

Enhancement of microhomology-mediated genomic rearrangements by transient loss of mouse Bloom syndrome helicase

Ayako Yamanishi,¹ Kosuke Yusa,² Kyoji Horie,¹ Masahiro Tokunaga,¹ Kohji Kusano,³ Chikara Kokubu,^{1,4} and Junji Takeda^{1,4}

¹Department of Social and Environmental Medicine, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan; ²Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; ³Center for Genetic Resource Education and Development, Kyoto Institute of Technology, Kyoto 606-8585, Japan

Bloom syndrome, an autosomal recessive disorder of the *BLM* gene, confers predisposition to a broad spectrum of early-onset cancers in multiple tissue types. Loss of genomic integrity is a primary hallmark of such human malignancies, but many studies using disease-affected specimens are limited in that they are retrospective and devoid of an appropriate experimental control. To overcome this, we devised an experimental system to recapitulate the early molecular events in genetically engineered mouse embryonic stem cells, in which cells undergoing loss of heterozygosity (LOH) can be enriched after inducible down-regulation of *Blm* expression, with or without site-directed DNA double-strand break (DSB) induction. Transient loss of BLM increased the rate of LOH, whose breakpoints were distributed along the chromosome. Combined with site-directed DSB induction, loss of BLM synergistically increased the rate of LOH and concentrated the breakpoints around the targeted chromosomal region. We characterized the LOH events using specifically tailored genomic tools, such as high-resolution array comparative genomic hybridization and high-density single nucleotide polymorphism genotyping, revealing that the combination of BLM suppression and DSB induction enhanced genomic rearrangements, including deletions and insertions, whose breakpoints were clustered in genomic inverted repeats and associated with junctional microhomologies. Our experimental approach successfully uncovered the detailed molecular mechanisms of as-yet-uncharacterized loss of heterozygosities and reveals the significant contribution of microhomology-mediated genomic rearrangements, which could be widely applicable to the early steps of cancer formation in general.

[Supplemental material is available for this article.]

Genomic instability, which has been implicated as a major underlying mechanism of human carcinogenesis, is considered to be driven largely by dysfunction of the so-called genome “caretaker” genes (Kinzler and Vogelstein 1997). Analogous to Darwinian evolution, carcinogenesis involves the step-by-step accumulation of genomic alterations at multiple loci and subsequent natural selection of the affected cells in the somatic environment. The initial steps of the genomic alterations are, therefore, difficult to assess retrospectively in already-diagnosed cancer samples, and how caretaker gene dysfunction contributes to each step remains unknown.

Germline deleterious mutations in the Bloom syndrome gene (*BLM*), one of the genome caretaker genes belonging to the RecQ family of DNA helicases, cause the rare autosomal recessive disorder Bloom syndrome. This syndrome predisposes affected individuals to a wide variety of malignancies (German 1997) that are associated with increased somatic loss of heterozygosity (LOH), including mitotic crossover (CO), noncrossover (NCO), and regional chromosome deletion (DEL) (Luo et al. 2000). The wide variety of Bloom syndrome-associated cancers implies that the initial steps of carcinogenesis in the *BLM*-deficient background are common to many cell types. The cancer predisposition phenotype

with increased rates of somatic LOH has been recapitulated in the *Blm*-mutant mouse model (Luo et al. 2000). We have previously generated a *Blm*^{tet/tet} mouse embryonic stem (ES) cell line in which the expression of *Blm* can be transiently switched off using the doxycycline-inducible (Dox-inducible) system (Fig. 1A; Yusa et al. 2004; Horie et al. 2011), allowing detailed investigation of the immediate impact of *Blm* deficiency on genomic integrity. The genome of mouse ES cells is relatively stable, and the spontaneous mutation rate is significantly lower than that in other somatic cells, such as mouse embryonic fibroblasts (Cervantes et al. 2002). Accordingly, mouse ES cells offer a powerful experimental platform in which de novo mutations and rearrangements are detectable efficiently. In addition, the F1-hybrid nature of the *Blm*^{tet/tet} ES cell line, which is derived from crosses between two divergent inbred mouse strains, C57BL/6 (B6) and 129S4/SvJae (129) (Eggan et al. 2001), together with the recent development of genome-wide high-density single nucleotide polymorphism (SNP) markers provided by next-generation sequencing (Keane et al. 2011; Yalcin et al. 2011), facilitates high-resolution mapping of LOH breakpoints and definition of the allelic origin of each chromosomal segment.

A previous study demonstrated that site-directed DNA double-strand break (DSB) induction enhanced the rate of CO-mediated LOH in *Blm*-deficient (*Blm*^{tet/tet}) ES cells (Larocque et al. 2011), suggesting that the DSB is a component of the mechanism of somatic LOH in Bloom syndrome patients. To provide a comprehensive view of the early processes of LOH in the absence of BLM, here, we present a molecular characterization of the genomic

⁴Corresponding authors

E-mail ckokubu@mr-envi.med.osaka-u.ac.jp

E-mail takeda@mr-envi.med.osaka-u.ac.jp

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.152744.112>.

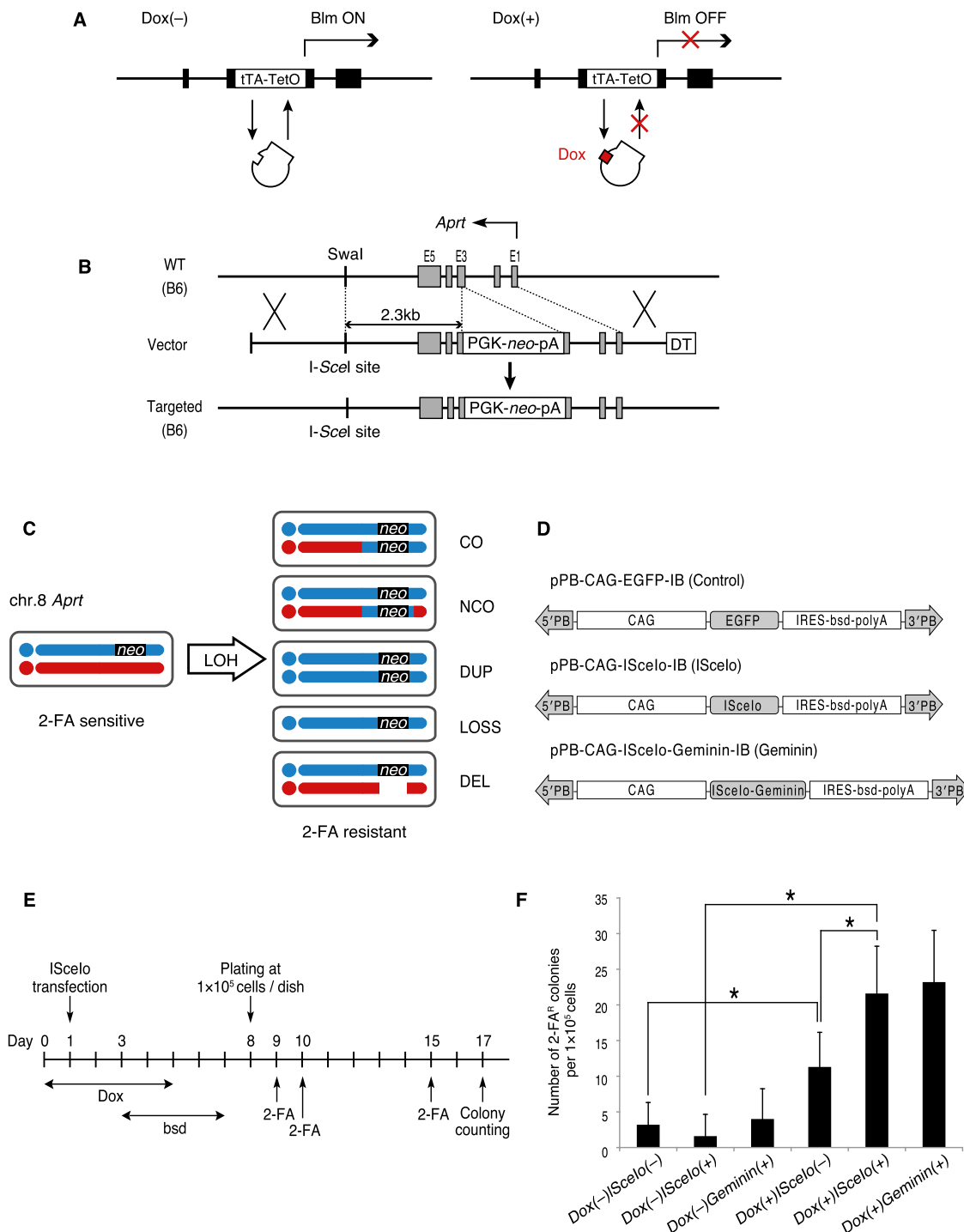


Figure 1. Experimental design and synergistic enhancement in the rate of LOH by the combination of transient loss of *Blm* and DSB induction. (A) Schematic representation of the Dox-inducible knockdown construct for the *Blm*^{tet/tet} ES cell line (Horie et al. 2011). A Tet-off cassette, consisting of tetracycline-regulated transactivator (tTA) and a basal promoter containing Tet operator sequences (TetO), was inserted into both alleles of the *Blm* gene in mouse F1 (B6 × 129) ES cells. (Left) In the absence of Dox, the tTA protein binds to the tetO sequence and activates *Blm* gene expression. (Right) The application of Dox prevents the binding of tTA to tetO and represses *Blm* expression in a reversible manner. (B) Targeting strategy. A wild-type neomycin resistance (*neo*) gene cassette was inserted into exon 3 of the *Aprt* gene on the B6-derived chromosome 8 of *Blm*^{tet/tet} ES cells, accompanied by the introduction of an *I-SceI* recognition sequence into the *Swal* site 2.3 kb centromeric to exon 3, resulting in *Blm*^{tet/tet} *Aprt*^{neo/+} ES cells. (DT) Diphtheria toxin A fragment, (WT) wild type. (C) A negative selection scheme for LOH using 2-fluoroadenine (2-FA). In theory, five LOH patterns—crossover (CO), non-crossover (NCO), chromosomal duplication (DUP), chromosomal loss (LOSS), and partial chromosomal deletion (DEL)—could be recovered. (D) Expression vectors used in this study. 5'PB and 3'PB are terminal repeats of the piggyBac transposon. (*EGFP*) Enhanced green fluorescent protein, (*I-SceI*) codon-optimized *I-SceI*, (*I-SceI*-Geminin) a fusion gene of *I-SceI* and *geminin*, (CAG) CAG promoter, (IRES) internal ribosomal entry site, (*bsd*) blasticidin resistance gene, (polyA) polyadenylation signal. (E) Schematic representation of the cell culture protocol. See text for details. (F) Total number of 2-FA-resistant colonies generated from 1 × 10⁵ *Blm*^{tet/tet} *Aprt*^{neo/+} ES cells subjected to different treatments as indicated: Dox(+) or Dox(-), 5-d culture in the presence or absence of Dox to switch *Blm* expression “off” or “on”, respectively; *I-SceI*(+) or *I-SceI*(-), transfection of the *I-SceI* vector for DSB induction or the control vector, respectively; Geminin(+), transfection of the Geminin vector. The error bars represent the standard deviation of four to seven independent experiments. (*) *P* < 0.05, Student's *t*-test.

structural changes after transient loss of BLM and/or DSB induction, based on a combination of high-density polymorphism typing and high-resolution array comparative genomic hybridization (array-CGH) analysis.

Results

Transient loss of BLM increases the rate of LOH with synergistic enhancement by site-directed DSB induction

By engineering *Blm*^{tet/tet} ES cells (Fig. 1A; Horie et al. 2011), we devised an experimental system to recapitulate the early molecular events affecting genomic integrity in proliferating cells in vitro. We focused on the genomic region encompassing the adenine phosphoribosyl transferase (*Aprt*) gene on mouse distal chromosome 8, because the *Aprt* gene itself could serve as a negative selection marker for LOH, as described below. The B6-derived allele of the *Aprt* gene in *Blm*^{tet/tet} ES cells was disrupted by targeted insertion of a wild-type *neo*-containing vector into the third exon (Fig. 1B; Supplemental Fig. S1A). The purpose of this targeting vector was twofold: negative selection for LOH and site-directed DSB induction. First, disruption of the endogenous *Aprt* gene provides a negative selection system for LOH events. The *Blm*^{tet/tet}*Aprt*^{neo/+} ES cells, in which the B6-derived copy of *Aprt* is disrupted, are still sensitive to 2-fluoroadenine (2-FA), because products from the other copy of *Aprt* convert 2-FA into cytotoxic metabolites (Hong et al. 2006). However, LOH events involving the *Aprt* gene locus, which could include any of CO, NCO, whole chromosome duplication (DUP), whole chromosome loss (LOSS), or DEL (Fig. 1C), could result in disruption of both alleles of *Aprt* and consequent resistance to 2-FA. Second, the targeting vector allows site-directed DSB induction by introducing an I-SceI recognition sequence centromeric to the *Aprt* gene on chromosome 8 (Fig. 1B). To deliver the yeast-derived I-SceI endonuclease to the I-SceI recognition sequence in vivo in *Blm*^{tet/tet}*Aprt*^{neo/+} mouse ES cells, a piggyBac (PB) transposon-based expression vector of a mammalian codon-optimized version of the *I-SceI* gene (*ISceI*) was constructed (referred to as pPB-CAG-*ISceI*-IB) (Fig. 1D; Supplemental Fig. S1B). The codon optimization resulted in increased efficiency of cleavage at the I-SceI restriction site in mammalian cells (Supplemental Fig. S1C). The efficiency of cleavage was further evaluated by loss of the I-SceI site caused by imprecise nonhomologous end-joining in *Blm*^{tet/tet}*Aprt*^{neo/+} ES cells, indicating ~40% cleavage at the I-SceI site in vivo (Supplemental Fig. S1D).

Using this *Blm*^{tet/tet}*Aprt*^{neo/+} ES cell system, we examined the rate of LOH arising at the *Aprt* gene locus in *Blm*-deficient and *Blm*-proficient ES cells, with or without site-specific DSB induction. The *Blm*^{tet/tet}*Aprt*^{neo/+} ES cells were cultured for 5 d in the absence or presence of Dox to continue, or transiently switch off, *Blm* expression, respectively (Fig. 1E). On day 1 of the culture, the cells were transfected with the *ISceI* expression vector or a control vector and later, from day 3 to day 7, subjected to Blasticidin (bsd) selection to enrich for transfected cells. On day 8 of the culture, an equal number (1×10^5) of bsd-resistant cells were plated into a 10-cm dish and subjected to 2-FA selection to compare the number of resistant colonies. The number of Dox(+)*ISceI*(+) and Dox(+)*ISceI*(-) colonies was significantly higher than that of Dox(-)*ISceI*(+) and Dox(-)*ISceI*(-) colonies, respectively, demonstrating that transient loss of BLM elevated the rate of LOH both in DSB-induced and uninduced conditions (Fig. 1F). The *ISceI*-mediated DSB induction did not significantly increase the rate of LOH in *Blm*-proficient cells [Fig. 1F, Dox(-)*ISceI*(+) vs. Dox(-)*ISceI*(-)].

However, under transient loss of BLM, DSB induction resulted in synergistic enhancement of the rate of LOH [Fig. 1F, Dox(+)*ISceI*(+) vs. Dox(+)*ISceI*(-)], which is consistent with a previous study (Larocque et al. 2011).

The occurrence of somatic LOH could be dependent on the cell-cycle phase. Thus, we constructed an S/G2/M phase-specific *ISceI* expression vector by fusing the *ISceI* gene with the gene encoding 110 amino acid N terminus of human Geminin (Sakaue-Sawano et al. 2008), resulting in pPB-CAG-*ISceI*-Geminin-IB (Geminin) (Fig. 1D). Although the Geminin vector did not exert the expected cell-cycle specificity (for undefined reasons) (Supplemental Fig. S2A), transfection of Dox-treated *Blm*^{tet/tet}*Aprt*^{neo/+} ES cells with this vector produced a similar number of 2-FA-resistant colonies, as did the *ISceI* vector (Fig. 1F).

LOH breakpoints are uniformly distributed throughout the chromosome

We next performed SNP analyses of LOH arising at the *Aprt* gene locus in *Blm*-deficient or *Blm*-proficient ES cells. LOH profiling of the isolated 2-FA-resistant clones with a panel of 12 SNP markers on chromosome 8 revealed that, without DSB induction, transient loss of BLM resulted in a fairly even distribution of LOH breakpoints (Fig. 2A; Supplemental Table S1). Such a chromosome-wide distribution was neither chromosome-specific nor selection method-specific, because we observed a similar distribution pattern in an additional two *Blm*^{tet/tet}-derived mutant ES cell lines, *Blm*^{tet/tet}*Fas*^{neo/+} and *Blm*^{tet/tet}*Fas*^{neo/+}*Nanog*^{puroΔTK/+}, in which a positive selection marker for LOH was inserted at the *Fas* locus on distal chromosome 1 (Supplemental Fig. S3A,B; Supplemental Table S1) and a negative selection marker for LOH was inserted at the *Nanog* locus on distal chromosome 6 (Supplemental Fig. S3C-E; Supplemental Table S1), respectively. This pattern was also observed in *Blm*-proficient conditions (data not shown), suggesting that the chromosomal distribution of LOH breakpoints is *Blm*-independent.

Site-directed DSB induction concentrates LOH breakpoints on a focused region of the genome in *Blm*-deficient ES cells

The combination of transient loss of BLM with site-directed DSB induction dramatically altered the distribution of LOH breakpoints, such that they were more localized to the vicinity of the *Aprt* gene locus (Fig. 2B, left). Out of 64 resistant clones shown in Figure 2B, 55 clones underwent LOH with retention of the B6 allele (denoted hereafter as LOH^{B6}) at the SNP12 locus, telomeric to the *Aprt* locus (Fig. 2B,C, Group I), while eight clones displayed no LOH patterns and one clone (clone 68) displayed LOH with retention of the 129 allele (denoted hereafter as LOH¹²⁹) at the SNP12 locus (Fig. 2B,C, Group II). Twenty-seven (49%) of the 55 clones in group I retained B6/129 heterozygosity at the SNP11 locus, representing the predominant occurrence of LOH near the *Aprt* locus compared to the LOH pattern in *ISceI*-nontreated cells ($P < 0.01$, χ^2 test) (Fig. 2A-C). To determine the boundaries of LOH with higher resolution, we analyzed the genomic DNA of these 27 clones with an additional five SNP markers between the SNP11 and SNP12 loci (Supplemental Table S1). Among the 27 clones, 13 clones underwent LOH^{B6} at the SNP u-50k locus, which is 50 kb centromeric to the *Aprt* locus (Fig. 2B,C; Supplemental Table S1). Meanwhile, the other 14 clones retained B6/129 heterozygosity at the SNP u-50k locus, indicating that their LOH breakpoints were between u-50k and *Aprt* (Fig. 2D, Group I). Nine clones in group II were likewise analyzed by the same SNP panel and demonstrated

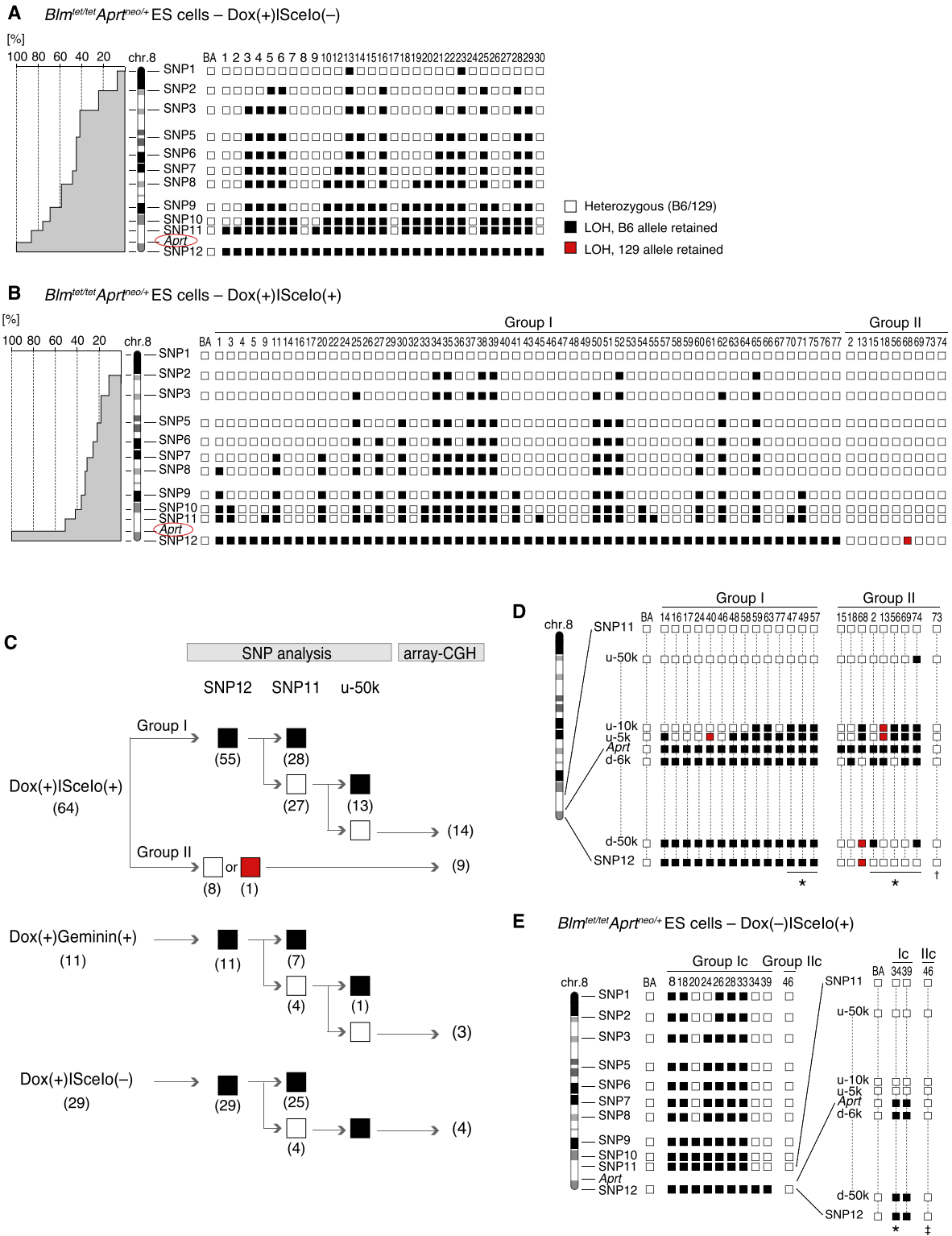


Figure 2. Polymorphic marker analyses of chromosome 8 in *Blm^{tet/tet}Aprt^{neo/+}* ES cell clones selected by 2-FA for LOH at the *Aprt* locus. (A) (Right) Twenty-nine independent Dox(+)|Scelo(-) ES cell clones resistant to 2-FA (numbered as indicated at the top) were genotyped for 12 SNPs (SNP1 to SNP12; SNP4 is not shown because of the low specificity of the PCR reaction) covering chromosome 8 (Supplemental Table S1). Open squares indicate the heterozygous genotype (B6/129). Closed black and red squares indicate LOH with retention of the B6 and 129 alleles, respectively. (BA) Control 2-FA-unselected Dox(-)|Scelo(-) *Blm^{tet/tet}Aprt^{neo/+}* ES cells. (Left) Histogram representation of LOH ratio (total number of clones with LOH/total number of clones examined \times 100 [%]) at each marker locus. (B) (Right) Sixty-four independent Dox(+)|Scelo(+) ES cell clones resistant to 2-FA (numbered as indicated at the top) were genotyped for SNP1 to SNP12. The 64 clones are divided into two groups: Group I, 55 clones carrying LOH with retention of the B6 allele (LOH^{B6}) at the SNP12 locus; Group II, the other nine clones carrying no LOH or LOH¹²⁹ at the SNP12 locus. (Left) Histogram representation of LOH ratio (%) at each marker locus of Group I clones. (C) Flow diagram of the SNP analysis performed in this study. The number of independent clones analyzed is shown in parentheses. (D) Detailed SNP analysis of the region between SNP11 and SNP12 in the clones shown in B. An additional five SNPs (u-50k, u-10k, u-5k, d-6k, and d-50k) were genotyped (Supplemental Table S1). (E) Ten independent Dox(-)|Scelo(+) ES cell clones resistant to 2-FA (numbered and categorized into Group Ic or Ilc as indicated) were genotyped for SNP1 to SNP12. In the following experiments, clones indicated by asterisks at the bottom (D and E) carry DELs involving this region. Clone II-73 (dagger) carries a missense mutation at exon 1 of the 129-derived *Aprt* gene. Clone Ilc-46 (double dagger) retains heterozygosity at the *Aprt* locus, and the mutation in the 129 allele is undefined.

complex patterns of LOH within the 50 kb surrounding the *Aprt* locus (Fig. 2D, Group II).

In addition, Geminin vector-mediated DSB induction also resulted in the isolation of a variety of 2-FA-resistant *Blm*^{tet/tet}*Aprt*^{neo/+} ES cell clones that had undergone LOH (Fig. 2C). All eleven independent clones isolated displayed LOH^{B6} at the SNP12 locus. Among them, four clones (36%) located the LOH breakpoints between the SNP11 and SNP12 loci: two clones within 5 kb and one clone within 10 kb from the *Aprt* locus (Supplemental Fig. S2B).

High-resolution array-CGH with high-density SNP analysis allows discrimination of DEL-associated LOH from other LOH patterns

The distinct clustering of LOH breakpoints in the vicinity of the *Aprt* gene locus, promoted by I-SceI-mediated DSB induction, allowed for in-depth analyses of LOH. We designed a custom high-resolution array-CGH platform covering the 8-Mb region from the telomere of mouse chromosome 8 with an average resolution of 300 bp, which encompasses the *Aprt* gene locus in the middle. With the custom array-CGH, we analyzed 30 independent clones with complex LOH patterns: 14 Group-I clones heterozygous for SNP u-50k, nine Group-II clones, three Geminin-transfected clones heterozygous for SNP u-50k, and four clones heterozygous for SNP11 as controls (Fig. 2C). The combination of array-CGH and SNP data allowed clear discrimination of DEL-associated LOH from the other LOH patterns (Fig. 1C) and prompted us to group the examined ES cell clones into three categories: clones with no detectable DEL (non-DEL), clones with DEL on the 129 allele (129-DEL), and clones with DEL on the B6 allele (B6-DEL) (Fig. 3; Supplemental Fig. S4). The “non-DEL” group contained 21 clones: 17 with ISceI transfection (I-14, -16, -17, -24, -40, -46, -48, -58, -59, -63, -77, II-15, -18, -68, -73, G-10, and -11) and four controls without ISceI transfection (8, 17, 24, and 30) (Fig. 2A,B). The other seven clones (clones I-47, -49, -57, II-2, -56, -69, and -74) carried a variety of large DELs on the 129 allele, ranging from 9.5 kb to 352 kb, and were categorized into the “129-DEL” group (Figs. 2D, 3A,B; Supplemental Table S2). The remaining two clones, II-13 (Fig. 2D) and G-4 (Supplemental Fig. S2B), were categorized into the “B6-DEL” group, in which DELs on the targeted B6 allele were accompanied by complex insertions at the breakpoints (Fig. 3C; Supplemental Table S2). Overall, given that the DELs were detected in the clones with DSB induction (9/26), but not in the controls with no DSB induction (0/4), it is likely that the DELs occurred during the process of DSB repair in *Blm*-deficient ES cells.

Blm-deficient ES cells with LOH but no detectable DELs undergo mitotic interhomolog recombination resulting in CO frequently associated with gene conversion

Among the above 17 non-DEL clones with ISceI transfection, only clone 73 displayed no CO points within the region (Fig. 2D). PCR amplification and direct sequencing revealed that this clone was a compound heterozygote (*Aprt*^{neo/-}) at the *Aprt* locus, with a *neo* insertion on the B6 chromosome and a spontaneous missense mutation on the 129 chromosome, which together conferred 2-FA resistance (data not shown). We then further analyzed the remaining 16 non-DEL clones with an additional nine SNP markers covering the 5-kb region between SNP u-5k and *Aprt* (Supplemental Fig. S5; Supplemental Table S1), demonstrating that at least one mitotic CO event occurred within this proximal flanking region of *Aprt* (Supplemental Fig. S5, arrows). Intriguingly,

the knocked-in I-SceI sites of ten clones (I-24, -40, -46, -48, -58, -59, II-15, -18, -68, and G-11) completely recovered the 129-derived wild-type genomic sequence after DSB induction (Supplemental Fig. S5), while those of the other six clones (I-14, -16, -17, -63, -77, and G-10) remained either intact (I-14, -16, -63, and G-10) or slightly modified (I-17 and -77) on the B6 allele (data not shown). The recovery of the *I-SceI* allele to the wild type can be explained by interhomolog NCO recombination in *Blm*-deficient ES cells, as explained in a previous paper (Larocque et al. 2011), and relatively short extents of gene conversion were observed (Supplemental Fig. S5, brackets). Similar interhomolog NCO recombination events were also detected in two other clones (I-47 and II-56) belonging to the 129-DEL group, whose rearrangement patterns are further illustrated in the following section (Fig. 3A,B).

Microhomology is present at the vast majority of genomic rearrangement breakpoints in *Blm*-deficient ES cells

Custom high-resolution array-CGH identified large DELs in seven 129-DEL clones and two B6-DEL clones. Among the seven 129-DEL clones, three clones (I-47, -49, and -57) carried DELs (ranging from 9.5 kb to 13 kb) at the flanking centromeric region of the *Aprt* locus on the 129 chromosome, associated with interhomolog CO of the targeted B6 chromosome distal to the DEL interval (Fig. 3A). In clone I-47, two 129-derived genomic DNA fragments were inserted into the DEL interval (Fig. 3A; Supplemental Fig. S6). The other four clones (II-2, -56, -69, and -74) carried DELs (ranging from 18 kb to 352 kb) encompassing the *Aprt* locus on the 129 chromosome with NCO (Fig. 3B). Notably, in clone II-56, a B6-derived short DNA fragment (<1.6 kb) flanked the centromeric breakpoint of the DEL on the 129-derived chromosome (Fig. 3B; Supplemental Fig. S6). A paucity of informative SNPs usually hampers detection of short interhomolog gene conversion tracts, leading to underestimation of the frequency of such DEL-associated gene conversions. The two B6-DEL clones (II-13 and G-4) carried 15-kb and 10-kb DELs on the B6 chromosomes, associated with interhomolog NCO and CO recombination of the *Aprt*-targeted B6 chromosomes, respectively (Fig. 3C). PCR cloning and sequencing of the DEL breakpoints revealed that a 432-bp DNA fragment from the neighboring *Galns* gene locus on the 129 chromosome was inserted in the inverted orientation into the DEL interval on the B6 chromosome in clone II-13, whereas a 132-bp mosaic DNA fragment with a seamless fusion of the 129- and B6-alleles of the *Aprt-Galns* intergenic region was inserted in the inverted orientation into the DEL interval on the B6 chromosome in clone G-4 (Fig. 3C; Supplemental Fig. S6). The presence of these breakpoint insertions is consistent with DNA copy number changes of the corresponding regions identified by array-CGH (Supplemental Fig. S4, red arrows).

At least nine large DELs and 13 rearrangement breakpoints were identified in the 26 independent *Blm*^{tet/tet}*Aprt*^{neo/+} ES cell clones that had been isolated by 2-FA selection after transient suppression of BLM and site-directed DSB induction. Remarkably, nine out of the 13 breakpoints (9/13, 69%) exhibited short junctional microhomologies of 2–7 bp that are shared with the proximal and distal reference sequences (Fig. 3; Supplemental Fig. S6). Moreover, although no microhomology was observed at the DEL breakpoints in clones I-49 and II-13 (the telomeric breakpoint), the other two breakpoints in clones I-57 and II-74 contained short insertion sequences of 3 and 8 bp, respectively, that appeared to have been templated from sequences near the breakpoints and again showed junctional microhomology between the original

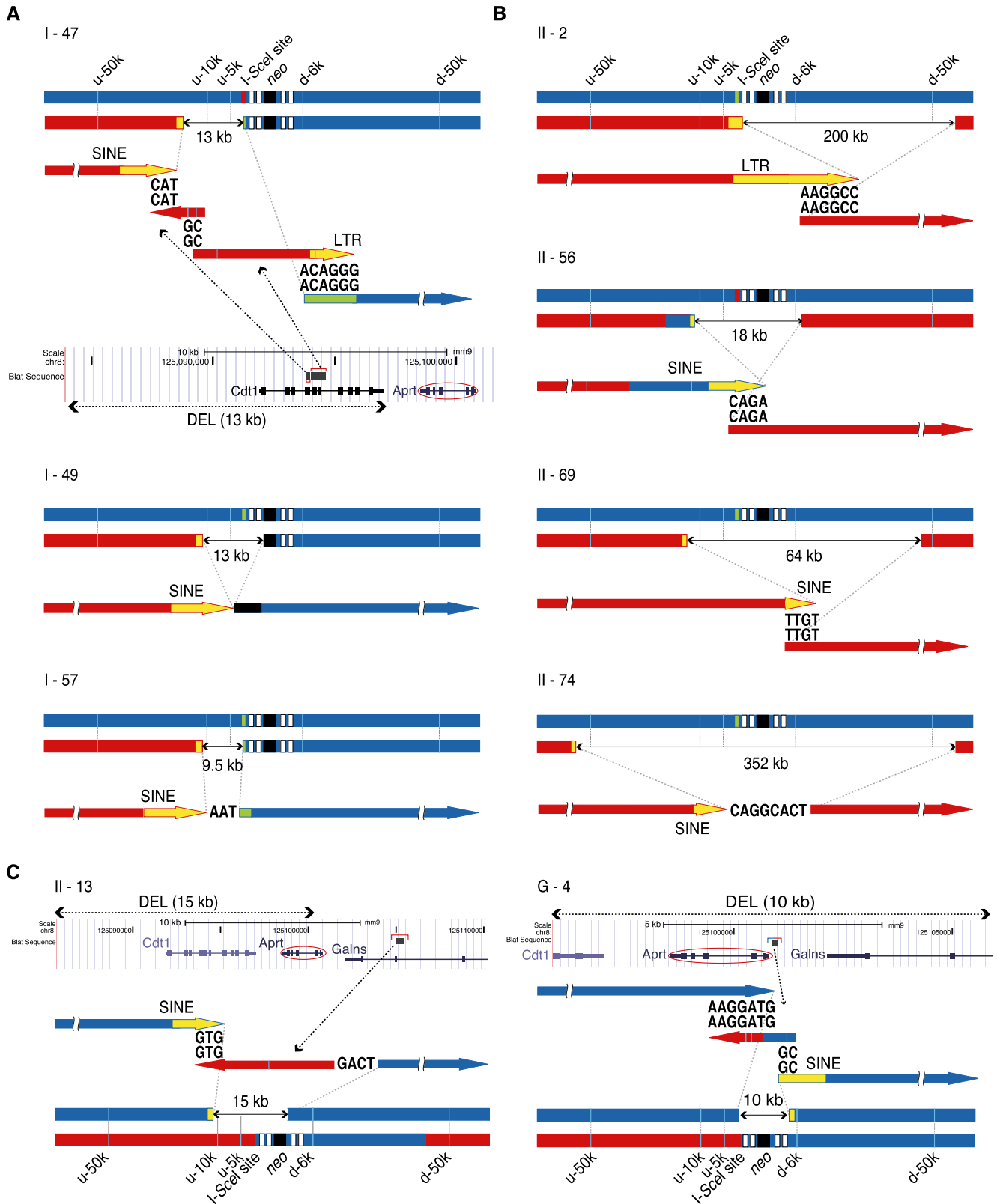


Figure 3. Schematic diagrams of DEL-associated LOH identified at the *Aprt* locus in *Blm^{tet/tet}Aprt^{neo/+}* ES cell clones that were subjected to transient BLM suppression and DSB induction. High-resolution array-CGH and a high-density SNP panel (Supplemental Table S1), followed by direct sequencing of the rearrangement breakpoints, were used. (Thick blue bar) B6-derived chromosome 8. (Thick red bar) 129-derived chromosome 8. (Black rectangle) *neo*-resistance gene cassette. (White rectangles) Exons of the *Aprt* gene. Blue (B6) and red (129) arrows represent a magnified view of the breakpoint boundaries. The direction of the arrows corresponds to chromosomal orientation from centromere to telomere in the mouse reference genome sequence (NCBI Build 37/mm9). Yellow segments represent genomic repetitive sequences such as a short interspersed nuclear element (SINE) or long terminal repeat (LTR) retrotransposon. Light green segments represent part of the I-SceI recognition site. Microhomology regions, shared by the two ends of the junctions, are indicated by double-stranded DNA sequences, while short DNA insertions are indicated by ssDNA sequences. (A,B) Clones carrying DEL on the 129 allele (129-DEL) associated with interhomolog crossover (CO) of the distal B6 chromosome (A) or noncrossover (NCO) (B). (C) Clones carrying DEL on the B6 allele (B6-DEL). The lengths of the genomic segments are depicted arbitrarily. Short 129-derived (red) segments embedded within the B6-derived (blue) chromosomes (A, I-47; B, II-56) represent interhomolog NCO events that copy the wild-type segment to the I-SceI DSB site, probably by gene conversion.

and the inserted sequence ends. Similar events have been observed in human and *Drosophila melanogaster* germline genome rearrangements (Yu and McVey 2010; Liu et al. 2011, 2012), and this is described in more detail in the following section. The vast majority (11/13, 85%) of the genomic rearrangement breakpoints arising under transient loss of BLM displayed microhomology.

In contrast to the *Blm*-deficient condition, a relatively small number of 2-FA-resistant colonies from Dox(-)IScelo(+) cell populations hampered the collective analysis of LOH breakpoints in *Blm*-proficient ES cells as controls (Fig. 2E). Indeed, we isolated only 10 independent clones (21%) from 47 *Blm*-proficient cell populations (Supplemental Fig. S7A), while isolating 43 independent clones (91%) from as many *Blm*-deficient cell populations (data not shown). Moreover, among the ten isolated clones (Fig. 2E), only one (clone Ic-34) carried a genomic rearrangement flanking the *Aprt* locus (Supplemental Fig. S7B). PCR cloning and sequencing

detected no breakpoint microhomology around the DEL breakpoint (Supplemental Fig. S7C; Supplemental Table S2). Thus, taken together, these results suggested that transient loss of BLM enhances microhomology-mediated genomic rearrangement, contributing, at least in part, to the high rate of somatic LOH.

Some short insertions at DEL breakpoints in *Blm*-deficient ES cells can be accounted for by the loop-out and/or snap-back models

As shown above, our sequence analysis identified 3- or 8-nt insertions at the DEL breakpoints in clones I-57 and II-74, respectively (Fig. 3A,B; Supplemental Fig. S6). Interestingly, these insertions formed at least part of short direct or inverted repeats that appeared on both sides of, or partially overlapped with, the DEL breakpoints. We propose a stepwise mechanism for these rearrangements (Fig. 4).

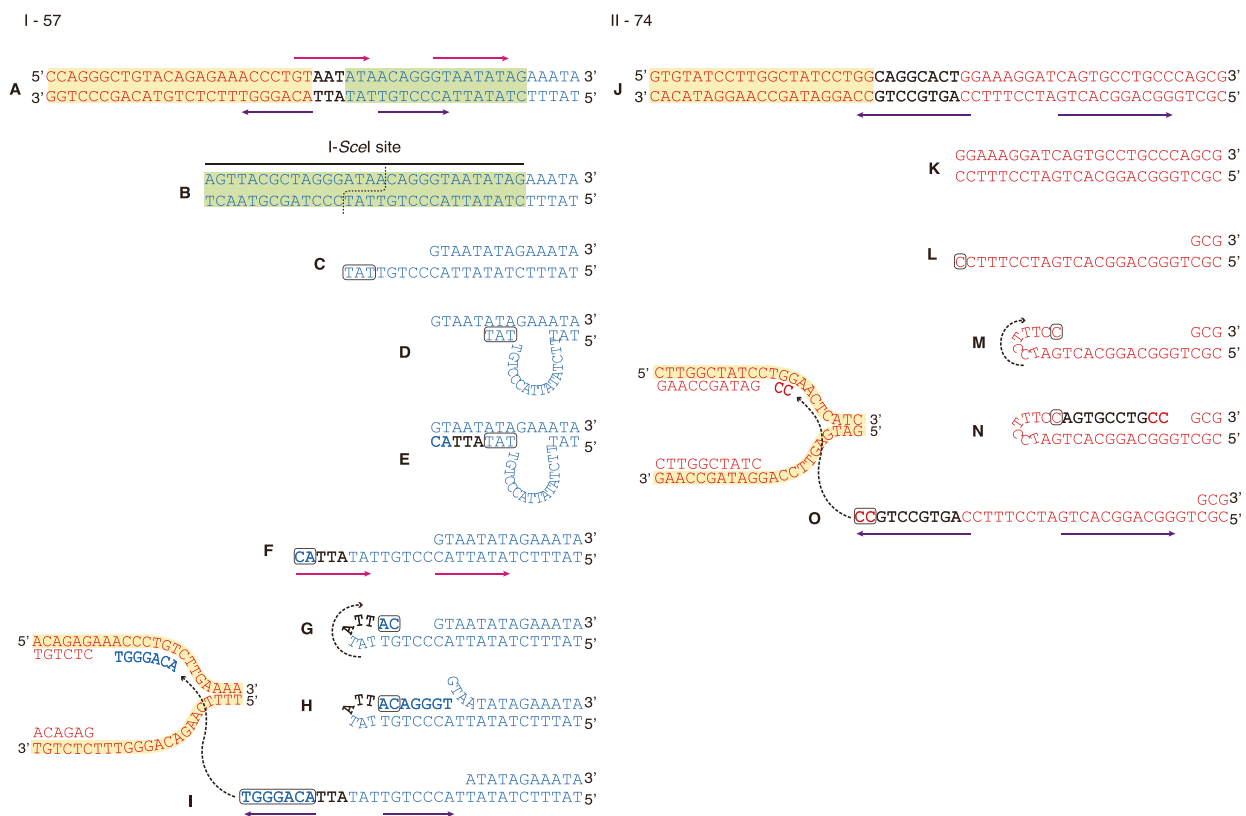


Figure 4. Templatd insertions at DEL breakpoints. (A) Sequence of the DEL breakpoint in clone I-57 (Fig. 3A). A net insertion of a 3-bp DNA stretch (5'-AAT-3', black bold typeface) was found between the 129-derived (red) 5' breakpoint (chr8: 125087464) and the B6-derived (blue) 3' breakpoint (I-SceI site inserted at chr8: 125097008) of the DEL interval. Part of the I-SceI recognition site is highlighted in light green. Pink and purple arrows indicate direct and inverted repeats, respectively. (B–I) Multi-step model based on the loop-out and snap-back models proposed in synthesis-dependent microhomology-mediated end joining (Yu and McVey 2010). (B) The I-SceI site on the B6 allele is cleaved by IScelo transfection. (C) The 5' end of the DSB site is resected to create a 3' ssDNA tail (bottom strand). (D) The 3' ssDNA tail end (5'-TAT-3', boxed) seeks out its annealing partner on the residual top strand (5'-ATA-3') by forming a transient looping-out structure and (E) primes strand extension along the template strand, resulting in (F) the formation of short direct repeats (5'-TATATTAC-3', pink arrows). (G) The extended 3' tail end (bottom) is then “snapped back” to form a hairpin structure for annealing of the two-base oligonucleotides (5'-AC-3', boxed) to the complementary sequence (5'-GT-3') on the same strand. (H) A template-dependent DNA polymerase extends the 3' end, removing some nucleotides on the top strand. (I) The hairpin structure is unwound, and the short inverted repeats (5'-ACCCTGT-3', purple arrows) are formed. The newly synthesized 3' end is then annealed to ssDNA 9.5 kb centromeric to the DSB site via the microhomology (5'-ACAGGGT-3', blue bold typeface). Interestingly, the annealing partner resides within a genomic inverted repeat (highlighted in yellow) as shown in Figure 5. (J) Sequence of the DEL breakpoint in clone II-74 (Fig. 3B). A net insertion of an 8-bp DNA stretch (5'-CAGGCACT-3', black bold typeface) is found between the 129-derived 5' breakpoint (chr8: 124905418) and 3' breakpoint (chr8: 125257681) of the DEL interval. (K–O) Multistep model: (K) A one-ended DSB is produced 160 kb telomeric to the *Aprt* locus by an undefined mechanism. (L) The 5' end of the DSB site is resected to create a 3' ssDNA tail (bottom strand). (M) The 3' tail end is snapped back to form a hairpin structure via one-base microhomology (boxed), and (N) primes strand extension along the template strand. (O) The hairpin structure is unwound, and the short inverted repeats (5'-CAGTGCCTGCC-3', purple arrow) are formed. The newly synthesized 3' end is then annealed to ssDNA within a genomic repeat (highlighted in yellow), 192 kb centromeric to the *Aprt* locus via microhomology (5'-CC-3', boxed).

The complicated junctional sequences can be adequately accounted for by the “loop-out” and/or “snap-back” mechanisms described in a recently proposed DSB repair model in *Drosophila melanogaster*, termed synthesis-dependent microhomology-mediated end joining (Yu and McVey 2010). In clone I-57, for example, the induction of I-SceI endonuclease in Dox-treated *Blm^{tet/tet}Aprt^{neo/+}* ES cells cleaves the knocked-in I-SceI recognition site centromeric to *Aprt* on the B6 allele, yielding DSBs of which the 5' ends are resected and 3' single-stranded DNA (ssDNA) tails are formed (Fig. 4A–C). The exogenous DNA sequence of each DSB end does not have its counterpart to which to anneal on the homologous chromosome; therefore, normal interhomolog DSB repair by homologous recombination may be unfavorable. Instead, the 3' ssDNA tail can seek out its annealing partner by taking advantage of the first few nucleotides, which mediate formation of transient secondary structures by “looping-out” or “snapping-back” motions and prime strand extension, resulting in the templated insertion-mediated junction formation (Fig. 4D–I). The snap-back model is also applicable to the junctional sequence observed in clone II-74, although the cause of the initial DSB at the telomeric DEL breakpoint remains unclear (Fig. 4J–O). In either case, after scanning the flanking DNA strands in this way, the newly synthesized 3' end is joined via microhomology to a sequence somewhat distant from the DSB. Intriguingly, the other ends of these DELs were both located within genomic inverted repeats, as shown in the next section.

Breakpoints of DELs in *Blm*-deficient ES cells are largely clustered in genomic inverted repeats

We sought to dissect the mechanism of DEL formation further by mapping the breakpoint sequences to mouse chromosome 8. This revealed that the centromeric DEL breakpoints in the majority (6/9) of the 129-DEL or B6-DEL clones (Fig. 3) were located within a distinct 4.2-kb repeat-rich region between SNP markers u-15k and u-10k (Fig. 5A). A self-alignment dot plot of the 4.2-kb repeat-rich region generated using BLAST2 (Tatusova and Madden 1999) demonstrated that all the centromeric breakpoints of the DELs were located in intrachromosomal inverted repeats consisting of short interspersed nuclear elements (SINES) or long terminal repeat (LTR) retrotransposons (Fig. 5A). Likewise, the telomeric breakpoint of the DEL in clone G-4 was located within two tandemly inverted (i.e., perfectly palindromic) 135-bp SINE repeats (Fig. 5B). Thus, the majority (7/9) of the 129-DEL and B6-DEL clones, in which *Blm* expression had been transiently switched off, underwent DELs, and at least one of the breakpoints was located in an inverted repeat.

Discussion

We present an in-depth experimental analysis of the immediate effects of functional loss of BLM on genomic integrity, which are difficult to assess retrospectively in affected cancer cells. Transient suppression of *Blm* expression in mouse ES cells increased the rate of LOH (Fig. 1F), the breakpoints of which were uniformly distributed along the chromosome (Fig. 2A; Supplemental Fig. S3B,E). The elevated rate of LOH was further enhanced by I-SceI-mediated site-directed DSB induction (Fig. 1F). DSB is a common event in any genomic region in living cells, resulting from exposure to intrinsic or environmental stress; therefore, such spontaneous DSBs may underlie the wide range of cancers arising in Bloom syndrome individuals. This motivated us to examine the genomic alterations at higher resolution. Conveniently, site-directed DSB induction

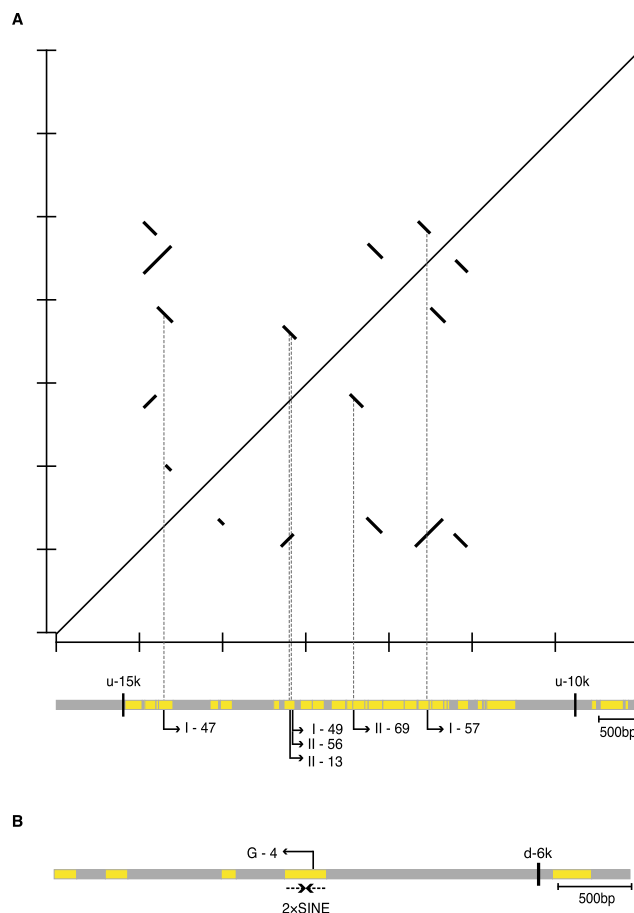


Figure 5. Clustering of DEL breakpoints in genomic inverted repeats. (A) Centromeric DEL breakpoints (rightward bent arrows) in six out of nine 129-DEL or B6-DEL clones (Fig. 3) are located within a distinct 4.2-kb repeat-rich region between SNP markers u-15k and u-10k (Supplemental Table S1). Yellow rectangles represent repetitive sequences. A self-alignment dot plot using BLAST2 (Tatusova and Madden 1999) of a DNA fragment (chr8: 125082686–125090184) containing the 4.2-kb repeat-rich region is shown. Lines with a slope of +1 indicate direct repeats, while lines with a slope of –1 indicate inverted repeats. (B) The telomeric DEL breakpoint (leftward bent arrow) in clone G-4 is localized within tandemly inverted (i.e., perfectly palindromic) $2 \times$ 135-bp SINE repeats. (d-6k) SNP marker located 6 kb telomeric to the *Aprt* locus (Supplemental Table S1).

was found to concentrate LOH events in a specific region in *Blm*-deficient ES cells (Fig. 2B), facilitating the detailed characterization of LOH by a combination of region-specific high-resolution array-CGH, high-density SNP analyses, and direct sequencing of the rearrangement breakpoints.

Possible molecular mechanisms for genomic rearrangements induced by a combination of transient loss of BLM and site-directed DSB induction

Site-directed DSB induction during temporary loss of *Blm* expression resulted in varying patterns of LOH affecting the *Aprt* gene locus on distal chromosome 8 in mouse *Blm^{tet/tet}Aprt^{neo/+}* ES cells. A plausible explanation for the molecular mechanisms underlying such genomic rearrangements is shown in Figure 6. In the initiating step of DSB repair, the 5' ends of the DSB are resected to yield 3' ssDNA tails, which invade the homologous DNA duplex forming

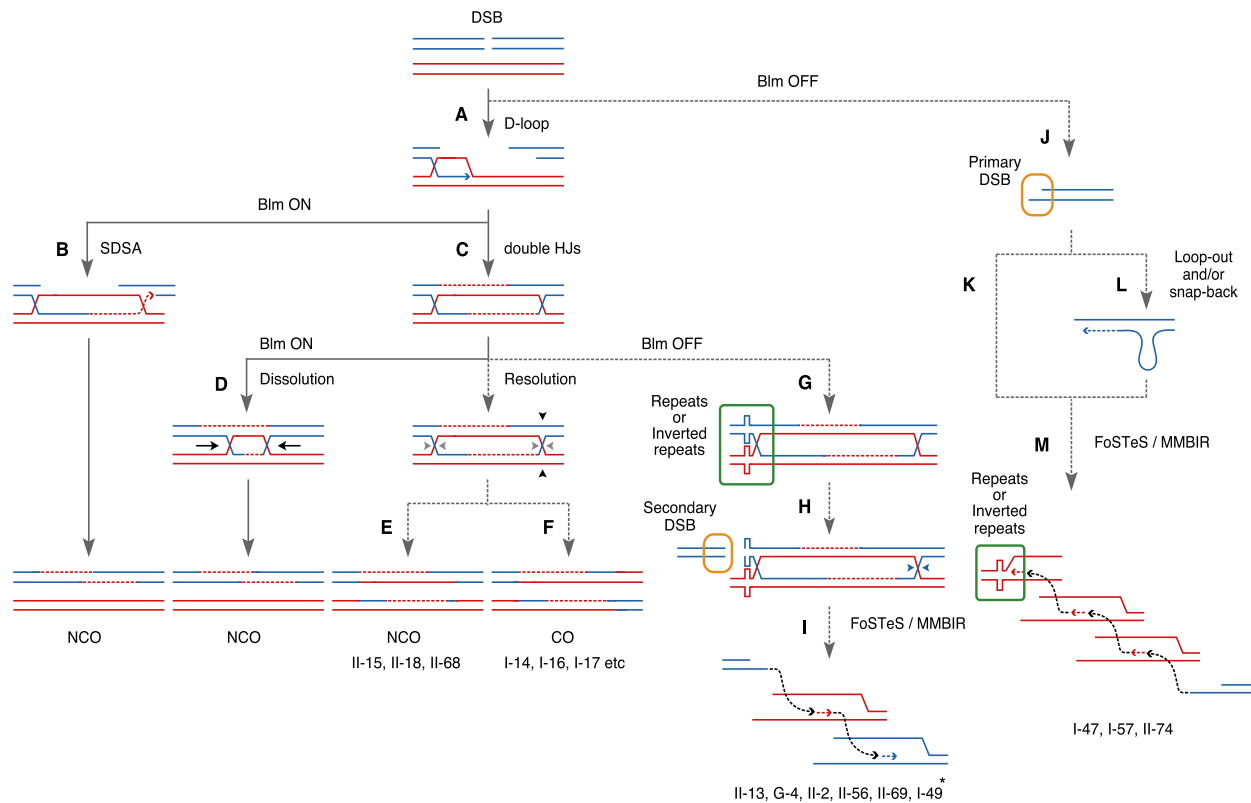


Figure 6. Proposed molecular mechanisms for genomic rearrangements observed in the present study. Blue and red bars represent the B6-derived and 129-derived chromosomes, respectively. (Top) A site-directed DSB is induced at the knocked-in I-SceI site near the *Aprt* locus on the B6-derived chromosome 8, leading to the activation of cellular DSB repair pathways. The following steps (A–M) are described in detail in the text. Solid arrows represent pathways that are dominant in the *Blm*-proficient (*Blm* ON) conditions, while dotted arrows represent pathways that become predominant in the *Blm*-deficient (*Blm* OFF) conditions. Corresponding clone names are indicated at the bottom. (*) Clone devoid of apparent breakpoint microhomology.

a displacement-loop (D-loop) (Fig. 6A; Chen et al. 2007). The invading strand is extended by DNA synthesis and often displaced from the template strand to anneal to the other 3' end of the DSB, leading to the exclusive formation of NCO products. This pathway, termed synthesis-dependent strand-annealing (SDSA), is, however, suppressed by *Blm* deficiency at least in *Drosophila melanogaster* (Fig. 6B; Adams et al. 2003). Instead, the D-loop intermediate is converted to double Holliday junctions (double HJs) (Fig. 6C). The BLM helicase, which forms a complex with DNA topoisomerase III α , acts to enhance “convergent migration,” resulting in “dissolution” of double HJs and generation of NCO products (Fig. 6D; Wu and Hickson 2003). In contrast, the absence of BLM would suppress the dissolution pathway and increase the relative contribution of an alternative “resolution” pathway of double HJs, whereby DNA strands at one of the double HJs are cleaved and religated to generate either NCO (Fig. 6E, II-15, -18, and -68) or CO (Fig. 6F, I-14, -16, and -17) products depending on the orientation of the cleavage strands. The extent of branch migration of HJs could be limited by encountering an unusual secondary DNA structure generated by genomic repeats or genomic inverted repeats (Fig. 6G). Then, the stalled HJ would be disrupted by an as-yet-unidentified endogenous endonuclease to yield an unpaired DSB end (referred to as a “secondary DSB” in Fig. 6H) either on the I-SceI-cleaved B6 chromosome (B6-DEL: clones II-13 and G-4) or on the template 129 chromosome (129-DEL: clones I-49, II-2, -56, and -69). Considering the presence of microhomology at most of the breakpoints, one can assume that these secondary DSB ends are engaged

in microhomology-mediated genomic rearrangements. Two major models, replication-independent and replication-dependent, have been suggested for the microhomology-mediated rearrangement mechanism. The former includes alternative nonhomologous end joining (alt-NHEJ; also known as microhomology-mediated end joining) (McVey and Lee 2008; Simsek and Jasin 2010), while the latter includes fork stalling and template switching (FoSTeS) (Lee et al. 2007) and its derivative, microhomology-mediated break-induced replication (MMBIR) (Hastings et al. 2009; Zhang et al. 2009). It is difficult to distinguish these two major models based only upon the observation of breakpoint sequences, and they are not necessarily mutually exclusive. For example, the DEL rearrangements observed in clones II-2, II-69, and II-74 (Fig. 3B) affect only the 129-derived chromosomal stretches and could be explained by alt-NHEJ. It would be the case if two unpaired DSB ends arise at the same time in the same cell, and this possibility is not ruled out by the present analysis. Another more parsimonious explanation, however, would be provided by the FoSTeS or MMBIR model (Fig. 6I; Arlt et al. 2012). Importantly, previous studies have shown that BLM is required for efficient replication-fork restarting, and a defect in BLM leads to increased replication fork stalling (Davies et al. 2007; Rao et al. 2007; Machwe et al. 2011), which is a prerequisite for the FoSTeS mechanism. Accordingly, in the FoSTeS/MMBIR model, the secondary DSB end is annealed via microhomology to ssDNA arising elsewhere under replication fork stalling and is extended along the given template strand. The microhomology annealing occurs in a chromosomal orientation-independent

manner, which could account for the presence of inverted insertion segments at the breakpoints, as observed in clones I-47, II-13, and G-4 (Fig. 3A,C), and also in recently published works on human germline and somatic genome rearrangements (Conrad et al. 2010; Arlt et al. 2011; Mills et al. 2011). The template-switching and extension can be repeated until the repair process is completed.

Besides the secondary DSB end discussed above, a primary DSB end generated by I-SceI cleavage (Fig. 6J) could directly provide a 3' ssDNA tail, which seeks other ssDNA templates to which to anneal via microhomology and to extend along using a replication mechanism instead of HJ formation (Fig. 6K, clone I-47). The I-SceI breakpoint on the B6-derived chromosome does not have its counterpart to which to anneal on the homologous 129-derived chromosome; therefore, the 3' ssDNA tail could scan the flanking DNA strands and occasionally form transient loop-out or snap-back structures, which may be maintained for longer in a BLM helicase-inactive state (Fig. 6L, clones I-57 and II-74). During the formation of transient structures, a short stretch of DNA is synthesized along a template strand and makes a breakpoint insertion, the microhomology end of which mediates a rearrangement with another ssDNA end arising elsewhere in repeats or inverted repeats (Fig. 6M). The occasional finding of templated insertions at breakpoints (Fig. 4) is readily attributed to DNA polymerase activity, further supporting the replication-dependent FoSTeS/MMBIR model, rather than replication-independent end-joining models, as a comprehensive interpretation of the rearrangement mechanism in the *Blm*-deficient situation. FoSTeS/MMBIR is thought to involve the repair mechanism of one-ended DSBs arising from collapsed replication forks (Liu et al. 2012), whereas, in the present experimental study, the rearrangement mechanism should involve two-ended DSBs generated by I-SceI cleavage, which has not been implicated in previously proposed break-induced replication (BIR) models, with the exception of one-ended DSBs induced by I-SceI cleavage of prokaryotic circular genomes (Shee et al. 2012). So far, the limited scale of the current experiment has hampered the complete tracing of the two DNA ends generated by a single I-SceI cleavage and has resulted in fate tracing of either end of the break (Fig. 3A, clones I-47 and I-57), which does not necessarily rule out the possibility that the two ends are independently repaired by the replication-dependent mechanism. Hicks et al. recently suggested the presence of a novel microhomology-mediated SDSA pathway in yeast (Hicks et al. 2010), which could be applicable to interpret the potential replication-dependent repair mechanism of two-ended DSBs observed in mammalian genomes.

Within the current experimental framework, there is a practical trade-off between short-range high-resolution genomic analyses and wide-range low-resolution genomic analyses. In the present study, the negative selection scheme by 2-FA resistance allowed focused high-resolution analyses of LOH involving the *Aprt* locus on chromosome 8 and detailed characterization of the LOH patterns arising in almost all clones tested (Fig. 2). However, this was at the expense of genome-wide characterization of rearrangement events, including interchromosomal translocations and more complex alterations. Array-CGH is a reliable method for screening for copy-number variations, and our custom high-resolution array-CGH is sensitive for discerning the difference at least between one, two, and three copies of genomic segments on distal chromosome 8, which was cross-validated by direct sequencing (Supplemental Fig. S4). Nevertheless, in principle, it cannot detect balanced chromosomal inversions and translocations that often occur at deletion breakpoints. Next-generation sequencing technology could contribute to the detection of such balanced rearrangements

and more complex alterations. In fact, recent studies have shown that extraordinarily complex somatic rearrangements like chromothripsis display less junctional microhomology than germline rearrangements (Stephens et al. 2011; Malhotra et al. 2013), and germline balanced rearrangements display less junctional microhomology than unbalanced ones (Chiang et al. 2012). Genome-wide rearrangement analyses using next-generation sequencing platforms are currently in progress for our Dox-inducible *Blm*-knockdown cell lines.

The localization of DEL breakpoints within annotated genomic repeats (Fig. 5) can also be seen in *Drosophila melanogaster Blm* mutants (Garcia et al. 2011), suggesting the existence of a common mechanism for DEL formation in a *Blm*-deficient background. BLM helicase acts to unwind secondary DNA structures at the telomere (Sfeir et al. 2009). The loss of BLM may therefore contribute to prolonged maintenance of the unusual secondary DNA structures that appear in regions of repetitive sequence, such as inverted repeats, which may hamper normal replication fork progression and/or HJ migration.

Potential relevance of transient loss of BLM to human carcinogenesis

Bloom syndrome, an autosomal recessive human disorder of the *BLM* gene, predisposes patients to a wide variety of cancers (German 1997). The disorder is an inherited and lifelong condition; therefore, the exact contribution of BLM dysfunction to the time course of cancer formation is difficult to discern in clinical human specimens. The Dox-controlled shutoff of the *Blm* gene expression in our mouse *Blm^{tet/tet}* ES cell system, however, enabled us to dissect the immediate effects of transient loss of BLM on genomic integrity. When combined with DSB introduction, the transient loss of BLM resulted in microhomology-mediated genomic rearrangements within a period as short as several days, representing a significant fraction of the induced LOH events. This might point to an early contribution of BLM dysfunction to cancer development. Moreover, this underscores the importance of such genomic rearrangements, in addition to increased generation of homozygosity for tumor-suppressor gene mutation, as a mechanism of cancer predisposition in Bloom syndrome. The *BLM* gene is a typical housekeeping gene and is constitutively expressed in most normal somatic cells. However, the BLM protein is a target of the proteolytic enzyme caspase 3 and can be cleaved into a truncated form that has lost its capability to interact with topoisomerase III α (Freire et al. 2001). Our in-depth genomic analysis will provide a valuable foundation for understanding early genomic alterations not only in cancers associated with Bloom syndrome, but also in a wide range of human cancers.

Methods

Vector construction

Construction of all the plasmid vectors used in this study is described in the Supplemental Methods.

Cell lines

Mutant mouse ES cell lines, *Blm^{tet/tet}FasI^{neo/+}*, *Blm^{tet/tet}FasI^{neo/+}Nanos^{fluro Δ TK/+}*, and *Blm^{tet/tet}Aprt^{neo/+}*, were generated by targeted homologous recombination. Detailed procedures are described in the Supplemental Methods.

Cell culture, transfection, and polymorphic marker analysis

Blm^{tet/tet}Aprt^{neo/+} ES cells were pretreated with 0.1 mM adenine and 20 μ g/mL alanosine-containing medium for 4 d to eliminate pre-existing *Aprt*-deficient cells (Hong et al. 2006). Subsequent procedures for culture, transfection, and colony isolation are shown in Figure 1E and detailed in the Supplemental Methods. To avoid double counting, each clone was isolated from a different culture dish. Genomic DNA was extracted by a standard phenol/chloroform method, and SNPs were genotyped by direct sequencing of PCR products (Supplemental Table S1).

Functional assay of I-SceI endonuclease activity

The pIRES-TK-EGFP reporter plasmid (Ogiwara et al. 2011) (kindly provided by Dr. T. Kohno, National Cancer Center Research Institute, Tokyo, Japan) was cotransfected with either pPB-CAG-IB (Control), pPB-CAG-I-SceI-IB (ISceI), or pPB-CAG-I-SceI-IB (ISceI) into 293FT cells using Lipofectamine 2000 (Life Technologies). EGFP-expressing cells were analyzed by BD FACSCanto II (BD Biosciences) 1 d after transfection. Alternatively, *Blm^{tet/tet}Aprt^{neo/+}* ES cells were transfected with the ISceI expression vectors and selected by blasticidin (100 μ g/mL) for 4 d. The resistant cells were harvested to extract genomic DNA, and the 3.7-kb region encompassing the I-SceI cleavage site was amplified by PCR using primers ISceI-1K-U and PGKpro-L2. The resulting PCR products were digested in vitro with I-SceI endonuclease (New England Biolabs) to detect loss of the I-SceI site, indicative of imprecise nonhomologous end-joining occurring after I-SceI cleavage in vivo. The sequences of the primers used in this assay are: ISceI-1K-U, 5'-CTCACTATGGAGGTGGTCTGTGCAAGGATG-3', and PGKpro-L2, 5'-CGGGCTGCTAAAGCGCATGCTCCAGACTG-3'.

Analysis of vector-mediated I-SceI expression associated with the cell cycle

293FT cells were transfected with HA-tagged *I-SceI* expression vectors pPB-CAG-ISceI-IB (ISceI) or pPB-CAG-ISceI-Geminin-IB (Geminin) using Lipofectamine 2000 (Life Technologies). After 2 d of culture, the cells were harvested by trypsinization, fixed/permeabilized with BD Cytotfix/Cytoperm solution and BD Perm/Wash buffer (BD Biosciences), and incubated with a monoclonal antibody against HA (12CA5; Roche Diagnostics). The cells were then incubated with an Alexa Fluor488-conjugated goat anti-mouse IgG secondary antibody, followed by staining with Hoechst 33342 (both from Life Technologies). Flow cytometric analysis was performed using a BD FACSAria II (BD Biosciences), and the data were analyzed using FlowJo software (TreeStar).

High-resolution array-CGH analysis

High-density probes with an average resolution of 300 bp were designed within the 8-Mb telomeric region of chromosome 8 in triplicate for an Agilent 4 \times 180 K CGH array (Agilent Technologies). The array also includes a control grid (IS-180880-4-V2_4by180K_CGH_Mm_20090225), a normalization probe group (Mouse_CGH_3k_Agilent Normalization Probe Group), and a replicate probe group (Mouse_CGH_1k_Agilent Replicate Probe Group). The remaining features were filled with probes from the 4 \times 180 K mouse CGH probe group. The array design was deposited with EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under the accession number A-MEXP-2217. Genomic DNA was labeled with either Cy3 (reference) or Cy5 (sample), and hybridized to the arrays. The arrays were scanned with an Agilent microarray scanner, and raw data were generated by Agilent Feature Extraction software. CGH calls were made with Agilent DNA Analytics software using the

ADM2 algorithm (6.0 threshold) with a minimum of three probes in the region as a filter.

Self-alignment dot plot analysis

The DNA sequence was compared with itself using the BLAST2 program (Tatusova and Madden 1999) with default parameters (<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>).

Data access

The array-CGH data have been deposited with EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-1116.

Acknowledgments

We thank Koichi Kawakami (National Institute of Genetics, Mishima, Japan) for providing plasmid pT2AL200R150G; Allan Bradley (The Wellcome Trust Sanger Institute) for plasmids pPB-SB-SA- β geo and mPB; Masaru Okabe (Osaka University, Japan) for plasmid pCAGGS-EGFP; and Takashi Kohno (National Cancer Center Research Institute, Tokyo, Japan) for plasmid pIRES-TK-EGFP. We also thank M. A. Li and C. McGee (The Wellcome Trust Sanger Institute) for CGH analysis; Yasushi Kuromi and Junko Yoshida (Osaka University, Japan) for help with plasmid preparation; and Vincent W. Keng (The Hong Kong Polytechnic University, Hong Kong) for critical reading of the manuscript. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (no. 24116709); and a Wellcome Trust grant (no. 098051). The authors declare that they have no conflicts of interest.

References

- Adams MD, McVey M, Sekelsky JJ. 2003. *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* **299**: 265–267.
- Arlt MF, Ozdemir AC, Birkeland SR, Lyons RH Jr, Glover TW, Wilson TE. 2011. Comparison of constitutional and replication stress-induced genome structural variation by SNP array and mate-pair sequencing. *Genetics* **187**: 675–683.
- Arlt MF, Rajendran S, Birkeland SR, Wilson TE, Glover TW. 2012. *De novo* CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. *PLoS Genet* **8**: e1002981.
- Cervantes RB, Stringer JR, Shao C, Tischfield JA, Stambrook PJ. 2002. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci* **99**: 3586–3590.
- Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: Mechanisms, evolution and human disease. *Nat Rev Genet* **8**: 762–775.
- Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren AM, Rudiger SR, et al. 2012. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet* **44**: 390–397.
- Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, Turner DJ, Hurler ME. 2010. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat Genet* **42**: 385–391.
- Davies SL, North PS, Hickson ID. 2007. Role for BLM in replication-fork restart and suppression of origin firing after replicative stress. *Nat Struct Mol Biol* **14**: 677–679.
- Eggan K, Akutsu H, Loring J, Jackson-Grusby L, Klemm M, Rideout WM III, Yanagimachi R, Jaenisch R. 2001. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci* **98**: 6209–6214.
- Freire R, d'Adda Di Fagagna F, Wu L, Pedrazzi G, Stagljar I, Hickson ID, Jackson SP. 2001. Cleavage of the Bloom's syndrome gene product during apoptosis by caspase-3 results in an impaired interaction with topoisomerase III α . *Nucleic Acids Res* **29**: 3172–3180.
- Garcia AM, Salomon RN, Witsell A, Liepkalns J, Calder RB, Lee M, Lundell M, Vijg J, McVey M. 2011. Loss of the bloom syndrome helicase increases

- DNA ligase 4-independent genome rearrangements and tumorigenesis in aging *Drosophila*. *Genome Biol* **12**: R121.
- German J. 1997. Bloom's syndrome. XX. The first 100 cancers. *Cancer Genet Cytogenet* **93**: 100–106.
- Hastings PJ, Ira G, Lupski JR. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* **5**: e1000327.
- Hicks WM, Kim M, Haber JE. 2010. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* **329**: 82–85.
- Hong Y, Cervantes RB, Stambrook PJ. 2006. DNA damage response and mutagenesis in mouse embryonic stem cells. *Methods Mol Biol* **329**: 313–326.
- Horie K, Kokubu C, Yoshida J, Akagi K, Isotani A, Oshitani A, Yusa K, Ikeda R, Huang Y, Bradley A, et al. 2011. A homozygous mutant embryonic stem cell bank applicable for phenotype-driven genetic screening. *Nat Methods* **8**: 1071–1077.
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**: 289–294.
- Kinzler KW, Vogelstein B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**: 763.
- Larocque JR, Stark JM, Oh J, Bojilova E, Yusa K, Horie K, Takeda J, Jasin M. 2011. Interhomolog recombination and loss of heterozygosity in wild-type and Bloom syndrome helicase (BLM)-deficient mammalian cells. *Proc Natl Acad Sci* **108**: 11971–11976.
- Lee JA, Carvalho CM, Lupski JR. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**: 1235–1247.
- Liu P, Erez A, Nagamani SC, Dhar SU, Kolodziejaska KE, Dharmadhikari AV, Cooper ML, Wiszniewska J, Zhang F, Withers MA, et al. 2011. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **146**: 889–903.
- Liu P, Carvalho CM, Hastings PJ, Lupski JR. 2012. Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev* **22**: 211–220.
- Luo G, Santoro IM, McDaniel LD, Nishijima I, Mills M, Youssoufian H, Vogel H, Schultz RA, Bradley A. 2000. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat Genet* **26**: 424–429.
- Machwe A, Karale R, Xu X, Liu Y, Orren DK. 2011. The Werner and Bloom syndrome proteins help resolve replication blockage by converting (regressed) Holliday junctions to functional replication forks. *Biochemistry* **50**: 6774–6788.
- Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM. 2013. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* **23**: 762–776.
- McVey M, Lee SE. 2008. MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet* **24**: 529–538.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* **470**: 59–65.
- Ogiwara H, Ui A, Otsuka A, Satoh H, Yokomi I, Nakajima S, Yasui A, Yokota J, Kohno T. 2011. Histone acetylation by CBP and p300 at double-strand break sites facilitates SWI/SNF chromatin remodeling and the recruitment of non-homologous end joining factors. *Oncogene* **30**: 2135–2146.
- Rao VA, Conti C, Guirouilh-Barbat J, Nakamura A, Miao ZH, Davies SL, Sacca B, Hickson ID, Bensimon A, Pommier Y. 2007. Endogenous γ -H2AX-ATM-Chk2 checkpoint activation in Bloom's syndrome helicase-deficient cells is related to DNA replication arrested forks. *Mol Cancer Res* **5**: 713–724.
- Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, et al. 2008. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**: 487–498.
- Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**: 90–103.
- Shee C, Gibson JL, Rosenberg SM. 2012. Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. *Cell Rep* **2**: 714–721.
- Simsek D, Jasin M. 2010. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol* **17**: 410–416.
- Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, et al. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**: 27–40.
- Tatusova TA, Madden TL. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247–250.
- Wu L, Hickson ID. 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**: 870–874.
- Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, Nellaker C, Goodstadt L, Nicod J, Bhomra A, et al. 2011. Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**: 326–329.
- Yu AM, McVey M. 2010. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res* **38**: 5706–5717.
- Yusa K, Horie K, Kondoh G, Kouno M, Maeda Y, Kinoshita T, Takeda J. 2004. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**: 896–899.
- Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* **41**: 849–853.

Received November 29, 2012; accepted in revised form May 14, 2013.