README for files contained within UCSC_SUBTYPE.zip

`README.doc, README.pdf`	This file
`BASAL_SUBTYPE.cys`	Basal subtype phenotype
`CLAUDIN_LOW_SUBTYPE.cys`	Claudin Low subytpe phenotype
`ERBB2_AMP_SUBTYPE.cys`	Erbb2 Amplified subtype phenotype
`LUMINAL_SUBTYPE.cys`	Luminal subtype phenotype

These files are *phenotype integrated pathway activity feature graphs* (PIPAF) graphs contained in their respective files. These derived from the cell line's expression and copy number data analyzed into global pathways by the UCSC *PARADIGM* system [Vaske]. The result was then processed through our pipeline that uses Stanford *Significance Analysis of Microarrays* (SAM) system [Efron] and then visualized into graphs using the Cytoscape software system [Shannon]. The Cytoscape files have been tested with Cytoscape version 2.8.0 (current version as of February 9, 2011. You may download Cytoscape for many popular computer platforms from:

http://www.cytoscape.org/download.php

NOTE ON GRAPHICAL PRESENTATION

SHAPE
Each file contains two networks: with and without drug information. The circular nodes are gene features, which may include non-coding RNA genes. Square nodes are abstract processes. Triangular nodes are drugs. Hexagonal nodes are complexes.

LABEL
After Cytoscape loads the file, all nodes are initially unlabeled. Please click on the menu item **View**, and then click on **Show Graphics Details** to see the labels. Only the gene features (circular nodes) are labeled in the interest of graphical presentation. However, you may click on any node and see the name and the score in the Cytoscape data panel.

LINE
Dashed lines are protein links. Solids lines are transcriptional regulators. Activating ends in an arrow, deactivating ends in a T shape.

COLOR
Red nodes are features that are statistically more likely to occur in the phenotype, Blue nodes are gene features that are more likely to be expressed occur in cell lines that are not in the phenotype. The color intensity and size of the complex and gene nodes are proportional to these SAM scores [Efron]. The score for drug nodes are arbitrarily set for graphical presentation purposes.

References

Efron and Tibshirani. *On testing the significance of sets of genes*. Annals of Applied Statistics (2007) vol. 1 (1) pp. 107-129.

Shannon. *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks*. Genome Research (2003) vol. 13 (11) pp. 2498-2504.

Vaske et al. *Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM*. BIOINFORMATICS (2010) pp. 1-9.