Plan Overview

A Data Management Plan created using DMPTool

Title: Suppression of duplication-mediated genome rearrangements by protein sumoylation

Creator: Huilin Zhou

Affiliation: University of California, San Diego (ucsd.edu)

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-185

Grant: GM116897

Template: NIH-Default DMSP

Project abstract:

Genome rearrangements are mutations that cause numerous genetic diseases including cancer, immune deficiencies, and developmental disorders. Genome rearrangements are caused by defects in DNA replication and repair pathways, as well as the presence of numerous "at-risk" sequences in the human genome that are prone to mutations. Protein sumoylation is reversible and involves the opposing actions of two families of enzymes; the E3 ligases that catalyze the attachment of Small Ubiquitin-like MOdifier (SUMO) to substrates, and the SUMO specific proteases that remove them. Using the yeast Saccharomyces cerevisiae as a model organism, our prior studies have established a major role of these enzymes in preventing genome rearrangements and showed that site-specific sumoylation of the Mini-Chromosome Maintenance (MCM) complex plays a hitherto unknown role in regulating its loading at DNA replication origins. Remarkably, defects in this control cause impaired DNA replication, resulting in a drastic accumulation of gross chromosomal rearrangements (GCRs). Because inherited mutations of the SUMO pathway genes cause genome instability syndromes in mammals, our study will impact human health for two major reasons: 1) a comprehensive understanding of the genetic consequences that arise from mutations to the SUMO pathways, with regards to DNA replication, will impact the development of assays for cancer diagnosis. 2) Identifying the mechanism by which SUMO regulates DNA replication will lead to new therapeutic interventions of human diseases. Our proposed studies will pursue three specific aims. First, we will perform genetic analysis of the MCM subunits to examine their roles in maintaining cell growth and suppressing GCRs. Second, we will investigate the function of MCM sumoylation during DNA replication, focusing on its role in MCM loading and formation of the replicative DNA helicase. Third, we will perform biochemical reconstitutions of SUMO dependent MCM loading to understand its mechanisms, using cell-free assays that we have developed.
 Altogether, the goal is to understand how SUMO modification regulates MCM loading and how cells utilize this new pathway to prevent genome rearrangements.

Start date: 04-01-2024

End date: 03-31-2028

Last modified: 05-30-2023

Copyright information:

Suppression of duplication-mediated genome rearrangements by protein sumoylation

In this proposed project, data will be generated via the following methods: real-time quantitative polymerase chain reaction (PCR), and mass spectrometry (MS). This data will be collected from a minimum of 3 independent experiments, with each independent experiment consisting of 3 groups. The total size of the data collected is projected to be 100 GB.

Specifically, we expect to generate MS data about peptide sequencing and quantification. The raw MS data are analyzed by open source softwares commonly used in the proteomics community, including TPP and Mascot. We expect to submit both the raw MS data and processed data, which include protein identity and quantification, to public repositories once the research is published. Real time PCR data are analyzed by Excel.

In this proposed project, the cleaned, item-level spreadsheet data for all variables will be shared openly, along with example quantifications and transformations from initial raw data. Final files used to generate specific analyses to answer the Specific Aims and related results will also be shared. The rationale for sharing only cleaned data is to foster ease of data reuse. All our published data tare preserved and shared with the research community via publishers. Our raw and processed proteomics data will be submitted to public repositories with identifier reported in our published work.

We do not expect to generate any Metadata in our work.

Our proteomics data are analyzed by open source softwares including TPP and Mascot, which are public available. All other softwares used in our work are accessible via subscription fees. For instance, we pay annual subscription to access Lasergene to analyze our DNA sequencing data. All softwares used in our study are described in our published work.

We do not expect to generate any imaging data.

In accordance with FAIR Principles for data, we will use open file formats (e.g. JPEG, PDF, HTML, etc.) and persistent unique identifiers (PIDs) such as RRIDs for resources (e.g., organisms, plasmids, antibodies, software tools, and databases). We calculate standard deviation and obtain calibration curve for multiple measurements to evaluate the significance of our measurements, including quantitative PCR and quantitative MS results. The statistic findings (p-value, etc.) are reported in our published results.

Our data will be made available immediately following the publication of our work through journals and public repositories in case MS data is used for publication, which will be described in the method section of our work.

Repository used in our study will provide identifiers for our proteomics dataset, which will be described in the method section of our published work. Data will be discoverable online through standard web search. The other data will be available from the journals and papers that we publish.

All scientific data generated from this project will be made available as soon as possible, and no later than the time of publication or the end of the funding period, whichever comes first. The duration of preservation and sharing of

the data will be a minimum of 10 years after the funding period.

There are no anticipated factors or limitations that will affect the access, distribution or reuse of the scientific data generated by the proposal.

Controlled access will not be used. The data that is shared will be shared by unrestricted download.

Not applicable to our study.

Lead PI __Huilin Zhou__, ORCID: _0000-0002-1350-4430_, will be responsible for the day-to-day oversight of lab/team data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI and Co-I team as part of general [campus(es)] stewardship, reporting, and compliance processes.