

Supplemental Data

Human Chromosomal Translocations at

CpG Sites and a Theoretical Basis for

Their Lineage and Stage Specificity

Albert G. Tsai, Haihui Lu, Sathees C. Raghavan, Markus Muschen, Chih-Lin Hsieh, and Michael R. Lieber

SUPPLEMENTAL INTRODUCTION

Overview of Recurrent Translocations in Leukemias and Lymphomas

A tumor's cell of origin has great clinical importance, and therefore, hematopoietic tumors are classified by lineage and stage (Jaffe et al., 2001). However, it also indicates which oncogenic processes have occurred or are occurring. Myeloid neoplasms are categorized as chronic myeloproliferative diseases (CMPD), myelodysplastic/myeloproliferative diseases (MDS/MPD), myelodysplastic syndromes (MDS), and acute myeloid leukemias (AML). Lymphoid neoplasms are categorized as B-cell, T and NK-cell, and Hodgkin lymphoma (HL), but for the purposes of this discussion will be divided into immature (pre-B or pre-T stage) or mature malignancies.

Chronic myelogenous leukemia (CML) and acute myelogenous leukemia (AML) are the primary myeloid neoplasms, with incidences of ~1-1.5/100,000 people per year and 3/100,000 people per year, respectively (Jaffe et al., 2001). Ninety to ninety-five percent of CMLs feature the Philadelphia chromosome, t(9;22)(q34;q11) BCR-ABL translocation, as do some AMLs and ALLs. CMLs can progress to AMLs or ALLs through a blast crisis. BCR-ABLs in CMLs tend to involve the major breakpoint cluster region (M-BCR) while those in primary ALLs tend to involve the minor breakpoint cluster region (m-BCR). In AMLs, the most common translocations are the t(8;21)(q22;q22) AML1-ETO in 15% of cases, and the t(15;17)(q22;q12) PML-RAR α in 5%. Translocations involving the MLL locus on 11q23 are less common overall, but occur in 50% of infant AML.

Pre-B and pre-T cancers usually take the form of acute lymphoblastic leukemia (ALL)/acute lymphoblastic lymphoma (LBL), with an overall incidence roughly similar to CML (Jaffe et al., 2001). They are divided into B-ALLs, with B-cell origin and surface immunoglobulin; T-ALLs, with a T-cell origin and surface T-cell receptor; and common ALLs, with a nonmarking lymphoblast or pre-B origin and rearranged immunoglobulin heavy chain. The most common rearrangements found in B-ALL and pre-B ALL are the t(12;21)(p12;q22) TEL-AML1 (also known as ETV6-RUNX1) in 20% of pre-B ALLs, then the t(1;19)(q23;p13) E2A-PBX1 in 5%, then the various MLL translocations in 5%, which include t(4;11)(q21;q23) MLL-AF4 and t(9;11)(p22;q23) MLL-AF9. MLL translocations are especially frequent in infant ALL, present in 85% of cases. TEL-AML1 and AML1-ETO translocations involve the same AML1 gene on 21q22, but of the sequenced breakpoints, the closest are 28 kb apart. Many normal T lymphocytes feature inv(7)(p15;q35), an inversion of chromosome 7 between the TCR β and TCR γ loci. Other than this, the most common rearrangement found in T-ALLs is the SCL-SIL interstitial deletion, del(1)(p32), in about 25% of cases. The interstitial p16 deletion del(9)(p21) occurs in a small percentage of both T-ALLs and pre-B ALLs / B-ALLs. Less common are SCL-TCR, LMO2-TCR, HOX11-TCR, and TTG1-TCR translocations.

Worldwide, mature lymphoid cancers constitute greater than 90% of lymphoid neoplasms and 4% of all cancers, and greater than 6% of all cancers in the United States (Jaffe et al., 2001). They are divided into Hodgkin lymphoma (HL) and non-Hodgkin lymphomas (NHL), with the latter composing about 90% of the group. HL appear to be mostly activated B-cells but recurrent

translocations have not been reported. About 85% of NHL are B-cell lymphomas, the major types with recurrent translocations being diffuse large B-cell lymphoma (DLBL) in 30.6% of cases; follicular lymphoma (FL), 22.1%; marginal zone lymphoma (MZL), 9.4%; mantle cell lymphoma, 6.0%; Burkitt lymphoma, 2.5%; and anaplastic large cell lymphoma (ALCL), 2.4%.

Strikingly, the t(14;18)(q32;q21) bcl-2 translocation is found in about half of NHL, including 80-90% of FL and 20-30% of DLBL (Vega and Medeiros, 2003). About 50% of bcl-2 breakpoints occur in the 175 bp major breakpoint region (MBR) in the exon encoding the 3' UTR located on the centromeric side of the bcl-2 oncogene; 13% in the 105 bp intermediate cluster region (icr) about 18 kb centromeric to the MBR; and 5% in the 561 bp minor cluster region (mcr) about 11 kb further centromeric to the icr (Weinberg et al., 2007). The t(11;14)(q13;q32) bcl-1 translocation is found in almost all MCL, with about 30% of breakpoints in the 150 bp major translocation cluster (MTC) about 120 kb telomeric to the bcl-1 oncogene (Bertoni et al., 2004). The remaining 70% are spread 100 kb on either side of the MTC, with two minor cluster regions each constituting a few percent. Both of these translocations join their respective oncogenes to the immunoglobulin RSS at a J segment on 14q32, juxtaposing them with the strong immunoglobulin heavy chain μ enhancer.

Almost all Burkitt lymphomas contain the t(8;14)(q24;q32) myc translocation. Ten to twenty percent of DLBLs contain the t(3;14)(q27;q32) bcl-6 translocation. Both of these translocations join their respective oncogenes to the repetitive immunoglobulin switch sequences on 14q32. An additional 5-10% join bcl-6 to various other loci. In MZL, the most common translocation is t(11;18)(q21;q21) API2-MALT1 in 30% of cases, though few junctions are available. A quick assessment of them revealed no CpG proximity and a lack of nucleotide additions. The most common translocation in T-cell lymphomas is the t(2;5)(p23;q35) ALK-NPM1 in 40-70% of anaplastic large cell lymphoma (ALCL), for which no translocation junctions are available (Vega and Medeiros, 2003).

SUPPLEMENTAL RESULTS

CpG Hotspots are not General to all Translocations

By far, the most common rearrangement known in pre-B lymphoblastic leukemia is the t(12;21)(p12;q22) TEL-AML1 (ETV6-RUNX1) translocation. The 53 TEL breakpoints analyzed here spread across 12.9 kb, only one occurring at a CpG ($p>0.8$), with an overall average of 41.1 bp to the nearest CpG, close to the 41.6 bp if randomly distributed ($p>0.6$). The 53 AML1 breakpoints are more dispersed, occurring over a 165 kb region, with only one occurring at a CpG ($p>0.8$), and an overall average of 61.7 bp to the closest CpG, virtually identical to the 61.7 bp if randomly distributed ($p>0.4$). Thus, neither TEL nor AML1 breakpoints appear to involve CpG hotspots.

The second most common rearrangement in pre-B lymphoblastic leukemia is the E2A-PBX1 translocation analyzed previously. Translocations involving the MLL locus on 11q23 are the third most common, and adjoin MLL with more than 20 partners across many chromosomes in cases of AML and ALL. The collection of 350 MLL breakpoints – 24 from primary AML, 291 from primary ALLs, and 35 leukemias secondary to topoisomerase II-inhibitor treatment – are mostly cases of t(4;11)(q21;q23) MLL-AF4 and t(9;11)(p22;q23) MLL-AF9, and fall within an 8.6 kb window on MLL. Of the primary AMLs, none occur at CpGs ($p=1$), averaging 64.4 bp to the nearest CpG, compared to 85.4 if randomly distributed ($p>0.5$). Of the primary ALL breakpoints, 12 occur at CpGs ($p>0.1$), averaging 79.0 bp to the closest CpG overall, not too dissimilar to the 84.9 bp if randomly distributed ($p>0.4$). Analysis of the B-lineage subset of primary ALLs also yielded no significant CpG proximity, as did the secondary leukemias. Thus, MLL breakpoints do not appear to involve CpG hotspots.

BCR-ABL translocations t(9;22)(q34;q11) result in Philadelphia chromosomes seen in chronic myelogenous leukemias (CML). All 35 BCR breakpoints collected fall into the 2.8 kb major breakpoint cluster region (M-BCR). Zero occur at CpGs ($p=1$), with an average of 28.7 bp to the nearest CpG, compared to 21.4 if distributed randomly ($p>0.9$). Breakpoints on ABL scatter over more than a megabase and often fall into repetitive DNA, thus not all could be located. A 125 kb window containing 25 of the 27 locatable breakpoints was used for analysis.

Again, none of the breakpoints occur at CpGs ($p=1$), with an average of 63.0 bp to the closest CpG, versus 50.3 if randomly distributed ($p>0.8$). Thus, neither BCR nor ABL breakpoints in CMLs appear to involve CpG hotspots. BCR-ABL translocations in primary ALLs (not including ALLs resulting from CML blast crisis) appear to behave similarly, though only nine junctions are available – too few for a decent analysis.

The t(8;14)(q24;q32) c-myc-IgH switch translocation in sporadic Burkitt lymphoma is thought to occur as an error of CSR in mature B cells. The mechanism of double-strand breakage at the highly repetitive switch regions of 14q32 is an area of intense investigation, while the mechanism at c-myc is speculated to be similar but unknown. The 125 breakpoints found on c-myc span 4.1 kb, averaging 8.8 bp to the nearest CpG, compared to 8.9 bp if randomly distributed ($p>0.7$). It is interesting to note that c-myc has a higher CpG density (244 CpGs in 4,099 bp, or 16.8 bp/CpG) than either the bcl-2 MBR (5 CpGs, 150 bp, 35 bp/CpG) or bcl-1 MTC (7 CpGs, 150 bp, 21.4 bp/CpG), yet only 17% (14 breakpoints) occur at CpGs ($p>0.5$) compared to 43% for the MBR and 37% for the MTC. Thus, c-myc breakpoints do not appear to involve CpG hotspots.

Similarly, the t(3;14)(q27;q32) bcl-6-IgH switch translocation joins the first intron of the bcl-6 gene to the switch regions of 14q32 in a variety of mature B-cell lymphomas. The 2 kb covering the 37 breakpoints obtained is, like c-myc, very CpG-rich (2045 bp, 102 CpGs = 20.0 bp/CpG). However, only seven breakpoints occur at CpGs ($p>0.2$), and the overall average distance to the nearest CpG is 8.3 bp, as opposed to 10.9 bp if randomly distributed ($p>0.1$). Thus, bcl-6 breakpoints do not appear to involve CpG hotspots.

Due to the high density of CpGs in the c-myc and bcl-6 regions, a relatively high proportion of breakpoints occur at CpGs just by random chance, but that proportion is not any *greater* than would be expected by random chance ($p>0.5$ with $N=125$ at c-myc, $p>0.13$ with $N=37$ at bcl-6). In contrast, bcl-2 MBR and bcl-1 MTC breakpoints occur at CpGs *far* more often than would be expected by random chance ($p<10^{-41}$ with $N=487$ at the MBR, $p<10^{-8}$ with $N=104$ at the MTC). Translocations to c-myc and bcl-6 occur in functionally-important regions, suggesting that they cluster due to these functions rather than to the CpGs. The c-myc translocation region includes the promoter and CpG island, while the bcl-6 translocation region includes conserved portions of the gene and a small CpG island.

When we measure distance to the closest CpG, we observe that bcl-2, bcl-1, and E2A breakpoints are much closer to CpGs than expected according to a random distribution. On average, bcl-2 breakpoints occur 7.2 bp away from the closest CpG, versus 59.1 bp if randomly distributed; bcl-1 breakpoints average 3.1 bp away, versus 45.3 bp if random; and E2A breakpoints average 5.1 bp away, versus 29.1 bp if random.

These findings are best illustrated by plotting the proportions of breakpoints at various distance intervals from CpG, and comparing breakpoints sequenced from human cells with a random distribution (where breakpoints are uniformly-distributed throughout the region) (Fig. S3). In human cells with translocations involving bcl-2, bcl-1, and E2A, the highest percentage of breakpoints occurs at a CpG site, that is, at a distance of 0 bp from CpG. And as one moves farther and farther away from CpG, one finds fewer and fewer breakpoints. In contrast, relatively few breakpoints are found at CpGs when they are uniformly-distributed throughout the region, and the highest proportion occur at distances >8 bp from CpG. A similar analysis with CAC, which is the minimal motif to support V(D)J recombination, reveals a pattern more consistent with random distribution (Fig. S3). That is, bcl-2, bcl-1, and E2A breakpoints sequenced from human cells do not focus around CAC in the same way they do about CpG. A summary of these results is shown in Table S1.

T Cell and Some Pre-B Cell Rearrangements Tend to Use a V(D)J-Type Rather than a CpG-Type Rearrangement Mechanism

In normal V(D)J recombination at the immunoglobulin and TCR loci, the RAG complex binds an RSS and makes a nick 5' of CAC. After synapsis with a partner RSS, they use the free hydroxyl group at the nick to attack the opposite strand, thereby creating a double-strand break – with one hairpin DNA end and one blunt DNA end – at each RSS. The blunt ends containing the RSS are termed signal ends, later joined by the ligase IV:XRCC4:XLF complex without loss to form a signal joint. The hairpin ends, also called coding ends, because they form the coding

sequence of the immune receptor, are opened and endonucleolytically recessed by Artemis:DNA-PK_{CS} before terminal deoxynucleotidyl transferase (TdT) adds nucleotides template-independently and the ligase IV:XRCC4:XLF complex joins them together. Recessing generally follows a distribution which falls with distance from the free end, and is usually less than eight nucleotides in human cells (Gauss and Lieber, 1996). Therefore, the coding joint breakpoint plot shows a large spike 5' of CAC which falls as one looks further 5'.

The most common T-ALL rearrangements are the intrachromosomal deletions, del(1)(p32) SCL-SIL and del(9)(p21) Δ p16. p16 deletions also occur in B-ALLs. While the furthest SCL-SIL breakpoints are 89 kb apart, 204 of 209 breakpoints map to one of three CACs (two shown in Figs. 1D and 1E). 35% of SCL-SIL breakpoints occur at CAC ($p < 10^{-74}$), with an average distance of 4.5 bp from the nearest CAC, compared to 40.4 bp if breakpoints were randomly distributed ($p < 10^{-107}$). CpG scores significantly due to proximity to these CACs ($p < 10^{-14}$) but not nearly on the same level. Individually analyzed, each of the three breakpoint peaks is also most significant for CAC(A), and all but one of the 209 breakpoints are compatible with a typical coding joint recombination. Interestingly, the one exception occurs at a CpG. Thus, SCL-SIL deletions appear to be almost entirely V(D)J-type events.

More than seven widely-spaced cRSS spanning 231 kb are used in 41 breakpoints from p16 deletions (Kitagawa et al., 2002). They average 11.3 bp to the nearest CAC, compared to 46.0 bp if randomly distributed ($p < 10^{-13}$), with seven occurring at a CAC ($p < 10^{-5}$). We find three breakpoints at CpG ($p = 0.13$) and an average distance of 63.7 bp to the nearest CpG, versus 96.7 bp if randomly distributed ($p = 0.05$). Individually analyzed, 28 breakpoints are compatible with V(D)J-type coding joint events but not CpG-type; three are compatible with CpG-type but not V(D)J-type; seven with both types; and three with neither. Thus p16 deletions appear to be V(D)J-type events. Separate analysis of the B-ALL subset results in the same conclusion.

The SCL locus also translocates to the TCR loci in a small percentage of T-ALLs, but none of the 20 translocation breakpoints we obtained use the same sites as SCL-SIL deletions. They span 64 kb, with five at CAC / GTG ($p = 0.0007$), and average 10.2 bp away, compared to 14.6 bp if randomly distributed ($p = 0.04$). A small cluster of seven breakpoints centered around a CACA is likely to be V(D)J-mediated, with two precise signal joints and three precise coding joints. However, CpG is more significant by measures of proximity, averaging 32.1 bp away, compared to 51.7 bp if randomly distributed ($p = 0.001$), but only four breakpoints occur at CpG ($p = 0.01$). This is complicated by the fact that, at many of the relevant CACs, CpGs are found close by. A breakpoint-by-breakpoint analysis indicates only nine of the 20 could be V(D)J-type and six are incompatible with either V(D)J-type or CpG-type breakage. The most significant motif ends up being CACA/TGTG; but as a group, SCL translocations cannot be confidently assigned as any one type and could be a mixture which includes random-type breakage.

A small percentage of T-ALLs translocate one of the TCR loci to LMO2, with breakpoints spanning over 40 kb. Of 31 breakpoints, nine occur at CAC/GTG ($p < 10^{-6}$), overall averaging 4.58 bp to the closest CAC / GTG, versus 16.3 bp if randomly distributed ($p < 10^{-8}$). CpG is somewhat significant, with breakpoints averaging 26.2 bp to the nearest CpG, compared to 47.1 bp if randomly distributed ($p = 0.04$), but only three occur at CpG ($p = 0.16$). Individual breakpoint analysis indicates 7 breakpoints compatible with signal joints but not CpG, 11 with coding joints but not CpG, seven with CpG but not V(D)J, three with both, and three with neither. On the whole, LMO2 translocations appear to be V(D)J-type recombination events, though a third or more could be other types.

The t(10;14)(q24;q11) HOX11-TCR and t(11;14)(p15;q11) TIG1-TCR translocations appear to represent more examples of cRSS recombination. Directly 5' of the main HOX11 CACAG is the sequence CGCG, complicating the analysis; but in TIG1, the closest CpG to the CACAGTG is more than 20 bp away.

Graphs of the breakpoint distributions at various distance intervals confirm these observations (available upon request as a resource). A summary of these results is shown in Table S1.

Coexistence of Multiple Distinct Breakage Mechanisms

p16 deletions in pre-B ALLs seem to use a cryptic V(D)J-type mechanism, like their pre-T counterparts, indicating that these mechanisms coexist with, but are dwarfed by the CpG-type

mechanism at *bcl-2* and *bcl-1*, which occur at the pre-B stage. Ikaros deletions in ALLs containing BCR-ABL translocations also appear to use a cryptic V(D)J-type mechanism (Mullighan et al., 2008). V(D)J-type and CpG-type hotspots appear to be very special cases where an active breakage mechanism creates double-strand breaks within small windows of specific regions in the genome.

Even within one translocation, the break on one chromosome may occur by one mechanism and the break on the other chromosome by another mechanism. This is evidenced by CpG-type breakages at *bcl-1* and *bcl-2* joining to V(D)J-type breakages at the heavy chain locus, and experimentally by I-SceI restriction cuts with V(D)J-type and class switch-type *ex vivo* and *in vivo*, respectively (Weinstock et al., 2007; Zarrin et al., 2007).

Unlike the interstitial deletions at SCL-SIL and p16, a large proportion of the SCL and LMO2 translocations in pre-T cells do not use V(D)J-type breakage – even if they join to the RSS at the TCR loci. CpG was at least marginally significant in these two rearrangements, thus CpG-type breakage might account for some of them. However, a remaining fraction escape either the V(D)J-type or CpG-type classification and for now are assumed to be random-type breakage.

We postulate that this “any-break-to-any-break” joining also occurs in the E2A-PBX1 translocation, where breakpoints at the E2A cluster at two CpG hotspots within a 17 bp region of a 3.2 kb intron, but those on PBX1 are distributed across 107 kb without CpG or CAC propensity, presumably due to random-type breakage.

Alternate Hypotheses for CpG Hotspots, Lineage- or Stage-Specificity, and Other Clustering Effects

Our proposed theory invokes the RAG complex for two primary reasons: (1) CpG-type rearrangements only occur in the developmental stage during which the RAG complex is expressed – not in the previous stage, and not in the following – and, (2) the width and shape of CpG-type clusters is consistent with the structure-specific endonuclease activity of the RAG complex.

Our proposed theory invokes AID for two primary reasons: (1) CpG-type rearrangements only occur in the cell lineage where AID is expressed, and (2) the enzymatic activity of AID is cytidine deamination – which is also the basis for point mutations at CpGs in cancers and CpG suppression through evolutionary time.

In the spirit of Occam’s razor, many if not most of the plausible alternatives will be derivatives of our own, because the RAG complex and AID each independently explain two aspects of CpG-type hotspots.

Without the RAG complex, one is hard-pressed to explain how AID alone can determine stage-specificity without some additional, and likely ad hoc, factor. After all, AID is most highly expressed in the mature/activated B-cell stage; thus if AID were sufficient, one would expect to see a higher level of CpG hotspots in the *c-myc* and *bcl-6* translocations, rather than the baseline random level we actually observe. One also loses the benefit of its structure-specific nuclease activity.

Without AID, one is hard-pressed to explain how RAGs alone can determine lineage-specificity without some additional, and likely ad hoc, factor. One also loses the benefit of its cytidine deaminase activity.

Unknown nucleases and deaminases and other such components are always possibilities, and can only be ruled out by direct experimentation. However, they would have to impart stage-specificity, lineage-specificity, and/or have a structure-specific activity. As it stands, AID and the RAG complex are the most logical candidates.

Many accessory factors exist which have been proposed to promote DSBs and/or DSB clustering in other rearrangements, such as replication-associated damage, transcription, and chromatin accessibility and nucleosome positioning.

Because RAG complex activity is high during the G_0 and G_1 phases and low during the G_2 , S, and M phases, it is thought that V(D)J recombination occurs mainly during G_0 and G_1 (Schlüssel et al., 1993). Since the *bcl-2* and *bcl-1* translocations join the *bcl-2* and *bcl-1* loci to V(D)J-generated DSBs, it is most likely that they occur during G_0 and G_1 , not during S phase when replication occurs. Replication-associated damage, per se, does not impart any stage-, lineage-, or CpG specificity.

While *bcl-2* translocations occur in a transcribed region of the genome, there are no identified transcribing units around the *bcl-1* MTC, despite extensive searches during the time when the identity of the *bcl-1* oncogene, now known to be *CCND1* or the gene for cyclin D1, was unknown (Rabbitts et al., 1988). *CCND1* is more than 100 kb telomeric to the MTC and transcribes in the opposite direction. If a more proximal gene were present, it should have been discovered due to juxtaposition with the strong immunoglobulin μ enhancer after translocation. On the UCSC genome browser, the closest gene centromeric to the MTC is more than 100 kb away as well. Moreover, there is no known case of or reason for transcription predisposing to DSB at CpGs versus sites a few bases away. The known cases of transcription-associated DSBs – immunoglobulin switch regions in CSR – have kilobases-long break windows, not the 150 bp windows of the MBR or MTC.

Accessible chromatin structure is likely required for enzymes to break and/or repair the DNA at the MBR and MTC. However, nucleosome positioning is not known to be so precise as to determine accessibility down to a few base pairs. Again, there is no known case of or reason for chromatin predisposing to DSB at CpGs versus sites a few bases away.

Value of the Database as a Resource

Our database of breakpoints, mapped to the hg18 build of the human genome and annotated with junctional and clinical data, will be made available in a computer-readable format. This will allow others to repeat and extend our work, and rapidly test their own hypotheses on real patient data in silico before devoting laboratory resources and time to experiments.

Additionally, we will provide general algorithms for statistical assessment of the breakpoint data, in the hopes that a standard and logical set of statistical methods will create consistency and comparability between studies.

Ideally, one should be able to use our data to ask questions such as, “Do breakpoints from topoisomerase II inhibitor-associated MLL translocations in AMLs and ALLs occur closer to topoisomerase II nick sites than expected by random chance?” or given some chromatin immunoprecipitation data, “Do *c-myc* translocation breakpoints occur closer to regions of high RNA polymerase II binding than would be expected by random chance?”

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Breakpoint References

Breakpoints from *bcl-2* translocations, many in Figures 1A-C, come from the following references: (Akasaka et al., 1998; Albinger-Hegyí et al., 2002; Aster et al., 1992; Bakhshi et al., 1985; Bakhshi et al., 1987; Cleary et al., 1986a; Cleary and Sklar, 1985; Cleary et al., 1986b; Cotter et al., 1990; Crescenzi et al., 1988; Delage et al., 1997; Dolken et al., 1996; Eick et al., 1990; Fuscoe et al., 1996; Galoin et al., 1996; Gauwerky et al., 1988; Hostein et al., 2001; Jager et al., 2000; Jenner et al., 2002; Ji et al., 1995; Kneba et al., 1995; Ladetto et al., 2003; Limpens et al., 1991; Limpens et al., 1995; Liu et al., 1994; Luthra et al., 1997; Marculescu et al., 2002a; Matolcsy et al., 1997; Ngan et al., 1989; Price et al., 1991; Rauzy et al., 1998; Roulland et al., 2003; Schmitt et al., 2006; Soubeyran et al., 1999; Summers et al., 2001; Takacs et al., 2000; Tsujimoto et al., 1985a; Tsujimoto et al., 1988; Wang et al., 1998; Willis et al., 1997; Wyatt et al., 1992; Young et al., 2008).

Breakpoints from SCL-SIL deletions, many in Figures 1D-E, come from the following references: (Aplan et al., 1990; Aplan et al., 1992a; Bash et al., 1993; Bernard et al., 1991; Breit et al., 1993; Brown et al., 1990; Janssen et al., 1993).

Breakpoints from *bcl-1* translocations, many in Figure S1A, come from the following references: (Baron et al., 1993; Degan et al., 2002; Galiegue-Zouitina et al., 1994; Meeker et al., 1991; Nyvold et al., 2007; Pott et al., 1998; Rabbitts et al., 1988; Rimokh et al., 1994; Segal et al., 1995; Stamatopoulos et al., 1999; Tsujimoto et al., 1985b; Welzel et al., 2001; Williams et al., 1993; Willis et al., 1997) and unpublished breakpoints on GenBank, DQ912821, DQ400340, DQ401134, DQ241758.

Breakpoints from E2A-PBX1 translocations, some shown in Figure S1B, are all from one reference: (Wiemels et al., 2002a).

Breakpoints from TEL-AML1 translocations come from the following references: (Andersen et al., 2001; Ford et al., 1998; McHale et al., 2003; Pine et al., 2003; Wiemels et al., 2000; Wiemels and Greaves, 1999).

Breakpoints from MLL translocations were separated into primary ALLs, primary AMLs, and therapy-related ALLs and AMLs, from the following references: (Akao and Isobe, 2000; Bizarro et al., 2007; Borkhardt et al., 1997; Chervinsky et al., 1995; Felix et al., 1999; Felix et al., 1997; Fu et al., 2005; Gale et al., 1997; Gillert et al., 1999; Gu et al., 1994; Gu et al., 1992; Hayette et al., 2000; Kuefer et al., 2003; Langer et al., 2003; Marschalek et al., 1995; Megonigal et al., 1998; Meyer et al., 2006; Negrini et al., 1993; Raffini et al., 2002; Reichel et al., 2001; Reichel et al., 1999; Reichel et al., 1998; Super et al., 1997; Vieira et al., 2006; Wechsler et al., 2003) and unpublished breakpoints on Genbank, DQ284496 to DQ284498, DQ249868, AY839727 to AY839754, AY288756 to AY288777.

Breakpoints from AML1-ETO translocations come from the following references: (Erickson et al., 1992; Wiemels et al., 2002b; Xiao et al., 2001; Zhang et al., 2002).

Breakpoints from BCR-ABL translocations in CMLs come from the following references: (Chissoe et al., 1995; de Klein et al., 1986; Heisterkamp et al., 1985; Jeffs et al., 1998; Koduru et al., 1993; Litz et al., 1993; Mills et al., 1992; Sowerby et al., 1993; Zhang et al., 1995).

Breakpoints from c-myc-IgH switch translocations in sporadic Burkitt lymphomas come from the following references: (Apel et al., 1992; Battey et al., 1983; Busch et al., 2004; Busch et al., 2007; Care et al., 1986; Dyson and Rabbitts, 1985; Gauwerky et al., 1988; Gelmann et al., 1983; Muller et al., 1995; Murphy et al., 1986; Saito et al., 1983; Showe et al., 1985; Wilda et al., 2004; Wiman et al., 1984).

Breakpoints from bcl-6-IgH switch translocations come from the following references: (Akasaka et al., 2000; Baron et al., 1993; Chaganti et al., 1998; Chen et al., 2006; Kerckaert et al., 1993; Nakamura et al., 1997; Sanchez-Izquierdo et al., 2001; Schwindt et al., 2006; Yang et al., 2006; Ye et al., 1995; Yoshida et al., 1999).

Breakpoints from lymphoid p16 deletions come from the following references: (Cayuela et al., 1997; Duro et al., 1996; Kitagawa et al., 2002).

Breakpoints from SCL translocations come from the following references: (Aplan et al., 1992b; Begley et al., 1989; Bernard et al., 1992; Bernard et al., 1993; Bernard et al., 1990; Chen et al., 1990a; Chen et al., 1990b; Fitzgerald et al., 1991; Jonsson et al., 1991; Xia et al., 1992).

Breakpoints from LMO2-TCR translocations come from the following references: (Boehm et al., 1988b; Cheng et al., 1990; Dik et al., 2007; Fischer et al., 2007; Garcia et al., 1991; Marculescu et al., 2002b; Van Vlierberghe et al., 2006; Yoffe et al., 1989) and unpublished breakpoints on GenBank, EF450258, EF450768.

Breakpoints from HOX11-TCR translocations come from the following references: (Kagan et al., 1989; Kagan et al., 1994; Lu et al., 1990; Park et al., 1992; Salvati et al., 1999; Zutter et al., 1990).

Breakpoints from TTG1-TCR translocations come from the following references: (Boehm et al., 1988a; Boehm et al., 1991; McGuire et al., 1989).

Statistical Methods

The goal of using statistics was to establish the proximity of breakpoints to CpGs. The first and most important test is an intuitive eyeball evaluation of the data. After that, p-values are calculated to reduce concerns about subjectivity. It is important to note that p-values themselves are single numbers funneled down from many complicated factors. Therefore, unfair, unnatural, and/or ad hoc treatment of these factors can bias p-values. We have tried to ensure the fairest appraisal of the data by calculating p-values by multiple different methods. The major statistical considerations are discussed below.

The first question is, what are some natural ways to measure whether breakpoints occur next to CpGs? One way is to count the number of breakpoints occurring at CpGs, the number not occurring at CpGs, and then compare this to “random.” Another way is to measure the distances between breakpoints and the closest CpGs, and then compare this to “random.” But what is “random”? “Random” might be the distribution of breakpoints we would see if we removed all the CpGs in those regions from a population of humans and waited a hundred years for them to develop lymphomas. Without such an experiment, the p-values will all be theoretical.

Most simply, “random” is depicted as entirely equal breakage at each base pair from the most 5’ to the most 3’ breakpoint in a chosen region.

This is not a difficult case to make for the bcl-2 and bcl-1 translocations, which place one of the strongest enhancers next to but not within their respective oncogenes. Breakpoints are known to occur over tens or hundreds of kilobases of apparently empty sequence and all result in essentially the same tumor. However, this assumes that breakpoints do not naturally cluster around CpG-rich regions for other reasons. To deal with this issue we further analyze the subclusters. Uniform breakage as a representation of “random” is not as good an assumption in translocations which result in fusion genes, but this is still reasonably fair. In these cases, we worry about selection issues such as stability of the fusion mRNA, critical protein binding regions, cryptic splice sites, etc. Top and bottom strands were treated separately and calculated together and separately, but because all the examined motifs were symmetrical except CAC, this is not an important factor.

Given this definition of “random,” the calculation is relatively simple. The binomial statistic groups breakpoints into those which occur at CpG and those that do not (i.e. 1 or 0). The p-value is the probability that one would see the observed number of events at CpG, or more given random chance, and is calculated as:

$$p\text{-value} = \sum_{x=k}^N \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \quad (\text{standard binomial distribution})$$

N = total number of events

k = observed number of events at motif

p = random probability of an event occurring at the motif, calculated as the number of sites at motifs divided by the total number of sites from the most 5’ to the most 3’ breakpoint

A significant p-value indicates that breakpoints occur at CpGs more often than by random chance. The main weakness of this calculation is the definition of random chance. Another potential weakness is that breakage might occur at CpGs but get recessed by Artemis:DNA PK_{CS}, in which case these CpG-caused breakpoints would not score in the binomial statistic.

To deal with this issue, we analyzed the distributions of distance from the breakpoint to the closest CpG, comparing the actual breakpoints with the simulated “random”/uniform breakpoints using two different statistical tests. “Distance” to a CpG was calculated as the number of nucleotides between the breakpoint and the closest CpG site. Distance distributions for uniformly or “randomly” distributed breakpoints were generated by traversing the region from the most 5’ to the most 3’ breakpoint, treating each strand independently, and calculating the distance to the closest CpG site for each site along the way.

The Student’s t-test compares the averages of the distributions, relative to a shared standard deviation. First, all the distances were subjected to a standard Box-Cox log(x+1) transformation to correct for the non-normality of the distributions. Next, the average and standard deviation were calculated for each set of breakpoints. Finally, the t statistic was calculated as:

$$t\text{-value} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} \quad (\text{Student's t-test})$$

\bar{x}_1 = average of log(x+1)-transformed actual breakpoint distances

\bar{x}_2 = average of log(x+1)-transformed random breakpoint distances

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

n_1 = number of actual breakpoints

n_2 = number of random breakpoints
 s_1 = standard deviation of actual breakpoints
 s_2 = standard deviation of random breakpoints

Using Excel, the t-value was converted to a p-value, which is the probability of seeing the observed average distance to CpG or smaller, given random chance. A significant p-value indicates that, on average, breakpoints are significantly closer to CpGs than would be expected by random chance. However, the t-test works best when comparing two sets of data which are normally-distributed.

Therefore, we used the Mann-Whitney U-test, which is independent of the shape of each individual distribution, to compare how actual breakpoints and random breakpoints are distributed relative to one another. First, the distances from both groups were combined and sorted from least to greatest. Next, each distance value was ranked from 1 to n_1+n_2 , where n_1 is the number of actual breakpoints and n_2 is the number of random breakpoints. In cases where there were many ranks for a particular distance, the ranks were averaged. Going back to the original groupings, the ranks for each distance value in the actual breakpoint group were summed into the value R_1 . The U statistic was calculated as:

$$U = R_1 - \frac{n_1(n_1 + 1)}{2}$$

For large samples, a z-value can be derived as:

$$z = \frac{U - m_v}{\sigma_v}$$

$$m_v = \frac{n_1 \cdot n_2}{2}$$

$$\sigma_v = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Excel was used to derive p-values from z-values. The p-value is the probability of seeing the observed sum rank or lower, given random chance. A significant p-value indicates that the actual breakpoints are nonrandomly ordered closer to CpG than the random breakpoints.

At least in our set, these distance statistics behave more or less the same and have various strengths and limitations. They work best when the motif being tested is >1% of the breakpoint window and spread evenly. Otherwise, the distance distributions can be affected by local fluctuations of motif density. If translocation to, say, a CpG island more often results in deregulation of an oncogene or tumor suppressor, then breakpoints may test significantly for CpG proximity even though frequency versus distance graphs show they are mostly far away. When breakpoints are clustered tightly, neighboring motifs may be pulled along, resulting in statistical significance, depending on the motif density. In these situations, significant p-values still indicate that breakpoints are closer to a motif than they should be given random chance – which is true – but there can be a number of different reasons for proximity, and we are mostly interested in breakage directly at those motifs. These calculations are necessarily insensitive to sequences outside the breakpoint window. If a given window is very dense for a specific motif and breakpoints gravitate towards but do not occur at the motif, these algorithms will not detect it. Widening the window will solve this, but makes the implicit and dangerous assumption that breakpoints could occur outside the window and still result in the same tumor.

There are a number of things one can do to increase the signal-to-noise ratio, such as using a binomial with a window of 5 bp, making an upper limit of 20 bp for distance measurements; while some of these may have a logical basis, we felt it best to avoid such ad hoc adjustments and arbitrary cutoffs.

RAG Nicking Assays

GST-core RAGs and oligonucleotides were purified as described previously (Yu et al., 2002). Substrates were prepared by annealing 12% PAGE-purified, 5'-labeled oligonucleotide with 1.5-fold excess of 12% PAGE-purified cold oligonucleotide(s) in TE with 100 mM NaCl in screw-cap tubes, according to standard boiling and slow cooling protocol in a beaker with ≥ 1 L water. The nicking assay was done as described previously, except that no competitor oligonucleotide was added. Briefly, substrates were incubated with purified RAG complexes and HMG1 for one hour, and the reaction stopped with EDTA, SDS, and proteinase K. Following phenol-chloroform extraction and precipitation, comparable levels of radioactivity were run on 8% denaturing PAGE. Lanes without RAGs were incubated with HMG1 alone. The oligonucleotides used for the various substrates are as follows, listed as labeled top strand-bottom strand: 12 signal, AT188-AT189; CG duplex from Fig. S3A, AT190-AT191; TA duplex, AT216-AT217; TG mismatch, AT216-AT191; TT mismatch, AT216-AT193; CT mismatch, AT190-AT193; CA mismatch, AT190-AT217; CG duplex from Fig. S3C, AT191-AT190; nick, AT191-AT218 & AT244; one-nucleotide gap, AT191-AT218 & AT240; two-nucleotide gap, AT191-AT235 & AT244; three-nucleotide gap, AT191-AT235 & AT240; one-nucleotide loop, AT191-AT192; one-nucleotide flap, AT191-AT244 & AT245; three-nucleotide flap, AT191-AT243 & AT245. The oligonucleotide sequences are as follows:

AT188,			
ggattaccgatgctggactgggtattatacccacagtgctcagagtccaacaaaacccatccctggg;	AT189,		
cccagggatgggtttgtggactctgagacactgtgggtataataaccagtccaagcatcgtaatacc;	AT190,		
ggattaccgatgctggactgggtattataccgctcagctctcagagtccaagatttacgcatccctggg;	AT191,		
cccagggatgcgtaaatctggactctgagagactgacggtataataaccagtccaagcatcgtaatacc;	AT193,		
cccagggatgcgtaaatctggactctgagagactgactgtataataaccagtccaagcatcgtaatacc;	AT216,		
ggattaccgatgctggactgggtattatactgtcagctctcagagtccaagatttacgcatccctggg;	AT217,		
cccagggatgcgtaaatctggactctgagagactgacagtataataaccagtccaagcatcgtaatacc;	AT218,		
ggattaccgatgctggactgggtattatacc;	AT235,	ggattaccgatgctggactgggtattat;	AT240,
gctcagctctcagagtccaagatttacgcatccctggg;	AT243,	accgctcagctctcagagtccaagatttacgcatccctggg;	
AT244,	cgtcagctctcagagtccaagatttacgcatccctggg;	AT245,	ggattaccgatgctggactgggtattatacc.

Methylation Analysis

pre-B cells were FACS-sorted from the bone marrow of five healthy individuals (without any evidence of leukemia or lymphoma) into 100% FCS using CD19 (BD Biosciences, Heidelberg, Germany) and VpreB (4G7) (Beckmann Coulter, Krefeld, Germany), and the genomic DNA extracted using the Qiagen blood miniprep kit. Bisulfite treatment and DNA purification were done using the Zymo Research EZ DNA Methylation Kit, according to manufacturer's directions. PCR was done using AT301, ggatgttattggttattgaggagt, and AT302, tcaaaaatcattcattcctta, for bcl-2; and AT297, ggggtgttttaagtttggtatt, and AT298, ccaatacccaaattcctta, for bcl-1. Conditions were 40 cycles of 95° C, 0:30; 56° C, 0:45; 72° C, 1:00; with an initial denaturation 95° C, 2:00 and a final extension 72° C, 5:00.

SUPPLEMENTAL REFERENCES

Akao, Y., and Isobe, M. (2000). Molecular analysis of the rearranged genome and chimeric mRNAs caused by the t(6;11)(q27;q23) chromosome translocation involving MLL in an infant acute monocytic leukemia. *Genes Chromosomes Cancer* 27, 412-417.

Akasaka, H., Akasaka, T., Kurata, M., Ueda, C., Shimizu, A., Uchiyama, T., and Ohno, H. (2000). Molecular anatomy of BCL6 translocations revealed by long-distance polymerase chain reaction-based assays. *Cancer Res* 60, 2335-2341.

Akasaka, T., Akasaka, H., Yonetani, N., Ohno, H., Yamabe, H., Fukuhara, S., and Okuma, M. (1998). Refinement of the BCL2/immunoglobulin heavy chain fusion gene in t(14;18)(q32;q21) by polymerase chain reaction amplification for long targets. *Genes Chromosomes Cancer* 21, 17-29.

Albinger-Hegy, A., Hochreutener, B., Abdou, M. T., Hegyi, I., Dours-Zimmermann, M. T., Kurrer, M. O., Heitz, P. U., and Zimmermann, D. R. (2002). High frequency of t(14;18)-translocation

- breakpoints outside of major breakpoint and minor cluster regions in follicular lymphomas: improved polymerase chain reaction protocols for their detection. *Am J Pathol* 160, 823-832.
- Andersen, M. T., Nordentoft, I., Hjalgrim, L. L., Christiansen, C. L., Jakobsen, V. D., Hjalgrim, H., Pallisgaard, N., Madsen, H. O., Christiansen, M., Ryder, L. P., *et al.* (2001). Characterization of t(12;21) breakpoint junctions in acute lymphoblastic leukemia. *Leukemia* 15, 858-859.
- Apel, T. W., Mautner, J., Polack, A., Bornkamm, G. W., and Eick, D. (1992). Two antisense promoters in the immunoglobulin mu-switch region drive expression of c-myc in the Burkitt's lymphoma cell line BL67. *Oncogene* 7, 1267-1271.
- Aplan, P. D., Lombardi, D. P., Ginsberg, A. M., Cossman, J., Bertness, V. L., and Kirsch, I. R. (1990). Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science* 250, 1426-1429.
- Aplan, P. D., Lombardi, D. P., Reaman, G. H., Sather, H. N., Hammond, G. D., and Kirsch, I. R. (1992a). Involvement of the putative hematopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia. *Blood* 79, 1327-1333.
- Aplan, P. D., Raimondi, S. C., and Kirsch, I. R. (1992b). Disruption of the SCL gene by a t(1;3) translocation in a patient with T cell acute lymphoblastic leukemia. *J Exp Med* 176, 1303-1310.
- Aster, J. C., Kobayashi, Y., Shiota, M., Mori, S., and Sklar, J. (1992). Detection of the t(14;18) at similar frequencies in hyperplastic lymphoid tissues from American and Japanese patients. *Am J Pathol* 141, 291-299.
- Bakhshi, A., Jensen, J. P., Goldman, P., Wright, J. J., McBride, O. W., Epstein, A. L., and Korsmeyer, S. J. (1985). Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. *Cell* 41, 899-906.
- Bakhshi, A., Wright, J. J., Graninger, W., Seto, M., Owens, J., Cossman, J., Jensen, J. P., Goldman, P., and Korsmeyer, S. J. (1987). Mechanism of the t(14;18) chromosomal translocation: structural analysis of both derivative 14 and 18 reciprocal partners. *Proc Natl Acad Sci U S A* 84, 2396-2400.
- Baron, B. W., Nucifora, G., McCabe, N., Espinosa, R., 3rd, Le Beau, M. M., and McKeithan, T. W. (1993). Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas. *Proc Natl Acad Sci U S A* 90, 5262-5266.
- Barreto, V., Marques, R., and Demengeot, J. (2001). Early death and severe lymphopenia caused by ubiquitous expression of the Rag1 and Rag2 genes in mice. *Eur J Immunol* 31, 3763-3772.
- Bash, R. O., Crist, W. M., Shuster, J. J., Link, M. P., Amylon, M., Pullen, J., Carroll, A. J., Buchanan, G. R., Smith, R. G., and Baer, R. (1993). Clinical features and outcome of T-cell acute lymphoblastic leukemia in childhood with respect to alterations at the TAL1 locus: a Pediatric Oncology Group study. *Blood* 81, 2110-2117.
- Bassing, C. H., Ranganath, S., Murphy, M., Savic, V., Gleason, M., and Alt, F. W. (2008). Aberrant V(D)J recombination is not required for rapid development of H2ax/p53-deficient thymic lymphomas with clonal translocations. *Blood* 111, 2163-2169.
- Batley, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983). The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34, 779-787.
- Begley, C. G., Aplan, P. D., Davey, M. P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M. S., Haynes, B. F., Cohen, D. I., Waldmann, T. A., and *et al.* (1989). Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A* 86, 2031-2035.
- Bernard, O., Azogui, O., Lecointe, N., Mugneret, F., Berger, R., Larsen, C. J., and Mathieu-Mahul, D. (1992). A third tal-1 promoter is specifically used in human T cell leukemias. *J Exp Med* 176, 919-925.
- Bernard, O., Barin, C., Charrin, C., Mathieu-Mahul, D., and Berger, R. (1993). Characterization of translocation t(1;14)(p32;q11) in a T and in a B acute leukemia. *Leukemia* 7, 1509-1513.
- Bernard, O., Guglielmi, P., Jonveaux, P., Cherif, D., Gisselbrecht, S., Mauchauffe, M., Berger, R., Larsen, C. J., and Mathieu-Mahul, D. (1990). Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. *Genes Chromosomes Cancer* 1, 194-208.

- Bernard, O., Lecointe, N., Jonveaux, P., Souyri, M., Mauchauffe, M., Berger, R., Larsen, C. J., and Mathieu-Mahul, D. (1991). Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene. *Oncogene* 6, 1477-1488.
- Bertoni, F., Zucca, E., and Cotter, F. E. (2004). Molecular basis of mantle cell lymphoma. *Br J Haematol* 124, 130-140.
- Bizarro, S., Cerveira, N., Correia, C., Lisboa, S., Peixoto, A., Norton, L., and Teixeira, M. R. (2007). Molecular characterization of a rare MLL-AF4 (MLL-AFF1) fusion rearrangement in infant leukemia. *Cancer Genet Cytogenet* 178, 61-64.
- Boehm, T., Baer, R., Lavenir, I., Forster, A., Waters, J. J., Nacheva, E., and Rabbitts, T. H. (1988a). The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C delta locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *Embo J* 7, 385-394.
- Boehm, T., Buluwela, L., Williams, D., White, L., and Rabbitts, T. H. (1988b). A cluster of chromosome 11p13 translocations found via distinct D-D and D-D-J rearrangements of the human T cell receptor delta chain gene. *Embo J* 7, 2011-2017.
- Boehm, T., Foroni, L., Kaneko, Y., Perutz, M. F., and Rabbitts, T. H. (1991). The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc Natl Acad Sci U S A* 88, 4367-4371.
- Borkhardt, A., Repp, R., Haas, O. A., Leis, T., Harbott, J., Kreuder, J., Hammermann, J., Henn, T., and Lampert, F. (1997). Cloning and characterization of AFX, the gene that fuses to MLL in acute leukemias with a t(X;11)(q13;q23). *Oncogene* 14, 195-202.
- Breit, T. M., Mol, E. J., Wolvers-Tettero, I. L., Ludwig, W. D., van Wering, E. R., and van Dongen, J. J. (1993). Site-specific deletions involving the tal-1 and sil genes are restricted to cells of the T cell receptor alpha/beta lineage: T cell receptor delta gene deletion mechanism affects multiple genes. *J Exp Med* 177, 965-977.
- Brown, L., Cheng, J. T., Chen, Q., Siciliano, M. J., Crist, W., Buchanan, G., and Baer, R. (1990). Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. *Embo J* 9, 3343-3351.
- Busch, K., Borkhardt, A., Wossmann, W., Reiter, A., and Harbott, J. (2004). Combined polymerase chain reaction methods to detect c-myc/IgH rearrangement in childhood Burkitt's lymphoma for minimal residual disease analysis. *Haematologica* 89, 818-825.
- Busch, K., Keller, T., Fuchs, U., Yeh, R. F., Harbott, J., Kloese, I., Wiemels, J., Novosel, A., Reiter, A., and Borkhardt, A. (2007). Identification of two distinct MYC breakpoint clusters and their association with various IGH breakpoint regions in the t(8;14) translocations in sporadic Burkitt-lymphoma. *Leukemia* 21, 1739-1751.
- Care, A., Cianetti, L., Giampaolo, A., Sposi, N. M., Zappavigna, V., Mavilio, F., Alimena, G., Amadori, S., Mandelli, F., and Peschle, C. (1986). Translocation of c-myc into the immunoglobulin heavy-chain locus in human acute B-cell leukemia. A molecular analysis. *Embo J* 5, 905-911.
- Castor, A., Nilsson, L., Astrand-Grundstrom, I., Buitenhuis, M., Ramirez, C., Anderson, K., Strombeck, B., Garwicz, S., Bekassy, A. N., Schmiegelow, K., *et al.* (2005). Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nat Med* 11, 630-637.
- Cayuela, J. M., Gardie, B., and Sigaux, F. (1997). Disruption of the multiple tumor suppressor gene MTS1/p16(INK4a)/CDKN2 by illegitimate V(D)J recombinase activity in T-cell acute lymphoblastic leukemias. *Blood* 90, 3720-3726.
- Chaganti, S. R., Rao, P. H., Chen, W., Dyomin, V., Jhanwar, S. C., Parsa, N. Z., Dalla-Favera, R., and Chaganti, R. S. (1998). Deregulation of BCL6 in non-Hodgkin lymphoma by insertion of IGH sequences in complex translocations involving band 3q27. *Genes Chromosomes Cancer* 23, 328-336.
- Chen, Q., Cheng, J. T., Tasi, L. H., Schneider, N., Buchanan, G., Carroll, A., Crist, W., Ozanne, B., Siciliano, M. J., and Baer, R. (1990a). The tal gene undergoes chromosome translocation in T cell leukemia and potentially encodes a helix-loop-helix protein. *Embo J* 9, 415-424.
- Chen, Q., Yang, C. Y., Tsan, J. T., Xia, Y., Ragab, A. H., Peiper, S. C., Carroll, A., and Baer, R. (1990b). Coding sequences of the tal-1 gene are disrupted by chromosome translocation in human T cell leukemia. *J Exp Med* 172, 1403-1408.
- Chen, Y. W., Hu, X. T., Liang, A. C., Au, W. Y., So, C. C., Wong, M. L., Shen, L., Tao, Q., Chu, K. M., Kwong, Y. L., *et al.* (2006). High BCL6 expression predicts better prognosis, independent of

BCL6 translocation status, translocation partner, or BCL6-deregulating mutations, in gastric lymphoma. *Blood* 108, 2373-2383.

Cheng, J. T., Yang, C. Y., Hernandez, J., Embrey, J., and Baer, R. (1990). The chromosome translocation (11;14)(p13;q11) associated with T cell acute leukemia. Asymmetric diversification of the translocational junctions. *J Exp Med* 171, 489-501.

Chervinsky, D. S., Sait, S. N., Nowak, N. J., Shows, T. B., and Aplan, P. D. (1995). Complex MLL rearrangement in a patient with T-cell acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 14, 76-84.

Chisoe, S. L., Bodenteich, A., Wang, Y. F., Wang, Y. P., Burian, D., Clifton, S. W., Crabtree, J., Freeman, A., Iyer, K., Jian, L., and et al. (1995). Sequence and analysis of the human ABL gene, the BCR gene, and regions involved in the Philadelphia chromosomal translocation. *Genomics* 27, 67-82.

Cleary, M. L., Galili, N., and Sklar, J. (1986a). Detection of a second t(14;18) breakpoint cluster region in human follicular lymphomas. *J Exp Med* 164, 315-320.

Cleary, M. L., and Sklar, J. (1985). Nucleotide sequence of a t(14;18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. *Proc Natl Acad Sci U S A* 82, 7439-7443.

Cleary, M. L., Smith, S. D., and Sklar, J. (1986b). Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. *Cell* 47, 19-28.

Cotter, F., Price, C., Zucca, E., and Young, B. D. (1990). Direct sequence analysis of the 14q+ and 18q- chromosome junctions in follicular lymphoma. *Blood* 76, 131-135.

Crescenzi, M., Seto, M., Herzig, G. P., Weiss, P. D., Griffith, R. C., and Korsmeyer, S. J. (1988). Thermostable DNA polymerase chain amplification of t(14;18) chromosome breakpoints and detection of minimal residual disease. *Proc Natl Acad Sci U S A* 85, 4869-4873.

de Klein, A., van Agthoven, T., Groffen, C., Heisterkamp, N., Groffen, J., and Grosveld, G. (1986). Molecular analysis of both translocation products of a Philadelphia-positive CML patient. *Nucleic Acids Res* 14, 7071-7082.

Degan, M., Doliana, R., Gloghini, A., Di Francia, R., Aldinucci, D., Mazzocut-Zecchin, L., Colombatti, A., Attadia, V., Carbone, A., and Gattei, V. (2002). A novel bcl-1/JH breakpoint from a patient affected by mantle cell lymphoma extends the major translocation cluster. *J Pathol* 197, 256-263.

Delage, R., Roy, J., Jacques, L., Bernier, V., Delage, J. M., and Darveau, A. (1997). Multiple bcl-2/Ig gene rearrangements in persistent polyclonal B-cell lymphocytosis. *Br J Haematol* 97, 589-595.

Di Noia, J. M., and Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 76, 1-22.

Dik, W. A., Nadel, B., Przybylski, G. K., Asnafi, V., Grabarczyk, P., Navarro, J. M., Verhaaf, B., Schmidt, C. A., Macintyre, E. A., van Dongen, J. J., and Langerak, A. W. (2007). Different chromosomal breakpoints impact the level of LMO2 expression in T-ALL. *Blood* 110, 388-392.

Dolken, G., Illerhaus, G., Hirt, C., and Mertelsmann, R. (1996). BCL-2/JH rearrangements in circulating B cells of healthy blood donors and patients with nonmalignant diseases. *J Clin Oncol* 14, 1333-1344.

Duro, D., Bernard, O., Della Valle, V., Leblanc, T., Berger, R., and Larsen, C. J. (1996). Inactivation of the P16INK4/MTS1 gene by a chromosome translocation t(9;14)(p21-22;q11) in an acute lymphoblastic leukemia of B-cell type. *Cancer Res* 56, 848-854.

Dyson, P. J., and Rabbitts, T. H. (1985). Chromatin structure around the c-myc gene in Burkitt lymphomas with upstream and downstream translocation points. *Proc Natl Acad Sci U S A* 82, 1984-1988.

Eick, S., Krieger, G., Bolz, I., and Kneba, M. (1990). Sequence analysis of amplified t(14;18) chromosomal breakpoints in B-cell lymphomas. *J Pathol* 162, 127-133.

Erickson, P., Gao, J., Chang, K. S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to *Drosophila* segmentation gene, runt. *Blood* 80, 1825-1831.

- Felix, C. A., Hosler, M. R., Slater, D. J., Megonigal, M. D., Lovett, B. D., Williams, T. M., Nowell, P. C., Spinner, N. B., Owens, N. L., Hoxie, J., *et al.* (1999). Duplicated regions of AF-4 intron 4 at t(4;11) translocation breakpoints. *Mol Diagn* 4, 269-283.
- Felix, C. A., Kim, C. S., Megonigal, M. D., Slater, D. J., Jones, D. H., Spinner, N. B., Stump, T., Hosler, M. R., Nowell, P. C., Lange, B. J., and Rappaport, E. F. (1997). Panhandle polymerase chain reaction amplifies MLL genomic translocation breakpoint involving unknown partner gene. *Blood* 90, 4679-4686.
- Fischer, S., Mann, G., Konrad, M., Metzler, M., Ebetsberger, G., Jones, N., Nadel, B., Bodamer, O., Haas, O. A., Schmitt, K., and Panzer-Grumayer, E. R. (2007). Screening for leukemia- and clone-specific markers at birth in children with T-cell precursor ALL suggests a predominantly postnatal origin. *Blood* 110, 3036-3038.
- Fitzgerald, T. J., Neale, G. A., Raimondi, S. C., and Goorha, R. M. (1991). c-tal, a helix-loop-helix protein, is juxtaposed to the T-cell receptor-beta chain gene by a reciprocal chromosomal translocation: t(1;7)(p32;q35). *Blood* 78, 2686-2695.
- Ford, A. M., Bennett, C. A., Price, C. M., Bruin, M. C., Van Wering, E. R., and Greaves, M. (1998). Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* 95, 4584-4588.
- Frederico, L. A., Kunkel, T. A., and Shaw, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* 29, 2532-2537.
- Fu, J. F., Hsu, H. C., and Shih, L. Y. (2005). MLL is fused to EB1 (MAPRE1), which encodes a microtubule-associated protein, in a patient with acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 43, 206-210.
- Fusco, J. C., Setzer, R. W., Collard, D. D., and Moore, M. M. (1996). Quantification of t(14;18) in the lymphocytes of healthy adult humans as a possible biomarker for environmental exposures to carcinogens. *Carcinogenesis* 17, 1013-1020.
- Gale, K. B., Ford, A. M., Repp, R., Borkhardt, A., Keller, C., Eden, O. B., and Greaves, M. F. (1997). Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A* 94, 13950-13954.
- Galiegue-Zouitina, S., Collyn-d'Hooghe, M., Denis, C., Mainardi, A., Hildebrand, M. P., Tilly, H., Bastard, C., and Kerckaert, J. P. (1994). Molecular cloning of a t(11;14)(q13;q32) translocation breakpoint centromeric to the BCL1-MTC. *Genes Chromosomes Cancer* 11, 246-255.
- Galoin, S., al Saati, T., Schlaifer, D., Huynh, A., Attal, M., and Delsol, G. (1996). Oligonucleotide clonospecific probes directed against the junctional sequence of t(14;18): a new tool for the assessment of minimal residual disease in follicular lymphomas. *Br J Haematol* 94, 676-684.
- Garcia, I. S., Kaneko, Y., Gonzalez-Sarmiento, R., Campbell, K., White, L., Boehm, T., and Rabbitts, T. H. (1991). A study of chromosome 11p13 translocations involving TCR beta and TCR delta in human T cell leukaemia. *Oncogene* 6, 577-582.
- Gauss, G. H., and Lieber, M. R. (1996). Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol* 16, 258-269.
- Gauwerky, C. E., Haluska, F. G., Tsujimoto, Y., Nowell, P. C., and Croce, C. M. (1988). Evolution of B-cell malignancy: pre-B-cell leukemia resulting from MYC activation in a B-cell neoplasm with a rearranged BCL2 gene. *Proc Natl Acad Sci U S A* 85, 8548-8552.
- Gelmann, E. P., Psallidopoulos, M. C., Papas, T. S., and Dalla-Favera, R. (1983). Identification of reciprocal translocation sites within the c-myc oncogene and immunoglobulin mu locus in a Burkitt lymphoma. *Nature* 306, 799-803.
- Gillert, E., Leis, T., Repp, R., Reichel, M., Hosch, A., Breitenlohner, I., Angermuller, S., Borkhardt, A., Harbott, J., Lampert, F., *et al.* (1999). A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. *Oncogene* 18, 4663-4671.
- Gu, Y., Alder, H., Nakamura, T., Schichman, S. A., Prasad, R., Canaani, O., Saito, H., Croce, C. M., and Canaani, E. (1994). Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in acute leukemia. *Cancer Res* 54, 2327-2330.
- Gu, Y., Cimino, G., Alder, H., Nakamura, T., Prasad, R., Canaani, O., Moir, D. T., Jones, C., Nowell, P. C., Croce, C. M., and *et al.* (1992). The (4;11)(q21;q23) chromosome translocations in acute leukemias involve the VDJ recombinase. *Proc Natl Acad Sci U S A* 89, 10464-10468.

- Hayette, S., Tigaud, I., Vanier, A., Martel, S., Corbo, L., Charrin, C., Beillard, E., Deleage, G., Magaud, J. P., and Rimokh, R. (2000). AF15q14, a novel partner gene fused to the MLL gene in an acute myeloid leukaemia with a t(11;15)(q23;q14). *Oncogene* 19, 4446-4450.
- Heisterkamp, N., Stam, K., Groffen, J., de Klein, A., and Grosveld, G. (1985). Structural organization of the bcr gene and its role in the Ph' translocation. *Nature* 315, 758-761.
- Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., Green, J., Colman, S., Piacibello, W., Buckle, V., *et al.* (2008). Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* 319, 336-339.
- Hostein, I., Menard, A., Soubeyran, I., Eghbali, H., Debled, M., Gastaldello, B., and Soubeyran, P. (2001). A 1-kb Bcl-2-PCR fragment detection in a patient with follicular lymphoma and development of a new diagnostic PCR method. *Diagn Mol Pathol* 10, 89-94.
- Hotfilder, M., Rottgers, S., Rosemann, A., Schrauder, A., Schrappe, M., Pieters, R., Jurgens, H., Harbott, J., and Vormoor, J. (2005). Leukemic stem cells in childhood high-risk ALL/t(9;22) and t(4;11) are present in primitive lymphoid-restricted CD34+CD19- cells. *Cancer Res* 65, 1442-1449.
- Jaffe, E. S., Harris, N. L., Stein, H., and Vardiman, J. W. (2001) Pathology and genetics of tumours of haematopoietic and lymphoid tissues (Lyon; Washington, D.C.: IARC Press).
- Jager, U., Bocskor, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C., and Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* 95, 3520-3529.
- Jansen, M. W., Corral, L., van der Velden, V. H., Panzer-Grumayer, R., Schrappe, M., Schrauder, A., Marschalek, R., Meyer, C., den Boer, M. L., Hop, W. J., *et al.* (2007). Immunobiological diversity in infant acute lymphoblastic leukemia is related to the occurrence and type of MLL gene rearrangement. *Leukemia* 21, 633-641.
- Janssen, J. W., Ludwig, W. D., Sterry, W., and Bartram, C. R. (1993). SIL-TAL1 deletion in T-cell acute lymphoblastic leukemia. *Leukemia* 7, 1204-1210.
- Jeffs, A. R., Benjes, S. M., Smith, T. L., Sowerby, S. J., and Morris, C. M. (1998). The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Hum Mol Genet* 7, 767-776.
- Jenner, M. J., Summers, K. E., Norton, A. J., Amess, J. A., Arch, R. S., Young, B. D., Lister, T. A., Fitzgibbon, J., and Goff, L. K. (2002). JH probe real-time quantitative polymerase chain reaction assay for Bcl-2/IgH rearrangements. *Br J Haematol* 118, 550-558.
- Ji, W., Qu, G. Z., Ye, P., Zhang, X. Y., Halabi, S., and Ehrlich, M. (1995). Frequent detection of bcl-2/JH translocations in human blood and organ samples by a quantitative polymerase chain reaction assay. *Cancer Res* 55, 2876-2882.
- Jonsson, O. G., Kitchens, R. L., Baer, R. J., Buchanan, G. R., and Smith, R. G. (1991). Rearrangements of the tal-1 locus as clonal markers for T cell acute lymphoblastic leukemia. *J Clin Invest* 87, 2029-2035.
- Kagan, J., Finger, L. R., Letofsky, J., Finan, J., Nowell, P. C., and Croce, C. M. (1989). Clustering of breakpoints on chromosome 10 in acute T-cell leukemias with the t(10;14) chromosome translocation. *Proc Natl Acad Sci U S A* 86, 4161-4165.
- Kagan, J., Joe, Y. S., and Freireich, E. J. (1994). Joining of recombination signals on the der 14q-chromosome in T-cell acute leukemia with t(10;14) chromosome translocation. *Cancer Res* 54, 226-230.
- Kerckaert, J. P., Dewindt, C., Tilly, H., Quief, S., Lecocq, G., and Bastard, C. (1993). LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. *Nat Genet* 5, 66-70.
- Kitagawa, Y., Inoue, K., Sasaki, S., Hayashi, Y., Matsuo, Y., Lieber, M. R., Mizoguchi, H., Yokota, J., and Kohno, T. (2002). Prevalent involvement of illegitimate V(D)J recombination in chromosome 9p21 deletions in lymphoid leukemia. *J Biol Chem* 277, 46289-46297.
- Kneba, M., Eick, S., Herbst, H., Pott, C., Bolz, I., Dallenbach, F., Hiddemann, W., and Stein, H. (1995). Low incidence of mbr bcl-2/JH fusion genes in Hodgkin's disease. *J Pathol* 175, 381-389.
- Koduru, P. R., Goh, J. C., Pergolizzi, R. G., Lichtman, S. M., and Broome, J. D. (1993). Molecular characterization of a variant Ph1 translocation t(9;22;11) (q34;q11;q13) in chronic myelogenous leukemia (CML) reveals the translocation of the 3'-part of BCR gene to the chromosome band 11q13. *Oncogene* 8, 3239-3247.

- Kuefer, M. U., Chinwalla, V., Zeleznik-Le, N. J., Behm, F. G., Naeve, C. W., Rakestraw, K. M., Mukatira, S. T., Raimondi, S. C., and Morris, S. W. (2003). Characterization of the MLL partner gene AF15q14 involved in t(11;15)(q23;q14). *Oncogene* 22, 1418-1424.
- Ladetto, M., Mantoan, B., Ricca, I., Astolfi, M., Drandi, D., Compagno, M., Vallet, S., dell'Aquila, M., Alfano, A., Rossatto, P., *et al.* (2003). Recurrence of Bcl-2/IgH polymerase chain reaction positivity following a prolonged molecular remission can be unrelated to the original follicular lymphoma clone. *Exp Hematol* 31, 784-788.
- Langer, T., Metzler, M., Reinhardt, D., Viehmann, S., Borkhardt, A., Reichel, M., Stanulla, M., Schrappe, M., Creutzig, U., Ritter, J., *et al.* (2003). Analysis of t(9;11) chromosomal breakpoint sequences in childhood acute leukemia: almost identical MLL breakpoints in therapy-related AML after treatment without etoposides. *Genes Chromosomes Cancer* 36, 393-401.
- Liao, M. J., Zhang, X. X., Hill, R., Gao, J., Qumsiyeh, M. B., Nichols, W., and Van Dyke, T. (1998). No requirement for V(D)J recombination in p53-deficient thymic lymphoma. *Mol Cell Biol* 18, 3495-3501.
- Limpens, J., de Jong, D., van Krieken, J. H., Price, C. G., Young, B. D., van Ommen, G. J., and Kluin, P. M. (1991). Bcl-2/JH rearrangements in benign lymphoid tissues with follicular hyperplasia. *Oncogene* 6, 2271-2276.
- Limpens, J., Stad, R., Vos, C., de Vlaam, C., de Jong, D., van Ommen, G. J., Schuurin, E., and Kluin, P. M. (1995). Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. *Blood* 85, 2528-2536.
- Litz, C. E., McClure, J. S., Copenhaver, C. M., and Brunning, R. D. (1993). Duplication of small segments within the major breakpoint cluster region in chronic myelogenous leukemia. *Blood* 81, 1567-1572.
- Liu, M., Duke, J. L., Richter, D. J., Vinuesa, C. G., Goodnow, C. C., Kleinstein, S. H., and Schatz, D. G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451, 841-845.
- Liu, Y., Hernandez, A. M., Shibata, D., and Cortopassi, G. A. (1994). BCL2 translocation frequency rises with age in humans. *Proc Natl Acad Sci U S A* 91, 8910-8914.
- Lovec, H., Grzeschiczek, A., Kowalski, M. B., and Moroy, T. (1994). Cyclin D1/bcl-1 cooperates with myc genes in the generation of B-cell lymphoma in transgenic mice. *Embo J* 13, 3487-3495.
- Lu, M., Dube, I., Raimondi, S., Carroll, A., Zhao, Y., Minden, M., and Sutherland, P. (1990). Molecular characterization of the t(10;14) translocation breakpoints in T-cell acute lymphoblastic leukemia: further evidence for illegitimate physiological recombination. *Genes Chromosomes Cancer* 2, 217-222.
- Luthra, R., McBride, J. A., Hai, S., Cabanillas, F., and Pugh, W. C. (1997). The application of fluorescence-based PCR and PCR-SSCP to monitor the clonal relationship of cells bearing the t(14;18)(q32;q21) in sequential biopsy specimens from patients with follicle center cell lymphoma. *Diagn Mol Pathol* 6, 71-77.
- Mao, C., Jiang, L., Melo-Jorge, M., Puthenveetil, M., Zhang, X., Carroll, M. C., and Imanishi-Kari, T. (2004). T cell-independent somatic hypermutation in murine B cells with an immature phenotype. *Immunity* 20, 133-144.
- Marculescu, R., Le, T., Bocskor, S., Mitterbauer, G., Chott, A., Mannhalter, C., Jaeger, U., and Nadel, B. (2002a). Alternative end-joining in follicular lymphomas' t(14;18) translocation. *Leukemia* 16, 120-126.
- Marculescu, R., Le, T., Simon, P., Jaeger, U., and Nadel, B. (2002b). V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med* 195, 85-98.
- Marschalek, R., Greil, J., Lochner, K., Nilson, I., Siegler, G., Zweckbronner, I., Beck, J. D., and Fey, G. H. (1995). Molecular analysis of the chromosomal breakpoint and fusion transcripts in the acute lymphoblastic SEM cell line with chromosomal translocation t(4;11). *Br J Haematol* 90, 308-320.
- Matolcsy, A., Warnke, R. A., and Knowles, D. M. (1997). Somatic mutations of the translocated bcl-2 gene are associated with morphologic transformation of follicular lymphoma to diffuse large-cell lymphoma. *Ann Oncol* 8 Suppl 2, 119-122.

- McDonnell, T. J., Deane, N., Platt, F. M., Nunez, G., Jaeger, U., McKearn, J. P., and Korsmeyer, S. J. (1989). bcl-2-immunoglobulin transgenic mice demonstrate extended B cell survival and follicular lymphoproliferation. *Cell* 57, 79-88.
- McGuire, E. A., Hockett, R. D., Pollock, K. M., Bartholdi, M. F., O'Brien, S. J., and Korsmeyer, S. J. (1989). The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. *Mol Cell Biol* 9, 2124-2132.
- McHale, C. M., Wiemels, J. L., Zhang, L., Ma, X., Buffler, P. A., Guo, W., Loh, M. L., and Smith, M. T. (2003). Prenatal origin of TEL-AML1-positive acute lymphoblastic leukemia in children born in California. *Genes Chromosomes Cancer* 37, 36-43.
- Meeker, T. C., Sellers, W., Harvey, R., Withers, D., Carey, K., Xiao, H., Block, A. M., Dadey, B., and Han, T. (1991). Cloning of the t(11;14)(q13;q32) translocation breakpoints from two human leukemia cell lines. *Leukemia* 5, 733-737.
- Megonigal, M. D., Rappaport, E. F., Jones, D. H., Williams, T. M., Lovett, B. D., Kelly, K. M., Lerou, P. H., Moulton, T., Budarf, M. L., and Felix, C. A. (1998). t(11;22)(q23;q11.2) In acute myeloid leukemia of infant twins fuses MLL with hCDCrel, a cell division cycle gene in the genomic region of deletion in DiGeorge and velocardiofacial syndromes. *Proc Natl Acad Sci U S A* 95, 6413-6418.
- Meyer, C., Schneider, B., Jakob, S., Strehl, S., Attarbaschi, A., Schnittger, S., Schoch, C., Jansen, M. W., van Dongen, J. J., den Boer, M. L., *et al.* (2006). The MLL recombinome of acute leukemias. *Leukemia* 20, 777-784.
- Mills, K. I., Sproul, A. M., Ogilvie, D., Elvin, P., Leibowitz, D., and Burnett, A. K. (1992). Amplification and sequencing of genomic breakpoints located within the M-bcr region by Vectorette-mediated polymerase chain reaction. *Leukemia* 6, 481-483.
- Muller, J. R., Janz, S., and Potter, M. (1995). Differences between Burkitt's lymphomas and mouse plasmacytomas in the immunoglobulin heavy chain/c-myc recombinations that occur in their chromosomal translocations. *Cancer Res* 55, 5012-5018.
- Mullighan, C. G., Miller, C. B., Radtke, I., Phillips, L. A., Dalton, J., Ma, J., White, D., Hughes, T. P., Le Beau, M. M., Pui, C. H., *et al.* (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 453, 110-114.
- Murphy, W., Sarid, J., Taub, R., Vasicek, T., Battey, J., Lenoir, G., and Leder, P. (1986). A translocated human c-myc oncogene is altered in a conserved coding sequence. *Proc Natl Acad Sci U S A* 83, 2939-2943.
- Nakamura, Y., Miki, T., Kawamata, N., Ohashi, K., Hirose, S., Kobayashi, H., Maseki, N., Kaneko, Y., Saito, K., Enokihara, H., and Furusawa, S. (1997). Two Burkitt-type lymphoma/leukemia-derived cell lines presenting 3q27 translocations and immunoglobulin/BCL6 chimeric transcripts. *Leukemia* 11, 1993-1994.
- Negrini, M., Felix, C. A., Martin, C., Lange, B. J., Nakamura, T., Canaani, E., and Croce, C. M. (1993). Potential topoisomerase II DNA-binding sites at the breakpoints of a t(9;11) chromosome translocation in acute myeloid leukemia. *Cancer Res* 53, 4489-4492.
- Ngan, B. Y., Nourse, J., and Cleary, M. L. (1989). Detection of chromosomal translocation t(14;18) within the minor cluster region of bcl-2 by polymerase chain reaction and direct genomic sequencing of the enzymatically amplified DNA in follicular lymphomas. *Blood* 73, 1759-1762.
- Nilsen, H., Stamp, G., Andersen, S., Hrivnak, G., Krokan, H. E., Lindahl, T., and Barnes, D. E. (2003). Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene* 22, 5381-5386.
- Nyvold, C. G., Bendix, K., Brandsborg, M., Pulczynski, S., Silkjaer, T., and Hokland, P. (2007). Multiplex PCR for the detection of BCL-1/IGH and BCL-2/IGH gene rearrangements--clinical validation in a prospective study of blood and bone marrow in 258 patients with or suspected of non-Hodgkin's lymphoma. *Acta Oncol* 46, 21-30.
- Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 197, 1173-1181.
- Park, J. K., Le Beau, M. M., Shows, T. B., Rowley, J. D., and Diaz, M. O. (1992). A complex genetic rearrangement in a t(10;14)(q24;q11) associated with T-cell acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 4, 32-40.

- Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* 38, 3468-3477.
- Pine, S. R., Wiemels, J. L., Jayabose, S., and Sandoval, C. (2003). TEL-AML1 fusion precedes differentiation to pre-B cells in childhood acute lymphoblastic leukemia. *Leuk Res* 27, 155-164.
- Pott, C., Tiemann, M., Linke, B., Ott, M. M., von Hofen, M., Bolz, I., Hiddemann, W., Parwaresch, R., and Kneba, M. (1998). Structure of Bcl-1 and IgH-CDR3 rearrangements as clonal markers in mantle cell lymphomas. *Leukemia* 12, 1630-1637.
- Price, C. G., Tuszyński, A., Watt, S. M., Murdoch, S. J., Lister, T. A., and Young, B. D. (1991). Detection of additional JH/BCL2 translocations in follicular lymphoma. *Leukemia* 5, 548-554.
- Rabbitts, P. H., Douglas, J., Fischer, P., Nacheva, E., Karpas, A., Catovsky, D., Melo, J. V., Baer, R., Stinson, M. A., and Rabbitts, T. H. (1988). Chromosome abnormalities at 11q13 in B cell tumours. *Oncogene* 3, 99-103.
- Raffini, L. J., Slater, D. J., Rappaport, E. F., Lo Nigro, L., Cheung, N. K., Biegel, J. A., Nowell, P. C., Lange, B. J., and Felix, C. A. (2002). Panhandle and reverse-panhandle PCR enable cloning of der(11) and der(other) genomic breakpoint junctions of MLL translocations and identify complex translocation of MLL, AF-4, and CDK6. *Proc Natl Acad Sci U S A* 99, 4568-4573.
- Raghavan, S. C., Gu, J., Swanson, P. C., and Lieber, M. R. (2007). The structure-specific nicking of small heteroduplexes by the RAG complex: implications for lymphoid chromosomal translocations. *DNA Repair (Amst)* 6, 751-759.
- Rauzy, O., Galoin, S., Chale, J. J., Adoue, D., Albarede, J. L., Delsol, G., and al Saati, T. (1998). Detection of t(14;18) carrying cells in bone marrow and peripheral blood from patients affected by non-lymphoid diseases. *Mol Pathol* 51, 333-338.
- Reichel, M., Gillert, E., Angermüller, S., Hensel, J. P., Heidel, F., Lode, M., Leis, T., Biondi, A., Haas, O. A., Strehl, S., et al. (2001). Biased distribution of chromosomal breakpoints involving the MLL gene in infants versus children and adults with t(4;11) ALL. *Oncogene* 20, 2900-2907.
- Reichel, M., Gillert, E., Breitenlohner, I., Repp, R., Greil, J., Beck, J. D., Fey, G. H., and Marschalek, R. (1999). Rapid isolation of chromosomal breakpoints from patients with t(4;11) acute lymphoblastic leukemia: implications for basic and clinical research. *Cancer Res* 59, 3357-3362.
- Reichel, M., Gillert, E., Nilson, I., Siegler, G., Greil, J., Fey, G. H., and Marschalek, R. (1998). Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. *Oncogene* 17, 3035-3044.
- Reitmair, A. H., Schmits, R., Ewel, A., Bapat, B., Redston, M., Mitri, A., Waterhouse, P., Mittrucker, H. W., Wakeham, A., Liu, B., and et al. (1995). MSH2 deficient mice are viable and susceptible to lymphoid tumours. *Nat Genet* 11, 64-70.
- Rimokh, R., Berger, F., Delsol, G., Digonnet, I., Rouault, J. P., Tigaud, J. D., Gadoux, M., Coiffier, B., Bryon, P. A., and Magaud, J. P. (1994). Detection of the chromosomal translocation t(11;14) by polymerase chain reaction in mantle cell lymphomas. *Blood* 83, 1871-1875.
- Roulland, S., Lebailly, P., and Gauduchon, P. (2003). Correspondence re: Welzel et al, *Cancer Res* 61: 1629-1636. *Cancer Res* 63, 1722-1723.
- Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Tonegawa, S. (1983). Activation of the c-myc gene by translocation: a model for translational control. *Proc Natl Acad Sci U S A* 80, 7476-7480.
- Salvati, P. D., Watt, P. M., Thomas, W. R., and Kees, U. R. (1999). Molecular characterization of a complex chromosomal translocation breakpoint t(10;14) including the HOX11 oncogene locus. *Leukemia* 13, 975-979.
- Sanchez-Izquierdo, D., Siebert, R., Harder, L., Marugan, I., Gozzetti, A., Price, H. P., Gesk, S., Hernandez-Rivas, J. M., Benet, I., Sole, F., et al. (2001). Detection of translocations affecting the BCL6 locus in B cell non-Hodgkin's lymphoma by interphase fluorescence in situ hybridization. *Leukemia* 15, 1475-1484.
- Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M., and Peng, A. (1993). Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev* 7, 2520-2532.
- Schmitt, C., Balogh, B., Grundt, A., Buchholtz, C., Leo, A., Benner, A., Hensel, M., Ho, A. D., and Leo, E. (2006). The bcl-2/IgH rearrangement in a population of 204 healthy individuals:

occurrence, age and gender distribution, breakpoints, and detection method validity. *Leuk Res* 30, 745-750.

Schwindt, H., Akasaka, T., Zuhlke-Jenisch, R., Hans, V., Schaller, C., Klapper, W., Dyer, M. J., Siebert, R., and Deckert, M. (2006). Chromosomal translocations fusing the BCL6 gene to different partner loci are recurrent in primary central nervous system lymphoma and may be associated with aberrant somatic hypermutation or defective class switch recombination. *J Neuropathol Exp Neurol* 65, 776-782.

Segal, G. H., Masih, A. S., Fox, A. C., Jorgensen, T., Scott, M., and Braylan, R. C. (1995). CD5-expressing B-cell non-Hodgkin's lymphomas with bcl-1 gene rearrangement have a relatively homogeneous immunophenotype and are associated with an overall poor prognosis. *Blood* 85, 1570-1579.

Showe, L. C., Ballantine, M., Nishikura, K., Erikson, J., Kaji, H., and Croce, C. M. (1985). Cloning and sequencing of a c-myc oncogene in a Burkitt's lymphoma cell line that is translocated to a germ line alpha switch region. *Mol Cell Biol* 5, 501-509.

Smith, D. P., Bath, M. L., Harris, A. W., and Cory, S. (2005). T-cell lymphomas mask slower developing B-lymphoid and myeloid tumours in transgenic mice with broad haemopoietic expression of MYC. *Oncogene* 24, 3544-3553.

Soubeyran, P., Hostein, I., Debled, M., Eghbali, H., Soubeyran, I., Bonichon, F., Astier-Gin, T., and Hoerni, B. (1999). Evolution of BCL-2/IgH hybrid gene RNA expression during treatment of T(14;18)-bearing follicular lymphomas. *Br J Cancer* 81, 860-869.

Sowerby, S. J., Kennedy, M. A., Fitzgerald, P. H., and Morris, C. M. (1993). DNA sequence analysis of the major breakpoint cluster region of the BCR gene rearranged in Philadelphia-positive human leukemias. *Oncogene* 8, 1679-1683.

Stamatopoulos, K., Kosmas, C., Belessi, C., Kyriazopoulos, P., Papadaki, T., Anagnostou, D., and Loukopoulos, D. (1999). Molecular analysis of bcl-1/IgH junctional sequences in mantle cell lymphoma: potential mechanism of the t(11;14) chromosomal translocation. *Br J Haematol* 105, 190-197.

Summers, K. E., Goff, L. K., Wilson, A. G., Gupta, R. K., Lister, T. A., and Fitzgibbon, J. (2001). Frequency of the Bcl-2/IgH rearrangement in normal individuals: implications for the monitoring of disease in patients with follicular lymphoma. *J Clin Oncol* 19, 420-424.

Super, H. G., Strissel, P. L., Sobulo, O. M., Burian, D., Reshmi, S. C., Roe, B., Zeleznik-Le, N. J., Diaz, M. O., and Rowley, J. D. (1997). Identification of complex genomic breakpoint junctions in the t(9;11) MLL-AF9 fusion gene in acute leukemia. *Genes Chromosomes Cancer* 20, 185-195.

Takacs, I., Zeher, M., Urban, L., Szegedi, G., and Semsei, I. (2000). Diagnostic value of the detection of t(14;18) chromosome translocation in malignant hematological and immunopathological diseases using polymerase chain reaction. *Acta Med Okayama* 54, 185-192.

Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E., and Croce, C. M. (1985a). The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science* 229, 1390-1393.

Tsujimoto, Y., Jaffe, E., Cossman, J., Gorham, J., Nowell, P. C., and Croce, C. M. (1985b). Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation. *Nature* 315, 340-343.

Tsujimoto, Y., Louie, E., Bashir, M. M., and Croce, C. M. (1988). The reciprocal partners of both the t(14; 18) and the t(11; 14) translocations involved in B-cell neoplasms are rearranged by the same mechanism. *Oncogene* 2, 347-351.

Van Vlierberghe, P., van Grotel, M., Beverloo, H. B., Lee, C., Helgason, T., Buijs-Gladdines, J., Passier, M., van Wering, E. R., Veerman, A. J., Kamps, W. A., *et al.* (2006). The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood* 108, 3520-3529.

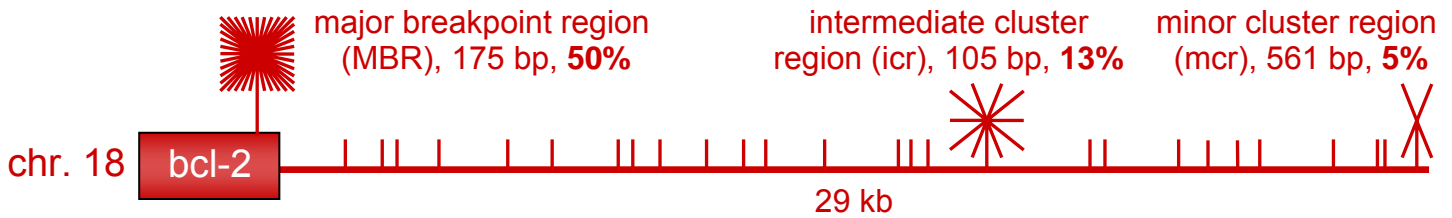
Vega, F., and Medeiros, L. J. (2003). Chromosomal translocations involved in non-Hodgkin lymphomas. *Arch Pathol Lab Med* 127, 1148-1160.

Vieira, L., Sousa, A. C., Matos, P., Marques, B., Alaiz, H., Ribeiro, M. J., Braga, P., da Silva, M. G., and Jordan, P. (2006). Three-way translocation involves MLL, MLLT3, and a novel cell cycle control gene, FLJ10374, in the pathogenesis of acute myeloid leukemia with t(9;11;19)(p22;q23;p13.3). *Genes Chromosomes Cancer* 45, 455-469.

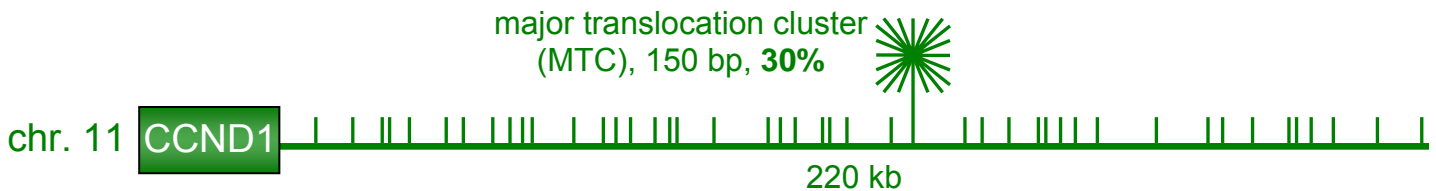
- Wada, M., Bartram, C. R., Nakamura, H., Hachiya, M., Chen, D. L., Borenstein, J., Miller, C. W., Ludwig, L., Hansen-Hagge, T. E., Ludwig, W. D., and et al. (1993). Analysis of p53 mutations in a large series of lymphoid hematologic malignancies of childhood. *Blood* 82, 3163-3169.
- Wang, Y. L., Addya, K., Edwards, R. H., Rennert, H., Dodson, L., Leonard, D. G., and Wilson, R. B. (1998). Novel bcl-2 breakpoints in patients with follicular lymphoma. *Diagn Mol Pathol* 7, 85-89.
- Wechsler, D. S., Engstrom, L. D., Alexander, B. M., Motto, D. G., and Roulston, D. (2003). A novel chromosomal inversion at 11q23 in infant acute myeloid leukemia fuses MLL to CALM, a gene that encodes a clathrin assembly protein. *Genes Chromosomes Cancer* 36, 26-36.
- Weinberg, O. K., Ai, W. Z., Mariappan, M. R., Shum, C., Levy, R., and Arber, D. A. (2007). "Minor" BCL2 breakpoints in follicular lymphoma: frequency and correlation with grade and disease presentation in 236 cases. *J Mol Diagn* 9, 530-537.
- Weinstock, D. M., Brunet, E., and Jasin, M. (2007). Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70. *Nat Cell Biol* 9, 978-981.
- Weizel, N., Le, T., Marculescu, R., Mitterbauer, G., Chott, A., Pott, C., Kneba, M., Du, M. Q., Kusec, R., Drach, J., et al. (2001). Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. *Cancer Res* 61, 1629-1636.
- Wiemels, J. L., Alexander, F. E., Cazzaniga, G., Biondi, A., Mayer, S. P., and Greaves, M. (2000). Microclustering of TEL-AML1 translocation breakpoints in childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 29, 219-228.
- Wiemels, J. L., and Greaves, M. (1999). Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res* 59, 4075-4082.
- Wiemels, J. L., Leonard, B. C., Wang, Y., Segal, M. R., Hunger, S. P., Smith, M. T., Crouse, V., Ma, X., Buffler, P. A., and Pine, S. R. (2002a). 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.
- Wiemels, J. L., Xiao, Z., Buffler, P. A., Maia, A. T., Ma, X., Dicks, B. M., Smith, M. T., Zhang, L., Feusner, J., Wiencke, J., et al. (2002b). In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* 99, 3801-3805.
- Wilda, M., Busch, K., Klose, I., Keller, T., Woessmann, W., Kreuder, J., Harbott, J., and Borkhardt, A. (2004). Level of MYC overexpression in pediatric Burkitt's lymphoma is strongly dependent on genomic breakpoint location within the MYC locus. *Genes Chromosomes Cancer* 41, 178-182.
- Williams, M. E., Swerdlow, S. H., and Meeker, T. C. (1993). Chromosome t(11;14)(q13;q32) breakpoints in centrocytic lymphoma are highly localized at the bcl-1 major translocation cluster. *Leukemia* 7, 1437-1440.
- Willis, T. G., Jadayel, D. M., Coignet, L. J., Abdul-Rauf, M., Treleaven, J. G., Catovsky, D., and Dyer, M. J. (1997). Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction. *Blood* 90, 2456-2464.
- Wiman, K. G., Clarkson, B., Hayday, A. C., Saito, H., Tonegawa, S., and Hayward, W. S. (1984). Activation of a translocated c-myc gene: role of structural alterations in the upstream region. *Proc Natl Acad Sci U S A* 81, 6798-6802.
- Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A. M., Lipkin, M., and Edelmann, W. (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* 99, 14937-14942.
- Wyatt, R. T., Rudders, R. A., Zelenetz, A., Delellis, R. A., and Krontiris, T. G. (1992). BCL2 oncogene translocation is mediated by a chi-like consensus. *J Exp Med* 175, 1575-1588.
- Xia, Y., Brown, L., Tsan, J. T., Yang, C. Y., Siciliano, M. J., Crist, W. M., Carroll, A. J., and Baer, R. (1992). The translocation (1;14)(p34;q11) in human T-cell leukemia: chromosome breakage 25 kilobase pairs downstream of the TAL1 protooncogene. *Genes Chromosomes Cancer* 4, 211-216.
- Xiao, Z., Greaves, M. F., Buffler, P., Smith, M. T., Segal, M. R., Dicks, B. M., Wiencke, J. K., and Wiemels, J. L. (2001). Molecular characterization of genomic AML1-ETO fusions in childhood leukemia. *Leukemia* 15, 1906-1913.

- Yang, X., Lee, K., Said, J., Gong, X., and Zhang, K. (2006). Association of Ig/BCL6 translocations with germinal center B lymphocytes in human lymphoid tissues: implications for malignant transformation. *Blood* 108, 2006-2012.
- Ye, B. H., Chaganti, S., Chang, C. C., Niu, H., Corradini, P., Chaganti, R. S., and Dalla-Favera, R. (1995). Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. *Embo J* 14, 6209-6217.
- Yoffe, G., Schneider, N., Van Dyk, L., Yang, C. Y., Siciliano, M., Buchanan, G., Capra, J. D., and Baer, R. (1989). The chromosome translocation (11;14)(p13;q11) associated with T-cell acute lymphocytic leukemia: an 11p13 breakpoint cluster region. *Blood* 74, 374-379.
- Yoshida, S., Kaneita, Y., Aoki, Y., Seto, M., Mori, S., and Moriyama, M. (1999). Identification of heterologous translocation partner genes fused to the BCL6 gene in diffuse large B-cell lymphomas: 5'-RACE and LA-PCR analyses of biopsy samples. *Oncogene* 18, 7994-7999.
- Young, K. H., Xie, Q., Zhou, G., Eickhoff, J. C., Sanger, W. G., Aoun, P., and Chan, W. C. (2008). Transformation of follicular lymphoma to precursor B-cell lymphoblastic lymphoma with c-myc gene rearrangement as a critical event. *Am J Clin Pathol* 129, 157-166.
- Yu, K., Taghva, A., and Lieber, M. R. (2002). The cleavage efficiency of the human immunoglobulin heavy chain VH elements by the RAG complex: implications for the immune repertoire. *J Biol Chem* 277, 5040-5046.
- Zarrin, A. A., Del Vecchio, C., Tseng, E., Gleason, M., Zarin, P., Tian, M., and Alt, F. W. (2007). Antibody class switching mediated by yeast endonuclease-generated DNA breaks. *Science* 315, 377-381.
- Zhang, J. G., Goldman, J. M., and Cross, N. C. (1995). Characterization of genomic BCR-ABL breakpoints in chronic myeloid leukaemia by PCR. *Br J Haematol* 90, 138-146.
- Zhang, M., and Swanson, P. C. (2008). V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. *J Biol Chem* 283, 6717-6727.
- Zhang, Y., Strissel, P., Strick, R., Chen, J., Nucifora, G., Le Beau, M. M., Larson, R. A., and Rowley, J. D. (2002). Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. *Proc Natl Acad Sci U S A* 99, 3070-3075.
- Zutter, M., Hockett, R. D., Roberts, C. W., McGuire, E. A., Bloomstone, J., Morton, C. C., Deaven, L. L., Crist, W. M., Carroll, A. J., and Korsmeyer, S. J. (1990). The t(10;14)(q24;q11) of T-cell acute lymphoblastic leukemia juxtaposes the delta T-cell receptor with TCL3, a conserved and activated locus at 10q24. *Proc Natl Acad Sci U S A* 87, 3161-3165.

A. *bcl-2* MBR, *icr*, and *mcr*.



B. *bcl-1* MTC.



C. E2A cluster.

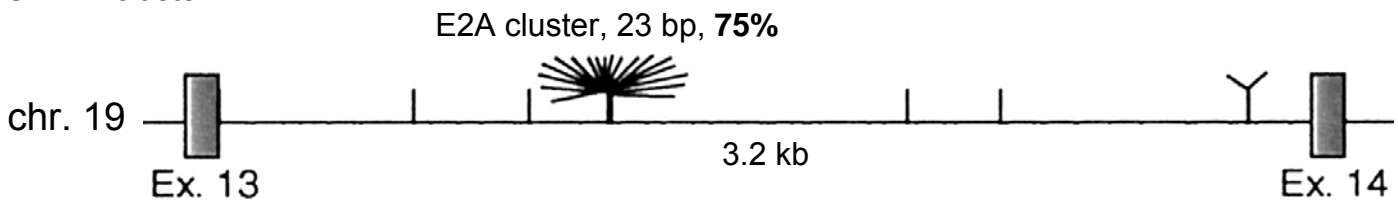


Figure S1. Schematics of breakpoint cluster regions.

Schematics of the *bcl-2*, *bcl-1*, and E2A regions illustrate clustering of breakpoints within the various identified cluster regions. The breakpoints that do not fall into cluster regions are plotted randomly for illustrative purposes, as most of them are never sequenced and often only mapped to general regions by Southern blotting.

(A) depicts relative proportions of breakpoints at the *bcl-2* MBR, *icr*, and *mcr* cluster regions. The third exon of the *bcl-2* gene, boxed, contains the MBR (within the 3' UTR region), while the centromeric 29 kb contains the *icr* and *mcr*. Short lines above the gene diagram mark the approximate locations and relative abundance of patient breakpoints. The 175 bp MBR, 105 bp *icr*, and 561 bp *mcr* account for about 50%, 13%, and 5% of *bcl-2* translocation breakpoints, respectively.

The *bcl-1* MTC, represented in (B), is located about 110 kb from *CCND1*, the gene for the cyclin D1 oncoprotein. The 150 bp MTC contains about 30% of breakpoints, whereas the remaining 70% of events are distributed widely over the surrounding 220 kb.

(C) shows a diagram of intron 13 of the E2A gene, taken from (Wiemels et al., 2002a). 75% of breakpoints occur in the 23 bp E2A cluster, while the surrounding 3 kb only account for 25%.

A. bcl-1 MTC.



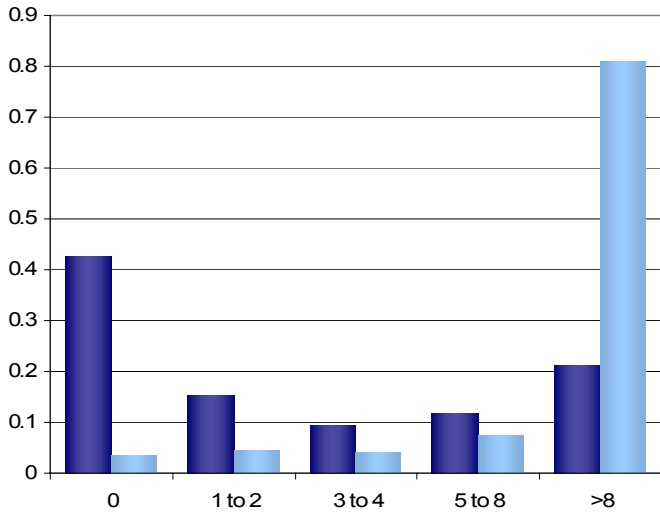
B. E2A cluster.



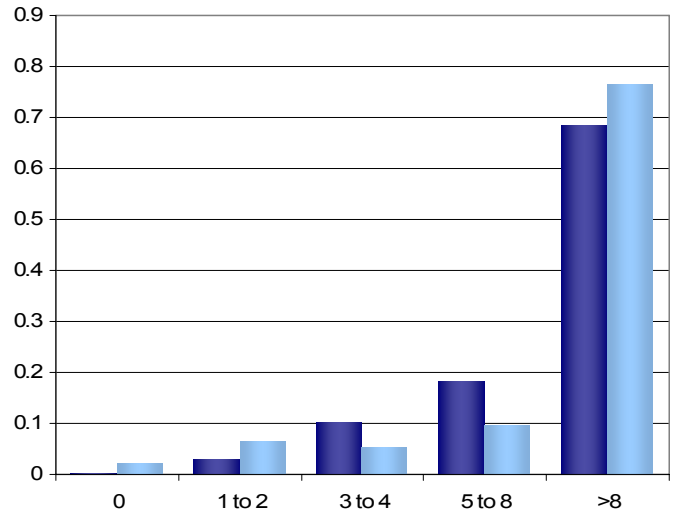
Figure S2. Breakpoint plots for the bcl-1 MTC and E2A cluster regions.

Breakpoint distributions are plotted for the (A) bcl-1 MTC, running telomeric to centromeric, with der(14) breakpoints above and der(11) breakpoints below; and (B) E2A cluster, running centromeric to telomeric, with der(19) breakpoints above and der(1) breakpoints below. Each breakpoint is represented as a triangle adjoining the breakpoint site.

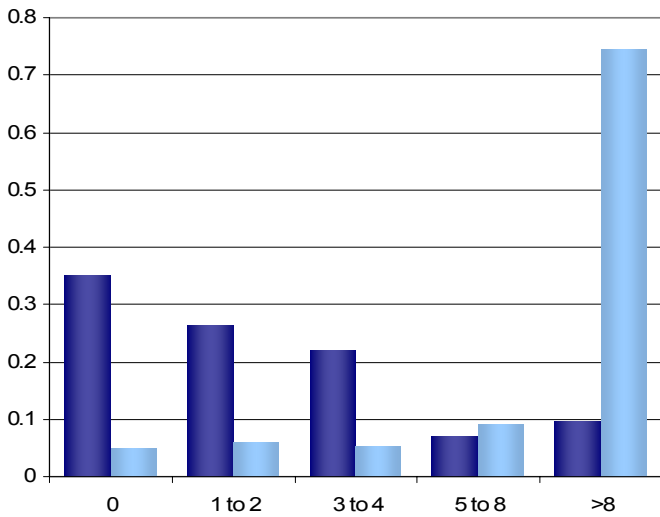
bcl-2 (MBR + icr + mcr + unclustered), CpG



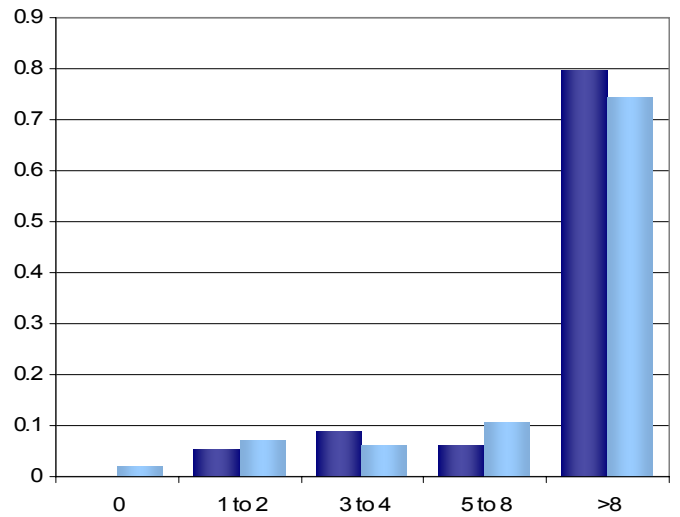
bcl-2 (MBR + icr + mcr + unclustered), CAC



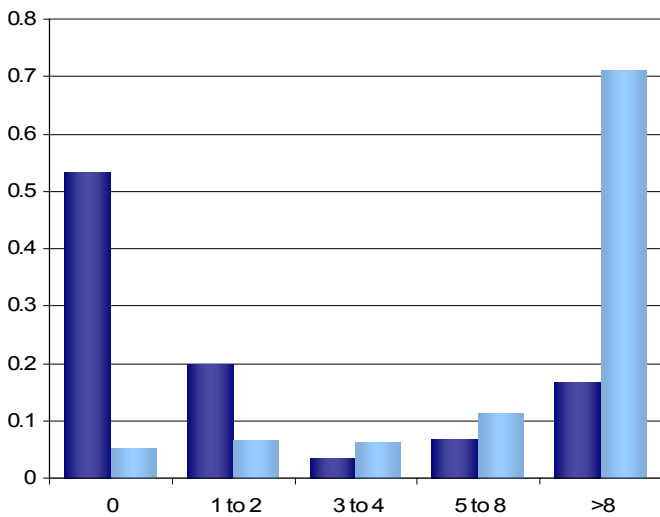
bcl-1 (MTC + unclustered), CpG



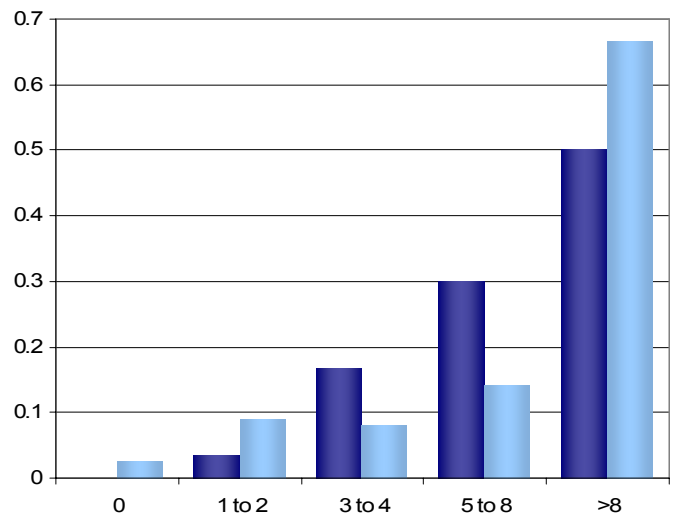
bcl-1 (MTC + unclustered), CAC



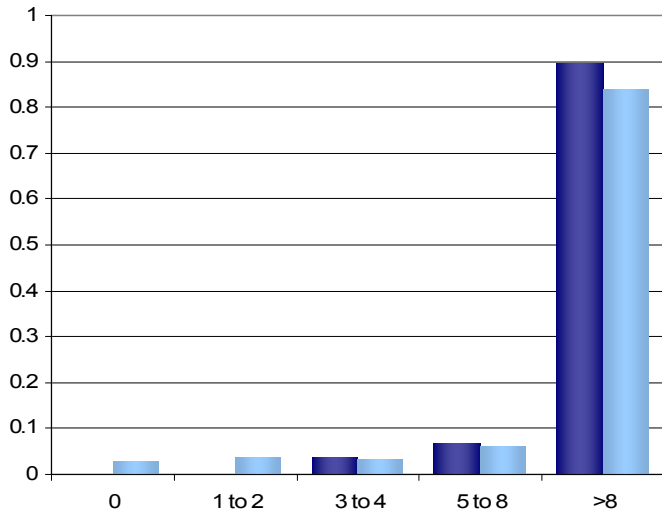
E2A (cluster + unclustered), CpG



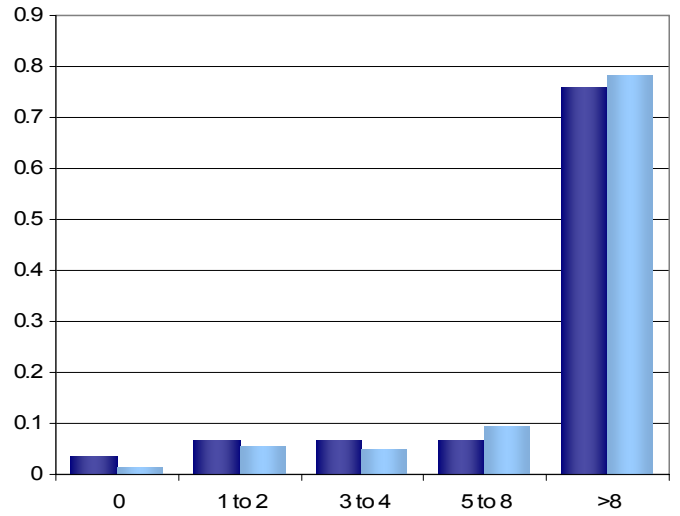
E2A (cluster + unclustered), CAC



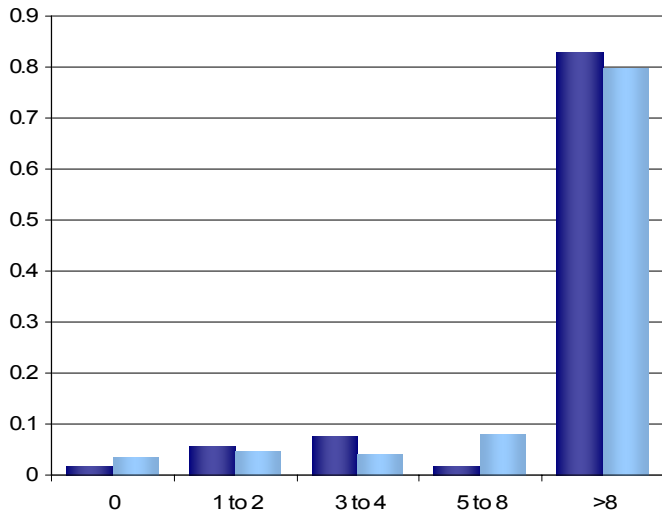
PBX1 from E2A-PBX1, CpG



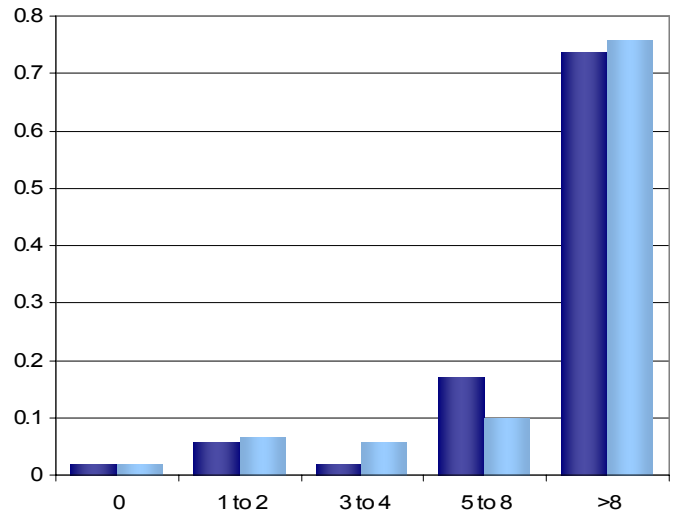
PBX1 from E2A-PBX1, CAC



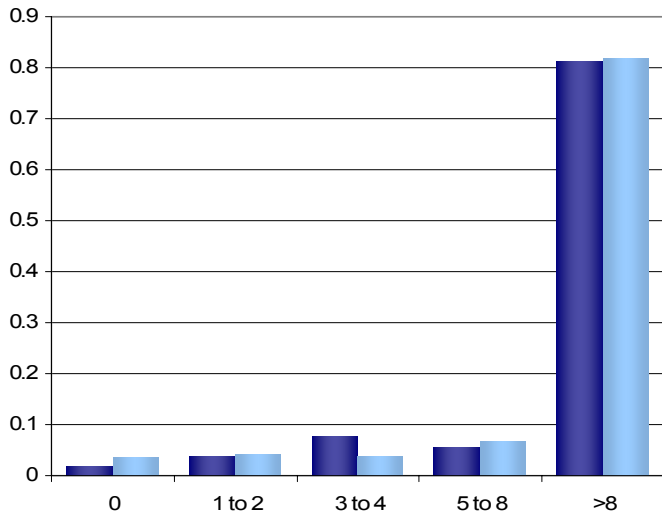
TEL from TEL-AML1, CpG



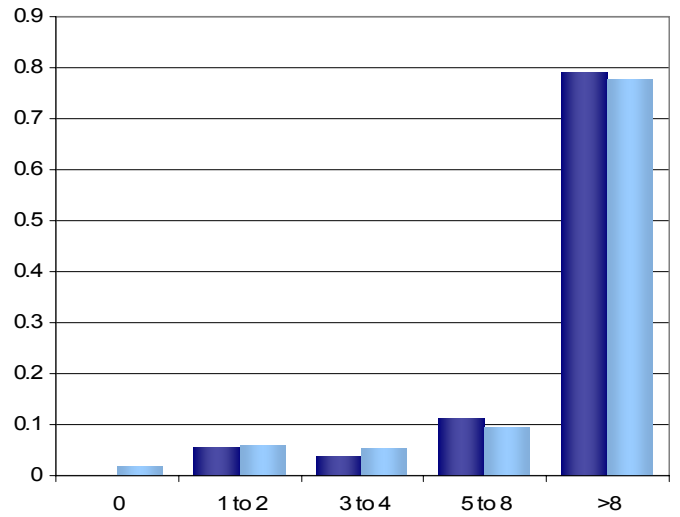
TEL from TEL-AML1, CAC



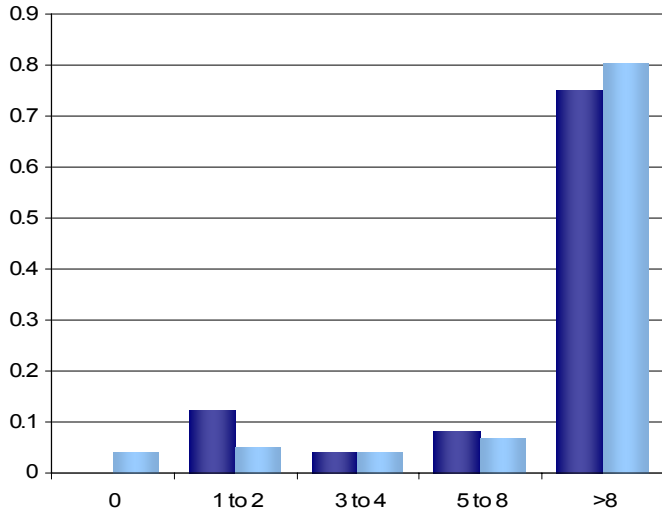
AML1 from TEL-AML1, CpG



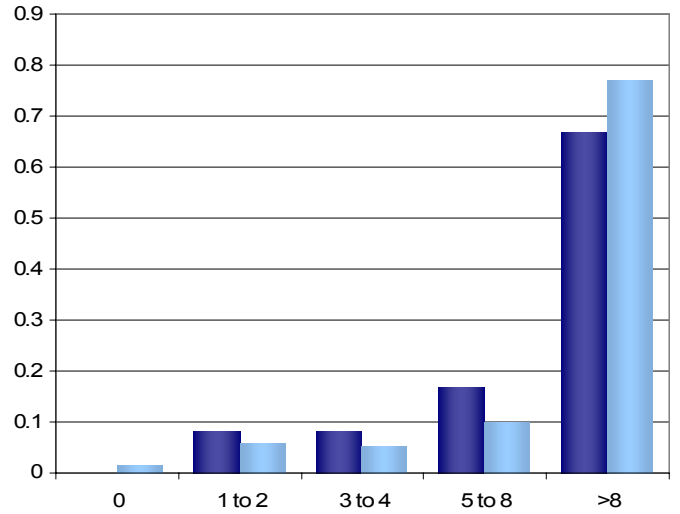
AML1 from TEL-AML1, CAC



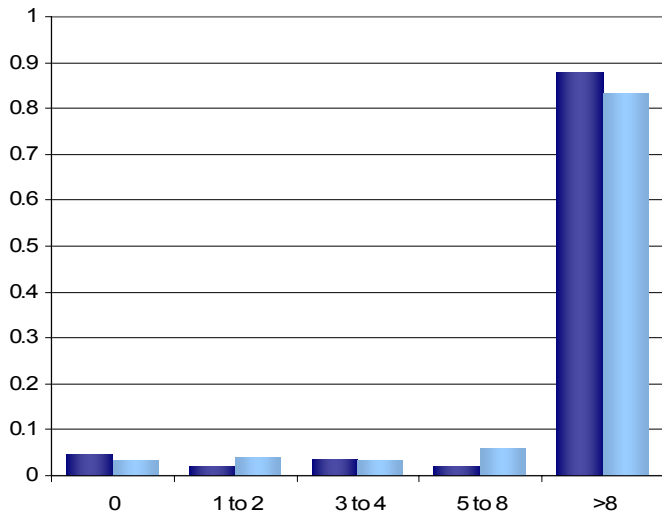
MLL translocations from AMLs, CpG



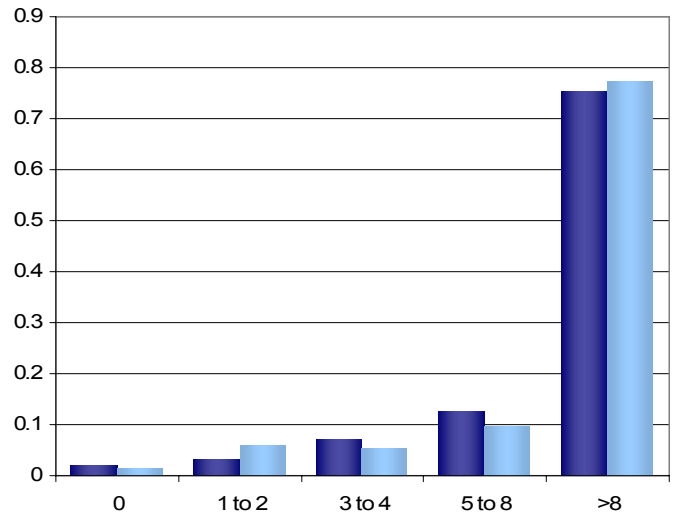
MLL translocations from AMLs, CAC



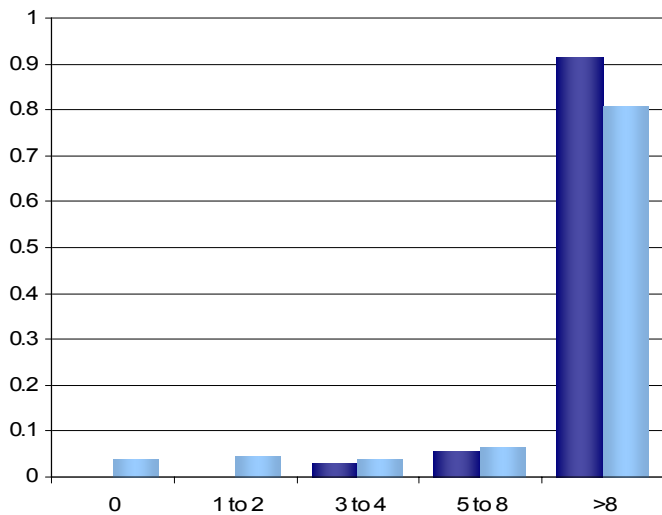
MLL translocations from ALLs, CpG



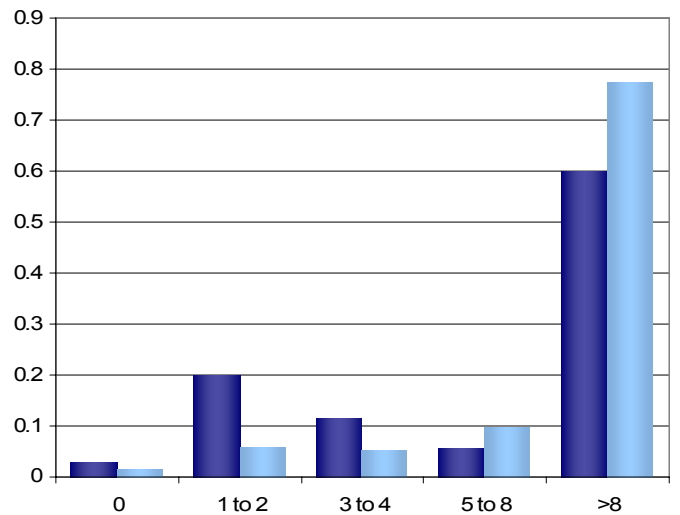
MLL translocations from ALLs, CAC



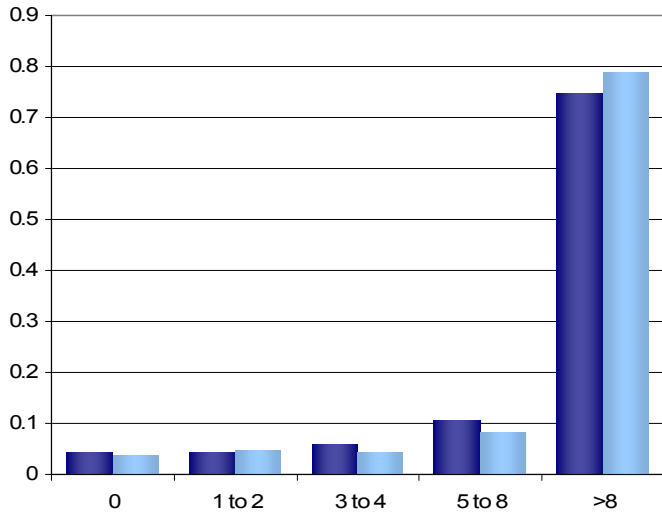
MLL translocations from AMLs and ALLs secondary to topoisomerase II-inhibitor therapy, CpG



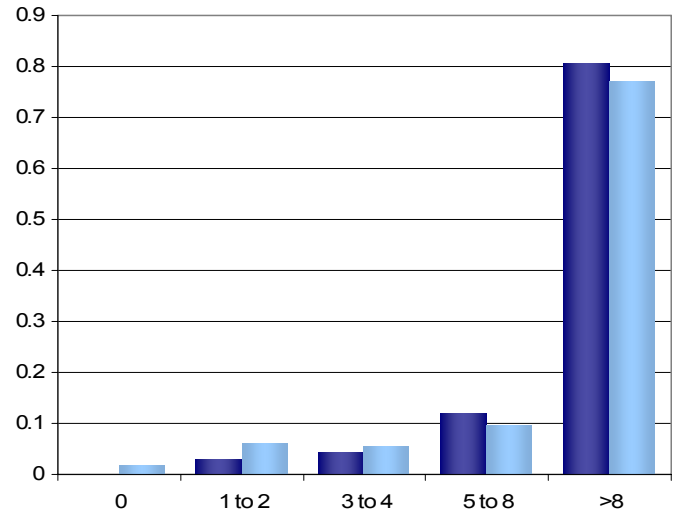
MLL translocations from AMLs and ALLs secondary to topoisomerase II-inhibitor therapy, CAC



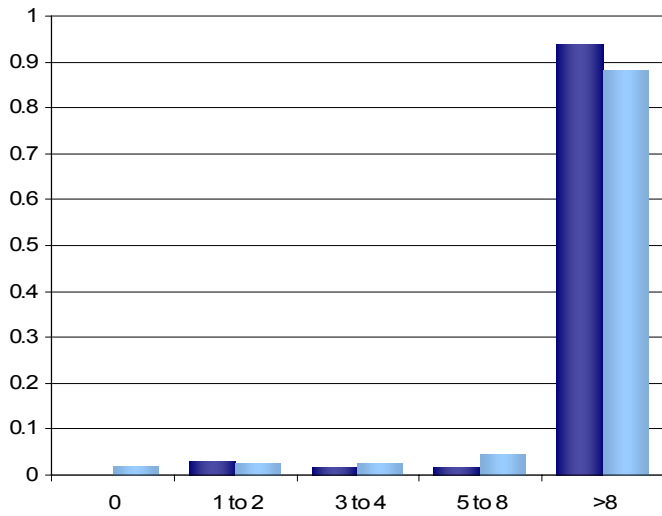
AML1 from AML1-ETO, CpG



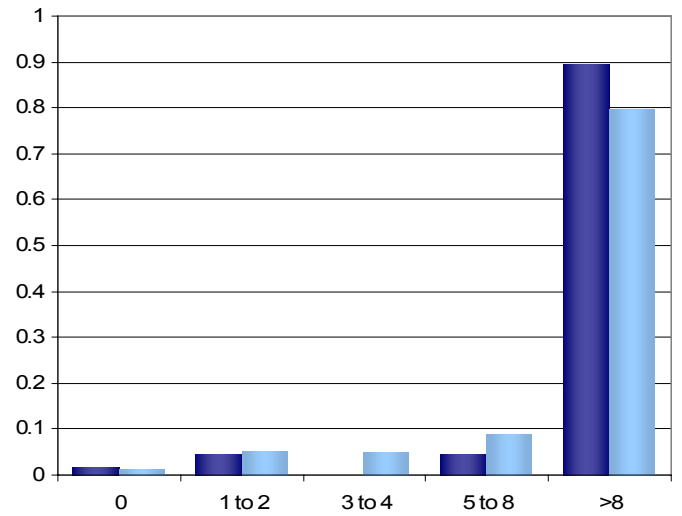
AML1 from AML1-ETO, CAC



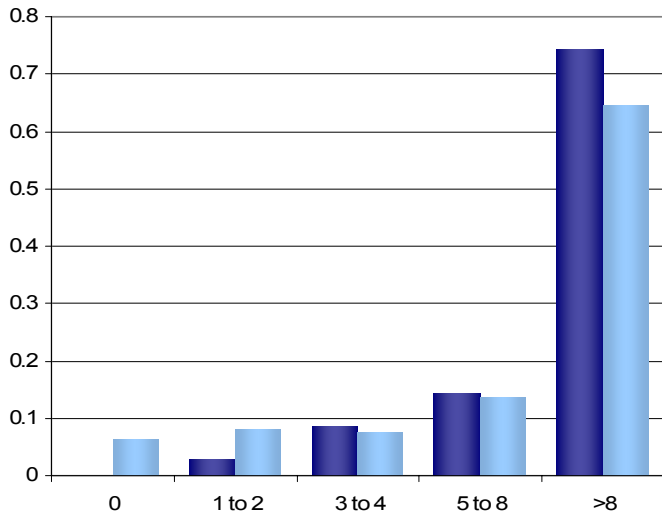
ETO from AML1-ETO, CpG



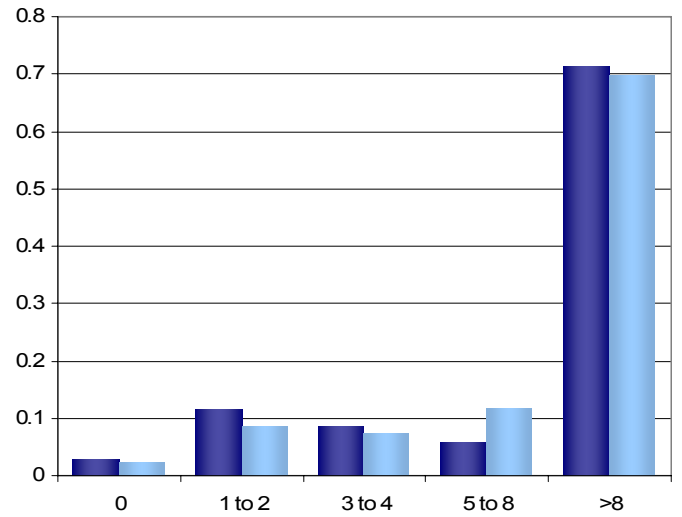
ETO from AML1-ETO, CAC



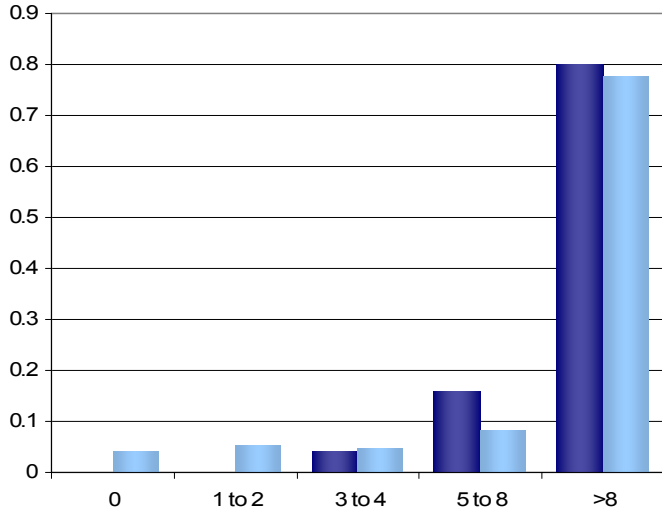
BCR from BCR-ABL from CML, CpG



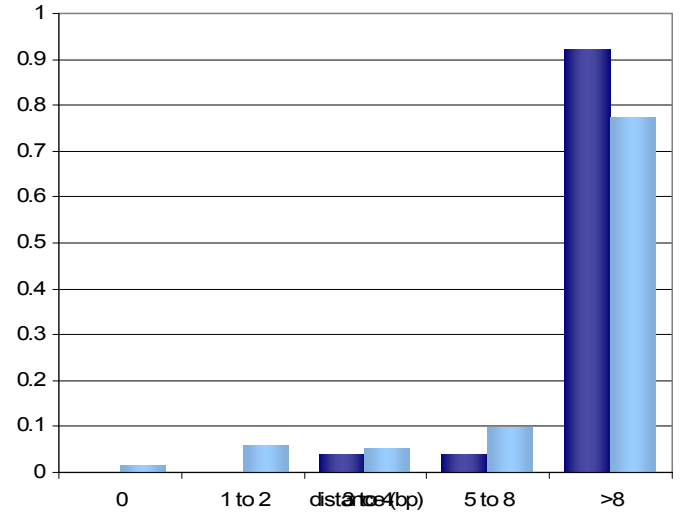
BCR from BCR-ABL from CML, CAC



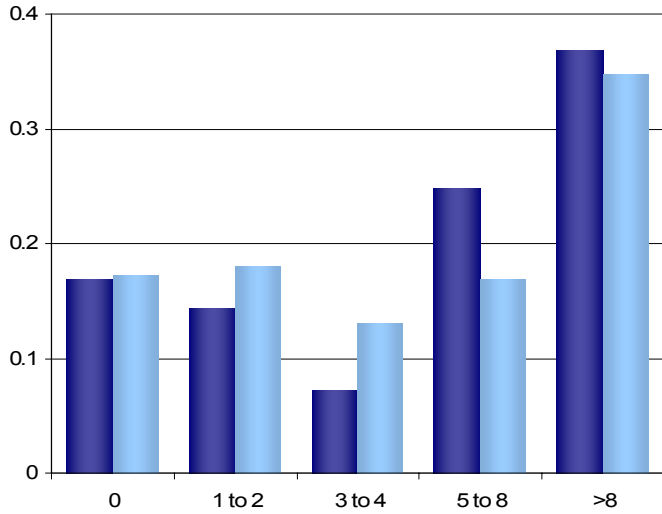
ABL from BCR-ABL from CML, CpG



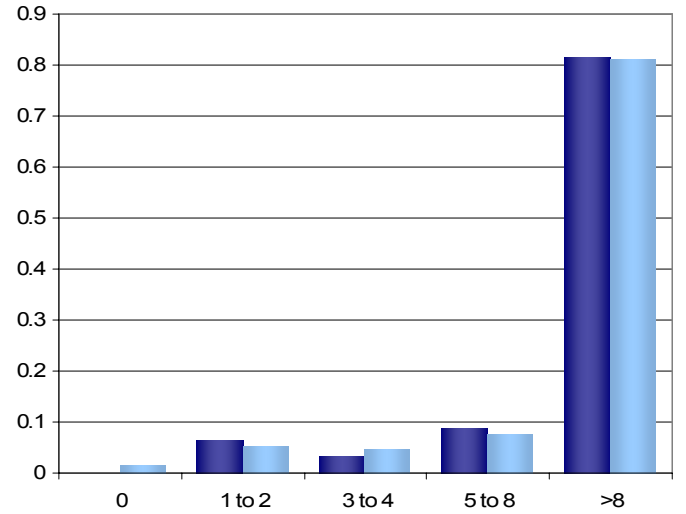
ABL from BCR-ABL from CML, CAC



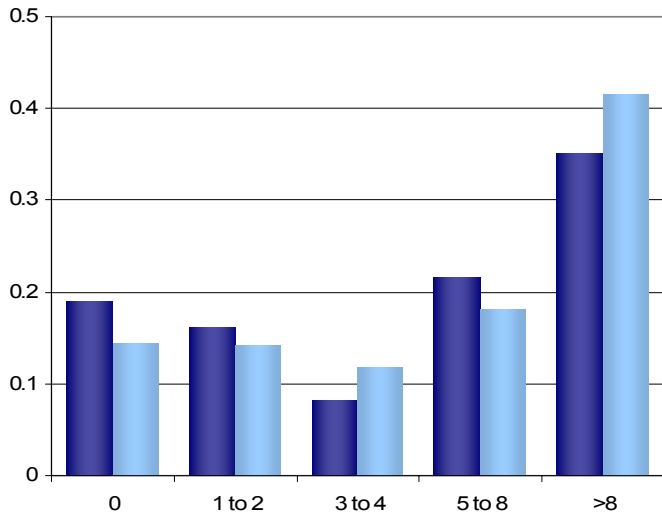
c-myc, CpG



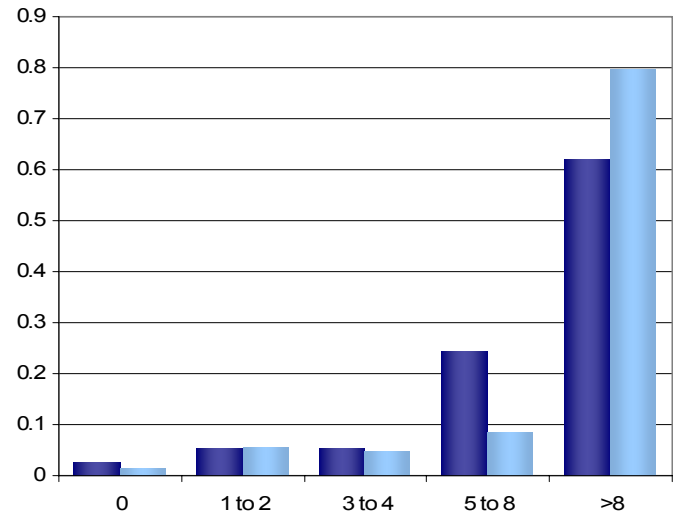
c-myc, CAC



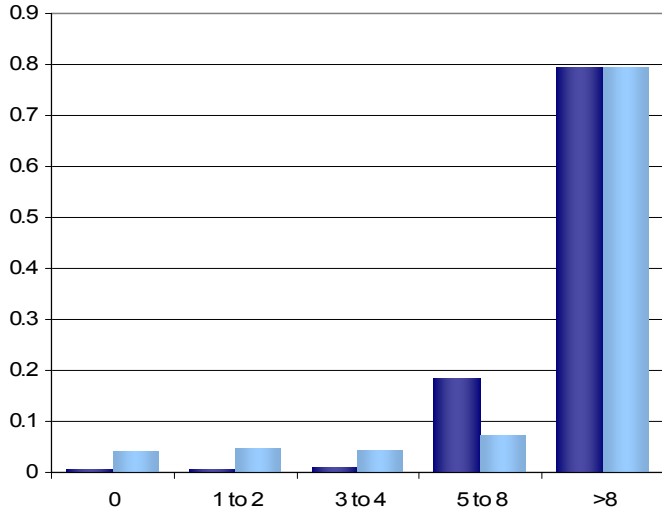
bcl-6, CpG



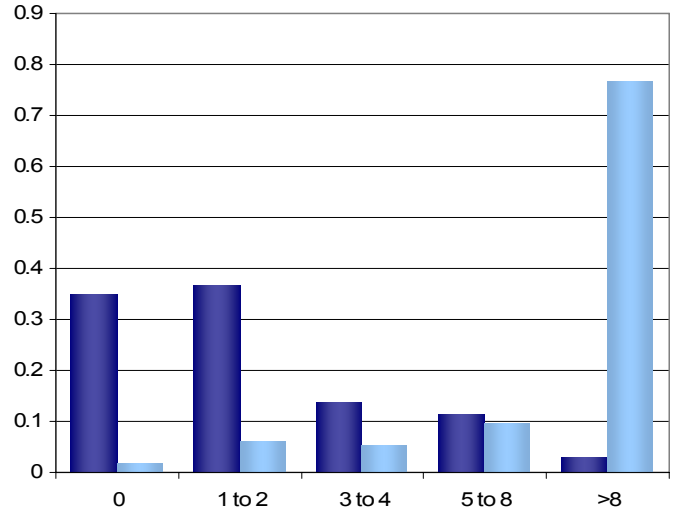
bcl-6, CAC



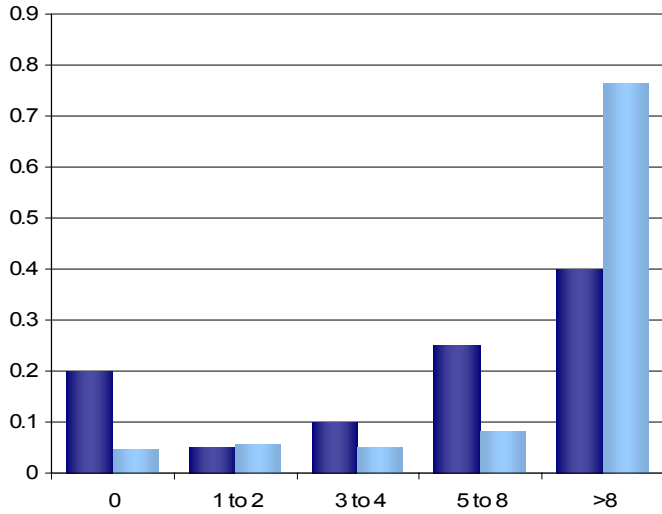
SCL-SIL, CpG



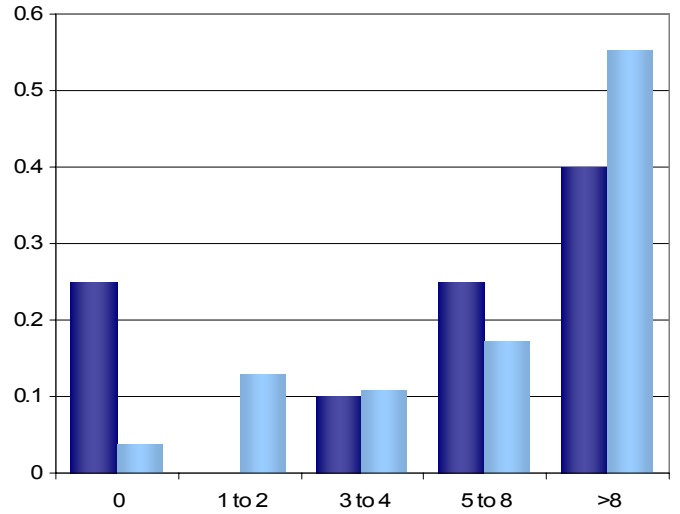
SCL-SIL, CAC



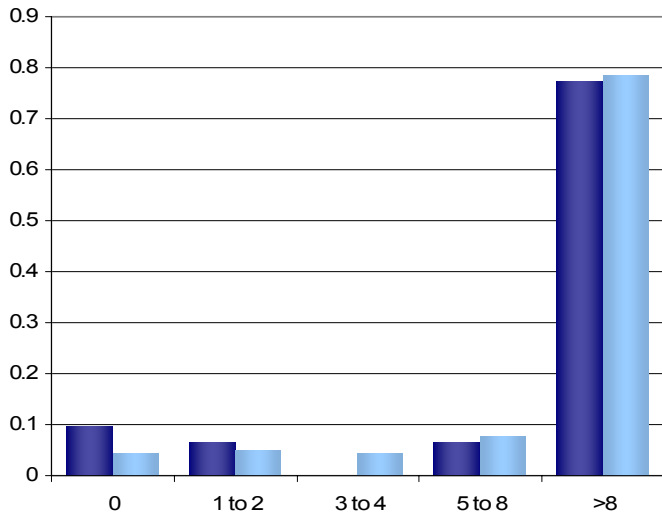
SCL from SCL translocations, CpG



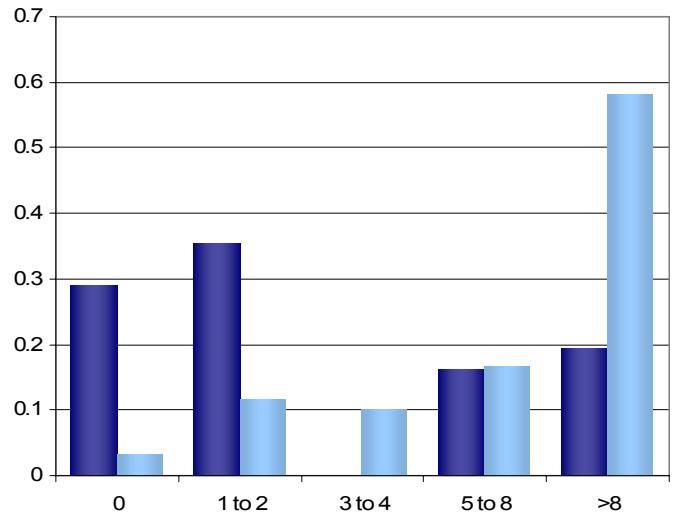
SCL from SCL translocations, CAC



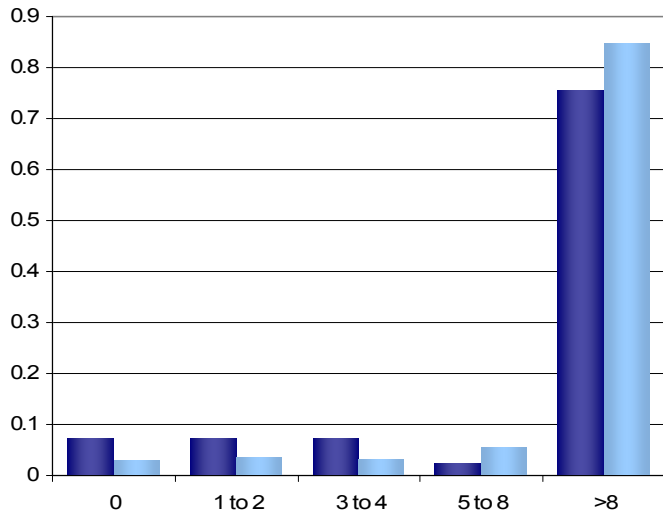
LMO2, CpG



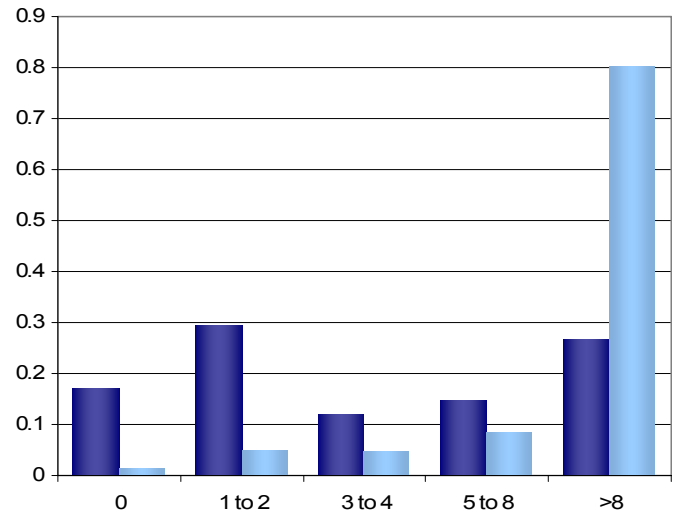
LMO2, CAC



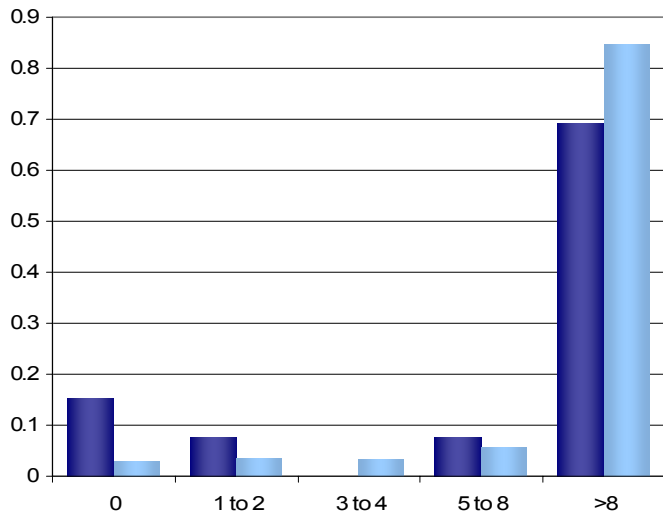
$\Delta p16$ from pre-B and pre-T leukemias, CpG



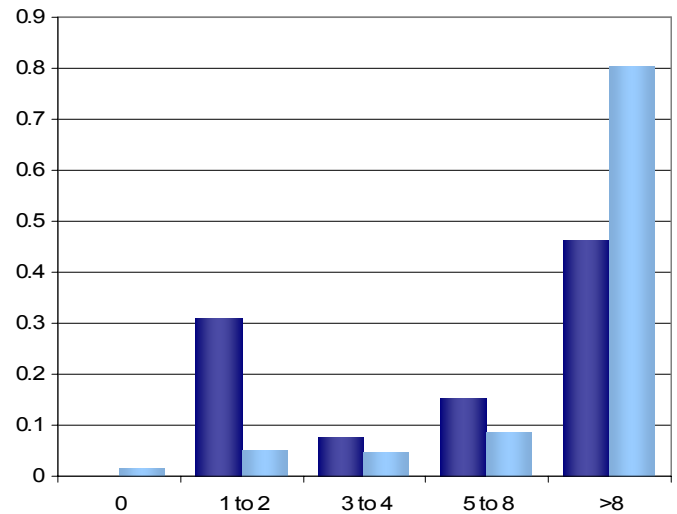
$\Delta p16$ from pre-B and pre-T leukemias, CAC



$\Delta p16$ from pre-B leukemias, CpG



$\Delta p16$ from pre-B leukemias, CAC



■ actual lymphoma breakpoints

■ breakpoints uniformly distributed from most 5' to most 3' actual breakpoints

Figure S3. Frequency-distance plots for CpG and CAC at various breakpoint regions.

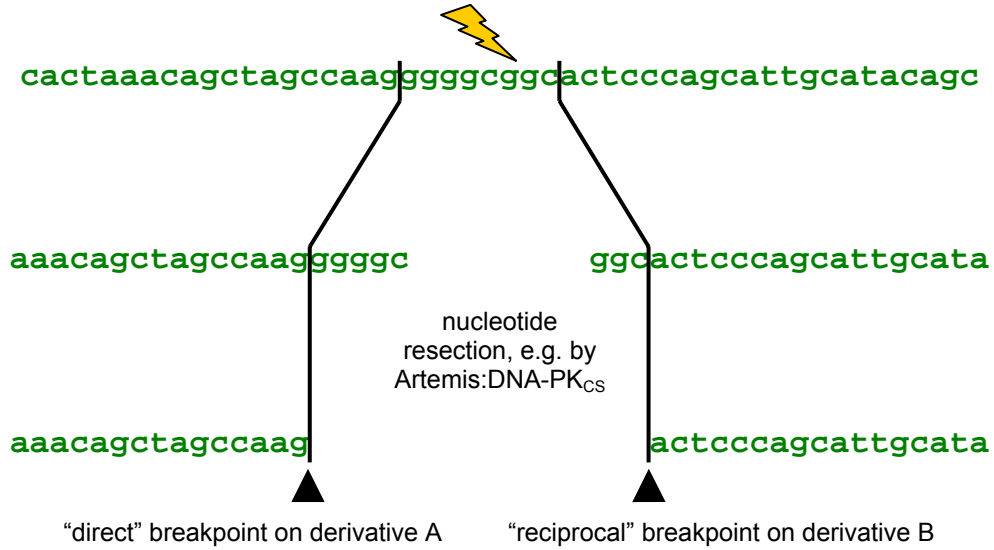
Proportions of breakpoints at distances of 0 bp, 1 to 2 bp, 3 to 4 bp, 5 to 8 bp, and greater than 8 bp from either CpG or CAC are graphed. The distribution for actual lymphoma or leukemia breakpoints is in dark blue, while that for a random distribution between the farthest breakpoints is in light blue.

The heights of the light blue bars, relative to one another, are affected by motif density and motif spacing within the region. For instance, *c-myc* and *bcl-6* are highly CpG-dense, thus a higher proportion of randomly-distributed breakpoints fall from 0 to 8 bp compared to *ABL*, which is not so CpG-dense. Also, many CpGs in *c-myc* and *bcl-6* are spaced very close together – within 4 bp of one another (e.g. CGNNCG, CGNNNCG, CGNNNNCG) – explaining why there is a greater proportion of randomly-distributed breakpoints 1-2 bp from a CpG than 3-4 bp from a CpG.

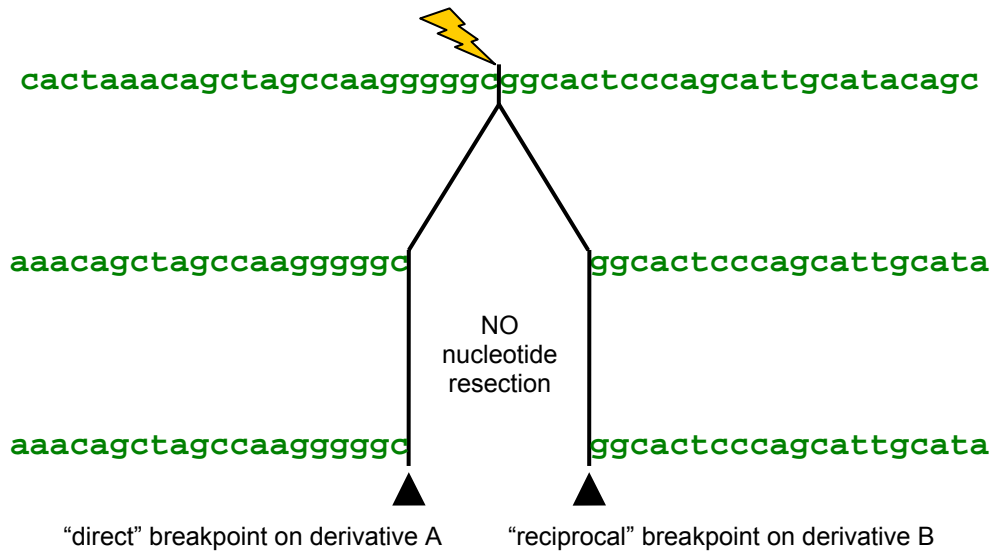
If the dark blue and light blue bars parallel one another, then the patient breakpoints appear random in their distribution relative to the specified motif. However, when they follow opposite trends, i.e. the light blue bars rise with increasing distance from the specified motif while the dark blue bars fall, then the breakage process appears to concentrate around the motif.

Breakpoints at *bcl-2*, *bcl-1*, and *E2A* appear to concentrate around CpG but not CAC. In contrast, breakpoints from pre-T rearrangements (*SCL-SIL* interstitial deletions, *SCL* translocations, *LMO2* translocations, and $\Delta p16$ interstitial deletions) are highly focused to CAC. The remainder, mostly from lymphoid-myeloid precursors and a couple from mature/activated B cells, do not focus to either CpG or CAC.

A. Balanced translocations indicate a window within which initial breakage likely occurred.



B. Balanced translocations with no nucleotide loss indicate a precise position at which initial breakage likely occurred.



C. Balanced translocations in the bcl-2 MBR.



D. Balanced translocations in the bcl-2 mcr.



E. Balanced translocations in the bcl-1 MTC.



F. Balanced translocations in the E2A cluster.

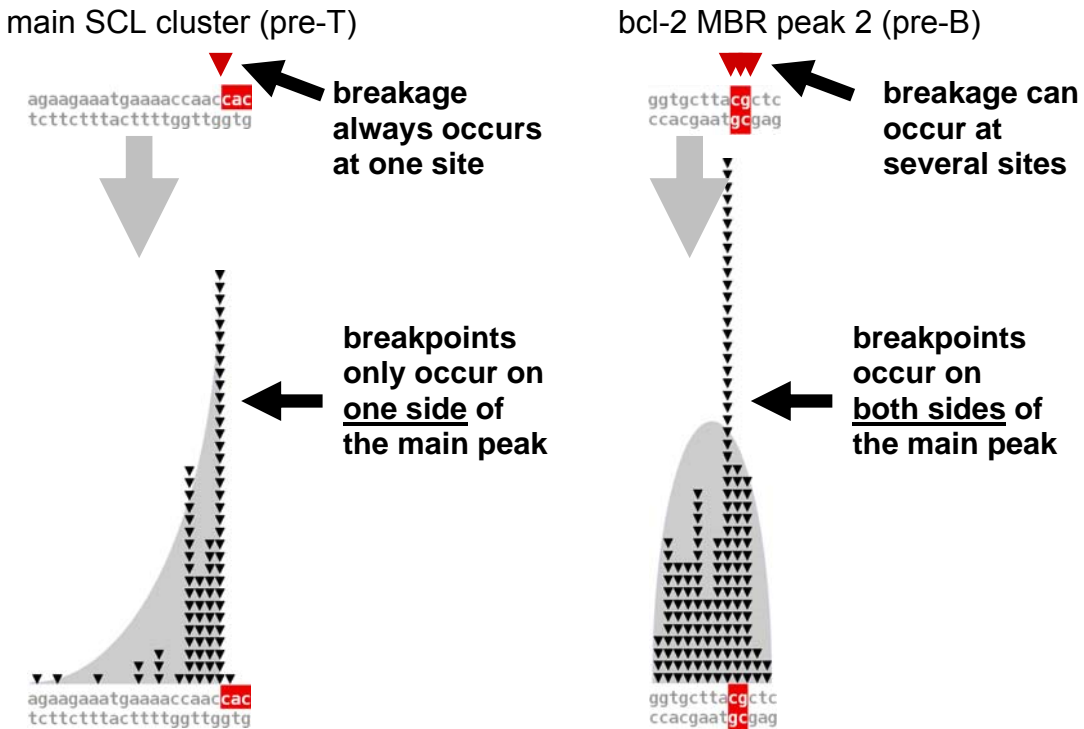


Figure S4. Balanced translocations

(A) and (B) diagram how the final breakpoint positions on the two derivatives of a balanced translocation derive from the initial DSB. When nucleotide resection occurs, as in (A), tracing these positions back to the original genomic sequence provides a window within which the initial DSB likely occurred. When there is no nucleotide resection, as in (B), both breakpoints map to the same position, and indicate the precise position where the initial DSB occurred.

Breakpoint distributions for balanced translocations are plotted similarly to Fig. 1 for the (C) bcl-2 MBR, (D) bcl-2 mcr, and to Fig. S2 for the (E) bcl-1 MTC and (D) E2A cluster. Each breakpoint triangle has a partner breakpoint triangle on the opposite strand. Open triangles denote balanced translocations where the direct and reciprocal breakpoints occurred at different positions, as in (A). Closed triangles denote balanced translocations where they occurred at the same position, as in (B). All ten cases without nucleotide resection are at CpGs, strongly supporting the notion that initial breakage occurs at or near CpG edges.

A. The breakpoint distribution at CpG-type rearrangement hotspots is suggestive of a structure-specific endonuclease.



B. The RAG complex is only expressed in the only stage during which CpG-type rearrangement hotspots occur.

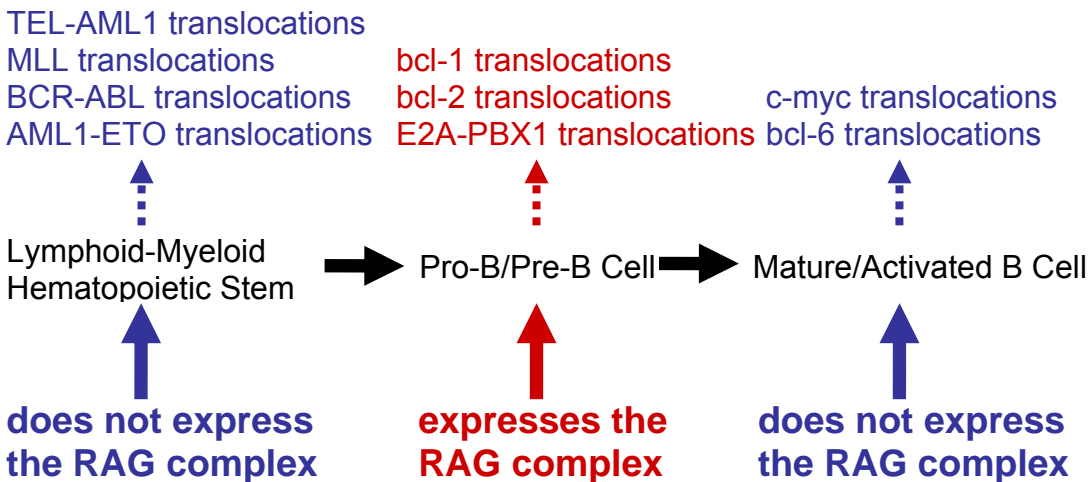


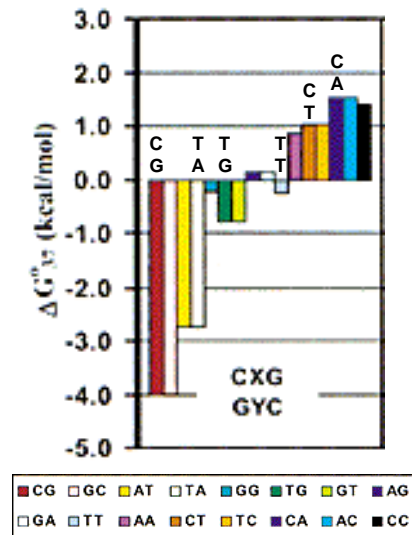
Figure S5. The RAG complex is the most logical candidate endonuclease.

In (A), comparison of V(D)J-type and CpG-type hotspots suggests they occur by different mechanisms. The main SCL cluster is an example of a V(D)J-type mechanism, in which a site-specific break, followed by limited nucleotide resection, results in a “spike-and-single-sided-tail” shape breakpoint distribution. The bcl-2 MBR peak 2 is an example of a CpG-type hotspot, which features a more bell-shaped distribution with breakpoints on either side of the main peak.

Such a distribution is indicative of heterogeneous initial breakage, typical of structure-specific nuclease activity. We have previously shown that the RAG complex possesses such an activity, and extend it to small deviations such as one-base mismatches, nicks, gaps, and flaps in Figure 5. Another reason to favor the RAG complex as the nuclease at CpG-type hotspots is that, in the B-cell lineage, it is only expressed in the only stage in which CpG-type hotspots are observed (B).

Additionally, the majority of der(14) or “top strand” breakpoints tend to be somewhat 5' of the CpG(s). What is likely occurring is that an initial breakage occurs at a CpG, and both the telomeric end (the left end in Fig. 2A), and the centromeric end (the right end in Fig. 1A) are recessed away from the CpG. The telomeric end becomes part of der(14), and the centromeric end part of der(18). However, the top strand only represents der(14) breakpoints, and therefore the top strand only shows the recessing on the telomeric or left end. Hence, if there were an equal number of der(18) bottom strand breakpoints, and they were reflected onto the top strand, then the distribution would be a symmetrical bell-shape centered at the CpG, rather than shifted slightly off.

Substrate	Nicking Efficiency
12 signal	66.7%
cacTgtg	1.8%
T-G mismatch	0.7%
T-T mismatch	0.8%
C-T mismatch	4.1%
C-A mismatch	4.0%

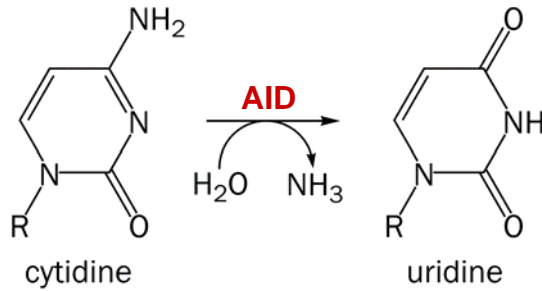


Peyret et al., 1999

Figure S6. The RAG complex recognizes the mismatch structure rather than its particular sequence.

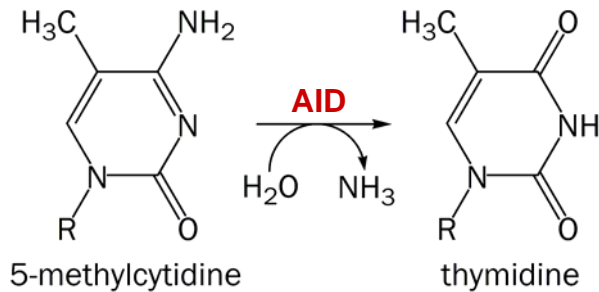
Quantitation shows that nicking efficiency increases with the ΔG of the mismatch (Peyret et al., 1999). That is, the RAG nicking is more efficient for pairs (matched or mismatched base pairs) that have a higher (more positive) ΔG , and RAG nicking is less efficient for pairs that have a lower (more negative) ΔG . This further demonstrates that the RAG complex recognizes the mismatch *structure* rather than its particular sequence.

A. Activity of AID at single-stranded cytidines.



not persistent due to efficient excision by UDG

B. Activity of AID at single-stranded 5'-methylcytidines.



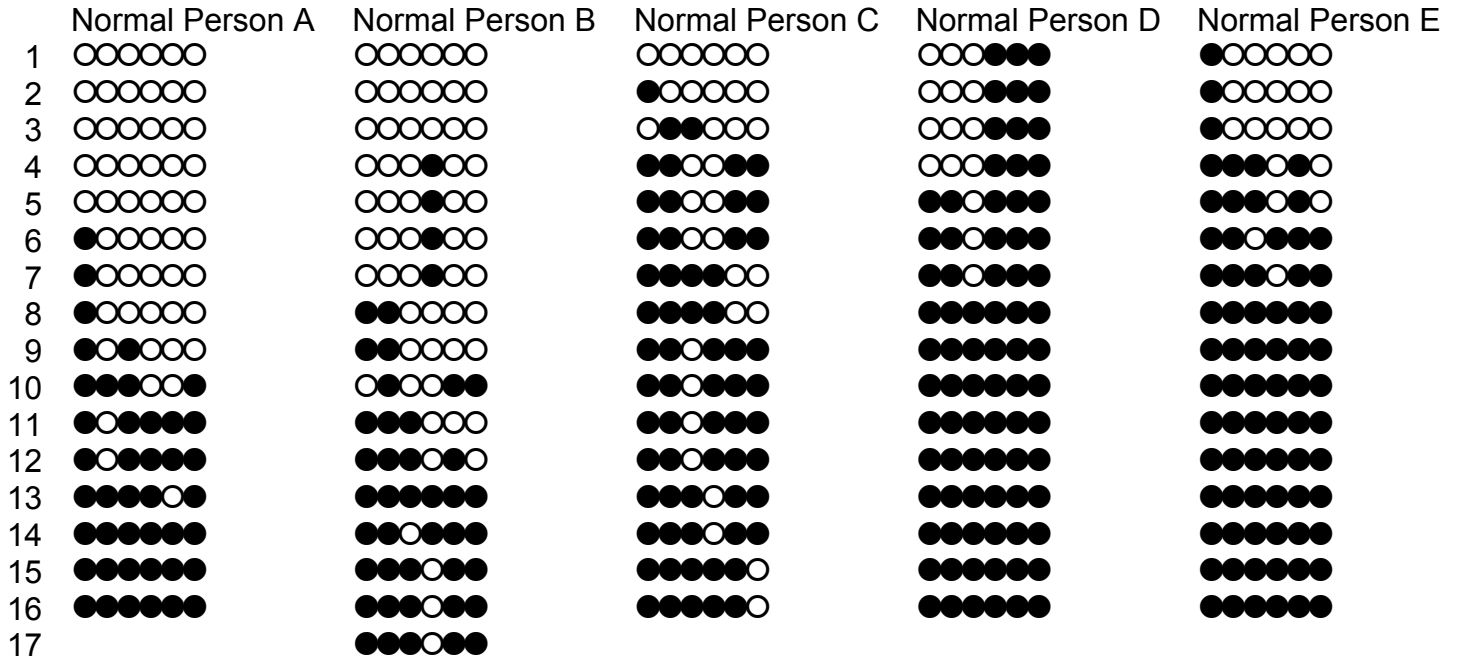
not efficiently excised by MBD4 and TDG (>2000-fold less efficient than UDG), therefore much more persistent

only at CpGs

Figure S7. Enzymatic activities of AID.

AID can deaminate both (A) unmethylated cytosines to uracils, and (B) methylated cytosines to thymines. However, in vitro repair at T:G mismatches is less efficient than repair at U:G mismatches, because TDG and MBD4, which excise thymines from T:G mismatches, are >2000-fold less efficient than UDG, which excises uracils. As a result, 5-methylcytosine deaminations are far more persistent than deaminations at unmethylated cytosines. Because methylcytosines in vertebrates only occur at CpGs, this explains why an abnormally high proportion of point mutations in p53 occur at CpGs, why vertebrates are depleted for CpGs in nonfunctional portions of the genome, and perhaps why breakpoints in pro-B/pre-B stage translocations are extremely focused to CpGs.

A. *bcl-2* MBR.



B. *bcl-1* MTC

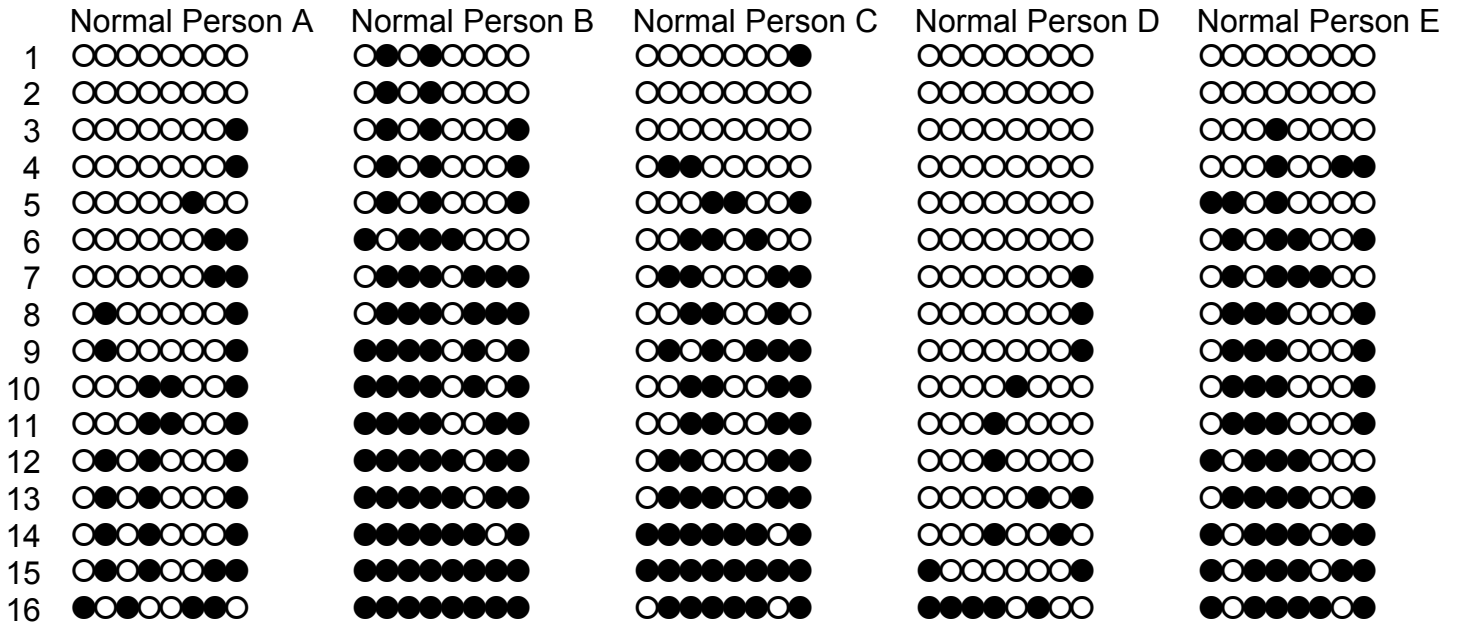


Figure S8. Methylation status of the *bcl-2* MBR and *bcl-1* MTC in FACS-sorted pre-B cells from five normal individuals.

Genomic DNA of FACS-sorted pre-B cells from the bone marrow of five normal individuals (without any evidence of leukemia or lymphoma) was bisulfite sequenced to determine the methylation pattern at the (A) *bcl-2* MBR and (B) *bcl-1* MTC. Each line of circles represents the CpGs from a single bisulfite sequenced molecule, where a filled circle represents a methylated CpG and an empty circle represents an unmethylated CpG. In (A), the five CpGs within the MBR are represented by the second through sixth circles, with the first circle being the closest CpG telomeric of the MBR. In (B), the second through seventh CpGs within the MTC are represented by the first through sixth circles, with the seventh circle being the closest CpG centromeric of the MTC. The first CpG of the MTC appears to be a C→A/G→T SNP in almost all of the samples and thus is not plotted.

Table S1. Statistical analysis of CpG and CAC proximity in various translocations.

Locus	Percentage at CpG	Actual Average Distance to CpG (bp)	Random Average Distance to CpG (bp)	p-value at CpG (binomial distribution)	p-value proximity to CpG (Student's t-test)	p-value proximity to CpG (Mann-Whitney U-test)	Percentage at ∇ CAC	Actual Average Distance to ∇ CAC (bp)	Random Average Distance to ∇ CAC (bp)	p-value at ∇ CAC (binomial distribution)	p-value proximity to ∇ CAC (Student's t-test)	p-value proximity to ∇ CAC (Mann-Whitney U-test)
bcl-2 MBR	43%	4.39	11.2	$1.8 \cdot 10^{-96}$	$2.3 \cdot 10^{-50}$	$1.2 \cdot 10^{-42}$	0.21%	21.2	22.1	1.0	0.45	0.27
bcl-2 icr	73%	0.818	21.5	$1.5 \cdot 10^{-8}$	$1.2 \cdot 10^{-12}$	$3.4 \cdot 10^{-7}$	0%	51.4	52.1	1.0	0.79	0.49
bcl-2 mcr	74%	0.632	40.4	$5.4 \cdot 10^{-18}$	$2.3 \cdot 10^{-31}$	$7.6 \cdot 10^{-13}$	0%	10.2	30.4	1.0	$5.6 \cdot 10^{-5}$	$5.6 \cdot 10^{-5}$
bcl-2 unclustered	19%	45.1	62.2	0.0019	0.0016	0.024	0%	44.0	37.9	1.0	0.79	0.82
bcl-1 MTC	37%	2.55	7.80	$7.0 \cdot 10^{-9}$	$1.8 \cdot 10^{-13}$	$1.1 \cdot 10^{-12}$	0%	21.8	22.8	1.0	0.68	0.56
bcl-1 unclustered	20%	8.80	45.3	0.083	0.00048	0.00090	0%	38.3	32.8	1.0	0.64	0.69
E2A	63%	1.08	3.25	0.00011	0.00020	0.00052	0%	11.3	17.2	1.0	0.068	0.019
PBX1	0%	93.7	64.3	1.0	0.96	0.96	3.4%	30.1	37.0	0.37	0.21	0.27
TEL in TEL-AML1	1.9%	41.1	41.6	0.86	0.66	0.66	1.9%	38.5	33.4	0.63	0.61	0.55
AML1 in TEL-AML1	1.9%	61.7	61.7	0.85	0.49	0.41	0%	42.7	37.5	1.0	0.87	0.89
MLL in primary ALL	4.7%	79.0	84.9	0.16	0.65	0.58	2.0%	32.2	35.6	0.37	0.26	0.23
MLL in primary AML	0%	64.4	85.4	1.0	0.41	0.47	0%	23.4	33.9	1.0	0.12	0.094
AML1 in AML1-ETO	4.5%	48.6	46.7	0.45	0.61	0.70	0%	38.0	35.7	1.0	0.77	0.67
ETO in AML1-ETO	0%	103	88.9	1.0	0.99	0.99	1.5%	52.8	41.6	0.62	0.99	1.0
BCR in CML	0%	28.7	21.4	1.0	0.97	0.96	4.3%	26.5	26.5	0.27	0.35	0.43
ABL in CML	0%	63.0	50.3	1.0	0.91	0.89	2.9%	30.0	25.1	0.56	0.68	0.72
c-myc	17%	8.85	8.90	0.59	0.71	0.75	0%	39.4	38.7	1.0	0.33	0.36
bcl-6	19%	8.32	10.9	0.28	0.13	0.15	1.7%	41.2	57.3	0.41	0.027	0.019
SCL-SIL	0.48%	16.6	60.9	1.0	$8.4 \cdot 10^{-15}$	$5.9 \cdot 10^{-26}$	35%	2.16	37.3	$7.6 \cdot 10^{-75}$	$3.7 \cdot 10^{-185}$	$8.7 \cdot 10^{-108}$
lymphoid Δ p16	7.3%	63.7	96.7	0.13	0.048	0.11	17%	11.3	46.0	$1.5 \cdot 10^{-6}$	$3.4 \cdot 10^{-20}$	$4.4 \cdot 10^{-14}$
SCL translocations*	20%	32.1	53.5	0.014	0.00029	0.0012	25%	10.2	12.5	0.00070	0.015	0.040
LMO2*	9.7%	26.2	47.1	0.16	0.060	0.036	29%	4.58	16.3	$4.8 \cdot 10^{-7}$	$3.6 \cdot 10^{-12}$	$1.6 \cdot 10^{-9}$

Statistical measures of CpG and CAC proximity are compared. Values related to CpG-type translocations in pro-B/pre-B stage cells are highlighted in cardinal. Values related to V(D)J-type translocations in immature lymphocytes are highlighted in gold. p-values are calculated with "random" defined as a uniform breakpoint distribution from the most 5' to the most 3' breakpoint, and depend on the number of breakpoints as well as the degree of clustering and the regional motif density, therefore precise values should not be compared between different clusters. In highly-clustered translocations, i.e. CpG-type or V(D)J-type, nearby motifs may score significantly but not to the same level as CG or CAC, respectively. More complete tables, including all dinucleotide combinations and individual analysis of top and bottom strands, can be made available upon request. Calculations are described briefly in the Experimental Procedures, and in more detail in the supplement.

* Mixture of signal and coding joint events, therefore calculated for ∇ CAC and GTG ∇ ; see supplemental text for details.