

ROBUST VARIANT CALLING WITH SRSLY[®] USING THE HORIZON ONCOSPAN cfDNA REFERENCE STANDARD

INTRODUCTION

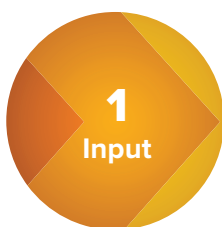
Cell-free DNA (cfDNA) in blood originates predominantly from apoptosis of varying blood cell lineages^{1,2}. In patients with cancer, a small fraction of the cfDNA known as circulating tumor DNA (ctDNA) originates from dying tumor cells³. Recently, cfDNA variant calling has emerged as a noninvasive next-generation sequencing (NGS) method to monitor disease progression and treatment efficacy.

Claret Bioscience has developed an easy and efficient NGS preparation method called SRSLY that features an optional unique molecular Identifier (UMI) add-on step for accurate detection of rare ctDNA variants⁴. UMIs are short molecular barcodes that are used to uniquely tag each molecule in a sequencing library to aid in the detection of single nucleotide polymorphisms (SNPs) and resolve polymerase chain reaction (PCR) duplicates^{5,6}. In analyses of NGS data from cancer samples, this allows true pathogenic variant alleles to be distinguished from variants introduced by copying errors anywhere in the library preparation and sequencing process.

Here we show that SRSLY library preparation can be used to detect common cancer variants at low allele fractions from varying sample input amounts. We use Horizon Discovery's OncoSpan cfDNA standard, a well-characterized genomic reagent that can be used for detection of variants with known variant allele frequency (VAFs) and Twist Bioscience's custom Pan Cancer targeted-enrichment Panel. Targeted enrichment allows for cost effective sequencing of the portions of the genome of clinical interest, including low frequency variants.

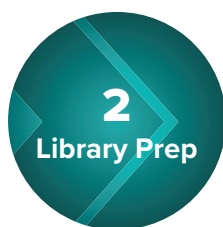
Highlights

- SRSLY is a unique NGS library prep kit from Claret Bioscience that renders all input DNA single-stranded prior to adapter ligation and captures all DNA strands even when nicked or damaged
- Here we showcase SRSLY's UMI add-on kit's ability to call low AF ctDNA variants using Horizon's OncoSpan cfDNA standard and Twist's PanCancer enrichment panel
- The Horizon OncoSpan cfDNA is a valuable tool for evaluation of NGS library preparation methods
- Compared to traditional dsDNA library preps SRSLY exhibits increased library complexity and similar UMI error rate and VAF calling sensitivity at all input amounts tested



horizon
a PerkinElmer company

Oncospan cfDNA
5-20ng



CLARETBIO

SRSLY[®] + UMI or
DS Prep + UMI



T W I S T
BIOSCIENCE

Pan Cancer Panel
(800kb)



Illumina[®]
2x150bp



UMI correction
Complexity
SNP Calling

METHODS

Library Preparation and Sequencing

The Horizon OncoSpan cfDNA standard is a well characterized cell line-derived reference standard. This standard contains >375 variants from over 152 key cancer genes present at between 1-100% variant allele fraction (VAF), 25 of which are verified by ddPCR. These verified variants include 8 indels and 17 SNPs. These 25 variants were analyzed in this application note.

Using the Horizon OncoSpan cfDNA standard, eight libraries were prepared for NGS. Three libraries were prepared using 5, 10, 20 ng of the cfDNA standard with Claret Bioscience’s SRSLY Library Prep kit with UMI Addition (Figure 2). For the same input amounts, libraries were also prepared with a traditional double-stranded library prep kits with UMIs (DS prep). The libraries were then pooled and enriched using a custom Pan-Cancer panel from Twist Biosciences® which spans ~100kb. Of the 25 variants found in the Oncospan standard, 21 are covered by the Twist enrichment panel. Enriched libraries were sequenced on an Illumina HiSeq X at a read length of 2 x 151 bp following manufacturer’s instructions to a depth of >20 million reads per library (average of x fold coverage for the 800Kb panel).

Informatic Processing

Sequencing data was converted to FASTQ format with bcl2fastq 2.20.0.422, reads were downsampled to 20 million reads per library, and adapters were trimmed with SeqPrep2. Reads were aligned to hg38 with bwa mem

0.7.15 for paired end reads. UMI tools 1.0.1 and was used to correct and group UMIs prior to duplicate marking with GATK PicardTools version 2.23.8. Correcting and grouping reads by UMI was also performed with fgbio 1.3.0 to determine UMI error rate. Target panel analysis was performed using GATK Picard 2.23.8, and variants were called using lofreq version 2.1.5. Figure 3 shows the pipeline inputs and outputs.

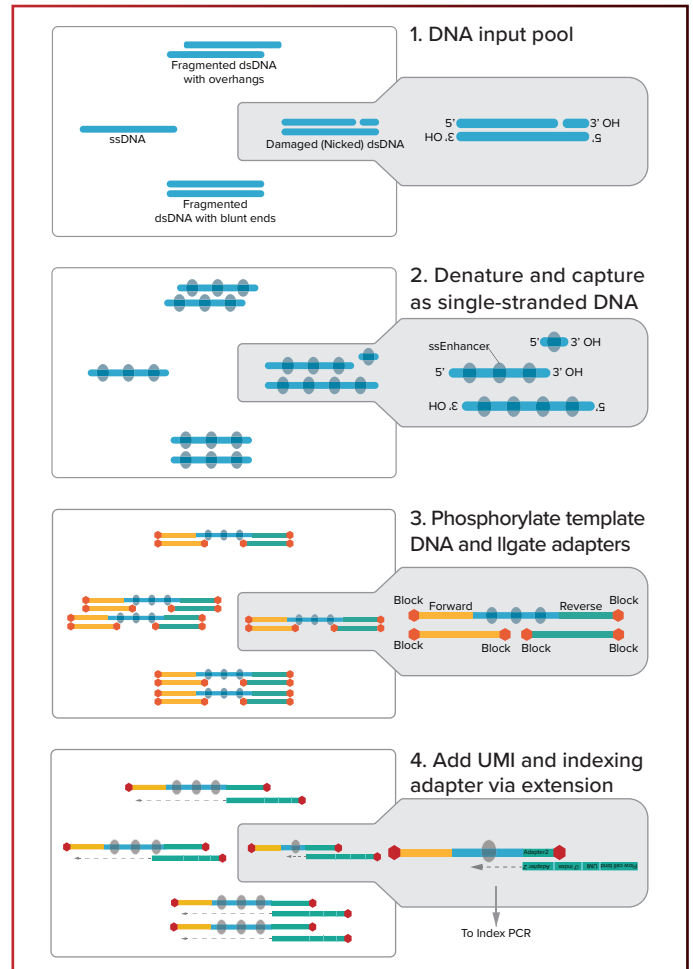


Figure 2. SRSLY library preparation protocol with UMI Addition. Libraries are generated in 4 simple steps (excluding Index PCR)

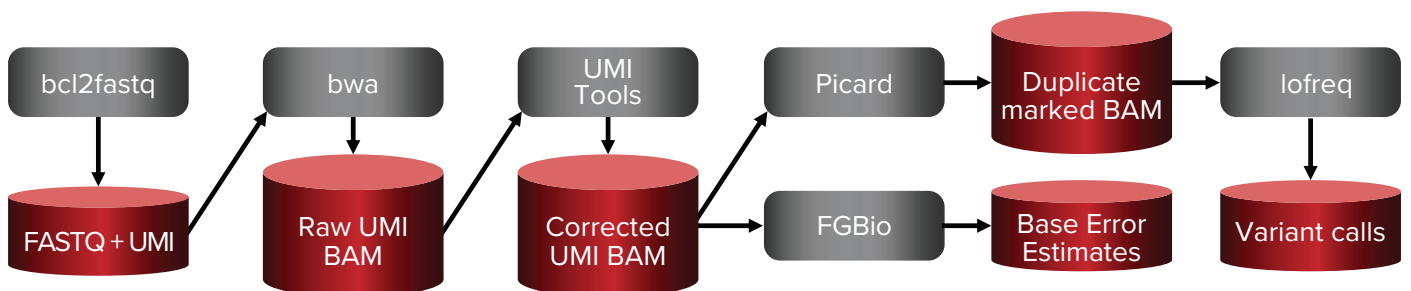


Figure 3 – Bioinformatic pipeline

RESULTS

Library complexity

To demonstrate the ability of SRSLY to perform in ctDNA variant analysis, we prepared UMI libraries from Oncospan cfDNA with a range of inputs representative of amount used in clinical analyses using both SRSLY and DS prep, followed targeted enrichment. The libraries were sequenced and processed by the bioinformatic pipeline described in Figure 3. GATK Picard Tools was used to estimate the number of unique molecules present in the library based on the number of paired end molecules observed and the number of unique pairs observed. Estimated library size increased with input amount for both prep methods, but SRSLY showed a marked increase at higher inputs, with approximately double the estimated library size of the DS prep at 10 and 20 ng of input (Figure 4).

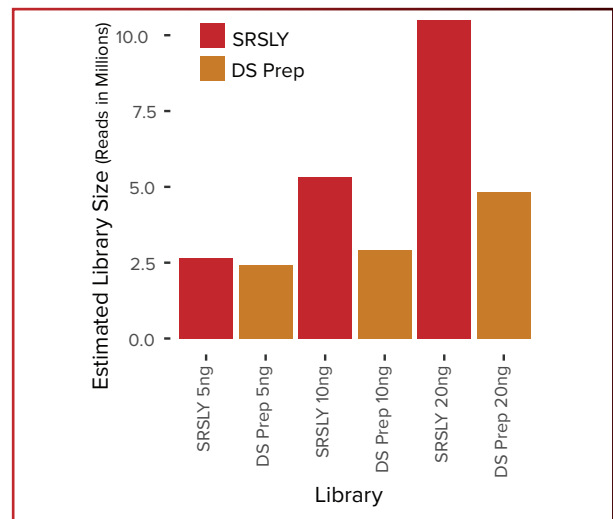


Figure 4 – Estimated library size from GATK Picard tools. Libraries at four different inputs were downsampled to 20 millions reads for analyses. Picard estimates library size based on the number of paired end molecules observed and number of unique pairs observed. The estimated library size is the estimated number of unique molecules present in the library.

Reads per UMI

UMI tools was used to group reads by UMI and mapping location to determine the number of reads sequenced for each UMI (Figure 5). At lower inputs, both SRSLY and DS preps have a higher proportion of UMIs with multiple reads per UMI, showing that higher cfDNA inputs are ideal for variant calling. However, the higher input libraries have more UMIs with lower number of reads per UMI. A lower number of reads per UMI implies a higher diversity of reads sequenced and therefore increased library complexity. For all inputs tested, SRSLY has a higher number of UMIs with only one read, and a lower maximum number of reads per UMI, due to increased library complexity (Figure 5).

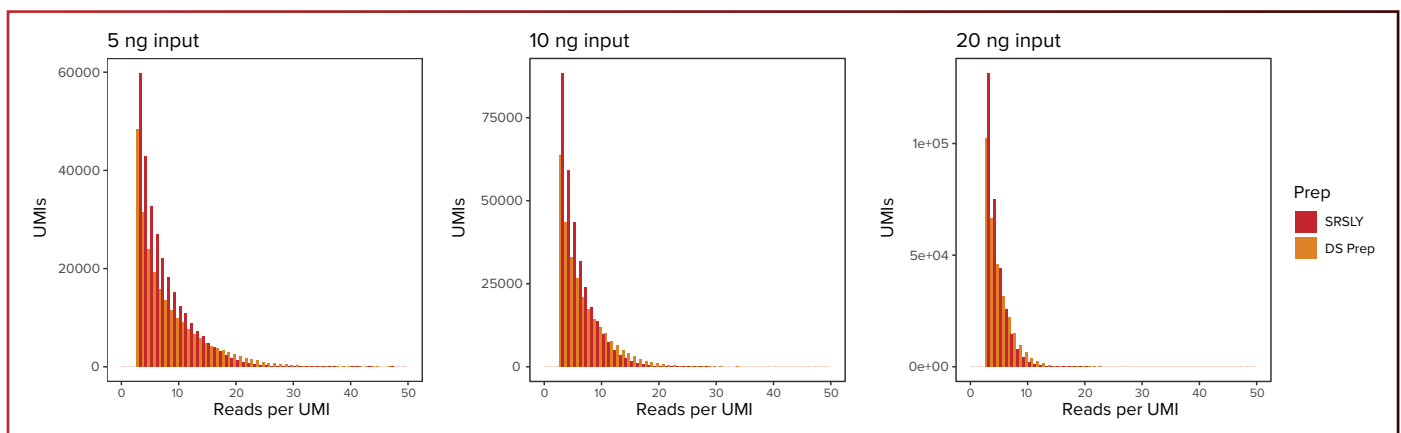


Figure 5 – Number of reads per UMI and the count of UMIs with that number of reads across prep method and input amounts. Reads per UMI is defined as the number of reads sequenced containing a given UMI, while count is the number of UMIs with that number of reads per UMI. A lower number of reads for a given UMI is preferred.

Error Rate

After UMI correction and grouping reads by UMI, FgBio was used to determine the consensus sequence for each UMI that had three or more reads. The mean error rate for each sample was then calculated from the error rates of the reads with the same UMI (Figure 6). At lower inputs, the mean error rates are roughly similar between SRSLY and DS prep, however at the two higher inputs SRSLY has a lower error rate.

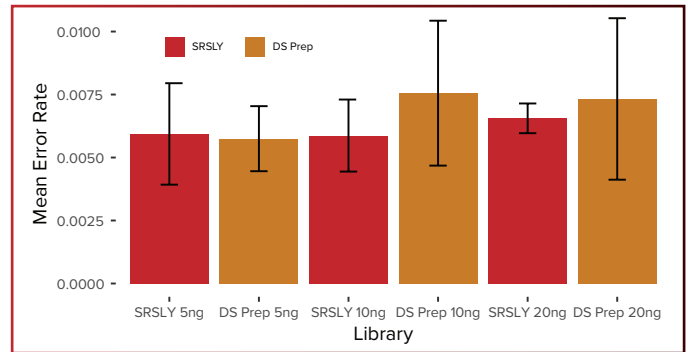


Figure 6 – Mean error rate from FgBio for each prep type and input level. Mean error rate is determined by first calling consensus sequences from reads with the same UMI, and then determining the error rate of each read from that consensus sequence.

Variant calling

After sequencing, mapping, and UMI correction, variants were called using lofreq v 2.1.5 and output in variant call format (VCF). VCFs from each sample were merged with the Horizon Gold Standard VCF containing the truth calls to determine accuracy (Figure 7). SRSLY was able to accurately call both SNVs and indels at all input levels. At every input, SRSLY had read data supporting all confirmed VAF variants in the Horizon standard, though not enough reads to support a call by lofreq. Of the 25 ddPCR verified variants in the Horizon standard, two variants with very low VAF are not called by lofreq in any sample for either SRSLY or DS prep. For the two variants that were not called in any sample by lofreq, VAF was calculated from the number of reads containing the variant aligned to the site and the total number aligned to that site. This method was also able to detect additional variants in the Oncospan standard that were not confirmed by ddPCR (not shown).

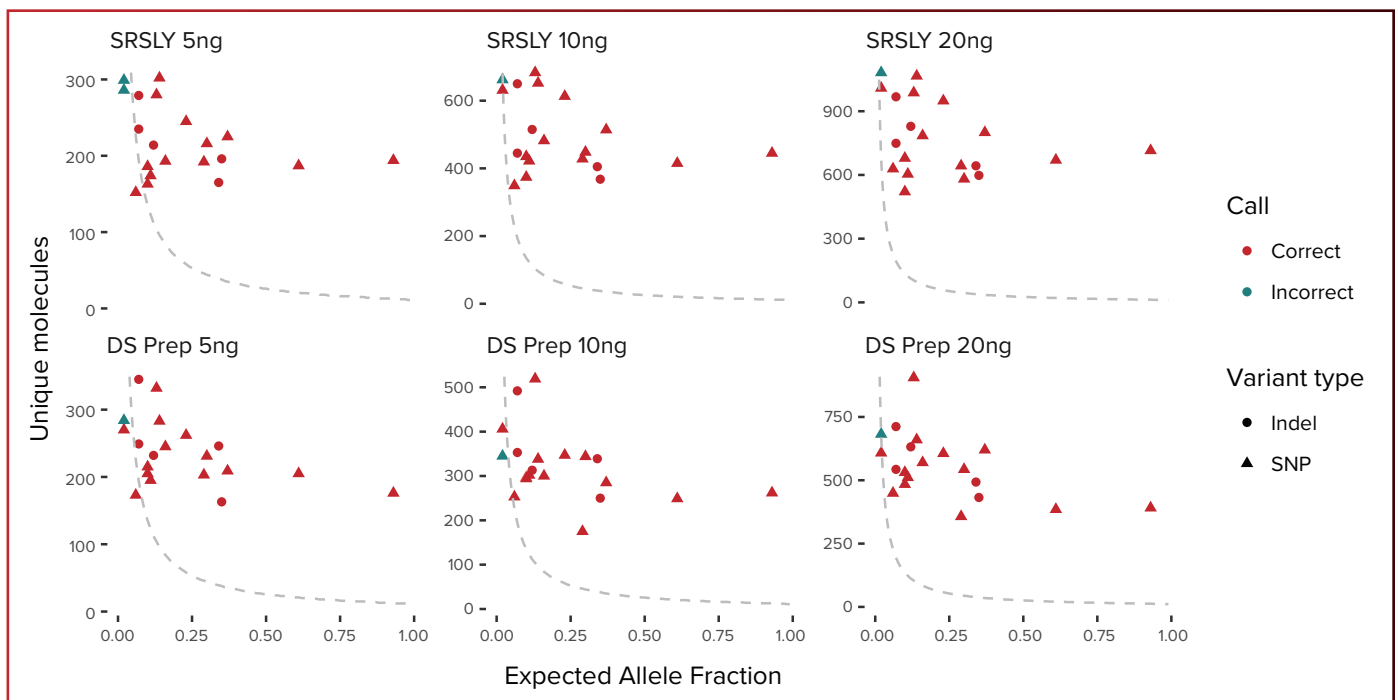


Figure 7 - Depth of coverage and variant allele fraction in the reference standard color coded by whether the variant is called correctly. Point shape indicates variant type. Depth of coverage is the total number of reads at the variant site regardless of whether they contain the variant. Variant allele fraction is provided by the horizon reference standard. The dashed grey line indicates, for each allele fraction, the number of unique molecules required to have 80% power to get sequence 10 reads with the alternate allele.

CONCLUSION

SRSLY's UMI add-on protocol demonstrates excellent performance in variant detection applications. Compared to traditional library preparations, SRSLY libraries have slightly higher complexity and better polymerase fidelity during PCR at higher inputs and provide similar variant calling performance at identical sequencing depths. These data also demonstrate that the Oncospan cfDNA can be used as standard to evaluate and deploy SRSLY for cfDNA-based variant calling in research use and clinical settings.

REFERENCES

1. Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin Chim Acta* 313, 139-142, doi:10.1016/s0009-8981(01)00665-9 (2001).
2. Stroun, M., Anker, P., Lyautey, J., Lederrey, C. & Maurice, P. A. Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol* 23, 707-712 (1987).
3. Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 11, 426-437, doi:10.1038/nrc3066 (2011).
4. Troll, C. J. et al. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics* 20, 1023, doi:10.1186/s12864-019-6355-0 (2019).
5. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* 2010, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).
6. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nature methods* 9, 72-74, doi:10.1038/nmeth.1778 (2011).



Claret Bioscience's SRSLY[®] NGS library preparation kits for Illumina[®] provide reagents and proprietary adapters that convert a variety of fragmented DNA inputs (as low as 10 nt in length) to sequence-ready molecules. The kits are provided in a modular format that allow maximum flexibility in the choice of Index PCR primers and DNA purification strategies.

CLARETBIO PRODUCT INFORMATION:

To order the SRSLY kits email at orders@claretbio.com

| Kit | Reactions | Catalog Number |
|--------------------------------|--------------------|----------------------------|
| SRSLY PicoPlus (Input < 10 ng) | 24 or 96 reactions | CBS-K250B-24, CBS-K250B-96 |
| SRSLY NanoPlus (Input > 10 ng) | 24 or 96 reactions | CBS-K150B-24, CBS-K150B-96 |
| Unique Dual Index Primers | 24 or 96 reactions | CBS-UD-24, CBS-UD-96 |
| Clarefy Beads | 24 or 96 reactions | CBS-BD-24 or CBS-BD-96 |



Horizon Discovery offers a catalogue of reference standards including FFPE, cfDNA and gDNA controls for diverse applications in next generation sequencing.

HORIZON PRODUCT INFORMATION:

To order the Horizon kits email orders@horizondiscovery.com

| Kit | Amount | Catalog Number |
|----------------|--------|----------------|
| Oncospan cfDNA | 350 ng | HD833 |