

Decoding genomic rearrangements for cancer driver discovery

Enyoung Seo^{1,2,†}, Sooyeon Park^{1,2,†}, Inho Park³, and Jinhyuk Bhin^{1,2,4,*}

¹Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, Republic of Korea, ²Department of Biomedical Sciences, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul 03722, Republic of Korea, ³Department of Pathology and Center for Precision Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul 03722, Republic of Korea, ⁴Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

*Corresponding author. jbhin@yuhs.ac

<https://doi.org/10.1016/j.mocell.2025.100272>

ABSTRACT

Somatically acquired genomic rearrangements are common genomic alterations that contribute to malignancy by altering the expression or activity of cancer-related genes in human cancer. Genomic rearrangements play a crucial role in tumor development by contributing to driver events in approximately 25% of cancer patients. Most rearrangements are non-recurrent and lack functional impact. However, some rearrangements produce functional transcripts and act as cancer drivers that may be therapeutic targets. The growing availability of whole-genome and matched RNA-sequencing data from large patient cohorts offers tremendous opportunities to identify novel, clinically relevant drivers arising from genomic rearrangements. In this review, we summarize current knowledge of driver rearrangements as therapeutic targets and highlight recent discoveries of functional transcripts such as intergenic fusions generated by noncanonical rearrangements. We also discuss computational approaches to decode rearrangement patterns and leverage large-scale whole-genome data to discover novel drivers.

© 2025 The Authors. Published by Elsevier Inc. on behalf of Korean Society for Molecular and Cellular Biology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Genomic rearrangements, Noncanonical rearrangements, Whole-genome sequencing

INTRODUCTION

Next-generation sequencing technologies have revolutionized cancer research over the past decade by enabling high-throughput, cost-effective profiling of tumor genomes. Early large-scale initiatives, such as The Cancer Genome Atlas (TCGA), pioneered the systematic characterization of somatic mutations across multiple cancer types (Bailey et al., 2018). These efforts extensively cataloged driver mutations and their prevalence, serving as an unprecedented resource to discover and validate novel oncogenic variants with potential clinical utility. However, TCGA studies primarily relied on whole-exome sequencing with limited information on genomic rearrangements formed through other mechanisms, including DNA replication errors, repair defects, and genomic instability.

Genomic rearrangements are large-scale structural changes in the genome, which include deletions, insertions, duplications, inversions, translocations, and more complex events, such as fold-back inversions, chromoplexy, and chromothripsis (Cosenza et al., 2022). These rearrangements are also dis-

tinguished into canonical types, which produce in-frame fusions between 2 known coding genes through well-characterized mechanisms, and noncanonical types, which involve atypical breakpoints such as intergenic regions, antisense orientation, complex multistep rearrangements, or regulatory element repositioning (Vellichirammal et al., 2020; Voronina et al., 2020; Yin et al., 2022). Notably, genomic rearrangements can alter local copy numbers by introducing extra copies or deleting genomic segments around breakpoints (Li et al., 2020b). In cancer, somatically acquired rearrangements usually affect cancer genes by disrupting gene structure, generating functional gene fusions, or inducing copy number variations. Whole-genome sequencing (WGS) in large cancer cohorts has become increasingly feasible due to reduced sequencing costs and it enables comprehensive characterization of genomic rearrangements. Notable consortia, such as the International Cancer Genome Consortium, the Pan-Cancer Analysis of Whole Genomes, and the UK 100,000 Genome Project, revealed different patterns of genomic rearrangement and exposed novel oncogenic drivers (Aaltonen et al., 2020; Zhang et al., 2019; Murugaesu et al., 2022). The International Cancer Genome Consortium's UK breast cancer project sequenced

[†] The authors contributed equally to this work.

560 whole genomes to reveal 6 structural variant (SV) signatures and potential driver rearrangements that directly affect cancer-related genes (Nik-Zainal et al., 2016). As a follow-up to this work, the Pan-Cancer Analysis of Whole Genomes consortium analyzed 2,600 primary cancers across 38 tumor types to expand the number of SV signatures from 6 to 16 (Li et al., 2020b). Cancer patients harbor an average of 4.6 somatic driver variants that include 1.3 driver rearrangements, which suggests their substantial contribution to tumor development (Aaltonen et al., 2020). More recently, the UK 100,000 Genome Project performed WGS on 2,023 colorectal cancers to provide a detailed landscape of driver variants with hotspot rearrangements that contribute to tumorigenesis (Cornish et al., 2024). Therefore, there have been substantial advancements in the identification of oncogenic alterations arising from genomic rearrangements due to recent large-scale WGS analyses.

Our review outlines the current landscape of precision medicine that targets oncogenic rearrangements. The review also emphasizes the clinical relevance of both canonical and noncanonical genomic rearrangements with potential therapeutic implications. While discussing the underlying sources of genomic rearrangements, we elaborate on the genomic features that result in the formation of SVs. Finally, we introduce the different computational approaches that can identify signals of positive selection against a nonrandom background shaped by these features.

PRECISION THERAPEUTICS FOR GENOMIC REARRANGEMENTS IN CLINICAL ONCOLOGY

As of 2024, according to OncoKB Level 1/2, which includes Food and Drug Administration (FDA)-recognized biomarkers (Level 1) and National Comprehensive Cancer Network guideline-based standard-of-care biomarkers (Level 2), there are approved targeted therapies for 12 types of genomic rearrangements in clinical oncology: *ABL1*, *ROS1*, *NTRK1/2/3*, *RET*, *ALK*, *FGFR1/2/3*, *PDGFRA/B*, *KMT2A*, *RARA*, *BRAF*, *NRG1*, and *JAK2* (Table 1). Collectively, 29 FDA-approved drugs target these rearrangements. Notably, these clinical rearrangements result in gene fusions that produce chimeric transcripts. Most involve kinase genes that generate kinase-domain-preserving, in-frame fusions that drive constitutive kinase activation. In addition to rearrangements, these kinases also harbor hotspot mutations or gene amplifications, suggesting multiple mechanisms of oncogenic activation (Stransky et al., 2014; Zehir et al., 2017). However, not all alterations confer equal clinical importance. This underscores the need to understand the functional significance of distinct oncogenic events for precision oncology (Berger and Mardis, 2018; Drilon et al., 2017; Katoh, 2019).

The specificity of fusion partners varies significantly among rearrangements. Some are highly partner-restricted, eg, *ABL1* that exclusively forms *BCR-ABL1* fusions in chronic myeloid leukemia is treated with kinase inhibitors including Imatinib and Dasatinib (Massimino et al., 2020). *RARA* fusion, which is predominantly involved in *PML-RARA* fusion in acute promyelocytic leukemia, responds to all-trans retinoic acid and arsenic trioxide (Zhang et al., 2021). However, rare *RARA* fusions, such

as *PLZF-RARA*, *NPM1-RARA*, and *TFG-RARA* fusions, which conserve the DNA-binding domain and ligand-binding capacity, vary in their responses to these therapies (Zhang et al., 2021). In bladder cancer, *FGFR3* rearrangements predominantly involve *TACC3* as they occur primarily through tandem duplications due to their genomic proximity (Ascione et al., 2023). This prevalence suggests selective oncogenic advantages conferred specifically by the *FGFR3-TACC3* fusion (Costa et al., 2016).

However, some clinically targetable rearrangements exhibit notable fusion partner promiscuity, such as *FGFR2*, *ALK*, and *NTRK1/2/3* (De Luca et al., 2020; Shreenivas et al., 2023; Westphalen et al., 2021). This suggests that oncogenicity primarily stems from intrinsic kinase activation rather than the fusion partner because these rearrangements fuse with diverse genomic loci, including intergenic regions. A recent study revealed that loss of exon 18 (E18), rather than specific fusion partners, is the key determinant of the oncogenic potency of *FGFR2* (Zingg et al., 2022). E18-lacking *FGFR2* fusions exhibit stronger tumorigenic potential than full-length fusions. This contrasts with *FGFR3*, where fusions occur predominantly with *TACC3*, suggesting fundamental mechanistic differences in oncogenicity between the 2 receptors (Krook et al., 2021). Similarly, *ALK* and *NTRK* rearrangements occur with various nonspecific partners and reinforce the therapeutic focus on targeting the kinase domains (Chen et al., 2021; Dai et al., 2022). Irrespective of specific partner identity, FDA-approved inhibitors ensure broad applicability by directly targeting kinase domains across various fusion contexts. Therefore, kinase inhibition is a central strategy in precision oncology to treat cancers driven by genomic rearrangements.

ONCOGENIC NONCANONICAL INTERGENIC REARRANGEMENTS IN CANCER

Traditional gene fusion studies have overlooked many rearrangements involving intergenic regions and mainly focused on rearrangements between protein-coding genes (Gao et al., 2018; Liu et al., 2025). However, recent studies have highlighted the significance of rearrangements beyond the gene body in cancer progression and treatment (Yun et al., 2020). The WGS analysis of 268 TCGA tumors showed that 62% (166/268) of rearrangements involved at least 1 intergenic breakpoint to reveal the prevalence of intergenic rearrangements (Yun et al., 2020). Many clinical cases of intergenic rearrangements in targetable kinase genes, such as *ALK*, *RET*, and *ROS1*, which showed good response to tyrosine kinase inhibitors, were identified through target panel sequencing (Cai et al., 2021; Li et al., 2020a; Paratala et al., 2018; Shlien et al., 2016; Yao et al., 2022; Zhai et al., 2025; Liao et al., 2022). Similarly, in a targeted sequencing analysis of 30,450 lung cancer patients, kinase fusions were identified in 3,411 patients, of which 538 (16%) of the 3,411 patients harbored intergenic rearrangements (Yao et al., 2022). In total, 624 kinase-intergenic rearrangements were detected in 538 patients. The most frequent kinase-intergenic rearrangement was *ALK* (117/624, 19%), followed by *EGFR* (54/624, 9%), *RET* (40/624, 6%), *ERBB2* (37/624, 6%), and *ROS1* (26/624, 4%). Of these, 316 were clinically targetable, suggesting the production of functional proteins with

preserved kinase activity, and most of these (219/316, 69.3%) retained the kinase domain. Notably, 3 out of 6 patients with intergenic-*ALK/ROS1* rearrangements responded favorably to Crizotinib, implying their therapeutic relevance (Yao et al., 2022).

The transcriptional consequences of kinase-intergenic rearrangements vary (Li et al., 2020a; Yun et al., 2020). In many cases especially when the rearrangement breakpoint occurs in the upstream region of the 3'-partner gene, canonical, in-frame fusion transcripts are produced through splicing events that skip intergenic sequences (Fig. 1A, left). This mechanism has been frequently observed in several known oncogenic fusions, such as *TMPRSS2-ETV4* and *TMPRSS2-ERG* in prostate cancer, *PTPRK-RSPO3* in colorectal cancer, and *TPM3-ROS1* in lung cancer (Yun et al., 2020). Alternatively, such rearrangements cause overexpression of the 3'-partner gene through enhancer hijacking (Fig. 1A, right). The overexpression of *IGF2BP3* in thyroid cancer has been attributed to this pattern that has been

recurrently identified in rearrangements between the upstream intergenic region of *IGF2BP3* (3'-gene) and intragenic regions of *THADA* or *WARS* (5'-gene) (Yun et al., 2020). Kinase-intergenic rearrangements also involve multiple intergenic breakpoints depending on the complexity of the underlying rearrangement, as exemplified by *EML4-ALK* fusions (Li et al., 2020a) (Fig. 1B). The resulting chimeric transcripts can be translated into fusion proteins, which may represent effective targets for kinase inhibitors. Some kinase-intergenic rearrangements are not expressed or translated, which results in nonproductive alterations due to either the loss of the transcription start site, reading frame, and splice sites or due to degradation by nonsense-mediated decay. Several studies assessed the expression of fusion genes using targeted RNA-seq panels. However, such panels are limited in detecting fusions with unexpected partners or noncanonical breakpoints. Therefore, it is still unclear whether these kinase-intergenic rearrangements are truly unexpressed or simply missed due to

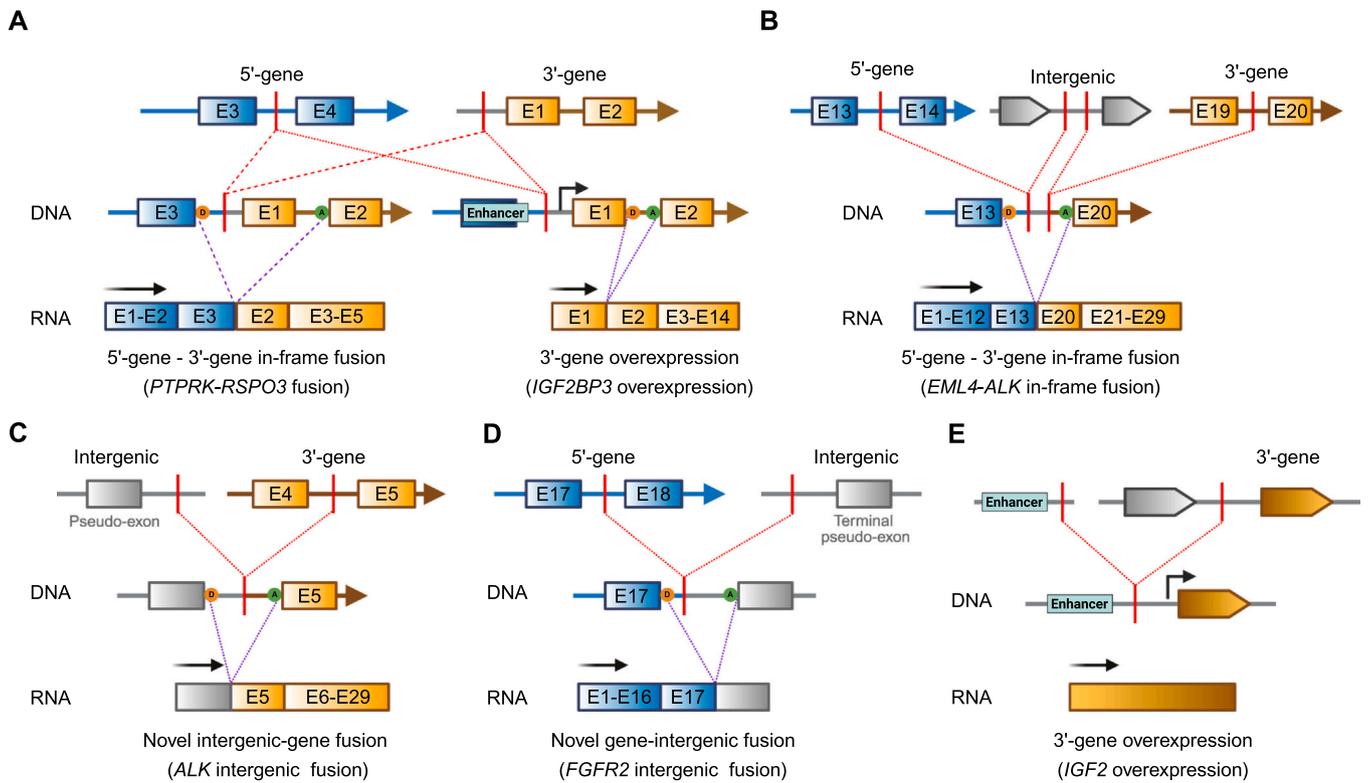


Fig. 1. Scenarios of noncanonical intergenic rearrangements forming functional transcripts. (A) Intergenic rearrangements resulting in either a canonical in-frame fusion transcript (left) or overexpression of an intact mRNA from a 3'-gene via enhancer hijacking (right). These events are often driven by a breakpoint occurring in the upstream intergenic region of the 3'-gene. Only partial exon structures are shown for both the 5'-gene (blue boxes and arrow) and the 3'-gene (yellow boxes and arrow) in the DNA. Gray line in the upstream of the 3'-gene denotes intergenic regions. Yellow circles labeled "D" and green circles labeled "A" in DNA indicate splice donor and splice acceptor sites, respectively. Red bars represent rearrangement breakpoint loci in the DNA. Red edges link breakpoint pairs forming the rearranged DNA structure, while purple edges connect exons in the resulting mRNA. Black arrows above RNA structures indicate the direction of transcription. "E" in the gene box indicates an exon. (B) Rearrangements involving multiple intergenic breakpoints resulting in a canonical in-frame fusion transcript. Gray arrow boxes represent genes located near an intergenic breakpoint. (C) Intergenic rearrangements resulting in a fusion transcript consisting of an intergenic sequence-derived pseudo-exon containing an upstream start codon (gray box) and exons following a breakpoint in the 3'-gene. (D) Intergenic rearrangements resulting in a fusion transcript consisting of exons upstream of the breakpoint in the 5'-gene and a pseudo-exon derived from the intergenic sequence containing a stop codon (gray box). (E) Intergenic rearrangements resulting in gene overexpression through enhancer hijacking, in which a pseudo-enhancer derived from the intergenic sequence induces an overexpression of the 3'-gene.

the expression of noncanonical transcripts containing novel intergenic pseudo-exons. Novel pseudo-exons derived from intergenic sequences provide splicing acceptor or donor sequences that preserve reading frames or generate novel junctions and, thus, contribute to the formation of functional transcripts (Fig. 1C and D). A previous study reported an *ALK* rearrangement that involves a breakpoint in the intergenic region between *Linc00308* and *D21S2088E*, which results in the expression of a fusion transcript comprising a novel exon derived from the intergenic region and *ALK* exon 20 (Zhang et al., 2020) (Fig. 1C). *ALK* protein level was also confirmed with *ALK* D5F3 antibody that binds to an epitope in the C-terminal portion of the *ALK* protein. Notably, the patient with this fusion showed sensitivity to Crizotinib and proved that this noncanonical *ALK* rearrangement can be a therapeutically targetable variant (Zhang et al., 2020). Pan-cancer *FGFR2* analysis also revealed several forms of noncanonical rearrangements that conserve the kinase domain but lack E18. The loss of E18 can be induced not only by canonical gene-gene fusions but also by non-canonical fusions, such as out-of-frame, out-of-strand, and intergenic rearrangements (Zingg et al., 2022) (Fig. 1D). At the RNA level, some of these fusions show supporting evidence to suggest the expression of functional transcripts lacking E18. Importantly, pharmacogenomic datasets confirmed that the loss of *FGFR2* E18 is a clinically actionable biomarker for *FGFR*-targeted therapies (Zingg et al., 2022). Another example is the *UBE3C*-intergenic fusion, a C-terminal truncating gene-intergenic rearrangement that incorporates a terminal pseudo-exon derived from an intergenic sequence, which has been implicated in distal hereditary motor neuropathies (Cutrupi et al., 2023). Enhancer hijacking or promoter swapping can generate increased expression of the intact 3'-gene mRNA when a breakpoint occurs in the upstream intergenic region of the 5'-gene and another in the downstream intergenic region of the 3'-gene (Fig. 1E). Several cancer-associated genes, including *IGF2* in colorectal cancer, *ETV1* in prostate cancer, *IGF2BP3* in thyroid cancer, and *FUT5* in breast cancer, are upregulated through this mechanism (Yun et al., 2020).

Once not considered as relevant, noncanonical rearrangements are emerging as important drivers of cancer progression and therapeutic response. Genomic and transcriptomic studies increasingly show that these rearrangements can generate functional transcripts that preserve drug-targetable domains, underscoring their clinical relevance. These findings support the broad inclusion of noncanonical rearrangement events in diagnostic and therapeutic strategies, especially for kinase-driven cancers.

GENOMIC FEATURES INFLUENCING THE FORMATION OF REARRANGEMENTS

Recent large-cohort WGS studies have facilitated systematic identification of oncogenic drivers mediated by genomic rearrangements (Rheinbay et al., 2020; Murugaesu et al., 2022). Similar to hotspot point mutations, the accumulation of recurrent breakpoints near or within cancer-related genes may be interpreted as a sign of positive selection for tumor evolution. These rearrangements can result in gene fusions, enhancer hijacking,

or loss of regulatory domains to promote clonal expansion that enhances proliferation, survival, or therapy resistance (Yun et al., 2020; Zingg et al., 2022). Therefore, mapping breakpoint hotspots is pivotal to discover novel oncogenic drivers. Genomic rearrangements are nonrandomly distributed across the genome, as their occurrence is influenced by the local genetic contexts (Du et al., 2019; Kidd et al., 2010). Neutral events arising from intrinsic genomic instability rather than positive selection could be the source of many recurrent rearrangements. Correcting for such background biases is essential to accurately detect functional drivers.

Due to their potential to increase the frequency of DNA double-strand breaks (DSBs) and to promote erroneous recombination between nonallelic sites, repetitive elements, including long interspersed nuclear element (LINE), short interspersed nuclear element (SINE), and long terminal repeat (LTR), are major contributors to genomic instability (Fig. 2A). LINE-1 is a transposable element comprising approximately 17% of the human genome (Beck et al., 2011; Paul et al., 2025). In cancer, it is often reactivated due to the loss of epigenetic silencing or functional p53, causing widespread insertional mutagenesis and rearrangements through a copy-and-paste mechanism (Beck et al., 2011; Paul et al., 2025). SINEs, such as Alu elements, also promote insertional mutagenesis by hijacking the enzymatic machinery of LINE-1 (Ade et al., 2013). Diverse genomic rearrangements occur due to the high sequence homology of these repeat elements that facilitate non-allelic homologous recombination (NAHR) following DSBs (Kim et al., 2016). Since NAHR is a key mechanism that underlies the formation of various rearrangements, the widespread distribution of such repetitive elements makes them prominent contributors to genomic instability (Carvalho and Lupski, 2016; Chen et al., 2014). Like Alu elements, LTR elements also contribute to rearrangement through NAHR between dispersed copies (Trombetta et al., 2016). Consequently, these repeat sequences serve as natural hotspots for DNA breakage and recombination. Therefore, it is pivotal to account for these repeat sequences in the identification of driver rearrangements.

Another important determinant of genomic rearrangements is the nucleotide composition of DNA (Fig. 2B). Genomic regions with high guanine-cytosine (GC) content are often associated with NAHR because they are usually enriched in low-copy repeats and segmental duplications, which thereby increase susceptibility to rearrangement (Jurka et al., 2004; Meunier and Duret, 2004). Additionally, high GC regions can form stable secondary structures, such as G-quadruplexes, which can cause replication fork stalling and collapse (De Magis et al., 2019). When replication stress remains unresolved, cells may enter mitosis with unrepliated DNA to form DSBs (Wilhelm et al., 2020). These DSBs are often repaired by error-prone mechanisms such as nonhomologous end-joining (NHEJ) or microhomology-mediated end-joining (MMEJ) (Owens et al., 2019; Sandoval and Labhart, 2004). These pathways are rapidly activated without using homologous templates, they increase the chance of rearrangements (Schimmel et al., 2019).

Replication timing also significantly influences the formation of rearrangements. Late-replicating regions, which are often associated with condensed, transcriptionally inactive chromatin,

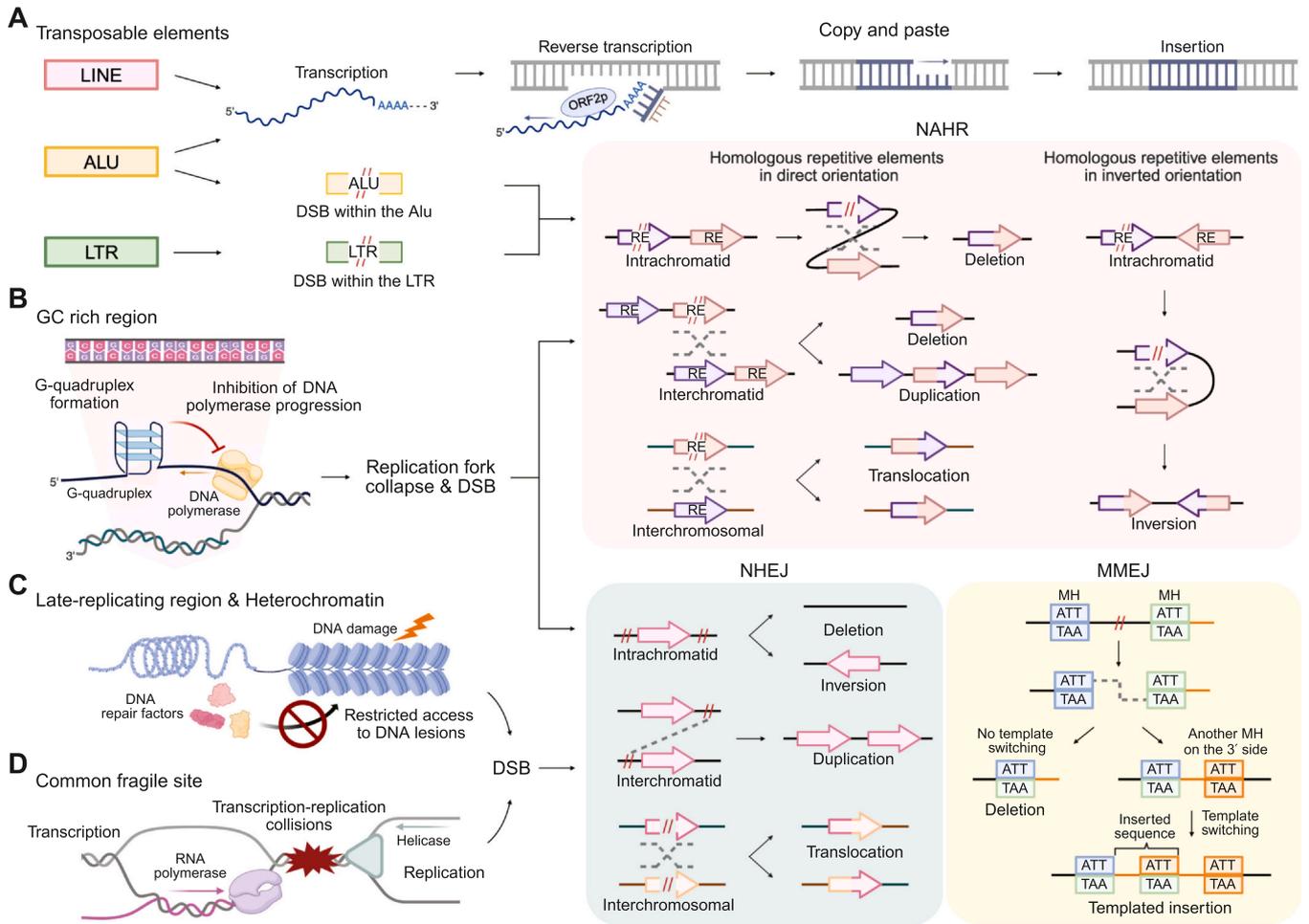


Fig. 2. Mechanisms of formation of genomic rearrangements mediated by genomic covariates. (A) Rearrangements associated with repeat elements. LINE-1 elements transpose through a copy-and-paste mechanism using their own reverse transcriptase and endonuclease, whereas SINEs such as Alu elements hijack these functions from the LINE-1 open-reading frame 2 (ORF2) protein. Structural rearrangements are promoted by sequence homology between repetitive elements, such as Alu elements or LTRs, which facilitate NAHR. (B) Rearrangements associated with GC content. High GC regions form stable secondary structures such as G-quadruplexes that stall replication forks, leading to DSBs and NAHR-, NHEJ-, and MMEJ-mediated rearrangements. (C) Rearrangements associated with replication timing and chromatin status. Late-replicating regions, often enriched in heterochromatin, exhibit restricted accessibility to DNA repair machinery with increased reliance on error-prone NHEJ and MMEJ pathways. (D) Rearrangements associated with CFSs. Under replication stress, transcription at CFSs can interfere with ongoing DNA replication to cause replication-transcription collisions. These collisions stall replication forks and could trigger fork collapse, generating DSBs that are often repaired by error-prone NHEJ or MMEJ, thereby promoting rearrangement formation. Arrow boxes labeled “RE” in the NAHR-mediated rearrangement represent repeat elements. “MH” in the MMEJ pathway denotes microhomology. MH located in different regions is distinguished by box colors, and an example MH sequence (ATT/TAA) is used to illustrate the mechanism of MMEJ-mediated rearrangement. Orange lines indicate inserted sequences. The gray dotted line indicates a homologous recombination in NAHR, end-joining in NHEJ, and end-joining through annealing of complementary MH sequences in MMEJ. For interchromosomal rearrangements, the 2 chromosomes are represented with different line colors. CFS, common fragile sites; GC, guanine-cytosine; MMEJ, microhomology-mediated end-joining; NHEJ, nonhomologous end-joining.

are more susceptible to DSBs and subsequent rearrangements (Du et al., 2019) (Fig. 2C). The limited accessibility of DNA repair machinery and homologous templates in these regions favors the use of NHEJ or MMEJ pathways (Du et al., 2019; Fortuny and Polo, 2018). Interestingly, deletions are often observed in late-replicating regions, whereas tandem duplications and unbalanced translocations are found in early-replicating regions that are often associated with euchromatic regions (Donley and Thayer, 2013). Notably, the association between replication timing, chromatin state, and rearrangement patterns

varies across cancer types, underlying the distinct rearrangement landscapes observed in different tumor types (Li et al., 2020b).

Common fragile sites (CFSs) are genomic regions prone to DSBs under replication stress, primarily due to late replication timing and paucity of replication origins. There is increased dependence on NHEJ and MMEJ pathways because DSBs at CFSs often occur in late S/G2 phase or during mitosis when homologous recombination is less active (Schwartz et al., 2005; Truong et al., 2013). CFSs are also susceptible to replication-

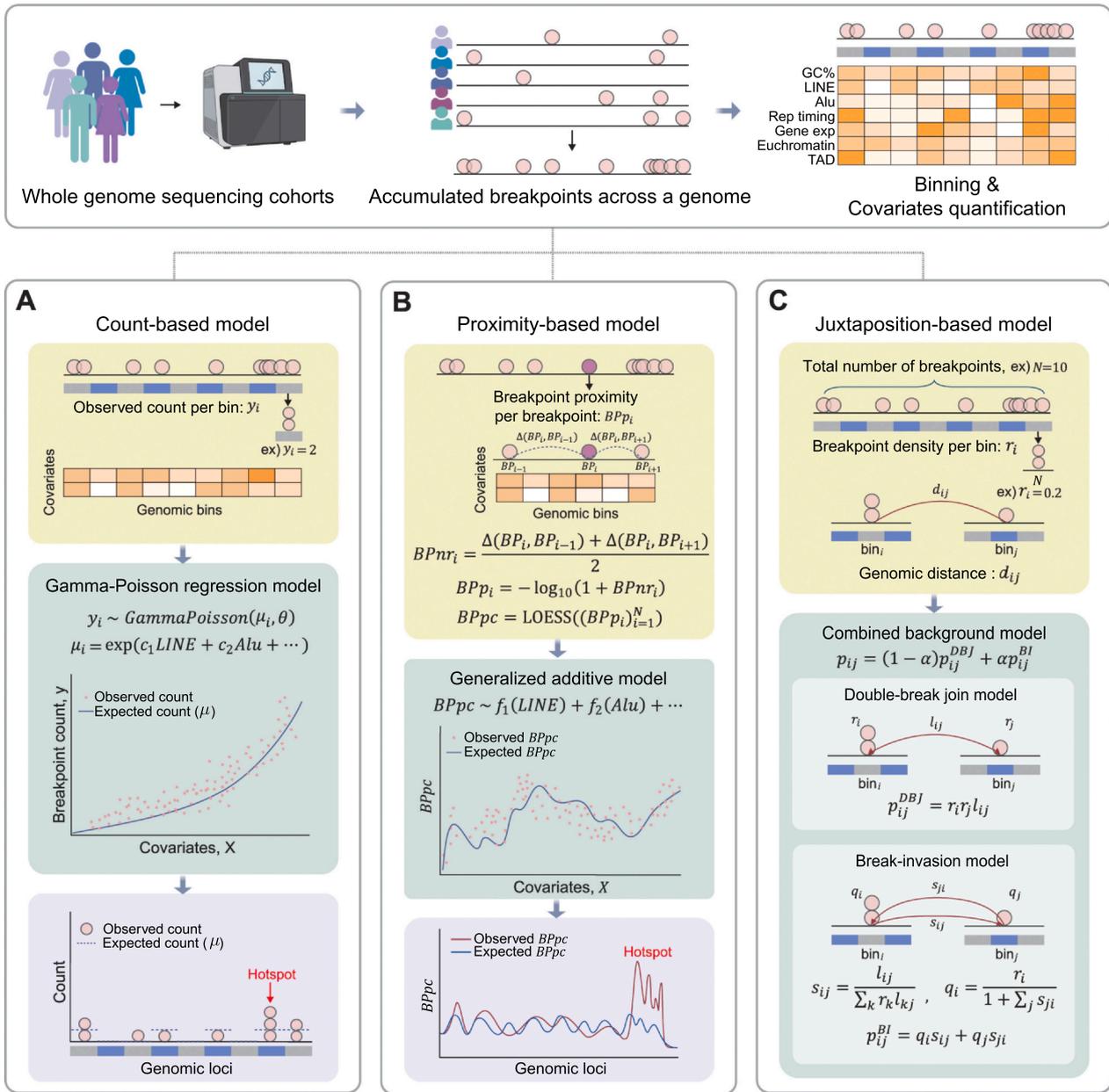


Fig. 3. Computational approaches to identify breakpoint hotspot regions. Rearrangement breakpoints are identified using WGS data of a cancer cohort. Publicly available genomic covariates, such as GC content (GC%), LINE, Alu, replication timing (Rep timing), gene expression (Gene exp), chromatin states, and topologically associated domains (TAD), are quantified at a bin-level resolution. (A) Count-based model quantifies breakpoint accumulations at bin-level resolution. Gamma-Poisson regression is used to estimate expected breakpoint counts for bin i (μ) and the dispersion by modeling the relationship between observed counts y_i and genomic covariates, with a coefficient c_j for each covariate j . Hotspots are defined as bins or segments with significantly higher observed breakpoints than expected. (B) Proximity-based model computes a proximity score (BPp_i) for each breakpoint (BP_i) as the $-\log_{10}$ of its average distance to adjacent breakpoints. d_{ij} denotes the distance between breakpoints. Locally Estimated Scatterplot Smoothing (LOESS) then smooths these values to obtain $BPpc$. Covariates, assigned to each breakpoint using covariate-specific bins, are modeled against $BPpc$ with a generalized additive model using covariate-specific smooth functions f_j to capture linear or nonlinear relationships. Hotspots are defined as regions that significantly deviate from the expected $BPpc$. (C) Juxtaposition-based model estimates the probability of background rearrangement between bins i and j (p_{ij}) as a weighted sum of p_{ij}^{DBJ} and p_{ij}^{BI} components). In the double-break join model, r_i and r_j are computed as the number of breakpoints in each bin divided by the total breakpoint count N . l_{ij} denotes the fraction of all rearrangement pairs whose distances fall within the same interval as the distance between bins i and j (d_{ij}). In the break-invasion model, s_{ij} , derived from l_{ij} , represents the conditional probability of a breakpoint at bin i invading bin j , while s_{ji} represents the reverse scenario. q_i and q_j are breakage probabilities that account for breakage events propagated from other bins. α is a weight parameter that controls the contribution of each model component.

transcription conflicts that can exacerbate replication stress and contribute to the formation of DSBs (Li and Wu, 2020) (Fig. 2D). Notably, CFSs may give rise to potentially pathogenic rearrangements in disease-associated genes, which suggests that CFSs contribute to genomic instability by shaping a landscape of rearrangements that encompass both neutral passenger and driver events (Mitsui et al., 2010).

Since genomic features inherently influence the formation of rearrangements in cancer-associated genes, background biases should be judiciously incorporated in statistical models to detect true oncogenic driver events among passenger events. It is pivotal to incorporate such features into background models to avoid misinterpreting passenger rearrangements as signals of positive selection.

COMPUTATIONAL MODELS FOR DETECTING ONCOGENIC REARRANGEMENTS

To identify oncogenic drivers mediated by genomic rearrangements from cohort-based WGS, accurate modeling of the heterogeneous mutational background is required to distinguish true driver events from passenger alterations. Recent computational efforts based on machine-learning frameworks have attempted to model the distribution of background breakpoints by capturing relationships between genomic features and rearrangement breakpoint frequencies. These efforts can be categorized as: (1) count-based modeling, (2) proximity-based modeling, and (3) juxtaposition-based modeling (Fig. 3).

Count-based modeling segments the genome into fixed-size bins and quantifies breakpoint counts per bin from a cohort to estimate local breakpoint densities (Fig. 3A). This approach aims to model background breakpoint rates that vary across the genome by incorporating genomic covariates, such as GC content, repeat elements, replication timing, chromatin accessibility, and CpG/TpC ratios (Glodzik et al., 2017; Imielinski et al., 2017). One example is FishHook that employs a Gamma-Poisson regression framework to model bin-specific background breakpoint counts as a function of these covariates (Imielinski et al., 2017). The expected counts derived from this model serve as a null distribution to identify bins with a statistically significant number of breakpoints. Genomic regions containing such bins are considered potential hotspot regions that result from positive selection (Dubois et al., 2022; Rheinbay et al., 2020; Zhou et al., 2022). Since the results of this model depend on the fixed bin size, the model may miss hotspot regions that emerge at different binning resolutions. Piecewise constant fitting, a method frequently used to detect copy number alterations, may be further applied to segment the genome based on shifts in breakpoint density to enable more flexible hotspot detection (Glodzik et al., 2017; Nik-Zainal et al., 2016; Rustad et al., 2020; Cornish et al., 2024).

Breakpoint proximity-based modeling utilizes physical distances between breakpoints to estimate local breakpoint clustering (Fig. 3B). CSVDriver computes a breakpoint proximity curve ($BPpc$) based on a smoothed curve of breakpoint neighbor reachability ($BPnr$), which reflects the distances between breakpoints (Martínez-Fundichely et al., 2022). A generalized additive model is used to compute expected $BPpc$ by

modeling the nonlinear relationship between genomic covariates and $BPpc$. Significant breakpoint clusters are then identified by comparing observed vs expected $BPpc$ values and computing a peak recurrence score to assess recurrence across samples (Martínez-Fundichely et al., 2022). This framework enables the discovery of hotspots without relying on fixed-size bins and allows for the modeling of both linear and nonlinear relationships with covariates.

Juxtaposition-based modeling focuses on identifying pairs of loci that are recurrently connected by genomic rearrangements rather than detecting breakpoint hotspots across samples within a cohort (Fig. 3C). SVSig-2D is a statistical framework that models the background frequency of rearrangements between bins i and j (p_{ij}), which is computed as a weighted mixture of 2 models, double-break join and break-invasion models (Zhang et al., 2023). In the double-break join model, the rearrangement is assumed to arise from the fusion of 2 independently formed breakpoints, as seen in NHEJ or MMEJ. This model estimates the rearrangement probability (p_{ij}^{DBJ}) as the product of the breakpoint densities at bins i (r_i) and j (r_j), and a length factor (l_{ij}) reflecting the distance between them. In contrast, the break-invasion model reflects NAHR-mediated rearrangements by assuming one breakpoint invades another. This model estimates the rearrangement probability (p_{ij}^{BI}) as the sum of 2 directional products: breakage probability at bin i (q_i) and the conditional probability of invasion into bin j (s_{ij}), and vice versa. A binomial framework is used to statistically test observed breakpoint pairs against this background to identify significantly recurrent juxtaposition hotspots (Zhang et al., 2023). This provides a more comprehensive view of SV formation by considering both initial breakage and the rearrangement process. However, further refinement is needed to fully incorporate interacting factors such as DNA repair pathways, replication timing, and 3D genome architecture into a unified model that captures the underlying mechanisms of rearrangements.

CONCLUSIONS AND FUTURE PERSPECTIVES

In the era of WGS for cancer genome profiling, the mutational processes, functions, and mechanisms of structural rearrangements have been widely explored. Ongoing large-cohort WGS projects, led by UK's 100,000 Genome Project and the Hartwig Medical Foundation, together with combined multi-cohort analyses, are expected to further advance our understanding of driver alterations mediated by genomic rearrangements. Analyses of WGS integrated with matched transcriptome and/or mass-spectrometry-based proteomics data can facilitate the identification of functionally expressed drivers arising from noncanonical mechanisms, including pseudo-exon usage from intergenic regions, frameshift events, and antisense rearrangements. Recent advances in long-read DNA- and RNA-sequencing technologies are poised to accelerate the discovery and validation of these noncanonical drivers.

Significant challenges persist although the use of machine-learning frameworks to identify novel drivers through breakpoint distribution analysis is a promising approach. Detecting true

signals of positive selection requires accurate modeling of the background distribution of breakpoint densities across the genome, which is typically predicted from genomic features. However, some of these features, such as replication timing, chromatin status, and topologically associated domains, vary within the same genomic regions depending on the cell and tissue types. This approach is limited in certain cancer types because these features are not available for all cell and tissue types. Therefore, expanding genomic feature data across diverse cellular and tissue contexts is crucial to improve the accuracy and applicability of machine-learning frameworks. Other challenges include the ambiguity in selecting an appropriate bin size for hotspot detection and the spatial correlations between adjacent bins, which are often simplistically assumed to be independent of each other in current machine-learning models.

Recent advances in deep learning models offer potential solutions to several of these challenges. Deep learning models have already been applied to structural variation research, particularly for improving SV calling (Cai et al., 2021; Lin et al., 2022; Liu et al., 2021) and genotyping (Linderman et al., 2024; Popic et al., 2023) by capturing complex, nonlinear patterns and discriminating true signals from noise in heterogeneous sequencing data. These capabilities could be leveraged to detect SV hotspots. For example, transformer-based models can learn long-range, multivariate interactions between nucleotide sequences and diverse genomic covariates (Choi and Lee, 2023). This enables inference of genomic contexts that predispose to SV formation. In addition, variational autoencoders can model the shared correlation structure of covariates into a latent space and identify anomalies as deviations from these learned patterns to provide a framework for hotspot discovery (Sado et al., 2024). Future advances in robust statistical modeling, powered by deep learning algorithms and expanding genomic feature datasets across diverse tissue types, will be critical to improve the identification of novel drivers.

AUTHOR CONTRIBUTIONS

Sooyeon Park: Writing – review & editing, Visualization.
Enyoung Seo: Writing – review & editing, Writing – original draft, Visualization, Supervision.
Inho Park: Writing – review & editing.
Jinhyuk Bhin: Writing – review & editing, Writing – original draft, Visualization, Supervision, Funding acquisition, Conceptualization.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation (NRF) of Korea grant funded by the Korean government (MIST) (RS-2024-00342142, RS-2023-00261820, and RS-2024-00406281) and a faculty research grant of Yonsei University College of Medicine (6-2024-0041).

ORCID

Enyoung Seo <https://orcid.org/0009-0006-3649-6767>
Sooyeon Park <https://orcid.org/0009-0006-5338-1726>
Inho Park <https://orcid.org/0000-0003-2250-6507>
Jinhyuk Bhin <https://orcid.org/0000-0001-8898-4531>

Received April 30, 2025
Revised August 17, 2025
Accepted August 24, 2025
Available online 30 August 2025.

REFERENCES

- Aaltonen, L.A., Abascal, F., Abeshouse, A., Aburatani, H., Adams, D.J., Agrawal, N., Ahn, K.S., Ahn, S.-M., Aikata, H., Akbani, R., et al. (2020). Pan-cancer analysis of whole genomes. *Nature*, *578*, 82-93.
- Ade, C., Roy-Engel, A.M., and Deininger, P.L. (2013). Alu elements: an intrinsic source of human genome instability. *Curr. Opin. Virol.* *3*, 639-645.
- Ascione, C.M., Napolitano, F., Esposito, D., Servetto, A., Belli, S., Santaniello, A., Scagliarini, S., Crocetto, F., Bianco, R., and Formisano, L. (2023). Role of FGFR3 in bladder cancer: treatment landscape and future challenges. *Cancer Treat. Rev.* *115*, 102530.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell*, *173*, 371-385.e18.
- Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* *12*, 187-215.
- Berger, M.F., and Mardis, E.R. (2018). The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* *15*, 353-365.
- Cai, C., Tang, Y., Li, Y., Chen, Y., Tian, P., Wang, Y., Gong, Y., Peng, F., Zhang, Y., Yu, M., et al. (2021). Distribution and therapeutic outcomes of intergenic sequence-ALK fusion and coexisting ALK fusions in lung adenocarcinoma patients. *Lung Cancer*, *152*, 104-108.
- Carvalho, C.M.B., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* *17*, 224-238.
- Chen, L., Zhou, W., Zhang, L., and Zhang, F. (2014). Genome architecture and its roles in human copy number variation. *Genomics Inform.* *12*, 136-144.
- Chen, T., Wang, Y., Goetz, L., Corey, Z., Dougher, M.C., Smith, J.D., Fox, E.J., Freiberg, A.S., Flemming, D., and Fanburg-Smith, J.C. (2021). Novel fusion sarcomas including targetable NTRK and ALK. *Ann. Diagn. Pathol.* *54*, Article 151800.
- Choi, S.R., and Lee, M. (2023). Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology (Basel)*, *12*, 1033.
- Cornish, A.J., Gruber, A.J., Kinnersley, B., Chubb, D., Frangou, A., Caravagna, G., Noyvert, B., Lakatos, E., Wood, H.M., Thorn, S., et al. (2024). The genomic landscape of 2,023 colorectal cancers. *Nature*, *633*, 127-136.
- Cosenza, M.R., Rodriguez-Martin, B., and Korbel, J.O. (2022). Structural variation in cancer: role, prevalence, and mechanisms. *Annu. Rev. Genomics Hum. Genet.* *23*, 123-152.
- Costa, R., Carneiro, B.A., Taxter, T., Tavora, F.A., Kalyan, A., Pai, S.A., Chae, Y.K., and Giles, F.J. (2016). FGFR3-TACC3 fusion in solid tumors: mini review. *Oncotarget*, *7*, 55924-55938.

- Cutrupi, A.N., Narayanan, R.K., Perez-Siles, G., Grosz, B.R., Lai, K., Boyling, A., Ellis, M., Lin, R.C.Y., Neumann, B., Mao, D., et al. (2023). Novel gene-intergenic fusion involving ubiquitin E3 ligase UBE3C causes distal hereditary motor neuropathy. *Brain*, *146*, 880-897.
- Dai, Y., Liu, P., He, W., Yang, L., Ni, Y., Ma, X., Du, F., Song, C., Liu, Y., and Sun, Y. (2022). Genomic features of solid tumor patients harboring ALK/ROS1/NTRK gene fusions. *Front. Oncol.* *12*, Article 813158.
- De Luca, A., Esposito Abate, R., Rachiglio, A.M., Maiello, M.R., Esposito, C., Schettino, C., Izzo, F., Nastì, G., and Normanno, N. (2020). FGFR fusions in cancer: from diagnostic approaches to therapeutic intervention. *Int. J. Mol. Sci.* *21*, 6856.
- De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O., and Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* *116*, 816-825.
- Donley, N., and Thayer, M.J. (2013). DNA replication timing, genome stability and cancer. *Semin. Cancer Biol.* *23*, 80-89.
- Drilon, A., Siena, S., Ou, S.-H.I., Patel, M., Ahn, M.J., Lee, J., Bauer, T.M., Farago, A.F., Wheler, J.J., Liu, S.V., et al. (2017). Safety and antitumor activity of the multitargeted Pan-TRK, ROS1, and ALK inhibitor entrectinib: combined results from two phase I trials (ALKA-372-001 and STARTRK-1). *Cancer Discov.* *7*, 400-409.
- Du, Q., Bert, S.A., Armstrong, N.J., Caldon, C.E., Song, J.Z., Nair, S.S., Gould, C.M., Luu, P.-L., Peters, T., Khoury, A., et al. (2019). Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat. Commun.* *10*, 416.
- Dubois, F.P.B., Shapira, O., Greenwald, N.F., Zack, T., Wala, J., Tsai, J.W., Crane, A., Baguette, A., Hadjadj, D., Harutyunyan, A.S., et al. (2022). Structural variants shape driver combinations and outcomes in pediatric high-grade glioma. *Nat. Cancer*, *3*, 994-1011.
- Fortuny, A., and Polo, S.E. (2018). The response to DNA damage in heterochromatin domains. *Chromosoma*, *127*, 291-300.
- Gao, Q., Liang, W.-W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.-W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al. (2018). Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* *23*, 227-238.e3.
- Glodzik, D., Morganella, S., Davies, H., Simpson, P.T., Li, Y., Zou, X., Diez-Perez, J., Staaf, J., Alexandrov, L.B., Smid, M., et al. (2017). A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* *49*, 341-348.
- Imielinski, M., Guo, G., and Meyerson, M. (2017). Insertions and deletions target lineage-defining genes in human cancers. *Cell*, *168*, 460-472.e14.
- Jurka, J., Kohany, O., Pavlicek, A., Kapitonov, V.V., and Jurka, M.V. (2004). Duplication, coclustering, and selection of human Alu retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* *101*, 1268-1272.
- Katoh, M. (2019). Fibroblast growth factor receptors as treatment targets in clinical oncology. *Nat. Rev. Clin. Oncol.* *16*, 105-122.
- Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K., and Eichler, E.E. (2010). A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell*, *143*, 837-847.
- Kim, S., Cho, C.-S., Han, K., and Lee, J. (2016). Structural variation of Alu element and human disease. *Genomics Inform.* *14*, 70-77.
- Krook, M.A., Reeser, J.W., Ernst, G., Barker, H., Wilberding, M., Li, G., Chen, H.-Z., and Roychowdhury, S. (2021). Fibroblast growth factor receptors in cancer: genetic alterations, diagnostics, therapeutic targets and mechanisms of resistance. *Br. J. Cancer*, *124*, 880-892.
- Li, S., and Wu, X. (2020). Common fragile sites: protection and repair. *Cell Biosci.* *10*, 29.
- Li, W., Liu, Y., Li, W., Chen, L., and Ying, J. (2020a). Intergenic breakpoints identified by DNA sequencing confound targetable kinase fusion detection in NSCLC. *J. Thorac. Oncol.* *15*, 1223-1231.
- Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., et al. (2020b). Patterns of somatic structural variation in human cancer genomes. *Nature*, *578*, 112-121.
- Liao, S., Sun, H., Wu, J., Lu, H., Fang, Y., Wang, Y., and Liao, W. (2022). Case report: two novel intergenic region-ALK fusions in non-small-cell lung cancer resistant to alectinib: a report of two cases. *Front. Oncol.* *12*, 916315.
- Lin, J., Wang, S., Audano, P.A., Meng, D., Flores, J.I., Kusters, W., Yang, X., Jia, P., Marschall, T., Beck, C.R., et al. (2022). SVision: a deep learning approach to resolve complex structural variants. *Nat. Methods*, *19*, 1230-1233.
- Linderman, M.D., Wallace, J., Van Der Heyde, A., Wieman, E., Brey, D., Shi, Y., Hansen, P., Shamsi, Z., Liu, J., Gelb, B.D., et al. (2024). NPSV-deep: a deep learning method for genotyping structural variants in short read genome sequencing data. *Bioinformatics*, *40*, Article btae129.
- Liu, S.V., Nagasaka, M., Atz, J., Solca, F., and Müllauer, L. (2025). Oncogenic gene fusions in cancer: from biology to therapy. *Sig. Transduct. Target Ther.* *10*, 111.
- Liu, Y., Huang, Y., Wang, G., and Wang, Y. (2021). A deep learning approach for filtering structural variants in short read sequencing data. *Brief Bioinform.* *22*, Article bbaa370.
- Martínez-Fundichely, A., Dixon, A., and Khurana, E. (2022). Modeling tissue-specific breakpoint proximity of structural variations from whole-genomes to identify cancer drivers. *Nat. Commun.* *13*, 5640.
- Massimino, M., Stella, S., Tirrò, E., Pennisi, M.S., Vitale, S.R., Puma, A., Romano, C., Gregorio, S.D., Tomarchio, C., Raimondo, F.D., et al. (2020). ABL1-directed inhibitors for CML: efficacy, resistance and future perspectives. *Anticancer Res.* *40*, 2457-2465.
- Meunier, J., and Duret, L. (2004). Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* *21*, 984-990.
- Mitsui, J., Takahashi, Y., Goto, J., Tomiyama, H., Ishikawa, S., Yoshino, H., Minami, N., Smith, D.I., Lesage, S., Aburatani, H., et al. (2010). Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *Am. J. Hum. Genet.* *87*, 75-89.
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole genome sequences. *Nature*, *534*, 47-54.
- Owens, D.D.G., Caulder, A., Frontera, V., Harman, J.R., Allan, A.J., Bucakci, A., Greder, L., Codner, G.F., Hublitz, P., McHugh, P.J., et al. (2019). Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res.* *47*, 7402-7417.
- Paratala, B.S., Chung, J.H., Williams, C.B., Yilmazel, B., Petrosky, W., Williams, K., Schrock, A.B., Gay, L.M., Lee, E., Dolfi, S.C., et al. (2018). RET rearrangements are actionable alterations in breast cancer. *Nat. Commun.* *9*, 4821.
- Paul, P., Kumar, A., Parida, A.S., De, A.K., Bhadke, G., Khatua, S., and Tiwari, B. (2025). p53-mediated regulation of LINE1 retrotransposon-derived R-loops. *J. Biol. Chem.* *301*, Article 108200.
- Popic, V., Rohlicek, C., Cunial, F., Hajirasouliha, I., Meleshko, D., Garimella, K., and Maheshwari, A. (2023). Cue: a deep-learning

- framework for structural variant discovery and genotyping. *Nat. Methods*, *20*, 559-568.
- Rheinbay, E., Nielsen, M.M., Abascal, F., Wala, J.A., Shapira, O., Tiao, G., Hornshøj, H., Hess, J.M., Juul, R.I., Lin, Z., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*, *578*, 102-111.
- Rustad, E.H., Yellapantula, V.D., Glodzik, D., Maclachlan, K.H., Diamond, B., Boyle, E.M., Ashby, C., Blaney, P., Gundem, G., Hultcrantz, M., et al. (2020). Revealing the impact of structural variants in multiple myeloma. *Blood Cancer Discov.* *1*, 258-273.
- Sado, I.T., Fitime, L.F., Pelap, G.F., Tinku, C., Meudje, G.M., and Bouetou, T.B. (2024). Early multi-cancer detection through deep learning: an anomaly detection approach using variational autoencoder. *J. Biomed. Inform.* *160*, Article 104751.
- Sandoval, A., and Labhart, P. (2004). High G/C content of cohesive overhangs renders DNA end joining Ku-independent. *DNA Repair*, *3*, 13-21.
- Schimmel, J., Van Schendel, R., Den Dunnen, J.T., and Tijsterman, M. (2019). Templated insertions: a smoking gun for polymerase theta-mediated end joining. *Trends Genet.* *35*, 632-644.
- Schwartz, M., Zlotorynski, E., Goldberg, M., Ozeri, E., Rahat, A., Sage, C.L.E., Chen, B.P.C., Chen, D.J., Agami, R., and Kerem, B. (2005). Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. *Genes Dev.* *19*, 2715-2726.
- Shlien, A., Raine, K., Fuligni, F., Arnold, R., Nik-Zainal, S., Dronov, S., Mamanova, L., Rosic, A., Ju, Y.S., Cooke, S.L., et al. (2016). Direct transcriptional consequences of somatic mutation in breast cancer. *Cell Rep.* *16*, 2032-2046.
- Shreenivas, A., Janku, F., Gouda, M.A., Chen, H.-Z., George, B., Kato, S., and Kurzrock, R. (2023). ALK fusions in the pan-cancer setting: another tumor-agnostic target? *npj Precis. Onc.* *7*, 1-20.
- Stransky, N., Cerami, E., Schalm, S., Kim, J.L., and Lengauer, C. (2014). The landscape of kinase fusions in cancer. *Nat. Commun.* *5*, 4846.
- Trombetta, B., Fantini, G., D'Atanasio, E., Sellitto, D., and Cruciani, F. (2016). Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. *Sci. Rep.* *6*, Article 28710.
- Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.-H., He, J., Wang, H., Razavian, N., Berns, M.W., and Wu, X. (2013). Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Aca. Sci.* *110*, 7720-7725.
- Vellichirammal, N.N., Albahrani, A., Banwait, J.K., Mishra, N.K., Li, Y., Roychoudhury, S., Kling, M.J., Mirza, S., Bhakat, K.K., Band, V., et al. (2020). Pan-cancer analysis reveals the diverse landscape of novel sense and antisense fusion transcripts. *Mol. Ther. Nucleic Acids*, *19*, 1379-1398.
- Voronina, N., Wong, J.K.L., Hübschmann, D., Hlevnjak, M., Uhrig, S., Heilig, C.E., Horak, P., Kreutzfeldt, S., Mock, A., Stenzinger, A., et al. (2020). The landscape of chromothripsis across adult cancer types. *Nat. Commun.* *11*, 2320.
- Westphalen, C.B., Krebs, M.G., Le Tourneau, C., Sokol, E.S., Maund, S.L., Wilson, T.R., Jin, D.X., Newberg, J.Y., Fabrizio, D., Veronese, L., et al. (2021). Genomic context of NTRK1/2/3 fusion-positive tumours from a large real-world population. *npj Precis. Onc.* *5*, 1-9.
- Wilhelm, T., Said, M., and Naim, V. (2020). DNA replication stress and chromosomal instability: dangerous liaisons. *Genes (Basel)*, *11*, 642.
- Yao, Y., Yu, Z., Ma, Y., Ou, Q., Wu, X., Lu, D., and Li, X. (2022). Characterizing kinase intergenic-breakpoint rearrangements in a large-scale lung cancer population and real-world clinical outcomes. *ESMO Open*, *7*, Article 100405.
- Yin, L., Han, Z., Feng, M., Wang, J., Xie, Z., Yu, W., Fu, X., Shen, N., Wang, X., Duan, A., et al. (2022). Chimeric transcripts observed in non-canonical FGFR2 fusions with partner genes' breakpoint located in intergenic region in intrahepatic cholangiocarcinoma. *Cancer Genet.* (266-267), 39-43.
- Yun, J.W., Yang, L., Park, H.-Y., Lee, C.-W., Cha, H., Shin, H.-T., Noh, K.-W., Choi, Y.-L., Park, W.-Y., and Park, P.J. (2020). Dysregulation of cancer genes by recurrent intergenic fusions. *Genome Biol.* *21*, 166.
- Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* *23*, 703-713.
- Zhai, X., Wang, M., Zhang, Q., Li, D., Wu, Y., Liang, Z., Liu, J., Wang, W., Liu, Y., Che, G., et al. (2025). Identifying the intergenic ALK fusion LOC388942-ALK as a driver of non-small cell lung cancer. *MedComm*, *6*, Article e70154.
- Zhang, J., Zou, C., Zhou, C., Luo, Y., He, Q., Sun, Y., Zhou, J., and Ke, Z. (2020). A novel Linc00308/D21S2088E intergenic region ALK fusion and its enduring clinical responses to crizotinib. *J. Thorac. Oncol.* *15*, 1073-1077.
- Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* *37*, 367-369.
- Zhang, S., Kumar, K.H., Shapira, O., Yao, X., Wala, J., Dubois, F., Gold, R., Haber, J.E., Cherniack, A., Imieliński, M., et al. (2023). Detecting significantly recurrent genomic connections from simple and complex rearrangements in the cancer genome. *bioRxiv*, 561748.
- Zhang, X., Sun, J., Yu, W., and Jin, J. (2021). Current views on the genetic landscape and management of variant acute promyelocytic leukemia. *Biomark. Res.* *9*, 33.
- Zhou, M., Ko, M., Hoge, A.C.H., Luu, K., Liu, Y., Russell, M.L., Hannon, W.W., Zhang, Z., Carrot-Zhang, J., Beroukham, R., et al. (2022). Patterns of structural variation define prostate cancer across disease states. *JCI Insight*, *7*, 161370.
- Zingg, D., Bhin, J., Yemelyanenko, J., Kas, S.M., Rolfs, F., Lutz, C., Lee, J.K., Klarenbeek, S., Silverman, I.M., Annunziato, S., et al. (2022). Truncated FGFR2 is a clinically actionable oncogene in multiple cancers. *Nature*, *608*, 609-617.
- Murugaesu, N., Sosinsky, A., Ambrose, J., Cross, W., Turnbull, C., Henderson, S., Jones, J., Hamblin, A., Arumugam, P., Chan, G., et al. (2022). Insights for precision healthcare from the 100,000 Genomes Cancer Programme. *Research Square*.