NEBNext Direct® Custom Ready Panels overcome challenges associated with targeted re-sequencing

Andrew J. Barry[1], Sarah K. Bowman[2], Kruti M. Patel[2], Susan Corbett[2], Amy B. Emerman[2], Charles Elfe[2], Evan Mauceli[2], Jonathon Dunn[2], Brendan Desmond[2], Fiona Stewart[1], Salvatore Russello[1], Cynthia L. Hendrickson[2] Theodore B. Davis[1]

[1]New England Biolabs, Ipswich, MA, USA. [2]Directed Genomics, Ipswich, MA, USA.

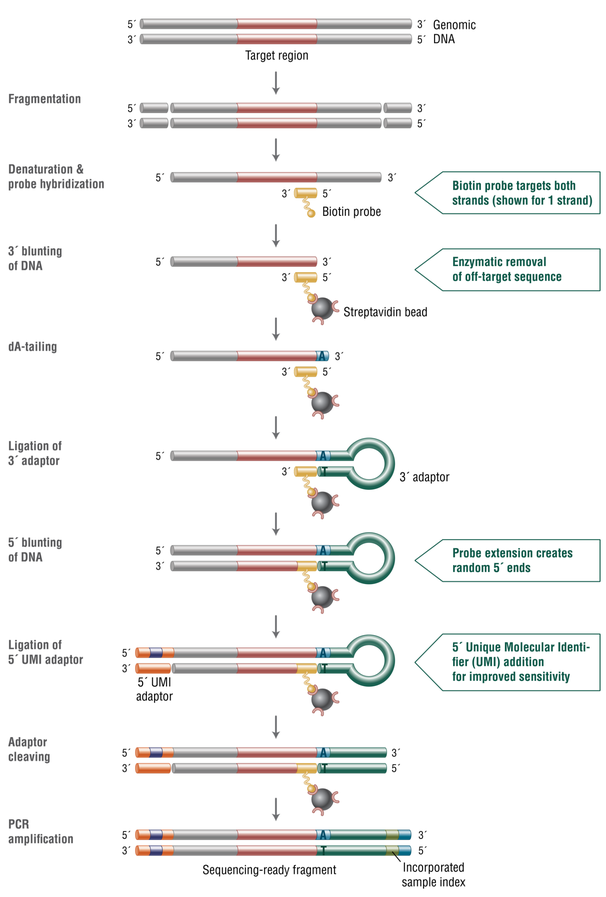
ABSTRACT

Efficient utilization of targeted gene panels for oncology research is challenged by the wide variation in gene constituents specific to a given study. While focused gene panels efficiently provide the necessary depth of coverage for low frequency variant detection, the high costs and design challenges associated with *de novo* panel design present challenges.

The NEBNext Direct® technology utilizes a novel approach to selectively enrich nucleic acid targets ranging from a single gene to several hundred genes, without sacrificing specificity. The approach rapidly hybridizes both strands of genomic DNA with biotinylated probes prior to streptavidin bead capture, enzymatic removal of off-target sequence, and conversion of captured molecules into sequencer-ready libraries. This results in a unique read coverage profile that results in uniform coverage across a given target. Unlike alternative hybridization methods, the approach does not necessitate upfront library preparation, and instead converts the captured molecules into dual-indexed Illumina sequencer compatible libraries containing an 8 basepair sample ID and a 12bp Unique Molecule Index (UMI). The UMI individually tags each molecule prior to the final PCR amplification of the library, enabling identification of PCR duplicate molecules. The result is a 1-day protocol that enables the preparation of sequence-ready libraries from purified genomic DNA specific to the gene content included in the panel.

We have designed and developed baits specific to the full exonic content of ~850 genes with clinical significance across a variety of disease areas. These are designed, balanced, and pooled on a per gene basis, and can be combined into customized panels, allowing rapid turnaround of specific custom gene subsets. Here, we present the ability to rapidly deploy custom gene panels across a variety of panel sizes and content, while maintaining high specificity, uniformity of coverage across target content, and sensitivity to detect nucleic acid variants to drive translational research applications.

1. Workflow



Libraries were prepared using the NEBNext Direct approach from a blend of 24 HapMap DNAs, Horizon® TruQ7 1% Variant Control DNA, and matched pairs of fresh frozen and FFPE DNA from liver, sheared to 200 bp with a Covaris. Denatured fragments of genomic DNA were hybridized to biotinylated baits targeting 1 to 445 genes from NEBNext Direct's Predesigned Content. Hybridized fragments were bound to streptavidin beads and separated from unbound fragments, then 3' off-target sequences were enzymatically removed. A 3' adapter and a 5' adapter containing a 12-base unique molecular identifier (UMI) were ligated to the samples, then the loop adapter was cleaved. Target strands were PCR amplified using a PCR primer to add an 8-base sample index. After sequencing the libraries on an Illumina® Miseq by PE150, the reads were aligned using BWA-MEM, and PCR duplicates were filtered using the UMIs and fgbio tools.

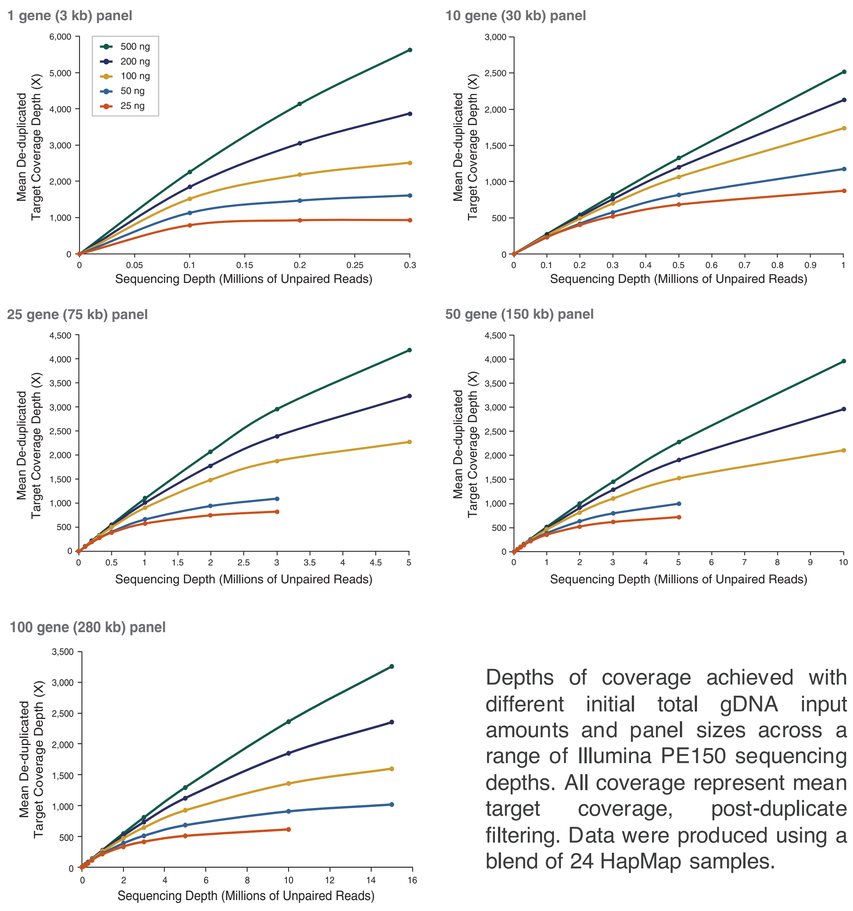
2. Materials and Methods

445 genes Custom Ready Genes associated with cancer were selected at random. Bait pools were created from 1, 10, 25, 50, 100, and all 445 genes and were used in the capture of 10 ng to 500 ng of DNA using the NEBNext Direct method. 100ng of DNA, representing a blend of 24 well-characterized HapMap Samples. Following enrichment, resulting libraries were sequenced on a MiSeq using 2 x 150 bp sequencing chemistry.

ABL1	CD274	EPHA5	GALNT12	KMT2C	NSD1	RAF1	SOX9
ABL2	**CD79A**	EPHA7	GATA1	KMT2D	NTHL1	RARA	SPEN
ACVR1B	CD79B	EPHB1	**GATA2**	**KRAS**	NTRK1	RASAL1	SPINK1
AIP	CD82	**ERBB2**	GATA3	LIG4	NTRK2	**RB1**	SPOP
AKT1	**CD83**	ERBB3	GATA4	LMO1	NTRK3	RBM10	SPRED1
AKT2	CDC73	**ERBB4**	GATA6	LRP1B	NUP93	RECQL4	SPTA1
AKT3	**CDH1**	ERCC2	GDNF	LRRFIP2	PAK3	**RET**	**SRC**
ALK	CDH2	ERCC3	GID4	LYN	**PALB2**	RFFL	**SRSF2**
AMER1	CDH4	ERCC4	GLI1	LZTR1	PALLD	RHEB	STAG2
APC	CDK12	ERCC5	**GNA11**	MAGI1	PARK2	RICTOR	STAT3
AR	**CDK4**	ERG	GNA13	MAGI2	**PAX5**	RINT1	STAT4
ARAF	CDK6	ERRFI1	**GNAQ**	**MAP2K1**	PBRM1	RIT1	**STK11**
ARFRP1	CDK8	**ESR1**	**GNAS**	MAP2K2	PDCD1LG2	RNASEL	SUFU
ARID1A	CDKN1A	ETV1	GPC3	MAP2K4	PDGFR	RNF2	SYK
ARID1B	CDKN1B	ETV4	GPR124	MAP3K1	**PDGFRA**	RNF43	TAF1
ARID2	**CDKN1C**	ETV5	**GREM1**	**MAPK1**	PDK1	ROS1	TBX3
ASIP	**CDKN2A**	ETV6	GRIN2A	MAX	PHF6	RPL11	TCF12
ASXL1	CDKN2B	EWSR1	GRM3	MC1R	**PHOX2B**	RPL26	TERC
ATM	CDKN2C	EXO1	GSK3B	MCL1	PICK1	RPL35A	TERT
ATR	**CEBPA**	**EZH2**	**H3F3A**	MDH2	PIK3C2B	**RPL5**	TERTP
ATRX	CEP112	**FAM175A**	H3F3B	MDM2	**PIK3CA**	RPS10	TET2
AURKA	CEP57	FAM46C	HGF	MDM4	PIK3CB	RPS19	TGFBR2
AURKB	CHD2	FANCA	HIST1H3B	MED12	PIK3CG	RPS24	TINF2
AXIN1	CHD4	FANCB	**HNF1A**	MEF2B	PIK3R1	RPS26	**TMEM127**
AXIN2	CHEK1	FANCC	HNF1B	MEN1	PIK3R2	RPS7	TMPRSS2
AXL	**CHEK2**	FANCD2	**HOXB13**	**MET**	PIM1	RPTOR	TNFAIP3
BAP1	CIC	FANCE	**HRAS**	MITF	PLCG2	RSPO2	TNFRSF14
BARD1	CREBBP	FANCF	HSD3B1	**MLH1**	PMS1	RUNX1	TOP1
BCL2	CRKL	FANCG	HSP90AA1	MLH3	PMS2	RUNX1T1	TOP2A
BCL2L1	CRLF2	FANCI	**IDH1**	**MPL**	**POLD1**	SBDS	**TP53**
BCL2L2	**CSF1R**	FANCL	**IDH2**	**MRE11A**	**POLE**	SDHA	TSC1
BCL6	CSF3R	FANCM	IGF1R	MRPL36	POLH	SDHAF2	TSC2
BCOR	CTCF	FAS	IGF2	**MSH2**	POT1	SDHB	TSHR
BCORL1	CTNNA1	**FBXW7**	IKBKE	MSH3	PPM1D	**SDHC**	TYR
BLM	**CTNNB1**	FGF10	IKZF1	**MSH6**	PPP2R1A	**SDHD**	TYRP1
BMPR1A	CTRC	FGF14	IL7R	MSR1	PRDM1	SET	VEGFA
BRAF	CUL3	FGF19	INHBA	MTOR	PREX2	SETBP1	**VHL**
BRCA1	CUX1	FGF23	INPP4B	**MUTYH**	PRF1	SETD2	WAS
BRCA2	CYLD	FGF3	IRF2	MXI1	**PRKAR1A**	SF3B1	WISP3
BRD2	DAXX	**FGF4**	IRF4	**MYC**	PRKCI	SLIT2	WRN
BRD3	DDB2	FGF6	IRS2	MYCL	PRKDC	SLTM	WT1
BRD4	**DDR2**	**FGFR1**	JAK1	MYCN	PRPF8	SLX4	XPA
BRIP1	DICER1	**FGFR2**	**JAK2**	**MYD88**	PRSS1	SMAD2	XPC
BTG1	DIS3L2	**FGFR3**	**JAK3**	**NBN**	PRSS8	SMAD3	XPO1
BTK	DKC1	FGFR4	JUN	**NF1**	PTCH1	**SMAD4**	**XRCC2**
BTNL2	DNMT3A	FH	KAT6A	NF2	PTCH2	SMARCA4	XRCC3
BUB1B	DNMT3B	FLCN	KDM5A	NFE2L2	**PTEN**	**SMARCB1**	ZBTB2
CALR	DOT1L	FLT1	KDM5C	NFKBIA	**PTPN11**	SMARCE1	ZFHX3
CARD11	**EGFR**	FLT3	KDM6A	NHP2	QKI	SMC1A	ZNF217
CASR	EGLN1	FLT4	**KDR**	NKX2-1	RAB35	SMC3	ZNF703
CBFB	ELAC1	**FOXL2**	KEAP1	NOP10	RAC1	**SMO**	ZNRF3
CBL	ELAC2	FOXP1	KEL	**NOTCH1**	RAD21	SMOX	**ZRSR2**
CCND1	ENG	FRS2	KIF1B	NOTCH2	**RAD50**	SNCAIP	
CCND2	EP300	FUBP1	**KIT**	NOTCH3	**RAD51**	SOCS1	
CCND3	**EPCAM**	FZR1	KLHL6	**NPM1**	**RAD51C**	SOX10	
CCNE1	EPHA3	GABRA6	KMT2A	**NRAS**	**RAD51D**	SOX2	

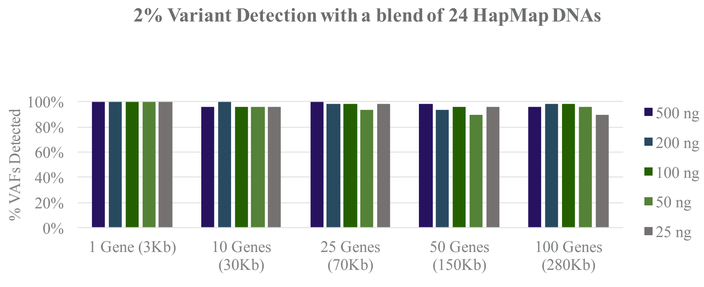
In All Panels | In 25 Gene through 445 Gene Panels | In 100 Gene and 445 Gene Panels | In 10 Gene through 445 Gene Panels | In 50 Gene through 445 Gene Panels | In 445 Gene Panel Only

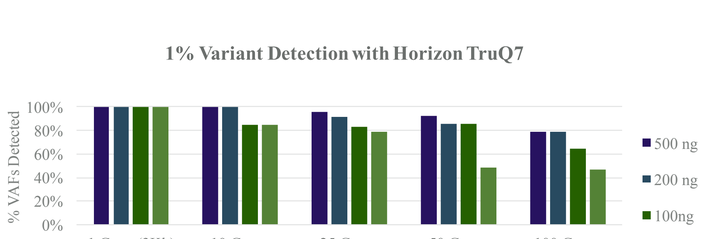
3. Results: Coverage after duplicate removal



Depths of coverage achieved with different initial total gDNA input amounts and panel sizes across a range of Illumina PE150 sequencing depths. All coverage represent mean target coverage, post-duplicate filtering. Data were produced using a blend of 24 HapMap samples.

4. Results: Sensitivity to detect variants





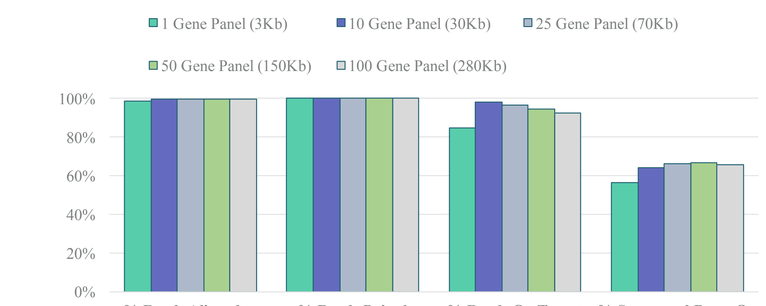
Detection of 2% known variants from the 24 HapMap DNA blend and detection of 1% known variants from TruQ7 control DNA was performed with Mutect and Vardict.

Cosmic Variants Detected within Fresh Frozen and FFPE Libraries

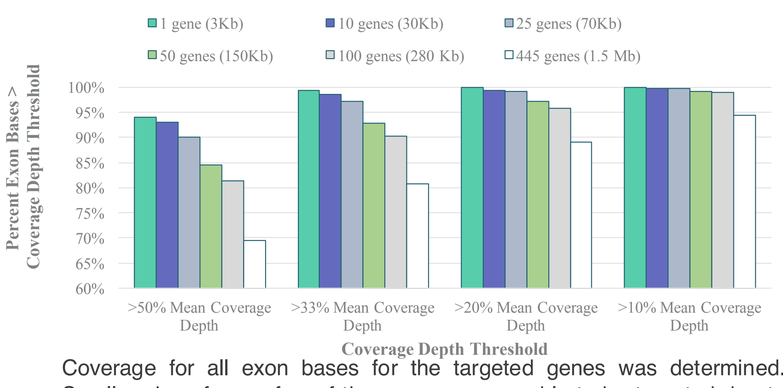
Gene	Chr	Pos	Ref	Alt	COSM ID	Confirmed Somatic	100 ng Frozen VAF	100 ng FFPE VAF	50 ng FFPE VAF	25 ng FFPE VAF
PTEN	chr10	87960892	A	T	COSM13731	Yes	3.49%	4.04%	4.17%	5.71%
PTEN	chr10	87960992	C	T	COSM921142	No	0.64%	0.66%	n.d.	n.d.
MRE11A	chr11	94479764	TA	T	COSM253028	Yes	5.49%	8.23%	7.76%	12.20%
KRAS	chr12	25209871	C	T	COSM41307	Yes	1.43%	2.01%	1.36%	n.d

Variant detection was performed with Mutect and Vardict. Cosmic variants were identified in 100 ng of fresh frozen DNA and in 25 ng to 100 ng of FFPE DNA

5. Results: Specificity and uniformity vs. panel size

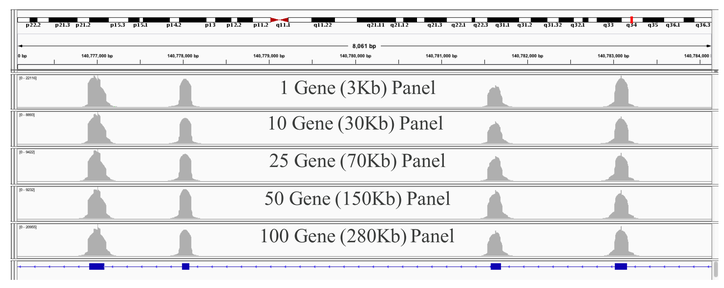


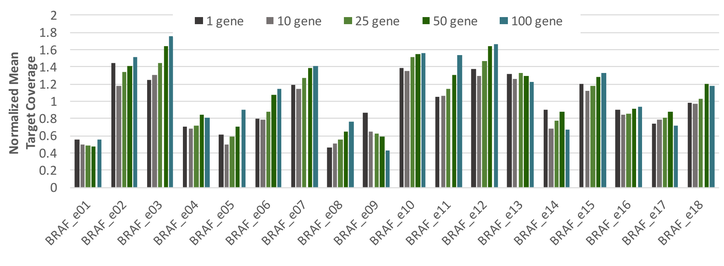
Greater than 99% of the reads aligned as pairs and 85-98% of the reads mapped to the targets, depending on the panel.



Coverage for all exon bases for the targeted genes was determined. Small regions from a few of the genes were unable to be targeted due to repetitive sequences, resulting in some decrease in coverage with the larger panels.

6. Results: Retention of target behavior





IGV images of 4 BRAF exons and mean target coverage across all BRAF exons demonstrate that BRAF retains it's coverage profile across panels.

7. Conclusions and references

Unique panels generated from subsets of Custom Ready genes display predictable capture performance with high specificity, coverage uniformity, and sensitivity across a wide range of panel sizes. Thus, we can rapidly produce cost-effective, highly scalable, custom gene panels to target specific genes for a wide range of genomic research and translational applications.

1) Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN]
2) Robinson, J. T., et al. (2011) Integrative genomics viewer. Nat. Biotechnology. 29:24–26.
3) Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P.(2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 14:178–192.
4) Cibulskis, K. et al. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnology. 31:213-219.
5) Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, Johnson J, Dougherty B, Barrett JC, and Dry JR. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016, pii: gkw227.
6) The fgbio tools are available from https://github.com/fulcrumgenomics/fgbio.