

# Module 3: Introduction

## Module 3



## Overview

This module is all about the differential expression pipeline, the series of steps we take to determine genes whose expression uniquely characterizes a specific sample group. In our case, we are interested in genes that are expressed at significantly different levels in female versus male placenta samples. There are many methods for determining differentially expressed genes, but we are going to use the limma-voom pipeline chosen for the placenta sex differences manuscript (Olney et al 2022) so that we can contextualize the effect trimming using previous work. In this module, you will learn about the statistics and visualization techniques used to determine and present differential gene expression between two sample groups first conceptually and then hands-on by running the R Markdown for differential expression on the untrimmed data. Pay attention because you will be running this pipeline on a trimmed data set in the next module.

If you have questions as you go through the learning materials, please post in the course Slack channel.

## Learning Objectives

By the end of this module, you'll be able to:

### 3.1 Biology/Stats (2 hours)

1. Describe how linear modeling can be used to identify differentially expressed genes
2. Describe how to visualize sample groupings by gene expression profiles
3. Describe why we conduct transformation and normalization of RNAseq data

4. Describe the effects of trimming on data

### **3.2 Coding** (2+ hours, dep on coding experience)

1. Run Rmd of differential expression analysis for default (untrimmed data)
2. Visualize gene expression differences with volcano plots and box plots
3. Visualize trends in samples based on their gene expression profiles with MDS plots

### **3.3 Research/Professional development** (30 min)

1. Identify ways debug errors in your code