When Detecting Somatic Variants in Cancer Genomes, Is it Necessary to Sequence Matched Normal Tissue?

#A025

Michael James Clark, Elena Helman, Sean Michael Boyle, Selene Virk, Ravi Alla, Shujun Luo, Gabor Bartha, Jason Harris, Nan Leng, Christian Haudenschild, Martina Lefterova, Massimo Morra, Sarah Garcia, John West, Richard Chen

Personalis, Inc. | 1330 O'Brien Dr, Menlo Park, CA 94025

Contact:
michael.clark@personalis.com

Abstract

Next-generation sequencing is commonly used to identify somatic variation in cancer. The technical act of somatic variant calling involves calling variants present in the tumor and removing those present in the normal sample. However, the detection of somatic variants in cancer is complicated by a number of issues unique to cancer samples.

Tumors are a heterogeneous mix of cells and commonly contain a number of cell populations bearing their own unique mix of somatic variations. Tumor biopsies are often contaminated with adjacent non-cancer tissue resulting in a reduction in purity. These two issues compound to result in reduced variant allele frequencies and varied allele fractions for somatic variants. On top of those challenges, cancer samples are often formalin fixed at high temperatures, causing damage to the nucleic acid that is detectable by next-gen sequencing. These features function to make somatic variant calling in real cancer samples substantially more challenging than germline calling in normal tissues.

Another challenge is that the matched normal sample from the patient may not always be readily available. Here we present a method for somatic variant calling in the absence of the matched normal which utilizes a proxy normal sample to reduce platform bias, a panel of normal samples from unrelated individuals to reduce analytical biases, and a set of filters that enriches for true somatic variation.

Method

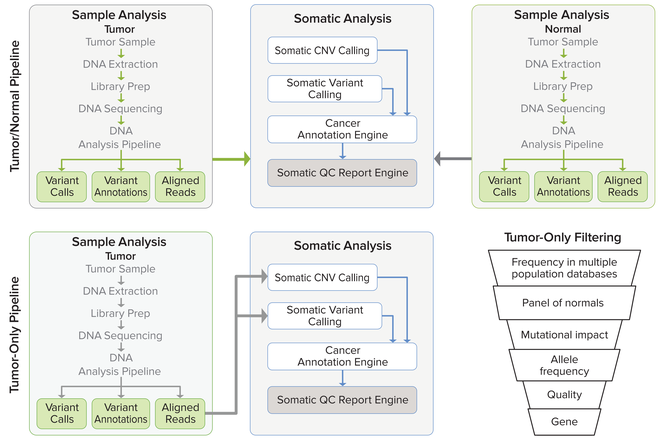
- Collected **80** tumor samples with matched normal samples
- Tumor samples were from **eight** tumor types

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

- Included formalin-fixed (FFPE) tumors, fresh frozen tumors and cancer cell lines
- Analyzed samples using two different bioinformatics pipeline approaches:
- Sequenced on two different cancer sequencing platforms:

ACE Extended Cancer Panel
- Covers > 1,600 cancer genes
- Includes actionable genes, commonly mutated genes, major cancer pathways
- Sequenced to over 500X depth
- Augmented targeting strategy

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

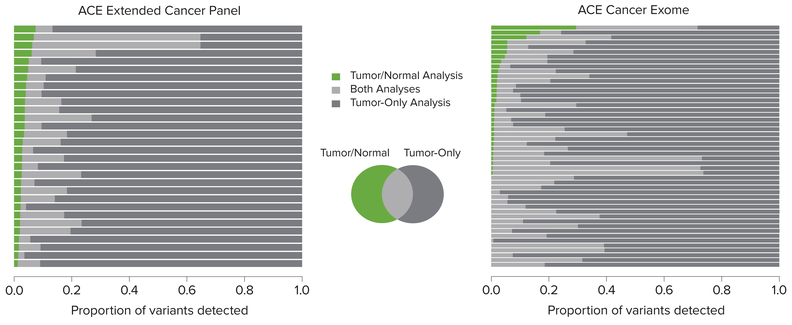


Results

Many more variants are called with tumor-only analysis when compared to tumor/normal analysis.

Using both the ACE Extended Cancer Panel and the ACE Cancer Exome, we observed the same trend of far more variants called by tumor-only than tumor/normal analysis.

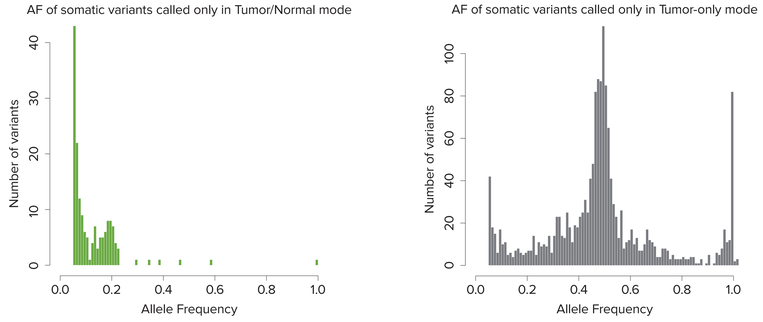
The few variants unique to tumor/normal analysis were largely filtered out of tumor-only analysis due to the stringent filters used to increase specificity in tumor-only analysis.

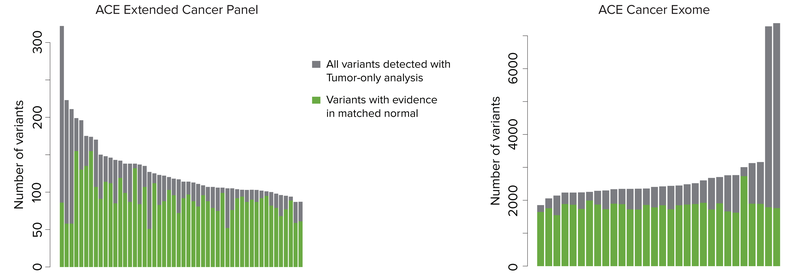


Variants called in tumor-only but not tumor/normal analysis show a likely germline allele fraction distribution.

By plotting the allele fraction (AF) of variants specific to tumor/normal analysis (blue plot) and tumor-only analysis (orange plot) respectively, we saw dramatically different trends.

The tumor-only variants show substantial enrichment near 0%, 50%, and 100% AFs, strongly suggestive that they are actually germline rather than somatic variants.





For most tumor samples, the majority of tumor-only variants are actually germline variants. Although stringent filters were applied to the tumor-only analysis to reduce germline calls, most somatic variants are still actually germline.

Filtering Criteria		Tumor/Normal	Both	Tumor-Only
MyCancerGenome	372 variants	0	25	0
Jones et al., 2015	47 genes	4	116	108
TARGETdb	135 genes	10	252	228

Different definitions of "actionable" change the number of tumor-only variants reported.

Using the most narrow definition of "actionable", the 372 variants of the MyCancerGenome database, there is no difference between tumor/normal and tumor-only calls. Using more broad definitions of "actionable" leads to substantially more tumor-only variants, and a small number of additional tumor/normal-only variants.

Including the paired normal sample does bring with it some additional considerations.

We also noted that in our cohort, four of our patients had germline cancer variants in *VHL*, *BRCA1*, and *BRCA2*. These findings have consent and genetic counseling consequences, demonstrating that including the paired normal sample brings with it an additional set of considerations beyond those when performing tumor-only sequencing.

Conclusion

In summary, we found that while including the matched normal sample reduces the number of somatic variant calls, there are filtering approaches that can be applied to tumor-only variant calls which can substantially reduce the number of germline calls. We found that the actionability criteria used in clinical reporting has a dramatic effect on the reported variants as well, and that a more unified definition of actionability would assist in optimizing somatic variant calling in tumors without matched normal samples. Moreover, there are consent and counseling implications involved in sequencing the matched normal tissue that must be considered.

Personalis
Pioneering Genome-Guided Medicine

100-736-A