LETTER

doi:10.1038/nature13968

Role of *TP53* mutations in the origin and evolution of therapy-related acute myeloid leukaemia

Terrence N. Wong[1]*, Giridharan Ramsingh[2]*, Andrew L. Young[3]*, Christopher A. Miller[4], Waseem Touma[1], John S. Welch[1,5], Tamara L. Lamprecht[1], Dong Shen[6], Jasreet Hundal[4], Robert S. Fulton[4], Sharon Heath[1], Jack D. Baty[7], Jeffery M. Klco[8], Li Ding[1,5], Elaine R. Mardis[4,5,9], Peter Westervelt[1,5], John F. DiPersio[1,5], Matthew J. Walter[1,5], Timothy A. Graubert[1,5], Timothy J. Ley[1,5], Todd E. Druley[3], Daniel C. Link[1,5] & Richard K. Wilson[4,5,9]

Therapy-related acute myeloid leukaemia (t-AML) and therapy-related myelodysplastic syndrome (t-MDS) are well-recognized complications of cytotoxic chemotherapy and/or radiotherapy[1]. There are several features that distinguish t-AML from *de novo* AML, including a higher incidence of *TP53* mutations[2,3], abnormalities of chromosomes 5 or 7, complex cytogenetics and a reduced response to chemotherapy[4]. However, it is not clear how prior exposure to cytotoxic therapy influences leukaemogenesis. In particular, the mechanism by which *TP53* mutations are selectively enriched in t-AML/t-MDS is unknown. Here, by sequencing the genomes of 22 patients with t-AML, we show that the total number of somatic single-nucleotide variants and the percentage of chemotherapy-related transversions are similar in t-AML and *de novo* AML, indicating that previous chemotherapy does not induce genome-wide DNA damage. We identified four cases of t-AML/t-MDS in which the exact *TP53* mutation found at diagnosis was also present at low frequencies (0.003–0.7%) in mobilized blood leukocytes or bone marrow 3–6 years before the development of t-AML/t-MDS, including two cases in which the relevant *TP53* mutation was detected before any chemotherapy. Moreover, functional *TP53* mutations were identified in small populations of peripheral blood cells of healthy chemotherapy-naive elderly individuals. Finally, in mouse bone marrow chimaeras containing both wild-type and $Tp53^{+/-}$ haematopoietic stem/progenitor cells (HSPCs), the $Tp53^{+/-}$ HSPCs preferentially expanded after exposure to chemotherapy. These data suggest that cytotoxic therapy does not directly induce *TP53* mutations. Rather, they support a model in which rare HSPCs carrying age-related *TP53* mutations are resistant to chemotherapy and expand preferentially after treatment. The early acquisition of *TP53* mutations in the founding HSPC clone probably contributes to the frequent cytogenetic abnormalities and poor responses to chemotherapy that are typical of patients with t-AML/t-MDS.

t-AML and t-MDS are clonal haematopoietic disorders that typically develop 1–5 years after exposure to chemotherapy or radiotherapy[1]. To understand better how prior cytotoxic therapy contributes to the high incidence of *TP53* mutations and karyotypic abnormalities in t-AML/t-MDS, we sequenced the genomes of 22 cases of t-AML, including one case that has been previously reported[5]. These data were compared to whole-genome sequence data previously reported for *de novo* AML[6] and secondary AML (s-AML) arising from MDS for which patients did not receive chemotherapy except hydroxyurea[7,8]. Of the sequenced t-AML cases, 23% had rearrangements of MLL (also known as *KMT2A*), 23% had complex cytogenetics, and 36% had normal cytogenetics (Extended Data Table 1 and Supplementary Table 1).

We predicted that DNA damage induced during exposure to cytotoxic therapy would manifest itself in t-AML genomes with an increased mutation burden. However, the total number of validated somatic single-nucleotide variants (SNVs) and genic (tier 1) somatic SNVs identified was similar to that for *de novo* AML and s-AML (Fig. 1a, b). Likewise, the number of small insertions or deletions (indels) in genic regions was similar in t-AML, *de novo* AML and s-AML (Fig. 1c). A previous study

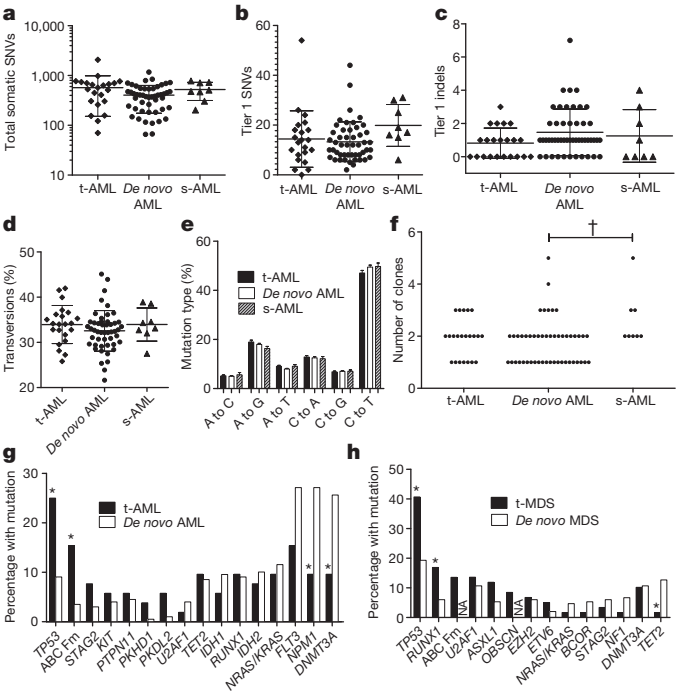


Figure 1 | The mutational burden in t-AML is similar to *de novo* AML. **a**, Total number of validated tier 1–3 somatic SNVs in t-AML ($n = 22$), *de novo* AML ($n = 49$) and s-AML ($n = 8$). The mean ages of the t-AML, *de novo* AML and s-AML cohorts were 55.7, 51 and 54.6 years, respectively. **b**, Number of validated tier 1 somatic SNVs. **c**, Number of validated tier 1 small indels. **d**, Percentage of tier 1–3 somatic SNVs that are transversions. **e**, Mutational spectrum for all validated tier 1–3 somatic SNVs. **f**, Number of distinct clones per sample inferred from the identification of discrete clusters of mutations with distinct variant allele frequencies. **g**, Percentage of cases of t-AML ($n = 52$) or *de novo* AML ($n = 199$) harbouring non-synonymous mutations of the indicated gene. **h**, Percentage of cases of t-MDS ($n = 59$) or *de novo* MDS ($n = 150$) harbouring non-synonymous mutations of the indicated gene. ABC Fm, ABC family genes; NA, not available. †$P < 0.05$ by one-way analysis of variance (ANOVA). *$P < 0.05$ by Fisher's exact test. Data represent the mean ± standard deviation (s.d.).

[1]Department of Medicine, Division of Oncology, Washington University, St Louis, Missouri 63110, USA. [2]Department of Medicine, Jane Anne Nohl Division of Hematology, University of Southern California, Los Angeles, California 90089, USA. [3]Department of Pediatrics, Division of Hematology/Oncology, Washington University, St Louis, Missouri 63110, USA. [4]The Genome Institute, Washington University, St Louis, Missouri 63110, USA. [5]Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. [6]AstraZeneca, Gaithersburg, Maryland 20878, USA. [7]Division of Biostatistics, Washington University, St Louis, Missouri 63110, USA. [8]Department of Pathology and Immunology, Washington University, St Louis, Missouri 63110, USA. [9]Department of Genetics, Washington University, St Louis, Missouri 63110, USA.

*These authors contributed equally to this work.

showed that transversions are specifically enriched in relapsed AML after chemotherapy[9]. However, the percentage of transversions, and in fact of all six classes of SNVs, was similar in all three cohorts (Fig. 1d, e). Structural variants and somatic copy number alterations were uncommon in these t-AML cases (Supplementary Table 2 and Extended Data Fig. 1a). Moreover, the number of identifiable subclones in t-AML was similar to that observed in *de novo* AML (Fig. 1f and Extended Data Fig. 1b). Collectively, these data show that the mutation burden of t-AML genomes is similar to that of *de novo* AML genomes.

We next asked whether the pattern of genes frequently mutated in t-AML/t-MDS is distinct from that observed in *de novo* AML/MDS. Whole-genome sequencing identified an average of 10.2 ± 7.1 missense, nonsense, in-frame indel or frameshift mutations per t-AML genome (Supplementary Table 3). To define better the frequency of specific mutations in t-AML/t-MDS, we sequenced a panel of 149 AML/MDS-related genes in an additional 89 patients with t-AML or t-MDS (Supplementary Table 4). We combined the whole-genome sequence data with the extension series to report on 52 cases of t-AML and 59 cases of t-MDS. Abnormalities of chromosome 5 or 7 or complex cytogenetics were present in 55.0% of cases (Extended Data Table 2 and Supplementary Table 1). The t-AML/t-MDS data were compared to 199 previously reported *de novo* AML genomes or exomes[6], or 150 previously reported cases of *de novo* MDS in which extensive candidate gene sequencing was performed[8]. As reported previously, *TP53* mutations are significantly enriched in t-AML/t-MDS compared with *de novo* AML/MDS (Fig. 1g, h and Supplementary Table 5). Interestingly, mutations of ABC transporter genes, a subset of which have been implicated in chemotherapy resistance, are also enriched in t-AML versus *de novo* AML. On the other hand, several well-defined driver gene mutations (that is, *DNMT3A* and *NPM1*) were significantly less common in t-AML. Thus, although the total mutation burden is similar, a distinct subset of mutated genes is present in t-AML/t-MDS.

TP53 is the most commonly mutated gene in t-AML/t-MDS, with 33.3% of patients affected in our series (Fig. 1g, h); the vast majority of these mutations have previously been identified as pathogenic[10]. Multivariate analysis revealed that *TP53* mutations were associated with poor risk cytogenetics and a worse prognosis (Supplementary Tables 6, 7 and Extended Data Fig. 2), both hallmarks of t-AML/t-MDS. These observations suggest a central role for *TP53* mutations in the pathogenesis of many cases of t-AML/t-MDS. However, the mechanism by which *TP53* mutations are selectively enriched in t-AML/t-MDS is unclear. The mutation burden in the genomic region containing *TP53* (including silent tier 1, and any tier 2 or tier 3 mutations) is similar between t-AML and *de novo* AML (Extended Data Fig. 1c). Thus, it is not likely that chemotherapy directly induces *TP53* mutations. We recently reported that individual HSPCs accumulate somatic mutations as a function of age, such that by age 50, there are on average five coding gene mutations per HSPC[11]. On the basis of these data and on current estimates that there are approximately 10,000 haematopoietic stem cells (HSCs) in humans[12], we predict that 44% of healthy individuals at 50 years of age may have at least one HSPC that carries a randomly generated, functional *TP53* mutation (see Methods). *TP53* has a central role in regulating cellular responses to genotoxic stress[13–17], and loss of *TP53* provides a selective advantage for neoplastic growth[18]. Together, these observations suggest a model in which rare HSPCs carrying age-related *TP53* mutations are resistant to chemotherapy and expand preferentially after treatment (Extended Data Fig. 3).

This model suggests the following testable predictions: (1) in patients with t-AML containing clonal *TP53* mutations, HSPCs harbouring the specific *TP53* mutation will be present long before the development of overt t-AML; (2) somatic *TP53* mutations will be present in the HSPCs of some healthy individuals never exposed to cytotoxic therapy; and (3) HSPCs harbouring *TP53* mutations will expand under the selective pressure of chemotherapy.

To test the first prediction, we identified seven cases of t-AML/t-MDS with specific *TP53* mutations for which we had leukapheresis or bone marrow specimens banked 3–8 years before the development of t-AML/t-MDS (Extended Data Table 3). Of note, in all the cases, the *TP53* mutation was clonal in the t-AML/t-MDS diagnostic sample. Current next-generation sequencing technology is limited in the detection of rare variant alleles owing to an intrinsic sequencing error rate of ~ 0.1% (ref. 19). To overcome this limitation, we introduced random barcodes during production of the sequencing libraries, such that sequence 'read families' containing unique barcodes are generated (Extended Data Fig. 4a). Using tumour DNA with a known *TP53* mutation, we show that this assay can detect a variant allele with a frequency of 0.009% (Extended Data Fig. 4b, c).

The specific *TP53* mutation present in the diagnostic t-AML/t-MDS sample was identified in previously banked specimens in four out of the seven cases tested (see Supplementary Notes for case presentations). In

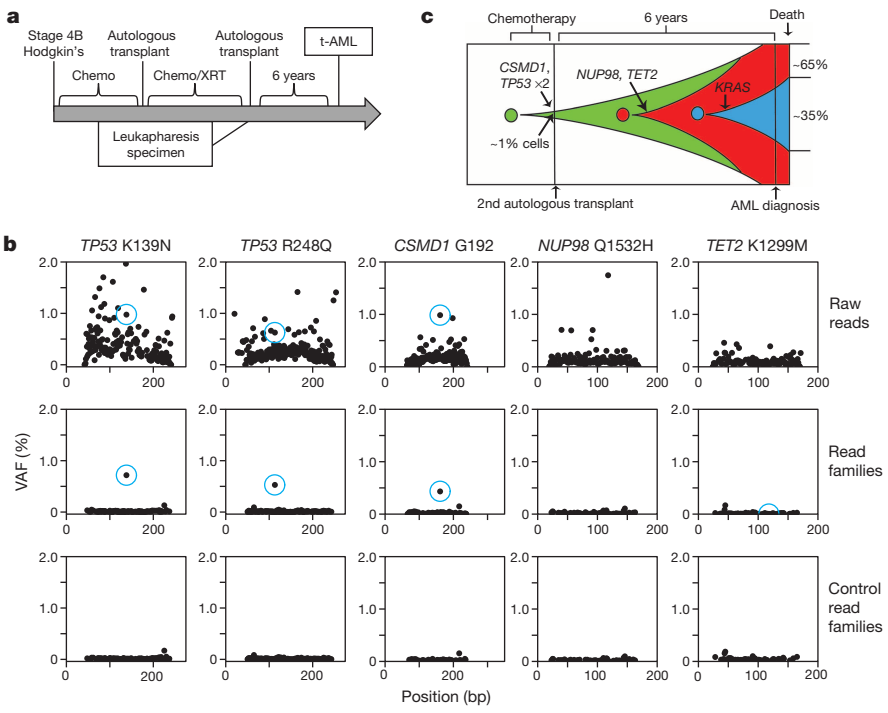


Figure 2 | Biallelic *TP53* mutations are early mutational events in the AML cells of UPN 530447. a, Clinical course of case 530447. Chemo, chemotherapy; XRT, radiation therapy. **b**, Unique adaptor sequencing of a leukapheresis sample obtained 6 years before the diagnosis of t-AML for each of the five clonal somatic SNVs identified in the diagnostic t-AML sample. Genomic DNA from a patient lacking these variants served as a control. Blue circles indicate the position of the variant SNV. **c**, Proposed model of clonal evolution to t-AML in this case.

the other three cases, we were unable to detect the diagnostic *TP53* mutation in the previously banked blood or bone marrow sample; it is not clear whether these mutations were present but below our limit of detection or were truly absent. Patient 530447 developed t-AML after an autologous stem cell transplant for refractory Hodgkin's lymphoma (Fig. 2a). The diagnostic t-AML sample carried biallelic mutations of *TP53*, missense mutations of *TET2* and *NUP98*, a silent mutation of *CSMD1*, and a subclonal *KRAS* mutation. Analysis of a leukapheresis sample obtained 6 years before the development of t-AML revealed that both *TP53* mutant alleles were present with a variant allele fraction (VAF) of approximately 0.5% (Fig. 2b). The *CSMD1* mutation was also present at the same VAF and is probably a passenger mutation. However, two potential driver mutations (*TET2* and *NUP98*) were not detectable in the previously banked sample. Thus, these data show that, in this patient, the biallelic *TP53* mutations preceded the development of t-AML by at least 6 years and antedated the development of the *TET2* and *NUP98* mutations (Fig. 2c). In a second case (unique patient number (UPN) 341666), a heterozygous *TP53* R196* mutation was identified in mobilized peripheral blood leukocytes 3 years before the development of t-MDS at a frequency of 0.1%, preceding the acquisition of a *RUNX1* mutation (Extended Data Fig. 5).

In two of the four cases, the previously banked sample was obtained before the initiation of chemotherapy. Patient 967645 developed t-AML 5 years after the diagnosis of marginal zone lymphoma (Fig. 3a). The diagnostic t-AML sample contained a homozygous *TP53* Y220C mutation. Using a droplet digital polymerase chain reaction (ddPCR) assay, we identified the same *TP53* Y220C mutation in a bone marrow sample obtained before any chemotherapy at a frequency of 0.0027% (average of two independent experiments) (Fig. 3b). We next asked whether other mutations in the diagnostic t-AML sample were also present in this previously banked sample (Supplementary Table 8). We focused on the G155S mutation in *SNAP25*; this mutation is probably non-pathogenic as *SNAP25* is not expressed in AML samples[6]. Indeed, we identified the *SNAP25* G155S mutation in the previously banked bone marrow sample with a similar VAF (0.0029%) as that for *TP53* Y220C (Fig. 3c). Of note, deletion (del)(5q) and del(7q) were subclonal at diagnosis (present in 54% and 38% of metaphases, respectively) (Supplementary Table 1). Collectively, these data provide evidence that an HSPC harbouring a *TP53* Y220C mutation preferentially expanded after chemotherapy with the subsequent acquisition of del(5q) and then del(7q) (Fig. 3d). Of note, we found two other cases of t-AML/t-MDS with clonal *TP53* mutations but subclonal del(5q), del(5) and/or del(7) (UPNs 756582 and 837334, Supplementary Table 1). Together, these data suggest that *TP53* mutations precede the development of these characteristic cytogenetic abnormalities of t-AML/t-MDS.

In a second case, patient 895681 developed t-MDS 5.5 years after the initiation of chemotherapy for non-Hodgkin's lymphoma (Fig. 3e). The diagnostic t-MDS sample contained a clonal *TP53* H179L mutation. Using ddPCR, we identified *TP53* H179L at a VAF of 0.05% in a bone marrow sample taken before the initiation of cytotoxic therapy (Fig. 3f). Thus, as with patient 967645, an HSPC carrying a functional *TP53* mutation was present before cytotoxic therapy exposure, later giving rise to the malignant t-AML/t-MDS clone (Fig. 3g).

To determine whether HSPCs harbouring *TP53* mutations are present in healthy individuals, we analysed peripheral blood leukocytes from 20 elderly (68–89 years old) cancer-free donors who had not received prior cytotoxic therapy. We limited our sequencing to exons 4–8 of *TP53* since the majority of pathogenic mutations in *TP53* are located in these exons. Using our unique adaptor sequencing assay, we identified *TP53* mutations in 9 of 19 evaluable cases, with VAFs ranging from 0.01% to 0.37% (Extended Data Table 4). Of note, since we did not sequence the entire coding region of *TP53*, it is likely that our study underestimates the true frequency of healthy elderly individuals harbouring HSPCs with *TP53* mutations. ddPCR confirmed the presence of the *TP53* mutation in all three cases that were tested (Extended Data Fig. 6). Interestingly, the majority of the *TP53* mutations identified are known pathogenic

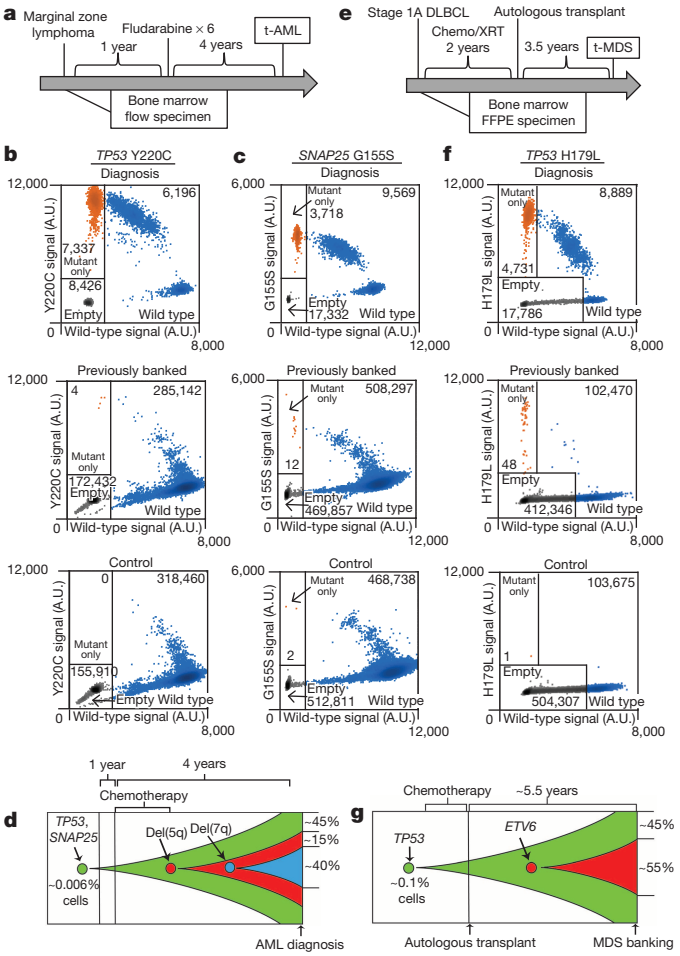


Figure 3 | HSPC clones harbouring somatic *TP53* mutations are detected in patients before cytotoxic therapy exposure. **a**, Clinical course of case 967645. **b**, Dot plots of ddPCR of a diagnostic t-AML sample from case 967645, a bone marrow sample from this patient obtained 5 years before the development of t-AML (before any cytotoxic therapy), or a control sample from a patient lacking a mutation in *TP53*. Droplets containing only the *TP53* Y220C allele are highlighted in orange, droplets containing wild-type *TP53* (with or without *TP53* Y220C) are highlighted in blue; empty droplets are grey. The number of droplets in each gate is indicated. Data are representative of two independent experiments. **c**, Dot plots of ddPCR data for *SNAP25* G155S using the same genomic DNA as in **b**. **d**, Proposed model of clonal evolution to t-AML in case 967645. **e**, Clinical course of case 895681. Chemo, chemotherapy; DLBCL, diffuse large B-cell lymphoma; FFPE, formalin-fixed parafin-embedded; XRT, radiotherapy. **f**, Dot plots of ddPCR data of the diagnostic t-MDS sample from case 895681, a bone marrow FFPE sample from this patient obtained 5.5 years before the development of t-MDS (before any cytotoxic therapy), or a control FFPE sample obtained from a patient lacking a mutation in *TP53*. The labelling scheme is the same as in **b**. **g**, Proposed model of clonal evolution to t-MDS in case 895681; the diagnostic t-MDS sample contained a subclonal *ETV6* mutation.

mutations previously implicated in cancer. These data suggest that functional *TP53* mutations may confer (even in the absence of cytotoxic therapy) a subtle competitive advantage that results in modest HSPC expansion over time.

To test directly the hypothesis that functional *TP53* mutations confer a competitive advantage after chemotherapy, we generated mixed bone marrow chimaeras containing both wild-type and $Tp53^{+/-}$ cells (Fig. 4a). In mice treated with vehicle control, we observed a non-significant trend towards an increased $Tp53^{+/-}$ donor contribution to haematopoiesis (Fig. 4b–e). Whether longer follow-up would confirm a subtle competitive advantage, as suggested by the expansion of *TP53* mutant HSPC clones in elderly healthy individuals, will require additional study. Regardless,

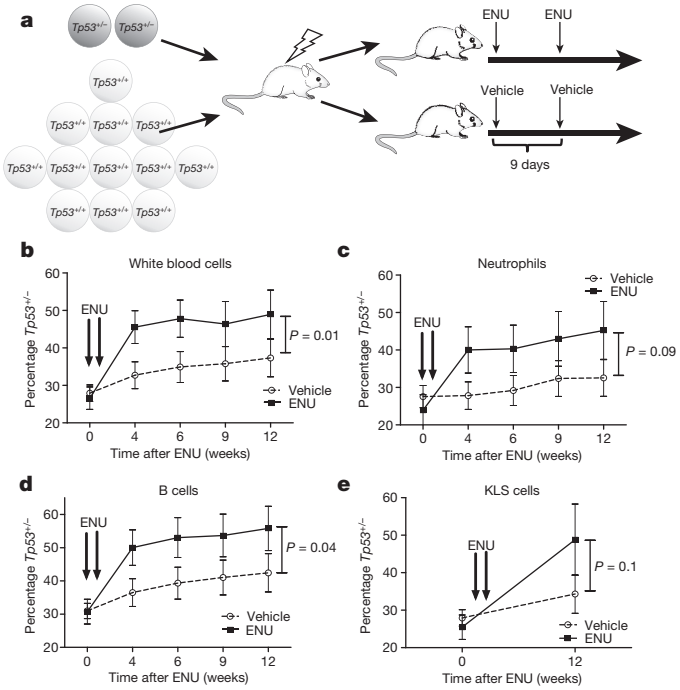


Figure 4 | Heterozygous loss of *TP53* confers a clonal advantage to HSCs after exposure to ENU. **a**, Experimental schema. Bone marrow chimaeras were generated by transplanting a 7 to 1 ratio of wild-type to $Tp53^{+/-}$ bone marrow into irradiated syngenic recipients. After haematopoietic reconstitution (5 weeks), mice were treated with ENU or vehicle control as indicated. **b**–**d**, Shown is the percentage of total leukocytes (**b**), Gr-1^{+} neutrophils (**c**) or B220^{+} B cells (**d**) that were derived from $Tp53^{+/-}$ cells. **e**, Percentage of Kit^{+} lineage^{-} Sca^{+} (KSL) cells in the bone marrow 12 weeks after ENU exposure that were derived from $Tp53^{+/-}$ cells. Data represent the mean ± standard error of the mean (s.e.m.) of 11 and 14 mice in the ENU and vehicle cohorts, respectively. Peripheral chimaerism was analysed using two-way ANOVA and KLS chimaerism was analysed using an analysis of covariance (ANCOVA).

upon treatment with *N*-ethyl-*N*-nitrosourea (ENU), $Tp53^{+/-}$ HSPCs show a competitive advantage. Importantly, a previous study similarly showed that $Tp53^{+/-}$ HSCs also have a competitive advantage after irradiation, which appeared to be due, at least in part, to reduced irradiation-induced senescence in $Tp53^{+/-}$ HSCs[20,21].

There is increasing evidence that cancers undergo clonal evolution under the selective pressure of chemotherapy[22]. For example, the clonal architecture of *de novo* AML is dynamic, with certain (often minor) subclones becoming dominant at relapse after chemotherapy[9]. We show that HSPCs that acquire heterozygous *TP53* mutations as a function of normal ageing are also subject to Darwinian selection upon exposure to cytotoxic therapy, ultimately resulting in the expansion of HSPCs with these mutations. The high frequency (nearly 50%) of elderly individuals with detectable heterozygous *TP53* mutations in their circulating leukocytes far exceeds the prevalence of AML or MDS in this age group. Clearly, additional mutations, including mutation of the second *TP53* allele, are needed for transformation to AML or MDS. Consistent with this observation, only a minority of patients with Li–Fraumeni syndrome, most of whom harbour germline heterozygous *TP53* mutations, develop AML or MDS[23]. This model provides a potential mechanism for the high incidence of *TP53* mutations in t-AML/t-MDS[24]. The *TP53* mutation in the founding clone probably contributes to the frequent cytogenetic abnormalities and poor response to chemotherapy that are typical of t-AML/t-MDS. For t-AML/t-MDS cases that do not harbour *TP53* mutations, it will be important to determine whether different age-related mutations also confer a competitive advantage to HSPCs that are exposed to cytotoxic therapy, and to define the nature of these mutations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 December 2013; accepted 13 October 2014.
Published online 8 December 2014.

1. Godley, L. A. & Larson, R. A. Therapy-related myeloid leukemia. *Semin. Oncol.* **35,** 418–429 (2008).
2. Christiansen, D. H., Andersen, M. K. & Pedersen-Bjergaard, J. Mutations with loss of heterozygosity of p53 are common in therapy-related myelodysplasia and acute myeloid leukemia after exposure to alkylating agents and significantly associated with deletion or loss of 5q, a complex karyotype, and a poor prognosis. *J. Clin. Oncol.* **19,** 1405–1413 (2001).
3. Shih, A. H. *et al.* Mutational analysis of therapy-related myelodysplastic syndromes and acute myelogenous leukemia. *Haematologica* **98,** 908–912 (2013).
4. Kayser, S. *et al.* The impact of therapy-related acute myeloid leukemia (AML) on outcome in 2853 adult patients with newly diagnosed AML. *Blood* **117,** 2137–2145 (2011).
5. Link, D. C. *et al.* Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. *J. Am. Med. Assoc.* **305,** 1568–1576 (2011).
6. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368,** 2059–2074 (2013).
7. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366,** 1090–1098 (2012).
8. Walter, M. J. *et al.* Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* **27,** 1275–1282 (2013).
9. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
10. Petitjean, A. *et al.* Impact of mutant p53 functional properties on *TP53* mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **28,** 622–629 (2007).
11. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150,** 264–278 (2012).
12. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells *in vivo*. *Blood* **117,** 4460–4466 (2011).
13. Carvajal, L. A. & Manfredi, J. J. Another fork in the road—life or death decisions by the tumour suppressor p53. *EMBO Rep.* **14,** 414–421 (2013).
14. Chen, X., Ko, L. J., Jayaraman, L. & Prives, C. p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev.* **10,** 2438–2451 (1996).
15. Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science* **336,** 1440–1444 (2012).
16. Zhang, X. P., Liu, F., Cheng, Z. & Wang, W. Cell fate decision mediated by p53 pulses. *Proc. Natl Acad. Sci. USA* **106,** 12245–12250 (2009).
17. Zhang, X. P., Liu, F. & Wang, W. Two-phase dynamics of p53 in the DNA damage response. *Proc. Natl Acad. Sci. USA* **108,** 8990–8995 (2011).
18. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253,** 49–53 (1991).
19. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
20. Bondar, T. & Medzhitov, R. p53-mediated hematopoietic stem and progenitor cell competition. *Cell Stem Cell* **6,** 309–322 (2010).
21. Marusyk, A., Porter, C. C., Zaberezhnyy, V. & DeGregori, J. Irradiation selects for p53-deficient hematopoietic progenitors. *PLoS Biol.* **8,** e1000324 (2010).
22. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
23. Birch, J. M. *et al.* Relative frequency and morphology of cancers in carriers of germline *TP53* mutations. *Oncogene* **20,** 4621–4628 (2001).
24. Casás-Selves, M. & Degregori, J. How cancer shapes evolution, and how evolution shapes cancer. *Evolution* **4,** 624–634 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Schmidt, B. McKethan and R. Miller for technical assistance, and K. Odell and J. Tucker-Davis for animal care. We thank P. Goodfellow and J. Ivanovich for providing peripheral blood leukocyte genomic DNA from cancer-free individuals. This work was supported by National Institutes of Health grants PO1 CA101937 (D.C.L.) and U54 HG003079 (R.K.W.) and by a grant from the Leukemia & Lymphoma Society (D.C.L.).

Author Contributions T.N.W. and G.R. designed and performed the research, analysed the data, and wrote the manuscript. A.L.Y. and T.E.D. developed and optimized the amplicon-based random primer sequencing assay. C.A.M., D.S., J.H., R.S.F., L.D., E.R.M. and R.K.W. contributed to the generation and analysis of the whole-genome or targeted sequencing. W.T., T.L.L., S.H., J.M.K. and P.W. collected and processed clinical data and tissue samples. J.D.B. performed statistical analyses of the clinical data. J.S.W., J.F.D., M.J.W., T.A.G. and T.J.L. contributed to data analysis. D.C.L. supervised all of the research and edited the manuscript, which was approved by all co-authors.

Author Information Sequence information on the 22 t-AML whole-genome sequencing patients and one exome sequencing patient (UPN 967645) has been deposited in dbGaP under accession number phs000159. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.L. (dlink@dom.wustl.edu).

METHODS

Patient characteristics. For the whole-genome sequencing study, we intentionally selected the original 22 cases of t-AML to have minimal numbers of cytogenetic abnormalities. However, the additional 89 cases of t-AML/t-MDS were randomly selected from those samples with sufficient tumour and skin DNA. All patients were selected from a larger cohort of adult AML and MDS patients enrolled in a single institution tissue banking protocol that was approved by the Washington University Human Studies Committee (WU HSC#01-1014). Written informed consent for whole-genome sequencing was obtained from all study participants. Patients were treated in accordance with National Comprehensive Cancer Network (NCCN) guidelines (http://www.nccn.org) with an emphasis on enrolment in therapeutic clinical trials whenever possible. Clinical data for all patients, including the pre-existing condition requiring cytotoxic therapy, the cytotoxic therapy received before the t-AML/t-MDS diagnosis, cytogenetics, treatment approach and outcomes data, are presented in Extended Data Tables 1 and 2 and Supplementary Table 1. Peripheral blood leukocyte genomic DNA from cancer-free individuals (median age = 75.3 ± 6.6 years) was obtained as part of a Washington University Institutional Review Board-approved protocol. All subjects had no previous history of invasive cancer or treatment with cytotoxic therapy, as determined by the medical history.

Whole-genome sequencing and variant detection. A previously described procedure[25] was followed for library construction and whole-genome sequencing. Briefly, Illumina DNA sequencing was used to generate sequence that covered the haploid reference at a depth between 30.51 and 72.60 (Supplementary Table 9). Sequence data were aligned to reference sequence build NCBI human build 36 using BWA v.0.5.5 (ref. 26) (params: -t 4) then merged and deduplicated using Picard v.1.29. We detected SNVs using the intersection of SAMtools v.r963 (ref. 27) (params: -A -B) and Somatic Sniper v.0.7.3 (ref. 28) (params: -q 1 -Q 15), and filtered to remove false positives (params: min-base-quality 15, min-mapping-quality 40, min-somatic-score 40). Indels were detected using GATK version 5336 (ref. 29) unioned with Pindel v.0.5 (ref. 30). Somatic copy number alterations were detected using copyCat v.1.5 (http://github.com/chrisamiller/copycat). We detected structural variants using BreakDancer v.1.2 (ref. 31) and SquareDancer v.0.1 (https://github.com/genome/genome/blob/master/lib/perl/Genome/Model/Tools/Sv/SquareDancer.pl), followed by assembly with Tigra-SV (https://github.com/genome/tigra-sv). SciClone (in review; http://github.com/genome/sciclone) was used to infer the subclonal architecture of all whole-genome sequencing samples.

Validation and extension sequencing with variant detection. We used custom sequence capture arrays from Roche Nimblegen that targeted variants detected by whole-genome sequencing and extended this array to cover all coding exons from an additional 149 genes of interest (Supplementary Table 4). Libraries were prepared, sequence was generated, and somatic alterations identified as described for whole-genome sequencing with the addition of VarScan v.2.2.6 (ref. 32) (params: -min-var.-freq 0.08-p-value 0.10-somatic-p-value 0.01 -validation) as a variant caller for both SNVs and indels. On average, genes were covered with a depth of 58.3 (Supplementary Table 10). Biallelic *TP53* mutations in case 530447 were confirmed with PCR amplification of the genomic region containing both somatic mutations from the diagnostic t-AML sample. The resulting amplicons were cloned into the pCR-TOPO plasmid vector (Life Sciences) and sequenced using Sanger sequencing.

Statistical analyses. Fisher's exact tests were used to evaluate the association between pairs of dichotomous variables, with a significant right-sided *P* value indicating a positive relationship and a significant left-sided *P* value indicating a negative relationship. The relationship between overall survival and each discrete measure was tested with Kaplan–Meier survival analyses with separate analyses for the AML and MDS groups. Age at diagnosis was discretized into quartiles for each group. Multivariate proportional-hazards regression models were created separately for the AML and MDS groups. All variables with log-rank *P* values of 0.20 or less in the Kaplan–Meier analyses were included in the first step. In successive steps, the variable with the largest *P* value was removed and the model re-run until all remaining variables had *P* values of 0.05 or less. Two-way interactions among the remaining variables were examined. Variables removed in earlier steps were added back to the model one at a time to determine if they significantly improved the final model. The proportionality assumption was evaluated for each variable in the final models.

Rare variant detection using unique adaptor next-generation sequencing. Amplicons approximately 200 bp in length were prepared from patient genomic DNA samples using primers designed to amplify genomic regions harbouring known tumour-specific SNVs (Supplementary Table 11). These amplicons were prepared for next-generation sequencing (NGS) using the Illumina TruSeq DNA Sample Preparation Kit (Illumina Catalog #FC-121-2001) replacing the kit adapters with adapters containing a random nucleotide index sequence. Libraries were quantified using the Agilent qPCR NGS Library Quantification Kit, Illumina GA (Agilent Technologies Catalog #G4880A). Using this quantification, each library was diluted to ensure that each random index would be observed in multiple sequenced reads[33,34]. Each diluted library was amplified and sequenced on the Illumina MiSeq platform. Sequenced reads containing the same index sequence were grouped together creating 'read families' in a manner similar to established methods[33,34]. Reads within a read family were aligned against each other to filter out stochastic sequencing errors generating an error-corrected read family consensus sequence. Each consensus sequence was locally aligned to UCSC hg19/GRCh37 using bowtie2 (ref. 35) with the default settings. The aligned read families were processed with Mpileup[27] using the parameters -BQ0 -d 10000000000000. Next, variants were called with VarScan[32] using the parameters -min-coverage 10000-min-reads2 10-min-avg-qual 0-min-var.-freq 0-p-value 1. Variant allele frequencies for the expected mutations and the background error rate were visualized using IGV[36] and graphically represented using ggplot2 (ref. 37). Variant coordinates are displayed in hg18/GRCh36.

Detection of somatic *TP53* mutations in cancer-free subjects. Amplicons were prepared from healthy control genomic DNA samples using primers designed to amplify exons 4–8 of *TP53* (Supplementary Table 11). Patient-specific barcodes, 6 nucleotides in length, were appended to the 5′ end of each primer to enable pooling of multiple samples for sequencing. Amplicons generated from each *TP53* exon/patient sample combination were generated as previously described and purified products were pooled in equimolar amounts. The pooled barcoded amplicons were prepared for error-corrected sequencing as previously described. Sequencing was completed on the Illumina Hi-Seq 2500 platform. Sequenced reads were demultiplexed based on the known patient-specific barcode sequences using a 2-nucleotide hamming distance. Demultiplexed sequence reads were organized into read families based on their random oligonucleotide index sequence and error-corrected as outlined previously. Read families composed of three reads or more were used for analysis. A binomial distribution of the substitution rate at each covered base in *TP53* was used to identify individuals with somatic *TP53* mutations. A variant was called if it met the following criteria: (1) the binomial *P* value was less than 10^{-6}; (2) the VAF was greater than 1:10,000; (3) at least 10,000 unique read families were sequenced at the position of interest; (4) at least 10 read families called the variant; and (5) the VAF in the individual was greater than five times the mean VAF for all individuals with greater than 10,000× coverage at that specific nucleotide. Read families from one patient sample (barcode GTACGGC) were removed from analysis due to a high error rate. All somatic mutations were identified in this manner except for *TP53* Y220C, which received closer manual inspection due to the large number of these mutations observed in our t-AML cohort.

Extraction of genomic DNA from FFPE samples. Genomic DNA was extracted from FFPE samples with the QIAamp DNA FFPE Tissue Kit. Because of the effects of formalin fixation (cross-linking, DNA fragmentation, and so on), the amount of amplifiable DNA per sample was less than would be expected with Qubit fluorometric quantitation. As such, ddPCR was used to quantify the amount of amplifiable genomic DNA per sample such that the numbers of amplifiable domains tested were comparable between experimental and control samples.

ddPCR. All primers and probes for ddPCR were designed by Bio-Rad as per MIQE guidelines[38]. In the case of *TP53* Y220C, the *TP53* region of interest in exon 6 was amplified with the following primers: 5′-TTTTCGACATAGTGTGGTG-3′ and 5′-CTGACAACCACCCTTAAC-3′. The 5′-Hex/TGCCCTATGAGCCGCCT/Iowa Black FQ-3′ probe was used to detect the wild-type allele and the 5′-FAM/CCCT GTGAGCCGCCTGA/Iowa Black FQ-3′ probe was used to detect the mutant allele. All reagents were purchased from Bio-Rad. ddPCR was performed as previously described[39]. Specifically, quantitative PCR was performed with 900–1,800 nM forward and reverse primers, 250 nM mutant and wild-type genomic probes, and 2–4 ng μl^{-1} genomic DNA. Quantitative PCR was performed with annealing/extension temperatures of 55.5–60 °C for 40 cycles. For droplet generation and analysis, we used the Bio-Rad QX100 and QX200 Droplet Digital PCR Systems.

Owing to the fact that DNA degradation with time (that is, guanosine oxidation, cytosine deamination) is known to interfere with rare allele detection[34], we only identified variant alleles present in droplets also lacking the reference allele. This greatly increased the specificity of our calls by removing droplets in which one of the two DNA strands may have been chemically altered. At low variant allele frequency, it was assumed that only a single variant allele was present in these 'mutant only' drops. Droplet allele distribution follows a Poisson distribution such that the number of droplets only containing a single allele (either variant or reference) can be determined from the percentage of empty droplets. Of note, droplets showing evidence of template independent amplification (that is, observed in 'no template controls') were counted as empty droplets. The VAF was determined from the fraction of the single allele droplets containing the variant allele. When appropriate, control samples were used to subtract potential background signal. VAFs calculated in this method were highly concordant with VAFs obtained through unique-adaptor NGS.

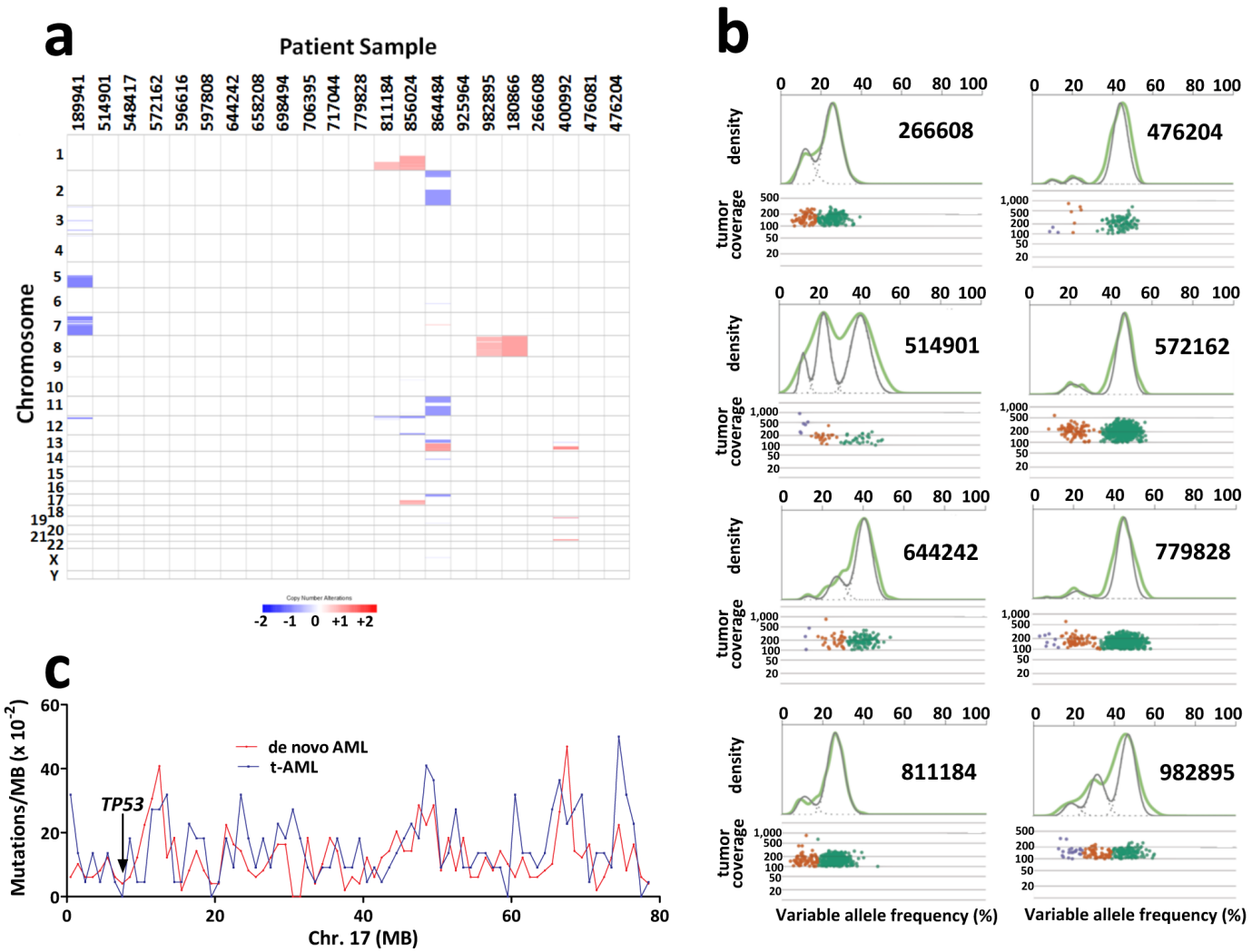
Generation and analysis of *Tp53*$^{+/-}$ bone marrow chimaeras. *Tp53*$^{+/-}$ and wild-type mice were inbred on a C57BL/6 strain. Bone marrow from *Tp53*$^{+/-}$ mice expressing *Ly5.2* was mixed at a 1:7 ratio with bone marrow from wild-type mice expressing *Ly5.1* and transplanted retro-orbitally into lethally irradiated *Ly5.1/5.2*

recipients. $Tp53^{+/-}$ and wild-type donors were both age (6–12 weeks) and sex matched (female). A total of 3×10^6 cells were injected per recipient mice. Recipient mice were conditioned with 1,000–1,100 cGy from a ^{137}caesium source at a rate of approximately 95 cGy min^{-1} before transplantation. Prophylactic antibiotics (trimethoprim-sulfamethoxazole; Alpharma) were given during the initial 2 weeks after transplantation. Five weeks after transplantation, mice were given two doses of ENU (100 mg kg^{-1}; Sigma-Aldrich) or vehicle alone intraperitoneally 9 days apart. Mice were stratified according to $Tp53^{+/-}$ chimaerism and then randomly distributed into the ENU and vehicle controls such that both cohorts had similar levels of $Tp53^{+/-}$ chimaerism at baseline. ENU and placebo were delivered in a final solution with 10% DMSO, 90 mM sodium citrate, and 180 mM sodium phosphate, pH 5.0. Peripheral blood chimaerism was measured before ENU administration and 4–12 weeks after ENU administration. The investigator was not blinded. Mice were euthanized and bone marrow chimaerism analysed 12 weeks after ENU administration. The desired cohort size was determined based on observations from previously reported experiments[20], and two independent experiments were performed. Mice were maintained under standard pathogen-free conditions according to methods approved by the Washington University animal studies committee.

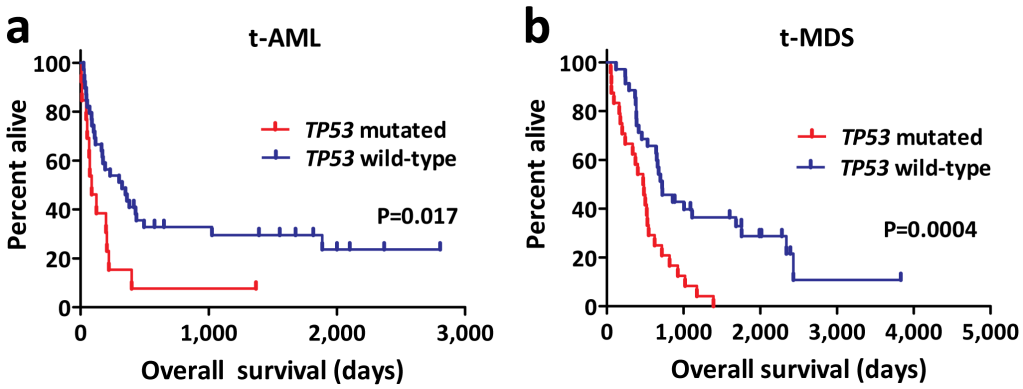
Flow cytometry. Flow cytometry data were collected on a Gallios 10-colour, 3-laser flow cytometer (Beckman Coulter) and analysed with FlowJo software (Treestar). Cells were stained by standard protocols with the following antibodies (eBiosciences unless otherwise noted): Ly5.1 (A20, CD45.1), Ly5.2 (104, CD45.2), Ly6C/G (RB6-8C5, Gr-1), CD3e (145-2C11), CD45R (RA3-6B2, B220), CD11c (N418), TER-119, CD41 (MWReg30), CD117 (ACK2, c-Kit) and Ly-6A/E (D7, Sca).

Estimation of *TP53* mutation frequency in ageing stem cells. The frequency and profile of somatic single-nucleotide mutations in the HSCs of normal individuals have been previously measured[11]. The somatic mutational burden is ageing-related, and the estimated rate of mutagenesis obtained from this study was 3.2×10^{-9} mutations per nucleotide per year (95% confidence interval $2.4–4.0 \times 10^{-9}$) for the average nucleotide in the exome. Thus, we would predict an average 50 year old to have 1.6×10^{-7} mutations per position. These mutations would not be randomly distributed but biased (in particular towards C to T/G to A transitions). It has been previously proposed that an individual possesses approximately 10,000 distinct HSCs[12]. We used a randomized Monte Carlo simulation to model the prevalence of somatic single-nucleotide mutations in healthy 50 year olds with 10,000 HSCs given a normal somatic mutational profile and mutation rate. Repeated simulation ($n = 100{,}000$) allowed us to predict the distribution of ageing-induced *TP53* (NM_000546) somatic mutations. As expected, this simulation modelled a Poisson process. We classified *TP53* mutations as likely to be functional if they fulfilled both of the following criteria. First, we analysed the mutations using the SIFT program (http://sift.jcvi.org) and required a SIFT score ≤ 0.05. Second, we required that the somatic mutations be reported at least once if a nonsense mutation or at least twice if a missense mutation in the International Agency for Research on Cancer *TP53* database (http://p53.iarc.fr). On the basis of this simulation, we predict that 44% of 50-year-old individuals harbour one or more HSCs with a functional *TP53* mutation.

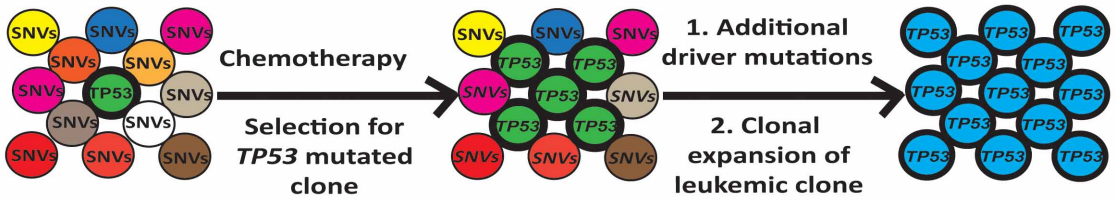
25. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361,** 1058–1066 (2009).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
27. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
28. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28,** 311–317 (2012).
29. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
30. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865–2871 (2009).
31. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6,** 677–681 (2009).
32. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22,** 568–576 (2012).
33. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108,** 9530–9535 (2011).
34. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109,** 14508–14513 (2012).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
36. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).
37. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
38. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55,** 611–622 (2009).
39. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83,** 8604–8610 (2011).



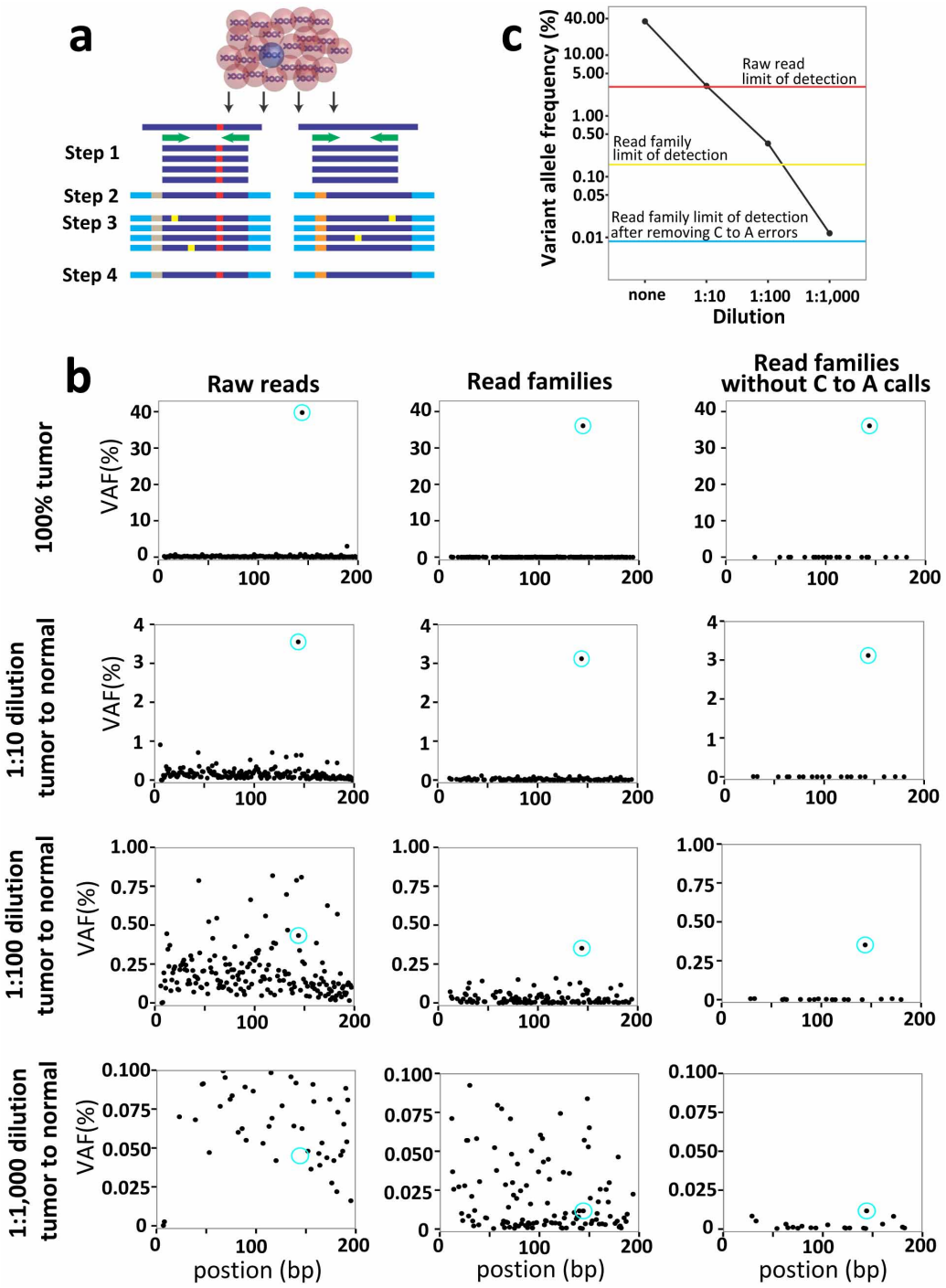
Extended Data Figure 1 | Whole-genome sequencing analysis of t-AML. **a**, Somatic copy number alterations in the 22 cases of t-AML. Blue indicates copy number loss; red indicates copy number gain. **b**, Representative clonality plots for 8 cases of t-AML are shown. Scatter plots (bottom) show variant allele frequency and read depth in the tumour sample. Variant alleles in the founding clone are depicted in green, while variants in subclones are depicted in orange or purple. Top, kernel density plots of the VAF data (green line) along with the posterior predictive densities (grey line) from the mathematical model used to segregate clusters. **c**, Frequency of tier 1 silent, tier 2, and tier 3 mutations in 1 Mb increments across chromosome 17 in *de novo* AML and t-AML. The *TP53* genomic locus is identified.



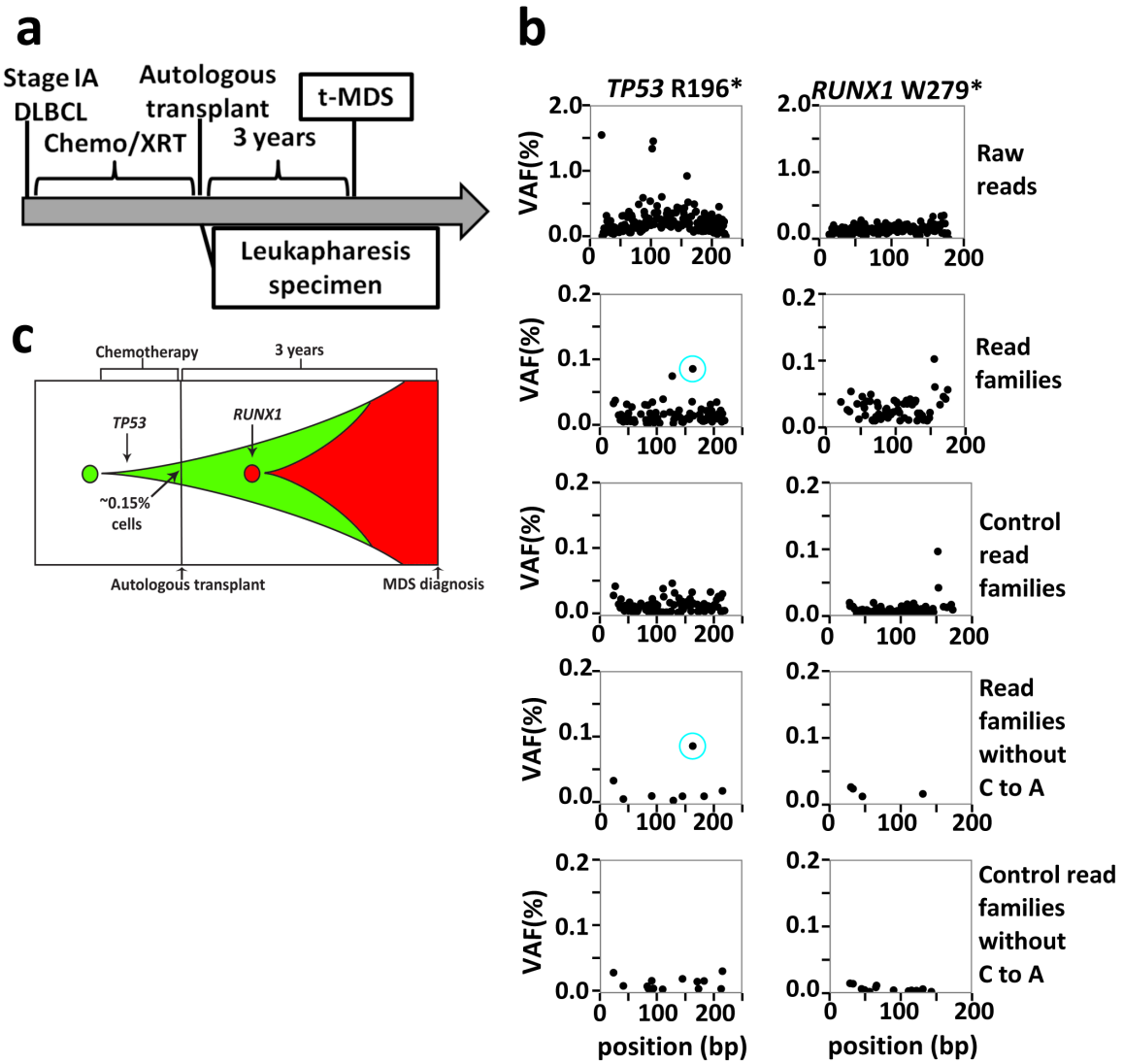
Extended Data Figure 2 | *TP53* mutations are associated with decreased overall survival in t-AML/t-MDS. **a**, Overall survival in *TP53* mutated ($n = 13$) and *TP53* wild-type ($n = 39$) t-AML patients. **b**, Overall survival in *TP53* mutated ($n = 24$) and *TP53* wild-type ($n = 35$) t-MDS patients.



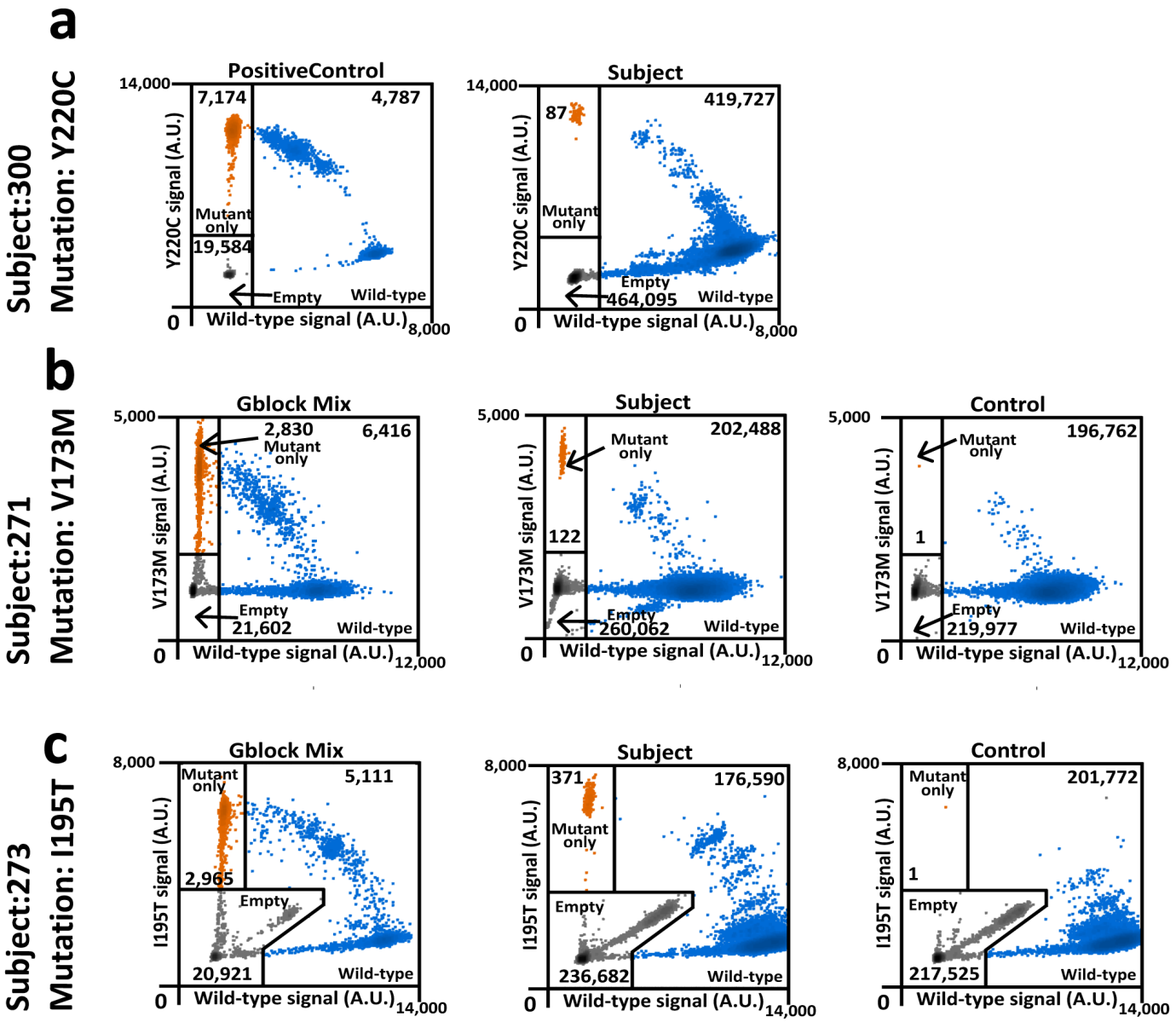
Extended Data Figure 3 | Model of how cytotoxic therapy shapes clonal evolution in t-AML/t-MDS. Age-related mutations in HSPCs result in the production of a genetically heterogeneous population of HSPCs, including rare HSPCs with heterozygous *TP53* mutations in some individuals. During chemotherapy and/or radiotherapy for the primary cancer, HSPC clones harbouring a *TP53* mutation have a competitive advantage, resulting in expansion of that clone. Subsequent acquisition of additional driver mutations results in transformation to t-AML/t-MDS. Of note, the presence of *TP53* mutations probably accounts for the high incidence of cytogenetic abnormalities in t-AML/t-MDS and poor response to chemotherapy.



Extended Data Figure 4 | Validation of the unique adaptor sequencing method. **a**, Unique adaptor sequencing approach. Step 1: genomic DNA is amplified with *TP53*-specific primers (green) with subpopulation-specific variant alleles highlighted in red. Step 2: randomly indexed adapters (tan and grey) are ligated to each amplicon. Step 3: the indexed amplicons are amplified to generate multiple reads possessing the same barcode (that is, read families). Step 4: after sequencing, reads are aligned and grouped by read families to generate an error-corrected consensus sequence. Sequencing errors (yellow) are randomly distributed amongst read families, while true variant alleles (red) are present in all members of a given read family. **b**, A tumour sample (UPN 895681) with a known *TP53* somatic mutation (chromosome 17: 7519119 T to A) at a VAF of ~37% was mixed with normal genomic DNA sample at the indicated ratio, and conventional (left) or unique adaptor next-generation sequencing (middle and right) was performed, as described in Methods. DNA degradation with time may result in errors that are then amplified during PCR, providing a source of false-positive calls. This is particularly true for C to A transversions. Since none of the *TP53* mutations analysed in this study were C to A transversions, we also analysed the data after removing C to A calls (right). The *TP53* variant allele is circled in blue. **c**, The threshold of detection for the variant allele with each sequencing method is shown.



Extended Data Figure 5 | Clonal evolution in case 314666. **a**, Clinical course of case 341666. Chemo, chemotherapy; DLBCL, diffuse large B-cell lymphoma; XRT, radiotherapy. **b**, Unique adaptor sequencing was performed on genomic DNA derived from leukapharesis samples obtained 3 years before the diagnosis of t-MDS for the two clonal mutations present in the diagnostic t-MDS sample. Genomic DNA from a patient lacking these variants was used as a control. The blue circle indicates the position of the variant SNV. **c**, Proposed model of clonal evolution to t-MDS in this case.



Extended Data Figure 6 | ddPCR verification of selected somatic *TP53* mutations identified in peripheral blood of cancer-free individuals. **a**–**c**, ddPCR was performed on genomic DNA isolated from the peripheral blood of cancer-free individuals (middle) for whom unique-read adaptor sequencing suggested the presence of the indicated *TP53* mutation. Controls represent peripheral blood DNA from cancer-free elderly individuals with VAFs not above background levels for the mutation of interest (right); the negative control for *TP53* Y220C is shown in Fig. 3b. **a**, The diagnostic t-AML sample from patient 967645 was used as a positive control for *TP53* Y220C. **b**, **c**, For *TP53* V173M (**b**) and *TP53* I195T (**c**) double-stranded genomic blocks (gBlocks) were synthesized containing the mutation of interest and mixed with gBlocks of wild-type sequence. Droplets containing only the variant *TP53* allele are highlighted in orange, droplets containing the wild-type *TP53* allele (with or without the variant *TP53* allele) are highlighted in blue; empty droplets are grey. The number of droplets in each gate is indicated.

Extended Data Table 1 | Clinical summary of the 22 t-AML whole-genome sequencing cases

Age	Median	56.5 years (26-80)
Gender	Male	36.4%
	Female	63.6%
Prior Disease	Breast	45.5%
	Non-Hodgkin's Lymphoma	13.6%
	Multiple Sclerosis	9.1%
Known Previous Treatment	Other	31.8%
	Alkylator	45.5%
	Topoisomerase inhibitor	63.6%
	Radiation	68.2%
	Autologous Transplant	9.1%
Latency	Median	3.2 years (0.9-13.3)
Cytogenetics	complex	22.7%
	MLL rearrangement	22.7%
	non-complex non-MLL	63.6%
% Blasts in the bone marrow	Median	79% (19-95%)
Most intensive t-AML/t-MDS treatment regimen	Allogeneic transplant	27.3%
	Myeloablative	40.9%
	Non-myeloablative	13.6%
	Other/unknown	18.2%
Remisison	Yes	50%
	No	40.9%
Overal Survival	Median	140.5 days (8-2000)

Latency is defined as the time from the original cancer diagnosis to the development of t-AML/t-MDS.

Extended Data Table 2 | Clinical summary of the combined 111 t-AML/t-MDS cases

Age	Median	61 years (18-82)
Gender	Male	53.2%
	Female	46.8%
Prior Disease	Breast	29.7%
	Non-Hodgkin's Lymphoma	24.7%
	Hodgkin's Disease	22.5%
	Other	37.8%
Known Previous Treatment	Alkylator	55.9%
	Topoisomerase inihbitor	50.5%
	Radiation	63.1%
	Autologous Transplant	21.6%
Latency	Median	6.25 years (0.4-40.7)
Diagnosis	Leukemia	46.8%
	MDS	53.2%
Cytogenetics	deletion 5	26.1%
	deletion 7	28.8%
	complex	45.0%
	MLL rearrangement	5.4%
	other/unknown	41.4%
% Blasts in the bone marrow	Median	13% (0-95%)
Most intensive AML treatment regimen	Allogeneic transplant	41.4%
	Myeloablative	17.1%
	Non-myeloablative	24.3%
	Other/unknown	17.1%
Remisison	Yes	49.5%
	No	43.2%
Overal Survival	Median	414 days (8-3831)

Latency is defined as the time from the original cancer diagnosis to the development of t-AML/t-MDS.

Extended Data Table 3 | Previously banked tissue samples in patients with t-AML/t-MDS with clonal *TP53* mutations

	TP53 mutation in t-AML/t-MDS			Prior banked tissue sample		
Patient	Position (Chr 17)	Mutation	Coding change	Year of Banking	Year of Diagnosis	Prior Banked Tissue
236041	7,518,261	T to A	R249W	2007	2011	BM FFPE
341666	7,518,988	G to A	R196*	2002	2005	Pharesis
530447	7,519,238	C to G	K139N	2001	2007	Pharesis
530447	7,518,263	C to T	R248Q	2001	2007	Pharesis
648904	7,514,759	C to T	Exon 9 splice site	2001	2004	Pharesis
756582	7,519,015	C to T	Exon 6 splice site	1999	2007	Pharesis
895681	7,519,119	T to A	H179L	2000	2006	BM FFPE
967645	7,518,915	T to C	Y220C	2005	2010	BM Flow

All patients had one or more clonal *TP53* mutations in their diagnostic t-AML/t-MDS samples (530447 had biallelic mutations). Cases in which the previously banked sample had detectable *TP53* mutated cells are highlighted in red. See Supplementary Table 1 for the clinical and molecular features of these cases. BM FFPE, formalin-fixed parafin-embedded sample; BM flow, snap-frozen bone marrow leukocyte pellet.

Extended Data Table 4 | Somatic *TP53* mutations in 19 cancer-free individuals

Sample	Chr	Exon	Start	Stop	Ref	Var	Amino acid	COSMIC ID	Var count	Total read family count	VAF (read-family	VAF (ddPCR)
34	17	7	7518230	7518230	T	G	D259A	none	13	33085	0.039%	N.D.
99	17	7	7518273	7518273	C	T	G245S	COSM6932	18	41836	0.043%	N.D.
99	17	8	7517849	7517849	C	T	V272M	COSM10891	26	81015	0.032%	N.D.
269	17	8	7517845	7517845	C	T	R273H	COSM10660	489	420026	0.12%	N.D.
271	17	5	7519138	7519138	C	T	V173M	COSM11084	177	182809	0.097%	0.081%
271	17	5	7519174	7519174	C	T	A161T	COSM10739	25	164591	0.015%	N.D.
271	17	NA	7520035	7520035	A	T	SPLICING	COSM1522474	23	165672	0.014%	N.D.
271	17	NA	7517934	7517934	C	T	INTRONIC	none	36	333996	0.011%	N.D.
273	17	6	7518990	7518990	A	G	I195T	COSM11089	57	15540	0.37%	0.28%
300	17	6	7518915	7518915	T	C	Y220C	COSM10758	91	316765	0.029%	0.029%
324	17	8	7517819	7517819	G	A	R282W	COSM10704	51	86090	0.059%	N.D.
335	17	7	7518264	7518264	G	C	R248G	COSM11564	245	218077	0.11%	N.D.
338	17	7	7518264	7518264	G	A	R248W	COSM10656	188	51001	0.37%	N.D.

Coverage statistics are as follows. In the amplicon targeting exon 4, 17/19 subjects had >10,000 coverage in 100% of the amplicon. In the amplicon targeting exon 5, 17/19 subjects had >10,000 coverage in 100% of the amplicon. In the amplicon targeting exon 6, 5/19 subjects had >10,000 coverage in 100% of the amplicon, and 11/19 subjects had >10,000 coverage in at least 75% of the amplicon. In the amplicon targeting exon 7, 17/19 subjects had >10,000 coverage in 100% of the amplicon. In the amplicon targeting exon 8, 18/19 subjects had >10,000 coverage in 100% of the amplicon. See Supplementary Table 11 for the primers used to make the amplicons from genomic DNA. N.D., not determined.