TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Physiologie

Profiling of circulating small RNAs and their potential as transcriptomic biomarkers to detect anabolic drug abuse in bovines

Melanie Spornraft

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. W. Windisch
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. M. W. Pfaffl
	2.	apl. Prof. Dr. R. Kühn

Die Dissertation wurde am 18.08.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 10.11.2015 angenommen.

Table of content

Abbreviations

A	adenine
AC	adenylate cyclase
AnaBull	name of the presented animal trial
ATP	adenosine triphosphate
bp	base pair
bta	bos taurus
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CLEN	clenbuterol-hydrochloride treated group
CON	control group
Cq	quantitative cycle
CRE	cAMP responsive element
CREBP	cAMP responsive element binding protein
d+	treatment day
DA	discriminant analysis
ddPCR	droplet digital PCR
EMBL	European Molecular Biology Laboratory
EU	European Union
exRNA	extracellular RNA
FIFA	Fédération Internationale de Football Association
FU	fluorescence unit
G	gauge
HCA	hierarchical cluster analysis
HRE	hormone responsive element
hsa	homo sapiens
HSP	heat-shock protein
KDK	Kontrollgemeinschaft deutsches Kalbfleisch e.V.
lncRNA	long non-coding RNA
miRNA, miR	microRNA
mRNA	messenger RNA
NADA	National Anti-Doping-Agency

NGS	Next-Generation Sequencing
nts	nucleotides
OPLS	orthogonal partial least squares
P+EB	progesterone (100mg) plus estradiol benzoate (10mg) treated group
PCA	principal component analysis
PCR	polymerase chain reaction
piRNA, piR	PIWI-interacting RNA
PIWI	P-element induced wimpy testes
PKC	protein kinase C
PLS	partial least-squares
POL II	RNA polymerase II
QC	quality control
RBC	red blood cell
rcf	relative centrifugal force
RGI	reference gene index
RIN	RNA Integrity Number
rpm	reads per million
rRNA	ribosomal RNA
RT	reverse transcription
RT-qPCR	reverse transcription quantitative polymerase chain reaction
SD	standard deviation
SGS	Société Générale de Surveillance
small RNA-Seq	small RNA-Sequencing
smexRNA	circulating extracellular small RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
tRNA	transfer RNA
U	uridine
UTR	untranslated region
WADA	World Anti-Doping Agency

Zusammenfassung

In der Fleischproduktion werden in vielen Ländern der Erde hormonelle und pharmazeutische Wachstums- und Leistungsförderer appliziert. In der Europäischen Union ist es verboten, anabole Substanzen in der Tiermast einzusetzen. Dadurch werden Verbraucher vor potentiellen gesundheitsschädlichen Nebenwirkungen geschützt und unabsichtliches Doping im Leistungssport über den Verzehr von kontaminiertem Fleisch verhindert. Die ökonomischen Anreize sind dennoch gegeben, weswegen die Einhaltung dieses Verbotes durch Kontrollsysteme streng überwacht werden muss. Um den illegalen Einsatz von Wachstumsförderern zuverlässig und betrugssicher nachweisen zu können, ist ein neuer und innovativer Weg zu beschreiten. Eine Alternative zu substanzspezifischen Nachweismethoden ist die Untersuchung der physiologischen Veränderungen eines Organismus nach der Gabe von anabolen Agentien auf molekularem Level. Genexpressionsveränderungen auf der Transkriptomebene können durch hochtechnologisches molekulares Profiling, dem Next-Generation Sequencing (NGS), höchst sensibel detektiert werden. Nicht-Protein-kodierende, kurze aber post-transkriptional agierende RNAs (small RNAs) ziehen seit Kurzem großes wissenschaftliches Interesse auf sich, da gezeigt wurde, dass sich diese Transkript-Moleküle ausgezeichnet als Biomarker in der klinischen Praxis eignen. Zum einen zirkulieren sie in Körperflüssigkeiten und zum anderen weisen sie eine erhöhte Stabilität auf, was Bio-Flüssigkeiten zur idealen Probe machen. Um die Gesamtheit der zirkulierenden small RNAs mittels NGS im Hochdurchsatz zu charakterisieren, wurden experimentelle Voraussetzungen geschaffen, um erstmalig diese Anwendung in der Spezies Rind durchzuführen. Um biologisch interpretierbare Information aus den extrem umfangreichen generierten NGS-Datensätzen zu ziehen, wurde eine bioinformatische Datenanalyse-Pipeline kreiert. Nachdem diese Bedingungen erfüllt waren, konnte ein spezifischer „Fingerabdruck“ der zirkulierenden small RNAs in bovinem Plasma und Vollblut gesunder Tiere präsentiert werden. Eine vergleichende Analyse wies eine leberspezifische miRNA im Plasma auf, deren verstärkte Präsenz einen Hinweis darauf gibt, dass regulatorische small RNAs von Geweben ins Blut transportiert werden und dort zirkulieren. Außerdem wurde das Potenzial des small RNA Transkriptoms als Quelle für neue Biomarker-Kandidaten validiert. Im Rahmen einer Tierstudie konnte gezeigt werden, dass die Summe der sequenzierten miRNAs die unbehandelten von den mit Anabolika behandelten Tieren diskriminiert. In Kombination mit der gewählten multivariaten Datenanalysemethode eignen sich zirkulierende miRNAs den Missbrauch von Tierarzneimitteln und verbotenen anabolen Substanzen in der Rindermast zu detektieren.

Abstract

Veterinary drug abuse is a common committed crime in food-producing animal husbandry with far-reaching impact on the animal´s health, human nutrition and thereby competitive sports. To prevent the consumer from possible health risks by consumption of contaminated meat, reliable and fraud-resistant controls need to be established. Application of anabolic drugs influences molecular regulations of the transcriptome leading to an enhancement of muscle growth. These genetic alterations on the transcriptomic level can be traced using the state-of-the-art molecular profiling technique called Next-Generation Sequencing (NGS). Recently, microRNAs (miRNA) came into sharp scientific focus, since miRNA biomarkers were successfully developed for multiple clinical applications. Liquid biopsies serve as beneficial matrix not requiring invasive sampling practices and free circulating small RNAs, including miRNAs and piRNAs, exhibit increased stability characteristics. However, blood and plasma are methodically critical sample types and exact quantification reaches limits due to restricted circulating RNA concentrations. Hence, to screen global expression of the circulating small RNA transcriptome in bovines, fundamental experimental setups need to be developed. Therefore, as pre-sequencing conditions, the isolation process and library preparation were primarily optimized for bovine applications to perform high-quality ultrahigh-throughput screening by NGS. To reveal biological interpretable information that is hidden behind raw sequencing output, a stringent bioinformatic data evaluation and analysis pipeline was developed downstream to NGS. Since NGS delivers massive data sets, multivariate projection methodologies are tailored to meet the challenge of the data explosion of '-omics' studies to uncover relevant information. On this basis, the unique footprint of the physiological miRNome and piRNome in bovine plasma and whole blood could be described for healthy animals. Comparative analysis revealed the elevated presence of a liver-specific miRNA in plasma. This indicated that several regulatory miRNAs circulate in the bloodstream, originating from tissue sources in the body and not from blood cells. Additionally, the potential of the small RNA transcriptome as innovative source of biomarker candidates was explored. Investigating the applicability for veterinary drug abuse control, circulating miRNAs showed the best potential to uncover anabolic drug misuse.

1. Introduction

1.1. The problem of substance residues in meat

1.1.1. Discrepancies between official EU legislation and unofficial suspicions

Globally seen, the application of growth and performance enhancing drugs in food producing animals aiming to increase weight gain and feed conversion rates is common practice. For example, hormonal implants for animal fattening are used in Australia, Canada, Japan, Mexico, New Zealand, South Africa, South America and the USA. Nevertheless, the EU pursues a clear strategy to prevent consumers from possible health risks caused by residue carryover [1]. The EU Council Directive 96/22/EC prohibited growth promoting β-agonists for the administration to animals whose meat or products are intended for human consumption. Furthermore, the use of hormones was restricted and substances with hormonal activity were entirely banned. Also, the import of products derived from hormone-treated cattle is legally forbidden in the EU.

It was shown repeatedly that monitoring of chemical contaminations, species fraud and product mislabeling in food is a complex task for control laboratories. Recent pan-European food safety affairs, for example the horsemeat scandal in 2013, underline the need for sophisticated and reliable analytical methods as well as sufficiently frequent routine investigations in food producing animals. To maximize economic profit, 'black sheep' in agricultural industry try to circumvent supervisory authorities. This can either be done by applying alternative compounds like synthetic designer drugs or xenobiotic substances. Targeted routine controls only screen for known compounds implying that unknown substances fall through the analytical grid. Another fraudulent method is an alternative application scenario administering drug cocktails with not-detectable low concentrated single-doses of each compound. Due to the highly appraised commercial success, the use of illicit substances in meat production can frequently be suspected [2].

1.1.2. Unintended clenbuterol doping in athletes, travelers and citizens by consumption of contaminated beef meat

The following example highlights very descriptively that illegal 'doping' practices in livestock farming and meat production affect public health as well as the world of sports. In May 2011, Mexico´s national soccer team came under suspicion initiated by five adverse analytical findings concerning the anabolic drug clenbuterol. Non-compliance with anti-doping regulations led The Fédération Internationale de Football Association (FIFA) to investigations of those reportings, precisely because Mexico was the host country for the FIFA U-17 World Cup 2011. Therefore, FIFA decided to support clarification measurements in advance and during the ongoing championships. In seven Mexican venue cities, meat samples of served food were collected at the team hotels and additionally urine of soccer players was sampled. As a result, 30% of meat samples and 52% of collected urine specimens were tested positive for clenbuterol [3]. The German soccer team remained unaffected by the doping accusations. As the German national Anti-Doping-Agency (NADA) was aware of potentially contaminated food in Mexico, they issued a warning to the German Soccer Federation (Deutscher Fußball-Bund, e.V.). Therefore, the team´s own chef imported 'safe' meat for the championships. Based on this inquiry outcome, all U-17 players were exempted from penalties per decisions of FIFA and the World Anti Doping Agency (WADA). Similar to Mexico, growth promoting practices in animal husbandry are widely spread in South America. It remains unclear, if contaminated meat also represents a risk at the Olympic Games in 2016 in Rio de Janeiro, Brazil. Dimitrij Ovtcharov experienced a similar case. The suspension of the German table tennis player was annulled as proof was presented that he ate contaminated meat at the China Open event venue. Therefore, athletes and their teams need to strictly control their food intake to prevent stepping into the 'doping trap'. The differentiation between deliberate sports doping with clenbuterol and the unintended uptake with food products will be an analytical challenge for anti-doping research. Moreover, better information should be supplied by responsible authorities for normal citizens and travelers, as unintended uptake of anabolic residues via food poses a severe health risk. Especially in China and Mexico, excessive clenbuterol use is common practice despite its official prohibition for food-producing animals. This supports a study that was conducted by Guddat and coworkers [4]. In fact, 79% of travelling volunteers showed positive clenbuterol findings after their return from China. Only 6 out of 28 persons were tested negative, amongst them one vegetarian. Additionally, no correlation was observed between clenbuterol concentrations in urine and the visited area, except for Hong Kong where rather low concentrations were found.

This section highlights that the use of β-agonists, such as clenbuterol, in both cattle fattening and unintended sports doping is closely related in other continents.

1.2. Common practice in livestock farming: anabolic drugs to enhance growth by activating molecular signaling cascades

1.2.1. The sympathomimetic drug group of β-adrenergic agonists: clenbuterol

β-adrenergic agonists like salbutamol, terbutaline or clenbuterol induce smooth-muscle relaxation, bronchodilation and cardiac contraction by binding their associated receptors ($\beta_{1/2/3}$- adrenergic receptors). These are ubiquitous in the mammalian organism and belong to the G-protein coupled receptors. The binding of the specific agonist induces the activation of a signaling cascade via adenylate cyclase (AC) that produces the second messenger molecule cyclic adenosine monophosphate (cAMP) from adenosine triphosphate (ATP). This in turn activates protein kinase C (PKC) and a subsequent protein phosphorylation cascade that induces DNA transcription mediated by cAMP responsive element binding protein (CREBP) phosphorylation and -binding to cAMP responsive element (CRE) (Figure 1, [1.]).

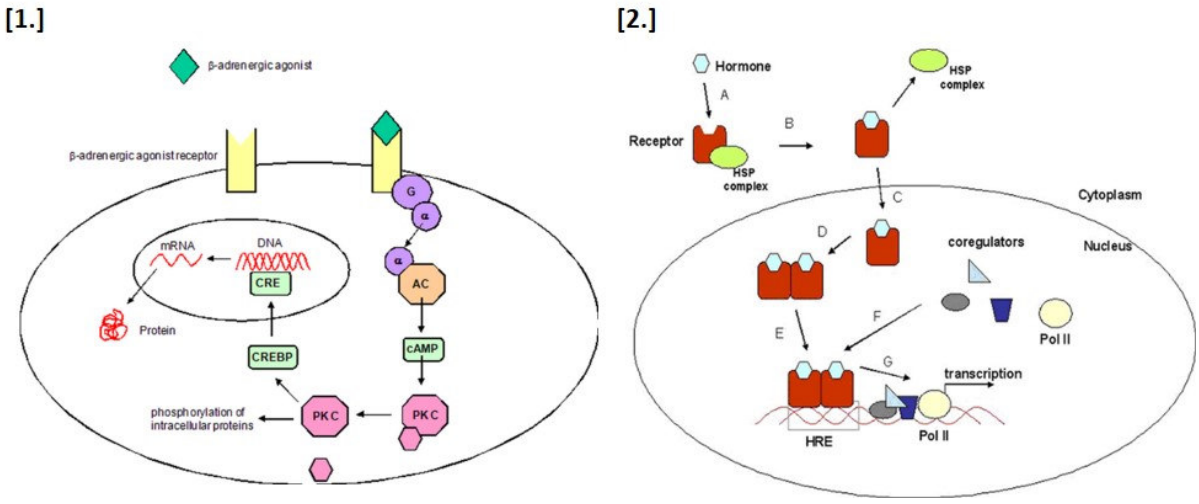


Figure 1: [1.] Schematic overview of the molecular signaling steps initiated by β-adrenergic agonists leading to DNA transcription and protein biosynthesis [5]. [2.] Summary of the physiological mechanisms induced by steroid hormones via binding to the inactivated steroid hormone receptor/HSP complex (A). After dissociation of the HSP (B), the activated hormone receptor complex diffuses to the cell nucleus (C), dimerizes (D) and binds to the HRE (E). Interacting with co-regulators and Pol II (F), a distinct gene is transcribed (G) [5].

Today, it is well known that β-adrenergic agonist administration depresses adipose tissue accretion by the combined effect of inhibiting de novo fatty acid biosynthesis and lipogenesis. In addition, postnatal muscle hypertrophy is promoted by decelerated protein turnover and an increased myofibrillar protein synthesis. β-adrenergic agonists are capable of increasing the cross-sectional area of bovine muscle fiber up to 8-40% [6]. Even though these effects are well known in animal husbandry, the exact molecular mechanisms are still not fully understood [6].

Clenbuterol is the most prominent example coming from the group of β-adrenergic agonists in the field of animal fattening, as it gained a lot of negative attention in the 1990s. Clenbuterol is not approved for the use in meat production, but the illegal feeding to cattle already led to severe food poisonings in humans. Adverse side effects could be anxiety, hypertension, increased heart rate and essential tremors, both in animals and consumers [7].

The medical application of β-adrenergic agonists is restricted in human sports by the WADA due to claimed anabolic effects and also, for the same reasons, in cattle fattening in the EU.

1.2.2. The substance group of anabolic steroid hormones: estradiol and progesterone

Steroid hormones are endogenously synthesized from cholesterol primarily in the gonads, adrenal cortex and during pregnancy in the placenta. Besides glucocorticoids and mineralocorticoids, the group of steroid hormones includes the reproductive hormone groups androgens, gestagens and estrogens. Estrogens are the main drivers for the development of the female reproductive organs, particularly the ovary, the mammary gland, the functions of the uterus and the behavioral changes during estrus. Progesterone is mainly synthesized in the corpus luteum in the ovaries, but also in the adrenal cortex and placenta. It converts the endometrium to its secretory stage to prepare the uterus for implantation of the embryo and maintains pregnancy.

Steroid hormones bind their respective receptors in the cytoplasm or in the nucleus [8,9]. Steroid receptors are stabilized by heat-shock proteins (HSP), if no specific ligand is present. After dissociation of the HSPs, the activated hormone-receptor-complex is translocated in the nucleus. There, the hormone-receptor-dimer binds to a specific DNA sequence motif in the promotor region of the steroid hormone regulated genes, called hormone responsive elements (HRE). After recruiting numerous tissue specific co-regulators and RNA polymerase II (Pol II), the hormone-receptor-complex activates gene expression of various RNA classes (Figure 1, [2.]).

The steroidal hormone implant that was used in the animal trail analyzed in this thesis, contained estradiol benzoate and progesterone, a powerful combination that is responsible for induction of muscle hypertrophy [6]. The steroidal spectrum of activity in male cattle has not yet been elucidated.

1.3. Genes do not lie: transcriptomics as holistic molecular profiling technique

Veterinary drug residue analysis claims fraud-proof and effective detection methods as chromatographic systems reach their limits. New and innovative techniques have emerged in veterinary medicine during the last years and allow the detection of veterinary drug abuse via endogenous molecular biomarkers on the transcriptomic, proteomic or metabolomic level [10]. Rapid technological advancements in these '-omics' sciences allow a comprehensive high-throughput and holistic screening for differentially expressed biomarkers. Thus, according to a physiological condition, disease status or drug application, a reliable biomarker signature is capable of revealing specific biological traits or a measurable change in the organism [11]. From the molecular point of view, the transcription of genes is a fast and highly dynamic process that adapts to environmental stimuli, such as drug application. This renders the transcriptome as promising candidate for the discovery of transcriptional biomarkers (Figure 2).

The monitoring of messenger RNA (mRNA) expression ratios was already proven to be a powerful tool for biomarker development to trace growth-promotor abuse [12], [13], [14]. The 'RNA universe' of the transcriptome, however, covers a multitude of different RNA subclasses besides mRNA, e.g. transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), long non-coding RNA (lncRNA) and also the group of small RNAs. Especially the small RNAs are very promising and up to date a highly investigated new source in the search for biomarkers. Their characteristics and advantages are presented in the following chapter.

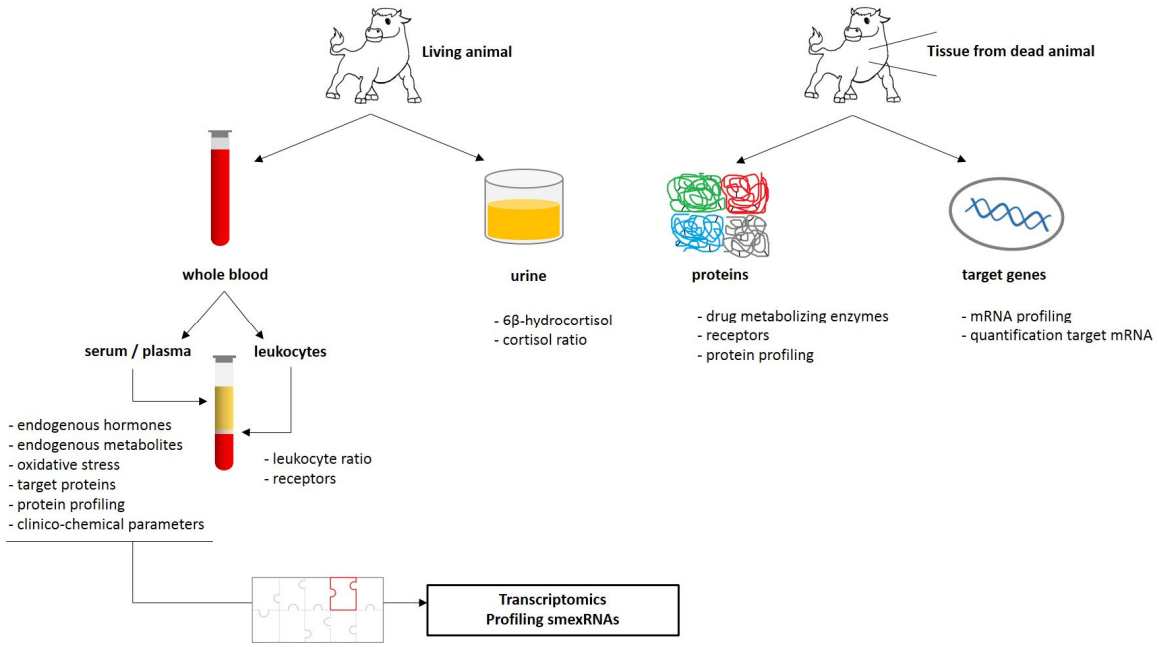


Figure 2: Profiling the transcriptional landscape of smexRNAs to discover veterinary drug residues via biomarkers is the missing puzzle piece (image adapted from [15]).

1.4. New source of transcriptomic biomarkers: small RNAs

1.4.1. Mighty regulators of post-transcriptional gene expression: microRNAs

MicroRNAs (miRNAs) were initially described in the year 1993 in the nematode model organism *Caenorhabditis elegans* [16]. Seven years later, the detection of miRNA in humans virtually initiated a scientific gold rush. The scientific community very rapidly explored the steps in miRNA biogenesis. It was also proven that miRNAs interact with their target mRNAs and are therefore mighty regulators of post-transcriptional protein expression. Friedmann and coworkers estimated that >60% of human protein-coding genes are confident targets of miRNAs, which explains their importance in the involvement in nearly all cell-biological pathways and functions, including differentiation, proliferation, apoptosis and even oncogenesis [17]. The potential of miRNAs as innovative biomarkers was demonstrated for various diseases, for example in cancer research in the year 2002 or for cardio-vascular disease in 2006 (as reviewed in [18]).

It is presumed that around 1-3% of the human genome encodes for miRNAs [19]. The mature single-stranded miRNA transcripts are located in the cytoplasm after precursor pri- and pre-miRNA processing in the nucleus and have a length of 19-25 nucleotides (nts). MiRNAs are non-coding and capture their target mRNA by perfect complementarity of the 'miRNA seed sequence', which is represented by 6-8nts at the miRNA 5´end. miRNAs mainly bind the mRNA at the mRNAs

3´untranslated region (UTR) or in the coding region [20]. As a result, the protein-coding message is impacted post-transcriptionally, depending on the degree of seed sequence complementarity. In the case of a perfect match, the mRNA transcript is destabilized and finally degraded, and if the matching is partly, mRNA translation is repressed [21].

In 2008, miRNAs were also detected as free, extracellular RNA (exRNA) in the bloodstream [22] and the potential usability of circulating nucleic acids as biomarkers was promptly recognized and investigated. Since then, circulating extracellular small RNAs (smexRNAs) have been detected in other human body fluids, e.g. milk, saliva, tears, cerebrospinal fluid, urine etc. [23]. The presence of smexRNAs in whole blood, plasma and serum as well as other liquid biopsies set off the second wave in miRNA biomarker research.

Multiple advantages made and still make the smexRNAs to goldmines in biomarker development:

- Accessibility – SmexRNAs are found in non- or minimal-invasive sample specimens. Compared to tissue biopsies, sampling is simplified and less harming for the patient [23], [24].
- Stability – smexRNAs were demonstrated to be reasonably stable in the living organism and after sample collection. Variations in pH-value and temperature, repeated freeze-and-thaw cycles and even RNase digestion did hardly affect the integrity of clinical samples [24].
- Novelty – working with this newly discovered RNA classes still allows biological breakthroughs, which increases the attractiveness of this research field.
- Successfulness – the applicability of smexRNA biomarkers for different diseases, e.g. in different cancer types [24], [25], myocardial infarction [26], [27] or central nervous system diseases [28] was already demonstrated.
- Practicability – more and more skilled diagnostic tools and technologies as well as bioinformatics evolve to analyze smexRNAs from different sample matrices and species.

The importance and timeliness of miRNA research is reflected by the increased number of published miRNA sequences in the miRBase registry (http://www.mirbase.org/) and by the miRNA publication record in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) over the past years (Figure 3, [A]) [29]. Up to date (August 2015), the online miRNA repository miRBase (Release 21) provides information about the location, sequence and annotation of 35.828 mature miRNA entries from 223 species, whereof 2.588 miRNAs were classified as human and 793 as bovine [30]. Therefore, the known human miRNA sequences have more than doubled since 2010.

1.4.2. New kids on the small RNA block: PIWI-interacting RNAs (piRNAs)

PiRNAs are a newly identified class of small non-coding RNAs that are slightly longer than miRNAs (25-32 nts) and can show specific traits. According to the piRNA biogenesis model, the so-called ping-pong model, piRNAs can be divided into primary and secondary piRNAs [31]. Primary piRNAs have a strong tendency for a 5` bias for uridine (U) and do not have nucleotide bias at position 10, while secondary piRNAs have a bias for adenine (A) at position 10, but do not show a 5` bias. PiRNAs bind to the piwi- (P-element induced wimpy testes)/argonaute protein family, whose function was primarily discovered in a *Drosophila*-mutant in 1998 [32]. However, as miRNAs are the most intensively studied small RNA class so far and piRNA molecules rather occupy a subordinate role, their importance as additional post-transcriptional regulators of gene expression is emerging nowadays. PiRNAs play key roles in the development of all animals. Studies in diverse model organisms have suggested that piRNAs have major functions in the germline development [33], [34], [35], [36] in twofold aspects. On the one hand, piRNAs regulate transposable elements in the germline [37], [38] and on the other hand, they form PIWI-piRNA complexes that regulate histone modifications [39]. Recently, roles in developmental biology were very clearly demonstrated in bovines by Roovers and coworkers [40]: sequencing small RNAs isolated from bovine testis and oocytes revealed that these tissues strongly express piRNAs, but relatively few miRNAs, which is in accordance with findings in human, mouse and rat. Moreover, more than 90% of all sequenced testis piRNAs come from the same loci. In the ovary, 84% of all sequenced piRNAs can be produced by their cluster loci, which are mainly testicular piRNA cluster loci. Additionally, they could show similar piRNA expression in both male and female germ-cell differentiation and highlighted parallels between oocytes maturation and spermatogenesis. Although research is focusing on the investigation of gamete small RNAs, the presence of circulating miRNAs and piRNAs could already be proven in bovine plasma and whole blood [41] (Appendix II). Tissue-specific smexRNAs could be found in the circulation giving strong indication that they are not locally restricted to the tissue they were originally expressed, but could also be released and transported throughout the organism. Hence, the function of piRNAs and their interactions in forming PIWI-protein complexes is not yet fully understood. Generating landscapes of piRNAs also in somatic tissues and in body fluids facilitated comparative studies and biomarker discoveries, e.g. in human bladder cancer [42] and in human saliva [43].

In the last years, the number of published research articles about piRNAs increased in the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed/?term=piRNA) (Figure 3, [B]), which signifies the trend in exploring new targets. Currently, 32.046 entries are compiled in the piRNA sequence collection of the nucleotide database of NCBI Genbank ([44], Release 201).

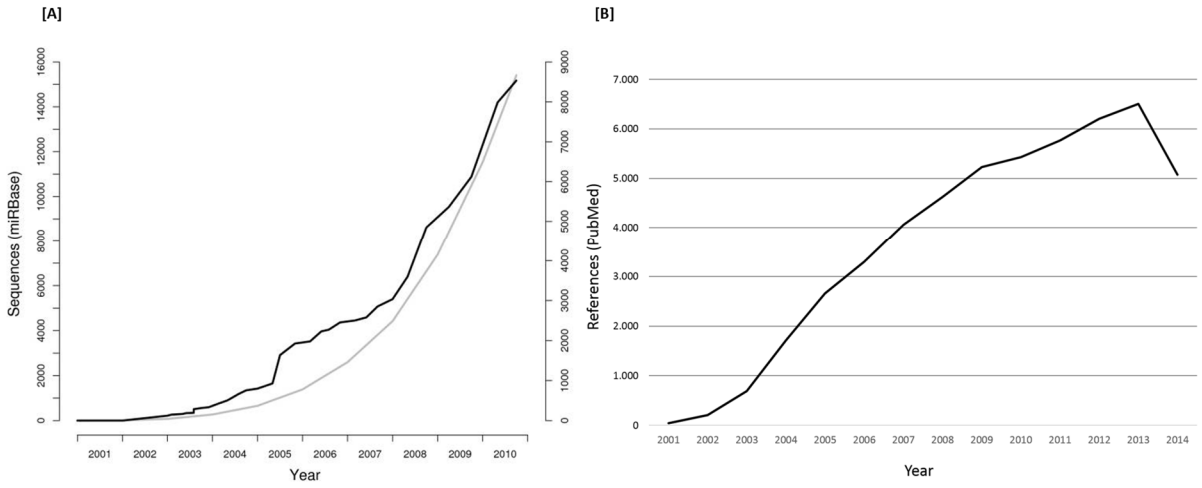


Figure 3: [A] Development of the known miRNA sequences in the miRBase database (black line) and the miRNA references in PubMed (grey line) [29]. [B] Growing number of piRNA references in PubMed.

1.5. RNA-Sequencing facilitates global profiling of the transcriptional landscape

Next-Generation Sequencing (NGS) realizes the holistic sequencing-by-synthesis analysis of whole transcriptomes. If the aim is to gain global sequence information of a samples` characteristic small RNA spectrum, small RNA-Sequencing (small RNA-Seq) is the state-of-the-art technology. This highly sophisticated method allows the sequencing in a high-throughput approach by simultaneously measuring up to 24 samples. However, high-throughput experiments per se do not lead to biological findings. Often massive seq-data is hoarded, but the following bioinformatical data processing steps are neglected or underestimated. Only essential processing (e.g. quality controls (QC), data trimming and of course alignments and mappings) provides access to further interpretation of the collected data.

1.6. Analysis of transcriptomic data aims to point out key small RNAs as potential biomarkers

Ultrahigh-throughput profiling experiments, like NGS studies, result in a data output of tremendous volume and highly multivariate character (k variables ≫ n observations). Therefore, the main goal in biomarker identification is the extraction of a characteristic signature that is hidden behind rich data sets. To identify a biomarker that precisely discovers a certain treatment necessitates highly elaborated multivariate projection methodologies, like orthogonal partial least squares discriminant analysis (OPLS-DA). This algorithm allows visualization and interpretation of highly multivariate and extensive data sets and facilitates carving out the class separation information between a control group and treated individuals [45]. OPLS-DA was implemented within this thesis to circumvent this essential bottleneck in the search for biomarker signatures [46] (Appendix IV).

1.7. Aim of the dissertation

This thesis is focusing on the discovery of innovative detection and profiling techniques to identify meat-producing cattle that was illegally treated with anabolic growth promotors. Therefore, an animal trial (AnaBull study) was conducted to simulate the real environment during a potential drug abuse situation using two different groups of anabolic drugs: steroid hormones and β-agonists. As mRNAs were already successfully validated as transcriptional biomarker candidates, the usability of small RNAs should be investigated, as those act as direct regulators of mRNAs. To be more precise, smexRNAs in bovine plasma should be the subject of analysis due to diagnostic advantages, such as increased stability and resistance as well as easy sampling of liquid biopsies (Figure 4).

Within this doctoral project, two sequencing approaches were applied with different goals. First, the physiological status quo of the global smexRNA transcriptome in plasma and whole blood of healthy calves was profiled to establish an initial basis for comparative studies. And second, in the context of the AnaBull study, smexRNAs from liquid biopsies were sequenced and assessed with multivariate projection methodologies to examine and discuss their potential as novel source of biomarkers in veterinary diagnostics (Figure 4).

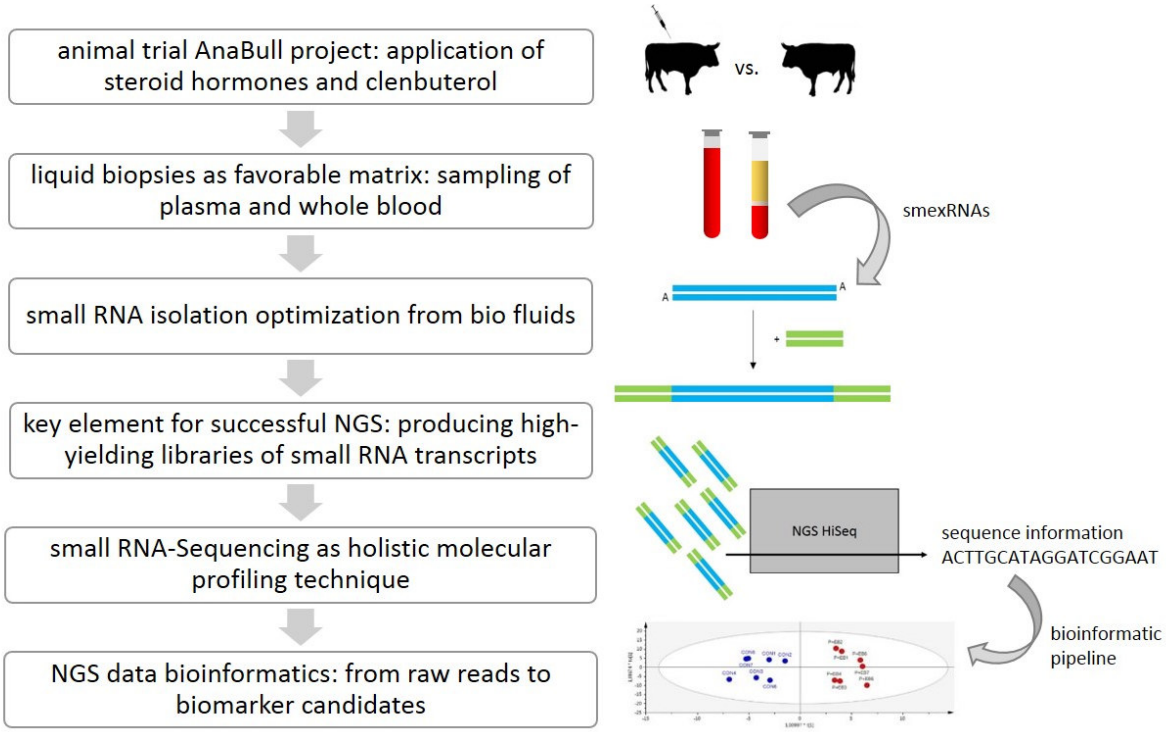


Figure 4: Milestones to identify transcriptional biomarker candidates for veterinary drug abuse.

2. Materials and methods

2.1. Animal experiment 'AnaBull project': Application of anabolic substances to bull calves

To study the influence of anabolic substances on gene expression profiles in meat-producing livestock, an animal trial was conducted to simulate the real environment during a potential drug abuse situation. This study is hereinafter referred to as 'AnaBull Project' and was approved by the ethical committee of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (Recklinghausen, Germany) (permit number 84-02.04.2012.A040). Experimental animals were housed and fed according to good animal attendance practice. The trial was organized and conducted in collaboration with the Kontrollgemeinschaft Deutsches Kalbfleisch e.V. (KDK, Bonn, Germany), Société Générale de Surveillance (SGS, Geneva, Switzerland) and Denkavit (Warendorf, Germany) who provided the experimental animals and handled livestock husbandry. Animals were housed in an appropriate and separated stable at the fattening farm run by Jürgen Wickentrup in Nordrhein-Westfalen (Wadersloh, Germany). Animal health monitoring, implantations and blood sampling was supported by the local veterinary practice owned by Stefan Lüllmann (Löningen, Germany).

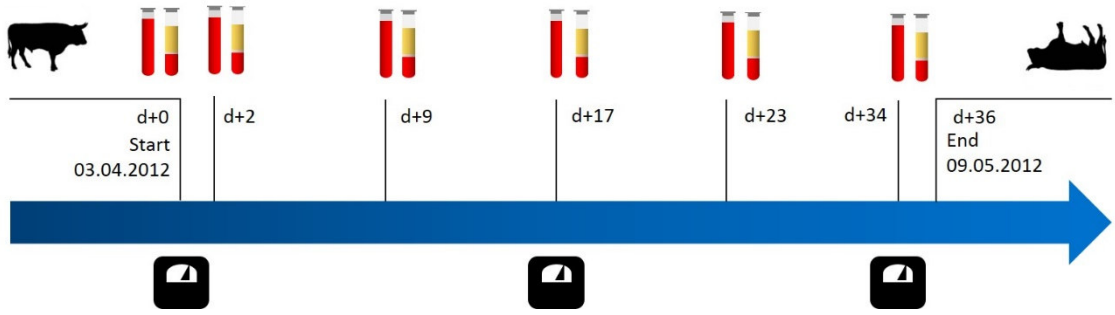


Figure 5: Sampling schedule for the AnaBull project. The animal trial started with blood sampling (indicated by red tube), plasma sampling (red and yellow tube) and weighing (balance symbol) of all 21 calves. Over the experimental period, liquid biopsies were taken as indicated and weight gain was monitored in the middle of the trial and at the end.

In this study, 21 male Friesian Holstein veal calves (*bos taurus*, bta) were randomly divided into three groups of 7 animals each (n=7). All the animals had a similar age (161 ± 15 days) and an average body weight of 151.4 ± 19.2 kg. One group remained completely untreated and served as control group

(CON group). The second group was treated with an oral dose of clenbuterol-hydrochloride (10 µg/kg body weight) (Boehringer Ingelheim, Ingelheim am Rhein, Germany) in daily intervals (CLEN group). The drug-containing liquid was orally gavaged into the animal´s mouth using a single-use plastic syringe. To prevent a contamination of personnel, equipment and animals from other groups, special care was taken in handling clenbuterol by e.g. wearing gloves during clenbuterol application, preventing clenbuterol-treated animals to touch other animals, using an external stable away from the regular calf fattening stable, prohibiting access of unauthorized persons and wearing clothes that were used in that stable only. The third group was treated with Component E-C (IVY Animal Health, Kansas, USA), a hormonal implant consisting of a combination of 100 mg progesterone plus 10 mg estradiol benzoate (P+EB group). The implant was deposited between the skin and the cartilage on the backside of the middle third of the pinna of the ear. The depot capsules released their hormonal content over a fixed period of time. The steroidal compounds were injected by a trained veterinarian according to good veterinary practice. To monitor the animals weight gains, all animals were weighed prior to the experiment as well as at d+17 and d+34 (Figure 5). The animal trial was terminated at d+36. Animals were sacrificed at the slaughterhouse of the Bayerische Landesanstalt für Tierzucht (Poing-Grub, Germany) and condemned material was properly disposed.

2.2. Liquid biopsies as favorable matrix: sampling of plasma and whole blood

For plasma generation, peripheral blood was taken from the *vena jugularis* from all calves at d+0, d+2, d+9, d+17, d+24 and d+34 (Figure 5) and processed as described before [47] (Appendix I). Whole blood was collected with PAXgene Blood RNA Tubes (PreAnalytiX, Germany) that guaranteed immediate stabilization of RNAs. PAXgene tubes were stored at -20°C until RNA extraction.

2.3. Isolation of high-quality nucleic acids from bio fluids

2.3.1. Optimized extraction technique for a challenging sample matrix: bovine plasma

There are several protocols and commercial kits for the extraction of circulating RNAs from plasma, but due to the marginal amount of cell-free RNAs in bovine plasma samples, the total RNA yield is insufficient to perform small RNA-Seq. For the holistic characterization of the small RNA transcriptome in bovines, the extraction of smexRNAs from plasma samples had to be optimized. The extraction optimization and the successful generation of high quality sequencing data was performed as published in Spornraft et al. [47] (Appendix I).

2.3.2. Total RNA isolation procedure from bovine whole blood

Total RNA purification including smexRNAs from whole blood was performed using the PAXgene Blood miRNA Kit (PreAnalytiX, Germany). Since whole blood contains cellular components and therefore a sufficient concentration of total RNA for measurement, the manufacturer´s protocol could be followed without limitations.

2.4. Ascertaining RNA quantity and quality as control checkpoint prior to sequencing

Subsequently to RNA isolation, yields were measured using the Qubit 2.0 Fluorometer (Life Technologies, Germany) in combination with the RNA Assay Kit (Life Technologies, Germany) according to the manufacturer's guidelines. To resolve the smexRNA fraction in the size range from 6 to 150 nts, a Bioanalyzer 2100 (Agilent Technologies, Germany) run using the Small RNA Kit (Agilent Technologies, Germany) was performed. Until library preparation, the extracted RNA was stored at -80°C [47] (Appendix I).

2.5. The key element for successful NGS: producing high-yielding libraries of small RNA transcripts

For small RNA-Seq, the isolated transcripts were converted into barcoded complementary DNA (cDNA) libraries as described by Spornraft et al. [47] (Appendix I).

2.6. The state-of-the-art profiling technique for the analysis of the small RNA transcriptome: Small RNA-Sequencing

The quality-checked sequence libraries of RNA transcripts were subjected to the Illumina small RNA-Seq sequencing pipeline as reported before [47] (Appendix I). Samples from d+17 were chosen for sequencing, since a pilot study revealed this time point as best suited for the surveillance of hormone misuse in meat producing animals [48]. Sequencing was carried out at the Genomics Core Facilities of the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

2.7. NGS data analysis: from raw sequencing read information to annotation and biological interpretation

The data evaluation and mapping pipelines were developed in-house and published in Spornraft et al. [47] (Appendix I). A schematic overview is given in Figure 6. All miRNA and piRNA readcounts were normalized to the sequenced library sizes in reads per million (rpm).

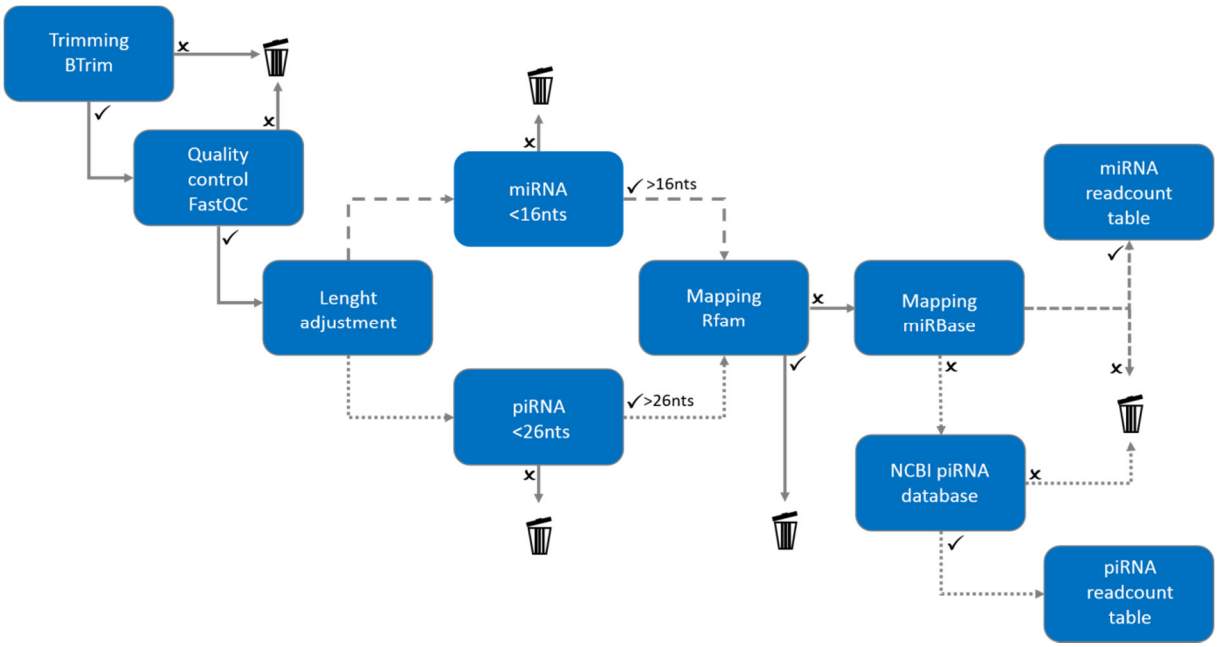


Figure 6: Overview of the QC and successive mapping processes to generate precise miRNA and piRNA readcount tables. Reads that passed a quality checkpoint were further processed in downstream data evaluation and pathways were indicated with checkmarks (✓), reads that failed were excluded and marked with a cross (x).

2.8. Highly sensitive miRNA screening with reverse transcription quantitative polymerase chain reaction (RT-qPCR) to identify a specific time profile

NGS and the downstream bioinformatics revealed the holistic spectrum of miRNAs in the sequenced plasma samples in the middle of the AnaBull study at d+17. The major expressed constituents were selected for a profiling of potential expression changes over the time course from the beginning of the trial (d+0) until the end (d+34). For the verification, each high scoring miRNA was measured using RT-qPCR with specific primer assays.

2.8.1. Purification of miRNAs for the expression analysis in plasma: including all time points of the AnaBull study

For miRNA profiling using RT-qPCR, total RNA was isolated with the miRCURY RNA Isolation Kit – Biofluids (Exiqon, Denmark). In brief, plasma was treated prior to RNA isolation by an additional centrifugation step to guarantee cell-free specimens (3.000 x rcf, 5 minutes, room temperature). Recommended usage of spike-ins (RNA Spike-In Kit, miRCURY LNA Universal RT microRNA PCR, Exiqon, Denmark) for monitoring RNA isolation efficiency as QC parameter and carrier RNAs to optimize RNA extraction procedures was followed. Bio fluid particles were lysed using the provided lysis solution. Proteins were precipitated and isopropanol (Sigma-Aldrich, Germany) was added to the transferred supernatant after phase separation. To exploit the maximum volume of starting material that is supported by the kit system, 900 µl of plasma was utilized. Therefore, volumes of lysis solution, protein precipitation solution and isopropanol were adequately scaled up. Loading the column was done repeatedly in aliquots of 700 µl. The optional on-column DNase digest was performed as described in the manual to remove residual amounts of genomic DNA contamination that could affect sensitive downstream RT-qPCR measurements. Bound RNAs were washed with supplied washing buffers. In order to maximize yields, elution with RNase free water was done in two steps eluting 25 µl each. Purified RNA samples were snap-frozen in liquid nitrogen before long-term storage at -80°C.

2.8.2. Optimization of the input volume of transcript molecules into RT reaction to inhibit PCR inhibition

Since the levels of total RNA found in plasma are very low, carrier RNA (MS2 bacteriophage total RNA, Roche, Germany) was applied to ensure robust isolation. Determination of the RNA concentration of given samples after RNA extraction is therefore impossible. Consequently, the volume of starting material and not the concentration was used as a measure of the input amount. This procedure was also recommended by the manufacturers of the used kit system.

RNA inhibitors in plasma of bovines could prevent appropriate cDNA synthesis and therefore the amplification of miRNAs. Before starting the profiling experiment, the quantity of extracted RNA that needs to be inserted in the RT reaction without experiencing qPCR inhibition, was determined. Individual assays with UniSp2-/ UniSp4-/ UniSp5- and miR-122 primers were run in triplicates with different amounts of RNA input volumes: 2 µl, 4 µl, 6 µl, 8 µl, 10 µl, 12 µl and 14 µl per 20 µL cDNA synthesis reaction.

2.8.3. Amplification and simultaneous quantification of miRNAs in AnaBull plasma specimens: RT-qPCR

The miRCURY LNA Universal RT microRNA PCR system (Exiqon, Denmark) is a two-part protocol consisting of the first-strand cDNA synthesis and the real-time PCR amplification. The individual assays were performed as described in the manufacturer´s instruction manual. For the RT reaction master mix, 4 µl 5x reaction buffer, 1 µl nuclease-free water, 1 µl RNA spike-in template (UniSp6) and 2 µl enzyme mix were combined on ice before individually adding 12 µl of template RNA eluate. Reagents were gently mixed and spun down. The RT setup was incubated at 42°C for 60 minutes with a following heat-inactivation of the enzyme at 5°C for 95 minutes and a cooling step to 4°C, using the Mastercycler gradient (Eppendorf, Germany). cDNA templates were diluted with nuclease-free water (1:40) in DNA LoBind tubes (Eppendorf, Germany) immediately before qPCR. Design LNA-enhanced microRNA PCR primers (Exiqon, Denmark) were customized and resuspended as described in the manufacturer´s protocol. Target sequences (5`-3`) are listed in Table 1. The PCR mix/primer working solution (5 µl PCR master mix + 1 µl PCR Primer) was prepared on ice and dispensed into 384 well plates (Frame Star, 4titude, UK) before adding 4 µl of cDNA template.

Table 1: Primers used in the profiling experiment and their ordering information.

Primer name	Sequence	Accession number miRBase	Product number Exiqon
bta-miR-122-5p	UGGAGUGUGACAAUGGUGUUUG	MI0005063	205664
bta-miR-423-3p	AGCUCGGUCUGAGGCCCCUCAGU	MIMAT0003831	204488
bta-miR-423-5p	UGAGGGGCAGAGAGCGAGACUUU	MIMAT0012537	205624
bta-miR-320a	AAAAGCUGGGUUGAGAGGGCGA	MIMAT0003534	204154
bta-miR-99a-5p	AACCCGUAGAUCCGAUCUUGU	MIMAT0003537	205945
bta-miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU	MIMAT0009383	204258
bta-miR-193a-5p	UGGGUCUUUGCGGGCGAGAUGA	MIMAT0003794	204665
bta-miR-192	CUGACCUAUGAAUUGACAGCCAG	MIMAT0003820	206999

Quantification of miRNAs was carried out using the CFX384 real-time PCR detection system (Bio-Rad, Germany) with the following qPCR cycle conditions: The process was initiated by an incubation step at 95°C for 10 minutes followed by 40 repetitive amplification cycles at 95°C for 10 seconds and 60°C for 1 minute. Melting curve analysis was done by increasing temperature from 50°C to 85°C with an

increment of 0.5°C/5 seconds. Expression levels were recorded as the quantitative cycle (Cq) that was derived using the second derivative maximum method identifying the Cq of a sample as the cycle number where the sample´s fluorescence is detected above the background noise and the amplification is in its exponential phase [49]. The obtained melting curves and quantification data were analyzed with the CFX384 Manager software (Bio-Rad, Germany). Relative gene expression data was normalized and analyzed according to the Pfaffl equation [50]. All animals from d+0 as well as the untreated animals from d+2 to d+34 were used as controls. If more than three values were missing from RT-qPCR in one treatment group (n=7) and day, the time point was excluded from analysis. The NormFinder algorithm [51] in the GenEx software bundle (Version 5) (MultiD Analyses, Tattaa Biocenter, Sweden) was used to determine the optimal reference gene set.

2.9. Explosion of data in -omics sciences: Mastering complex data sets with sophisticated statistical suits

2.9.1. Univariate data analysis for testing statistical significance

Statistical significance of miRNA expression levels of treated groups compared to the controls was tested by calculating p-values using paired t-tests in Sigma Plot (Systat Software Inc., USA, Version 11.0).

2.9.2. Multivariate data analysis for empowering cluster identification

To explore, analyze and extract valuable information from big data sets, the powerful multivariate data analysis package, SIMCA 13.0.3.0 software (Umetrics AB, Sweden) was used as described in Spornraft et al. [46] (Appendix IV). SIMCA is equipped with extensive implements for dedicated multivariate methods to cluster data sets and generate models.

3. Results and discussion

3.1. Verification of the effectiveness of anabolic treatment: monitoring the weight gain

To test the phenotypic effectiveness of anabolic substances during the AnaBull project, the animal´s weight gains were supervised (Figure 7). Weight gain was monitored after the first half of the experiment (d+0 to d+17) and after the second half (d+17 to d+34) as well as after the overall treatment period (d+0 to d+34). In summary, there was not a statistically significant difference between the weights of animals from the control group and the P+EB group ($p=0.4$) or the CLEN group ($p = 0.6$) prior to treatments. The weight gain of animals from the P+EB group ($p = 0.028$) and the CLEN group ($p = 0.019$), respectively, was significantly different at d+17. After the second half, the weight gain differed significantly between the CON and both treatment groups (P+EB: $p=0.0005$; CLEN: $p=0.004$). A highly significant increase in weight was shown in both treatment groups throughout the entire period of the experiment compared to the CON animals (P+EB: $p=0.0015$; CLEN: $p=0.0021$). Therefore, it could be concluded that anabolic treatments were successful over the whole experimental period. These results were already published by our group [14] (Appendix III).

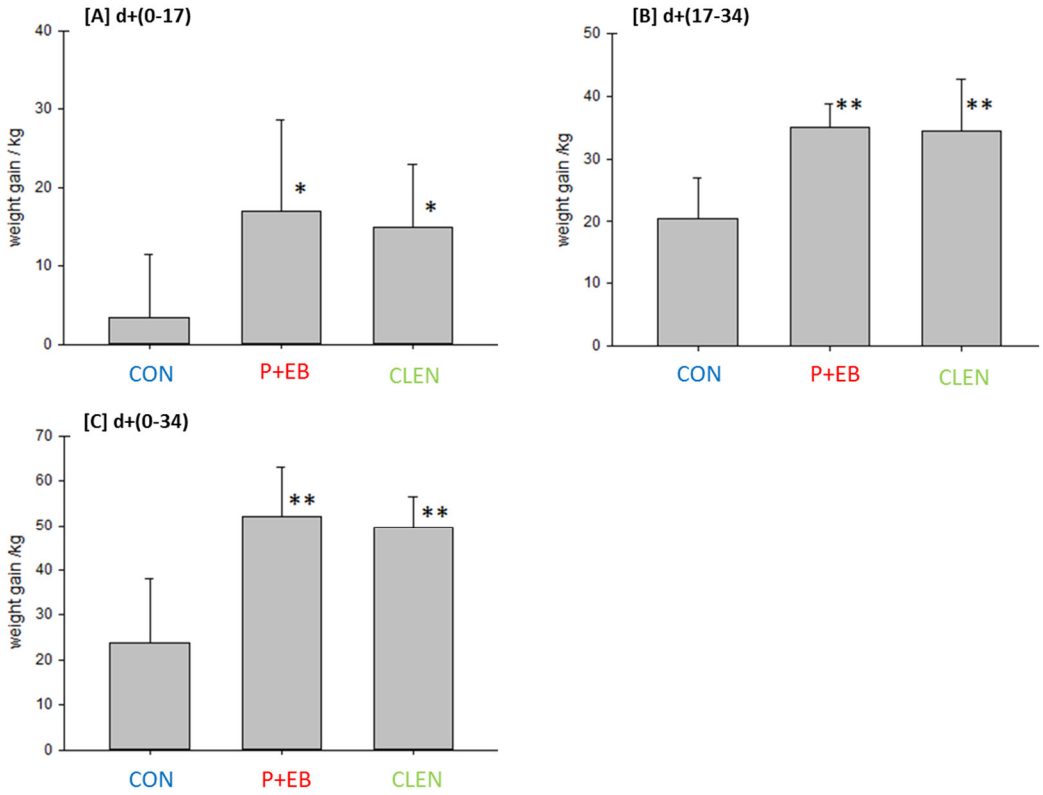


Figure 7: Weight gain development of AnaBull animals was monitored from the start of the trial until d+17 [A], from d+17 until d+34 [B] and image [C] shows the overall weight gain. Bars marked with an asterisk (*) show a significant weight gain difference between the CON group and the treatment group with $p < 0.05$. Two asterisks (**) signify highly significant differences ($p < 0.01$).

3.2. The crucial first step: extracting sufficient concentrations of smexRNAs from the methodically challenging bovine plasma sample matrix

RNA yields of plasma samples that were extracted with commercially available kit systems were not quantifiable by spectrophotometry due to unmeasurably low RNA concentrations.

Since there was no fundamental work performed on technological feasibility of smexRNA extraction from bovine body fluids, suitable methods needed to be studied first. Therefore, the isolation process was optimized to fulfill pre-Seq quantity and quality features [47] (Appendix I).

Applying an upgraded input volume of 6 ml using the miRNeasy Serum/Plasma Kit in combination with the QIAvac system, the Bioanalyzer Small RNA Series II Assay (Agilent Technologies, Germany) initially showed measurable signals. Best results were achieved, if extraction was performed with a starting plasma volume of 9 ml (Figure 8) [47] (Appendix I).

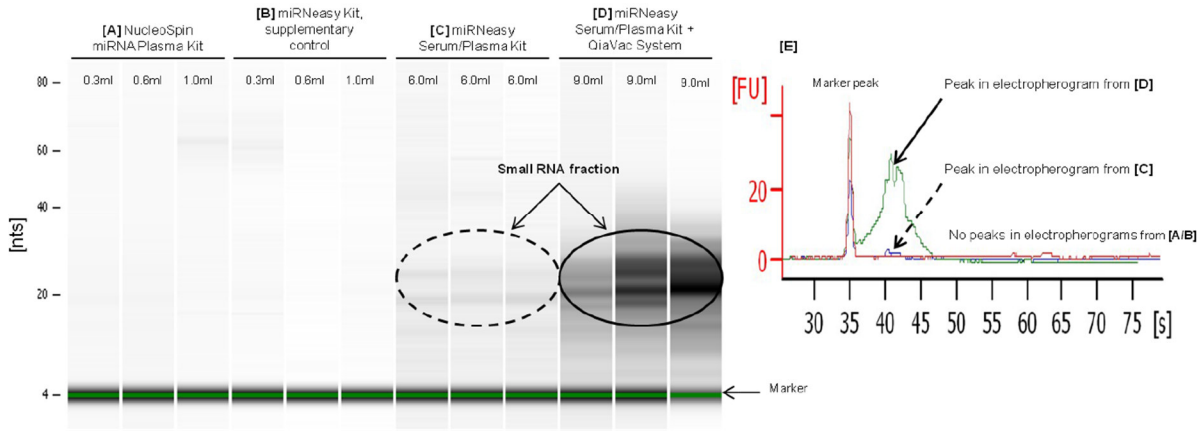


Figure 8: By increasing the initial plasma volume [A-C] up to 9 ml (image [D]), the quantification assay confirmed the successful isolation of measurable RNA concentrations that were prerequisite for NGS. Image [E] mirrors the results from the gel image as electropherogram to illustrate RNA quantity [47].

3.3. The second step prior NGS: generating high-quality and size-controlled transcript libraries

When the samples were sequenced in 2012, a minimum input amount of 1-10 µg of total RNA was necessary for the library preparation step prior to sequencing. This relatively high input amount was mandatory due to quality checkpoints during library preparation such as Bioanalyzer 2100 (Agilent Technologies, Germany) runs or concentration measurements using the Qubit 2.0 fluorometer (Life Technologies, Germany). In 2015, 100 ng to 1 µg are needed for the same library preparation kit or even less with newly developed systems. For example, 10 ng are used in the NEBNext Ultra RNA Library Prep Kit for Illumina from New England Biolabs (New England Biolabs, USA) and the TruSeq RNA Access Library Prep Kit (Illumina Inc., USA). Therefore, the trend is going towards using less and less starting material, which is realizable as we achieved high-quality seq data with 35 ng extracted RNA as avant-garde starting material for bovine applications [47] (Appendix I). Accurate size and quantity assessments of library molecules was critical to prevent biased signals and unused sequencing capacity. The generation of sequencing libraries was not impaired by the reduced inserted transcript density and target RNAs were successfully amplified (Figure 9) [47] (Appendix I).

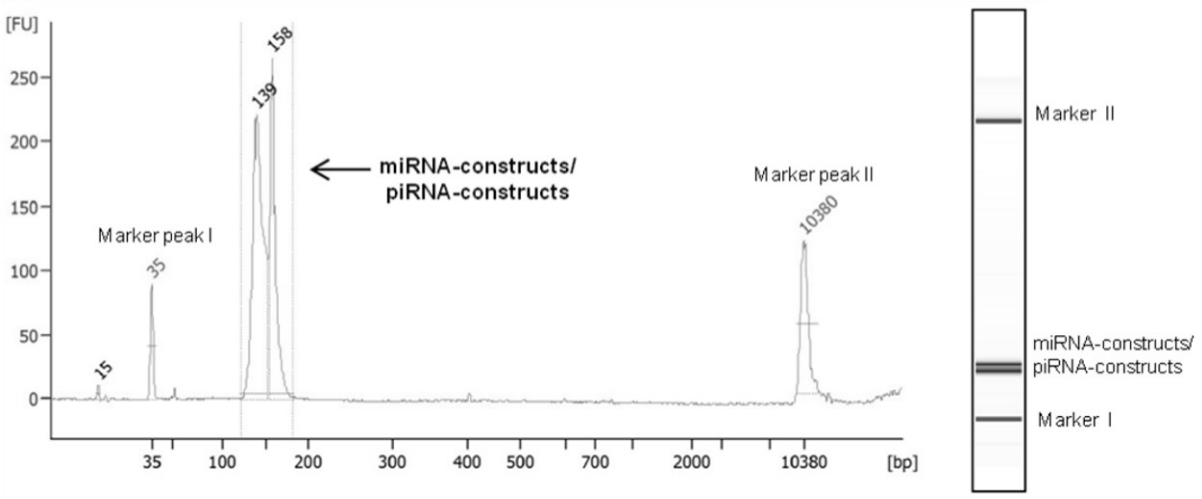


Figure 9: Capillary electrophoresis as final quality and quantity validation after library preparation. All samples showed target constructs in adequate sizes [47].

3.4. Technical feasibility of transcriptome profiling using small RNA-Seq

After prearrangements were optimized, bovine plasma could be fingerprinted using a holistic and integrated NGS approach. Although the pre-sequencing procedures deviated from the standard protocols, small RNA-Seq yielded high quality data [47] (Appendix I). The in-house data evaluation, including mapping and annotation of sequenced target transcripts, proofed the experiment to be effective [47] (Appendix I). NGS is a highly advanced and fast developing technology. Manufacturers offer more and more specialized and improved kit systems for diverse applications. Nevertheless, pre-sequencing steps and bioinformatics in the downstream data analysis are still not designed and certified for other samples than human. To pave the way to NGS for diverse research projects, there is still a need for more specific applications. To render further achievements in veterinary drug control and food safety, sequencing applicability needs to be extended for animal sciences in general and especially for livestock research.

3.5. Comparison of the miRNome and piRNome of bovine whole blood and plasma

The smexRNAs of bovine plasma and whole blood were analyzed using NGS with the objective to quantify, profile and compare the miRNA and piRNA signatures in both bio fluids. Before screening the AnaBull samples for biomarkers, the status quo of healthy individuals was investigated to set a basis for future comparative studies. With knowing the healthy physiological footprint of smexRNAs, altered signatures could help to detect e.g. ruminant pathologies, track physiological changes (e.g. pregnancy) or proof the illegal administration of anabolic substances or growth promotors.

3.5.1. Differences in data evaluation and smexRNA abundance between whole blood and plasma

In whole blood, more reads passed the data trimming pipeline and annotation as described in Spornraft et al. [41] (Appendix II). Blood data wasn´t subject of great variation, compared to higher variability in plasma. This was highlighted in the obviously more unified picture of blood parameters (Figure 10). In plasma from untreated control animals, 5.0 ± 2.9% of miRNAs could be identified, while 38.2 ± 3.4% could be identified in whole blood of the same individuals (Figure 11, [A]). The share of piRNAs in both sample specimens was nearly the same: 1.4 ± 0.8% of piRNAs in plasma and 1.9 ± 0.8% in whole blood (Figure 11, [A]). The AnaBull study revealed significant differences in miRNA expression. Steroid treatment led to a significant decrease of miRNAs compared to the CON group ($p=0.047$) (Figure 11, [B]). Additionally, there was a statistically significant difference between the treatment groups ($p=0.042$) (Figure 11, [B]), while clenbuterol treatment tended to promote miRNA up-regulation. Therefore, the two diverse veterinary drug substance classes seemed to have an opposite impact on miRNA genesis.

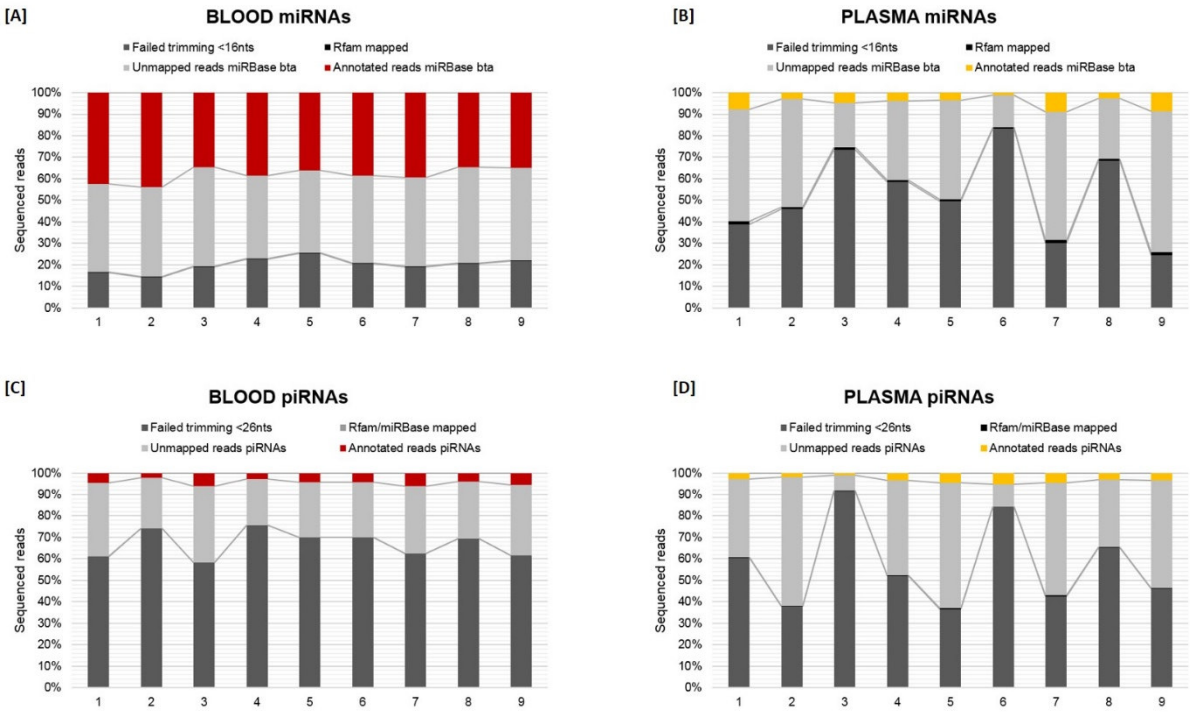


Figure 10: Data evaluation of size-filtered and successively mapped reads of sequenced healthy animals. [A-B] Data analysis pipeline for miRNAs in whole blood (red) and plasma (yellow). [C-D] Data processing for piRNAs [41].

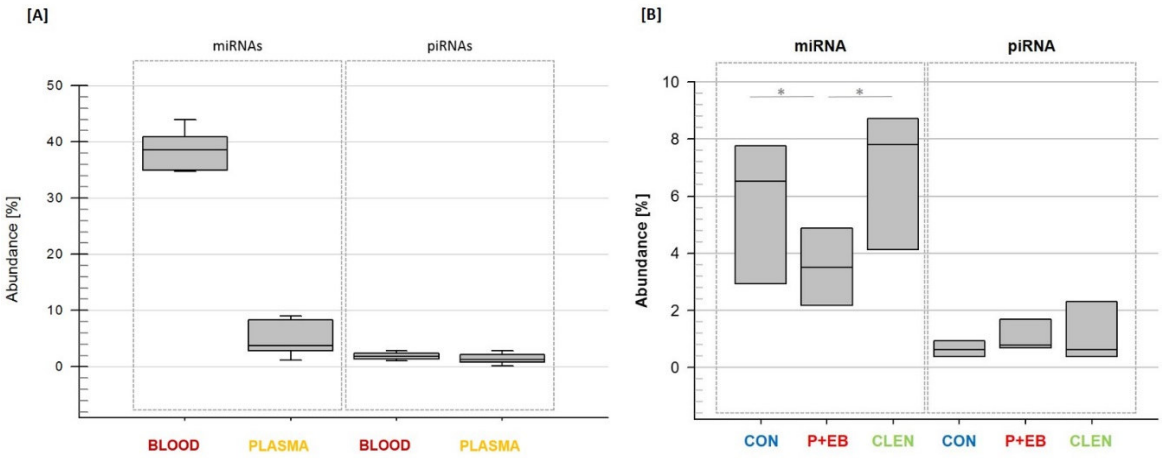


Figure 11: [A] Abundance of annotated smexRNAs in plasma and whole blood of untreated bovines (n = 9) [41]. [B] Box plots illustrate the abundance of smexRNA proportions in plasma of CON animals, P+EB and CLEN treated animals (n=7). Steroid hormones decreased miRNA expression (p=0.047) and clenbuterol application resulted in up-regulated miRNA genesis (p=0.042) [46].

3.5.2. Physiological differences in small RNA transcriptomes

Transcriptomes are often screened to find differences in expression ratios that are linked with diseases or altered physiological states. Besides our publication [41] (Appendix II), no information is given in literature about the physiological basic levels of miRNomes and piRNomes in bovines. After sorting reads according to readcount frequency of occurrence, comparative top 10 lists were created as previously reported [41] (Appendix II). Two miRNAs (bta-miR-486 and bta-miR-92a) and seven piRNAs (piR-43772, piR-43771, piR-60565, piR- 43786, piR-31068, piR-35982 and piR-3315) were similarly detectable in whole blood and in plasma, although the ranking order varied (Figure 12). The major component of the blood profile was bta-miR-451 with a share of 27.5%, followed by bta-miR-25 and bta-miR-191. The plasma signature was dominated by miR-122, occupying 54.0% of the top 10 readcounts. The two subtypes of bta-miR-423 were ranked second and third. The piRNome was determined by three piRNAs: piR-31068, piR-35982 and piR-33151 were mainly expressed in whole blood and piR-43772, piR-43771 and piR-60565 in plasma.

Data sets were matched for pairs that were present in both tissue matrices. As presented in Figure 13, several miRNAs as well as piRNAs in plasma exceeded their level in whole blood. Especially bta-miR-122 was found in increased abundances in plasma compared to whole blood. In literature, the release mode and distribution system of miRNAs are strongly discussed. It is supposed that miRNAs are transported throughout the body in shedding vesicles, in apoptotic bodies, in high-density lipoprotein particles, incorporated in exosomes or associated to proteins of the Argonaute family [52], but still many questions remain. Even less is known about the original sources of smexRNAs in the body and their destination in an organism. Whole blood contains blood cells that have their own capacity of RNAs, while plasma in contrast is a cell-free matrix that exclusively covers circulating, non-cell-bound RNAs. The presented contrast pointed out that the sum of smexRNAs in plasma was not merely the product of disrupted blood cells that released their RNA content. The signatures of matching miRNA and piRNA pairs in both bio fluids raised evidence that individual smexRNAs cannot exclusively originate from cellular blood components, such as red blood cells (RBC), thrombocytes or leukocytes. Interestingly, miR-122 is very well known to be liver-specific [53], [54]. Hence, its increased presence in plasma seems not to be the consequence of a release from blood cells. Furthermore, footprinting of a healthy cohort revealed that the most abundant miRNA in the human circulation was hsa-miR-122 [55], which evidently emphasized the importance of this miRNA. In literature, the role of miR-122 is well described as key player in maintenance of liver homeostasis as well as in hepatic metabolic, anti-inflammatory and anti-tumorigenic mechanisms [56]. However, a map that explains the allocation of miRNAs and piRNAs throughout an organism isn`t existing yet. This comparative study is a first step in reaching that goal for bovine research.

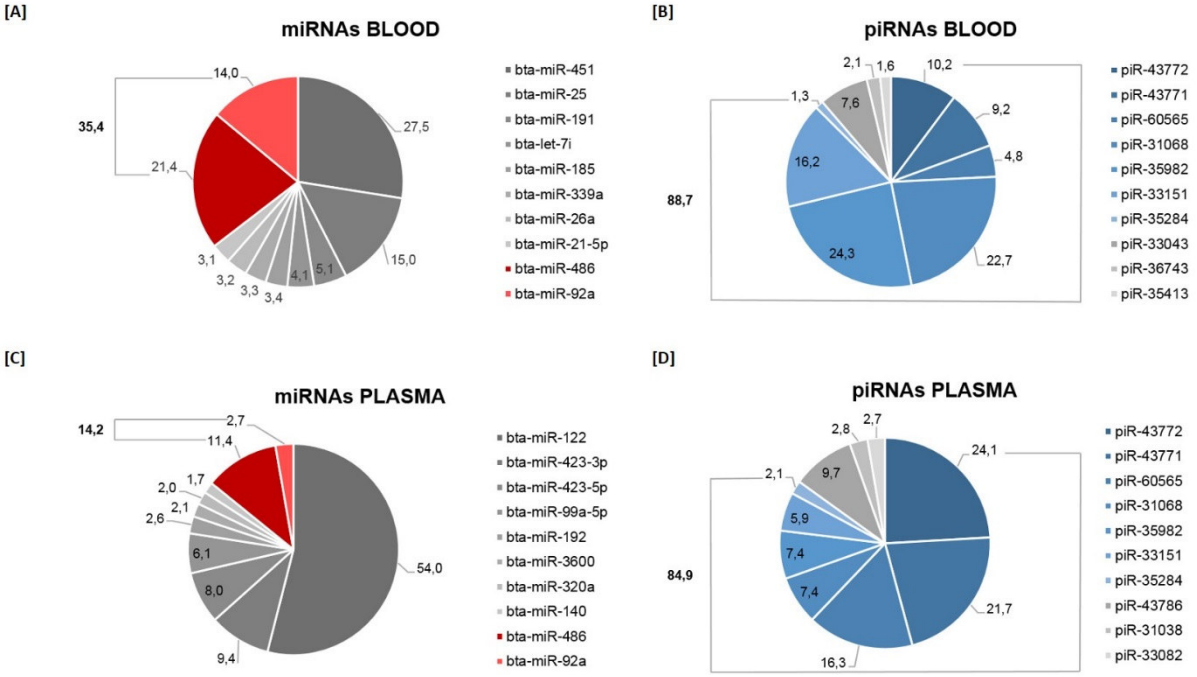


Figure 12: The ten most abundant miRNAs and piRNAs in whole blood [A, B] and in plasma [C, D]. MiRNAs that were found in both bio fluids were colored in red, identical piRNAs were marked in blue [41].

There is not yet a universally accepted data normalization strategy for miRNA quantification, posing a challenge for comparative data analysis. In the literature, several human reference miRNAs are recommended (miR-16, miR-92a, miR-144, miR-451, miR-486) [57], [58], [59], [60]. However, their applicability could not be attested in our data sets, as these miRNAs were marginally expressed.

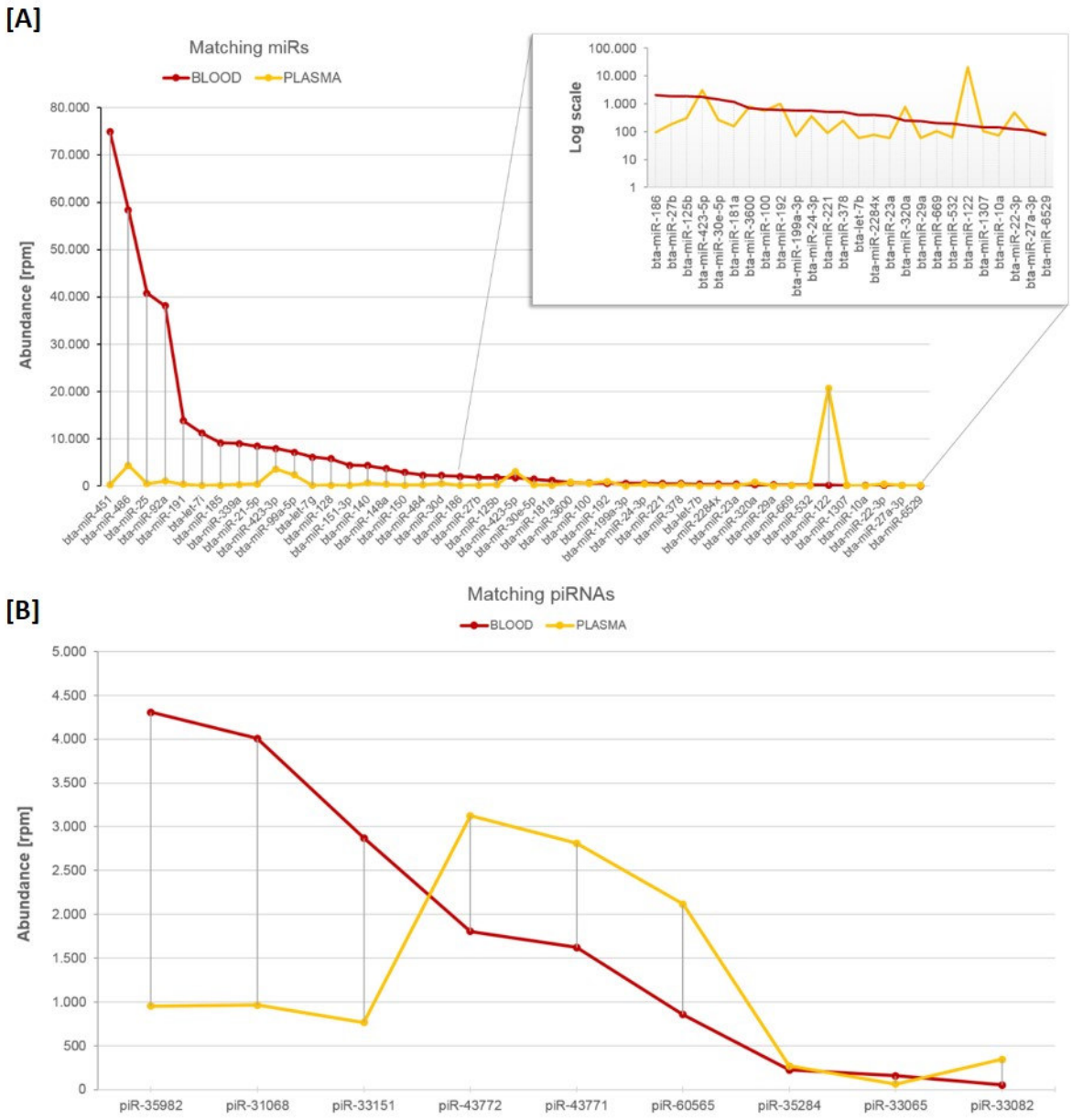


Figure 13: [A] Matching miRNA pairs highlighted the differential distribution of miRNAs in whole blood (red) and in plasma (yellow). [B] Matching piRNA pairs revealed a bidirectional expression effect [41].

This basic footprinting of the miRNome and piRNome provided primary information concerning the composition of bovine whole blood and plasma and described the presence of piRNAs. In clinical research, circulating miRNAs are already highly investigated as 'circulating biomarkers' for multiple applications [61]. This highlights the future potential usability of piRNAs in this context. Hence, smexRNAs could also be beneficial for investigating targets for veterinary medicine studies.

3.6. The potential of smexRNAs in veterinary diagnostics: identifying biomarker signatures by multivariate data analysis

3.6.1. Differential abundance of smexRNAs in AnaBull animals

Total sequenced reads (9.378.527 ± 3.289.748 reads) were quality checked, size pre-filtered and successively mapped as it was described in Spornraft et al. [47] (Appendix I). In the AnaBull study, 783 bta miRNAs were successfully mapped with 54 hits exceeding the noise background. Regarding piRNAs, mapping resulted in 126 hsa-piRNA hits including 15 piRNAs higher than the noise cut-off. Processing readcount tables to detect unique miRNAs revealed a specific expression of four miRNAs in the P+EB group (miR-127-3p, 151a-3p, 215-5p, 3168) and six in the CLEN group (miR-127-3p, 151a-3p, 186-5p, 193b-5p, 106b-3p, 93-5p). No statistically significant expression changes could be calculated for the aligned smexRNAs. The proportion of miRNAs and piRNAs on the total sequenced library size were calculated for all 21 AnaBull animals [46] (Appendix IV) (Figure 11, [B]). The resulting abundances are comparable with those from non-treated healthy bovines (Figure 11, [A]) underlining reproducibility of sequencing and stability of a distinct smexRNA level in plasma.

3.6.2. The largest proportion of the data sets was reflected by the top 10 small RNAs

Sorting according to rpm-normalized readcounts, expressed miRNAs and piRNAs were listed (Table 2). Summarizing all findings, it can be stated that the main percentage of the smexRNA signature is composed of the top 10 small RNA candidates (Table 2) and this pattern is robustly expressed without statistically significant expression ratios [46] (Appendix IV). Therefore, the investigated smexRNAs in animals that were treated with anabolic substances were expressed more stable than assumed. It was however already proven that the treatments were effective, as animals showed significant differences in the weight gain development and a potential gene expression biomarker signature was identified on the mRNA level in liver samples in the course of the AnaBull study [14] (Appendix III).

Table 2: Comparison of the master miRNAs and piRNAs in the AnaBull study. Checkmarks indicate the affiliation to the 'top 10 small RNA list' and superscript numbers give ranking information. Pie charts highlight the percentage on the total annotated reads [46].

miRNA	CON	P+EB	CLEN	piRNA	CON	P+EB	CLEN
bta-miR-122	✓ 1	✓ 1	✓ 1	piR-33151	✓ 1	✓ 3	✓ 4
bta-miR-423-3p	✓ 2	✓ 2	✓ 2	piR-33082	✓ 2		✓ 8
bta-miR-423-5p	✓ 3	✓ 3	✓ 3	piR-43772	✓ 3	✓ 1	✓ 1
bta-miR-486	✓ 4	✓ 6	✓ 5	piR-43771	✓ 4	✓ 2	✓ 2
bta-miR-99a-5p	✓ 5	✓ 5	✓ 6	piR-43786	✓ 5	✓ 4	✓ 5
bta-miR-320a	✓ 6	✓ 4	✓ 4	piR-60565	✓ 6	✓ 5	✓ 3
bta-miR-92a	✓ 7	✓ 7	✓ 9	piR-35982	✓ 7	✓ 7	✓ 7
bta-miR-3600	✓ 8		✓ 8	piR-31068	✓ 8	✓ 6	✓ 6
bta-miR-193a-5p	✓ 9	✓ 8	✓ 7	piR-31038	✓ 9		✓ 9
bta-miR-192	✓ 10	✓ 9	✓ 10	piR-35284	✓ 10		✓ 10
bta-miR-140		✓ 10		piR-36256		✓ 8	
				piR-36255		✓ 9	
				piR-36243		✓ 10	
% Top 10	84.2	76.0	83.8		91.3	82.2	92.3
							

3.6.3. Comparative data management strategies for discriminative analysis to identify veterinary drug misuse

The discriminative power of OPLS-DA was investigated to identify the ideal performance settings. As displayed in Figure 14, the discrimination of the non-treated animals from the treated ones was superior, if data sets including all sequenced reads (all reads) were used for the analysis (Figure 14, [A-C]). OPLS-DA utilizing merged data sets could not perfectly separate the CLEN-treated animals from the CON group, while acceptance of all reads resulted in a model that allowed separation of P+EB classes with high quality (Figure 14, [A]). Using only the full miRNA data set delivered the best results concerning discrimination, goodness of fit and prediction (Figure 14, [B]). OPLS-DA including exclusively piRNA reads could not cluster the treated animals nor meet quality standards (Figure 14, [C]). Detailed information on the DA and quality assessment can be found in Spornraft et al. [46] (Appendix IV).

MiRNA readcounts and hence the available coverage for statistical analyses surpass piRNA abundance (Figure 11 and Tables S1, S2). Therefore, OPLS-DA could operate with additional data supporting improved prediction ability and discrimination for miRNAs. Concerning the AnaBull trial, it

became obvious that OPLS-DA is best suited for full datasets of circulating miRNAs. In the search for biomarkers to trace veterinary drug abuse, multivariate data analysis tools managed to detect treatment-dependent differences at the miRNA level.

As described before, the top 10 miRNA pattern was robustly expressed and not subject of anabolic treatment impact. Therefore, the multiplicity of low-abundant miRNAs was rather responsible for the generation of well performing OPLS-DA than the master miRNAs. To surpass this limit, an increased sequencing depth could help to promote deeper sequencing of minor expressed smexRNAs. The AnaBull study gave a first indication that circulating miRNAs could be favorable biomarker candidates to identify anabolic misuse, if laboratory conditions are further enhanced. Furthermore, since blood is in permanent contact with the entire organism, the smexRNA profile is affected by constant substantial systemic alterations. As anabolic treatment had an impact, maybe there was an effect on the smexRNA signature, but could not be captured at the sampling time point due to rapid circulatory turnover.

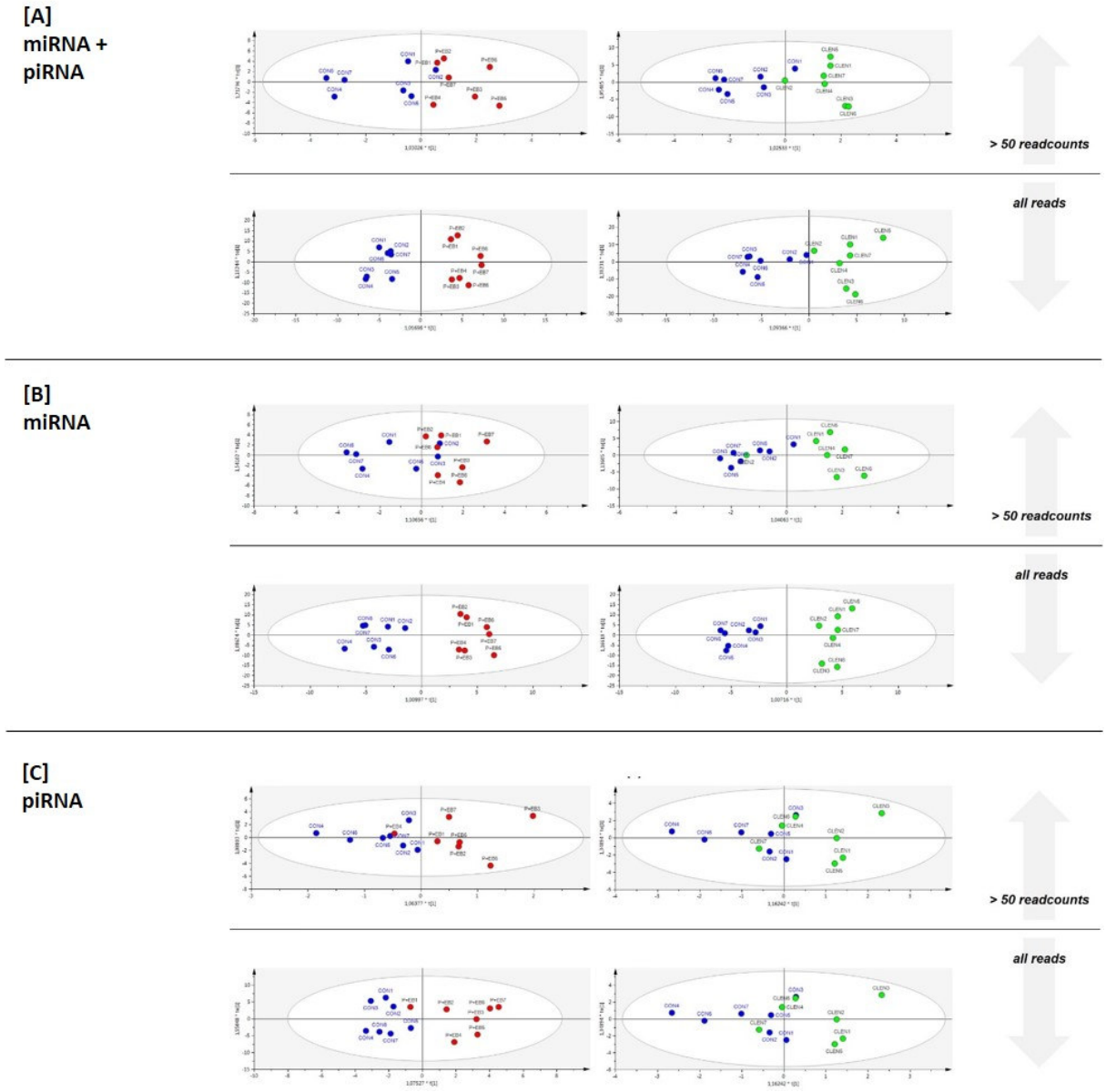


Figure 14: OPLS-DA using combined miRNA and piRNA data sets [A], miRNAs [B] and piRNAs [C] alone. CON animals: blue; P+EB: red, CLEN: green [46].

3.7. RT-qPCR measurement of master miRNAs identified by miRNome profiling: completing the time profile over the AnaBull treatment period

3.7.1. Determination of the optimal input volume to avoid PCR inhibition

To achieve the best RT performance without an inhibiting effect on downstream qPCR, the optimal input volume was determined prior to the profiling experiment. By increasing the RNA input in the RT reaction, the Cq values are decreasing until inhibition sets in, which is accompanied by an increase in Cq values. This effect was observed at the threshold input volume of 12µl, which was defined as the best practice input volume where lowest Cq values were reached with the primer assays UniSp2, UniSp4, UniSp5 and miR-122 (Figure 15). Therefore, the profiling of miRNAs with RT-qPCR was performed with an initial transcription volume of 12 µl of isolated RNA in the RT reaction.

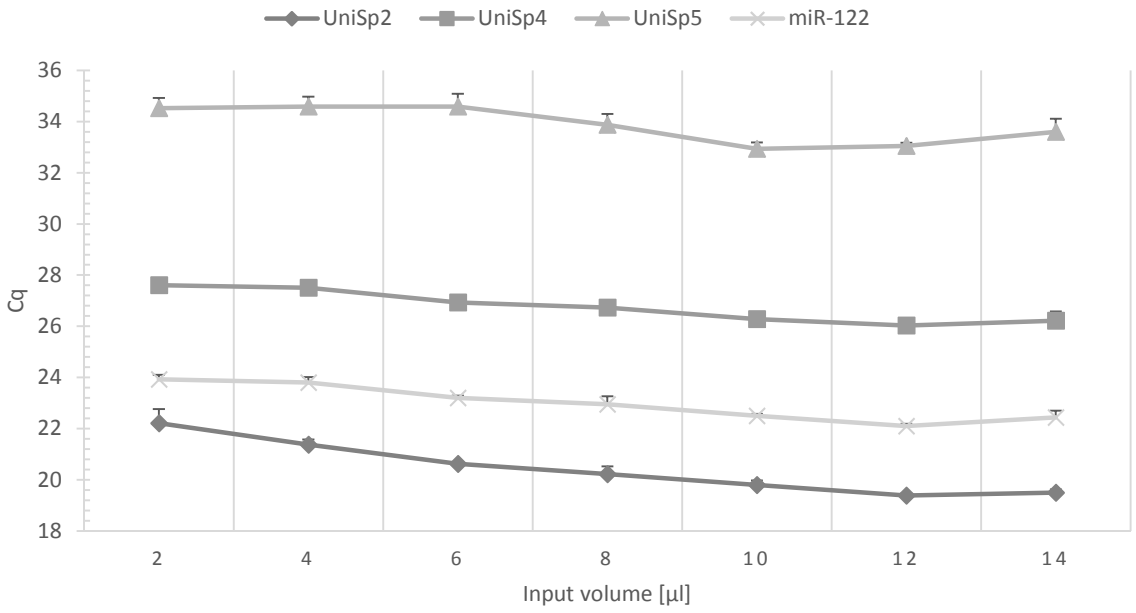


Figure 15: Optimization of the input volume in cDNA synthesis reaction. Different input volumes in the range from 2-14 µl of extracted RNA were tested for the optimal input in cDNA synthesis prior qPCR profiling assays. Lowest Cq values were achieved for the input volume of 12 µl.

3.7.2. Identification of the best normalization strategy for AnaBull plasma and investigating the time profile of the master miRNAs

Small RNA-Seq resulted in a global perspective on the bovine small RNA spectrum and identified major expressed miRNAs. These miRNAs were validated via RT-qPCR. Seven primer assays could be successfully designed, while three assays delivered not suitable raw data due to concentration deficiency. Plasma as cell-free material is a matrix with minimum RNA concentrations. Measuring the correct RNA concentrations is very difficult and increased RNA losses during isolation processes have to be considered. RT-qPCR measurement is a sensitive approach for the quantification of miRNA expression ratios. However, to reduce technical variation and determine accurate biological changes, an appropriate normalization strategy is inevitable. As described before, several reference miRNAs are recommended in literature. Due to marginal expression, these miRNAs could not be used in our data sets. The global mean, as the mean value encompassing all measured assays per sample, was confirmed in this individual case as optimum normalization strategy due to lowest standard deviations (SD) (Figure 16). This analysis method has also proven to outperform other current normalization strategies in literature [62], [63] and is in accordance with our NormFinder search for the optimum reference (Figure 16).

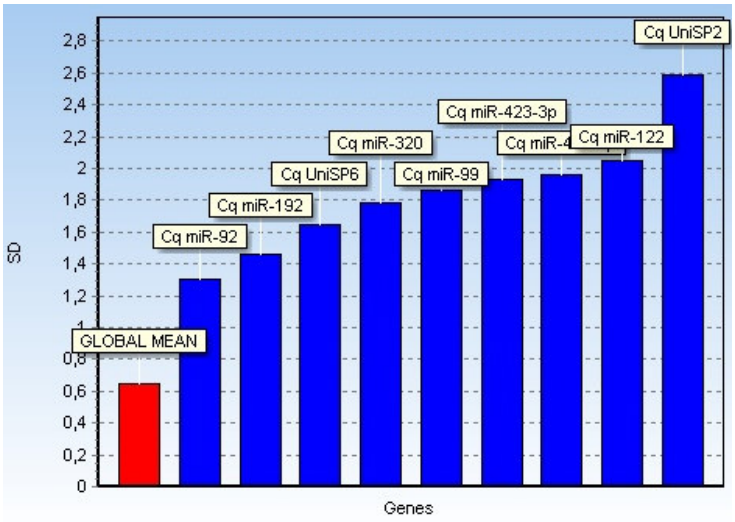
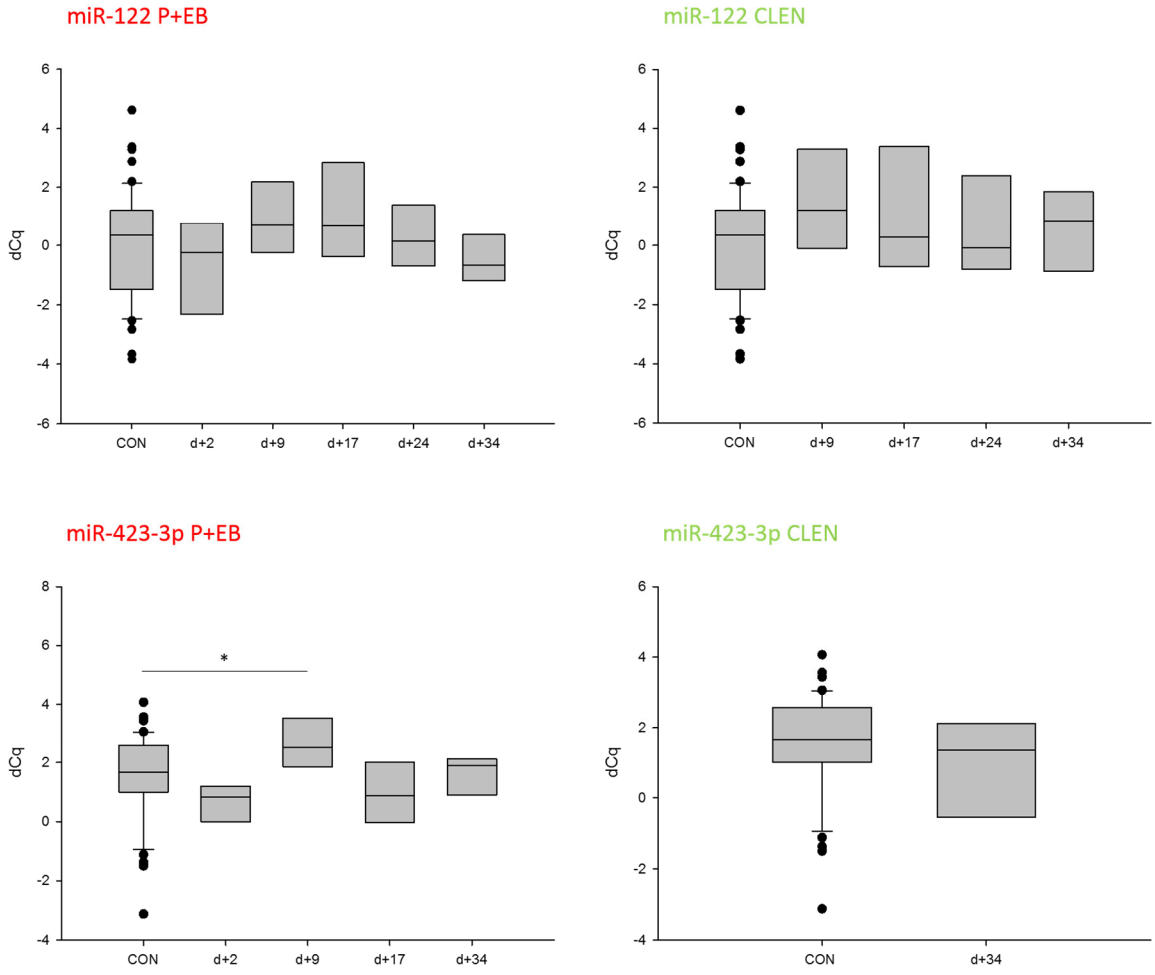


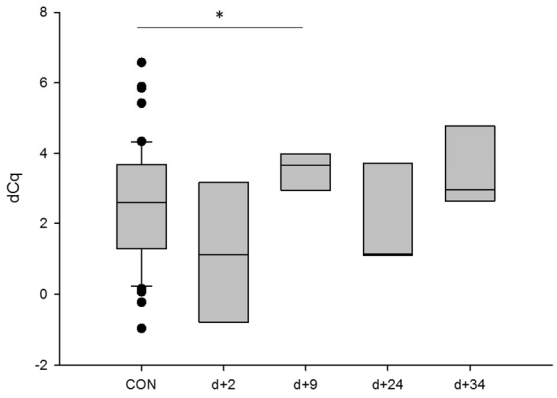
Figure 16: The NormFinder algorithm [51] was applied for finding the optimum reference gene(s). The best normalization strategy was using the global mean value (highlighted in red) as reference gene index (RGI) in the ΔCq analysis method [50]. UniSp2 and UniSp6 were spike-in control assays (RNA Spike-In Kit, Exiqon).

To complete the small RNA-Seq picture of the AnaBull plasma samples from d+17, the other time points (Figure 5) were individually measured with RT-qPCR (Figure 17). Screening the most abundant miR-122 resulted in an unvarying profile. No significant expression changes were detected between the ΔCq values from controls (n=56) and treated samples (P+EB: d+2 p=0.505, d+9 p=0.186, d+17 p=0.061, d+24 p=0.873, d+34 p=0.774; CLEN: d+9 p=0.058, d+17 p=0.164, d+24 p=0.497, d+34 p=0.216). In the CLEN group, the profiling of miR-423-3p and 5p was incomplete (CLEN: 3p: d+34

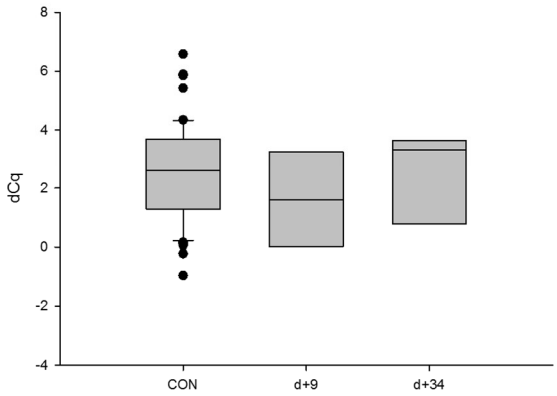
p=0.572; 5p: d+9 p=0.154, d+34 p=0.998). For the P+EB samples, the measurements revealed a significantly higher expression level of miR-423-3p and miR-423-5p at nine days after implantation of steroidal compounds (P+EB: 3p: d+2 p=0.101, d+9 p=0.041*, d+17 p=0.203, d+34 p=0.990; 5p: d+2 p=0.130, d+9 p=0.038*, d+24 p=0.347, d+34 p=0.195). MiR-192 was constantly expressed in both treatment groups (P+EB: d+2 p=0.751, d+9 p=0.477, d+17 p=0.425, d+24 p=388, d+34 p=0.638; CLEN: d+2 p=0.477, d+9 p=0.113, d+17 p=0.519, d+24 p=0.904, d+34 p=0.309). Although no significant expression changes could be observed for the P+EB group (P+EB: d+2 p=0.061, d+9 p=0.751, d+17 p=0.735, d+24 p=0.289, d+34 p=0.768), miR-320 was significantly up-regulated at d+17 and d+24 in the CLEN group (CLEN: d+2 p=0.562, d+9 p=0.251, d+17 p=0.031*, d+24 p=0.033*, d+34 p=0.922). MiR-99 was significantly down-regulated in both groups rather at the beginning of the trial (P+EB: d+2 p=0.030*, d+9 p=0.945, d+17 p=0.930, d+34 p=0.109; CLEN: d+9 p=0.037*, d+34 p=0.238). miR-92 showed a significant up-regulation at d+24 in the P+EB group (P+EB: d+2 p=0.269, d+9 p=0.412, d+17 p=0.451, d+24 p=<0.001*, d+34 p=0.109) and at d+2, d+17 and d+24 in the CLEN group (CLEN: d+2 p=0.018*, d+9 p=0.751, d+17 p=0.033*, d+24 p=<0.001*, d+34 p=0.853).



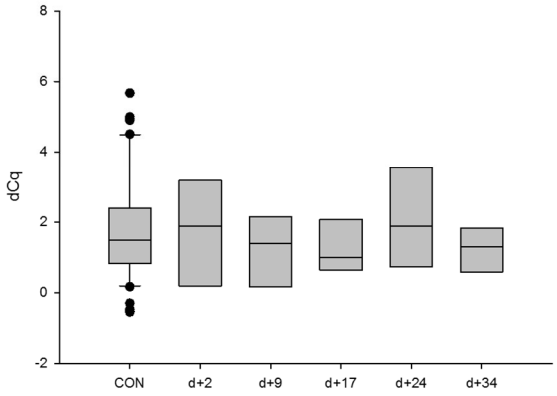
miR-423-5p P+EB



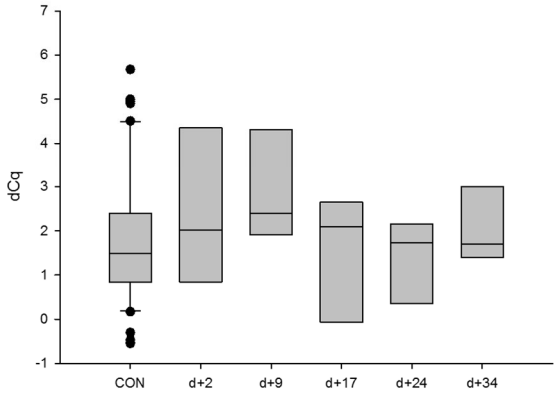
miR-423-5p CLEN



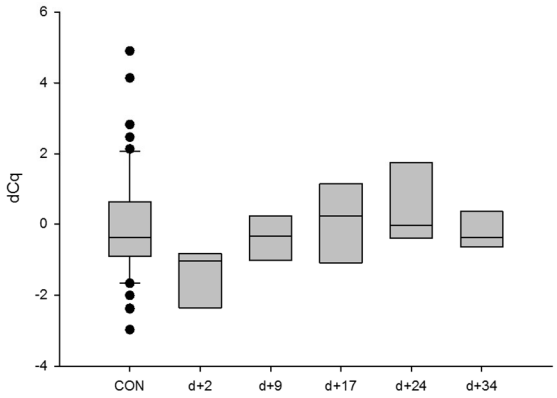
miR-192 P+EB



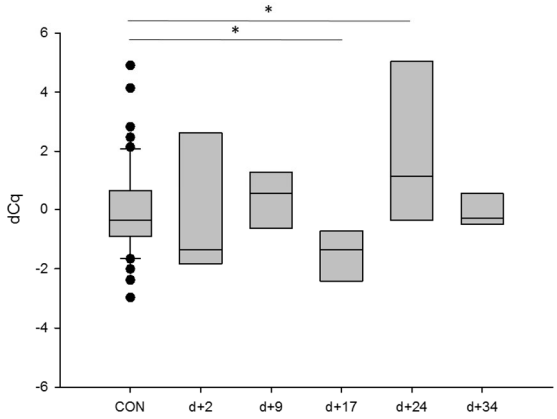
miR-192 CLEN



miR-320a P+EB



miR-320a CLEN



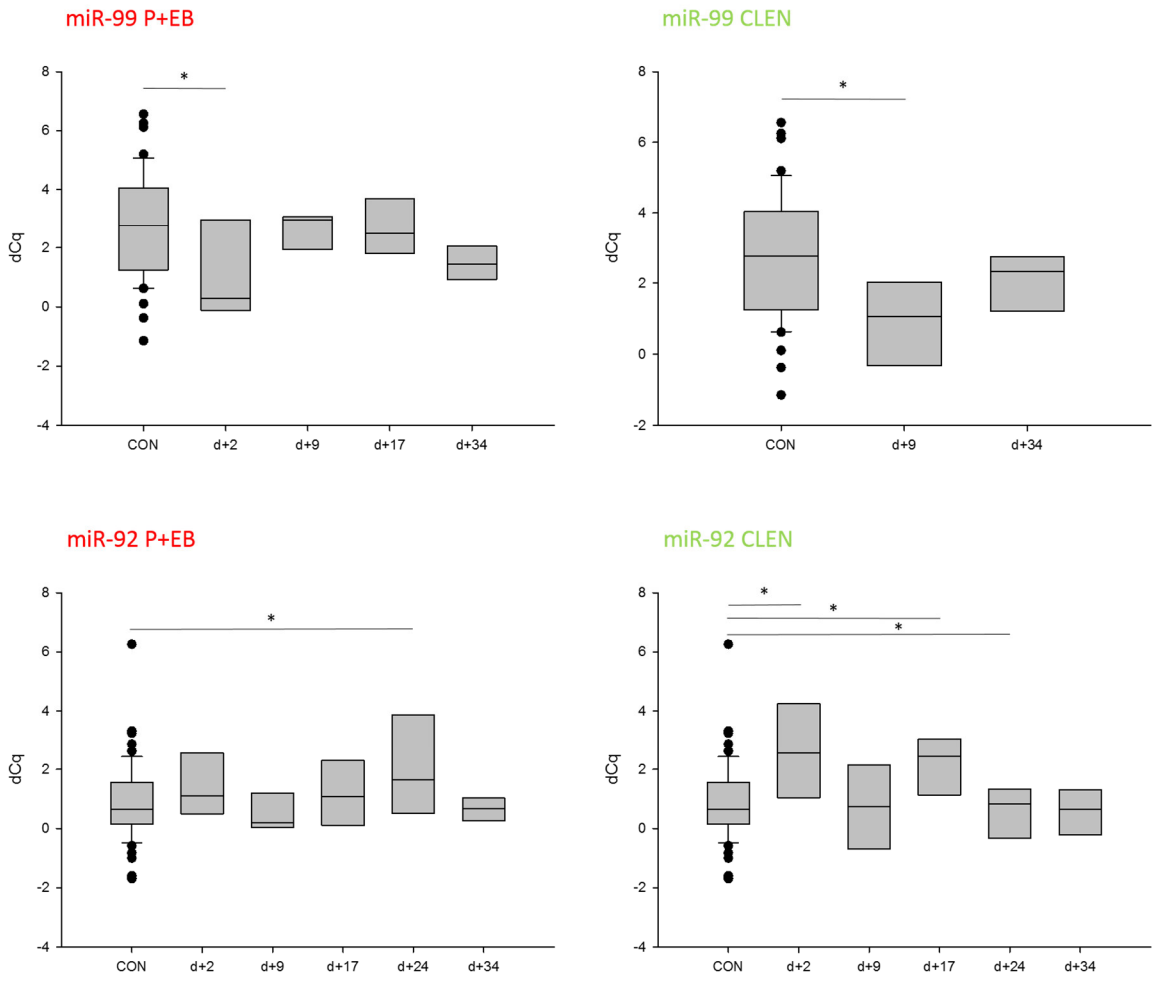


Figure 17: RT-qPCR profiling of AnaBull plasma samples at time-points d+2, d+9, d+17, d+24 and d+34 (n=7 per treatment group). Primer assays were selected on the basis of NGS data. Bars that were marked with an asterisk (*) showed a significant expression change between the controls and the treatment group with $p < 0.05$.

In the P+EB group, the expression of four miRNAs was significantly regulated, three were up-regulated, one was down-regulated. In the CLEN group, two miRNAs were up-regulated and one down. No significant changes could be calculated using NGS readcounts, but RT-qPCR profiling revealed indeed expressional differences between the CON group and the treated AnaBull groups (Figure 17). However, as reported before, small RNA-Seq results demonstrated that the sum of the data is suited to discriminate treated from non-treated animals (Figure 14). The use of a routine small RNA-Seq analysis for the detection of illegal application of veterinary drugs clearly raises the questions, if this is an affordable and convenient practice. Even if miRNA based-models could separate CON animals from treated ones (Figure 14, [B]), the exact quantification and a specific verification of a confident biomarker signature in bovine plasma is limited. 3/10 primer assays failed to measure Cq-values due to marginal concentrations of certain miRNAs. PiRNA abundance is even reduced by a power of ten in this matrix (Appendix V: Table S1, S2) and an individual validation with RT-qPCR is impossible, since a 10-fold volume was used in NGS compared to RT-qPCR validation. Moreover, companies just now start to provide piRNA primer designs. At the experiment´s time point (October 2014), only Exiqon (Denmark) sold customized piRNA primers that were, however, not certified for bta. Therefore, experimental conditions for the RT-qPCR quantification of piRNAs need to be further improved.

Yet, basic feasibility and significant expressional changes were demonstrated for several miRNAs at diverse time-points of the AnaBull study (Figure 17). However, no significant miRNA combination with a parallel altered expression at a same day, which could therefore be used as a stable biomarker pattern, was found. Furthermore, RT-qPCR experiments lack high dynamic ranges as well as sensitivity and reproducibility for the measurements of transcripts at very low concentrations [64]. The prerequisite of a sufficient transcript molecule density was only given in seven cases. MiRNAs, however, that were responsible for the discriminative effects in OPLS-DA did not fulfill this precondition due to a very low concentration of regulated smexRNAs in bovine plasma.

4. Conclusions

The total RNA yield that is extracted using commercially available isolation kits for smexRNAs from body fluids is not sufficient to perform NGS. The presented extraction method was successfully optimized for bovine plasma to enable Next-Generation small RNA Sequencing with high-quality read output. Running the specially designed bioinformatics pipeline guaranteed optimal data acquisition in terms of stringent trimming, gradual size filtering, mapping and annotation. Therefore, the percentage of smexRNAs could be described for the first time in bovine plasma.

Based on the achieved technological possibilities, the bovine miRNome and piRNome of whole blood and plasma were sequenced. Status quo signatures of the composition of smexRNAs in both bio fluids could be presented which established the basis for comparative analyses. Several miRNAs and piRNAs in plasma exceeded their concentration in whole blood. This strengthened evidence that the smexRNA signature in plasma was not only affected by cellular blood components and certain smexRNAs could originate from other sources in the body, e.g. tissues.

SmexRNAs qualified as potential candidates in the identification of transcriptomic biomarkers with the ability to uncover forbidden anabolic drug abuse in veterinary diagnostics. Using the global small RNA-Seq data, multivariate projection methodologies like OPLS-DA have proven the best potential to generate discriminative miRNA models favoring the differentiation of treatment groups. PiRNAs were expressed at lower copy numbers and data sets did not ensure statistical robustness and significance. NGS data revealed that those smexRNAs that contributed the major part to the global signature were constantly expressed and therefore not altered by anabolic stimulation.

Due to the extremely low concentration of smexRNAs in plasma, the quantitative validation of the top 10 miRNAs via RT-qPCR reached the technological limits. NGS data revealed highly expressed smexRNAs that could be subsequently measured with RT-qPCR, but the discriminative power to separate treatment and non-treated groups lied in the sum of all sequenced observations. Single RT-qPCR measurements allowed a specific look at miRNAs, but this validation could not identify a biomarker pattern.

5. Perspectives

Although the total of sequenced plasma miRNAs discriminated treated from non-treated animals, steroid hormones and clenbuterol have other direct targets in the organism than blood. Therefore, it seems promising to sequence the small RNA transcriptome of explicit target tissues from the AnaBull study, e.g. heart, lung and liver. On the mRNA level, the description of a validated biomarker pattern in these tissues was successfully accomplished by our group (paper submitted, [65]). As a consequence, the profiling of miRNAs, as direct regulators of mRNAs, is a promising approach to find miRNA biomarkers in tissues. Additionally, prediction tools for mRNA-miRNA target interaction could identify suitable miRNAs that regulate candidate mRNA translation.

Of course, the presented data is a snapshot of the smexRNA profiles in calves under a particular treatment regime in a limited time frame and at a specific age, gender and breed. If a potential smexRNA or mRNA biomarker pattern is found as part of the AnaBull study, this trial needs to be further elaborated. Only this strategy could identify a long-term and reliable pattern under varying circumstances (older animals, other gender, breed or anabolic substances).

Due to the methodical limits and the lack of commonly accepted normalization solutions in the RT-qPCR-validation of single miRNAs, the latest technology in nucleic acid detection and quantification could be used for most precise quantification. Applying droplet digital PCR (ddPCR) could detect and absolutely quantify even the low abundant circulating nucleic acids [66], [67].

Still, the striking advantages of liquid biopsies remain. Therefore, other bio fluids besides plasma like urine, milk or saliva could be possible matrices for investigating smexRNA signatures of bovines under anabolic influence.

6. References

1. Serratosa J, Blass A, Rigau B, Mongrell B, Rigau T et al. (2006) Residues from veterinary medicinal products, growth promoters and performance enhancers in food-producing animals: a European Union perspective. Revue scientifique et technique (International Office of Epizootics) 25 (2): 637–653.
2. Stephany RW (2010) Hormonal growth promoting agents in food producing animals. Handbook of experimental pharmacology (195): 355–367.
3. Thevis M, Geyer L, Geyer H, Guddat S, Dvorak J et al. (2013) Adverse analytical findings with clenbuterol among U-17 soccer players attributed to food contamination issues. Drug Test. Analysis 5 (5): 372–376.
4. Guddat S, Fußhöller G, Geyer H, Thomas A, Braun H et al. (2012) Clenbuterol - regional food contamination a possible source for inadvertent doping in sports. Drug Test. Analysis 4 (6): 534–538.
5. Riedmaier I, Becker C, Pfaffl MW, Meyer, Heinrich H D (2009) The use of omic technologies for biomarker development to trace functions of anabolic agents. Journal of chromatography. A 1216 (46): 8192–8199.
6. Johnson G, Nour AA, Nolan T, Huggett J, Bustin S (2014) Minimum information necessary for quantitative real-time PCR experiments. Methods Mol. Biol. 1160: 5–17.
7. Fragkaki AG, Georgakopoulos C, Sterk S, Nielen MWF (2013) Sports doping: Emerging designer and therapeutic β2-agonists. Clin. Chim. Acta.
8. Picard D, Kumar V, Chambon P, Yamamoto KR (1990) Signal transduction by steroid hormones: nuclear localization is differentially regulated in estrogen and glucocorticoid receptors. Molecular Biology of the Cell 1 (3): 291–299.
9. Ribeiro RC, Kushner PJ, Baxter JD (1995) The nuclear hormone receptor gene superfamily. Annual review of medicine 46: 443–453.
10. Riedmaier I, Pfaffl MW (2013) Transcriptional biomarkers--high throughput screening, quantitative verification, and bioinformatical validation methods. Methods 59 (1): 3–9.
11. Pfaffl MW (2013) Transcriptional biomarkers. Methods 59 (1): 1–2.
12. Riedmaier I, Pfaffl MW, Meyer, Heinrich H D (2011) The analysis of the transcriptome as a new approach for biomarker development to trace the abuse of anabolic steroid hormones. Drug Test. Analysis 3 (10): 676–681.
13. Riedmaier I, Pfaffl MW, Meyer HHD (2012) The physiological way: monitoring RNA expression changes as new approach to combat illegal growth promoter application. Drug Test. Analysis 4 Suppl 1: 70–74.
14. Riedmaier I, Spornraft M, Pfaffl MW (2014) Identification of a potential gene expression biomarker signature in bovine liver to detect the abuse of growth promoters. Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment 31 (4): 641–649.
15. Nebbia C, Urbani A, Carletti M, Gardini G, Balbo A et al. (2011) Novel strategies for tracing the exposure of meat cattle to illegal growth-promoters. Veterinary journal (London, England : 1997) 189 (1): 34–42.
16. Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75 (5): 843–854.
17. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19 (1): 92–105.
18. van Rooij E (2011) The art of microRNA research. Circ. Res 108 (2): 219–234.
19. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nature reviews. Molecular cell biology 10 (2): 126–139.
20. Bartels CL, Tsongalis GJ (2009) MicroRNAs: novel biomarkers for human cancer. Clinical Chemistry 55 (4): 623–631.
21. Doench JG, Sharp PA (2004) Specificity of microRNA target selection in translational repression. Genes Dev 18 (5): 504–511.

22. Chim, Stephen S C, Shing, Tristan K F, Hung, Emily C W, Leung T, Lau T et al. (2008) Detection and characterization of placental microRNAs in maternal plasma. Clinical Chemistry 54 (3): 482–490.
23. Weber JA, Baxter DH, Zhang S, Huang DY, How Huang K et al. (2010) The MicroRNA Spectrum in 12 Body Fluids. Clinical Chemistry 56 (11): 1733–1741.
24. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK et al. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. U.S.A 105 (30): 10513–10518.
25. Tang D, Shen Y, Wang M, Yang R, Wang Z et al. (2013) Identification of plasma microRNAs as novel noninvasive biomarkers for early detection of lung cancer. Eur. J. Cancer Prev 22 (6): 540–548.
26. Xu J, Zhao J, Evan G, Xiao C, Cheng Y et al. (2012) Circulating microRNAs: novel biomarkers for cardiovascular diseases. J Mol Med 90 (8): 865–875.
27. Li C, Pei F, Zhu X, Duan DD, Zeng C (2012) Circulating microRNAs as novel and sensitive biomarkers of acute myocardial Infarction. Clin. Biochem 45 (10-11): 727–732.
28. Jin X, Wu N, Wang L, Li J (2013) Circulating microRNAs: a novel class of potential biomarkers for diagnosing and prognosing central nervous system diseases. Cell. Mol. Neurobiol 33 (5): 601–613.
29. Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42 (D1): D68.
30. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34 (Database issue): D140-4.
31. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128 (6): 1089–1103.
32. Cox DN, Chao A, Baker J, Chang L, Qiao D et al. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 12 (23): 3715–3727.
33. Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442 (7099): 199–202.
34. Grivna ST, Beyret E, Wang Z, Lin H (2006) A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 20 (13): 1709–1714.
35. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129 (1): 69–82.
36. Kalmykova AI, Klenov MS, Gvozdev VA (2005) Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic Acids Res 33 (6): 2052–2059.
37. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. Science (New York, N.Y.) 322 (5906): 1387–1392.
38. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A et al. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 22 (7): 908–917.
39. Yin H, Lin H (2007) An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. Nature 450 (7167): 304–308.
40. Roovers EF, Rosenkranz D, Mahdipour M, Han C, He N et al. (2015) Piwi Proteins and piRNAs in Mammalian Oocytes and Early Embryos. Cell reports 10 (12): 2069–2082.
41. Spornraft M, Kirchner B, Pfaffl MW, Riedmaier I (2015) Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing. Biotechnology letters.
42. Chu H, Hui G, Yuan L, Shi D, Wang Y et al. (2015) Identification of novel piRNAs in bladder cancer. Cancer Lett 356 (2 Pt B): 561–567.
43. Bahn JH, Zhang Q, Li F, Chan T, Lin X et al. (2015) The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. Clinical Chemistry 61 (1): 221–230.
44. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ et al. (2013) GenBank. Nucleic Acids Res 41 (Database issue): D36-42.

45. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E et al. (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J. Chemometrics 20 (8-10): 341–351.
46. Spornraft M, Kirchner B, Pfaffl MW, Riedmaier I (2015) The Potential Of Circulating Extracellular Small RNAs (smexRNA) In Veterinary Diagnostics – Identifying Biomarker Signatures by Multivariate Data Analysis.
47. Spornraft M, Kirchner B, Haase B, Benes V, Pfaffl MW et al. (2014) Optimization of Extraction of Circulating RNAs from Plasma - Enabling Small RNA Sequencing. PLoS ONE 9 (9): e107259.
48. Riedmaier I, Tichopad A, Reiter M, Pfaffl MW, Meyer, Heinrich H D (2009) Identification of potential gene expression biomarkers for the surveillance of anabolic agents in bovine blood cells. Anal. Chim. Acta 638 (1): 106–113.
49. Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H (2006) Quantitative real-time RT-PCR data analysis: current concepts and the novel \"gene expression's CT difference\" formula. Journal of molecular medicine (Berlin, Germany) 84 (11): 901–910.
50. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29 (9): e45.
51. Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer research 64 (15): 5245–5250.
52. Turchinovich A, Weiz L, Langheinz A, Burwinkel B (2011) Characterization of extracellular circulating microRNA. Nucleic Acids Res 39 (16): 7223–7233.
53. Long J, Chen H (2009) Identification and characteristics of cattle microRNAs by homology searching and small RNA cloning. Biochemical genetics 47 (5-6): 329–343.
54. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W et al. (2002) Identification of tissue-specific microRNAs from mouse. Curr. Biol. 12 (9): 735–739.
55. Duttagupta R, Jiang R, Gollub J, Getts RC, Jones KW (2011) Impact of cellular miRNAs on circulating miRNA biomarker signatures. PLoS ONE 6 (6): e20769.
56. Hsu S, Wang B, Kota J, Yu J, Costinean S et al. (2012) Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. The Journal of clinical investigation 122 (8): 2871–2883.
57. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P et al. (2013) Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods 59 (1): S1-6.
58. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ et al. (2012) Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila) 5 (3): 492–497.
59. Rasmussen KD, Simmini S, Abreu-Goodger C, Bartonicek N, Di Giacomo M et al. (2010) The miR-144/451 locus is required for erythroid homeostasis. The Journal of experimental medicine 207 (7): 1351–1358.
60. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP et al. (2011) Haemolysis during sample preparation alters microRNA content of plasma. PLoS ONE 6 (9): e24145.
61. Hydbring P, Badalian-Very G (2013) Clinical applications of microRNAs. F1000Research 2: 136.
62. Mestdagh P, van Vlierberghe P, Weer A de, Muth D, Westermann F et al. (2009) A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol. 10 (6): R64.
63. D'haene B, Mestdagh P, Hellemans J, Vandesompele J (2012) miRNA expression profiling: from reference genes to global mean normalization. Methods Mol. Biol. 822: 261–272.
64. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics 10 (1): 57–63.
65. Riedmaier I, Spornraft M, Börger N, Schiemann I, Pfaffl MW (2015) Detection of a gene expression biomarker signature for the abuse of anabolic substances in bovine heart and lung. Journal of Nutritional Health and Food Science.

66. Miotto E, Saccenti E, Lupini L, Callegari E, Negrini M et al. (2014) Quantification of circulating miRNAs by droplet digital PCR: comparison of EvaGreen- and TaqMan-based chemistries. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 23 (12): 2638–2642.
67. Ferracin M, Lupini L, Salamon I, Saccenti E, Zanzi MV et al. (2015) Absolute quantification of cell-free microRNAs in cancer patients. Oncotarget.

Acknowledgements

I would like to thank Prof. Dr. Michael W. Pfaffl for giving me the possibility to do my PhD thesis at the Chair of Physiology. I would like to thank him for all his scientific input, the short official channels and for his sovereign management of the Physiology Institute in times of change.

Sincerely I want to thank Dr. Irmgard Riedmaier for her excellent supervision and support. Thank you so much for the always open and understanding ears, the scientific inspiration and guidance and also for the wonderful atmosphere in the working group.

Additionally, I want to thank Jürgen Wickentrup (head of the veal farm in Wadersloh, Germany) and veterinarian Marina Doum (from the veterinary practice Stefan Lüllmann) for their great help during the AnaBull study and KDK, SGS and Denkavit for financial support.

Special thanks go to Dr. Vladimir Benes and Bettina Haase for their invitation to EMBL and their experienced support in library preparation and NGS. Additionally, I want to thank Dr. Leen van Ginkel and Dr. Marco Blokland for their invitation to RIKILT - Institute for Food Safety (Netherlands) and the support in statistical data evaluation.

I want to thank my colleague Benedikt Kirchner for his cooperation in bioinformatics and data analysis and Enrique Viturro for creative help with figure and table design.

I would like to thank Dr. Christiane Becker from the Chair of Animal Nutrition who supported my work as my TUM Graduate School mentor.

Many thanks go to all my colleagues from the Chair of Physiology. Helping hands and a friendly working atmosphere was always guaranteed.

Especially I want to thank my sister Julia, my parents Robert and Waltraud and my grandmother Uli that they kept loads away so I could concentrate on my thesis. Thank you for your caring support.

Last but not least, special thanks go to my friends. They shared good and bad times and spending time together was always great.

List of scientific communications

Research papers

Melanie Spornraft, Benedikt Kirchner, Michael W. Pfaffl, Irmgard Riedmaier

The Potential Of Circulating Extracellular Small RNAs (smexRNA) In Veterinary Diagnostics – Identifying Biomarker Signatures by Multivariate Data Analysis

Biomolecular Detection and Quantification, Volume 5, 19. September 2015, Special Issue: Advanced Molecular Diagnostics for Biomarker Discovery – Part I, DOI: 10.1016/j.bdq.2015.08.001

Melanie Spornraft, Benedikt Kirchner, Michael W. Pfaffl, Irmgard Riedmaier

Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing

Biotechnology Letters, 21. February 2015, DOI: 10.1007/s10529-015-1788-2

Melanie Spornraft, Benedikt Kirchner, Bettina Haase, Vladimir Benes, Michael W. Pfaffl, Irmgard Riedmaier

Optimization of Extraction of Circulating RNAs from Plasma – Enabling small RNA Sequencing

PLoS One eCollection, 17. September 2014, DOI: 10.1371/journal.pone.0107259

Irmgard Riedmaier, Melanie Spornraft, Michael W. Pfaffl

Identification of a potential gene expression biomarker signature in bovine liver to detect the abuse of growth promoters

Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment, 3. April 2014, DOI: 10.1080/19440049.2014.886341

Irmgard Riedmaier, Melanie Spornraft, Nora Börger, Isabel Schiemann, Michael W. Pfaffl

Development of an uniform biomarker signature in bovine heart and lung to detect the abuse of different anabolic growth promoters

Journal of Nutritional Health and Food Science, Special Issue Food Safety & Hygiene, submitted

Poster presentations

Melanie Spornraft, Benedikt Kirchner, Michael W. Pfaffl, Irmgard Riedmaier-Sprenzel

Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing

7th international qPCR & NGS Event – Advanced Molecular Diagnostics for Biomarker Discovery, 23.-27. March 2015, Freising-Weihenstephan, Germany

Melanie Spornraft, Benedikt Kirchner, Bettina Haase, Vladimir Benes, Michael W. Pfaffl, Irmgard Riedmaier-Sprenzel

Circulating Small RNAs As Biomarker Candidates For The Detection Of Anabolic Misuse In Cattle

7th International Symposium on Hormone and Veterinary Drug Residue Analysis, 2.-5. June 2014, Ghent, Belgium

Melanie Spornraft, Bettina Haase , Benedikt Kirchner, Vladimir Benes , Michael W. Pfaffl, Irmgard Riedmaier-Sprenzel

Ready to Go for Next-Generation Sequencing: Extraction of Circulating microRNAs from Bovine Plasma

6th international qPCR & NGS Event - Next Generation Thinking in Molecular Diagnostics, 18.-22. March 2013, Freising-Weihenstephan, Germany

Melanie Spornraft, Irmgard Riedmaier

The Physiological Way: A new Approach for Biomarker Detection against Anabolic Misuse

Heinrich HD Meyer Symposium, 19. Oktober 2012, Freising-Weihenstephan, Germany

Book chapters

Melanie Spornraft, Irmgard Riedmaier, Michael W. Pfaffl

Transkriptionelle Biomarkersignaturen in der Lebensmittelüberwachung

5. Fachtagung Gentechnik, Band 8 der Schriftenreihe Gentechnik für Umwelt und Verbraucherschutz Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit (LGL), ISBN: 978-3-945332-08-5

Scientific talks

Melanie Spornraft

RNA Profiling in Bovine Body Fluids using Next-Generation Sequencing

Advances in Genome Science: Next-Generation Sequencing Illumina Seminar, Helmholtz Zentrum München, 23. April 2015, München, Germany

Melanie Spornraft

RNA Quality and Primer Design as Prerequisites for Successful qPCR Experiments

Centre for Veterinary Medicine, Faculty of Veterinary Medicine of Ludwig-Maximilian-University Munich, 25. August 2014, Oberschleißheim, Germany

Curriculum vitae

Melanie Spornraft

Date of birth	September, 18th 1986
Place of birth	Mainburg
Nationality	German
12/2011 – 09/2015	Research assistant and PhD student Technical University Munich
10/2009 – 08/2011	Studies in Molecular Biology Technical University Munich, Center of Life Sciences Weihenstephan Degree: Master of Science, M.Sc.
10/2007 – 09/2009	Studies in Biology Technical University Munich, Center of Life Sciences Weihenstephan Degree: Bachelor of Science, B.Sc.
09/1997 – 06/2006	Graduation Allgemeine Hochschulreife Hallertau Gymnasium Wolnzach
09/1993 – 07/1997	Primary School Au in der Hallertau

Appendix

Appendix I:

Melanie Spornraft, Benedikt Kirchner, Bettina Haase, Vladimir Benes, Michael W. Pfaffl, Irmgard Riedmaier

Optimization of Extraction of Circulating RNAs from Plasma – Enabling small RNA Sequencing

PLoS One eCollection, 17. September 2014, DOI: 10.1371/journal.pone.0107259

Appendix II:

Melanie Spornraft, Benedikt Kirchner, Michael W. Pfaffl, Irmgard Riedmaier

Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing

Biotechnology Letters, 21. February 2015, DOI: 10.1007/s10529-015-1788-2

Appendix III:

Irmgard Riedmaier, Melanie Spornraft, Michael W. Pfaffl

Identification of a potential gene expression biomarker signature in bovine liver to detect the abuse of growth promoters

Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment, 3. April 2014, DOI: 10.1080/19440049.2014.886341

Appendix IV:

Melanie Spornraft, Benedikt Kirchner, Michael W. Pfaffl, Irmgard Riedmaier

The Potential Of Circulating Extracellular Small RNAs (smexRNA) In Veterinary Diagnostics – Identifying Biomarker Signatures by Multivariate Data Analysis

Biomolecular Detection and Quantification, Volume 5, 19. September 2015, Special Issue: Advanced Molecular Diagnostics for Biomarker Discovery – Part I, DOI: 10.1016/j.bdq.2015.08.001

Appendix V:

Table S1: miRNA data evaluation of all 21 sequenced samples.

Table S2: piRNA data evaluation of all 21 sequenced samples.

Appendix I

OPEN ACCESS Freely available online

PLOS ONE

Optimization of Extraction of Circulating RNAs from Plasma – Enabling Small RNA Sequencing

Melanie Spornraft[1]*¤, Benedikt Kirchner[1], Bettina Haase[2], Vladimir Benes[2], Michael W. Pfaffl[1], Irmgard Riedmaier[1]

1 Physiology Weihenstephan and ZIEL Research Center for Nutrition and Food Sciences, Technical University Munich, Freising, Germany, **2** Genomics Core Facility, EMBL European Molecular Biology Laboratory, Heidelberg, Germany

Abstract

There are several protocols and kits for the extraction of circulating RNAs from plasma with a following quantification of specific genes via RT-qPCR. Due to the marginal amount of cell-free RNA in plasma samples, the total RNA yield is insufficient to perform Next-Generation Sequencing (NGS), the state-of-the-art technology in massive parallel sequencing that enables a comprehensive characterization of the whole transcriptome. Screening the transcriptome for biomarker signatures accelerates progress in biomarker profiling for molecular diagnostics, early disease detection or food safety. Therefore, the aim was to optimize a method that enables the extraction of sufficient amounts of total RNA from bovine plasma to generate good-quality small RNA Sequencing (small RNA-Seq) data. An increased volume of plasma (9 ml) was processed using the Qiagen miRNeasy Serum/Plasma Kit in combination with the QIAvac24 Plus system, a vacuum manifold that enables handling of high volumes during RNA isolation. 35 ng of total RNA were passed on to cDNA library preparation followed by small RNA high-throughput sequencing analysis on the Illumina HiSeq2000 platform. Raw sequencing reads were processed by a data analysis pipeline using different free software solutions. Seq-data was trimmed, quality checked, gradually selected for miRNAs/piRNAs and aligned to small RNA reference annotation indexes. Mapping to human reference indexes resulted in 4.8±2.8% of mature miRNAs and 1.4±0.8% of piRNAs and of 5.0±2.9% of mature miRNAs for *bos taurus*.

Citation: Spornraft M, Kirchner B, Haase B, Benes V, Pfaffl MW, et al. (2014) Optimization of Extraction of Circulating RNAs from Plasma – Enabling Small RNA Sequencing. PLoS ONE 9(9): e107259. doi:10.1371/journal.pone.0107259

Editor: Christophe Antoniewski, CNRS UMR7622 & University Paris 6 Pierre-et-Marie-Curie, France

Received February 18, 2014; **Accepted** August 12, 2014; **Published** September 17, 2014

Funding: The project was funded by internal budget of the Chair of Physiology, TUM. This work was supported by the German Research Foundation (DFG) and the Technische Universität München within the funding programme Open Access Publishing. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: melanie.spornraft@wzw.tum.de

¤ Current address: Physiology Weihenstephan and ZIEL Research Center for Nutrition and Food Sciences, Technical University Munich, Freising, Germany

Introduction

Since the discovery of microRNAs (miRNAs) in *c.elegans* in the year 1993, the research in the world of small non-coding RNAs (ncRNAs) increased drastically. Due to the finding of novel RNA classes that have no protein-coding potential and do not overlap with already characterized non-coding RNA species, there are new transcript populations with various sizes and functions. Their role in RNA modification and silencing needs to be further unraveled. NcRNAs which are important for protein translation like tRNA and rRNA or RNAs responsible for RNA function and maturation e.g. small nuclear RNAs (snRNA) or small nucleolar RNAs (snoRNA) are well known and intensively investigated for decades. In the last years, additional species of ncRNAs have been discovered and their examination is very popular in recent science. MiRNAs and PIWI-interacting RNAs (piRNAs) as members of the RNA interference (RNAi) group play a major role in this context.

MiRNAs have a length of 19–25 nucleotides (nts) and negatively regulate gene expression by translational inhibition of mRNAs or by mRNA degradation. In the year 2000, miRNAs were detected in humans and over the years, correlations between miRNAs and diseases like cancer and cardio-vascular diseases were discovered [1]. In 2008, miRNAs were found in plasma and one year later they were used as plasma markers for cardiac disease in clinical diagnostics [1].

PiRNAs are longer than miRNAs (25–32 nts) and bind to the piwi- (P-element induced wimpy testes)/argonaute protein family, whose functions were primarily discovered in a *Drosophila*-mutant [2]. The PIWI-piRNA-complex is mandatory in gametogenesis as it is linked to germline and stem cell development as well as to gene silencing by regulating transposons [3]. However, there are few publications about the appearing of piRNAs in organisms, the piRNA distribution in different organs and fundamental research needs to decode their biological function.

Profiling of serum or plasma and/or other body fluids (e.g. urine, breast milk, saliva) for regulating small RNAs can provide new biomarkers for a broad range of diseases and biological processes. Their analysis offers various advantages which could make them a goldmine in the identification of novel biomarkers:

- Small RNAs are found in non- or minimally invasive specimens simplifying sampling for both the clinician and the patient compared to tissue sampling via biopsies.
- They are relatively stable in clinical samples regarding RNase digestion, temperature variation and multiple freeze-thaw cycles [4], [5].

- The detection of small RNA biomarkers is already proven in different illnesses, e.g. in cancer [5], [6], cardiovascular diseases [7], myocardial infarction [8] and central nervous system diseases [9].
- Nowadays, analysis of ncRNAs is a highly updated research field. The interest in the survey of their post-translational actions is massive as they show great promise for new insights in cell biology [10].

In the search for circulating nucleic acids plasma biomarkers, mainly miRNAs are in the focus of investigation. The commonly used strategy therefore is using microarrays but mainly RT-qPCR [11], [12]. However, the number of analyzed miRNAs is limited when working with these methods and the RNAs of interest need to be known before the experiment. NGS offers researchers a completely detailed view into their samples. All transcripts are captured in an integral picture of the sample. In terms of throughput, sequence output, data generation rate and data quality, sequencing performance is continuously increasing while costs, hands-on time in the lab and sequencing durations are decreasing. That's why NGS is considered to be the state-of-the-art technology when it comes to holistic gene expression profiling. However, detecting circulating nucleic acids biomarker profiles in plasma by NGS and especially by small RNA-Seq is in its infancy.

The amount of RNA which can be extracted with purchasable kit systems for circulating RNAs from body fluids is only sufficient for a limited number of RT-qPCR analyses, but not for the application of small RNA-Seq. Therefore, small RNA-Seq biomarker profiling in body fluids is very challenging. Commercially available kits for RNA isolation from bio-fluids use low starting volumes: 50–200 µl (miRNeasy Serum/Plasma Kit, Qiagen, Germany) or 300 µl (NucleoSpin miRNA Plasma Kit, Macherey-Nagel, Germany). Standard methods for measuring RNA yield with UV-Vis spectrophotometric instruments like Nano Drop (Thermo Scientific, Germany) are not optimally suited for measurements below the detection limit of 2 ng/µl of RNA. Moreover, using the Small RNA Analysis Kit on the Bioanalyzer 2100 (Agilent Technologies, Germany) to resolve and quantify the small nucleic acid fraction of extracted RNAs does also not provide signals in electropherograms. Furthermore, reverse transcription reactions request a consistent input amount of extracted RNA, which is difficult as long as the concentration cannot be reliably measured. In addition, contamination of plasma samples with cellular RNA from lysed cells (e.g. apoptotic epithelial cells in urine or ruptured red blood cells during sampling) can alter the RNA expression profile and disturb robust biomarker detection [13].

Besides profiling of clinical samples for disease-specific biomarkers, the screening for distinct and convincing RNA signatures is an applied method in the surveillance of food safety and anti-doping control systems in food-delivering livestock as well as in human or equine competitive sports [5], [14], [15]. Including small RNAs in the screening for differentially expressed genes on the transcriptional level could lead to multiple RNA biomarkers that are more robust, reliable and failure-safe. Here, an adaption of an RNA isolation method is presented that enabled the holistic analysis of circulating cell-free miRNAs and piRNAs in bovine plasma via small RNA-Seq.

Materials and Methods

Plasma sampling

Blood was taken from *vena jugularis* from nine male Holstein Friesian calves at the age of 6 months. For blood drawing, 9 ml K3E K_3EDTA-Vacuette tubes (Greiner bio-one, Germany) and single use needles (20G x 1½", Greiner bio-one, Germany) were used. Plasma was separated from other blood components directly after blood sampling (15 min for 3500×rcf at room temperature with transportable centrifuges EBA20, Hettich, Germany) and stored at -80°C until RNA extraction. Blood collection was approved by the ethical committee of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (Recklinghausen, Germany) (permit number 84-02.04.2012.A040). Animals were housed and fed according to good animal attendance practice and all efforts were made to minimize suffering.

Total RNA extraction

Frozen plasma samples were used for small RNA extraction. It is important to consider the existence of contaminations containing RNAs such as intact cells, apoptotic cells or cell fragments. The presence of cells or lysis of blood cells with a following release of their content to plasma can alter the RNA expression profile. Thus, after thawing, plasma samples were centrifuged at 3000×rcf for 5 min at room temperature to pellet debris.

To extract circulating RNA species from plasma samples, the miRNeasy Serum/Plasma Kit (Qiagen, Germany) was used with modifications. Instead of 200 µl of initial sample material, a volume of 9 ml was applied. To improve the handling of increased volumes and to prevent multiple column loadings, the resulting column clogging, the need for bigger consumables, e.g. centrifugation tubes, and more needed time, the vacuum device QIAvac24 Plus (Qiagen, Germany) was used. 9 ml were separated in 3×3 ml fractions and transferred to 10 ml plastic centrifuge tubes with push caps (Sarstedt, Germany). The amount of cell lysis reagent QIAzol (Qiagen, Germany) and chloroform (AppliChem, Germany) was raised accordingly to 9 ml and 6 ml, respectively. 3 ml of QIAzol were added to each centrifuge tube, mixed by vortexing and incubated at room temperature for 5 min. 2 ml chloroform were added to each tube, vortexed for 15 s and incubated at room temperature for 3 min. An EBA20 benchtop centrifuge (Hettich, Germany) was used in a 4°C cooling chamber for the phase separation steps (15 min, 3500×rcf). After centrifugation, the mixture separates clearly into three phases: the upper aqueous, transparent phase containing RNA, the white interphase and the lower organic, reddish phase with proteins. The aqueous phases of the three separated samples were pooled in a 50 ml reaction tube followed by addition of 1.5 volumes of ethanol (100%) and transfer to one spin column. A vacuum pump was connected to the QIAvac24 Plus vacuum manifold equipped with luer plugs, RNeasy MinElute spin columns and tube extenders (Qiagen, Germany). Washing and binding buffers were applied twice. RNA was eluted with an increased volume of RNase-free water (50 µl) by applying two times 25 µl. To compare RNA yields from the presented method with those from purchasable kits, RNA was extracted from plasma samples using the miRNeasy Serum/Plasma Kit (Qiagen, Germany), the NucleoSpin miRNA Plasma Kit (Macherey Nagel, Germany) and the miRNeasy Mini Kit (Qiagen, Germany) with their recommended supplementary miRNA extraction protocol. If an initial plasma volume other than stated in the manufacturer's specifications was tested, multiple column loadings were necessary to process samples.

RNA Quantity and Quality Check

Subsequently, RNA yields were measured using the Qubit 2.0 Fluorometer (Life Technologies, Germany) in combination with the RNA Assay Kit (Life Technologies, Germany) according to the manufacturer's protocol. The maximum volume of sample input was used (20 µl), standards were freshly prepared and the Qubit

was equilibrated after the manufacturer's instructions. A Bioanalyzer 2100 (Agilent Technologies, Germany) run using the Small RNA Kit (Agilent Technologies, Germany) was performed afterwards for the analysis and quantification of RNA eluates resolving small RNAs in the range from 6 to 150 nts length. The extracted RNA was stored at −80°C until further analysis.

Library Preparation and small RNA Sequencing

For sequencing, the small RNA transcripts were converted into barcoded cDNA libraries. Library preparation was performed with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs Inc., USA) followed by small RNA-Seq analysis on the Illumina HiSeq2000 platform (Illumina Inc., USA). Limited RNA quantity led to library preparation with 35 ng of RNA as starting material. Multiplex adaptor ligations, reverse transcription primer hybridization, reverse transcription reaction and the PCR amplification were processed with regard to the protocol for library preparation (Protocol E7330, New England BioLabs Inc., USA). When working with lower RNA input, the protocol offers modifications at several steps, for example a longer incubation time and reduced temperatures in the adaptor ligation step. These modifications are increasing the ligation efficiency of methylated RNAs such as piRNAs. Therefore, the reaction was carried out for 18 h at 16°C instead of 1 h at 25°C. After PCR pre-amplification, the cDNA constructs were purified with the MinElute PCR Purification Kit (Qiagen, Germany) and loaded on the Bioanalyzer 2100 (Agilent, Germany) using the DNA 1000 Kit (Agilent, Germany) according to the manufacturer's protocol. For the size selection of amplified cDNA libraries, PCR products were loaded on an agarose gel (4%), appropriate bands of approximately 135 bp to 160 bp in size were cut out and passed on to gel extraction with the MinElute Gel Extraction Kit (Qiagen, Germany). A concluding Bioanalyzer 2100 run with the High Sensitivity DNA Kit (Agilent Technologies, Germany) that allows the analysis of DNA libraries regarding size, purity and concentration completed the workflow of library preparation. The obtained sequence libraries were subjected to the Illumina sequencing pipeline, passing through clonal cluster generation on a single-read flow cell (Illumina Inc., USA) by bridge amplification on the cBot (TruSeq SR Cluster Kit v3-cBOT-HS, Illumina Inc., USA) and 50 cycles sequencing-by-synthesis on the HiSeq2000 (Illumina Inc., USA).

Software, Statistics and Mapping

To assess overall NGS data quality, adaptor sequences were trimmed from the 3′end using Btrim [16] and reads without detectable adaptors were excluded from the data set. Sequence length distribution as well as the base calling accuracy as indicated by the phred quality score (Q score) were calculated with high throughput sequence data quality control software FastQC (Babraham Bioinformatics, UK, Version 0.10.1). To avoid distortion and generation of false positive mappings by degraded RNA material and other small reads, the data set was depleted of all reads with a sequencing length of less than 16 nts for miRNA and 26 nts for piRNA analysis respectively. In addition, all rRNA, tRNA, snRNA as well as snoRNA reads were omitted using their corresponding bovine sequences obtained from Rfam database [17] prior to mapping to further improve specificity of the read data.

All mappings including the previous Rfam comparison were done via Bowtie short read aligner [18] using the default parameters with the exception of the "best" alignment algorithm and only allowing one mismatch in the first 15 nts. Aligned reads were then sorted and indexed by SAMtools [19] and final readcounts were generated by calling the sum of hits per miRNA or piRNA sequence respectively. To obtain miRNA readcounts, the trimmed and filtered reads were aligned to the most recent miRBase database of mature miRNAs ([20], release 20) for *bos taurus* (bta) and *homo sapiens* (hsa). Likewise for piRNA readcounts, reads were first subjected to an additional filtering on the complete miRBase database of mature miRNAs. To generate a bowtie index for piRNA alignment, the nucleotide database of NCBI Genbank ([21], Release 201) was searched for human piRNA entries (32.046 entries in February 2014) that were merged to a file of short piRNA sequences. In addition, reads mapped on this index were checked for piRNA specific traits. A sequence motif analysis was done to evaluate 5′-T-bias by creating a positional weight matrix of all mapped piRNA reads using R and the packages ShortRead, Biostrings and seqLogo (http://www.bioconductor.org/). Furthermore, reads were mapped on the bovine genome (NCBI UMD 3.1, [22]) using bowtie and tested for ping-pong-cycle formation using a python script developed by Antoniewski [23]. To detect piRNA clusters in the bovine genome, piRNA reads were mapped with SeqMap [24] allowing no mismatches and resulting Eland3-format data from all nine animals was combined (NGS tools for the novice, http://www.uni-mainz.de/FB/Biologie/Anthropologie/487_ENG_HTML.php). ProTRAC software was used with the default settings but random base composition was chosen and the minimal score for accumulation of loci with typical length was set to zero to analyze and visualize clusters [25].

Although bovine plasma was used, the analysis of reads that were mapped to the human reference databases is necessary. First, hsa mapping results are crucial for the evaluation of piRNAs as there are -up to now- no annotated piRNAs for bovines. Second, the mapping output for hsa consisted of 2.578 entries for mature miRNAs and of 32.046 entries for piRNAs, whereas the bta output consisted of 783 entries for mature miRNAs and of none for the bovine piRNAs. Sequenced reads were deposited in the European Nucleotide Archive ENA (http://www.ebi.ac.uk/ena/, accession numbers PRJEB6683/ERP006244).

Results

RNA Quantity and Quality

RNA yield of plasma samples that were extracted with a choice of commercially available kits were not quantifiable due to a too low amount of RNA (Table 1). Although the amount of sample input was increased in the tested kits during the assay optimization process, RNA was still not measureable. Only by using the miRNeasy Serum/Plasma Kit (Qiagen, Germany), experiments with a raised plasma input (6 ml) showed quantifiable RNA concentrations (Table 1). After optimization, the miRNeasy Serum/Plasma Kit coupled with the QIAvac System was applied to isolate total RNA from 9 ml of nine different bovine plasma samples. Yields were 77.4 ng±24.9 ng which were determined by fluorometric quantification (Table 1).

The Bioanalyzer Small RNA Series II Assay (Agilent Technologies, Germany) resolves the small RNA fraction in the size range from 6 to 150 nts. No RNA could be detected in all the samples where Qubit measurement was not possible, irrespective of the initial plasma volume (Figure 1, [A] and [B]). Small peaks in the range of approximately 20 nts could be observed when 6 ml of plasma were extracted with the miRNeasy Serum/Plasma Kit (Figure 1, [C]). Lanes in Figure 1 [D] showed bands in the small RNA area when the presented method with 9 ml of plasma as initial sample input and the miRNeasy Serum/Plasma Kit in combination with the QIAvac system was utilized.

Table 1. Optimization process of total RNA extraction.

Isolation system	Plasma input [ml]	RNA yield [ng]	Sample
Assay optimization			
NucleoSpin miRNA Plasma Kit (Macherey-Nagel, Germany)	0.3	not quantifiable	
NucleoSpin miRNA Plasma Kit (Macherey-Nagel, Germany)	0.6	not quantifiable	
NucleoSpin miRNA Plasma Kit (Macherey-Nagel, Germany)	1.0	not quantifiable	
miRNeasy Kit, supplementary control (Qiagen, Germany)	0.3	not quantifiable	
miRNeasy Kit, supplementary control (Qiagen, Germany)	0.6	not quantifiable	
miRNeasy Kit, supplementary control (Qiagen, Germany)	1.0	not quantifiable	
miRNeasy Serum/Plasma Kit (Qiagen, Germany)	6.0	37.2	
miRNeasy Serum/Plasma Kit (Qiagen, Germany)	6.0	61.0	
Analyzed samples			
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	57.6	Animal 1
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	36.3	Animal 2
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	72.9	Animal 3
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	38.4	Animal 4
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	45.0	Animal 5
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	51.9	Animal 6
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	112.0	Animal 7
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	71.9	Animal 8
miRNeasy Serum/Plasma Kit + QIAvac System (Qiagen, Germany)	9.0	210.4	Animal 9

The table compiles the tested plasma input volumes, the used isolation systems and the resulting yields of extracted total RNA [ng] measured with the Qubit 2.0 Fluorometer.

doi:10.1371/journal.pone.0107259.t001

As the applicability of total RNA quantification on the one hand and the analysis of the small RNA fraction on the Bioanalyzer 2100 (Agilent Technologies, Germany) on the other hand was approved after optimization, nine bovine plasma samples were extracted using the isolation procedure as presented. For small RNA-Seq, an input amount of 1–5 µg total RNA is required, due to several quality control checkpoints during library preparation. When working with plasma samples and the aim is the analysis of circulating RNAs, the problem of low abundance of RNAs has to be faced. Due to the limited RNA quantity, library preparation was performed with 35 ng of RNA as starting material as this RNA yield was the least common denominator in RNA concentrations (Table 1).

Validation of Library Preparation

During library preparation, adaptors were hybridized to RNAs to bind the resulting sequencing library to oligonucleotides on the flow cell and index sequences were ligated to enable multiplexing

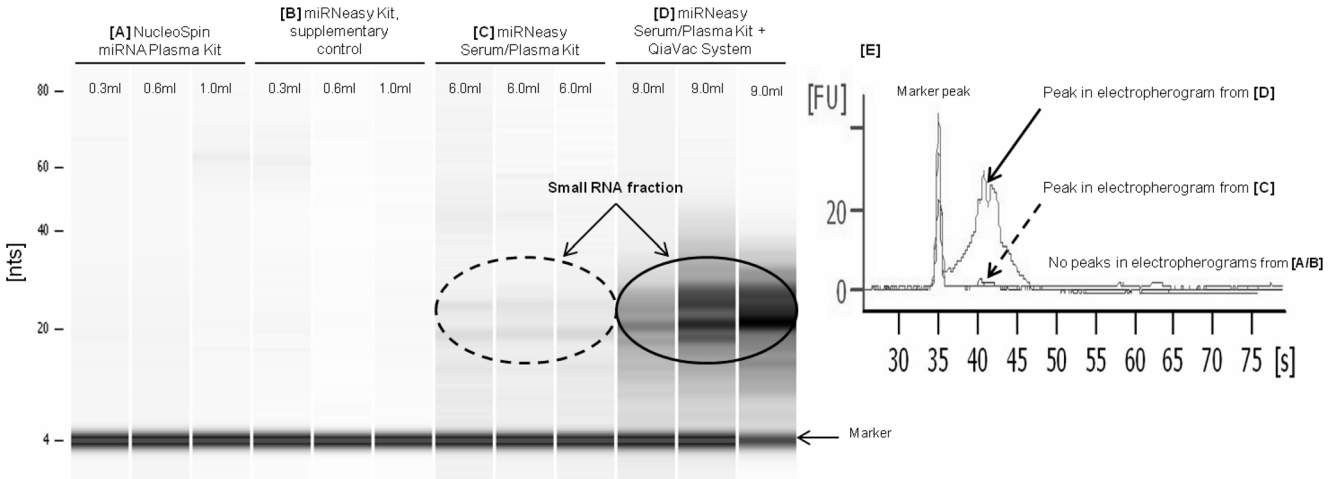


Figure 1. Small RNA Series II Bioanalyzer assay as checkpoint for RNA quantity after extraction procedure. Intensity of bands in the gel images ([A–D]) depicts the quantity of eluted RNAs with different isolation systems and varying plasma input volumes (0.3–9.0 ml). The first signal at 4 nts is the marker that is included in each run. Image [E] shows an overlay of representative electropherograms to illustrate size proportions. The fluorescence unit (FU) represents the signal intensity of small RNAs.

doi:10.1371/journal.pone.0107259.g001

of samples. The construct of the 3′-adaptor and the 5′-adaptor with the index sequence has the length of 119 nts. Therefore, depending on the RNA-insert length, a successfully amplified cDNA library depicts miRNA-/adaptor-constructs of approximately 138–144 nts and piRNA-/adaptor-constructs of 144–151 nts. A concluding High Sensitivity DNA Assay on the Bioanalyzer 2100 (Agilent Technologies, Germany) revealed the generation of cDNA libraries with adaptor-ligated constructs in the correct size, which signified the successful amplification of mature miRNAs and piRNAs (Figure 2).

Sequencing Quality Criteria

For the nine sequenced sample outputs the total phred scores and the average phred scores were calculated. The total phred score is assigned to each nucleotide of the reads and the average phred score is calculated for an entire sequence. For 98.3% of the sequenced nucleotides and for 98.7% of all reads, a Q score >30 was calculated. A Q score of 30 that is assigned to a base implies a probability of an incorrect base call in 1 of 1000 bp [26]. If more than 90% of the reads show an average quality score ≥30, the obtained data is considered to be of high quality regarding the base calling on the HiSeq platform [27]. Additionally, reads were analyzed regarding their proportions of sequences with and without ambiguities. 99.5% of all reads did not contain ambiguities.

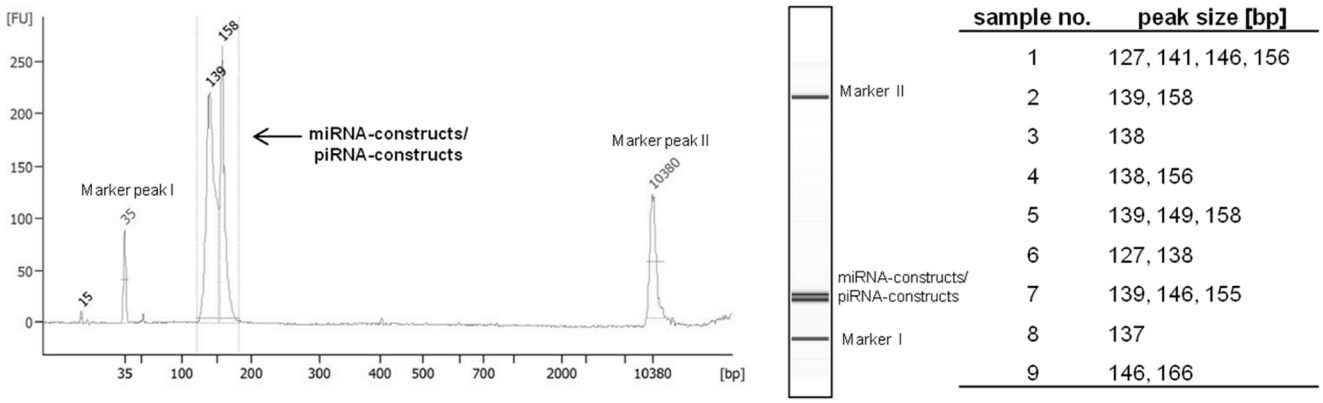
Evaluation of Readcounts

After trimming, data was analyzed in terms of sequence length distribution. The profile in Figure 3 displayed a bimodal pattern with two distinct peaks at 21 nts and 31 nts. The peak in the range of 19 to 25 nts represented the miRNA fraction, while piRNAs were depicted by read lengths between 26 and 32 nts. More readcounts showed a length that is specific for miRNAs (1.728.441±46.787 reads) compared to the number of readcounts that were in the size range of piRNAs (1.505.090±191.338 reads).

The sequenced samples contained on average 10.465.348± 3.638.968 reads (Table 2). 4.508.918±1.490.660 sequences passed the trimming quality threshold of 16 nts length, while 5.956.429±3.874.215 sequences failed the trimming process and were not further included in the data analysis steps (Table 2). 114.566±43.523 sequences were mapped to Rfam database, removing RNAs that were neither miRNAs nor piRNAs, but rRNAs, snRNAs, snoRNAs and tRNAs. Contrarily, sequences that were not mapped to Rfam database (4.394.352±1.467.749 reads) were processed in the search for annotated miRNAs using miRBase. miRBase sequence database provides 2.578 entries representing human mature miRNAs and 783 entries for bovine mature miRNAs. Mapping data to the hsa entries in miRBase, 471.640±269.100 reads were annotated, being a proportion of 4.8±2.8% of the total sequenced reads (Figure 4). Setting the threshold of sequenced readcounts to 50 reads at an average over all animals, 99.2% of annotated human miRNAs were abundant with more than the defined threshold. 482.072±274.637 reads could be mapped to annotated bta miRNA sequences accounting for a 5.0±2.9% share of the total readcounts (Figure 4). 99.5% of annotated reads showed more than 50 readcounts at an average over all sequenced samples. 3.467.882±1.212.523 reads remained unmapped to no database. A detailed summary of evaluated miRNA reads is given in Table 2 and Figure 5 [A] and [B] clarifies the proportions of trimmed, annotated and remaining readcounts.

Trimming data sets with the length threshold of 26 nts resulted in 1.792.803±1.125.965 reads that passed and 8.672.545± 4.295.024 reads that failed (Table 3). Sequences that were not mapped to Rfam and miRBase were mapped to 32.046 entries in NCBI's piRNA database, resulting in 128.333±69.954 matching human piRNA reads. Compared to the total number of sequenced reads, 1.4±0.8% of reads were assigned to be piRNAs (Figure 4). 98.5% of annotated piRNAs showed at an average over all animals more than 50 readcounts. Due to the fact that no bovine piRNAs are reported yet, there is no matching annotation reference library available. A detailed summary of evaluated piRNA reads is given in Table 3 and Figure 5 [C] clarifies the proportions of trimmed, annotated and remaining readcounts. MiRNA and piRNA readcounts normalized to library sizes are presented in Table S1.

According to the piRNA biogenesis model, the ping-pong model, piRNAs can be divided into primary and secondary piRNAs [28]. Primary piRNAs have a strong tendency for 5′ bias for uridine and do not have nucleotide bias at position 10, while secondary piRNAs have a bias for adenine (A) at position 10 but do not show a 5′ bias. Results of the sequence motif analysis are displayed in Figure 6. 69.1% of analyzed piRNA readcounts showed a thymin (T) at the first nucleotide and 5.2% an A at position 10. The human piRNA reference that was used for piRNA alignment also showed 5′-T bias (79.0%) that mirrors



sample no.	peak size [bp]
1	127, 141, 146, 156
2	139, 158
3	138
4	138, 156
5	139, 149, 158
6	127, 138
7	139, 146, 155
8	137
9	146, 166

Figure 2. High Sensitivity DNA Bioanalyzer assay as checkpoint for correct size selection during library preparation. All nine samples showed adaptor/RNA/adaptor-constructs in appropriate sizes. One electropherogram is shown as representative example. The lengths of adaptor-ligated constructs from all nine samples were reported as indicated in the column *peak size [bp]*. The initial peak at 35 bp and the final peak at 10.380 bp are marker peaks that are system inherent included in all runs.
doi:10.1371/journal.pone.0107259.g002

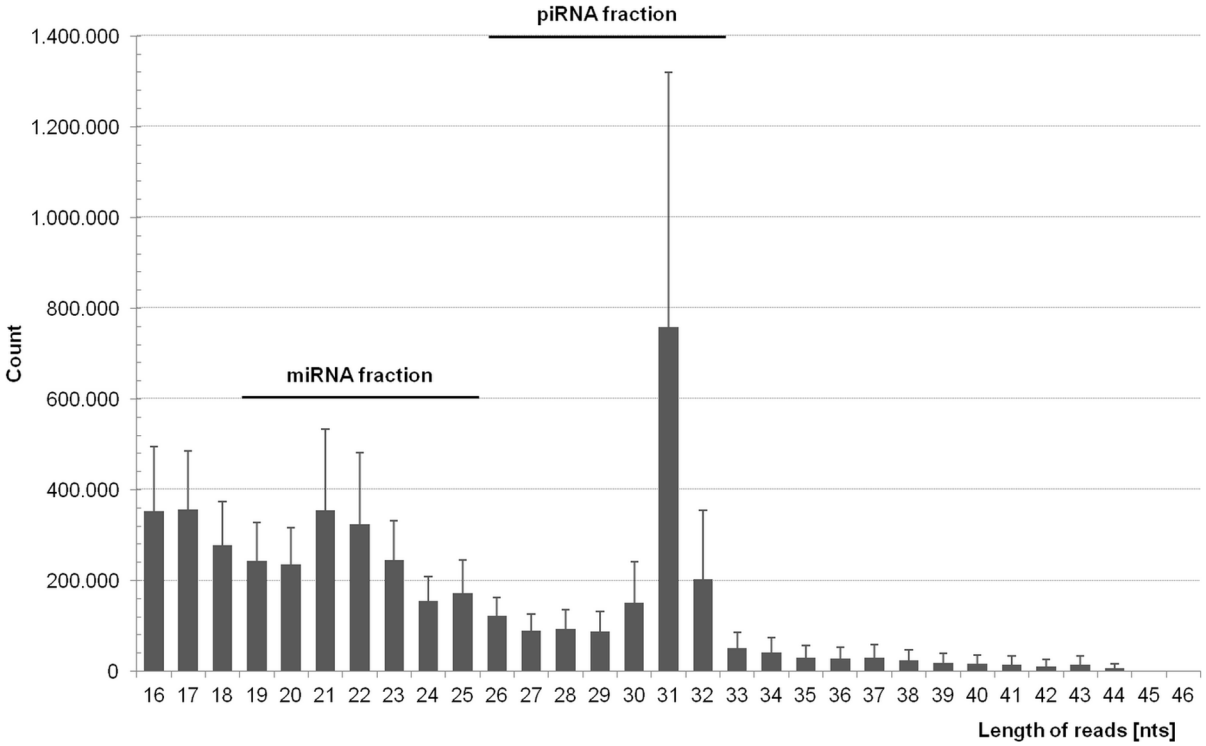


Figure 3. Sequence length distribution pattern analyzed by FastQC software. Calculating the average number of sequences with a certain length for all nine plasma samples, the profile displayed a bimodal pattern. The first peak includes sequences with a length between 20 nts and 24 nts reflecting miRNAs. The second peak indicates the piRNA fraction with sequences of 29 nts to 33 nts in size.
doi:10.1371/journal.pone.0107259.g003

expediency of this dataset. However, reference piRNAs were not weighted compared to the sequenced samples. The sequence motif with the highest probability is arising from the top 4 piRNAs that account for 66.2% of mapped piRNAs (Table S2). Computing sense and antisense piRNA overlaps, no ping-pong signatures could be identified. Using proTRAC software, no clusters could be identified in the bovine genome.

Discussion

Extraction Procedure

MiRNAs have shown to be powerful candidates in biomarker profiling for the detection of diseases or altered health conditions. Circulating RNAs hold great promise in finding new biomarkers not only in tissues but also in other body matrices like the easy-to-sample plasma. Analyzing circulating RNAs compared to cell-bound RNAs is a new approach in the screening for biomarkers. For the analysis of the transcriptome in plasma, a suitable isolation method is needed to perform reliable and holistic small RNA-Seq experiments. Several prevalent RNA isolation systems on the market were given a trial for the extraction of circulating RNAs from plasma samples. The initially tested NucleoSpin miRNA Plasma Kit from Macherey-Nagel as well as the Qiagen miRNeasy Mini Kit with a supplementary protocol could not provide measurable RNA yields. Extracting the starting volume of 9 ml with the NucleoSpin miRNA Plasma Kit was not performed, because the columns were not compatible with the vacuum device. However, plasma input was raised up to a fourfold than recommended, but higher volumes led to column clogging. Furthermore, during optimization process, Qiagen launched an adapted isolation kit for RNAs especially from plasma or serum matrices. Since the extraction method with this miRNeasy Serum/Plasma Kit using 9 ml of plasma provided sufficient RNA yields and good quality NGS data, it was omitted to test the obsolete miRNeasy Mini Kit system with the supplementary protocol.

Yields of the extracted RNAs fluctuated between the analyzed samples (Table 1). It has to be considered, that RNA was extracted from nine different bull calves and total RNA content may vary between animals due to a different health status, immune situation, genetic variation or other stimuli.

Analyzing the small RNA fraction on the Bioanalyzer 2100 (Agilent Technologies, Germany), double-peaks or shifted peaks towards a higher nucleotide length, as it can be seen in Figure 1 [D], indicated the presence of not solely miRNAs but also of slightly longer RNAs. Library validation reinforced this assumption (Figure 2) and mapping finally confirmed the presence of miRNAs as well as piRNAs (Table 2, Table 3).

cDNA Library Preparation

The produced cDNA libraries were tested for the right size and purity using a High Sensitivity DNA chip on the Bioanalyzer 2100 (Agilent Technologies, Germany). The adaptor-ligated constructs varied in size, depending on the length of the initial RNA fragment. Inserts derived from miRNAs with a length of approximately 19–25 nts result in constructs of 138–143 bp. For piRNAs, the corresponding fragment length would be 144–151 bp, as the insert RNA has 25–32 nts. Although there was not a marked peak in the size of >138 bp in all of the samples (Figure 2), peaks had a broad basis covering the range not only of miRNAs but also of slightly longer small RNAs. However, despite of the correct size of amplified constructs, no statement about the variety of the containing RNAs in a sample could be made at that point.

Table 2. miRNA data analysis shows the composition of evaluated reads from nine animals generated by computer data analysis pipeline using free software tools.

Data evaluation - pipeline miRNA	Animal 1	Animal 2	Animal 3	Animal 4	Animal 5	Animal 6	Animal 7	Animal 8	Animal 9	MEAN	SD
Total sequences	9.694.743	9.585.690	18.089.452	11.225.200	9.210.634	13.046.396	9.871.768	8.900.340	4.563.907	10.465.348	3.638.968
Passed trimming >16 nts	5.942.478	5.189.172	4.797.229	4.656.921	4.655.609	2.164.606	6.894.507	2.826.235	3.453.508	4.508.918	1.490.660
Failed trimming <16 nts	3.752.265	4.396.518	13.292.223	6.568.279	4.555.025	10.881.790	2.977.261	6.074.105	1.110.399	5.956.429	3.874.215
Rfam mapped	155.108	95.216	207.425	102.996	96.956	78.487	132.335	92.173	70.402	114.566	43.523
Rfam unmapped	5.787.370	5.093.956	4.589.804	4.553.925	4.558.653	2.086.119	6.762.172	2.734.062	3.383.106	4.394.352	1.467.749
Annotated reads miRBase bta	764.079	283.405	840.542	430.832	334.847	162.123	885.241	237.019	400.560	482.072	274.637
% (/total sequences)	7.9	3.0	4.6	3.8	3.6	1.2	9.0	2.7	8.8	5.0	2.9
Unmapped reads miRBase bta	5.023.291	4.810.551	3.749.262	4.123.093	4.223.806	1.923.996	5.876.931	2.497.043	2.982.546	3.912.280	1.269.337
Annotated reads miRBase hsa	752.137	281.446	827.746	416.990	323.844	159.927	861.539	233.927	387.208	471.640	269.100
% (/total sequences)	7.8	2.9	4.6	3.7	3.5	1.2	8.7	2.6	8.5	4.8	2.8
Unmapped reads miRBase hsa	5.035.233	4.812.510	3.762.058	4.136.935	4.234.809	1.926.192	5.900.633	2.500.135	2.995.898	3.922.711	1.273.750

doi:10.1371/journal.pone.0107259.t002

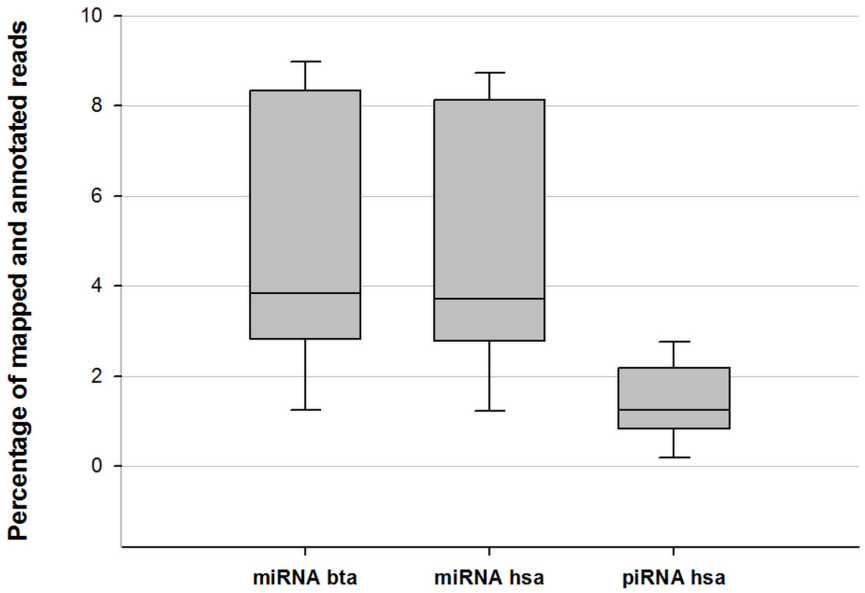


Figure 4. Percentage of mapped and annotated miRNA/piRNA reads compared to the total number of sequences.
doi:10.1371/journal.pone.0107259.g004

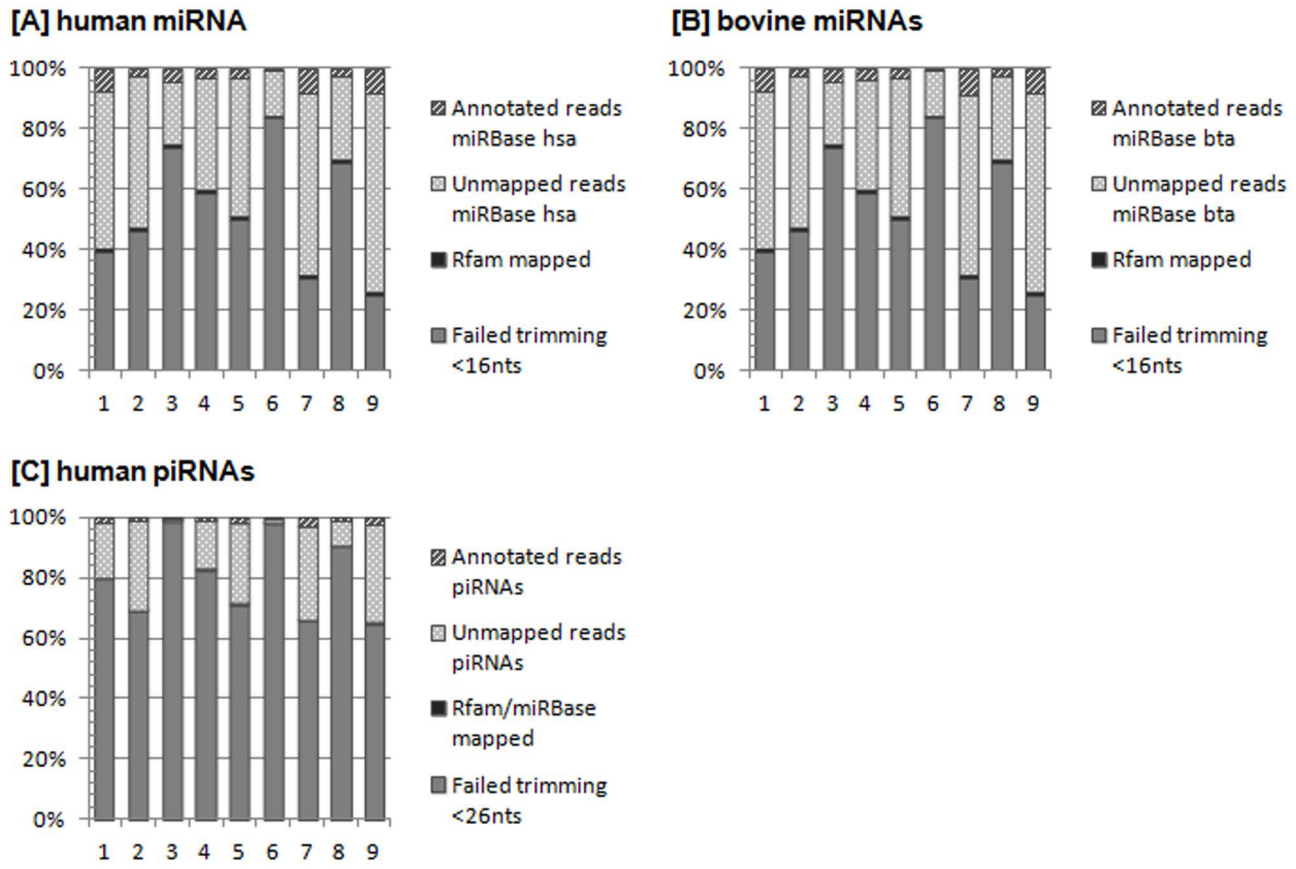


Figure 5. The proportions of trimmed, annotated and non-annotated reads. The total number of sequenced reads (100%) is divided into reads that failed trimming and reads that passed trimming and were mapped to Rfam database. Reads that were not mapped to Rfam database, were mapped to miRBase. Reads separated into annotated reads in miRBase and in reads that failed miRNA annotation. Image [A] displays miRNA results from mapping to human reference indexes. Image [B] presents miRNA results from mapping to bovine references. Regarding piRNAs (Image [C]), reads that could not be mapped to miRBase were aligned to piRNA database. They separate into annotated piRNAs and unmapped piRNAs.
doi:10.1371/journal.pone.0107259.g005

Table 3. piRNA data analysis shows the composition of evaluated reads from nine animals generated by computer data analysis pipeline using free software tools.

Data evaluation - pipeline piRNA	Animal 1	Animal 2	Animal 3	Animal 4	Animal 5	Animal 6	Animal 7	Animal 8	Animal 9	MEAN	SD
Total sequences	9.694.743	9.585.690	18.089.452	11.225.200	9.210.634	13.046.396	9.871.768	8.900.340	4.563.907	10.465.348	3.638.968
Passed trimming >26 nts	1.992.987	3.002.846	313.300	1.977.291	2.675.451	303.727	3.385.648	871.336	1.612.639	1.792.803	1.125.965
Failed trimming <26 nts	7.701.756	6.582.844	17.776.152	9.247.909	6.535.183	12.742.669	6.486.120	8.029.004	2.951.268	8.672.545	4.295.024
Rfam/miRBase mapped	24.229	27.254	10.254	15.375	23.495	4.166	34.196	9.794	14.273	18.115	9.723
Rfam/miRBase unmapped	1.968.758	2.975.592	303.046	1.961.916	2.651.956	299.561	3.351.452	861.542	1.598.366	1.774.688	1.116.766
Annotated reads piRNAs	144.667	89.786	33.649	140.601	190.090	101.363	271.987	77.665	105.186	128.333	69.954
% (/total sequences)	1.5	0.9	0.2	1.3	2.1	0.8	2.8	0.9	2.3	1.4	0.8
Unmapped reads piRNAs	1.824.091	2.885.806	269.397	1.821.315	2.461.866	198.198	3.079.465	783.877	1.493.180	1.646.355	1.065.235

doi:10.1371/journal.pone.0107259.t003

Sequencing and Mapping

To screen the abundance of piRNAs and mature miRNAs in bovine plasma, sequencing data were mapped to miRBase and NCBI's piRNA database according to the presented data analysis strategies. This revealed the presence of 4.8±2.8% of annotated human miRNAs, 5.0±2.9% of annotated bovine miRNAs and 1.4±0.8% of annotated human piRNAs in plasma samples (Figure 4, Table 2 and Table 3). It is known that piRNAs act mainly in Piwi-dependent transposon silencing, heterochromatin modification and in germ cell maintenance [29]. They were discovered in rat testes [30], but their presence was also confirmed in somatic tissues of fruit fly, mouse and *rhesus macaque* [31]. Thus, since their discovery in 2006, verification of their presence in various tissues and their role in different, not only germ-line affecting functions is gaining significance. It was recently shown that piRNAs were also found in the circulation, namely in exosomes, endosome-derived membrane microvesicles that contain specific RNA transcripts and are thought to be involved in cell-cell communication. Compared to results in human exosomes, which contained 1.31% of piRNAs, our findings in bovines are in the same dimension (Table 3) [32]. Exosomal miRNAs were represented in humans as the major part (76.2% of mappable miRNAs) [32], while circulating miRNAs in bovine plasma showed a presence of approximately 5% (Figure 4). Noticing that these are results from different species, the question why the miRNA content differs between exosomes and pure plasma gives rise to several assumptions. Plasma as analysis matrix is very sensitive to RNase digestion. MiRNAs are not only present in secure exosomes but also free in circulation and/or bound to proteins. As many reads did not pass the trimming due to a too short length (<16 nts), it is conceivable that the RNA degradation was already in an advanced stage (Figure 5, Table 2). However, the analysis of circulating miRNAs or piRNAs is still potential, due to a substantial number of mapped readcounts for each miRNAs and piRNAs (Table 2, Table 3). Hence, it still needs to be elucidated, if there is a balanced level of circulating small RNAs and if the miRNA/piRNA concentration in plasma is actively regulated by organisms. Moreover, clarification is needed to what extend miRNAs/piRNAs are present as apoptotic by-products or through active release.

Williams et al. reported a top 10 list of the most abundant circulating miRNAs in plasma samples from human volunteers using small RNA sequencing [33]. Comparing these results with our data collected by small RNA-Seq of bovine plasma samples, five out of ten miRNAs (miR-486, miR-21, miR-22, miR-25 and miR-92) were as well present in our top 20 list of most abundant miRNAs, either for the analyzed miRNA profile through mapping to hsa and bta miRBase entries (Table S3). Running the BLASTN algorithm (http://blast.ncbi.nlm.nih.gov, version 2.2.29, human genomic plus transcript database) revealed high query coverages (bta-miR-486: 46%, bta-miR-21: 100%, bta-miR-22 100%, bta-miR-25 100%, bta-miR-92: 95%) and therefore high similarity of bta-miRNAs aligned to human transcripts. The human most abundant miRNA that was reported by Williams et al. [33] is miR-451 (~50%), a red blood cell specific miRNA which is present in our dataset to 0.5% (Table S1 [A]). The second (miR-486) and third (miR-92a) most abundant human miRNAs are found in the top 10 list of sequenced bovine miRNAs (Table S3). The major miRNA in bovine plasma (miR-122) was not ranked amongst the top 10 human miRNAs. MiR-423-3p and miR-423-5p were ranked number 3 and 4 and were as well not present in the human top 10. Apart from pathological processes like hepatic cell death [34], heart failure [35] and type-2 diabetes [36], no relations to physiological conditions could be found in literature

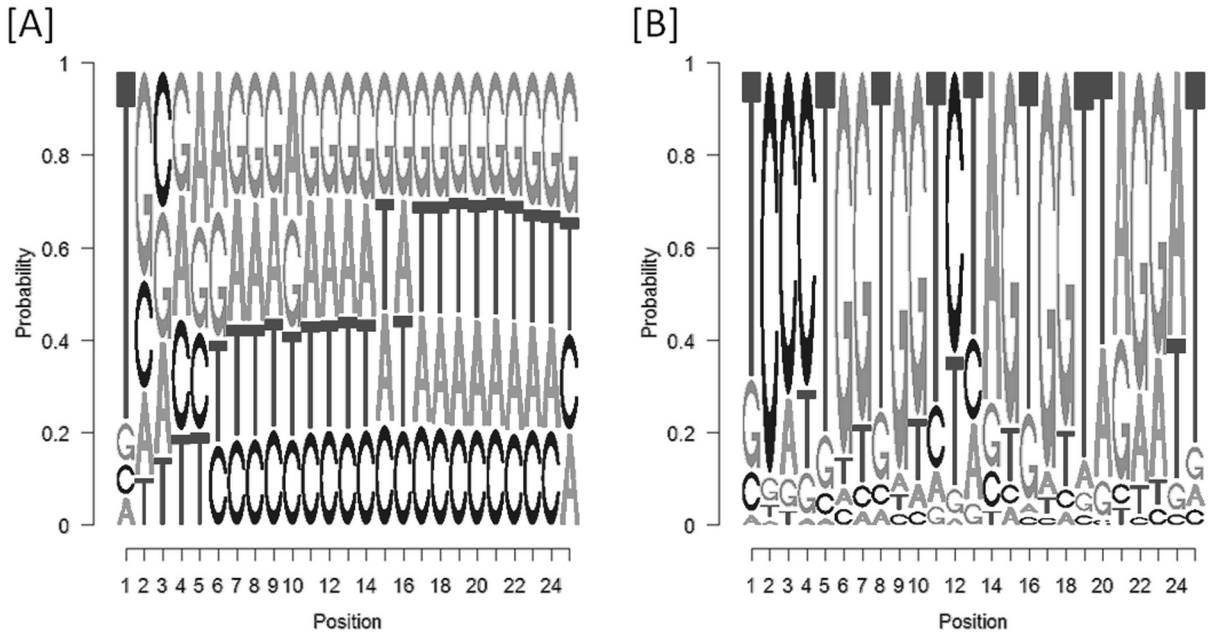


Figure 6. Sequence motif analysis to evaluate piRNA 5′-T-bias. Image [A] shows the piRNA reference that was used for alignment and Image [B] represents the nucleotide composition of mapped piRNAs pointing out the T-bias at 5′.
doi:10.1371/journal.pone.0107259.g006

that could explain altered miRNA compositions between humans and bovines. It remains unclear to what extend findings in humans can be transferred to bovines and vice versa due to great differences in their physiologies, e.g. the digestive system.

Clearly, there is an inconsistency in the number of mapped and annotated readcounts compared to reads that have the appropriate length of miRNAs and piRNAs, respectively (Figure 3, Table 2 and Table 3). Furthermore, there is a considerable part of a) sequenced reads that failed the trimming criterions and b) of small RNAs that could not be mapped at all (Figure 5). First, possibly there are more small RNAs in the samples that are not annotated yet. Second, as mentioned earlier, RNAs are exposed to a high content of RNases in plasma. Even though miRNAs are known to be relatively stable in plasma [4], other circulating RNAs, e.g. from apoptotic cells or active release, could be degraded to pieces with an equivalent size range and superimpose the length distribution profile.

Although no clusters could be found along the genome and neither a bias for an A at position 10 nor ping-pong structures could be observed, the indication for primary piRNAs remain considerable. Besides appropriate length distribution in combination with mapping to piRNA reference datasets, the piRNA sequence motif analysis raised evidence that bovine plasma contained primary piRNAs with a characteristic T nucleotide enrichment at 5′ (Figure 6 [B]). In addition, a 10 nt A bias was primarily observed in *drosophila melanogaster*, but not in mammals like *rhesus macaque* or *mus musculus* [31]. This signifies the knowledge gap in the role of piRNAs and their specific features in the mammalian circulation.

Conclusions

In conclusion, the amount of total RNA that is extracted using commercially available isolation kits for circulating RNAs from body fluids is only sufficient for RT-qPCR measurements but not for NGS analyses. Therefore, the presented optimized extraction procedure for plasma samples was developed to enable Next Generation small RNA Sequencing. The miRNeasy Serum/Plasma Kit (Qiagen, Germany) was used in combination with the QIAvac vacuum device to process the increased sample volumes (starting material = 9 ml plasma). Plasma is low RNA concentration material. This has to be kept in mind during the experimental design of animal trials to ensure a sufficient available sample volume.

NGS is a very fast developing and yet a highly advanced technology. Companies offer a multitude of isolation kits for diverse applications (e.g. specialized kits for different sample origins, RNA types that are to be analyzed, starting sample amount, pre-analysis storage of samples, etc.) and library preparation kits provide increasingly improving protocols for faster procedures with shorter hands-on times in the lab and less required starting material. Sample preparation and bioinformatics in the downstream data analysis are yet mostly not designed and certified for other samples than human. Hence, there is still a need for more specific requirements. However, the presented method enables the small RNA-Seq analysis of circulating, cell-free miRNAs and piRNAs in bovine plasma with good performance data and a substantial number of further usable readcounts, e.g. for differential gene expression profiling, although the proportion of miRNAs together with piRNAs on total readcounts is not higher than 10% (Figure 4, Table 2 and Table 3).

As mentioned above, plasma as minimally invasive sample could make the biomarker profiling highly attractive for example in molecular diagnostics, risk assessment or food safety [37]. This experiment exhibited new insights in the composition of bovine circulating small RNAs and described the presence of piRNAs. Consequently, better knowledge about piRNAs and their analysis could potentially lead to find biomarkers on other RNA levels than mRNAs or miRNAs. Besides plasma, other body fluids like urine, milk or saliva could also be suitable non-invasive biomarker matrices. Investigating the miRNA and/or the piRNA profile in bio-fluids via small RNA-Seq could be a new option to detect novel biomarker signatures.

Supporting Information

Table S1 [A] bovine miRNA readcounts, [B] human miRNA readcounts and [C] human piRNA readcounts normalized to library sizes in reads per million (rpm).
(XLSX)

Table S2 Top 15 piRNAs mainly contributing to the sequence motif analysis in Figure 6 [B].
(XLSX)

Table S3 Top 20 ranking list displays the most abundant circulating miRNAs in nine bovine plasma samples using small RNA-Seq. Data was compared to findings by Williams et al. [33]. Same miRNAs are marked in bold type. Table S3 [A] reports data from mapping to bovine entries in miRBase, Table S3 [B] lists results for mapping to human reference index.
(XLSX)

Author Contributions

Conceived and designed the experiments: MS BH IR. Performed the experiments: MS BH. Analyzed the data: MS BK. Contributed reagents/materials/analysis tools: VB MWP. Wrote the paper: MS.

References

1. van Rooij E (2011) The art of microRNA research. Circ. Res 108 (2): 219–234.
2. Cox DN, Chao A, Baker J, Chang L, Qiao D, et al. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 12 (23): 3715–3727.
3. Seto AG, Kingston RE, Lau NC (2007) The coming of age for Piwi proteins. Mol. Cell 26 (5): 603–609.
4. Etheridge A, Lee I, Hood L, Galas D, Wang K (2011) Extracellular microRNA: a new source of biomarkers. Mutat. Res 717 (1–2): 85–90.
5. Mitchell CJ, Nelson AE, Cowley MJ, Kaplan W, Stone G, et al. (2009) Detection of Growth Hormone Doping by Gene Expression Profiling of Peripheral Blood. Journal of Clinical Endocrinology & Metabolism 94 (12): 4703–4709.
6. Tang D, Shen Y, Wang M, Yang R, Wang Z, et al. (2013) Identification of plasma microRNAs as novel noninvasive biomarkers for early detection of lung cancer. Eur. J. Cancer Prev 22 (6): 540–548.
7. Xu J, Zhao J, Evan G, Xiao C, Cheng Y, et al. (2012) Circulating microRNAs: novel biomarkers for cardiovascular diseases. J Mol Med 90 (8): 865–875.
8. Li C, Pei F, Zhu X, Duan DD, Zeng C (2012) Circulating microRNAs as novel and sensitive biomarkers of acute myocardial Infarction. Clin. Biochem 45 (10–11): 727–732.
9. Jin X, Wu N, Wang L, Li J (2013) Circulating microRNAs: a novel class of potential biomarkers for diagnosing and prognosing central nervous system diseases. Cell. Mol. Neurobiol 33 (5): 601–613.
10. Morillon A, Andrau J (2013) Editorial for "Diversity of the non-coding transcriptomes revealed by RNA-seq technologies". Methods 63 (1): 1–2.
11. D'Alessandra Y, Devanna P, Limana F, Straino S, Di Carlo A, et al. (2010) Circulating microRNAs are new and sensitive biomarkers of myocardial infarction. Eur. Heart J 31 (22): 2765–2773.
12. Kroh EM, Parkin RK, Mitchell PS, Tewari M (2010) Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). Methods 50 (4): 298–301.
13. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, et al. (2012) Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila) 5 (3): 492–497.
14. Riedmaier I, Pfaffl MW, Meyer HHD (2012) The physiological way: monitoring RNA expression changes as new approach to combat illegal growth promoter application. Drug Test Anal 4 Suppl 1: 70–74.
15. Bailly-Chouriberry L, Noguier F, Manchon L, Piquemal D, Garcia P, et al. (2010) Blood cells RNA biomarkers as a first long-term detection strategy for EPO abuse in horseracing. Drug Test Anal 2 (7): 339–345.
16. Kong Y (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98 (2): 152–153.
17. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, et al. (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41 (Database issue): D226–32.
18. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (3): R25.
19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16): 2078–2079.
20. Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research 42 (D1): D68.
21. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, et al. (2013) GenBank. Nucleic Acids Res. 41 (Database issue): D36–42.
22. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, et al. (2009) A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol 10 (4): R42.
23. Antoniewski C (2014) Computing siRNA and piRNA overlap signatures. Methods Mol. Biol. 1173: 135–146.
24. Jiang H, Wong WH (2008) SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics 24 (20): 2395–2396.
25. Rosenkranz D, Zischler H (2012) protract–a software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics 13: 5.
26. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8 (3): 186–194.
27. Illumina Inc. (2011) Quality Scores for Next-Generation Sequencing. Assessing sequencing accuracy using Phred quality scoring. Technical Note: Sequencing. Available: http://res.illumina.com/documents/products/technotes/technote_q-scores.pdf.
28. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128 (6): 1089–1103.
29. Mei Y, Clark D, Mao L (2013) Novel dimensions of piRNAs in cancer. Cancer Lett 336 (1): 46–52.
30. Grivna ST, Beyret E, Wang Z, Lin H (2006) A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 20 (13): 1709–1714.
31. Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, et al. (2011) Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res 39 (15): 6596–6607.
32. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, et al. (2013) Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics 14: 319.
33. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, et al. (2013) Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc. Natl. Acad. Sci. U.S.A. 110 (11): 4255–4260.
34. Selth LA, Tilley WD, Butler LM (2012) Circulating microRNAs: macro-utility as markers of prostate cancer. Endocrine Related Cancer 19 (4): R99.
35. Tijsen AJ, Creemers EE, Moerland PD, Windt LJ de, van der Wal AC, et al. (2010) MiR423-5p as a circulating biomarker for heart failure. Circ. Res. 106 (6): 1035–1039.
36. Ortega FJ, Mercader JM, Moreno-Navarrete JM, Rovira O, Guerra E, et al. (2014) Profiling of Circulating MicroRNAs Reveals Common MicroRNAs Linked to Type 2 Diabetes That Change With Insulin Sensitization. Diabetes Care 37 (5): 1375–1383.
37. Riedmaier I, Pfaffl MW (2013) Transcriptional biomarkers–high throughput screening, quantitative verification, and bioinformatical validation methods. Methods 59 (1): 3–9.

Appendix II

Biotechnol Lett (2015) 37:1165–1176
DOI 10.1007/s10529-015-1788-2

ORIGINAL RESEARCH PAPER

Comparison of the miRNome and piRNome of bovine blood and plasma by small RNA sequencing

Melanie Spornraft · Benedikt Kirchner · Michael W. Pfaffl · Irmgard Riedmaier

Received: 10 December 2014 / Accepted: 5 February 2015 / Published online: 21 February 2015

Abstract
Objectives The small RNAs of bovine plasma and whole blood were analysed using next-generation sequencing to quantify, profile and compare the microRNAs (miRNA) and piRNA signatures in both bio fluids.
Results Evaluating read-count data resulted in a proportion of 5.0 ± 2.9 % of miRNAs in plasma while 38.2 ± 3.4 % were identified in whole blood. Regarding piRNAs, the percentages in both matrices were nearly the same: 1.4 ± 0.8 % of piRNAs in plasma and 1.9 ± 0.8 % in whole blood. Investigation of the ten most abundant miRNAs and piRNAs in both bio fluids revealed that two miRNAs and seven piRNAs were identical. Comparing the read-count values of these matching pairs highlighted that miRNA and piRNA levels in blood exceeded the abundance of their corresponding miRNAs and piRNAs in plasma, except liver-specific miR-122 and three piRNAs.
Conclusions The data strengthened evidence that the circulating small RNA signature in plasma is not only influenced by hematocytes and certain small RNAs could originate from other sources than cellular blood components.

Keywords Blood analysis · microRNA · Next-generation sequencing (NGS) · piRNA · Plasma analysis · Small RNAs · Small RNA-Seq

M. Spornraft (✉) · B. Kirchner · M. W. Pfaffl · I. Riedmaier
Physiology Weihenstephan, TUM School of Life Sciences Weihenstephan, Technical University Munich, Weihenstephaner Berg 3, 85354 Freising, Germany
e-mail: melanie.spornraft@wzw.tum.de

B. Kirchner
e-mail: benedikt.kirchner@wzw.tum.de

M. W. Pfaffl
e-mail: michael.pfaffl@wzw.tum.de

I. Riedmaier
e-mail: irmgard.riedmaier@wzw.tum.de

Introduction

The presence of free-circulating, non-cell-bound microRNAs (miRNAs) has already been verified for various bio fluids, e.g. urine, serum, plasma, saliva, tears, cerebrospinal fluid, breast milk, colostrum etc. (Weber et al. 2010). Non- or minimal-invasive sample matrices clearly offer multiple advantages compared to biopsies and additionally, the potential of circulating miRNAs as biomarkers for different diseases and physiological conditions has already been successfully demonstrated. Therefore, they earn great attention by scientists in the search for novel biomarkers. In addition, compared to other RNAs, circulating

miRNAs are highly stable, resistant to variations in temperature, pH value and multiple thaw and freezing cycles (Mitchell et al. 2008), which makes them suitable candidates for routine biomarker profiling in laboratories. Circulating miRNAs are small non-coding molecules with a length of 18–25 nucleotides (nts) that are either transported throughout the circulation encapsulated in microvesicles (e.g. exosomes) or associated and therefore protected by proteins, e.g. argonaute 2 (Arroyo et al. 2011) or nucleophosmin (Wang et al. 2010) or high density lipoproteins (Vickers et al. 2011). They act on the post-transcriptional level as modulators of mRNAs by either repressing their translation or destabilizing the corresponding mRNA targets and are therefore capable of regulating gene expression in manifold biological processes. In the search for miRNA biomarkers, altered levels of distinct miRNAs or miRNA signatures are investigated, but not much is known about the normal physiological level and composition of miRNomes, the collectivity of all miRNAs, in either tissues or bio fluids.

microRNAs are involved in numerous cell functions such as differentiation, proliferation and apoptosis and they can act both as promotors (e.g. oncogenes) or as suppressors (e.g. tumor suppressor genes) (as shown in Vasilatou et al. 2010). Hence, investigating the normal and healthy physiological landscape of miRNAs could lead to a better understanding of dysregulations diverging from ordinary physiology. MiRNAs in plasma are thought to have key roles in almost every stage of hematopoiesis (Ramkissoon et al. 2006) and additionally, plasma and the different hematocytes seem to be characterized by a specific miRNA profile. For example, miR-451 and miR-16 are contained in red blood cells (RBCs) at high levels but at low concentrations in leucocytes and thrombocytes (Ramkissoon et al. 2006; Kirschner et al. 2011).

Since miRNAs, as comparable class of posttranscriptional regulatory molecules, are very successful in the search for novel biomarkers, PIWI-interacting small non-coding RNAs (piRNAs) are also traded as new potential goldmine in molecular diagnostics (Qi and Du 2013). PiRNAs have a length of 25–32 nt and were initially detected in germ-line cells, where they maintain genome stability, are involved in RNA silencing and regulate gene expression (as reviewed in Gou et al. 2014). Hence, they act as genetic and epigenetic regulatory factors in the germ-line branch of the RNAi aspect. Nevertheless, differentially expressed small RNAs could also be identified in tumor cells, e.g. colon, lung, breast or gastric cancer (Cheng et al. 2011). Additionally, circulating piRNAs in bio fluids were already identified as biomarkers for several cancer types, e.g. gastric cancer (Cui et al. 2011). Consequently, besides the functions of piRNAs in germ-line, their role in somatic tissues or body fluids is up to date not well understood.

When the aim is developing a significant biomarker signature, it is important to consider that each body fluid appears to have a normal spectrum of miRNAs or piRNAs, which is presumably the reflection of healthy ordinary physiological processes (Kirschner et al. 2011). Therefore, besides screening for differentially expressed biomarkers, it is no less interesting to profile healthy individuals to increase the knowledge about the miRNA and piRNA status quo of certain samples. This paper describes the miRNA and piRNA signatures of plasma and whole blood of bovines and compares the two specimens regarding the miRNA and piRNA profiles, raising evidence that blood cells or blood vessel epithelial cells are contributing to the plasma community, but there must be further factors adding miRNA and piRNAs to the circulation. Yet little is known about the origin of circulating mi-/piRNAs in either healthy or pathogen conditions and which factors influence their expression levels.

High-throughput next-generation sequencing (NGS) was applied in this study to profile the totality of the "small RNA universe" in the analyzed samples. Small RNA-Sequencing (small RNA-Seq) is the technology of choice in sequencing entire small RNA transcriptomes. The benefits lie in high data output, massive parallel sequencing ability by multiplexing samples, decreasing costs and shorter hands-on-time during sample preparation and sequencing itself.

Materials and methods

Blood and plasma sampling

Peripheral blood was taken from *vena jugularis* from male Holstein–Friesian calves (n = 9). For blood drawing, 9 ml K3E K_3EDTA-Vacuette tubes (Greiner bio-one, Germany) and single-use needles (20

G × 1½″, Greiner bio-one, Germany) were used. Plasma was separated from other blood components directly after blood sampling (15 min for 3500×*g* at room temperature with transportable centrifuges) and stored at −80 °C until RNA extraction. Whole blood was collected with PAXgene Blood RNA Tubes (PreAnalytiX, Germany) that guarantee immediate stabilization of RNAs. PAXgene tubes were kept at −20 °C until RNA extraction. The ethical committee of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (Recklinghausen, Germany) permitted specimen sampling (permit number 84-02.04.2012.A040). Taking care of the animals according to good livestock farming practices was as granted as avoiding suffering.

Total RNA isolation

Circulating RNAs from plasma (starting volume = 9 ml) were isolated as presented by our group by an optimized method that enabled small RNA-Seq (Spornraft et al. 2014). Total RNA purification of whole blood was done by using the PAXgene Blood miRNA Kit (PreAnalytiX, Germany) and RNA eluates were stored at −80 °C until further usage.

Small RNA sequencing, data evaluation, mapping and annotation

The workflow starting with library preparation (NEBNext Multiplex Small RNA Library Prep Set for Illumina, New England BioLabs Inc., USA) until small RNA-Seq on a HiSeq 2000 high-throughput sequencing system (Illumina Inc., USA) was performed as previously described by our group (Spornraft et al. 2014). In brief, 35 ng total plasma RNA and 1 µg total blood RNA were adapter-ligated, reverse transcribed, PCR amplified and barcoded for multiplexing. Fractions were size-selected via gel electrophoresis and DNA libraries were controlled regarding correct size, purity and concentration with a High Sensitivity DNA chip on the Bioanalyzer 2100 (Agilent). Single-read sequencing-by-synthesis was carried out in 50 cycles. Seq-data was processed as previously described in (Spornraft et al. 2014). In brief, to generate miRNA and piRNA read-count tables, the data analysis pipeline basically consisted of adaptor trimming, quality controls, size assortment and stepwise mapping of reads to different databases to improve specificity. Data sets were gradually reduced of reads with a sequencing length of <16 nt for miRNAs and <26 nt for piRNA analysis. Reads with appropriate sequencing lengths were mapped to Rfam database (Burge et al. 2013) to exclude rRNA, tRNA, snRNA as well as snoRNA reads prior to aligning them to miRBase (Kozomara and Griffiths-Jones 2013, release 20), a high confidence miRNA repository to generate mature *Bos taurus* (bta) miRNA read-counts. To obtain piRNA read-counts, size-selected reads were first removed by additional filtering steps on Rfam and the complete miRBase database. To generate a bowtie index (Langmead et al. 2009) for piRNA alignment using reads that did not match either to Rfam nor miRBase, the nucleotide database of NCBI Genbank (Benson et al. 2013, Release 201) was searched for human piRNA entries. Although bovine plasma was used, the mapping of reads to human piRNA references was essential, as there are no piRNAs annotated yet for *bos taurus*. All read-counts were normalized to library sizes in reads per million (rpm). As a noise cut-off, reads with <50 rpm at an average were excluded from read-count tables. Sequencing data was recorded at the European Nucleotide Archive (ENA) with the study accession numbers PRJEB7882/ERP008868 (http://www.ebi.ac.uk/ena).

Results

Data evaluation

Sequencing the plasma samples resulted in $10.47 \times 10^6 \pm 3.64 \times 10^6$ reads and $16.16 \times 10^6 \pm 5.48 \times 10^6$ reads for whole blood samples (Table 1) (Total sequences). Raw sequencing reads were quality checked and size pre-filtered prior to mapping. To evaluate miRNAs, sequences that were shorter than 16 nt (Failed trimming <16 nt) were eliminated from the sequenced library of each animal. As a consequence, the number of sequences that passed the trimming process were further subjected to an alignment to Rfam database to further exclude unwanted RNAs such as rRNAs, tRNAs, snoRNAs and snRNAs (Rfam mapped). Unmapped sequences to Rfam database were aligned to bovine entries in miRBase as a second mapping step (annotated reads miRBase bta) (Table 1, plasma miRNA and blood miRNA). Sequences that

Table 1 Data evaluation. Compilation of evaluated read data in the process of generating miRNA and piRNA read-count lists of plasma and blood samples of all sequenced bovines (*animals 1–9*)

Data evaluation	Animal 1	Animal 2	Animal 3	Animal 4	Animal 5	Animal 6	Animal 7	Animal 8	Animal 9	Mean	SD
PLASMA miRNA											
Total sequences	9,694,743	9,585,690	18,089,452	11,225,200	9,210,634	13,046,396	9,871,768	8,900,340	4,563,907	10,465,348	3,638,968
Passed trimming >16nt	5,942,478	5,189,172	4,797,229	4,656,921	4,655,609	2,164,606	6,894,507	2,826,235	3,453,508	4,508,918	1,490,660
Failed trimming <16nt	3,752,265	4,396,518	13,292,223	6,568,279	4,555,025	10,881,790	2,977,261	6,074,105	1,110,399	5,956,429	3,874,215
Rfam mapped	155,108	95,216	207,425	102,996	96,956	78,487	132,335	92,173	70,402	114,566	43,523
Rfam unmapped	5,787,370	5,093,956	4,589,804	4,553,925	4,558,653	2,086,119	6,762,172	2,734,062	3,383,106	4,394,352	1,467,749
Annotated reads miRBase bta	764,079	283,405	840,542	430,832	334,847	162,123	885,241	237,019	400,560	482,072	274,637
% (/Total sequences)	7.9	3.0	4.6	3.8	3.6	1.2	9.0	2.7	8.8	5.0	2.9
Unmapped reads miRBase bta	5,023,291	4,810,551	3,749,262	4,123,093	4,223,806	1,923,996	5,876,931	2,497,043	2,982,546	3,912,280	1,269,337
PLASMA piRNA											
Passed trimming >26nt	1,992,987	3,002,846	313,300	1,977,291	2,675,451	303,727	3,385,648	871,336	1,612,639	1,792,803	1,125,965
Failed trimming <26nt	3,030,304	1,807,705	3,435,962	2,145,802	1,548,355	1,620,269	2,491,283	1,625,707	1,369,907	2,119,477	722,835
Rfam/miRBase mapped	24,229	27,254	10,254	15,375	23,495	4166	34,196	9794	14,273	18,115	9723
Rfam/miRBase unmapped	1,968,758	2,975,592	303,046	1,961,916	2,651,956	299,561	3,351,452	861,542	1,598,366	1,774,688	1,116,766
Annotated reads piRNAs	144,667	89,786	33,649	140,601	190,090	101,363	271,987	77,665	105,186	128,333	69,954
% (/Total sequences)	1.5	0.9	0.2	1.3	2.1	0.8	2.8	0.9	2.3	1.4	0.8
Unmapped reads piRNAs	1,824,091	2,885,806	269,397	1,821,315	2,461,866	198,198	3,079,465	783,877	1,493,180	1,646,355	1,065,235
BLOOD miRNA											
Total sequences	17,203,127	18,668,797	17,901,634	16,751,614	17,327,518	26,844,388	7,795,014	11,257,801	11,700,467	16,161,151	5,479,241
Passed trimming >16nt	14,386,825	16,003,311	14,515,834	12,985,148	12,955,457	21,324,626	6,315,968	8,959,206	9,165,248	12,956,847	4,447,406
Failed trimming <16nt	2,816,302	2,665,486	3,385,800	3,766,466	4,372,061	5,519,762	1,479,046	2,298,595	2,535,219	3,204,304	1,212,401
Rfam mapped	62,634	59,681	79,111	58,106	68,831	91,119	28,551	43,609	40,401	59,116	19,548
Rfam unmapped	14,324,191	15,943,630	14,436,723	12,927,042	12,886,626	21,233,507	6,287,417	8,915,597	9,124,847	12,897,731	4,429,380
Annotated reads miRBase bta	7,319,241	8,214,217	6,215,328	6,468,948	6,251,894	10,343,552	3,071,147	3,917,106	4,106,122	6,211,951	2,289,608
% (/Total sequences)	42.5	44.0	34.7	38.6	36.1	38.5	39.4	34.8	35.1	38.2	3.4
Unmapped reads miRBase bta	7,004,950	7,729,413	8,221,395	6,458,094	6,634,732	10,889,955	3,216,270	4,998,491	5,018,725	6,685,781	2,206,304
BLOOD piRNA											
Passed trimming >26nts	2,720,173	1,992,599	3,415,681	1,577,184	1,984,923	3,255,951	1,209,619	1,532,528	1,922,463	2,179,013	777,826
Failed trimming <26nts	4,284,777	5,736,814	4,805,714	4,880,910	4,649,809	7,634,004	2,006,651	3,465,963	3,096,262	4,506,767	1,616,907

Table 1 continued

Data evaluation	Animal 1	Animal 2	Animal 3	Animal 4	Animal 5	Animal 6	Animal 7	Animal 8	Animal 9	Mean	SD
Rfam/miRBase mapped	505	148	316	137	334	548	223	118	338	296	156
Rfam/miRBase unmapped	2,719,668	1,992,451	3,415,365	1,577,047	1,984,589	3,255,403	1,209,396	1,532,410	1,922,125	2,178,717	777,713
Annotated reads piRNAs	321,782	182,975	509,887	177,167	287,874	468,794	193,777	205,525	273,720	291,278	123,573
% (/Total sequences)	1.9	1.0	2.8	1.1	1.7	1.7	2.5	1.8	2.3	1.9	0.6
Unmapped reads piRNAs	2,397,886	1,809,476	2,905,478	1,399,880	1,696,715	2,786,609	1,015,619	1,326,885	1,648,405	1,887,439	663,048

The proportion of miRNA and piRNAs were calculated as percentage of annotated miRNAs or piRNAs to total sequences [*%(total sequences)*]

passed the length filtering (>26 nt) and were not aligned to Rfam database as well as miRBase database were mapped to piRNA reference indexes to generate annotated piRNA read-counts (annotated reads piRNAs) (Table 1, PLASMA piRNA and BLOOD piRNA). In plasma, more reads failed the size trimming and first selective mapping compared to blood that resulted in more annotated miRNAs in whole blood (Fig. 1a, b). Plasma data was subject of greater variations compared to the obviously more unified picture of blood parameters (Fig. 1a–d). The proportion of annotated miRNAs and piRNAs on the total sequences was calculated (%/total sequences). In plasma, 5 ± 2.9 % of miRNAs could be identified, while 38.2 ± 3.4 % could be identified in whole blood (Fig. 2, left panel). Concerning piRNAs, the share of piRNAs in both matrices was nearly the same: 1.4 ± 0.8 % of piRNAs in plasma and 1.9 ± 0.8 % in whole blood (Fig. 2, right panel).

Profiling of Top 10 miRNAs and piRNAs

All annotated miRNA and piRNA reads were normalized to library sizes in rpm and organized according to their abundances. Top 10 lists for the most abundant miRNAs and piRNAs respectively, in both plasma and whole blood were created by sorting read-counts according to MEAN values over all nine sequenced animals (Table 2). The sum of selected miRNAs/piRNAs was set to 100 % and respective percentages were calculated (% Top 10). Visualizing the Top 10 miRNA profiles in plasma compared to whole blood in pie charts easily allowed the comparison: two miRNAs (bta-miR-486 and bta-miR-92a) were present in both bio-fluids while the other eight differed between plasma and whole blood (Fig. 3a, c). Taking the shares of bta-miR-486 and bta-miR-92a together, their proportion of the Top 10 is 14.2 % for plasma and 35.4 % for whole blood (Fig. 3a, c). In both cases, bta-miR-486 appeared in second place in terms of occurrence. The situation was different when the piRNA profiles of plasma and whole blood were opposed: seven out of ten piRNAs were found in both sample types (piR-43772, piR-43771, piR-60565, piR-43786, piR-31068, piR-35982 and piR-33151) (Fig. 3b, d). The matching piRNAs in plasma accounted for a portion of 84.9 % and of 88.7 % in whole blood. However, the order of identified piRNAs differed (Table 2).

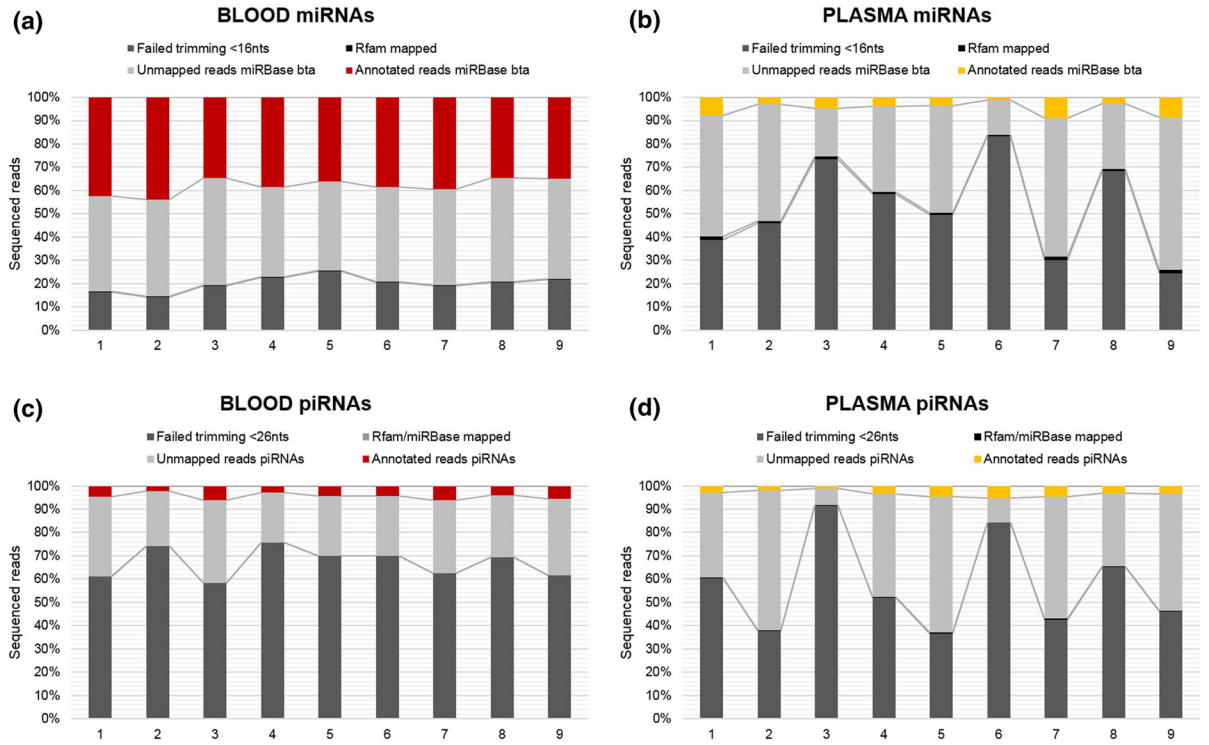


Fig. 1 Data evaluation of size-selected and stepwise mapped reads of all nine sequenced animals. **a**, **b** Data analysis pipeline for miRNAs. Image **a** displays the profile for miRNAs in whole blood and image [b] shows results for plasma miRNAs. **c**, **d** Data analysis pipeline for piRNAs. Image **c** shows the number of reads that were rejected in each step until piRNA annotation in whole blood and in plasma (**d**)

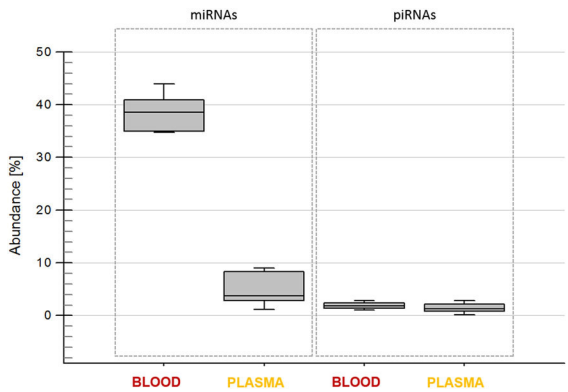


Fig. 2 Annotated miRNAs (*left panel*) as well as piRNAs (*right panel*) in PLASMA and BLOOD. Their abundance was calculated for each animal as percentage of annotated reads on the total sequenced reads and then averaged (n = 9)

By far the major component of the plasma miRNome was bta-miR-122 with a share of 54 %, followed by the two subtypes of bta-miR-423 that originate from opposite arms of the identical pre-miRNA (bta-miR-423-3p and -5p, together 17.4 %). The blood profile was as well dominated by one miRNA, which was bta-miR-451 that accounted for 27.5 %. With a fraction of 15 %, bta-miR-25 was ranked second while bta-miR-191 was ranked third with 5.1 % (Table 2; Fig. 3). The piRNome of plasma was mainly determined by piR-43772, piR-43771 and piR-60565 with a share of approximately 20 % each and 62.1 % if the Top 3 were put together. Likewise in whole blood, the piRNome was composed of three main piRNAs, however they differed from those in plasma: piR-31068, piR-35982 and piR-33151 were present with approximately 20 % each and 63.2 % in combination.

Spectrum of blood and plasma miRNAs and piRNAs

The evaluation of read-counts between miRNAs and piRNAs, which were present in whole blood and its

Table 2 Annotated read-counts [rpm] and proportional contributions [%] to the top 10 signature of the most abundant miRNAs and piRNAs in plasma and in blood

BLOOD miRNAs	Read-counts [rpm]	% Top 10	PLASMA miRNAs	Read-counts (rpm)	% Top 10	BLOOD piRNAs	Read-counts (rpm)	% Top 10	PLASMA piRNAs	Read-counts [rpm]	% Top 10
bta-miR-451	74,917	27.5	bta-miR-122	20,670	54.0	piR-35982	4304	24.3	piR-43772	3123	24.1
bta-miR-486	58,400	21.4	bta-miR-486	4378	11.4	piR-31068	4008	22.7	piR-43771	2812	21.7
bta-miR-25	40,788	15.0	bta-miR-423-3p	3608	9.4	piR-33151	2871	16.2	piR-60565	2116	16.3
bta-miR-92a	38,089	14.0	bta-miR-423-5p	3056	8.0	piR-43772	1805	10.2	piR-43786	1251	9.7
bta-miR-191	13,807	5.1	bta-miR-99a-5p	2333	6.1	piR-43771	1620	9.2	piR-31068	963	7.4
bta-let-7i	11,134	4.1	bta-miR-92a	1049	2.7	piR-33043	1340	7.6	piR-35982	954	7.4
bta-miR-185	9136	3.4	bta-miR-192	1002	2.6	piR-60565	855	4.8	piR-33151	767	5.9
bta-miR-339a	8965	3.3	bta-miR-3600	785	2.1	piR-36743	366	2.1	piR-31038	357	2.8
bta-miR-26a	8823	3.2	bta-miR-320a	756	2.0	piR-35413	290	1.6	piR-33082	345	2.7
bta-miR-21-5p	8364	3.1	bta-miR-140	656	1.7	piR-35284	223	1.3	piR-35284	267	2.1

Values were displayed as mean values over nine sequenced samples

cell-free byproduct plasma, by annotation and normalizing strategies allowed a comparison of the abundance of similarly present miRNAs (Fig. 4) and piRNAs (Fig. 5). As a cut-off, the average value of 50 read-counts has emerged in NGS data analysis to not include noise data. 58 plasma miRNAs (96.6 % of all plasma miRNA read-counts), 104 blood miRNAs (99.5 % of all blood miRNA read-counts) and 14 piRNAs in both whole blood and plasma passed the noise cut off. These 14 piRNAs accounted for 94.2 % of all read-counts in plasma and for 97.0 % in whole blood. The normalized read-count data sets were filtered for miRNAs and piRNAs, which were present in both plasma and whole blood. These matching miRNA pairs (n = 45 of 58) and piRNA pairs (n = 9 of 14) were taken for further analyses. The 45 matching miRNAs were represented by 92.3 % of all plasma read-counts and by 85.6 % of all whole blood read-counts. Regarding piRNAs, 81.2 % of plasma read-counts and 91.1 % of blood read-counts accounted for the nine matching pairs.

Figure 4 highlights that several matching miRNAs were highly expressed in whole blood but at lower levels in plasma. However, there were miRNAs in plasma whose expression ratio exceeded those of whole blood miRNAs, which was conspicuous especially for bta-miR-122. Besides the diversely abundant miRNAs, the majority of matching miRNAs that were annotated in this experiment were expressed in the same dimension in the two compared matrices. For piRNAs, a flic-flac effect could be observed for the 6 main matching pairs. Three highly expressed piRNAs in whole blood (piR-35982, piR-31068, piR-33151) were less abundant in plasma and vice versa for piR-43772, piR-43771 and piR-60565. Another three piRNAs that were existent with more than 50 read-counts in both plasma and whole blood were piR-35284, piR-33065 and piR-33082. Their number of reads were comparable.

Discussion

Pre-small RNA-Seq workflow

Since plasma is a low RNA concentration material, the analysis of circulating small RNAs is challenging. To

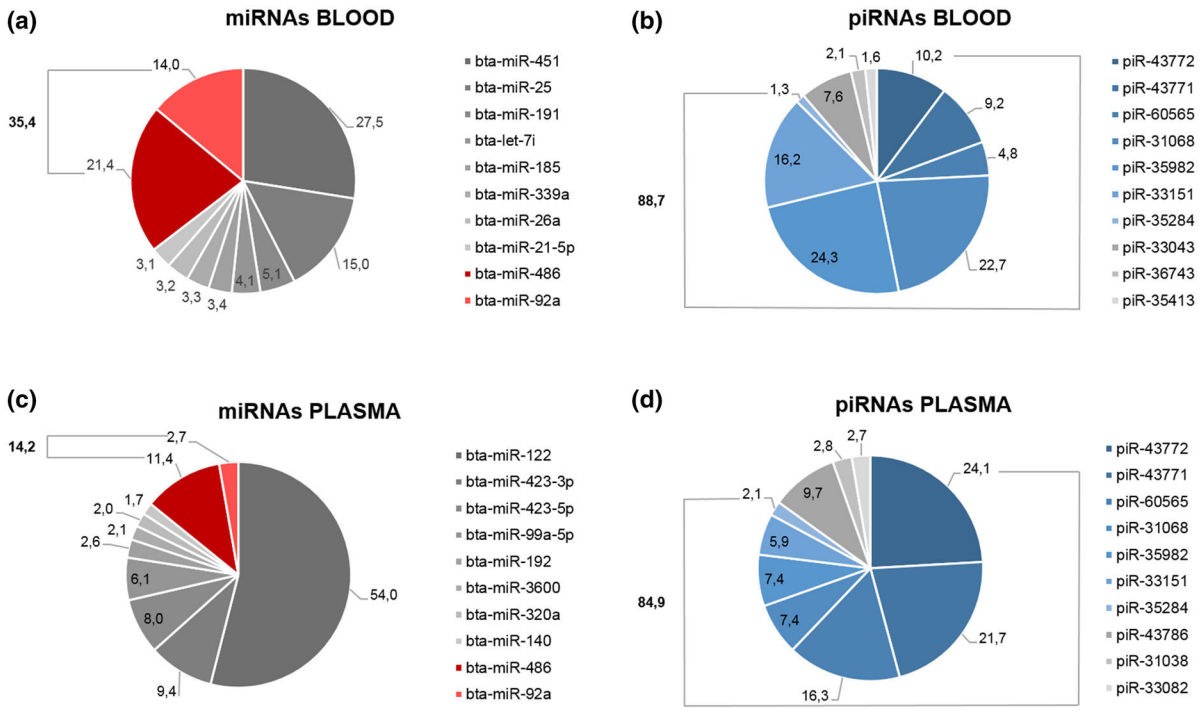


Fig. 3 The ten most abundant miRNAs and piRNAs in BLOOD (**a**, **b**) and PLASMA (**c**, **d**). MiRNAs that were found in both bio fluids were colored in *red* (pie charts **a**, **c**). Equally found piRNAs were marked in *blue colors* (pie charts **b**, **d**)

isolate sufficient RNA concentrations from plasma, an increased volume had to be used as starting material followed by an extensive extraction technique (Spornraft et al. 2014). Quality controls regarding the RNA integrity of samples is essential (Fleige et al. 2006; Bustin et al. 2009) prior to library preparation. It was not feasible in the case for circulating RNAs in plasma, as the concentration was too low to perform a RNA quality control analysis, e.g. using a Bioanalyzer 2100 RNA 6000 Nano Assay (Agilent). However, as shown elsewhere (Spornraft et al. 2014), the high-volume extraction method and data analysis pipeline resulted in high quality sequencing data from plasma. Contrarily, whole blood is rich of cellular components and therefore, the isolated RNA concentration allowed the RNA Integrity Number (RIN) measurement. The result was a RIN of 8.6 ± 0.3 that stands for high integrity of RNA samples.

Extraction of low amounts of RNAs from a complex matrix is difficult, especially when plasma had to be generated prior to RNA purification by separation of cellular parts. This processing step could lead to a prolonged exposure to RNases, which could promote degradation. In comparison, PAX-gene tubes contain stabilizers for RNAs in whole blood and could be used directly for RNA isolation without any additional manual steps. Therefore, extraction of total RNAs from whole blood resulted in increased RNA yields than isolation of plasma circulating total RNAs, which is reflected in higher RIN values and the more robust sequencing data (Fig. 1). These may also be the reasons why inter-individual differences appeared not as stable as in blood-based data.

Small RNA data analysis workflow

To generate read-count tables, in a first step, for sequencing essential adaptor sequences were trimmed from the 3′-end of sequenced reads and reads that did not show detectable adaptors were excluded. The data set was in a second step further depleted of sequenced reads that were too short to be miRNAs. This length cut-off of reads with <16 nt reduced the possibility of false positive mappings by degraded RNA fragments. Reads with appropriate sequencing lengths were mapped to Rfam database (Burge et al. 2013) to

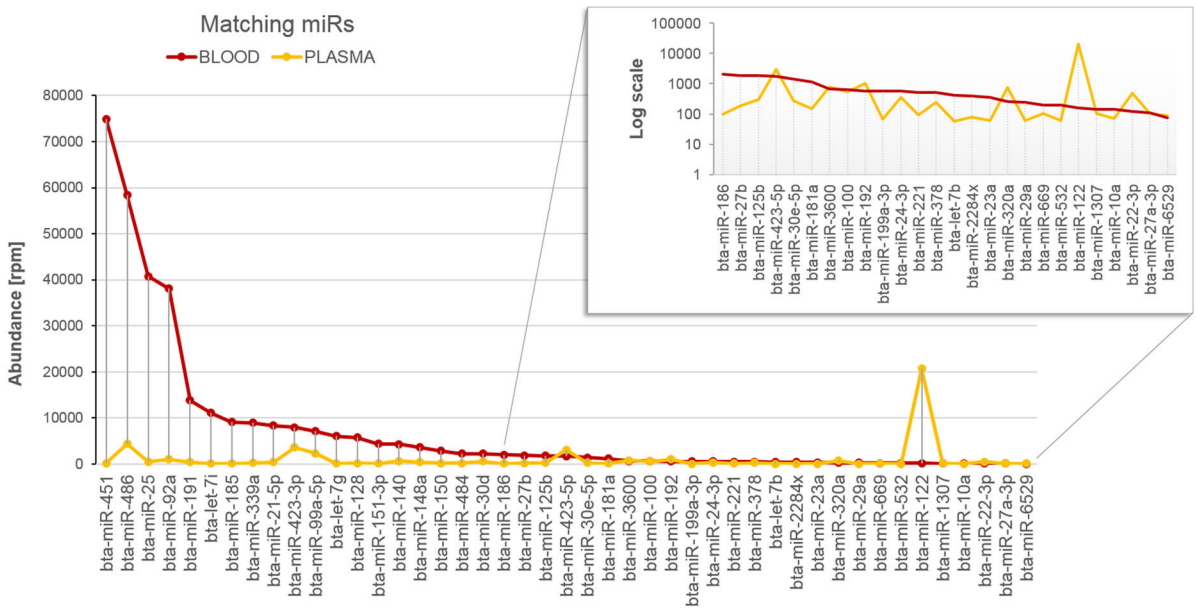


Fig. 4 Matching miRNAs. Zooming into the read-count lists allowed insights into the differential distribution of miRNAs and piRNA which were expressed higher than the noise ratio of 50 read-counts in blood (*red lines*) and in plasma (*yellow lines*)

exclude rRNA, tRNA, snRNA as well as snoRNA traits. Doing so, the initial data set was gradually diminished of complexity to improve specificity and accuracy in the final mapping to miRBase (Kozomara and Griffiths-Jones 2013) to better identify mature bta miRNAs. For the analysis of piRNAs, the length cut-off was set to 26 nt, to reduce the maximum number of reads that could lead to distortion from the start. The remaining sequences were aligned to Rfam database and to all subsets of the 206 species in miRBase database as an additional filtering step to exclude further reads that did not show piRNA-specific criteria. This sequential reduction of unsuitable sequences led to the highest achievable specificity and accuracy of the detected miRNAs and piRNAs, respectively. This pipeline used the most recent database versions and since all software tools are non-commercially available online, it gives the opportunity to replicate the presented data management and analyses.

Physiological differences between plasma and whole blood small transcriptomes

The encouraging results of research dedicated to find circulating miRNA biomarkers that exist today are very promising. miRNomes are profiled to find differences in expression ratios that are associated with specific diseases and their progression (Nair et al. 2014). Although this screening strategy works well in biomarker identification, not much is known about the origin of circulating small RNAs and their fate or destination in the circulatory system. There is still a lack of knowledge about the status quo of miRNA and piRNA distribution in healthy individuals. That applies in particular for piRNAs, as they are a younger class of investigated small RNAs and besides research in *Drosophila melanogaster*, the model organism where they were primarily described, and reproductive tissue, the organs where they were primarily found, fewer is examined regarding piRNA distribution in other tissues and species. Furthermore, it remains unclear which factors are regulating the concentration of small RNAs in plasma or if they are actively regulated at all. Can the circulation be seen as a repository, where miRNAs come and go, as all organs are permanently in direct contact to the blood stream? Additionally, it would be of interest if there is a basis level of certain small RNAs in a healthy body.

Our small RNA-Seq profiling of healthy randomly selected bovines indicated that there are core components of miRNAs and piRNAs in plasma and whole blood. The presented profiling enabled an insight into the basis miRNA and piRNA pattern of healthy

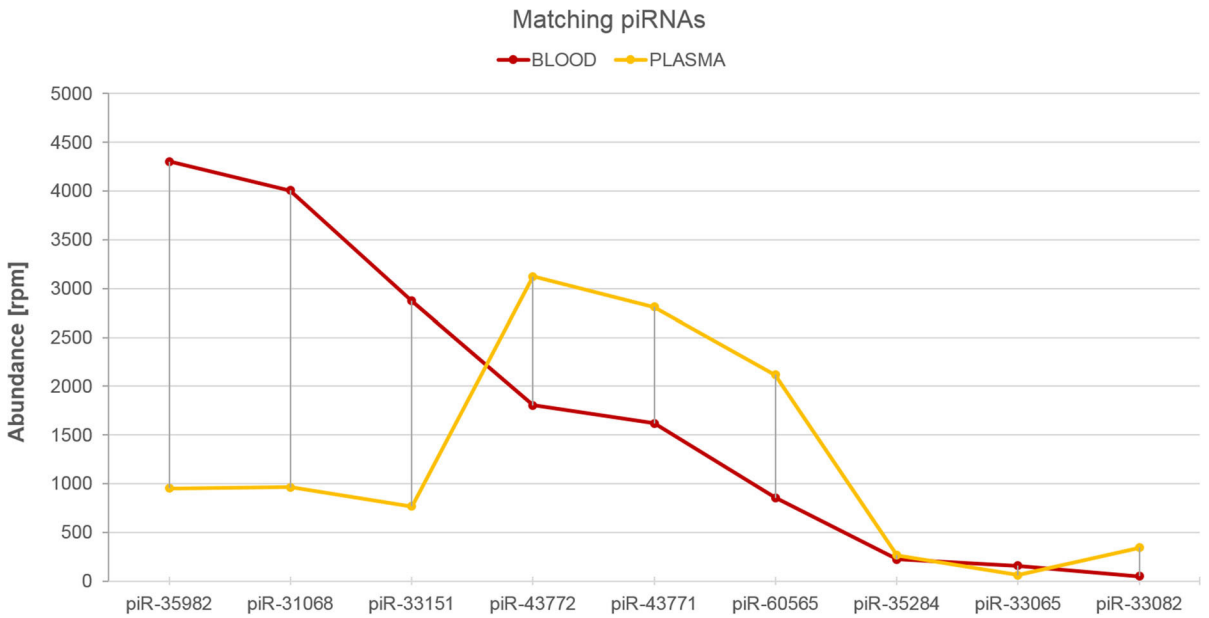


Fig. 5 Matching piRNAs. A flic-flac effect could be observed: three highly abundant piRNAs in whole blood were less expressed in plasma and vice versa. Three minor expressed piRNAs did not differ in their content in the two compared body fluids

bovines. As all the individuals showed an expression of these small RNAs at the same sampling time point, it may lead to the assumption that those play leading roles in either blood or plasma. It can be speculated that, based on the high abundance of specific miRNAs and piRNAs, the sequencing results mirrored a picture of the key small RNA modulators, at least of bovines at that specific age and gender. The analyses indicated that the sum of circulating small RNAs was not merely the product of disrupted blood cells, but other tissues than blood must contribute to the plasma small RNA pattern. Analysis of matching RNAs raised evidence that individual miRNAs and piRNAs cannot exclusively originate from cellular components of blood (e.g. RBCs, thrombocytes, leukocytes) as especially the content of bta-miR-122, piR-43772, piR-43771 and piR-60565 in plasma exceeded the levels of their counterparts in whole blood. Moreover, miR-122 is very well known to be liver-restricted (Long and Chen 2009; Lagos-Quintana et al. 2002) and in the initially investigated mouse model, Lagos-Quintana et al. (2002) found out that miR-122 accounted for 72 % of all investigated liver miRNAs. A profiling of a healthy human cohort (n = 18) revealed that the most abundant miRNA in the circulation was hsa-miR-122 (Duttagupta et al. 2011). These findings go in accordance with our data and emphasize the role of this miRNA. It is well known that miR-122 is relevant in maintenance of liver homeostasis and has essential metabolic, anti-inflammatory and anti-tumorigenic functions in liver (Hsu et al. 2012). Although miR-122 is a liver-related miRNA, recent research focusses other diseases in conjunction with it, e.g. sepsis (Wang et al. 2012). However, a detailed map of allocated miRNAs throughout an either healthy or dysfunctional organism is still missing and, regarding the highly expressed piRNAs, no information about their functions could be found in the literature.

MiR-451 and miR-144 are erythrocyte markers (Rasmussen et al. 2010) that were found in our blood data set at 19.6 and 0.02 % respectively. In plasma, miR-451 held a proportion of 0.4 % and of 0.005 % for miR-144. Furthermore, a commonly used normalization miRNA, miR-16 (Kirschner et al. 2011) was marginally expressed in both sample matrices. Other RBC markers are miR-486 and miR-92a that were present in our analyzed blood samples (15.3 and 10 %, respectively) and plasma samples (8.8 and 2.1 %, respectively). In summary, not all RBC miRNA markers (miR-16, miR-92a, miR-144, miR-451 and miR-486) that are commonly applied (Blondal et al. 2013; Pritchard et al. 2012) could be used as such in our data sets. Therefore, the usability of RBC miRNA markers in normalizing strategies, e.g. in RT-qPCR

profilings, need to be determined in each individual case. In a human plasma profiling study via small RNA-Seq (Williams et al. 2013), the miRNA pattern seems to be comparable: miR-486, miR-92a, miR-21, miR-25 and miR-22 were also found in our top 20. However, their most abundant miR-451, which comprises approximately 50 % of the RBC miRNA content in humans, was not as abundant in our dataset (0.4 %). An important fact that has to be considered in explaining alterations between miRNA spectrums in humans and bovines is the different erythropoiesis. In cattle, the complete erythropoiesis is located in the bone marrow and only mature erythrocytes are released into the circulation. Since erythropoiesis in humans is not limited to the bone marrow, erythrocyte precursor cells (reticulocytes) are present in blood. Their miRNAs could cause alterations in the human profile compared to bovines.

Conclusions

A comprehensive NGS approach was chosen for the screening of whole blood and plasma small RNA specimens in nine healthy bovines. The applied data evaluation pipeline offered to analyze data regarding quality and allowed annotation and generation of read-count tables by aligning sequencing reads to reference databases (Table 1). The combination of small RNA-Seq with the presented way of data processing is valuable for a holistic screening of the miRNA and piRNA spectrum to make statements about the composition of matrices (Fig. 1), abundance of miRNAs or piRNAs (Fig. 2) and compare characteristics (Figs. 4, 5).

The data strengthened evidence that blood cell miRNAs could contribute to the plasma circulating small RNA profile. The present specific blood cell markers that exist in literature were found in plasma, indicating that the plasma profile could be a down-sized reflection of the whole blood profile. However, data also gave an indication that certain miRNAs circulate in the bloodstream and originate from other sources than cellular blood components or exhibit increased stability in plasma (Figs. 4, 5).

Taken together, plasma is a technically challenging bio fluid for small RNA measurements for several reasons. First, the concentration of circulating small RNAs is very low and second, there are so far no standardized protocols for extraction, library preparation, sequencing itself and data analysis procedures nor are there comparable checklists for the accurate publication of small RNA-Seq experiments like there are the MIQE guidelines for qPCR experiments (Bustin et al. 2009). Moreover, contamination with blood cell derived RNAs could be a severe disturbance in the highly investigated research field of circulating small RNA biomarkers.

Nevertheless, this study could have laid a foundation for future comparative studies by providing information about the small RNA signature of bovine whole blood and plasma. Investigating the status quo of miRNomes and also piRNomes of healthy individuals provided a basis to explore the potentially altered footprints of other parameters, for example, comparison of different ages, gender, organs, physiological conditions and/or other species. Moreover, the presented read-count tables could serve as template for literature-based, targeted gene expression profiling studies via candidate gene RT-qPCR approaches.

The data pointed out that the circulating miRNA spectrum was more than just a by-product of dead cells that remained in extracellular space as it was hypothesized in the past years (Turchinovich et al. 2011). Finally, this work provided a footprint of miRNAs and piRNAs revealed by high-throughput sequencing that could facilitate further biomarker studies in bovine bio fluids, e.g. saliva or urine. Therefore, the door was opened to screen the transcriptome to detect small RNA changes as a result of ruminant maladies, track physiological changes (e.g. pregnancy) or proof the illegal administration of anabolic substances or growth promotors.

References

Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, Stirewalt DL, Tait JF, Tewari M (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci USA 108:5003–5008

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acid Res 41:D36–D42

Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, Dahlsveen IK (2013) Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods 59:S1–S6

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acid Res 41:D226–D232

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–622

Cheng J, Guo J, Xiao B, Miao Y, Jiang Z, Zhou H, Li Q (2011) piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clin Chim Acta 412:1621–1625

Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, Yu X, Xiao B, Wang W, Guo J (2011) Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. Clin Biochem 44:1050–1057

Duttagupta R, Jiang R, Gollub J, Getts RC, Jones KW (2011) Impact of cellular miRNAs on circulating miRNA biomarker signatures. PLoS One 6:e20769

Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW (2006) Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. Biotechnol Lett 28:1601–1613

Gou L, Dai P, Liu M (2014) Small noncoding RNAs and male infertility. Wiley Interdiscip Rev RNA 5:733–745

Hsu S, Wang B, Kota J, Yu J, Costinean S, Kutay H, Yu L, Bai S, La Perle K, Chivukula RR, Mao H, Wei M, Clark KR, Mendell JR, Caligiuri MA, Jacob ST, Mendell JT, Ghoshal K (2012) Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. J Clin Invest 122:2871–2883

Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, Reid G (2011) Haemolysis during sample preparation alters microRNA content of plasma. PLoS One 6:e24145

Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acid Res 42:D68

Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T (2002) Identification of tissue-specific microRNAs from mouse. Curr Biol 12:735–739

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Long J, Chen H (2009) Identification and characteristics of cattle microRNAs by homology searching and small RNA cloning. Biochem Genet 47:329–343

Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M (2008) Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA 105:10513–10518

Nair VS, Pritchard CC, Tewari M, Ioannidis, John PA (2014) Design and analysis for studying microRNAs in human disease: a primer on -omic technologies. Am J Epidemiol 180:140–152

Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, Tait JF, Tewari M (2012) Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res (Phila) 5:492–497

Qi P, Du X (2013) The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol 26:155–165

Ramkissoon SH, Mainwaring LA, Ogasawara Y, Keyvanfar K, McCoy JP, Sloand EM, Kajigaya S, Young NS (2006) Hematopoietic-specific microRNA expression in human cells. Leuk Res 30:643–647

Rasmussen KD, Simmini S, Abreu-Goodger C, Bartonicek N, Di Giacomo M, Bilbao-Cortes D, Horos R, von Lindern M, Enright AJ, O'Carroll D (2010) The miR-144/451 locus is required for erythroid homeostasis. J Exp Med 207:1351–1358

Spornraft M, Kirchner B, Haase B, Benes V, Pfaffl MW, Riedmaier I (2014) Optimization of extraction of circulating RNAs from plasma—enabling small RNA sequencing. PLoS One 9:e107259

Turchinovich A, Weiz L, Langheinz A, Burwinkel B (2011) Characterization of extracellular circulating microRNA. Nucleic Acid Res 39:7223–7233

Vasilatou D, Papageorgiou S, Pappa V, Papageorgiou E, Dervenoulas J (2010) The role of microRNAs in normal and malignant hematopoiesis. Eur J Haematol 84:1–16

Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol 13:423–433

Wang K, Zhang S, Weber J, Baxter D, Galas DJ (2010) Export of microRNAs and microRNA-protective protein by mammalian cells. Nucleic Acids Res 38:7248–7259

Wang H, Zhang P, Chen W, Feng D, Jia Y, Xie L (2012) Serum microRNA signatures identified by Solexa sequencing predict sepsis patients' mortality: a prospective observational study. PLoS One 7:e38885

Weber JA, Baxter DH, Zhang S, Huang DY, How Huang K, Jen Lee M, Galas DJ, Wang K (2010) The microRNA spectrum in 12 body fluids. Clin Chem 56:1733–1741

Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T (2013) Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc Natl Acad Sci USA 110:4255–4260

Appendix III

Food Additives & Contaminants: Part A, 2014
http://dx.doi.org/10.1080/19440049.2014.886341

Taylor & Francis
Taylor & Francis Group

Identification of a potential gene expression biomarker signature in bovine liver to detect the abuse of growth promoters

Irmgard Riedmaier*, Melanie Spornraft and Michael W. Pfaffl

Physiology Weihenstephan, and ZIEL Research Center for Nutrition and Food Sciences, Technical University Munich, Freising, Germany

(*Received 21 November 2013; accepted 14 January 2014*)

The misuse of anabolic agents in animal husbandry is a ubiquitous problem. The ban of growth promoters in food producing animals in the European Union is well controlled, but there are still application regimes, such as new designed drugs or hormone cocktails, that are difficult to detect. Therefore, the idea of identifying molecular biomarkers that are based on the physiological effect of treatment has come into focus. In a previous study we identified mRNA biomarker candidates in liver samples that enable the separation of untreated animals from animals treated with a combination of androgens plus estrogens. In the present study those candidates were validated in calves treated with a combination of progesterone plus estradiol or clenbuterol, respectively. Therefore, the candidate genes were quantified in liver samples of those calves via RT-qPCR. Using dynamic principal component analysis (PCA), a signature of 11 genes could be selected. This set of genes enabled the separation of treated and control animals independent of the applied drug. Additional quantification of these genes in a set of control samples from another animal trial resulted in a PCA that also showed a separation of those samples from treated animals. This study showed that gene expression biomarkers have a high potential to enable the detection of physiological changes caused by the application of growth-promoting substances independent of the given drug, but further studies are necessary to broaden the spectrum of anabolic substance groups for which those biomarker candidates can be used.

Keywords: steroids; clenbuterol; gene expression; biomarker signature

Introduction

The identification of molecular biomarkers in different research fields has come into focus during the last years. Mainly in the field of clinical diagnostics, research has focused on the identification of biomarkers for different diseases on different molecular levels, such as the genome, the epigenome, the transcriptome, the proteome or the metabolome (Riedmaier & Pfaffl 2013). But biomarker research is not limited to molecular medicine, but is also done in other fields where physiological changes can be monitored; the search for transcriptional biomarkers is promising.

The misuse of anabolic substances in sports and animal husbandry is an ubiquitous problem. In sports, anabolic substances are applied to increase muscle mass and/or increase sporting performance. In order to ensure fairness in sports the misuse of anabolic substances is effectively controlled. The use of growth-promoting agents is also attractive in animal husbandry to increase muscle mass and to decrease fat content (Moran et al. 1991; Lange et al. 2001). Whereas the use of specific growth promoters is allowed in different countries, such as, for example, the United States, Canada and South Africa, the use of any anabolic substances is forbidden in the European Union due to proven side-effects of residues of these substances in meat for the consumer (Directives 96/22/EC, 96/23 EC, 2008/97/EC). This ban is strictly controlled by detecting hormone residues in different matrices. But the detection of natural steroids, new designed drugs or substance cocktails is very difficult (Cantiello et al. 2007), even if the current detection methods are getting more and more sensitive. Therefore, the search for alternative methods to detect the illegal use of new anabolic practices is important. The biomarker approach is very promising in this area (Riedmaier, Becker et al. 2009; Pinel et al. 2010). Independent of which growth promoter is applied, the physiological effect is the same: increased muscle growth and mobilisation of fat stores. Therefore the approach to detect effects based on physiological marker molecules which are independent of the applied substance and of the way of application is very promising (Ludwig et al. 2013). There are already different publications available that deal with the detection of such biomarkers at the transcriptomic, proteomic or metabolomic level, but identified biomarkers are up to now still substance specific (Riedmaier, Tichopad, et al. 2009; Riedmaier et al. 2011, 2012; Becker, Riedmaier, Reiter, Tichopad, Groot, et al. 2011; Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011; Dervilly-Pinel et al. 2012; Ludwig et al. 2012, 2013). The goal would be to identify markers that are independent of breed, age, gender, applied anabolic substances or even species.

*Corresponding author. Email: irmgard.riedmaier@wzw.tum.de

Most of those studies work with the so-called targeted approach, which means that only a limited number of biomarker candidates are quantified. On the contrary is the untargeted approach, where the expression of all genes, proteins or the presence of metabolites in a specific matrix or tissue is monitored in a holistic way (Pinel et al. 2010) using so-called 'omic' technologies. What is common in most of the published studies is that not the detection of a single biomarker, but rather a signature of markers is the most promising way. Biostatistical tools for pattern recognition like principal component analysis (PCA) or hierarchical cluster analysis (HCA) enable the separation of treated and untreated animals (Riedmaier, Becker, et al. 2009).

In a former study, we investigated the potential of analysing gene expression changes in bovine liver caused by treatment with trenbolone acetate plus estradiol using the targeted approach and could identify a set of 15 potential biomarker candidates (Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011). In order to find more candidate genes, we evaluated the potential of RNA-sequencing technology to detect new gene expression biomarker candidates in the same liver samples. Here we could identify a set of 20 new candidate genes that allowed the separation of heifers that were treated with a combination of trenbolone acetate plus estradiol from untreated heifers using PCA and HCA. Those biomarker candidates could then be validated in boars where also a combination of trenbolone acetate plus estradiol was applied, and even in calves treated with a cocktail of androgens plus estrogens via a pour-on application, which indicated that the identified biomarker signature is independent of age, species and gender (Riedmaier et al. 2012). In the present study those candidate genes should be validated in liver samples from another trial, where veal calves where treated with a combination of progesterone plus estradiol benzoate or with clenbuterol, respectively. Additionally, potential candidate genes from the targeted approach were also quantified. As in the targeted approach studies the receptors of the substances have shown to be regulated by their ligand the receptors for the applied substances were also quantified.

Materials and methods

Animal experiments

Treatment experiment

Twenty-one male, 6-month-old Holstein Friesian calves were separated into three groups of seven animals each. One group was not treated with anabolic agents, serving as control group. The animals of the second group were treated with Component E-C® (100 mg progesterone plus 10 mg estradiol benzoate; IVY Animal Health, Overland Parks, KS, USA) by implantation into the middle third of the pinna of the ear; and the animals of the third group were treated with 10 µg clenbuterol (kindly provided by Boehringer Ingelheim, Ingelheim, Germany) per day via oral application. To test if treatment was successful for anabolic purpose, the weights of the animals were measured before treatment start and at days 17 and 34 of the trial.

After a treatment time of 36 days, animals were slaughtered, liver samples were collected, directly frozen in liquid nitrogen and stored at –80°C until analysis.

The animal attendance was done according to practice.

Additional control samples

To obtain a bigger group of untreated controls, 42 liver samples of male Holstein Friesian calves that were also 6 months old from an animal trial that was previously performed and where the animals were not treated with any growth-promoting agents were also analysed (Masanetz et al. 2011). The identified biomarker signature was also quantified in those samples.

Gene expression analysis

Gene expression analysis was performed according to the minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines (Bustin et al. 2009).

RNA was extracted from liver samples using the miRNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA purity was calculated using the $OD_{260/280}$ ratio. RNA quality was determined using the Eukaryotic total RNA Nano Assay on the 2100 Bioanalyzer (Agilent Technology, Palo Alto, CA, USA).

For cDNA synthesis, constant amounts of 500 ng integer total RNA were reverse transcribed, as already described (Riedmaier et al. 2012).

To analyse the expression of candidate genes, qPCR analysis was performed using the iQ5 (Bio-Rad, Munich, Germany). Quantitative real-time PCR was performed using the SensiFAST™ SYBR & Fluorescein Kit (Bioline GmbH, Luckenwalde, Germany) by a standard protocol recommended by the manufacturer.

The master mix was prepared as follows: for one sample it is 7.5 µl SensiFast™ SYBR & Fluorescein Mix (2×), 0.6 µl forward primer (10 pmol μl^{-1}), 0.6 µl reverse primer (10 pmol μl^{-1}) and 4.8 µl RNase free water. For qPCR analysis, 1.5 µl cDNA were added to 13.5 µl master mix. qPCR was performed in 96 well plates (4titude, Wotton, UK).

The following real-time PCR cycling protocol was employed for all investigated factors: denaturation for 2 min at 95°C, 40 cycles of a three-segmented amplification and quantification programme (denaturation for 5 s at 95°C, annealing for 10 s at primer specific annealing temperature listed in Table 1 and elongation for 5 s at 72°C), a melting step by slow heating from 60°C to 95°C with a dwell time of 10 s and continuous fluorescence measurement. Cycle of quantification (C_q) and melting curves were acquired by

Table 1. Primer pairs used for RT-qPCR analysis.

Acc no.	Gene		Sequence [5' → 3']	Product length (bp)	T_M (°C)
NM_174231	ADRB2	for	ATT TCA GGA TTG CCT TCC AG	120	64
		rev	ATC CAC TCT GTT CCC CTG TG		
NM_001077933	AK3L1	for	GTTGAAACATAGCACCAGGAGG	197	64
		rev	TGGAGCACGGATACAGAGAAG		
NM_001077933.1	AK4	for	TCATGCTGTTGGAGTTGGAG	247	66
		rev	TCACCAGTGACGTCATCCAT		
NM_001037480	APOA4	for	TTTCTGAAGGCTGTGGTCCTG	187	60
		rev	CTTGGAAGAGGGTGTTGAGC		
NM_174242	APOC2	for	GGG TTT CTC ATC CTC CTG GT	215	63
		rev	AAT CCC TGC ATA GGT GGT CAC		
AY862875	AR	for	CCT GGT TTT TCA ATG AGT ACC GCA TG	172	62
		rev	TTG ATT TTT CAG CCC ATC CAC TGG A		
NM_001014391	CLDN4	for	CTTCTTCCATCCTTCCTGCTC	164	64
		rev	TCTAAACCTGTCCGTCCACTC		
NM_001192597	CUX2	for	AGCGATTTGATCTACGGCGAC	243	64
		rev	GGCCTCCTGACTTCTCTTAC		
NM_001046551	CXCL10	for	GCCCACGTGTCGAGATTATT	140	64
		rev	CTCTTTCCGTGTTCGAGGAG		
NM_174530	CYP2E1	for	GACCTCCTTCCCTCCAATC	230	64
		rev	ACACACACCCGCTTTCCTG		
AF177936	ERa	for	AGG GAA GCT CCT ATT TGC TCC	233	60
		rev	GGT GGA TGT GGT CCT TCT C		
NM_174051	ERb	for	TTA GCC ATC CAT TGC CAG CC	248	64
		rev	GCC TTA CAT CCT TCA CAC GAC		
NM_001083444	FADS2	for	AGACCGCTGAGGACATGAAC	161	64
		rev	AGAGGTAGCAAGGACGAAGG		
NM_001037613.1	FBXO39	for	CAAGTGCCACATTCATGACC	107	62
		rev	CCCGCTCAAAGAAGAAGTTG		
NM_001076400	GDPD1	for	TGGTGAAGCAGTACAGACGAG	139	64
		rev	AACAGGCCAAGAACGAGCAG		
NM_001080354	GSTM1	for	ACATTCACCGCATATTTGAGC	172	64
		rev	CCAGAGCACTACTTGTTGC		
NM_173917	HBB	for	ACACTTGCTTCTGACACAACC	126	64
		rev	TCACCACCAACTTCATCCAC		
NM_001014389	Histon	for	ACTGCTACAAAAGCCGCTC	233	62
		rev	ACTTGCCTCCTGCAAAGCAC		
NM_001045883.1	HMGCoAS	for	GAT GGT CGC TAT GCA CTG GT	246	60
		rev	GCC CTC TCT CGA GGA CCA GA		
NM_174097	HOPX	for	TGAGTGCAGATCCGTCACAG	130	64
		rev	GGGACATAACAACGCTACACTG		
NM_001077828	IGF-1	for	CAT CCT CCT CGC ATC TCT TC	238	62
		rev	CTC CAG CCT CCT CAG ATC AC		
BC126514	IGF2	for	ACC CTC CAG TTT GTC TGT GG	166	54
		rev	ACA CAT CCC TCT CGG ACT TG		
AJ320235	IRa	for	TCC TCA AGG AGC TGG AGG AGT	89	62
		rev	TTT CCT CGA AGG CCT GGG GAT		
AJ320235	IRb	for	TCC TCA AGG AGC TGG AGG AGT	111	60
		rev	TAG CGT CCT CGG CAA CAG G		
NM_174366	ISG15	for	CAGAACCCACGGCCATGG	130	64
		rev	GGCACATTGATCTTCTGGGC		
NM_173941	MX2	for	CACCTACCGCAACATTACG	230	64
		rev	GTGCCTTGATCTGCAGTC		
NM_001046269	PON1	for	CGGCATTAGCACATTCACAG	176	64
		rev	CCACAGCGACAAGGTCATT		
NM_001104967	PRODH2	for	TGGGGTGAAGTTGGTACGAG	189	64
		rev	TTGTGGGAAGCCACCATGAG		
NM_001102326	SECTM1	for	TCCAATGCCTTCTCCCACATC	194	64
		rev	TGTAGACTCCACTTGTACTGCC		
NM_001102054	SERPINI2	for	TGCCAATGCCCTCTACCTTC	221	64
		rev	TTACCAGAACCAGCCGAGTC		

(*continued*)

Table 1. Continued.

Acc no.	Gene		Sequence [5' → 3']	Product length (bp)	T_M (°C)
NM_001098858	SHBG	for	ACT TGG GAT CCA GAG GGA GT	188	62
		rev	TCC CCA TGG ATC TTC ACT TCc		
NM_001082610	SOD3	for	CACTCCTGAAACCATGTTGCC	187	64
		rev	CATCATCTCCTGCCAGATCTC		
BT021798	TAT	for	ACC CTT GTG GGT CAG TGT TC	167	60
		rev	ACA GGA TGG GGA CTT TGC TG		
NM_214230.1	TNC	for	GAGAAGACTGCTCCCAGGTG	168	66
		rev	CTCCAGGGACTCTGAACTGC		
NM_174814.2	YWHAZ	for	CAG GCT GAG CGA TAT GAT GAC	141	60
		rev	GAC CCT CCA AGA TGA CCT AC		

using the iQ5 Optical System software 2.1 (Bio-Rad). Only genes with clear melting curves and single specific product peaks were taken for further data analysis. Samples that showed irregular melting peaks were excluded from the quantification procedure.

Candidate genes that were quantified have already been shown to be biomarker candidates for the use of anabolic substances in bovine liver (Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011; Riedmaier et al. 2012). A list of quantified genes including the corresponding primer pairs is shown in Table 1.

Data analysis

Weight gain

To determine significant differences in weight gain between control and both treatment groups, an unpaired *t*-test was applied using Sigma Plot 12.5 Software (SPSS Inc., Chicago, IL, USA). *P*-values < 0.05 were regarded as statistically significant.

Gene expression

Gene expression data analysis was performed as described by Riedmaier et al. (2012).

To visualise the multivariate response of the selected classifiers to the treatment, dynamic PCA was employed using GenEx v.5 (MultiD Analyses AB, Gothenburg, Sweden). Dynamic PCA is a useful tool to select a biomarker pattern out of a number of genes, based on the level of significance or the distance of expression level and the variance of expression, enabling the best visualisation of sub-clusters.

The selected sub-cluster of genes was also quantified in the additional control liver samples and a PCA analysis using those samples and the samples obtained from treated animals was performed.

Results

Weight gain

In order to determine treatment performance, animals were weighed before treatment started and at days 17 and 34 of the trial. At treatment start there was no significant difference in weight between controls and the steroid group ($p = 0.4$) and controls and the clenbuterol-treated group ($p = 0.6$). After 17 days of treatment, weight gain in the treatment groups was significantly higher compared with the control groups ($p = 0.028$ for the steroid group and $p = 0.019$ for the clenbuterol group). After 34 days of treatment, weight gain in the treatment groups was again significantly higher compared with the control groups ($p = 0.0005$ for the steroid group and $p = 0.004$ for the clenbuterol group). Regarding the whole trial, weight gain between days 0 and 34 was determined. Weight gain over the whole trial was significantly increased in the treatment groups compared with the control group ($p = 0.0015$ for the steroid group and $p = 0.0021$ for the clenbuterol group). This indicates that the anabolic treatment was successful. An overview of weight gain development is given in Figure 1.

RNA quantity and quality

RNA quantity and purity was measured using the NanoDrop Photometer (PeqLab). Mean RNA concentration was 1323.6 ± 561.1 ng μl^{-1}. As an indicator for RNA purity, the $OD_{260/280}$ ratio was determined, where a ratio >1.8 is generally assumed as adequate for gene expression experiments. The $OD_{260/280}$ ratio was 2.08 ± 0.03.

RNA integrity was determined using the Eukaryotic total RNA Nano Assay on the 2100 Bioanalyzer (Agilent Technologies). The mean RNA integrity number (RIN) value of the liver samples from the treatment trial was 8.3 ± 0.5 and from the additional control samples the mean RIN was 7.2 ± 0.4, indicating intact RNA.

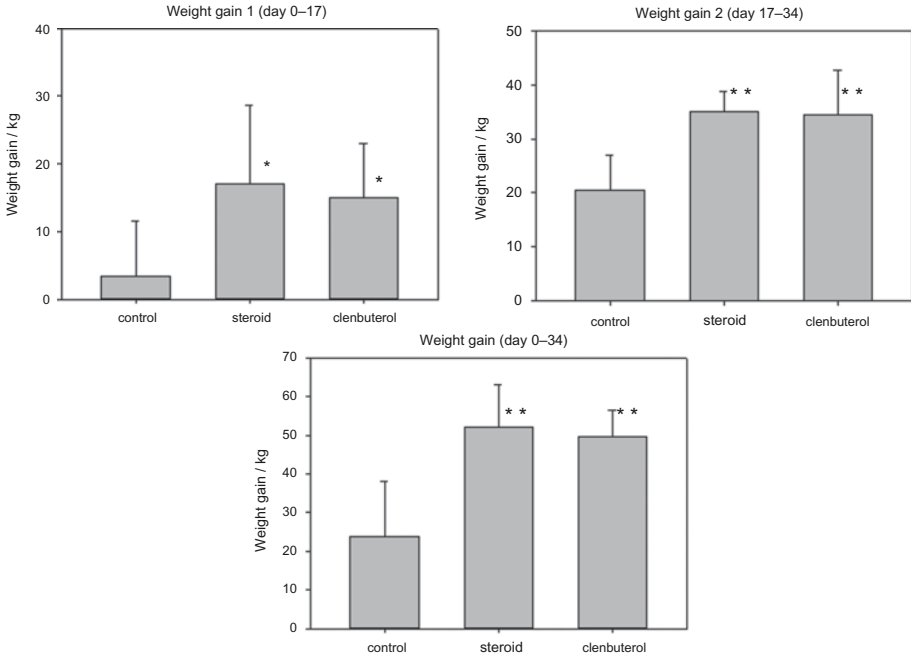


Figure 1. Weight gain of the treatment groups during the trial.

RT-qPCR analysis

In the steroid-treated group, seven genes showed significant changes compared with untreated controls and in the clenbuterol-treated group six genes were significantly regulated. Three of those genes could be shown to be influenced by both treatments. Steroid treatment up-regulated the expression of the adrenergic beta-2 receptor (ADRB2) by 2.11-fold ($p = 0.026$), of the insulin receptor beta (IRβ) by 1.61-fold ($p = 0.001$), of paraoxonase 1 (PON1) by 1.63-fold ($p = 0.014$) and of the serpin peptidase inhibitor 2 (SerpinI2) by 34.09-fold ($p = 0.007$). Clenbuterol treatment resulted in an up-regulation of adenylate kinase 3 (AK3L1) by 4.75-fold ($p = 0.008$) and tyrosine aminotransferase (TAT) by 1.41-fold ($p = 0.036$) and down-regulated the expression of the F-box protein 39 (FBXO39) by 0.36-fold ($p = 0.030$). Both treatments up-regulated the expression of apolipoprotein A4 (APOA4) – steroids by 4.94-fold ($p = 0.024$) and clenbuterol by 5.46-fold ($p = 0.043$) – and of 3-hydroxy-3-methylglutaryl-CoA synthase (HMGCoAS) – steroids by 1.86-fold ($p = 0.037$) and clenbuterol by 2.05-fold ($p = 0.029$) – and down-regulated the expression of the estrogen receptor beta (ERβ) – steroids by 0.40-fold ($p = 0.010$) and clenbuterol by 0.47-fold ($p = 0.022$). A summary of all analysed genes including-fold changes and p-value is given in Table 2.

Dynamic PCA was applied to see which set of genes results in the best separation of treated and untreated animals. Black dots represent control samples, grey triangles steroid samples and silver diamonds represent samples from clenbuterol-treated animals. Dynamic PCA was performed with all measured genes in all three treatment groups, whereat the gene signature of ADRB2, AK3L1, APOA4, APOC2, ERβ, FBXO39, GDPD1, HMGCoAS, IRβ, PON1 and TAT showed the best results (Figure 2A). It was also tested if this set of genes is also suitable to separate control and steroid group and control and clenbuterol group separately. In both PCAs a clear separation between control and treated animals could be observed (Figures 2B and C).

The expression of the selected subgroup of genes was also analysed in the additional control liver samples ($n = 42$) and a PCA analysis using those samples and the samples obtained from treated animals was performed (Figure 3). White dots represent control samples, grey triangles steroid samples and silver diamonds represent samples from clenbuterol-treated animals. Here a separation between control and treated samples is also visible.

Discussion

The identification of molecular biomarkers on different levels using “omic” technologies is a highly discussed approach in order to develop additional detection methods for the abuse of growth promoting agents in animal husbandry. Even if the contemporary MS-based techniques that are used for residue screening enable a rapid and sensitive analysis, unknown compounds cannot be detected and so the development of additional effect based methods could help to fill in this detection gap.

Table 2. Overview of *p*-values and *x*-fold regulations of all quantified genes.

	Component E-C		Clenbuterol	
Gene	*p*-value	Fold regulation	*p*-value	Fold regulation
ADRB2	**0.026**	**2.11**	0.059	1.90
AK3L1	0.287	1.75	**0.008**	**4.75**
AK4	0.463	0.93	0.064	1.83
APOA4	**0.024**	**4.94**	**0.043**	**5.46**
APOC2	0.289	1.42	0.154	1.98
AR	0.818	0.96	0.979	1.06
CLDN4	0.600	0.93	0.428	1.65
CUX2	0.472	1.13	0.654	1.55
CXCL10	0.832	1.08	0.509	1.27
CYP2E1	0.332	0.88	0.433	0.92
ERα	0.866	1.01	0.191	1.43
ERβ	**0.010**	**0.40**	**0.022**	**0.47**
FADS2	0.723	1.35	0.061	0.45
FBXO39	0.085	0.49	**0.030**	**0.36**
GDPD1	0.168	0.74	0.176	0.74
GSTM1	0.827	1.17	0.920	1.41
HBB	0.594	0.91	0.985	1.31
HMGCoAS	**0.037**	**1.86**	**0.029**	**2.05**
HOPX	0.185	0.70	0.858	1.60
IGF-1	0.435	1.61	0.877	1.39
IGF2	0.448	1.25	0.812	1.25
IRα	0.870	1.01	0.510	1.44
IRβ	**0.001**	**1.61**	0.307	1.40
ISG15	0.684	2.47	0.848	1.03
MX2	0.901	2.90	0.278	0.75
PON1	**0.014**	**1.63**	0.192	1.44
PRODH2	0.376	0.85	0.755	1.45
SECTM1	0.934	1.62	0.802	1.11
SerpinI2	**0.007**	**34.09**	0.537	1.40
SHBG	0.157	1.35	0.348	1.32
SOD3	0.490	2.50	0.868	1.02
TAT	0.519	1.14	**0.036**	**1.41**
TNC	0.203	1.94	0.616	1.66

Note: Significantly regulated genes are emboldened.

In this study, previously determined gene expression biomarker candidates for the application of a combination of androgens plus estrogens that were detected using RNA-sequencing technology (Riedmaier et al. 2012) were validated in animals treated with a combination of progesterone plus estradiol benzoate or clenbuterol, respectively. Additionally, genes that were also found to be potential candidate genes using the targeted approach were also quantified (Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011). Dynamic PCA was applied to choose a set of genes that shows the best separation between treated and untreated animals. The selected set of genes consists of all genes that were significantly regulated between the control group and one of the treatment groups (ADRB2, AK3L1, APOA4, ERβ, FBXO39, HMGCoAS, IRβ, PON1 and TAT), except of SerpinI2. Regarding the Cq values of this gene, it can be observed that the variance of expression within the groups is very high. That is why SerpinI2 has been excluded by the dynamic PCA algorithm. There are also two genes – GDPD1 and APOC2 – that are not significantly regulated but were included in the gene set anyway. Regarding the fold regulation values of those genes, it is obvious that GDPD1 is down-regulated in both groups by 0.74-fold which counts for a down-regulation of approximately 30%. Regarding the regulation values of the other genes, it is not often the case that fold regulations in both groups are that similar, what might be the reason that GDPD1 is included in the selected gene set. Another gene was not significantly regulated was APOC2, but was regulated nearly two-fold in the clenbuterol-treated group and 1.42-fold in the steroid-treated group. The regulation is not significant, but the relatively high fold regulation values compared with the other genes might be the reason for including this gene in the selected gene set that supports the separation of control and treated animals best. Regarding the source of choosing the genes as potential candidate genes, it could be observed that three genes were selected from the targeted approach (Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011), four genes were selected using RNA

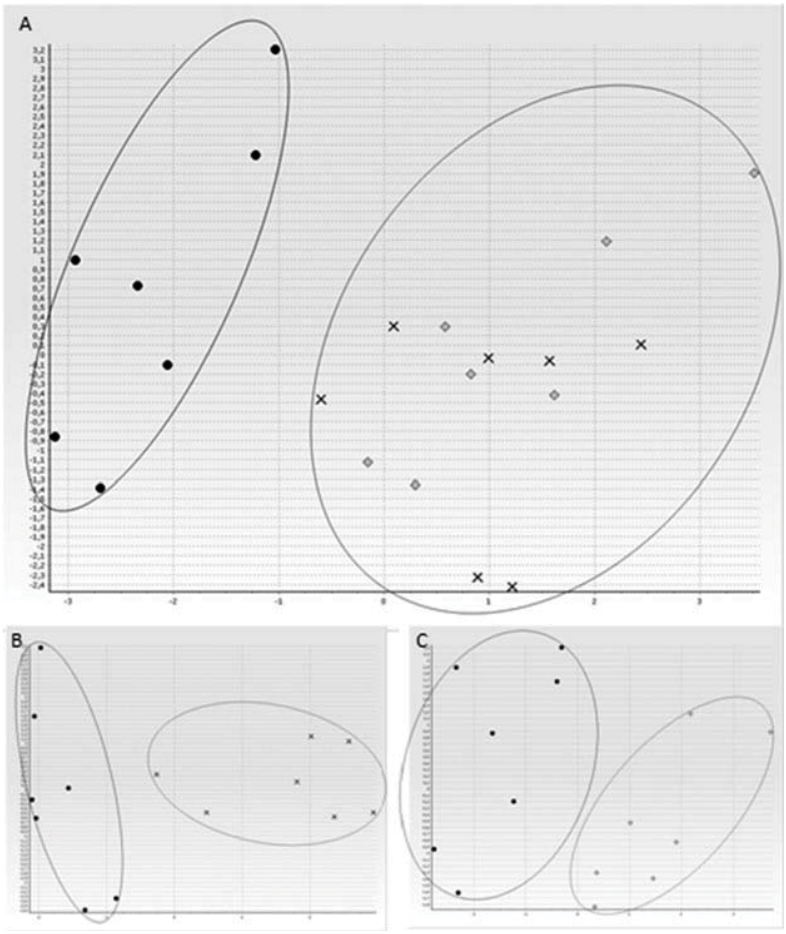


Figure 2. Principal component analysis (PCA). Figure 2A represents the PCA including all three treatment groups. Figures 2B and C show PCA analyses including the control group and the steroid group (B) and the control group and the clenbuterol group (C) separately. Animals of the control group are represented by black dots; animals of the steroid-treated groups are represented by grey triangles; and clenbuterol-treated animals are shown by silver diamonds. Ellipses are drawn to separate the treatment groups.

sequencing (Riedmaier et al. 2012), and two genes encode for receptors for the applied substances. This indicates that using RNA sequencing in combination with the targeted approach is the most promising way for candidate gene research.

Including only one treatment group into the PCA analysis using the same set of genes, a separation between control and treated animals is visible for both anabolic substances.

This set of genes was selected in cattle treated with a combination of androgens plus estrogens (Becker, Riedmaier, Reiter, Tichopad, Pfaffl, et al. 2011; Riedmaier et al. 2012) and could here be validated in samples obtained from animals that were treated with a combination of a gestagen plus an estrogen, which indicates that this set of genes is sufficient to separate animals treated with steroid hormones from non-treated animals. The fact that this gene set is also sufficient to separate animals treated with clenbuterol, a beta-adrenergic substance that acts via a totally different mechanism than steroid hormones (Riedmaier, Becker, et al. 2009), lets us assume that this way of biomarker detection is really based on the physiological, anabolic effect and independent of the kind of molecular interaction of the applied substance. The development of such a biomarker-based screening method presupposes a high number of untreated samples obtained from animals of different age, breed, feeding regimes, etc. (Riedmaier, Becker, et al. 2009). To test if the selected set of genes is also sufficient to separate the treated animals from a bigger pool of untreated animals, the expression of those genes was quantified in liver samples from another feeding trial with animals of nearly the same age and the same breed that were guaranteed not to be treated with any growth promoting agents. PCA showed that the treated animals still separate from the untreated individuals, except of one steroid-treated animal which would be a false-negative individual in a routine test. But as in one farm not only one single animal is tested and the other steroid-treated

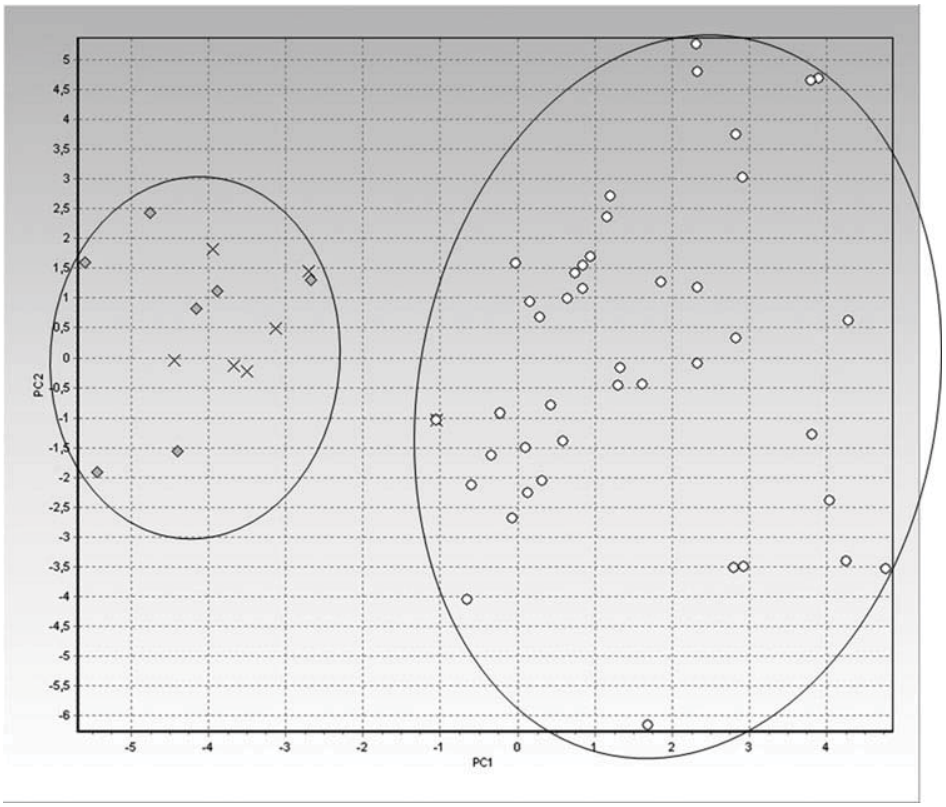


Figure 3. Principal component analysis (PCA) including both treatment groups and the 42 additional control animals. White dots represent untreated animals; animals of the steroid-treated groups are represented by grey triangles; and clenbuterol-treated animals are shown by silver diamonds.

animals clearly separate from the controls, it would be obvious that animals in that farm are treated with growth promoters. These results indicate that this gene signature is a good candidate gene pool for further tests.

But as the biomarker pattern chosen here by the dynamic PCA algorithm is not the same set of genes published by Riedmaier et al. (2012), animals of different breed age and gender should be treated with other anabolic components or hormone cocktails to test if the presented set of candidate genes is stable.

Conclusions

In this study, previously identified gene expression biomarker candidates to detect the misuse of a combination of an androgen plus an estrogen for anabolic purposes in liver tissue could be validated for a combination of a gestagen plus an estrogen and even for the use of clenbuterol. This indicates that this approach is very promising to find a biomarker signature for anabolic substances that is really based on the physiological effect, independent of the way of receptor action and of the applied substance. Selection of candidate genes using a holistic process such as RNA sequencing seems to be a good way, but as there are also some genes in the selected gene pattern that were chosen by a literature search for physiological effects caused by anabolic substances, a combination of both processes for candidate gene selection seems to be an even better strategy.

To verify those candidates, more validation studies are necessary using other anabolic substances or hormone cocktails. Another important point would be to test, if the separation is still valid, if different breeds and animals of different age and gender are used in the treatment trials. Combining significantly regulated genes in additional tissues, e.g. muscle tissue, would also be a potential way to have more biomarker candidates that enable a more solid separation of treatment groups.

Acknowledgements

A special thanks to the Kontrollgemeinschaft Deutsches Kalbfleisch, the Société Générale de Surveillance (SGS) and Denkavit for supporting the animal study. Thanks too to Boehringer Ingelheim for providing the applied clenbuterol.

Funding

This study was supported by the Deutsche Forschungsgemeinschaft [grant number DFG Ri 2129/1-1].

References

Becker C, Riedmaier I, Reiter M, Tichopad A, Groot MJ, Stolker AA, Pfaffl MW, Nielen MF, Meyer HH. 2011. Influence of anabolic combinations of an androgen plus an estrogen on biochemical pathways in bovine uterine endometrium and ovary. J Steroid Biochem Mol Biol. 125:192–201.

Becker C, Riedmaier I, Reiter M, Tichopad A, Pfaffl MW, Meyer Heinrich HD. 2011. Effect of trenbolone acetate plus estradiol on trqnscriptional regulation of metabolism pathways in bovine liver. Horm Mol Biol Clin Investig. 2:257–265.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 55:611–622.

Cantiello M, Carletti M, Cannizzo FT, Nebbia C, Bellino C, Pié S, Oswald IP, Bollo E, Dacasto M. 2007. Effects of an illicit cocktail on serum immunoglobulins, lymphocyte proliferation and cytokine gene expression in the veal calf. Toxicology. 242:39–51.

Dervilly-Pinel G, Courant F, Chéreau S, Royer AL, Boyard-Kieken F, Antignac JP, Monteau F, Le BB. 2012. Metabolomics in food analysis: application to the control of forbidden substances. Drug Test Anal. 4:59–69.

Lange IG, Daxenberger A, Meyer HH. 2001. Hormone contents in peripheral tissues after correct and off-label use of growth promoting hormones in cattle: effect of the implant preparations Finaplix-H, Ralgro, Synovex-H and Synovex Plus. APMIS. 109:53–65.

Ludwig SK, Smits NG, Cannizzo FT, Nielen MW. 2013. Potential of treatment-specific protein biomarker profiles for detection of hormone abuse in cattle. J Agric Food Chem. 61:4514–4519.

Ludwig SK, Smits NG, van d G., Bremer MG, Nielen MW, Koomen JM. 2012. Multiple protein biomarker assessment for recombinant bovine somatotropin (rbST) abuse in cattle. PLoS One. 7:e52917.

Masanetz S, Preissinger W, Meyer HH, Pfaffl MW. 2011. Effects of the prebiotics inulin and lactulose on intestinal immunology and hematology of preruminant calves. Animal. 5:1099–1106.

Moran C, Quirke JF, Prendiville DJ, Bourke S, Roche JF. 1991. The effect of estradiol, trenbolone acetate, or zeranol on growth rate, mammary development, carcass traits, and plasma estradiol concentrations of beef heifers. J Anim Sci. 69:4249–4258.

Pinel G, Weigel S, Antignac J-P, Mooney MH, Elliott C, Nielen MWF, Bizec B. 2010. Targeted and untargeted profiling of biological fluids to screen for anabolic practices in cattle. TrAC Trends Analyt Chem. 29:1269–1280.

Riedmaier I, Becker C, Pfaffl MW, Meyer HH. 2009. The use of omic technologies for biomarker development to trace functions of anabolic agents. J Chromatogr A. 1216:8192–8199.

Riedmaier I, Benes V, Blake J, Bretschneider N, Zinser C, Becker C, Meyer HH, Pfaffl MW. 2012. RNA-sequencing as useful screening tool in the combat against the misuse of anabolic agents. Anal Chem. 84:6863–6868.

Riedmaier I, Pfaffl MW. 2013. Transcriptional biomarkers–high throughput screening, quantitative verification, and bioinformatical validation methods. Methods. 59:3–9.

Riedmaier I, Reiter M, Tichopad A, Pfaffl MW, Meyer HH. 2011. The potential of bovine vaginal smear for biomarker development to trace the misuse of anabolic agents. Exp Clin Endocrinol Diabetes. 119:86–94.

Riedmaier I, Tichopad A, Reiter M, Pfaffl MW, Meyer HH. 2009. Identification of potential gene expression biomarkers for the surveillance of anabolic agents in bovine blood cells. Anal Chim Acta. 638:106–113.

Appendix IV

Biomolecular Detection and Quantification 5 (2015) 15–22

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

journal homepage: www.elsevier.com/locate/bdq

Research paper

The potential of circulating extracellular small RNAs (smexRNA) in veterinary diagnostics—Identifying biomarker signatures by multivariate data analysis

Spornraft Melanie*, Kirchner Benedikt, Michael W. Pfaffl, Riedmaier Irmgard

Chair of Physiology, Technical University Munich, TUM School of Life Sciences Weihenstephan, Weihenstephaner Berg 3, 85354 Freising-Weihenstephan, Germany

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 3 August 2015
Accepted 7 August 2015
Available online 19 September 2015

Keywords:
Biomarker signatures
Circulating small RNAs
Multivariate data analysis
Small RNA-Sequencing
Transcriptomics
Veterinary diagnostics

ABSTRACT

Worldwide growth and performance-enhancing substances are used in cattle husbandry to increase productivity. In certain countries however e.g., in the EU, these practices are forbidden to prevent the consumers from potential health risks of substance residues in food. To maximize economic profit, 'black sheep' among farmers might circumvent the detection methods used in routine controls, which highlights the need for an innovative and reliable detection method. Transcriptomics is a promising new approach in the discovery of veterinary medicine biomarkers and also a missing puzzle piece, as up to date, metabolomics and proteomics are paramount. Due to increased stability and easy sampling, circulating extracellular small RNAs (smexRNAs) in bovine plasma were small RNA-sequenced and their potential to serve as biomarker candidates was evaluated using multivariate data analysis tools.

After running the data evaluation pipeline, the proportion of miRNAs (microRNAs) and piRNAs (PIWI-interacting small non-coding RNAs) on the total sequenced reads was calculated. Additionally, top 10 signatures were compared which revealed that the readcount data sets were highly affected by the most abundant miRNA and piRNA profiles. To evaluate the discriminative power of multivariate data analyses to identify animals after veterinary drug application on the basis of smexRNAs, OPLS-DA was performed. In summary, the quality of miRNA models using all mapped reads for both treatment groups (animals treated with steroid hormones or the β-agonist clenbuterol) is predominant to those generated with combined data sets or piRNAs alone. Using multivariate projection methodologies like OPLS-DA have proven the best potential to generate discriminative miRNA models, supported by small RNA-Seq data. Based on the presented comparative OPLS-DA, miRNAs are the favorable smexRNA biomarker candidates in the research field of veterinary drug abuse.

1. Introduction

Monitoring of chemical contaminations, species fraud and product mislabelings in food is a complex task for control laboratories. Recent pan-European food safety affairs, for example the horse-meat scandal in 2013, underline the need for sophisticated and reliable analytical methods as well as sufficiently frequent routine investigations in food producing animals [1]. For official laboratories, the conventional methods for screening for forbidden veterinary drug compounds are RIA (radio immuno assay) and ELISA (enzyme-linked immunosorbent assay), and for the confirmation, it is mass spectrometry (MS) combined with gas (GC–MS) or liquid chromatography (LC–MS) [2]. These verifying approaches persue the direct tracking of targeted chemical compounds and/or their metabolites in various food, feed or biological samples. As corresponding analytical protocols are based on the direct detection of the target substance in a sample matrix, the chemical and physical properties of this substance must be known in advance. For example, to test the compliance with regulations in antibiotics surveillance, a maximum threshold of antibiotic residues may not be exceeded in the detection window. However, in the

Abbreviations: CLEN, treated group with clenbuterol-hydrochloride; CON, control group; DA, discriminant analysis; EU, European Union; exRNA, extracellular RNA; miRNA, microRNA; OPLS, orthogonal partial least-squares; PCA, principal component analysis; P+EB, treated group with steroid hormone implant: progesterone plus estradiol benzoate; piRNA, PIWI-interacting small non-coding RNA; PLS, partial least-squares projection; rpm, reads per million; small RNA-Seq, small RNA-Sequencing; smexRNA, circulating extracellular small RNA.

* Corresponding author.
E-mail addresses: melanie.spornraft@wzw.tum.de (S. Melanie), benedikt.kirchner@wzw.tum.de (K. Benedikt), michael.pfaffl@wzw.tum.de (M.W. Pfaffl), irmgard.riedmaier@wzw.tum.de (R. Irmgard).

http://dx.doi.org/10.1016/j.bdq.2015.08.001

case of an illegal abuse, where unknown substances or undefined drug cocktails with low-dose single compounds were administered, chromatographic systems are limited. This is especially the case when the substance itself has already been metabolized (but the physiological effect is still existent), or due to signal to noise ratio in MS and the unknown mass of the applied drug(s). Next to chromatographical methods or immunoassays, new and innovative techniques have emerged in veterinary medicine in the last years. Since recently, veterinary drug abuse can be detected by finding endogenous molecular biomarkers on the transcriptomic, proteomic or metabolomic level that indirectly indicate exogenous physiological modifications [3]. With the objective of controlling veterinary drug abuse, metabolomics approaches have so far shown to be effective in detecting growth-promotor abuse in bovines [4–6], and in racehorses [7,8]. Rapid technological advancements in these "-omics" sciences allow now a comprehensive high-throughput screening for differentially expressed biomarkers. Thus, according to a physiological condition, disease status, or drug application, the biomarker signature is capable of revealing specific biological traits or a measurable change in the organism [9]. Seen from the genetic point of view, the transcription of genes is a fast and highly dynamic process that adapts to environmental stimuli, such as medication, making the transcriptome ideally suitable for the discovery of new biomarkers. The transcriptome covers inter alia a RNA class called microRNAs (miRNA). These small, non-protein coding molecules with a length of typically 18 to 25 nucleotides act as modulators of mRNA targets on the post-transcriptional level. By suppressing the mRNA translation or promoting mRNA destabilization, miRNAs play key roles in regulating gene expression in a multitude of healthy and pathologic biological processes [10]. The successful identification of miRNA biomarkers is already evident in clinical diagnostics, such as early disease detection, progression monitoring and prognosis [11]. In veterinary drug analysis, it was also possible to establish miRNA supported biomarkers in bovine liver to detect anabolic steroid treatment [12]. In the year 2008, miRNAs were also detected as free, extracellular RNA (exRNA) in the bloodstream [13] and the potential usability of circulating nucleic acids as biomarkers was promptly recognized and investigated. Since then, circulating extracellular small RNAs (smexRNA) have been detected in other human body fluids, e.g., milk, saliva, tears, cerebrospinal fluid, urine etc. [14]. Among these smexRNAs are also a recently very emerging class of transcriptional molecules, the PIWI-interacting small non-coding RNAs (piRNAs). They are slightly longer than miRNAs (25 to 32 nucleotides), but also show post-transcriptional regulatory functions. Initially detected in the germ line of drosophila, piRNAs are involved in RNA silencing and therefore in gene expression regulation (as reviewed in [15]). In biomarker development, it was already verified that circulating piRNAs own the potential to serve as human biomarkers for several cancer types, for example gastric cancer [16]. Focusing animal sciences, proteomics and metabolomics are now gradually finding their way into veterinary medicine and food safety analyses, but transcriptomics and especially the analysis of small RNAs and/or smexRNAs have not yet fully arrived.

Worldwide growth and performance-enhancing substances are used in cattle husbandry to increase productivity. Livestock farming strives to promote faster weight gains, increased feed conversion efficiencies and heavier carcasses to maximize economic profit. However, the use of anabolic agents is prohibited in certain countries, including the European Union (EU) since the EU Council Directive 88/146/EC188 entered into force in 1988. From that year on, all growth promoting agents including steroid hormones and β-adrenergic agonists have been prohibited from animal breeding practices across European markets. This ban was mainly due to precautionary food safety reasons to prevent consumers from possible health risks caused by residue carryover [17]. Also the import of products derived from hormone-treated cattle is legally forbidden in the EU. Due to financial benefits, an abuse by application of illicit substances is still frequently suspected in meat production [18]. To circumvent supervisory authorities and positive test results, alternative compounds as well as application scenarios have emerged. Applying transcriptomics in the field of food safety constitutes a new innovative screening strategy for a reliable and effective control method to maintain legislation. First studies demonstrated that the monitoring of mRNA expression ratios has already proven to be a useful tool for biomarker development to trace growth-promotor abuse [19–21] however, far less is known about the applicability of smexRNAs in this context.

If the aim is to measure small non-coding RNAs in a high-throughput approach, small RNA-Sequencing (small RNA-Seq) is the strategy of choice. This allows the holistic and parallel sequencing-by-synthesis analysis of the whole transcriptome of multiplexed samples. To study the influence of anabolic substances on the gene expression profiles at the small RNA level in meat-producing livestock, an animal trial was conducted to simulate the real environment during a potential drug abuse situation. In general, ultrahigh-throughput studies result in immensely huge data output that is highly multivariate (k variables $\gg n$ observations). To get the most value out of complex small RNA-Seq data and reveal knowledge that is hiding behind, we implemented multivariate projection methodologies to circumvent this bottleneck in biomarker development. The aim is to find a valid and stable biomarker signature, which explicitly leads back to the treatment. Thereby, treated or diseased subjects will be compared with untreated control samples. To select the most significant single biomarkers and combine this pattern to a biomarker signature, the applicability of multivariate projection methodologies in omics studies is beneficial and productive [22]. Most applied multivariate projection methodologies are principal-component analysis (PCA), hierarchical clustering (HCA), and partial least-squares (PLS) projections to latent structures [22]. Recently, orthogonal partial least squares (OPLS) demonstrated to be a useful discriminant analysis (DA) tool for complex data structures [23,24]. The goal of OPLS-DA is to establish a model that is able to distinguish the classes of observations (non-treated from treated), to visualize large-volume data sets and to highlight meaningful interpretation possibilities.

The OPLS algorithm [25] is an improved and complexity-reduced interpretation of PLS regression models with an integrated orthogonal correction filter [26], allowing easier interpretation and augmenting classification performance [27]. Therefore, systemic variation from the input data set X, which is not correlating with the response set Y, is eliminated [25]. High-quality OPLS-DA models have the ability to separate the modelled variation in X into two parts, one that is correlated to Y and therefore predictive, and another that is orthogonal to Y. Thus, the correlated and therefore predictive variation in X is displayed by the predictive components and represents the variation between classes (non-treated animals and treatment groups). The variation in X that is orthogonal to Y is modeled by the orthogonal components and reflects the variation within classes [28].

Not only miRNAs but also piRNAs were investigated in this study to evaluate the potential of both smexRNA biomarker candidates. The decisive advantages of smexRNAs in bio-fluids compared to RNAs sampled from tissue are easy accessibility and an increased stability in the body and after sample collection [29]. We examine and discuss the potential of smexRNAs as novel source of biomark-

ers in veterinary diagnostics, to battle against illegal application of growth and performance enhancing substances to bovines.

2. Material and methods

2.1. Design of the animal study

In this study, 21 male Friesian Holstein veal calves (*bos taurus*) were randomly divided into three groups of 7 animals each ($n = 7$). All animals had a similar age (161 ± 15 days) and an average body weight of 151.4 ± 19.2 kg at the beginning of the trial. One group remained completely untreated and served as control group (CON). The second group was treated with Component E–C (IVY Animal Health, USA), a hormonal implant that consisted of a combination of 100 mg of progesterone plus 10 mg of estradiol benzoate (steroid hormone group, P + EB). One implant per animal was deposited between the skin and the cartilage on the backside of the middle third of the pinna of the ear. The third group received an oral dose of clenbuterol-hydrochloride (clenbuterol group, CLEN) (10 μg/kg body weight) (Boehringer Ingelheim, Germany) in daily intervals for 36 days. This animal study was approved by the ethical committee of the German Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (permit number 84-02.04.2012.A040). The animals were housed and fed according to good animal attendance practice and all efforts were made to prevent suffering.

2.2. Plasma sampling

To generate plasma, peripheral whole blood was collected from the jugular vein using 9 ml K3E K_3EDTA-Vacuette tubes (Greiner bio-one, Germany) and single use needles (Greiner bio-one, Germany) with a subsequent separation of cellular components by centrifugation for 15 min at 3500 × rcf at room temperature with transportable centrifuges (EBA20, Hettich, Germany). Plasma was stored at −80 °C until RNA extraction. Samples were taken at d + 17 after the initial treatments.

2.3. Total RNA isolation

ExRNAs from plasma were isolated by an optimized method that enabled small RNA-Seq as previously implemented by our group [30]. RNA eluates were stored at −80 °C until further usage.

2.4. Small RNA Sequencing, Data Evaluation, Mapping and Annotation

The sample pre-processing pipeline, analytical small RNA-Seq steps on a HiSeq sequencing platform (Illumina, USA) and the bioinformatics steps to generate annotated readcount tables of 21 bovine plasma samples were realized as described and discussed by our group [30]. As inter-sample normalization strategy, total readcounts were adjusted to library sizes in reads per million (rpm) to correct differences in library sizes [31].

2.5. Univariate and multivariate data analysis

SigmaPlot 11.0 (Systat Software Inc., USA) was used for statistical data analysis and SIMCA 13.0.3.0 software (Umetrics AB, Sweden) for running the multivariate data analysis. For model generation, library size-normalized data sets were first logarithmically transformed and then pareto-scaled [24]. Different miRNA and piRNA models, depending on the input data quantities, were built. These were either readcount tables with all annotated reads

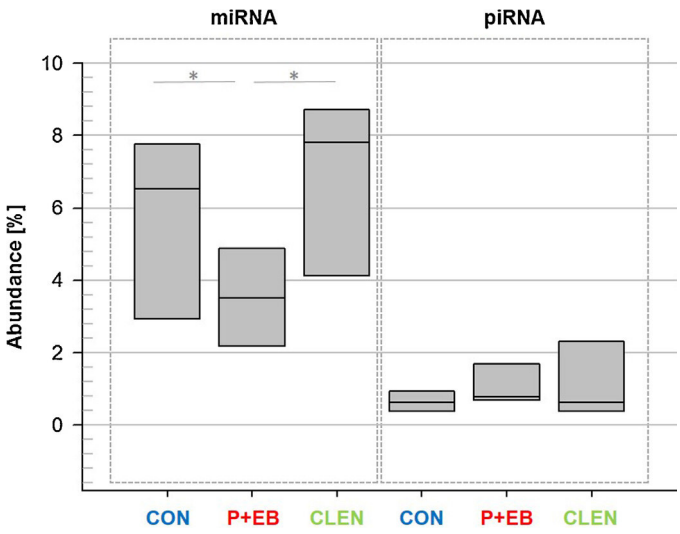


Fig. 1. Abundance of circulating miRNAs and piRNAs. Box plots illustrate the circulating miRNA and piRNA proportions in plasma of untreated control animals (CON), steroid hormone- (P + EB) and clenbuterol (CLEN)-treated animals ($n = 7$ each). Steroid hormones decreased the miRNA quantity ($p = 0.047$) and clenbuterol application resulted in increased miRNA concentrations ($p = 0.042$).

(all reads) or with more than 50 rpm at an average (>50 readcounts). 50 rpm was set as a noise cut-off that is commonly used in small RNA-Seq data analysis. Discriminative model results were shown in scores scatter plots, displaying the CON group in blue, the P + EB group in red and the CLEN group in green. The quality of OPLS-DA models was controlled by evaluating R2 and Q2 values. The R2(cum) value represents the cumulative percentage of the modelled variation in *Y*, using the *X* model. Therefore, the R2(cum) value is the measure of fit and describes how well the model fits the *X* data. A large value close to 1 is a requisite condition for good models. The Q2(cum) value is the cumulative percentage of the variation in *Y* that can be predicted by the model according to cross validation using the *X* model. Q2(cum) is the measure of predictability and explains how well the generated model predicts new data. A large value (>0.5) indicates good predictability [28].

3. Results and discussion

3.1. Abundance of smexRNAs

After passing the sequencing quality checkpoints and successful alignment, miRNA and piRNA data from all sequenced 21 animals were library-size normalized and compiled in readcount tables. Analysis of the proportion of miRNAs on the total sequenced reads resulted in a content of $5.7\% \pm 2.4$ (SD) (median = 6.6) in the CON group, $3.5\% \pm 1.7$ (SD) (median = 3.6) in the P ± EB group and $6.6\% \pm 2.7$ (SD) (median = 7.9) in the CLEN group (Fig. 1). There is a statistically significant difference between the CON group and the steroid hormone treated group ($p = 0.047$) and also between the two treatment groups ($p = 0.042$). Concerning piRNAs, CON contained $0.7\% \pm 0.4$ (SD) (median = 0.6), P ± EB $1.0\% \pm 0.7$ (SD) (median = 0.8) and CLEN $1.1\% \pm 1.0$ (SD) (median = 0.6) without statistically significant differences (Fig. 1).

As the magnitude of piRNAs was comparable to previously published data of nine healthy bovines [32], the proportion of piRNAs seemed not to alter even under the influence of anabolic stimulants. It was a recognizable effect that the steroid treatment led to a significant decrease of miRNA quantity compared to the CON group. In the CLEN-treated animals, gene expression changes towards an upregulation of miRNAs were noticed compared to the P + EB treated

Table 1
Comparison of the top 10 expressed miRNAs and piRNAs in the three analyzed groups: (CON) control group, (P + EB) steroid hormone-treated group, (CLEN) clenbuterol-treated group. Checkmarks signify presence of matching small RNAs and superscript numbers give ranking information. Pie charts depict the percentage of the top 10 on the total annotated miRNAs and piRNAs, respectively.

miRNA	CON	P+EB	CLEN	piRNA	CON	P+EB	CLEN
bta-miR-122	✔ 1	✔ 1	✔ 1	piR-33151	✔ 1	✔ 3	✔ 4
bta-miR-423-3p	✔ 2	✔ 2	✔ 2	piR-33082	✔ 2		✔ 8
bta-miR-423-5p	✔ 3	✔ 3	✔ 3	piR-43772	✔ 3	✔ 1	✔ 1
bta-miR-486	✔ 4	✔ 6	✔ 5	piR-43771	✔ 4	✔ 2	✔ 2
bta-miR-99a-5p	✔ 5	✔ 5	✔ 6	piR-43786	✔ 5	✔ 4	✔ 5
bta-miR-320a	✔ 6	✔ 4	✔ 4	piR-60565	✔ 6	✔ 5	✔ 3
bta-miR-92a	✔ 7	✔ 7	✔ 9	piR-35982	✔ 7	✔ 7	✔ 7
bta-miR-3600	✔ 8		✔ 8	piR-31068	✔ 8	✔ 6	✔ 6
bta-miR-193a-5p	✔ 9	✔ 8	✔ 7	piR-31038	✔ 9		✔ 9
bta-miR-192	✔ 10	✔ 9	✔ 10	piR-35284	✔ 10		✔ 10
bta-miR-140		✔ 10		piR-36256		✔ 8	
				piR-36255		✔ 9	
				piR-36243		✔ 10	
% Top 10	84.2	76.0	83.8		91.3	82.2	92.3
							

individuals. Therefore, the different kinds of treatment substances seemed to have an opposite impact on miRNA translation.

3.2. Top 10 abundance lists

Rpm-normalized miRNA reads were sorted according to their decreasing readcount numbers to generate top 10 abundance lists. This revealed that the largest proportion of the data sets was reflected by the top 10 ranks: 84.2% in the CON group, 76.0% in the P + EB group and 83.8% in the CLEN group (Table 1). By comparing the most abundant CON miRNAs with the P + EB and CLEN treatment group, it could be stated that the composition is nearly the same (CON vs. P + EB) or exactly matching (CON vs. CLEN). The top 10 signature of the CON group confirmed miR-3600, which was substituted with miR-140 in the P + EB group. To evaluate piRNAs that were high ranking in terms of abundance, the same data organization was conducted. The top 10 piRNA list accounted for 91.3% of the total reads data set in the CON group, for 82.2% in the P + EB group and for 92.3% in the CLEN group. The CON and CLEN piRNA list exhibited the same pattern and the P + EB group varied in three piRNAs: piR-31038, piR-35284 and piR-33082. A statistically significant difference between the groups could not be detected. Therefore, in summary for both treatment groups, the readcount data sets were highly affected by the most abundant miRNA and piRNA profiles. Moreover, top 10 expressed data did not show significant deviations from the CON group, indicating that the smexRNA profiles of treated animals were not subject of fluctuations as great as assumed and the major components were constantly expressed.

3.3. Differential expression

To evaluate the discriminative power of multivariate data analyses to identify animals after veterinary drug application on the basis of smexRNAs, OPLS-DA was performed after data pre-processing. miRNA and piRNA scores scatter plots were analyzed regarding between class variation (horizontal direction) and within class variation (vertical direction) depending on the read input of either all aligned reads (all reads) or size-filtered data sets with reads that had averagely more than 50 rpm (>50 readcounts) (Figs. 2–4), ,). Fig. 5 gives an overview over model quality parameters of all examined discriminative analyses.

First, the discriminative power of combined data sets including miRNAs and piRNAs was examined. As shown in Fig. 2[A] and [B], the separation between the CON animals and the treated groups, based on miRNA observations >50 readcounts, is imperfect. Although moderate goodness of fit and prediction could be attested for the CLEN study model (CLEN: R2(cum) = 0.752, Q2(cum) = 0.458), it was not feasible for the P + EB study (P + EB: R2(cum) = 0.319, Q2(cum) = 0.001). Better discriminative and quality results were accomplished with models that included all reads (Fig. 2[C] and [D]). DA could not manage to perfectly separate the

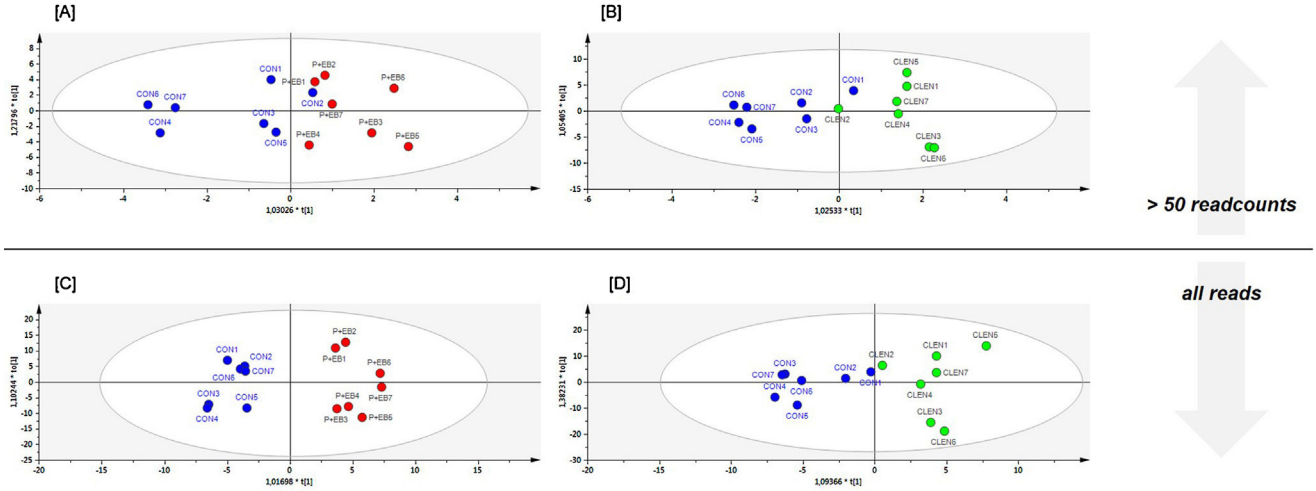


Fig. 2. Combined miRNA and piRNA data set**.** OPLS-DA of sequenced plasma samples using full [C and D] and readcount-filtered datasets [A and B]. [A and C] represent scores scatter plots discriminating control animals (CON, blue) from steroid hormone-treated animals (P + EB, red). [B] and [D] display samples from the CON and the clenbuterol-treated population (CLEN, green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

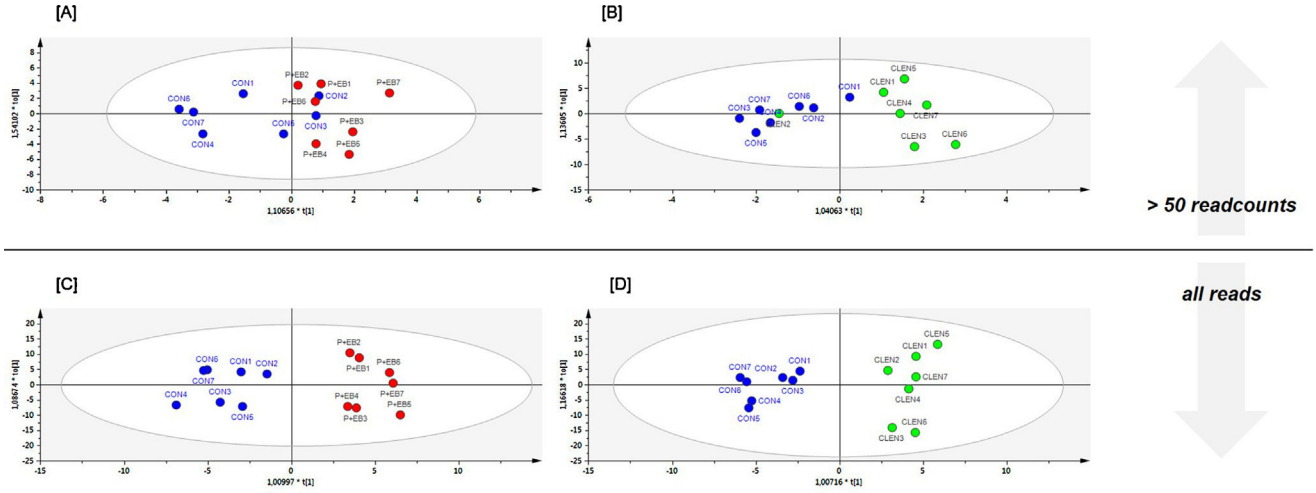


Fig. 3. MiRNA data set. OPLS-DA of sequenced plasma samples using full [C and D] and readcount-filtered datasets [A and B]. [A and C] represent scores scatter plots discriminating control animals (CON, blue) from steroid hormone-treated animals (P + EB, red). [B] and [D] display samples from the CON and the clenbuterol-treated population (CLEN, green).

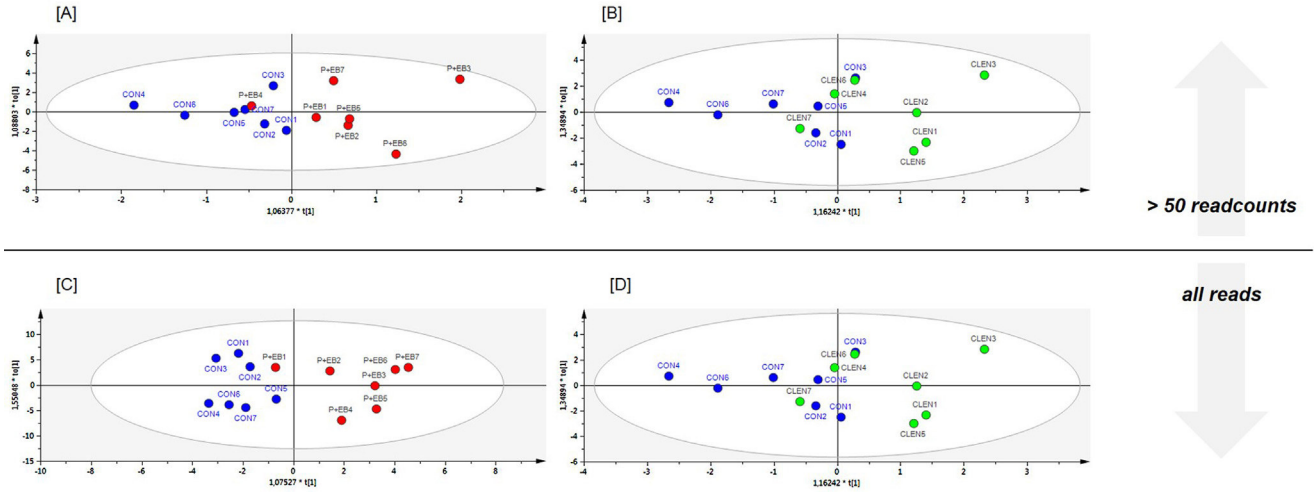


Fig. 4. PiRNA data set. OPLS-DA of sequenced plasma samples using full [C and D] and readcount-filtered datasets [A and B]. [A and C] represent scores scatter plots discriminating control animals (CON, blue) from steroid hormone-treated animals (P + EB, red). [B] and [D] display samples from the CON and the clenbuterol-treated population (CLEN, green).

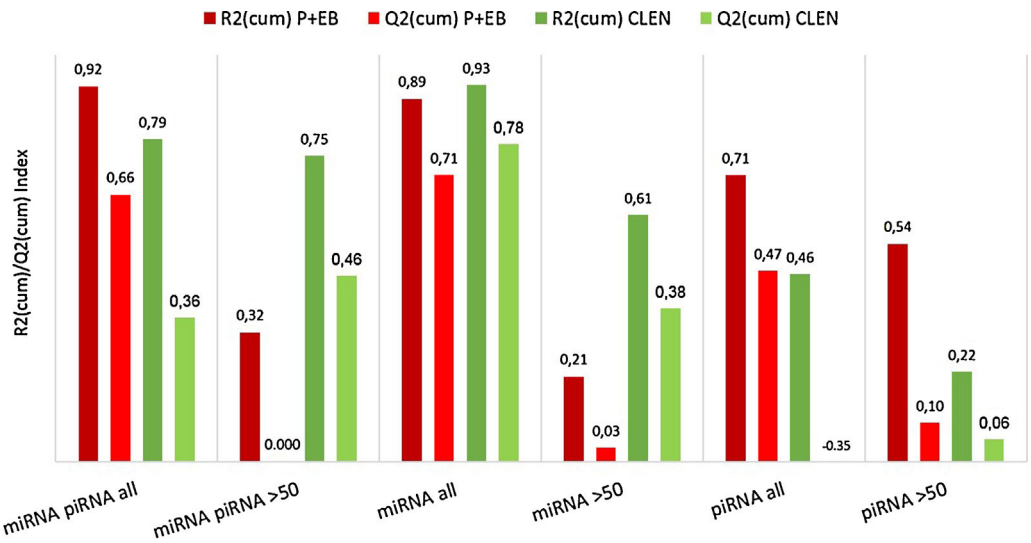


Fig. 5. Model quality overview. The R2(cum) value (dark colored bars) reflects the goodness of fit and the Q2(cum) value (light colored bars) the goodness of prediction. Quality parameters were evaluated for the data set with all reads and with reads over averagely more than 50 readcounts (>50 readcounts). Red colored bars display the values from the P + EB study and green colored bars the values from the CLEN study.

CLEN-treated animals from the CON group (CLEN: R2(cum) = 0.794; Q2(cum) = 0.355), while acceptance of all reads in the discriminative analysis led to a model that allowed separation of P + EB classes with high quality (P + EB: R2(cum) = 924, Q2(cum) = 0.657).

In a second step, to examine a potentially improved discriminative power of data sets with uniquely miRNAs or piRNAs respectively, equivalent OPLS-DA models were generated. As shown in Fig. 3[A] and [B], the separation between the CON animals and the treated groups, based on miRNA observations >50 readcounts, is not more specific. The model quality parameters attest better fit and prediction for the CLEN study (CLEN: R2(cum) = 0.607, Q2(cum) = 0.377; P + EB: R2(cum) = 0.210, Q2(cum) = 0.035) but still, best discriminative and quality results were accomplished with models that included all available reads. For both multivariate data analyses studies, the miRNA models were able to distinguish between classes (Fig. 3[C] and [D]), which was reflected by good fit (CLEN: R2(cum) = 0.927; P + EB: R2(cum) = 0.893) as well as good predictability (CLEN: Q2(cum) = 0.782; P + EB: Q2(cum) = 0.706). Generally, large R2 and Q2 values at the level of 0.5 or above are necessary for high quality OPLS-DA models. Therefore, the quality parameters of the miRNA OPLS-DA models for the P + EB and the CLEN-treated animals indicated that the models fitted the data very well and that new variables could be predicted. Using all reads, the scores scatter plots illustrated a grouping of the CON and the treated animals, highlighting that multivariate data analysis tools were clearly capable to reveal treatment-dependent differences at the miRNA level. Moreover, fusion of data sets did not deliver better fitting and predicting results compared to miRNAs only, neither for the different treatment groups nor for the two compared data inputs.

Besides miRNAs, OPLS-DA models were generated and evaluated for piRNA data only (Fig. 4). Here again, a separation of the treatment groups was not feasible for the data set >50 readcounts. According to that effect, R2 and Q2 could not meet quality standards (CLEN: R2(cum) = 0.221, Q2(cum) = 0.055; P + EB: R2(cum) = 0.536, Q2(cum) = 0.096). Compared to the miRNA models, the piRNA models with all reads could not cluster the treated animals. For the P + EB group, a better OPLS-DA model could be generated than for the CLEN group, also regarding quality (CLEN: R2(cum) = 0.461, Q2(cum) = −0.346; P + EB: R2(cum) = 0.706, Q2(cum) = 0.47). The DA of piRNAs could not present acceptable models, as they could not explain the variation of the variables nor could they predict. Obviously, the miRNA abundance and thus the utilizable read numbers for statistical analyses exceed that of piRNAs (Fig. 1). Therefore, miRNA OPLS-DA could be based on increased data volumes supporting a better prediction ability and discrimination. Fusion of data sets delivered better fitting and predicting results compared to piRNAs only, when using all reads. For the CLEN study, the combined model also presented better fit and prediction for the >50 readcount model. For the miRNAs from the P + EB study, merging data sets resulted in slightly increased R2 values using all reads. Best fit and prediction in the CLEN study were achieved while using all miRNA reads (Fig. 3[D]). In summary, adding miRNAs to improve a piRNA based model resulted in better discrimination, fit and predictability, whereas miRNAs alone (all reads) provided best results for both treatment modalities. Referring to this trial, it can be concluded that DA is improving the more data (reads) are fed into OPLS-DA. It became apparent that OPLS-DA is best suited for full datasets of circulating miRNAs in the search for veterinary drug abuse biomarkers.

3.4. Comparison of analysis models

Taking all findings together, it can be stated that the main percentage of the miRNA and piRNA signature is composed of the top 10 candidates (Table 1). Furthermore, the miRNA and piRNA top 10 signature is almost the same, if the CON group was compared with the P + EB animals, or even identical (CON vs. CLEN). No statistically significant expression ratio could be determined. Therefore, the investigated smexRNAs were expressed more stable than assumed. It was though already published that the treatments were effective, as animals showed a significant weight gain (d0 to d + 34) and a potential gene expression biomarker signature was identified on the mRNA level in liver samples in the course of the same animal trial [21]. Nevertheless, overall variation in the expressed miRNA profiles is sufficient for the generation of good-quality OPLS-DA models, but this variation was not explained by the main components of the data set, but rather by the multiplicity of low-abundance miRNAs. Increasing the sequencing depth to exemplary 24-fold (one sample per flow cell lane) could thereby help to

improve the detection of low-abundant smexRNAs, the sequencing of more reads and finally the manifestation of differences in low-abundance circulating small RNAs. As explained before, there is a discriminative effect lying in the data, however hiding behind the major expressed circulating miRNAs.

The quality of miRNA models (all reads) for both treatment groups is predominant to those generated with combined data sets or piRNAs alone. Therefore, the presented results highlighted that miRNAs were superior biomarker candidates to piRNAs regarding the annotated number of reads, model quality, data fit and predictive ability. Therefore, based on the presented comparative OPLS-DA data analyses, miRNAs are the favorable smexRNA biomarker candidates in the research field of veterinary drug abuse. As treated animals could be separated from untreated controls, this study was a first hint, that circulating miRNAs could be beneficial biomarker candidates for anabolic misuse in the future, if sequencing depth is chosen properly.

Although there were clear differences between the control and the treated animals revealed by small RNA-Seq (scores scatter plots in Figs. 2 and 3), the verification and validation of a confident biomarker signature is technically very difficult for smexRNAs in bovine plasma. To experimentally manifest the quantitative expression of candidates in plasma via RT-qPCR, a detectable read-count number is prerequisite. As described before, those miRNAs that fulfilled this precondition were stably expressed and did not underlie variation that is owed to the anabolic treatments. Therefore, due to a very low concentration of smexRNAs in bovine plasma, a sufficient sample volume, an efficient RNA isolation method and appropriate sequencing strategies need to be united for successful screening and validation experiments.

In summary, smexRNAs could be seen as potential candidates in the identification of biomarkers with the ability to uncover illegal drug application in veterinary diagnostics. Using multivariate projection methodologies like OPLS-DA have proven the best potential to generate discriminative miRNA models, supported by small RNA-Seq data. PiRNAs were expressed with low copy numbers, which is not ensuring statistical robustness and significance. OPLS-DA enabled insights into the complex structure of sequencing data and clarified that value could be gained from the presented experiments, namely information about differentiation of treatment groups. However, the quantitative analysis in plasma is challenging as the content of miRNAs or piRNAs seemed robust in bovine plasma (Fig. 1, [32]) and differences in the abundance of minor expressed smexRNAs could not be revealed. Therefore, the assumed modifications of the smexRNA profile by growth-promoting substances was overestimated. Yet, it must not be forgotten that blood and hence plasma underlie extreme systemic influences. The bloodstream is permanently in direct contact with the complete organ system. Hence, the circulating small RNA profile could be heavily altered only throughout one circulation through the body. As described before, steroidal hormone implants as well as the oral clenbuterol doses were effective, but it might be the case that the potential alterations in the small RNA signatures could not be captured due to rapid turnover of the circulation system. To the present date, it could not be described in literature, that smexRNAs are the direct targets of stimulants like steroid hormones or β-2-adrenergic agonist. Therefore, for transcriptional biomarker development, a long-term and permanent miRNA pattern needs to be detected, which is not influenced by the animal circulatory system.

4. Conclusion

In transcriptional biomarker discovery, easy collectable sample specimen like whole blood, serum or plasma offer advantages in veterinary routine diagnostics but meet methodological difficulties, mainly due to matrix complexity, low RNA concentration and bioinformatical challenges. As miRNAs in plasma were tested to be highly stable and resistant to degradation [29], smexRNAs are seen as very potential candidates in the search for the next generation of transcriptional biomarkers. The presented experimental pipeline offered to analyze circulating miRNAs and piRNAs in bovines under anabolic stimulation. The usability of small RNA-Seq in the search for novel miRNA biomarkers in veterinary medicine was demonstrated here, as OPLS-DA discriminative models could be successfully created that also showed high goodness of fit and predictability. Next steps in the experimental lineup would be to deeper sequence plasma samples in order to provide detailed information about the composition of the smexRNA profile at low-abundance levels. Furthermore, RNA isolation systems from bio fluids still need enhanced performances to extract sufficient concentrations of smexRNAs that can be measured with RT-qPCR.

Acknowledgments

The study was organized and conducted in cooperation with KDK (Kontrollgemeinschaft Deutsches Kalbfleisch e.V., Germany), SGS (Société Générale de Surveillance, Switzerland) and DENKAVIT Ingredients (Germany). Boehringer Ingelheim (Germany) supported the study by supplying clenbuterol-hydrochloride.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bdq.2015.08.001.

References

[1] J. Premanandh, Food Control 34 (2013) 568–569.
[2] B. Le Bizec, G. Pinel, J.P. Antignac, J. Chromatogr. A 1216 (2009) 8016–8034.
[3] I. Riedmaier, M.W. Pfaffl, Methods 59 (2013) 3–9.
[4] In: R. Schilt (Ed.), Residues of veterinary drugs in food, Proceedings of the Euroresidue VII Conference, Egmond ann Zee, The Netherlands, 14–16 May, 2012, National Institute of Public Health and the Environment, Egmond aan Zee, 2012.
[5] P. Regal, M.H. Blokland, C.A. Fente, S.S. Sterk, A. Cepeda, Leen A. van Ginkel, J. Agric. Food Chem. 63 (1) (2015) 370–378.
[6] S. Dervilly-Pinel, A. Weigel, S. Lommen, L. Chereau, M. Rambaud, J.P. Antignac, Nielen, W.F. Michel, B. Le Bizec, Anal. Chim. Acta 700 (2011) 144–154.
[7] E. Barrey, M. Triba, F. Messier, L. Le Moyec, Equine Vet. J. 46 (2014) 2–3.
[8] F. Boyard-Kieken, G. Dervilly-Pinel, P. Garcia, A.C. Paris, M.A. Popot, B. Le Bizec, Y. Bonnaire, J. Sep. Sci. 34 (2011) 3493–3501.
[9] M.W. Pfaffl, Methods 59 (2013) 1–2.
[10] C. Price, J. Chen, Genes Dis. 1 (2014) 53–63.
[11] H. Schwarzenbach, N. Nishida, G.A. Calin, K. Pantel, Nat. Rev. Clin. Oncol. 11 (2014) 145–156.
[12] C. Becker, I. Riedmaier, M. Reiter, A. Tichopad, M.W. Pfaffl, Heinrich H.D. Meyer, Analyst 136 (2011) 1204–1209.
[13] Stephen S.C. Chim, Tristan K.F. Shing, Emily C.W. Hung, T.-Y. Leung, T.K. Lau, Rossa W.K. Chiu, Y.M. Dennis Lo, Clin. Chem. 54 (2008) 482–490.
[14] J.A. Weber, D.H. Baxter, S. Zhang, D.Y. Huang, K. How Huang, M. Jen Lee, D.J. Galas, K. Wang, Clin. Chem. 56 (2010) 1733–1741.
[15] L.T. Gou, P. Dai, M.F. Liu, Wiley Interdiscip. Rev. RNA 5 (2014) 733–745.
[16] L. Cui, Y. Lou, X. Zhang, H. Zhou, H. Deng, H. Song, X. Yu, B. Xiao, W. Wang, J. Guo, Clin. Biochem. 44 (2011) 1050–1057.
[17] J. Serratosa, A. Blass, B. Rigau, B. Mongrell, T. Rigau, M. Tortadès, E. Tolosa, C. Aguilar, O. Ribó, J. Balagué, Revue Scientifique Et Technique (Int. Off. Epizoot.) 25 (2006) 637–653.
[18] R.W. Stephany, Handbook of Experimental Pharmacology, 2010, 355–367.
[19] I. Riedmaier, M.W. Pfaffl, Heinrich H.D. Meyer, Drug Test Anal. 3 (2011) 676–681.
[20] I. Riedmaier, V. Benes, J. Blake, N. Bretschneider, C. Zinser, C. Becker, H.H.D. Meyer, M.W. Pfaffl, Anal. Chem. 84 (2012) 6863–6868.
[21] I. Riedmaier, M. Spornraft, M.W. Pfaffl, Food additives & contaminants. Part A, Chem. Anal. Control Expos. Risk Assessment 31 (2014) 641–649.
[22] L. Eriksson, H. Antti, J. Gottfries, E. Holmes, E. Johansson, F. Lindgren, I. Long, T. Lundstedt, J. Trygg, S. Wold, Anal. Bioanal. Chem. 380 (2004) 419–429.
[23] J. Boccard, D.N. Rutledge, Anal. Chim. Acta 769 (2013) 30–39.
[24] P. Regal, S. Anizan, J.P. Antignac, B. Le Bizec, A. Cepeda, C. Fente, Anal. Chim. Acta 700 (2011) 16–25.
[25] J. Trygg, S. Wold, J. Chemometr. 16 (2002) 119–128.

[26] S. Wold, H. Antti, F. Lindgren, J. Öhman, Chemometr. Intell. Lab. Sys. 44 (1998) 175–185.
[27] M. Bylesjö, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, J. Chemometr. 20 (2006) 341–351.
[28] User Guide to SIMCA By MKS Umetrics Version 13 downloaded from: http://131.130.57.230/clarotest190/claroline/backends/ download.php?url=L1N0YXQvU0lNQ0EvVXNlciBHdWlkZSB0byBTSU1DQSAxMy5wZGY%3D&cidReset=true&cidReq=300152WS13 1992-2012 MKS Umetrics AB User guide edition date: April 23, 2012 Sweden Phone: +46 (0)40 664 2580 Email: info@umetrics.com.
[29] P.S. Mitchell, R.K. Parkin, E.M. Kroh, B.R. Fritz, S.K. Wyman, E.L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K.C. O'Briant, A. Allen, D.W. Lin, N. Urban, C.W. Drescher, B.S. Knudsen, D.L. Stirewalt, R. Gentleman, R.L. Vessella, P.S. Nelson, D.B. Martin, M. Tewari, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 10513–10518.
[30] M. Spornraft, B. Kirchner, B. Haase, V. Benes, M.W. Pfaffl, I. Riedmaier, PLoS One 9 (2014) e107259.
[31] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Nat. Methods 5 (2008) 621–628.
[32] M. Spornraft, B. Kirchner, M.W. Pfaffl, I. Riedmaier, Biotechnol. Lett. 37 (June (6)) (2015) 1165–1176, http://dx.doi.org/10.1007/s10529-015-1788-2, Epub 2015 February 21.

Appendix V

Table S1: miRNA data evaluation of all 21 sequenced samples. The table presents summarized readcount information about the number of reads that failed the trimming and size selection process (<16nts), reads that were mapped to Rfam, reads that could not be aligned to miRBase and reads that were annotated as mature, bovine miRNAs.

Data evaluation - pipeline miRNA	**Animal 1**	**Animal 2**	**Animal 3**	**Animal 4**
Total sequences	9694743	9585690	18089452	6411243
Passed trimming >16nts	5942478	5189172	4797229	3677865
Failed trimming <16nts	3752265	4396518	13292223	2733378
Rfam mapped	155108	95216	207425	176035
Rfam unmapped	5787370	5093956	4589804	3501830
Annotated reads miRBase bta	764079	283405	840542	543465
% (/total sequences)	7.9	3.0	4.6	8.5
Unmapped reads miRBase bta	5023291	4810551	3749262	2958365
Annotated reads miRBase hsa	752137	281446	827746	540062
% (/total sequences)	7.8	2.9	4.6	8.4
Unmapped reads miRBase hsa	5035233	4812510	3762058	2961768
Data evaluation - pipeline miRNA	**Animal 5**	**Animal 6**	**Animal 7**	**Animal 8**
Total sequences	6499294	7423432	8601041	11225200
Passed trimming >16nts	3170317	5010301	5780894	4656921
Failed trimming <16nts	3328977	2413131	2820147	6568279
Rfam mapped	179729	177234	200088	102996
Rfam unmapped	2990588	4833067	5580806	4553925
Annotated reads miRBase bta	162725	503797	568690	430832
% (/total sequences)	2.5	6.8	6.6	3.8
Unmapped reads miRBase bta	2827863	4329270	5012116	4123093
Annotated reads miRBase hsa	161224	491981	560443	416990
% (/total sequences)	2.5	6.6	6.5	3.7
Unmapped reads miRBase hsa	2829364	4341086	5020363	4136935
Data evaluation - pipeline miRNA	**Animal 9**	**Animal 10**	**Animal 11**	**Animal 12**
Total sequences	9210634	13046396	5579989	4152707
Passed trimming >16nts	4655609	2164606	2569336	1587216
Failed trimming <16nts	4555025	10881790	3010653	2565491
Rfam mapped	96956	78487	160643	172079
Rfam unmapped	4558653	2086119	2408693	1415137
Annotated reads miRBase bta	334847	162123	145620	90648
% (/total sequences)	3.6	1.2	2.6	2.2
Unmapped reads miRBase bta	4223806	1923996	2263073	1324489
Annotated reads miRBase hsa	323844	159927	145211	90926
% (/total sequences)	3.5	1.2	2.6	2.2
Unmapped reads miRBase hsa	4234809	1926192	2263482	1324211

Data evaluation - pipeline miRNA	Animal 13	Animal 14	Animal 15	Animal 16
Total sequences	13179852	13196152	9871768	8900340
Passed trimming >16nts	6544568	6920379	6894507	2826235
Failed trimming <16nts	6635284	6275773	2977261	6074105
Rfam mapped	453767	149547	132335	92173
Rfam unmapped	6090801	6770832	6762172	2734062
Annotated reads miRBase bta	800431	646841	885241	237019
% (/total sequences)	6.1	4.9	9.0	2.7
Unmapped reads miRBase bta	5290370	6123991	5876931	2497043
Annotated reads miRBase hsa	788532	645849	861539	233927
% (/total sequences)	6.0	4.9	8.7	2.6
Unmapped reads miRBase hsa	5302269	6124983	5900633	2500135

Data evaluation - pipeline miRNA	Animal 17	Animal 18	Animal 19	Animal 20
Total sequences	8638527	7571565	4563907	10712637
Passed trimming >16nts	3690893	3794789	3453508	4546348
Failed trimming <16nts	4947634	3776776	1110399	6166289
Rfam mapped	540868	412431	70402	509336
Rfam unmapped	3150025	3382358	3383106	4037012
Annotated reads miRBase bta	356691	597167	400560	470206
% (/total sequences)	4.1	7.9	8.8	4.4
Unmapped reads miRBase bta	2793334	2785191	2982546	3566806
Annotated reads miRBase hsa	356137	590453	387208	468602
% (/total sequences)	4.1	7.8	8.5	4.4
Unmapped reads miRBase hsa	2793888	2791905	2995898	3568410

Data evaluation - pipeline miRNA	Animal 21	MEAN	SD
Total sequences	10794489	9378527	3289748
Passed trimming >16nts	6598935	4498672	1580066
Failed trimming <16nts	4195554	4879855	2872908
Rfam mapped	468271	220530	153581
Rfam unmapped	6130664	4278142	1555267
Annotated reads miRBase bta	991866	486514	264635
% (/total sequences)	9.2	5.3	2.5
Unmapped reads miRBase bta	5138798	3791628	1349201
Annotated reads miRBase hsa	984389	479456	260770
% (/total sequences)	9.1	5.2	2.5
Unmapped reads miRBase hsa	5146275	3798686	1352576

Table S2: piRNA data evaluation of all 21 sequenced samples. The table presents summarized readcount information about the number of reads that failed the trimming and size selection process (<26nts), reads that were mapped to Rfam and miRBase, reads that could not be aligned and reads that were annotated as human piRNAs.

Data evaluation - pipeline piRNA	**Animal 1**	**Animal 2**	**Animal 3**	**Animal 4**
Total sequences	9694743	9585690	18089452	6411243
Passed trimming >26nts	1992987	3002846	313300	1825777
Failed trimming <26nts	7701756	6582844	17776152	4585466
Rfam/miRBase mapped	24229	27254	10254	16900
Rfam/miRBase unmapped	1968758	2975592	303046	1808877
Annotated reads piRNAs	144667	89786	33649	33082
% (/total sequences)	1.5	0.9	0.2	0.5
Unmapped reads piRNAs	1824091	2885806	269397	1775795
Data evaluation - pipeline piRNA	**Animal 5**	**Animal 6**	**Animal 7**	**Animal 8**
Total sequences	6499294	7423432	8601041	11225200
Passed trimming >26nts	1402275	3120110	3445637	1977291
Failed trimming <26nts	5097019	4303322	5155404	9247909
Rfam/miRBase mapped	16106	19896	17256	15375
Rfam/miRBase unmapped	1386169	3100214	3428381	1961916
Annotated reads piRNAs	53905	46996	32637	140601
% (/total sequences)	0.8	0.6	0.4	1.3
Unmapped reads piRNAs	1332264	3053218	3395744	1821315
Data evaluation - pipeline piRNA	**Animal 9**	**Animal 10**	**Animal 11**	**Animal 12**
Total sequences	9210634	13046396	5579989	4152707
Passed trimming >26nts	2675451	303727	1000628	472233
Failed trimming <26nts	6535183	12742669	4579361	3680474
Rfam/miRBase mapped	23495	4166	17145	10153
Rfam/miRBase unmapped	2651956	299561	983483	462080
Annotated reads piRNAs	190090	101363	39920	29119
% (/total sequences)	2.1	0.8	0.7	0.7
Unmapped reads piRNAs	2461866	198198	943563	432961
Data evaluation - pipeline piRNA	**Animal 13**	**Animal 14**	**Animal 15**	**Animal 16**
Total sequences	13179852	13196152	9871768	8900340
Passed trimming >26nts	2333271	531655	3385648	871336
Failed trimming <26nts	10846581	12664497	6486120	8029004
Rfam/miRBase mapped	40496	6697	34196	9794
Rfam/miRBase unmapped	2292775	524958	3351452	861542
Annotated reads piRNAs	223261	13540	271987	77665

% (/total sequences)	1.7	0.1	2.8	0.9
Unmapped reads piRNAs	2069514	511418	3079465	783877

Data evaluation - pipeline piRNA	**Animal 17**	**Animal 18**	**Animal 19**	**Animal 20**
Total sequences	8638527	7571565	4563907	10712637
Passed trimming >26nts	733025	1121639	1612639	1043583
Failed trimming <26nts	7905502	6449926	2951268	9669054
Rfam/miRBase mapped	28945	23734	14273	34672
Rfam/miRBase unmapped	704080	1097905	1598366	1008911
Annotated reads piRNAs	54194	29155	105186	35759
% (/total sequences)	0.6	0.4	2.3	0.3
Unmapped reads piRNAs	649886	1068750	1493180	973152

Data evaluation - pipeline piRNA	**Animal 21**	**MEAN**	**SD**
Total sequences	10794489	9378527	3289748
Passed trimming >26nts	2725107	1709055	1046433
Failed trimming <26nts	8069382	7669471	3582037
Rfam/miRBase mapped	34651	20461	10105
Rfam/miRBase unmapped	2690456	1688594	1041008
Annotated reads piRNAs	68538	86433	70709
% (/total sequences)	0.6	1.0	0.7
Unmapped reads piRNAs	2621918	1602161	1008878