#2690T

Flying Blind: Building a Cancer Genomic Standard

Michael James Clark, Jennifer Yen, Elena Helman, Sean Michael Boyle, Ravi Alla, Martina Lefterova, Massimo Morra, Shujun Luo, Deanna Church, Jason Harris, Gabor Bartha, Sarah Garcia, Richard Chen, John West
Personalis, Inc. | 1330 O'Brien Drive, Menlo Park, CA 94025

Contact:
michael.clark@personalis.com

Overview

A major challenge in gaging the accuracy of genome sequencing assays and analyses is the choice of a gold standard reference that can be used as "ground truth". While great strides have been made in creating a reference standard for germline genome sequencing, these "normal" reference standards are not adequate for cancer sequencing and analysis. Cancer has unique features, including tumor heterogeneity, sample purity issues, fixation with formalin, and unusual variant types such as gene fusions and complex copy number alterations (CNAs).

In the creation of our own large NGS cancer panel, the ACE Cancer Panel, we sought to assess sensitivity, accuracy, and precision of variant calling in the context of the unique features of cancer samples. We therefore created a cancer genomic standard set using well described cancer cell lines. Heterogeneity and sample purity are major factors in cancer testing, and these must be assessed in limit of detection (LOD) tests. We mixed cell lines at particular ratios to simulate these LOD for small variant analysis. For CNAs, we mixed cancer cell lines with known CNAs with their paired normal cell lines as well.

This study defines the particular needs of a cancer genomic standard set and demonstrates the complexity involved in fully assessing the accuracy of a cancer NGS assay.

The ACE Platform

The impetus for creating a cancer genomic standard came when we were attempting to clinically validate our ACE sequencing platform, described here.

Accuracy and Content Enhanced Platform

Many cancer genes contain gaps in sequencing coverage with standard NGS. We developed an augmented target enrichment strategy optimized to fully sequence all targeted genes. This approach includes multiple enrichments and custom oligonucleotides to deal with high-GC content and challenging regions. In the augmented exome, over 8,000 medically relevant genes are enhanced for improved coverage. This augmented approach is also applied in our cancer gene panel, including over 1300 cancer genes.

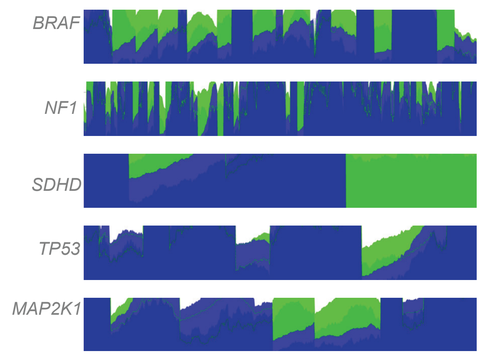

Figure 1. Sequencing coverage of select cancer genes with a standard (blue) and our augmented exome (green).

Augmented Exome and Targeted Gene Panel

While an augmented exome is suited for discovery projects, a targeted panel provides a cost-effective method to investigate relevant genes at high depth. We created a targeted cancer panel that covers over 1300 cancer genes and 200 miRNA at high depth (>500X). This panel interrogates all known cancer genes with therapeutic implications, frequently mutated genes, and genes in commonly perturbed cancer pathways. We integrate RNA analysis to both augmented exome and targeted panel approaches to enhance discovery.

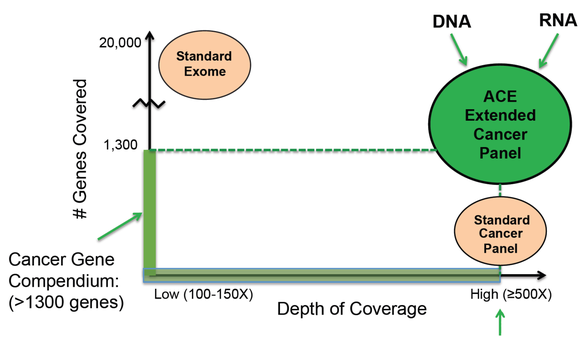

Figure 2. Depth of coverage versus gene content (number of genes) trade-off when considering cancer NGS.

The Current State of Commercial Reference Standards

There are a number of commercially available reference standards for cancer, largely focused on small variants at various variant allele frequencies (VAFs). For a preliminary investigation of the accuracy of our ACE Cancer Panel, we used Horizon Diagnostics FFPE reference standards because they had been exposed to formalin and were more representative of samples we would process in our labs.

Horizon (FFPE)

REF
- 20 variants (16 SNVs, 4 indels) at frequencies ranging from 5% to 33.5%
- Numerous major clinical variants, including BRAF V600E at 10.5%, KRAS G13D at 15%, KRAS G12D at 6%

KRAS
- 13 variants (12 SNVs, 1 indel) at frequencies ranging from 5% to 50%
- 4 major KRAS variants (G13D, G12D, Q61H, A146T) at 5%
- 2 major NRAS variants (G12V, Q61K) at 5%

EGFR
- 8 variants (8 SNVs) at frequencies ranging from 5% to 66.7%
- 4 major EGFR variants (L861Q, L858R, T790M, G719S) at 5%

RKO
- 8 variants (8 SNVs) at frequencies ranging from 1% to 66.7%
- 4 major EGFR variants (L861Q, L858R, T790M, G719S) at 1%

ONC434
- 1 EGFR exon 19 deletion (E746-A750) at 50%

We detected all known variants in these reference standards, leading us of a sensitivity of 100%. While standards of this type have some utility, they are not adequate for truly evaluating the limits of detection because they only include a relatively small number of variants, and they only include small variants to boot. Thus, there is a need for samples containing the full spectrum of mutation types and frequencies to represent the context of actual biopsies and samples received for processing.

Building a Comprehensive Cancer Reference Standard

We saw a need to create a cancer reference standard set after finding public- and commercially-available reference standards to fall short of our requirements. This is how we created a comprehensive cancer reference standard that accounted for the plethora of cancer mutation types and tested limits of detection.

Test Set Contents

We procured a number of cell lines containing known variants for validation:

Cancer Cell Lines	Variant Type
28	Small Variants
19	Copy Number Alterations (CNAs)
15	Gene Fusions

Tumor Heterogeneity (LOD)

To simulate tumor heterogeneity, cell lines containing small variants were mixed in the following ratios:

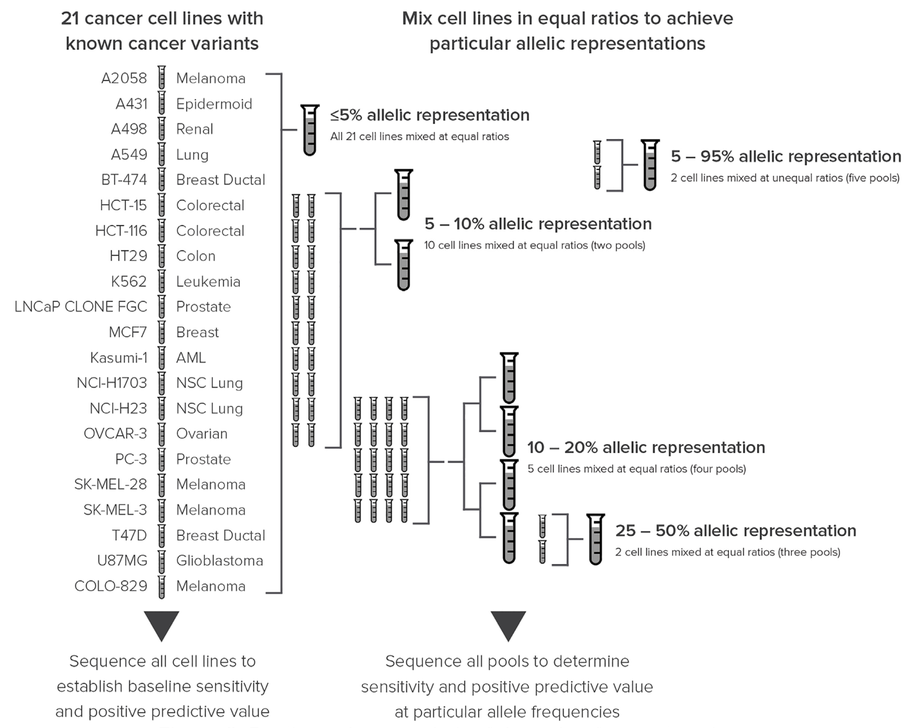

Figure 3. Pooling scheme – limit of detection for small variants.

Tumor Purity

Tumor biopsies are frequently contaminated with surrounding normal tissue. To simulate different levels of normal contamination, cell lines containing CNAs were diluted according to the following:

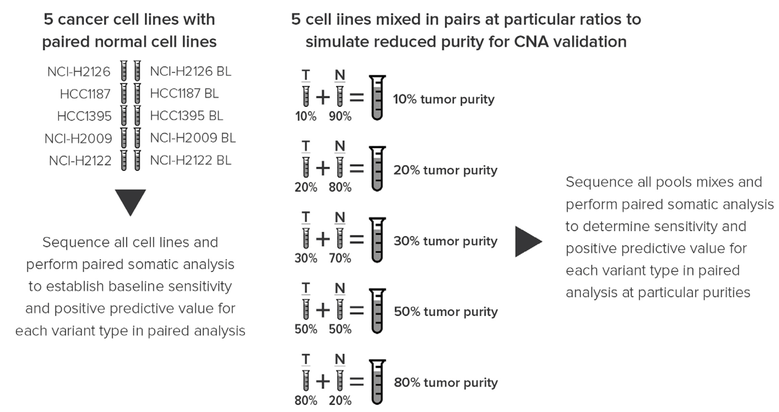

Figure 4. Dilution scheme – limit of detection for copy number alterations.

Gene Fusions

We identified fifteen cell lines known to contain fusion events that had also been identified by at least two independent studies. These events included ALK fusions, BCR-ABL1 and ROS1 fusions. The cell lines and tumor tissue are:

Cell Lines Used for Gene Fusion Events							
Cell Line	**Cancer Type**	**Cell Line**	**Cancer Type**	**Cell Line**	**Cancer Type**	**Cell Line**	**Cancer Type**
A431	Epidermoid	HCC78	Lung	LNCaP CLONE FGC	Prostate	SR	Lymphoma
A498	Renal	K562	Lukemia	MCF7	Breast	THP-1	AML
A673	Ewing's Sarcoma	Kasumi-1	AML	NCI-H2228	Lung	U937	Histolytic Lymphoma
HCC38	Breast Ductal	LA-2/ad	Lung	SJCRH30	Rhabdomyosarcoma		

Results of Testing Against a Comprehensive Cancer Reference Standard

Calculating Limits of Detection

Analytical Sensitivity Analytical Sensitivity = TP*100/(TP+FN)
LOD Sensitivity Positive Predictive Agreement (PPA) = TP/(TP+FN)
LOD Sensitivity Positive Predictive Value (PPV) = TP/(TP+FP)

Overall Validation Results Summary

Sensitivity			
	Base Substitutions	MAF ≥5%	>99%
	Indels	MAF ≥5%	>99%
	Copy Number Alterations (copy number ≥ 8 or 0)		96%
	Gene Fusions		95%
Specificity			>99%

These high percentages for sensitvity and specificity enabled us to validate our ACE Cancer Panel platform for detection of all major cancer variant types at a confidence level far beyond that with available reference standards.

Personalis®
Pioneering Genome-Guided Medicine