4-1-2020

Analytical performance of an immunoprofiling assay based on RNA models

Ian Schillebeeckx
Cofactor Genomics, Inc.

Jon R Armstrong
Cofactor Genomics, Inc.

Jason T Forys
Cofactor Genomics, Inc.

Jeffrey Hiken
Cofactor Genomics, Inc.

Jon Earls
Cofactor Genomics, Inc.

See next page for additional authors

The Journal of Molecular Diagnostics, Vol. 22, No. 4, April 2020

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Analytical Performance of an Immunoprofiling Assay Based on RNA Models

Ian Schillebeeckx,* Jon R. Armstrong,* Jason T. Forys,* Jeffrey Hiken,* Jon Earls,* Kevin C. Flanagan,* Tiange Cui,* Jarret I. Glasscock,* David N. Messina,* and Eric J. Duncavage†

From the Cofactor Genomics, Inc., San Francisco, California; and the Department of Pathology and Immunology,† Washington University School of Medicine, St. Louis, Missouri*

Accepted for publication January 14, 2020.

Address correspondence to Jon R. Armstrong, M.Sc., Cofactor Genomics, Inc., 300 Brannan St, San Francisco, CA 94107. E-mail: jon_armstrong@cofactorgenomics.com.

As immuno-oncology drugs grow more popular in the treatment of cancer, better methods are needed to quantify the tumor immune cell component to determine which patients are most likely to benefit from treatment. Methods such as flow cytometry can accurately assess the composition of infiltrating immune cells; however, they show limited use in formalin-fixed, paraffin-embedded (FFPE) specimens. This article describes a novel hybrid-capture RNA sequencing assay, ImmunoPrism, that estimates the relative percentage abundance of eight immune cell types in FFPE solid tumors. Immune health expression models were generated using machine learning methods and used to uniquely identify each immune cell type using the most discriminatively expressed genes. The analytical performance of the assay was assessed using 101 libraries from 40 FFPE and 32 fresh-frozen samples. With defined samples, ImmunoPrism had a precision of ±2.72%, a total error of 2.75%, and a strong correlation ($r^2 = 0.81$; $P < 0.001$) to flow cytometry. ImmunoPrism had similar performance in dissociated tumor cell samples (total error of 8.12%) and correlated strongly with immunohistochemistry (CD8: $r^2 = 0.83$; $P < 0.001$) in FFPE samples. Other performance metrics were determined, including limit of detection, reportable range, and reproducibility. The approach used for analytical validation is shared here so that it may serve as a helpful framework for other laboratories when validating future complex RNA-based assays. *(J Mol Diagn 2020, 22: 555–570; https://doi.org/10.1016/j.jmoldx.2020.01.009)*

Cancer pathogenesis has traditionally been viewed as a multistep process through which normal cells progressively acquire tumorigenic traits, the so-called hallmarks of cancer.[1,2] Specifically, genetic and epigenetic alterations have been considered the predominant drivers of cancer pathogenesis. Increased tumor-infiltrating lymphocytes have been associated with improved outcomes in a broad range of human cancers, including melanoma, colorectal cancer, and triple-negative breast cancer. More recently, the presence of immune cells in the tumor microenvironment has been shown to play a role in the progression of cancer and the response of the patient to therapy.[3–7]

With this increased understanding of the important interplay between cancer development and the immune system, there have been new therapeutic efforts aimed at modulating the immune system to improve patient outcomes. Most salient is the checkpoint inhibitor class of drugs. This class of drugs works by blocking immune checkpoints, which normally act to moderate immune responses by suppressing immune cell activation. By lifting this suppression, checkpoint inhibitors can reverse the blockade of an immune response against cancer cells, which is often imposed in the tumor microenvironment. Checkpoint inhibitors, such as ipilimumab and pembrolizumab, humanized monoclonal antibodies that block activation of cytotoxic T-lymphocyte-associated protein 4 and programmed cell death protein 1, respectively, have revolutionized the treatment of multiple cancer types, including melanoma, non–small-cell lung

Supported by Cofactor Genomics, Inc.
I.S. and J.R.A. contributed equally to this work.
Disclosures: All studies were performed by Cofactor Genomics, Inc. I.S., J.R.A., J.T.F., J.H., J.E., K.C.F., T.C., J.I.G., and D.N.M. are employees of Cofactor Genomics, Inc., and have common stock interests in Cofactor Genomics, Inc. J.R.A., J.I.G., and D.N.M. are members of the board of Cofactor Genomics, Inc. E.J.D. provides paid consulting services to Cofactor Genomics, Inc.

https://doi.org/10.1016/j.jmoldx.2020.01.009

cancer, and several other indications.[8,9] However, despite durable responses in some patients, the objective response rate for these treatments is typically <30%, demonstrating a clear need to understand their underlying mechanisms and develop tools to identify patients who are likely to benefit from them.[9,10] For example, studies suggest that checkpoint inhibitor therapies are less effective when tumors lack infiltrating lymphocytes in general, and are most effective in patients with preexisting programmed death-ligand 1–inhibited tumor infiltrating T cells.[11,12]

The early clinical success of checkpoint inhibitors has led to the development of other immuno-oncology therapy modalities, including adoptive cell transfer[13,14] and cancer vaccines.[15,16] These therapies all share the same intent: to improve the ability of the immune system to detect cancer cells, recruit immune cells to the site of tumors, and ultimately promote the cytotoxic functions of these immune cells to destroy cancer cells. An example of adoptive cell transfer is the use of chimeric antigen receptor T cell therapy. Representative of chimeric antigen receptor T therapies, one US Food and Drug Administration–approved chimeric antigen receptor T therapy demonstrates high overall remission rate of 81%; however, 73% of patients experienced serious adverse events.[17] Because of the high cost of chimeric antigen receptor T cell therapies and the high risk of serious adverse effects, there is a strong need for understanding which patients will benefit from treatment even when overall response rates are high.

Given the central role of immune cells in oncology and their connection to positive patient outcomes, a clear need is demonstrated for the quantification of immune cell presence, especially in solid tumor biopsies.[8] The current gold standard method, immunohistochemistry (IHC), has limited throughput and exhibits variability.[18,19] To try to address these weaknesses, clinical methods, such as Omniseq,[20] and some research methods, such as the Nanostring PanCancer IO 360 Gene Panel[21] and Cibersort,[22] have used RNA expression to profile the immune response in tumors. However, these qualitative assays rely on rank ordered gene lists or single-gene identifiers to generate cell scores that have not been validated or weakly correlate with immune cell presence. Furthermore, the qualitative cell presence generated by methods such as Nanostring is not comparable within samples. For example, T-cell scores cannot be compared with B-cell scores in the same sample. As such, these methods are not able to accurately or precisely quantify tumor-infiltrating lymphocytes in the tumor microenvironment. Therefore, new quantitative methods and technologies are required to quantify tumor-infiltrating lymphocytes in the tumor microenvironment to provide a platform for diagnostics and, ultimately, help drive drug development.

To address this need, a new RNA-based approach was developed to accurately detect the relative amount, or percentage, of immune cells in a heterogeneous cell mixture. The approach uses a database of immune health expression models (iHEMs), which are composed of the most discriminatively expressed genes that identify each of the eight immune cell types. These gene expression models afford several advantages over the single marker or ranked gene list approaches cited previously. Of note, the use of iHEMs enables more robust quantification and is less susceptible to molecular noise. The molecular and informatic pipelines of this approach have been standardized into the ImmunoPrism assay (Cofactor Genomics, Inc., San Francisco, CA). ImmunoPrism quantifies the presence of eight different immune cell types from as little as 40 ng of RNA extracted from a formalin-fixed, paraffin-embedded (FFPE) tumor sample. ImmunoPrism enriches for the small fraction of gene constituents of the iHEMs to maximize the analytical performance of immune characterization (see *Materials and Methods*). By enriching for and leveraging iHEMs, the ImmunoPrism assay provides quantitative immune profiling of tumor samples, with the goal of facilitating drug development, clinical studies, and patient care in oncology. This article describes the validation employed to characterize ImmunoPrism's analytical performance in our College of American Pathologists–accredited and Clinical Laboratory Improvement Amendments–certified laboratory. Robust assay performance is demonstrated, and an experimental framework is provided that others can use to validate complex RNA-based assays.

Materials and Methods

Performance Metrics and Statistical Analysis

There is little guidance from the College of American Pathologists on how to measure analytical performance in RNA-based assays, especially those that use machine learning methods to generate RNA models and provide percentages as read outs. Herein, how performance was measured in this new type of assay is explained.

ImmunoPrism estimates the relative abundance of eight immune cell types (eg, 10% of the cells in a processed FFPE tumor sample are $CD4^+$ T cells). Central to the analytical validation in this work are samples generated with known immune cell ratios. These samples were processed with ImmunoPrism molecular and analysis pipelines to estimate the abundance of immune cells. These estimations were compared with the known values of these samples to understand the error of ImmunoPrism estimations and derive performance metrics.

The assay's error and the subsequent analytical performance metrics are described in the same unit as the assay's estimations: percentage. As such, these metrics consider the absolute error in percentage points (of 100), as opposed to a relative error (depending on the known value). For example, if CD8 cells are known to be 10% of the cells in a sample and ImmunoPrism estimates it to be 9%, a 1% absolute error is measured, not a 10% relative error. This was chosen to simplify evaluation of errors, estimations, and performance metrics, especially across a wide reportable range of values.

For the figures in this work, absolute error metrics are presented. However, for select figures, relative error metrics are also provided.

The ImmunoPrism assay employs an analysis algorithm that is different than common genetic assays because it does not estimate a binary classification (eg, the presence or absence of a particular single-nucleotide polymorphism). Instead, this assay estimates continuous values. Therefore, diagnostic metrics, such as precision, recall, sensitivity, and specificity, are not appropriate to describe the performance of this assay. Instead, analytical metrics are used, sometimes of the same name (eg, sensitivity or limit of detection). See Saah and Hoover[23] for a complete discussion comparing diagnostic and analytic metrics.

In lieu of diagnostic metrics, this study adopted analytical performance metrics inspired by International Organization for Standardization 5725:1994 accuracy (trueness and precision) of measurement methods and results—part 1: general principles and definitions. In particular, performance was characterized by trueness, precision, and accuracy. Trueness is measured as the average error across all samples. This metric describes the bias of estimates (ie, "is it consistently too low or too high"?). Precision is measured as the first SD of error across considered samples. This metric gives a sense for the distribution of how much samples may deviate from a known value. Accuracy is defined as the root mean squared error across all samples. This metric gives a single, high-level measure of how well the assay estimates cell presence. In effect, it combines the different ways that trueness and precision describe error into a single metric. Trueness, precision, and accuracy describe the assay's error, and therefore, these should ideally be 0%. The potentially confusing semantics this may cause are noted (eg, a low accuracy value, such as 1.2%, indicates highly accurate performance). Typically, the trueness, precision, and accuracy are calculated across all cell types and all samples. To provide further resolution to the performance of the assay, these metrics are also calculated across all samples per cell type.

In addition to these analytical metrics, more traditional correlation statistics are also shown when appropriate. These include the coefficient of determination, denoted as r^2, and the two-tailed null hypothesis significance test, denoted as p. The r^2 value was calculated by squaring the sample Pearson correlation coefficient r. For reporting, r^2 was chosen over r for its ease of interpretation: it is the proportion of the variance in the dependent variable that is predictable from the independent variable. p was calculated via the t-test from r.

Generating iHEMs

ImmunoPrism leverages eight iHEMs. Each iHEM is a distinguishing pattern of gene expression that characterizes an immune cell type. For this assay, iHEMs are used that describe the identity of $CD4^+$ T cells, $CD8^+$ T cells, regulatory T cells (Tregs), M1 and M2 macrophages, monocytes, natural killer (NK) cells, and B cells. Briefly, these iHEMs were generated using machine learning methods to mine the RNA expression data of purified immune cells (origin described below) and other databases. This section describes how the iHEMs are generated and ultimately used to estimate immune cell percentages in tumor samples.

Processing of RNA Sequencing Data

FASTQ files were preprocessed with trim_galore/cutadapt version 0.4.1 to remove adapter sequences as well as reads with PHRED quality scores <20 and reads that were <20 bp. The trimmed reads were aligned to the human genome GRCh38 with STAR[24] version 2.5.2a using the two-pass method. Read counts were generated using htseq-count[25] version 0.9.1 and annotation from Gencode version 22.

The Data-Driven Approach to Defining iHEMs

The genes comprising each iHEM were selected in a data-driven way, according to the following three criteria: i) genes must have a low variability for a specific cell type, but ii) a high variability across cell types, and iii) be uncorrelated to normal or diseased tissue. Genes comprising each iHEM were identified using the following steps:

1. Six donors from CD8, CD4, M1, and M2 pure immune cell types and seven donors from CD19, Treg, CD56, and CD14 pure immune cell types (52 total donors) were sequenced using whole transcriptome RNA sequencing, and the sequence data were processed using the methods described above.
2. Immune cell donors were grouped by cell type before differential expression analysis using DESeq2[26] version 1.10.1, and $\log^2$ fold changes were computed for all protein coding genes for every pairwise comparison between immune cell types using DESeq2.

The data for the genes from each pairwise comparison were used for further filtering, as below:

a. The average counts per million mapped reads was calculated for each cell type using the following equation:

$$\overline{CPM} = \frac{\sum\left(\frac{X_i}{n}\right) * 10^6}{N} \quad (1)$$

where X_i is the number of reads mapped to a feature (eg, gene, transcript, or exon), n is the total number of mapped reads for the sample, and N is the total number of donors in the cell type.

b. Genes exhibiting highly variable expression with a coefficient of variance >25%, within a cell type, were filtered from further analysis.
c. Genes displaying expression levels of <15 counts per million, which could contain undesired sequencing noise, were filtered from further analysis.

d. From the remaining genes, log2 differential expression values for each pairwise comparison between cell types (32 comparisons) were sorted, from high to low, and the top five genes with the highest log fold change were selected for inclusion.

3. To reduce the potential for cross talk with nonimmune cell types, the average gene expression was calculated across nonhematopoietic cell lines in the Cancer Cell Line Encyclopedia (968 cell lines) and the Genotype Tissue Expression (27 tissues; version 7), and genes with average expression >2 reads per kilobase of transcript per million mapped reads were removed from further analysis.

Taken together, these intraimmune and interimmune filtering approaches yielded 125 genes. These genes are listed in Supplemental Table S1. The function of these genes, pathways, and their associated tissue expression were further investigated using Reactome Pathway Analysis version 3.6 (release 70)[27] and Molecular Signatures Database version 7.0[28] tools. This information can be found in Supplemental Figure S1 and Supplemental Tables S2 and S3. In brief, the Molecular Signatures Database showed 65 genes in gene families consisting of tumor suppressors, oncogenes, translocated cancer genes, protein kinases, cell differentiation markers, homeodomain proteins, transcription factors, cytokines, and growth factors, with >50% of the genes comprising cell differentiation markers. Reactome Pathway Analysis showed 52 genes involved in immune pathways, including adaptive and innate immune pathways.

Capture probes were designed to enrich total RNA for the resulting 125 genes, and using this reduced capture, the same pure cell donor samples were resequenced. For the chosen 125 genes, the mean count per million values were calculated, observed with the reduced capture method, across all donors for each cell type. The mean values of these 125 genes define each iHEM. Thus, informatically speaking, each iHEM is a 125-gene vector.

Estimation of Correction Factors

RNA sequencing indicates the relative expression level of different genes; however, each cell type has varying levels of RNA content. To translate transcript quantification into the relative amount of cell present for immune profiling, a corrective factor is necessary. In *Using iHEMs for Immune Profiling*, corrective factors are used that were derived to enable immune profiling in terms of cell percentage. Briefly, pure immune cell types were combined in known quantities to generate several artificial cell mixtures. RNA from these mixtures was sequenced, and immune mRNA content was estimated using the procedure described below. Correction factors were estimated via Powell nonlinear optimization. The optimization minimized the total squared error of known values and corrected values of all cell types and samples.

Using iHEMs for Immune Profiling

The task of immune profiling is to determine the relative amounts of each cell type in a sample. The expression of immune cells was characterized via iHEMs. However, a heterogeneous tissue sample will have a diverse mix of different immune cell types and nonimmune cells, and therefore a heterogeneous expression for a set of genes. Immune profiling thus seeks to solve for the relationship between the heterogeneous expression that is sequenced and the iHEMs that define immune cells.

This relationship can be modeled as a linear combination of the gene expression of each cell type present in the bulk reduced capture RNA sequencing data:

$$B = S * F \tag{2}$$

where B is a vector representing the gene expression of the 125 genes from a heterogeneous sample, S is a 125 by 8 matrix of iHEMs, and F is a vector of length 8 that represents the estimated mRNA fractions of each immune cell type present in the heterogeneous sample. For every sample, S is known, B is sequenced, and immune profiling thus solves for F.

Raw counts of an input sample were normalized to counts per million, and linear epsilon support vector regression was used to solve equation 1, yielding estimated mRNA fractions of the immune cells represented in the iHEMs. Different immune cell types generate differing amounts of mRNA, so a final operation is needed to generate a final cell type percentage estimation. To do this, cell type−specific correction factors were applied to the mRNA fractions, and the resulting corrected fractions were then scaled such that the sum of the corrected coefficients equaled the sum of the coefficients from the original mRNA.

Analytical Validation

The materials and methods that went into the analytical validation are detailed in the following sections. Where applicable, these materials and methods also apply to iHEM generation and definition.

Cells and Specimens

Cryopreserved human peripheral blood mononuclear cells (PBMCs) from normal healthy donors and cryopreserved human $CD4^+$ T cells (enriched by negative selection) from normal healthy donors were purchased from StemExpress (Folsom, CA) and Astarte Biologics (Bothwell, WA), and were stored in liquid nitrogen on receipt. Cryopreserved human $CD56^+$ NK cells from normal healthy donors (enriched by negative selection) were obtained from StemExpress. Cryopreserved human Tregs from normal healthy donors (enriched first by negative selection of $CD4^+$ T cells, followed by positive selection of $CD25^+$ cells) were obtained from StemExpress. Fresh human $CD14^+$ peripheral blood monocytes from normal healthy

donors were purchased from StemExpress and received within 20 hours of donor apheresis. The PC3 human prostate cancer cell line was purchased from Sigma-Aldrich (St. Louis, MO), and was maintained in RPMI 1640 media supplemented with 10% fetal bovine serum, 10 mmol/L HEPES buffer, 1× GlutaMAX, and 50 μg/mL gentamicin. Cell culture reagents for PC3 maintenance were purchased from Gibco/Thermo Fisher (Waltham, MA). FFPE samples were acquired from Discovery Life Sciences (Huntsville, AL), Cureline (Brisbane, CA), and House of Tissues (Janesville, WI).

Isolation of $CD4^+$ T Cells, $CD8^+$ T Cells, $CD14^+$ Monocytes, $CD19^+$ B Cells, and $CD56^+$ NK Cells from PBMCs

Cryopreserved PBMCs (50 to 100 million cells) were removed from liquid nitrogen storage and thawed in a 37°C water bath with gentle hand shaking until only a small piece of ice remained. The cells were gently transferred to a 50-mL conical centrifuge tube. Prewarmed medium [RPMI 1640 (no phenol red) supplemented with 10% fetal bovine serum, 10 mmol/L HEPES buffer, 1× GlutaMAX, and 50 μg/mL gentamicin; 1 mL] was added dropwise to the cells. Prewarmed medium (15 mL) was then slowly added to the cells. Cells were centrifuged at 200 × *g* for 10 minutes at room temperature. The supernatant was aspirated, and the cells were resuspended in 1.7 mL fluorescence-activated cell sorting (FACS) buffer [calcium- and magnesium-free Hanks' balanced salt solution (HBSS) supplemented with 2% fetal bovine serum and 0.1% sodium azide] and placed on ice. Aliquots (100 μL) were transferred to each of 10 tubes containing 100 μL PBMC antibody panel [allophycocyanin (APC) anti-human CD127 (clone A019D5), phosphatidylethanolamine (PE) anti-human CD25 (clone M-A251), PE/Cy7 anti-human CD194, Brilliant Violet 421 anti-human CD3, Brilliant Violet 785 anti-human CD4, PerCP anti-human CD8a, Brilliant Violet 510 anti-human CD14, APC/Fire 750 anti-human CD19, Alexa Fluor 700 anti-human CD45, PE/Dazzle 594 anti-human CD56; BioLegend, San Diego, CA], and incubated at 4°C for 20 minutes. The cells were washed twice with 1 mL cold FACS buffer by centrifugation at 350 × *g*, 5 minutes, 4°C. Pellets were each resuspended in 100 μL cold FACS buffer and then pooled. SYTOX Green dead cell stain (Thermo Fisher; 1 μL) was added to pooled cells. FACS sorting was performed using the BD Biosciences (San Jose, CA) Aria Fusion at the Flow Cytometry Research Core Facility at Saint Louis University School of Medicine (St. Louis, MO). Compensation was established using Anti-Mouse Ig, κ/Negative Control Compensation Particles Set (BD Biosciences) for conjugated antibodies and PBMCs for SYTOX Green dead cell stain. To start, 500,000 $CD14^+$ monocytes were sorted using a 100-μm nozzle into a 15-mL conical polypropylene tube containing 3 mL cold sort buffer (HBSS and 20% fetal bovine serum). Then, the nozzle was switched to the 70-μm size and 500,000 each of $CD4^+$ T cells, $CD8^+$ T cells, $CD19^+$ B cells, and $CD56^+$ NK cells were simultaneously sorted into 1.5-mL tubes containing 300 μL cold sort buffer. $CD14^+$ monocytes were centrifuged in a swinging bucket rotor for 5 minutes at 1000 × *g* at 4°C. Lymphocytes were microcentrifuged at 100 × *g* at 4°C for 1 minute, and then, without removal of the tubes, microcentifuged an additional 4 minutes at 1000 × *g*. The initial low speed spin prevented trailing of the cell pellet up the side of the tube. Supernatants were aspirated and pellets were lysed in 350 μL Buffer RLT Plus (Qiagen, Germantown, MD) supplemented with 1:100 volume ß-mercaptoethanol. RNA was extracted using the RNeasy Plus Micro Kit (Qiagen), according to the manufacturer's instructions. RNA concentration was measured using the Qubit assay (Thermo Fisher), according to the manufacturer's instructions.

Isolation of Tregs from $CD4^+$ T Cells

Cryopreserved $CD4^+$ T cells (5 to 15 million) were thawed and processed for FACS sorting, as described above for cryopreserved PBMCs, except that thawed cells were resuspended in 300 μL FACS buffer; and 100-μL aliquots were transferred to each of three tubes for staining with PBMC antibody panel. $CD45^+/CD3^+/CD8^-/CD4^+/CD19^-/CD127^{low}/CD25^+/CCR7^+$ cells were FACS sorted into 1.5-mL tubes containing 300 μL sort buffer, and RNA was isolated as described above.

In Vitro Macrophage Differentiation

Fresh $CD14^+$ peripheral blood monocytes were received within 20 hours of donor apheresis. Cells were washed in calcium- and magnesium-free HBSS supplemented with 2% human serum albumin (Sigma-Aldrich) by centrifugation at 200 × *g* for 10 minutes at room temperature. Pellets were resuspended to a concentration of 1×10^6 cells/mL in ImmunoCult-SF Macrophage Medium (STEMCELL Technologies, Vancouver, BC, Canada) supplemented with 50 ng/mL human recombinant macrophage colony-stimulating factor (M-CSF; STEMCELL Technologies). Cells were then seeded (day 0) at 2.5 mL per well of 6-well plates, 5 mL per T-25 cm^2 flask, or 15 mL per T-75 cm^2 flask and incubated at 37°C. On day 4, cells were fed with one half volume of media supplemented with 50 ng/mL M-CSF. For M1-like differentiation, on day 6, cells were treated with media supplemented with 50 ng/mL M-CSF, 10 ng/mL lipopolysaccharide (Sigma-Aldrich), and 50 ng/mL interferon-γ (STEMCELL Technologies). For M2-like differentiation, on day 6, cells were treated with media supplemented with 50 ng/mL M-CSF and 10 ng/mL IL-4 (STEMCELL Technologies). After the day 6 additions, incubation at 37°C was continued, and cells were harvested on day 8.

For flow cytometry, M1 macrophages were detached by replacing media with Accutase solution (Sigma-Aldrich). M2 macrophages were detached by replacing media with 2.5 mmol/L EDTA in HBSS. Flasks were incubated at 37°C for

15 minutes. Cells were then triturated off the flasks numerous times, and then transferred to 15-mL centrifuge tubs. Tubes were centrifuged 5 minutes at room temperature at 200 × *g*. Pellets were resuspended in macrophage FACS buffer (HBSS supplemented with 0.5% human serum albumin; Sigma-Aldrich), 5 ng/mL M-CSF, 2 mmol/L EDTA, and 0.1% sodium azide. Fc receptors were blocked using Human TruStain FcX, according to manufacturer's instructions (BioLegend), before staining with PE anti-CD80 (2D10), PE anti-CCR7 (G043H7), Brilliant Violet 421 anti-CD206 (15-2), or PE anti-CD209 clone 9E9A8 (BioLegend).

For RNA extractions, M1 and M2 macrophages were differentiated in 6-well dishes and harvested by aspirating media and lysing cells with 1 mL Buffer RLT Plus supplemented with 1:100 volume ß-mercaptoethanol.

For cell mixes, M1 and M2 macrophages were differentiated in Nunc UpCell 6 Multidishes (Thermo Fisher) and detached according to manufacturer's instructions. Macrophages were pelleted by centrifugation at 200 × *g* for 5 minutes at room temperature. Pellets were resuspended in HBSS supplemented with 0.5% human serum albumin, 5 ng/mL M-CSF, and 5 mmol/L EDTA. Cells were counted using a hemocytometer.

SM-PC3 Mixes

Cryopreserved PBMCs (50 million cells; Astarte Biologics) were thawed and stained for flow cytometry using PBMC antibody panel, as described above. Percentages of $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, $CD19^+$ B cells, $CD56^+$ NK cells, and Tregs were measured. Then, 1×10^6 viable $CD45^+$ cells were sorted using a 100-µm nozzle, as described above. The volume of collected cells was measured to calculate the final cell concentration. A supermix (SM) of 96% viable $CD45^+$ cells, 2% M1 macrophages, and 2% M2 macrophages was then prepared. This super mix was then combined with PC3 cells (counted with a hemocytometer) at varying percentages. All cells and cell mixes were kept ice cold before centrifuging at 4°C for 10 minutes at 200 × *g*. Pellets were lysed in 350 µL Buffer RLT Plus supplemented with 1:100 volume ß-mercaptoethanol, and RNA was extracted as described above.

Barycentric Cell Mixes Used in Determining Cell Corrections and Analytical Performance

Duplicate vials of cryopreserved PBMCs (50×10^6 cells/vial) from six different normal healthy donors were obtained from StemExpress. One vial from each donor was processed for staining with PBMC antibody panel, as described above. $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, $CD19^+$ B cells, and $CD56^+$ NK cells (500,000 cells each) were FACS sorted, as described above. The second vial from each donor was likewise stained with PBMC antibody panel, but Tregs were FACS sorted from the entire sample. At least 300,000 Tregs were obtained from each donor. FACS sorted cells were pelleted and lysed at a concentration of 500,000 cells/mL in Buffer RLT Plus supplemented with 1:100 volume ß-mercaptoethanol. Lysates were stored at −80°C.

Fresh $CD14^+$ peripheral blood monocytes from six different normal healthy donors were obtained from StemExpress. M1 and M2 macrophages were differentiated in Nunc UpCell 6 Multidishes, as described above (triplicate wells per donor). Cells from the triplicate wells were harvested, pooled, and pelleted. Pellets were resuspended in 1 mL HBSS supplemented with 0.5% human serum albumin, 5 ng/mL M-CSF, and 5 mmol/L EDTA and counted using a hemocytometer. Cells were pelleted and then lysed in 1 mL in Buffer RLT Plus supplemented with 1:100 volume ß-mercaptoethanol. Lysates were stored at −80°C.

The barycentric cell mixes were assembled by combining lysates from the individual cell types at ratios to yield the equivalent cell percentages shown in Supplemental Table S4. RNA was extracted from the combined lysate mixes using the RNeasy Plus Micro Kit (Qiagen), according to the manufacturer's instructions.

Dissociated Tumor Cells

Cryopreserved dissociated tumor cells from three indications (ovarian adenocarcinoma, lung adenocarcinoma, and melanoma) were obtained from Discovery Life Sciences (Huntsville, AL). Cells were processed and stained for FACS analysis, as described above for cryopreserved PBMCs, except that before antibody staining, Fc receptors were blocked using Human TruStain FcX, according to manufacturer's instructions (BioLegend).

The estimates of M1, M2, and CD14 cell types could not be evaluated in the dissociated tumor cell samples because acceptable flow markers do not exist for M1 and M2 macrophages, and flow cytometry of $CD14^+$ dissociated tumor cells measures all cells exhibiting CD14 proteins (monocytes, M1 macrophages, M2 macrophages, dendritic cells, and, at a lower level, neutrophils), whereas the ImmunoPrism assay specifically measures $CD14^+$ monocytes. In addition, technical challenges were encountered with the CD56 antibody for flow cytometry measurements; thus, $CD56^+$ cells were omitted from the analysis in the lung and melanoma samples.

Immunohistochemistry

Immunohistochemistry staining, imaging, and analysis were performed by Mosaic Laboratories (Lake Forest, CA). Sections from 10 FFPE blocks (4 breast cancer, 2 lung cancer, and 4 head and neck cancer; Discovery Life Sciences) were mounted on slides and stained with anti-CD20 (Dako clone L26) or costained with anti-CD8 (Dako clone C8/144B) and anti-FoxP3 (Abcam clone 236A/E7). Adjacent sections were analyzed using ImmunoPrism. Stained whole slides were imaged using an Aperio slide scanner (Leica Biosystems, Buffalo Grove, IL). Cells were quantitated for the entire tissue area using Halo software (Indica

Labs, Albuquerque, NM). B cells were $CD20^+$ and $CD8^+$, T cells were $CD8^+$, and Tregs were $FoxP3^+/CD8^-$.

Control RNA

The ImmunoPrism positive control was made by fragmenting FirstChoice Human Spleen Total RNA (Thermo Fisher). Total spleen RNA was fragmented using NEBNext Magnesium RNA Fragmentation Module (New England Biolabs, Ipswich, MA). Total RNA (2 μg) was fragmented at 94°C in a total volume of 20 μL, according to the manufacturer's protocol. After 3 minutes, the fragmentation was stopped by placing the tube on ice, adding 2 μL 10× RNA Fragmentation Stop Solution (New England Biolabs, Ipswich, MA), and mixing by pipetting. The solution was increased to a volume of 100 μL with molecular-grade water and the reaction was cleaned up with a Zymo RNA Clean and Concentrator-5 column (Irvine, CA). RNA was eluted in 22 μL molecular-grade water. Concentration was determined using a High Sensitivity RNA Qubit (Thermo Fisher), and distribution value 200 (DV_{200}) was determined using a bioanalyzer (Agilent Technologies, Santa Clara, CA). The RNA was diluted to 10 ng/μL with molecular-grade water, aliquoted, and stored at −80°C until use.

The FFPE WT ALK-RET-ROS1 Fusion Negative FFPE RNA Reference Standard was purchased from Horizon Discovery (Cambridge, UK) and used as a negative control. This fusion-negative FFPE RNA reference standard was chosen because it was generated from cell lines with no immune component. The reproducibility and negligible immune cell content were confirmed using three replicate 40-ng samples prepared on different days and by different operators.

FFPE RNA Extraction

All FFPE samples were assessed for tumor purity by a board-certified pathologist (E.J.D.). Extractions were performed using the RNAstorm kit, according to the manufacturer's instructions (Cell Data Sciences, Fremont, CA). RNA quantity was assessed by the High Sensitivity RNA Qubit assay (Thermo Fisher). A predefined yield of 40 ng FFPE RNA was established to increase the rate of successful library generation. Quality of the RNA was assessed using a bioanalyzer (Agilent Technologies, Santa Clara, CA), but no analytical threshold for DV_{200} was established.

Tissue Cellularity Limitations

ImmunoPrism requires 40 ng of total RNA from an FFPE specimen. If this amount is unattainable from a sample, this would be considered a quality control failure (quality not sufficient failure), and the sample would not be characterized by the assay. Therefore, the performance of our assay is only guaranteed with this minimum amount of starting RNA. From unpublished experiments, FFPE samples were found to yield 0.94 to 4.64 pg of RNA per cell. Therefore, one can reason that ImmunoPrism may require 42,560 to 8633 cells as input to be successfully profiled. Under these assumptions, paucicellular samples may not be able to be characterized by ImmunoPrism.

Library Preparation and Sequencing

Libraries were prepared using the ImmunoPrism Kit (Cofactor Genomics, Inc., St. Louis, MO), according to the manufacturer's instructions. The workflow for the ImmunoPrism Kit is summarized in Figure 1A. RNA input to library preparation was 20 ng for high-quality RNA samples and 40 ng for FFPE samples. A precapture library threshold requirement of 200 ng was defined. Any samples with <200 ng of material following library construction are not moved forward into hybridization and are noted as failing library construction. For whole transcriptome RNA sequencing, the xGen Exome Research Panel biotinylated oligonucleotide pool (Integrated DNA Technologies, Coralville, IA) was substituted for the custom pool included in the ImmunoPrism kit. Final libraries were sequenced as single-end 75-bp reads on a NextSeq500 (Illumina, San Diego, CA), following the manufacturer's protocols.

Calculation

Limit of detection (LOD) was calculated by iteratively considering the titrations (from highest immune content to lowest) for each cell type individually. At each step, error was calculated and the SD of that error was determined, or the precision for replicates of that step. Two SDs were subtracted from the mean error. Iteration stopped if that difference was <0%. At that point, linear interpolation was used to find the previous known amount at which this difference is 0%. If the difference was not <0%, then iteration continued to the next group of replicates with lower abundance.

Results

Assay Validation Design

Validating an RNA-based assay is challenging in and of itself, owing to potential changes in RNA content during sample processing, which can cause degradation and chemical modification of RNA. A clinical RNA assay needs to be robust to these potential changes. Analytically validating the ImmunoPrism assay is additionally challenging because of the inability to experimentally control levels of various immune cell types in a tumor sample. To address these challenges, the performance of the assay is analytically validated by progressively testing less controlled and more complicated and more highly processed samples. First, samples composed of purified immune cells are characterized. Second, controlled interfering substances are titrated

into samples with known immune cell content. Third, testing is conducted on fresh (unfixed) tumor tissue. Finally, performance in FFPE samples is measured. Each progressive step adds new evidence in a more complicated sample, while corroborating performance in previously tested, better controlled samples. By chaining these successive results, we are able to rigorously test the analytical performance to quantitatively measure the presence of immune cells in complicated and heavily processed samples. The overall validation approach is visualized in Figure 1B.

This validation was performed using 86 FFPE, fresh-frozen, and control samples. The samples were selected to reflect the anticipated specimen types for the assay (ie, FFPE human solid tumor tissue), and an appropriate number was obtained to provide a reasonable assessment of the assay performance. In addition, positive and negative control samples of commercially available human spleen and cancer cell line RNA, respectively, were included. All samples were processed with the ImmunoPrism tumor profiling molecular workflow and analysis pipelines to establish the analytical performance characteristics of the assay. Sequencing information on these samples is included in Supplemental Table S5. When possible, flow cytometry was used to measure a known presence of immune cells in samples. These samples include the immune cell mixes as well as the fresh-frozen tumor samples. For some FFPE samples, ImmunoPrism estimations are compared with IHC measurements.

Immune Profiling in Immune Cell Only Mixes

The ImmunoPrism assay is able to robustly measure the relative presence of eight different immune cell types: $CD4^+$ T cell (CD4), $CD8^+$ T cell (CD8), $CD14^+$ monocyte (CD14), $CD19^+$ B cell (CD19), $CD56^+$ natural killer cell (CD56), M1 macrophage (M1), M2 macrophage (M2), and Treg. As a first step in validating the assay, its ability to measure levels of each of the eight assayed cell types in the absence of any nonassayed cell types was evaluated. Immune cell type percentages were measured using ImmunoPrism and flow cytometry in mixes consisting of variable

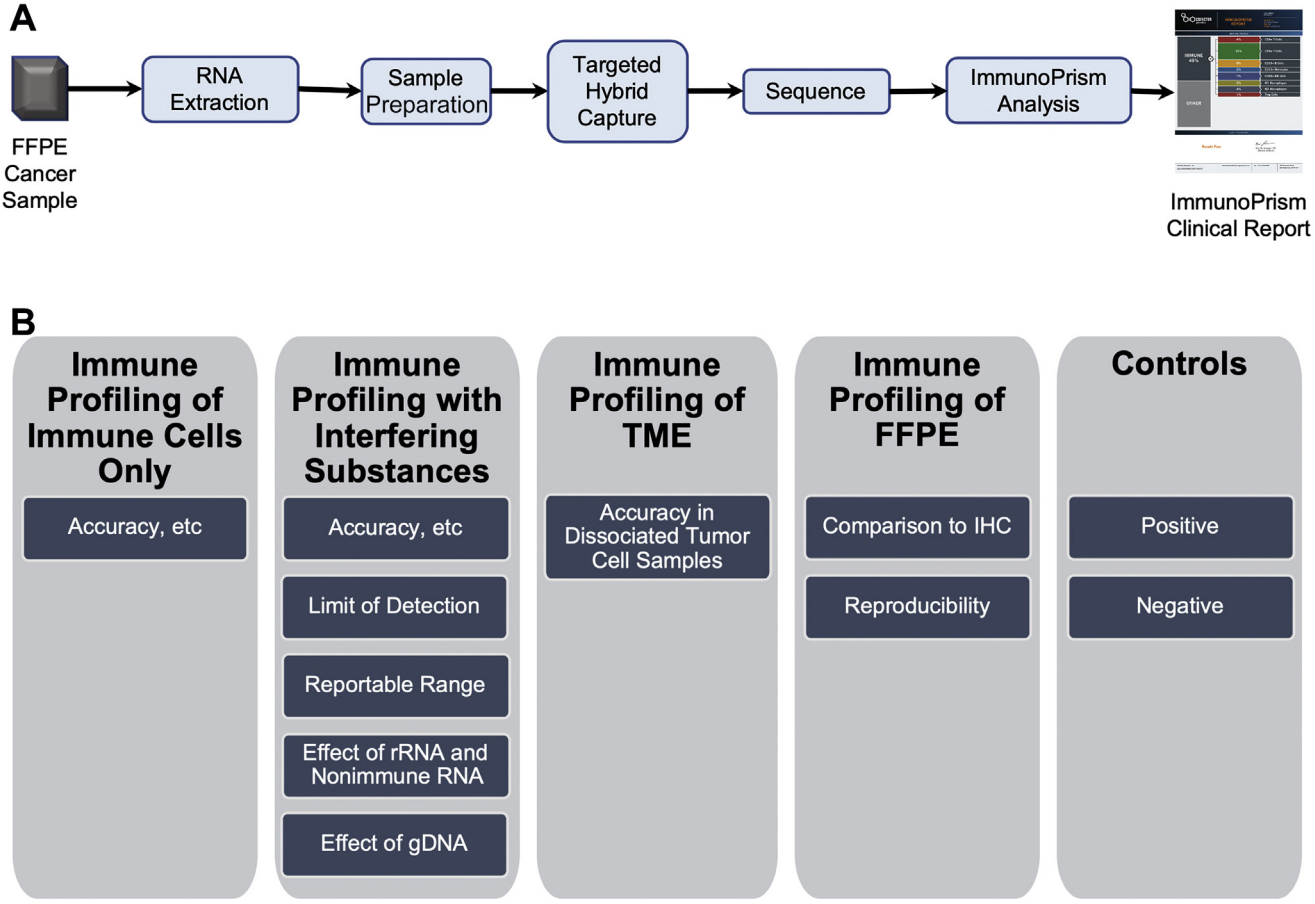

Figure 1 ImmunoPrism workflow. **A:** The assay is designed to use multiple types of cancer samples, including formalin-fixed, paraffin-embedded (FFPE) and fresh-frozen tissue. RNA is extracted from the tissue, and selected RNA transcripts are isolated for next-generation sequencing. Once data are generated, they are analyzed using the ImmunoPrism cloud-based analysis pipeline to generate cell percentage estimations and a clinical report is generated for each patient sample. **B:** The assay was evaluated for multiple performance parameters, including trueness, precision, accuracy, limit of detection, and reproducibility, using various types of samples: mixes of immune cells only, mixes of immune cells and potentially interfering substances, dissociated tumor cells from fresh tumor, and FFPE tumor. gDNA, genomic DNA; IHC, immunohistochemistry; TME, tumor microenvironment.

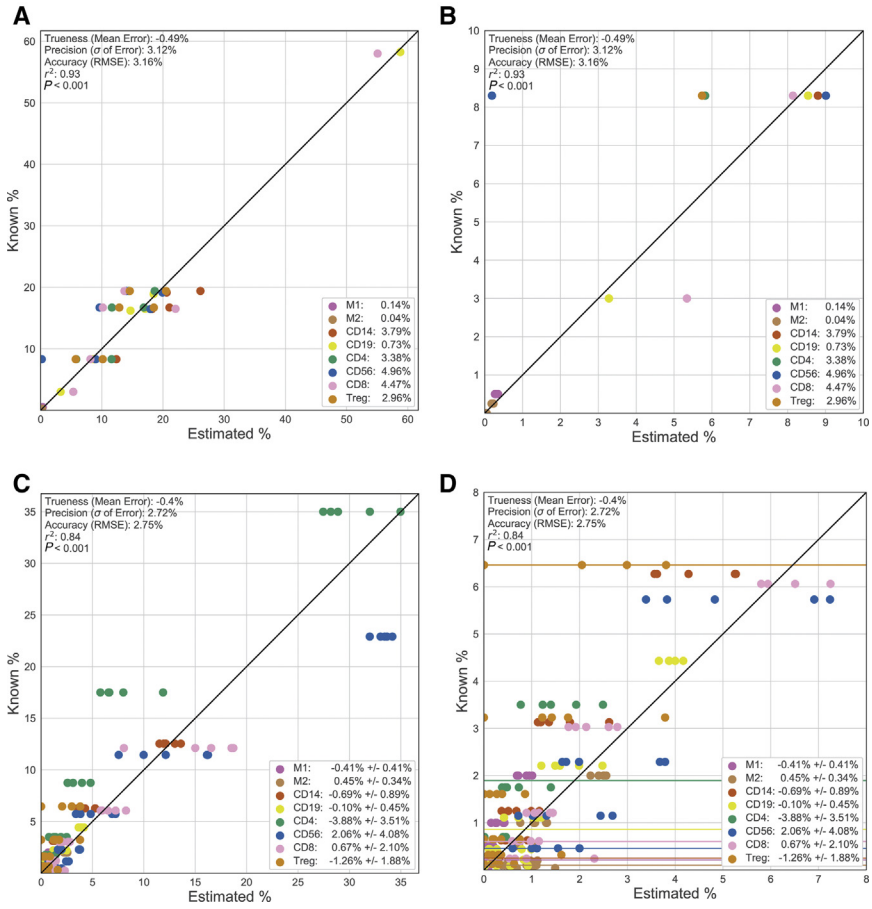

Figure 2 Assay performance compared with flow cytometry. **A:** Scatter plot of known values, from flow cytometry, is shown on the *y* axis, and values estimated by the ImmunoPrism assay are shown on the *x* axis, for each of the eight cell types for all six test mixes. The performance across all cell types is shown in the upper left of the graph. Statistics were calculated across all data points. On the bottom right, the trueness (mean error) is shown for each cell type individually. **B:** The same data are presented; however, the graph is zoomed in at the 0% to 10% range to show cell percentage estimates at ≤10%. **C** and **D:** Immune profiling with potentially interfering substances, prostate cancer 3 cell line. **C:** Scatter plot of known values, from flow cytometry, is shown on the *y* axis, and values estimated by the ImmunoPrism assay are shown on the *x* axis, for each of the eight cell types in six mixes across five technical replicates. The performance across all cell types is shown in the upper left of the graph. Statistics were calculated across all data points. On the bottom right, the trueness and precision measurements are shown for each cell type individually. **D:** The same data are presented; however, the graph is zoomed in at the 0% to 8% range to show cell percentage estimates at ≤8%. The limit of detection, or minimum known value where the precision of a cell type is retained, was calculated for each cell type and is visualized as a **horizontal line** in graph. Trueness, the mean error; precision, the first SD around the average error (σ of error); accuracy, the root mean squared error (RMSE); r^2, the coefficient of determination. $n = 30$ (**C**). Treg, regulatory T cell.

ratios of the eight cell types. Samples were generated to permute different relative ratios of cells and to cover both high and low abundance of each cell type. The percentage of each cell type (as measured by flow cytometry), as well as the amount of spiked-in M1 and M2 cells, are shown in Supplemental Table S4. The assay was optimized on 17 training samples (Supplemental Figure S2), and the performance of the assay was evaluated on 6 testing samples.

Estimations for the six testing samples are shown in Figure 2, A and B. A trueness (mean error) of −0.86% is observed across all cell types, which indicates little bias in the assay estimations. The precision (SD of error) was 3.12% and the accuracy (root mean squared error) was 3.16%. This experiment suggests that the performance of capturing individual cell types is unencumbered by the value of other cell types. Our assay correlated well to flow cytometry, with an r^2 of 0.93 and $P < 0.001$. The accuracy per cell type is as follows: M1, 0.14%; M2, 0.04%; CD14, 3.79%; CD19, 0.73%; CD4, 3.38%; CD56, 4.96%; CD8, 4.47%; and Treg, 2.96%. In Supplemental Figure S3, the relative estimation errors for training and testing sample sets is shown. These results show that the ImmunoPrism is able to differentiate between the relative amount of different immune cells. Of note, because cell presence is measured as a percentage, instead of a unitless score, one can use these measurements to directly compare the relative presence of two or more cell types (eg, CD8 versus CD4) in a single sample, as well as across samples. As the next step of validation, the performance of the assay in more complicated and clinically relevant samples is shown.

Immune Profiling in the Presence of Interfering Substances

Clinical samples have a diverse set of immune and nonimmune cell types present, including organ tissue, cancer, stroma, and others. In addition, input samples may have varying levels of relevant and nonrelevant RNA (eg, ribosomal RNA) and, if proper measures are not taken, contaminating DNA after extraction. To evaluate the performance of the ImmunoPrism assay when challenged with confounding factors, the assay was applied to samples with varying levels of interfering substances: nonimmune RNA, ribosomal RNA, and DNA.

ImmunoPrism was first tested in the presence of RNA derived from nonimmune cells and ribosomal RNA. The

performance of the assay was assessed using samples that were titrated to vary the ratio of immune and nonimmune cells. Immune cells were procured from PBMCs (six cell types) or differentiated *in vitro* (two cell types). Starting from an SM of these eight cell types, prostate cancer cell line (PC3) was titrated into the SM in increasing amounts to generate six samples (SM-PC3) with 100%, 50%, 25%, 10%, 5%, and 2% immune content. RNA was extracted from these six samples, split into quintuplicate aliquots and prepared for the ImmunoPrism assay. Samples had varying levels of ribosomal RNA content, which is discussed later. A total of 30 SM-PC3 samples were processed using ImmunoPrism and compared with known values derived by flow cytometry and controlled mixing.

Analytical Performance

Results of cell percentages estimated for all SM-PC3 samples are shown in Figure 2, C and D. For these samples, the assay has high trueness (mean error of −0.41%), high precision (SD of error of 2.72%), and high accuracy (root mean squared error of 2.75%) across all cell types. The trueness and precision per cell type is as follows: M1, −0.41% ± 0.41%; M2, 0.45% ± 0.34%; CD14, −0.69% ± 0.89%; CD19, −0.10% ± 0.45%; CD4, −3.88% ± 3.51%; CD56, 2.06% ± 4.08%; CD8, 0.67% ± 2.10%; and Treg, −1.26% ± 1.88%. The assay performs better for some cell types than others. CD4 and CD56 seem to be underestimated and overestimated, respectively. However, CD8, CD19, CD14, M1, and M2 cell types are called with low bias. Correlation remained high (r^2 of 0.83) and significant ($P < 0.001$), indicating that the assay is accurate at calling absolute percentage values, and thus practical for intrasample and intersample comparisons of cell type presence. In Supplemental Figure S4, the relative estimation error for these samples is shown as well. From these results, one can see that the assay performs well despite two potential sources of interference.

Limit of Detection

Using these same SM-PC3 estimations, the LOD was determined with the aim of characterizing how little of a cell type can be reliably measured by the assay. The LOD is defined as the minimum known value for which the precision of a cell type can be retained. For more details, see *Materials and Methods*. The LOD was calculated for each cell type and is represented in Figure 2D and detailed in Table 1. The LOD for each cell type is as follows: CD56, 0.46%; Treg, 6.46%; CD4, 1.89%; CD14, 0.25%; M1, 0.21%; M2, 0.10%; CD8 0.61%; and CD19, 0.86%. Cell types like CD14, M1, and M2 have a low LOD (≤0.25%), whereas Treg has the highest LOD (6.46%.) These LODs demonstrate the ability of our assay to reliably estimate even low presence of different immune cells.

Reportable Range

The reportable range for each cell type (Table 1) was set using the calculated LOD and the highest tested values in the SM-PC3 samples and the immune cell only samples. The reportable range for each cell type is as follows: CD56, 0.46% to 22.90%; Treg, 6.46% to 19.40%; CD4, 1.89% to 35.00%; CD14, 0.25% to 19.40%; M1, 0.21% to 2.00%; M2, 0.10% to 2.00%; CD8 0.61% to 58.00%; and CD19, 0.86% to 58.25%. Of note, the assay was optimized with 17 samples (Supplemental Figure S2) that had all individual cell types at higher relative abundance than the samples considered for determining the reportable range and so the assay is expected to also perform well at these higher levels of individual cell presence. This is especially true for cell types such as M1 and M2.

Effect of Ribosomal and Nonimmune RNA

The SM-PC3 samples contain two types of interfering substances that may be present in clinical samples: ribosomal RNA and nonimmune RNA. Ribosomal RNA is inherent to the sample because total RNA is used as the starting input to the ImmunoPrism assay. As such, ribosomal RNA contributes reads to the sequencing data. Typical samples had 0% to 25% reads aligning to the ribosome, and rarely had as high as 80% if appropriate mitigating steps are not followed (data not shown). Of interest, in the SM-PC3 samples, the ribosomal RNA was observed to be roughly inversely proportional to the total immune RNA (ie, on-target RNA) in the hybridization (results not shown). Similarly, nonimmune RNA may also have an effect on the hybridization of capture probes designed to target immune cell RNA. Clinical samples may have varying levels of total immune cells and thus varying amounts of ribosomal and nonimmune RNA. Therefore, the robustness of the assay was evaluated as a function of these potentially interfering substances.

The effect of ribosomal content on the assay is shown in Figure 3A. A null hypothesis that estimate errors are not

Table 1 Reportable Range for Individual Cell Types

Variable	CD56	Treg	CD4	CD14	M1	M2	CD8	CD19
Limit of detection, %	0.46	6.46	1.89	0.25	0.21	0.10	0.61	0.86
Maximum level of detection, %	22.90	19.40	35.00	19.40	2.00	2.00	58.00	58.25

Treg, regulatory T cell.

correlated with ribosomal content was tested. For all cell types, the P values are as follows: M1, 0.225; M2, 0.563; CD14, 0.505; CD19, 0.709; CD4, 0.474; CD56, 0.061; CD8, 0.034; and Treg, 0.421. All these cell types except CD8, $P > 0.05$, and so the null hypothesis cannot be rejected. Bolstering this result are slopes close to 0 and small r^2 values. This result suggests that ribosomal content cannot predict the error. Therefore, one can conclude that the assay performs independent of the ribosomal content.

Figure 3B shows the assay's error, but now as a function of nonimmune content in the sample. A similar statistical experiment is performed to understand if error is correlated with nonimmune content. For all cell types except M2 and CD14, we see significant correlations, with $P < 0.05$. The slopes of all correlations are as follows: M1, −0.012; M2, −0.003; CD14, −0.002; CD19, −0.008; CD4, −0.052; CD56, 0.096; CD8, 0.03; and Treg, −0.048. Despite significant correlations, most cell types have small slopes, which indicate a minimal impact. The largest relationship is seen between error and the ratio of immune and nonimmune content in CD56. Overall, these ribosomal and nonimmune results suggest a limited to no effect of nonimmune content to our estimations, down to the LOD of individual cell types.

Effect of gDNA

The first step of the ImmunoPrism protocol, when working from FFPE cancer samples, is to extract or purify total RNA from the sample. During this procedure and under certain conditions (eg, using more FFPE material than recommended or insufficient DNase treatment), genomic DNA (gDNA) species may contaminate the purified RNA preparation. To better understand the effect of gDNA contamination on cell percentage calls, four FFPE cancer samples were generated with increasing ratio of gDNA versus total RNA. gDNA was mixed with FFPE cancer sample RNA, which was devoid of gDNA, in increasing percentages of 0%, 5%, 10%, 20%, and 40%. As expected, increasing gDNA had an individual effect on cell percentage estimates at higher ranges of gDNA contamination (Supplemental Figure S5). Thus, the threshold for maximum gDNA contamination is currently set at 10%. Of 145 recent FFPE samples prepared for the ImmunoPrism assay in our laboratory, 123 (approximately 85%) of the RNA samples exhibited <10% genomic DNA (mean, 4.89%; median, 1.22%). For the samples that exhibited gDNA contamination >10%, a second DNase treatment was performed, resulting in 144 (approximately 99%) of the samples with gDNA contamination below threshold (data not shown).

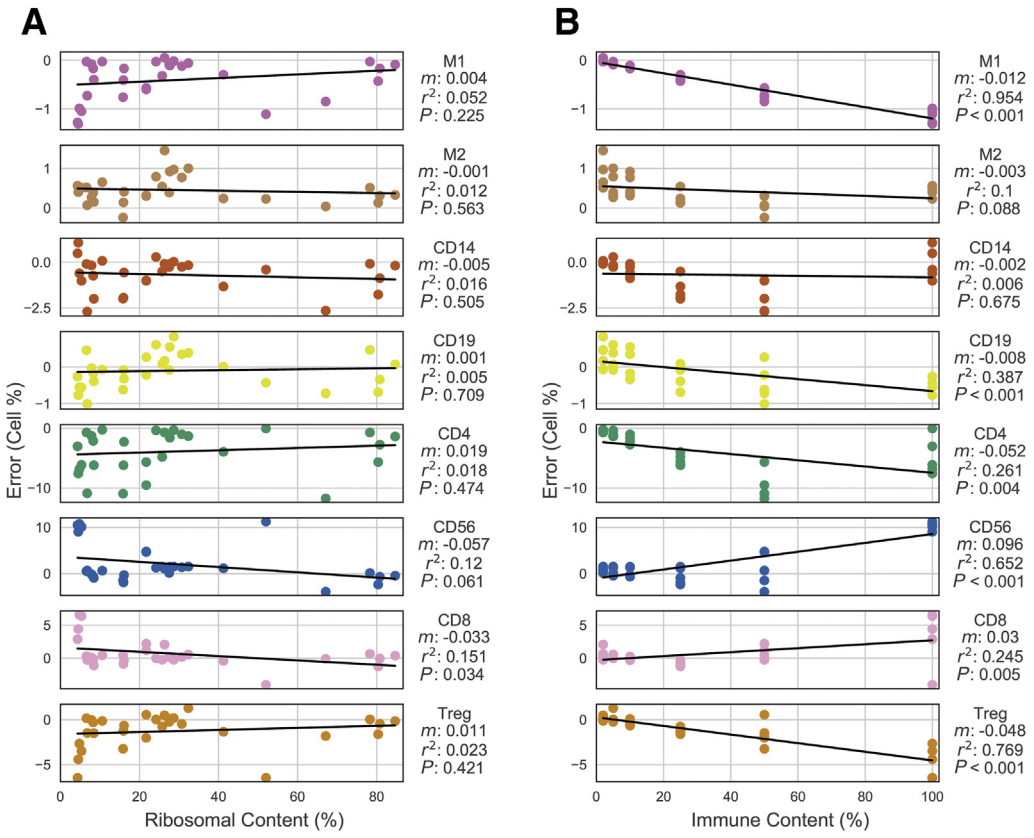

Figure 3 Immune profiling with potentially interfering substances, ribosomal RNA and nonimmune content. **A:** The estimation error of measured cell percentages is shown on the *y* axis, and the ribosomal content is shown on the *x* axis, for each cell type across supermix prostate cancer 3 cell line (SM-PC3) samples. **B:** The estimation errors of measured cell percentages are shown on the *y* axis, and the percentage of immune content is shown on the *x* axis, for each cell type across SM-PC3 samples. The **black line** is the best fit to the data. This fit was statistically tested to see if it was significantly different than a slope of 0. Slope (m), coefficient of determination (r^2), and *P* value are shown to the right of each cell type subplot. $n = 30$ (**A** and **B**).

The experiments in this section show the performance of the assay under a wide range of cell values and with several controlled sources of potential interference. In the next section, the ability of the assay to measure the immune profile at the site of a tumor is validated.

Immune Profiling of Tumor Environment

The ImmunoPrism assay measures the presence of various adaptive and innate immune cells in the tumor environment. To investigate the performance of the assay when profiling tumor samples, dissociated tumor cells were quantified with ImmunoPrism. This type of sample starts as a piece of fresh tumor and is processed to isolate single cells. Dissociated tumor cells serve as an approximation of clinical FFPE tumor samples, while still retaining the ability to measure cell type abundance with flow cytometry.

The performance of ImmunoPrism was evaluated by estimating the presence of immune cells in three dissociated tumor cell samples originating from ovarian adenocarcinoma, lung adenocarcinoma, and melanoma tumor tissue. Results for these samples across five cell types are shown in Figure 4A. Please see *Materials and Methods* for details on how limitations with flow cytometry prevented the evaluation of the other three cell types. The accuracy (root mean squared error) across all indications and cell types was 8.12%, and the accuracies per indication were as follows: lung, 7.40%; melanoma, 9.25%; and ovarian, 7.70%. The accuracies per cell type were as follows: CD19, 7.40%; CD4, 8.37%; CD56, 0.40%; CD8, 10.56%; and Treg, 7.02%. The overall accuracy (root mean squared error) of these samples was lower than SM-PC3 (8.12% versus 2.75%). Regardless, these results corroborate previous high accuracies measured and demonstrate that ImmunoPrism is able to identify and measure the presence of immune cells that have infiltrated a tumor.

Immune Profiling of FFPE Samples

The current gold standard method for storing and preserving tumor samples is FFPE treatment. Countless cancer samples, with extensive clinical data, are currently preserved this way and are inputs to most retrospective studies. ImmunoPrism was optimized for FFPE treated RNA and as such can evaluate archived FFPE samples. This may enable clinicians and researchers to better understand the immune component of the tumor microenvironment in all their archived samples.

To demonstrate the effectiveness of ImmunoPrism with FFPE samples, 28 samples of various tissue/cancer types were quantified using ImmunoPrism. Supplemental Table S6 lists these samples, detailing the disease, percentage tumor cellularity, DV_{200}, total RNA extracted, whether IHC data are available, and whether the samples passed molecular quality control. In all, 27 of 28 ($\approx$96%) passed molecular quality control (see *Materials and Methods*). Sample F00065618 failed molecular quality control because library generation did not produce enough material to perform a hybrid capture. If RNA is highly modified, as happens as a result of formalin fixation, it will be more inaccessible to reverse transcription, resulting in lower precapture library yield. Therefore, the likely reason a sample may fail library construction is formalin overfixation.

Comparison to Immunohistochemistry

In 10 of these FFPE samples, CD8, CD19, and Treg cell percentages were measured using IHC and compared with immune cell percentages reported by ImmunoPrism. These results are shown in Figure 4, B–D. CD8 and CD19 significantly ($P < 0.001$) and strongly (r^2 of 0.81 and 0.85, respectively) correlate with IHC. Treg exhibits a lower correlation (r^2 of 0.34) with lower significance ($P = 0.078$). It is notable that all linear regressions have slopes close to that of the doubling slope (ie, when IHC cell counts double, the percentage of estimated cells doubles). These results show that ImmunoPrism estimations correlate well with IHC measurements, suggesting that ImmunoPrism can accurately profile the presence of immune cells in FFPE samples.

Reproducibility

Assay variability can occur in different areas of the protocol or analytical pipeline. This variability reduces the ability for the assay to discern true biological variability across samples and minimizes the reproducibility of past estimations. To assess the reproducibility of the ImmunoPrism assay, RNA was extracted for one of the FFPE samples (F00020350) and split into six replicates. Two people each processed three replicates on different days and on different sequencing machines. In Figure 5A, the differences in estimates for these replicates is shown. The SD for each of the cell types is as follows: CD14, 0.09%; M1, 0.04%; CD56, 0.22%; M2, 0.15%; CD4, 1.61%; Treg, 3.20%; CD19, 3.56%; and CD8, 4.13%. Cell types with higher means (CD8, CD19, and Treg) seem to have a higher variance than those with low means (CD4, M2, CD56, M1, and CD14). Table 2 shows the reproducibility of the assay by the same operator, and between operators on different machines. Among replicates performed by the same operator, the average SD across all cell types is 0.70% and 3.65%. Similarly, when comparing the replicates performed by two different operators on two different machines on two different days, the average difference of SDs across all cell types is 2.65%. These results indicate a high level of reproducibility for the assay.

The experiments in this section show how the assay correlates with IHC and performs reliably in FFPE. In the next section, the controls for ImmunoPrism are described.

Controls

Control materials were developed to ensure the ImmunoPrism assay is performing as expected, regardless of test sample

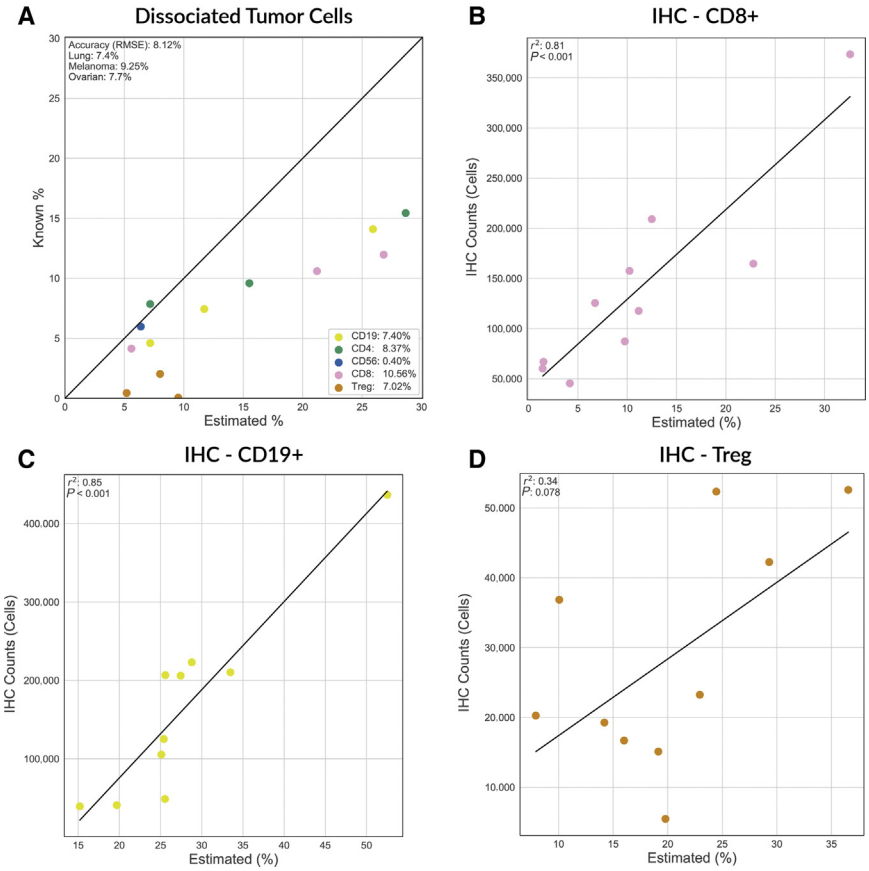

Figure 4 Validation of assay using dissociated tumor cells (DTCs) from tumors and performance comparison to immunohistochemistry (IHC). **A:** Cell percentage measurements of DTCs from three cancer types (lung adenocarcinoma, melanoma, and ovarian adenocarcinoma) were assessed using the ImmunoPrism assay and compared with flow cytometry of the same samples. The known value from flow cytometry is shown on the *y* axis, and the estimated value from the assay is shown on the *x* axis. In the upper left, the accuracy [calculated as the root mean squared error (RMSE)] is shown over all cells for all samples and for each disease individually (lung adenocarcinoma, melanoma, and ovarian adenocarcinoma). The accuracy for individual cell type measurements, across disease types, is shown in the lower right. The **black line** represents the slope of 1. **B**–**D:** Cell percentage measurements from ImmunoPrism compared with IHC for three cell types [CD8, CD19, and regulatory T cell (Treg)] across 10 formalin-fixed, paraffin-embedded samples. Cellular IHC counts are shown on the *y* axis, and the estimated cell percentage from the assay is on the *x* axis. The **black line** represents the best fit regression line of the data. The coefficient of determination (r^2) of this fit is presented in the upper left corner. These fits were statistically tested to see if they are significantly different than a slope of 0, and the *P* value is shown in the upper left corner.

composition. Human spleen RNA was chosen because all eight cell types estimated their LOD on average. In addition, the human spleen RNA was fragmented to allow processing in an identical manner to degraded FFPE RNA.

The reproducibility of the fragmented spleen RNA was measured to confirm it as a suitable positive control. To do this, six replicate samples were prepared and processed by two different operators on two different days. In Figure 5B, the differences in estimations are shown. The SD for each of the cell types is as follows: M1, 0.13%; M2, 0.13%; Treg, 0.45%; CD8, 0.35%; CD4 0.58%; CD56, 0.63%; CD14, 1.18%; and CD19, 1.14%. The deviation of samples across all cell types is low. In addition, the mean measurements of all cell types except Treg are above the respective LODs. Therefore, fragmented spleen RNA is an appropriate positive control because ImmunoPrism reproducibly measures all eight cell types.

It is challenging to develop a true negative control for targeted capture molecular methods because oligo capture probes are designed to capture and enrich for targets with high sequence homology to the probes. Moreover, the success of the molecular steps, following capture (before sequencing), relies on the release of genetic material from the capture probes. A true negative control (containing no sequence homology to the oligo probes in the capture) would end up releasing no material and subsequently failing library generation. Our goal was to develop a negative control that could be employed through the molecular and analytic methods with appropriate outputs. The commercially available ALK-RET-ROS1 Fusion Negative FFPE sample was chosen as a negative control because it lacks immune cell content but has sufficient on-target transcripts to generate successful libraries and sequencing data. To confirm the suitability of ALK-RET-ROS1 Fusion Negative FFPE material as a negative control, three replicate samples were prepared and quantified by the assay on different days. Figure 5C shows the differences in estimates for these samples for each of the cell types. The mean estimated values for the cell types are at or close to 0%: M1, 0.00%; CD4, 0.04%; M2, 0.08%; Treg, 0.56%; CD14, 0.64%; CD8, 0.69%; CD19, 1.23%; and CD56, 1.24%. Of the eight cell types, four (M1, CD4, M2, and Treg) exhibited mean cell percentages at or lower than the respective LOD. These characteristics confirm the suitability of ALK-RET-ROS1 Fusion Negative FFPE RNA as a negative control.

Discussion

We believe that the analytical validation approach presented in this work potentially serves as a reference for validation of other intricate RNA-based assays. The progressive nature of this approach allows one to characterize the analytical

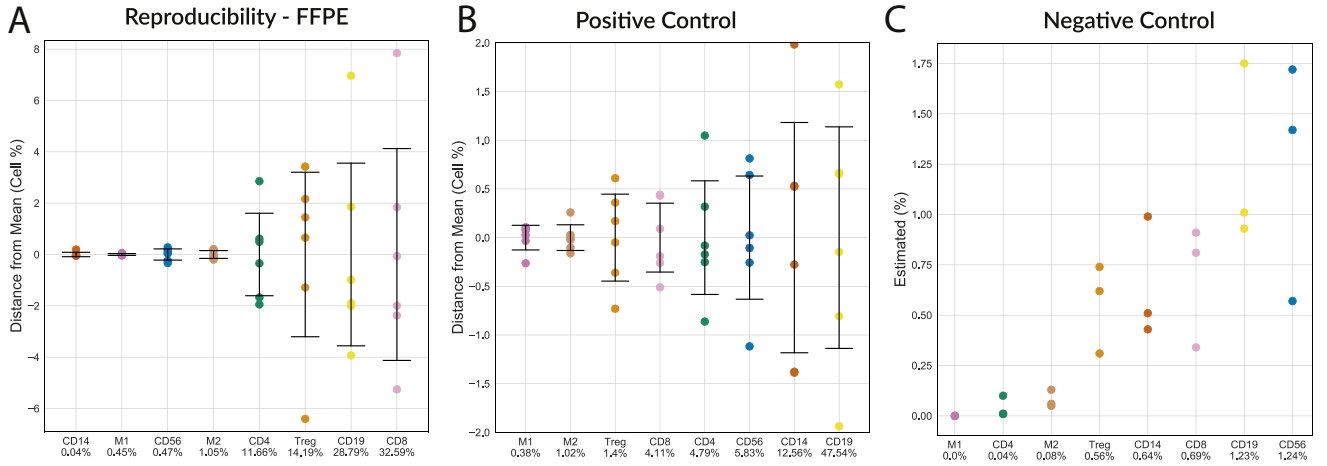

Figure 5 Reproducibility and control materials. **A:** Reproducibility of estimates made with ImmunoPrism. RNA from a formalin-fixed, paraffin-embedded (FFPE) sample (F00020350) was split into six replicates after extraction. Replicates were divided and processed by two people on different days and on different sequencing machines. The *y* axis shows the mean centered estimates for each of the cell replicates. Each cell type is listed on the *x* axis. **B:** Reproducibility of the fragmented spleen RNA positive control. RNA from the standard control sample was split into six replicate samples and processed with the assay by two different operators on 6 different days. Results are presented as they are in **A**. **C:** Reproducibility of the ALK-RET-ROS1 Fusion Negative FFPE material as a negative control. RNA from the negative control samples was split into three replicate samples and processed with the assay on different days. The *y* axis shows the estimated percentage reported by the assay for each of the cell types on the *x* axis. **A** and **B:** For each cell type, the **short, black horizontal lines** connected by a vertical line represent the upper and lower first SDs. The mean values for each cell type are displayed in the *x*-axis tick labels. **A**–**C:** Cell types on the *x* axis are sorted by their mean estimated values from low to high. Treg, regulatory T cell.

performance of the assay in well-controlled samples, and iteratively corroborate with samples more akin to FFPE tumor samples. This approach addresses several challenges of evaluating RNA-based assays on primary tissue. For example, dissociated tumor cell samples are useful because they may contain immune cells from a tumor that can be quantified via flow cytometry. However, it is infeasible to procure sufficient dissociated tumor cell samples to adequately characterize performance, such as precision and LOD, across a diverse set of immune cell combinations. Similarly, FFPE samples represent the typical input for the assay, but the immune content can only be orthogonally validated by methods such as IHC. IHC must be used because FFPE processing of cancer samples prevents validation by flow cytometry. Even with IHC, there are challenges in orthogonal validation because it is infeasible to use IHC and the ImmunoPrism assay on the same slice of FFPE tissue. Two slices of a heterogeneous tissue block must be used, and despite being adjacent, can have different populations of immune cells. This fact limits our ability to draw strong conclusions about ImmunoPrism performance solely on comparisons to IHC. The analytical validation presented attempts to mitigate the challenges of testing RNA-based assays on primary tissue, and we hope it may serve as a roadmap to researchers interested in future validation of complex RNA-based clinical assays.

Table 2 Intraoperator and Interoperator Variability of the ImmunoPrism Assay

Cell	Operator 1 mean, %	Operator 1 SD, %	Operator 2 mean, %	Operator 2 SD, %	Difference of SDs, %
CD4	12.21	2.27	11.10	1.29	0.97
CD8	31.11	1.24	34.06	6.56	5.32
CD19	27.16	0.56	30.42	5.45	4.89
CD14	0.08	0.13	0.00	0.00	0.13
CD56	0.64	0.12	0.31	0.22	0.09
M1	0.46	0.04	0.44	0.04	0.00
M2	1.11	0.22	0.99	0.10	0.13
Treg	16.54	1.00	11.85	3.65	2.65
Mean		0.70		3.65	2.65

Treg, regulatory T cell.

This assay was designed to specifically profile the immune response by reporting the relative presence of eight different immune cells. Other cell types that would be valuable to measure include mast cells, neutrophils, dendritic cells, and more. Similarly, minor immune cell subtypes (eg, $CD8^+$ effector memory and $CD8^+$ central memory) would be potentially valuable to consider in future versions of ImmunoPrism. Our assay is complementary to other pieces of information about the immune system. It has been shown that characterizing T cell receptor diversity and clonality can help inform how well a host has identified a cancer. In addition, tumor mutational burden has been implicated in clinical outcomes using checkpoint inhibitors.[29] Assays that measure T cell receptor diversity and clonality or tumor mutational burden could be complementary to ImmunoPrism in describing the state of the immune response in a patient. Interestingly, all these different characteristics tell a different part of the story of immune response. It is exciting to think about the potential

of bringing these disparate measurements together to form a single, multidimensional biomarker. This could be powerful to use for predictive and prognostic use for particular therapies, particularly checkpoint inhibitors.

ImmunoPrism was designed to identify and measure the abundances of leukocytes in and around the tumor, with the goal to facilitate immuno-oncology research and treatment. However, information about immune cell presence could be useful in other areas of human health as well. For example, understanding the presence of different immune cells could help inform dysfunctional or autoreactive immune response in diseases such as rheumatoid arthritis, inflammatory bowel disease, and psoriasis, as well as conditions such as asthma and allergies. For these diseases and conditions, it may be less practical to consider tissue samples, but the immune response could be profiled from the blood. Conceptually, RNA harvested from blood should be of higher quality than RNA from FFPE tissues and so should be an appropriate input to ImmunoPrism. Regardless, ImmunoPrism may prove to be useful in any health condition that has an immune component.

In this work, the analytical validation of the ImmunoPrism assay is presented. This is the first validation of a clinical assay that uses multidimensional RNA models of immune cells to identify immune cells in the tumor microenvironment. The assay was characterized in controlled samples and was shown to have high trueness, precision, accuracy, and reproducibility. The assay continued to perform well in the presence of potentially interfering substances. Experiments with dissociated tumor cells, which mimic the tumor microenvironment, and FFPE samples corroborated the analytical performance of the assay. In all, experimental results of controlled and uncontrolled samples show that ImmunoPrism is a robust assay for measuring the presence of eight immune cell types in FFPE tissue.

Acknowledgments

We thank the entire Cofactor Genomics, Inc. team for expertise and support in executing and writing this analytical validation.

Author Contributions

The study was designed by I.S., J.R.A., J.T.F., D.N.M., and E.J.D.; sample procurement and quality control was performed by J.T.F., J.H., and K.C.F.; individual experiments were planned and supervised by J.R.A. and J.H.; experiments were performed by J.T.F., J.H., and K.C.F.; data analysis was performed by I.S., J.E., and T.C.; the manuscript was written by I.S. and J.R.A.; the study concept and manuscript were critically reviewed by J.T.F., J.H., K.C.F., J.I.G., and E.J.D.; and all authors read and approved the final manuscript.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2020.01.009*.

References

1. Hanahan D, Weinberg RA: The hallmarks of cancer. Cell 2000, 100: 57–70
2. Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. Cell 2011, 144:646–674
3. Gooden MJM, de Bock GH, Leffers N, Daemen T, Nijman HW: The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. Br J Cancer 2011, 105:93–103
4. Denkert C, Loibl S, Noske A, Roller M, Müller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, von Törne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel M, von Minckwitz G: Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 2010, 28:105–113
5. Lee HJ, Seo J-Y, Ahn J-H, Ahn S-H, Gong G: Tumor-associated lymphocytes predict response to neoadjuvant chemotherapy in breast cancer patients. J Breast Cancer 2013, 16:32
6. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al: Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors. Adv Anat Pathol 2017, 24:311–335
7. Fortes C, Mastroeni S, Mannooranparampil TJ, Passarelli F, Zappalà A, Annessi G, Marino C, Caggiati A, Russo N, Michelozzi P: Tumor-infiltrating lymphocytes predict cutaneous melanoma survival. Melanoma Res 2015, 25:306–311
8. Gong J, Chehrazi-Raffle A, Reddi S, Salgia R: Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. J Immunother Cancer 2018, 6:8
9. Darvin P, Toor SM, Sasidharan Nair V, Elkord E: Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med 2018, 50:165
10. Marconcini R, Spagnolo F, Stucci LS, Ribero S, Marra E, Rosa F De, Picasso V, Di Guardo L, Cimminiello C, Cavalieri S, Orgiano L, Tanda E, Spano L, Falcone A, Queirolo P; Italian Melanoma Intergroup (IMI) for the IMI: Current status and perspectives in immunotherapy for metastatic melanoma. Oncotarget 2018, 9:12452–12470
11. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V, West AN, Carmona M, Kivork C, Seja E, Cherry G, Gutierrez AJ, Grogan TR, Mateus C, Tomasic G, Glaspy JA, Emerson RO, Robins H, Pierce RH, Elashoff DA, Robert C, Ribas A: PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014, 515:568–571
12. Herbst RS, Soria J-C, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, Kohrt HEK, Horn L, Lawrence DP, Rost S, Leabman M, Xiao Y, Mokatrin A, Koeppen H, Hegde PS, Mellman I, Chen DS, Hodi FS: Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature 2014, 515:563–567
13. Newick K, O'Brien S, Moon E, Albelda SM: CAR T cell therapy for solid tumors. Annu Rev Med 2017, 68:139–152
14. Ö Met, Jensen KM, Chamberlain CA, Donia M, Svane IM: Principles of adoptive T cell therapy in cancer. Semin Immunopathol 2019, 41: 49–58

15. Guo C, Manjili MH, Subjeck JR, Sarkar D, Fisher PB, Wang X-Y: Therapeutic cancer vaccines. Adv Cancer Res 2013, 119:421–475
16. Banchereau J, Palucka K: Immunotherapy: cancer vaccines on the move. Nat Rev Clin Oncol 2017, 15:9–10
17. Maude SL, Laetsch TW, Buechner J, Rives S, Boyer M, Bittencourt H, Bader P, Verneris MR, Stefanski HE, Myers GD, Qayed M, De Moerloose B, Hiramatsu H, Schlis K, Davis KL, Martin PL, Nemecek ER, Yanik GA, Peters C, Baruchel A, Boissel N, Mechinaud F, Balduzzi A, Krueger J, June CH, Levine BL, Wood P, Taran T, Leung M, Mueller KT, Zhang Y, Sen K, Lebwohl D, Pulsipher MA, Grupp SA: Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. N Engl J Med 2018, 378:439–448
18. Anagnostou VK, Welsh AW, Giltnane JM, Siddiqui S, Liceaga C, Gustavson M, Syrigos KN, Reiter JL, Rimm DL: Analytic variability in immunohistochemistry biomarker studies. Cancer Epidemiol Biomarkers Prev 2010, 19:982–991
19. Micke P, Johansson A, Westbom-Fremer A, Backman M, Djureinovic D, Patthey A, Isaksson-Mettävainio M, Gulyas M, Brunnstrom H: PD-L1 immunohistochemistry in clinical diagnostics: inter-pathologist variability is as high as assay variability. J Clin Oncol 2017, 35:e20637
20. Conroy JM, Pabla S, Glenn ST, Burgher B, Nesline M, Papanicolau-Sengos A, Andreas J, Giamo V, Lenzo FL, Hyland FCL, Omilian A, Bshara W, Qin M, He J, Puzanov I, Ernstoff MS, Gardner M, Galluzzi L, Morrison C: Analytical validation of a next-generation sequencing assay to monitor immune responses in solid tumors. J Mol Diagn 2018, 20:95–109
21. Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, Geller MA, Odunsi K, Beechem J, Fling SP: Gene expression markers of tumor infiltrating leukocytes. J Immunother Cancer 2017, 5:18
22. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015, 12:453–457
23. Saah AJ, Hoover DR: "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. Ann Intern Med 1997, 126:91–94
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013, 29:15–21
25. Anders S, Pyl PT, Huber W: HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics 2015, 31:166–169
26. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014, 15:550
27. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, D'Eustachio P, Stein L, Hermjakob H: Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 2017, 18:142
28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP: Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011, 27:1739–1740
29. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R: Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther 2017, 16:2598–2608