The relationship between cfDNA input, achievable duplex depth and sensitivity in liquid biopsy sequencing assays

JONNA GRIMSBY[1], MATTHEW DEFELICE[1], MARK FLEHARTY[1], MADELEINE DURAN[1], JUNKO TSUJI[1], MICAH RICKLES-YOUNG[1], MICHELLE CIPICCHIO[1], LISA GREEN[1], KATIE LARKIN[1], SOPHIE LOW[1], NICHOLAS FITZGERALD[1], MICHAEL NASUTI[1], JUSTIN ABREU[1], MAURA COSTELLO[1], TOM HOWD[1], CARRIE CIBULSKIS[1], STACEY GABRIEL[1], NIALL LENNON[1], BRENDAN BLUMENSTIEL[1]

[1]Broad Institute of MIT and Harvard

Introduction

Sequencing of cell-free DNA (cfDNA) from whole blood liquid biopsy is a promising molecular diagnostic tool that may enable non-invasive tumor profiling in patients and provide information for patient-specific treatments. Accurate detection of low allele fraction mutations from deeply sequenced targeted sequencing data is crucial for this process.

In an effort to better understand how sample quality and lab processing can impact our ability for effectively calling low allele fraction variants from targeted sequencing data, we sought to understand the relationships between cfDNA input into NGS library construction, sample quality, library yield, raw sequencing depth, and achievable duplex depth (read depth based on UMIs [unique molecular identifiers]).

Methods

Cell-free DNA from blood plasma was extracted from 146 patients. NGS libraries were constructed using duplex UMI adapters (Fig 1).

Targeted regions were selected through hybridization and capture (xGen Hybridization and Wash Kit, IDT) using array-synthesized probes (Twist Bioscience). 83 libraries were selected for a 665 kb target region, 35 libraries were selected for a 290 kb target region, and 28 libraries were selected for a 2Mb target region.

Libraries for each target panel were sequenced to a mean raw read depth of 19,043x (± 5884), 61,870x (± 20,288), and 26,482 (± 4891) respectively. Reads were de-duplicated by UMIs, and duplex consensus reads were generated.

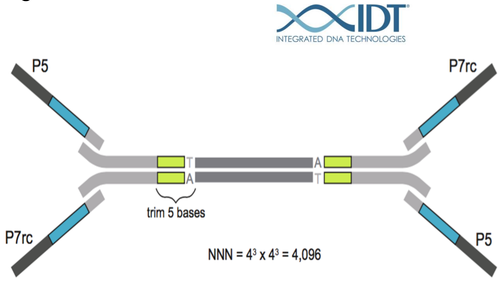

Fig. 1. IDT's "Stubby-Y" duplex adapters incorporate in-line double stranded UMIs allowing for assembly of duplex consensus reads from both top and bottom strands of the original molecule.

Duplex Consensus Read Generation

1. Used fgbio GroupReadsByUmi to group reads that appeared to have come from the same original molecule (this groups reads by template as well as by UMI sequence.)

2. Used fgbio CallDuplexConsensusReads to create duplex consensus reads by combining the evidence from a top and bottom single stranded consensus read.

A consensus read required at least one observation from both strands of the original molecule.

Disagreements in consensus called bases typically result in a base call of N. The resulting reads have high base qualities that are typically q55 or better

Raw Depth and Duplex Depth

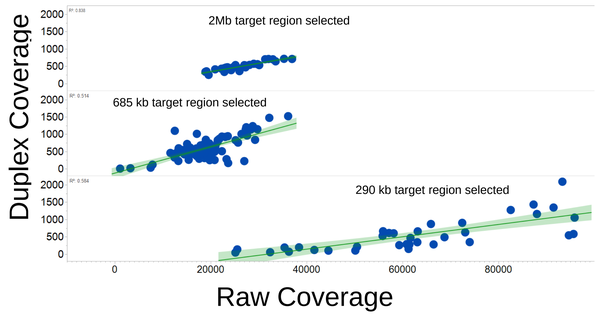

Fig. 3. Duplex depth as a function of raw depth. Raw depth is calculated as PF Aligned Bases x Percent Selected Bases / Target Territory. Duplex depth is calculated based on UMIs.

gDNA Contamination in cfDNA

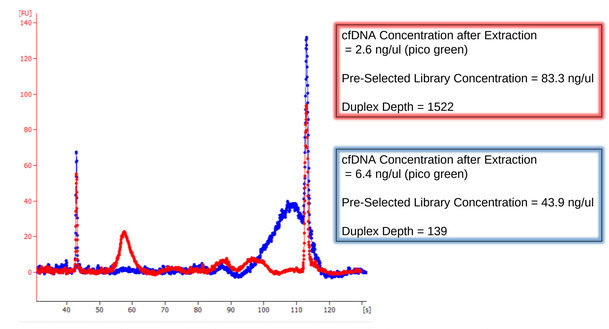

Fig. 5. High Sensitivity BioAnalyzer (Agilent) traces of cfDNA samples before library construction. One sample exhibits high gDNA contamination (blue) while the other (red) does not. The sample with high gDNA contamination resulted in lower library yield and lower duplex depth.

Duplex Depth Drives Sensitivity

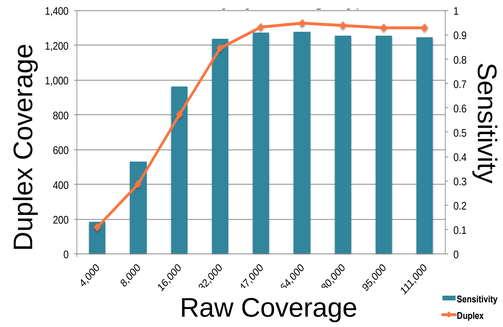

Fig. 2. In a previous experiment, cfDNA from two healthy donors was used to create a 2.5% spike-in library (20 ng DNA input into library construction) and a 2Mb targeted panel was selected through hybridization and capture. Observed MAF was 0.9%. The library was sequenced to >100,000x raw coverage and down-sampled to asses sensitivity.

DNA Input and Duplex Depth

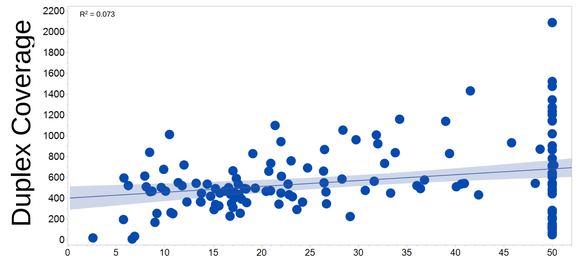

Fig. 4. Duplex depth as a function of ng cfDNA going into library construction. Duplex depth is duplex mean target coverage based on UMIs. cfDNA input into library construction was a maximum of 50 ng. Extracted DNA quantified as >1 ng/ul was normalized to 1 ng/ul before library construction.

Library Conc. and Duplex Depth

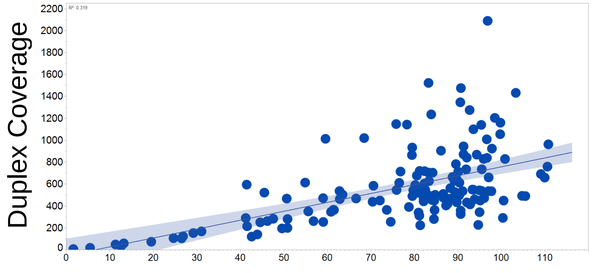

Fig. 6. Duplex depth as a function of library concentration (ng/ul) after library construction.

Conclusions

There are several considerations necessary when planning cfDNA liquid biopsy sequencing experiments:

- Maximizing your library complexity (i.e. the number of unique molecules in your library) would allow for greater duplex depth and more sensitivity in calling rare variants.
- There is a limit to duplex depth, where all unique molecules have been observed and more sequencing will not provide more information.
- The amount of cfDNA input into library construction influences duplex depth, but cfDNA concentration alone may be misleading, as gDNA can contaminate cfDNA samples, reducing the actual number of cfDNA molecules available for library construction.
- Pre-target capture library yield can serve as a useful predictor for ultimate duplex depth.

These considerations are useful when designing liquid biopsy experiments, for assessing costs, and for estimating the amount of sequencing required to achieve a given sensitivity in calling low allele fraction mutations.

Acknowledgements

We acknowledge Viktor Adalsteinsson and the Blood Biopsy Team in the Cancer Program at the Broad Institute, in addition to multiple project managers, product managers and research associates who organized or processed these samples.

The data used in this poster was generated at the Broad Institute. For more information please visit: http://genomics.broadinstitute.org/