Epigenetic Database

What Database is needed for the AI?

DNA Methylation Database:

- **The Cancer Genome Atlas (TCGA):** Contains extensive cancer-related data, including DNA methylation profiles, which are often used in epigenetics.

- **ENCODE (Encyclopedia of DNA Elements):** Provides comprehensive data about regulatory regions of the genome, including DNA methylation and chromatin states.

- **MethBase:** A database of human epigenetic information, specifically focusing on methylation data from cancer and normal tissues.

Histone Modification Database:

- **ROADMAP Epigenomics:** Provides data about histone modifications, chromatin states, and DNA accessibility. It can be used to map changes in gene expression due to epigenetic alterations.

RNA and Epigenetic Database:

- **miRBase:** A repository for microRNA sequences and annotations, which can be useful for linking epigenetic regulation with gene silencing and cancer progression.

DNA Accessibility Databases:

- **ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing):** Databases like Cistrome DB provide chromatin accessibility data, which could be useful for understanding epigenetic control of gene expression.

Synthetic Data for Machine Learning:

- **EpiSCORE:** Provides large-scale synthetic datasets for training AI models in epigenetic biomarker discovery. Synthetic data can also be useful for

supplementing real-world data.

Combining Database:

Combining these databases would give access to different aspects of epigenetic changes (DNA methylation, histone modification, chromatin states, etc.), which could feed into the AI model to detect ultra-early signs of cancer. It may also need specific tools for processing and integrating this data, such as bioinformatics pipelines (e.g., Bismark, DeepTools) to handle epigenetic data and preprocess it for AI training.