Supplemental RNA Sequencing Analysis Methods

Step 0: Trim sequence reads

- Remove adapters from raw FASTQ sequence reads with cutadapt 1.16.

```
cutadapt \
       -m 30      \
       -a <ADPT_R1> \
       -A <ADPT_R2> \
       -o <FASTQ.1> \
       -p <FASTQ.2> \
       --trim-n  \
       --cores 4 \
       <IN_FASTQ.1> \
       <IN_FASTQ.2>
```

Step 1: Sequence Alignment with Star 2.5.3a and samtools 1.1

```
STAR --runThreadN 5   \
  --genomeDir <ref_genome_dir>1 \
  --outSAMtype BAM SortedByCoordinate \
  --readFilesCommand zcat \
  --twopassMode  Basic \
  --chimOutType  SeparateSAMold \
  --outFilterMismatchNoverLmax 0.04 \
  --chimSegmentMin 10 \
  --sjdbGTFfile <gene_model_gtf>2  \
  --readFilesIn <IN_FASTQ.1> <IN_FASTQ.2> \
  --outReadsUnmapped Fastx

samtools index Aligned.sortedByCoord.out.bam
samtools index Aligned.toTranscriptome.out.bam
```

[1]hs37d5 was used in this study.
[2]RefGene gene models were used in this study.

Step 2: Gene level quantification using HTSeq 0.6.1

```
samtools sort -n -m 8G Aligned.sortedByCoord.out.bam sample.star_sortByName

htseq-count -f bam -m union -s reverse \
  -i gene_id sample.star_sortByName.bam <gene_model_gtf> \
  >sample.star.cnt
```

Step 3: Expression normalization with IRON (PMC3651355)

```
# Read counts normalized against median sample SL278658:
iron_generic --rnaseq --norm-iron="SL278658" expression.txt \
  > norm_expression.txt
```