

Thang Dao & Michael Woo  
Professor Ruby  
CSC325  
13 October 2019

## Milestone 1

### **Part 1:**

#### **Description**

The dataset we chosen was “Transcriptomics in Yeast”. This dataset displays the condition in which the yeast cell was placed in, the cellular component involved, the molecular function and the biological processes of a gene exhibited by the yeast. In simple terms, the cellular component is where the gene is located within a cell, molecular function is what the gene is doing in the cell and biological processes is the genes involved in the process. The author of this dataset was attempting to determine if there is a possibility for improvement in the genetic classification by modeling the data.

#### **1. Conditions Annotations**

ID - Identification of condition

Primary - primary condition

Secondary - secondary condition

Additional Information - any additional information that needed to be noted in conditioning.

#### **2. Yeast-Genome**

ALL\_GENE\_ID

#### **3. Biological Processes**

GeneID - Identification of the gene

Validation -Is the gene verified, uncharacterized or dubious

Localization - where in the yeast cell is the gene being expressed in

#### **4. Cellular Component**

GeneID - Identification of the gene

Validation- Is the gene verified, uncharacterized or dubious

Localization - Primary localization

Loc2- other location

Loc3- other location

Loc4- other location

Loc5-other location

Loc6 -other location

#### **5. Molecular Function**

GeneID - Identification of the gene

Verified - Is the gene verified, uncharacterized or dubious

Molecular Function - what the gene is doing in the yeast cell

#### **6. Expression Values**

X-axis is the condition in which the yeast cell was placed in (92 columns)

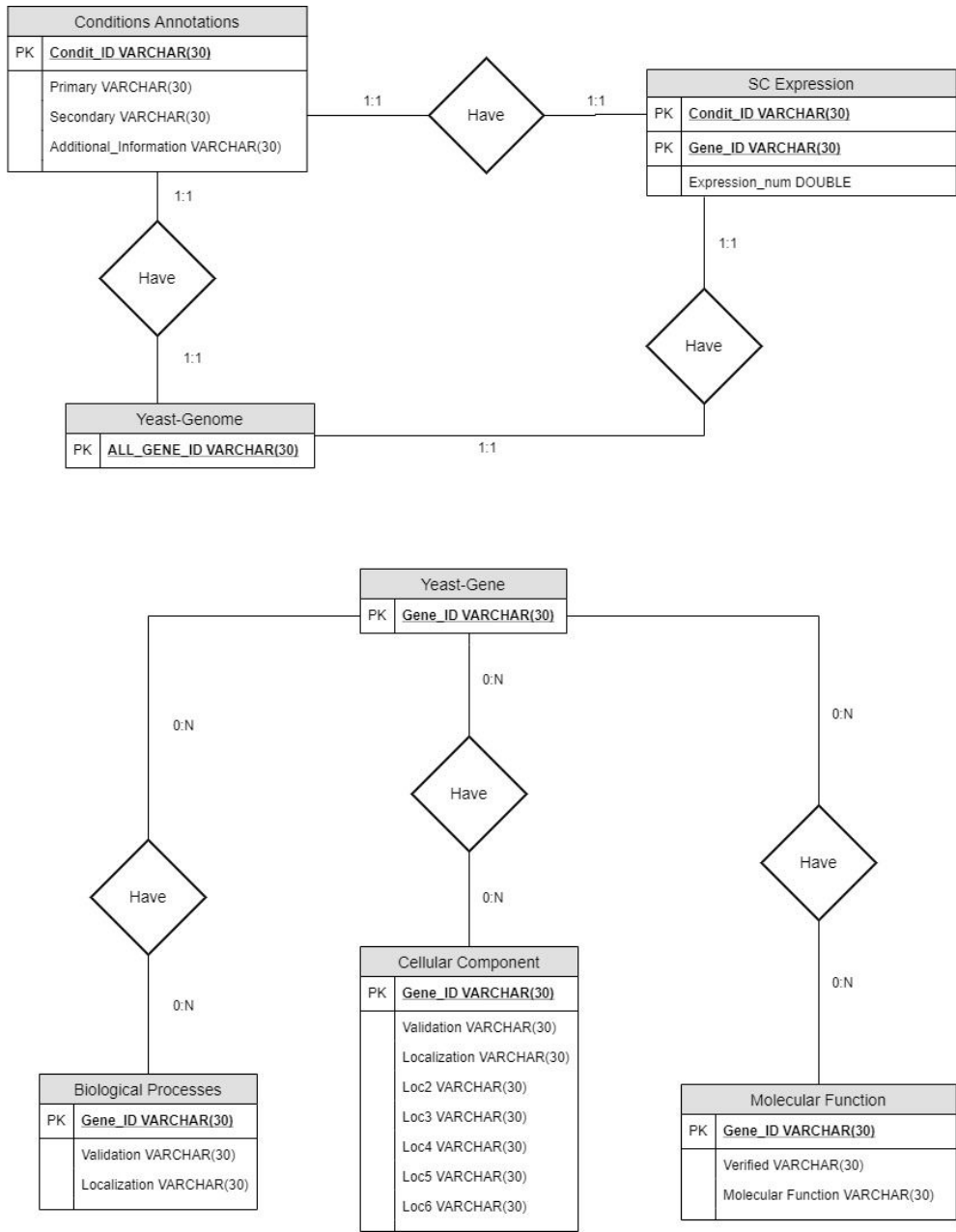
Y-axis is the ID of the gene (6071 genes) per yeast cell (6071 rows)

Note: Evaluation was done on gene level. Each gene has its own row.

Expression levels were normalized by TPM (transcripts per million), as a

default normalization procedure. Raw counts would have been integers, normalized they are floats

ER-Diagram:



The Conditions Annotations must have only one Yeast-Genome: 1:1  
The Yeast-Genome must have only one Condition Annotations. 1:1

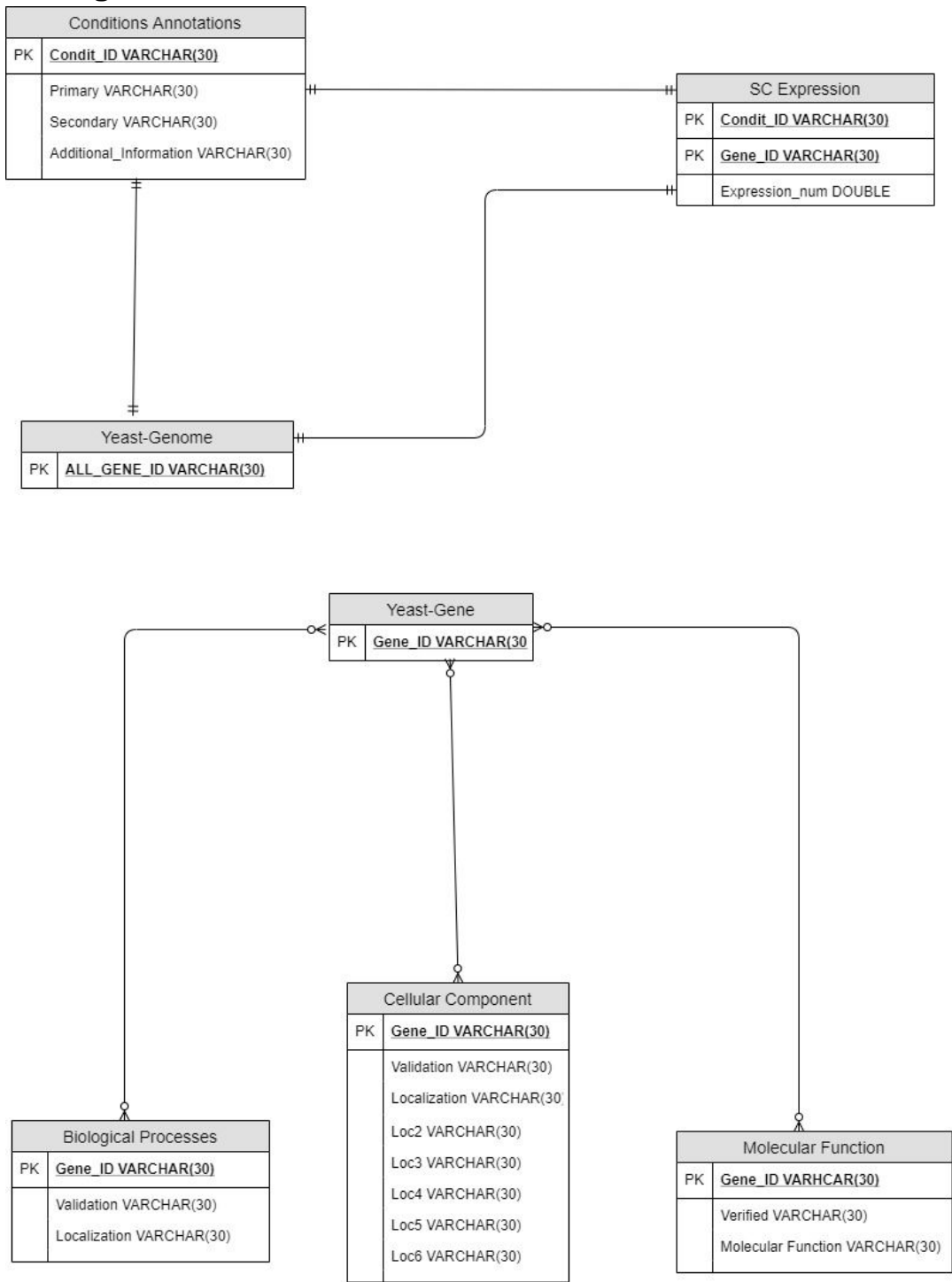
The Condition Annotations and Yeast-Genome must only one SC Expression. 1:1  
The SC Expression must have only one Condition Annotation and Yeast-Genome. 1:1

A Yeast-Gene can have many Biological Process. 0:N  
The Biological Process can have many yeast-genes. 0:N

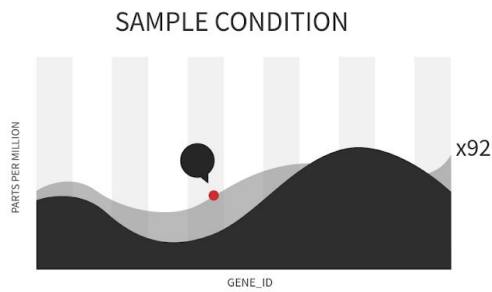
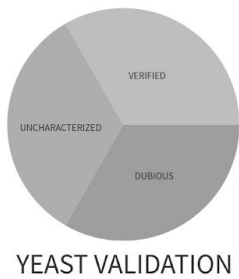
A Yeast-Gene can have many Cellular Components. 0:N  
The Cellular Components can have many Yeast-Genes. 0:N

A Yeast-Gene can have many Molecular Functions. 0:N  
The Molecular Functions can have many Yeast-Genes. 0:N

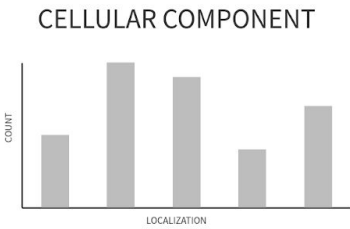
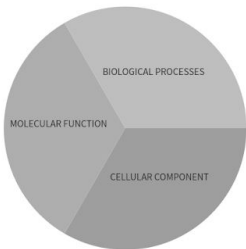
RM Diagram:



Mock Dashboard:



OVERALL YEAST GENOME FUNCTIONALITIES



YEAST GENOME (CC, BP, MF)

