

Comparing Cost-Effective Gene Expression Phenotyping Methods In Cattle

By: Ruwaa Mohamed¹

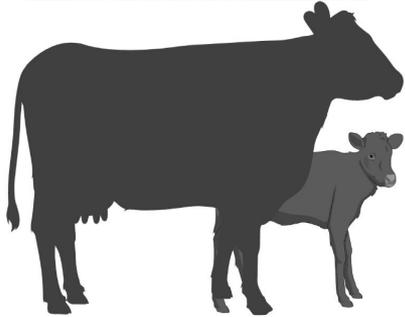
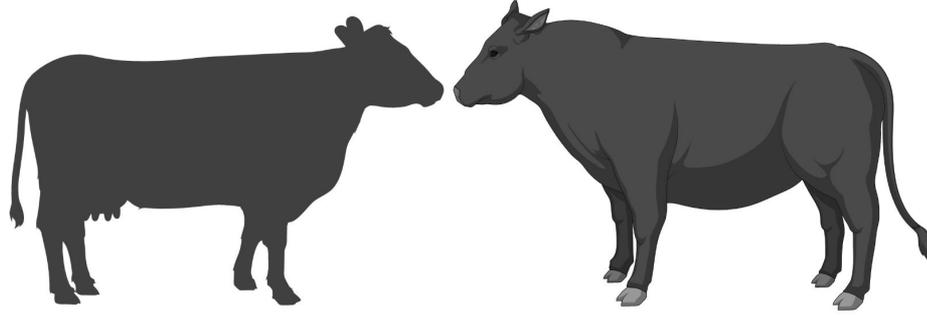
Supervisor: Dr. Troy Rowan²

¹Genome Science and Technology - UTORII,

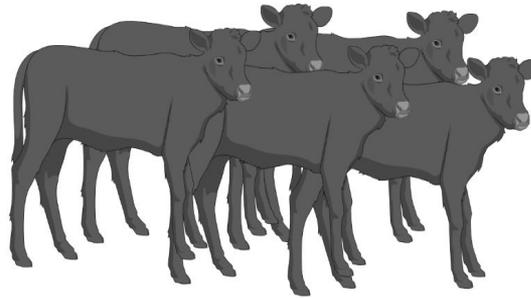
²Animal Science Department, UTK



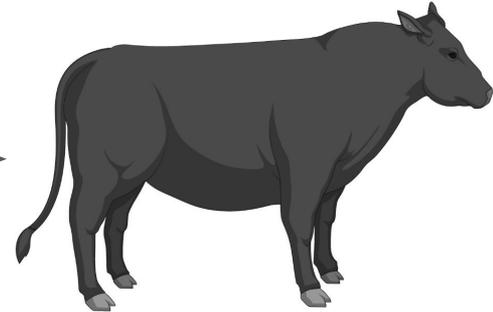
Calf Life Cycle (Beef Production)



1) Cow/Calf



2) Stocker



3) Feedlot & Slaughter house ²

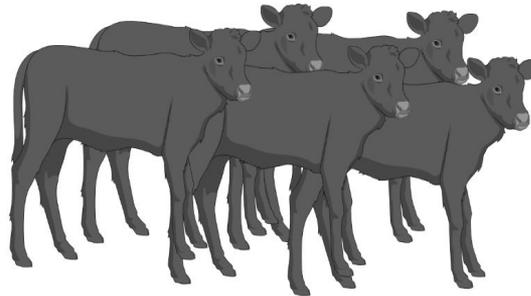
Which one(s) to pick? Desirable traits?

Easily-measured traits

breed, coat color, frame score, height, weaning weight, horns, ... etc.

Complex traits

meat quality, methane emission, susceptibility to diseases, ...etc.



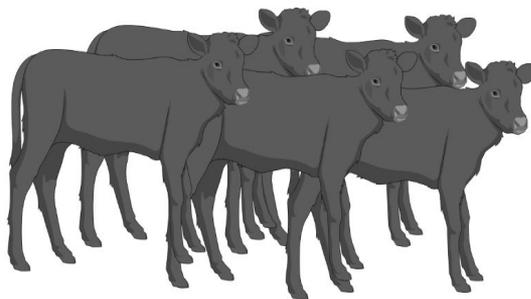
2) Stocker

Which one(s) to pick? Desirable traits?

**Bovine Respiratory
Diseases (RBD)**

\$\$\$\$

Direct cost: \$54.12 million
Indirect cost = \$5 billion
(Annually)



2) Stocker

Complex traits

meat quality, methane
emission, susceptibility to
diseases, ...etc.

**Phenotyping by gene expression
Transcriptome
(Sequencing cost)**

\$\$\$

Seed Grant Proposal

Developing a cost-effective method for collecting informative, population-level molecular phenotypes

Troy Rowan, Jon Beever, Kurt Lamour, Liesel Schneider

[Download Narrative](#) 

We propose using a sub-\$5 targeted gene expression approach as a molecular phenotype in beef cattle. This proposal aims to computationally identify 500 high-information genes that will be assayed in ~1,500 beef stocker calves. We will explore the utility of using these expression counts to predict future calf performance & health outcomes and as a latent phenotype. While we focus on beef cattle, we expect that this technology could be applied across species and genome to phenome applications.

Grant Goal: Find high information genes to perform cheap **targeted sequencing**.
(find the hub genes)

1) Gene Co-expression cluster analysis of secondary data (FAANG database) ~ WGCNA R-package

2) Primary data analysis ~ Heat stress study design (Dr. Jon Beever).

```
graph TD; Root[ ] --> Box1[1) Gene Co-expression cluster analysis of secondary data (FAANG database) ~ WGCNA R-package]; Root --> Box2[2) Primary data analysis ~ Heat stress study design (Dr. Jon Beever).]; Box1 --> List1[• Not enough variation to create 500 separate clusters.  
• Insufficient meta-data  
• Accuracy measures?]; Box2 --> List2[• Problematic targeted sequencing library prep with a lot of off-target hits  
• 3' RNA-seq data available];
```

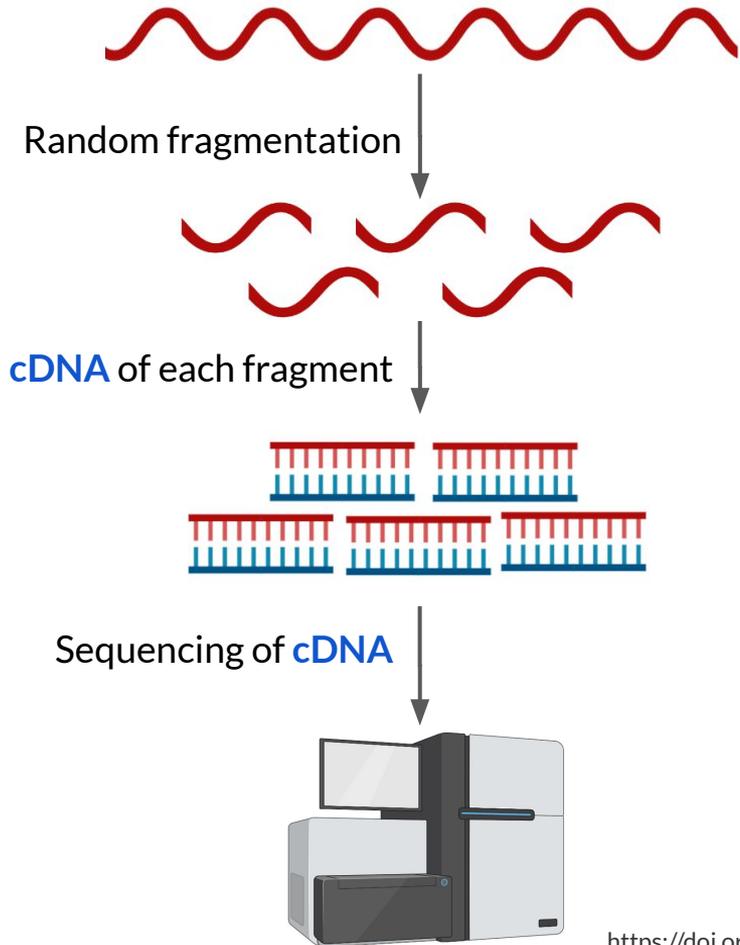
1) Gene Co-expression cluster analysis of secondary data (FAANG database) ~ WGCNA R-package

- Not enough variation to create 500 separate clusters.
- Insufficient meta-data
- Accuracy measures?

2) Primary data analysis ~ Heat stress study design (Dr. Jon Beever).

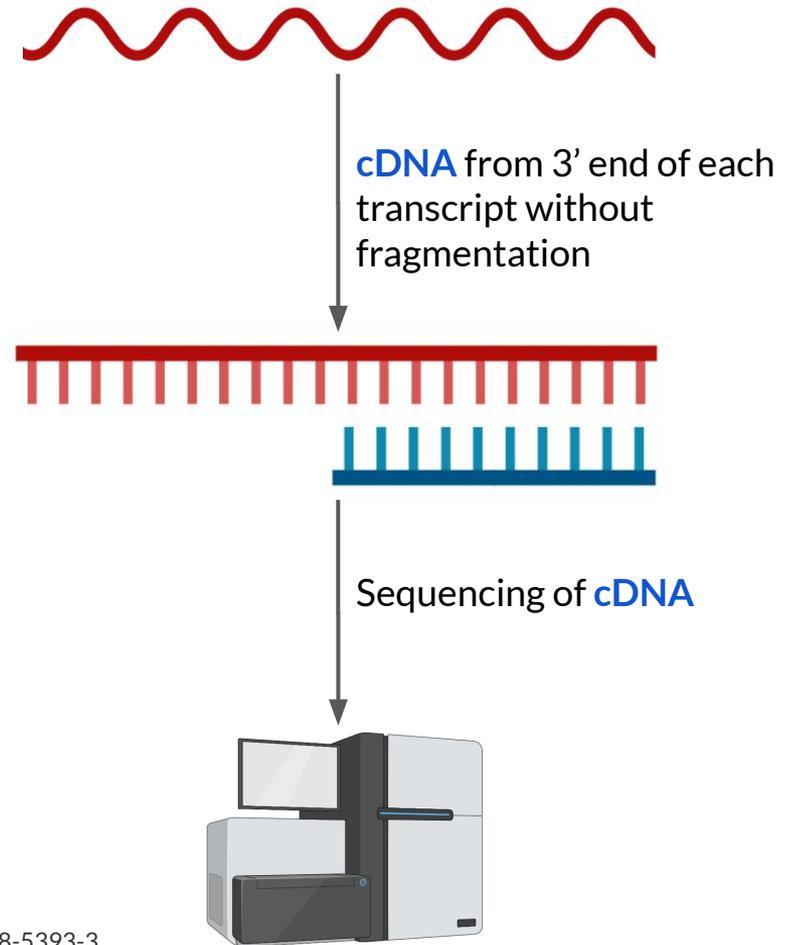
- Problematic targeted sequencing library prep with a lot of off-target hits
- 3' RNA-seq data available

Total RNA-Seq



3' RNA-Seq Technology

mRNA



Grant Goal: Find high information genes to perform cheap **targeted sequencing**.
(find the hub genes)

New Goal: Comparing Cost-Effective Gene
Expression Phenotyping Methods In Cattle

(3' RNA-Seq vs. Targeted RNA-Seq)

Heat Stress Study Design (78 Samples ~ 31 Calves)

Study Design

Day 1  Day 5

	Heat Stress	No Heat
Saline	Morning & Night Samples	Morning & Night Samples
Dexa-methasone	Morning & Night Samples	Morning & Night Samples

	Heat Stress	No Heat
Saline	Morning & Night Samples	Morning & Night Samples
Dexa-methasone	Morning & Night Samples	Morning & Night Samples

1

3' RNA-Seq & Targeted Sequencing Gene Expression Methods Identify Similar Patterns of Expression

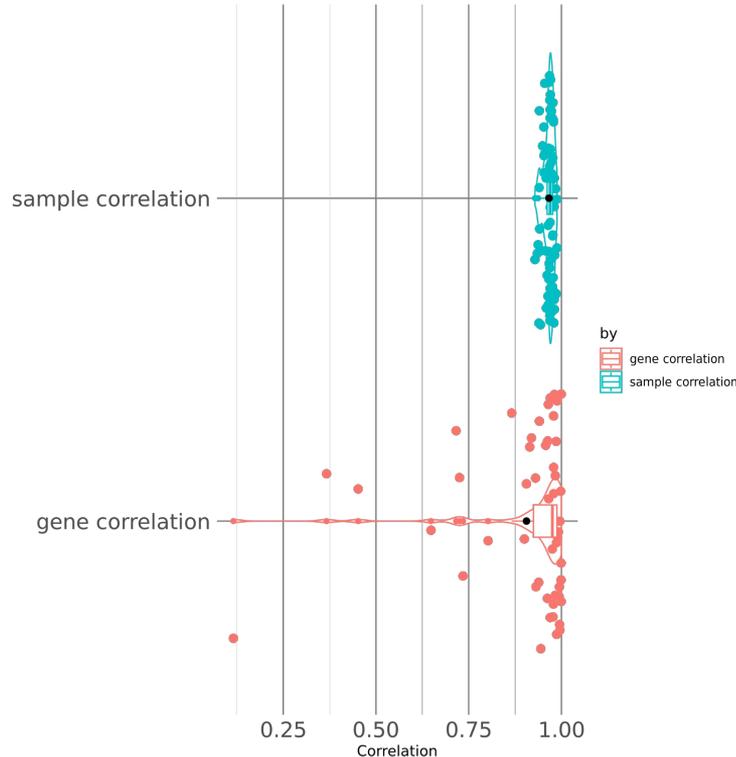


Principal component analysis (PCA) of selected 51 genes from 78 samples shows that 3' RNA-sequencing sample IDs match targeted sequencing sample IDs. The variance explained by each PC is similar between both sequencing methods.

2

Per Gene & Per Individual Gene Expression Correlations Are High

	Per Sample	Per Gene
Min.	0.93	0.12
1st Qu.	0.96	0.92
Median	0.97	0.98
Mean	0.97	0.91
3rd Qu.	0.98	0.99
Max	0.99	1.00



Correlation analysis on a per-sample basis shows a high correlation between both sequencing methods with a mean r^2 of 0.97.

High correlation is observed to on a per-gene basis with a mean r^2 of 0.91 with some exceptions.

Both sequencing methods are equally capable of representing the transcription of selected genes.

3 3` RNA-Seq has Promising Variant Calling Results

	SNPs	InDels
Min.	25949	1754
1st Qu.	56936	4752
Median	69257	5646
Mean	70665	5636
3rd Qu.	84526	6562
Max	120930	8861

Variant calling on 3` RNA-seq data is promising as shown in the density plots. The mean raw variants called per sample are 70,000 SNPs and 5,600 InDels. 3` RNA-sequencing could be used to call variants and used for imputing variants across the whole transcriptome.

4

Differential Gene Expression Analysis

Day 1  Day 5

	Heat Stress	No Heat
Saline	Morning & Night Samples	Morning & Night Samples
Dexa-methasone	Morning & Night Samples	Morning & Night Samples

	Heat Stress	No Heat
Saline	Morning & Night Samples	Morning & Night Samples
Dexa-methasone	Morning & Night Samples	Morning & Night Samples

4

Differential Gene Expression Analysis

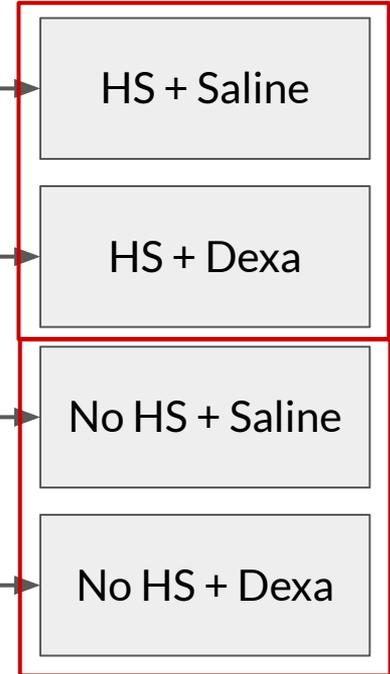
Day 1

	Heat Stress	No Heat
Saline	Morning & Night Samples	Morning & Night Samples
Dexa-methasone	Morning & Night Samples	Morning & Night Samples

Morning Samples

Control (HS and No HS; Saline and Dexa)

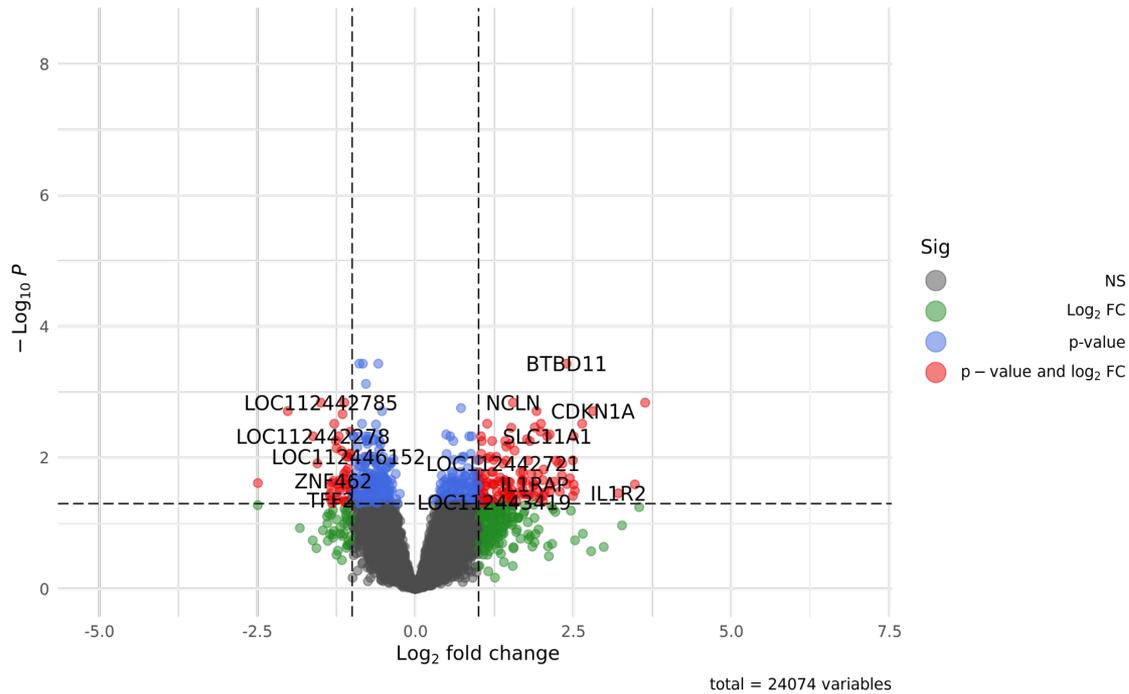
Night Samples



4

Differential Gene Expression Analysis of 3` RNA-Seq Is More Inclusive Than Targeted Sequencing

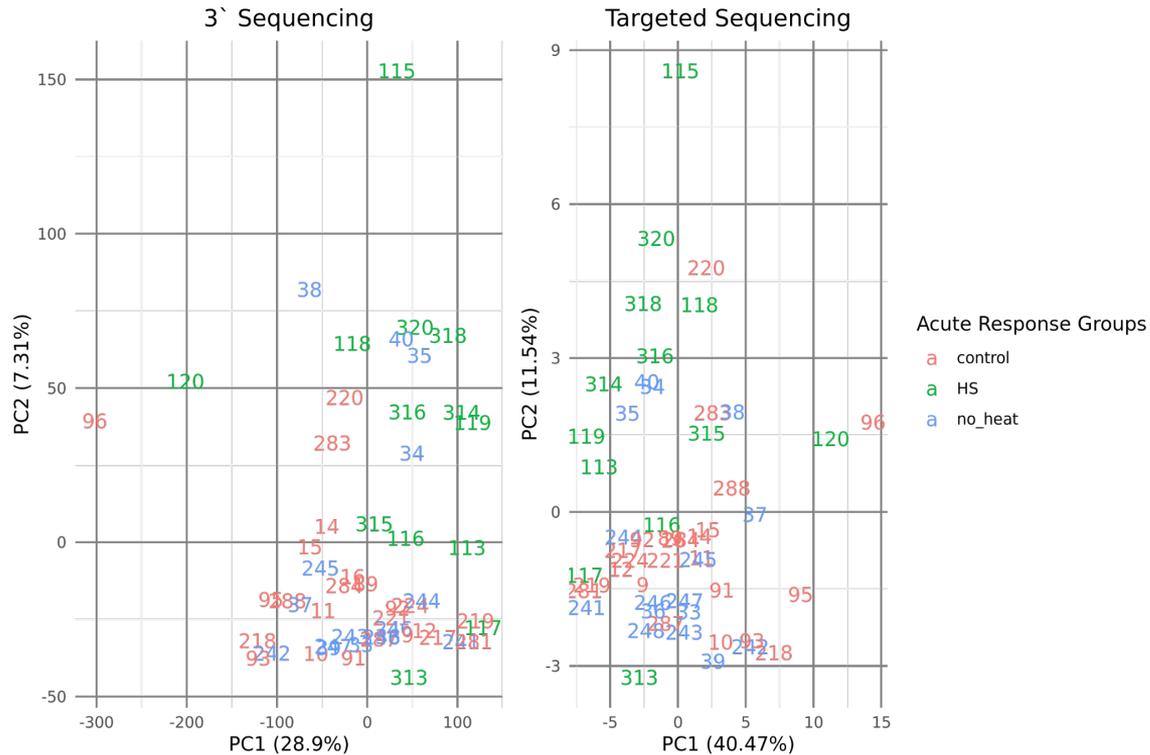
All Differentially Expressed Genes Using 3` RNA-Sequencing
EnhancedVolcano



Out of 24,074 expressed genes captured via 3` RNA-Seq, acute response to heat stress results in up-regulation of 256 and down-regulation of 264 genes with $\text{adj.p-val} < 0.05$. Only 2 of those were captured by the targeted sequencing approach (B2M, HSF2)

5

3` RNA-Seq Captures More Information than Targeted Sequencing



PCA of all captured 24,074 genes from 3' RNA-Seq samples shows the acute response of cattle to heat stress (HS) and Dexamethasone (Dexa) treatment.

Conclusion

	Targeted RNA-Sequencing	3` RNA-Sequencing	Total RNA-Sequencing
Transcript Quantification	Up to 500 transcripts	Whole transcriptome	Whole transcriptome
Variant Calling	Only for selected transcripts	3` Biased Variants	Whole transcriptome
Alternative Splicing	✓ (selected genes)	✗	✓
Cost	\$	\$\$	\$\$\$\$

Future Work

400 Calves (Samples to be sequenced + Phenotypic data).

Model for selection stocker calves.



Thank you!

Dr. Troy Rowan

GST family and friends

Dr. Jon Beever

GST colloquium attendees

Dr. Sonia Moisa

Rowan's lab

Claire E. Hunkler

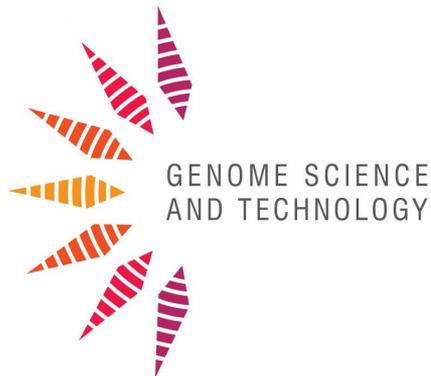
Garrett Franklin

Cassidy Catrett

Mohamed Mahrous

Animal Science Graduate Students

Madison Henniger



GRC conference, Feb. 2023, Ventura, CA, USA

Questions...?