**Pipeline for TCR analysis**

I have set up a directory “C:\TCR” with subdirectories *input* and *output* so as to facilitate processing data and speed up running. All scripts and necessary accessories are in Decombinator folder on github

1. Download Illumina output files from GridScaler to *input*. If these are the Illumina demultiplex files combine them into three files (one for each read) using **Packagor.R**.
2. Upload **Packagor.R** output to RawData folder on GridScaler. Also move files from *output* to *input.*
3. Check file names; the first read should be R1, the second should be R2 and contain index (small file) and the third should be R3 and contain reverse read. NOTE this nomenclature differs in different sequence runs. But **Packagor.R**  always gives this output.
4. Run **DemultiplexoR.R** . User-set variables include name of experiment; path for index files; input and output path (default is *input* and *output)*; position of barcode (depends on ligation oligo). And potentially also position of second index (default is 7-12 on read 1).
5. Upload output to GridScaler, folder *demultiplex\_BC*
6. Use **Z-zip** to unzip all files from *output* a folder containing **Decombinator**V2, preferably on C drive (for faster running).
7. Run **Decombinate.bat** in this folder, removing all switches except of – TRUE (keep out-of-frame sequences). Remove plots and also complete translation etc. If analyzing mouse sequences, use **decombinatem.bat**.
8. Move output folders (one for each index combination, i.e. biological sample) out of Decombinator folder into a new analysis folder. Process each file using **CollapsinatoR.R**; this script removes PCR duplicates using the barcode information, and calculates unique TCR RNA molecules; and also unique TCR numbers. The script saves all output files (unique TCR RNA molecules as a six part descriptor, together with a barcode) together in one list file for future use.
9. Process all files output from **Collapsinator** using CDR3extractoR, which adds two columns containing full length aminoacid sequences and CDR3.