

Outline

- Gene regulation
- RNA-seq
- Regular expressions, grep, and sed

---

---

---

---

---

---

---

---

### Regulation of Gene Expression

**Regulation of transcription:**

- Transcription factors
- Histone modifications
- DNA methylation

**Regulation of RNA processing:**

- Polyadenylation
- Splicing
- Capping
- RNA export

**Regulation of translation:**

- mRNA decay
- Translational repression
- Sequestration

**Posttranslational regulation:**

- Chemical modifications (e.g. phosphorylation)
- Protein turnover (proteolysis)

**RNA-seq measures steady state mRNA levels**

Fu et al. (2014)

---

---

---

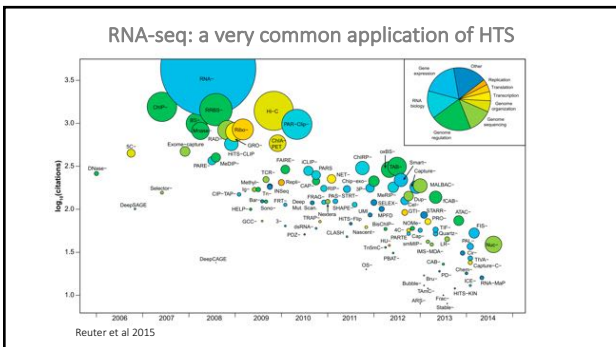
---

---

---

---

---




---

---

---

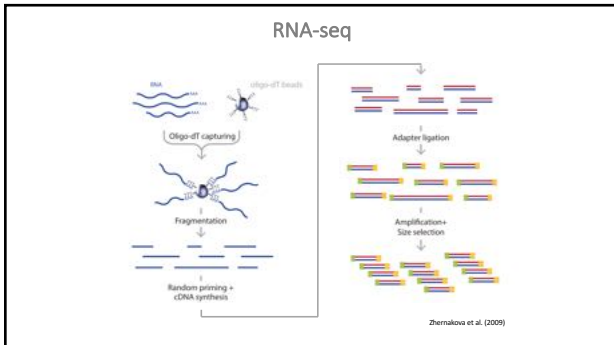
---

---

---

---

---




---

---

---

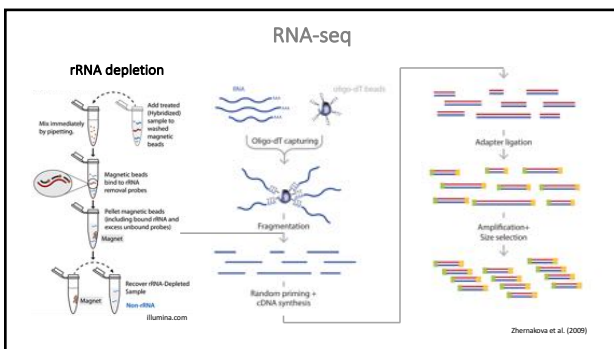
---

---

---

---

---




---

---

---

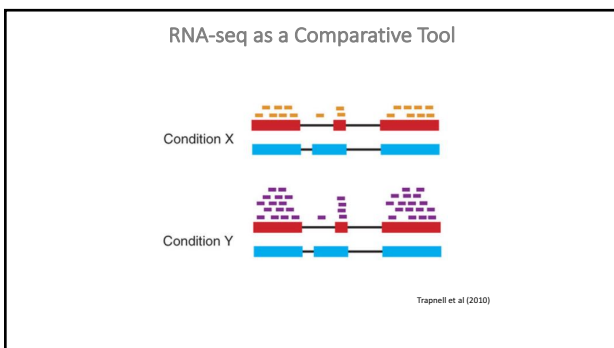
---

---

---

---

---




---

---

---

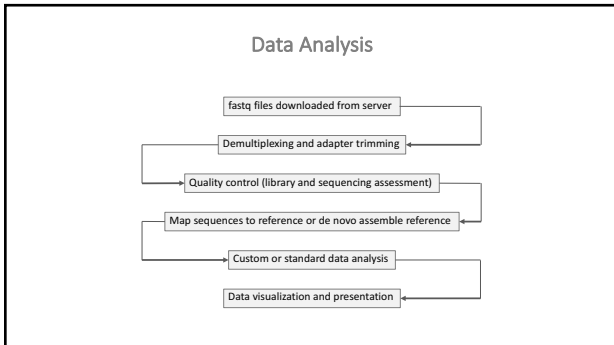
---

---

---

---

---




---

---

---

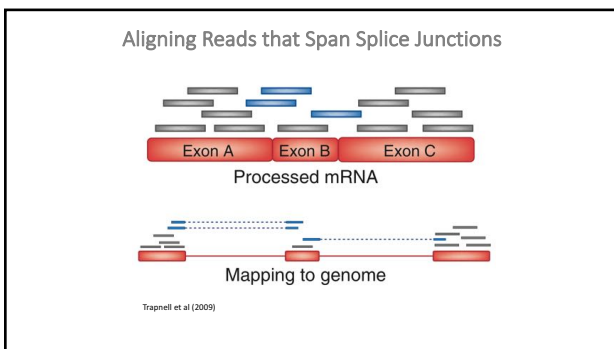
---

---

---

---

---




---

---

---

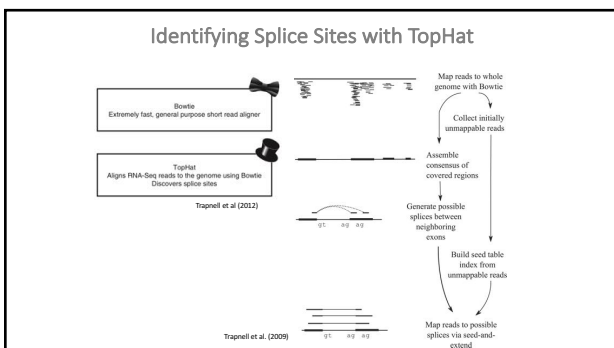
---

---

---

---

---




---

---

---

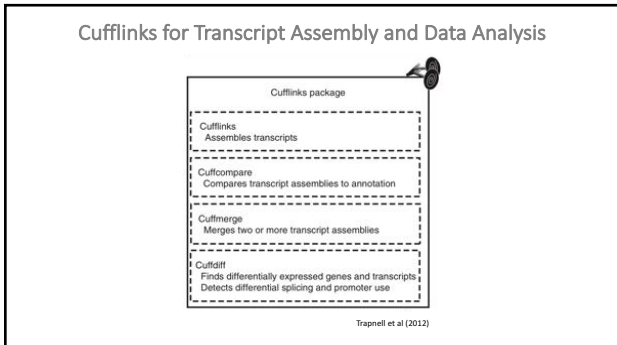
---

---

---

---

---




---

---

---

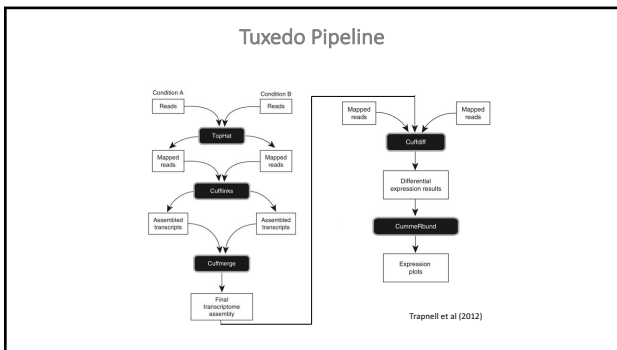
---

---

---

---

---




---

---

---

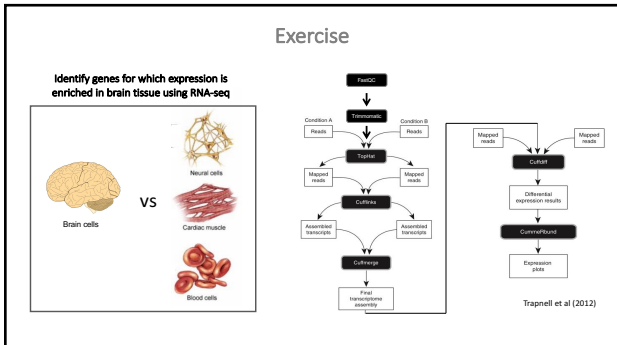
---

---

---

---

---




---

---

---

---

---

---

---

---

### Files Needed for RNA-seq

**High-throughput sequencing data (2 files per library if paired-end data)**

- fastq format

**Genome sequence, if available**

- fasta format

**mRNA coordinates, if available**

- gff format

---

---

---

---

---

---

---

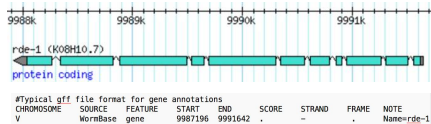
---

---

---

### GFF Format

**Generic Feature Format (gff3): 9 tab-delimited columns**




---

---

---

---

---

---

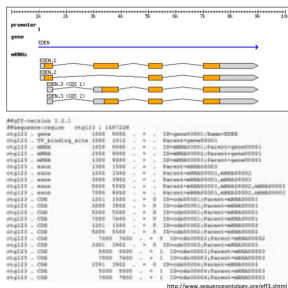
---

---

---

---

### GFF Format




---

---

---

---

---

---

---

---

---

---

### Grep, Sed, and Regular Expressions

Genomics data files, such as reference sequences and annotations, often need to be filtered or manipulated before using them as references for data analysis, such as RNA-seq

**Regular expressions:** character combinations that represent a particular pattern.

**grep:** a command line tool for searching for patterns within a file and returning lines containing the pattern.

```
$ grep "pattern" file
```

**sed:** a command line tool for making substitutions in a file. Allows pattern matching with regular expressions.

```
$ sed 's/"old_pattern"/"new_pattern"/g' input_file
```

---

---

---

---

---

---

---

---