Spatial protein and RNA analysis on the same tissue section using MICS technology

Authors

Emily Neil[1], Dongju Park[1], Rebecca C. Hennessey[1], Eric C. DiBiasio[1], Michael DiBuono[1], Hanna Lafayette[1], Erica Lloyd[1], Hsinyi Lo[1], Julia Femel[2], Alex Makrigiorgos[1], Sameh Soliman[1], Dominic Mangiardi[1], Paurush Praveen[2], Silvia Rüberg[2], Fabian Staubach[2], Ryan Hindman[3], Thomas Rothmann[2], Hansueli Meyer[1], Tanya Wantenaar[2], Jinling Wang[1], Werner Müller[2], Robert Pinard[1], and Andreas Bosio[2]*

*Corresponding author

Affiliations

[1]Miltenyi Biotec Inc, Waltham, Massachusetts, USA
[2]Miltenyi Biotec B.V. & Co. KG, Bergisch Gladbach, Germany
[3]Miltenyi Biotec Inc, Gaithersburg, Maryland, USA

Abstract

Spatial Biology has evolved from the molecular characterization of microdissected cells to high throughput spatial RNA and protein expression analysis at scale. The main limitation of spatial technologies so far is the inability to resolve protein and RNA information in the same histological section. Here, we report for the first time the integration of highly multiplexed RNA and protein detection on the same tissue section. We developed a new, automated, spatial RNA detection method (RNAsky™), which is based on targeted rolling circle amplification and iterative staining. We combine RNAsky with MACSima™ Imaging Cyclic Staining (MICS) based protein analysis and show compatibility with subsequent standard hematoxylin and eosin (H&E) staining. Using both, open-source tools and our recently developed software suite MACS® iQ View, we demonstrate our multiomics MICS workflow by characterizing key immune-oncology markers at subcellular resolution across normal and diseased tissues.

Introduction

Understanding the composition and architecture of cells within tissues is fundamental to studying tissue function, and more broadly, human biology. The complex dynamics of cellular functions and interactions between cell populations drive normal physiology, with dysregulation of these mechanisms underlying disease progression. Advances in bulk and single cell sequencing have facilitated the establishment of cell atlas initiatives, which aim to map all cells in the human body in healthy and diseased states [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Although sequencing methods allow for transcriptome wide analyses of individual cells with high sensitivity and accuracy, they require dissociation of cells, thereby destroying all contextual information [11, 12].

Image-based, spatially resolved transcriptomics methods have been extensively reviewed, and fall primarily into two categories: fluorescent *in situ* hybridization (FISH) based or rolling circle amplification (RCA) based approaches [13, 14, 15, 16, 17]. With FISH based approaches, complementary probes are hybridized to mRNAs of interest, and repeatedly interrogated to generate optical barcodes associated with each gene [18, 19, 20, 21, 22, 23]. While optical barcoding allows for high multiplexing, it requires long imaging times and advanced algorithms for barcode decoding and downstream analyses. For RCA approaches, a circular template is required for amplification and is achieved either by using targeted padlock probes or circularizing templates *in situ* [24, 25, 26, 27, 28, 29]. Rolling circle amplicons can be interrogated in a number of ways, including sequencing by ligation or hybridization-based readouts. A combination of low plex *in situ* fluorescence mRNA analysis with MALDI mass spectrometry imaging or immunofluorescent protein detection has been reported in literature [30, 31, 32]. However, spatially resolved and highly multiplexed RNA detection methods have so far shown limited compatibility for same-slide iterative immunofluorescence-based protein detection methods.

Spatial proteomics is a vital counterpart to spatial transcriptomics, as it provides direct assessments of functional information. Numerous multiplex proteomics technologies have been developed, including iterative and simultaneous detection methods [33, 34, 35, 36, 37, 38]. One such advance is MACSima™ imaging cyclic staining (MICS), which is a fully automated, cyclic immunofluorescence technology performed on the MACSima platform utilizing directly conjugated antibodies combined with gentle signal erasure methods, thereby allowing the application of hundreds of antibodies to a single sample [33]. Due to the non-destructive nature of this process, the MACSima platform is an ideal system for combining multiplex proteomics with spatial transcriptomics.

Here, we introduce RNAsky™, a targeted, RCA-based gene expression profiling assay designed for the MACSima platform. We combined this assay with multiplex protein imaging to generate multiomic datasets in normal and diseased tissues. To demonstrate the capabilities of our multiomics platform, we profiled multiple tissues using a targeted RNAsky Immuno-Oncology (I-O) Core panel with additional customized gene targets and a selection of protein markers.

Results

The MACSima Platform enables same-section multiplex analysis of proteins and mRNA

A comprehensive understanding of the complex interactions between cells in a spatial context requires multiple information modalities. Recently, MICS technology was introduced for spatial analysis of epitopes such as proteins, glycolipids, and complex sugars at subcellular resolution using hundreds of immunofluorescent antibodies on the same specimen [33]. To add the capability of transcript detection, we developed RNAsky, which is based on the same principle of iterative staining, sample washing, multi-field imaging, and signal erasure. Like protein detection, the RNAsky assay is performed on the MACSima Platform, a fully automated, fluorescence microscopy instrument. We specifically optimized transcript amplification and detection chemistries to be compatible with subsequent epitope detection in order to combine the analysis of RNA and protein targets on a single formalin-fixed paraffin-embeded (FFPE) tissue section (Fig. 1a, Methods). We chose a workflow starting with RNA target probe amplification, followed by iterative detection probe hybridization, and ending with immunofluorescent detection of epitopes. Similar to directly-conjugated antibodies, each gene displayed a unique emission spectrum during a single cycle via the hybridization of gene-specific detection probes. Transcripts were detected across four wavelengths (FITC, PE, APC, near infrared), and proteins were detected with three wavelengths (FITC, PE, APC). This technology supports simultaneous analyses of protein markers and RNA transcripts, enabling the characterization of the spatial architecture and expression topography of tissues.

RNAsky is a reproducible, sensitive, specific, and spatially resolved gene expression assay

RNAsky is a spatial gene expression assay designed to complement the proteomic capabilities of the MACSima platform. Briefly, the assay is based on targeted padlock probe hybridization, ligation, and amplification. Padlock probes are oligonucleotides whose ends are complementary to adjacent target sequences [24, 39]. During hybridization, the two binding portions of the padlock must perfectly hybridize to the target to bring the 5' and 3' ends into contact, allowing circularization of the probes by ligation. Rolling circle amplification then generates multiple copies of the circularized probes, producing DNA nanoballs. To assess the performance of RNAsky, normal tonsil samples were interrogated with a Human I-O Core Panel, composed of 21 target genes, two housekeeping genes, and negative controls. Regions of interest (ROI) containing a minimum of two follicles across three serial sections were evaluated to determine intra-sample and inter-sample reproducibility, as well as cross technology comparison (Fig. 1b-e). To correct for sample-to-sample variation, RNAsky counts were normalized to the higher expressed housekeeping gene in the panel, *PSMB4*. RNAsky showed high concordance across serial sections of the same block (r=0.9522 or higher) (Fig. 1b, Supplementary Fig. 1a) with an average of 22,782 cells profiled across the three sections. RNAsky was also highly concordant between two independent tissue samples with similar regions of interest assayed (Tonsil 1= 22,782 cells, Tonsil 2 = 21,313 cells, r=0.9370) (Fig. 1c). We then compared RNAsky data with two independent RNA quantification technologies, bulk RNA-Seq and single-cell RNA-Seq to benchmark the assay's performance (Fig. 1d-e, Supplementary Fig. 1b). RNAsky showed high correlation with both methods (Bulk RNA-Seq correlation r= 0.9329, scRNA-Seq correlation r= 0.8953).

We developed RNAsky as a flexible assay, where the number and composition of targets can be customized according to study needs. To highlight this flexibility and expand cell type characterization, four additional gene targets were added to the I-O Core Panel (*BCL6*, *CD68*, *CD14*, *LAMP3*, referred to as extended I-O panel) in cancer samples of interest. Specificity was evaluated by calculating the false discovery rate (FDR) across

multiple replicates per tissue type (Fig. 1f). RNAsky demonstrated highly specific performance, with all FDR's falling below 0.3%. The number of counts per cell varied across tissue types, with an average of 10.8 counts per cell in all tissues (Fig. 1g, Supplementary Fig. 1c). To benchmark the sensitivity of RNAsky, the ratio of mean counts per cell for each gene was calculated in both RNAsky and scRNA-Seq tonsil data (Supplementary Fig. 1d). RNAsky showed higher sensitivity for low and medium expressed genes, but this sensitivity dropped for highly expressed transcripts. For highly expressed genes, switching from a discrete count-based quantification to continuous intensity-based quantification provided a robust alternative to assess transcript abundance (Supplementary Fig. 2). The median sensitivity was calculated for all genes across three tonsil blocks, and showed higher sensitivity compared to scRNA-Seq (Fig. 1h). Overall, RNAsky provides spatially resolved quantification of transcripts at subcellular resolution with high reproducibility, specificity, and sensitivity (Fig. 1i).

The combination of RNA and protein analyses identifies major cell types within normal tonsil tissue

The ability to capture tissue architecture on a transcriptional level was investigated by assessing the gene expression patterns of the extended I-O panel in normal tonsil (Fig. 2). The spatial distribution of detected genes demonstrated major characteristics of tonsil histology, including the germinal center and marginal zone of lymphoid follicles (Fig. 2, Region 1), the interfollicular zone (Fig. 2, Region 2), and squamous epithelium (Fig. 2, Region 3). The patterns of gene expression were localized across all panel targets (Supplementary Figs. 3-5).

Because RNA and protein data were acquired on the same section, the relationship between RNA and protein expression profiles could be assessed within a spatial context (Fig. 3a). For genes such as *CD3D* and *MKI67*, expression was enriched in cells that also expressed the corresponding proteins markers. To identify canonical cell populations of the tonsil based on gene expression profiles, a stringent gating strategy using normalized transcript counts (see methods) was applied to pairwise gene relationships that define different cell types (Fig. 3b, Supplementary Fig. 6a). B cells were defined by the presence of *MS4A1*, while T cells were identified by the absence of this gene and presence of *CD3D* and/or *CD8A*. Then, additional selection gates were applied to further characterize these two types of lymphocytes. Germinal center cells were characterized by double expression of *MS4A1* and *BLC6*, and proliferating cells within this region were further identified with the presence of *MKI67*. Cytotoxic T cells showed a combination of *CD3D* and *CD8A* expression, while remaining T cells were characterized by *CD3D* expression alone. Finally, epithelial cells were defined by the expression of three hallmark genes *EGFR*, *EPCAM*, and *ERBB2* and the absence of *MS4A1*.

To investigate cell population dynamics of the tonsil, major cell types were determined using hierarchical clustering based on protein signal intensities (Fig. 3c, Supplementary Fig. 6b). Eight major cell types were identified including B cells, T cells, germinal center B cells, proliferating B cells, cytotoxic T cells, plasma cells, macrophages, and epithelial cells. Clustering based on both transcript counts and protein intensities further improved cell population annotation by refining the boundaries between populations (Fig. 3d, Supplementary Fig. 6c). A heatmap of intensity-based quantification of select protein and mRNAs allowed for the visualization of expression differences between spatially distinct cell type clusters (Fig. 3e). For example, one of the five CD4 T cell groups (second column) may be proliferative with higher protein and RNA levels of the proliferation markers PCNA and Ki67. Additionally, this heatmap exhibited correlations of RNA and protein to define cell types, such as plasma cells. This cell type showed a high correlation of *IGHG1* and *IGKC*

transcripts, and CD79a and CD138 marker expression. Thus, using the intensity-based detection method, plasma cells were visualized by mRNA expression with cells expressing either *IGHG1* (red arrows), *IGKC* (green arrows), or both (yellow arrows) (Fig. 3f, Supplementary Fig. 7).

Multiomic analyses capture the dynamics of the tumor microenvironment and expression signatures across different cancer systems

The extended I-O panel enabled robust transcriptional analyses across cancer samples. To determine the panel's ability to capture canonical cancer and immune signatures, gene expression profiles were assessed across four cancer samples (Fig. 4, Supplementary Fig. 8). High *GATA3* expression is commonly seen in well differentiated breast cancers (BC), and was observed in the triple positive breast cancer sample (Fig.4a) [40]. The non-keratinizing squamous cell carcinoma (NKSCC) sample showed high *CDKN2A* expression (Fig. 4b). This is in alignment with cBioPortal for Cancer Genomics data sets showing high *CDKN2A* levels in 97% of patients with HPV+ head and neck squamous cell carcinomas [41, 42, 43]. Immune infiltration is a key factor in understanding cancer progression and prognostic outcomes. Three of the tumor ROIs exhibited the presence of an immune infiltrate. In the breast cancer sample (Fig. 4a), T and B cells can be observed throughout the tumor. The NKSCC (Fig 4b) and colorectal cancer (CRC) (Fig. 4c), showed organization of T and B cells into tertiary lymphoid structures (TLSs), with presence of *BCL6* indicating germinal center formation (Fig. 4c) [44].

The combination of high-resolution transcriptomics and proteomics data enables characterization of the cell interactions and molecular mechanisms occurring within the spatial context of the tumor microenvironment (Fig. 5). A tonsil control was included in these analyses as a normal tissue reference. Each cancer system showed a distinct profile, highlighting the various levels of gene expression in different tumors (Fig. 5a, Supplementary Fig. 9). Following the Seurat pipeline, dimension reduction, clustering, and curated cluster annotation was performed (Fig. 5b, see methods) [45]. Five cell populations were defined based on gene expression signatures, including cancer/epithelial cells, T cells, B cells, plasma cells, and macrophages/monocytes. (Fig. 5b bottom). Using the cluster annotations, cell population dynamics, both quantitatively and spatially, were evaluated (Fig. 5c-d). The tonsil showed a low presence of the cancer/epithelial cell group, representing the tonsillar crypt epithelial cells, while the tumor ROIs showed a significant amount of malignant epithelial cells. This is reflected in the enrichment of the epithelium population within all tumor ROIs (unpaired t-test with Welch's correction, $p<0.05$).

Adding protein information to RNA showed the functional organization of the tumor microenvironment (Fig. 5e, Supplementary Figs. 10-12). The tonsil and nonkeratinizing squamous cell carcinoma showed strong expression for cytokeratin 5 (CK5) and high molecular weight cytokeratin (CK HMW) in the tonsillar crypt epithelium and malignant squamous epithelium respectively (Supplementary Fig. 10). The breast carcinoma sample showed a characteristic cytokeratin 7 (CK7) positive, cytokeratin 20 (CK20) negative cytokeratin profile and the colorectal carcinoma the inverse (CK7 negative / CK20 positive) as expected. As seen in Figure 5e, malignant epithelium (purple and green) was identified in the upper two quadrants of the CRC image, while normal epithelium (green) was located on the bottom left quadrant. Two TLSs (red and blue) were located in the bottom right quadrant of the image, as characterized by their RNA and protein expression and histological appearance. Using hierarchical clustering based on protein expression, ten dominant cell populations were identified representing epithelial cells, cancer cells, proliferating cancer cells, stromal cells, and immune cells (Fig. 5f). Because erasure methods used were gentle, samples remained intact at the completion of image acquisitions, enabling downstream

hematoxylin and eosin (H&E) staining on the same tissue section. The H&E images were registered with DAPI images from the RNA and protein image acquisition. Once registered, RNA, protein, and histological data could be simultaneously viewed in the MACS iQ View software platform (Fig. 5g, Supplementary Fig. 13), allowing for direct comparison of classic histology with spatially valuable transcriptomic and proteomic information on the same section.

Discussion

In this study, we introduced a true multiomics approach, where the chemistry for RNA detection was compatible with subsequent immunofluorescent epitope detection, enabling the combined analysis of RNA and protein targets on a single, histological tissue section. Here, RNAsky showed a high reproducibility for serial sections, which now serves as the basis for spatially unbiased results from MICS technology on the same tissue section. Additionally, because the methods used were gentle, samples remained intact at the completion of all image acquisitions, enabling downstream histology staining (H&E) on the same section. Transcriptomic and proteomic approaches have identified a substantial degree of variation in both, protein and mRNA levels in most, if not all types of cancer [46, 47, 48]. This observed heterogeneity often increases as tumors progress. In addition, kinetics and intensity of RNA and protein expression of the same gene can differ due to post transcriptional control. For example, miRNAs negatively regulate the expression of a large proportion of mRNAs [49, 50] and translational regulation/post translational modifications alter protein stability and abundance [51, 52]. The visualization of protein and RNA in the context of a canonical diagnostic stain provides the link between mRNA and protein expression and qualitative image analysis/interpretation (H&E). This increases the level of confidence when determining the correlation or the divergence between all three established information modalities. It also enables more accurate phenotyping, determination of cell states, tumor microenvironment profiling, cell segmentation, and clustering, all of which are key to capturing molecular changes contributing to the progression and cellular response of disease. Finally, it can also be used to study the mechanisms of gene regulation in a spatial context.

The approach described here is based on *in situ* hybridization and amplification of probes for RNA transcript detection. The padlock probes in RNAsky were designed to anneal directly to mRNA, eliminating the need for cDNA synthesis therefore improving the detection sensitivity. While we have used a first set of padlocks to evaluate the technology, the addition of new gene targets to a panel is straightforward. As each gene is detected by a single detection probe sequence independently of all other targets, users can readily combine or modify gene panels. Additionally, the size of a panel is flexible and the upper limit is only restricted by the processing time. When we assessed the FDR, calculated across multiple replicates, for the transcript detection approach, the proportion of false positive genes in the extended I-O panel was extremely low (mainly less than <0.1%). This low FDR indicates a high degree of specificity of RNAsky, which requires the hybridization of the two arms of the padlock probes to occur, and a ligation event to generate the amplified molecule underlying the observed punctate signal. Our punctate detection approach shows a higher sensitivity for low and medium expressed genes compared to scRNA-Seq, but a lower sensitivity for highly expressed transcripts. Intuitively, this observation is in alignment with higher transcript expression being more prone to optical and physical overcrowding. To mitigate this issue, we developed a computational approach, integrated into the MACS iQ View analysis software, where the discrete count-based quantification is supplemented with a continuous intensity-based quantification, allowing for a robust quantification and an increase in the dynamic range of detection of RNAsky. Future technological iterations of the MACSima platform will

enable z-stack acquisition, where multiple images are taken at different focal distances. By acquiring three-dimensional data, the composite image will have a greater depth of field and will better resolve punctate amplicon signal.

The spatial multiomics analyses performed on the same tissue section revealed the dynamics of cell populations in both healthy and diseased tissues. Distinct profiles of expression and immune signatures based on RNA, protein or a combination could be drawn. Using stringent gating strategies or hierarchical clustering followed by cluster annotation, cell population dynamics were evaluated both quantitatively and spatially. In tonsil, the corresponding mRNA and protein marker expression were enriched in specific cell type cells, allowing the identification of canonical cell populations such as T cells, germinal center B cells, and epithelial cells.

Although scRNA-Seq captures the transcriptional activity of individual cells in tissues, it lacks spatial information. Here, we investigated the spatial expression patterns of RNA and proteins in the microenvironment of tumor. The tumor microenvironment is a complex ecosystem that is composed of various cell types, such as tumor, immune, and stromal cells. This unique environment varies by tumor type and tumor progression over time [53, 54]. TLS are heterogenous ectopic lymphoid organs that form at inflammatory and cancerous sites. Without the power of spatial immunostaining and transcriptomics information, it is difficult to assess the activation and maturation stage of tumor-associated TLSs. Spatial analyses can identify aggregates of T cells, B cells, and dendritic cells that are characteristic of these structures, and additional protein and transcriptional data can elucidate maturation characteristics.

Given the complexity of studying internal tumor heterogeneity, combining proteomic and transcriptomic data on the same tissue section would increase our understanding of cancer dynamics and biological variations, including the discrepancy observed in mRNA to protein ratio shown recently in multiple cancer types [46, 47, 48]. The degree to which mRNA levels scale with protein abundances and the implications in cases where this dependency breaks due to dysregulation, could lead to new avenues for cancer research. While we concentrated in our study on the analysis of proteins and mRNAs it is worth noting that the technology can in theory be expanded to other types of molecules. This could include long non-coding RNAs and all kinds of epitopes detectable by antibodies such as cytokines, glycolipids, carbohydrates, or neurotransmitters. The combinatorial approach described here, using the MACSima platform, overcomes a key limitation of current technologies, and could potentially lead to the identification of new clusters or multimodal patterns of expression, further enabling the discovery and development of predictive and prognostic cancer biomarkers.

Methods

Biospecimen Samples

PSN1 cell pellet FFPE blocks and normal tonsil FFPE blocks were purchased from AMSBIO (3120-0410, AMS6022 respectively). Cancer blocks were multi-omics grade biospecimens purchased from Indivumed Services.

Colorectal cancer (CRC) – The sample is a low-grade adenocarcinoma of the ascending colon.

Breast cancer (BC) – The sample is an ER+, PR+, HER2 3+ invasive carcinoma of the breast.

Non-keratinizing squamous cell carcinoma (NKSCC) – The sample is an HPV+, poorly differentiated non-keratinizing squamous cell carcinoma of the oropharynx.

Metastatic lymph node (Met. LN) – The sample is a lymph node metastasis of a colorectal adenocarcinoma.

Tissue Processing

For all tissue blocks, 5 µm sections were acquired on a HM355 S automatic microtome (Epredia), collected on high resolution slides (Miltenyi Biotec), and baked for one hour at 60°C in a symphony™ incubator (VWR) from each of the tissue blocks. The slides were deparaffinized in Histoclear II (National Diagnostics), following the manufacturer's instructions, and rehydrated in nuclease free water. The slides then underwent antigen retrieval in tris-EDTA citrate (TEC) buffer pH9 for 20 min at 98°C in a PT Module (Epredia). After washing twice with nuclease free water, the slides were mounted into MACSwell Imaging Frames (Miltenyi Biotec).

Hematoxylin and Eosin (H&E) Staining

H&E stains were performed according to the kit's instructions (Vector Laboratories, H-3502) for pathological assessment and informed Region of Interest (ROI) selection. After H&E staining, coverslips were mounted with Cytoseal XYL (Epredia). The samples that underwent H&E staining after RNA and protein acquisition were processed in the same method as serial sections.

RNAsky probes and Immuno-Oncology Core Panel design

RNAsky utilizes targeted padlock probes for gene identification. Each probe contains two arms which hybridize to two adjacent sequences of an mRNA of interest, as well as a gene specific sequence used for detection. The mRNA targeting regions of the probes were designed with Miltenyi Biotec's proprietary probe design pipeline. The pipeline was optimized to produce specific and sensitive probes by filtering candidates on multiple parameters including melting temperature, GC content, secondary structure formation, homopolymers, repetitive sequences, and off target events. The I-O Core Panel includes 21 gene targets representing canonical and actionable hallmarks across different cancer systems, two housekeeping genes, and three negative control probes. The two housekeeping genes were

selected for their expression levels as previously described [55], and were used for data normalization. The negative control probes were designed to not bind anywhere in the human transcriptome and provide a measure of specificity of the system. In addition to the I-O Core panel, four custom gene targets were designed and included: *BCL6*, *CD14*, *CD68*, and *LAMP3*. A list of gene targets can be found in Supplementary Table 1.

RNAsky Sample Preparation

RNAsky (Miltenyi Biotec) sample preparation began immediately after MACSwell Imaging Frame assembly, following Miltenyi's RNAsky user manual for FFPE sections. In brief, probes were hybridized overnight at 45°C in a humidified chamber. Following hybridization, samples were washed with a stringency buffer, and then probes were ligated overnight at 37°C. After ligation, samples were again washed, and an amplification primer was annealed. Rolling circle amplification (RCA) was performed at 30°C overnight. Finally, samples were crosslinked for one hour at room temperature to stabilize amplicons prior to loading samples on the MACSima system.

MACSima Imaging Cyclic Staining (MICS) Technology

Cyclic imaging on the MACSima System was performed as previously described [56, 57, 58] with the addition of cyclic RNA acquisition to generate multiomic data sets. The MACSima Platform combines widefield microscopy with fully automated liquid handling to perform cyclic imaging. The instrument iteratively stains analytes of interest, acquires multispectral images, and erases the signal at the end of each round. For RNAsky rounds, signal was erased with Miltenyi's proprietary RNAsky release reagent, while antibody rounds were either optically bleached or signal was removed using the REAlease® Release Reagent (Miltenyi Biotec). Briefly, before loading samples onto the instrument, regions of interest were selected based on a pathologist's assessment of H&E sections. ROIs were selected by identifying well preserved tissue representative of the anatomical structure and the pathology encountered, including normal adjacent tissue and associated tertiary lymphoid structures when present. Areas of overt necrosis, pigmentation, hemorrhage and artifacts were avoided. The size of ROI for all tonsil samples and metastatic lymph node is 1.6x1.3mm, 2.6x1.8 mm for the NKSCC and breast cancer, and 2.1x2.2 mm for the CRC. A list of gene targets and antibodies can be found in Supplementary Table 2.

Image Analysis

All datasets contained RNA and protein images and were analyzed in the MACS iQ View analysis software (Miltenyi Biotec). Every field of view was assessed across two exposure times during image preprocessing. Flatfield and distortion corrections were applied based on calibration data from the instrument [33]. In the event of pixel overexposure with the highest exposure time, the signal was normalized with the lower exposed image to generate a high dynamic range image (HDR). Then, DAPI images were used for stitching and registration across the entire ROI for each cycle. Finally, signal images had the background subtracted based on background images acquired for every cycle. Segmentation was performed based on DAPI images in a two-step process, with nuclear segmentation based on the watershed method followed by a donut based cytoplasmic segmentation. A gaussian smoothing was performed on DAPI images to improve cell segmentation for tumor tissues. After segmentation cell features are measured across all channels.

Transcript Detection

Utilizing the integrated algorithm in MACS iQ View, transcripts were detected with default parameters. With this algorithm, transcripts are detected in two stages for enhanced accuracy. First, a multi-scale object detection strategy is employed utilizing the Laplacian of Gaussian (LoG) filter. This process evaluates the maximum response over a range of Gaussian kernel sizes, allowing for the identification of objects of various sizes and across multiple focal planes. The outcome of this process is an image characterized by the Laplacian transformation, which is then thresholded to create a collection of high-response regions where punctate signal has been detected. In the second stage, in order to obtain an exact list of object positions, wavelet filtering of the preprocessed images is performed followed by local peak detection. Any local maxima not found within the high-response regions from stage 1 are discarded as false positives, while the remainder form our final curated set of detected transcripts.

RNA gating cell population analysis

After transcript detection, gene counts were normalized using the MinMax algorithm in MACS iQ View. Briefly, the MinMax normalization algorithm transforms all values for a given feature into the range [0,1], where the minimum and maximum value for that feature is 0 and 1 respectively. Then, pairwise scatterplots were generated, and selection gates were established based on desired gene relationships. For example, when using gene A and gene B, a user can select a gate for double positive expression (A and B), or an exclusion gate (A not B). A very stringent approach was used where cells were not allowed to be represented in more than one population group.

Protein clustering analysis

For protein clustering, the fluorescence intensity of each marker was measured and a Z-normalization used across all protein measurements. Then, principal component analysis for 4 dimensions was performed. These were then used as the input for the hierarchical clustering within MACS iQ View with the following parameters: agglomerative hierarchical clustering algorithm, number of clusters 30, Euclidean distance (L2 norm), non-squared, and Ward's method.30 clusters were generated, and cell types were assigned to a cluster based on the five most expressed markers. Clusters were combined when the same cell type was assigned in separate clusters. The identified cell types were colored and plotted in an XY plot using the nuclear segmentation mask as the underlying map. Expression heatmaps were generated displaying both protein and RNA intensity values across the annotated spatially grouped protein-based cell type clusters.

Multiomic clustering analysis

RNA count data and protein intensity data were used as input features for the agglomerative hierarchical clustering algorithm in MACS iQ View with the following parameters: number of clusters 20, Euclidean distance (L2 norm), non-squared, and Ward's method. After initial clustering, clusters were filtered and annotated to show cell types.

Intensity-based Transcript Quantification

The overall intensity of signals from detected RNA transcripts in segmented cells was measured in the same way the amount of signal from fluorescent antibodies is assessed. The intensity-based method is compatible with software developed for fluorescence microscopy and uses the same methodology as the antibody staining method. For the comparison of both count- and intensity-based transcript quantification, the DAPI image and the RNAsky image of the *MS4A1* (CD20) were imported into QuPath [59]. The cell nuclei were

segmented using the DAPI channel and a defined area around the nuclei was used as the cytoplasm. Using the experimental organelle detection method within QuPath, the detected RNA transcripts were segmented and counted per cell. From these data both the intensity and the count data were extracted and exported into a data table and further analyzed.

MICS and H&E registration

H&E staining was performed following RNAsky and MICS run on MACSima as described above, and whole slide images were acquired. For image registration of H&E image to DAPI channel, selected ROI area was cropped and then stacked with inverted cycle 0 DAPI image in Fiji. The stacked images were aligned using Fiji's plugin, Linear Stack Alignment with SIFT [60] and merged by Z projection in Fiji. To combine transcripts, protein and H&E images, the registered H&E image was imported into MACS iQ View analysis software. The stacked images were aligned using Fiji's plugin, Linear Stack Alignment with SIFT [60] and merged by Z projection in Fiji. To combine transcripts, protein and H&E images, the registered H&E image was imported into MACS iQ View and overlaid with detected transcripts and protein images acquired from MACSima.

<u>RNAsky performance and cross-technology comparisons</u>

For the following analyses, *IGHG1*, *IGKC* counts were excluded due to their very high expression.

False Discovery Rate (FDR) Calculation

FDR (%), a measure of specificity, is the rate of calls from negative control probes (n=3) per total number of calls for a region of interest (ROI). The FDR calculation was determined based on the I-O Core Panel for Tonsils 1 and 2 and the extended I-O Panel for all remaining tissues.

Bulk Sequencing Correlation

Pearson correlation with 95% confidence intervals was determined on the I-O Core Panel genes between the mean counts per ROI of a representative tonsil sample and nTPM of a bulk RNA-Seq tonsil dataset. Two-tailed P value was also determined.

Cell Line Correlation

The counts or TPM for each gene were normalized to *PSMB4* counts or TPM. Pearson correlation with 95% confidence intervals was determined on the I-O Core Panel genes between the average of two PSN1 RNAsky normalized counts and a PSN1 Cancer Cell Line Encyclopedia (CCLE) bulk RNA-Seq dataset. Two-tailed P value was also determined.

Single Cell Sequencing (scRNA-Seq) Correlation

Pearson correlation with 95% confidence intervals was determined on the I-O Core Panel genes between the mean counts per cell of a representative tonsil sample and mean counts per cell of a scRNA-Seq healthy young adult tonsil dataset. Two-tailed P value was also determined.

In situ Sensitivity Calculations

Both RNAsky and scRNA-Seq are able to measure transcripts per cell allowing for the determination of how sensitive RNAsky *in situ* transcript detection is relative to scRNA-Seq transcript detection. In healthy tonsil RNAsky datasets and a scRNA-Seq healthy young adult tonsil dataset, transcripts/cell was determined for each gene followed by the ratio of RNAsky

to scRNA-Seq for each gene. Median sensitivity was determined from the ratios all the I-O Core Panel genes.

Gene expression analyses

Transcript count preprocessing: filtering and transcript count normalization

For each sample, cells with less than three RNAsky genes detected were filtered out, and RNAsky genes detected in less than three cells were excluded from downstream analysis. The raw transcript counts from all RNAsky samples were first merged to a single Seurat object by Seurat (version 4.3.0) function 'merge' [45]. To normalize transcript counts across all cells, the raw gene transcript count for each cell was divided by the total transcript count of that cell, multiplied by scale factor 10,000, and then natural log transformed with pseudo count set as 1. Seurat function NormalizeData(normalization.method="LogNormalize", scale.factor=1e4) was used for the log normalization. Subsequently, the log normalized counts were transformed to Z-score by 'Seurat::ScaleData()', and finally gene expression variance was stabilized by 'Seurat::SCTransform' .

RNAsky transcript count dimensionality reduction and batch effect correction

Principle Component Analysis was performed on the merged object as described above with all RNAsky genes included and total number of PCs set as 26 (Seurat::RunPCA(npcs=26)). To correct batch effect, all samples in the merged object were integrated by the fast, sensitive, and scalable integration tool Harmony, and performed by Seurat wrapper 'RunHarmony' [61, 62]. Cells from individual samples were considered as different batches (specified by RunHarmony parameter group.by). The integration process generated harmony components, which were all used subsequently as input for UMAP (Seurat::RunUMAP(dims=1:26)) and clustering analysis.

Clustered heatmap plotting, cell type annotation and spatial reconstruction of cell types

The single cell clustered heatmap was generated based on the above log normalized counts, which was scaled between 0 and1. The plotting was performed by the R function dittoSeq::dittoHeatmap(scaled.to.max=TRUE).

To cluster cells for cell type annotation, all the harmony components were utilized to construct a shared nearest neighbor (SNN) graph. This was performed by Seurat::FindNeighbors(dims=1:26) function with the other parameters kept at their default values. Subsequently Louvain algorithm was employed to identify clusters by Seurat::FindClusters(algorithm=1, resolution=0.1). The "resolution" parameter for Louvain clustering determines the size and number of detected clusters. In this study, we tested resolution values ranging from 0.1 to 0.5. When resolution was set at 0.1, a total of seven clusters were identified, which aligns well with the expected number of cell types.

The resulting clusters were automatically annotated by R package SingleR (version 2.0.0) against two general-purpose built-in references: Human Primary Cell Atlas and Blueprint/ENCODE,which are provided by R package celldex (version 1.8.0) [63]. Additionally, we annotated the clusters by R package Azimuth (version 0.4.6) function 'RunAzimuth' against the Human tonsil reference [45]. Marker genes for each cluster were identified by Seurat::FindAllMarkers(test.use = "wilcox", min.pct=0.1) function with log normalized counts as input and the default Wilcoxon rank sum test. The average of log normalized counts for each cluster was visualized by R function 'pheatmap::pheatmap' (version 1.0.12). The final annotation was determined by considering both automatic cell annotation and marker gene expression level.

The spatial reconstruction of cell types was achieved by the 'scatter' method in Python library plotly.express (Plotly Technologies Inc.). The input data is the cell coordinates generated by cell segmentation.

Datasets

The gene expression for all 23 targets in the RNAsky Immuno-Oncology Core Panel was retrieved from the following data sources:

Human Protein Atlas (HPA) Healthy Tonsil RNASeq

The RNA HPA tissue gene data set was downloaded from the Human Protein Atlas (http://www.proteinatlas.org) and is based on the Human Protein Atlas version 23.0 and Ensembl version 109 [8, 64].

Single Cell Tonsil Data Set

Single cell RNA sequencing data was downloaded from the deposited Seurat objects as described in the tonsil atlas publication [65]. The non-cell-hashed dataset from three healthy young adult donors ages 26, 33, and 35 were used for the analysis in this study. Single cell RNA sequencing data was downloaded from the deposited Seurat objects as described in the tonsil atlas publication [65].

Cancer Cell Line Encyclopedia (CCLE) RNASeq

The CCLE data set was downloaded from the DepMap Public 22Q2 release on the DepMap portal (http://depmap.org). The PSN1 cell line was selected as a representative cancer system for marker assessment [66].

Statistics

Unless otherwise specified, all bar graphs show the mean with error bars representing the standard error of the mean. Pearson correlation coefficients with 95% confidence intervals and two-tailed P values were determined for all correlations. Unpaired t tests with Welch's correction with two-tailed P values were performed for comparisons. Statistical analysis performed using GraphPad Prism 10.

Acknowledgements

We thank Dr. Christian Wöhle for bioinformatics support. We thank Rami Hayajneh for histology image acquisition support. We thank the Molecular Chemistry team in Waltham, MA for their support in RNAsky reagent and protocol developments. This publication is part of the Human Cell Atlas – www.humancellatlas.org/publications/.

Author contributions

E.N., D.P., R.C.H., R.P., W.M., T.W., A.B, J.F., S.R. designed studies.

E.N., D.P., R.C.H., M.D., E.D., H.L., E.L., H.L., S.S. performed experiments.

E.N., D.P., R.C.H., R.P., J.W., W.M., T.W., A.B., A.M. wrote manuscript.

E.N., D.P., R.C.H., J.W., W.M., T.W., A.M., F.S., D.M., P.P., H.L., E.L., J.F. analyzed the data and interpreted results

E.N., D.P., R.C.H., J.W., W.M., T.W., R.P., A.B., J.F., H.M., T.R., R.H. revised and edited the manuscript

Materials & Correspondence

Correspondence and material requests should be addressed to Robert Pinard (robertpi@miltenyi.com) or Andreas Bosio (Andreasbo@miltenyi.com).

Competing Interests

The authors declare the following competing interests: All authors are employees of Miltenyi Biotec.

Data availability

As examples, the RNAsky and antibody-stained original stitched images of the Tonsil and the CRC tissue shown in Figures 4 and 5, are uploaded to Zenodo in the OME.TIF file format and the links will be provided after acceptance of the manuscript.

Supplementary Information

In attached Supplementary Information file.

References

1. Consortium H. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187-192 (2019).

2. Cool J, Conroy RS, Hanlon SE, Hughes SK, Roy AL. Spatial and temporal tools for building a human cell atlas. *Mol Biol Cell* **30**, 2435-2438 (2019).

3. Regev A*, et al.* The Human Cell Atlas. *Elife* **6**, (2017).

4. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci* **27**, 233-244 (2018).

5. Thul PJ*, et al.* A subcellular map of the human proteome. *Science* **356**, (2017).

6. Uhlen M*, et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, (2017).

7. Börner K*, et al.* Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nat Cell Biol* **23**, 1117-1128 (2021).

8. Karlsson M*, et al.* A single-cell type transcriptomics map of human tissues. *Sci Adv* **7**, (2021).

9. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).

10. Consortium AA. Aging Atlas: a multi-omics database for aging biology. *Nucleic Acids Res* **49**, D825-D830 (2021).

11. Svensson V*, et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat Methods* **14**, 381-387 (2017).

12. Ziegenhain C*, et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* **65**, 631-643.e634 (2017).

13. M C*, et al.* Spatially resolved transcriptomics: A comprehensive review of their technological advances, applications, and challenges. *Journal of genetics and genomics = Yi chuan xue bao*, (2023).

14. Moffitt JR, Lundberg E, Heyn H. The emerging landscape of spatial profiling technologies. *Nat Rev Genet* **23**, 741-759 (2022).

15. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol* **41**, 773-782 (2023).

633
634 16. Walker BL, Cang Z, Ren H, Bourgain-Chang E, Nie Q. Deciphering tissue structure and function
635 using spatial transcriptomics. *Commun Biol* **5**, 220 (2022).

636
637 17. Cheng M*, et al.* Spatially resolved transcriptomics: a comprehensive review of their
638 technological advances, applications, and challenges. *J Genet Genomics* **50**, 625-640 (2023).

639
640 18. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly
641 multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

642
643 19. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput single-cell gene-
644 expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc*
645 *Natl Acad Sci U S A* **113**, 11046-11051 (2016).

646
647 20. Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells Reveals Spatial
648 Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342-357 (2016).

649
650 21. Eng CL*, et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature*
651 **568**, 235-239 (2019).

652
653 22. Goh JJL*, et al.* Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat Methods*
654 **17**, 689-693 (2020).

655
656 23. He S*, et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by
657 spatial molecular imaging. *Nat Biotechnol* **40**, 1794-1806 (2022).

658
659 24. Ke R*, et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10**,
660 857-860 (2013).

661
662 25. Lee JH*, et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in
663 intact cells and tissues. *Nat Protoc* **10**, 442-458 (2015).

664
665 26. Liu S*, et al.* Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel
666 in situ analyses. *Nucleic Acids Res* **49**, e58 (2021).

667
668 27. Wang X*, et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states.
669 *Science* **361**, (2018).

670
671 28. Chen X, Sun YC, Church GM, Lee JH, Zador AM. Efficient in situ barcode sequencing using
672 padlock probe-based BaristaSeq. *Nucleic Acids Res* **46**, e22 (2018).

673
674 29. Janesick A*, et al.* High resolution mapping of the breast cancer tumor microenvironment
675 using integrated single cell, spatial and in situ analysis of FFPE tissue. (2022).

676

677 30. Bourceau P, *et al.* Visualization of metabolites and microbes at high spatial resolution using
678 MALDI mass spectrometry imaging and in situ fluorescence labeling. *Nat Protoc* **18**, 3050-
679 3079 (2023).

680
681 31. Ball JB, *et al.* Combining RNAscope and immunohistochemistry to visualize inflammatory
682 gene products in neurons and microglia. *Front Mol Neurosci* **16**, 1225847 (2023).

683
684 32. Li X, *et al.* Combined single-molecule fluorescence. *STAR Protoc* **2**, 100555 (2021).

685
686 33. Kinkhabwala A, *et al.* MACSima imaging cyclic staining (MICS) technology reveals

687 combinatorial target pairs for CAR T cell treatment of solid tumors *Scientific Reports* **12**, 1-16
688 (2022).

689
690 34. Giesen C, *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by
691 mass cytometry. *Nat Methods* **11**, 417-422 (2014).

692
693 35. Lin JR, *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors
694 using t-CyCIF and conventional optical microscopes. *Elife* **7**, (2018).

695
696 36. Manesse M, Patel KK, Bobrow M, Downing SR. The InSituPlex. *Methods Mol Biol* **2055**, 585-
697 592 (2020).

698
699 37. Black S, *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat Protoc*
700 **16**, 3802-3835 (2021).

701
702 38. Keren L, *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast
703 Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e1319 (2018).

704
705 39. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock
706 probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085-2088
707 (1994).

708
709 40. Perou CM, Børresen-Dale AL. Systems biology and genomics of breast cancer. *Cold Spring*
710 *Harb Perspect Biol* **3**, (2011).

711
712 41. Cerami E, *et al.* The cBio cancer genomics portal: an open platform for exploring
713 multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).

714
715 42. Gao J, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the
716 cBioPortal. *Sci Signal* **6**, pl1 (2013).

717
718 43. de Bruijn I, *et al.* Analysis and Visualization of Longitudinal Genomic and Clinical Data from
719 the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res*, (2023).

720
721 44. Ruffin AT, *et al.* B cell signatures and tertiary lymphoid structures contribute to outcome in
722 head and neck squamous cell carcinoma. *Nat Commun* **12**, 3349 (2021).

723
724 45. Hao Y, *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529
725 (2021).

726
727 46. Arad G, Geiger T. Functional Impact of Protein-RNA Variation in Clinical Cancer Analyses. *Mol*
728 *Cell Proteomics* **22**, 100587 (2023).

729
730 47. Ben-Chetrit N, *et al.* Integration of whole transcriptome spatial profiling with protein markers.
731 *Nat Biotechnol* **41**, 788-793 (2023).

732
733 48. Chen G, *et al.* Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell*
734 *Proteomics* **1**, 304-313 (2002).

735
736 49. Michlewski G, Cáceres JF. Post-transcriptional control of miRNA biogenesis. *RNA* **25**, 1-16
737 (2019).

738
739 50. Landgraf P, *et al.* A mammalian microRNA expression atlas based on small RNA library
740 sequencing. *Cell* **129**, 1401-1414 (2007).

741
742 51. Robichaud N, Sonenberg N. Translational control and the cancer cell response to stress. *Curr*
743 *Opin Cell Biol* **45**, 102-109 (2017).

744
745 52. Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J, Topisirovic I. Targeting the
746 translation machinery in cancer. *Nat Rev Drug Discov* **14**, 261-278 (2015).

747
748 53. Chen X, Song E. The theory of tumor ecosystem. *Cancer Commun (Lond)* **42**, 587-608 (2022).

749
750 54. de Visser KE, Joyce JA. The evolving tumor microenvironment: From cancer initiation to
751 metastatic outgrowth. *Cancer Cell* **41**, 374-403 (2023).

752
753 55. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet* **29**, 569-574
754 (2013).

755
756 56. Kinkhabwala A, *et al. Scientific Reports* **12**, 1-16 (2022).

757
758 57. Schäfer D, *et al.* Identification of CD318, TSPAN8 and CD66c as target candidates for CAR T
759 cell based immunotherapy of pancreatic adenocarcinoma. *Nature Communications* **12**, 1-18
760 (2021).

761
762 58. Guilliams M, *et al.* Spatial proteogenomics reveals distinct and evolutionarily conserved
763 hepatic macrophage niches. *Cell* **185**, 379-396.e338 (2022).

59. Bankhead P, *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).

60. Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* **60**, 91-110 (2004).

61. Korsunsky I, *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).

62. Tran HTN, *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* **21**, 12 (2020).

63. Aran D, *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).

64. Sjöstedt E, *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, (2020).

65. Massoni-Badosa R, *et al.* An Atlas of Cells in the Human Tonsil. *bioRxiv*, 2022.2006.2024.497299 (2022).

66. Ghandi M, *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503-508 (2019).

Figure Legends

Fig. 1. Schematic workflow and performance of RNAsky.
a Overview of MACSima Imaging Cyclic Staining (MICS), an automated, multiomic approach to integrate spatial protein and RNA analysis on a single tissue section. **b** Representative correlation of RNAsky counts between independent runs (serial section replicates) of a representative normal tonsil tissue. Rep. 1 vs Rep. 2 Pearson correlation coefficient (PCC) = 0.9522 (95% confidence intervals = 0.8865 to 0.9803); P value (two-tailed) <0.0001. Remaining correlations in Supplementary Fig. 1. **c** Correlation of RNAsky counts obtained in two different normal tonsil blocks (n=3 for each block). PCC = 0.9370 (95% confidence intervals = 0.8486 to 0.9745); P value (two-tailed) <0.0001. **d** Correlation of a normal tonsillar bulk RNA-Seq dataset (Human Protein Atlas) and RNAsky counts of a representative normal tonsil. PCC = 0.9329 (95% confidence intervals = 0.8392 to 0.9728); P value (two-tailed) <0.0001. **e** Correlation of a normal young adult tonsillar scRNA-Seq dataset and RNAsky counts of a representative normal tonsil. PCC = 0.8953 (95% confidence intervals = 0.7558 to 0.9571); P value (two-tailed) <0.0001. **f** False discovery rate (FDR) of RNAsky in all tissue (BC – breast cancer, NKSCC – non-keratinizing squamous cell carcinoma, CRC – colorectal cancer, Met. LN – metastatic lymph node) analyzed. Data represents mean ± SEM (n=3 for tonsil, BC, NKSCC, Met. LN; n=2 for CRC). **g** Average transcripts per cell in all tissues analyzed. Data represents mean ± SEM (n=3 for tonsil, BC, NKSCC, met. LN; n=2 for CRC; n = 20 for all tissues). **h** Median *in situ* sensitivity of RNAsky in three normal tonsils relative to a normal young adult tonsillar scRNA-Seq dataset. The line at 100% indicates the ratio where the sensitivity of scRNA-Seq and RNAsky are equal. **i** Subcellular resolution of RNAsky transcript detection in a representative NKSCC ROI. Colored dots indicate the genes in extended I-O panel and grey lines indicate cell segmentation. Image (scale bar = 200 µm; left) and close-up region (scale bar = 20 µm; right).

Fig. 2. Spatial RNA distribution in normal human tonsil.
RNAsky gene expression exhibits histological structure of the tonsil. Numbered boxes on top left image correspond to the three zoomed in images (top right and bottom row). Colored dots indicate location of detected transcripts and grey outlines show cellular segmentation. Scale bar = 200 µm; magnified region scale bar = 50 µm.

Fig. 3. Multiomic analyses by combining spatial information of RNA and protein expression in human tonsil.
a Display of RNA transcripts with respective proteins on the same section of a human tonsil. Box on left image is location of zoomed in image on right. Colors of protein immunostaining and detected RNAs (dots) indicated. DAPI counterstaining shown in blue. **b** Cell type annotation based on known gene relationships using stringent RNA gating strategy in human tonsil section. Cell types for all analysis strategies (b, c, d) shown below images. **c** Agglomerative hierarchical clustering using protein immunostaining intensity in human tonsil section. **d** Agglomerative hierarchical clustering using protein immunostaining intensity and RNA counts in human tonsil section. **e** Heatmap of selected protein and RNA intensities in spatially grouped, protein-based cell type clusters. The y-axis contains spatially-group clusters, thus there can be more than one cluster of the same cell type, with number of cells in each cluster. **f** Intensity-based transcript detection of IG genes, in a manner similar to antibody-stained tissues, allows for the identification of plasma cells expressing either *IGHG1* mRNA (red arrows), *IGKC* mRNA (green arrows), or both (yellow arrows). Scale bar = 200 µm; close-up region scale bar = 50 µm.

Fig. 4. Spatial RNA distribution in multiple different cancer types.
RNAsky on a representative **a** breast cancer **b** non-keratinizing squamous cell carcinoma and **c** colorectal cancer. Box on left image is location of zoomed in image on right. Colored dots indicate location of selected transcripts (legend below each image), and grey outlines show cellular segmentation. Scale bar = 200 µm; close-up region scale bar = 50 µm.

Fig. 5. Multiomic analyses across cancer samples.
a Clustered heat map of normalized gene expression of all genes in the extended I-O panel across all tissue types (BC – breast cancer, NKSCC – non-keratinizing squamous cell carcinoma, CRC – colorectal cancer, Met. LN – metastatic lymph node). **b** UMAP plots of all tissue types analyzed. Annotated cell populations labeled in bottom image. **c** Cross sample quantification of annotated cell populations. Data represents mean ± SEM (n=3 for BC, NKSCC, Met. LN; n=2 for tonsil, CRC). **d** Spatial reconstruction of identified cell populations based on the locations of transcription expression. **e** RNAsky transcript detection and immunostaining exhibit distinct features in colorectal cancer. Colors of detected proteins and RNAs (dots) indicated in legend below image. **f** Agglomerative hierarchical clustering using protein immunostaining intensity in colorectal cancer. Clusters and associated coloring in legend below image. **g** H&E performed on the same tissue section after combined MICS run is registered and combined with DAPI counterstaining (grayscale), then overlaid with RNAsky transcripts and protein immunostaining. Colors of detected proteins and RNAs (dots) indicated in legend below image. Scale bar = 200 µm.

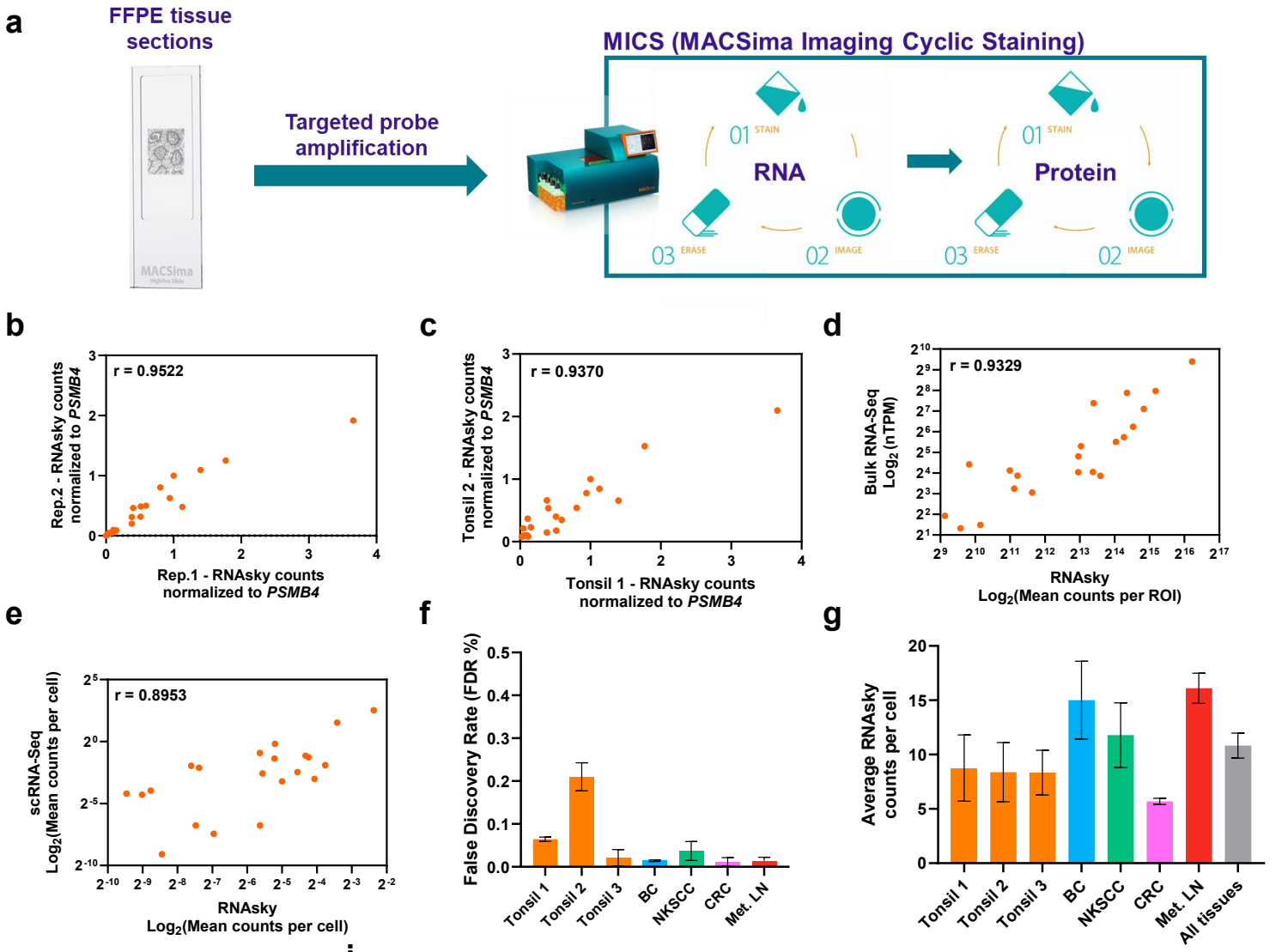

a
FFPE tissue sections
Targeted probe amplification
MICS (MACSima Imaging Cyclic Staining)
01 STAIN
RNA
02 IMAGE
03 ERASE
01 STAIN
Protein
02 IMAGE
03 ERASE
b
r = 0.9522
Rep.2 - RNAsky counts normalized to PSMB4
Rep.1 - RNAsky counts normalized to PSMB4
c
r = 0.9370
Tonsil 2 - RNAsky counts normalized to PSMB4
Tonsil 1 - RNAsky counts normalized to PSMB4
d
r = 0.9329
Bulk RNA-Seq Log2 (nTPM)
RNAsky Log2(Mean counts per ROI)
e
r = 0.8953
scRNA-Seq Log2(Mean counts per cell)
RNAsky Log2(Mean counts per cell)
f
False Discovery Rate (FDR %)
Tonsil 1
Tonsil 2
Tonsil 3
BC
NKSCC
CRC
Met. LN
g
Average RNAsky counts per cell
Tonsil 1
Tonsil 2
Tonsil 3
BC
NKSCC
CRC
Met. LN
All tissues
h
Median sensitivity relative to scRNA-Seq (%)
Tonsil 1
Tonsil 2
Tonsil 3
i

Figure 1

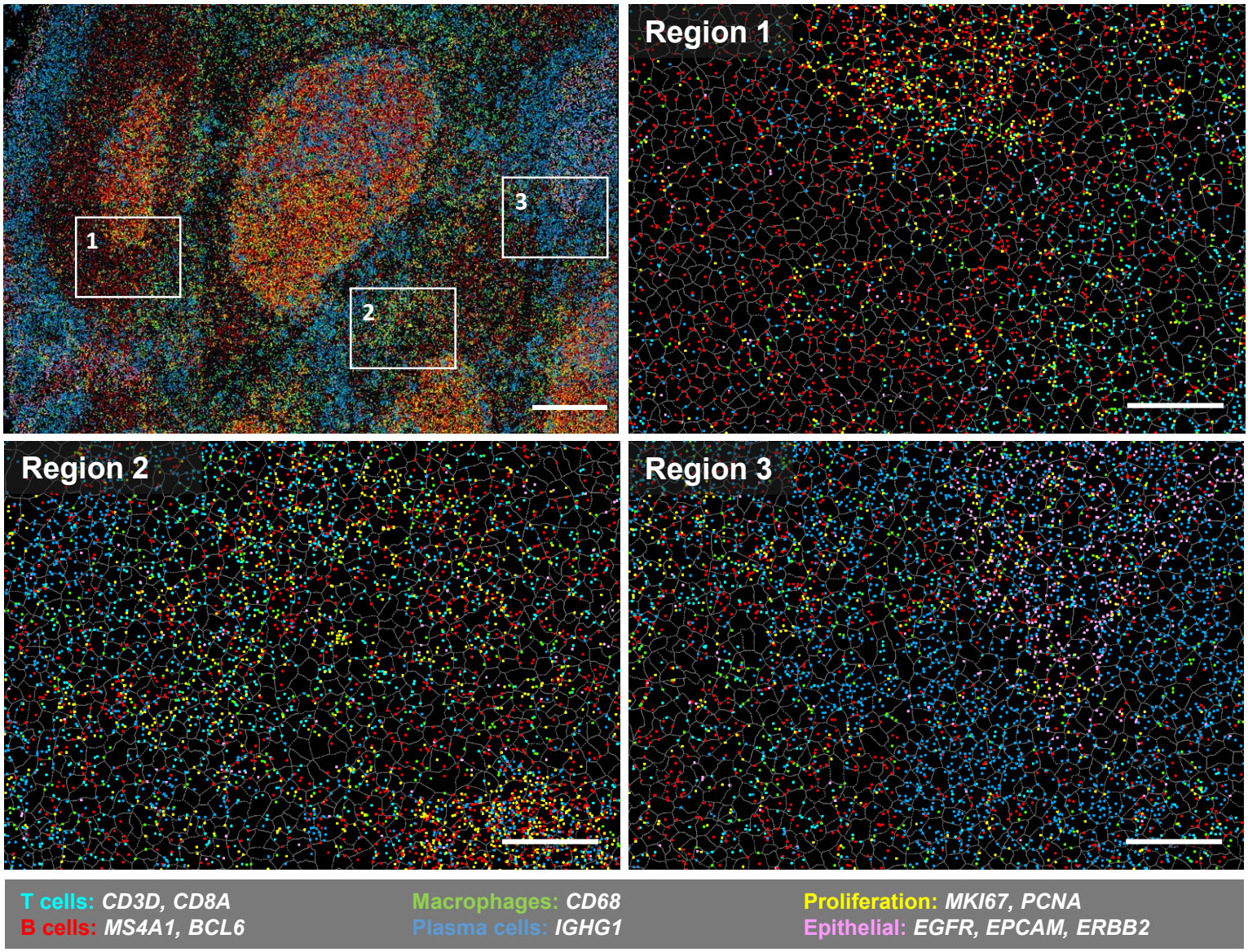

1
2
3
Region 1
Region 2
Region 3
T cells: CD3D, CD8A
B cells: MS4A1, BCL6
Macrophages: CD68
Plasma cells: IGHG1
Proliferation: MKI67, PCNA
Epithelial: EGFR, EPCAM, ERBB2

Figure 2

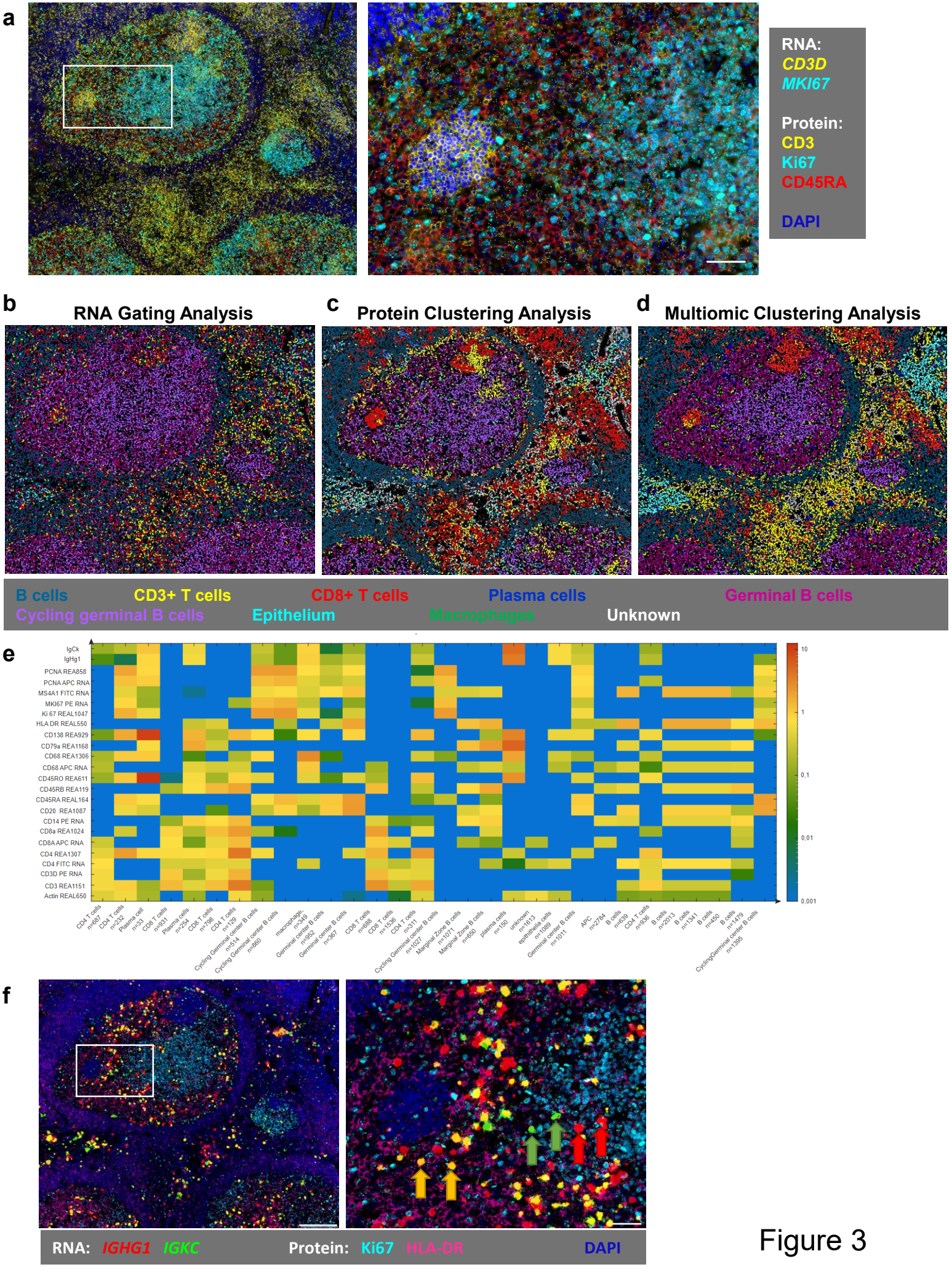

a
RNA:
CD3D
MKI67
Protein:
CD3
Ki67
CD45RA
DAPI
b
RNA Gating Analysis
c
Protein Clustering Analysis
d
Multiomic Clustering Analysis
B cells
CD3+ T cells
CD8+ T cells
Plasma cells
Germinal B cells
Cycling germinal B cells
Epithelium
Macrophages
Unknown
e
IgCk
IgHg1
PCNA REA858
PCNA APC RNA
MS4A1 FITC RNA
MKI67 PE RNA
Ki 67 REAL1047
HLA DR REAL550
CD138 REA929
CD79a REA1168
CD68 REA1306
CD68 APC RNA
CD45RO REA611
CD45RB REA119
CD45RA REAL164
CD20 REA1087
CD14 PE RNA
CD8a REA1024
CD8A APC RNA
CD4 REA1307
CD4 FITC RNA
CD3D PE RNA
CD3 REA1151
Actin REAL650
10
1
0,1
0,01
0,001
f
RNA: IGHG1 IGKC
Protein: Ki67 HLA-DR
DAPI

Figure 3

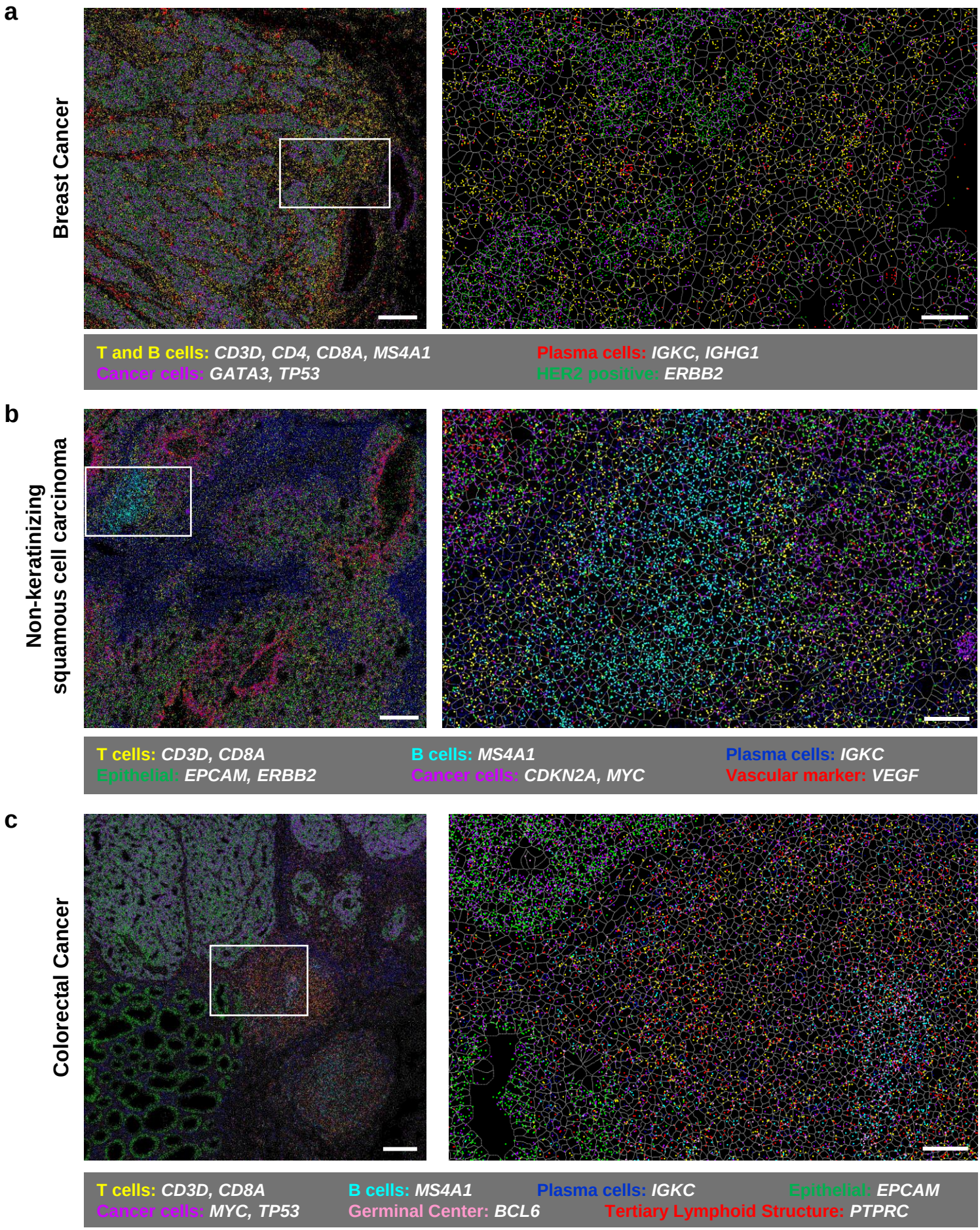
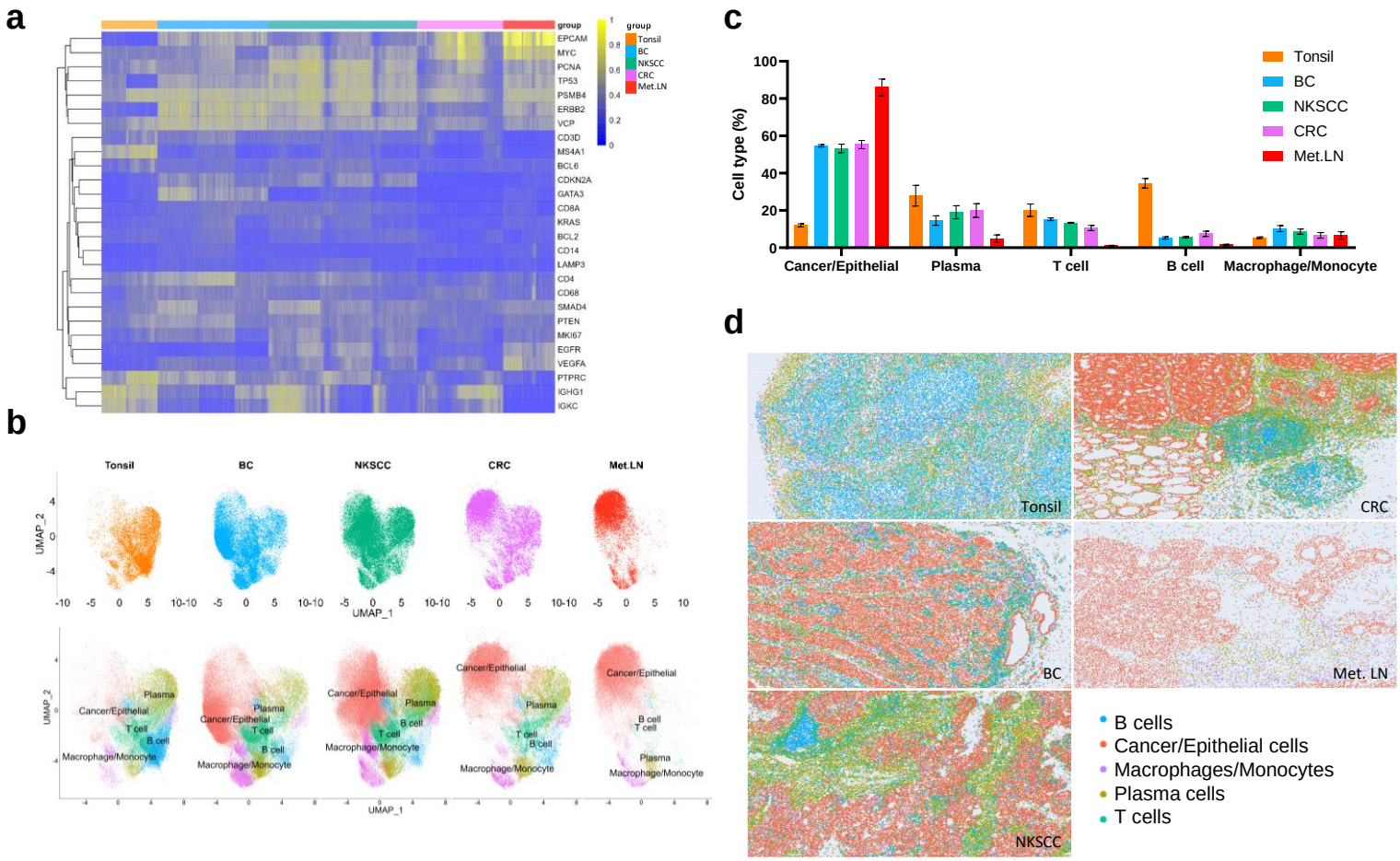

a
Breast Cancer
T and B cells: CD3D, CD4, CD8A, MS4A1
Plasma cells: IGKC, IGHG1
Cancer cells: GATA3, TP53
HER2 positive: ERBB2
b
Non-keratinizing squamous cell carcinoma
T cells: CD3D, CD8A
B cells: MS4A1
Plasma cells: IGKC
Epithelial: EPCAM, ERBB2
Cancer cells: CDKN2A, MYC
Vascular marker: VEGF
c
Colorectal Cancer
T cells: CD3D, CD8A
B cells: MS4A1
Plasma cells: IGKC
Epithelial: EPCAM
Cancer cells: MYC, TP53
Germinal Center: BCL6
Tertiary Lymphoid Structure: PTPRC

Figure 4


a
group
Tonsil
BC
NKSCC
CRC
Met.LN
EPCAM
MYC
PCNA
TP53
PSMB4
ERBB2
VCP
CD3D
MS4A1
BCL6
CDKN2A
GATA3
CD8A
KRAS
BCL2
CD14
LAMP3
CD4
CD68
SMAD4
PTEN
MKI67
EGFR
VEGFA
PTPRC
IGHG1
IGKC
b
Tonsil
BC
NKSCC
CRC
Met.LN
UMAP_1
UMAP_2
Cancer/Epithelial
Plasma
T cell
B cell
Macrophage/Monocyte
c
Cell type (%)
Cancer/Epithelial
Plasma
T cell
B cell
Macrophage/Monocyte
d
Tonsil
CRC
BC
Met. LN
NKSCC
B cells
Cancer/Epithelial cells
Macrophages/Monocytes
Plasma cells
T cells

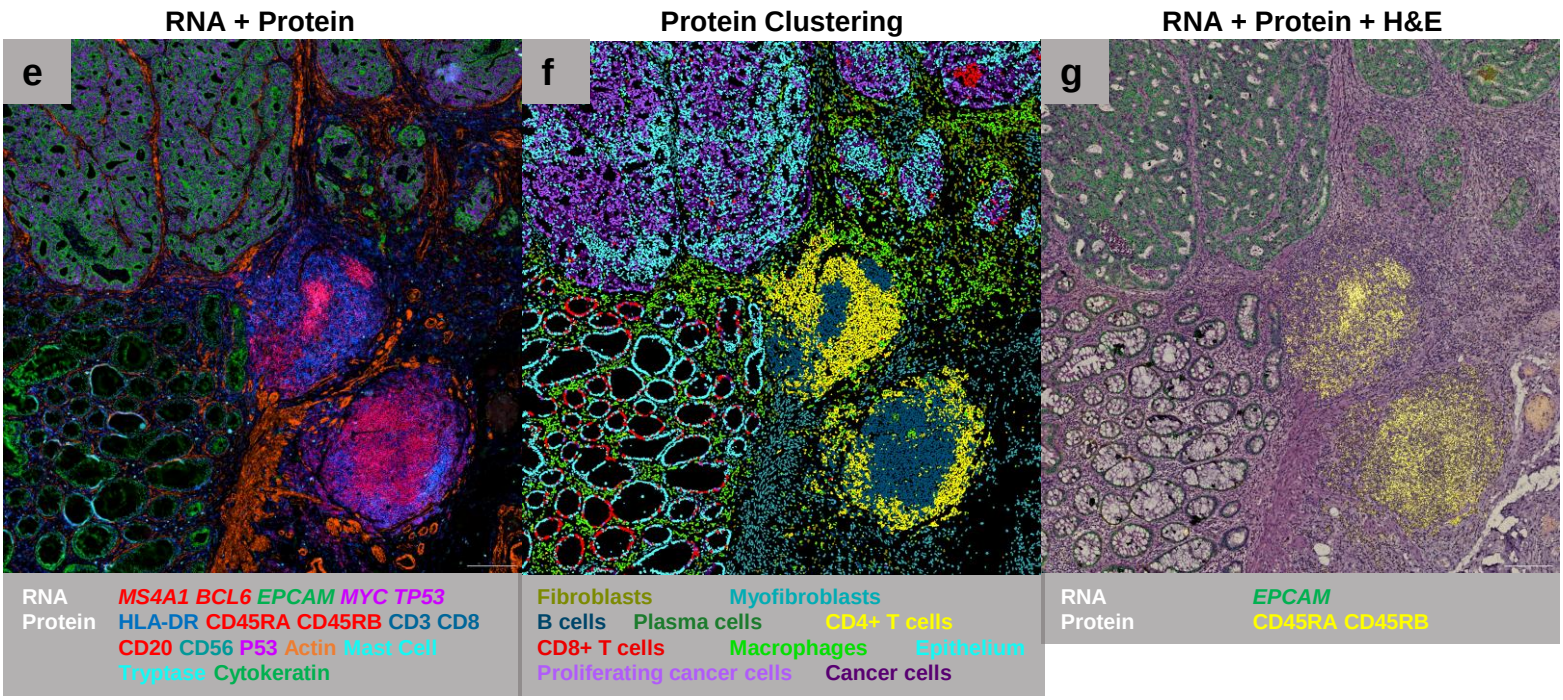

RNA + Protein
Protein Clustering
RNA + Protein + H&E
e
f
g
RNA MS4A1 BCL6 EPCAM MYC TP53
Protein HLA-DR CD45RA CD45RB CD3 CD8 CD20 CD56 P53 Actin Mast Cell Tryptase Cytokeratin
Fibroblasts Myofibroblasts
B cells Plasma cells CD4+ T cells
CD8+ T cells Macrophages Epithelium
Proliferating cancer cells Cancer cells
RNA EPCAM
Protein CD45RA CD45RB

Figure 5