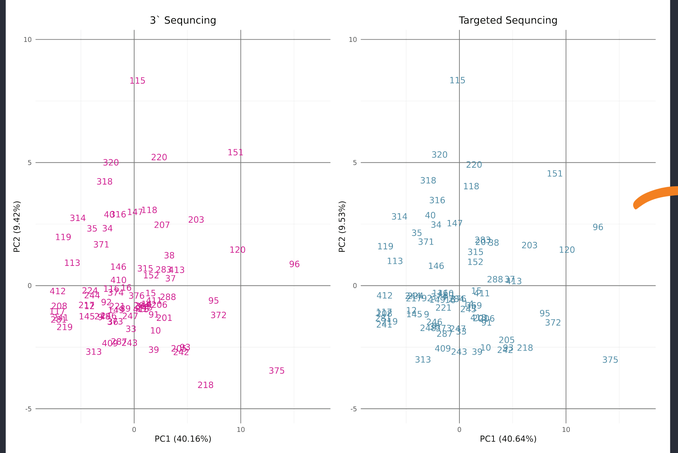
COMPARING COST-EFFECTIVE GENEEXPRESSION PHENOTYPING METHODS IN CATTLE

R Mohamed, JE Beever, S Moisa, TN Rowan

ABSTRACT

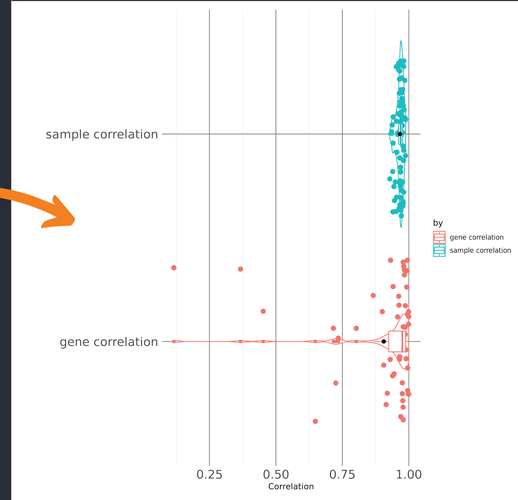
Genetic and genomic selection programs require large numbers of phenotypes observed for animals in shared environments. Direct measurements of phenotypes like meat quality, methane emission, and disease susceptibility are difficult and expensive to measure at scale but are critically important to beef production. Our work leans on our understanding of the "Central Dogma" of molecular genetics to use molecular intermediates as cheaply-measured proxies of organism-level phenotypes. The rapidly-declining cost of next-generation sequencing presents opportunities for population-level molecular phenotyping. While the cost of whole transcriptome sequencing has declined recently, its required depth of sequencing makes it an expensive choice for wide-scale molecular phenotyping. Our work aims to use and optimize two innovative gene expression quantification methods for use in creating proxy phenotypes for beef cattle from easy-to-collect tissue samples (i.e., whole blood). We compare two cost-effective methods for measuring gene expression: 3′ RNA-seq that sequences only the 3′ end of transcripts, and a targeted sequencing approach that allows for ultra-cheap expression quantification for a subset of up to 500 genes. Using 78 samples with both 3′ and targeted RNA-seq data we identified high correlations between expression counts for both methods (mean r^2 for the 52 genes = 0.906 and r^2 for 78 samples = 0.967). Further, we were able to call an average of 70,000 SNPs and 5,600 indels per sample from 3′ RNA-seq reads. This encouraging result suggests that 3' RNA-sequencing may have dual utility in genotyping when used in conjunction with imputation. Ongoing work is exploring the utility of these two approaches in performing phenotypic prediction of disease incidence in a population of stocker calves (n = 400). We anticipate numerous opportunities to leverage these developing technologies to assay intermediate molecular phenotypes across quantitative genetics and genome-to-phenome applications.

01 3` RNA-Seq & Targeted Gene Expression Methods Identify Similar Patterns of Experssion



Principle 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.

02 Per Gene & Per Individual Gene Expression Correlations Are High

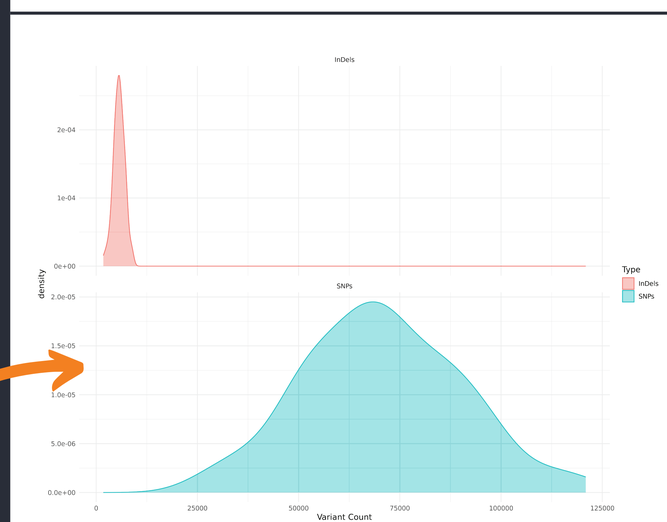


Correlation analysis on a per-sample basis shows a high correlation between both sequencing methods with a mean r^2 of 0.97. High correlation is observed to on a per-gene basis with a mean r^2 of 0.91 with some exceptions. Both sequencing methods are equally capable of representing the transcription of selected genes.

Summary of Results

	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

03 3` RNA-Seq Has promising Variant Calling Results

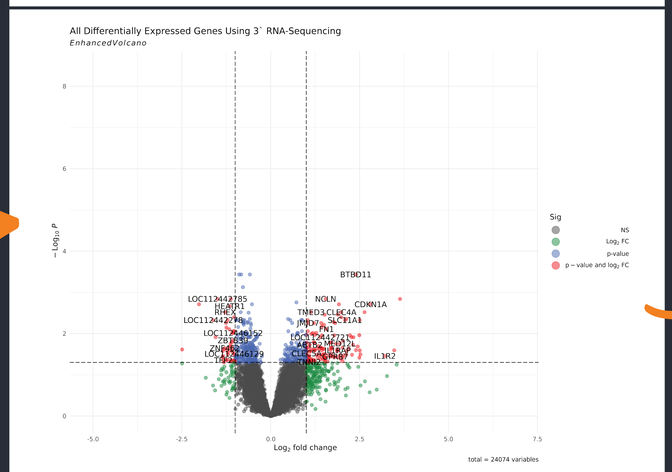


Variant calling on 3` RNA-seq data is promising as shown in the above density plots. The mean raw variants called per sample are 70,000 SNPs and 5,600 InDels. 3` RNA-sequencing could be used to call variants and used for imputing variants across the whole transcriptome.

Summary of Results

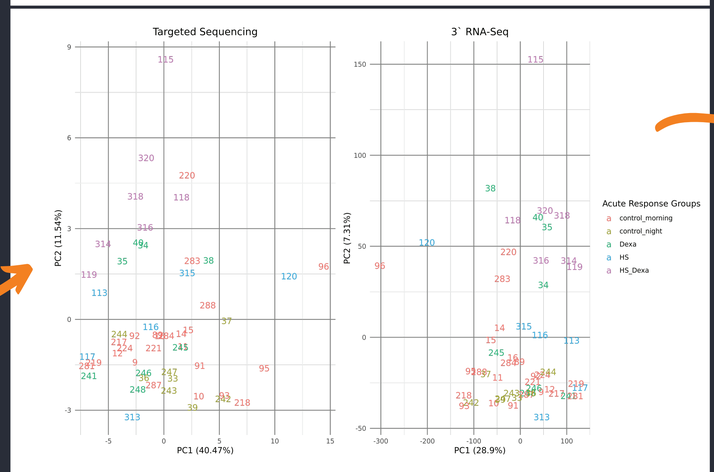
	SNPs	InDels
Min.	25949	1754
1st Qu.	56936	4752
Median	69257	5646
Mean	70665	5636
3rd Qu.	84526	6562
Max	120930	8861

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



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 adj.p-val < 0.05. Only 2 of the those were captured by the targeted sequencing approach (B2M, HSF2)

05 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 Sequencing	3` RNA-Sequencing	Whole Transcriptome Sequencing
Transcript Quantification	Up to 500 transcripts	Whole transcriptome	Whole transcriptome
Variant Calling	Only for selected transcripts	3` Biased Variants	Whole transcriptome
Cost	$	$$	$$$$$

UTIA INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE | Real. Life. Solutions.™

AG2PI Agricultural Genome to Phenome Initiative

The AG2PI project is supported by USDA-NIFA awards 2020-70412-32615 and 2021-70412-35233