Help Conquer Cancer project update, April 2014

Results:

How is the data being generated by the volunteers being used by your team? Are you analyzing the results yet or will analysis begin once the computations are done? Have you gained any insight from such analysis (any interesting findings, surprises, disappointments)? If the analysis is in process, how long do you anticipate this will take? Any way of quantifying and articulating how the data is useful to your work is highly valuable and motivating for volunteers.

Biologists and medical researchers use protein 3D structure to design drugs and understand protein function. Solving a protein's 3D structure requires a long and difficult sequence of steps. The *Help Conquer Cancer* project (HCC) addressed a bottleneck in this pipeline: recognizing protein crystals (or other experimental outcomes) in a protein-crystallization image. The HCC project on World Community Grid analyzed hundreds of millions of these images, but these results need to be processed further in order to generate reliable automatic image classifiers, discover trends in data, and ultimately improve our understanding how proteins form crystals.

Analysis is in progress, and there are some exciting results we will be reporting on next time. However, over the last year considerable energy and resources were devoted to the new project on WCG – Mapping Cancer Markers project (MCM), and other cancer-gene-signature projects, in which our research group is involved. To help with both priorities and directions, our team expanded and we have a new Post-Doctoral Fellow (Dr. Lisa Yan) helping us with HCC research.

Public access to data:

Has the data been made available to the public? What are your plans to do so? How soon will the data be made available to other researchers? If already available (e.g., in a public database), do you have any details on how the data is being used and in what context?

We have not yet decided the time-frame or the exact form of any *Help Conquer Cancer* data we will make available to the public. Thanks to World Community Grid members, our project's terabytes of raw image data have been transformed into terabytes of computed image features (morphological image properties used in automated image classification). The identity of proteins in the crystallization trials is largely unknown to us and partially unknown even to the Hauptman Woodward Institute (HWI, the source of the images). The features we have computed do not directly relate to crystallization outcomes or human-understandable image labels. A classifier is required to translate computed features to meaningful human labels or experimental outcomes. We have trained multiple image classifiers so far, but are confident that we can improve them. It is essential (and practical) that we finish this part of research, and release only useful data.

Paper publications:

Talk about any papers you're currently working on, where they might be in the review and publication process, anticipated timescales for publication and whatever you're comfortable sharing about the nature and significance of the paper. (Once a paper is accepted for publication, we'll work with you to build separate communications to promote it to the volunteers and wider community).

The Grid-computed results of the *Help Conquer Cancer* have yet to be fully analyzed. Once complete, we intend to publish one or more papers based on the analysis, but cannot currently estimate a time-

frame.

Collaborations:

If you collaborate with other research groups, volunteers would be interested in hearing about this. How are your teams collaborating? What are the dependencies between your teams? How might your collaborators be using your data? This would help volunteers understand the scientific research process and potential benefit of the data they generate beyond your immediate research team.

The High-Throughput Screening Lab at HWI supplied the original protein-crystallization image data, and indeed continues to generate more. Both HWI and the scientists who send them protein samples will benefit from the research in two ways: better systems for automatically classifying protein-crystallization images (saving time and manual labour), and better understanding of the protein crystallization process.

Team news:

Communicate any changes in your team structure and celebrate your team members' successes (e.g. competition of a thesis).

Dr. Lisa Yan joined our lab as a post-doctoral fellow, and is currently working on the HCC project.

Funding and grants:

Discuss any funding challenges and opportunities you're comfortable sharing. This will help volunteers to understand what a time consuming yet essential part of your role this is.

The work remains partially funded by grants from NIH and Ontario Research Fund (salary support for Christian, Mike and Lisa). Computing infrastructure is primarily supported by Canada Foundation for Innovation, with additional support from IBM.

Thank you all for your highly appreciated contributions to our HCC project

HCC team