The benefits and burdens of assaying matched normal tissue when sequencing cancer genomes

#3169

Elena Helman, Michael J Clark, Ravi Alla, Sean Boyle, Selene Virk, Deanna Church, Shujun Luo,
Nan Leng, Parin Sripakdeevong, Mirian Karbelashvili, Christian Haudenschild, Richard Chen, John West
Personalis, Inc. | 1330 O'Brien Dr, Menlo Park, CA 94025

Contact:
Elena.Helman@personalis.com

Introduction

Targeted sequencing assays are increasingly used to identify tumor mutations that guide therapeutic decisions. Interpretation of a cancer variant's origin and therapeutic impact poses analytical challenges. Recent studies have indicated that jointly analyzing a tumor with its matched normal can accurately discriminate between tumor-specific (somatic) and inherited (germline) mutations. Moreover, a NHGRI/NCI Clinical Sequencing Exploratory Research Consortium Tumor Working Group just released a set of guidelines recommending that laboratories performing cancer sequencing tests should include germline variants. However, procurement of a matched sample is often logistically impractical. In the absence of a matched normal, large databases and analytical techniques are currently used to identify cancer variants in tumor sequencing data. Whether the benefits outweigh the additional burden of sequencing the matched normal for accurate detection of cancer-relevant mutations remains an open question.

Methods

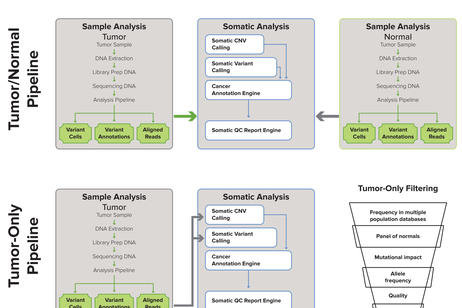
We collected 57 tumor samples where a matched normal was available — either blood, adjacent normal, or both. These samples consisted of 8 tumor types and were either fresh-frozen, FFPE and or cell lines. We sequenced these samples using our ACE augmented exome, which is a full exome with augmented coverage over cancer content. Sequencing data from each sample was analyzed using two bioinformatic pipelines: Tumor/Normal analysis and Tumor-Only analysis.

- 57 tumors
- Formalin-fixed (FFPE) tumors, fresh frozen tumors and cancer cell lines

Melanoma | GBM | Breast | Ovarian | PCC | CRC | BCC | RCC

ACE Cancer Exome

- Covers all genes in the human exome
- Augmented coverage of all cancer genes and relevant genomic content
- Sequenced to over 150X depth



For our tumor-only analysis, instead of incorporating aligned reads from the matched normal, we employed a series of filters and annotations to the variants called in the tumor to approximate somatic status. These include a variety of population frequency metrics, a panel of unmatched normals, mutational impact, and others.

Results

Consistent number of tumor-only variants are private germline across samples

We ran all our samples through both tumor/normal and tumor-only pipelines and compared variants detected by each (**Figure 1**). The variants that are unique to tumor/normal were likely filtered out of the tumor-only analysis due to the stringent filters that are necessary for our tumor-only analysis.

We cross-referenced the tumor-only variants with the matched normal. A wide rage (7–93%) of the proportion of variants detected in tumor-only mode are actually present in the matched normal and represent private germline variants not seen across a large population database (**Figure 2A**). Absolute numbers of variants however average 550 germline variants, regardless of the total number of variants called (**Figure 2B**).

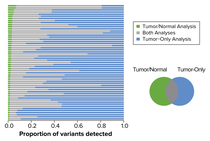


Figure 1: Proportion of variants detected by tumor/normal mode alone (blue), by tumor-only mode alone (orange), and by both modes (grey).

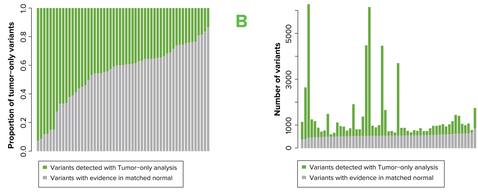


Figure 2: **A**. Each bar represents all variants called in tumor-only mode for each sample, and the maroon shows the proportion of these variants that have evidence in the matched normal. **B**. The height of each bar shows the number of variants called in tumor-only mode for each sample and the maroon shows the number of these variants that have evidence in the matched normal.

We plotted the tumor allele frequency of variants unique to tumor/normal and variants unique to tumor-only. The additional variants called by tumor-only analysis are likely germline, by their allele fraction distribution with peaks at 0.5 and 1.0.

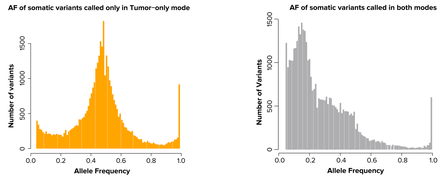


Figure 3: Allele frequency of variants detected by only tumor-only mode (left) and variants detected by both tumor-only and tumor/normal modes (right).

Clinically Actionable Variants detected depends on "actionable" definition

We examined variants in 'actionable' genes as defined by MyCancerGenome, by Jones et al. 2015, and by the TARGET database. Across 57 samples, we found 34 MyCancerGenome variants with both Tumor/Normal and Tumor-Only analyses. The Tumor-Only analysis did not identify 'extra' (germline) actionable variants. Using the two broader sets of actionable genes, we do see that the Tumor-Only analysis calls additional, germline, variant.

FILTERING CRITERIA		TUMOR/ NORMAL	BOTH	TUMOR-ONLY
MyCancer Genome	372 variants	0	34	0
Jones et al. 2015	47 genes	1	107	41
TARGET db	135 genes	7	233	93

Benefit of Large Population Databases

The use of newly available large datasets, such as ExAC, substantially decreases the number of miscalled somatic variants in the absence of a matched normal. We used 20 different population-based genomic studies to filter out common variants in our Tumor-Only pipeline, and based threshold stringency on size and quality of each population database.

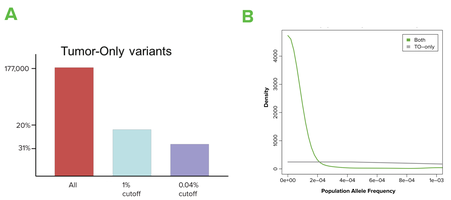


Figure 4: **A**. Number of variants detected in tumor-only mode using various population frequency filters. **B**. Distribution of maximum population frequency for each variant detected in tumor-only mode alone (red) and detected by both modes (blue).

Burden of Secondary Findings

The American College of Medical Genetics recommends that pathogenic findings in a set of 56 genes should be reported. We sequenced 30 matched normals on our augmented cancer panel (which covers 25 of the 56 ACMG genes) and 20 matched normal samples on the exome, and found almost 1200 germline variants called in these genes. Variant classification scientists went through the filtering and classification process for all of these variants and the result was four pathologic variants in three genes. We find that the burden of germline classification for secondary findings is high.

Matched normal tissue type requires consideration

Adjacent normal tissue may contain some tumor contamination, which can skew results. We took a set of 11 tumors where we had both matched blood and adjacent tissue available, and ran our Tumor/Normal analysis using each of the normals in turn. Some variants called in blood-normal analysis are filtered out when using the adjacent normal because there was evidence for the mutation in the 'normal' sample. Varying filters are necessary to account for possible tumor-in-normal contamination.

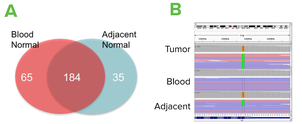


Figure 5: **A**. Number of variants called using tumor/normal mode with a matched blood normal (red) or a matched adjacent normal (blue). **B**. IGV plot of reads supporting ALT variant in tumor (top), blood normal (middle), and adjacent normal (bottom).

Conclusion

The effects of administering targeted therapies to patients with germline mutations in the relevant gene are largely unknown. Mutations of putative germline origin may be important for hereditary cancer knowledge and tumor treatment, and should be reported as such. For NGS-based cancer interpretation to guide clinical decisions in a practical and cost-effective manner, highly optimized tumor-only and tumor/normal analyses must be available with proper attention to germline consent, classification and education.

Personalis®
Pioneering Genome-Guided Medicine