# Function:Troilkatt/FAQ

**From FunctionWiki**

## Contents

- 1 DATA
- 2 BIOLOGY
- 3 ALGORITHM
- 4 EVALUATION
- 5 SYSTEM
- 6 SOFTWARE

## DATA

- Current information in .pcl file is enough? More needed information from yeast to human, for things like tissue, disease, cell line, gene family (later)?
- Big picture about the pipeline, including what it does, how to use it, how to expand on (customize) it if necessary. What about time series data? Is there a way to distinguish them?
- ArrayExpress data, easy to obtain? how difficult to integrate with GEO datasets? how to eliminate or deal with overlap between the two sources?
- has the human dataset gone through SVD? what about yeast datasets? is the SVD done in pipeline?
- Can dataset be broken to column subsets according to conditions? Should we analyze only a column subset of the dataset, since a dataset may include columns representing different disease conditions and tissue information? Is analyzing by column better than analyzing the whole dataset?
- Need filter according to tissue/disease? Especially at the beginning stage, what about narrowing our focus to one tissue one disease?
- Only microarray for the first step? What are the differences we should notice for RNA Seq and others?
- Can we calculate correlations for gene pairs in advance?

To store gene-pair correlations for human we need about 2.5TB (724*30,000*30,000*4). There is a tradeoff between space and time.

## BIOLOGY

- Should we connect gene expression to pathways?
- Should we integrate other experimental results, such as protein-protein interactions, chip-binding experiments, etc?
- Which question does our system answer? Helping biologists confirm current pathway genes, or help biologists discover novel pathway genes? or help biologists navigate the expression diversity of a given gene?

## ALGORITHM

- Some detail descriptions in the current algorithm, like label?
- For future consideration, like scalability (beyond 100 concurrent execution) and larger data processing (TB or more), the current iteration algorithm will be efficient? Can we estimate what efficiency it can process?

## EVALUATION

- Evaluate some typical tissue/disease?
- Is there any cancer curated pathway or process data that we can use as gold standard?

# SYSTEM

# SOFTWARE

- where to get the newest version?

cvs.cs.princeton.edu

- Is there a demo?

http://mmxserver.cs.princeton.edu/piger.html

Retrieved from "http://incendio.princeton.edu/functionwiki/index.php/Function:Troilkatt/FAQ"

- This page was last modified 14:38, 6 April 2010.
- This page has been accessed 24 times.
- Privacy policy
- About FunctionWiki
- Disclaimers