Transcriptome augmentation provides accurate and sensitive quantification of genes associated with the tumor microenvironment

146

Fabio Navarro, Pamela Milani, Koichi Hashikawa, Shruti Bhide, Upasana Dutta, Dave Delano, Charles W. Abbott, Robin Li, Zeid Rusan, John West, John Lyle, Sean M. Boyle, Richard O. Chen
1. Personalis, Inc. | 1330 O'Brien Dr., Menlo Park, CA 94025

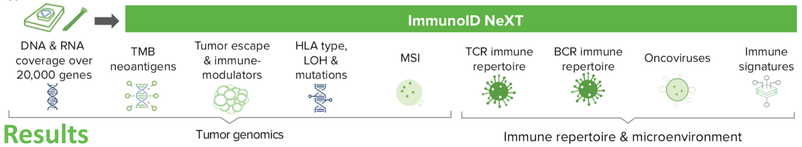
Contact:
Fabio.Navarro@personalis.com
Sean.Boyle@personalis.com

Background

Tumors harbor a complex and dynamic ecosystem of malignant, immune, and stromal cells. While malignant cells dictate much of the tumor biology, there is evidence that the tumor microenvironment (TME) also plays a significant role in disease progression and response to therapy. The role of the immune cells is particularly relevant in immunotherapy, and multiple transcriptome-based biomarkers have shown utility in predicting the efficacy of immune checkpoint blockade. However, little is known about the benefits of enhancing the depth and uniformity of transcriptome sequencing coverage for quantifying the TME cell type composition.

Methods

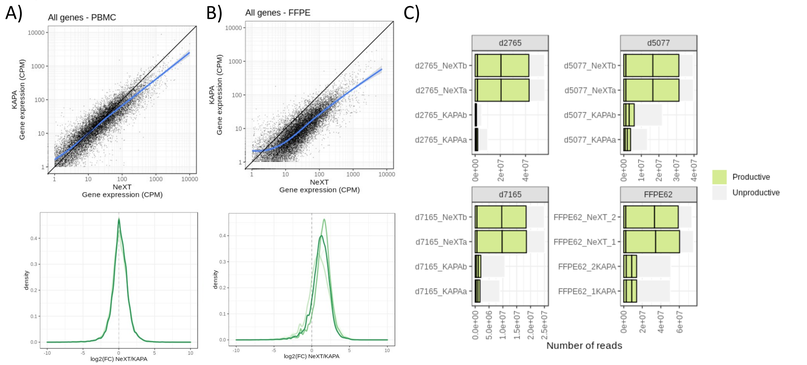
We have developed the ImmunoID NeXT Platform®, which combines high-quality exome and transcriptome sequencing with advanced informatics designed for immune-oncology to comprehensively characterize the tumor and TME from formalin-fixed paraffin-embedded (FFPE) tumor sample. Proprietary augmentation technology bolters sequencing depth in regions of low coverage across approximately 20,000 genes, enhancing transcriptome coverage uniformity. We processed and sequenced 32 Peripheral blood mononuclear cells (PBMC) samples, in-vitro cell mixtures (CD8, CD4, Tregs, B-cells), and over 100 purified cell types to assess the biases and performance of gene expression quantification using the augmented transcriptome. Immune cell composition calling was validated using flow cytometry. Using purified cell types, we applied differential expression analysis to identify novel preferentially expressed genes in target cell types. Finally, we confirmed the augmented transcriptome identifies well-established cell-type marker genes and novel cell-type enrichment genes fit for bulk tissue cell-type deconvolution.



Results

Transcriptome augmentation rescues exonic reads from challenging samples

Ribosomal depletion and augmented transcriptome RNA-seq were compared head-to-head in PBMC (A) and FFPE samples (B). We observed that the NeXT Platform benefits read coverage for most genes. Further investigation suggests that while we observe a high correlation between PBMC transcriptome quantifications in both NeXT and ribosomal depletion libraries (cor=0.956, p-value < 2.2e-16), the NeXT platform is particularly beneficial for challenging FFPE samples in which we observed a significant improvement in counts per million (CPM) read-out. The read-out improvement is a consequence of a better ratio between productive (exonic) and unproductive reads (C). Furthermore, we observed a significant improvement in the absolute number of productive reads, potentially contributing to greater accuracy of gene expression quantification in FFPE samples.

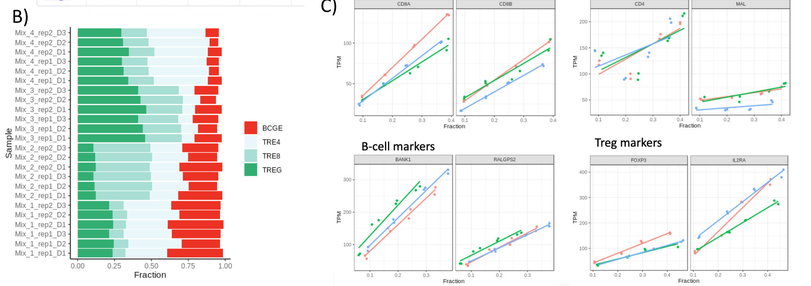


Augmented transcriptome maintains the co-linearity between gene expression and cell fractions

We next sought to understand if transcriptome augmentation would preclude bulk tissue cell type deconvolution by introducing insurmountable biases in gene expression quantification. We created in-vitro mixtures of 4 immune cell types from 3 donors at known proportions (A). Mixtures were then assessed with flow cytometry to accurately quantify B-cell, CD4, CD8, and T_{regs} (B). Flow cytometry cell fractions were correlated to well-established marker genes (C). We observed strong linear correlation between CD8 markers and T-cell CD8 fractions (CD8A corr=0.947 p-value=2.55e-12). Similar results were observed for the remaining cell types. Together, these results suggest that ImmunoID NeXT rescues more productive reads from challenging FFPE samples while maintaining the co-linearity between gene expression and cell fractions.

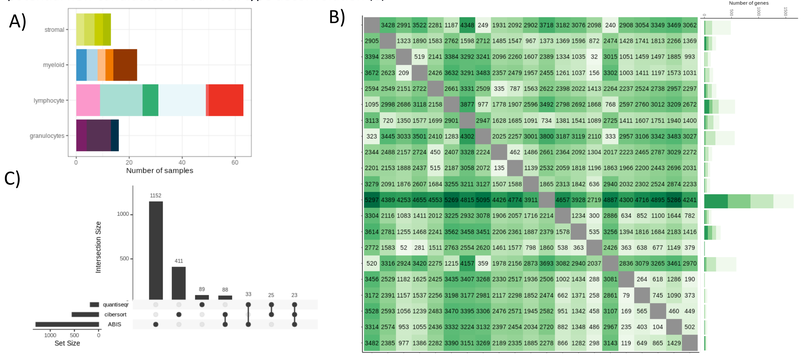
A) **Mixture Matrix**

Cell type	Mixture 1	Mixture 2	Mixture 3	Mixture 4
B-cell	40	30	20	10
CD8+	30	20	10	40
CD4+	20	10	40	30
Tregs	10	40	30	20



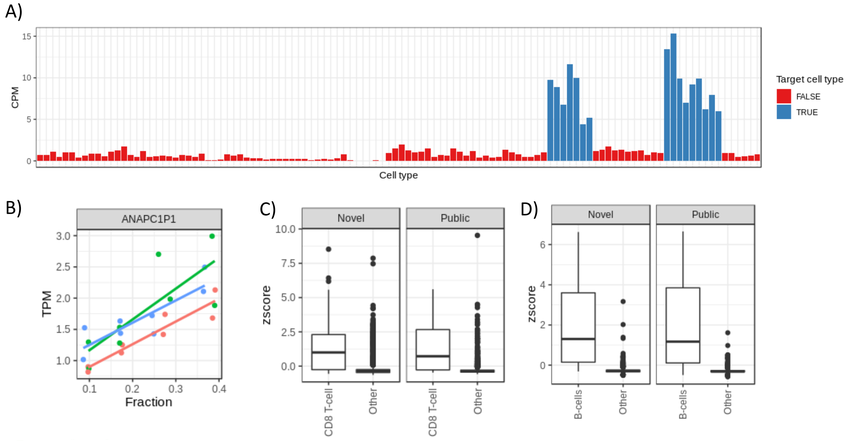
Identification of candidate marker genes for bulk tissue cell type deconvolution.

Using ImmunoID NeXT we characterized over 100 cell-type enriched samples building one of the most comprehensive cell-type expression references using augmented transcriptome (A). The majority of the reference samples are derived from immune cells such as T-cells CD8, CD4, and B-cells. However, we also characterized granulocytes, myeloid, and stroma cells to great depth. Furthermore, using pair-wise differential expression analysis, we identified genes preferentially expressed in over 20 cell types frequently present in the tumor microenvironment (B) and assessed their presence in published deconvolution signatures. Many of the newly identified preferentially expressed genes have been previously described in published signatures such as CIBERSORT (LM22)[1], quantiseqr[2], and ABIS[3]. However, the majority of preferentially expressed genes are potential new candidates for bulk cell type deconvolution (C)



Candidate marker genes display strong preferential expression in reference samples

We then evaluated the applicability of genes displaying preferential cell type expression for bulk tissue cell type deconvolution. As expected, the majority of candidate genes display a cell type-specific expression profile. For instance, we identified ANAPC1P1, a non-coding gene with strong preferential expression in CD8 T-cells (A). After further investigation, we identified ANAPC1P1 as a pseudogene sharing a bi-directional promoter with the CD8B gene. ANAPC1P1 also displayed co-linearity with CD8 fractions in our in-vitro mixtures of immune cells experiments (B). Together, these results suggest that, anecdotally, this strategy can identify novel genes fit to inform cell fractions in bulk tissue cell type deconvolution strategies. Finally, we selected genes preferentially expressed in CD8 T-cells (C) and B-cells (D) and split genes according to their occurrence in bulk cell type deconvolution studies (Novel and Public). We found that novel candidate genes show a remarkably similar z-score enrichment to genes previously described as used for bulk tissue cell type deconvolution.



Conclusion

We show that ImmunoID NeXT® accurately captures a more complete transcriptome of PBMC and FFPE samples. In particular, NeXT is able to rescue more productive (exonic) reads from challenging FFPE samples without introducing cell-fraction/expression biases. Applying this to the over 100 purified cell type samples frequently present in the TME benefited the identification of bulk tissue cell type deconvolution gene candidates. Moreover, our analyses suggest that the newly identified genes potentially fit to bulk cell type deconvolution show similar expression enrichment to genes previously described in the literature.

Reference:

1. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 12, 453–457 (2015).
2. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med 11, 34 (2019).
3. Monaco, G. et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Reports 26, 1627-1640.e7 (2019).

Personalis®

DOC-002282(A)