REVIEW ARTICLE

Open Access

Patterns and mechanisms of structural variations in human cancer

Kijong Yi[1] and Young Seok Ju [1]

Abstract

Next-generation sequencing technology has enabled the comprehensive detection of genomic alterations in human somatic cells, including point mutations, chromosomal rearrangements, and structural variations (SVs). Using sophisticated bioinformatics algorithms, unbiased catalogs of SVs are emerging from thousands of human cancer genomes for the first time. Via careful examination of SV breakpoints at single-nucleotide resolution as well as local DNA copy number changes, diverse patterns of genomic rearrangements are being revealed. These "SV signatures" provide deep insight into the mutational processes that have shaped genome changes in human somatic cells. This review summarizes the characteristics of recently identified complex SVs, including chromothripsis, chromoplexy, microhomology-mediated breakage-induced replication (MMBIR), and others, to provide a holistic snapshot of the current knowledge on genomic rearrangements in somatic cells.

Introduction

Cancer genomics has contributed to medical oncology by providing the genomic landscape and catalog of somatic mutations of human cancers. This information holds clinically actionable targets that may be used for personalized oncology and the development of new therapeutics. In addition, because the catalog of somatic mutations is a cumulative archeological record of all the mutational processes a cancer cell has experienced throughout the lifetime of a patient, it provides a rich source of information for biologists to understand the DNA damage and repair mechanisms that function in human somatic cells[1].

Genomic alterations in cancer cells consist of two major categories: (1) small variations that include single-nucleotide variants and short indels, and (2) large variations known as chromosomal rearrangements or structural variations (SVs). SVs are rearrangements of large DNA segments (for example, chromosomal translocations), occasionally accompanying DNA copy number

alterations. Although there is no rule that clearly distinguishes the "small" and "large" variation categories, researchers currently regard 50 bp as the tentative cutoff criteria[2]. Before the era of whole-genome sequencing (WGS), tentatively regarded as prior to 2010, the comprehensive detection of SV "breakpoints" (qualitative changes) was not feasible in cancer genomes. CNAs (quantitative changes) were relatively easier to assess using classical technologies, such as comparative genomic hybridization (CGH) and genotyping microarrays[3].

Because high-throughput DNA sequencing technologies produce unbiased sequences from whole genomes within a reasonable timeframe and at a reasonable cost (i.e., < 2000 USD and < 1 week for the production of 30 × WGS data from a tumor and paired normal tissue, as of Nov 2017), many research groups, in particular, two large international consortia (The International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA)), have produced large-scale WGS data sets from a variety of common and rare tumor types during the last decade[4,5]. Various computational algorithms and tools have been developed for the sensitive and precise detection of SVs from the WGS data (reviewed in ref. [6,7]). These efforts have enabled the identification of driver SV events with
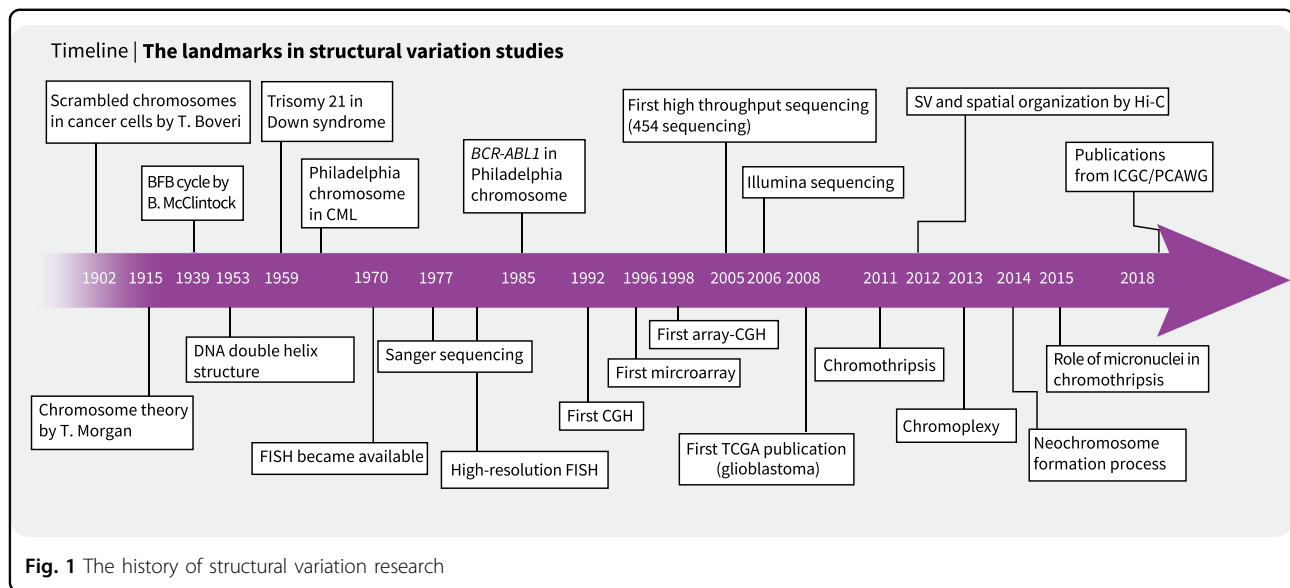
Correspondence: Young Seok Ju (ysju@kaist.ac.kr)
[1]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

Fig. 1 The history of structural variation research

remarkable functional consequences[8–14] and mechanistic patterns of SVs, which could not be identified by classical technologies. For example, the chromothripsis[15] mechanism, exhibiting a massive number of localized SV breakpoints with extensive oscillation of two DNA copy number states, was observed in cancer genome sequences, and elucidation of its molecular mechanisms followed[16–20]. However, many features remain unexplored, such as the frequency, activating conditions, and molecular machineries that are associated with the complex event. Understanding the diverse patterns of SVs observed in genome sequences is the first step to answering these questions.

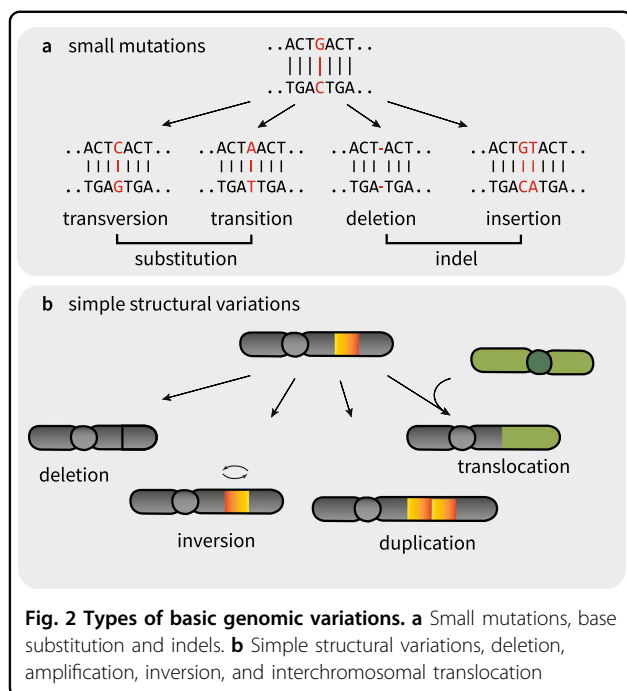## Historical overview of SVs in cancers: from cytogenetics to array CGH

The first insights into SVs in cancer cells were provided by Theodor Boveri in the early twentieth century[21] (Fig. 1). By examining dividing cancer cells under a microscope, he observed the presence of scrambled chromosomes associated with uncontrolled cell division. Following the discovery of the double helix DNA structure (1952)[22], abnormalities of the genome were proposed to cause many human diseases. For example, the trisomy of chromosome 21 in Down syndrome (1959)[23] and the recurrent translocation between chromosomes 9 and 22 (known as the Philadelphia chromosome; 1960) in chronic myelogenous leukemia (CML) were found using cytogenetics technologies[24]. As the resolution of florescence in situ hybridization (FISH) technology improved, the CML-causing *BCR-ABL1* fusion gene in the Philadelphia chromosome was identified[25]. In parallel, quantitative FISH analyses showed that some genetic loci are markedly amplified from the normal two copies in cancer cells[26,27]. Further technical improvements, such as CGH[28], array

CGH[29] and genotyping microarray[30,31], enabled genome-wide screening of CNAs in the 1990s and 2000s[27,28,32,33]. Many cancer genes have been found to be frequently amplified (i.e., *MCL1*, *EGFR*, *MYC,* and *ERBB2*) or deleted (i.e., *CDKN2A/B*, *RB1*, and P*TEN*) in cancer cells[34,35]. Indeed, genomic instability is one of the hallmarks of cancers[36,37].

Advances in hybridization technologies increased the resolution of CNA detection to ∼ 1000 base pairs. However, regardless of the resolution, these methods only approximate the genomic locations of CNAs without giving an accurate determination of the breakpoint sequences. Moreover, detection of novel copy number-neutral SVs (for example, balanced inversions and translocations) is fundamentally impossible when using array technologies. In addition, hybridization technologies are not adequate for exploring repetitive genome sequences (i.e., transposable elements)[3]. In the 2010s, advances in sequencing technologies finally enabled comprehensive, fine-scaled SV detection[4,38,39].

Patterns and mechanisms of SVs

Conventionally, cytogenetic technologies categorized SVs into four simple types: (large) deletions, duplications (amplifications), translocations, and inversions (Fig. 2). By definition, deletions and duplications are accompanied by CNAs. By contrast, inversions and translocations can be copy number neutral (balanced inversion or translocation). However, whole-genome analysis has shown that many SVs are not independent events but are acquired by a "single-hit" event and are therefore complex genome rearrangements. In this section, we introduce typical patterns of complex rearrangements found in cancers.

Fig. 2 Types of basic genomic variations. a Small mutations, base substitution and indels. **b** Simple structural variations, deletion, amplification, inversion, and interchromosomal translocation

Chromothripsis

Chromothripsis is a pattern of complex chromosomal rearrangement that is affected by a massive number of SV breakpoints, sometimes > 100, which are densely clustered in mostly one or a few chromosomal arms[40] (Fig. 3a). The term chromothripsis means "chromosome shattering into pieces" and was identified in 2011[15]. In general, chromothripsis is found in ~ 3% of all tumors and is frequently found in bone tumors (osteosarcoma and chordoma; 25%) and brain tumors (10%)[15]. However, an accurate description of its prevalence and cancer type specificity remains largely elusive.

In the typical case of chromothripsis localized in a chromosome arm, a massive number of SV elements (breakpoints) consist of similar proportions of all intra-chromosomal rearrangement types (i.e., deletion type, tandem duplication type, and head-to-head and tail-to-tail inversion types). The copy number of the involved chromosome arm usually oscillates between the normal and deleted copy number states. In addition, loss-of-heterozygosity (LOH) is frequently observed in the low-DNA copy number regions. The simplest model for explaining the chromothripsis pattern is that a single "catastrophic hit" shatters one or a few chromosome arms into hundreds of DNA segments simultaneously in an ancestral region of cancer cells, and DNA repair pathways (presumably non-homologous end-joining) reassemble the fragments in an incorrect order and orientation[15]. DNA segments that are not rejoined during the repair process result in deletions. Although such a scenario explains the features of chromothripsis, the nature of the

catastrophic hit is not fully understood. At present, two non-mutually exclusive mechanisms have been experimentally shown: (1) telomere crisis with telomere shortening and end-to-end chromosomal fusions followed by the formation of a chromatin bridge[41], and (2) micronuclei formation due to mis-segregated chromosomes during mitosis[18].

A telomere is the DNA sequence region at the end of a chromosome that protects the chromosome. When telomeres are shortened, the ends of chromosomes (chromatids) can be fused, forming a dicentric chromosome that fails to segregate into daughter cells during mitosis. The fused sites are then stretched during the anaphase of mitosis[41], forming a chromatin bridge. Under certain circumstances, the bridge induces a partial rupture of the nuclear membrane in anaphase, and the nuclease activity of the 3′ repair exonuclease 1 (*TREX1*) generates extensive single-strand DNA and bridge breakage[42]. The frequently observed SV spectrums in the daughter cells are genomic rearrangements recapitulating known features of chromothripsis combined with localized hyper-point mutations (*kataegis*)[42]. This mechanism explains why chromothripsis frequently occurs in the vicinity of telomeric regions.

Alternatively, a physical isolation of chromosomes in aberrant nuclear structures (micronuclei) was proposed as a possible mechanism of chromothripsis[18,19]. Micronuclei are frequently caused by errors in cell division, such as mis-segregation of intact chromosomes during mitosis[43] and acentric genome fragments from abnormal DNA replication/repair processes[19,20,44]. Molecular processes in micronuclei are known to be error prone; thus, isolated genetic materials are massively broken into pieces and reassembled[18,19,45]. The rejoined DNA fragments, showing chromothripsis-like features, can be fixed in a daughter cell.

Chromoplexy

Chromoplexy is another pattern of complex rearrangements that has many interdependent SV breakpoints (mostly interchromosomal translocations) but usually fewer than chromothripsis. This phenomenon was identified in prostate cancer genomes[46]. Chromoplexy mechanisms frequently disrupt tumor suppressor genes (i.e., *PTEN*, *TP53*, and *CHEK2*) and activate oncogenes by the formation of fusion genes (i.e., *TMPRSS2-ERG*) in the cancer type. The prevalence in prostate cancer is ~ 90%, but chromoplexy has not yet been explored in other cancer types. Conceptually, chromoplexy is an extended version of balanced translocation that reshuffles multiple chromosomes (rather than two chromosomes, as in balanced translocations) in a new scrambled configuration (Fig. 3b). Therefore, SVs in a chromoplexy event usually involve multiple chromosomes (usually > 3), and its
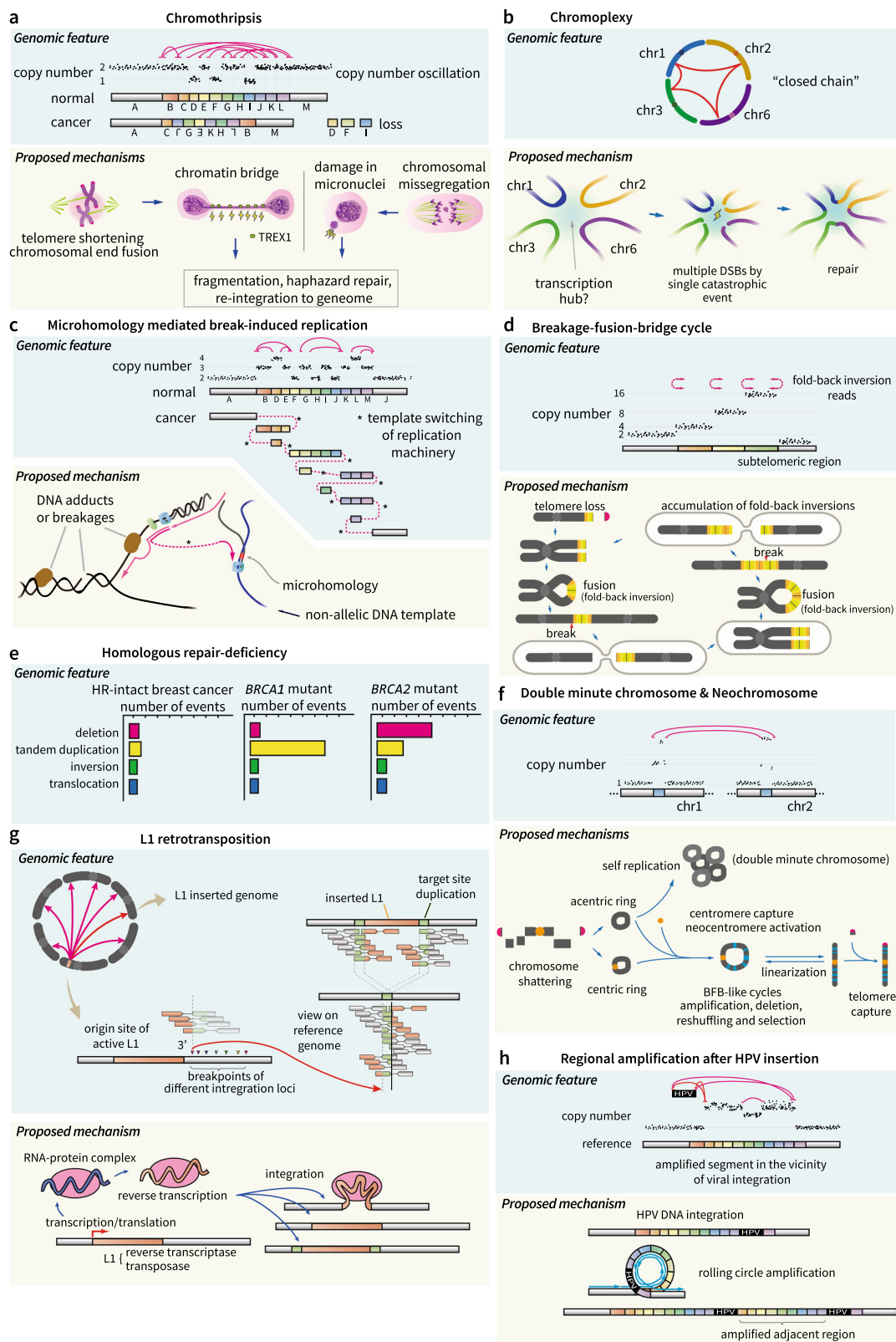
Fig. 3 (See legend on next page.)

Fig. 3 Patterns and proposed mechanisms of structural variations. a Chromothripsis, showing a shattering and subsequent repair process. Telomere crisis and/or micronuclei by chromosome mis-segregation may induce chromothripsis. **b** Chromoplexy, showing a "closed chain" (upper) in the Circos plot. This is a multi-chromosomal translocation (lower). **c** MMBIR by template switching of the replication machineries. **d** BFB cycle, showing subtelomeric copy number increases and fold-back inversions. Proposed mechanisms are shown below. **e** Different patterns of SVs in *BRCA1*- and *BRCA2*-mutant breast cancers. **f** Patterns and formation of DMs and neochromosomes. DNA fragments can self-ligate, forming a ring structure, and are amplified (DMs). Fragments capturing centromeres and telomeres become neochromosomes. **g** Patterns and processes of L1 retrotransposition in the cancer genome. **h** HPV integration and regional rolling-circle amplification

rearrangement pattern resembles a "closed chain". Although small deletions can occasionally be combined in the vicinity of the breakpoints as a form of "deletion bridge", a large fraction of SVs in a chromoplexy event is copy number neutral. Like chromothripsis, chromoplexy is readily explained by the presence of a catastrophic hit that produces multiple DNA double-strand breaks (DSBs). Unlike chromothripsis, multiple DSBs in chromoplexy are not confined to a chromosome arm but are rather distributed across many chromosomes[46].

Although the phenomenon is found in many common cancers (including prostate cancers, non-small cell lung cancers, head and neck cancers, and melanomas[46]) and rare solid cancers[47], the molecular basis of the catastrophic hit is unclear. The genome-wide distribution of DSBs in a chromoplexy event is not random but is enriched in actively transcribed and open chromatin regions[48–50]. This suggests that a nuclear transcription hub wherein many co-regulated genomic regions are spatially aggregated is fragmented by the catastrophic blow in chromoplexy[46].

Microhomology-mediated break-induced replication

The basic mechanisms of chromothripsis and chromoplexy are massive "shatter-and-stitch" processes of the genome. In these mechanisms, copy number gains of DNA segments are rarely observed. Cancer genomes frequently harbor another pattern of complex rearrangements, demonstrating a massive number of interspersed copy number gains (amplifications) of one parental allele without evidence of LOH These amplicons are directly interconnected with frequent templated insertions and common microhomologies (2–15 bps) at breakpoint junctions. These features suggest a replication-based mechanism for the acquisition of extra DNA copies, with frequent template switching of the DNA replication complex for the rearrangement (Fig. 3c). The replication-based model, termed microhomology-mediated break-induced replication (MMBIR), was initially suggested to explain the patterns of germline CNAs[51,52]. Presumably, translesion DNA polymerases, such as Polζ and Rev1, are responsible for MMBIR[53].

The cellular conditions that induce MMBIR are not fully understood. Presumably, collapse of a replication fork due to a single-strand DNA break and/or a bulky DNA adduct in the template DNA (collectively referred to as replication stress) interferes with normal DNA replication and stimulates template switching[54,55]. Normally, the template switching contributes to the repair of broken replication forks using a sister chromatid. However, the process is a double-edged sword that may lead to chromosomal rearrangements when non-allelic chromosomal regions are selected as the template. A lack of Rec/RAD proteins (e.g., RAD51) due to persistent replication stress has been reported to trigger MMBIR[51,56].

Breakage-fusion-bridge cycle

The breakage-fusion-bridge (BFB) cycle, first discovered by Barbara McClintock[57] in 1939, is a recursive cycle of generation of the dicentric chromosome by telomere fusions and breaks when the two centromeres are pulled apart in anaphase (Fig. 3d). As multiple DSBs occur in random positions in the middle of the two centromeres over a few cell cycles, the BFB cycle leaves typical patterns of rearrangements, including (1) the stair-like increase in subtelomeric regions[58] (reviewed in ref. [41]) and (2) the enrichment fold-back inversions in the breakpoints. BFB cycle-mediated SVs have been well demonstrated in a subtype of acute lymphoblastic leukemia, which exhibits intrachromosomal amplification of chromosome 21 involving *RUNX1* gene alteration[59,60].

Homologous recombination repair defect

Homologous recombination (HR) is a basic cellular mechanism to repair DSBs using identical or similar DNA sequences[61]. The basic steps of HR are (1) resection of the 5′ extremes of DSBs, (2) invasion of overhanging 3′ ends to a similar or identical DNA segment, and (3) DNA repair using one of two pathways—double-Holliday junction (reviewed in ref. [62]) or synthesis-dependent strand annealing (reviewed in ref. [62,63]).

The defect of HR (for example, *BRCA1* and *BRCA2* inactivation) causes genomic instability and increases the incidence of breast and ovarian cancers[64,65]. Complete

inactivation of *BRCA1* and/or *BRCA2* genes are found in 7% of all breast cancers[66], with an enrichment in the triple-negative breast cancer subtype[67]. BRCA gene-mutant breast cancers have a much higher burden of genome-wide SVs compared to ordinary breast cancers[68]. Interestingly, specific patterns of SVs are found according to the inactivated genes (Fig. 3e). For example, BRCA1-inactive cancers dominantly harbor short (< 10 kb) tandem duplications, but BRCA2-mutant cancers primarily show deletions[68]. Generally, BRCA1 recognizes DNA double-strand breaks along with ATM, TP53, and CHEK2 in the HR pathway. BRCA2 has an important role in the loading of RAD51[69,70], which is necessary for strand invasion after 5′-end resection[71].

The HR defect has been of interest in clinical research fields because HR-defective cancers are susceptible to targeted therapies (PARP inhibitors) that inhibit the base excision repair pathway. This strategy aims to trigger additional genomic instability in HR-defective cancer cells (but not in normal cells), which leads to cancer cell death[72]. Breast cancer patients with germline BRCA1/BRCA2 mutations are responding well to PARP inhibitor therapy[73,74].

Double-minute chromosome and neochromosome

Double-minute chromosomes (DMs) are aberrant genomic segments in a small circular form that are self-replicable but lack a centromere (Fig. 3f). DMs are often massively amplified in various solid and hematologic cancer cells[75]. DMs are detected in ~ 40% of glioblastomas, and some oncogenes, such as *CDK4*, *MDM2*, and *EGFR*, are frequently co-amplified in DMs[76,77]. DMs are important in tumorigenesis and tumor clonal evolution[78,79]. DM segments can be derived from DNA fragments that fail to be reassembled during chromothripsis[15].

Neochromosomes are aberrant genomic segments in either circular or linear forms. Unlike DMs, neochromosomes harbor a centromeric structure and (if linear) `telomeric regions (Fig. 3f). Neochromosomes are observed in ~ 3% of all cancers and are especially frequent in a subset of mesenchymal tumors, including parosteal osteosarcomas (90%), atypical lipomatous tumors (85%), dedifferentiated liposarcomas (82%), and dermatofibrosarcoma protuberans (67%)[80]. The formation process of neochromosomes has been elucidated in detail from liposarcoma genomes[81]. Like DMs, neochromosomes begin as circular DNA structures. The intermediate structures subsequently capture centromeres and are finally linearized by the acquisition of telomeres at both ends due to concurrent rearrangements, including chromothripsis- and BFB cycle-like processes.

Transposition of mobile elements

Transposable elements (TEs) are repetitive DNA sequences that occupy 45% of the human genome[82]. In the human genome, these elements are successful parasitic units that have important roles in genome evolution by generating SVs via "cutting-and-pasting" (DNA transposons) or "copying-and-pasting" themselves (retrotransposons)[83]. Most of the TEs in human genomes are now truncated and inactive in both germline and somatic lineages. For example, of the 500,000 copies of the L1 retrotransposons[84,85] in the human genome, only ~ 100 L1 copies have intact open reading frames and are potentially capable of retrotransposition. In cancer cells, retrotranspositions of L1 are frequently observed (Fig. 3g)[86,87] in ~ 50% of pan-cancer tissues[86,88], with a high enrichment in esophageal cancers (> 90%), colon cancers (> 90%) and squamous cell lung cancers (> 90%)[86,88]. L1 retrotransposition is carried out by transcription, processing, reverse transcription, and novel insertion[89]. In some cases, hundreds of somatic retrotranspositions are observed in a cancer cell. In addition, L1 retrotranspositions occasionally carry adjacent non-repetitive DNA sequences (termed transduction), which can widely scatter genes, exons and regulatory elements across the genome[86]. The functional impacts of retrotranspositions in the pathogenesis of cancers are emerging[90]. The retrotranspositional insertion sites are enriched in the heterochromatin and hypomethylated regions[91], and cancer-related genes are sometimes affected[87,90,92−94].

Insertion of external DNA sequences

In addition to reshuffling of the nuclear genomes mentioned above, cancer cells may acquire completely new extranuclear DNA sequences from viruses, mitochondria[95,96] and bacteria[97,98]. For example, the vast majority of uterine cervical cancers (> 95%) and a substantial fraction of head and neck cancers (12%) contain human papillomavirus (HPV) DNA sequences in their genome[99]. HPV genome integration is involved in direct tumorigenesis (i.e., inhibition of the p53 pathway by the HPV oncoprotein E6[100]) and in the induction of genomic instability[101]. For example, the insertional sites of HPV are frequently amplified[102] by the "loop-mediated mechanism"[101] (Fig. 3h). If brief, the insertional regions tend to form a loop structure, which is susceptible to amplification during DNA replication. As a result, genomic DNA segments flanked by viral insertions can be massively amplified, occasionally by > 50 copies, which leads upregulation of the viral oncoprotein and co-amplified adjacent gene products[101].

Intracellular nuclear transfers of full or partial mitochondrial DNA sequences are also observed in cancer genomes[95,103−105]. The prevalence of this event is ~ 2% of

Table 1 Selected examples of genes altered by structural variation in cancers

Type	Malignancy	Affected gene (prevalence)
Deletion	Retinoblastoma	*RB* (~ 100%)[131]
	Renal cell carcinoma	*VHL* (90% of clear cell type)
Amplification	Invasive breast carcinoma	*ERBB2* (18–25%)
	Neuroblastoma	*MYCN* (20–25%)
	Diffuse large B cell lymphoma	*BCL2* (31%)
	Acute myeloid leukemia	*MLL, ALL1* (5–10%)
	Gastric adenocarcinoma	*FGF4* (7%)
Fusion gene	Lung adenocarcinoma	*ALK* (3.5%)
	Prostatic adenocarcinoma	*TMPRSS2-ETS family* (29–43%)
	Chronic myelogenous leukemia	*ABL-BCR* (~ 100%)
	Burkitt lymphoma	*MYC-IGH* (~ 100%)
	Ewing Sarcoma	*EWSR1-FLI1* (~ 100%)
Enhancer hijacking	Medulloblastoma	*GFI1* family gene activation[132]
	Salivary gland adenoid cystic carcinoma	*MYB* gene overexpression[133]
	T-lymphoblastic leukemia/lymphoma	*TAL1* overexpression (30%)[134,135]
	Lung squamous cell carcinoma	*IRS4* overexpression[119]

all cancers, with an enrichment in skin, lung, and breast cancers[96]. However, the molecular mechanism by which mitochondrial DNA is mobilized and inserted into nuclear genomes has not been fully elucidated. Most somatic nuclear integrations of mitochondrial DNA do not occur alone but are frequently combined with other complex rearrangements, suggesting that mitochondrial DNA fragments could be used as a "filler material" or a string for weaving broken nuclear DNA segments into the DNA repair processes in somatic cells[106].

Comprehensive signatures of SVs

Beyond the rearrangement patterns mentioned above, additional mechanisms presumably remain undetermined. Many ongoing efforts are being carried out to reveal comprehensive SV mutational signatures in cancer genomes. For example, > 30 mutational signatures have been revealed for point mutations from the statistical analysis of large catalogs of mutations[107]. Similar concepts have been applied to SVs in breast cancer genomes by clustering genome-wide SVs according to their features, such as local proximities, rearrangement class (tandem duplication, deletion, inversion, and translocation), and rearrangement size[68]. The analysis yielded six rearrangement signatures: (1) large (> 100 kb) tandem duplication, (2) dispersed translocation, (3) small tandem duplication, (4) clustered translocation, (5) deletion, and (6) other clustered rearrangements. Among these signatures, tandem duplications (SV signatures 1 and 3) are thought to

occur due to HR deficiency[108]. In a similar manner, Li et al.[109] identified nine SV signatures from a cohort of > 2500 cancer genomes. Using this classification, they inferred that a considerable proportion of rearrangements are caused by replication-based mechanisms.

Large-scale genome studies have revealed that SVs are not evenly distributed across the genome. The density of SVs is affected by local genome and epigenome features as well as by 3D genome conformation[110–112]. For example, local rearrangement rates are affected by replication time, transcription rate, GC content, methylation status[113,114], and chromosomal fragile sites[115,116], including chromosome loop anchor sites[117]. More systematic analyses combining genome and epigenome features from a larger cohort will likely yield a better definition of the structural variation signatures and additional mutational processes in human cancers.

Functional consequences of SVs

SVs have functional consequences in tumorigenesis and clonal evolution via at least four direct mechanisms (Table 1): (1) truncation of genes (for example, deletion or gene disruption)[8,118], (2) amplifications of whole genes and their expression levels by the "dosage effect", (3) fusion gene formation (for example, *BCR-ABL* in CML and *EML4-ALK* in lung cancers) and (4) mobilization of gene-regulatory element organization ('enhancer hijacking')[2]. The first three mechanisms are conventional, and evidence for the fourth mechanism is actively emerging.

Examples of enhancer hijacking, which alters gene expression of cancer genes, including *IRS4*, *SMARCA1*, and *TERT*, have been reported[119]. In breast cancers, breast tissue-specific regulatory regions are recurrently duplicated[120], suggesting that positive selection pressures are strongly present. Similarly, many non-coding SVs may affect the gene expression of adjacent or distant genes by mobilizing many regulator regions or expressional quantitative trait loci[121,122]. More specifically, an experiment has shown that rearrangement involving the genomic topologically associating domain boundary can alter gene expression by altering the 3D genome structures that are involved in regulating gene expression[123].

Future direction and conclusion

The revolution of WGS provides an unbiased and comprehensive catalog of SVs in human cancer cells. Via a systematic, in-depth analysis of SV breakpoints, unique patterns and their underlying mutational processes are now emerging. However, current predominant WGS platforms producing short reads (< 500 bp) provide disintegrated data that are limited in the direct phasing of SV breakpoints. Despite many bioinformatic and statistical algorithms, the seamless reconstruction of final reassembled chromosomes is sometimes impossible with short read sequences, especially when the SVs are highly complex. In addition, SVs involved in highly repetitive regions (for example, telomeres, centromeres, and simple repeats) cannot be fully explored using these technologies. To this end, the combination of long read sequences (for example, from the PacBio platform) and high-resolution cytogenetics data will be helpful. Alternatively, Hi-C can be used to detect SVs in a high-throughput manner[124], although the cost efficiency could be an issue. If culturing somatic cells in vitro is possible, the Strand-seq[125] technique can provide fully phased data even if the subjects are not diploid. Single-cell genome sequencing is also a promising technology. For example, single-cell whole-genome sequencing could determine the exact timing of an SV per cell cycle[126].

Apart from the technical limitations of DNA sequencing, the accurate molecular mechanisms of SVs are difficult to elucidate because tissue sequencing primarily reflects only the terminal results of SVs. Although we can observe DSBs under a microscope[127] or with special sequencing technology (e.g., END-seq[128]), observing the DSBs and final rearrangements (outcome) at the sequence level in the same cell is currently impossible to. Well-designed experiments and analyses are needed to bridge this gap.

Understanding the functional consequences of SVs and their association with drug efficacy are important for precision medicine. For accurate functional analyses, the "genome sequencing-only" approach is limited, and the integration of multiomics data, such as genome, transcriptome, and epigenome data, are needed. Data representing the association between gene expression and the variation in the genome are being collected in the GTEx project[121]. Information on the regulatory region of the genome and the genomic regions interacting with it is actively accumulating in the ENCODE[129] and FANTOM projects[130,131]. By integrating these data sets, we will be able to comprehensively interpret the functional consequences of genome SVs and further advance precision oncology in the near future.

Conflict of interest
The authors declare that they have no conflict of interest.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
2. Carvalho, C. M. B. & Lupski, J. R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **17**, 224–238 (2016).
3. Trask, B. J. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat. Rev. Genet.* **3**, 769–778 (2002).
4. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
5. Joly, Y., Dove, E. S., Knoppers, B. M., Bobrow, M. & Chalmers, D. Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLoS. Comput. Biol.* **8**, e1002549 (2012).
6. Guan, P. & Sung, W.-K. Structural variation detection using next-generation sequencing data: a comparative technical review. *Methods* **102**, 36–49 (2016).
7. Tattini, L., D'Aurizio, R. & Magi, A. Detection of genomic structural variants from next-generation sequencing data. *Front. Bioeng. Biotechnol.* **3**, 92 (2015).
8. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).
9. Morris, S. W. et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
10. Takeuchi, K. et al. KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. *Clin. Cancer Res.* **15**, 3143–3149 (2009).
11. Parker, B. C. & Zhang, W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin. J. Cancer* **32**, 594–603 (2013).
12. Uguen, A. & De Braekeleer, M. ROS1 fusions in cancer: a review. *Future Oncol.* **12**, 1911–1928 (2016).
13. Kumar-Sinha, C., Kalyana-Sundaram, S. & Chinnaiyan, A. M. Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med.* **7**, 129 (2015).

14. Ju, Y. S. et al. A transforming KIF5B and RET gene fusion in lung adeno-carcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res.* **22**, 436–445 (2012).

15. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).

16. Kloosterman, W. P. et al. Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. *Cell Rep.* **1**, 648–655 (2012).

17. Malhotra, A. et al. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res.* **23**, 762–776 (2013).

18. Zhang, C.-Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).

19. Crasta, K. et al. DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).

20. Hoffelder, D. R. et al. Resolution of anaphase bridges in cancer cells. *Chromosoma* **112**, 389–397 (2004).

21. Boveri, T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J. Cell Sci.* **121**, 1–84 (2008).

22. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

23. Lejeune, J., Gautier, M. & Turpin, R. [Study of somatic chromosomes from 9 mongoloid children]. *C. R. Hebd. Seances Acad. Sci.* **248**, 1721–1722 (1959).

24. Nowell, P. C. & Hungerford, D. A. Chromosome studies on normal and leukemic human leukocytes. *J. Natl. Cancer Inst.* **25**, 85–109 (1960).

25. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. & Grosveld, G. Structural organization of the bcr gene and its role in the Ph' translocation. *Nature* **315**, 758–761 (1985).

26. Cheung, V. G. et al. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953–958 (2001).

27. Sebat, J. et al. Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).

28. Kallioniemi, A. et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).

29. Pinkel, D. et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**, 207–211 (1998).

30. Chee, M. et al. Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).

31. Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. & Collins, F. S. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.* **14**, 441–447 (1996).

32. Iafrate, A. J. et al. Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).

33. Redon, R. et al. Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).

34. Beroukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).

35. Scheble, V. J. et al. ERG rearrangement is specific to prostate cancer and does not occur in any other common tumor. *Mod. Pathol.* **23**, 1061–1067 (2010).

36. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability–an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).

37. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

38. Consortium, GenomesProject et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). 1000.

39. Kim, J.-I. et al. A highly annotated whole-genome sequence of a Korean individual. *Nature* **460**, 1011–1015 (2009).

40. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).

41. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175–186 (2017).

42. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).

43. Janssen, A., van der Burg, M., Szuhai, K., GJPL, Kops & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333**, 1895–1898 (2011).

44. Terradas, M., Martín, M., Tusell, L. & Genescà, A. Genetic activities in micronuclei: is the DNA entrapped in micronuclei lost for the cell? *Mutat. Res.* **705**, 60–67 (2010).

45. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154**, 47–60 (2013).

46. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).

47. Lee, J. K. et al. Complex chromosomal rearrangements by single catastrophic pathogenesis in NUT midline carcinoma. *Ann. Oncol.* **28**, 890–897 (2017).

48. Marnef, A., Cohen, S. & Legube, G. Transcription-coupled DNA double-strand break repair: active genes need special care. *J. Mol. Biol.* **429**, 1277–1288 (2017).

49. Sima, J. & Gilbert, D. M. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. *Curr. Opin. Genet. Dev.* **25**, 93–100 (2014).

50. Chiarle, R. et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).

51. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS. Genet.* **5**, e1000327 (2009).

52. Liu, P. et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **146**, 889–903 (2011).

53. Sakofsky, C. J. et al. Translesion polymerases drive microhomology-mediated break-induced replication leading to complex chromosomal rearrangements. *Mol. Cell* **60**, 860–872 (2015).

54. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).

55. Minca, E. C. & Kowalski, D. Replication fork stalling by bulky DNA damage: localization at active origins and checkpoint modulation. *Nucleic Acids Res.* **39**, 2610–2623 (2011).

56. Sakofsky, C. J., Ayyar, S. & Malkova, A. Break-induced replication and genome stability. *Biomolecules* **2**, 483–504 (2012).

57. McClintock, B. The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl. Acad. Sci. USA* **25**, 405–416 (1939).

58. Greenman, C. D., Cooke, S. L., Marshall, J., Stratton, M. R. & Campbell, P. J. Modeling the evolution space of breakage fusion bridge cycles with a stochastic folding process. *J. Math. Biol.* **72**, 47–86 (2016).

59. Robinson, H. M., Harrison, C. J., Moorman, A. V., Chudoba, I. & Strefford, J. C. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes Chromosomes Cancer* **46**, 318–326 (2007).

60. Li, Y. et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).

61. Weiner, A., Zauberman, N. & Minsky, A. Recombinational DNA repair in a cellular context: a search for the homology search. *Nat. Rev. Micro* **7**, 748–755 (2009).

62. Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551–564 (2009).

63. Sekelsky, J. DNA repair in drosophila: mutagens, models, and missing genes. *Genetics* **205**, 471–490 (2017).

64. Friedman, L. S. et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat. Genet.* **8**, 399–404 (1994).

65. Wooster, R. et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).

66. Antoniou, A. et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* **72**, 1117–1130 (2003).

67. Sharma, P. et al. Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing. *Breast Cancer Res. Treat.* **145**, 707–714 (2014).

68. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

69. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell* **69**, 457–470 (1992).

70. Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265**, 1241–1243 (1994).

71. Baumann, P., Benson, F. E. & West, S. C. Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* **87**, 757–766 (1996).

72. Weil, M. K. & Chen, A. P. PARP inhibitor treatment in ovarian and breast cancer. *Curr. Probl. Cancer* **35**, 7–50 (2011).

73. Von Minckwitz, G. et al. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): a randomised phase 2 trial. *Lancet Oncol.* **15**, 747–756 (2014).

74. Sikov, W. M. et al. Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). *J. Clin. Oncol.* **33**, 13–21 (2015).

75. Thomas, L., Stamberg, J., Gojo, I., Ning, Y. & Rapoport, A. P. Double minute chromosomes in monoblastic (M5) and myeloblastic (M2) acute myeloid leukemia: two case reports and a review of literature. *Am. J. Hematol.* **77**, 55–61 (2004).

76. Sanborn, J. Z. et al. Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res.* **73**, 6036–6045 (2013).

77. 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. USA* **101**, 11368–11373 (2004).

78. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).

79. deCarvalho, A. C. et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat. Genet.* **50**, 708–717 (2018).

80. Garsed, D. W., Holloway, A. J. & Thomas, D. M. Cancer-associated neochromosomes: a novel mechanism of oncogenesis. *Bioessays* **31**, 1191–1200 (2009).

81. Garsed, D. W. et al. The architecture and evolution of cancer neochromosomes. *Cancer Cell.* **26**, 653–667 (2014).

82. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

83. Piégu, B., Bire, S., Arensburger, P. & Bigot, Y. A survey of transposable element classification systems–a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet. Evol.* **86**, 90–109 (2015).

84. Hancks, D. C. & Kazazian, H. H. Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22**, 191–203 (2012).

85. Clapp, J. et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 264–279 (2007).

86. Tubio, J. M. C. et al. Mobile DNA in cancer. Extensive transduction of non-repetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343 (2014).

87. Miki, Y. et al. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).

88. Rodriguez-Martin B., et al. Pan-cancer analysis of whole genomes reveals driver rearrangements promoted by LINE-1 retrotransposition in human tumours. *BioRxiv* 2017. https://doi.org/10.1101/179705.

89. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annu. Rev. Genom. Hum. Genet.* **12**, 187–215 (2011).

90. Shukla, R. et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**, 101–111 (2013).

91. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967–971 (2012).

92. Ewing, A. D. et al. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* **25**, 1536–1545 (2015).

93. Solyom, S. et al. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* **22**, 2328–2338 (2012).

94. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* **24**, 1053–1063 (2014).

95. Ju, Y. S. et al. Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res.* **25**, 814–824 (2015).

96. Yuan Y., et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *BioRxiv* 2017. https://doi.org/10.1101/161356.

97. Riley, D. R. et al. Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. *PLoS. Comput. Biol.* **9**, e1003107 (2013).

98. Robinson, K. M., Sieber, K. B. & Dunning Hotopp, J. C. A review of bacteria-animal lateral gene transfer may inform our understanding of diseases like cancer. *PLoS. Genet.* **9**, e1003877 (2013).

99. Tang, K.-W., Alaei-Mahabadi, B., Samuelsson, T., Lindh, M. & Larsson, E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun.* **4**, 2513 (2013).

100. Lehoux, M., D'Abramo, C. M. & Archambault, J. Molecular mechanisms of human papillomavirus-induced carcinogenesis. *Public Health Genom.* **12**, 268–280 (2009).

101. Akagi, K. et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* **24**, 185–199 (2014).

102. Peter, M. et al. Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. *J. Pathol.* **221**, 320–330 (2010).

103. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).

104. Adams, K. L. & Palmer, J. D. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29**, 380–395 (2003).

105. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).

106. Ju, Y. S. Intracellular mitochondrial DNA transfers to the nucleus in human cancer cells. *Curr. Opin. Genet. Dev.* **38**, 23–30 (2016).

107. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

108. Willis, N. A. et al. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**, 590–595 (2017).

109. Li Y., et al. Patterns of structural variation in human cancer. *BioRxiv* 2017. https://doi.org/10.1101/181339.

110. Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A. & Misteli, T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat. Genet.* **34**, 287–291 (2003).

111. Zhang, Y. et al. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).

112. Wala J. A., et al. Selective and mechanistic sources of recurrent rearrangements across the cancer genome. *BioRxiv* 2017. https://doi.org/10.1101/187609.

113. Li, J. et al. Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. *PLoS. Genet.* **8**, e1002692 (2012).

114. Drier, Y. et al. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**, 228–235 (2013).

115. Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. & Debatisse, M. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* **89**, 215–225 (1997).

116. Coquelle, A., Rozier, L., Dutrillaux, B. & Debatisse, M. Induction of multiple double-strand breaks within an hsr by meganucleaseI-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. *Oncogene* **21**, 7671–7679 (2002).

117. Canela, A. et al. Genome organization drives chromosome fragility. *Cell* **170**, 507–521.e18 (2017).

118. Zheng J. Oncogenic chromosomal translocations and human cancer (Review). *Oncol. Rep.* 30, 2011–2019 (2013).

119. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2017).

120. Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* **49**, 341–348 (2017).

121. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017)..

122. Chiang, C. et al. The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699 (2017).

123. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).

124. 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).

125. Sanders, A. D., Falconer, E., Hills, M., Spierings, D. C. J. & Lansdorp, P. M. Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. *Nat. Protoc.* **12**, 1151–1176 (2017).

126. Voet, T. et al. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res.* **41**, 6119–6138 (2013).

127. Wang, H. et al. A perspective on chromosomal double strand break markers in mammalian cells. *Jacobs J. Radiat. Oncol.* **1**, 003 (2014).

128. Canela, A. et al. DNA breaks and end resection measured genome-wide by end sequencing. *Mol. Cell* **63**, 898–911 (2016).

129. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

130. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A. R. R. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

131. Rushlow, D. E. et al. Characterisation of retinoblastomas without RB1 mutations: genomic, gene expression, and clinical studies. *Lancet Oncol.* **14**, 327–334 (2013).

132. Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

133. Drier, Y. et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat. Genet.* **48**, 265–272 (2016).

134. Aplan, P. D. et al. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science* **250**, 1426–1429 (1990).

135. Liau, W. S. et al. Aberrant activation of the GIMAP enhancer by oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Leukemia* **31**, 1798–1807 (2017).