NGUYEN Hoang Alexis

Chicken Muscle Shear Force RNA-seq Analysis: A reproduction study

This document is a report concerning the RNA-seq analysis made.

The data used for this analysis seems to be affiliated to the article from Piórkowska, et al. (2016) who realized a "whole transcriptome analysis of breast muscles in broiler chicken groups differing in shear force".

The whole summary is publicly available (Piórkowska et al., 2015)[1]

As the complete article is not publicly available, the purpose of this present analysis and report is not to confirm the potential results or informations since I didn't have access to it, but rather show at the present moment my ability to conduct a transcriptome analysis from a set of data and make interpretations from it and the limited context available.

NB : This report and analysis is entirely personal. By no means this report is meant to be considered as the point of view of the authors from the original article the data is associated with. This only represent my personal results and interpretations of the public available data.

Shear force is a critical measure of meat quality in chickens, particularly in evaluating the tenderness of the meat. Studies have shown that the shear force value can be influenced by various factors, including the method of cooking and the presence of myopathies such as wooden breast (WB) and white striping (WS) myopathies (Bordignon et al., 2022)[2].

White striping (WS) and wooden breast (WB) myopathies are significant issues in the poultry industry, affecting the quality and economic value of chicken breast meat. These conditions have been the subject of several studies, which have provided insights into their characteristics, prevalence, and potential causes (Praud et al., 2020)[3].

- **White Striping (WS)**: This condition is characterized by the appearance of white striations parallel to the muscle fibers on the breast, thigh, and tender muscles of broilers. WS affects the visual aspect, nutritional value, and processing yields of the meat.34
- **Wooden Breast (WB)**: WB is characterized by muscle hardening, heterogeneity of color, excessive superficial exudate, and loss of muscle elasticity. It significantly impacts the quality of breast meat, making it tougher and less appealing to consumers.

The data used here has been retrieved from the BioProject accession number PRJNA297364

Gallus gallus breed:Hubbard Flex (chicken) Accession: PRJNA297364 ID: 297364

Gallus gallus breed:Hubbard Flex Transcriptome or Gene expression

In the present study, a whole transcriptome analysis of pectoralis muscles was conducted in 8 broiler chicken groups differing in shear force (4 characterized low shear force, 4 characterized high shear force). More...

See Genome Information for Gallus gallus

NAVIGATE ACROSS

1717 additional projects are related by organism.

Accession	PRJNA297364
Data Type	Transcriptome or Gene expression
Scope	Multiisolate
Organism	**Gallus gallus** [Taxonomy ID: 9031] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes; Phasianidae; Phasianinae; Gallus; Gallus gallus
Submission	Registration date: 29-Sep-2015 **National Research Institute of Animal Production**
Relevance	Agricultural

From this accession the different SRA data containing the reads in fastq format have been obtained thanks to their accession number.

The experimental design is composed of four biological replicates per condition (two conditions studied : high-shear force and low-shear force)

The different samples are organized as the following :

- High shear force (4 replicates):
  - SRR2554364 (replicate 1)
  - SRR2554365 (replicate 2)
  - SRR2554366 (replicate 3)
  - SRR2554367 (replicate 4)
- Low shear force (4 replicates):
  - SRR2554362 (replicate 1)
  - SRR2554363 (replicate 2)
  - SRR2554345 (replicate 3)
  - SRR2554344 (replicate 4)

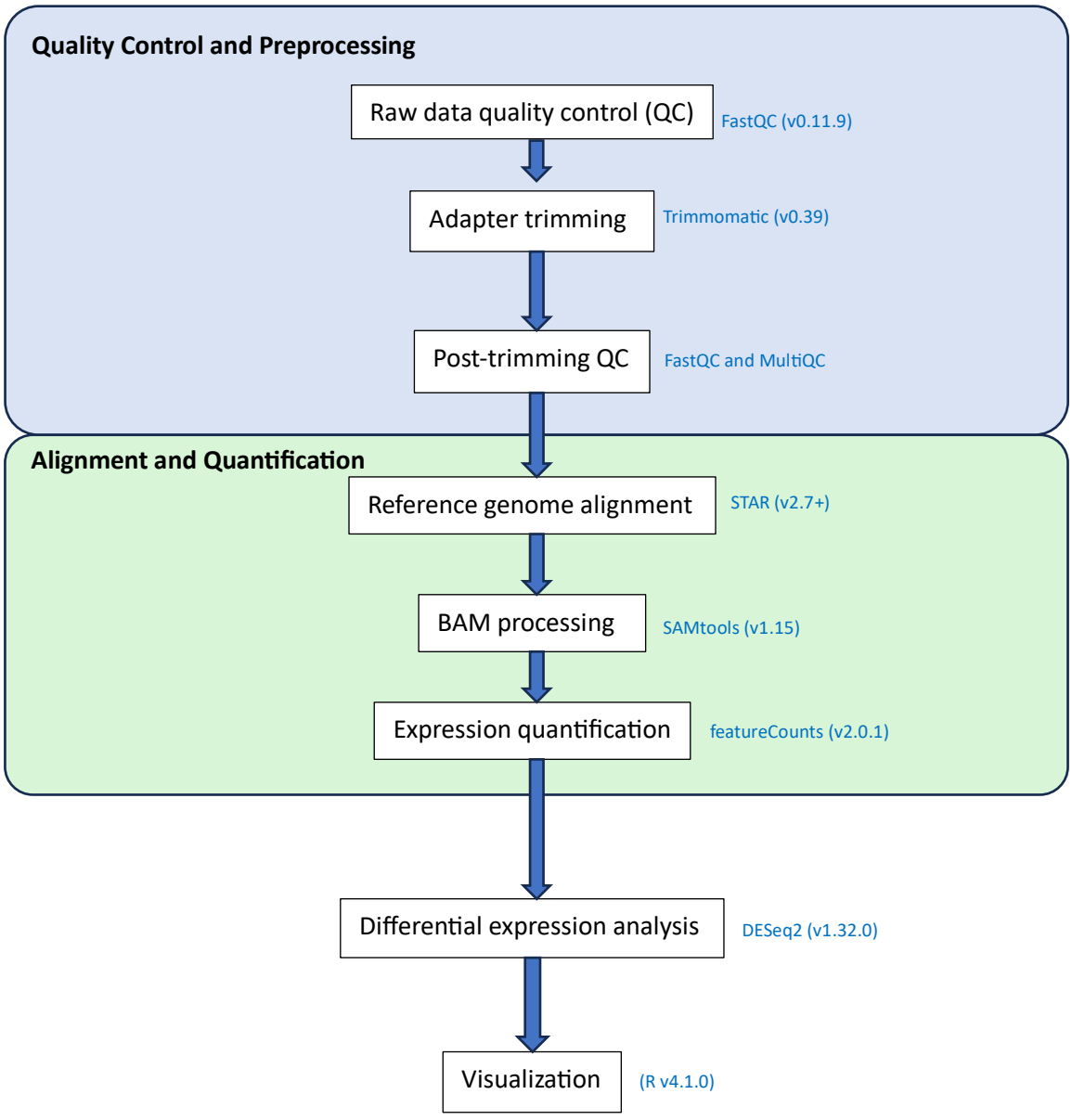
Other characteristic of interest are listed there :

- **Experimental Design :** Four biological replicates per condition
- **Sequencing Platform :** Illumina HiScanSQ
- **Library Type :** RNA-seq, single-end reads

NGUYEN Hoang Alexis

Analysis pipeline

The analysis pipeline made is presented below:



NGUYEN Hoang Alexis

Description of tools used in the pipeline

This pipeline integrates several widely used tools for quality control, alignment, quantification, and statistical analysis of RNA-seq data. Below is a summary of the tools used, along with their roles and links to their respective GitHub repositories.

- **FastQC**

Purpose: Raw sequencing read quality control
Description: FastQC provides quick assessments of sequence quality, including base quality scores, GC content, duplication levels, and adapter contamination.

- **Trimmomatic**

Purpose: Adapter trimming and quality filtering
Description: Trimmomatic removes adapter sequences and low-quality regions from Illumina reads, improving downstream alignment performance.

- **MultiQC**

Purpose: Consolidated QC reporting
Description: MultiQC aggregates results from FastQC and other tools into a single interactive HTML report for easy comparison across all samples.

- **STAR (Spliced Transcripts Alignment to a Reference)**

Purpose: RNA-seq read alignment
Description: STAR is a highly efficient and accurate RNA-seq aligner that handles splice junctions and large genomes. It produces high-quality alignment outputs in BAM format.

- **featureCounts**

Purpose: Gene expression quantification
Description: featureCounts, part of the Subread package, assigns aligned reads to genomic features (e.g., genes) to generate count matrices for differential expression analysis.

- **SAMtools**

Purpose: Processing of aligned reads
Description: SAMtools is used for manipulating alignment files, including sorting, indexing, and filtering of BAM files.

- **DESeq2**

Purpose: Differential expression analysis
Description: DESeq2 uses negative binomial models to identify differentially expressed genes while accounting for biological variability and count distribution properties.

- **R (v4.1.0)**

Purpose: Statistical analysis and data visualization
Description: R is used to run DESeq2, generate plots (PCA, MA, heatmaps), and produce interpretable figures and tables from RNA-seq data.

NGUYEN Hoang Alexis

Results

MultiQC report

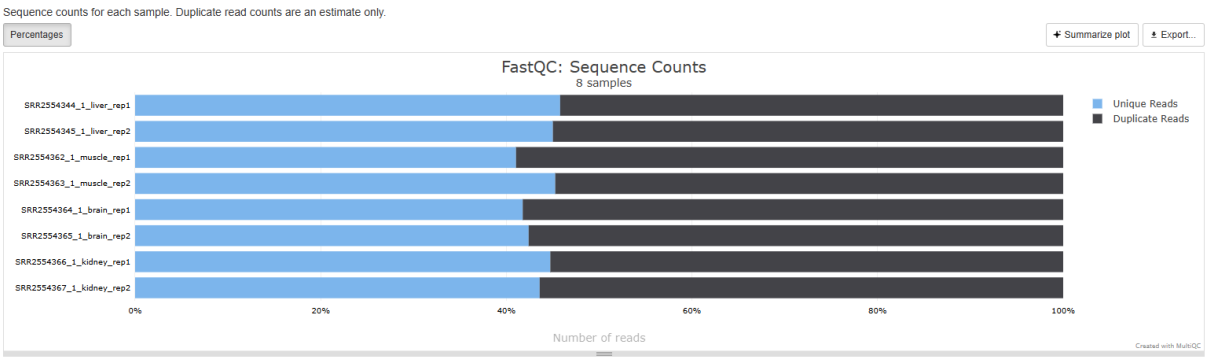
The multiQC report (available in the Github repository) is presented below :

General Statistics

Sample Name	Dups	GC	Seqs
SRR2554344_1_liver_rep1	54.2 %	49.0 %	7.8 M
SRR2554345_1_liver_rep2	55.0 %	49.0 %	10.1 M
SRR2554362_1_muscle_rep1	59.0 %	49.0 %	9.0 M
SRR2554363_1_muscle_rep2	54.7 %	49.0 %	8.9 M
SRR2554364_1_brain_rep1	58.2 %	49.0 %	9.7 M
SRR2554365_1_brain_rep2	57.6 %	49.0 %	10.6 M
SRR2554366_1_kidney_rep1	55.3 %	50.0 %	10.4 M
SRR2554367_1_kidney_rep2	56.4 %	49.0 %	9.5 M

Dups : % duplicate reads / GC : average %GC content / Seq : total sequences (Millions)

Sequence Counts

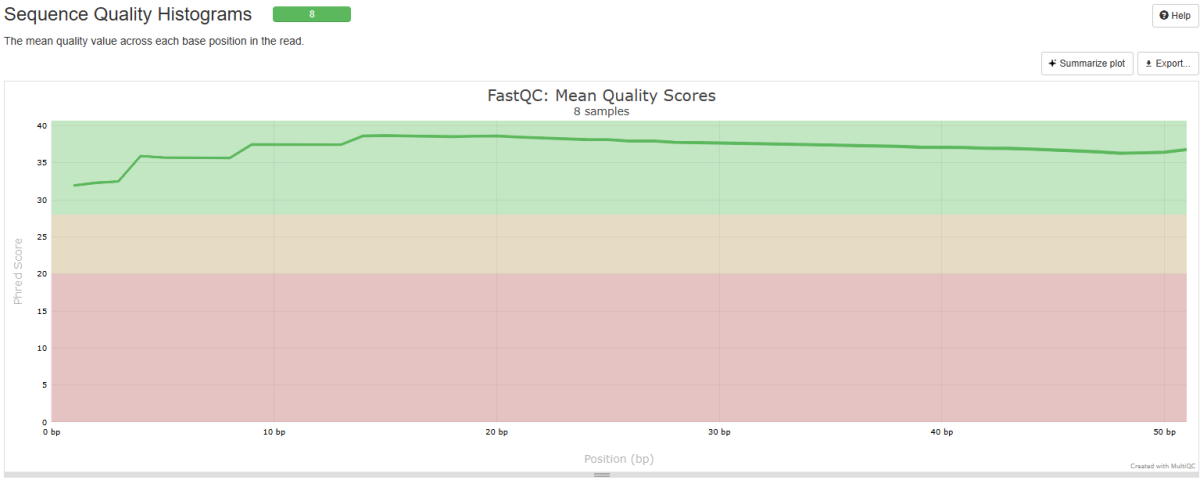


For every sample, nearly half of reads are marked as **duplicate**, while the remaining ones are reported as unique. The duplicate rate is fairly **uniform across tissues**, which suggests a **systematic feature** (not an anomaly in one sample).

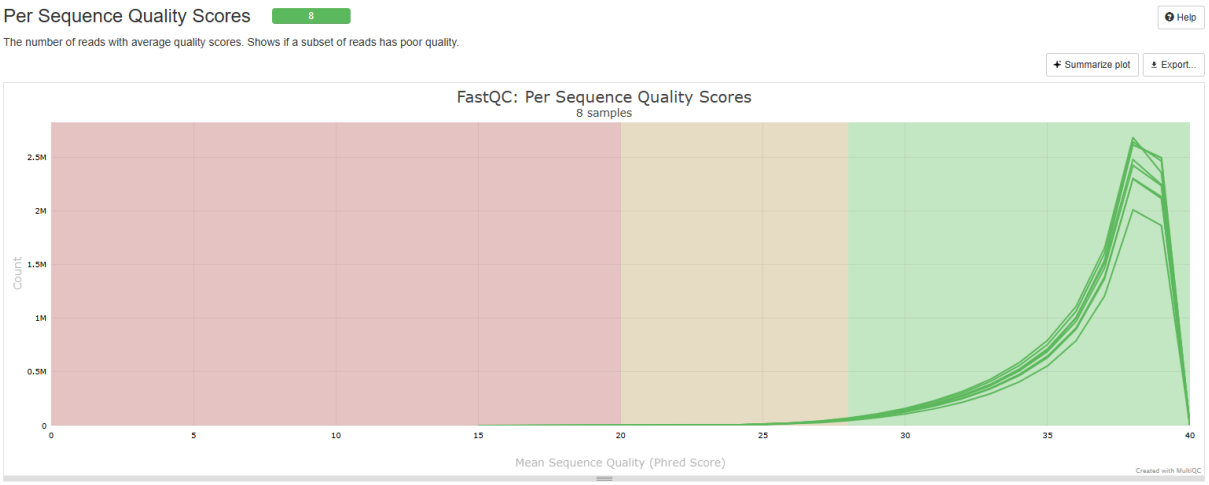
In RNA-seq, high duplication rate is expected, as the differential expression is based on the differential coverage of the reads on the genome reference.

NGUYEN Hoang Alexis

Sequence Quality Histograms



Per Sequence Quality Scores



Each sample has a passing Per Sequence Quality score meaning good average quality.

Per Base Sequence Content

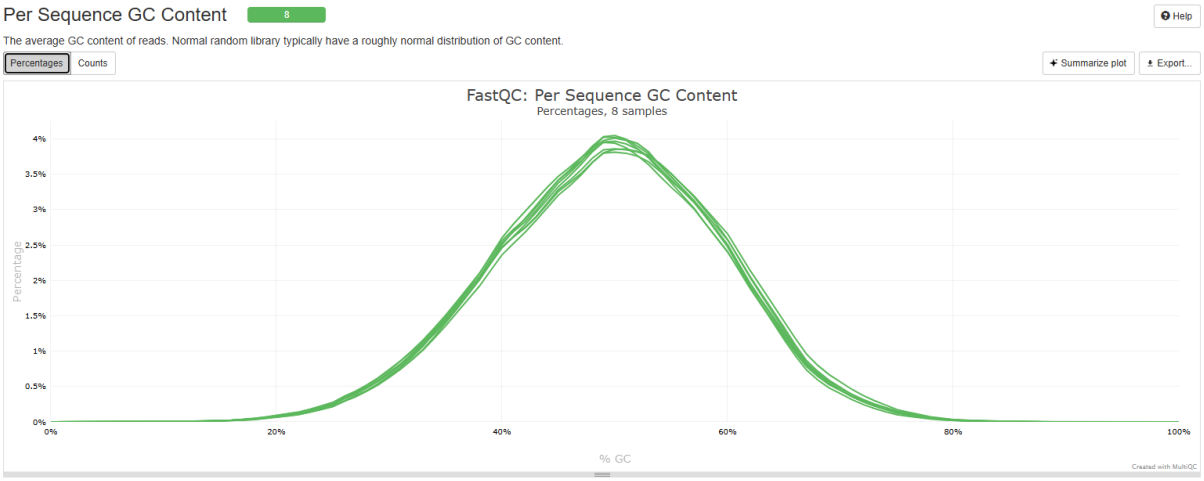
Base composition (A, T, G, C) at **each position** across the read length in the sequencing data. It checks whether the proportions of the four nucleotides are roughly balanced across the read — something expected in random, high-complexity libraries like RNA-seq.

All the samples failed this test, the reason being RNA-seq libraries, especially **single-end**, **non-stranded**, or **poly-A enriched**, often exhibit **non-random base composition at the start of reads** due to:

- Primer bias
- Adapter remnants
- Poly-A tailing
- Biased fragmentation

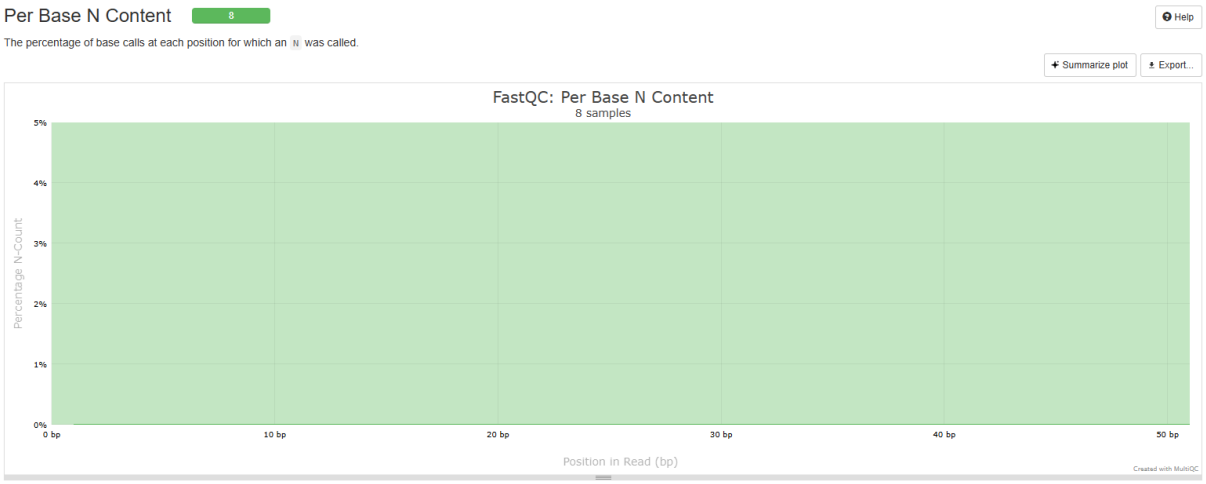
This **technical artifact** leads to **non-uniform base frequencies**.

Per Sequence GC Content



The figures represent **the GC content across the whole length of each sequence and compares it to a modelled normal distribution of GC content**. In a normal random library, you would expect to see a roughly normal distribution of GC content where the central peak corresponds to the overall GC content of the underlying genome.

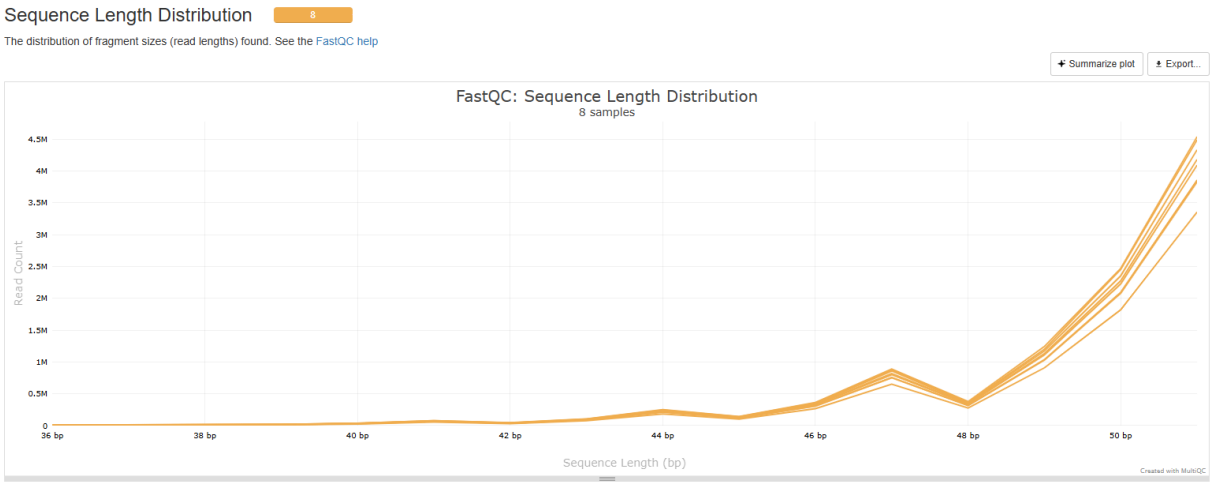
Per Base N Content



There is no “N” bases, which means no unknown base has been detected.

NGUYEN Hoang Alexis

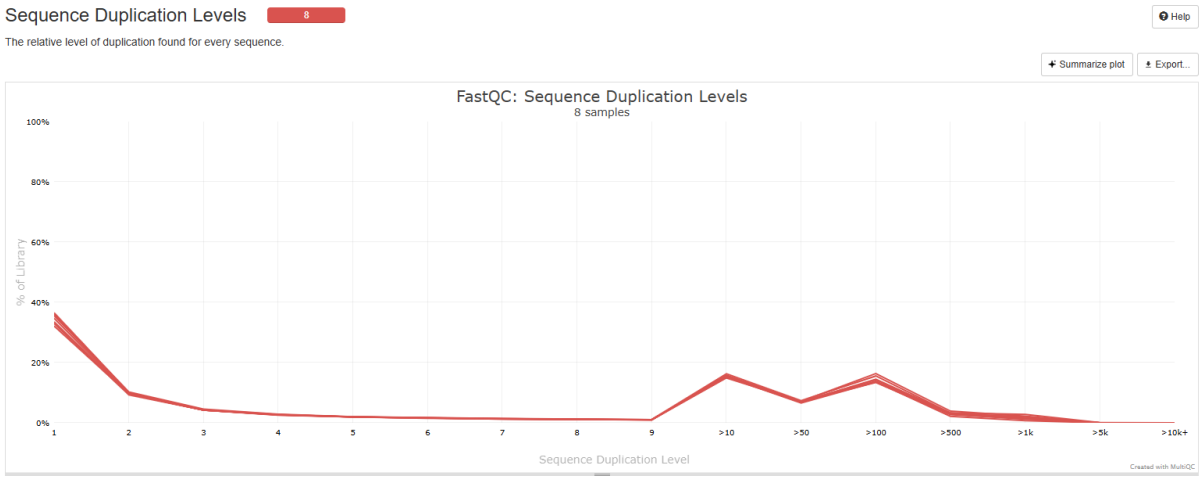
Sequence Length Distribution



A majority of reads have a sequence length of more than 50 pb, which is expected from single-end sequencing.

A minor part with reduced length of 47 pb could be explained by adaptor trimming.

Sequence Duplication Levels

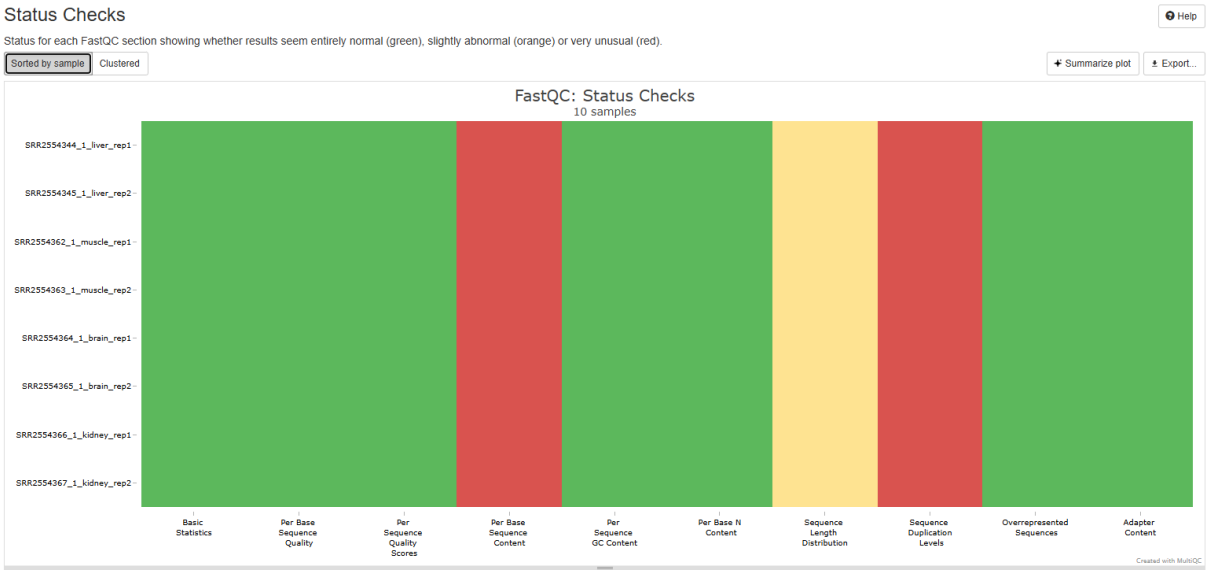


There seems to be an important number of sequences duplications levels.

As explained before high duplication rate is expected, especially in RNA-seq, as the differential expression is based on the differential coverage of the reads on the genome reference.

NGUYEN Hoang Alexis

Summary status checks

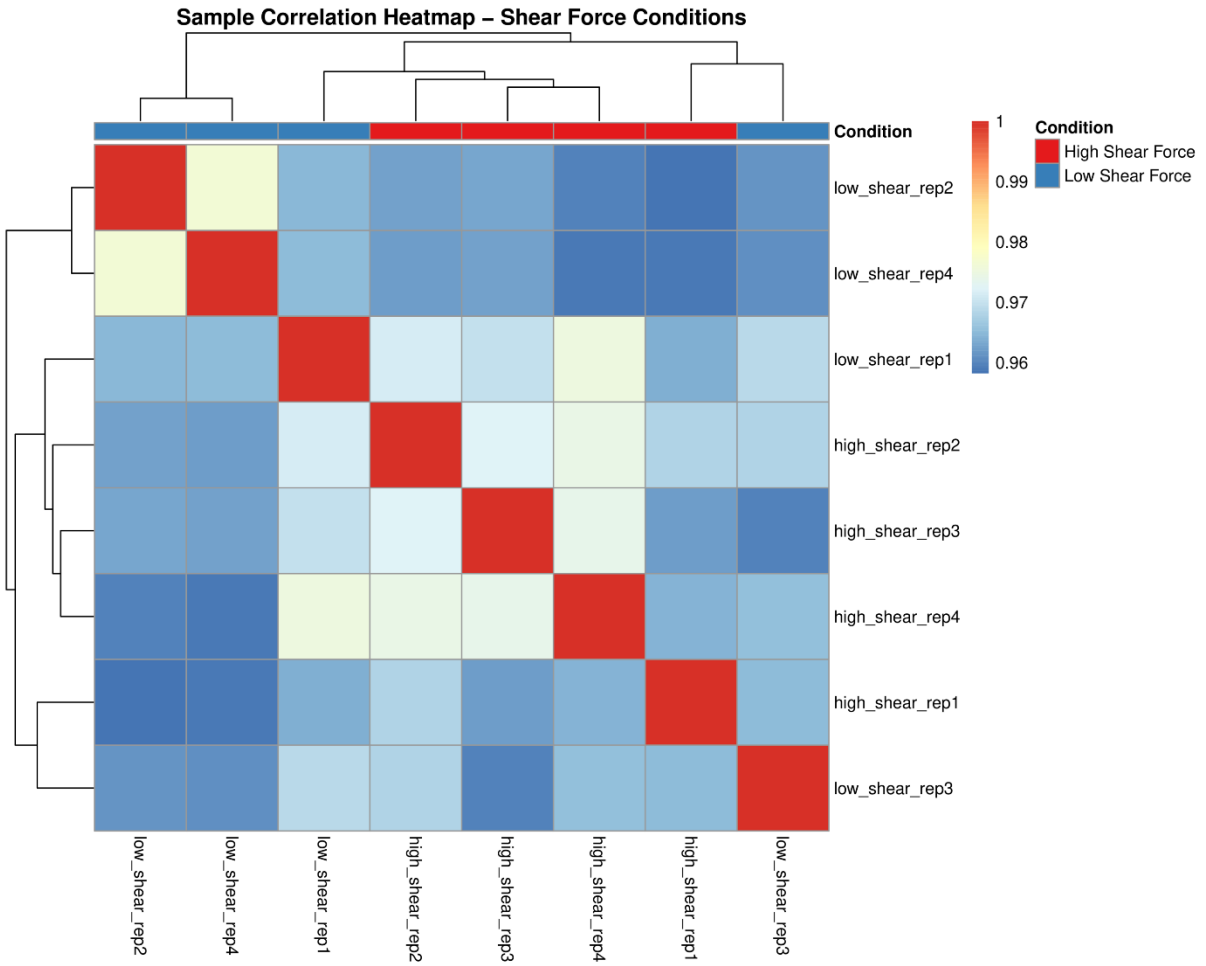


Visualizations

The different visualizations produced by the R script are presented below :

1. Sample Correlation Analysis

The correlation heatmap shows the pairwise correlation coefficients between samples from different shear force conditions:



This visualization helps identify:

- Clear clustering of samples by shear force condition
- High correlation between replicates within each condition
- Distinct expression patterns between high and low shear force groups

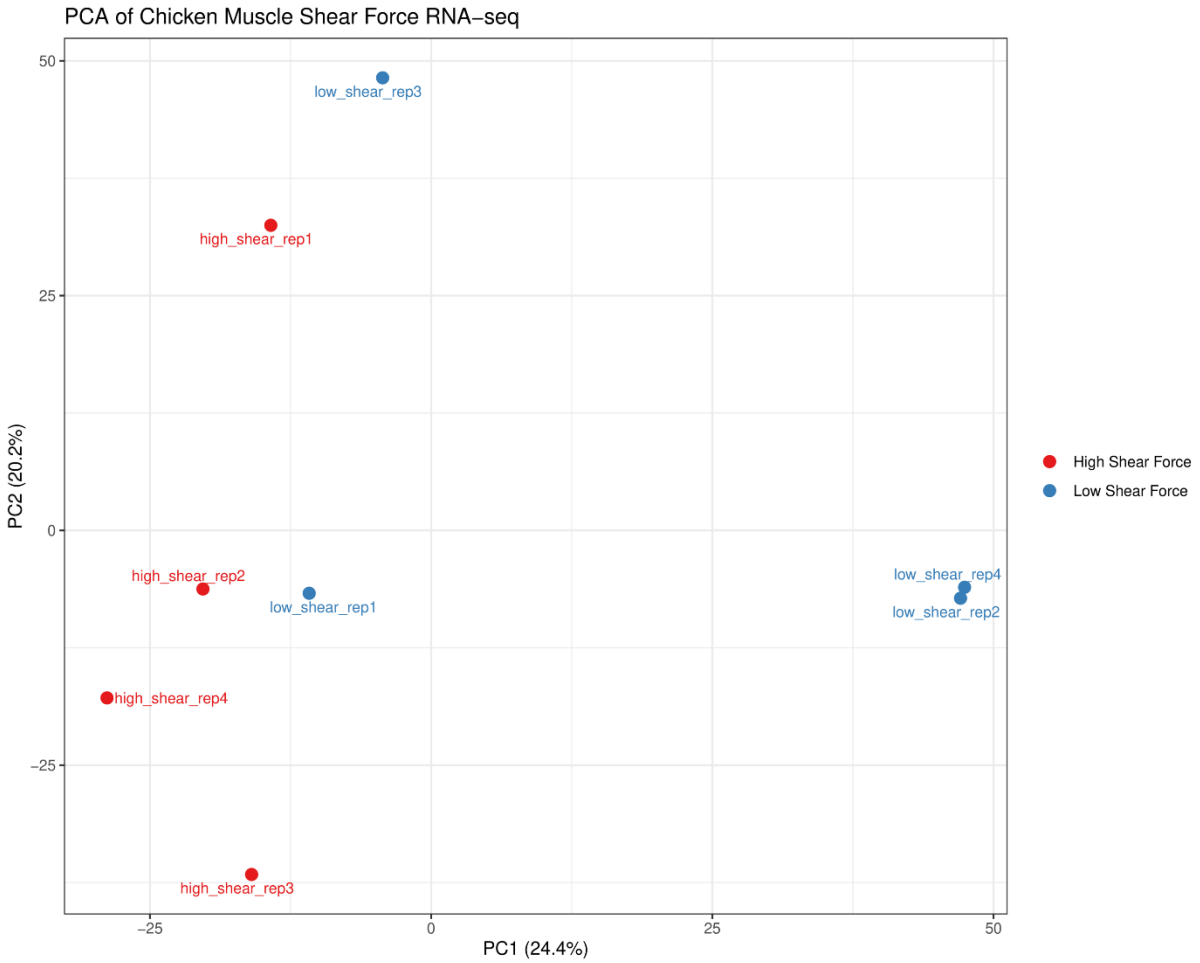
There seems to be a slightly higher correlation between low_shear_rep1, high_shear_rep2, high_shear_rep3 and high_shear_rep4. However, the colour scale used in the legend must be considered, as the correlations, since the correlation values are between 0.96 and 1 which are in any case very high and close values.

The hierarchical clustering located above the heatmap shows that many samples of the same condition are grouped together, especially the high-shear condition. A note can be made regarding the low_shear_rep3 since it is observed far from other replicates of its condition.

In any case, given the scale of the legend, one must be careful about any interpretation regarding this heatmap.

2. Principal Component Analysis

The PCA plot demonstrates the separation between high and low shear force groups in reduced dimensional space:



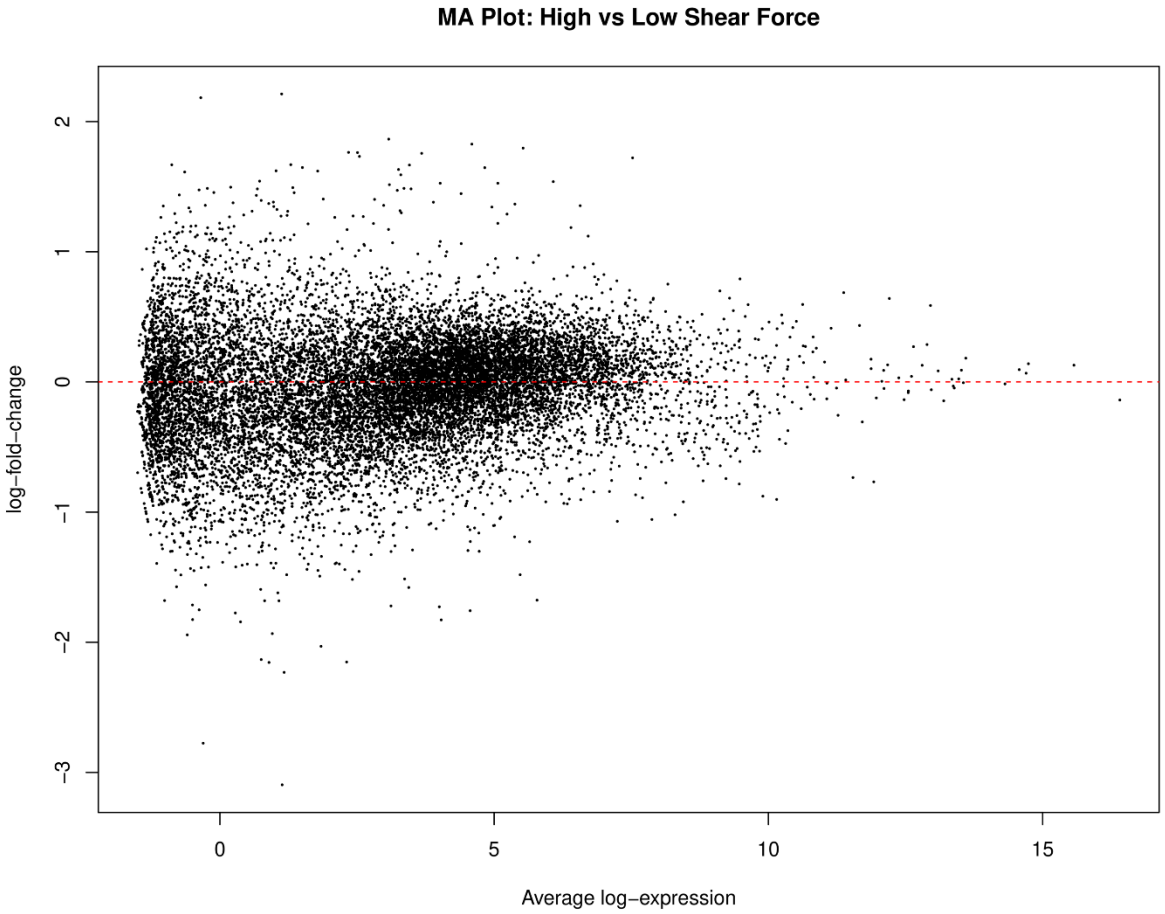
The PCE seems to apply a clear segregation between two groups:

- One on the left regrouping all of the sample from the high shear force condition and two of the samples from the low shear force conditions (replicates 1 and 3)
- One on the right with the two remaining samples from the low-shear force condition.

These observations are consistent with these made before regarding the heatmap, however the replicates 1 and 3 from the low shear force conditions are the ones located most to the right within the left group and closer to the replicates of the same condition.

3. Differential Expression Analysis

MA plots showing the relationship between mean expression and log fold change, it shows genes that are differentially expressed comparing the two conditions :



This plot shows the distribution of up and down-regulated genes in the high shear force condition comparing to the low shear force condition.

We can see many outliers which have an value of the log-fold-change above 1, and others below -1 : we can then state that these outliers correspond to genes that are respectively upregulated and down-regulated, the most differentially regulated being the ones the furthest from the red line.

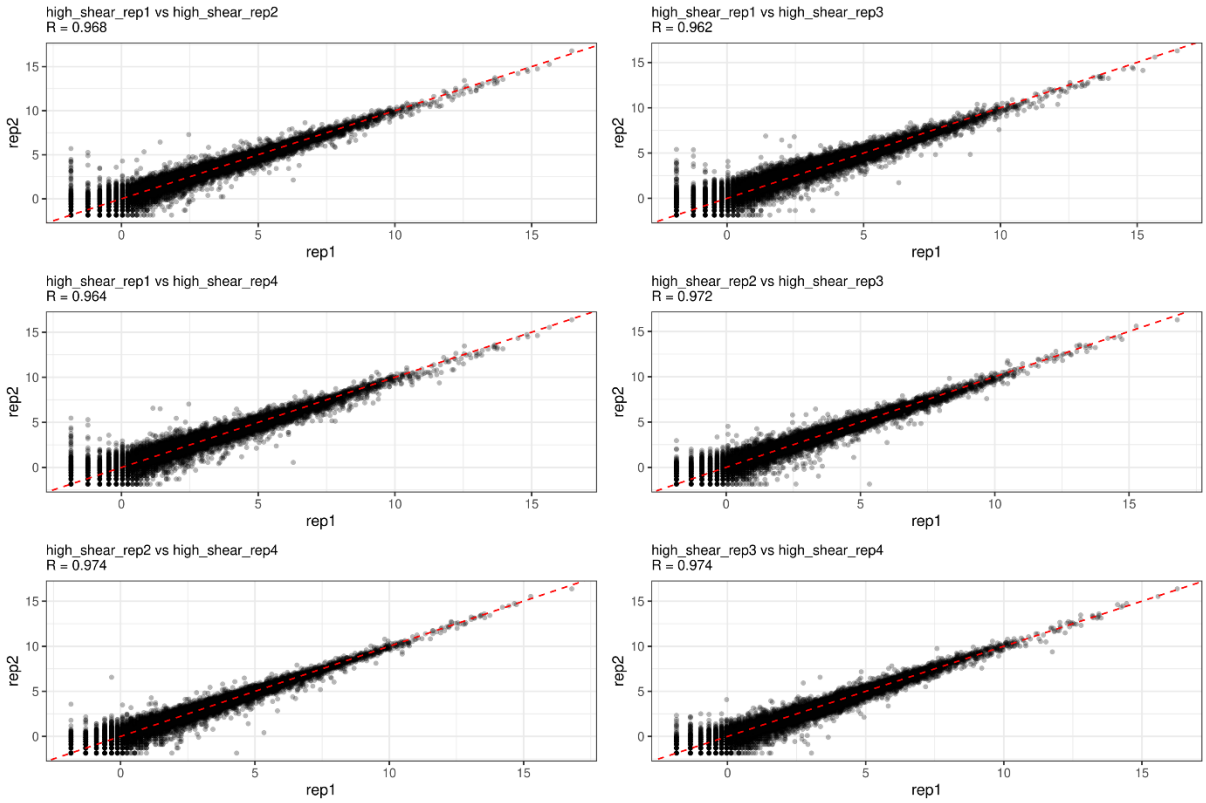
Another analysis related to the results from MA plot stated that among the 30,108 genes analyzed :

- 827 genes showed significantly higher expression in high shear force samples
- 184 genes showed significantly higher expression in low shear force samples

4. Technical Reproducibility

Correlation plots between biological replicates demonstrate the quality and reproducibility of the data:

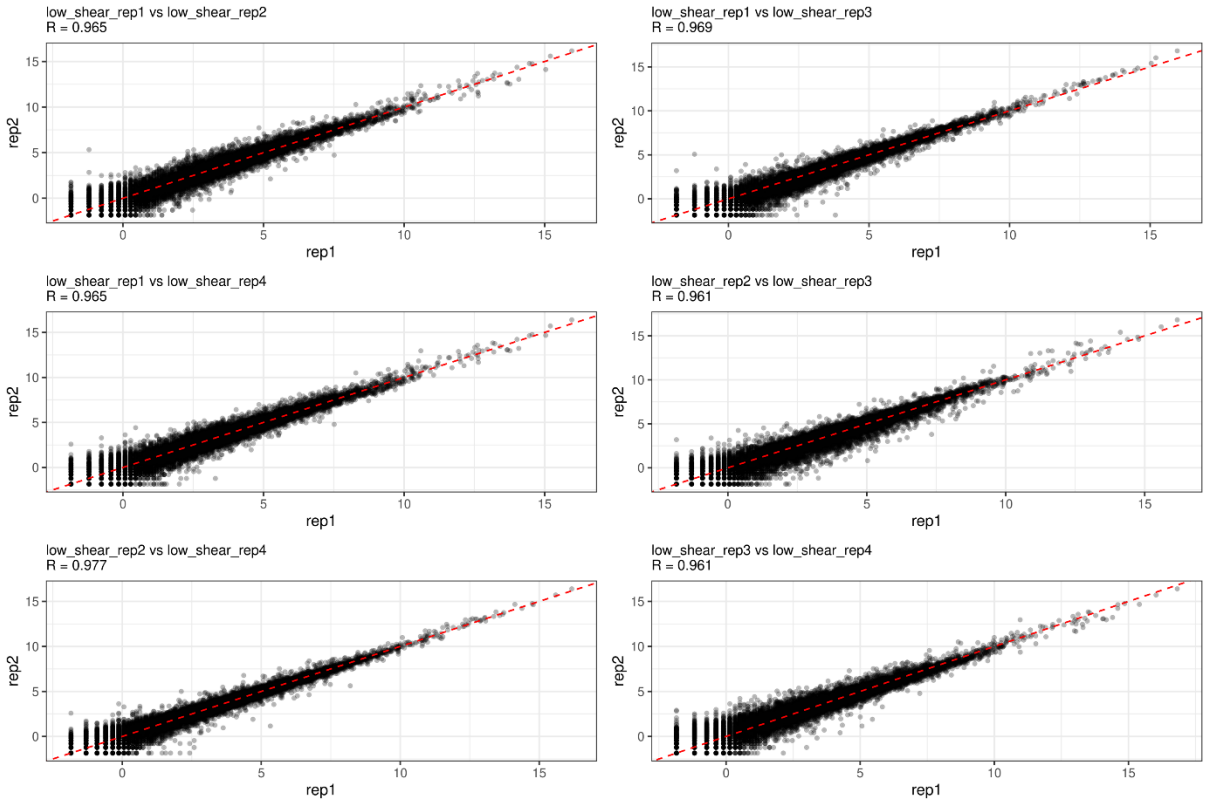
High Shear Force Replicates



The correlations show:

- Pairwise comparisons between all high shear force replicates
- Strong correlation coefficients ($R^2 > 0.95$)
- Consistent expression patterns across replicates

Low Shear Force Replicates



The correlations demonstrate:

- Pairwise comparisons between all low shear force replicates
- High technical reproducibility ($R^2 > 0.95$)
- Consistent expression patterns across biological replicates

Extraction of the genes most affiliated to each condition

High shear-force :

- ENSGALG00010029558 : Novel gene – no description
- ENSGALG00010019253 : HBAD - hemoglobin alpha subunit D
- ENSGALG00010003868 : HBBA - hemoglobin subunit epsilon 1
- ENSGALG00010019732 : HBA1 - hemoglobin subunit alpha 1
- ENSGALG00010025644 : PI16 - peptidase inhibitor 16

Low shear-force :

- ENSGALG00010014488 : ADPRHL1 - ADP-ribosylhydrolase like 1
- ENSGALG00010016987 : ASB2 - ankyrin repeat and SOCS box containing 2
- ENSGALG00010016978 : FAM181A - family with sequence similarity 181 member A
- ENSGALG00010028364 : MHM2 - male hypermethylated region 2
- ENSGALG00010007900 : FHOD3 - formin homology 2 domain containing 3

Genes Associated with the *High Shear Force* Condition (Tougher Meat)

The genes most highly expressed in muscles with high shear force include several hemoglobin subunits (*HBAD*, *HBBA*, *HBA1*), as well as *PI16* and a gene annotated as "novel."

- **Hemoglobins (HBAD, HBBA, HBA1)**: The high expression of these genes may indicate a response to local hypoxia or increased oxidative stress in muscle fibers. Such a response could be linked to structural disorders like myopathies (e.g., wooden breast), where tissue oxygenation is impaired (Emami et al., 2021)[3].
- **PI16 (Peptidase Inhibitor 16)**: This gene codes for a protease inhibitor involved in extracellular matrix remodeling and inflammatory processes. Its overexpression may reflect a fibrotic response or altered muscle turnover (Hazell et al., 2016)[4].
- **Novel gene (ENSGALG00010029558)**: This unannotated gene could represent a muscle-specific sequence or a conditional regulation related to stress. Further functional analysis (BLAST, protein domains, etc.) would be necessary.

These genes suggest a potential disruption of muscle homeostasis in high shear force samples, characterized by metabolic adaptation and an inflammatory or structural response.

Genes Associated with the *Low Shear Force* Condition (Tender Meat)

The genes most expressed in the low shear force condition are primarily involved in normal muscle development, cytoskeletal remodeling, and intracellular signaling.

- **ASB2**: This gene plays a key role in the degradation of filamins, structural proteins of the muscle cytoskeleton (Heuzé et al., 2008)[4]. Its expression aligns with active muscle remodeling that favors better tenderness. It is important to state that Piórkowska et al. (2015) also detected that gene in the upregulated group from the low shear force condition.
- **FHOD3**: This gene codes for a regulator of actin involved in the formation of functional muscle fibers. Its expression is consistent with healthy, structured cytoskeleton organization.
- **ADPRHL1**: Specifically expressed in muscle tissues, this gene is involved in cardiac and muscle development, suggesting normal muscle function. It seems that this gene is associated with growth traits in the chicken as well (Li et al., 2021)[5]
- **FAM181A and MHM2**: While their functions are not fully defined, their overexpression in tender meat samples could reflect secondary epigenetic or neuronal regulations that favor optimal muscle quality.

These findings suggest that tender meat is characterized by regulated muscle activity, a well-organized cytoskeleton, and efficient remodeling, in contrast to the stressed or disorganized response of muscles with high shear force.

References

1. Piórkowska, K., Żukowski, K., Nowak, J., Połtowicz, K., Ropka-Molik, K., & Gurgul, A. (2015). Genome-wide RNA-Seq analysis of breast muscles of two broiler chicken groups differing in shear force. *Animal Genetics*, *47*(1), 68–80. https://doi.org/10.1111/age.12388
2. Bordignon, F., Xiccato, G., Cabrol, M. B., Birolo, M., & Trocino, A. (2022). Factors affecting breast myopathies in broiler chickens and quality of defective meat: A Meta-Analysis. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/fphys.2022.933235
3. Emami, N. K., Cauble, R. N., Dhamad, A. E., Greene, E. S., Coy, C. S., Velleman, S. G., Orlowski, S., Anthony, N., Bedford, M., & Dridi, S. (2021). Hypoxia further exacerbates woody breast

myopathy in broilers via alteration of satellite cell fate. Poultry Science, 100(7), 101167. https://doi.org/10.1016/j.psj.2021.101167

4. Hazell, G. G. J., Peachey, A. M. G., Teasdale, J. E., Sala-Newby, G. B., Angelini, G. D., Newby, A. C., & White, S. J. (2016). PI16 is a shear stress and inflammation-regulated inhibitor of MMP2. Scientific Reports, 6(1). https://doi.org/10.1038/srep39553
5. Praud, C., Jimenez, J., Pampouille, E., Couroussé, N., Godet, E., Bihan-Duval, E. L., & Berri, C. (2020). Molecular phenotyping of white striping and wooden breast myopathies in chicken. *Frontiers in Physiology*, *11*. https://doi.org/10.3389/fphys.2020.00633
6. Heuzé, M. L., Lamsoul, I., Baldassarre, M., Lad, Y., Lévêque, S., Razinia, Z., Moog-Lutz, C., Calderwood, D. A., & Lutz, P. G. (2008). ASB2 targets filamins A and B to proteasomal degradation. *Blood*, *112*(13), 5130–5140. https://doi.org/10.1182/blood-2007-12-128744
7. Li, T., Chen, B., Wei, C., Hou, D., Qin, P., Jing, Z., Ma, H., Niu, X., Wang, C., Han, R., Li, H., Liu, X., Xu, H., Kang, X., & Li, Z. (2021). A 104-BP structural variation of the ADPRHL1 gene is associated with growth traits in chickens. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.691272

Github

https://github.com/atomemeteore/chicken_RNAseq_analysis.git