

Plan Overview

A Data Management Plan created using DMPTool

Title: Systems biology analysis of multiomic data to develop predictive biomarkers of personalized (immuno-)therapy in multiple myeloma

Creator: Ariosto Silva

Affiliation: Moffitt Cancer Center (moffitt.org)

Principal Investigator: Kenneth Shain, Ariosto Silva

Data Manager: Ariosto Silva

Funder: United States Department of Defense (DOD) (defense.gov)

Funding opportunity number: HT9425-23-PRCRP-TTSA

Template: Department of Defense (DOD)

Project abstract:

Our primary objective is to investigate cancer evolutionary dynamics, to develop patient-specific strategies for overcoming therapy resistance. As cancer model, we have focused on multiple myeloma (MM)^{1,2}. MM is a cancer of bone marrow-resident plasma cells disproportionately affecting select groups of military personnel^{3,4}, including males, African Americans, as well as individuals with exposure agent orange, water contaminants at the Marine Corp Base Camp Lejeune (NC), or radiation-risk activities (Amchitka Island, AK and multiple diffusion plants in KY, OH, and TN). Unfortunately, MM remains an all but incurable. Even in the face of a growing armamentarium of agents, curative intent therapy fails, secondary to the evolution of therapy resistant disease over multiple lines of treatment^{2,5,6}. Although a heterogeneous array cytogenetic abnormalities, mutations, and epigenetic dysregulation have been linked to this cancer⁷⁻⁹, we still do not understand the mechanisms central to therapy resistance, in part due to this inter-patient heterogeneity⁷.

To address this issue, we have collected a unique cohort of primary samples from 890 MM patients across the disease spectrum. Analysis of whole exome and RNA sequencing revealed that aberrant transcriptomic programming is central to both malignant transformation and therapy resistance¹⁰. Importantly, these reprogramming events are conserved across different genetic backgrounds. Similar transcriptional pathway alterations were observed in sequential biopsies from individuals, even when cytogenetic and mutational traits remained unaltered. These findings suggested to us that transcriptional dysregulation as a unifying

biological feature in the evolution of therapy resistance -even in the heterogeneous genetic background of MM. In order to better elucidate the regulatory mechanisms of the aberrant gene expression, we developed a topology of the MM transcriptome. From these data, we discovered two major epigenetically regulated regions. A large gene supercluster associated with histone 3 lysine 27 trimethylation (H3K27me3) that was enriched in genes under-expressed during malignant transformation and a second supercluster of genes associated with H3K27 acetylation (H3K27ac) that were over-expressed in refractory disease.

To investigate the biology of specific drugs, we next utilized a subgroup of 300 MM samples with paired WES, RNA-seq and ex vivo drug sensitivity data. In so doing, we identified drug-specific sensitivity transcriptomic “footprints”, with both mechanistic and predictive applications. These findings inspired the development of novel evolutionary-based therapeutic strategies of sequential therapy, such as the combination of the immunotherapeutic daratumumab (DARA), and selective inhibitor of nuclear export selinexor (SELI), which we predict to double of PFS, which we confirmed in examining patients in SELI-based BOSTON clinical trial.

Collectively, our patient specimen-derived analysis indicated that the evolution of therapy resistance and response to therapy, is governed in great part by epigenetic dysregulation of critical biological pathways, which in turn are subject to targeting and translation to the clinic. Accordingly, we propose to investigate the hypothesis that development of patient/tumor-specific computational models integrating multiomics, clinical data and ex vivo patient avatars will identify the central biology associated with the development of therapy resistance and novel patient-specific management strategies based on predictive biomarkers.

Last modified: 08-08-2023

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

Systems biology analysis of multiomic data to develop predictive biomarkers of personalized (immuno-)therapy in multiple myeloma

Types of data produced

The types of data, software, curriculum materials, and other materials to be produced in the course of the project that are publicly releasable.

- a) Genomic Data: To be generated for the prospective cohort of MM patients:
 - RNA Sequencing Data: Bulk RNA sequencing data for 270 MM patients, in the form of tab-delimited text files (TPM, FPKM).
 - Whole Exome Sequencing Data: Whole exome sequencing data for 270MM patients in the form of text files (vcf and maf) .
 - scMultiomic (scRNA-seq/scATAC-seq) Data: Estimated 100 MM samples.
 - FISH/Cytogenetic Data: Cytogenetic abnormalities to be abstracted for all MM patient biopsies collected in this study.
- b) Clinical Data: De-identified and abstracted clinical data for all patient with biopsies collected for this study, including treatment history and response to therapy.
- c) Ex Vivo Imaging and Data: Image files (tiff format) from ex vivo experiments (brightfield and fluorescence) as well as post-image analysis dose response curves for the samples tested ex vivo against panel of drugs.

Data and metadata standards

The standards to be used for data and metadata format and content.

1. Bulk RNA-seq: text files with TPM/FPKM. This is the format through which we (Moffitt) receive transcriptomic data from ORIEN/AVATAR consortium. We will use the same format, to ensure compatibility with our pre-existing cohort of ~1,000 RNA-seq from MM biopsies.
2. WES: Vcf and maf file formats. VCF is the standard format we (Moffitt) receive WES data from ORIEN/AVATAR consortium, and we plan to use the same standard to keep compatibility with our pre-existing cohort of MM patients. As part of our pipeline, we convert these files to maf format, which is standard for most bioinformatics tools, so we will also share in this format.
3. scMultiomic: These data are generated by 10xGenomics platform, and will be shread in format .loupe or h5. The first is usable for 10xGenomics visualization tool, while the second is commonly used by single cell analysis platforms, such as Seurat.
4. FISH/cytogenetics: These data will be shared as spreadsheets.
5. Clinical data: These data will be shared in the format RedCap, which consists of de-identified spreadsheets exported by the tool Redcap, used to abstract clinical data.
6. Ex vivo imaging and ex vivo data: Images will be shared in their original tiff format, while dose response

curves will be shared as tab-delimited text files.

Conditions for access and sharing

Conditions for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

- **Ethics and Privacy:** This research relies on human subjects and abstracted electronic health records, thus only de-identified information will be shared. No germline genomic data will be provided, and all dates will be removed and replaced by age of patient at time of event. See Human Subjects section for more details.
- **Intellectual Property:** Moffitt Cancer Center reserves the right to any IP generated from this research, as well as ownership of the data for commercial use (i.e. non-research).
- **Storage and security:** All data used in this project will be de-identified, and linkage to actual patient data (e.g. acquisition of new patient data from EHR) will be conducted through an honest broker for Moffitt's NTRO department. All data will be stored in the following password protected databases: RedCap (EHR and ex vivo results), MySQL (unified relational database with molecular, ex vivo and de-identified EHR), LIMS (inventory of samples collected), SQLServer (pathology database with FISH/cytogenetics).

Conditions and provisions for reuse, redistribution, and derivatives

Conditions and provisions for reuse, redistribution, and the creation of derivative works.

Data use for others: The post-analysis data may be useful for researchers who plan to conduct a study in drug resistance in MM, as this will be part of the largest and more detailed dataset with clinical, molecular and ex vivo drug response data in multiple myeloma. Additionally, we believe that the predictive models generated will support further research in the translational field of personalized therapy in MM.

Data limitations for secondary use: While the data involve human subjects, only completely de-identified data will be available and used in the proposed study. Secondary data use is not expected to be limited, given the permission obtained to use the data from Moffitt Cancer Center, through the data use agreement with PIs.

Plans for archiving and preservation

Plans for archiving datasets, or data samples, and other digitally formatted scientific data, and for preservation of access thereto. Explicitly describe how the data that underlies scientific publications will be available for discovery, retrieval, and analysis. In accordance with OSTP Memorandum, digitally formatted scientific data resulting from unclassified, publicly releasable research supported wholly or in part by DoD funding should be stored and publicly accessible to search, retrieve, and analyze to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; and U.S. national, homeland, and economic security.

All experimental data generated (ex vivo, molecular, clinical) will be preserved in electronic format in RedCap

(description of ex vivo experiments), excel (actual ex vivo raw readings and dose response curves), ORIEN/AVATAR DNANexus platform (RNA-seq/WES), MySQL database in Moffitt Cancer Center's Azure/VM system (integrative database of matched ex vivo, clinical, molecular data).

R scripts, Matlab files, downstream analysis data (e.g. scMultiomic analysis), predictive models implemented in MATLAB, will be stored in Moffitt online shared folders allocated to the PIs of the proposal, with daily backup provided by Moffitt IT.

Non-digital data (e.g. western blots) will be digitalized (scanned/photographed) and stored in Moffitt LabArchives platform with no limit of expiration.

Justification for the restriction of data

If, for legitimate reasons, the data cannot be preserved and made available for public access, the plan will include a justification citing such reasons.

No justifications are made, as PIs plan to store all data generated in this proposal in digital format for future use.

Planned Research Outputs

Dataset - "Systems biology analysis of multiomic data to develop predictive biomarkers of personalized (immuno-)therapy in multiple myeloma"

This dataset will compile all the data generated during this project and referenced by publications.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Systems biology analysis of multiomic data to deve ...	Dataset	2028-01-07	Restricted	Synapse		Creative Commons Attribution 4.0 International	Synapse Controlled Vocabularies	No	No