Am J Blood Res 2011;1(1):1-12
www.AJBlood.us /ISSN: 2160-1992/AJBR1103001

Original Article

Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer

Eric Kowarz[1], Jennifer Merkens[1], Michael Karas[2],Theo Dingermann[1], Rolf Marschalek[1]

[1]Institute of Pharmaceutical Biology, DCAL, ZAFES, CEF, JWG-University of Frankfurt, Biocenter, Max-von-Laue-Str.9, D-60438 Frankfurt, Main, Germany; [2]Institute of Pharmaceutical Chemistry, ZAFES, CEF, JWG-University of Frankfurt, Biocenter, Max-von-Laue-Str. 9, D-60438 Frankfurt, Main, Germany.

Received March 14, 2011; accepted April 4, 2011; Epub April 7, 2011; published June 1, 2011

Abstract: So far, about 800 different chromosomal translocations have been characterized in hemato-malignant and solid tumors. Chromosomal translocations mostly result in the expression of chimeric fusion proteins associated with enhanced proliferation and/or malignant transformation. Here, we demonstrate that genes frequently involved in such genetic rearrangements exhibit a unique feature: premature transcriptional termination. These early-terminated RNA molecules have an abundance of 10-20% when compared to their cognate full-length transcripts. They exhibit an unsaturated splice donor site that gives rise to trans-splicing events, leading to RNAs displaying exon repetitions or chimeric fusion RNAs. These arbitrary fusion RNAs mimic the presence of a chromosomal translocation in genetically unaffected cells. Based on our and published data, we propose the hypothesis that these artificial "chimeric fusion transcripts" may influence DNA repair processes, resulting in the generation of *de novo* chromosomal translocations. This idea provides a rational explanation why different individuals suffer from nearly identical genetic rearrangements.

Keywords: Acute leukemia, chromosomal translocations, early-terminated transcripts, transsplicing, DNA repair, RNA-templated DNA repair

Introduction

Chromosomal translocations (CT) – associated with different human cancer types - arise due to DNA double-strand breaks and subsequent DNA repair processes, predominantly executed by the non-homologous end joining (NHEJ) pathway [1, 2]. DNA double-strand breaks occur more frequently in specific chromosomal regions that exhibit an intrinsic genetic instability. Such a property was exemplarily investigated for the human *MLL* gene and could be linked to specific chromatin features, e.g. SAR/MAR structures [3-6], DNase I hypersensitive sites [7], Topoisomerase II binding sites [8,9], gene internal transcription initiation [10] and site-specific DNA cleavage during early apoptosis [11-13]. These features - or any combination thereof - seem to define recombination hot spots that seem not randomly distributed in the human genome. Recombination events at those sites may result in genetic rearrangements with the potential to generate novel oncogenes and the onset of pre-cancerous cells. Potent oncogenes arise when proto-oncogenes are recombined with cell-type specific enhancer regions (type I CTs) or by the creation of chimeric fusion genes (type II CTs). In the hematopoietic system, type I rearrangements are associated with a lymphoma disease phenotype, while type II rearrangements are associated with more aggressive leukemias. Nearly 800 different chromosomal rearrangements have been characterized that are associated with the development of solid or hematological malignancies. Noteworthy, 83 different genes have been yet identified in solid tumors [14], a list which is rapidly growing due to the many efforts to precisely characterize cancer genomes of different tumor entities.

However, the question remains why different individuals suffer from nearly "identical" genetic rearrangements, described as "recurrence".

Here, we asked the question whether chromosomal translocations are generated by pure accident combined with subsequent positive selection processes, or whether there exists a molecular mechanism that could explain the phenomenon of recurrence as described for different human cancers.

A first hint comes from older experimental observations where specific chimeric fusion mRNA transcripts - known from chromosomal translocations - were identified at the RNA level in cells deriving from healthy individuals. For example, the *MLL-AF4* chimeric fusion transcript and *MLL* -partial tandem duplications (PTDs) were diagnosed at low levels in hematopoietic stem or cord blood cells of healthy individuals [15-17]. Similar observations were made for chimeric fusion transcripts of *BCR-ABL* [18, 19], *TEL-AML1* and *AML1-ETO* [20, 21], *PML-RAR* [22], *NPM-ALK* and *ATIC-ALK* [23, 24]. Most importantly, the investigated individuals were without any sign of a disease phenotype and displayed no genomic rearrangement of the corresponding gene loci. These findings were controversially discussed in the scientific community; however, different laboratories provided conclusive experimental evidence for their existence. Thus, these experimental observations were classified as biological curiosity, with no explanation for their biological function or any other convincing explanation why they occur at all.

Another important observation was the discovery of chromosome territories in interphase nuclei [25-28]. Based on the 3-dimensional architecture of interphase chromosomes, it was experimentally demonstrated that genes known to participate in illegitimate recombination events are located in close spatial proximity in the nucleus. These are e.g. *BCR* and *ABL*, *PML* and *RAR* [29, 30], *MYC*, *BCL2* and *IGH* [31]. Some of these genes were even transcribed in the same transcription factor (e.g. *MYC* and *IGH* [32]; for review see Gingeras, 2009 [33]). Similarly, *MLL* and at least *AF4* and *ENL* are closely located in the 3-dimensional space of nuclei [34]. All these findings may suggest that chromosomal translocations are not created by chance rather than by specific properties of the involved genes, and in particular by their specific localization within the 3-dimensional space of the nucleus.

For the purpose of our studies we chose the well-characterized human *MLL* gene. This gene is frequently rearranged with a large variety (n>60) of different translocation partner genes (TPGs). About 40% of the yet characterized 64 TPGs were recurrently diagnosed in *MLL*-mediated leukemias [35]. However, the most frequently diagnosed fusion partners are *AF4* (42%), *AF9* (16%), *ENL* (11%) and *ELL* (4%) which account for the large majority of all diagnosed *MLL* rearrangements in human acute leukemias. Therefore, our studies focused on these 5 genes to investigate potential mechanisms that could provide a rational explanation for the phenomenon of recurrence that is typically found in all human cancers.

Material and methods

Source of samples and RNA isolation

PBMCs were prepared from peripherial blood of the healthy volunteers using Ficoll-Reagent (GE-Healthcare) according to the manufacturers instruction. About 1 x 10^7 cells were used for RNA preparation using the RNeasy-Kit (Qiagen).

cDNA sysnthesis and PCR experiments

Five µg total RNA were reverse transcribed (Invitrogen) with either random hexamer primers (for inverse RT-PCR experiments), oligo-dT anchored primer (obtained from the GeneRacer-Kit from Invitrogen for 3'-RACE) or gene specific primers (AF4.Y32 5'-CG-AT-GA-CG-TT-CC-TT-GC-TG-AG-AA-TT-TG-AG-TG-AG-3' and MLL.CR2 5'-GT-CC-CA-GG-CA-CT-CA-GG-GT-GA-TA-GC-TG-TT-TC-GG-3' for MLL-AF4 and AF4-MLL nested PCR; ALK.E20R1 5'-GG-TT-GT-AG-TC-GG-TCA-TG-AT-GG-TC-GA-GGT-3' for NPM-ALK nested PCR).

Using the Genracer3' and the Generacer3'-nested primers 3'-RACE nested PCR was performed for MLL, AF4 (AF4.E3F 5'-CG-GA-GG-AC-TA-TC-GA-CA-GC-AG-AC-CT-TT-GAA-3' and AF4.X3 5'-AC-CA-AA-CT-GA-AG-AT-GC-CT-TC-TC-AG-TC-AG-TT-GAG-3'), AF9 (AF9.E3F 5'-AT-CT-GG-GT-AT-GC-TG-GT-TT-CA-TT-TT-GC-CA-AT-3' and AF9.E4F 5'-AT-CC-AC-CA-GT-GA-AT-CA-CC-TC-CG-CT-GT-GA-AA-3'), ENL (ENL.E1F1 5'-GG-CG-GC-GC-TT-GA-CA-GA-CA-AT-GA-GG-3' and ENL.E1F2 5'-GG-GC-GC-CA-GC-CA-TG-GA-CA-AT-CAG-3') and ELL (ELL.E1F 5'-AA-GG-AG-GA-TA-GG-AG-CT-AC-GG-GC-TG-TC-GT-3' and ELL.E2F 5'-CC-AT-CT-AT-CC-GA-TT-TC-AA-GG-AA-GC-CA-AGG-3'). The second round of PCR was subjected to gel-

electrophoresis and resulting bands were cut out from the gel, extracted (Qiagen gel extraction kit) and cloned into the pGEM-T vector (Promega) for sequence analysis.

QRT-PCR experiments

Quantification of early-terminated transcripts was performed on a MiniOpticon Realtime PCR Detection System (BioRad): For the MLL gene: MLL.E7F2 5'-AA-GC-CT-AC-CT-GC-AG-AA-GC-AA-3', MLL.I8R1 5'-TC-CT-GC-TT-TC-AA-AT-GC-TG-TTT-3', MLL.E5F 5'-TA-AG-CC-CA-AG-TT-TG-GT-GG-TC-3' and MLL.E8R 5'-CT-TG-GG-CT-CA-CT-AG-GA-GT-GG-3'; for the AF4: gene: AF4.E3F, AF4.I3R 5'-GC-CA-AA-AA-GA-AT-TC-CC-CC-TA-3' and AF4.E5R 5'-AA-GG-AA-AC-TT-GG-AT-GG-CT-CA-3'; for the AF9 gene: AF9.E4F 5'-CA-AC-AA-CC-CC-AC-AG-AG-GA-CT-3', AF9.E5bR 5'-GG-AT-TC-GA-AT-TC-TT-GC-TC-TG-TC-3', and AF9.E6R 5'-GC-AG-GA-CT-GG-GT-TG-TT-CA-GA-3'; for the ELL gene: ELL.E1F 5'-GA-GA-GA-TG-GT-CG-CA-AG-AT-GG-3' and ELL.I2R 5'-CT-AT-CC-TG-GG-GG-CC-TA-GA-AC-3', and ELL.E3R 5'-AC-TG-CT-GG-AT-GC-AG-TC-GA-AG-3'; for the ENL gene: ENL.E2F 5'-CG-TC-CA-GG-TG-AG-GT-TA-GA-GC-3', ENL.E3bR 5'-TG-TT-CC-CA-GC-AG-AT-GT-CA-AG-3' and ENL.E4R 5'-GT-GG-GG-TT-GT-TG-AA-GG-TG-AG-3'.

Inverse RT-PCR experiments

The following primers were used for inverse PCR experiments: MLL.R1 5'-GA-CA-TT-CC-CT-TC-TT-CA-CT-CT-TT-TC-CTC-3' and MLL.F1A 5'-CC-AC-CT-AC-TA-CA-GG-AC-CG-CA-AG-AA-AA-3'; AF4.E3F and AF4.E3R 5'-CA-TG-GA-GA-CT-TG-GC-AT-TG-GT-TC-AG-TT-CT-TG-3'; AF9.E5F 5'-GA-TC-CC-AA-TG-AT-TC-AG-AT-GT-GG-AG-GA-GA-AT-3' and AF9.E5R 5'-TG-TG-AG-GC-TT-TG-AA-AA-AC-TG-GT-AC-TA-CT-GC-TG-3'; ELL.E2F and ELL.E2R 5'-TG-AA-AT-CG-GA-TA-GA-TG-GC-CT-CA-GT-GA-AA-CA-3'; ENL.E2F and ENL.E2R 5'-CA-AA-CA-CC-AT-CC-AG-TC-GT-GA-GT-GA-ACC-3'. The PCR-products were subjected to gel electrophoresis and the resulting bands were cut out of the gel, extracted (Qiagen gel extraction kit) and cloned into the pGEM-T vector (Promega) for sequencing.

Nested PCR experiments

MLL-AF4 fusion transcripts were detected with nested PCR using the primers MLL.F1A with AF4.Y30 5'-GG-TT-TT-GG-GT-TA-CA-GA-AC-TG-AC-AT-GC-TG-AG-AG-3' (first PCR) and MLL.NF1 5'-CC-AA-AA-CC-AC-TC-CT-AG-TG-AG-CC-CA-AGA-3' with AF4.Y29 5'-GT-AT-TG-CT-GT-CA-AA-GG-AG-GC-GG-CC-AT-GA-AT-GG-3' (second PCR); MLL.CR2 with AF4.E3F and MLL.R6 5'-CA-AA-AC-TT-GT-GG-AA-GG-GC-TC-AC-AA-CA-GA-CT-TGG-3' with AF4.X3 for the reciprocal AF4-MLL fusion transcript.

NPM-ALK fusion transcripts were detected with nested PCR using the primers NPM.E4F1 5'-GC-AA-AG-GA-TG-AG-TT-GC-AC-AT-TG-TT-GA-AGC-3' with ALK.E20R1 5'-GG-TT-GT-AG-TC-GG-TC-AT-GA-TG-GT-CG-AG-GT-3' (first PCR) and NPM.E4F2 5'-AA-GC-AG-AG-GC-AA-TG-AA-TT-AC-GA-AG-GC-AG-TC-3' with ALK.E20R2 5'-AG-GT-GC-GG-AG-CT-TG-CT-CA-GC-TT-GT-ACT-3' (second PCR).

All primers used for experiments testing several housekeeping genes (*GAPDH*, *ACTB*, *HSPCB*, *CCND3* and *RPL3*) and *MLL2* can be made available upon request.

Results

Genes participating in chromosomal translocations generate early-terminated transcripts

All members of the ALF gene family (*AF4*, *LAF4*, *FMR2* and the later discovered *AF5q31/MCEF* gene) produce "early-terminated transcripts" that have been described in the literature [36, 37]. In case of the *AF4* gene, the *FelC* transcript encodes only the first 3 exons of *AF4* and is terminated within *AF4* intron 3 at a poly A site residing about 1.350 nucleotides downstream of the 3'-end of *AF4* exon 3. For the other investigated gene (*MLL, AF9*, *ELL* and *ENL*) no such early-terminated RNA transcripts were described so far. Therefore, we analyzed all these genes by 3'-RACE experiments using oligo-dT primers and an oligonucleotide that specifically binds to the first exon 5' to the breakpoint cluster region.

As shown in **Figure 1A,** all 5 investigated genes produce transcripts that were terminated within corresponding breakpoint cluster regions (*MLL-I8T*, *AF4-I3T*, *AF9-I5T, ELL-I2T and ENL-I3T*). Similar experiments performed for several introns of housekeeping genes (*GAPDH*, *ACTB*, *HSPCB*, *CCND3* and *RPL3*) or the *MLL2* gene revealed no such early-terminated transcripts. A precise quantification of these early-terminated transcripts is summarized in **Figure 1B**, indicating that their relative abundance compared to

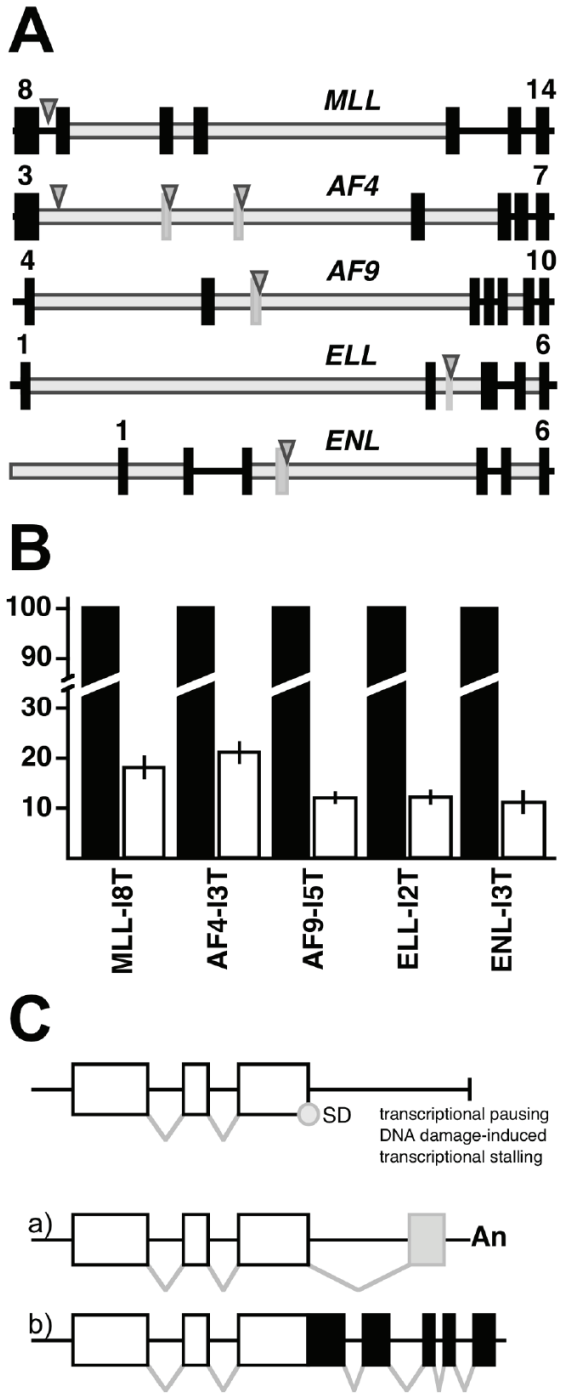

Figure 1. Mapping of early-terminated transcripts in breakpoint cluster regions of *MLL* and *MLL* translocations partner genes **A.** Breakpoint cluster regions of the investigated genes *MLL*, *AF4*, *AF9*, *ELL* and *ENL*. Black boxes: exons separated by introns (black lines). Light grey horizontal bars: known breakpoint cluster regions (bcr's); numbers indicate exons flanking the individual bcr's. Grey vertical boxes: cryptic exons identified in this study. Triangles: identified cryptic poly-A sites in early-terminated transcripts. **B.** Real-time quantification of early-terminated transcripts. RNA transcripts derived from 10 healthy volunteers were reverse transcribed and equal amounts were mixed for analysis using three replicates. Full-length transcripts were set to 100% (black bars) and the relative amount of early-terminated transcripts is indicated (white bars). **C.** Early-terminated transcripts may arise by different mechanisms (artificial poly-adenylation; transcriptional pausing; DNA-damage-induced transcriptional stalling). In all these cases, the final exon carries an unsaturated splice donor site able to participate in splicing reactions. a) the unsaturated splice donor site initiates a splice reaction to a cryptic exons located in the downstream intron; b) the unsaturated splice donor site initiates a trans-splicing reaction into another transcript.

their cognate full-length transcript is in the range of 10-20%. Thus, the presence of early-terminated transcripts seems to be a novel and specific property of genes involved in illegitimate recombination events. Since all 3'-RACE were performed with oligo-dT primers, the cloned RNA species ended all with a poly-A tail. Surprisingly, many transcripts contained yet unknown cryptic exons (e.g. 2 cryptic exons within *AF4* intron 3, 1 cryptic exon in *AF9* intron 5, 1 cryptic exon in *ELL* intron 2 and 1 cryptic exon in *ENL* intron 3). These cryptic exons are normally not used during splicing of regular transcripts deriving from these genes. These transcripts may represent only a minor fraction of early-terminated transcripts, because premature termination also occurs due to other processes, like e.g. transcriptional pausing or stalling of RNA polymerase. This may result in RNA molecules without poly-A tails, however, could not be identified by our applied strategy.

Early-terminated transcripts participate in trans-splicing events resulting in chimeric fusion transcripts

Early-terminated transcripts exhibit a non-saturated splice donor site at their final exon-intron borderline (see **Figure 1C**). This unsaturated splice donor site is *per se* able to participate in splicing reactions. This may result in the use of arbitrary, cryptic exons to saturate the splice donor site as already discussed above (see **Figure 1Ca**). Alternatively, a trans-splicing reaction may occur (see **Figure 1Cb**). Trans-splicing events may involve transcripts deriving

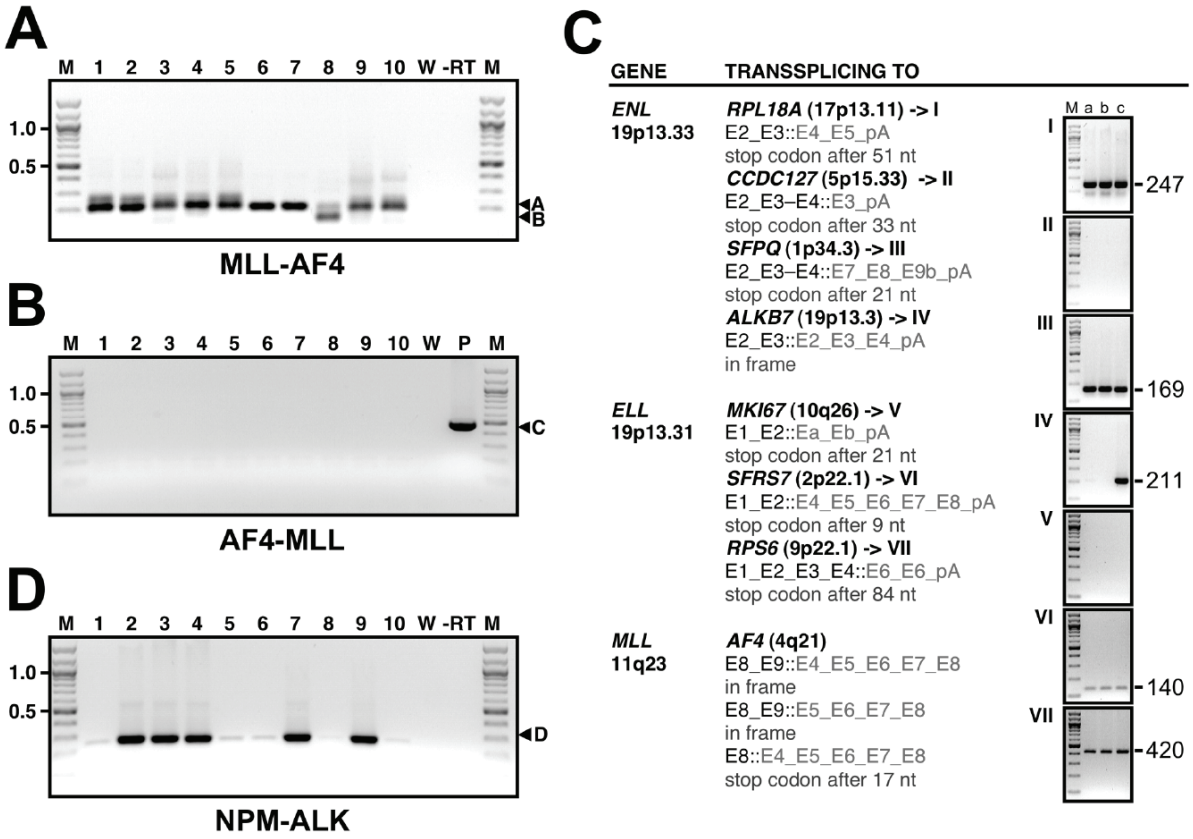

Figure 2. Intergenic trans-splicing events. **A.** Nested RT-PCR experiments revealed *MLL-AF4* fusion transcripts in 10 different healthy volunteers (1-10). Sequence analysis of these PCR amplimers revealed exon-exon fusions between *MLL* exon 9 and *AF4* exon 4 (A). In patient 8, *MLL* exon 8 was fused to *AF4* exon 4 (B). W: water control; -RT: minus RT control; a 100 bp ladder was used as size marker. **B.** Nested RT-PCR experiments revealed no *AF4-MLL* fusion transcripts in the same volunteers (1-10). W: water control; P: positive control (C). **C.** 3-RACE experiments were used to identify intergenic transplicing products. For 3 out of the 5 tested genes trans-splicing products were identified. Left: trans-splicing donor gene; middle: identified trans-splicing partner, listing the gene name (chromosome) and panel; the exon-exon-fusions between the trans-splicing partners are indicated by "::"; lower dashes between single exons indicate regularly spliced exon-exon boundaries in the identified transcripts. Exons designated with "a" or "b" represent cryptic exons normally not used for regular splicing events; right panels I-VII: RT-PCR experiments to validate the presence of the identified intergenic trans-splicing events in 3 additional healthy volunteers (a-c); M: size marker (100 bp ladder). **D.** Presence of the *NPM-ALK* fusion transcript in investigated healthy volunteers (1-10; D).

from the same gene (intragenic trans-splicing), or RNA molecules derived from other genes transcribed in vicinity or in the same transcription factory (intergenic trans-splicing). By using nested PCR experiments in combination with cloning and sequencing, we were able to demonstrate a predominant intergenic trans-splicing product of transcripts deriving from the *MLL* gene in all investigated healthy individuals: an *MLL-AF4* chimeric fusion RNA (in-frame fusion of *MLL* exon 9 with *AF4* exon 4 (A); see **Figure 2A**). A minor chimeric fusion RNA (fusing *MLL* exon 8 with *AF4* exon 4 (B)) was identified in volunteer #8 - and after cloning and sequencing - also in the remaining 9 volunteers. An additional *MLL* exon 9 / *AF4* exon 5 chimeric fusion RNA was infrequently detected. In none of the investigated samples, a corresponding reciprocal *AF4-MLL* chimeric fusion transcript could be detected (see **Figure 2B**), thereby excluding the presence of a genomic *MLL* rearrangement. It is worth mentioning that all investigated samples derived from healthy individuals that never had a leukemic disease. We also excluded that the investigated material carried chromosomal translocations of the *MLL* gene by using a recently established and highly sensitive method to characterize *MLL* rearrangements at the ge-

nomic DNA level [38].

Additional 3'-RACE experiments were performed to investigate the spectrum of trans-spliced RNA molecules and which are summarized in **Figure 2C**; the identified trans-spliced RNA species were mostly out of frame, and thus, not able to encode a chimeric fusion protein. However, the *ENL-ALKB7* fusion transcript was in frame, indicating that this specific fusion transcript could potentially be translated into a fusion protein. We also validated these findings by conventional RT-PCR experiments using 3 additional healthy volunteers (a-c). As shown in **Figure 2C** (right panels), 3 out of 4 intergenic trans-splicing events of the *ENL* gene and 2 out of 3 intergenic trans-splicing events of the *ELL* gene could be validated in independent PBMC samples, indicating that these arbitrary fusion transcripts can be readily identified in healthy volunteers. The exception was the in frame *ENL-ALKB7* fusion transcript that was identified only in 1 out of 3 healthy volunteers.

Several chimeric fusion transcripts have been described to be present in hematopoietic cells of healthy individuals (see above). We tested our samples for these fusion transcripts that potentially derive from trans-splicing events, but only the *NPM-ALK* fusion RNA could be identified (see **Figure 2D**). It is interesting to note that 5 out of 10 healthy volunteers displayed high abundance of this in frame fusion transcript, while the other 5 volunteers showed only a weak PCR band. This may indicate that there are interindividual differences in the production of these trans-spliced fusion RNAs. The chimeric *NPM-ALK* fusion transcript was first identified in patients suffering from anaplastic large cell lymphoma (ALCL) and derives from the chromosomal translocation t(2;5)(p23;q35) [39]. The reason why we were unable to detect the other chimeric fusion RNAs (e.g. *TEL-AML1, AML1-ETO, BCR-ABL* and *PML-RAR*) could be due to fact that all former studies have used hematopoietic stem or cord blood cells [15-21], while we investigated only peripheral blood mononuclear cells (PBMCs).

Exon-repetitions in mRNAs species are an indicator for the presence of early-terminated transcripts

Intragenic trans-splicing reactions between early-terminated and regular transcripts of the same gene result in another phenomenon termed "exon-repetition", also known as "partial tandem duplication" (PTD). By simple standard RT-PCR experiments (not nested) using inverse oligonucleotides that specifically bind to a single exon located 5' to the breakpoint cluster region (see **Figure 3A**), we were able to identify exon-repetitions for all 5 investigated genes (see **Figure 3B**). Sequence analyses revealed the repetition of single or multiple exons. This has several consequences, because *MLL* partial tandem duplications are used as diagnostic readout to identify AML patients with genomic tandem duplications of *MLL* exons 3-9. In order to identify such patients, fusion transcripts between *MLL* exon 9 and exon 3 were diagnosed by RT-PCR experiments. Exactly this *MLL* exon-exon fusion was identified also in healthy cells (see **Figure 3C**, RNA species a), indicating that even routine diagnostic methods may be compromised by these trans-splicing events. Analogous experiments performed on the above mentioned housekeeping genes and *MLL2* remained again negative. The high frequency of intragenic trans-splicing was surprising, because only 263 (corresponding to 178 human genes) out of ~6.25 million human ESTs display exon repetitions [40], indicating that PTDs are rarely identified in human transcripts. Noteworthy, thirteen out of these 178 genes (~7,5 %) have been reported to participate in chromosomal translocations [14]. In conclusion, genes that produce early-terminated transcripts can be easily identified by exon-repetitions, using the above outlined RT-PCR method.

Discussion

Recurrent chromosomal translocations are a genetic hallmark for many types of human cancer. They are mostly generated by NHEJ-mediated pathways of DNA double-strand breaks that occur in selected genes of our genome and which seem to be prone to DNA damage (genetic hot spots). Here, we demonstrate that genes known to be involved in chromosomal translocations display a genuine and novel feature: early-termination of transcriptional processes. Early termination was observed in those regions of the investigated genes that have already been defined as breakpoint cluster regions. The abundance of these early terminated transcripts was in the range of 10-20% when compared to their corresponding full-length transcripts. These shorter transcripts

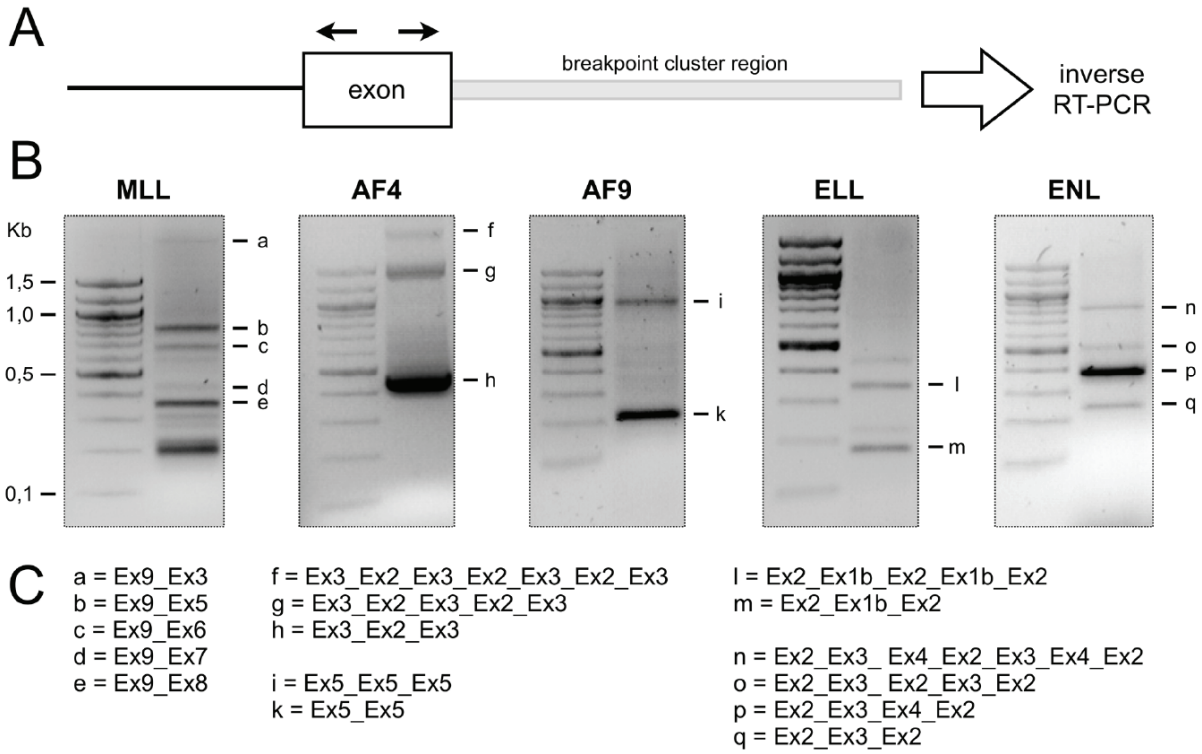

Figure 3. Intragenic trans-splicing reactions. **A.** Inverse RT-PCR experiments were performed for all investigated genes. Two oligonucleotides - pointing into opposite directions - were designed that specifically bind to a single exon located upstream of the bcr's. Subsequent RT-PCR experiments were performed using cDNAs of healthy volunteers. **B.** For all investigated genes, intragenic trans-splicing products were identified, cut-out from the gel, cloned and sequenced. Left lane in all panel: size marker; Right lane in all panels: intragenic trans-splicing products (a-q). **C.** Sequence analyses of PCR species a-q revealed the exon-exon-fusions of all intragenic trans-splicing products (lower dashes between single exons indicate correctly spliced exon-exon boundaries in the identified trans-spliced transcripts. Exons designated with additional letters (e.g. Ex1b) are cryptic exons normally not used during regular splicing.

seem to participate in intragenic and intergenic trans-splicing events, resulting in the creation of RNA species that exhibit exon-repetitions or resemble chimeric fusion RNAs. All our data were obtained by using human peripheral blood mononuclear cells (PBMCs) deriving from healthy individuals. These data confirm earlier studies made by others [15-24], however, we present for the first time a rational explanation for this unusual observation.

In first instance, the production of early-terminated transcripts seems to provide negative effects to cells, because it uses predominantly transcripts of the same gene to saturate the unique splice donor site, and thus, is lowering the transcript abundance of such genes. Moreover it also disrupts transcripts of other genes due to trans-splicing events. On the other hand, trans-splicing of RNA molecules remind us of the exon-shuffling mechanism that turned out to be a useful mechanism to create novel genes from existing ones during evolution [41].

However, the question remains what is causing early-termination of transcription. Cryptic poly-A sites, transcriptional pausing/stalling due to specific chromatin features or the presence of repetitive DNA elements may provide possible explanations. Another explanation is provided by microRNAs. A recently published study identified microRNAs encoded at a genomically unstable region [42]. Such microRNAs can specifically bind to their own gene transcripts and probably cause cleavage. The outlined mechanisms or any combination thereof possibly explain the creation of early-terminated transcripts, exhibiting a non-saturated splice donor site, and therefore, produce molecule species that are able to initiate *in cis* splicing processes to cryptic exons

(if available) or participates in trans-splicing reactions.

This study demonstrated that early-terminated transcripts seem to be the molecular source for the creation of trans-spliced RNA species. Most trans-spliced RNA species result in exon-repetitions (intragenic trans-splicing; identified in RT-PCR experiments) while only few cause the formation of chimeric fusion transcripts (intergenic trans-splicing; identified in nested RT -PCR experiments). In addition, most identified trans-spliced fusion RNAs displayed no continuous open reading frame (*ENL-RPL18A*, *ENL-CCDC127*, *ENL-SFPQ*, *ELL-MKI67*, *ELL-SFRS7*, *and ELL-RPS6*). By contrast, we identified *ENL-ALKB7*, *MLL-AF4* and *NPM-ALK* with fused open reading frames in PBMCs of healthy individuals. Noteworthy, the latter two are identical to chimeric fusion transcripts that are regularly identified in tumor cells carrying the specific chromosomal translocations t(4;11)(q21;q23) and t (2;5)(p23;q35), These translocations are all associated with acute lymphoblastic leukemia and anaplastic large cell lymphoma (ALCL), respectively. Therefore, we suggest defining these chimeric fusion RNAs produced in non-rearranged cells as “pro-neoplastic RNA molecules”.

The existence of pro-neoplastic RNA molecules in healthy individuals is not restricted to hematopoietic cells. Recent findings suggest that chimeric fusion RNAs can be produced in healthy (by trans-splicing) and in cancer cells carrying indeed this particular genetic rearrangements. Examples are the *JAZF1-JJAZ1* chimeric fusion RNA in normal endometrial cells [43], or the androgen-induced *SLC45A3-ELK4* chimeric fusion RNA in normal prostate cells [44]. The JAZF1-JJAZ1 fusion protein provides anti-apoptotic activity, indicating that such a trans-spliced fusion RNA even provide a benefit to those cells. Moreover, the chimeric fusion RNAs seem to be produced in a cell-type specific manner [43, 44]. This could be explained by a cell-type specific localization of chromosome territories in the nucleus; because chromatin loops deriving from different chromosomes must be attracted to a common transcription factory [33, 45-48] in order to create such trans-spliced fusion RNAs.

These findings are leading back to our initial question whether there exists a molecular mechanism that could explain how illegitimate recombination events are created and why the phenomenon of recurrence exists in human tumor samples. Are pro-neoplastic RNA molecules somehow involved in the creation of illegitimate recombination? If so, the trans-spliced fusion RNA molecules must be able to influence DNA repair processes. This idea is supported by three recent findings:

a. Storici and coworkers demonstrated for the first time that DNA repair in fission yeast could be templated by small RNA molecules [49]. The existence of “RNA-templated DNA repair” in fission yeast implies that RNA has *per se* the capability to influence DNA repair processes, most likely by basepairing to DNA sequences flanking a damaged DNA region.

b. Another study revealed that a component of the NHEJ machinery, DNA ligase IV, has the ability to catalyze a ligation reaction between single-stranded DNA ends [50]. Thus, the opposite DNA strand - not basepairing to a complementary, chimeric fusion RNA - could be ligated by using the single-stranded DNA ends of a broken DNA molecule.

c. Thirdly, the ENCODE project [51] has already demonstrated that hnRNA molecules seem to be generated from most parts of the human genome, including long intergenic regions. Unspliced hnRNA molecules represent perfect copies of genomic regions and are easily accessible for DNA repair processes, without the need to find appropriate DNA sequences on a homologous chromosome.

Based on these data and our findings, we propose a hypothesis that might explain the onset of chromosomal translocations: (a) transcription of genes in local proximity, e.g. in a common transcription factory, (b) the presence of early-terminated transcripts in certain genes of our genome, (c) the creation of trans-spliced fusion transcripts, (d) the occurrence of DNA double-strand breaks in both genes involved in genetic recombination events, and finally (e) the formation of RNA/DNA hybrid structures that guide subsequent DNA repair process. The latter process is depicted in **Figure 4**. It is important to note that the trans-spliced chimeric RNA molecules, in this case the *MLL-AF4* fusion RNA, is not a substrate for DNA repair polymerases rather than forcing an "RNA/DNA repair structure". Due to the enzymatic activity of DNA ligase IV, the genomic fusion may occur at any

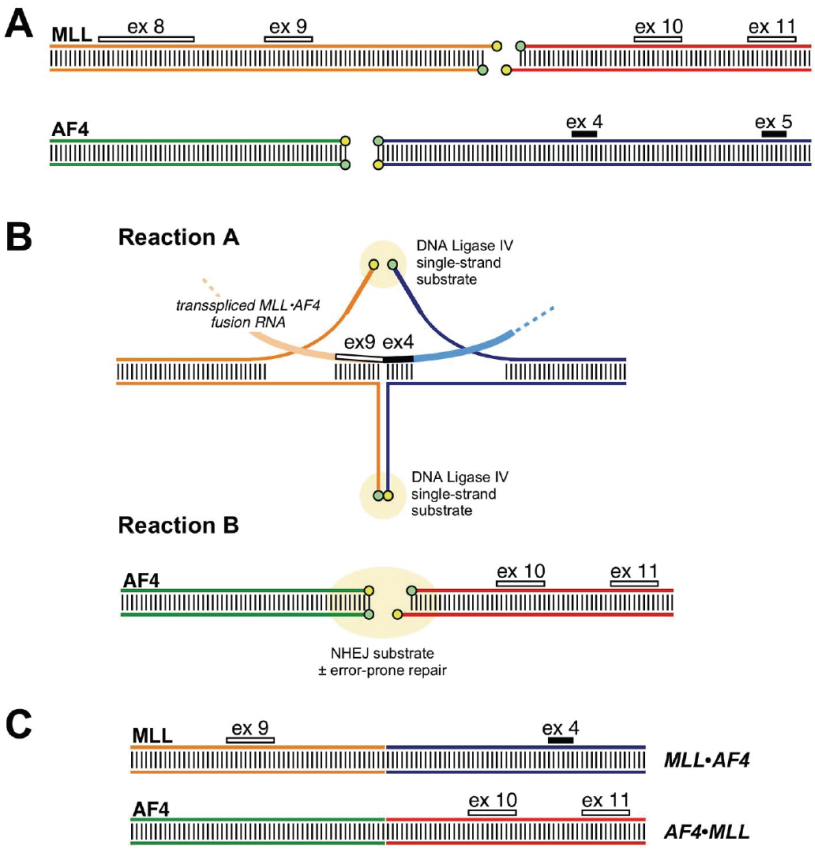

Figure 4. Proposed hypothesis for the RNA-guided DNA repair mechanism. **A.** Double-stranded DNA breaks downstream of *MLL* exon 9 and upstream of *AF4* exon 4 (or exon 5) are a genetic prerequiste for misguided DNA repair and the onset of CTLs. **B.** The transspliced RNA, in that case an *MLL-AF4* fusion RNA, binds to the non-coding strand and aligns the two chromosomes via *MLL* exon 9 and *AF4* exon 4 sequences. The protruding single strand ends of both chromosomes are substrates for the ss-Ligase activity of DNA Ligase IV (part of the NHEJ system). The first single-strand ligation step is resulting in the *MLL-AF4* fusion allele (Reaction A). In a second step, reaction B is carried out by components of the NHEJ DNA repair system. Reaction B is a non-directed but necessary step to allow cell survival. **C.** The final product is a typical t(4;11) translocation as diagnosed in most t (4;11) patients.

sequence downstream of *MLL* exon 9 (e.g. *MLL* intron 9) and upstream of *AF4* exon 4 (e.g. *AF4* intron 3). This is exactly where most t(4;11) leukemia patients display their chromosomal fusion site. Thus, the most abundant trans-spliced *MLL (exon9): AF4(exon4)* fusion RNA is explaining also the breakpoint distribution that has been described for acute leukemia patients bearing t(4;11) translocations [52].

The existence of an "RNA-mediated proof-editing process" would not only be interesting for the onset of specific genetic lesions, but could be a novel and fundamental mechanism for maintaining genome integrity by using the existent hnRNA molecules. We are currently conducting experiments that aim to validate this novel hypothesis and to investigate these emerging data in more detail.

Acknowledgements

This study was supported by the grant III L 4-518/55.004 from LOEWE priority program (Oncogenic Signaling Frankfurt, OSF), funded by the Hessisches Ministerium für Wissenschaft und Kunst (HMWK). RM is supported by the Center of Excellence Frankfurt on Macromolecular Complexes (CEF-MC).

Address correspondence to: Dr. Rolf Marschalek, Institute of Pharmaceutical Biology, JWG-University Frankfurt, Biocenter, N230, 3.03, Max-von-Laue-Str. 9, D-60438 Frankfurt/Main Tel: +49-69-798-29647 Fax: +49-69-798-29662, Email: rolf.marschalek@em.uni-rankfurt.de

References

[1] Reichel M, Gillert E, Nilson I, Siegler G, Greil J, Fey GH, Marschalek R. Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. Oncogene 2009;17:3035-3044.

[2] Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. Nature 2000;405:697-700.

[3] Broeker PL, Super HG, Thirman MJ, Pomykala H, Yonebayashi Y, Tanabe S, Zeleznik-Le N, Rowley JD. Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in de

novo acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consensus binding sites. Blood 1996;87:1912-1922.
[4] Strissel PL, Strick R, Tomek RJ, Roe BA, Rowley JD, Zeleznik-Le NJ. DNA structural properties of AF9 are similar to MLL and could act as recombination hot spots resulting in MLL/AF9 translocations and leukemogenesis. Hum Mol Genet 2000;9:1671-1679.
[5] Hensel JP, Gillert E, Fey GH, Marschalek R. Breakpoints of t(4;11) translocations in the human MLL and AF4 genes in ALL patients are preferentially clustered outside of high-affinity matrix attachment regions. J Cell Biochem 2001;82:299-309.
[6] Strick R, Zhang Y, Emmanuel N, Strissel PL. Common chromatin structures at breakpoint cluster regions may lead to chromosomal translocations found in chronic and acute leukemias. Hum Genet 2006;119:479-495.
[7] Strissel PL, Strick R, Rowley JD, Zeleznik-Le NJ. An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. Blood 1998;92:3793-803.
[8] Aplan PD, Chervinsky DS, Stanulla M, Burhans WC. Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors. Blood 1996;87:2649-2658.
[9] Megonigal MD, Cheung NK, Rappaport EF, Nowell PC, Wilson RB, Jones DH, Addya K, Leonard DG, Kushner BH, Williams TM, Lange BJ, Felix CA. Detection of leukemia-associated MLL-GAS7 translocation early during chemotherapy with DNA topoisomerase II inhibitors. Proc Natl Acad Sci USA 2000;97:2814-2819.
[10] Scharf S, Zech J, Bursen A, Schraets D, Oliver PL, Kliem S, Pfitzner E, Gillert E, Dingermann T, Marschalek R. Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. Oncogene 2007;26:1361-1371.
[11] Betti CJ, Villalobos MJ, Diaz MO, Vaughan AT. Apoptotic triggers initiate translocations within the MLL gene involving the nonhomologous end joining repair system. Cancer Res 2001;61:4550-4555.
[12] Stanulla M, Chhalliyil P, Wang J, Jani-Sait SN, Aplan PD. Mechanisms of MLL gene rearrangement: site-specific DNA cleavage within the breakpoint cluster region is independent of chromosomal context. Hum Mol Genet 2001;10:2481-2491.
[13] Sim SP, Liu LF. Nucleolytic cleavage of the mixed lineage leukemia breakpoint cluster region during apoptosis. J Biol Chem 2001;276:31590-31595.
[14] Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007; 7:233-245.
[15] Uckun FM, Herman-Hatten K, Crotty ML, Sensel MG, Sather HN, Tuel-Ahlgren L, Sarquis MB, Bostrom B, Nachman JB, Steinherz PG, Gaynon PS, Heerema N. Clinical significance of MLL-AF4 fusion transcript expression in the absence of a cytogenetically detectable t(4;11)(q21;q23) chromosomal translocation. Blood 1998;92:810-821.
[16] Marcucci G, Strout MP, Bloomfield CD, Caligiuri MA. Detection of unique ALL1 (MLL) fusion transcripts in normal human bone marrow and blood: distinct origin of normal versus leukemic ALL1 fusion transcripts. Cancer Res 1998;58:790-793.
[17] Caldas C, So CW, MacGregor A, Ford AM, McDonald B, Chan LC, Wiedemann LM. Exon scrambling of MLL transcripts occurs commonly and mimics partial genomic duplication of the gene. Gene 1998;208:167-176.
[18] Biernaux C, Loos M, Sels A, Huez G, Stryckmans P. Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals. Blood 1995; 86:3118-3122.
[19] Bose S, Deininger M, Gora-Tybor J, Goldman JM, Melo JV. The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease. Blood 1998; 92:3362-3367.
[20] Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, Hows JM, Navarrete C, Greaves M. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci USA 2002;99:8242-8247.
[21] Eguchi-Ishimae M, Eguchi M, Ishii E, Miyazaki S, Ueda K, Kamada N, Mizutani S. Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals. Blood 2001; 97:737-743.
[22] Quina AS, Gameiro P, Sá da Costa M, Telhada M, Parreira L. PML-RARA fusion transcripts in irradiated and normal hematopoietic cells. Genes Chrom Cancer 2000;29:266-275.
[23] Maes B, Vanhentenrijk V, Wlodarska I, Cools J, Peeters B, Marynen P, de Wolf-Peeters C. The NPM-ALK and the ATIC-ALK fusion genes can be detected in non-neoplastic cells. Am J Pathol 2001; 158:2185-2193.
[24] Beylot-Barry M, Groppi A, Vergier B, Pulford K, Merlio JP. Characterization of t(2;5) reciprocal transcripts and genomic breakpoints in CD30+ cutaneous lymphoproliferations. Blood 1998; 91:4668-4676.
[25] Cremer T, Kurz A, Zirbel R, Dietzel S, Rinke B, Schrock E, Speicher MR, Mathieu U, Jauch A, Emmerich P, Scherthan H, Ried T, Cremer C,

Lichter P. Role of chromosome territories in the functional compartmentalization of the cell nucleus. Cold Spring Harb Symp Quant Biol 1993;58:777-792.
[26] Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, Turner BM, Zink D. Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J Cell Biol 1999;146:1211-1226.
[27] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001;2:292-301.
[28] Pederson T. Gene territories and cancer. Nat Genet 2003;34:242-243.
[29] Kozubek S, Lukásová E, Marecková A, Skalníková M, Kozubek M, Bártová E, Kroha V, Krahulcová E, Slotová J. The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t (9, 22) translocations and in the pathogenesis of t (9,22) leukemias. Chromosoma 1999; 108:426-435.
[30] Neves H, Ramos C, da Silva MG, Parreira A, Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood 1999; 93:1197-1207.
[31] Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet 2003;34:287-291.
[32] Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P. Myc Dynamically and Preferentially Relocates to a Transcription Factory Occupied by Igh. PLoS Biol 2007;5:e192.
[33] Gingeras TR. Implications of chimeric non-co-linear transcripts. Nature 2009; 461:206-211.
[34] Gué M, Sun JS, Boudier T. Simultaneous localization of MLL, AF4 and ENL genes in interphase nuclei by 3D-FISH: MLL translocation revisited. BMC Cancer 2006;6:20.
[35] Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J, Ben Abdelali R, Macintyre E, De Braekeleer E, De Braekeleer M, Delabesse E, de Oliveira MP, Cavé H, Clappier E, van Dongen JJ, Balgobind BV, van den Heuvel-Eibrink MM, Beverloo HB, Panzer-Grümayer R, Teigler-Schlegel A, Harbott J, Kjeldsen E, Schnittger S, Koehl U, Gruhn B, Heidenreich O, Chan LC, Yip SF, Krzywinski M, Eckert C, Möricke A, Schrappe M, Alonso CN, Schäfer BW, Krauter J, Lee DA, Zur Stadt U, Te Kronnie G, Sutton R, Izraeli S, Trakhtenbrot L, Lo Nigro L, Tsaur G, Fechina L, Szczepanski T, Strehl S, Ilencikova D, Molkentin M, Burmeister T, Dingermann T, Klingebiel T, Marschalek R. The MLL recombinome of acute leukemias. Leukemia 2009;23:1490-1499.
[36] Nilson I, Reichel M, Ennas MG, Greim R, Knorr C, Siegler G, Greil J, Fey GH, Marschalek R. Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. Br J Hematol 1997;98:157-169.
[37] Estable MC, Naghavi MH, Kato H, Xiao H, Qin J, Vahlne A, Roeder RG. MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein. J Biomed Sci 2002;9:234-245.
[38] Meyer C, Schneider B, Reichel M, Angermueller S, Strehl S, Schnittger S, Schoch C, Jansen MW, van Dongen JJ, Pieters R, Haas OA, Dingermann T, Klingebiel T, Marschalek R. Diagnostic tool for the identification of MLL rearrangements including unknown partner genes. Proc Natl Acad Sci USA 2005;102:449-454.
[39] Ladanyi M, Cavalchire G, Morris SW, Downing J, Filippa DA. Reverse transcriptase polymerase chain reaction for the Ki-1 anaplastic large cell lymphoma-associated t(2;5) translocation in Hodgkin's disease. Am J Pathol 1995; 145:1296-300.
[40] Dixon RJ, Eperon IC, Hall L, Samani NJ. A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species. Nucl Acids Res 2005; 33:5904-5913.
[41] Gilbert W, de Souza SJ, Long M. Origin of genes. Proc Natl Acad Sci USA 1997; 94:7698-7703.
[42] Huppi K, Volfovsky N, Runfola T, Jones TL, Mackiewicz M, Martin SE, Mushinski JF, Stephens R, Caplen NJ. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. Mol Cancer Res 2008;6:212-221.
[43] Li H, Wang J, Mor G, Sklar J. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. Science 2008;321:1357-1361.
[44] Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, Rubin MA. SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer. Cancer Res 2009;69:2734-2738.
[45] Jackson DA, Hassan AB, Errington RJ, Cook PR. Visualization of focal sites of transcription within human nuclei. EMBO J 1993;12:1059-1065.
[46] Iborra FJ, Pombo A, Jackson DA, Cook PR . Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. J Cell Sci 1996;109:1427-1436.
[47] Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, Fraser P. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet

2004;36:1065-1071.
[48] Branco MR. Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 2006;4:e138.
[49] Storici F, Bebenek K, Kunkel TA, Gordenin DA, Resnick MA. RNA-templated DNA repair. Nature 2007;447:338-341.
[50] Gu J, Lu H, Tsai AG, Schwarz K, Lieber MR. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. Nucl Acids Res 2007; 35:5755-5762.
[51] ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799-816.
[52] Reichel M, Gillert E, Angermüller S, Hensel JP, Heidel F, Lode M, Leis T, Biondi A, Haas OA, Strehl S, Panzer-Grümayer ER, Griesinger F, Beck JD, Greil J, Fey GH, Uckun FM, Marschalek R. Biased distribution of chromosomal breakpoints involving the MLL gene in infants versus children and adults with t(4;11) ALL. Oncogene 2001;20:2900-2907.