleavage sites, in the absence of inhibitors. In all four cases, topo IA-specific DNA fragments were immunoprecipitated and sequenced using the Illumina Genome Analyzer Iix.

Being able to contrast binding and cleavage, as with topo IV [16,17], revealed a similar disconnect between binding and catalytic activity. Whilst binding was shown genome wide and somewhat evenly distributed, for topo IA, the cleavage profiles were far more region specific. Multiple regions displayed topo IA binding enrichment where cleavage was completely absent. In vitro assays demonstrated that the enriched sequence was cleavable. However, topo IA inhibition by Nucleoid Associated Proteins (NAPs) measured in vitro suggests that the inhibition of topo IA cleavage in vivo could arise from the presence of NAPs. A unique feature amongst type IA topoisomerases is the mycobacterial topo IA sequence preference for both binding and cleavage, which was identified in previous in vitro studies [51–53]. Analysis of the genome-wide cleavage sites of *M. smegmatis* topo IA revealed a similar consensus sequence as previously established for topo IA in vitro, indicating that topo IA binds and cleaves DNA in a sequence-dependent manner in vivo.

As with the previous study looking at topo IA and gyrase co-localization in *M. tuberculosis*, topo IA in *M. smegmatis* exhibited increased occupancy and cleavage associated with RNAP and highly transcribed genes. However, in *M. smegmatis* both cleavage and binding were detected at the termination site, a result not previously seen, implicating topo IA in chromosome segregation and suggesting that, like mycobacterial gyrase, topo IA also has dual activities in both relaxation and decatenation. In most bacterial species the dedicated type IA topo involved in decatenation is topo III, an enzyme absent in mycobacterial species. In vitro assays performed in this study did indeed show that *M. smegmatis* topo IA had significant catenation and decatenation activity, and therefore may also be involved in chromosome segregation along with gyrase, in vivo.

1.8. DNA Topoisomerase IB

Currently only one study has mapped genome-wide activity of human topo IB, a type IB topoisomerase [19]. Using HCT116 human colon cancer cells, topo IB binding was assayed using a ChIP-seq method, and cleavage by a method termed Top1-seq. Top1-seq involved the use of camptothecin, a topo IB inhibitor, to generate covalent cleavage complexes, which were enriched using immunoprecipitation and then sequenced.

Topo IB binding enrichment was highly correlated with transcription, with 67% of DNA-bound topo IB interacting with transcribed genes both upstream of the TSS and downstream of the transcription termination site. By quantifying the transcription level of genes using RNA-seq, a method developed to allow sequencing of the transcriptome [54], increased topo IB occupancy was shown to correlate with increased levels of transcription, providing further evidence for its role in transcriptional topology regulation in eukaryotic cells, much like topo IA in prokaryotes. Comparison of the binding landscape of topo IB to RNAPII revealed their localization was highly correlated, in a manner that suggested a direct protein-protein interaction, rather than topo IB being solely bound to regions of torsional stress within the genome. To explore this, cellular extracts were subjected to a pull-down assay, which confirmed the direct interaction between topo IB and RNAPII in the absence of DNA, indicating that topo IB is an integral member of the transcriptional machinery and is recruited to sites of RNAPII engagement.

As mentioned, this study used a technique called Top1-seq, which was designed to capture catalytically engaged topo IB, using a brief camptothecin treatment. Whilst binding of topo IB was high at promoters, particularly in concert with RNAPII, catalytically engaged topo IB was diminished at TSS but enriched within gene bodies, again with a positive correlation with transcription level. Cleavage was also biased towards the template strand, indicating that topo IB activity is coupled to the movement of the transcriptional machinery. Using complementary *in vitro* assays, RNAPII was found to enhance topo IB relaxation activity and increase processivity under physiological ionic conditions, whilst none of the other general transcription factors seemed to have an effect on topo IB relaxation activity. The interaction between RNAPII and topo IB was mapped to the CTD of RNAPII and the NTD of topo IB. An interaction that increased the time topo IB spends in the cleaved state, enhancing the processive removal of supercoils.

2. Conclusions

Since the advent of NGS technologies in 2005, numerous techniques have been developed to explore genome-wide DNA binding and cleavage landscapes, which has both answered and posed new questions concerning *in vivo* topoisomerase behavior. Encouraging evidence in support of the widely accepted twin-domain model, has established a greater understanding of topoisomerase activity during transcription and to a lesser extent replication. Topo II and topo I in eukaryotes exhibit higher levels of activity in the vicinity of highly transcribed genes [14,19], as has also been shown for mycobacterial topo I and gyrase [48]. Topo IV, on the other hand, demonstrates a clear preference for the *dif* site in *E. coli* [16,17], supporting the idea that decatenation is its primary cellular function [55].

Using NGS as a means to explore topo cleavage across the genome, encouraged innovation in protocol development to circumvent issues which arise as a result of the covalent linkage between the protein and the DNA backbone. Proteinase K will remove the majority of the protein but will not cleave the phosphotyrosyl bond, leaving the adduct to interfere with sequencing adapter ligation. Two enzymatic solutions have been used to remove the adduct, namely CIP and TDP2, the first of which is very inefficient and the second far less commercially accessible. Using a single-strand DNA library preparation kit, as Sutormin and colleagues did when exploring DNA gyrase activity [18], offers another elegant solution by only sequencing the free 3' DNA strand.

One particularly intriguing result has come from contrasting topoisomerase binding with cleavage activity. For some proteins, e.g., topo IV and topo IA, the correlation between them is so low that most binding sites do not constitute cleavage sites and vice versa. The nature of the techniques used to explore binding versus cleavage (such as using cleavage complex stabilizing agents which bias cleavage preference) and the platforms used to sequence the fragments, may contribute to this discrepancy. However, *in vitro* biochemical data also suggest that the relationship between cleavage and binding is not so clear. For example, topo II α whilst being more catalytically active on positively supercoiled DNA, maintains higher levels of cleavage on negatively supercoiled DNA [56]. This raises questions as to the determinants of topo binding and cleavage sites; how do the factors known to influence binding and activity including DNA topology and sequence play off against protein-mediated recruitment or inhibition *in trans*, and what other factors may influence topo binding and cleavage *in vivo*. It is clear the answer is not straightforward, with topo IB being directly recruited as part of the transcriptional machinery [19] and mycobacterial topo IA [48] and prokaryotic DNA gyrase [18] exhibiting more substantial DNA sequence preferences. As with all cellular processes, the mechanisms governing DNA topology maintenance are often not only protein specific but also species specific. A difficulty with collecting such vast amounts of data is the interpretation thereof, which has been challenging in some cases due to the complexity of the cellular environment. One possible approach would be the use of *in vitro* NGS techniques to map topo binding and cleavage in which DNA topology and sequence can be systematically controlled in addition to the use of specific proteins and precise buffer conditions, including inhibitor addition. Such an approach would complement genome-wide studies and could provide insights that could aid in the interpretation of genome-wide results. NGS is undoubtedly

a very powerful technique, both in isolation and as part of an experimental repertoire, which will hopefully be further implemented in the field of DNA topoisomerases.

Funding: This research was funded by the Intramural Program of the National Heart, Lung, and Blood Institute of the National Institutes of Health (K.C.N. and S.J.M.), and through a Wellcome Trust and National Institutes of Health fellowship to S.J.M.

Acknowledgments: The authors thank Rachel Kim for comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, J.C. DNA topoisomerases. *Annu. Rev. Biochem.* **1996**, *65*, 635–692. [[CrossRef](#)]
2. Bush, N.G.; Evans-Roberts, K.; Maxwell, A. DNA Topoisomerases. *EcoSal Plus* **2015**, *6*. [[CrossRef](#)]
3. Anderson, V.E.; Osheroff, N. Type II topoisomerases as targets for quinolone antibacterials: Turning Dr. Jekyll into Mr. Hyde. *Curr. Pharm. Des.* **2001**, *7*, 337–353. [[CrossRef](#)] [[PubMed](#)]
4. Nitiss, J.L. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat. Rev. Cancer* **2009**, *9*, 338–350. [[CrossRef](#)] [[PubMed](#)]
5. Pommier, Y.; Sun, Y.; Huang, S.N.; Nitiss, J.L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 703–721. [[CrossRef](#)] [[PubMed](#)]
6. Ratain, M.J.; Rowley, J.D. Therapy-related acute myeloid leukemia secondary to inhibitors of topoisomerase II: From the bedside to the target genes. *Ann. Oncol.* **1992**, *3*, 107–111. [[CrossRef](#)] [[PubMed](#)]
7. Pui, C.H.; Relling, M.V. Topoisomerase II inhibitor-related acute myeloid leukaemia. *Br. J. Haematol.* **2000**, *109*, 13–23. [[CrossRef](#)] [[PubMed](#)]
8. Schmidt, D.; Wilson, M.D.; Spyrou, C.; Brown, G.D.; Hadfield, J.; Odom, D.T. ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* **2009**, *48*, 240–248. [[CrossRef](#)]
9. Crosetto, N.; Mitra, A.; Silva, M.J.; Bienko, M.; Dojer, N.; Wang, Q.; Karaca, E.; Chiarle, R.; Skrzypczak, M.; Ginalski, K.; et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **2013**, *10*, 361–365. [[CrossRef](#)]
10. Canela, A.; Sridharan, S.; Sciascia, N.; Tubbs, A.; Meltzer, P.; Sleckman, B.P.; Nussenzweig, A. DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol. Cell* **2016**, *63*, 898–911. [[CrossRef](#)]
11. Canela, A.; Maman, Y.; Huang, S.N.; Wutz, G.; Tang, W.; Zagnoli-Vieira, G.; Callen, E.; Wong, N.; Day, A.; Peters, J.M.; et al. Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. *Mol. Cell* **2019**, *75*, 252–266.e8. [[CrossRef](#)]
12. Pan, J.; Sasaki, M.; Kniewel, R.; Murakami, H.; Blitzblau, H.G.; Tischfield, S.E.; Zhu, X.; Neale, M.J.; Jasin, M.; Socci, N.D.; et al. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **2011**, *144*, 719–731. [[CrossRef](#)]
13. Rigby, P.W.; Dieckmann, M.; Rhodes, C.; Berg, P. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **1977**, *113*, 237–251. [[CrossRef](#)]
14. Baranello, L.; Kouzine, F.; Wojtowicz, D.; Cui, K.; Przytycka, T.M.; Zhao, K.; Levens, D. DNA break mapping reveals topoisomerase II activity genome-wide. *Int. J. Mol. Sci.* **2014**, *15*, 13111–13122. [[CrossRef](#)] [[PubMed](#)]
15. Gittens, W.H.; Johnson, D.J.; Allison, R.M.; Cooper, T.J.; Thomas, H.; Neale, M.J. A nucleotide resolution map of Top2-linked DNA breaks in the yeast and human genome. *Nat. Commun.* **2019**, *10*, 4846. [[CrossRef](#)] [[PubMed](#)]
16. El Sayyed, H.; Espeli, O. Mapping *E. coli* Topoisomerase IV Binding and Activity Sites. *Methods Mol. Biol.* **2018**, *1703*, 87–94. [[CrossRef](#)] [[PubMed](#)]
17. El Sayyed, H.; Le Chat, L.; Lebailly, E.; Vickridge, E.; Pages, C.; Cornet, F.; Cosentino Lagomarsino, M.; Espeli, O. Mapping Topoisomerase IV Binding and Activity Sites on the *E. coli* Genome. *PLoS Genet.* **2016**, *12*, e1006025. [[CrossRef](#)] [[PubMed](#)]
18. Sutormin, D.; Rubanova, N.; Logacheva, M.; Ghilarov, D.; Severinov, K. Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the *Escherichia coli* genome. *Nucleic Acids Res.* **2019**, *47*, 1601. [[CrossRef](#)]
19. Baranello, L.; Wojtowicz, D.; Cui, K.; Devaiah, B.N.; Chung, H.J.; Chan-Salis, K.Y.; Guha, R.; Wilson, K.; Zhang, X.; Zhang, H.; et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **2016**, *165*, 357–371. [[CrossRef](#)]

20. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [[CrossRef](#)]
21. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, *2012*, 251364. [[CrossRef](#)] [[PubMed](#)]
22. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945. [[CrossRef](#)]
23. Van Dijk, E.L.; Auger, H.; Jaszczyszyn, Y.; Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **2014**, *30*, 418–426. [[CrossRef](#)] [[PubMed](#)]
24. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376–380. [[CrossRef](#)] [[PubMed](#)]
25. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141. [[CrossRef](#)] [[PubMed](#)]
26. Friedberg, E.C. DNA damage and repair. *Nature* **2003**, *421*, 436–440. [[CrossRef](#)] [[PubMed](#)]
27. Jackson, S.P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071–1078. [[CrossRef](#)]
28. Keeney, S. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dyn. Stab.* **2008**, *2*, 81–123. [[CrossRef](#)]
29. Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **2007**, *89*, 285–307. [[CrossRef](#)]
30. Muslimovic, A.; Nystrom, S.; Gao, Y.; Hammarsten, O. Numerical analysis of etoposide induced DNA breaks. *PLoS ONE* **2009**, *4*, e5859. [[CrossRef](#)]
31. Uuskula-Reimand, L.; Hou, H.; Samavarchi-Tehrani, P.; Rudan, M.V.; Liang, M.; Medina-Rivera, A.; Mohammed, H.; Schmidt, D.; Schwalie, P.; Young, E.J.; et al. Topoisomerase II β interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **2016**, *17*, 182. [[CrossRef](#)] [[PubMed](#)]
32. Yang, X.; Li, W.; Prescott, E.D.; Burden, S.J.; Wang, J.C. DNA topoisomerase II β and neural development. *Science* **2000**, *287*, 131–134. [[CrossRef](#)] [[PubMed](#)]
33. Li, Y.; Hao, H.; Tzatzalos, E.; Lin, R.K.; Doh, S.; Liu, L.F.; Lyu, Y.L.; Cai, L. Topoisomerase II β is required for proper retinal development and survival of postmitotic cells. *Biol. Open* **2014**, *3*, 172–184. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, Y.L.; Yu, C.; Ji, S.Y.; Li, X.M.; Zhang, Y.P.; Zhang, D.; Zhou, D.; Fan, H.Y. TOP2 β is essential for ovarian follicles that are hypersensitive to chemotherapeutic drugs. *Mol. Endocrinol.* **2013**, *27*, 1678–1691. [[CrossRef](#)] [[PubMed](#)]
35. Meyer-Ficca, M.L.; Lonchar, J.D.; Ihara, M.; Meistrich, M.L.; Austin, C.A.; Meyer, R.G. Poly(ADP-ribose) polymerases PARP1 and PARP2 modulate topoisomerase II β (TOP2B) function during chromatin condensation in mouse spermiogenesis. *Biol. Reprod.* **2011**, *84*, 900–909. [[CrossRef](#)] [[PubMed](#)]
36. Daev, E.; Chaly, N.; Brown, D.L.; Valentine, B.; Little, J.E.; Chen, X.; Walker, P.R. Role of topoisomerase II in the structural and functional evolution of mitogen-stimulated lymphocyte nuclei. *Exp. Cell Res.* **1994**, *214*, 331–342. [[CrossRef](#)] [[PubMed](#)]
37. Yu, X.; Davenport, J.W.; Urtishak, K.A.; Carillo, M.L.; Gosai, S.J.; Kolaris, C.P.; Byl, J.A.W.; Rappaport, E.F.; Osheroff, N.; Gregory, B.D.; et al. Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res.* **2017**, *27*, 1238–1249. [[CrossRef](#)]
38. Gomez-Herreros, F.; Schuurs-Hoeijmakers, J.H.; McCormack, M.; Grealley, M.T.; Rulten, S.; Romero-Granados, R.; Counihan, T.J.; Chaila, E.; Conroy, J.; Ennis, S.; et al. TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function. *Nat. Genet.* **2014**, *46*, 516–521. [[CrossRef](#)]
39. Kuempel, P.L.; Henson, J.M.; Dircks, L.; Tecklenburg, M.; Lim, D.F. dif, a recA-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol.* **1991**, *3*, 799–811.
40. Ahmed, W.; Sala, C.; Hegde, S.R.; Jha, R.K.; Cole, S.T.; Nagaraja, V. Transcription facilitated genome-wide recruitment of topoisomerase I and DNA gyrase. *PLoS Genet.* **2017**, *13*, e1006754. [[CrossRef](#)]
41. Glaziou, P.; Floyd, K.; Raviglione, M.C. Global Epidemiology of Tuberculosis. *Semin. Respir. Crit. Care Med.* **2018**, *39*, 271–285. [[CrossRef](#)] [[PubMed](#)]
42. Nagaraja, V.; Godbole, A.A.; Henderson, S.R.; Maxwell, A. DNA topoisomerase I and DNA gyrase as targets for TB therapy. *Drug Discov. Today* **2017**, *22*, 510–518. [[CrossRef](#)] [[PubMed](#)]

43. Liu, L.F.; Wang, J.C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7024–7027. [[CrossRef](#)]
44. Tsao, Y.P.; Wu, H.Y.; Liu, L.F. Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell* **1989**, *56*, 111–118. [[CrossRef](#)]
45. Lockshon, D.; Morris, D.R. Positively supercoiled plasmid DNA is produced by treatment of *Escherichia coli* with DNA gyrase inhibitors. *Nucleic Acids Res.* **1983**, *11*, 2999–3017. [[CrossRef](#)]
46. Pruss, G.J.; Drlica, K. Topoisomerase I mutants: The gene on pBR322 that encodes resistance to tetracycline affects plasmid DNA supercoiling. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8952–8956. [[CrossRef](#)]
47. Manjunatha, U.H.; Dalal, M.; Chatterji, M.; Radha, D.R.; Visweswariah, S.S.; Nagaraja, V. Functional characterisation of mycobacterial DNA gyrase: An efficient decatenase. *Nucleic Acids Res.* **2002**, *30*, 2144–2153. [[CrossRef](#)]
48. Rani, P.; Nagaraja, V. Genome-wide mapping of Topoisomerase I activity sites reveal its role in chromosome segregation. *Nucleic Acids Res.* **2019**, *47*, 1416–1427. [[CrossRef](#)]
49. Bhat, A.G.; Leelaram, M.N.; Hegde, S.M.; Nagaraja, V. Deciphering the distinct role for the metal coordination motif in the catalytic activity of *Mycobacterium smegmatis* topoisomerase I. *J. Mol. Biol.* **2009**, *393*, 788–802. [[CrossRef](#)]
50. Godbole, A.A.; Ahmed, W.; Bhat, R.S.; Bradley, E.K.; Ekins, S.; Nagaraja, V. Targeting *Mycobacterium tuberculosis* topoisomerase I by small-molecule inhibitors. *Antimicrob. Agents Chemother.* **2015**, *59*, 1549–1557. [[CrossRef](#)]
51. Bhaduri, T.; Bagui, T.K.; Sikder, D.; Nagaraja, V. DNA topoisomerase I from *Mycobacterium smegmatis*. An enzyme with distinct features. *J. Biol. Chem.* **1998**, *273*, 13925–13932. [[CrossRef](#)] [[PubMed](#)]
52. Sikder, D.; Nagaraja, V. Determination of the recognition sequence of *Mycobacterium smegmatis* topoisomerase I on mycobacterial genomic sequences. *Nucleic Acids Res.* **2000**, *28*, 1830–1837. [[CrossRef](#)] [[PubMed](#)]
53. Sikder, D.; Nagaraja, V. A novel bipartite mode of binding of *M. smegmatis* topoisomerase I to its recognition sequence. *J. Mol. Biol.* **2001**, *312*, 347–357. [[CrossRef](#)] [[PubMed](#)]
54. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
55. Peng, H.; Marians, K.J. Decatenation activity of topoisomerase IV during oriC and pBR322 DNA replication in vitro. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8571–8575. [[CrossRef](#)]
56. McClendon, A.K.; Rodriguez, A.C.; Osheroff, N. Human topoisomerase IIalpha rapidly relaxes positively supercoiled DNA: Implications for enzyme action ahead of replication forks. *J. Biol. Chem.* **2005**, *280*, 39337–39345. [[CrossRef](#)]

