

Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells

Bjoern Schwer^{a,b,c,1,2}, Pei-Chi Wei (魏珮琪)^{a,b,c,1}, Amelia N. Chang^{a,b,c}, Jennifer Kao^{a,b,c}, Zhou Du^{a,b,c}, Robin M. Meyers^{a,b,c}, and Frederick W. Alt^{a,b,c,2}

^aHoward Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115; ^bDepartment of Genetics, Harvard Medical School, Boston, MA 02115; and ^cDepartment of Pediatrics, Harvard Medical School, Boston, MA 02115

Contributed by Frederick W. Alt, December 25, 2015 (sent for review December 19, 2015; reviewed by David Ferguson, Barry Sleckman, and Kefei Yu)

High-throughput, genome-wide translocation sequencing (HTGTS) studies of activated B cells have revealed that DNA double-strand breaks (DSBs) capable of translocating to defined bait DSBs are enriched around the transcription start sites (TSSs) of active genes. We used the HTGTS approach to investigate whether a similar phenomenon occurs in primary neural stem/progenitor cells (NSPCs). We report that breakpoint junctions indeed are enriched around TSSs that were determined to be active by global run-on sequencing analyses of NSPCs. Comparative analyses of transcription profiles in NSPCs and B cells revealed that the great majority of TSS-proximal junctions occurred in genes commonly expressed in both cell types, possibly because this common set has higher transcription levels on average than genes transcribed in only one or the other cell type. In the latter context, among all actively transcribed genes containing translocation junctions in NSPCs, those with junctions located within 2 kb of the TSS show a significantly higher transcription rate on average than genes with junctions in the gene body located at distances greater than 2 kb from the TSS. Finally, analysis of repair junction signatures of TSS-associated translocations in wild-type versus classical nonhomologous end-joining (C-NHEJ)-deficient NSPCs reveals that both C-NHEJ and alternative end-joining pathways can generate translocations by joining TSS-proximal DSBs to DSBs on other chromosomes. Our studies show that the generation of transcription-associated DSBs is conserved across divergent cell types.

nonhomologous end-joining | neural stem cells | transcription | translocation | alternative end-joining

The integrity of mammalian genomes is constantly challenged by DNA double-strand breaks (DSBs) that are generated by cell-intrinsic processes such as DNA replication, transcription, and oxidative stress (1, 2). In addition, during early lymphocyte development the RAG1/2 endonuclease specifically generates DSBs at antigen receptor variable region gene segments; these segments then are assembled to generate variable region exons (3). Likewise, in activated, mature B lymphocytes, activation-induced cytidine deaminases (AID) initiate DSB formation in the noncoding “switch” regions of the Ig heavy chain (IgH) locus to exchange IgH constant region loci to effect IgH class switch recombination (CSR) (3). Recently, our studies of neural stem/progenitor cells (NSPCs) further revealed that recurrent DSB clusters (RDCs) occur within the bodies of a set of at least 27 long genes that mostly are involved in neural functions (4). DSBs around transcription start sites (TSSs) have been implicated in the transcriptional activation of nuclear hormone receptor genes in tumor cell lines (5, 6) and are required for the transcription of a set of early-response genes in activated neurons in culture (7). We also have found that activated B cells in culture generate genomic DSBs at active TSSs, as measured by global nuclear run-on sequencing (GRO-seq) (8) and that these TSS-associated DSBs can translocate to recurrent “bait” DSBs introduced in the body of the *c-Myc* gene (9).

One of the two major, well-characterized DSB repair pathways in mammalian cells is classical nonhomologous end-joining (C-NHEJ). C-NHEJ is active throughout the cell cycle but is particularly critical in noncycling cells in which the other major DSB repair pathway, homologous recombination, is not available (10). C-NHEJ joins blunt DSB ends directly and also can join ends with several base pairs of short microhomology (MH) to form MH-mediated joins (11). C-NHEJ is required for V(D)J recombination joining and plays a major role in CSR joining (11). C-NHEJ also appears to play a significant role in joining NSPC RDC DSBs, because their translocation frequency increases in its absence (4). In addition to blocking lymphocyte development, C-NHEJ deficiency results in widespread apoptotic death of newly generated postmitotic neurons in mice (12–16). However, because neuronal death in the absence of C-NHEJ results from a p53 tumor-suppressor response to DSBs, such death can be obviated in vivo by p53 deficiency (14, 17). Likewise, p53 deficiency rescues the viability of cultured C-NHEJ-deficient NSPCs (4). In the absence of C-NHEJ (and p53), genomic DSBs can persist unjoined and can be substrates for more frequent translocation (3). In WT NSPCs, RDC-associated DSBs can be fused to other DSBs in the genome by C-NHEJ. In C-NHEJ-deficient NSPCs, such RDC

Significance

DNA double-strand breaks (DSBs) occur in all cells, including neural stem/progenitor cells (NSPCs) that give rise to the brain. We previously found that developing neural cells lacking a major DSB end-joining pathway are subject to widespread death. Because DSBs may result from gene transcription, we assayed for DSBs near active transcription start sites (TSSs) genome-wide in NSPCs. DSBs occur near TSSs of highly transcribed genes involved in general cellular processes but occur less often near neural-specific TSSs. These TSS-associated DSBs can translocate to other DSBs by both the normal and alternative DSB repair pathways. We report similar findings in B lymphocytes, suggesting that highly transcribed genes involved in general cellular processes are subject to TSS-associated DSBs in divergent cell types.

Author contributions: B.S., P.-C.W., and F.W.A. designed research; B.S., P.-C.W., A.N.C., J.K., Z.D., and R.M.M. performed research; B.S., P.-C.W., A.N.C., Z.D., R.M.M., and F.W.A. analyzed data; and B.S., P.-C.W., and F.W.A. wrote the paper.

Reviewers: D.F., University of Michigan; B.S., Weill Cornell Medical College; and K.Y., Michigan State University.

The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the Gene Expression Omnibus database (accession no. [GSE74356](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74356)).

¹B.S. and P.-C.W. contributed equally to this work.

²To whom correspondence may be addressed. Email: Bjoern.Schwer@childrens.harvard.edu or alt@enders.tch.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1525564113/-DCSupplemental.

translocations are mediated by an alternative end-joining (A-EJ) pathway that is biased toward the use of longer MHs (4).

C-NHEJ also requires the ataxia telangiectasia-mutated (ATM)-dependent DNA damage response to generate fully efficient DSB repair via end-joining (3, 18). ATM has been shown to have a direct role in DSB end-joining (3, 19, 20) and also is required for the p53-dependent G1 check point response to unrepaired DSBs (21, 22). Because of its dual role in C-NHEJ and DSB checkpoint responses, ATM deficiency can lead to long-term DSB persistence, leading, for example, to V(D)J recombination breaks in early B-cell development that persist developmentally into mature B cells (23–25). Although these ATM roles likely contribute to immunodeficiency and lymphoid cancers associated with ATM deficiency, whether such ATM functions are involved with the neurodegenerative phenotype associated with ATM deficiency in humans remains unknown (21, 22).

We previously developed an unbiased, high-throughput, genome-wide, translocation sequencing (HTGTS) approach to identify DSBs via their translocation to bait DSBs at defined genomic sites in primary progenitor and activated mature B cells, NSPCs, and human cell lines (4, 9, 26–28). Because of cellular heterogeneity in 3D genome organization (3, 27, 29), HTGTS detects, with great sensitivity, recurrent DSBs in any given genomic location (4, 9, 26–29) or recurrent classes of DSBs, such as DSBs at TSSs, that may occur at much lower frequency in any given location (ref. 9 and this study). Because of increased contact frequency, translocation frequency is increased further between sequences on the same *cis* chromosome (3, 28, 29). HTGTS detects only those DSBs that translocate. Given that most DSBs are rejoined or repaired locally to other DSBs within the same topological domain, HTGTS from DSBs on a different chromosome or even in a different topological domain on the same chromosome provides only a minimal estimate of the actual frequency of DSBs at a given site or region (3, 4).

In the current study, we used GRO-seq and HTGTS to identify TSS-associated DSBs in NSPCs, to quantify the frequency of such DSBs relative to that of RDC-associated DSBs, and to elucidate the class of genes in which most such DSBs occur and compare

them with the class of genes that incurs TSS-associated DSBs in B lymphocytes.

Results

DSBs and Translocations Are Enriched Around Active TSSs in NSPCs.

To assess whether transcription-associated DSBs occur and form translocations in NSPCs, we first identified actively transcribed genes by performing GRO-seq (8, 30, 31) in *Xrcc4*^{-/-}*p53*^{-/-} NSPCs. In these analyses, a Pearson correlation coefficient analysis showed high reproducibility among three independent experiments (Fig. S1A, Left). Aggregate patterns of transcription profiles showed enrichment of reads within 1.5 kb of TSSs genome-wide (Fig. S1B, Right), consistent with previously published GRO-seq results obtained in other cell types (30, 31). To perform HTGTS for DSB identification, we used a Cas9:sgRNA-based approach to introduce HTGTS bait DSBs into primary NSPCs isolated from *Xrcc4*^{-/-}*p53*^{-/-} mice, as previously described (4). More specifically, we used single-guide RNAs (sgRNAs) targeting Cas9-mediated DSBs to a nontranscribed intergenic region ~50 kb telomeric of *N-myc* on chromosome 12 (Chr12-sgRNA-1) (Fig. 1A) or to the *c-Myc* gene on chromosome 15 (Chr15-*Myc*-sgRNA) (Fig. S2). Primary NSPCs were collected 4.5 d after the introduction of Cas9:sgRNA expression constructs, and genomic DNA was isolated and processed for HTGTS. For both Cas9:sgRNA bait DSBs used, we used HTGTS primers that allowed genome-wide identification of endogenous “prey” DSBs that joined to centromeric broken ends of the bait DSB (Fig. 1A, Upper).

To examine potential relationships between transcriptionally active NSPC genes and DSBs, we calculated the distance between genome-wide HTGTS junctions generated from either Chr12-sgRNA-1 DSBs or Chr15-*Myc*-sgRNA DSBs and the nearest TSS of active or inactive genes (lower panels in Fig. 1A and Fig. S2), as described (9). For these analyses, we excluded junctions within 1 Mb on each side of the break site to eliminate junctions that occur via bait DSB rejoining following resection (9) and also eliminated Cas9:sgRNA-mediated off-target hotspot DSBs defined in our separate studies (Fig. 1A and Fig. S2) (4).

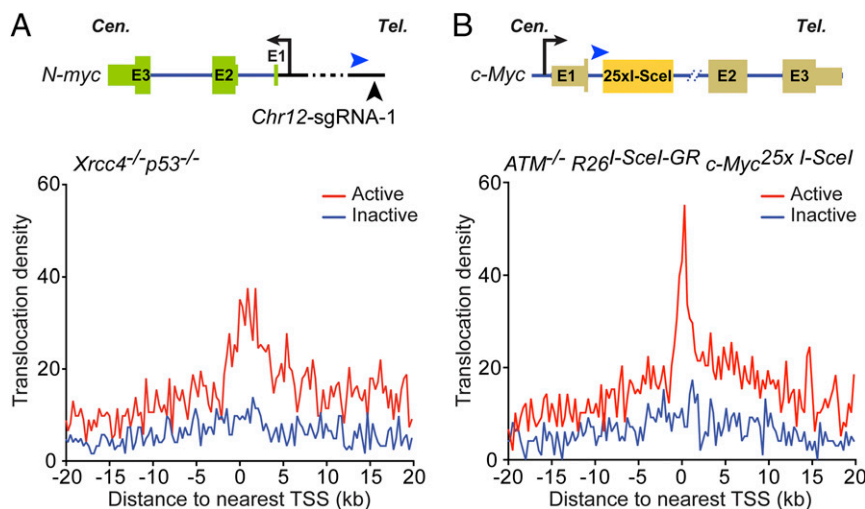


Fig. 1. DSBs and translocations are enriched around active TSSs in NSPCs. (A, Upper) Diagram of the *N-myc* locus and sgRNA target site (Chr12-sgRNA-1; vertical black arrowhead) and the location and orientation of HTGTS primer (horizontal blue arrowhead). Cen, centromere; E, exon; Tel, telomere. (Lower) Distribution of HTGTS junctions relative to TSSs genome-wide. Junctions from Chr12-sgRNA-1-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs ($n = 12,236$) are plotted in relation to the nearest RefGene (42) TSS (0 on the x axis). Negative numbers on the x axis indicate locations upstream of TSS; positive numbers on the x axis indicate locations downstream of TSS. The translocation density per 10,000 junctions (bin size 300 bp) relative to the TSS of active (red) or inactive (blue) genes, as determined by GRO-seq (4), is shown. (B, Upper) Diagram of the 25xl-SceI *c-Myc* locus; the HTGTS primer and orientation are indicated by the horizontal blue arrowhead. (Lower) Distribution of HTGTS junctions relative to TSSs genome-wide. Junctions from TA-treated *ATM*^{-/-} *R26*^{l-SceI-GR} *c-Myc*^{25xl-SceI} NSPCs ($n = 9,788$) are plotted in relation to the nearest RefGene (42) TSS; details are as in A.

Although Chr12-sgRNA-1 targets a nontranscribed intergenic region, and Chr15-*Myc*-sgRNA targets the transcribed *c-Myc* gene, we found that translocation junctions for both were significantly enriched in active genes, with a clear peak spanning the TSS (Fig. 1A and Fig. S2). We estimate that translocations between a DSB located within 2 kb of an active TSS and a bait Chr12-sgRNA-1 DSB occur in about 1 in 167 *XRCC4/p53*-deficient NSPCs (i.e., in about 0.4% of cells, based on the observed translocation rate of bait DSBs to DSBs located within 2 kb of an active TSS on the break-site chromosome, multiplied by the total number of chromosomes) (Table S1). We found similar levels of DSBs located within 2 kb of active TSSs for the Chr15-*Myc*-sgRNA bait DSBs (Table S1).

To test whether this DSB/active TSS correlation occurs for a different type of bait DSB end and in a different DNA repair-deficient background, we performed HTGTS and GRO-seq analyses in *ATM*-deficient (*ATM*^{-/-}) NSPCs, in which I-SceI-mediated bait DSBs can be induced in a 25×I-SceI knock-in cassette located in *c-Myc* (*c-Myc*^{25×I-SceI}) (Fig. 1B, Upper) (9). GRO-seq experiments in the *ATM*^{-/-} NSPCs again gave high reproducibility across three independent experiments, and the resulting aggregate patterns of transcription profiles indicated enrichment of reads within 1.5 kb of TSSs genome-wide (Fig. S1 A and B, Right). Similarly, as in the two different Cas9:sgRNA-based bait DSB analyses described above, I-SceI-based HTGTS

revealed a significant enrichment of translocation junctions around the TSSs of active genes (Fig. 1B, Lower). Based on the observed translocation rate of bait DSBs to DSBs located within 2 kb of active TSSs on the break-site chromosome, multiplied by the total number of chromosomes, we estimate that, genome-wide, TSS-associated DSBs that translocate to bait DSBs occur in about 1 in 73 (1.37%) *ATM*-deficient NSPCs, consistent with higher bait DSB translocation frequency in the *ATM*^{-/-} background (Table S1) (4).

To investigate further the set of transcribed genes that contained TSS-proximal junctions in NSPCs, we performed a comparative analysis of GRO-seq transcription profiles of NSPCs and those previously generated for activated B cells (31). This analysis yielded sets of genes transcribed in NSPCs but not in B cells (class 1), genes commonly transcribed in both NSPCs and B cells (class 2), and genes transcribed in B cells but not in NSPCs (class 3). Notably, examination of transcription rates among these classes revealed that the transcription rate of class 2 genes was significantly higher than that of either class 1 or class 3 genes, in all datasets examined (Fig. 2A and B and Fig. S3). We then evaluated how active genes with translocation junctions located within 2 kb of their TSS in NSPCs were distributed among the different classes of genes. Although ~2.6% of total genes in each class contained TSS-proximal junctions, we found that the majority of active genes with TSS-proximal junctions

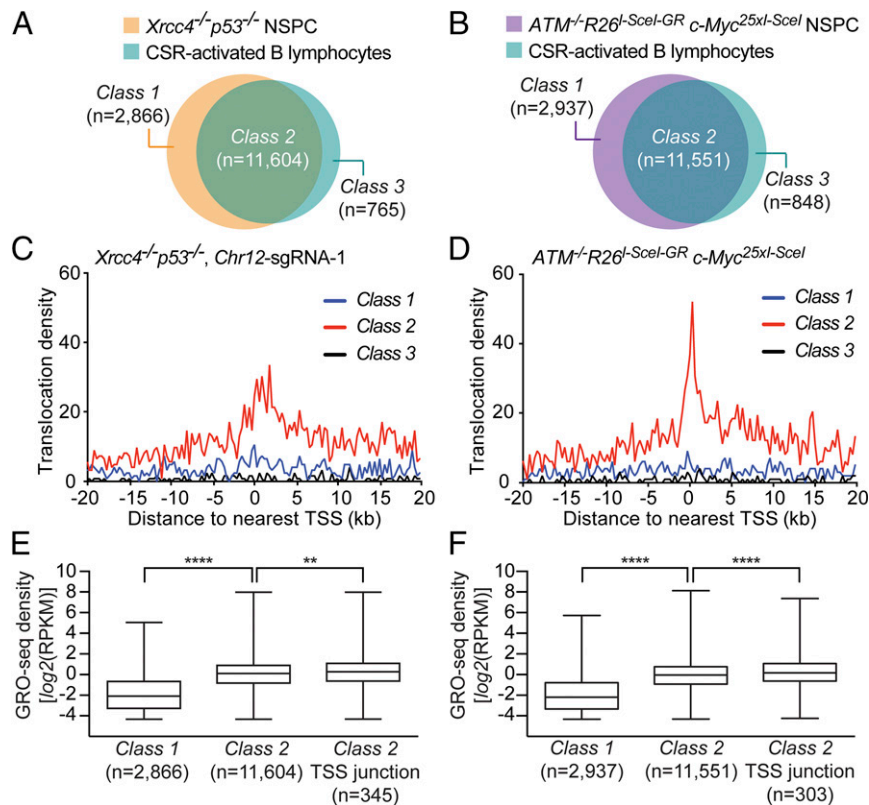


Fig. 2. TSS-proximal junctions occur in genes that are highly and commonly transcribed in NSPCs and B lymphocytes. (A) Venn diagram illustrating the three classes (1–3) of genes transcribed in WT CSR-activated B lymphocytes and *Xrcc4*^{-/-}*p53*^{-/-} NSPCs. (B) Venn diagram showing class 1–3 genes transcribed in WT CSR-activated B lymphocytes and *ATM*^{-/-}*R26*^{I-SceI-GR} *c-Myc*^{25xI-SceI} NSPCs. Class 1 genes are actively transcribed in NSPCs but are inactive in B cells. Class 2 genes are actively transcribed in both NSPCs and B cells. Class 3 genes are transcribed in B cells but not in NSPCs. Gene activity [reads per kilobase of transcript per million reads mapped (RPKM) ≥ 0.05] was determined by GRO-seq analysis. Numbers of genes per category are listed. (C and D) Plots showing the relative distance of junctions from Chr12-sgRNA-1-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs (C) or *ATM*^{-/-}*R26*^{I-SceI-GR} *c-Myc*^{25xI-SceI} NSPCs (D) to the closest TSS of class 1 (blue trace), class 2 (red trace), and class 3 (black trace) genes. Figures are organized as described in Fig. 1A. (E) Box-and-whiskers plots of the transcription activity (GRO-seq density) of class 1 (Left) and class 2 (Center) genes and genes in class 2 containing HTGTS junctions located within 2 kb of the TSS from Chr12-sgRNA-1-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs or *ATM*^{-/-}*R26*^{I-SceI-GR} *c-Myc*^{25xI-SceI} NSPCs (Right). (F) The number of genes in each group in E is indicated. The upper and lower box edges represent the 25th and 75th percentile, respectively; the horizontal line denotes the median. *****P* < 0.0001; ***P* < 0.01 (Mann-Whitney *u* test).

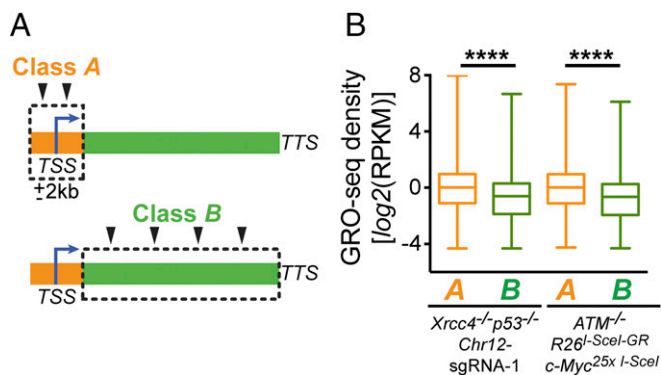


Fig. 3. Genes with TSS-proximal junctions show higher transcriptional activity than genes with junctions in the gene body located more than 2 kb from the TSS. (A) Active genes were determined by GRO-seq and were classified as class A (active genes with junctions located within 2 kb of the TSS) or class B (active genes with junctions within the gene body located more than 2 kb from the TSS). TTS, transcription termination site. (B) Box-and-whisker plots of transcriptional activity (GRO-seq density) of class A and class B genes in Chr12-sgRNA-1-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs from four independent experiments or in TA-treated *ATM*^{-/-} *R26*^{f-SceI-GR} *c-Myc*^{25x I-SceI} NSPCs from four independent experiments. Numbers of genes within each group: Chr12-sgRNA-1 *Xrcc4*^{-/-}*p53*^{-/-}: class A = 440 genes; class B = 2,377 genes; *ATM*^{-/-} *R26*^{f-SceI-GR} *c-Myc*^{25x I-SceI}: class A = 311 genes; class B = 1,807 genes. The upper and lower box edges represent the 25th and 75th percentile, respectively; the horizontal line denotes the median. *****P* < 0.0001 (Mann-Whitney *u* test).

(345/440; 78.4%) in experiments based on the Chr12-sgRNA-1 bait DSB represented class 2 genes (Fig. 2C). Similar results were obtained in experiments based on Chr15-*Myc*-sgRNA bait (Fig. S3A). Moreover, in experiments based on I-SceI-bait DSB 85.4% of active genes with TSS-proximal junctions (303/355) were class 2 genes (Fig. 2D). Notably, the class 2 genes that contained TSS-proximal junctions showed significantly higher average transcription levels than all class 2 genes on average, irrespective of genotype, in both Cas9:sgRNA- and I-SceI-based experiments (Fig. 2E and F and Fig. S3B) in NSPCs as well as in I-SceI-based experiments in CSR-activated B cells (Fig. S3C). These results suggest an association between transcription rate

and TSS-proximal DSBs. Consistent with their common expression in NSPCs and B cells, most class 2 genes with HTGTS junctions are involved in general biological processes (Table S2).

To evaluate further the potential relationship between transcription and the generation of DSBs that translocate to HTGTS bait DSBs, we compared the transcription rates of active genes with junctions located within 2 kb of an active TSS (class A) (Fig. 3A) with those of active genes with junctions only in the gene body located at distances greater than 2 kb from the TSS (class B) (Fig. 3A). Analysis of GRO-seq densities of class A and class B genes in Chr12-sgRNA-1- or Chr15-*Myc*-sgRNA-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs (Fig. 3B and Fig. S4) or triamcinolone acetone (TA)-treated *ATM*^{-/-} *R26*^{f-SceI-GR} *c-Myc*^{25x I-SceI} NSPCs (Fig. 3B) revealed a significantly higher transcriptional activity of class A genes (*P* < 0.0001; Mann-Whitney *u* test). Together, these findings are consistent with the notion that increased overall rates of transcription initiation may positively influence the frequency of DSB formation in the vicinity of TSSs.

Both C-NHEJ and A-EJ Processes Can Mediate TSS-Associated Translocations in NSPCs.

HTGTS also provides a powerful DSB end-joining assay. In this regard, junctional repair signatures in HTGTS libraries can be analyzed at the nucleotide level by calculating the difference between the end coordinate of the bait alignment and the start coordinate of the prey alignment. In this calculation, a value of 0 corresponds to a “direct” junction, whereas negative values represent short nucleotide homologies or MHs. To investigate the repair junction structures of translocations between bait DSBs and prey DSBs located within 2 kb of TSSs (either annotated TSSs or “active” TSSs identified by GRO-seq; see *Materials and Methods*), we examined HTGTS data from Chr12-sgRNA-1-based experiments or Chr15-*Myc*-sgRNA-based experiments in WT or *Xrcc4*^{-/-}*p53*^{-/-} NSPCs (Fig. 4). Four independent experiments in *Xrcc4*^{-/-}*p53*^{-/-} NSPCs expressing Chr12-sgRNA-1 yielded 509 junctions located within 2 kb of an annotated TSS (Fig. 4A). Seven independent experiments in WT cells expressing Chr12-sgRNA-1 yielded a total of 366 junctions located within 2 kb of an annotated TSS (Fig. 4A). Notably, ~35% of TSS-proximal translocation junctions in WT cells were direct, whereas only ~10% of such junctions were direct in *Xrcc4*^{-/-}*p53*^{-/-} NSPCs (Fig. 4A). Moreover, the TSS-proximal

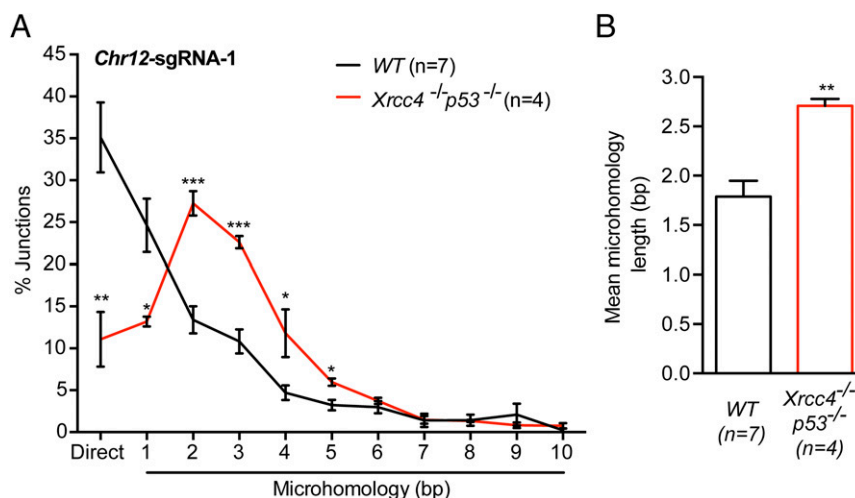


Fig. 4. Repair junction signatures of TSS-proximal HTGTS junctions. (A) TSS-proximal HTGTS junction profiles (located within 2 kb of all TSSs) from Chr12-sgRNA-1-expressing NSPCs of the indicated genotypes. Joins with 0 (direct or “blunt” joins) to 10 bp of junctional MH were identified and plotted as a percentage of total junctions; insertions were excluded from the dataset. The number of independent experiments is indicated. Total junction numbers analyzed: *n* = 366 WT and *n* = 509 *Xrcc4*^{-/-}*p53*^{-/-} NSPCs. (B) The mean MH length of TSS-proximal HTGTS junctions indicated in bar graphs was determined in the indicated number of independent experiments. Data represent mean ± SEM; *****P* < 0.001; ****P* < 0.01; **P* < 0.05 (unpaired two-tailed *t* test).

translocation junctions in Chr12-sgRNA-1-expressing *Xrcc4*^{-/-}*p53*^{-/-} NSPCs displayed significantly longer junctional homologies (mean MH length of 2.71 ± 0.07 bp in *Xrcc4*^{-/-}*p53*^{-/-} NSPCs versus 1.79 ± 0.16 bp in WT NSPCs) (Fig. 4B). Examination of junction structures in Chr15-*Myc*-sgRNA-based experiments in WT or *Xrcc4*^{-/-}*p53*^{-/-} NSPCs yielded similar results (Fig. S5). Together, these findings indicate that both the C-NHEJ and A-EJ pathways can mediate translocations arising from transcription-associated DSBs.

Discussion

This study, along with our additional recent study (4), identifies various classes of recurrent DSBs in NSPCs. Beyond the TSS-associated DSBs we define here, these other DSB classes in NSPCs include widespread, low-level DSBs that may be caused by cell-intrinsic processes such as replication or oxidative stress (1, 2) and RDCs in long neural genes that mostly are associated with mild replication stress in NSPCs (4). The estimated overall frequency of TSS-associated DSBs corresponds to about one in five NSPCs having a TSS-proximal DSB that translocates to a bait DSB (Table S1). This frequency is about 30-fold and 60-fold lower, respectively, than the frequency per NSPC of low-level widespread and RDC-associated DSBs that translocate to the same bait in the same *Xrcc4*^{-/-}*p53*^{-/-} background. Again, these numbers represent the minimal numbers of DSBs per NSPC in each class, because most DSBs are repaired locally, even in an XRCC4-deficient background (4). Based on these estimated numbers, the overall impact of TSS-associated DSBs on genomic stability in NSPCs is likely less than that of the other two classes. In this regard, consistent with these estimates, the great majority of DSBs in long transcribed neural RDC genes are in gene bodies and are not associated with TSSs, also indicating that these two classes of DSBs are independent (4). Finally, also consistent with this notion, TSS-associated DSBs tend to occur in a class of genes (class 2) involved in more general cellular processes that also are expressed and translocated in activated B cells, as opposed to cell-type-specific genes.

We find that DSBs are enriched in transcribed versus non-transcribed genes in NSPCs and are further enriched near the TSSs of generally expressed genes with transcription rates that are higher on average than those of most neural- or B-cell-specific genes. Transcription may contribute to this class of HTGTS-detectable DSBs mainly by promoting their occurrence (2, 5, 7, 9, 32, 33). Formation of transcription-associated DSBs may be related to torsional stress associated with transcription initiation, for example, that is caused by the binding of transcription factors to promoters (32). DSBs at gene promoters have been proposed to induce topological alterations and subsequent changes in chromatin organization that facilitate transcription initiation (32). In this regard, DSBs generated by DNA topoisomerase IIβ (TopoIIβ) during transcription initiation have been mechanistically implicated in the transcriptional activation of genes in various cell types, including postmitotic neurons (5, 7). Another proposed cause of TSS-associated DSBs is the formation of R-loops (2). We note, however, that our findings do not rule out a role for increased proximity of these generally expressed genes resulting from their localization to transcription factories (34). Enhanced interaction in transcription factories

also could facilitate the formation of specific translocations of genes that break recurrently because of transcription-associated DSBs in specific cell types (6, 35).

Our analysis of repair junction profiles of TSS-proximal translocations in WT and C-NHEJ-deficient NSPCs shows that both the C-NHEJ and A-EJ pathways can mediate such translocations, as we recently have shown for DSB translocation junctions in RDC genes and in junctions involving translocations of more widespread low-level DSBs (4). A-EJ has been speculated to be a major contributor to oncogenic translocations (36), possibly because of some specific aspect of the A-EJ pathway or simply reflecting increased DSB persistence (and availability for translocation) in the absence of C-NHEJ (37). Although our studies do not rule out a more dominant role for A-EJ in contributing to oncogenic or other translocation junctions, they clearly suggest that C-NHEJ also can participate in such events.

Materials and Methods

Primary NSPC Isolation, Culture, and DSB Induction. Primary NSPCs were prepared from postnatal day (P) 8–14 frontal brains and were cultured as described (38). All work involving mice was approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital (Protocol 14-10-2790R). Passage 0 dissociated cells at 0 d in vitro (DIV) were seeded into ultra-low-attachment six-well plates (Corning) at 4 × 10⁵ cells/mL. DIV 4.5 cultures were dissociated into single cells, and 5 × 10⁶ NSPCs were nucleofected with 5 μg of Cas9:sgRNA expression vector *pSpCas9(BB)* (Addgene plasmid 42230) (39) by using the Mouse Neural Stem Cell Nucleofector reagent (VPG-1004; Lonza). Cas9:sgRNA expression vectors were constructed by ligating annealed oligonucleotides into BbsI-digested *pSpCas9(BB)* (4). GR-I-SceI-mediated bait DSBs (9) were induced by the addition of 10 μM TA (Sigma) on DIV 5.5. Cells were collected and processed for GRO-seq or HTGTS analyses on DIV 9.

Global Run-On Sequencing. Three biological replicate GRO-seq libraries per genotype (*ATM*^{-/-}*R26*^{Cre}*Scel-GR**c-myc*^{25xI-SceI} or *Xrcc4*^{-/-}*p53*^{-/-}) were prepared and analyzed as previously described (4, 30, 31). In brief, Bowtie2 (40) was used to align GRO-seq data to the mouse genome (mm9/NCBI37), HOMER (41) was used for de novo identification of transcripts, and gene transcription rates were estimated as described (31).

HTGTS and Bioinformatic Analyses. HTGTS was performed and data were analyzed as described (4, 9, 27). In brief, libraries (500–1,000 bp fragment size) were purified and sequenced on the Illumina MiSeq platform. After demultiplexing of FASTQ output files, unique reads were aligned to the mouse genome (mm9/NCBI37) by Bowtie2 (40) and were processed further through a custom HTGTS pipeline (27). The distances of HTGTS junctions to the closest TSS were calculated by a custom script. Translocation rates and DSB frequency were estimated (4), and Cas9:sgRNA off-target sites were identified (4, 27) as described. Nucleotide-level analysis of HTGTS junctional repair signatures was performed by calculating the difference between the end coordinate of the bait alignment and the start coordinate of the prey alignment (4). In this calculation, a value of 0 corresponds to a direct junction, and negative values represent MHs.

ACKNOWLEDGMENTS. We thank members of the F.W.A. laboratory for stimulating discussions, Drs. Chunguang Guo, Monica Gostissa, and Jiazhi Hu for experimental advice; and Drs. Yi Zhang, Li Shen, and Fei-Long Meng for assistance with high-throughput sequencing. This work in the F.W.A. laboratory was supported by the Porter Anderson Fund from Boston Children's Hospital and the Howard Hughes Medical Institute. B.S. is a Martin D. Abeloff Scholar of The V Foundation for Cancer Research and is supported by National Institute on Aging/NIH Grant K01AG043630. P.-C.W. is supported by a National Cancer Center postdoctoral fellowship.

1. Aguilera A, Garcia-Muse T (2013) Causes of genome instability. *Annu Rev Genet* 47:1–32.
2. Kim N, Jinks-Robertson S (2012) Transcription as a source of genome instability. *Nat Rev Genet* 13(3):204–214.
3. Alt FW, Zhang Y, Meng FL, Guo C, Schwer B (2013) Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* 152(3):417–429.
4. Wei PC, et al. Long neural genes harbor recurrent DNA break clusters in neural stem/progenitor cells. *Cell*, 10.1016/j.cell.2015.12.039.
5. Ju BG, et al. (2006) A topoisomerase IIβ-mediated dsDNA break required for regulated transcription. *Science* 312(5781):1798–1802.

6. Lin C, et al. (2009) Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139(6):1069–1083.
7. Madabhushi R, et al. (2015) Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* 161(7):1592–1605.
8. Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322(5909):1845–1848.
9. Chiarle R, et al. (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147(1):107–119.
10. Lieber MR (2010) The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. *Annu Rev Biochem* 79:181–211.

11. Boboila C, Alt FW, Schwer B (2012) Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv Immunol* 116:1–49.
12. Barnes DE, Stamp G, Rosewell I, Denzel A, Lindahl T (1998) Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr Biol* 8(25):1395–1398.
13. Gao Y, et al. (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* 95(7):891–902.
14. Frank KM, et al. (2000) DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol Cell* 5(6):993–1002.
15. Gu Y, et al. (2000) Defective embryonic neurogenesis in Ku-deficient but not DNA-dependent protein kinase catalytic subunit-deficient mice. *Proc Natl Acad Sci USA* 97(6):2668–2673.
16. McKinnon PJ (2013) Maintaining genome stability in the nervous system. *Nat Neurosci* 16(11):1523–1529.
17. Gao Y, et al. (2000) Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* 404(6780):897–900.
18. Nussenzweig A, Nussenzweig MC (2010) Origin of chromosomal translocations in lymphoid cancer. *Cell* 141(1):27–38.
19. Zha S, et al. (2011) ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. *Nature* 469(7329):250–254.
20. Kumar V, Alt FW, Oksenyich V (2014) Functional overlaps between XLF and the ATM-dependent DNA double strand break response. *DNA Repair (Amst)* 16:11–22.
21. McKinnon PJ (2012) ATM and the molecular pathogenesis of ataxia telangiectasia. *Annu Rev Pathol* 7:303–321.
22. Shiloh Y, Ziv Y (2013) The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* 14(4):197–210.
23. Callén E, et al. (2007) ATM prevents the persistence and propagation of chromosome breaks in lymphocytes. *Cell* 130(1):63–75.
24. Hu J, Tepsuporn S, Meyers RM, Gostissa M, Alt FW (2014) Developmental propagation of V(D)J recombination-associated DNA breaks and translocations in mature B cells via dicentric chromosomes. *Proc Natl Acad Sci USA* 111(28):10269–10274.
25. Tepsuporn S, Hu J, Gostissa M, Alt FW (2014) Mechanisms that can promote peripheral B-cell lymphoma in ATM-deficient mice. *Cancer Immunol Res* 2(9):857–866.
26. Dong J, et al. (2015) Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. *Nature* 525(7567):134–139.
27. Frock RL, et al. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* 33(2):179–186.
28. Hu J, et al. (2015) Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell* 163(4):947–959.
29. Zhang Y, et al. (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148(5):908–921.
30. Min IM, et al. (2011) Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* 25(7):742–754.
31. Meng FL, et al. (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159(7):1538–1548.
32. Ju BG, Rosenfeld MG (2006) A breaking strategy for topoisomerase IIbeta/PARP-1-dependent regulated transcription. *Cell Cycle* 5(22):2557–2560.
33. Klein IA, et al. (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147(1):95–106.
34. Sutherland H, Bickmore WA (2009) Transcription factories: Gene expression in unions? *Nat Rev Genet* 10(7):457–466.
35. Mani RS, et al. (2009) Induced chromosomal proximity and gene fusions in prostate cancer. *Science* 326(5957):1230.
36. 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(4):410–416.
37. Zhang Y, et al. (2010) The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv Immunol* 106:93–133.
38. Brewer GJ, Torricelli JR (2007) Isolation and culture of adult neurons and neurospheres. *Nat Protoc* 2(6):1490–1498.
39. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.
40. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
41. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–589.
42. Pruitt KD, et al. (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* 42:D756–D763.
43. Mi H, Muruganujan A, Casagrande JT, Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 8(8):1551–1566.
44. Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 41(Database issue):D377–D386.