Supporting Neoantigen Identification for Personalized Cancer Vaccines Through Analytical Validation of an Augmented Content Enhanced (ACE) Exome

Ravi K. Alla, Jennifer Yen, Sean M. Boyle, Shujun Luo, Richard Chen, John West
Personalis, Inc. 1330 O'Brien Dr, Menlo Park, CA 94025

Contact:
ravi.alla@personalis.com

Introduction

The identification of neoantigens has become a critical step in the development of neoantigen-based personalized cancer vaccines and other immunotherapy applications. Since neoantigens can be generated from tumor specific mutations in any expressed gene, the first step in identification of neoantigens typically involves deep exome and transcriptome sequencing on the tumor and exome sequencing of the matched normal. As personalized vaccines enter clinical trials with the potential for clinical use, there is a growing need for strong analytical validation of these platforms.

To address this we have developed our ACE Exome (~200X) and Transcriptome platforms for neoantigen identification which utilitize an augmented exome approach designed to increase sensitivity for neoantigens in low complexity, traditionally hard to sequencing regions. To enable this platform for neoantigen based personalized cancer vaccines, we have performed a validation of both our ACE Exome (tumor and normal) and ACE transcriptome platforms for detecting DNA-based SNVs and Indels, as well as for RNA based small variant and fusion calls. These are variant types are especially important for neoantigen identification. We describe our validation strategy for our ACE Exome.

Methods

Cancer Reference Standards

Comprehensive cancer reference standards were created to account for Single Nucleotide Variants (SNVs) and Insertions/Deletions (Indels). We procured 11 well characterized cancer cell lines along with their matched normal cell lines (Table 1) . We identified a gold set of variants, 875 SNVs and 19 Indels that were previously validated in these 11 cell lines (COSMIC, CCLE and Sanger Sequencing confirmed variants). These gold set variants were used to calculate our analytical sensitivity (percent of gold variants detected across the 11 cell line pairs using our assay and analysis).

Table 1: Cancer Cell Lines Used for Validation

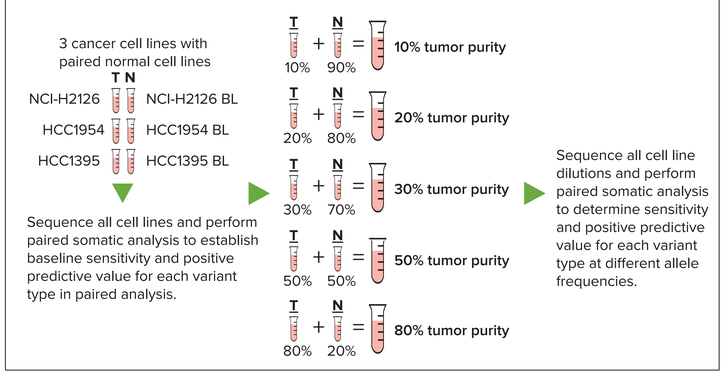
Cell Line	Matched Normal	Cancer Type
NCI-H2126	NCI-BL2126	Lung
HCC1187	HCC1187 BL	Lung
HCC1395	HCC1395 BL	Lung
NCI-H2009	NCI-BL2009	Colorectal
NCI-H2122	NCI-BL2122	Colorectal
COLO-829	COLO-829 BL	Colorectal
HCC-1599	HCC1599 BL	Breast
HCC-1954	HCC-1954 BL	Lung
HCC-2157	HCC-2157 BL	Lung
NCI-H128	NCI-BL128	Lung
NCI-H1770	NCI-BL1770	Lung

Methods — Continued

Limit of Detection (LOD) for Small Variants

To validate the LOD of small variant detection, we mixed three cancer cell lines with their matched normal cell line to generate dilutions with tumor purity ranging from 5-80%. These dilutions represent >7000 SNVs and >380 Indels with minor allele frequencies (MAF) ranging from ≥5% to <95%. The cell lines and dilution scheme are shown in Figure 1.

Figure 1: Dilution Scheme - Limits of Detection



Genomic DNA was isolated from the cell lines and the dilutions. We then created an indexed genomic library using our proprietary ACE library preparation protocol. These libraries were then pooled and enriched using ACE enrichment technology. The resulting enriched pools were then sequenced on next generation sequencers from Illumina using paired-end read technology. To ensure the quality of the library preparation and sequencing we implemented acceptance criteria as detailed in Table 2.

Table 2: DNA Quality Metrics

200x mean alignment coverage depth
≥90% Reads mapped
<0.3 Read pair duplication
≥Q30 Average base quality
≥0.35 Capture specificity

Results

Calculating Limits of Detection

Analytical Sensitivity = TP*100/(TP+FN)
LOD Sensitivity Positive Predictive Agreement (PPA) = TP*100/(TF+FN)
LOD Specificity Positive Predictive Value (PPV) = TP*100/(TP+FP) = 100 - False Discovery Rate (FDR)%

ACE Cancer Exome Performance for Small Variants

The analytical sensitivity for small variants was calculated as 98% for SNVs and 95% for Indels (Table 3).
The number of SNVs reported to be present in the LOD experiment across the whole exome were 7072 at ≥10% MAF, and 3,964 at ≥20% MAF respectively. When we compared the SNVs detected to those predicted to be present in the undiluted cell lines, we found the concordance to be very high (Table 4).
Indels were defined as 1-50 nucleotides in length. The number of indels represented in this experiment were 378 and 196 at ≥10% and ≥20% MAF respectively. When we compared the Indels detected to those predicted in the pure cell lines, we found a high concordance (Table 4).

Table 3: Analytical Sensitivity for the ACE Exome

	SNVs	InDels
Analytical Sensitivity	98%	95%

Table 4: Limits of Detection Results for the ACE Exome

	PPA (Sensitivity)		PPV (Specificity)	
MAF	10%	20%	10%	20%
Small Variants	97%	99%	98%	99%
SNVs	97%	99%	98%	99%
InDels	87%	94%	97%	97%

Conclusion

Robust validation of our comprehensive augmented exome demonstrated high sensitivity and specificity. We demonstrate that the ACE "Tumor Normal" Exome assay is highly accurate for identification of SNVs and Indels in cancer exomes. With high analytical sensitivity, PPA and low FDR we believe this assay provides augmented ability to detect cancer driver and potential neoantigen generating mutations across various tumor types.

Personalis®
Pioneering Genome-Guided Medicine