

Q: Question, **A:** Answer, **O:** Opinion

Rebecca Talk - 1

Uma Talk

Philip Talk

- **O:** Work on use cases that exceed the capacity of a login node, like obtain reads in certain locations of a BAM file (slicing a BAM file).
- **O:** Should we have region specific (for e.g., chr specific) BAM files to enhance performance? – Mike Barmada
- **O:** Or split the slicing job across Frank and PSC? – Adrian Lee

Mike Barmada Talk

- **Q:** Is there a way to request on node over another? Like a node with infiniband connection?
- **Q:** Any plans for RStudio to be installed in Frank? **A:** No plans for the moment as it's a GUI.
- **O:** RStudio server installation in SAM is an option though – Roger Day.

GenomOncology – Clinton Talk

Rebecca Talk – 2 (NCI Cloud Pilots and CGC)

Q: How to run TCGA data on Cloud? **A:** You install the tools in Cloud and then run it

Q: Will they let us use their raw BAM files? **A:** Yes!

O: Overlaying cBio or GenomOncology over TCGA data would be very good for visualization and analysis.

Q: The methods for harmonizing data? **A:** Left up to the centers. Hence, no right way of doing it.

Paul Cantalupo Talk

- **Q:** Actual process involved? **A:** All BAM files were downloaded using GeneTorrent from CGHub. Reads that do not map to human genome were extracted and went through DB's like UCSC genes, RefSeq, NCBI to map the unmapped reads to different versions Human genome. Then, these unmapped reads were taken to map to virus genome using Bowtie2/BWA.

Xinghua Lu Talk

- **Q:** Is CNV data from the array data that TCGA provides? **A:** Yes

Kevin McDade Talk

Nolan Priedigkeit Talk

Kevin Levine Talk

- **Q:** Where the inconsistencies in the TCGA data for ductal carcinoma resolved? **A:** No. They have not fixed all of them.
- **Q:** How others found any other discrepancies? **A:** Clinical data outcomes are dissimilar between Pitt and TCGA. TCGA clinical data update is a bit static.
- **O:** The working group of TCGA stopped updating data.
- **O:** Da Yang suggests using TCGA working group data freeze (NCI wiki) or PANCAN instead of control access. Do we want to provide PANCAN data as a part of PGRR?

Da Yang Talk

Better way to collect clinical/outcome data?

- Depends on the end user.
- Start off with Breast and Lung cancer outcomes data as a part of CURE and then expand on other datasets.
- Which data elements are the most important to capture and how much accuracy do we need?
- Cancer registry would be useful for PGRR.
- Volunteers – Adrian Lee, Xia

Having a resource available to do intermediate analysis.

- Someone could upload a bed file to portal and in the backend a job could be run and return the sliced BAM files.
- Need more use cases to come up with a concrete solution.
- Design something more generic.

Making the Project results available

- A proper system to share project results, like Driver mutations from Dr. Xinghua Lu.
- Something like cBio.
- Publish pipelines or data? Worth looking into. But ideally we need a system where the user can just easily search data.
- Volunteers - Da Yang, Xinghua Lu

Better wiki to communicate

- We need a better way to share information or ask questions.
- A face to face meeting would be better.

How frequently can we update the PGRR?

- Data refresh as there are small increments especially BRCA as people use it most frequently.

Need CNV discrete data for BRCA

- Haven't downloaded it yet. Will download this level 4 data soon (within 2 weeks).

Should accumulate data from UPMC

- We need to accumulate clinical data and research data within UPMC.
- We already have UPMC warehouse working on it though. The current state: genomics side is almost finished, but the clinical data is not done yet.