

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1S88W>

Title: EVALUATION OF MARKERS ASSOCIATED WITH PROTEOLYTIC PROCESSING IN BIOLOGICAL SAMPLES OF PATIENTS WITH MELANOMA

Creator: André Zelanis - **ORCID:** [0000-0003-3066-6527](https://orcid.org/0000-0003-3066-6527)

Affiliation: Universidade Federal de São Paulo (unifesp.br)

Principal Investigator: André Zelanis

Data Manager: André Zelanis

Project Administrator: André Zelanis

Funder: São Paulo Research Foundation (fapesp.br)

Funding opportunity number: 58508

Template: Digital Curation Centre

Project abstract:

Melanoma is an aggressive skin cancer and a lethal melanocytic neoplasm with increasing annual number of cases (faster than any other solid tumor). Since the protein composition (proteome) of tumoral cells may be regarded as a description of cellular status, the evaluation of protein expression patterns in biopsy samples may lead to the discovery of cancer-related markers and new drug targets. In the context of tumoral development, proteolytic processing plays an important role, favoring the dispersion of tumor cells to distant sites (metastasis) as well as in mediating irreversible proteolytic signaling events. The main goal of this project is to probe the Biobank of the São Paulo of Cancer Institute (ICESP) aiming at validating prognostic markers that

had already been identified in our previous research. Data obtained at this work will provide a systems-wide profile of proteolytic signaling in melanoma, with important translational potential, contributing to the understanding of the progression of melanoma.

Start date: 10-01-2020

End date: 09-02-2024

Last modified: 04-22-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

EVALUATION OF MARKERS ASSOCIATED WITH PROTEOLYTIC PROCESSING IN BIOLOGICAL SAMPLES OF PATIENTS WITH MELANOMA

Data Collection

What data will you collect or create?

Experimental approaches will generate 3 types of data:

- (1) Data regarding the proteome profile of isolated exosomes from human melanoma plasma samples;
- (2) Proteomics data derived from shotgun proteomics approach on formalin-fixed paraffin-embedded melanoma tissues;
- (3) Proteomics data from targeted analysis of plasma proteins derived from melanoma patients

Mass spectrometric data format will be made available under the original vendor (RAW) format as well as will be converted to the universal data format 'mzML'. Assembled contig data will be made available in '.fasta' format. Proteomics data will be submitted to public repositories such as the ProteomeXchange Consortium (<http://www.proteomexchange.org/>). In addition, for bioinformatic analysis purposes, data from the above repositories might also be interrogated.

How will the data be collected or created?

Biological samples will be submitted to standard proteomics protocols.

Briefly, Mass spectrometric (RAW) data will be analyzed within the Trans Proteomics Pipeline platform (v.4.8 ; Build 201411201551-6764) with Comet search engine (version 2014.02, rev. 2) against the UniProt/SwissProt database restricted to the taxonomy 'Homo sapiens' (latest release). Proteolytic signaling events (N-terminomic data) will be analyzed using WebPICS, a web-based platform freely available (<http://clipserve.clip.ubc.ca/pics/>) and the TopFIND knowledge base, a database and analysis resource for protein termini and protease processing (<http://clipserve.clip.ubc.ca/topfind>). Search results will be further filtered with PeptideProphet to a > 99% confidence interval, corresponding to a False Discovery Rate (FDR) of less than 1%. For both proteomics data, custom-made scripts will be designed and the analyses will be performed in R scripting and statistical environment.

Documentation and Metadata

What documentation and metadata will accompany the data?

Proteomics data, annotation files (eg. Microsoft Excel spreadsheets) will be provided together with the corresponding RAW files. In the case of protein identification by mass spectrometry, these annotation files are commonly derived from the output of searching algorithms and contain relevant information on the instrument used, measured error, accuracy, quantitative values (in the case of quantitative proteomics), the database used and so on (it depends on the searching algorithm used).

Ethics and Legal Compliance

How will you manage any ethical issues?

The experimental protocol that will be used in this project was already submitted to and have the approval of the National Research Ethics Commission (CEP/UNIFESP: 1477/2016).

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

There will be no restrictions on the reuse of third-party data.

Storage and Backup

How will the data be stored and backed up during the research?

In addition to the public repositories mentioned above, data will be stored locally in a server located at the Functional Proteomics Laboratory, at the Institute of Science and Technology of the Federal University of São Paulo, at São José dos Campos, Brazil. A backup will be done on a weekly basis, using external hard drives.

How will you manage access and security?

Data in public repositories are available to the general public. Ideally, our local storage will mirror the public repositories.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All data will be retained and used in throughout the development of this and other research projects at the laboratory.

What is the long-term preservation plan for the dataset?

At this time (December 2020), there is no budget limitation for time and effort to preserve data, as we have plenty of space on our server. However, as the project starts and data will come, it will be necessary a minimum of financial support for ordinary maintenance. However, as mentioned before, all data will be made public through public repositories of 'omics' data.

Data Sharing

How will you share the data?

All data will be made public through public repositories of 'omics' data.

Are any restrictions on data sharing required?

No

Responsibilities and Resources

Who will be responsible for data management?

As of today, all data handling in our server is managed by graduate students and myself. Additionally, IT personnel from the University are available for assisting in any issue.

What resources will you require to deliver your plan?

In fact, financial support is of paramount importance for the ordinary activities of data handling and hardware updating. The IT personnel of the University is constantly assisting us in any issue that may occur with our server or with data handling.
