Targeted sequencing of a 1 Mb carrier screening panel using molecular inversion probes

Patrick C Saunders, Jack M Amaral, Arjun D Patel, Joseph M Vieira, Gregory J Porreca, Eric D Boyden

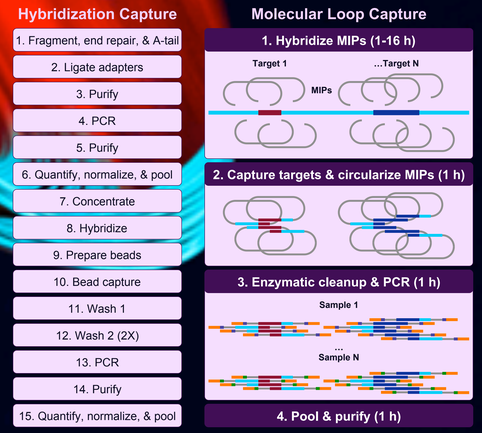
Molecular Loop Biosciences, Inc.
info@molecularloop.com

Background

Common methods of target capture for next generation sequencing (NGS) include hybridization capture, which suffers from a complicated and labor-intensive workflow, and PCR capture, which is typically simpler but prone to poor uniformity and allele dropout. We have developed and optimized an alternative method based on molecular inversion probes (MIPs) that combines the robust data quality of hybridization capture with a workflow that is even simpler than that of PCR capture. The superior performance of our chemistry is enabled by dense bi-stranded MIP tiling that ensures high coverage uniformity and resistance to allele dropout, while benefitting from a 4-step addition-only workflow that eliminates the creation of a shotgun library and all sample-level purifications. Sequencing adapters and unique dual indices (UDIs) are incorporated into the PCR primers, which allows multiplexing of hundreds to thousands of samples without risk of index cross-talk. Our protocol is highly scalable, easily automatable, requires less than 2 hours of hands-on labor, and may be performed using either a 1-day or overnight sample-to-sequencer workflow (Figure 1).

Historically MIP capture panels have commonly targeted relatively small regions (e.g. <100 kb or up to dozens of genes) with modest numbers of probes. We developed and evaluated a high performance full gene expanded carrier screening (ECS) panel to demonstrate that neither target size nor probe number is technologically limited.

Figure 1. Capture workflow comparison



Methods & Results

We designed a set of >40,000 MIPs to capture over 1 Mb of coding exons, splice sites, and 5' UTRs for 339 genes commonly targeted in ECS panels, with an average of 5 MIPs per target base (Table 1). Each MIP captures an element that is nominally 225 bp, which is ideal for PE150 sequencing and enables optional stitching of overlapping paired reads into longer single reads for improved indel sensitivity. We evaluated the performance of our panel using both 1-day and overnight workflows (Figure 2).

With either workflow the fraction of reads that aligned and represent unique capture events was very high (Figure 3). Nearly all aligned bases overlapped the tiled MIP footprint ("on bait") and more than half were on target (Figure 4). Coverage uniformity was excellent across both the full target and the subset that excludes GC-rich 5' UTRs (Figure 5). Importantly, with only 10 million reads/sample (sufficient to sequence 96 samples on a NovaSeq SP flow cell), coverage depth over most of the target was high (Figure 6), with >98% or >96% of all target bp sequenced to at least 20X or 50X, respectively (Figure 7).

This work demonstrates the efficiency and clinical utility of our ECS panel, and that target size is not a limitation of MIP capture. Even exome capture may be feasible, and would greatly improve on the simplicity and scalability of current methods without compromising data quality.

This study was funded by Molecular Loop Biosciences.

Figure 2. Experimental design & analysis

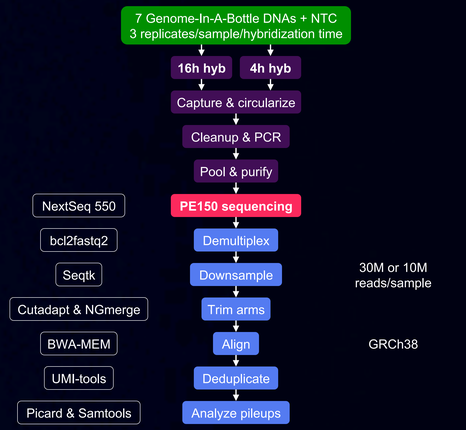
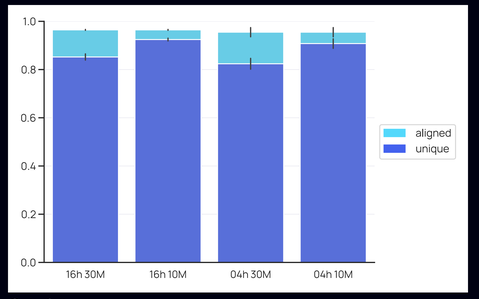


Table 1. Targeted ECS genes

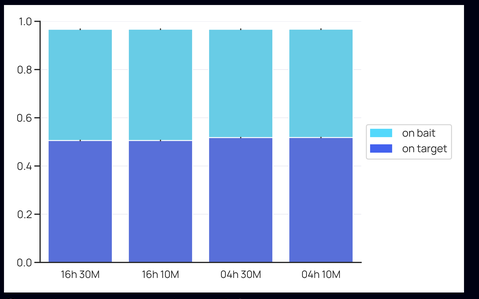
ABCB11	ATP7A	COX15	ERCC5	GBE1	HMOX1	MEFV	NPHS2	RAB23	SLC7A7
ABCC8	ATP7B	CPS1	ERCC6	GCDH	HOGA1	MESP2	NR0B1	RAG2	SMARCAL1
ABCD1	ATRX	CPT1A	ERCC8	GFM1	HPD1	MFSD8	NR2E3	RAPSN	SMN1
ACAD9	BBS1	CPT2	ESCO2	GHRHR	HPS3	MKS1	NTRK1	RARS2	SMPD1
ACADM	BBS10	CRB1	ETFA	GJB1	HSD17B4	MLC1	OAT	RDH12	STAR
ACADS	BBS12	CTNS	ETFDH	GJB2	HSD3B2	MMAA	OPA3	RMRP	SUMF1
ACADVL	BBS2	CTSA	ETHE1	GJB6	HYAL1	MMAB	OTC	RPE65	SURF1
ACAT1	BCKDHA	CTSC	EVC	GLA	HYLS1	MMACHC	PAH	RPGRIP1L	TAT
ACOX1	BCKDHB	CTSK	EVC2	GLB1	IDS	MMADHC	PANK2	RS1	TCIRG1
ACSF3	BCS1L	CYBA	EYS	GLDC	IDUA	MMUT	PC	RTEL1	TECPR2
ADA	BLM	CYBB	F11	GLE1	IL2RG	MPI	PCCA	SACS	TFR2
ADAMTS2	BSND	CYP11B1	F2	GNE	IVD	MPL	PCCB	SAMHD1	TGM1
ADGRG1	BTD	CYP11B2	F5	GNPTAB	KCNJ11	MPV17	PCDH15	SEPSECS	TH
AGA	CAPN3	CYP17A1	F9	GNPTG	LAMA2	MTHFR	PDHA1	SERPINA1	TMEM216
AGL	CBS	CYP19A1	FAH	GNS	LAMA3	MTM1	PDHB	SGCA	TPP1
AGPS	CCN6	CYP1B1	FAM161A	GP1BA	LAMB3	MTRR	PEX1	SGCB	TRMU
AGXT	CDH23	CYP21A2	FANCA	GP1BB	LAMC2	MTTP	PEX10	SGCD	TSFM
AIRE	CEP290	CYP27A1	FANCC	GP9	LCA5	MYO7A	PEX12	SGCG	TTC37
ALDH3A2	CERKL	DBT	FANCG	GRHPR	LDLR	NAGLU	PEX2	SGSH	TTPA
ALDOB	CFTR	DCLRE1C	FH	GSS	LDLRAP1	NAGS	PEX26	SLC12A3	TYMP
ALG6	CHM	DHCR7	FKRP	GUSB	LHX3	NBN	PEX6	SLC12A6	TYR
ALMS1	CHRNE	DHDDS	FKTN	HADHA	LIFR	NDRG1	PEX7	SLC17A5	USH1C
ALPL	CIITA	DLD	FMR1	HAX1	LIPA	NDUFAF2	PFKM	SLC19A2	USH2A
AMT	CLN3	DMD	FOXRED1	HBA1	LIPH	NDUFAF5	PHGDH	SLC22A5	VPS13A
AQP2	CLN5	DNAH5	FUCA1	HBA2	LOXHD1	NDUFS4	PKHD1	SLC25A13	VPS13B
ARG1	CLN6	DNAI1	G6PC1	HBB	LPL	NDUFS6	PMM2	SLC25A15	VPS45
ARSA	CLN8	DNAI2	G6PD	HEXA	LRPPRC	NDUFS7	POMGNT1	SLC25A20	VRK1
ARSB	CLRN1	DPYD	GAA	HEXB	MAN2B1	NDUFV1	PPT1	SLC26A2	VSX2
ASL	CNGB3	DYSF	GALC	HFE	MANBA	NEB	PROP1	SLC26A4	WNT10A
ASNS	COL27A1	EDA	GALK1	HGD	MCCC1	NEU1	PRPS1	SLC35A3	WRN
ASPA	COL4A3	EDAR	GALNS	HGSNAT	MCCC2	NLRP7	PSAP	SLC37A4	XPA
ASS1	COL4A4	EIF2B5	GALT	HJV	MCOLN1	NPC1	PTS	SLC39A4	XPC
ATM	COL4A5	ELP1	GAMT	HLCS	MECP2	NPC2	PUS1	SLC4A11	ZFYVE26
ATP6V1B1	COL7A1	EMD	GBA	HMGCL	MED17	NPHS1	PYGM	SLC6A8	

Figure 3. Fraction of reads aligned/unique



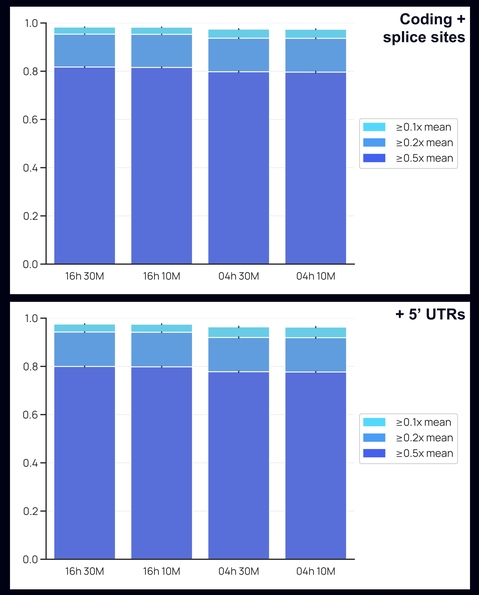
Only fully disarmed reads may align. Unique alignments represent individual capture events rather than PCR duplicates. With 30M or 10M reads, respectively, >80% or >90% of reads aligned uniquely.

Figure 4. Fraction of bases on bait/target



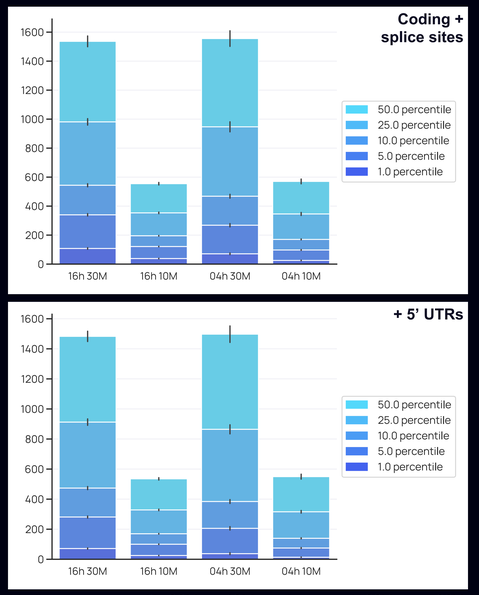
On bait bases overlap the MIP footprint but not necessarily the target footprint when MIPs are located at target boundaries. >95% of aligned bases were on bait and >50% were on target.

Figure 5. Fraction of target ≥X of mean



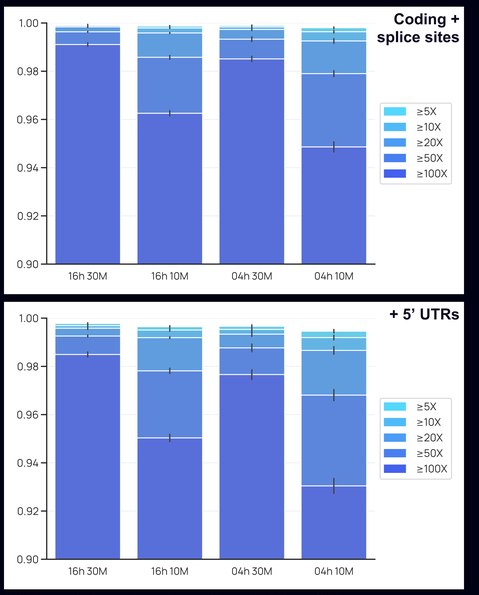
Coverage uniformity was measured as the fraction of target bp within 0.1X/0.2X/0.5X of the mean coverage depth. Uniformity was similar regardless of whether GC-rich 5' UTRs were considered, indicating that they did not have a substantial impact on uniformity.

Figure 6. Percentile target coverage



Median coverage with 30M or 10M reads was ~1500X or ~500X, respectively. With 30M reads, 95% of target bp were covered to a depth of ≥200X. With 10M reads, 95% of target bp were covered to a depth of ≥100X or ≥50X using a 16h or 4h hyb, respectively.

Figure 7. Fraction of target ≥X coverage



30M or 10M reads yielded at least 50X or 20X coverage depth, respectively, across >99% of target bp (excluding 5' UTRs) or >98% of target bp (including 5' UTRs). An overnight hybridization modestly improved coverage, especially with 10M reads.