

Multiclonal Profiling of FLT3-ITD in AML Using MinION Sequencing: A Tailored Clustering Approach to Enhance Subclonal Detection

Jordi Martínez-Serra¹, Aser Alonso-Carballo¹, Ángel Horrillo¹, Paula Gómez¹, Oliver Vögler², Antonio Gutiérrez¹, Antonia Sampol¹

¹Hematology Department, Son Espases University Hospital, IdISBa, Palma, Spain; ²Group of Advanced Therapies and Biomarkers in Clinical Oncology, IUNICS, University of the Balearic Islands, Palma, Spain

Correspondence: Jordi Martínez-Serra, Molecular Diagnostic Laboratory, Son Espases University Hospital, Palma de Mallorca, Spain, Email jordimserra@gmail.com

Background: Internal tandem duplications (FLT3-ITDs) in the FLT3 gene constitute a key driver mutation in acute myeloid leukemia (AML), strongly associated with poor prognosis and therapeutic resistance. Although general-purpose structural variant callers such as Sniffles have been used to detect FLT3-ITDs, their limitations in resolving clonal diversity and low-frequency variants can lead to underrepresentation of minor clones. These shortcomings highlight the need for a dedicated bioinformatics pipeline.

Materials and Methods: We developed a custom clustering-based pipeline to overcome the constraints of generic SV callers, leveraging Oxford Nanopore's MinION for sequencing. Our method focuses on FLT3-ITDs by grouping near-identical insertions into biologically meaningful subclones, thereby allowing accurate variant detection of even low-frequency events. The pipeline was benchmarked against capillary electrophoresis (CE) and Sniffles at various thresholds (including 10%, 20%, and 50% allele-frequency cutoffs), with results validated via IGV inspection and cross-mapping.

Results: The pipeline successfully detected FLT3-ITDs across all tested samples, including low-frequency variants and diverse subclones that Sniffles overlooked. Analyses uncovered complex multiclonal architectures composed of dominant clones (~20–25% of reads) plus multiple minor subclones differing in length, sequence, and breakpoint. Crucially, our approach identified duplications as short as 15 bp—events often dismissed by conventional SV callers. Comparative analyses showed that Sniffles failed to call several biologically validated ITDs detected by our custom pipeline.

Conclusion: Third-generation sequencing combined with a tailored clustering strategy enhances the detection of FLT3-ITDs and clonal diversity in AML compared to generic variant callers. This method provides critical insights into subclonal populations driving relapse and therapeutic resistance—particularly in relapsed/refractory AML—underscoring the importance of specialized pipelines for precision medicine in leukemia.

Keywords: acute myeloid leukemia, AML, third-generation sequencing, TGS, FLT3-ITD, minION

Background

Acute myeloid leukemia (AML) is a clinically heterogeneous malignancy characterized by the rapid accumulation of abnormal myeloid progenitor cells. Among the genetic abnormalities implicated in AML, mutations in the FLT3 gene—especially internal tandem duplications (FLT3-ITDs) and tyrosine kinase domain mutations (TKDs)—stand out for their high prevalence and strong correlation with adverse outcomes, making them clinically significant therapeutic targets.^{1–3} FLT3-ITDs typically appear in the juxtamembrane domain, disrupting normal receptor function and promoting constitutive signaling that drives leukemogenesis. Notably, co-occurrence of FLT3-ITDs with other mutations (eg, DNMT3A, TET2, NPM1, and partial MLL duplications) can worsen prognosis and highlights the genetically complex nature of AML.^{4–6}

Third-generation sequencing (TGS) platforms such as the Oxford Nanopore MinION generate long reads that span repetitive or duplicated regions in full, thereby overcoming many of the limitations inherent in short-read platforms. In addition to fast turnaround times, TGS enables long-read sequencing that can span repetitive regions, providing advantages for structural variant detection. While its subclonal resolution is promising, it does not exceed second-generation sequencing platforms in terms of cost-effectiveness or throughput. Therefore, the benefit lies more in read architecture than in absolute sequencing depth.^{7,8} Despite its potential, TGS-based FLT3-ITD detection requires sufficiently high coverage to capture minor variants. Limited depth may fail to delineate low-frequency events, underestimating the complexity of FLT3-ITD burdens. Detection of FLT3-ITD from MinION sequencing (MS) can be facilitated by structural variant callers such as Sniffles,⁹ although these tools are primarily optimized for large genomic rearrangements and may underperform with smaller duplications or low-frequency subclonal events in repetitive regions.^{10,11}

When analyzing the thousands of reads generated by TGS, small sequence discrepancies—whether from biological variation or basecalling errors—often produce a swarm of slightly different ITD variants around a major clone.⁷ Without post-processing, this can yield multiple small peaks that obscure the true clonal structure of a sample. Clustering algorithms tackle this by merging near-identical insertions into biologically meaningful sets (clones), thus revealing discrete FLT3-ITDs that share length and high sequence identity. This approach (i) consolidates artifactual variability by grouping minor sequencing errors into true consensus sequences, (ii) reduces noise by limiting the effect of random mismatches or small indels, and (iii) differentiates major from minor clones, allowing a more granular view of ITD complexity that could be relevant for future studies on AML risk stratification and personalized treatment, although its current clinical impact remains to be demonstrated. Based on empirical inspections in IGV and iterative alignments, a 90% identity threshold serves as a practical compromise; minor clones with low read counts can still be misclassified at this threshold, but higher coverage helps ensure adequate representation of less abundant variants.

Conventional FLT3-ITD detection with capillary electrophoresis (CE) offers only fragment-size profiles, masking subtle subclonal diversity. Minor differences in ITD length or breakpoint usage often merge into a single band or go undetected if below resolution limits. Thus, CE may underestimate the genuine complexity of FLT3-ITD–positive AML. Short-read sequencing technologies (eg, Illumina) can theoretically capture these minor variations but may rely on parameters that filter out low-frequency insertions or struggle with alignments in the highly repetitive juxtamembrane domain of FLT3. Furthermore, the high costs and extended runtimes required for deep short-read coverage can be prohibitive in routine clinical settings.¹² We aim to develop a TGS pipeline to analyze FLT3-ITD testing and compare it with conventional detection based on CE.

Studies consistently highlight that high FLT3-ITD burden (allelic ratio >0.5–0.7) correlates with worse overall and disease-free survival, as well as an increased relapse risk. Co-existing mutations (eg, NPM1, DNMT3A, or TET2) can compound this adverse prognosis. Indeed, FLT3-ITD is now recognized as a cornerstone in personalized AML management, prompting routine incorporation of allelic ratio measurements into risk stratification models (eg, ELN).³ Moreover, longer FLT3-ITDs have also been associated with inferior clinical outcomes.^{13,14} Importantly, widely used SV callers often fail to detect small (5–50 bp) ITDs, especially in repetitive domains, leading to underdetection of clinically significant subclones. By carefully examining each insertion event—rather than focusing solely on large structural rearrangements—our clustering strategy addresses this gap and enables the detection of ITDs as short as 15 bp.

Materials and Methods

Sample Collection and RNA Extraction

Bone marrow aspirates were collected from five AML patients. All were diagnosed with AML, and three had been previously confirmed as FLT3-ITD–positive by capillary electrophoresis (CE). The study was conducted in accordance with national and international ethical guidelines, including the Declaration of Helsinki, and informed consent was obtained under protocols approved by the *Ethics Committee for Drug Research of the Balearic Islands (CEIm-IB)*. CEIm-IB was selected as the ethics committee with the relevant expertise and authority to oversee research activities in the region. All study protocols were reviewed and approved by the committee, in order to ensure compliance with local

regulatory and ethical standards. Total RNA was extracted from fresh or cryopreserved cells using the RNeasy Mini Kit (Qiagen), following the manufacturer's instructions.

cDNA Synthesis and FLT3 Amplification

Complementary DNA (cDNA) was synthesized and amplified in a one-step RT-PCR reaction using the SensiFAST One-Step RT-PCR Kit (Bioline), which includes reverse transcription and PCR amplification in a single tube. The reaction targeted the juxtamembrane domain of FLT3—a hotspot for internal tandem duplications (ITDs)—and was carried out using the following primers:

- Forward primer: 5'-GATCCTAGTACCTTCCCAAAGTC-3'.
- Reverse primer: 5'-GGTGTCGACGTACTCTAAACAGT-3'.

Thermal cycling conditions were as follows:

- Reverse transcription: 45 °C for 15 min.
- Initial denaturation: 95 °C for 2 min.
- 30 cycles of:
 - Denaturation at 94 °C for 30s.
 - Annealing at 56 °C for 45s.
 - Extension at 72 °C for 30s.
- Final extension: 72 °C for 7 min.

Amplicon sizes (~500–700 bp) were confirmed via agarose gel electrophoresis and purified prior to library preparation.

MinION Library Preparation and Sequencing

Sequencing libraries were prepared with the Ligation Sequencing Kit V14 (SQK-LSK114) following the manufacturer's protocol. Purified FLT3 amplicons plus Lambda phage DNA (internal control) were ligated to Oxford Nanopore adapters, then loaded on Flongle or MinION flow cells depending on the desired throughput. Sequencing was run for up to two hours to achieve sufficient coverage. Basecalling employed the Dorado suite (Oxford Nanopore), configured for duplex analysis to improve accuracy by utilizing both DNA strands. Reads with Q-scores <10 were discarded. The Lambda control served to monitor run performance and validated that the amplicon data were processed under identical conditions.

Alignment and Preliminary Insertion Detection

Reads were aligned to the FLT3 reference sequence (XM_054374312.1) via minimap2 (v2.24) with the “map-ont” preset. Resulting BAM files were sorted, indexed (samtools v1.15), and visually inspected using IGV. A custom Python script parsed CIGAR strings to identify insertions (“I”) ≥15 bp in length, recording insertion start position, size, and the inserted sequence.

Clustering ITD Variants

To cluster insertion sequences, we first applied a similarity-based grouping using Python's difflib.SequenceMatcher, which calculates a normalized alignment score based on internal heuristics. Sequences sharing ≥90% similarity were clustered together, a threshold empirically chosen to balance MinION's ~10% base error rate with the detection of genuine subclonal divergence. For each cluster, we then computed a consensus ITD sequence using Biopython's pairwise2.align.globalxx, a global alignment function with uniform match scoring (match = 1, mismatch = 0, no gap penalties). This dual-step approach allowed efficient clustering of noisy reads followed by accurate consensus refinement. Supporting read counts were recorded per cluster.

Custom FLT3-ITD Detection and Clustering Pipeline

We implemented a Python pipeline to automate the detection, clustering, and consensus refinement of FLT3-ITDs following minimap2 alignment. In brief, each aligned read (parsed via pysam) was examined for insertions above a user-defined length. The identified insertions were then grouped into clusters if they shared at least 90% sequence identity (evaluated by pairwise alignment). A consensus for each cluster was generated by progressively aligning all members; positions with mismatches were temporarily replaced with “N” before selecting the most frequent base at each position. This targeted approach prioritizes the known FLT3 hotspot for internal tandem duplications, in contrast to general-purpose SV callers that may miss insertions failing to meet their minimum size or support criteria.

Validation and Visualization

Manual IGV inspection was performed by tracing representative consensus sequences back to the raw reads in order to confirm authentic breakpoints. Finally, the percentage of reads supporting each cluster was calculated relative to the total number of high-quality, FLT3-mapped reads.

Results

MinION Sequencing vs Capillary Electrophoresis

In all three clinically confirmed FLT3-ITD–positive AML cases, MinION sequencing accurately detected the ITDs. Insertion sizes and allelic ratios largely mirrored those obtained by conventional capillary electrophoresis (CE) (Figures 1 and 2), underscoring the suitability of third-generation sequencing for capturing the broader FLT3-ITD landscape, including variants that pose challenges for short-read methods (Figure 1).

Insertion Analysis Across Diagnostic Samples

We evaluated five diagnostic samples: control1, control2, pos1, pos2, and pos3. A 90% similarity threshold was applied to cluster near-identical insertions. Table 1 summarizes key metrics.

Notably, pos1 displayed the highest rate of insertion-bearing reads (59.81%), while pos2 and pos3 showed moderate insertion frequencies (17.78% and 13.50%, respectively). Both control samples lacked insertions.

Clustering Reveals Closely Related Subclones and Divergent ITDs

Applying the 90% identity threshold across FLT3-ITD–positive samples yielded 95 clusters encompassing 251 grouped insertions. Most samples had a single dominant clone, plus multiple minor subclones with small length or breakpoint differences. Sample pos3, however, exhibited a distinct biclonal configuration: besides a major ~80-bp clone, a second

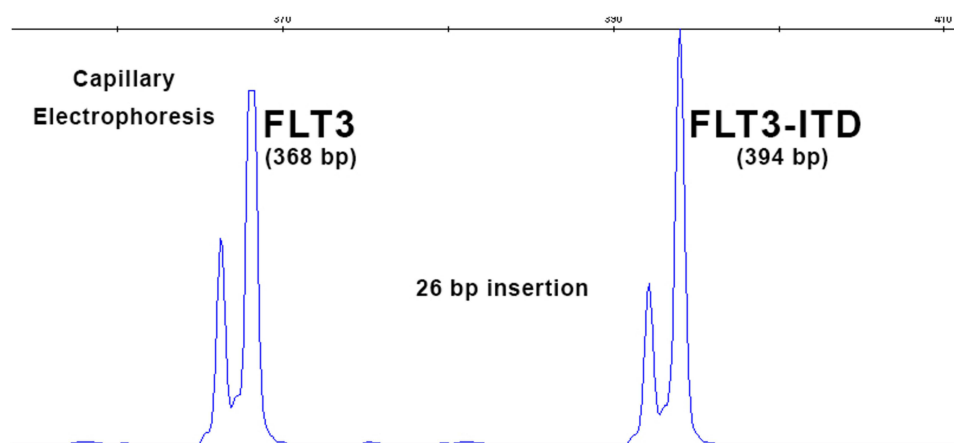


Figure 1 Representative example of a capillary electrophoresis (CE) profile displaying an FLT3-ITD band whose measured size corresponds closely to the ITD insertion identified by our MinION-based pipeline. The alignment of these two measurements confirms both the reliability of the traditional CE approach for initial screening and the capacity of third-generation sequencing to pinpoint precise insertion lengths and breakpoints.



Figure 2 IGV Visualization of FLT3-ITD Insertions in pos3 vs Control. Top panel: Control sample (no FLT3-ITD) showing wild-type alignment of FLT3 with no insertions. Bottom panel: pos3 sample displaying multiclonal ITD architecture, including dominant ~80 bp duplications and a minor ~30 bp insertion cluster (blue reads).

notable ~30-bp clone emerged from a separate breakpoint region, indicating a more divergent branch rather than a minor tweak of the same ancestral ITD. In addition, numerous smaller subclones (<10 reads each, <5% frequency) were detected (Figure 3).

Shared Breakpoint Hotspot Suggests a Common Ancestral ITD

All major and minor clones mapped to a similar region within the FLT3 juxtamembrane domain.¹⁵ This finding supports the idea of an original ancestral duplication followed by incremental variations. High mapping quality (MAPQ ≥20) across these variants reinforced their authenticity.^{16,17} Representative clusters were traced back (IGV) to raw reads to confirm consistent breakpoints without alignment artifacts. Whereas pos1 displayed a broader distribution of insertion-carrying reads, pos3 had a more structured subclonal makeup, including at least four significant groups (Table 2).

The presence of multiple ~80-bp clones sharing partially overlapping breakpoints, combined with a distinct ~30-bp event in a different region, suggests parallel or branched evolutionary pathways in pos3.

Addressing Threshold Selection and Limitations

While clustering at 90% identity reduces artifactual noise, coverage depth remains a limiting factor. Minor subclones supported by few reads might be underrepresented or merged with dominant clones. Stricter thresholds (≥95%) could

Table 1 Summary of Sequencing Metrics Across Control and FLT3-ITD–Positive Samples

Sample	Total Reads Analyzed	Reads with Insertions (>15 bp)	% with Insertions	Total Bases (Post-filtering)	Q20 (%)	Q30 (%)
Control1	1807	0	0.00%	3,565,349	64.77%	37.24%
Control2	3612	0	0.00%	6,343,452	71.65%	47.12%
pos1	2548	1524	59.81%	4,483,871	78.84%	57.24%
pos2	3470	617	17.78%	5,947,752	69.15%	39.37%
pos3	1541	208	13.50%	2,079,739	73.88%	47.93%

Note: Listed for each sample are the total reads analyzed, number and percentage of reads bearing insertions exceeding 15 bp, total bases post-filtering, and the respective Q20 and Q30 percentages, reflecting overall sequencing quality.

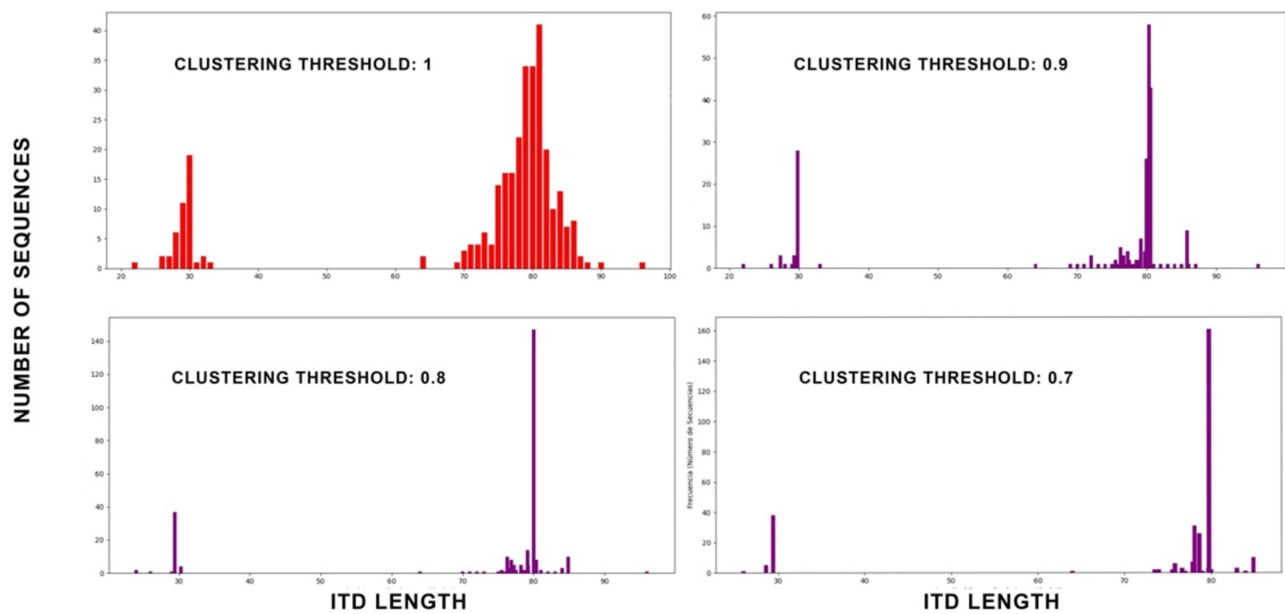


Figure 3 Clustering and ITD Length Distribution in the three FLT3-ITD–positive AML samples (pos3). Visualization of clustering dynamics across sequence similarity thresholds, ranging from 100% (1.0) to a minimum of 70% (0.7). Each panel shows a histogram representing the size of insertions detected using our clustering-based pipeline. Dominant clones typically appear as major peaks, while minor clones manifest as smaller satellite peaks. As the similarity threshold decreases, the total number of clones is reduced, reflecting the merging of closely related sequences. However, lower thresholds (eg. 0.7) enable the differentiation of distinct subclones that are otherwise grouped under higher similarity criteria, highlighting divergent evolutionary trajectories within the FLT3-ITD population. Color intensity corresponds to the relative abundance of each cluster. These visualizations reinforce that AML samples often harbor both dominant clones and a constellation of lower-frequency subclones. By improving resolution of this multiclonal environment, our method refines prognostic models and supports precision medicine strategies.

resolve slight variations yet risk fragmenting legitimate subclones if coverage is insufficient. Conversely, lowering the threshold (<80%) may unify closely related clones but also risk grouping genuinely distinct events. As such, deeper sequencing can mitigate these concerns, balancing over-splitting and over-merging.

To visualize how subclones transition across multiple thresholds, we generated a Sankey diagram (Figure 4) for the pos3 sample, ranging from a highly stringent cutoff (1.0) down to a more relaxed one (0.5). Each column in the diagram represents the clusters identified at a given threshold, and the width of each flow corresponds to the number of reads shared between successive thresholds. Notably, several smaller clusters at Th1.0 and Th0.9 converge into a handful of larger, unified clusters by Th0.5. This highlights that seemingly distinct subclones at higher similarity requirements can in fact share enough identity to merge under more lenient criteria. Consequently, the Sankey visualization both confirms the multiclonal nature of pos3 and illustrates the threshold’s pivotal role in defining the boundaries between closely related variants. Importantly, the persistence of two dominant clusters—one centered around an ~80 bp ITD and another around ~30 bp—across all thresholds suggests the presence of biclonal architecture, potentially reflecting parallel evolutionary trajectories within the leukemic population.

Detection Limitations of Conventional SV Callers in Short Insertions

In our analysis, we observed that Sniffles, a commonly used structural variant caller, exhibited limitations in detecting insertions of approximately 30 base pairs in length (Table 3). This oversight highlights a critical gap in the tool’s

Table 2 Summary of the Main Clonal Groups Detected in Sample pos3

Clonal Group	Reads (Approx. %)	Insertion Length	Breakpoints
1	56 (~19.05%)	~80.34 bp	1812, 1814–1816
2	43 (~14.63%)	~80.53 bp	1814–1816, 1818–1820
3	27 (~9.18%)	~29.81 bp	1853, 1856–1859
4	26 (~8.84%)	~79.92 bp	1815–1816

Note: Each clone is characterized by its approximate frequency (percentage of total FLT3-ITD reads), average insertion length, and specific breakpoint positions.

Sankey diagram (pos3)

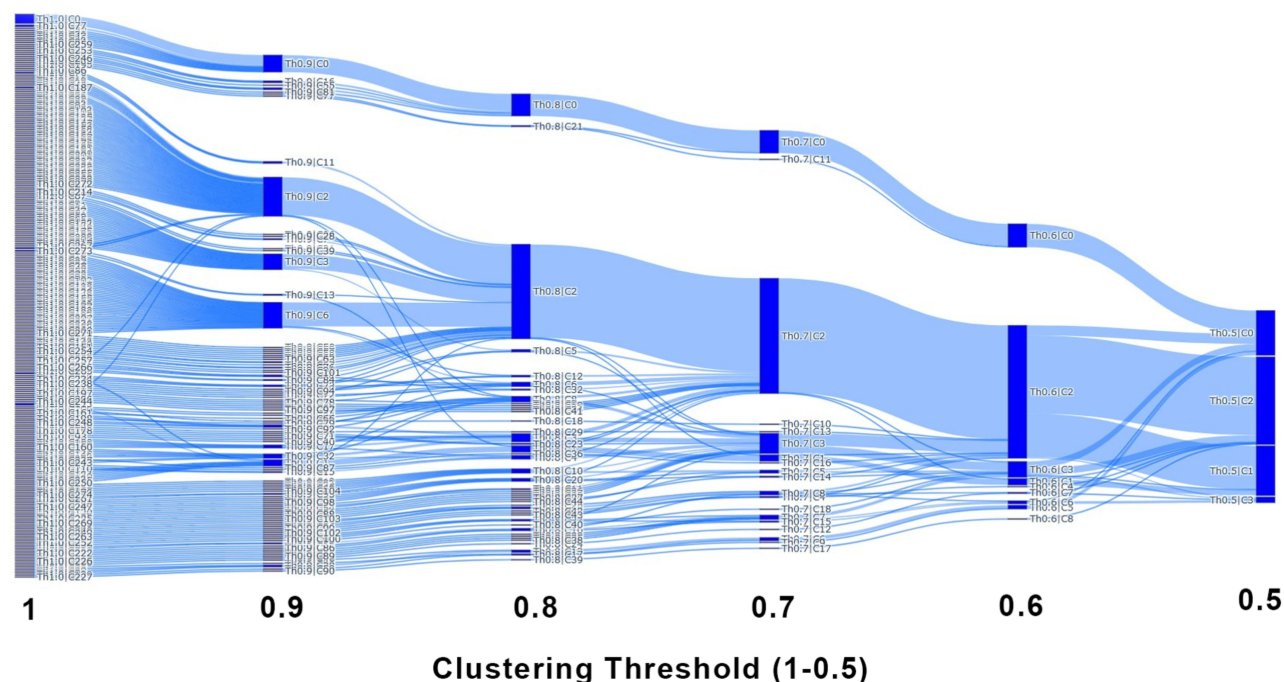


Figure 4 Sankey diagram illustrating clonal merging at different similarity thresholds in the pos3 sample. The leftmost column (Th1.0) shows many small clusters. As the threshold decreases (Th0.9, 0.8, 0.7, 0.6, 0.5), these clusters merge—often uniting several previously distinct sets of reads into a smaller number of larger, unified clusters. Flow width indicates the relative number of reads transferred from one threshold's cluster(s) to the next. This visualization emphasizes the importance of threshold selection: small differences in insertion sequence can be preserved at high cutoffs but are grouped under more relaxed criteria.

sensitivity, particularly for small insertions within our sample set, which could impact the overall accuracy of variant detection in genomic studies. Further investigation into these discrepancies would be necessary to refine the tool's performance or to consider alternative methods for capturing these subtle genetic changes.

Challenges with Duplex Basecalling in Low Clonality Scenarios

Despite employing duplex basecalling with Dorado, this approach presents a significant drawback when dealing with very low clonal fractions. The tool tends to dismiss these events as potential sequencing errors rather than true biological variants. This issue is not unique to nanopore sequencing; similar challenges are encountered with technologies like Illumina, where low-frequency variants might be filtered out during data processing. In the case of Sanger sequencing, such events might not even produce a detectable peak on the electropherogram, rendering them invisible to traditional analysis methods. This underscores the need for more sensitive detection algorithms or alternative strategies to accurately capture and verify these rare variants in clinical and research settings.

Table 3 Comparison of FLT3-ITD Detection Results Across Three Positive AML Samples

Sample	Clustering-Based Method	Sniffles Result
pos1	Detected (one clone)	Not Detected
pos2	Detected (one clone)	Not Detected
pos3	Detected (two clones)	Not Detected

Note: While our clustering-based method identified insertions in all cases, Sniffles failed to detect any structural variant. These results highlight the limitations of generic SV callers in identifying short duplications within repetitive genomic regions such as FLT3. Detected | Not Detected.

Discussion

Our results confirm that third-generation Nanopore sequencing enhances FLT3-ITD detection over traditional CE by revealing a broader range of major and minor subclones. While CE effectively screens for primary FLT3 mutations, it can overlook minor variants if they are too close in size or below detection limits.^{18,19} Multiclonality in FLT3-ITD underscores the heterogeneous nature of AML, as demonstrated by dominant clones frequently representing ~20–25% of ITD-bearing reads alongside multiple minor variants with distinct lengths and breakpoints.^{4,20} Conventional CE may underestimate this subclonal complexity when insertion sizes are similar or below detection thresholds.^{21–23} Capturing both high- and low-frequency events may provide a more comprehensive view of clonal complexity in AML, which could be relevant for future studies assessing prognosis and treatment response.^{24–26} Incorporating a clustering-based approach into routine diagnostics could thus reveal a richer subclonal landscape, provided that sequencing depth is sufficient.²⁷

Although clustering at 90% identity reduces noise, coverage remains a key limitation. Additionally, the presence of PCR artifacts or sequencing chimeras cannot be entirely ruled out, especially for low-frequency events supported by few reads. While we used high-fidelity enzymes and size selection to mitigate such artifacts, future implementations of this workflow could benefit from Unique Molecular Identifiers (UMIs) to confirm the biological origin of minor subclones and enhance confidence in their interpretation. Minor clones supported by only a few reads risk merging with dominant clones or being lost entirely.²⁸ Based on our empirical observations, the current pipeline reliably detects subclonal ITDs down to ~5% variant allele frequency (VAF) in regions with coverage exceeding 500–1000×. Below this threshold, sensitivity is reduced and detection becomes more stochastic, especially in the absence of strand-paired reads or UMIs. Variants supported by fewer than 10 reads were considered low-confidence and not included in final outputs. Future improvements—including UMI incorporation and synthetic benchmarking—could help extend the limit of detection toward the 1% VAF range, enabling even finer resolution of minor subclones. More stringent thresholds (≥95%) can split genuine events if coverage is inadequate, while less stringent thresholds (<80%) may collapse truly distinct clones. Furthermore, we observed that outside of the ITD regions, aligned reads showed high base-level concordance with the reference, indicating that sequencing errors were not uniformly distributed but rather concentrated around insertions. This pattern lends additional support to the biological validity of the detected ITDs and the appropriateness of the 90% clustering threshold, even in the absence of synthetic benchmarking.

Our findings also highlight that multiple coexisting FLT3-ITDs point to branching evolution in AML. Even minor variants initially considered silent may expand during therapy, emphasizing the need for early detection to guide more aggressive or combination treatments before relapse occurs.^{29,30} As shown in [Figure 3](#), we illustrate the ITD length distributions and genomic positions for the three positive AML samples, reinforcing that AML samples often harbor both dominant clones and a constellation of lower-frequency subclones. By improving the resolution of this multiclonal environment, our method refines prognostic models and supports precision medicine strategies. While our study was not powered to draw direct correlations between low-frequency subclones and patient outcomes due to the small cohort size, previous reports have demonstrated that minor FLT3-ITD variants—particularly those initially present at low VAF—can expand under therapeutic pressure and contribute to relapse.¹³ Such subclones may harbor differential sensitivity to FLT3 inhibitors or coexist with other mutations (eg, NPM1, DNMT3A) that modulate disease progression.²⁶ Therefore, early identification of these minor ITD-bearing clones could inform risk-adapted treatment decisions, especially in the context of post-remission monitoring or targeted therapies. Larger prospective studies integrating clinical follow-up and therapeutic response data will be essential to validate these associations.

Moreover, our targeted pipeline differs from generic SV callers such as Sniffles—though Sniffles is effective for large rearrangements, it often struggles with small (<30 bp) duplications.^{11,28} By focusing on known FLT3 hotspots and employing a specialized clustering approach, we detect short events that might otherwise go unnoticed. Future enhancements could include higher sequencing depth, adaptive thresholding to refine cluster boundaries, and larger, longitudinal cohort studies to correlate subclonal patterns with clinical outcomes. Integration with co-occurring mutations (eg, NPM1, DNMT3A) would offer an even broader genomic context. By pairing deeper coverage with specialized clustering, subsequent research can more accurately delineate the evolutionary trajectories of AML, refine prognostic tools, and optimize individualized therapies.

This study demonstrates that MinION sequencing, combined with a tailored bioinformatics approach based on clustering, significantly enhances the detection of FLT3-ITD subclones in AML patients, even for low-frequency variants and short duplications (≥ 15 bp) that generic tools like sniffles could overlook. The strategy of clustering insertions with a 90% similarity threshold aligns with the historical error profile of MinION ($\sim 10\%$), allowing differentiation between technical artifacts and genuine biological variants. Moreover, the use of duplex basecalling via Dorado increased the overall read accuracy. However, it is important to note that Oxford Nanopore sequencing does not sequence both strands of each DNA molecule in a deterministic or paired manner. Rather, strands arrive stochastically, and many reads are derived from only one strand. In such cases, Dorado's consensus model may either fail to form a duplex or may downweight strand-specific variants as noise. This limitation can lead to the underrepresentation or loss of rare subclonal variants that are only observed in one strand, especially in regions with high homology or repetitive content such as FLT3.

Although the results are promising, the study has inherent limitations due to its preliminary design. Firstly, the small cohort size ($n=5$), moderate number of reads analyzed and coverage in some samples might underestimate the true subclonal diversity, particularly for variants present in $<5\%$ of reads. While our results demonstrate clear technical advantages of the MinION-based clustering pipeline, we acknowledge that the limited sample size—particularly the number of FLT3-ITD-positive patients—constrains broader clinical generalization. This study was designed as a methodological proof-of-concept to validate the detection and resolution of subclonal ITDs rather than to provide outcome-linked statistical models. Ongoing efforts are now focused on expanding the patient cohort through multicenter collaboration to enable more comprehensive clinical correlation and validation.

Future work should prioritize larger cohorts with deeper sequencing ($\geq 1000\times$) to validate the method's sensitivity in real clinical contexts. Secondly, while the 90% threshold was empirically chosen to balance MinION's technical error and biological diversity, its optimization would require benchmarks with synthetic controls or validation through orthogonal methods (eg, Sanger sequencing of individual clones). Lastly, the current pipeline does not comprehensively assess the impact of computational resources, a critical aspect for its implementation in diagnostic settings with limited infrastructure.

Despite these limitations, the proposed approach lays the groundwork for exploring clonal heterogeneity in AML with unprecedented resolution. Integrating this method into longitudinal studies could reveal evolutionary dynamics during relapse or therapeutic resistance, while its extension to other genes prone to ITDs (eg, CEBPA) would broaden its clinical utility. Collectively, this methodology represents an advancement towards the precise characterization of genomic complexities that escape conventional techniques, establishing nanopore sequencing as a viable tool for personalized medicine in oncological hematology.

Acknowledgments

We thank all collaborating clinicians and laboratory staff for their contributions to patient sample collection and technical support.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Garg M, Nagata Y, Kanojia D, et al. Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood*. 2015;126(22):2491–2501. doi:10.1182/blood-2015-05-646240
2. Patel JP, Gönen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1079–1089. doi:10.1056/NEJMoa1112304
3. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447. doi:10.1182/blood-2016-08-733196
4. Schranz K, Hubmann M, Harin E, et al. Clonal heterogeneity of FLT3 -ITD detected by high-throughput amplicon sequencing correlates with adverse prognosis in acute myeloid leukemia. *Oncotarget*. 2018;9(53):30128–30145. doi:10.18632/oncotarget.25729
5. Fröhling S, Schlenk RF, Breitnick J, et al. Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML study group Ulm. *Blood*. 2002;100(13):4372–4380. doi:10.1182/blood-2002-05-1440

6. Schlenk RF, Döhner K, Krauter J, et al. Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N Engl J Med.* 2008;358(18):1909–1918. doi:10.1056/NEJMoa074306
7. Levis M, Perl A, Altman J, et al. A next-generation sequencing-based assay for minimal residual disease assessment in AML patients with FLT3-ITD mutations. *Blood Adv.* 2018;2(8):825–831. doi:10.1182/bloodadvances.2018015925
8. Kim B, Kim S, Lee S, et al. FLT3 internal tandem duplication in patients with acute myeloid leukemia is readily detectable in a single next-generation sequencing assay using the pindel algorithm. *Ann Lab Med.* 2019;39(3):327–329. doi:10.3343/alm.2019.39.3.327
9. Dayer N, Schlenk R, Russell N, et al. Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia.* 2019;33(2):299–312. doi:10.1038/s41375-018-0357-9
10. Cumbo C, Minervini CF, Orsini P, et al. Nanopore targeted sequencing for rapid gene mutations detection in acute myeloid leukemia. *Genes.* 2019;10(12):1026. doi:10.3390/genes10121026
11. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods.* 2018;15(6):461–468. doi:10.1038/s41592-018-0001-7
12. Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics, and applications. *Nature Biotechnol.* 2021;39(11):1348–1365. doi:10.1038/s41587-021-01108-x
13. Arriba-Tutusa P, Mack TS, Bullinger L, et al. Impact of FLT3-ITD location on sensitivity to TKI-therapy in vitro and in vivo. *Leukemia.* 2016;30(5):1220–1225. doi:10.1038/leu.2015.292
14. Liu S-B, Dong H-J, Bao X-B, et al. Impact of FLT3 -ITD length on prognosis of acute myeloid leukemia. *Haematologica.* 2019;104(1):e9–e12. doi:10.3324/haematol.2018.191809
15. Knight TE, Edwards H, Meshinchi S, Taub JW, Ge Y. “FLipping” the story: FLT3-mutated acute myeloid leukemia and the evolving role of FLT3 inhibitors. *Cancers.* 2022;14(14):3398. doi:10.3390/cancers14143398
16. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352
17. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–1303. doi:10.1101/gr.107524.110
18. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnol.* 2018;36(4):338–345. doi:10.1038/nbt.4060
19. Smith CC, Wang Q, Chin CS, et al. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukemia. *Nature.* 2020;485(7397):260–263. doi:10.1038/nature11016
20. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature.* 2012;481(7382):506–510. doi:10.1038/nature10738
21. Cumbo C, Orsini P, Anelli L, et al. Nanopore sequencing sheds a light on the FLT3 gene mutations complexity in acute promyelocytic leukemia. *Leukemia Lymphoma.* 2021;62(5):1219–1225. doi:10.1080/10428194.2020.1856838
22. Minervini CF, Cumbo C, Orsini P, et al. Nanopore sequencing in blood diseases: a wide range of opportunities. *Front Genetics.* 2020;11:76. doi:10.3389/fgene.2020.00076
23. Spencer DH, Abel HJ, Lockwood CM, et al. Detection of FLT3 internal tandem duplication in targeted, short-read-length, next-generation sequencing data. *J Mol Diagn.* 2013;15(1):81–93. doi:10.1016/j.jmoldx.2012.08.001
24. Levis M, Small D. FLT3 tyrosine kinase inhibitors. *Int J Hematol.* 2005;82(2):100–107. doi:10.1532/IJH97.05079
25. Kennedy VE, Smith CC. FLT3 targeting in the modern era: from clonal selection to combination therapies. *Int J Hematol.* 2024;120(5):528–540. doi:10.1007/s12185-023-03681-0
26. Gale RE, Green C, Allen C, et al. The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood.* 2008;111(5):2776–2784. doi:10.1182/blood-2007-08-109090
27. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature.* 2018;562(7728):526–531. doi:10.1038/s41586-018-0623-z
28. De Coster W, D’Hert S, Schultz DT, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnol.* 2019;36(4):338–345.
29. Thiede C, Steudel C, Mohr B, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia (AML). *Blood.* 2002;99(12):4326–4335. doi:10.1182/blood.V99.12.4326
30. Kivioja J, Malani D, Kumar A. FLT3-ITD allelic ratio and HLF expression predict FLT3 inhibitor efficacy in adult AML. *Sci Rep.* 2021;11(1):23565. doi:10.1038/s41598-021-03010-7

OncoTargets and Therapy

Publish your work in this journal

OncoTargets and Therapy is an international, peer-reviewed, open access journal focusing on the pathological basis of all cancers, potential targets for therapy and treatment protocols employed to improve the management of cancer patients. The journal also focuses on the impact of management programs and new therapeutic agents and protocols on patient perspectives such as quality of life, adherence and satisfaction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/oncotargets-and-therapy-journal>

Dovepress
Taylor & Francis Group