CANB 7640 : RNAseq Analysis Example

Jihye Kim & Aik Choon Tan

Data

- GSE80802 (DeGregori Lab, PMID: 27663586)
  - Gefitinib, AZ1366 (tankyrase inhibitor) or combination on NSCLC (HCC4006)
  - Chr 7 ONLY for our RNAseq example analysis
  - No Drug (Sample1) vs Gefitinib (Sample5)

- Data
  - Amc-genesis1/Volumes/Data/Users/guest00
  - sample1_chr7.fastq, sample5_chr7.fastq

Methods

- Tophat
  - https://ccb.jhu.edu/software/tophat/index.shtml
  - Tophat 1.4.1 for our example
- Cufflinks
  - http://cole-trapnell-lab.github.io/cufflinks/
  - Cufflinks 2.1.1 for our example
- Mapping Library
  - HG19 (human 37.2)
  - GTF, Fasta (index)
  - amc-genesis1:/Volumes/Data/CANB7640/Bowtie_Index
    - *.ebwt
    - *.fa
    - *.fa.fai
    - *.gtf

```
tophat -p 8 --library-type fr-unstranded -z 0 --GTF
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2.gtf -o sample1
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2 sample1_chr7.fastq


cufflinks -p 8 --GTF-guide
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2.gtf --frag-bias-correct
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2.fa --multi-read-correct --
library-type fr-unstranded --min-isoform-fraction 0.1 --min-frags-per-transfrag 50 --no-
update-check -o sample1 sample1/accepted_hits.bam


cuffdiff -p 8 -L CTRL,Gefitinib -o ./CTRLvsGEF -b
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2.fa -u
/Volumes/Data/CANB7640/Bowtie_index/illumina_human_37_2.gtf
sample1/accepted_hits.bam sample5/accepted_hits.bam
```

Outputs :
  sample1/
  sample5/
  CTRLvsGEF/

```
amc-genesis1:CTRLvsGEF kimji$ ls -l
total 77408
-rw-r--r--+ 1 kimji  staff   165419 Oct 24 10:33 bias_params.info
-rw-r--r--+ 1 kimji  staff   841413 Oct 24 10:33 cds.count_tracking
-rw-r--r--+ 1 kimji  staff  1556057 Oct 24 10:33 cds.diff
-rw-r--r--+ 1 kimji  staff  2177026 Oct 24 10:33 cds.fpkm_tracking
-rw-r--r--+ 1 kimji  staff  1688422 Oct 24 10:33 cds.read_group_tracking
-rw-r--r--+ 1 kimji  staff  2226885 Oct 24 10:33 cds_exp.diff
-rw-r--r--+ 1 kimji  staff  2067153 Oct 24 10:33 gene_exp.diff
-rw-r--r--+ 1 kimji  staff   780099 Oct 24 10:33 genes.count_tracking
-rw-r--r--+ 1 kimji  staff  2045949 Oct 24 10:33 genes.fpkm_tracking
-rw-r--r--+ 1 kimji  staff  1559447 Oct 24 10:33 genes.read_group_tracking
-rw-r--r--+ 1 kimji  staff  3277239 Oct 24 10:33 isoform_exp.diff
-rw-r--r--+ 1 kimji  staff  1369748 Oct 24 10:33 isoforms.count_tracking
-rw-r--r--+ 1 kimji  staff  3302702 Oct 24 10:33 isoforms.fpkm_tracking
-rw-r--r--+ 1 kimji  staff  2747770 Oct 24 10:33 isoforms.read_group_tracking
-rw-r--r--+ 1 kimji  staff  2035658 Oct 24 10:33 promoters.diff
-rw-r--r--+ 1 kimji  staff      220 Oct 24 10:33 read_groups.info
-rw-r--r--+ 1 kimji  staff      326 Oct 24 10:33 run.info
-rw-r--r--+ 1 kimji  staff  2386355 Oct 24 10:33 splicing.diff
-rw-r--r--+ 1 kimji  staff  2418681 Oct 24 10:33 tss_group_exp.diff
-rw-r--r--+ 1 kimji  staff   933818 Oct 24 10:33 tss_groups.count_tracking
-rw-r--r--+ 1 kimji  staff  2355321 Oct 24 10:33 tss_groups.fpkm_tracking
-rw-r--r--+ 1 kimji  staff  1869572 Oct 24 10:33 tss_groups.read_group_tracking
-rw-r--r--+ 1 kimji  staff  1778162 Oct 24 10:10 var_model.info
```

```
amc-genesis1:sample1 kimji$ ls -l
total 258048
-rw-r--r--+ 1 kimji  staff  47029276 Oct 19 11:17 accepted_hits.bam
-rw-r--r--+ 1 kimji  staff     12872 Oct 19 11:17 deletions.bed
-rw-r--r--+ 1 kimji  staff   1451493 Oct 19 13:04 genes.fpkm_tracking
-rw-r--r--+ 1 kimji  staff      5513 Oct 19 11:17 insertions.bed
-rw-r--r--+ 1 kimji  staff   2517564 Oct 19 13:04 isoforms.fpkm_tracking
-rw-r--r--+ 1 kimji  staff    486693 Oct 19 11:17 junctions.bed
-rw-r--r--+ 1 kimji  staff        68 Oct 19 10:49 left_kept_reads.info
drwxr-xr-x+ 21 kimji  staff       714 Oct 19 11:22 logs
-rw-r--r--+ 1 kimji  staff         0 Oct 19 12:55 skipped.gtf
-rw-r--r--+ 1 kimji  staff  80508470 Oct 19 13:04 transcripts.gtf
-rw-r--r--+ 1 kimji  staff     87877 Oct 19 11:17 unmapped_left.fq
```
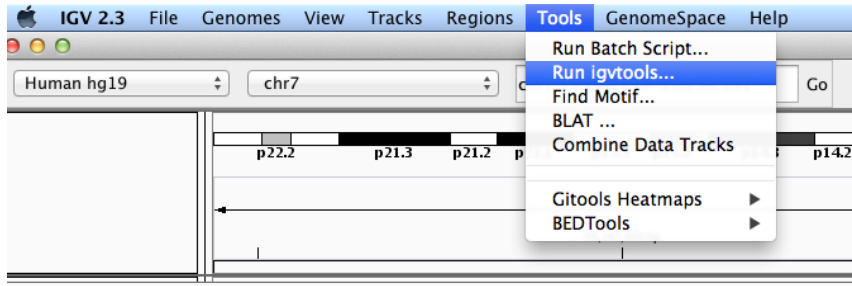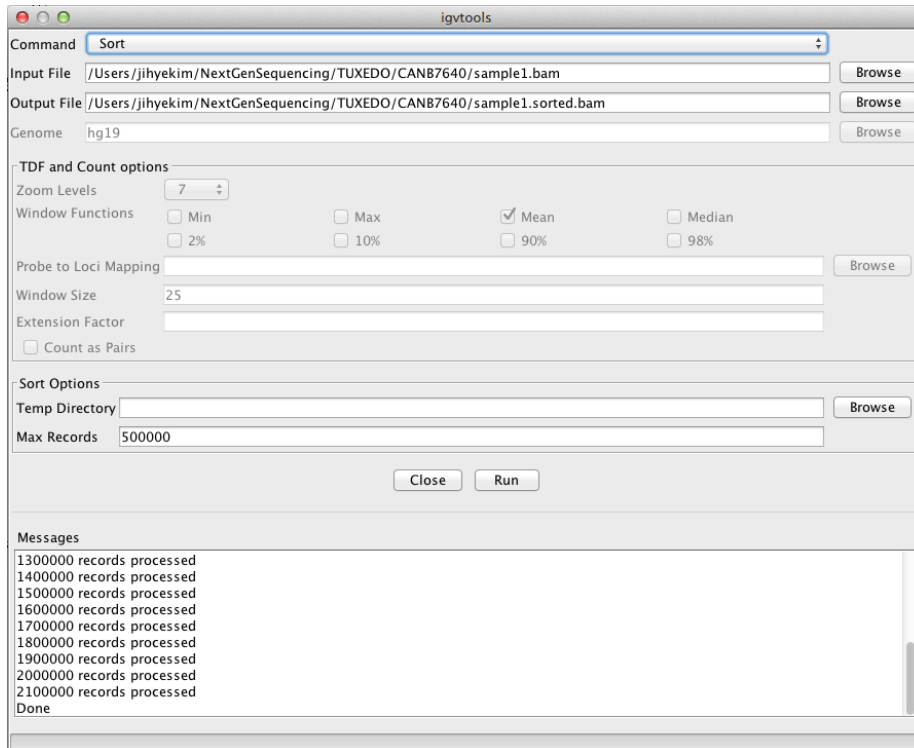
In CTRLvsGEF/

Open gene_exp.diff in excel, and sort by Status/Significance/P_value/Q_value

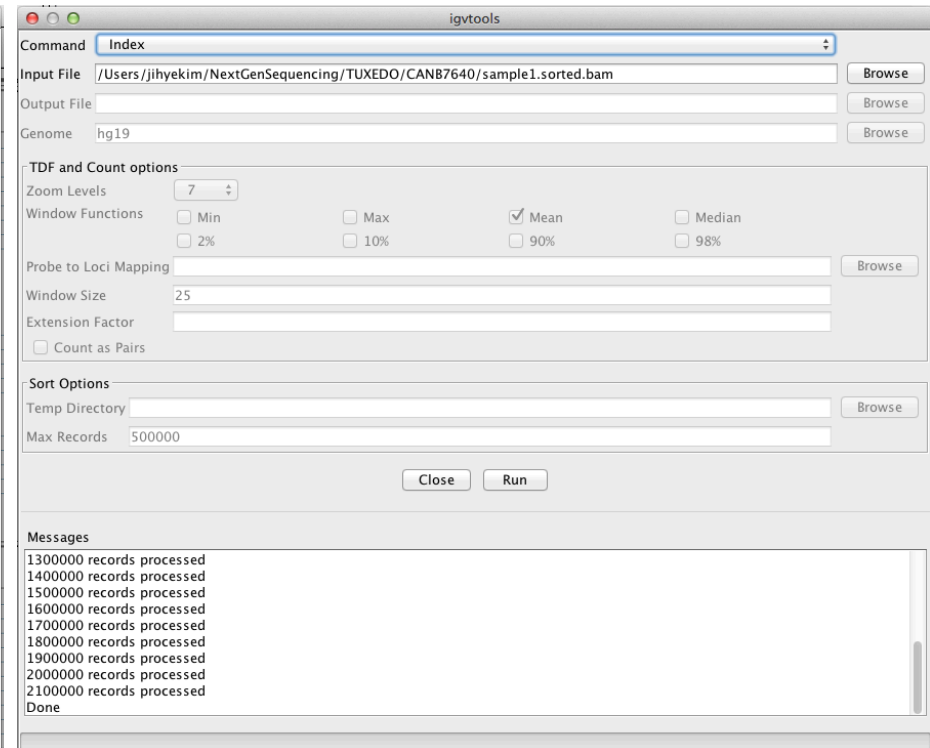| test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_ch	test_stat	p_value	q_value	significant
LOC1001283	LOC1001283	LOC1001283	7:55814589-	CTRL	Gefitinib	OK	0	8.10506	inf	nan	5.00E-05	0.017175	yes
LOC1005072	LOC1005072	LOC1005072	7:55811791-	CTRL	Gefitinib	OK	34.2939	0	#NAME?	nan	5.00E-05	0.017175	yes
OR2A9P	OR2A9P	OR2A9P	7:143996613	CTRL	Gefitinib	OK	7.88008	2.39781	-1.71649	-2.16835	0.13325	0.997001	no
RAB19	RAB19	RAB19	7:140103842	CTRL	Gefitinib	OK	8.78744	3.69744	-1.24892	-1.82393	0.15695	0.997001	no
WBSCR26	WBSCR26	WBSCR26	7:73149398-	CTRL	Gefitinib	OK	29.1613	12.2771	-1.24808	-3.0066	0.15695	0.997001	no
STAG3L2	STAG3L2	STAG3L2	7:74299266-	CTRL	Gefitinib	OK	10.763	4.58908	-1.2298	-2.31616	0.10025	0.997001	no
HOXA6	HOXA6	HOXA6	7:27179982-	CTRL	Gefitinib	OK	12.4756	6.17026	-1.01571	-1.67521	0.4082	0.997001	no
LOC1005058	LOC1005058	LOC1005058	7:5566777-5	CTRL	Gefitinib	OK	4976.77	2708.43	-0.87775	-0.501384	0.62995	0.997001	no
LOC1005055	LOC1005055	LOC1005055	7:916190-93	CTRL	Gefitinib	OK	11.1085	6.15308	-0.852283	-0.286303	0.77215	0.997001	no
LOC1005064	LOC1005064	LOC1005064	7:157129709	CTRL	Gefitinib	OK	116.803	67.3118	-0.795145	-0.594647	0.55605	0.997001	no
DPY19L2P4	DPY19L2P4	DPY19L2P4	7:89748713-	CTRL	Gefitinib	OK	9.35422	5.43378	-0.783661	-1.65466	0.09735	0.997001	no
DLX6	DLX6	DLX6	7:96597826-	CTRL	Gefitinib	OK	10.1665	5.925	-0.778937	-1.53142	0.1253	0.997001	no
PTN	PTN	PTN	7:136912087	CTRL	Gefitinib	OK	10.5727	6.56279	-0.687959	-1.45859	0.14355	0.997001	no

In IGV, Tools -> Run igvtools,

Sort sample1.bam

Index sample1.sorted.bam

In IGV, File – Load from file – sample1.sorted.bam & sample5.sorted.bam
Type "EGFR" and go.
(Remember , we have Chr7 only)