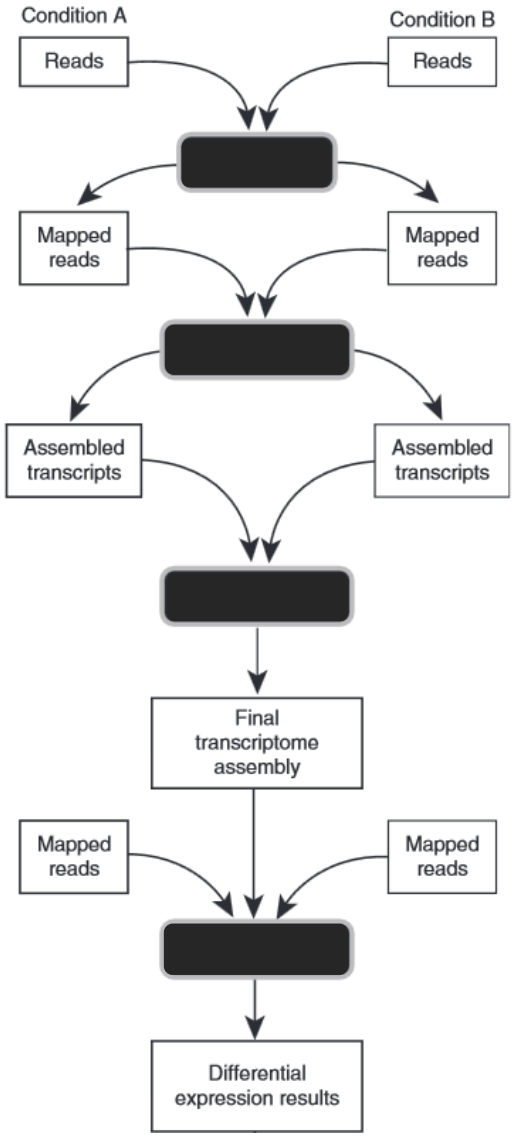
Student: _____

PRACTICAL – EXIT TICKET

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



FORMAT: Fastq

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 there are over expressed or under expressed genes in the adrenal and brain tissue samples

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 come from paired-end reads from illumina. There are 2 files of adrenal and brain tissue reads because each one has the forward and reverse strand. The files contain the sequence and the read quality of the data

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

Because the intronic sequence between splice sites are still present in a genome's sequence and to use bwa we should modify the genome sequence, while Tophat aligns RNA-Seq reads to mammalian-sized genomes and then analyzes the mapping results to identify splice junctions between exons

TopHat outputs a file with the *splice junctions*. What are they?

Splice junctions are the points in the gene where exons are joined together. This points are found where introns are removed during RNA splicing