01.23.2017 Melissa Pespeni

Info Update Rubric:

* Outline
* Infor update ~20 minutes
* Learning/engaging activity
* Use board effectively
* Take home message(s)
* Examples from literature

What is Next-Gen sequencing? (Outline)

1. Advances in sequencing technology
   1. Human Genome Project (finished ~2001-2003)
      1. Mostly with Sanger Sequencing (ABI)
      2. 15 years to complete
      3. 1 person’s genome
      4. Cost ~$3billion
   2. Illumina HiSeq X Ten (2014)
      1. 1 day 🡪 45 whole genomes of a human
      2. $1000/each genome
2. Range of applications
   1. Whole genome sequencing
   2. RNA sequencing
   3. ChIP Sequencing (Chromatin Immunoprecipitation Sequencing)
      1. Using antibodies recognizing-proteins and can illustrate what proteins are being transcribed
      2. Transcription networks
      3. Protein-DNA interactions
   4. Targeted/Captured Sequencing
      1. Design probes (synthesized DNA) and hybridize with sample
      2. Pull targeted genes using probe
      3. Need to have prior information about organism’s genome
3. General library preparation workflow
   1. What may influence what technology you chose to use?
      1. Where is the targeted genetic variation?
      2. Scale/# samples
      3. Model vs. not model organism
      4. \*\*Is there genomic information available?
      5. What are you looking for?
         1. Demographic history
         2. Adaptive genetic variation
         3. Gene expression variation
      6. Length/number/distribution of reads
   2. Workflow
      1. Extraction (DNA, RNA 🡪cDNA)
      2. Fragment sample (break into smaller chunks)
      3. Ligate adaptors on (individual barcodes)
      4. Add sequence adaptors
      5. PCR
4. Sequencing-by-Synthesis (SBS)
   1. Sample loaded on flow cell into a lane
   2. Sample washes over oligos and attach
   3. Bend over (bridge amplification)
   4. Amplifies in tight clusters of the same sequence (generates cluster)
   5. Labeled Dntp (ATCG) build up in sequence order
5. Other technologies
   1. SMRT (see below)
6. Learning activity
7. **Take Home Messages**
   1. **Whatever your application is, you’ll probably use Illumina (SBS)**
      1. **90% of sequences world-wide at this point**

**Glossary:**

* Reads
  + Short reads (50bp)
  + Long reads (100, 150, 300,…10,000-60,000 bp [SMRT])
    - Long reads are easier to assemble for non-model organisms (de-novo assembly)
    - Read length is based on approach used
  + Single vs. paired end
* Reduced representation
  + RNA
    - Coding regions
  + GBS/RAD-Seq
    - Near restriction sites
    - Non-coding regions
* SMRT (Single Molecule Real Time)
  + Reduced accuracy
  + Increased amount of data

**Paper Discussion:**

*Genome sequencing and population genomics in non-model organisms* (Hans Ellegren 2013)

\*See Mendeley PDF

Homework:

-Sign up for 1 info updates and one paper discussion 1