**RESOURCE SHARING PLAN: GENOMIC, PHENOTYPIC, AND CLINICAL DATA TO NDA**

1. **Summary of data to be shared and associated documentation**

Our genomic study will be [registered with dbGaP](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=HowToSubmit.pdf), and our whole genome sequencing data and derived data will be submitted to the NIMH Data Archive (NDA). Phenotypic and clinical data will be collected and deposited in NDA using the data dictionaries outlined in section B. All research subjects will be consented to allow broad data sharing for all genomic data and data will be collected in compliance with the NIH Genomic Data Sharing Policy.

The Institutional Certification will be submitted during the dbGaP registration process once we have been told that a grant award is likely. Within the first six months following the award, we will submit the Data Submission Agreement to NDA and will create the Data Expected section in our new NDA Collection.

1. **Standards/Data Dictionaries to be Used**

Genotypic data undergo an extensive automated data cleaning process in the laboratory. Our replication plan for observed associations is outlined in the Research Strategy. While all data from this proposal will be sequenced using Illumina pipelines, differences in read depth and primer libraries between studies will require joint re-calling of all genotypes from raw read files to yield the highest possible quality calls and a harmonized dataset for future use in follow-up and unrelated studies. Using the Broad Institute’s Genome Analysis Toolkit (GATK), we will apply standard Best Practices workflows for single nucleotide variant and InDel discovery from whole genome sequencing raw reads (SAM/BAM files). These steps should ensure that final association results are representative of “true” genotypes rather than miscalls or confounded genotypes that are unlikely to replicate in independent populations.

In compliance with NOT-MH-20-067, the following data will be collected to facilitate aggregation of this data set with other data sets:

1. Age (ndar\_subject01)
2. Sex at Birth (ndar\_subject01)
3. DSM Crosscutting (dsm5crossa0)
4. WHODAS 2.0 (whodas201)
5. PHQ-9 (phq901)
6. GAD-7 (cde\_gad701)

As described in the Research Plan, the additional phenotypic and clinical information will be collected using the indicated data dictionaries from the NIMH Data Archive:

1. Genomics Subject (genomics\_subject02)
2. Structured Clinical Interview for DSM-V (scidv\_dep01)
3. MATRICS Consensus Cognitive Battery (matrics01)

The sequence data will be stored in standard formats FASTQ, SAM/BAM, BED, and VCF. All raw and derived data will be deposited into NDA. Raw data will be submitted using the NDA genomics\_sample03 data structure and derived data will be submitted upon creation of an NDA Study to report research results. Experimental protocols will be described in NDA Experiments associated with our NDA Collection.

1. **Validation Schedule**

The NDA Validation and Upload tool will be used for quality control on newly collected phenotypic and clinical data every two weeks.

1. **Sharing Results**

Analyzed data associated with published research findings will be shared using the NDA Study feature and each NDA Study will be shared when a manuscript is accepted, prior to publication. The doi for each NDA Study will be referenced in the associated publication.