Supplementary material

QuantISH: RNA in situ hybridization image analysis framework for quantifying cell type-specific target RNA expression and variability

Sanaz Jamalzadeh¹, Antti Häkkinen¹, Noora Andersson¹, Kaisa Huhtinen², Anna Laury¹, Sakari Hietanen³, Johanna Hynninen³, Jaana Oikkonen¹, Olli Carpén¹,²,⁴, Anni Virtanen⁴,⁵, Sampsa Hautaniemi¹,⁶
Supplementary methods

RNA in situ hybridization and imaging

Chromogenic RNA-ISH: We used freshly cut 3 μm sections of the TMAs and RNAscope 2.5 HD detection kit-BROWN (#322310, ACDBio, Milano, IT) for target mRNA detection. TMA sections were baked for 1 h at 56 °C, deparaffinized and treated with hydrogen peroxide for 10 min at room temperature. Target retrieval was performed for 15 min at 98 °C, followed by protease plus treatment for 15 min at 40 °C. Target probes Hs-DDIT3-C1 (#311131), Hs-CCNE1 (#470021), as well as the positive and negative control probes Hs-PPIB (#313901) and DapB (#310043), respectively, were hybridized for 2 h at 40 °C followed by signal amplification steps. The samples were incubated for 60 min with AMP 5 reagent. The sections were next treated with DAB for 10 min at room temperature followed by counterstaining with 50% hematoxylin. The sections were dipped in ammonium water and dehydrated before mounting. All TMAs were digitalized using 3DHISTECH Pannoramic 250 FLASH II digital slide scanner.

Fluorescent RNA-ISH: RNA in situ hybridization was performed on freshly cut 3 μm formalin-fixed paraffin embedded (FFPE) tissue sections using RNAscope Multiplex Fluorescent Reagent Kit Version 2 for target detection (#323100, Advanced Cell Diagnostics) according to the manual. In short, tissue sections were baked for 1 h at 60 °C, then deparaffinized and treated with hydrogen peroxide for 10 min at room temperature (RT). Target retrieval was performed for 15 min at 98 °C, followed by protease plus treatment for 15 min at 40 °C. All RNAscope probes (Table S1) were hybridized for 2 h at 40 °C followed by signal amplification, and developing of HRP channels was performed according to the manual. TSA Plus fluorophores fluorescein (1:750 dilution), Cyanine 3 (1:1500 dilution), and Cyanine 5 (1:3000 dilution) (NEL744001KT, Perkin Elmer) were used for signal detection. The sections were counterstained with DAPI, and mounted with ProLong Gold Antifade Mountant (P36930, Invitrogen). Images were generated using 3DHISTECH Pannoramic 250 FLASH II digital slide scanner at Genome Biology Unit supported by HiLIFE and the Faculty
of Medicine, University of Helsinki, and Biocenter Finland. All samples were scanned using 40x magnification with extended focus and 7 focus levels.

**Table S1.** The probe used in fluorescent RNA *in situ* imaging and its RNA scope catalog number.

| Probe name | Catalog number |
|------------|----------------|
| Hs-DDIT    | 311131         |

**Quantitative analysis of whole slide RNA *in situ* hybridization images**

*Channel separation:* We used CaseViewer (version 2.3.0) to read the MRXS immunofluorescence image, and to separate the different channels into the DAPI staining (nuclei), and fluorescein (FITC 38 HE), cyanine 3 (TRITC 48 HE), and cyanine 5 (Cy5) target RNA channels for gene expression quantification (Supplementary Figure 2).

*Tiling the DAPI images:* As the full whole slide images are too resource demanding for CellProfiler segmentation, the DAPI channel images were cropped into four tiles, each of which was segmented separately and assembled back into a whole slide segmentation for downstream analysis.

*Cell segmentation:* We used CellProfiler software (version 3.1.8) for segmenting the DAPI staining. The non-default parameters that were determined experimentally were as follows: typical diameter 18 to 56 pixels, thresholding using adaptive Otsu’s method (i.e. local not global), clumped object detection and splitting using shape (i.e. not intensity), and low-resolution speedups were disabled. An example is shown in Supplementary Figure 5A.

*Cell type classification:* The segmented objects were classified into cancer, immune, and stromal cells using the DAPI staining and its segmentation. For this, we extracted the area, the mean nucleus stain intensity, and the eccentricity of each segmented object (Supplementary Figure 5B).
These features were selected based on the same logic as for the chromogenic images. Subsequently, we trained a supervised quadratic classifier using 402 cells with the properties mentioned above and the desired cell types (Supplementary Figure 5C). Afterwards, the cell types were predicted in untrained images (Supplementary Figure 5D). In order to improve the classification accuracy, the classification involved one artifact class in addition to the cancer, immune, and stromal classes, which was trained to represent the small segmentation artifacts in classification. The classifier was implemented in MATLAB and was trained with uniform class priors. The classification was filtered using spatial information from the neighboring cells, as described for the chromogenic images in the main manuscript.

RNA signal quantification: Cross-channel fluorescence bleed of Cy5, FITC, and TRITC staining was reduced by finding a suitable basis for the intensity data of all pixels near the principal axes (channels) using power iteration. Note that this is done at the pixel level, not at the cellular level. Since some RNA signals are localized in the cytoplasm of the cells, we also expanded the segmentation corresponding to the cancer, immune, and stromal nuclei to include the cellular cytoplasm, as described for the chromogenic images in the main manuscript. Next, the fluorescence intensity signal was quantified using the negative response of a Laplacian of Gaussian filter with standard deviation of unity. The value was tuned manually, and the kernel width roughly corresponds to the diameter of an observed RNA spot in our images. This procedure filters out background variations and cellular autofluorescence, leaving intensity blobs of the specified size.

Variability factor formulation
The weighted mean expression and its weighted variance for each TMA (whole slide) spot/patient are computed as follows:
\[ m = \sum_i \left( \frac{\text{Area}_i}{\sum \text{Area}_i} \right) \left( \frac{\text{Intensity}_i}{\text{Area}_i} \right) = \frac{\sum \text{Intensity}_i}{\sum \text{Area}_i} \]

\[ v = \sum_i \left( \frac{\text{Area}_i}{\sum \text{Area}_i} \right) \left( \frac{\text{Intensity}_i}{\text{Area}_i} - m \right)^2 \]

where \( m \) is the mean expression in each spot/patient. \( \text{Area}_i \) is the area of each individual cell, and the \( \text{Intensity}_i \) represents the total (sum) intensity inside each cell. To compute the variability factor, we regressed out the mean from variability. First, we fit a linear regression model of expression on the raw variance in logarithmic spaces:

\[ \log v \sim \alpha_1 \log m + \alpha_0 \]

and obtained the estimated parameters \( \alpha_1, \alpha_0 \) of the linear regression. Following this, the expected log-variance due to the mean is:

\[ \hat{\log v} = \hat{\alpha}_1 \log m + \hat{\alpha}_0 \]

which is then regressed out from the variability measurements:

\[ \log v_{\text{factor}} := \log v - \hat{\log v} \]

that is:

\[ v_{\text{factor}} \sim v / m^{\hat{\alpha}_1} \]

which is a generalization of the Fano factor \((\hat{\alpha}_1 = 1)\), and also of the coefficient of variation \((\hat{\alpha}_1 = 2)\), and the variance \((\hat{\alpha}_1 = 0)\), with arbitrary power-law relationship between the variance and the mean. Once the relationship has been configured, the variability factor can be computed for individual samples, just like the mean and the variance.
Supplementary Figures

Supplementary Figure 1. Pre-processing of the TMA images for downstream analysis

A. A linear transformation is used to align the frames from the slide scanner and to remove the overlapping areas in the MRXS image tiles. B. For tissue microarray images, the TMA spots are annotated and cropped from the whole slide image for downstream analysis.

Supplementary Figure 2. Pre-processing of the whole slide images for downstream analysis

Channel separation of immunofluorescence whole slide images to the DAPI, FITC, TRITC, and Cy5 channels.

Supplementary Figure 3. Overview of cell type classification features and training set for TMA analysis

A. Feature extraction for the cell classification, including area, mean-intensity, and eccentricity for each object inside TMA. B. Scatter plot of the training data in the feature space and a decision boundary plot including the different cell type classes.

Supplementary Figure 4. Selection of the reliable TMAs for survival analysis using positive control images

Otsu’s thresholding method for two class classification of PPIB intensities resulted in separating low and high intensity groups in the PPIB intensity is used to select reliable images for quality control purposes.

Supplementary Figure 5. QuantISH image analysis pipeline steps from pre-processing to RNA quantification for whole slide images
A. Cell nucleus segmentation of the separated DAPI channel B. Extracting the segment features for training the cell type classifier. C. Decision boundary plot representing well-separated classifier features in the training data. D. Cell-type classification of a whole slide image.

Supplementary Figure 6. *DDIT3* average expression is uninformative of survival in HGSOC patients

*DDIT3* cancer cell average expression is not associated with patients survival across the 92 HGSOC patients in this cohort. This in contrast to the *CCNE1* average expression (cf. Figure 2C) or the *DDIT3* variability (cf. Figure 4C), both of which correlate with the platinum-free interval.

Supplementary Figure 7. *CCNE1* variability factor does not significantly change over three *CCNE1* expression groups

Variability factor does not significantly differ both between low and medium *CCNE1* expression \( (p=0.67) \) and medium and high *CCNE1* expression groups \( (p=0.95) \).
Figure S3

A.
- Area
- Mean Intensity
- Eccentricity

B.
- Figure with data points and distribution:
  - Log(Area) vs. Mean Intensity
  - Log(Area) vs. Log(Intensity)

Legend:
- Increasing eccentricity
Figure S5

A. 

B. 

C. 

D. 

Cancer cells

Immune cells

Stromal cells
Figure S7

- $p = 0.67$
- $p = 0.95$

Groups:
- low-exp
- medium-exp
- high-exp