SNPs affect RNA protein interactions or binding sites of Alzheimer's disease and related dementia

Benyu Zhou

Katherine Tu

Mintao Lin

Qinrui Wu

Xiaojing Peng

BOSTON
UNIVERSITY

Study Background

❖ Alzheimer's Disease

➢ A progressive disease that destroys memory and other important mental functions.

❖ RNA binding proteins(RBP)

➢ proteins that bind to the double or single stranded RNA in cells and participate in forming ribonucleoprotein complexes

➢ RBP normally are present in the nucleus. Under conditions of cell stress, these RBP translocate to the cytoplasm where they form stress granules, which function in part to sequester specialized transcript and promote translation of protective proteins.
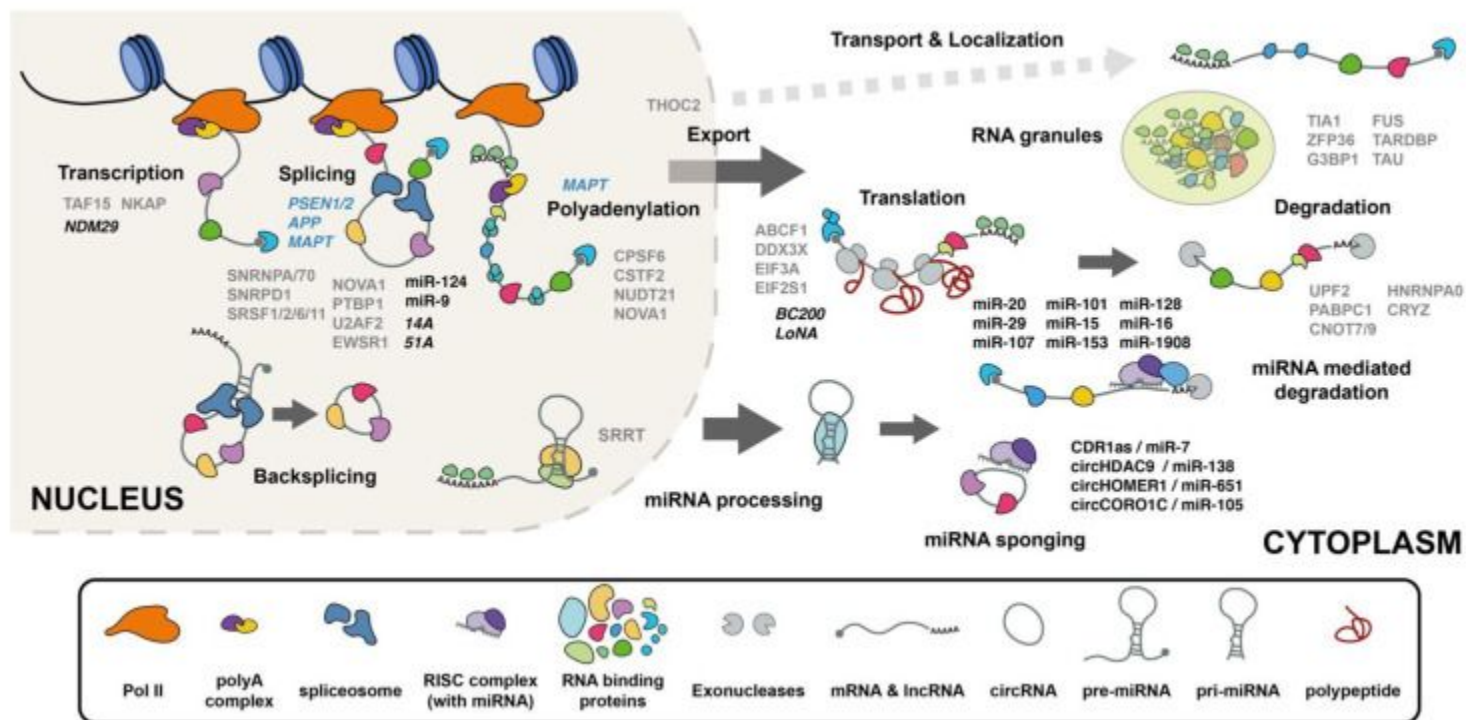
Figure.1. Rybak-Wolf, Agnieszka, and Mireya Plass. "RNA Dynamics in Alzheimer's Disease." Molecules (Basel, Switzerland) vol. 26,17 5113. 24 Aug. 2021, doi:10.3390/molecules26175113

Study Background

❖ Locus

➢ A locus is a specific, fixed position on a chromosome where a particular gene or genetic marker is located.

❖ SNP

➢ A single-nucleotide polymorphism is a germline substitution of a single nucleotide at a specific position in the genome

❖ eCLIP

➢ An enhanced version of the crosslinking and immunoprecipitation (CLIP) assay, and is used to identify the binding sites of RNA binding proteins (RBPs)

Study Background

❖ **Genome-wide association studies (GWAS)** have identified >50 susceptibility loci showing robust evidence for association with late-onset Alzheimer's disease (AD).

❖ A striking fraction of these loci implicate RNA metabolism. **Some loci code for RNA binding proteins (RBPs)**, which regulate RNA splicing.

❖ Understanding the effect of these same-sense SNPs has wide-ranging implications for understanding AD.

Questions to be answered:

❖ Investigate the relationship between SNPs and their effect on RBP binding
❖ Number of loci on each specific RNA and the function of each Loci
❖ Do AD-related SNPs impact on RNA-binding protein?

Database Working Pipeline

❖ Gather and extract data from ENCODE

❖ Create database

❖ Design the web interface

❖ HTML/Javascript/AJAX

❖ Use query to access the data in the database

❖ Create CGI program

Datasource

1. Single nucleotide polymorphisms (SNPs) data comes from The **International Genomics of Alzheimer's Project (IGAP)** (Alzheimer's disease GWAS of Kunkle et al. , Nat Genet, 2019).
2. RNA-binding proteins (RBPs) data is download from **ENCSR456FVU**(Reference: A large-scale binding and functional map of human RNA-binding proteins, Nature)
3. Extra DATA: to find the targeted genes, we use bamtools to find the RBP target genes
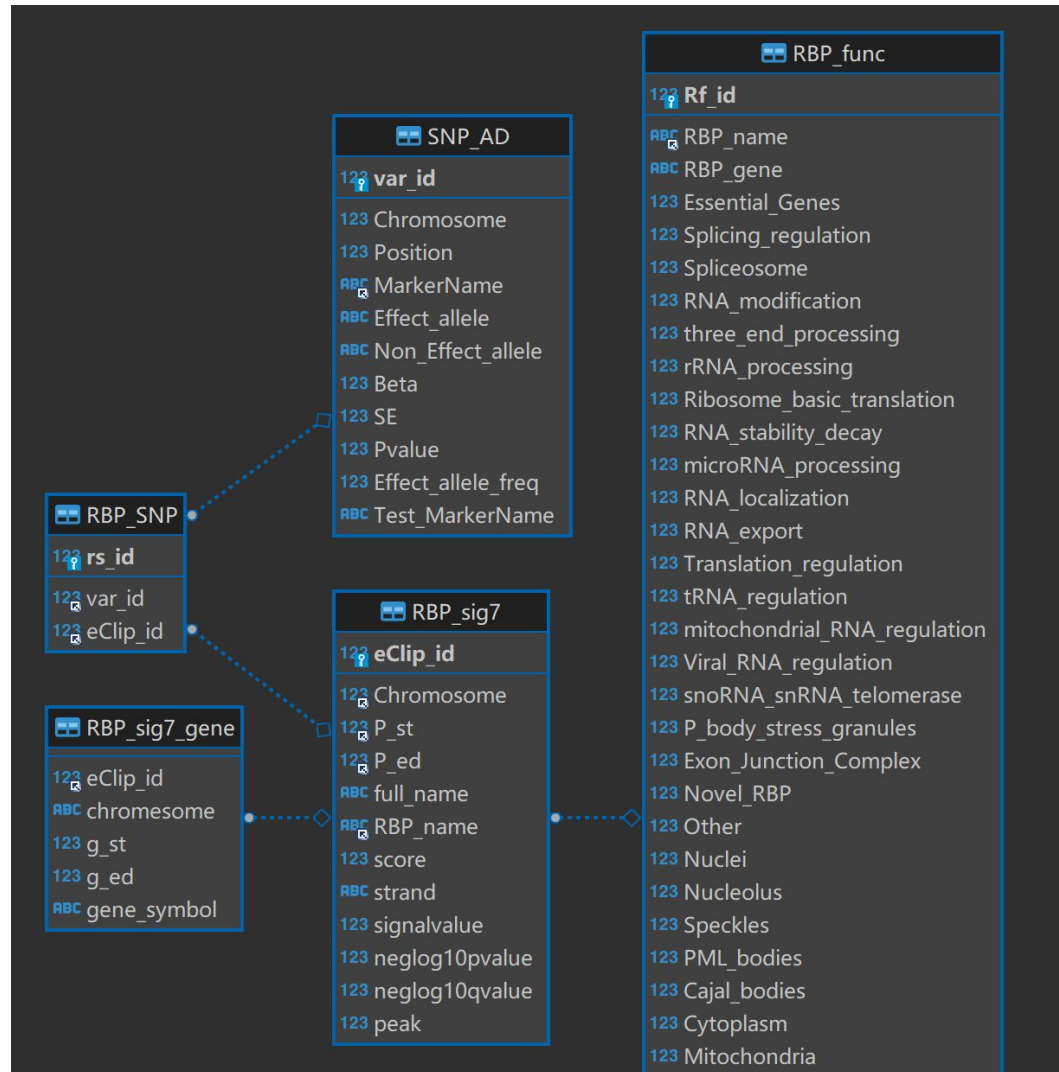
Scripts:
https://github.com/Rey572/BF768_FinalProject

Build Database
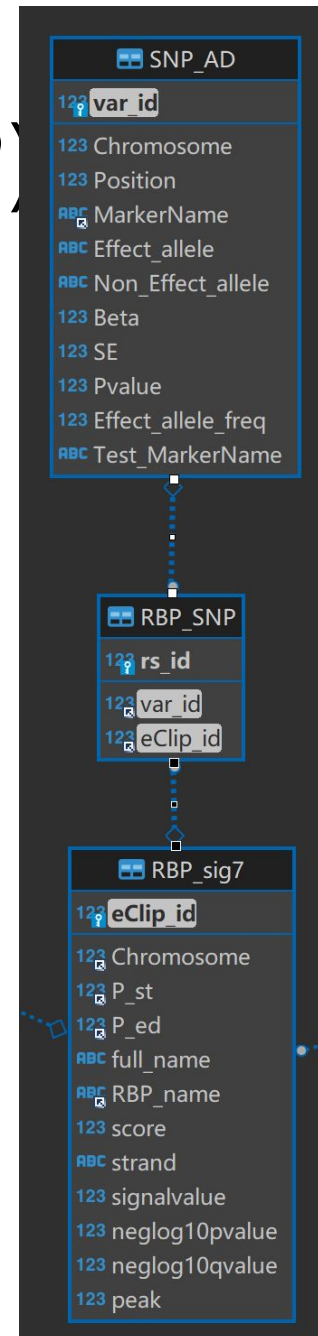
Relationship 1(SNP and RBP)

Many to Many Relationships

Both **partial participation**

One SNP could be on a Gene regulated by **more than one** RBP

The genes regulated by **one** RBP could have **more than one** SNPs

Not every SNP is regulated by a RBP, Not every RBP regulate a gene having SNP

SNP_AD
- var_id
- Chromosome
- Position
- MarkerName
- Effect_allele
- Non_Effect_allele
- Beta
- SE
- Pvalue
- Effect_allele_freq
- Test_MarkerName

RBP_SNP
- rs_id
- var_id
- eClip_id

RBP_sig7
- eClip_id
- Chromosome
- P_st
- P_ed
- full_name
- RBP_name
- score
- strand
- signalvalue
- neglog10pvalue
- neglog10qvalue
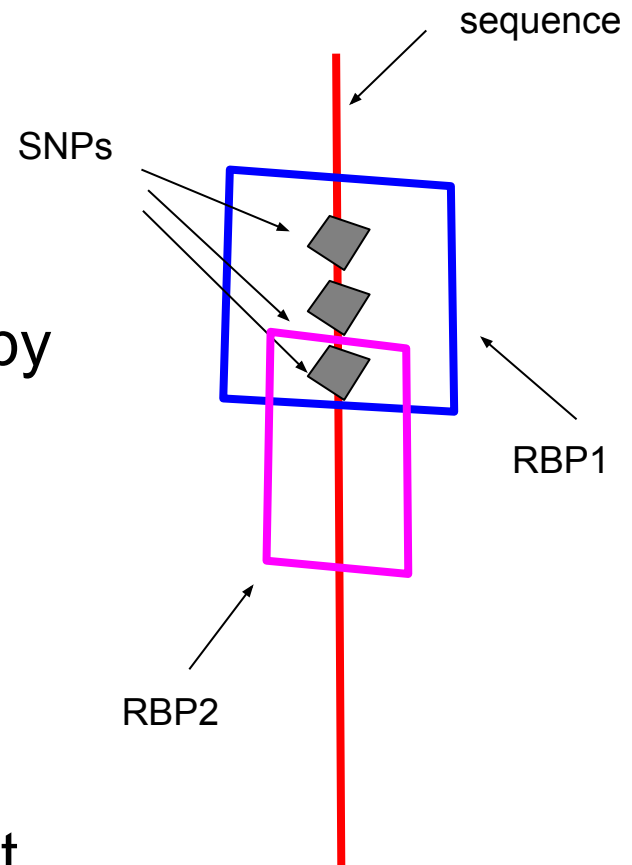- peak

Relationship 1(SNP and RBP)

Many to Many Relationships

Both **partial participation**

One SNP could be on a Gene regulated by **more than one** RBP

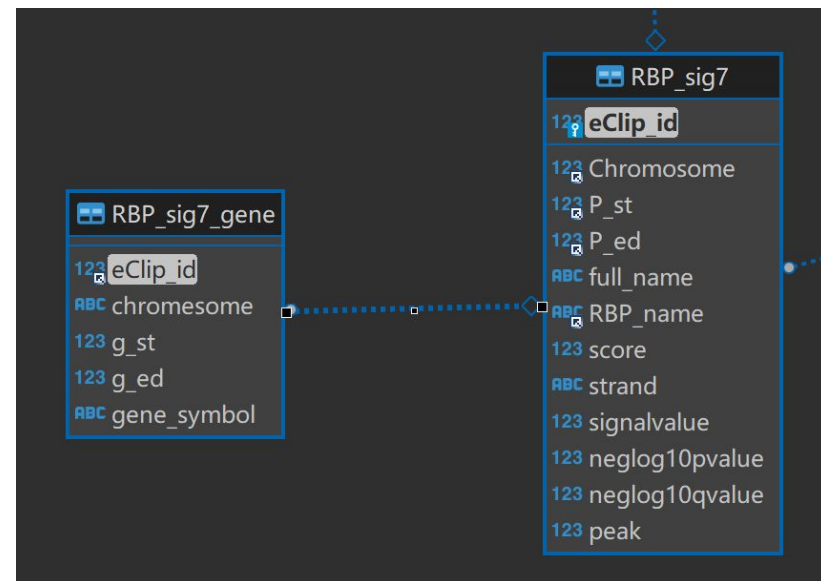The genes regulated by **one** RBP could have **more than one** SNPs

Not every SNP is regulated by a RBP, Not every RBP regulate a gene having SNP

sequence

SNPs

RBP1

RBP2

11

Relationship 2 (RBP and its target genes)
one to many relationship

A targeted gene could be regulated by different RBPs.

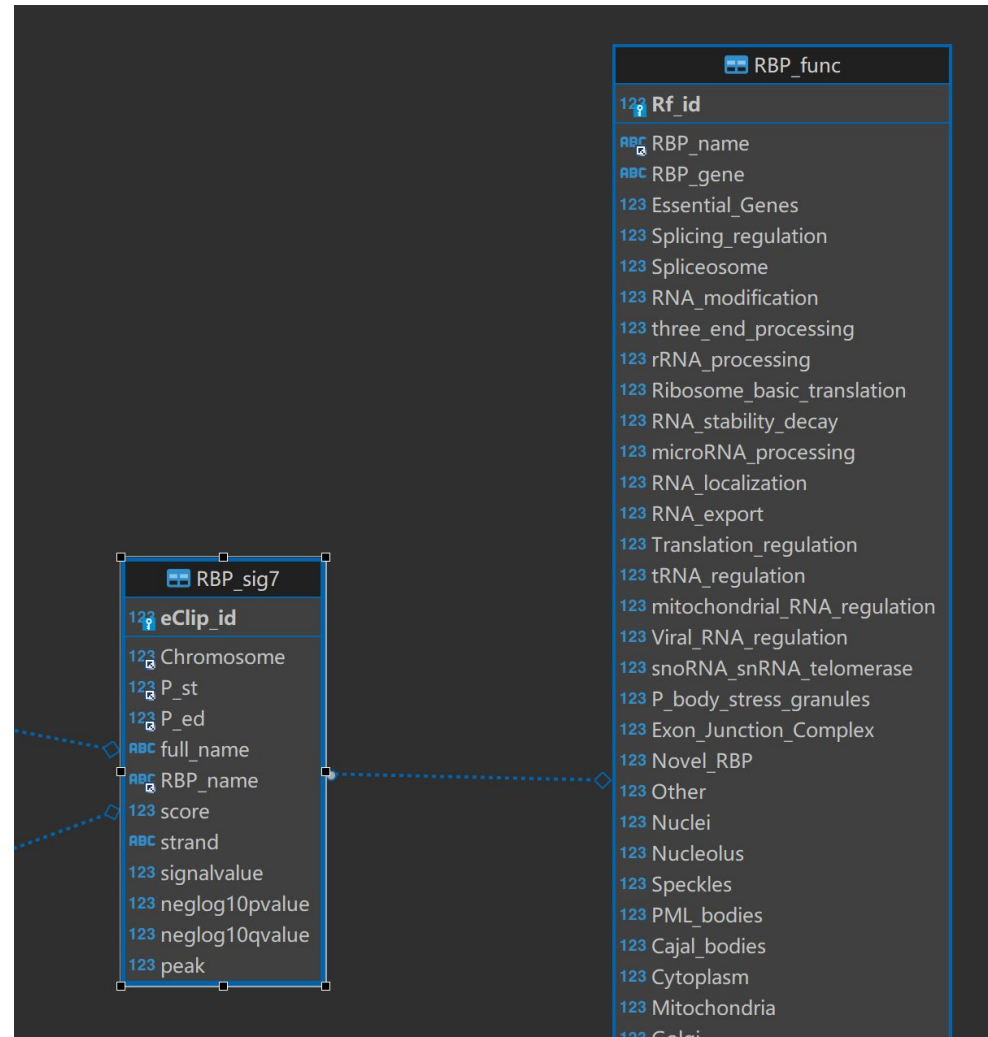# Relationship 3 (RBP and its function)

one to many
Relationship

A RBP must belongs to
a function cluster

A function cluster could
have **more than one**
RBP

Something interesting we met

How to build the many to many relationship for SNP and RBP?
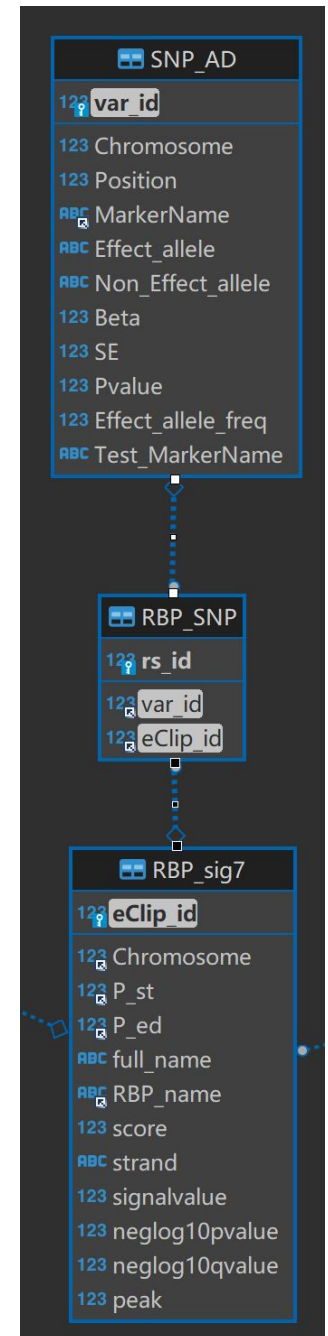
If we use sql to build the relationship table

Data size:

SNP: 676,450 rows; RBP: 6,517,506 rows

If we use join directly(using index at the same time)

the complexity is O(nlogm)= 676,450 log(6,517,506)= 10 ^ 8

Theoretically, it would not take more than 1 seconds

SNP_AD
- var_id
- Chromosome
- Position
- MarkerName
- Effect_allele
- Non_Effect_allele
- Beta
- SE
- Pvalue
- Effect_allele_freq
- Test_MarkerName

RBP_SNP
- rs_id
- var_id
- eClip_id

RBP_sig7
- eClip_id
- Chromosome
- P_st
- P_ed
- full_name
- RBP_name
- score
- strand
- signalvalue
- neglog10pvalue
- neglog10qvalue
- peak

Something interesting we met

How to build the many to many relationship for SNP and RBP?

However, it takes a lot of time using sql.(more than three hours and we did not receive any result)

We finally decide to build a Balanced tree to get the relationship table.


```cpp
#include <cstdlib>
#include <iostream>

using namespace std;
struct tree
{
  int val,fix,weight,size;
  tree *left,*right;
};
int resize(tree *tr)
{
  if (tr==NULL) return 0;
  else return tr->size;
}
void turnleft(tree *&x)
{
  tree *y=x->right;
  x->right=y->left;
  y->left=x;
  x=y;
  y=x->left;
  y->size=resize(y->right)+resize(y->left)+y->weight;
  x->size=resize(x->left)+resize(x->right)+x->weight;
}
void turnright(tree *&x)
{
  tree *y=x->left;
  x->left=y->right;
  y->right=x;
  x=y;
  y=x->right;
  y->size=resize(y->right)+resize(y->left)+y->weight;
  x->size=resize(x->left)+resize(x->right)+x->weight;
}
void insert(tree *&head,int item)
{
  if (head==NULL)
  {
    head=new tree;
    head->val=item;
    head->right=NULL;
```

BU

Code available:
https://github.com/Rey572/BF768_FinalProject Tree

15

Webpage

BF768 Final Project - Group B

Home | AD-related SNP-RBP binding | Loci Number | AD-related RBP function | DIY | Help

SNP-RNA Protein Interactions for Alzheimer's Disease and Related Dementia

Genome-wide association studies (GWAS) have identified more than 50 susceptibility loci showing robust evidence for association with late-onset Alzheimer's disease (AD). A striking fraction of these loci implicates RNA metabolism. Some loci code for RNA binding proteins (RBPs), which regulate RNA splicing. Other susceptibility loci contain polymorphisms that alter splicing of the gene or contain altered splicing patterns associated with AD. Overall, more than 90% of SNPs identified from GWAS occur in non-coding regions. Non-coding SNPs can appear in noncoding RNAs, introns, or in 5' and 3' untranslated regions (UTRs). Our goal is to understand the effect of these non-coding or same-sense SNPs and their wide-ranging implications for understanding AD.

About us

This page is developed at Boston University as part of BF768 Biological Database Systems, Spring 2022 by students Benyu Zhou, Katherine Tu, Mintao Lin, Qinrui Wu, and Xiaojing Peng and guided by instructor G. Benson and professor Xiaoling Zhang from the Department of Medicine at Boston University.

16

Webpage

BF768 Final Project - Group B

Home | AD-related SNP-RBP binding | Loci Number | AD-related RBP function | DIY | Help

AD-related SNP-RBP binding

Select p-value
- 0.05
- 0.01

submit

Description | Table | Plot

Description summary:
the number of SNP,
the number of RBP

17

Webpage

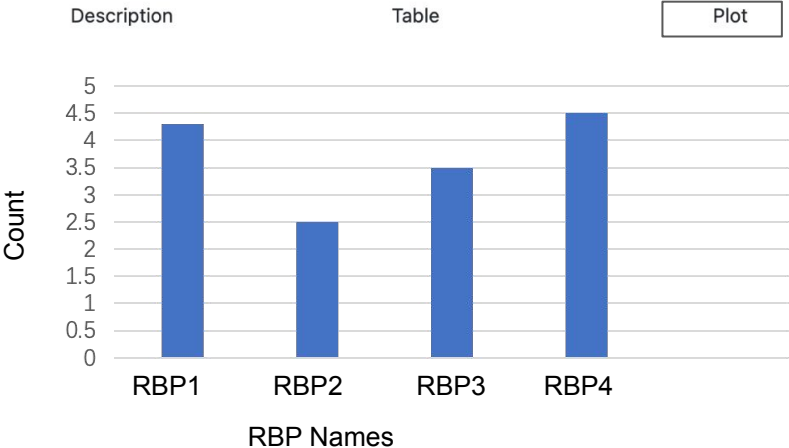BF768 Final Project - Group B

Home | AD-related SNP-RBP binding | Loci Number | AD-related RBP function | DIY | Help

AD-related SNP-RBP binding

Select p-value
- ● 0.05
- ○ 0.01

submit

Description   Table   Plot

SNP-chr	SNP position	SNP marker name	SNP P-value	RBP Name	P-value	Output gene

Download table

18

Webpage

Webpage

BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

Loci Number

Input RBP name (e.g. BUD13): [   ]

submit

Description   Table   Plot

Description summary:
    The number of SNP,
    Genes annotated to loci

20

Webpage

BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

Loci Number

Input RBP name (e.g. BUD13): [   ]

submit

Description   Table   Plot

SNP chromosome	SNP position	SNP marker name	P-value for SNP	Target Gene
				APOBEC3C

Download table

Webpage

BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

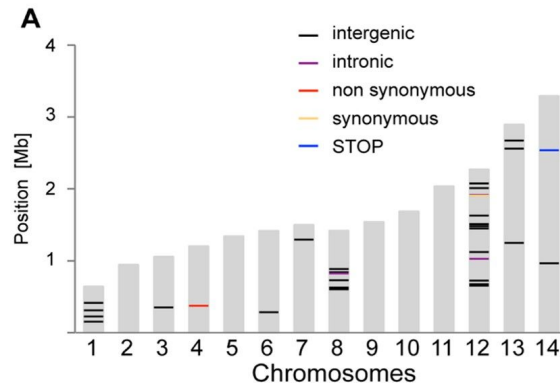## Loci Number

Input RBP name (e.g. BUD13): [   ]

submit

Description   Table   Plot

Example graph:
Bopp et al. PlOS Genetics
2013;9(2):e1003293)

Webpage

BOSTON UNIVERSITY BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

AD-related RBP function

Input SNP: [   ]
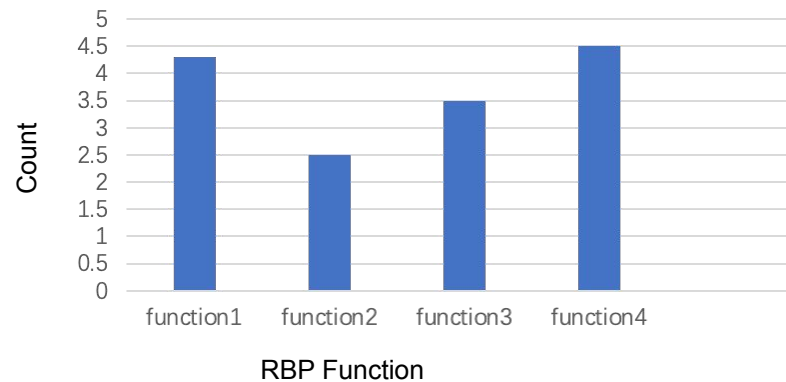
[submit]

[Description]   Table   Plot

Description summary:
    Beta value of the SNP
    Whether a Risk/Protective SNP
    The number of RBPs that target the SNP

Webpage

Webpage

Webpage

BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

DIY query

Input query:

[text input box]

Submit

Table

Webpage

BF768 Final Project - Group B

Home   AD-related SNP-RBP binding   Loci Number   AD-related RBP function   DIY   Help

Under construction…

CGI

❖ **APIDB.py**

  ➢ Application Programming Interface to our database.
  ➢ In this py file, a class named APIDB is defined that connects to the database on bioed and performs all queries based on selectors.

❖ **api_main.py**

  ➢ This py file contains main CGI structure that receives html form input and and use the APIDB class to returns query results in json form to html.

Reference

1. Van Nostrand, E.L., Freese, P., Pratt, G.A. et al. A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719 (2020). https://doi.org/10.1038/s41586-020-2077-3
2. Van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B., Rabano, I., Stanton, R., Sundararaman, B., Wang, R., Fu, X. D., Graveley, B. R., & Yeo, G. W. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. Genome biology, 21(1), 90. https://doi.org/10.1186/s13059-020-01982-9
3. Wolozin, Benjamin, and Daniel Apicco. "RNA binding proteins and the genesis of neurodegenerative diseases." Advances in experimental medicine and biology vol. 822 (2015): 11-5. doi:10.1007/978-3-319-08927-0_3
4. Wikipedia contributors. (2022, February 17). RNA-binding protein. In Wikipedia, The Free Encyclopedia. Retrieved 00:16, April 27, 2022, from https://en.wikipedia.org/w/index.php?title=RNA-binding_protein&oldid=1072461325
5. Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, et al. (2013) Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. PLoS Genet 9(2): e1003293. https://doi.org/10.1371/journal.pgen.1003293