# Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** AI-based computational pathology applied to digitized patient biopsies

**Creator:**Cyril Puyo

**Principal Investigator:** Cyril Puyo

**Project Administrator:** Carlos de Andrea

**Affiliation:** Other

**Template:** DCC Template

**Project abstract:**

The advent of spatial proteomics technologies has unveiled a plethora of previously unexplored possibilities. By leveraging these technologies, we can now integrate the multidimensional network of relationships that govern the behavior of biological cells and tissues. This offers the potential for future advancements in delivering timely, efficient, and personalized diagnoses to patients.

In this study, we employ SPIAT (Spatial Image Analysis of Tissues), an R Bioconductor package specifically designed for single-cell spatial proteomics platforms. Our objective is to investigate the tumor microenvironment in advanced urothelial carcinomas, explore its correlation with treatment response to anti-PDL1 immune checkpoint blockade therapy, and evaluate its prognostic implications.

To achieve this, we analyzed 100 TIF images from formalin-fixed, paraffin-embedded (FFPE) tissue microarrays (TMAs) of the tumor microenvironment (TME) in urothelial cancer patients scheduled for treatment with Atezolizumab (Tecentriq). Before treatment, the samples were stained using the Opal tyramide signal amplification multiplexing method to label CD3, CD8, LAG-3, TIM-3, PD-1, and CK-positive cells. These TMAs were subsequently digitally processed using inFORM (by Akoya Biosciences) and QuPath (Bankhead, P. et al.). Finally, we utilized SPIAT to analyze the X- and Y-coordinates, as well as the mean marker intensity matrix, for in-depth investigation.

Spatial statistics have emerged as a promising tool in the diagnosis, prognosis, and classification of various cancers. However, the interplay between spatial statistics and immunofluorescence workflows requires further refinement to account for unexplained variability within analyses. This includes the annotation of cardinal directions during tumor sample extraction and the development of new statistical tools to manage multiple comparisons involving spatial data.

**ID:** 132088

**Start date:** 26-05-2023

**End date:** 22-09-2023

**Last modified:** 02-09-2023

**Grant number / URL:** https://www.unav.edu/documents/23943127/38459486/22-23_03_AI-based+computational+pathology+applied+to+digitized+patient+biopsies_C+de+Andrea+%26+J+Echeveste.pdf

# AI-based computational pathology applied to digitized patient biopsies

## Data Collection

We will create data based on formalin fixed tumor samples from urothelial carcinomas dating from 1950s. Employing segmentation algorithms, we will proceed to extract the marker intensities of numerous proteins contained withing the cells of said tumors.

The data will be collected in the form of an intensity matrix containing the unique name of the image, the ID of the patient they images comes from as well as spatial data such as XY coordinates for each cell and a matrix of immune marker intensities.

## Documentation and Metadata

Data were collected from formalin-fixed tumor samples derived from urothelial carcinomas. These samples underwent processing via an immunofluorescent multiplex assay, utilizing seven Opal dyes from Akoya Biosciences for the following markers: CK, PD1, TIM3, LAG3, CD3, CD8, and DAPI.

The images were captured using multispectral imaging systems, specifically the Vectra Polaris. These images were then reconstituted, taking into account autofluorescence and individual channel capture.

Lastly, the images were meticulously processed and annotated to identify regions of interest (ROI) within the tumor's central core. This was performed by expert pathologists using QuPath software. Subsequent segmentation was carried out using the DeepCell Kiosk ImageJ plugin integrated with QuPath, resulting in the generation of an intensity matrix.

## Ethics and Legal Compliance

In managing ethical issues, we have taken a multi-faceted approach to ensure the utmost integrity and security throughout the study. First and foremost, informed consent for both data preservation and tumor sample extraction has been duly obtained from all participants. This ensures that all individuals involved are aware of how their data and samples will be used. To protect the identity of participants, we employ a system of secondary IDs, which are only accessible to members of the study team. This adds an extra layer of anonymity, safeguarding against any unauthorized access to sensitive information. Additionally, data handling is strictly limited to personnel who have undergone security profiling and is confined to secure facility locations. This ensures that sensitive data is not only stored but also transferred in a secure environment, minimizing the risk of any data breaches or unauthorized access.

In addressing copyright and Intellectual Property Rights (IPR) issues, we adhere to a well-defined set of guidelines to ensure proper management and protection of the data. Ownership of the data is vested entirely in the Clinica Universidad de Navarra, and it is not available for publishing, teaching, marketing, or any form of reuse outside the specific scope of this master's thesis. This exclusivity ensures that the data remains secure and its integrity is maintained. Any data sharing or external use is strictly restricted and will only be considered within the context of the thesis reviewers. This approach allows us to maintain full control over the data, thereby safeguarding against unauthorized use or dissemination. Furthermore, we have clear policies in place regarding the reuse of third-party data, ensuring that we adhere to any restrictions imposed by the original data owners. This comprehensive strategy allows us to manage copyright and IPR issues effectively, while also fulfilling our ethical obligations.

## Storage and Backup

Data backup is systematically carried out by the principal investigator and stored securely in the cloud. To further enhance data integrity, version control is maintained through GitHub, where each update is duly commented for clarity and traceability.

Access to these backups and version-controlled data is strictly limited to the principal investigator and the project manager, ensuring that only authorized personnel can interact with the data. In the event of an incident, our backup and recovery protocols are designed to quickly restore the data, with the principal investigator and project manager being responsible for overseeing this process.

## Selection and Preservation

All data will be retained until one month after the project's conclusion, after which it will be reviewed for either preservation or destruction based on its relevance and potential for future research. This time frame allows for any necessary reviews, audits, or follow-up studies to be conducted, while also providing a window for data to be securely destroyed if it no longer serves a purpose.

The data for this project will be securely stored in GitHub, a well-known and trusted repository. Normally, GitHub would charge $4 per month for the storage services we require; however, due to the benefits of an alumni account, the cost for us will be zero.

## Data Sharing

The principal investigator and project manager have full access to the data, ensuring that they can oversee all aspects of the research and data management. For the duration of the review process, reviewers will be granted branch privileges, allowing them to access specific segments of the data relevant to their evaluation.

The data will be stored on GitHub, which serves as both a secure repository and a mechanism for handling data access requests.

To overcome or minimize restrictions on data access and sharing, we have implemented a tiered access system that allows for varying levels of data availability based on roles and responsibilities. This ensures that only authorized personnel have access to sensitive or critical data, thereby maintaining data integrity while still allowing for necessary collaboration and review.

## Responsibilities and Resources

The responsibility for implementing the Data Management Plan (DMP) falls primarily on the principal investigator, who will also ensure that the plan is regularly reviewed and revised as needed to adapt to any changes in the project scope or regulatory landscape. The principal investigator will work closely with the project manager to oversee each data management activity, from data collection and storage to backup and recovery.

No additional expertise or training for existing staff is required. The team is already well-equipped with the necessary skills and knowledge to manage the data effectively.