DATA MANAGEMENT AND SHARING PLAN

Element 1: Data Type

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

Summarize the types and estimated amount of scientific data expected to be generated in the project,

The data will be generated from acute myeloid leukemia (AML) human bone marrow biopsy samples from 8 patients and 3 AML human cell lines. The data will include those generated with western blotting, droplet digital polymerase chain reaction, and lentiviral techniques done in the biopsy samples and the cell lines. The data from these laboratory techniques will be shared as digital image files (TIFF). Genomic data will be derived from single-cell RNA sequencing (scRNA-seq) in bone marrow biopsy samples and sleeping beauty (SB) transposon technique in AML cell lines, which will be in FASTQ files. Phenotypic/Clinical data will include relevant demographic data, diagnosis, response to treatment and treatment regimen used in the tabulated form as text in the publication.

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

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

We will preserve and share genomic data, phenotypic/clinical data, and data from other laboratory techniques in our study.

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

Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

We will publish detailed data on methods we use for our research in any publications associated with this work.

Element 2: Related Tools, Software and/or Code:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

The digital image files TIFF will be shared.
The raw sequencing data will be collected and stored as compressed FASTQ data.
Other relevant phenotypic/clinical data will be available in the publication.

Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that

will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Data standards: Data generated from scRNA-seq and SB will be shared as FASTQ files. Digital image files will be shared as TIFF files. Phenotypic/clinical data will be in the published work in the tabulated form in text.

Metadata standards: Metadata standards: Deposits in dbGap are accompanied with a sequencing metadata spreadsheet.

Element 4: Data Preservation, Access, and Associated Timelines

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

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository).

We will use dbGaP as data repository for genomic data. Data generated from laboratory techniques, clinical/phenotypic data, and methods related to the study will be published. Original digital image files generated from the study will be with the research team and will be provided on request from other researchers. A digital object identifier (DOI) of the dataset(s) will be linked to the publication and vice versa.

B. How scientific data will be findable and identifiable:

Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Our genomic data will be available in dbGaP repository, which will be findable and identifiable using record accession numbers from dbGaP. We will publish data from laboratory techniques, phenotypic/clinical data, and methods related to the study. Some images will be published. All the original digital image files generated from the study will be with the research team and will be provided on request. A DOI of the dataset(s) will be linked to the publication and vice versa.

C. When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Data will be shared at the time of publication or the end of award period. Duration of retention and access to the data is dependent on the policies of dbGaP.

Element 5: Access, Distribution, or Reuse Considerations

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

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

We will remove all HIPPA designated identifiers from data to be shared.

B. Whether access to scientific data will be controlled:

State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).)

Access to individual-level genomic data deposited dbGaP will be available only after approval from study's data access committee. Phenotypic/clinical data will be available in the published data. While

some of the images will be published, all the original digital image files generated from the study will be with the research team and will be provided on request.

C. Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

HIPPA identifiers will be removed from data to preserve privacy and confidentiality. Individual-level genomic data will be shared only through controlled access via dbGap repository.

Element 6: Oversight of Data Management and Sharing:
Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Data will be submitted by Dr. Dhakal (the PI). He will ensure the data is submitted to the repository on time as described by this Data Management and Sharing Plan. His mentoring team will provide oversight and ensure compliance by Dr. Dhakal with the plan during regular meetings.