1. **Availability of work**

Our estimate for the complete run time for this project is 2 years. Over the period of this project, we plan to analyze four cancers: lung, ovarian, prostate, and sarcoma. Currently, members are processing lung cancer work units.

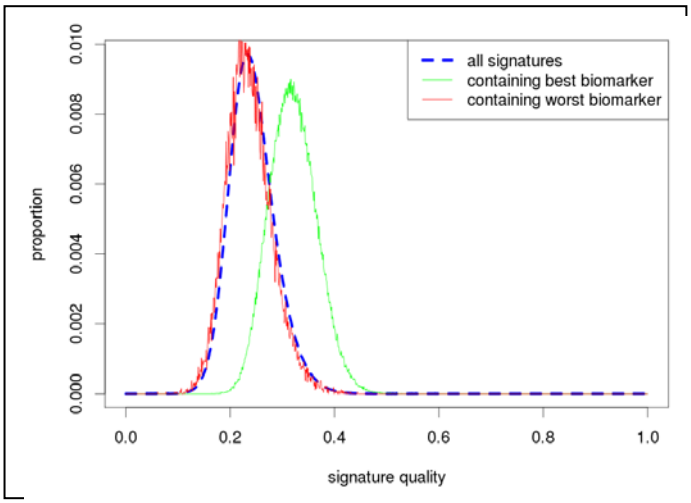
Over the past several months, members may have experienced shorter or longer than usual run-times. The core algorithm of the MCM engine, used to evaluate each potential gene signature, has a run-time that is highly dependent on the statistical characteristics of each signature. The search space targeted by a single work unit can sometimes contain time-consuming signatures, which together lead to a longer total run-time. A similar issue occurs with the size of MCM results. A typical work unit will evaluate tens of thousands of potential gene signatures, many of which are of low quality. Signatures below a certain quality threshold are removed from the returned results. The search space targeted by a single work unit can sometimes contain a high proportion of high-quality gene signatures, producing a larger result file.

There are currently two planned phases for the MCM project. In the first phase (described more in “Results”), we are attempting to set a benchmark for further experiments. The second phase is geared towards finding clinically useful molecular signatures, initially focusing on prognostic gene signatures. We expect a smooth transition between the two phases.

2. **Results**

Every year, numerous research groups from around the world develop and publish interesting molecular signatures for various diseases, including multiple cancers. One of the challenges of interpreting these different findings is that many of these reports are not directly comparable to each other. The first phase of the MCM project is designed to set a standard benchmark so that we and other groups can estimate how well individual signatures perform. For example, what we are trying to do might be similar to designing an IQ test. By establishing a standard test and scoring system, we can evaluate any person's intelligence. The results from the first phase of MCM will allow us to create such a test for existing and future gene signatures.

Our preliminary analysis of the work units processed (roughly 26 billion gene signatures) so far is focused on the nature of genes in the signatures, measuring their quality by assessing how accurately they contribute to identifying bad prognosis patients. One of the goals of the first phase is to understand if some genes might be better than others in their predictive ability. We took 0.1% of the top scoring gene signatures and identified the individual genes that make up each signature. For each gene, we looked at how many times it occurred within top scoring signatures and plotted the scores of those signatures (see Figure). The blue line shows the average of all of the genes together. The red line highlights the worst performing single gene while the green line indicates our best performing gene. The average of all the genes performs as well as the worst single gene. This is not surprising as most genes are likely to have poor predictive ability. We are looking for the few genes that separate from the field. In other words, if we have 1 million potential gene signatures, and we look at the top 1,000 scoring signatures, our green lined gene (among others) may have better predictive ability.



3. **Public access to data**: We are still in the result accumulation and preliminary analysis stage of the project. After the data is properly processed, we will make the results and data publicly available.

4. **Lab or further testing**: Validation experiments will be planned in later stages of the project, after comprehensive analyses are finalized.

5. **Paper publications**: No publications directly from MCM project in progress.

6. **Collaborations:** Besides existing collaborations in lung and ovarian cancer, we have ran a pilot study for prognostic signature optimization in leukemia (although at present, computation was done outside of WCG).

7. **Upcoming milestones or important dates**: We are nearing the end of the “landscaping” portion of the project which will determine characteristics of our search space. The next phase will involve more targeted approaches to gene signature discovery.

8. **Team news**: Mike had a 3rd daughter last year.

9. **Upcoming events:** Team Ian Ride – http://www.team-ian.org will be in mid-August.

10. **Technical updates:** We have completed the initial design meetings for porting the MCM code to GPUs. On the analytics side, we have been evaluating the use of IBM's InfoSphere Streams package to aid with post-processing of MCM results.

11. **Funding and grants:** Applications in progress to Ontario Research Fund; Canada Foundation for Innovation; Department of Defense – to both leverage contributions from WCG users, and fully utilize findings of the MCM computation with the primary focus on lung and ovarian cancer.

12. **Media and awards:** Interview with Globe & Mail technical editor on May 20, 2014.