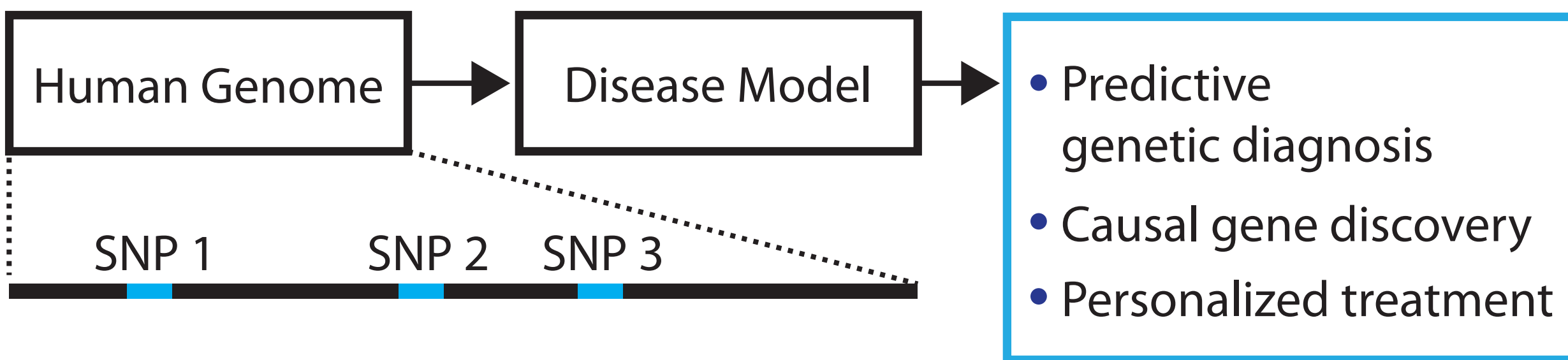




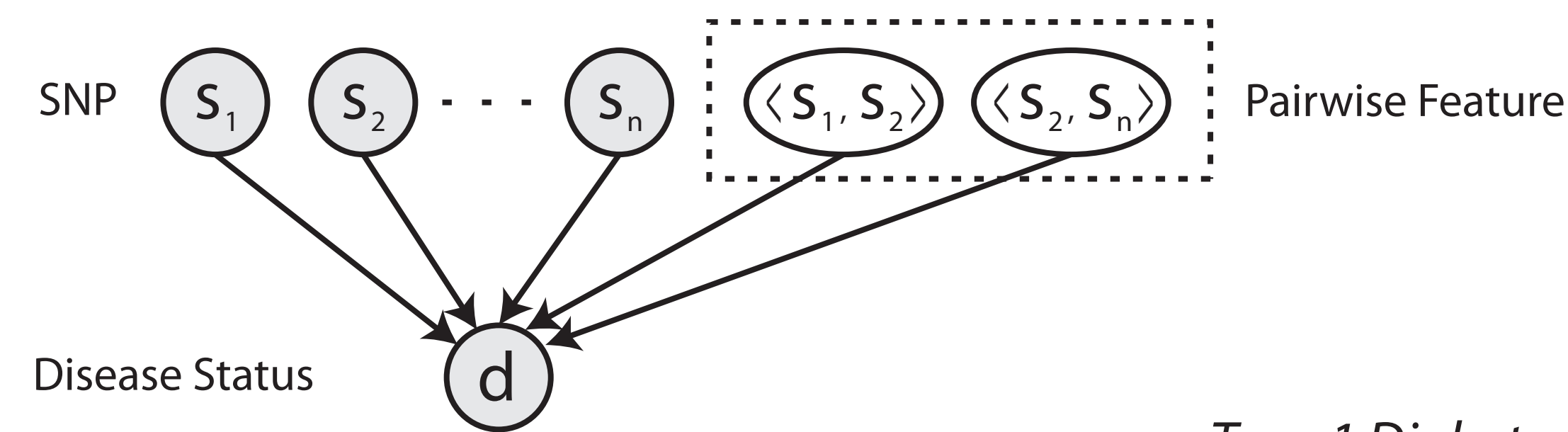
## Motivation

### Personalized Genomics



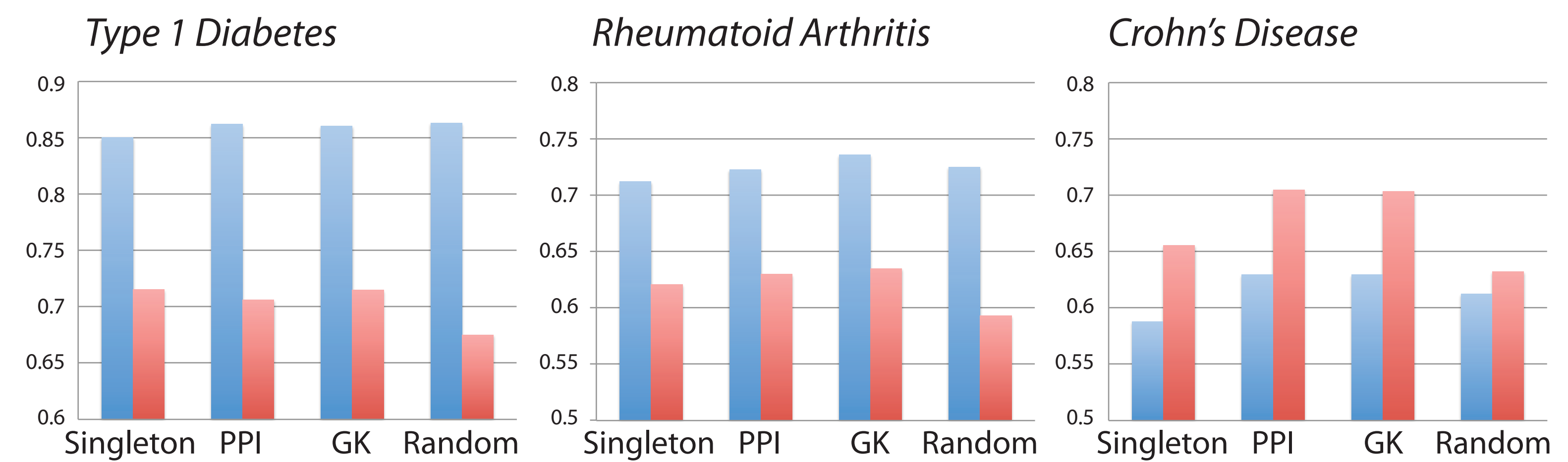
- Conventional method considers **each SNP independently**, which **does not capture genetic interactions** and oversimplifies the cause of human diseases.
- Heritability of diseases is not being fully explained** with current genetic models. Capturing genetic interaction may help.
- A model that better represents the underlying process will improve numerous clinical applications.

## Effect of Pairwise Features



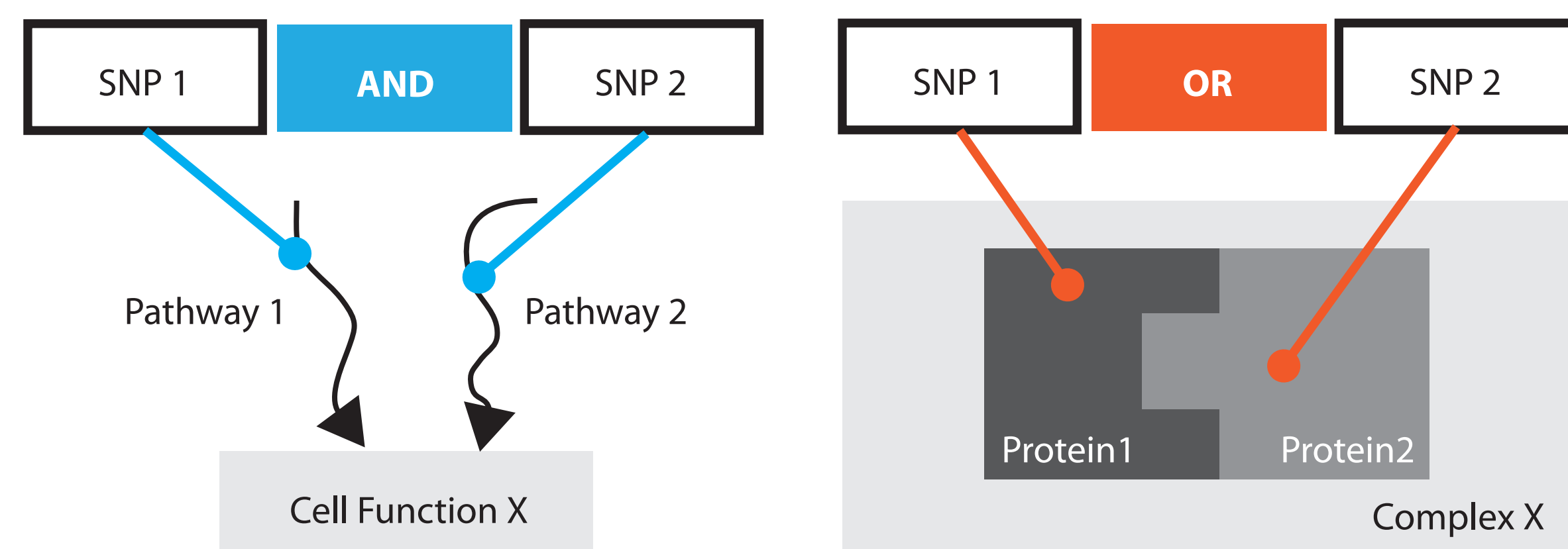
- We compare the performance of logistic regression model with and without pairwise features.
- We also test how well the regression weights predict genes that are known to be related with the disease.

▶ **Addition of pairwise features always improve classification performance**, regardless of which method is used for feature selection. Randomly selecting pairs fails in gene prediction, while PPI and GK features often improve it.



## Pairwise Feature

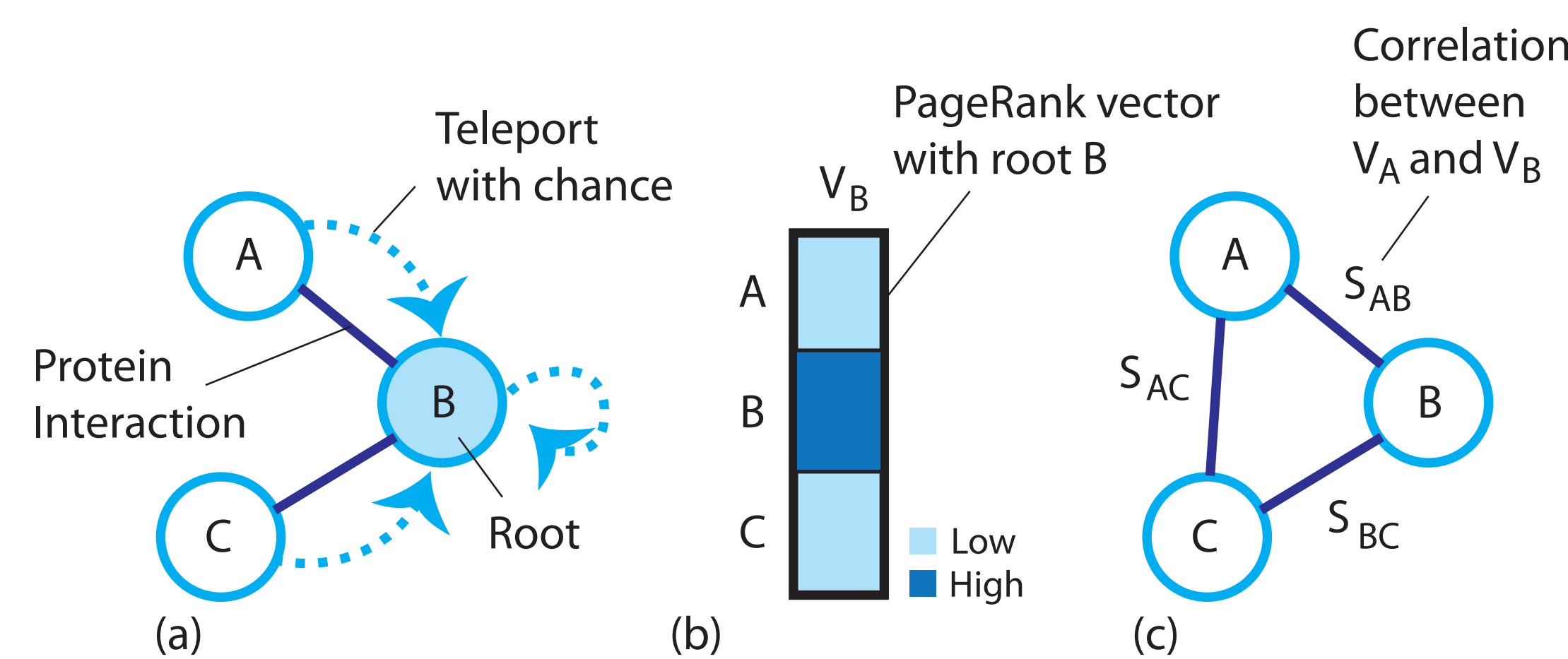
We introduce pairwise features between SNPs that resembles **AND and OR logical operations** to capture genetic interactions.



- For two redundant pathways, AND feature indicates **whether the function is entirely lost**.
- For a complex with two proteins, OR feature indicates **whether there is any defect**.

## Graph Kernel (GK) Using Personalized PageRank

We **infer genetic interaction** between any pair of genes by computing connectivity with Personalized PageRank (PPR).

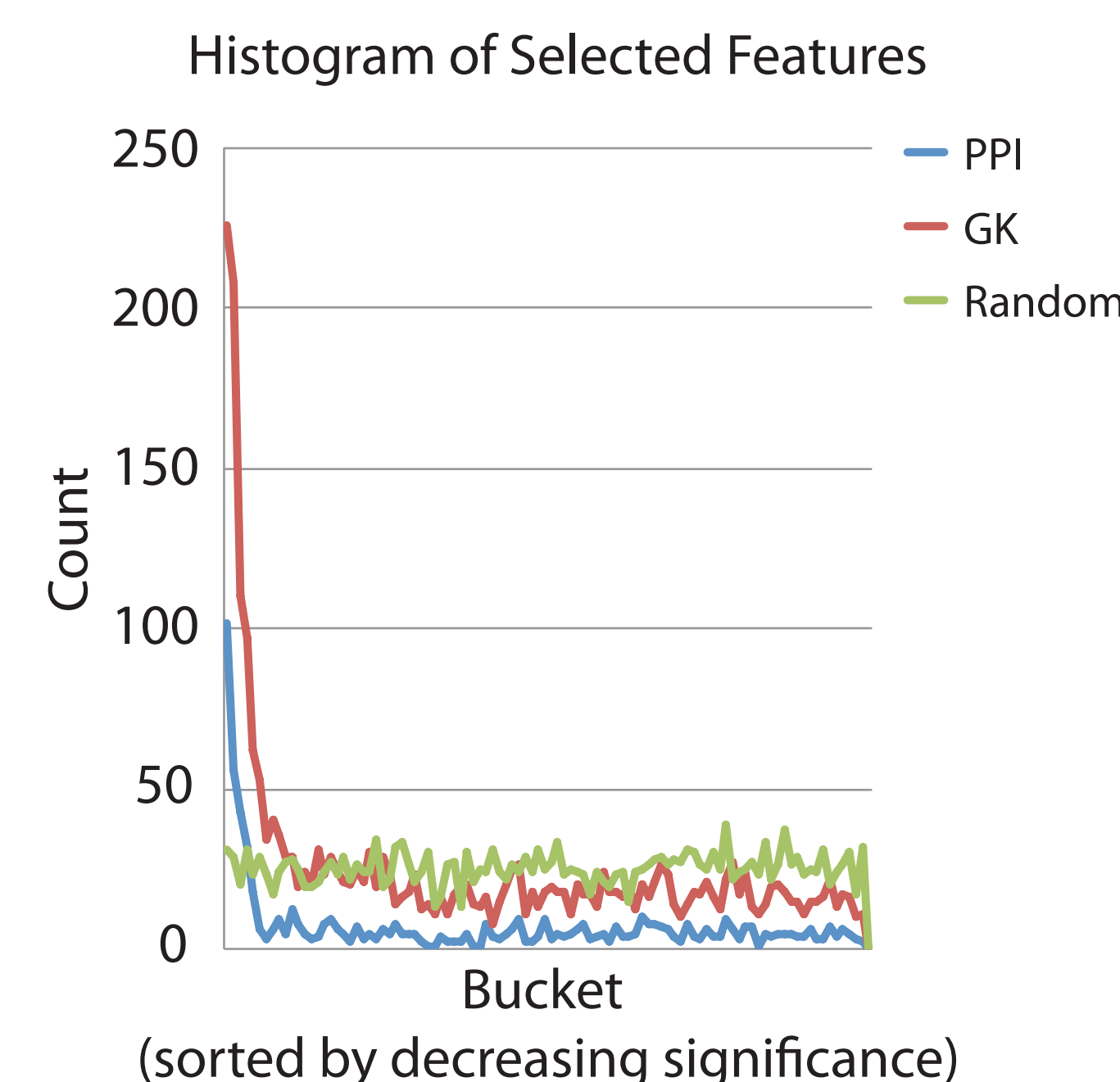


- (a) PPR is PageRank with a chance of teleporting back to root after each iteration. (b) We run PPR on each gene. (c) Correlation between PageRank vectors represents connectivity.

## Selectivity

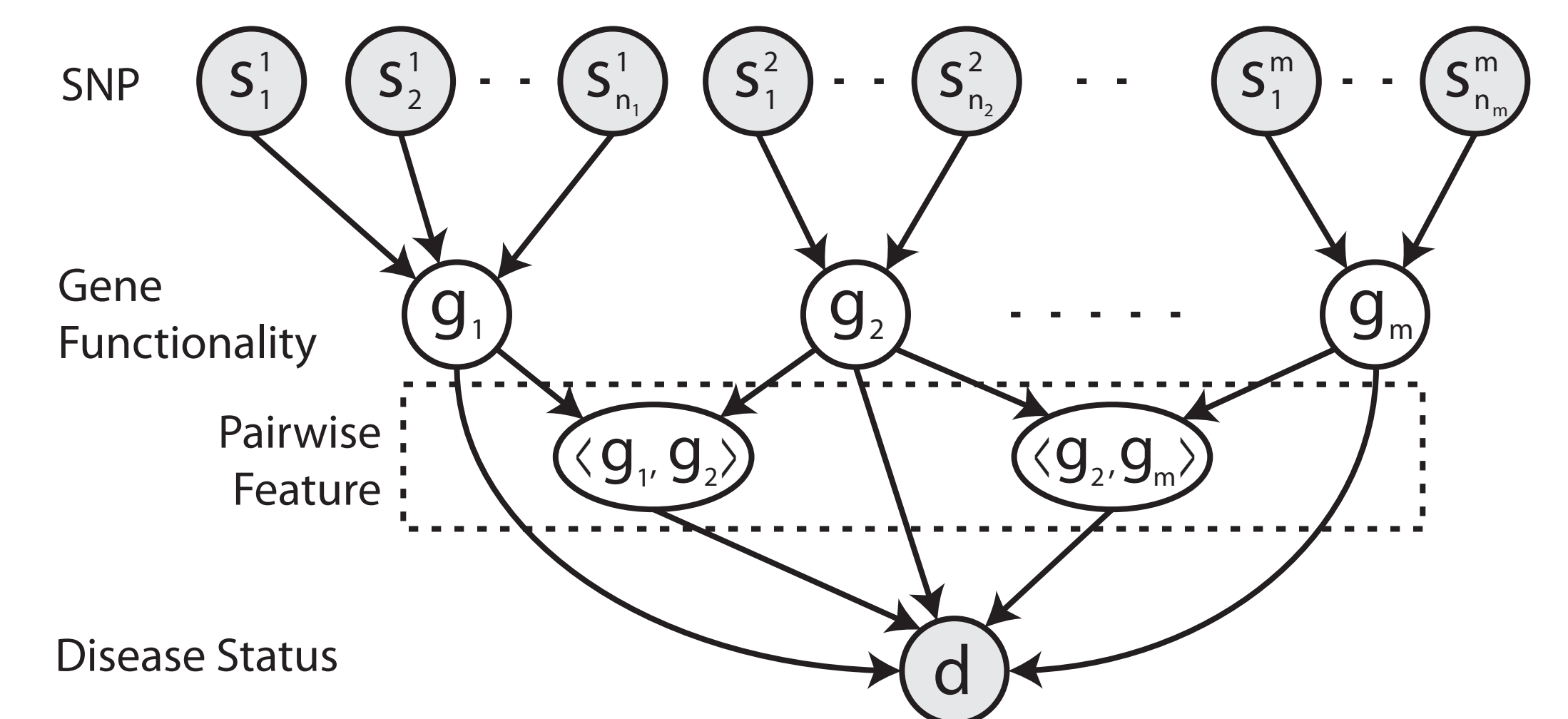
We test every pair for genetic interaction and check the quality of the subset.

**Using gene networks picks out significant pairs with selectivity.**



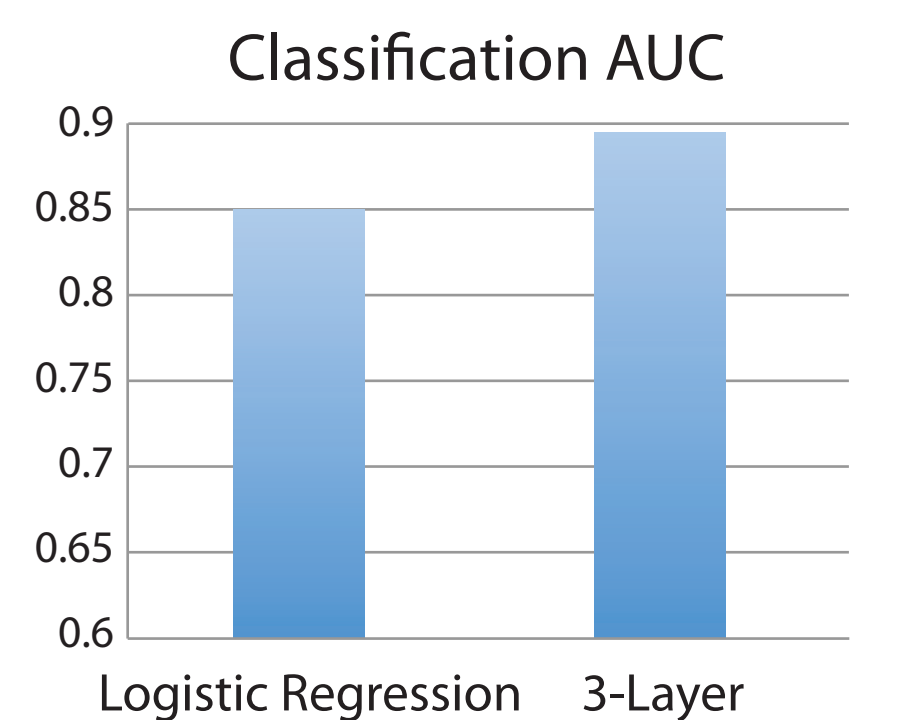
## 3-Layer Model

\* Work in progress



- It is more intuitive to capture genetic interactions as gene pairs than SNP pairs.

3-layer model without pairwise features still shows a **boost in classification performance** compared to logistic regression model when trained on type 1 diabetes.



## Gene Network

There are **too many pairwise features** to consider. We use gene networks to guide our selection process.

### Protein-Protein Interaction (PPI) Network

Edges between genes encoding **proteins that are known to directly interact**.

## Conclusions

- Including pairwise interactions in logistic regression model improves both classification and gene prediction performance.
- Using gene networks can help us find significant pairwise features with selectivity.
- 3-layer model achieves better classification performance and has potential to be a better model for encoding genetic interactions.