Application note

Comprehensive AML MRD Detection: Integrating µCaler™ Technology with Variant Analyzer

Tomáš Hron[1], Tereza Valterová[2], Lenka Žejšková[2], Markéta Kalinová[2]

[1] Data analysis and education in genomics and sequencing technologies, email: tomas.hron@bioprojects.cz, phone: +420 605 325 350

[2] Central Laboratories, Laboratory of genomic medicine, Faculty Hospital Kralovske Vinohrady, Prague, Czech Republic

Abstract

Acute Myeloid Leukemia (AML) diagnostics and Minimal Residual Disease (MRD) monitoring require the detection of variants at extremely low allele frequencies. This study evaluates a bundled solution combining Nanodigmbio µCaler™ AML MRD library preparation with the Variant Analyzer (VARAN) platform.

Use of VARAN's UMI consensus deduplication and specific parameter optimizations enables the detection of ultra-low frequency variants. We provide two different analysis workflows balancing analytical sensitivity while maintaining a low number of false positives. This enables efficient detection of variants down to 0.05% VAF, approaching the technical limits of the NGS approach. This makes the solution uniquely suitable for MRD monitoring.

VARAN solution for AML MRD diagnostics

- Combination of VARAN analysis platform and Nanodigmbio µCaler™ libraries enables detection of ultra-low frequency variants suitable for Minimal Residual Disease (MRD) monitoring.

Ultimate detection limit and sensitivity

- We provide two different analysis workflows balancing sensitivity and specificity. Our approach enables efficient detection of variants down to 0.05% VAF, approaching the technical limits of the NGS method.

Recommended experimental setups

- **Sensitive (>0.05% VAF):** Sensitive wf, 5M read pairs/sample
- **Sensitive cost-effective (>0.05% VAF, compromised specificity):** Sensitive wf, 1M read pairs/sample
- **Balanced (>0.45% VAF, high specificity):** Balanced wf, 1M read pairs/sample
- **Balanced cost-effective (>0.5% VAF, high specificity):** Balanced wf, 500K read pairs/sample

Introduction

Acute Myeloid Leukemia (AML) is a complex hematologic malignancy characterized by the rapid growth of abnormal cells in the bone marrow. Minimal Residual Disease (MRD) refers to the small number of cancer cells that remain in the patient during or after treatment, serving as a critical prognostic indicator for relapse and treatment response monitoring. Detecting these rare clones requires Next-Generation Sequencing (NGS) technologies with high analytical sensitivity. The combination of Nanodigmbio enrichment technology and the VARiant ANalyzer (VARAN) platform enables detection of ultra-low frequency variants essential for Minimal Residual Disease assessment.

The μCaler™ AML Panel v1.0 from Nanodigmbio company is engineered specifically for the rigorous demands of adult acute myeloid leukemia monitoring. The panel targets 32 clinically significant genes, covering a ~42.5 Kb genomic region optimized for MRD monitoring. It is designed to simultaneously enrich for multiple mutation types, including single nucleotide substitutions, small insertions/deletions (indels), and critical gene fusions. The hybridization-based target enrichment method is specifically tuned to capture rare malignant clones even when they represent only a fraction of a percent of the total cell population.

VARAN is a sophisticated web platform provided by Altium International that is designed to transform raw sequencing data into actionable clinical and research insights, with a focus on security and customizability. All calculations are performed on powerful, secure cloud servers located in the Czech Republic, ensuring the protection of sensitive clinical data. VARAN utilizes the Sentieon® toolkit for variant calling, which has been recognized as a world leader in accuracy. Analysis procedures can be tailored to specific sequencing libraries, platforms and diagnostic needs. For MRD applications, VARAN supports the processing of Unique Molecular Identifiers (UMIs) to correct sequencing artifacts and detect low-frequency variants.

Materials and Methods

Samples

A reference (Testing Sample) was diluted into a Buffy Coat pool (10 germinal samples) to create a series of diluted samples: Sample 1 (1:9 dilution of Testing sample), Sample 2 (1:99 dilution), Sample 3 (1:999 dilution).

Library Preparation and Sequencing

Libraries were generated using the μCaler™ AML MRD comprehensive solution and sequenced on DNBSEQ-T1+ sequencer (MGI) with 2x150 bp chemistry. The genes covered by μCaler AML Panel v1.0 are shown in the Table 1. To evaluate the impact of read depth, data were subsampled from 50M down to 10K read pairs producing 8 parallel datasets (50M, 10M, 5M, 1M, 500K, 100K, 50K, 10K). Sequencing quality reached Q30 for more than 96% and Q40 for more than 90% of bases.

Data Analysis in VARAN

Data were processed using the Sentieon toolkit integrated in the VARAN platform. To mitigate sequencing artifacts and improve specificity at low VAF, Unique Molecular Identifiers (UMIs) were used to create duplex consensus reads. Four workflows were compared: i) No UMI (standard somatic calling without consensus creation), ii) Normal (standard somatic calling using UMI-corrected reads, iii) Balanced (optimized for low VAF with balanced specificity), and iv) Sensitive (optimized for extremely low VAF).

Variant Set Definition

To rigorously evaluate the analytical performance of the bundled solution, a gold-standard variant set was established. The initial genetic profiles of the Buffy Coat and Testing Sample were characterized using an

Table 1. The genes covered by μCaler AML Panel v1.0

ASXL1 Exon 12,13	*BRINP3* Exon 3,8	*CBL* Exon 8,9	*CEBPA** Full CDS	*DNMT3A* Exon 8-23	*EZH2* Exon 4-6,8,13-20	*FLT3* Exon 14,15,20	*GATA2* Exon 3-6	*HNRNPK* Exon 4-6,10,12,15,16	*IDH1* Exon 4	*IDH2* Exon 4
JAK2 Exon 14	*KIT* Exon 8,17	*KMT2A*† Intron 8-10	*KRAS* Exon 2,3	*MYH11*† Intron 32	*NPM1* Exon 10,11	*NRAS* Exon 2,3	*PHF6** Full CDS	*PTEN* Exon 5,7	*PTPN11* Exon 3,13	*RAD21** Full CDS
*RUNX1** Full CDS	*SF3B1* Exon 14,15	*SMC1A* Exon 2,9,11,13,15-17,22		*SMC3* Exon 9,10,13,19,24,25,27		*SRSF2* Exon 1	*STAG2* Exon 5,6,8-10,15,17,19,20,25,27-31		*TET2** Full CDS	*TP53** Full CDS
U2AF1 Exon 2	*WT1* Exon 7-10									

** Indicates that the gene is covered across the entire coding sequence (CDS) region.*
† Indicates that the gene is covered in fusion-related intronic regions.

independent NGS custom panel (SureSelect target enrichment HS XT) and analysis tool (Varsome Clinical). Any variants with a VAF >3% located within the µCaler™ AML panel regions were designated as true positives, resulting in a total set of 45 variants.

For the diluted samples (Sample 1, 2, and 3), the expected VAF for each variant was calculated as the weighted sum of the VAFs from the Buffy Coat and Testing Sample, adjusted for the specific dilution factor.

False positive variant set was predicted for VAF levels >0.1% by differentiating between sample dilutions. Specifically, false positive variant was defined as a variant present in Sample 3 with VAF > 0.1% and at the same time absent from Sample 1 (VAF limit 1%).

Sensitivity and FPR Calculation

Sensitivity was calculated based on the total set of all variant-sample combinations across Sample 1, Sample 2, and Sample 3. In this model, a single variant is counted multiple times if it is independently called across different samples and dilution levels, providing a comprehensive view of detection limits.

To provide a standardized metric for specificity, the False Positive Rate (FPR) was calculated as the number of false positive calls per kilobase (kbp) of the panel's target region.

Results and Discussion

Noise reduction by UMI consensus

The detection of ultra-low frequency single nucleotide variants in MRD monitoring is primarily limited by the technical noise floor of the NGS process. While modern sequencing platforms reach relatively high quality—with a sequencing-specific error contribution as low as 0.01% (Q40) —the primary bottleneck remains the PCR reactions during library processing. This accumulated noise can create artificial variants with a VAF reaching close to 1%.

While variant calling tools are, in general, capable of low-frequency variant detection, they cannot efficiently distinguish true variants from this background error rate of the method. To overcome this, VARAN utilizes UMI-assisted duplex consensus read creation (UMI consensus) to efficiently reduce the noise in the data and distinguish true biological variants from stochastic technical artifacts.

As demonstrated in Figure 1, the workflow with UMI consensus (UMI cons.) and the normal (No UMI) workflow show similar performance when detecting variants with a Variant Allele Frequency (VAF) above 1.0%. However, as the detection threshold moves below this limit, the standard pipeline shows a sharp increase in the number of called variants. This divergence corresponds directly to the expected cumulative error rates of PCR amplification and sequencing chemistry.

The UMI consensus workflow effectively flattens the noise curve in the low VAF region, maintaining high analytical sensitivity while drastically reducing the False Positive Rate (FPR). This demonstrates that UMI consensus deduplication is an essential component for reliable variant detection in sensitive mode, providing the necessary specificity to identify rare clones.

VARAN workflow sensitivity

To address the diverse requirements of clinical research and MRD monitoring, two specialized workflows were developed within VARAN platform: Balanced and Sensitive. These workflows balance detection limits against the False Positive Rate (FPR).

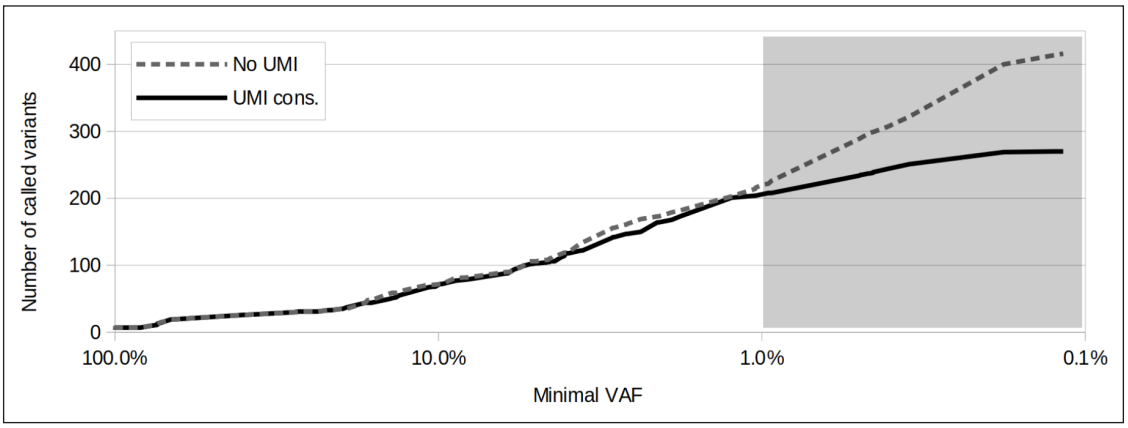

Figure 1. Comparison of variant calling performance with and without UMI consensus. *The graph illustrates the number of variants called across a range of minimal VAF thresholds for all tested samples subsampled to 1M read pairs per sample. The dashed line represents standard somatic calling (No UMI), while the solid line shows the results after UMI consensus deduplication (UMI cons.). The shaded grey area highlights the critical VAF region where technical noise (PCR/sequencing errors) becomes the dominant factor in "No UMI" workflows.*

The performance evaluation was conducted using datasets subsampled to 1M read pairs per sample, providing approximately 6,400x raw coverage. The trade-off between these two workflows is visualized in Figure 2.

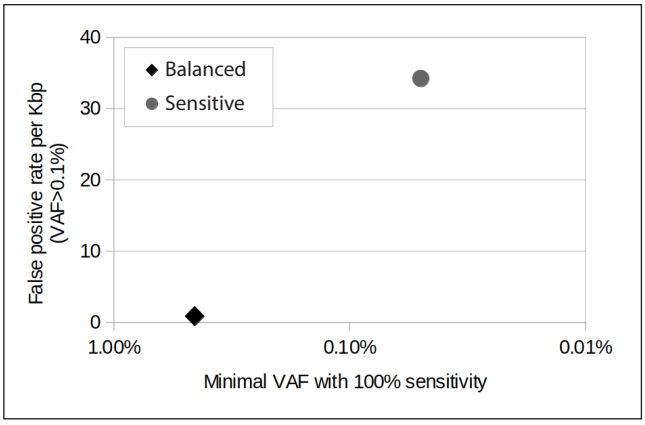

Figure 2. Analytical performance of VARAN Sensitive workflows. *The scatter plot illustrates the relationship between the Minimal VAF required to reach 100% sensitivity and the resulting False Positive Rate (FPR) per kilobase of the target panel. Data points represent the Balanced workflow and the Sensitive workflow based on 1M read pairs of sequencing data.*

Balanced workflow is balanced to provide high sensitivity with low FP background and is recommended for most cases. It provides resolution of 0.45% VAF while maintaining an exceptionally low false positive rate (less than 1 FP per kbp).

Sensitive workflow is optimized for ultra-low frequency detection in MRD monitoring. It achieves resolution of 0.05% VAF. This increased sensitivity comes at the cost of an elevated FPR (approximately 34 FP per kbp), which is acceptable in clinical contexts where missing a residual malignant clone is more critical than the manual review of potential artifacts.

Sequencing depth and analytical sensitivity trade-off

To determine the optimal sequencing coverage for MRD monitoring, we performed a subsampling analysis ranging from 10K to 50M read pairs per sample. The results demonstrate that analytical sensitivity is strictly dependent on sequencing depth up to a specific inflection point, after which additional sequencing provides diminishing returns.

As visualized in Figure 3, sensitivity improves rapidly as sequencing depth increases from 10K to 500K read pairs. However, a distinct plateau is reached between 500K and 1M read pairs. Increasing sequencing depth beyond 1M read pairs results in a significant increase in the PCR duplicate rate, which reaches 96.56% at 50M read pairs (Table 2). This "over-sequencing" does not translate into higher clean sequencing coverage because the library becomes limited by the number of original genomic equivalents captured during the hybridization-based enrichment. While these duplicates are useful for UMI-based error correction, they do not provide new information to increase the resolution beyond the 0.05% VAF threshold.

For both workflows we therefore recommend 1M read pairs per sample as an optimal sequencing output which corresponds to ~ 6 400x raw coverage. Sufficient sensitivity is provided at 0.45% VAF for Balanced workflow, and at 0.05% VAF for Sensitive workflow. For Balanced workflow, user can consider to lower sequencing output up to 500K read pairs expecting the considerably similar detection limit.

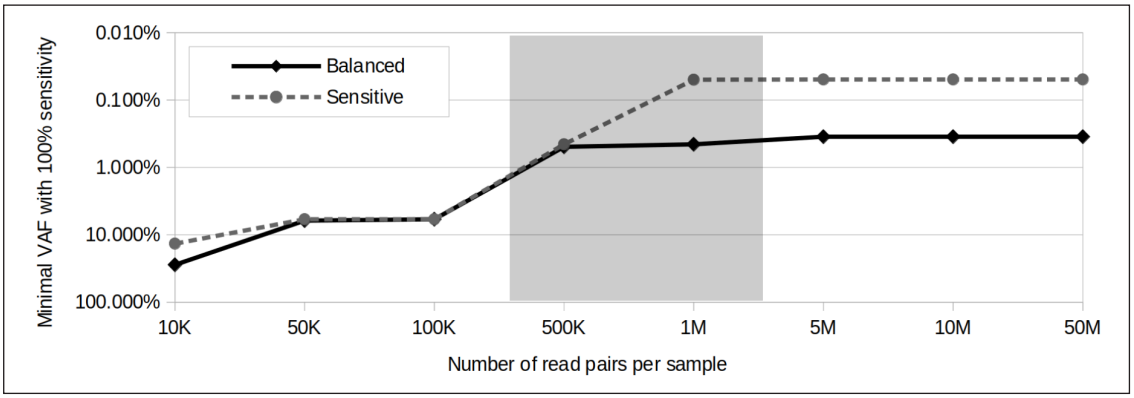

Figure 3. Minimal VAF with 100% sensitivity across different sequencing depths. *The plot demonstrates the saturation of sensitivity for both workflows at approximately 500K-1M read pairs. The shaded grey area highlights the recommended optimal sequencing range.*

Table 2. Theoretical vs. achieved VAF detection limits by sequencing depth.

Sequencing depth [read pairs]	Raw sequencing coverage	PCR duplicates [%]	Clean sequencing coverage	Theoretical VAF limit with 95% sensitivity*	Achieved VAF (Sensitive wf)	Achieved VAF (Balanced wf)
50M	320,031	96,56%	11,003	0,03%	0,05%	0,35%
10M	64,064	85,30%	9,401	0,03%	0,05%	0,35%
5M	32,032	75,16%	7,946	0,04%	0,05%	0,35%
1M	**6,406**	**40,97%**	**3,779**	**0,08%**	**0,05%**	**0,45%**
500K	3,203	26,70%	2,347	0,13%	0,45%	0,50%
100K	640	7,17%	594	0,50%	5,87%	5,87%
50K	320	3,69%	308	0,97%	5,87%	6,16%
10K	64	0,61%	64	4,70%	13,54%	27,88%

** Assuming that a single read is sufficient for variant detection and sequencing coverage is uniform. Poisson distribution was used to estimate 95% probability of obtaining at least one variant read from a given sequencing coverage.*

Conclusion

The combination of µCaler™ enrichment and VARAN analysis provides a robust framework for AML MRD monitoring. For optimal performance, we recommend a sequencing depth of 1M read pairs per sample, which provides the best balance of sensitivity and cost-efficiency. While 500K read pairs may suffice for the Balanced workflow (detecting ~0.5% VAF), the Sensitive workflow requires 1M read pairs to reach its maximum resolution of 0.05% VAF. It is important to note that the Sensitive workflow nearly touches the absolute technical limits of current hybridization-based NGS methodology; further increases in resolution are practically limited by the stochastic noise of the library preparation process itself.

References

Altium International s.r.o is the maker, owner and distributor of the analysis platform VARiant ANalyzer (VARAN). The company has a strong heritage acting as a trusted authorized distributor of software, equipment and chemical kits for applications in molecular biology and genomics including the product portfolio produced by Nanodigmbio.
altium.net/cz

VARiant ANalyzer (VARAN), is a web platform focused on evaluating genetic variations in the human genome from next-generation sequencing (NGS) data. The platform enables flexible data processing, interpretation of results, and expansion with advanced analyses for specific user needs.
variantanalyzer.eu

Nanodigmbio Pte. Ltd., is the manufacturer of complete series of products in targeted NGS sequencing, providing optimized kits for applications in precision medicine and research.