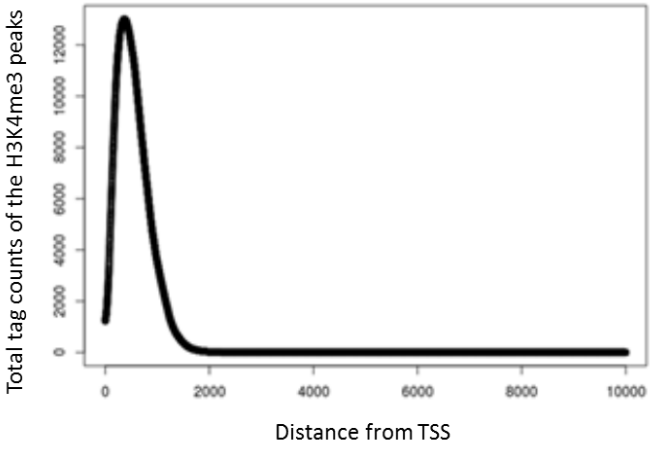
The landscape of histone modifications in the rainbow trout genome: preliminary data for FAASG

Moh Salem, Middle Tennessee State University

The FAASG consortium "Functional Analysis of All Salmonid Genomes" has been established to study the functional genomic basis of phenotypic variation in all salmonids (https://www.faasg.org/). The outcomes will help in improving the sustainability of aquaculture industry in the U.S. An annotated (RefSeq) trout genome sequence is now available (GenBank Accession GCA_002163495). It is composed of a 2.17 GB genome assembly, containing 139,726 scaffolds with N50 > 1.7Mb. Using dense SNP genetic maps, ~88% of the assembly was anchored and ordered within chromosomes. A gene models approach was used to annotate the genome sequence, predicting 153,807 transcripts. We also annotated the genome with four other transcriptome drafts. 1) A transcriptome reference sequence using a 19X coverage of Sanger and 454-pyrosequencing data (Salem et al. 2010). 2) A transcriptome sequenced representing responses to several stressors affecting the aquaculture production environments (Sanchez et al. 2011). 3) A transcriptome using 4,333X sequence coverage for tissue gene expression atlas (Salem et al. 2015). 4) A lncRNA transcriptome reference (Al-Tobasei et al. 2016).

FAASG aims to annotate the rainbow trout genome for various functional elements such as enhancer/cis-reg. elements, promoter flank, and TSS. Also, we aim to annotate the genome for chromatin histone modifications, DNase hypersensitivity, and transcription factors, by integrating data from RNA-seq, DNAse-seq, and ChIP-seq. However, so far, no data are available from the latter two techniques in rainbow trout.

To establish protocols for the ChIP-seq histone modifications, recently, we generated high-resolution maps of the permissive transcription H3K4me3 and repressive H3K27me3 histone marks during *in vitro* myogenesis (data not published). Two samples were collected from 3 conditions: 1) day-4 myoblasts, 2) day-8 nascent myotubes and 3) day-8 serum-starved myotubes. Data from ChIP-seq and RNA-seq data were integrated. An initial data analysis showed distribution of the H3K4me3 marks near the active promotors (figure). Also, correlations between the histone marks and differential gene expression between different conditions were observed.



Here we propose to add two more histone marks to the study: 1) H3K27ac which is associated with high activation of transcription and defined as active enhancer mark; 2) H3K9me3 which is found in constitutively repressed genes. The total cost is $14,200, we are requesting $10,000 and the rest of the budget will be covered by MTSU internal grant. Libraries will be prepared and sequenced at Epigentek Group Inc. Sequence reads will be mapped to the genome and Chip-Seq analysis will be done using MACS2 pipeline. These data will be necessary for the FAASG work and will provide valuable genome annotations that will be available to the salmonid community. Also, this NRSP8 fund will serve as a seed to generate preliminary data and show support to FAASG members in applying for additional funds (e.g. NIFA).