EXAMPLE DATA MANAGEMENT AND SHARING PLAN
(in compliance with SF-424 Forms H)

EXAMPLE FOR RAW WHOLE GENOME SEQUENCING DATA

ELEMENT 1: DATA TYPE

A. Types and amount of scientific data expected to be generated in the project:

We will obtain RNA-seq data through whole genome sequencing of livers and white adipose tissue of 10-week-old Pax5+/+ (WT) and Pax5-/- C57BL6 mice. Individual array samples were generated by pooling equal amounts of total RNA derived from livers and white adipose tissue of 5 mice of each genotype and sex.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

The proposed research will generate gene expression data from the samples described above obtained from RNA-seq analysis and will be deposited into the NCBI Gene Expression Omnibus (GEO) to be accessible to the broader scientific community

C. Metadata, other relevant data, and associated documentation:

Curated metagenomic data, including source name, tissue, strain, cell type, genotype, and treatment will be uploaded to the NCBI Biosample database and will be submitted to GenBank. Metadata will comply with and exceed the minimum requirements of the Genome Standards Consortium's (GSCs) Minimum Information about Metagenome Sequences (MIMS).

ELEMENT 2: RELATED TOOLS, SOFTWARE, AND/OR CODE

Raw data files will be made available and will be submitted to the Sequence Read Archive (SRA) repository via the GEO. The data will not require the use of specialized tools to be accessed or manipulated. Processed sequence data will also be made available through GEO and will be traceable using public accession numbers and chromosome coordinates. The reference assembly used will be included in the metadata files.

ELEMENT 3: STANDARDS

Data will be compliant with the MIAME and MINSEQE standards, as per the GEO submission procedures.

ELEMENT 4: DATA PRESERVATION, ACCESS, DISTRIBUTION AND ASSOCIATED TIMELINES

A. Repository where scientific data and metadata will be archived:

All data will be deposited to the GEO data repository and will be made publicly available upon publication of the associated manuscript, or the end of the grant performance period, whichever comes first.

B. How scientific data will be findable and identifiable:

Data will be discoverable through a GEO accession number, which will be cited in the corresponding publication and findable through search queries of GEO data sets. The data may also be findable through searches for keywords associated with the study. GEO datasets can be programmatically accessed using Entrez Programming Utilities, which are server-side programs that provide a stable interface to search, and retrieve GEO DataSets. Specific databases in which the data are stored

include Entrez Geo DataSets, which contains descriptive and accession information for all records, and Entrez GEO profiles, which contains gene annotation and synoptic/visual data for each expression profile.

C. When and how long the scientific data will be made available:
Data will be made available at the time of associated publication or at the end of the performance period, whichever comes first. Once the data are submitted to GEO, that archive will control the long-term persistence of the data set.

ELEMENT 5: ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

Data deposited into the GEO repository is publicly available.

B. Whether access to scientific data will be controlled

There are no controls on access to data in GEO.

C. Protections for privacy, rights, and confidentiality of human research participants:

Not applicable.

ELEMENT 6: OVERSIGHT OF DATA MANAGEMENT AND SHARING

John Doe, PhD will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the Data Management and Sharing Plan when necessary.