

Using .gtf files from iGenomes: "not sorted!"

Hi all,

I am trying to use Flux Simulator with a .gtf file from Illumina's [iGenomes collection](#). Specifically, I am using the genes.gtf in the Drosophila_melanogaster /UCSC/dm3/Annotation/Genes subdirectory of this archive: [dm3](#). I have set REF_FILE_NAME appropriately. However, when I run flux-simulator, I get an warning indicating that the GTF file isn't sorted. Next, flux "sorts" the GTF file for me, but then yields an error saying that the sorted file isn't sorted:

```
$ ./flux-simulator/bin/flux-simulator -p parameters/example.par
Flux-Simulator v1.2.1 (Flux Library: 1.22)

[INFO] No mode selected, executing the full pipeline (-x -l -s)
[INFO] I am collecting information on the run.
[INFO] Reading error model 76 bases model

[INFO] Checking GTF file
Checking GTF *[WARN] Unsorted in line 4 transcript id NM_175941 used twice, on: chr2L,chr2L
[GTF FILE] The GTF reference file given is not sorted, sorting it right now...
      sorting GTF file OK (00:00:20)
[GTF FILE] The Simulator will use /Users/langmead/git/tornado/tools/flux_sim/parameters/genes_sorted.gtf
[GTF FILE] You might want to update your parameters file
[PROFILING] I am assigning the expression profile
Checking GTF *[WARN] Unsorted in line 3496 transcript id NR_073697 used twice, on: chr2L,chr2L

[ERROR] The reference annotation GTF is not sorted!
java.lang.RuntimeException: The reference annotation GTF is not sorted!
    at barna.flux.simulator.Profiler.createGTFReader(Profiler.java:756)
    at barna.flux.simulator.Profiler.readAnnotation(Profiler.java:202)
    at barna.flux.simulator.Profiler.call(Profiler.java:127)
    at barna.flux.simulator.SimulationPipeline.call(SimulationPipeline.java:429)
    at barna.flux.simulator.SimulationPipeline.call(SimulationPipeline.java:54)
    at barna.commons.launcher.Flux.main(Flux.java:198)
```

I'm guessing this has something to do with how the .gtf files from iGenomes are formatted. Apparently, they are formatted to contain information that Cufflinks likes so that it can do differential analysis w/r/t promoters and coding sequences, in addition to transcripts:

<http://cufflinks.cbc.umd.edu/igenomes.html>

Can you help?

Best,

Ben