

Build Folder Structures From Sample Sheet for rcbio NGS Workflows

- Login to O2
- Start an interactive session
- Make a folder to work in
- Copy testing sample sheet to work directory
- Examine sample sheet on local computer
- Load related modules
- Build folder structure from the sample sheet
- Look at the folder structure
- Now you are ready to run an rcbio workflow

Note: You can copy and paste all the text to your Linux command line to run. Anything with "#" is comment, and will be IGNORED by Linux.

Login to O2

```
# replace user123 with your eCommons ID
ssh user123@o2.hms.harvard.edu
```

Start an interactive session

```
# this command requests a job in the interactive partition, with one
processor for 2 hours
srun --pty -p interactive -t 0-02:0:0 --mem 2000MB -n 1 /bin/bash
```

Make a folder to work in

```
# Make a directory in scratch2 file system and work there. We recommend
creating separate folders for each project.
mkdir -p /n/scratch2/$USER/test && cd /n/scratch2/$USER/test
```

Note: Each user has 10 TB /n/scratch2 space. There is no backup for data saved in/n/scratch2, and files will be deleted if they are not accessed for a month. You can read more about /n/scratch2 on the [Filesystems](#) page.

Copy testing sample sheet to work directory

```
cp /n/shared_db/misc/rcbio/data/fruitFlyFastq/sampleSheet.xlsx .
```

Examine sample sheet on local computer

The sample sheet is in Microsoft Excel format. You can look at this file by transferring it to your local computer, and opening in Excel. Programs that can be used to transfer the sample sheet include Filezilla or WinSCP. For help on transferring files to or from the O2 cluster, please read the File Transfer wiki page.

Group Number (must be 1 2 3 and so on)	Sample Name (could be any name without space)	Library Name (could be any name without space)	Lane (could be any name without space)	Path to library read file without space	Read1 file name without space	Read2 file name without space
1	tumor1	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
1	tumor1	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830	twoMillionReads_1.fq	twoMillionReads_2.fq
1	tumor2	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
1	tumor2	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855	twoMillionReads_1.fq	twoMillionReads_2.fq
1	tumor3	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
1	tumor3	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211	twoMillionReads_1.fq	twoMillionReads_2.fq
2	normal1	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
2	normal1	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208	twoMillionReads_1.fq	twoMillionReads_2.fq
2	normal2	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
2	normal2	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210	twoMillionReads_1.fq	twoMillionReads_2.fq
2	normal3	lib1	lane1	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212	anotherTwoMillionReads_1.fq	anotherTwoMillionReads_2.fq
2	normal3	lib1	lane2	/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212	twoMillionReads_1.fq	twoMillionReads_2.fq

Load related modules

```
module load gcc/6.2.0 python/2.7.12 rcbio/1.1

#####
note: rcbio/1.1 is not available yet. If you would like to try the version
still in development stage, you can run:
export PATH=/home/ld32/rcbioDev/bin:$PATH
```

Build folder structure from the sample sheet

```
buildSampleFoldersFromSampleSheet.py sampleSheet.xlsx
```

Look at the folder structure

```
ls -l group*/*/
group2/normal3/:
total 16K
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212/twoMillionRead
s_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212/twoMillionRead
s_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212/anotherTwoMill
ionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558212/anotherTwoMill
ionReads_1.fq

group2/normal2/:
total 16K
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210/twoMillionRead
s_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210/twoMillionRead
s_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210/anotherTwoMill
ionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558210/anotherTwoMill
ionReads_1.fq

group2/normal1/:
total 16K
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208/twoMillionRead
s_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208/twoMillionRead
s_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208/anotherTwoMill
ionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group2/ERR558208/anotherTwoMill
ionReads_1.fq

group1/tumor3/:
total 16K
```

```
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211/twoMillionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211/twoMillionReads_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211/anotherTwoMillionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR558211/anotherTwoMillionReads_1.fq
```

group1/tumor2/:

total 16K

```
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855/twoMillionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855/twoMillionReads_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855/anotherTwoMillionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435855/anotherTwoMillionReads_1.fq
```

group1/tumor1/:

total 16K

```
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830/twoMillionReads_2.fq
lrwxrwxrwx 1 ld32 ld32 80 Jan 16 11:59 lib1_lane2_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830/twoMillionReads_1.fq
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_2.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830/anotherTwoMillionReads_2.fq
```

```
lrwxrwxrwx 1 ld32 ld32 87 Jan 16 11:59 lib1_lane1_1.fq ->
/n/shared_db/misc/rcbio/data/fruitFlyFastq/group1/ERR435830/anotherTwoMill
ionReads_1.fq
```

Now you are ready to run an rcbio workflow

To instead run workflow on your own data, transfer the sample sheet to your local machine following this [wiki page](#) and modify the sample sheet. Then you can transfer it back to O2 under your account, then go to the [build folder structure](#) step.