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-