Aim: To perform RNA seq data analysis of the buffalo genome in 2- celled, 8- celled and blastula stages of the embryo with respect to IVF and Cloned replicates.

Methodology:

1) **Filtering the raw reads:** For quality control, trimming of adapters, filtering by quality, and read pruning of the paired sequences a Fastp software tool was used initially and hence obtained trimmed files.

2) **Aligning the reads:** Hisat-2 is the software tool which was used on the trimmed fastp reads and thereby the whole process of alignment and mapping with the reference genome i.e, with *Bubalas bubalis* was performed resulting, Sequence alignment map (SAM) files.

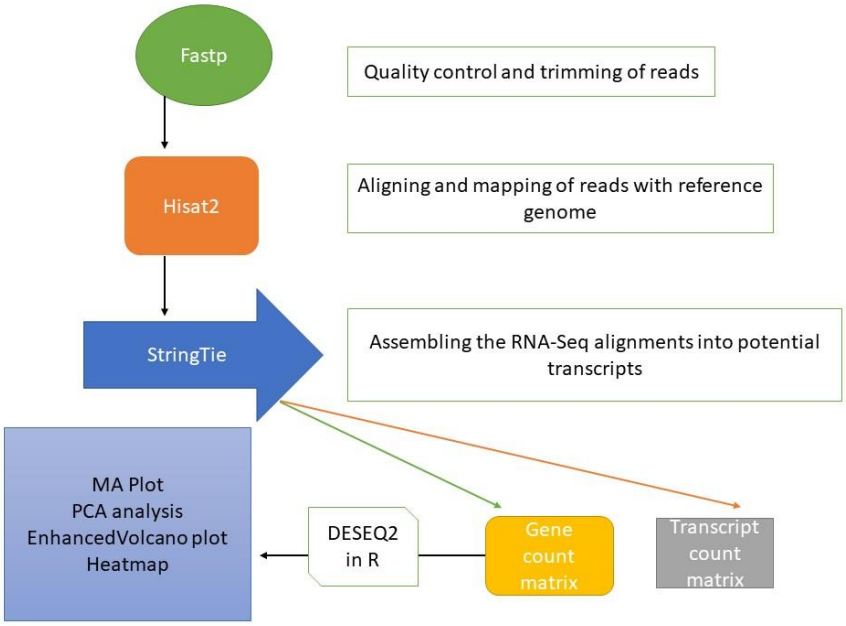
3) **Reading the SAM and BAM files:** SAMtools software was used to read the BAM, SAM files after Hisat-2.

4) **Obtaining gtf files:** StringTie is a software tool used on the BAM files and hence individual Gene transfer files (GTF) were obtained for all the replicates and eventually merged them using StringTie merge function resulting in a StringTie merged GTF file which holds all the information about mRNA transcripts and exons present and helps us to find out novel transcripts present if any, in the obtained data set. A total gene count matrix and transcript count matrix was obtained finally, upon using StringTie yet again on the StringTie merged re-estimated GTF files.

5) **GFFcompare:** GFF compare software was used on the StringTie merged gtf file to obtain the total information about the amount of RNA transcripts present and hence novel transcripts were identified.

6) **DESEQ2 analysis:** Upon using DESEQ2 package on R on the total gene count matrix obtained we could able to normalize the gene counts by applying an adjusted p-value of <0.05 with a Log fold change ratio 2. This is how all the significant genes were obtained depending on the fold change ratios and adjusted p-values in a xlsx file.

7) **R-plots:** Using the xlsx data, plotted MA plot, PCA analysis, Enhanced Volcano plot and Heat map for top 20 significant genes.



***1** Tuxedo pipeline (HISAT2-StringTie-DESEQ2*

References and websites used:

https://academic.oup.com/bioinformatics/article/34/17/i884/5093234

https://www.htslib.org/doc/samtools.html

http://daehwankimlab.github.io/hisat2/manual/

https://ccb.jhu.edu/software/stringtie/index.shtml?t=manual

https://ccb.jhu.edu/software/stringtie/gffcompare.shtml#:~:text=The%20GffCompare%20utility%20The%20program%20gffcompare%20can%20be,with%20a%20reference%20annotation%20%28also%20provided%20as%20GFF%29.