Title

Diversity upon diversity: linking DNA double-strand break repair to blood cancer health disparities

Journal

Trends in Cancer, 8(4)

ISSN

2405-8033

Authors

Sterrenberg, Jason N
Folkerts, Melissa L
Rangel, Valeria
et al.

Publication Date

2022-04-01

DOI

10.1016/j.trecan.2022.01.003

Peer reviewed

HHS Public Access

Author manuscript

Trends Cancer. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Trends Cancer. 2022 April ; 8(4): 328–343. doi:10.1016/j.trecan.2022.01.003.

Diversity upon diversity: linking DNA double-strand break repair to blood cancer health disparities

Jason N. Sterrenberg[1], **Melissa L. Folkerts**[2], **Valeria Rangel**[2], **Sarah Eugenie Lee**[3], **Nicholas R. Pannunzio**[1,2,4,*]

[1]Department of Medicine, Division of Hematology/Oncology, University of California at Irvine, Irvine, CA 92697, USA

[2]Department of Biological Chemistry, University of California at Irvine, Irvine, CA 92697, USA

[3]School of Biological Sciences, University of California at Irvine, Irvine, CA 92697, USA

[4]Chao Family Comprehensive Cancer Center, University of California at Irvine, Irvine, CA 92697, USA

Abstract

Chromosomal translocations arising from aberrant repair of multiple DNA double-strand breaks (DSBs) are a defining characteristic of many cancers. DSBs are an essential part of physiological processes in antibody-producing B cells. The B cell environment is poised to generate genome instability leading to translocations relevant to the pathology of blood cancers. These are a diverse set of cancers, but limited data from under-represented groups have pointed to health disparities associated with each. We focus on the DSBs that occur in developing B cells and propose the most likely mechanism behind the formation of translocations. We also highlight specific cancers in which these rearrangements occur and address the growing concern of health disparities associated with them.

Where do DNA DSBs come from, and how do we fix them?

Maintaining genome integrity is required for cell survival and longevity because DNA damage disrupts normal cellular functions and underlies the etiology of several human diseases. DSBs in DNA are an especially dangerous insult because they are lethal if not repaired, but unfaithful repair leads to insertions, deletions, and large-scale genome rearrangements such as **chromosomal translocations** (see Glossary). Even in the case of a reciprocal translocation with no net loss of genetic material, translocations can be highly disruptive to cellular integrity because they lead to altered transcriptional regulation of genes or the formation of oncogenic gene fusions.

Chromosomal translocations are the result of DNA DSBs, which can either be physiological (planned) or pathological (unplanned). Physiological DSBs occur as part of a programmed

*Correspondence: nrpann@hs.uci.edu (N.R. Pannunzio).

Declaration of interests

The authors have no interests to declare.

or enzymatically controlled process and include those induced in B cells during antibody maturation [1], Spo11-induced DSBs are required for recombination and homolog pairing in meiosis [2], and DSBs are initiated by type 2 topoisomerases [3]. Pathological DSBs can arise from endogenous or exogenous sources of damage including ionizing radiation, reactive oxygen species, DNA replication failure, or rogue nuclease activity. Acting alone or together, these can cause the multiple DNA DSBs that result in translocation formation.

Irrespective of whether the DNA DSBs are pathological or physiological, cells must repair them quickly and with minimal disruption to genome integrity to ensure survival [4]. Eukaryotic cells have evolved two primary DSB repair processes – **homologous recombination (HR)** and non-homologous end joining (NHEJ) (Box 1) – of which NHEJ is the predominant repair pathway in humans. If the NHEJ pathway is compromised, DSBs can be repaired by two more loosely defined pathways that represent a continuum between NHEJ and HR based upon their requirement for greater end **resection** and sequence homology: **alternative end-joining (Alt-EJ)** and **single-strand annealing (SSA)** [5,6].

B cells undergo two events that induce DNA DSBs by different mechanisms at different developmental stages at the **immunoglobulin heavy chain (IgH)** locus. These physiological processes, variable–(diversity)–joining [V(D)J] recombination (Box 2) occurring in pro- and pre-B cells and **class-switch recombination (CSR)** that occurs in mature B cells, are necessary to generate a diverse antibody population [7–10] (Figure 1). Although B cell cancers are a heterogeneous group, many have recurrent chromosomal translocations that are a direct result of physiological *IgH* DSBs and NHEJ repair. Formation of recurrent chromosomal translocations is often the primary transformative event and is linked to the dynamic environment of developing B cells that actively introduce mutations in genes encoding the antibodies of the adaptive immune system [1]. In addition, some of these translocations have been disproportionately found in under-represented patient populations and correlate with less favorable responses to treatment and a sharply reduced survival rate. We describe here the mechanisms by which aberrant DSB repair results in several translocations that form the basis of oncogenesis in several B cell malignancies, and how these mechanisms could be one factor that drives health disparities in patients.

Identification of human fragile zones (HFZs) associated with translocation in pre-B cells

A major insight into the etiology of chromosomal translocations occurring in B cell cancers came from mapping DNA breakpoints in human patients. Using sequencing data, a database was generated of DNA breakpoints at translocation junctions mapped from over 2000 cancer patients covering nearly all hematopoietic malignancies [1,11–13]. This analysis not only revealed that many B cell malignancies involve a DSB at the *IgH* locus, specifically a DSB between the D_H and J_H regions, but also that the non-*IgH* DNA DSBs are clustered in a non-random pattern. We refer to these non-*IgH* regions of focal DNA DSB formation as HFZs [14]. HFZs are not akin to chromosomal fragile sites (FRAs). FRAs consist of kilo- or megabase-long DNA regions and are prone to DSBs under replication stress [15], whereas HFZs are highly focal, ranging from 20 to 600 bp in length, and are not impacted

by replication stress. We discuss here six of these HFZs (Figure 2), but it is likely that others exist and will be found through further analysis of sequencing data [1].

HFZs are within regions associated with recurrent translocations. For example, patient DNA DSBs have been mapped to a 29 kb region of the *BCL2* locus that includes the 3′ untranslated region (3′-UTR) and downstream non-coding sequence (Figure 2). A DNA DSB anywhere in this 29 kb region can join with a broken and unrepaired V(D)J DSB to generate the t(14;18)(q32;q21) *BCL2–IgH* translocation. However, 50% of all the DSBs in this 29 kb region occur within a 175 bp cluster called the major breakpoint region (MBR), representing a 300-fold enrichment of DSB formation. In general, DSBs are ten-to 1000-fold more likely to occur in an HFZ compared to the surrounding chromosomal regions. Importantly, there is no evidence to suggest that recurrent translocations involving HFZs lead to cancers that are more aggressive or proliferative than those forming due to DSBs outside HFZs, ruling out selection as a reason for the focal nature of HFZ DSBs [11,12,16,17].

Mechanism of HFZ DSB formation

Examination at the nucleotide level of patient DNA breakpoints at translocation junctions revealed that most DSBs occur at CpG sites. CpGs are located throughout the genome, and why CpG sites within HFZs are targeted is an active area of research. The B cell-specific mutator activation-induced cytidine deaminase (AID) – that is highly expressed in mature B cells undergoing CSR and **somatic hypermutation (SHM)** – is also expressed at low levels in pre-B cells concurrent with V(D)J recombination [18–26] (Figure 1). AID targets cytosines within single-stranded DNA (ssDNA) at preferred sequence motifs (WG**C**W > WR**C** > R**C**G; where W = A or T, and R = A or G) [27], and converts cytosine to uracil (C➔U). Thus, AID recognizes CpG sites at HFZs, although with reduced efficiency than at WGCW sites arrayed within *IgH* switch regions [28]. CpG sites are also sites of DNA methylation where AID converts 5-methylcytosine to thymine (^{me}C➔T) [27,29]. How DNA methylation might exacerbate AID-induced DSB has been discussed in more depth elsewhere [14,30].

C➔U deamination by AID leads to DSBs by hijacking the base excision repair (BER) mechanism. The U is recognized by a highly efficient uracil glycosylase and is removed to create an abasic site, which is subsequently cleaved by an apurinic/apyrimidinic (AP) endonuclease [31,32]. This activity occurring on both DNA strands in close proximity results in a DSB. AID activity at HFZs provides an important clue about the mechanism behind HFZ DSBs because AID only targets these motifs in ssDNA. Thus, what makes HFZs prone to DNA DSBs is their inherent nature to form transient regions of ssDNA.

Currently, the cause of this ssDNA formation leading to DSBs is unknown, but several possibilities have been tested. The propensity of HFZs to form transient ssDNA has been confirmed through native chemical probing with sodium bisulfite or ammonium bisulfite [30] and by circular dichroism [33]. One possibility considered was that the formation of R-loops – RNA:DNA hybrids that displace one DNA strand to generate a region of ssDNA – at HFZs leads to AID recruitment, similar to what occurs in mature B cells [34–37].

However, treatment with RNase H, which digests the RNA in an RNA:DNA hybrid to collapse an R-loop, had no effect on ssDNA formation in HFZ sequences [30,38]. The formation of triplex or G-quadruplex structures has also been suggested [39,40], but it has not yet been demonstrated that such structures are able to unwind sufficient duplex DNA to form *in vivo* or that the structures are directly linked to DSB formation at HFZs mapped to patient DNA breakpoints.

Because HFZ DSBs occur in early B cells concurrent with V(D)J recombination, one logical culprit was the recombination-activating gene (RAG) complex itself. Cryptic recombination signal sequences (RSSs) that diverge from the conserved RSSs and can be cleaved by the RAG complex are found throughout the genome, and these types of DSBs have already been implicated in several human T cell malignancies and a small percentage of B cell malignancies [39–42]. However, the HFZs do not break at cryptic RSS sites or CAC motifs [43], making it unlikely that the RAG complex is directly involved in HFZ DSB formation, although an indirect mechanism cannot be dismissed [44]. Most of the translocations involve a joining between an HFZ and a D or J region of the *IgH* locus, demonstrating that the availability of an unrepaired RAG-induced DSB is required for translocation formation. There may even be a more important requirement for V(D)J recombination that would make it a causal factor for HFZ DSBs. NHEJ repair of RAG-induced DSBs requires activation of Artemis endonuclease activity for hairpin opening (Box 2). If NHEJ fails to repair this break, it is likely that Artemis will remain active. Because the DNA substrate for Artemis is ssDNA to double-stranded DNA (dsDNA) boundaries [44,45], which would form at the transient ssDNA structures of HFZs, activated Artemis could potentially cleave HFZs to generate DSBs.

HFZs adopt a DNA structure that transitions between B-form and A-form DNA (B/A intermediate) [30]. It is likely that shifting between these two states is what leads to ssDNA. One model that fits the available data is the formation of slipped-strand DNA [46,47] (Figure 3). In this model, when the DNA strands separate, they reanneal out of register, thereby looping out regions on both the top and bottom DNA strands that can act as an ssDNA substrate for AID [14]. Typically, these structures would quickly reform the more thermodynamically stable DNA duplex, but both the frequency and duration of slipped-stand DNA could contribute to increased AID activity at HFZs. AID, reactive oxygen species, and increased torsional stress have been shown to increase DSBs at HFZs, supporting the theory that the HFZs form ssDNA along with secondary DNA structures, and the formation of these structures makes these sequences prone to damage and DSBs [48].

Bringing several of these aspects together, we can form a theoretical model for how HFZ DSBs, and thus chromosomal translocations, are generated in early B cells (Figure 4,Key figure) – (i) the RAG complex mediates V(D)J recombination at the *IgH* locus; (ii) following RAG cleavage and DNA hairpin formation, autophosphorylation of **DNA-dependent protein kinase catalytic subunit (DNA-PKcs)** results in Artemis phosphorylation and activation of its endonuclease activity to open the hairpin; (iii) failure to repair the V(D)J DSB leads to sustained Artemis activity; (iv) dynamic movement of DNA through B/A intermediate transitions generates ssDNA at HFZs and creates a substrate for AID; (v) AID deamination results in DNA mismatches that prevent the DNA duplex from fully reforming,

leaving bubble structures with ssDNA/dsDNA boundaries; (vi) Artemis cleavage at ssDNA/dsDNA boundaries results in HFZ DSBs with unrepaired V(D)J DSBs; finally, (vii) NHEJ repairs the four DNA DSB ends by pairing each *IgH* DSB to an HFZ DSB end, thus forming the reciprocal chromosomal translocations associated with several B cell malignancies.

HFZ-linked translocations and cancer health disparities

Large increases in genomic data have greatly expanded our understanding of the etiology of many of these cancers. However, there is a serious deficit in the diversity of data from different ethnicities and races that affects our ability to address cancer health disparities. In the field of blood cancers, this issue is compounded by the fact that incidence and survival rates are often not deaggregated between different leukemias and lymphomas, or even between different subtypes of a specific cancer, such as the highly diverse group of acute lymphoblastic leukemias (ALLs). Averaging the few data we do have has the result of masking distinctive disparities within under-represented populations. The mechanism of translocation formation described in the preceding text is complex, and genetic or epigenetic differences within populations can affect translocation formation at any step in the process. Although much work lies ahead, we discuss here the etiology of a group of blood cancers where translocations form in early B cells and where health disparities exist within subtypes of these cancers because of higher incidence and lower survival rates (Table 1).

This is a good time to point out the complexity of the health disparities issue and how it is linked to historical events and nomenclature. The term ‘Hispanic’ is used in the scientific community and by the US Census Bureau to classify a wide swath of people spanning multiple countries and continents including Spain, Mexico, Central America, South America, Cuba, and Puerto Rico. What defines a person as Hispanic, Latino, or Latina is highly debated. The term Latinx, widely gaining use in academic circles to support gender fluidity, has not developed widespread use within the community. Throughout this text, we use the term ‘Hispanic’ to refer to a myriad of different peoples. This is a consequence of the fact that the majority of scientific studies group all these peoples together, thus we have few data on people from specific regions. Furthermore, most people only have the option to self-identify as Hispanic or non-Hispanic white on medical forms. Until we begin to add a diversity of options to consent forms for clinical studies and define people by culture and country of origin, cancer health disparities within these communities will remain pervasive or, worse, fully masked.

Follicular lymphoma (FL)

FL is the second most common adult lymphoma representing ~20% of all non-Hodgkin lymphoma cases [49]. It is slow-growing and typically occurs in older adults. Although currently incurable, the introduction of the CD20 monoclonal antibody rituximab into treatment regiments has increased survival rates. The primary translocation leading to FL, that is present in nearly >85% of cases, involves a chromosomal translocation between a pathological DSB at the *BCL2* locus and an unrepaired physiological DSB at the *IgH* locus, resulting in the t(14;18) product [50]. The *BCL2–IgH* t(14;18) translocation places the *mu* enhancer (E_μ) of *IgH* proximal to *BCL2*, dysregulating transcription of the anti-apoptotic

gene. Pairing of the *BCL2 DSB* to a D_H or J_H region of *IgH* indicates that this translocation occurs in precursor B cells. However, although almost all FL cases have the *BCL2–IgH* translocation, this rearrangement alone is not sufficient for disease presentation, and the B cell continues to mature, acquiring additional mutations before presenting as a mature B cell lymphoma.

BCL2 DSBs that lead to the translocation with *IgH* occur in a 29 kb region consisting of the 3′-UTR of *BCL2* and downstream sequence (Figure 2). Within this region are three DSB clusters that fit our definition of HFZs. One half of DSBs mapped in human patients occur within the 175 bp MBR described previously, 13% occur in the intermediate cluster region (ICR), and 5% occur within the minor cluster region (MCR) [11]. It has been estimated that 50–70% of 'healthy' individuals carry circulating B cells with the *BCL2–IgH* translocation, meaning that they have B cells in a pre-FL state and are at risk of acquiring additional mutations that can lead to full FL [51]. Although most of these healthy individuals do not progress to full FL, it is possible that the incurability and the high risk of relapse in those with FL is related to having a reservoir of pre-FL cells already present in their blood.

FL is more prevalent in developed countries and is most common in non-Hispanic whites. Although the rate of FL in Asians and Pacific Islanders is lower overall, the incidence is higher in those born in Western countries versus those born in Asia, underscoring potential environmental and lifestyle risk factors [49]. Furthermore, agricultural workers are more likely to carry the *BCL2–IgH* translocations due to exposure to pesticides [51]. Considering the number of Hispanic migrant farm workers employed in the USA, it is likely that increased cases of FL will begin to appear in these populations as they age. This highlights that there are several non-genetic factors associated with cancer health disparities (Box 3).

Diffuse large B cell lymphoma (DLBCL)

DLBCL is the most common non-Hodgkin lymphoma diagnosed in the USA and has many subtypes. DLBCL not otherwise specified (DLBCL-NOS) is the catch-all term for the DLBCLs that do not easily fall into one of the known subtypes. DLBCL-NOS can be further subdivided into those that affect germinal center B cells (GCB subtype) or activated B cells (ABC subtype). Although the *BCL2–IgH* t(14;18) translocation is a hallmark of FL, it occurs in ~20% of all DLBCLs, but 40% of DLBCL-NOS GCBs [52]. Furthermore, the presence of *BCL2–IgH* in this latter subtype is indicative of poor response to the standard DLBCL treatment of rituximab, cyclophosphamide, hydroxydaunomycin (doxorubicin), Oncovin (vincristine sulfate), and prednisone (R-CHOP), indicating that early detection of this rearrangement can lead to a better standard of care for some DLBCL patients [53].

Among the DLBLC subtypes is one linked specifically to Epstein–Barr virus (EBV). If pre-existing immunodeficiency is eliminated, there seems to be a higher prevalence of this subtype among Asian and Hispanic populations [52]. Given the higher prevalence of HFZ-linked translocations linked to the Hispanic population, it would be interesting to uncover whether an underlying immune system defect is the cause of this and is exacerbated by viral infection.

Mantle cell lymphoma (MCL)

MCL is a rare but aggressive form of non-Hodgkin lymphoma. Older white males are the highest risk group for MCL by almost 10% over other ethnic groups, and MCL often has a poor prognosis and a high rate of relapse [54,55]. Like FL, the majority of MCLs begin with formation of a translocation during the pro-B cell stage involving an unrepaired RAG-induced break at the *IgH* locus. In MCL, over 95% of cases involve pathological DSBs on chromosome 11 adjacent to the *CCND1* gene encoding cyclin D1 [52]. Formation of the *CCND1–IgH* t(11;14) translocation results in juxtaposition of *IgH* E_{μ} upstream of *CCND1*, causing increased and constitutive expression [50], and thus disrupts the G1/S cell-cycle transition and enables hyperproliferation. The *CCND1–IgH* translocation is the first (but not the only) step in carcinogenesis, and additional mutations occur as the B cell develops until MCL presents in mature B cells.

Pathological DSBs at the *CCND1* locus that result in translocations with *IgH* have been mapped in human patients to a 340 kb region. Of these DSBs, 30% occur within the 150 bp major translocation cluster (MTC) HFZ ~100 kb upstream of *CCND1* [11] (Figure 2). Like other HFZs, the DSBs mapped to MTC are predominately at CpG sites. Not only does this provide a target for AID, but studies have also shown that hypomethylation within the MTC HFZ may render the locus more accessible to AID [56].

Multiple myeloma (MM)

MM is a malignancy characterized by uncontrolled growth of plasma cells (terminally differentiated B cells). MM is the second most common hematopoietic malignancy after non-Hodgkin lymphoma and accounts for ~10% of cases. Although rare, the median survival range for MM is only 5.5 years [52]. Sixteen percent of MM cases involve the *CCND1–IgH* translocation described previously, indicating that an early event in developing B cells is the primary cause of MM in mature B cells in these cases. An additional 24% of MM cases involve a translocation between *IgH* and another gene (*MAF*, *MMSET*, *CCND3*, *MAFB*, *MAFA*, or *CCND2*).

The incidence rate of MM in Blacks of African descent is twice that of whites, indicating a clear cancer health disparity. Blacks are also much more likely to have MM with translocations involving *IgH*. Furthermore, epidemiological data clearly show that socioeconomic factors and low recruitment to clinical trials worsens the prognosis of MM in this population [57,58] (Box 3).

B cell ALL

B cell ALL is a heterogeneous cancer affecting immature B cells that leads to overproduction of these cells in the bone marrow and increased levels in the blood. ALL occurs in both children and adults, and the prognosis is typically worse in older patients. Different chromosomal abnormalities are associated with the different ALL subtypes. These not only include chromosomal translocations but also changes in chromosome copy number (hyper- and hypodiploidy). Identifying the chromosomal abnormality is important in classifying the ALL subtype because this dictates the severity of the disease and the best treatment course.

Hispanics, particularly those with Native American ancestry, have a significantly greater incidence rate of ALL (10–30%), and this contributes to a major cancer health disparity [59–64]. Determining what genetic markers are the cause of this disparity is an active area of research [65]. Two of the translocations associated with two different ALL subtypes (*CRLF2–IgH* and *TCF3–PBX1*) involve HFZs and show increased occurrence in Hispanic patients, indicating that factors in these heterogeneous populations can influence the HFZ DSB mechanism and increase susceptibility to specific chromosomal translocations [66–68].

An HFZ has been identified on chromosome 19 at the transcription factor 3 (*TCF3*) locus, also called *E2A*, leading to a *TCF3–PBX1* t(1;19) translocation and a fusion gene product that disrupts transcriptional activation. The *TCF3* HFZ is typically mapped to intron 16 of the gene, although this can vary depending on the splice variant. Patient DSBs have been mapped to a 3.2 kb region within this intron, but a striking 75% of DSBs occur at the 23 bp HFZ [30], making this the most focal of the HFZs (Figure 2). Unlike other HFZs, the second DSB required for the translocation is not a physiological DSB at the *IgH* locus but a second pathological DSB at the *PBX1* locus. The *PBX1* DSBs do not occur at an HFZ but have been mapped to a 100 kb region of *PBX1* intron 1. Although not involving *IgH*, there is evidence that this translocation forms concurrently with V(D)J recombination because the repair junctions show evidence of non-templated nucleotides inserted by the X-family DNA polymerase terminal deoxynucleotidyl transferase (TdT) (Box 2) which is only active in early B cells [13,69].

In one genome-wide association study (GWAS) involving 940 Hispanic childhood ALL patients and 681 ancestry-matched ALL-free controls, a mutant allele of the *ERG* gene, encoding an erythroblast transformation-specific transcription factor, was implicated as a significant ALL risk factor [68]. This genetic marker was under-represented in the most common ALL subtype carrying a *ETV6–RUNX1* translocation, but was significantly enriched in the rarer *TCF3–PBX1*-bearing ALL subtype. This correlation also held for a smaller cohort of Guatemalan children with ALL where the presence of the risk allele was highly correlated with Native American ancestry.

Philadelphia chromosome-like B cell ALL (Ph-like ALL) has recently emerged as a distinct ALL subtype [70–73] characterized by a poor response to therapy, a high risk of relapse, and a peak onset in adolescents and young adults, where the 5-year overall survival rate in the latter was only 25% [74]. The name of this subtype derives from a gene expression profile similar to that of Philadelphia chromosome-positive ALL, but it lacks the eponymous t(9;22) translocation carrying the *BCR–ABL* fusion [73]. Although Ph-like ALL is associated with a heterogeneous mutation spectrum [75,76], nearly 65% of cases carry a rearrangement in the cytokine receptor-like factor 2 (*CRLF2*) gene leading to upregulation of the JAK2/STAT3 pathway, often leading to additional *JAK2* mutations that sustain this upregulation [77].

CRLF2 rearrangements typically involve either an intrachromosomal deletion on the X or Y chromosome, placing *CRLF2* under the control of the *P2RY8* promoter, or a chromosomal translocation with chromosome 14 to create the *CRLF2–IgH* product. This latter rearrangement is caused by a DNA DSB that forms at an HFZ upstream of the *CRLF2*

gene, thereby allowing joining to an unrepaired V(D)J DSB at *IgH* to form the t(X/Y;14) *CRLF2–IgH* translocation. DSBs at *CRLF2* have been mapped to a 25 kb region, but 32% of DSBs occur in a 311 bp cluster where DSBs are associated with CpG sites (Figure 2). Among *CRLF2* rearrangements, the *CRLF2–IgH* translocation is the most common in the USA and is often associated with a poor prognosis [77].

Ph-like ALL also presents clear Hispanic cancer health disparities. A recent study involving 155 ALL patients showed that, among this cohort, 68% of Ph-like ALL cases were Hispanics versus 23% whites, 5% Asians, and 4% African Americans [67]. Furthermore, Hispanic patients respond less favorably to current treatment regimens than non-Hispanics [77]. *CRLF2–IgH* translocations are two- to threefold more likely to occur within Hispanic and Native American populations and correlate with less favorable responses to treatment and a sharply reduced survival rate [69–71].

Enhanced understanding of the etiology of HFZ-linked translocations and the mechanisms resulting in HFZ DSBs will be crucial in addressing these major cancer health disparities. A significant roadblock is the lack of studies that specifically focus on under-represented ethnicities or that deaggregate Hispanic and Latino populations in large datasets. Similar issues result from placing all Blacks or Asian and Pacific Islanders in broad categories, or by not distinguishing white Europeans from Middle Easterners. As more studies begin to make these distinctions, it is very likely that more health disparities that are currently hidden in aggregated data will begin to present themselves.

Concluding remarks

Pathological DSBs occurring at HFZs significantly contribute to the etiology of B cell malignancies. Translocations that arise from HFZ DSBs occur in several blood cancers that affect both immature and mature B cells. Their common link is that these translocations form in the pre-B cell environment concurrent with V(D)J recombination (Figure 4). In most cases, this primary translocation is not sufficient for carcinogenesis, and additional mutations are necessary to fully transform a normal cell to a cancerous one. Nevertheless, the translocations described here are typically the primary and necessary rearrangement required for many B cell malignancies. Thus, further understanding of the genetic and environmental circumstances that lead to these translocations will greatly enhance our ability to predict and diagnose these cancers. It also raises the possibility that these early translocations involving HFZs can be used to enhance early detection of these cancers, thereby leading to decreased mortality (see Outstanding questions). Furthermore, several B cell malignancies disproportionately affect the Hispanic and Black populations, and further research into the mechanism of HFZ DSBs will allow us to address these major cancer health disparities.

Key figure

Model for formation of oncogenic translocations originating in pro-B cells

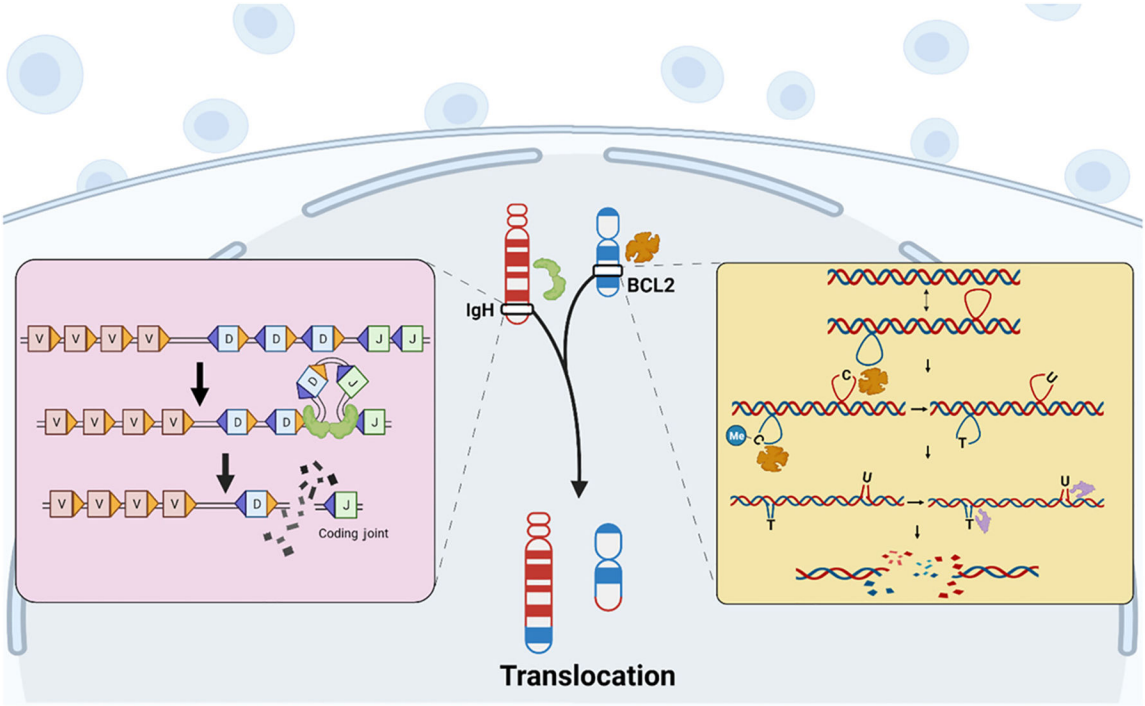

Figure 4.

Failure to repair physiological variable–(diversity)–joining [V(D)J] double-strand breaks (DSBs) induced by the recombination-activating gene (RAG) complex and the creation of pathological human fragile zone (HFZ) DSBs mediated by the formation of transient single-stranded (ss)DNA and activation-induced cytidine deaminase (AID) activity can lead to oncogenic translocations when the two occur simultaneously because non-homologous end joining (NHEJ) links incorrect DNA ends together. The *BCL2* locus carrying the major breakpoint region (MBR), intermediate cluster region (ICR), and minor cluster region (MCR) HFZs is depicted, but the same model applies to *CRLF2*, the BCL1 major translocation cluster (MTC) at *CCND1*, and the *TCF3* HFZs. In the case of *TCF3*, the HFZ DSB is AID-mediated, whereas the other DSB involved at the *PBX1* locus is likely random.

Acknowledgments

This work was supported by funds from the American Cancer Society (IRG-16-187-13), the University of California Cancer Research Coordinating Committee (C22CR), and the Chao Family Comprehensive Cancer Center. Figures 1–4, Box 1 Figure I, and Box 2 Figure I were created with BioRender.com.

Glossary

Alternative end joining (Alt-EJ)

an error-prone double-strand break (DSB) repair pathway used when non-homologous end joining (NHEJ) is inactive. Alt-EJ uses 2–20 bp of microhomology following minimal end resection

Chromosomal translocations
genomic rearrangements that occur when two DNA DSBs occur simultaneously on non-homologous chromosomes. Instead of repair resulting in joining the two DNA ends from the same chromosome to regenerate the original parent molecule, aberrant repair joins ends from different chromosomes resulting in an exchange of chromosome arms and the formation of a new derivative product. Translocations can either be reciprocal, where both new products are retained, or non-reciprocal, where only one derivative product is retained

Class-switch recombination (CSR)
a process initiated in germinal center B cells by high AID expression. DSBs generated lead to recombination events that change the default μ and δ constant regions at the *IgH* locus, that encode IgM and IgD antibodies, to γ, α, αnd ε constant regions that encode IgG, IgA, and IgE antibodies that have different functions and the ability to travel to different areas of the body

DNA-dependent protein kinase catalytic subunit (DNA-PKcs)
a protein serine/threonine kinase with homology to the phosphoinositide 3-kinase family of enzymes that are involved in DNA damage sensing. DNA-PKcs has high affinity for Ku-bound DNA ends, and binding of DNA-PKcs to Ku creates the full DNA–PK complex. Autophosphorylation of DNA-PKcs leads to the activation of other NHEJ factors, notably Artemis

Homologous recombination (HR)
a DNA DSB repair pathway that uses >100 bp of sequence homology for high-fidelity repair. HR requires extensive end resection and is typically active in the S/G2 phases of the cell cycle when a sister chromatid is available as a repair template

Immunoglobulin heavy chain (IgH)
antibodies have two heavy chains and two light chains. The larger heavy chains carry the variable and constant domains, while the smaller light chains consist of variable regions that have undergone variable (V)/junction (J) rearrangements at the *Igκ* or *Igλ* loci [light chains have no diversity (D) cassettes]. The *IgH* locus undergoes V(D)J recombination at the 5′ portion of the gene in early B cells and CSR in the 3′ portion of the gene in mature B cells that generate physiological DSBs

Insertions and deletions (indels)
mutations stemming from nucleotide insertions or deletions. These information scars provide evidence of previous DSBs at a site

Resection
enzymatic processing of DNA ends by nucleases and helicases to produce 3′ single-stranded DNA tails for homology use during repair

Single-strand annealing (SSA)
an error-prone DSB repair pathway that usually results in deletions and typically uses >20 bp of homology following moderate end resection

Somatic hypermutation (SHM)

a process that further mutates the variable regions of antibody-encoding genes in mature B cells to increase antigen affinity. This process is also initiated by AID, but instead of leading to DSBs, the process results in single-nucleotide mutations

References

1. Liu D and Lieber MR (2021) The mechanisms of human lymphoid chromosomal translocations and their medical relevance. Crit. Rev. Biochem. Mol. Biol Published online December 7, 2021. 10.1080/10409238.2021.2004576
2. Keeney S (2001) Mechanism and control of meiotic recombination initiation. Curr. Top. Dev. Biol 52, 1–53 [PubMed: 11529427]
3. Wang JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. Nat. Rev. Mol. Cell Biol 3, 430–440 [PubMed: 12042765]
4. Berger AH et al. (2011) A continuum model for tumour suppression. Nature 476, 163–169 [PubMed: 21833082]
5. Chang HHY et al. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat. Rev. Mol. Cell Biol 18, 495–506 [PubMed: 28512351]
6. Bhargava R et al. (2016) Regulation of single-strand annealing and its role in genome maintenance. Trends Genet. 32, 566–575 [PubMed: 27450436]
7. Hesslein DGT and Schatz DG (2001) Factors and forces controlling V(D)J recombination. Adv. Immunol 78, 169–232 [PubMed: 11432204]
8. Schatz DG and Swanson PC (2011) V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet 45, 167–202 [PubMed: 21854230]
9. Han JH et al. (2007) Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. Immunity 27, 64–75 [PubMed: 17658280]
10. Yu K and Lieber MR (2019) Current insights into the mechanism of mammalian immunoglobulin class switch recombination. Crit. Rev. Biochem. Mol. Biol 54, 333–351 [PubMed: 31509023]
11. Tsai AG et al. (2008) Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell 135, 1130–1142 [PubMed: 19070581]
12. Greisman HA et al. (2012) IgH partner breakpoint sequences provide evidence that AID initiates t(11;14) and t(8;14) chromosomal breaks in mantle cell and Burkitt lymphomas. Blood 120, 2864–2867 [PubMed: 22915650]
13. Wiemels JL et al. (2002) Site-specific translocation and evidence of postnatal origin of the t(1;19) E2A–PBX1 fusion in childhood acute lymphoblastic leukemia. Proc. Natl. Acad. Sci. U. S. A 99, 15101–15106 [PubMed: 12415113]
14. Pannunzio NR and Lieber MR (2018) Concept of DNA lesion longevity and chromosomal translocations. Trends Biochem. Sci 43, 490–498 [PubMed: 29735400]
15. Debatisse M et al. (2012) Common fragile sites: mechanisms of instability revisited. Trends Genet. 28, 22–32 [PubMed: 22094264]
16. Buchonnet G et al. (2002) Distribution of BCL2 breakpoints in follicular lymphoma and correlation with clinical features: specific subtypes or same disease? Leukemia 16, 1852–1856 [PubMed: 12200703]
17. Lieber MR (2016) Mechanisms of human lymphoid chromosomal translocations. Nat. Rev. Cancer 16, 387–398 [PubMed: 27220482]
18. Swaminathan S et al. (2015) Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. Nat. Immunol 16, 766–774 [PubMed: 25985233]
19. Mao C et al. (2004) T cell-independent somatic hypermutation in murine B cells with an immature phenotype. Immunity 20, 133–144 [PubMed: 14975236]
20. Cantaert T et al. (2015) Activation-induced cytidine deaminase expression in human B cell precursors is essential for central B cell tolerance. Immunity 43, 884–895 [PubMed: 26546282]

21. Ueda Y et al. (2007) T-independent activation-induced cytidine deaminase expression, class-switch recombination, and antibody production by immature/transitional 1 B cells. J. Immunol 178, 3593–3601 [PubMed: 17339456]
22. Kuraoka M et al. (2011) Activation-induced cytidine deaminase mediates central tolerance in B cells. Proc. Natl. Acad. Sci. U. S. A 108, 11560–11565 [PubMed: 21700885]
23. Kuraoka M et al. (2009) Activation-induced cytidine deaminase expression and activity in the absence of germinal centers: insights into hyper-IgM syndrome. J. Immunol 183, 3237–3248 [PubMed: 19667096]
24. Kumar S et al. (2013) Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. Genes Dev. 27, 2439–2444 [PubMed: 24240234]
25. Umiker BR et al. (2014) Production of IgG autoantibody requires expression of activation-induced deaminase in early-developing B cells in a mouse model of SLE. Eur. J. Immunol 44, 3093–3108 [PubMed: 25044405]
26. Kelsoe G (2014) Curiouser and curiouser: the role(s) of AID expression in self-tolerance. Eur. J. Immunol 44, 2876–2879 [PubMed: 25308427]
27. Pham P et al. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103–107 [PubMed: 12819663]
28. Zhang ZZ et al. (2014) The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. Cell Rep. 8, 557–569 [PubMed: 25017067]
29. Schmutte C et al. (1995) Base excision repair of U:G mismatches at a mutational hotspot in the p53 gene is more efficient than base excision repair of T:G mismatches in extracts of human colon tumors. Cancer Res. 55, 3742–3746 [PubMed: 7641186]
30. Liu D et al. (2021) Mechanistic basis for chromosomal translocations at the E2A gene and its broader relevance to human B cell malignancies. Cell Rep. 36, 109387 [PubMed: 34260910]
31. Rogier M et al. (2021) Fam72a enforces error-prone DNA repair during antibody diversification. Nature 600, 329–333 [PubMed: 34819671]
32. Feng Y et al. (2021) FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. Nature 600, 324–328 [PubMed: 34819670]
33. Raghavan SC et al. (2004) A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature 428, 88–93 [PubMed: 14999286]
34. Duquette ML et al. (2005) AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene 24, 5791–5798 [PubMed: 15940261]
35. Yu K and Lieber MR (2003) Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. DNA Repair 2, 1163–1174 [PubMed: 14599739]
36. Yu K et al. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol 4, 442–451 [PubMed: 12679812]
37. Yang Y et al. (2014) Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. Mol. Cell 53, 484–497 [PubMed: 24507716]
38. Tsai AG et al. (2009) Conformational variants of duplex DNA correlated with cytosine-rich chromosomal fragile sites. J. Biol. Chem 284, 7157–7164 [PubMed: 19106104]
39. Nambiar M et al. (2011) Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. Nucleic Acids Res. 39, 936–948 [PubMed: 20880994]
40. Williams JD et al. (2015) Sites of instability in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures in vitro. Front. Genet 6, 177 [PubMed: 26029241]
41. Bassing CH et al. (2002) The mechanism and regulation of chromosomal V(D)J recombination. Cell 109, 45–55
42. Lee J and Desiderio S (1999) Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair. Immunity 11, 771–781 [PubMed: 10626899]
43. Hu J et al. (2015) Chromosomal loop fomains direct the recombination of antigen receptor genes. Cell 163, 947–959 [PubMed: 26593423]

44. Cui X et al. (2013) Both CpG methylation and activation-induced deaminase are required for the fragility of the human BCL-2 major breakpoint region: implications for the timing of the breaks in the t(14;18) translocation. Mol. Cell. Biol 33, 947–957 [PubMed: 23263985]
45. Chang HHY and Lieber MR (2016) Structure-specific nuclease activities of Artemis and the Artemis:DNA-PKcs complex. Nucleic Acids Res. 44, 4991–4997 [PubMed: 27198222]
46. Sinden RR (1994) DNA Structure and Function. Academic Press
47. Pearson CE et al. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in $(CTG)_n \cdot (CAG)_n$ repeats from the myotonic dystrophy locus. Nucleic Acids Res. 26, 816–823 [PubMed: 9443975]
48. Pannunzio NR and Lieber MR (2017) AID and reactive oxygen species can induce DNA breaks within human chromosomal translocation fragile zones. Mol. Cell 68, 901–912 [PubMed: 29220655]
49. Carbone A et al. (2019) Follicular lymphoma. Nat. Rev. Dis. Primers 5, 83 [PubMed: 31831752]
50. Kumar E et al. (2021) Pathogenesis of follicular lymphoma: genetics to the microenvironment to clinical translation. Br. J. Haematol 194, 810–821 [PubMed: 33694181]
51. Brisou G et al. (2021) The premalignant ancestor cell of $t(14;18)^+$ lymphoma. Hemasphere 5, e579 [PubMed: 34095762]
52. Swerdlow SH et al. (2017) WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (4th edn), World Health Organization
53. Visco C et al. (2013) Patients with diffuse large B-cell lymphoma of germinal center origin with BCL2 translocations have poor outcome, irrespective of MYC status: a report from an International DLBCL Rituximab-CHOP Consortium Program Study. Haematologica 98, 255–263 [PubMed: 22929980]
54. Dreyling M et al. (2018) Treatment for patients with relapsed/refractory mantle cell lymphoma: European-based recommendations. Leuk. Lymphoma 59, 1814–1828 [PubMed: 29172868]
55. Wang Y and Ma S (2014) Racial differences in mantle cell lymphoma in the United States. BMC Cancer 14, 764 [PubMed: 25315847]
56. Lu Z et al. (2015) Human lymphoid translocation fragile zones are hypomethylated and have accessible chromatin. Mol. Cell. Biol 35, 1209–1222 [PubMed: 25624348]
57. Castañeda-Avila MA et al. (2021) Differences in survival among multiple myeloma patients in the United States SEER population by neighborhood socioeconomic status and race/ethnicity. Cancer Causes Control 32, 1021–1028 [PubMed: 34089470]
58. Gormley N et al. (2021) Recommendations on eliminating racial disparities in multiple myeloma therapies: a step toward achieving equity in healthcare. Blood Cancer Discov. 2, 119–124 [PubMed: 34179821]
59. Yamamoto JF and Goodman MT (2008) Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997–2002. Cancer Causes Control 19, 379–390 [PubMed: 18064533]
60. Walsh KM et al. (2013) Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. Leukemia 27, 2416–2419 [PubMed: 23615557]
61. Fidler MM et al. (2017) Cancer incidence and mortality among young adults aged 20–39 years worldwide in 2012: a population-based study. Lancet Oncol. 18, 1579–1589 [PubMed: 29111259]
62. Qian M et al. (2019) Novel susceptibility variants at the ERG locus for childhood acute lymphoblastic leukemia in Hispanics. Blood 133, 724–729 [PubMed: 30510082]
63. Paulozzi LJ et al. (2020) A disparity beneath a paradox: cancer mortality among young Hispanic Americans in the US–Mexico border region. J. Racial Ethn. Health Disparities 8, 1556–1562 [PubMed: 33156479]
64. Bencomo-Alvarez AE et al. (2021) Ethnic and border differences on blood cancer presentation and outcomes: a Texas population-based study. Cancer 127, 1068–1079 [PubMed: 33616915]
65. Wiemels JL et al. (2018) GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes. Nat. Commun 9, 286 [PubMed: 29348612]
66. Konoplev S et al. (2017) CRLF2-positive B-cell acute lymphoblastic leukemia in adult patients: a single-institution experience. Am. J. Clin. Pathol 147, 357–363

67. Jain N et al. (2017) Ph-like acute lymphoblastic leukemia: a high-risk subtype in adults. Blood 129, 572–581 [PubMed: 27919910]
68. Qian M et al. (2019) Genome-wide association study of susceptibility loci for T-cell acute lymphoblastic leukemia in children. J. Natl. Cancer Inst 111, 1350–1357 [PubMed: 30938820]
69. Hein D et al. (2019) The preleukemic TCF3–PBX1 gene fusion can be generated in utero and is present in ~0.6% of healthy newborns. Blood 134, 1355–1358 [PubMed: 31434706]
70. Den Boer ML et al. (2009) A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. Lancet Oncol. 10, 125–134 [PubMed: 19138562]
71. Mullighan CG et al. (2009) Rearrangement of CRLF2 in B-progenitor– and Down syndrome–associated acute lymphoblastic leukemia. Nat. Genet 41, 1243–1246 [PubMed: 19838194]
72. Arber DA et al. (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391–2405 [PubMed: 27069254]
73. Shiraz P et al. (2020) The current genomic and molecular landscape of Philadelphia-like acute lymphoblastic leukemia. Int. J. Mol. Sci 21, 2193
74. Roberts KG et al. (2014) Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. N. Engl. J. Med 371, 1005–1015 [PubMed: 25207766]
75. Tasian SK et al. (2017) Philadelphia chromosome-like acute lymphoblastic leukemia. Blood 130, 2064–2072 [PubMed: 28972016]
76. Roberts KG (2018) Why and how to treat Ph-like ALL? Best Pract. Res. Clin. Haematol 31, 351–356 [PubMed: 30466746]
77. Harvey RC et al. (2010) Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. Blood 115, 5312–5321 [PubMed: 20139093]
78. Bhargava R et al. (2018) C-NHEJ without indels is robust and requires synergistic function of distinct XLF domains. Nat. Commun 9, 2484 [PubMed: 29950655]
79. Bétermier M et al. (2014) Is non-homologous end-joining really an inherently error-prone process? PLoS Genet. 10, e1004086 [PubMed: 24453986]
80. Lieber MR (2008) The mechanism of human nonhomologous DNA end joining. J. Biol. Chem 283, 1–5 [PubMed: 17999957]
81. Jackson SP (2002) Sensing and repairing DNA double-strand breaks. Carcinogenesis 23, 687–696 [PubMed: 12016139]
82. Mimitou EP and Symington LS (2010) Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. EMBO J. 29, 3358–3369 [PubMed: 20729809]
83. West RB et al. (1998) Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. Mol. Cell. Biol 18, 5908–5920 [PubMed: 9742108]
84. Meek K et al. (2008) DNA-PK: the means to justify the ends? Adv. Immunol 99, 33–58 [PubMed: 19117531]
85. Bebenek K et al. (2014) Structure–function studies of DNA polymerase λ. Biochemistry 53, 2781–2792 [PubMed: 24716527]
86. Moon AF et al. (2014) Sustained active site rigidity during synthesis by human DNA polymerase μ. Nat. Struct. Mol. Biol 21, 253–260 [PubMed: 24487959]
87. Pryor JM et al. (2015) Essential role for polymerase specialization in cellular nonhomologous end joining. Proc. Natl. Acad. Sci. U. S. A 112, E4537–E4545 [PubMed: 26240371]
88. Gilfillan S et al. (1993) Mice lacking TdT: mature animals with an immature lymphocyte repertoire. Science 261, 1175–1178 [PubMed: 8356452]
89. Ma Y et al. (2004) A biochemically defined system for mammalian nonhomologous DNA end joining. Mol. Cell 16, 701–713 [PubMed: 15574326]
90. Bertocci B et al. (2006) Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. Immunity 25, 31–41 [PubMed: 16860755]

91. McElhinny SAN et al. (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. Mol. Cell 19, 357–366 [PubMed: 16061182]
92. Gu J et al. (2007) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. EMBO J. 26, 1010–1023 [PubMed: 17290226]
93. Dominguez O (2000) DNA polymerase mu (Pol μ), homologous to TdT, could act as a DNA mutator in eukaryotic cells. EMBO J. 19, 1731–1742 [PubMed: 10747040]
94. Adachi N et al. (2003) Hypersensitivity of nonhomologous DNA end-joining mutants to VP-16 and ICRF-193: implications for the repair of topoisomerase II-mediated DNA damage. J. Biol. Chem 278, 35897–35902 [PubMed: 12842886]
95. Kurosawa A et al. (2008) The requirement of Artemis in double-strand break repair depends on the type of DNA damage. DNA Cell Biol. 27, 55–61 [PubMed: 17941805]
96. Riballo E et al. (2004) A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to γ-H2AX foci. Mol. Cell 16, 715–724 [PubMed: 15574327]
97. Pawelczak KS and Turchi JJ (2010) Purification and characterization of exonuclease-free Artemis: implications for DNA-PK-dependent processing of DNA termini in NHEJ-catalyzed DSB repair. DNA Repair 9, 670–677 [PubMed: 20347402]
98. Poinsignon C et al. (2004) The metallo-β-lactamase/β-CASP domain of Artemis constitutes the catalytic core for V(D)J recombination. J. Exp. Med 199, 315–321 [PubMed: 14744996]
99. Ma Y et al. (2005) The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. DNA Repair 4, 845–851 [PubMed: 15936993]
100. Ma Y et al. (2002) Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108, 781–794 [PubMed: 11955432]
101. Goodarzi AA et al. (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. EMBO J. 25, 3880–3889 [PubMed: 16874298]
102. Ma Y et al. (2005) The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human artemis. J. Biol. Chem 280, 33839–33846 [PubMed: 16093244]
103. Li S et al. (2014) Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5′-exonuclease activity. J. Biol. Chem 289, 7825–7834 [PubMed: 24500713]
104. Zhao B et al. (2020) NAD^+ is not utilized as a co-factor for DNA ligation by human DNA ligase IV. Nucleic Acids Res. 48, 12746–12750 [PubMed: 33264406]
105. Grawunder U et al. (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388, 492–495 [PubMed: 9242410]
106. Zhao B et al. (2019) The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. Nat. Commun 10, 3588 [PubMed: 31399561]
107. Dai Y et al. (2003) Nonhomologous end joining and V(D)J recombination require an additional factor. Proc. Natl. Acad. Sci. U. S. A 100, 2462–2467 [PubMed: 12604777]
108. Ahnesorg P et al. (2006) XLF interacts with the XRCC4–DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell 124, 301–313 [PubMed: 16439205]
109. Buck D et al. (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell 124, 287–299 [PubMed: 16439204]
110. Ochi T et al. (2015) PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. Science 347, 185–188 [PubMed: 25574025]
111. Hesse JE et al. (1989) V(D)J recombination: a functional definition of the joining signals. Genes Dev. 3, 1053–1061 [PubMed: 2777075]
112. Ramsden DA et al. (1996) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. EMBO J. 15, 3197–3206 [PubMed: 8670820]
113. Lewis SM (1994) The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv. Immunol 56, 27–150 [PubMed: 8073949]
114. Feeney AJ et al. (2004) Many levels of control of V gene rearrangement frequency. Immunol. Rev 200, 44–56 [PubMed: 15242395]

115. Bednarski JJ and Sleckman BP (2019) At the intersection of DNA damage and immune responses. Nat. Rev. Immunol 19, 231–242 [PubMed: 30778174]

116. Zhang Y et al. (2019) The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. Nature 573, 600–604 [PubMed: 31511698]

117. Zhao B et al. (2020) The molecular basis and disease relevance of non-homologous DNA end joining. Nat. Rev. Mol. Cell Biol 21, 765–781 [PubMed: 33077885]

118. Jones JF et al. (1991) Severe combined immunodeficiency among the Navajo. I. Characterization of phenotypes, epidemiology, and population genetics. Hum. Biol 63, 669–682 [PubMed: 1916741]

119. Kwong PC et al. (1999) Oral and genital ulceration: a unique presentation of immunodeficiency in Athabascan-speaking American Indian children with severe combined immunodeficiency. Arch. Dermatol 135, 927–931 [PubMed: 10456341]

120. Buckley RH et al. (1997) Human severe combined immunodeficiency: genetic, phenotypic, and functional diversity in one hundred eight infants. J. Pediatr 130, 378–387 [PubMed: 9063412]

121. Agrawal A and Schatz DG (1997) RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell 89, 43–53 [PubMed: 9094713]

122. Lee GS et al. (2004) RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. Cell 117, 171–184 [PubMed: 15084256]

Box 1.

NHEJ repairs most DNA DSBs in humans

NHEJ repairs DSBs from multiple sources throughout the cell cycle, and the DNA end configurations dictate which NHEJ factors are required [5]. NHEJ factors can also engage and re-engage each broken DNA end multiple times until ligation by DNA ligase IV (Lig4) at juxtaposed DNA ends (Figure I). Although the iterative nature of NHEJ ensures repair of most DNA DSBs, it also potentiates mutations at the DSB site, leading to nucleotide **insertions and deletions (indels)**. Although it has been demonstrated that NHEJ can perfectly repair DSBs [78,79], particularity those with blunt ends, it is much more typical for indels to occur during end processing, leaving information scars where DSBs occurred [80]. As listed in the following text, NHEJ can be divided into three phases.

(i) End protection

The first step is the binding of the Ku heterodimer, consisting of Ku70 and Ku80 [81], to each DNA DSB end. Ku is highly abundant and has a high affinity for DSB ends, thus ensuring a high probability that Ku will be a first responder at DSBs [5] and preventing the extensive end resection required for HR, Alt-EJ, and SSA repair pathways [82]. In addition, Ku enables the recruitment and loading of additional factors involved in end processing and end ligation including the DNA damage response factor DNA-PKcs [83,84].

(ii) End processing

The purpose of end processing is to create a ligatable substrate for Lig4. NHEJ commonly achieves this by using up to 4 bp of complementary nucleotides to anneal the two ends together. These nucleotides can be generated by template-independent X-family DNA polymerases that add nucleotides to one strand of DNA to create homology [85–93] or by nucleases, such as Artemis, that resect the ends to reveal existing DNA homology. Artemis also plays a crucial role in processing incompatible DNA ends or DNA ends with chemical modifications that can prevent ligation [94–103] (Figure I).

(iii) End ligation

Of the three DNA ligases in humans, LIG4 functions exclusively in NHEJ [104] and works in complex with XRCC4; the XRCC4•LIG4 complex is essential for NHEJ [105]. Ligation requires stabilization of the DNA ends in a synaptic configuration that allows phosphodiester bond formation. In addition to using annealed DNA homology, NHEJ can also utilize protein scaffolding to hold the DNA ends together for synapsis [106]. Both XRCC4-like factor (XLF) and paralog of XRCC4 and XLF (PAXX) share structural similarity with XRCC4 and can increase the efficiency of stably positioning DNA DSB ends in ligation-proficient configurations [107–110].

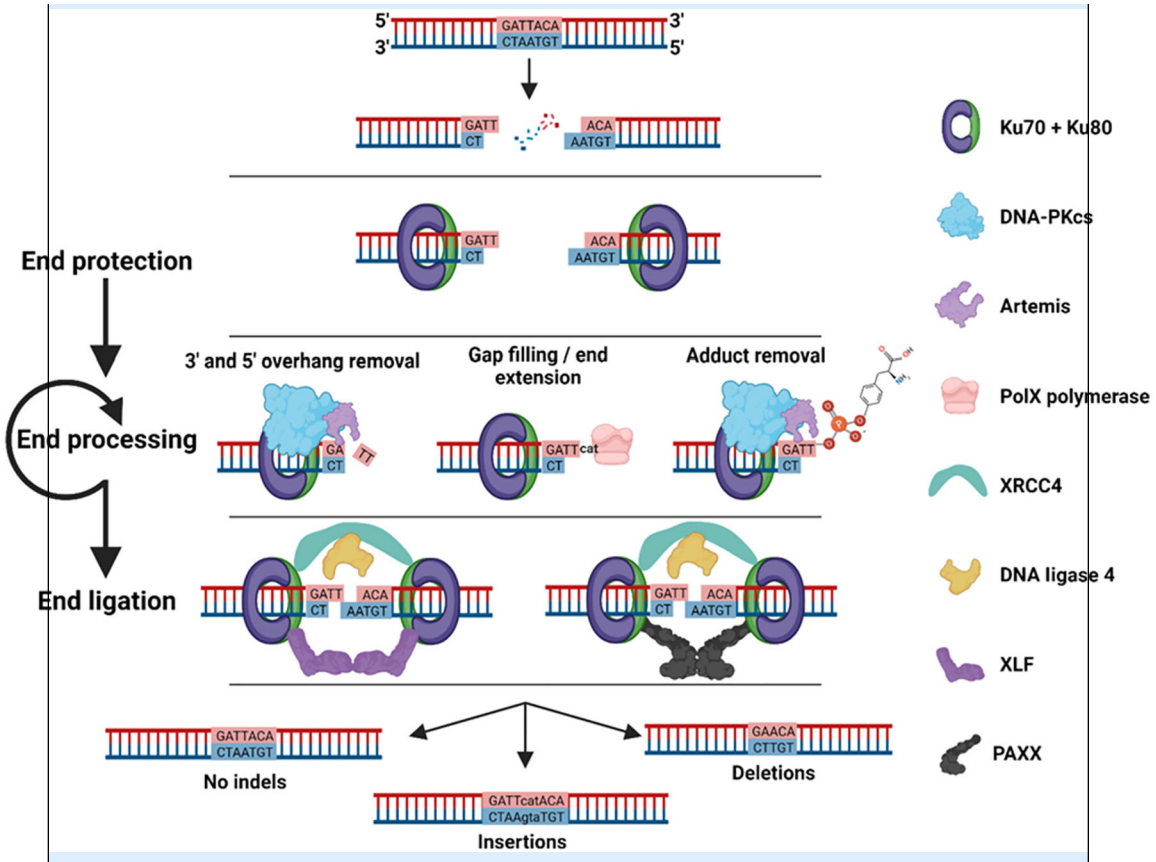

Figure I.

The end-protection, end-processing, and end-ligation steps of non-homologous end joining (NHEJ).

Box 2.

V(D)J recombination occurs in pro- and pre-B cells

V(D)J recombination is initiated by the recombination-activating gene (RAG) complex (consisting of RAG1 and RAG2) that induces DSBs at the variable (V), diversity (D), and joining (J) regions of the *IgH* locus. The RAG complex recognizes recombination signal sequences (RSSs) between the V, D, and J segments. RSSs consist of a conserved heptamer (5′-CACAGTG-3′) and nonamer (5′-ACAAAAACC-3′) separated by 12 or 23 bp of non-conserved sequence [111], and RAG cleavage takes place at one 12RSS and one 23RSS [112–116]. The orientation of RSSs ensures that a D and a J region will join first, followed by a V to DJ joining (Figure I).

RAG cutting generates two DSBs comprising four DNA ends. The two joined ends that lead to expression of the rearranged VDJ region form the coding joint, while the two joined ends that create a closed circle bearing the intermediate region form the signal joint; this closed circle is lost from the genome. Coding and signal ends are both repaired by NHEJ, but the ends are not equivalent. RAG binding at either a 12RSS or 23RSS generates a nick in one DNA strand that results in a 3′-hydroxyl that breaks the phosphate backbone of the opposite DNA strand by nucleophilic attack. This creates a DNA hairpin at each coding end and blunt signal ends. Artemis therefore becomes essential because it is the only nuclease known to open these RAG-generated hairpins. Indeed, mutations that ablate Artemis function result in severe compromised immunodeficiency (SCID) due to inability to complete V(D)J recombination [45,117–120]. Other than this essential requirement for Artemis, NHEJ proceeds as described in Box 1. One feature that is also unique to B cells is that, in addition to DNA polymerases Polμ and Polλ, another X-family DNA polymerase, terminal deoxynucleotidyl transferase (TdT), adds nucleotides to coding ends in a fully template-independent manner.

The early B cell environment is poised to create indels through Artemis and TdT activity during NHEJ. Blunt signal ends are often precisely joined, but this is probably because the RAG complex remains bound [121,122] and that joining the two ends of a single DNA molecule is more efficient than joining two separate molecules. Overall, the implication is that NHEJ repair in pro-B/pre-B cells is deliberately inaccurate to ensure junctional diversity at immunoglobulin loci, indicating why translocations involving *IgH* and several non-*IgH* loci are prevalent in B cell malignancies.

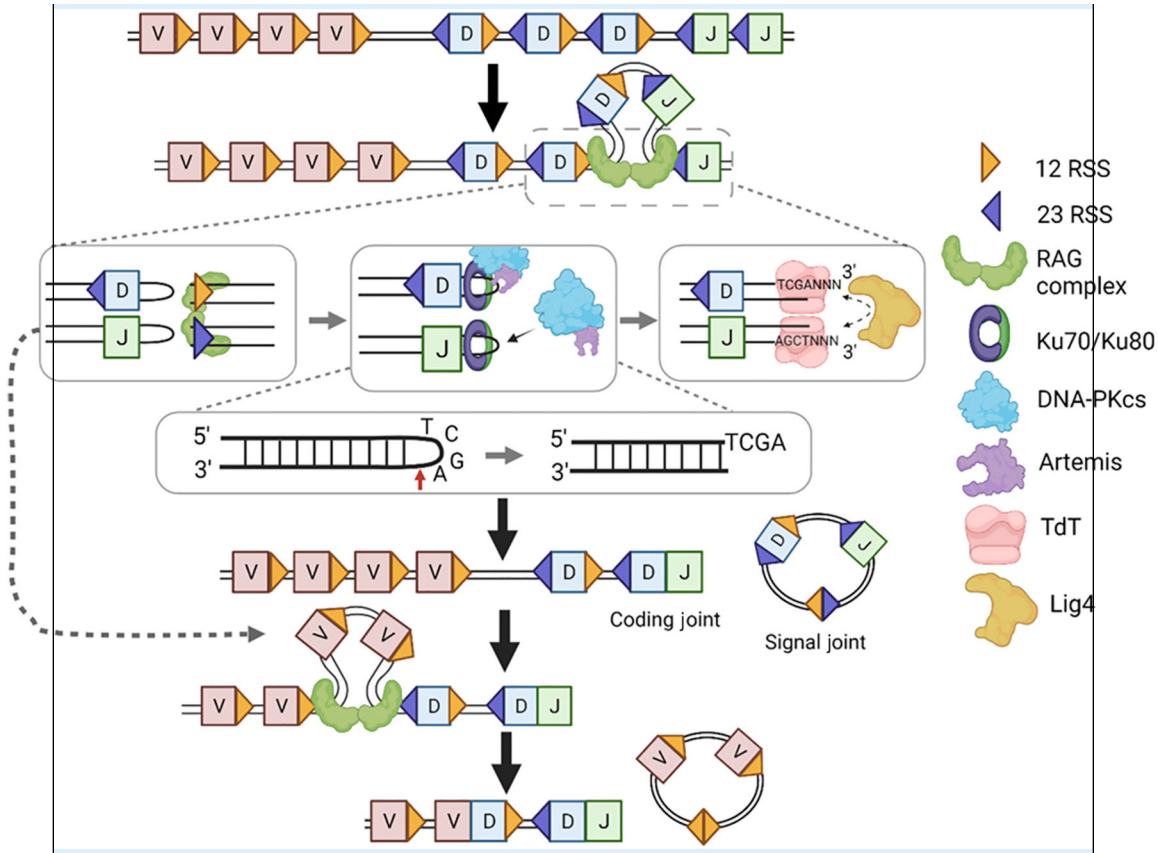

Figure I.
Variable (V), diversity (D), joining (J) recombination in pro-B cells.

Box 3.

Non-biological factors resulting in cancer health disparities

Discussion of the genetic and medical factors leading to cancer health disparities only conveys part of the problem. Another is that there are enormous inequities in our societal and healthcare systems that disproportionately affect minorities and the poor. For example, although the incidence and mortality rates for B cell malignancies are higher in Hispanics versus non-Hispanics living near the USA/Mexico border, for many other cancers, the incidence rate is lower in Hispanics, but the mortality rate is higher [63]. Thus, even though Hispanics are less likely to have some cancers, they are also less likely to survive if they get those cancers, likely because they have less access to healthcare and fewer opportunities for early diagnosis. Lack of access can result from concerns about the financial cost of cancer treatment, language barriers, distance from an NCI-Designated Comprehensive Cancer Center, or concerns about documentation status [64]. The latter point is also significant for cancers where exposure to toxins is a risk factor – there is a link between agricultural pesticides and FL development. An undocumented migrant farm worker may be less likely to speak up about unsafe work conditions or take time off to go to the doctor if they get sick.

Although several of these barriers also exist for African Americans, it is also clear that African Americans do not have access to the most advanced treatments for many cancers. A major factor behind this is that African Americans are under-represented in clinical trials for novel cancer treatments [58]. This has a twofold effect: African Americans are excluded from the most advanced cancer treatments, and drug response is not evaluated in African Americans – for these reasons, it is unknown whether new therapeutic options would have a beneficial or negative effect.

Remedying these issues will require building a bridge between clinical science and these communities to create trust and improve awareness of new treatment programs. A major component will be to diversify the healthcare and scientific workforce by supporting K-12 scientific education in underserved areas, as well as to ensure that more under-represented minorities pursue careers in science and medicine. It also requires a health insurance system that is not tied to work status so as to overcome the financial hardships associated with cancer treatment. There are obviously no easy fixes to these issues, but it is important to emphasize that the more we work to understand cancer disparities, the greater the benefit will be for all.

Outstanding questions

Does the physical location of HFZs in the nucleus play a role in determining how likely it is that a broken V(D)J DSB will interact with an HFZ to create a translocation?

Do HFZs occur in chromatin areas that are more accessible in early B cells?

Are there factors present in B cells that prevent aberrant DNA repair between *IgH* and non-*IgH* loci, and are these defective in cancer patients?

What epigenetic factors, such as DNA methylation, differ between whites, Hispanics, and Blacks that would make HFZ DSBs more prevalent?

Once deaggregated, are there polymorphisms within NHEJ repair genes that make Hispanic and Black individuals more prone to some DNA rearrangements?

Do diet and lifestyle play a role in the risk of B cell malignancies?

Do differences between how an individual identifies culturally and their biological ethnicity or race affect the risk of particular B cell malignancies as a result of diet and environment?

What advances will be necessary to allow us to move beyond a one-size-fits-all approach to cancer diagnosis and treatment to a system where we treat an individual?

Because the formation of HFZ-linked translocations is often a very early event in several cancers, can we develop better ways to detect translocations that are present at a low frequency as an early warning system in high-risk populations?

Highlights

Nearly half of all B cell cancers have a chromosomal translocation between physiological DNA double-strand breaks (DSBs) that create antibody diversity and pathological DSBs at focal human fragile zones (HFZs).

HFZs are dynamic DNA sequences that create transient single-stranded DNA (ssDNA). In early B cells, low levels of activation-induced cytidine deaminase (AID), active variable–(diversity)–joining [V(D)J] recombination, and error-prone non-homologous end joining (NHEJ) create the perfect environment for oncogenic genome rearrangements.

The lack of diversity in genomic data makes it difficult to identify genetic and epigenetic factors leading to disparities among under-represented groups. In blood cancers, this is particularly clear for Hispanics, Latinos, and Blacks. B cell malignancies are also highly diverse, and failure to link specific cancer subtypes to specific ethnicities has masked cancer disparities.

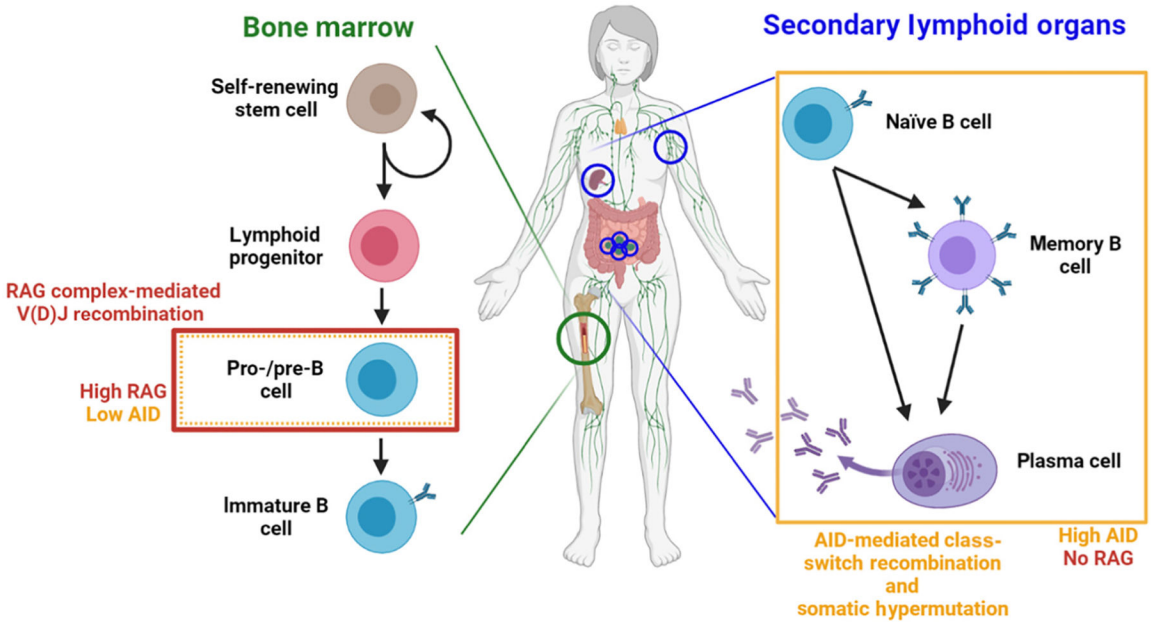

Figure 1. Variable–(diversity)–joining [V(D)J] recombination and human fragile zone (HFZ) double-strand breaks (DSBs) occur within the pro-B/pre-B cell stage of B cell maturation in the bone marrow.

High expression of recombination-activating gene (RAG) components in pro-B cells induces DNA DSBs at the *IgH* locus that are necessary for V(D)J recombination. Low levels of activation-induced cytidine deaminase (AID) in these cells result in DSBs at HFZs, indicating that this is the crucial stage during which many oncogenic translocations involved in B cell malignancies form. Depending on the translocation, cells will either become hyperproliferative at the immature B cell stage or carry the translocation to maturity where additional mutations lead to carcinogenesis. High expression of AID in mature B cells drives somatic hypermutation (SHM) and class-switch recombination (CSR). Physiological CSR DSBs at IgH are also associated with translocations involving the *MYC* and *BCL6* loci. Although the pathological DSB mechanism at *MYC* and *BCL6* is not completely random, the large region over which DSBs occur excludes them from being considered as highly focal HFZs.

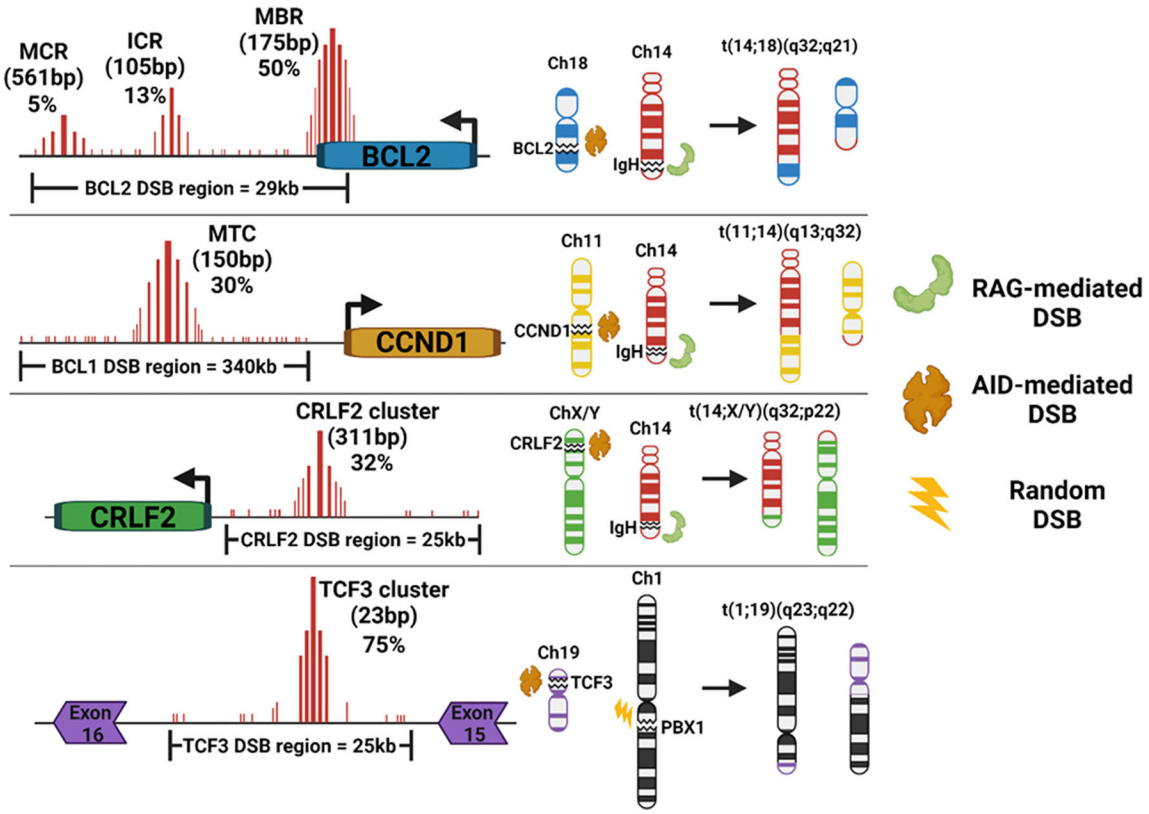

Figure 2. Human fragile zones (HFZs) represent regions or naturally occurring double-strand breaks (DSBs) involved in oncogenic chromosomal translocations.

Translocation junctions from >2000 human patients with B cell leukemia or lymphoma were sequenced, and the locations of translocation breakpoints were compiled into a database. Each red line indicates a mapped DSB from a human patient, and the increased amplitude indicates clustering of DSBs at HFZs. The size of the HFZ and the percentage of DSBs that fall within that HFZ out of the total number of DSBs mapped to that locus are indicated. The coding region of relevant genes as well as the direction of transcription is depicted. Note that most HFZs occur outside coding regions. The chromosomes involved in the oncogenic translocation and whether the DSB is mediated by the recombination-activating gene (RAG) complex, by activation-induced cytidine deaminase (AID), or randomly is indicated. The resulting translocation products are also depicted. Abbreviations: Ch, chromosome; ICR, intermediate cluster region; MBR, major breakpoint region; MCR, minor cluster region.

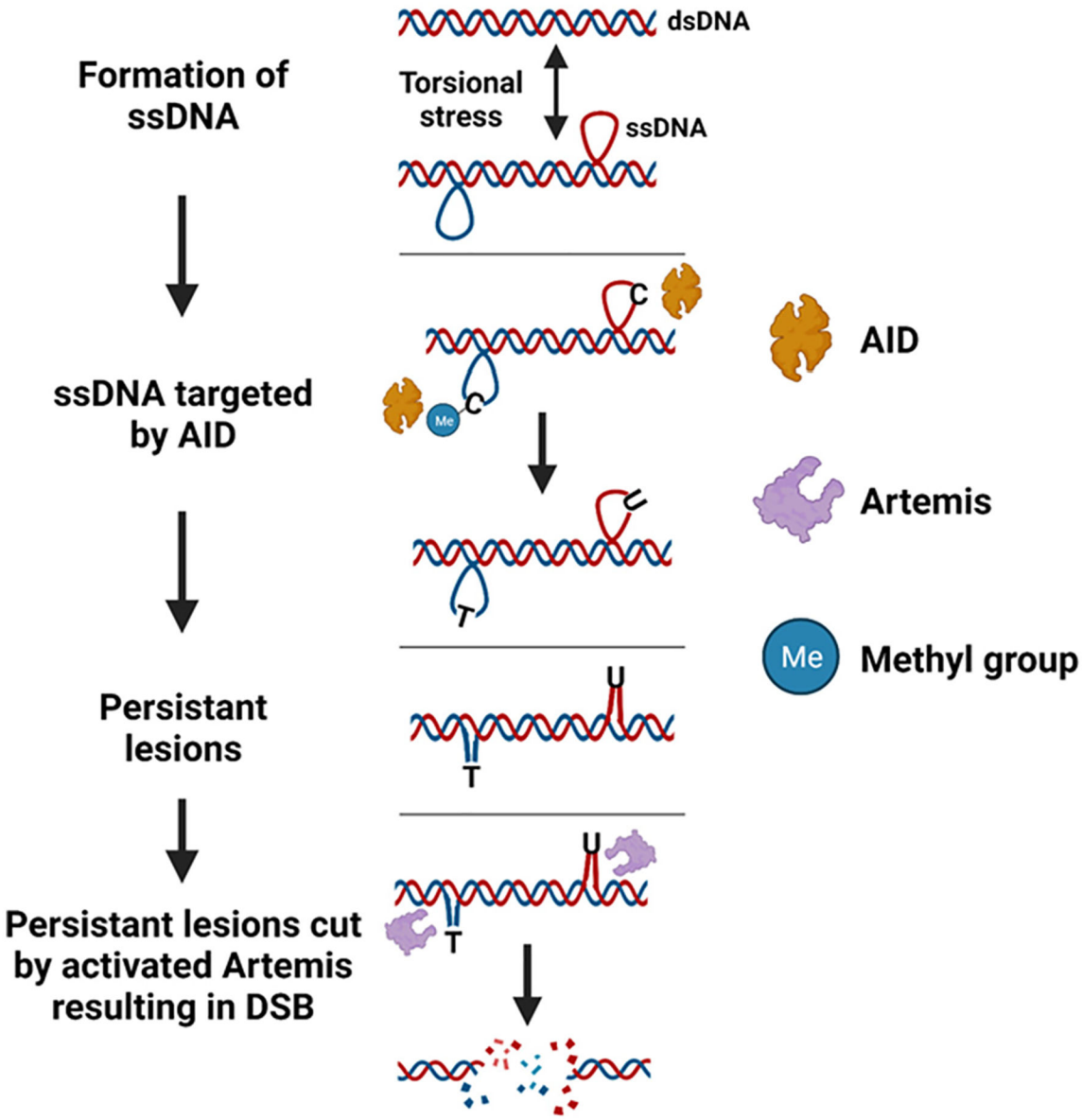

Figure 3. Double-strand break (DSB) formation at human fragile zones (HFZs).

Transient single-stranded (ss)DNA forms at HFZs as a result of multiple factors including torsional stress and shifting between A- and B-form DNA structures. Owing to the repetitive nature and CG richness of HFZs, a strong possibility is that melting of the DNA duplex and reannealing out of register creates slipped-strand DNA that generates ssDNA on each strand. Activation-induced cytidine deaminase (AID) acts on this ssDNA to deaminate it from C➔U or ^{me}C➔T, creating mismatches that produce ssDNA/double-stranded (ds)DNA boundaries. These boundaries activate Artemis, which is required during variable–(diversity)–joining [V(D)J] recombination, and Artemis attacks these to create DSBs at HFZs. The broken DNA ends can then interact with unrepaired V(D)J DNA ends to generate translocations.

Table 1.

HFZs associated with chromosomal translocations and their occurrence in B cell cancers

HFZ	Translocation	Associated B cell cancer	Cancer disparity/risk
BCL2 MBR, ICR, and MCR	*IgH–BCL2* t(14;18) (q32;q21)	>85% of all follicular lymphomas	Higher incidence in non-Hispanic whites, but migration to developed countries increases risk
		20% of diffuse large B cell lymphoma	Higher incidence rate in Asians and Hispanics with EBV infection
BCL1 MTC	*CCND1–IgH* t(11;14) (q13;q32)	>95% of all mantle cell lymphomas	Highest incidence in non-Hispanic white males
		Multiple myeloma	Higher incidence in Blacks of African descent
CRLF2	*IgH–CRLF2* t(14;X/Y) (q32;p22)	65% of Ph-like ALL 3% of all B cell ALL	Higher incidence in Hispanics and those of Native American descent
TCF3 (E2A)	*PBX1–TCF3* t(1;19) (q23;q22)	5% of all B cell ALL	Higher incidence in Hispanics and those of Native American descent