Solving Genomic Assay Trade-offs with an Optimized, Extended Cancer Gene Panel for Research and Clinical Applications

Poster
4744/10

Michael James Clark, Sean Boyle, Elena Helman, Ravi Alla, Shujun Luo, Gabor Bartha, Massimo Morra, Anil Patwardhan, Christian Haudenschild, Mirian Karbelashvili, Parin Sripakdeevong, Jason Harris, Deanna Church, Steve Chervitz, John West, Richard Chen

Personalis, Inc. | Menlo Park, CA

Contact: michael.clark@personalis.com

Choosing a genomic assay for cancer research is complicated by trade-offs. Cancer gene panels are a common choice, but they target mutational hotspots in a relatively small number of genes, often for cancers that are most commonly tested and that have common genetic etiologies. A common alternative is exome sequencing, which includes all the coding genes but, due to its larger genomic footprint, cannot reach the same depths as panels and therefore is less able to deal with low tumor purity and heterogeneity. Whole genome sequencing trades off very shallow depth and coverage over vast regions of uninterpretable genomic sequence in exchange for the identification of intergenic variants and structural variant breakpoints. All of these assays can be supplemented with RNA sequencing in order to capture gene fusions, allelic expression, splice isoforms, and gene expression. RNAseq comes with its own costs: the need to extract RNA from the same tissue, the need to perform a second assay, and the need to analyze a very different type of data from DNA sequencing.

The trade-offs generally come down to three major issues: depth of sequencing, specific genes targeted, and cost. To solve these, we designed an extended, optimized cancer gene panel facilitating high depth sequencing at low cost. We started by identifying a comprehensive list of over 1,300 cancer genes. These genes were chosen through exhaustive cancer gene database and literature curation, and include genes from all major cancer pathways and from the Cancer Gene Census. We then took this gene list and applied an augmented targeting design strategy that we have previously used to create an augmented exome enrichment platform which fills in gaps that standard technical exomes miss.

To validate the panel and analysis, we identified test samples including well-described cancer cell lines, cell line mixtures with engineered cancer variations, and formalin-fixed neoplastic tissues. We then performed a series of tests with these samples to measure the panel's small variant sensitivity and specificity, gauge its limits of detection, validate the detection of gene fusions, and demonstrate its ability to identify copy number alterations and loss of heterozygosity. In engineered cell lines, we detected 100% of small variants down to 5% allele frequency. We also mixed the cancer cell lines in various ratios and found similarly high sensitivity as well as very high specificity for small variant detection. We further compared our structural variation calls in the DNA and our fusion calls in the RNA with known data and found that we had very high concordance with known variations. These studies demonstrate that an extended, augmented cancer gene panel strategy solves many genomic assay trade-offs and leads to high accuracy and variant yield for cancer research applications.

Accuracy and Content Enhanced Sequencing to Fill in Coverage Gaps

Standard Genomic Assays Miss Important Genetic Regions

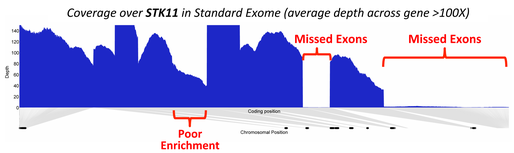

FIGURE 1: Standard enrichment-based genomic assays (exome, panels) have variable performance based on a number of factors. Coverage of STK11 by standard exome sequencing is represented in blue. Regions of low complexity or high GC content have reduced coverage due to less than optimal molecular protocols for those types of elements. Some platforms avoid targeting these regions because they underperform. Even whole genome sequencing performs poorly across these types of elements.

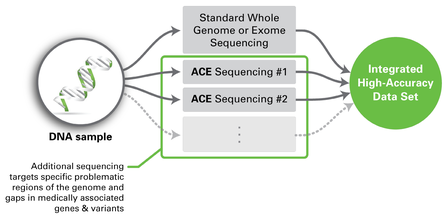

FIGURE 2: Successfully targeting and enriching genes evenly such that the entire span of the genes is covered to high depth necessitates custom probes designed to specifically enrich challenging regions. Moreover, because different reagents and parameters must be used during sample preparation, multiple enrichment steps must be performed and then sequenced.

Additional targeting also allows for capturing genomic regions outside the protein coding exons, including cancer-associated non-coding variants, promoter elements, and other regulatory content. This is applied to both DNA and RNA assays.

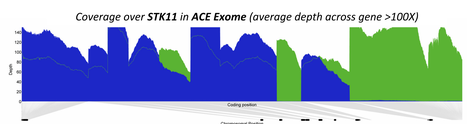

FIGURE 3: Using the Accuracy and Content Enhanced (ACE) design, protein coding regions that are missed by standard genomic assays are filled in. The regions STK11 that are missed by standard exome (blue) are enriched and sequenced to high depth via ACE exome sequencing (green).

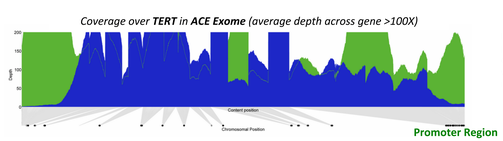

FIGURE 4: ACE enrichment enables the sequencing of important non-protein coding regions, including regulatory elements and promoters. For example, coverage of the promoter region of TERT is substantially improved when using ACE, enabling accurate variant calling in this important, clinically relevant region.

Designing an Optimized, Extended Cancer Gene Panel

Determine the Genes and Genomic Elements to Target

1. Identify **genes containing clinically actionable variants** based on the literature, clinical databases, drug databases, and clinical trials databases (**TABLE 1**).

ABL1	CBFB	DNMT3A	FLT4	MAPK1	NF2	PTCH1	SYK
AKT1	CCND2	EGFR	FOXO1	MAPK3	NFKBIA	PTEN	TET2
AKT2	CCND3	EPHB4	GNA11	MCL1	NKX2-1	RAB35	TERT
AKT3	CCNE1	ERBB2	GNAQ	MDM2	NOTCH1	RAD21	TGFBR1
ALK	CDH1	ERBB3	GNAS	MDM4	NOTCH2	RAF1	TMPRSS2
APC	CDK1	ERBB4	HRAS	MECOM	NPM1	RARA	TNFSF13B
AR	CDK2	ERG	IDH1	MED12	NRAS	RB1	TP53
ARAF	CDK4	ERRFI1	IDH2	MEN1	NTRK1	RBM15	TSC1
ASXL1	CDK6	ESR1	IGF1R	MET	NTRK2	RET	TSC2
ATM	CDKN1A	ETV1	IL6R	MITF	NTRK3	RHEB	U2AF1
ATR	CDKN1B	ETV4	JAK1	MKL1	NUP214	ROS1	VEGFA
AURKA	CDKN2A	ETV5	JAK2	MLH1	PDCD1	RPN1	VEGFB
AURKB	CDKN2B	ETV6	JAK3	MLLT3	PDGFRA	RUNX1	VHL
AURKC	CEBPA	EWSR1	KDR	MLLT10	PDGFRB	RUNX1T1	WT1
BAP1	CHEK1	EZH2	KIT	MPL	PGR	SF3B1	XPO1
BCL2	CREBBP	FGFR1	KMT2A	MSH2	PIK3CA	SMAD2	ZRSR2
BCOR	CRLF2	FGFR2	KRAS	MSH6	PIK3CB	SMAD4	
BCR	CRTC1	FGFR3	LYN	MTOR	PIK3CD	SMARCA4	
BRAF	CSF1R	FGFR4	MAP2K1	MYC	PIK3CG	SMO	
BRCA1	CTNNB1	FLCN	MAP2K2	MYD88	PIK3R1	SRC	
BRCA2	DDR2	FLT1	MAP2K4	MYH11	PML	SRSF2	
BTK	DEK	FLT3	MAP3K1	NF1	PRKCB	STAG2	

TABLE 1: Genes containing clinically actionable variants. These are included in the ACE Extended Cancer Gene Panel along with many other cancer-relevant genes.

2. Cross-reference **known cancer gene pathways** and include all major genes known to be associated with cancer (**TABLE 2**).

Pathway	Panel 1	Panel 2	ACE Extended Cancer Panel
Notch	5	9	15
PI3K-Akt	67	67	106
P53	18	19	25
TGF-beta	11	12	24
MAPK	37	46	70
JAK-STAT	26	33	45
Apoptosis	17	18	28
Cell Cycle	28	31	44
Transcriptional Misregulation	30	35	91
miRNA	51	53	175

TABLE 2: Canonical cancer gene pathways (and microRNAs) targeted by cancer gene panels. The ACE Extended Cancer Gene Panel that we developed to capture all cancer genes includes a greater number of genes across all pathways than other commercially available panels.

3. Include **known fusion partners, pharmacogenetics genes and additional genomic targets of interest.**

Balance the Final, Optimized Gene List with the Desired and Total Necessary Sequencing

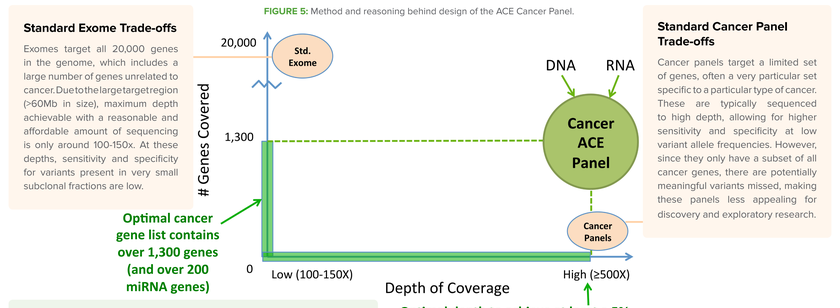

FIGURE 5: Method and reasoning behind design of the ACE Cancer Panel.

Standard Exome Trade-offs

Exomes target all 20,000 genes in the genome, which includes a large number of genes unrelated to cancer. Due to the large target region (>60Mb in size), maximum depth achievable with a reasonable and affordable amount of sequencing is only around 100-150x. At these depths, sensitivity and specificity for variants present in very small subclonal fractions are low.

Standard Cancer Panel Trade-offs

Cancer panels target a limited set of genes, often a very particular set specific to a particular type of cancer. These are typically sequenced to high depth, allowing for higher sensitivity and specificity at low variant allele frequencies. However, since they only have a subset of all cancer genes, there are potentially meaningful variants missed, making these panels less appealing for discovery and exploratory research.

Optimized Design: ACE Cancer Panel

The optimized panel design of the ACE Cancer Panel includes all genes with known associations to cancer (over 1,300 cancer genes). This gives it a substantially larger footprint than most cancer panels, which typically target tens to hundreds of genes. Due to the high throughput of modern sequencing instruments, the ACE Cancer Panel is sequenced to the same level as a typical exome, leading to substantially higher limits of detection than achieved by exome (500x or higher). Moreover, with ACE technology as a base, targeted genes are more completely covered than by standard exomes or panels.

Choosing an Assay: Depth Trade-offs

One of the major differences between whole genome, exome and panel augmented sequencing assays is the total depth of sequencing. With higher depth sequencing, there is substantially higher sensitivity for low representation variant alleles. For this reason, when variants with very low variant allele frequency are of high interest, which is often the case in low purity, high heterogeneity tumor samples, an augmented cancer panel is the best choice.

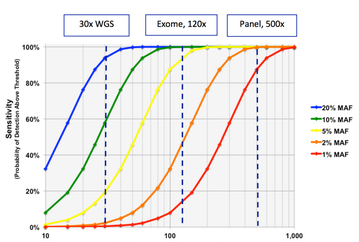

FIGURE 6: Theoretical sensitivity for alleles at particular variant allele frequencies.

Choosing a Genomic Assay Depends on the Application

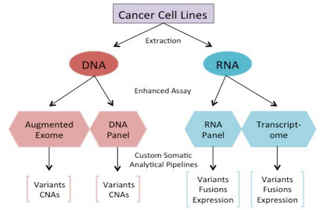

FIGURE 7: Flowchart showing the methods applied to cancer cell lines in dilution experiments to validate accuracy of ACE Cancer Panel for DNA and RNA and ACE Exome and Transcriptome. This flow shows the standard flow that would be run on any given sample. DNA and RNA are extracted from the same sample and processed either through the panel track or the whole exome track.

	VAF	Sensitivity	PPV
SNVs	≥ 5%	> 99.5%	> 99.9%
Indels	≥ 10%	> 99.9%	96%

TABLE 3: Sensitivity and positive predictive value (PPV) for SNVs with a variant allele frequency (VAF) >5% and indels with a VAF >10%.

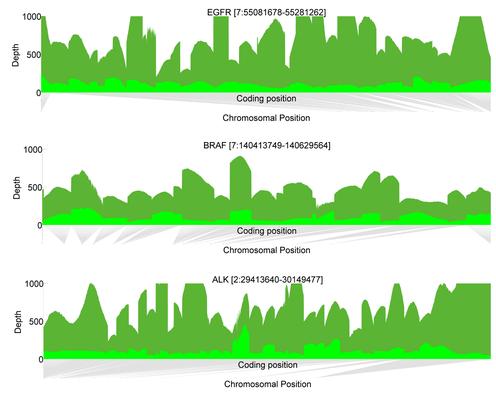

FIGURE 8: These coverage plots show the difference in depth across major cancer genes between the ACE Exome and the ACE Cancer Panel with 12Gb of sequencing. Both platforms show coverage over all exons in these genes. ACE Exome shows a mean depth of over 100x, while ACE Cancer Panel shows a mean depth of over 500x.

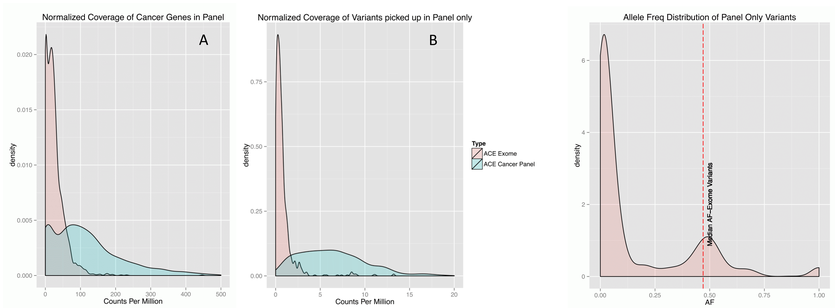

FIGURE 9: Normalized coverage of all 1,300+ cancer genes is represented by the two density plots in these figures. Figure 9A shows the coverage of all genes by ACE Exome (pink) and ACE Cancer Panel (blue) respectively. The blue ACE Cancer Panel curve shifting and extending towards higher read counts shows that there are many positions at substantially higher depth than by ACE Exome. Figure 9B shows normalized coverage over variants detected only in the panel. These variants show a substantially higher depth on average than they do by exome, showing that exome is unable to resolve them.

FIGURE 10: The allele frequency for variants detected only by Cancer ACE Panel but not by exome across a series of test samples is shown. The vast majority of variants called only by panel are at very low variant allele frequencies (below 5%).

AACR 2015 | Booth 2156

Personalis®
Pioneering Genome-Guided Medicine

www.personalis.com | info@personalis.com | +1 855-436-6634 | +1 650-752-1300

POS-ACMC-1.0 APR 2015