SureSeq

An Analytical Pipeline for UHS NGS MRD Assessment and Prediction of Relapse in AML

Harry Petch-Smith, Venu Pullabhatla, Bronwen Miller, Lucy Georgieva, James Reid, Jolyon Holdstock, Graham Speight Oxford Gene Technology (OGT).

A Sysmex Group Company

What binds us, makes us.

Introduction

Ultra-high sensitivity next-generation sequencing (UHS-NGS) enables the detection of variants at extremely low variant allele frequencies (VAFs),supporting highly sensitive measurable residual disease (MRD) assessment and monitoring.

As variant frequencies approach the limit of detection, background sequencing noise and technical artefacts can reduce confidence in distinguishing true variants from sequencing error.

To address this, we have developed an MRD-focused variant monitoring module for OGT's NGS Software Interpret, that leverages site-specific error modelling to characterise background noise, suppress false-positive calls, and improve confidence in variant detection.

Here, we demonstrate the ability of this approach to support reliable variant detection down to 0.01% VAF while maintaining high specificity.

Methods

Sample Cohort

A total of 84 technical replicates of germline NA12878 (from Coriell Institute) were processed using an UHS-NGS capture based workflow. 64 of these were used in error modelling , and the remaining replicates used for performance evaluation.

To assess variant calling performance, 67 contrived samples containing Revvity Mimix Myeloid Reference DNA were analysed. These samples contain 5 SNVs, 2 indels, and a 300 bp *FLT3*-ITD, and were serially diluted down to 0.005% VAF.

Dilution	Number of Samples
0	84
0.005	15
0.01	37
0.02	15

Table 1: Samples used during development and testing of the analytical pipeline.

Sequencing Panel

Sequencing was performed using OGT's SureSeq™ Myeloid MRD Plus NGS Panel, which incorporates 16 key genes relevant to AML research.

Gene	Exons	Gene	Exons	Gene	Exons	Gene	Exons
CSF3R	Exons 13-17	*KIT*	Exons 2, 8-11, 13 and 17	*FLT3*	Exons 13, 14 and 20	*RUNX1*	Exons 5-9
MPL	Exon 10	*NPM1*	Exon 11	*IDH2*	Exons 4 and 5	*CEBPA*	Exon 1
SF3B1	Exons 13-16	*JAK2*	Exons 12 and 14	*TP53*	Exons 2-11	*KRAS*	Exons 2 and 3
IDH1	Exon 4	*NRAS*	Exons 2 and 3	*CALR*	Exon 9	*WT1*	Exons 7 and 9

Table 2: The SureSeq Myeloid MRD Plus NGS Panel targets SNVs, indels and *FLT3*-ITDs in 52 hotspot exons across 16 genes relevant to AML research.

Base Specific Model Generation

For each genomic position and possible non-reference allele (SNV) in the panel, a statistical model was fitted to the observed VAFs across the negative control samples, characterising the underlying site- and substitution-specific error profile.

The resulting model parameters were used to derive position-specific limits of blank (LoBs) representing the lowest detectable VAF when comparing a test sample to the negative controls (see figure 1).

Variant Calling Workflow

Samples were processed using Interpret, aligned to GRCh38, and corrected using unique molecular identifiers (UMIs) before variant calling, with a family-size filter of 2.

Observed VAFs were assessed against site- and allele-specific error models; variants meeting the site-specific threshold ($p < 0.05$) were classified as true signals.

Minimum supporting read thresholds of three reads for SNPs and two for indels were set to further reduce false-positive calls.

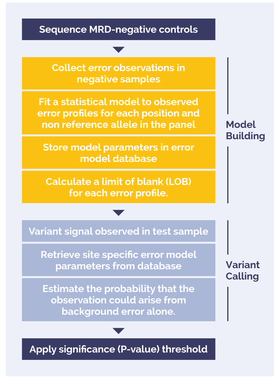


Figure 1: MRD variant calling workflow for low frequency detection.

Results I

Ultra-High-Quality Sequencing Coverage Supports Accurate Error Profiling

Negative control samples were sequenced to ultra-high depths; with 93% of panel positions achieving >80,000× coverage. This extensive and consistent sequencing coverage supported robust characterisation of site-specific error profiles across the panel, providing a reliable foundation for statistical error modelling and downstream ultra-low-frequency variant detection (see figure 2).

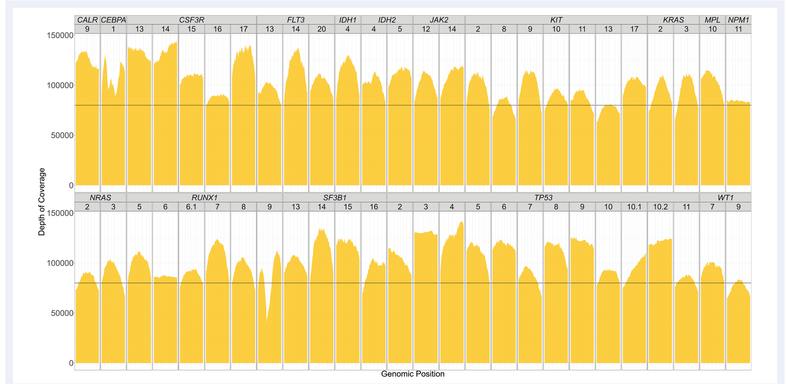


Figure 2: Average coverage profile of target regions in the SureSeq MRD Plus panel in the 64 negative samples used to build the error models. The horizontal line indicates 80,000x depth of coverage.

Results II

High Theoretical Sensitivity at 0.01% Across Targeted Regions

On average, the panel exons demonstrated a limit of blank < 0.01% in 92% of their genomic positions (see figure 3).

The highest LoBs were observed in *CEBPA* exon 1 and *RUNX1* exon 8, both GC-rich and repetitive regions. Elevated LoBs in these exons indicate increased background sequencing noise in the controls. Site specific error models capture this pattern, leading to better control of false positives arising from these challenging regions.

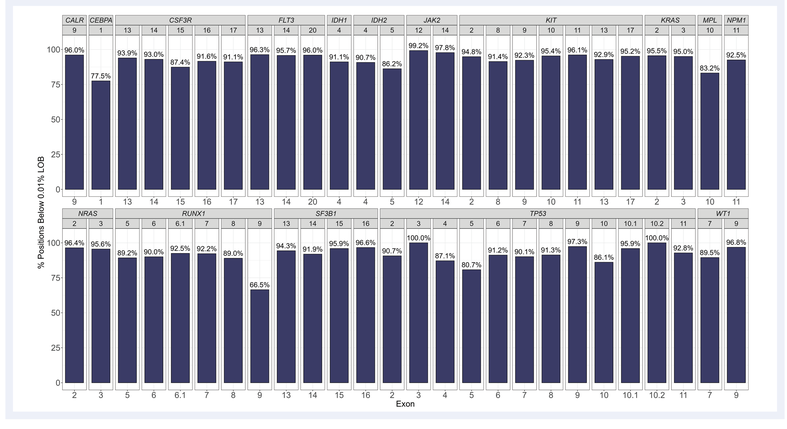


Figure 3: Percentage of targeted bases within each gene and exon achieving an average LoB < 0.01% VAF across the SureSeq Myeloid MRD Plus NGS Panel.

High Theoretical Sensitivity at 0.01% Across Hotspot Variants

Limits of blank (LoBs) were established for the Revvity Mimix variants, demonstrated high theoretical sensitivity, with a mean LoB of 0.0076% across the assessed SNVs. Clinically relevant hotspot variants, *FLT3*:c.2503G>T and *IDH2*:c.515G>A, achieved LoBs of 0.005% and 0.009% respectively, both below the 0.01% threshold. These results support reliable detection of clinically relevant variants at ultra-low allele frequencies.

Mimix SNVs
FLT3:c.2503G>T
IDH2:c.515G>A
JAK2:c.1849G>T
SF3B1:c.2219G>A
TP53:c.722C>T

Results III

Error Modelling Reduces False Positive Variant Calls in Negative Samples

The reduction in false positive calls was assessed using dilutions of the Mimix standard (5 SNVs, *JAK2* del, *NPM1* dup and 300 bp *FLT3*-ITD) by calling variants with and without the LoB model. The false-positive rate decreased from 18.8% to 4.4% in 20 independent negative replicates when applying the LoB model, representing a 65% reduction in false-positive observations (see figure 4 for the reduction of false positive calls in left figure compared to the right figure).

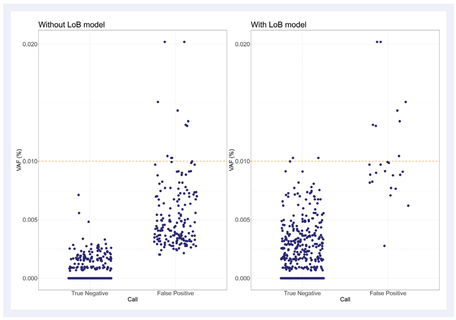


Figure 4: Mimix standard variants across all negative control samples. Specificity for 20 independent samples was 81.13% with no LoB model (left) and 95.59% when the LoB model was applied (right).

Error Modelling Maintains High Sensitivity While Improving Overall Accuracy

Error-model filtering resulted in a modest reduction in sensitivity, with the largest impact at the lowest dilution (0.005% VAF). Sensitivity decreased from 95.0% to 85.0% at 0.01% VAF and from 96.7% to 94.2% at 0.02% VAF with most false negatives occurring below 0.01% VAF (94.3%).

Assessment of cumulative sensitivity and specificity across observed VAFs, demonstrated improved overall accuracy following the application of the LoB model. Both metrics exceeded 97.5% at VAFs slightly below 0.01 (see figure 5).

Overall, site-specific error modelling substantially improved specificity while maintaining high sensitivity for clinically relevant ultra-low frequency variant detection.

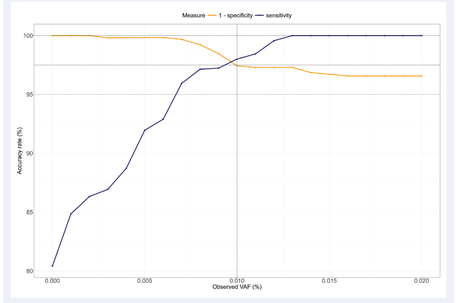


Figure 5: Sensitivity (blue) and specificity (yellow) after application of the LoB model across observed VAFs from 0-0.02%. Both metrics exceeded 97.5% at VAFs approaching 0.01%.

Conclusions

- Ultra-high sequencing depth and uniform coverage across the SureSeq Myeloid MRD Plus NGS Panel supported robust site- and allele-specific error modelling.
- Site-specific error models achieved a LoB of ≤0.01% VAF across 92% of panel bases, including known variants; most clinically relevant hotspots demonstrate LoBs below 0.01% VAF.
- Application of statistical error modelling reduced false-positive observations in MRD-negative samples by 65% while retaining high sensitivity for ultra-low frequency variant detection.
- Overall, integration of site-specific error modelling within Interpret improves confidence in variant detection at ultra-low allele frequencies and supports highly sensitive AML MRD monitoring.

Further information **contact@ogt.com | ogt.com | Oxford Gene Technology Ltd., Unit 5, Oxford Technology Park, 4A Technology Drive, Kidlington, Oxfordshire, OX5 1GN, UK**

SureSeq: For Research Use Only; Not for Diagnostic Procedures. © Oxford Gene Technology IP Limited – 2026.