

Participants:

Fran Lewitter, WHI
Michael Rebhan, Novartis
Brent Richter, Partners Healthcare
David Sexton, Vanderbilt
David LapointeUMass Medical
Steve Jennings U of Arkansas, Little Rock
Charlie Whitaker MIT Cancer Research
Somitt from Mayo clinic
Amir Bouers from Harvard University.
chen xui—(did not obtain affiliation)
John Rux Upenn
Mauman Maqbool
Aaron Noll, Stowers Institute
Hemant Kelkar UNC
David Tom UNC
Jianping Zen, UNC
Monika Wood Promega
Laxman Iyer Tufts University
Dawei Lin UC Davis

Bioinfo-core BoF's at ISMB (July 19 - 23).

N.B. We have had to change the second BoF session from Monday to the Tuesday lunch hour. And—if there are additional volunteers who would like to present, please contact Fran or David soon (we would like to have preliminary slide decks to us by July 8 for review). Please see end of these minutes for further details of the agenda and dinners.

FL--We have organized 2 meetings in Toronto to be held during the Birds of a Feather lunchtime sessions. Both are allotted times by ISMB for 1.5 hour each from 12:40 to 2:15. We have planned to start our sessions at 1pm to allow time to grab some food.

We've planned to discuss 2 topics—Support of next gen sequencing and best practices in running a bioinformatics core. These will take place on Sunday July 20th and Monday respectively. On this call we would like to discuss topics for each session and ask for volunteers or nominees to present a short topic.

It was discussed during the call that the time allotted could accommodate 3 – 4 presenters in each session for 10 minutes each. This will provide plenty of time for discussion and questions to follow. The talks should be informal and practical in nature.

What topics should be included in each session? There was a brief discussion considering the difference between the bioinfo-core BoF and the SIG. The talks selected for the SIG are

presentations of algorithms. The topics for the core meetings, considered below, are more pragmatic—there will not be an substantive overlap.

Support of Next-Gen Sequencing

LI—volunteered to present a report on the next-gen sequencing SIG.

DL—their group has developed a web based LIMS system for Solexa. She can present on this. Some questions raised on the call for Dawei:

Will this LIMS be released under an open source license?

Can a URL for demo version be distributed to the list?

have solexa from August.

HK—volunteered to speak on how to get started with analysis of the data. Will include a discussion of end user tools and how the UNC group interacts with end users. This will be experience based on the GA2 platform over the last 4-5 months.

Slightly off topic:

DS--Are there different ways to run your core with regard to Solexa? What's the data flow?

These questions will be discussed at ISMB. A discussion on the conference call focused on what data is being delivered to the investigator. David Sexton is not delivering raw data at all. They work with the data through to analysis. This is the general case with the rest of the callers as well.

BR—Partners has a mix of delivery. Mostly it is analyzed data to the investigator, but in some circumstances, the investigator is developing their own algorithms and running the pipeline themselves.

DS--What are the workloads and some of the examples for different groups? It will be good to understand the different ways in which the community is using their next-gen instruments.

There has been found to be a concentration effect in running the ga2. Need to look out for this. Some cores are looking at % error in order to determine ideal conc. This is empirically derived to start.

back on topic:

DS—David volunteered to discuss analysis tools and tool development. He's had some experience with UCLA's software suite and we can discuss some of the ISV's who are doing genome alignments.

ISCB: Regarding getting started with next gen sequencing: there is a column in the journal PLoS Computational Biology regarding how to get started in a particular area. A possible topic could be how to setup a core for handling Solexa data. If you are interested in participating, please send mail to Fran.

Best Practices for Running a Bioinformatics Core Facility.

Topics to be discussed will include the hiring challenge, customer service, funding & chargebacks, practical aspects (buy in, budget, etc). Ideally presentations will come from different sources including someone running an established core, a newer core, and industry, state, research institute, public and private university perspectives. Topics to touch on include outreach, training, education.

FL--volunteered to talk about Whitehead core, a small US research institution.

DS—volunteered to provide the mid-sized private university perspective.

DL—The large state university perspective of UC Davis.

SA—will speak about the Babraham Institute in the UK.

MR—has been chosen to speak from an industry perspective.

Some other discussion points on this call:

We can use the bioinfo-core wiki for adding discussion points to the meeting, or sending them around to the list. Fran/Michael/Brent/David will also be happy to take a note of a discussion point prior to the BoF so that we can ensure the point brought into the meeting.

Part II of call: ePHI:

How does your institution balance need for security and promote collaborative research?

Some organizations simply deal with the end product of the data--Data comes de-identified. The data is sufficiently de-identified such that it cannot be traced back to an individual. Most data is stored in oracle databases that maintains authorization through controlled password management.

DS—There is a need to ascertain individuals outside of a healthcare information, for recruitment purposes. These are circumstances where the primary data is collected during prospective studies where patient identifiable data is submitted. Data is collected and stored using oracle. If the form is web based, data transfers are performed using SSL and in general, no data or passwords are transmitted in clear text.

For those collecting identifiable information, data is always kept within servers housed in institutional data centers, behind the firewall.

DS--SSH is only opened internally within the institution, and not externally. Access into the Vanderbilt network is restricted.

HK--UNC has 2 networks, a separate medical firewall walls off the clinical side. The research enterprise is outside this firewall.

BR--Partners has 1 big network with small pockets, but in general clinic and research share the same network. SSH is selectively enable through the firewall, but controlled.

What about collaborators? Data going out to collaborators is always de-identified prior to sharing.

Cloud computing and amazon. Any experience in hosting ePHI offsite?

There are several circumstances that are becoming more common in requests to IRB's—can an investigator use a third party to host data. Some examples are Amazon, but also google docs to share information with collaborators, survey monkey to collect information, etc.

DS—Base camp hosts it's own servers and you can choose SSL encryption for your project. All their servers are in a data center with all of the physical security you would expect. You would still use caution in permanently storing PHI here.

23 and me—they are doing everything in house. If using genotypes themselves, there is no personal information on Amazon's cloud or S3. Their data is encrypted possibly using a 3rd party encryption company.

DS--Vanderbilt has come up with own solution to host data.

BR—Partners provides central facilities to store/host information. If hosted by investigator, Partners will request a security audit from the information security group, but implementation is left up to the investigator in collaboration with the IRB.

DL--UC Davis has published guidelines.

guidelines for slides:

deadline for slide deck to organizers: July 8th

slides to David Sexton for Sunday talks, and Fran for Monday talks.

12:40 - 2:15 is allotted time, sessions will run from 1-2:15.

10 minutes presentation, 8-10 slides absolute max.

Publication of agendas will be via the list serve prior to the meeting.

There will be 2 dinners organized, one to take place on Saturday and the other on Monday, both following the end of the meeting. There are no formal organized ISMB events during these evenings. If this changes, we can reschedule. Announcements and meeting places will be made via the list.