This documents lists the pipeline used for the assembly of 8 mouse macrophage RNAseq samples:

1. Sample information: The study contained RNASeq samples from three conditions:

2. control exosome: Contained three biological replicates with each replicate having two technical replicates.

3. pIC exosome: Contained three biological replicates with each replicate having two technical replicates.

4. PBS : Contained two biological replicates with each replicate having two technical replicates. However one of the biological replicate showed contamination in capture, and was left out of the analysis/

5. Reference information: The mouse mm10 reference genome and annotation files (UCSC) were downloaded from the Illumina igenomes, genome repository. Mus musculus (mm10) ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Mus_musculus/UCSC/mm10/Mus_musculus_UCSC_mm10.tar.gz

6. TopHat: Tophat was run using the following commands:

7. Control M1:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/Control_Exo_M1_MC1-
31664696/Control_ExoM1_UCSC ~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome
~/Control_Exo_M1_MC1-31664696/Control-Exo-M1-MC1_S5_L001_R1_001.fastq.gz,~/Control_Exo_M1_MC1-
31664696/Control-Exo-M1-MC1_S5_L002_R1_001.fastq.gz ~/Control_Exo_M1_MC1-31664696/Control-Exo-M1-
MC1_S5_L001_R2_001.fastq.gz,~/Control_Exo_M1_MC1-31664696/Control-Exo-M1-
MC1_S5_L002_R2_001.fastq.gz
```

2. Control M2:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/Control_Exo_M2_MC1-
31646815/Control_ExoM2_UCSC ~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome
~/Control_Exo_M2_MC1-31646815/Control-Exo-M2-MC1_S6_L001_R1_001.fastq.gz,~/Control_Exo_M2_MC1-
31646815/Control-Exo-M2-MC1_S6_L002_R1_001.fastq.gz ~/Control_Exo_M2_MC1-31646815/Control-Exo-M2-
MC1_S6_L001_R2_001.fastq.gz,~/Control_Exo_M2_MC1-31646815/Control-Exo-M2-
MC1_S6_L002_R2_001.fastq.gz
```

3. Control M3:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/Control_Exo_M3_MC1/Control_ExoM3_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome
~/Control_Exo_M3_MC1/Control_Exo_M3_MC_S20_L001_R1_001.fastq.gz,~/Control_Exo_M3_MC1/Control_Exo_M
3_MC_S20_L002_R1_001.fastq.gz
~/Control_Exo_M3_MC1/Control_Exo_M3_MC_S20_L001_R2_001.fastq.gz,~/Control_Exo_M3_MC1/Control_Exo_M
3_MC_S20_L002_R2_001.fastq.gz
```

4. pIC M1:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M1_MC1-31656710/pIC_exoM1_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome ~/pIC_exo_M1_MC1-31656710/pIC-exo-M1-
MC1_S7_L001_R1_001.fastq.gz,~/pIC_exo_M1_MC1-31656710/pIC-exo-M1-MC1_S7_L002_R1_001.fastq.gz
~/pIC_exo_M1_MC1-31656710/pIC-exo-M1-MC1_S7_L001_R2_001.fastq.gz,~/pIC_exo_M1_MC1-31656710/pIC-
exo-M1-MC1_S7_L002_R2_001.fastq.gz
```

5. pIC M2:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M2_MC1-31646822/pIC_exoM2_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome ~/pIC_exo_M2_MC1-31646822/pIC-exo-M2-
MC1_S8_L001_R1_001.fastq.gz,~/pIC_exo_M2_MC1-31646822/pIC-exo-M2-MC1_S8_L002_R1_001.fastq.gz
~/pIC_exo_M2_MC1-31646822/pIC-exo-M2-MC1_S8_L001_R2_001.fastq.gz,~/pIC_exo_M2_MC1-31646822/pIC-
exo-M2-MC1_S8_L002_R2_001.fastq.gz
```

6. pIC M3:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M3_MC1/pIC_exoM3_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome ~/pIC_exo_M3_MC1/pIC-exo-M3-
MC1_S17_L001_R1_001.fastq.gz,~/pIC_exo_M3_MC1/pIC-exo-M3-MC1_S17_L002_R1_001.fastq.gz
~/pIC_exo_M3_MC1/pIC-exo-M3-MC1_S17_L001_R2_001.fastq.gz,~/pIC_exo_M3_MC1/pIC-exo-M3-
MC1_S17_L002_R2_001.fastq.gz
```

7. PBS M1:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/PBS_M1_MC1-31658713/PBS_M1_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome ~/PBS_M1_MC1-31658713/PBS-M1-
MC1_S1_L001_R1_001.fastq.gz,~/PBS_M1_MC1-31658713/PBS-M1-MC1_S1_L002_R1_001.fastq.gz ~/PBS_M1_MC1-
31658713/PBS-M1-MC1_S1_L001_R2_001.fastq.gz,~/PBS_M1_MC1-31658713/PBS-M1-
MC1_S1_L002_R2_001.fastq.gz
```

8. PBS M2:

```
tophat2 --no-coverage-search --b2-very-sensitive --no-novel-juncs -p 4 -G
~/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/PBS_M2_MC1-31644830/PBS_M2_UCSC
~/Mus_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome ~/PBS_M2_MC1-31644830/PBS-M2-
MC1_S2_L001_R1_001.fastq.gz,~/PBS_M2_MC1-31644830/PBS-M2-MC1_S2_L002_R1_001.fastq.gz ~/PBS_M2_MC1-
31644830/PBS-M2-MC1_S2_L001_R2_001.fastq.gz,~/PBS_M2_MC1-31644830/PBS-M2-
MC1_S2_L002_R2_001.fastq.gz
```

4. Cufflinks:
5. Control M1:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/Control_Exo_M1_MC1-
31664696/Control_ExoM1_UCSC_cuff ~/Control_Exo_M1_MC1-
31664696/Control_ExoM1_UCSC/accepted_hits.bam
```

2. Control M2:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/Control_Exo_M2_MC1-
31646815/Control_ExoM2_UCSC_cuff ~/Control_Exo_M2_MC1-
31646815/Control_ExoM2_UCSC/accepted_hits.bam
```

3. Control M3:

```
 /projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
 /data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o
 ~/Control_Exo_M3_MC1/Control_ExoM3_UCSC_cuff
 ~/Control_Exo_M3_MC1/Control_ExoM3_UCSC/accepted_hits.bam
```

4. pIC M1:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M1_MC1-
31656710/pIC_exoM1_UCSC_cuff ~/pIC_exo_M1_MC1-31656710/pIC_exoM1_UCSC/accepted_hits.bam
```

5. pIC M2:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M2_MC1-
```

```
31646822/pIC_exoM2_UCSC_cuff ~/pIC_exo_M2_MC1-31646822/pIC_exoM2_UCSC/accepted_hits.bam
```

6. pIC M3:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/pIC_exo_M3_MC1/pIC_exoM3_UCSC_cuff
~/pIC_exo_M3_MC1/pIC_exoM3_UCSC/accepted_hits.bam
```

7. PBS M1:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/PBS_M1_MC1-
31658713/PBS_M1_UCSC_cuff ~/PBS_M1_MC1-31658713/PBS_M1_UCSC/accepted_hits.bam
```

8. PBS M2:

```
/projects/home/sravishankar9/tools/tuxedo_suite/cufflinks/cufflinks -p 4 -G
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -o ~/PBS_M2_MC1-
31644830/PBS_M2_UCSC_cuff ~/PBS_M2_MC1-31644830/PBS_M2_UCSC/accepted_hits.bam
```

5. Cuffmerge:
6. Prepare text file with paths to cufflinks transcripts files:

```
cat ~/Control_Exo_M1_MC1-31664696/Control_ExoM1_UCSC_cuff/transcripts.gtf ~/Control_Exo_M2_MC1-
31646815/Control_ExoM2_UCSC_cuff/transcripts.gtf
~/Control_Exo_M3_MC1/Control_ExoM3_UCSC_cuff/transcripts.gtf ~/pIC_exo_M1_MC1-
31656710/pIC_exoM1_UCSC_cuff/transcripts.gtf ~/pIC_exo_M2_MC1-
31646822/pIC_exoM2_UCSC_cuff/transcripts.gtf ~/pIC_exo_M3_MC1/pIC_exoM3_UCSC_cuff/transcripts.gtf
~/PBS_M1_MC1-31658713/PBS_M1_UCSC_cuff/transcripts.gtf ~/PBS_M2_MC1-
31644830/PBS_M2_UCSC_cuff/transcripts.gtf > ~/assemblies.txt
```

2. Run cuffmerge to prepare merged transcript file:

```
cuffmerge -g /data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genes.gtf -s
/data/db/Mus_musculus/UCSC/mm10/Annotation/Genes/genome.fa -p 8 ~/assemblies.txt
```

6. Cuffdiff:
7. Run cuffdiff between control exosome and pIC exosome:

```
cuffdiff -o controlvspIC -b /data/db/Mus_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa -p
4 -L Control,pIC -u ~/merged_asm/merged.gtf ~/Control_Exo_M1_MC1-
31664696/Control_ExoM1_UCSC/accepted_hits.bam,~/Control_Exo_M2_MC1-
31646815/Control_ExoM2_UCSC/accepted_hits.bam,~/Control_Exo_M3_MC1/Control_Exo_M3_outputs_UCSC/acc
epted_hits.bam ~/pIC_exo_M1_MC1-31656710/pIC_M1_UCSC/accepted_hits.bam,~/pIC_exo_M2_MC1-
31646822/pIC_M2_UCSC/accepted_hits.bam,~/pIC_exo_M3_MC1/pIC-Exo-M3_outputs_UCSC/accepted_hits.bam
```

2. Run cuffdiff between PBS and pIC exosome:

```
cuffdiff -o PBSvspIC -b /data/db/Mus_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa -p 4 -
L PBS,pIC -u ~/merged_asm/merged.gtf ~/PBS_M1_MC1-
31658713/PBSM1_UCSC/accepted_hits.bam,~/PBS_M2_MC1-31644830/PBSM2_UCSC/accepted_hits.bam
~/pIC_M1_UCSC/accepted_hits.bam,~/pIC_M2_UCSC/accepted_hits.bam
```

7. Gage:
8. Aggregating the bamfiles:

```
cp experiment2/condition1/Control_Exo_M1_MC1-31664696/Control_ExoM1_UCSC/accepted_hits.bam
bamfiles/controlm1.bam
cp experiment2/condition1/Control_Exo_M2_MC1-31646815/Control_ExoM2_UCSC/accepted_hits.bam
bamfiles/controlm2.bam
```

```
cp experiment2/condition1/Control_Exo_M3_MC1/Control_Exo_M3_accepted_hits.bam
bamfiles/controlm2.bam
cp experiment2/condition1/Control_Exo_M3_MC1/Control_Exo_M3_outputs_UCSC/accepted_hits.bam
bamfiles/controlm3.bam
cp experiment2/condition2/pIC_exo_M1_MC1-31656710/pIC_M1_UCSC/accepted_hits.bam bamfiles/picm1.bam
cp experiment2/condition2/pIC_exo_M2_MC1-31646822/pIC_M2_UCSC/accepted_hits.bam bamfiles/picm2.bam
cp experiment2/condition2/pIC_exo_M3_MC1/pIC-Exo-M3_outputs_UCSC/accepted_hits.bam
bamfiles/picm3.bam
cp experiment1/condition1/PBS_M2_MC1-31644830/PBSM2_UCSC/accepted_hits.bam bamfiles/pbsm2.bam
```

2. Indexing all bamfiles:

```
cd bamfiles
samtools index *
```

3. Count reads mapped to each gene:

```
source("http://bioconductor.org/biocLite.R")
biocLite(c("pathview", "gage", "gageData", "GenomicAlignments",
"TxDb.Mmusculus.UCSC.mm10.knownGene"))
library(TxDb.Mmusculus.UCSC.mm10.knownGene)
exByGn <- exonsBy(TxDb.Mmusculus.UCSC.mm10.knownGene, "gene")
library(GenomicAlignments)
fls <- list.files("bamfiles/", pattern="bam$", full.names =T)
bamfls <- BamFileList(fls)
flag <- scanBamFlag(isSecondaryAlignment=FALSE, isProperPair=TRUE)
param <- ScanBamParam(flag=flag)
gnCnt <- summarizeOverlaps(exByGn, bamfls, mode="Union", ignore.strand=TRUE, singleEnd=FALSE,
param=param)
hnrnp.cnts=assay(gnCnt)
```

8. Figures and plots: The following R code plots the expression of the genes mentioned in "~/GenesList.txt" to create the expression plots

```
library(ggplot2)
library(Hmisc)
library(gridExtra)
library(grid)
library(reshape2)
genes <- read.table('~/GenesList.txt', header=TRUE,sep='\t')
genel <- as.vector(genes$Gene.name)
genel <- capitalize(tolower(genel))
names(genes)[1] <- "gene"
genes$gene <- capitalize(tolower(genes$gene))
difexp <- read.table('~/controlvspIC/gene_exp.diff',header=TRUE,sep='\t')
difexp <- subset(difexp,select = c(gene,value_1,value_2))
difexp2 <- read.table('~/PBSvspICM2/gene_exp.diff', header=TRUE,sep='\t')
difexp2 <- subset(difexp2, select=c(gene,value_1))
names(difexp)[2] <- "Con_exo"
names(difexp)[3] <- "pIC_exo"
names(difexp2)[2] <- "PBS"
difexp <- merge(difexp,difexp2)
difexp$Con_exo <- log2(difexp$Con_exo+1)
difexp$pIC_exo <- log2(difexp$pIC_exo+1)
difexp$PBS <- log2(difexp$PBS+1)
dmg <- difexp[is.element(difexp$gene,genel),]
dmg <- merge(dmg, genes)
dmg <- dmg[order(dmg[,5],dmg[,1]),]
row.names(dmg) <- dmg$gene
dmgl <- dmg[,2:5]
dmgl <- dmgl[order(dmgl[,4],dmgl[,1],decreasing=TRUE),]
dmgl <- dmgl[,1:3]
```

```
dmg1$Gene <- row.names(dmg1)
dmg1.m <- melt(dmg1)
names(dmg1.m)[1] <- "Gene"
names(dmg1.m)[2] <- "Sample"
names(dmg1.m)[3] <- "log2_FPKM"
dmg1.m$Gene <- factor(dmg1.m$Gene,levels=dmg1.m$Gene)
df.m <- melt(difexp)
names(df.m)[3] <- "log2_FPKM"
names(df.m)[2] <- "Condition"
scatter <- ggplot(data = dmg1.m, aes(x = Gene, y = log2_FPKM, shape=Sample,  fill=Sample,
aplha=0.9, stroke=1 )) + geom_point(size=5) + coord_flip() + scale_shape_manual(values =
c(21,21,21)) + scale_fill_manual(values=c("grey","black","white"))  + theme_bw() +
theme(legend.position=c(1,0.8),legend.justification=c(1,1),axis.text.x=element_text(face="bold"
), axis.text.y=element_text(face="italic"), axis.line=element_line(size=1, linetype = "solid"),
panel.background=element_blank(), panel.border =element_blank(), panel.grid.major =
element_blank(), panel.grid.minor = element_blank()) + labs(y=bquote('Gene
Expression('~Log[2]~')'),x=element_blank())
density <- ggplot(df.m, aes(x=log2_FPKM, fill=Condition,stroke=12)) + geom_density(alpha=.6) +
scale_fill_manual(values=c("grey","black","white")) +theme_bw() +theme(legend.position =
"none", axis.title.y=element_blank(), axis.title.x=element_blank(), panel.grid.major =
element_blank(), panel.grid.minor = element_blank()) +scale_x_continuous(limits =
range(dmg1.m$log2_FPKM))
scat <- ggplot_gtable(ggplot_build(scatter))
dens <- ggplot_gtable(ggplot_build(density))
maxWidth = unit.pmax(scat$widths[2:3], dens$widths[2:3])
scat$widths[2:3] <- maxWidth
dens$widths[2:3] <- maxWidth
pdf('ExpressionPlots.pdf')
grid.arrange(dens, scat, heights = c(3, 16))
dev.off()
```

Example plot:

![Expression plot][plot] [plot]:ExpressionPlots.png