# Body fluid identification of blood, saliva and semen using second generation sequencing of micro-RNA

Petersen, Christel H.; Hjort, Benjamin Benn; Tvedebrink, Torben; Kielpinski, Lukasz Jan; Vinther, Jeppe; Morling, Niels

# Body fluid identification of blood, saliva and semen using second generation sequencing of micro-RNA

CrossMark

Christel H. Petersen [a], Benjamin B. Hjort [a,*], Torben Tvedebrink [b], Lukasz J. Kielpinski [c], Jeppe Vinther [c], Niels Morling [a]

[a] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[b] Department of Mathematical Sciences, Aalborg University, Denmark
[c] Department of Biology, Faculty of Science, University of Copenhagen, Denmark

## ARTICLE INFO

## ABSTRACT

We report a new second generation sequencing method for identification micro-RNA (miRNA) that can be used to identify body fluids and tissues. Principal component analysis of 10 miRNAs with high expression in 16 samples of blood, saliva and semen showed clear differences in the expression of miRNA in the investigated tissues. We expect that the method can also be used for identification of other miRNAs that can be used for identifying other body fluids and tissues. We also expect that the method can be used for identification of body fluids and tissues in practical forensic genetic case work.

© 2013 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

When an STR profile is obtained from a crime scene sample, additional knowledge of the body fluid from which the sample was obtained, may be of interest. Many such tests are available. We have explored the use of second generation sequencing of miRNA for the identification of body fluids and tissues within the field of forensic genetics.

## 2. Materials and methods

### 2.1. Samples

Four blood, five saliva and seven semen samples were obtained from one female and nine male volunteers. RNAlater® (Ambion, USA) was added to the blood samples according to the instructions of the manufacturer. Saliva samples were collected in Oragene®-RNA RE-100 collection tubes (DNA Genotek, Ottawa, ON, Canada). Semen ejaculates were collected in Oragene®-RNA RE-100 collection tubes (DNA Genotek, Ottawa, ON, Canada). The samples were stored at −20 °C until investigation.

### 2.2. Extraction of RNA, amplification and sequencing

miRNA/RNA was extracted with RiboPure™-Blood Kit (Invitrogen, USA) from blood, with mirVana™ miRNA Isolation Kit (Invitrogen, USA) from saliva and with NucleoSpin® miRNA extraction kit (Macherey-Nagel, Germany) from semen. RNA was poly-adenylated using the Poly-A Polymerase Tailing Kit (Epicentre Biotechnologies, Wisconsin, USA). RNA was reverse transcribed into cDNA using PrimeScript Reverse Transcriptase™ (Takara Bio Inc., Otsu, Shiga, Japan). Purification and size selection of cDNA was performed by gel electrophoresis using 10% polyacrylamide gels followed by overnight phenol extraction. The cDNA was ligated to a linker using T4 DNA Ligase (Thermo Scientific, Massachusetts, USA). PCR was performed with Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Massechusetts, USA) to amplify the library. PCR products were purified using AMPure XP (Beckman-Coulter). Sequencing by synthesis was performed on Illumina Hiseq 2000 following the recommendations of the manufacturer.

### 2.3. Data analysis

Data analysis was done using Flexbar [1] and MiRanalyser [2]. Massively parallel sequencing with SGS generates large numbers of data, some of which is irrelevant. Therefore, the sequence data was trimmed by the following procedures: (1) Sequences were inspected from the 3′end and recorded when the first base with a Phred score > 19 was observed; (2) thereafter, the poly-A tails, (3)

* Corresponding author. Tel.: +45 28756123.
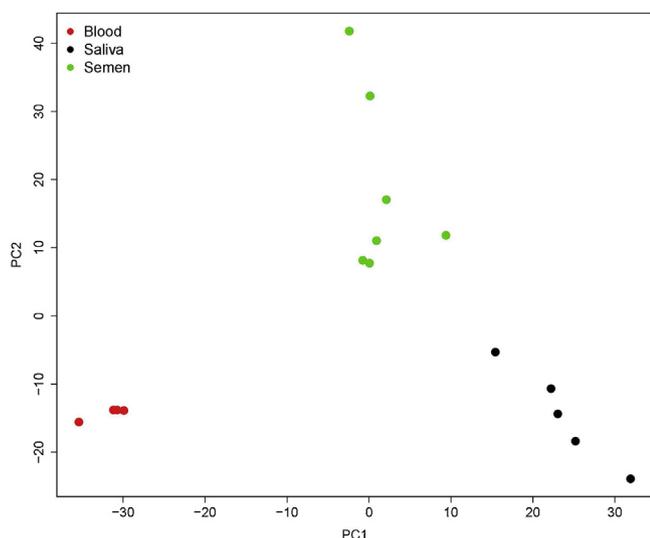E-mail address: benjamin.hjort@sund.ku.dk (B.B. Hjort).

**Fig. 1.** Principal component analysis of 10 miRNAs in 16 samples of blood, saliva and semen.

reads with <18 bases (4) and/or with one or more bases with Phred scores < 20 were removed. The remaining high-quality reads were Bowtie mapped to the miRBase [3]. The principal component analysis was made in R. We used Logit transformed rates between the specific and total numbers of high quality miRNA reads.

## 3. Results and discussion

For each miRNA detected in a sample, the rate of the specific, high-quality miRNAs compared to the total number of detected high-quality miRNAs of the sample was calculated. A total of 10 miRNAs with rates >10% for at least one of the samples were selected (miR-4301, miR-451, miR486-5p, miR-223, miR-29c, miR-4286, miR-16, miR 205, miR-1246 and miR-1274a) for further analysis. Fig. 1 shows a principal component analysis of Logit

transformed rates of the 10 miRNAs in 16 samples of blood, saliva and semen.

## 4. Conclusions

The expression levels of the chosen set of miRNAs efficiently separated the investigated tissues (Fig. 1), indicating that the miRNA expression profiles of blood, semen and saliva are very different. The method is a promising tool for identification and quantification of miRNAs that can be used for body fluid and tissue identification. It seems likely that blood, semen and saliva can be identified in forensic genetic case work with a quite low number of miRNA markers. Moreover, identification of additional body fluids and tissues is anticipated to be possible with a limited number of miRNA markers.

## References

[1] M. Dodt, J.T. Roehr, R. Ahmed, et al., FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms, MDPI Biol. 1 (2012) 895–905.
[2] M. Hackenberg, M. Sturm, D. Langenberger, et al., miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments, Nucl. Acids Res. 37 (2009) W68–W76.
[3] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, Nucl. Acids Res. 39 (2011) D152–D157.