Single cell analysis of clonal architecture in acute myeloid leukaemia

Nicola Potter[1], Farideh Miraki-Moud[2], Luca Ermini[1], Ian Titley[1], Gowri Vijayaraghavan[1], Elli Papaemmanuil[3], Peter Campbell[4], John Gribben[2], David Taussig[5], Mel Greaves[1]

[1]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

[2]Barts Cancer Institute, Queen Mary University of London, London, UK

[3]Memorial Sloan Kettering Cancer Center, New York, USA

[4]Wellcome Sanger Institute, Hinxton, UK

[5]Royal Marsden Hospital, Sutton, UK

Corresponding author:

Professor Mel Greaves FRS, Centre for Evolution and Cancer, The Institute of Cancer Research, Brookes Lawley Building, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK
Tel: +44 (0)20 8722 4073; E: mel.greaves@icr.ac.uk

Running title: Clonal architecture in AML

Key words: subclones; nucleophosmin1; stem cells

The authors declare no conflict of interest.

ABSTRACT

We used single cell Q-PCR on a micro-fluidic platform (Fluidigm) to analyse clonal, genetic architecture and phylogeny in acute myeloid leukaemia (AML) using selected mutations. Ten cases of *NPM1*c mutant AML were screened for 111 mutations that are recurrent in AML and cancer. Clonal architectures were relatively simple with one to six sub-clones and were branching in some, but not all, patients. *NPM1* mutations were secondary or sub-clonal to other driver mutations (*DNM3TA, TET2, WT1* and *IDH2*) in all cases. In three of the ten cases, single cell analysis of enriched $CD34^{+}/CD33^{-}$ cells revealed a putative pre-leukaemic sub-clone, undetectable in the bulk $CD33^{+}$ population that had one or more driver mutations but lacked *NPM1*c. Cells from all cases were transplanted into NSG mice and in most (8/10), more than one sub-clone (#2-5 sub-clones) transplanted. However, the dominant regenerating sub-clone in 9/10 cases was *NPM1*$^{+}$ and this sub-clone was either dominant or minor in the diagnostic sample from which it was derived. This study provides further evidence, at the single cell level, for genetic variegation in sub-clones and stem cells in acute leukaemia and demonstrates both a preferential order of mutation accrual and parallel evolution of sub-clones.

INTRODUCTION

Although almost all cancers originate in a single cell, the sequential acquisition of necessary additional mutations fuels sub-clonal diversity which is then a substrate for positive or negative selection within the tissue ecosystems and with therapy [1, 2]. This process frequently results in complex cell population structures and highly variegated genetics [3-5].

The genomics of AML have been described in considerable detail, revealing multiple sub-types [6, 7] and sequential transition between clinically silent pre-leukaemia and overt disease [8, 9]. Sub-clonal architectures in diagnostic samples have been inferred from allele

50 burdens [10] and appear to be relatively simple compared to that observed in many common
51 cancers [11, 12].
52 The extent of sub-clonal diversity and phylogenetic architecture is, however, best
53 derived from single cell analysis. This is challenging however in terms of accuracy and
54 depth, but has proven illuminating in some solid tumours [13] and ALL [3].
55 Only a few studies to date have reported single cell genetics and inferred sub-clonal
56 phylogenies in AML. Paguirigan *et al* [14] used single cell, multiplexed Q-PCR to investigate
57 patterns of segregation of two concurrent mutations in AML – *FLT3*-ITD and *NPM1*c
58 mutations. The data revealed significantly more sub-clonal diversity than could be inferred
59 from analysis of the bulk population. Klco *et al* [15] fractionated immuno-phenotypically
60 distinct cell populations from a patient with AML and sequenced the amplified DNA from
61 single cells for ten known mutations. From these data, they could infer a branching sub-
62 clonal architecture. Jan *et al* used a Q-PCR assay on colonies derived from sorted single cells
63 derived from two cases of AML and were able to infer an ordered sequence of mutations [8].
64 Quek *et al* screened single cells for targeted mutations in immunophenotypically-defined
65 subsets and identified putative clonal sequences and mutation order in six cases [16].
66 In our previous studies in ALL, we used multi-colour FISH or multi-plexed Q-PCR in
67 a micro-fluidic platform (Fluidigm) to detect sub-clonal variegation and clonal architecture [3,
68 17]. In this study, we sought to replicate our observations on ALL for AML, selecting the
69 subset of cases with *NPM1*c mutations. *NPM1c*$^+$ cases constitute around 27% of adult AML
70 with a variable but overall intermediate risk [7]. The questions posed included the extent of
71 sub-clonal complexity that was discernible, sequential order of mutations and whether stem
72 cells or leukaemia propagating cells, assayed by xeno-transplantation, were genetically
73 variable.
74

MATERIALS AND METHODS

Sample cohort

A total of ten well-characterized *NPM1* mutant AML samples [18] were selected for further study according to engraftment potential. Blood and marrow samples were collected from patients with AML after written informed consent at St Bartholomew's Hospital. The protocol was approved by the East London and City Research Ethics Committee. All studies comply with the rules of the revised Helsinki protocol. These had all been found to successfully transplant in NOD/SCID mice [18]. This selection criterion may have biased our analysis towards poorer prognosis cases [7, 18]. Available peripheral blood was collected prior to treatment at presentation (n = 10) and from matched relapse (n = 3). Mononuclear cells were obtained by density gradient centrifugation. Details of the patient samples are listed (Table 1).

FACS cell sorting according to immunophenotype details can be found in Supplementary Information.

Mutation analysis

A targeted screening approach investigating 111 genes (Table 2) was used to identify mutations and DNA coding region alterations in each *NPM1*c AML as previously described [7] that could potentially be tracked in single cells. The analysis is based on variants that can be classified as recurrent driver mutations, using widely accepted genetic criteria. These included non-synonymous base substitutions and small (<200-bp) insertions or deletions (indels). Table 3 lists the probes used for mutant versus wild type sequences and PCR primers.

Sequencing data

100 For the targeted mutation screening of each leukaemia, two populations of interest were
101 stained and sorted as described in Supplementary Information and DNA extracted (Qiagen®
102 DNA blood kit according to manufacturers' instructions): peripheral blood T-cells
103 ($CD3^+/CD33^-$) (as a control) and mononuclear blast cells ($CD3^-/CD33^+$). The latter had <1%
104 $CD34^+$ cells and we refer to this population as $CD34^-$.

105 For details of library preparation, sequencing, alignment and analysis, please refer to
106 Supplementary Information.

107

108 **Xeno-transplantation**

109 NOD/SCID ($Il2rg^{-/-}$) mice (Jackson Laboratory, Bar Harbor, ME) were injected intravenously
110 (3 mice per AML sample) with 9-10 million AML cells after T-cell depletion by Easysep T-
111 cell enrichment cocktail (Stem Cell Technologies). Mice were bled by tail veins at 12-14
112 weeks and blood leukocytes investigated by FACS (as described in Supplementary
113 Information and Supplementary Figures 2 and 3) using anti-human and anti-mouse CD45
114 antibodies to determine the percentage of leukaemic cell engraftment. For details of how
115 successful/undetectable/minimal grafts were managed and serial transplantations were carried
116 out, please refer to Supplementary Information.

117

118 **Single cell sorting and multiplex Q-PCR analysis**

119 Single cell sorting was carried out (see Supplementary Information and Supplementary
120 Figure 1) according to our established published Q-PCR single cell (Fluidigm) protocol [17].
121 Briefly, from each case single AML cells (either $CD33^+/CD34^-/CD3^-$ (blast population),
122 $CD3^+/CD33^-$ (internal control), $CD34^+/CD33^-$ (putative stem cell), $CD45^+$ (human cells post-
123 transplant) or cord blood cells (normal diploid control) were sorted into individual wells of a
124 96 well plate, lysed and DNA target amplification completed for regions of interest

125 encompassing patient specific mutations or DNA alterations. Allelic discrimination Q-PCR
126 assays were designed specifically for each mutation in every patient. Standard Q-PCR assays
127 targeting unique *FLT3-ITDs* were designed for each positive patient. Genes targeted in each
128 case are listed in Table 1. The *ß2M* locus, located in a diploid region of the genome, was
129 used as a control. Q-PCR completed using the 48 x 48 dynamic array and the BioMark™ HD
130 from Fluidigm.

131 Several approaches were adopted during this experiment to optimise and confirm the
132 presence of a single cell and ensure all assays performed efficiently under experimental
133 conditions [17]; a brief description can be found in Supplementary Information and
134 Supplementary Figure 4.

135

136 **Maximum parsimony**

137 Maximum parsimony searches for sub-clonal phylogenies were conducted using heuristic
138 searches as previously described [17]; a brief description can be found in Supplementary
139 Information.

140

141 **RESULTS**

142 Our targeted exomic screening approach identified a number of common or recurrent driver
143 SNV mutations in each patient's diagnostic sample (Table 1); similar to those previously
144 described for *NPM1c* AML [7, 19-21]. Five of the ten cases had both *DNM3TA* and *TET2*
145 mutations, reflecting the selection of driver mutations that cooperate to confer fitness
146 advantage of haemopoietic stem cells [22]. Allele frequencies varied greatly suggested that
147 many mutations were probably sub-clonally distributed.

148 Individual cells sorted as $CD34^+/CD33^-$ or $CD33^+/CD3^-$ were assayed by multiplex Q-
149 PCR for each driver mutation identified in that patient's sample. We similarly assessed

150 individual cells (unsorted) from NSG mice in which T-cell depleted AML cells from each
151 patient had been transplanted. From those single cell data, we are able to infer a probable
152 clonal phylogeny for each case with genetically distinct sub-clones, the immunophenotype
153 and the clonal derivation of leukaemia that regenerated in NSG mice. We take the latter as a
154 read-out of sub-clones with self-renewal or stem cell activity.
155
156 *Clonal architectures*
157 Figure 1 summarises the data from all ten cases (see Supplementary Information for more
158 detailed data). This includes an identifier (t = transplant) of sub-clones that successfully
159 transplanted into mice (t1, t2 and t3 refer to individually transplanted mice using diagnostic
160 material from each patient). The phylogenetic or sub-clonal architectures inferred are
161 relatively simple and either linear or branching (three patients). The analyses are relatively
162 insensitive however with minor clones below 5% being difficult to detect. It is very likely
163 that we are significantly under-estimating clonal complexity and will have missed minor sub-
164 clones that could be clinically relevant, emerging at relapse [23].
165 In two patients (#1 and #3; Fig 1), there were more than one equally parsimonious
166 phylogenetic trees (illustrated by alternative dotted lines connecting sub-clones). We depict
167 all equally parsimonious trees for patients #1 and #3 in Supplementary Figures 5 and 6.
168 The number of identifiable sub-clones varied from one to six. In four patients (#4, #5,
169 #7 and #8) the small, putative stem cell $CD34^+/CD33^-$ fractions contained a genetic sub-clone
170 that was not discernible in the large $CD33^+$ blast population. These cells had fewer
171 mutations, lacked *NPM1*c mutation and could represent pre-leukaemic cells [8, 9].
172 We did not detect *DNM3TA* or other putative founder mutations in the T cells by
173 single cell analysis. However, in most cases reported by Shlush *et al* [24], the mutant *DNM3TA*
174 allele frequency in T cells was low and so could have been missed in our samples in which

175 only a maximum of 48 single T cells were assayed. In the total or bulk population of *NPM1c*
176 AML-derived T cells that were subject to targeted sequencing in our series of patients, the
177 calculated allele frequency for *DNM3TA* mutations and other putative driver mutations in
178 AML cells ranged from 0.64% to 4.35% in the T cell population.

179 *NPM1c* mutations were always preceded by mutations previously considered as
180 possible founders; *DNMT3A*, *IDH2*, *WT1*, *TET2* as well as some additional mutations that are
181 less well validated as early events in *NPM1c* AML including *NRAS*, *ZRSR2* and *CBL*. *FLT3*
182 mutations and *FLT3-ITDs* were found to occur both before and after the acquisition of
183 *NPM1c* but were always sub-clonal to putative founder mutations.

184

185 *Match relapsed cases*

186 In the three *NPM1*c AMLs with matched relapse samples (#8, 9, 10), we found high levels of
187 *NPM1*c sub-clones in the $CD34^{+}/CD33^{-}$ population at diagnosis ranging from 43% -100%.
188 In the single case in which the $CD34^{+}/CD33^{-}$ population could be assessed at relapse (patient
189 #8), the size of the *NPM1*c sub-clone had increased from 43% to 95%. In these AMLs it was
190 also possible to identify sub-clones at relapse or in the mice after transplant of the diagnostic
191 or relapsed material that had acquired more mutations in addition to those found in the major
192 clone at diagnosis (in #9, #10; see Table 1). Some of these mutations could not be tracked by
193 Q-PCR but were identified by direct sequencing (Table 1). Patient #10 had two *FLT3* sub-
194 clonal mutations (detected by sequencing), one at diagnosis (10.03%) rising to 41.64% whilst
195 the other was only detected at relapse (41.16%). Neither of these *FLT3* mutations could be
196 tracked, so they do not appear in patient #10 clonal structure (Fig 1).

197

198 *Reiterative mutations*

199 Reiterated mutations in individual driver genes were identified in some cases. In patient #3,

200 the two distinctive *FLT3* mutations were segregated in distinctive sub-clones. Similarly, in
201 patient #1, the two distinctive *TET2* mutations were present in separate (minor/major) sub-
202 clones. In contrast, in patient #7 the two *TET2* mutations were in the same sub-clone and
203 probably bi-allelic. Phylogenetic architectures suggested that *NPM1*c mutations may also
204 have been reiterative in some cases, for example with patient #3 (and possibly patient #1) but
205 the invariant nature of this mutation makes this more ambiguous.
206
207 *Stem cell read-outs in transplants*
208 The single cell genetics of regenerated leukaemias in mice (see t1,t2,t3 % in Fig 1) allowed
209 us to infer the sub-clonal origins of leukaemias and hence the genetic composition and its
210 variation in the stem or leukaemia propagating cell compartment of these AML. The clonal
211 read-outs in the transplants were diverse but some patterns emerged.
212 In eight cases (patients #1, #2, #3, #4, #5, #6, #8 at diagnosis, #10 at relapse) two to
213 five sub-clones present in the diagnostic sample regenerated in the mice. However in each
214 case, one sub-clone was dominant, proportionally and this sub-clone always contained
215 *NPM1c*.
216 In one patient (#7), only one sub-clone was present at low levels (0.39% $CD34^+$ cells)
217 in a single mouse and, surprisingly, this corresponded to the most ancestral sub-clone in the
218 diagnostic sample which had *DNMT3A* as its sole identifier mutation. These are most likely
219 pre-leukaemic cells. In patient #8, two sub-clones read-out in mice from the diagnostic
220 sample. The dominant or largest sub-clone in all three mice harboured not only a *NPM1c* but
221 also a *TET2* mutation; this clone was below the detection limit in the diagnostic sample itself
222 (indicated by dotted circle in Fig 1). The relapse sample from patient #8 contained only one
223 *NPM1c* sub-clone corresponding to the major sub-clone seen at diagnosis. However, in the
224 transplant of this sample, a *NPM1c*-negative sub-clone, ancestral to the relapse sub-clone,

225 represented 100% of the regenerated leukaemia.

226 Finally, in patient #9, there was only one clone discernible both at diagnosis and 227 relapse and this clone read-out consistently in transplants of diagnostic and relapse samples.

228

229 **DISCUSSION**

230 These single cell data provide definitive identification of clonal architectures and preferential 231 order of mutations, furthering endorse the concept of sub-clonal complexity in myeloid 232 leukaemia [7, 14-16]. However, the current limits of single cell screening means that we will 233 have under-estimated the extent of sub-clonal genetic diversity that can be revealed by ultra-234 deep sequencing [25] and by new technologies that allow interrogation of thousands of cells [26]. 235 This has implications for clonal architecture and phylogeny. For example, in diagnostic 236 samples from several patients (#2, #3, #4, #8, #9), the sub-clone with the most simple genetic 237 composition at the base of the phylogenetic tree harboured more than one mutation. The 238 phylogenetic structure is therefore likely to have missed earlier, sequential (pre-leukaemic) 239 clones [8].

240 Different driver mutations have epistatic or synergistic functional impacts in AML [7, 241 22, 27] and the order of mutation accrual may impact on stem/progenitor cell function and 242 clinical features [28]. Our data provides direct evidence that *NPM1c* mutation is a sub-clonal 243 and therefore secondary mutation rather than a truncal or initiating lesion, as previously 244 suggested [29]. This concurs with the observations of Shlush *et al* [9] who found (in ten patients 245 with AML) that *DNMT3A* mutations in AML were present in differentiation competent 246 haemopoietic stem cells and putative pre-leukaemic clones. *NPM1* mutations, in contrast, 247 were absent from such cells but present in blasts cells with a myeloid progenitor cell 248 phenotype presumed to be descended from the *DNMT3A* mutant clones. Similarly, Corces-249 Zimmerman *et al* [30] found that *NPM1c* mutations were absent in purified haemopoietic stem

250 cells, in contrast to putative founder mutations including *DNMT3A, IDH1, IDH2* and *ASXL1*.
251 In cases of AML analysed at the single cell level, Jan *et al* [8] (one case) and Quek *et al* [16]
252 (three cases) documented that *NPM1c* was sub-clonal or secondary to *TET2* mutations.
253 However, Quek *et al* [16] also identified, in two cases, very rare $CD34^+$ cells that had *NPM1*
254 mutations but not other mutations found in the bulk leukaemic cells raising the possibility
255 that *NPM1* might occasionally be a founder mutation in pre-leukaemic cells. The
256 preservation of diagnostic *DNMT3A* but not *NPM1c* mutations in remission [9, 31] and in a
257 small minority of relapses is also commensurate with the predominantly secondary, sub-
258 clonal nature of *NPM1c* [32, 33]. As is the presence of *DNMT3A* and *TET2* but not *NPM1c*
259 mutations in covert pre-malignant clones of normal, ageing adults [34].

260 A preferential order of mutation may reflect genetic network or cell context
261 dependencies. *NPM1c* (and *FLT3* mutations) might be potent drivers only when arising in
262 myeloid progenitor cells with enhanced self-renewal provided by mutations in epigenetic
263 mutations such as *DNM3TA* or *TET2*.

264 In the bulk blast cell population, *DNMT3A* and *NPM1c* mutations were present at
265 similar high allele burden suggesting these were concurrent in the same cells [9]. In another
266 study however, *NPM1c* allele burden was consistently less than that of other drivers including
267 *DNMT3A* commensurate with a sub-clonal origin [19]. In our series, the allele burden for
268 *NPM1c* was consistently less than that of other putative founder mutations including
269 *DNMT3A*, *TET2* and *IDH2* (Fig 1). The existence of clones ancestral to those with *NPM1c*
270 mutations was clearly evident (in 8/10 cases) in the minor population sub-fractionated as
271 $CD34^+/CD33^-$. This again accords with the data of Shlush *et al* [9].

272 Mouse models with transgene or knock-in *NPM1c* have been developed to assess the
273 role of *NPM1* in leukaemogenesis [35]. By itself *NPM1* expressed in haemopoietic stem cells
274 produces a myeloproliferative disorder and a low penetrance of late occurring AML. A high

275 frequency of AML does develop in *NPM1c* mice subjected to insertional mutagenesis [36] or in
276 compound mutant mice with both *NPM1c* and *FLT3-ITD* [37, 38]. These modelling data testify
277 to the functional impact of *NPM1c* on myeloid cells and leukaemogenesis but underscore that
278 it is, at best, a weak initiating or founder lesion for AML.

279 The order of mutations and their position in the phylogenic tree is relevant to the
280 selection of mutated gene for targeted therapy [4]. In the cases of *NPM1c*$^+$ AML, the
281 phylogenetic studies highlight *DNMT3A* and *TET2* as truncal mutations as reported
282 previously [8, 9]. Effective therapeutic targeting of either *NPM1c* or *FLT3* mutations might be
283 expected to debulk the leukaemia but with only transient benefit. However, persistence or
284 increase of MRD in AML via detection of *NPM1c* transcripts is strongly predictive of relapse
285 [39] and in the great majority (>95%) of cases of *NPM1c*$^+$ AML that relapse, the relapsing
286 clone is *NPM1c*$^+$ [40]. In contrast, persistence of founder mutations (*DNMT3A, TET2, ASXL1*)
287 or pre-leukaemic clones, is not predictive of relapse [41] This reflects the strong driver status of
288 *NPM1c* mutations and the malignant potential of *NPM1c* sub-clones which is likely
289 contingent upon the genetic background of founder (truncal) mutations (i.e. by epistasis) and
290 additional co-existing sub-clonal mutations (e.g. in *FLT3*). Effective targeting of *NPM1c*
291 could, therefore, be very beneficial in restraining progression of disease.

292 There was evidence for reiterated driver mutations in sub-clones of several cases in
293 this study. This has been described before in ALL [3] and other cancers [42]. Mutations that are
294 highly recurrent between patients with a sub-type of leukaemia (or any cancer) might be
295 expected to occur more than once within a leukaemia from single patients. Functionally, this
296 could reflect either the fitness advantage of bi-allelic mutations of the same gene in the same
297 cells or convergent evolution of sub-clones contingent upon prevalent selective pressures or
298 preferential, epistatic partnership with earlier, common mutations [43].

299 A comparison of clonal structures in three cases of matched diagnosis and relapse

300 samples (#8, #9, #10) allowed us to infer the possible sub-clonal origins of the relapses. In
301 one patient (#9), there was only one clone detectable at diagnosis and that same clone was the
302 only clone observed at relapse. In case #8, the single relapse detected corresponded to one of
303 two clones present at diagnosis. However, sequencing also revealed a *TET2* mutation at low
304 allele burden (1.03%) at relapse. The allele burden for this mutation at diagnosis was
305 undetectable. However, when the diagnostic sample was transplanted into mice, a sub-clone
306 with that 'relapse' *TET2* mutation was the dominant clone (refer to Figure 1 for case #8).
307 In patient #10, there were three sub-clones at diagnosis and all three were present in
308 the relapse sample. These data raise the possibility that relapse in AML is not necessarily
309 more clonal and this should be further explored as it has important implications for the basis
310 of drug resistance.
311 Xeno-transplant read-outs depend upon the genetic background of the immuno-
312 deficient mice [15] and may not faithfully reflect the true diversity of propagating cells in AML.
313 Furthermore, we made no attempt to titrate leukaemia propagating activity by varying the
314 number of cells transplanted or by serial transplantation (except in patient #1). We note
315 however that replicate mice provide very similar read-outs which suggest intrinsic, functional
316 properties of AML sub-clones are being registered. The only conclusion we wish to draw
317 from these limited transplant experiments is that multiple sub-clones from individual patients
318 transplant indicating, as we showed previously for ALL [3] and glioblastoma [44], that individual
319 leukaemia's contain several, genetically distinct cells with self-renewing or leukaemia
320 propagating activity. These cells will provide a diverse pool of cells distributed throughout
321 the phylogenetic tree and from which relapse or drug resistance can emerge as recently
322 demonstrated by Shlush *et al* [24]. As such they function as cellular units of evolutionary
323 selection [45, 46]. However, sub-clones have variable repopulating capacity [47] and as previously
324 reported in AML [15], one *NPM1c* sub-clone dominated leukaemia regeneration in mice. This

325 may reflect the increased malignant potential of this sub-clone and the contribution of
326 *NPM1c*$^{+}$ cells to relapse in >95% of cases [40]. In all our six cases where the diagnostic clone
327 had both *NPM1c* and *FLT3* ITD or *FLT3* mutations, the dominant sub-clone in transplant
328 readouts had both mutations. Competitiveness of sub-clones with this genotype in a
329 xenotransplant context might be relevant to the very poor prognosis of AML cases that
330 harbour a combination of mutants in *DNMT3A*, *NPM1c* and *FLT3* [7].
331

332 **ACKNOWLEDGEMENTS**
333 This work was supported by Bloodwise, the Kay Kendall Leukaemia Fund and Gabrielle's
334 Angel Foundation UK.
335

336 **CONFLICT OF INTEREST**
337 There are no conflicts of interest to declare.
338

339 **AUTHOR CONTRIBUTIONS**
340 NP completed all laboratory work and analysis except initial patient sample sequencing and
341 animal experiments/care/transplants, assisted with manuscript writing and editing. FM-M
342 completed all animal experiments/care/transplants. IT assisted with FACS experiments. LE
343 completed all phylogenetic analysis. GV assisted with FACS experiments. EP completed the
344 targeted sequencing and analysis of each patient sample with support of PC. JG provided
345 access to patient samples. DT co-designed experiments, managed animal experiments and
346 contributed to writing of the manuscript. MG co-designed experiments, supervised primary
347 work and co-wrote the manuscript.
348
349

350 Supplementary information accompanies this paper on the Leukemia website
351 (http://www.nature.com/leu).
352
353
354

355 **REFERENCES**

356 1. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; **481:** 306-313.

357 2. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976; **194:** 23-28.

358 3. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM*, et al.* Genetic
359 variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011; **469:** 356-
360 361.

361 4. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in
362 cancer evolution. *Cancer Cell* 2015 Jan 12; **27**(1)**:** 15-26.

363 5. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012; **13:** 795-806.

364 6. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.*
365 Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*
366 2013 May 30; **368**(22)**:** 2059-2074.

367 7. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.*
368 Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016 Jun
369 9; **374**(23)**:** 2209-2221.

370 8. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR*, et al.* Clonal
371 evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia.
372 *Sci Transl Med* 2012 Aug 29; **4**(149)**:** 149ra118.

373 9. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V*, et al.* Identification of
374 pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 2014 Feb 20; **506**(7488)**:**
375 328-333.

376 10. Miller CA, Wilson RK, Ley TJ. Genomic landscapes and clonality of de novo AML. *N Engl J*
377 *Med* 2013 Oct 10; **369**(15)**:** 1472-1473.

378 11. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L*, et al.* Spatial and
379 temporal diversity in genomic instability processes defines lung cancer evolution. *Science*
380 2014 Oct 10; **346**(6206)**:** 251-256.

381 12. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y*, et al.* Mutational landscape and
382 clonal architecture in grade II and III gliomas. *Nat Genet* 2015 May; **47**(5)**:** 458-468.
383 13. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X*, et al.* Clonal evolution in breast
384 cancer revealed by single nucleus genome sequencing. *Nature* 2014 Aug 14; **512**(7513)**:** 155-
385 160.
386 14. Paguirigan AL, Smith J, Meshinchi S, Carroll M, Maley C, Radich JP. Single-cell genotyping
387 demonstrates complex clonal diversity in acute myeloid leukemia. *Sci Transl Med* 2015 Apr
388 1; **7**(281)**:** 281re282.
389 15. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'Laughlin M*, et al.* Functional
390 heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* 2014
391 Mar 17; **25**(3)**:** 379-392.
392 16. Quek L, Otto GW, Garnett C, Lhermitte L, Karamitros D, Stoilova B*, et al.* Genetically
393 distinct leukemic stem cells in human CD34- acute myeloid leukemia are arrested at a
394 hemopoietic precursor-like stage. *J Exp Med* 2016 Jul 25; **213**(8)**:** 1513-1535.
395 17. Potter NE, Ermini L, Papaemmanuil E, Cazzaniga G, Vijayaraghavan G, Titley I*, et al.*
396 Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Res* 2013 Dec;
397 **23**(12)**:** 2115-2125.
398 18. Taussig DC, Vargaftig J, Miraki-Moud F, Griessinger E, Sharrock K, Luke T*, et al.*
399 Leukemia-initiating cells from some acute myeloid leukemia patients with mutated
400 nucleophosmin reside in the CD34(-) fraction. *Blood* 2010 Mar 11; **115**(10)**:** 1976-1984.
401 19. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J*, et al.* Prognostic
402 relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012 Mar
403 22; **366**(12)**:** 1079-1089.
404 20. Peterlin P, Renneville A, Ben Abdelali R, Nibourel O, Thomas X, Pautas C*, et al.* Impact of
405 additional genetic alterations on the outcome of patients with NPM1-mutated cytogenetically
406 normal acute myeloid leukemia. *Haematologica* 2015 May; **100**(5)**:** e196-199.
407 21. Rose D, Haferlach T, Schnittger S, Perglerova K, Kern W, Haferlach C. Subtype-specific
408 patterns of molecular mutations in acute myeloid leukemia. *Leukemia* 2017 Jan; **31**(1)**:** 11-17.

409 22. Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ*, et al.* DNMT3A and TET2 compete and
410 cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat*
411 *Genet* 2016 Sep; **48**(9)**:** 1014-1023.
412 23. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS*, et al.* Clonal evolution in
413 relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012 Jan
414 26; **481**(7382)**:** 506-510.
415 24. Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A*, et al.* Tracing the
416 origins of relapse in acute myeloid leukaemia to stem cells. *Nature* 2017 Jul 6; **547**(7661)**:**
417 104-108.
418 25. Griffith M, Miller CA, Griffith OL, Krysiak K, Skidmore ZL, Ramu A*, et al.* Optimizing
419 cancer genome sequencing and analysis. *Cell Syst* 2015 Sep 23; **1**(3)**:** 210-223.
420 26. Pellegrino M, Sciambi A, Treusch S, Durruthy-Durruthy R, Gokhale K, Jacob J*, et al.* High-
421 throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet
422 microfluidics. *Genome Res* 2018 Aug 7.
423 27. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC*, et al.* The origin and
424 evolution of mutations in acute myeloid leukemia. *Cell* 2012; **150:** 264-278.
425 28. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J*, et al.* Effect of mutation
426 order on myeloproliferative neoplasms. *N Engl J Med* 2015 Feb 12; **372**(7)**:** 601-612.
427 29. Martelli MP, Pettirossi V, Thiede C, Bonifacio E, Mezzasoma F, Cecchini D*, et al.* CD34+
428 cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and
429 generate leukemia in immunocompromised mice. *Blood* 2010 Nov 11; **116**(19)**:** 3907-3922.
430 30. Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic
431 mutations in human acute myeloid leukemia affect epigenetic regulators and persist in
432 remission. *Proc Natl Acad Sci, USA* 2014 Feb 18; **111**(7)**:** 2548-2553.
433 31. Parkin B, Londono-Joshi A, Kang Q, Tewari M, Rhim AD, Malek SN. Ultrasensitive
434 mutation detection identifies rare residual cells causing acute myelogenous leukemia relapse.
435 *J Clin Invest* 2017 Sep 1; **127**(9)**:** 3484-3495.

436 32. Kim T, Yoshida K, Kim YK, Tyndel MS, Park HJ, Choi SH, *et al.* Clonal dynamics in a
437 single AML case tracked for 9 years reveals the complexity of leukemia progression.
438 *Leukemia* 2016 Feb; **30**(2)**:** 295-302.
439 33. Kronke J, Bullinger L, Teleanu V, Tschurtz F, Gaidzik VI, Kuhn MW, *et al.* Clonal evolution
440 in relapsed NPM1-mutated acute myeloid leukemia. *Blood* 2013 Jul 4; **122**(1)**:** 100-108.
441 34. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, *et al.* Age-related
442 clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014 Dec 25; **371**(26)**:**
443 2488-2498.
444 35. Sportoletti P, Varasano E, Rossi R, Mupo A, Tiacci E, Vassiliou G, *et al.* Mouse models of
445 NPM1-mutated acute myeloid leukemia: biological and clinical implications. *Leukemia* 2015
446 Feb; **29**(2)**:** 269-278.
447 36. Vassiliou GS, Cooper JL, Rad R, Li J, Rice S, Uren A, *et al.* Mutant nucleophosmin and
448 cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet* 2011
449 May; **43**(5)**:** 470-475.
450 37. Mallardo M, Caronno A, Pruneri G, Raviele PR, Viale A, Pelicci PG, *et al.* NPMc+ and
451 FLT3_ITD mutations cooperate in inducing acute leukaemia in a novel mouse model.
452 *Leukemia* 2013 Nov; **27**(11)**:** 2248-2251.
453 38. Mupo A, Celani L, Dovey O, Cooper JL, Grove C, Rad R, *et al.* A powerful molecular
454 synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in
455 mice. *Leukemia* 2013 Sep; **27**(9)**:** 1917-1920.
456 39. Shayegi N, Kramer M, Bornhauser M, Schaich M, Schetelig J, Platzbecker U, *et al.* The level
457 of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and
458 survival in AML. *Blood* 2013 Jul 4; **122**(1)**:** 83-92.
459 40. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A, *et al.* Assessment of
460 Minimal Residual Disease in Standard-Risk AML. *N Engl J Med* 2016 Feb 4; **374**(5)**:** 422-
461 433.

462 41. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A*, et al.*
463 Molecular minimal residual disease in acute myeloid leukemia. *N Engl J Med* 2018 Mar 29;
464 **378**(13)**:** 1189-1199.
465 42. Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer:
466 evolution within a lifetime. *Annu Rev Genet* 2014; **48:** 215-236.
467 43. Greaves M. Evolutionary determinants of cancer. *Cancer Discovery* 2015 Aug; **5**(8)**:** 806-
468 820.
469 44. Piccirillo SGM, Colman S, Potter NE, van Delft FW, Lillis S, Carnicer M-J*, et al.* Genetic
470 and functional diversity of propagating cells in glioblastoma. *Stem Cell Reports* 2015; **4:** 7-
471 15.
472 45. Greaves M. Cancer stem cells as 'units of selection'. *Evol Appl* 2013; **6**(1)**:** 102-108.
473 46. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell* 2014 Mar 6; **14**(3)**:**
474 275-291.
475 47. Wang K, Sanchez-Martin M, Wang X, Knapp KM, Koche R, Vu L*, et al.* Patient-derived
476 xenotransplants can recapitulate the genetic driver landscape of acute leukemias. *Leukemia*
477 2017 Jan; **31**(1)**:** 151-158.
478

479 **FIGURE LEGENDS**

480 **Figure 1.** Clonal phylogenies, inferred by maximum parsimony, and sub-clone genotypes in
481 10 patients.

482 Genetically distinct sub-clone percentages (as a fraction of the total population) are indicated
483 next to each clone; e.g. patient 1, most primitive sub-clone, $CD34^+/CD33^-$ first and
484 $CD33^+/CD34^-/CD3^-$ second percentages indicated as 18%/5% respectively. This indicates
485 that this sub-clone was found in 18% of the total $CD34^+/CD33^-$ cells investigated and 5% of
486 the total $CD33^+/CD34^-/CD3^-$ cells investigated (for the relapse samples of patients 9 and 10
487 only bulk cells without phenotype consideration could be sorted, as the samples available
488 were from fixed cytogenetic preparations; the sub-clone is shown as a single percentage).
489 Those sub-clones that grew in mice are indicated with horizontal black arrows. t1-3 (%). T,
490 transplant. 1-3 individual mice. % fraction of human cells in mouse bone marrow. Sub-
491 clone denoted by dotted circle is below detection limit in diagnostic sample but present in
492 mouse transplant read-out. Dotted arrows lines between sub-clones (case #1 and #3)
493 indicates alternative clonal phylogenies. In case #3, there are 4 possible equally
494 parsimonious phylogenetic trees (details in Supplementary Information Figures 5 and 6).
495 Further details on each of the individual 10 patients' clonal analyses are given in
496 Supplementary Information.

497