RESEARCH BRIEF

Elevated Mutational Age in Blood of Children Treated for Cancer Contributes to Therapy-Related Myeloid Neoplasms

Eline J.M. Bertrums[1,2,3], Axel K.M. Rosendahl Huber[1,2], Jurrian K. de Kanter[1,2], Arianne M. Brandsma[1,2], Anaïs J.C.N. van Leeuwen[1,2], Mark Verheul[1,2], Marry M. van den Heuvel-Eibrink[1,4], Rurika Oka[1,2], Markus J. van Roosmalen[1,2], Hester A. de Groot-Kruseman[1], C. Michel Zwaan[1,3], Bianca F. Goemans[1], and Ruben van Boxtel[1,2]

ABSTRACT Childhood cancer survivors are confronted with various chronic health conditions like therapy-related malignancies. However, it is unclear how exposure to chemotherapy contributes to the mutation burden and clonal composition of healthy tissues early in life. Here, we studied mutation accumulation in hematopoietic stem and progenitor cells (HSPC) before and after cancer treatment of 24 children. Of these children, 19 developed therapy-related myeloid neoplasms (t-MN). Posttreatment HSPCs had an average mutation burden increase comparable to what treatment-naïve cells accumulate during 16 years of life, with excesses up to 80 years. In most children, these additional mutations were induced by clock-like processes, which are also active during healthy aging. Other patients harbored mutations that could be directly attributed to treatments like platinum-based drugs and thiopurines. Using phylogenetic inference, we demonstrate that most t-MN in children originate after the start of treatment and that leukemic clones become dominant during or directly after chemotherapy exposure.

SIGNIFICANCE: Our study shows that chemotherapy increases the mutation burden of normal blood cells in cancer survivors. Only few drugs damage the DNA directly, whereas in most patients, chemotherapy-induced mutations are caused by processes similar to those present during normal aging.

INTRODUCTION

Chemotherapy is the major treatment modality for cancer and has led to the curative treatment of an increasingly large number of patients (1). Most chemotherapeutic drugs act by fatally damaging or blocking DNA replication in malignant cells (2). However, chemotherapy can also be highly mutagenic to malignant cells that survive the cytotoxic effects (3).

Indeed, whole-genome sequencing (WGS) analyses of more than 3,500 cancer metastases revealed several mutational signatures, which are a direct consequence of exposure to specific chemotherapeutic drugs, such as platinum-based compounds and 5-fluorouracil (5-FU; refs. 4, 5). In addition, experimental strategies have defined and confirmed mutational signatures induced by chemotherapy *in vitro*, such as

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands. [2]Oncode Institute, Utrecht, the Netherlands. [3]Department of Pediatric Oncology, Erasmus Medical Center – Sophia Children's Hospital, Rotterdam, the Netherlands. [4]Utrecht University, Utrecht, the Netherlands.

E.J.M. Bertrums, A.K.M. Rosendahl Huber, and J.K. de Kanter contributed equally to this article.

Corresponding Author: Ruben van Boxtel, Princess Máxima Center for Pediatric Oncology, Heidelberglaan 25, 3584 CS, Utrecht, the Netherlands. Phone: 0031 (0)889727272; E-mail: r.vanboxtel@prinsesmaximacentrum.nl

temozolomide, platinum-based compounds, cyclophospha-mide, 5-FU, and 6-mercaptopurine (6-MP; refs. 6–8).

During chemotherapeutic treatment, the toxic effects on normal tissues are often dose limiting, with the hematopoietic system being especially vulnerable (9). Besides these acute toxic effects, cancer survivors are confronted with a variety of chronic health conditions later in life as a result of chemotherapy, such as cardiac problems, infertility, and secondary malignancies (10–12). Childhood cancer survivors especially suffer from these long-term adverse effects, which collectively resemble accelerated aging (12). Indeed, although the long-term survival rate of children treated for cancer approaches 80%, their bodies are still in development during treatment, and survivors can develop adverse effects even decades after their initial diagnosis (12). Chemotherapy-induced mutagenesis and clonal expansions in healthy tissues may be responsible for inducing some of these long-term adverse effects, in particular secondary malignancies. Thus far, the impact of chemotherapy exposure in normal blood has been inferred from mutational landscapes of therapy-related myeloid neoplasms (t-MN; refs. 5, 13) as well as of cases of clonal hematopoiesis (CH) in exposed patients (14). t-MN comprises two disease types, namely, therapy-related acute myeloid leukemia (t-AML) and therapy-related myelodysplastic syndrome (t-MDS; refs. 15, 16). In t-MN genomes, high numbers of clonal platinum- and thiopurine-induced mutations could be observed, which indicated that in these cases clonal expansion of the hematopoietic cell founding the leukemia started after the initiation of exposure (5). Indeed, cancer therapy preferentially selects for cells that harbor mutations in DNA damage response genes, such as *TP53*, *CHEK2*, and *PPM1D*, ultimately resulting in CH and t-MN (5). In adults, preleukemic CH can often already be observed at the time of the primary cancer diagnosis and before exposure to treatment (5, 17). However, a recent report on three pediatric neuroblastoma patients showed that most mutations present in the CH clone in these children could be linked to platinum-associated signatures (16). These previous studies focused on studying mutational landscapes of preleukemic clonal expansions and/or t-MN. However, the mutational consequences of chemotherapy exposure in normal hematopoietic stem and progenitor cells (HSPC) and how this relates to mutations observed in t-MN are unknown. In addition, the origin of t-MN with respect to the timing of chemotherapy exposure in children seems different from adults yet remains understudied.

Here, we characterized the mutational consequences of chemotherapy in normal HSPCs of pediatric cancer patients before and after receiving treatment. We included patients who developed t-MN to determine how mutagenesis and clonal evolution in healthy blood contribute to the development of secondary malignancies. We found that the mutation burden of normal HSPCs was increased after chemotherapy. Remarkably, chemotherapy-associated mutagenesis in most patients was caused by processes resembling those active during healthy aging. Only few compounds, such as platinum-based drugs and thiopurines, had a direct mutagenic impact. In contrast to the effect of thiopurines, our data suggest that the effect of platinum-based drugs is independent of cell division. By combining mutational signature analysis and phylogenetic inference, we demonstrate that both induction of driving events and subsequent selection occur during chemotherapeutic exposure, which can ultimately lead to t-MN.

RESULTS

Cataloging Somatic Mutations in Chemotherapy-Exposed HSPCs of Children

To assess the mutational consequences of chemotherapy exposure in normal tissues, we determined the somatic mutations present in the HSPCs of children before and after receiving cancer treatment. We focused on the hematopoietic system, since it is highly sensitive to chemotherapy exposure (9) and because the lifelong mutation accumulation in healthy HSPCs has been determined (18–20). In total, we assessed 24 patients, who underwent treatment for different pediatric cancer types (Fig. 1A). Of these, 19 patients developed t-MN (Fig. 1B), of which 18 were t-AML and one was t-MDS (UPN012). A detailed list of diagnosis, age, and treatment information for all patients is provided in Supplementary Table S1. The latency between the start of chemotherapy and t-MN onset among patients ranged from 1.1 to 10 years (Fig. 1B).

Depending on the available patient material, we performed WGS on individual HSPCs at the time of the primary diagnosis (DX1), after complete remission of this first cancer (FU; post-treatment), and at the time of t-MN diagnosis (DX2; Fig. 1B and C). In addition, we sequenced the t-MN genomes. Of note, the t-MN of patient UPN021 was excluded because of poor sequencing quality (Methods; Fig. 1A; Supplementary Table S2). We sorted single HSPCs using flow cytometry (Methods), clonally expanded these cells *in vitro* to obtain sufficient DNA for WGS (Fig. 1C; Supplementary Fig. S1), and cataloged all clonal somatic mutations per HSPC (Methods; Supplementary Fig. S2A).

In total, we assessed 135 HSPC clones (28 before and 107 after treatment) and identified 46,831 single base substitutions, 2,658 small insertions and deletions (indel), and 346 double-base substitutions. Most of the t-MN cases were driven by gene fusions (16 out of 18 cases; Fig. 1D). Of these, 11 patients (69%) harbored an MLL fusion (*KMT2A* rearrangement), two patients (13%) a *RUNX1* fusion, and two patients (13%) a MECOM fusion (Fig. 1D; Supplementary Fig. S2B). Five MLL breakpoints were present in the 11-bp topoisomerase II inhibitor (TOP2i)–related hotspot positioned within the MLL breakpoint cluster region, and all of these patients received TOP2i (refs. 21, 22; Supplementary Fig. S2C). In four patients, the MLL fusion was the sole t-MN driver. The frequency of MLL fusions in these pediatric t-MN patients is considerably higher than previously reported in primary pediatric and infant AML (13% and 38%, respectively; ref. 23) and in adult t-AML (23%; ref. 24). Nonetheless, we did identify genetic similarities between primary pediatric AML (pAML) and t-AML, such as a higher prevalence of *RAS* mutations in MLL-rearranged (MLL^r) AML (Fig. 1D). Finally, despite the use of alkylating agents in many of our patients and its previously described association with monosomy 5(q) and 7(q) (13, 24), only three t-MNs (17%) presented with this aberration. The lack of t-MN cases with these monosomies in our cohort is explainable, as we predominantly assessed t-AML (17 out of 18 cases). Indeed, it is known that monosomy 5 or 7 is more often associated with a disease that is preceded by t-MDS (15, 25).

Chemotherapy Exposure Increases the Mutation Burden *In Vivo*

During healthy life, HSPCs accumulate mutations in a linear fashion at a rate of 14 to 15 single base substitutions
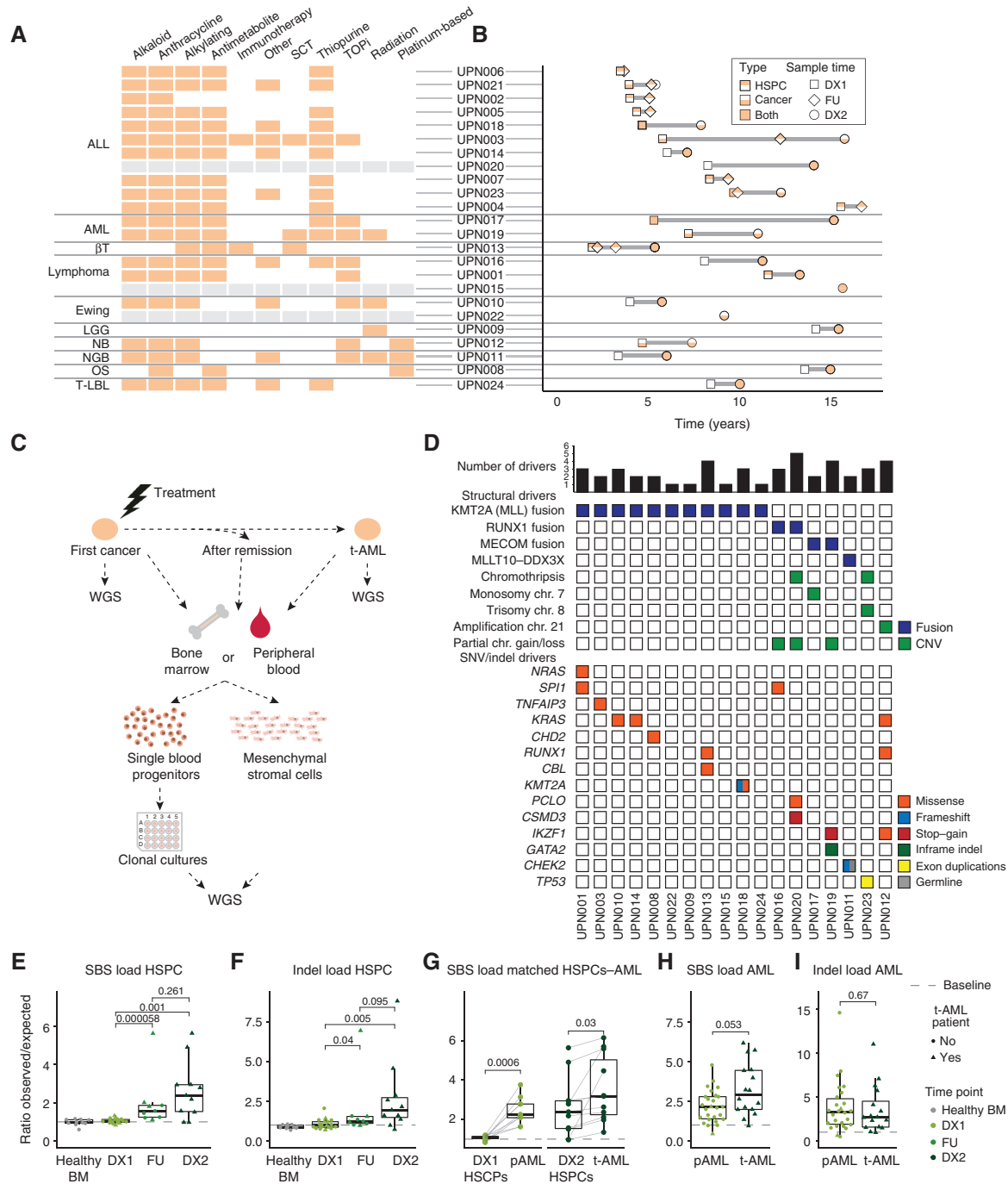
Figure 1. HSPCs and t-MN blasts in a chemotherapy-treated patient cohort display an increased mutation load. **A,** A table depicting the different treatment categories that patients from each tumor type received. Rows per patient match with time lines in **B**. ALL, acute lymphoblastic leukemia; βT, beta-thalassemia; LGG, low-grade glioma; NB, neuroblastoma; NGB, neuroganglioblastoma; OS, osteosarcoma; SCT, stem cell transplantation; T-LBL, T-cell lymphoblastic lymphoma; TOPi, topoisomerase inhibitor. **B,** The per-patient time lines of sample collection and the type of material that was sequenced. **C,** A schematic overview of the experimental setup of this study. In short, biopsies at the time of the primary cancer, follow-up (after remission), and t-MN were collected. Blasts were enriched by FACS, mesenchymal stromal cells were expanded *in vitro*, and both were sequenced in bulk. FACS was also used to sort single HSPCs into 96-well plates, which were then clonally expanded to obtain sufficient DNA for WGS, after which the mutation catalogs of the original HSPCs could be obtained. **D,** The clonal driver events were identified in the t-MN samples. The bar plots on top indicate the number of driver events identified in each sample. chr., chromosome; CNV, copy-number variation; SNV, single-nucleotide variant. **E,** The mutation load of single base substitutions (SBS) per time point per HSPC, normalized to the HSPC baseline consisting of healthy bone marrow (BM) HSPCs. The HSPCs were averaged per patient per time point. Two-sided Wilcox test, FDR-corrected. Here and in all other figures, the box plots depict the median (center line), 25th and 75th percentiles (box), and the largest values no more than 1.5* the interquartile range (whiskers). **F,** similar to **E** but for indels in HSPCs. **G,** The mutation load of single base substitutions of primary and therapy-related AML blasts and the mean value of matched HSPCs from the same time point per patient. Connecting lines represented matched AML and HSPCs. Two-sided paired *t* test, FDR-corrected. **H,** Similar to **E** but for t-AML blasts. **I,** Similar to **F** but for t-AML blasts.

and about one indel per year (18, 19). We compared the mutation burden of the pre- and posttreatment HSPCs to this baseline to correct for age-related mutation accumulation (Fig. 1E and F; Supplementary Fig. S2D). The pretreatment HSPCs showed a mutation burden that was similar to this baseline. In contrast, the posttreatment HSPCs showed an increased number of single base substitutions and indels, corresponding to mutational ages (26) up to 94 years for single base substitutions [mean increase of 16; 95% confidence interval (CI), 13–19] and 90 years for indels (mean increase of 16; 95% CI, 12–19; Fig. 1E and F). Furthermore, the t-MN blasts showed an increased mutation load compared with the baseline but also compared with matched posttreatment HSPCs (Fig. 1G). This latter observation indicates that at the time of t-MN initiation, the leukemic cell of origin suffered more from mutagenesis than the average exposed HSPC. Nonetheless, similar to previous reports, the mutation load in t-AML blasts was only slightly higher than *de novo* pediatric AML, and this difference was not significant in our analysis ($P = 0.053$, Fig. 1H and I; refs. 13, 16, 19). Together, our data suggest that for pediatric AML to arise, albeit related to treatment or naïve, a minimal mutation burden seems to be required, which is higher than the average number of somatic mutations observed within the healthy HSPC population.

Mutational Signatures Induced by Chemotherapy Exposure

To identify the processes that underlie the increased mutation burden in posttreatment HSPCs and t-MN blasts, we analyzed mutation spectra and signature contributions (27). As expected, the mutation spectra of pretreatment HSPCs were similar to treatment-naïve cells and could be predominantly explained by the HSPC signature (18, 20, 28) and to a lesser extent Catalogue Of Somatic Mutations In Cancer (COSMIC) signatures SBS1 and SBS5 (ref. 29; Fig. 2A). Indeed, the contribution of these three clock-like signatures increases with age in healthy HSPCs (18, 19). In the posttreatment HSPCs and t-MNs, we additionally identified SBS31 and SBS87, which are caused by platinum-based drugs and thiopurines, respectively (7, 30). SBS31 and DBS5, also caused by platinum-based drugs (30), were present in all cells of platinum-exposed patients ($N = 3$; Fig. 2A and B; Supplementary Fig. S3A and S3B). In t-MN patients that received thiopurine therapy ($N = 9$), all but one t-MN genome harbored SBS87 mutations (Fig. 2A and C). In contrast to platinum-based exposure, only some posttreatment HSPCs displayed SBS87, indicating that thiopurine exposure is not always mutagenic to all cells. Furthermore, we identified three novel signatures that likely represent distinct mutational processes, as they could not be accurately decomposed by three or fewer existing signatures (Fig. 2A and D; Supplementary Fig. S3C–S3F; ref. 31). Of these, SBSA was recently shown to be caused by the antiviral nucleoside analogue ganciclovir (32) and was present in the previously reported t-MN of patient UPN003 after exposure (Supplementary Fig. S4A and S4B). The other two signatures (SBSB and SBSC) were observed in all posttreatment samples of single patients (UPN015 and UPN013, respectively). SBSB was associated with single thymidine deletions at short T-repeats, whereas SBSC samples harbored large deletions at locations with microhomology (Supplementary Fig. S5A). The

T>N and C>T changes that contributed to SBSC displayed a wide sequence context preference for guanine at the −2 position of the mutated base (Supplementary Fig. S5B). Although for patient UPN015 no treatment data were available, patient UPN013 received the alkylating drugs thiotepa and treosulfan as part of a conditioning treatment for hematopoietic stem cell transplantation. This patient was treated with multiple hematopoietic stem cell transplantations—which partly failed—to treat beta-thalassemia (Supplementary Table S1). Posttreatment HSPCs of this patient displayed an increasing contribution of SBSC mutations after each consecutive round of transplantation (Fig. 2A). This suggests a causative role for the conditioning treatments in inducing SBSC mutations. Indeed, a pretreatment HSPC of this patient did not harbor SBSC mutations. We treated cord blood–derived HSPCs with thiotepa and treosulfan *in vitro*, after which we performed WGS on exposed cells, as described previously (33). The resulting profiles showed similarities to SBSC but could not fully explain the signature (Supplementary Fig. S5C and S5D).

Although the mutation burden of most assessed t-MN cases was higher than the healthy baseline (Fig. 1H), only half of these cases displayed evidence of direct mutagenesis by chemotherapeutic compounds. To study the likelihood that leukemic driver mutations were directly caused by such compounds, we used a previously established method to calculate the probability that a certain mutation can be attributed to a signature (34, 35). This analysis showed that only seven of 17 identified driving single base substitutions had high probability (>70%) to be attributed to a treatment-related signature, whereas nine drivers were best explained by clock-like signatures (Fig. 2E). Therefore, in contrast to the identified fusion genes, which were likely the result of TOP2i, additional driver mutations in t-MN could only partly be explained by direct mutagenic consequences of chemotherapy treatment.

Direct and Indirect Induction of Mutations after Chemotherapy

Although we identified chemotherapy-associated signatures in a subset of posttreatment HSPCs and t-MN samples, most samples showed mutation spectra similar to pretreatment HSPCs, which could be fully explained by clock-like signatures (cosine similarity 0.97; Supplementary Fig. S6A and S6B). Given this difference, we defined two categories: t-HSCPs, which have a spectrum that is >20% explained by the contribution of a treatment-related signature; n-HSPCs, which have a normal spectrum that is similar to healthy cells. Surprisingly, the mutation burden not only in t-HSPCs but also in n-HSPCs was elevated compared with age-matched treatment-naïve HSPCs (Fig. 3A and B). The mutation load increase of n-HSPCs at FU and DX2 was 1.47-fold (95% CI, 1.25–1.69) and 1.91-fold (1.34–2.47) compared with the baseline, respectively. Only the posttreatment HSPCs of patients UPN017 and UPN020 did not display elevated mutation burdens compared with the healthy baseline (Fig. 2A; Supplementary Fig. S6C). The absence of an increased mutation burden in the HSPCs of these patients was not explained by a lack of exposure to chemotherapy, as UPN017 was treated for pAML and UPN020 for acute lymphoblastic leukemia (ALL). Indeed, the t-MN blasts of UPN020 harbored SBS87
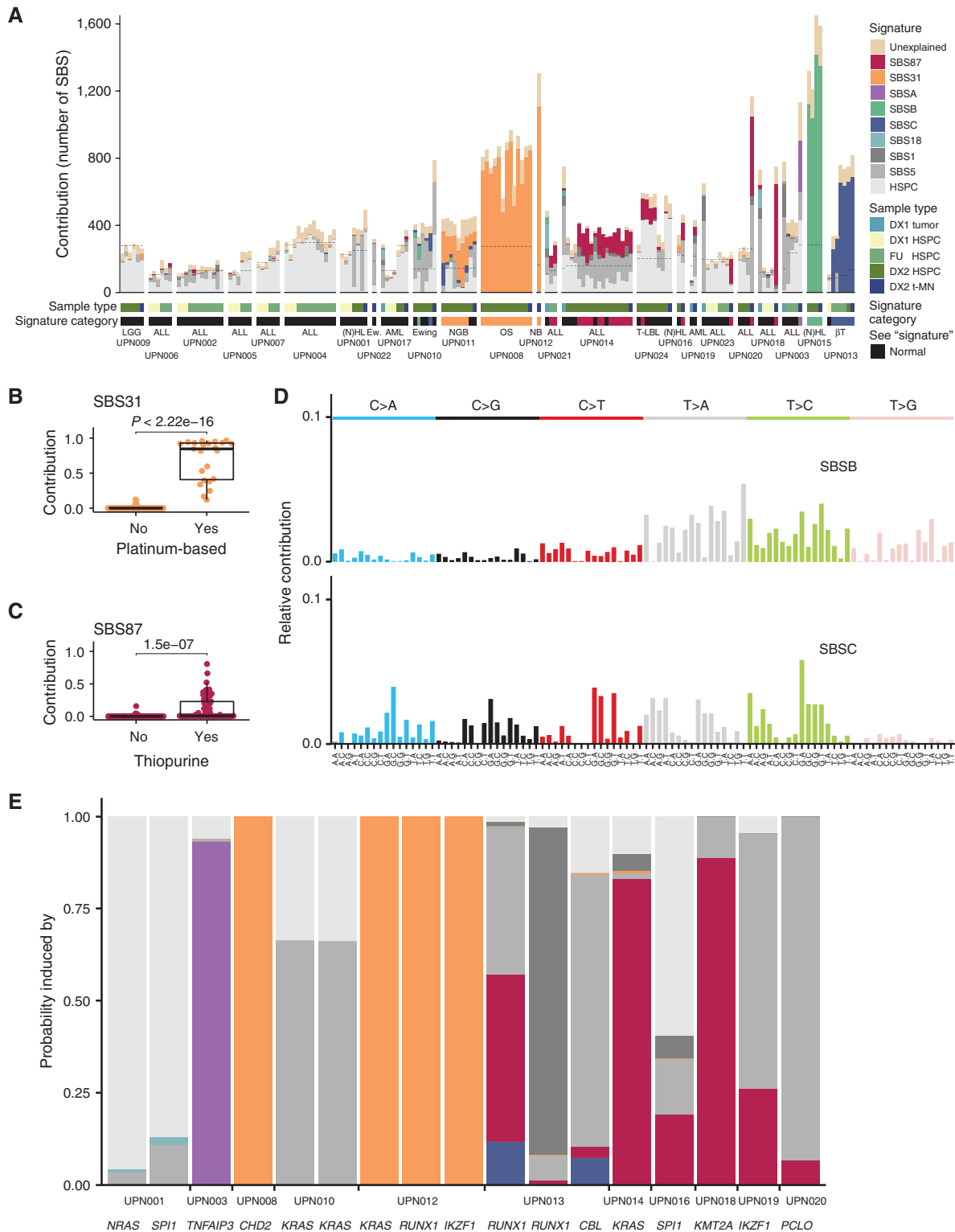
Figure 2. Thiopurines and platinum-based compounds have direct mutagenic effects in posttreatment HSPCs. **A,** The contribution of each SBS signature to each sample, as obtained after refitting of signatures that were extracted by non-negative matrix factorization. The first row of bars below the plot indicates the timing of each sample. The second row indicates which treatment signature had more than 20% contribution in each sample; these HSPCs were termed t-HSPCs. HSPCs without more than 20% contribution of any of these signatures were termed n-HSPC. ALL, acute lymphoblastic leukemia; βT, beta-thalassemia; Ew., Ewing; LGG, low-grade glioma; NB, neuroblastoma; NGB, neuroganglioblastoma; (N)HL, (non-)Hodgkin lymphoma; OS, osteosarcoma; SBS, single base substitutions; T-LBL, T-cell lymphoblastic lymphoma. **B,** The contribution of SBS31 to samples treated or not treated by platinum-based drugs (two-sided Wilcox test). **C,** The contribution of SBS87 to samples treated or not treated by thiopurines (two-sided Wilcox test). **D,** The 96-trinucleotide single base substitution profiles of SBSB and SBSC. **E,** The probability that different driver mutations were caused by treatment-related or clock-like signatures.
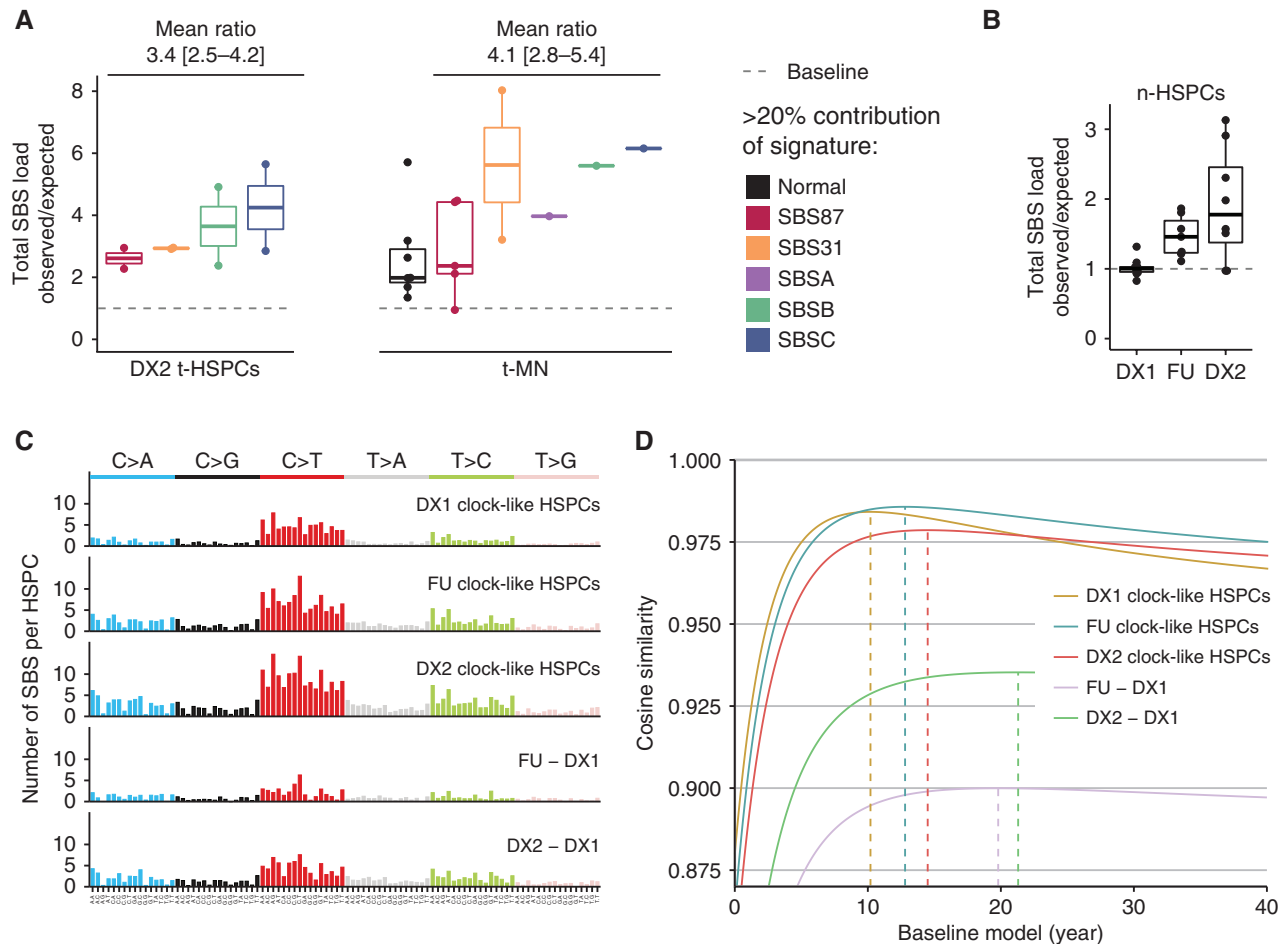
Figure 3. Chemotherapies induce a direct and indirect increase of mutations. **A,** The mutation load of single base substitutions (SBS) for DX2 HSPCs and t-MN grouped by signature category if 20% or more mutations can be explained by signature SBS87, SBS31, SBSA, SBSB, or SBSC. **B,** The baseline-normalized mutation load of the DX1, FU, and DX2 n-HSPCs. **C,** The average 96-trinucleotide SBS profile of n-HSPCs per time point. "FU − DX1" depicts the average profile of the FU n-HSPCs with the DX1 n-HSPCs subtracted. "DX2 − DX1" is similar but for DX2 with DX1 subtracted. **D,** The cosine similarity of each profile in **C** with a model of the baseline. Each dotted line indicates the age of the baseline at which the correlation of that profile is most similar to the model.

mutations as well as an increased number of indels (Fig. 2A; Supplementary Fig. S3A). Interestingly, both patients showed a longer latency time to t-MN development compared with the rest of the cohort. The posttreatment HSPCs of UPN017 were isolated with the longest latency after the end of treatment (9.3 years vs. 0.2–4.7 years in the rest of the cohort). The end of treatment for UPN020 was unknown, but the latency time between primary cancer and t-MN was 5.8 years (Supplementary Table S1). These observations may suggest that years after treatment, HSPCs with a lower mutation load may preferentially contribute to hematopoiesis, similar to what has been reported in bronchial cells of ex-smokers (36). Indeed, the risk for developing t-MN after treatment of a solid tumor peaks at 2 years after treatment and has been reported to return to a baseline population risk in 10–15 years (37).

As SBS1 and SBS5 are mainly active during fetal hematopoietic development and the HSPC signature postnatally, the mutation spectrum of healthy HSPCs changes during the first years of life (18, 19, 29, 38, 39). To estimate if similar mutations accumulated during treatment as during aging, we determined the similarity between the mean SBS profile

of the n-HSPCs to the mutation accumulation baseline in healthy HSPCs (Fig. 3C and D). Compared with the profile of DX1 HSPCs, the n-HSPC profile at the time of FU1 and DX2 was more similar to the profile of older, healthy HSPCs. This observation was even more apparent for the profile of additional mutations in posttreatment HSPCs, showing that not only the mutation burden but also the mutation spectra of these HSPCs are similar to those of healthy individuals of an older age. These analyses suggest that chemotherapy exposure can lead to an increased mutational age of normal HSPCs *in vivo*. Importantly, this indirect mutagenic effect of chemotherapy exposure may contribute to the accumulation of t-MN driver mutations (Fig. 2E).

Phylogenetic History of t-MN

To time the mutagenic effect of chemotherapy during t-MN development, we delineated the phylogenetic history using somatic mutations that were shared between cells of the same patient (18, 20, 40). Although most posttreatment HSPCs did not harbor cancer driver mutations, we identified some HSPCs with structural rearrangements. Three HSPCs in two

patients harbored genetic fusion genes that were not shared by the t-MN (Supplementary Fig. S6D), and, importantly, in four t-MN cases we identified phenotypically HSPC-like cells, which shared the MLL rearrangement with the t-MN blasts (Fig. 4). For patient UPN008, all 12 posttreatment HSPC clones and the t-MN blasts predominantly harbored SBS31 mutations (Fig. 4A and B). Indeed, this patient was initially treated for osteosarcoma, which included platinum-based drugs (Supplementary Table S1). We identified six MLL^r HSPCs that also shared all the clonal somatic mutations present in the t-MN (α branch) and additionally harbored 71 to 101 unique mutations each (β branches), of which some were subclonally present in the bulk t-MN sample (Fig. 4A). These unique mutations were predominantly attributed to SBS5 and SBS1, with low similarity to SBS31 (Fig. 4B). This indicated that they were acquired after cisplatin exposure and thus that the t-MN expanded after cisplatin treatment. Based on the mutation data, these MLL^r "HSPCs" were part of the leukemic blast population despite their HSPC-like phenotype and lack of CD33 expression, which was the blast-defining marker that was used to sort the blast population. This phenomenon resembles a previous report of MLL^r infant ALL, where an HSPC-like blast population was reported that lacked expression of the characteristic B-cell marker CD19 (41). We also observed this phenomenon in patient UPN024, treated for T-cell lymphoblastic lymphoma (T-LBL), where one HSPC-like cell shared all mutations with the t-MN, thus genetically characterizing it as a leukemic cell. Interestingly, similar to UPN008, we found an MLL rearrangement as the sole genetic driver of the t-MN of UPN024. This rearrangement was shared by an additional four HSPC-like cells that did not harbor all other t-MN mutations in their genomes (Fig. 4C and D). This observation suggests that additional nongenetic hits are required for full malignant transformation.
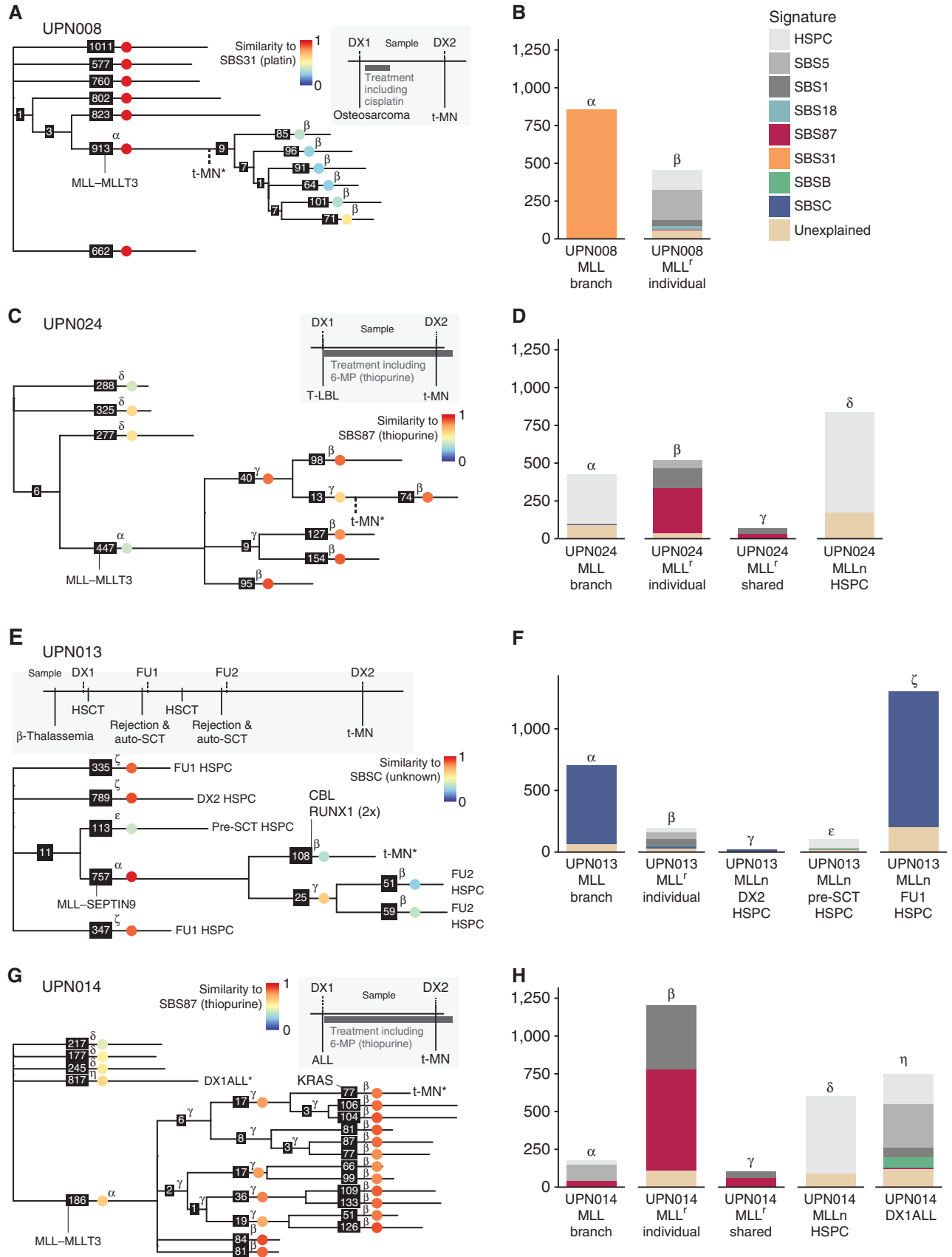
In patients UPN013 and UPN014 (Fig. 4E–H), the t-MN had additional driver mutations that were not shared with the MLL^r HSPCs (RUNX1/CBL and KRAS^{G12A}, respectively), indicating these were preleukemic cells that separated from the t-MN lineage before the leukemic cell of origin started expanding. In patient UPN013, treated for beta-thalassemia (see above), the mutations shared between MLL^r HSPCs and t-MN blasts were all attributed to SBSC [Fig. 4E (α branch)]. In contrast, the clone-specific mutations (β branches) were mostly attributed to SBS1 and SBS5, indicating the MLL^r HSPCs separated from the t-MN lineage at the end of mutagen exposure, similar to UPN008 (Fig. 4F). In patient UPN014, who developed t-MN after a first diagnosis of ALL, the 13 MLL^r HSPCs shared 186 single base substitutions/indels with the t-MN [Fig. 4G (α

branch)]. The MLL^r branch had an estimated length of 8.0 years (Methods), whereas the patient was 7.1 years old at t-MN diagnosis. In this branch, 37 mutations could be attributed to SBS87 (Fig. 4H). These observations suggest that the first detectable division of the MLL^r cell occurred during thiopurine exposure. As the timing of MLL rearrangement within this branch cannot be further determined, it is unclear if this initial t-MN driver event in this patient was acquired before or after the initiation of treatment. The unique mutations in the t-MN and MLL^r HSPCs (β branches) were predominantly attributed to SBS87 and SBS1, whereas three posttreatment non-MLL^r HSPCs showed only the HSPC signature (δ branches), similar to what was observed in patient UPN024 (Fig. 4D and H). The lack of SBS87 mutations in these latter cells is likely explained by the quiescent state of normal HSPCs (42) and the dependency of thiopurine-induced mutagenesis on replication (43). In contrast, the t-MN and MLL^r HSPCs of UPN014 and UPN024 did harbor SBS87 mutations (Fig. 2A; Supplementary Fig. S6E), suggesting that their predecessors were replicating during thiopurine treatment. This idea is further supported by the presence of SBS1 mutations in these cells, which is a signature that has been associated with cell division (ref. 29; Fig. 4D and H). The data together suggest that cell division, which may have been propagated by MLL rearrangement (44), during thiopurine therapy results in the accumulation of passenger and driver mutations (Fig. 2E).

DISCUSSION

Previous studies have reported the mutational effects of chemotherapy exposure in cell culture systems, metastasized cancers, or colonic crypts (3, 6, 20). Here, we report the first systematic analysis of chemotherapy-associated mutation accumulation in normal blood cells of pediatric cancer patients. Our t-MN patient cohort includes a large variety in clinical characteristics, such as age at first cancer diagnosis, type of first cancer, and treatment regimen. Due to the relative rarity of the disease, and thus limited availability of samples in individual centers and even countries, international collaboration is essential to building larger cohorts and learning more about the specific mechanisms behind pediatric t-MN development. Despite these limitations, we could systematically confirm hypotheses that chemotherapy mutates normal HSPCs (5). Furthermore, t-MN blasts showed an even higher mutation load compared with DX2 HSPCs, resulting in similar mutation numbers to pAML. Phylogenetic analyses allowed us to time t-MN, which elucidated different timing of t-MN expansion between treatments. Lastly, patients who

Figure 4. MLL^r HSPCs in MLL^r t-MN patients. **A,** Phylogenetic tree of DX2 HSPCs and bulk t-MN blasts of patient UPN008. Branches without a label represent DX2 HSPCs. Colored dots indicate the similarity of the 96-trinucleotide profile of each branch with more than 10 mutations with SBS31. The numbers indicate the number of single base substitutions (SBS) and indels in that branch. Sample marked with an asterisk is the only one that harbored blast markers. Top right, schematic overview of the disease, treatment, and sample collection time line for this patient. In this patient, the t-MN blasts had no unique mutations. platin, cisplatin. **B,** The signature contribution of the mutations in the corresponding lineage trees on the left. **C,** Similar to **A** but for patient UPN024; similarity to SBS87 is depicted in the colored dots. Sample marked with an asterisk is the only one that harbored blast markers. **D,** Similar to **B** but for patient UPN024. MLLn, MLL normal. **E,** Similar to **A** but for patient UPN013; similarity to SBSC is depicted in the colored dots. Sample marked with an asterisk is the only one that harbored blast markers. Auto-SCT, autologous stem cell transplantation; HSCT, hematopoietic stem cell transplantation; SCT, stem cell transplantation. **F,** Similar to **B** but for patient UPN013. **G,** Similar to **A** but for patient UPN014; similarity to SBS87 is depicted in the colored dots. Samples marked with an asterisk are the only ones that harbored blast markers. All other samples were sorted on HSPC markers. **H,** Similar to **B** but for patient UPN014. α, the MLL rearrangement–containing branch; β, the aggregate of the unique mutations in the MLL^r samples; γ, the shared mutations of the MLL^r samples after the MLL-containing branch; δ, MLL-normal DX2 HSPC; ε, MLL-normal pre-SCT HSPC; ζ, MLL-normal FU1 and DX2 HSPCs; η, the primary ALL.

had the longest latency to t-MN development showed a lower mutation load, which could mean that in those patients HSPCs with fewer mutations took over the blood system.

To elaborate, we found that posttreatment HSPCs of childhood cancer patients harbored a mutation burden comparable with HSPCs of adults. Collectively, these HSPCs showed a mean increase of 16 years of mutational age with excesses up to 80 years. In some patients, this increment in mutation load could be attributed to direct mutagenesis by thiopurines or platinum-based drugs, as reflected by mutational signatures (SBS87 and SBS31, respectively). Whereas in previous literature of pediatric t-MN, both SBS31 and SBS35 have been linked to platinum therapy (13, 16), we here mainly identified SBS31. Although the originally extracted non-negative matrix factorization (NMF) signatures did show characteristics of SBS35, subsequent refitting revealed a larger cosine similarity with SBS31 (0.98). In contrast to these clear therapy-related signatures in some cells, we show that in most HSPCs, the increase in mutation burden upon chemotherapy exposure could not be explained by direct chemotherapeutic drug–induced mutagenesis. Their mutational profiles were more similar to those of older, healthy individuals, indicating that treatment predominantly causes indirect mutagenesis in exposed HSPCs. This indirect mutagenesis could be attributed to SBS5 and HSPC signatures, but not to SBS1, which is also observed in treatment-naïve HSPCs during healthy aging (18, 20) and in line with the predominant quiescent state of these cells after birth (42). Moreover, hematopoietic stem cell transplantation in patients with leukemia does not result in increased HSPC mutation loads (32), making bone marrow repopulation after therapy an unlikely cause for the observed increase in mutation load. The lack of direct chemotherapy-associated signatures in exposed HSPCs corresponds to recent data on environmental carcinogens in various mouse tumors (31), suggesting similar mechanisms may be active in other tissues. Thus, the origin for this indirect mutagenesis may be replicative stress (45) or stress-induced mutagenesis (46).

We found that direct mutagenesis by chemotherapeutic drugs may have varying dependencies. Whereas thiopurine-induced mutagenesis critically depended on cell division, platinum-based drugs were mutagenic to all assessed cells of exposed patients. The cisplatin-induced mutations in normal HSPCs support an earlier hypothesis that nonmalignant cells are first damaged by chemotherapy before developing into t-MN (5). Our observations that most HSPCs do not harbor SBS87 mutations after thiopurine treatment are in line with previous literature reporting on the lack of 5-FU–related mutations in exposed t-MN cases, which was believed to be caused by quiescence of normal cells at the time of treatment (5). Therefore, our data suggest that for the mutagenic action of cisplatin, cell proliferation is not required. Indeed, cisplatin covalently binds to base residues in double-stranded DNA and was previously reported in WGS data to induce mutations in all exposed t-MNs (5, 13). In addition, our findings imply that MLL[r] cells associated with DNA cross-linking treatment can only divide after the end of exposure, whereas MLL[r] cells can start dividing during treatment with the thiopurine base analogues.

In four cases, we observed that HSPCs acquired an MLL fusion (either before or during treatment) and gave rise to a pool of (pre-)leukemic, HSPC-like cells that started dividing during or directly after chemotherapy exposure. The additional driver mutations in two t-MN genomes indicate that the leukemic cell of origin started expanding and became dominant after the additional hit, which according to our data might be a nongenetic event. We also identified two cases in which MLL[r] HSPCs were genetically indistinguishable from the t-MN, similar to reports of earlier described leukemic stem cells (47, 48). Unfortunately, for our patients, we did not have multiple longitudinal samples available from the period between first diagnosis and t-MN to further assess the clonal dynamics preceding t-MN. Deep sequencing of such retrospective samples could, in the future, shed more light on the timing and evolution of t-MN development (16). Interestingly, the shared mutations of all MLL[r] cells in UPN008 do not harbor a driver mutation and were completely explained by SBS31 (platinum induced). This finding is in line with a previous report on three pediatric neuroblastoma patients, in whom CH, mainly consisting of platinum-induced mutations and no drivers, preceded the development of t-MN that arose after the acquisition of drivers (16). This is in stark contrast to CH in adult cancer patients, in whom no platinum-induced mutations were found after treatment (5, 49). In conclusion, we showed that chemotherapy can be mutagenic in at least three ways: directly to all exposed cells by DNA cross-linking, directly to dividing cells by base analogue incorporation, and indirectly by mimicking clock-like processes. All these mechanisms ultimately result in increased mutagenesis, which can contribute to t-MN development through induction of cancer driver mutations.

METHODS

Patient Samples

All bone marrow and peripheral blood samples were obtained via the biobank of the Princess Máxima Center for Pediatric Oncology with ethical approval under proposals OC2018-07, PMCLAB2018.026, and PMCLAB2020.151 in accordance with the Declaration of Helsinki. The mutational spectra from UPN003 were previously reported (32). Patients' written informed consents were obtained by the University Medical Center Utrecht and the Princess Máxima Center. This study was approved by the Biobank Research Ethics Committee of the University Medical Center Utrecht and the Biobank and Data Access Committee of the Princess Máxima Center. Five patients were first diagnosed with primary ALL and were in remission at the time that the follow-up (FU) sample was taken. These were UPN002, UPN004, UPN005, UPN006, and UPN007. The other patients had diverse primary diagnoses and developed a t-MN (t-AML, $N = 18$, t-MDS, $N = 1$) later in life. One t-MN blast sample (UPN021 DX2AML) was excluded, as no clonal mutations were found in this sample, indicating that the sorted populations were not purely blasts. UPN009 received radiotherapy, but not chemotherapy as a treatment for the primary cancer; therefore, the DX2 samples of this patient were excluded from mutation load analyses and posttreatment signature analyses.

FACS and HSPC Culture

Bone marrow mononuclear cells were stained for FACS after thawing. HSPCs were identified using the following surface markers: Lin⁻CD11c⁻CD16⁻CD34⁺, CD38⁻/CD45RA⁻ (Supplementary Fig. S1). We defined (t-)MN blasts from both first and second diagnoses based on diagnostic immunophenotyping data if available. In most cases, these blasts were CD33, CD38, and/or CD34 positive. ALL blasts from the first diagnosis were defined based on diagnostic immunophenotyping data if available (mostly these were CD10, CD19, or CD7 positive).

Blasts and HSPCs were purified on an SH800S Cell Sorter (Sony, RRID:SCR_018066). First, blasts were sorted in bulk for DNA isolation, after which HSPCs were index sorted in a flat-bottom, 384-well plate prepared with 75 μL HSPC culture medium per well. HSPC culture medium consisted of StemSpan SFEM medium (STEMCELL Technologies) supplemented with SCF (100 ng/mL), FLT3 ligand (100 ng/mL), IL6 (20 ng/mL), IL3 (10 ng/mL), TPO (50 ng/mL), UM729 (500 nmol/L), and Stemregenin (750 nmol/L).

For five samples (UPN001DX2, UPN023DX1, UPN002DX1, UPN005DX1, and UPN004DX1), the obtained sample was depleted for monocytic, pro–T cell, or pro–B cell blasts (marked by anti-CD14, CD7, and CD10, respectively) using the EasySep anti-APC kit, following the manufacturer's instructions. After blast deletion, we plated mesenchymal stromal cells (MSC) and sorted HSPCs following the same procedure as with all other samples.

HSPCs were cultured for 4 to 7 weeks at 37°C in 5% CO_2 before collection. MSCs were cultured from a fraction of bone marrow cells by plating bulk cells in 12-well culture dishes with DMEM-F12 medium (Gibco) supplemented with 10% FBS. The medium was refreshed every other day to remove nonadherent cells, and MSCs could be harvested when confluent (after approximately 2 to 3 weeks).

FACS Antibodies

All antibodies were obtained from BioLegend except for CD13 (Biosciences). Antibodies used for (t-)MN blast and HSPC populations were as follows: CD34-BV421 (clone 561, 1:20, RRID:AB_11147951), lineage (CD3/CD14/CD19/CD20/CD56)-FITC (clones UCHT1, HCD14, HIB19, 2H7, HCD56, 1:20, RRID:AB_10644012), CD38-PE (clone HIT2, 1:50, RRID:AB_314357), CD90-APC (clone 5E10, 1:200, RRID:AB_893440), CD45RA-PerCP/Cy5.5 (clone HI100, 1:20, RRID: AB_893358), CD33-PE/Cy7 (clone WM53, 1:100, RRID:AB_2734264), CD49f-PE/Cy7 (clone GoH3, 1:100, RRID:AB_2561704), CD16-FITC (clone 3G8, 1:100, RRID:AB_314205), CD11c-FITC (clone 3.9, 1:20, RRID:AB_314173), CD123-Pe/Cy7 (clone 6H6, 1:100, RRID:AB_493577), CD13-PerCP/Cy5.5 (Biosciences, clone WM15, 1:20, RRID:AB_10645787), and CD14-APC (HCD14, RRID:AB_830680). Additional antibodies used for depleting ALL blast populations were CD10-APC (clone HI10a, 1:100, RRID:AB_314920) and CD7-APC (clone CD7-6B7, 1:100, RRID:AB_1877156).

Cord Blood Chemotherapy Exposure

We used a previously established protocol (33) to treat cord blood–derived HSPCs with approximately IC_{50} concentrations of treosulfan and thiotepa (4 and 12.5 μmol/L, respectively). Thiotepa treatment was combined with liver enzymes to support conversion to the active metabolites (6). Used concentrations for these additional compounds were 0.25% S9 fraction (Aroclor-1254–induced male Sprague Dawley rat liver), 3 mmol/L NADP (Sigma), and 15 mmol/L DL-isocitric acid trisodium salt hydrate (Sigma).

DNA Isolation and WGS

DNA was isolated from cell pellets of blasts, MSCs, and clonally expanded HSPCs using the DNeasy DNA Micro Kit (Qiagen) according to the instructions provided by the manufacturer. We modified this protocol slightly by adding 2 μL RNase A (Qiagen) during the lysis step and eluting DNA in 50 μL low EDTA TE buffer (10 mmol/L Tris, 0.1 mmol/L EDTA, G Biosciences).

For each sample, DNA libraries for Illumina sequencing were generated from at least 35 ng genomic DNA using standard protocols. The libraries were sequenced on NovaSeq 6000 sequencers (RRID:SCR_016387; 2 × 150 bp) at a depth of 15 to 30×. Two t-MN blast samples (UPN018 and UPN023) were sequenced to 90× coverage, as only a DNA pellet was available, and blast purity was 15% and 22%, respectively. Reads were mapped to the human reference genome GRCh38 using the Burrows-Wheeler Aligner v0.7.17 mapping tool

with settings "bwa mem –M –c100" (50). Sambamba v0.6.8 (51) was used to mark duplicate sequencing reads, and GATK v.4.1.3.0 (52) was used to perform base recalibration. See https://github.com/UMCUGenetics/NF-IAP for a full description and all code of the pipeline.

Mutation Calling and Filtering

Mutation calling and filtering were performed on multisample VCF files generated using HaplotypeCaller from GATK v.4.1.3.0. GATK's VariantFiltration was used for variant quality evaluation with options: "–filter-expression "QD < 2.0" –filter-expression "MQ < 40.0" –filter-expression "FS > 60.0" –filter-expression "HaplotypeScore > 13.0" –filter-expression "MQRankSum <–12.5" –filter-expression "Read-PosRankSum <–8.0" –filter-expression "MQ0 > = 4 && ((MQ0/(1.0 * DP)) > 0.1)" –filter-expression "DP < 5" –filter-expression "QUAL < 30" –filter-expression "QUAL > = 30.0 && QUAL < 50.0" –filter-expression "SOR > 4.0" –filter-name "SNP_LowQualityDepth" –filter-name "SNP_MappingQuality" –filter-name "SNP_StrandBias" –filter-name "SNP_HaplotypeScoreHigh" –filter-name "SNP_MQRankSumLow" –filter-name "SNP_ReadPosRankSumLow" –filter-name "SNP_HardToValidate" –filter-name "SNP_LowCoverage" –filter-name "SNP_VeryLowQual" –filter-name "SNP_LowQual" –filter-name "SNP_SOR" -cluster 3 -window 10."

For two impure t-MN blast samples that were sequenced to 90×, the "SNP_LowQualityDepth" filter was lowered to "QD < 1.0."

Subsequently, SNPEffFilter (53), SNPSiftDbnsfp (database dbNSFP3.2a (54), GATK VariantAnnotator (database COSMIC v.89), and SNPSiftAnnotate (database GoNL release 5) were used for variant annotation.

Finally, to obtain catalogs of high-quality somatic mutation calls, we applied postprocessing filtering steps, per patient, as described below (all scripts are available at https://github.com/ToolsVanBox/SMuRF).

Briefly, only variants were considered that (i) were present on autosomal chromosomes; (ii) passed VariantFiltration with a GATK phred-scaled quality score ≥100; (iii) had a base coverage of at least 10× (30× samples) or 5× (15× samples) in the clonal and paired control sample; (iv) had a mapping quality (MQ) score of 60; (v) did not overlap with single-nucleotide polymorphisms (SNP) in the Single Nucleotide Polymorphism Database v146 and a panel of unmatched normal human MSC and fetal genomes (BED-file available upon request); (vi) had a GATK genotype score (GQ) of 99 (indel/single base substitution in clonal sample, or indel in paired control) or higher than 10 (single base substitution in paired control); (vii) had a variant allele frequency of ≥0.3 (single base substitution/indel in 30× coverage sample or indel in 15× sample) or ≥0.15 (single base substitution in 15× coverage sample) or 0.07 (single base substitution/indel in the two 90× t-MN samples, see above) to exclude *in vitro* accumulated mutations; and (viii) did not have any evidence from a paired control sample (MSCs isolated from the same bone marrow) if available. For patients for whom no matched MSC control was available, or when the control was contaminated with blast cells (UPN008/UPN014), instead of step (viii) a mutation was filtered out when it (a) was clonally present in all samples that passed QC for that mutation, (b) was subclonally present in any sample, or (c) was not confidently absent in at least one sample.

One HSPC clone (UPN013DX2 HSPC 1B23) was excluded, as it did not match the fingerprint of the other samples of UPN013 and was likely a surviving donor cell from one of the two unsuccessful stem cell transplantations that were administered prior to sample collection. For UPN008 and UPN014, bulk MSC samples were excluded, as they showed clear contamination with t-MN blasts as evidenced by reads supporting the t-MN driving fusion and subclonal presence of t-MN single-nucleotide variant/indel mutations.

Driver Events

Single base substitutions or indels were considered driver events when they (i) had an MQ of 60, and a GQ of 10 or higher in both

the (t-)MN and the paired control sample (if available) and minimal base coverage of 10× in both the (t-)MN and the paired control sample (if available); (ii) had a variant allele frequency higher than 0.3; (iii) were present in driver genes (either COSMIC Cancer Gene Consensus, version of 9/5/2019) or one of the frequently mutated genes in primary pediatric (t-)MN (23); (iv) were a missense, frameshift, stop–gain, insertion, or deletion; (v) had either a high or moderate expected effect as annotated by SnpEff; (vi) were not present in the Single Nucleotide Polymorphism Database v146 and a panel of unmatched normal human MSC and fetal genomes (BED-file available upon request); and (vii) had no evidence in the paired control samples (if available; ref. 55).

Structural variant and chromosomal copy-number alteration calling was performed using the GRIDSS-purple-linx pipeline developed at the Hartwig Medical Foundation (56). All structural variants were validated by hand using the Integrative Genomics Viewer (IGV; ref. 57), and false-positive results were excluded. All whole chromosome duplications and deletions were considered driver events, as well as partial chromosomal (arm) gains and losses (reported as one category). Finally, translocation events resulting in fusion genes that involved at least one known AML driver gene [in this data set, KMT2A (MLL), RUNX1, MECOM, or MLLT10] were considered drivers.

Comparison with the Baseline

When comparing mutation load, single base substitution and indel counts were normalized to GATK CallableLoci's CALLABLE length. The baseline data from previous publications were used (18, 19). As described before, a linear mixed-effects model was used to calculate the slope and intercept of the baseline while taking donor dependency into account using lme4 package in R (58). Mutational ages were calculated based on the expected rate of mutation accumulation over time in HSPCs of healthy individuals, previously defined as baseline (18), as mutational_age = (number_of_mutations – baseline_intercept)/baseline_slope.

Mutational Signature Extraction and Refitting

For the analysis of mutational patterns and signatures, the in-house developed R package MutationalPatterns v3.0.1 (59) was used. For single base substitution analysis, first, the 96-trinucleotide profiles per sample were extracted. Then, NMF was applied on these data combined with previously published mutational patterns of healthy tissues (18) to extract nine signatures ("extract_signatures" with options "rank = 9, nrun = 100"). These signatures were compared with the COSMIC mutational signature database v3.1 (55) and a previously established clock-like signature in HSPCs (HSPC signature; ref. 60). Signatures with a cosine similarity of >0.8 to one of the known signatures were replaced by that signature (1, 5, 18, 31, 87, HSPC). Of note, the signature replaced by SBS31 was very similar to this signature (cosine similarity of 0.98) but also had some characteristics of SBS35 (cosine similarity of 0.73). The other signatures were named SBSB and SBSC.

Only from one sample that had more than 100 mutations, the cosine similarity between the reconstructed profile derived from these signatures and the original profile had a cosine similarity below 0.8 (UPN003 t-MN, 1,078 mutations; Supplementary Fig. S4A). We have previously reported this sample and showed that it had the contribution of SBSA, a signature caused by ganciclovir. Therefore, we have added SBSA to the mutational signature repertoire, which resulted in a cosine similarity of UPN003 t-MN to 0.98 (Supplementary Fig. S4B).

Fitting the Signatures to the Mutational Profiles of Samples

The resulting set of signatures was used to perform bootstrapped fitting using "fit_to_signatures_bootstrapped" with options "n_boots = 100, max_delta = 0.05." The bootstrap results were averaged per sample to get the contribution of each signature to the profile of each sample. Finally, per sample the number of single base substitutions in the reconstructed 96-trinucleotide profile was subtracted from the original number of single base substitutions and added as "unexplained" mutations.

The same steps were taken when refitting on the per-branch profiles of phylogenetic trees (Fig. 4). The aggregate profiles (Fig. 3C) were acquired by first making an average profile of HSPCs per time point per patient and then taking the mean from the resulting profiles per time point.

Determining Signature Categories

Cells with a contribution of more than 20% from signature SBSA, SBSB, SBSC, SBS31, or SBS87 were assigned as t-HSPC and grouped to the corresponding signature category. Cells with more than 20% contribution of more than one of these signatures were grouped to the signature with the highest contribution. Finally, all remaining cells were assigned to the category n-HSPC, as most of their mutations could be attributed to SBS1, SBS5, and the HSPC signature. SBS18 was not a category, as in none of the HSPCs or t-MN samples did SBS18 have a contribution of 20% or more.

Genomic Age Estimation

The healthy 96-trinucleotide mutation data were obtained from previously sequenced HSPC clonal cultures (18, 32). For each trinucleotide category, a linear mixed-effects model was applied to determine the age-related increase. Predictions for the 96-trinucleotide profiles of healthy aging were made for each time point at a resolution of 0.1 year. For each time point of our data set (DX1, FU, and DX2), the n-HSPC profiles were merged. Each resulting profile was compared with all baseline profiles using cosine similarity. The mutational age of each of the three n-HSPC profiles was set to the age of the baseline profile with the highest cosine similarity.

Constructing Phylogenetic Trees

To construct phylogenetic trees, all samples from one patient were compared among one another. To obtain only high-confident mutations and to include mutations that arose during early development, filtering was slightly adjusted compared with previous analyses. If a control MSC sample was available, mutations that were subclonally (variant allele frequency <0.3) present in the control were considered. Mutations that were subclonally present in any other sample were filtered out. To still account for germline mutations, mutations that were clonally present in all samples were filtered out. In addition, all samples needed to have passed QC filters as described above (among others, sufficient coverage and MQ), not only the sample in which the mutation was found. Finally, for patient UPN013, 35 mutations were removed that were detected in all samples but the primary ALL and that were present in locations with loss of heterozygosity in the ALL. All shared mutations were manually inspected in IGV, and false-positive results were filtered out. A binary mutation table was constructed from the mutations that passed these criteria, and a tree was constructed using the ape v5.5 R package (61).

As filtering to obtain the tree is very strict, mutations that were filtered out due to failed QC in one or more samples were reconsidered. These mutations were added to a branch only if the variant allele frequency of that mutation was ≥0.15 for all samples in that branch and if the variant allele frequency was 0 for all samples not in the branch. In addition, mutations only found in one sample were only considered if that sample passed QC.

The mutations per branch were extracted using the binary mutation table, and a cosine similarity to one of the NMF-extracted or COSMIC signatures was calculated. Then, the per-branch mutations were merged into categories, and refitting was performed on the resulting mutation catalogs as described above.

Potential Impact of Mutational Signatures

Calculating the probability of a mutation being caused by the signatures that contributed to that sample was done similarly to that done by Morganella and colleagues (35). In short, the contributions of each signature to the sample were multiplied by the chance of each signature to induce a mutation of the mutation type and trinucleotide context of the driver mutation. These values were summed. The fraction that each signature contributed to the summed value was multiplied by 100 to get a probability in percentages.

Extended Context

To determine the extended context of the mutations of posttreatment samples from patient UPN013, we extracted the −4/+4 context of each unique mutation in all samples. Next, we pulled all mutations for each of the six mutation types and plotted the sequence logos with the R package *ggseqlogo* v.0.1 (62).

Statistical Analysis

Due to limited primary material availability, no sample-size calculation was performed. No randomization was performed. Regarding the included t-MN patients, all patients for whom material at time of t-MN was available in the biobank were included in our study. For most samples, at least three (with up to 16) HSPCs were sequenced. As specifically the t-MN material was scarce, we collected and processed all available samples to obtain this unique data set. For the comparisons of mutation burden and signature contribution between groups, a two-sided Wilcox test was used.

Data and Code Availability

The data sets generated during this study are available at the European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) under accession number EGA:EGAS00001005141. Most of the scripts used during this study are available at https://github.com/ToolsVanBox/ and in the MutationalPatterns R package (https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html). Other scripts are available upon request.

Authors' Disclosures

A.K.M. Rosendahl Huber, A.J.C.N. van Leeuwen, and R. van Boxtel report a patent for means and methods for assessing genotoxicity pending. R. van Boxtel reports grants from the European Research Council (ERC) and the Dutch Research Council (NWO) during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

E.J.M. Bertrums: Data curation, formal analysis, validation, investigation, methodology, writing–original draft, project administration, writing–review and editing. **A.K.M. Rosendahl Huber:** Conceptualization, data curation, software, formal analysis, validation, investigation, methodology, writing–original draft, project administration, writing–review and editing. **J.K. de Kanter:** Data curation, software, formal analysis, investigation, visualization, writing–original draft, writing–review and editing. **A.M. Brandsma:** Investigation, methodology. **A.J.C.N. van Leeuwen:** Investigation. **M. Verheul:** Investigation. **M.M. van den Heuvel-Eibrink:** Data curation, supervision, project administration. **R. Oka:** Software. **M.J. van Roosmalen:** Software. **H.A. de Groot-Kruseman:** Data curation. **C.M. Zwaan:** Data curation, supervision, project administration. **B.F. Goemans:** Data curation, supervision, project administration. **R. van Boxtel:** Conceptualization, supervision, funding acquisition, writing–original draft, project administration, writing–review and editing.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

REFERENCES

1. Chabner BA, Roberts TG Jr. Timeline: chemotherapy and the war on cancer. Nat Rev Cancer 2005;5:65–72.
2. Hurley LH. DNA and its associated processes as targets for cancer therapy. Nat Rev Cancer 2002;2:188–200.
3. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 2019;575:210–6.
4. Pich O, Muiños F, Lolkema MP, Steeghs N, Gonzalez-Perez A, Lopez-Bigas N. The mutational footprints of cancer therapies. Nat Genet 2019;51:1732–40.
5. Pich O, Cortes-Bullich A, Muiños F, Pratcorona M, Gonzalez-Perez A, Lopez-Bigas N. The evolution of hematopoietic cells under cancer therapy. Nat Commun 2021;12:4803.
6. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A compendium of mutational signatures of environmental agents. Cell 2019;177:821–36.
7. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. Blood 2020;135:41–55.
8. Christensen S, Van der Roest B, Besselink N, Janssen R, Boymans S, Martens JWM, et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. Nat Commun 2019;10:4571.
9. Crawford J, Dale DC, Lyman GH. Chemotherapy-induced neutropenia: risks, consequences, and new directions for its management. Cancer 2004;100:228–37.
10. Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. Nat Rev Cancer 2014;14:61–70.
11. Choi DK, Helenowski I, Hijiya N. Secondary malignancies in pediatric cancer survivors: perspectives and review of the literature. Int J Cancer 2014;135:1764–73.
12. Cupit-Link MC, Kirkland JL, Ness KK, Armstrong GT, Tchkonia T, LeBrasseur NK, et al. Biology of premature ageing in survivors of cancer. ESMO Open 2017;2:e000250.
13. Schwartz JR, Ma J, Kamens J, Westover T, Walsh MP, Brady SW, et al. The acquisition of molecular drivers in pediatric therapy-related myeloid neoplasms. Nat Commun 2021;12:985.
14. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat Genet 2020;52:1219–26.
15. McNerney ME, Godley LA, Le Beau MM. Therapy-related myeloid neoplasms: when genetics and environment collide. Nat Rev Cancer 2017;17:513–27.
16. Coorens THH, Collord G, Lu W, Mitchell E, Ijaz J, Roberts T, et al. Clonal hematopoiesis and therapy-related myeloid neoplasms following neuroblastoma treatment. Blood 2021;137:2992–7.
17. Takahashi K, Wang F, Kantarjian H, Doss D, Khanna K, Thompson E, et al. Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. Lancet Oncol 2017;18:100–11.

18. Osorio FG, Rosendahl Huber A, Oka R, Verheul M, Patel SH, Hasaart K, et al. Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. Cell Rep 2018;25:2308–16.

19. Brandsma AM, Bertrums EJM, van Roosmalen MJ, Hofman DA, Oka R, Verheul M, et al. Mutation signatures of pediatric acute myeloid leukemia and normal blood progenitors associated with differential patient outcomes. Blood Cancer Discov 2021;2:484–99.

20. Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, et al. Population dynamics of normal human blood inferred from somatic mutations. Nature 2018;561:473–8.

21. Le H, Singh S, Shih SJ, Du N, Schnyder S, Loredo GA, et al. Rearrangements of the MLL gene are influenced by DNA secondary structure, potentially mediated by topoisomerase II binding. Genes Chromosomes Cancer 2009;48:806–15.

22. Mirault ME, Boucher P, Tremblay A. Nucleotide-resolution mapping of topoisomerase-mediated and apoptotic DNA strand scissions at or near an MLL translocation hotspot. Am J Hum Genet 2006;79:779–91.

23. Bolouri H, Farrar JE, Triche T Jr., Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med 2018;24:103–12.

24. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature 2015;518:552–5.

25. Pedersen-Bjergaard J, Pedersen M, Roulston D, Philip P. Different genetic pathways in leukemogenesis for patients presenting with therapy-related myelodysplasia and therapy-related acute myeloid leukemia. Blood 1995;86:3542–52.

26. Robinson PS, Coorens THH, Palles C, Mitchell E, Abascal F, Olafsson S, et al. Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. Nat Genet 2021;53:1434–42.

27. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3:246–59.

28. Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. Nat Commun 2019;10:2969.

29. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.

30. Boot A, Huang MN, Ng AWT, Ho SC, Lim JQ, Kawakami Y, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. Genome Res 2018;28:654–65.

31. Riva L, Pandiri AR, Li YR, Droop A, Hewinson J, Quail MA, et al. The mutational signature profile of known and suspected human carcinogens in mice. Nat Genet 2020;52:1189–97.

32. de Kanter JK, Peci F, Bertrums E, Rosendahl Huber A, van Leeuwen A, van Roosmalen MJ, et al. Antiviral treatment causes a unique mutational signature in cancers of transplantation recipients. Cell Stem Cell 2021;28:1726–39.

33. Rosendahl Huber A, van Leeuwen A, Peci F, de Kanter JK, Bertrums EJM, van Boxtel R. Whole-genome sequencing and mutational analysis of human cord-blood derived stem and progenitor cells. STAR Protoc 2022;3:101361.

34. Brady SW, Ma X, Bahrami A, Satas G, Wu G, Newman S, et al. The clonal evolution of metastatic osteosarcoma as shaped by cisplatin treatment. Mol Cancer Res 2019;17:895–906.

35. Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, et al. The topography of mutational processes in breast cancer genomes. Nat Commun 2016;7:11383.

36. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature 2020;578:266–72.

37. Radivoyevitch T, Sachs RK, Gale RP, Molenaar RJ, Brenner DJ, Hill BT, et al. Defining AML and MDS second cancer risk dynamics after diagnoses of first cancers treated or not with radiation. Leukemia 2016;30:285–94.

38. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet 2015;47:1402–7.

39. Hasaart KAL, Manders F, van der Hoorn ML, Verheul M, Poplonski T, Kuijk E, et al. Mutation accumulation and developmental lineages in

40. Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature 2014;513:422–5.

41. Chen C, Yu W, Alikarami F, Qiu Q, Chen CH, Flournoy J, et al. Single-cell multiomics reveals increased plasticity, resistant populations and stem-cell-like blasts in KMT2A-rearranged leukemia. Blood 2021;139:2198–211.

42. Bowie MB, McKnight KD, Kent DG, McCaffrey L, Hoodless PA, Eaves CJ. Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. J Clin Invest 2006;116:2808–16.

43. Ling YH, Nelson JA, Cheng YC, Anderson RS, Beattie KL. 2′-Deoxy-6-thioguanosine 5′-triphosphate as a substrate for purified human DNA polymerases and calf thymus terminal deoxynucleotidyltransferase in vitro. Mol Pharmacol 1991;40:508–14.

44. van der Linden MH, Willekes M, van Roon E, Seslija L, Schneider P, Pieters R, et al. MLL fusion-driven activation of CDK6 potentiates proliferation in MLL-rearranged infant ALL. Cell Cycle (Georgetown, Tex) 2014;13:834–44.

45. Flach J, Bakker ST, Mohrin M, Conroy PC, Pietras EM, Reynaud D, et al. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature 2014;512:198–202.

46. Cipponi A, Goode DL, Bedo J, McCabe MJ, Pajic M, Croucher DR, et al. MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer. Science 2020;368:1127–31.

47. Chopra M, Bohlander SK. The cell of origin and the leukemia stem cell in acute myeloid leukemia. Genes Chromosomes Cancer 2019;58:850–8.

48. Jordan CT. The leukemic stem cell: best practice and research Clin Haematol 2007;20:13–8.

49. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P, et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. Cell Stem Cell 2017;21:374–82.

50. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 2009;25:1754–60.

51. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 2015;31: 2032–4.

52. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.

53. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front Genet 2012;3:35.

54. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80–92.

55. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2019;47:D941–d7.

56. Cameron DL, Baber J, Shale C, Papenfuss AT, Valle-Inclan JE, Besselink N, et al. GRIDSS, PURPLE, LINX: unscrambling the tumor genome via integrated analysis of structural variation and copy number. BioRxiv 781013 [Preprint]. 2019. Available from: https://doi.org/10.1101/781013.

57. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.

58. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1–48.

59. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 2018;10:33.

60. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature 2016;538:260–4.

61. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–8.

62. Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 2017;33:3645–7.

normal and Down syndrome human fetal haematopoiesis. Sci Rep 2020;10:12991.