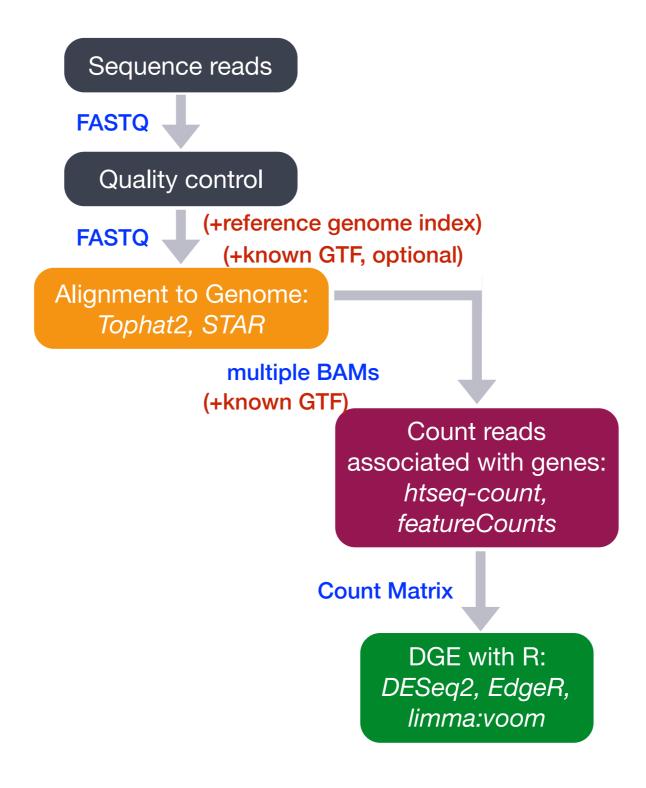
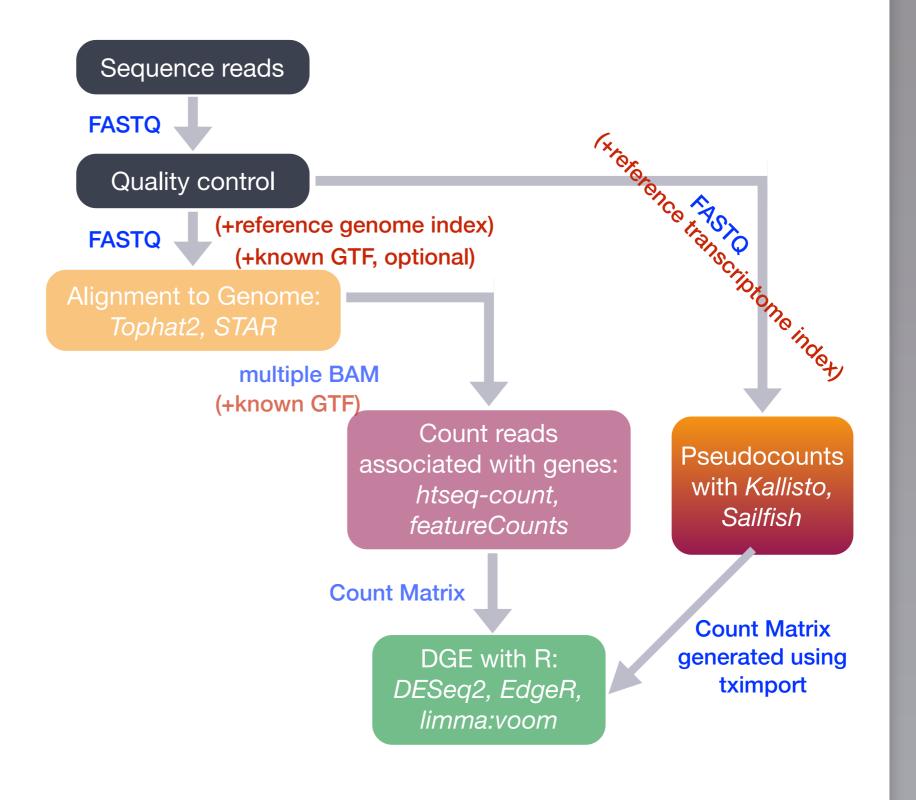
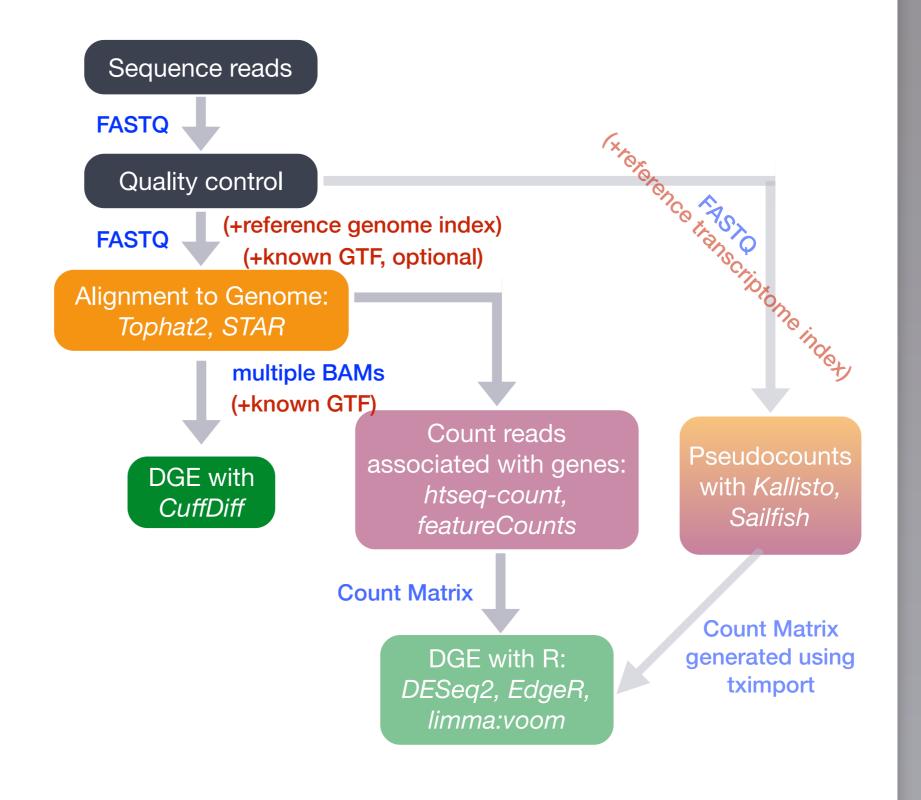
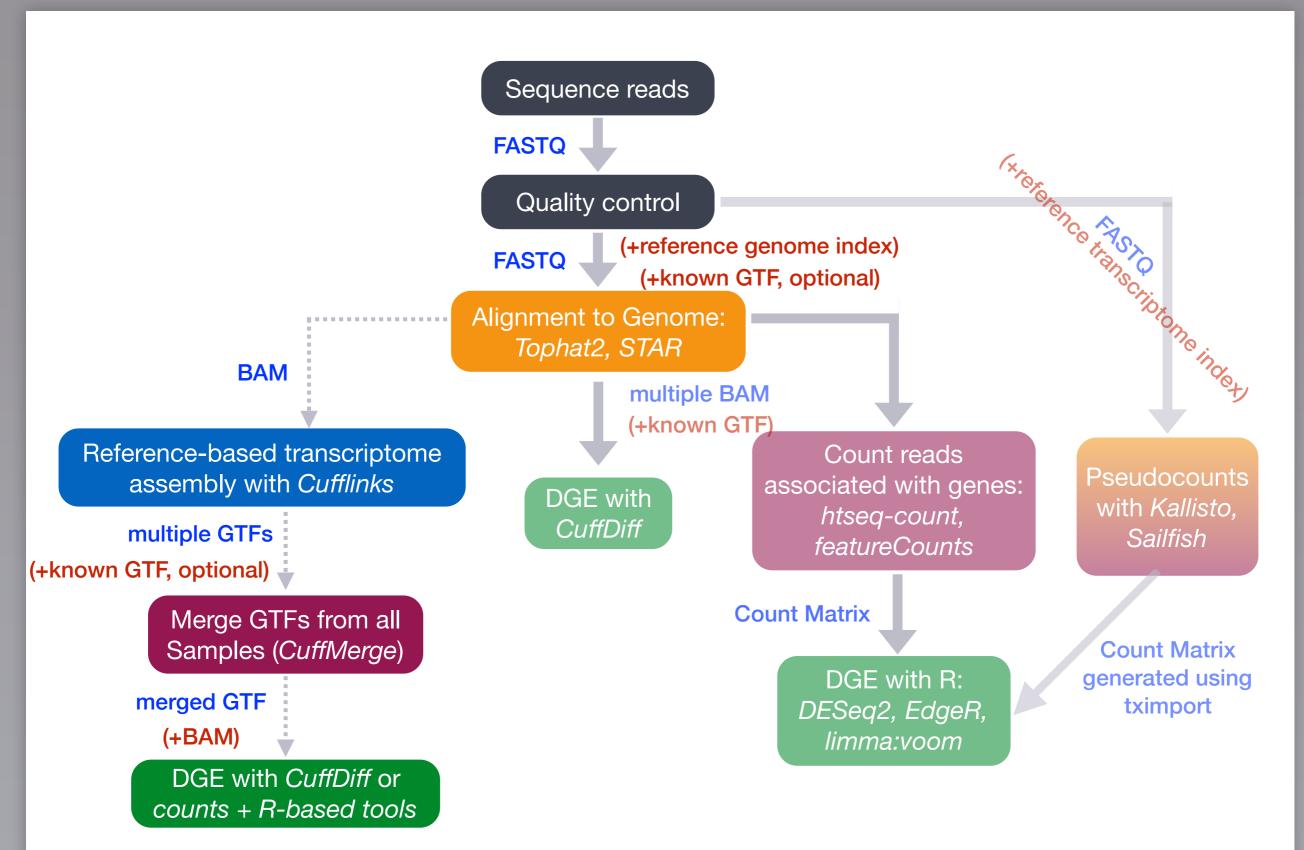
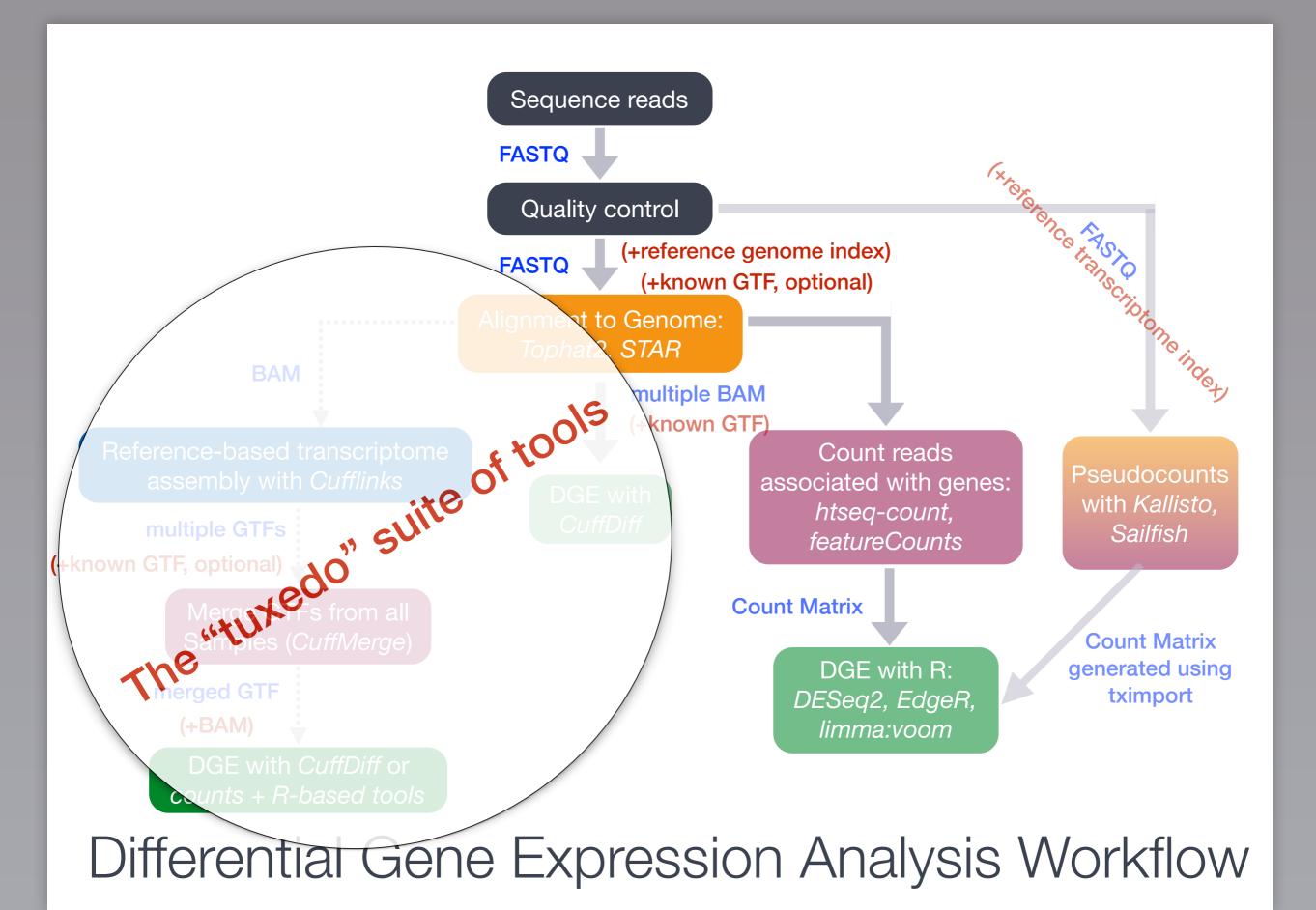
## RNA-Seq: Analysis methods













Bowtie Extremely fast, general purpose short read aligner

Bowtie/Bowtie2 both use Burrows-Wheeler indexing for aligning reads.

TopHat
Aligns RNA-Seq reads to the genome using Bowtie
Discovers splice sites

Tophat uses either Bowtie or Bowtie2 to align reads in a splice-aware manner and aids the discovery of new splice junctions

Cufflinks package

Cufflinks

Assembles transcripts

Cuffcompare

Compares transcript assemblies to annotation

Cuffmerge

Merges two or more transcript assemblies

Cuffdiff

Finds differentially expressed genes and transcripts
Detects differential splicing and promoter use

The <u>Cufflinks package</u> has 4 components, the 2 major ones are listed below -

Cufflinks does reference-based transcriptome assembly

Cuffdiff does statistical analysis and identifies differentially expressed transcripts in a simple pairwise comparison, and a series of pairwise comparisons in a time-course experiment

#### When is it appropriate to use Cufflinks/CuffDiff?

You are looking to identify as yet unknown genes/isoforms

Your experiment is a pairwise comparison

You are looking for isoform-level differential expression

#### Cons:

Complex experimental designs are not supported

False positives are rampant with novel discovery methods

# When is it appropriate to use R-based analysis methods using raw counts?

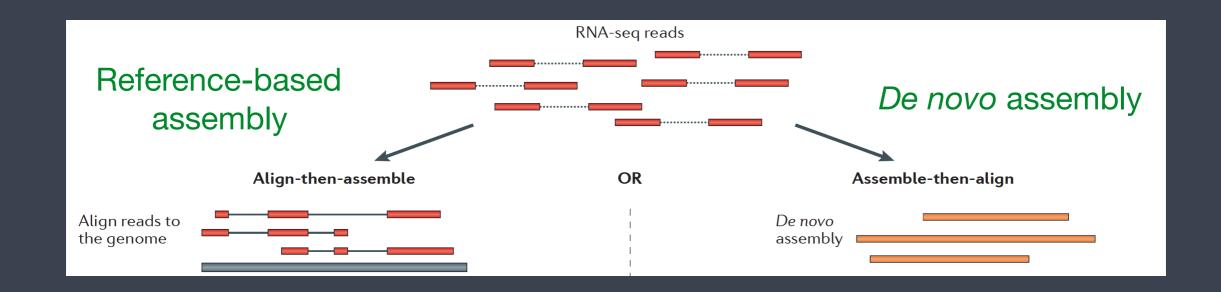
For almost all experiments where you are looking to identify differentially expressed genes (known genes), irrespective of the complexity of the experimental design

#### Cons:

A basic knowledge of R programming is required

Not useful for isoform-specific differential expression analysis

Undercounting due to discarded multi-mappers



Reference-based assembly

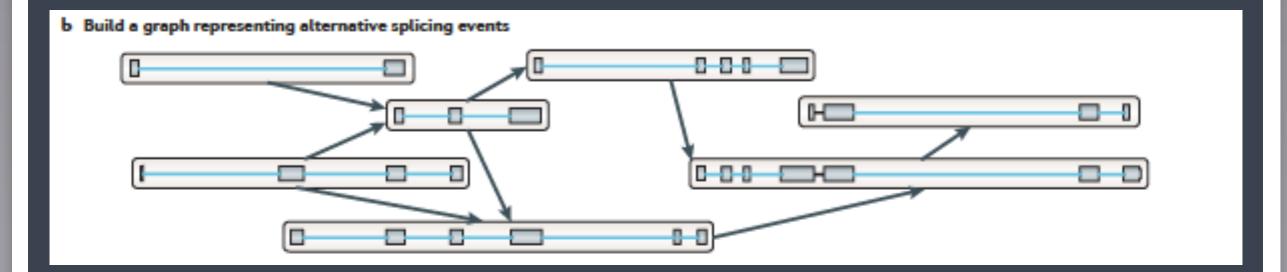
This type of assembly is used when the genome sequence is known.

- Transcriptome data are not available
- Transcriptome information available is not good enough, i.e. missing isoforms of genes, or unknown non-coding regions
- ♦ The existing transcriptome information is for a different tissue type or state
- Cufflinks and Scripture are two reference-based transcriptome assemblers
- Annotation of any newly-discovered genes or isoforms will be performed downstream

#### Reference-based assembly a Splice-align reads to the genome 176,800 kb 176,802 kb 176,804 kb 176,806 kb 176,808 kb b Build a graph representing alternative splicing events 0 0 0 0-0-0

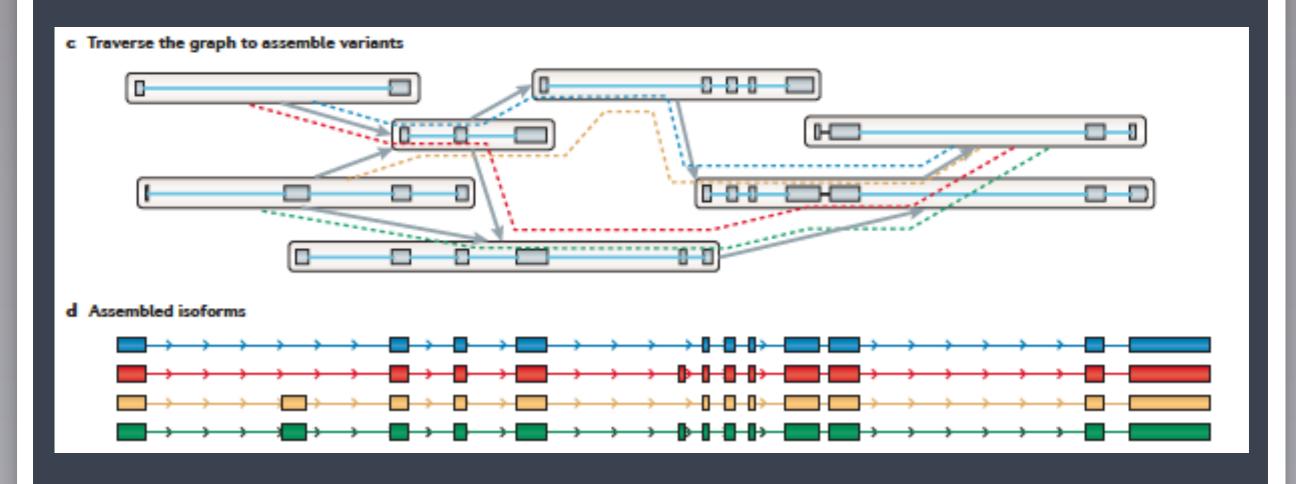
Martin J.A. and Wang Z., Nat. Rev. Genet. (2011) 12:671–682

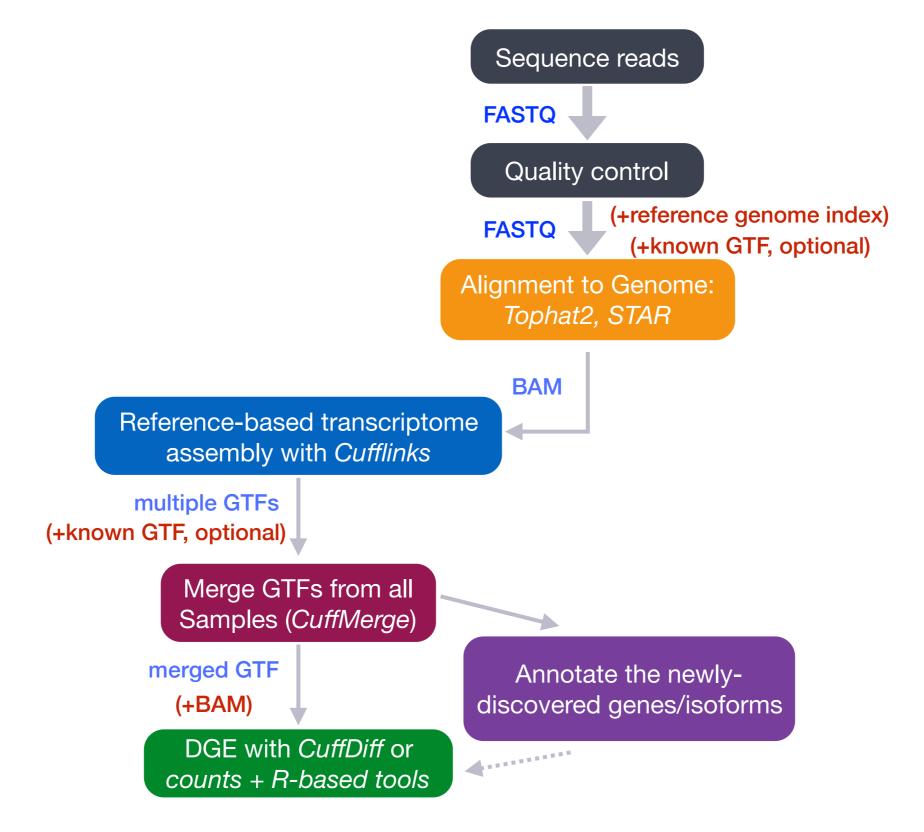
#### Reference-based assembly



# Reference-based assembly b Build a graph representing alternative splicing events 0 0 0 c Traverse the graph to assemble variants 0 0 0

#### Reference-based assembly





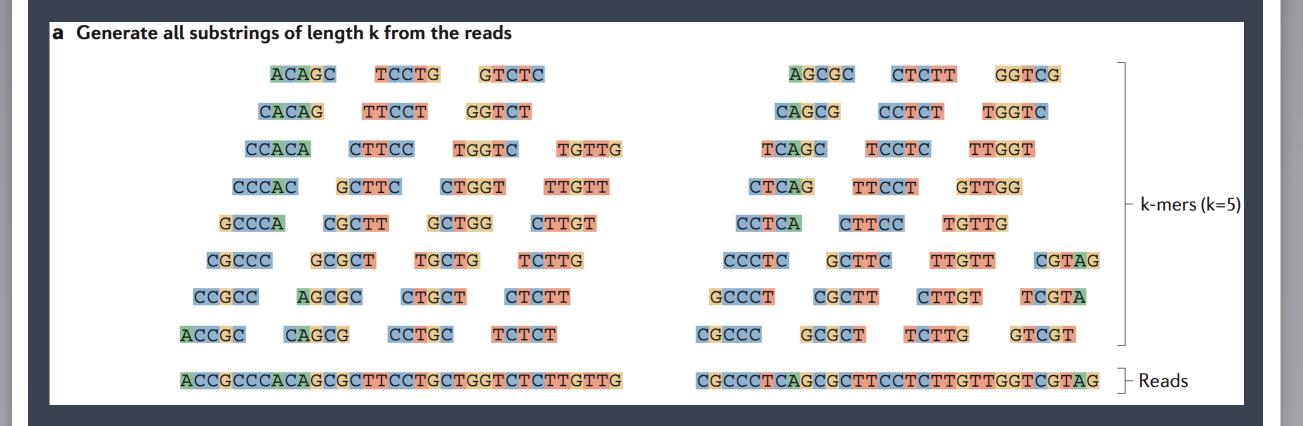
#### De novo assembly

This type of assembly is used when very little information is available for the genome

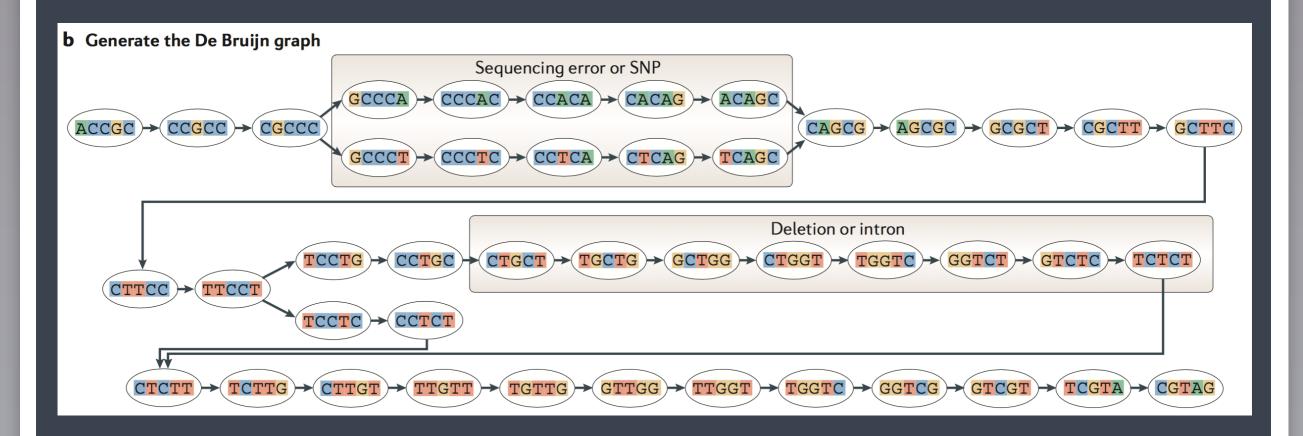
- An assembly of this type is often the first step in putting together information about an unknown genome
- Amount of data needed for a good de novo assembly is higher than what is needed for a reference-based assembly
- Assemblies of this sort can be used for genome annotation, once the genome is assembled
- Oases, TransABySS, Trinity are examples of well-regarded transcriptome assemblers, especially Trinity
- Annotation of any newly-discovered genes or isoforms will be performed downstream

It is not uncommon to use both methods and compare and combine the assemblies, even when a genome sequence is known, especially for a new genome.

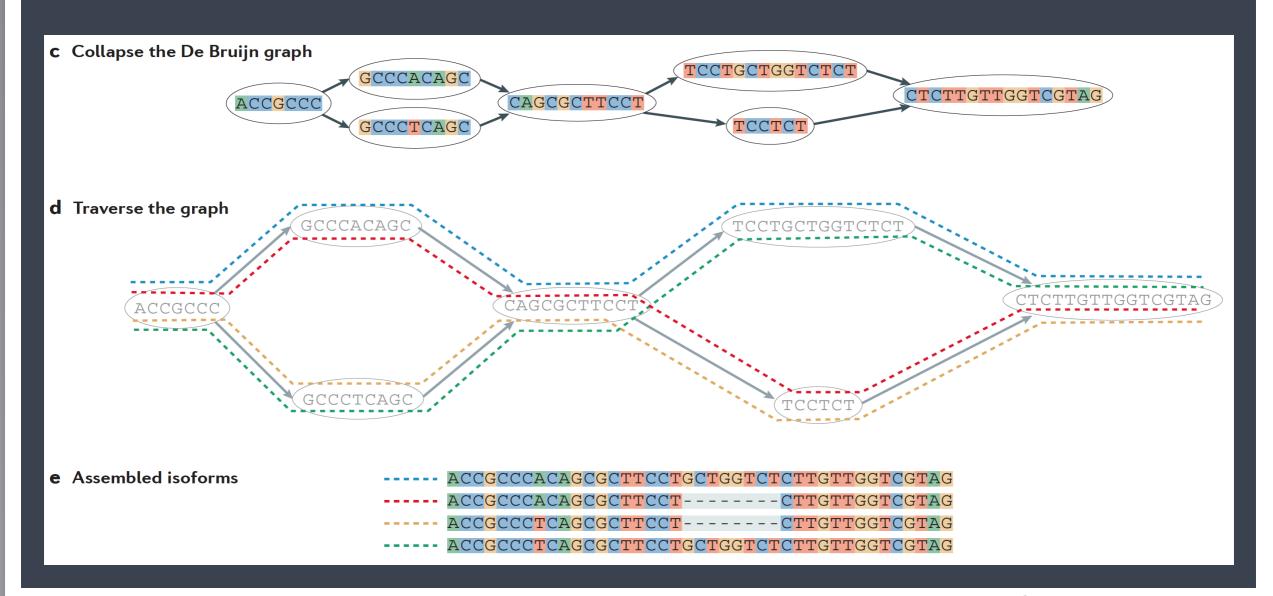
#### De novo assembly

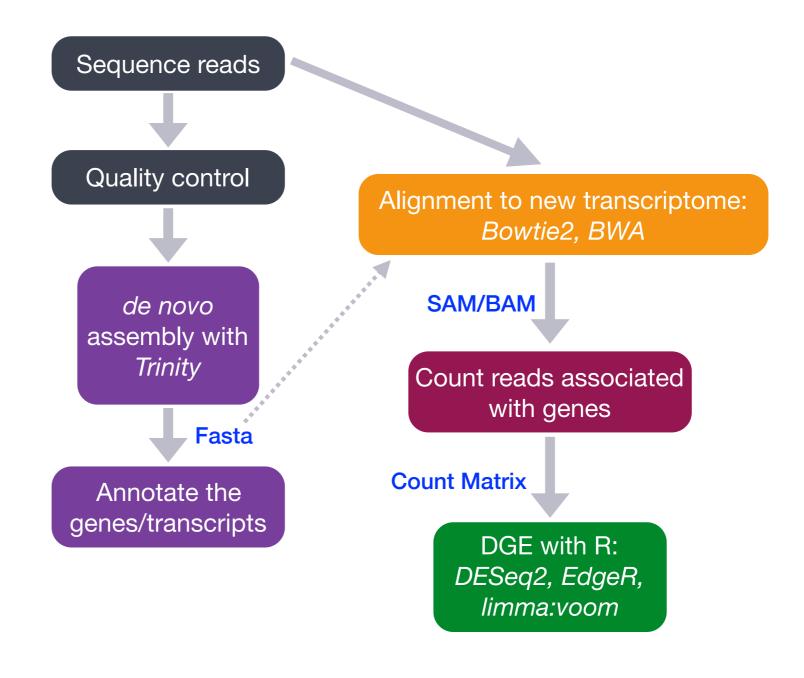


#### De novo assembly (De Bruijn graph construction)



#### De novo assembly (De Bruijn graph construction)





Irrelevant kd (knock down)

Replicate1

Replicate2

Replicate3

Mov10 kd (knock down)

Mov10 oe (over expression)

Replicate2

Replicate1

Replicate2

Replicate3

Replicate3

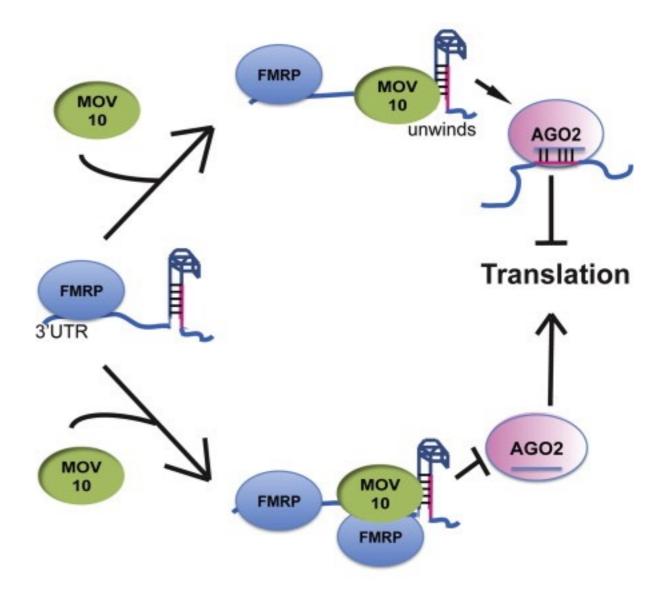
### Dataset for Class

MOV10, is a putative RNA helicase that is also associated with FMRP in the context of the microRNA pathway.

FMRP and **MOV10** associate and regulate the translation of a subset of RNAs

#### Our questions:

- What patterns of expression can we identify with the loss/gain of MOV10?
- Are there any genes shared between the two conditions?



### Biological Question

These materials have been developed by members of the teaching team at the <u>Harvard Chan Bioinformatics Core (HBC)</u>. These are open access materials distributed under the terms of the <u>Creative Commons Attribution license (CC BY 4.0)</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

