Lab Meeting 5/9/19

Pre-meeting

1. Sam discussed the need for users of the Nanopore/flow cells to enter basic statistical info (data, identify values, where data are stored on PC, # of reads, run time) into an Excel spreadsheet after they are finished. This information can be sent to/used by Nanopore to work with us as QC data.
2. As the amount of data we generate ↑, we will need to develop a naming system for our files
3. We will have summer students and Penn students working with us, so we will need to work on technical documentation.

Presentation: Yan Gao - IsoCirc: Circular RNA profiling and analysis using Nanopore long-read sequencing

1. Sequence Workflow/Computational Pipeline
    - Figure has minor inaccuracies: arrows need to be remade, and there are some typos
2. CircRNA Species and Isoform Statistics
    - Need to work out nomenclature for several terms used in slides: species, overlapping gene, truncation
    - Rather than Venn diagrams, consider presenting data in other formats, e.g., pie charts with raw data plus percentages
    - Large numbers should be presented with commas for thousands place
    - For graphs, numbers should be presented as whole numbers rather than scientific notation (e.g., 6,000,000 rather than 6.0e6)
3. Slide with Venn diagrams comparing "All circRNAs" to "circRNA isoforms": switch headings
4. Union of datasets: Most datasets include testing datasets, different experiments, leading to a high rate of false positives. Ruijiao is rerunning with new datasets to improve data.
5. Number of circRNA species among overlapping species: #s look too high, should check
6. Internal splice sites
    - CircRNA may come from different sites, and may not be confirmed by short-read data. No good standards exist for internal splice junctions without full-length sequence, and the short-read data (known splice site) is not sufficient.
    - One alternative would be to check the splice-site score (e.g., package from Jing Yao lab). A high score would indicate a true splice site.
    - For rate-mapped data, could map flanking sequencing for true splice sequence
7. CircRNA diversity among all isoforms: Check cumulative #s - appear to be too low
8. CircRNA introns returned are shorter
    - Data have an inherent bias - can't actually resolve because of the large length of circRNAs
9. Correlation between circRNA block # and total length
    - Data go above y-axis
    - Short-read instruments cannot provide these data, highlighting the utility of this approach
10. Mapped genomic region of circRNAs
    - Flip the figures so that exons are given priority
11. Flanking elements have more repetitive elements – Typo in title
12. Yi mentioned a paper by Lin Chang in Cell on ncRNAs, circRNAs in immunoresponse