- RNA-Seq Goals:massively parallel cDNAsequencing
    - Precisely defines TSS, reveals small ncRNAs/ long ncRNAs, alternative splice products/sites, and estimate transcript levels
- Reference-based RNA-seq analysis
    - RNA-seq data can be aligned to a reference genome to determine location of origin for RNA molecule.
    - Tools available to accomplish this include TopHatsuite and Cufflinks
- Alignment issues
    - cDNA comes from mRNA which has already been processed and therefore the introns have been removed from that sequence, meaning that sequences can have large gaps in between them, which the alignment software has to account for.
    - Paired-end sequencing becomes difficult due to these gaps as well, since the gaps could lead to ends being very separated.
    - TopHat allowsfor transcript assembly rather than genome assembly to account for this obstacle.
- Cufflinks, Cuffcompare, and Cuffdiff
    - Cufflink: probabilisticmodel to assemble reads of transcripts, using accepted_hits.bamfile from a TopHatrun of the sequence. Produces GTF file that includes gene name/loci.
    - Cuffcompare: compares 2+ GTF files to examine gene expr. differences between cell types, dev. stages, or disease states.
    - Cuffdiff: statistical analysis of differential expression, including different transcript isoforms, TSS, and exons expressed.
- Other tools for RNA-seq data analysis include Diffsplice, DESeq2, edgeR, EBSeq, and baySeq.
- *De novo*based analysis
    - Align RNA-seq data together based on overlapping segments
    - Most popular program = Trinity, others include Salmon, Sailfish, RSEM, and kallisto
- RNA-Seq output:
    - location and expression level of known/unknown transcripts.
    - Comparison of gene expression across different cell types and condition.
    - Advantage over microarrays which need very specific probes and are not as well suited to detect different splice site expression.