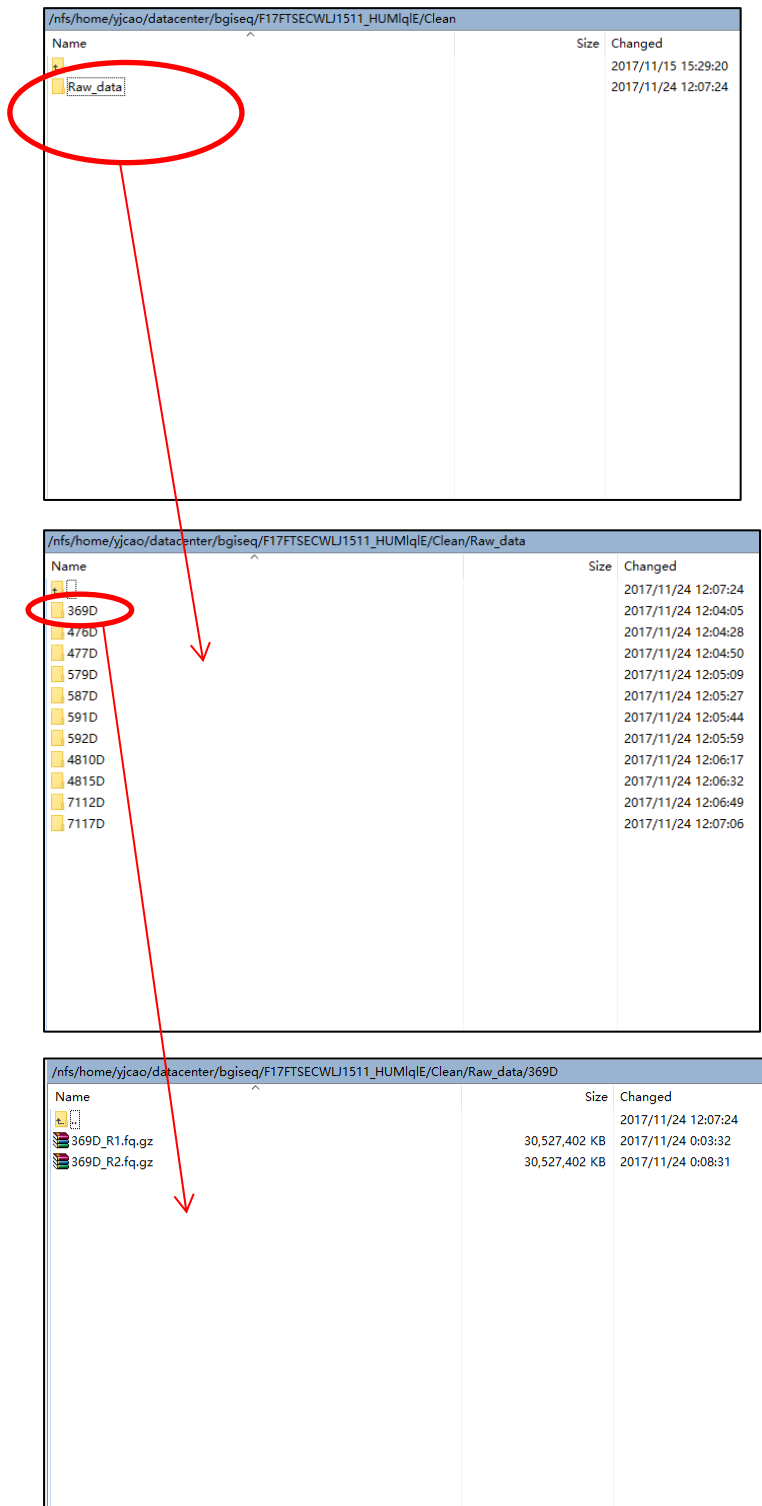


Easy WES tools by Cao Yujie 2017-11-29

- Upload your Raw_data to the project folder naming the folder Raw_data as show below:



- Your file name should be in the format of SampleID+R1orR2.fq.gz, otherwise you need to rename file name using bash command `mv`.

- Upload Easy_WES folder to the project folder, same folder as Raw_data folder as show below:

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean			
Name	Size	Changed	
..		2017/11/15 15:29:20	r
Easy_WES		2017/11/24 12:12:59	r
Raw_data		2017/11/24 12:07:24	r

- Then, cd Easy_WES

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean/Easy_WES			
Name	Size	Changed	
..		2017/11/24 12:12:57	
PBSModels		2017/11/24 12:12:59	
Easy_WES.py	3 KB	2017/11/23 19:38:47	
Easy_WES_configuration.txt	2 KB	2017/11/23 19:09:45	
mammual.docx	252 KB	2017/11/24 12:12:20	
supportingFun.py	5 KB	2017/11/23 19:14:28	

- Then, change configures in file Easy_WES_configuration.txt, such as working path, software path, sample name list and PBSscripts lists:

```

1  ### Configuration file for EasyWES ###
2
3  ##### Users must manually set-up the below information for using EasyWES! #####
4
5  ##project folder##
6  project_folder=/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean
7
8  ## Software paths (please use absolute path with the software executables) ##
9  BWA_path=/home/yjcao/software/bwa-0.7.15/bwa
10 samtools_path=/home/yjcao/software/samtools-1.4/samtools
11 picard_path=/home/yjcao/software/picard.jar
12 GATK_path=/home/yjcao/software/GenomeAnalysisTK.jar
13 GATK_bundle_path=/home/yjcao/reference/GATKbundle/hg19/
14 ANNO_path=/home/yjcao/software/annovar/table_annovar.pl
15
16 ##Your email to get data processing information from CGS root#####
17 Email=boxyjcao@connect.hku.hk
18
19 #####Sample Name List(with each line corresponding to one sample)#####
20 sample1=369D
21 sample2=476D
22 sample3=477D
23 sample4=579D
24 sample5=587D
25 sample6=591D
26 sample7=592D
27 sample8=4810D
28 sample9=4811D
29 sample10=7112D
30 sample11=7117D
31
32 #####PBSfile Name List(with each line corresponding to one sample)#####
33 PBSfile1=Easy_WES_phase1_pre_processing_model.pbs
34 PBSfile2=Easy_WES_phase2_Step1_RunHC_model.pbs
35 PBSfile3=Easy_WES_phase2_Step2_Gting_model.pbs
36 PBSfile4=Easy_WES_phase2_Step3_v2_VQSR_model.pbs
37 PBSfile5=Easy_WES_phase2_Step3_v1_HardFiltering_model.pbs
38
39


```


- Then, cd Easy_WES, run the scripts:


python Easy_WES.py



```
[12:22:26] yjcao@statgenpro:/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Easy_WES $pwd
/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Easy_WES
[12:22:29] yjcao@statgenpro:/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Easy_WES $ll
total 674K
-rw-r--r-- 1 yjcao wlyang 252K Nov 24 12:12 mammual.docx
drwxr-xr-x 2 yjcao wlyang 291 Nov 24 12:12 PBSModels
-rw-r--r-- 1 yjcao wlyang 1.3K Nov 24 12:19 Easy_WES_configuration.txt
-rw-r--r-- 1 yjcao wlyang 4.3K Nov 24 12:20 supportingFun.pyc
-rw-r--r-- 1 yjcao wlyang 4.9K Nov 24 12:21 supportingFun.py
-rw-r--r-- 1 yjcao wlyang 2.4K Nov 24 12:22 Easy_WES.py
[12:22:30] yjcao@statgenpro:/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Easy_WES $python Easy_WES.py
[12:22:38] yjcao@statgenpro:/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Easy_WES $
```



- You will get:



/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean			
Name	Size	Changed	^
		2017/11/15 15:29:20	
Raw_data		2017/11/24 12:07:24	
Bam_file		2017/11/24 12:22:38	
Easy_WES		2017/11/24 12:22:38	
VCF file		2017/11/24 12:22:38	

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Bam_file			
Name	Size	Changed	^
		2017/11/24 12:22:38	
369D		2017/11/24 12:22:38	
476D		2017/11/24 12:22:38	
477D		2017/11/24 12:22:38	
579D		2017/11/24 12:22:38	
587D		2017/11/24 12:22:38	
591D		2017/11/24 12:22:38	
592D		2017/11/24 12:22:38	
4810D		2017/11/24 12:22:38	
4811D		2017/11/24 12:22:38	
7112D		2017/11/24 12:22:38	
7117D		2017/11/24 12:22:38	





/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqLE/Clean/Bam_file/369D			
Name	Size	Changed	^
		2017/11/24 12:22:38	
Easy_WES_phase1_pre_processing_369D.pbs	4 KB	2017/11/24 12:22:38	

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean/VCF_file			
Name	Size	Changed	^
		2017/11/24 12:22:38	
369D		2017/11/24 12:22:38	
476D		2017/11/24 12:22:38	
477D		2017/11/24 12:22:38	
579D		2017/11/24 12:22:38	
587D		2017/11/24 12:22:38	
591D		2017/11/24 12:22:38	
592D		2017/11/24 12:22:38	
4810D		2017/11/24 12:22:38	
4811D		2017/11/24 12:22:38	
7112D		2017/11/24 12:22:38	
7117D		2017/11/24 12:22:38	
Joint		2017/11/24 12:22:38	
 Easy_WES_phase2_Step2_GTing.pbs	3 KB	2017/11/24 12:22:38	

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean/VCF_file/369D			
Name	Size	Changed	^
		2017/11/24 12:22:38	
 Easy_WES_phase2_Step1_RunHC_369D.pbs	2 KB	2017/11/24 12:22:38	

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean/VCF_file/Joint			
Name	Size	Changed	^
		2017/11/24 12:22:38	
 Easy_WES_phase2_Step3_v1_HardFiltering.pbs	5 KB	2017/11/24 12:22:38	

- And under each Raw_data sample folder, you will get fastqc pbs files, such as this case, under /Raw_data/369D/:

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMlqlE/Clean/Raw_data/369D					
Name	Size	Changed	Rights	Owner	
		2017/11/28 23:26:07	rw-r--r--	yjcao	
 Easy_WES_phase0_fastqc_369D.pbs	2 KB	2017/11/29 9:30:40	rw-r--r--	yjcao	
 369D_R2.fastq.gz	11,063,664 KB	2017/11/29 2:17:29	rw-----	yjcao	
 369D_R1.fastq.gz	10,495,216 KB	2017/11/29 1:32:30	rw-----	yjcao	

- If necessary, run fastqc for each sample. In the reporter of fastqc, you need to make sure base quality for fastq files should be Illumina 1.8+ Phred+33(https://en.wikipedia.org/wiki/FASTQ_format). Otherwise you need to correct it use tools like seqtk.

- Submit your tasks to CGS queen

/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqlE/Clean					
Name	Size	Changed	Rights	Owner	
Raw_data		2017/11/15 15:29:20	rw-r--r--	yjcao	
Bam_file		2017/11/28 23:26:07	rw-r--r--	yjcao	
Easy_WES		2017/11/29 9:30:40	rw-r--r--	yjcao	
VCF file		2017/11/29 9:30:40	rw-r--r--	yjcao	
submit.sh	1 KB	2017/11/29 9:36:09	rw-r--r--	yjcao	

```

1  for i in 369D 476D 477D 579D 587D 591D 592D 4810D 4815D 7112D 7117D
2  do
3      cd /nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUMLqlE/Clean/Bam_file/$i
4      qsub Easy_WES_phase1_pre_processing "$i".pbs
5  done
6

```

- Only manually change the highlighted folders or prefix, and run the following bash commands:

sh submit.sh

in the following orders:

```
cd /path/to/projectfolder/Raw_data/$i
qsub Easy_WES_phase0_Fastqc "$i".pbs
```

```
cd /path/to/projectfolder/Bam_file/$i
qsub Easy_WES_phase1_pre_processing "$i".pbs
```

```
cd /path/to/projectfolder/VCF_file/$i
qsub Easy_WES_phase2_Step1_RunHC "$i".pbs
```

Under VCF_file folder, qsub Easy_WES_phase2_Step2_GTing.pbs, then under /VCF_file/Joint there will be a file named All_Genotyping.vcf.

Then if your sample size is large than 30, qsub the Easy_WES_phase2_Step3_v2_VQSR.pbs, otherwise qsub Easy_WES_phase2_Step3_v1_HardFiltering.pbs Both of them will perform filtering on the callset, and do the genotype

refinement, followed by annotate the callset.

```
621659.statgenpro    ...re_processing yjcao    23:30:09 R medium
621660.statgenpro    ...re_processing yjcao    27:18:01 R medium
621661.statgenpro    ...re_processing yjcao    28:35:35 R medium
621662.statgenpro    ...re_processing yjcao    21:20:56 R medium
621663.statgenpro    ...re_processing yjcao    20:58:07 R medium
621664.statgenpro    ...re_processing yjcao    15:45:18 R medium
621665.statgenpro    ...re_processing yjcao    21:22:53 R medium
621666.statgenpro    ...re_processing yjcao    18:19:53 R medium
621667.statgenpro    ...re_processing yjcao    22:10:55 R medium
621668.statgenpro    ...re_processing yjcao    17:04:38 R medium
621669.statgenpro    ...re_processing yjcao    0 Q medium

[15:17:21] yjcao@statgenpro:/nfs/home/yjcao/datacenter/bgiseq/F17FTSECWLJ1511_HUNlqlE $
```

