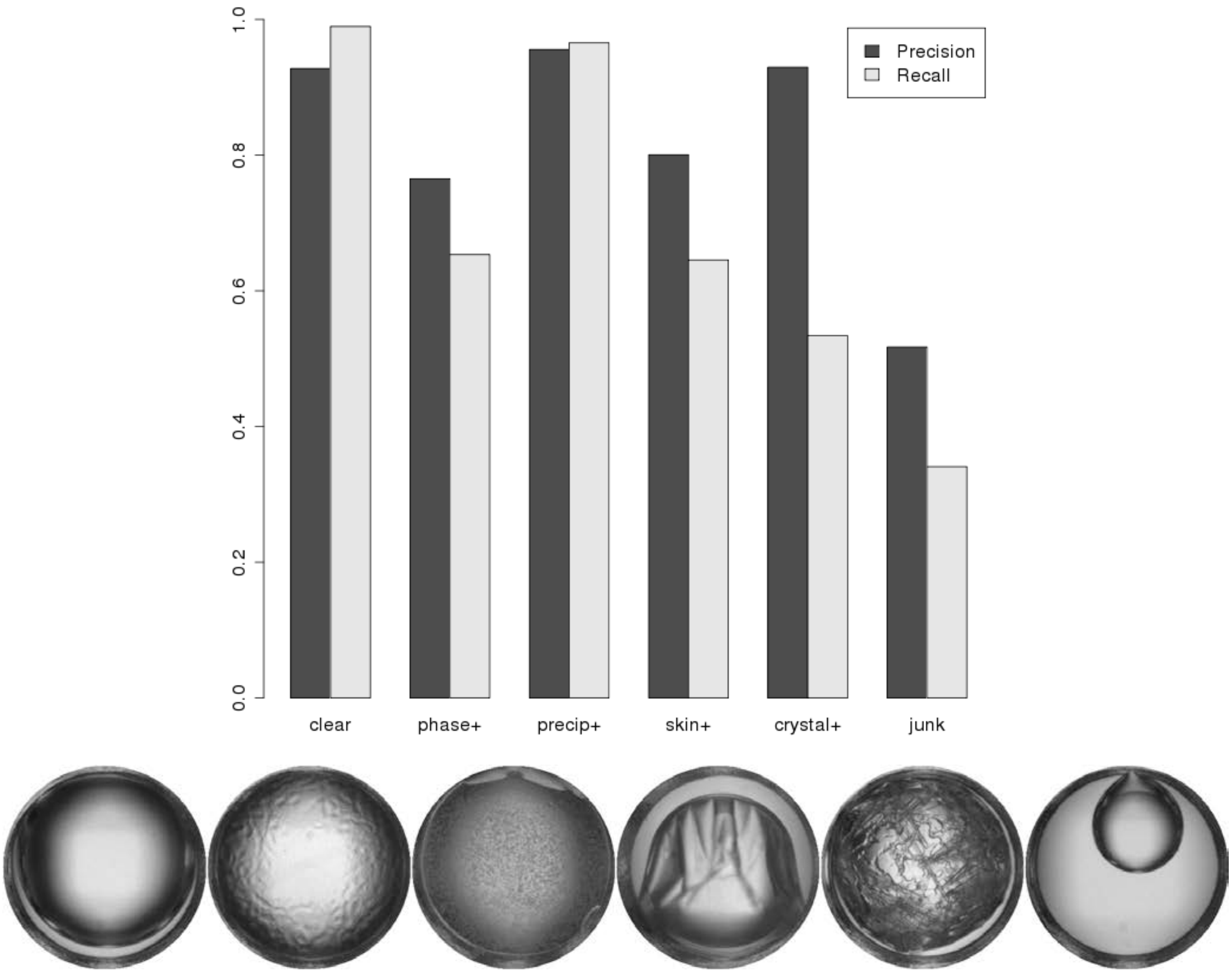
Help Conquer Cancer April 2013 Update

Thanks to overwhelming support by World Community Grid members, and their contributions of CPU and GPU cycles, we are nearing the end of the grid-computation phase of the Help Conquer Cancer project. April 4, 2013, marked the completion of 300 million images.

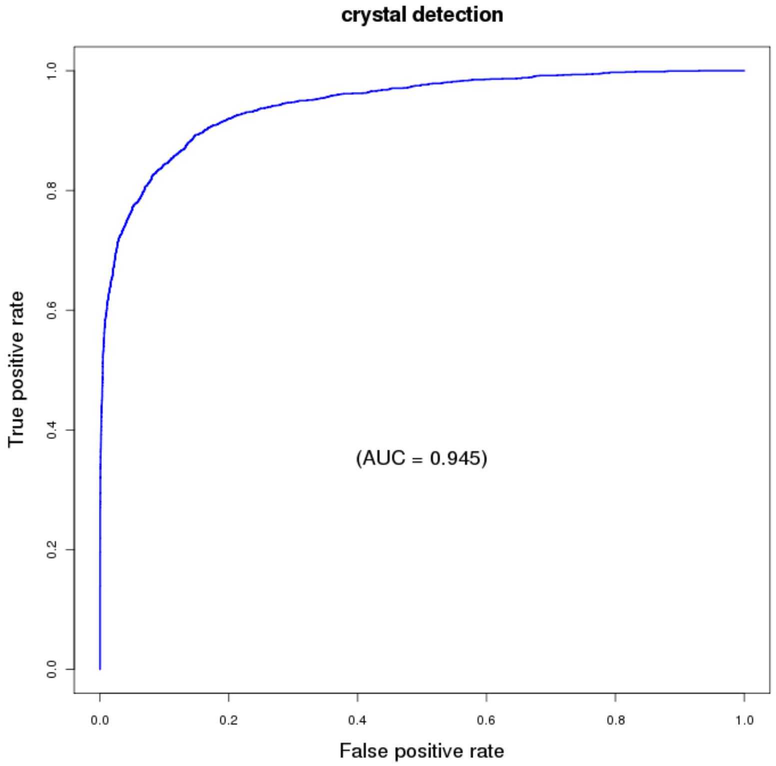
Thank you all for continued support. It is exciting to finalize this long project and start getting results we have been waiting for. At the same time – we are embarking on another, even more exciting project.

New Binary Classifiers

We recently used Help Conquer Cancer results to train a new set of binary classifiers to complement our existing suite of Random Forest classifiers. Unlike the 11-way classifiers in the last update (one model distinguishes between 11 experimental outcomes), each of these classifiers detects a different event in protein crystallization trials: is-clear, has-phase-separation, has-precipitation, has-skin-effect, has-crystal, and junk. (The *junk* detector is used to label experiments where an obvious experimental error has occurred, such as a missing droplet, or contamination by dust.) Classification results for each classifier are shown below:



Of course, the most interesting classifier in the group is the *crystal* detector. It detects crystals with 93% precision and 53% recall, meaning that in our test set, 93% of the images chosen by the crystal-detector contained a true crystal, but that only 53% of crystals overall were detected. Classifiers from earlier updates and publications showed lower precision but higher recall (80%). A plot of the ROC curve of the crystal classifier is shown below.



Towards Project Completion and Future Work

Processing of all work units is nearly complete, with just a few weeks remaining. Much work remains to be done in our lab and in our compute cluster, compiling, analyzing, and mining the results, generating new models, and propagating the advances back into the protein crystallization pipeline. Work will carry on.

Meanwhile, the end of *Help Conquer Cancer* will not be the end of our involvement with the World Community Grid. We are busy developing a new project that will search for prognostic and predictive signatures (sets of genes, proteins, microRNAs, etc.) that help predict patient survival and response to treatment. The project will target multiple cancers. The title, launch date, and further details will be announced at a later date.

Some of the relevant related papers:

Cirilo, P. D. R., Marchi, F. A., Filho, M. C. B., Rocha, R. M., Domingues, M. A. C., Jurisica, I., Pontes, A., Rogatto, S. R. An integrative genomic and transcriptomic analysis rReveals potential targets associated with cell proliferation in uterine leiomyomas, *PLoS One*, **8**(3):e57901, 2013.

Vafaee, F., Rosu, D., Broackes-Carter, F. and Jurisica, I. Novel semantic similarity measure improves an integrative approach to predicting gene functional associations, *BMC Sys Biol*, **7**(1):22, 2013.

Goswami, R. S., Atenafu, E. G., Xuan, Y., Waldron, L., Pintor dos Reis, P., Sun, T., Datti, A., Xu, W., Kuruvilla, J., Good, D. J., Lai, R., Church, A. J., Lam, W., Baetz, T., LeBrun, D. P., Sehn, L. H., Farinha, P., Jurisica, I., Bailey, D. J., Gascoyne, R. D., Crump, M., and Kamel-Reid, S. A microRNA signature obtained from the comparison of aggressive to indolent non-Hodgkin lymphomas can be used in the prognosis of mantle cell lymphoma, *J Clin Oncol*, 2013. In press /2012/453050

Berger T, Ueda T, Arpaia E, Chio II, Shirdel EA, Jurisica I, Hamada K, You-Ten A, Haight J, Wakeham A, Cheung CC, Mak TW. Flotillin-2 deficiency leads to reduced lung metastases in a mouse breast cancer model. *Oncogene*, doi: 10.1038/onc.2012.499. In press.

Pastrello, C., Otasek, D., Fortney, K., Agapito, G., Cannataro, M., Shirdel, E., Jurisica, I. Visual data mining of biological networks: one size does not fit all, *PLoS Comp Biol*, **9**(1): e1002833. doi:10.1371/journal.pcbi.1002833, 2013.

Fortney, K., Xie, W., Kotlyar, M., Griesman, J., Kotseruba, J., Jurisica, I. NetwoRx: Connecting drugs to networks and phenotypes in S. Cerevisiae, *Nucl Acids Res*, **41**(D1): D720-7, 2013.

Starmans MH, Pintilie M, John T, Der SD, Shepherd FA, Jurisica I, Lambin P, Tsao MS, Boutros PC. Exploiting the noise: improving biomarkers with ensembles of data analysis methodologies, *Genome Med*, **4**(11):84, 2012.

Hammerman, P. S., et al., The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. **489**, 519–525, 2012.

McKee, C.M., Xu, D., Cao, Y., Kabraji, S., Allen, D., Kearsmans, V., Beech, J., Smart, S., Hamdy, F., Ishkanian, A., Sykes, J., Pintile, M., Milosevic, M., Kwast, T. van der, Zafarana, G., Ramnarine, R., Jurisica, I., Mallof, C., Lam, W., Bristow, R.G., Muschel, R.J. Protease Nexin 1 modulates prostate adenocarcinoma by regulating the Hedgehog pathway. *J Clin Invest*, **122**(11):4025-36, 2012.

Kotlyar, M., Fortney, F. and Jurisica, I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods*, **57**(4): 477-485, 2012.

Heifets, A., Jurisica, I. Construction of new medicines via game proof search. *26th American Association for Artificial Intelligence Conference on Artificial Intelligence (AAAI-12)*, AAAI Press, Menlo Park, 1564-1570, 2012.

Wong, S., Cercone,N., Jurisica, I. Characterizing healthy and disease states by systematically comparing differential correlation networks in lung. *Advances in Health Informatics Conference*, 2012.