MethylDB: An interactive visualization application of Differential Methylation Region

Xin Zhou C11909180   Ruichao Hou C11888117

December 3, 2014

1 Background And Motivation

The recent released TCGA data portal has contained large amount of Infinium HumanMethylation450 array (the 450k array) data, which provides a high-throughout assay to quantify DNA methylation at 480,000 loci across a range of genomic feature.

Since CpG's methylation status is so important for different tumor formation and cancer diagnosis, we will construct a database application to integrate differential methylated positions, differential methylated regions, red and green probes' .idat file, CpG probe archive and different CpG sites' related gene information. And then, since TCGA is a huge database which not only contains DNA methylation data but also many other resource such as gene expression data, patient clinical data and so on. As the result, we can build a comprehensive object-oriented databse and a related visualization application to help cancer researchers access the processed data from TCGA.

2 Tools and Database UML model

Here, we will provide a suite of computational tools that incorporate state-of-the-art statistical techniques for the analysis of raw DNAm data, and another Shiny-based web application to visualize the processed data from the object-oriented database.

2.1 Tools

To build a web application, we will use 'Shiny' to build our interface.

2.1.1 Shiny

Shiny is a web application framework for R, which can turn your analyses into interactive web applications and no HTML, CSS, or JavaScript knowledge required. Actually, author of Shiny package related CSS, javascripts and HTML in shiny, but for us we should only focus on application's logic and structure.
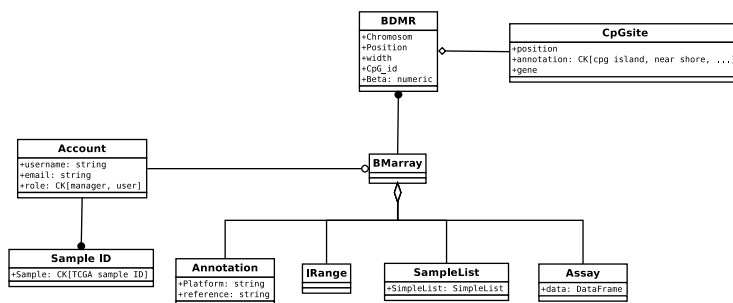
Figure 1: MethylDB's UML object model

In Shiny, we should define 2 or 3 mandatory .R files. They are ui.R, server.R and help.R. Here I will describe the structure of ui.R and server.R

```R
# ui.R
library(shiny)
library(knitr)
library(RMySQL)

shinyUI(pageWithSidebar(
    headerPanel("..."),
    sidebarPanel("..."),
    mainPanel("...")
))
```

```R
# server.R
shinyServer(function(input, output) {
  ...
})
```

In this example, we can seperate our application into 2 parts, ui.R is for interface design and serve.R is for database operation and other functions.

2.2 Model

In our design, we create a data model to communicate with TCGA's Illumina 450K database. Firstly, we create and keep an account object in our MethylDB to store each user's access information. Account object is an object to store each user's original information, besides attributes such as username, password, email, user account object is also connected to several TCGA Illumina 450k sample objects of TCGA. And the 450k sample objects store information of each sample's original 450k microArray information.
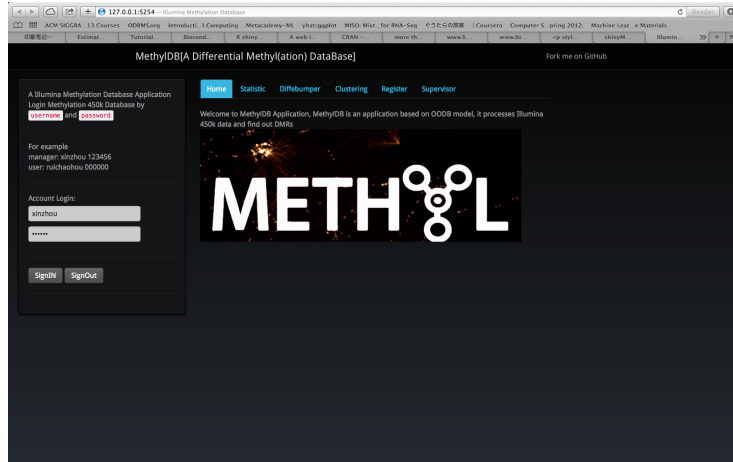
Figure 2: Main Interface of MethylDB

Additionally, our work is an advanced analysis of CpG sites from every account's 450k microArray information dataset, therefore, there exists an association between our comprehensive BMarray object and user account. Once a user creates his account, our system begins to generate a BMarray object based on his account object and TCGA sample object. And the multiplicity of this association is one or zero. It shows that every account has no more than one BMarray comprehensive object.

After our backend program successfully generates the BMarray object for specific user account, the user account of MethylDB begins to work. And then, we are going to find out all significant DMR(Differential Methylation Region) obejcts by this BMarray object of each account. Since a BMarray is object on whole genome, and each DMR object is cluster of several CpGsites which are closed. In a word, a BMarray object has meveral DMRs objects, and each DMR is build by one or many CpGsites objects. CpGsites object is a description of each CpG site, including attributes such as location, gene name, cpg island id and so on.

All these objects and relationships are described by the model in Figure 1. Based on this model, we can create our MethylDB application.

3   Result

Figure 2 shows the main page of our MethylDB. Our web application constructed by two panel: one is the side panel which contains log in interface and the other is the main panel which contains different components for different functions.
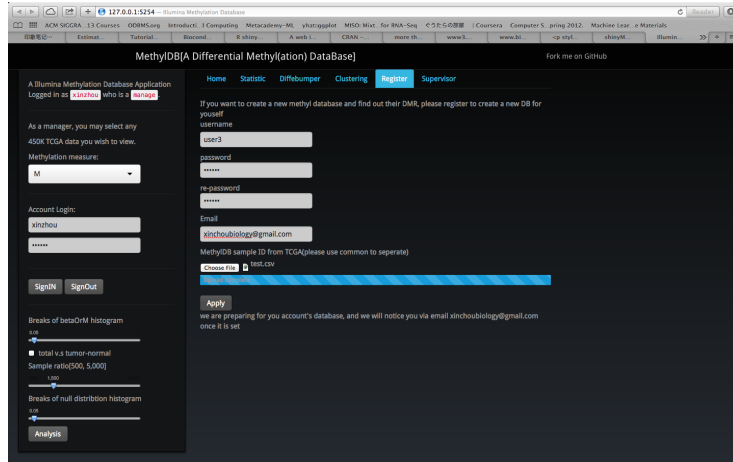
And our MethylDB has 2 main functions:

Figure 3: Register page of MethylDB

3.1 User account register and login

Register page offers a register entrance of our MethylDB, you need to submit your account's username, password, email and initial TCGA sample ID you prefered to our system. And then, system will help you generate account related BMarray object. And this step is a time consuming step, as the result, our web application will notice you via you email address after we prepare the BMarray object for your account.

Once user create an account, the system's account database will be modified, we can observed this change in our manager account.

When user submits his account registration information to MethylDB, system will insert his informaion into account object's database, but the attribute 'status' of the new account is defined as 'processing' and user can not log in his account when his account's status is 'processing', and after BMarray objects generated, user can log in his account.

3.2 BMarray object and DMRs object visualization

There are 3 tabs for BMarray object and DMRs object visualization: 'Statistic' tab is for BMarray's visualization. It show basic information of BMarray object.

And the the 'Diffbumper' tab is for the DMRs object's visualization, since different BMarray will generate different DMRs set, our MethylDB will help user to generate a beautiful report of k most signifcant DMR from user specific BMarray object. Here, k is the number of DMRs user want to get.

'Clustering' tab is also built for DMRs objects summarization, it will represent each DMR of each sample ID via a heatmap. The heatmap implies DMRs' correlationship. And with the help of clustering heatmap and report about top k significant DMRs, user can find DMRs they are interested in for their future
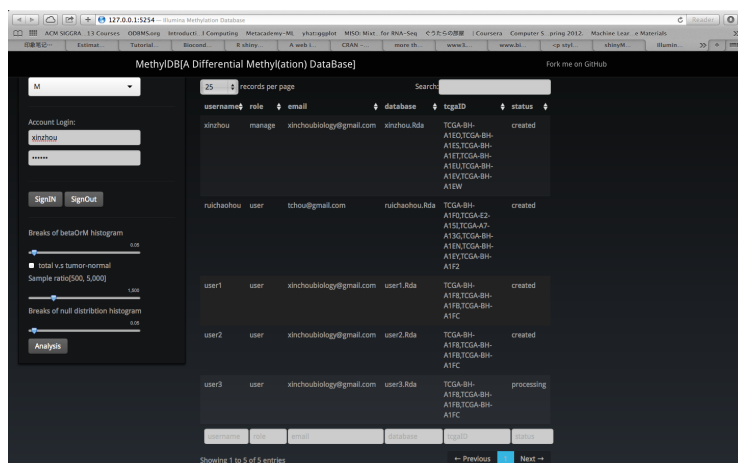
Figure 4: New account registration will insert new user information into account database
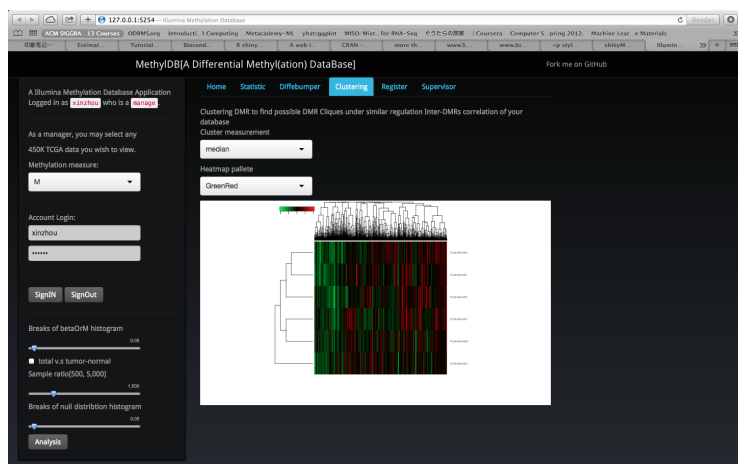

Figure 5: Clustering tab for DMRs object

research or diagnosis.

4 Conclusion

MethylDB application is created by cooperation between Xin Zhou and Ruichao Hou. Ruichao mainly works for MethylDB's obejct model and Xin mainly works for MethylDB's creation. MethylDB is a database web application and can dynamically respond to user's command. It is also an analysis tool for TCGA database.