Théo Lassale

GeoMx DSP: report

Context and goal of this report: In the project I have been working on during my time in Curie (MODICED project), we expected to receive spatial proteomics data generated from the GeoMx DSP platform developed by Nanostring Technologies. My role was then to subsequently process and analyse the resulting data, but due to technical issues, we were not able to retrieve these from the software. However, I thought it would be valuable to emphasise the upstream work I have done, to prepare for the data release. This includes an outline of the purpose of this platform and of its workflow **(1)**, a detailed description of the different profiling/segmentation methods **(2),** a review of the main steps of data processing and analysis **(3)**, an outline of multiple assets offered by the technology and of its main drawback **(4)**, an example of the application of the technology to discover biomarkers predictive of ICB response **(5)**, the inclusion of some links to scripts/tutorials for data processing and analysis **(6)**, and finally, the suggestion of additional papers focused on others applications of the technology in cancer and dementia research **(7)**.

1. Goal of the technology and workflow description

1.1. What is GeoMx designed for ?

GeoMx DSP is a platform, developed by Nanostring Technologies, which enables to generate the gene and protein expression profiles of bulk tissue samples, while keeping the spatial information.

1.2. How to retrieve quantitative profiling data from a single tissue biopsy ?

A workflow of 5 steps is required to obtain data. Firstly, you need to stain your single tissue section, by combining two types of reagents: **the imaging reagents** (up to 4 colours) used to image your tissue, and the **profiling reagents** used to profile it. Typically, **the imaging reagents** can be either fluorescent antibodies for protein detection, or RNAscope probes for mRNA detection. **The profiling reagents are** either ISH probes (for mRNA) or antibodies (for protein), which are tagged with oligos via a UV-photocleavable linker molecule.
Once the sample preparation is done, the slide will go onto the GeoMx system, to get scanned. The fluorescence of the resulting image will then be leveraged by the

user to select the areas of the tissue he/she wants to profile. Thereafter, on each selected area, a digital micromirror device will shine a UV light, that will release photocleavable linkers, and in turn, the barcodes. Those will then be aspired by a microcapillary tube , which will dispense them into a single well on a 96-well plate, where each well is indexed back to the XY coordinates of the region that was selected.
After the system went through all the regions selected on the tissue, the oligos from the plate will be either counted on the nCounter platform or sequenced on an Illumina sequencer.

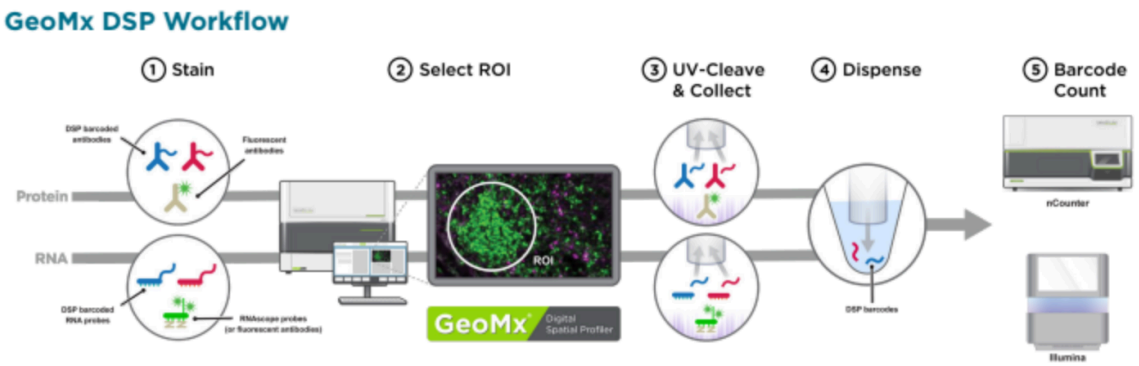

Figure 1. Schematic view of the GeoMx DSP workflow. The 5 sequential steps include: **1) Staining process**, **2) Segmentation method, 3) UV light exposure and collection of the released barcodes, 4) Store them in the 96-well plate, 5) Readout of the oligos on the nCounter system or on an Illumina sequencer.**

2. GeoMx DSP ROI Selection

What are the different strategies that the user of the platform can employ to select the regions he/she wants to profile ?

Several strategies are available for the user of the platform, to select the areas of the tissue he/she wants to profile. It is, in fact, the biological question you want to answer, that will inform your region selection. For example, if you want to assess the heterogeneity of expression in different regions of my tissue section, I will opt for **geometric profiling**, by placing geometric shapes (such as circles) around large areas of your tissue.
However, in the case, you only want to retrieve the expression profile of distinct tissue compartments, such as the cancer cells and its surrounding environment, you would need to further divide a region of interest of your tissue (i.e. a ROI) into individual segments (i.e. AOIs = for areas of interest), using the combination of the fluorescent markers (**Segment profiling**). Similarly, the latter would allow you as well to segment a specific cell population, if you aim to profile a specific cell type (**Cell-Type specific profiling**).

Furthermore, you may want to examine how the proximity to a histological landmark affects the biological response. To that end, you can draw concentric circles around this biological feature, known as **Contour profiling.**
Finally, you might choose to grid across your entire tissue section, to make a digital map of the expression profile of the tumour (**Gridded profiling**).

An illustration of these 5 profiling methods is shown below:

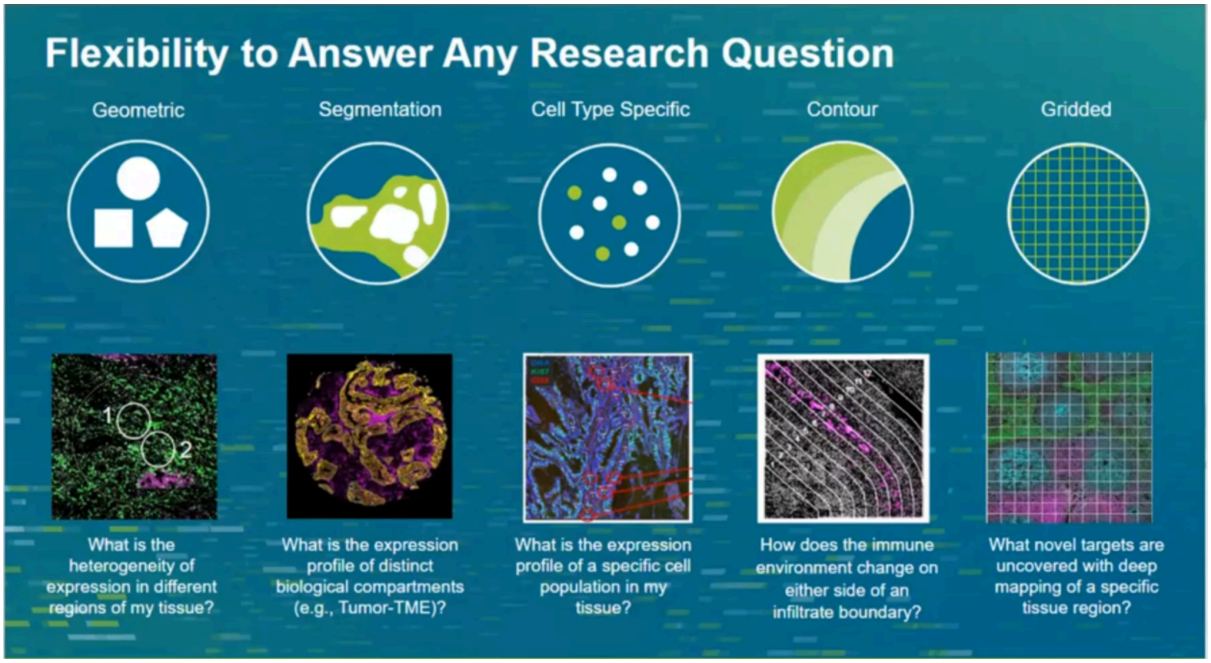

Figure 2. ROI/AOI selection: 5 different strategies. **1) Geometric profiling**: to compare gene/protein expression across broad areas of the tissue. 2) **Segment profiling**: to compare gene/ protein expression across distinct biological compartments of the tissue. **3) Cell Type Specific profiling:** to get the gene/protein expression profile of a specific cell type. **4) Contour profiling:** to evaluate how the proximity to a biological feature affects the biological response. **5) Gridded profiling:** to grid across the entire tissue section (expands on the **geometric profiling** method), especially useful for an unbiased analysis.

3. GeoMx Data Processing and Analysis

3.1. QC of segments and targets

After you retrieve the barcoded oligos from the plate, you have two options, you can

either choose to count them using nCounter platform, or you can decide to sequence them on an Illumina sequencer. In both cases, you will be able to retrieve a matrix of counts, for downstream analysis, using either a Data Analysis suite embedded to the GeoMx, or a programming language, such as R as an example.

But, regardless of the method you use, it is essential to do a quality control and normalisation of your raw data, to remove any technical variability and to focus on the true biological signal of interest.

As a first step, you need to filter out segments with low target detection (i.e. with poor data). To that end, you keep segments with a frequency (user-defined value) of targets above an expression threshold. This expression can either be defined by the user or can be regarded as the LOQ (i.e. signal to noise ratio).
After filtering out the poor quality segments, you are ready now to get rid of the targets with low detection across the remaining segments. Again, the standard rule consists in keeping targets with an expression above the expression threshold, at a specified frequency.

3.2. Normalisation

Once the quality control steps are done, you are now ready to normalise your data. There are 4 choices for normalisation: **housekeeper**, **background**, **region**, or **nuclei normalisation**. To choose the right normalisation strategy for your data, it is important to do a QC of the 4 potential normalisation factors (listed above). For more information about the details of the QC process, you can follow the link below: https://nanostring.com/wp-content/uploads/MK2593_GeoMx_Normalization-Protein.pdf

3.3. Data Exploration

Assuming that you performed the QC and normalisation steps, you can now start exploring your data. As an example, if you want to visualise the distribution of gene or protein expression across multiple tissue compartments (e.g. tumour vs stroma), you can generate some figures such as heatmaps, boxplots, or bar graphs. While if you are interested in discovering statistical changes in expression between experimental groups, you might opt for a volcano or PCA plot.

A few examples of figures, the user of the platform can make, using the GeoMx Data Analysis suite:

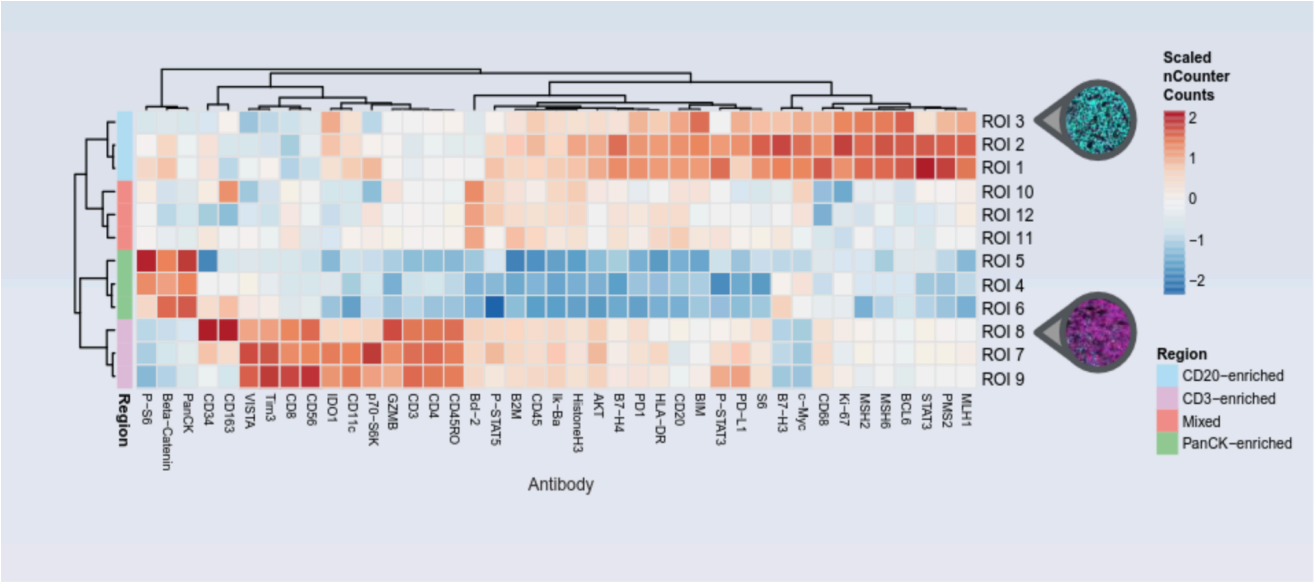

Figure 3. Heatmap showing the expression of 41 proteins across 12 ROIs. Four morphology markers were used for ROIs selection. The released oligo tags were readout on the nCounter system, and scaled across the different ROIs. Clusters of ROIs and proteins are noticeable.

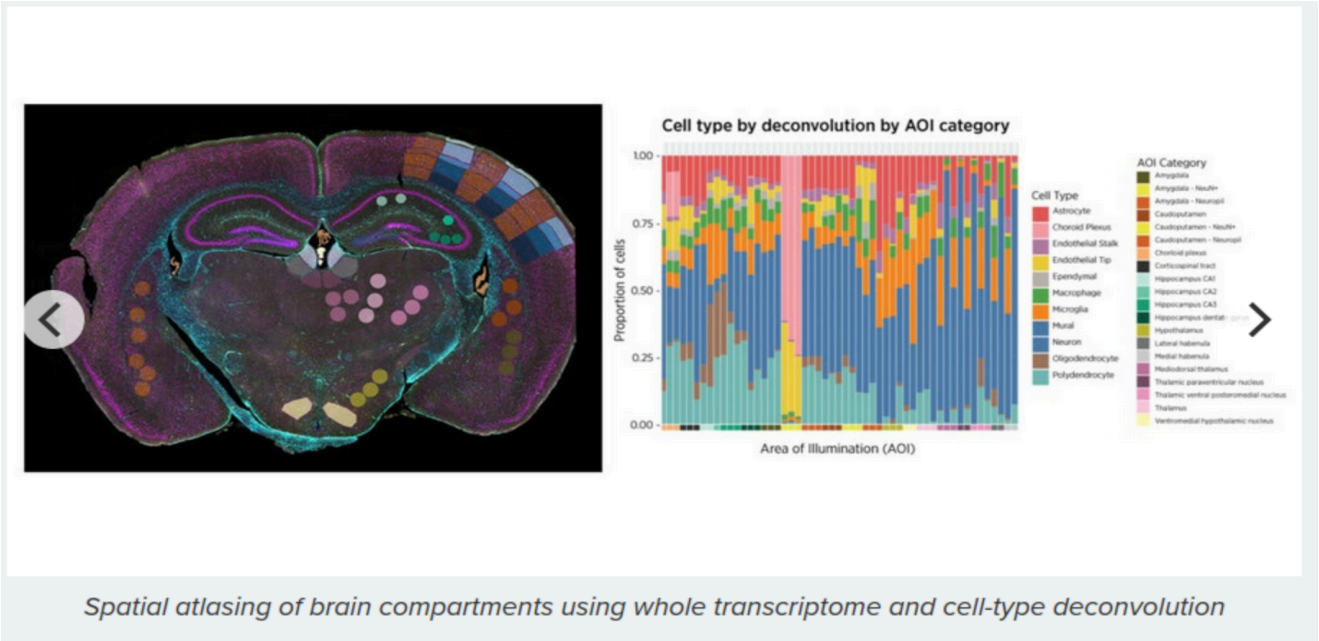

Figure 4. Stacked bar chart, to represent the proportion of different cell types across multiple brain compartments.

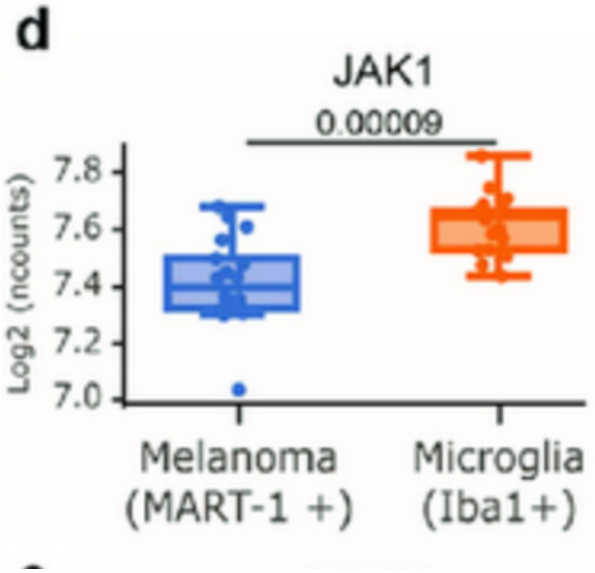

Figure 5. Boxplots showing the expression of the **Jak1** gene across MART-1 and Iba-1 positive ROIs (18 ROIs in total - 3 tissue slides). ***(Izraely et al., 2023)***.

3.4. Statistical Tests for Differential Expression

The choice you make for employing the right statistical test, depends on multiple factors: the number of ROIs to compare, the number of samples (i.e. of tissue sections), the distribution of your data.

Assuming that your data is normally distributed, and that you only have two ROIs to compare, you will choose a standard t-test, which will be paired (if the 2 ROIs come from the same slide), or unpaired (if they come from different tissue sections).

However, if the data is skewed, you can employ a Wilcoxon signed-rank test, if your two ROIs come from the same slide, or a Mann-Whitney U test if they come from different ones.

Finally, if you need to compare gene or protein expression across multiple (more than 2) ROIs, you can fit a linear mixed effect model to your data, with the ability to correct for batch effect (where your tissue is your batch variable).

An illustration of the differents tests you can use for differential expression, from the Data analysis suite panel, is shown below:

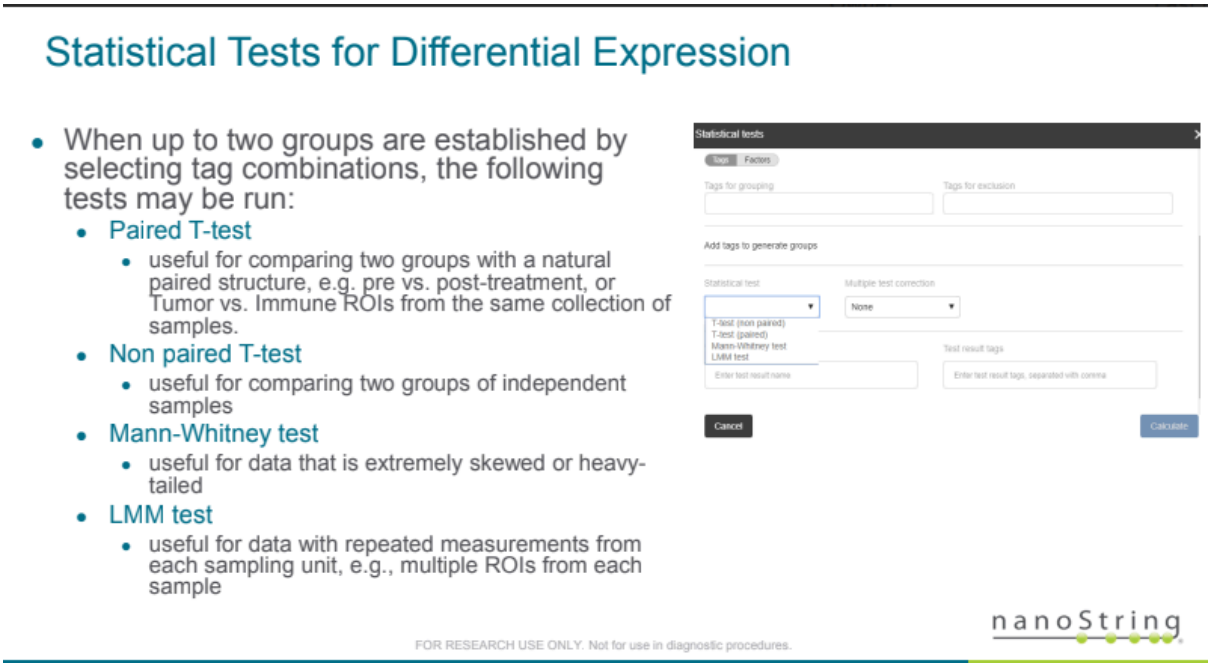

Figure 6. Statistical tests you have the choice to use, from the Data Analysis suite panel. This illustration guides you in the choice of the right statistical test to employ, if you are not very familiar with statistical testing.

4. GeoMx DSP: Assets and a major limitation

4.1. Assets

As a first main asset, the GeoMx DSP works by directly counting mRNA or protein molecules, eliminating bias in the expression data, as you might have with reverse transcription or PCR amplification in the sequencing process.

Secondly, you are able to spatially profile the whole transcriptome of human and mouse as well as more than 570 protein targets, separately or simultaneously, from your choice of inputs: whole tissue sections (FFPE or fresh frozen), tissue microarrays (TMAs), or organoids.

Furthermore, it enables you to process more than 10 sections per day (--> very fast).

One of the other major advantages of the GeoMx DSP, is the flexibility with which you can select regions of the tissue to profile, based on the biological question you want to answer.

In addition, the interface of the Data Analysis suite provides you with both the profiling and imaging data, and allows you to easily navigate between the counts and their tissue location. To that end, you only need to move your cursor to the heatmap (i.e. showing the distribution of counts), and this shows you straight where the counts come from on your tissue section (**Figures 7 and 8**).

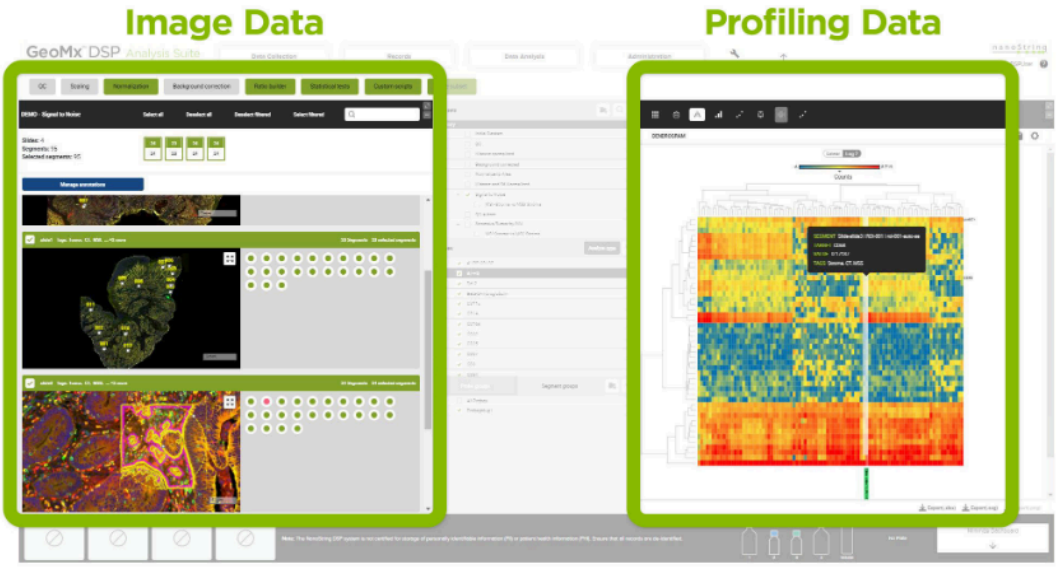

Figure 7. Interface of the GeoMx DSP Data analysis software showing the profiling data visualisation alongside the tissue slide image.

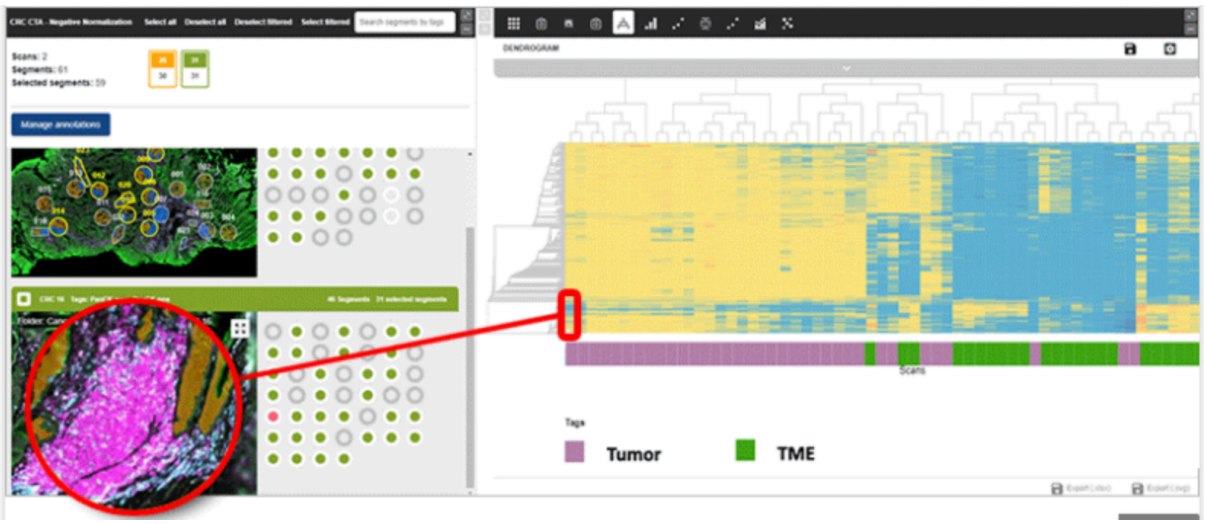

Figure 8. Direct interaction (in real time) between the counts and the ROI/AOI from which they derive.

Finally, the platform has many biological applications, you can use it, for example, to discover biomarkers predictive of therapeutic response, to profile tumour cells and its surrounding environment (i.e. stroma and immune cells), to uncover drug mechanisms of action, or to reveal new molecular subtypes of disease.

All the applications of the GeoMx DSP are illustrated below:

Figure 9. Biological applications of the GeoMx DSP.

4.2. Main limitation of the technology

Similarly to the Visium technology from 10X Genomics, the **GeoMx** is not conducive to single-cell analysis, although a ROI can be as small as 10 micrometres (i.e. the size of a single cell), you need at least 20 cells in a region you selected to get a reliable quantification of transcripts and proteins. And, in fact, to overcome this drawback, the **CosMx SMI** was designed, but its throughput is limited, since it enables quantification of up to 1000 mRNA molecules or up to 100 proteins.

5. Example of GeoMx DSP application: Biomarker discovery

Paper of choice: **Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma (Amaria et al., 2018)**.

Contextual Background: The current standard of care for patients with clinical stage III melanoma is upfront surgery with an adjuvant therapy. However, preclinical studies suggest that neoadjuvant ICB is associated with improved survival and enhanced anti-tumor immune response compared to the same treatment administered in the adjuvant setting. But, so far, we were unable to validate this observation, since there were no clinical studies done to assess the effect of the neoadjuvant ICB in melanoma. It is in this context that Rodabe Amaria and her team conducted the first randomised neoadjuvant trial in patients with stage III and IV melanoma. This trial included 23 patients, where 12 were randomised to nivolumab and the eleven others to nivolumab+ipilimumab.

Methods of interest: To better characterise the tumour molecular and immune microenvironment of these patients before and after treatment, they took tumour biopsies from them prior and after therapy, and passed them onto the GeoMx system to do the profiling.

Results: At first, they showed that the combination therapy was associated with improved survival over the monotherapy. But more importantly, in each treatment group, they noted that some patients responded better than others, and interestingly, when they compared the immune/tumour cell profiling between the responders and non-responders, they observed a greater expression of PD-L1 and of some immune markers (Granzyme B, CD4, FoxP3, CD20, and PD-1), as well as higher CD8+ T cell infiltrate, in responders over the non-responders, both at baseline and early-on treatment. These observations are illustrated in the bar plots below (**Figure 10**), where we can clearly observe a greater density of CD8+ T cells and a higher expression of PD-L1 in responders compared with the non-responders, both at baseline and after treatment. The heatmap in **Figure 11** shows that the signal of the additional immune markers: Granzyme B, CD4, FoxP3, CD20 and PD-1, is much higher in responders than in non-responders, both at baseline and after therapy, regardless of the treatment arm. Finally, the volcano plot in **Figure 12,** is in support of these findings, since all the immune markers quantified by the GeoMx are significantly more expressed in responders than in non-responders, prior and after combination therapy.

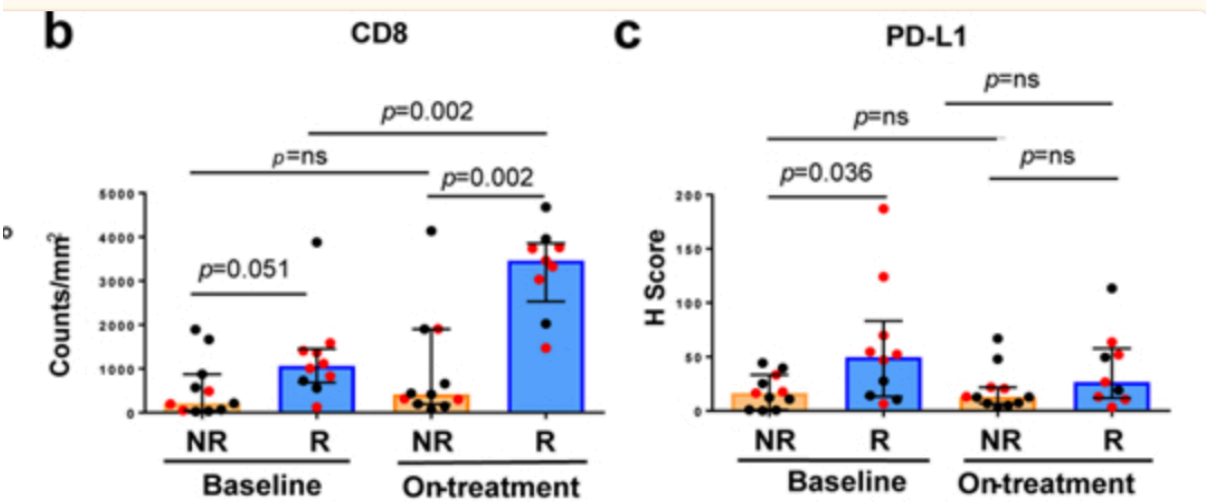

Figure 10. Quantification by immunohistochemistry of CD8 cell count density (n=11 NR and 10 R at baseline, n=11 NR and 9 R on-treatment) and PD-L1 H score (n=11 NR and 10 R at baseline, n=11 NR and 9 R on-treatment).

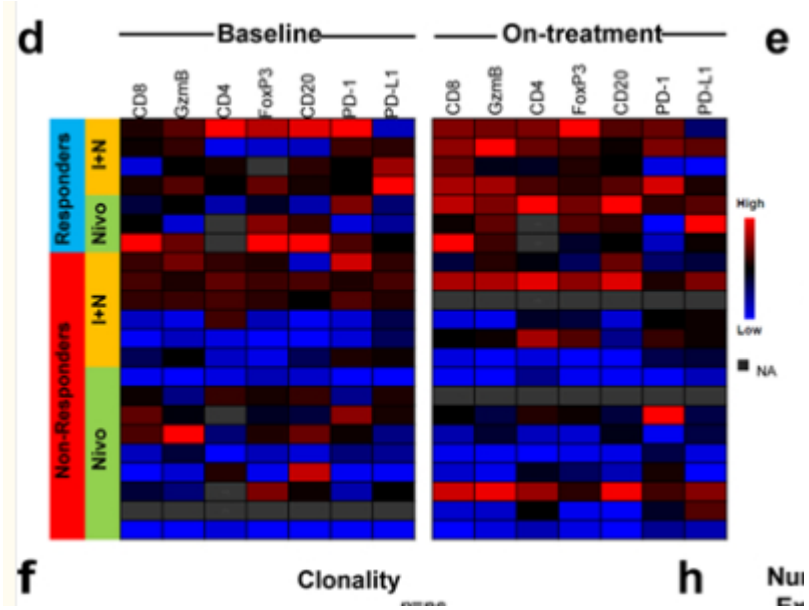

Figure 11. Supervised hierarchical clustering summarises expression of additional lymphoid markers (Granzyme B, CD4, FoxP3, CD20, PD-L1) using standard immunohistochemistry with blue indicating lower count density and red higher count density (or H-score for PD-L1).

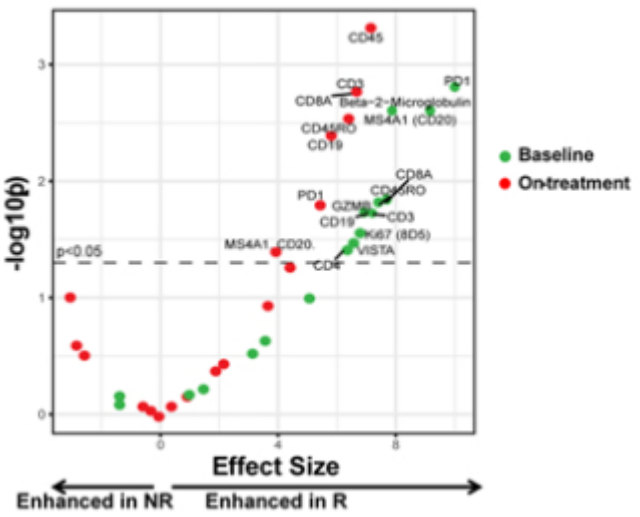

Figure 12. Volcano plots of two-sided pairwise Mann-Whitney comparisons of multiplex immunohistochemistry marker expression between NR and R in baseline (green) and on-treatment (red) samples in the combined treatment arms (n=11 NR and 11 R at baseline, n=10 NR and 6R on-treatment)**.**

Conclusions

This paper is a good example of the use of the GeoMx DSP for biomarker discovery. Through this study, they potentially found tumour molecular and immune markers that could be used to predict the success of the neoadjuvant immunotherapy in patients with high-risk resectable melanoma.

Limitations of the study

We need to be cautious in the interpretation of their findings, since this study has several limitations, including relatively small sample size (only 23 patients included in the trial), an early closure of the latter as 2 patients in the monotherapy group had disease progression preventing surgery and 8 patients that received the combination, had severe adverse events. Furthermore, the inclusion of patients with metastatic stage IV melanoma, makes the study population heterogeneous and therefore decreases the generalizability of the results to a stage III population. Finally, these data are limited to patients with the cutaneous melanoma subtype, so we do not know if we can really generalise these findings to other subtypes, such as mucosal or acral ones.

6. Supplementary materials for data processing and analysis

The following tutorial can be used for QC and normalisation (RNA + Protein): https://www.bioconductor.org/packages/release/bioc/vignettes/GeomxTools/inst/doc/Protein_in_GeomxTools.html

Additionally, you can retrieve the Rscript from the GeoMx hub, allowing you to QC the potential normalisation factors for protein and RNA expression data, by following this link: https://nanostring.com/products/geomx-digital-spatial-profiler/geoscript-hub/zzz#script-evaluate-normalization-for-protein

The following vignette also can be useful to have a better understanding of the QC and normalisation steps for DSP protein data: https://nanostring.com/wp-content/uploads/MK2593_GeoMx_Normalization-Protein.pdf

If you are provided with GeoMx RNA expression data, you can use the following tutorials, it guides you through all the steps of a comprehensive analysis:

1. https://bioconductor.org/packages/release/workflows/vignettes/GeoMxWorkflows/inst/doc/GeomxTools_RNA-NGS_Analysis.html#4_QC__Pre-processing
2. https://davislaboratory.github.io/GeoMXAnalysisWorkflow/articles/GeoMXAnalysisWorkflow.html

Finally, for other scripts of interest for analysis, when you are provided with the data, you can follow this link, that will take you straight to the GeoScript Hub:

https://nanostring.com/products/geomx-digital-spatial-profiler/geoscript-hub/

7. Examples of other papers of interest

Tumour heterogeneity (intra/inter)

- https://www.nature.com/articles/s41467-021-21615-4: ***Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling***
- https://pmc.ncbi.nlm.nih.gov/articles/PMC10158028/: ***Exploration of spatial heterogeneity of tumor microenvironment in nasopharyngeal carcinoma via transcriptional digital spatial profiling***
- *https://www.biorxiv.org/content/10.1101/2021.05.27.445912v1.full*: ***Digital spatial profiling of metastatic clear cell carcinoma reveals intra-tumor heterogeneity in epithelial-mesenchymal gradient***

Prognostic biomarkers

- **https://pmc.ncbi.nlm.nih.gov/articles/PMC9013045/**: ***Identification and prognostic analysis of biomarkers to predict the progression of pancreatic cancer patients***
- https://pubmed.ncbi.nlm.nih.gov/32313087/: ***Spatially-resolved quantification of proteins in triple negative breast cancers reveals differences in the immune microenvironment associated with prognosis***
- **https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-023-04395-x**: ***Identifying prognostic markers in spatially heterogeneous breast cancer microenvironment***

Application of the GeoMx DSP in dementia (example of Alzheimer's disease)

- **https://pubmed.ncbi.nlm.nih.gov/37777848/**: ***Spatial proteomics of hippocampal subfield-specific pathology in Alzheimer's disease and primary age-related tauopathy***