APPLICATION NOTE

NGS Liquid Biopsy Solution for Highly Sensitive MRD Detection

INTRODUCTION

Liquid biopsies of blood or other bodily fluids have emerged as a minimally invasive alternative to traditional tissue biopsies. In the context of cancer management, analysis of cell-free DNA (cfDNA) in blood samples offers information regarding a tumor's genetic makeup, providing valuable insights into disease status and treatment response. Among the most promising applications of these analyses is the detection of minimal residual disease (MRD). MRD refers to the small number of cancer cells that may remain in the body after primary treatment. Early detection of MRD can be critical to informing treatment strategies, predicting relapse risk, and improving patient outcomes.[1,2]

IntegraGen, an OncoDNA group company specializing in genomic services, empowers researchers and clinicians with cutting-edge tools and expertise to advance translational research and personalized medicine for various diseases including cancer. IntegraGen offers a comprehensive suite of services, including high-throughput sequencing and sophisticated bioinformatic analysis.

Twist Bioscience, a partner of IntegraGen, leverages a proprietary semiconductor-based synthetic DNA manufacturing process to develop highly customized NGS panels and efficient library preparation kits that can be seamlessly employed in various genetic sequencing workflows, including MRD detection. Twist Custom Panels, alongside Twist library preparation and target enrichment kits, enable effective and high-throughput sequencing even with low DNA input.

This application note describes a comprehensive MRD analysis workflow developed by IntegraGen using various Twist Bioscience products. It highlights the performance and utility of using a targeted NGS-based approach to MRD detection.

APPLICATION SPOTLIGHT

- **Ultra-sensitive MRD detection:** Achieved a detection limit of 0.003% ctDNA using an NGS-based workflow.
- **High-efficiency workflow:** >75% library conversion across samples with efficient multiplexing.
- **Customized, patient-specific panels:** Targeting up to 119 tumor-specific variants per patient.
- **Reliable and validated results:** Leveraged UMI-based duplex sequencing for high specificity and error correction.

METHODOLOGY

Panel Design

To design the enrichment panels, whole exome sequencing was performed on germline and tumor samples of eight patients. Using IntegraGen's proprietary Mercury software solution, relevant tumor variants (single nucleotide variants (SNVs), substitutions, and indels with a maximum length of 5 bp) were selected based on specific filters regarding somatic score, tumor variant allele frequency (VAF), and low population frequency. Selected tumor variants were cleaned for clonal hematopoiesis and sequencing artifacts. Depending on the patient, 24 to 119 tumor variants were selected. There were 40 polymorphic single nucleotide polymorphisms (SNPs) added for sample identification, and 42 variants from reference samples (30 SNVs of each mutation type, 5x6 A>G/C>T/A>C/A>T/G>C/G>T and 12 small indels) for assessment of analytical capacities with the Twist cfDNA Pan-Cancer Reference Standards v2 (**Figure 1**).

Validation

Each tumor variant was validated by comparing the frequency of the detected variant to that of unexpected variants from the 40 sample ID regions (around 5000 genomic positions) corresponding to background sequencing noise. For each plasma sample, a p-value was calculated using a binomial probability distribution. For plasma samples with at least two detected variants having a p-value $<10^{-3}$ when compared with background error rate (i.e. $<10^{-6}$ for at least two variants) were considered circulating tumor DNA (ctDNA) positive.[3] ctDNA quantification was performed by counting sequencing reads carrying tumor variants among all sequencing reads overlapping the genomic coordinates of the tumor-specific SNVs identified in the patient's tumor tissue.

Limit of Detection

One of the key objectives of this experiment was to precisely define the Limit of Detection (LoD) that can be achieved with this workflow, along with the associated Limit of Blank (LoB). The concept of LoD for an MRD test can be represented probabilistically according to the Poisson distribution, considering the input cfDNA in genome-equivalent copies (GE), the threshold ctDNA fraction (VAF), and the number of selected tumor variants (**Figure 2**).

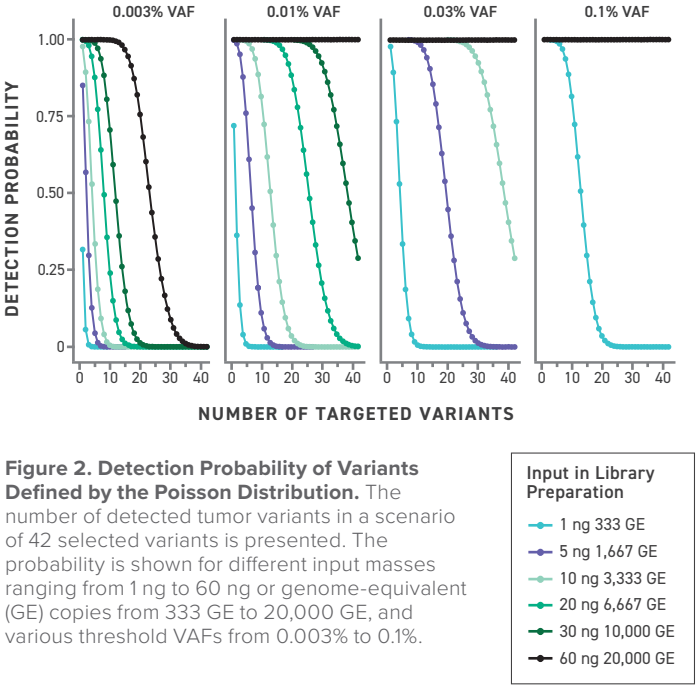


Figure 2. Detection Probability of Variants Defined by the Poisson Distribution. The number of detected tumor variants in a scenario of 42 selected variants is presented. The probability is shown for different input masses ranging from 1 ng to 60 ng or genome-equivalent (GE) copies from 333 GE to 20,000 GE, and various threshold VAFs from 0.003% to 0.1%.

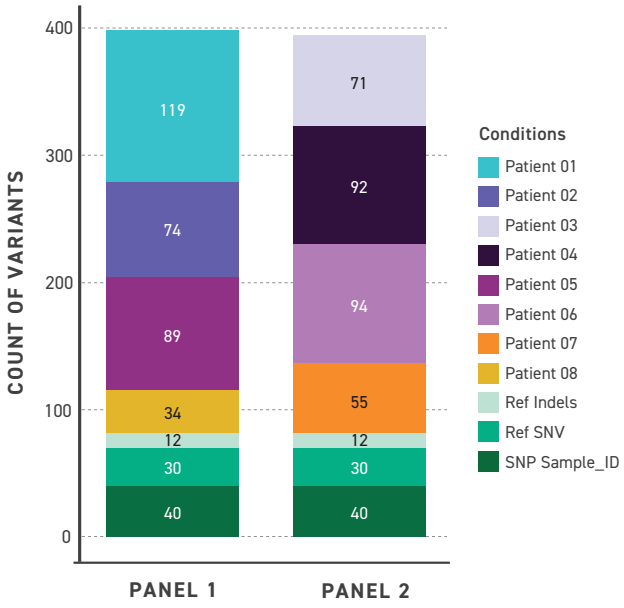


Figure 1. Overview of the Two Customized Twist Rapid MRD 500 Panels.

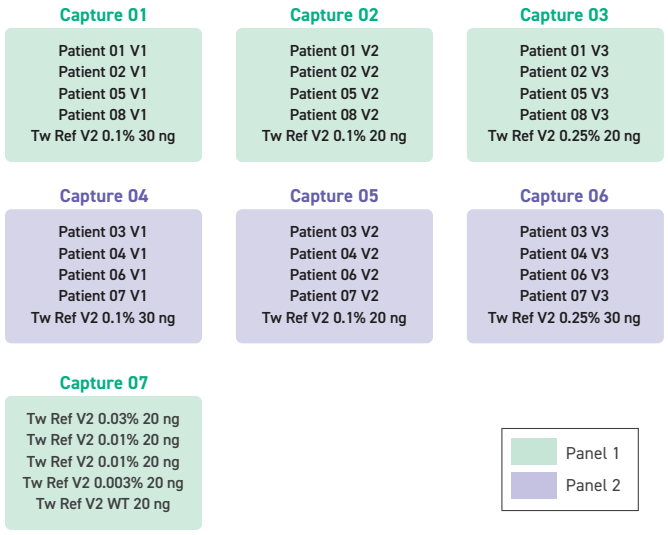


Figure 3. Overview of the Seven 5-plex Target Enrichments.

MRD Workflow

For MRD analysis, 24 samples were taken from eight patients with digestive cancer at three time points: V1 (before treatment), V2 (4 weeks after treatment), and V3 (after treatment and relapse). Each 3 ml plasma sample had cfDNA extracted using the QIAamp Circulating Nucleic Acid Kit. Additionally, tumor DNA extracted from FFPE tissue samples and germline DNA isolated from PBMCs were also collected. Library preparation and target enrichment were performed with the Twist cfDNA Library Preparation and Hyb Mix Kit using the manufacturer's recommendations. cfDNA input into library preparation varied from 6-30 ng (20 ng average) and library input into target enrichment was 400 ng for each patient sample. Two different custom Twist MRD Rapid panels were used for target enrichment (Panel 1 for four patients and Panel 2 for another four patients; **Figure 3**). The panels were designed to cover tumoral SNVs and small indel variants.

Six 5-plex target enrichment captures were performed based on patient group (Panel 1 vs Panel 2) and sample time point (**Figure 3**). Each capture included a control (Twist cfDNA Pan-Cancer Reference Standard v2 0.1% and 0.25%). A seventh 5-plex enrichment capture was performed with the diluted Twist cfDNA Pan-Cancer Reference Standard v2 control (0.1% in WT) for LoD assessment of the method. All enriched libraries were then sequenced 2x150 on an Illumina NovaSeq X. For variant calling, data processing involved molecular barcode management with fgbio (v. 2.3.0), alignment with BWA (v. 2.2.1), and variant calling using samtools mpileup (v. 1.9). FASTQ files were downsampled to a maximum of 150 million reads. UMI error correction used a molecular collapse methodology. For more details regarding the analysis methodology please see the publication here.[3]

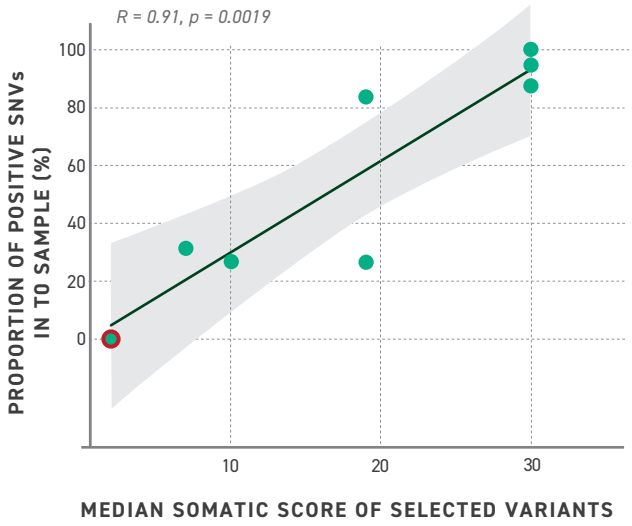


Figure 4. Proportion of Found Variants in the Eight Plasma V1 Samples as a Function of Median Somatic Score of Selected Variants From FFPE Tumor Tissues (red circle for Patient 06).

RESULTS

Library Preparation

Library conversion rates were >75% for the vast majority of samples (only the sample for Patient 06 at time point V2 had a low conversion rate of 32%), showing that the Twist cfDNA Library Preparation and Hyb Mix Kit generated optimal quantities of pre-capture libraries (**Table 1**). The enriched libraries had an average on-target rate of 72%. After sequencing was completed, each library generated an average of 140 million paired-end reads.

Variant Detection

Tumor variants were detected in all patients' plasma and at all follow-up points, except for Patient 06, due to poor quality of the tumor DNA sample. Thus, for Patient 06, no variant calling could be performed at time point V1 and V2. For Patient 06's plasma sample at V1, the 6 tumor variants detected didn't reach the validation threshold of at least 2 variants detected with a p-value $<10^{-3}$ (**Figure 4**). Patient 06's plasma sample at V2 had low library diversity, likely due to hemoglobin inhibition of enzymatic reactions. The tube containing the library had red coloration throughout the entire library preparation process.

Additionally, for Patient 03 at time point V3, sample identification QC was performed by genotyping 40 SNP IDs and it was determined that the plasma in this tube was discordant with the original germline. Therefore, due to a potential sample handling issue prior to lab acceptance, variant calling could not be performed for Patient 03 at time point V3

Except for some plasma samples derived from Patient 06 and Patient 03 as described above, the VAF levels of specific SNVs were detected in all patients and time points (**Figure 5**). Additionally, ctDNA detected in patient plasmas had proportions ranging from 0.005% to 26.59% (**Figure 6**). The results are consistent with the expected clinical trend for V1 before treatment, V2 after 4 weeks of treatment, and V3 after relapse. Note that, unlike the others, Patient 08 showed an increase in detected ctDNA and VAFs at V2 (this observation was confirmed by ddPCR). In this case, the treatment may not have delivered the expected results.

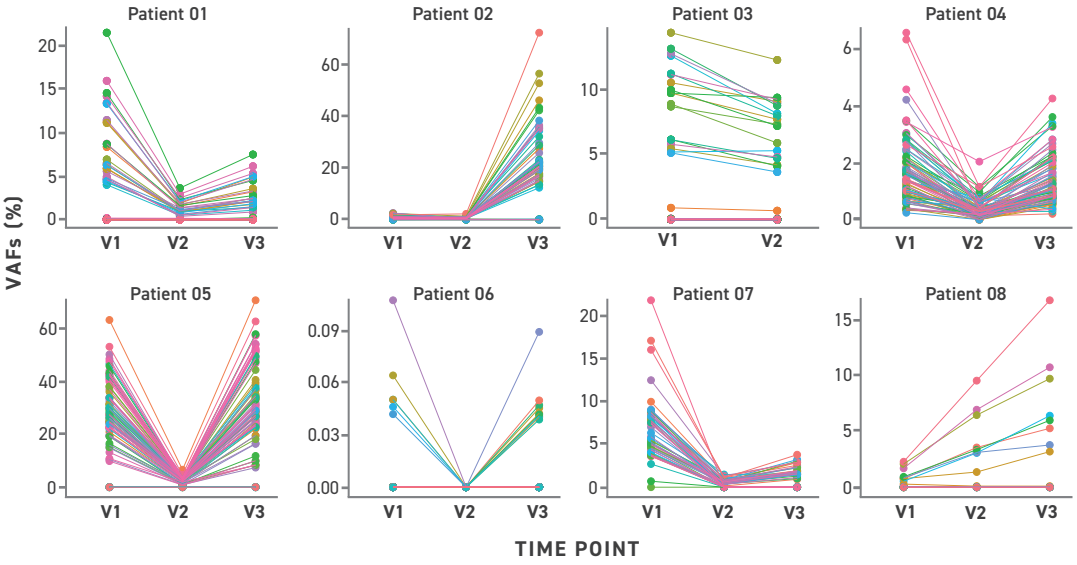


Figure 5. Percent VAFs for Specific SNVs in Each Patient Over Time.

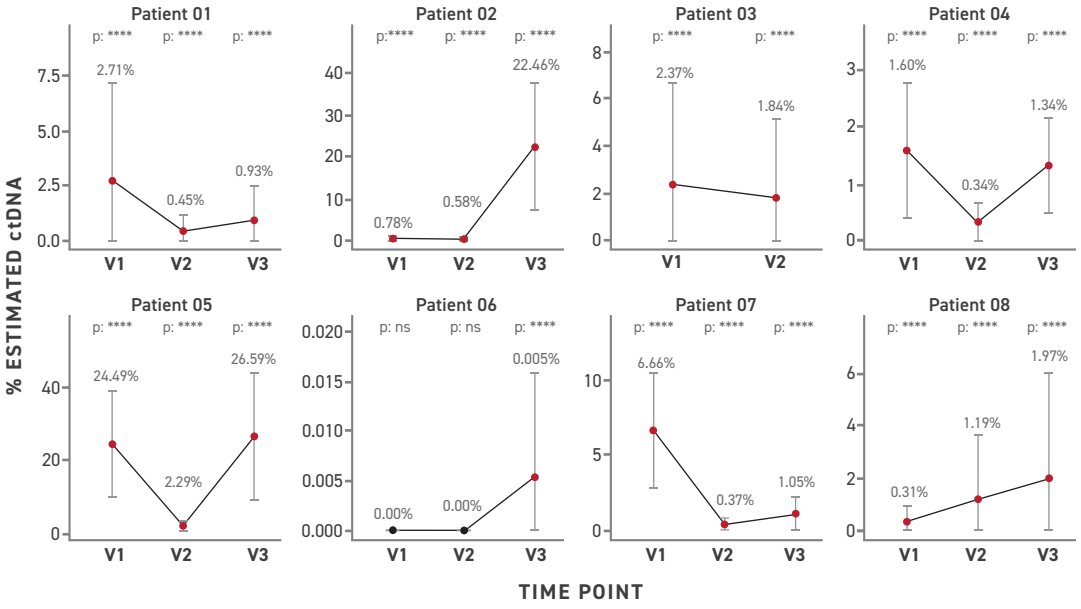


Figure 6. Percent ctDNA Detected in Each Patient Over Time. Error bars correspond to standard deviation of tumor VAF. Red points indicate statistically significant values ($< 10^{-6}$) for ctDNA detection, while black points are not significant.

**** denotes a $p<10^{-6}$;
ns denotes non-significant.

Limit of Detection and Limit of Blank Assessment

For LoB, frequencies of sequencing errors are calculated with molecular consensus analysis on the 40 regions designed for Sample ID outside the central polymorphic SNP +/- 60 bp (i.e., 40 x 119 = 4760 positions). Mean error frequency was $2x10^{-5}$ and there were 3 times more errors on C/G than on A/T when sequencing on the NovaSeq X. LoB was defined as the square of the 95% confidence interval for sequencing errors for each nucleotide type. To consider a sample as positive, at least 2 expected variants are required. Consequently, the LoB ranges from 6.3×10^{-7} to 2.2×10^{-5} depending on the nucleotide bases analyzed (**Figure 7**).

For LoD, variant calling and ctDNA quantification were performed using the method described above. From the 20 ng input (as well as 30 ng, not shown), all Twist Ref samples with VAF ranging from 0.003% to 0.1% were classified as mutant DNA (mutDNA) positive (**Figure 8**). An LoD of 0.003% was achieved, enabling the significant detection of 4 tumor variants out of the 42 included in the panel. These results were consistent with the Poisson distribution described in the methodology section.

Furthermore, the absence of detection in the WT reference sample confirmed the optimal specificity of the test and the LoB.

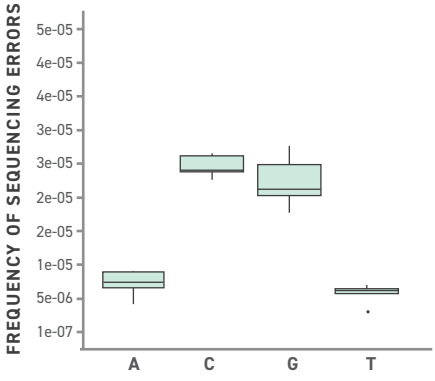


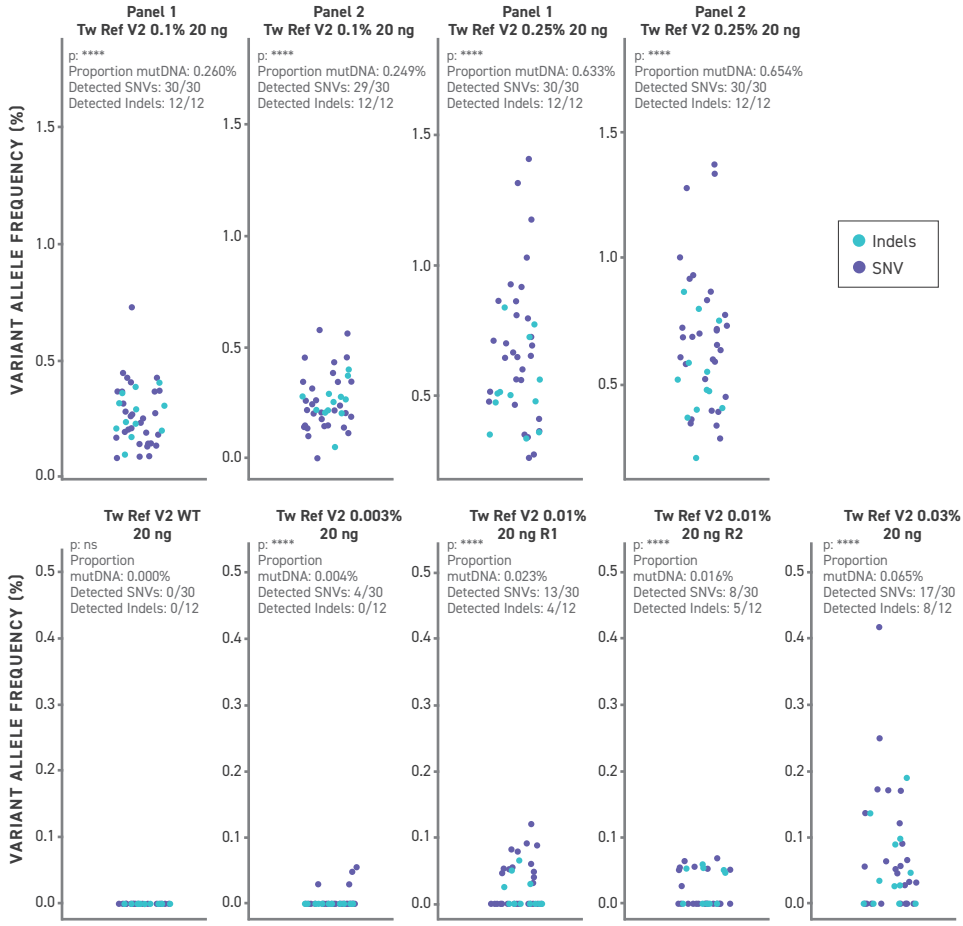
Figure 7. Frequency of Sequencing Errors on A / C / G / T

PANEL	CAPTURE	PATIENT ID	NG INPUT (ng)	LIB QTY (ng)	LCR	# READS
PANEL 1	Capture 1	Patient 01_V1	30	2969	>75%	117 895 051
	Capture 1	Patient 02_V1	14	1580	>75%	124 761 622
	Capture 1	Patient 05_V1	8	944	>75%	134 007 560
	Capture 1	Patient 08_V1	20	2100	>75%	120 520 617
	Capture 1	**Twist Ref V2 0.1%**	**30**	**2683**	**>75%**	**143 517 347**
	Capture 2	Patient 01_V2	30	3171	>75%	175 710 661
	Capture 2	Patient 02_V2	17	1613	>75%	178 795 461
	Capture 2	Patient 05_V2	12	1937	>75%	162 562 441
	Capture 2	Patient 08_V2	14	1204	>75%	189 320 321
	Capture 2	**Twist Ref V2 0.1%**	**20**	**2090**	**>75%**	**182 668 907**
	Capture 3	Patient 01_V3	30	2879	>75%	145 202 033
	Capture 3	Patient 02_V3	28	2792	>75%	163 604 895
	Capture 3	Patient 05_V3	30	3466	>75%	156 058 143
	Capture 3	Patient 08_V3	6	1346	>75%	141 609 238
	Capture 3	**Twist Ref V2 0.25%**	**20**	**2245**	**>75%**	**148 590 527**
PANEL 2	Capture 4	Patient 03_V1	22	2019	>75%	135 292 489
	Capture 4	Patient 04_V1	9	1277	>75%	133 037 941
	Capture 4	Patient 06_V1	10	1294	>75%	115 003 947
	Capture 4	Patient 07_V1	19	2264	>75%	123 865 021
	Capture 4	**Twist Ref V2 0.1%**	**30**	**2683**	**>75%**	**147 175 993**
	Capture 5	Patient 03_V2	30	3165	>75%	173 946 136
	Capture 5	Patient 04_V2	9	1356	>75%	143 710 610
	Capture 5	***Patient 06_V2***	***20***	***1101***	***32%***	***90 646 938***
	Capture 5	Patient 07_V2	15	1771	>75%	126 629 173
	Capture 5	**Twist Ref V2 0.1%**	**20**	**2090**	**>75%**	**137 438 617**
	Capture 6	Patient 03_V3	30	2996	>75%	138 306 535
	Capture 6	Patient 04_V3	15	1921	>75%	152 423 396
	Capture 6	Patient 06_V3	13	1278	>75%	136 359 836
	Capture 6	Patient 07_V3	22	2286	>75%	152 907 770
	Capture 6	**Twist Ref V2 0.25%**	**20**	**2245**	**>75%**	**139 708 558**
PANEL 1	Capture 7	**Twist Ref V2 0.03%**	**20**	2285	**>75%**	189 904 579
	Capture 7	**Twist Ref V2 0.01%**	**20**	2332	**>75%**	207 279 563
	Capture 7	**Twist Ref V2 0.01%**	**20**	2314	**>75%**	183 340 920
	Capture 7	**Twist Ref V2 0.003%**	**20**	2416	**>75%**	179 275 874
	Capture 7	**Twist Ref V2 WT**	**20**	2219	**>75%**	166 892 260

Table 1. Library Preparation Metrics for Each Sample.

Figure 8. Variant Detection Results and mutDNA Calculation for 20 ng Twist Ref From 0.003% to 0.25% VAF.

**** denotes $p<10^{-6}$;
ns denotes non-significant.



CONCLUSION

This MRD workflow, developed and provided by IntegraGen, is highly efficient and effective. The results and data presented here demonstrate that IntegraGen's MRD workflow is capable of identifying and detecting many personalized gene variants in a high-throughput manner. Two custom DNA panels were developed to target unique mutations specific to eight FFPE tissue samples. In total, the two panels together targeted 792 gene variants.

Using Twist Bioscience's high-quality DNA synthesis technology, highly customized and specific DNA panels can be developed to target a large number of personalized gene variants. The Twist cfDNA Library Preparation and Hyb Mix Kit used in this workflow has a very high library conversion rate and allows for accurate MRD analysis even with low cfDNA mass input. The proof of concept for this workflow established an LoD at 0.003% using Twist cfDNA Pan-Cancer Reference Standard v2 as controls, with duplex sequencing through the use of the Twist UMI Adapter System enabling the necessary sensitivity and specificity to detect low variant allele frequencies. Furthermore, validation on clinical samples confirmed the ability to reliably detect tumor variants in all tested patients and at follow-up time points, in concordance with clinical observations and ddPCR results.

The capacity for multiplex capture and sequencing allows for parallel analysis of multiple samples, facilitating rapid turnaround times for critical clinical information. Furthermore, the data analysis pipeline described here has undergone rigorous validation and the variant calling results attest to its accuracy and capability in providing timely clinical information. Overall, this MRD workflow, with its various high-quality components, offers a robust solution for MRD monitoring and analysis. When available, this MRD workflow, implemented by IntegraGen, a subsidiary of the OncoDNA Group, can be found under the brand name OncoFOLLOW®.

The Twist cfDNA Library Preparation and Hyb Mix Kit is for research use only.
This product is not intended for the diagnosis, prevention, or treatment of a disease or condition. Twist Bioscience assumes no liability regarding use of the product for applications in which it is not intended. The results are specific to the institution to which they were obtained. The results presented are customer-specific and should not be interpreted as indicative of performance across all institutions.

REFERENCES

1. Honoré, N. et al. Liquid Biopsy to Detect Minimal Residual Disease: Methodology and Impact. Cancers (Basel) 13, 5364 (2021). https://doi.org/10.3390/cancers13215364

2. Chen, J. et al. Measurable residual disease (MRD)-testing in haematological and solid cancers. Leukemia 38, 1202–1212 (2024). https://doi.org/10.1038/s41375-024-02252-4

3. Ben Sassi, M. et al. Improved Tumor-Type informed compared to Tumor-Informed Mutation Tracking for ctDNA Detection and Microscopic Residual Disease Assessment in Epithelial Ovarian Cancer. Preprint at https://doi.org/10.21203/rs.3.rs-6031886/v1 (2025).