**Pipeline for TCR analysis**

1. Download raw data files (3 fastqz files) from GridScaler onto a folder on C drive (for fast running).
2. Open new working folder (e.g. in R folder) which will contain output
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.
4. Set path to output, data; position of barcode, and position of second index in **DemultiplexoR.R** script in github folder *Decombinator*.
5. Run the script. Run summary script which gives file with frequencies.
6. Use **Z-zip** to unzip all files output from Demultiplexor to the folder containing **Decombinator** , 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.