

Tissue Specificity in *Syngnathus floridae*

Coley Tosto

2023-11-07

```
#Create the DESeq dataset
dds_FL <- DESeqDataSetFromTximport(tx1.salmom,
                                      colData = samples,
                                      design = ~ group)

#only keeping rows that have at least 10 reads total
keep <- rowSums(counts(dds_FL)) >= 10
dds_FL <- dds_FL[keep, ]

#Generate the expression values
dds_FL_exp <- DESeq(dds_FL)
```

To estimate tissue specificity, the TPM estimates are needed which is the number of transcripts from a particular gene normalized first by gene length, and then by sequencing depth (in millions) in the sample. The output quant.sf files from salmon have the following format:

Name | Length | EffectiveLength | TPM | NumReads

From these salmon outputs we can then pull out the TPM values for each sample.

```
files <- list.files(pattern = ".sf", path = "data/floridae_expression_files",
                      full.names = TRUE)

tpms <- do.call(cbind, lapply(files, function(file){
  #browser()
  dat <- read.delim(file, row.names = "Name")
  tpm <- dat["TPM"]
  colnames(tpm) <- gsub("data/floridae_expression_files/(.*)_quant.sf", "\\\1", file)

  return(tpm)
}))
```

Once all the TPMs were gathered for the different samples I filtered some of the rows out based on results from DESeq2. It should be noted that the DESeq2 filtering was done based off of the count data and not the TPMs, but as TPMs are just normalized counts they should be correlated and something that was had low gene counts should also have a low TPM and anything that was considered an outlier in the counts could also be an outlier in terms of TPM, but it may not be perfect.

```
#Only keeping the rows that weren't filtered out due to low counts
tpms <- tpms[rownames(tpms) %in% rownames(dds_FL), ]
```

```

#Pulling out the geneIDs for genes that were categorized as "outliers" by DESeq2
#Calculating the Cooks threshold that would have been used
np <- length(resultsNames(dds_FL_exp))
nsamp <- ncol(dds_FL_exp)
cooks_thresh <- qf(0.99, df1 = np, df2 = 29-np)

out_ids <- names(mcols(dds_FL_exp)$maxCooks[mcols(dds_FL_exp)$maxCooks > cooks_thresh])

#Removing the rows in the tpm dataset that were deemed "outliers" by DESeq2
tpms <- tpms[!(rownames(tpms) %in% out_ids), ]

```

These TPM estimates can then be used to estimate τ , a tissue specificity estimator that can range from 0 to 1. τ is calculated for each tissue, i , as follows:

$$\tau = \frac{\sum_i [1 - \ln(TPM_i)/\ln(TPM_{max})]}{N - 1}$$

where TPM_{max} is the maximum average TPM for a given tissue type, and TPM_i is the average TPM for tissue i . If $\tau = 0$, that gene is evenly expressed across tissues; if $\tau = 1$, the gene is expressed in an entirely tissue-specific fashion. Because TPM values approach 0 are impacted by sampling stochasticity, any tissues that had an average expression of 0 for a given gene were set to $TPM = 2$.

```

est_tau<-function(geneDat,colDat){

  tissue_dat<-data.frame(cbind(colDat,
                                 geneDat))
  #browser()
  tissue_dat$geneDat[tissue_dat$geneDat<1]<-2

  #tpmMax<-max(tissue_dat$geneDat)
  tissue_avgs<-tapply(tissue_dat$geneDat,tissue_dat$Organ,mean)
  tpmMax <- max(tissue_avgs, na.rm=TRUE)

  #tissue_avgs[tissue_avgs<1]<-2
  if(length(unique(tissue_dat$Organ)) == 3){
    tau <- sum(1-(log(tissue_avgs[unique(tissue_dat$Organ)]))/log(tpmMax)))/(length(unique(tissue_dat$Organ))-1)

    return(tau)
  }

  tau<-sum(1-(log(tissue_avgs)/log(tpmMax)))/(length(unique(tissue_dat$Organ))-1)

  return(tau)
}

tau <- apply(tpms, 1, est_tau, colData=colData)

tau_fem<-apply(tpms[,which(colData$Sex=="F")], 1, est_tau,
               colData[which(colData$Sex=="F"),])

```

```
tau_mal<-apply(tpms[,which(colData$Sex=="M")], 1, est_tau,
                 colData[which(colData$Sex=="M"),])
```

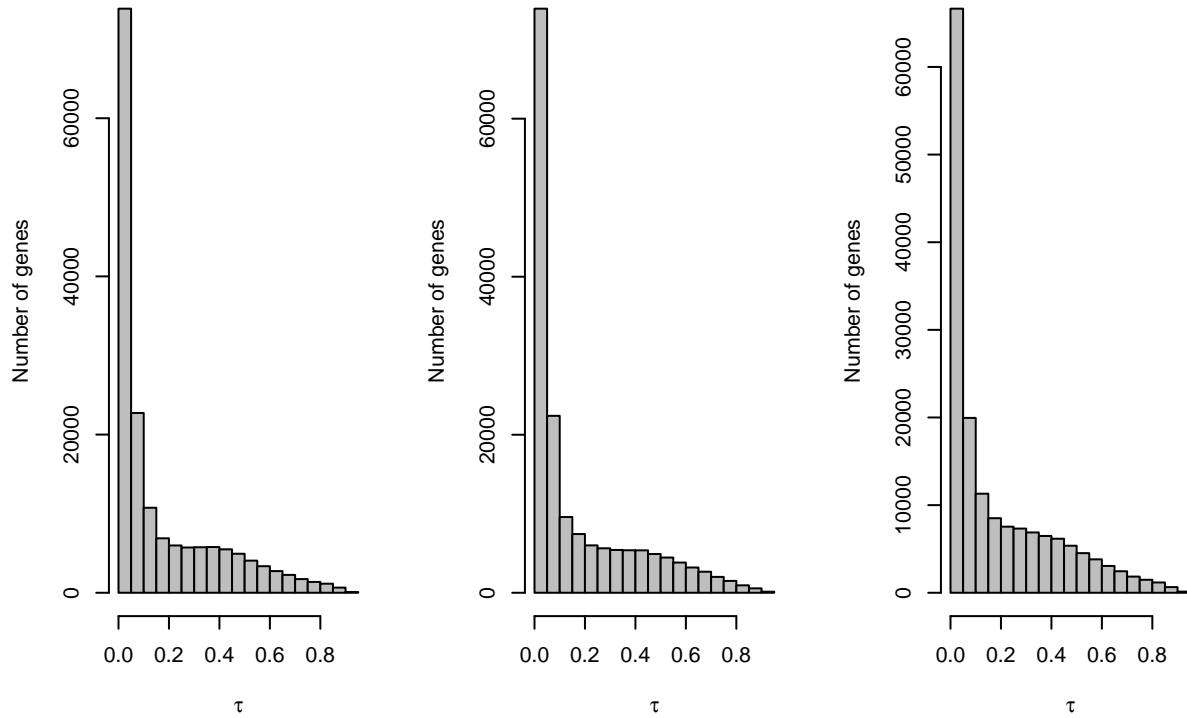


Figure 1: Distribution of tissue specificity estimates for only the male samples (left) versus only the female samples (middle) versus the male samples and the female samples (right).

Let's now investigate some of the plot counts for the genes with a high tissue specificity index.

```
## named numeric(0)

## TRINITY_DN4937_c1_g1
##               0.9400684

##          FLG2F7   FLG3F1   FLG3F2   FLG3M5   FLG3M7   FLG3M8   FLG4M3
## TRINITY_DN79971_c0_g1      0 0.22401 1.467807      0      0 0.524378 1.357382
##          FLG4M4   FLG8F3   FLL2F7   FLL3F1   FLL3F2   FLL3F4
## TRINITY_DN79971_c0_g1 2.201918 0.123275 2.618225 1.054227 2.438477 0.835499
##          FLL3M5   FLL3M7   FLL3M8   FLL4M3   FLL4M4   FLL8F3
## TRINITY_DN79971_c0_g1 0.068457 0.665254 0.580137 7.961344 118.5182 1.335837
##          FL02F7   FL03F1   FL03F2   FL03F4   FL08F3   FLT2M3   FLT3M5
## TRINITY_DN79971_c0_g1      0      0 0.353455      0      0 0.172507      0
##          FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN79971_c0_g1 0.3445      0      0
```

```

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8
## TRINITY_DN33844_c0_g1 2.173862 1.58947 2.009153 0.758988 0.847319 5.510411
##          FLG4M3  FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4
## TRINITY_DN33844_c0_g1 2.528495 59.9265      0 1.105366 1.418612      0 1.619898
##          FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4  FLL8F3
## TRINITY_DN33844_c0_g1      0 0.381859 1.095565 0.524579 0.862077 0.604758
##          FL02F7  FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3
## TRINITY_DN33844_c0_g1 0.863839 0.658435      0 1.173996 2.537471 0.793188
##          FLT3M5  FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN33844_c0_g1 1.885087      0      0 0.885463

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3  FLG4M4
## TRINITY_DN32505_c0_g2      0      0      0      0      0      0      0      0
##          FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4  FLL3M5  FLL3M7
## TRINITY_DN32505_c0_g2      0      0      0 11.72874 81.76168      0      0
##          FLL3M8  FLL4M3  FLL4M4  FLL8F3  FL02F7  FL03F1  FL03F2  FL03F4
## TRINITY_DN32505_c0_g2      0      0      0      0      0      0      0      0
##          FL08F3  FLT2M3  FLT3M5  FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN32505_c0_g2      0 0.336207      0      0 1.232252 0.662413

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3
## TRINITY_DN17266_c0_g2 2.746994 172.7351      0 3.219645      0 1.901685 9.51915
##          FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4  FLL3M5  FLL3M7
## TRINITY_DN17266_c0_g2      0 13.24765      0      0      0      0      0      0
##          FLL3M8  FLL4M3  FLL4M4  FLL8F3  FL02F7  FL03F1  FL03F2  FL03F4
## TRINITY_DN17266_c0_g2      0      0      0      0      0      0      0 0.103716
##          FL08F3  FLT2M3  FLT3M5  FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN17266_c0_g2 1.120032      0      0 0.069393 0.715632      0

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3
## TRINITY_DN71047_c0_g1      0      0      0      0      0 0.93237 0.5407
##          FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4
## TRINITY_DN71047_c0_g1 0.527475 0.15202 3.498205 1.270062 1.820162 1.393168
##          FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4  FLL8F3
## TRINITY_DN71047_c0_g1 0.166199 2.193122 788.7528 27.96898 0.202476 11.25624
##          FL02F7  FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3  FLT3M5
## TRINITY_DN71047_c0_g1      0 0.173856      0      0 0.415763      0      0
##          FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN71047_c0_g1      0      0      0

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3  FLG4M4
## TRINITY_DN927_c6_g2      0      0      0      0      0      0      0      0

```

```

##          FLG8F3  FLL2F7 FLL3F1   FLL3F2   FLL3F4 FLL3M5 FLL3M7
## TRINITY_DN927_c6_g2      0 0.386974      0 13.04647 162.5054      0      0
##          FLL3M8 FLL4M3  FLL4M4 FLL8F3   FL02F7 FL03F1 FL03F2 FL03F4
## TRINITY_DN927_c6_g2      0      0 0.92337      0 1.732552      0      0      0
##          FL08F3   FLT2M3 FLT3M5 FLT4M4  FLT5M3   FLT8M7
## TRINITY_DN927_c6_g2      0 0.637396      0      0      0 0.646855

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5 FLG3M7 FLG3M8 FLG4M3 FLG4M4
## TRINITY_DN37495_c0_g1    0 0.479101      0      0      0      0      0      0
##          FLG8F3  FLL2F7  FLL3F1   FLL3F2   FLL3F4 FLL3M5
## TRINITY_DN37495_c0_g1    0 0.89563 88.99382 0.335787 0.296697 0.197344
##          FLL3M7  FLL3M8  FLL4M3   FLL4M4   FLL8F3 FL02F7
## TRINITY_DN37495_c0_g1 0.631658 2.494548 1.463013 2.514473 3.961195      0
##          FL03F1   FL03F2 FL03F4  FL08F3  FLT2M3 FLT3M5 FLT4M4
## TRINITY_DN37495_c0_g1 1.698259 0.753874      0      0      0      0      0
##          FLT5M3   FLT8M7
## TRINITY_DN37495_c0_g1      0      0

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5 FLG3M7 FLG3M8 FLG4M3
## TRINITY_DN2652_c2_g1 1.089246 0.185818 2.297256 1.367329      0 0.207352      0
##          FLG4M4  FLG8F3  FLL2F7 FLL3F1   FLL3F2   FLL3F4 FLL3M5
## TRINITY_DN2652_c2_g1      0 1.627541 0.517962      0 0.530727 109.5318      0
##          FLL3M7  FLL3M8  FLL4M3   FLL4M4   FLL8F3 FL02F7 FL03F1
## TRINITY_DN2652_c2_g1 0.124085      0      0      0 7.187611      0      0
##          FL03F2   FL03F4  FL08F3  FLT2M3 FLT3M5 FLT4M4  FLT5M3
## TRINITY_DN2652_c2_g1      0 0.963106 0.881694      0      0      0      0
##          FLT8M7
## TRINITY_DN2652_c2_g1      0

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5 FLG3M7 FLG3M8
## TRINITY_DN29035_c1_g1 1.911056 7.444657 3.207478 2.052365 5.227144 41.86098
##          FLG4M3  FLG4M4  FLG8F3 FLL2F7 FLL3F1   FLL3F2
## TRINITY_DN29035_c1_g1 0.764862 3.383421 0.832816      0      0 0.711857
##          FLL3F4  FLL3M5  FLL3M7   FLL3M8 FLL4M3 FLL4M4
## TRINITY_DN29035_c1_g1 2.099761 0.906164 1.008221 4.493387      0      0
##          FLL8F3 FL02F7  FL03F1   FL03F2   FL03F4  FL08F3
## TRINITY_DN29035_c1_g1 0.786662      0 1.716189 1.202728 0.803588      0
##          FLT2M3   FLT3M5 FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN29035_c1_g1 1.050478 1.314504      0 1.312749 1.048797

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5 FLG3M7 FLG3M8
## TRINITY_DN138910_c0_g2 6.398822 281.7957      0 5.431731      0 3.725921
##          FLG4M3  FLG4M4  FLG8F3 FLL2F7 FLL3F1   FLL3F2   FLL3F4
## TRINITY_DN138910_c0_g2 14.56353      0 32.2533      0      0 0.518673

```

```

## FLL3M5 FLL3M7 FLL3M8 FLL4M3 FLL4M4 FLL8F3 FL02F7 FL03F1
## TRINITY_DN138910_c0_g2      0      0      0      0      0      0      0      0
##          FL03F2 FL03F4   FL08F3 FLT2M3 FLT3M5  FLT4M4   FLT5M3
## TRINITY_DN138910_c0_g2      0      0 1.037258      0      0 0.19479 0.333056
##          FLT8M7
## TRINITY_DN138910_c0_g2      0

## using 'avgTxLength' from assays(dds), correcting for library size

## FLG2F7   FLG3F1 FLG3F2   FLG3M5   FLG3M7 FLG3M8
## TRINITY_DN123871_c1_g1 1.380002 0.704705      0 1.360599 2.212473      0
##          FLG4M3   FLG4M4   FLG8F3   FLL2F7   FLL3F1   FLL3F2
## TRINITY_DN123871_c1_g1 0.726456 57.26973 0.346797 1.442465 2.242362 1.072718
##          FLL3F4   FLL3M5 FLL3M7   FLL3M8   FLL4M3   FLL4M4 FLL8F3
## TRINITY_DN123871_c1_g1      0 0.696165      0 0.246113 1.05008 0.293099      0
##          FL02F7   FL03F1   FL03F2   FL03F4   FL08F3 FLT2M3
## TRINITY_DN123871_c1_g1 1.638058 0.471651 1.886876 2.036188 1.707548      0
##          FLT3M5   FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN123871_c1_g1      0 0.540945 1.223018 2.800842

## using 'avgTxLength' from assays(dds), correcting for library size

## TRINITY_DN4937_c1_g1    TRINITY_DN3390_c5_g1 TRINITY_DN168500_c0_g1
##          0.9400684      0.9328661      0.9304992
## TRINITY_DN125448_c0_g1    TRINITY_DN2737_c10_g1 TRINITY_DN150289_c0_g1
##          0.9303219      0.9255532      0.9254920

## FLG2F7   FLG3F1 FLG3F2   FLG3M5   FLG3M7 FLG3M8
## TRINITY_DN4937_c1_g1 0.244273 1.387541 0.144996 0.173663 2.953393 1.531152
##          FLG4M3   FLG4M4   FLG8F3   FLL2F7   FLL3F1   FLL3F2
## TRINITY_DN4937_c1_g1 0.328049 0.037401 0.281063 7860.621 8947.962 10201.43
##          FLL3F4   FLL3M5 FLL3M7   FLL3M8   FLL4M3   FLL4M4
## TRINITY_DN4937_c1_g1 8286.969 9197.222 10789.22 8872.272 12427.72 16293.61
##          FLL8F3   FL02F7   FL03F1   FL03F2   FL03F4   FL08F3
## TRINITY_DN4937_c1_g1 15376.87 0.951264 0.556749 1.404592 1.156154 1.092096
##          FLT2M3   FLT3M5   FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN4937_c1_g1 0.254501 1.373361 0.221112 1.367545 0.093597

## using 'avgTxLength' from assays(dds), correcting for library size

```

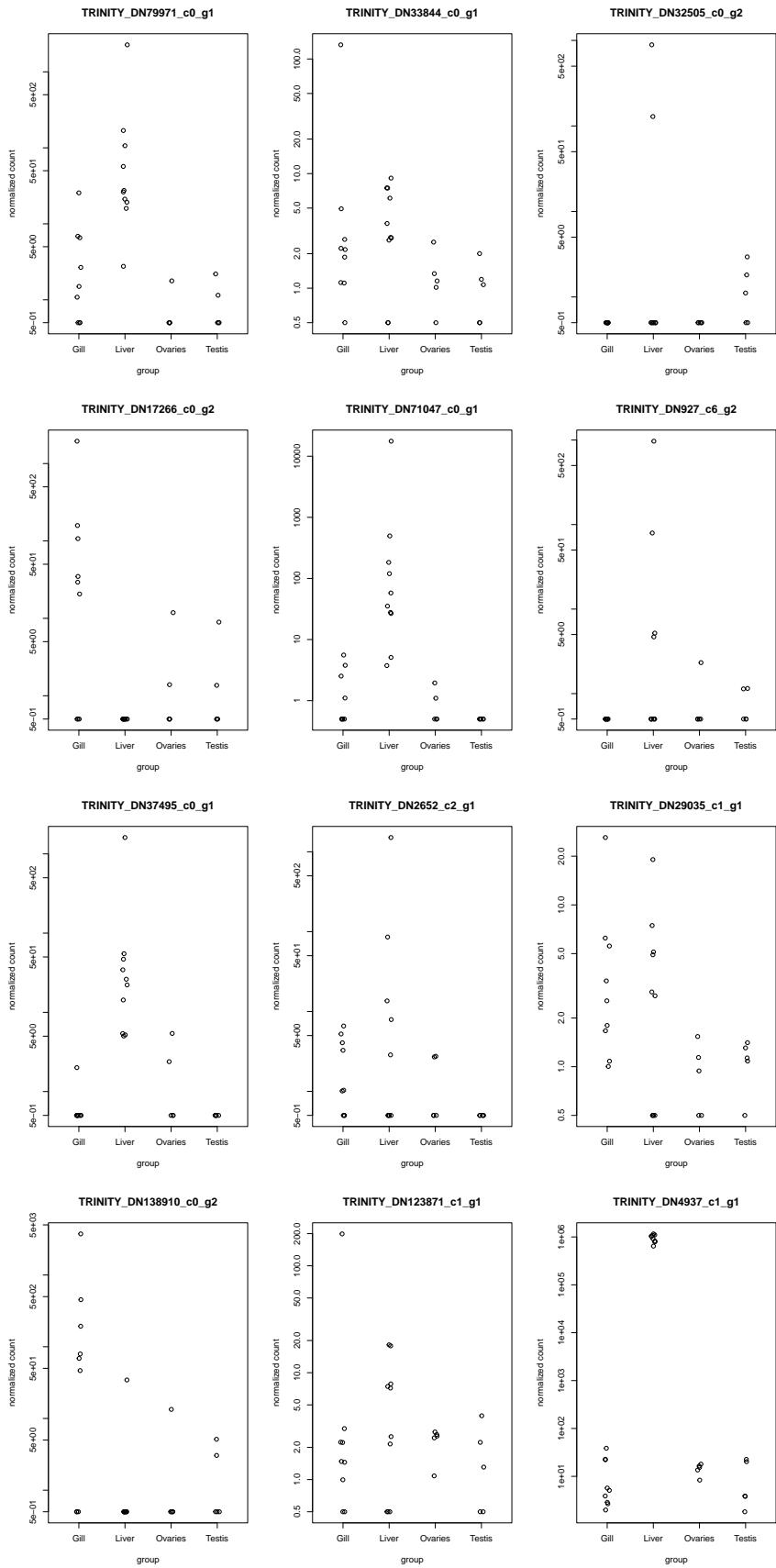


Figure 2: Plots of the counts from the DESeq2 Dataset for all of the genes that had a tau greater than 1

```

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8
## TRINITY_DN3390_c5_g1 1.915322 1.191699 1.374606 1.283282 0.648988 0.293973
##          FLG4M3  FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4
## TRINITY_DN3390_c5_g1 0.086935 0.47156   0       0 0.752944 0.166046   0
##          FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4  FLL8F3  FL02F7
## TRINITY_DN3390_c5_g1 0.648803   0 0.167483 0.07965   0       0 1.339849
##          FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3  FLT3M5
## TRINITY_DN3390_c5_g1 1.243675 1.719132 1.212814 0.894546 797.6304 1766.4
##          FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN3390_c5_g1 5728.606 9424.542 1389.621

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8
## TRINITY_DN168500_c0_g1   0 0.942199 1.553451 0.288835 1.454043 2.409941
##          FLG4M3  FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2
## TRINITY_DN168500_c0_g1   0       0 0.234851 7007.593 6960.66 8191.516
##          FLL3F4  FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4
## TRINITY_DN168500_c0_g1 7611.868 8648.97 10103.08 6255.89 10745.12 14742.49
##          FLL8F3  FL02F7  FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3
## TRINITY_DN168500_c0_g1 16709.12 1.533473   0 1.228352 2.847534   0       0
##          FLT3M5  FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN168500_c0_g1 1.11452    0 0.398892 0.356016

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3
## TRINITY_DN125448_c0_g1   0 0.313606 1.413377 0.186909 1.835993 3.2061
##          FLG4M3  FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2
## TRINITY_DN125448_c0_g1 1.621366 0.423442 1.131618 18283.77 22189.16 15420.11
##          FLL3F4  FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4
## TRINITY_DN125448_c0_g1 22303.16 7284.046 14130.52 16768.33 17550.15 20049.69
##          FLL8F3  FL02F7  FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3
## TRINITY_DN125448_c0_g1 22558.86 0.619148   0 2.096597   0       0 0.456869
##          FLT3M5  FLT4M4  FLT5M3  FLT8M7
## TRINITY_DN125448_c0_g1   0       0   0 0.221689

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7  FLG3F1  FLG3F2  FLG3M5  FLG3M7  FLG3M8  FLG4M3
## TRINITY_DN2737_c10_g1   0       0 1.253771   0 2.89316 1.865081   0
##          FLG4M4  FLG8F3  FLL2F7  FLL3F1  FLL3F2  FLL3F4
## TRINITY_DN2737_c10_g1   0       0 12006.16 17338.61 14526.94 17942.09
##          FLL3M5  FLL3M7  FLL3M8  FLL4M3  FLL4M4  FLL8F3
## TRINITY_DN2737_c10_g1 9629.504 13219.28 15319.45 23167.47 9311.478 15759.04
##          FL02F7  FL03F1  FL03F2  FL03F4  FL08F3  FLT2M3  FLT3M5  FLT4M4
## TRINITY_DN2737_c10_g1   0       0 2.6697   0 0.59965   0       0       0
##          FLT5M3  FLT8M7
## TRINITY_DN2737_c10_g1   0       0

```

using 'avgTxLength' from assays(dds), correcting for library size

```

##          FLG2F7 FLG3F1 FLG3F2 FLG3M5 FLG3M7 FLG3M8 FLG4M3 FLG4M4
## TRINITY_DN150289_c0_g1      0      0      0      0      0      0      0      0
##          FLG8F3 FLL2F7 FLL3F1 FLL3F2 FLL3F4 FLL3M5 FLL3M7 FLL3M8
## TRINITY_DN150289_c0_g1      0      0      0      0      0      0      0      0
##          FLL4M3 FLL4M4 FLL8F3   FL02F7   FL03F1   FL03F2   FL03F4
## TRINITY_DN150289_c0_g1      0      0      0 14452.83 7087.196 7599.357 12631.54
##          FL08F3   FLT2M3   FLT3M5   FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN150289_c0_g1 13082.83 0.286077 0.346457      0      0 0.569997

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7 FLG3F1 FLG3F2 FLG3M5 FLG3M7 FLG3M8 FLG4M3
## TRINITY_DN3_c0_g1 0.096025 0.0326 0.027709 0.080982 0.018051 0.10282 0.032907
##          FLG4M4 FLG8F3 FLL2F7 FLL3F1 FLL3F2 FLL3F4 FLL3M5
## TRINITY_DN3_c0_g1 0.040528 0.013272 0.029741 0.025211 0.50325 0.631 0.47946
##          FLL3M7 FLL3M8 FLL4M3 FLL4M4 FLL8F3   FL02F7
## TRINITY_DN3_c0_g1 0.159477 0.041036 0.008677 0.051655 0.158885 0.095594
##          FL03F1   FL03F2   FL03F4   FL08F3   FLT2M3   FLT3M5   FLT4M4
## TRINITY_DN3_c0_g1 0.125921 0.175559 0.241228 0.104543 453.2654 614.722 1815.225
##          FLT5M3   FLT8M7
## TRINITY_DN3_c0_g1 2541.909 445.8102

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7 FLG3F1 FLG3F2 FLG3M5 FLG3M7 FLG3M8 FLG4M3
## TRINITY_DN105340_c0_g1 7472.732 2082.841 1893.741 3040.507 6876.184 6735.64
##          FLG4M3 FLG4M4 FLG8F3 FLL2F7 FLL3F1 FLL3F2 FLL3F4
## TRINITY_DN105340_c0_g1 2266.581 2.794158 550.1212      0      0      0      0
##          FLL3M5 FLL3M7 FLL3M8 FLL4M3 FLL4M4 FLL8F3   FL02F7   FL03F1
## TRINITY_DN105340_c0_g1      0      0      0      0      0      0      0      0
##          FL03F2   FL03F4   FL08F3   FLT2M3   FLT3M5   FLT4M4   FLT5M3
## TRINITY_DN105340_c0_g1      0      0      0      0 1.990756      0      0
##          FLT8M7
## TRINITY_DN105340_c0_g1 0.872494

## using 'avgTxLength' from assays(dds), correcting for library size

##          FLG2F7 FLG3F1 FLG3F2 FLG3M5 FLG3M7 FLG3M8 FLG4M3 FLG4M4
## TRINITY_DN30747_c0_g1      0      0      0      0      0      0      0      0
##          FLG8F3 FLL2F7 FLL3F1 FLL3F2 FLL3F4 FLL3M5 FLL3M7 FLL3M8
## TRINITY_DN30747_c0_g1      0      0      0      0      0      0      0 0.570495
##          FLL4M3 FLL4M4 FLL8F3   FL02F7   FL03F1   FL03F2   FL03F4
## TRINITY_DN30747_c0_g1 0.019432      0      0 0.18223 0.402248 0.198879      0
##          FL08F3   FLT2M3   FLT3M5   FLT4M4   FLT5M3   FLT8M7
## TRINITY_DN30747_c0_g1 0.016844 0.107019 0.017097 0.030756 0.180107 0.296205

## using 'avgTxLength' from assays(dds), correcting for library size

```

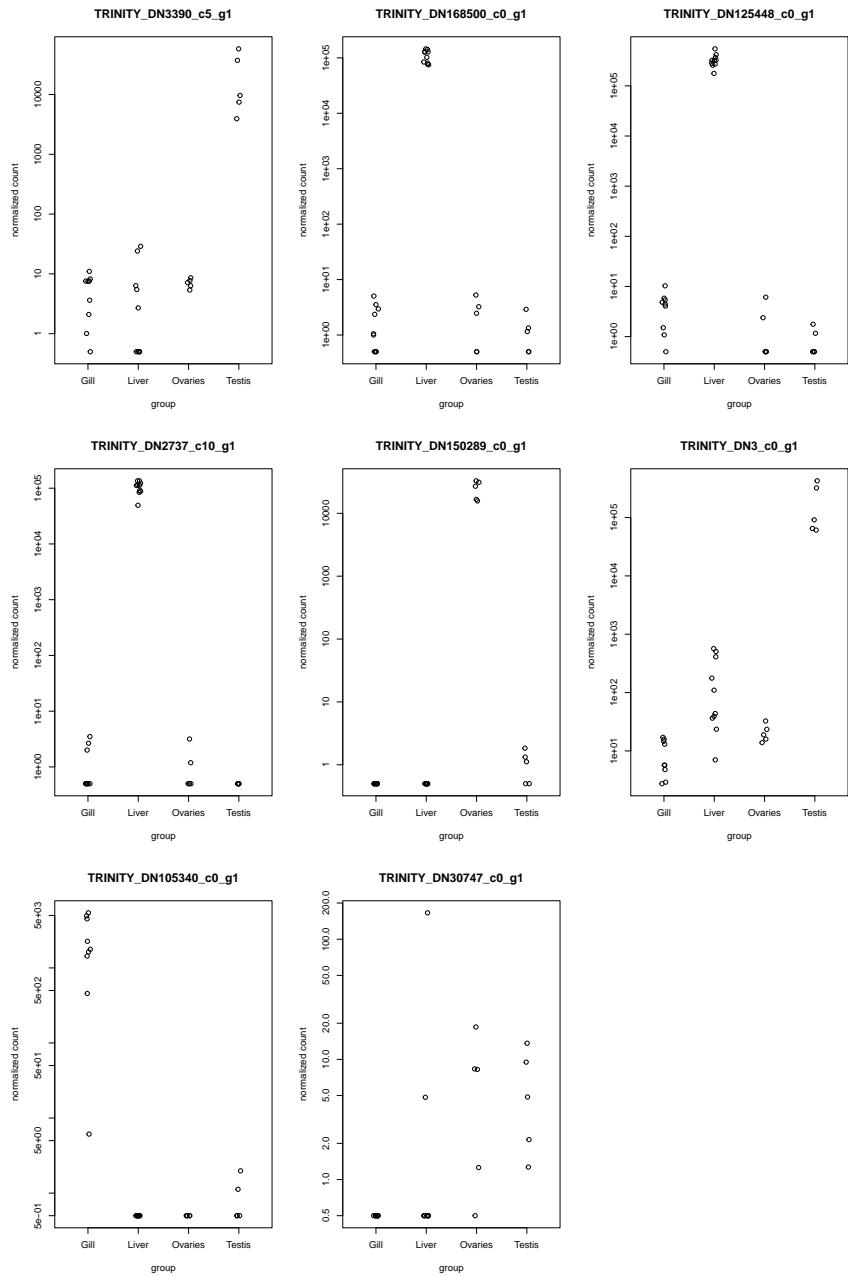


Figure 3: Plots of the counts from the DESeq2 Dataset for all of the genes that had a tau greater than 10⁻¹⁰

Sex Bias and Tissue Specificity

From the DESeq2 analysis we can pull out the FC information for all of the sex biased genes and plot that against the τ calculations.

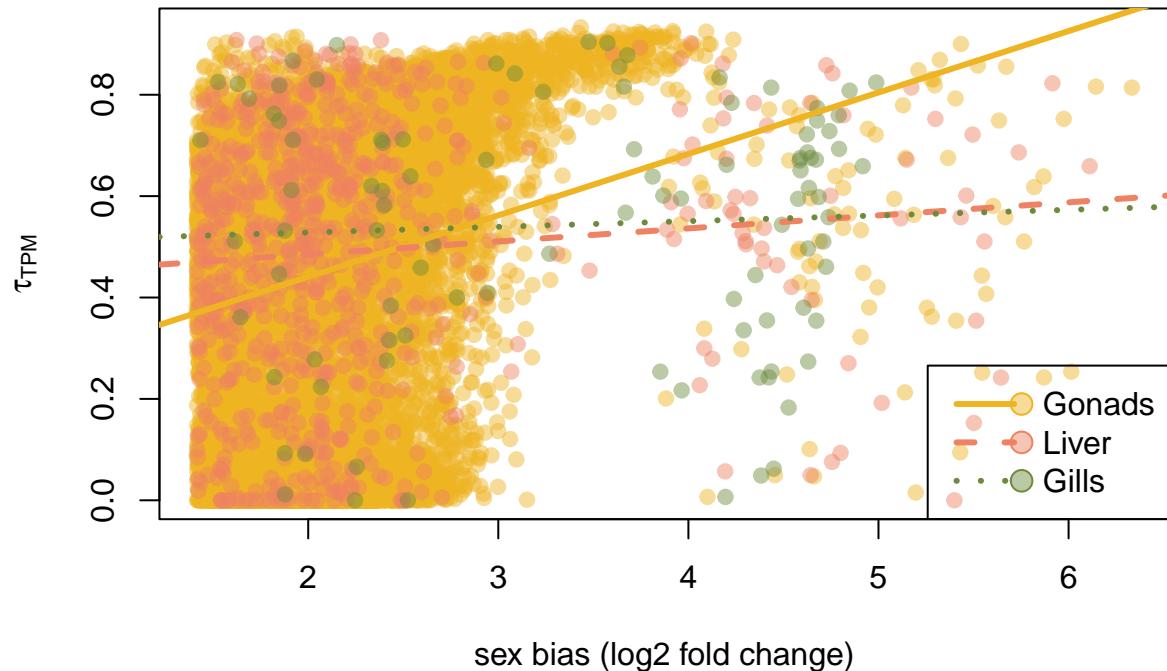


Figure 4: Sex bias (in terms of the absolute value of the log2FoldChange) versus tissue specificity (τ) for all genes that are sex biased in the gonads, liver, and gills.

```
## Warning in cor.test.default(gonad_bias_tau$tau,
## abs(gonad_bias_tau$log2FoldChange), : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: gonad_bias_tau$tau and abs(gonad_bias_tau$log2FoldChange)
## S = 2.7911e+11, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2120436

## Warning in cor.test.default(liver_bias_tau$tau,
## abs(liver_bias_tau$log2FoldChange), : Cannot compute exact p-value with ties

##
```

```

##  Spearman's rank correlation rho
##
## data: liver_bias_tau$tau and abs(liver_bias_tau$log2FoldChange)
## S = 70362832, p-value = 0.0003511
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.1272572

## Warning in cor.test.default(gill_bias_tau$tau,
## abs(gill_bias_tau$log2FoldChange), : Cannot compute exact p-value with ties

##  Spearman's rank correlation rho
##
## data: gill_bias_tau$tau and abs(gill_bias_tau$log2FoldChange)
## S = 158225, p-value = 0.8328
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.02149048

```

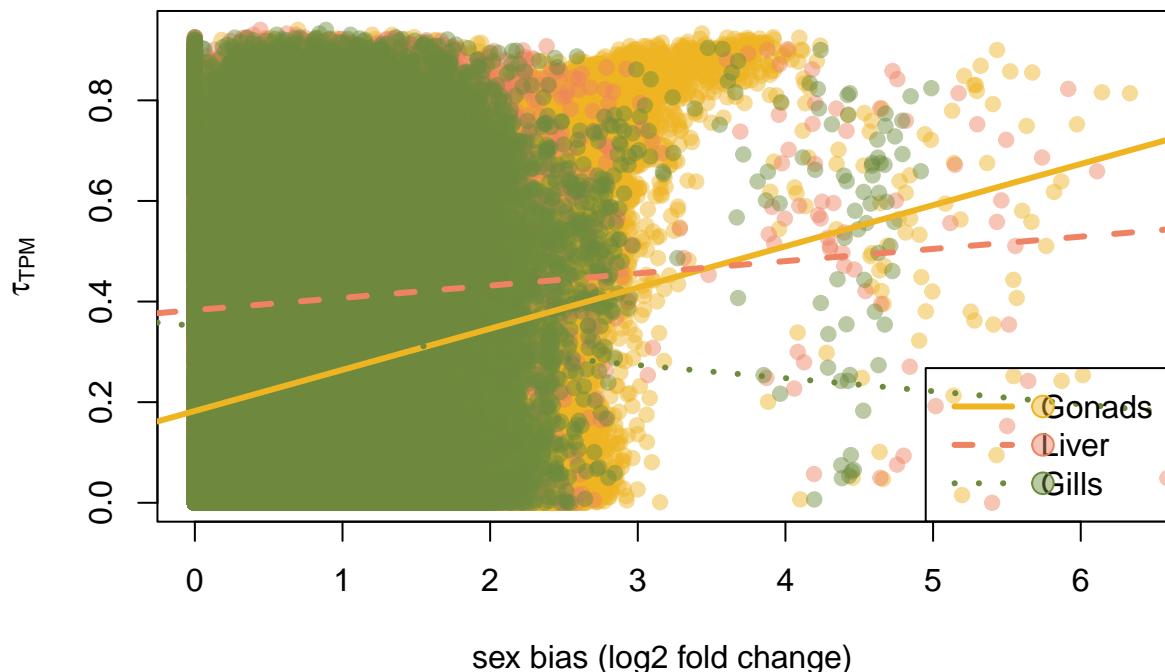


Figure 5: Sex bias (in terms of the absolute value of the log2FoldChange) versus tissue specificity (τ) for all genes that are significantly expressed in the gonads, liver, and gills.

```

## Warning in cor.test.default(gonad_bias_tau_all$tau,
## abs(gonad_bias_tau_all$log2FoldChange), : Cannot compute exact p-value with
## ties

##
## Spearman's rank correlation rho
##
## data: gonad_bias_tau_all$tau and abs(gonad_bias_tau_all$log2FoldChange)
## S = 8.699e+13, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.1772599

## Warning in cor.test.default(liver_bias_tau_all$tau,
## abs(liver_bias_tau_all$log2FoldChange), : Cannot compute exact p-value with
## ties

##
## Spearman's rank correlation rho
##
## data: liver_bias_tau_all$tau and abs(liver_bias_tau_all$log2FoldChange)
## S = 8.5996e+12, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.04882487

## Warning in cor.test.default(gill_bias_tau_all$tau,
## abs(gill_bias_tau_all$log2FoldChange), : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: gill_bias_tau_all$tau and abs(gill_bias_tau_all$log2FoldChange)
## S = 3.7499e+13, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.09779226

## estimating size factors

## using 'avgTxLength' from assays(dds), correcting for library size

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

```

```

## fitting model and testing

## -- replacing outliers and refitting for 37383 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)

## estimating dispersions

## fitting model and testing

```

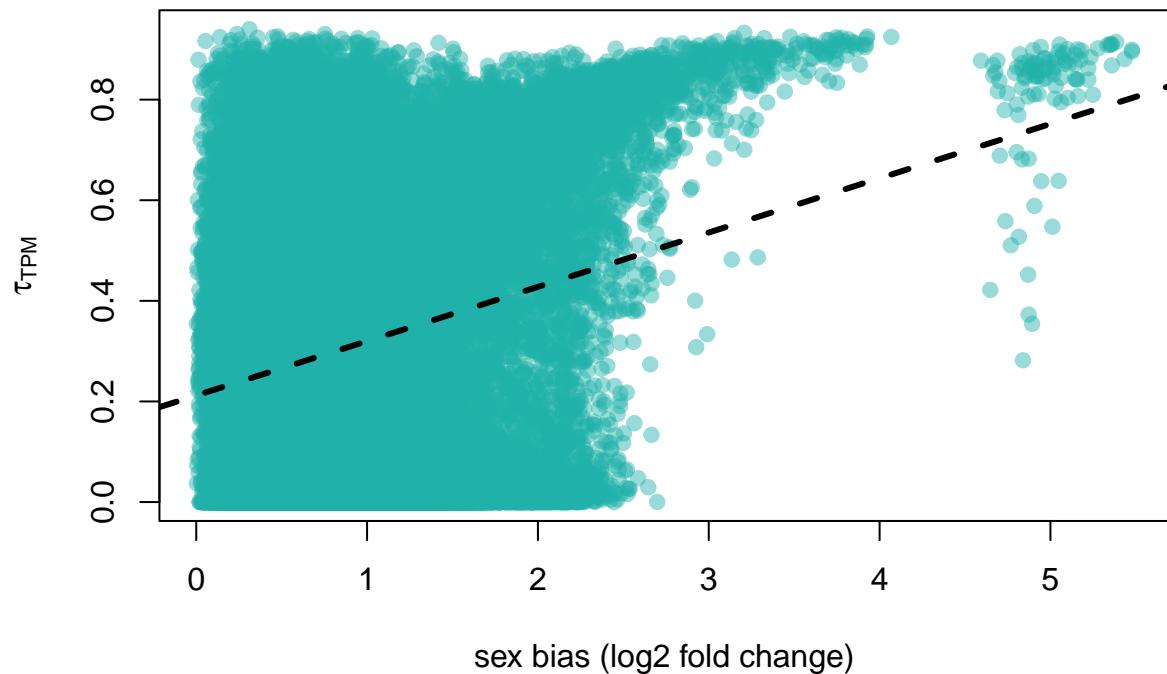


Figure 6: Sex bias (in terms of the absolute value of the log2FoldChange) versus tissue specificity (τ) for all genes that are significantly expressed. The fold change in this plot is a difference between males and females regardless of tissue type.

```

## Warning in cor.test.default(MFbias_tau_all$tau,
## abs(MFbias_tau_all$log2FoldChange), : Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: MFbias_tau_all$tau and abs(MFbias_tau_all$log2FoldChange)
## S = 3.6718e+13, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:

```

```
##      rho
## 0.1254565
```

To further investigate this I categorized the sex biased genes based on a series of logFC thresholds to see if we can get a better idea about what relationship may lie between sex bias and τ . After that I can plot τ against the bias level.

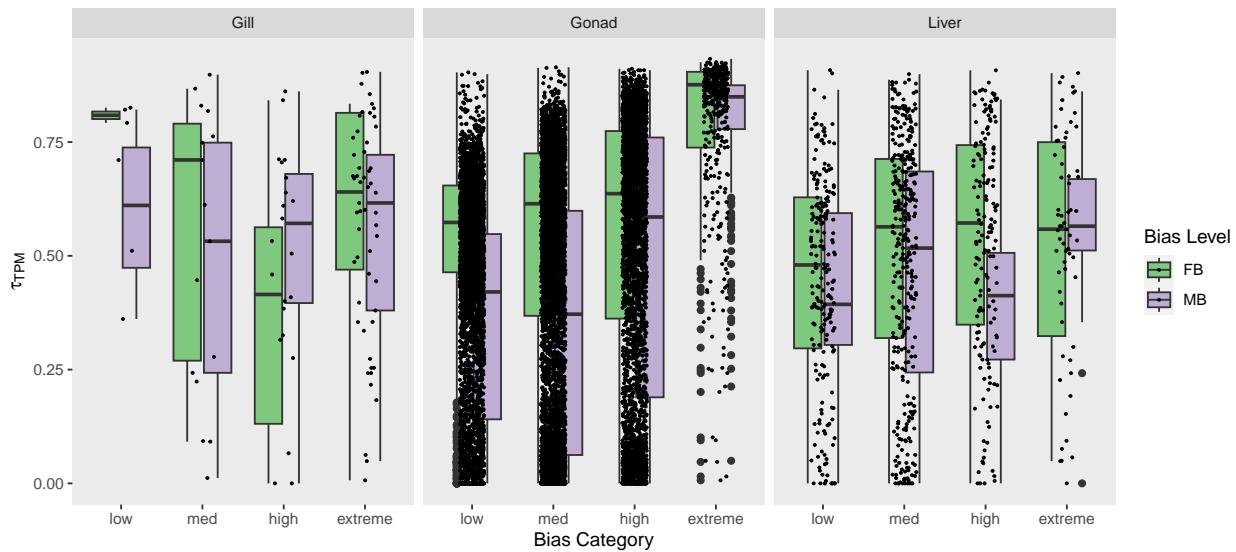
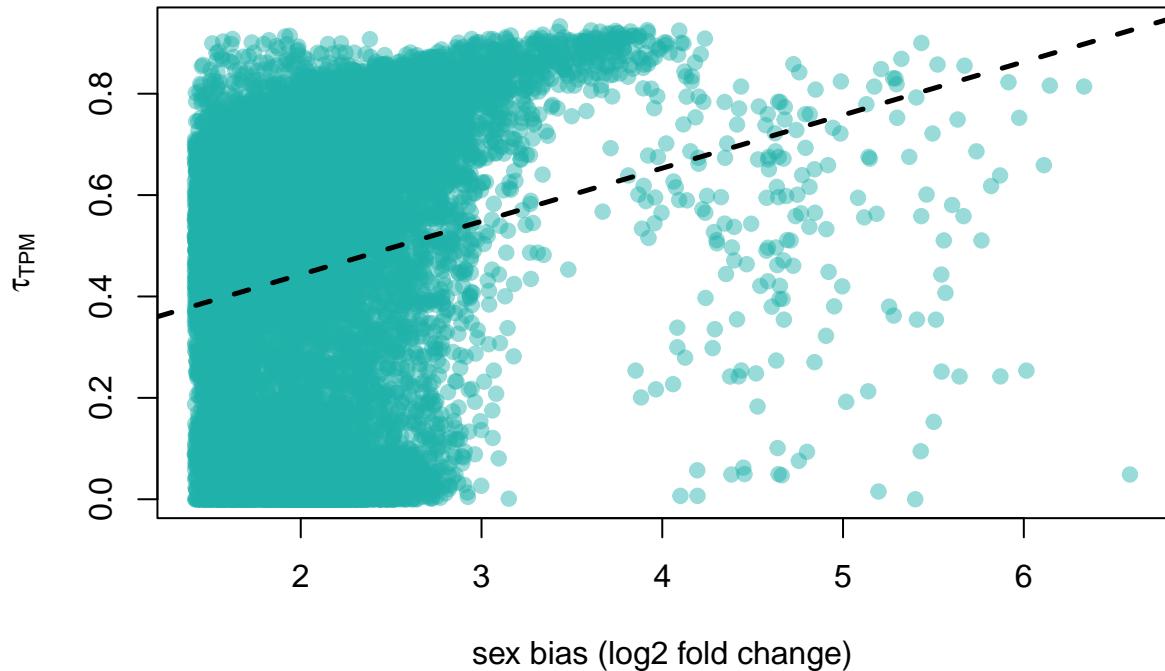


Figure 7: Tissue specificity (τ) across the different bias levels. Color denotes female-biased versus male biased and jitters were added to show all of the raw τ values.

```
plot(sqrt(abs(logFC_long$logFC)),
  logFC_long$tau,
  pch=19,
  col = "#20B2AA75",
  xlab="sex bias (log2 fold change)",
  ylab=expression(tau["TPM"]))
abline(lm(logFC_long$tau ~ sqrt(abs(logFC_long$logFC))), lwd = 2.5, lty = 2)
```



```
cor.test(logFC_long$tau, sqrt(abs(logFC_long$logFC)), method = "spearman")
```

```
## Warning in cor.test.default(logFC_long$tau, sqrt(abs(logFC_long$logFC)), :
## Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: logFC_long$tau and sqrt(abs(logFC_long$logFC))
## S = 3.4264e+11, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.207623
```

