Combination of Multivariate Analysis and KEGG Data Mapping in GeneSpring 13

Technical Overview

Authors

Nilanjan Guha, Deepak SA,
Arunkumar Padmanaban, Syed Lateef,
and Seetaramanjaneyulu Gundimeda
Agilent Technologies

Arkasubhra Ghosh, Ashwin Mallipatna
Narayana Nethralaya

Introduction

This Technical Overview highlights some of the added features of GeneSpring 13 (GS 13) as applied to retinoblastoma, a childhood eye cancer affecting children usually less than 5 years of age. Cancer development is driven by inactivation of both copies of the RB1 gene in a child's retina. In this study we looked into the transcriptomic profiling of mRNA and miRNA of human retinoblastomas.

Designed specifically for biologists, GS 13 offers an interactive computing environment that promotes investigation and enables understanding of transcriptomics, metabolomics, proteomics, and next generation sequencing data within a biological context. The tool provides a comprehensive suite of statistical tests so that differential analysis can be applied robustly on a variety of experimental designs. Multivariate analysis is available to test the effects and interaction of each factor, as well as the changes in expression across experimental conditions. The software permits access to pathway data from KEGG, WikiPathways, and BioCyc. It also allows researchers to import and view pathways in the BioPAX exchange (Level II and Level III) and GPML format.

We performed pathway analysis of mRNA and miRNA microarrays gene expression data looking for gene enrichment that would enable functional characterization of tumors. We were able to identify many of the key pathways known to be involved in the progression of retinoblastoma including cell cycle pathway. The study also revealed novel miRNAs not previously implicated in the disease. Many of the individual biological responses were corroborated by observations in Rb positive and Rb negative cell lines. Metabolomics experiments were performed in these cell lines along with transcriptomics, followed by integrated analysis of transcriptomics and metabolomics data.

We had tumor samples from enucleated eyes from a total of nine patients and two deceased controls, whose cause of death was not due to any eye related disease. There were five male and four female patients comprised of both unilateral and bilateral cases. The age range of the patients was from 4 months to 2.6 years. Clinical and pathological risks were scored as high or low and a combined risk number was assigned. The numbers ranged from 1–4, with 1 being highest combined risk and 4 the lowest.

Agilent Technologies

Metadata Analysis

The metadata analysis framework is a newly added feature of GeneSpring GX/MPP that allows researchers to visualize tacit dependencies between characteristics of the subjects or samples and their gene. Figure 1 shows a clustered heat map of a subset of differentially regulated cell cycle checkpoint genes. The bar chart indicates the age of the individual samples. As is evident from the gene expression pattern, the expression of many of these genes varies widely from patient to patient. There is no plausible correlation between expression and age of the patients. Adding the risk profile of the individual patients, one can observe the change in the expression pattern with their risk parameter.
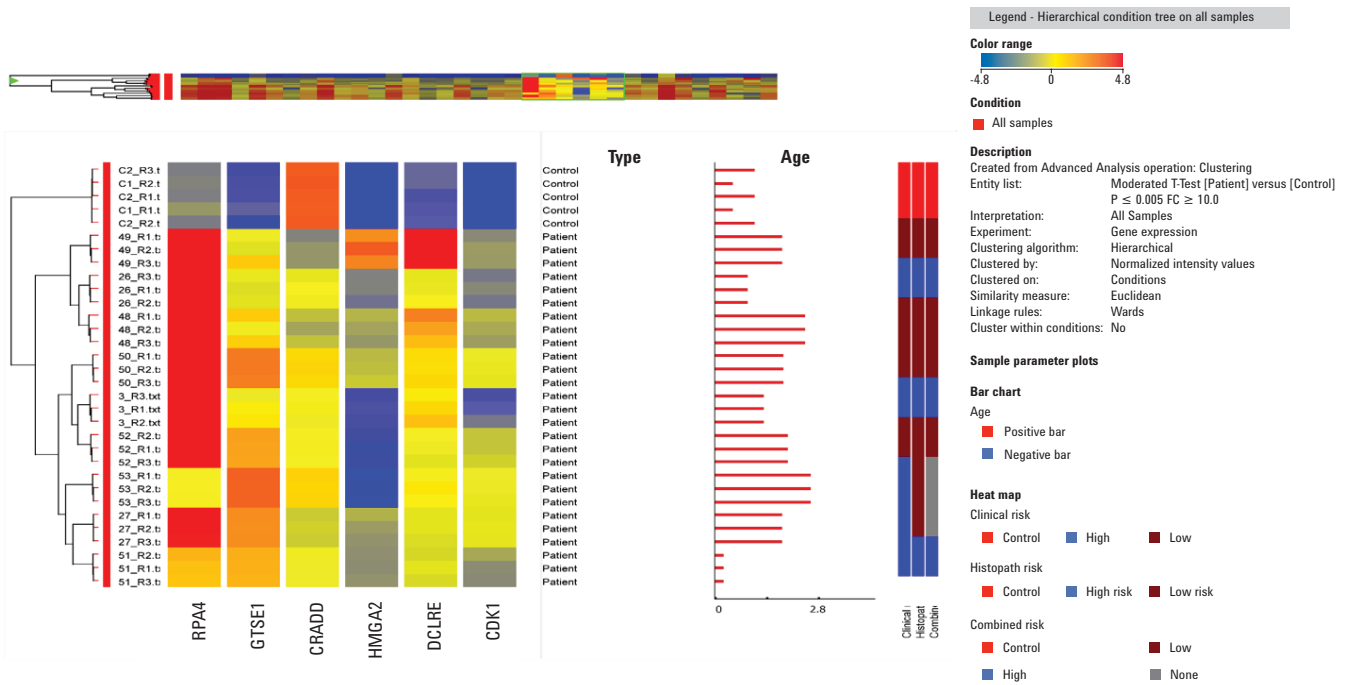

Figure 1. Hierarchical conditioning tree on cell cycle checkpoint genes. The risk classification of the samples are depicted by heat map.

Pathway Architect

Biological processes involve complex interactions between entities that are often best observed and understood in the context of pathways. The pathway architect tool of GS 13 permits the mapping of differential entities into matching pathways from KEGG, WikiPathways, and BioCyc. It also allows researchers to import and view pathways in the BioPAX exchange (Level II and Level III) and GPML format. Figure 2 shows the oxidative phosphorylation pathway to be significantly affected in retinoblastoma patients who have high risk of metastasis. The differentially regulated entities in this pathway are highlighted in yellow, and the histograms near entities indicate the relative combined levels of the gene in controls and patients.
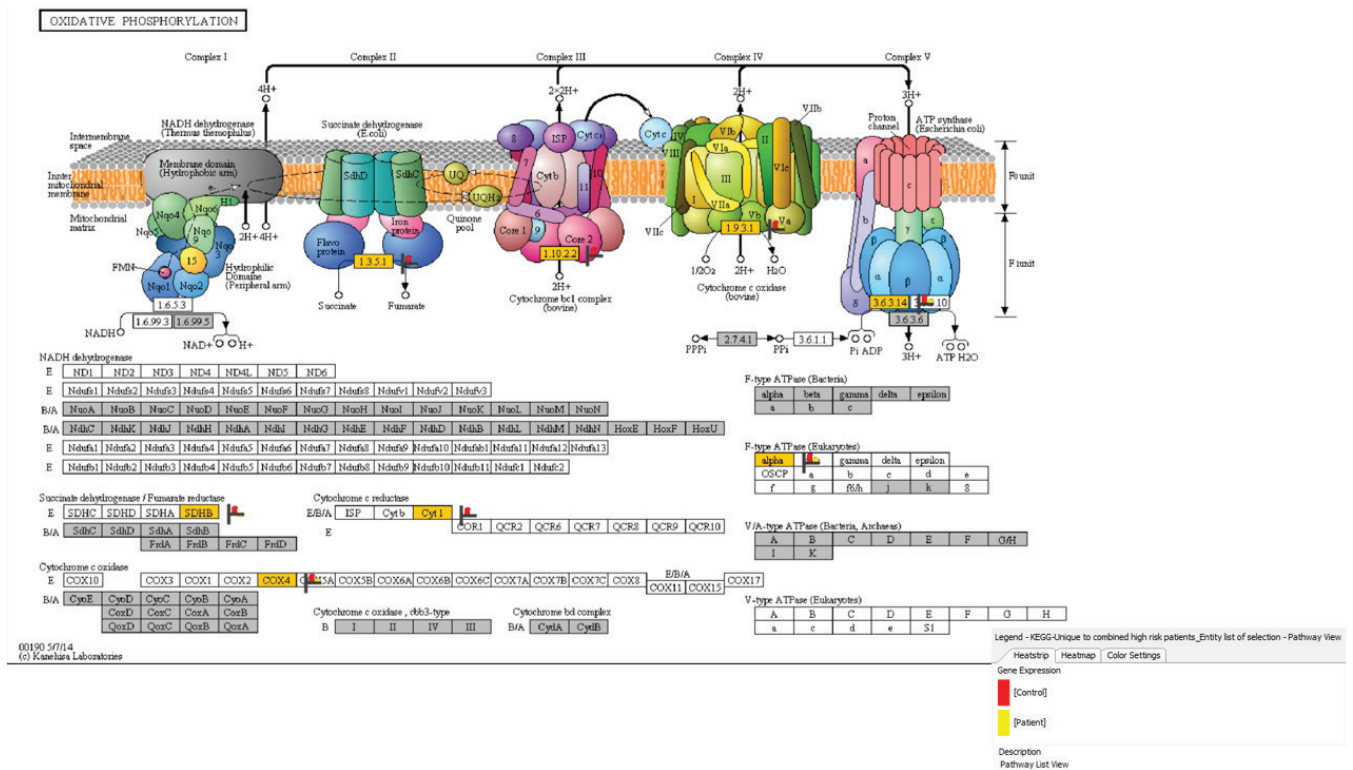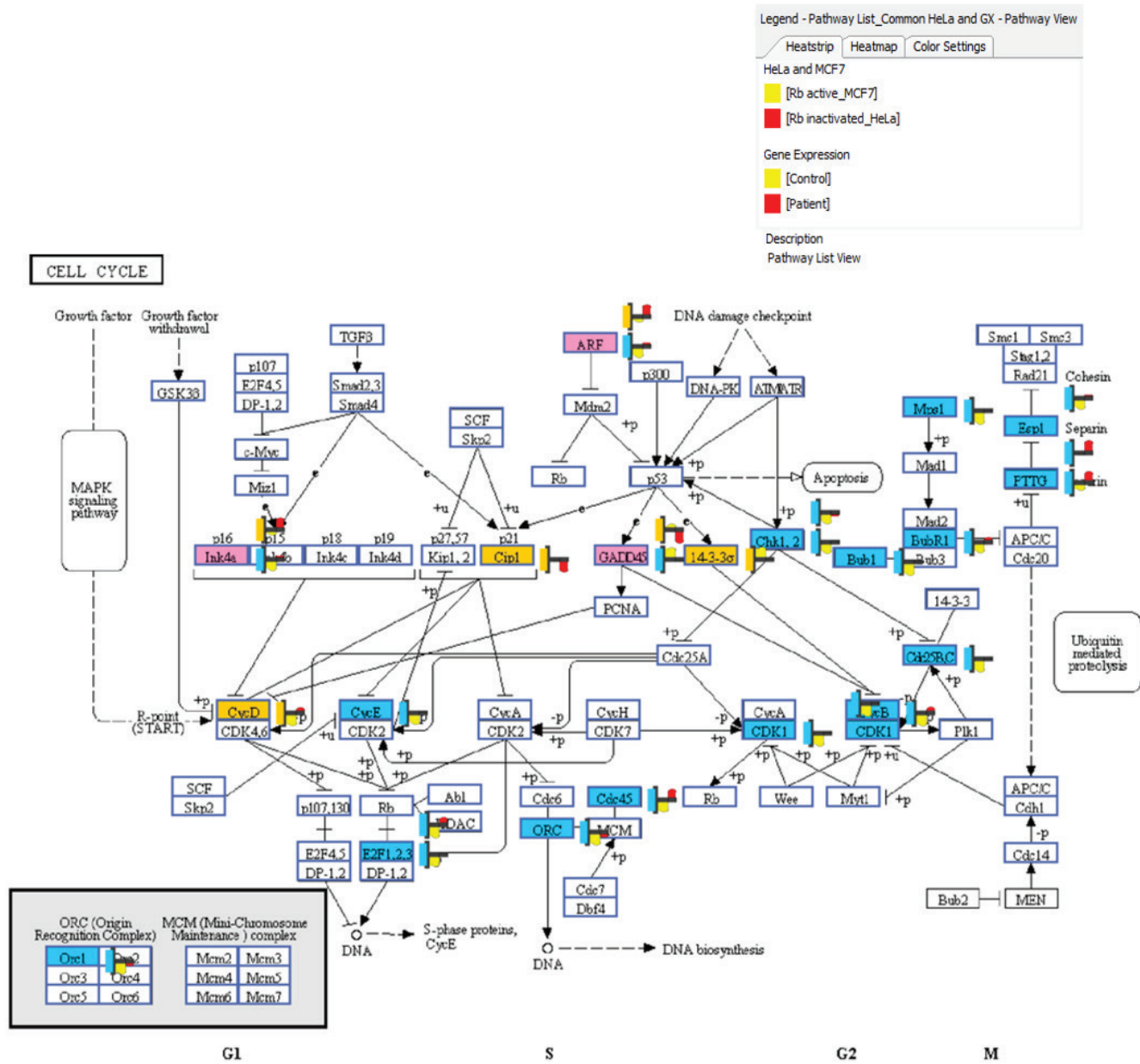

Figure 2. Oxidative phosphorylation pathway is one of the significant pathways affected in high risk retinoblastoma patients. The affected entities of the pathway are highlighted in yellow.

Analysis of Multi-omics Data

To gain a better understanding of the interrelationship of changes in expression between each biochemical entity in a given pathway, GS 13 allows multi-omic analysis, and enables co-visualization of the data. Figure 3 shows how cell cycle pathway entities behave in cell lines versus human samples. MCF7 is Rb positive and HeLa is Rb negative. The pink colored boxes indicate that the genes are differentially expressed in both cell lines and human samples. The histograms next to the genes indicate their expression levels in both experiments.

Figure 3. Multi-omics analysis with the gene expression profiles of patient samples versus cell lines shows overlap of three differentially regulated genes (pink boxes) in the cell cycle pathway.

Correlation Analysis

Correlation analysis can be used as a method of showing relationships among the expression levels of entities over a series of samples, or the expression levels of samples. GS 13 permits correlation analysis on either entities or samples. In the resulting heat map, entities are arranged according to their correlation coefficient values.

Hierarchical clustering in GeneSpring GX 13 allows alignment and observation of individual entities. The clustered heat map in Figure 4A has been generated using 18 differentially expressed miRNAs. This tool is valuable in determining correlation of expression among individual entities.

GS 13 also permits the generation of correlation coefficient scatter plots. This provides a graphical representation of the empirical data values used to calculate the correlation coefficient between an entity pair. Figure 4B shows an example of a scatter plot between two differentially regulated miRNAs. The regression fit and equation in the plot display the direction and strength of the dependency between the pair of entities.
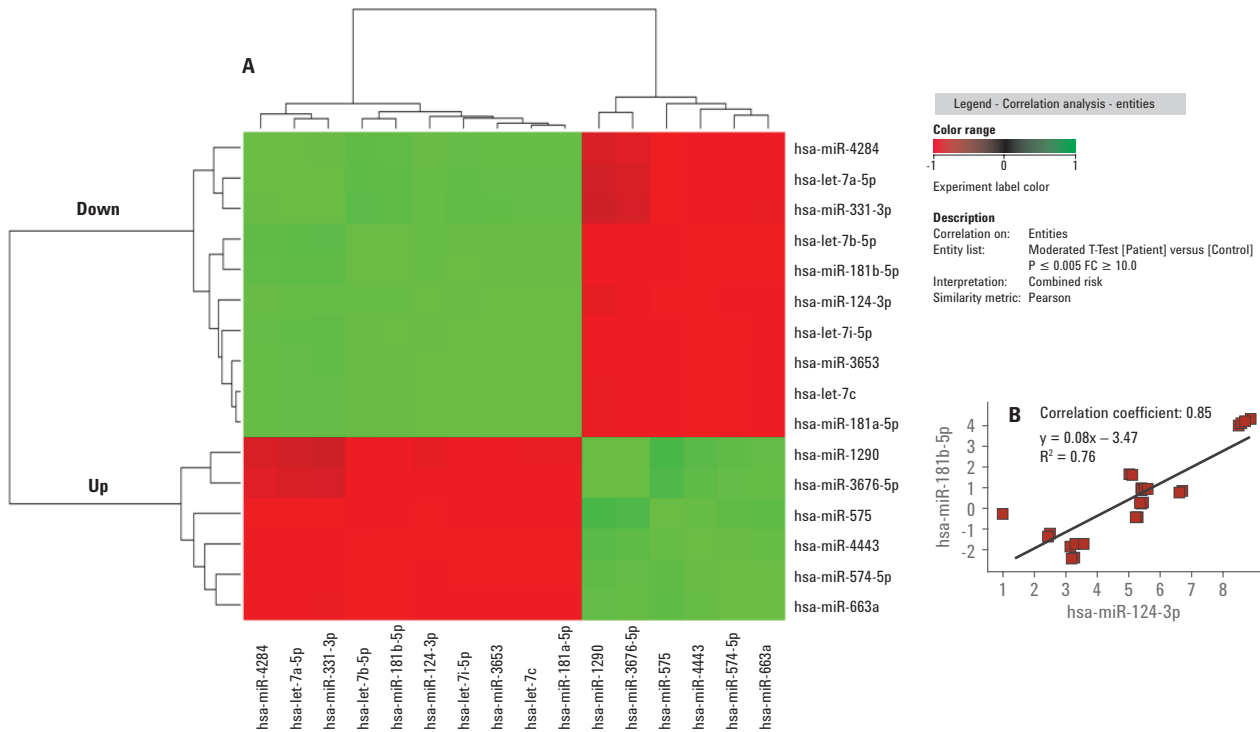

Figure 4. A) Correlation analysis among the 18 differentially regulated miRNAs. B) Scatter plot of two differentially regulated miRNAs.

GS 13 also enables correlation analysis among the entities of two experiments performed by two different technologies including mRNA/miRNA pairing. Figure 5A shows an example of a correlation plot between differentially expressed miRNAs and differentially expressed mRNAs that overlapped with miRNA target list. The purpose of correlation is to look for associations of entities or samples, which can be observed as heat maps. This helps in discovering patterns of variability and co-dependence of entity expression data. One can also see the dependency between a pair comprising miRNA and mRNA with the help of correlation coefficient scatter plots (Figure 5B)
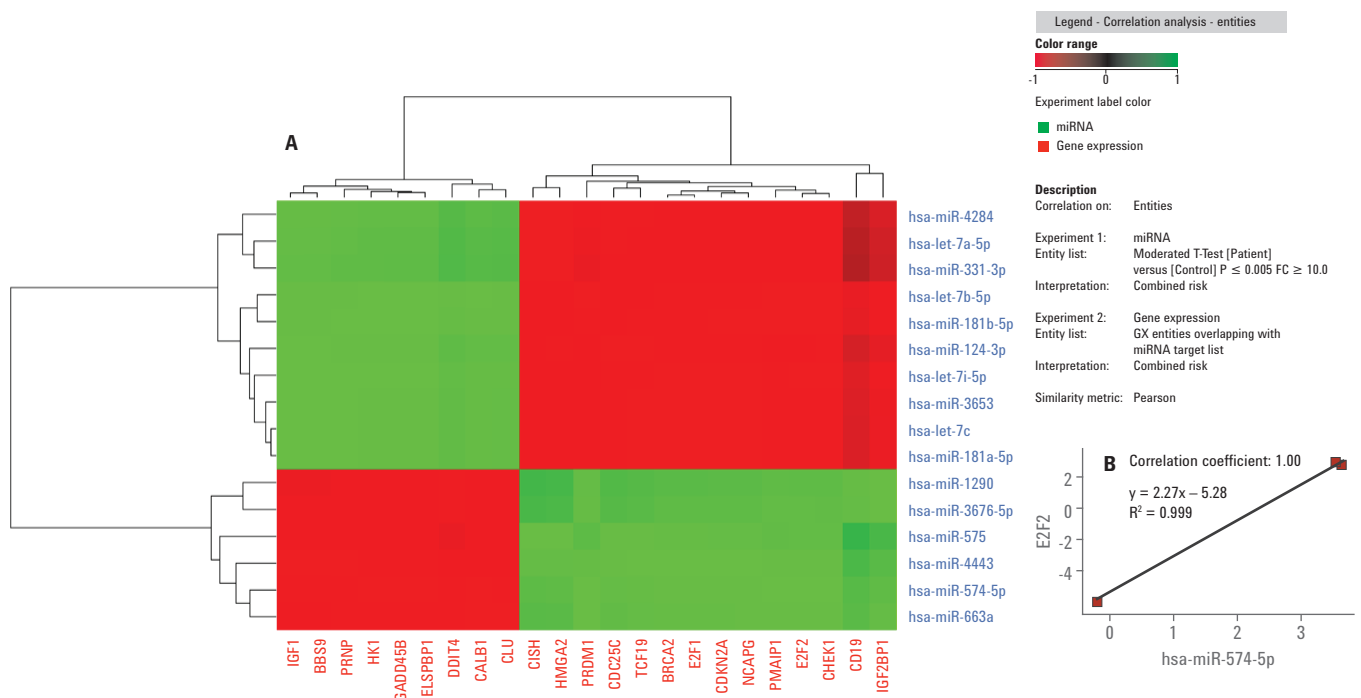

Figure 5. A) Correlation analysis showing mRNA/miRNA co-regulation pattern. The mRNAs from mRNA microarray dataset overlapping with miRNA target list is plotted with the differential miRNA list. B) Scatter plot showing the relationship between a E2F2 and hsa-miR-574-5p.

Integrated Analysis

Since biological processes involve complex interactions between entities that are often best observed and understood as pathways, this type of analysis is a critical component of multi-omics data integration. We performed a multi-omic pilot experiment using HeLa and MCF7 integrating mRNA transcriptomics with metabolomics data. Figure 6 shows glutathione metabolism as one of key common pathways. The levels of glutathione (GSH) are significantly down-regulated in HeLa cells. GSH plays an important role in a number of cellular processes, including cell differentiation, proliferation, and apoptosis. Disturbances in GSH homeostasis are involved in the etiology and progression of many human diseases including cancer. The level of expression of glutathione peroxidases, mainly GPX2 and GPX3, that generate glutathione disulfide during the reduction of peroxides and organic hydroperoxides are also significantly low in HeLa as compared to MCF7. This may be one of the contributing factors leading to low GSH.
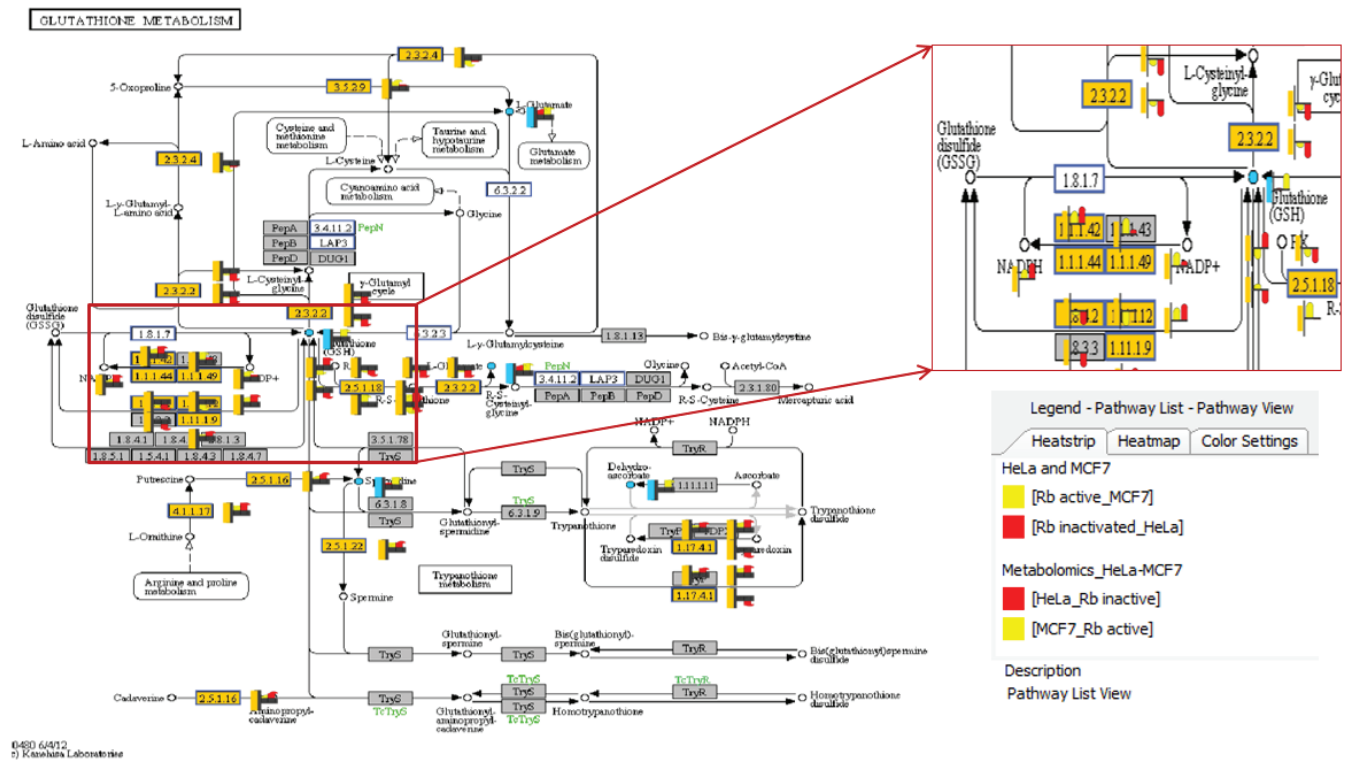

Figure 6. Combined analysis of transcriptomics and metabolomics data show the glutathione metabolism pathway. The differentially regulated genes are marked yellow and the metabolites are marked as blue circles. The corresponding levels of genes (yellow bar) and metabolites (blue bar) are next to the differential genes and metabolites.

Conclusion

The newly added features of GeneSpring 13 enable researchers to perform discovery experiments by:

- Correlating gene expression pattern with metadata

- Looking at the dependencies of individual entities within the same (mRNA-mRNA) or cross technology (mRNA-miRNA)

- Co-analyzing data from complementary technologies such as genomics, transcriptomics, and metabolomics

The software permits new biological insights, leading to a comprehensive understanding of the biological system.

www.agilent.com/chem

Agilent Technologies