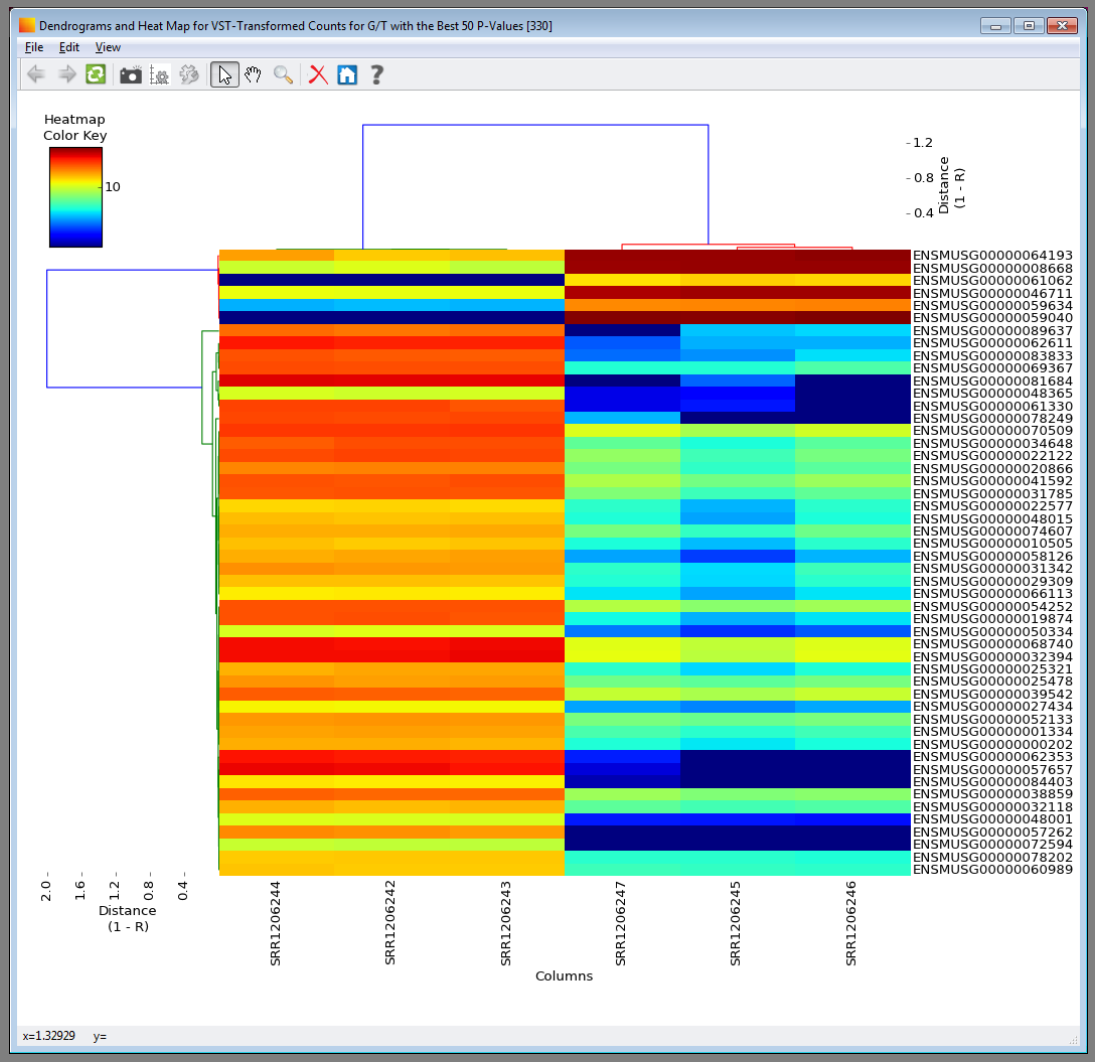



# RNASeq Functionality in SVS Using Public Data

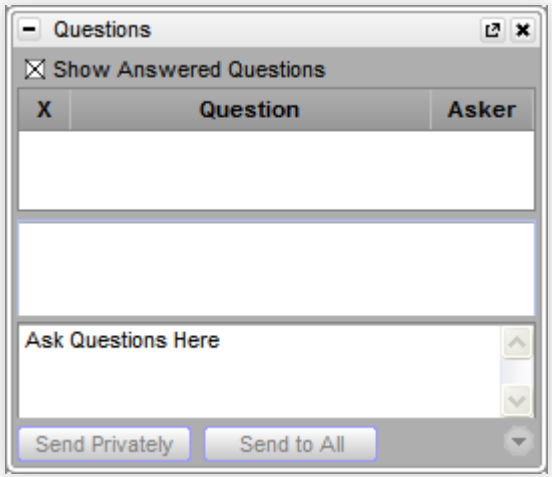
September 23, 2014

Ashley Hintz
Field Application Scientist

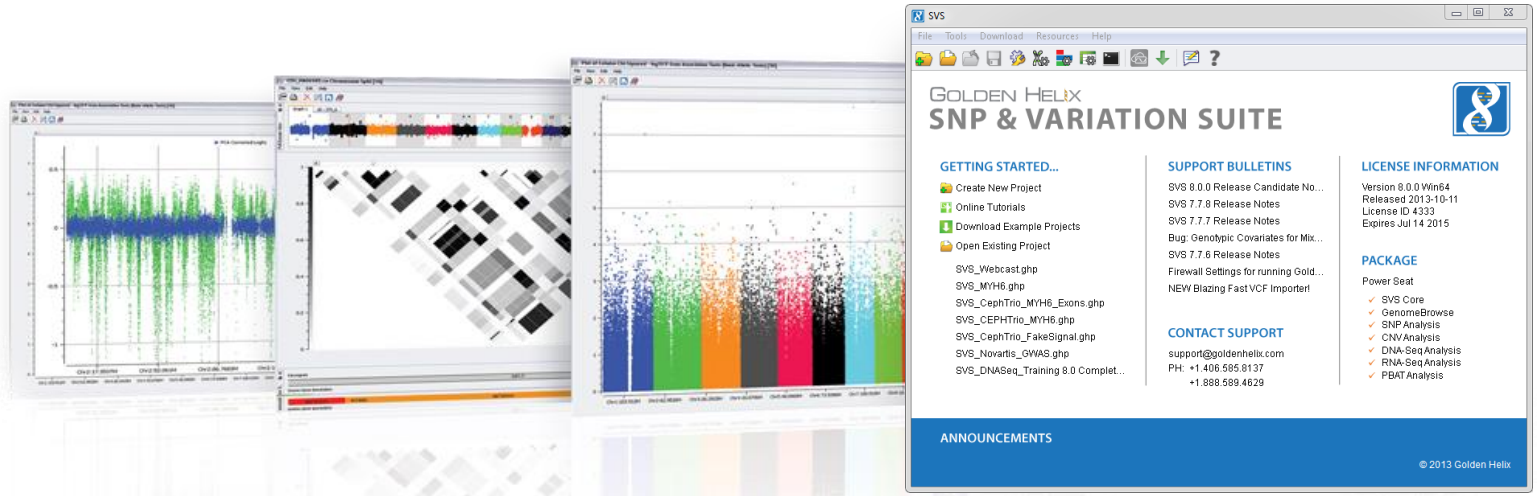
GOLDEN HELIX
Accelerating the Quest for Significance™

# Questions during the presentation

Use the Questions pane in your GoToWebinar window



# SNP & Variation Suite (SVS)




## Core Features

- Powerful Data Management
- Rich Visualizations
- Robust Statistics
- Flexible
- Easy-to-use

## Applications

- Genotype Analysis
- DNA sequence analysis
- CNV Analysis
- RNA-seq differential expression
- Family Based Association

# Agenda

1. RNASeq Introduction
2. GEO Dataset
3. Demonstration
4. Conclusions

# Agenda

1. RNASeq Introduction
2. GEO Dataset
3. Demonstration
4. Conclusions

# NGS RNA-seq Analysis

**Primary Analysis**

- Analysis of hardware generated data, on-machine real-time stats.
- Production of sequence reads and quality scores
- Typical product is "**FASTQ**" file

**Secondary Analysis**

- Recalibrating, de-duplication, QA and clipping/filtering reads
- Alignment/Assembly of reads
- Read quanitification by gene or transcript
- Typical products are "**BAM**" files and/or count data

**Tertiary Analysis**

**"Sense Making"**

- QA and filtering of count data
- Visualization of sequence data in genomic context
- Sample and population summary statistics
- Differential expression analysis

# Introduction

- What is RNASeq?
  - Uses next-generation sequencing technology to capture a snapshot of RNA presence and quantity from the genome at a given moment of time in a specific tissue also called the transcriptome.

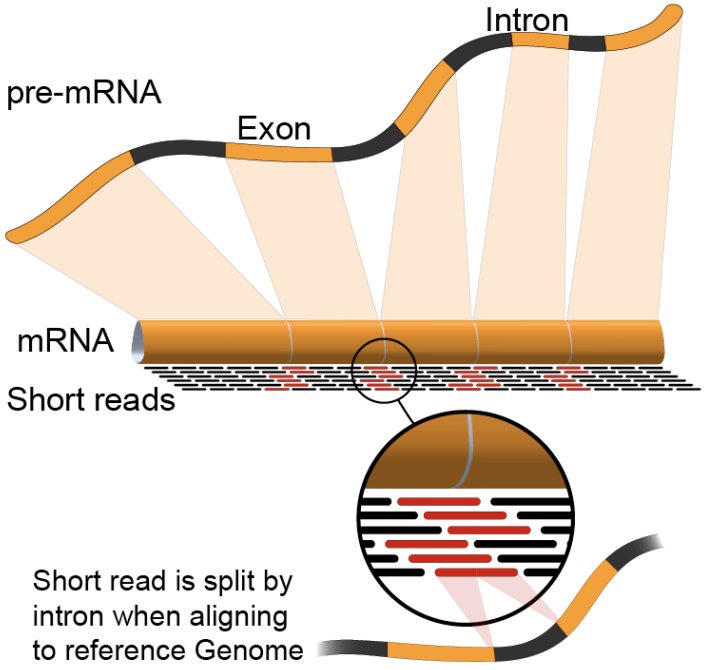



Image Courtesy Rgocs:
http://en.wikipedia.org/wiki/RNA-Seq

# Introduction

- What are the products of the secondary analysis pipeline?
  - **Counts**: simply the number of reads overlapping a given feature, such as a gene, in the genome.
  - **RPKM**: Reads Per Kilobase of exon per Million: term was developed before paired-end sequencing techniques and counting reads would effectively double the number of sequenced molecules
  - **FPKM**: Fragments Per Kilobase of exon per Million: term "fragment" used to accommodate the paired-end nature of sequencing. They are normalized by dividing by the total length of all exons in the gene (or transcript)
  - **TPM**: Transcripts Per Million – simple normalization, essentially states that out of a million transcripts found in a cell, how many would be from this gene?

# Agenda

1. RNASeq Introduction
2. GEO Dataset
3. Demonstration
4. Conclusions

# Dataset Overview

- Downloaded from Gene Expression Omnibus (GEO)
- Maeda *et al.* 2014, accession number GSE56284
- Spinal Muscular Atrophy (SMA) caused by mutation in SMN1 or SMN2
- Ubiquitously expressed so utilizing RNASeq technology to identify expression differences between mice with these mutations vs control

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

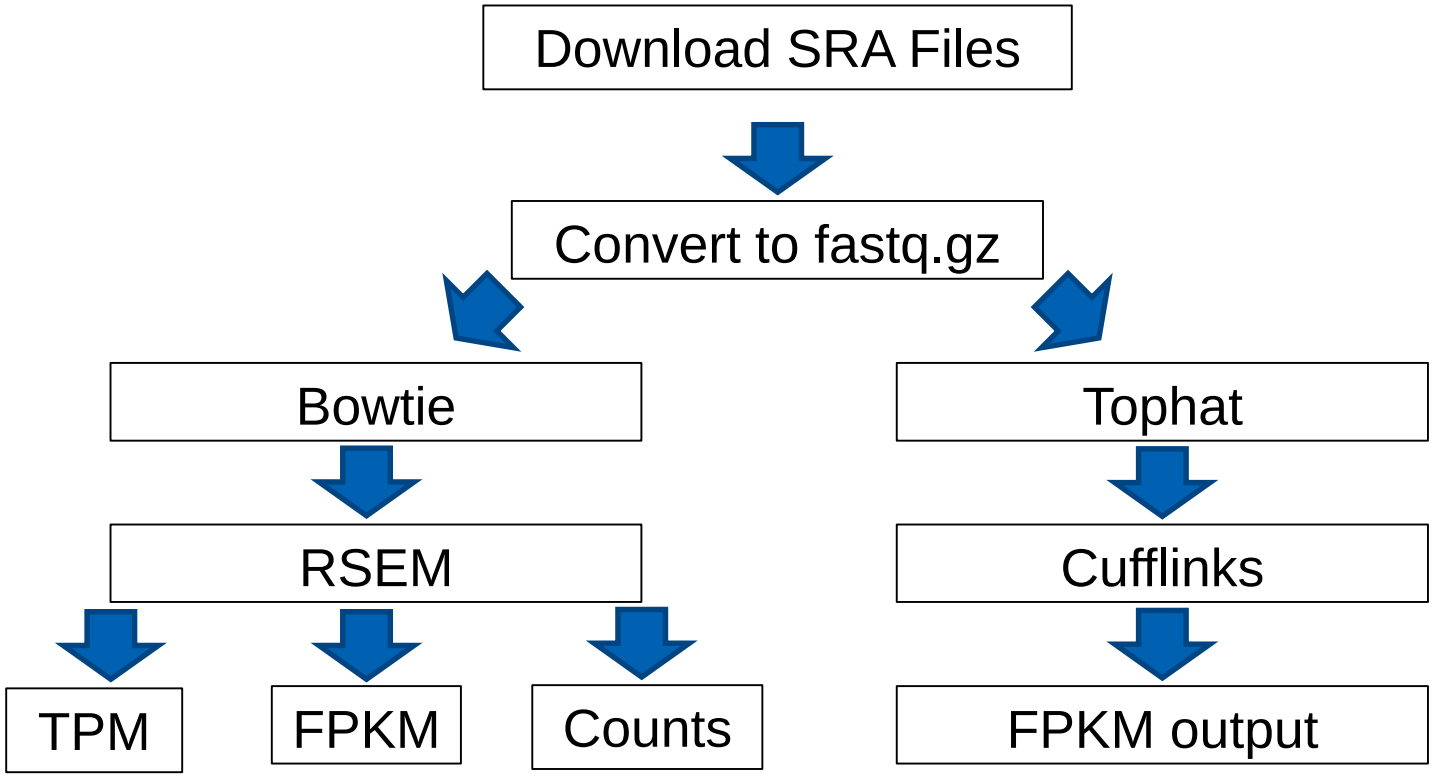
NCBI > GEO > **Accession Display** | Not logged in | Login

Scope: Self | Format: HTML | Amount: Quick | GEO accession: GSE56284 | GO

**Series GSE56284** — Query DataSets for GSE56284

| | |
|---|---|
| Status | Public on Sep 08, 2014 |
| Title | Transcriptome profiling of severe spinal muscular atrophy mouse embryonic stem cell-derived motor neurons |
| Organism | Mus musculus |
| Experiment type | Expression profiling by high throughput sequencing |
| Summary | Proximal spinal muscular atrophy (SMA) is an early onset, autosomal recessive motor neuron disease caused by loss of or mutation in SMN1 (survival motor neuron 1). Despite understanding the genetic basis underlying this disease, it is still not known why motor neurons (MNs) are selectively affected by the loss of the ubiquitously expressed SMN protein. Using a mouse embryonic stem cell (mESC) model for severe SMA, the RNA transcript profiles (transcriptomes) between control and severe SMA (SMN2+/+;mSmn-/-) mESC-derived MNs were compared in this study using massively parallel RNA sequencing (RNA-Seq). The MN differentiation efficiencies between control and severe SMA mESCs were similar. RNA-Seq analysis identified 3094 upregulated and 6964 downregulated transcripts in SMA mESC-derived MNs when compared against control cells. Pathway and network analysis of the differentially expressed RNA transcripts showed that pluripotency and cell proliferation transcripts were significantly increased in SMA MNs while transcripts related to neuronal development and activity were reduced. The differential expression of selected transcripts such as Crabp1, Crabp2 and Nkx2.2 was validated in a second mESC model for SMA as well as in the spinal cords of low copy SMN2 severe SMA mice. Furthermore, the levels of these selected transcripts were restored in high copy SMN2 rescue mouse spinal cords when compared against low copy SMN2 severe SMA mice. These findings suggest that SMN deficiency affects processes critical for normal development and maintenance of MNs. |
| Overall design | RNA profiles were generated from FACS-purified control and SMA mESC-derived motor neurons (n=3/genotype) by deep sequencing using Illumina HighSeq 2500 |
| Contributor(s) | Butchbach ME |
| Citation(s) | Maeda M, Harris AW, Kingham BF, Lumpkin CJ et al. Transcriptome profiling of spinal muscular atrophy motor neurons derived from mouse embryonic stem cells. *PLoS One* 2014;9(9):e106818. PMID: 25191843 |
| Submission date | Mar 27, 2014 |

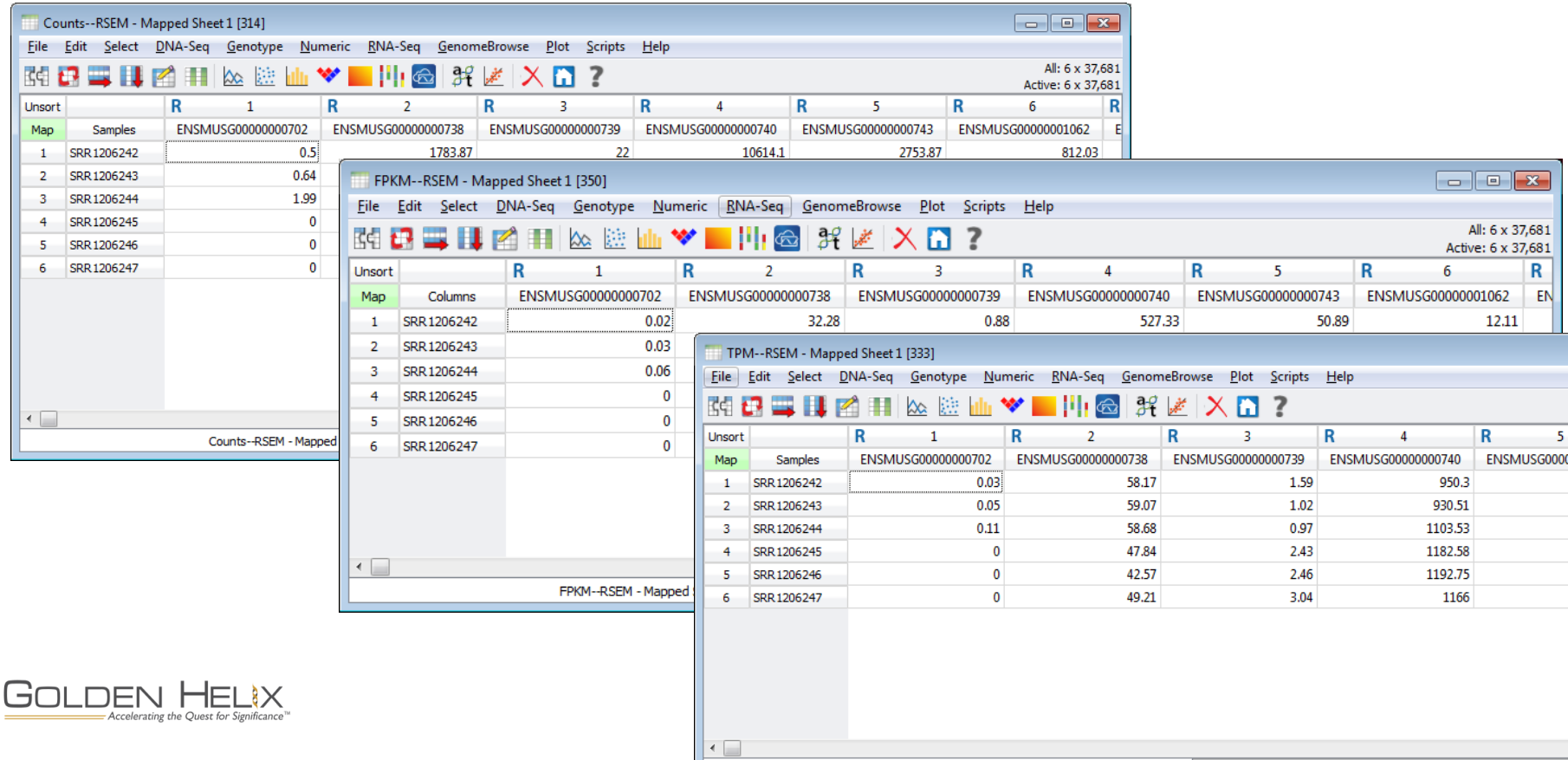
# Data Processing




Both alignments based on mm9/NCBIm37 genome and Ensembl v65 transcripts

# Dataset

- A case/control study with 3 cases knocked-out for the gene (SMN) that causes the SMA phenotype (A2 group).
- 3 control mice were used for comparison (Hb9 group)

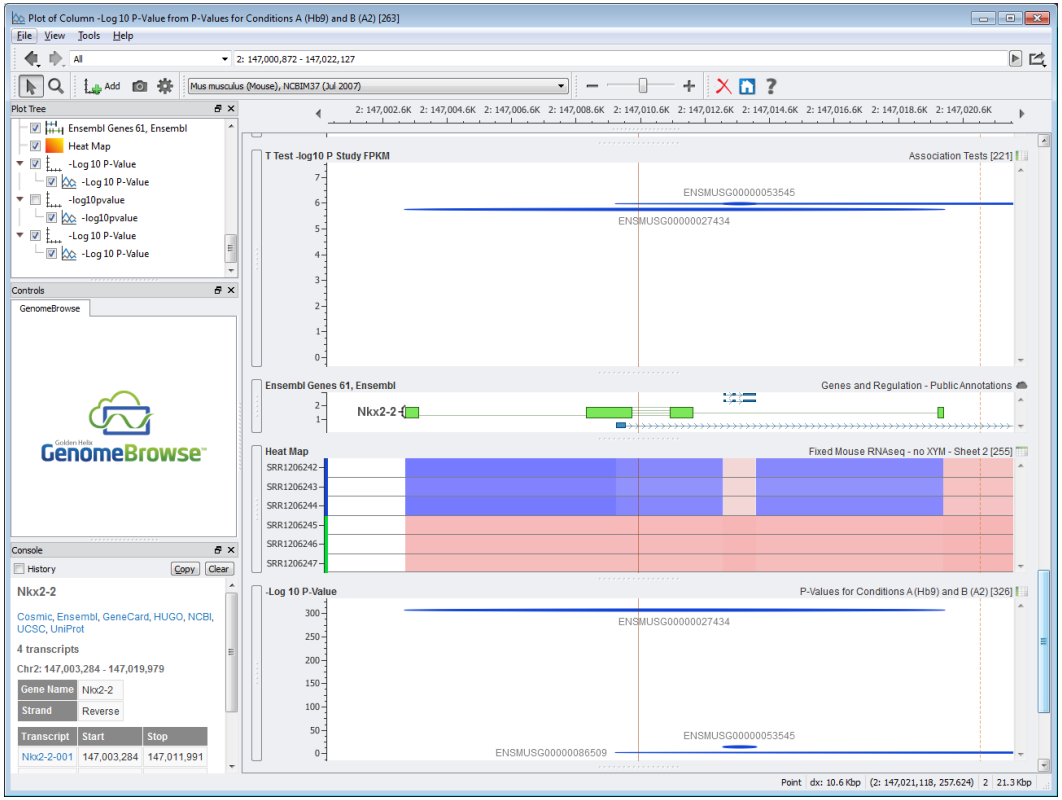


# Agenda

1. RNASeq Introduction
2. GEO Dataset
3. Demonstration
4. Conclusions

# Demonstration

- Importing RNAseq data
- Quality Assurance and Sample Statistics
- PCA
- Association Testing
- DESeq with Counts and FPKM
- Heat Map
- Visualizing Results in Genome Browse



GOLDEN HELIX

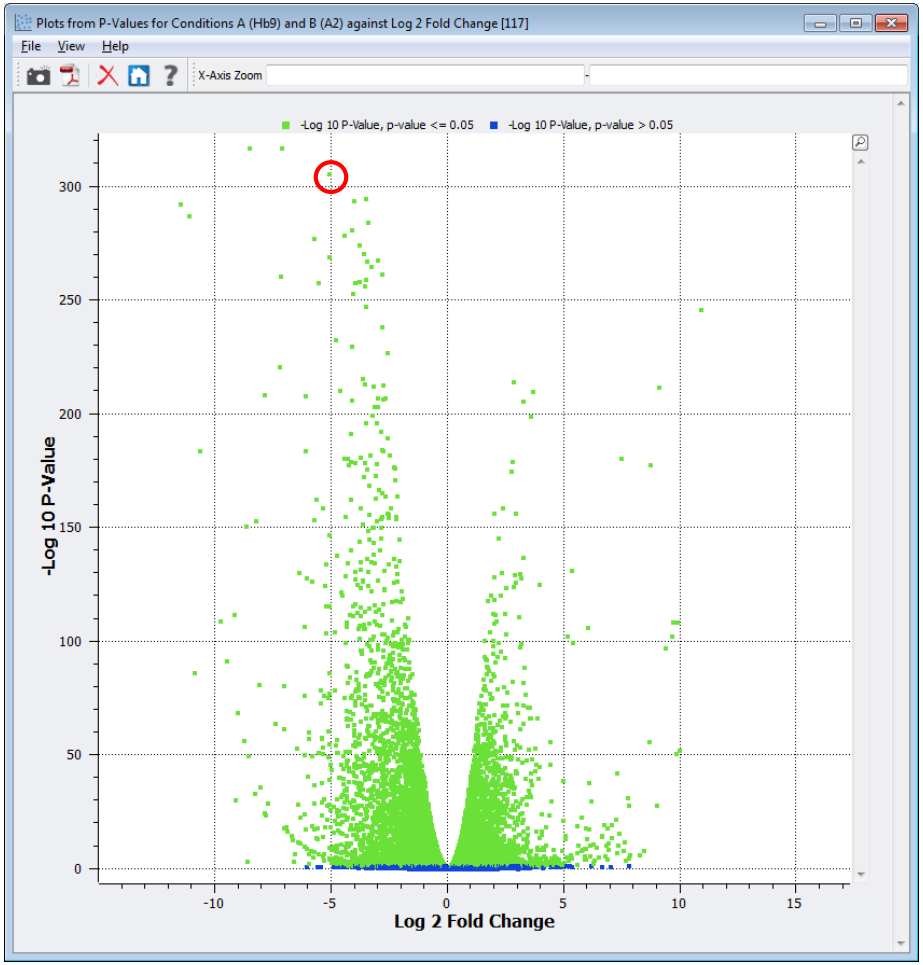
# SNP & VARIATION SUITE

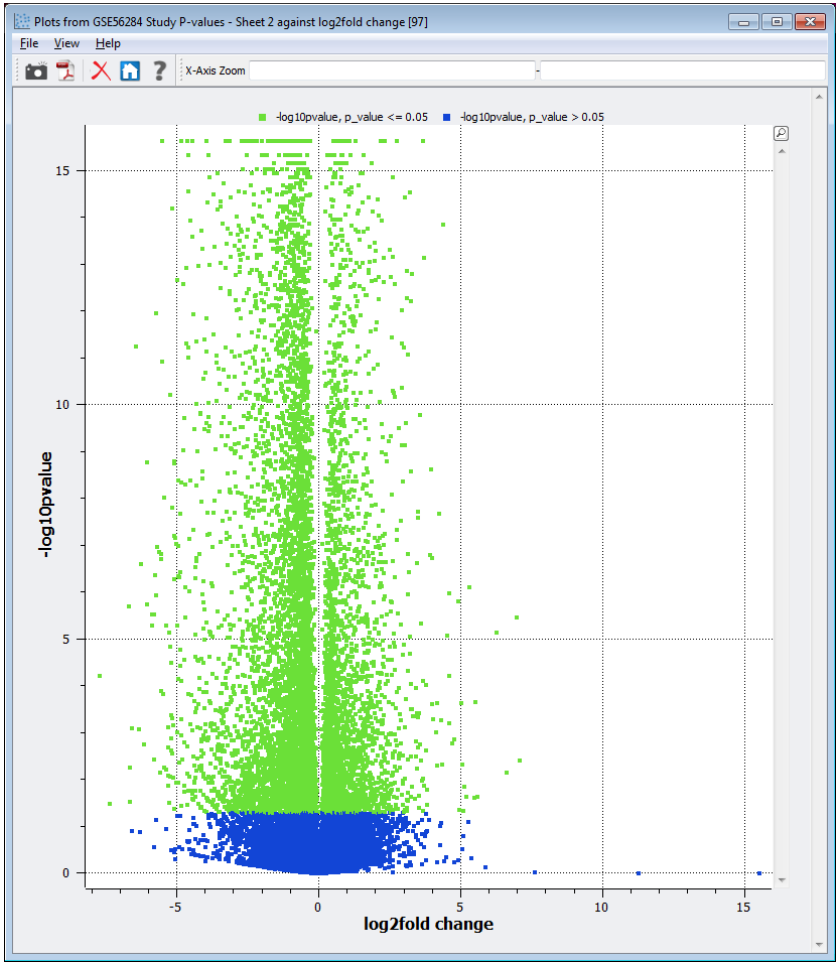
[Demonstration]

# Agenda

1. RNASeq Introduction
2. GEO Dataset
3. Demonstration
4. Conclusions

# Conclusions



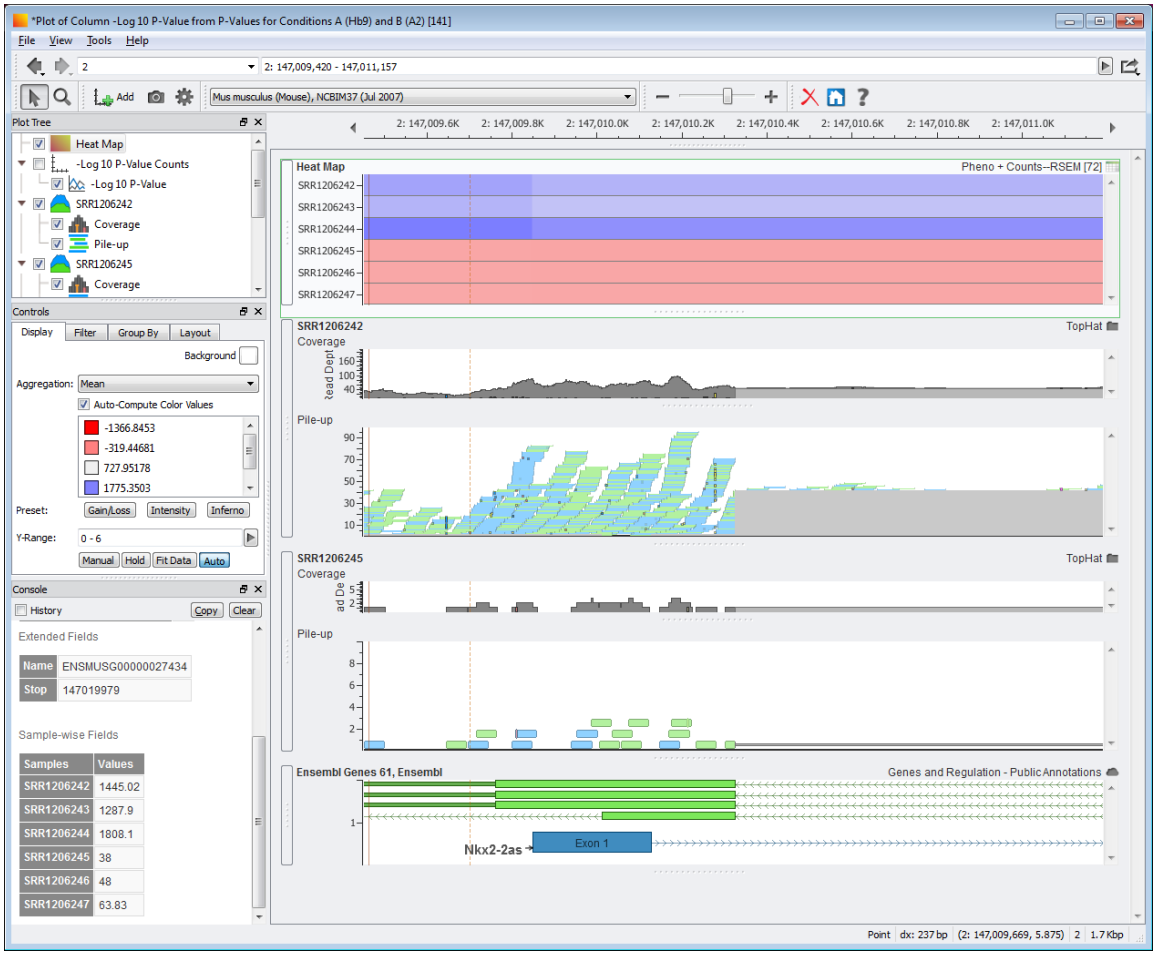
GOLDEN HELIX

# SNP & VARIATION SUITE

[Demonstration]

# Conclusions

- DESeq in SVS is best preformed with Count data
- FPKM, TPM data can be analyzed with Association Testing but is not ideal of DESeq in SVS



# Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com