Genetic Validation of a Specific Intervertebral Disc Cell Population

Sarah Leigh Frunz

Coach: Prof. Dr. Fabian Ille
Expert: PD, Dr. Philipp Stämpfli

Bachelor Thesis Medical Engineering

Bachelor Thesis submitted as part of the requirements for the BSc in Medical Engineering at the School of Engineering & Architecture, Lucerne University of Applied Sciences and Arts.

1st January 2022

Bachelor-Thesis an der Hochschule Luzern - Technik & Architektur

Titel	**Genetic Validation of a Specific Intervertebral Disc Cell Population**
Diplomandin/Diplomand	**Frunz, Sarah Leigh**
Bachelor-Studiengang	**Bachelor Medizintechnik**
Semester	**HS21**
Dozentin/Dozent	**Ille, Fabian**
Expertin/Experte	**Stämpfli, Philipp**

Abstract Deutsch

Die Degeneration der Bandscheibe (IVD) und ihre medizinischen Folgen sind nach wie vor ein Thema in der heutigen Gesellschaft. Mangelnde Kenntnisse über die zelluläre Zusammensetzung der IVD behindern Fortschritte bei regenerativen Therapien. Um hier Abhilfe zu schaffen, bestimmten Calió et al. (2021) die zelluläre Zusammensetzung des Nucleus Pulposus (NP) und des Annulus Fibrosus (AF). Fünf eindeutig exprimierte Gene, die ausschliesslich in Zellcluster zwei der identifizierten NP-Zellen exprimiert werden, wurden zur Validierung durch Genexpressionsanalyse mittels quantitative polymerase chain reaction (qPCR) ausgewählt. Ziel dieses Projekts war die Entwicklung von qPCR-Primern für die fünf eindeutig exprimierten Gene: SFRP5, ADIRF, S100A2, ACTG1 und TUBB. Anschliessend wurden die Primer an RNA getestet, die extrahiert und in komplementäre DNA umgeschrieben worden war. Zur Validierung der Gene wurden eine Spezifitätsanalyse mittels Schmelzkurvenanalyse und Gelelektrophorese sowie eine Effizienzanalyse anhand einer Standardkurve durchgeführt. Schliesslich wurde eine relative Quantifizierung durchgeführt, um den Unterschied in der Expression der Gene in der distalen IVD im Vergleich zur proximalen IVD zu bestimmen. Das Ergebnis zeigt, dass es möglich war, Primer für SFRP5, ADIRF und S100A2 zu etablieren, nicht aber für ACTG1 und TUBB. SFRP5 amplifizierte unspezifische Produkte und wurde als nicht funktionierender Primer ausser Acht gelassen. Darüber hinaus ist ADIRF von den fünf eindeutig exprimierten Genen das einzige, das die Validierungskriterien erfüllt. Ausserdem zeigt die relative Quantifizierung, dass ADIRF und S100A2 in der distalen IVD in geringerem Masse exprimiert werden als in der proximalen IVD. Schliesslich zeigen die Gesamtergebnisse, dass ADIRF und S100A2 während der Genexpressionsanalyse vorhanden und aktiv waren. Somit kann das Vorhandensein von ADIRF und S100A2 in der RNA der Rinder-IVD bestätigt werden.

Abstract Englisch

Degeneration of the intervertebral disc (IVD) and its medical repercussions remains an issue in today's society. A lack of knowledge of the cellular composition of the IVD hinders advances in regenerative therapies. To assist in this matter, Calió et al. (2021) determined the cellular composition of the Nucleus Pulposus (NP) and Annulus Fibrosus (AF). Five uniquely expressed genes that are exclusively expressed in cell cluster two of the identified NP cells, were selected for validation through gene expression analysis using quantitative polymerase chain reaction (qPCR). This project aimed to design qPCR primers for the five

uniquely expressed genes: SFRP5, ADIRF, S100A2, ACTG1 and TUBB. Following that, the primers were tested on RNA that had been extracted and reverse transcribed into complementary DNA. Specificity analysis via melt curve analysis and gel electrophoresis and efficiency analysis using a standard curve were used to validate the genes. Lastly, relative quantification was performed to determine the difference in expression of the genes in distal IVD compared to proximal IVD. The outcome indicates that it was feasible to establish primers for SFRP5, ADIRF and S100A2 but not for ACTG1 and TUBB. SFRP5 amplified non-specific products and was disregarded as a non-working primer. Furthermore, of the five uniquely expressed genes, ADIRF is the only one that passes validation criteria. Additionally, relative quantification shows that ADIRF and S100A2 are expressed at lower levels in the distal IVD than in the proximal IVD. Lastly, the overall findings show that ADIRF and S100A2 were present and active during the gene expression analysis. Hence, the presence of ADIRF and S100A2 in bovine IVD RNA may be validated.

Ort, Datum Luzern, 1 January 2022

List of Abbreviations

ACTG1	Actin Gamma 1
ADIRF	Adiposis Regulatory Factor
AF	Annulus Fibrosus
bp	Base Pairs
cDNA	Complementary DNA
Ct	Threshold Cycle
DNA	Deoxyribonucleic Acid
gDNA	Genomic DNA
GO	Gene Ontology
GTP	Guanosine Triphosphate
RNA	Ribonucleic Acid
dsDNA	Double-Stranded DNA
IVD	Intervertebral Discs
IVDD	Intervertebral Disc Degeneration
mRNA	Messenger RNA
NP	Nucleus Pulposus
NTC	No Template Control
qPCR	Quantitative Polymerase Chain Reaction
PCR	Polymerase Chain Reaction
SFRP5	Secreted Frizzled-Related Protein 5
ssDNA	Single-Stranded DNA
S100A2	S100 Calcium-Binding Protein A2
TUBB	Tubulin Beta Class I
Wnt	Wingless-Related Integration Site

Table of Contents

1. Introduction

1.1 Topic

Topic of this thesis is the genetic validation of a specific interverbal disc cell population.

1.2 Rationale

The intervertebral discs (IVD) are components of the spinal column and flexibly connect the vertebrae.

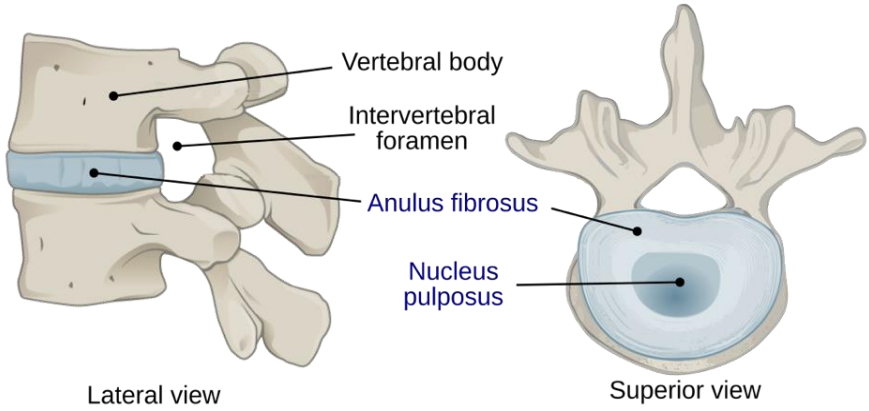


Figure 1: Intervertebral Disc

The IVD is known as the most extensive avascular tissue in the body consisting of 3 structures: the collagenous Annulus Fibrosus (AF) that surrounds the highly hydrated, gelatinous Nucleus Pulposus (NP), the cartilaginous endplate, which is the interface between the IVD and the vertebrae (Figure 1). Under physical stress, the spine is exposed to constant movement and impact. The IVD acts as a non-compressible, elastic pressure cushion that absorbs shocks in the longitudinal direction of the spine and distributes pressure evenly to all sides.

Due to degenerative changes during life, the water-binding capacity of the NP decreases. Together with cracks in the AF, it shows up radiologically as a reduction in the height of the intervertebral space. The degeneration leads to loosening of the motion segment with consecutive instability. Due to the tears in the AF, parts of the nucleus pulposus can escape from the intervertebral space, also known as disc hernia or nucleus pulposus hernia. These changes may also cause severe chronic lower back pain, which is experienced by astronauts that have displayed signs of intervertebral disc degeneration (IVDD) attributed to microgravity (Penchev et al., 2021).

Furthermore, the IVD can be counted among the bradytrophic tissues and is therefore very weakly regenerative. As degeneration progresses, the diminished buffering function of the disc leads to increased stress on the vertebral body endplates. The resulting sclerosis is referred to as osteochondrosis. Recent experimental studies have explored various biological strategies to address IVDD. They include applying growth factors and cytokines, gene therapy, scaffold, and cell transplantation of disc or mesenchymal stem cells. However, a lack of knowledge of the cellular composition of the IVD hinders advances in regenerative therapies (Illien-juenger, 2010).

1.3 Initial Situation

Calió et al. (2021) deconvoluted bovine IVD using single-cell RNA sequencing in conjunction with bulk RNA sequencing to determine the cellular composition of the NP and AF.

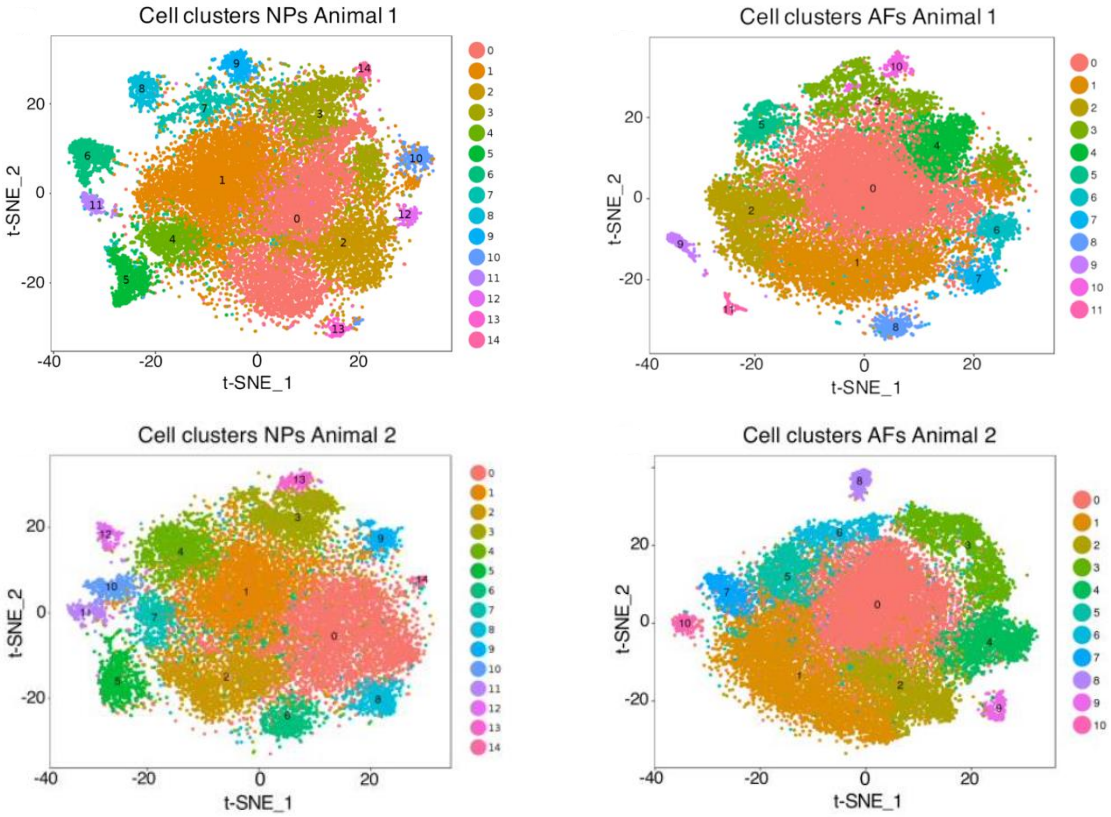


Figure 2: Cell Cluster Distribution in the Nucleus Pulposus and Annulus Fibrosus

Figure 2 depicts the cell cluster distribution between the proximal and distal NPs and AFs of animals one and two during the single-cell RNA sequencing technique. The entire analysis resulted in 15 distinct NP cell clusters along with 27 NP structure-specific genes and 12 distinct AF cell clusters, including 24 AF structure-specific genes.

For this project, cell cluster two has been selected from the 15 identified NP cell clusters for validation. Five uniquely expressed genes that are exclusively expressed in this cell type and displayed the most significant fold changes were selected for validation through gene expression analysis using quantitative polymerase chain reaction (qPCR). For this method qPCR primers will need to be designed for the five genes, which are as follows: Secreted Frizzled-Related Protein 5 (SFRP5), Adiposis Regulatory Factor (ADIRF), S100 Calcium-Binding Protein A2 (S100A2), Actin Gamma 1 (ACTG1), and Tubulin Beta Class I (TUBB).

1.4 Aim

The project aims to design and establish qPCR primers for the five uniquely expressed genes identified in cell cluster 2. The intention is that these primers will be used in the qPCR technique to validate the genes when put to the test on extracted RNA from bovine IVD.

1.5 Hypothesis

It is hypothesized that ideal qPCR primers for each uniquely expressed gene will be designed to validate the genes in cell cluster 2.

1.6 Research Questions

The following research questions are derived.

1. What are the optimum parameters to establish ideal qPCR primers?
2. Is it feasible to establish qPCR primers for each gene?
3. Do the primers amplify only the target of interest and not any unintended targets such as primer-dimers and are therefore specific?
4. Do the primers meet the acceptable range of efficiency?
5. Do the primers meet the acceptable range of linearity and are the replicates reproducible?
6. What is the difference in expression of the genes in distal IVD compared to proximal IVD?

1.7 Methodology

Firstly, a well-designed assay begins with understanding the gene of interest. Databases such as Ensembl and Genebank will be used to obtain vital information about the genes. The focus will also be to uncover published primers and utilise those for validation. If published primers are not found, they will be designed via software like NCBI BLAST and analysed for primer dimers, hairpins, and secondary structures. Afterward, cDNA will be reverse transcribed from RNA that has be isolated out of bovine IVD. Then, the primers will be tested using qPCR. The assays will be validated by conducting specificity and efficiency analysis, using the generated qPCR amplification and melt curves, gel electrophoresis, and standard curves. Relative quantification will be performed to determine the different expression levels of the genes in distal and proximal IVD from different bovine tails. Finally, the results of all the processes will be presented and discussed respectively. Ultimately, a conclusion will be drawn, taking into consideration the hypothesis.

2. Literature Review

This section provides an insight into the five uniquely expressed genes, the qPCR technique, essential factors that can affect the qPCR, and the amplification curve and threshold cycle (Ct) generated by the qPCR in instrument reporting. Afterward, the process to validate the assay through specificity and efficiency analysis will be covered. Lastly, data analysis is described.

2.1 Uniquely Expressed Genes in Cluster 2

As mentioned in 1.3, qPCR primers need to be established for the following genes. However, before beginning that process, some basics of the genes are covered here including the gene ontology (GO).

2.1.1 SFRP5

SFRP5 is a member of the SFRP family and is a protein-coding gene that functions as modulators of wingless-related integration site (Wnt) signaling through direct interaction with Wnts. Additionally, it plays a role in regulating cell growth and differentiation in specific cell types. SFRP5 could play a role in determining the polarity of photoreceptors and other retinal cells.

This gene is highly expressed in the retinal pigment epithelium, bone marrow, subcutaneous adipose tissue, vas deferens, and weakly expressed in the retina, brain, heart, liver, kidney, testis, and muscle (Chang et al., 1999). Table 1 comprises the GO.

Table 1: Gene Ontology of SFRP5

Gene Ontology of SFRP5	
Molecular function	Wnt-protein binding
Cellular component	None
Biological process	Canonical Wnt signaling pathway
	Cell differentiation
	Multicellular organism development
	Negative regulation of cell population proliferation
	DNA-binding transcription factor activity is negatively regulated.
	Protein kinase B signaling is regulated negatively.
	The Wnt signaling pathway is regulated in a negative manner.
	Negative control of the Wnt signaling pathway, which is involved in morphogenesis of the gut tract

	Regulation of BMP signaling pathway

2.1.2 ADIRF

ADIRF, also known as adipose-specific 2, is a protein-coding type of gene. Initially, it was identifiable as the second most abundant transcript in adipose tissue. It up-regulates the quantity of enhancer-binding protein alpha and PPAR-gamma while promoting adipogenic differentiation at an early stage of adipogenesis. Furthermore, during differentiation, an adipocyte lineage-specific nuclear factor could influence the master adipogenesis transcription factors. It also regulates glucose transport in adipocytes and the number of preadipocytes.

ADIRF is highly expressed in bone marrow, adipose tissue, subcutaneous adipose tissue, vas deferens, moderately expressed in brain, colon, heart, and weakly expressed in liver, muscle tissue, testis. Table 2 comprises the GO.

Table 2: Gene Ontology of ADIRF

Gene Ontology of ADIRF	
Molecular function	None
Cellular component	Nucleus
Biological process	Positive regulation of fat cell differentiation

2.1.3 S100A2

S100A2 is a protein-coding gene that contributes to cellular calcium signaling by acting as a calcium sensor and modulator. It also interacts with other proteins, such as TPR-containing proteins, and plays an indirect role in various physiological processes, including tumor cell growth suppression. S100A2 is highly expressed in the urethra, surface of the tongue, vas deferens, subcutaneous adipose tissue; moderately expressed in bone marrow, aorta, and weakly expressed in retina and hypothalamus. Table 3 comprises the GO.

Table 3: Gene Ontology of S100A2

Gene Ontology of S100A2	
Molecular function	Calcium-dependent protein binding
	Calcium ion binding
	Identical protein binding
	Transition metal ion binding
Cellular component	None

Biological process	Endothelial cell migration

2.1.4 ACTG1

ACTG1 is a highly conserved protein found in all eukaryotic cells that has a function in a range of cell motility types. It is highly expressed in the placenta, ammon's horn, lung, bone marrow, subcutaneous adipose tissue and weakly expressed in the liver and longissimus thoracis muscle. Table 4 comprises the GO.

Table 4: Gene Ontology of ACTG1

Gene Ontology of ACTG1	
Molecular function	ATP binding
	The postsynaptic actin cytoskeleton's structural constituent.
Cellular component	None
Biological process	Cellular response to interferon-gamma
	Sarcomere organization
	Synaptic vesicle endocytosis

2.1.5 TUBB

TUBB belongs to the protein tubulin family and is constituents of microtubules. It is known for binding two moles of Guanosine triphosphate (GTP) together, one from the exchangeable site on the beta chain and the other from the non-exchangeable site on the alpha chain. It is highly expressed in ammon's horn, vas deferens, subcutaneous adipose tissue, bone marrow, and weakly expressed in liver and spermatocyte. Table 5 comprises the GO.

Table 5: Gene Ontology of TUBB

Gene Ontology of TUBB	
Molecular function	GTPase activity
	GTP binding
	Structural constituent of cytoskeleton
Cellular component	Microtubule
Biological process	Microtubule cytoskeleton organization
	Mitotic cell cycle
	Regulation of synapse organization
	Spindle assembly

2.2 Quantitative Polymerase Chain Reaction

qPCR allows amplification, and simultaneous quantification of a targeted DNA strand as the reaction progresses in real-time (Bustin & Huggett, 2017).

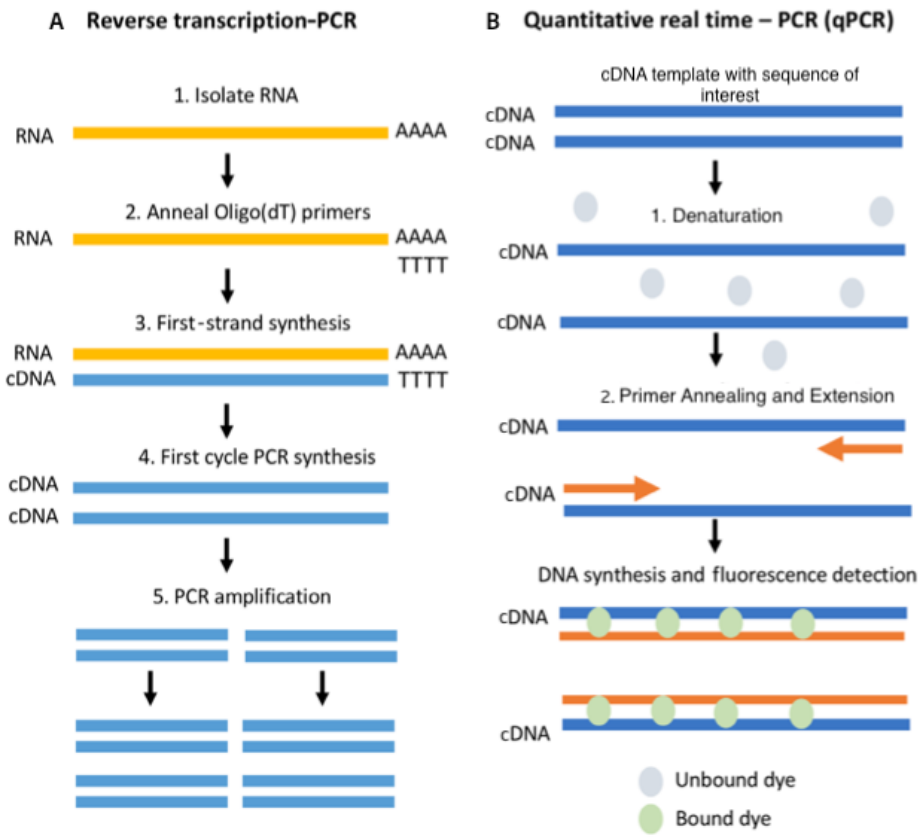


Figure 3: Two-Step Reaction (Adams, 2020)

Figure 3 provides a general overview of the two-step reaction in this project. Firstly, RNA is isolated. Thereafter complementary DNA (cDNA) is generated via reverse transcription as the template for qPCR to quantify the gene expression products (A). B starts with the double-stranded cDNA that is denatured by heating the template to separate into two single strands. Afterward, the temperature is lowered to enable the DNA primers to attach to the template and elongate. The new strand of DNA is synthesised when the DNA polymerase enzyme binds to the primer and begins adding nucleotides. This process is repeated 40 times, with each cycle doubling the quantity of DNA strands and resulting in more than 1 trillion copies (Newton & Graham, 1994).

The quantitation of the target sequence is established by measuring the emission of a fluorescent reporter dye during each cycle of the qPCR experiment - the more DNA generated in the qPCR reaction, the more detected fluorescence. SYBR® Green, double-stranded DNA

(dsDNA) binding dye, will be used because of its ease of use, cost-efficiency, and generic detection (Pabinger et al., 2014). It intercalates non-specifically into newly synthesised dsDNA and, upon binding, emits a fluorescent signal that is detected by the thermal cycler.

qPCR is known for its sensitivity, speed, and simplicity of a reaction, and it is widely accepted as the "gold standard" for gene expression analysis because of its efficiency to detect and precisely quantify the target, even at low expression levels (Rocha et al., 2016). The five uniquely genes mentioned in this project exhibit the highest expression levels and fold changes in this cell, so utilizing the qPCR for its efficiency will allow accurate validation of these genes. That is why this project has chosen qPCR as the method and basis.

2.3 Critical Factors Regarding qPCR

Technical deficiencies that affect qPCR and assay performance include the following:

a) Inadequate sample storage, preparation, and DNA or RNA quality, yielding highly variable results.
b) Inadequate reverse transcription and qPCR primers and probes, resulting in inefficient and unreliable test performance.
c) Inappropriate data and statistical analyses generate highly misleading results (Bustin et al., 2009).

Furthermore, efficient qPCR studies necessitate a proper normalization method to eliminate non-specific variance among cDNA samples. As a result, using qPCR with target genes and reference genes is critical to avoid errors in RNA extraction or contamination during sample processing (Rocha et al., 2016).
To generate consistent results and eliminate qPCR data interpretation problems, certain steps must be taken:

a) A precise primer design in terms of specificity and efficacy
b) mRNA purified to remove impurities such as carbohydrates, proteins, and phenols.
c) A careful selection of a reference gene

The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) are guidelines that have been published to assist in the process of qPCR design and execution (Bustin et al., 2009). It supplies a set of quality criteria and details 85 parameters

starting from experimental design over sample processing along with assay validation and data analysis.

2.4 Primers and Amplicons

To achieve a successful qPCR, essential properties of primers include having a balanced G/C content, similar melting temperatures at which 50% of the oligonucleotide is hybridized for the forward and reverse primer, and no preference for self-complementarity or hairpin structure. In addition, the primer pair should only amplify the target of interest, not other targets. Additionally, the pair's absolute and relative binding positions to the target sequence need to fulfill constraints. Furthermore, the primers have to harmonize with universal reaction conditions regarding the "master mix" and the reaction's time and temperature profile (Burger et al., 2018).

Subsequently, amplicons should range between 100 and 150 base pairs (bp) so that a high level of fluorescence can be obtained without compromising the PCR efficiency.

2.5 Instrument Reporting

2.5.1 Amplification Curve

The qPCR machine reads the fluorescence signal from the assays in real-time and is displayed as an amplification curve. On the y-axis is the detected fluorescence and on the x-axis is the cycle number (Figure 4). The amplification curve has an initial, exponential, and plateau phase. Quantitation is based on this curve.

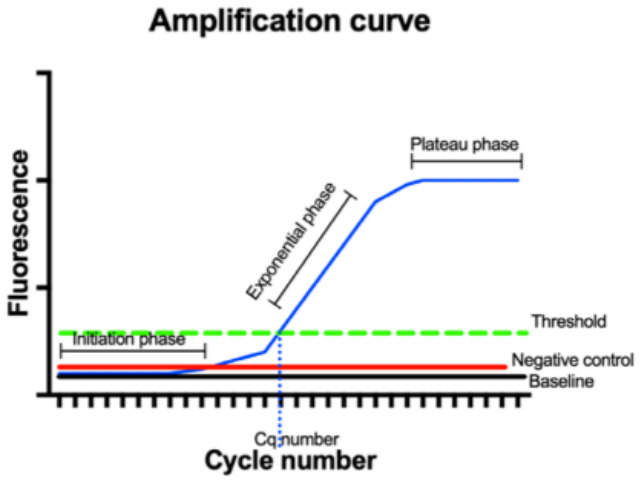


Figure 4: Amplification Curve (Adams, 2020)

The level of fluorescence is low when amplification begins, and this is utilized to establish the baseline level of fluorescence. Fluorescence reaches a much greater level than the baseline when the reaction advances towards exponential development, referred to as the threshold level. The point at which the sample exceeds this threshold is recorded as the Ct (Adams, 2020). Typically, qPCR instrument software sets the threshold at ten times the standard deviation of the baseline fluorescence value. However, the threshold can be set at any position during the exponential phase of PCR.

2.5.2 Ct Value

The Ct value is the number of cycles at which the reaction's fluorescence signal crosses the threshold. Since the Ct value is inversely proportional to the starting amount of the target, it is utilized to compute the initial DNA copy number as well to generate a standard curve after a dilution series.

2.6 Assay Validation

2.6.1 Specificity Analysis

The assay’s specificity must be validated to ensure that only the sequence of interest is detected. This is performed by combining melt curve and size analysis.

2.6.1.1 Melt Curve

The primers and reaction conditions used in a qPCR experiment determine its specificity. However, even well-designed primers can produce primer-dimers or amplify a non-specific product. When using qPCR, it is also possible that the RNA sample contains genomic DNA (gDNA), which could also be amplified (Adams, 2020). Melting curve analysis will validate the specificity of the qPCR reaction.

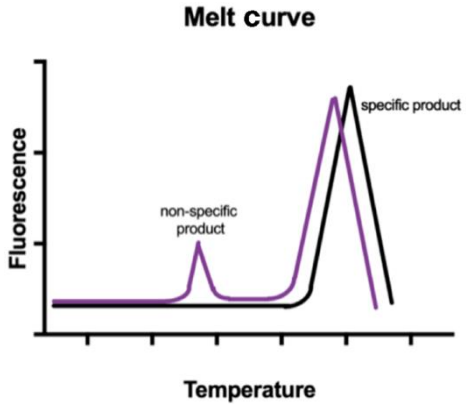


Figure 5: Melt Curve (Adams, 2020)

A melt curve plots the change in fluorescence (y-axis) observed as dsDNA containing dye, here SYBR® Green, dissociates or "melts" into single-stranded DNA (ssDNA) as the temperature (x-axis) of the reaction is elevated from 60 to 95 °C (Figure 5).

Since the temperature of dissociation is proportional to the length and composition of the amplicon, it is possible to determine how many amplification products are present in the well. A favourable melt curve should result in a unique peak. If primer-dimers are amplified, a small product is observed around 70 °C. The primer dimer peak can even be higher than the particular product's peak in some cases. The primer-dimer peak will be visible even with the no template control (NTC), confirming that the amplification is non-specific.

2.6.1.2 Gel Electrophoresis

By performing size analysis of a qPCR product on agarose gels, one can detect if any non-specific products have formed because gel electrophoresis separates DNA fragments based on the size and charge. At one end of the gel, DNA samples are placed into wells, and an electric current is used to pull them through the gel. When the gel is stained with a DNA-binding dye, the DNA fragments become visible as bands, indicating whether the correct gene has been amplified (Makovets, 2013).

2.6.2 Efficiency Analysis

A dilution series can be performed by diluting the cDNA over a certain amount of points of dilution and thereafter, generating a standard curve to measure reaction efficiency.

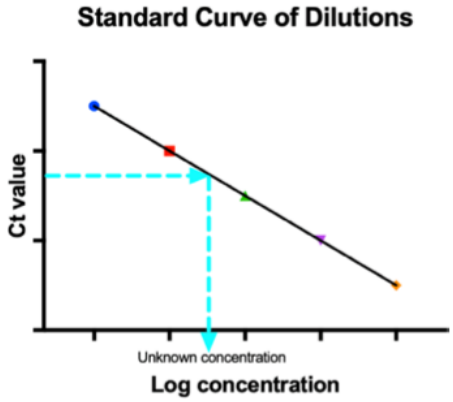


Figure 6: Standard Curve of a Dilution Series (Adams,2020)

In the dilution series, the log of each known concentration is displayed against the Ct value of that concentration (Figure 6). A slope of -3.32 indicates that the efficiency is 100% where each cycle yields a doubling of the product (Adams, 2020). The curve also indicates other information about the reaction's performance, such as slope, y-intercept, assay sensitivity, assay reproducibility and correlation coefficient.

The correlation coefficient (R^2) is determined by how well the data points lie on the line, hence the linearity of the PCR assay and where the reproducibility parameter is determined by the replicates. The standard deviation between the replicates have to be calculated.. The assay's sensitivity is determined by comparing the Ct values of the same sample at the same threshold level. The lower the Ct value is, the higher the assay sensitivity

2.7 Data Analysis

The number of cycles essential for the fluorescence signal to reach the threshold is used to measure gene expression. During amplification, a gene with a higher expression level will approach the threshold sooner than a gene with a lower expression levels. Quantification can be done either by absolute quantification or relative quantification. In absolute quantification, the copy numbers of DNA is estimated. In relative quantification, the expression of the gene of interest in a sample is expressed relative to another gene, another sample, used as a reference, to determine the amount of qPCR product generated (Schmittgen & Livak, 2008). Additionally, the difference between the Ct values is measured. The equation employed is determined by the similarities or variations in the reactions' efficiency and the number of reference genes used (Adams, 2020).

3. Methodology

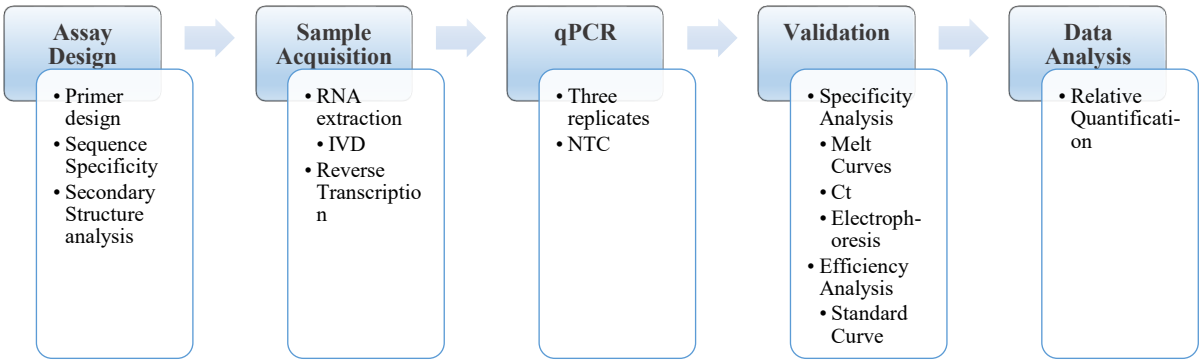


Figure 7: Methodological Process

Figure 7 displays the methodological steps that occurred during the project. This comprises of the assay design requirements, design tools and process, reverse transcription, qPCR primer testing, gel electrophoresis, 2-fold dilution series and lastly, qPCR using cDNA from 3 different bovine tails segmented into proximal and distal. RNA extraction was conducted separately and will not be featured in this report.

3.1 Assay Design Requirements

Before designing the primers, the following requirements were pre-defined:

1. The primer pairs shall have a length of 18 to 25 bp with an optimum of 20 bp.
2. The primer shall have a GC content between 30 to 80 % with an optimum of 40 to 60%.
3. The melting temperature of the primer pairs shall be between 59 to 65°C, ideally 62°C.
    a. The melting difference between the forward and reverse primer shall not be greater than 4°C.
4. Runs of identical nucleotides shall be avoided, especially of 4 or more Gs or Cs at the 3’ end.
5. Secondary structures at the binding site shall be avoided.
6. The primers shall be designed intron spanning or intron flanking to prevent amplification of contaminating gDNA.
7. Complementarities between the primers shall be avoided to avoid primer dimers, especially of 2 or more bases at the 3’ end of the primers.
    a. Self dimer for both primers: avoid if maximum delta G is less than -9 kcal/mole.
    b. Hetero dimer: avoid if delta G is less than -9 kcal/mole

    c. Delta G = 10 is also tolerated.
8. Complementarities within the primer shall be avoided to avoid hairpins.
    a. Avoid if delta G is less than -9 kcal/mole.
9. Mismatches between the primers and targets shall be avoided, especially towards the 3’ end of the primer.
10. At the 3’ end, a T base shall be avoided.
11. The primers shall be unique and specific utilizing the NCBI BLAST search.

Additionally, the following requirements for the amplicon were established:

1. The amplicon shall have a length of 80 to 150 bp; 80 to 200 bp is also tolerated.
2. The amplicon shall have a G/C content between 30 to 80 % with an optimum of 40 to 60 %.
3. Long repeats of single bases, hence more than 4 of the identical nucleotides, shall be avoided.
4. Avoid stable secondary structures at the primer binding sites. Check with mfold, with a setting of 50 mM Na+, 3 mM Mg2+ at a primer annealing at 60 °C.

3.2 Design Tools and Process

The primer design software used during this process comprises NCBI Primer-BLAST, IDT's OligoAnalyzer, and UNAFold's mfold web server. These tools include algorithms for designing primers that are specifically optimized for qPCR, ensuring that the primer pairs are specific for the target sequence, free of internal secondary structures, and avoid complementary hybridization at the 3' ends of each primer and among themselves.

The first step involved accumulating as much information as possible from sequence databases like NCBI and Ensemble about the genes. These databases segment their sequences based on genomic differences and nomenclature. Here, the Bos taurus mRNA sequences, prefixed as NM_, for the specified genes on NCBI were located. The desired target sequence could directly be transferred to the NCBI BLAST as the template, and that parameters were manually adjusted to fulfill the requirements mentioned in 3.1. After that, multiple sets of primer pairs were generated.

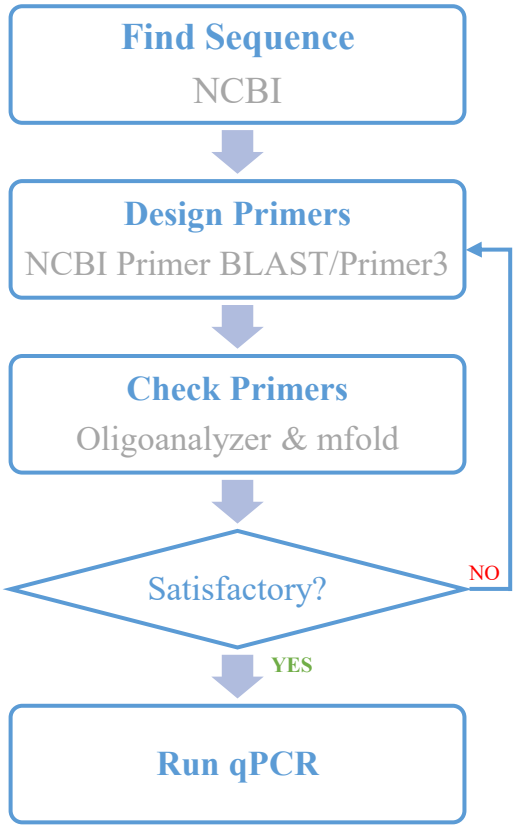


Figure 8: Primer Design Process

Subsequently, secondary structure analysis was performed on the primer pairs to test for potential harmful secondary structures, hairpins, and primer dimers that could hamper the performance of the qPCR. Tools used for this part were OligoAnalyzer and mfold.

Figure 8 displays the process graphically. If the primer pairs did not satisfy the pre-defined requirements, the process reverted to step two, where the design parameters would be adjusted or optimized to generate primer pairs for the gene.

Lastly, two of the best primers pairs for each gene that fulfilled the requirements were selected. However, only one was ordered and tested in the process mentioned below. These are mentioned in 4.3.

3.3 Reverse Transcription

RNA extracted out of the bovine IVD was reverse transcribed into cDNA using SuperScript III reverse transcriptase and 20-fold dilution in this process step. Furthermore, Hexanucleotide random primers were used to provide an equal representation of all targets in the qPCR application. A detailed protocol and reagents used to convert RNA into cDNA is provided in Appendix A: Reverse Transcription Protocol.

3.4 Dilute Primers

The primers arrived as soluble pellets that had to be solubilized by a specific amount of water before further use in the qPCR reaction. Once the entire primer was solubilzed with water, 10 µl of the stock primer was diluted further with 90 µl ddH_2O and immediately used for primer testing. A detailed description of this process is provided in Appendix B: qPCR Primer Testing Protocol.

3.5 Primer testing

In this process, three replicates of SFRP5, ADIRF and S100A2, were prepared in conjunction with an NTC. Furthermore, normalization to the two reference genes, 18S and B2M, was used as a correction method for the qPCR reaction. The master mix composed of ddH_2O, SYBR® Green dye, and the forward and reverse primer resulted in 16 µl and then 4 µl cDNA was added to each master mix in the wells. A detailed protocol and reagents used to prepare the samples for this step are provided in Appendix B: qPCR Primer Testing Protocol.

The thermal profile started with denaturation at 95 °C, followed by 40 PCR-cycles including three steps: denaturation at 95 °C, annealing of primer at 60 °C, elongation of DNA-polymerase at 72 °C, and measurement of fluorescence, and ended with melt curve analysis consisting of denaturation at 95 °C, and increment steps from 65 °C to 95 °C. The qPCR run was performed on the iQ5 Bio-Rad machine. Amplification and melt curves were generated and the threshold was set to approximately 100 RFU.

3.6 Gel Electrophoresis

After the qPCR, gel electrophoresis was used to check the assay's specificity. This process involved loading the samples from qPCR 96-well plate and a DNA-binding dye into the gel wells. The DNA ladder, a standard reference that contains DNA fragments of known length, was placed in the first well. After that, a current of 65 mA and ran through the gel while a voltage of ≅ 100 V was applied. After that, the samples were identifiable as bands, and the size could be determined. A detailed protocol and reagents used in this step are provided in Appendix C: Gel Electrophoresis Protocol.

3.7 Dilution Series

A 2-fold dilution series using the initial 20-fold cDNA reverse transcribed for the primer testing was carried out to establish a standard curve for ADIRF and S100A2. Furthermore, to assess the performance of the reaction. The cDNA utilised in this process had an initial concentration of 5% and included 5 points of dilution in three replicates. A detailed description of the protocol is provided in Appendix D: Dilutions Series Protocol. The same thermal cycling protocol and iQ5 Bio-Rad machine mentioned above in 3.5 was utilised in this step. Amplification and melt curves were generated and the threshold was set to approximately 100 RFU. After dilution was carried out, a standard curve was generated by taking the log of each known concentration in the dilution series on the x-axis and plotting it against the Ct value for the concentration on the

y-axis. Information about the assay's efficiency, linearity, sensitivity, and reproducibility from the standard curve could be determined.

The PCR efficiency was determined by applying the slope of the standard curve into the following equation:

$$Efficiency = 10^{\left(-\frac{1}{slope}\right)} - 1$$

- PCR efficiency between 90 – 110% is accepted, corresponding to a slope ranging between - 3.1 and -3.58. However, a PCR efficiency ≥ 87% is tolerated.
- An R^2 greater than 0.985 is acceptable and the reproducibility parameter was determined by the replicates. The standard deviation between the replicates was calculated. Replicates that have more than 0.5 Ct of difference should be avoided.
- The assay's sensitivity was determined by comparing the Ct values of the same sample at the same threshold level.

3.8 Relative Quantification

RNA extracted out of 3 different bovine IVD were segmented into proximal and distal. IVD's 1 to 4 was grouped together representing proximal and IVD's 5 to 8 representing distal (Table 6). Thereafter, the RNA was reverse transcribed into 6 pools of cDNA using SuperScript III reverse transcriptase and further diluted to an end concentration of 1.25 ng/µl. Furthermore, Hexanucleotide random primers were used to provide an equal representation of all targets in the qPCR application. A detailed protocol and reagents used to convert RNA into cDNA is provided in Appendix F: qPCR with 6 pools of cDNA.

Table 6: Segmentation of RNA

	Tail 1	**Tail 2**	Tail 3
Proximal	IVD 1 – 4	IVD 1 – 4	IVD 1 – 4
Distal	IVD 5 – 8	IVD 5 – 8	IVD 5 – 8

Thereafter, three replicates of ADIRF and S100A2 in conjunction with an NTC were prepared. Furthermore, normalization to the two reference genes, 18S and B2M, and a control, proximal, was used as a correction method for the qPCR reaction. The master mix composed of ddH_2O, SYBR® Green dye, and the forward and reverse primer resulted in 16 µl and then 4 µl cDNA

was added to each master mix in the wells. A detailed protocol and reagents used to prepare the samples for this step are provided in Appendix F: qPCR with 6 pools of cDNA. The same thermal cycling protocol and iQ5 Bio-Rad machine mentioned above in primer testing was utilised in this step. Amplification and melt curves were generated and the threshold was set to 100 RFU.

The generated Ct values were used in data analysis for relative quantification, where the expression of ADIRF and S100A2 in the cDNA from distal IVD is expressed relatively to 18S and B2M in the cDNA from proximal IVD. The raw data ran through a pre-established software that grouped the data where r represented the reference gene, g the target gene, s the sample (distal in this case), and c the control (proximal in this case). After a few steps of filtration, the gene expression would be calculated. This also included the calculating the amplification efficiency and Ct values of 18S, B2M, ADIRF and S100A2. Subsequently, the mean and stand deviation would be calculated for the ADIRF and S100A2. The mean indicated the expression ratio. These values were then used to generate a barplot representing the ratio in gene expression differences distal to proximal.

4. Results

4.1 Primer Pairs

Three Primer pairs were developed, fulfilling the requirements; for example, primer pairs could be developed for the genes SFRP5, ADIRF, and S100A2 (Table 7). However, for ACTG1 and TUBB, no primer pairs could be generated that fulfilled the requirements mentioned in 3.1.

Table 7: qPCR Primers List

qPCR Primers used to assess gene expression										
Gene symbol	**Accession number**	**Primer Pair**	**Sequence**	**bp**	**T_m (°C)**	**G/C %**	**Amplicon length**	**Primer dimer (kcal/mole)**		
								Hairpin	**Self-dimer**	**Hetero-dimer**
SFRP5	NM_174461.3	1	FP: AACGACCTCTGCATCGCTGT RP: ACCACGAAGTCGCTGGAACA	20 20	62.16 62.02	55 55	137	- 1.5 - 3.16, - 1.67	- 7.05 - 5.19	- 6.53
		3	FP: AGCAGATGTGTTCCAGCGACT RP: CCTCTTGGTGTCCTTGCGCT	21 20	62.05 62.74	52.38 60	143	-0.32, 0.13, 0.47, 0.55, 0.55 -0.05, 0.13, 0.34	- 3.61 - 9.89	- 8.3
ADIRF	NM_001114513.2	3	FP: ACCCTTAACCACTCGCAGACC RP: TGCTTCTGTGGCCTGATCCA	21 20	62.05 61.49	57.14 55	145	-0.15, -0.01, 0.01, 0.15 -0.66, -0.5, -0.3	- 4.85 - 9.28	- 6.37
		4	FP: GGCTACCCTTAACCACTCGCA RP: TTCTGTGGCCTGATCCACCA	21 20	62.11 61.13	57.14 55	146	-0.1, 0.15, 0.21 -3	- 4.85 - 9.28	- 6.37
S100A2	NM_001034367.2	1	FP: GAGTCTCTCTCTGCCCTCTGG RP: GAGTACTTGTGGAAGGTGGCG	21 21	61.02 60.94	61.9 57.14	97	-1.46, -0.47 -1.14	- 3.17 - 6.84	- 6.21
		3	FP: TTCCTGAGTCTCTCTCTGCCCT RP: GAGTACTTGTGGAAGGTGGCGA	22 22	61.97 62.24	54.55 54.55	102	- 2.02 - 1.14	- 3.17 - 6.84	- 6.59

4.2 Secondary Structure Analysis of the Amplicons

When analysing the amplicons for secondary structures at the primer binding sites, mfold generated the circular structure plots below.

4.2.1 SFRP5

These are the circular structure plots for primer pairs one and three of SFRP5.

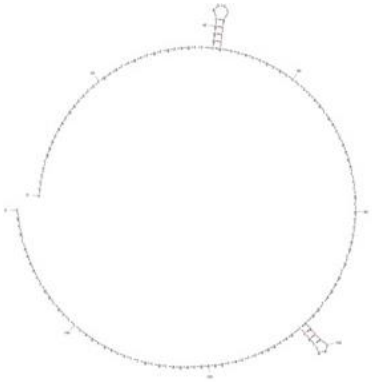


Figure 9: SFRP5 | Primer Pair 1| dG = -2.67 kcal/mole

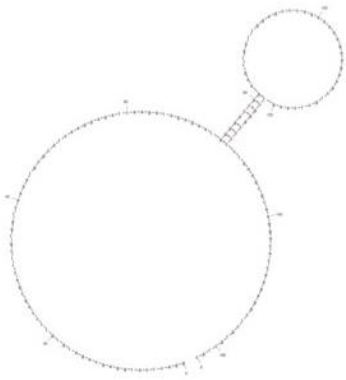


Figure 10: SFRP5 | Primer Pair 3 | dG = -0.45 kcal/mole

No secondary structures appeared at the primer binding sites of either primer pairs and delta G value was a relatively high negative value, less negative, indicating that if secondary structures were to form it would not be very stable and annealing of the primers would not be hampered.

4.2.2 ADIRF

These were the circular structure plots for primer pairs three and four of ADIRF.

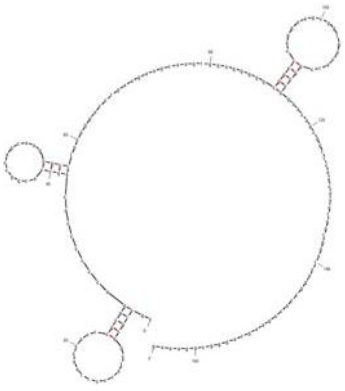


Figure 11: ADIRF | Primer Pair 3 | dG = -1.74 kcal/mole

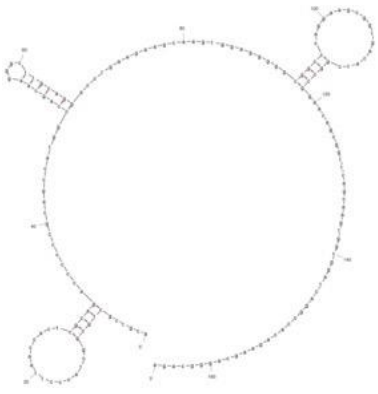


Figure 12: ADIRF | Primer Pair 4 | dG = -1.82

Secondary structures appeared at the primer binding sites in primer pair three and four. However, delta G was less negative for both, indicating that if secondary structures were to form it would not be very stable and annealing of the primers would not be hampered.

4.2.3 S100A2

Figure 13 and Figure 14 were the circular structure plots for primer pairs one and three of S100A2.

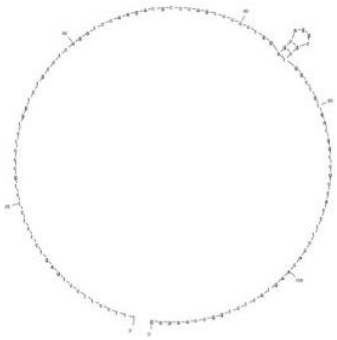


Figure 13: S100A2 | Primer Pair 1 | dG = 0.00 kcal/mole

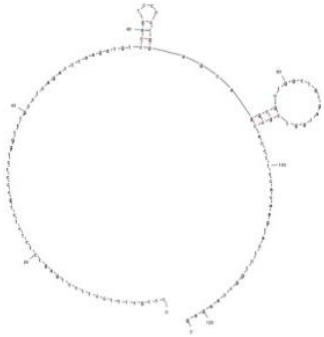


Figure 14: S100A2 | Primer Pair 3 | dG = -0.50 kcal/mole

No secondary structures appeared at the primer binding sites of either primer pairs and delta G is less negative for both, indicating that if secondary structures were to form it would not be very stable and annealing of the primers would not be hampered.

4.3 Primer Testing

As mentioned above, in 3.2, only one primer pair each was ordered to be tested in this process these were primer pairs one, three and one respectively for SFRP5, ADIRF, and S100A2. Amplification, melt and melt peak curves were generated. The amplification curve displays the fluorescence detected against cycles. The melt curve displays the fluorescence against temperature. To achieve a clear view of the melting dynamics, the -ΔF/ΔT (change in fluorescence/change in temperature) curve was plotted against temperature for the melt peak curve.

4.3.1 SFRP5

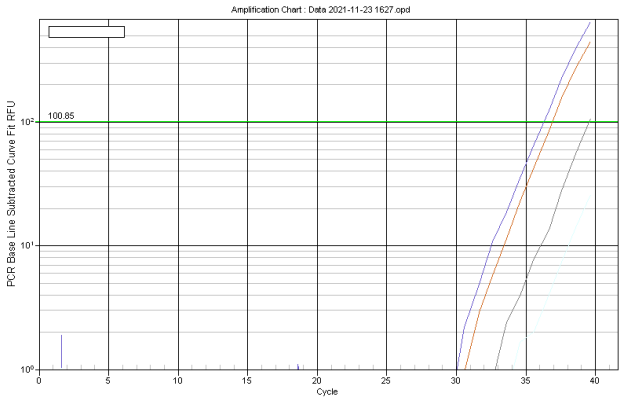


Figure 15 a: SFRP5 | Amplification Curve

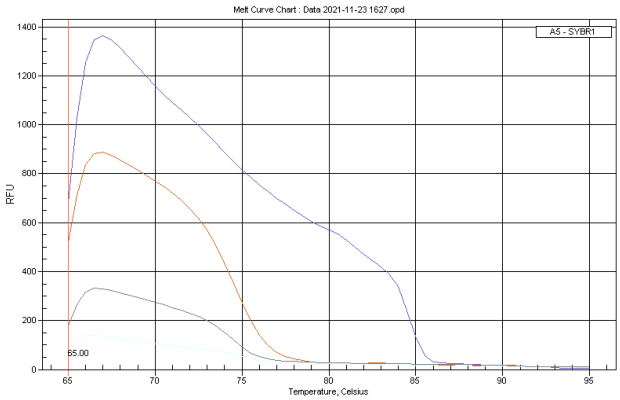


Figure 15 b: SFRP5 | Melt Curve

Only two out three replicates crossed the threshold at cycles 37.64 and 38.25, whereas the third replicate did not cross the threshold (Figure 15 a). In addition, the three replicates do not dissociate evenly into ssDNA, as a wide spacing between curves can be observed (Figure 15 b).

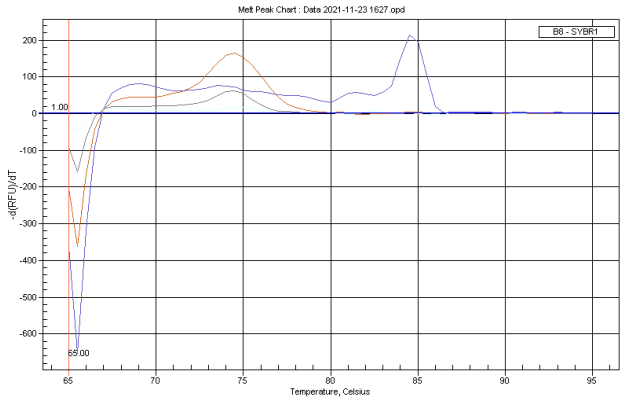


Figure 15 c: SFRP5 | Melt Peak Curve

The melt peak curve showed that primer-dimers were amplified as there is no unique peak rather a peak at around 74 °C (Figure 15 c).

SFRP5 was tested a second time to determine if the result would be the same as the first.

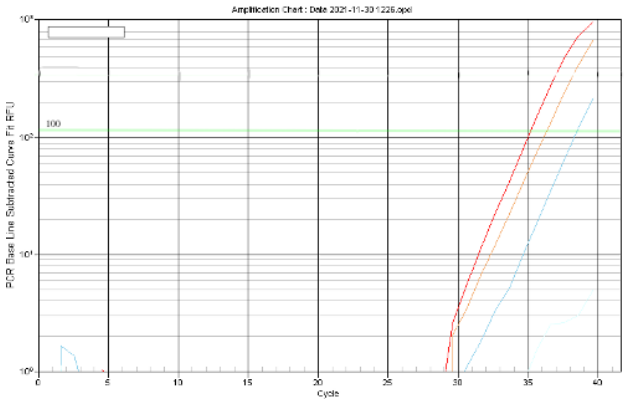


Figure 16 a: SFRP5 | Amplification Curve

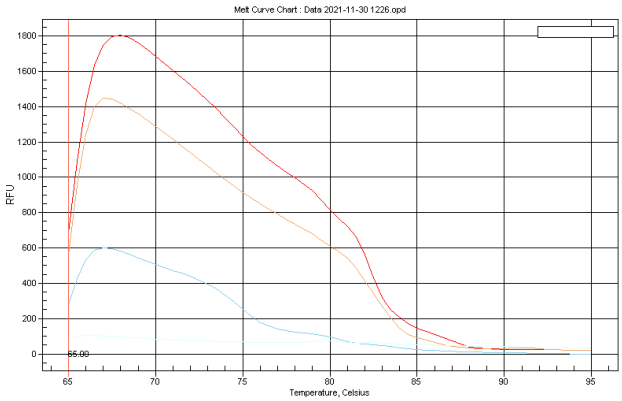


Figure 16 b: SFRP5 | Melt Curve 2

All three replicates passed the threshold and had Ct values of 38.24, 34.87 and 35.95 (Figure 16 a). Again, the three replicates do not dissociate evenly into ssDNA, as a wide spacing between curves can be observed (Figure 16 b).

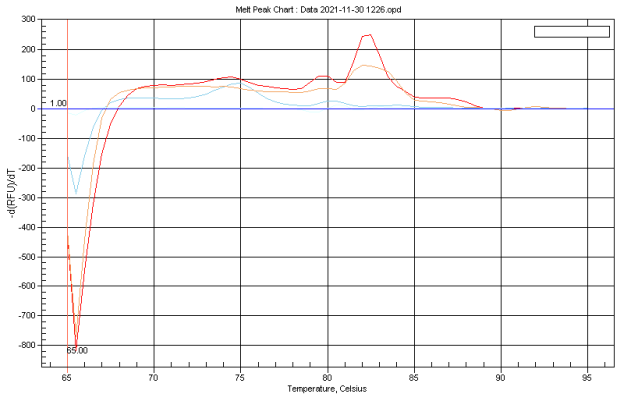


Figure 16 c: SFRP5 | Melt Peak Curve 2

The melt peak curve was less jumpy than the first primer test. However, there was no unique peak to be found, indicating instability and a strong likelihood that primer-dimers were enhanced at 75 °C (Figure 16 c).

4.3.2 ADIRF

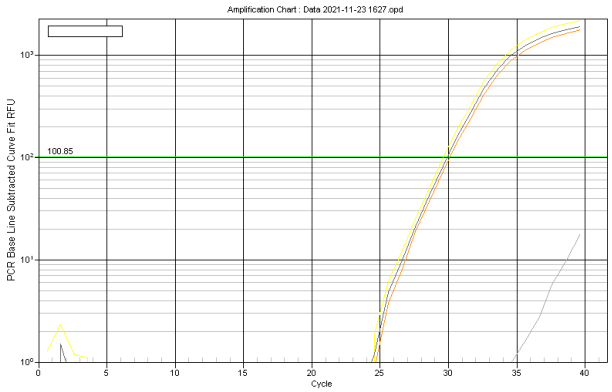


Figure 17 a: ADIRF | Amplification Curve

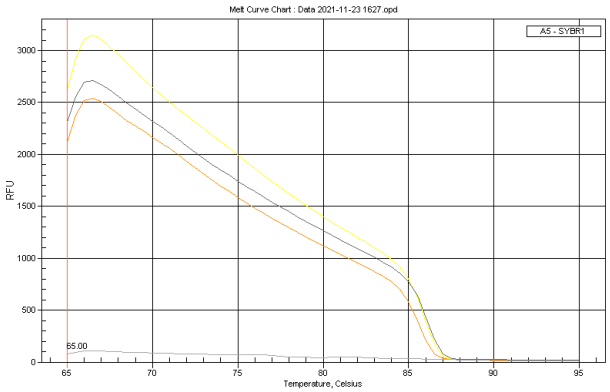


Figure 17 b: ADIRF | Melt Curve

All three replicates passed the threshold at cycles 30.98, 31.58, and 31.28 (Figure 17 a). The melt curve showed all three curves running parallel to each other. The three replicates uniformly dissociate into ssDNA, and there is little spacing between curves (Figure 17 b).

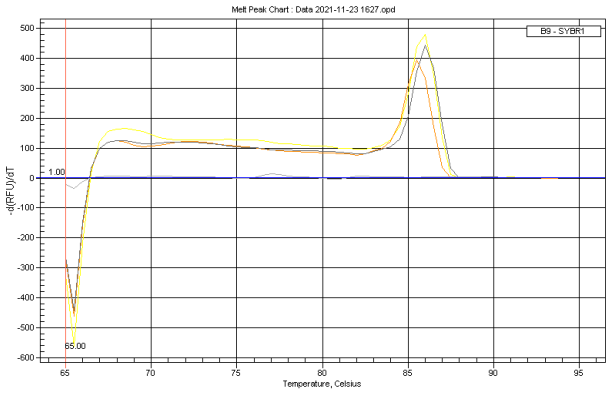


Figure 17 c: ADIRF | Melt Peak Curve

The melt peak curve showed that no primer-dimers were amplified and that all curves reach a unique peak at approximately 81.5 °C (Figure 17 c).
ADIRF was also tested a second time, with results that were quite similar to the first.

4.3.3 S100A2

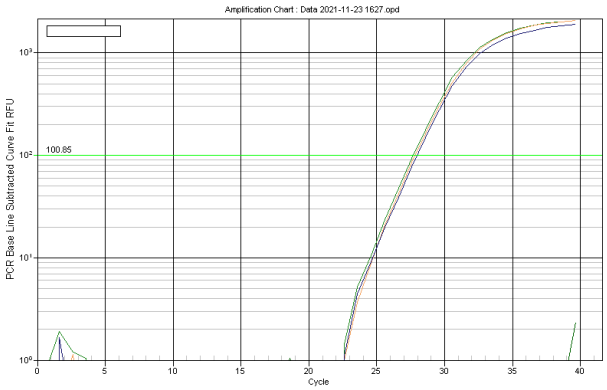


Figure 18 a: S100A2 | Amplification Curve

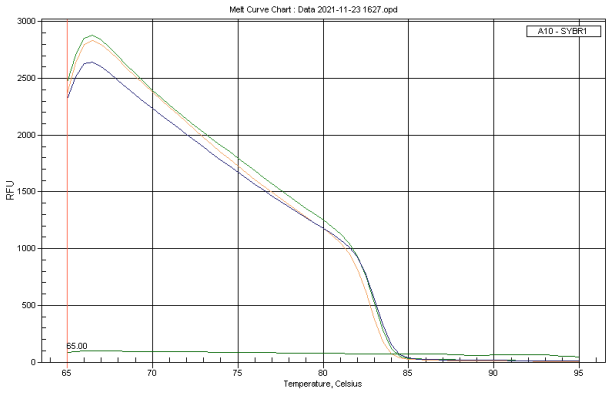


Figure 18 b: S100A2 | Melt Curve

All three replicates passed the threshold at cycles 28.98, 29.12, and 29.27 (Figure 18 a). The melt curve shows all three curves running parallel to each other. The three replicates uniformly dissociate into ssDNA, and there is little spacing between curves (Figure 18 b).

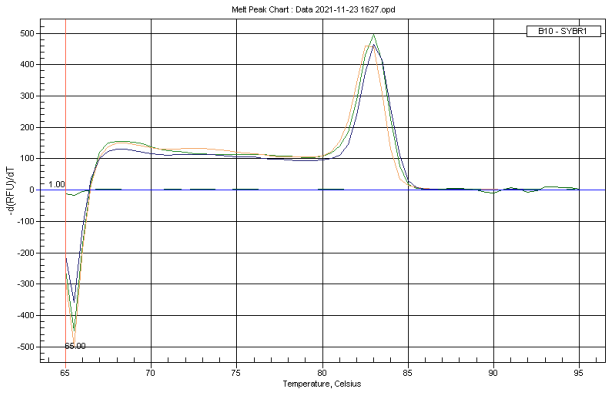


Figure 18 c: S100A2 | Melt Peak Curve

The melt peak curve showed that no primer-dimers were amplified and that all curves reach a unique peak at approximately 83°C.

S100A2 was also tested a second time, with results that were quite similar to the first.

4.3.4 Gel Electrophoresis

The ladder, which is prominently marked for referencing in the first well on the left, is made up of ten chromatography-purified individual DNA fragments in the following bp: 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100. After that, triplicate SFRP5 ADIRF and were loaded in the wells.

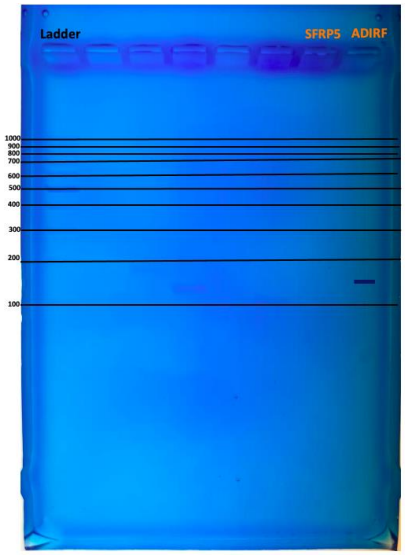


Figure 19 a: Gel Electrophoresis | Primer Test 1

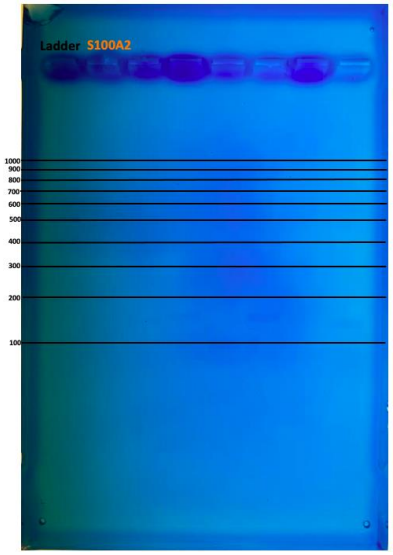


Figure 19 b: Gel Electrophoresis | Primer Test 1

SFRP5 and ADIRF are located in the seventh and eighth well of the first gel (Figure 19 a), and S100A2 is located in the second well of the second gel (Figure 19 b). For SFRP5 and S100A2,

no bands were detected in the first electrophoresis test. A band at around 150 bp was detected for ADIRF.

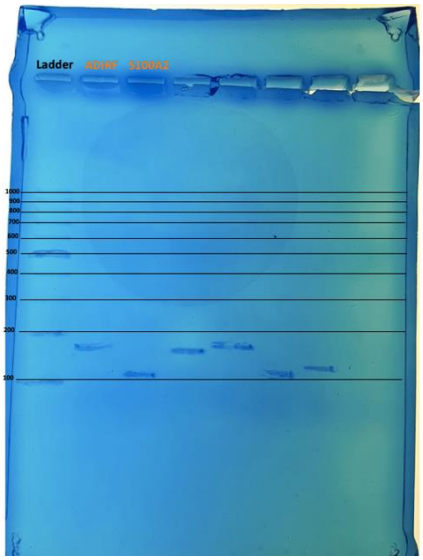


Figure 19 c: Gel Electrophoresis | Primer Test 2

In the second electrophoresis test, ADIRF and S100A2 are located in the second and third wells of the gel (Figure 19 c). A band at around 150 bp was detected for ADIRF, and a band at approximately 100 bp was detected.

4.4 Standard Curves

The standard curves display the log of the concentration against the Ct values, derived after the dilution series only for ADIRF and S100A2 as they are working primers that did not amplify primer-dimers. The individual curves from the different dilution points of for both ADIRF and S100A2 can be found in Appendix D: Dilutions Series Protocol.

4.4.1 ADIRF

The log of concentration scaled from 0 to minus 3 in increments of 0.5, and the Ct values 0 to 36 but focus on cycles 29 to 36. The blue points represent the ADIRF samples of the 6 points of dilution, and the trendline represents the linear correlation between samples.

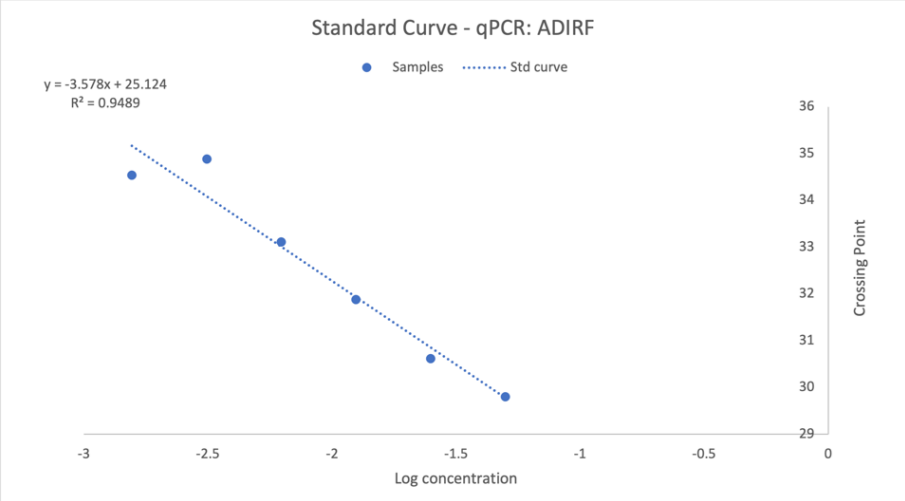


Figure 20 a: ADIRF | Standard Curve

The standard curve of ADIRF revealed a slope of -3.57 and a y-intercept of 25.12. Therefore, R^2 was approximately 0.95, as two points between -2.5 and -3 do not lie directly on the line (Figure 20 a).

$$Efficiency = 10^{\left(-\frac{1}{slope}\right)} - 1 = \mathbf{90.32\%}$$

An efficiency of 90.32 % of the ADIRF assay was calculated.

4.4.2 S100A2

The log of concentration scaled from 0 to minus 3 in increments of 0.5, and the Ct values 0 to 35 in increments of 5. The blue points represent the S100A2 samples of the 6 points of dilution, and the trendline represents the linear correlation between samples.

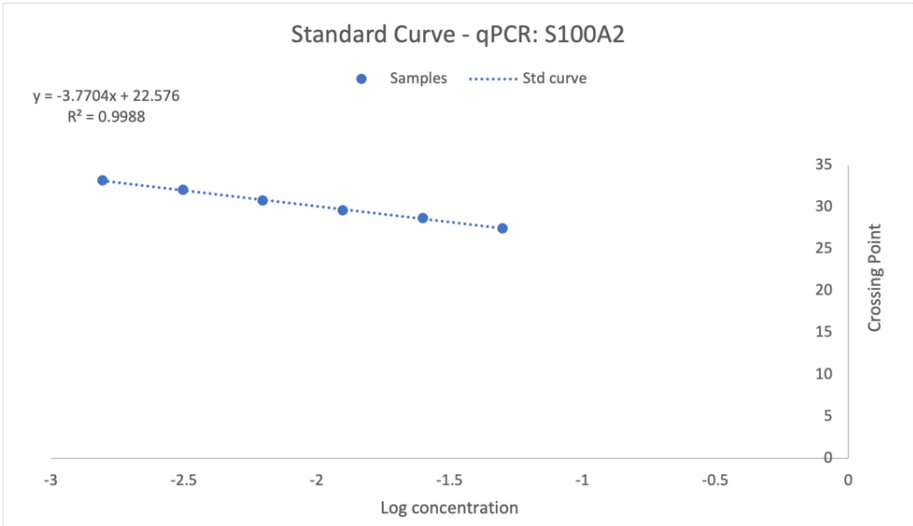


Figure 20 b: S100A2 | Standard Curve

The standard curve of S100A2 revealed a slope of -3.77 and a y-intercept of 22.57. Therefore, R^2 was approximately 0.99, as the samples lie directly on the line (Figure 20 b).

$$Efficiency = 10^{\left(-\frac{1}{slope}\right)} - 1 = \mathbf{84.17\%}$$

An efficiency of 84.17 % of the S100A2 assay was calculated.

4.5 Relative Quantification

In relative quantification, the expression of ADIRF and S100A2 in the cDNA from distal IVD was expressed relatively to 18S and B2M in the cDNA from proximal IVD. The mean and standard deviation were used to generate a barplot representing the gene expression differences of ADIRF and S100A2 from the three different tails, distal to proximal (Table 8).

Table 8: Expression Ratio of ADIRF and S100A2 | Animal 1 to 3

	ADIRF_A1	**S100A2_A1**	ADIRF_A2	S100A2_A2	ADIRF_A3	S100A2_A3
Mean	0.8841226	0.16338487	0.9796285	0.9751257	1.102507	1.4893927
Std.dev	0.0721531	0.06659597	0.1238958	0.1114421	0.215409	0.3403629

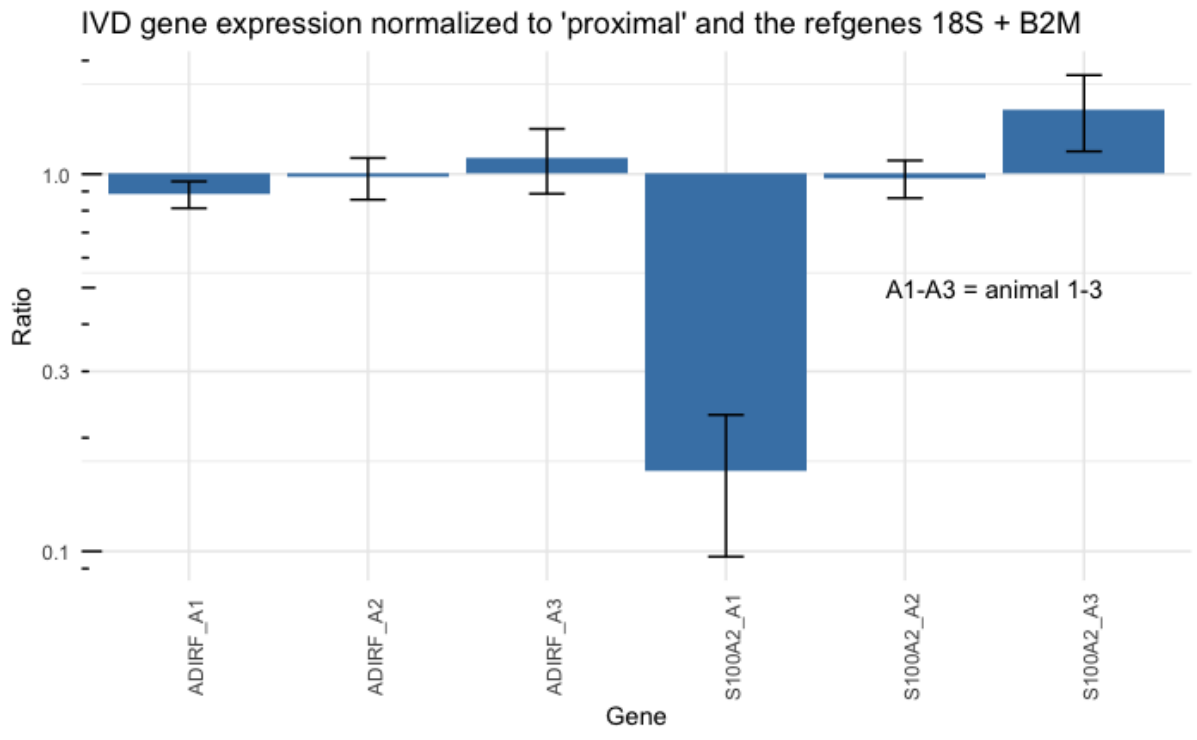


Figure 21: IVD Gene Expression Normalized to "Proximal" and the reference genes 18S + B2M

The ratio at logarithm base 10, from 0.1 to 1.0, was plotted on the y-axis. Furthermore, the genes were plotted on the x-axis in Gene_Animal (Figure 21). Each gene's bar height is equal to the mean also know as ratio. The error bars indicate one standard deviation above and below the tops of the bars. The baseline was set at 1.0.

The barplot displayed the fold increase and decrease of ADIRF and S100A2 in the test sample, distal, relative to the control sample, proximal, and was normalised to the expression of 18S and B2M. The results indicated that ADIRF had an expression ratio of 0.88, 0.97 and 1.10 for animals 1 to 3. Additional, S100A2 had an expression ratio of 0.16, 0.97 and 1.48 for animals 1 to 3.

5. Discussion

The purpose of this project is to design and establish qPCR primers that would validate the uniquely expressed genes SFRP5, ADIRF, S100A2, ACTG1 and TUBB from cell cluster 2. The following section discusses the results in relation to the research questions that were set for the project.

The research questions were:

1. What are the optimum parameters to establish ideal qPCR primers?
2. Is it feasible to establish qPCR primers for each gene?
3. Do the primers amplify only the target of interest and not any unintended targets such as primer-dimers and are therefore specific?
4. Do the primers meet the acceptable range of efficiency?
5. Do the primers meet the acceptable range of linearity and are the replicates reproducible?
6. What is the difference in expression of the genes in distal IVD compared to proximal IVD?

The optimum parameters to establish ideal qPCR primers are mentioned in 3.1. They cover pre-defined requirements for primer and amplicon that had to be adhered.

Primers for SFRP5 established by Zhao et al. (2009), Mahmoud et al. (2021), Shi et al. (2021), and Wu et al. (2017), primers for S100A2 established by in Ying et al. (2016) and Forcella et al. (2020), and primers for ACTG1 by Gao et al. (2019) were created for human cells and could not be used due to species constraints. The primers established for SFRP5 by Tan et al. (2006) for ACTG1 by Hamdi et al. (2019) and Bernier-Dodier et al. (2010), and for TUBB by Somers et al. (2006) were intended for bovine cells. However, the amplicon length exceeded the limtation of 200 bp and in some cases these primers could not be traced via their job ID on the NCBI BLAST website. An amplicon length over 200 bp would reduce the efficiency of the qPCR. Additionally, no published primers ADIRF that could be found. Since no ideal published primers could be found, they were established from scratch.

It was feasible to establish primers for genes SFRP5, ADIRF and S100A2. However, no primer pairs could not be developed for genes ACTG1 and TUBB. The reason for this is that for ACTG1, after carrying out the assay process exactly as in 3.2, no primers were generated on

the NCBI BLAST website. However, for TUBB all ten primer pairs that were generated would potentially bind to unintended templates. Hence the primers were non-specific, and the chance of primer-dimer formation would be high. Therefore, it was not feasible to qPCR primers for ACTG1 and TUBB.

From the primer testing one and two, it was evident that SFRP5 amplified unintended targets, even though no secondary structure were observed on the primer binding sites in Figure 13 of the secondary structure analysis. In Figure 15 c and Figure 16 c, a small product was visible at 74.5 °C, indicating that a non-specific product had been amplified, and not just the target of interest. Hence the primer pair for SFRP5 was not specific. Furthermore, no bands were visible after gel electrophoresis.
In regards to ADIRF, the melt peak curve (Figure 17 c) displayed a unique peak at 81.5 °C, indicating that only the target of interest had been amplified and no non-specific products. This primer pair was specfic. Even though the band, at approximately 150 bp after gel electrophoresis, appeared slightly higher than anticipated, the amplicon length of 147bp for ADIRF was validated. Moreover, S100A2 displayed a single peak at 83°C on the melt peak curve (Figure 18 a), showing that only the target of interest had been amplified and no non-specific products. This primer was specific as well. Again, the band was slightly higher than expected, at approximately 100 bp after gel electrophoresis, the amplicon length of 97 bp for S100A2 was validated.

Since SFRP5 was disregarded as a non-working primer, the dilution series was only conducted with ADIRF and S100A2. The accepted range of efficiency is between 90 and 110 percent. The efficiency of ADIRF was determined to be 90.32 percent. This primer satisfied the requirement. S100A2, on the other hand, had an efficiency of 84.17 percent and hence failed to fulfil the requirement. If an assay has a low efficiency, the reliability of the experiment results deminishes and affects assay sensitivity.

Furthermore, an R^2 greater than 0.985 is acceptable. ADIRF's linearity was determined to be 0.9489 and indicated that the replicates are non-reproducible. The condition was not met by this primer. S100A2, on the other hand, met the criteria with a linearity of 0.9988 and indicated that the replicates were reproducible.

The final qPCR conducted was to determine the difference in expression of ADIRF and S100A2 in distal IVD compared to proximal IVD through relative quantification (Figure 21). These results indicated that distal IVD from animals 1 and 2, expressed ADIRF at a 0.88-fold and 0.97-fold lower level than proximal IVD. However, distal IVD from animal 3 expressed ADIRF at a 1.10-fold higher level than proximal IVD. In regards to S100A2, distal IVD from animals 1 and 2 expressed the gene at a 0.16-fold and 0.97-fold lower level than proximal IVD. Lastly, distal IVD from animal 3 expressed S100A2 at a 1.48-fold higher level than proximal IVD.

All of the research questions could be answered during the project. Subsequently, the outcome signifies the occurrence and activity of ADIRF and S100A2 during the gene expression analysis. Hence, the presents of ADIRF and S100A2 in RNA from bovine IVD is validated.

6. Conclusion

The aim of the project was to design and establish qPCR primers for the five uniquely expressed genes identified in cell cluster 2. The intention was that these primers would be used for qPCR to validate the genes when put to the test on extracted RNA from bovine IVD. The aim could not be met for all the genes during the project.

Furthermore, it was hypothesised that ideal qPCR primers for each uniquely expressed gene would be designed to validate the genes in cell cluster 2. A practical design process has been established, together with the optimum parameters for the primers. The outcome indicates that it was feasible to establish primers for SFRP5, ADIRF and S100A2 but not for ACTG1 and TUBB. The hypothesis is partially confirmed.

Validating the genes included ruling out non-specific amplification via specificity analysis and ensuring that the signal emitted originated from the expected target. For ADIRF and S100A2 primers, this is apparent. There were no non-specific products amplified. Unfortunately, a minor product at 74°C was amplified for SFRP5. The primer pair can be optimised to reduce primer-dimer formation. In addition, the efficiency had to be examined to ensure that the assays met the requirements of 90 -110 percent. After a 2-fold dilution series, the efficiency of ADIRF and S100A2 was assessed. The efficiency of ADIRF was estimated to be 90.32 percent and that of S100A2 to be 84.78 percent. Efficiency can be improved by better design and specificity of the assay or optimisation.
Subsequently, two out of three samples indicate that both ADIRF and S100A2 is expressed lower in distal IVD than proximal IVD.

To conclude, ADIRF is the only gene out of the five uniquely expressed genes that meets validation criteria. However, the findings signify the occurrence and activity of ADIRF and S100A2 during the gene expression analysis. Hence, the presents of ADIRF and S100A2 in RNA from bovine IVD can be validated.

List of Figures

List of Tables

References

Adams, G. (2020). A beginner's guide to RT-PCR, qPCR and RT-qPCR. *The Biochemist*, *42*. https://doi.org/10.1042/BIO20200034

Bernier-Dodier, P., Delbecchi, L., Wagner, G. F., Talbot, B. G., & Lacasse, P. (2010). Effect of milking frequency on lactation persistency and mammary gland remodeling in mid-lactation cows1. *Journal of Dairy Science*, *93*(2), 555–564. https://doi.org/10.3168/jds.2009-2320

Burger, F., Angioni, M., Russo, G., Schad, M., & Kallarackal, J. (2018). PCRdrive: The largest qPCR assay archive to date and endless potential for lab workflow revitalization. *BMC Bioinformatics*, *19*(1), 447. https://doi.org/10.1186/s12859-018-2452-3

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., & Wittwer, C. T. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, *55*(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Bustin, S., & Huggett, J. (2017). QPCR primer design revisited. *Biomolecular Detection and Quantification*, *14*, 19–28. https://doi.org/10.1016/j.bdq.2017.11.001

Calió, M., Gantenbein, B., Egli, M., Poveda, L., & Ille, F. (2021). The Cellular Composition of Bovine Coccygeal Intervertebral Discs: A Comprehensive Single-Cell RNAseq Analysis. *International Journal of Molecular Sciences*, *22*(9), 4917. https://doi.org/10.3390/ijms22094917

Chang, J. T., Esumi, N., Moore, K., Li, Y., Zhang, S., Chew, C., Goodman, B., Rattner, A., Moody, S., Stetten, G., Campochiaro, P. A., & Zack, D. J. (1999). Cloning and Characterization of a Secreted Frizzled-Related Protein that is Expressed by the

Retinal Pigment Epithelium. *Human Molecular Genetics*, *8*(4), 575–583. https://doi.org/10.1093/hmg/8.4.575

Forcella, M., Lau, P., Oldani, M., Melchioretto, P., Bogni, A., Gribaldo, L., Fusi, P., & Urani, C. (2020). Neuronal specific and non-specific responses to cadmium possibly involved in neurodegeneration: A toxicogenomics study in a human neuronal cell model. *NeuroToxicology*, *76*, 162–173. https://doi.org/10.1016/j.neuro.2019.11.002

Gao, B., Li, S., Tan, Z., Ma, L., & Liu, J. (2019). ACTG1 and TLR3 are biomarkers for alcohol-associated hepatocellular carcinoma. *Oncology Letters*, *17*(2), 1714–1722. https://doi.org/10.3892/ol.2018.9757

Hamdi, M., Sánchez-Calabuig, M. J., Rodríguez-Alonso, B., Arnal, S. B., Roussi, K., Sturmey, R., Gutiérrez-Adán, A., Lonergan, P., & Rizos, D. (2019). Gene expression and metabolic response of bovine oviduct epithelial cells to the early embryo. *Reproduction*, *158*(1), 85–94. https://doi.org/10.1530/REP-18-0561

Illien-juenger, S. (2010). *Use of Intervertebral Disc Whole Organ Culture for Degenerative and Regenerative Studies.* University of Bern , Graduate School for Cellular and Biomedical Sciences . Bern: University Library of Bern.

Mahmoud, D. E., Kaabachi, W., Sassi, N., Mokhtar, A., Moalla, M., Ammar, L. B., Jemmali, S., Rekik, S., Tarhouni, L., Kallel-Sellami, M., Cheour, E., & Laadhar, L. (2021). SFRP5 Enhances Wnt5a Induced-Inflammation in Rheumatoid Arthritis Fibroblast-Like Synoviocytes. *Frontiers in Immunology*, *12*, 2356. https://doi.org/10.3389/fimmu.2021.663683

Makovets, S. (Ed.). (2013). *DNA Electrophoresis: Methods and Protocols* (Vol. 1054). Humana Press. https://doi.org/10.1007/978-1-62703-565-1

Newton, C. R., & Graham, A. (1994). *PCR.*

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., & Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification*, *1*(1), 23–33. https://doi.org/10.1016/j.bdq.2014.08.002

Penchev, R., Scheuring, R. A., Soto, A. T., Miletich, D. M., Kerstman, E., & Cohen, S. P. (2021). Back Pain in Outer Space. *Anesthesiology*, *135*(3), 384–395. https://doi.org/10.1097/ALN.0000000000003812

Rocha, A. J., Miranda, R. de S., Sousa, A. J. S., & Silva, A. L. C. da. (2016). Guidelines for Successful Quantitative Gene Expression in Real- Time qPCR Assays. In *Polymerase Chain Reaction for Biomedical Applications*. IntechOpen. https://doi.org/10.5772/65850

Schmittgen, T., & Livak, K. (2008). Schmittgen TD, Livak KJAnalyzing real-time PCR data by the comparative C(T) method. Nat Protocols 3(6): 1101-1108. *Nature Protocols*, *3*, 1101–1108. https://doi.org/10.1038/nprot.2008.73

Shi, Z., Xu, M., Chen, X., Wang, J., Zhao, T., & Zha, D. (2021). The regulatory role of SFRP5/WNT5A axis in allergic rhinitis through inhibiting JNK pathway activation and lowering mucin generation in human nasal epithelial cells. *Experimental and Molecular Pathology*, *118*, 104591. https://doi.org/10.1016/j.yexmp.2020.104591

Somers, J., Smith, C., Donnison, M., Wells, D. N., Henderson, H., McLeay, L., & Pfeffer, P. L. (2006). Gene expression profiling of individual bovine nuclear transfer blastocysts. *Reproduction*, *131*(6), 1073–1084. https://doi.org/10.1530/rep.1.00967

Tan, S. H., Reverter, A., Wang, Y., Byrne, K. A., McWilliam, S. M., & Lehnert, S. A. (2006). Gene expression profiling of bovine in vitro adipogenesis using a cDNA microarray. *Functional & Integrative Genomics*, *6*(3), 235–249. https://doi.org/10.1007/s10142-005-0016-x

Wu, X. -q., Wang, Y. -q., Xu, S. -m., Liu, J. -f., Bi, X. -y., Wang, Z. -q., & Zhang, J. -p. (2017). The WNT/β-catenin signaling pathway may be involved in granulosa cell apoptosis from patients with PCOS in North China. *Journal of Gynecology Obstetrics and Human Reproduction*, *46*(1), 93–99. https://doi.org/10.1016/j.jgyn.2015.08.013

Ying, J., Wang, J., Ji, H., Lin, C., Pan, R., Zhou, L., Song, Y., Zhang, E., Ren, P., Chen, J., Liu, Q., Xu, T., Yi, H., Li, J., Bao, Q., Hu, Y., & Li, P. (2016). Transcriptome analysis of phycocyanin inhibitory effects on SKOV-3 cell proliferation. *Gene*, *585*(1), 58–64. https://doi.org/10.1016/j.gene.2016.03.023

Zhao, C., Bu, X., Zhang, N., & Wang, W. (2009). Downregulation of SFRP5 expression and its inverse correlation with those of MMP-7 and MT1-MMP in gastric cancer. *BMC Cancer*, *9*, 224. https://doi.org/10.1186/1471-2407-9-224

Appendices

Appendix A: Reverse Transcription Protocol

The following reaction yields 20 µl volume of cDNA

1. mRNA must be diluted to attain a concentration of either 250 ng/µl or 500 ng/ µl with a volume of 50 µl or 100 µl. To calculate the needed amounts of mRNA (250 ng or 500 ng) and water the following equation can be used:

$$c_1V_1 = c_2V_2$$

2. Dilute mRNA with water. For primer testing 1300 ng/ µl was diluted with 9.4 µl of water to attain 250 ng/ µl mRNA.
3. Prepare mix (13 µl total volume), work on ice:
    - 2 µl RNA (250 ng/ul)
    - 9 µl ddH2O
    - 1 µl random primers (100 ng/µl or 10 mM (TBC), Hexanucleotide Primers, Sigma)
    - 1 µl dNTP (10 mM, dNTP Mix, Thermo Scientific)
4. Incubate for 5 min at 65°C (Gene Amp PCR System 2400, user "tw", program "reverseT_start")
5. Incubate on ice for at least 1 min
6. Centrifuge briefly
7. Add to RNA solution (final volume 20 µl) and mix gently with pipet:
    - 4 µl 5X First Strand Buffer (Invitrogen)
    - 1 µl 0.1 M DTT (Invitrogen)
    - 1 µl RiboLock RNase Inhibitor (40 U/µl, 2500 U, Thermo Scientific)
    - 1 µl SuperScript III Reverse Transcriptase (200 units/µl, 10'000 U, Invitrogen)
8. Incubate for 5 min at 25°C (Gene Amp PCR System 2400, user "tw", program "reverseT_cDNA")
9. Incubate for 60 min at 50°C
10. Inactivate reaction by heating for 15 min at 70°C

For qPCR

1. dilute cDNA 1:20 by adding 380 µl of ddH2O
2. Keep on ice or store at -20°C.
3. Use 5 ng (i.e., 4 µl of dilution) for qPCR.

Appendix B: qPCR Primer Testing Protocol

Essential tips and steps prior to starting:

- All necessary files like pipetting scheme with calculations, iQ5 plate, CSV-labelling file etc. can be prepared beforehand.
- Three replicates were used for the samples. A 10% pipetting loss were considered when calculating volumes.
- All reagents must be defrosted on ice and SYBR® Green mix must be kept away from light.
- Filtered tips should be used to avoid contamination.
- Inaccurate pipetting and bad mixing of fluids can weaken the data considerably. Therefore:
    - Before and after every pipetting step, the volume in the tip is to be checked visually.
    - all fluids (primers, mastermix etc.) must be mixed well, but only prior to use.
- When sealing the plate, fingerprints, and marks on the top of the film must be avoided.
- The plate can be spun down before placing it into the machine to ensure that the reagents are properly mixed and collected at the bottom of the reaction tube.

Solubilise Primer Pellet from Microsynth:

1. Thaw DNase free water
2. Label primer tubes (stock solutions) on the lid with the primer number according to the Primer Inventory list.
3. Do a short spin at max. speed in a centrifuge so that the pellets will accumulate at the bottom of the tube
4. Add required amount of water (DNase- and DEPC-free) for a 100 µM solution according Microsynth delivery sheet

   The primer pairs were diluted with the following quantity of water:

Table 9: Primer Dilution - ddH2O Quantities

Gene	primer	Water content (µl)
SFRP5	Forward primer	265.8

	Reverse primer	332.9
ADIRF	Forward primer	427.2
	Reverse primer	506.1
S100A2	Forward primer	455.9
	Reverse primer	405.1

5. Heat the tube for 5 minutes at 65 °C
6. Vortex briefly
7. Consider preparing aliquots
8. Store primers at -20°C

Aliquot and Dilute Primers to 10 µM:

1. Thaw DNase free water
2. Label tubes for aliquot with primer name, concentration, and date: e.g., bov NOTCH-1 F2, 10 µM, 26.02.2019
3. Dilute primers 1:10. Put 90 µl water to the tubes and add 10 µl of stock primer suspension, mix well by pipetting
4. Vortex briefly and spin down
5. Store at -20°C

Preparation of 96-well plate and start of iCycler:

Reagents consisted of: SYBR® Green, ddH_2O, forward and reverse primers, and cDNA.

Table 10: Master Mix for Primer Testing

Mastermix			**1x**	**5x**
SYBRmix	µl		10.0	50.0
Primer F	µl		0.5	2.5
Primer R	µl		0.5	2.5
ddH2O	µl		5.0	25.0
Total MM	**µl**		**16.0**	**80.0**
cDNA	µl		4.0	36.0
Total Volumen	µl		20.0	116.0

1. Turn on ice machine
2. Check, that all ingredients are present (pipetting protocol)

3. Start qPCR machine and Computer (switch on qPCR before computer, qPCR lid needs 10min warm-up time)
4. Check, that on the qPCR screen the user is: HOST_USER, otherwise the run cannot be started!
5. Make/Copy the plate setup (on iQ5), check that the reaction volume is “20 µl” and the seal type is “Film”. Edit/select the (cycling) protocol (on iQ5) if required.
6. Thaw and vortex components of master mix (use DNase- and DEPC-free water)

    - work on ice, keep SYBR® Green away from light

7. Prepare master mix, start with the largest volume
8. Vortex and spin down master mix
9. Mix cDNA by tilting the tube a few times (do NOT vortex, it shreds DNA)
10. Pipette first master mix and then cDNA into reaction wells (work on cooling element). Cover plate with film and seal it firmly using the scraper. Do not touch the foil without gloves (measurement through foil).
11. Check that protocol and layout are correct (on iQ5)
12. Place plate in qPCR cycler.
13. Click on “Run”
14. Click on “Begin run” (ignore calibration messages and click on “yes”)
15. Browse to the desktop and save the data file there (in case of connection loss to server)
16. After the run, check that all volumes in the single wells are equal, note down deviations (e.g., evaporation)
17. Transfer the data file to the desired location on the server.

Appendix C: Gel Electrophoresis Protocol

For this process 3% agarose electrophoresis precast gel (8-well) is used

1. Place gel in the electrophoresis rig with the wells closest to the negative (black) end and cover with 1x TAE buffer.

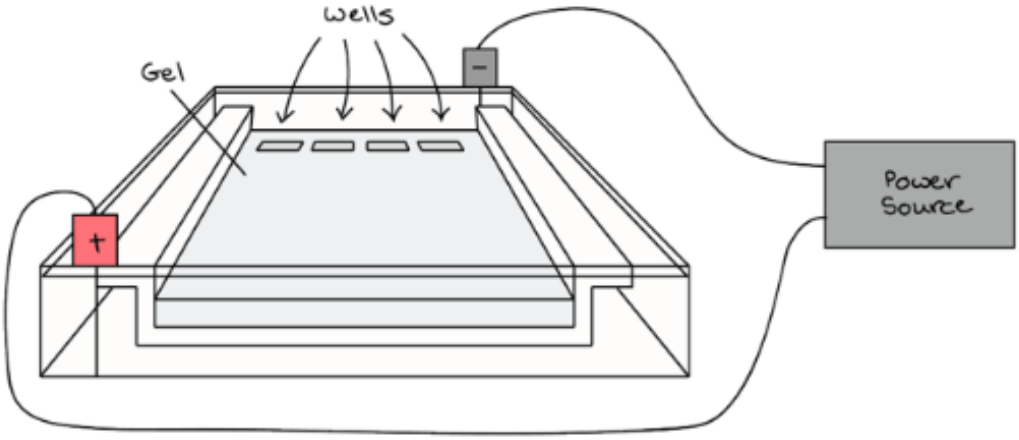


Figure 22 : Gel Electrophoresis Rig

2. Add the appropriate amount of 6x loading dye to each of the DNA samples.
   Here, triplets of the DNA were taken, hence, 5 µl of each sample (total 15 µl) and plus 3 µl of dye.
3. Load 6 µl of the GeneRuler 100bp DNA Ladder (ready-to use, Thermo scientific) in the first lane of the gel
4. Load 15 µl of the DNA samples respectively into the wells.
5. Check that the wells of the agarose gel is near the black (-) electrode and the base of the gel
6. Run the gel at 65 mA and ≅ 100 volts until the samples have run 3/4th the way down the gel (approx. 50 min)
7. Analyse the gel and take a clear picture.

Staining and washing of the Agarose gel 100x Fast Blast DNA Stain

1. Prepare 100x Fast Blast DNA stain by diluting 100 ml of 500x Fast Blast stain with 400 ml of deionized water in appropriately sized flask and mix.
2. Cover the flask and store at room temperature until ready to use.
3. Pour 100x stain into a gel staining tray.
4. Remove the gel from the gel tray and carefully slide it into the staining tray containing the stain.

5. Stain the gel for 2 – 3 min
6. Using a funnel, pour the 100x stain into a storage bottle and save it for future use (the stain can be reused at least 7 times)
7. Transfer the gel into a container with 500 – 700 ml warm, 40 – 55 °C, tap water.
8. Gently shake the gel in the water for approx. 10 sec to rinse.
9. Transfer the gel into a large container with 500 – 700 ml of clean, warm tap water.
10. Gently rock or shake the gel on a rocking platform at speed 17 for 5 min
11. Perform a second wash as in steps 7 – 10
12. Examine the stained gel for expected DNA bands on an illumination device.

Appendix D: Dilutions Series Protocol

Essential tips and steps prior to starting were adhered to as in Appendix B: qPCR Primer Testing Protocol

Preparation of 96-well plate and start of iCycler:

Reagents consisted of: SYBR® Green, ddH_2O, forward and reverse primers, and cDNA.

Table 11: Master Mix for Dilution Series

Mastermix		**1x**	**25x**
SYBRmix	µl	10.0	250.0
Primer F	µl	0.5	12.5
Primer R	µl	0.5	12.5
ddH2O	µl	5.0	125.0
Total MM	**µl**	**16.0**	**400.0**
cDNA	µl	4.0	36.0
Total Volumen	µl	20.0	436.0

Table 12: Points of Dilution with cDNA

Dilution	**cDNA(µl)**	**ddH2O (µl)**	**[cDNA] (%)**
1	64 (from 5% stock)	0	5
2	32 (from dilution 1)	32	2.5
3	32 (from dilution 2)	32	1.25
4	32 (from dilution 3)	32	0.625
5	32 (from dilution 4)	32	0.313
6	32 (from dilution 5)	32	0.156

1. Turn on ice machine
2. Check, that all ingredients are present (pipetting protocol)
3. Start qPCR machine and Computer (switch on qPCR before computer, qPCR lid needs 10min warm-up time)
4. Check, that on the qPCR screen the user is: HOST_USER, otherwise the run cannot be started!
5. Make/Copy the plate setup (on iQ5), check that the reaction volume is “20 µl” and the seal type is “Film”. Edit/select the (cycling) protocol (on iQ5) if required.
6. Thaw and vortex components of master mix (use DNase- and DEPC-free water)

    - work on ice, keep SYBR Green away from light

7. Prepare master mix, start with the largest volume

8. Vortex and spin down master mix
9. Mix cDNA by tilting the tube a few times (do NOT vortex, it shreds DNA)
10. Label 7 Eppendorf tubes corresponding to the points of dilution. The last tube it to remove the excess volume.
11. Pipette calculated amount of 5% cDNA into 0 µl of ddH20. Mix well better continuing. Avoid generating bubbles.
12. Pipette ½ of dilution 1 (from previous step) into equal amount of ddH20 to create dilution 2. Mix well before continuing. Avoid generating bubbles.
13. Pipette ½ of dilution 2 (from previous step) into equal amount of ddH20 to create dilution 3. Mix well before continuing. Avoid generating bubbles.
14. Pipette ½ of dilution 3 (from previous step) into equal amount of ddH20 to create dilution 4. Mix well before continuing. Avoid generating bubbles.
15. Pipette ½ of dilution 4 (from previous step) into equal amount of ddH20 to create dilution 5. Mix well before continuing. Avoid generating bubbles.
16. Pipette ½ of dilution 5 (from previous step) into equal amount of ddH20 to create dilution 6. Mix well before continuing. Avoid generating bubbles.
17. Discard ½ from dilution 6 to obtain the correct volume for the final dilution.
18. Pipette first master mix and then cDNA into reaction wells (work on cooling element). Cover plate with film and seal it firmly using the scraper. Do not touch the foil without gloves (measurement through foil).
19. Check that protocol and layout are correct (on iQ5)
20. Place plate in qPCR cycler.
21. Click on “Run”
22. Click on “Begin run” (ignore calibration messages and click on “yes”)
23. Browse to the desktop and save the data file there (in case of connection loss to server)
24. After the run, check that all volumes in the single wells are equal, note down deviations (e.g., evaporation)
25. Transfer the data file to the desired location on the server.

Appendix E: Curves from Dilution Series

I. ADIRF

a. ADIRF 1, Ct Mean = 31.97

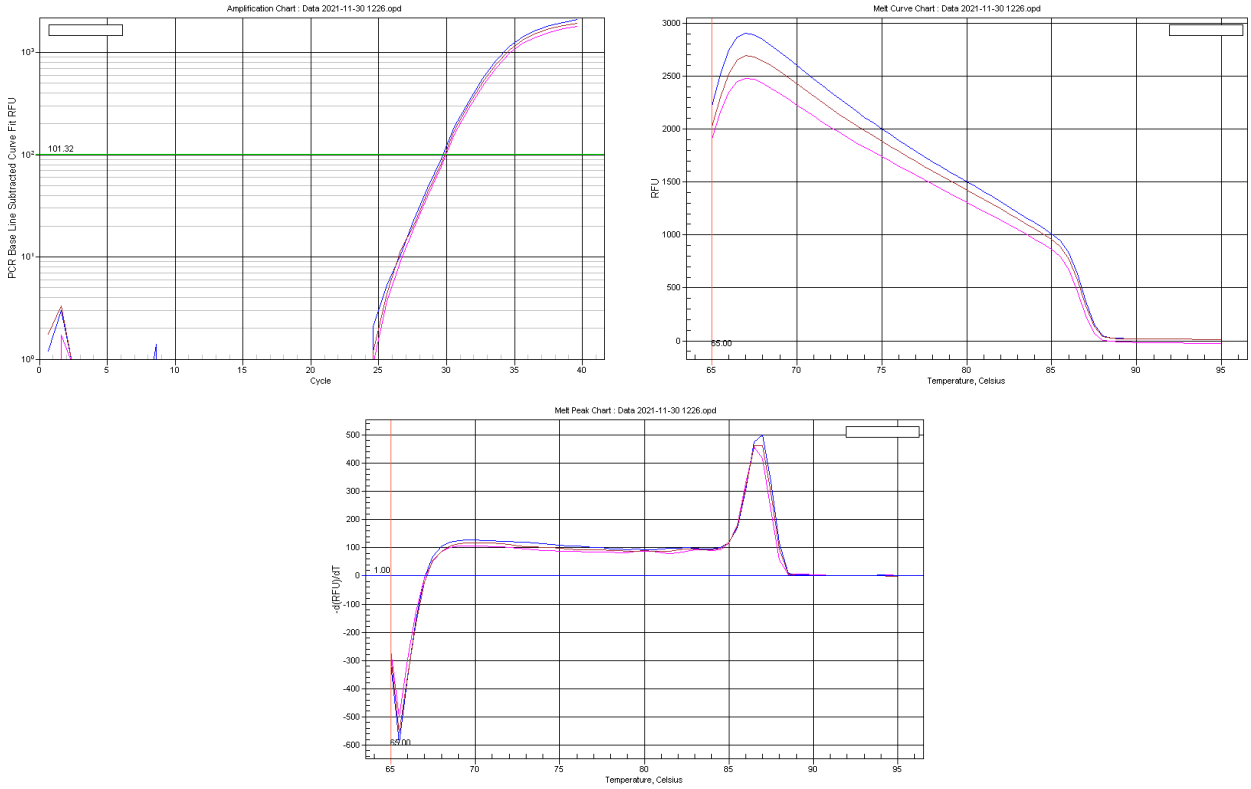
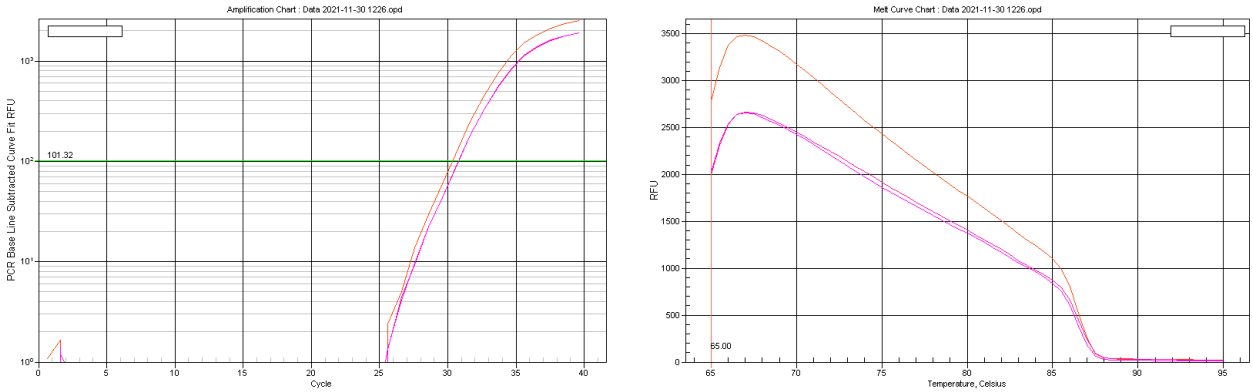


Figure 23 a: ADIRF 1st Point of Dilution

b. ADIRF 2, Ct Mean = 32.67



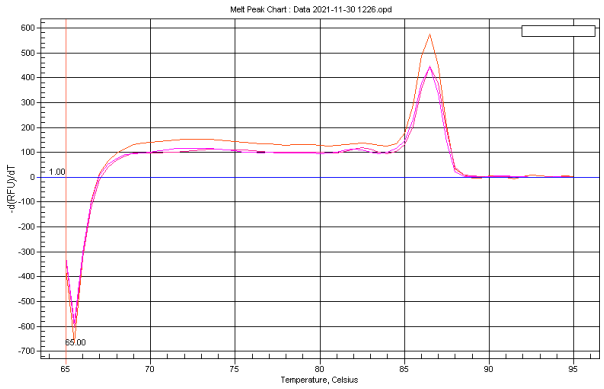


Figure 23 b: ADIRF 2nd Point of Dilution

c. ADIRF 3, Ct Mean = 33.91

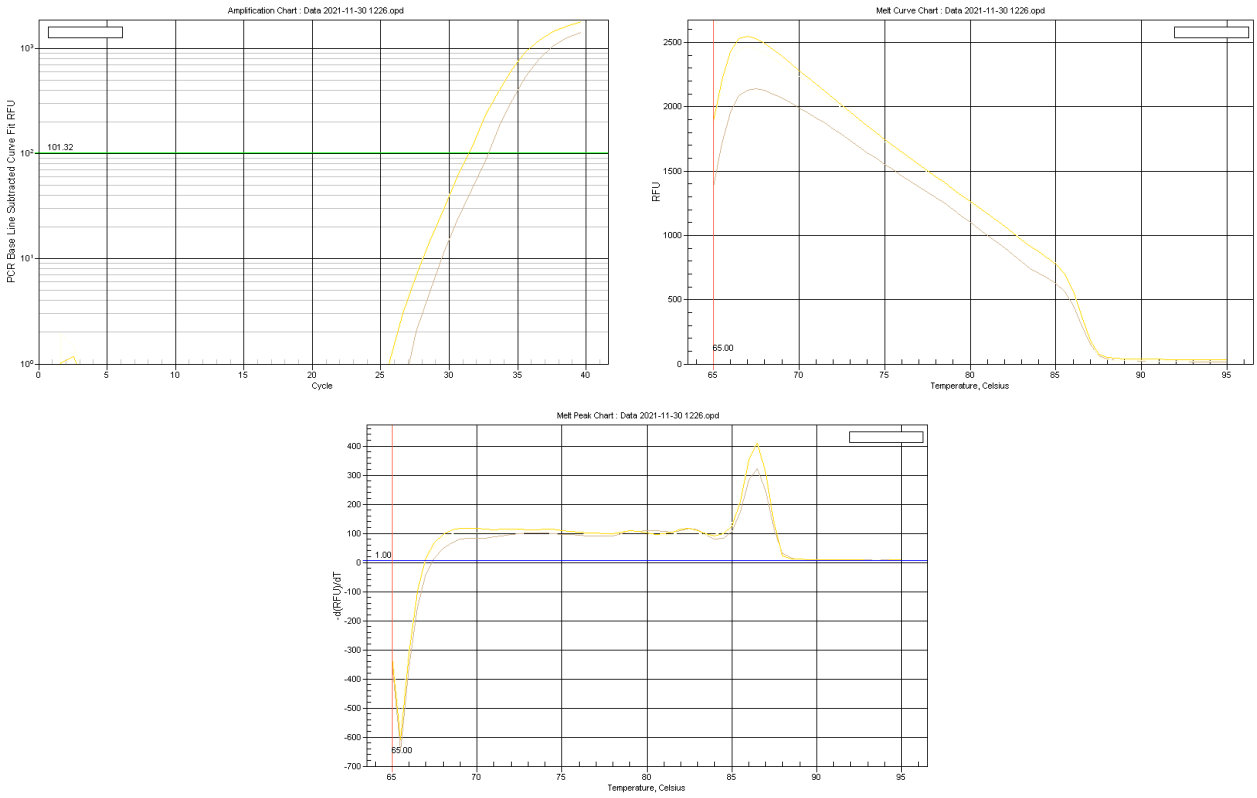


Figure 23 c: ADIRF 3rd Point of DIlution

d. ADIRF 4, Ct Mean = 35.12

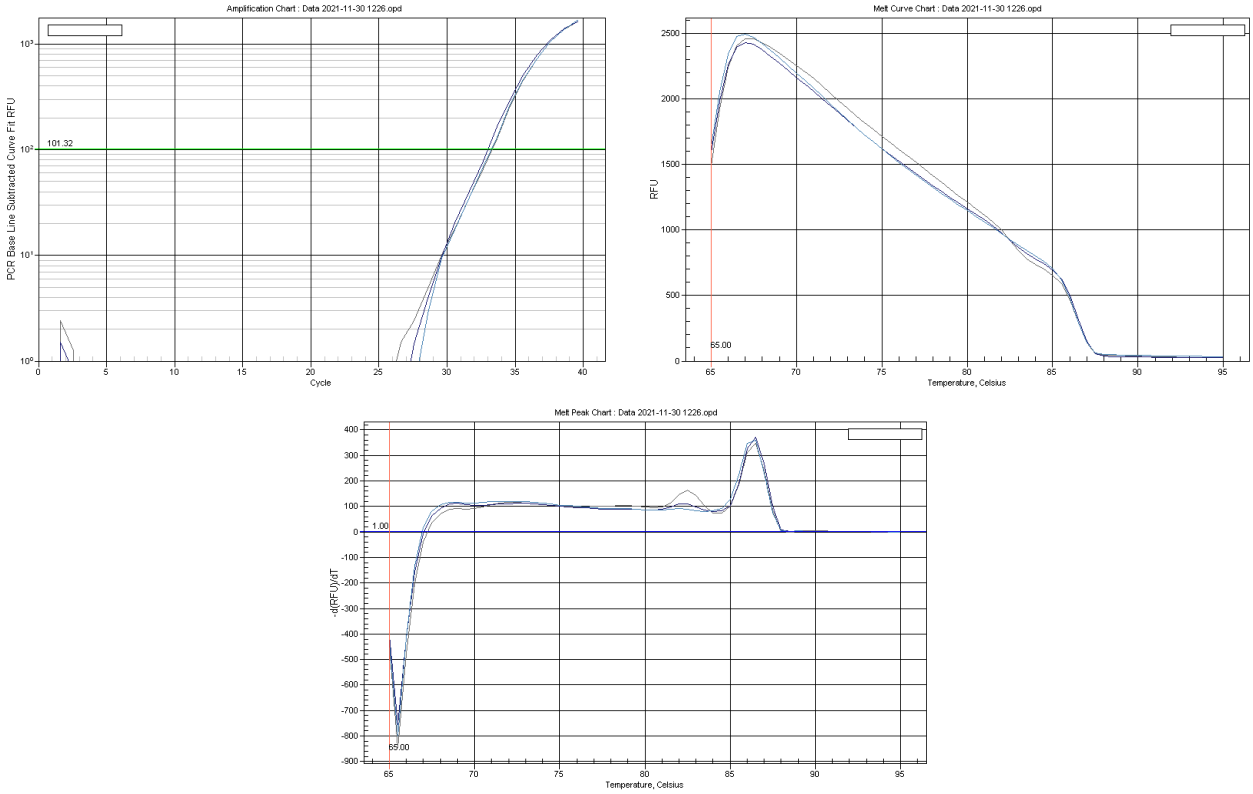
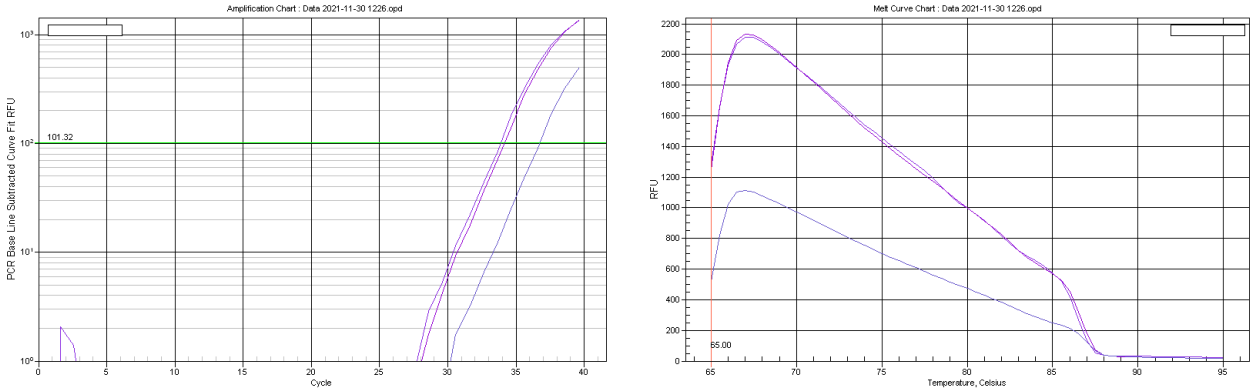


Figure 23 d: ADIRF 4th Point of Dilution

e. ADIRF 5, Ct Mean = 35.94



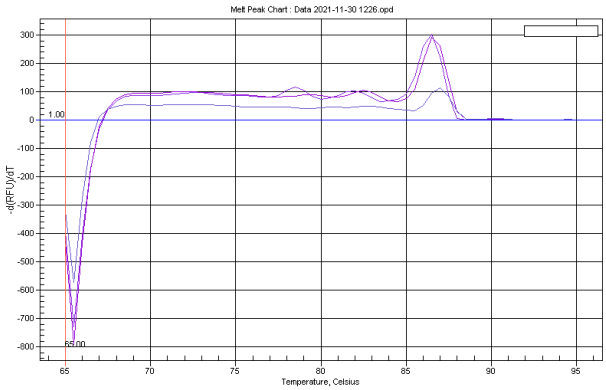


Figure 23 e: ADIRF 5th Point of Dilution

f. ADIRF 6, Ct Mean = 36.60

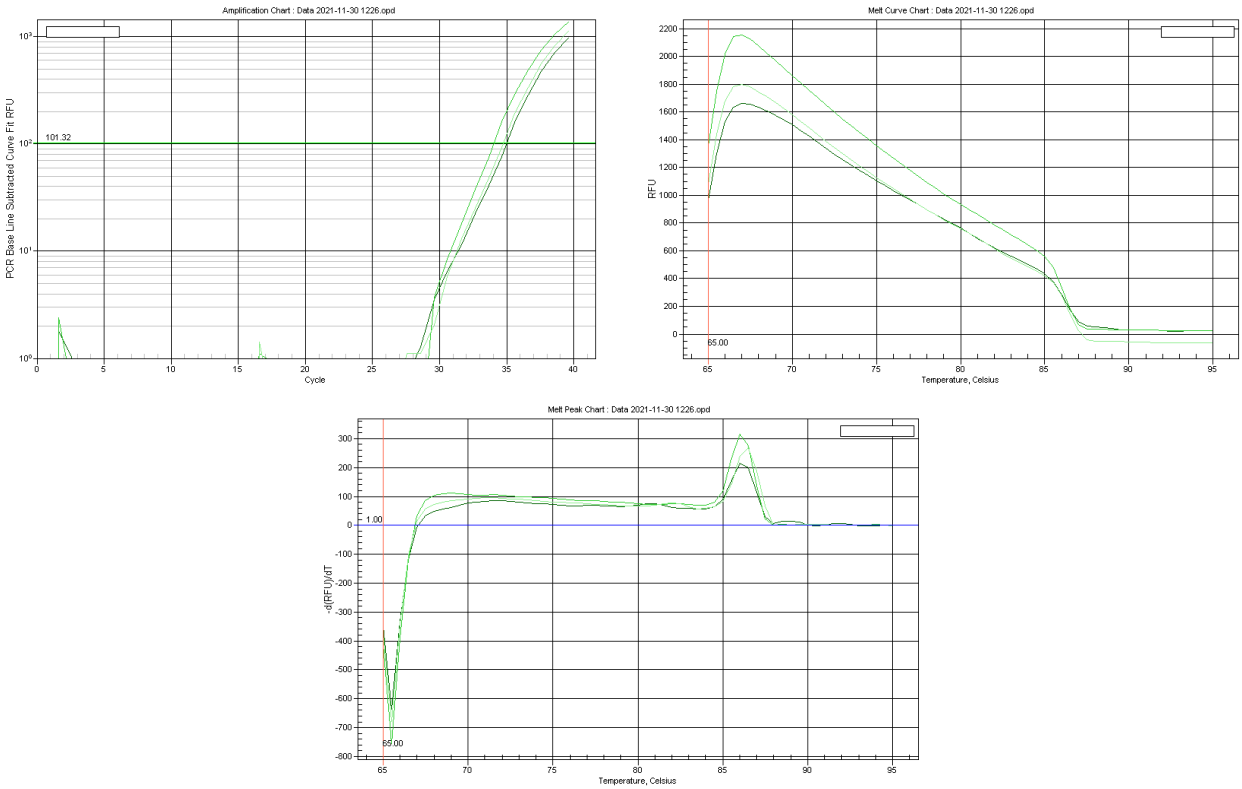


Figure 23 f: ADIRF 6th Point of Dilution

II. S100A2

a. S100A2 1, Ct Mean = 29.63

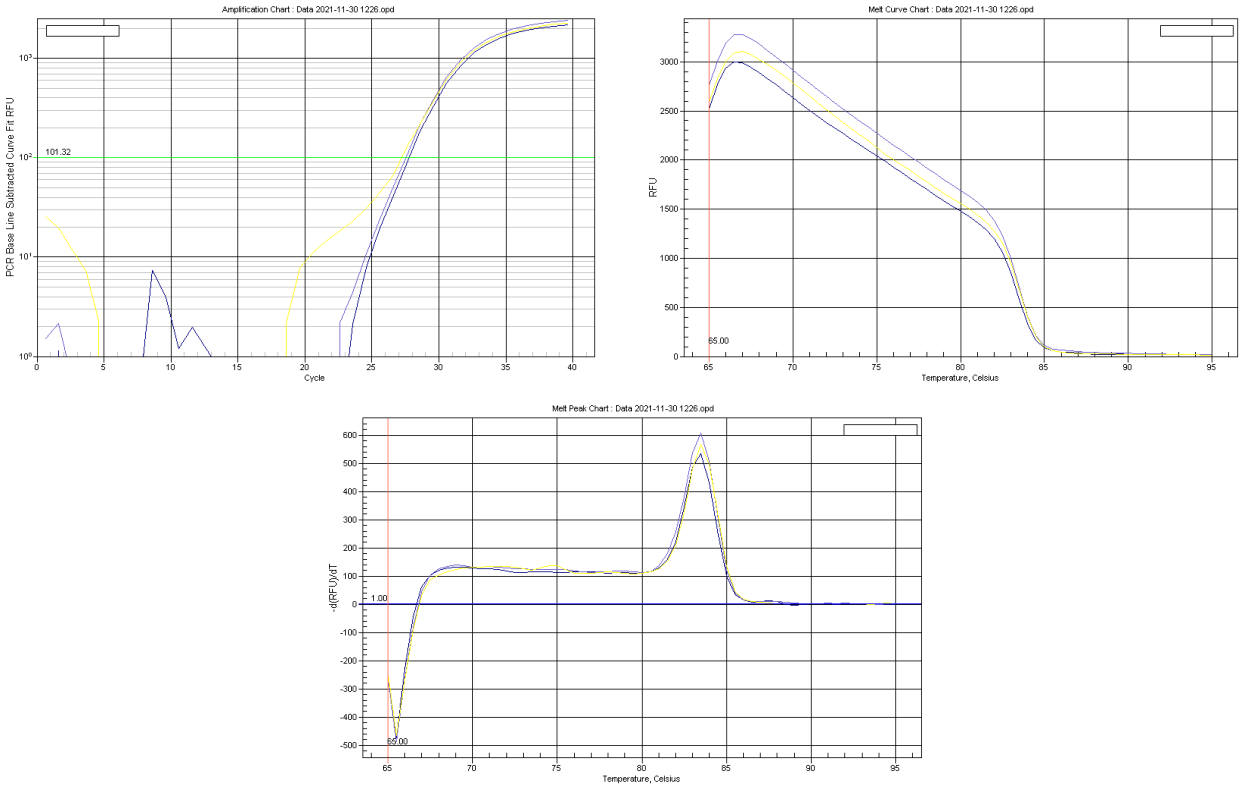
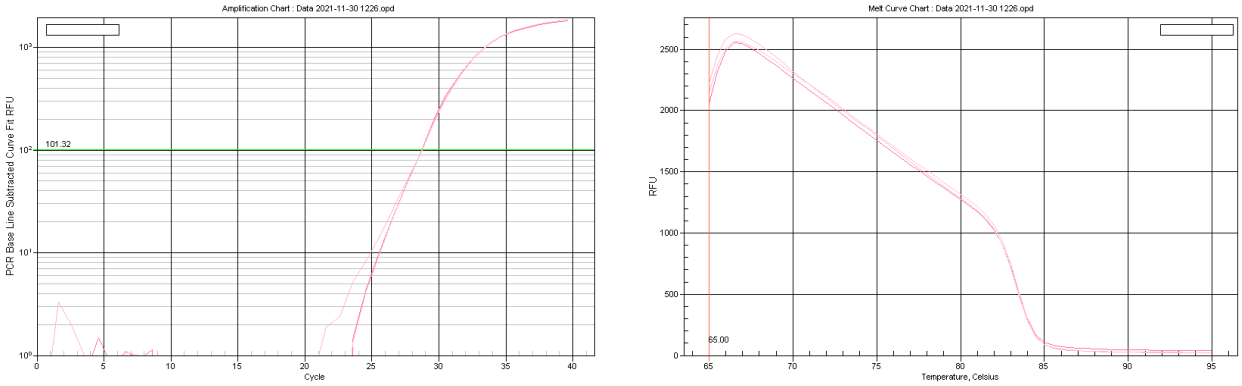


Figure 24 a: S100A2 1st Point of Dilution

b. S100A2 2, Ct Mean = 30.76



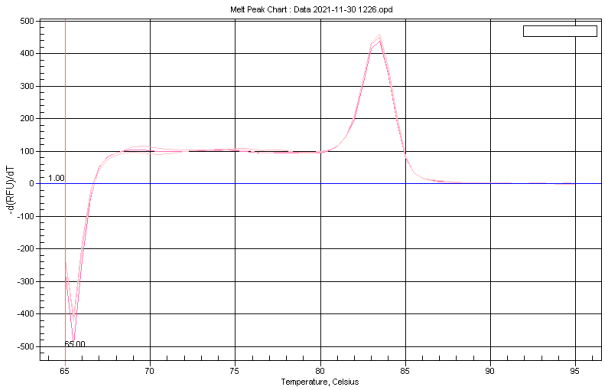


Figure 24 b: S100A2 2nd Point of Dilution

c. S100A2 3, Ct Mean = 31.70

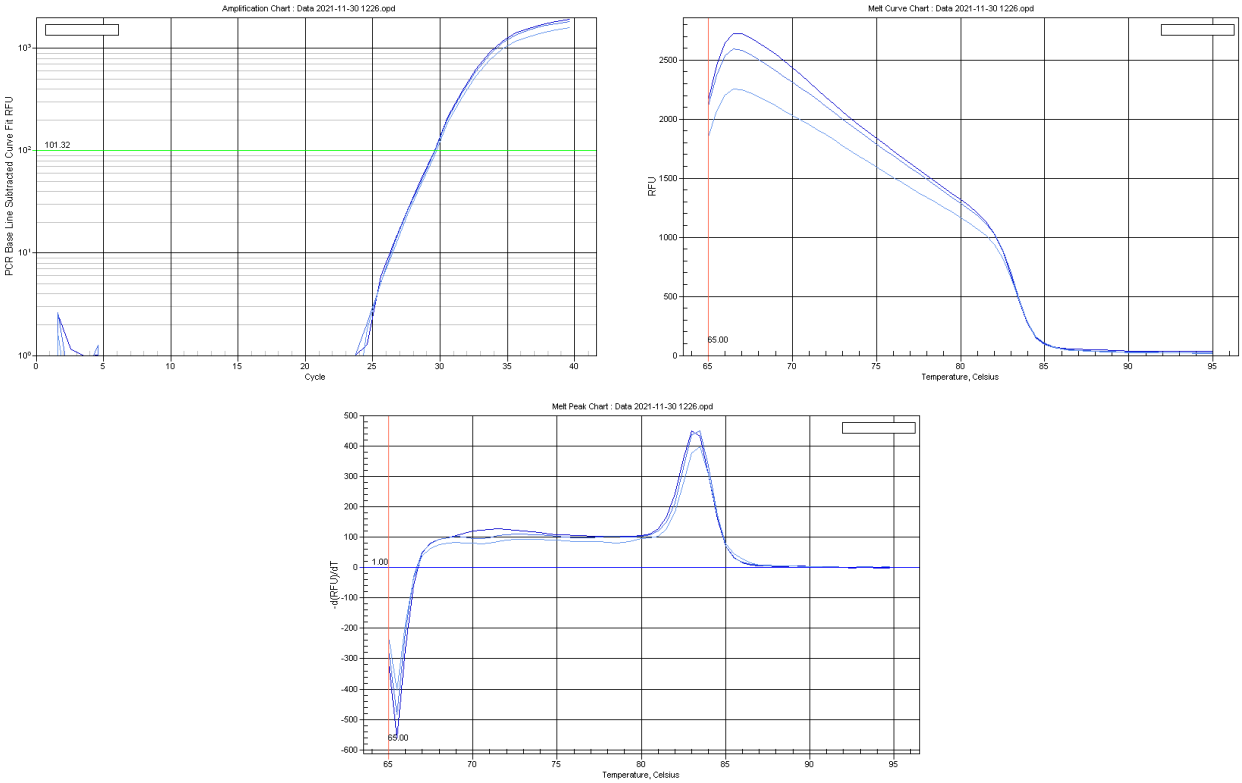


Figure 24 c: S100A2 3rd Point of Dilution

d. S100A2 4, Ct Mean = 32.90

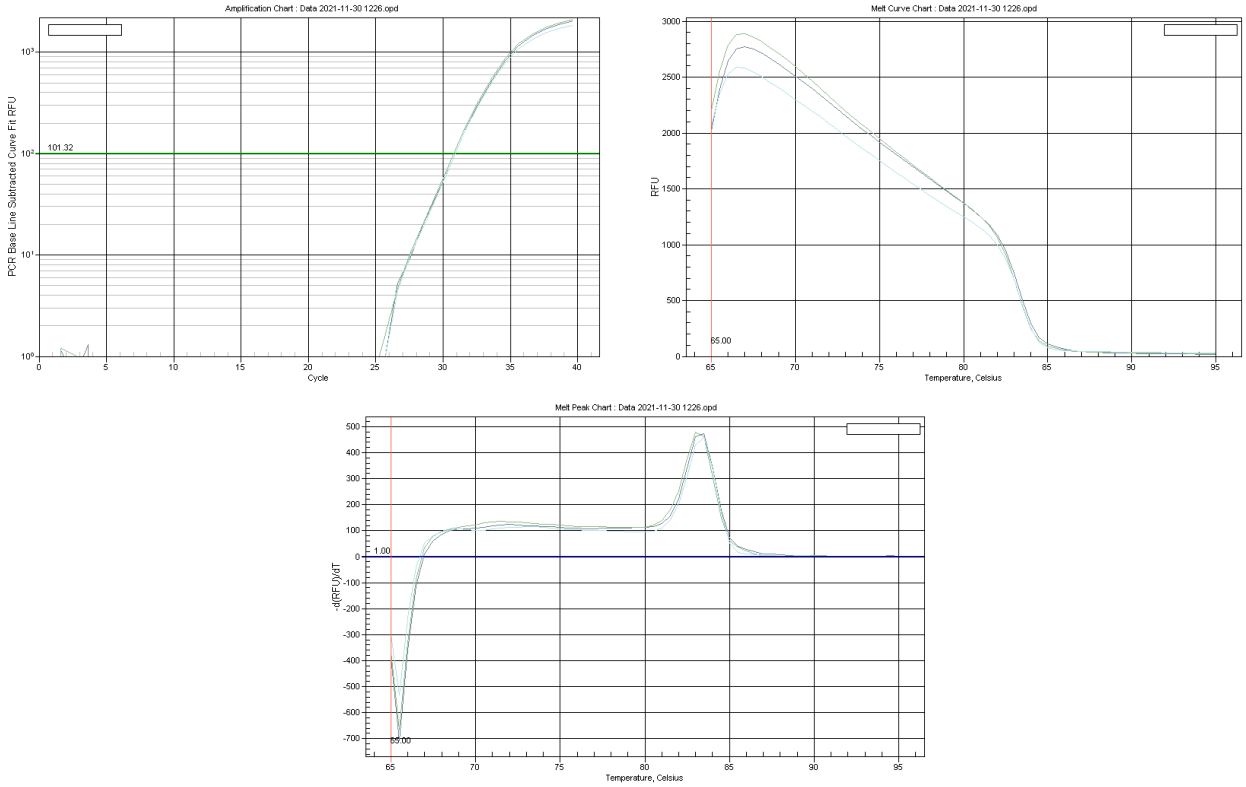
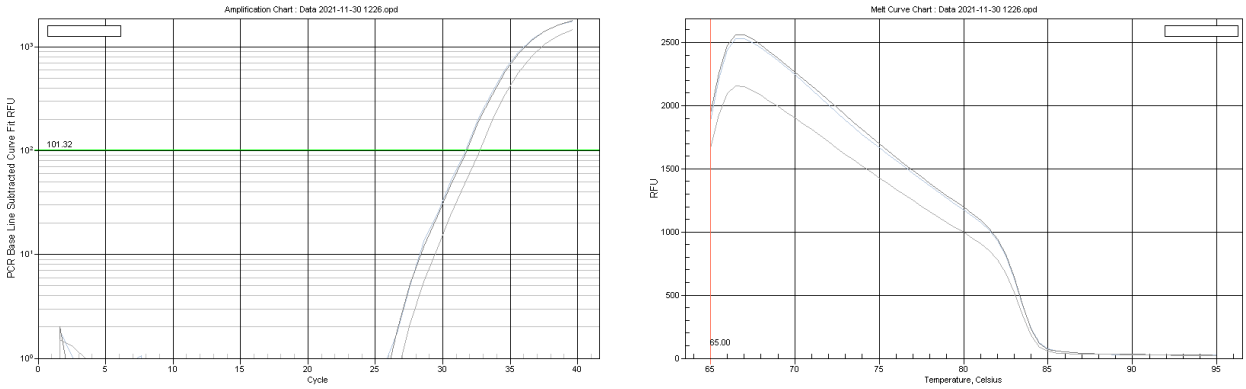


Figure 24 d: S100A2 4th Point of Dilution

e. S100A2 5, Ct Mean = 34.06



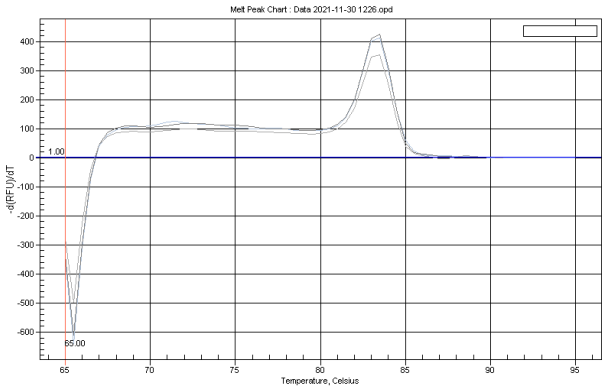


Figure 24 e: S100A2 5th Point of Dilution

f. S100A2 6, Ct Mean = 35.13

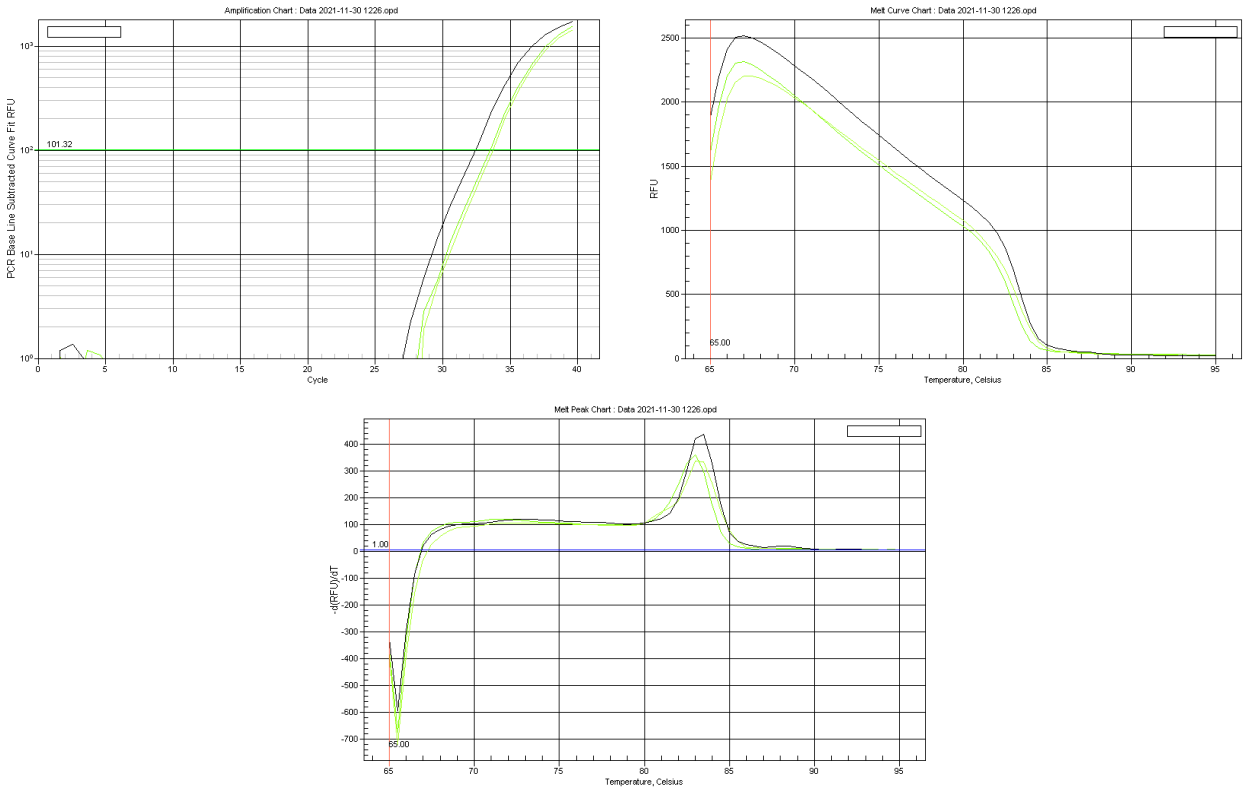


Figure 24 f: S100A2 6th Point of Dilution

Appendix F: qPCR with 6 pools of cDNA

Reverse Transcription

RNA extracted out of the NP from 3 different bovine tails were segmented into proximal and distal. IVD's 1 to 4 was grouped together representing proximal and IVD's 5 to 8 representing distal. However, with some consideration like extracting a certain amount from each RNA to attain a 1.25 ng/µl concentration of cDNA. Table 13 presents the amount of RNA taken from each IVD.

Thereafter, the RNA was reverse transcribed into 6 pools of cDNA according to the protocol in Appendix A: Reverse Transcription Protocol, here 500 ng RNA was reserve transcribed instead of 250 ng.

Table 13: Segmentation of RNA from IVD's

IVD		Tail 1			Tail 2			Tail 3		
		Nucleic acid conc. (ng/µl)	Pools	D.	Nucleic acid conc. (ng/µl)	Pools	D.	Nucleic acid conc. (ng/µl)	Pools	D.
Proximal	1	12.4	25.4µl/IVD + 0.8µl ddH2O → 7 tubes	+580µl ddH2O	67.8	4.3µl/IVD + 4.8µl ddH2O	+680µl ddH2O	13.6	11.2µl/IVD + 10.2µl ddH2O	+620µl ddH2O
	2	16.0			52.2			12.7		
	3	n/a			48.8			14.5		
	4	7.1			38.3			39.6		
Distal	5	85.9	9.8µl/IVD + 2.4 µl ddH2O →2 tubes	+680µl ddH2O	n/a	4.2µl/IVD + 13.6µl ddH2O	+680µl ddH2O	n/a	24.0µl/IVD + 9µl ddH2O →	+660µl ddH2O
	6	6.2			143.8			37.6		
	7	n/a			n/a			n/a		
	8	n/a			68.5			n/a		

qPCR Procedure

Essential tips and steps prior to starting were adhered to as in Appendix B: qPCR Primer Testing Protocol

Preparation of 96-well plate and start of iCycler:

Reagents consisted of: SYBR® Green, ddH_2O, forward and reverse primers, and 6 pools cDNA (T1 P, T1 D, T2 P, T2 D, T3 P, T3 D)

Table 14: Master mix for qPCR with 6 cDNA pools

Mastermix		**1x**	**27x**
SYBRmix	µl	10.0	270.0
Primer F	µl	0.5	13.5
Primer R	µl	0.5	13.5
ddH2O	µl	5.0	135.0
Total MM	**µl**	**16.0**	**432.0**
cDNA	µl	4.0	36.0
Total Volumen	µl	20.0	468.0

The remaining steps were adhered to as stated in Appendix B: qPCR Primer Testing Protocol