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

Elena Helman, Michael J Clark, Ravi Ailta, Sean Boyle, Selene Vik, Deanna Church, Shujun Luo, Nan Leng, Paivi Sipakdeevong, Mirian Karbelashvili, Christian Haudenschild, Richard Chen, John West Personalis, Inc. 1330 O’Brien Dr, Menlo Park, CA 94025

#3169

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/NIH 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.

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 range (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).

Clinically Actionable Variants detected depends on “actionable” definition
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 frequency distribution with peaks at 0.5 and 1.0.

Benefit of Large Population Databases
The use of any available large databases, such as ExAC, substantially decreases the number of miscalled somatic variants in the absence of a matched normal. We used 20 different cancer sequencing databases to filter out common variants in our Tumor-Only pipeline, and based on the variation in the size and quality of each population database.

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 an 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 these 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. Variants filtering are necessary to account for possible tumor-in-normal contamination.

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 NSG-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.