Comprehensive analysis of IncRNAs and mRNAs in skeletal muscle of rainbow trout (Oncorhynchus mykiss) exposed to estradiol

Jian Wang1,3, Prasanthi P. Koganti1, Jianbo Yao1, Shuo Wei1 & Beth Cleveland2

Estradiol (E2) is a steroid hormone that negatively affects muscle growth in rainbow trout (Oncorhynchus mykiss), but the mechanisms directing with this response are not fully understood. To better characterize the effects of E2 in muscle, we identified differentially regulated mRNAs and IncRNAs in juvenile rainbow trout exposed to E2. Here, we performed next-generation RNA sequencing and comprehensive bioinformatics analyses to characterize the transcriptome profiles, including mRNAs and long noncoding RNAs (IncRNAs), in skeletal muscle of rainbow trout injected with E2. A total of 226 IncRNAs and 253 mRNAs were identified as differentially regulated. We identified crucial pathways, including several signal transduction pathways, hormone response, oxidative response and protein, carbon and fatty acid metabolism pathways. Subsequently, a functional IncRNA-mRNA co-expression network was constructed, which consisted of 681 co-expression relationships between 164 IncRNAs and 201 mRNAs. Moreover, a IncRNA-pathway network was constructed. A total of 65 key IncRNAs were identified that regulate 20 significantly enriched pathways. Overall, our analysis provides insights into mRNA and IncRNA networks in rainbow trout skeletal muscle and their regulation by E2 while understanding the molecular mechanism of IncRNAs.

The recent advent of next generation RNA sequencing (RNA-Seq) and publication of reference genomes for many organisms have allowed researchers to study transcription profiles. The majority of the mammalian genome (up to 80%) is transcribed, of which only 2–3% are protein-coding RNAs (mRNAs) and the rest are noncoding RNAs. Most transcribed ncRNAs are larger than 200 nucleotides and defined as long non-coding RNAs (IncRNA) with some further processed to generate small RNAs.1 IncRNA have fewer, longer exons when compared to coding genes and exhibit cell-type specific expression. These long non-coding RNAs are reported to play critical roles in various biological processes, including chromatin modification, regulation of transcription, influence of nuclear architecture and regulation of gene expression at post-transcriptional and post-translational levels, and are also reported to interact with DNA, RNA and proteins during these processes.2 Availability of sequencing techniques led to the discovery of non-coding transcripts involved in regulation of gene expression, as well as novel protein coding genes.

Studies in humans (Homo sapiens) and mice (Mus musculus) support that IncRNA are involved in transcriptional regulation of skeletal muscle development and differentiation.3 For instance, ChiP-Seq studies identified two different IncRNAs (5032RNA and 5355RNA) that transcribe from the enhancer region of myogenic transcription factor MyoD2,3 to increase expression of MyoD and myogenin. An additional study identified a number of muscle specific IncRNAs regulated by Yin Yang 1 protein that function as both positive or negative regulators of muscle differentiation.4 Expanding the investigation of IncRNAs as myogenic regulators in additional species is critical to both 1) establish whether previously reported mechanisms are conserved and 2) identify novel mechanisms within a species.

1Division of Animal and Nutritional Sciences, West Virginia University, Morgantown, WV, 26506, USA. 2USDA/ARS National Center for Cool and Cold Water Aquaculture, Kearneysville, WV, 25430, USA. 3Department of Computational Biology, St. Jude Children’s Research Hospital, Memphis, TN, 38105, USA. 4Department of Biological Sciences, University of Delaware, Newark, DE, 19716, USA. Jian Wang and Prasanthi P. Koganti contributed equally to this work. Correspondence and requests for materials should be addressed to B.C. (email: beth.cleveland@ars.usda.gov)
Rainbow trout (*Oncorhynchus mykiss*) is a fish species valued as a research model, food source, and recreational fish. This species demonstrates a pattern of indeterminate growth, or the ability to grow in length and continually accumulate muscle throughout its lifespan. Muscle regulatory factors (MRF) such as MyoD and myogenin that regulate hyperplasia and hypertrophy in mammalian muscle are also important for these processes in salmonids. However, unique MRF expression patterns occur during differentiation of rainbow trout myogenic precursor cells (MPCs), suggesting that these mechanisms are significant for continued hyperplastic capacity and indeterminate growth.

Determining how lncRNAs are regulated by biological factors that affect muscle growth is central for identifying lncRNA-related mechanisms important for myogenesis. In salmonids, estradiol (E2) is a biological factor that negatively effects muscle growth, partially through direct effects on MPC proliferation and protein turnover. Plasma E2 levels increase during sexual maturation and its catabolic effects in muscle are important for directing nutrients away from muscle growth and toward gonad development, an energy intensive process that compromises muscle quality. Our previous findings support a role for noncoding RNAs in the short-term E2-response, specifically through regulation of miRNAs that directly affect expression of genes related to proteolysis, cell differentiation, and cell proliferation. The present study aims to identify the differentially expressed lncRNAs and mRNAs in skeletal muscle exposed to E2 and create a network of the mRNAs and lncRNAs involved in the process. Here we use an RNA-Seq approach to comprehensively investigate temporal gene expression differences in skeletal muscle of rainbow trout after a single injection of E2. We also construct a functional lncRNA-mRNA regulatory network that is associated with skeletal muscle response to E2 treatment, and conduct a pathway enrichment analysis. This comprehensive analysis helps identify novel lncRNA and their related mechanisms with emphasis on regulation of muscle growth in rainbow trout under the influence of E2.

### Results

**Generation of a muscle transcriptome reference.** To provide a comprehensive understanding of the effects of E2 treatment on muscle physiology, the lncRNA and mRNA transcriptomes were analyzed in skeletal muscle from E2-treated and control rainbow trout 24 and 72 hours post-injection. A total of 789,485,036 paired-end raw reads were generated from 16 samples with 101-bp read length. After quality control including removal of ambiguous nucleotides, low-quality bases and ribosomal RNA sequences, a total of 749,490,934 cleaned reads (94.9%) were harvested for further analysis. The number of cleaned reads of each sample ranged from 34.5 to 60.9 million (Table 1).

| Group | Time (h) | Replicate | Reads | Clean reads | Mapped reads | Mapping ratio (%) |
|-------|---------|-----------|-------|-------------|--------------|------------------|
| CTRL  | 24      | 1         | 36,802,826 | 35,432,594  | 31,145,250  | 87.9             |
|       |         | 2         | 61,291,464 | 52,228,282  | 46,274,258  | 88.6             |
|       |         | 3         | 59,111,132 | 56,868,138  | 50,328,302  | 88.5             |
|       |         | 4         | 50,601,976 | 48,340,122  | 42,974,368  | 88.9             |
| CTRL  | 72      | 1         | 63,019,266 | 60,702,946  | 54,389,838  | 89.6             |
|       |         | 2         | 60,452,644 | 58,114,660  | 52,186,986  | 89.8             |
|       |         | 3         | 51,642,380 | 49,378,116  | 44,045,278  | 89.2             |
|       |         | 4         | 63,821,958 | 60,981,886  | 53,786,022  | 88.2             |
| E2    | 24      | 1         | 61,743,118 | 59,403,804  | 52,394,128  | 88.2             |
|       |         | 2         | 37,513,122 | 35,786,668  | 30,186,188  | 84.3             |
|       |         | 3         | 42,203,620 | 40,341,512  | 34,895,386  | 86.5             |
|       |         | 4         | 47,655,922 | 45,563,586  | 40,232,838  | 88.3             |
| E2    | 72      | 1         | 36,221,406 | 34,809,018  | 30,144,800  | 86.6             |
|       |         | 2         | 41,421,664 | 39,634,688  | 34,878,528  | 88.0             |
|       |         | 3         | 38,926,574 | 37,367,184  | 31,874,208  | 85.3             |
|       |         | 4         | 37,055,964 | 34,537,730  | 29,586,296  | 85.6             |
| Total |         |           | 789,485,036 | 749,490,934 | 659,322,674 |                  |

Table 1. Summary of samples and RNA-Seq data. CTRL indicates control group. E2 represents estradiol treatment group.

Rainbow trout (*Oncorhynchus mykiss*) is a fish species valued as a research model, food source, and recreational fish. This species demonstrates a pattern of indeterminate growth, or the ability to grow in length and continually accumulate muscle throughout its lifespan. Muscle regulatory factors (MRF) such as MyoD and myogenin that regulate hyperplasia and hypertrophy in mammalian muscle are also important for these processes in salmonids. However, unique MRF expression patterns occur during differentiation of rainbow trout myogenic precursor cells (MPCs), suggesting that these mechanisms are significant for continued hyperplastic capacity and indeterminate growth.

Determining how lncRNAs are regulated by biological factors that affect muscle growth is central for identifying lncRNA-related mechanisms important for myogenesis. In salmonids, estradiol (E2) is a biological factor that negatively effects muscle growth, partially through direct effects on MPC proliferation and protein turnover. Plasma E2 levels increase during sexual maturation and its catabolic effects in muscle are important for directing nutrients away from muscle growth and toward gonad development, an energy intensive process that compromises muscle quality. Our previous findings support a role for noncoding RNAs in the short-term E2-response (24 hr post-injection), specifically through regulation of miRNAs that directly affect expression of genes related to proteolysis, cell differentiation, and cell proliferation. The present study aims to identify the differentially expressed lncRNAs and mRNAs in skeletal muscle exposed to E2 and create a network of the mRNAs and lncRNAs involved in the process. Here we use an RNA-Seq approach to comprehensively investigate temporal gene expression differences in skeletal muscle of rainbow trout after a single injection of E2. We also construct a functional lncRNA-mRNA regulatory network that is associated with skeletal muscle response to E2 treatment, and conduct a pathway enrichment analysis. This comprehensive analysis helps identify novel lncRNA and their related mechanisms with emphasis on regulation of muscle growth in rainbow trout under the influence of E2.

**Identification of differentially expressed genes.** The pipeline reported in our previous study was used to identify all lncRNAs from the expressed transcriptome reference. Two biological replicates (CTRL &
than 0.99 and FDR less than 0.05 were selected for network construction. The co-expression network consisted normalized expression values of each of the lncRNA-mRNA pairs. The lncRNA-mRNA pairs with a PCC greater obtained by using DESeq222. Subsequently, Pearson's correlation coefficient (PCC) was calculated between the lncRNA-mRNA co-expression network, the normalized expression values of the lncRNAs and mRNAs were.

Table 2. Statistics of transcriptome assembly.

| Assembly                  | Number of components | 203,148 |
|---------------------------|----------------------|---------|
|                           | Number of contigs    | 243,509 |
|                           | Maximum contig length| 20,635 bp|
|                           | Minimum contig length| 201 bp  |
|                           | Average contig length| 660 bp  |
|                           | Median contig length | 365 bp  |
|                           | N50 length           | 1,076 bp|

| Filtered Assembly         | Number of components | 31,419 |
|---------------------------|----------------------|--------|
|                           | Number of contigs    | 63,181 |
|                           | Maximum contig length| 20,635 bp|
|                           | Minimum contig length| 201 bp  |
|                           | Average contig length| 1,466 bp|
|                           | Median contig length | 1,189 bp|
|                           | N50 length           | 1,982 bp|

To construct the differentially expressed lncRNA-mRNA co-expression network, the normalized expression values of the lncRNAs and mRNAs were obtained by using DESeq222. Subsequently, Pearson’s correlation coefficient (PCC) was calculated between the normalized expression values of each of the lncRNA-mRNA pairs. The lncRNA-mRNA pairs with a PCC greater than 0.99 and FDR less than 0.05 were selected for network construction. The co-expression network consisted.

Experimental validation of lncRNA and mRNA. To confirm the sequencing results, seven lncRNAs and five mRNAs that were differentially expressed were randomly selected for validation using real time RT-PCR. Each transcript was confirmed as differentially expressed in E2-treated fish in the same direction as indicated by the RNAseq analysis (Fig. 2A,B).

Differentially expressed genes represent several important pathways. To obtain the functions of the differentially expressed mRNAs and connections among them, we performed GO term and KEGG pathway enrichment analyses. To comprehensively interpret the GO term enrichment result, treemap was constructed based on the result of semantic similarity analysis. In each treemap, related terms were joined into a ‘supercluster’ with the same color and the most significant term as the representative of the group. As shown in Fig. 3a,b, in response to E2, skeletal muscle tissue showed notable up-regulations related to oxidative stress response, hormone response, protein ubiquitination, cysteine biosynthesis and DNA repair, and down-regulations mostly related to mesenchyme morphogenesis and cAMP biosynthesis in biological process. Notably, the term of response to leptin (GO:0044321) was identified to be enriched (Supplementary Fig. S1). Up-regulated genes in molecular function relate to kinase activity, hormone receptor activity, amino-acyl transferase activity and ubiquitin conjugating enzyme activity. Down regulated genes related to molecular function were related to muscle structural protein binding, motor activity, pyruvate carboxylase activity and glucose-6-phosphate activity (Supplementary Fig. S2a,b). The enrichment of up-regulated genes in cellular component largely related to chromosome and ubiquitin conjugating enzyme complex, and down-regulated genes related to skeletal muscle structural protein complex (Supplementary Fig. S3a,b).

The KEGG pathway analyses mapped the DEGs to KEGG reference pathways to infer systemic biological behaviors. Comparing to the control group, we observed significant KEGG pathway enrichment of DEGs in skeletal muscle in response to E2 stress. Of the top 20 over-represented KEGG pathways (Fig. 4), four were involved in signal transduction, including Rap1 signaling pathway, Jak-STAT signaling pathway, calcium signaling pathway and AMPK signaling pathway. An additional four pathways were associated with the endocrine system, which are PPAR signaling pathway, insulin resistance, estrogen signaling pathway and adipocytokine signaling pathway. Finally, the remaining 12 over-represented KEGG pathways were classified into different functional groups, including amino acid and carbohydrate metabolism, cellular community, protein synthesis and digestion.

Construction of lncRNA-mRNA co-expression network. To construct the differentially expressed lncRNA-mRNA co-expression network, the normalized expression values of the lncRNAs and mRNAs were obtained by using DESeq222. The KEGG pathway analyses mapped the DEGs to KEGG reference pathways to infer systemic biological behaviors. Comparing to the control group, we observed significant KEGG pathway enrichment of DEGs in skeletal muscle in response to E2 stress. Of the top 20 over-represented KEGG pathways (Fig. 4), four were involved in signal transduction, including Rap1 signaling pathway, Jak-STAT signaling pathway, calcium signaling pathway and AMPK signaling pathway. An additional four pathways were associated with the endocrine system, which are PPAR signaling pathway, insulin resistance, estrogen signaling pathway and adipocytokine signaling pathway. Finally, the remaining 12 over-represented KEGG pathways were classified into different functional groups, including amino acid and carbohydrate metabolism, cellular community, protein synthesis and digestion.

The KEGG pathway analyses mapped the DEGs to KEGG reference pathways to infer systemic biological behaviors. Comparing to the control group, we observed significant KEGG pathway enrichment of DEGs in skeletal muscle in response to E2 stress. Of the top 20 over-represented KEGG pathways (Fig. 4), four were involved in signal transduction, including Rap1 signaling pathway, Jak-STAT signaling pathway, calcium signaling pathway and AMPK signaling pathway. An additional four pathways were associated with the endocrine system, which are PPAR signaling pathway, insulin resistance, estrogen signaling pathway and adipocytokine signaling pathway. Finally, the remaining 12 over-represented KEGG pathways were classified into different functional groups, including amino acid and carbohydrate metabolism, cellular community, protein synthesis and digestion.
of 681 co-expression relationships between 164 lncRNAs and 201 mRNAs (Fig. 5). We then considered the node degree of the network, as a higher degree indicated that the nodes were likely to be hubs and therefore involved in more competing interactions. As a more stringent threshold, the top 5% (top 20) of the nodes were defined as hubs (Table 3). These 20 hub nodes that contain 14 lncRNAs co-expressed with more than 40% of the nodes in the network, implying the centrality of these nodes.

A lncRNA-pathway network was constructed to identify the possible functions of lncRNAs, in which nodes represent lncRNAs or over-represented pathways. Connections between nodes were made if mRNAs that co-express with a lncRNA associate with a common pathway, thus supporting that the pathway is regulated by the corresponding lncRNAs (Fig. 6). Sixty-five lncRNAs (Supplementary Table S3) were linked with 20 enriched pathways (Table 4, Supplementary Table S4) in the lncRNA-pathway network, suggesting these lncRNAs play a central role in regulating the E2 response.

Discussion

In rainbow trout, E2 is a maturation-related signal that negatively affects muscle growth. The present study identified lncRNAs and mRNAs regulated by E2 in rainbow trout skeletal muscle, thereby identifying mechanisms contributing to E2-induced muscle catabolism. Association analysis suggests that lncRNAs are central regulators of multiple components within a single pathway, supporting lncRNAs as important mechanisms for muscle physiology and the E2 response.

LncRNAs are emerging as regulators of diverse biological functions. We identified 226 lncRNAs that were differentially expressed, suggesting that lncRNAs are regulated by E2 and likely contribute to the catabolic effects of this steroid in skeletal muscle. Only three of the differentially expressed lncRNAs (Inc-OM206, 3272 & 7282) were previously reported in a rainbow trout lncRNA reference transcriptome pooled from 15 tissues, including muscle. This lack of overlap suggests that lncRNAs are temporospatially expressed, introducing the need for greater coverage of the rainbow trout lncRNA transcriptome across various tissues, periods of development, and treatment conditions. In particular, the sensitivity of lncRNAs (and mRNAs) to treatment conditions are supported by this study with the dramatic reduction in number of DEGs between 24 and 72 hr post-injection.

The functional roles of lncRNAs are poorly characterized. Co-expression models assist in large-scale prediction of lncRNA functions through integrating expression profiles of protein-coding genes and lncRNAs (coding-non-coding gene co-expression network). In this study a similar co-expression analysis was used to predict lncRNA functions based on their association with co-expressed genes, based on the fact that genes with similar expression patterns have a greater tendency to be involved in the same pathways. The resultant co-expression network contained a total of 164 differentially expressed lncRNAs and 201 differentially co-expressed mRNAs.
expressed mRNAs. This network provides a global view of possible lncRNA-coding gene expression associations that are influenced by E2 treatment in salmonids. In this network, the top 5% (20) large degree nodes were all lncRNAs, of which, 14 lncRNAs co-expressed with more than 40% of the nodes. These observations suggest that lncRNAs have functional significance in the E2-induced reduction of muscle growth in rainbow trout.

To examine the key lncRNAs and their potential functions, lncRNA-pathway network was constructed based on pathway enrichment analysis. The lncRNA-pathway network contains 65 lncRNAs linked with 20 significantly enriched pathways. The top pathway containing the most lncRNAs linked was focal adhesion (FA). Focal adhesions are integrin-containing, multi-protein structures that mediate the regulatory effects of a cell in response to extracellular matrix (ECM) adhesion which also contribute to ECM degradation. Our observation that E2 induces regulation of lncRNAs associating with FAs may contribute to changes in the structure and functional capacity of the skeletal muscle ECM. In mammals skeletal muscle ECM adapts to conditions that affect muscle growth such as injury or disease, conditions that produce similar negative effects on muscle growth induced by E2 in salmonids. Additionally, spawning female rainbow trout exhibiting muscle atrophy display reduced expression of ECM and FA-related genes, supporting that regulation of related proteins represents muscle restructuring.

In addition to FA, we also found that calcium signaling pathway was linked with 15 lncRNAs in our lncRNA-pathway network. Calcium ions are important for cellular signaling, which plays a pivotal role in almost all cellular processes. Calpain is a protein belonging to the family of non-lysosomal cysteine proteases that is activated by calcium ions. Calpains function in various biological processes in skeletal muscle, including apoptosis, proteolysis of myofibrillar proteins, and myogenic events. Increased caplain expression
occurs concurrently with elevated E2 during spawning in rainbow trout, however, regulation of calpastatins, the calpain inhibitors, also occurs during this period, which is under direct effect of E2. Therefore, regulation of lncRNAs affecting calcium signaling capacity is likely significant for E2-induced elevations in muscle protein degradation. Probably further affecting calcium signaling was down-regulation of calsequestrin, a calcium-binding protein that regulates calcium homeostasis. Our previous findings suggest that target genes of miR-23a-3p are involved in mitochondrial outer membrane permeability; calcium signaling also plays a

Figure 3. GO analysis of differentially expressed mRNAs. (a) Treemap of up-regulated representative GO terms in biological processes. (b) Treemap of down-regulated representative GO terms in biological processes. Each rectangle is a single cluster representative. The representative GO terms were joined into 'superclusters' of loosely related terms with same color.
vital role in mitochondrial permeability which results in cell death. Additional lncRNAs regulated by E2 in skeletal muscle have functional roles in nutrient metabolism, including amino acids, fatty acids, and pyruvate, as well as protein degradation. As E2 shifts nutrient partitioning from muscle accretion in support of gonad development, the energy requirements of muscle are expected to change. These data support a role for lncRNA as regulators of these processes. Although the results of the present study require further experimental verification, they provide insight into the significant role that lncRNAs have in the physiological and biochemical response to E2 in skeletal muscle of rainbow trout.

Various genes functioning as molecular factors, cellular components and those involved in biological processes were up-regulated in rainbow trout skeletal muscle under the influence of E2. These proteins function to regulate different signaling pathways including JAK/STAT pathway, response to oxidative stress, cell-cell signaling and intracellular estrogen receptor signaling pathways with positive regulation of various receptor biosynthesis and their binding activity (insulin receptor binding, insulin like growth factor binding, integrin binding etc). Besides, important protein cross-linking molecules involved in histone ubiquitination, protein autoubiquitination, and their responsible enzymes like ubiquitin conjugating enzyme and its complex were also up-regulated. Elevated expression of important metabolic enzymes, including aldehyde dehydrogenase and succinate dehydrogenase, important for energy synthesis was observed.

Estradiol induces the activation of estrogen receptors that are functional through genomic and non-genomic pathways. Besides activation of estrogen receptors, E2 is recognized as a regulator of the IGF pathway in salmonids. IGFBP-5, an IGF binding protein, was reported to be highly conserved among humans, rats (*Rattus norvegicus*), mice and frog (*Xenopus laevis*), and is abundantly expressed in skeletal muscle during myoblast differentiation in mouse. This response inhibits muscle cell differentiation by binding to the IGF receptor and impeding IGF activity. In the current study, real time PCR and RNAseq analysis both support increased IFGBP-5 expression after 24 hours of E2 treatment, and a similar response was observed in a previous rainbow trout E2 injection study. However, long-term E2 exposure (30 days) in rainbow trout had the opposite effect. The differential time-dependent response could reflect the complex and dynamic nature of the IGF endocrine system, in which relative concentrations of both IGF and IGF binding proteins regulate IGF binding to receptors. Collectively these observations support that regulation of IGFBP-5 in muscle is important for effects of E2 on muscle cell differentiation.

Production of reactive oxygen species leading to oxidative stress as a response to E2 was reported in Japanese medaka (*Oryzias latipes*) and in Japanese sea bass (*Lateolabrax japonicus*). DNA damage induced by reactive oxygen species was reported in fathead minnows (*Pimephales promelas*) and Japanese sea bass exposed to E2.

Figure 4. KEGG pathway analysis of differentially expressed genes. The rich factor was calculated using the gene count divided by the expected gene count.
Reactive oxygen species (ROS) contribute to autophagy and muscle atrophy by activating muscle degrading pathways involving calpains, caspase-3, ATG4b, MuRF-1 and atrogin-1. Supporting this concept is our previous report of increased expression of caspase-9, caspase-3 and atrogin-1 by E2 in rainbow trout. In vivo and in vitro experiments in rainbow trout and rainbow trout primary myocytes showed that exposure to E2 increased expression of atrogin-1/fbxo32 and MURF genes and increased protein degradation. Atrogin-1 is a ubiquitin ligase responsible for ubiquitinating specific proteins such as MyoD and myogenin for proteasomal degradation, and increased atrogin-1 expression is typically associated with catabolic conditions in both fish and mammals. Supporting E2-induced reductions in muscle protein retention in rainbow trout was increased expression of autophagy related 4b cysteine peptidase (ATG4b) and previous literature reports a similar response for additional components of the autophagy-lysosome system. Collectively, these findings indicate the regulation of E2 on the skeletal myogenic program and mechanisms regulating protein retention and muscle accretion.

Figure 5. IncRNA-mRNA co-expression network. The differentially expressed IncRNA-mRNA co-expression network consisted of 681 co-expression relationship between 164 IncRNAs and 201 mRNAs. Blue circles represent IncRNAs, and red circles denote mRNAs.
Characterized GO terms that showed reduced expression in E2-treated skeletal muscle after 24 hours are mainly involved in structural make up of skeletal muscle. Significantly down regulated proteins are primarily involved in actin cytoskeleton, actin binding, motor activity and mesenchymal morphogenesis. Additionally, GO terms related to cAMP biosynthesis, cristae formation, regulation of smoothened signaling pathway, and RNA mediated DNA polymerase activity were differentially regulated. In humans and mouse mesenchymal stromal cells are multipotent cells that differentiate to different cell lineages including mesodermal cells considered as precursors for myoblasts or MPCs. The bone marrow derived mesenchymal cells with myogenic markers migrate to sites of muscle regeneration. In the current study, the observed decrease in the

| Gene ID                      | Degree | Gene name     | Symbol     |
|------------------------------|--------|---------------|------------|
| TRINITY_DN61993_c9_g1        | 29     | Lnc-OM9822    |            |
| TRINITY_DN89620_c0_g1        | 25     | Lnc-OM9785    |            |
| TRINITY_DN61638_c0_g1        | 18     | Lnc-OM9852    |            |
| TRINITY_DN62639_c0_g4        | 16     | Lnc-OM9744    |            |
| TRINITY_DN60962_c3_g3        | 16     | Lnc-OM9748    |            |
| TRINITY_DN57241_c1_g3        | 15     | Lnc-OM9787    |            |
| TRINITY_DN38889_c0_g1        | 14     | Lnc-OM9826    |            |
| TRINITY_DN33630_c0_g1        | 14     | Lnc-OM9694    |            |
| TRINITY_DN57241_c2_g3        | 14     | ADP/ATP translocase 2 ADT2 |
| TRINITY_DN60357_c0_g5        | 13     | Lnc-OM9743    |            |
| TRINITY_DN58760_c7_g3        | 13     | Lnc-OM9759    |            |
| TRINITY_DN46607_c0_g1        | 12     | Lnc-OM9725    |            |
| TRINITY_DN48565_c0_g1        | 12     | Sestrin 2 SESN2 |
| TRINITY_DN63539_c4_g1        | 12     | Actin alpha 1 ACTA1 |
| TRINITY_DN59565_c3_g3        | 12     | Calsequestrin-2 CASQ2 |
| TRINITY_DN62931_c4_g5        | 12     | Lnc-OM9693    |            |
| TRINITY_DN32746_c0_g2        | 11     | Lnc-OM9780    |            |
| TRINITY_DN2730_c1_g2         | 11     | E3 ubiquitin-protein ligase TRIM39 |
| TRINITY_DN55042_c2_g2        | 11     | Anion exchange protein 3 SLC4A3 |
| TRINITY_DN60070_c2_g5        | 11     | Lnc-OM9778    |            |

Table 3. Top 20 nodes with the highest degree in differential IncRNA-mRNA co-expression network.

Figure 6. IncRNA-pathway network. Blue circles represent IncRNAs, and orange circles denote pathways.
genes responsible for mesenchymal cell migration and morphogenesis may have negative effects on muscle development. Wnt signaling in rat is necessary for differentiation of mesenchymal cells to myogenic origin [57], which also inhibits their differentiation to adipogenic origin by decreasing the expression of CCAAT enhancer binding protein alpha and peroxisome proliferator-activated receptor gamma (PPARγ) [57]. Peroxisome proliferator-activated receptor gamma coactivator 1 alpha (PPARGC1A) is a coactivator of PPARγ and highly expressed in mouse skeletal muscle [58,59]. Studies also indicate that PPARGC1A and PPARGC1B play prominent roles in estrogen receptor signaling pathway by acting as cofactors to enhance transactivation of estrogen receptor α [60–62]. In our study, PPARGC1A exhibited increased expression in skeletal muscle within 24 hours of E2 exposure, supporting that effects of E2 in muscle extend to regulating mechanisms affecting mesenchymal cell differentiation and fiber type.

In summary, our comprehensive analyses provide novel knowledge of mRNAs and lncRNAs at the transcriptomic level regarding the influence of E2 on rainbow trout skeletal muscle within a relatively short period after steroid injection (24 hr). These results and conclusions will serve as important resources for future experiments that further investigate the role and regulation of lncRNAs in rainbow trout. The dynamic expression response observed in the current study may be driven by a dramatic reduction in plasma E2 between 24 and 72 hrs post-injection due to high E2 clearance. A previous study using an identical injection approach reported a greater than 3-fold reduction in plasma E2 between these sampling periods (121 ng/ml vs 36 ng/ml) [14]. Therefore, an experimental model that enables sustained elevated plasma E2 would be especially valuable to mimic the steady-state elevation in E2 production during sexual maturation and further coordinate the gene expression data with changes in body weight and dynamics of muscle accretion.

### Methods

**Ethics statement.** All animal experiments in this study were performed at the USDA/ARS National Center for Cool and Cold Water Aquaculture (NCCCWA) and approved by the NCCCWA Institutional Animal Care and Use Committee (protocol #50).

**Experimental design.** A total of forty fish weighing approximately 40 g were randomly assigned to four experimental tanks (n = 10 fish per tank). The study consisted of two treatments, including intraperitoneal injections of E2 and the delivery vehicle to serve as the control. Treatments were applied to fish in duplicate tanks. Estradiol was resuspended (10 μg/μL) in 95% ethanol and diluted to 2.5 μg/μL with vegetable oil. The control treatment contained an equal ratio of ethanol: vegetable oil as compared to E2 suspension. Estradiol and the vehicle injection methodology was adapted from previously published procedures used in tilapia [63]. Feed was withheld the day of E2 injection and throughout the study period. Fish were anesthetized with tricaine methanesulphonate (MS-222, 100 mg/L), weighed, and received intraperitoneal injections (2.0 μL/g body weight) of E2 (5.0 μg/g body weight) or the vehicle. Fish in one tank per treatment were each harvested 24 and 72 hours post-injection using lethal dose of MS-222 (300 mg/L). Skeletal muscle was dissected and immediately frozen using liquid nitrogen for further processing.

| Regulatory lncRNAs (n) | Pathway                                      |
|------------------------|----------------------------------------------|
| 30                     | Focal adhesion                               |
| 28                     | Regulation of actin cytoskeleton              |
| 17                     | Cardiac muscle contraction                    |
| 15                     | Calcium signaling pathway                     |
| 15                     | Rap1 signaling pathway                        |
| 13                     | PPAR signaling pathway                        |
| 10                     | Adipocytokine signaling pathway               |
| 9                      | AMPK signaling pathway                        |
| 7                      | Carbon metabolism                             |
| 6                      | Jak-STAT signaling pathway                    |
| 6                      | Protein processing in endoplasmic reticulum   |
| 5                      | Arginine and proline metabolism               |
| 5                      | Fatty acid metabolism/degradation             |
| 5                      | Beta-Alanine metabolism                       |
| 4                      | Citrate cycle (TCA cycle)                     |
| 4                      | Protein digestion and absorption              |
| 3                      | Regulation of autophagy                       |
| 2                      | Estrogen signaling pathway                    |
| 2                      | Insulin resistance                            |
| 2                      | Pyruvate metabolism                           |

**Table 4.** Key pathways in E2 treated skeletal muscle.
RNA extraction and quality control. Total RNA was extracted from skeletal muscle samples of fish treated with E2 and controls at 24 and 72 hours (6 biological replicates each) using TRIZOL reagent (Invitrogen, Carlsbad, CA) per the manufacturer’s suggested protocol. Quality and quantity of RNA was estimated using the A260/A280 ratio. Integrity and size distribution of RNA was evaluated using a Bioanalyzer 2100 (Agilent technologies, Santa Clara, CA). Four replicates of each treatment were subject to RNAseq.

cDNA library construction and sequencing. Library construction (poly(A)-enriched) and sequencing was completed by the Johns Hopkins University Genetic Resources Core facility High Throughput Sequencing Center (Baltimore, MD). cDNA libraries were constructed using TruSeq RNA library preparations and high throughput sequencing was completed using the Illumina HiSeq platform (100 bp single reads).

Sequence data processing, de novo assembly and differential expression analysis. Adaptor sequences were trimmed and ambiguous and low quality bases were removed, then read lengths less than 50 were removed. TRINITY17 was used to assemble all cleaned reads with default parameters. CD-HIT-EST was used to remove the shorter redundant transcripts when they were 100% covered by other transcripts with more than 99% identity18. All the cleaned reads were mapped to the assembled transcriptome by Bowtier26. RSEM was used to estimate and quantify gene expression levels from RNA-Seq data18. The final counts matrix file was used as input for the R package edgeR19. The exact test in edgeR was completed to discover the DEGs between the treatment groups. False discovery rate was used for multiple test correction. Any genes with a fold change greater than 2.0 and FDR of less than 0.05 were defined as DEG.

Validation of sequencing results. Sequencing results were validated using real time RT-PCR. RNA from 24 hour samples (6 replicates from each treatment) was used for cDNA synthesis using miScript II (Qiagen, Valencia, CA). Expression of mRNA and IncRNA was normalized to β-actin, which was not identified as a DEG by RNASeq analysis. Real time RT-PCR followed by melt curve analysis was performed to determine the specificity of the amplicons. Standard curves with 10-time serial dilutions of a pooled cDNA sample were generated to calculate the efficiency of qPCR. Five μl of diluted cDNA (1:1024) was used with IQ™ SYBR® Green Supermix (Bio-Rad) and 300 nM of forward and reverse primer in a final volume of 25 μl for the reaction. Cycle conditions were the same for all the primers except the annealing temperatures for different primers. Initial denaturation at 95 °C for 3 minutes followed by a denaturation at 95 °C for 40 sec, annealing for 30 sec, extension at 72 °C for 30 seconds for 40 cycles and a final extension for 10 min at 72 °C. Melt curve analysis was performed with an increase of 0.5 °C every cycle. Quantification cycle (Cq) values were used for quantification of expression using the log-linear equation of standard curve. Relative fold changes were calculated by setting the values of the controls to 1.0 and comparing the respective treatment groups. Statistical analysis was performed using Student’s t-test and those with a P-value < 0.05 were considered statistically significant.

LncRNA identification and GO and KEGG enrichment analysis of mRNAs. All DEGs were mapped to the rainbow trout genome65 using BLAT66. We used the pipeline reported in our previous study21 for lncRNA identification to detect differentially expressed IncRNAs. All differentially expressed mRNAs were subjected to similarity search against NCBI non-redundant (nr) protein database using BLASTx67 with an e-value cutoff of 1e-10. Gene names and GI were assigned to each mRNA based on the BLASTx result. ID mapping was performed using our in-house script to extract all associated GO terms for each mRNA. KEGG pathways were assigned to each mRNA using the online KEGG Automatic Annotation Server (KASS)68. The R package GOstats was used to run hypergeometric testing on GO and KEGG terms69. Redundant GO terms were removed by REVIGO, a Web semantic similarity measurement based clustering algorithm70.

Co-expression analysis. To identify co-expressed lncRNA-mRNA pairs, Pearson’s correlation coefficients were calculated based on the normalized expression value between every differentially expressed lncRNA and mRNA pair. Only the strong correlations (≥0.99) were selected to construct the network. The threshold of FDR was < 0.05. Cytoscape was used to construct the co-expression network22.

Data availability. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References
1. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1842, 1910–1922, https://doi.org/10.1016/j.bbadis.2014.03.011 (2014).
2. Nie, M., Deng, Z.-L. & Liu, J. & Wang, D.-Z. Noncoding RNAs, Emerging Regulators of Skeletal Muscle Development and Diseases. BioMed Research International 2015, 17, https://doi.org/10.1155/2015/767575 (2015).
3. Mousavi, K. et al. eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Molecular Cell 51, 606–617, https://doi.org/10.1016/j.molcel.2013.07.022 (2013).
4. Lu, L. et al. Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. The EMBO Journal 32, 2575–2588, https://doi.org/10.1038/emboj.2013.182 (2013).
5. Froehlich, J. M., Galt, N. J., Charging, M. J., Meyer, B. M. & Biga, P. R. In vitro indeterminate telost myogenesis appears to be dependent on Pax3. In vitro cellular & developmental biology. Animal 49, 371–385, https://doi.org/10.1007/s11626-013-9616-2 (2013).
6. Steinbacher, P. et al. MyoD and Myogenin expression during myogenic phases in brown trout: a prerequisite onset of mosaic hyperplasia is a prerequisite for fast somatic growth. Developmental dynamics: an official publication of the American Association of Anatomists 236, 1106–1114, https://doi.org/10.1002/dvdy.21103 (2007).

7. Rossi, G. & Messina, G. Comparative myogenesis in teleosts and mammals. Cellular and molecular life sciences: CMLS 71, 3081–3099, https://doi.org/10.1007/s00018-014-1604-3 (2014).

8. Cleveland, B. M. & Weber, G. M. Effects of steroidal treatment on growth, nutrient partitioning, and expression of genes related to growth and nutrient metabolism in adult triploid rainbow trout (Oncorhynchus mykiss). Domest Anim Endocrinol 56, 1–12, https://doi.org/10.1016/j.domaniend.2016.01.001 (2016).

9. Olin, T., Nazar, D. S. & von der Decken, A. Response of epaxial muscle and liver to 17β-estradiol in fed and starved Atlantic salmon (Salmo salar). Aquat 99, 179–191 (1991).

10. Cleveland, B. M. & Manor, M. L. Effects of phytoestrogens on growth, related and lipogenic genes in rainbow trout (Oncorhynchus mykiss). Comp Biochem Phys C 170, 28–37, https://doi.org/10.1016/j.cbpc.2015.02.001 (2015).

11. Davidson, J. Effects of steroid treatment on growth, nutrient partitioning, and expression of genes related to growth and nutrient metabolism in adult triploid rainbow trout (Oncorhynchus mykiss). Marine biotechnology (New York, N.Y.) 15, 132–142, https://doi.org/10.1007/s10273-013-0211-6 (2013).

12. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 24, 2572–2573, https://doi.org/10.1093/bioinformatics/btq476 (2008).

13. Olin, T. & van der Decken, A. Estrogen treatment and its implication on vitellogenin and myosin synthesis in salmon (Salmo salar). Physiol Zool 60, 346–351 (1987).

14. Cleveland, B. M. & Weber, G. M. Effects of sex steroids on expression of genes regulating growth-related mechanisms in rainbow trout (Oncorhynchus mykiss). Gen Comp Endocrinol 216, 103–115, https://doi.org/10.1016/j.ygcend.2014.11.018 (2015).

15. Davidson, J. et al. Growth performance, fillet quality, and reproductive maturity of Rainbow Trout (Oncorhynchus mykiss) cultured to 5 kilograms within freshwater recirculating systems. Journal of Aquaculture Research & Development 2014, doi:https://doi.org/10.4172/2155-9546.1000238 (2014).

16. Aussanasuwannakul, A. et al. Effect of Sexual Maturation on Thermal Stability, Viscoelastic Properties, and Texture of Female Rainbow Trout, Oncorhynchus mykiss, Fillets. J Food Sci 77, 577–583, https://doi.org/10.1111/j.1750-3841.2012.02512.x (2012).

17. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29, 644–652, https://doi.org/10.1038/nbt.1883 (2011).

18. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659, https://doi.org/10.1093/bioinformatics/btl158 (2006).

19. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140, https://doi.org/10.1093/bioinformatics/btp616 (2010).

20. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10, R25, https://doi.org/10.1186/gb-2009-10-3-r25 (2009).

21. Wang, J. et al. Identification and Functional Prediction of Large Intergenic Noncoding RNAs (lincRNAs) in Rainbow Trout (Oncorhynchus mykiss). Marine biotechnology (New York, N.Y.) 18, 271–282, https://doi.org/10.1007/s10273-016-9689-5 (2016).

22. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550, https://doi.org/10.1186/s13059-014-0550-8 (2014).

23. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. Cell 136, 629–641, https://doi.org/10.1016/j.cell.2009.02.006 (2009).

24. Liao, Q. et al. Large-scale prediction of long non-coding RNA functions in a coding non-coding gene co-expression network. Nucleic acids research 41, e35, https://doi.org/10.1093/nar/gks967 (2013).

25. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America 95, 14863–14868 (1998).

26. Chen, C. S., Alonso, J. L., Ostuni, E., Whitesides, G. M. & Ingber, D. E. Cell shape provides global control of focal adhesion assembly. Biochemical and biophysical research communications 307, 355–361 (2003).

27. Wang, Y. & McNiven, M. A. Invasive matrix degradation at focal adhesions occurs via protease recruitment by a FAK-p130Cas complex. The journal of cell biology 196, 375–385, https://doi.org/10.1083/jcb.201015513 (2013).

28. Gillies, A. R. & Lieber, R. L. Structure and function of the skeletal muscle extracellular matrix. Muscle Nerve 44, 318–331, https://doi.org/10.1002/mus.22094 (2011).

29. Norbeck, L. A. & Sheridan, M. A. An in vitro model for evaluating peripheral regulation of growth in fish: effects of 17β-estradiol and testosterone on the expression of growth hormone receptors, insulin-like growth factors, and insulin-like growth factor type I receptors in rainbow trout (Oncorhynchus mykiss). General and Comparative Endocrinology 173, 270–280, https://doi.org/10.1016/j.ygcen.2011.06.009 (2011).

30. James, P. L., Jones, S. R., Busby, W. H., Clemmons, D. R. & Rotwein, P. A highly conserved insulin-like growth factor-binding protein (IGFBP-5) is expressed during myoblast differentiation. Journal of Biological Chemistry 268, 22305–22312 (1993).
41. Mukherjee, A., Wilson, E. M. & Rotwein, P. Insulin-Like Growth Factor (IGF) Binding Protein-5 Blocks Skeletal Muscle Differentiation by Inhibiting IGF Actions. Molecular Endocrinology 22, 206–215, https://doi.org/10.1210/me.2007-0336 (2008).

42. Cavaleri, E., Frenkel, K., Liehr, J. G., Rogan, E. & Roy, D. Estrogens as Endogenous Genotoxic Agents—DNA Adducts and Mutations. INCI Monographs 2000, 75–90 (2000).

43. Woo, S., Won, H., Lee, A. & Yun, S. Oxidative stress and gene expression in diverse tissues of Oryzias javanicus exposed to 17β-estradiol. Molecular & Cellular Toxicology 8, 263–269, https://doi.org/10.1007/s12373-012-0032-6 (2012).

44. Thilagam, H., Gopalakrishnan, S., Qu, H.-D., Bo, J. & Wang, K.-J. 17β estradiol induced ROS generation, DNA damage and enzymatic responses in the hepatic tissue of Japanese sea bass. Ecotoxicology 19, 1258–1267, https://doi.org/10.1007/s10646-010-0510-3 (2010).

45. Filby, A. L. et al. Health Impacts of Estrogens in the Environment, Considering Complex Mixture Effects. Environmental Health Perspectives 115, 1704–1710, https://doi.org/10.1289/ehp.10443 (2007).

46. Wirawan, E., Berghoe, T. V., Lippens, S., Agostinis, P. & Vandenabeele, P. Autophagy: for better or for worse. Cell Research 22, 43–61, https://doi.org/10.1038/cr.2011.152 (2012).

47. Yoshida, T. et al. Molecular mechanisms and signaling pathways of angiogenesis II-induced muscle wasting: potential therapeutic targets for cardiac cachexia. The international journal of biochemistry & cell biology 45, 2322–2332, https://doi.org/10.1016/j.biocel.2013.05.035 (2013).

48. Cleveland, B. M. & Weber, G. M. Effects of sex steroids on indices of protein turnover in rainbow trout (Oncorhynchus mykiss) white muscle. General and Comparative Endocrinology 174, 132–142, https://doi.org/10.1016/j.ygcen.2011.08.011 (2011).

49. Jogo, M. & Shiraishi, S. & Tamura, T.-a. Identification of MAFbx as a myogenin–engaged F-box protein in SCF ubiquitin ligase. FEBS Letters 583, 2715–2719, https://doi.org/10.1016/j.febslet.2009.07.033 (2009).

50. Lagirand-Cantaloube, J. et al. Inhibition of Atrogin-1/MAFbx Mediated MyoD Proteolysis Prevents Skeletal Muscle Atrophy In Vivo. PLoS ONE 4, e4897, https://doi.org/10.1371/journal.pone.0004897 (2009).

51. Cleveland, B. M. & Evenhuis, J. P. Molecular characterization of atrogin-1/F-box protein-32 (FBXO32) and F-box protein-25 (FBXO25) in rainbow trout (Oncorhynchus mykiss): Expression across tissues in response to feed deprivation. Comp Biochem Physiol B Biochem Mol Biol 157, 248–257, doi:S1096-4590%2810%2900177-6 [pii] 10.1016/j.cbpb.2010.06.010 (2010).

52. Poletta, V. C., White, L. J., Larsen, A. E., Leger, B. & Russell, A. P. The role and regulation of MAFbx/atrogen-1 and MuRF1 in skeletal muscle atrophy. Pflugers Arch 461, 325–335, https://doi.org/10.1007/s00424-010-0919-9 (2011).

53. Sassoli, C. et al. Defining the role of mesenchymal stromal cells on the regulation of matrix metalloproteinases in skeletal muscle cells. Experimental Cell Research 323, 297–313, https://doi.org/10.1016/j.yexcr.2014.03.003 (2014).

54. Pittenger, M. F. et al. Multilineage Potential of Adult Human Mesenchymal Stem Cells. Science 284, 143–147 (1999).

55. Ferrari, G. et al. Muscle Regeneration by Bone Marrow-Derived Myogenic Progenitors. Science 279, 1528–1530 (1998).

56. Parker, M. H., Seale, P. & Rudnicki, M. A. Looking back to the embryo: defining transcriptional networks in adult myogenesis. Nat Rev Genet 4, 497–507 (2003).

57. Shang, Y.-c. et al. Wnt3a signaling promotes proliferation, myogenic differentiation, and migration of rat bone marrow mesenchymal stem cells. Acta Pharmacol Sin 28, 1761–1774 (2007).

58. Puigserver, P. & Spiegelman, B. M. Peroxisome Proliferator-Activated Receptor-γ Coactivator 1α (PGC-1α): Transcriptional Coactivator and Metabolic Regulator. Endocrine Reviews 24, 78–90, https://doi.org/10.1210/er.2002-0012 (2003).

59. Lin, J. et al. Transcriptional co-activator PGC-1[alpha] drives the formation of slow-twitch muscle fibres. Nature 418, 797–801 (2002).

60. Wirtenberger, M. et al. Associations of genetic variants in the estrogen receptor coactivators PPARGC1A, PPARGC1B and EP300 with familial breast cancer. Carcinogenesis 27, 2201–2208 (2006).

61. Kressler, D., Schreiber, S. N., Knutti, D. & Krali, A. The PGC-1-related Protein PERC Is a Selective Coactivator of Estrogen Receptor α. Journal of Biological Chemistry 277, 13918–13925 (2002).

62. Tcherepanova, I., Puigserver, P., Norris, J. D., Spiegelman, B. M. & McDonnell, D. P. Modulation of Estrogen Receptor–α Transcriptional Activity by the Coactivator PGC-1. Journal of Biological Chemistry 275, 16302–16308 (2000).

63. Davis, L. K. et al. Induction of three vitellogenins by 1β-estradiol with concurrent inhibition of the growth-hormone-insulin-like growth factor 1 axis in a euryhaline teleost, the tilapia (Oreochromis mossambicus). Biol Reprod 77, 614–625, doi:10.1095/biolreprod.107.060947 [pii] 10.1095/biolreprod.107.060947 (2007).

64. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bioinformatics 26, 564–574, https://doi.org/10.1093/bioinformatics/btq036 (2010).

65. Falcón, S. & Gentleman, R. Using GOstas to test gene lists for GO term association. Bioinformatics (Oxford, England) 23, 257–258, https://doi.org/10.1093/bioinformatics/btl567 (2007).

66. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6, e21800, https://doi.org/10.1371/journal.pone.0021800 (2011).

67. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13, 2498–2504, https://doi.org/10.1101/gr.123903.2003.

Acknowledgements

The authors acknowledge invaluable contributions from Greg Weber, Josh Kretzer, Lisa Radler, and Kyle Jenkins. Mention of trade names is solely to provide accurate information and does not imply endorsement by the USDA. The USDA is an equal opportunity employer and provider. This project was funded in part by NIH grant #GM114105.

Author Contributions

B.C. conceived, designed, and performed the experiments; J.W. analyzed the data and P.K. validated the data, both prepared figures and manuscript; B.C., J.Y. and S.W. reviewed the manuscript. All authors contributed to the manuscript at various stages.
