

## MYELOID NEOPLASIA

Terminal deoxynucleotidyl transferase promotes acute myeloid leukemia by priming *FLT3*-ITD replication slippageJulian Borrow,<sup>1,2</sup> Sara A. Dyer,<sup>1,2</sup> Susanna Akiki,<sup>1,3</sup> and Michael J. Griffiths<sup>1,2</sup><sup>1</sup>West Midlands Regional Genetics Laboratory, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, United Kingdom; <sup>2</sup>Institute of Cancer and Genomic Studies, University of Birmingham, Birmingham, United Kingdom; and <sup>3</sup>Diagnostic Genomic Division, Department of Laboratory Medicine and Pathology, Hamad Medical Corporation, Doha, Qatar

## KEY POINTS

- *FLT3*-ITDs unexpectedly show junctional N-nucleotides with properties consistent with synthesis by TdT.
- Off-target TdT activity in AML is proposed to promote *FLT3*-ITD formation by priming replication slippage.

***FLT3*-internal tandem duplications (*FLT3*-ITDs) are prognostic driver mutations found in acute myeloid leukemia (AML). Although these short duplications occur in 25% of AML patients, little is known about the molecular mechanism underlying their formation. Understanding the origin of *FLT3*-ITDs would advance our understanding of the genesis of AML. We analyzed the sequence and molecular anatomy of 300 *FLT3*-ITDs to address this issue, including 114 ITDs with additional nucleotides of unknown origin located between the 2 copies of the repeat. We observed anatomy consistent with replication slippage, but could only identify the germline microhomology (1-6 bp) anticipated to prime such slippage in one-third of *FLT3*-ITDs. We explain the paradox of the "missing" microhomology in the majority of *FLT3*-ITDs through occult microhomology: specifically, by priming through use of nontemplated nucleotides (N-nucleotides) added by terminal deoxynucleotidyl transferase (TdT). We suggest that TdT-mediated nucleotide addition in excess of that required**

**for priming creates N-regions at the duplication junctions, explaining the additional nucleotides observed at this position. *FLT3*-ITD N-regions have a G/C content (66.9%), dinucleotide composition ( $P < .001$ ), and length characteristics consistent with synthesis by TdT. AML types with high TdT show an increased incidence of *FLT3*-ITDs (M0;  $P = .0017$ ). These results point to an unexpected role for the lymphoid enzyme TdT in priming *FLT3*-ITDs. Although the physiological role of TdT is to increase antigenic diversity through N-nucleotide addition during V(D)J recombination of IG/TCR genes, here we propose that illegitimate TdT activity makes a significant contribution to the genesis of AML. (*Blood*. 2019;134(25):2281-2290)**

## Introduction

*FLT3* encodes a receptor tyrosine kinase that governs proliferation of early progenitor hematopoietic cells.<sup>1</sup> *FLT3*-internal tandem duplications (*FLT3*-ITDs) are 1 of the commonest mutations found in acute myeloid leukemia (AML),<sup>1,2</sup> and also occur occasionally in acute lymphoblastic leukemia (ALL).<sup>3,4</sup> In AML, they confer a poor prognosis, particularly at higher allelic loads<sup>5,6</sup> arising from acquired isodisomy,<sup>7</sup> although development of effective *FLT3* inhibitors may improve this outlook. *FLT3*-ITDs vary in size from 3 to over 200 bp, invariably remain in-frame,<sup>6</sup> and target exon 14, resulting in constitutive kinase activation. Both start and end points are variable, with the majority of duplications considered unique.<sup>8</sup> Patients have been reported with up to 5 independent ITDs,<sup>9-11</sup> consistent with a mutator mechanism that promotes their formation.

Genomic rearrangements can be characterized through examination of junction sequences for recombinogenic motifs and

homology.<sup>12-15</sup> Despite the importance of *FLT3*-ITDs in AML, the mechanism underlying their formation has not been addressed beyond proposition of a replication error.<sup>16</sup> Replication-based models (microhomology-mediated replication dependent recombination [MMRDR]) encompass both simple replication slippage, typically involving change of a few bases at a repetitive sequence, and more complex models requiring breakage of the template strand.<sup>12,17</sup> If a replication model of *FLT3*-ITD genesis is to be accepted, then the microhomology required for priming slippage should be identified. Moreover, any proposed model must also account for the frequent insertion of nucleotides of unknown origin at the point of duplication, referred to as filler or foreign DNA.<sup>2,6,18</sup>

Failure to identify germline microhomology might argue against a replication-based origin for *FLT3*-ITDs. However, a suggestion for how microhomology can be created comes from the double-strand break repair process termed nonhomologous end joining

(NHEJ). Microhomology is not essential for NHEJ to join ends, but enhances ligation if present. In rare instances, this microhomology can be provided through nontemplated nucleotides (N-nucleotides) added by polymerase X family members (Pol  $\mu$ , Pol  $\lambda$ , or terminal deoxynucleotidyl transferase [TdT]).<sup>19,20</sup> As this homology is not present in the original sequences it usually escapes detection, and is therefore known as occult microhomology. However, occult microhomology remains poorly understood, with no identified role in MMRDR.

Here, we investigate the origin of *FLT3*-ITDs through examination of their molecular anatomy and breakpoint sequences. We propose a model of replication slippage primed extensively by occult microhomology synthesized by TdT, with a supporting role for germline microhomology. Longer TdT additions are revealed as N-regions at the duplication junction. We propose that the purposefully mutagenic enzyme TdT plays a significant role in the genesis of AML.

## Methods

### Patient cohort

Genomic DNA or complementary DNA *FLT3*-ITD sequences were identified using PubMed and Google publications from 1996 to 2013 (supplemental Table 1, available on the *Blood* Web site). Exclusion criteria were uncertain reference sequence, suspected contamination, and repeat publication. Numbering was converted to the *FLT3* coding reference sequence LRG\_457t1. Breakpoint positions were harmonized to the Human Genome Variation Society (HGVS) 3' rule, and to recognize triplications or intronic end points. Three hundred ITDs (AML, n = 273; chronic myelomonocytic leukemia, n = 6; myelodysplastic syndrome, n = 4; acute mixed lineage leukemia, n = 1; ALL, n = 16) were identified from 275 patients, representing 271 primary samples and 4 cell lines. For 19 of 114 ITDs with filler, the ITD position and filler length were available but the sequence of the filler was not stated. These ITDs were retained to preserve the proportion of patients with fillers. The sequence of 7 of 95 of the remaining fillers was partially deduced by reverse translation. Five additional *FLT3*-ITDs from ALL were considered only with respect to filler nucleotide incidence. French-American-British (FAB) type was available for 220 of 273 AML *FLT3*-ITDs.

### Statistics and homology searches

Cumulative binomial probabilities and Fisher exact test 2-sided *P* values were calculated at [www.danielsoper.com/statcalc](http://www.danielsoper.com/statcalc). The Spearman rank correlation was performed at [www.wessa.net/rwasp\\_spearman.wasp/](http://www.wessa.net/rwasp_spearman.wasp/). *P* values of <.05 were considered significant, unless adjusted by Bonferroni correction. Basic Local Alignment Search Tool (BLAST) searches were performed at National Center for Biotechnology Information (NCBI; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

## Results

### Molecular anatomy and incidence of *FLT3*-ITDs

The 300 *FLT3*-ITD sequences were divided according to molecular anatomy (Figure 1). Types A-C were duplications, each starting within exon 14, and ending within exon 14, intron 14, and exon 15, respectively (length range, 18-240 bp; mean, 57 bp). Type D showed insertion of filler sequence (length range,

3-18 bp; mean, 10 bp), without loss or gain of *FLT3* material, suggesting there is no critical region of *FLT3* that must be duplicated. Type E were deletions, although 3 of 12 appeared larger than wild type due to insertion of filler that exceeded the deleted material in length. A single-type E mutation was length neutral, with deletion of *FLT3* sequence balanced in length by gain of filler. Types F and G represented more complex events, including full or partial triplications. MMRDR could account for all types via backward slippage (duplication), forward slippage (deletion), and repeat slippage (more complex events including triplications). We therefore looked for the obligatory junction microhomology required to substantiate a replication-based model.

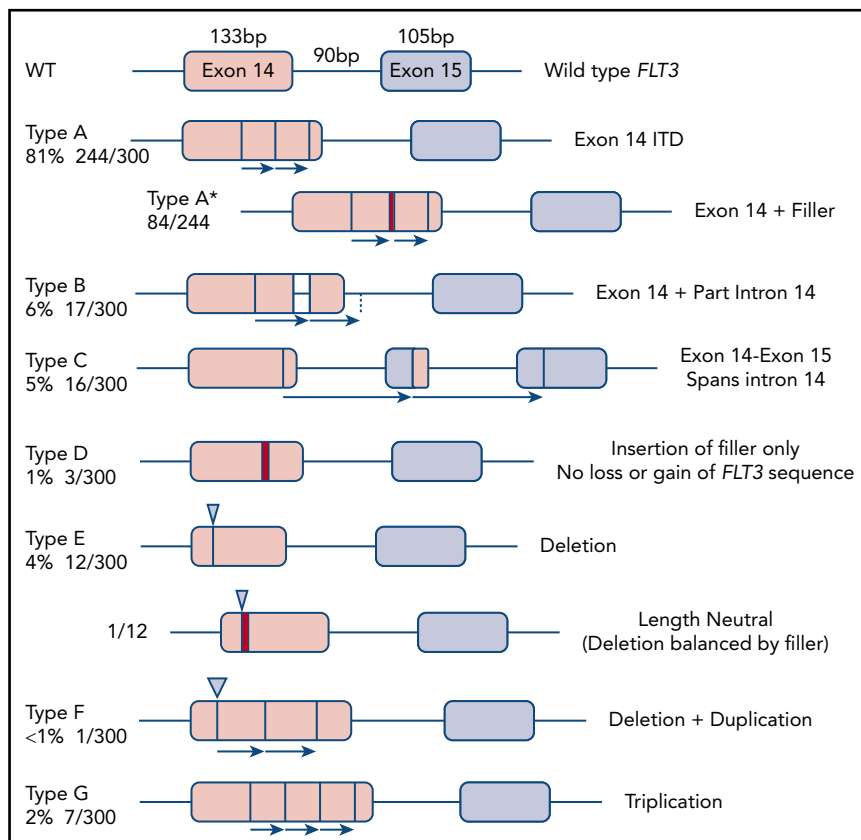
### Identification of 3 junction categories

Breakpoint examination did not identify universal flanking microhomology. Instead, 3 different junction categories were identified: 38% of ITDs showed the anticipated germline microhomology (1-6 bp), 38% showed addition of nontemplated filler nucleotides of unknown origin, and 24% lacked either microhomology or filler nucleotides (supplemental Table 2). Each junction type was identified across the mutational spectrum (duplications, deletions, and triplications), arguing against separate mechanistic origins for the different ITD types. All 3 junction types were identified in both myeloid- and lymphoid-derived ITDs (supplemental Table 2). The choice of junction type could vary between multiple independent ITDs identified from a single patient; 1 patient showed examples of each category. Junction category was not necessarily conserved at both junctions found in triplications. Each junction category is explored in the following 3 sections.

**Germline microhomology-mediated junctions** Figure 2A shows how microhomology can prime MMRDR, consistent with the 1 to 6 bp of microhomology observed flanking 38% of junctions (supplemental Tables 2 and 3). Although some microhomology would be expected by chance, there is an excess of microhomology at each length over that expected by chance (supplemental Figure 1). The observed excess increases with increasing microhomology length.

We asked whether identical ITDs existed, driven by longer microhomology blocks. The majority (52 of 72; 72%) of type A-D ITDs identified more than once in the cohort showed visible microhomology, with 17 different recurrent ITDs with visible microhomology observed (supplemental Table 4A). The 2 most highly recurrent were a 21-bp c.1780\_1800dup driven by 4-bp TGAT microhomology, and a 21-bp c.1784-1804dup with 3-bp TCA microhomology. These 2 duplications account for 6% of all ITDs and create a spike at 21 bp in the *FLT3*-ITD length distribution (supplemental Figure 2). There is a positive correlation between recurrence and microhomology length ( $\rho = 0.991$ ;  $P = <.0001$ ; Spearman rank correlation) (supplemental Figure 3). Start position c.1770 was also unexpectedly associated with recurrence in the absence of microhomology (supplemental Table 4B) for reasons explored further under "Triggering the replication error." Few other microhomology block pairings >3 bp exist that could cause in-frame *FLT3*-ITDs between the type A/B start and end point ranges (supplemental Table 5). Visible germline microhomology is therefore an important, but not universal, determinant of *FLT3*-ITD genesis, with many of the potential longer blocks of microhomology used.

**Figure 1. Molecular anatomy and incidence of different types of FLT3-ITD.** Only types A, B, and C meet the HGVS definition of a duplication, unless a filler (N-region, indicated by a red bar in the diagram and an asterisk (\*) in the nomenclature) is present at the junction. All types can have N-nucleotides, type D by definition.



**Filler junctions: identification of fillers as TdT-synthesized N-nucleotides** Insertion of filler nucleotides at the junction between the 2 repeats has been noted in approximately one-third of FLT3-ITDs.<sup>2,5,6,18,21</sup> We reasoned that understanding the origin of these nucleotides might help decipher FLT3-ITD genesis.

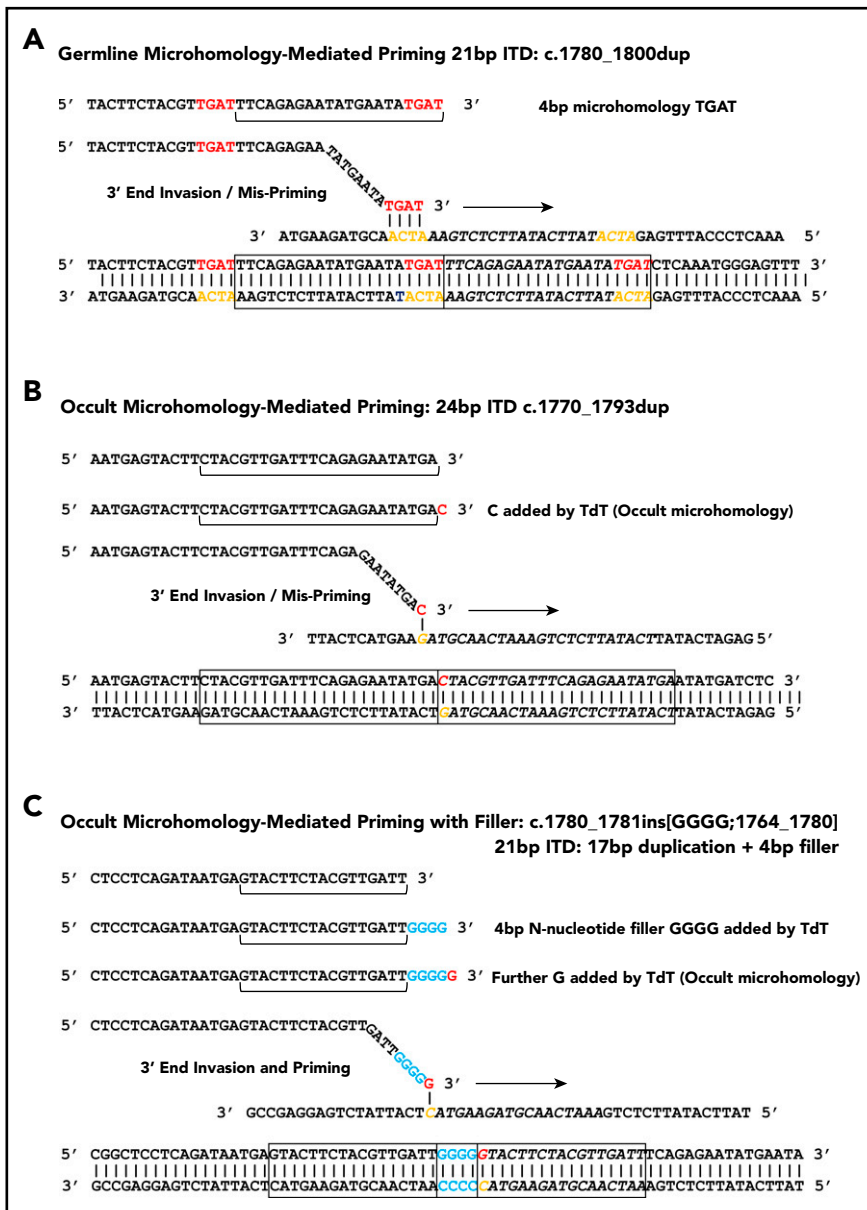
Determining the origin of short insertions is not necessarily straightforward as such fragments may provide chance matches, and even a short insertion can originate from 2 or more separate tracts through sequential replication errors.<sup>22</sup> We considered whether fillers had originated from exons 14 to 15 of FLT3,  $\pm 1$ -kb flanking DNA, aligning only fillers of  $\geq 7$  bp to reduce chance matches. Matches between 3 of 27 fillers were identified, each 7 to 8 bp, and attributed to chance. These results argue against a serial replication slippage origin within a single replication fork for the majority of filler sequences. Alternatively, the fillers could have originated elsewhere in the genome, through replication switching outside of the immediate replication fork or oligonucleotide capture.<sup>23</sup> BLAST searches were performed using filler sequences  $>20$  bp. Only 2 such fillers were available (28 bp and 36 bp), and neither was identified elsewhere in the genome as a single tract.

We alternatively considered whether the fillers might represent template-independent syntheses by a member of the polymerase X family, including TdT. TdT might seem an unlikely candidate as a myeloid mutagen, whereas both Pol  $\mu$  and Pol  $\lambda$  are widely expressed; TdT is regarded as a lymphoid-specific enzyme. However, up to 55% of AML patients are known to be TdT<sup>+</sup>,<sup>24</sup> and other patients may have downregulated TdT by

diagnosis. Moreover, only TdT invariably polymerizes in a template-independent fashion. We present below 5 lines of evidence suggesting that TdT is responsible for filler synthesis, although a minor contribution from other enzymes cannot be excluded.

First, we considered the G/C content of the filler. TdT is biased toward addition of G and C nucleotides.<sup>25-28</sup> The G/C percentage of N-nucleotides synthesized by TdT is typically 57% to 70%,<sup>25,29,30</sup> whereas there is no increase in the G/C content of fillers originating through other mechanisms.<sup>25</sup> For comparison, the G/C content of the human genome and FLT3 exon 14 are 41% and 38%, respectively. The G/C content of a total of 492 bp of FLT3-ITD filler nucleotides (from 95 ITDs) was 66.9% (supplemental Table 6), consistent with synthesis by TdT. This result also argues against involvement of Pol  $\mu$ , which displays a preference for deoxythymidine triphosphate and deoxycytidine triphosphate.<sup>27</sup>

Second, we considered the length and size range of FLT3-ITD fillers. The mean length of N-regions from recombination activating gene (RAG)-mediated events is 3 to 6 nt,<sup>25,29,30</sup> with a range of 1 to 13 nt at antigen receptor loci<sup>25,29</sup> (supplemental Table 6). Longer N-nucleotide tracts (up to 21 nt) are occasionally reported at RAG-mediated events at loci other than IGH and TCR, including deletions of BTG1 in B-ALL.<sup>30</sup> As TdT-mediated N-regions increase in length, they decrease in frequency.<sup>25</sup> In contrast, filler fragments from other origins routinely exceed 13 nt in length, and show a distinct peak of 1-nt additions with a broadly flat distribution from 2 to 40 nt.<sup>25</sup> Examination of 114 FLT3-ITD filler lengths showed a mean length



**Figure 2. Microhomology for priming *FLT3*-ITD replication slippage.** (A) Misalignment and priming using germline (visible) microhomology. The recurrent c.1780-1800dup primed by TGAT microhomology (red against yellow) is shown. (B) Occult (invisible) microhomology-mediated priming, where TdT adds a base to the misaligning 3' end. This base provides the homology for priming, but is not detected in the final sequence as it matches an existing base. The 24-bp c.1770-1793dup is shown. (C) Occult microhomology-mediated priming following addition of multiple N-nucleotides (blue). Only the terminal N-nucleotide matches the target strand, hence the previous N-nucleotides are visible as filler DNA between the 2 copies of the repeat. A 21-bp *FLT3*-ITD is shown (17-bp c.1764\_1780 duplication, with 4 additional N-nucleotides). These models are applicable to either simple replication slippage or break-induced models of MMRDR. The latter can occur following replication across a single-stranded nick and resulting replication fork collapse; fork collapse creates a single-ended double-stranded break, and the 3' end must invade the template strand to reinitiate replication.

of 5.6 nt with a range of 1 to 36 nt. Only 7 exceeded 13 nt. The distribution of filler lengths followed that expected for TdT-mediated N-nucleotides at antigen receptor loci, with a minimal number of longer tracts as observed at illegitimate targets (supplemental Figure 4).

Third, we examined dinucleotide content. Lacking a template, TdT stacks the incoming dNTP onto the existing base at the 3'OH, disposing to runs of homopurine or homopyrimidine.<sup>31</sup> Purine-purine (RR) and pyrimidine-pyrimidine (YY) dinucleotides are therefore overrepresented in TdT syntheses at both legitimate<sup>31</sup> and illegitimate<sup>32</sup> RAG-mediated events. Among the 8 RR and YY dinucleotides, the 4 homopolymers (GG, AA, TT, and CC) are the most highly overrepresented.<sup>31</sup> Conversely, RY and YR dinucleotides are underrepresented. We tested whether a total of 404 dinucleotides from *FLT3*-ITD fillers showed such biases. For example, G accounts for 33.7% of all *FLT3*-ITD filler nucleotides, and hence the dinucleotide GG would be expected at a frequency of  $0.337 \times 0.337 = 0.114$ . Approximately 46

occurrences would therefore be predicted within the 404 dinucleotides, but we observed a significantly higher figure of 80 ( $P < .001$ ). Table 1 shows that 5 of 8 of the RR and YY dinucleotides were observed at a higher level than anticipated, 3 of them significantly (AA, CC, and GG; all homopolymers). Conversely, 7 of 8 of the RY and YR dinucleotides were present at a lower level than anticipated, 3 of them significantly (CA, CG, and GC). These results strongly implicate TdT.

As a control, we analyzed 327 genomic dinucleotides spanning *FLT3* exons 14 to 15 (supplemental Table 7). Global genomic dinucleotide analysis in humans has previously shown a threefold to approximately fivefold depletion of CG dinucleotides due to deamination of 5-methylcytosine to thymine, and consequently smaller increases in both TG and CA dinucleotides.<sup>33</sup> TA is also globally depleted for reasons that are unclear.<sup>33</sup> In the control *FLT3* dinucleotides, both CG and TA were significantly depleted as expected (supplemental Table 7). The 2 most highly over-represented dinucleotides were TG and CA as predicted, but

**Table 1. Dinucleotide analysis of *FLT3*-ITD N-regions**

Dinucleotide	No. obs	No. exp	Obs/exp ratio	P	Significant at P = .05
<b>RR*</b>					
GG	80	45.88	1.74	<.001	Yes
AA	29	16.65	1.74	.003	Yes
GA	29	27.64	1.05	.409	No
AG	21	27.64	0.76	.921	No
<b>YY*</b>					
CC	82	44.26	1.85	<.001	Yes
TT	11	6.62	1.66	.063	No
CT	11	17.12	0.64	.954	No
TC	15	17.12	0.88	.722	No
<b>RY†</b>					
GC	16	45.06	0.36	<.001	Yes
GT	13	17.43	0.75	.172	No
AT	14	10.50	1.33	.891	No
AC	20	27.15	0.74	.092	No
<b>YR†</b>					
CG	22	45.06	0.49	<.001	Yes
CA	17	27.15	0.63	.023	Yes
TG	14	17.43	0.80	.217	No
TA	10	10.50	0.95	.519	No

Calculated using frequencies of individual nucleotides: G = 0.337, A = 0.203, T = 0.128, and C = 0.331.

Exp, expected; Obs, observed.

\*For the RR and YY dinucleotides, the P value reflects the cumulative binomial probability of the observed value exceeding or equaling the expected value.

†For RY and YR dinucleotides, the P value is for observed values less than or equal to the expected value.

neither P value reached significance (supplemental Table 7). The significant depletion of CA from the *FLT3*-ITD filler dinucleotide data set contrasts to its overrepresentation in germline DNA, suggesting that the former has not been exposed to evolutionary time scales, consistent with neosynthesis by TdT.

Fourth, we compared the incidence of filler in *FLT3*-ITDs with the level of TdT positivity across the AML FAB types. Levels of TdT are highest in immature leukemias, especially FAB M0, and low overall in M3 (acute promyelocytic leukemia [APL]).<sup>24,34-36</sup> FAB type was available for 211 *FLT3*-ITD<sup>+</sup> AMLs with type A-D duplications (supplemental Table 8). The incidence of filler in all FAB types was 35% (74 of 211), but was significantly higher in M0 (10 of 12; 83%) ( $P = .0017$ ; Fisher exact test; M0 against all other FAB types bar M3), and significantly lower in M3 (1 of 24; 4%) ( $P = .0009$ ; Fisher exact test; M3 against all other FAB types bar M0), consistent with synthesis by TdT. In APL, the low level of filler in *FLT3*-ITDs is reflected by significantly higher use of visible microhomology (17 of 24 types A-D), particularly  $\geq 2$  bp (50% [12 of 24], cf 16% [29 of 187] in all other duplications;  $P = .0003$ , Fisher exact test). This results in a reduced palette of *FLT3*-ITDs in APL patients, slanted toward use of recurrent microhomology-driven ITDs. In the 24 APL duplications, the level of the recurrent 21-bp c.1780-1800dup reached 25% (6 of 24), and there were 3 examples of an 18-bp c.1790-1807dup driven by AAT microhomology.

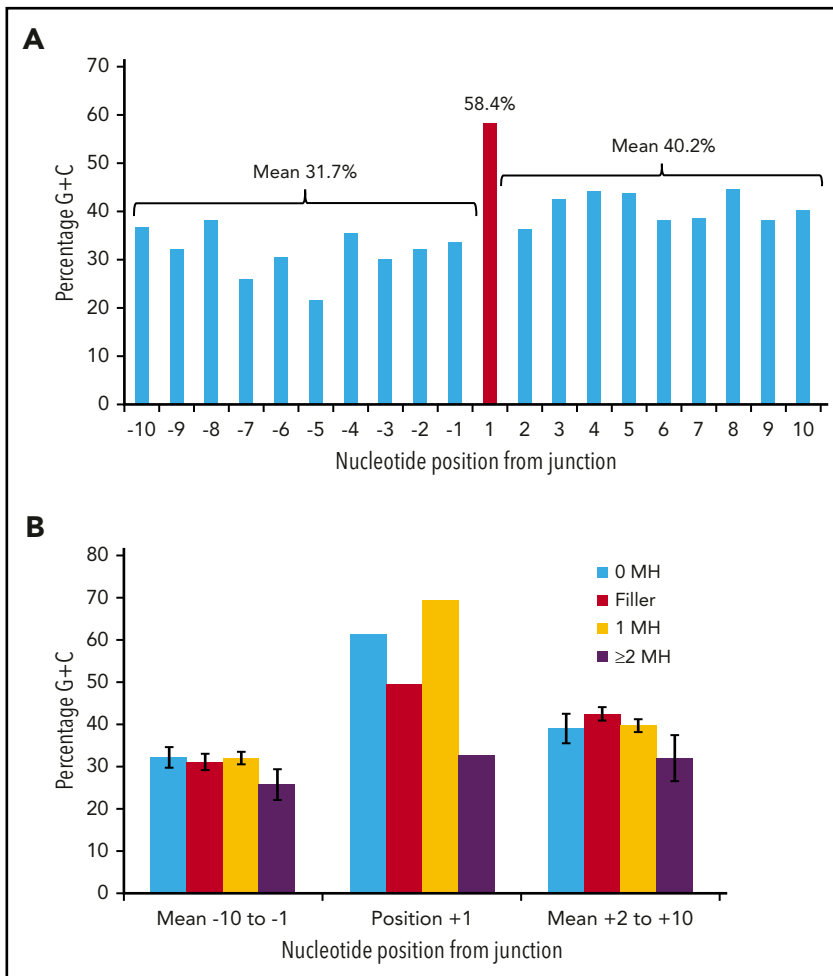
Finally, we determined whether there was a higher incidence of filler in *FLT3*-ITDs in ALL compared with AML because ALL shows a higher level of TdT positivity.<sup>36,37</sup> Filler incidence was

37% (104 of 284) and 63% (10 of 16) in AML and ALL, respectively ( $P = .061$ ; Fisher exact test) (supplemental Table 2). This result did not reach significance, possibly due to the small number of lymphoid *FLT3*-ITDs. The sequences of a further 5 *FLT3*-ITDs from patients with ALL were obtained from a reference not initially identified.<sup>38</sup> The revised data showed that 15 of 21 (71%) ALL ITDs had filler, significantly higher than myeloid *FLT3*-ITDs ( $P = .002$ ).

**Occult microhomology junctions** The data presented in the previous 2 sections suggest that TdT synthesizes the short junctional fillers, hereafter referred to as N-regions. We reasoned that TdT could also add 1 or more additional N-nucleotides, thereby creating the microhomology required for priming MMRDR. As these bases would by definition match the existing *FLT3* sequence, their addition would not be apparent by inspection of the final sequence (occult microhomology).

Occult microhomology could also extend to those *FLT3*-ITDs lacking either visible germline microhomology or N-regions. Here, we envisage that TdT adds 1 or more nucleotides, but the bases fully match the target sequence and hence leave no apparent evidence of TdT involvement. This would allow all *FLT3*-ITDs to be attributed to MMRDR, satisfying the requirement for microhomology with either preexisting germline or polymerase-generated microhomology (Figure 2B-C).

The *FLT3*-ITD data set provides an opportunity to test for occult microhomology. Any bases synthesized by TdT and used as microhomology for priming should still show a G/C bias. This



**Figure 3. Occult microhomology G/C content fingerprint in *FLT3*-ITD junctions.** (A) Percentage G/C content of the 20-nt positions flanking the *FLT3*-ITD repeat junction derived from 226 duplication junctions (N-nucleotide filler,  $n = 97$ ; no visible microhomology,  $n = 70$ ; 1 bp of microhomology,  $n = 59$ ). Junctions with  $\geq 2$ -bp microhomology were excluded, as their visible germline homology may be considered sufficient to prime MMRDR, and only a limited number of reintegration sites are available. A significant G/C spike is visible at position +1 (red). (B) Junction G/C content for individual junction categories. Error bars show standard error of the mean. There is no increase in G/C content for junctions with  $\geq 2$ -bp microhomology ( $n = 49$ ). MH, microhomology.

should manifest as a peak in the G/C content of the *FLT3* sequence (when measured across multiple ITDs) at a limited number of bases immediately flanking the repeat junction (minimally positions +1 or -1) (supplemental Figures 5 and 6). Such a peak should be visible in ITDs showing N-nucleotide addition or lacking visible microhomology, and in an unknown proportion of those patients showing 1 bp of microhomology (as this homology may be present by chance, with a further occult base added by TdT). In contrast, ITDs showing 2 or more bases of visible germline microhomology might not be expected to show such a peak.

Figure 3A shows the G/C percentage across the 20 junction positions for 226 duplication *FLT3*-ITDs with either N-regions, no germline microhomology, or 1 bp of germline microhomology. There is a spike of 58.4% G/C at position +1, in contrast to means of 31.7% G/C for positions -10 to -1, and 40.2% for positions +2 to +10, consistent with the concept of occult microhomology ( $P = .003$ , Fisher exact test; position 1 vs 108-bp start region). In contrast, there was no evidence of a G/C spike at position +1 in 49 ITDs showing  $\geq 2$  bp of microhomology (Figure 3B). Neither set of results was affected by removal of repeat ITDs (data not shown). We do not attribute these results to a general requirement for G/C-rich microhomology priming, as the overall G/C content of the 1- to 6-nt germline microhomology was only 25.1% (supplemental Table 3). These data suggest that occult microhomology generated by TdT may

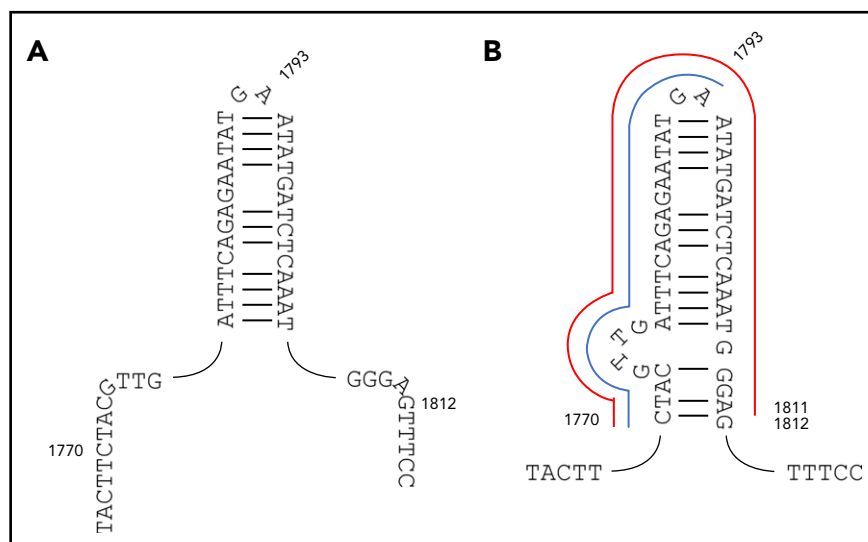
prime at least 80% of all *FLT3*-ITDs, excluding only those primed by  $\geq 2$  bases of germline microhomology, and that a single base of occult microhomology will suffice.

### Triggering the replication error

A 30-bp imperfect palindrome (c.1778-1807, centered 1792/3) was previously suggested to promote *FLT3* replication errors.<sup>16</sup> Our data set showed no evidence of overall increased use of the palindrome start, hairpin tip, or end points (supplemental Table 9; supplemental Figure 7). Moreover, ITD start and end points were both found within or either side of the palindrome (supplemental Table 9).

We show herein that the majority (52 of 72) of recurrent ITDs are driven by germline microhomology. We further reasoned that some of the remaining recurrent ITDs lacking germline microhomology might instead relate to secondary structure responsible for triggering replication slippage. Notably, a group of 12 ITDs all shared a common start point of c.1770, coupled to 1 of 3 end points (c.1793,  $n = 7$ ; c.1811,  $n = 3$ ; and c.1830,  $n = 2$ ) (supplemental Table 4B). Importantly, the c.1793 end point corresponds to the tip of the c.1778-1807 palindrome previously identified.<sup>16</sup> However, positions 1770 and 1811 do not correspond to the palindrome start and end points. We therefore propose an extended c.1770\_1812 palindrome (Figure 4). In this structure, the recurrent c.1770\_1793dup and c.1770\_1811dup

**Figure 4. Original and revised palindromes proposed to destabilize *FLT3* exon 14 and trigger MMRDR.** (A) The 30-bp c.1778-1807 imperfect palindrome originally proposed. (B) The revised c.1770\_1812 structure proposed here. The recurrent c.1770\_1793dup ITD corresponding to start → hairpin tip and start → end duplication is delineated in blue, and the recurrent c.1770\_1811dup start → end duplication in red. *FLT3*-ITDs start before or within this revised structure, and end within or after, with only rare exceptions.



ITDs correspond to start → hairpin tip and start → end duplications (Figure 4).

The significance of this structure was further assessed through comparison of the 278 *FLT3*-ITD type A-D start and end points to positions c.1770 and c.1812 (supplemental Table 9). One hundred forty-four of 278 start points (51.8%) occurred prior to the extended palindrome and the remainder within the palindrome. The last start position corresponded exactly to the last base of the palindrome, and no start points were observed after this position. A comparable converse pattern was observed for the end points; 2 of 278 (0.7%), 158 of 278 (56.8%), and 118 of 278 (42.4%) were observed before, within, and after the extended palindrome, respectively. The 2 end points found before the palindrome were close to the start point. Fifty of 278 ITDs (18.0%) started before and ended after the palindrome, showing that many ITDs span the palindrome, and that a breakpoint within the palindrome is not required. *FLT3*-ITDs, therefore, almost invariably start before or within the revised palindrome and end within or after the revised palindrome, confirming the significance of this structure. We suggest that this palindrome triggers MMRDR, with misalignment promoted by TdT.

## Discussion

The genetic landscape of AML includes many cytogenetic and molecular lesions now exploited for diagnostic, prognostic, monitoring, and therapeutic purposes.<sup>39</sup> Understanding how these mutations occur is also critical. For example, the recurrent translocations and inversions in AML arise following chromosome breakage and NHEJ-mediated repair. However, the breakage sites are not necessarily random, and in therapy-related AML often occur at topoisomerase II-binding sites.<sup>40</sup> Analysis of base substitutions in AML has identified just 2 causative mutational signatures, deamination of 5-methyl-cytosine, reflecting age,<sup>41</sup> and the ubiquitous mutational signature 5.<sup>42</sup> Such studies have left the genesis of *FLT3*-ITDs unresolved.

Here, we explore the origin of *FLT3*-ITDs in AML. We identify an unexpected role for TdT, proposing that the majority of *FLT3*-ITDs occur following addition of N-nucleotide(s) by TdT during

MMRDR. The physiological function of TdT is to add N-nucleotides to single-stranded DNA during V(D)J recombination to enhance antigen receptor diversity.<sup>28</sup> This ability to synthesize DNA in the absence of a template strand is unusual among polymerases,<sup>28</sup> with a clear potential for off-target activity to cause neoplasia. However, TdT mutagenesis has not previously been identified as carcinogenic. Expression of TdT is essentially restricted to lymphoid cells to help limit illegitimate activity, but its frequent expression in myeloid stem cells may risk the development of AML. DNA processes requiring extension from a 3' end, such as MMRDR, may be exquisitely sensitive to TdT activity as the 3' tip is critical for alignment.<sup>43</sup> TdT is believed to have ready access to free DNA ends in the nucleus.<sup>44</sup>

We did not set out to implicate TdT in leukemogenesis, but instead explore how *FLT3*-ITDs were primed. We initially identified 1 to 6 bp of germline microhomology suitable for priming a minority of ITDs. To explain how the remaining ITDs are primed, we propose that TdT adds short runs of N-nucleotides to the 3' end of the misaligning strand. The last nucleotide added is used for priming (occult microhomology), and any previous nucleotides appear as N-regions at the duplication junction. We suggest that this activity by TdT both permits priming at an incorrect site and inhibits alignment at the correct site. We support this model by identifying the unique footprint of TdT neosynthesis at *FLT3*-ITD repeat junctions. This footprint, conferred by the unusual properties of this polymerase, is most easily visualized within the N-regions. TdT's bias toward addition of G and C nucleotides results in an elevated G/C content, whereas its lack of a template creates a predisposition toward nucleotide stacking, resulting in a uniquely skewed dinucleotide composition. We are unaware of any other polymerases capable of similar nontemplated syntheses. Moreover, analysis of junction sequences from *BCOR*-ITDs (duplications of comparable size found in specific solid tumors that lack expression of TdT<sup>45</sup>) failed to reveal equivalent G/C-rich insertions (J.B., unpublished data, 10 August 2019). Additional to the N-region analysis, we identify a G/C-rich spike representing occult microhomology at a position immediately adjacent to the duplication junction. Furthermore, our results do not exclude the possibility that some ITDs apparently primed by

germline microhomology still occur following N-nucleotide addition by TdT, as such addition would still inhibit correct re-alignment. The failure to detect a G/C-rich spike at position +1 in ITDs with germline microhomology of  $\geq 2$  bases may reflect the limited availability of matching sites.

The occurrence of multiple *FLT3*-ITDs in a single patient suggests that *FLT3* exon 14 is prone to rearrangement. This may represent both the destabilizing effect of the extended palindrome and the action of TdT. Unselected out-of-frame *FLT3*-ITDs are also expected to occur. Although the effect of the palindrome may be uniform, progenitor cells transforming with high TdT might be predicted to have a high *FLT3*-ITD incidence, with an increased proportion showing N-regions. We support our model by showing an association across FAB types between TdT levels and the incidence of *FLT3*-ITD N-regions. This increase is also seen in *FLT3*-ITDs from ALL, where TdT levels are markedly high.

The varying incidence of *FLT3*-ITDs across AML cytogenetic types allows further examination of this hypothesis. The overall incidence of *FLT3*-ITDs in AML is 25%, but lower (7% to 9%) in patients with t(8;21) or inv(16), and lower again (2% to 4%) in patients with a complex karyotype.<sup>10,11</sup> In contrast, 90% of patients with t(6;9) *DEK-NUP214*<sup>3,42</sup> or t(5;11) *NUP98-NSD1*<sup>46,47</sup> are *FLT3*-ITD<sup>+</sup>. These differences could reflect preferential cooperation between mutations and/or differential activity of a mutator. Our model suggests that t(6;9) and t(5;11) stem cells might express significant levels of TdT. The t(6;9) is indeed recognized to arise from an early hematopoietic precursor associated with high levels of TdT,<sup>48</sup> although the TdT status of t(5;11) leukemias is unknown. Furthermore, in t(15;17) *PML-RARA* (APL), the incidence of *FLT3*-ITDs is 35% overall,<sup>9,10</sup> but ranges from 23% in hypergranular M3 to 65% in the rarer hypogranular M3v.<sup>10</sup> Hypergranular APL arises from a myeloid committed progenitor and typically lacks lymphoid antigens,<sup>49</sup> whereas M3v arises from an earlier CD34<sup>+</sup> progenitor and often coexpresses lymphoid antigens.<sup>49-52</sup> TdT is only rarely detected in hypergranular M3, but more commonly in M3v.<sup>51-53</sup> These data are consistent with the concept that M3v cases occur in a progenitor with high TdT and are more likely to acquire a *FLT3*-ITD. As the distinction between M3 and M3v was not clear throughout our cohort, this idea requires confirmation. Overall, APL patients may show a higher reliance on germline microhomology.

In our accompanying manuscript, we confirm and extend this AML TdT-mutator model to the genesis of *NPM1* mutations, which we also propose require priming by TdT.<sup>54</sup> We suggest that TdT may be a significant cause of AML, and that additional examples of TdT mutagenesis in select neoplasms will emerge.

## Acknowledgments

The authors thank David Schatz and Joydeep Banerjee (School of Medicine, Yale University) for helpful discussions, and Joanne Mason (West Midlands Regional Genetics Laboratory) for critical reading of the manuscript.

## Authorship

Contribution: J.B. conceived the study, assembled the *FLT3*-ITD cohort, analyzed data, performed statistical analyses, and wrote the draft manuscript; S.A.D., S.A., and M.J.G. supervised the project; and all authors provided intellectual input and revised and gave final approval to the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profile: M.J.G., 0000-0001-5112-2882.

Correspondence: Julian Borrow, West Midlands Regional Genetics Laboratory, Birmingham Women's and Children's NHS Foundation Trust, Mindelsohn Way, Edgbaston, Birmingham, B15 2TG, United Kingdom; e-mail: j.borrow@nhs.net.

## Footnotes

Submitted 14 April 2019; accepted 30 September 2019. Prepublished online as *Blood* First Edition paper, 17 October 2019; DOI 10.1182/blood.2019001238.

Data may be found in supplemental Figures 1-7 and supplemental Tables 1-9.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

## REFERENCES

- Gilliland DG, Griffin JD. The roles of FLT3 in hematopoiesis and leukemia. *Blood*. 2002; 100(5):1532-1542.
- Nakao M, Yokota S, Iwai T, et al. Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia*. 1996; 10(12):1911-1918.
- Xu F, Taki T, Yang HW, et al. Tandem duplication of the *FLT3* gene is found in acute lymphoblastic leukaemia as well as acute myeloid leukaemia but not in myelodysplastic syndrome or juvenile chronic myelogenous leukaemia in children. *Br J Haematol*. 1999; 105(1):155-162.
- Armstrong SA, Mabon ME, Silverman LB, et al. *FLT3* mutations in childhood acute lymphoblastic leukemia. *Blood*. 2004;103(9): 3544-3546.
- Thiede C, Studel C, Mohr B, et al. Analysis of *FLT3*-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood*. 2002; 99(12):4326-4335.
- Schnittger S, Bacher U, Haferlach C, Alpermann T, Kern W, Haferlach T. Diversity of the juxtamembrane and TKD1 mutations (exons 13-15) in the *FLT3* gene with regards to mutant load, sequence, length, localization, and correlation with biological data. *Genes Chromosomes Cancer*. 2012;51(10): 910-924.
- Griffiths M, Mason J, Rindl M, et al. Acquired isodisomy for chromosome 13 is common in AML, and associated with *FLT3*-itd mutations. *Leukemia*. 2005;19(12): 2355-2358.
- Yokota S, Kiyoi H, Nakao M, et al. Internal tandem duplication of the *FLT3* gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia*. 1997;11(10):1605-1609.
- Gale RE, Green C, Allen C, et al; Medical Research Council Adult Leukaemia Working Party. The impact of *FLT3* internal tandem duplication mutant level, number, size, and interaction with *NPM1* mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood*. 2008;111(5): 2776-2784.



10. Schnittger S, Schoch C, Dugas M, et al. Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood*. 2002;100(1):59-66.
11. Kottaridis PD, Gale RE, Frew ME, et al. The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood*. 2001;98(6):1752-1759.
12. Chen J-M, Cooper DN, Férec C, Kehrer-Sawatzki H, Patrinos GP. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*. 2010;20(4):222-233.
13. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009;10(8):551-564.
14. Jennes I, de Jong D, Mees K, Hogendoorn PCW, Szuhai K, Wuys W. Breakpoint characterization of large deletions in EXT1 or EXT2 in 10 multiple osteochondromas families. *BMC Med Genet*. 2011;12:85.
15. Vissers LELM, Bhatt SS, Janssen IM, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet*. 2009;18(19):3579-3593.
16. Kiyoi H, Towatari M, Yokota S, et al. Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product. *Leukemia*. 1998;12(9):1333-1337.
17. Chauvin A, Chen J-M, Quemener S, et al. Elucidation of the complex structure and origin of the human trypsinogen locus triplication. *Hum Mol Genet*. 2009;18(19):3605-3614.
18. Ma L, Feng D-R, Zhong M-H, et al. Analysis of ITD characteristics in acute myeloid leukemia patients with FLT3-ITD positive [in Chinese]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*. 2011;19(5):1161-1165.
19. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*. 2010;79:181-211.
20. Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J*. 2007;26(4):1010-1023.
21. Meshinchi S, Stirewalt DL, Alonzo TA, et al. Structural and numerical variation of FLT3/ITD in pediatric AML. *Blood*. 2008;111(10):4930-4933.
22. Sheen CR, Jewell UR, Morris CM, et al. Double complex mutations involving F8 and FUNDC2 caused by distinct break-induced replication. *Hum Mutat*. 2007;28(12):1198-1206.
23. Roth DB, Proctor GN, Stewart LK, Wilson JH. Oligonucleotide capture during end joining in mammalian cells. *Nucleic Acids Res*. 1991;19(25):7201-7205.
24. Drexler HG, Sperling C, Ludwig WD. Terminal deoxynucleotidyl transferase (TdT) expression in acute myeloid leukemia. *Leukemia*. 1993;7(8):1142-1150.
25. Roth DB, Chang XB, Wilson JH. Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. *Mol Cell Biol*. 1989;9(7):3049-3057.
26. Repasky JAE, Corbett E, Boboila C, Schatz DG. Mutational analysis of terminal deoxynucleotidyltransferase-mediated N-nucleotide addition in V(D)J recombination. *J Immunol*. 2004;172(9):5478-5488.
27. Domínguez O, Ruiz JF, Laín de Lera T, et al. DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J*. 2000;19(7):1731-1742.
28. Motea EA, Berdis AJ. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim Biophys Acta*. 2010;1804(5):1151-1166.
29. Bangs LA, Sanz IE, Teale JM. Comparison of D, JH, and junctional diversity in the fetal, adult, and aged B cell repertoires. *J Immunol*. 1991;146(6):1996-2004.
30. Waanders E, Scheijen B, van der Meer LT, et al. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLoS Genet*. 2012;8(2):e1002533.
31. Gauss GH, Lieber MR. Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol*. 1996;16(1):258-269.
32. Champagne DP, Shockett PE. Illegitimate V(D)J recombination-mediated deletions in Notch1 and Bcl11b are not sufficient for extensive clonal expansion and show minimal age or sex bias in frequency or junctional processing. *Mutat Res*. 2014;761:34-48.
33. Simmen MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics*. 2008;92(1):33-40.
34. Venditti A, Del Poeta G, Buccisano F, et al. Prognostic relevance of the expression of Tdt and CD7 in 335 cases of acute myeloid leukemia. *Leukemia*. 1998;12(7):1056-1063.
35. Huh YO, Smith TL, Collins P, et al. Terminal deoxynucleotidyl transferase expression in acute myelogenous leukemia and myelodysplasia as determined by flow cytometry. *Leuk Lymphoma*. 2000;37(3-4):319-331.
36. Kaleem Z, Crawford E, Pathan MH, et al. Flow cytometric analysis of acute leukemias. Diagnostic utility and critical analysis of data. *Arch Pathol Lab Med*. 2003;127(1):42-48.
37. Paietta E, Racevskis J, Bennett JM, Wiernik PH. Differential expression of terminal transferase (TdT) in acute lymphocytic leukaemia expressing myeloid antigens and TdT positive acute myeloid leukaemia as compared to myeloid antigen negative acute lymphocytic leukaemia. *Br J Haematol*. 1993;84(3):416-422.
38. Zhang J, Ding L, Holmfeldt L, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 2012;481(7380):157-163.
39. Grimwade D, Ivey A, Huntly BJP. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood*. 2016;127(1):29-41.
40. Cowell IG, Austin CA. Mechanism of generation of therapy related leukemia in response to anti-topoisomerase II agents. *Int J Environ Res Public Health*. 2012;9(6):2075-2091.
41. Alexandrov LB, Nik-Zainal S, Wedge DC, et al; ICGC PedBrain. Signatures of mutational processes in human cancer [published correction appears in *Nature*. 2013;502(7470):258]. *Nature*. 2013;500(7463):415-421.
42. Volinia S, Druck T, Paisie CA, Schrock MS, Huebner K. The ubiquitous "cancer mutational signature" 5 occurs specifically in cancers with deleted FHIT alleles. *Oncotarget*. 2017;8(60):102199-102211.
43. Viguera E, Canceill D, Ehrlich SD. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*. 2001;20(10):2587-2595.
44. Sandor Z, Calicchio ML, Sargent RG, Roth DB, Wilson JH. Distinct requirements for Ku in N nucleotide addition at V(D)J- and non-V(D)J-generated double-strand breaks. *Nucleic Acids Res*. 2004;32(6):1866-1873.
45. Ueno-Yokohata H, Okita H, Nakasato K, et al. Consistent in-frame internal tandem duplications of BCOR characterize clear cell sarcoma of the kidney. *Nat Genet*. 2015;47(8):861-863.
46. Hollink IHIM, van den Heuvel-Eibrink MM, Arentsen-Peters STCJM, et al. NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood*. 2011;118(13):3645-3656.
47. Akiki S, Dyer SA, Grimwade D, et al. NUP98-NSD1 fusion in association with FLT3-ITD mutation identifies a prognostically relevant subgroup of pediatric acute myeloid leukemia patients suitable for monitoring by real time quantitative PCR. *Genes Chromosomes Cancer*. 2013;52(11):1053-1064.
48. Oyarzo MP, Lin P, Glassman A, Bueso-Ramos CE, Luthra R, Medeiros LJ. Acute myeloid leukemia with t(6;9)(p23;q34) is associated with dysplasia and a high frequency of flt3 gene mutations. *Am J Clin Pathol*. 2004;122(3):348-358.
49. Grimwade D, Enver T. Acute promyelocytic leukemia: where does it stem from? *Leukemia*. 2004;18(3):375-384.
50. Lin P, Hao S, Medeiros LJ, et al. Expression of CD2 in acute promyelocytic leukemia correlates with short form of PML-RARalpha transcripts and poorer prognosis. *Am J Clin Pathol*. 2004;121(3):402-407.
51. Takenokuchi M, Kawano S, Nakamachi Y, et al. FLT3/ITD associated with an immature immunophenotype in PML-RARα leukemia. *Hematol Rep*. 2012;4(4):e22.
52. Chapiro E, Delabesse E, Asnafi V, et al. Expression of T-lineage-affiliated transcripts and TCR rearrangements in acute promyelocytic leukemia: implications for the cellular

target of t(15;17). *Blood*. 2006;108(10):3484-3493.

53. Rizzatti EG, Portieres FL, Martins SL, Rego EM, Zago MA, Falcão RP. Microgranular and

t(11;17)/PLZF-RARalpha variants of acute promyelocytic leukemia also present the flow cytometric pattern of CD13, CD34, and CD15 expression characteristic of PML-RARalpha gene rearrangement. *Am J Hematol*. 2004;76(1):44-51.

54. Borrow J, Dyer SA, Akiki S, Griffiths MJ. Molecular roulette: nucleophosmin mutations in AML are orchestrated through N-nucleotide addition by TdT. *Blood*. 2019;134(25):2291-2303.