



## OPEN ACCESS

EDITED BY  
Tad Stewart Sonstegard,  
Acceligen, United States

REVIEWED BY  
Christopher K. Tuggle,  
Iowa State University, United States  
Nathaniel J. Szweczyk,  
Ohio University, United States

\*CORRESPONDENCE  
Brittney N. Keel,  
brittney.keel@usda.gov

SPECIALTY SECTION  
This article was submitted to Livestock  
Genomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 30 June 2022  
ACCEPTED 24 August 2022  
PUBLISHED 19 September 2022

CITATION  
Keel BN and Lindholm-Perry AK (2022),  
Recent developments and future  
directions in meta-analysis of  
differential gene expression in  
livestock RNA-Seq.  
*Front. Genet.* 13:983043.  
doi: 10.3389/fgene.2022.983043

COPYRIGHT  
© 2022 Keel and Lindholm-Perry. This is  
an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Recent developments and future directions in meta-analysis of differential gene expression in livestock RNA-Seq

Brittney N. Keel\* and Amanda K. Lindholm-Perry

USDA-ARS, Roman L. Hruska U.S. Meat Animal Research Center, Clay Center, NE, United States

Decreases in the costs of high-throughput sequencing technologies have led to continually increasing numbers of livestock RNA-Seq studies in the last decade. Although the number of studies has increased dramatically, most livestock RNA-Seq experiments are limited by cost to a small number of biological replicates. Meta-analysis procedures can be used to integrate and jointly analyze data from multiple independent studies. Meta-analyses increase the sample size, which in turn increase both statistical power and robustness of the results. In this work, we discuss cutting edge approaches to combining results from multiple independent RNA-Seq studies to improve livestock transcriptomics research. We review currently published RNA-Seq meta-analyses in livestock, describe many of the key issues specific to RNA-Seq meta-analysis in livestock species, and discuss future perspectives.

## KEYWORDS

RNA-seq, meta-analysis, livestock, *p*-value combination, gene expression

## 1 Introduction

The development of RNA-sequencing (RNA-Seq) as a high-throughput gene expression quantification technology has been crucial in the advancement of livestock genomics research. In the last decade, the number of annually published livestock transcriptome studies has nearly tripled (Figure 1). For most of these studies, raw sequence data is stored in public repositories, such as National Center for Biotechnology Information Sequence Read Archive (NCBI SRA). To-date, NCBI SRA houses RNA-Seq data from over 3,100 different livestock projects, comprised of approximately 59,000 samples (Figure 2). In addition, multiple consortiums, including Functional Annotation of Animal Genomes (FAANG; Giuffra et al., 2019) and Farm Animal Genotype-Tissue Expression (FarmGTEx; Liu et al., 2021), are continuing to collect, curate, and house RNA-Seq data for multiple livestock species. Data from these resources is easily accessible and can be utilized for integrated analyses to generate new knowledge and scientific findings.

Despite its widespread use, challenges remain in RNA-Seq data analysis. One major issue that has been observed in gene expression studies, in livestock and other species, is non-reproducibility of results (Shi et al., 2008). Due to sequencing costs, livestock RNA-

Seq experiments are typically performed on a small number of biological replicates, which limits their power to detect differentially expressed genes (DEG) between experimental conditions. In addition to small numbers of replicates, inter-study variability due to technical differences (e.g., sample preparation, library protocols, batch effects) as well as biological differences (e.g., environmental, management, and genetic effects) also contributes to reproducibility issues.

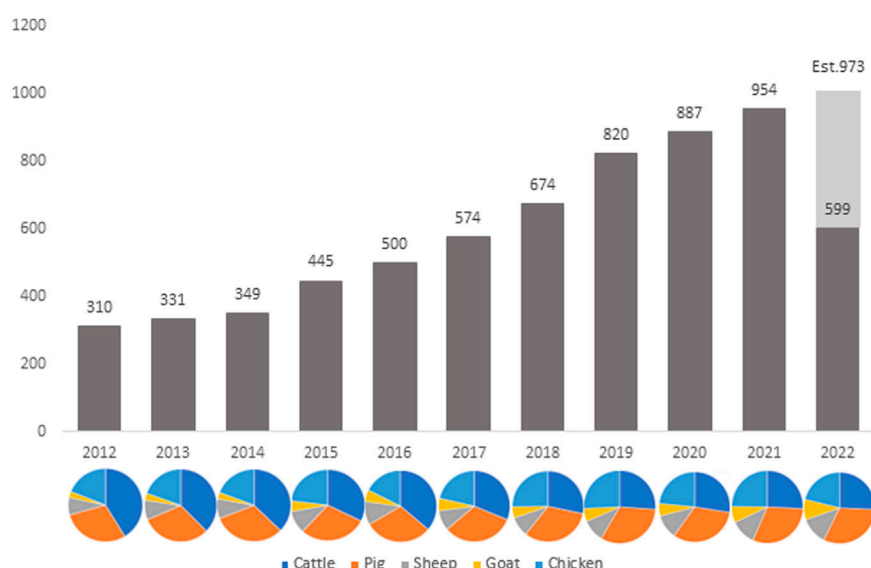
One way to improve reproducibility of RNA-Seq is by integrating data from multiple independent studies via meta-analysis. It should be noted that the term meta-analysis is an all-encompassing term, used to describe a systematic synthesis of quantitative results from different empirical studies. Hence, meta-analysis can be used to describe a large breadth of analyses with different goals. In this review, we focus specifically on meta-analysis procedures for integrating RNA-Seq data across independent studies to identify differential gene expression.

Meta-analysis procedures increase the sample size by incorporating samples from different studies, increasing both statistical power and robustness of the results. Recently, the number of publications involving differential gene expression meta-analysis has increased drastically, with most studies conducted on human samples (Toro-Domínguez et al., 2021). In humans, meta-analysis has been widely used in identification of biomarkers related cancer (Chen et al., 2014; Bell et al., 2017; Kori and Arga, 2018) and other diseases, including hypertension (Huan et al., 2015), Alzheimer's disease (Su et al., 2019; Wan et al., 2020), autoimmune disorders (Kröger et al., 2016), and schizophrenia (Piras et al., 2019).

To-date, few livestock RNA-Seq studies have utilized meta-analysis procedures to analyze differential gene expression. This is likely due to several issues, including, but not limited to, high levels of inter-study variation, availability of metadata, and limited technical guidance for conducting a meta-analysis. In this review, we discuss many of the key issues specific to RNA-Seq meta-analysis in livestock species. The first two issues are related to data acquisition and pre-processing, while the remaining issues relate to choosing a meta-analysis procedure and interpretation of results. Lastly, we discuss other points related to meta-analysis of livestock RNA-Seq data, including additional applications of meta-analysis procedures outside of differential expression analysis, new methodologies for large scale RNA-Seq, and single-cell RNA-Seq.

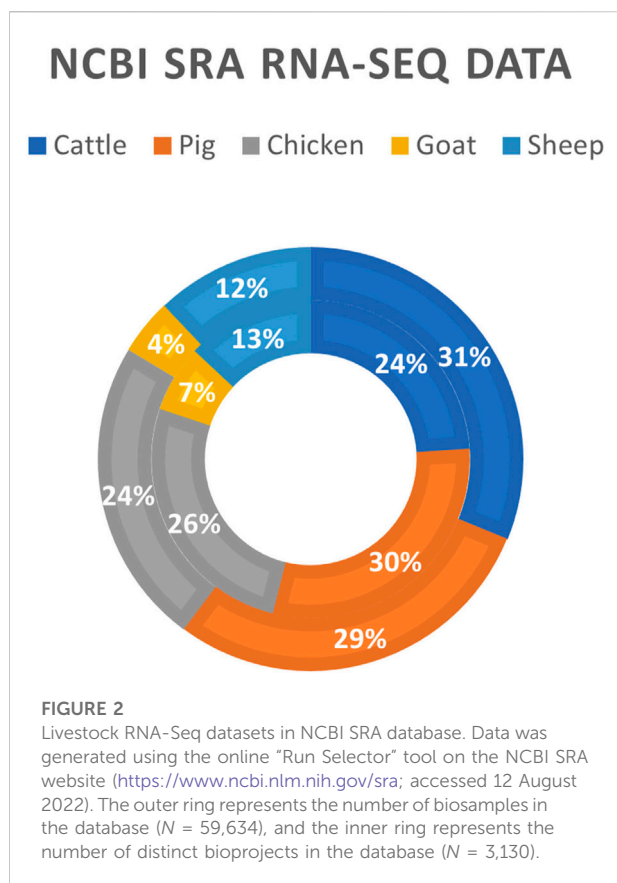
## 2 Applications of RNA-Seq differential gene expression meta-analysis in livestock

As described by Toro-Domínguez et al. (2021), a wide range of contexts exist for meta-analysis of gene expression data. In this work, we will focus on the use of meta-analysis of RNA-Seq data across multiple studies with the same phenotypic groups in order to increase the statistical power to detect genes showing consistent differences between groups. To-date, four livestock studies have reported the use of RNA-Seq meta-analysis procedures for differential expression analysis, all of which are cattle studies (Keel et al., 2018; Lindholm-Perry et al., 2020; Ghahramani et al., 2021; Lindholm-Perry et al., 2022). Traits of



**FIGURE 1**

Number of published livestock transcriptome studies by year since 2012. Data was compiled via PubMed search: "TRANSCRIPTOME" and "species" and "year" (accessed 12 August 2022).



interest in these studies include those related to feed efficiency and lactation.

There have been numerous transcriptome studies that aim to identify DEG related to feed efficiency in beef cattle. Minimal overlap in the results from these studies is an ongoing issue. In order to identify DEG predictions that would be more robust across the beef cattle industry, companion studies in skeletal muscle (Keel et al., 2018) and mesenteric fat (Lindholm-Perry et al., 2020) identified DEG associated with body weight gain and feed intake in beef steers across five different cohorts of crossbred animals reared at the same facility. Data used in these studies included nineteen beef breeds from the GPE population (Retallick et al., 2017), as well as both fall and spring seasons over 3 years. Several of the DEG in both skeletal muscle and mesenteric fat had been previously identified as candidate genes for feed efficiency or DEG associated with feed efficiency in livestock.

In addition to the aforementioned studies, where data from multiple cohorts reared in the same facility were utilized, meta-analysis techniques have been used to combine data across independent studies with common aim to detect DEG associated with cattle feed efficiency. Differentially expressed genes in the rumen epithelium of beef steers with high and low residual feed intake (RFI) phenotypes (Lindholm-Perry et al., 2022) were identified from a meta-analysis of two unrelated and physically

distant populations, one located in the U.S. and the other in Canada. A total of 83 DEG were identified in the meta-analysis, compared to 12 and 119 DEG in the individual U.S. and Canadian studies, respectively. Twenty of the DEG from the meta-analysis were classified as robust, meaning they passed a jackknife sensitivity test (for details on jackknife sensitivity testing see Section 4.2). These robust DEG were not identified in either of the individual analyses. Several DEG from the meta-analysis were involved in TORC2 signaling and proteasomal ubiquitin-independent protein catabolic biological processes. While gene ontology and pathway analyses of DEG in the individual studies did not identify these mechanisms, protein turnover via mTOR and ubiquitin-proteasome pathways have been identified as mechanisms involved in RFI in the rumen tissue of beef cattle in another study (Elohimy et al., 2019). This suggests that the meta-analysis approach can facilitate the discovery of more robust DEG through increased statistical power.

Transcriptomic studies related to bovine lactation have also taken advantage of meta-analysis approaches. Ghahramani et al. (2021) combined meta-analysis data from both publicly available RNA-Seq and microarray to investigate DEG associated with *E. coli* mastitis in dairy cattle. Separate meta-analyses were conducted for RNA-Seq (2 datasets) and microarray (6 datasets). A total of 360 DEG were common between the two meta-analyses. Common DEG were subjected to multiple downstream analyses, including ontology, protein-protein interaction (PPI) network analysis, and co-expression network analysis. Many of the significant biological pathways and DEG that were hubs in the PPI networks had been previously associated with mastitis in the literature, but no single study was able to identify all of them at once, indicating the meta-analysis provides a more robust biosignature.

### 3 Key issues in RNA-Seq meta-analysis workflow

In this section we discuss key steps and important considerations in the RNA-Seq meta-analysis workflow.

#### 3.1 Identifying data sets and data acquisition

Eligibility criteria for inclusion of data sets is dependent on the aim of the study. Typical biological criteria include species, breed, disease status, treatment, tissue, age, and/or phenotypic group. Public repositories, such as the NCBI SRA database, have search options to easily identify data sets linked to specific keywords and ontologies. Careful attention should be given to the metadata associated with each data set and how different parameters could affect the results of the study. For example, if the goal of a given meta-analysis is to investigate differential gene expression associated with feed efficiency in a specific tissue,

using a data set where animals have been inoculated with a bacteria or virus may not be appropriate due to feeding behavior differences associated with the disease state of the animal. In addition, diet (i.e., forage versus concentrate diets for cattle) and sampling location within a large tissue like the liver may also significantly affect gene expression.

One major issue with data in publicly available omics repositories, which is not unique to livestock, is the incompleteness of dataset descriptions, called metadata. Metadata can include many different pieces of information, including sources of data, dates of data collection, methods used, etc., (Koltes et al., 2019). Although stakeholders have come together to design and jointly endorse a concise and measurable set of guidelines for metadata, referred to as FAIR Data Principles (Wilkinson et al., 2016), the degree to which the research community follows these principles is varied. Recently, Rajesh et al. (2021) evaluated the completeness of public metadata accompanying transcriptomic data of patients with sepsis and corresponding controls (3,125 samples across 29 data sets). They found that, on average, only 65% of clinical phenotypes were reported in the publication and/or public repository, with 35% of the information thus being lost from the publication to the repository.

Metadata standards and infrastructure are crucial for meta-analysis across data sets. Under the Open Science initiative (Nosek et al., 2015), funding agencies and journals have begun to require that data used in a study be made publicly available, but in most databases the number of metadata entries required is still minimal. In order to incentivize authors to submit high-quality metadata, a prototype streamlined workflow for conversion of European Nucleotide Archive (ENA) genomic metadata into a data manuscript has been proposed (Dimitrova et al., 2021). Currently, this workflow is focused on genomic data, but future plans are to expand the workflow to include other repositories and data types.

To support submission of standardized rich metadata in animal genomics, the Functional Annotation of Animal Genomes (FAANG) consortium has developed the FAANG Data Portal (Harrison et al., 2021). The Data Portal offers open access to a wealth of data following FAIR Data Principles produced by an ever-growing number of FAANG consortia. To ensure that data submissions are of high-quality with complete metadata, the portal includes a contextual metadata validation. Changes to the metadata standards in the Data Portal can be proposed by anyone in the research community via their GitHub page. Adoption of metadata practices like those of ENA and FAANG, described herein, should be considered by other public repositories to facilitate the reuse of omics datasets.

### 3.2 Data preprocessing

Data preprocessing is an important step in the meta-analysis process. Unfortunately, there is no optimal pipeline for the

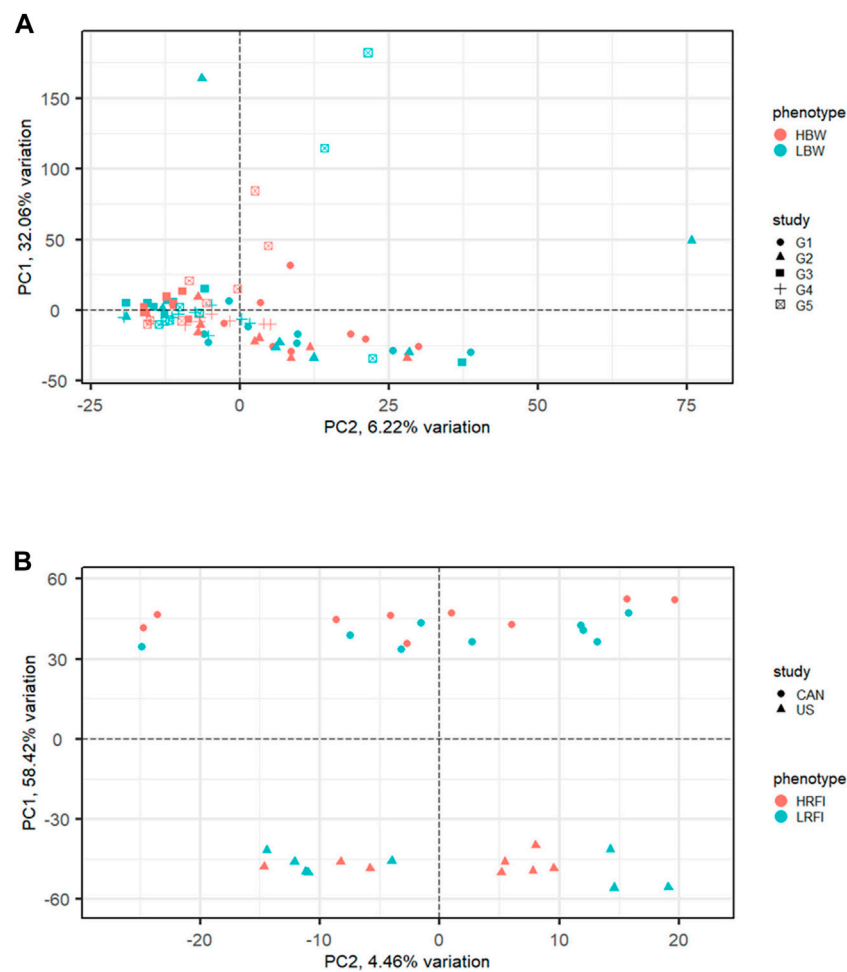
variety of different applications and analysis scenarios in which RNA-Seq can be used, and preprocessing protocols may vary greatly from study to study (Conesa et al., 2016). For meta-analysis, data preprocessing should be as standardized as possible between studies in order to minimize technical heterogeneity. Although preprocessed data is often available in public repositories, it is best to start from the raw data and process all data sets using a unified pipeline. Conesa et al. (2016) provide an extensive review of the major steps involved in processing RNA-seq data.

Standard quality control, read mapping, and quantification procedures should be performed on raw RNA-Seq data at the start of any meta-analysis (Conesa et al., 2016). Briefly, raw reads should be analyzed for sequence quality, GC content, adaptor presence, overrepresented reads, and duplicated reads. Acceptable levels of duplicated content differ by organism but should be homogeneous for samples in the same experiment. It is recommended that outliers with more than 30% disagreement be discarded (Conesa et al., 2016). After initial quality checking of raw sequence reads, reads should be trimmed to remove adaptor sequences, low-quality reads, and poor-quality bases. Trimmed reads should then be mapped to the reference genome, and libraries with low read mapping percentage (organism dependent) removed from downstream analysis. After read mapping, quantification of transcript/gene expression should be performed.

One important, often overlooked, aspect of data preprocessing for meta-analysis is how to handle data generated using different RNA library preparation protocols. During RNA-Seq library preparation, the highly abundant ribosomal RNA (rRNA), which constitute over 90% of the RNA in the cell, are removed. The two most utilized methods for rRNA removal are polyA-selection and rRNA-depletion, which generate distinct fractions of the transcriptome (Sultan et al., 2014; Bush et al., 2017). Using two ovine RNA-Seq data sets, identical except for RNA selection method, Bush et al. (2017) demonstrate that although expression levels estimated by the two methods were correlated, rRNA depleted libraries systematically produced lower estimates of the relative expression of protein-coding genes. Using a common processing pipeline, in particular a common threshold for filtering lowly expressed genes, for data sets produced using differing RNA selection methods would then result in incompatible downstream data. However, equivalent expression levels between polyA-selected and rRNA-depleted libraries can be achieved using a combination of reference transcriptome filtering and a ratio-based correction.

### 3.3 Choosing a meta-analysis procedure

Meta-analysis procedures for gene expression data have been utilized since the early 2000s, with the earliest methods being

**FIGURE 3**

Examples of RNA-Seq data with low and high inter-study variability exhibited via principal components analysis (PCA). **(A)** Low inter-study variability in the muscle transcriptome of high (HBW) and low (LBW) body weight gain from five cohorts of steers reared at the U.S. Meat Animal Research Center (Keel et al., 2018). **(B)** High inter-study variability in the rumen transcriptome of cattle with high (HRFI) and low (LRFI) residual feed intake from a Canadian population and a United States population (Lindholm-Perry et al., 2022).

proposed in the context of microarrays (Tseng et al., 2012). For microarray data, proposed meta-analysis methods include  $p$ -value combination (Marot et al., 2009), combining effect sizes (Choi et al., 2003), and ranking genes within each study (Breitling et al., 2004). Several reviews and comparisons of these methods are available (Hu et al., 2006; Hong and Breitling, 2008; Tseng et al., 2012).

Many of the meta-analysis techniques used for microarray data are not suited for RNA-Seq due to differences in the underlying structure of the data (Rau et al., 2014). Microarray data analyses utilize standard or moderated  $t$ -tests, which assume that data are continuous and can be approximated by Gaussian distributions after log-transformation (Smyth, 2004; Jaffrézic et al., 2007). RNA-Seq data, which come in the form of gene read counts, are often modeled by overdispersed Poisson (Auer

and Doerge, 2011) or negative binomial distributions (Anders and Huber, 2010; Hardcastle and Kelley, 2010; Robinson et al., 2010). Under these models, calculation and interpretation of effect sizes is not straightforward (Rau et al., 2014).

Commonly used meta-analysis procedures for RNA-Seq data include  $p$ -value combination and generalized linear models (GLM) with a fixed study effect. Performance of these methods has been linked to inter-study variability and the number of studies included in the analysis. Figure 3 illustrates examples of meta-analysis with low (Figure 3A) and high (Figure 3B) inter-study variability. Using both real data from human melanoma cells and simulated data, Rau et al. (2014) evaluated the performance of these methods for differing levels of inter-study variability, number of studies, and number of biological replicates per study. The methods considered



included per-study  $p$ -value combination using the inverse normal (Liptak, 1958) and Fisher (1932) methods and a negative binomial GLM with fixed study effect. Results from the two  $p$ -value combination techniques were nearly identical. For low inter-study variability, the results from  $p$ -value combination were very close to those of the global GLM with study effect. As the inter-study variability increased, however, the gains in performance for  $p$ -value combination were significant compared to the global GLM, particularly for the analysis of data from more than two studies. Given these results, the use of  $p$ -value combination is likely the best choice for most meta-analyses. In some cases, it may be useful to utilize the union of results from the  $p$ -value and global GLM methods. In their meta-analysis of human melanoma samples, Rau et al. (2014) found that the sets of genes uniquely identified by Fisher  $p$ -value combination and global GLM, as well as the set of genes found in common, all appeared to include genes related to cancer or melanoma.

Additional data filtering is required if using the  $p$ -value combination approach. A critical underlying assumption for the statistics in  $p$ -value combination is that  $p$ -values for all genes in the per-study differential analyses are uniformly distributed under the null hypothesis (Rau et al., 2014). This assumption is, however, not always satisfied for RNA-Seq data; in particular, a peak is often observed for  $p$ -values close to one due to the discretization of  $p$ -values for very low counts (Rau et al., 2013). To circumvent this issue, weakly expressed genes should be filtered from the analysis. Data-based methods, such as HTSFilter (Rau et al., 2013), are preferred over ad-hoc filtering procedures, as they can account for differences between studies resulting from sequencing depth, intra-condition variability, and other technical factors.

## 4 Interpreting meta-analysis results

In this section, we discuss key considerations for interpreting the results of RNA-Seq meta-analysis.

### 4.1 Fold changes

It is not uncommon for genes in a meta-analysis to exhibit conflicting expression patterns among studies. For example, Lindholm-Perry et al. (2022) identified nearly 50% of rumen epithelial DEG (37 of 83 DEG) as discordant in their direction of expression between two unrelated, geographically distant populations of cattle. A similar phenomenon was observed in meta-analyses of muscle and mesenteric fat tissues across five cohorts of related cattle that were reared in a common facility (Keel et al., 2018; Lindholm-Perry et al., 2020).

Fold changes must be handled differently for global GLM and  $p$ -value combination techniques. For global GLM with study

effect, fold changes are readily calculated using software such as DESeq2 (Love et al., 2014) and edgeR (Robinson et al., 2010). These fold changes have the same interpretation as those in single study analyses, as they are computed by averaging gene expression values across the conditions being compared. On the other hand, fold changes for the  $p$ -value combination are calculated on a per-study basis and can exhibit differing patterns between studies.

In the context of microarrays, approaches for overcoming conflicting expression patterns between studies for the inverse normal (Marot et al., 2009) and Fisher (Owen, 2009)  $p$ -value combination methods have been proposed. Both methods rely on the use of a two-tailed gene expression distribution, where under- and over-expressed genes reside in the tails of the distribution. However, for RNA-Seq data, where the data follows an overdispersed Poisson or negative binomial distribution, under- and over-expressed genes cannot be separated into distribution tails.

It has been suggested that DEG with differing expression patterns between studies be removed post hoc (Rau et al., 2014). However, genes with both concordant and discordant gene expression patterns across studies could be of interest. For example, sample collection time points will often be different between studies. The expression of some genes may be higher at earlier sampling time points and lower at later sampling time points. Hence, these genes would still be of biological interest and should be considered. Recently, there have been methodologies proposed for identifying and exploring discordant DEG sets (Lai et al., 2016; Ye et al., 2021), but additional research is needed to improve our understanding of underlying causes of these differences and how they contribute to the mechanisms of complex phenotypes.

### 4.2 Robustness of differentially expressed genes in $p$ -value combination

Differential expression results, especially those derived from small sample sizes, are known to be susceptible to heterogeneity (Cui et al., 2021). As a result, reproducibility of DEG is often poor from study to study. For this reason, robustness of DEG arising from meta-analysis should be measured by replication validity rather than in independent data, such as cross-validation (Keel et al., 2018). Jackknife sensitivity analysis has been employed to measure the robustness of meta-analysis DEG (Ch'ng et al., 2015; Keel et al., 2018; Lindholm-Perry et al., 2020; Lindholm-Perry et al., 2022).

Jackknife sensitivity analysis consists of repeating the meta-analysis procedure multiple times, each time removing a single study from the analysis (Miller, 1974). Suppose a meta-analysis consists of  $N$  studies. A gene is said to pass a jackknife sensitivity test if it is identified as a DEG in the jackknife analysis. Genes that pass all  $N$  jackknife analyses can be considered highly robust, as their statistical significance is spread across studies, i.e., they are not being driven by any one study. Genes failing multiple

jackknife studies can also be interpreted as robust, with a higher number of failed jackknife tests corresponding to greater robustness. The interpretation is derived as follows. A gene that fails only one jackknife analysis indicates that the meta  $p$ -value was being driven by the  $p$ -value of a single study, i.e., there is a significant amount of study bias. Similarly, genes that fail multiple jackknife analyses are being driven by  $p$ -values from multiple studies. This indicates a reduced level of study bias, i.e. a more robust result.

### 4.3 Downstream analysis

In RNA-Seq studies, DEG lists are subjected to several different downstream analyses with the intention of identifying gene signatures that link to the phenotype of interest. The most popular type of downstream analysis is gene set enrichment analysis (GSEA). In GSEA, the list of DEG is compared to a background gene set with known biological processes, such as gene ontology (GO) or biological pathways. An enrichment score, which indicates the degree by which a gene set is overrepresented in the list of DEG, is used to identify biological processes potentially associated with the phenotype (Subramanian et al., 2005).

There are three primary approaches to GSEA, overrepresentation analysis, functional class scoring, and pathway topology-based methods (Khatri et al., 2012). Over-representation analysis (ORA) approaches statistically evaluate the fraction of genes in a particular ontology/pathway found among the set of DEG. The most commonly used tests in ORA are based on the hypergeometric, chi-square, or binomial distribution (Huang et al., 2009). ORA methodologies only utilize lists of DEG, disregarding other quantitative measures of the genes such as fold change. In functional class scoring (FCS) methods, use a similar approach to ORA but adds in quantitative information from the genes (Mootha et al., 2003). A gene score is calculated for each gene, and individual gene scores are used to calculate a gene set score. Significance of gene set scores is assessed and differentially enriched gene sets are reported. Topology-based (TB) methods for GSEA go a step further by utilizing the network structure of a gene pathway to quantify a gene's importance to a given pathway (Draghici et al., 2007).

Currently, researchers are using the same downstream analyses for DEG sets coming from meta-analysis as are used for single studies. While ORA tools are directly applicable to meta-analysis results, current FCS and TB methodologies are not since they are designed to be used with a single set of information for each gene, e.g., a single fold change. Integration of DEG expression profiles across studies in a meta-analysis could shed light on the molecular mechanisms governing phenotypes across environments, especially for DEG exhibiting discordant fold changes across studies. To our knowledge, there are currently no bioinformatic approaches that integrate meta-analysis results in this way. Future development of algorithms to incorporate this type of data would help harness the full potential of gene expression meta-analyses.

## 5 Additional discussion points

### 5.1 Other applications of RNA-Seq meta-analysis in livestock

As mentioned in the introduction, the term meta-analysis is used to encompass many different types of analyses, all with an underlying common goal of synthesizing results across different empirical studies. In this review, we focused specifically on meta-analysis procedures for analyzing differential gene expression. In addition to this approach, meta-analysis procedures have been used to integrate analyses of functional genome information with large-scale GWAS data to discover trait- and disease-relevant tissues and cell types.

In an effort to establish connections at the RNA level between tissue/cell types and complex traits, Fang et al. (2020) uniformly assembled and analyzed, via meta-analysis techniques, over 700 bovine RNA-Seq data sets, encompassing 91 tissues and cell types. Tissue- and cell-specific genes were detected and classified in terms of their biological characteristics such as biological function, DNA methylation, and evolution. Tissue-specific genes were integrated with large scale GWAS data to detect candidate genes for 45 complex, economically relevant traits via transcriptome-wide analysis study (TWAS). This study laid the groundwork for the FarmGTE<sub>x</sub> project, which seeks to characterize tissue-specific gene expression and genetic regulation in livestock.

In the seminal manuscript from the FarmGTE<sub>x</sub> project, Liu et al. (2021) built the Cattle GTE<sub>x</sub> (<http://cgtex.roslin.ed.ac.uk/>) utilizing 11,642 bovine RNA-Seq data sets. Transcriptome landscape across over 100 tissues was described, and gene expression in different tissues was linked to 43 economically relevant traits via TWAS and colocalization analyses. Similar efforts in swine (PigGTE<sub>x</sub>) are ongoing in the FarmGTE<sub>x</sub> project (Fang, personal communication, 24 May 2022). With more transcriptomics data becoming available across diverse tissues in livestock in the near future, from projects such as FarmGTE<sub>x</sub> and FAANG (Giuffra et al., 2019), the use of this type of meta-analysis will be powerful in providing novel insights into the genetic and biological mechanisms underpinning traits and thus enhancing genomic improvement programs.

In addition to its use in integrating multiple RNA-Seq experiments with GWAS data, network meta-analysis has been utilized for differential expression analysis across independent livestock studies. Fonseca et al. (2020) identified DEG between high- and sub-fertile cows in two independent studies and utilized network meta-analysis (Winter et al., 2019) to obtain combined test-statistics for each of the genes. These test-statistics are like those described in Section 3.3. The use of network meta-analysis for RNA-Seq data sets is relatively new and, to our knowledge, has not been systematically compared to the meta-analysis procedures described in Section 3.

## 5.2 New RNA-Seq protocols to reduce costs in large-scale expression studies

There are many different options for library preparation and sequencing that users must consider when designing an RNA-Seq experiment. One of the most critical and costly steps in RNA-Seq is the construction of the cDNA library (Hou et al., 2015). In the classic whole-transcriptome approach, cDNA are generated from reverse transcribing randomly sheared fragments of the extracted mRNA. Although these approaches are generally considered unbiased, there are some subtle biases that are introduced, such as differentially expressed genes being more likely to be enriched for longer transcripts (Oshlack & Wakefield, 2009). Recently, 3' RNA-Seq have been introduced to address this bias. In addition, 3' RNA-Seq methods, such as QuantSeq (Moll et al., 2014), have been shown to significantly reduce associated financial and labor costs associated with library preparation. These cost reductions will make large-scale expression studies more feasible, especially in livestock where budget is typically the most limiting factor.

The most widely deployed 3' RNA-Seq method, QuantSeq, produces fragments for sequencing close to the 3' end of polyadenated [poly(A)] mRNA, generally from the last exon or the 3' untranslated region (UTR) (Moll et al., 2014). Total RNA is used as input with no prior poly(A) enrichment or rRNA depletion. QuantSeq differs from traditional RNA-Seq in that it sequences a smaller part of the transcript and produces only one fragment per transcript. Therefore, less sequencing is required. In fact, the QuantSeq vendor, Lexogen, recommends sequencing only 10 million reads per sample for mammalian transcriptomics (Corley et al., 2019). It has been shown that gene expression levels in QuantSeq and Illumina TruSeq are strongly correlated (Corley et al., 2019). The major limitation of QuantSeq is that it is restricted to assessing gene expression changes of poly(A) mRNA. It does not provide information regarding mutations or novel transcripts. For longer transcripts with many isoforms or non-poly(A) transcripts, traditional RNA-Seq may be more appropriate. To date, QuantSeq has been deployed in transcriptomic studies of multiple livestock species, including cattle (Pascottini et al., 2021; Beak and Baik, 2022; Busato et al., 2022; Pedroza et al., 2022), sheep (Kubik et al., 2018), chicken (Ibrahim et al., 2021) and pigs (Dong et al., 2018; Kroeske et al., 2021; Schroyen et al., 2021; Kramer et al., 2022).

One drawback to QuantSeq, as well as traditional RNA-Seq library preparation, is that the user needs to process each sample on an individual basis. To address this limitation, early multiplexing protocols, which label individual samples during the reverse transcription reaction, have been developed (Alpern et al., 2019). Once individual samples have been labeled, they are pooled, and the remainder of the library processing steps are performed in bulk. This shortens the time and cost of library preparation. Early multiplexing methodologies are available for traditional RNA-Seq library preparation (BRB-seq; Alpern et al., 2019), as well as 3' RNA-Seq library preparation (3'Pool-seq; Sholder et al., 2020).

Utilization of these approaches results in a significant reduction in cost per library for library preparation and sequencing. In fact, 3'Pool-seq can reduce the per-library cost of library preparation and sequencing from over \$160 in traditional RNA-Seq to under \$15 (Sholder et al., 2020).

The technologies mentioned above have many strengths, including reducing cost and streamlining experimental procedures. They will facilitate new opportunities for future large-scale transcriptomics studies in livestock. Future work should focus on evaluating and comparing these methodologies in terms of various robustness metrics, including gene expression quantification accuracy and DEG detection. Understanding how data generated from these approaches compares with traditional RNA-Seq data will be crucial for conducting meta-analyses across studies with differing RNA-Seq technologies.

## 6 Conclusion

Although the cost of next-generation sequencing technologies continues to decrease, livestock transcriptomic studies are often performed on a small number of samples due to financial constraints. Small sample sizes result in decreased detection power for DEG. Utilization of meta-analysis procedures to combine and analyze data across multiple related studies can increase statistical power and robustness of results. The aim of this paper was to present and discuss several key considerations for meta-analysis of livestock RNA-Seq data.

The use of meta-analysis in livestock transcriptomic data should provide identification of DEG that underly complex phenotypes as they account for some of the potentially confounding issues that may influence gene expression in a single study such as, sire lines, environment, and management. Differentially expressed genes identified from analysis across multiple populations will likely be more robust biological markers. Additional research is needed to develop techniques for downstream analysis of meta-DEG that integrate expression profiles across studies.

## Author contributions

Both authors conceived the manuscript. BK drafted the manuscript, and AP contributed to the writing and editing. Both authors read and approved the final manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alpern, D., Gardeux, V., Russeil, J., Mangeat, B., Meireles-Filho, A. C. A., Breyse, R., et al. (2019). BRB-Seq: Ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.* 20, 71. doi:10.1186/s13059-019-1671-x
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi:10.1186/gb-2010-11-10-r106
- Auer, P., and Doerge, R. (2011). A two-stage Poisson model for testing RNA-seq data. *Stat. Appl. Genet. Mol. Biol.* 10 (26), 1–26. doi:10.2202/1544-6115.1627
- Beak, S.-H., and Baik, M. (2022). Comparison of transcriptome between high- and low-marbling fineness in longissimus thoracis muscle of Korean cattle. *Anim. Biosci.* 35 (2), 196–203. doi:10.5713/ab.21.0150
- Bell, R., Barraclough, R., and Vasieva, O. (2017). Gene expression meta-analysis of potential metastatic breast cancer markers. *Curr. Mol. Med.* 17, 200–210. doi:10.2174/1566524017666170807144946
- Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: A simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573, 83–92. doi:10.1016/j.febslet.2004.07.055
- Busato, S., Ford, H. R., Abdelatty, A. M., Estill, C. T., and Bionaz, M. (2022). Peroxisome proliferator-activated receptor activation in precision-cut bovine liver slices reveals novel putative PPAR targets in periparturient dairy cows. *Front. Vet. Sci.* 9, 931264. doi:10.3389/fvets.2022.931264
- Bush, S. J., McCulloch, M. E. B., Summers, K. M., Hume, D. A., and Clark, E. L. (2017). Integration of quantitated expression estimates from polyA-selected and rRNA-depleted RNA-seq libraries. *BMC Bioinforma.* 18, 301. doi:10.1186/s12859-017-1714-9
- Chen, R., Khatri, P., Mazur, P. K., Polin, M., Zheng, Y., Vaka, D., et al. (2014). A meta-analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma. *Cancer Res.* 74 (10), 2892–2902. doi:10.1158/0008-5472.CAN-13-2775
- Ch'ng, C., Kwok, W., Rogic, S., and Pavlidis, P. (2015). Meta-analysis of gene expression in autism spectrum disorder. *Autism Res.* 8 (5), 593–608. doi:10.1002/aur.1475
- Choi, J. K., Yu, U., Kim, S., and Yoo, O. J. (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* 19 (1), 84–90. doi:10.1093/bioinformatics/btg1010
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13. doi:10.1186/s13059-016-0881-8
- Corely, S. M., Troy, N. M., Bosco, A., and Wilkins, M. R. (2019). QuantSeq. 3' sequencing combined with salmon provides a fast, reliable approach for high throughput RNA expression analysis. *Sci. Rep.* 9, 18895. doi:10.1038/s41598-019-55434-x
- Cui, W., Xue, H., Wei, L., Jin, J., Tian, X., and Wang, Q. (2021). High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis. *Hum. Genomics* 15, 7. doi:10.1186/s40246-021-00308-5
- Dimitrova, M., Meyer, R., Buttigieg, P., Georgiev, T., Zhelezov, G., Demirov, S., et al. (2021). A streamlined workflow for conversion, peer review, and publication of genomics metadata as omics data papers. *GigaScience* 10 (5), giab034. doi:10.1093/gigascience/giab034
- Dong, Q., Lunney, J. K., Nguyen, Y., Tuggle, C., Reecy, J., Rowland, R. R., et al. (2018). "The effects of vaccination and WUR genotype on blood gene expression response to co-infection with PRRSV and PCV2 in pigs," in Proceedings, World Congress of Genetics Applied to Livestock Production 11, 993.
- Draghici, S., Khatri, P., Tarca, A. L., Amin, K., Done, A., Voichita, C., et al. (2007). A systems biology approach for pathway level analysis. *Genome Res.* 17, 1537–1545. doi:10.1101/gr.6202607

## Author disclaimer

Mention of a trade name, proprietary product, or specified equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable. The USDA is an equal opportunity provider and employer.

- Elolimy, A. A., Abdel-Hamied, E., Hu, L., McCann, J. C., Shike, D. W., and Loor, J. J. (2019). Rapid Communication: Residual feed intake in beef cattle is associated with differences in protein turnover and nutrient transporters in ruminal epithelium. *J. Anim. Sci.* 97, 2181–2187. doi:10.1093/jas/skz080
- Fang, L., Cai, W., Liu, S., Canela-Xandri, O., Gao, Y., Jiang, J., et al. (2020). Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res.* 30 (5), 790–801. doi:10.1101/gr.250704.119
- Fisher, R. A. (1932). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fonseca, P. A. S., Suárez-Vega, A., and Cánovas, A. (2020). Weighted gene correlation network meta-analysis reveals functional candidate genes associated with high- and sub-fertile reproductive performance in beef cattle. *Genes* 11 (5), 543. doi:10.3390/genes11050543
- Ghahramani, N., Shodja, J., Rafat, S. A., Panahi, B., and Hasanpur, K. (2021). Integrative systems biology analysis elucidates mastitis disease underlying functional modules in dairy cattle. *Front. Genet.* 12, 712306. doi:10.3389/fgene.2021.712306
- Giuffra, E., Tuggle, C. K., and Faang, C. (2019). Functional annotation of animal genomes (FAANG): Current achievements and roadmap. *Annu. Rev. Anim. Biosci.* 7, 65–88. doi:10.1146/annurev-animal-020518-114913
- Hardcastle, T., and Kelly, K. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinforma.* 11, 422. doi:10.1186/1471-2105-11-422
- Harrison, P. W., Sokolov, A., Nayak, A., Fan, J., Zerbino, D., Cochrane, G., et al. (2021). The FAANG data portal: Global, open-access, "FAIR", and richly validated genotype to phenotype data for high-quality functional annotation of animal genomes. *Front. Genet.* 12, 639238. doi:10.3389/fgene.2021.639238
- Hong, F., and Breitling, R. (2008). A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* 24 (3), 374–382. doi:10.1093/bioinformatics/btm620
- Hou, Z., Jiang, P., Swanson, S. A., Elwell, A. L., Nguyen, B. K. S., Bolin, J. M., et al. (2015). A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Sci. Rep.* 5, 9570. doi:10.1038/srep09570
- Hu, P., Greenwood, C. M., and Beyene, J. (2006). Statistical methods for meta-analysis of microarray data: A comparative study. *Inf. Syst. Front.* 8, 9–20. doi:10.1007/s10796-005-6099-z
- Huan, T., Esko, T., Peters, M. J., Pilling, L. C., Schramm, K., Schurmann, C., et al. (2015). A meta-analysis of gene expression signatures of blood pressure and hypertension. *PLoS Genet.* 11 (3), e1005035. doi:10.1371/journal.pgen.1005035
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using David bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211
- Ibrahim, M. A. M., Nelson, J. R., Archer, G. S., and Athrey, G. (2021). Effects of monochromatic lighting during incubation and vaccination on the splenic transcriptome profiles of chicken. *Front. Genet.* 12, 628041. doi:10.3389/fgene.2021.628041
- Jaffrézic, F., Marot, G., Degrelle, S., Hue, I., and Foulley, J. L. (2007). A structural mixed model for variances in differential gene expression studies. *Genet. Res.* 89, 19–25. doi:10.1017/S0016672307008646
- Keel, B. N., Zarek, C. M., Keele, J. W., Kuehn, L. A., Snelling, W. M., Oliver, W. T., et al. (2018). RNA-Seq Meta-analysis identifies genes in skeletal muscle associated with gain and intake across a multi-season study of crossbred beef steers. *BMC Genomics* 19, 430. doi:10.1186/s12864-018-4769-8
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Comput. Biol.* 8 (2), e1002375. doi:10.1371/journal.pcbi.1002375

- Koltes, J. E., Cole, J. B., Clemmens, R., Dilger, R. N., Kramer, L. M., Lunney, J. K., et al. (2019). A vision for development and utilization of high-throughput phenotyping and big data analytics in livestock. *Front. Genet.* 10, 1197. doi:10.3389/fgene.2019.01197
- Kori, M., and Arga, K. Y. (2018). Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective. *PLoS One* 13 (7), e0200717. doi:10.1371/journal.pone.0200717
- Kramer, L. M., Teng, J., Lim, K. S., Gao, Y., Yin, H., Bai, L., et al. (2022). "Large-scale cis-eQTL analysis of gene expression in blood of young healthy pigs using PigGTEx," in Proceedings, World Congress on Genetics Applied to Livestock Production, Rotterdam, Netherlands.
- Kroeske, K., Sureda, E. A., Uerlings, J., Deforce, D., Van Nieuwerburgh, F., Heyndrickx, M., et al. (2021). The impact of maternal and piglet low protein diet and their interaction on the porcine liver transcriptome around the time of weaning. *Vet. Sci.* 8 (10), 233. doi:10.3390/vetsci8100233
- Kröger, W., Mapiye, D., Domelevo Entfellner, J.-B., and Tiffin, N. (2016). A meta-analysis of public microarray data identifies gene regulatory pathways deregulated in peripheral blood mononuclear cells from individuals with Systemic Lupus Erythematosus compared to those without. *BMC Med. Genomics* 9, 66. doi:10.1186/s12920-016-0227-0
- Kubik, R. M., Tietze, S. M., Schmidt, T. B., Yates, D. T., and Petersen, J. L. (2018). Investigation of the skeletal muscle transcriptome in lambs fed  $\beta$  adrenergic agonists and subjected to heat stress for 21 d. *Transl. Anim. Sci.* 2, S53–S56. doi:10.1093/tas/txy053
- Lai, Y., Zhang, F., Nayak, T. K., Modarres, R., Lee, N. H., and McCaffrey, T. A. (2016). Detecting discordance enrichment among a series of two-sample genome-wide expression data sets. *BMC Genomics* 17 (11), 1050. doi:10.1186/s12864-016-3265-2
- Lindholm-Perry, A. K., Freetly, H. C., Oliver, W. T., Rempel, L. A., and Keel, B. N. (2020). Genes associated with body weight gain and feed intake identified by meta-analysis of the mesenteric fat from crossbred beef steers. *PLoS One* 15 (1), e0227154. doi:10.1371/journal.pone.0227154
- Lindholm-Perry, A. K., Meyer, A. M., Kern-Lunbery, R. J., Cunningham-Hollinger, H. C., Funk, T. H., and Keel, B. N. (2022). Genes involved in feed efficiency identified in a meta-analysis of rumen tissue from two populations of beef steers. *Animals*. 12 (12), 1514. doi:10.3390/ani12121514
- Liptak, T. (1958). On the combination of independent tests. *Akad. Mat. Kut. Intezetenek Kozlemenyei* 3, 171
- Liu, S., Gao, Y., Canela-Xandri, O., Wang, S., Yu, Y., Cai, W., et al. (2021). A comprehensive catalogue of regulatory variants in the cattle transcriptome. *bioRxiv*. doi:10.1101/2020.12.01.406280
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Marot, G., Foulley, J. L., Mayer, C. D., and Jaffrézic, F. (2009). Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics* 25 (20), 2692–2699. doi:10.1093/bioinformatics/btp444
- Miller, R. G. (1974). The jackknife: A review. *Biometrika* 61, 1–15. doi:10.2307/2334280
- Moll, P., Ante, M., Seitz, A., and Reda, T. (2014). QuantSeq 3' mRNA sequencing for RNA quantification. *Nat. Methods* 11, i–iii. doi:10.1038/nmeth.f.376
- Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34, 267–273. doi:10.1038/ng1180
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). SCIENTIFIC STANDARDS. Promoting an open research culture. *Science* 348 (6242), 1422–1425. doi:10.1126/science.aab2374
- Owen, A. B. (2009). Karl Pearson's meta-analysis revisited. *Ann. Stat.* 37 (6), 3867–3892. doi:10.1214/09-AOS697
- Pascottini, O. B., Koster, J. D., Van Nieuwerburgh, F., Van Poucke, M., Peelman, L., Fievez, V., et al. (2021). Effect of overconditioning on the hepatic global gene expression pattern of dairy cows at the end of pregnancy. *J. Dairy Sci.* 104 (7), 8152–8163. doi:10.3168/jds.2020-19302
- Pedroza, G. H., Lanzon, L. F., Rabaglino, M. B., Walker, W. L., Vahmani, P., and Dincol, A. C. (2022). Exposure to non-esterified fatty acids *in vitro* results in changes in the ovarian and follicular environment in cattle. *Anim. Reprod. Sci.* 238, 106937. doi:10.1016/j.anireprosci.2022.106937
- Piras, I. S., Manchia, M., Huentelman, M. J., Pinna, F., Zai, C. C., Kennedy, J. L., et al. (2019). Peripheral biomarkers in schizophrenia: A meta-analysis of microarray gene expression data sets. *Int. J. Neuropsychopharmacol.* 22 (3), 186–193. doi:10.1093/ijnp/ppy103
- Rajesh, A., Chang, Y., Abedalthagafi, M. S., Wong-Beringer, A., Love, M. I., and Mangul, S. (2021). Improving the completeness of public metadata accompanying omics studies. *Genome Biol.* 22, 106. doi:10.1186/s13059-021-02332-z
- Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* 29 (17), 2146–2152. doi:10.1093/bioinformatics/btt350
- Rau, A., Marot, G., and Jaffrézic, F. (2014). Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinforma.* 15, 91. doi:10.1186/1471-2105-15-91
- Retallick, K. J., Bormann, J. M., Weaber, R. L., MacNeil, M. D., Bradford, H. L., Freetly, H. C., et al. (2017). Genetic variance and covariance and breed differences for feed intake and average daily gain to improve feed efficiency in growing cattle. *J. Anim. Sci.* 95, 1444–1450. doi:10.2527/jas.2016.1260
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Schroyen, M., Li, B., Sureda, E. A., Zhang, Y., Lebouis, J., Deforce, D., et al. (2021). Pre-weaning inulin supplementation alters the ileal transcriptome in pigs regarding lipid metabolism. *Vet. Sci.* 8 (10), 207. doi:10.3390/vetsci8100207
- Shi, L., Jones, W. D., Jensen, R. V., Harris, S. C., Perkins, R. G., Goodsaid, F. M., et al. (2008). The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinforma.* 9 (9), S10. doi:10.1186/1471-2105-9-S9-S10
- Sholder, G., Lanz, T. A., Moccia, R., Quan, J., Aparacio-Prat, E., Stanton, R., et al. (2020). 3'Pool-seq: An optimized cost-efficient and scalable method of whole-transcriptome gene expression profiling. *BMC Genomics* 21, 64. doi:10.1186/s12864-020-6478-3
- Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 3. doi:10.2202/1544-6115.1027
- Su, L., Chen, S., Zheng, C., Wei, H., and Song, X. (2019). Meta-analysis of gene expression and identification of biological regulatory mechanisms in Alzheimer's disease. *Front. Neurosci.* 13, 633. doi:10.3389/fnins.2019.00633
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralsler, M., et al. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* 15 (1), 675. doi:10.1186/1471-2164-15-675
- Toro-Domínguez, D., Villatoro-García, J. A., Martorell-Marugán, J., Román-Montoya, Y., Alarcón-Riquelme, M. E., et al. (2021). A survey of gene expression meta-analysis: Methods and applications. *Brief. Bioinform.* 22 (2), 1694–1705. doi:10.1093/bib/bbaa019
- Tseng, G. C., Ghosh, D., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for micro-array meta-analysis. *Nucleic Acids Res.* 40 (9), 3785–3799. doi:10.1093/nar/gkr1265
- Wan, Y.-W., Al-Ouran, R., Mangleburg, C. G., Perumal, T. M., Lee, T. V., Allison, K., et al. (2020). Meta-analysis of the alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep.* 32 (2), 107908. doi:10.1016/j.celrep.2020.107908
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi:10.1038/sdata.2016.18
- Winter, C., Kosch, R., Ludlow, M., Osterhaus, A. D. M. E., and Jung, K. (2019). Network meta-analysis correlates with analysis of merged independent transcriptome expression data. *BMC Bioinforma.* 20, 144. doi:10.1186/s12859-019-2705-9
- Ye, Z., Ke, H., Chen, S., Cruz-Cano, R., He, X., Zhang, J., et al. (2021). Biomarker categorization in transcriptomic meta-analysis by concordant patterns with application to pan-cancer studies. *Front. Genet.* 12, 651546. doi:10.3389/fgene.2021.651546