Appendix 1: NIMH Example Data-Resource Sharing Plan

1. **Grantee Information:**
   * **Project Title:** Genome Mapping and Functional Analysis in Schizophrenia.
   * **Principal Investigator:** John D. Smith, Ph.D.
   * **Institution:** University of Michigan, Ann Arbor, MI, USA
   * **Grant Number**: 1R01MH220716-01
   * **Project Period:** 01-01-2015 to 12-31-2018
2. **Summary of Project:** The objective of this study is to (1) employ unbiased and genome-wide approaches to identify the risk architecture of Schizophrenia (SCZ) and (2) and integrate genomic, and transcriptomic and epigenomic data to map quantitative trait loci (QTL), including expression QTLs (eQTLs), to map non-coding regulatory elements in associated with SCZ. In total, 1000 SCZ cases and 1000 age-matched control subjects will undergo structured clinical assessments with functional MRI; blood samples will be obtained from each subject for generation of lymphoblastoid cells lines and isolation of DNA for genome-wide association studies (GWAS). For follow-up functional analysis, skin punches will be obtained from 50 of these case and 50 control subjects for establishment of fibroblast cell lines and reprogramming to induced pluripotent stem cell (iPSC) lines. These iPSC lines will be differentiated to cultures highly enriched in layer 2/3 glutamatergic pyramidal neurons and analyzed using RNA-seq and ChIP-seq, along with analysis of structural and functional properties using robotic platforms for morphometry, multi-electrode array and other assays. These results will be compared with analogous data generated from layer 2/3 pyramidal neurons obtained from 100 post-mortem brains in a separate grant.
3. **Data/Biospecimen Descriptions and Submission Timeline:** The types of data and biospecimens to be collected are described below. These will be submitted to the NRGR (or other designated repository) and made available for public distribution on the following estimated schedule:

|  |  |  |
| --- | --- | --- |
| **Resource** | **Submission to Repositories** | **Release to public** |
| Established source cell lines for iPSC reprogramming (e.g., fibroblast cell lines) | * Not applicable; skin biopsies will be converted to fibroblasts by NRGR; fibroblast aliquots returned in kind to PI (or subcontractor) for reprogramming. | Not applicable. |
| Reprogrammed cell lines (e.g., iPSCs, iNSC, iGPC, iOPC) | * Sample type: iPSC * Sample number: 100 subject lines x 3 clonal lines/subject = 300 clonal lines * Reprogramming site(s): Rutgers (25 subjects) or Cellular Dynamics International (25 subjects). * QC data: outside-derived lines will be accompanied by relevant culture information to inform the use of cells (e.g., reprogramming factors, vectors, media, passage number). * Submission start date: 4/1/2015 * Submission end date: 12/1/2017 * Submission frequency: every 3 months for lines derived by CDI. | Released within 12 months after final delivery from contractor to project team, but not later than project end date (12/31/2018) or at time of publication, whichever occurs first. QC data released with respective iPSC lines. |
| GWAS data with accompanying study documents | * Data type: GWAS * Sample number: 2000 (DNA from LCL) * Platform: Illumina Human Exome + beadchip with 10,000 custom SNPs * Performance site(s): University of Michigan * Submission start date: 5/1/2017 * Submission end date: 11/30/2017 * Submission frequency: After data cleaning and quality control, data will be submitted at 6 month intervals. | Up to 6 months after data submission is initiated or at the time of acceptance of initial publication, whichever occurs first. |
| Sequence data (e.g., DNA sequence alignments to a reference sequence or de novo assembly, RNA expression profiling) | * Data type: Whole genome sequencing (WGS), ChIP-Seq * Sample number: 500 WGS (250 cases, 250 controls); 100 ChIP-Seq (50 cases, 50 controls) * Platform: WGS with Illumina HiSeq 2500, ChIP-Seq with H3K4me1, H3K4me3 * Performance site(s): University of Michigan * File formats: raw (fastq) and called genomes (SAM) * Submission start date: 12/31/2015 * Submission end date: 3/31/2017 * Submission frequency: After data cleaning and quality control, data will be submitted at 6 month intervals. | Up to 6 months after data submission is initiated or at the time of acceptance of initial publication, whichever occurs first. |
| Other genetic/genomic data (e.g., methylome, transcriptome, CNV) | * Data type: Transcriptome, Epigenome (from iPSC-derived neurons) * Sample number: 400 Transcriptome; 1200 Epigenome * Platform: Illumina HiSeq 2500 * Performance site(s): University of Michigan * File formats: raw (fastq) and called genomes (SAM) * Submission start date: 12/31/2015 * Submission end date: 3/31/2017 * Submission frequency: After data cleaning and quality control, data will be submitted at 6 month intervals. | Up to 6 months after data submission is initiated or at the time of acceptance of initial publication, whichever occurs first. |
| Analyzed data | * Summary analysis files of association tests * Results * Data from custom or proprietary clinical assessments/measures | Data released with publication. |