Log of changes pipeline script

Блокнот: Laboratory Journal - Protein Kitchen
Создана: 15.07.2014 13:20 **Изменена:** 13.08.2014 12:01

Used OS :
> uname -a
Linux tux-meerkat 3.13.0-30-generic #55-Ubuntu SMP Fri Jul 4 21:40:53 UTC 2014 x86_64 x86_64 x86_64 GNU/Linux

How to run script. Now just create a folder in your HOME directory where will locate all data. Named as "data". In this folder you could create three folders - "RUN", "OUT" and "SRC".
In RUN folder place the follow scripts: *DamID-seq_pipeline.sh* , *align_local_fq.sh* , *reads2GATC.sh* . And also .gff&config file: *parameterfile* , *DmelGATCfragments-r5_LP120507.gff*
Into *parameterfile* place this text:

> #by example, please change to your species

> SPECIES=fly
> FASTQ_FILES='~/data/SRC/*.fastq.gz'
> #by example, please change to your assembly
> ASSEMBLY=dm3
> OUTPUT_DIR=~/data/OUT

Please run the follow commands:
> cd ~/data/RUN
> chmod +x *.sh
> ./DamID-seq_pipeline.sh parameterfile

Then you need to install some program's:

> sudo apt-get update
> sudo apt-get install bowtie bowtie2 fastx-toolkit samtools python-htseq python-pip

Add this line to the end of file /etc/apt/sources.list:
deb http://cran.r-project.org//bin/linux/ubuntu trusty/

Then:
> gpg --keyserver hkp://keyserver.ubuntu.com:80 --recv-keys E084DAB9
> gpg -a --export E084DAB9 | sudo apt-key add -

> sudo apt-get update
> sudo apt-get install r-base r-bioc-iranges r-bioc-biocgenerics parallel python-dev libevent-dev
> sudo pip install upgrade pip
> sudo pip unstall cutadapt

Some changes that I made in the script *align_local_fq.sh* :

line 40-44:

```
BOWTIE2=bowtie2
CUTADAPT=cutadapt
FASTX_REVCOM=fastx_reverse_complement
BOWTIE2_INDEXES=/home/anton/data/indexes/
```

line 65-66: (remove -e option of 'echo' command)

```
ADPTR_SHORT_3=`echo ">\n${ADPTR_SHORT_5}" | ${FASTX_REVCOM} | awk 'NR > 1'`
ADPTR_LONG_3=`echo ">\n${ADPTR_LONG_5}" | ${FASTX_REVCOM} | awk 'NR > 1'`
```

line 114-120: (remove before tmp_* primary slash and point)**

```
TMP_BAM_INNER=`mktemp tmp_bam_inner.XXXXXXXXXX`
TMP_BAM_EDGE=`mktemp tmp_bam_edge.XXXXXXXXXX`
TMP_STATS_INNER=`mktemp tmp_stats_inner.XXXXXXXXXX`
TMP_STATS_EDGE=`mktemp tmp_stats_edge.XXXXXXXXXX`
TMP_FQ=`mktemp tmp_fq.XXXXXXXXXX`
TMP_FQ_EDGE=`mktemp tmp_fq_edge.XXXXXXXXXX`
TMP_FQ_INNER=`mktemp tmp_fq_inner.XXXXXXXXXX`
```

line 181: (commented line "${CAT} ${IN_FQ} |", and changed variable)

```
${CUTADAPT} -g "${ADPTR_LONG_5}" -a "${ADPTR_LONG_3}" -O "${LEN_THRES}" --match-read-wildcards --discard-trimmed ${IN_FQ} -o "${TMP_FQ}" > ${CLIP_STATS}
```

line 193-196: (-p option, change the number of cores your processor)

```
#Get number of CPU Cores
CORE=`lscpu | grep 'CPU(s):' | sed -n '1p' | rev | cut -c 1`
#Run Bowtie
BOWTIE_PAR="-k 3 -p ${CORE} -t --phred33 --local -x ${BOWTIE2_INDEXES}${ASSEMBLY}"
```