

# UCLA

## UCLA Previously Published Works

### Title

Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD).

### Permalink

<https://escholarship.org/uc/item/2xk7r93t>

### Journal

Leukemia, 31(1)

### ISSN

0887-6924

### Authors

Sun, Q-Y  
Ding, L-W  
Tan, K-T  
[et al.](#)

### Publication Date

2017

### DOI

10.1038/leu.2016.160

Peer reviewed

## ORIGINAL ARTICLE

## Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD)

Q-Y Sun<sup>1,10</sup>, L-W Ding<sup>1,10</sup>, K-T Tan<sup>1,10</sup>, W Chien<sup>1</sup>, A Mayakonda<sup>1</sup>, D-C Lin<sup>1,2</sup>, X-Y Loh<sup>1</sup>, J-F Xiao<sup>1</sup>, M Meggendorfer<sup>3</sup>, T Alpermann<sup>3</sup>, M Garg<sup>1</sup>, S-L Lim<sup>1</sup>, V Madan<sup>1</sup>, N Hattori<sup>1</sup>, Y Nagata<sup>4</sup>, S Miyano<sup>5,6</sup>, AEJ Yeoh<sup>1</sup>, H-A Hou<sup>7</sup>, Y-Y Jiang<sup>1</sup>, S Takao<sup>1</sup>, L-Z Liu<sup>1</sup>, S-Z Tan<sup>1</sup>, M Lill<sup>2</sup>, M Hayashi<sup>8</sup>, A Kinoshita<sup>8</sup>, HM Kantarjian<sup>9</sup>, SM Kornblau<sup>9</sup>, S Ogawa<sup>4</sup>, T Haferlach<sup>3</sup>, H Yang<sup>1</sup> and HP Koefler<sup>1,2</sup>

Partial tandem duplication of *MLL* (*MLL-PTD*) characterizes acute myeloid leukemia (AML) patients often with a poor prognosis. To understand the order of occurrence of *MLL-PTD* in relation to other major AML mutations and to identify novel mutations that may be present in this unique AML molecular subtype, exome and targeted sequencing was performed on 85 *MLL-PTD* AML samples using HiSeq-2000. Genes involved in the cohesin complex (*STAG2*), a splicing factor (*U2AF1*) and a poorly studied gene, *MGA* were recurrently mutated, whereas *NPM1*, one of the most frequently mutated AML gene, was not mutated in *MLL-PTD* patients. Interestingly, clonality analysis suggests that *IDH2/1*, *DNMT3A*, *U2AF1* and *TET2* mutations are clonal and occur early, and *MLL-PTD* likely arises after these initial mutations. Conversely, proliferative mutations (*FLT3*, *RAS*), typically appear later, are largely subclonal and tend to be unstable. This study provides important insights for understanding the relative importance of different mutations for defining a targeted therapeutic strategy for *MLL-PTD* AML patients.

Leukemia (2017) 31, 1–10; doi:10.1038/leu.2016.160

## INTRODUCTION

Acute myeloid leukemia (AML) is a highly malignant hematopoietic disorder affecting annually ~13 000 adults in the United States.<sup>1,2</sup> It is a heterogeneous group of diseases classified by distinct morphologies, karyotypes and molecular subtypes. A particularly interesting AML subtype is the AML cases with partial tandem duplication of the *MLL* gene (*MLL-PTD*), characterized by the internal partial tandem duplication of exons 3–9 or 3–11 in the *MLL* gene. *MLL-PTD* often occurs in elderly patients and consists of 3–5% of *de novo* AML. In particular, patients carrying *MLL-PTD* often have a bad prognosis.<sup>3–6</sup> *MLL-PTD* is generally believed to act as an oncogenic driver by modulating expression of *HOX* genes. However, mice carrying *Mll-PTD* alone do not develop spontaneous leukemia,<sup>7</sup> unless they are crossed with those harboring another major leukemogenic driver (for example, *Flt3-ITD*),<sup>8</sup> suggesting that by itself *Mll-PTD* is not sufficient to transform hematopoietic cells. In addition, the mutational landscape of *MLL-PTD* AML has not been fully explored as this subtype constitutes only 9 of the 200 TCGA-AML samples.<sup>9,10</sup> Identification of collaborating driver mutations and, subsequently, the underlying molecular mechanisms driving this AML subtype will enhance our understanding of this devastating disease.

AML is viewed as a disease formed by sequential acquisition of a number of mutations in long-lived, self-renewing hematopoietic stem cells.<sup>10–12</sup> Investigators have recently proposed that the order of mutational acquisition has a substantial impact on clinical response in myeloid neoplasms.<sup>13</sup> In addition, the order of acquiring mutations has profound implications for the design of

therapeutic strategies, where the greatest therapeutic benefits are most likely gained by targeting the earliest initiating mutation. In the context of *MLL-PTD* AML, deciphering the evolutionary timeline of the acquisition of this gene is important. A strategy based on DOT1L inhibition was recently proposed for treating *MLL*-rearranged (*MLL* fusion/*MLL-PTD*) leukemia,<sup>14,15</sup> and the effectiveness of this strategy is reliant on an underlying assumption that the alteration of *MLL* occurs as an early initiating, clonal mutation. To examine the ordering of different mutations in this rare but clinically important AML subtype and to identify novel uncharacterized mutations, we performed whole-exome and targeted sequencing on 85 *MLL-PTD* AML patients.

## MATERIALS AND METHODS

## Exome sequencing and target capture sequencing

AML samples were purified using Ficoll paque to enrich for mononuclear cells. DNA samples were processed for whole-exome sequencing using SureSelect XT or XT2 (for targeted sequencing) Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol. Sequence data were generated on Illumina HiSeq-2000 platform. Sequence reads were aligned to human reference genome hg19 for both whole-exome and targeted sequencing libraries using Novoalign V3.01.02 (<http://www.novocraft.com/>) with parameter -F ILMFQ to reflect properly base sequencing quality of our data sets. SAMtools (Version 0.1.19) was used to remove PCR duplicates.

<sup>1</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore; <sup>2</sup>Division of Hematology/Oncology, Cedars-Sinai Medical Center, UCLA School of Medicine, Los Angeles, CA, USA; <sup>3</sup>MLL Munich Leukemia Laboratory, Munich, Germany; <sup>4</sup>Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; <sup>5</sup>Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; <sup>6</sup>Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; <sup>7</sup>Division of Hematology, National Taiwan University Hospital, National Taiwan University, Taiwan; <sup>8</sup>Department of Pathology, Keio University School of Medicine, Tokyo, Japan and <sup>9</sup>Department of Leukemia, University of Texas, MD Anderson Cancer Center, Houston, USA. Correspondence: Dr L-W Ding, Cancer Science Institute of Singapore, National University of Singapore, #13 North Core, MD6, Singapore 117599, Singapore. E-mail: csidlw@nus.edu.sg

<sup>10</sup>These authors contributed equally to this work.

Received 2 December 2015; revised 21 May 2016; accepted 25 May 2016; accepted article preview online 8 June 2016; advance online publication, 8 July 2016

### Somatic SNV detection (with matched remission samples)

To identify somatic single-nucleotide variants (SNV) in patients with matched remission samples, software MuTect (Version 1.1.4)<sup>16</sup> was utilized with default parameters. Annotation files corresponding to COSMIC database (hg19-cosmic-v54-120711.vcf) and dbSNP database (dbsnp-132-b37.leftAligned.vcf) were retrieved from MuTect website (<http://www.broadinstitute.org/cancer/cga/mutect-download>) and utilized with MuTect software for variant calls. Identified SNV passing the filtering criteria were retained and used for downstream analysis.

### Generation of false SNV list from panel of normal samples

Artifact detection mode of Mutect program was utilized to generate a list of potential sequencing artifacts, mapping errors and common SNPs from our in-house database of 240 normal samples (manually curated). Variants that were positively identified in the normal samples by Mutect were compiled to generate a putative list of false somatic SNVs. Somatic SNVs detected from each normal-tumor pair were discarded if they were similarly found in this list of false somatic SNVs.

See Supplementary data for a full description of Materials and Methods.

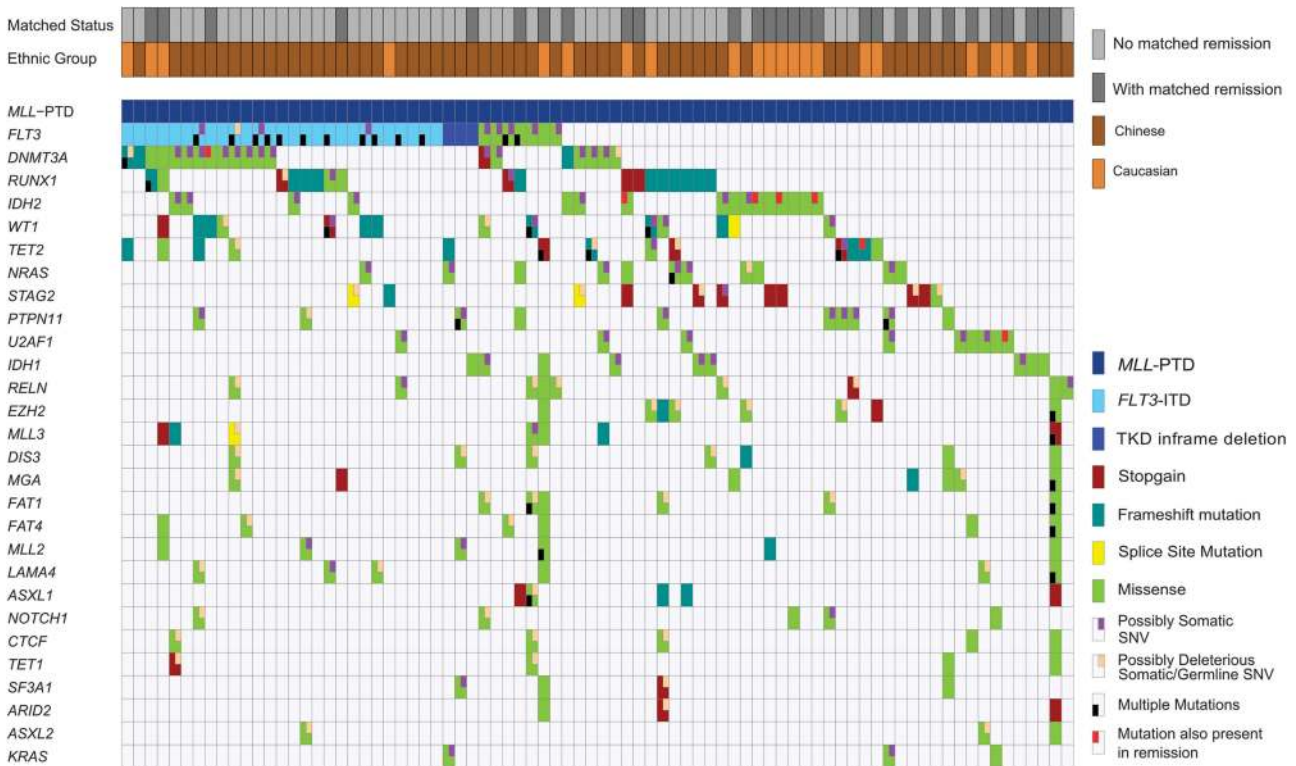
## RESULTS

### Mutational landscape of MLL-PTD AML

To generate a mutational spectrum for MLL-PTD AML, we initially searched for cooperative driving events by performing exome sequencing of genomic DNA of five MLL-PTD AML patients at diagnosis, relapse and at complete remission (Supplementary Table 1). Driver mutations (for example, *FLT3*, *DNMT3A*) were clearly identified in all patients (Supplementary Table 2). The analysis of the mutational spectrum indicates the dominance of C>T transitions, which is in line with the finding of TCGA-AML and is likely caused by spontaneous deamination of methylcytosine<sup>17</sup> related with aging (Supplementary Figure 1a).

The potential clonal architecture and clonal evolution were subsequently inferred from the mutational data (Supplementary Figures 1b and c). All of the patients shared a similar clonal evolutionary pattern: despite achieving complete clinical hematologic remission, the original founder clone of AML cells or the pre-malignant clone survived chemotherapy, gained additional mutations and reemerged at relapse (Supplementary Figure 1c).

We next extended our study to a larger group of MLL-PTD AML patients (prevalence cohort). Mutated genes identified from the discovery cohort and selected leukemia/cancer-related genes (Supplementary Table 3, 530 genes) were chosen for targeted sequencing on an additional 80 AML patients carrying the MLL-PTD mutation (Supplementary Tables 4–6). A 259X mean sequencing read depth was achieved for the prevalence cohort (Supplementary Figure 2). A dominant C>T transition mutation pattern was similarly found in the prevalence cohort, and mutation signature 1b was identified as the dominant mutational signature (Supplementary Figure 3, Alexandrov *et al.*<sup>18</sup>). At least one well-known oncogenic driver mutation was identified in over 90% of the MLL-PTD patients (Figure 1, Supplementary Figure 4). Recurrent hotspot mutations were noted in *IDH1/2*, *U2AF1*, *NRAS* and *FLT3*, suggesting a potential gain-of-function effect of these alterations. On the other hand, mutations of *RUNX1*, *TET2*, *WT1*, *STAG2* and *ASXL1* were largely scattered throughout these genes and tended to be inactivating, implying their potential tumor suppressor roles in AML. Compared with the mutational landscape of 200 TCGA-AML samples,<sup>9</sup> several well-known leukemic drivers (Figure 2, Supplementary Table 7) as well as a few novel genes like *MGA* (9%) and members of the NOTCH family were enriched in MLL-PTD-positive samples (Supplementary Figure 5). Remarkably, no mutations of *NPM1* were found in our patient cohort, although it is one of the most frequently mutated AML genes identified by TCGA (27%).<sup>9</sup> The frequently mutated genes that we identified



**Figure 1.** Mutational landscape of 80 MLL-PTD patients. Types of alterations are color-coded (lower right side of the figure). *FLT3*-ITD: *FLT3* internal tandem duplication. TKD: *FLT3* tyrosine kinase domain mutations. For patients with diagnosis and relapse samples, only the diagnosis mutations are shown (mutations of complete diagnosis-relapse trio patients are shown in Supplementary Figure 4). Only canonical hotspots or fs/stop-gain mutations found in both diagnosis and remission were considered as 'mutation also present at remission'.

Mutant Genes	Frequency in MLL-PTD (%)	Frequency in TCGA AML (%)	MLL-PTD vs TCGA AML
<i>FLT3</i>	46	28	**
<i>IDH1/2</i>	31	20	
<i>DNMT3A</i>	25	26	
<i>RUNX1</i>	23	10	*
<i>NRAS/KRAS</i>	16.3	12	
<i>STAG2</i>	16.3	3	***
<i>TET2</i>	16.3	8	
<i>WT1</i>	16	6	**
<i>PTPN11</i>	13	4	*
<i>U2AF1</i>	12.5	4	*
<i>EZH2</i>	10	1.5	**
<i>ASXL1</i>	6	2.5	
<i>KIT</i>	5	4	
<i>NPM1</i>	0	27	***

\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

**Figure 2.** Differing rates of mutation in AML: our MLL-PTD cohort versus AMLs from TCGA. Comparison of the mutation rate of our 80 MLL-PTD AMLs and 200 AMLs from TCGA (includes nine MLL-PTD samples). Mutated *NPM1* gene was found only in TCGA AMLs.

were classified into several categories: epigenetic regulators, RAS-RTK pathway, transcriptional factors, cohesin complex and splicing factors.

#### Mutations of epigenetic regulators

In line with earlier sequencing studies of other AML subtypes and the TCGA-AML-sequencing project,<sup>9</sup> *DNMT3A* was the most often mutated epigenetic regulator (25%): the well-known hotspot mutation R882H was found in 12 patients; S714C/F mutation was detected in three patients. A dominant-negative role of these missense mutations has been found and a tumor-suppressor role of *DNMT3A* has been recently proposed.<sup>19–21</sup> In addition, frameshift and stop-gain *DNMT3A* mutations were found in six patients. *IDH1/IDH2* hotspot mutations were identified in 31% of patients (R140Q (13 cases) or R172K (3 cases) in *IDH2* and at R132 in *IDH1* (4 cases)). The TET family was the third most prominently mutated epigenetic regulator (*TET1* (5%), *TET2* (16.3%, six frame-shift and three stop-gain)). In line with previous observations,<sup>3</sup> mutually exclusive mutational patterns were noted between *IDH1/2* and *TET1/2* (Supplementary Figure 6). Mutations of epigenetic regulators also occurred in polycomb-associated proteins (*EZH2*, *ASXL* family members, Supplementary Figure 7), chromatin remodelers (*ARID2*, *ARID1A*), genes associated with histone acetylation (*CREBBP*, *EP300*, *KAT6A*, *KAT6B*) and histone methylation (*MLL2*, *MLL3*, *MLL4*).

Mutations in the RAS-FLT3 pathway are subclonal and tend to be unstable

The proliferation-related pathway was extensively mutated, with 54 of 80 MLL-PTD patients (67.5%) carrying at least one mutation of the proliferative genes (Figure 3a). Specifically, *FLT3* mutations were found in 46% of patient samples. These mutations included: #1, internal tandem-duplications (ITD) in exon 14 (41 ITD mutations in 27 patients, Supplementary Table 8); #2, well-known hotspot mutations located in the kinase domain (D835/D839); #3, recurrent in-frame (p.836\_837) deletion in the kinase domain (Figure 3b); #4, novel missense mutations in the juxta-membrane domain. Notably, some FLT3-ITD patients had more than one type of ITD insertion, which probably reflects the existence of multiple subclones in their leukemia (Supplementary Table 8). The presence of several ITD mutations in different subclones

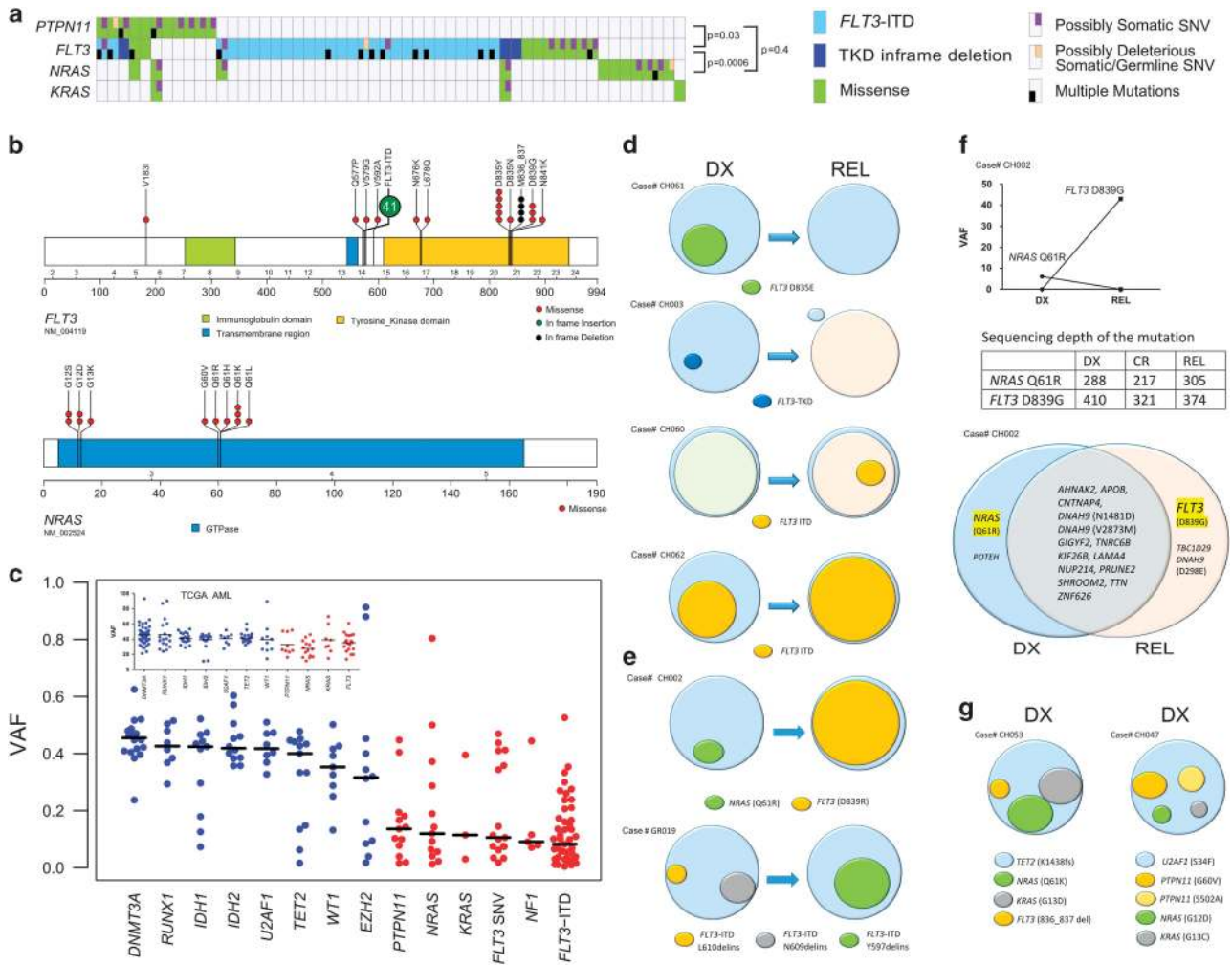
signifies the prominent role of FLT3-ITD in accelerating clonal expansion.

Importantly, mutations in *FLT3* and other proliferation-related genes were largely subclonal with median variant allele-frequencies (VAFs) ranging from ~0.10–0.15 (Figure 3c). Comparatively, mutations of epigenetic regulators (*DNMT3A*, *IDH1*, *IDH2*, *TET2*), transcriptional factors (*RUNX1*, *WT1*), as well as a splicing factor *U2AF1* tended to have relatively high VAF (~0.35–0.45, Figure 3c, Supplementary Figure 8), suggesting they occurred earlier during leukemic evolution. Notably, clones carrying *FLT3* mutations were often found to be unstable and frequently expanded or diminished during progression of disease (6 of 13 patients who had matched relapse samples for analysis, Figure 3d, Supplementary Figure 9). Interestingly, mutations driving proliferation also appeared to be interchangeable during the evolution of disease. In patient CH002, for example, the clone carrying the Q61R *NRAS* mutation was readily observed at diagnosis, but declined beyond the limits of detection in the relapse sample (Figure 3e). In contrast, the clone carrying the *FLT3* D839R mutation was identified in the relapse sample while it was below the limits of detection (~400X) in the diagnosis sample (Figure 3e and f). Similarly, in patient GR019, two *FLT3*-ITDs (L610delins and N609delins) that were initially observed at diagnosis, were not found at relapse. Coupled with this, a relatively strong surge of a clone carrying another *FLT3*-ITD (Y597delins) occurred in the relapse (Figure 3e). Furthermore, multiple subclonal mutations carrying different proliferation driving mutations could often be found in the patients' AML (Figure 3g). Collectively, these results support a model in which diverse proliferative drivers were independently acquired in different subclones derived from the founder clone.

#### Recurrent mutations of cohesin genes and splicing factors

Besides the well-established role in the control of separation of the sister chromatids during cell division, cohesin proteins also have important roles in the maintenance of genome stability, DNA damage repair and regulation of gene transcription. Recurrent mutations of genes in the cohesin complex have been reported in a variety of cancers, including myelodysplastic syndromes, myeloproliferative neoplasms and AML.<sup>22,23</sup> Herein, we also found highly prevalent mutations of the cohesin genes (*STAG2* (16.3%), *SMC1A* (6.3%), *SMC3* (1%) and *RAD21* (1%)) and *CTCF* (6%, a transcription factor associated with the cohesin complex). In totality, the cohesin pathway is more frequently mutated in MLL-PTD patients (26%) than the AML samples from either TCGA (13%,  $P < 0.01$ ) or a meta-analysis of 1000 AML (9.1%,  $P < 0.0001$ ).<sup>24</sup> Remarkably, an extremely high proportion of the mutations had a strong tendency to disrupt the coding sequence (eight stop-gain, two frameshift and two splice-site mutations) in *STAG2* (Figure 4a), emphasizing its crucial tumor-suppressor role in this AML subtype (16.3% in MLL-PTD vs 3% in TCGA-AML ( $P < 0.01$ ) and 3.2% in a meta-analysis of 1102 AML<sup>24</sup> ( $P < 0.0001$ )). Interestingly, although *STAG2* mutations were found as early clonal mutations in our cohort with mean VAFs of around 0.4, these mutations were lost at relapse in two of the patients (Figure 4b-d), suggesting that maintenance of the *STAG2* mutation may not be required for relapse. Consistent with the notion of a functional overlap between mutations of the cohesin genes, a mutually exclusive pattern was noted (Figure 4e and f). Interestingly, a mutually exclusive pattern was also noted between *STAG2* and the splicing factor *U2AF1*. The underlying mechanism behind this observation is currently unknown and requires further studies. In addition, a mutually exclusive pattern was noted between *STAG2* and the other major AML drivers (except for *IDH2* which tended to co-occur with *STAG2*) (Supplementary Figure 6).

The RNA processing pathway was also strikingly altered in MLL-PTD patients. The most prominently mutated genes within this category were the splicing factors. They included *U2AF1* (12.5%,

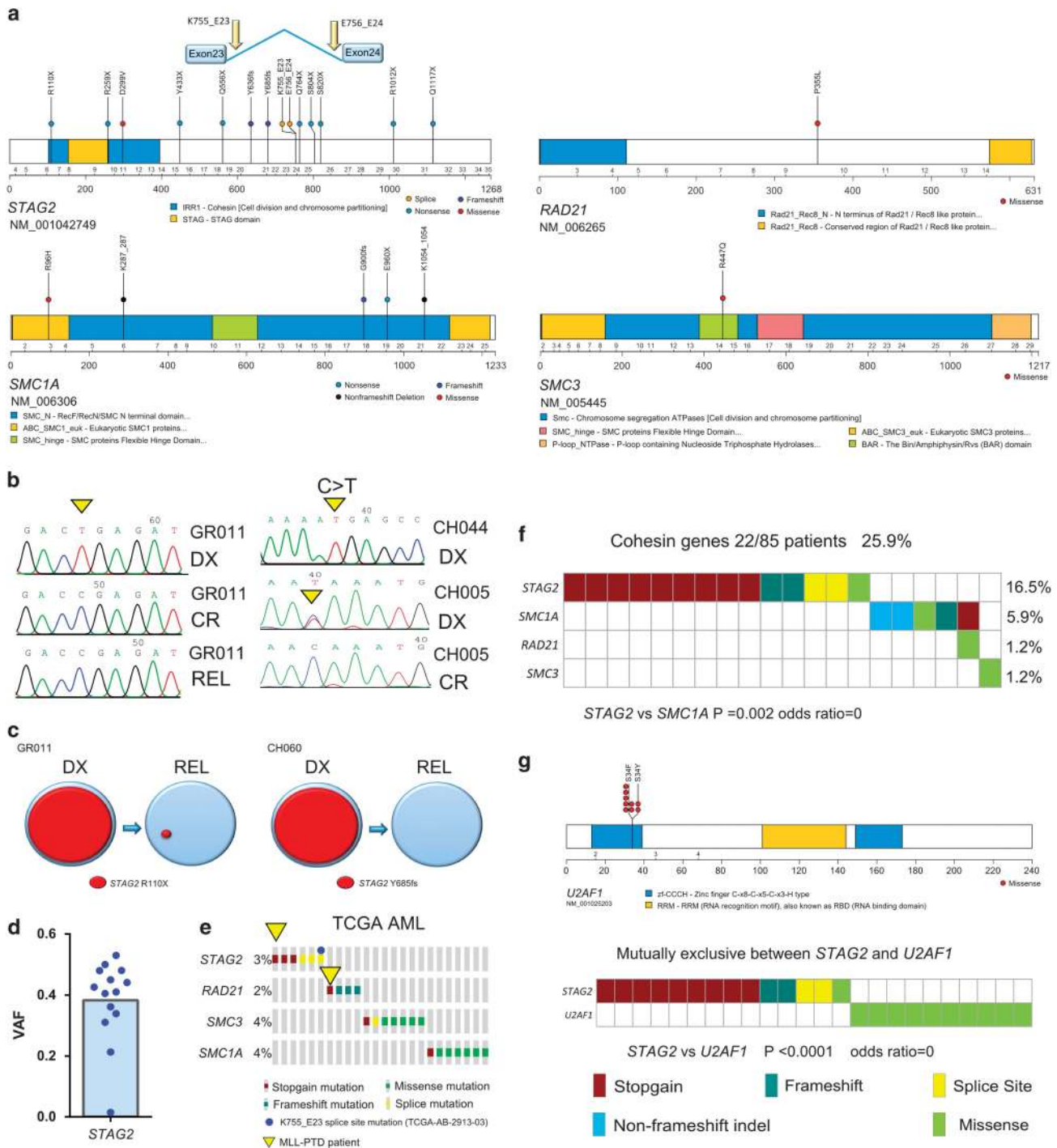


**Figure 3.** Mutations in the RAS-FLT3 pathway are subclonal and tend to be unstable. **(a)** Mutational landscape of patients carrying *PTPN11*, *FLT3*, *NRAS* and *KRAS* mutations. **(b)** Mutational diagram of *FLT3* and *NRAS* in 80 MLL-PTD patients at diagnosis. In total, 41 *FLT3*-ITD mutations were detected in 27 patients. **(c)** Variant allele frequency (VAF) plot of the frequently mutated genes. Mutations in epigenetic regulator genes (blue) tend to have high VAF, whereas proliferation-related genes (red) tend to have low VAF, indicating that SNV mutations in proliferation-related genes are generally subclonal and occur later than the genes involved in epigenetic regulation. *FLT3* SNV (single nucleotide variant) and *FLT3*-ITD are labeled separately. TCGA-AML result are shown as an insert. **(d)** *FLT3* mutations are unstable during disease progression. Schematic diagrams demonstrating the proportion and progression of the *FLT3* mutation carrying subclones inferred from their VAF in several diagnosis (DX) and relapse (REL) pairs are shown. Sequencing read depth of each mutation position are displayed in Supplementary Figure 9. **(e)** Mutations driving cell proliferation are replaceable during diagnosis and relapse. Upper pair, in patient CH002, mutant *NRAS* carrying subclone was found at leukemic diagnosis, this subclone subsequently was eliminated by chemotherapy. However, the founding clone survived during the chemotherapy, gained a *FLT3* D839G mutation, and this alternative *FLT3* carrying subclone (*FLT3* D839G) expanded and became the dominant clone at relapse. Lower pair, in patient GR019, leukemia at diagnosis harbored two different *FLT3*-ITD carrying subclones, both were eliminated by chemotherapy. The founding clone survived the chemotherapy, and an alternative subclone derived from the founding clone harbored a different *FLT3*-ITD mutation, which subsequently appeared at relapse. **(f)** Diagram depicting VAF and sequencing read depth of *NRAS* and *FLT3* mutations in patient CH002. Venn diagram indicates mutations that are either shared or distinct between the DX and REL samples. **(g)** Two case examples indicating the presence of multiple subclones carrying different proliferative driver mutations in patients' AML cells.

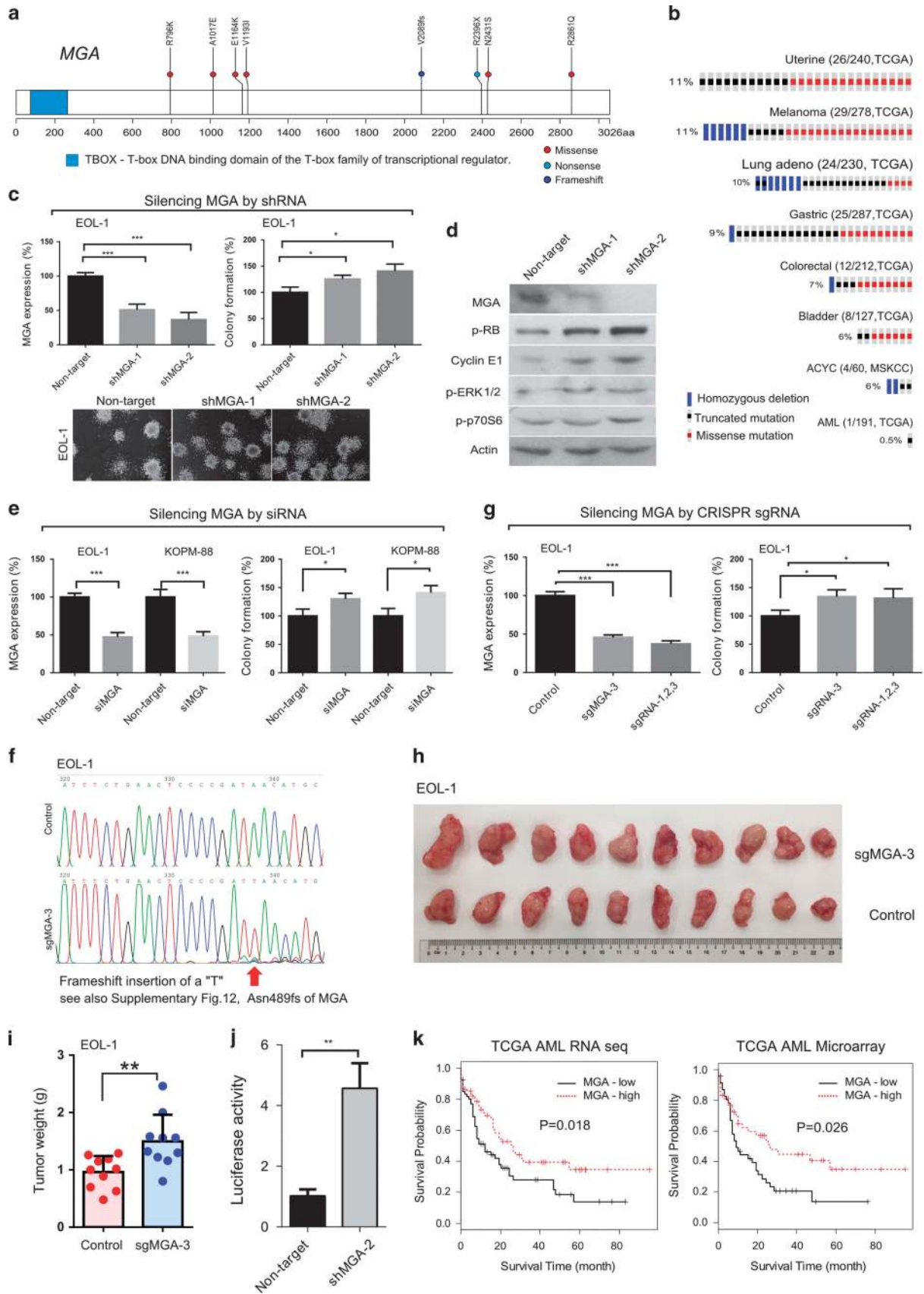
S34F/Y), *SRSF2* (2.5%, hotspot position P95, Supplementary Figure 10), *SF3A1* (5%), *ZRSR2* (2.5%), *DHX15* (1%) and *CWC22* (1%). We and others have hallmarked their importance by noting their high frequency (>60%) in myelodysplastic syndromes.<sup>25,26</sup> Genes encoding RNA helicases (*DDX10* (5%) and *HELZ* (3%)) were also frequently altered in the prevalence cohort. This family of proteins is involved in the regulation of RNA splicing, ribosome/spliceosome assembly and initiation of translation.<sup>27</sup> Mutations were also identified in RNA processing and degradation proteins *DIS3*<sup>28</sup> (9%) and *SMG1* (2.5%).

*MGA* is a potential tumor suppressor

Another highly mutated novel gene was *MGA*, an incompletely studied gene with a mutation frequency of 9% in MLL-PTD AML (Figure 5a). It encodes a MAX-interacting protein and is believed to act as a transcription factor that suppresses MYC binding to its target. By *in silico* analysis, we found that *MGA* was expressed in normal myeloid hematopoietic cells and AML (Supplementary Figure 11). Further data mining of TCGA sequencing projects revealed a high frequency of inactivating mutations of the *MGA* gene in a variety of cancers (Figure 5b), suggesting a potential pan



**Figure 4.** Members of the cohesin pathway are highly mutated in MLL-PTD AML. **(a)** Mutational diagram of the cohesin pathway genes in MLL-PTD patients. **(b)** Representative Sanger sequencing tracing showing the *STAG2* (X-chromosome gene) mutation in patients GR011, CH044 and CH005. A complete C > T alteration was noted in patients GR011 and CH044 (both males), whereas heterozygous double peaks were noted in patient CH005 (female). **(c)** Clones carrying *STAG2* mutation were present as dominant clone at diagnosis, but either disappeared or shrank drastically at relapse. GR011 at relapse had 7 of 146 *STAG2* mutant reads detected by NGS sequencing (below the detection threshold of Sanger sequencing, see Figure 6a and Supplementary Figure 13 where the number of mutant reads of *STAG2* in patients GR011 and CH060 are indicated). **(d)** Average VAF of *STAG2* mutations in diagnosis samples corrected for gender of patients (X-chromosome gene). **(e)** Mutational heatmap of cohesin pathway genes in TCGA-AML. Two patients had both *MLL-PTD* and cohesin mutations (highlighted with a yellow triangle). One of the TCGA patients harbored a mutation of *STAG2* in the same splicing site found in this study (highlighted with blue circle). **(f)** Mutual exclusivity of mutations of genes in the cohesin pathway within our MLL-PTD AML cohort. **(g)** *STAG2* mutations are mutually exclusive with the splicing factor *U2AF1* mutations as noted in our MLL-PTD AML cohort. Upper panel, mutational diagram of splicing factor *U2AF1*. Lower panel, mutually exclusive pattern between the samples carrying either *STAG2* or *U2AF1* mutations.



cancer tumor suppressor role of this gene. To interrogate functionally its role in leukemogenesis, lentiviral constructs containing either shRNA or CRISPR-sgRNA targeted to different regions of the *MGA* gene were generated (Supplementary Table 9). MLL-PTD containing AML cell line EOL-1<sup>29</sup> was infected with the shRNA/sgRNA viral particles. Stable cell lines were selected with puromycin, and the silencing ability of the shRNAs towards *MGA* was assessed by western blot and real-time PCR (Figure 5c and d, Supplementary Table 10). Methylcellulose colony-formation assay was subsequently employed to evaluate the effect of silencing *MGA*. An increase of colony number (~30%) was observed in *MGA* silenced cell lines (Figure 5c). Similarly, silencing of *MGA* by either siRNA or CRISPR-sgRNA modestly increased the colony formation (Figure 5e-g). Congruently, silencing of *MGA* by CRISPR-sgRNA consistently enhanced the *in vivo* xenograft cell growth in NSG mice (Figure 5h and i, Supplementary Figure 12). In addition, western blot analyses revealed the elevation of p-RB and Cyclin E1 in the *MGA* silenced cells (Figure 5d, Supplementary Figure 12), indicative of a proliferative advantage conferred by the silencing of *MGA*.

As mentioned above, *MGA* may be a potential regulator of the MYC pathway. We, therefore, examined whether silencing of *MGA* alters MYC transcriptional activity. Luciferase reporter assay was carried out in 293FT cells stabilized with either scramble or shRNA-targeting *MGA*. A fourfold increase in luciferase activity was observed in *MGA* silenced cells when compared with non-targeting shRNA controls (Figure 5j). Furthermore, Kaplan–Meier survival analysis of the TCGA-AML patients suggests that *MGA* mRNA expression level is correlated with overall patient survival. AML patients with lower levels of *MGA* in their leukemic samples had a worse outcome compared with those whose leukemic cells expressed higher levels of *MGA* (Figure 5k). Collectively, our results suggest that *MGA* may function as a potential tumor-suppressor in AML.

#### Initiating mutations often persist at remission

Strikingly, we observed that some early clonal mutations in *DNMT3A*, *TET2*, *IDH1/2*, *U2AF1* (herein termed remission-persisting mutations) were retained with high VAF in a large proportion of patients' remission samples (29%, 8 of 28 patients); in contrast,

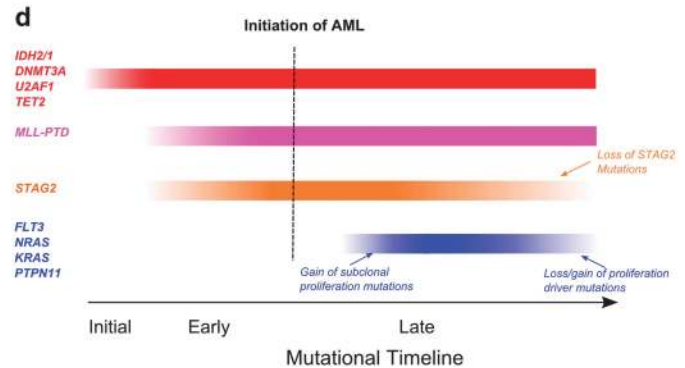
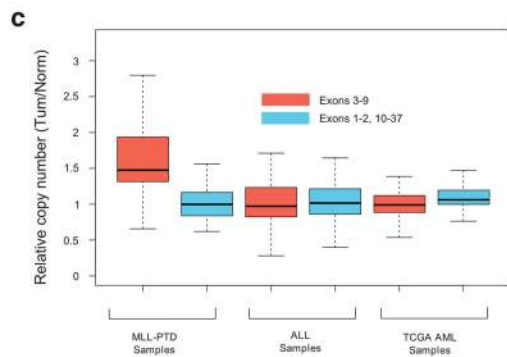
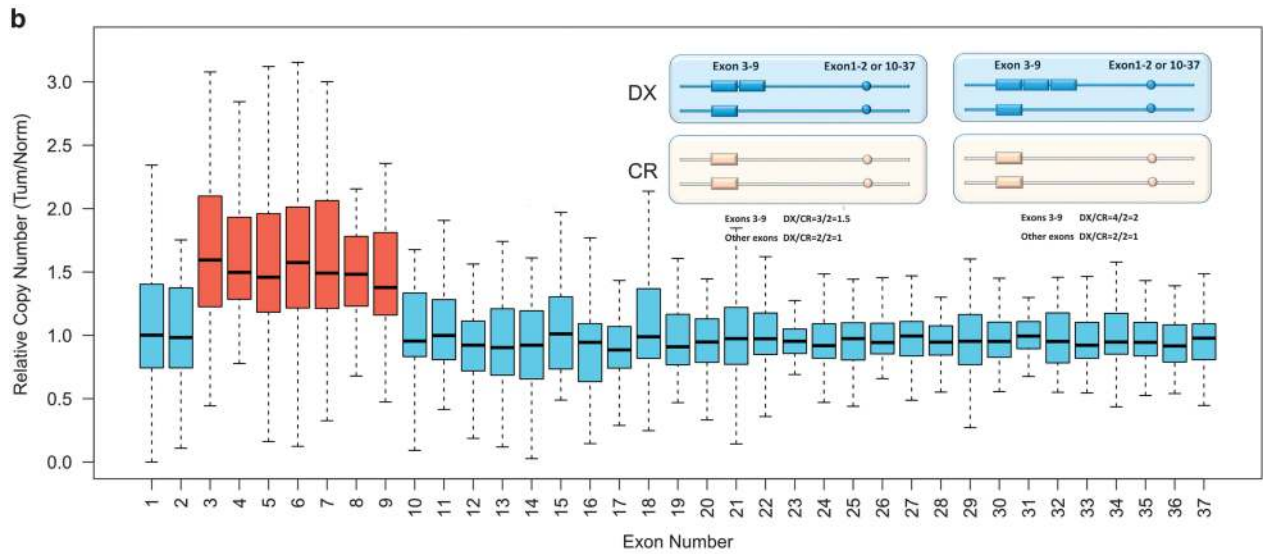
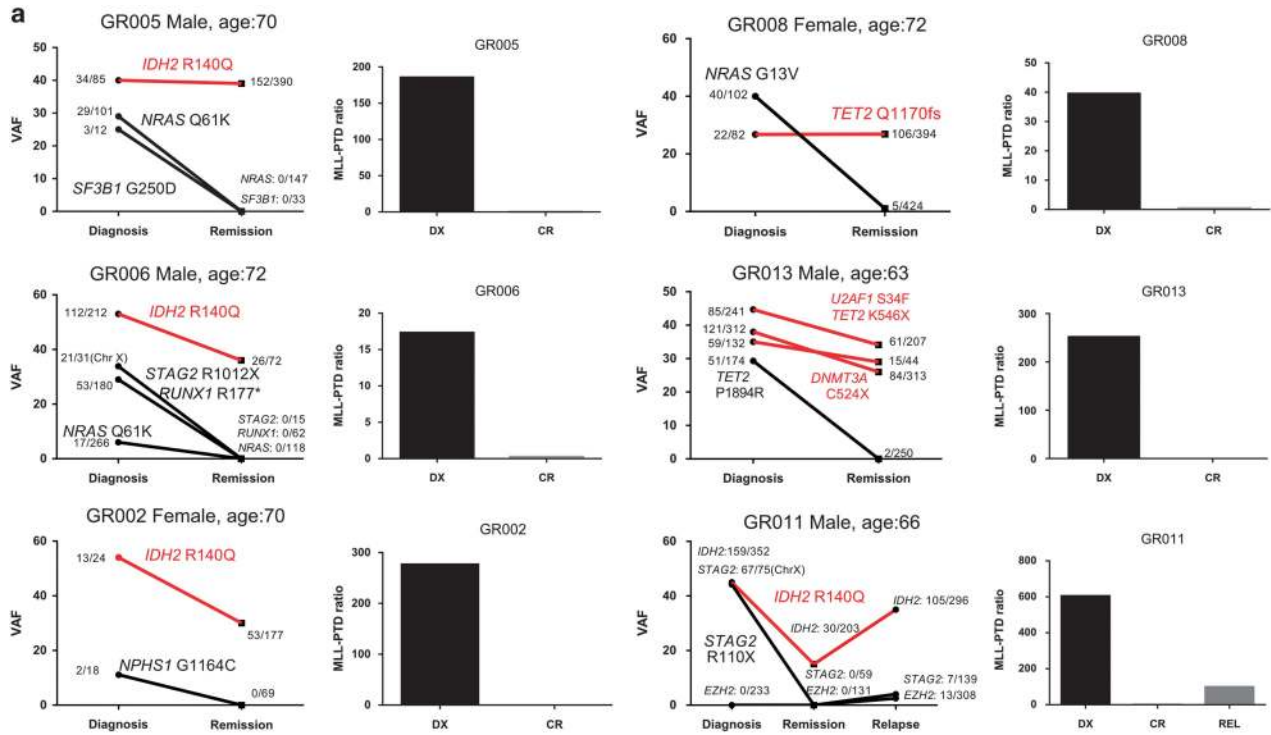
neither the *MLL*-PTD (determined by real-time RT-PCR (sensitive range from 1:100-1:10 000, see Weisser *et al.*<sup>30</sup>) and by relative copy number (RCN) analysis) as well as those associated with a relatively later subclonal mutations (for example, *FLT3/NRAS*, Figure 6a and Supplementary Figure 13) were detected. The fact that *DNMT3A/TET2/IDH1/2/U2AF1* mutations were detected at both diagnosis and complete clinical morphological remission (determined by normal morphology/cytogenetics, % blasts and further supported by the disappearance of other mutations found at diagnosis) suggests that these mutations are most likely present prior to malignant transformation.<sup>10,12,31</sup> Notably, recent studies from several research groups have revealed a high prevalence of hematologic cancer-related gene (*DNMT3A*, *TET2*, *ASXL1*) mutations in normal individuals as they aged<sup>31–35</sup> and these mutations are associated with an increased risk for the development of hematologic cancer. Collectively, our observations suggest that chemotherapy eliminated the malignant leukemic cells and reverted these patients back to a pre-leukemic (clonal hematopoiesis) instead of a completely normal state (polyclonal hematopoiesis). These remission-persisting mutations carrying clones may act as a reservoir for subsequent relapse.<sup>36</sup>

#### Ordering of mutations in MLL-PTD AML

To identify the mutation order of the *MLL*-PTD aberration in relation to other mutations, we performed copy number analysis of *MLL*-PTD. The tandem duplication of *MLL* gene leads to a copy number gain of the affected exons. By analyzing the sequencing read depth of each exon (normalized with the local and global capturing efficiency, Supplementary Figures 14–16), the RCN of each exon of the *MLL* gene can be calculated and the proportion of mutation carrying cells inferred (see Supplementary Results for full discussion). Guided by this notion, we calculated the RCN value of each exon of the *MLL* gene for 28 paired MLL-PTD samples with matched remission control. For most of the samples, we found a mean RCN of ~1.5 for exons 3–9 and a RCN value of ~1 in the other exons (Figure 6b and c). At the same time, a control analysis was performed with TCGA-AML patient's exome-sequencing data (excluding the nine MLL-PTD patients) and 200 ALL patients targeted sequencing data (Supplementary Figure 15),

**Figure 5.** *MGA* is a potential tumor-suppressor mutated in MLL-PTD AML. **(a)** Diagram showing the mutational type and position within the *MGA* gene identified in MLL-PTD samples. Missense mutations R796K and V1193I of *MGA* occurred in the same sample. **(b)** High proportion of inactivating mutations were observed in *MGA* in a variety of cancers. Adeno, adenocarcinoma; ACYC, adenoid cystic carcinoma. **(c)** Enhanced colony formation in EOL-1 cells after silencing *MGA* with either shMGA-1 or shMGA-2. Left panel, real-time PCR result shows the silencing effect of *MGA* by two shRNAs; right panel, colony formation result of EOL-1 cell after silencing *MGA* with two shRNA. Mean  $\pm$  s.d.,  $n=3$ ,  $*P < 0.05$ ;  $***P < 0.001$ . Lower panel, representative photographs of EOL-1 colonies formed in methylcellulose of non-target shRNA (left photos) and shRNA silenced *MGA* (middle and right photos). **(d)** Silencing of *MGA* in EOL-1 cells increased protein levels of Cyclin E1 and phospho-RB (S807, phosphorylation inhibits RB target binding and allow cell cycle progression). A representative western blot of four independent experiments is shown. **(e)** Enhanced colony formation in EOL-1 and KOPM88 cells after silencing *MGA* with siRNA pool (SMARTpool: ON-TARGETplus *MGA* siRNA, Dharmacon). siRNA was delivered using Nucleofector™ electroporation device (Lonza). Transfection efficiency was > 80% in both cell lines as determined by the co-transfected GFP. Left panel, real-time PCR result showing the silencing effect of *MGA* by siRNA pool in EOL-1 and KOPM88 cells; right panel, colony formation result of EOL-1 and KOPM88 cells after silencing *MGA* with siRNA pool. Mean  $\pm$  s.d.,  $n=3$ .  $*P < 0.05$ ;  $***P < 0.001$ . **(f)** Sanger sequencing tracings depicting the frameshift insertion of a T nucleotide induced by CRISPR sgMGA-3 in EOL-1 cell line (see Supplementary Figures 12c and d for detailed sequencing tracing and BLAST result). **(g)** Enhanced colony formation in EOL-1 after silencing *MGA* with CRISPR-sgRNA. Left panel, real-time PCR result indicating the silencing effect of *MGA* by a single CRISPR guide RNA (sgMGA-3) or a mixture of three sgRNAs targeting *MGA* (sgMGA-1,2,3); right panel, the colony-formation result of EOL-1 cell after silencing *MGA* with sgMGA-3 or a mixture of three sgRNAs designed to target *MGA* (sgMGA-1,2,3). Real-time PCR and colony-formation assay were performed 15 days after the lentiviral infection (3 days for virus infection, 7 days for puromycin selection and 5 days with normal medium). Mean  $\pm$  s.d.,  $n=3$ .  $*P < 0.05$ ;  $***P < 0.001$ . **(h)** and **(i)** Enhanced tumor mass of *MGA* knockout EOL-1 cell in an *in vivo* xenograft model. Control EOL-1 cells or EOL-1 cells silenced with *MGA* CRISPR sgRNAs were injected into both flanks of NSG mice, and tumor masses were harvested 21 days after injection. **(h)** Photo of tumor masses dissected from the NSG mice, which were injected with EOL-1 cells after silencing *MGA* expression by CRISPR sgMGA-3 (upper panel) or control EOL-1 cells (lower panel); **(i)** Barplots indicating the weight of the tumor mass. Mean  $\pm$  s.d.,  $n=10$ ,  $***P < 0.01$ . **(j)** Increased MYC activity in 293FT cells after stable silencing of *MGA* by shRNA. Luciferase activities were measured 48 h after transfection of cells with MYC activity reporter pMyc4ElbLuc (Addgene, #53246) and normalized to the corresponding co-transfected Renilla luciferase activity. Mean  $\pm$  s.d.,  $n=3$ ,  $***P < 0.01$ . **(k)** Kaplan–Meier plot of overall survival: comparison of cases with highest versus lowest expression of *MGA* in AML patients.  $P$ -values were calculated by log-rank test. *MGA* expression data and patient survival data were retrieved from TCGA-AML patients RNA seq (left panel), or microarray (70 AML patients) (right panel). The *MGA* expression 'high' and 'low' groups were defined by 15% higher than the median or 15% lower than the median, respectively.





**Figure 6.** MLL-PTD is a clonal mutation that occurs before mutations of the RAS-RTK pathway, but later than the *IDH2/DNMT3A/U2AF1/TET2* mutations. (a) AML mutations of either *IDH2*, *TET2*, *DNMT3A*, *U2AF1* (colored in red) persist in the remission samples of eight patients with matched remission control (plot of two patients are shown in Supplementary Figure 13). Sequencing read depth of each mutation at different time point is indicated (mutant reads/total read of the indicated position). MLL-PTD level at each time point was determined by quantitative RT-PCR to calculate the MLL-PTD ratio (normalized copy number of MLL-PTD to ABL1 ratio ((MLL-PTD/ABL) × 100) according to the method described in Weisser *et al.*<sup>30</sup>). (b) Diagram illustrating the relative copy number of each exon of the *MLL* gene. Inner picture depicts the expected relative copy number between DX and CR samples for exons affected by MLL-PTD (exons 3–9), and those unaffected by the aberration (other exons). Gain of exons 3–9 of the *MLL* gene could be observed in 28 MLL-PTD patients with matched remission control. (c) Diagram illustrating the relative copy number of exons 3–9 or exons 1–2, 10–37 of the *MLL* gene in 28 pair MLL-PTD patients, 200 pediatric ALL patients (Supplementary Figure 15b) and 120 TCGA-AML patients (the 9 MLL-PTD samples were excluded). (d) Mutational timeline model for the sequence of acquisition of *IDH2/DNMT3A/U2AF1/TET2*, MLL-PTD and the *FLT3/NRAS/KRAS/PTPN11* mutations in MLL-PTD AML patients.

which revealed a RCN value of ~1 in all *MLL* exons (Figure 6c). Taken together, the observed RCN of duplicated MLL-PTD exons of ~1.5 (Figure 6c) suggests the presence of MLL-PTD as a clonal mutation in most of the MLL-PTD patients.

MLL-PTD is a highly specific AML driver

Our observation of MLL-PTD as an early clonal driver mutation in AML raises an intriguing question of whether MLL-PTD may serve similar roles in other cancer types. MLL-PTD manifests as a tandem duplication of exons 3–9 of *MLL* gene. The length of exons 3–9 of *MLL* is ~12.5 kb, suggesting that SNP-arrays can identify the MLL-PTD alteration. Indeed, canonical MLL-PTD involving exons 3–9 was identified in the EOL-1 (known to carry MLL-PTD mutation<sup>29</sup>), as well as the Kasumi-6 AML cell line (Supplementary Figures 15c and d). Furthermore, AML samples known to have MLL-PTD can also be detected by analysis of the TCGA SNP-array database (Supplementary Figure 15e), suggesting that SNP-array is a valid platform for the detection of MLL-PTD. Upon further examination of 1000 different cancer cell lines in CCLE, 10 000 different cancer samples in TCGA SNP-array database (Broad-Firehose-standard-data-run2015\_04\_02) and a panel of 2792 samples of different cancer types in the Tumorscape SNP-array database, we conclude that the gain of exon 3–9 in *MLL*, as a clonal mutation, is likely only present in AML, suggesting that MLL-PTD is a rare and highly specific AML driver.

## DISCUSSION

In this study, we found that multiple mutations co-occur with MLL-PTD and mutations are usually acquired in a sequential manner. A potential ordering for acquisition of mutations was proposed (*IDH2/DNMT3A/U2AF1/TET2* → MLL-PTD → RAS-RTK) based on the following reasons: #1, real-time-PCR (Figure 6a) showed that MLL-PTD was absent in remission while mutations of *IDH2*, *DNMT3A*, *TET2* and *U2AF1* were still retained with a high VAF. This suggests that MLL-PTD was acquired after mutations of *IDH2*, *DNMT3A*, *TET2* and *U2AF1*; #2, MLL-PTD is highly stable during disease progression as compared with mutations of the RAS-RTK. In a previous study characterizing 46 AML patients carry MLL-PTD mutation at diagnosis, MLL-PTD persisted in 43 of these individuals at relapse.<sup>30</sup> On the other hand, as shown in Figure 3, RAS-RTK mutations frequently exist as subclonal mutations and tend to be unstable during disease progression. These observations support a notion that MLL-PTD was acquired prior to RAS-RTK. Taken together, the above reasoning suggests MLL-PTD is acquired after those remission-persisting, initiating mutations (*IDH2*, *DNMT3A*, *TET2* and *U2AF1*), but prior to lesions of the proliferation-related driver (Figure 6d), highlighting its critical role in the development of full-blown AML. Indeed, our observations and inference of clonal early/late mutations are consistent with recent reports of healthy elderly individuals carrying functionally inactivating *DNMT3A/TET2* mutations,<sup>31–34</sup> as well as previous studies from the *DNMT3A* and *TET2* knockout mouse model.<sup>37–39</sup> These studies point to a model where a

pre-malignant founder clone is formed following acquisition of an initiating mutation and where the subsequent gain of cooperating mutations in these pre-leukemic cells leads to the onset of leukemogenesis. In this paradigm, MLL-PTD can thus be viewed as one of the early cooperating mutations. Interestingly, one of the most frequently mutated AML genes, *NPM1*, is mutually exclusive with MLL-PTD. Mutations of *NPM1* are usually present as early clonal mutations in AML blasts, but are absent in normal appearing elderly individuals who can carry mutations of *DNMT3A* and *TET2*.<sup>32–34,40</sup> This observation implies that, similar to MLL-PTD, *NPM1* mutations can also be considered as one of the early cooperating mutations. In this regards, MLL-PTD and *NPM1* may be involved in similar pathways to trigger AML as inferred from their mutually exclusive pattern.

Based on our observations and previous studies of normal elderly individuals,<sup>32–34,40</sup> a potential molecular boundary can be defined to distinguish the leukemic and pre-leukemic stage: the initiating mutations (*DNMT3A* and so on) were detected in both pre- and bonafide leukemia as well as in 10% of normal individuals over the age of 65,<sup>32–34,40</sup> whereas the early mutations of *RUNX1*, *WT1*, *NPM1*, *STAG2* and so on were usually only found in frank leukemic blast cells (and in myelodysplastic syndromes/myeloproliferative neoplasms). Therefore, these early mutations (*RUNX1*, *WT1*, *NPM1* and so on) rather than the initiating mutations, will likely be the most informative markers for the measurement of minimal-residual disease.<sup>41</sup>

In the design of targeted therapeutic strategies for cancer patients, high priority should be accorded to the ‘trunk’ mutations that are present in the majority of the cancer cells. The molecular timeline model established here suggests that, despite the observed high prevalence of RAS-RTK mutations in MLL-PTD AML, most of these mutations are present in only a fraction of the cancer cells. Thus, targeting these late subclonal events may have only limited therapeutic value.

During submission of our manuscript, another article on MLL-rearranged AML appeared;<sup>42</sup> this article identified LOC100289656 as a specific MLL fusion AML marker and focused on analysis of the transcriptome and mutation of MLL fusion AML patients. Conversely, our study focuses on MLL-PTD instead of MLL fusion, which distinguishes our study from their report.

In summary, our analysis of the mutational landscape of 85 MLL-PTD AML patients revealed the sequential pattern of mutational acquisition and a number of co-occurring mutations. This study provides a detailed mutational profile and approximate timeline for a unique AML subtype, which may guide the development of more effective therapeutic approaches and further pave the way for AML personalized therapeutic strategies.

## CONFLICT OF INTEREST

Torsten Haferlach is employed by and partly owns the MLL Munich Leukemia Laboratory. Tamara Alpermann and Manja Meggendorfer are employed by the MLL Munich Leukemia Laboratory. The remaining authors declare no competing financial interests.

**ACKNOWLEDGEMENTS**

We thank Professor Patrick Tan and Dr Zhijiang Zang for sharing relevant facilities. This work was funded by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research (STaR) Investigator Award, NIH grant R01CA026038-35, the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centers of Excellence Initiatives. This study was also partially supported by a generous donation from the Melamed family and the Eleanor and Glenn Padnick Discovery Fund in Cellular Therapy.

**REFERENCES**

1 Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K *et al*. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008; **456**: 66–72.

2 Estey E, Döhner H. Acute myeloid leukaemia. *Lancet* 2006; **368**: 1894–1907.

3 Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J *et al*. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012; **366**: 1079–1089.

4 Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K *et al*. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. *Leukemia* 2014; **28**: 1586–1595.

5 Dohner K, Tobis K, Ulrich R, Frohling S, Benner A, Schlenk RF *et al*. Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. *J Clin Oncol* 2002; **20**: 3254–3261.

6 Horne G, Brewin J, Chevassut T. Genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012; **366**: 2321 author reply 2321–2322.

7 Dorrance AM, Liu S, Yuan W, Becknell B, Arnoczky KJ, Guimond M *et al*. Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *J Clin Invest* 2006; **116**: 2707–2716.

8 Zorko NA, Bernot KM, Whitman SP, Siebenaler RF, Ahmed EH, Marcucci GG *et al*. Mll partial tandem duplication and Flt3 internal tandem duplication in a double knock-in mouse recapitulates features of counterpart human acute myeloid leukemias. *Blood* 2012; **120**: 1130–1136.

9 Network TCGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.

10 Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC *et al*. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 2012; **150**: 264–278.

11 Passegue E, Jamieson CH, Ailles LE, Weissman IL. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci USA* 2003; **100** (Suppl 1): 11842–11849.

12 Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR *et al*. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* 2012; **4**: 149ra118.

13 Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J *et al*. Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med* 2015; **372**: 601–612.

14 Daigle SR, Olhava EJ, Therkelsen CA, Majer CR, Sneeringer CJ, Song J *et al*. Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* 2011; **20**: 53–65.

15 Kuhn MW, Hadler MJ, Daigle SR, Koche RP, Krivtsov AV, Olhava EJ *et al*. MLL partial tandem duplication leukemia cells are sensitive to small molecule DOT1L inhibition. *Haematologica* 2015; **100**: e190–e193.

16 Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213–219.

17 Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS *et al*. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012; **481**: 506–510.

18 Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV *et al*. Signatures of mutational processes in human cancer. *Nature* 2013; **500**: 415–421.

19 Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE *et al*. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 2010; **363**: 2424–2433.

20 Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V *et al*. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 2014; **506**: 328–333.

21 Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R *et al*. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* 2014; **25**: 442–454.

22 Thota S, Viny AD, Makishima H, Spitzer B, Radivoyevitch T, Przychodzen B *et al*. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood* 2014; **124**: 1790–1798.

23 Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y *et al*. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat Genet* 2013; **45**: 1232–1237.

24 Leeke B, Marsman J, O'Sullivan JM, Horsfield JA. Cohesin mutations in myeloid malignancies: underlying mechanisms. *Exp Hematol Oncol* 2014; **3**: 13.

25 Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R *et al*. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011; **478**: 64–69.

26 Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J *et al*. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* 2012; **44**: 53–57.

27 Abdelhaleem M, Maltais L, Wain H. The human DDX and DHX gene families of putative RNA helicases. *Genomics* 2003; **81**: 618–622.

28 Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC *et al*. Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011; **471**: 467–472.

29 Quentmeier H, Reinhardt J, Zaborski M, Drexler H. MLL partial tandem duplications in acute leukemia cell lines. *Leukemia* 2003; **17**: 980–981.

30 Weisser M, Kern W, Schoch C, Hiddemann W, Haeflrich T, Schnittger S. Risk assessment by monitoring expression levels of partial tandem duplications in the MLL gene in acute myeloid leukemia during therapy. *Haematologica* 2005; **90**: 881–889.

31 Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci USA* 2014; **111**: 2548–2553.

32 Genovesi G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF *et al*. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014; **371**: 2477–2487.

33 Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG *et al*. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014; **371**: 2488–2498.

34 Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC *et al*. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* 2014; **20**: 1472–1478.

35 Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z *et al*. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* 2012; **44**: 1179–1181.

36 Kico JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S *et al*. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. *JAMA* 2015; **314**: 811–822.

37 Mayle A, Yang L, Rodriguez B, Zhou T, Chang E, Curry CV *et al*. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* 2015; **125**: 629–638.


38 Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C *et al*. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011; **20**: 11–24.

39 Li Z, Cai X, Cai CL, Wang J, Zhang W, Petersen BE *et al*. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 2011; **118**: 4509–4518.

40 McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J *et al*. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep* 2015; **10**: 1239–1245.

41 Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A *et al*. Assessment of minimal residual disease in standard-risk AML. *N Engl J Med* 2016; **374**: 422–433.

42 Lavalley VP, Baccelli I, Kros J, Wilhelm B, Barabe F, Gendron P *et al*. The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias. *Nat Genet* 2015; **47**: 1030–1037.

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on the Leukemia website (<http://www.nature.com/leu>)