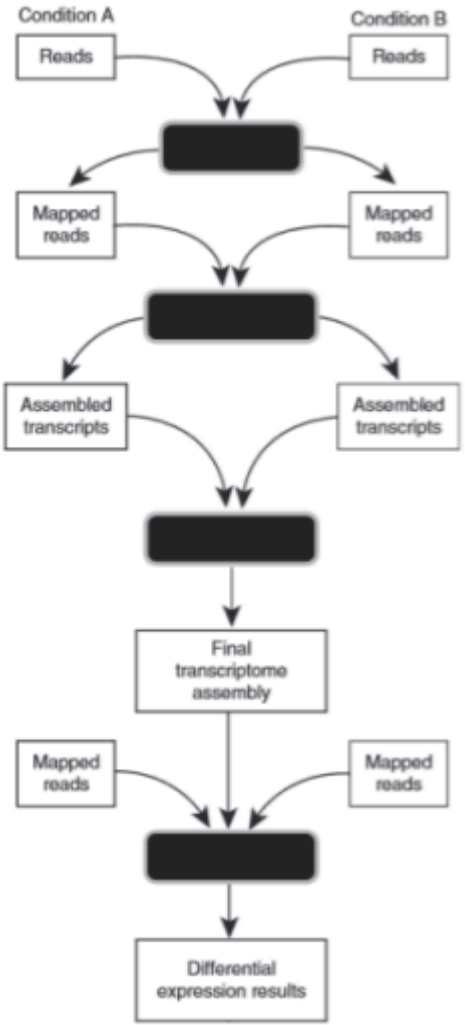
PRACTICAL EXIT TICKET

We applied this workflow in the practical. Fill it with the format file and software used in each step:



FORMAT: Fastaq

SOFTWARE: Tophat

FORMAT: bam

SOFTWARE: Cufflinks

FORMAT: GTF

SOFTWARE: Cuffmerge

FORMAT: GTF

FORMAT: bam

SOFTWARE: Cuffdiff

FORMAT: Tabular

Which biological question are we addressing in this practical?

If the brain and adrenal samples contain any overexpressed or underexpressed genes

Why initial data for each of the two samples come in the form of two separate files? What is the nature of the initial data?

Because they are derived from Illumina's paired-end reads. The forward and reverse strands of the tissue reads from the brain and adrenal glands are separated into two files. The files include the data's sequence and read quality.

Would BWA be a choice for mapping the reads? Why did we use TopHat instead?

Tophat matches RNA-Seq reads to genomes the size of mammals, then evaluates the mapping outcomes to find splice junctions between exons. This is necessary because the intronic sequence between splice sites is still present in a genome's sequence and we should alter the genome sequence to use BWA.

TopHat outputs a file with thesplice junctions. What are they?

The locations in the gene where exons are joined together are known as splice junctions. Where introns are cut during RNA splicing is where these points are located.