Step 1: Download the two files bam from the links given in the webpage.
Step 2: Download StringTie and Tablemaker.
Step 3: We assemble the RNA-seq alignments to potential transcripts using StringTie for both samples. Using the long-read mode (-L) and the bam files as the input, we generate two .gtf files as the output. For reference input, we used the Grcm38 reference gtf file from Ensemble v94 (as stated in the data section of the paper.
Step 4: Now to prepare the Cufflinks output for Ballgown we used Tablemaker. However, when we ran the tool with the gtf file and the bam file as the input, we got the following error

> “BAM record error: found spliced alignment without XS attribute”

To resolve this issue, I went through the documentation of Tablemaker which said that “Tablemaker relies on Cufflinks version 2.1.1 (released April 2013) to estimate transcript FPKMs”. This prompted me to look through Cufflinks documentation from where I got the following information

> _"Cufflinks takes a text file of SAM alignments as input. For more details on the SAM format, see the specification . The RNA-Seq read mapper TopHat produces output in this format, and is recommended for use with Cufflinks. However Cufflinks will accept SAM alignments generated by any read mapper. Here's an example of an alignment Cufflinks will accept: s6.25mer.txt-913508 16 chr1 4482736 255 14M431N11M * 0 0 \ CAAGATGCTAGGCAAGTCTTGGAAG IIIIIIIIIIIIIIIIIIIIIIIII NM:i:0 XS:A:-_ **_Note the use of the custom tag XS ._** _This attribute, which must have a value of "+" or "-", indicates which strand the RNA that produced this read came from. While this tag can be applied to any alignment, including unspliced ones, it must be present for all spliced alignment records (those with a 'N' operation in the CIGAR string)."_

In short, Cufflinks require a custom “XS” tag which was missing from the input files because the reads were mapped using minimap2 (link to the data section of the paper). Only Tophat includes the “XS” tag which explains the missing tag in our case. I was, however, unable to find out whether the data was stranded or unstranded.
To bypass this problem, I used only Stringtie to generate the assembled transcripts using the -L tag and passing the bam files and reference as input. In the **v1.3.5 release** of Stringtie, spliced alignments produced by minimap2 (in SAM format) is supported; there is no need to pre-process them in order to add the XS tag, the "ts" tag is now recognized as an alternative.

Step 5: To generate a global, unified set of transcripts (isoforms) across multiple RNA-Seq samples**,** we use the stringtie --merge option where it takes input a list of GTF/GFF files and merges/assembles these transcripts into a non-redundant set of transcripts.

Step 6: To prepare Ballgown input, I use Stringtie with the -B switch. With this option StringTie can be used as a direct replacement of the *tablemaker* program included with the Ballgown distribution.

Step 7: The *ctab files from the previous steps are placed in separate folders for the two different samples and Ballgown is run to generate the FPKM values. All the outputs as required by the tests, are present in the Github repo.