Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors

Highlights
- B lineage genes are γH2AX-marked DNA damage sites in transformed B cell precursors
- γH2AX hotspots occur at highly transcribed, H3K27ac-marked genes
- γH2AX hotspots relate to R loop sequences and convergent/divergent transcription
- Myeloid lineage induction in transformed B cells causes DNA damage at myeloid loci

Authors
Bryant Boulianne, Mark E. Robinson, Philippa C. May, ..., Jane F. Apperley, Justin Stebbing, Niklas Feldhahn

Correspondence
n.feldhahn@imperial.ac.uk

In Brief
B lineage-specific genes are frequently altered in B lymphoid leukemia. Boulianne et al. show that transformed B cell precursors exhibit transcription-coupled DNA damage at B lineage genes. Likewise, lymphoid-to-myeloid conversion of transformed B cells induces DNA damage at myeloid loci, suggesting that lineage-specific transcription predisposes lineage-specific genes to subsequent alterations as observed in human leukemia.

Accession Numbers
GSE80682
Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors

Bryant Boulianne,1,6 Mark E. Robinson,1,2,6 Philippa C. May,1,3,6 Leandro Castellano,2 Kevin Blighe,1,2 Jennifer Thomas,1,2 Alistair Reid,1,3 Markus Muschen,* Jane F. Apperley,1 Justin Stebbing,2 and Niklas Feldhahn1,6*

1Centre for Haematology, Department of Medicine, Imperial College London, W12 0NN London, UK
2Division of Cancer, Department of Surgery and Cancer, Imperial College London, W12 0NN London, UK
3Molecular Pathology, Imperial College Healthcare NHS Trust, W12 0NN London, UK
4Department of Systems Biology, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Pasadena, CA 91016, USA
5Co-first author
6Lead Contact
*Correspondence: n.feldhahn@imperial.ac.uk
http://dx.doi.org/10.1016/j.celrep.2017.01.057

SUMMARY

In human leukemia, lineage-specific genes represent predominant targets of deletion, with lymphoid-specific genes frequently affected in lymphoid leukemia and myeloid-specific genes in myeloid leukemia. To investigate the basis of lineage-specific alterations, we analyzed global DNA damage in primary B cell precursors expressing leukemia-inducing oncogenes by ChiP-seq. We identified more than 1,000 sensitive regions, of which B lineage-specific genes constitute the most prominent targets. Identified hotspots at B lineage genes relate to DNA-DSBs, affect genes that harbor genomic lesions in human leukemia, and associate with ectopic deletion in successfully transformed cells. Furthermore, we show that most identified regions overlap with gene bodies of highly expressed genes and that induction of a myeloid lineage phenotype in transformed B cell precursors promotes de novo DNA damage at myeloid loci. Hence, we demonstrate that lineage-specific transcription predisposes lineage-specific genes in transformed B cell precursors to DNA damage, which is likely to promote the frequent alteration of lineage-specific genes in human leukemia.

INTRODUCTION

Most cancers are genetically unstable (Lengauer et al., 1998) and exhibit genomic alterations that promote tumor progression (Vogelstein et al., 2013). Analyses of cancer genomes over the last decade have revealed highly mutated genomes with more than 100 alterations per tumor genome for solid tumors (Vogelstein et al., 2013), whereas liquid tumors such as leukemia are less mutated (Mullighan et al., 2007). The analyses of leukemia genomes, however, uncovered that the most prominent group of genes to become defective in B cell precursor leukemia (B-cell precursor acute lymphoblastic leukemia [B-ALL]) or B-lymphoid acute lymphoblastic leukemia (B-ALL) are regulators of B cell development (40% in B-ALL and 66%–68% in high-risk B-ALL (Mullighan et al., 2007, 2009; Zhang et al., 2011)), with the transcription factors PAX5 and IKZF1 being altered in ~32% and ~29% of cases in high-risk B-ALL, respectively (Mullighan et al., 2009). Deletion of IKZF1 is associated with particularly poor clinical outcomes (Mullighan et al., 2009), and recent studies based on a PAX5 restoration model clearly validated the powerful tumor-suppressive function of PAX5 in B lineage ALL (Liu et al., 2014). Similar to lymphoid malignancies, myeloid-lineage genes are often targets of genomic alteration in myeloid leukemia (e.g., CEBPA, GATA1, PU.1, and RUNX1; Rosenbauer and Tenen, 2007). Although this pattern of lineage-specific genetic lesions in lymphoid and myeloid leukemia cases has been frequently observed, the mechanistic basis underlying its appearance remained elusive.

DNA damage resulting from oncogenic stress was proposed as a cause of genome instability in transformed cells almost 10 years ago (Halazonetis et al., 2008). Malignant transformation of healthy cells resulting in cancer is frequently associated with the activation of an oncogene (e.g., BCR-ABL1 in B-ALL), and, similarly, oncogene activation can occur during tumor progression. Oncogene activation may cause cellular stress, including DNA damage, as a result of oxidative stress (Sattler et al., 2000; Vafa et al., 2002) and/or replicative stress (Di Micco et al., 2006). The latter has been associated with hyper-replication, aberrant replication origin firing, and an exhaustion of the cellular nucleotide pool that interferes with efficient replication (Bester et al., 2011). DNA damage in oncogene-expressing cells has been shown to promote malignant transformation and acquisition of genomic defects in cancer models (Gilad et al., 2010). Likewise, oncogene-induced DNA damage has been associated with genomic defects found in human cancer, including solid tumors (Miron et al., 2015) and B cell lymphoma (Barlow et al., 2013). However, the potential relevance of
oncogenic stress in promoting genomic defects found in patients with leukemia remained unclear and has been further brought into question by the identification of the antibody-diversifying enzymes AID and RAG1 as potential drivers of genomic defects in ETV6-RUNX1-positive B cell precursor ALL (Swami-nathan et al., 2015). Likewise, recurrent genomic deletions in B-ALL have been associated with aberrant RAG1 activity by others because of the presence of cryptic RAG1 cleavage sites at individual deletion breakpoints (Holmfeldt et al., 2013; Iaco-bucci et al., 2009; Mullighan et al., 2008).

To obtain insight into the origin of DNA damage that promotes genomic defects found in leukemia genomes, here we performed a genome-wide analysis of DNA damage in primary B cell precursors undergoing oncogene-induced leukemic transformation using chromatin immunoprecipitation sequencing (ChIP-seq). The cells analyzed by us do not express AID or RAG1; hence, identified DNA damage is thought to relate primarily to oncogenic stress present in these transformed cells.

RESULTS

Transformed B Cell Precursors Exhibit Genome-wide DNA Damage

B lymphoid leukemia arises from hematopoietic stem cells or B cell precursors that acquired a transforming, leukemia-initiating event. To analyze DNA damage related to leukemic transformation, we transduced primary murine bone marrow-derived B cell precursors with retroviral vectors encoding the leukemia-inducing oncogenes BCR-ABL1 or C-MYC or with an empty vector (EV) as a control (Figure 1A; Figure S1A). To increase the sensitivity of DNA damage detection, we used bone marrow from S3BP1−/− mice. As expected, oncogene expression resulted in an immediate response characterized by elevated proliferation and a temporary phase of cellular stress (Figure 1B).

Early-transformed cells further exhibited signs of DNA damage (Figure 1C; Figure S1B), which was unrelated to apoptosis (Figure 1D), or expression of AID or RAG1 (Figure S1C).

We next aimed to identify the precise genomic regions that experience DNA damage upon leukemic transformation. To do so, we performed ChIP-seq on apoptotic/dead cell-depleted fractions using antibodies for the DNA-bound DNA damage response protein γH2AX, as done previously by others (Barlow et al., 2013; Rodriguez et al., 2012; Seo et al., 2012). γH2AX regions were identified as described previously (Barlow et al., 2013), with minor modifications to optimize calling of broad γH2AX peaks.

γH2AX signals were highly reproducible in a repeat experiment (Figures S1D and S1E). In total, 889 and 960 γH2AX-regions were identified for BCR-ABL1 and C-MYC, respectively (Tables S1 and S2). Interestingly, γH2AX signal intensities at γH2AX-enriched regions were comparable for both BCR-ABL1- and C-MYC-expressing cells (Figure 1E). This may reflect that BCR-ABL1- and C-MYC-induced DNA damage arises in a similar manner (e.g., by replicative stress) or that BCR-ABL1-induced DNA damage mainly arises from activation of BCR-ABL1-induced expression of C-MYC (Sawyers, 1993). We therefore combined both region lists, yielding a merged list of 1,289 γH2AX sites that were used for further comparison (Table S3).

γH2AX sites were distributed over all chromosomes (Figure 1F), but the total coverage across chromosomes highly correlated with exon coverage per chromosome (Figure 1G).

γH2AX-Marked Genomic Regions Occur at B Cell Precursor-Specific Genes and Relate to Endogenous DNA Damage

We next analyzed the identified γH2AX regions for genes that overlap with them. We observed that many of the most prominent regions affected genes specifically expressed by B cell precursors (Figure 2A). Furthermore, most of the crucial genes related to the pre-B cell receptor complex, pre-B cell receptor signaling, and B cell precursor identity-related transcription overlapped with γH2AX-enriched regions present in transformed B cell precursors (Figure 2B; Figures S2A–S2C). Gene ontology (GO) pathway analysis of all γH2AX regions confirmed that they, in many cases, affected immune system-related processes (Figure S2D; Table S4). To verify that identified γH2AX hotspots at B lineage genes correlate to sites of active DNA damage, we next performed DNA fluorescence in situ hybridization (FISH) using “break-apart” probes for two B lineage gene-related γH2AX hotspots (BLNK and PAX5) and a γH2AX coldspot (FOXP2) as a control (Figure 2C). The FISH probes cover the flanking regions of the target genes and are physically separated when a DNA double-strand break (DSB) occurs (split signal) (Figure 2C, left).

In agreement with our ChIP-seq results, we could detect a significant increase in split FISH probe signals for the γH2AX hotspots in transformed B cell precursors compared with control cells (Figure 2C, right) but not for the γH2AX-negative region analyzed as a control. Interestingly, when analyzing a successfully transformed B cell precursor line (~50 days after oncogene induction) by γH2AX ChIP-seq, we observed that the most prominent γH2AX region at IGH was diminished (Figure 2D, top). To investigate whether this may relate to internal deletions at the IGH locus, we analyzed seven successfully transformed B cell precursor cell lines for genomic loss at IGH by PCR. We confirmed genomic loss in two of seven cultures, including the cell line initially analyzed by ChIP-seq (Figure 2D, bottom). In summary, our results show that B cell precursor-specific genes represent prominent γH2AX hotspots in transformed B cell precursors, that γH2AX enrichment at B lineage genes correlates to endogenous DNA damage as validated in a limited number of loci by DNA-FISH, and that γH2AX-enriched regions can be subject to genomic deletion during prolonged culture of transformed B cell precursors.

Identified Regions Locate to Genes Affected by Genomic Lesions in Human B Cell Precursor Leukemia

Within the most prominent γH2AX regions, we identified B cell precursor-specific genes but also non-B cell-specific genes known to become defective in leukemia (e.g., HIST1 and ERG) (Figure 3A). To evaluate to what extent genomic alterations in human B-ALL occur at genes that are characterized by γH2AX hotspots in transformed B cell precursors, we compared the list of genes overlapping with γH2AX regions with three datasets documenting gene defects in human leukemia. Only genomic alterations from these datasets that were present in B cell precursor leukemia and that related to genomic deletions or translocations were used for comparison. We found significant enrichment of...
genes associated with genomic deletions/translocations that similarly overlapped with γH2AX regions identified by us (Figure 3B; Table S5). γH2AX regions also located to genomic loci associated with alterations in B-ALL that were not identified by the above studies (e.g., BACH2 and MIR142). In total, 27 genes reported as defective in human B-ALL were present in our γH2AX region list (Figure S3A) and an additional eight genes related to B-ALL that displayed γH2AX signals but failed to pass our stringent region-calling threshold (Figure S3B).

**γH2AX Hotspots in Transformed B Cell Precursors Occur at Transcriptionally Active Loci**

We next analyzed the genomic location of the identified regions in more detail. 788 of 1,289 γH2AX regions identified for BCR-ABL1-expressing cells (blue), C-MYC-expressing cells (red), or for both ChIP-seq libraries (common regions = black) (for repeat experiments, see Figures S1D and S1E). The shaded area indicates regions with γH2AX enrichment above background under both conditions (percent of total regions indicated). Linear regression (red line) and Pearson’s correlation (p < 2.2 × 10^-16) were calculated from these common regions.

**Chromosomal distribution of γH2AX regions (blue bars) present in transformed B cell precursors.** Bar height represents γH2AX signals (RPM of the BCR-ABL1 library) for each region (range = 0–100 RPM).

**Comparison of γH2AX region versus exon coverage per chromosome of combined γH2AX regions for BCR-ABL1- and C-MYC-expressing cells.** The red line represents linear regression fit.

Figure 1. Transformed B Cell Precursors Exhibit Genome-wide DNA Damage

(A) Schematic of the experimental strategy.

(B) Diagram visualizing proliferation (left) and viability (right) of B cell precursor cultures transduced with the oncogenes BCR-ABL1 or C-MYC or empty vector (EV) as a control. Shown is one of more than five repeat experiments.

(C) Western blots for the DNA damage-associated phosphorylations γH2AX and p-P53S15 in transformed versus untransformed B cell precursors. Shown is one of more than five repeat experiments.

(D) Left: verification of the dead cell removal procedure by flow cytometry (7-AAD = dead cells, AnnexinV = apoptotic). Right: western blot as in (C) for dead/apoptotic cell-depleted fractions.

(E) Scatterplot of normalized γH2AX read counts per million reads (nRPM, normalized to input/isotype control library) within γH2AX regions identified for BCR-ABL1-expressing cells (blue), C-MYC-expressing cells (red), or for both ChIP-seq libraries (common regions = black) (for repeat experiments, see Figures S1D and S1E). The shaded area indicates regions with γH2AX enrichment above background under both conditions (percent of total regions indicated). Linear regression (red line) and Pearson’s correlation (p < 2.2 × 10^-16) were calculated from these common regions.

(F) Chromosomal distribution of γH2AX regions (blue bars) present in transformed B cell precursors. Bar height represents γH2AX signals (RPM of the BCR-ABL1 library) for each region (range = 0–100 RPM).

(G) Comparison of γH2AX region versus exon coverage per chromosome of combined γH2AX regions for BCR-ABL1- and C-MYC-expressing cells. The red line represents linear regression fit.
Figure 2. γH2AX-Marked Genomic Regions Occur at B Cell Precursor-Specific Genes and Relate to Endogenous DNA Damage

(A) Bar diagram of the top-ranked γH2AX regions identified for BCR-ABL1-transformed B cell precursors. Ranking is based on fold change (FC) of the γH2AX signal compared with control (Iso) signals. Genes that overlap with γH2AX regions are indicated. For multiple regions covering the same gene, regions are combined, and average FC values are shown (i.e., for HIST1 and HIST2 genes and for LY6E).

(B) Top: representative custom track images for genes with γH2AX regions related to the pre-B cell receptor, pre-BCR signaling, and B cell precursor identity. Grey boxes indicate gene body location for the indicated genes. Bottom: tables of γH2AX signal FC values and read density per region expressed as reads per kilobase per million (RPKM) of BCR-ABL1 (B/A) and C-MYC ChIP-seq libraries for genes related to respective gene categories. Values shown in gray did not pass our stringent empirical selection criteria using Statistical Model for Identification of ChIP-Enriched Regions (SICER). Related custom track images can be found in Figures S2A–S2C.

(C) Verification of DNA damage at γH2AX hotspots at B lineage genes using DNA FISH and break-apart probes. Left: schematic overview of signals detected. Right: results of five independent experiments for PAX5 and BLNK (γH2AX hotspots) versus FOXP2 (γH2AX coldspot) (data represent mean ± SEM).

(D) Top: custom track visualization of the γH2AX ChIP-seq signal at IGH in BCR-ABL1-transformed B cell precursors on day 7 p.i. versus in a BCR-ABL1-induced B cell precursor cell line (>50 days p.i.). Bottom: analysis of the IGH locus by PCR using genomic DNA as a template. JH4 proximal intronic and Eμ regions of IGH were assessed; CD74 was used as a non-IGH control region (one of three repeats is shown). BCR-ABL1-induced cell line #2 is the one analyzed by γH2AX ChIP-seq above.
cross-analysis of structural variations (SVs) previously identified in human leukemia also showed breakpoints to be enriched within gene bodies (Figure S4A). To assess whether γH2AX region formation relates to transcription, we next generated ChiP-seq libraries for the chromatin mark H3K27ac, which is present at transcriptionally active promoters and enhancers (Shlyueva et al., 2014). Analysis of H3K27ac distribution confirmed accumulation at transcription start sites (TSSs) in transformed B cell precursors (Figure S4B). Furthermore, most γH2AX regions exhibited H3K27ac at respective promoters, and abundance of these two histone marks at overlapping regions showed significant correlation (Figures 4B and 4C). Analysis by RNA sequencing (RNA-seq) confirmed γH2AX regions to be enriched among highly expressed genes (Figures 4D and 4E) and that highly expressed genes exhibit greater γH2AX intensities (Figure 4F). Furthermore, genes differentially expressed between BCR-ABL1 and C-MYC conditions displayed a concomitant change in γH2AX signal intensity (Figure 4G). In agreement with increased γH2AX signals at highly transcribed loci in transformed B cell precursors, we also observed that induced expression of a transgenic fluorochrome (DS-RED) in transformed B cell precursors resulted in elevated γH2AX accumulation at the related expression cassette (Figure S4C).

Analysis of Convergent and Divergent Transcription, R Loop-Forming Sequences, and Early-Replicating Fragile Sites at γH2AX Regions

To understand the origins of transcription-induced γH2AX regions in transformed B cell precursors, we analyzed whether γH2AX regions are associated with known causes of transcription-related fragility. We first analyzed γH2AX regions for their relation to convergent and divergent transcription, which have been associated with genomic instability (Barlow et al., 2013; Heinäniemi et al., 2016; Pannunzio and Lieber, 2016). γH2AX regions were enriched at gene pairs compared with isolated genes but not further enriched for convergent or divergent gene orientation compared with tandem orientation (Figure 5A, left). However, in agreement with increased fragility at regions with convergent and divergent transcription, we observed that γH2AX signals were elevated at convergent/divergent gene pairs compared with tandem genes (Figures 5A, right; p < 1 × 10^{-6} for both, convergent versus tandem and divergent versus tandem comparisons; Mann-Whitney test). We next assessed whether γH2AX regions may relate to R loop formation, which can promote fragility by exposure of single-stranded DNA (ssDNA) or interference with replication. To do so, we predicted R loop-forming sequences (RLFSs) for the mouse genome as described previously (Jenjaroenpun et al., 2015) and analyzed γH2AX regions for enrichment of RLFSs. Of note, RLFSs do not indicate the actual presence of R-loops but an increased propensity of a nucleotide sequence to promote R-loop formation. Likewise, detection of RNA-DNA hybrids by DNA:RNA immunoprecipitation sequencing (DRIP-seq) analysis has shown significant enrichment of R-loops at computationally predicted RLFSs (Heinäniemi et al., 2016). We found that genes marked by γH2AX were enriched for RLFSs compared with the gene population as a whole (Figure 5B, left) or when considering only highly expressed genes (Figure 5B, right), suggesting an increased likelihood of R loop formation at γH2AX regions. We next compared γH2AX regions to locations of early replication fragile sites (ERFSs) (Barlow et al., 2013), which accumulate DNA DSBs and γH2AX in response to hydroxyurea (HU)-induced replicative stress. ERFS hotspots significantly overlapped with the γH2AX regions identified by us (55% of ERFSs overlapped with 26% of γH2AX regions) (Figure 5C, left); however, HU-induced γH2AX signals at ERFSs differed from the signals present in transformed B cell precursors at overlapping regions (Figure 5C, right).
Figure 4. γH2AX Hotspots in Transformed B Cell Precursors Relate to Transcriptionally Active Genes

(A) Overlap of γH2AX regions with the indicated gene features for all γH2AX regions identified in transformed B cell precursors. Inset: distribution of γH2AX reads within gene bodies versus upstream and downstream regions for transformed B cell precursors by meta-gene analysis. Signals are normalized to Iso.

(B) A representative custom track of γH2AX versus H3K27ac read densities.

(C) Scatterplots of γH2AX versus H3K27ac intensities for γH2AX regions. Signals are background-subtracted (“normalized”) RPM values for γH2AX and H3K27ac within identified γH2AX regions from the merged γH2AX region list. For better identification of overlapping regions, regions were extended ± 10 kb as described in the Supplemental Experimental Procedures. Linear regression (red line) and Pearson’s correlation (p < 2.2 × 10^−16 for both) between γH2AX versus H3K27ac signal intensities is shown for regions that contain both γH2AX and H3K27ac signals (gray shaded area, percentage of total indicated). See Figure S4B for meta-gene analysis of H3K27ac.

(D) Relative gene expression levels ranked by fragments per kilobase of transcript per million reads (FPKM) for BCR-ABL1- and C-MYC-transformed B cell precursors. Genes overlapping γH2AX-enriched regions are highlighted as indicated.

(E and F) Genes were categorized as high- or low-expressed relative to the median expression level across the population of all expressed genes. Inset, left: kernel density plot of FPKM values; colors indicate expression categories. Relative distribution of γH2AX-enriched genes between expression categories (E) and distribution of normalized γH2AX values measured for extended gene regions between expression categories (F) are shown. The shaded area indicates kernel density overlaid with a standard boxplot, and stars indicate the highly significant difference between categories (p < 1 × 10^−15; Mann-Whitney U test).

(G) Scatterplot of the difference in γH2AX signal (nRPM) across extended gene regions between BCR-ABL1- and C-MYC-transformed B cell precursors. Differentially expressed genes are highlighted in blue or red as indicated. Linear regression (red line) and Pearson’s rho (r) are indicated.
Lineage-Specific Transcription Predisposes Lineage-Specific Genes to DNA Damage

To functionally investigate the potential link between lineage-specific transcription and DNA damage at lineage-specific genes, we next investigated whether a fundamental change in the cell-type-specific transcriptional program of transformed B cell precursors may cause a concomitant change in DNA damage hotspot localizations at lineage-specific genes. To do so, we used a similar model as before, in which B cell precursors are transformed by the BCR-ABL1 oncogene but with the addition of an inducible CEBPA transgene, which promotes myeloid-specific transcription and a corresponding myeloid lineage phenotype in B cell precursors (Chen et al., 2015; Figure 6A). This allowed us to examine de novo DNA damage-sensitive sites in transformed B cell precursors by γH2AX ChIP-seq that occur as a consequence of myeloid lineage-specific transcription. Indeed, induction of a myeloid phenotype (CEBPA + doxycycline [DOX]) was associated with de novo γH2AX signals at myeloid-specific genes that were not present in our lymphoid lineage controls (Figure 6B, bottom; Table S6). γH2AX signals at non-lineage-specific genes showed no difference between lymphoid and myeloid phenotypes (Figure 6B, top). γH2AX signals at B cell-specific genes only showed a minor reduction in CEBPA-expressing cells compared with controls, possibly reflecting an incomplete loss of the B cell phenotype in our experiments. To confirm DNA damage at myeloid γH2AX hotspots, we performed DNA FISH as described in Figure 2C but using probes for two myeloid hotspots (LYZ2 and THBS1). In agreement with our ChIP-seq results, we detected increased split probe signals for LYZ2 and THBS1 by DNA FISH in the myeloid compared with

|               | γH2AX regions: | Randomly expected: |
|---------------|----------------|-------------------|
| Tandem        | 24%            | 16%               |
| Convergent    | 16%            | 10%               |
| Divergent     | 11%            | 8%                |
| Isolated      | 30%            | 54%               |
| Overlapping   | 19%            | 12%               |
Figure 6. Induction of a Myeloid Lineage Phenotype in Transformed B Cell Precursors Causes DNA Damage at Myeloid-Specific Genes

(A) Left: Schematic of the experimental strategy for induction of a myeloid phenotype in transformed B cell precursors using DOX-inducible CEBPA expression. Right: DOX/CEBPA-mediated induction of the myeloid phenotype verified by flow cytometry using CD11B as a myeloid lineage and CD19 as a B lymphoid lineage marker. EV-transduced lymphoid leukemia cells were used as a control. Anti-CD11B magnetic bead based cell sorting (MACS) was performed on DOX-treated cells with inducible CEBPA to reach a purity of >90% CD11B-positive cells for ChIP-seq analysis.

(B) Custom track visualization of γH2AX read densities for non-lineage-specific (HIST1 cluster, top) and myeloid lineage-specific gene loci (LYZ2, bottom). Grey boxes indicate gene-body location of the indicated gene.

(C) Assessment of myeloid γH2AX ChIP-seq hotspots identified in CEBPA + DOX cells using DNA FISH and break-apart probes as described in Figure 2C. Results of five independent experiments for LYZ2 and THBS1 (myeloid γH2AX hotspots) versus FOXP2 (coldspot) are shown (data represent mean ± SEM).

(D) Table of the ten top-ranked genes with significant γH2AX signals specific for CEBPA + DOX cells (by fold change). Full references can be found in the Supplemental Information.

Analysis of the top-ranked γH2AX ChIP-seq regions specific for cells with the myeloid phenotype further showed that all of them related to myeloid lineage or function (Figure 6D). GO pathway analysis additionally confirmed high enrichment of myeloid-specific genes overlapping with γH2AX regions identified in cells with the myeloid phenotype (Table S7). These findings provide a functional validation of our hypothesis that lineage-specific transcriptional programs promote fragility of lineage-specific genes in transformed B cell precursors.
DISCUSSION

Oncogenic lesions in hematopoietic progenitor cells give rise to B lymphoid and myeloid malignancies that, at the time of diagnosis, often carry secondary genomic defects that contribute to the disease. The most prominent defects in lymphoid leukemia genomes affect lineage-specific genes (Mullighan et al., 2007, 2009; Zhang et al., 2011) that guide respective hematopoietic differentiation and function. Defects at lineage-specific genes are likely to contribute to the differentiation arrest present in leukemia cells, which is thought to represent a causative event for malignant transformation (Rosenbauer and Tenen, 2007). It is unclear, however, whether defects at lineage-specific genes arise from random mutation and selection or whether lineage-specific genes are more likely to become defective. Although our experiments focus on B cell precursors, our results argue for the latter. During malignant transformation of B cell precursors by the leukemia-inducing oncogenes C-MYC and BCR-ABL1, B lineage-specific genes become prominent γH2AX hotspots that relate to DNA DSBs, as shown by DNA FISH for PAX5 and BLNK. γH2AX accumulation is associated with increased transcription, and induction of myeloid-specific transcription in transformed B cell precursors by CEBPA-mediated reprogramming further causes de novo γH2AX hotspots at myeloid-specific genes that relate to DNA DSBs, as shown by DNA FISH for LY22 and THBS1. As we assessed only a relatively small number of γH2AX hotspots by DNA FISH, further genome-wide investigation using other tools than ChIP-seq, such as translocation capture sequencing (TC-seq) (Klein et al., 2011), DSB-Capture (Lensink et al., 2016), or DSB end sequencing (END-seq) (Canela et al., 2016) will be beneficial to corroborate DNA DSBs at all identified γH2AX hotspots. Such analyses would also allow a more detailed characterization of the precise breakage sites within identified γH2AX hotspots because γH2AX spreading can occur over extended distances around a DNA DSB (e.g., as shown by Bothmer et al., 2011). Nevertheless, our results show that lineage-specific transcription and the associated fragility of highly transcribed loci represents a major cause of instability at lineage genes in transformed B cell precursors. Such predisposition to DNA damage is likely to increase the occurrence of genomic alterations at B lineage-specific genes and to facilitate subsequent positive selection during clonal evolution as observed in human B-ALL. Notably, with the two experimental strategies we performed (oncogene induction in primary B cell precursors and CEBPA-induced reprogramming in BCR-ABL1-transformed B cell precursors), we analyzed DNA damage at an early stage of malignant transformation as well as at a late stage (primary B cell precursors were analyzed 7 days after oncogene induction [p.i.], whereas CEBPA-reprogrammed cells were analyzed >30 days p.i.). Because we observed fragility of lineage-specific genes in both models, our experiments indicate that this instability phenotype is not limited to the acute response to oncogene activation but represents a characteristic of both early- and long-term-transformed cells.

Although our results provide a link between the high frequency of genomic defects of lineage genes in hematopoietic cancer and transcription-related DNA damage observed in other experimental contexts, transcription can cause genomic fragility for many reasons. For example, transcription exposes ssDNA, which is more fragile and vulnerable to DNA damage. ssDNA can be formed at sites of premature transcription termination (Wang et al., 2014) or within R-loops that occur naturally during transcription (Santos-Pereira and Aguilar, 2015; Skourt-Stathaki and Proudfoot, 2014). R-loops have been shown to specifically associate with highly expressed genes in yeast (Chan et al., 2014), and absence of the DNA damage response protein BRCA1 has been shown to promote R loop-dependent DNA damage at transcriptional pause sites (Hatchi et al., 2015). R loop-associated DNA damage further causes γH2AX accumulation (Hatchi et al., 2015), and RLFSs have been recently described at structural variation hotspots (Heinäniemi et al., 2016). Similarly, we observed that transcription-associated γH2AX hotspots in transformed B cell precursors are strongly enriched for RLFSs. Although RLFSs do not indicate the actual presence of R-loops but an increased likelihood of R loop formation, RLFSs do have significant predictive power for R-loops (Heinäniemi et al., 2016; Jenjaroenpun et al., 2015). However, RLFSs are abundantly encoded throughout the genome within and outside of γH2AX regions (Figure S4D), but cell-type-specific transcription programs are likely to dictate which of these loci actively form R-loops and, consequently, which genomic regions exhibit fragility. In agreement with this, we showed that γH2AX hotspot formation and related DNA DSB induction can be altered by changing lineage-specific transcription.

Additional factors may further amplify transcription-induced fragility during leukemic transformation, such as divergent and/or convergent transcription, which, similarly, can lead to exposure of ssDNA and which have been associated with an increased risk of genomic instability (Barlow et al., 2013; Heinäniemi et al., 2016; Pannunzio and Lieber, 2016). Indeed, we observed elevated γH2AX signals at convergently and divergently transcribed genes. Alternatively, transcription-related DNA damage might also occur as a result of Topoisomerase 2 (TOP2)-mediated DNA DSB induction, which has been described to occur during rapid transcriptional activation of induced genes (Bunch et al., 2015; Madabhushi et al., 2015). Notably, γH2AX signals reported for TOP2-induced DNA damage exhibited a gene body-restricted pattern similar to the pattern present in transformed B cell precursors (Bunch et al., 2015; Madabhushi et al., 2015). However, we were unable to assess whether TOP2-related DNA damage may additionally promote γH2AX hotspot formation and consecutive DNA damage in our experimental model because of immediate induction of apoptosis in response to TOP2 inhibition (data not shown).

Besides transcription itself, interference between transcription and replication has been repeatedly reported as a potential reason for DNA damage (e.g., by Barlow et al., 2013, and Jones et al., 2013). For example fragile sites, which are especially prone to DNA damage during replicative stress, often coincide with sites of high transcription (common fragile sites [CFSs] often coincide with large and transcriptionally active genes [Helmrich et al., 2006], and ERFSs, similarly, associate with transcriptionally active DNA [Barlow et al., 2013]). In agreement with potential transcription-replication conflicts in transformed B cell precursors, we observed an overlap of 55% of reported ERFS hotspots (Barlow et al., 2013), with 26% of the γH2AX regions identified...
here in transformed B cell precursors. Similarly, a recent report linked replicative stress in RAS-expressing immortalized fibroblasts to transcription-replication conflicts (Kotsantis et al., 2016). However, RAS-induced transcriptional interference with replication in fibroblasts related to upregulation of the TATA box binding protein TBP, but we did not observe any increase of TBP expression in transformed B cell precursors by RNA-seq (data not shown). This is in agreement with several studies by others that showed that oncogenic stress and related DNA damage can significantly differ between cell types and individual oncogenes (Maya-Mendoza et al., 2015; Miron et al., 2015). Nevertheless, we observed that ~85% of the γH2AX regions identified here in transformed B cell precursors overlapped with early replication domains described for the murine B cell lymphoma line CH12 (1,091 of 1,289 regions, p < 1 × 10−7; permutation test) (Stamatoyannopoulos et al., 2012), suggesting interference of transcription with replication as a potential cause for amplification of transcription-induced fragility in transformed B cell precursors. However, because replication timing is highly correlated with transcriptional activity in higher eukaryotes (Schübeler et al., 2002), the relative importance of replication interference in transformed B cell precursors compared with the direct effects of transcription alone needs to be determined.

**EXPERIMENTAL PROCEDURES**

**Mice**

For experiments on early-transformed B cell precursors, 53BP1−/− mice (Ward et al., 2003) were used. For experiments involving CEBPA expression (Figure 6), BCR-ABL1-transformed B cell precursors from wild-type mice were used. 53BP1−/− mice were a kind gift from Dr. Simon Boulton (The Francis Crick Institute). All experiments were performed in agreement with the Animals (Scientific Procedures) Act 1986 (ASPA) guidelines and regulations and protocols approved by the Home Office.

**Cell Culture**

Bone marrow (BM) was isolated from femora and tibiae of mice and depleted of erythrocytes using ammonium-chloride-potassium (ACK) lysis buffer (Gibco). Primary B cell precursors were then enriched and maintained by ex vivo culture in Iscove’s Modified Dulbecco’s Medium (IMDM) containing 20% fetal calf serum (FCS) (PeproTech), 1% penicillin/streptomycin (Gibco), and 2 mM L-glutamine (Gibco), and 10 ng/mL of recombinant mouse interleukin-7 (IL-7) (PeproTech). DOX-inducible CEBPA and control vectors was performed by spinoculation on days 2 and 3 after start of the early replication domains described for the murine B cell lymphoma line CH12 (1,091 of 1,289 regions, p < 1 × 10−7; permutation test) (Stamatoyannopoulos et al., 2012), suggesting interference of transcription with replication as a potential cause for amplification of transcription-induced fragility in transformed B cell precursors. However, because replication timing is highly correlated with transcriptional activity in higher eukaryotes (Schübeler et al., 2002), the relative importance of replication interference in transformed B cell precursors compared with the direct effects of transcription alone needs to be determined.

**Virally Transduced**

Transduction of B cell precursors was performed as described previously (Feldhahn et al., 2012) but using X-tremeGENE 9 DNA transfection reagent (Roche) and MIGR1 vectors. Empty MIGR1 vector and human BCR-ABL1P210-MIGR1 were kind gifts from Danilo Perrotti (University of Maryland School of Medicine). Mouse C-MYC was sub-cloned into MIGR1 from C-MYC-MIG using XhoI (New England Biolabs). Transduction with MIGR1 vectors was performed by spinoculation on days 2 and 3 after start of the cell culture. Details on vectors, transduction, and experimental conditions for experiments involving inducible DS-RED expression can be found in the Supplemental Experimental Procedures.

**ChIP-Seq, ChIP-qPCR and RNA-Seq**

Cells were prepared for ChIP-seq or ChIP-qPCR analysis according to Yamane et al. (2013) with minor modifications. For each experiment, harvested cells were depleted of dead/apoptotic cells using the dead cell removal kit (Miltenyi Biotech) before further processing. For CEBPA experiments, myeloid cells were additionally enriched using anti-CD11B micro-beads (Miltenyi). Primary transformed and untransformed B cell precursors were typically harvested on day 7 after transduction; for CEBPA experiments, cells were harvested after 7 days of DOX addition. For library preparation, the NEBNext ChIP-Seq Library Prep Master Mix Set and NEBNext Multiplex Oligos for Illumina (NEB) were used. For RNA-seq, cells were isolated, cultured, and harvested as for ChIP-seq experiments. RNA was extracted in duplicate experiments from 5 × 10⁶ cells using the QIAGEN RNeasy mini kit. Libraries were generated from pre-enriched mRNA using the NEBNext mRNA Library PrepReagent Set for Illumina (NEB) and pre-amplified using the NEBNext Multiplex Oligos for Illumina (NEB). A detailed description of the experimental procedures, bioinformatic analyses, and statistics is available in the Supplemental Experimental Procedures.

**Statistical Analysis**

Statistical analysis of DNA FISH and qRT-PCR data was done by paired Student’s t test ± SEM using GraphPad Prism software. For comparison of γH2AX regions to published datasets on gene defects in B-ALL, two-by-two contingency table analysis was performed as by Swaminathan et al. (2015). For statistical analyses of ChIP-seq and RNA-seq data, please refer to the Supplemental Experimental Procedures.

**DNA FISH**

Cultured cells were depleted of dead/apoptotic cells using the dead cell removal kit (Miltenyi), 2–5 × 10⁶ cells were then cultured in fresh medium with 100 ng/mL Colcemid (Gibco) for 4 hr and processed as described previously (Barch et al., 1997). DNA FISH was performed using custom dual-color break-apart probes (Agilent Technologies) diluted 1:100 in ULTRAhyb (Ambion) and hybridized overnight. Automated FISH capture and semi-automated analysis were performed using CytoVision GSL-120 (Leica). Detailed information on DNA FISH probes will be provided upon request.

**qRT-PCR and PCR**

For qRT-PCR on IL-7-cultured B cell precursors, cells were typically harvested on day 7 after transduction, dead/apoptotic cell-depleted, and snap-frozen on dry ice. For thymus and BM, cells were isolated freshly from 53BP1−/− mice and snap-frozen on dry ice. Stimulated B cells were obtained from the spleen and cultured as described previously (Feldhahn et al., 2012). RNA was isolated as for RNA-seq, and cDNA was generated using the ReverTra Ace cDNA synthesis kit (Thermo Fisher Scientific). qPCR was performed using a SYBR green master mix (Sigma) with 1 × 10⁶ cells on a StepOnePlus platform (Thermo Fisher Scientific). Information of primers, PCR conditions, and analysis can be found in the Supplemental Experimental Procedures. Analysis of the IGH locus for genomic loss was done as by Jankovic et al. (2013) using the primers and conditions described.

**Flow Cytometry**

Survival indicated by forward scatter (FSC)/side scatter (SSC) as well as GFP or DS-RED expression of transduced cells was assessed by flow cytometry using FACSComp or LSRFortessa (BD Biosciences). For a more detailed analysis of viability and apoptosis induction, the Annexin V-PE (phycoerythrin) apoptosis detection kit (BD Biosciences) in combination with 7-AAD (BD Biosciences) was used. For cell surface staining of single-cell suspensions, the following anti-mouse antibodies were used: CD16/32 (Fc Block, clone 2.4G2, BD Biosciences) and CD19-APC (eBio103), B220-PerCP/Cy5.5 (RA3-6B2), IgM-APC780 (II/41), CD43-PE (eBioR2/60), CD24-FITC (30-F1), and CD11b-Pecy7 (M/170) (eBioscience).

**Western Blot**

Protein lysates were made from snap-frozen cell pellets using radioimmuno-precipitation assay (RIPA) buffer (50 mM Tris-HCl [pH 8.0], 0.5 mM EDTA, 0.1% SDS, 1% NP-40/IGEPAL (octylphenoxypolyethoxyethanol), 0.5% sodium deoxycholate, and 150 mM NaCl). 10–20 μg of proteins were analyzed...
by SDS-PAGE gel electrophoresis and western blot using the Mini-PROTEAN blotting module and Tris-Glycine Extended (TGE) stain-free pre-cast gels (Bio-Rad). For detection of specific proteins, the following antibodies were used: anti-phospho-histone H2AX (serine 139, γH2AX) clone JBW301 (Millipore), anti-phospho-P53 (serine 15, Cell Signaling Technology), anti-α-tubulin (Abcam), anti-LAMIN B1 (Abcam), and anti-AID (L7E7) (Cell Signal).

ACCESSION NUMBERS

The accession number for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE80682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.01.057.

AUTHOR CONTRIBUTIONS

B.B. performed the cell culture, ChIP-seq, qRT-PCR, and western blot. M.E.R. analyzed all deep sequencing data. P.C.M. performed RNA FISH, and ChIP qPCR. L.C. helped with RNA-seq processing. J.T. performed the DS-RED experiment. K.B. optimized DNA FISH. M.M., J.F.A., and J.S. contributed by helpful discussion. N.F. designed and supervised the study, wrote the manuscript, and performed some experimental work.

ACKNOWLEDGMENTS

This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (360911/P49318) (to N.F.), research support by LEUKA (P47415) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (630911/P49318) (to N.F.), and the Imperial College Experimental Cancer Medicine Centre (630911/P49318) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (R35CA197628). This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), and Cancer Research UK (to J.S.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F.), research support by LEUKA (P47415) (to N.F.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.).

REFERENCES

Barch, M.J., Knutsen, T., Spurbeck, J.L., and Technologists, A.G. (1997). The revised accession numbers for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE80682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.01.057.

AUTHOR CONTRIBUTIONS

B.B. performed the cell culture, ChIP-seq, qRT-PCR, and western blot. M.E.R. analyzed all deep sequencing data. P.C.M. performed RNA FISH, and ChIP qPCR. L.C. helped with RNA-seq processing. J.T. performed the DS-RED experiment. K.B. optimized DNA FISH. M.M., J.F.A., and J.S. contributed by helpful discussion. N.F. designed and supervised the study, wrote the manuscript, and performed some experimental work.

ACKNOWLEDGMENTS

This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (630911/P49318) (to N.F.), research support by LEUKA (P47415) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F. and J.S.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.), research support by LEUKA (P47415) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (R35CA197628). This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), and Cancer Research UK (to J.S.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F.), research support by LEUKA (P47415) (to N.F.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.).

REFERENCES

Barch, M.J., Knutsen, T., Spurbeck, J.L., and Technologists, A.G. (1997). The revised accession numbers for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE80682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.01.057.

AUTHOR CONTRIBUTIONS

B.B. performed the cell culture, ChIP-seq, qRT-PCR, and western blot. M.E.R. analyzed all deep sequencing data. P.C.M. performed RNA FISH, and ChIP qPCR. L.C. helped with RNA-seq processing. J.T. performed the DS-RED experiment. K.B. optimized DNA FISH. M.M., J.F.A., and J.S. contributed by helpful discussion. N.F. designed and supervised the study, wrote the manuscript, and performed some experimental work.

ACKNOWLEDGMENTS

This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (630911/P49318) (to N.F.), research support by LEUKA (P47415) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F. and J.S.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.).

REFERENCES

Barch, M.J., Knutsen, T., Spurbeck, J.L., and Technologists, A.G. (1997). The revised accession numbers for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE80682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.01.057.

AUTHOR CONTRIBUTIONS

B.B. performed the cell culture, ChIP-seq, qRT-PCR, and western blot. M.E.R. analyzed all deep sequencing data. P.C.M. performed RNA FISH, and ChIP qPCR. L.C. helped with RNA-seq processing. J.T. performed the DS-RED experiment. K.B. optimized DNA FISH. M.M., J.F.A., and J.S. contributed by helpful discussion. N.F. designed and supervised the study, wrote the manuscript, and performed some experimental work.

ACKNOWLEDGMENTS

This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (630911/P49318) (to N.F.), research support by LEUKA (P47415) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F. and J.S.), and the Imperial College Experimental Cancer Medicine Centre and Biomedical Research Centre. M.M. is an HHMI Faculty Scholar (HHMI-55108547) and supported by an NCI Outstanding Investigator Award (R35CA197628).

Received: April 22, 2016
Revised: November 29, 2016
Accepted: January 23, 2017
Published: February 14, 2017

REFERENCES

Barch, M.J., Knutsen, T., Spurbeck, J.L., and Technologists, A.G. (1997). The revised accession numbers for the RNA-seq and ChIP-seq data reported in this paper is NCBI GEO: GSE80682.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.01.057.

AUTHOR CONTRIBUTIONS

B.B. performed the cell culture, ChIP-seq, qRT-PCR, and western blot. M.E.R. analyzed all deep sequencing data. P.C.M. performed RNA FISH, and ChIP qPCR. L.C. helped with RNA-seq processing. J.T. performed the DS-RED experiment. K.B. optimized DNA FISH. M.M., J.F.A., and J.S. contributed by helpful discussion. N.F. designed and supervised the study, wrote the manuscript, and performed some experimental work.

ACKNOWLEDGMENTS

This work is supported by a Bennett Fellowship of Bloodwise (13016/P47463) (to N.F.), a Marie Curie Career Integration Grant from the European Commission (630911/P49318) (to N.F.), research support by LEUKA (P47415) (to N.F.), Action Against Cancer and the Hilary Craft Foundation (PS8508) (to N.F. and J.S.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.), and the Imperial College Experimental Cancer Medicine Centre (PS8508) (to N.F.).
Kotsantis, P., Silva, L.M., Irmischer, S., Jones, R.M., Folkes, L., Gromak, N., and Petermann, E. (2016). Increased global transcription activity as a mechanism of replication stress in cancer. Nat. Commun. 7, 13087.

Lengauer, C., Knizler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature 396, 643–649.

Lensing, S.V., Marsico, G., Hansel-Hertsch, R., Lam, E.Y., Tannahill, D., and Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. Nat. Methods 13, 855–857.

Liu, G.J., Cimmino, L., Jude, J.G., Hu, Y., Witkowski, M.T., McKenzie, M.D., Kartal-Kaess, M., Best, S.A., Tuohy, L., Liao, Y., et al. (2014). Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia. Genes Dev. 28, 1337–1350.

Madabhushi, R., Gao, F., Penning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.C., et al. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell 161, 1592–1605.

Maya-Mendoza, A., Ostrokova, J., Kosar, M., Hall, A., Dusko, P., Mistrlik, M., Merchut-Mayo, J.M., Hodny, Z., Bartkova, J., Christensen, C., and Bartek, J. (2015). Myc and Ras oncogenes engage different energy metabolism programs and evoke distinct patterns of oxidative and DNA repair stress. Mol. Oncol. 9, 601–616.

MIron, K., Golan-Lev, T., Divir, R., Ben-David, E., and Kerem, B. (2015). Oncogenes create a unique landscape of fragile sites. Nat. Commun. 6, 7094.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., et al. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukemia. Nature 446, 758–764.

Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.H., et al. (2008). BCR-ABL1 lymphoblastic leukemia is characterized by the deletion of Ikaros. Nature 453, 110–114.

Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., Schulman, B.A., et al.; Children’s Oncology Group (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N. Engl. J. Med. 360, 470–480.

Pannunzio, N.R., and Lieber, M.R. (2016). Dissecting the Roles of Divergent and Convergent Transcription in Chromosome Instability. Cell Rep. 14, 1025–1031.

Roberts, K.G., Morin, R.D., Zhang, J., Hirsh, M., Zhao, Y., Su, X., Chen, S.C., Payne-Turner, D., Churchman, M.L., Harvey, R.C., et al. (2012). Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell 22, 153–166.

Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. 8, 301–310.

Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nat. Rev. Immunol. 7, 105–117.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat. Rev. Genet. 16, 583–597.

Sattler, M., Verma, S., Shrikhande, G., Byrne, C.H., Pride, Y.B., Winkler, T., Greenfield, E.A., Salgia, R., and Griffin, J.D. (2000). The BCR/ABL tyrosine kinase induces production of reactive oxygen species in hematopoietic cells. J. Biol. Chem. 275, 24273–24278.

Sawyers, C.L. (1993). The role of myc in transformation by BCR-ABL. Leuk. Lymphoma 11 (Suppl 1), 45–46.

Schübel, D., Scazo, D., Kooperberg, C., van Steensel, B., Delrow, J., and Groudine, M. (2002). Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. Nat. Genet. 32, 438–442.

Seo, J., Kim, S.C., Lee, H.S., Kim, J.K., Shon, H.J., Saleh, N.L., Desai, K.V., Lee, J.H., Kang, E.S., Kim, J.S., and Choi, J.K. (2012). Genome-wide profiles of H2AX and γ-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. Nucleic Acids Res. 40, 5965–5974.

Shyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286.

Skourtis-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396.

Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaur, R., Canfield, T., et al.; Mouse ENCODE Consortium (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. 13, 418.

Swamianathan, S., Klemm, L., Park, E., Papamannui, E., Ford, A., Kweon, S.M., Trgeser, D., Hassellfelt, B., Henke, N., Moorster, J., et al. (2015). Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. Nat. Immunol. 16, 766–774.

Vafa, O., Wade, M., Kern, S., Beeche, M., Pandita, T.K., Hampton, G.M., and Wahl, G.M. (2002). c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability. Mol. Cell 9, 1031–1044.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science 339, 1546–1558.

Wang, X., Fan, M., Kailis, S., Wei, L., and Scharff, M.D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. Nat. Commun. 5, 4137.

Ward, I.M., Minn, K., van Deursen, J., and Chen, J. (2003). p53 Binding protein 3BP1 is required for DNA damage responses and tumor suppression in mice. Mol. Cell. Biol. 23, 2556–2563.

Yamane, A., Robbiano, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., and Casellas, R. (2013). RPA accumulation during class switch recombination represents 5′–3′ DNA-end resection during the S-G2/M phase of the cell cycle. Cell Rep. 3, 138–147.

Zhang, J., Mullighan, C.G., Harvey, R.C., Wu, G., Chen, X., Edmonson, M., Buetow, K.H., Carroll, W.L., Chen, I.M., Devidas, M., et al. (2011). Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children’s Oncology Group. Blood 118, 3080–3087.
Supplemental Information

Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors

Bryant Boulianne, Mark E. Robinson, Philippa C. May, Leandro Castellano, Kevin Blighe, Jennifer Thomas, Alistair Reid, Markus Müschen, Jane F. Apperley, Justin Stebbing, and Niklas Feldhahn
Figure S1. Analysis of DNA damage in transformed B-cell precursors by γH2AX-ChIP-Seq. Related to Figure 1

A) Analysis of B-cell precursor enrichment after 10 days of cell culture of total bone marrow in presence of IL7 by flow cytometry using antibodies indicated. BCR-ABL1, C-MYC and EV (empty vector) cells were additionally transduced on day 2 and 3 by respective MIGR1 encoding retrovirus. As a positive control for functionality of the anti-IgM antibody a staining of splenocytes is shown.

B) Western blot analysis of infected vs. non-infected (mock) B-cell precursor cells for DNA damage marks γH2AX and p-P53Ser15 using LAMIN B1 as loading control (independent experiment to Figure 1C/D).

C) qRT-PCR analysis for RAG1 (left) and AID (middle) expression in transformed vs. untransformed B-cell precursors (D7 post oncogene induction [p.i.]) using total thymus and splenic LPS/IL4-stimulated mature B-cells as positive controls for RAG1 and AID, respectively. Data represents three independent experiments, error bars indicate mean +/- SEM. (Right) Analysis of AID protein abundance in transformed B-cell precursors by Western blot using the antibodies indicated.

D) Comparison of γH2AX-ChIP-seq repeat experiments for BCR-ABL1 and MYC transduced B-cell precursors by scatter plot visualization. Normalized RPM (nRPM) values of the γH2AX signal within identified γH2AX regions (merged list) are shown for BCR-ABL1 (left) and MYC (right), red line represents linear regression fit.

E) Comparison of γH2AX-ChIP-seq repeat-experiments by custom track visualization. A representative image for BCR-ABL1-transduced B-cell precursors is shown. The isotype control library (Iso) visualizes the background ChIP-seq signal. Gene bodies and orientation is indicated with black bars.
Figure S2. Enrichment of the DNA damage response protein γH2AX at B-cell lineage genes, Related to Figure 2

A-C) γH2AX-ChIP-seq custom track images of B-cell precursor cells expressing C-MYC or BCR-ABL1 (B/A) for B-cell lineage genes. These include (A) genes forming the pre-B-cell receptor complex, (B) genes required for pre-B-cell receptor signal transduction, and (C) genes required for B-cell lineage commitment/differentiation. The isotype control library (Iso) visualizes the background ChIP-seq signal. Gene bodies of the related gene are indicated with black bars, neighboring genes in grey. D) Gene ontology (GO) pathway enrichment analysis of identified γH2AX regions: (Left) Pie charts showing GO enrichment categories. (Right) Bar diagram showing in more detail the GO pathways that resemble enrichment categories of Pie charts on the left.
Figure S3. Enrichment of γH2AX at genes known to become defective in B-ALL, Related to Figure 3

A) γH2AX-ChIP-seq custom track images of B-cell precursor cells expressing C-MYC or BCR-ABL1 (B/A) for 27 genes with significant γH2AX signals identified by SICER and that are reported to become defective in human B-ALL. The isotype control library (Iso) visualizes the background ChIP-seq signal. Gene bodies of related genes are indicated with black bars, neighboring genes in grey. B) Visualization as in (A) for 8 genes reported to become defective in B-ALL that exhibit γH2AX signals but did not pass our stringent selection criteria by SICER.
Figure S4. H3K27ac and γH2AX in transformed B-cell precursors, Related to Figure 4 and 5

A) Analysis of structural variation (SV) breakpoint (Bp) frequencies from SVs reported for human leukemia genomes for their distribution over genic and intergenic regions as indicated. Inset shows the sum of Bp’s for all data sets analyzed compared to the normal distribution of genic and intergenic DNA within the human genome. B) Distribution of H3K27ac reads within gene bodies vs. upstream and downstream regions for transformed B-cell precursors by meta-gene analysis. Signals are normalized to isotype-Ctrl ChIP-seq library (Iso). C) Analysis of γH2AX accumulation upon Doxycycline (DOX)-induced DS-RED expression in C-MYC-transformed B-cell precursors by ChIP qPCR. (Left) Schematic diagram of experimental strategy for the DS-RED experiment. (Middle) Verification of DOX-induced DS-RED expression in C-MYC-transformed B-cell precursors by flow cytometry. (Right) Bar diagram showing the increase of DS-RED sequence within the γH2AX-ChIP fraction normalized to input DNA used for ChIP, for +DOX vs. -DOX treated cells, as determined by ChIP qPCR. D) Custom track images of the γH2AX-Chip-seq signal at two genes with γH2AX regions in BCR-ABL1- and C-MYC-transformed B-cell precursors and respective locations of predicted R-loop forming sequences (RLFS) (mint).
Supplemental Table Legends

**Table S1. BCR-ABL1 γH2AX regions summary, related to Figure 1 and 2.** This table contains the lists of annotated γH2AX regions identified by SICER for BCR-ABL1-transformed B-cell precursors. Shown are the normalized reads per million (RPM; isotype background subtracted), fold change, reads per kilo base per million (RPKM) and gene annotations of regions enriched for the γH2AX ChIP-seq signal versus the isotype control library.

**Table S2. C-MYC γH2AX regions summary, related to Figure 1 and 2.** This table contains the lists of annotated γH2AX regions identified by SICER for C-MYC-transformed B-cell precursors. Shown are the normalized reads per million (RPM; isotype background subtracted), fold change, reads per kilo base per million (RPKM) and gene annotations of regions enriched for the γH2AX ChIP-seq signal versus the isotype control library.

**Table S3. Merged γH2AX regions summary, related to Figure 1, 4 and 5.** This table contains the merged list of annotated γH2AX regions. Shown are the annotations and chromosomal locations of γH2AX enriched regions present in BCR-ABL1-transformed B-cell precursors, C-MYC-transformed B-cell precursors or in both as indicated.

**Table S4. Merged γH2AX regions GO results, related to Figure S2D.** This table contains the output of the Gene Ontology (GO) analysis of γH2AX regions. Shown are the enriched biological process (BP) gene ontology (GO) categories for genes associated with the merged γH2AX enriched regions indicated by Table S3.

**Table S5. Gene defects in human B-ALL vs. γH2AX regions, related to Figure 3.** The table shows the comparisons of genes associated with γH2AX-enriched regions in transformed B-cell precursors from this study to genes associated with genomic deletions or translocations in human B-ALL identified by Mullighan et al., 2007, Mullighan et al., 2009 or Roberts et al., 2012 as indicated. Data used for comparison relates to the information provided by the supplementary information of these studies, as indicated in this table and the Supplemental Experimental Procedures section.

**Table S6. CEBPA+DOX specific γH2AX regions summary, related to Figure 6.** This table contains the lists of annotated γH2AX regions identified by SICER as specific for myeloid-reprogrammed BCR-ABL1-transformed B-cell precursors. Shown are fold change, reads per kilo base per million (RPKM) and gene annotations of regions enriched for the γH2AX ChIP-seq signal in DOX-treated vs. untreated CEBPA-transgenic BCR-ABL1-transformed B-cell precursors.

**Table S7. CEBPA+DOX specific γH2AX regions GO results, related to Figure 6.** This table contains the output of the Gene Ontology (GO) analysis of γH2AX regions identified by SICER as specific for myeloid-reprogrammed BCR-ABL1-transformed B-cell precursors. Shown are the enriched biological process (BP) gene ontology (GO) categories for genes associated with the γH2AX enriched regions in DOX-treated vs. untreated CEBPA-transgenic BCR-ABL1 transformed B-cell precursors.
Supplemental Experimental Procedures

Chromatin immunoprecipitation (ChIP): Ex-vivo cultured B-cell precursor cells were prepared for ChIP-seq analysis according to (Yamane et al., 2013) with minor modifications: All cells used for ChIP-seq were depleted of dead/apoptotic cells using the dead cell removal kit (Miltenyi). Formaldehyde used for crosslinking was from Alfa Aesar and cells were snap-frozen as cell pellets before start of the ChIP procedure. Before sonication, cell pellets were re-suspended in 130µL of sonication buffer (0.5% SDS, 10mM Tris·HCl pH7.6) supplemented with protease and phosphatase inhibitor (Roche) and incubated for 20 minutes on ice. Sonication was performed using microTUBE AFA 6x16mm on the Covaris S220 sonicator and the following settings to yield an average DNA fragment size of 300bp: peak power 105, duty factor 5 and cycles/burst 200 for 5 minutes. Sonicated cells were transferred to concentrated RIPA buffer (10mM Tris·HCl pH7.6, 1.15mM EDTA, 0.04% SDS, 1.15% TritonX-100, 0.115% sodium deoxycholate) to yield 1ml of 1X RIPA buffer described in (Yamane et al., 2013), and subjected to ChIP using 5µg ab/12-15x10^6 cells of the following antibodies: Anti-gamma H2A.X (phospho S139) antibody - ChIP Grade (ab2893) (Abcam), Anti-Histone H3 (acetyl K27) antibody – ChIP Grade (ab4729) (Abcam), polyclonal rabbit IgG (isotype control) (sc2027) (Santa Cruz Biotechnologies). All sonications were verified by agarose gel electrophoresis, and respective immunoprecipitations controlled by Western blot. ChIP pull-down DNA was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher).

ChIP-Seq library preparation: ChIP-seq libraries were generated from ChIP pull-down DNA or from input DNA of the same ChIP procedure in parallel as control using the NEBNext® ChIP-seq Library Prep Master Mix Set and NEBNext Multiplex Oligos for Illumina (NEB). Pre-amplified ChIP-seq libraries were then sequenced by the genomics facilities of the Medical Research Council (MRC) and Imperial College London (ICL).

ChIP-Seq libraries: In total, 15 ChIP libraries were generated: Two repeats of γH2AX-ChIP-seq from two individual and time-separated experiments for BCR-ABL1 and C-MYC transduced and IL7 cultured B-cell precursors harvested on day 7 post oncogene induction. Respective control libraries using isotype control antibodies for ChIP or input DNA were generated in parallel. In addition, one γH2AX-ChIP-seq library from long-term cultured (>D50) BCR-ABL1-transformed B-cell precursors was generated in parallel. These >D50 cells had been additionally treated for 2 hours with 10mM Hydroxyurea before harvest. For H3K27ac analysis four ChIP-seq libraries have been generated: One H3K27ac-ChIP-seq library each for BCR-ABL1 and C-MYC transduced IL7 cultured B-cell precursors harvested on day 7 post oncogene induction, and two respective input DNA control libraries. Five additional ChIP-seq libraries were generated on BCR-ABL1-transformed B-cell precursors (CEBPA experiment, Figure 6) cultured in the absence of IL7 and originally obtained from Markus M üsch en laboratory (Chen et al., 2015). These include one library each for cells carrying an inducible CEBPA transgene treated either with Doxycycline or not, and one library for control cells carrying the inducible vector without CEBPA (empty vector, EV) and treated with Doxycycline. For CEBPA +DOX and EV +DOX conditions input DNA control libraries were generated.

ChIP-Seq read processing and generation of custom tracks: Sequence reads of 50 bps were obtained through standard Illumina image-analysis and default quality filters. Reads were aligned to the Build 38 mouse genome assembly (GRCh38/mm10) using BWA-MEM with default settings. Non-uniquely aligned reads were removed using SAMtools (Li and Durbin, 2009) and PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard). To generate custom tracks for visualization, coverage within 200 bp windows was calculated and normalized to library and window size using deepTools (Ramirez et al., 2014) presented as reads per kilobase per million mapped reads (RPKM).

ChIP-Seq read enriched region calling: Region calling was done as described in (Barlow et al., 2013) with minor modifications to further optimize calling of γH2AX-enriched regions. In brief, sequence reads were pre-processed as described above and statistical enrichment of uniquely aligned reads mapping within non-overlapping 100bp windows was analyzed relative to a Poissonian random background using the SICER program (Zang et al., 2009) with parameters: e-value 10,000, gap size 300. Statistical enrichment of bins passing the first criterion was examined relative to negative control DNA from either input or isotype control ChIP reactions based on tag density fold increase determined from False Discovery Rate of 1 x 10^-5. Subsequently, identified enriched windows that were less than 10 kb apart were merged to delineate enriched genomic regions. Finally, an empirical filtering step was applied to eliminate any regions with low coverage (<1.5 RPKM), width (<5 kb) or
fold change (< 1.1), resulting in the final lists of highly enriched regions. While these settings were optimized for calling of γH2AX-enriched regions, the same settings were used for H3K27ac enriched region calling.

**Generation of a merged γH2AX-enriched region list:** Stringent lists of γH2AX-enriched regions from BCR-ABL1 and C-MYC were merged using the GRanges package (Lawrence et al., 2013) in R (R Core Team, 2015; https://www.R-project.org), taking the union of any overlapping regions.

**Chromosomal distribution of enriched regions:** RPM values for γH2AX-enriched regions from the merged stringent list were calculated using BEDTools (Quinlan and Hall, 2010). Ideograms in Figure 1F were plotted using the IdeoViz package in R and RPM values from BCR-ABL1 (run1). The percentage of total chromosome coverage for the merged region list shown in Figure 1G was calculated using Granges (Lawrence et al., 2013) and plotted against exon coverage estimates obtained from (Sakharkar et al., 2005).

**Correlation of γH2AX signals from different ChIP-Seq libraries:** RPM values for γH2AX-enriched regions from the merged stringent list were calculated using BEDTools (Quinlan and Hall, 2010) for BCR-ABL1 and C-MYC samples, with subtraction of the isotype signal. Scatter plots of log2 signal intensity were generated and overlapping/unique regions highlighted. Linear regression fit (red line) and Pearson’s rho values (r) were calculated for each comparison.

**Analysis of enriched regions for gene ontology (GO) pathway enrichment:** Genes overlapping γH2AX-enriched regions from BCR-ABL1 and C-MYC samples were identified using the R package ChIPpeakAnno (Zhu et al., 2010) and enriched biological process GO categories identified (P < 0.01). Enriched GO terms were associated with four empirically chosen parent categories: chromosome organization (GO:0051276), gene expression (GO:0010467), nucleic acid metabolic process (GO:0090304) and immune system process (GO:0002376). Pie charts were plotted for each enriched set before identifying common categories between BCR-ABL1 and C-MYC samples, removal of redundant GO terms and plotting of -log10 p-values.

**Genic distribution of enriched regions:** Meta-gene profiles of γH2AX and H3K27ac signal distribution in Inset of Figure 4A and in Figure S4B were plotted for all Ensembl protein-coding genes (n=22,533) plus 10kb up- and down-stream using the ngs.plot package (Shen et al., 2014). Signal profiles for all meta-gene profiles are normalized to gene length by spline fitting, to library size by RPM, and plotted as log2 fold change over isotype control ChIP-seq library. The intersection of enriched regions overlapping genic features shown in Figure 4A was determined using the ChIPSeeker R package (Yu et al., 2015) with TSS defined as +/- 1kb and default settings.

**Correlation of γH2AX and H3K27ac signals:** Stringent region lists for both H3K27ac and γH2AX were merged for BCR-ABL1 and C-MYC, and RPM values within these regions were calculated from BCR-ABL1, C-MYC and isotype control samples using BEDTools before performing background subtraction. Negative signal values were arbitrarily set to 0.01 RPM for visualization purposes and excluded during calculation of Pearson's correlation coefficients. Scatter plots of log10 signal intensity were generated and overlapping/unique regions highlighted. Red lines indicate the linear regression fit.

**Comparison of γH2AX regions to location of ERFS:** Identified γH2AX regions were compared to a previously published ChIP-seq dataset that characterized early replicating fragile sites (ERFS) in Hydroxyurea-treated mature B-cells (Barlow et al., 2013) (region list: Table S1, accession number: SRR648771). Published ERFS regions were converted from NCBI37/mm9 to GRCm38/mm10 (UCSC mm9ToMm10 chain file) for comparison to mm10 referenced γH2AX regions. FASTQ reads were extracted from SRA files using the SRA-toolkit and aligned against the mm10 genome as detailed for our own data. The observed overlap between regions was compared against the average number of permuted overlaps. Permutation models were performed with sex chromosomes and unmapable regions masked for 100,000 enumerations.

**Comparison of γH2AX regions to location of RLFS:** Locations of potential R-loop forming sequences (RLFS) were identified throughout the mm10 genome using QmRLFS-finder (Jenjaroenpun et al., 2015). Permutation testing was performed with the R package regioneR (Gel et al., 2016) with 10,000 enumerations per test. Analysis of RLFS enrichment within γH2AX regions was determined by re-sampling of genomic regions within masked mm10 genome, while enrichment within genes was determined through re-sampling of UCSC known-
genes. Enrichment of RLFS within highly expressed genes was performed in a similar manner but first restricting the total gene population to the top 1000 most high-expressed genes as ranked by mean expression across all conditions analyzed by RNA-seq.

**Comparison of γH2AX regions to genomic defects in leukemia:** For comparison of gene sets as shown in Figure 3B, datasets published in Mullighan et al 2007 (Mullighan et al., 2007) (Table S10 in respective article), Mullighan et al 2009 (Mullighan et al., 2009) (Table S4 in respective article) and Roberts et al 2012 (Roberts et al., 2012) (Table S3 and S7 in respective article) were used. Only defects that were reported for B-ALL and that relate to genomic deletion or translocation were used for comparison. Further, only those defects could be compared that were indicated to relate to annotated genes, and for which a respective counterpart could be found in the mouse genome. For example, iAmp21, Znf528, Ccdc26, Fam22f, C20orf94 and FLJ11273 from the Mullighan datasets were excluded from analysis because we could not relate them to a unique mouse gene. Also, lesions of extreme large sizes were not included into the analysis (e.g. ‘all 10p’ or ‘383 genes involved’). Reported defects affecting multiple genes (e.g. Atm/Rab39/Npat) were counted as one defect, and only one overlap of γH2AX regions was counted even if multiple genes within the defect exhibited overlap with respective γH2AX regions. This lead to a list of 39 alterations for comparison for the Mullighan et al 2007 dataset, 42 alterations for comparison for the Mullighan et al 2009 dataset, and 42 alterations for comparison for the Roberts et al 2012 dataset as indicated in detail in Supplemental Table S5.

**Analysis of genic vs. intergenic distribution of structural variant (SV) breakpoints in leukemia genomes:** To analyze the genic vs. intergenic distribution of SV breakpoints identified in leukemia genomes as shown in Figure S4A, published breakpoint coordinates from Mullighan et al 2007 (Mullighan et al., 2007) (Table S10 in respective article), Papaemmanuil et al 2014 (Papaemmanuil et al., 2014) (Table S4 in respective article), Andersson et al 2015 (Andersson et al., 2015) (Table S7 in respective article), Holmfeldt et al 2013 (Holmfeldt et al., 2013) (Table S8 in respective article), and Paulsson et al 2015 (Paulsson et al., 2015) (Table S11 in respective article) were used. SV breakpoint coordinates were mapped to gene regions in R (exon, intron, promoter = 3kb upstream, downstream = 3kb downstream, distal = >3kb from gene boundaries). Significance of enrichment in genic regions was tested by hypergeometric distribution test.

**Comparison of γH2AX regions to replication timing:** γH2AX regions were categorized as early or late replicating by comparison with ENCODE Repli-chip data (Mouse et al., 2012). A zero-crossing algorithm was applied to wavelet-smoothed Repli-chip data from the CH12 cell line to identify early-replicating regions, and only regions called as early replicating in both biological replicates used to annotate γH2AX regions. Permutation models were performed as described previously.

**RNA-Seq analysis:** Reads were aligned to the mouse genome (GRCm38/mm10) using STAR v2.5.0 (Dobin et al., 2013), indexes were supplemented with UCSC known gene reference transcript assemblies. Alignment was performed with the following settings: --outFilterType BySJout, --outFilterMultimamNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --outFilterMismatchNoverLmax 0.04, --alignIntronMin 20, --alignIntronMax 1000000. FPKM values were calculated with StringTie v1.3.0 (Pertea et al., 2015). Differential expression analysis was performed in R (R Core Team, 2016; https://www.R-project.org) with DESeq2 v1.12.4 (Love et al., 2014).

**Comparison of RNA-Seq and γH2AX-ChIP-seq data:** All UCSC known genes were categorized into low or high expression groups relative to the median FPKM value of expressed genes for each condition. The proportion of genes overlapping γH2AX regions falling into high or low expression categories was then determined. To compare γH2AX signals in high vs. low expression categories and in BCR-ABL1 vs. C-MYC differentially expressed comparison, all UCSC-known gene regions were extended by 10 kb and the γH2AX signal within these regions determined using the R package bamsignals v1.4.3, normalized to library size, and background signal from isotype control libraries subtracted.

**DS-RED ChIP-qPCR:** For analysis of DNA damage upon inducible expression of the fluorochrome DS-RED, the expression vector TRMPVIR (Addgene) was modified by digestion with EcoRV and Xhol to remove the mVenus and rtTA cassettes. MYC induced B-cell precursor cells used for the experiments were generated from 53BP1/-/- mice and first transduced with the TET3G expression plasmid (Addgene) followed by selection using 500µg/mL Geneticin/G418 (Sigma). Pre-selected cells were then transduced in triplicate with 24h time between
each round of transduction using retrovirus encoding the modified TRMPVIR vector. Three days after the last transduction, 1µg/mL DOX was added to the media. After three days of DOX treatment cells were harvested for analysis by ChIP-qPCR. Cells were processed similar to cells designated for ChIP-seq including analysis by flow cytometry and dead cell removal by MACS. ChIP DNA was analysed in triplicate using SYBR Green Jumpstart Taq ReadyMix (Sigma) on the StepOnePlus (Applied Biosciences). For quantification of DS-RED the following primers were used: CGAGTTCATGCCTTCAAGG as forward primer and GTCACCTTCAGCTTCACGCT as reverse primer.

qRT-PCR. For IL-7 cultured control and transformed cells, cells were harvested at day 7 p.i. as described, dead cells were removed by dead cell removal kit (MACS, Miltenyi), and then snap frozen on dry ice. For thymus and BM, cells were isolated fresh from 53BP1−/− mice and snap frozen on dry ice. Stimulated B-cells were obtained fresh from 53BP1−/− mice and cultured as previously described (Feldhahn et al., 2012). RNA was isolated from cells using a column-based kit (Qiagen RNeasy) and cDNA was generated using RevertAid cDNA Synthesis Kit (Thermo Fisher). qPCR was performed using a SYBR green master mix (Sigma JumpStart) on a StepOnePlus platform (Thermo Fisher) and the following primer pairs: RAG1 forward AGGCCTGTGGAGCAAGGTAG and reverse TTTCATCGGTGCAGAACTGA at 63ºC annealing; AID forward CCCTTGTACGAAGTCGATGAC and reverse ATCACGTGTGACATTCCAGGAG at 60ºC annealing (Crouch et al., 2007) HPRT forward CCCAGGTGTGATTAGC and reverse GGAATAAACACTTTTTCCAAAT at 60ºC annealing. Reactions were performed in triplicate and the average Ct and reaction efficiency were analyzed using the Pfaffl method (Pfaffl, 2001) to determine relative expression of AID and RAG1 using HPRT as the endogenous control. Expression levels of genes in experimental cells were plotted relative to expression in resting, naïve splenic B-cells.
Supplemental References

Andersson, A.K., Ma, J., Wang, J., Chen, X., Gedman, A.L., Dang, J., Naktandwe, J., Holmfeldt, L., Parker, M., Easton, J., et al. (2015). The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet 47, 330-337.

Barlow, J.H., Faryabi, R.B., Callen, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell 152, 620-632.

Berger, M., Krebs, P., Crozat, K., Li, X., Croker, B.A., Siggs, O.M., Popkin, D., Du, X., Lawson, B.R., Theofilopoulos, A.N., et al. (2010). An Sflg2 mutation causes lymphoid and myeloid immunodeficiency due to loss of immune cell quiescence. Nat Immunol 11, 335-343.

Chen, Z., Shojaee, S., Buchner, M., Geng, H., Lee, J.W., Klemm, L., Titz, B., Graeber, T.G., Park, E., Tan, Y.X., et al. (2015). Signalling thresholds and negative B-cell selection in acute lymphoblastic leukaemia. Nature 521, 357-361.

Clausen, B.E., Burkhardt, C., Reith, W., Renkawitz, R., and Forster, I. (1999). Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic Res 8, 265-277.

Crouch, E.E., Li, Z., Takizawa, M., Fichtner-Feigl, S., Gourzi, P., Montano, C., Feigenbaum, L., Wilson, P., Janz, S., Papavasiliou, F.N., et al. (2007). Regulation of AID expression in the immune response. J Exp Med 204, 1145-1156.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

Feingold, K.R., Kazemi, M.R., Magra, A.L., McDonald, C.M., Chui, L.G., Shigenaga, J.K., Patzek, S.M., Chan, Z.W., Londos, C., and Grunfeld, C. (2010). ADRP/ADFP and Mal1 expression are increased in macrophages treated with TLR agonists. Atherosclerosis 209, 81-88.

Feldhahn, N., Feretti, E., Robbiani, D.F., Callen, E., Deroubaix, S., Selleri, L., Nussenzweig, A., and Nussenzweig, M.C. (2012). The hSSB1 orthologue Obfc2b is essential for skeletogenesis but dispensable for the DNA damage response in vivo. EMBO J 31, 4045-4056.

Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289-291.

Holmfeldt, L., Wei, L., Diaz-Flores, E., Walsh, M., Zhang, J., Ding, L., Payne-Turner, D., Churchman, M., Andersson, A., Chen, S.C., et al. (2013). The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet 45, 242-252.

Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S.P., and Kuznetsov, V.A. (2015). QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. Nucleic Acids Res 43, W527-W534.

Kim, Y., Schulz, V.P., Satake, N., Gruber, T.A., Teixeira, A.M., Halene, S., Gallagher, P.G., and Krause, D.S. (2014). Whole-exome sequencing identifies a novel somatic mutation in MMP8 associated with a t(1;22)-acute megakaryoblastic leukemia. Leukemia 28, 945-948.

Lawrence, M., Huber, W., Pages, H., Abyoung, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS computational biology 9, e1003118.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25, 1754-1760.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., et al. (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome biology 13, 418.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., et al. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 446, 758-764.

Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., Schulman, B.A., et al. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med 360, 470-480.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet 46, 116-125.

Park, Y.J., Yoon, S.J., Suh, H.W., Kim, D.O., Park, J.R., Jung, H., Kim, T.D., Yoon, S.R., Min, J.K., Na, H.J., et al. (2013). TXNIP deficiency exacerbates endotoxic shock via the induction of excessive nitric oxide synthesis. PLoS Pathog 9, e1003646.

Paul, S.P., Taylor, L.S., Stansbury, E.K., and McVicar, D.W. (2000). Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2. Blood 96, 483-490.

Paulsson, K., Lilljebjorn, H., Biloglav, A., Olsson, L., Rissler, M., Castor, A., Barbany, G., Fogelstrand, L., Nordgren, A., Sjogren, H., et al. (2015). The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. Nat Genet 47, 672-676.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33, 290-295.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research 29, e45.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research 42, W187-191.

Roberts, K.G., Morin, R.D., Zhang, J., Hirst, M., Zhao, Y., Su, X., Chen, S.C., Payne-Turner, D., Churchman, M.L., Harvey, R.C., et al. (2012). Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell 22, 153-166.

Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R., and Kangeuane, P. (2005). An analysis on gene architecture in human and mouse genomes. In silico biology 5, 347-365.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC genomics 15, 284.
Veillette, A., Thibaudeau, E., and Latour, S. (1998). High expression of inhibitory receptor SHPS-1 and its association with protein-tyrosine phosphatase SHP-1 in macrophages. J Biol Chem 273, 22719-22728.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., et al. (2013). RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. Cell reports 3, 138-147.

Yokoyama, T., Kanno, Y., Yamazaki, Y., Takahara, T., Miyata, S., and Nakamura, T. (2010). Trib1 links the MEK1/ERK pathway in myeloid leukemogenesis. Blood 116, 2768-2775.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382-2383.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25, 1952-1958.

Zhao, Y., Xiong, Z., Lechner, E.J., Klenotic, P.A., Hamburg, B.J., Hulver, M., Khare, A., Oriss, T., Mangalmurti, N., Chan, Y., et al. (2014). Thrombospondin-1 triggers macrophage IL-10 production and promotes resolution of experimental lung injury. Mucosal Immunol 7, 440-448.

Zhou, X., Sun, L., Bastos de Oliveira, F., Qi, X., Brown, W.J., Smolka, M.B., Sun, Y., and Hu, F. (2015). Prosaposin facilitates sortilin-independent lysosomal trafficking of progranulin. J Cell Biol 210, 991-1002.

Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC bioinformatics 11, 237.