Illumina based *ex-situ* whole-transcriptome workflow providing large area capture substrate, up to 7.5 cm² active area, and a scalable end-to-end software solution.

Kareem Ahmad, Connor Allen, Joel Sher, Reza Riahi, Erin Sarocco, Andrea Manzo, Sarah Nainar, Angela Nicholson-Shaw, Thu Nguyen, Shaobo Zhang, Asli Yildirim, Samuel Clamons, Sujai Chari, Chris Edlund, Allie Duchnak, Adam White, Tim Merkel, Jeff Fisher, Joachim Schmid, Niranjan Vissa, Bret Langham, Ben Moore, Hean Koon Koay, Darren Segale, Mark Wang, Cande Rogert, Emily Parker

[1]Core R&D, Illumina Inc

ABSTRACT

Spatial omics, by preserving the architecture of the tissue, enables new insight into the biological processes within their native microenvironments. A large capture area can allow for large tissues to be studied on a single substrate, without the need to break tissues up into multiple separate runs. However, the ability to study large tissues is challenging, as it requires a large capture area and the capability to process the correspondingly large amount of data.

Illumina spatial technology is built for scalability and efficiency by leveraging Illumina's best-in-class software ecosystem, enabling an end-to-end spatial analysis workflow. Powered by DRAGEN™ Illumina spatial solution provides a scalable system for researchers to analyze large datasets, with user-friendly interfaces for image processing, secondary analysis, spatial visualization and tertiary analysis. Combining with tissue imaging, our technology has integrated cell-based binning through nuclei identification and cell border expansion, yielding improved cell typing and marker gene identification. In addition, Illumina spatial technology gives the flexibility for users to export data for use with third-party spatial analysis tools.

To demonstrate this large capture area and end-to-end analysis capability, we present kidney on a single spatial substrate and processed the data with Illumina spatial software. We identified ~3 million cells and were able to process the entire dataset, including secondary and tertiary analysis, within 14 hours. Additionally, using 1 NovaSeq™ X 25B 200-cycle kit we detected a median of 1964 molecules and 991 genes per cell. Using 3 NovaSeq™ X 25B 200 cycle kits, we detected a median of 3100 molecules and 1365 genes per cell. Further, Illumina Connected Multiomics (ICM) provides accelerated discovery of biological insights via an intuitive interface and scalable tertiary analysis functionality. This shows the ability of Illumina spatial technology end-to-end solution to analyze large tissue samples and accelerate discoveries with scalable analysis software.

OVERVIEW

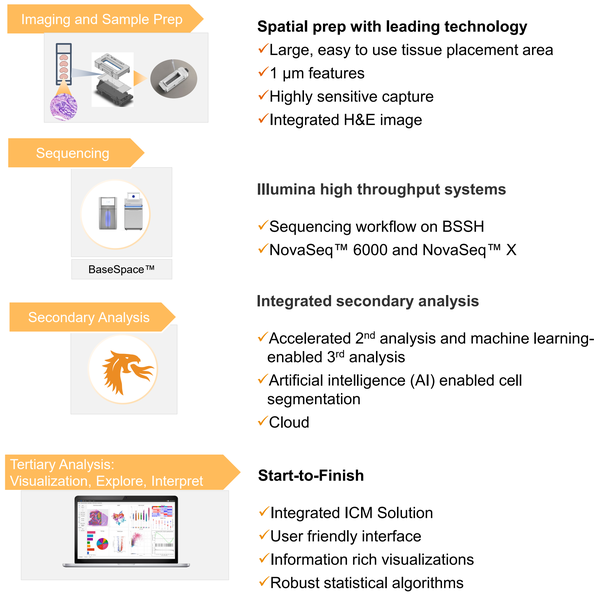


Figure 1. End-to-end workflow.

MATERIALS & METHODS

Large area spatial substrate

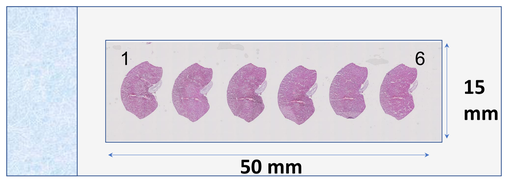


Figure 2. Six mounted kidney sections stained with Hematoxylin and Eosin.

- 50 x 15mm open placement active area has 750 million features encoded into it
- High density of address spots giving micron level resolution

Assay overview

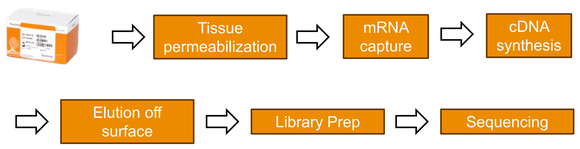


Figure 3. Assay workflow.

- *Ex-situ* spatial transcriptomics with polyA capture
- Compatible with a variety of Fresh Frozen tissues

Software overview

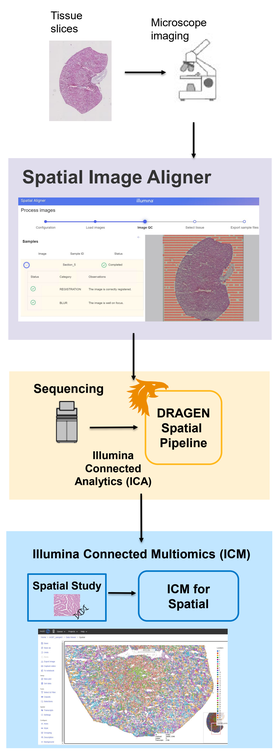


Figure 4. Software workflow.

Spatial Image Aligner

- **Image registration** enabling downstream alignment of transcript data with microscopy images
- **Image QC check** ensuring high quality images prior to tissue digestion
- **Automatic tissue border detection** to enable downstream tissue mask filtering

DRAGEN Spatial Pipeline

- Rapid RNA alignment using **DRAGEN**
- **Automated cell segmentation** using a custom trained machine learning model
- Pipeline outputs cell-binned transcript data in formats compatible with ICM and open-source tools for downstream analysis

Illumina Connected Multiomics (ICM)

- Users can easily view their spatial data overlaid on their microscopy images
- Quickly filter and/or make selection, in real time
- Perform additional tertiary analysis with a few clicks, such as differential expression, marker gene identification, cell typing

RESULTS

Whole substrate heatmap

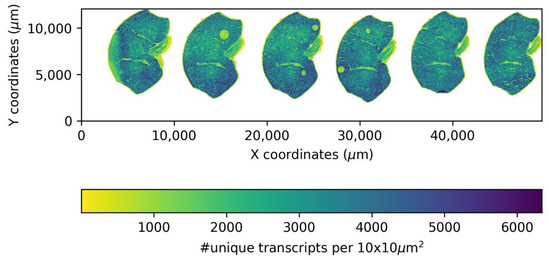
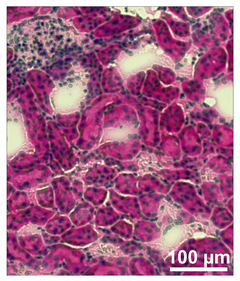


Figure 5. Whole substrate spatial heatmap. 6 slices of mouse kidney on a single substrate achieved > 3000 unique molecules per 10 µm x 10 µm bin.

Cell segmentation



Tissue slice number	Number of cells for each tissue slice
1	311,953
2	323,073
3	336,201
4	330,099
5	344,286
6	283,889

Figure 6. Example image cell segmentation result. Secondary analysis pipeline performs cell segmentation by nuclei identification and cell border expansion.

Analysis time

Runtime	Median UMI/cell	Median gene/cell
13.5 hours	1964	991
22 hours	3100	1365

- The pipeline automatically outputs cell-binned transcript counts for downstream viewing in ICM.

Sequencing saturation curves

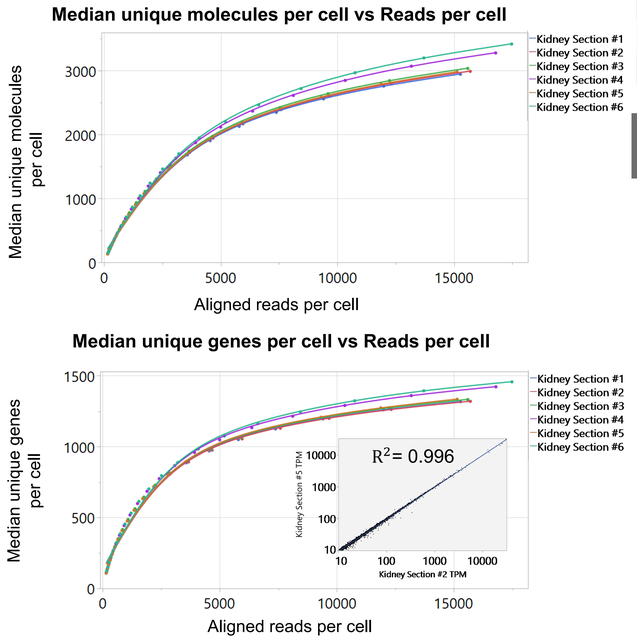


Figure 7. Sequencing saturation curves. Down sampled per cell unique molecule and gene counts plotted versus reads. Inset on graph is transcripts per million (TPM) for kidney section 2 vs section 5 showing high concordance between nearby tissue slices.

RESULTS (Continued)

Illumina Connected Multiomics (ICM) visualization

Easily view spatial data and perform customized analysis interactively. Investigate regions of interest to help answer biological questions.

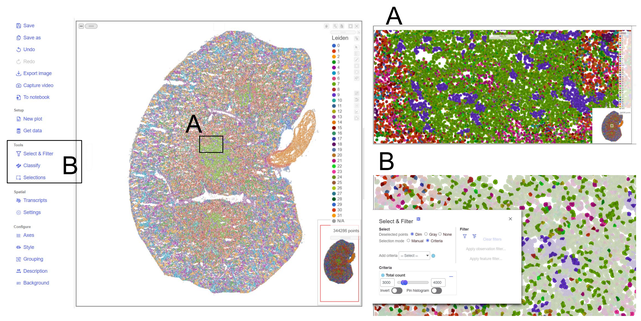
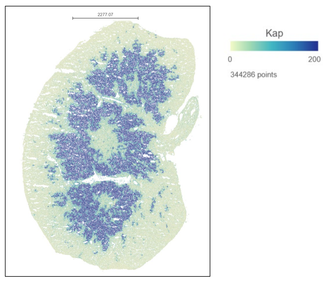


Figure 8. Visualization of tissue slice in ICM. A) Zoom in to areas of interest B) Filter on gene-specific counts, total counts or Leiden cluster

Figure 9. Visualization of cell counts for kidney marker gene (Kap) in ICM. ICM allows for quickly filtering to visualize gene specific counts of key marker genes.



Replicate experiments

Number of tissue slices	Total number of cells (millions)
5	1.5
6	1.9
6	2.1
6	2.5
7	2.6
7	2.8
Total	12

We have performed this experiment numerous times with varying number of mouse kidney slices on substrates and have achieved 1.5–3 million cells.

CONCLUSIONS

We used our spatial workflow offering high sensitivity, high resolution and broad unbiased coverage to generate spatial transcriptomics maps for a total of 12 million cells in 37 kidney sections across 6 experiments.

Secondary analysis, nuclear identification and cellular border expansion are all automated to output unique transcripts associated with each cell. On average, our approach identified 3100 UMI/cell and 1365 genes/cell.

Finally, cells and transcripts are visualized with Illumina Connected Multiomics for localization of cell types and marker genes.

The combination of high sensitivity and high spatial resolution coupled with an integrated analysis and visualization workflow make Illumina spatial technology a powerful tool for discovery applications.

SP-00652

For Research Use Only. Not for use in diagnostic procedures.

illumina®