

1 **Title (100/150):** *UBTF* Tandem Duplications in Pediatric MDS and AML: Implications for
2 Clinical Screening and Diagnosis

3

4 **Short Title (46/50):** *UBTF*-TDs in Pediatric MDS and AML

5

6 **Authors and Affiliations:**

7 Juan M. Barajas^{1,+}, Masayuki Umeda^{1,+}, Lisett Contreras¹, Mahsa Khanlari¹, Tamara
8 Westover¹, Michael P. Walsh¹, Emily Xiong¹, Chenchen Yang², Brittney Otero², Marc
9 Arribas-Layton², Sherif Abdelhamed¹, Guangchun Song¹, Xiaotu Ma³, Melvin E.
10 Thomas 3rd¹, Jing Ma¹, Jeffery M. Klco^{1,*}

11 + Equal first authors

12 * Corresponding author

13 1. Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN,
14 USA.

15 2. Mission Bio, South San Francisco, CA

16 3. Department of Computational Biology, St. Jude Children's Research Hospital,
17 Memphis, TN, USA.

18 **Contact Information/Corresponding authors:**

19 *Correspondence: jeffery.klco@stjude.org

20 Jeffery M. Klco: Mail Stop 342, Room D4047B, St. Jude Children's Research Hospital,
21 262 Danny Thomas Place, Memphis, TN 38105-3678, Phone: (901) 595-6807, Fax:
22 (901) 595-5947, Email: jeffery.klco@stjude.org

23 **Word Count (includes Methods/Introduction/Results/Discussion):** 3,090/4,500

24

25 **Figures:** 5/7

26

27 **Key Points:**

- 28 • Largest cohort of pediatric *UBTF*-TD in myeloid neoplasms reported to date.
- 29 • Use of single-cell DNA+protein sequencing technology in 3 *UBTF*-TD samples
30 reveals a clonal evolution pattern characterized by sequential acquisition of
31 *WT1* and *FLT3* mutations and a more stem cell-like protein expression pattern.
- 32 • Pediatric MDS and AML patients with *UBTF*-TD alterations dysplastic features
33 with an increase erythroid precursors.
- 34 • Tandem duplications in exon 9 of *UBTF* represent a rare but functionally
35 equivalent subgroup of *UBTF*-TD myeloid neoplasms.

36

37 **Impact Statement:**

38 *UBTF* tandem duplications (TD) are subtype-defining genomic alterations in adult and
39 pediatric myeloid neoplasms. Here, we provide a comprehensive characterization of the
40 largest cohort of pediatric *UBTF*-TD cases to date, including the recognition of
41 additional *UBTF* alterations that mimic the exon 13 duplications in pediatric AML.

42

43 **Abstract, (200/200 words):**

44 Recent genomic studies in adult and pediatric acute myeloid leukemia (AML)
45 demonstrated recurrent in-frame tandem duplications (TD) in exon 13 of upstream
46 binding transcription factor (*UBTF*). These alterations, which account for ~4.3% of AMLs
47 in childhood and up to 3% in adult AMLs under 60, are subtype-defining and associated
48 with poor outcomes. Here, we provide a comprehensive investigation into the
49 clinicopathological features of *UBTF*-TD myeloid neoplasms in childhood, including 89
50 unique pediatric AML and 6 myelodysplastic syndrome (MDS) cases harboring a
51 tandem duplication in exon 13 of *UBTF*. We demonstrate that *UBTF*-TD myeloid tumors
52 are associated with dysplastic features, low bone marrow blast infiltration, and low white
53 blood cell count. Furthermore, using bulk and single-cell analyses, we confirm that
54 *UBTF*-TD is an early and clonal event associated with a distinct transcriptional profile,
55 whereas the acquisition of *FLT3* or *WT1* mutations is associated with more stem cell-
56 like programs. Lastly, we report rare duplications within exon 9 of *UBTF* that phenocopy
57 exon 13 duplications, expanding the spectrum of *UBTF* alterations in pediatric myeloid
58 tumors. Collectively, we comprehensively characterize pediatric AML and MDS with
59 *UBTF*-TD and highlight key clinical and pathologic features that distinguish this new
60 entity from other molecular subtypes of AML.

61

62 **Key Words:**

63 *UBTF* Tandem Duplications; Pediatric and Adult Acute Myeloid Leukemia; Myeloid
64 Neoplasms

65

66 **Introduction:**

67 Pediatric acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) are
68 characterized by unique genetic backgrounds when compared to those in adults (1-3).
69 Recurrent tandem duplications (TD) of exon 13 of upstream binding transcription factor
70 (*UBTF*) were only recently identified as potential initiating events in pediatric AML (4-7),
71 accounting for about 4% of newly diagnosed pediatric AML. PCR-based screening
72 covering exon 13 of *UBTF* also identified *UBTF*-TD alterations in large adult AML
73 cohorts (8, 9). These studies significantly contributed to the accumulation of evidence of
74 *UBTF*-TD alterations in adult AML. However, PCR-based methods potentially
75 underestimate partial tandem duplications (PTD) extending outside the regions covered
76 by amplicons or possible alterations not involving exon 13 (9). Also, data on *UBTF*
77 alterations in pediatric AML is limited to screening of relatively small cohorts (4, 5), and
78 further efforts are needed to accumulate more knowledge about the biology and
79 clinicopathologic features of this disease entity.

80

81 *UBTF* encodes the *UBTF/UBF* protein known to regulate ribosomal RNA (rRNA)
82 transcription and nucleolar formation (10, 11). We previously reported that expression of
83 exon 13 *UBTF*-TD in cord blood CD34+ (cbCD34+) cells is sufficient to induce cellular
84 proliferation, increase clonogenic activity and the establishment of a transcriptional
85 signature that recapitulates that observed in *UBTF*-TD AML patient samples (4). Our
86 previous analyses also demonstrated that *UBTF*-TD do not occur with other canonical
87 alterations in pediatric AML, but that *UBTF*-TD AMLs often harbor additional somatic
88 mutations, such as internal tandem duplications in *FLT3* (*FLT3*-ITD) and frameshift

89 mutations in *WT1*. The acquisition of these cooperating mutations can likely contribute
90 to the stepwise progression of the disease and clonal evolution. However, our
91 understanding of how these cooperating mutations contribute to the cellular and disease
92 status remains to be elucidated.

93

94 To bridge these knowledge gaps, we present an extended pediatric cohort of 89 AML
95 and 6 MDS samples with exon 13 *UBTF*-TD, showing that *UBTF*-TD is strongly
96 associated with dysplastic features while acquisition of *FLT3*-ITD is associated with
97 progression from MDS to AML. By leveraging bulk RNA-sequencing and single cell
98 genomics, we show that the co-occurrence of *FLT3*-ITD and *WT1* mutations is
99 associated with stem cell-like phenotypes. Furthermore, we identified tandem
100 duplications within exon 9 of *UBTF* in two cases transcriptionally resembling exon 13
101 *UBTF*-TD AML. Exon 9 *UBTF*-TDs contain hydrophobic leucine-rich sequences and can
102 induce leukemic phenotypes in cbCD34+ cells, which recapitulate findings in exon 13
103 *UBTF*-TD, suggesting that they likely have a shared mechanism and should be
104 classified as the same molecular entity. These findings offer valuable insights to inform
105 future diagnostic strategies and understanding of the molecular basis of *UBTF*-TD
106 myeloid neoplasms.

107

108 **Materials/Subjects and Methods:**

109 *Cell Culture and Analysis of cord-blood CD34+ Cell Models*

110 *UBTF*-TD cbCD34+ models were generated as previously described (4). Freshly
111 isolated cord blood CD34+ (cbCD34+) cells were transduced with lentiviral particles

112 from MND-PGK-mCherry constructs expressing N-terminus HA-tagged UBTF-wild-type
113 (WT), N-terminus HA-tagged UBTF-TD, N-Terminus HA-tagged UBTF-TD78-exon9 or
114 N-Terminus HA-tagged UBTF-TD153-exon9. Transduced cells were sorted for mCherry
115 positivity and expanded as previously described(4). Molecular experiments were
116 performed at 40-60 days after sorting unless otherwise noted. For cytopins, 100,000
117 cells were washed with 1X PBS and spun onto Superfrost Plus Microscope slides (12-
118 660-16, Fisher Scientific) at 800 rpm for 5 min.

119

120 *DNA Single-Cell Sequencing and Analysis*

121 Single-cell targeted sequencing was performed using Tapestri System from Mission Bio
122 (missionbio.com). A panel of 162 PCR amplicons with an average size of 260bp was
123 designed using Tapestri Designer from Mission Bio (designer.missionbio.com), as well
124 as a manually designed amplicon targeting the *UBTF* exon13 TD region. This panel
125 consists of amplicons targeting *UBTF* tandem duplicated regions, common mutations in
126 *UBTF*-TD AML, and mutations specific to cases presented in this manuscript
127 (**Supplemental Table 2**). Cryopreserved UBTF-TD primary AML samples were thawed
128 and subjected to dead cell removal using the EasySep™ Dead Cell Removal Kit
129 (STEMCELL Technologies, cat# 17899). Live cells were then subjected to the Mission
130 Bio DNA+Protein protocol per manufacturer's instructions (missionbio.com). Libraries
131 were then sequenced on the Novaseq platform (100M read pairs for DNA libraries and
132 225M read pairs for protein libraries). Bam files, loom files, h5 files, and QC metrics
133 were produced via a customization of the Tapestri pipeline developed by Mission Bio
134 (support.missionbio.com/hc/en-us). Analysis of the samples was completed using the

135 Mosaic package v3.0.1 (missionbio.github.io/mosaic/). Reads for the *UBTF* and *FLT3*-
136 ITD calls were isolated from bam files using the pysam python package v0.21.0
137 (github.com/pysam-developers/pysam). Reads were then realigned to ITD contigs
138 reported in previous studies (4) using the BWA aligner v0.7.15-r1140 (12, 13). When
139 necessary, mutation VAFs were adjusted using pysam.

140

141 *RNA-sequencing and genomic profiling*

142 As we previously performed (1), RNA reads from newly sequenced samples and from
143 publications were mapped to the GRCh37/hg19 human genome assembly using the
144 StrongARM pipeline (14). Chimeric fusion detection was carried out using CICERO
145 (v0.3.0)(15). For somatic mutations calling from RNA-seq BAM files, we applied
146 Bambino (v1.07)(16) for SNV and RNAindel (v3.0.4) (17, 18) focusing on 87 predefined
147 genes recurrently mutated in pediatric AML and myelodysplastic syndrome. UBTF-TD
148 screening was performed as we have reported (4). For the cases with TWIST capture
149 sequencing, we called mutations as previously described (4). We also collected
150 mutation calls with DNA data from referenced publication in Supplemental Table 5.

151

152 *Transcriptome analysis*

153 Gene expression analysis was performed as previously described (1). Briefly, an RNA-
154 seq cohort was established by integrating UBTF cases with RNA-seq data in this study
155 (n=96) and AML in other categories from the published study (1) (n=837) and cbCD34+
156 cells (n=5). Reads from aligned RNA-Seq BAM files were assigned to genes and
157 counted using HTSeq (v0.11.2) (19) with the GENCODE human release 19 gene

158 annotation. The count data were transformed to log₂-counts per million (log₂CPM)
159 using Voom available from R package Limma (v3.50.3) (20). The top variable genes
160 were selected using the "vst" method in Seurat package (21). The expression data were
161 then scaled, and PCA (Principal Component Analysis) was performed on the scaled
162 data using the top 265 variable genes. Dimension reduction was performed using
163 UMAP (Uniform Manifold Approximation and Projection) (22) with the top 100 principal
164 components. Differential gene expression analysis was performed by Limma between
165 groups as indicated in each figure, and we set Log₂ CPM = -1 if it is < -1 based on the
166 Log₂ CPM data distribution. *P* values were adjusted by the Benjamini-Hochberg method
167 to calculate the false discovery rate (FDR) using R function `p.adjust`. Genes with
168 absolute fold change > 2 and FDR < 0.05 were regarded as significantly differentially
169 expressed. Gene Set Enrichment Analysis (GSEA) was performed by GSEA (v4.2.3)
170 using MSigDB gene sets `c2.all` (v7.5.1) (23). Permutations were done 1000 times
171 among gene sets with sizes between 15 and 1500 genes.

172

173 *Statistics*

174 Details about statistical comparisons are provided in each figure legend. All the
175 computations were done using R or GraphPad Prism, and all *P* values are 2-sided.

176 **Results:**

177 *UBTF-TD in pediatric myeloid neoplasms.*

178 Our previous study described the molecular features of 27 pediatric AML cases with
179 tandem duplications in exon 13 of *UBTF* (*UBTF*-TD) (4). To expand the cohort and
180 better understand the biology, we screened RNA-sequencing data of pediatric MDS and

181 AML cases and identified an additional 68 cases from available datasets and previously
182 published studies (2, 24) and routine clinical service at St. Jude Children's Research
183 Hospital. All 95 cases (median age = 14 yrs., range = 2.4-27.4) possessed exon 13
184 *UBTF*-TDs encoding a consensus hydrophobic leucine-rich ELTRLLARM motif within
185 the duplications (**Figure 1A, Supplemental Table 1**). The duplications resulted in an
186 increased size of exon 13 (median size = 60 bp, range = 45-339) (**Figure 1B**).
187 Consistent with previous findings, *UBTF*-TD did not co-occur with other subtype-
188 defining alterations and showed high variant allele frequencies (VAFs, median = 36.3%,
189 Interquartile range (IQR) = 10.2%), further supporting our previous assertion that *UBTF*-
190 TD alterations are early clonal events (1, 4) (**Figure 1C**). We further investigated the
191 mutational background of this *UBTF*-TD cohort, confirming a strong association with a
192 normal or trisomy 8 karyotype (**Figure 1C-E**), as well as with internal tandem
193 duplications (ITD) in *FLT3* (n=55, 57.9%) and mutations in *WT1* (n=39, 41.1%), which
194 are highly co-occurring (**Figure 1F**, $P=0.011$, Fisher's exact test) with 30.1% (n=29)
195 cases harboring both alterations. In addition, 26.6% (n=25) of cases also had at least
196 one mutation in Ras-MAPK pathway genes: *NRAS* (n=17, 17.9%); *PTPN11* (n=5, 5.3%);
197 *RIT1* (n=5, 5.3%); *NF1* (n=4, 4.2%); *CBL* (n=2, 2.1%); and *KRAS* (n=2, 2.1%). Other
198 recurrent mutations in myeloid malignancies were rarely observed, including *IDH1/IDH2*
199 (n=3, 3.2%), *BCOR* (n=2, 2.1%), and *RUNX1* (n=1, 1.1%). Although our previous
200 studies only evaluated AML samples, we also identified *UBTF*-TD in 6 cases of MDS
201 with normal karyotype. These include 3 out of 46 primary MDS cases (6.2%) from our
202 previously published cohort (2). Five of the six MDS cases were classified as childhood
203 MDS with increased blasts according to the current WHO classification (25) and lacked

204 known germline predispositions to pediatric MDS or monosomy 7/del7q. No *FLT3*-ITD
205 mutations were detected in the 6 MDS cases, while *WT1* mutations were present in 3
206 out of the 6 MDS cases, suggesting clonal evolutionary patterns initiating with *UBTF*-TD
207 alterations.

208

209 To address clonal evolution in *UBTF*-TD myeloid neoplasms, we utilized a droplet-
210 based single cell multi-omics platform from MissionBio (**Figure 2**)(26). This platform
211 enables the concurrent detection of *UBTF*-TD alterations and somatic mutations present
212 by a custom targeted DNA panel and cell classification by using DNA-oligo conjugated
213 antibodies targeting cell surface markers at the single cell level (**Supplemental Table**
214 **2**). In a single timepoint AML case, we found a clonal *UBTF*-TD alteration in the myeloid
215 population and subclonal mutations in *NRAS* (p.G12D), *FLT3* (p.V592D), and *WT1*
216 (p.359fs), suggesting *UBTF*-TD as an initiating event (**Supplemental Table 3, Figure**
217 **2A-B**). A *WT1* mutation was broadly present in the myeloid population, while *NRAS* or
218 *FLT3* mutations were subclonal to *WT1* in a mutually exclusive manner, representing
219 branched evolution of the *UBTF*-TD-*WT1*⁺ clone. We also found that cells with
220 *WT1*⁺*FLT3*⁺ mutations were associated with high stem cell marker expression (CD34,
221 CD117, or CD123) compared with the *UBTF*-TD-only population, whereas the
222 *WT1*⁺*NRAS*⁺ population was characterized by low expression of these markers. In a
223 diagnosis and relapse-paired case, we found that the identical *UBTF*-TD was retained
224 through disease progression along with a *FLT3*-ITD alteration (**Figure 2C-D**).
225 Interestingly, a minor *WT1*⁺ (p.R375fs) subclone at diagnosis was eradicated after
226 chemotherapy, whereas a different *WT1*⁺ (p.R353fs) clone became dominant at relapse,

227 showing high expression of CD34, CD117, and CD123. These data collectively confirm
228 that *UBTF*-TD is an early initiating event, while somatic mutations are subclonal to
229 *UBTF*-TD, possibly contributing to disease progress toward subclones with unique
230 expression profiles.

231

232 *Clinical features of UBTF-TD pediatric myeloid neoplasms*

233 *UBTF*-TD AML showed a variety of morphologic features associated with cellular
234 differentiation, as evidenced by variable FAB (French-American-British) classifications
235 (**Figure 3A, Supplemental Table 4**). Although AML with maturation (FAB M2) was the
236 most common (19/43, 44.2%), cases with FAB M6 (erythroid) features were also
237 observed, as was also recently described for *UBTF*-TD AMLs in adults(9). Morphologic
238 assessments also revealed that *UBTF*-TD cases often displayed pleomorphic blasts
239 (**Figure 3B**), accompanied by background multilineage dysplasia and increased
240 erythroid precursors. *UBTF*-TD AMLs showed lower white blood cell count (median =
241 $8.4 \times 10^9/L$, IQR = $41.4 \times 10^9/L$) and bone marrow blast percentage (mean = 33%, IQR =
242 41.4%) when compared to other AMLs, including those with similar transcriptional
243 profiles like AML with *NUP98*-rearrangements or *NPM1* mutations(1, 4) (**Figure 3C-D**).
244 *FLT3*-ITD, but not *WT1* mutations, were associated with a higher WBC count and bone
245 marrow blasts in *UBTF*-TD AMLs (**Figure 3E**). Despite the presence of dysplastic
246 features, cytogenetic studies commonly found either a normal karyotype (58/95, 61.1%)
247 or trisomy 8 (27/95, 28.4%), and myelodysplasia-related chromosomal changes or
248 myelodysplasia-related mutations were overall rare, suggesting that *UBTF*-TD itself
249 contributes to dysplastic features (**Supplemental Table 4**). Considering these overall

250 features, the majority of *UBTF*-TD AMLs (83/89, 93.3%) are best classified as “Acute
251 myeloid leukemia with other defined genetic alterations” in the current WHO
252 classification(25) (**Supplemental Table 4**).

253

254 *Transcriptional features of UBTF-TD myeloid neoplasms*

255 We and others have previously shown that AML with *UBTF*-TD is characterized by high
256 *HOXA* and *HOXB* cluster gene expression, similar to *NPM1*-mutated or *NUP98::NSD1*
257 AML (1, 4, 8). To further define the unique expression profiles of *UBTF*-TD, we
258 established an RNA-seq cohort consisting of various AML subtypes (1) (n=837), cord
259 blood CD34+ samples from healthy donors (n=5), and *UBTF*-TD AML and MDS
260 samples (n=94: 1 *UBTF*-TD case did not have RNA-seq data available) (**Figure 4A**).

261 Consistent with previous data, *UBTF*-TD cases clustered with *NPM1*-mutated and
262 *NUP98::NSD1* AMLs. However, *UBTF*-TD displayed a significantly higher expression of
263 a subset of *HOXB* cluster genes (e.g., *HOXB8*, *HOXB9*) compared with *NUP98::NSD1*
264 AML (**Figure 4B**). We also observed uniquely high expression of histone genes (e.g.,
265 *HIST1H4F* and *HIST1H2B1*) compared to *NPM1*-mutated AML (**Figure 4C**), suggesting
266 transcriptional mechanisms unique to *UBTF*-TD AML. Within *UBTF*-TD samples, those
267 with *FLT3*-ITD and *WT1* mutations showed unique distribution on the UMAP cluster
268 (**Figure 4D**), and each mutation group demonstrated differential gene expression
269 against *UBTF*-TD samples without either mutation (**Figure 4E**). Co-occurrence of *WT1*
270 and *FLT3*-ITD was associated with stemness-related genes (e.g., *CD34* and *DNTT*,
271 **Figure 4E-F**), and Gene Set Enrichment Analysis (GSEA) confirmed enrichment of
272 stemness or cell cycle-related gene expression in *WT1*⁺*FLT3*-ITD⁺ *UBTF*-TD samples

273 **(Figure 4G)**. These results show the unique expression profile of *UBTF*-TD AML and
274 the specific influence of additional cooperating mutations, which can likely influence
275 patterns of clonal evolution.

276

277 *Exon 9 tandem duplications in UBTF*

278 Given the recurrent *UBTF* exon 13 alterations duplicating specific hydrophobic amino
279 acid sequences, we hypothesized that *UBTF* alterations outside exon 13 resulting in
280 similar amino acid sequences could be found in cases without defining alterations but
281 with a similar expression signature. By close inspection of the *UBTF* gene using RNA-
282 sequencing data, we found two pediatric AML cases without exon 13 *UBTF*-TD or other
283 driver alterations that instead have in-frame tandem duplications (lengths of 78 and
284 153bp) in exon 9 of *UBTF* (TD-exon9), encoding short hydrophobic amino acid
285 sequences (**Figure 5A-B**). These cases express *HOXA/B* cluster genes comparably to
286 exon 13 *UBTF*-TD (**Supplemental Figure 1**), and one had a *WT1* mutation
287 (**Supplemental Table 5**). To test whether these exon 9 duplication events could lead to
288 leukemic transformation, we expressed both exon 9 duplications in cbCD34+ cells using
289 lentiviral vectors and assessed their impact on cell proliferation, clonogenic potential,
290 and cellular morphology in comparison with control conditions and exon 13 *UBTF*-TD
291 (**Figure 5C-D**). Colony forming unit (CFU) assay revealed that expression of both
292 *UBTF*-TD-exon9 increased total colony number (**Figure 5E**). After the second round of
293 replating, cells with *UBTF*-TD-exon9 showed an immature morphology along with
294 erythroid features, similar to exon 13 *UBTF*-TD expressing cells (**Figure 5F**).
295 Furthermore, cells expressing *UBTF*-TD-exon9 as well as *UBTF*-TD-exon13 showed a

296 proliferative advantage compared to *UBTF*-WT and vector controls (**Figure 5G**).
297 Collectively, these data highlight a tandem duplication in *UBTF* exon 9 as a defining
298 alteration functionally equivalent to exon 13 tandem duplications.

299

300 **Discussion:**

301 In this study, we extended our cohort of pediatric myeloid malignancies with exon 13
302 *UBTF*-TD, which now includes 95 pediatric cases. Similar to studies in adults (9),
303 pediatric myeloid tumors with *UBTF*-TD are associated with lower bone marrow blast
304 infiltration and lower white blood cell count, suggesting a similar clinical presentation. In
305 this pediatric cohort, we observed that *UBTF*-TD is associated with an M2 morphology
306 subtype (44.2%) according to FAB classification (27). These morphological features
307 align with findings in adult cases, which showed a high prevalence of FAB M6 and M2
308 cases (9). Furthermore, the presence of dysplastic features and the observation that
309 *UBTF*-TD occurs in cases of pediatric MDS suggest that MDS and AML could be part of
310 a continuum driven by *UBTF*-TD. Progression to AML may require the acquisition of
311 cooperating mutations such as *FLT3*-ITD, supported by the finding that none of the
312 MDS cases in this cohort harbor a *FLT3*-ITD alteration.

313

314 To evaluate the impact of these cooperating mutations on cellular status, we analyzed
315 RNA-seq data of *UBTF*-TD cases with or without mutations, showing that the co-
316 occurrence of *FLT3*-ITD and *WT1* is strongly associated with stem cell features and
317 represented by CD34 expression or quiescent states. These findings were further
318 supported by single cell studies. Given that *WT1* mutations are often subclonal

319 alterations that expand at relapse and that the co-occurrence of these mutations is
320 associated with particularly dismal outcomes (4), acquiring these stem-like features can
321 contribute to resistance to conventional chemotherapy. Furthermore, a recent study
322 showed that UBTF-TD localizes to HOXA/B genomic regions and that Menin inhibition is
323 sufficient to promote terminal differentiation (28). How these other mutations co-operate
324 with UBTF-TD to maintain stem cell features requires further evaluation.

325

326 Given that exon 13 *UBTF*-TD has been underappreciated in standard computational
327 pipelines, we investigated other possible *UBTF* alterations in pediatric AML cases
328 without a defined driver event and demonstrated that in-frame tandem duplications in
329 exon 9 of *UBTF* are possible driver alterations. Although rare within cases with *UBTF*
330 alterations (2/97 *UBTF*-TD cases, 2.1%), cases with tandem duplications in exon 9
331 show similar transcriptional profiles to exon 13 *UBTF*-TD AML. We further show that
332 exon 9 alterations can induce leukemic changes, including hematopoietic cell growth
333 and increased clonogenicity in cbCD34+ cells similar to exon 13 alterations. At the
334 amino acid level, exon 9 tandem duplications contain amino acid sequences
335 (LKDKFDGLTI) that share hydrophobic α -helix structure with recurrently duplicated
336 sequences in exon 13 tandem duplications (LTRLLARM), suggesting a shared
337 mechanism. Importantly, PCR-based screening can underestimate PTD or these exon 9
338 alterations since they are outside the covered regions(9), and we propose that
339 enhanced sequencing-based strategies are needed to accurately diagnose this entity.
340 The findings presented here will help build on our understanding of UBTF-TD myeloid

341 neoplasms and further support it's recognition as a distinct entity in future classification
342 systems.

343

344

345 **Acknowledgements:**

346 The work was funded by the American Lebanese and Syrian Associated Charities of St.
347 Jude Children's Research Hospital and funds from the US NIH, including F32 HL154636
348 (JMB), U54 CA243124 and R01 CA276079 (JMK). The content, however, does not
349 necessarily represent the official views of the NIH and is solely the responsibility of the
350 authors. The studies were also funded by the Jane Coffin Childs Fund (JMB). JMK
351 holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund.
352 Support was also provided by Shared Resources provided through the St. Jude
353 Comprehensive Cancer Center (P30-CA21765), Flow Cytometry and Cell Sorting,
354 Comparative Pathology Core, and Genome Sequencing (Hartwell Center).

355

356 **Author Contributions:**

357 Contributions: Conceptualization, J.M.K; Methodology, J.M.B., L.C., E.X.; Software,
358 Validation, J.M.B., L.C., E.X.; Formal Analysis, J.M.B., M.U., B.O., M.A.L., M.K.;
359 Investigation, J.M.B., M.U., L.C., M.P.W., M.K., M.E.T., X.M.; Resources, J.M.K.; Data
360 Curation, T.W., G.S.; Writing – Original Draft, J.M.B., M.U.; Writing – Review & Editing,
361 all authors were involved in review & editing of the manuscript; Visualization, J.M.B.,
362 M.U.; Supervision, J.M.K., Project Administration, J.M.B., M.U., T.W., S.A., J.M.K.;
363 Funding Acquisition, J.M.B., J.M.K.

364

365 **Competing Interests:** C.Y., B.O., and M.A. are employed by Mission Bio, Inc. No other
366 authors have conflicts to declare.

367

368 **Data Availability Statement:**

369 Samples from patients with MDS or AML from St. Jude Children’s Research Hospital
370 tissue resource core facility were obtained with written informed consent using a
371 protocol approved by the St. Jude Children’s Research Hospital institutional review
372 board (IRB). Studies were conducted in accordance with the International Ethical
373 Guidelines for Biomedical Research Involving Human Subjects. The expression data
374 newly generated in this study (RNA-Seq: n=3) and scDNA + protein sequencing (n=3)
375 have been deposited in the European Genome-Phenome Archive (EGA) which is
376 hosted by the European Bioinformatics Institute (EBI), under accession
377 [EGAS00001005760](https://ega-archive.org/studies/EGAS00001005760). The remaining RNA-Seq data are available via EGA, St. Jude
378 Cloud or TARGET as defined in **Supplemental Table 6**. Information about TARGET
379 can be found at <http://ocg.cancer.gov/programs/target>. Other data generated in this
380 study are available in the Supplemental tables or upon request to the corresponding
381 author.

382 **Citations:**

383

384 1. Umeda M, Ma J, Westover T, Ni Y, Song G, Maciaszek J, et al. Proposal of a
385 new genomic framework for categorization of pediatric acute myeloid leukemia
386 associated with prognosis. Res Sq. 2023.

- 387 2. Schwartz JR, Ma J, Lamprecht T, Walsh M, Wang S, Bryant V, et al. The
388 genomic landscape of pediatric myelodysplastic syndromes. *Nat Commun.*
389 2017;8(1):1557.
- 390 3. Bolouri H, Farrar JE, Triche T, Jr., Ries RE, Lim EL, Alonzo TA, et al. The
391 molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural
392 alterations and age-specific mutational interactions. *Nat Med.* 2018;24(1):103-12.
- 393 4. Umeda M, Ma J, Huang BJ, Hagiwara K, Westover T, Abdelhamed S, et al.
394 Integrated genomic analysis identifies UBTF tandem duplications as a recurrent lesion
395 in pediatric acute myeloid leukemia. *Blood Cancer Discov.* 2022.
- 396 5. Kaburagi T, Shiba N, Yamato G, Yoshida K, Tabuchi K, Ohki K, et al. UBTF-
397 Internal tandem duplication as a novel poor prognostic factor in pediatric acute myeloid
398 leukemia. *Genes Chromosomes Cancer.* 2022.
- 399 6. Stratmann S, Yones SA, Mayrhofer M, Norgren N, Skaftason A, Sun J, et al.
400 Genomic characterization of relapsed acute myeloid leukemia reveals novel putative
401 therapeutic targets. *Blood Adv.* 2021;5(3):900-12.
- 402 7. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer
403 genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours.
404 *Nature.* 2018;555(7696):371-6.
- 405 8. Georgi JA, Stasik S, Eckardt JN, Zukunft S, Hartwig M, Rollig C, et al. UBTF
406 tandem duplications are rare but recurrent alterations in adult AML and associated with
407 younger age, myelodysplasia, and inferior outcome. *Blood Cancer J.* 2023;13(1):88.
- 408 9. Duployez N, Vasseur L, Kim R, Largeaud L, Passet M, L'Haridon A, et al. UBTF
409 tandem duplications define a distinct subtype of adult de novo acute myeloid leukemia.
410 *Leukemia.* 2023.
- 411 10. Sanij E, Hannan RD. The role of UBF in regulating the structure and dynamics of
412 transcriptionally active rDNA chromatin. *Epigenetics.* 2009;4(6):374-82.
- 413 11. Moss T, Mars JC, Tremblay MG, Sabourin-Felix M. The chromatin landscape of
414 the ribosomal RNA genes in mouse and human. *Chromosome Res.* 2019;27(1-2):31-
415 40.
- 416 12. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler
417 transform. *Bioinformatics.* 2010;26(5):589-95.
- 418 13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The
419 Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
- 420 14. Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, et al. The genomic landscape
421 of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat*
422 *Genet.* 2014;46(5):444-50.

- 423 15. Tian L, Li Y, Edmonson MN, Zhou X, Newman S, McLeod C, et al. CICERO: a
424 versatile method for detecting complex and diverse driver fusions using cancer RNA
425 sequencing data. *Genome Biol.* 2020;21(1):126.
- 426 16. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH.
427 Bambino: a variant detector and alignment viewer for next-generation sequencing data
428 in the SAM/BAM format. *Bioinformatics.* 2011;27(6):865-6.
- 429 17. Hagiwara K, Edmonson MN, Wheeler DA, Zhang J. indelPost: harmonizing
430 ambiguities in simple and complex indel alignments. *Bioinformatics.* 2022;38(2):549-51.
- 431 18. Hagiwara K, Ding L, Edmonson MN, Rice SV, Newman S, Easton J, et al.
432 RNAIndel: discovering somatic coding indels from tumor RNA-Seq data. *Bioinformatics.*
433 2020;36(5):1382-90.
- 434 19. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-
435 throughput sequencing data. *Bioinformatics.* 2015;31(2):166-9.
- 436 20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
437 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic
438 Acids Res.* 2015;43(7):e47.
- 439 21. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of
440 single-cell gene expression data. *Nat Biotechnol.* 2015;33(5):495-502.
- 441 22. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al.
442 Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.*
443 2018.
- 444 23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et
445 al. Gene set enrichment analysis: a knowledge-based approach for interpreting
446 genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50.
- 447 24. Fornerod M, Ma J, Noort S, Liu Y, Walsh MP, Shi L, et al. Integrative Genomic
448 Analysis of Pediatric Myeloid-Related Acute Leukemias Identifies Novel Subtypes and
449 Prognostic Indicators. *Blood Cancer Discov.* 2021;2(6):586-99.
- 450 25. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th
451 edition of the World Health Organization Classification of Haematolymphoid Tumours:
452 Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia.* 2022;36(7):1703-19.
- 453 26. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, et
454 al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature.*
455 2020;587(7834):477-82.
- 456 27. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al.
457 Proposed revised criteria for the classification of acute myeloid leukemia. A report of the
458 French-American-British Cooperative Group. *Ann Intern Med.* 1985;103(4):620-5.

459 28. Barajas JM, Rasouli M, Umeda M, Hiltenbrand RL, Abdelhamed S, Mohnani R,
460 et al. Acute myeloid leukemias with *UBTF* tandem duplications are sensitive to Menin
461 inhibitors. *Blood*. 2023.

462

463 **Figure Legends:**

464 **Figure 1. Characterization of Tandem Duplications in Exon 13 of *UBTF*.** **A.** Genomic
465 representation of duplicated regions (grey) from 95 cases with *UBTF* exon 13 duplications. A
466 common duplicated region is outlined in red with the alignment of the resulting amino acid
467 sequence. **B.** Size assessment of exon 13 resulting from the duplicated regions and associated
468 insertion/deletions. For partial tandem duplications (n=13) extending beyond exon 13, sizes of
469 duplicated exon 13 with insertions or deletions are included. **C.** Genomic landscape of *UBTF*-TD
470 cases. For each case, variant allele frequency (VAF), diagnosis, karyotype, and sex are shown.
471 **D.** Distribution of cytogenetic changes among *UBTF*-TD cases. **E.** Ribbon plot depicting
472 association of cytogenetics with mutations common in *UBTF*-TD myeloid neoplasms. Genes
473 coding for proteins involved in RAS/MAPK pathway are grouped together (*NF1*, *PTPN11*,
474 *NRAS*, *KRAS*, *RIT1*, *CBL*). All other mutations in *FLT3* not classified as ITD were grouped into
475 *FLT3*-other. **F.** Ribbon plot depicting the association of co-occurring mutations in the *UBTF*-TD
476 cohort. Genes with only one occurrence were excluded.

477

478 **Figure 2. Clonal Dynamics of *UBTF*-TD Leukemias.** **A.** Single-cell DNA sequencing coupled
479 with surface marker expression of *UBTF*-TD case at diagnosis. Heatmap depicting the presence
480 of mutant allele (bottom) and relative protein abundance (top). **B.** Schematic of clonal structure
481 in *UBTF*-TD case from A. Percentages are calculated as a proportion of total cells with a
482 somatic mutation. **C.** Single-cell DNA sequencing of a *UBTF*-TD case with diagnosis (left) and
483 relapse (right) paired samples. Heatmap depicts the presence of a mutant allele and protein

484 expression. **D.** Schematic of clonal dynamics in diagnosis/relapse case from C. Percentages are
485 calculated as a proportion of total cells with a somatic mutation.

486

487 **Figure 3. Morphological and Biological Assessment of *UBTF*-TD.** **A.** Available FAB
488 (French-American-British) classification of *UBTF*-TD AMLs (n=43). **B.** Representative Wright-
489 Giemsa Staining of bone marrow (BM) aspirates and peripheral blood (PB) smears from 4
490 unique *UBTF*-TD cases. Case ID, FAB/WHO classification, along with *FLT3* and *WT1* mutation
491 status, are labeled above. **UBTF-TD-3-Childhood MDS with increased blasts:** increased
492 erythroid cells with dysplastic morphologic features (top), dysplastic myeloid cells with salmon-
493 colored granules, and dysplastic small megakaryocytes (bottom). **UBTF-TD-78-AML with other
494 defined genetic alterations (Acute myeloid leukemia with maturation):** characteristic blasts
495 and myeloid precursor cells with salmon-colored granules (top) and blasts with small Auer rods
496 (bottom). **UBTF-TD-91-AML with other defined genetic alterations (Acute myelomonocytic
497 leukemia):** blasts with both myeloid and monocytic/monoblastic morphologic features. PB in this
498 patient is one of the few cases in our cohort showing hyperleukocytosis with many circulating
499 blasts (bottom). **UBTF-TD-90-AML with other defined genetic alterations (Acute
500 myelomonocytic leukemia):** dysplastic megakaryocytes, increased erythroid cells with
501 dysplastic morphologic features, and blasts, many with monocytic/monoblastic morphologic
502 features. **C.** White blood cell counts (WBC) for a pediatric AML cohort stratified by oncogenic
503 driver subtypes. **D.** Bone marrow blast percentage for a pediatric AML cohort stratified by
504 oncogenic driver subtypes. **E.** WBC count and BM blast percentage among *UBTF*-TD cases
505 with different *WT1* and *FLT3* mutation status.

506

507 **Figure 4. Transcriptional characterization of *UBTF*-TD leukemias.** **A.** UMAP (Uniform
508 Manifold Approximation and Projection) of expression profiles across a cohort of *UBTF*-TD

509 (n=94) and pediatric AML (n=839), adapted from a previous study(4). Each dot is colored by
510 subtype-defining alterations. The black box outlines a cluster of cases with *HOXA/HOXB*
511 dysregulation, which includes all *UBTF*-TD cases. **B.** Differentially expressed genes between
512 *UBTF*-TD AML and *NUP98::NSD1* pediatric AML. Representative genes are annotated. **C.**
513 Differentially expressed genes between *UBTF*-TD AML and *NPM1*-mutated pediatric AML. **D.**
514 Distribution of *UBTF*-TD cases within the cluster with *HOXA/B* dysregulation. Each data is
515 colored by the mutational status of *UBTF*-TD, *FLT3*-ITD, or *WT1*. MDS cases are depicted with
516 triangles, and AML cases are with circles. **E-top.** Schematic depicting the transcriptional
517 comparison of *UBTF*-TD cases by mutational status. **-bottom.** Venn diagram showing overlap
518 of differentially expressed genes in *FLT3*-ITD only cases and *FLT3*-ITD+/*WT1*+ cases
519 compared with cases without these mutations. **F.** Expression *CD34* and *DNTT* with respect to
520 *FLT3* and *WT1* mutational status. **G.** GSEA of *UBTF*-TD cases based on mutational status.

521

522 **Figure 5. Exon 9 *UBTF* Tandem Duplications in Pediatric AML.** **A.** Identification of *UBTF*-TD
523 cases with tandem duplications in exon 9 of *UBTF*. *UBTF*-TD-exon9 on the UMAP plot of the
524 Pediatric AML cohort (red circles), adapted from a previous study(4). **B.** Schematic of *UBTF*
525 protein, highlighting amino acid sequences encoded in exon 9 of *UBTF*-WT, *UBTF*-TD78-exon9,
526 *UBTF*-TD153-exon9. Duplications resulting in short hydrophobic sequences are labeled in red.
527 Hydrophobic residues are underlined. **C.** Experimental design to evaluate the transforming
528 potential of *UBTF*-TD-exon9 mutants **D.** Immunoblot of lysates from cbCD34+ transduced with
529 *UBTF*-TD-exon9 vectors and corresponding controls. Antibodies against HA-tag, *UBTF*, and β -
530 actin were used. **E.** Colony forming unit assay. 2-way ANOVA with two-sided Dunnett's test was
531 applied with *UBTF*-WT as control **F.** Cytospins isolated from CFU assay at round 2 of replating
532 (Wright-Giemsa Staining, 400X magnification). **G.** In vitro growth curves of cbCD34+ cells
533 expressing controls or *UBTF*-TD-e9 mutants. 2-way ANOVA with Dunnett's test was applied
534 with *UBTF*-WT as control.

535

536

537

538

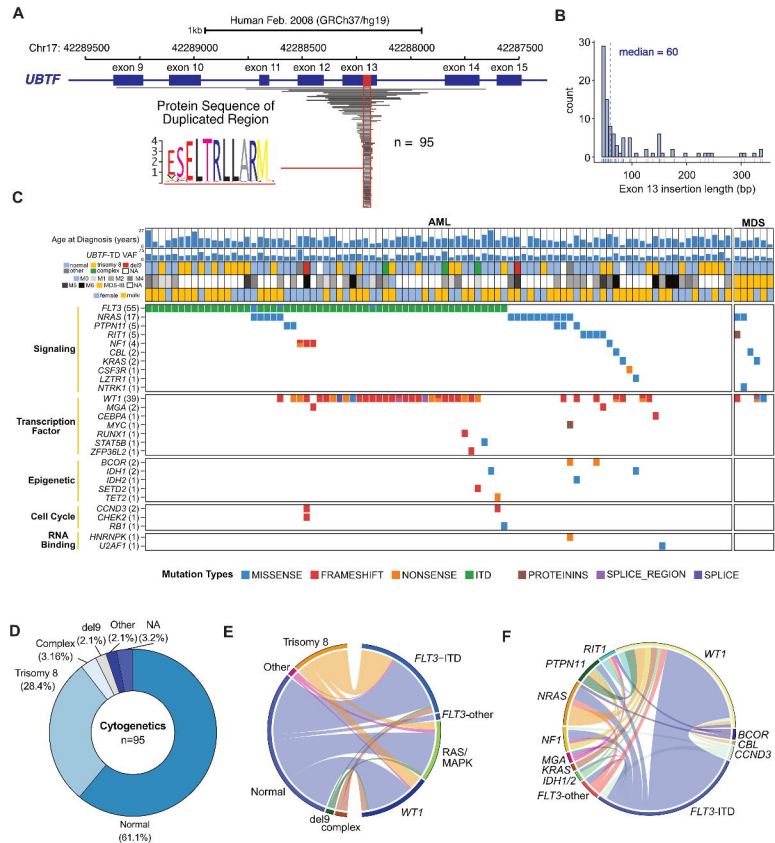
Figure 1


Figure 2

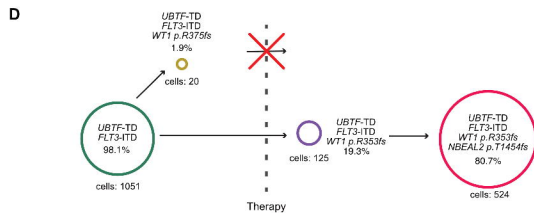
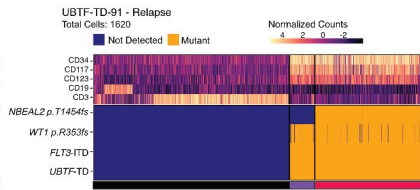
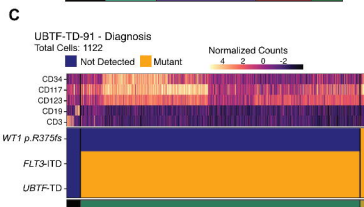
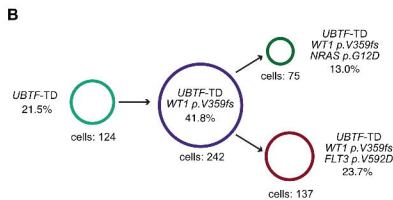
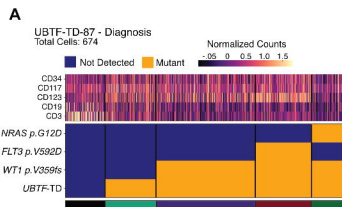


Figure 3

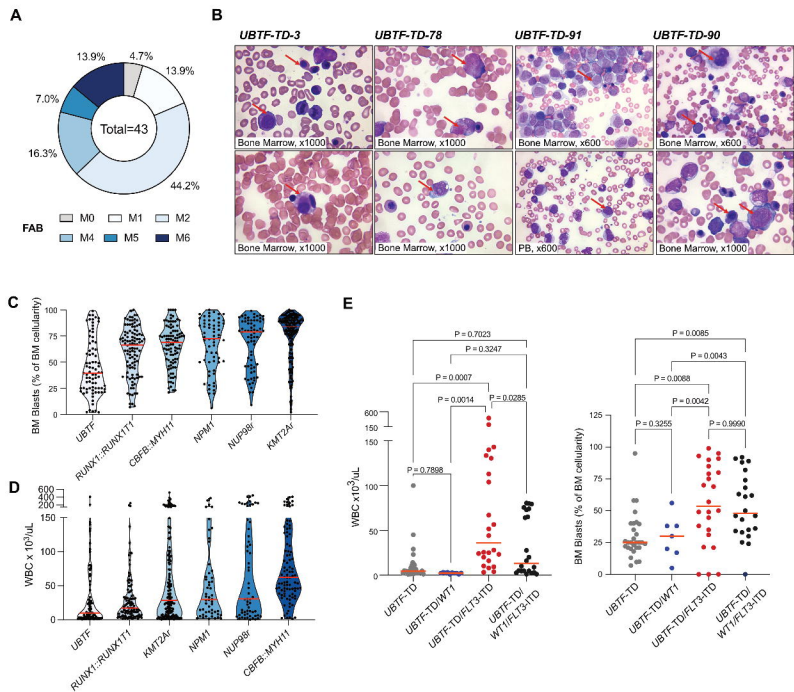


Figure 4

