ANNUAL REVIEWS

Annual Review of Cell and Developmental Biology

Diversification of Antibodies: From V(D)J Recombination to Somatic Exon Shuffling

Mikhail Lebedin[1,2] and Kathrin de la Rosa[1,3]

[1]Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany; email: kathrin.delarosa@mdc-berlin.de

[2]Charité – Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[3]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

Annu. Rev. Cell Dev. Biol. 2024. 40:265–81

The *Annual Review of Cell and Developmental Biology* is online at cellbio.annualreviews.org

https://doi.org/10.1146/annurev-cellbio-112122-030835

Keywords

antibody diversity, B cell diversification, exon shuffling, recombination, genomic mobility, mechanisms of diversification

Abstract

Antibodies that gain specificity by a large insert encoding for an extra domain were described for the first time in 2016. In malaria-exposed individuals, an exon deriving from the leukocyte-associated immunoglobulin-like 1 (*LAIR1*) gene integrated via a copy-and-paste insertion into the immunoglobulin heavy chain encoding region. A few years later, a second example was identified, namely a dual exon integration from the leukocyte immunoglobulin-like receptor B1 (*LILRB1*) gene that is located in close proximity to *LAIR1*. A dedicated high-throughput characterization of chimeric immunoglobulin heavy chain transcripts unraveled, that insertions from distant genomic regions (including mitochondrial DNA) can contribute to human antibody diversity. This review describes the modalities of insert-containing antibodies. The role of known DNA mobility aspects, such as genomic translocation, gene conversion, and DNA fragility, is discussed in the context of insert-antibody generation. Finally, the review covers why insert antibodies were omitted from the past repertoire analyses and how insert antibodies can contribute to protective immunity or an autoreactive response.

Contents

INTRODUCTION

The discovery of leukocyte-associated immunoglobulin-like 1 (LAIR1)-containing antibodies was the kick start to unraveling a new layer of antibody diversification. Until recently, functional examples of antibodies with extra domains were exclusive to malaria, and they were shown to result from the integration of exons that encode for domains with inherent specificity for the *Plasmodium* parasite. A *LAIR1* exon can integrate into the variable, diversity, and joining [V(D)J] junction via copy and pasting, giving rise to antibodies with an additional domain on the antibody tip (Tan et al. 2016). Furthermore, one *LAIR1* (Pieper et al. 2017) or two leukocyte immunoglobulin-like receptor B1 (*LILRB1*) exons (Chen et al. 2021) between the J-to-constant junction create bispecific antibodies with additional binding sites between the variable and constant heavy 1 (CH1) immunoglobulin (Ig) domains. Beyond *LAIR1* and *LILRB1* exons, high-throughput Ig RNA analysis identified multiple exons and introns from diverse genes as well as intergenic regions integrating into antibody transcripts independent of malaria (Lebedin et al. 2022). While the latter study could shed some light on the contribution of inserts to antibody diversity, their general impact on antibody specificity remains to be determined. From a molecular point of view, there are key unanswered questions. What is the underlying molecular mechanism of insertion events, and do these events emerge via a true mechanism or rather accidentally? In this review, we strive to elaborate on the potential molecular basis, give an overview of the current state of insertion antibody research, and survey the directions of further inquiries.

INSERT ANTIBODIES: A MECHANISM OR AN ACCIDENT?

The question of if the insertion of extra elements into the antibody heavy chain represents a defined mechanism involves a philosophical disputation (Nicholson 2012). Therefore, we start with developing a definition of the term mechanism. A mechanism may satisfy at least one of the

following criteria: (*a*) Its product—here the genetic event—occurs significantly more frequently than coincidentally due to selective pressure imposed on its generation (selection-induced bias) or (*b*) it involves a dedicated cellular machinery. Since biological mechanisms emerged as chance events that endured through evolution by imparting increased reproductive fitness, these definitions allow us to set a boundary between a valid mechanism and a by-product of various events.

In this context and as outlined in more detail below, insert acquisition does not involve a dedicated cellular machinery. It rather relies on players of diverse molecular processes. The criterion of selection-induced bias is more difficult to address. While the insert acquisition of a single B cell clone is a rare event in an individual, it occurs with significant frequency in the population and becomes relevant under a given selection pressure: More than 5% of malaria-exposed individuals generate LAIR1-containing antibodies that can dominate the specific immune response (Pieper et al. 2017). Thus, rare events that are not based on a dedicated cellular machinery might be beneficial and potentially provide a selective advantage for an individual. It helps to compare the phenomenon of insert antibodies to an established genetic mechanism, namely exon shuffling, which is a well-described process of genetic diversification. Exon shuffling serves as an umbrella term comprising exon integration via (*a*) homologous recombination (HR) (Patthy 1999), (*b*) the transposition of mobile genetic elements (MGEs) (Moran et al. 1999), (*c*) illegitimate recombination (Kuhn et al. 2009), and, in some studies, (*d*) alternative splicing (Babushok et al. 2007). The first occurs between sister chromatids or highly homologous genomic loci, creating new gene alleles through crossover during meiosis or repairing the DNA lesions in somatic cells. By contrast, the transposition of MGEs and illegitimate recombination involve regions of zero to low homology and generate novel and highly distinct sequences. As exons occupy around just 1% of the eukaryotic genome by length, recombination frequently happens in the introns or intergenic sequences, which leads to the rearrangement of either whole intact exons or genes. Exon boundaries were shown to coincide with the protein domain borders with an unexpectedly high frequency that allows the eukaryotic organism to generate novel proteins in a modular fashion (Liu & Grigoriev 2004). In addition, the PR domain containing 9 (PRDM9) protein in mammals directs the meiotic HR away from the promoters and H3K4 trimethylated regions to prevent transcription-HR conflicts and the potential mutagenic effect of the programmed double-strand breaks (DSBs) (Brick et al. 2012). This manifests in the positional bias of germ line recombination.

In conclusion, exon shuffling involves a diverse cellular machinery. Furthermore, it was induced by evolutionary selection pressure. It thus satisfies the above-mentioned criteria for a biological mechanism. The generation of insert-containing antibodies via exon shuffling may be regarded as a somatic counterpart of germ line exon shuffling. It remains open, however, if insert-containing antibodies conferred an evolutionary advantage.

Chromosomal translocations involving the *IGH* locus and oncogenes like *c-Myc*, *CEBP*, and *CRLF2* serve as examples of detrimental side effects or accidents of dedicated mechanisms of diversification via recombination-activating gene (RAG) proteins and activation-induced cytidine deaminase (AID) (Akasaka et al. 2007, Battey et al. 1983, Bemark & Neuberger 2000, Gostissa et al. 2009, Yoda et al. 2010). Every mechanism may generate accidental by-products. In that respect, insertions in the *IGH* that do not confer a function to the antibody may be regarded as accidents.

Finally, LAIR1- or LILRB1-containing antibodies can emerge as random integrations facilitated by the increased fragility of *LAIR1/LILRB1* loci and *IGH* locus rearrangement, which were selected due to malaria exposure. These events reflect how somatic B cell genome mobility, similar to exon shuffling, can result in a novel gene that contributes specific immunity. In the following section, we aim to pinpoint molecular players that potentially contribute to insert-antibody generation.

ANTIBODY DIVERSIFICATION BY INSERT ACQUISITION

A vast repertoire of T cell receptors (TCRs) and B cell receptors (BCRs) is created by three processes orchestrated by dedicated molecular players: (*a*) V(D)J rearrangement, the reshuffling of the variable (V), diversity (D), and joining (J) segments in Ig or TCR chain-encoding genes induced by RAG; (*b*) somatic hypermutation (SHM), the introduction of single-nucleotide (nt) exchanges in heavy and light chains of Ig genes mediated by AID; and (*c*) class switch recombination (CSR), the AID-mediated deletion of constant exons of the Ig heavy chain, switching the isotype of an antibody (Jung et al. 2006). *LAIR1*, *LILRB1*, and a diverse set of genomic fragments integrate into Ig heavy chains, further diversifying the repertoire. These insertions originate from various sources of insert substrates and are incorporated in either of two acceptor break sites: in between the V, D, and J segments (VDJ insertion) or between the J segment and the constant gene (J-CH1 insertion) (Lebedin et al. 2022, Pieper et al. 2017, Tan et al. 2016) (**Figure 1**). VDJ insertions are unlikely to be removed by alternative splicing and are forced to be expressed on the apex of the antibody variable domain, disrupting the original binding site. J-CH1 insertions are genomically located in the J-to-constant intron and are included in the mature transcript only if they bear splicing

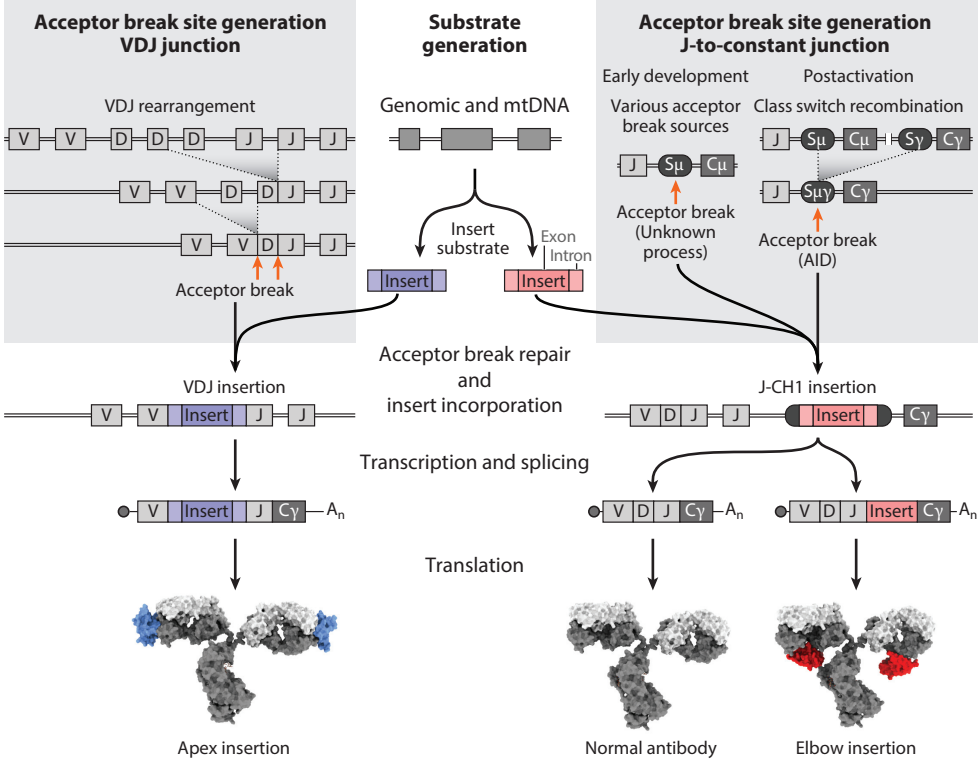

Figure 1

Insert antibodies are generated through the DNA repair of acceptor breaks generated by VDJ rearrangement, CSR, and other processes. Incorporated fragments originate from various fragility sources and comprise exons, introns, and intergenic sequences of genomic and mtDNA. In-frame VDJ insertions are expressed on the apex of the antibody variable domain, while J-CH1 insertions are either removed by alternative splicing or present as a so-called elbow insertion. Abbreviations: AID, activation-induced cytidine deaminase; Cγ, constant gamma; CH1, constant heavy 1; Cμ, constant mu; CSR, class switch recombination; D, diversity; J, joining; mtDNA, mitochondrial DNA; V, variable.

sites. That makes them subject to alternative splicing, making it possible for a B cell to express the insert-free antibody. On the protein level, the J-CH1 insertion is located in the so-called elbow region and does not necessarily destroy the original binding site (**Figure 1**).

Based on the concepts of insert substrates and acceptor break sites, we identified four insert classes with distinct combinations of these features: (*a*) mitochondrial DNA (mtDNA) and (*b*) nuclear DNA detected at VDJ junctions named mitochondrial VDJ (mtVDJ) and nuclear VDJ (nucVDJ) insertions, correspondingly. Furthermore, inserts originating from (*c*) telomere proximal genes and (*d*) early replicating fragile sites (ERFSs) incorporate between J-to-constant junctions—telomere J-to-constant (telJC) and nuclear J-to-constant (nucJC) insertions.

To further characterize the insert antibodies, it is helpful to consider the formation of acceptor break sites and insertion substrates separately.

RAG AND AID ENZYMES HAVE THE POTENTIAL TO GENERATE BOTH ACCEPTOR BREAKS AND INSERT SUBSTRATES

On the one hand, RAG and AID can provide acceptor break sites for insert integration. On the other hand, due to their potential to target genes outside the Ig locus, they may deliver insert substrates. RAG1/2 enzymes specifically recognize recombination signal sequences (RSSs) flanking the V, D, and J segments in Ig loci. This can lead to break generation that results in segment rearrangement and takes place during B and T cell maturation. VDJ insertions are exclusively and precisely located between VDJ joints, thereby preserving the segments. This suggests that RAG-mediated DSBs are the sole acceptor break sites for this insertion type. Of note, VDJ inserts are acquired early in B cell development (Lebedin et al. 2022). Off-target activity of RAG enzymes is often associated with such detrimental consequences as genomic translocation leading to acute lymphoblastic leukemia (Papaemmanuil et al. 2014). Deficiency in the proteasomal degradation of the RAG enzymes at the G1-S transition was also implicated in lymphoid malignancies (Zhang et al. 2011). Various studies indicated that cryptic RSSs flank the most common RAG translocation partners (Blumenberg & Skok 2015, Holmfeldt et al. 2013, Mijušković et al. 2015, Mullighan et al. 2008, Waanders et al. 2012, Zhang & Swanson 2008). Nevertheless, VDJ insertions did not overlap with the documented RAG off-targets. Furthermore, RSS analysis revealed no cryptic RAG recognition sites adjacent to the VDJ insertion substrates. In conclusion, although RAG can theoretically serve both purposes, its contribution is likely limited to the acceptor break generation.

The AID enzyme is expressed in B cells upon antigen encounter and mediates breaks in the antibody switch regions located in the J-to-constant introns. Therefore, it potentially contributes to the acceptor break generation for J-CH1 insert antibodies. However, some J-CH1 insertions are detected in pre–, pro–, and naive B cells. Although AID was demonstrated to be expressed in RAG2+ immature B cells, these cells are deleted through apoptosis (Cantaert et al. 2015), suggesting that during early B cell development J-CH1 insertions (telJC) are acquired through an AID-independent process, the exact nature of which remains elusive. J-CH1 insertions detected in memory and especially in activated B cells (nucJC) are markedly distinct from J-CH1 insertions in naive B cells, which points to a distinct mechanism of acquisition and a potential integration into AID-mediated acceptor break sites. As is the case for RAG enzymes, AID was shown to bind to non-Ig genomic loci, which suggests it might be able to supply insertion substrates (Yamane et al. 2011). AID off-targets were shown to be enriched in actively transcribed genes associated with replication stalling and convergent transcription (Pavri et al. 2010, Yamane et al. 2011), which are the features they share with the fragments found in insert antibodies. However, as AID binding was demonstrated to be insufficient for DSB induction (Yamane et al. 2011), insertions were

compared to AID-mediated translocations (Klein et al. 2011), which revealed a negligible overlap. In summary, AID might provide the acceptor break in activated cells, yet there is no evidence of AID-mediated insert substrate generation. AID-independent acceptor breaks in naive cells might result from various mechanisms of fragility, which are discussed further in the next section.

THE GENERATION OF INSERT SUBSTRATES

Any region of human genomic DNA is subject to various exogenous or endogenous sources of DNA damage. DSBs represent the most relevant type of DNA lesions in the context of insert antibodies. Although oxidative damage induced by ultraviolet radiation and reactive oxygen species was shown to elicit DSBs through attempted base excision repair (Greinert et al. 2012, Yang et al. 2004, 2006), the majority of the DSBs can be attributed to DNA replication and transcription or their combination (Syeda et al. 2014). The replication fork was shown to encounter various impediments, including DNA-binding proteins, secondary structures, and transcription complexes that frequently lead to DNA breaks manifesting in genomic instability (Mehta & Haber 2014, Mirkin & Mirkin 2007, Prado & Aguilera 2005). Transcription was previously shown to facilitate insertions originating from aberrant DNA lesions frequently detected in tumor genomes (Min et al. 2023). We observed that one prominent feature of loci donating Ig insertions was a high level of transcription, pointing to the potential role of transcription-associated fragility in insert generation.

One of the most characterized structures linking transcription and genomic fragility is the R-loop (Castillo-Guzman & Chédin 2021). R-loops represent three-stranded RNA/DNA heteroduplexes formed via the reannealing of the nascent RNA into the transcribed DNA strand, exposing the nontemplate DNA stretch in its single-stranded form (Masukata & Tomizawa 1990). This phenomenon was demonstrated to be crucial in replication initiation in bacterial plasmids, viral genomes, and mtDNA (Belanger & Kreuzer 1998, Drolet & Brochu 2019, Lee & Clayton 1998). R-loops were also observed in switch regions in murine B cells under physiological conditions, associated with programmed DSBs necessary to trigger CSR (Yu et al. 2003). Later it was shown that R-loops play an important role in transcription termination in human genes (Skourti-Stathaki & Proudfoot 2014, Skourti-Stathaki et al. 2011). It has become clear that R-loops occur naturally in numerous conserved genomic loci of various organisms, including humans, and contribute to normal cellular processes (Aguilera & García-Muse 2012, Niehrs & Luke 2020, Sanz et al. 2016). However, failure to timely remove the R-loops may exacerbate transcription-replication collisions and promote nuclease-mediated DSBs, eventually causing translocations (Castillo-Guzman & Chédin 2021, Hamperl et al. 2017, Niehrs & Luke 2020, Sollier et al. 2014). R-loops were previously detected at common translocation partners of Ig loci, such as *c-MYC*, an oncogenic rearrangement frequently associated with Burkitt lymphoma (Yang et al. 2014). AID-mediated DSBs were also shown to depend on R-loop formation (Ruiz et al. 2011). Relative proximity to R-loop–forming loci was demonstrated for insertion substrates, especially for the J-CH1 insertions (Lebedin et al. 2022). Furthermore, J-CH1 insertions were found to be prominently GC enriched—a feature they share with R-loop-forming sequences (Ginno et al. 2012). In conclusion, insertions may originate from R-loop-forming loci.

Several regions, called common fragile sites (CFSs), were found to be particularly prone to replication-associated damage in various cell types (Fungtammasan et al. 2012). CFSs were first described as genomic loci that frequently generate breaks in human lymphocyte metaphase chromosomes upon aphidicolin-induced replication inhibition (Glover et al. 1984). Characteristic features of CFSs include secondary structure formation in AT-rich stretches (Irony-Tur Sinai et al. 2019, Kaushal et al. 2019), late replication (Beau et al. 1998), replication origin paucity

(Letessier et al. 2011, Sugimoto et al. 2018), and association with large genes (Helmrich et al. 2011). Insertions detected in the chimeric antibodies originate from large genes (Lebedin et al. 2022), especially the fragments incorporated in VDJ modality. Although proximity analysis did not reveal a substantial overlap between insert origins and CFSs, it is important to highlight that CFSs, in discordance to their name, are not only cell type–specific but also distinct for every fragile site inducer (Murano et al. 1989).

In contrast to CFSs, fragments incorporated in J-CH1 insert antibodies in in vitro–activated B cells overlap significantly with ERFSs (Lebedin et al. 2022). ERFSs were revealed by chromatin immunoprecipitation with an anti-single-stranded DNA antibody upon hydroxyurea treatment, oncogene induction, and spontaneous replication stress. They were mapped to actively transcribed gene clusters, enriched in GC, and featured early replication timing (Barlow et al. 2013). Similar to CFSs, ERFSs result from replication fork misregulation and are controlled by ataxia telangiectasia and Rad3-related kinase. However, while increased origin firing at CFSs might hamper replication completion (Jones et al. 2013), the major source of DNA damage in ERFSs is intensified replication-transcription conflict (Barlow et al. 2013). ERFSs can contribute to oncogenesis as more than 50% of common amplifications/deletions in human diffuse large B cell lymphoma overlapped with these fragile sites. The observation of ERFSs donating insertions to J-CH1 chimeric antibodies in in vitro–activated B cells suggests that intensive proliferation might create replicative stress in lymphocytes, causing transcribed B cell–specific genes to accumulate DSBs and insert into Ig loci.

One of the most striking features of the insertions detected in J-CH1 modality is the telomeric proximity of their origins (Lebedin et al. 2022). Human telomeres consist of numerous repeats of the TTAGGG sequence, which were shown to form R-loops. As previously discussed, R-loops might serve as a source of telomere fragility, contributing to insert substrates (Palm & Lange 2008). Although telomeric repeats were revealed in the switch region upon genomic DNA sequencing (C. Vázquez-García & K. de la Rosa, unpublished data), most of the telomere-proximal J-CH1 insertions map to subtelomeric regions rather than to chromosome ends. Human subtelomeres contain a high density of genes and MGEs and serve as hot spots for recombination (Calderón et al. 2014, Linardopoulou et al. 2005). The *IGH* locus itself is located in the subtelomeric region of chromosome 14, featuring numerous insertion/deletion polymorphisms generated through nonhomologous recombination (Cook et al. 1994, Walter & Cox 1991). J-CH1 insert antibody bias toward subtelomeric regions suggests an ongoing recombination between the *IGH* locus and other subtelomeres.

Drivers of mutagenesis and obvious candidates for insert-antibody generation include MGEs, such as transposable elements (TEs). TEs are characterized as mobile DNA sequences present in the germ line genome that are capable of mobilization and insertion independent of the host DNA replication (Wells & Feschotte 2020). TEs are ubiquitous in eukaryotic genomes, sometimes constituting up to 85% of the genome (Wegrzyn et al. 2013), and serve as indispensable drivers of genetic diversity, regulation, and evolution (Ågren & Wright 2011, Batzer & Deininger 2002, Chuong et al. 2017). TEs are classified into two major classes: retrotransposons, which replicate through an RNA intermediate and reverse transcription (RT) (class I), and DNA transposons, which excise themselves and insert into a different position in the genome (Finnegan 1989). Among both classes, two types are delineated based on mobilization capability: autonomous elements, which encode for necessary transposition enzymes, and nonautonomous elements, which exploit other elements' machinery to mobilize (Wells & Feschotte 2020). Although some TEs were found to have an insertion site preference for sites in which modification would not influence the normal cell function (Eickbush & Eickbush 2015), they were also implicated in large-scale genome translocations and oncogenesis (Anwar et al. 2017, Hoang et al. 2010, Mazoyer 2005).

Elements of one of the autonomous retrotransposon classes, long interspersed nuclear element-1 (LINE-1), were shown to contribute to VDJ inserts, while less prominent overlap was detected for nonautonomous short interspersed Alu nuclear elements (SINEs/Alu) (Lebedin et al. 2022). In conclusion, the active mobilization of the genome, along with the sporadic DNA damage stemming from fragile sites, R-loops, and telomeres, can contribute to the generation of insert antibodies.

Around 14% of the fragments found in insert antibodies originate from mtDNA. It is important to mention that the mtDNA-derived inserts share higher homology with the mitochondrial genome than the nuclear mitochondrial pseudogenes. This suggests that mtDNA inserts might emerge through de novo nuclear mtDNA numtogenesis (Singh et al. 2017) as was also demonstrated for genomic polymorphisms (Onozawa et al. 2014) and mtDNA patches repairing DSBs in yeast (Ricchetti et al. 1999) and human cancer cells (Singh et al. 2017). mtVDJ inserts originate predominantly from the D-loop region, which bears the transcription start site and the leading-strand origin of replication (Lee & Clayton 1998). Furthermore, mtDNA replication is coupled to transcription through R-loop formation (Holt 2022), which can contribute to its fragility (Sollier et al. 2014). Recently, mtDNA-nucleus transfer was shown to be exacerbated by mtDNA fragmentation following oxidative stress. This transfer can be mitigated by cytoplasmic exonucleases, suggesting that mtDNA fragments leak through the mitochondrial membrane and enter the nucleus, potentially serving as insert substrates (Wu et al. 2023).

SIMILAR OBSERVATIONS

Several processes of copying and pasting were described as contributing to large insert polymorphisms, including templated sequence insertions (TSIs) (Onozawa & Aplan 2016). TSIs emerge predominantly as germ line events and are classified into two types: (*a*) RT-mediated insertions and (*b*) DNA patches copied from a distant genomic region (Onozawa & Aplan 2016). The first class was observed in vitro, as an RT-mediated insertion of LINE-1 retrotransposon sequences to repair DSBs in yeast (Teng et al. 1996). These kinds of events were later shown to be independent of LINE-1 endonuclease activity, further supporting a copy-and-paste mechanism (Morrish et al. 2002). RT-mediated repair could potentially explain the overrepresentation of highly transcribed sequences incorporated in insert antibodies.

In somatic cells, DNA patches up to several kilobases long were detected in repaired DSBs (Yu & Gabriel 1999). These patches included mtDNA. Later, large insertions from Ty retrotransposons, ribosomal DNA, and other fragile regions, but not mtDNA, were found to be copied in the endonuclease-induced DSBs generated in Dna2-deficient yeast (Yu et al. 2018). J-CH1 insertions are frequently represented by several exons and are thus likely to span multiple kilobases in the genome, potentially resembling large TSIs.

A recently discovered theta-mediated end joining (TMEJ) pathway uses distant DNA stretches bearing microhomology to the damaged locus to repair the DSB (Schimmel et al. 2019). Furthermore, RNA can directly contribute to the repair serving as a template for DNA polymerases (Storici et al. 2007). However, both TMEJ and RNA-templated repair were shown to be dependent on the sequence microhomology, which was not observed between the majority of insert substrates and the *IGH* locus (Pieper et al. 2017).

In conclusion, at VDJ junctions, the acquisition of (*a*) mtDNA (mtVDJ inserts) may depend on numtogenesis. (*b*) Nuclear DNA at VDJ junctions (nucVDJ inserts) likely derives from LINE-1 MGEs or highly transcribed genes that are prone to genomic fragility or serve for RNA-mediated DSB repair. Inserts detected at J-to-constant junctions originate from (*c*) telomere proximal genes (telJC inserts) and (*d*) ERFSs (nucJC inserts). An association with R-loop-forming regions was

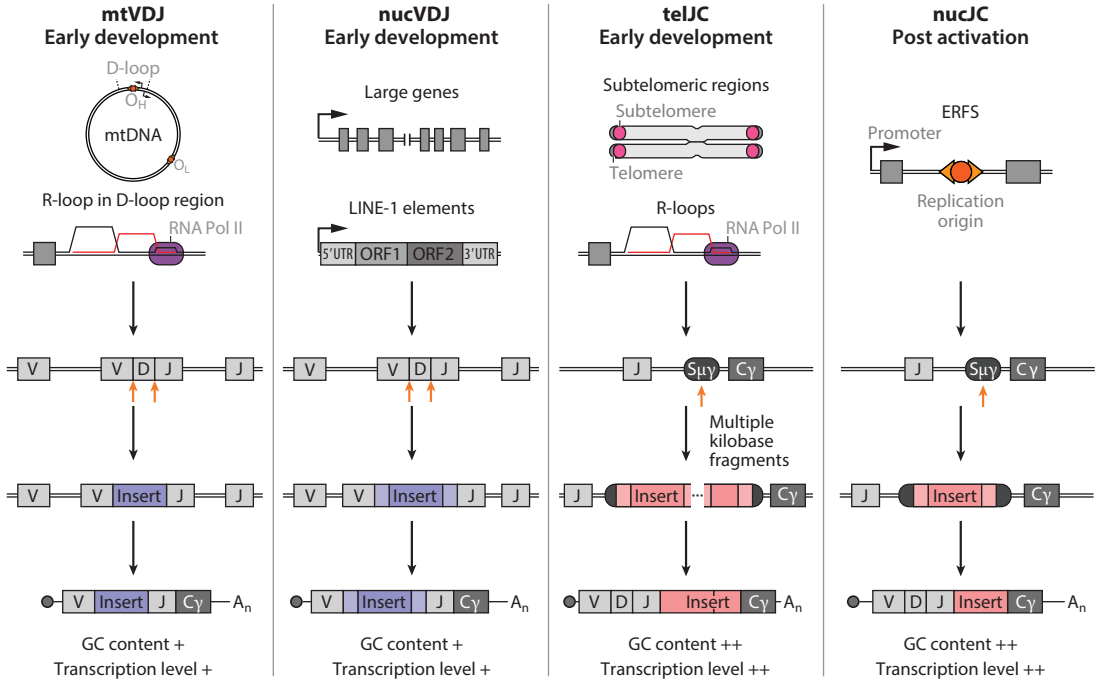

Figure 2

Insert substrate generation. High GC content and increased transcription are characteristic of all detected insert donor sites. Substrates of mtVDJ inserts are predominantly generated by D-loop regions of mtDNA. NucVDJ insertions are enriched in fragments from large genes and LINE-1 elements. TelJC inserts are often represented by multiple-exon insertions originating from subtelomeric regions and R-loops. NucJC insertions feature ERFS fragments. Abbreviations: Cγ, constant gamma; D, diversity; ERFS, early replicating fragile site; J, joining; LINE-1, long interspersed nuclear L1 element; mtDNA, mitochondrial DNA; mtVDJ, mitochondrial VDJ; nucJC, nuclear J-to-constant; nucVDJ, nuclear VDJ; Pol, polymerase; telJC, telomere J-to-constant; UTR, untranslated region; V, variable.

observed for both mtDNA and telJC insertions. While telJC inserts depend on B cell activation, all other insert classes arise early during B cell development. The contribution of substrate sources to different classes of Ig inserts is summarized in **Figure 2**.

LAIR1- AND LILRB1-CONTAINING ANTIBODIES: SOMATIC EXON SHUFFLING GENERATES FUNCTIONAL IMMUNITY

Insert antibodies were first revealed in screens designed to identify broadly neutralizing antibodies against *Plasmodium falciparum* (Tan et al. 2016). Out of 557 screened donors, 4 samples bound erythrocytes infected with at least five parasite isolates. Sequence analysis of subcloned B cells from the reactive donors showed that broadly reactive antibodies bear a large >100 amino acid insertion encoded by a fragment of *LAIR1*. Of note, the insert contained SHMs that diminished LAIR1 binding to collagen but increased binding to the repetitive interspersed family (RIFIN) *P. falciparum* antigen. Analysis of the associated VDJ insert junctions revealed that, for every individual, LAIR1 antibodies emerged in a single B cell clone and subsequently expanded upon malaria stimulation.

Later, the original research was expanded to a larger group of malaria-exposed individuals and European donors (Pieper et al. 2017). In this study, donors were analyzed for the binding of infected erythrocytes. More than 5% of donors living in malaria-endemic regions expressed

LAIR1-containing antibodies, while this type of Ig was not detected in European donors. Of note, in all individuals studied, the antimalaria immune response was dominated by clones expanded from the LAIR1-containing founder B cell. However, no evidence for enhanced protection was found.

Insertions of the extracellular LAIR1 domain maintained the antibody structure. Antibodies therefore seem to be permissive for additional domains, and LAIR1 might not be the only example of a functional receptor antibody. Indeed, further investigation based on an educated guess approach yielded an additional example of functional insertion-containing antibodies that incorporated two domains of another RIFIN-binding receptor—LILRB1 (Chen et al. 2021). This discovery highlights the functional importance of receptor-based antibodies and suggests that binding domains of other proteins might be found to be integrated in antibodies. For instance, apart from RIFIN antigens, the malaria parasite expresses *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) on the surface of infected red blood cells (iRBC), contributing to iRBC rosetting and parasite immune evasion (Barfod et al. 2011, Chen et al. 1998). PfEMP1 was shown to be a target of natural and vaccine-induced immunity (Chan et al. 2019, Olsen et al. 2018), suggesting that its host receptors, including intercellular adhesion molecule 1 and endothelial protein receptor C, can be protective if inserted as additional antibody domains (Tan et al. 2018).

TECHNICAL LIMITATIONS OBSTRUCTING THE DETECTION OF CHIMERIC ANTIBODIES

The low frequency of insert antibodies in the repertoire presents a general problem for their detection given no preexisting information on their sequence or function. Currently available means of insertion detection include in vitro hybridization, genome sequencing, and targeted sequencing of the locus of interest. Hybridization-based methods like fluorescence in situ hybridization bear the advantage of unskewed variant representation, as they do not entail DNA amplification. However, due to technical limitations of microscopy and image analysis, hybridization-based methods feature throughput that might not be high enough for the detection of insertions in the *IGH* locus. Furthermore, a hybridization-based approach is based on sequence-specific probe annealing, precluding the unbiased search for insertions of unknown sequence. Genome sequencing is suitable for the identification of novel sequences incorporated in the Ig locus, but it requires vast resources to detect rare events, potentially represented by a single clone. Long-read switch region sequencing, employed by Pieper et al. (2017), is limited to the detection of genomic events that occur in a narrow region (J-CH1). The combination of these factors begs for gene-targeted RNA-based amplification. As described below, RNA-based methods can be designed to profile both VDJ and spliced J-CH1 insertions.

SUPPRESSION POLYMERASE CHAIN REACTION TARGETS VARIANTS OF DESIRED LENGTH

The amplification of a heterogeneous template mixture by polymerase chain reaction (PCR) is inherently size biased (Dabney & Meyer 2012, Huber et al. 2009). Short amplicons are multiplied with higher probability despite the high processivity of recombinant DNA polymerases routinely used for PCR. In combination with the low frequency of insert-containing antibodies, their analysis represents a technical challenge. Droplet-based single-cell technologies have the potential to overcome the PCR bias. However, currently employed techniques include several rounds of amplification after droplet breakage, which might introduce the bias (Zheng et al. 2017). We addressed this problem by adapting a technique called suppression PCR, originally used for the generation of complementary DNA libraries from a low-concentration starting material, to select

the amplicons of the desired length (Lukyanov et al. 1995). The method is based on introducing inverted terminal repeats (ITRs) during the first amplification step and further amplification with a primer targeting these ITRs. Competition between the normal primer annealing and ITR-mediated intramolecular duplex formation is a size- and concentration-dependent process, which allows one to shift the range of amplified fragments' length by adjusting the primer concentration. We optimized the procedure to select for the antibody heavy chain transcripts containing an ~300-nt insertion. This method amplifies a region encompassing the variable domain and the constant gene exon, allowing one to profile both VDJ and J-CH1 insertions (Lebedin et al. 2022). It is important to point out that the devised technique was optimized to profile a certain length range in *IGH* transcripts, which narrows the scope of the chimeric antibodies detected. Therefore, although the suppression PCR technique enabled sensitive detection of insertion events, further improvements in suppression PCR (or deploying alternative methods) are required to shed light on the real frequency and repertoire of insert-containing antibodies.

INSERTION-MEDIATED DIVERSIFICATION: A UNIQUE PHENOMENON IN THE ANTIBODY HEAVY CHAIN?

Genes encoding for TCR chains and light chains of the BCR are subject to RAG-mediated recombination. This suggests that insertions may be acquired at the respective genomic loci. Nevertheless, insertions were not detected in any human Ig locus apart from *IGH*. While antibodies bind to a broad spectrum of epitopes formed by diverse molecules ranging from haptens to proteoglycans, the TCR interacts with peptides embedded in major histocompatibility complexes (Laethem et al. 2012). This mode of binding constrains the TCR structure, possibly excluding the positive selection of extra domain TCRs. Thus in-frame TCR transcripts containing an insert were not expected. Out-of-frame TCR transcripts containing an insert were also absent from the sequencing libraries. Of note, it needs to be considered that nonsense-mediated decay (NMD) of the *IGH* transcript is distinct from other genes. While we detected numerous conventional out-of-frame *IGH* transcripts, this was not the case for conventional TCR transcripts. Normally, transcripts with premature termination codon are rapidly removed by NMD (Li & Wilkinson 1998). In contrast, nonproductive *IGH* alleles are transcribed along with the productive alleles throughout B cell development until they encounter an antigen (Tinguely et al. 2012). This might render *IGH* exceptionally permissible for the detection of insertions and may provide an explanation for the negative results regarding TCR and Ig light chains.

INSERT-MEDIATED AUTOIMMUNITY AND REDEMPTION

When most frequently detected inserts have been grafted onto recombinant antibodies, insert acquisition has been associated with an increase of polyreactivity (Lebedin et al. 2022). Moreover, *LAIR1* encodes for a collagen-binding domain, which would render the chimeric antibody self-reactive unless it is mutated to abrogate this specificity (Tan et al. 2016). LILRB1 normally binds to human leukocyte antigen class I (HLA-I), yet LILRB1-Ig contains domains not responsible for HLA-I binding (Chen et al. 2021). As LAIR1 and LILRB1 may resemble nucVDJ and nuclear J-CH1 insert classes, they are likely to integrate early during B cell development, giving rise to autoreactive naive B cells. Self-reactivity might drive these cells into an anergic state (Duty et al. 2009). However, anergic self-reactive B cells were shown to be recruited in the germinal center upon antigen exposure in the context of T cell help (Duty et al. 2009). Consequently, B cells expressing self-reactive insert-containing BCRs might undergo clonal redemption through SHM (Burnett et al. 2019), diminishing their self-reactivity while keeping or enhancing pathogen binding (Tan et al. 2016).

Unlike entire domains with dedicated binding sites, various insertions detected in Igs might simply imbue the antibody with ultralong complementarity-determining region (CDR) loops, resembling bovine antibodies. The latter were previously proven to be a promising tool in broad-neutralizing antibody design (Svilenov et al. 2021).

Finally, in case of tolerance breakdown and a failure to acquire SHMs that abrogate self-specificity, insert-containing antibodies might contribute to autoimmunity such as collagen-associated autoimmune disorders.

FURTHER DIRECTIONS

LAIR1- and LILRB1-containing Igs are natural examples of a new class of antibodies. Their generation relies on transcription of the respective donor genes in B cells (Lebedin et al. 2022). Therefore, the natural repertoire of insert antibodies is expected to be limited but might be expanded artificially. Through the integration of a host receptor that is most relevant for pathogenicity, hybrid antibodies could surpass reactivity profiles of conventional antibodies. As B cells of any individual have the intrinsic ability to engineer the antibody locus with distant gene elements, the studies presented above may provide ground for future developments such as novel insert-containing antibodies and the engineering of B cells independent of nucleases.

Beyond its potential for applied science, many unanswered basic questions linger on. So far, information for both the transcript and the underlying genomic architecture is only available for *LAIR1* and *LILRB1* insertions, which narrows the mechanistic insights. For naive B cells that represent single clones, it remains particularly challenging to match the analysis of RNA and genomics at high throughput. This is most relevant when addressing the question of whether all inserts underly a copy and pasting or whether genomic deletions of donor sequences can be observed.

Genomic inserts deriving from subtelomeric regions are likely to cover multiple kilobases because multiple-exon insertions have been detected in antibody transcripts. The application of unbiased approaches for targeted genomic DNA sequences, such as linear amplification-mediated high-throughput genomic translocation sequencing (Hu et al. 2016) combined with long-read sequencing, might shed light on the original presplicing structure of the insertions and thus the cellular machinery involved in their emergence.

As inserts resembling the LAIR1 type have been detected in naive B cells (Lebedin et al. 2022), it remains to be determined how an autoreactive clone containing a collagen-reactive domain like LAIR1 would get positively selected. May similar events rely on chronic antigen exposure? Would selection during an acute infection occur? Despite a break of tolerance, the likelihood of activating a rare B cell clone increases under chronic stimulation.

So far, large templated insertions at the *IGH* locus have been exclusively found in humans. It is still unclear whether such events occur in mice, whether insertional hot spots are conserved, and whether AID- and RAG-deficient mice are suitable models to address further mechanistic insights. Interestingly, in humans and mice, short templated fragments have been described as contributing diversity during the germinal center reaction (Dale et al. 2019). As these events occur postactivation and were excluded from the bioinformatics pipeline used by Lebedin et al. (2022) due to a selection of events >20 nt and a primer design that omits CDR1 and CDR2, it remains open how these events compare.

CONCLUSION

Current data suggest that at least four types of insert classes are contributing to the human antibody repertoire. These include, at the VDJ junctions, (*a*) mtDNA and (*b*) nuclear DNA and, at the

J-to-constant junction, insertions deriving from (*c*) telomere-proximal regions and (*d*) fragile sites. In rare cases, insert-containing antibodies are expressed as functional proteins, e.g., LAIR1 and LILRB1 fragments, that can provide broad specificity against malaria antigens. Beyond that, several instances of in-frame insertions have been revealed, some of which acquired SHMs, pointing toward their affinity maturation. Curiously, insertions can introduce binding to self- or irrelevant targets, which may contribute to autoreactivity.

Despite hints that molecular processes are involved in insert acquisition, deeper insights in the molecular machineries remain elusive. Nevertheless, we suggest that there is no exclusive way for their emergence. As for the acceptor break, current evidence speaks for RAG-mediated DSBs providing the destination point for VDJ insertions, while AID's role is less defined and might contribute to both the insert substrate and acceptor break site. Similar to germ line exon shuffling, various genome mobility features may contribute inserts to antibodies expressed by single B cells, representing somatic exon shuffling.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors were supported by the German Research Foundation (394523286) and the European Research Council (948464). We thank our colleagues Emre Mert İpekoğlu, Casper Silvis, and Clara Vázquez García for critically reading the manuscript.

LITERATURE CITED

Ågren JA, Wright SI. 2011. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosom. Res.* 19(6):777–86

Aguilera A, García-Muse T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* 46(2):115–24

Akasaka T, Balasas T, Russell LJ, Sugimoto K, Majid A, et al. 2007. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood* 109(8):3451–61

Anwar SL, Wulaningsih W, Lehmann U. 2017. Transposable elements in human cancer: causes and consequences of deregulation. *Int. J. Mol. Sci.* 18(5):974

Babushok DV, Ostertag EM, Kazazian HH. 2007. Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell. Mol. Life Sci.* 64(5):542–54

Barfod L, Dalgaard MB, Pleman ST, Ofori MF, Pleass RJ, Hviid L. 2011. Evasion of immunity to *Plasmodium falciparum* malaria by IgM masking of protective IgG epitopes in infected erythrocyte surface-exposed PfEMP1. *PNAS* 108(30):12485–90

Barlow JH, Faryabi RB, Callén E, Wong N, Malhowski A, et al. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152(3):620–32

Battey J, Moulding C, Taub R, Murphy W, Stewart T, et al. 1983. The human *c-myc* oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34(3):779–87

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3(5):370–79

Beau MML, Rassool FV, Neilly ME, Espinosa R, Glover TW, et al. 1998. Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.* 7(4):755–61

Belanger KG, Kreuzer KN. 1998. Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. *Mol. Cell* 2(5):693–701

Bemark M, Neuberger MS. 2000. The c-*MYC* allele that is translocated into the *IgH* locus undergoes constitutive hypermutation in a Burkitt's lymphoma line. *Oncogene* 19(30):3404–10
Blumenberg L, Skok JA. 2015. RAG off-target activity is in the loop. *Trends Mol. Med.* 21(12):733–35
Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485(7400):642–45
Burnett DL, Reed JH, Christ D, Goodnow CC. 2019. Clonal redemption and clonal anergy as mechanisms to balance B cell tolerance and immunity. *Immunol. Rev.* 292(1):61–75
Calderón M del C, Rey M-D, Cabrera A, Prieto P. 2014. The subtelomeric region is important for chromosome recognition and pairing during meiosis. *Sci. Rep.* 4(1):6488
Cantaert T, Schickel J-N, Bannock JM, Ng Y-S, Massad C, et al. 2015. Activation-induced cytidine deaminase expression in human B cell precursors is essential for central B cell tolerance. *Immunity* 43(5):884–95
Castillo-Guzman D, Chédin F. 2021. Defining R-loop classes and their contributions to genome instability. *DNA Repair* 106:103182
Chan J-A, Boyle MJ, Moore KA, Reiling L, Lin Z, et al. 2019. Antibody targets on the surface of *Plasmodium falciparum*-infected erythrocytes that are associated with immunity to severe malaria in young children. *J. Infect. Dis.* 219(5):819–28
Chen Q, Barragan A, Fernandez V, Sundström A, Schlichtherle M, et al. 1998. Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite *P. falciparum*. *J. Exp. Med.* 187(1):15–23
Chen Y, Xu K, Piccoli L, Foglierini M, Tan J, et al. 2021. Structural basis of malaria RIFIN binding by LILRB1-containing antibodies. *Nature* 592(7855):639–43
Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18(2):71–86
Cook GP, Tomlinson IM, Walter G, Riethman H, Carter NP, et al. 1994. A map of the human immunoglobulin V_H locus completed by analysis of the telomeric region of chromosome 14q. *Nat. Genet.* 7(2):162–68
Dabney J, Meyer M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52(2):87–94
Dale GA, Wilkins DJ, Bohannon CD, Dilernia D, Hunter E, et al. 2019. Clustered mutations at the murine and human IgH locus exhibit significant linkage consistent with templated mutagenesis. *J. Immunol.* 203(5):1252–64
Drolet M, Brochu J. 2019. R-loop-dependent replication and genomic instability in bacteria. *DNA Repair* 84:102693
Duty JA, Szodoray P, Zheng N-Y, Koelsch KA, Zhang Q, et al. 2009. Functional anergy in a subpopulation of naive B cells from healthy humans that express autoreactive immunoglobulin receptors. *J. Exp. Med.* 206(1):139–51
Eickbush TH, Eickbush DG. 2015. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol. Spectr.* 3(2):mdna3-0011-2014
Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5(4):103–7
Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA, Makova KD. 2012. A genome-wide analysis of common fragile sites: What features determine chromosomal instability in the human genome? *Genome Res.* 22(6):993–1005
Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45(6):814–25
Glover TW, Berger C, Coyle J, Echo B. 1984. DNA polymerase α inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* 67(2):136–42
Gostissa M, Yan CT, Bianco JM, Cogné M, Pinaud E, Alt FW. 2009. Long-range oncogenic activation of *Igh/c-myc* translocations by the *Igh* 3′ regulatory region. *Nature* 462(7274):803–7
Greinert R, Volkmer B, Henning S, Breitbart EW, Greulich KO, et al. 2012. UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages. *Nucleic Acids Res.* 40(20):10263–73
Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. 2017. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* 170(4):774–86.e19

Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* 44(6):966–77
Hoang ML, Tan FJ, Lai DC, Celniker SE, Hoskins RA, et al. 2010. Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. *PLOS Genet.* 6(12):e1001228
Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, et al. 2013. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* 45(3):242–52
Holt IJ. 2022. R-loops, methods and protocols. *Methods Mol. Biol.* 2528:173–202
Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL. 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* 11(5):853–71
Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Welch DBM. 2009. Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ. Microbiol.* 11(5):1292–302
Irony-Tur Sinai M, Salamon A, Stanleigh N, Goldberg T, Weiss A, et al. 2019. AT-dinucleotide rich sequences drive fragile site formation. *Nucleic Acids Res.* 47(18):9685–95
Jones RM, Mortusewicz O, Afzal I, Lorvellec M, García P, et al. 2013. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* 32(32):3744–53
Jung D, Giallourakis C, Mostoslavsky R, Alt FW. 2006. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* 24:541–70
Kaushal S, Wollmuth CE, Das K, Hile SE, Regan SB, et al. 2019. Sequence and nuclease requirements for breakage and healing of a structure-forming (AT)n sequence within fragile site FRA16D. *Cell Rep.* 27(4):1151–64.e5
Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, et al. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147(1):95–106
Kuhn GCS, Teo CH, Schwarzacher T, Heslop-Harrison JS. 2009. Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. *Heredity* 102(5):453–64
Laethem FV, Tikhonova AN, Singer A. 2012. MHC restriction is imposed on a diverse T cell receptor repertoire by CD4 and CD8 co-receptors during thymic selection. *Trends Immunol.* 33(9):437–41
Lebedin M, Foglierini M, Khorkova S, García CV, Ratswohl C, et al. 2022. Different classes of genomic inserts contribute to human antibody diversity. *PNAS* 119(36):e2205470119
Lee DY, Clayton DA. 1998. Initiation of mitochondrial DNA replication by transcription and R-loop processing. *J. Biol. Chem.* 273(46):30614–21
Letessier A, Millot GA, Koundrioukoff S, Lachagès A-M, Vogt N, et al. 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470(7332):120–23
Li S, Wilkinson MF. 1998. Nonsense surveillance in lymphocytes? *Immunity* 8(2):135–41
Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437(7055):94–100
Liu M, Grigoriev A. 2004. Protein domains correlate strongly with exons in multiple eukaryotic genomes – evidence of exon shuffling? *Trends Genet.* 20(9):399–403
Lukyanov KA, Launer GA, Tarabykin VS, Zaraisky AG, Lukyanov SA. 1995. Inverted terminal repeats permit the average length of amplified DNA fragments to be regulated during preparation of cDNA libraries by polymerase chain reaction. *Anal. Biochem.* 229(2):198–202
Masukata H, Tomizawa J. 1990. A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* 62(2):331–38
Mazoyer S. 2005. Genomic rearrangements in the BRCA1 and BRCA2 genes. *Hum. Mutat.* 25(5):415–22
Mehta A, Haber JE. 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* 6(9):a016428
Mijušković M, Chou Y-F, Gigi V, Lindsay CR, Shestova O, et al. 2015. Off-target V(D)J recombination drives lymphomagenesis and is escalated by loss of the Rag2 C terminus. *Cell Rep.* 12(11):1842–52
Min J, Zhao J, Zagelbaum J, Lee J, Takahashi S, et al. 2023. Mechanisms of insertions at a DNA double-strand break. *Mol. Cell* 83(14):2434–48.e7

Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.* 71(1):13–35

Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* 283(5407):1530–34

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, et al. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31(2):159–65

Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, et al. 2008. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 322(5906):1377–80

Murano I, Kuwano A, Kajii T. 1989. Fibroblast-specific common fragile sites induced by aphidicolin. *Hum. Genet.* 83(1):45–48

Nicholson DJ. 2012. The concept of mechanism in biology. *Stud. Hist. Philos. Sci. C* 43(1):152–63

Niehrs C, Luke B. 2020. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat. Rev. Mol. Cell Biol.* 21(3):167–78

Olsen RW, Ecklu-Mensah G, Bengtsson A, Ofori MF, Lusingu JPA, et al. 2018. Natural and vaccine-induced acquisition of cross-reactive IgG-inhibiting ICAM-1-specific binding of a *Plasmodium falciparum* PfEMP1 subtype associated specifically with cerebral malaria. *Infect. Immun.* 86(4):e00622-17

Onozawa M, Aplan PD. 2016. Templated sequence insertion polymorphisms in the human genome. *Front. Chem.* 4:43

Onozawa M, Zhang Z, Kim YJ, Goldberg L, Varga T, et al. 2014. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *PNAS* 111(21):7729–34

Palm W, de Lange T. 2008. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* 42:301–34

Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, et al. 2014. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in *ETV6-RUNX1* acute lymphoblastic leukemia. *Nat. Genet.* 46(2):116–25

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling—a review. *Gene* 238(1):103–14

Pavri R, Gazumyan A, Jankovic M, Virgilio MD, Klein I, et al. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143(1):122–33

Pieper K, Tan J, Piccoli L, Foglierini M, Barbieri S, et al. 2017. Public antibodies to malaria antigens generated by two *LAIR1* insertion modalities. *Nature* 548(7669):597–601

Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* 24(6):1267–76

Ricchetti M, Fairhead C, Dujon B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402(6757):96–100

Ruiz JF, Gómez-González B, Aguilera A. 2011. AID induces double-strand breaks at immunoglobulin switch regions and *c-MYC* causing chromosomal translocations in yeast THO mutants. *PLOS Genet.* 7(2):e1002009

Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, et al. 2016. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* 63(1):167–78

Schimmel J, van Schendel R, den Dunnen JT, Tijsterman M. 2019. Templated insertions: a smoking gun for polymerase theta-mediated end joining. *Trends Genet.* 35(9):632–44

Singh KK, Choudhury AR, Tiwari HK. 2017. Numtogenesis as a mechanism for development of cancer. *Semin. Cancer Biol.* 47:101–9

Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* 28(13):1384–96

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42(6):794–805

Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* 56(6):777–85

Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. 2007. RNA-templated DNA repair. *Nature* 447(7142):338–41

Sugimoto N, Maehara K, Yoshida K, Ohkawa Y, Fujita M. 2018. Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. *Nucleic Acids Res.* 46(13):6683–96

Svilenov HL, Sacherl J, Protzer U, Zacharias M, Buchner J. 2021. Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design. *Nat. Commun.* 12(1):6737

Syeda AH, Hawkins M, McGlynn P. 2014. Recombination and replication. *Cold Spring Harb. Perspect. Biol.* 6(11):a016550

Tan J, Piccoli L, Lanzavecchia A. 2018. The antibody response to *Plasmodium falciparum*: cues for vaccine design and the discovery of receptor-based antibodies. *Annu. Rev. Immunol.* 37:225–46

Tan J, Pieper K, Piccoli L, Abdi A, Foglierini M, et al. 2016. A *LAIR1* insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* 529(7584):105–9

Teng S-C, Kim B, Gabriel A. 1996. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* 383(6601):641–44

Tinguely A, Chemin G, Péron S, Sirac C, Reynaud S, et al. 2012. Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol. Cell. Biol.* 32(1):107–17

Waanders E, Scheijen B, van der Meer LT, van Reijmersdal SV, van Emst L, et al. 2012. The origin and nature of tightly clustered *BTG1* deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLOS Genet.* 8(2):e1002533

Walter MA, Cox DW. 1991. Nonuniform linkage disequilibrium within a 1,500-kb region of the human immunoglobulin heavy-chain complex. *Am. J. Hum. Genet.* 49(5):917–31

Wegrzyn JL, Lin BY, Zieve JJ, Dougherty WM, Martínez-García PJ, et al. 2013. Insights into the loblolly pine genome: characterization of BAC and fosmid sequences. *PLOS ONE* 8(9):e72439

Wells JN, Feschotte C. 2020. A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* 54:539–61

Wu J, Liu Y, Ou L, Gan T, Zhangding Z, et al. 2023. Genome editing induces the transfer of mitochondrial DNA into the nuclear genome. bioRxiv 2023.07.19.549443. **https://doi.org/10.1101/2023.07.19.549443**

Yamane A, Resch W, Kuo N, Kuchen S, Li Z, et al. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* 12(1):62–69

Yang N, Chaudhry MA, Wallace SS. 2006. Base excision repair by hNTH1 and hOGG1: a two edged sword in the processing of DNA damage in γ-irradiated human cells. *DNA Repair* 5(1):43–51

Yang N, Galick H, Wallace SS. 2004. Attempted base excision repair of ionizing radiation damage in human lymphoblastoid cells produces lethal and mutagenic double strand breaks. *DNA Repair* 3(10):1323–34

Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol. Cell* 53(3):484–97

Yoda A, Yoda Y, Chiaretti S, Bar-Natan M, Mani K, et al. 2010. Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *PNAS* 107(1):252–57

Yu K, Chedin F, Hsieh C-L, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* 4(5):442–51

Yu X, Gabriel A. 1999. Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4(5):873–81

Yu Y, Pham N, Xia B, Papusha A, Wang G, et al. 2018. Dna2 nuclease deficiency results in large and complex DNA insertions at chromosomal breaks. *Nature* 564(7735):287–90

Zhang L, Reynolds TL, Shan X, Desiderio S. 2011. Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. *Immunity* 34(2):163–74

Zhang M, Swanson PC. 2008. V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. *J. Biol. Chem.* 283(11):6717–27

Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8(1):14049