2018

Elucidation of the molecular mechanism through which estradiol and trenbolone acetate improve skeletal muscle growth in beef cattle

Kara Thornton-Kurth
Utah State University, kara.thornton@usu.edu

Expected Data Type

This project will generate both *in vitro* cell culture data as well as *in vivo* cattle data. The *in vitro* cell culture data collected will consist of four different datasets: 1) proliferation rate, 2) protein synthesis rate, 3) RNAseq data, and 4) 2DE proteomic data. In addition, images of different treatments provided to both proliferating and fused cultures will also be collected. The *in vivo* cattle data will also consist of four different datasets: 1) cattle growth performance data, 2) RNAseq data, 3) 2DE proteomics data, and 4) carcass data. All data collected and used in this project will be primary data. *In vivo* data will all be generated through lab work and *in vitro* experiments will come from a combination of lab work (RNAseq and 2DE proteomics data) and field work (cattle growth data and carcass data).

Data Format

The data collected from both the *in vivo* and *in vitro* experiments will be in the following formats:

- ***Proliferation data:*** All proliferation data collected will be recorded as total BrdU incorporation and stored in an excel file that contains the basic experiment information (bovine satellite cell isolate number, replicate number, date).
- ***Protein synthesis data:*** All protein synthesis data will be recorded as total amount of protein synthesized and stored in an excel file that contains the basic experiment information (bovine satellite cell isolate number, replicate number, date).
- ***RNAseq data:*** Transcriptome analysis creates the following file types and formats: 1) DNA sequence and quality files for each sample (fastq format); 2) Binary read mapping files for each sample (BAM format); 3) Binary bed files of read coverage for genome browser tracks (BigWig format); 4) Transcriptome definition files for individual samples and a unified non-redundant file for an experimental series of cell culture ot cattle experiments (gtf/gff format); 5) Cuffdiff2 data sets for importation to R/Bioconductor/CummeRbund (series of 23 tab-delimited files).
- ***2DE proteomics data:*** The first stage of the 2DE proteomic analysis will result in a relative densitometry value for each spot on each gel. This relative value will be obtained used the DeCyder software. The second stage of the 2DE proteomic analysis will use mass spectrometry to identify exactly what the specific protein is within some of the spots. The resulting peptide mass and the associated fragmentation spectra will be submitted to a GPS Explorer workstation equipped with a MASCOT search engine to search the database of the National Center for Biotechnology Information non-redundant. This will result in the protein name as well as a confidence interval % for that protein.
- ***Cattle growth data:*** General growth data on all cattle will be collected throughout the feedlot trial. This data will include bi-weekly weights, daily intakes, average daily gains, and feed efficiency. This data will be stored in an excel file.
- ***Carcass data:*** General carcass data will be collected on cattle once they go to harvest at JBS in Hyrum, UT. This data will be collected by the plant and emailed in an excel file that will contain hot carcass weight, marbling score, quality grade, yield grade, and ribeye area. This data will be stored in an excel file.

Data Storage and Preservation

All data collected from both the *in vivo* and *in vitro* experiments as well as the accompanying metadata will be stored and managed on the Utah State University "Box" system. Additionally,

all data stored on Box will be backed up using "Fortress" at Purdue University by Dr. Bidwell. Fortress is designed for long-term storage but has slower access. Currently, there are no storage limits or costs associated with using Fortress. More detailed descriptions of computational and data storage facilities are provided in the Facilities and other Resources section. Furthermore, all research data obtained will also be stored on several different hard drives/servers in the labs of the individuals responsible for collecting that particular piece of data. All cell culture data (i.e. proliferation rates, protein synthesis rates, and images) will be stored on the hard drive of the lab computer, the hard drive of the laboratory technician, as well as the hard drive of Dr. Thornton, the PD. RNA seq data will initially be stored on a server within the genomics core at Utah State University and subsequently backed-up onto Box. This data will be transferred to Purdue University through an FTP server to the "Data Depot" at Purdue University. The Data Depot has redundant daily back-up and can be directly accessed by any of the community clusters for data analysis. Sequence data as well as processed data files (below) will reside on Data Depot. All RNAseq data will also be backed-up to Fortress. The 2DE proteomic data will be stored in the Center for Integrated Biosystems Proteomic core, and on the hard-drives of both the PD and her laboratory technician in addition to being stored on Box. All of the cattle growth and carcass data will be stored on Box as well as on the hard-drive of both the PD and her technician. All of the data generated in this project will be stored long-term on both Box and Fortress.

Data Sharing and Public Access

Permanent storage and public access to all of the data collected will occur through submission of data to the Gene Expression Omnibus and Sequence Read Archive (GEO/SRA) at the National Center for Biotechnology Information. Specifically, public access to the RNA sequencing data and the essential gene expression results (gene_name, ensembl_id, genome, locus, FPKM/RPKM values by biological replicate) will be stored as metadata and include all results obtained in both the cell culture and the cattle experiments that originally produced the RNA. GEO/SRA prefers original read sequence files which will be provided along with the essential gene expression results and metadata for the cell cultures or animals that produced the RNA. The metadata will also consist of proliferation rate, protein synthesis rate, 2DE proteomic data, and images in the *in vivo* experiments and individual cattle growth data, carcass data and 2DE proteomic data in the *in vitro* experiments.

Roles and Responsibilities

The PD will be responsible for both coordinating and ensuring that all data generated in this project is properly stored, backed-up and shared. In addition, as a back-up, Dr. Bidwell will store all data in Fortress at Purdue University as a safe-guard to having all files stored on Box. In addition, during the time in which RNAseq analysis is being done, all RNAseq data will be stored on Data Depot. Storage space on Data Depot is purchased in 1 Tb increments for $150/year; this cost has been included in both the budget request and the budget narrative.