

Web portal to visualize and retrieve pan-cancer methylome data

Internship summary report

Long Tian

Supervisor: Dr. Zhifu Sun

08/07/2018

Acknowledgement

- Dr. Zhifu Sun
- Mayo Clinic

Outline

- Introduction & Rationale
- Objectives
- Methods
- Results
- Future plans

Introduction & Rationale

- Exploring DNA methylation contributes to early cancer detection, tumor progression monitoring and potential drug discovery
- Large genomic data – half a million to a few million CpGs
- Large sample size – patient cohort ranges from hundred to tens of thousand
- The Cancer Genome Atlas (TCGA) is one of the resources that hosts methylome data for ~10k patients on 450k CpG sites

Objectives

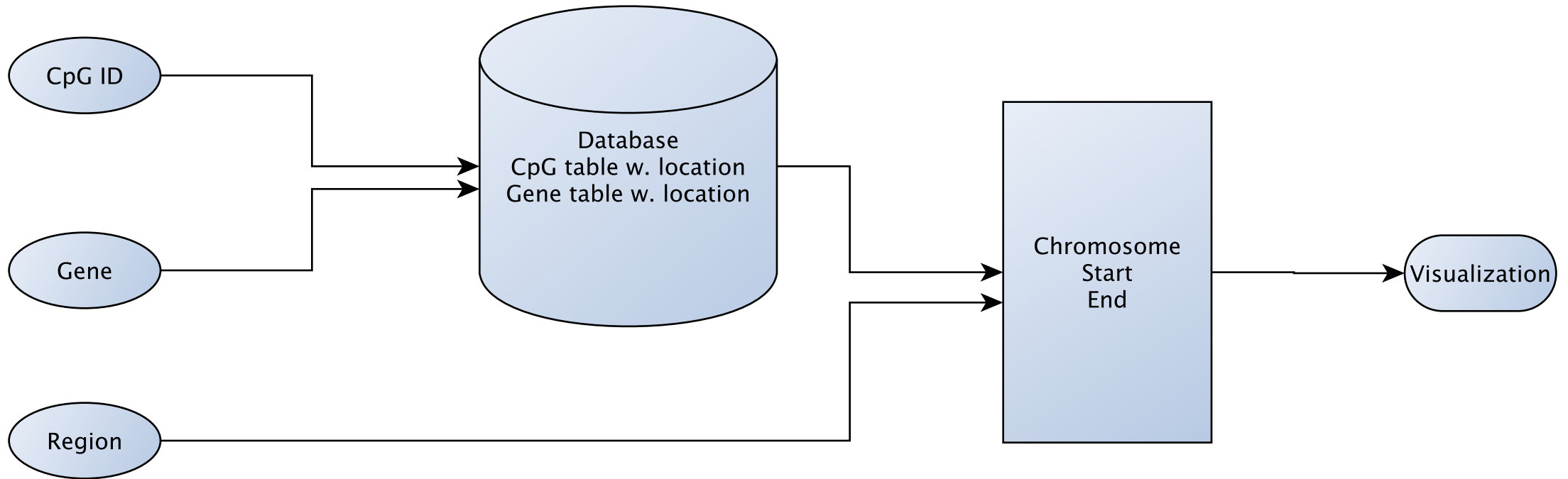
- Build a Web server to host data
- Create a Web interface to query data by CpG ID, gene, and genomic region
- View CpGs in gene context with a user-friendly way
- Visualize CpG methylation distribution within and across tumor types (normal and tumor samples)
- Perform differential methylation analysis between tumor and normal samples for each tumor type
- Enable the downloading of selected sub-dataset

Methods

- Web server
 - Microsoft Azure Virtual Machine
 - LAMP stack
 - Linux (Ubuntu 17.04) + **A**pache2 + **M**ySQL + **P**HP
- Database
 - MySQL
 - Metadata of samples (patients): tumor type, status, etc.
 - Metadata of CpG sites: genomic location, gene, etc.
 - Physical files on disk
 - DNA methylation data file for each CpG site

Method (Cont'd)

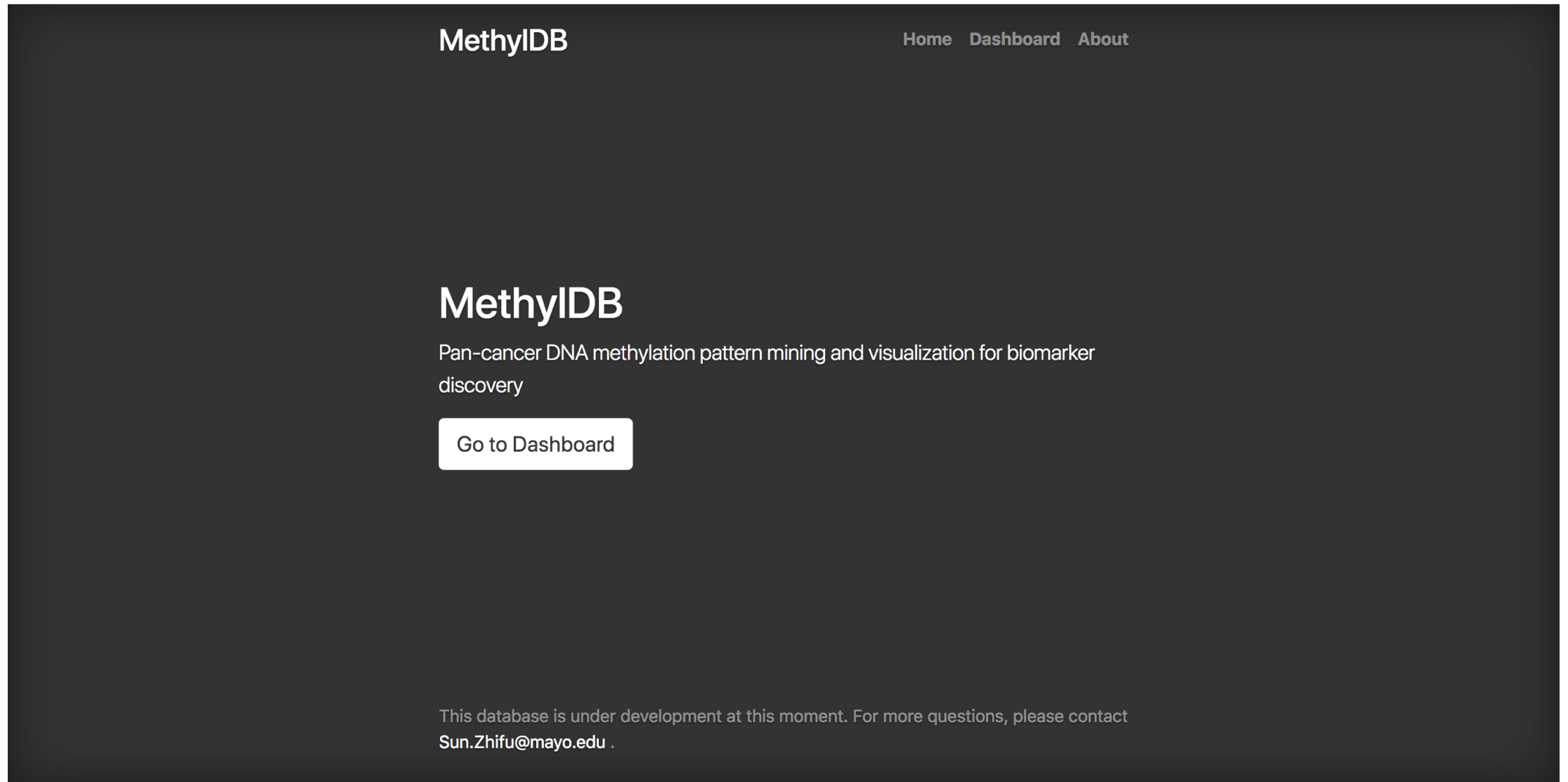
- Query data



Method (Cont'd)

- Dynamically visualize the distribution of DNA methylation data
 - Use igv.js to provide a genome viewer with customized CpG site information added
 - Use JavaScript and JQuery to “listen” and respond to the events (mouse clicking)
 - Use plotly.js, a D3.js-based JavaScript charting library to visualize the distribution of DNA methylation data
- Statistical analysis of DNA methylation data
 - T-test on each tumor type
 - Save results into tables and read upon CpG selection

Results



Future plans

- Make it available to connect to different datasets
- Further analytical functions
 - Associate CpGs with clinical results, e.g. survival analysis
 - Discover most differentially expressed markers in tumor types
- Publication

Questions?