1. **下載sra file**

# prefetch SRR975551 SRR975552 …

1. **當拿到reads資料(.sra)時，須將sra轉檔為fastq (sample.sra ->** **sample.fastq)**

若此reads為single-end(SE) : fastq-dump sample.sra

**一次轉多個檔**

# bash dump\_SE\_multi.sh ~/ncbi/public/sra/

若此reads為paired-end(PE) : fastq-dump --split-3 sample.sra

(切割成upstream,downstream)

**一次轉多個檔**

# bash dump\_PE\_multi.sh ~/ncbi/public/sra/

1. **(已將多支程式彙整在RSEM.sh，執行RSEM.sh即可算出表現量並匯入資料庫)**

**一次算一個fastq**

# bash ~/RNA\_data/RSEM.sh \

~/RNA\_data/RNAseq/sample.fastq \

trim ( or untrim) \

single-end ( or paired-end)

(# bash RSEM.sh 序列原始檔 要不要去除adapter SE或PE)

**一次算多個fastq**

# bash ~/RNA\_data/RSEM\_multi.sh \

~/RNA\_data/RNAseq/sample.fastq \

~/RNA\_data/RNAseq/RNA\_GFP.fastq \

~/RNA\_data/RNAseq/RNA\_GFPQ1.fastq \

trim (or untrim) \

single-end (or paired-end)

**以此例子為例，資料庫將產生**

sample\_genome\_genes RNA\_GFP\_genome\_genes

sample\_genome\_isoform RNA\_GFP\_genome\_isoform

sample\_transcriptome\_genes RNA\_GFP\_transcriptome\_genes

sample\_transcriptome\_isoform RNA\_GFP\_transcriptome\_isoform

RNA\_GFPQ1\_genome\_genes

RNA\_GFPQ1\_genome\_isoform

RNA\_GFPQ1\_transcriptome\_genes

RNA\_GFPQ1\_transcriptome\_isoform

以輸入~/RNA\_data/RNAseq/sample.fastq為例，將輸出下列資料夾與檔案

~/EXP\_output/sample\_EXP\_folder

|--sample\_fastqc\_report

|--sample\_fastqc.html

|--sample\_fastqc.zip

|--sample\_cutadapt\_report

If trim

|--cutadapt\_report

|--sample\_fastqc\_report\_trimmed

|--sample\_trimmed\_fastqc.html

|--sample\_trimmed\_fastqc.zip

|--sample\_result\_genome

|--report

|--sample.genes.results

|--sample.isoforms.results

|--sample.stat

|--sample.transcript.bam

|--sample\_result\_transcriptome

|--report

|--sample.genes.results

|--sample.isoforms.results

|--sample.stat

|-- sample.transcript.bam

~/RNA\_data

|--RNAseq

|--sample.fastq (raw data)

|--RNAseq\_trimmed

|--sample\_trimmed.fq

1. **執行edgeR**

**(已將多支程式彙整在edgeR.sh，執行edgeR.sh即可算出表現量並匯入資料庫)**

# bash ~/Django/edward\_project/libraries/edgeR/edgeR.sh \

-C sample1 sample2 sample3 …….\ #控制組之檔案

-E sample4 sample5 sample6 …….\ #實驗組之檔案

-P GSE50760 \ #paper名稱

-M normal\_colon primary\_colorectal\_cancer \ #condition1 , condition2

-L genome\_isoforms #mapping leveld

**以上述為例，將輸出下列檔案與表格**

~/EXP\_output/edgeR/GSE50760

|--genome\_isoforms

|--ExperimentalvsControl\_AddGene (#isoform level才有)

|--ExperimentalvsControl\_rawResult

|--ExperimentalvsControl\_pVal0.05

|--ExperimentalvsControl\_cpm

MySQL

|--edward\_database

|--更新table name 至 Mutual\_Relationship

|--edward\_EdgeR\_database

|--edgeR\_GSE50760\_normal\_primary\_genome\_isoforms

1. **將所有RSEM跑完之結果抽出FPKM存入資料庫(畫boxplot)**

#python ~/Django/edward\_project/libraries/RSEM\_FPKM\_gather/RSEM\_FPKM\_gather.py \

-G ~/EXP\_output/SRR975551\_EXP\_folder/SRR975551\_result\_genome/SRR975551.genes.results ……..\(primary) ~/EXP\_output/SRR975568\_EXP\_folder/SRR975568\_result\_genome/SRR975568.genes.results ….....\(normal)

~/EXP\_output/SRR975586\_EXP\_folder/SRR975586\_result\_genome/SRR975586.genes.results ……..\ (metastasized)

-SG 18-18-18 \

-T ~/EXP\_output/SRR975551\_EXP\_folder/SRR975551\_result\_transcriptome/SRR975551.genes.results ……..\(primary)

~/EXP\_output/SRR975568\_EXP\_folder/SRR975568\_result\_transcriptome/SRR975568.genes.results ……..\(normal)

~/EXP\_output/SRR975586\_EXP\_folder/SRR975586\_result\_transcriptome/SRR975586.genes.results ……..\ (metastasized)

-ST 18-18-18 \

-P GSE50760 \

-C primary normal metastasized \

-L genes

檔案所屬之condition須按照-C 內之condition排序

以此範例為例，資料庫產生GSE50760\_gather\_genes

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| genes | primary\_  genome\_  genes | normal\_  genome\_  genes | metastasized\_  genome\_  genes | primary\_  transcriptome\_  genes | normal\_  transcriptome \_genes | metastasized\_  transcriptome\_  genes |
| A1BG  .  .  . | 0,0,0,0,0,0  ,0,0,0,0,0,0  ,0,0,0,0,0,0 | 0,1,0,5,0,0  ,1,0,0,4,0,3  ,0,0,5,0,0.5,0 | 0.8,1.2,0,5,0.7,0  ,1,0,0,4,0,3  ,0,0,5,0,0.5,0 |  |  |  |