



# Global Gene Map for Cancer Reveals Pathway Hotspots

Pankaj Chopra,<sup>1</sup> Han Jun Shin<sup>2</sup>, Jaewoo Kang<sup>2,3</sup>  
 Department of Computer Science, North Carolina State University<sup>1</sup>,  
 Department of Computer Science and Engineering, Korea University,<sup>2</sup>  
 Department of Biostatistics, Korea University<sup>3</sup>  
 email: pchopra@ncsu.edu



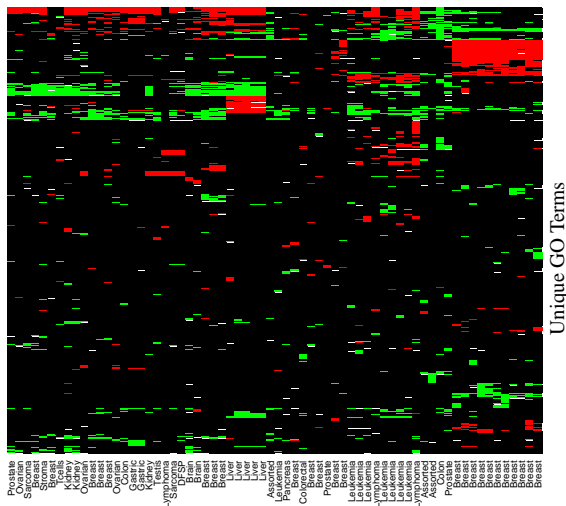
## I. Introduction

There have been a number of papers on meta-analysis of microarray datasets. Most of these papers have focused on genes to arrive at a prognostic signature. However, there has been no concerted meta-analysis of expression data in relation to ontologies and pathways.

## II. Objective

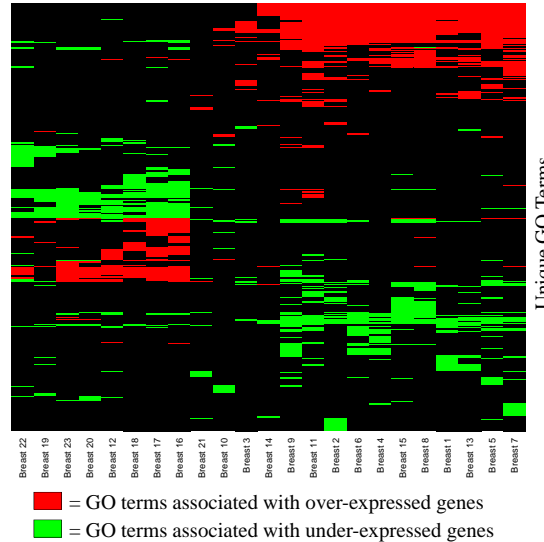
In our meta-analysis, we have studied ontologies (GO), pathways (KEGG) and protein families (PFAM) that are regulated/de-regulated in cancer. An analysis of these hotspots of activation/deactivation will enhance our understanding of different types of cancer.

## Significant GO Terms Associated with Cancer



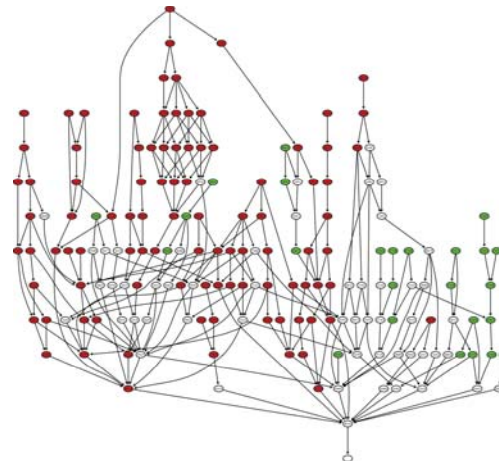
Red = GO terms associated with over-expressed genes  
 Green = GO terms associated with under-expressed genes

## Significant GO Terms Associated with Breast Cancer



Red = GO terms associated with over-expressed genes  
 Green = GO terms associated with under-expressed genes

## Significant GO Terms Mapped to GO Tree



Red = GO terms associated with over-expressed genes  
 Green = GO terms associated with under-expressed genes

## III. Methodology

- Sixty seven microarray datasets, comprising of ~ 4,000 tissue samples, were downloaded from SMD & GEO.
- These tissue samples covered twelve different types of cancer.
- Single channel values were converted to log<sub>2</sub> values.
- A gene was considered to be over (or under) expressed if it had a log<sub>2</sub> value greater than 1 in 80% of tissue samples in that dataset.
- For a set of over (or under) expressed genes, we evaluated if there were any GO/KEGG/PFAM terms that were over represented.
- We further analyzed all breast cancer datasets for pathway hotspots.

**Table 1: Top Significant KEGG Terms for Breast Cancer Data**

S. No.	KEGG ID	KEGG Terms
1	00592	Alpha-Linolenic acid metabolism
2	03050	Proteasome
3	04950	Maturity onset diabetes of the young
4	01040	Polyunsaturated fatty acid biosynthesis
5	00062	Fatty acid elongation in mitochondria

## IV. Results

- Different types of cancer show distinct pathway hotspots, i.e. pathways that are associated with up-regulated genes, and those associated with down-regulated genes.
- Many of the genes tied to these GO/KEGG/PFAM terms are already closely linked to breast cancer.