Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D14354

Title: Requirements for ZNF292 in cortical development and mechanisms of pathogenic mutation in neurodevelopmental disorders

Creator: Kristen Kroll - ORCID: 0000-0002-5450-6694

Affiliation: Washington University in St. Louis (wustl.edu)

Principal Investigator: Kristen Kroll, Susan Maloney

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-185

Template: NIH-Default DMSP

Project abstract:

Recent advances in human genetics have defined hundreds of causal variants for Autism Spectrum Disorder (ASD) and other neurodevelopmental disorders (NDDs). However, substantial effort is now required to define downstream disease processes and guide development of interventions. One such new NDD gene is ZNF292: while mutations have recently become associated with ASD and other NDDs in humans, the role of ZNF292 in brain development and circuit function has not been studied in any animal or human cellular model. Here, we propose a comprehensive mechanistic investigation of ZNF292. This project uses both established cutting-edge workflows and innovative new approaches to enable in-depth study of ZNF292 mutation and deficiency at the molecular, cellular, structural, and behavioral circuit levels. We use two complementary experimental systems, mouse models and human pluripotent stem cell (hPSC)-derived neurons, to define requirements for ZNF292 in brain development and identify the consequences and reversibility of ZNF292 deficiency. We focus initially on hPSC models carrying six pathogenic ZNF292 variants identified in patients with ASD, intellectual disability, and other NDD clinical phenotypes (e.g. microcephaly, epilepsy), as well as hPSC lines with constitutive or inducible ZNF292 deficiency. For this work, we derived isogenic pairs of control and ZNF292 variant hPSC models, either by knock-in of pathogenic ZNF292 variants into control lines and by deriving hPSC models from subjects with ZNF292 mutation, with versus without ZNF292 variant correction by genome engineering. We also developed conditional and non-conditional mouse knockout models of ZNF292 deficiency in the developing brain, to
study consequences of ZNF292 deficiency on brain development, structure, physiology, and behavioral circuit function in an intact animal. Further, cutting-edge gene therapy-like tools developed for both human cellular and mouse models enable us to investigate effects of rescuing ZNF292 gene function during brain development or in mature neurons. Related landmark experiments profoundly changed the understanding of other NDDs by demonstrating that a substantial proportion of chronic NDD phenotypes were reversible, thus spurring development of therapeutics based on rescuing gene expression or modulating altered neuronal function. In complementary efforts, we also continue ongoing characterization of this rapidly expanding patient population by assessing natural history and brain structural changes in these individuals. Based upon our preliminary data in the experimental models above, we hypothesize that both altered brain development and chronic disruption of neuronal function contribute to patient phenotypes, with some chronic consequences such as impaired synapse formation and function contributing substantially to NDDs in this patient population. These deficits could be tractable for molecular or pharmacological treatment to develop interventions. Together, the experiments performed here will elucidate requirements for and mechanisms by which ZNF292 normally controls brain development and function, will determine how these are disrupted by pathogenic ZNF292 mutation, and could also chart a course towards ZNF292-targeted therapies.

Start date: 01-01-2024

End date: 12-31-2029

Last modified: 12-22-2023

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.
Requirements for ZNF292 in cortical development and mechanisms of pathogenic mutation in neurodevelopmental disorders

Data Type

Types and amount of scientific data expected to be generated in the project: Summarize the types and estimated amount of scientific data expected to be generated in the project.

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

In this proposed project, data will be generated via the following methods in human pluripotent stem cell (hPSC) models of ZNF292 pathogenic mutation or deficiency: hPSC culture, neuronal differentiation, light and confocal microscopy, a range of cellular phenotyping approaches (e.g. quantification of neuronal morphology, neurite extension, neuronal migration, and functional assessment by electrophysiology), and molecular methods (RT-qPCR, DNA and RNA preparation). In parallel, data will be gathered from mouse models with constitutive or conditional ZNF292 knockout with various assessments (neuropathology, brain imaging, neurophysiology, and behavioral assessment).

Data will also be generated in a mouse model of ZNF292 deficiency by using the following methods: behavioral analysis (a range of assessments related to IDD phenotypes), neuropathology, neurophysiology, evoked seizure monitoring, and small animal brain MRI imaging.

Molecular datasets (e.g. transcriptomic, epigenomic, and genome-wide occupancy of transcription factors and chromatin proteins) will be generated by using RNA-seq, CUT&Tag, or CUT&RUN. These data types will be collected from mouse cortex from 2 or 3 genotypes (wild-type, and heterozygous and, at E14, also homozygous knockout) at two timepoints (E14, adult) across four biological replicate samples (~24 samples per data type). They will also be collected from 7 hPSC models carrying deficiency or pathogenic mutation for ZNF292 (14 models with paired isogenic controls) across two time points (progenitor, neuron) and across four biological replicates (~112 samples). The total size of the data collected is 500 GB.

We expect to generate the following data file types and formats during this project: microscopic image files (.CZI), image (.TIFF) or (.AI) files, tabular (.CSV), and plain text files (.TXT). Raw data files will be analyzed to generate CSV files containing processed data (e.g., total number of neurons expressing a particular molecular marker or showing differentially expressed genes) and to enable statistical analysis. For behavioral data, tabular (.CSV, .XLSX) files will be generated and used for analysis, up to <1MB per dataset. For MR image sets, we will generate ~360 Neuroimaging Informatics Technology Initiative open file format (.NII) files at 14.4MB per animal for an approximate total of 4.8 GB.

As an alternate subaim or future direction, the Washington University IDDRC's Clinical and Translational core will use the Brain Gene Registry and Biomaterials Acquisition program to collect information about patients
carrying ZNF292 mutations, including conducting natural history studies of their clinical presentation and neuropsychiatric diagnoses and collection of blood from some patients. These data and biomaterials may be provided to the investigators in a deidentified form for this study and used to construct hPSC models for phenotyping as described above.

Scientific data that will be preserved and shared, and the rationale for doing so: Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

In this proposed project, processed spreadsheet data for all variables tested will be shared openly, along with example quantifications and transformations from initial raw data; as requested by the journal where this work is published, raw values may also be provided. Final files used to generate specific analyses to answer the Specific Aims and related results will also be shared through inclusion in manuscript main or supplemental data, while large raw and processed transcriptomic and epigenomic datasets will be shared by deposition in the Gene Expression Omnibus. Scientific data will also be preserved on the WUSTL Box cloud server, which is in compliance with data retention guidelines.

For a planned future direction involving acquisition of deidentified clinical phenotype data from the Washington University Intellectual and Developmental Disability Research Center (IDDRC’s) Clinical Translational Core, we will share de-identified individual-participant level (IPD) data, which is deidentified and assigned a Global Unique Identifier (e.g. via the NIMH Data Archive). This work has been designated as exempt from human subjects research according to our institutional HRPO.

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

The main and supporting data files will be shared via scientific publication and will provide all method descriptions, equipment, and settings, information about resources such as cell lines, model organisms, plasmids, behavioral protocols, and other tools (e.g. software, databases, or services used) and the data described above.

All large datasets (e.g. transcriptomic and epigenomic data) generated in the course of this work will be deposited into the Gene Expression Omnibus (GEO) repository. The Gene Expression Omnibus has templates that define the metadata/documentation to be submitted with each type of transcriptomic or epigenomic data. This includes a detailed description of the experimental design, goal and methods, and an inventory of both the raw and processed data files that must be submitted.

To facilitate the interpretation and reuse of the behavioral and physiological data, a README file and data dictionary will be generated and deposited into a repository along with all shared datasets. The README file will include method description, instrument settings, RRIDs of resources such as tracking software, mouse models and ages, and protocol details. The data dictionary will define and describe all variables in the dataset.
Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

All imaging, transcriptomic, behavioral and epigenomic and other data will be made available in formats (e.g. as .TIFF or .JPG image files or .XLSX or .TXT files or spreadsheets) that do not require the use of specialized tools to be accessed or manipulated.

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Data will be stored in common and open formats, such as .JPG or .TIF files for image data, .XLSX or .TXT files for spreadsheet data, and .DOC or .PDF files for published data. Information needed to make use of this data [e.g., the meaning of variable names, codes, other metadata, etc.] along with references to the sources of those standardized names and metadata items will be included wherever applicable.

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository.

All large datasets (e.g. transcriptomic and epigenomic data) generated in the course of this work will be deposited into the Gene Expression Omnibus (GEO) repository. All other data will be shared as main or supporting datasets/spreadsheets, images, or charts/graphs in our published manuscripts or will be deposited in Washington University School of Medicine’s institutional repository, Digital Commons@Becker.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Datasets and documentation will be disseminated via the Gene Expression Omnibus. Upon deposition into
GEO, each sample submission is assigned a unique digital object identifier (DOI), while the entire dataset is given a SuperSeries DOI. Data submissions are curated and assessed for completeness of the meta-template describing the data, methods, and samples, and deposition of corresponding raw and processed data files by GEO personnel. Release of the data is made upon acceptance for publication of the accompanying manuscript, at which time users can freely download and access these data. There is no end date for retention of the data in GEO.

Digital Commons@Becker is optimized for fast and accurate indexing by Google and Google Scholar. Digital Commons@Becker assigns DOIs as persistent identifiers, and has a robust preservation plan to ensure long-term access.

We will use Persistent Unique Identifiers (PIDs) to improve data findability across all dissemination outputs. PIDs used will include ORCID iDs for people, DOIs for outputs (e.g., datasets, protocols), Research Resource IDentifiers (RRIDs) for resources, and Research Organization Registry (ROR) IDs and funder IDs for places, as much as possible to make data identifiable and findable. We will also use indexed metadata, such as MeSH terms with a unique URL to make scientific data easily findable. We will keep our ORCID Records up to date with DOIs for our datasets and publications, ROR, and funder IDs to increase findability.

**When and how long the scientific data will be made available:** Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Release of datasets deposited into the Gene Expression Omnibus is made upon acceptance for publication of the accompanying manuscript, at which time users can freely download and access these data. There is no end date for retention of the data in GEO. Other data will be shared upon publication.

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.

**Access, Distribution, or Reuse Considerations**

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

There are no anticipated factors or limitations that will affect the access, distribution, or reuse of the scientific data generated by the proposal.
Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

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

Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

This work has been declared human subjects exempt by the WUSM HRPO and therefore does not directly involve human research participants.

The WUSTL IDDRC operates the IRB under which samples are generated and provided in a deidentified form for this work. In order to ensure participant consent for data sharing, WUSM IDDRC IRB paperwork and informed consent documents will include language describing plans for data management and sharing of data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

To protect participant privacy and confidentiality, shared data will be de-identified by assignment of a unique global unique identifier (GUID). Research participant data that is deposited into the NIMH Data Archive receives a GUID prior to deposition.

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Lead PI, Kristen Kroll, ORCID: 0000-0002-5450-6694, 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 Washington University stewardship, reporting, and compliance processes.
Planned Research Outputs

Data paper - "Anticipated publication: Requirements for ZNF292 in cortical interneuron development"

CRISPRi mediated knockdown of ZNF292 will be performed during neuronal differentiation of pluripotent stem cells. Cellular phenotyping and transcriptomic and epigenomic data will be conducted and included in this publication.

Data paper - "Anticipated research paper: A murine model of ZNF292 loss of function reveals its roles in brain development"

This study will characterize mouse knockout models of ZNF292 deficiency.

Data paper - "Anticipated publication: Mechanisms of pathogenic ZNF292 mutation in neurodevelopmental disorders"

Human stem cell models carrying pathogenic ZNF292 mutations will be characterized using standard assays, including transcriptomic and epigenomic data.

Dataset - "Transcriptomic and epigenomic data for a human stem cell-derived model of ZNF292 deficiency"

Dataset - "Transcriptomic and epigenomic data for murine a model of ZNF292 deficiency"

Transcriptomic and epigenomic data from mouse models of ZNF292 deficiency will be deposited in the Gene Expression Omnibus and released concurrent data paper publication.

Dataset - "Transcriptomic and epigenomic data for human stem cell-derived models of ZNF292 pathogenic mutation"

Transcriptomic and epigenomic data for human stem cell-derived models of ZNF292 pathogenic mutation.

Planned research output details

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Anticipated release date</th>
<th>Initial access level</th>
<th>Intended repository(ies)</th>
<th>Anticipated file size</th>
<th>License</th>
<th>Metadata standard(s)</th>
<th>May contain sensitive data?</th>
<th>May contain PII?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Type</td>
<td>Anticipated release date</td>
<td>Initial access level</td>
<td>Intended repository(ies)</td>
<td>Anticipated file size</td>
<td>License</td>
<td>Metadata standard(s)</td>
<td>May contain sensitive data?</td>
<td>May contain PII?</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------</td>
<td>--------------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Anticipated publication: Requirements for ZNF292 i ...</td>
<td>Data paper</td>
<td>2024-12-31</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>1 GB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Anticipated research paper: A murine model of ZNF2 ...</td>
<td>Data paper</td>
<td>2025-12-31</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>500 MB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Anticipated publication: Mechanisms of pathogenic ...</td>
<td>Data paper</td>
<td>2024-04-30</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>500 MB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Transcriptomic and epigenomic data for a human ste ...</td>
<td>Dataset</td>
<td>2024-12-31</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>1 GB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Title</td>
<td>Type</td>
<td>Anticipated release date</td>
<td>Initial access level</td>
<td>Intended repository(ies)</td>
<td>Anticipated file size</td>
<td>License</td>
<td>Metadata standard(s)</td>
<td>May contain sensitive data?</td>
<td>May contain PII?</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------</td>
<td>--------------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Transcriptomic and epigenomic data for murine a mo ...</td>
<td>Dataset</td>
<td>2025-12-31</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>500 MB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Transcriptomic and epigenomic data for human stem ...</td>
<td>Dataset</td>
<td>2025-04-30</td>
<td>Open</td>
<td>Gene Expression Omnibus</td>
<td>500 MB</td>
<td>Creative Commons Attribution Non Commercial No Derivatives 4.0 International</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>