Work Study Projects

1. Develop web interface for national flow cytometry data repository

Background: Flow cytometry is a high-throughput procedure for the measurement of individual cell phenotype and function within large populations. It is used extensively in basic research and clinical trials to monitor complex immune system responses. We plan to develop a centralized public data repository for researchers to share their data using the MiFlowCyt standards recently accepted by the International Society for Analytical Cytometry (ISAC). We expect that if successful, researchers funded by NIH will be required to upload flow cytometry data at the time of any publication.

This project is loosely modeled after the existing repository for microarray data, and is expected to confer similar benefits to the research and clinical community. This work will be done in collaboration with Ryan Brinkman (University of British Columbia), who is the lead author of the MiFlowCyt standards.

Goal: Develop a web interface for researchers to upload and annotate flow cytometry data sets, with a mechanism for efficient database coupling.

Skills: It is likely that we will use a dynamic language (Ruby, Python) web framework to implement the interface, with AJAX features to enhance the user experience. Knowledge of HTML, CSS, JavaScript and a relevant scripting language are required. Experience designing database schemas is not required but welcome.

2. Improve existing software for flow cytometry statistics and visualization

Background: Existing software for flow cytometry falls into two disjoint classes - user-friendly commercial software with minimal statistical capability, and advanced statistical packages that require programming expertise to use. There is a need to put powerful statistical tools in the hands of the end-user as the current methods of manual analysis are time-consuming, subjective and error-prone.

An early version of the software has already been written, with the front-end coded using the wxPython GUI library, and hooks to statistical routines in C++/R. The major work here will be the integration of the software with the R BioConductor statistical packages for flow cytometry and the GGobi visualization library for advanced graphics, as well as packaging the software for easy distribution and installation.

Goal: Build an application framework that will provide advanced statistics and visualization capabilities in a user-friendly package.

Skills: Expert Python coding and familiarity with the R statistics package are required. Experience with GUI design and software packaging will be very useful.

3. Implement pipeline for high throughput flow cytometry statistical analysis

Background: Modern flow cytometry data sets are huge and we are developing advanced statistical methods based on Bayesian mixture modeling for the automatic clustering and analysis of the data. These techniques are extremely computer-intensive, and would benefit from the use of multiple cores/Beowulf clusters/Grid systems to speed up processing.

Goal: Develop a pipeline in C++ that takes raw flow cytometry data, and apply the appropriate transformations and statistical processing, automatically exploiting multiple cores/processors where available.

Skills: Knowledge of modern object-oriented C++ and design patterns is required. Experience with parallel programming would be highly useful. A sophisticated statistical background is not required, as this project is focused on building the infrastructure for computation and not the actual analysis itself.

Ongoing Collaborations

Flow cytometry

All these projects will be done in close collaboration with Kent Weinhold and Janet Ottinger, who run the Center for Aids Research (CFAR)/Duke Translational Research Institute (DTRI) flow cytometry core.

Statistics

We also collaborate with Mike West and Chunlin Ji from the Department of Statistical Sciences (DSS) for advanced statistical methodology.