HHS Public Access

Author manuscript

Nat Rev Cancer. Author manuscript; available in PMC 2023 August 13.

Published in final edited form as:

Nat Rev Cancer. 2022 September ; 22(9): 533–546. doi:10.1038/s41568-022-00488-9.

Structural Variations in Cancer and the 3D Genome

Frank Dubois[1,2], **Nikos Sidiropoulos**[3,4], **Joachim Weischenfeldt**[3,4,5,*], **Rameen Beroukhim**[1,2,6,7,*]

[1]Departments of Cancer Biology and Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[2]Harvard Medical School, Boston, MA 02215, USA

[3]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Ole Maaløes Vej 5, 2200, Copenhagen N, Denmark

[4]The Finsen Laboratory, Rigshospitalet, Ole Maaløes Vej 5, 2200, Copenhagen N, Denmark

[5]Department of Urology, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[7]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

Abstract

Structural variations (SVs) affect more of the cancer genome than any other type of somatic genetic alteration but difficulties in detecting and interpreting them have limited our understanding. Clinical cancer sequencing also increasingly aims to detect SVs, leading to a widespread necessity to interpret them. Recently, analyses of large whole-genome sequencing datasets revealed features that impact rates of SVs across the genome in different cancers. A striking feature has been the extent to which, in both their generation and their impacts on the selective fitness of cancer cells, SVs are more specific to individual cancer types than other genetic alterations such as single nucleotide variants. This review discusses how the folding of the 3D genome, and differences in its folding across cell types, affect observed SV rates in different cancers.

1. Importance and tissue-type specificity of cancer SVs

Among the three types of genetic alteration in the cancer genome – single nucleotide variants (SNVs), structural variants (SVs), and exogenous DNA – SVs affect by far the largest fraction of the cancer genome. SVs broadly represent two classes of alteration: whole-chromosome or genome-scale aneuploidies, and rearrangements between distant genomic loci. All copy-number alterations result from SVs, and the first three genetically targeted therapeutics in cancer targeted oncogenes activated by SVs: PML-RARA in the case of ATRA[1] amplified ERBB2 in the case of trastuzumab[2], and BCR-ABL1[3] in the

[*]**Corresponding authors** (these authors jointly supervised this work): Joachim Weischenfeldt: joachim.weischenfeldt@bric.ku.dk; Rameen Beroukhim: rameen_beroukhim@dfci.harvard.edu.

case of imatinib. All three of these SVs have a direct effect (gene fusion or amplification) on the DNA sequence of the oncogene, which spurred their detection and interpretation. While this approach was consequently expanded to develop targeted therapies against oncogenes with SNVs in their coding sequence, the vast majority of oncogenic SVs lie outside of the exome and remained mostly overlooked until now. With recent improvements in sequencing technologies (Box 1) and in our understanding of the epigenetic features that shape the three-dimensional genome, however, we are now detecting more SVs[4-6] and more oncogenic SVs[7-9]. This is important both for cancer genome discovery--developing a better understanding of the oncogenic events that lead to cancer--and for clinical diagnostics, SVs can serve as important biomarkers to indicate patients who are likely to respond to targeted therapeutics. Indeed, whole-genome sequencing is now beginning to enter clinical practice [10,11,12].

Here, we focus on rearrangements rather than aneuploidies, and use “SV” to represent only the former. These SVs can be further classified as deletions, duplications, inversions, interchromosomal translocations, and insertions (Figure 1A), as well as any combination of these. As long as the SV remains isolated it can be fairly easy to assign it to one of these classes. However, a large share of SVs involve several simultaneous breaks and re-ligation. This creates complex SVs, which are harder to interpret (Figure 1B, Box 1).

In contrast to SNVs, SVs can often affect multiple genes, often through resulting copy number aberrations, which affect one-third of the genome in the average cancer[13]. However, the vast majority of SVs occur outside of the small fraction of the genome that is represented in exome sequencing. Sensitive detection requires high coverage whole genome sequencing (WGS) – and even then, standard short-read sequencing can result in high false-negative rates, particularly for complex SVs and SVs in genomic loci that are difficult to map to the genome. However, recent advances in WGS, long- and linked-read sequencing, and availability of orthogonal epigenetic data are beginning to shed light on their prevalence and distribution (Box 1). Even after successful detection, the interpretation of SVs as driver vs passenger events has been complicated due to unknown rates at which SVs naturally occur across the genome due to mechanistic biases (their “background rate”). Additionally, even individual simple SVs can have a vast array of consequences including copy-number changes that extend over many megabases and simultaneous disruption of protein sequences and cis-regulatory elements, for example resulting in the simultaneous disruption of tumor suppressors and activation of oncogenes[8]. These same considerations, however--the manifold and extensive effects of individual SVs--add to the importance of understanding them. This is increasingly possible as we gain the ability to resolve and interpret the mechanistic and selective processes that determine where and why these alterations are observed. Only recently, however, have we begun to understand the role of epigenetics in shaping these mechanisms and the SVs that result[8,14,15].

While many SVs in germline DNA[16,17] have been linked to disease[18,19] the SV landscape in a typical cancer genome is distinctly influenced by somatic processes[5,20,21]. Therefore, we will exclusively focus on somatically acquired SVs in this review. Germline SVs have recently been reviewed elsewhere[22].

In contrast to SNVs, the rates at which SVs are observed in cancer appear to be more heavily influenced by the 3D genome[5]. The very nature of SVs, connecting distant loci, suggests such a relationship: the frequency with which genetic loci are fused by SVs is inversely proportional to the linear distance between them, which in the context of normal DNA has been modeled by a knot-free 3D conformation called a fractal globule[20,23,24]. Their consequences are also often due in part to their effects on the 3D conformation of DNA. For example, driver fusion events, which rely largely on the hijacking of regulatory components to spur oncogene expression, have been observed in over a quarter of cancers[21]. In addition, extrachromosomal DNA (ecDNA), formed by SVs, has been detected in almost half of cancers[25,26] and involves changes in DNA conformation to bring tissue-type specific regulatory regions in close vicinity to oncogenes[15,27].

Perhaps for these reasons, specific 3D features of individual cell types appear to matter more for SVs. For example, whereas BRAF V600E mutations are frequently observed across many cancer types, oncogenic BRAF-KIAA1549 rearrangements have almost exclusively been observed in a single cancer type: juvenile pilocytic astrocytoma[5]. Likewise, whereas TMPRSS2-ERG rearrangements are common in prostate cancers, they are almost never observed in other cancers[28]. The major genomic differences between the cell types that generate these cancers are epigenetic (meaning “molecules and mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence”[29]). Much of the epigenetic state of a cell is represented by its chromatin conformation and three-dimensional organization of DNA. This review describes mechanisms by which this 3D conformation of the genome affects rates at which SVs are observed–the way the 3D genome “shapes” SVs in cancer–including through its impact on rates of SV formation and on SV effects on cellular fitness.

2. An emerging understanding of the 3D genome

a. The pre-cancerous genome

Recently, methodologies have made it possible to infer the nuclear distance between any two loci in the genome. For example, sequencing-based methods such as **Hi-C**[30], a technique that assesses chromatin interactions by proximity-mediated ligation and sequencing, have now made it possible to study nuclear organization genome-wide at fine scale. It has also become possible to obtain microscopy-based distance-metrics for several regions simultaneously[31], which can provide insights into the extent to which regions that are distant at the linear scale but close in nuclear distance tend to form SVs.

Analyses of these data have revealed general folding principles of the genome, which are important to understanding how its 3D structure impacts SV formation in cancer. Overall, the likelihood that two genomic loci are in contact is inversely proportional to the linear distance between them [24]. This has been attributed to a “fractal globule” (alternatively called “crumpled globule”) 3D conformation of DNA, which forms through local regions crumpling iteratively across each chromosome[32]. Hi-C has also revealed higher-resolution structures of DNA, including chromatin compartments and **topologically associated domains (TADs)**, both of which affect the frequency of SVs in cancer.

The genome can be separated into large **chromosome territories** – areas occupied by particular chromosomes[33] – and smaller chromatin compartments into which they subdivide. When evaluating the latter, Rao et al identified 'A' and 'B' compartments corresponding to gene-rich, actively transcribed euchromatin associated with early replication timing, primarily located in the centre of the nucleus; and gene-poor, repressive, heterochromatin, preferentially located at the nuclear periphery (Figure 2A)[24,34-36]. The separation of A and B compartments has been proposed to reflect **phase separation** between active and inactive chromatin based on their inherent physical properties[37],

TADs are chromatin domains that are typically 100 kb-1Mb in size and demarcated by architectural proteins including CCCTC binding factor (CTCF), and the cohesin complex (RAD21, SMC1, SMC3, and STAG1/2). TADs contain extensive chromatin looping such as promoter-enhancer interactions around these CTCF/cohesin "**chromatin loop anchors**", that are likely to facilitate the regulation of gene expression[38-40]. The orientation of the CTCF motif dictates the probability with which nearby loci are spatially juxtaposed. Hi-C data have shown that most CTCF-associated loops are in a convergent orientation (Figure 2B)[38,41-43], which is compatible with a **loop-extrusion model**, by which DNA is pulled through a ring-shaped structure composed of CTCF-Cohesin complex[44]. Interestingly, recent studies have shown that a substantial proportion of enhancer-promoter loops occur independently of CTCF binding. The rules governing how such interactions occur are not well-understood

Although the 3D conformation of the genome is not equal to its epigenetic state, they are heavily interdependent. Histone marks have classically been used to identify epigenetic features such as enhancers, promoters, and open and closed chromatin[45]. In Hi-C data, interaction loops appear as connections in 3D between elements such as enhancers and promoters[46]. Chromatin regions represented as A and B compartments can also be identified in Hi-C data by their characteristic patterns[24] as well as using open chromatin data and methylation data [47]. For these reasons, our review describes both explicit relations between SVs and the 3D genome, and implicit relations indicated by relationships between SVs and epigenetic features of the genome.

b. Effects of SVs on the 3D genome

The 3D organization of the genome exhibits substantial and dynamic heterogeneity at the single-cell and even allelic level, which is shaped both by cellular states and stochasticity[48,49] and can be deregulated in cancer through SVs.

A major role of SVs in shaping the 3D genome is by altering interactions between CTCF-dependent and independent DNA loops. This can occur in three ways (Figure 2C). First, if a CTCF anchor is deleted, the two adjacent TADs will fuse. Second, duplications of CTCF anchors will create new TADs ("neo-TADs"). Third, inversions encompassing CTCF anchors will swap DNA territory between TADs. Translocations between chromosomes can either delete CTCF sites, thereby fusing TADs across chromosomes, or maintain CTCF sites, thereby swapping DNA territory between TADs on these different chromosomes. Dysregulation of CTCF leads to altered enhancer-promoter looping and gene expression in cancer[50,51].

SVs may also alter the global positioning of DNA territories within the nucleus. CRISPR engineering has shown that removal of cis-elements can lead to changes in compartment structure and DNA activation or repression[52,53] phenomena that have been associated with SVs in cancer[54] and reviewed in[55]. DNA movement within the nucleus can also alter its relation to the nuclear membrane, thereby contributing to transcriptional repression reviewed in[56]; it is interesting to speculate that adjacency to the nuclear membrane may also contribute to further DNA damage. Excess curvature of nuclear membranes in micronuclei and anaphase bridges has been linked to DNA damage and complex SVs[57-59]

These effects of SVs in cancer genomes can modify transcription in multiple ways, including by altering copy-number and changing DNA compartment structure. The effects of SVs on chromatin looping also often modify interactions between enhancers, promoters, and other transcriptional regulatory elements, as discussed in detail below.

3. The two major determinants of observed rates of SVs in cancer: mechanistic biases and effects on evolutionary selection

The frequency with which SVs are observed in cancer is determined both by the rates at which they occur across cells in the body (mechanistic biases) and by their effects on cellular fitness, which lead cancer genetic studies to observe them more or less frequently than this underlying rate (selection biases). UnderstandingUnderstanding this better requires an appreciation of how genetic alterations arise and are enriched during cancer development, which we discuss below. In the following section, we will see how each of these factors, mechanism and cellular fitness, are influenced by the 3D conformation of the genome.

Cancer originates when a normal cell acquires a set of genetic alterations that lead it to grow uncontrollably[60]. These oncogenic, or “**driver**”, alterations arise through a random process of mutagenesis that typically also generates genetic alterations that are either evolutionarily neutral or disadvantageous (Figure 3A). In a simplistic view, tumor cells will continue to grow if the net sum of their mutations on cellular fitness is positive. As a consequence, cells usually carry both neutral and even disadvantageous alterations, maintained as “**passenger events**”--and indeed these may outnumber the driver events by orders of magnitude due to the far greater numbers of genetic alterations with neutral or negative fitness effects. Both driver and passenger events include SNVs, SVs, and sometimes exogenous nucleic acids (e.g. introduced by oncogenic viruses). The driver events are oncogenic either due to changes to the protein-coding sequence of oncogenes or tumor suppressors or altered expression of the encoded proteins (or both).

The prevalence of an SV in cancer is thus determined both by the **mechanistic biases** that determine the probability with which the alteration occurs in the cancer’s cell of origin and by the extent to which effects on fitness in the context of **evolutionary selection** enrich the alteration in cells that reach cancer. The mechanistic biases shape the fraction of cells in a person’s body that acquire a given alteration. The effects on evolutionary fitness impact which of these cells will continue to grow, potentially forming a cancer. Likewise, once a cancer has formed, mechanistic biases continue to shape the fraction of subclones that

generate any specific genetic alteration, whereas the fitness effects impact which of these subclones will grow to dominate the tumor.

Observing a genetic alteration at high frequency in cancer thus may suggest that the alteration was a driver event that increased cellular fitness, but does not provide proof. It is also possible that the alteration occurs frequently during cell division, including in cells that do not become cancer[5,61,62].

4. Mechanistic biases shape where SVs form

a. SV formation understood on the linear genome

Mechanisms of SV formation fall into two major classes: **replication-based mechanisms** such as microhomology-mediated break-induced replication (**MMBIR**), in which a DNA polymerase jumps to a distant locus; and **fusion-based mechanisms** such as non-homologous end-joining **(NHEJ)**, homologous recombination (**HR**), and microhomology-mediated end-joining (**MMEJ**), in which a double-strand break (**DSB**) in DNA is repaired incorrectly, fusing it to a distant locus.

Moreover, these mechanisms can generate single isolated SVs (Figure 1A) or multiple SVs, typically clustered in specific genomic loci, in a single traumatic event (Figure 1B). Such complex events often create DNA topologies that are difficult to interpret, with the potential for multiple distant loci brought into juxtaposition. These complex events can be generated by multiple mechanisms, including **chromothripsis**, **breakage-fusion bridge cycles**, **extrachromosomal amplification**, **chromoanasynthesis**, **templated insertions**, and **chromoplexy**[63].

Some of these complex events, such as chromothripsis, and the related formation of extrachromosomal amplifications and breakage-fusion-bridge events, can reflect DNA breaks that later fuse. In addition, complex events can form due to replication-based mechanisms. An example is **chromoanasynthesis**, which can form from repeated microhomology-mediated template switching and therefore is hypothesized to originate from MMBIR[64-66].

Many genomic features impact which loci break and fuse, including sequence characteristics and replication timing. However, the impact of the 3D structure of the genome on the development of rearrangements is becoming increasingly clear.

b. The genome’s 3D structure impacts where SVs form and fuse

In the context of SVs, mechanistic biases determine both which loci are prone to double-strand breakage and which pairs of loci are prone to combine.

The importance of 3D nuclear distance in cancer SV formation was identified in early, seminal observations. The fusion generating the Philadelphia chromosome[67], which joins *BCR* in chromosome 9 and *ABL1* in chromosome 22 in myeloid leukemias, was demonstrated to be catalyzed at least in part by the 3D proximity of the two loci specifically in hematopoietic cells [68,69]. In contrast, these two loci are more distant in other cell

types, where the fusion gene is not observed. Another well-established example is IGH-MYC, present in 9 out of 10 Burkitt's lymphomas[70]. However, a systematic, genome-wide assessment of the role played by the 3D organization was impossible until only recently.

The organization of the genome has impacts both on which loci form DSBs, and which pairs of loci join to form SVs. Indeed, DSBs are an intrinsic component of the maintenance of the 3D conformation of DNA during transcription. These DSBs are generated by the endonuclease topoisomerase 2B (TOP2B) to relieve tension on the double helix throughout the genome but especially adjacent to CTCF **chromatin loop anchors** (Figure 3B). Perhaps for this reason, sites of active transcription have been shown to be more prone to SV formation[5,71-74]. These actively transcribed sites are also prone to selection, suggesting that the prevalence of SVs near open, active chromatin in cancer genomes may be a consequence of both mechanistic biases and their special effects on cellular fitness. In contrast, SNVs have been shown to primarily accumulate in late-replicating, repressed chromatin [75].

Cell culture systems that trace DSBs through the cell cycle have shed light on processes that make these sites more prone to SVs. Untranscribed regions, heterochromatic chromatin loop anchors of the genome were found to be more often repaired during G1 through error-prone NHEJ, which is the most common source of SVs in cancer[20,76]. In contrast, transcriptionally active sites of DSBs, associated with **"A" compartment**, long-range chromatin interactions, and histone marks of active transcription and HR-activity such as H3K79me2 and H3K36me3 were shown to be preferentially repaired by HR late during S-phase[77,78]. Hence, DNA breaks at sites of active transcription can form clusters of long-range interactions in the nucleus, which are preferentially repaired during replication. This can be reconciled with a **phase-separation** model containing droplets with the accumulation of transcription factors, replication machinery, gH2AX, Ubiquitin, and 53BP1, as well as homologous recombination machinery. Indeed, 53BP1 forms liquid-like clusters around DNA breaks[79] and stabilizes the 3D conformation at sites of DSB by recruiting RIF1 to the TAD boundaries, which may serve to contain and concentrate DNA repair enzymes at sites of DSB[80]. The choice of breakpoint fusion events during replication stress is also thought to be shaped by nuclear proximity[81]. Replication forks that are distant on the linear genome but close in space often exchange templates and form SVs [82,83]. Indeed, Hi-C based analysis of the 3D chromatin before and after induction of *chromothripsis* in a human cell line showed that SVs were significantly enriched between genomically distant loci with similar replication timing zones and compartments, in particular early replication timing and A compartments[84]–which tend to be in close physical proximity[24,34,85].

Perhaps the most important factor influencing which pairs of DSBs fuse is the distance between them (the "span" of the SV). Across all SVs, the frequency with which rearrangements are observed decreases as the reciprocal of their spans. This distribution is the same as the distribution of distances between joined DNA fragments in Hi-C data–which has been interpreted to indicate a fractal globule three-dimensional structure of the genome [5,20,23]. This suggests that the three-dimensional conformation of DNA that determines which loci ligate in Hi-C experiments also determines which DSBs are likely to ligate in SVs in cancer, which corroborates findings from early high-resolution imaging [70,86].

However, rearrangement spans are also influenced by local chromatin compaction (Fig. 3C). Once a DSB forms, its “search” for a site to ligate with appears to be enabled by the mobility of chromatin within the nucleus[87]. DSBs can increase this mobility[88]. Several chromatin changes occur upon DSB, which are thought to have a direct impact on the choice of DSB partner. Following HR, poly(ADP-ribose) polymerase 1 (PARP1) is recruited to the DSB and interacts with the chromatin remodeler *ALC1* to relax the chromatin[89]. ALC1 is frequently upregulated in cancer and is associated with resistance to PARP inhibitors[90]. ATM and ATR kinases phosphorylate the histone variant H2AX (also known as gH2AX) at and around DSB sites. This occurs in conjunction with replacement of histone H1, widespread ubiquitination, and recruitment of the 53BP1 protein[91]. These chromatin changes affect DNA compaction and are thought to induce a change in the flexibility of the chromatin fiber, allowing for a more promiscuous search for partner locus (Fig. 3D). Indeed, sites with gH2AX were shown to partake in more chromatin interactions, both intra- and inter-chromosomal, and these sites were able to form clusters of chromatin[92]. Topologically associated domains are also known to influence somatic SV formation in the non-oncogenic process of VDJ recombination in developing B cells. Here, disruption of CTCF-mediated looping has been found to skew recombination [93,94]

Flexible genomic loci may also allow DSBs both to form more often and ligate with more distant loci. For example, common fragile sites (CFSs) are both prone to SV formation and tend to contain highly flexible regions. Other features enriched in these sites include AT-rich, G-negative chromosomal bands, ALU repeats, and late replication [95]. CFSs tend to involve late-replicating regions with highly transcribed large genes that span TAD boundaries. Indeed, by using polymer physics to describe interactions between chromatin fibers and chromatin-binding proteins without prior knowledge of nuclear proximity, it has been possible to predict many of the DNA interactions and 3D conformations observed in wild-type mouse cells and cells with acquired SVs from experimental Hi-C data [96]. Although this model was only applied to one well-studied 6 Mb locus associated with cancer-related SVs, it does suggest DNA folding shapes where SVs form. Therefore, the extent to which SVs occur in cancer genomes is to some extent dependent on the pre-cancerous tissue-specific 3D conformation[23,84,97,98]. Although CTCF is associated with 75-95% of all TAD anchors [41,99], adding CTCF binding sites to the model did not improve the correlation with experimental Hi-C data. Recently, however, a deep learning approach using Hi-C data found CTCF binding sites helped predict tissue and species-specific genome folding and SVs [100].

In addition to topological tension, transcription-coupled processes can contribute to incomplete repair and SV formation (Figure 3B-C)[101,102]. A prominent example is the fusion of the 5’ regulatory region of *TMPRSS2* to the 2nd intron of the transcription factor proto-oncogene *ERG*, one of the most recurrent SVs in cancer. *TMPRSS2* and *ERG* are located 2 Mb apart on chromosome 21 and chromatin conformation experiments have shown that these two genes are brought in close nuclear distance through chromatin looping in the prostate[5,103,104]. TOP2B is recruited to relieve DNA tension by creating DSBs, resulting in *TMPRSS2-ERG* fusions due to illegitimate recombination between the 5’ end of *TMPRSS2* and *ERG*. This leads to high-level upregulation of *ERG* through the activity of the *TMPRSS2* promoter and enhancer. Curiously, in prostate cancer, SVs near regions

associated with transcription factor binding, active transcription, and long-range chromatin looping, including those leading to *TMPRSS2:ERG* fusion, were shown to be highly enriched in younger patients[72]. The androgen receptor (AR) is a transcription factor that facilitates long-range interactions, and higher levels of AR in younger men might explain this observation. *TMPRSS2:ERG* formation has also been associated with **chromoplexy** formation involving one or several other chromosomes in the fusion formation[81].

5. Fitness consequences of SVs

Most cancers are thought to evolve through a stepwise Darwinian process by sequential acquisition of somatic alterations, each associated with negative, neutral, or positive effects on cellular fitness. In contrast, the many rearrangements generated in a single complex SV can separately and simultaneously have positive and negative effects on fitness, providing a mechanism by which cancers undergo "**punctuated evolution**"[81,105-107]. This is best exemplified by chromothripsis, hypothesised to be a stochastic process in terms of location and occurrence[108,109]. Chromothripsis can cause several hundred SVs in a single event, including loss of multiple genomic loci. Essential genes—which comprise approximately a tenth of all genes[110])—will often be disrupted by chance, resulting in strong negative selection. On rare occasions, some of these tens to hundreds of SVs can also have strong positive effects on fitness, resulting in increased overall fitness and survival of the tumour cells through this single catastrophic event[26,111].

While neutral selection is hypothesised to be common for SNVs and InDels (insertions or deletions of only a few bases)[112], it is considered uncommon for large SVs. These SVs typically have direct effects on a larger span of the genome--either through resulting large-scale changes in copy-number or by altering DNA topology and interactions--and therefore are more likely to affect one or more gene(s) either directly or indirectly[61,113,114].

The epigenetic landscape and 3D chromatin organisation of a cell have a large impact on the fitness effects of individual SVs. Although SVs can alter coding sequences and therefore gene composition, the major effects of SVs on selection are mediated through changes in gene dosage and in chromatin topology, both resulting in altered gene expression. These can be through trans (indirect) effects by affecting chromatin modifiers such as CTCF itself (discussed above). In multiple myeloma (MM), a translocation of the histone methyltransferase, *NSD2* causes high-level upregulation, leading to pervasive methylation of H3K36. This was associated with changes in A/B compartment and TAD structure, which was linked with gene expression dysregulation in these tumors[115]. However, the main consequences of SV-mediated changes in 3D folding on selection rely on modifying direct interactions between cis-regulatory elements (CREs) and their target genes.

a. The importance of cis-regulatory elements

Transcriptional regulation is often retained within TADs, and early work demonstrated that disruption of TADs could lead to dramatic changes in the transcriptional regulation[9,116]. Although one might expect that simply altering TAD boundaries would be the major mechanism by which SVs alter gene expression, this is usually not the case. According to a pan-cancer "PCAWG" survey across 2,700 tumor genomes, SVs and especially duplications

often span TAD boundaries[117] However, only a minority of these TAD-disrupting SVs were associated with marked changes in gene expression. This is in line with experiments showing minimal effect on gene expression following reduction of TADs through CTCF or Cohesin depletion[99,118]. In contrast, both the availability of open chromatin, enabling transcription factor binding, and the specific locus constituents are important factors in shaping the effects of SVs on transcription. However, the exact constituents mediating SV-associated mis-expression is a major open question.

Two important constituents appear to be 'active' promoters and enhancers that can translate transcription factor binding into altered expression. Both of these are considered to be cis-regulatory elements (CREs).: Promoters broadly serve as a landing platform to recruit transcription factors and RNA Polymerase II to the transcription start site (TSS). The classical model of gene activation involves long-range enhancer-promoter contact (abbreviated E:P from hereon) [119]. Enhancers have operationally been categorised into latent, primed, and active state[120], based on their ability to stimulate transcription, and are typically found to bind tissue-specific TFs and form long-range E:P looping. Promoters and enhancers are therefore classically thought to be involved in transcription initiation and transcription amplification, respectively, although these distinctions do not always hold and they can in some instances substitute for one another (see recent reviews[121,122]). Enhancers can form long-range contact with one or more promoters, depending upon their nuclear distance and the presence of cofactors such as transcription factors. By altering the distances between these regulatory elements, SVs can generate new contacts, which may result in SV-mediated oncogenic activation in *cis*. Experiments in engineered mouse embryos found that only SVs that directly 'flipped' active enhancers towards promoters caused significant expression changes[123], whereas simply deleting CTCF binding sites had minimal effects--findings that are consistent with the results from the PCAWG analysis of cancer genomes. A key requirement for SV-mediated gene expression upregulation in *cis* therefore appears to be *de novo* loop formation between CRE, and indeed the majority of SV-mediated gene expression upregulation in *cis* appears to involve gene-proximal promoter regions being placed in closer proximity to active, normally distal regulatory regions associated with enhancer activity[5], a mechanism coined *enhancer hijacking*[9].

These CREs, and especially enhancers, often display considerable variation across species, tissues, and differentiation stages [124]. Perhaps for this reason, oncogenic fusion events that exploit enhancer hijacking tend to be highly tissue- or even cell-type-specific [125] The exact enhancer altered to activate the proto-oncogene MYC, for example, is highly dependent on the tumor and tissue type and involves medulloblastoma, B-cell, T-cell, myeloid, endometrial, lung and colorectal cancer-specific enhancers[96,103,104].

b. Mechanisms by which SVs modify interactions between genes and cis-regulatory elements

Although the exact details remain to be identified, several mechanisms have emerged by which CRE-gene interactions are modified (Figure 4A). We have also made a website platform to visualise SVs and TADs across 22 tumor types.

SV-mediated Enhancer Juxtaposition.—The seminal study introducing the concept of enhancer hijacking detected a set of highly recurrent SVs on chromosomes 1 and 9 in pediatric group 3 and 4 medulloblastomas. By integrating SVs with chromatin data and mRNA expression, the authors identified *GFI1* and *GFI1B* as highly upregulated due to juxtaposition of distant, active enhancers[9]. Since then, similar examples of enhancer hijacking have been observed across multiple SVs and cancer types[7,126-128].

One of the oncogenes that are most often activated by this mechanism is *TERT*, the catalytic enzyme of telomerase. Almost all cancers require telomerase activation to prevent telomere attrition and induce immortalisation. *TERT* promoter-associated recurrent SVs were first found to be associated with high-level upregulation in lymphoma[129] and later across different cancer types[130,131,7,21,132]. Mapping of the distal breakpoints has identified a wide range of partner loci, which tend to be associated with open, active chromatin.

Although SV-mediated enhancer hijacking primarily involves enhancers from cancer-irrelevant genes, an example of 'enhancer swapping' between the two oncogenes *BCL6* and *MYC* was observed in B-cell lymphoma. Curiously, other B-cell lymphomas involved enhancer amplification of the *BCL6* super-enhancer, leading to its upregulation [133].

Although whether TADs are a functional unit is under debate, chromatin features associated with TAD boundaries likely prevent illegitimate enhancer:promoter (E:P) interactions and *de novo* TAD formation can create novel E:P interactions. A prominent example is in colorectal cancers, where a series of highly recurrent tandem duplications spanning both the oncogene *IGF2* and a strong, colon-specific enhancer in a neighbouring TAD was found to cause formation of a so-called neo-TAD, leading to more than 200-fold upregulation of *IGF2* [7].

SV-mediated CRE-gene Fusion.—In some cases, other CRE elements such as promoters are also hijacked along with the enhancers by the fusion event. *TMPRSS2:ERG* is an example of SV-mediated CRE-gene fusion, whereby the entire regulatory machinery of *TMPRSS2* is recruited to *ERG* to drive its expression. Of note, in addition to driving high-level upregulation of ERG, the fusion also removes a degron motif in the N-terminus of ERG, making it resistant to ubiquitin-mediated degradation [134,135]. To date, we know of no cases in which CRE elements are recurrently hijacked without including enhancers.

In some cases, single SVs are oncogenic through two or more such mechanisms. For example, inversions in chromosome 3 in leukemia activate the protooncogene *MECOM* by hijacking enhancers of the haploinsufficient tumor suppressor gene *GATA2*, and thereby simultaneously reduce *GATA2* expression [136]. Another intriguing example is the intrachromosomal deletion leading to fusion between *QKI* and *MYB* on chromosome 6 and concomitant upregulation of *MYB*, truncation of its negative regulatory domain, and partial loss of the tumor suppressor *QKI* [8,137].

In a minority of cases, these types of fusions can also generate proteins with *de novo* functions [5], which are the subject of other reviews[138]

SV-mediated Enhancer Looping.—Although the majority of SVs causing TAD disruption do not alter gene expression, certain events have been associated with *de novo*

long-range looping and oncogenic dysregulation of cancer genes. Indeed, CTCF binding sites represent some of the most frequently mutated non-coding sites in the cancer genome [139,140]. These mutations can disrupt CTCF-mediated insulator effects, generate novel E:P long-range loops, and activate oncogenes [141,142]. A systematic pan-cancer analysis of recurrent SVs that disrupt TADs identified several instances of oncogene activation through altering the 3D chromatin architecture [7]. The most prominent example was a set of highly recurrent SVs in lung squamous carcinoma that intersected a CTCF loop-anchor boundary and led to high-level upregulation of the oncogene *IRS4*, situated 100 kb away. SV-mediated disruption of the TAD boundary led to spreading of active chromatin, marked by H3K27ac, and ectopic looping between a normally silent enhancer and the *IRS4* promoter. Similar findings have been observed for cancers with mutations or deletions in CTCF binding sites, which can lead to permissive, long-range looping [143,144].

Many recurrent “hot spots” of SVs with unknown gene targets may indeed be explained by formation of de novo long-range enhancer-promoter contacts, but the distances can be several hundred kb. A prototypical example of SV mediated long-range looping was identified in medulloblastoma, where a series of complex SVs including tandem duplications and an inversion, all spanning a super-enhancer and a TAD boundary, caused the enhancer to ‘flip’ to the other side of the TAD boundary, leading to high-level upregulation of the histone methyltransferase *PRDM6*, residing more than 600kb away from the enhancer [126].

SV-mediated Enhancer amplification.—Copy number amplifications can amplify whole genes to cause gene expression dosage increase. However, amplification of strong, lineage-specific enhancers that control *MYC* have been identified in AML, T-ALL, lung, and endometrial carcinoma [145,146]. By analysing recurrent amplified non-protein-coding regions in the genome, Zhang et al conducted a systematic analysis of recurrently amplified non-protein-coding genomic loci and identified recurrent amplifications of four regulatory regions that were associated with increased expression of otherwise unaltered genes controlled by these enhancers, including *MYC* [147]. Similar to the enhancer hijacking examples, the enhancer amplifications were distinct and specific to each tumor type, again arguing for a strongly tissue-specific epigenetic mechanism.

A similar mechanism of enhancer amplification was identified in prostate cancer, where an androgen receptor-specific lineage-specific enhancer was found to be frequently amplified in response to androgen-deprivation, leading to re-activation of androgen receptor expression through increased E:P interaction [148].

Enhancer Looping in Extrachromosomal DNA.—Extrachromosomal circular DNA (ecDNA) structures can be present in hundreds of copies and have often been found to carry one or more oncogenes [25,149,150]. The upregulation of ecDNA-containing oncogenes are thought to be mediated primarily through dosage-effects, but recent discoveries suggest that increased chromatin accessibility in ecDNAs [27], as well as noncoding elements associated with enhancer activity from other regions of the genome, can be incorporated into the ecDNA, possibly causing novel E:P interactions with the oncogenes. Examples include ecDNA incorporating *EGFR* and *MYC* and an enhancer from an adjacent TAD in glioblastoma and neuroblastoma [15,151]. Recent discoveries also showed that ecDNA can act

in trans as enhancers for promoters on the linear genome[152] with other ecDNA molecules in a structure called the ecDNA hub, which can facilitate co-evolution and synergies between amplicons[153].

5. Conclusion

Much of cancer genome discovery to date has focused on the exome. However, SVs by nature tend to occur within the 99% of the genome that does not encode proteins, and which is the domain of gene regulatory elements. As technologies improve to characterize both SVs genome-wide and the characteristics of the regulatory elements across the genome, we are learning more about the essentiality of their interaction. Going forward, we anticipate that a relatively complete understanding of the specific epigenetic contexts within which each SV occurs will be essential to understanding both the mechanisms by which those SVs are formed, their effects on cancer evolution, and their clinical interpretation. An important consideration when linking distant SVs to gene expression in *cis*, such as enhancer hijacking, is to phase the SVs with expression data to the same allele, which is currently challenged by short read length and short haplotype blocks (see Box 1).

For this reason, long-read sequencing will likely help resolve both the genome topology resulting from SVs and their relations to epigenetic elements that impact chromatin conformation. Several additional technologies may also help shed light on the causal relations between 3D genome structure and SV formation. For example, single-cell DNA sequencing can be used to identify the mechanistic biases associated with SV formation before cell division and selection, by detecting DNA breaks and rearrangement formation at single-cell resolution before each cell undergoes cell division[57]. Complete telomere-to-telomere[154] and ancestry-aware human reference assemblies will also enable a more accurate and sensitive mapping and analysis of changes in DNA topology. Moreover, large-scale profiling of chromatin conformations across cell and tissue types may also support an improved understanding of the relations between tissue-specific DNA organization and SV rates. Similarly, sequential analyses of genomes during cancer formation and progression may support a more detailed understanding of how SVs form in loci with altered 3D organization, such as extrachromosomal DNA. Lastly, much larger numbers of cancers will have to undergo whole genome sequencing to enable robust detection of SVs that recur more frequently than current models would suggest[5].

As our knowledge and tools advance, we anticipate additional questions to emerge. Perhaps the most prominent and immediate of these is a full understanding of chromatin interactions in 3D and how these relate to somatic SVs. Although Hi-C and related technologies have been available for several years now, their high costs in both money and DNA have severely limited the contexts in which they have been applied. Recent discoveries have also highlighted the promiscuity of certain promoters with varied enhancers (e.g. *MYC*), whereas other promoter-enhancer interactions appear to be more restrictive. The reasons behind the selectivity of promoter-enhancer interactions are unclear.

Perhaps in the more distant future, we anticipate contextual questions to become more prominent. For example, the epigenetic contexts in which SVs occur evolve with cellular

differentiation, including within tumors. How do these changes in epigenetic context within a tumor modify the effects of the SVs in the tumor, and do these SV-epigenetic interactions, in turn, limit the cellular differentiation states within the tumor? Cancer cells tend to reside in progenitor-like cell states; does this create a common and unifying stem-like epigenetic background across cancers? Moreover, the vast majority of germline genomic variation occurs outside the exome, and its effects on phenotype are thought to occur largely through modifications of CREs. How do these germline variants affect SV formation and selective effects, and do these relate to cancer risk both within and across ancestral populations?

Lastly, how do we ultimately leverage this understanding of the role of epigenetics in SV formation and selective effects to develop new cancer preventative and therapeutic approaches? Existing therapeutics that target oncogenes activated by SVs, such as fusions of BCR-ABL1 and PML-RARα, and amplifications of ERBB2 and EGFR, have targeted the enzymatic activity of the oncogenes themselves. However, an improved understanding of changes in the 3D genome that occur with SVs may offer new opportunities. Altered interactions between oncogenes and cis-regulatory domains will not alter the amino acid sequence of the oncogene itself, but only the transcriptional regulation; epigenetically-directed therapies (reviewed in [155,156]) may have high specificity in these cases. For example, a recent study demonstrated that ecDNA- containing cancers were susceptible to the BET domain inhibitor JQ1, due to the high, focal concentration of transcriptional activity at ecDNA hubs[153]. In some cases, detection of SVs altering 3D chromatin structure may serve important diagnostic purposes. For example, the androgen receptor is often activated by amplification of its enhancer that do not include the gene itself[148,157]. Diagnostic and therapeutic approaches that can effectively detect and leverage the altered epigenetic interactions induced by SVs in cancer cells would be likely to have a major impact on patient outcomes.

Acknowledgment:

This work was supported by the Independent Research Fund Denmark (0134-00265B; JW and NS), the German Research Foundation (DFG) FD), and the NIH, the Pediatric Brain Tumor Foundation, the Fund for Innovative Cancer Informatics, and the Gray Matters Brain Cancer Foundation (RB).

Glossary:

Branching evolution
A pattern of tumor evolution in which subclones with distinct driver combinations emerge from the initial clonal outgrowth of the tumor, contributing to tumor heterogeneity

Breakage-Fusion Bridge cycles
A mechanism of amplification in cancer genomes. Chromosomes fuse due to telomere shortening and get broken apart during mitosis leaving one daughter cell with extra copies of genes.

Chromoanasynthesis
A mechanism of complex SV generation due to template switching during replication

Chromoplexy

Chains of SVs involving multiple chromosomes

Chromatin compartments
Subdivisions of chromosome territories divided into compartments with open and closed chromatin

Chromosome Territories
A discrete region within the nucleus containing a specific chromosome

Chromothripsis
A mechanism of SV generation leading to a complex pattern of step-like copy number loss and random relegation of DNA fragments resulting in approximately equal ratios of deletion to duplication and the two inversion types.

Driver events
Genetic variants resulting in increased evolutionary fitness of the affected cell.

Evolutionary Selection
The process of enrichment of clones with higher fitness in a population of cancer cells.

Extrachromosomal Amplification
A mechanism of amplification in cancer genomes potentially originating as a byproduct of chromothripsis. Extrachromosomal amplicons are small circular fragments of DNA that typically contain at least one oncogene but no centromere. Due to the lacking centromere, they are randomly distributed during mitosis. The daughter cell receiving more amplicons can gain evolutionary fitness leading to rapid high-level amplification of the affected oncogenes in the cancer.

Hi-C
A proximity ligation-based sequencing technology that, in principle, allows the detection of all 3D interactions between DNA segments in the nucleus.

MMBIR
Microhomology-Mediated Break-Induced Replication, a form of microhomology-mediated template switching during replication, which can lead to a complex SV pattern.

Passenger events
Genetic variants with no effect on the evolutionary fitness of the affected cell. Phase separation: A phenomenon of decreased mixing between molecules because of differences in the intermolecular interactions. Most familiar from the unmixing of oil and water resulting in two separate phases.

Punctuated evolution
Bursts of changes to the genome, with many variants acquired in one event, resulting in a dramatically changed evolutionary fitness. Often contrasted with the sequential acquisition and selection of individual variants in classic evolutionary theory.

Selective sweep

Outgrowth of a subclone with increased fitness that completely replaces the parental and all other clones in the population.

Templated Insertions

A form of SV caused by the DNA polymerase jumping to a different strand during replication.

Topologically associated domains (TADs)

Regions in the genome in the range of several 100kbps - few Mbp which are separated by boundary elements and show higher interaction frequencies in 3D within one TAD than between TADs.

References

1. Tallman MS et al. All-trans-retinoic acid in acute promyelocytic leukemia. N. Engl. J. Med 337, 1021–1028 (1997). [PubMed: 9321529]
2. Slamon DJ et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N. Engl. J. Med 344, 783–792 (2001). [PubMed: 11248153]
3. Druker BJ et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N. Engl. J. Med 344, 1031–1037 (2001). [PubMed: 11287972]
4. Wala JA et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res. 28, 581–591 (2018). [PubMed: 29535149]
5. Rheinbay E et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature 578, 102–111 (2020). [PubMed: 32025015]
6. Gong T, Hayes VM & Chan EKF Detection of somatic structural variants from short-read next-generation sequencing data. Brief. Bioinform 22, (2021).
7. Weischenfeldt J et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat Genet (2016).
8. Bandopadhayay P et al. MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism. Nat. Genet 48, 273–282 (2016). [PubMed: 26829751]
9. Northcott PA et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature vol. 511 428–434 (2014). [PubMed: 25043047]
10. Newman S et al. Genomes for Kids: The scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing. Cancer Discov. (2021) doi:10.1158/2159-8290.CD-20-1631.
11. Staaf J et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat. Med 25, 1526–1533 (2019). [PubMed: 31570822]
12. Duncavage EJ et al. Genome Sequencing as an Alternative to Cytogenetic Analysis in Myeloid Cancers. N. Engl. J. Med 384, 924–935 (2021). [PubMed: 33704937]
13. Zack TI et al. Pan-cancer patterns of somatic copy number alteration. Nat. Genet 45, 1134–1140 (2013). [PubMed: 24071852]
14. Cho SW et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell 173, 1398–1412.e22 (2018). [PubMed: 29731168]
15. Morton AR et al. Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. Cell 179, 1330–1341.e13 (2019). [PubMed: 31761532]
16. Chiang C et al. The impact of structural variation on human gene expression. Nat. Genet 49, 692–699 (2017). [PubMed: 28369037]
17. Audano PA et al. Characterizing the Major Structural Variant Alleles of the Human Genome. Cell 176, 663–675.e19 (2019). [PubMed: 30661756]
18. Merker JD et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet. Med 20, 159–163 (2018). [PubMed: 28640241]

19. Sebat J et al. Strong association of de novo copy number mutations with autism. Science 316, 445–449 (2007). [PubMed: 17363630]
20. Li Y et al. Patterns of somatic structural variation in human cancer genomes. Nature 578, 112–121 (2020). [PubMed: 32025012]
21. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature 578, 82–93 (2020). [PubMed: 32025007]
22. Spielmann M, Lupiáñez DG & Mundlos S Structural variation in the 3D genome. Nature Reviews Genetics vol. 19 453–467 (2018).
23. Fudenberg G, Getz G, Meyerson M & Mirny LA High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat. Biotechnol 29, 1109–1113 (2011). [PubMed: 22101486]
24. Lieberman-Aiden E et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science vol. 326 289–293 (2009). [PubMed: 19815776]
25. Kim H et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. Nat. Genet 52, 891–897 (2020). [PubMed: 32807987]
26. Turner KM et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature 543, 122–125 (2017). [PubMed: 28178237]
27. Wu S et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature 575, 699–703 (2019). [PubMed: 31748743]
28. Scheble VJ et al. ERG rearrangement is specific to prostate cancer and does not occur in any other common tumor. Mod. Pathol 23, 1061–1067 (2010). [PubMed: 20473283]
29. Cavalli G & Heard E Advances in epigenetics link genetics to the environment and disease. Nature 571, 489–499 (2019). [PubMed: 31341302]
30. Zheng H & Xie W The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol 20, 535–550 (2019). [PubMed: 31197269]
31. Bintu B et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362, (2018).
32. Mirny LA The fractal globule as a model of chromatin architecture in the cell. Chromosome Res. 19, 37–51 (2011). [PubMed: 21274616]
33. Meaburn KJ & Misteli T Cell biology: chromosome territories. Nature 445, 379–781 (2007). [PubMed: 17251970]
34. Rao SS et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell vol. 159 1665–1680 (2014). [PubMed: 25497547]
35. Boyle S et al. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum. Mol. Genet 10, 211–219 (2001). [PubMed: 11159939]
36. Szabo Q, Bantignies F & Cavalli G Principles of genome folding into topologically associating domains. Sci Adv 5, eaaw1668 (2019). [PubMed: 30989119]
37. Strom AR et al. Phase separation drives heterochromatin domain formation. Nature vol. 547 241–245 (2017). [PubMed: 28636597]
38. de Wit E et al. CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell 60, 676–684 (2015). [PubMed: 26527277]
39. Symmons O et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res vol. 24 390–400 (2014). [PubMed: 24398455]
40. Mifsud B et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet vol. 47 598–606 (2015). [PubMed: 25938943]
41. Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature vol. 485 376–380 (2012). [PubMed: 22495300]
42. Bell AC, West AG & Felsenfeld G The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387–396 (1999). [PubMed: 10458613]
43. Parelho V et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422–433 (2008). [PubMed: 18237772]

44. Sanborn AL et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A 112, E6456–65 (2015). [PubMed: 26499245]
45. Ernst J & Kellis M ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012). [PubMed: 22373907]
46. Dekker J, Marti-Renom MA & Mirny LA Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet 14, 390–403 (2013). [PubMed: 23657480]
47. Fortin J-P & Hansen KD Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol. 16, 180 (2015). [PubMed: 26316348]
48. Finn EH et al. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell 176, 1502–1515.e10 (2019). [PubMed: 30799036]
49. Cattoni DI et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat. Commun 8, 1753 (2017). [PubMed: 29170434]
50. Aitken SJ et al. CTCF maintains regulatory homeostasis of cancer pathways. Genome Biology vol. 19 (2018).
51. Hyle J et al. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer–promoter looping. Nucleic Acids Res. 47, 6699–6713 (2019). [PubMed: 31127282]
52. Sima J et al. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. Cell 176, 816–830.e18 (2019). [PubMed: 30595451]
53. Feng Y & Pauklin S Revisiting 3D chromatin architecture in cancer development and progression. Nucleic Acids Res. 48, 10632–10647 (2020). [PubMed: 32941624]
54. Fritz AJ et al. Wide-scale alterations in interchromosomal organization in breast cancer cells: defining a network of interacting chromosomes. Hum. Mol. Genet 23, 5133–5146 (2014). [PubMed: 24833717]
55. Hnisz D, Schuijers J, Li CH & Young RA Regulation and Dysregulation of Chromosome Structure in Cancer. Annu. Rev. Cancer Biol 2, 21–40 (2018).
56. van Steensel B & Belmont AS Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell 169, 780–791 (2017). [PubMed: 28525751]
57. Zhang C-Z et al. Chromothripsis from DNA damage in micronuclei. Nature 522, 179–184 (2015). [PubMed: 26017310]
58. Umbreit NT et al. Mechanisms generating cancer genome complexity from a single cell division error. Science 368, (2020).
59. Lemaître C et al. Nuclear position dictates DNA repair pathway choice. Genes Dev. 28, 2450–2463 (2014). [PubMed: 25366693]
60. Stratton MR, Campbell PJ & Futreal PA The cancer genome. Nature vol. 458 719–724 (2009). [PubMed: 19360079]
61. Martincorena I et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 173, 1823 (2018).
62. Lawrence MS et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218 (2013). [PubMed: 23770567]
63. Zhang C-Z & Pellman D Cancer Genomic Rearrangements and Copy Number Alterations from Errors in Cell Division. (2022) doi:10.1146/annurev-cancerbio-070620-094029.
64. Liu P et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell 146, 889–903 (2011). [PubMed: 21925314]
65. Zhang C-Z & Pellman D From Mutational Mechanisms in Single Cells to Mutational Patterns in Cancer Genomes. Cold Spring Harb. Symp. Quant. Biol 80, 117–137 (2015). [PubMed: 26968629]
66. Anand RP, Lovett ST & Haber JE Break-induced DNA replication. Cold Spring Harb. Perspect. Biol 5, a010397 (2013). [PubMed: 23881940]

67. Hungerford DA A minute chromosome in human chronic granulocytic leukemia. Science 132, 1497–1499 (1960).
68. Neves H, Ramos C, da Silva MG, Parreira A & Parreira L The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood vol. 93 1197–1207 (1999). [PubMed: 9949162]
69. Lukásová E et al. Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. Hum. Genet 100, 525–535 (1997). [PubMed: 9341866]
70. Roix JJ, McQueen PG, Munson PJ, Parada LA & Misteli T Spatial proximity of translocation-prone gene loci in human lymphomas. Nat. Genet 34, 287–291 (2003). [PubMed: 12808455]
71. Klein IA et al. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell vol. 147 95–106 (2011). [PubMed: 21962510]
72. Gerhauser C et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell 34, 996–1011.e8 (2018). [PubMed: 30537516]
73. Haffner MC, De Marzo AM, Meeker AK, Nelson WG & Yegnasubramanian S Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target? Clin Cancer Res vol. 17 3858–3864 (2011). [PubMed: 21385925]
74. Canela A et al. Genome Organization Drives Chromosome Fragility. Cell 170, 507–521.e18 (2017). [PubMed: 28735753]
75. Polak P et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature vol. 518 360–364 (2015). [PubMed: 25693567]
76. Yang L et al. Diverse Mechanisms of Somatic Structural Variations in Human Cancer Genomes. Cell vol. 153 919–929 (2013). [PubMed: 23663786]
77. Aymard F et al. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. Nat. Struct. Mol. Biol 24, 353–361 (2017). [PubMed: 28263325]
78. Pfister SX et al. SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell Rep. 7, 2006–2018 (2014). [PubMed: 24931610]
79. Kilic S et al. Phase separation of 53 BP 1 determines liquid-like behavior of DNA repair compartments. The EMBO Journal vol. 38 (2019).
80. Ochs F et al. Stabilization of chromatin topology safeguards genome integrity. Nature 574, 571–574 (2019). [PubMed: 31645724]
81. Baca SC et al. Punctuated evolution of prostate cancer genomes. Cell 153, 666–677 (2013). [PubMed: 23622249]
82. Sakofsky CJ et al. Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements. Mol Cell vol. 60 860–872 (2015). [PubMed: 26669261]
83. Zhang F et al. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat. Genet 41, 849–853 (2009). [PubMed: 19543269]
84. Sidiropoulos N et al. Somatic structural variant formation is guided by and influences genome architecture. Genome Res. 32, 643–655 (2022). [PubMed: 35177558]
85. Ryba T et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 20, 761–770 (2010). [PubMed: 20430782]
86. Parada LA, McQueen PG, Munson PJ & Misteli T Conservation of relative chromosome positioning in normal and cancer cells. Curr. Biol 12, 1692–1697 (2002). [PubMed: 12361574]
87. Marshall WF et al. Interphase chromosomes undergo constrained diffusional motion in living cells. Curr. Biol 7, 930–939 (1997). [PubMed: 9382846]
88. Miné-Hattab J & Rothstein R Increased chromosome mobility facilitates homology search during recombination. Nat. Cell Biol 14, 510–517 (2012). [PubMed: 22484485]
89. Polo SE & Jackson SP Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev vol. 25 409–433 (2011). [PubMed: 21363960]

90. Juhász S et al. The chromatin remodeler ALC1 underlies resistance to PARP inhibitor treatment. Sci Adv 6, (2020).
91. Clouaire T et al. Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. Mol. Cell 72, 250–262.e6 (2018). [PubMed: 30270107]
92. Lottersberger F, Karssemeijer RA, Dimitrova N & de Lange T 53BP1 and the LINC Complex Promote Microtubule-Dependent DSB Mobility and DNA Repair. Cell 163, 880–893 (2015). [PubMed: 26544937]
93. Ribeiro de Almeida C et al. The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. Immunity 35, 501–513 (2011). [PubMed: 22035845]
94. Guo C et al. CTCF-binding elements mediate control of V(D)J recombination. Nature 477, 424–430 (2011). [PubMed: 21909113]
95. Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA & Makova KD A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? Genome Res. 22, 993–1005 (2012). [PubMed: 22456607]
96. Bianco S et al. Polymer physics predicts the effects of structural variants on chromatin architecture. Nat. Genet 50, 662–667 (2018). [PubMed: 29662163]
97. Engreitz JM, Agarwala V & Mirny LA Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. PLoS One 7, e44196 (2012). [PubMed: 23028501]
98. De S & Michor F DNA secondary structures and epigenetic determinants of cancer genome evolution. Genome Biology vol. 11 P10 (2010).
99. Nora EP et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169, 930–944.e22 (2017). [PubMed: 28525758]
100. Fudenberg G, Kelley DR & Pollard KS Predicting 3D genome folding from DNA sequence with Akita. Nat. Methods (2020) doi:10.1038/s41592-020-0958-x.
101. Glover TW, Wilson TE & Arlt MF Fragile sites in cancer: more than meets the eye. Nat. Rev. Cancer 17, 489–501 (2017). [PubMed: 28740117]
102. Sarni D et al. 3D genome organization contributes to genome instability at fragile sites. Nat. Commun 11, 3613 (2020). [PubMed: 32680994]
103. Haffner MC et al. Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. Nat Genet vol. 42 668–675 (2010). [PubMed: 20601956]
104. Lin C et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell vol. 139 1069–1083 (2009). [PubMed: 19962179]
105. Zhang C-Z, Leibowitz ML & Pellman D Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. Genes Dev. 27, 2513–2530 (2013). [PubMed: 24298051]
106. Davis A, Gao R & Navin N Tumor evolution: Linear, branching, neutral or punctuated? Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1867, 151–161 (2017). [PubMed: 28110020]
107. Heng HHQ et al. The evolutionary mechanism of cancer. J. Cell. Biochem 109, 1072–1084 (2010). [PubMed: 20213744]
108. Cortés-Ciriano I et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat. Genet 52, 331–341 (2020). [PubMed: 32025003]
109. Stephens PJ et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell vol. 144 27–40 (2011). [PubMed: 21215367]
110. Nichols CA et al. Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. doi:10.1101/534529.
111. Stephens PJ et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144, 27–40 (2011). [PubMed: 21215367]

112. Williams MJ et al. Quantification of subclonal selection in cancer from bulk sequencing data. doi:10.1101/096305.
113. Davoli T et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell 155, 948–962 (2013). [PubMed: 24183448]
114. Solimini NL et al. Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential. Science (2012) doi:10.1126/science.1219580.
115. Lhoumaud P et al. NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains. doi:10.1101/587931.
116. Lupiáñez DG et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell vol. 161 1012–1025 (2015). [PubMed: 25959774]
117. Akdemir KC et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat. Genet 52, 294–305 (2020). [PubMed: 32024999]
118. Rao SSP et al. Cohesin Loss Eliminates All Loop Domains. Cell 171, 305–320.e24 (2017). [PubMed: 28985562]
119. Bonev B et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell 171, 557–572.e24 (2017). [PubMed: 29053968]
120. Spitz F & Furlong EEM Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics vol. 13 613–626 (2012).
121. Gasperini M, Tome JM & Shendure J Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet 21, 292–310 (2020). [PubMed: 31988385]
122. Andersson R & Sandelin A Determinants of enhancer and promoter activities of regulatory elements. Nat. Rev. Genet 21, 71–87 (2020). [PubMed: 31605096]
123. Despang A et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat. Genet 51, 1263–1271 (2019). [PubMed: 31358994]
124. Villar D et al. Enhancer evolution across 20 mammalian species. Cell 160, 554–566 (2015). [PubMed: 25635462]
125. Bahr C et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature 553, 515–520 (2018). [PubMed: 29342133]
126. Northcott PA et al. The whole-genome landscape of medulloblastoma subtypes. Nature 547, 311–317 (2017). [PubMed: 28726821]
127. Sin-Chan P et al. A C19MC-LIN28A-MYCN Oncogenic Circuit Driven by Hijacked Super-enhancers Is a Distinct Therapeutic Vulnerability in ETMRs: A Lethal Brain Tumor. Cancer Cell 36, 51–67.e7 (2019). [PubMed: 31287992]
128. Northcott PA, Pfister SM & Jones DT Next-generation (epi)genetic drivers of childhood brain tumours and the outlook for targeted therapies. Lancet Oncol vol. 16 e293–302 (2015). [PubMed: 26065614]
129. Schilling G et al. Molecular characterization of chromosomal band 5p15.33: a recurrent breakpoint region in mantle cell lymphoma involving the TERT-CLPTM1L locus. Leuk. Res 37, 280–286 (2013). [PubMed: 23137523]
130. Davis CF et al. The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell vol. 26 319–330 (2014). [PubMed: 25155756]
131. Peifer M et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature vol. 526 700–704 (2015). [PubMed: 26466568]
132. Barthel FP et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. Nat. Genet 49, 349–357 (2017). [PubMed: 28135248]
133. Ryan RJH et al. Detection of Enhancer-Associated Rearrangements Reveals Mechanisms of Oncogene Dysregulation in B-cell Lymphoma. Cancer Discov. 5, 1058–1071 (2015). [PubMed: 26229090]
134. Gan W et al. SPOP Promotes Ubiquitination and Degradation of the ERG Oncoprotein to Suppress Prostate Cancer Progression. Mol. Cell 59, 917–930 (2015). [PubMed: 26344095]
135. An J et al. Truncated ERG Oncoproteins from TMPRSS2-ERG Fusions Are Resistant to SPOP-Mediated Proteasome Degradation. Mol. Cell 59, 904–916 (2015). [PubMed: 26344096]

136. Gröschel S et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157, 369–381 (2014). [PubMed: 24703711]
137. Drier Y et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. Nat. Genet 48, 265–272 (2016). [PubMed: 26829750]
138. Latysheva NS & Babu MM Discovering and understanding oncogenic gene fusions through data intensive computational approaches. Nucleic Acids Res. 44, 4487–4503 (2016). [PubMed: 27105842]
139. Katainen R et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nature Genetics vol. 47 818–821 (2015). [PubMed: 26053496]
140. Kaiser VB, Taylor MS & Semple CA Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types. PLOS Genetics vol. 12 e1006207 (2016). [PubMed: 27490693]
141. Hnisz D, Day DS & Young RA Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell 167, 1188–1200 (2016). [PubMed: 27863240]
142. Liu EM et al. Identification of Cancer Drivers at CTCF Insulators in 1,962 Whole Genomes. Cell Systems vol. 8 446–455.e8 (2019). [PubMed: 31078526]
143. Hnisz D et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science (2016).
144. Liu Y et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat. Genet 52, 811–818 (2020). [PubMed: 32632335]
145. Shi J et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev. 27, 2648–2662 (2013). [PubMed: 24285714]
146. Herranz D et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat. Med 20, 1130–1137 (2014). [PubMed: 25194570]
147. Zhang X et al. Somatic super-enhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. Cancer Discov. CD–17–0532 (2017).
148. Takeda DY et al. A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. Cell 174, 422–432.e13 (2018). [PubMed: 29909987]
149. Vogt N et al. Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. Proc. Natl. Acad. Sci. U. S. A 101, 11368–11373 (2004). [PubMed: 15269346]
150. Kohl NE et al. Transposition and amplification of oncogene-related sequences in human neuroblastomas. Cell 35, 359–367 (1983). [PubMed: 6197179]
151. Koche RP et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat. Genet 52, 29–34 (2020). [PubMed: 31844324]
152. Zhu Y et al. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. Cancer Cell 39, 694–707.e7 (2021). [PubMed: 33836152]
153. Hung KL et al. ecDNA hubs drive cooperative intermolecular oncogene expression. Nature 1–6 (2021).
154. Nurk S et al. The complete sequence of a human genome. Science 376, 44–53 (2022). [PubMed: 35357919]
155. Mohammad HP, Barbash O & Creasy CL Targeting epigenetic modifications in cancer therapy: erasing the roadmap to cancer. Nat. Med 25, 403–418 (2019). [PubMed: 30842676]
156. Topper MJ, Vaz M, Marrone KA, Brahmer JR & Baylin SB The emerging role of epigenetic therapeutics in immuno-oncology. Nat. Rev. Clin. Oncol 17, 75–90 (2020). [PubMed: 31548600]
157. Viswanathan SR et al. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell 174, 433–447.e19 (2018). [PubMed: 29909985]
158. Logsdon GA, Vollger MR & Eichler EE Long-read human genome sequencing and its applications. Nat. Rev. Genet 21, 597–614 (2020). [PubMed: 32504078]
159. Chen Z et al. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. Genome Res. 30, 898–909 (2020). [PubMed: 32540955]

160. Luebeck J et al. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. Nat. Commun 11, 4374 (2020). [PubMed: 32873787]

161. Chen X et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222 (2016). [PubMed: 26647377]

162. Chong Z et al. novoBreak: local assembly for breakpoint detection in cancer genomes. Nat. Methods 14, 65–67 (2017). [PubMed: 27892959]

163. Hadi K et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. Cell 183, 197–210.e32 (2020). [PubMed: 33007263]

164. Deshpande V et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat. Commun 10, 392 (2019). [PubMed: 30674876]

165. McPherson AW et al. Correction to: ReMixT: clone-specific genomic structure estimation in cancer. Genome Biol. 18, 188 (2017). [PubMed: 28985744]

166. Harewood L et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol. 18, 125 (2017). [PubMed: 28655341]

167. Dixon JR et al. Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet 50, 1388–1398 (2018). [PubMed: 30202056]

168. Agnihotri S et al. The genomic landscape of schwannoma. Nat. Genet 48, 1339–1348 (2016). [PubMed: 27723760]

169. PCAWG Transcriptome Core Group et al. Genomic basis for RNA alterations in cancer. Nature 578, 129–136 (2020). [PubMed: 32025019]

170. McPherson A et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput. Biol 7, e1001138 (2011). [PubMed: 21625565]

Box 1:

Limitations to SV discovery and interpretation

Whole-genome sequencing (WGS) has limited sensitivity for SV detection. Standard contemporary sequencing generates paired 100-250bp reads that represent the ends of longer (usually ~500 bp) fragments. SVs are indicated by "**discordant reads**", whose paired ends align to genomic loci that are more distant than the typical fragment size, and by "**split reads**" that overlap SV breakpoints (Figure Box 1 **panel A**). Only part of a split read will align to each parental sequence, and these short segments are often unmappable. Complex SVs, with multiple breakpoints in close proximity, can further subdivide overlapping reads, increasing ambiguity.

Additional challenges involve untangling the linear genomic sequence generated by complex SVs (Figure Box 1 **panel B**). First, each individual SV in the complex event must be **phased**, determining which are on the same homologous chromosome. Second, the **linear ordering of different SVs** within a complex event can be ambiguous, especially when some SVs may have escaped detection. Complex SVs also often differ between cancer subclones, generating more ambiguity.

New computational and sequencing technologies are helping to solve these issues. **Long-read**[158], **linked-read**[159], and **optical-mapping**[160] generate reads or read sets that are more likely to contain unique sequences, reducing mapping ambiguity. Different SVs can be phased when they are captured on the same read pair or on separate reads that also contain sequences (such as germline variants) that are unique to a single haplotype (Figure Box 1 **panel B**). Single reads that capture two or more SVs can also topologically disambiguate the resulting linear sequence. SV detection methods based upon **genome assembly**[4,161,162] can also generate longer contigs that are more likely to align uniquely. Probabilistic methods are also being developed to integrate copy-number and SV data to reconstruct genome topologies[163-165], which sometimes disambiguates the topology of complex SVs[163]. And orthogonal datasets can further aid SV detection. For example, **Hi-C** data can detect juxtapositions between normally distant genomic loci by indicating their three-dimensional proximity[166,167] and **RNA sequencing** can detect novel fusion transcripts[168-170].

Request for NRC graphics help:

1. Please improve the quality of the figure's appearance. For example, color choices need to be better!
2. Please standardize appearance across subpanels and other figures

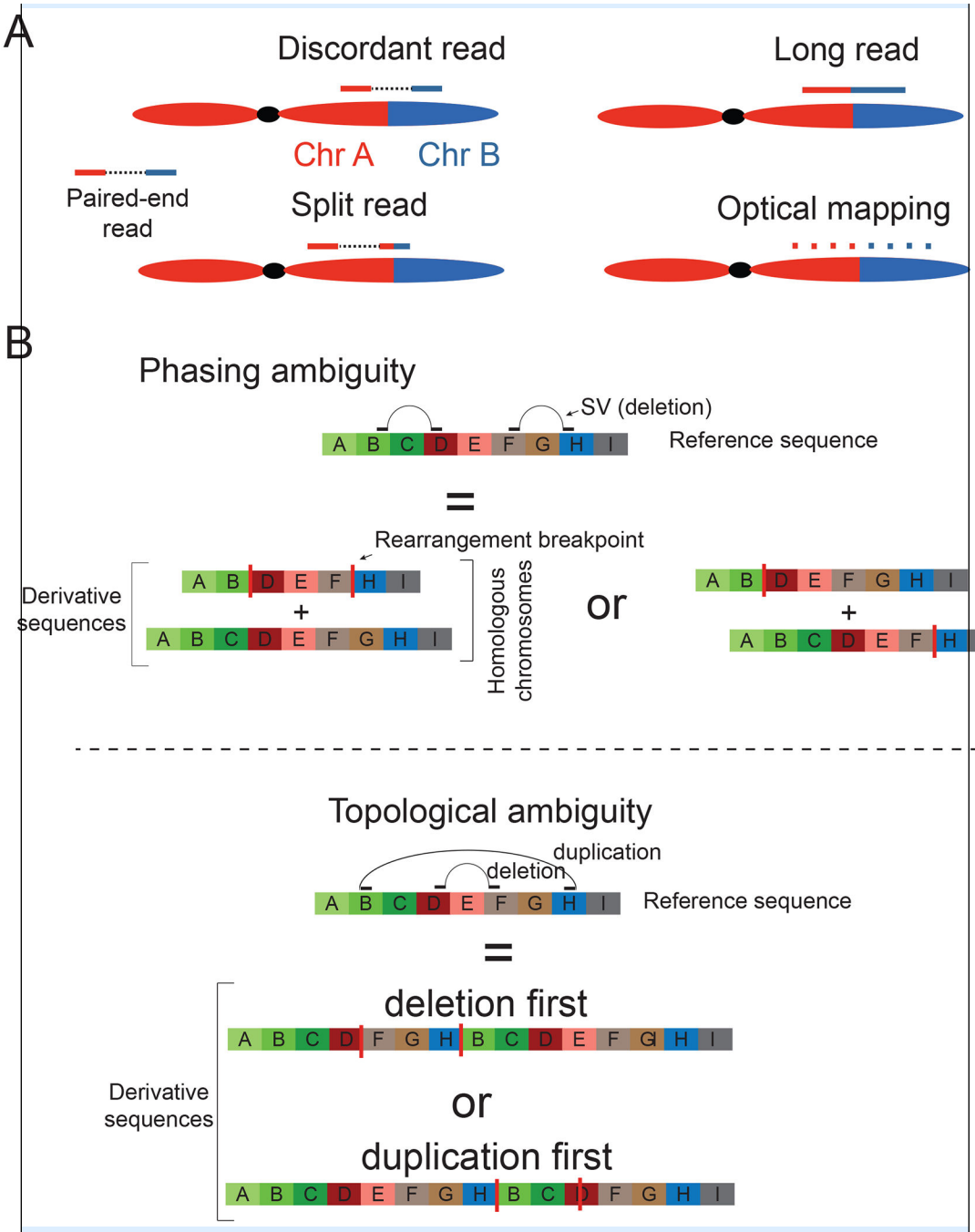

Figure Box 1.

Figure 1: Types of rearrangements.

A) Topological categories of simple SVs: translocations, deletions, duplications, inversions, and insertions. The initial DNA sequence is indicated in the center, with arrows pointing to sequences generated by each of these categories of SV. Underneath these SV-derived sequences are the copy-number profiles and SV calls that would be detected, projected onto the reference genome. In second-generation sequencing data, the copy-number profiles reflect the number of reads aligned at each locus. Here, “2” indicates the number of reads expected from loci with two copies. The SV calls are indicated by arrows connected by dotted lines. These represent the discordant and split reads in second-generation sequencing data, where part of the read maps to each side of the SV (the arrows). **B)** Mechanisms of generating complex SVs. These can originate in a single event (typical for chromothripsis) or sequence of events over multiple cell cycles (as in breakage-fusion-bridge cycles (BFBs) or with the accumulation and selection of double minutes). The resulting cancer genome as well as the pattern detected by second-generation sequencing are indicated analogous to (A). The detected patterns can be indicative of the SV generating process, (Amplification with inversions for BFB, focal high CN for DM, complex SVs across a chromosome with CN loss for chromothripsis) but are often overlayed by subsequent SVs in real cancer genomes. Additionally, one pattern of complex SVs can be mechanistically linked to another, increasing the probability of observing mixed patterns in real cancer genomes. The figures for this paper were produced by Nature Reviews Cancer graphics artists and are under copyright to Nature Reviews Cancer. The figures can be viewed at : Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R. Structural variations in cancer and the 3D genome. Nature Reviews Cancer. 2022 September;22(9):533–546. PMID: 35764888; DOI: 10.1038/s41568-022-00488-9.

Figure 2: Chromatin organization and its disruption by rearrangements in cancer.

A) The three-dimensional organization of DNA. Chromatin fibers can form regions of close interactions such as topologically associated domains (TADs). At the largest scale, chromatin is organized as a fractal globule, in which the likelihood that two sequences within a chromosome will be adjacent in three-dimensional space is proportional to the number of bases that separate them. **B)** TADs are highly interconnected DNA loops, as indicated by dense triangles in Hi-C maps, where higher interaction frequencies are represented by deeper red. Their boundaries are often defined by CTCF and cohesin binding, and genetic elements that regulate transcription (e.g. promoters, enhancers, and silencers) typically act upon genes within their TAD. **C)** Effects of SVs on TAD structure and characteristic resulting interaction maps. (top) Situation before the SV with two TADs separated by a boundary element. The ectopic contacts resulting from deletion (left), duplication (middle), and inversion (right) are shown in blue. As apparent in the interaction maps, all three can cause genes to interact with regulatory elements in the neighboring TAD. The reference genome projections (bottom) show characteristic interaction maps for each SV type, which enables their interpretation in Hi-C analysis. Figure panels modified from Spielmann M, Lupiáñez DG, Mundlos S. Nat Rev Genet. 2018.

Figure 3: Mechanistic biases of SV formation.

A) Cancer cells accumulate variants that are advantageous (positively selected drivers), or neutral or even deleterious (passengers). Even clones with neutral or deleterious events can expand if they coincide with drivers, or simply due to chance. **B)** The probability of DSB formation at a genomic locus depends on a combination of its characteristics, including proximity to active transcription (e.g. due to TOP2B-associated DSBs), repeat elements, GC content, and chromatin organization. **C)** The probability two loci fuse is primarily determined by their 3D distance. While this often correlates with genomic distance, DSBs in flexible regions can fuse to genomically distant loci due to 3D proximities. **D)** Following a DSB event multiple response factors including PARP1, ATM/ATR, and ALC1 are recruited to modify the surrounding histones. Collectively these changes result in a change of DNA topology towards increased mobility of the break ends.

The figures for this paper were produced by Nature Reviews Cancer graphics artists and are under copyright to Nature Reviews Cancer. The figures can be viewed at : Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R. Structural variations in cancer and the 3D genome. Nature Reviews Cancer. 2022 September;22(9):533–546. PMID: 35764888; DOI: 10.1038/s41568-022-00488-9.

Figure 4: Effects on fitness that shape SVs in cancer.

A) Five mechanisms by which SVs can increase expression of nearby oncogenes. (1) Enhancer juxtaposition. A lineage-specific enhancer, which promotes tissue-specific expression, is juxtaposed with an oncogene. (2) CRE-gene fusions. Additional cis-regulatory elements such as promoters are juxtaposed with the oncogene. (3) Enhancer de novo looping. The SV generates new long-range loops involving lineage-specific enhancers. (4) Enhancer amplification. A low-activity enhancer is amplified, multiplying its effects on expression of oncogenes within its TAD. (5) Extrachromosomal amplification. Combinations of the three prior mechanisms are active in extrachromosomal amplicons. These often contain enhancers hijacked from distant sites, and their circular topology circumvents insulators, allowing enhancers from the whole amplicon to interact with contained oncogenes. **B)** Positions of oncogenic SVs and their mechanisms of activation, indicated by the colored circles below affected oncogenes.

The figures for this paper were produced by Nature Reviews Cancer graphics artists and are under copyright to Nature Reviews Cancer. The figures can be viewed at : Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R. Structural variations in cancer and the 3D genome. Nature Reviews Cancer. 2022 September;22(9):533–546. PMID: 35764888; DOI: 10.1038/s41568-022-00488-9.