MRD Express: Rapid, Scalable, and High-Performance Custom Target Enrichment for Minimal Residual Disease Monitoring

Sean Tighe, Michael Bocek, Shawn Gorda, Rebecca Barnard, Elian Lee, and Siyuan Chen

Abstract

Detection of minimal residual disease (MRD) via liquid biopsy offers a high-sensitivity alternative to conventional radiographic imaging, enabling early detection of molecular relapse within clinically actionable decision windows. To achieve the necessary analytical sensitivity, MRD assays require tracking patient-specific circulating tumor DNA (ctDNA) variants using highly optimized, personalized NGS panels. However, the clinical utility of personalized MRD monitoring is often constrained by the complexity of panel design and lengthy manufacturing lead times.

We present the Twist Bioscience MRD Express Panel, a scalable target enrichment workflow engineered to streamline the transition from variant characterization to repeat analysis across multiple samples. Leveraging a silicon-based DNA synthesis platform and proprietary design algorithms, an automated pipeline could achieve a rapid turnaround time of as little as one business day from synthesis to shipment. To ensure robust performance across challenging genomic regions, we implemented an innovative LNA (Locked Nucleic Acid) boosting strategy, enhancing coverage uniformity for probes with extreme GC content. Additionally, we assessed the feasibility of using LNAs to perform allele-specific targeting in the context of this capture system.

This platform supports ultra-flexible panel sizes ranging from 1 to 5,000 probes, allowing researchers to perform highly focused MRD tracking or broader exploratory research strategies. Fully compatible with existing Twist NGS library preparation and hybrid capture workflows, MRD Express provides a high-throughput, rapid approach to panel synthesis that maintains comparable performance metrics to traditional hybrid capture panels, and enables highly personalized low-frequency variant discovery research.

Probe design strategy

Our standard panel process requires time-intensive manufacturing steps prior to obtaining a complete panel. We asked instead whether direct incorporation of biotin-containing phosphoramidites with Twist's proprietary printing technology could reduce manufacturing times.

To address capture variation, we exploited the use of locked nucleic acids or LNAs. LNAs contain an oxymethylene bridge between the 2' and 4' sugar residues, and strongly stabilize Watson-Crick interactions, increasing Tm (Koshkin *et al* 1998). Previous experiments have used individual, short (~20nt) LNA containing probes to enrich for long genomic fragments (Ma *et al* 2013), but to our knowledge this is the first use of LNAs in a conventional NGS hybrid capture panel. We evaluated 2 strategies for LNA incorporation - either within low GC (<=40% GC) probes, or within all probes, as well as a no-LNA control. Our results demonstrate that placing LNAs at low GC probes can effectively rescue capture coverage in these regions, but that LNAs in high GC probes drive additional off-target (Figure 1A).

To more clearly understand how these improvements in uniformity would translate into a cancer panel, we compared a direct LNA-containing design against a standard capture panel developed against the variant sites in the Twist cfDNA Pan-Cancer Reference Standard v2. We obtained increased total depth from the new design (Figure 1B), and found an increase in coverage for low GC sequences (Figure 1C), including at important recurrently altered loci such as those in ROS1 (Figure 1D) and PTEN (Figure 1E).

As LNAs are commonly used for allele-specific targeting in other applications (Latorra *et al* 2003), we evaluated the allele bias of probes containing LNA residues targeted against either the reference or alternate alleles (Figure 1F), finding that LNAs produced a small additional effect compared to the large effect of allele bias (Figure 1G).

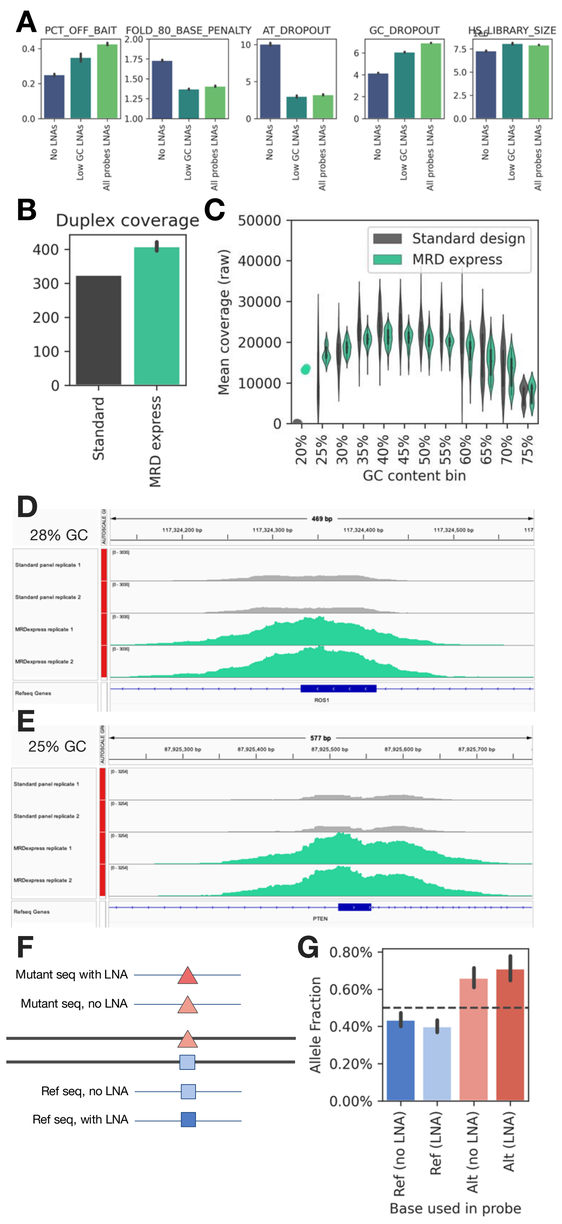

Figure 1: (A) Panel metrics obtained from three LNA placement strategies - No LNAs, LNAs placed only in low GC probes, and LNAs in all probes. (B) Comparison of total duplex molecule coverage between LNA-containing (MRD express) and non-LNA containing panels at 10,000x raw downsampling. (C) Effect of LNA incorporation and direct print on total coverage obtained across different GC bins for a capture panel targeting a number of recurrent cancer variants. Gray represents standard Twist panels, while green represents LNA-containing direct print (MRD Express) panels. (D) and (E) - genome browser views of two replicate captures at recurrent cancer-causing variants in ROS1 (D) or PTEN (E). (F) Schematic of LNA allele-bias experiment, and (G) Allele frequency bias obtained from reference and alternate probes, with and without LNAs on the Twist cfDNA Pan-Cancer Reference Standard v2 (0.5% VAF sample)

Comparison of tiling strategies

Total, non-redundant, on-target sequencing coverage is a critical factor in the overall sensitivity of a minimal residual disease monitoring assay. Increased coverage allows for greater statistical power, and lowers the potential limit of detection (LOD) for the assay. To this end, we sought to define a probe tiling strategy that would maximize unique coverage in a simulated MRD sequencing assay.

We used an internal capture panel developed against the variant sites in the Twist cfDNA Pan-Cancer Reference Standard v2. We evaluated four potential probe designs - either single or double-stranded probes (ssDNA vs dsDNA), and either probes placed directly over the variant sites, or "staggered" probes that target 30bp up- and downstream of the site (Figure 2A), as well as a standard panel design. Our results indicate that staggered designs, and particularly staggered designs with both strands present, dramatically expand the available unique sequencing depth, for total reads, unique reads, and duplex consensus reads (Figure 2B).

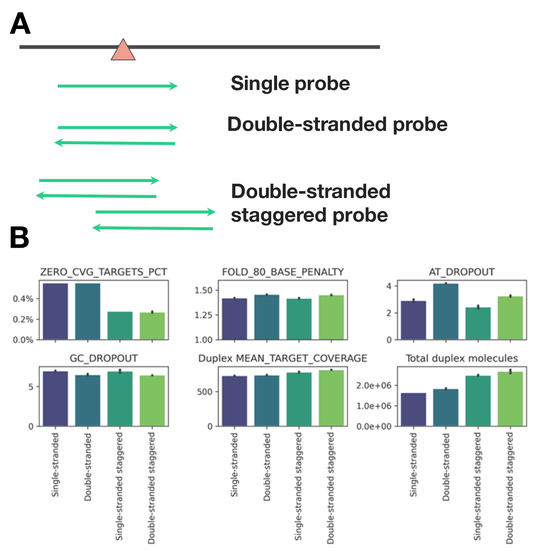

Figure 2: (A) Summary of differential expression experiment design. (B) Panel metrics collected from different design strategies

Our results demonstrate a vastly improved yield, both for total reads, and for duplex reads using a "staggered" strategy compared to using single probes. We suggest that "staggered" designs should be adopted in most MRD experiments using hybrid-capture. The primary trade-off at stake is bait efficiency, but if sequencing depth is not a limiting factor, these designs should demonstrate superior performance.

Variant detection in a simulated MRD experiment

We next asked how MRD Express probes compared to standard probes in a simulated MRD experiment. Using the Twist cfDNA Pan-Cancer Reference Standard v2, we created a series of diluted standards encompassing 1%, 0.5%, 0.2%, 0.1%, 0.01%, 0.005% (50 ppm), 0.001% (10 ppm), or 0% (WT) variant allele frequencies (VAFs). Total mass was 15g (5000 genome equivalents), meaning that all conditions below 0.02% contain less than one mutant molecule per site on average. Standards were captured with both panel configurations, and sequenced to high depth (80,000x) over the targets. After duplex UMI collapse, we ascertained the reference and alternate allele counts at each site.

MRD Express probes have comparable sensitivity to standard Twist capture probes, with mild improvement at very low GC sequences (Figure 3A). Pooling across capture conditions, this configuration effectively distinguishes between 10 ppm and blank (Figure 3B, p=0.001, one-sided exact permutation test). We also assessed the different rates of recall at SNPs and indels, finding that the two panel types perform comparably on both variant types in terms of recall (Figure 3C).

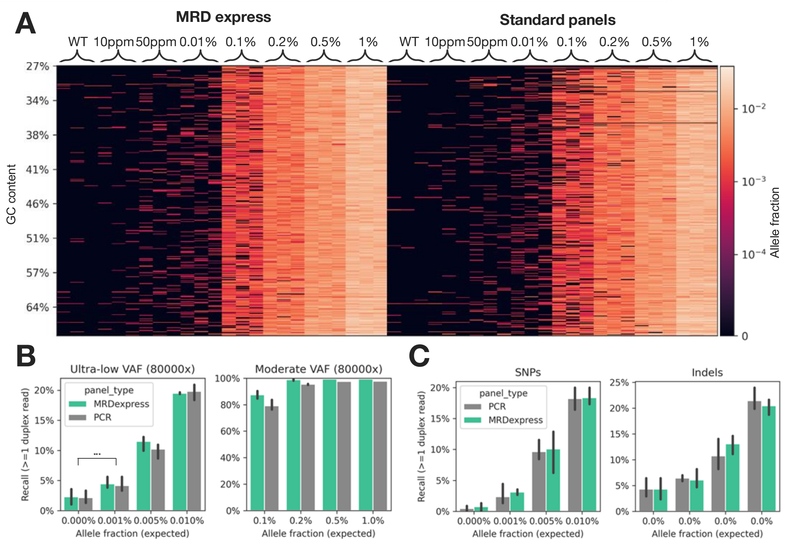

Figure 3: (A) - heatmap of detection events across all samples in the sequencing experiment. Rows represent sites sorted by nearby GC-content, while columns represent samples. VAF and panel type indicated above, and per-site alelle fraction at right. (B) Overall recall among all sites for MRD express or standard (PCR-based) panels. (C) Estimated recall for SNPs and Indels across different panel types.

Modeling the effect of distinct sites on recall

One advantage of large hybrid capture panels (compared to multiplexed amplicon panels, for example) is that they permit for a much larger number of sites to be interrogated. In principle, this allows us to look at a larger number of potential variant sites, and - in doing so - maximize the number of potential observations for rare variants in an MRD experiment. In addition, in principle a large number of potential variants allows for decisions to be made to prioritize site choice, if some sites have characteristics (i.e. GC content, hompolymers) that make them undesirable.

In principle, the flexibility that our LNA-based design provides should enable the user to target a larger number of sites than previously available (particualrly in low-GC sequences). However, we wanted to test what the expected analytical sensitivity to smaller number of sites would be. To do this, we simulated downsampling the number of available target sites (from the 396 total). Our results show that increased sites leads to more robust tumor fraction estimates (Figure 4A), with similar results for both panel types. Combining across types, larger numbers of samples that can be cleanly separated from the highest blank sample (Figure 4B), and greater statistical power in distinguishing the sample from blank (Figure 4C).,

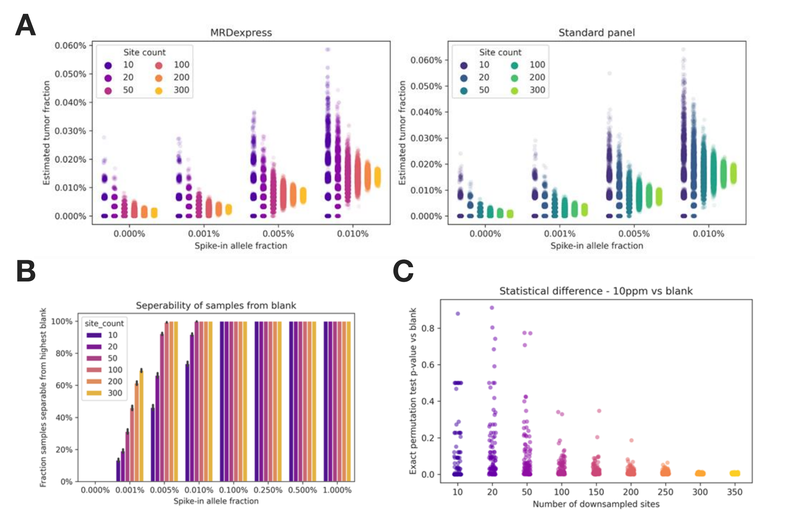

Figure 4: (A) Tumor fraction estimated from 1,000 randomly simulated draws of sites from the population of sequenced sites, for MRD express (left) or Standard (right) panels. Sites were filtered to avoid those which were recurrently detected across multiple negative controls. (B) Fraction of simulated draws where the number of detected sites exceeds the maximum number detected in the blank (0%) sample - again excluding recurrently detected negative sites. (C) P-values distinguishing between 10ppm and blank (0%) samples for different numbers of sampled sites (one-sided exact permutation test).

Turn around time and rapid synthesis

Existing methods for rare variant profiling are limited either in their sensitivity (amplicon-based methods), or their total required synthesis time (standard hybrid capture-based methods). MRD express panels can profile hundreds of variant sites simultaneously (Figure 4B), including those at extremes of GC content (Figure 1B, 1C, and 1D), offering both high sensitivity (~10 ppm, figure 3B) and rapid turnaround time (~24 hours).

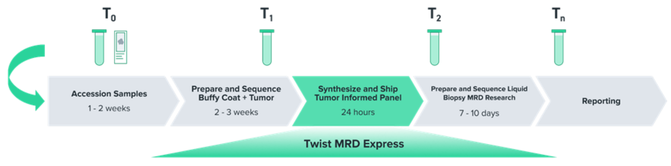

Figure 5: (A) Schematic of total workflow for an MRD experiment, including time for panel order (highlighted in green).

Materials and methods

Sequencing samples were prepared using the Twist cfDNA Pan-Cancer Reference Standard v2 control material, diluted to the specified concentration using additional 0% (WT) DNA from the standards. Libraries were created using the Twist cfDNA Library Preparation Kit, using 15ng of input DNA. Capture was then performed using the Twist Standard Hybridization protocol, using either capture probes synthesized by Twist's standard method, or synthesized directly on the writer with LNA incorporation (MRD Express). Libraries were sequenced using an Illumina NextSeq 550, or in the case of the VAF detection series experiment, an Illumina NovaSeq X.

After sequencing, we downsampled each set of samples to 10,000x raw coverage (80,000x raw coverage for the VAF series experiment). UMIs were extracted with fgbio, and reads were then aligned to human reference genome build hg38 using BWA (Li and Durbin 2009).. Post alignment, UMI families were collapsed using fgbio, which was also used to call duplex consensus reads from the UMI information and mapping position. Consensus reads were again aligned to hg38 using BWA. Allele detections were tabulated using samtools mpileup, with additional post-processing and plotting using standard Python data science libraries. Metrics were calculated either as a part of duplex collapse using fgbio, or, in the case of capture metrics, were calculated using GATK CollectHsMetrics.

References

Alexei A. Koshkin, Sanjay K. Singh, Poul Nielsen, Vivek K. Rajwanshi, Ravindra Kumar, Michael Meldgaard, Carl Erik Olsen, Jesper Wengel,LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition, Tetrahedron 1998. 54(14)

Ma Z, Axtell MJ (2013) Long-Range Genomic Enrichment, Sequencing, and Assembly to Determine Unknown Sequences Flanking a Known microRNA. PLoS ONE 2013. 8(12)

Latorra D, Campbell K, Wolter A, Hurley JM. Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. Hum Mutat. 2003 Jul;22(1)

Li H, Durbin R, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, Volume 25, Issue 14, 15 July 2009, Pages 1754–1760,

Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017 Mar;27(3):491-499.

Disclaimers

Conflict of interest statement:
All authors are employees and shareholders of Twist Bioscience