Defining risk factors for genomic instability in B cells: Novel insights from NGS-based technologies
Hogenbirk, M.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
**Genetic instability and cancer**

One in eight deaths worldwide are caused by cancer.¹ Most cancers arise from an accumulation of somatic mutations within cells from pre-malignant tumors. These mutations include subtle DNA nucleotide substitutions, insertions and deletions, copy number variations, and chromosomal aberrations. In this thesis, we focus on a chromosomal aberration known as chromosomal translocations which occur when DNA has been broken and is subsequently rejoined to a DNA segment from elsewhere in the genome (Figure 1A).² ³ In specific cases, the translocation involves two gene bodies, and a hybrid gene is created.⁴ ¹³ A chromosomal translocation can result in expression changes of genes flanking the fusion site.⁹ ¹₄ ¹₆ As translocations can lead to gene deregulation, they can drive oncogenic transformation if a cancer gene is involved.⁴ ¹₁ ¹₇ ¹₉ Although there is ample evidence that translocations are involved in the initiation of carcinogenesis, the molecular mechanism underlying these chromosomal aberrations remains speculative.⁴ Accordingly, defining novel risk factors that contribute to the origin of translocations will provide a deeper understanding of cancer development.

---

### Figure 1. Chromosomal translocations and cancer development.

Chromosomal translocations are genetic aberrations caused by rearrangement of parts between non-homologous chromosomes or between segments within a chromosome. Whether chromosomal translocations require the formation of a single or paired DSBs is unknown. Besides repair in cis, examples of reciprocal and non-reciprocal chromosomal translocation are shown. Figure is adapted from Nussenzweig et al.⁴⁴

### Table 1. Genes involved in rearrangements after applying TC-Seq on switch activated B cells. The associated translocations (or deletion) observed in human mature B cell lymphomas are shown (BCL). BCL types are abbreviated as follows: MALT, mucosa-associated lymphoid tissue lymphoma; DLBCL, diffuse large B cell lymphoma; FL, follicular lymphoma; MCL, mantle cell lymphoma; BL, Burkitt’s lymphoma; B-PLL, B cell prolymphocytic leukaemia. Table is adapted from Nussenzweig et al.⁴⁴.
Figure 2. Chromosomal translocation models. (A) The breakage-first model. Upon DNA damage on distant chromosomes, the broken chromosome ends roam the nuclear space by diffusion and undergo illegitimate joining when they encountering another DSB. This model implies that two DSBs occur first. (B) In the contact-first model translocations occur among chromosomes which are in spatial proximity to each other. Upon concurrent damage of neighboring chromosomes, the broken chromosome ends are mis-joined to form a translocation. (C) In the misrecombination model, homology dependent repair occurs when a DSB is misrepaired via homologous association with undamaged DNA. (D) In the transcription based model a DNA–topo I cleavage complex is formed in close proximity to the DSB. The complex subsequently interacts with topo-I bound to a transcription unit (within the same transcriptional factory) on a different chromosome. This leads to the formation of a cleavage complex at the latter site. The two DNA–topo I cleavage complexes are then reversed by joining to a 5-hydroxyl group on the opposing chromosome. By cleavage and ligation this leads to the formation of a reciprocal exchange event.
Chromosomal translocation models

The fact that chromosomal translocations are infrequent events has hampered understanding their origin. As a consequence, four chromosome translocation models are currently considered: (1) the breakage-first; (2) the contact-first; (3) the misrecombination; and (4) the transcription-based model. The breakage-first model, first postulated over 70 years ago, proposes that double-stranded DNA breaks (DSBs) occur first in distant chromosomes after which the broken chromosome ends roam the nuclear space by diffusion (Figure 2A). The two DSBs, which simultaneously exist in a single cell, subsequently undergo illegitimate joining leading to a mis-joined structure that can give rise to chromosomal translocation. In contrast, the contact-first model proposes that chromosomes need to be physically touching before mis-joining can take place (Figure 2B). It has been proposed that spatial proximity is likely the reason for the occurrence of typical translocations like the Philadelphia chromosome in chronic myeloid leukaemia which gives rise to the BCR-ABL fusion. The mis-recombination model postulates that homologous recombination between the DSB and an undamaged homologous (or near homologous) sequence on another chromosome results in mis-joining of DNA ends, resulting in a chromosomal translocation (Figure 2C). In the transcription-based translocation model, a DSB-end is brought in contact with a topoisomerase I (topo-I) molecule via transcription. These trapped DNA–topo-I cleavage complexes interact with another topo-I molecule on a temporarily inactive transcription unit, at the same transcription factory (Figure 2D). As topo-I enzymes can cut double-stranded DNA, and re-anneal the strands, DNA cleavage and strand exchange between the two topo-I cleavage complexes can result in a translocation. Although an unifying chromosomal translocation model is missing to date, it is generally accepted that DSBs are the critical initiating lesions leading to chromosomal translocations.

Chromosomal translocations and DNA breaks

Chromosomal translocation models can be classified on the basis of the number of DSBs that need to pre-exist before a translocation can be initiated. The breakage-first and contact-first models assume that two DSBs must pre-exist (Figure 2AB). This notion is based on biophysical experiments which induce DSBs by ionizing irradiation. When the dose-rate was kept constant, the yield of interchromosomal exchanges appeared proportional to the square of the radiation dose. This quadratic induction of exchanges is easily understood as the probability of inducing two DSBs simultaneously in two chromosomes increases as the radiation dose increases (Figure 3A). Accordingly, a ~80 fold increase in chromosomal translocations was observed when two, instead of one, I-SceI induced DSBs were generated. In contrast, the misrecombination and transcription-based model are based on the concept that translocation events predominantly arise from a single pre-existing DSB (Figure 2CD). These models are supported by experiments where a linear relationship between ultra-soft X-rays and chromosomal aberrations was observed (Figure 3B). Likewise a linear dose-effect relationship was reported when DSBs are induced by mitomycin C. Taken together, these results suggest that the levels by which DSBs are introduced determine the translocation rate.
When considering chromosomal translocations as second-order reactions, which is implied by the ‘breakage-first’ and ‘contact-first’ model on the origin of cTs, two breaks need to pre-exist before a translocation reaction takes place. According to these models, one expects that any additional DSB should accelerate the translocation reaction, and result in a disproportional increase in the occurrence chromosomal aberrations. (B) A linear increase is expected when a single DSB can provoke a chromosomal translocation event.

**Dsb repair and chromosomal translocations**

DSBs can arise from endogenous and exogenous sources. Upon their acquisition, factors are recruited to the DNA ends, which activate DNA damage response pathways to coordinate the cellular response and repair. Three pathways exist to repair DSBs: homologous recombination (HR), classical non-homologous end-joining (C-NHEJ), and alternative-end-joining (A-EJ). HR and C-NHEJ are genetically and biochemically well-defined. DSB repair by HR is rather error-free as it takes advantage of the intact genetic information of the sister chromatid which is present in S/G2 phase of the cell cycle. In contrast, C-NHEJ, the predominant DSB repair pathway in higher eukaryotes, involves processing of DNA ends and consequently is error-prone. If C-NHEJ is impaired, A-EJ can compensate, which is characterized by a frequent occurrence of microhomology at fusion sites. There is still no agreement as to whether A-EJ comprises one or more pathways. To date, A-EJ has been defined as any form of end-joining that preferentially occurs in the absence of C-NHEJ factors. Although C-NHEJ and HR ensure that DSBs are normally effectively repaired, when these strategies fail, illegitimate joining of two unrelated DNA ends can give rise to a chromosomal translocation. In fact, deficiency in C-NHEJ leads to an increased occurrence of translocations indicating that normally C-NHEJ suppresses translocations. Interestingly, as microhomology is frequently observed in the majority of translocation junctions, and A-EJ prefers ends with short microhomologies, strongly suggest a role for A-EJ in establishing a translocation. The lack of extensive homology at translocation junctions sites, excludes HR as a repair pathway in establishing translocations. The fact that A-EJ has not been characterized at the molecular level has hampered the understanding the exact mechanism leading to translocations. Hence, novel insights into the mechanism(s) of A-EJ are likely to provide novel clues on how translocations arise.
B cells as a model system to study chromosomal translocations

Chromosomal translocations are frequently observed within B cell malignancies\textsuperscript{5-17, 19} which is in line with the notion that the development of B cell malignancies occurs relatively frequent; about 20 new cases of lymphomas are diagnosed per 100000 people per year in the Western world of which 95\% are of B cell origin.\textsuperscript{43} B cells are therefore a favored model system in which chromosomal translocations are studied. Of note, most incipient cancer cells likely share fundamental mechanisms involved in the development of, and protection against, chromosomal translocations.\textsuperscript{44} The majority of non-Hodgkin’s lymphomas are of germinal center or post-germinal center origin.\textsuperscript{45-46} Germinal centers are sites within lymphoid organs where antigen selected B cells proliferate, differentiate, and further improve the antibody repertoire by initiating two processes: class switch recombination (CSR) and somatic hypermutation (SHM). Both processes require an enzyme known as activation-induced cytidine deaminase (AID).\textsuperscript{47-48}

Activation-induced cytidine deaminase

In 1999 Muramatsu and collaborators discovered AID by comparing subtracted cDNA from switch- and non-induced murine B lymphoma CH12F3-2 cells\textsuperscript{49}. An open reading frame of 1.2 kb in length was revealed, encoding the 198-residue protein with a molecular mass of 24kDa. Based on the crystal structure of a yeast ortholog called CDD1, AID’s molecular structure was predicted.\textsuperscript{50} AID’s catalytic function is to remove the amino group from a cytosine base, turning it into a uracil. Therefore it changes a C:G base pair into a U:G mismatch within DNA. Harris and Liddament provided a model of the DNA deamination reaction within the potential catalytic centre of AID where the histidine and cysteine residues within this catalytic centre likely coordinate a zinc ion that is necessary for its catalytic activity.\textsuperscript{51} The glutamic-acid residue further helps to produce the hydroxide ion that is required for amine-group removal. Due to its DNA modifying activity, AID is a potential threat for genomic integrity. Several mechanisms therefore exist to restrict AID’s activity in the genome. For example, its nuclear localisation signal (NLS) and nuclear export signal (NES) strongly restricts AID’s presence in the nucleus.\textsuperscript{52-53} As a consequence AID predominantly locates within the cytoplasm (>99\%).\textsuperscript{52-55} In addition, AID prefers to deaminate ssDNA within a preferred DNA sequence hotspot (WRCY motifs, W=adenine or thymine, R=purine, C=cytosine, Y=pyrimidine, or the inverse RGYW G=guanine) only after activation by PKA phosphorylation.

AID induces DSBs to enable CSR

B cells of the germinal center can change their antibody isotype (from IgM to IgA, IgE or IgG) by CSR (Figure 4). CSR requires the activation of B cells and defined external stimuli. Depending on the signals, specific intronic (\(I\)) promoters become activated which are located upstream of highly repetitive and conserved switch (\(S\)) regions. For example, when B cells are activated with LPS and IL-4, both the \(I_{\mu}\) and \(I_{\gamma 1}\) promoters become activated and subsequently the \(S_{\mu}\) and \(S_{\gamma 1}\) region become highly transcribed. Similarly, stimulation with IL-5 will elevate the expression of \(S_{\mu}\) and \(S_{\alpha}\). As a consequence of high transcriprional activity, single-stranded DNA is highly exposed in two independent switch regions which enable Replication Protein
A (RPA) to bind and stabilize it. RPA interacts with activation-induced cytidine deaminase (AID), which is highly expressed in B cells of the germinal center. Upon activation of AID by PKA mediated phosphorylation, AID can deaminate cytosine into uracils (Figure 5A). AID deamination preferentially occurs within ‘hot spots’ matching the RGYW/WRCY motif (R is G or A, Y is T or C, and W is A or T). The uracils generated can be hydrolyzed from the sugar phosphate backbone by the uracil-DNA glycosylase, thereby leaving apyrimidinic (AP) sites. The AP sites are recognized by AP endonucleases that break the phosphodiester bonds at the AP sites. If these processes occur on opposite strands, a DSB is generated in a S region. When two DSBs are generated in two individual switch regions, the intervening DNA between those S-regions is deleted as a switch circle which enables the substitution of Cμ with one of the downstream constant gene segments (Cγ, Cα or Cε) (Figure 4). The free ends of the DNA are rejoined by C-NHEJ or A-EJ to link the variable domain exon (VDJ segment) to the desired downstream constant domain exon of the antibody heavy chain. As a result, the VDJ segment is linked adjacent to a downstream constant region. CSR increases the flexibility of the humoral immune response as it allows to exploit the different capacities of immunoglobulin’s to activate appropriate effectors mechanisms.

Aid induces somatic hypermutation in gc b cells

Besides CSR, AID is essential for the induction of SHM. SHM enables B cells to generate point mutations in IgV regions and eventually change the affinity of the antibody for the cognate antigen. During SHM, AID deaminates cytosine to uracil within the V(D)J segment...
of IgH and IgL chain genes (Figure 4). As a consequence of deamination, a C:G basepair is converted into a U:G mismatch (Figure 5A). This lesion is normally repaired efficiently by the base-excision repair (BER) or mismatch repair (MMR) pathway. To establish mutations from this primary lesion, these error-free repair pathways are overruled by error-prone mutator pathways during SHM of Ig genes. If the U:G mismatch is not repaired, replication of the uracil instructs a template T and thereby generates transitions (C/G to T/A). If the uracil is processed by the uracil-DNA glycosylase (UNG) an abasic site is generated and error-prone replication by damage tolerant translesion synthesis polymerases can generate transitions and transversions (C/G to T/A and C/G to G/C or G/C to A/T). Alternatively, the uracil is detected as a U:G mismatch by the mismatch recognition system and a gap is formed around the initial U:G mismatch. Error-prone gap filling by error-prone translesion synthesis polymerases generate predominantly A/T mutations around the initial lesion. B cells of the GC can acquire AID dependent point mutations in their V(D)J regions at a rate of about $10^{-3}$ per generation, which is six orders of magnitude higher as compared to spontaneous mutations in house keeping genes. The introduction of these point mutations may ultimately culminate in the survival of B cells which generate antibody variants that are of higher affinity for the antigen. Within the germinal center, B cells with highest affinities for the antigen are selected and allowed to differentiate into antibody producing plasma cells or long-lived memory B cells. The latter contributing to an effective immune responses upon a recall infection.
**Alternative functions of aid**

As AID has been shown to be also expressed in non-B cells (albeit at lower levels) it may exert other functions besides inducing SHM and CSR. In line with this notion, AID has been identified as the founding member of the APOBEC gene family of cytosine deaminases. The APOBEC family of enzymes is an effective arm of innate immunity to viral elements. With the exception of RNA editing enzyme APOBEC-1, members of this family modify cytosine within viral DNA to control retro-elements for instance by retroviral hypermutation. Evolution appears to have tailored this defense system of innate immunity, to improve adaptive immunity by enabling SHM and CSR of Ig genes. Besides its contribution to immune responses, AID has been proposed to be involved in active DNA demethylation. After deaminating 5-methylcytosine, a T:G mismatch results (Figure 5B). This T:G mismatch is proposed to be subsequently processed by Mbd4 glycosylase. The resulting gap is finally filled by an unmethylated cytosine. However, the recent discovery and characterization of TET enzymes provided an alternative mode of action for active DNA demethylation, which has challenged the relevance of AID in this process.

**AID targeting in Ig genes and lymphogenesis**

The generation of AID-induced DSBs during CSR enables isotype switching which improves the humoral immune response to increase the organism fitness. At the same time, the generation of DSBs is unfavorable as these lesions, if not repaired, can be highly genotoxic. For instance, chromosomal translocations can be provoked by those DSBs and thereby CSR can risk the development of non-Hodgkin lymphomas. Indirect evidence supporting a role for AID in lymphomagenesis was based on characterization of a number of recurrent translocations in defined non-Hodgkin lymphomas. These studies revealed switch regions, the prime AID target, as frequent translocation partners. More direct evidence for a critical role of AID in the generation of translocations came from in vitro studies showing that AID promotes chromosomal translocations involving the proto-oncogenes c-myc and the S regions in mouse splenic B cells stimulated with LPS and IL-4. Furthermore, studies using an IL-6 transgenic mouse model showed that AID is required for the generation of IL-6 induced cMyc-IgH translocations in hyperplasic lymph nodes, which can lead to plasmacytomas. AID was also shown to be required for c-myc/IgH chromosome translocation in vivo. Besides inducing DSB, a AID mutation signature was recently found in the immunoglobulin locus of many B cell malignancies. Taken together, there is ample evidence that AID increases chromosomal translocation risk by inducing DSBs specifically in Ig loci.

**Aberrant targeting of aid and lymphomagenesis**

Besides specific targeting to Ig loci, it has been proposed that AID off-targeting further contributes to the exceptional genomic instability of germinal center B cells. Robbiani et al. were the first to provide evidence for the view that AID is required for the c-Myc DNA breaks required for IgH-c-Myc translocations. In addition, AID has also been implicated to induce many other unintentional DSBs throughout the genome, thereby increasing the chromosomal
translocation risk.\(^{18,83}\) Besides the initiation of DSBs, AID is thought to induce somatic point mutations in many non-Ig genes. This concept was fueled by the intriguing observation that the *Bcl6* proto-oncogene is mutated in GC and post-GC B cells of healthy individuals. Of note, the pattern of mutations found in *Bcl6* appeared similar as was found in hypermutated Ig genes.\(^{84-85}\) Subsequent studies further suggested that AID may initiate mutations in a number of non-Ig genes.\(^ {60,86-88}\) In contrast to the view of unintentional off-targeting, others proposed that AID functionally binds to Spt5 at stalled promoters throughout the genome.\(^ {89}\) A genome-wide function of AID has also been proposed in primordial germ cells, where AID was proposed to be involved in active DNA demethylation in gene bodies.\(^ {90}\) Although the extent by which AID targets throughout the genome is unclear, the mutagenic potential of AID critically clearly depends on its expression levels. Overexpression of AID causes a mutator phenotype in *E.coli*, hybridomas, fibroblasts, and AID transgenic mice develop T cell lymphomas and adenomas by introducing point mutations in non-Ig genes and oncogenes.\(^ {91}\) that in normal B cells are not subject to SHM.\(^ {92-94}\)
Aim of the project and its implications

Mature B cells are highly predisposed to neoplastic transformation. This project was initiated to improve our understanding of the underlying molecular factors contributing to their exceptional genetic instability. Given the mutagenic potential of AID, revealing the specificity of AID targeting throughout the genome became central to understand the neoplastic predisposition of B cells. A chromatin binding profile of AID may provide potential hotspots of AID-induced mutagenesis. Besides subtle genetic alterations like point mutations, hotspots for DSBs in non-Ig genes can be provided. The latter may provide novel insights in the origins of chromosomal translocations in B cells, and the role of AID in initiating these lesions. Given the relevance of chromosomal translocations in the generation of defined B cell malignancies, this project also aimed to identify AID-independent translocation risk factors.

Approach: integrative analyses of ngs-based technologies

At the onset of this project, studies on AID were restricted to specific genes, leaving the genome-wide impact of AID unknown. Technologies involving next-generation sequencing (NGS) appeared indispensable to define the genome-wide impact of AID. NGS-based techniques are powerful tools to unravel biological processes. However, as these NGS-based techniques generate huge data sets, their proper analyses are most complex, challenging and time-consuming. Nevertheless, to approach this project, we designed novel bioinformatics algorithms mainly using the programming language ‘R’. In addition we integrated publicly available omics data with novel experimental data to test our hypotheses. We integrated RNA-Seq, ChiP-Seq, MethylCap-Seq, 4C, HTGTS, TC-Seq, DamID and Dam-profiling all obtained from the B cell system.

Thesis outline

This project was initiated to define risk factors for genomic instability of B cells. We are the first to challenge the widely held belief that AID destabilizes the B cell genome by binding and deamination of cytosines throughout the genome. In chapter 2 we question previous conclusions on genome-wide AID targeting obtained with a ChiP-Seq approach on switch activated B cells. The authors concluded that AID preferentially binds ~6000 genes and generically binds to stalled promoters. However, applying ChiP-Seq to identify genomic targets of AID is not trivial, given that AID is not a true chromatin-associated protein. Most AID, which is encoded by the Aicda gene, is retained in the cytosol and only a small portion of the AID pool is actively shuttled between the cytoplasm and nucleus. Of the nuclear fraction, only a small proportion is expected to interact with chromatin. Hence, extensive controls are needed to ensure that the measured signal derives from a specific enrichment of AID bound chromatin in Aicda+/+ B cells. Based on our reanalysis we conclude that the ChiP-Seq experiment provided did not reveal AID binding and therefore did not support the authors’ conclusions.

Given the proposed role of AID in active DNA demethylation, we assessed a potential role of AID as an epigenetic eraser and transcriptional regulator in chapter 3. Well-controlled RNA-Seq experiments involving different B cell subsets revealed that Aicda−/− B cells are
developmentally affected. However as shown by RNA-Seq, MethylCap-Seq, and SNP analysis these transcriptome alterations unlikely relate to AID, but alternatively to a CBA mouse strain derived region around the targeted Aicda locus. Unexpected confounding parameters were revealed, which provide alternative, AID-independent interpretations on genotype-phenotype correlations previously reported in numerous studies on AID using the Aicda\(^{-/-}\) mouse strain.

In chapter 4 we assessed the degree by which genome-wide studies were controlled for the potential confounding variables identified between Aicda\(^{-/-}\) and Aicda\(^{+/+}\) B cell subsets. To test this in an unbiased manner, we developed a control-checklist for studies involving microarray- and/or NGS-based technologies. After applying this checklist on studies proposing genome-wide targeting and/or activity of AID we conclude that evidence for extensive genome-wide targeting of AID is lacking.

Since whole-genome targeting by AID appears irrelevant to genomic instability of B cells, our investigations on chromosomal translocations turned to experiments unrelated to AID. While transcription has long been associated with chromosomal translocations, the exact nature of its contribution remains unknown. In chapter 5 we integrated whole-genome Dam-profiling, RNA-Seq and High-Throughput Genome-wide Translocation Sequencing (HTGTS) data from the B cell system to distinguish between direct, transcription-related effect (gene activity) and indirect, transcription-associated effects (gene access).\(^{42}\) Based on our analyses, we conclude that beyond the formation of a primary DSB, gene access and not transcription per se determines the translocation risk of active genes. Hence, we are the first to explain why expressed genes become frequently involved in translocations.

Chapter 6 provides a critical viewpoint on the origin of chromosomal translocations. Accordingly, the role of AID in inducing translocations in B cells is reassessed. The frequent involvement of Ig loci in chromosomal translocations of B cell malignancies emphasizes the oncogenic potential of AID-induced DSBs in Ig genes. However, a mutagenic role for AID outside Ig loci appears rather limited. We propose that chromosomal translocations are predominantly initiated by DNA-ends from a single DSB, which in switch-induced B cells are scheduled in the immunoglobulin IgH locus.

REFERENCES


10 Ueda C, Akasaka T, Kurata M, Maesako Y, Nishikori M, Ichinohasama R, et al. The gene for interleukin-21 receptor is the partner of BCL6 in t(3;16)(q27;p11), which is recurrently observed in diffuse large B-cell lymphomas. Oncogene. 2002 Jan 17;21(3):368-76.


29 Natarajan AT, Obe G, van Zeeland AA, Palitti F, Meijers M, Verdegaal-Immerzeel EA. Molecular mechanisms involved in the production of chromosomal aberrations. II. Utilization of Neurospora endonuclease for the study of aberration production by X-rays in G1 and G2 stages of the cell cycle. Mutat Res. 1980 Feb;69(2):293-305.


