Application Note

Genomics

Agilent

Trusted Answers

Agilent SureSelect Targeted RNA-Seq Paired With iSort Enables Robust Enumeration of Cell Type Composition in Solid Tissue Samples

Author

Aki Nakao[1]
Aaron M. Newman[1,2]
Ash A. Alizadeh[1,2]
Maximilian Diehn[1,2]
Manjula Aliminati[3]
Mistuni Ghosh[3]
Jayati Ghosh[3]
Kristi Stephenson[3]
Ashutosh Ashutosh[3]

1. CiberMed, Inc., California, USA
2. Stanford University, Stanford, California, USA
3. Agilent Technologies Inc., California, USA

Abstract

CiberMed's iSort software suite is a widely used digital cytometry solution for profiling the cellular composition of complex tissues from bulk RNA-sequencing (RNA-seq) data. iSort Fractions 'tissue mode' quantifies the relative fractions of 22 functionally defined hematopoietic subsets along with three non-immune cell types—fibroblasts, endothelial cells, and epithelial cells—from bulk tissue expression data. To enable accurate enumeration of these 25 cell types from fresh/frozen (FF) or formalin-fixed paraffin-embedded (FFPE) samples, iSort Fractions includes specialized methods to remove noise and technical variation across diverse platforms and sample preservation conditions. In this study, we assess the accuracy, robustness, and reproducibility of the Agilent SureSelect CD CiberMed Tissue panel using iSort Fractions. This panel was developed to enable highly efficient and reliable enumeration of 25 cell types from FF or FFPE tissues.

Total RNA from paired FF and FFPE tumor specimens was extracted and prepared for sequencing using Agilent SureSelect XT HS2 RNA reagents and for target enrichment using the SureSelect CD CiberMed Tissue panel. The resulting sequencing data were analyzed using iSort Fractions tissue mode. Across all evaluable phenotypes (n = 25), cell type fractions determined by iSort were highly concordant between paired FF and FFPE tumor samples (rho = 0.93). This substantially outperformed whole-transcriptome profiling (rho = 0.78) while requiring approximately 20-fold less sequencing. Additionally, strong reproducibility was observed between technical replicates profiled by targeted sequencing (rho ≥ 0.97 for FF and rho ≥ 0.96 for FFPE). The SureSelect CD CiberMed Tissue panel is available through the Agilent Community Design program to enable highly accurate cellular profiling of fresh/frozen and fixed tissue specimens with iSort.

Introduction

Cellular heterogeneity and complex intercellular interactions underlie diverse physiological and pathological states, including various malignancies. Therefore, it is critically important to study the phenotypic and genotypic composition of cell subsets within the diseased milieu. It is also essential to monitor changes in their relative abundances during disease progression and in response to therapy. The importance of studying cellular heterogeneity and composition within the tumor microenvironment (TME) is well established.[1-3] Enumerating cell type composition has prognostic value and holds great promise as a potential predictive biomarker for therapy response.[4,5] Therefore, traditional methods such as flow/mass cytometry,[6,7] immunohistochemistry (IHC),[8] and immunofluorescence (IF)[9] are routinely employed for quantifying and characterizing tissue heterogeneity (Table 1). However, these methods can only interrogate a modest number of markers and there is often a trade-off between the number of markers that can be measured and the throughput of the assay. Cytometry by time-of-flight (CyTOF)[10] is destructive to the sample and does not enable co-interrogation of cell type fractions and cell type expression profiles across thousands of genes. In more recent years, single-cell RNA sequencing (scRNA-seq) approaches have been embraced as a means of characterizing cell type composition and gene expression at the single-cell level. However, scRNA-seq is limited by sample preparation artifacts, including dissociation-induced distortions in cellular composition, and remains cost prohibitive for large-scale cellular profiling.[11] Similarly, FFPE tissue specimens, which are collected as part of routine clinical care, cannot be dissociated into a cell suspension without disruption of cell type composition.

Given these limitations for characterizing FF and FFPE samples, deconvolution algorithms for determining cell type abundances from bulk tissue expression profiles have gained traction.[1,2,12-16] These methods enable dissection of cell-type-specific signals from bulk sequencing data. Comparative analyses of scRNA-seq, deconvolution of bulk expression data, and IHC revealed that deconvolution is free from artifacts arising from cell separation and tissue dissociation.[17] Of these deconvolution methods, CIBERSORT and CIBERSORTx, have emerged as robust and accurate tools for determining cell type proportions from blood and tissue samples.[3,17-28] In fact, *CIBERSORTx was recently identified as one of the five fastest-growing software tools in the biosciences.*[29]

CiberMed further optimized and standardized CIBERSORTx with proprietary enhancements and validated the improved algorithms with different sequencing platforms and sample types, including FF and FFPE preservation states. CiberMed currently offers two algorithms by way of two flagship products within the iSort digital cytometry suite (Figure 1). iSort Fractions reliably determines cell subset abundance from bulk tissue or blood expression data and iSort HiRes infers cell-type-specific gene expression profiles from bulk tissue or blood expression data.

Table 1. Comparison of cell profiling methods.

	Software-Based Deconvolution	Single-Cell RNA-Seq	Flow Cytometry/CyTOF	IHC
Throughput	+++++	+	++	+++
Requires Tissue Dissociation	No	Yes	Yes	No
Artifacts Introduced by Dissociation	No	Yes	Yes	No
Workflow Simplicity	+++++	+	++	++++
Manual Data Analysis Required	No	Yes	Yes	Yes
Number of Cell Types/Cell Type Resolutions	+++++	+++++	+++	+++

iSort Fractions 'blood mode' quantifies cell type abundances from bulk RNA-seq data by applying the well-established LM22 leukocyte gene signature matrix to distinguish 22 human hematopoietic cell subsets. LM22 has been validated in pure leukocyte subset titrations, blood samples, and tumors from multiple cancer types. CiberMed's iSort Fractions tissue mode uses the LM22 signature matrix plus tissue-specific signature profiles to quantify 25 cell types, including 22 immune subsets, fibroblasts, endothelial cells, and epithelial cells, from bulk tissue RNA-seq profiles of FF or FFPE biospecimens (see Table 2 for the complete list of cell subsets).[3,27,30,47,48,49]

In this application note, we demonstrate robust, accurate, and reproducible enumeration of cell subsets in solid tumor samples using iSort Fractions combined with the Agilent SureSelect target enrichment solution (Figure 2). The SureSelect CD CiberMed Tissue panel targets genes from the LM22 signature matrix along with additional genes for discriminating fibroblasts, endothelial cells, and epithelial cells with iSort Fractions tissue mode (see Table 3 for panel details). Four pairs of matched FF and FFPE samples from non-small cell lung cancer (NSCLC) tumor biopsies were included in the assessment. SureSelect XT HS2 RNA libraries were prepared in triplicate from total RNA extracted from each sample and enriched with the SureSelect CD CiberMed Tissue panel.

Table 2. Human cell types quantified by blood mode (22 cell types) and tissue mode (25 cell types) of iSort Fractions.

Parent Subsets	Cell Type Description	Blood	Tissue
B cells	B cells naïve	■	■
	B cells memory	■	■
Plasma cells	Plasma cells	■	■
CD8 T cells	T cells CD8	■	■
CD4 T cells	T cells CD4 naïve	■	■
	T cells CD4 memory resting	■	■
	T cells CD4 memory activated	■	■
	T cells follicular helper	■	■
	T cells regulatory (Tregs)	■	■
Gama delta T cells	T cells gamma delta	■	■
NK cells	NK cells resting	■	■
	NK cells activated	■	■
Monocytes and Macrophages	Monocytes	■	■
	Macrophages M0	■	■
	Macrophages M1	■	■
	Macrophages M2	■	■
Dendritic cells	Dendritic cells resting	■	■
	Dendritic cells activated	■	■
Mast cells	Mast cells resting	■	■
	Mast cells activated	■	■
Eosinophils	Eosinophils	■	■
PMNs	Neutrophils	■	■
	Fibroblasts		■
	Endothelial cells		■
	Epithelial cells		■

Table 3. Comparison of panel designs and the intended application for each.

	SureSelect CD CiberMed Tissue	SureSelect CD CiberMed Heme[†]
Intended Application	Enumeration of 25 cell subsets (Table 2) with iSort Fractions	Enumeration of 22 immune cell subsets (Table 2) with iSort Fractions
Sample Type	FF, FFPE	Whole blood, PBMCs
Signature Matrix	LM22 (immune subsets) and TR4 (non-immune cell types)	LM22 (immune cell types)
Number of Targets	1,423 genes	547 genes
Total Capture Size	5.7 Mb	1.8 Mb
Recommended Min Reads/Sample	2 M (1 M × 2, 150 bp)	500 k (250 k × 2, 150 bp)

[†]An application note detailing the Agilent SureSelect CD CiberMed Heme panel is available at www.agilent.com/cs/library/applications/ap-isort-sureselect-5994-6964en-agilent.pdf.

Across all 25 evaluable cell types, high concordance was observed between matched FF and FFPE samples. The concordance was markedly stronger for targeted RNA-seq with the SureSelect CD CiberMed Tissue panel than for whole-transcriptome sequencing applied to the same samples (Figure 3). There was also high reproducibility between technical replicates (Figure 4). Because of several proprietary enhancements, we found that iSort Fractions outperformed CIBERSORTx for the enumeration of cell type composition (Figure 3).

Notably, targeted enrichment with the SureSelect CD CiberMed Tissue panel reduced the sequencing requirement to only 2M (1M × 2) input reads per sample, as opposed to 40M (20M × 2) input reads per sample for whole-transcriptome sequencing, a reduction of 20-fold. Together, these results underscore the accuracy, reliability, and cost-effectiveness of the combined SureSelect/iSort assay for cell type abundance profiling from solid tissue samples.

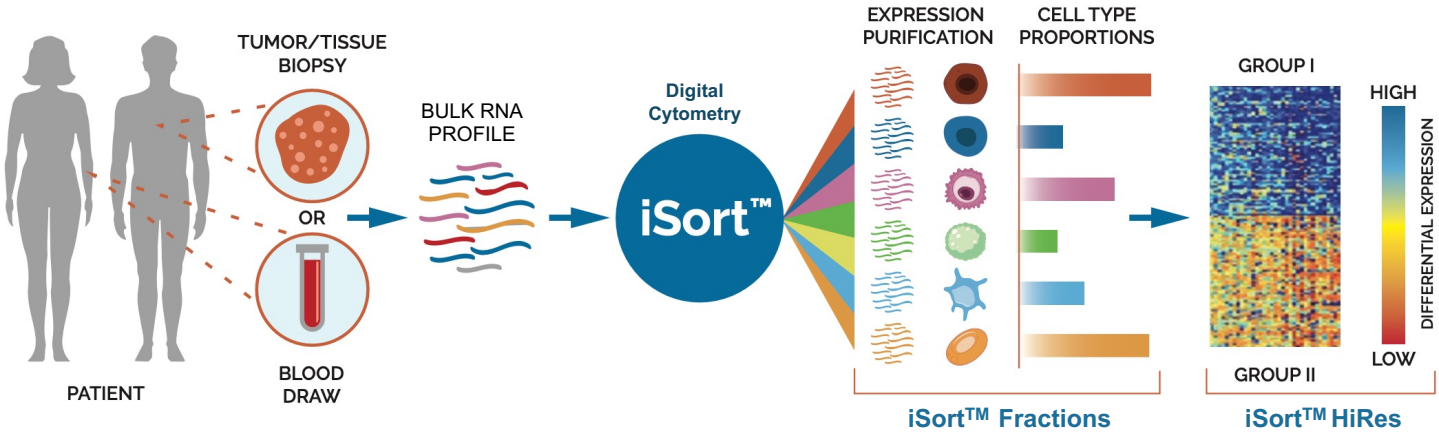

Figure 1. Schematic of iSort algorithms for digital cytometry. iSort Fractions determines cell type proportions and iSort HiRes determines cell-type-specific gene expression profiles from bulk tissue expression data.

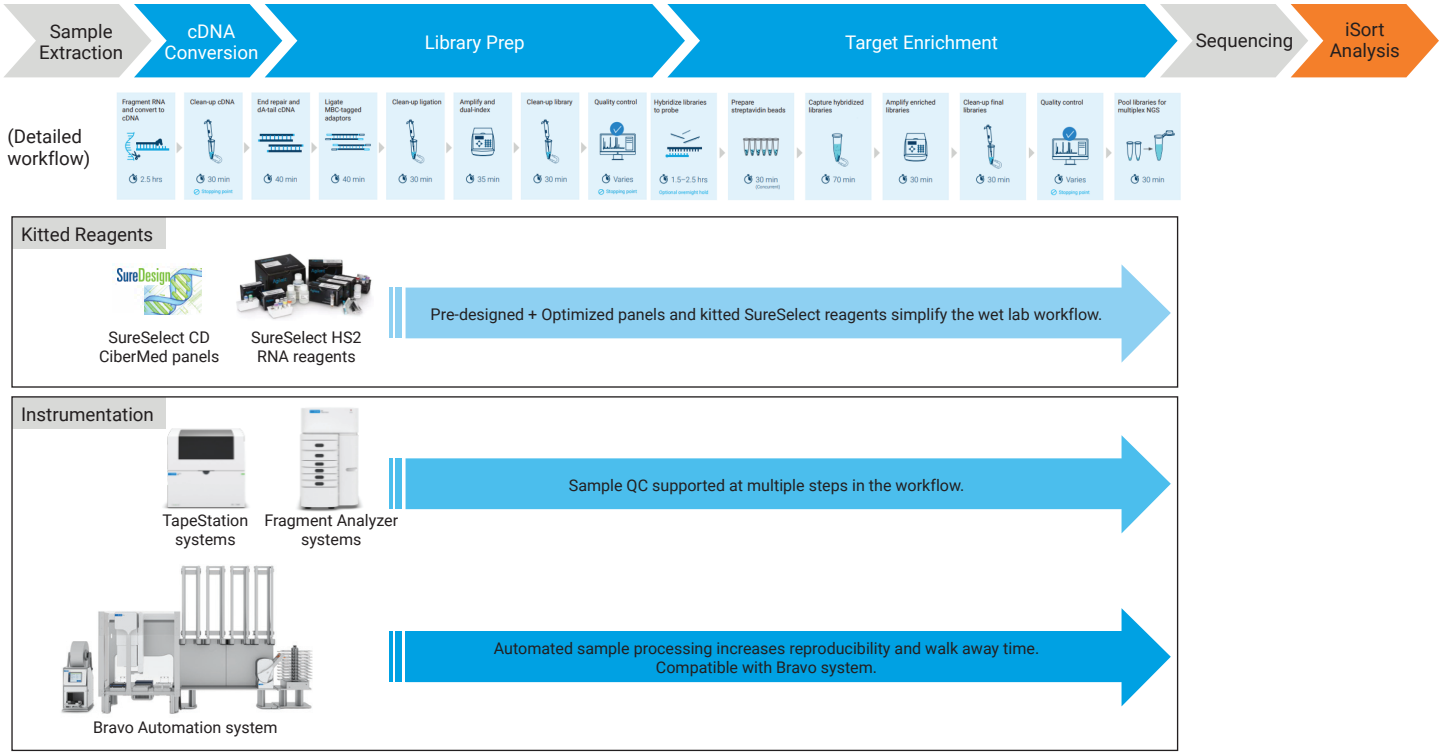

Figure 2. End-to-end workflow supported by kitted reagents and instrumentation from Agilent and an integrated analysis solution from CiberMed.

Materials and methods

Samples

Paired FF and FFPE tumor samples from four patients with NSCLC were procured from Proteogenex (Inglewood, CA, USA). Total RNA of the FF samples was extracted using the AllPrep DNA/RNA Micro Kit (Qiagen). The AllPrep DNA/RNA FFPE Kit (Qiagen) was used to extract the RNA of one FFPE sample (FFPE-1). The RNAstorm FFPE RNA extraction kit (Biotium, formerly Cell Data Sciences) was used to extract RNA from the remaining three FFPE samples. RNA quality was assessed using the Agilent RNA 6000 Pico kit (Agilent p/n 5067-1513) and RNA concentrations were determined using the Qubit RNA HS Assay kit (Thermo Fisher Scientific p/n Q32855). All FF samples had RIN > 6 whereas FFPE samples had DV200 values from 52 to 82.

Library preparation and target enrichment for targeted RNA sequencing

Library preparation and targeted enrichment were performed using an Agilent Bravo NGS Workstation Option B following the Agilent SureSelect XT HS2 RNA system user guide G9993-90010.[31] Agilent SureSelect XT HS2 RNA reagent kits G9991A and G9991B were employed to generate three technical replicate libraries per sample, each starting with an input of 30 ng total RNA. Agilent SureSelect XT HS2 RNA target enrichment kit part number G9994A was employed to enrich sample libraries with the SureSelect CD CiberMed Tissue panel.

Agilent SureDesign software was used to create the SureSelect CD CiberMed Tissue panel. Manual curation was performed to ensure full coverage of the coding region for every included gene.

Enriched libraries were sequenced as 150 bp × 2 paired-end reads on Illumina NovaSeq 6000 instrument. Each sample was sequenced to approximately 50M (25M × 2) total reads to allow thorough performance assessment at varying depths of coverage.

Data processing

As input to iSort, RNA sequencing reads (in FASTQ format) were first summarized to gene expression values in transcripts per million (TPM). While this process can be accomplished using any standard mapping/alignment approach, for the data presented here, we used Salmon v1.9[32] to map and quantify RNA-seq reads. The TPM values were then used as input to iSort Fractions v1.4, which determined the relative fractions of 25 immune subsets (Table 2) in each tissue sample.

CiberMed software tools, iSort Fractions and iSort HiRes, are currently available as docker-containerized tools that can be run locally. iSort Fractions is offered also by way of a user-friendly web interface run securely through AWS (Amazon Web Services) at https://isort.cibermed.com. For more details about the iSort digital cytometry suite, see the brochure at https://isort.cibermed.com/iSort_ProductsAndServices.pdf or contact CiberMed directly for questions at https://cibermed.com/contact.

Performance evaluation

The accuracy of iSort Fractions was evaluated using Spearman rho, Pearson r, and root mean squared error (RMSE) to quantify the concordance of estimated cell type proportions, both within and between paired FF and FFPE samples.

Read titration analysis

To determine the impact of the number of reads per sample on deconvolution accuracy and reproducibility, paired reads were randomly sampled to 5M, 2M, 1M, 500k, 100k, and 10k effective reads per sample. Effective reads are defined as the number of reads mapped “on-target” to the genes included in the SureSelect CD CiberMed Tissue panel. Input reads are calculated as [minimum effective reads]/[on-target map rate].

Results

All RNA-seq data generated in this study passed internal quality control requirements.

Accuracy

Four pairs of matched FF and FFPE samples, derived from NSCLC tumor biopsies, with three replicates each, were analyzed using iSort Fractions tissue mode, which quantifies 25 cell types, including 22 hematopoietic subsets and three non-immune cell types: fibroblast, endothelial, and epithelial cells.

Figure 3 summarizes the concordance of iSort Fractions results between FF and FFPE samples for 25 cell types (top) and 23 cell types (bottom) with the latter excluding fibroblasts and epithelial cells to better visualize the 0 to 10% fractional abundance range. While performance gains can be seen for iSort over CIBERSORTx (Figure 3b versus Figure 3a), targeted sequencing with the SureSelect CD CiberMed Tissue panel substantially outperformed whole transcriptome sequencing applied to the same samples (Figure 3c versus Figure 3b). These data indicate that targeted digital cytometry with SureSelect and iSort can overcome FFPE-related distortions to enable reliable cell profiling in fixed tissue specimens

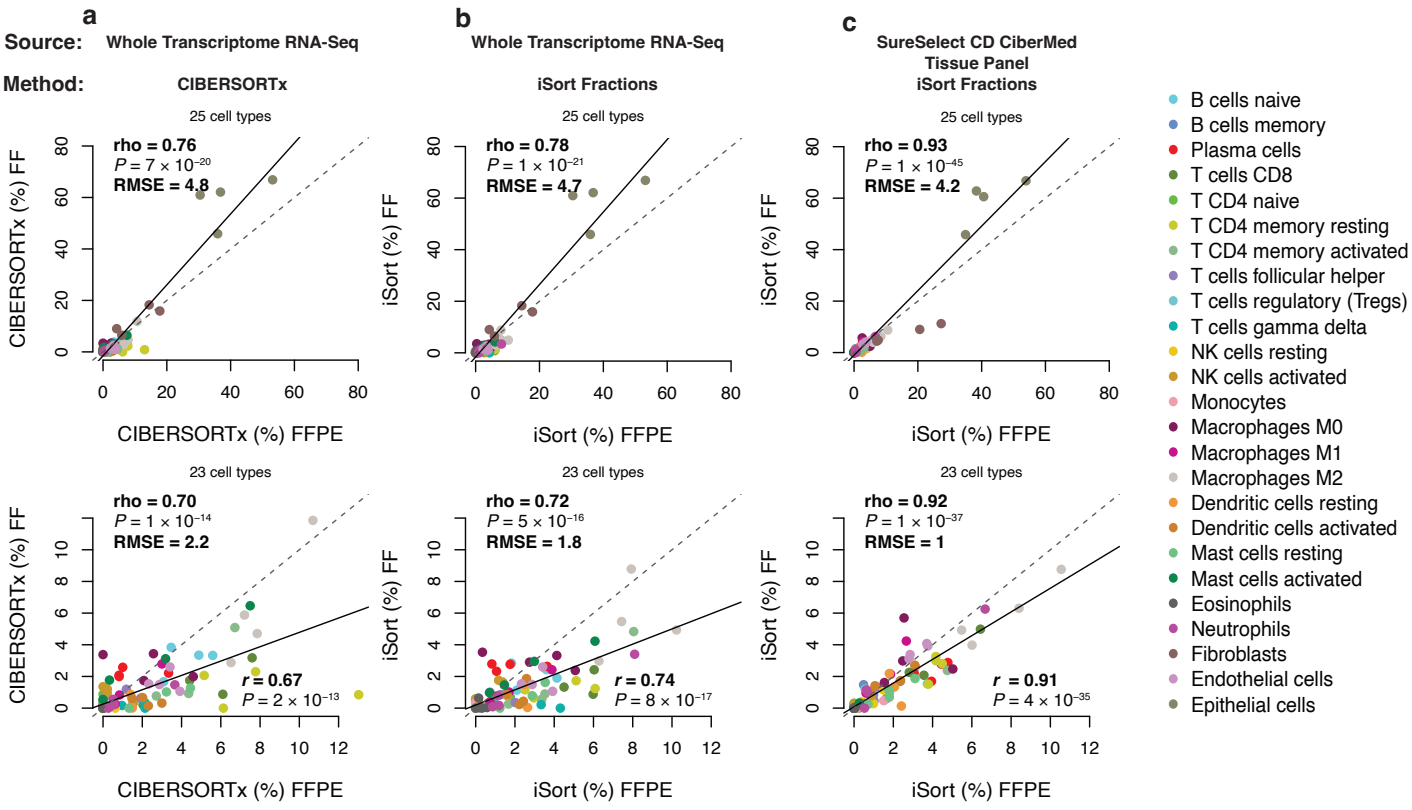

Figure 3. Scatterplots of cell type fractions determined by deconvolution of paired FF (y-axis) and FFPE (x-axis) bulk tumor expression profiles (n = 4 pairs); comparing (a) whole-transcriptome sequencing data analyzed by CIBERSORTx, (b) whole-transcriptome sequencing data analyzed by iSort Fractions, and (c) Agilent SureSelect CD CiberMed Tissue panel data analyzed by iSort Fractions. Results are shown for all 25 evaluable cell types (top) and 23 cell types (bottom) to expand the lower range of fractional abundances (0 to 10%). Each point represents a sample colored by cell type.

Figure 4 summarizes the sample-level concordance between FF and FFPE cell type fractions. The SureSelect CD CiberMed Tissue panel outperformed whole-transcriptome sequencing for all four samples.

Reproducibility

To assess reproducibility, RNA from all samples was processed and sequenced in triplicate and the concordance of iSort Fractions results across technical replicates was evaluated. All pairwise comparisons across the three replicates demonstrated strong reproducibility, with nearly perfect correlations obtained for all samples (Figure 5).

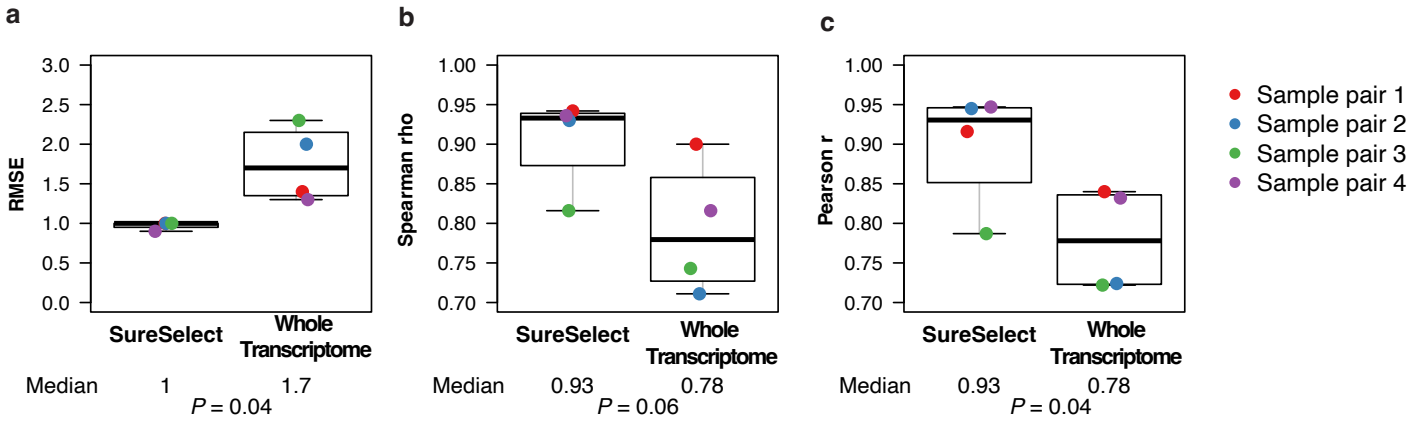

Figure 4. Box plots summarizing sample-level concordance of cell type fractions from paired FF and FFPE NSCLC specimens, analyzed by sample pair across 23 cell types and stratified by sequencing assay (Agilent SureSelect CD CiberMed Tissue panel versus whole-transcriptome sequencing). Fibroblasts and epithelial cells were excluded to focus on cell types in the range of 0 to 10% fractional abundance. Concordance was evaluated by (a) root mean squared error (RMSE), (b) Spearman rho, and (c) Pearson r. Statistical significance was determined with a two-sided paired t-test. All results were obtained using iSort Fractions tissue mode.

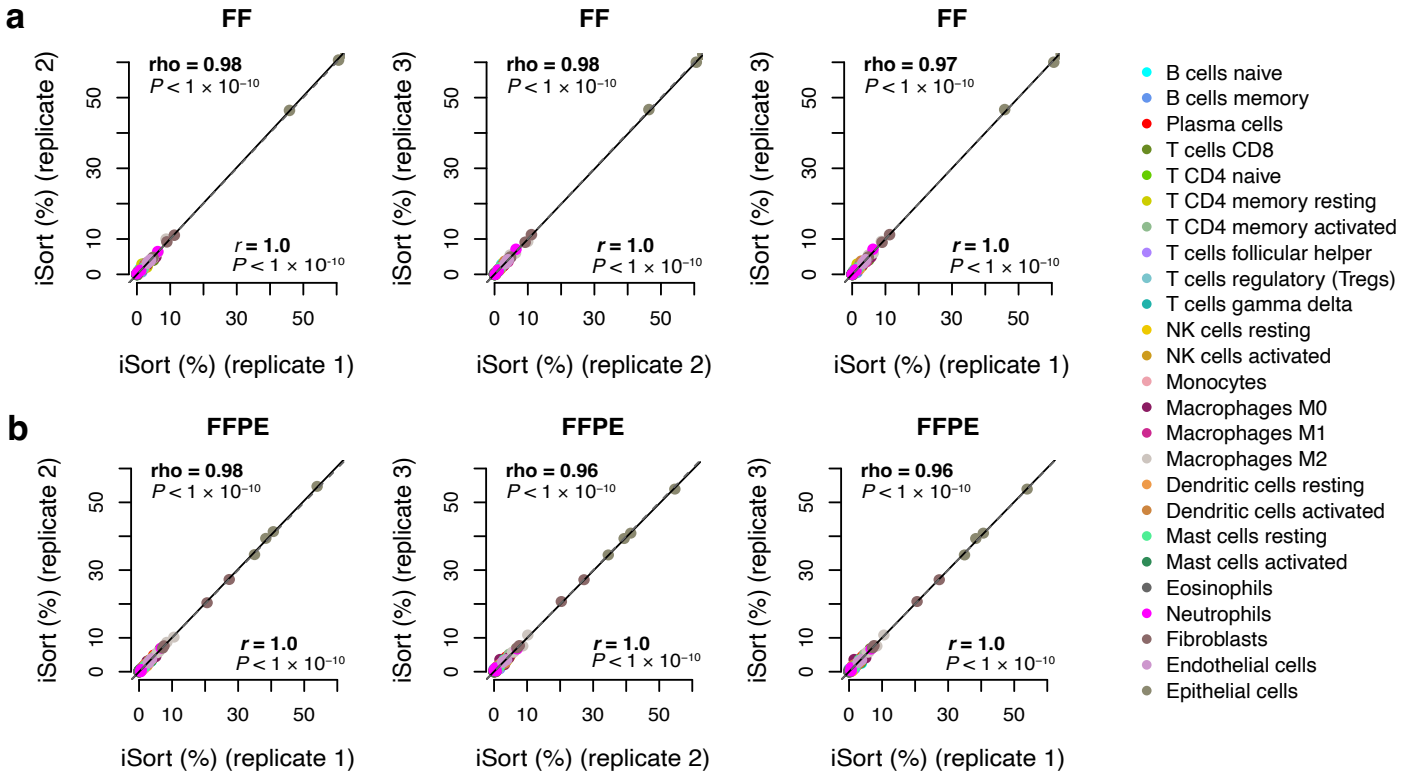

Figure 5. Scatterplots comparing iSort Fractions results between sequencing replicates, shown for (a) four FF samples and (b) four FFPE tumor samples. Each data point represents a sample colored by cell type. The Agilent SureSelect CD CiberMed Tissue panel was applied to all samples.

Sequencing requirement

Samples were sequenced to an average of 50M (25M × 2) reads for the SureSelect CD CiberMed Tissue panel.
To determine the minimum number of reads needed per sample without compromising performance, we performed a titration experiment. In this experiment, the number of effective reads per sample was down-sampled to predefined quantities before running iSort Fractions (where "effective reads" denotes on-target reads only). Figure 6 summarizes the concordance between FF and FFPE cell type fractions for all three replicates using pre-determined numbers of effective reads. All three replicates maintained accurate and stable performance down to 1M (500k × 2) effective reads (Figure 6). Figure 7 shows all data points from Replicate 1 before and after down-sampling to 1M (500k × 2) effective reads.

Based on these data, when using the SureSelect CD CiberMed Tissue panel, the total number of required input reads is expected to range from approximately 1.3M per sample (assuming a 75% on-target mapping rate) to approximately 2M per sample (assuming a 50% on-target rate). Therefore, a conservative projection of 2M (1M × 2) total input reads per sample should be sufficient. With the recommended input of at least 40M (20M × 2) reads for iSort Fractions from whole-transcriptome sequencing, the SureSelect CD CiberMed Tissue panel offers a 20-fold reduction in sequencing costs while achieving superior performance.

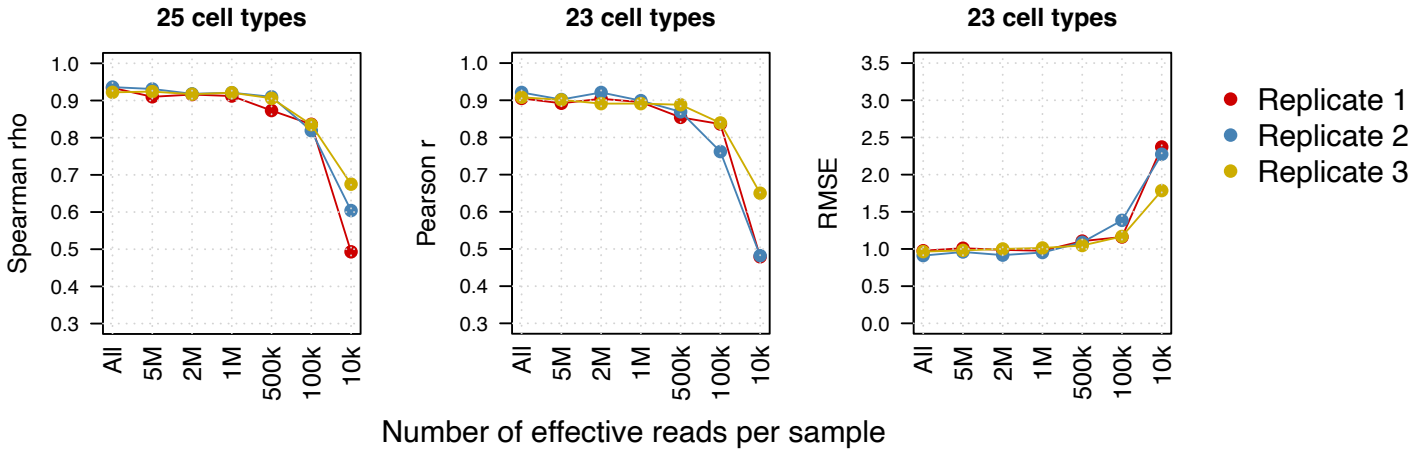

Figure 6. Impact of reads per sample on the concordance between FF and FFPE cell type proportions determined by iSort Fractions. "All" denotes all evaluable reads per sample. The Agilent SureSelect CD CiberMed Tissue panel maintained stable performance down to 1M (500k × 2) effective ("on-target") reads per sample for all three replicates. Because Pearson r and RMSE are heavily driven by abundant fibroblasts and epithelial cells, they are shown without these two cell types. RMSE, root mean squared error.

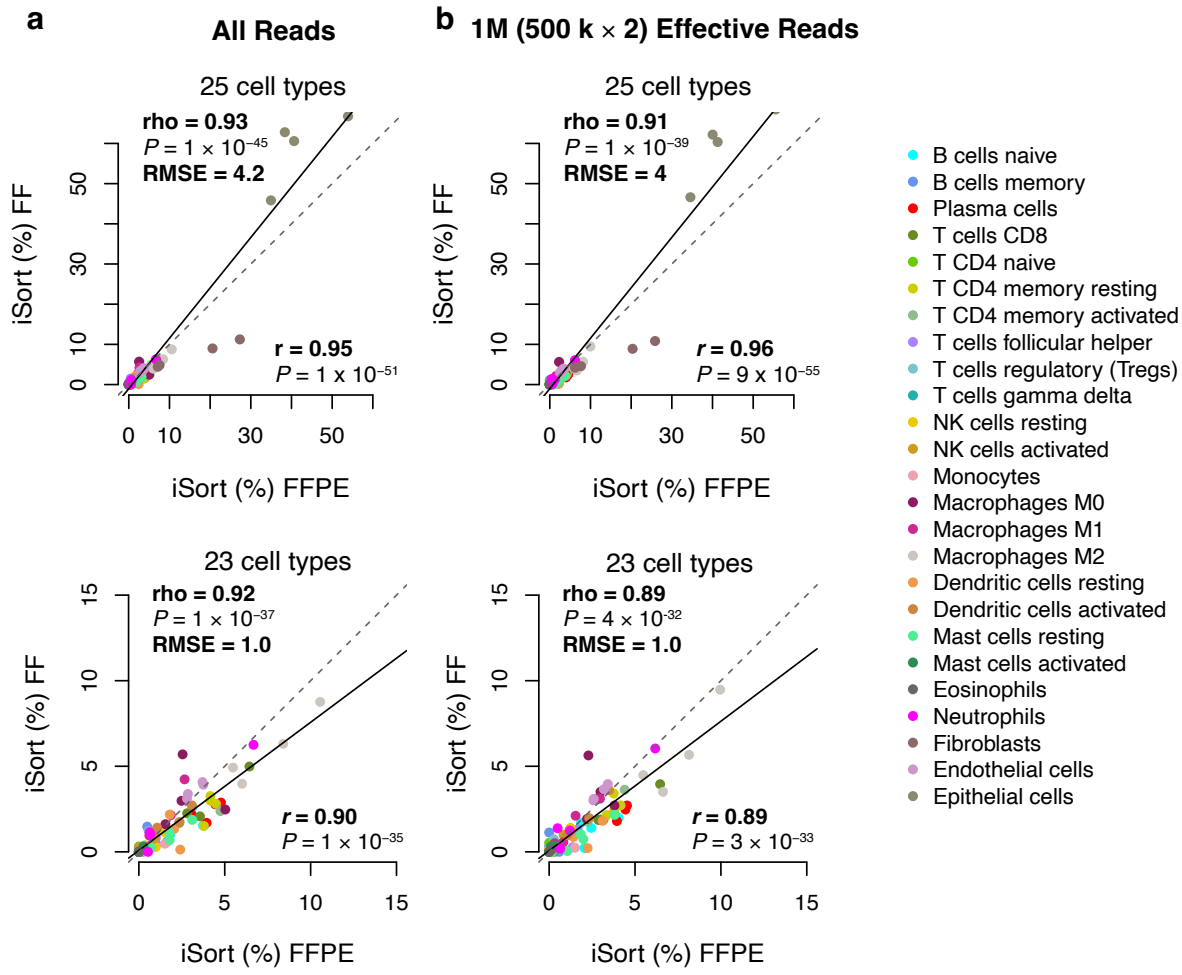

Figure 7. Impact of reads per sample on the concordance between cell type fractions in paired FF and FFPE samples, comparing (a) all evaluable reads per sample (no down-sampling) with (b) 1M (500k × 2) effective reads per sample (on-target reads per sample). The results are shown for all 25 cell types (top) and 23 cell types (bottom), with the latter expanding the fractional abundance range between 0 and 15%. Each data point denotes a sample colored by cell type. Results were calculated using the Agilent SureSelect CD CiberMed Tissue panel applied to four matching pairs of FF (y-axis) and FFPE (x-axis) tumor specimens.

Conclusion

In this application note, we describe the performance characteristics of the SureSelect CD Tissue panel for profiling both FF and FFPE tissue samples with iSort, a state-of-the-art software solution for digital cytometry offered by CiberMed. There are several applications that would benefit from the enhanced robustness, accuracy, and cost-effectiveness of this new joint assay for cytometry by sequencing. These include routine and translational tissue analysis,[35] retrospective characterization of bulk tissue expression data to derive new insights into cellular composition,[1] and large-scale validation of sequencing analyses.[36] Other applications include the general assessment of cell type composition under diverse physiological and pathological conditions, all without the need for antibodies, fresh specimens, viable material, or millions of cells.

As shown previously, by comparing FFPE samples to a matched FF reference, the methodology is robust to FFPE artifacts and recovers the cell type abundance profile of matched references. Moreover, the Agilent SureSelect CD Tissue panel substantially outperforms whole-transcriptome profiling with potential to greatly reduce sequencing cost.

The utility of the core iSort methodology has been demonstrated in different tissues and various cancer types.[27,30,47,48] It has also been demonstrated in multiple other contexts, including immuno-oncology,[37,38] organ transplantation,[39] cardiology,[40] Crohn's disease,[30] and neonatal sepsis.[41] The ability to deconvolve alternative genomic data types, including methylation and proteomic profiles, has also been demonstrated.[19,42] Furthermore, iSort has been analytically validated for whole blood samples and fresh, frozen, and fixed tumor specimens.[27,30] The core algorithm underlying iSort has become the standard methodology for deconvolution in large cancer studies and datasets, such as The Cancer Genome Atlas (TCGA), and is being applied in clinical trials.[43-46] Given the performance gains demonstrated here, the SureSelect/iSort assay for digital cytometry promises to facilitate many exciting and impactful future applications.

References

1. Thorsson, V.; Gibbs, D. L.; Brown, S. D.; Wolf, D.; Bortone, D. S.; Yang, T. H. O.; Porta-Pardo, E.; Gao, G. F.; Plaisier, C. L.; Eddy, J. A.; et al. The Immune Landscape of Cancer. *Immunity*. **2018**, *48* (4), P812-830. https://doi.org/10.1016/j.immuni.2018.03.023.

2. Farc, O.; Cristea, V. An Overview of the Tumor Microenvironment, from Cells to Complex Networks (Review). *Exp. Ther. Med*. **2020**, *21* (1). https://doi.org/10.3892/etm.2020.9528

3. Newman, A. M.; Liu, C. L.; Green, M. R.; Gentles, A. J.; Feng, W.; Xu, Y.; Hoang, C. D.; Diehn, M.; Alizadeh, A. A. Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* **2015**, *12* (5), 453–457. https://doi.org/10.1038/nmeth.3337.

4. Gentles, A. J.; Newman, A. M.; Liu, C. L.; Bratman, S. V.; Feng, W.; Kim, D.; Nair, V. S.; Xu, Y.; Khuong, A.; Hoang, C. D.; et al. The Prognostic Landscape of Genes and Infiltrating Immune Cells across Human Cancers. Nat. Med. 2015, 21 (8), 938–945. https://doi.org/10.1038/nm.3909.

5. Bentham, R.; Litchfield, K.; Watkins, T. B. K.; Lim, E. L.; Rosenthal, R.; Martínez-Ruiz, C.; Hiley, C. T.; Bakir, M. A.; Salgado, R.; Moore, D. A.; et al. TRACERx Consortium. Using DNA Sequencing Data to Quantify T Cell Fraction and Therapy Response. *Nature* **2021**, *597* (7877), 555–560. https://doi.org/10.1038/s41586-021-03894-5.

6. Herold, N. C.; Mitra, P. *Immunophenotyping; StatPearls* Publishing, **2023**.

7. Alfonso, B.-F.; Al-Rubeai, M. Flow Cytometry. In *Comprehensive Biotechnology*; Elsevier, **2011**; pp 559–578.

8. Dixon, A. R.; Bathany, C.; Tsuei, M.; White, J.; Barald, K. F.; Takayama, S. Recent Developments in Multiplexing Techniques for Immunohistochemistry. *Expert Rev. Mol. Diagn*. **2015**, *15* (9), 1171–1186. https://doi.org/10.1586/14737159.2015.1069182

9. Rashid, R.; Gaglia, G.; Chen, Y.-A.; Lin, J.-R.; Du, Z.; Maliga, Z.; Schapiro, D.; Yapp, C.; Muhlich, J.; Sokolov, A.; et al. Highly Multiplexed Immunofluorescence Images and Single-Cell Data of Immune Markers in Tonsil and Lung Cancer. *Sci. Data* **2019**, *6* (1), 1–10. https://doi.org/10.1038/s41597-019-0332-y.

10. Palit, S.; Heuser, C.; de Almeida, G. P.; Theis, F. J.; Zielinski, C. E. Meeting the Challenges of High-Dimensional Single-Cell Data Analysis in Immunology. *Front. Immunol*. **2019**, *10*. https://doi.org/10.3389/fimmu.2019.01515.

11. Lähnemann, D.; Köster, J.; Szczurek, E.; McCarthy, D. J.; Hicks, S. C.; Robinson, M. D.; Vallejos, C. A.; Campbell, K. R.; Beerenwinkel, N.; Mahfouz, A.; et al. Eleven Grand Challenges in Single-Cell Data Science. *Genome Biol*. **2020**, *21* (1). https://doi.org/10.1186/s13059-020-1926-6.

12. Aran, D.; Hu, Z.; Butte, A. J. xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol*. **2017**, *18* (1). https://doi.org/10.1186/s13059-017-1349-1.

13. Racle, J.; Gfeller, D. EPIC: A Tool to Estimate the Proportions of Different Cell Types from Bulk Gene Expression Data. In *Bioinformatics for Cancer Immunotherapy*; Springer US: New York, NY, **2020**; Vol. 2120, pp 233–248.

14. Li, B.; Li, T.; Liu, J. S.; Liu, X. S. Computational Deconvolution of Tumor-Infiltrating Immune Components with Bulk Tumor Gene Expression Data. In *Bioinformatics for Cancer Immunotherapy*; Springer US: New York, NY, **2020**; Vol. 2120, pp 249–262.

15. Becht, E.; Giraldo, N. A.; Lacroix, L.; Buttard, B.; Elarouci, N.; Petitprez, F.; Selves, J.; Laurent-Puig, P.; Sautès-Fridman, C.; Fridman, W. H.; de Reyniès, A. Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol*. **2016**, *17* (1). https://doi.org/10.1186/s13059-016-1070-5.

16. Shen-Orr, S. S.; Tibshirani, R.; Butte, A. J. Gene Expression Deconvolution in Linear Space. *Nat. Methods* **2012**, *9* (1), 9–9. https://doi.org/10.1038/nmeth.1831.

17. Newman, A. M.; Steen, C. B.; Liu, C. L.; Gentles, A. J.; Chaudhuri, A. A.; Scherer, F.; Khodadoust, M. S.; Esfahani, M. S.; Luca, B. A.; Steiner, D.; et al. Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol*. **2019**, *37* (7), 773–782. https://doi.org/10.1038/s41587-019-0114-2.

18. Bionetworks, S. Synapse. Synapse.org. https://www.synapse.org/ (accessed 2023-10-30).

19. Nadel, B. B.; Oliva, M.; Shou, B. L.; Mitchell, K.; Ma, F.; Montoya, D. J.; Mouton, A.; Kim-Hellmuth, S.; Stranger, B. E.; Pellegrini, M.; Mangul, S. Systematic Evaluation of Transcriptomics-Based Deconvolution Methods and References Using Thousands of Clinical Samples. *Brief. Bioinform*. **2021**, *22* (6). https://doi.org/10.1093/bib/bbab265.

20. Avila Cobos, F.; Alquicira-Hernandez, J.; Powell, J. E.; Mestdagh, P.; De Preter, K. Benchmarking of Cell Type Deconvolution Pipelines for Transcriptomics Data. *Nat. Commun*. **2020**, *11* (1). https://doi.org/10.1038/s41467-020-19015-1.

21. Le, T.; Aronow, R. A.; Kirshtein, A.; Shahriyari, L. A Review of Digital Cytometry Methods: Estimating the Relative Abundance of Cell Types in a Bulk of Cells. *Brief. Bioinform*. **2021**, *22* (4). https://doi.org/10.1093/bib/bbaa219.

22. Chen, B.; Khodadoust, M. S.; Liu, C. L.; Newman, A. M.; Alizadeh, A. A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. In *Methods in Molecular Biology*; Springer New York: New York, NY, **2018**; Vol. 1711, pp 243–259.

23. Sutton, G. J.; Poppe, D.; Simmons, R. K.; Walsh, K.; Nawaz, U.; Lister, R.; Gagnon-Bartsch, J. A.; Voineagu, I. Comprehensive Evaluation of Deconvolution Methods for Human Brain Gene Expression. *Nat. Commun*. **2022**, *13* (1), 1–18. https://doi.org/10.1038/s41467-022-28655-4.

24. Breen, M. S.; Ozcan, S.; Ramsey, J. M.; Wang, Z.; Ma'ayan, A.; Rustogi, N.; Gottschalk, M. G.; Webster, M. J.; Weickert, C. S.; Buxbaum, J. D.; Bahn, S. Temporal Proteomic Profiling of Postnatal Human Cortical Development. *Transl. Psychiatry* **2018**, *8* (1), 1–14. https://doi.org/10.1038/s41398-018-0306-4.

25. Li, L.; Shen, L.; Ma, J.; Zhou, Q.; Li, M.; Wu, H.; Wei, M.; Zhang, D.; Wang, T.; Qin, S.; Xing, T. Evaluating Distribution and Prognostic Value of New Tumor-Infiltrating Lymphocytes in HCC Based on a scRNA-Seq Study with CIBERSORTx. *Front. Med. (Lausanne)* **2020**, *7*. https://doi.org/10.3389/fmed.2020.00451.

26. Newman, A.M.; Gulati, S.G.; Clarke, M.F.; Sikandar, S.S. Methods Utilizing Single Cell Genetic Data for Cell Population Analysis and Applications Thereof. US202003700112 A1, November 26, **2020**.

27. Newman, A. M.; Nakao, A.; Li, K.; Liu, C.-L.; Mathi, K.; Sigal, N.; Maecker, H.; Diehn, M.; Alizadeh, A. A. Analytical Validation of Digital Cytometry (iSort) for Leukocyte Enumeration Using Stored Blood. *J. Clin. Oncol*. **2020**, *38* (15_suppl), 3542–3542. https://doi.org/10.1200/jco.2020.38.15_suppl.3542.

28. Chakravarthy, A.; Furness, A.; Joshi, K.; Ghorani, E.; Ford, K.; Ward, M. J.; King, E. V.; Lechner, M.; Marafioti, T.; Quezada, S. A.; Thomas, G. J.; Feber, A.; Fenton, T. R. Pan-Cancer Deconvolution of Tumour Composition Using DNA Methylation. *Nat. Commun*. **2018**, *9* (1). https://doi.org/10.1038/s41467-018-05570-1.

29. Hutson, M. Hunting for the Best Bioscience Software Tool? Check This Database. *Nature* **2023**. https://doi.org/10.1038/d41586-023-00053-w.

30. Chen, H.; Chen, C.; Yuan, X.; Xu, W.; Yang, M.-Q.; Li, Q.; Shen, Z.; Yin, L. Identification of Immune Cell Landscape and Construction of a Novel Diagnostic Nomogram for Crohn's Disease. *Front. Genet*. **2020**, *11*. https://doi.org/10.3389/fgene.2020.00423.

31. A, V. Agilent.com. https://www.agilent.com/cs/library/usermanuals/public/G9993-90010.pdf (accessed 2023-10-30).

32. Patro, R.; Duggal, G.; Love, M. I.; Irizarry, R. A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14* (4), 417–419. https://doi.org/10.1038/nmeth.4197.

33. Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

34. Li, B.; Dewey, C. N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* **2011**, *12* (1). https://doi.org/10.1186/1471-2105-12-323.

35. Orange, D. E.; Yao, V.; Sawicka, K.; Fak, J.; Frank, M. O.; Parveen, S.; Blachere, N. E.; Hale, C.; Zhang, F.; Raychaudhuri, S.; et al. RNA Identification of PRIME Cells Predicting Rheumatoid Arthritis Flares. *N. Engl. J. Med*. **2020**, *383* (3), 218–228. https://doi.org/10.1056/nejmoa2004114.

36. Gohil, S. H.; Iorgulescu, J. B.; Braun, D. A.; Keskin, D. B.; Livak, K. J. Applying High-Dimensional Single-Cell Technologies to the Analysis of Cancer Immunotherapy. *Nat. Rev. Clin. Oncol*. **2021**, *18* (4), 244–256. https://doi.org/10.1038/s41571-020-00449-x.

37. Nabet, B. Y.; Esfahani, M. S.; Moding, E. J.; Hamilton, E. G.; Chabon, J. J.; Rizvi, H.; Steen, C. B.; Chaudhuri, A. A.; Liu, C. L.; Hui, A. B.; et al. Noninvasive Early Identification of Therapeutic Benefit from Immune Checkpoint Inhibition. *Cell* **2020**, *183* (2), 363-376.e13. https://doi.org/10.1016/j.cell.2020.09.001.

38. Lozano, A. X.; Chaudhuri, A. A.; Nene, A.; Bacchiocchi, A.; Earland, N.; Vesely, M. D.; Usmani, A.; Turner, B. E.; Steen, C. B.; Luca, B. A.; et al. T Cell Characteristics Associated with Toxicity to Immune Checkpoint Blockade in Patients with Melanoma. *Nat. Med*. **2022**, *28* (2), 353–362. https://doi.org/10.1038/s41591-021-01623-z.

39. Huang, S.; Chen, H.; Guo, Z.; He, X. A Comprehensive Cibersort Study and Gene-Expression Based Transcriptomic Analysis on Patterns of Immune Infiltration in Ischemia-Reperfusion Injury Livers Post Liver Transplantation. *Transplantation* **2020**, *104* (S3), S195–S195. https://doi.org/10.1097/01.tp.0000699368.56252.17.

40. Journal-of-cardiology.com. https://www.journal-of-cardiology.com/article/S0914-5087(20)30289-6/fulltext (accessed 2023-10-30).

41. Jiang, Z.; Luo, Y.; Wei, L.; Gu, R.; Zhang, X.; Zhou, Y.; Zhang, S. Bioinformatic Analysis and Machine Learning Methods in Neonatal Sepsis: Identification of Biomarkers and Immune Infiltration. *Biomedicines* **2023**, *11* (7), 1853. https://doi.org/10.3390/biomedicines11071853.

42. Voss, M. H.; Buros Novik, J.; Hellmann, M. D.; Ball, M.; Hakimi, A. A.; Miao, D.; Margolis, C.; Horak, C.; Wind-Rotolo, M.; De Velasco, G.; et al. Correlation of Degree of Tumor Immune Infiltration and Insertion-and-Deletion (Indel) Burden with Outcome on Programmed Death 1 (PD1) Therapy in Advanced Renal Cell Cancer (RCC). *J. Clin. Oncol*. **2018**, *36* (15_suppl), 4518–4518. https://doi.org/10.1200/jco.2018.36.15_suppl.4518.

43. Metzger Filho, O.; Stover, D. G.; Asad, S.; Ansell, P. J.; Watson, M.; Loibl, S.; Geyer, C. E.; O'Shaughnessy, J.; Untch, M.; Rugo, H. S.; et al. Immunophenotype and Proliferation to Predict for Response to Neoadjuvant Chemotherapy in TNBC: Results from BrighTNess Phase III Study. *J. Clin. Oncol*. **2019**, *37* (15_suppl), 510–510. https://doi.org/10.1200/jco.2019.37.15_suppl.510.

44. *CTG labs - NCBI*. Clinicaltrials.gov. https://clinicaltrials.gov/study/NCT03979508 (accessed 2023-10-30).

45. Clinicaltrials.gov. https://classic.clinicaltrials.gov/ProvidedDocs/10/NCT03955510/Prot_000.pdf (accessed 2023-10-30).

46. Medjebar, S.; Richard, C.; Fumet, J.-D.; Malo, J.; Elkrief, A.; Blais, N.; Tehfe, M.; Florescu, M.; Boidot, R.; Truntzer, C.; et al. Angiotensin-Converting Enzyme Inhibitor Prescription Is Associated with Decreased Progression-Free Survival (PFS) and Overall Survival (OS) in Patients with Lung Cancers Treated with PD-1/PD-L1 Immune Checkpoint Blockers. *J. Clin. Oncol*. **2019**, *37* (15_suppl), e20512–e20512. https://doi.org/10.1200/jco.2019.37.15_suppl.e20512

47. Steen, C. B.; Liu, C. L.; Alizadeh, A. A.; Newman, A. M. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods in Molecular Biology* **2020**, 135–157. https://doi.org/10.1007/978-1-0716-0301-7_7

48. Craven, K. E.; Gökmen-Polar, Y.; Badve, S. S. CIBERSORT Analysis of TCGA and METABRIC Identifies Subgroups with Better Outcomes in Triple Negative Breast Cancer. *Scientific Reports* **2021**, *11* (1). https://doi.org/10.1038/s41598-021-83913-7

49. Newman, A.M. et al, *J. Clinical Oncology*, 38, Supp. e15243 (**2020**) https://ascopubs.org/doi/abs/10.1200/JCO.2020.38.15_suppl.e15243

www.agilent.com

PR7001-2303

Published in the USA, March 20, 2024
5994-7248EN

Agilent
Trusted Answers