Understanding RPKM values

Hi,

I ran flux capacitor on a human BAM file with 9902168 aligned paired end reads (19804336 aligned reads which equals the number of SAM records as I consider primary alignments only); I used the parameter file

```
  ANNOTATION_FILE ensembl_human_71.gtf

  COUNT_ELEMENTS [SPLICE_JUNCTIONS, INTRONS]

  ANNOTATION_MAPPING PAIRED
```

Now if I only consider the entries of the output GTF file with feature = "transcript" (ignoring intron and junction entries) and add up the counts-per-million (CPM) values:

   RPKM * length / 1000

over all these entries, then I obtain as sum 1568177.63124 although by definition the sum of the CPM values should be 1000000.

Can you comment on this? Why do I get signifcantly more normalized read counts than I should?

One way to deal with this is to renormalize the data so that the CPM add up to 1000000. Another question would be whether this renormalization also need to be applied to intron and junction counts.

Best regrads,

Sven