

1 **Rapid detection of myeloid neoplasm fusions using Single Molecule Long-Read Sequencing**

2 Olga Sala-Torra MD^{1,3}, Shishir Reddy³, Ling-Hong Hung³, Lan Beppu¹, David Wu², Jerald Radich,
3 MD^{1,2}, Ka Yee Yeung³, Cecilia CS Yeung, MD^{1,2}

4

5 ¹Fred Hutchinson Cancer Research Center, Seattle, WA

6 ²University of Washington, Seattle, WA

7 ³School of Engineering and Technology, University of Washington Tacoma, Tacoma, WA

8

9

10 Keywords (6): Long-read sequencing, CRISPR-Cas9, myeloid neoplasm, nanopore, CML,
11 sequencing

12

13 Short Title: **Long-Read Sequencing for Myeloid Neoplasm Fusions**

14

15 Address correspondence to:

16 Cecilia Yeung, MD

17 1100 Fairview Ave N, G7-910

18 Seattle, WA 98109

19 cyeung@fredhutch.org

20

21 **Key points:**

22 1) We describe a CRISPR-Cas9 enrichment Nanopore sequencing assay with streamlined
23 bioinformatics that outperforms other fusion detectors.

24 2) We successfully detected both fusion genes and specific breakpoints in CML, APL, and
25 AML in under 8 hours in 80% of patients.

26

27 **Visual Abstract** (Figure 1):

28 1) We successfully detected fusion genes in hematological malignancies with a fast and
29 efficient long-read sequencing workflow in under 8 hours. The method makes the
30 genomic characterization of *BCR-ABL1* DNA breakpoint in patients quick and simple,
31 which potentiates design of patient specific primers for personalized monitoring MRD
32 assays.

33 2) Our assay is based on a CRISPR-Cas9 non-amplification enrichment library preparation
34 strategy and uses Nanopore sequencing single stranded genomic DNA coupled with
35 streamlined bioinformatic workflow containing a novel fusion detector software which
36 outperforms current fusion detection software.

37

38 **Abstract**

39 Recurrent gene fusions are common drivers of disease pathophysiology in leukemias.
40 Identification of these structural variants helps stratify disease by risk and assists with therapy
41 choice. Current fusion detection methods require long turnaround time (7-10 days) or advance
42 knowledge of the genes involved in the fusions. To address the need for rapid identification of
43 clinically actionable fusion genes in heme malignancies without *a-priori* knowledge of the

44 genes, we describe a long-read sequencing DNA assay designed with CRISPR guides to select
45 and enrich for recurrent leukemia fusion genes. By applying rapid sequencing technology based
46 on nanopores, we sequenced long pieces of genomic DNA and successfully detected fusion
47 genes in cell lines and primary specimens (e.g., *BCR-ABL1*, *PML-RARA*, *CBFB-MYH11*, *KMT2A-*
48 *AF4*) using cloud-based bioinformatics workflows with novel custom fusion finder software. We
49 detected fusion genes in 100% of cell lines with the expected breakpoints and confirmed the
50 presence or absence of a recurrent fusion gene in 12 of 14 patient cases. With our optimized
51 assay and cloud-based bioinformatics workflow, these assays and analyses could be performed
52 in under 8 hours.

53

54 **Introduction**

55 Current classification of myeloid malignancies is largely based on the molecular and genetic
56 aberrations^{1,2}. Recurrent gene rearrangements are present in 30-40% of acute myeloid
57 leukemias (AML), and well described driver fusions that in some cases suffice to diagnose
58 leukemias with *PML-RARA*, *RUNX1-RUNX1T1*, etc. even when the blast percentage is below
59 20%¹. Recurrent fusion genes confer certain clinical and biological characteristics as drivers of
60 leukemogenesis, and their identification assists in prognosis stratification and inform treatment
61 decisions. Identification of driver fusion genes is especially relevant when targeted therapies
62 are available. Examples include tyrosine kinase inhibitors (TKI) in chronic myelogenous leukemia
63 (CML)³ that bind to the kinase domain in *ABL1* deregulated as a consequence of the fusion with
64 *BCR*, and differentiation therapy in acute promyelocytic leukemia (APL) in which identification
65 of the promyelocytic leukemia- retinoic acid receptor alpha (*PML-RARA*) fusion confirms the

66 diagnosis and indicates that the patient will likely respond to treatment with all-trans retinoic
67 acid (ATRA) therapy, a nontoxic and highly effective treatment⁴ that achieves 90% long term
68 survival rates when combined with anthracyclines and/or arsenic trioxide.⁵

69
70 Fusion genes can be detected through several techniques that include cytogenetics techniques
71 such as fluorescence *in situ* hybridization (FISH), and molecular testing such as polymerase
72 chain reaction (PCR) and next generation sequencing (NGS), which are the clinical gold
73 standard. Clinical molecular assays for *BCR-ABL1* and *PML-RARA* primarily target RNA
74 transcripts by RT-PCR because of the greater abundance of fusion gene transcripts compared to
75 the DNA copies per cell, and because of the large variability in the genomic sequence of the
76 fusion that can encompass large intronic region making routine DNA PCR impossible (e.g., *BCR-*
77 *ABL*).

78
79 Typical NGS reads are 150 to 250 bps in length which are not sufficiently long enough to extend
80 over DNA fragments that are adjoined because of a large genomic aberration thus hindering
81 both alignment and detection of large indels and other structural abnormalities. NGS is limited
82 by short read assembly mis-mapping, and amplification strategies instill imperfect quantitation
83 of the variant allele frequencies. The diversity of aberrations in myeloid neoplasms include
84 large genomic aberrations, including insertions and deletions, loss of heterozygosity, single
85 nucleotide polymorphism, mutations in homopolymer rich regions and highly repetitive regions
86 such as internal tandem duplications, that are difficult to detect by NGS and require different
87 assays in molecular and cytogenetics labs additionally to those used to confirm fusion genes.

88 Hence, currently a multiple assay approach is used to obtain a complete diagnostic molecular
89 picture in myeloid malignancies.

90

91 The advancement of long-read sequencing technologies has enabled the sequencing of
92 continuous single DNA or RNA molecules up to tens to hundreds of kilobases (kb) long.⁶

93 Ongoing improvements in Nanopore sequencing accuracy have reduced error rates to less than
94 5%,⁷ but remain higher than those for Illumina and Ion Torrent, which are used frequently in

95 clinical laboratories.⁸ This technology has already made an impact on the understanding of the
96 pathobiology of various diseases,^{9,10} and its impact will increase as the quality of sequencing

97 improves and becomes more accurate.^{7,11} Addition of CRISPR-Cas9 for targeted enrichment
98 concentrates the regions of interests prior to sequencing by Nanopore¹² without requiring

99 amplification steps, thus optimizing sequencing time and efficiency. Additionally, the portability
100 and affordability of the Nanopore sequencer MinION and Flongle, hold great promise to impact

101 the clinical field. However, a major limitation has been the lack of analytical software featuring
102 standardized parameters to aid in translation into clinical diagnostics.

103

104 Here we report our success in developing an amplification-free CRISPR-Cas9 targeted
105 enrichment sequencing protocol using Nanopore MinION and Flongles to detect fusions

106 relevant in the diagnosis and classification of CML and AMLs. The ONT Flongle is an adaptor for
107 the MinION that provides cost-effective (~\$90 per flow cell), real-time sequencing for smaller

108 assays. Our assays were designed to capture varied breakpoints of CML and APL, as well as
109 fusion genes resulting from *inv(16) (MYH11-CBFB)* and *t(4;11)(KMT2A-AF4)*. Simultaneous

110 interrogation of these targets is a first step to a rapid characterization of AMLs in a single assay
111 combining data that up to now required multiple different techniques and provide relevant
112 information promptly. In addition to this amplification-free CRISPR-Cas9 nanopore assay, we
113 extended our previously developed cloud-based nanopore data analysis pipeline¹³ to include
114 fusion detection and develop a custom breakpoint detection tool (see **Figure 1**). Using our
115 optimized assay and our custom breakpoint finder, we showed that we can reliably detect and
116 confirm fusion breakpoints in 80% of our specimens in under 3 hours of sequencing and data
117 analysis.

118

119 **Methods:**

120 **Cells lines and patient samples**

121 Our assay was optimized using 6 cell lines: three with the *BCR-ABL1* fusion (K562, KU812, and
122 KCL22), and NB4, MV4;11 and ME-1 that bear the *PML-RARA*, *KMT2A-AF4*, and *MYH11-CBFB*
123 fusions respectively. Residual mononuclear cells from primary specimens (6 specimens from 5
124 patients with CML, 6 specimens from 5 patients with suspected APL, and 2 acute myeloid
125 leukemia, not acute promyelocytic leukemia) were isolated using Ficoll® reagent (Millipore-
126 Sigma) and banked in liquid nitrogen until the time of the experiment. All specimens had been
127 originally tested in a CLIA certified laboratory according to standard clinical protocols.¹⁴ IRB
128 coverage was obtained for use of residual laboratory samples. Patient samples were de-
129 identified to the nanopore testing lab, and cytogenetic or molecular results were confirmed
130 after nanopore results were rendered. Characteristics and demographics of specimens and
131 patients are listed in **Table 1**.

132

133 **Library preparation and sequencing assay**

134 For the cell lines and 11/14 patient specimens the DNA was extracted with PureGene (Qiagen,
135 Germantown, MD, USA) following the standard protocol. Special caution, including use of wide
136 bore pipette tips and moderate centrifuge spin velocity was exercised to minimize fragmenting
137 DNA strands. Two DNA specimens were extracted with AllPrep DNA/RNA Kit (Qiagen,
138 Germantown, MD, USA), and one with QiAgen X-tractor with Reagent Pack DX (Qiagen,
139 Germantown, MD, USA). cRNA guides were designed to direct Cas9 to cut on genomic proximity
140 of each of the regions involved in each one of the translocations studied. When the target
141 region was big, guides were tiled across the region to maximize coverage. Guides were
142 designed to capture *PML-RARA*, *BCR-ABL1* p210, *KMT2A-AF4*, and *MYH11-CBFB*, including
143 different fusion isoforms. Guides were designed using Chopchop
144 (<https://chopchop.cbu.uib.no/>) with the CRISPR-Cas9 and nanopore enrichment settings and
145 previously described¹⁵.

146

147 We used 5 micrograms of DNA as input for each cell line and 2 to 5 micrograms for primary
148 specimens. Average DNA integrity number (DIN) was 9.2 (range: 7.5-9.8). **Figure 1** shows a
149 schematic of our workflow and details of the library prep are published¹⁵. Briefly, enrichment of
150 target regions was obtained using Oxford Nanopore Technologies “Targeted, amplification-free
151 DNA sequencing using CRISPR-Cas9” protocol¹². The different guides used in the assay were
152 pooled in equimolar amounts of each guide. Through an initial dephosphorylation step, the 5’
153 ends of the DNA becomes inaccessible to adapter ligation. Double stranded DNA breaks that

154 excise the region of interest are generated with the directional, target specific RNA guides
155 complexed with tracrRNA and Cas9 enzyme. The Cas9 complex remains bound to the 5' end of
156 the guide, and the resulting new DNA ends contain a phosphorylated 5' end that is available for
157 dA tailing and adapter ligation.¹² All libraries generated in this manner were run on a MinION
158 version 9.4. (Oxford Nanopore Technologies, Oxford, UK) nanopore sequencer using flow cells
159 or Flongles. Modifications for libraries sequenced on the Flongle were only at the library loading
160 step, in which the amount of Sequencing Buffer and library beads (both SQK-LSK 109, ONT) are
161 reduced to from 35 to 13 and 25.5 to 7.5UL respectively, and 0.5UL of SQT is added. QC
162 parameters tracked for each run are listed in **Table 2**.

163

164 **PCR/Sanger sequencing**

165 Primers specific to *BCR-ABL1* patient breakpoints were designed using Primer3 v. 0.4.0.¹⁶ in 2
166 cases. 100 ng of DNA were amplified, the PCR product was run on a 2% gel and Sanger
167 sequenced to confirm the genomic breakpoint.

168

169 **Biodepot-workflow-builder: Interactive and accessible front end for fusion detection**

170 We present a graphical, reproducible, and cloud-enabled fusion detection workflow consisting
171 of all the steps of the analyses, including base calling, alignment, fusion detection, and
172 visualization (see **Figure 1 screenshot of BwB interface**). In contrast to the NanoFG workflow
173 from Stangle et al¹⁷, our platform includes the computationally intensive base calling step, an
174 interactive graphical user interface, and can readily leverage GPUs and be deployed on the
175 cloud. Thus, analyses are fast, and our platform is accessible to biomedical and clinical

176 scientists. Specifically, we extended the Biodepot-workflow-builder (Bwb)¹⁸ platform in which
177 each computational task (or module) is represented by a graphical widget that calls a software
178 container in the back end. Software containers, such as Docker, include all software
179 dependencies and libraries required to execute the code. We have recently developed a Bwb
180 workflow¹³ to support the processing of Nanopore data that includes the use of base callers
181 Guppy¹⁹ and Bonito¹⁹, alignment using minimap2 and visualization of results using the
182 Integrative Genomics Viewer (IGV) and GRCh37 hg19 as reference genome QC data was
183 generated with PycoQC²⁰. Guppy was used as the base caller. Minimap2²¹ was used as the
184 aligner and variant caller. Fusions are visualized on IGV²² and confirmed on Blast. Most
185 importantly, we extend our previous work by adding support for fusion detection, including
186 LongGF²³ and our own custom software “Biodepot Fusion Finder” (BFF).

187 188 **Bioinformatics pipeline for fusion detection: LongGF**

189 LongGF²³ is a software tool for fusion detection optimized for the high base calling error rates
190 and alignment errors commonly found in long read sequencing data. LongGF takes as input a
191 BAM file containing alignments (generated by minimap2 in this pipeline) and a GTF file
192 containing the definitions of known genes. The output is a log file with detected gene fusions
193 and their supporting reads. We created a graphical widget for LongGF in the Bwb. **Figure 2**
194 shows a screenshot of the comparative workflows with output for LongGF versus BFF.

195 196 **Custom fusion detection tool: Biodepot Fusion Finder (BFF)**

197 Reads that span a fusion gene will align to coordinates in both parts of the fusion and provide a
198 specific breakpoint coordinate. We wrote a new software tool, the Biodepot Fusion

199 Finder, BFF that examines the alternate alignments for each read and identifies reads that map
200 to coordinates spanning a set of known breakpoints. This is similar to the strategy used
201 by LongGF except that we allow for errors in the alignment near the breakpoint. Accordingly,
202 the Breakpoint Finder identifies fusions that are not detected by LongGF which looks for
203 supporting reads where the breakpoints are exactly matched. The user can provide a panel of
204 breakpoint coordinates of interest. If no panel is provided, BFF will return candidate fusions
205 that span nonadjacent genomic regions. Incomplete coordinates for breakpoints in the panel
206 are supported – the user does not need to define the exact coordinates, nor do both
207 breakpoints need to be given. BFF will return all the reads that match the panel of breakpoints
208 in a text file for further inspection by the user if desired. The user can also provide a file with
209 guide coordinates to obtain additional enrichment metrics. A containerized widget was
210 developed that can be integrated with the Bwb workflows for processing nanopore data. Using
211 these workflows, we can directly analyze raw nanopore data and obtain lists of candidate
212 fusions.

213

214 **Benchmarking experiments**

215 We performed empirical experiments to benchmark the sensitivity and runtime required to
216 reliably detect fusion. For each individual sample, sequencing metrics including quality scores
217 and timestamps are obtained from the sequencing summary text file obtained as an output of
218 base calling using Guppy. Detected fusions are then acquired from the Breakpoint Finder as
219 well as LongGF. Specific fusion breakpoints are also provided in the data set from the BFF. All
220 samples are then combined into a dictionary and separated by patient and cell line data. Plots

221 are constructed for each category of data pertaining to the time to reach 3 fusions as well as
222 the number of reads required to reach 3 fusions. Finally, the total number of fusion reads
223 detected for each sample is compared between the BFF and LongGF as shown in **Figure 2**.

224

225 **Enrichment assessment**

226 Two enrichment metrics were computed by BFF and tracked for each sample. First is the *fusion-*
227 *specific enrichment* which is calculated with the following formula $[(\text{number of fusion reads}) /$
228 $(\text{mean coverage of the genome})]$. Second is the *on-target enrichment* which is calculated with
229 the following formula $[(\text{number of reads that originate from a guide RNA cut point that includes}$
230 $\text{the region of the breakpoint})/(\text{mean genome coverage})]$. Reads originating from a guide RNA
231 cut are distinguished by the guide sequence being at the beginning of a read. As initial electrical
232 signal data is generated by DNA passing through the nanopores (reads at the beginning of a
233 strand) are error prone, which affects base calling and therefore the alignment of the reads so
234 that the start sequence is often misaligned. Consequently, the BFF considers base pairs of the
235 sequence near the start (default within 50 bp) of a read that aligns near the coordinates of a
236 guide RNA cut site to have originated from a guide RNA cut. Specific reads cut by guides are
237 manually confirmed. The allowed error intervals are customizable. SamTools v1.13 is used to
238 sort and convert BAM files and determine the overall average coverage²⁴. Picard
239 CollectHsMetrics is used to generate the unique base pairs mapped metric for each sample²⁵.

240

241 **Results**

242 Sample sequencing and enrichment:

243 Details of the sample sequencing and enrichment metrics are included in Table 2. A range of
244 0.04 Gb – 5.47 gigabases of sequencing data was generated for each sample for an average
245 mean coverage of the human genome of 0.32-fold (range: 0.01 – 1.66). We adopted standard
246 quality metrics for nanopore workflows and tracked N50, which is a quality metric where half
247 the reads are above this length (range 4.65kb-32.2kb) and median read length for total reads
248 (range 0.87kb- 9.60kb). Percentage of reads aligned ranged from 72 - 96%. Fusion specific
249 enrichment was 135 - 837 fold for cell lines and 6 - 509 fold for patient samples. On target
250 enrichment was 849 - 5830 fold for cell lines and 535 - 3007 fold for patient samples.

251

252 Concordance of fusion detection with clinical results:

253 Concordance with expected results was 100% for cell lines as the expected fusion was detected
254 in 6/6 cell lines (3 *BCR-ABL1*, 1 *PML-RARA*, 1 *MYH11-CBFB*, 1 *KMT2A-AFF1*). Breakpoint
255 sequences detected for *BCR-ABL1* cell lines are the same as previously published ²⁶. We
256 correctly confirmed the presence or absence of fusions in 11/14 (78.5%) primary specimens
257 including both diagnostic and measurable residual disease (MRD) cases with a minimum of
258 three reads, however one case (APL6) showed only 2 fusion reads and was not counted as
259 confirmed. The 3 missed cases (CML3, APL1, and APL6 in tables 1 and 2) were comprised of low
260 disease burden, ~0%, 1% and <5%. We detected *BCR-ABL1* in 5/6 cases, *PML-RARA* in 2/6 cases,
261 *KMT2A-AF9* and *CBFB-MYH11* in 1/1 case each. *BCR-ABL1* genomic breakpoint was confirmed
262 by Sanger sequencing in CML1 and CML2 after designing patient specific primers.

263

264 In 3 specimens from 2 patients with suspected APL, we could not detect *PML-RARA* but
265 observed other findings. Clinical and laboratory details are listed in **Table 2**. For the first patient,
266 two specimens, one bone marrow and one peripheral blood (APL4 and APL 5) yielded no *PML-*
267 *RARA* fusion reads. This patient presented with an AML morphologically suggestive of APL and
268 an isochromosome 17q without t(15;17) detected by karyotype. While fusion detection
269 software did not detect a fusion, manual inspection showed an insertion in an intronic region of
270 *RARA* with TTMV viral genome; this case was previously reported¹⁵. Patient APL6 (presented
271 with 22% blasts on flow in a <5% marrow which on unblinding showed a complex karyotype
272 with t(11;17)(q23;q25) including the *KMT2A* gene. Two reads with *KMT2A-SEPT9* fusion were
273 detected in a suboptimal but acceptable run (N50 < 5000bp; on-target enrichment 650.66 fold),
274 confirming the lack of t(15;17) or *PML-RARA* fusion but the threshold was below the requisite 3
275 reads to confirm the *KMT2A-SEPT9* fusion.

276

277 Comparison of fusion detection tools:

278 A comparison of the bioinformatic workflows for data analysis using different fusion detection
279 widgets including LongGF vs Biodepot Fusion Finder (BFF) was conducted; the specific workflow
280 is demonstrated in **Figure 2**. Additionally, BFF computes the fusion enrichment and on target
281 enrichment statistics; these are summarized in **Table 2**. In most cell line and primary specimens,
282 LongGF shows a particular challenge in the detection of BCR-ABL1 and does not detect all fusion
283 reads that are identified by BFF. Using the BFF, the average sequencing, and data processing
284 time to 3 reads with fusions in the cell line experiments was 42.75 minutes (range: 18.87 - 77.65
285 min) and 188 minutes in the primary specimens where 3 reads were detected (range: 32 - 654

286 min) [see **Figure 3**, and **Supplemental Table 1**]. Cell line experiments took an average of 11,711
287 reads to identify 3 fusion reads (range: 1,321 – 43,326 reads) and 10,273 reads in primary
288 specimens (range: 1,790 - 24,999 reads) for confirmation of the fusion calls.

289

290 Comparison of Flongles and flow cells:

291 In five of our experiments (2 cell lines, 3 primary specimens), we used Flongles, while in the rest
292 we used Flow Cells. The performance of the affordable Flongles was inferior to the Flow Cells,
293 with lower expected average data output from the Flongles (based on manufacture
294 expectations of ~3GB), the N50 and median read length were smaller in Flongle reads (21,614
295 vs 19,166 and 6058 vs 5250 respectively), and significantly, the median Phred score for Flongle
296 reads being lower than that for the Flow Cells (9.2 vs 11.88). Despite the worse performance of
297 the Flongles, fusions were detected in 2 of 2 cell lines, and 2 of the 3 experiments with primary
298 specimens with at least 8 and 14 reads confirming the fusion and the specimen without fusion
299 confirmation was CML3 with pancytopenia and low disease burden.

300

301 **Discussion**

302 We developed a rapid assay to detect fusion genes in blood or marrow samples in less than 8
303 hours with the fastest time achieving 3 fusion reads was 5 hours. This was accomplished by
304 combining CRISPR-Cas9 enrichment during library preparation, nanopore long-read sequencing,
305 a cloud-based data analytic pipeline, with a novel BFF program developed and optimized for
306 finding fusions reads and breakpoints. RNA guides were designed to target genes involved in
307 recurrent fusions in myeloid malignancies and used to enrich an amplification-free library

308 preparation over 1600-fold. Our modular and containerized pipeline in Bwb allows users to
309 efficiently process raw FAST5 data on the cloud through an accessible graphical user interface
310 allowing for a very fast analysis step (average ~4.5 minutes for basecalling, alignment, and
311 fusion detection). To improve fusion calling, we developed the custom BFF that allow users to
312 identify fusion reads not detected by LongGF. We successfully confirmed published genomic
313 breakpoints in our series of cell lines and archival patient samples, which includes both
314 diagnostic and follow-up samples to test feasibility in confirming both common and novel
315 fusions over a range of tumor burdens. Our study includes 14 patient specimens and
316 demonstrates the usability of this method in primary specimens with 2microgram of DNA.

317
318 An advantage of the CRISRP-Cas9 based enrichment protocol is that it allows for targeting of
319 multiple common leukemia fusion genes by pooling multiple guide RNAs. Fusions are
320 particularly well suited for this method as they have large gene segments that aid in alignment
321 despite sequencing errors and wide variation where translocation breakpoints may occur.
322 Other labs have employed similar but different methods that detected fusions by targeting one
323 partner gene in the fusion¹⁷. In contrast, our assay targets both partners of the fusion thus our
324 approach allows for an expanded capability to detect known and novel fusions, such as in case
325 AML1 with t(9;11), when guides are designed to target t(4;11). However, in 2/14 patients the
326 assay did not detect fusion genes because both cases had a very low amount of fusion target. In
327 one case (CML3) because the BCR-ABL1 was <0.01%, the other in the context of a very
328 hypocellular sample.

329

330 Our work differs from standard RNA-based fusion detection assays, and instead interrogated
331 single molecules of DNA rapidly and accurately to detect specific translocation breakpoints.
332 Long-read sequencing technologies, like nanopore (Oxford Nanopore Technologies, Oxford,
333 U.K.), allow sequencing of unamplified, long unbroken fragments of DNA which are more likely
334 to span a breakpoint. This has potential clinical utility for personalized disease monitoring
335 when CML patients are on TKI therapy and suppressing RNA transcription²⁷; targeting DNA as a
336 monitoring target may be more robust and reproducible since DNA is stable and present in
337 constant numbers²⁸. However genomic breakpoints in *BCR-ABL1* are unique to individual
338 patients requiring patient-specific breakpoint characterization²⁸⁻³² as *ABL1* breakpoints occur
339 over an expansive region of about 150Kb, making this is an arduous endeavor previously
340 involving multiple primer sets and Sanger sequencing³³. Our method allows a single approach
341 spanning the *BCR* and *ABL1* breakpoint regions without the use of multiple primers and PCR
342 reactions. The sensitivity of DNA-based qPCR once the breakpoint is known can be as low as 10⁻⁷
343 ³⁴⁻³⁵. Specimens CML1 and CML5 are a BM and PB obtained from the same patient and
344 demonstrate high fidelity in confirming genomic breakpoints and the ability to use patient
345 specific primers for personalized MRD monitoring.

346
347 Advantages of long-read sequencing over current clinical diagnostic assays are speed and the
348 relatively low complexity of the assay when compared to cytogenetics and targeted NGS
349 panels. While long-read sequencing results could potentially have a turnaround time (TAT) of
350 less than 24 hours, full karyotype analysis TAT is generally longer with the fastest times at days
351 to a week and most targeted NGS panels require ~7-10 days from start of processing to result

352 report. The nanopore sequencing streams data simultaneously to our GPU-enabled data
353 analytic pipeline in the Bwb interface which resides on the cloud to help interpret and reliably
354 identify fusion reads within 5000 seconds(<2 hours) computational time in most specimens.
355 Building on our experience^{14,17}, we used 3 reads as a threshold for fusion confirmations. With
356 current simultaneous sequencing and data analysis workflow described here, 3 sequences are
357 detected in an average of 3 hours and 7 min (fastest at 30 min) in the 9 patient specimens
358 where a fusion could be confirmed. This means that a diagnostic result with a precise fusion
359 breakpoint with 3 fusion supporting reads would be possible in the same day.

360
361 Five specimens were sequenced on the less costly Flongle device, which has lower sequencing
362 capabilities (pore count ~60, sequencing life 24h), but 3 fusion reads were reached in 4/5
363 specimens (2 cell lines, 3 patient samples) with less resources. The sequencing quality is
364 significantly lower on the Flongles (median Phred 9.2 in Flongles vs 11.9 in Flow Cells), however
365 fusions were detected in all samples with adequate tumor burden. To achieve a cheaper
366 version of our fast and portable nanopore fusion assay, further challenges will need to be
367 addressed. We predict overall less sequencing can be achieved, even with additional
368 optimizations of the CRISPR library guides as multiplexing of the guide RNA appears to increase
369 efficiency and on-target fusion reads. Nanopores on the Flongles have poorer viability and
370 generate more read errors demonstrable by lower Flongle Phred quality on average and
371 therefore additional alignment challenges and a study of errors specific in Flongle bioinformatic
372 data compared to the flow cells are needed to understand potential compensation mechanisms
373 for data analysis.

374

375 **Conclusion**

376 We demonstrate the feasibility of using single molecule long-range sequencing assay to detect
377 fusion genes in heme malignancy (AML, CML and APL) patients. Inherent characteristics of
378 fusions make this assay a promising cost effective tool for rapid detection of recurrent fusions
379 that 1) does not require previous knowledge of the target, , 2) with a rapid TAT (8 hours in 80%
380 of samples) when multiplexing different assays and used with the specific data analysis and
381 fusion detection tools described in our manuscript, 3) can precisely map translocation genomic
382 breakpoints that allow for development of personalized markers for disease monitoring, and 4
383) can potentially allow discovery of novel/different fusion partners

384

385 **Contributions:** OST and CY designed the study and performed background research and
386 secured IRB protocol. KYY and LHH designed the informatics pipeline and algorithms. JR, DW
387 and CY identified specimens. LHH and SR implemented the software. SR, OST and CY tested the
388 software. OST, LB, SR, LHH, CY performed research and analyzed the data. KYY, JR, CY funded
389 the project and provided supervision. OST, SR, KYY, and CY drafted the manuscript. All authors
390 reviewed and edited the manuscript.

391

392 **Funding**

393 This study was supported in part by NCCN young investigator award and the Hyundai Hope on
394 Wheels Scholars Award, R01 CA175008-06, UG1 CA233338-02, and Adult Leukemia Research
395 Center Grant # P01 CA018029 as funded by the National Cancer Institute, National Institutes of

396 Health, Bethesda, MD. LHH, SR and KYY are supported by NIH grant R01GM126019. SR is also
397 supported by the Vicky L. Carwein and William B. Andrews Endowments for Graduate
398 Programs.

399
400 **Disclosures:** LHH and KYY also have equity interest in Biodepot LLC, which receives
401 compensation from NCI SBIR contract numbers 75N91020C00009 and 75N91021C00022.

402
403 **Acknowledgements:** The authors want to thank Dr Phillip E. Starshak, Kaiser Permanente
404 Oakland, for procuring specimens.

405
406 **References:**

- 407 1. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health
408 Organization classification of myeloid neoplasms and acute leukemia. *Blood*. May 19
409 2016;127(20):2391-405. doi:10.1182/blood-2016-03-643544
- 410 2. Quessada J, Cuccini W, Saultier P, Loosveld M, Harrison CJ, Lafage-Pochitaloff M.
411 Cytogenetics of Pediatric Acute Myeloid Leukemia: A Review of the Current Knowledge. *Genes*
412 (*Basel*). Jun 17 2021;12(6)doi:10.3390/genes12060924
- 413 3. Hehlmann R. Innovation in hematology. Perspectives: CML 2016. *Haematologica*. Jun
414 2016;101(6):657-9. doi:10.3324/haematol.2016.142877
- 415 4. Lo-Coco F, Cicconi L. History of acute promyelocytic leukemia: a tale of endless
416 revolution. *Mediterr J Hematol Infect Dis*. 2011;3(1):e2011067. doi:10.4084/MJHID.2011.067

- 417 5. Long ZJ, Hu Y, Li XD, et al. ATO/ATRA/anthracycline-chemotherapy sequential
418 consolidation achieves long-term efficacy in primary acute promyelocytic leukemia. *PLoS One*.
419 2014;9(8):e104610. doi:10.1371/journal.pone.0104610
- 420 6. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and
421 challenges in long-read sequencing data analysis. *Genome Biol*. Feb 7 2020;21(1):30.
422 doi:10.1186/s13059-020-1935-5
- 423 7. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome
424 with ultra-long reads. *Nat Biotechnol*. Apr 2018;36(4):338-345. doi:10.1038/nbt.4060
- 425 8. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of Next Generation Sequencing
426 Platforms. *Next Gener Seq Appl*. 2014;1doi:10.4172/jngsa.1000106
- 427 9. Jancuskova T, Plachy R, Zemankova L, et al. Molecular characterization of the rare
428 translocation t(3;10)(q26;q21) in an acute myeloid leukemia patient. *Mol Cytogenet*. 2014;7:47.
429 doi:10.1186/1755-8166-7-47
- 430 10. Jeck WR, Lee J, Robinson H, Le LP, lafrate AJ, Nardi V. A nanopore sequencing-based
431 assay for rapid detection of gene fusions. *J Mol Diagn*. Sep 28
432 2018;doi:10.1016/j.jmoldx.2018.08.003
- 433 11. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akesson M. Improved data analysis for the
434 MinION nanopore sequencer. *Nat Methods*. Apr 2015;12(4):351-6. doi:10.1038/nmeth.3290
- 435 12. Gilpatrick T, Lee I, Graham JE, et al. Targeted nanopore sequencing with Cas9-guided
436 adapter ligation. *Nat Biotechnol*. Apr 2020;38(4):433-438. doi:10.1038/s41587-020-0407-5

- 437 13. Reddy S, Hung LH, Sala-Torra O, Radich JP, Yeung CC, Yeung KY. A graphical, interactive
438 and GPU-enabled workflow to process long-read sequencing data. *BMC Genomics*. Aug 23
439 2021;22(1):626. doi:10.1186/s12864-021-07927-1
- 440 14. Yeung C, Qu X, Sala-Torra O, Woolston D, Radich J, Fang M. Mutational profiling in acute
441 lymphoblastic leukemia by RNA sequencing and chromosomal genomic array testing. *Cancer*
442 *Med*. Aug 2021;10(16):5629-5642. doi:10.1002/cam4.4101
- 443 15. Sala-Torra O, Beppu LW, Abukar FA, Radich JP, Yeung CC. TTMV-RARA fusion as a
444 recurrent cause of AML with APL characteristics. *Blood Adv*. Apr 14
445 2022;doi:10.1182/bloodadvances.2022007256
- 446 16. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. . Humana Press;
- 447 17. Stangl C, de Blank S, Renkens I, et al. Partner independent fusion gene detection by
448 multiplexed CRISPR-Cas9 enrichment and long read nanopore sequencing. *Nat Commun*. Jun 5
449 2020;11(1):2861. doi:10.1038/s41467-020-16641-7
- 450 18. Hung LH, Hu J, Meiss T, et al. Building Containerized Workflows Using the BioDepot-
451 Workflow-Builder. *Cell Syst*. Nov 27 2019;9(5):508-514 e3. doi:10.1016/j.cels.2019.08.007
- 452 19. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford
453 Nanopore sequencing. *Genome Biol*. Jun 24 2019;20(1):129. doi:10.1186/s13059-019-1727-y
- 454 20. Leger AL, T. pycoQC, interactive quality control for Oxford Nanopore Sequencing.
455 *Journal of Open Source Software*. 2019;4(34):1236.
- 456 21. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. Sep 15
457 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191

- 458 22. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat*
459 *Biotechnol.* Jan 2011;29(1):24-6. doi:10.1038/nbt.1754
- 460 23. Liu Q, Hu Y, Stucky A, Fang L, Zhong JF, Wang K. LongGF: computational algorithm and
461 software tool for fast and accurate detection of gene fusions by long-read transcriptome
462 sequencing. *BMC Genomics.* Dec 29 2020;21(Suppl 11):793. doi:10.1186/s12864-020-07207-4
- 463 24. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and
464 SAMtools. *Bioinformatics.* Aug 15 2009;25(16):2078-9. doi:10.1093/bioinformatics/btp352
- 465 25. Picard Toolkit. Broad Institute, GitHub repository.
- 466 26. Ross DM, Schafranek L, Hughes TP, Nicola M, Branford S, Score J. Genomic translocation
467 breakpoint sequences are conserved in BCR-ABL1 cell lines despite the presence of
468 amplification. *Cancer Genet Cytogenet.* Mar 2009;189(2):138-9.
469 doi:10.1016/j.cancergencyto.2008.10.010
- 470 27. Pagani IS, Dang P, Kommers IO, et al. BCR-ABL1 genomic DNA PCR response kinetics
471 during first-line imatinib treatment of chronic myeloid leukemia. *Haematologica.* Dec
472 2018;103(12):2026-2032. doi:10.3324/haematol.2018.189787
- 473 28. Ross DM, Branford S, Seymour JF, et al. Patients with chronic myeloid leukemia who
474 maintain a complete molecular response after stopping imatinib treatment have evidence of
475 persistent leukemia by DNA PCR. *Leukemia.* Oct 2010;24(10):1719-24.
476 doi:10.1038/leu.2010.185
- 477 29. Zhang JG, Lin F, Chase A, Goldman JM, Cross NC. Comparison of genomic DNA and cDNA
478 for detection of residual disease after treatment of chronic myeloid leukemia with allogeneic
479 bone marrow transplantation. *Blood.* Mar 15 1996;87(6):2588-93.

- 480 30. Mattarucchi E, Spinelli O, Rambaldi A, et al. Molecular monitoring of residual disease in
481 chronic myeloid leukemia by genomic DNA compared with conventional mRNA analysis. *J Mol*
482 *Diagn.* Sep 2009;11(5):482-7. doi:10.2353/jmoldx.2009.080150
- 483 31. Bartley PA, Latham S, Budgen B, et al. A DNA real-time quantitative PCR method suitable
484 for routine monitoring of low levels of minimal residual disease in chronic myeloid leukemia. *J*
485 *Mol Diagn.* Mar 2015;17(2):185-92. doi:10.1016/j.jmoldx.2014.10.002
- 486 32. Bartley PA, Ross DM, Latham S, et al. Sensitive detection and quantification of minimal
487 residual disease in chronic myeloid leukaemia using nested quantitative PCR for BCR-ABL DNA.
488 *Int J Lab Hematol.* Dec 2010;32(6 Pt 1):e222-8. doi:10.1111/j.1751-553X.2010.01236.x
- 489 33. Cumbo C, Anelli L, Specchia G, Albano F. Monitoring of Minimal Residual Disease (MRD)
490 in Chronic Myeloid Leukemia: Recent Advances. *Cancer Manag Res.* 2020;12:3175-3189.
491 doi:10.2147/CMAR.S232752
- 492 34. Burmeister T, Marschalek R, Schneider B, et al. Monitoring minimal residual disease by
493 quantification of genomic chromosomal breakpoint sequences in acute leukemias with MLL
494 aberrations. *Leukemia.* Mar 2006;20(3):451-7. doi:10.1038/sj.leu.2404082
- 495 35. Branford S. Molecular monitoring in chronic myeloid leukemia-how low can you go?
496 *Hematology Am Soc Hematol Educ Program.* Dec 2 2016;2016(1):156-163.
497 doi:10.1182/asheducation-2016.1.156
- 498
- 499
- 500
- 501

502 **Figure and Table legends:**

503

504 **Figure 1.** Chemistry and bioinformatics workflow of rapid single molecule long-read
505 sequencing.

506 Genomic DNA may contain a target fusion gene. The CRISPR-Cas9 system binds via
507 specific guideRNA (gRNA) designed to enrich for DNA containing regions of interests. The library
508 preparation does not undergo any amplification, and simply requires dA tailing and adapter
509 ligation and a clean-up prior to being loaded onto a sequencing flow cell. On the flow cell
510 libraries are sequenced by nanopores. Data from the sequencing devices are streamed
511 onto biodepot workflow builder and are analyzed starting from FAST5 files, through an initial
512 quality control, then base calling and alignment. After alignment different fusion finder tools
513 were tested including LongGF and BFF before visualizing on IGV for confirmation and
514 interpretation.

515

516 **Figure 2.** Screenshot of our bioinformatics workflow and output. **Panel A:** Bwb workflow
517 showing the workflow including our custom Biodepot Fusion Finder (BFF) and LongGF widgets.

518 **Panel B:** shows enrichment statistics as fusion enrichment and on target enrichment. **Panel C:**
519 alignment can be viewed on IGV based on a BAM file generated from minimap2. This case
520 shows a t(9;22) *BCR-ABL1* fusion in a primary specimen. Reads spanning the breakpoint are
521 colored with the same color on alignment to both genes.

522

523 **Figure 3.** Time required to obtain 3 fusion supporting reads.

524 Top panel shows the time to 3 fusion supporting reads seen in the cell lines samples. Bottom
525 panel shows time to 3 fusion supporting reads seen in patient samples. Details of specific times
526 are listed in supplementary table 1.

527

528 **Table 1.** Characteristics of the Primary Specimens analyzed

529

530 **Table 2.** Characteristics of the Runs

531

532 **Supplementary Table 1.** Times and total reads before 3 fusion reads could be confirmed in cell
533 lines and patient samples.

Figure 1.

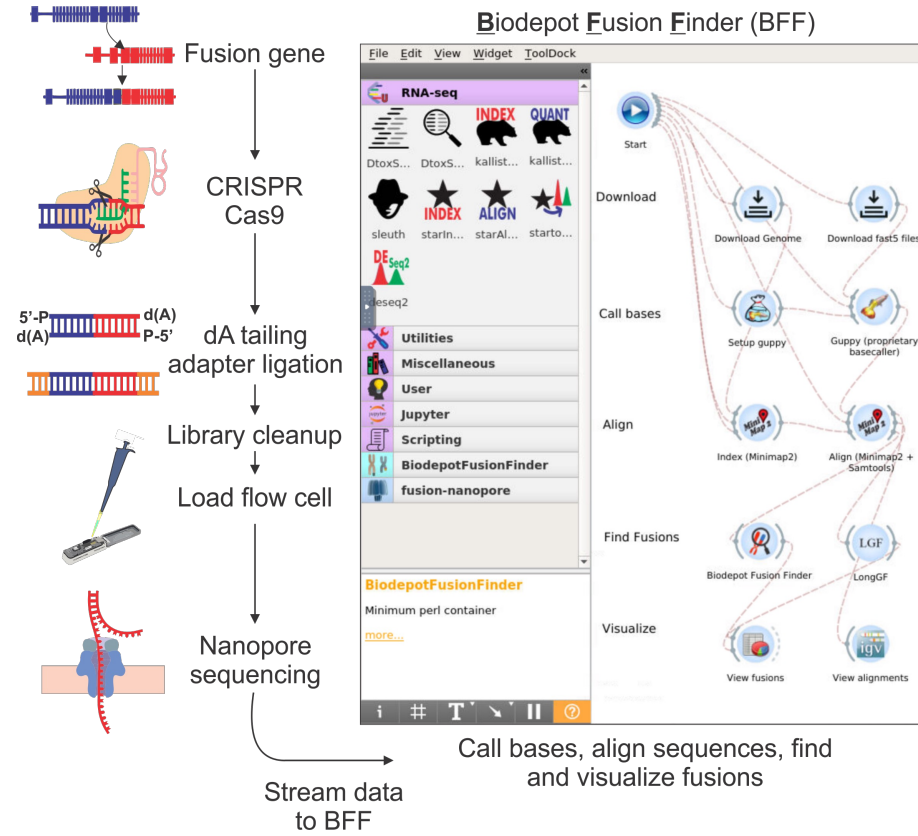
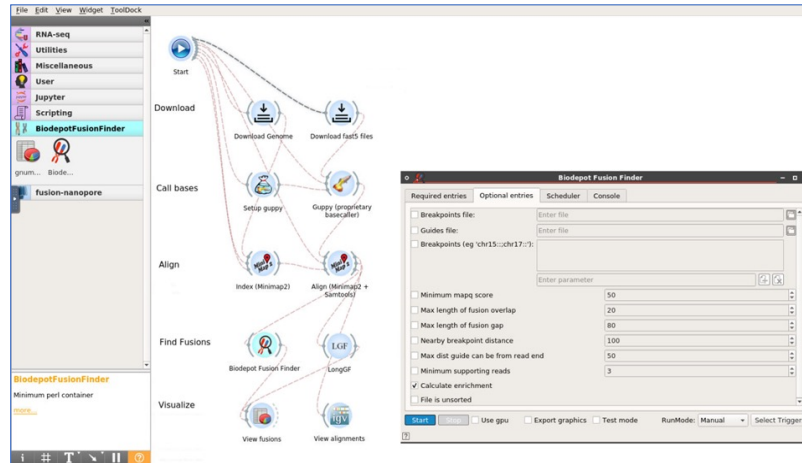


Figure 2.

A: Bwb workflow with LongGF & BFF



B: Enrichment statistics computation output

	A	B	C	D	E	F	G	H	I	J
1	Breakpoint	Gap/Overlap	Count	Nearby-count	ReadThru	NearbyReadThru	Breakpoint enrichment	Breakpoint nearby enrichment	Fraction on target	Fraction on target nearby
2	chr9:133607156:chr22:23632742	2	1	29	30	31	30.19597	875.68318	0.03333	0.93548
3	chr9:133607152:chr22:23632739	4	1	29	29	31	30.19597	875.68318	0.03448	0.93548
4	chr9:133607147:chr22:23632743	5	21	29	28	31	634.11541	875.68318	0.75000	0.93548
5	chr9:133607164:chr22:23632742	5	1	29	31	31	30.19597	875.68318	0.03226	0.93548
6	chr9:133607168:chr22:23632748	5	1	29	30	31	30.19597	875.68318	0.03333	0.93548
7	chr9:133607158:chr22:23632736	8	1	29	31	31	30.19597	875.68318	0.03226	0.93548
8	chr9:133607147:chr22:23632715	16	1	29	31	31	30.19597	875.68318	0.03226	0.93548
9	chr9:133607177:chr22:23632748	17	1	29	30	31	30.19597	875.68318	0.03333	0.93548
10	chr9:133607153:chr22:23632738	29	1	29	31	31	30.19597	875.68318	0.03226	0.93548
11	Coverage	MaxBpFromEnd	GuideReads	Enrichment						
12		0.03333	50	212	6403.546					

C: IGV image confirming fusion

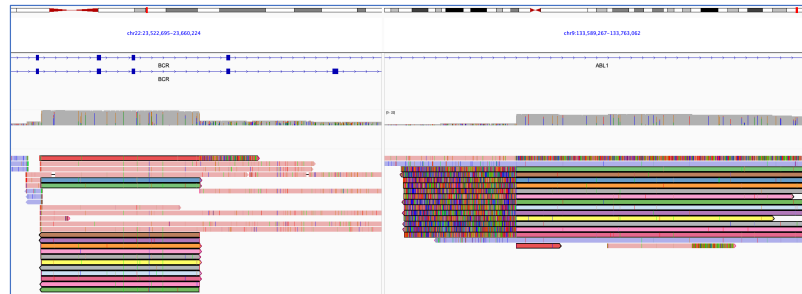


Figure 3.

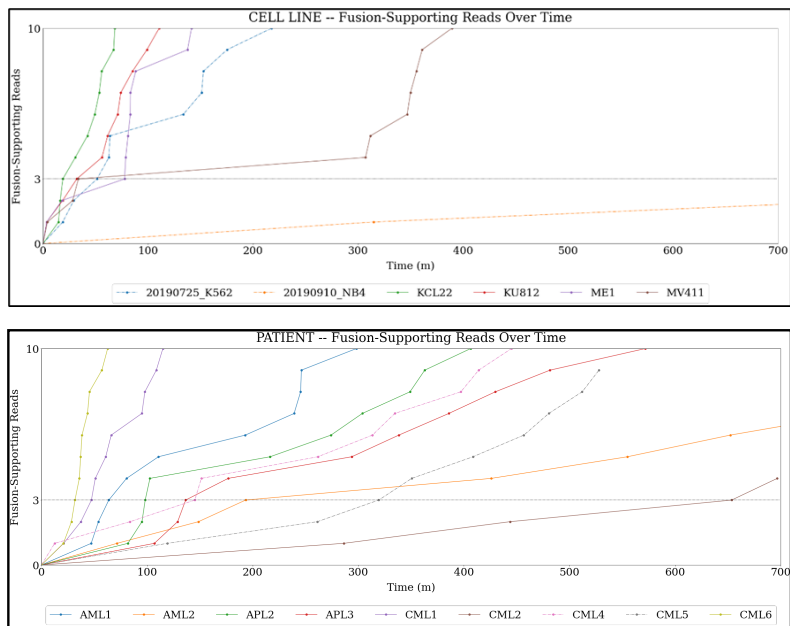


Table 1- Characteristics of the Primary Specimens analyzed

Patient Code	Age	Gender	Diagnosis	Specimen Source	Disease Burden	Cytogenetics/Molecular
CML1	61 - 65	Male	CML-AP Rlps	BM	2% blasts	46,XY,t(5;12)(q33;q15),t(9;22)(q34;q11.2)[2]/47,sl,+8[18]
CML2	41 - 45	Female	CML-BC	PB	NA	Outside cytogenetics confirmed t(9;22) but only in 3/21 karyotypes. Full karyotype not available.
CML3	56 - 60	Male	CML-AP	PB	NA	Patient with long standing p190 CML. At this time-point sample presents MDS/MPN with low level BCR-ABL1 by PCR.
CML4	21 - 25	Male	CML-CP	PB	3% blasts	46,XX,t(9;22)(q34;q11.2)[20]
CML5	61 - 65	Male	CML-AP Rlps	PB	3% blasts	46,XY,t(5;12)(q33;q15),t(9;22)(q34;q11.2)[2]/47,sl,+8[18]
CML6	65 - 70	Female	CML-BC	PB	16% blasts	46,XX,t(9;22)(q34;q11.2)[4]; Confirmed by molecular studies.
AML1	36 - 40	Male	AML	PB	90% blasts	46,XY,t(9;11)(q21;q23)[20]
AML2	36 - 40	Female	AML	PB	20% blasts	47,XX+8,inv(16)(p13.1q22)[17]/46,XX[3]
APL1	81 - 85	Male	APL	PB	<1% blasts	Patient with collision Multiple Myeloma and APL in BM with 30% blasts, not available for analysis. No circulating APL in PB.
APL2	31 - 35	Female	APL	BM	80% blasts	46,XX,der(2)t(2;17)(q33;q21)t(15;17)(q22;q21),der(15)t(15;17),der(17)t(2;17)[3]/47,sl+8[3]/48,sdl,+8[12]/46,XX[2]
APL3	21 - 25	Male	APL	PB	75% blasts	NA
APL4	36 - 40	Male	APL	BM	83% blasts	AML with isochromosome 17q
APL5	36 - 40	Male	APL	PB	79% blasts	AML with isochromosome 17q
APL6	41 - 45	Male	APL	BM	22% blasts in a <5% marrow	46, XY,t(1;7)(q21;q21), t(4;10)(q21;q25), add(8)(p23), del(9)(p22), t(11;17)(q23;q25)[4]/46, XY[16] from 3 months prior

CML= Chronic Myeloid Leukemia; AML= Acute Myeloid Leukemia; APL= Acute Promyelocytic Leukemia, AP= Accelerated Phase, BC= Blast Crisis, PB= Peripheral Blood, BM= Bone Marrow; NA= not available

Table 2. 2. Characteristics of the Sequencing Runs

Sample	DIN	Device	Gb of Data	Mean Coverage	All Reads	N50 All Reads	Median Read Length	Median P HRED	% Reads Aligned	Fusion Reads - LongGF	Fusion Reads - BFF	On-target Enrichment	Fusion Enrichment spanning breakpoint	Disease Burden
K562	9.8	Flongle	0.12	0.04	35943	5980	2200	9.516	87.57	0	29	5830.00	797.50	NA
KU812	8.7	Flowcell	0.36	0.11	86579	9690	1820	12.759	96.29	0	51	2108.33	467.50	NA
KCL22	9.7	Flowcell	0.69	0.21	50439	32200	6290	12.259	91.86	5	171	2118.70	817.83	NA
NB4	9.1	Flongle	0.07	0.02	23122	5890	1450	9.133	88.86	3	3	848.57	141.43	NA
MV4;11	8.6	Flowcell	0.28	0.08	40000	27100	1890	11.98	89.34	57	71	2698.93	836.79	NA
ME1	8.2	Flowcell	0.91	0.28	433846	16500	872	11.17	84.69	45	37	1305.49	134.18	NA
CML1	9.8	Flowcell	0.92	0.28	32723	24600	9600	12.309	91.74	0	142	1563.91	509.35	CML-AP Relapse
CML2	9.6	Flowcell	0.04	0.01	4026	22400	6520	11.148	72.08	0	4	2227.50	330.00	CML BC
CML3	9.8	Flongle	0.09	0.03	9857	19900	5860	9.45	85.54	0	0*	3006.67	NA	CML-AP
CML4	9.7	Flongle	0.93	0.28	109648	15200	1800	9.314	85.97	0	14	557.10	49.68	CML-CP
CML5	ND	Flongle	0.04	0.03	3907	22400	8090	8.95	91.06	0	8	1283.33	293.33	CML-AP Relapse
CML6	9.2	Flowcell	5	1.52	381552	30800	7090	11.94	90.73	0	in	1539.12	142.56	CML BC
AML1	9.3	Flowcell	5.47	1.66	639884	15100	6340	9.89	86.76	28	38	535.12	22.93	90% blasts
AML2	9.2	Flowcell	1.59	0.48	161010	19400	6540	12	89.95	0	0	649.62	NA	20% blasts
														0% blasts in PB, APL confirmed by FISH on BM sample
APL1	9.6	Flowcell	0.17	0.05	13486	24700	7450	12.942	92.49	0	0**	1242.35	NA	80% blasts
APL2	7.76	Flowcell	1.1	0.33	40919	24200	8770	12.12	92.69	8	10	618.00	30.00	80% blasts
APL3	9.7	Flowcell	0.31	0.09	13625	22500	6910	12.422	93.51	8	10	872.90	106.45	75% blasts
APL4	9.6	Flowcell	1.63	0.49	176665	25200	3740	12.38	93.27	0	0***	1785.64	NA	83% blasts
APL5	9.3	Flowcell	0.22	0.07	31863	24200	2450	11.82	91.49	0	0***	2520.00	NA	79% blasts
APL6	7.5	Flowcell	1.06	0.32	485374	4650	1230	11.7	84.27	0	2****	650.66	6.23	22% blasts on flow in a <5% marrow

* Pt had very low BCR-ABL1 levels that was only detected by qualitative PCR and not quantitative PCR (meaning below 0.01%)

** Pt with confirmed BM disease but PB sample that is used for this assay had <1% blast count.

*** Pt had an unusual fusion between a viral gene and RARA(reference: Sala-Torra, Blood Adv. Apr 14 2022), however a human reference genome was used on BFF to generate this data and thus this fusion could not be detected.

**** Pt was suspected to have APL, but subsequent cytogenetic results showed complex karyotype with an unexpected KMT2A fusion and this was a challenged sample with an N50 below 5000bp.

CML= Chronic Myeloid Leukemia; AML= Acute Myeloid Leukemia; APL= Acute Promyelocytic Leukemia, AP= Accelerated Phase, BC= Blast Crisis, PB= Peripheral Blood, BM= Bone Marrow; NA= not available

FISH = fluorescence in-situ hybridization