Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D15S4Z

Title: Improving the quality of stem-cell derived insulin-producing cells

Creator: Xing Jian - **ORCID:** 0000-0002-0043-8729

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

Principal Investigator: Xing Jian

Data Manager: Xing Jian

Project Administrator: Ruthsabel O'Lexy

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: RFA-DK-21-030

Template: NIH-GDS: Genomic Data Sharing

Project abstract:

Type 1 diabetes (T1D) is a chronic autoimmune disease characterized by the loss of pancreatic β-cells. It accounts for up to 10% of all current cases of diabetes in the United States, and its overall incidence has significantly increased since 2002. The insulin-producing cells (SC-β-like cells) derived from human induced pluripotent stem cells (iPSCs) are promising renewable resources for both β-cell replacement
T1D therapies and biomimetic devices to study human pancreatic islets' function ex vivo. However, SC-β-like cells generated from current protocols do not meet all the critical properties of mature β-cells and are therefore not fully functional. The cells can reach maturation stage after transplantation into mouse models; however, this procedure is slow and costly, and it may not predict responses in humans accurately. The proposed project attempts to develop a superior protocol that would generate better functional cells with less time, expense, and variability. The proposed research has significant potential to achieve its long-term goal, namely, generating highly functional β-cells from iPSCs to create a much-needed model system for biomimetic devices and β-cell replacement therapies.
This R03 application’s rationale is based on published data from SC-β-like cells and mouse models. Recent studies of SC-β-like cells indicate that the key mature β-cell-specific transcription factor MAFA is only transiently expressed in the differentiation process, and its loss in the final differentiation stage leads to reduced β-cell functionality. Previous studies also showed that inducing MAFA expression in these cells would improve cell properties. Mouse animal model and cell line studies identified two biological pathways that are used in β-

cells to upregulate MAFA levels. In the first pathway, m6A methylation on MAFA mRNA increases its stability and abundance. In the other pathway, high concentration of glucose increases MAFA expression by the hexosamine biosynthetic pathway. The objective of this project is to use small molecules related to these pathways to upregulate MAFA levels so that the critical functional properties of these cells can improve. To achieve its objective, this project will pursue two specific aims: aim 1 - using small molecule modulators of the m6A RNA methylation pathway to increase MAFA expression levels in SC-β-like cells; aim 2 - using small molecule modulators of hexosamine biosynthetic pathway to increase MAFA expression levels in SC-β-like cells. The six critical properties of human β-cells outlined by the Human Islet Research Network (HIRN) will be used as guidelines to characterize SC-βlike cells. Project data will be generated from the Coriell Institute's facilities, with the exception that the next-generation sequencing will be carried out at an external high-throughput sequencing core facility, and analyzed with help from Coriell's Genomics & Epigenomics Core and its Bioinformatics Core.

Start date: 05-01-2023

End date: 04-30-2025

Last modified: 01-18-2024

Improving the quality of stem-cell derived insulin-producing cells

This study will use de-identified human cell lines from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository (https://www.coriell.org/1/NIGMS) without unique identify information. High throughput transcriptomic sequencing data generated in this study will be generated at single cell level. Single-cell transcriptome libraries will be prepared by a 10X Genomics Controller, and be sequenced by Illumina sequencers in a core facility. The raw data will be processed by Cell Ranger pipeline (proprietary software by 10X Genomics) and the Seurat workflow (open-source software). The raw data (FASTQ and TSV format) and the processed data (TSV format), as well as the associated metadata (TXT format) will be deposited to an NIH-designated data repository (see the next section), and the total size of the data is estimated to be 5 Gb. The phenotype data for the the cell line doners are publicly available at Coriell Institute website (www.coriell.org). The methods section of publications from this study will provide technique details of the experimental procedures. Other researchers should be able to download and independently perform analysis on the data, and validate the results of publications from this study. No custom-created code or in-house software will be used for data analysis.

High throughput transcriptomic sequencing data generated in this study will be deposited into Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), which is an NIH-funded repository, and freely distributes these data to the broad scientific community. In order to meet the MIBBI (Minimum Information for Biological and Biomedical Investigations) standard, the metadata (TSV format) will be prepared from the GEO metadata spreadsheet (https://www.ncbi.nlm.nih.gov/geo/info/examples/seq_template.xlsx). To facilitated file transfer and save storage space, the raw and processed data will be stored in GEO as compressed files (GZ format or TAR.GZ format). Submitted data will conform to relevant data and terminology standards in the single-cell transcriptome research field. The accession number (GSE) assigned to the data by GEO will be used as a persistent unique identifier, so that the data are findable and identifiable to the general public. No unusual data formats are planning to be used.

I will comply with the broad and rapid data sharing as outlined in the NIH Genomic Data Sharing Policy and will adhere to the changes to NIH-wide data sharing policies, such as the expected 2023 NIH Data Management and Sharing Policy. I agree to deposit high throughput transcriptomic sequencing data into GEO repository as soon as possible but no later than the date of the completion of the funded project period for the parent award or upon acceptance of the data for publication, or public disclosure of a submitted patent application, whichever is earlier. After the initial release of data, I will update the deposited data bi-annually. Data will be available at GEO to the general public as long as possible.

Since this study will use de-identified human cell lines from the NIGMS Human Genetic Cell Repository without unique identify information, the IRB at Coriell Institute determines that this study is not a human-subject study, and no approval from IRB at Coriell Institute is required. I agree that I will not use any data generated from this study for identification purpose.

The donors of the cell lines used in this study are participants of the Personal Genome Project (https://www.personalgenomes.org/). These samples have been consented for commercial use and for public posting of Personally Identifying Genetic Information (PIGI), which allows open-access, public posting of extensive genetic data. Participants the Personal Genome Project are consented to publicly share their genomic and trait data in an integrated, publicly-accessible format using a CC0 waiver or equivalent public domain license. Other than purposes of reasonable cost recovery, the participates of the Personal Genome Project are consented not to sell or license participant data or tissues. Since the data are intended to have open-access to the general public, data security will be controlled by GEO and NIH.

Because of the broad consent of the doners of the cell lines used in this study, this study is designated as not sensitive for the purposes of access to GSR. Principle Investigator of this project, Xing Jian will serve as the data manager, and will be responsible for performing all the required data management and sharing activities and the execution the data management and sharing plan. The Director of Research Office at Coriell Institute, Ruthsabel O'Lexy will serve as the project administrator, and oversee the data management and sharing plan, and monitor the compliance with the plan quarterly.

Planned Research Outputs

Dataset - "Single Cell Transcriptome Analysis of Stem Cell Derived Insulin-Producing Cells"

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?
Single Cell Transcriptome Analysis of Stem Cell De ...	Dataset	2025-04-29	Open	Gene Expression Omnibus	5 GB	Creative Commons Attribution 4.0 International	MIBBI (Minimum Information for Biological and Biomedical Investigations)	No	No