Tophat- Cufflinks

Tophat

Tophat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.

Tophat 2.0.3 is installed on Fourierseq as of 06/11/12.

Very basic Tophat command:

```
nohup tophat -p 4 -r <mate-inner-distance> -G <gfffile>
<bowtie_index_prefix> <R1.fasta> <R2.fasta> &>tophat.log &
```

- `--bowtie1` can be added to ask tophat to use bowtie1 instead of bowtie2 (bowtie2 does not support colorspace data).
- `-C` for colorspace (provide csfasta and quality files when using the flag)

Example 1: For mouse:

```
nohup tophat -p 4 -r 130 -G
/usr/local/genome/references/mmu_ncbi37/mm9_ucsc-known.gff3
/usr/local/genome/references/mmu_ncbi37/mmu_masked_ncbi37.fasta
sim.test1.forward.fa sim.test1.reverse.fa &>tophat.log &
```

Example 2: For human:

```
tophat
--transcriptome-index=/usr/local/genome/references/hg19/bowtie_gtf_index/hg19.gtf
/usr/local/genome/references/hg19/bowtie_index/hg19.bs.bowtie
<reads.fq>
```

Cufflinks

Cufflinks 2.0.0 is installed on Fourierseq as of 05/24/12

Very basic cufflinks command:

```
nohup cufflinks accepted_hits.bam -o cufflinks_outputdir &>cufflinks.log &
```

- `-G reference.gtf` can be added to use reference annotation to assemble transcripts. This will not assemble novel transcripts.
- `-g reference.gtf` can be added to use the reference annotation as a guide to assemble transcripts. This will include reference transcripts and novel transcripts.
- `accepted_hits.bam`: bam file created by tophat

Errors encountered:

An error during the segment mapping step with tophat 2.0.2 (Error: segment-based junction search failed with err =1) has been encountered that is solved by not using multiple threads ( `-p` option).