Quantitative Medicine Project Meeting with Lundbeck
Wednesday October 16th, 2013
Webinar

Cambridge Healthtech Associates	**Quantitative Medicine**
Ernie Bush	John King
Dawn Van Dam	Scott Bodine *(not on call)*
Amid Zand	Geoff Hoare
	Josh Kangas
	David Demosthenes
Lundbeck	*Jim Parrino (not on call)*
Ulf Norinder	*Bob Murphy (not on call)*
Jorrit Hornberg	*Jamie Grooms (not on call)*

I) Welcome

Dawn welcomed Lundbeck and Quantitative Medicine to the meeting.

II) Quantitative Medicine Presentation

Josh discussed the agenda of this meeting:

1. Hepatotoxicity Study Updates

Josh shared the Hepatotoxicity study open questions from the previous meeting. They included:

- What is the effect of historical data on prediction accuracy?
- What is the effect of using only compound features on prediction accuracy?
- Can we effectively prioritize assays to predict toxicity? What is the effect on accuracy when assays are removed?
- What is the effect of utilizing neighboring compound and assay results on prediction accuracy?

Josh shared the data sources used in the Lundbeck study. Lundbeck provided the hepatotoxicity, HCS measurements and structural information was calculated from the SMILES provided. Finally, results of ~800 assays in which some of the compounds were previously tested were added to the study.

Josh described the prediction and selection methods. He then showed the graph of the hepatotoxicity simulation results. The results compared the *CoRE*, with RandomForest (using only HCS Results), the *CoRE* (no historical data), the *CoRE* (only structural data with *CoRE* selection), and the *CoRE* (only structural data and random selection).

The results showed that the effect of historical data on accuracy is not extremely high. There is a consistent improvement in using the experimental data and there is only a small decrease in efficiency when it is removed.

The results also showed a substantial value in the use of the *CoRE* when using structural data. Selection using the *CoRE* showed a higher accuracy of prediction as compared to random selection.

Finally, the results also showed that when HCS and historical data were not used, there was a substantial drop in accuracy.

Ulf: You only had historical data for a certain portion of the data, therefore there is not a big drop in accuracy when they are removed.

Josh: That is correct.

Ulf: Neighboring assessment can be done in different ways. How does your system decide on the right method?

Josh: We incorporate 5, 6, or 10 different ways of using neighboring information.

Jorrit: Can we use historical data with RandomForest and show the results?

Josh: Yes. The only challenge is that historical data is sparse and there may not be a standard approach for doing a RandomForest test.

Jorrit: It would be interesting to see how the system would work if the compounds were more closely related to each other.

Josh added that the slope of the purple line and the green line were similar and concluded that much of the learning in the system was from working with the structure and not from the HCS data or the external data.

Josh then continued to compare the accuracy of different methods.

Josh: Where do you see the most value in this technology?

Jorrit: It depends on the issue in hand. There is a lot of value in using structural information and neighboring compounds. We want to get the best prediction but need to look into best use cases.

Jorrit: We think of this as a research project. We are in TEC to do this in collaboration. It may be a challenge to help everyone understand the benefits of this technology.

John: Leveraging the TEC group is important in moving forward. Once we have the patent approved we can share more details about the technology. However, there are trade secrets in the algorithms as well. What do you suggest for the next step?

Jorrit: We can have a real live example and a publication. We have to decide how to share and release structural data with you. We can have an offline discussion with Josh.

III) Action Items

1) Ulf and Josh will discuss offline how to proceed with sharing compound information and design the next study (for example to assess the effect of neighboring information).
2) Ernie suggested that QM provide a summary of the findings to date to the steering committee to show the project progress.
3) Dawn and John will discuss how to move forward with communicating the results to the TEC group.