Pooling CaptureC samples without a pipeline !
Instructions by Jelena - 040816

 # Make a directory

```
mkdir pooling
```

 # Go to the directory

```
cd pooling
```

 # Make a directory for all your samples-to-be-pooled

```
mkdir E14-2 E14-3 E14-4
```

 # Copy the F6 gff files (if CC4 output) or F4 files (if CC3 output) to the folders you created :

```
cp /where/my/CC4/run/is/E14-2/F6*/*.gff E14-2/.
cp /where/my/CC4/run/is/E14-3/F6*/*.gff E14-3/.
cp /where/my/CC4/run/is/E14-4/F6*/*.gff E14-4/.
```

 # Check that you have them

```
tree
```

 # Add sample name to the file names :

```
cd E14-2

/home/molhaem2/telenius/CC/norm/VS100/addToName.sh E14-2

cd ..
cd E14-3

/home/molhaem2/telenius/CC/norm/VS100/addToName.sh E14-3

cd ..
cd E14-4

/home/molhaem2/telenius/CC/norm/VS100/addToName.sh E14-4

cd ..
```

 # Now we have unique file names, and can move all the files to same folder :

```
cp */*.gff .
```

 # Now we moved all the files away from the sample directories, and can delete those empty dirs :

```
rmdir E14-2
rmdir E14-3
rmdir E14-4
```

 # Now we can list what we have :

```
ls
```

- You should see all your gff files in same folder, named by the sample name (beginning of name), and by capture point (end of name), like this :

```
E14-2_COMBINED_A16.gff
E14-2_COMBINED_A18.gff
.
.
.
E14-4_COMBINED_A20.gff
```

 # Now we need to separate them based on the CAPTURE point :

```
mkdir A16 A18 A20
```

```
mv *A16.gff A16/.
mv *A18.gff A18/.
mv *A20.gff A20/.
```

 # Check that it worked :

```
tree
```

 # The following steps needs to be done for ALL the capture point folders

```
A16 A18 A20
```

 # Go to the folder :

```
cd A16
```

 # Check that you have your files :

```
ls
```

```
E14-2_COMBINED_A16.gff
E14-3_COMBINED_A16.gff
E14-3_COMBINED_A16.gff
```

 # First we will run the first parts of the normalisation with a script I wrote :

```
/home/molhaem2/telenius/CC/norm/VS100/run.sh &> normalisation.log
```

The above command normalises the data for 100 000 reads, using James' codes.

Now you should have file

```
ls | grep gfc
```

normalised.gfc.out

Check that you got no error messages in

```
cat normalisation.log
```

The subsequent steps depend on the ordering of columns in the file - so we prepare our data for those steps manually !

 # Generating the sum of all samples !

```
/home/molhaem2/telenius/CC/norm/VS100/sum.sh &> summing.log
```

Check that you got no error messages in

```
cat summing.log
```

 # The last step is to make the visualisation files for all our data !

We should have files :

```
ls | grep gfc
```

normalised.gfc.out
normalised_sum.gfc.out

We need to generate data hub for these tracks - to see them in USCS !

```
/home/molhaem2/telenius/CC/norm/VS100/hubPoolSamples.sh --genome mm9 --hubname testhub030816 --pf /public/telenius/capturetests/hub030816 &> hubbing.log
```

Check that you didn't get any error messages :

```
cat hubbing.log
```

The hub address is given in the end of the log file :

```
tail hubbing.log
```