**Fig. 4.** Global connection map of KEGG with mapped coefficients of the decision function obtained by applying a customary linear SVM (left) and using high-frequency eigenvalue attenuation (80% of high-frequency eigenvalues have been removed) (right). Spectral filtering divided the whole network into modules having coordinated responses, with the activation of low-frequency eigen modes being determined by microarray data. Positive coefficients are marked in red, negative coefficients are in green, and the intensity of the colour reflects the absolute values of the coefficients. Rhombuses highlight proteins participating in the Glycolysis/Gluconeogenesis KEGG pathway. Some other parts of the network are annotated including big highly connected clusters corresponding to protein kinases and DNA and RNA polymerase sub-units.

**5 DISCUSSION**

Our algorithm groups predictor variables according to highly connected "modules" of the global gene network. We assume that the genes within a tightly connected network module are likely to contribute similarly to the prediction function because of the interactions between the genes. This motivates the filtering of gene expression profile to remove the noisy high-frequency modes of the network.

Such grouping of variables is a very useful feature of the resulting classification function because the function becomes meaningful for interpreting and suggesting biological factors that cause the class separation. This allows classifications based on functions, pathways and network modules rather than on individual genes. This can lead to a more robust behaviour of the classifier in independent tests and to equal if not better classification results. Our results on the dataset we analysed shows only a slight improvement, although this may be due to its limited size. Therefore we are currently extending our work to larger data sets.

An important remark to bear in mind when analyzing pictures such as fig.4 and 5 is that the colors represent the weights of the classifier, and not gene expression levels. There is of course a relationship between the classifier weights and the typical expression levels of genes in irradiated and non-irradiated samples: irradiated samples tend to have expression profiles positively correlated with the classifier, while non-irradiated samples tend to be negatively correlated. Roughly speaking, the classifier tries to find a smooth function that has this property. If more samples were available, better non-smooth classifier might be learned by the algorithm, but constraining the smoothness of the classifier is a way to reduce the complexity of the learning problem when a limited number of samples are available. This means in particular that the pictures provide virtually no information regarding the over-