Effective and efficient targeted sequencing of circulating cell-free tumor DNA using molecular inversion probes

127 Western Ave
Boston, MA 02134
molecularloop.com

Eric D Boyden, Patrick C Saunders, Gregory J Porreca

Background

Liquid biopsy and sequencing of circulating cell-free tumor DNA (ctDNA) is a relatively non-invasive means of early cancer detection. However, ctDNA is present in blood plasma at low concentrations and is highly fragmented due to apoptotic degradation. Therefore, methods to enrich ctDNA for high throughput sequencing must be sensitive, especially for short DNA fragments.

Molecular Loop has developed target capture technology that may be ideal for ctDNA sequencing. Our platform combines sophisticated molecular inversion probe design, reaction chemistry optimized for low quantities of DNA, and a simple single-tube workflow to sensitively capture short DNA fragments in a tiled and overlapping manner and generate complex libraries that are insensitive to allele dropout (Figure 1). It also easily incorporates unique molecular identifiers (UMIs) that are crucial for reducing false positive genotypes from PCR and sequencing errors.

Methods

We designed probes that target 73 somatic cancer exons and hotspots in 28 genes (Table 1), with an amplicon length of 100 bp and 2x5 bp UMIs. Importantly, each of the 5216 target bases is nominally captured by 3 different amplicons. We then ran 3 replicate capture reactions on 2 amounts (40 ng or 10 ng) of 2 contrived ctDNA samples (SeraCare) that each carry 22 targeted variants at ~1% or ~0.25% variant allele fraction (VAF) respectively, and sequenced the 12 libraries on an Illumina HiSeq 2500. Data were analyzed using a custom pipeline that included downsampling to 10, 8, 4, 2, 1, 0.5, or 0.25 million reads per sample, alignment to GRCh38, removal of duplicate reads, arm trimming, and generation of variant pileups.

Results

Mean target coverage after normalizing to 10M reads per sample was >45000X without read deduplication, and >4000X (40 ng) or >1000X (10 ng) with deduplication (Figure 2). 4M (40 ng) or 1M (10 ng) reads were sufficient to achieve >95% capture saturation (Figure 2). Over 95% (40 ng) or 92% (10 ng) of reads aligned on target (Figure 3), and coverage uniformity measured using the Fold-80 penalty (fold additional sequencing required to raise 80% of target bases to the mean coverage) was <1.7 (Figure 4).

For the 40 ng ~0.25% VAF and 10 ng ~1% VAF samples, we generated variant pileups and calculated the mean±SD VAF across the 3 replicates for each of the 22 known variants after downsampling to 4M (40 ng) or 1M (10 ng) reads, which produced ~4000X or ~1000X mean target coverage respectively after read deduplication (Figure 2). Expected VAFs were derived by the vendor using a commercial PCR-based NGS assay. All variants were observed in each cohort with at least 0.19% or 0.69% VAF, respectively (Tables 2 and 3). We also calculated the mean±SD VAFs across all variants; for 40 ng ctDNA, the expected and observed VAFs were 0.25±0.12% and 0.36±0.13%; for 10 ng ctDNA, the expected and observed VAFs were 0.97±0.38% and 1.20±0.41%.

Conclusions

Molecular Loop's molecular inversion probe technology is an attractive platform for targeted capture and sequencing of ctDNA. With only 40 ng ctDNA and 4 million reads, or 10 ng ctDNA and 1 million reads, variants at ~0.25% or ~1% VAF respectively were identified with high sensitivity.

This study was funded by Molecular Loop Biosciences, Inc.

Figure 1. Capture workflow comparison

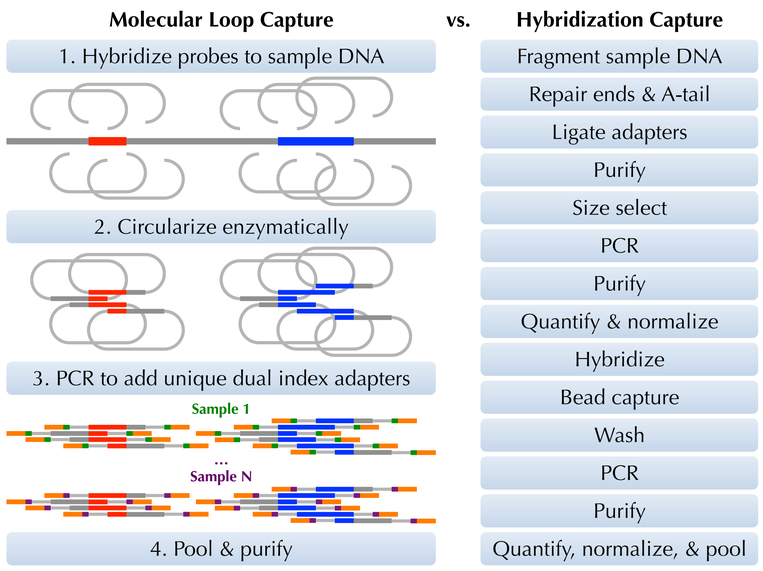

Figure 2. Mean target coverage

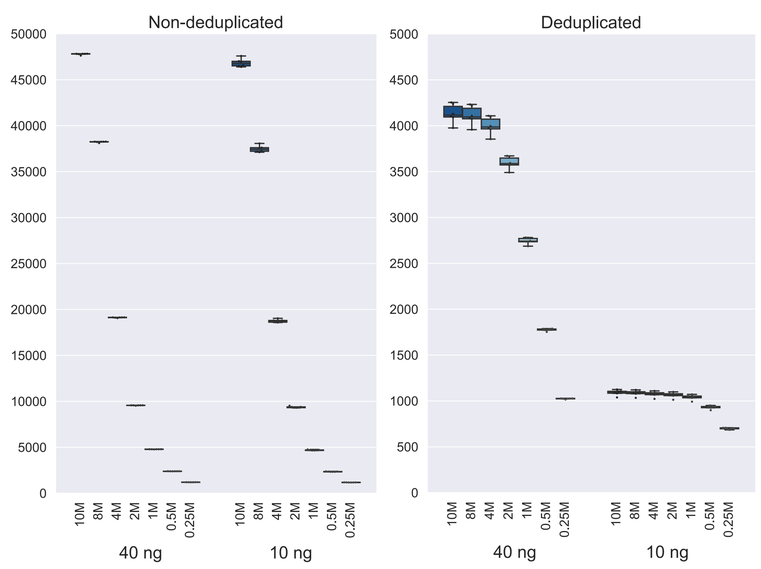

Figure 3. Fraction of reads aligned on target

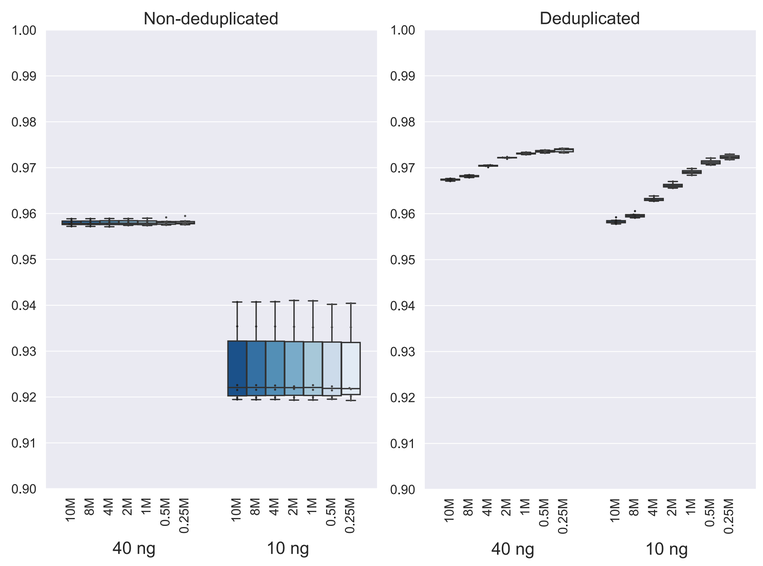

Figure 4. Fold-80 penalty

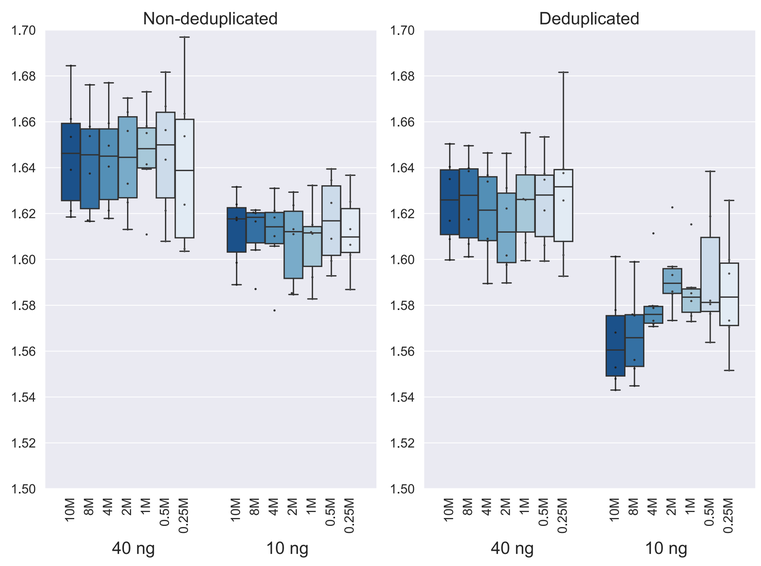

Table 1. Variants targeted (#present in ctDNA)

Gene	Targeted Exons and Amino Acids	Gene	Targeted Exons and Amino Acids
AKT1	E17K[1]	*KRAS*	G12-G13 (G12D[11]), Q61, S136-K147
ALK	C1156, Exon 23, A1266-A1274	*MAP2K1*	A52-G77, Exon 3
AR	A700-N706, I738-F748, E873-I899	*MAP2K2*	Exon 3
BRAF	D454-G474, V600E[2]	*MET*	Exon 14 splicing site
CTNNB1	W25-P44 (T41A[3])	*MTOR*	F2108, A2034, M2327
DDR2	S768	*NRAS*	G12-G13, Q61R[12]
EGFR	G459-K479, L704-G721, Exon 19 (E746-A750del[4]), Exon 20 (D770-N771insG[5], T790M[6]), L858R[7]	*NTRK1*	G595, G667
ERBB2	Exon 8, E770-R784 (Y772_A775dup[8])	*NTRK3*	D609-N626, K694-S701
ESR1	V376-A382, T460-V478, N532-L540	*PDGFRA*	Q559-E571 (S566Qfs*6[13]), I647-G662, V743-K753, V824-S851 (D842V[14])
FGFR1	Exon 13 N-terminus	*PIK3CA*	Exon 10 (E545K[15]), H1047R[16]
HRAS	G12-G13, Q61	*RET*	E632-T636, Exon 13, G798-R820, A876-K887, Exon 16 (M918T[17])
IDH1	R132C[9]	*ROS1*	I2024-L2035, K2099-N2112
IDH2	Exon 4	*SMAD4*	Exon 9
KIT	T500-K513, Exon 11, Exon 13, I808-G827 (D816V[10]), Exon 18	*TP53*	12 intervals encompassing full coding sequence & 5'UTR (S90Pfs*33[18], R175H[19], C242Afs*5[20], R248Q[21], R273H[22])

Table 2. % VAF, 40 ng ctDNA & 4M reads

Variant	Expected	Observed	Variant	Expected	Observed
1	0.17	**0.30±0.04**	12	0.36	**0.31±0.12**
2	0.21	**0.26±0.04**	13	0.16	**0.47±0.08**
3	0.13	**0.34±0.07**	14	0.30	**0.45±0.17**
4	0.37	**0.69±0.11**	15	0.57	**0.41±0.10**
5	0.26	**0.29±0.03**	16	0.28	**0.27±0.03**
6	0.36	**0.37±0.12**	17	0.38	**0.47±0.16**
7	0.16	**0.31±0.03**	18	0.34	**0.26±0.10**
8	0.08	**0.19±0.07**	19	0.28	**0.61±0.18**
9	0.39	**0.26±0.07**	20	0.15	**0.24±0.03**
10	0.16	**0.24±0.08**	21	0.17	**0.32±0.04**
11	0.26	**0.24±0.09**	22	0.05	**0.52±0.08**

Table 3. % VAF, 10 ng ctDNA & 1M reads

Variant	Expected	Observed	Variant	Expected	Observed
1	0.90	**0.71±0.06**	12	1.74	**1.08±0.18**
2	1.33	**0.90±0.07**	13	1.08	**1.28±0.16**
3	0.69	**1.26±0.11**	14	1.04	**1.54±0.70**
4	1.20	**2.59±0.42**	15	1.11	**0.73±0.30**
5	0.52	**1.28±0.51**	16	1.01	**0.94±0.04**
6	0.62	**0.69±0.14**	17	0.80	**1.33±0.82**
7	0.93	**1.01±0.31**	18	1.24	**1.41±0.45**
8	0.78	**0.83±0.28**	19	0.71	**1.61±0.60**
9	0.81	**0.88±0.11**	20	0.90	**1.13±0.36**
10	1.72	**1.41±0.29**	21	0.91	**1.17±0.30**
11	0.74	**1.17±0.37**	22	0.56	**1.41±0.48**