Associated Repos
CEABiGR Main - https://github.com/sr320/ceabigr

oyster-lnc - https://github.com/zbengt/oyster-lnc

Intro/Background

General lncRNA:

- Unique features of long non-coding RNA biogenesis and function - https://www.nature.com/articles/nrg.2015.10
    - Long non-coding RNAs (lncRNAs) are a diverse class of RNAs that engage in numerous biological processes across every branch of life. Although initially discovered as mRNA-like transcripts that do not encode proteins, recent studies have revealed features of lncRNAs that further distinguish them from mRNAs. Review describes special events in the lifetimes of lncRNAs and discusses how these events ultimately shape the unique characteristics and functional roles of lncRNAs.
- LncRNAs insights into functions - https://www.nature.com/articles/nrg2521
    - In mammals and other eukaryotes most of the genome is transcribed in a developmentally regulated manner to produce large numbers of long non-coding RNAs (ncRNAs). Rapidly advancing field of long ncRNAs, describing their conservation, their organization in the genome and their roles in gene regulation. Medical implications, and the emerging recognition that any transcript, regardless of coding potential, can have an intrinsic function as an RNA.
- Classification of lncRNAs - https://www.tandfonline.com/doi/full/10.4161/rna.24604
    - Summary of classification methods of lncRNAs according to their four major features, namely, genomic location and context, effect exerted on DNA sequences, mechanism of functioning and their targeting mechanism. In combination with the presently available function annotations, they explore potential relationships between different classification categories, and generalize and compare biological features of different lncRNAs within each category. Classifications of lncRNAs are of fundamental importance for lncRNA studies, helpful for further investigation of specific lncRNAs, for formulation of new hypothesis based on different features of lncRNA and for exploration of the underlying lncRNA functional mechanisms
- Idenfification and functions of lncRNAs - https://www.frontiersin.org/articles/10.3389/fncel.2013.00168/full
    - Recent studies of lncRNAs across different species have revealed a diverse population of RNA molecules of differing size and function. RNA sequencing studies suggest transcription throughout the genome, so there is a need to understand how sequence relates to functional and structural relationships amongst RNA molecules. Synthesis of recent studies suggests that neither size, presence of a poly-A tail, splicing, direction of transcription, nor strand specificity

are of importance to lncRNA function. Rather, relative genomic position in relation to a target is fundamentally important.

LncRNAs in marine organisms and molluscs

- Functional role of non-coding RNAs in molluscs - https://www.sciencedirect.com/science/article/pii/S0378111920309690?via%3Dihub
    - Focuses on the current knowledge of non-coding RNAs (ncRNAs) in molluscs. In this review, we provide an overview of long ncRNAs (lncRNA), microRNAs (miRNA) and piwi-interacting RNAs (piRNA), followed by evidence for the regulation of ncRNAs in variety of biological process in molluscs. Roles of ncRNAs in molluscs and suggests the future direction to fully understand the epigenetic regulatory network of molluscs.
- Shell growth - https://www.sciencedirect.com/science/article/pii/S2090123223003648?via%3Dihub
    - In bivalves with asymmetric shells, DEGs are more abundant in the flat valve mantle. LncRNA are new members of the bivalve biomineralization toolbox. Biomineralization-related genes in the mantle are under cis regulation by lnRNAs. Asymmetric Ostreidae family members share conserved regulatory gene modules. Gene modules regulate calcium carbonate crystal growth and spatial orientation.

LncRNAs in bivalves

- Patinopecten yessoensis
    - mRNA-lncRNA networks regulating melinization and biomineralization https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-023-09837-w
        - This research investigates the role of long non-coding RNAs (lncRNAs) in combating Polydora disease in Patinopecten yessoensis, a key shellfish in Asian aquaculture facing significant losses due to the disease. Through transcriptome analysis, it identifies numerous lncRNAs and demonstrates their involvement in crucial defense mechanisms like melanization and biomineralization, revealing differences between healthy and infected scallops. The study constructs an interactive network between lncRNAs and mRNAs, providing insights into the molecular defense against Polydora, and offers resources for developing disease-resistant mollusc varieties.
- Mytilus galloprovinciallus
    - DE in hemocytes under immune stimuli - https://www.mdpi.com/2073-4425/12/9/1393
        - Second characterization and study of lncRNAs in this bivalve species. In this work, we identified the potential repertoire of lncRNAs expressed in mussel hemocytes, and using RNA-Seq we analyzed the lncRNA profile of mussel hemocytes stimulated in vitro with three different immune stimuli: LPS, poly I:C, and β-glucans. Compared to unstimulated hemocytes, LPS induced the highest modulation of lncRNAs, whereas poly I:C and β-glucans induced a similar discrete response. Based on the potential cis-regulatory activity of the lncRNAs, identified the neighboring

protein-coding genes of the regulated lncRNAs to estimate—at least partially—the processes in which they are implicated. After applying correlation analyses, it seems that—especially for LPS—the lncRNAs could participate in the regulation of gene expression, and substantially contribute to the immune response.

- Ruditapes phillipinarium
    - Pigmentation related lncRNAs - https://www.sciencedirect.com/science/article/abs/pii/S0044848623006634?via%3Dihub
        - Long non-coding RNAs in Manila clams was performed using Illumina sequencing platform. Key pathways related to pigmentation in different shell color Manila clam were analyzed. The results uncover the molecular regulation of shell color diversity in R. philippinarum.
- Pinctada fucata martensii
    - Biomineralization - http://dx.doi.org/10.3389/fmars.2022.1014810
        - Cloned the complete sequence of a mantle-specific expressed long non-coding RNA (designated as LncMPEG1) from a pearl oyster, Pinctada fucata martensii. A quantitative real-time PCR analysis showed that LncMPEG1 expression was significantly high in early umbo larvae and juveniles, which would be the critical periods of shell development. LncMPEG1 was identified in the outer epithelium of the middle fold from the mantle edge, mantle pallial, and mantle center by using in situ hybridization. Additionally, the expression of LncMPEG1 was stimulated by shell damage, alien invasion, heat and cold temperature stress, and hypoxia stress. In the mantle, a decreased in LncMPEG1 expression was detected by RNA interference, which can cause the irregular growth of crystals on the inner surface of the prismatic layer and nacre in the shells.
- Crassostrea gigas
    - Regulation of glycogen content - https://www.sciencedirect.com/science/article/abs/pii/S0044848620308577?via%3Dihub
        - Non-coding transcriptome analysis of high glycogen content of Crassostrea gigas was studied for the first time. Networks was drawn based on correlation between non-coding RNAs associated with differences in glycogen content. These findings explained the molecular mechanism of C. gigas with high glycogen content from non-coding RNA aspect. The results filled the non-classical inheritance of quality traits in marine economic shellfish.
    - Antiviral immune response - https://www.sciencedirect.com/science/article/abs/pii/S1050464818304303?via%3Dihub
        - 101 differentially expressed lncRNAs during Ostreid herpesvirus infection were identified. LncRNAs are expressed in a time-dependent manner during viral challenge. LncRNAs may play a role in binding and cellular

process functions in cis mode. 14 lncRNAs regulate the expression of 17 antiviral immune-related mRNAs.

- Larval development - https://www.nature.com/articles/srep20796
    - First identification and characterization of lincRNAs in the Pacific oyster (Crassostrea gigas). We developed a pipeline and identified 11,668 lincRNAs in C. gigas based on RNA-Seq resources available. These lincRNAs exhibited many common characteristics with vertebrate lincRNAs: relatively short length, low exon numbers, low expression and low sequence conservation. 1,175 lincRNAs were expressed in a tissue-specific manner, with 35.2% preferentially expressed in male gonad. 776 lincRNAs were specifically expressed in juvenile during different developmental stages. In addition, 47 lincRNAs were found to be potentially related to oyster settlement and metamorphosis. Such diverse temporal and spatial patterns of expression suggest that these lincRNAs might function in cell differentiation during early development, as well as sex differentiation and reproduction. Based on a co-expression network analysis, five lincRNAs were detected that have an expression correlation with key hub genes in four modules significantly correlated with larval development.
- Shell pigmentation - https://www.nature.com/articles/s41598-018-19950-6
    - A total of 427 lncRNAs and 349 mRNAs were identified to differentially express among six pairwise groups, mainly involving in biomineralization and pigmentation through functional enrichment. A total of 6 mRNAs and their cis-acting lncRNAs were predicted to involve in synthesis of melanin, carotenoid, tetrapyrrole, or ommochrome. Of them, chorion peroxidase and its cis-acting lincRNA TCONS_00951105 are implicated in playing an essential role in the melanin synthetic pathway. First systematic characterization of lncRNAs catalog expressed in oyster mantle, which may facilitate understanding the molecular regulation of shell colour diversity and provide new insights into future selective breeding of C. gigas for aquaculture.

Methods

Data

Male and female eastern oysters (*Crassostrea virginica*) in the treatment group were exposed to elevated pCO_2 conditions (2500 µatm, Ωcalcite < 1) for comparison against oysters in the control group (500 µatm, Ωcalcite > 1), with a total of 26 samples sequenced. RNA-seq data were quality controlled and trimmed. The reference genome for *Crassostrea virginica* was obtained from NCBI. Long non-coding RNA tracks were downloaded from NCBI to identify previously discovered lncRNAs in the eastern oyster genome. A new GFF and FASTA were created to

Size Distribution, Location, and Comparison to Other Species

Code -
https://github.com/zbengt/oyster-lnc/blob/main/code/09-size-distribution-and-location.Rmd

Lengths of lncRNAs were retrieved from the FASTA and binned by every 500nt to display size distribution. BedTools intersect was used to match the locations of known lncRNAs with the genomic GFF provided of the species. Results were summarized to provide the location of lncRNAs relative to genome features (exons, genes, and mRNAs).

Comparison of lncRNAs with other bivalve species was completed using BLAST. LncRNA FASTAs for *Crassostrea gigas*, *Patinopecten yessoensis*, *Mytilus galloprovinciallus*, and *Ruditapes phillipinarium* were blasted against a *Crassostrea virginica* lncRNA database. Special attention was paid to those lncRNAs identified in previous literature related to pigmentation, biomineralization, and glycogen content. Each FASTA from the listed species was filtered to include only those sequences classified as ncRNA (NCBI standard classification). Two *Crassostrea gigas* genomes were used to create reference two separate FASTAs for this species, representing a commonly used reference as well as a more recent genome. BLAST results were joined into a single table for match summarization to *Crassostrea virginica*.

Differential Expression

lncRNA kallisto and DESeq2 -
https://github.com/zbengt/oyster-lnc/blob/main/code/10-count-matrices-DESeq2-final.Rmd
mRNA kallisto and DESeq2 -
https://github.com/zbengt/oyster-lnc/blob/main/code/11-mRNA-counts-DESeq2.Rmd

Kallisto was used to create indices from the FASTAs of lncRNAs and coding sequences. Trimmed reads and indices were used in kallisto to retrieve abundance values. Trinity generated count matrices for both lncRNAs and coding sequences from kallisto abundance estimates. The resulting count matrices were subsetted to obtain sex specific count matrices for males and females. DESeq2 was used to complete differential expression analysis on combined sex count matrices, with both sex and OA exposure treatment included in the model. Male and female specific count matrices were also run through DESeq2 to specifically examine differential expression of treatment without the impact of sex.

Weighted Gene Co-expression Network Analysis (WGCNA)

lncRNA WGCNA -
https://github.com/zbengt/oyster-lnc/blob/main/code/12-WGCNA-lncRNA.Rmd

WGCNA creates modules of RNAs with similar expression patterns using network analysis. Modules are then used to examine differences between conditions, like treatment, sex, and

RNA type. WGCNA settings reflect experimental design, desired sensitivity of analysis, and reasonable filtering to limit the scope of analysis. The settings reflect proportion of samples in which a RNA transcript must occur, minimum number of counts, minimum module size, and sensitivity of splitting for module generation. Co-expression of RNAs was examined by including both lncRNAs and mRNAs in this analysis.

Settings secisions were made to accurately reflect treatments within the experiment, reduce noise from lowly and infrequently expressed lncRNAs, and ensure module generation was sensitive but did not generate an unmanageable number of modules. Decisions were implemented in the three following steps.

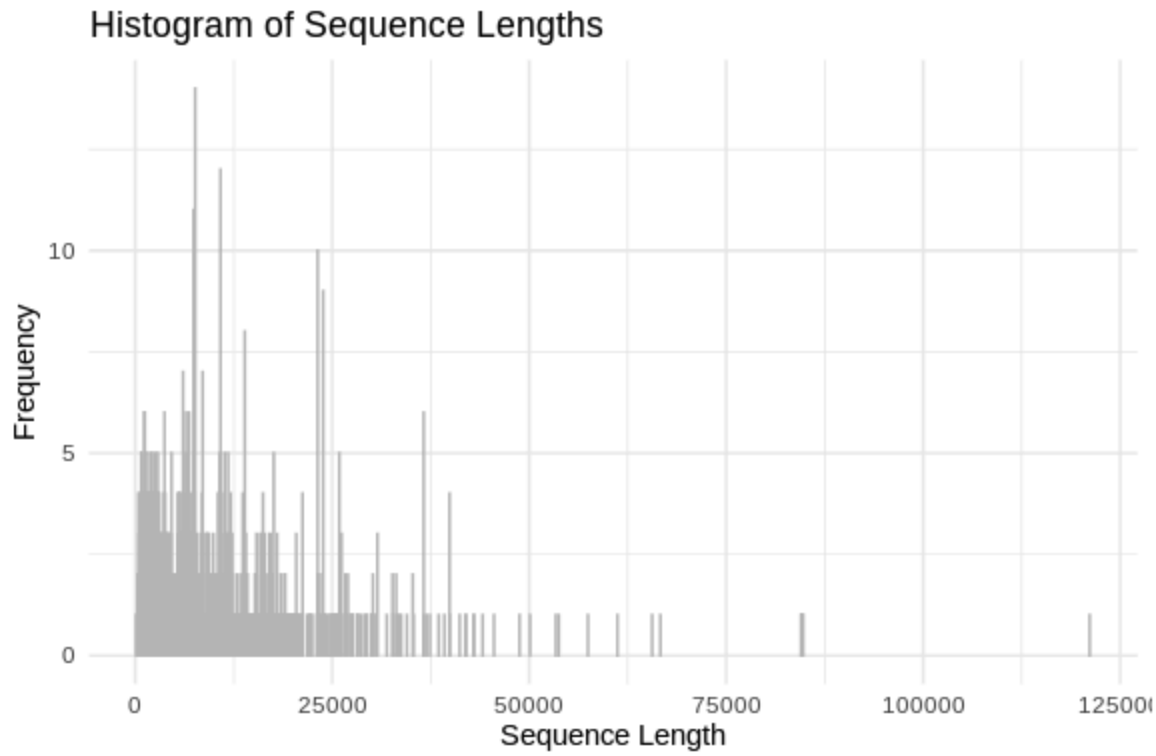
(1) pOverA(0.5, 20) - Setting pOverA to this means a lncRNA must be expressed in half of the samples to be included in the analysis with a count of at least 20. Half of samples are control and half are exposed to OA conditions, so we might expect similar expression patterns based on treatment. Limiting to at least 20 removes lowly expressed lncRNAs.
(2) Minimum module size: 40 - This means a module must include at least 40 lncRNAs. We tested multiple min module sizes and noted an increasing trend of more modules as we increased the number to 30. There was a drop of and stabilization of modules generated between 30 and 40. So we went with 40.
(3) deepSplit: 3 - deepSplit influences how aggressively the clustering dendrogram is cut into separate modules. It can take on values typically in the range of 0 to 4. 1 to 3 Indicates an increasingly aggressive approach to splitting, with higher values resulting in more and smaller modules. These intermediate values allow for a balance between sensitivity and specificity in module detection. We went with 3 to ensure sensitivity in module separation, but not so much as to inflate the number of modules created.

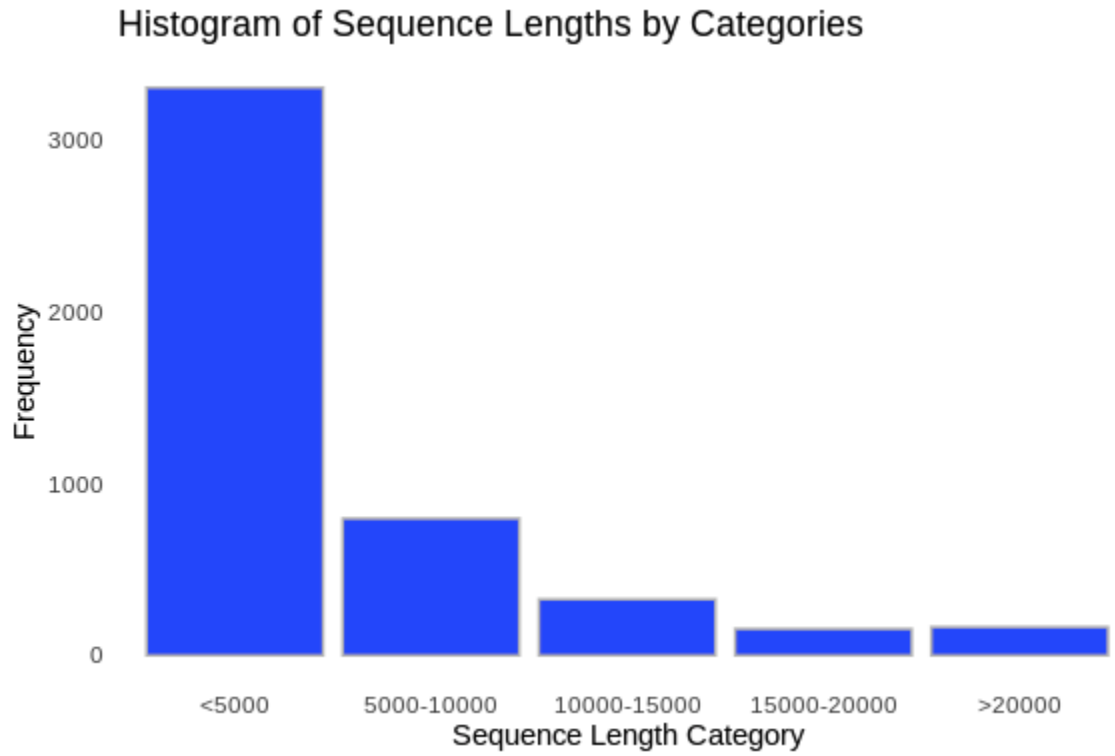
For males, we started with 4750 lncRNAs, removed rows in the count matrix with a sum of 0 (indicating no expression). This brought the number of lncRNAs down to 4515. After application of pOverA with the above settings, 3862 lncRNAs remained. After setting min module size and deepSplit, we were left with 31 modules of lncRNAs. This reduced the number of modules to a manageable size for interpretation of results.

Results

Size and Location

The majority of lncRNAs in the NCBI FASTA are over 1000nt and less than <5000nt.


Histogram of Sequence Lengths
Frequency
10
5
0
0
25000
50000
75000
100000
125000
Sequence Length


Histogram of Sequence Lengths by Categories
Frequency
3000
2000
1000
0
<5000
5000-10000
10000-15000
15000-20000
>20000
Sequence Length Category

Basics - Species Comparison

Species	ncRNA count
Crassostrea gigas	7247
Patinopecten yessoensis	2203
Mytilus trossulus	6173
Ruditapes phillipinarium	5853
*Magallana gigas**	7177

*Updated 2023 RNA transcripts from NCBI

After BLAST:

Species	Hits
Crassostrea gigas	11992
Patinopecten yessoensis	2207
Mytilus trossulus	6269
Ruditapes phillipinarium	6957
Magallana gigas	13895

Differential Expression

Differential expression accounting for treatment and sex shows many differences in expression of lncRNAs based on sex. There are 2370 differentially expressed lncRNAs between female and male sample groups. Sex specific DESeq2 runs identified far fewer differentially expressed lncRNAs, with five for males and one for females. After separation of DGE analysis based on sex, a small number of differentially expressed lncRNAs were found (1 for females and 5 for males). Differential expression of mRNAs followed a similar pattern after sex separation (107 for males and 25 for females).

lncRNAs

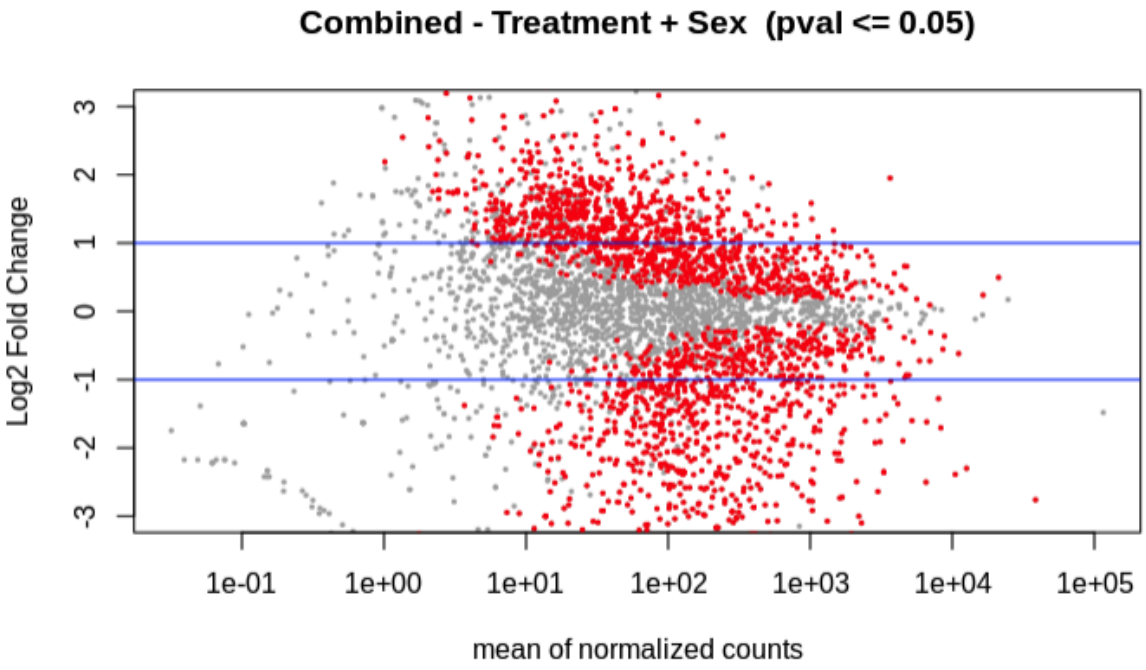


Fig. Volcano plot displaying 2370 differentially expressed lncRNAs from combined sex and treatment DESeq2.

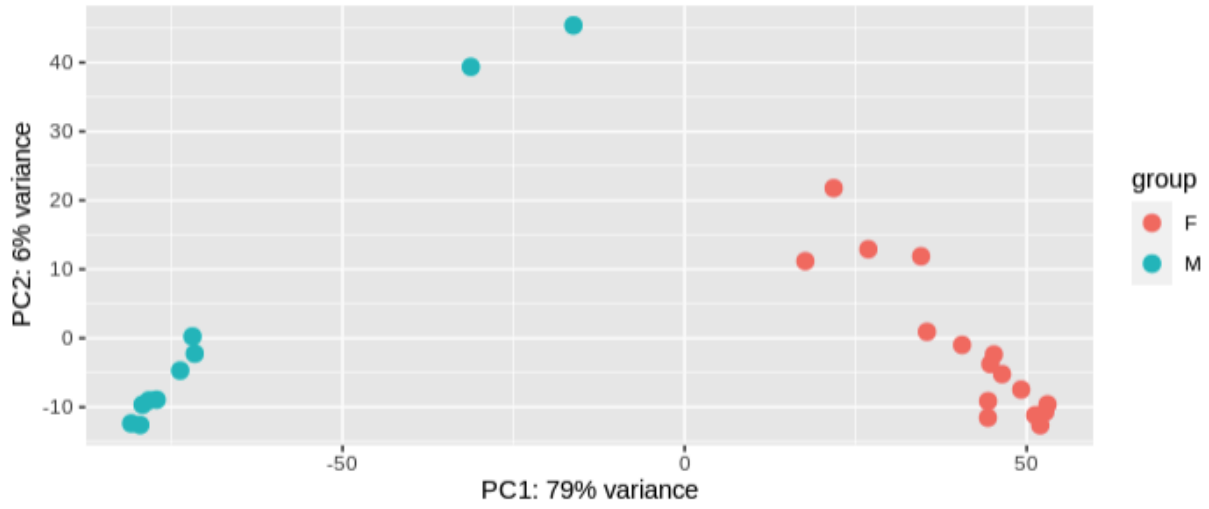


Fig. PCA from combined sex and treatment differential expression displaying sex. Data clusters by sex.

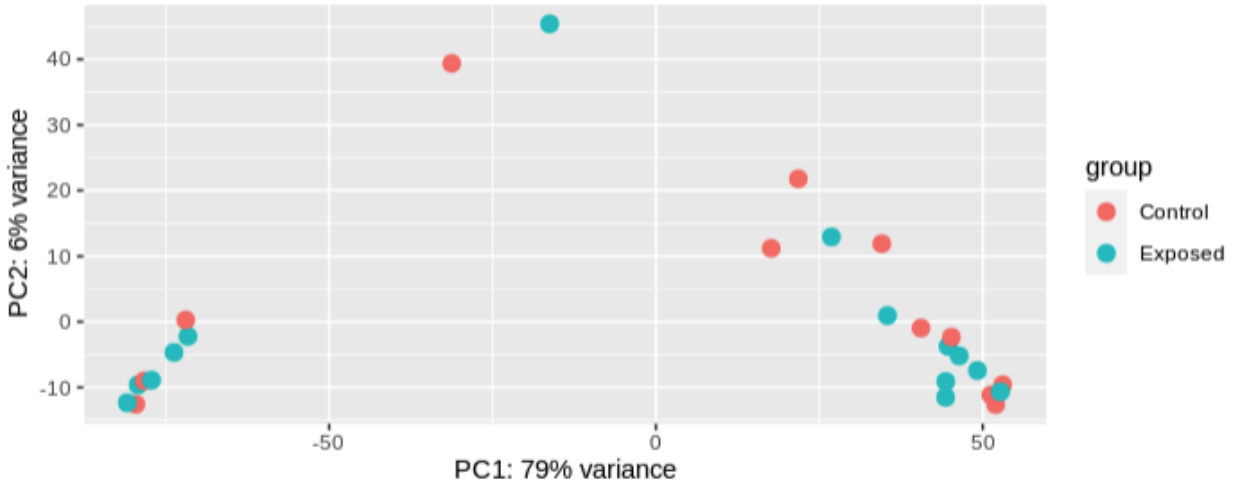
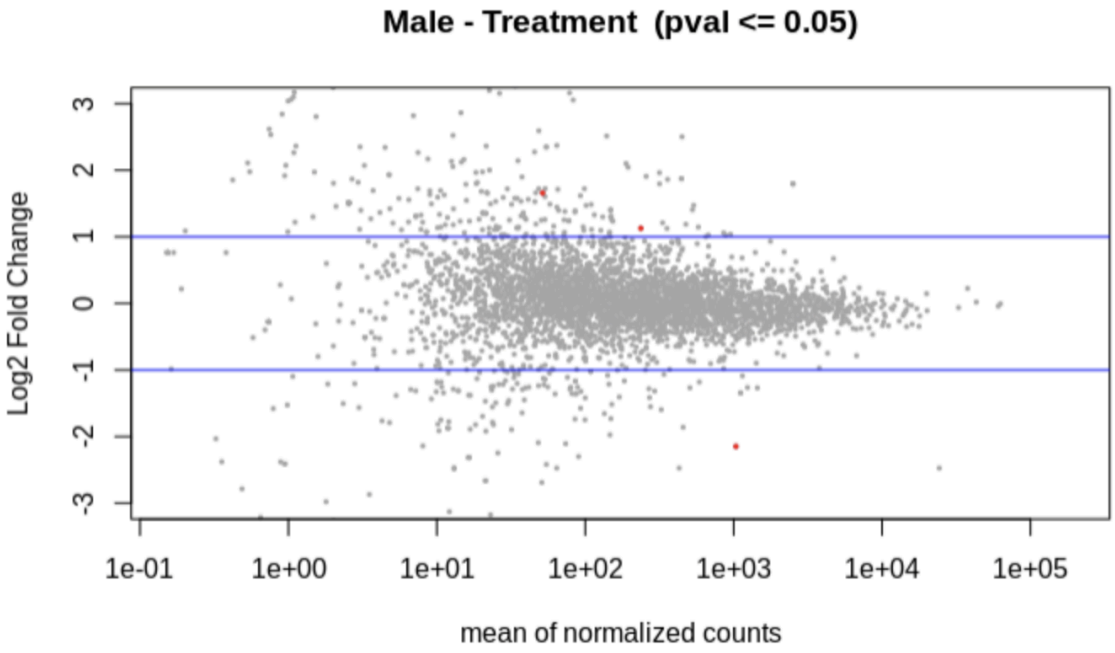


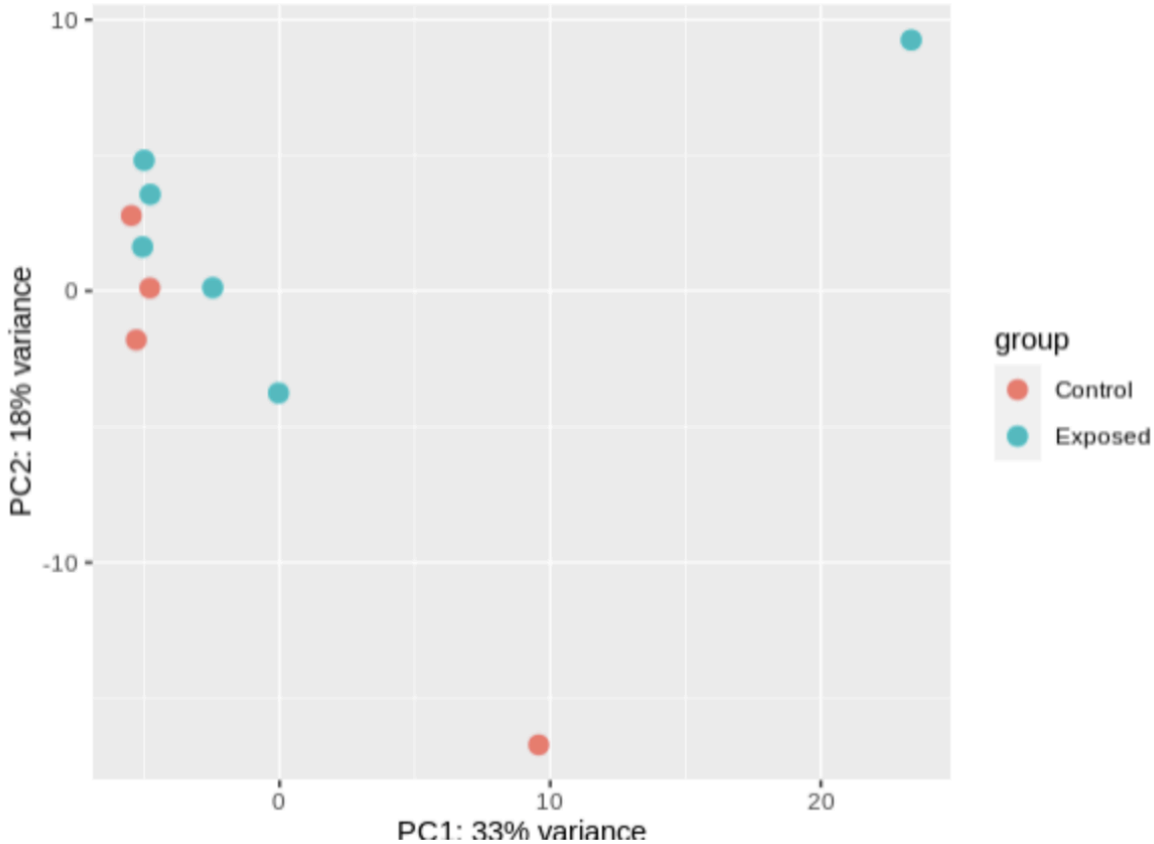
Fig. PCA from combined sex and treatment differential expression displaying treatment. No clustering based on treatment.

Males

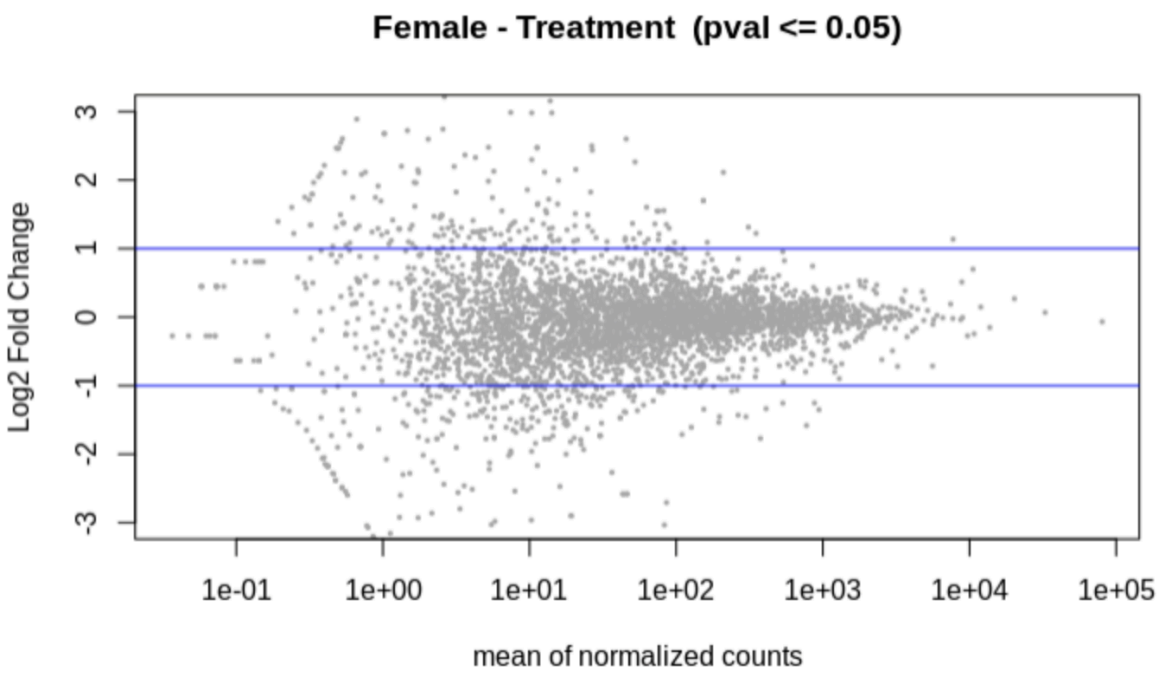


- 5 significantly differentially expressed lncRNAs

PCA - treatment

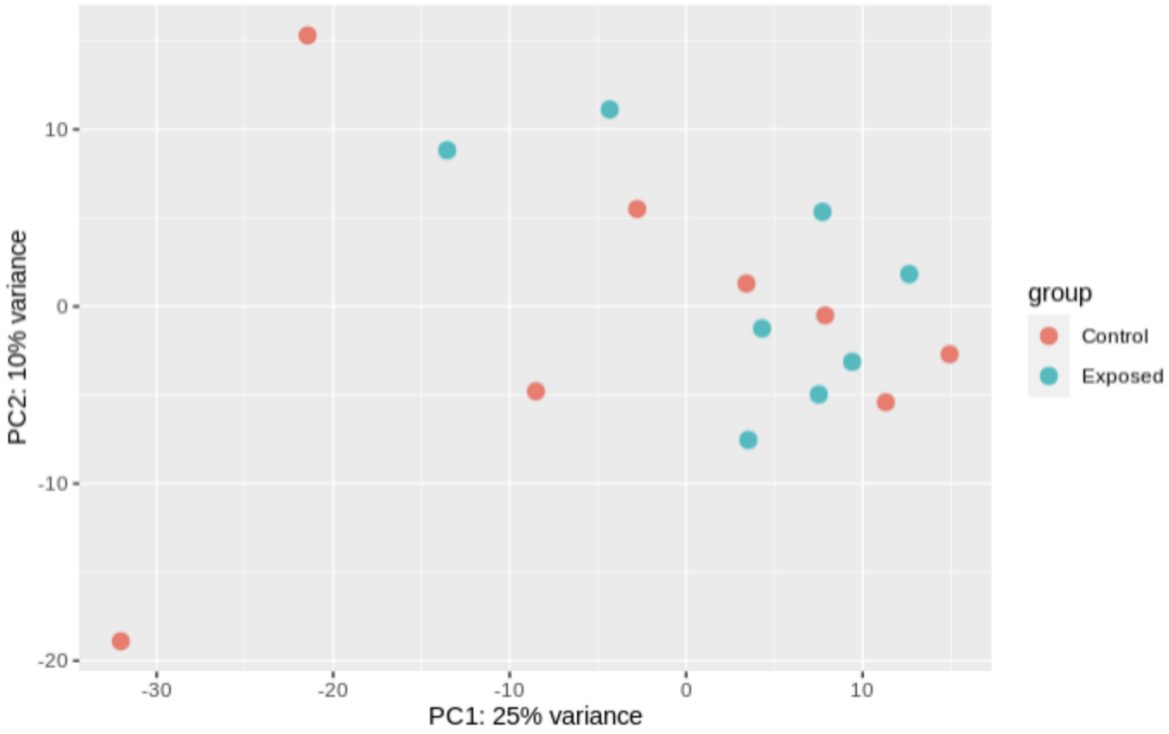


Female



- 1 significantly differentially expressed lncRNA

PCA - treatment



Differential expression of mRNAs revealed a similar pattern in sex differences. 36242 significantly differentially expressed coding sequences resulted from the combined sex and treatment DESeq2 run of mRNAs. In sex specific runs, DESeq2 identified 25 DEGs in females and 107 DEGs in males.

mRNAs

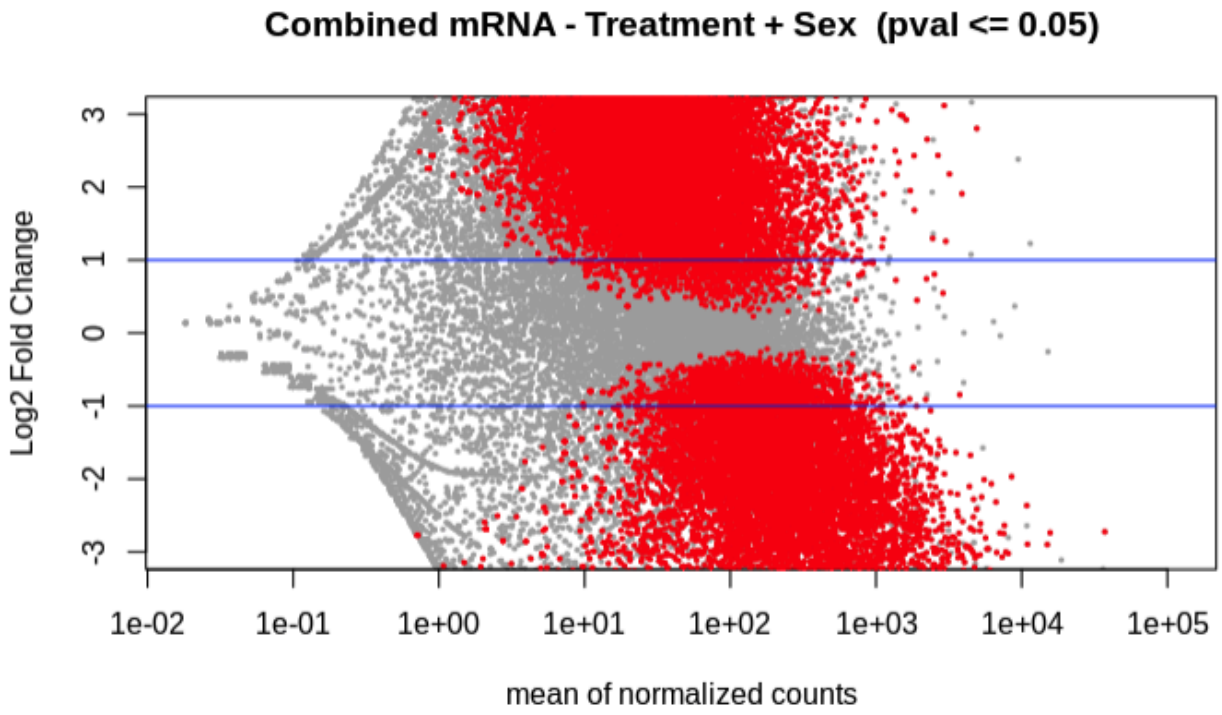
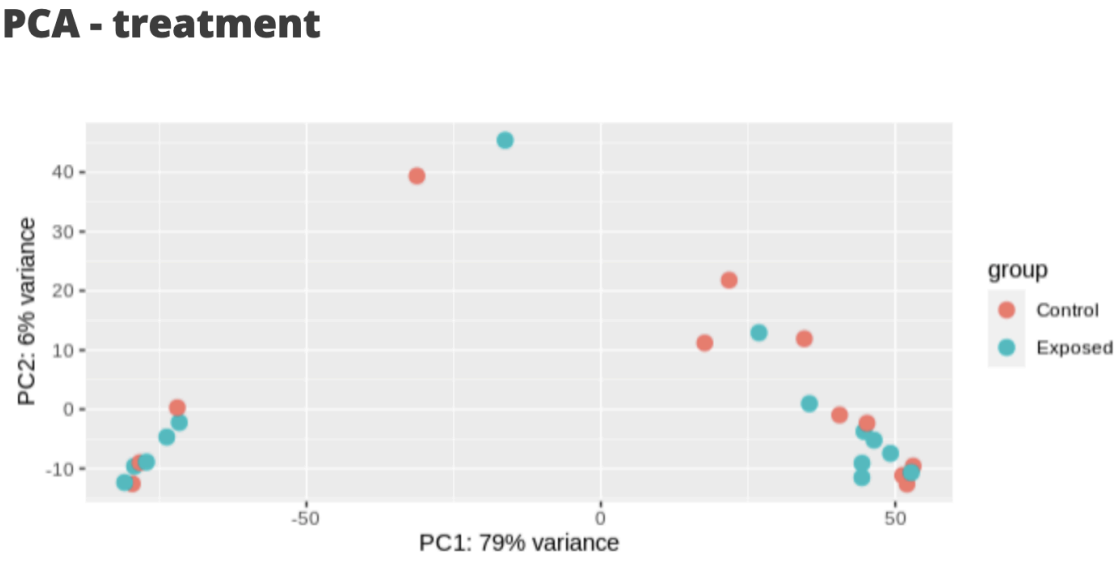


Fig. Volcano plot displaying 36242 DEGs from combined DESeq2 with sex and treatment.



PCA - sex

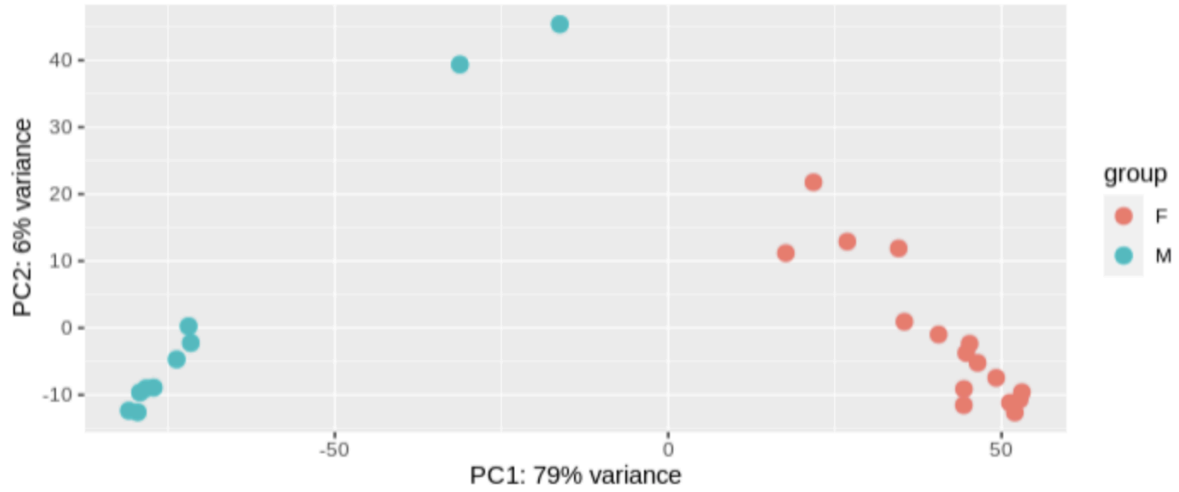
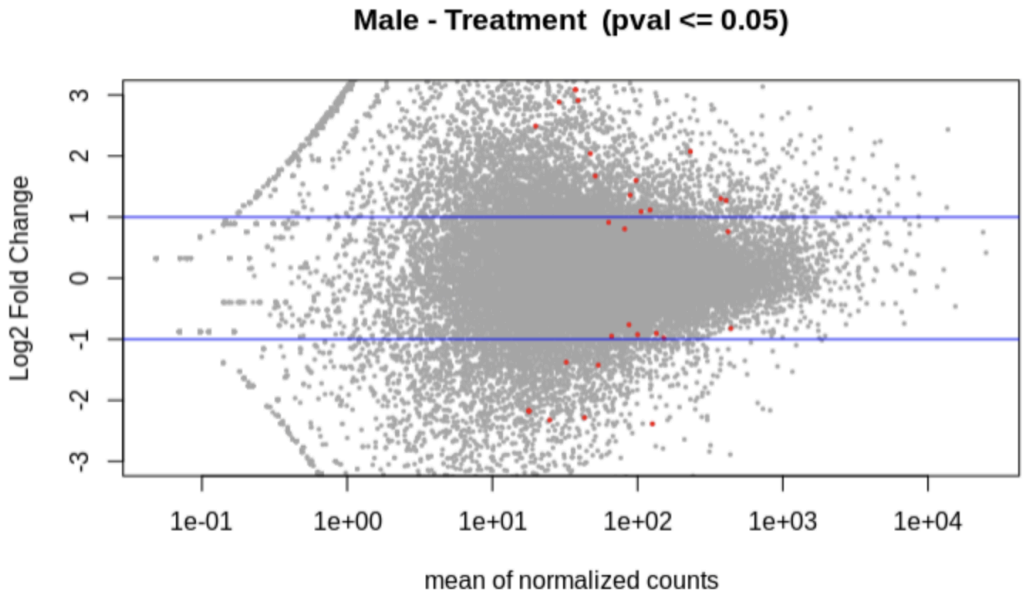


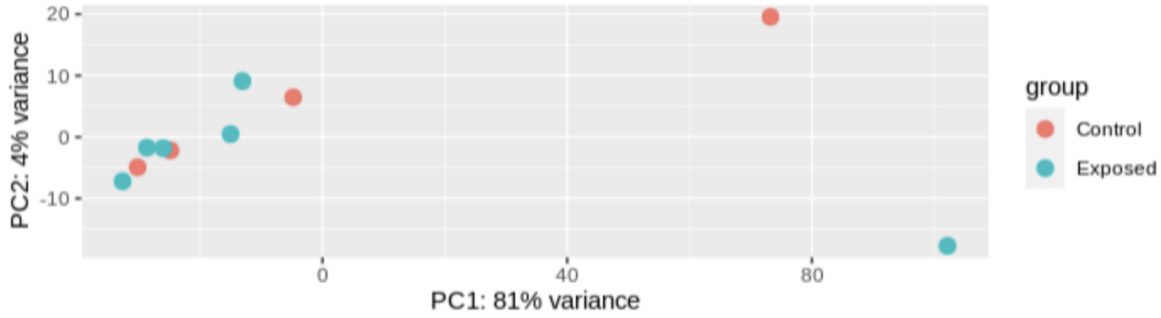
Fig. PCA showing clustering by sex and treatment from differential expression analysis.

Males

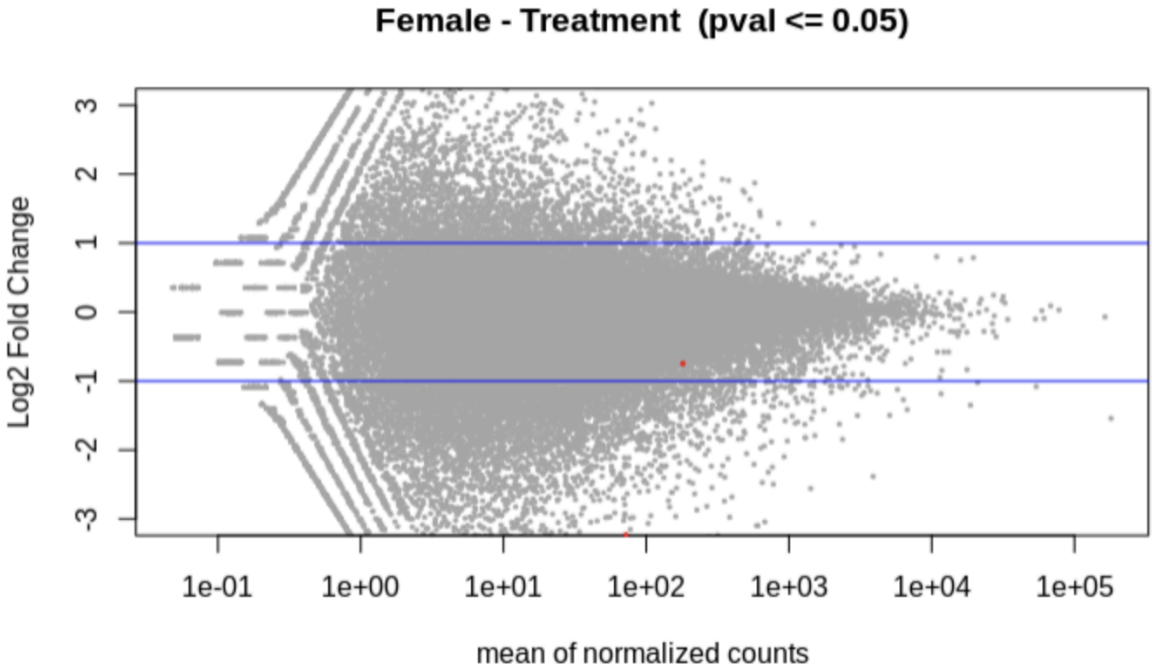


- 107 significantly differentially expressed mRNAs

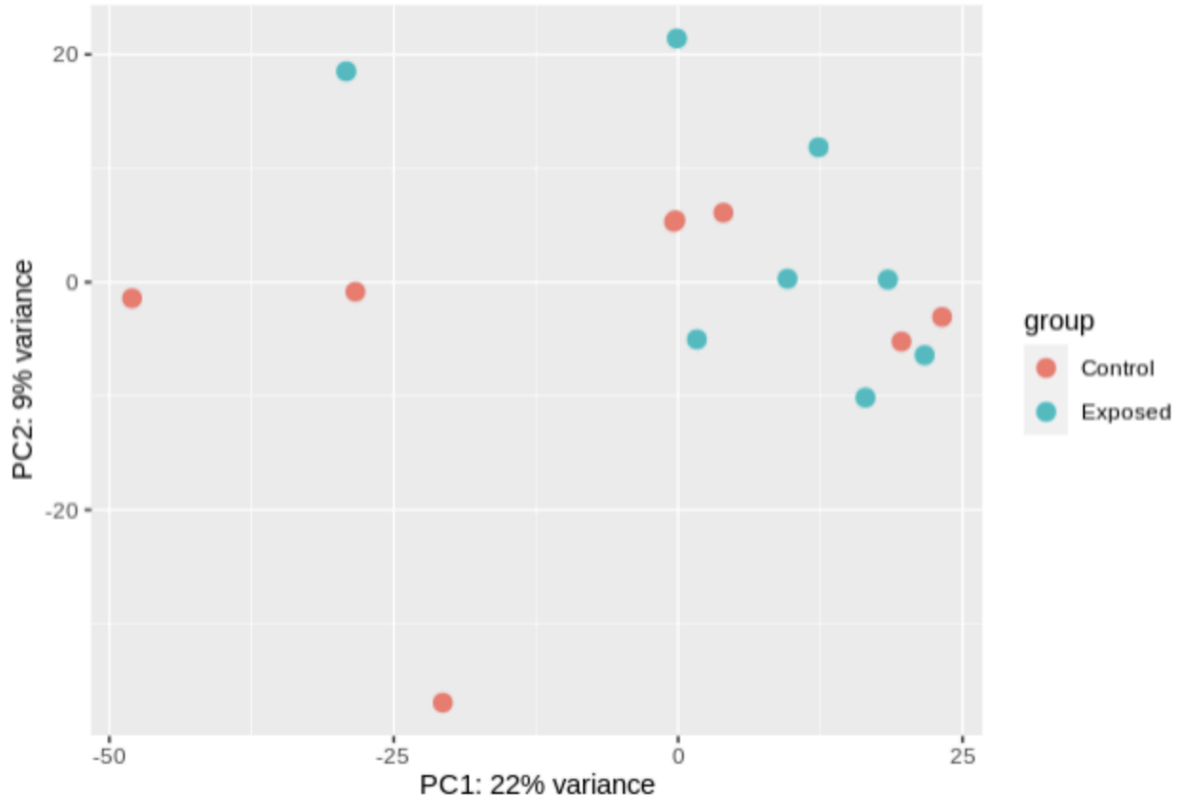
PCA - treatment



Females



- 25 significantly differentially expressed mRNAs



WGCNA

Males

Expression Plots
http://172.25.149.12:8787/files/github/oyster-lnc/output/M_expression_plot.pdf

Heatmap
http://172.25.149.12:8787/files/github/oyster-lnc/output/12-WGCNA-lncRNA/M_heatmap_updated.pdf

Expression Box Plots
http://172.25.149.12:8787/files/github/oyster-lnc/output/M_expression_plot_treatment.pdf

DRAFT

Introduction

Long non-coding RNAs (lncRNAs) are a diverse class of RNA transcripts that lack protein-coding potential but exert critical regulatory functions in eukaryotic genomes. Once regarded as transcriptional noise, lncRNAs are now recognized as important players in gene regulation, with distinct features of their biogenesis, expression, and evolutionary patterns that set them apart from messenger RNAs (mRNAs) [Quinn & Chang, 2016; Ponting et al., 2009]. They are typically characterized by low expression, limited sequence conservation, and developmental or tissue-specific transcription [Ulitsky & Bartel, 2013]. Classification systems based on genomic context, molecular mechanism, and regulatory target further underscore their functional heterogeneity [Kung et al., 2013]. Notably, genomic proximity often dictates lncRNA activity, as cis-acting interactions with neighboring genes can be as influential as broader trans-regulatory roles [Barry, 2014].

In marine invertebrates, particularly molluscs, lncRNAs are emerging as key regulators of physiological and developmental processes. Studies in scallops (*Patinopecten yessoensis*), mussels (*Mytilus galloprovincialis*), clams (*Ruditapes philippinarum*), and pearl oysters (*Pinctada fucata martensii*) have linked lncRNAs to immune defense, pigmentation, and biomineralization [Wang et al., 2020; Venier et al., 2021; Nie et al., 2023; Zhang et al., 2022]. For example, co-expression networks in scallops revealed lncRNAs involved in melanization and biomineralization during *Polydora* infection [Chen et al., 2023], while mussel hemocytes showed distinct lncRNA signatures in response to immune stimulation [Gerdol et al., 2021]. In asymmetric-shelled bivalves, mantle-expressed lncRNAs regulate calcium carbonate deposition and crystal orientation [Wang et al., 2023], placing them within the biomineralization regulatory toolkit.

Most progress in oyster lncRNA research has focused on the Pacific oyster, *Crassostrea gigas*, a global aquaculture species and molluscan genomic model. Transcriptomic surveys have identified thousands of lncRNAs with tissue- and stage-specific expression during larval development, reproduction, and shell pigmentation [Zhang et al., 2016; Yu et al., 2018]. Functional studies have further implicated *C. gigas* lncRNAs in glycogen metabolism [Wang et al., 2020], antiviral responses to Ostreid herpesvirus [Rosani et al., 2019], and mantle-driven shell formation [Yu et al., 2018]. These findings underscore the importance of lncRNAs as regulators of physiology and stress responses in oysters.

In contrast, the lncRNA landscape of the Eastern oyster, *Crassostrea virginica*, remains poorly characterized despite its ecological and economic significance along the Atlantic and Gulf coasts of North America. *C. virginica* faces mounting threats from ocean acidification (OA), which disrupts calcification, energy balance, and immune function in calcifying invertebrates. However, it is not yet known whether lncRNA expression in this species is sensitive to OA, nor which molecular pathways lncRNAs may influence under acidified conditions.

Here, we characterize the expression dynamics of lncRNAs in *C. virginica* under experimental OA exposure. We assess whether lncRNA transcription is responsive to reduced pH and use co-expression analyses with mRNAs to identify the pathways and processes in which lncRNAs may be involved. This approach provides insight into the potential regulatory roles of lncRNAs in environmental stress resilience, contributing to a broader understanding of how oysters respond and adapt to ocean change.

Methods

Experimental Design and Data Acquisition

Male and female eastern oysters (*Crassostrea virginica*) were exposed to elevated pCO_2 conditions (2500 µatm, Ω_calcite < 1) to assess transcriptomic responses relative to control oysters maintained at ambient conditions (500 µatm, Ω_calcite > 1). A total of 26 individuals were sequenced. RNA-seq data were subjected to quality control and adapter trimming prior to downstream analysis. The *C. virginica* reference genome was obtained from NCBI, and long non-coding RNA (lncRNA) annotations were downloaded from the same source to identify previously described lncRNAs. A custom GFF and FASTA were generated to enable downstream analyses.

Size Distribution and Genomic Context of lncRNAs

LncRNA sequence lengths were extracted from the FASTA file and binned at 500-nt intervals to characterize size distribution. Genomic coordinates of lncRNAs were compared with *C. virginica* gene models using BedTools intersect, enabling classification of lncRNAs relative to annotated exons, genes, and mRNAs.

Comparative analyses were performed against lncRNA datasets from other bivalves, including *Crassostrea gigas*, *Patinopecten yessoensis*, *Mytilus galloprovincialis*, and *Ruditapes philippinarum*. For each species, lncRNA FASTA files were filtered to retain only sequences annotated as ncRNA under NCBI classification. Two *C. gigas* assemblies were used to generate independent FASTA datasets to capture both a widely used reference and a more recent genome build. Cross-species homology searches were conducted using BLAST against the *C. virginica* lncRNA database. Results were consolidated into a single summary table, with special emphasis placed on lncRNAs implicated in pigmentation, biomineralization, and glycogen metabolism.

Differential Expression Analysis

Transcript abundances for both lncRNAs and coding sequences were estimated using kallisto, which generated indices from custom FASTA files. Trimmed reads were quantified against these indices, and abundance estimates were imported into Trinity to construct count matrices. Matrices were subset by sex, yielding male- and female-specific datasets. Differential expression analyses were conducted using DESeq2. The primary model incorporated both sex

and treatment as factors, while sex-specific analyses were also performed to assess treatment effects independent of sex.

mRNA–lncRNA Co-expression Analysis

Co-expression between mRNAs and lncRNAs was assessed using pairwise Pearson correlation coefficients (r) calculated across all samples. Correlation analyses were performed in R v4.3.2 using the `cor()` and `cor.test()` functions. To account for multiple testing, p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) procedure, and only highly significant correlations ($|r| \geq 0.8$, FDR < 0.05) were retained. The resulting correlation matrix was filtered to identify both positively and negatively associated mRNA–lncRNA pairs potentially involved in shared transcriptional regulation.

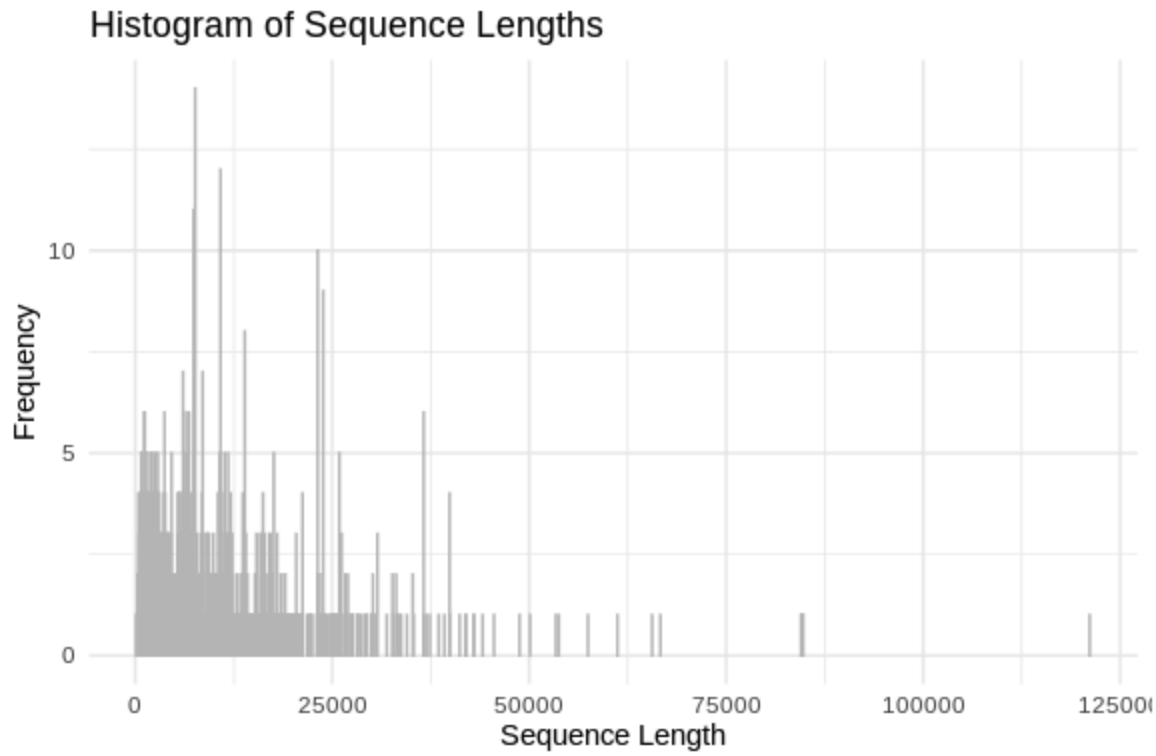
Network Visualization and Functional Annotation

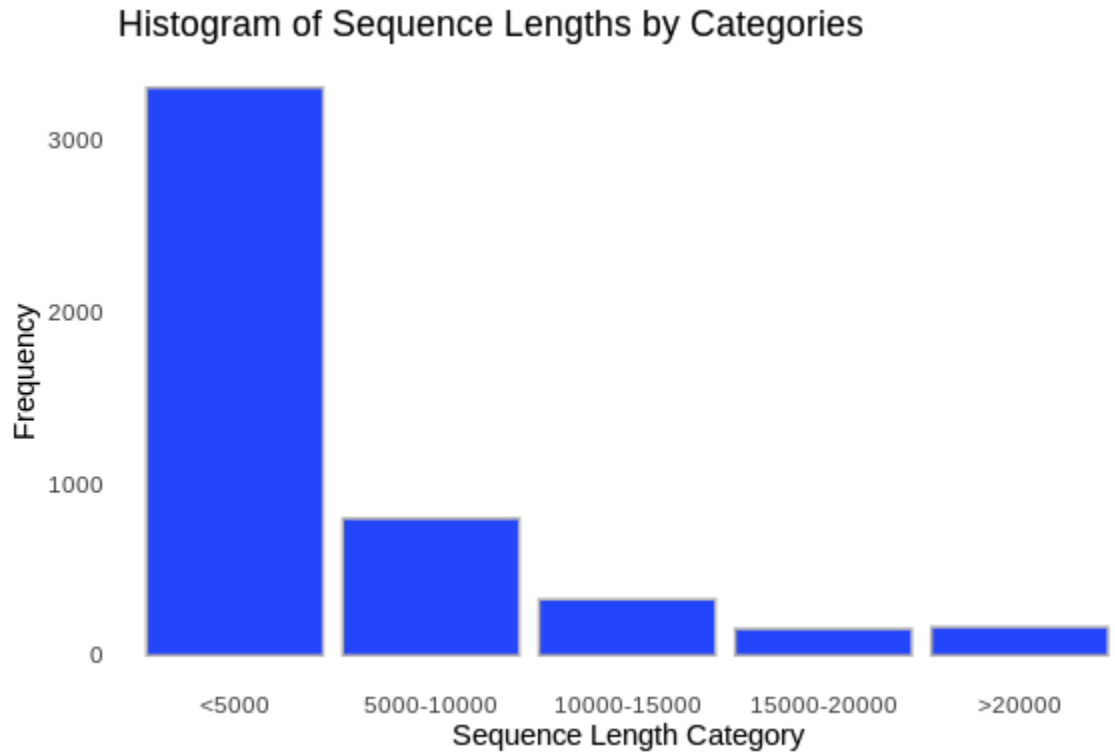
Significant mRNA–lncRNA co-expression relationships were visualized as undirected networks using Cytoscape v3.10.0, where nodes represented transcripts and edges indicated significant correlations. Network modules, or clusters of highly co-expressed transcripts, were identified using the MCODE plugin in Cytoscape. To infer potential biological functions of lncRNAs, Gene Ontology (GO) enrichment analyses were performed on the mRNA members of each module using the clusterProfiler v4.10.1 package in R. Enriched GO terms (FDR < 0.05) were interpreted as indicative of biological processes potentially influenced by co-expressed lncRNAs.

Results

Size and Location

The majority of lncRNAs in the NCBI FASTA are over 1000nt and less than <5000nt.


Histogram of Sequence Lengths
Frequency
10
5
0
0
25000
50000
75000
100000
125000
Sequence Length


Histogram of Sequence Lengths by Categories
Frequency
3000
2000
1000
0
<5000
5000-10000
10000-15000
15000-20000
>20000
Sequence Length Category

Basics - Species Comparison

Species	ncRNA count
Crassostrea gigas	7247
Patinopecten yessoensis	2203
Mytilus trossulus	6173
Ruditapes phillipinarium	5853
*Magallana gigas**	7177

*Updated 2023 RNA transcripts from NCBI

After BLAST:

Species	Hits
Crassostrea gigas	11992
Patinopecten yessoensis	2207
Mytilus trossulus	6269
Ruditapes phillipinarium	6957
Magallana gigas	13895

Differential Expression

Differential expression accounting for treatment and sex shows many differences in expression of lncRNAs based on sex. There are 2370 differentially expressed lncRNAs between female and male sample groups. Sex specific DESeq2 runs identified far fewer differentially expressed lncRNAs, with five for males and one for females. After separation of DGE analysis based on sex, a small number of differentially expressed lncRNAs were found (1 for females and 5 for males). Differential expression of mRNAs followed a similar pattern after sex separation (107 for males and 25 for females).

lncRNAs

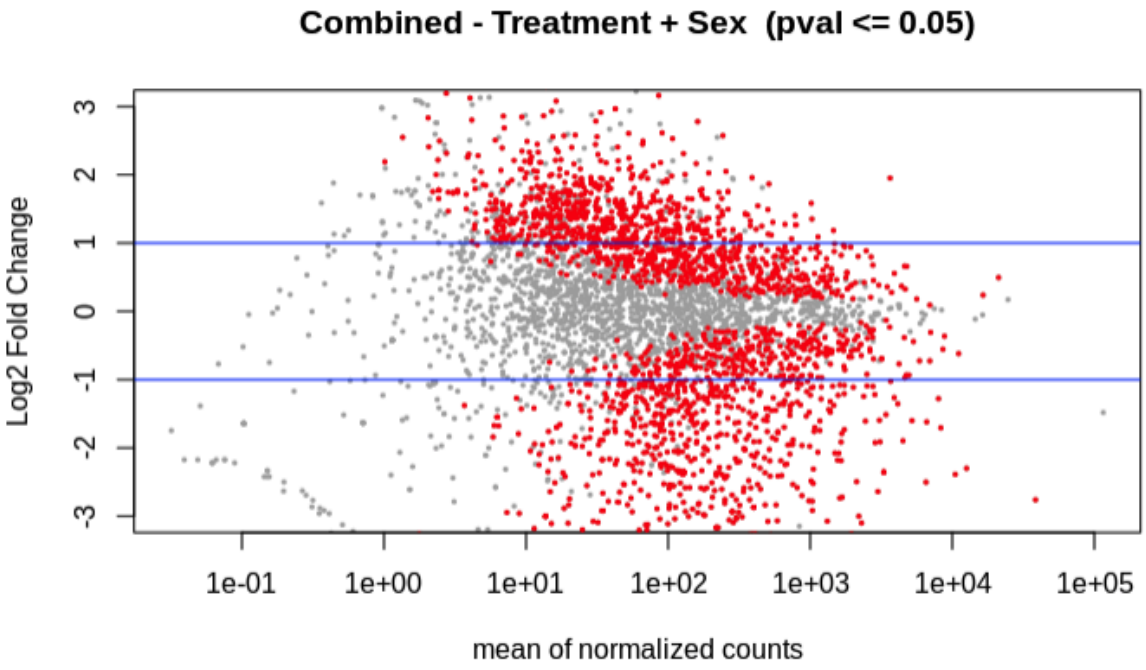


Fig. Volcano plot displaying 2370 differentially expressed lncRNAs from combined sex and treatment DESeq2.

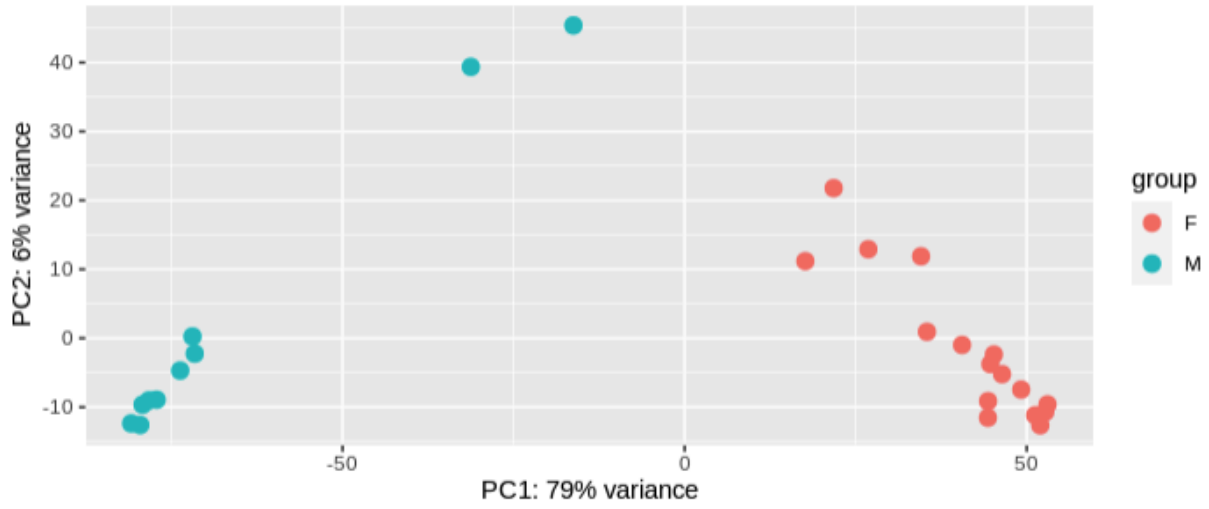


Fig. PCA from combined sex and treatment differential expression displaying sex. Data clusters by sex.

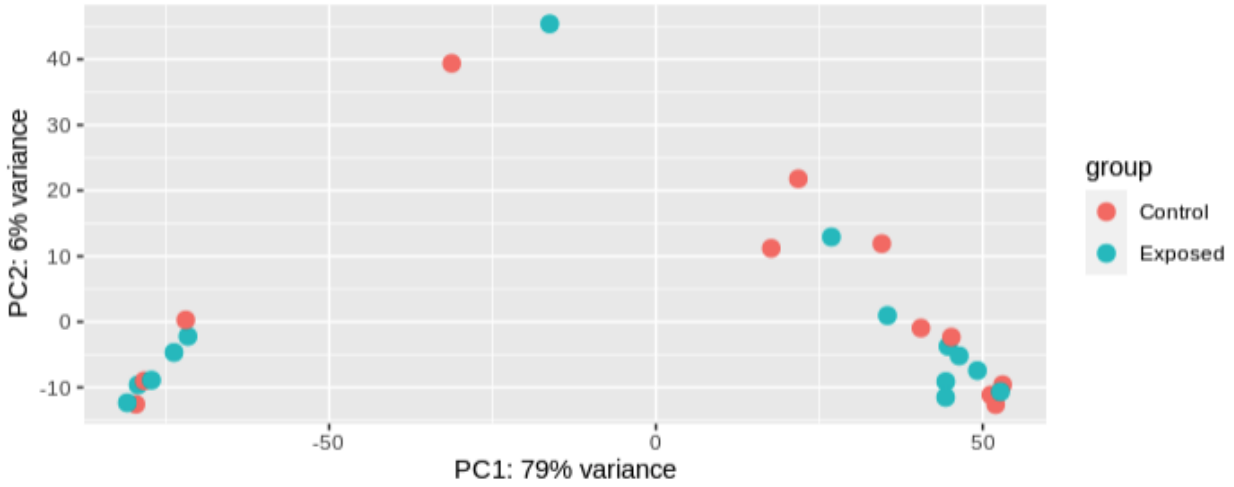
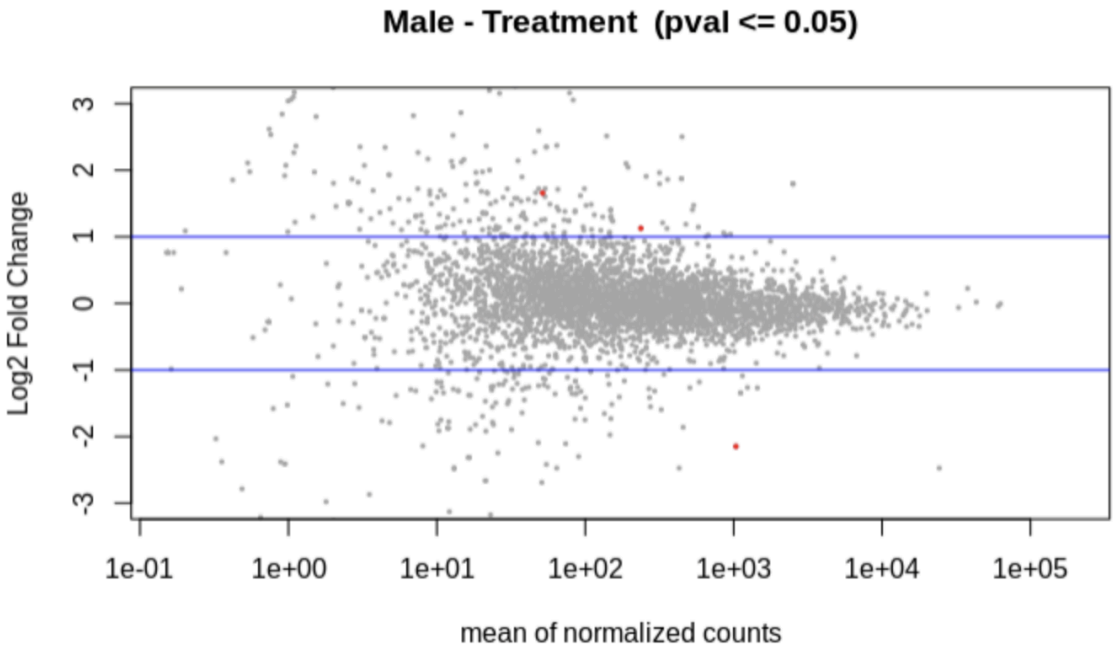


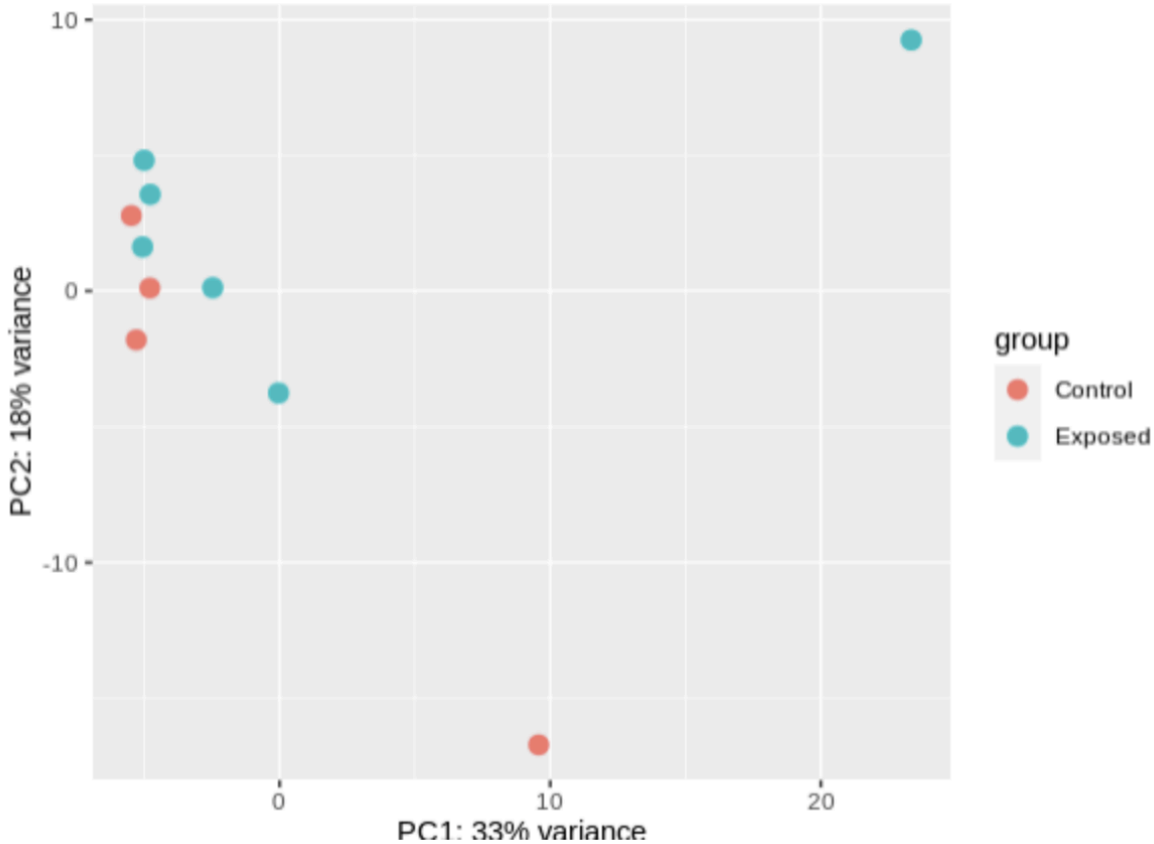
Fig. PCA from combined sex and treatment differential expression displaying treatment. No clustering based on treatment.

Males

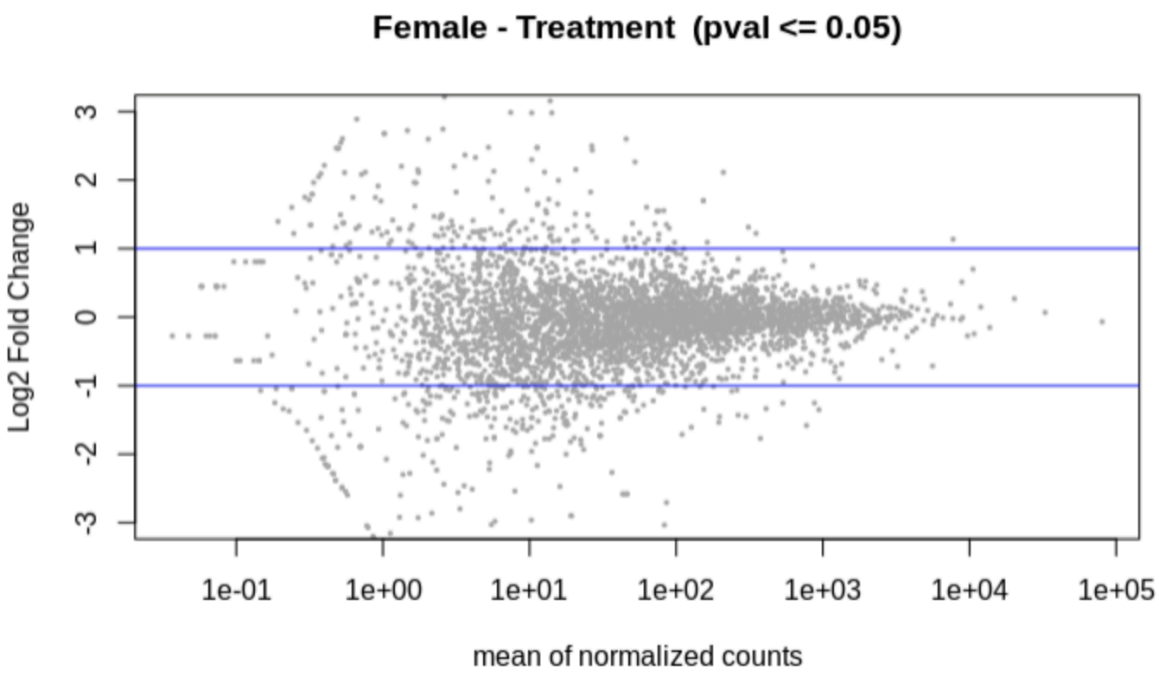


- 5 significantly differentially expressed lncRNAs

PCA - treatment

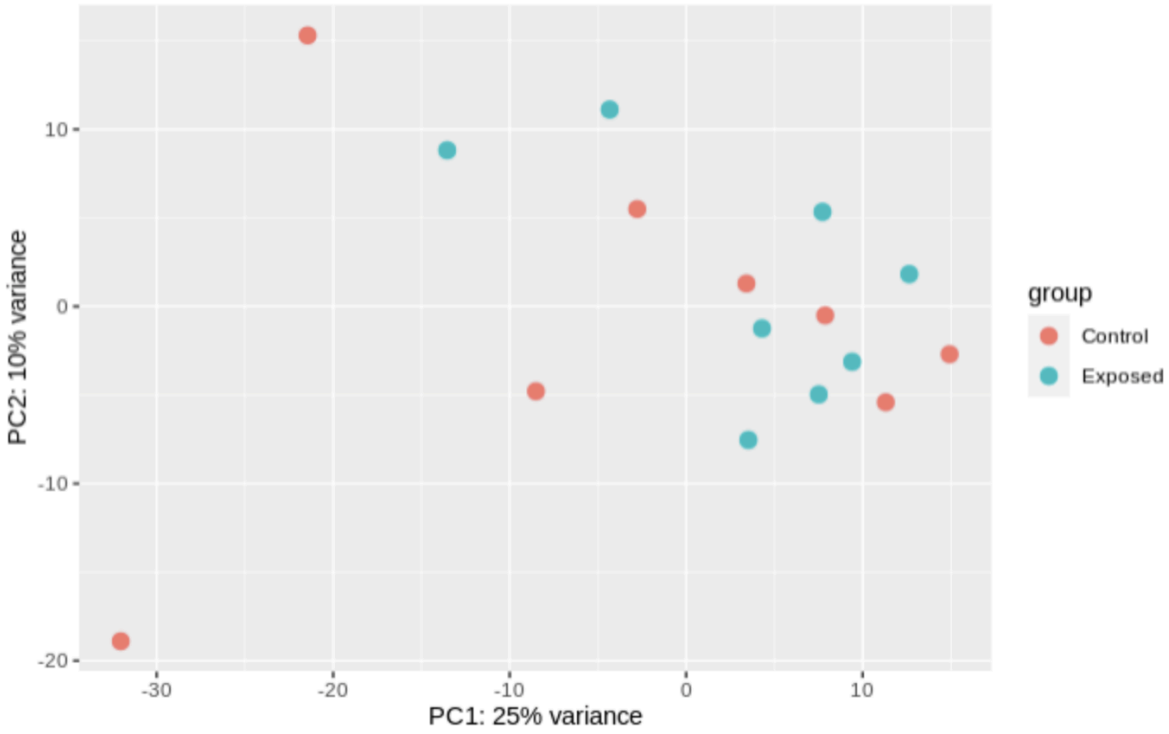


Female



- 1 significantly differentially expressed lncRNA

PCA - treatment



Differential expression of mRNAs revealed a similar pattern in sex differences. 36242 significantly differentially expressed coding sequences resulted from the combined sex and treatment DESeq2 run of mRNAs. In sex specific runs, DESeq2 identified 25 DEGs in females and 107 DEGs in males.

mRNAs

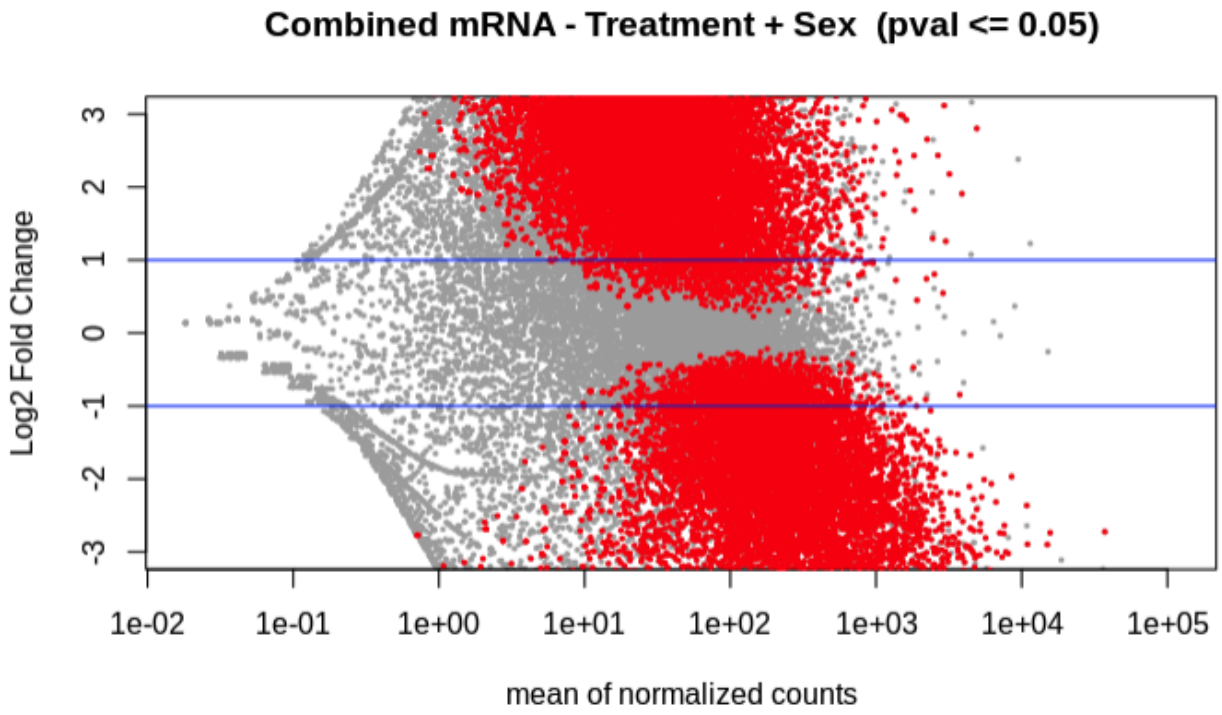
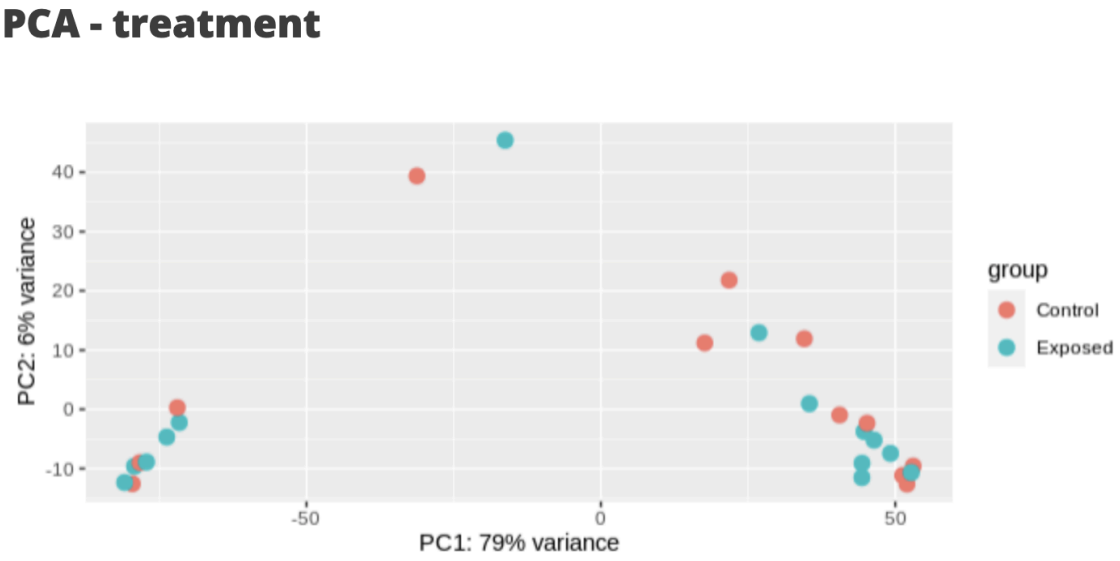


Fig. Volcano plot displaying 36242 DEGs from combined DESeq2 with sex and treatment.



PCA - sex

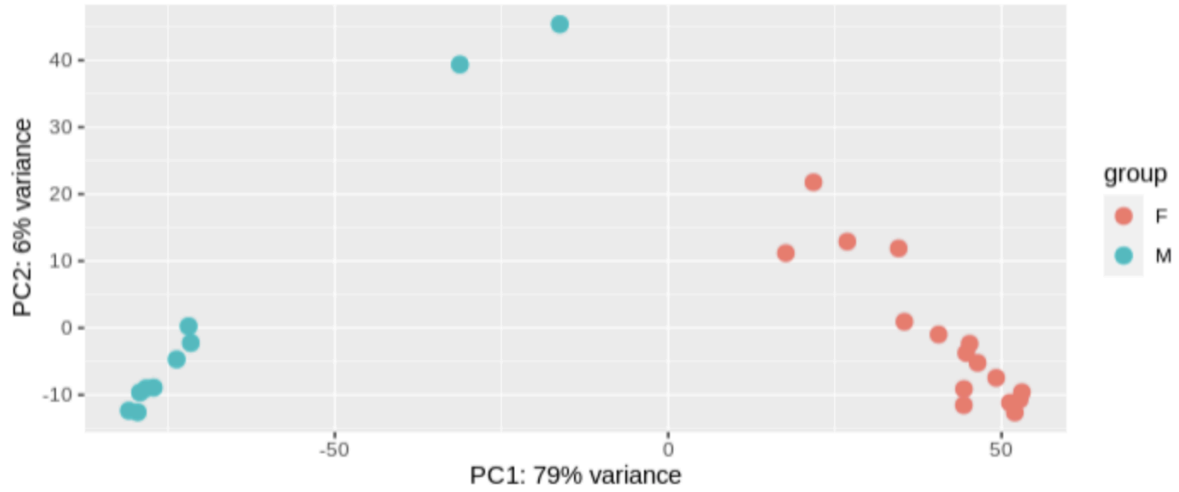
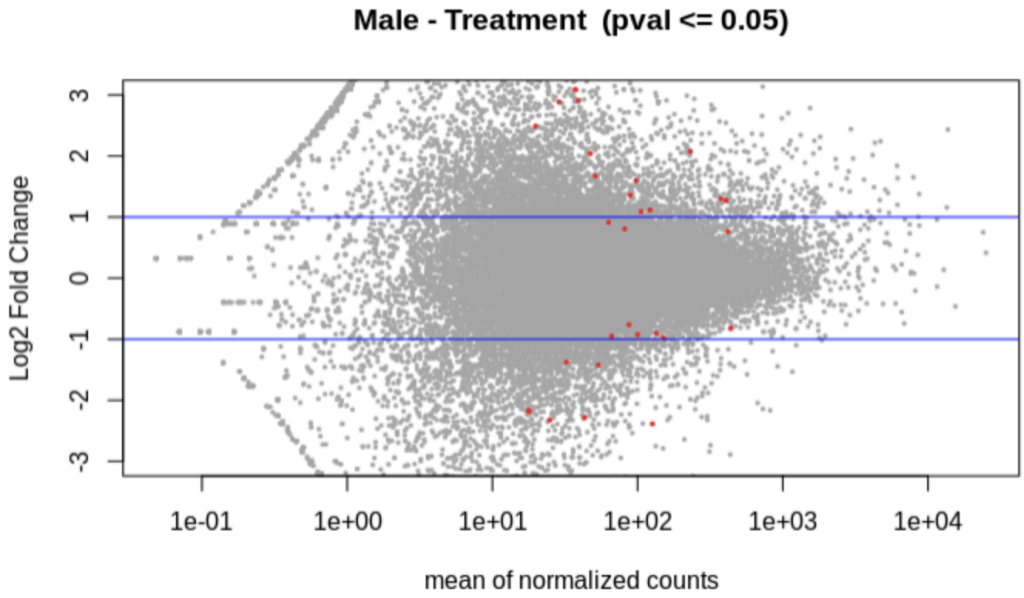


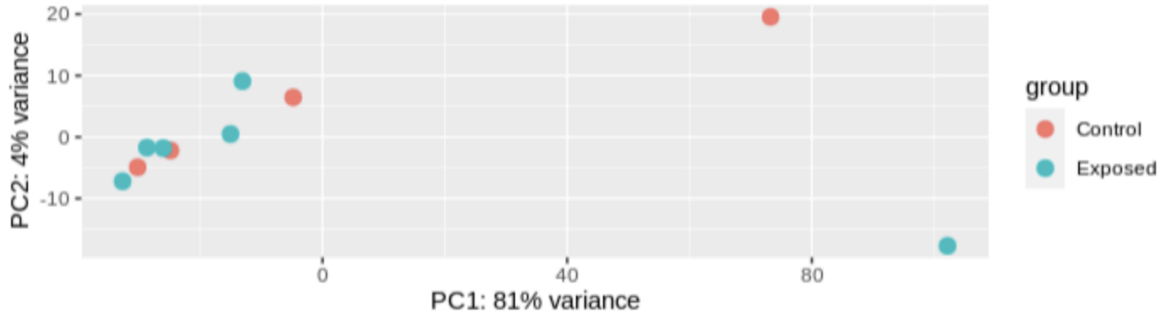
Fig. PCA showing clustering by sex and treatment from differential expression analysis.

Males

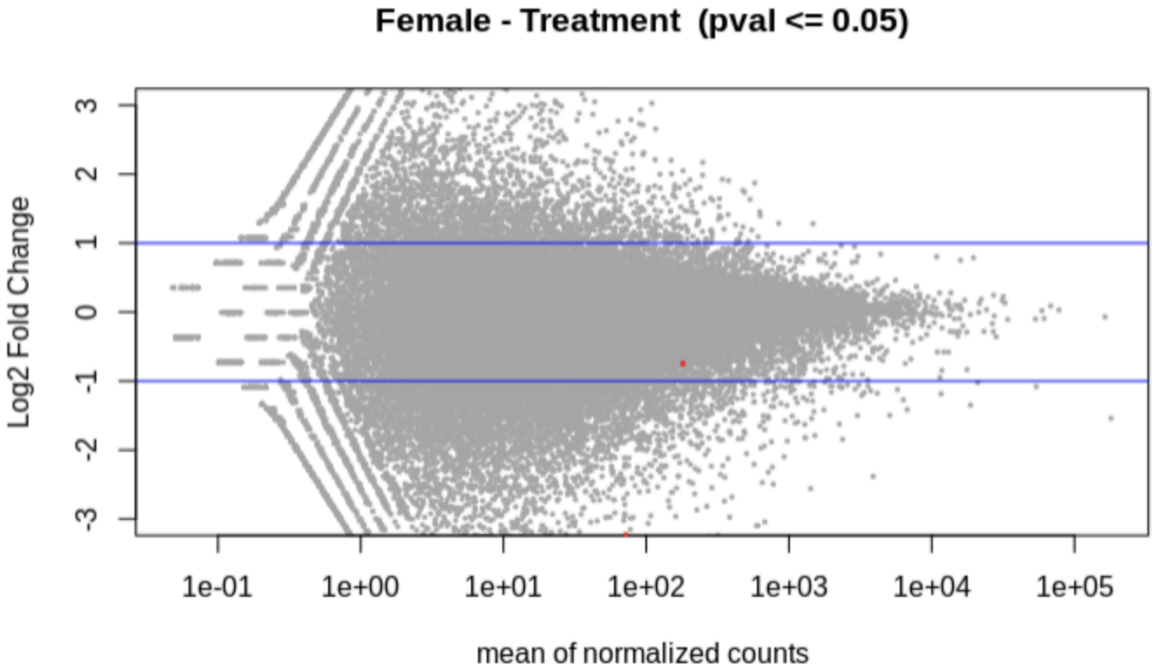


- 107 significantly differentially expressed mRNAs

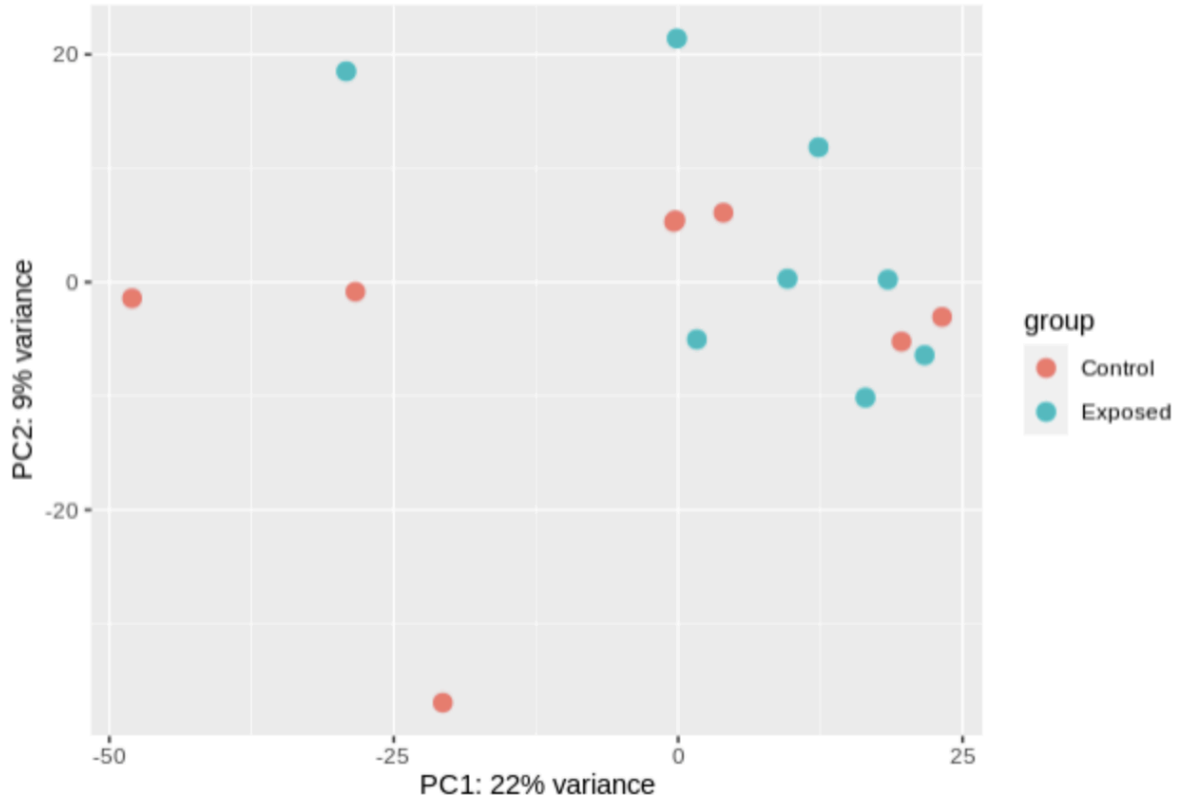
PCA - treatment



Females



- 25 significantly differentially expressed mRNAs



Size and Genomic Distribution of lncRNAs

Across the *Crassostrea virginica* reference genome, most annotated long non-coding RNAs (lncRNAs) were between 1,000 and 5,000 nucleotides in length, consistent with the expected size range for metazoan lncRNAs. These transcripts were broadly distributed throughout the genome, often located in intergenic regions or antisense to protein-coding genes, suggesting diverse potential regulatory roles.

Comparative Abundance of lncRNAs Across Bivalve Species

To place C. virginica lncRNAs in a broader evolutionary context, homologous non-coding transcripts were identified in several related bivalve species using BLAST searches against publicly available transcriptomes. Before BLAST filtering, lncRNA counts ranged from 2,203 in Patinopecten yessoensis to 7,247 in Crassostrea gigas. After BLAST, the number of matching transcripts increased across species, with M. gigas (13,895 hits) and C. gigas (11,992 hits) showing the highest similarity to C. virginica lncRNAs. Moderate homology was observed in Ruditapes philippinarum (6,957 hits) and Mytilus trossulus (6,269 hits), indicating partial conservation of lncRNA sequences among bivalves.

Differential Expression Patterns

Differential expression analyses incorporating both sex and treatment factors revealed extensive sex-specific transcriptional differences. Combined DESeq2 analyses identified 2,370 differentially expressed lncRNAs between male and female oysters (Fig. 1). Principal component analysis (PCA) confirmed strong clustering by sex but not by treatment (Fig. 2–3), suggesting that sex explains most of the variance in lncRNA expression.

When differential expression was reanalyzed separately by sex, the number of significantly expressed lncRNAs declined sharply, with only one detected in females and five in males. This pattern indicates that the primary lncRNA expression differences are driven by sex rather than environmental treatment effects.

Similarly, coding mRNAs displayed extensive sex-linked expression divergence. Combined DESeq2 analyses identified 36,242 differentially expressed mRNAs, whereas sex-specific analyses detected 25 DEGs in females and 107 DEGs in males (Fig. 4–6). These results mirror the lncRNA findings, emphasizing strong transcriptional dimorphism between sexes and relatively limited treatment effects.

Co-expression of mRNAs and lncRNAs

To explore potential regulatory interactions between coding and non-coding transcripts, we constructed a co-expression network based on pairwise Pearson correlations between normalized mRNA and lncRNA expression values. From all possible transcript pairs, those with strong correlations ($|r| \geq 0.8$, FDR < 0.05) were retained, resulting in a network comprising *[insert number]* lncRNAs, *[insert number]* mRNAs, and *[insert number]* significant edges. The distribution of correlation coefficients was bimodal, with the majority of relationships showing positive associations, suggesting coordinated transcriptional activity between lncRNAs and their putative target genes.

Network structure and module identification

Visualization of the co-expression network in Cytoscape revealed a modular topology with several clusters of highly interconnected transcripts (Fig. X). Application of the MCODE algorithm identified *[insert number]* discrete co-expression modules, each representing sets of mRNAs and lncRNAs exhibiting tightly correlated expression patterns across samples. Many modules were dominated by either sex-biased or treatment-associated transcripts, consistent with the strong sex effects observed in the differential expression analyses.

Functional enrichment of mRNA modules

To infer potential biological functions of co-expressed lncRNAs, Gene Ontology (GO) enrichment analysis was conducted on the mRNA members of each module using clusterProfiler. Significant enrichment (FDR < 0.05) was detected in processes related to gametogenesis, ion transport, signal transduction, and stress response, among others (Table X). Modules enriched for developmental and reproductive pathways were primarily associated

with sex-biased expression patterns, whereas smaller modules containing treatment-responsive transcripts showed enrichment for cellular stress and metabolic processes.