* Practical access to publicly available datasets- challenges thereto, and how this group might be able to facilitate- Radha to lead discussion
  + Methods to accelerate
  + Datasets that look at disease phenotype, across different scales
  + eMERGE (40 phenotypes into dbGap), TCGA, ExAC, Snyder?, Sage projects,
  + Sam- disease commons for NCI- place to deposit any type of data for a disease, and building out tools on top of that.
    - Starting with cancers
    - Looking for pilots
  + Note- current focus of omics types, shift to disease focus?
  + Benchmark data resource?
  + Cool sounding TCGA-esque project that I missed
  + Let's write a review paper! Interested parties: Radha, Sam, Bisakha, Panduka. Awesome.

Accessing genomic data for biomedical research; progress and challenges

Technological advancements over the last two decades have produced cost-effective, high throughput experimental techniques elucidating the genomic composition of the human genome. These techniques have been heavily utilized in biomedical research to understand the genetic contributions in human health and disease. Heavy utilization of high throughput sequencing techniques in diverse biomedical projects have led to an unprecedented rate of digital data generation. These data need to be shared efficiently across the scientific community in various formats depending on the user needs to facilitate scientific discovery. Several data repositories have been introduced to host and distribute data such as nucleic acid sequences, high level sequence analysis, biorepositories with Electronic Medical Records (EMR) data, and gene expression data. Here we present a survey of data repositories landscape, data access protocols, challenges faced by biomedical researchers and discuss methods to accelerate scientific discovery in the precision medicine era.

We will discuss a wide spectrum of data repositories hosting human DNA elements, expression assays of genetic regions, genome level sequence reads and high level genome-wide sequence analysis. We will elaborate on the classification of whole genome, exome, RNASeq sequencing data available through the repositories, and discuss mechanisms of access control.

Cell lines: encode geo

Emerge

DNA elements: encode

Human: WGS, RNASeq, WES – dbGaP, EBI-EGA,

Annotation tracks: UCSC, dbSNP

Eye on the future with sequencing in precision medicine era

Challenges: Access delays, download size and HPCF resources necessary for analysis once downloaded, moratoriums