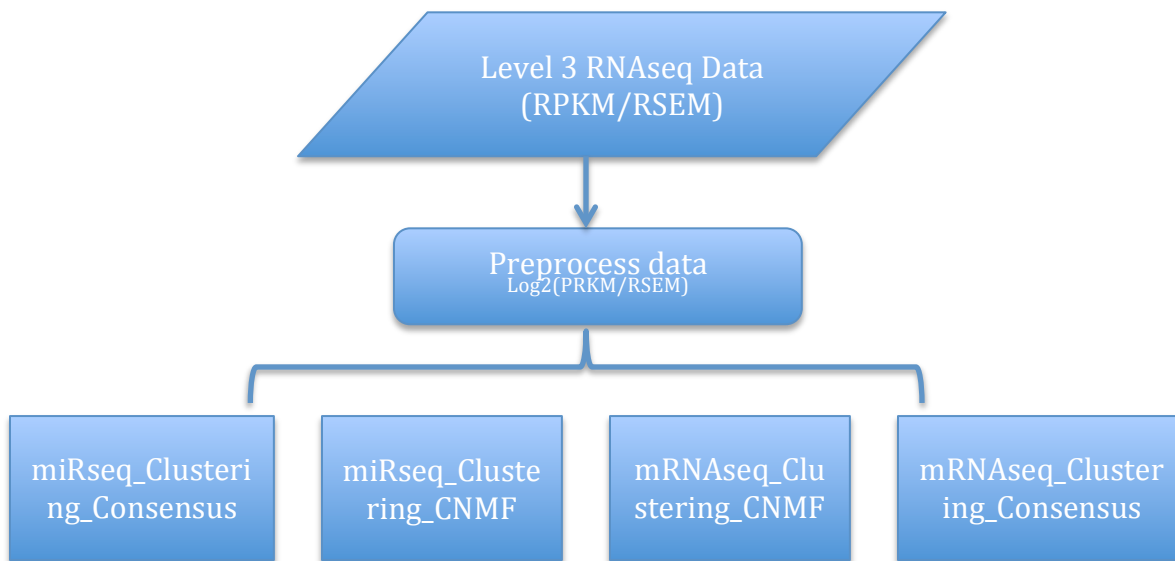


# The flowchart of RNAseq clustering pipeline



\*RPKM for a given GeneX is calculated by:  $(\text{raw read counts} \times 10^9) / (\text{total reads} \times \text{length of GeneX})$ . "Total reads" is the lane yield after removing poor quality reads and the length of GeneX is defined as the median length of all transcripts associated with GeneX.