

# Introduction Report

9.24.2013

开题报告

Yao Heng  
Rlibs Zhejiang University

# Introduction

- **Gene co-expression**
- **Domain interaction**
- **Shared annotation (GO)**
- **Co-localization**
- **Phylogenetic profile**
- **Homologous interactions**

# Gene Co-expression

- **Gene co-expression**

- 6 database
- Pre-normalization: robust multi-array average method
- 6 features.

Calculate the Pearson's correlation coefficients for each pair of proteins of each chosen database.

# Domain interaction

- **Data Collection**

- Pfam database for domain annotation.

- Other 9 databases for domain interaction data.

- 10 domain interaction features.

- Feature Selection According to each dataset, we counted the number of interacting domains in a pair of proteins as its feature value.

# Shared annotation

- **Shared annotation**

- GO

- 3 features

The fraction of proteins annotated to this shared parent term and all its child terms will be calculated for each aspects(molecular function, biological process, and cellular component)

# Co-localization

- **Co-localization**

- 1 database
- 1 feature

The co-localization feature value for a protein pair  $(A,B)$  was computed as the negative logarithm of the fraction of all proteins presented in the most specific location where both proteins  $A$  and  $B$  were observed to localize.

# Phylogenetic Profile

- **Phylogenetic profile**

- 1 database.
- 2 features

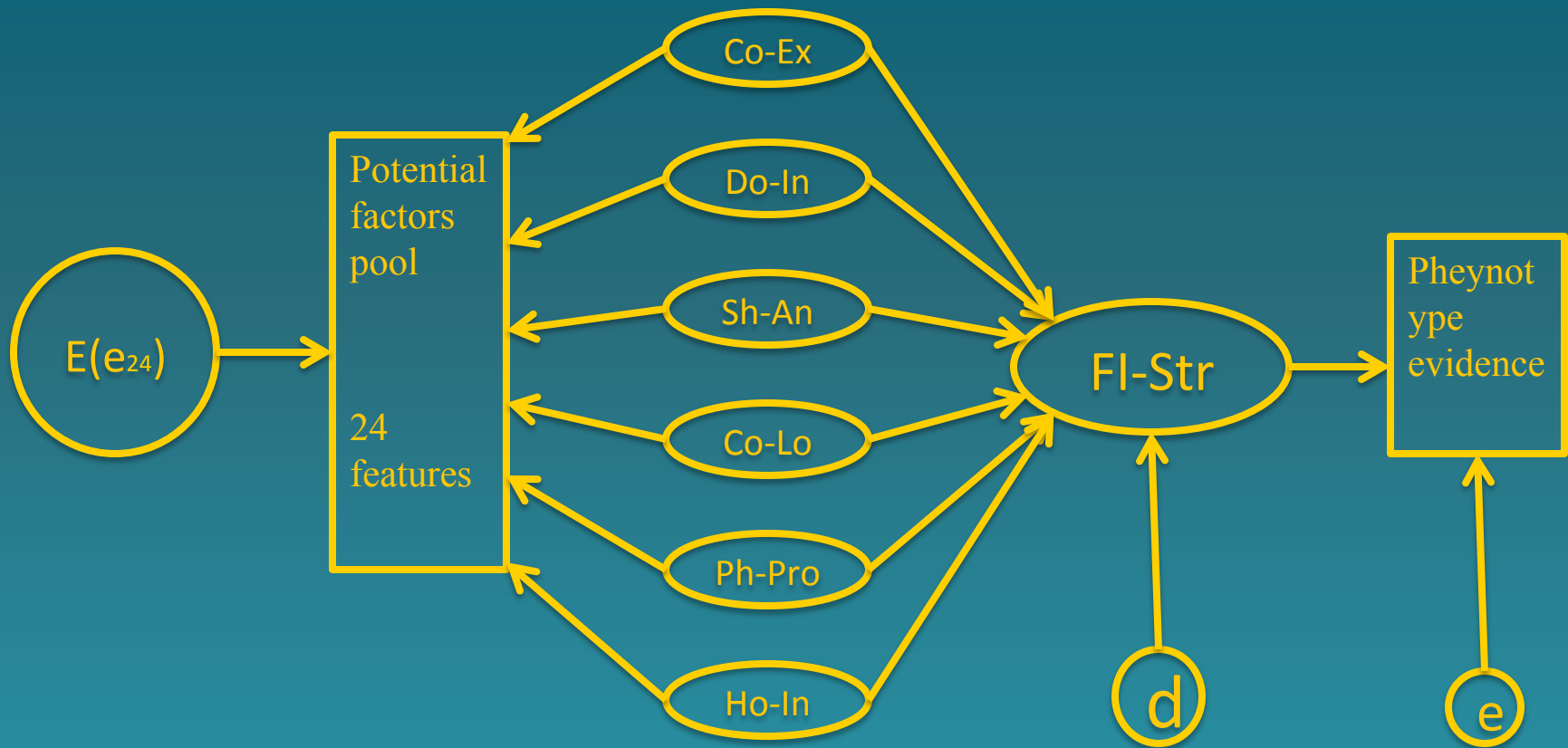
The degree of similarity between two profiles was measured with both the mutual information score and the Pearson's correlation coefficient

# Homologous Interactions

- **Homologous interactions**
  - 4 databases and PSI-blast for pre-processes.
  - 2 features



# SEM Structures



Q&A

Thank You