Data Management Plan (DMP)

Project Title: Personalized Medicine Approach for Hypertension Management: A Multi-Omics Study in a High-Risk Cohort

1. Data Types and Collection

Data Type	Volume/Format	Collection Method	Responsibility
Clinical Data	500 records (structured tabular)	Electronic Data Capture (REDCap/Survey)	Clinical Research Coordinator
Genomic Data	500 Whole-Exome Sequencing files (FASTQ, BAM, VCF)	Sequencer Output (External Vendor)	Bioinformatics Lead
Metabolomic Data	500 files (raw mass spectrometry data and processed CSV)	Mass Spectrometry Platform (Internal Core)	Metabolomics Core Manager
Analysis Code	R and Python scripts, Jupyter notebooks	GitHub/Institutional GitLab	Principal Investigator

2. Documentation and Metadata

All data will be thoroughly documented to ensure reproducibility and utility by other researchers.

- **Clinical Data:** Metadata includes definitions for all variables (e.g., units of measurement, data ranges, coding schemes for co-morbidities) documented in a README file alongside the final dataset.
- **Genomic Data:** All sequencing run parameters, alignment reference genomes (e.g., GRCh38), and variant calling pipeline details will be logged using the Common Workflow Language (CWL).
- **Metabolomic Data:** Standardized metabolite identification (PubChem IDs) and internal quality control (QC) procedures will be recorded.

3. Storage, Backup, and Security

Phase	Storage Solution	Security Measures	Retention Period
Active Analysis	Institutional High-Performance Computing (HPC) Cluster	Encrypted file system, restricted IP access, daily automated backups.	3 Years (Duration of Grant)
Archival (Long-term)	Institutional Data Archive (IDA) / Controlled Public Repository	Two-factor authentication, physical security controls, geographical separation of backups.	10 Years post-publication

Access Control: Access to linked, pseudonymous data is limited to the core research team. Publicly shared data (post-anonymization) will be stored in recognized controlled-access repositories, requiring researchers to submit an application and sign a data use agreement (DUA).

4. Data Sharing and Preservation

Anonymization Strategy

- **Direct Identifiers** (Name, DOB, Address) are removed immediately upon collection.
- **Indirect Identifiers** (e.g., rare genetic variants combined with detailed clinical history) will be reviewed by the IRB and the Data Access Committee prior to release.
- Genomic data will be stripped of specific single nucleotide polymorphisms (SNPs) that pose a re-identification risk, if necessary, to meet institutional guidelines.

Public Repository Details

Repository Name	dbGaP (Genomic Data) & Dryad (Clinical/Metabolomics)
Target Dataset	Final, fully anonymized multi-omics and corresponding clinical response data.
Timing of Release	Within 6 months of the final publication of the main results (Month 36).

Access Type	Controlled Access (Data Use Agreement required)

5. Roles and Responsibilities

Role	Responsibility
PI (Dr. A. Smith)	Overall responsibility for DMP compliance, final data submission, and DUA approval.
Bioinformatics Lead	Management of raw omics data, QC, pipeline execution, and metadata documentation.
Data Security Officer	Monitoring security protocols, managing access controls, and pseudonymization key management.