

File name convention

- [PD|AD|HD|HC]_[brainBankID]_<batchID>_<rep #>.[R1|R2].fastq<.gz>
 - Beginning with patient type, in abbreviation;
 - Use unique brainBankID rather than the lane# or date
 - batchID and rep# are optional, in format of (if any)
 - batchID: batch1, batch2 etc.
 - rep#: rep1 is for technique replicate, Rep1 for biological replicate
 - Use R1 or R2 to tell the pair-end sequencing
 - Use unzipped **or** zipped fastq, but NOT both.

Folder structure

- RNAseq | smallRNA | ChIPseq | DNAseq
 - rawfiles
 - Raw files from sequencing;
 - Making soft links for conventional file name
 - Log file (readme.txt or Excel)
 - filtered
 - Filtered files (e.g. adaptor removal/clip)
 - run_output
 - sample1
 - Output of Tophat/Cufflinks/htseq-count runs
 - sample2 etc.
 - for_display
 - Files used for display on UCSC / IGV, such as *.bam, *.bam.bai, *.bw, *.gtf etc.
 - Can be soft links to the output files

Project folder structure (Personal suggestion)

- \$HOME
 - bin
 - ...
 - projects
 - PD
 - ...
 - HD
 - src
 - data
 - result
 - document