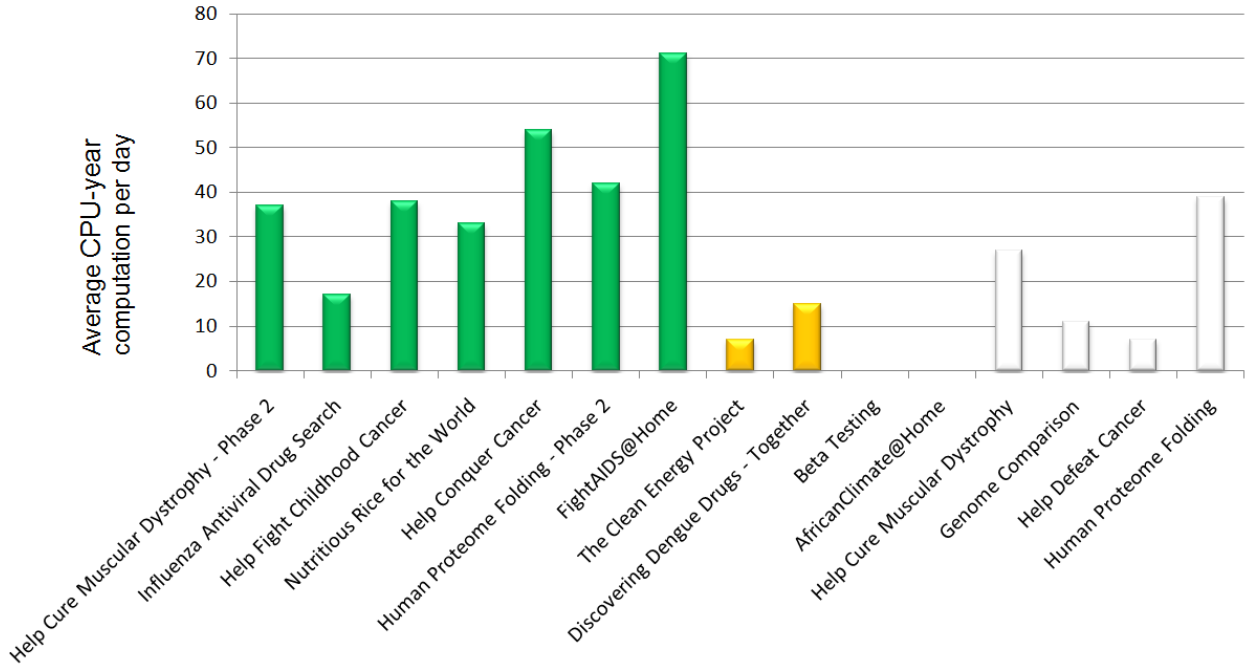
Update October 2009

Thank you for your continuing support of the Help Conquer Cancer project (HCC). We appreciate all the computing resources you donate to this and other projects at WCG.

Since the launch of HCC project in November 2007, WCG members contributed over 39,287 years of run time, averaging more than 54 years of computation per day. To date 52,902,799 results were returned (*Statistics Last Updated:* 10/26/09 12:06:08 (UTC)). We are thrilled and thankful to be the second "most CPU/day computation" project on WCG (in the graph below, green bars show current projects, orange bars represent intermittent projects, and white bars correspond to completed projects):



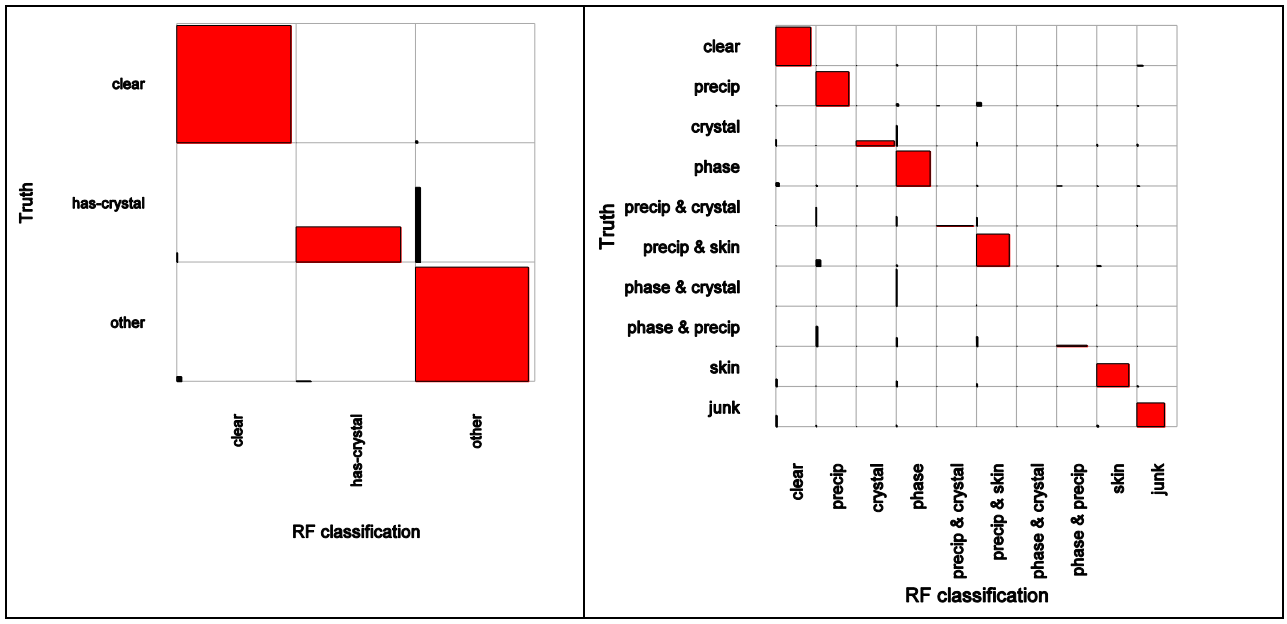
Summary

Since the project started, 32% of HCC work units have been processed on the WCG. Subset of results from new analyses are strengthening trends in data we have been following for a while now, while new trends are emerging due to increased diversity of proteins that have been processed. Some of these results have recently been summarized and submitted to the *Journal of Structural and Functional Genomics*.

We have also made a significant progress on the “back end” analysis – updating our NAViGaTOR tool for integrative analysis of protein interaction networks, and text mining to help us annotate our cancer targets from Pubmed. Both manuscripts are now published in *Bioinformatics*, and available online with free access[1] (note that these manuscript versions will still be edited).

Results

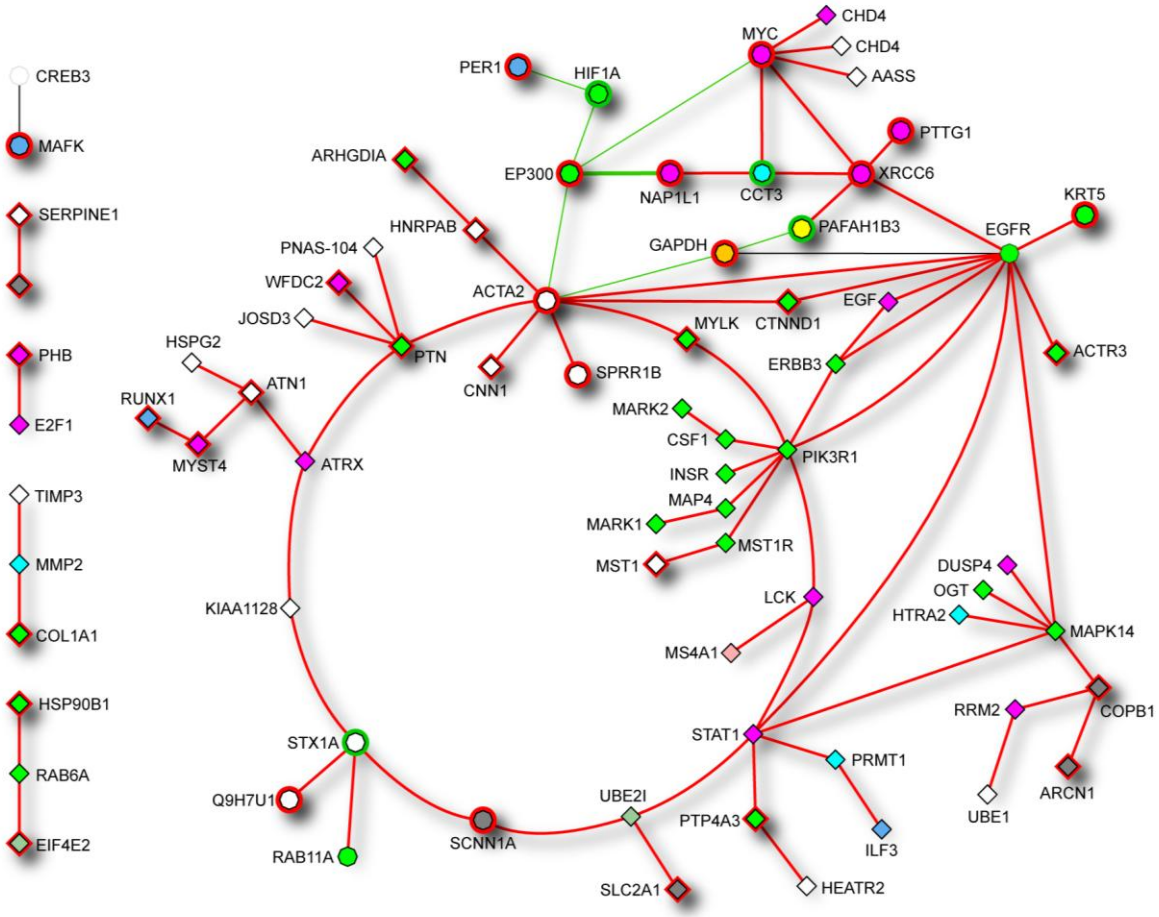
Data from WCG continue to be analyzed on our Linux cluster. Christian has enhanced performance of our image classification system by using Random Forest algorithm with improved training, and below we show new results for both 3- and 10-way classification. Impressive precision and recall especially for clear and other category makes the computer system equal or better to the human expert, who cannot be available 24/7 at 100% of her/his best performance.



Preliminary analysis using the probabilistic image scoring has been completed for 2,598 proteins. If we set a 95% confidence threshold, from the set of 2,598*1,536 = 3,812,352 images, we see 922 images >=95% likely containing crystal. After considering multiple time-points, these represent 830 unique protein/cocktail experiments from 171 unique proteins. These may provide very valuable leads for new structure-determination experiments.

[1] http://bioinformatics.oxfordjournals.org/cgi/reprint/btp602? ijkey=it0l1m7pDMy9pOP&keytype=ref
http://bioinformatics.oxfordjournals.org/cgi/reprint/btp595? ijkey=XmbJRIjzlxvPAWp&keytype=ref

We continue to analyze lung cancer data, with the goal to identify small sets of genes that can be used for accurate diagnosis and prediction of treatment outcome. These proteins then form subset of our “targets” at HCC. Preliminary results have been summarized in Zhu et al., *Clin Lung Cancer,* 2009. Combined network integrates results across 26 studies in lung cancer that identified small prognostic gene signatures (visualized in NAViGaTOR, (Brown et al., *Bioinformatics*, 2009); http://ophid.utoronto.ca/navigator), including our three and six gene signatures from (Lau et al., *JCO,* 2007) and (Boutros et al., *PNAS,* 2009). These three and six gene signatures (highlighted with green circles around the nodes) are currently being evaluated for prognostic value on an independent set of samples. Further, permutation analysis on the same dataset identified thousands of six gene signatures that performed equally well as the original six gene signature. Sixteen genes (highlighted in red circle around the node) were present in almost half of those six gene signatures, yet those 16 genes were not the most differential (Boutros et al., *PNAS,* 2009); and 11 of them mapped to this small network. Considering the length of the shortest path between individual members of 3, 6 and 16 gene signatures (i.e., their proximity on the network) suggests that they are involved in similar pathways and functions in lung cancer.



Recently published work

Crystallography

1. Cumbaa, C. A., Jurisica, I. protein crystallization analysis on the World Community Grid. Under review in *Journal of Structural and Functional Genomics*.

Cancer research

1. Agarwal R., Jurisica, I., Cheng K.W., Mills G.B. The emerging role of the Rab25 small GTPase in cancer, *Traffic*, 2009. E-Pub July 23, 2009. In Press.

2. Mills, G. B., Jurisica, I., Yarden, Y., Norman, J. C. Genomic amplicons target vesicle recycling in breast cancer. *J Clin Invest,* 119(8): 2123-7, 2009.

3. Zhu, C.Q., Pintilie, M., John, T., Strumpf, D., Shepherd, F.A., Der, S.D., Jurisica, I., Tsao, M.-S., Understanding Prognostic Gene Expression Signatures in Lung Cancer, *Clin Lung Cancer*, 10(5): 331-340, 2009.

4. Hui, A.B. Y., Shi, W., Boutros, P.C., Miller, N., Pintilie, M., Fyles, T., McCready, D., Wong, D., Gerster, K., Waldron, L., Jurisica, I., Penn, L.Z., Liu, F.F. Robust global micro-RNA profiling with formalin-fixed paraffin-embedded breast cancer tissues. *Lab Invest*, 89(5):597-606, 2009.

Tools and resources

1. Niu, Y., Otasek, D., Jurisica, I. Evaluation of linguistic features useful in extraction of interactions from PubMed; Application to annotating known, high-throughput and predicted interactions in I2D. *Bioinformatics*, 2009. doi: 10.1093/bioinformatics/btp602.

2. Brown, K.R., Otasek, D., Ali, M., McGuffin, M., Xie, W., Devani, B., van Toch, I. L., Jurisica, I. NAViGaTOR: Network analysis, visualization & graphing Toronto. *Bioinformatics*, 2009. doi: 10.1093/bioinformatics/btp595.

3. McGuffin, M, and Jurisica, I. Interaction techniques for selecting and manipulating subgraphs in network visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6): 937-944, 2009. [Honorable Mention at InfoVis'09].

Recent (selected) presentations

- Cumbaa, C. A., Jurisica, I. Crystallization image analysis on the World Community Grid. *Annual Meeting of the American Crystallographic Association*,

Toronto, ON, July 25-30, 2009.

- Jurisica, I. Rational prediction and analysis of protein interactions; Disease-specific interaction prediction, *HUPO PSI Spring Meeting*, Turku, Finland, April 25-30, 2009.

Thank you,

C. A. Cumbaa and I. Jurisica