

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-Gene

Biological Processes

GeneID - Identification of the gene

Validation -Is the gene verified, uncharacterized or dubious

Cellular Component

Molecular Function

SC Expression Values

3. Localization_ID

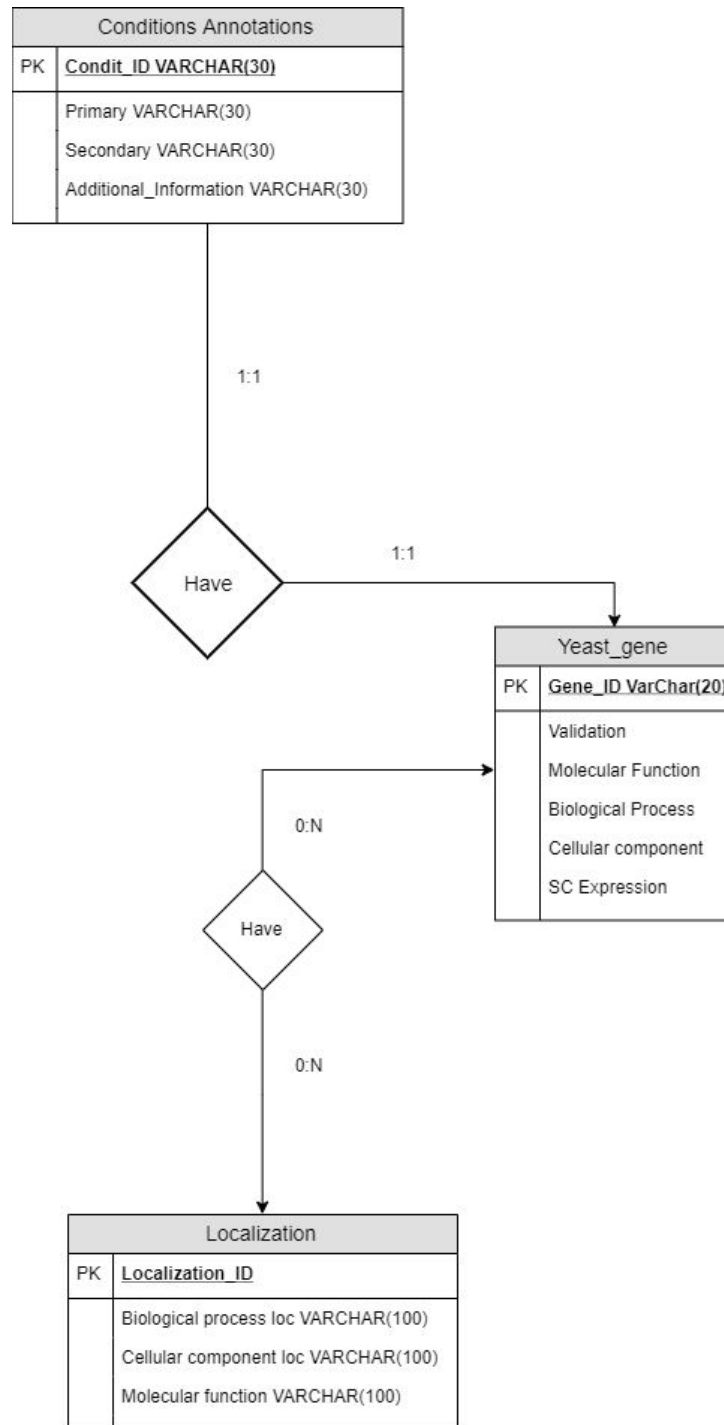
Biological process loc

Cellular component loc

Molecular function loc

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

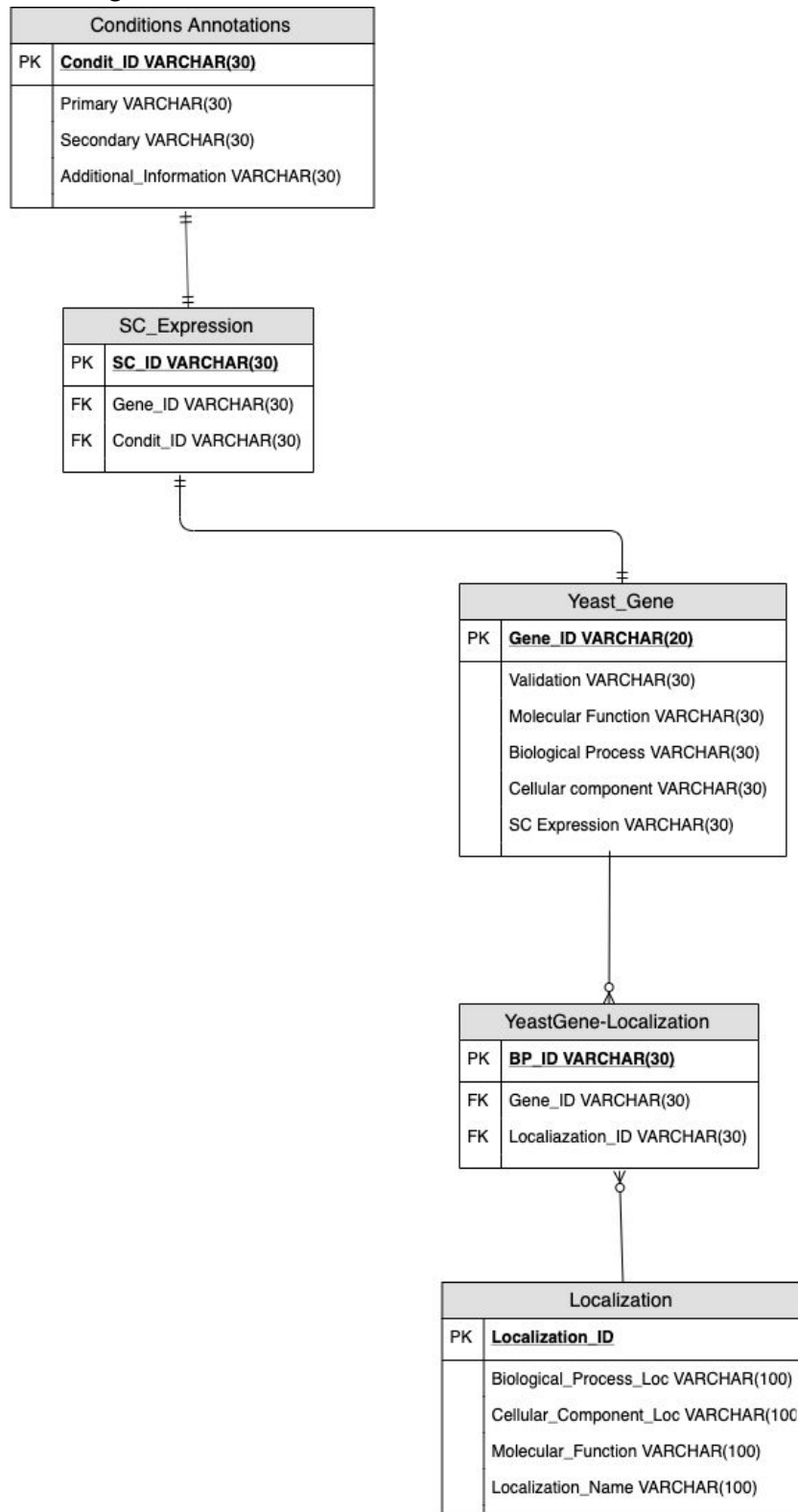
ER-Diagram:



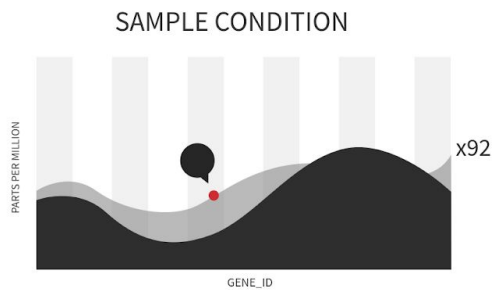
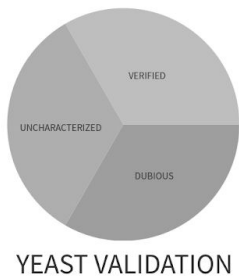
The Conditions Annotations must have only one Yeast-Gene: 1:1
The Yeast-Gene must have only one Condition Annotations. 1:1

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

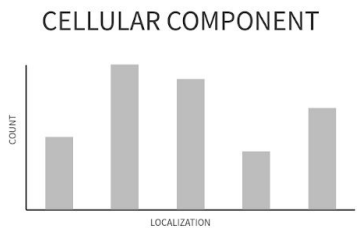
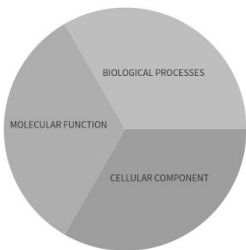
RM Diagram:



Mock Dashboard:



OVERALL YEAST GENOME FUNCTIONALITIES



YEAST GENOME (CC, BP, MF)

