microRNA Biomarker Discovery and High-Throughput DNA Sequencing Are Possible Using Long-term Archived Serum Samples

Trine B. Rounge1, Marianne Lauritzen1, Hilde Langseth1, Espen Enerly1, Robert Lyle2, and Randi E. Gislefoss1

Abstract

Background: The impacts of long-term storage and varying preanalytical factors on the quality and quantity of DNA and miRNA from archived serum have not been fully assessed. Preanalytical and analytical variations and degradation may introduce bias in representation of DNA and miRNA and may result in loss or corruption of quantitative data.

Methods: We have evaluated DNA and miRNA quantity, quality, and variability in samples stored up to 40 years using one of the oldest prospective serum collections in the world, the Janus Serumbank, a biorepository dedicated to cancer research.

Results: miRNAs are present and stable in archived serum samples frozen at −25°C for at least 40 years. Long-time storage did not reduce miRNA yields; however, varying preanalytical conditions had a significant effect and should be taken into consideration during project design. Of note, 500 µL serum yielded sufficient miRNA for qPCR and small RNA sequencing and on average 650 unique miRNAs were detected in samples from presumably healthy donors. Of note, 500 µL serum yielded sufficient DNA for whole-genome sequencing and subsequent SNP calling, giving a uniform representation of the genomes.

Conclusions: DNA and miRNA are stable during long-term storage, making large prospectively collected serum repositories an invaluable source for miRNA and DNA biomarker discovery.

Impact: Large-scale biomarker studies with long follow-up time are possible utilizing biorepositories with archived serum and state-of-the-art technology. Cancer Epidemiol Biomarkers Prev; 24(9): 1–7. ©2015 AACR.

Introduction

Biobanks have become an important resource in medical research. The collection of biobank samples may span a long period of time and include samples from different locations giving rise to variation in the collection protocol. Long-term stored samples are suitable for prospective studies requiring long follow-up times, hence quality and quantity of the samples and molecules, as well as comparability of sample and analyses should be evaluated before large-scale analyses.

miRNA and DNA are both valuable sources of biomarkers of disease. They are present in serum in minute amounts, originating from disrupted cells or secreted, bound to membranes, proteins, or in microvesicles (1, 2). Mature miRNAs, 19 to 24 nucleotides in length, post-transcriptionally regulate gene expression by binding to multiple target mRNAs, resulting in gene silencing through mRNA cleavage, translational repression, or deadenylation (3–5). miRNAs are in general conserved between invertebrates and vertebrates (6) and 2,588 mature miRNAs have been identified in humans to date (miRBase, Release 21; ref. 7). miRNAs are present in all body fluids, including blood plasma and serum (8). Tumor-associated miRNAs in serum were identified in 2008 (9) and it was subsequently shown that the expression levels were reproducible and consistent among individuals (10). miRNA expression profiles differ between cancer and adjacent normal tissue (reviewed in 11) and may accurately classify poorly differentiated cancer types (12). Multiple studies have explored the potential of miRNAs as diagnostic and prognostic cancer biomarkers (10, 13–18). Extracellular miRNAs have a heterogeneous cellular origin and exist in differently packaged forms: the majority is in AGO protein-bound form and the minority is microvesicle-associated (19, 20).

miRNAs remain stable even after being subjected to severe conditions, such as boiling, very low or high pH levels, storage, RNase A treatment, and 10 freeze–thaw cycles (10, 21). In contrast, synthetic unbound free circulating miRNA molecules degrade rapidly (19). miRNAs are preserved up to 10 years in archived serum (22). A small set of long-term stored samples from Janus Serumbank has been used successfully to produce miRNA profiles in noninvasive lung cancer patients (13). However, to our knowledge, comprehensive assessments of miRNA quality and quantity from samples stored beyond 10 years are not available.

Because all cells should be removed during blood clotting, it is expected that serum is deficient in DNA; however, DNA is readily detected in serum at varying amounts depending on extraction protocols and biologic variance. DNA yields vary dependent on
Cancer Epidemiol Biomarkers Prev; 24(9) September 2015

OF2
Cancer Epidemiology, Biomarkers & Prevention

Rounge et al.

Table 1. Sample groups from the Janus Serumbank repository

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample collection period</th>
<th>Sample source</th>
<th>Serum processing</th>
<th>Analyzed for miRNA yield (N)</th>
<th>Analyzed for DNA yield (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1972-1978</td>
<td>HE</td>
<td>Iodoactetate added</td>
<td>44</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>1979-1986</td>
<td>HE</td>
<td>No additives</td>
<td>44</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>1987-2004</td>
<td>HE</td>
<td>Separating gel tubes</td>
<td>49</td>
<td>25</td>
</tr>
<tr>
<td>4A</td>
<td>1973-1979</td>
<td>RCBD</td>
<td>No additives</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4B</td>
<td>1975-1979</td>
<td>RCBD</td>
<td>Lyophilization</td>
<td>42</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>1980-1990</td>
<td>RCBD</td>
<td>No additives</td>
<td>48</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>1997-2004</td>
<td>RCBD</td>
<td>No additives</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>2015-2014</td>
<td>Fresh</td>
<td>No additives</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>319</td>
<td>177</td>
</tr>
</tbody>
</table>

NOTE: The samples are collected as part of national health examinations (HE) or from Red Cross Blood Donors (RCBD). Preanalytical handling differs between groups. The last two columns show the number of samples with measured yield of miRNA and DNA, respectively.

Laboratory protocols

Total RNA was isolated as previously described by Keller and colleagues (13), using TRIzol LS (Life technologies), chloroform, and miRNeasy Serum/Plasma Kit (Qiagen) with an input volume of 500 μL serum. Glycogen was used as a carrier. We also evaluated a high-throughput miRNA extraction method using a phenol-free method on BioRobot Universal System (Qiagen) on 96 samples with 200 μL serum input. miRNA quantity was assessed using the Agilent 2100 Bioanalyzer and the small RNA kit (Agilent Technologies, Cat. No 5067-1548) as described by the supplier. qPCR of miR-16-1, including also a C. elegans miR-39 spike-in, and RT-qPCR control were used to evaluate the yield of miRNA in 96 phenol-free extractions. DNA sequencing was prepared using the ScriptMiher on 8 phenol extracts. Small RNA-Seq Library preparation Kit (Epicentre) sequenced on a MiSeq (Illumina) to confirm sufficient quantity and quality of the miRNA for high-throughput analyses.

DNA was isolated using Qiaamp DNA Blood Mini kit (Qiagen, cat. No. 51104) serum protocol according to the manufacturer’s instructions. The quantity of DNA was measured using the Qubit dsDNA HS Assay kit (Life Technologies). DNA quality was assessed using Agilent 2100 Bioanalyzer with the High Sensitivity DNA kit (Agilent Technologies, Cat. No 5067-4626). Whole-genome sequencing of one random sample from each group and the samples with highest DNA yield from some groups, in total 12 samples, were performed. DNA sequencing libraries were prepared using MicroPlex kits (Diagenode) according to the manufacturer’s instructions. The 12 samples were pooled and sequenced (75 bp single-end reads) on a NextSeq500 (Illumina).

Statistical analyses and bioinformatics

miRNA yield extreme values (>3 SD) were removed from the DNA and miRNA datasets. The Shapiro–Wilk test of normality showed non-normal distribution of DNA and miRNA yields, thus the data were log transformed. The t test was used to test differences in preanalytical treatment and one-way ANOVA was used to test variance of miRNA, percentage of miRNA versus small RNA, and DNA between groups. In conjunction with an ANOVA (post hoc analysis), the Tukey HSD test identified the means that were significantly different from each other.
significantly different from each other. All statistical analyses were performed using the statistical program R (27).

The sequences were filtered for adaptors, primers, and low-quality bases using Nesoni-clip (http://www.vichioinformatics.com/software.nesoni.shtml). miRNA counts were produced by aligning the sequences to the genome (hg19) using Novoalign (http://www.novocraft.com/: version V3.02.00) using the miRNA option (-m). Sequences aligned to miRNA in MiRbase version 20 were counted using FeatureCounts (http://bioinf.wehi.edu.au/featureCounts/, version v1.4.3-p1) allowing multiple matches (28).

DNA sequences were aligned to the human genome (hg19) using bwa (v0.7.10; ref. 29). Sequence coverage comparisons were made using DeepTools (v1.5.9.1; ref. 30), and SNPs were called using the samtools/bcftools package (http://samtools.sourceforge.net; v1.1; ref. 31).

Results

miRNA yields were significantly higher in lyophilized serum

miRNA yields from 500 µL serum ranged from 0.6 to 44 (median of 5.4) ng miRNA/500 µL serum assessed by the Agilent Small RNA kit (see ref. Fig. 1A; one outlier excluded). There is a statistically significant difference in miRNA yields between groups differing in storage time and preanalytical treatment [ANOVA F (6, 6288) = 17.68, P ≤ 0.001]. Comparisons of the means of the groups using the Tukey HSD test, indicated that the group 1 was significantly different from group 7, and group 4B was significantly different from all other groups, with exception of group 7 (Fig. 2A). The sample size of group 4A was too small to draw any conclusions from the comparisons; however, the miRNA yield is similar to the other groups. Ampoules with lyophilized serum (group 4B) produced higher miRNA mean yield than frozen

Figure 1.
Distribution of miRNA yields from 500 µL serum, divided into groups by time of collection and preanalytical handling. A, the boxplot shows the mean and distribution of miRNA yields assessed by Bioanalyzer. B, distribution of the percentage of miRNA yield relative to yield of small RNA (0–150 nt in length) for each group.

Figure 2.
Heatmap of adjusted P values from post hoc analysis of the Tukey HSD test identifying differences in means between the groups. P values lower than 0.01 are colored in dark grey, P values between 0.01 and 0.05 are colored gray, and P values above 0.05 are colored light grey. A, the adjusted P values for miRNA yield differences between groups; B, the adjusted P values for percentage of miRNA versus longer RNA differences between groups; C, the adjusted P value for DNA yield differences between groups.

www.aacrjournals.org Cancer Epidemiol Biomarkers Prev; 24(9) September 2015 OF3

Published OnlineFirst June 24, 2015; DOI: 10.1158/1055-9965.EPI-15-0289

Downloaded from cbp.aacrjournals.org on June 21, 2017. © 2015 American Association for Cancer Research.
serum in non-ampoules supported by an independent-sample t test, \( P \leq 0.05 \). The assumption of homogeneity of variances was not violated, as assessed by the Levene test for equality of variances \( (P = 0.077) \). The effect size, eta squared \( (\eta^2) \), was calculated: \( \eta^2 = 0.07 \) which is considered to be a moderate effect according to the Cohen criteria \( (32) \). The fraction of miRNA \((10–40 \text{ nt})\) of the total small RNA \((0–150 \text{ nt})\), illustrated on the bioanalyzer trace (Supplementary Fig. S1), was also significantly different between the groups \( (\text{ANOVA } F (8.880) = 9.601, P \leq 0.001) \). The comparison of means showed significantly different mean for the HE groups, whereas no differences were detected for RCBD (Fig. 2B). The HE samples with the longest storage time had higher miRNA/small RNA ratio (Fig. 1B) compared with more recent HE samples.

### Successful miRNA sequencing using archived serum samples

Eight miRNA samples, using 500 µL serum as input, were sequenced to assess the miRNA content. The small RNA sequencing produced on average 750,000 miRNA sequences \( (SD = 324,000) \) and on average 648 unique miRNAs \( (SD = 114) \) were identified in the samples (Table 2). There was no correlation between miRNA yield and number of unique miRNA identified. However, the number of unique miRNAs identified depended on sequencing depth (Supplementary Fig. S2). Low amounts of miRNA \( (4.38 \text{ ng}) \) resulted in identification of 724 unique miRNAs. Further analysis of the sequence data revealed that on average, 84% of the uniquely mapped reads were miRNA. Protein coding RNA and other groups of small RNA \((
\text{snoRNA, snRNA, etc.})\) were also identified (Supplementary Fig. S3). The automated miRNA extraction of 96 samples with 200 µL serum input resulted in a mean \( C_t \) value of 29 with a SD of 2.8 using a miR-16-1 assay. Thirty-four samples had \( C_t \geq 30 \) and 89 samples had \( C_t \geq 25 \). A miRNA qPCR panel \((\text{miRNome, QIAgen})\) successfully identified 441 and 461 unique miRNAs \( (C_t < 30) \) in two random selected samples, though production of small RNA sequencing libraries from these samples failed in our hands. Sequence data have been deposited at the European Genome-phenome Archive \((\text{EGA})\), which is hosted by the EBL, under accession number EGAS00001001270, with restricted access.

### Preanalytical treatment influences DNA yields

The median DNA yield was 10.65 ng/500 µL serum \( (SD = 10.99 \text{ ng/500 µL serum}; N = 177) \), excluding three extreme values \( (>3 \text{ SD}) \). The boxplot (Fig. 3A) shows differences between groups confirmed by the ANOVA analyses \( (F (5.835) = 19.365, P \leq 0.01, \text{log transformed}) \). The Tukey HSD test identified a significant higher mean DNA yield in HE groups 2 and 3 compared with RCBD group 5 and fresh serum (Fig. 2C). The mean DNA yield from HE serum samples with iodoacetate added (group 1) was

### Table 2. Overview of miRNA sequencing results from the Janus samples

<table>
<thead>
<tr>
<th>Sample id</th>
<th>Sample</th>
<th>Group</th>
<th>Input (µL)</th>
<th>miRNA yield (ng)</th>
<th>Reads mapping to miRNA</th>
<th>miRNAs (&gt;0 reads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>412</td>
<td>SM3</td>
<td>4B</td>
<td>1000</td>
<td>10.97</td>
<td>1023741</td>
<td>748</td>
</tr>
<tr>
<td>412</td>
<td>SM4</td>
<td>4B</td>
<td>500</td>
<td>4.38</td>
<td>1003809</td>
<td>724</td>
</tr>
<tr>
<td>038</td>
<td>SM1</td>
<td>5</td>
<td>1000</td>
<td>40.89</td>
<td>710476</td>
<td>602</td>
</tr>
<tr>
<td>038</td>
<td>SM2</td>
<td>5</td>
<td>500</td>
<td>26.15</td>
<td>801742</td>
<td>617</td>
</tr>
<tr>
<td>093</td>
<td>SM6</td>
<td>6</td>
<td>1000</td>
<td>18.7</td>
<td>378678</td>
<td>599</td>
</tr>
<tr>
<td>093</td>
<td>SM7</td>
<td>6</td>
<td>500</td>
<td>10.05</td>
<td>100023</td>
<td>439</td>
</tr>
<tr>
<td>214</td>
<td>SM5</td>
<td>3</td>
<td>500</td>
<td>12.65</td>
<td>1104383</td>
<td>763</td>
</tr>
<tr>
<td>Pool</td>
<td>SM8</td>
<td>QA</td>
<td>500</td>
<td>9.62</td>
<td>849473</td>
<td>698</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>16.68</td>
<td>746541</td>
<td>649</td>
</tr>
</tbody>
</table>

NOTE: Sample id, sample name, group, amount serum used in the extraction, and miRNA yield are shown. The sequences were mapped to the human genome \((\text{hg19})\) and sequences mapped to regions of miRbase miRNA were counted. The total counts of unique miRNA are shown.

---

Figure 3.

A. Distribution of DNA yields from 500 µL serum, divided into groups by time of collection and preanalytical handling. B. The miRNA yields correlate with the DNA yields, suggesting that a proportion of the miRNA originates from disrupted cells during preanalytical processing. The correlation between DNA and miRNA yields is statistically significant \( (R^2 = 0.45, P \leq 0.001) \).
Table 3. Overview of DNA sequencing results from the Janus samples

<table>
<thead>
<tr>
<th>ID</th>
<th>Group</th>
<th>Year of collection</th>
<th>DNA yield per 500 μL serum (ng)</th>
<th>Number of reads (millions)</th>
<th>Theoretical read depth</th>
<th>Mean read depth</th>
<th>Mean % SNPs in dbSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>234</td>
<td>1</td>
<td>1972</td>
<td>13.75</td>
<td>1.33</td>
<td>0.03</td>
<td>0.03</td>
<td>34,242</td>
</tr>
<tr>
<td>376</td>
<td>1</td>
<td>1973</td>
<td>187.50</td>
<td>39.32</td>
<td>0.98</td>
<td>0.54</td>
<td>32,771</td>
</tr>
<tr>
<td>341</td>
<td>2</td>
<td>1986</td>
<td>18.90</td>
<td>37.19</td>
<td>0.93</td>
<td>0.52</td>
<td>24,528</td>
</tr>
<tr>
<td>419</td>
<td>2</td>
<td>1985</td>
<td>33.80</td>
<td>32.85</td>
<td>0.82</td>
<td>0.48</td>
<td>27,099</td>
</tr>
<tr>
<td>214</td>
<td>3</td>
<td>1991</td>
<td>24.60</td>
<td>34.30</td>
<td>0.86</td>
<td>0.49</td>
<td>24,728</td>
</tr>
<tr>
<td>066</td>
<td>3</td>
<td>1987</td>
<td>41.40</td>
<td>34.28</td>
<td>0.86</td>
<td>0.50</td>
<td>29,099</td>
</tr>
<tr>
<td>055</td>
<td>4A</td>
<td>1976</td>
<td>7.85</td>
<td>33.50</td>
<td>0.83</td>
<td>0.47</td>
<td>30,867</td>
</tr>
<tr>
<td>304</td>
<td>4B</td>
<td>1974</td>
<td>35.80</td>
<td>36.97</td>
<td>0.92</td>
<td>0.52</td>
<td>29,353</td>
</tr>
<tr>
<td>445</td>
<td>4B</td>
<td>1973</td>
<td>146.00</td>
<td>37.26</td>
<td>0.93</td>
<td>0.52</td>
<td>29,286</td>
</tr>
<tr>
<td>431</td>
<td>5</td>
<td>1981</td>
<td>16.50</td>
<td>31.16</td>
<td>0.78</td>
<td>0.47</td>
<td>21,529</td>
</tr>
<tr>
<td>464</td>
<td>6</td>
<td>2003</td>
<td>16.10</td>
<td>36.47</td>
<td>0.91</td>
<td>0.51</td>
<td>32,742</td>
</tr>
<tr>
<td>369</td>
<td>6</td>
<td>2001</td>
<td>67.00</td>
<td>39.55</td>
<td>0.99</td>
<td>0.54</td>
<td>34,032</td>
</tr>
</tbody>
</table>

NOTE: Number of reads and SNPs called from low depth sequencing of 12 serum samples, all with the input volumes of 500 μL serum, show potential for genome-wide SNP calling with higher depth sequencing.

Successful DNA sequencing using archived serum samples

Twelve archived serum samples yielded sufficient quantity and quality DNA for use in genome sequencing, the oldest collected in 1972. We multiplexed these 12 samples in a single NextSeq500 run to provide low-pass sequence coverage (~0.5×). The sequencing data were sufficient to assess the evenness of coverage, identify any biases due to degradation, and the ability to call SNPs from the data. The average depth of mapped sequences was 0.47 (range, 0.03–0.54; Table 3). Sequence coverage was very even over the whole genome. 99% of the genomes had depth less than 5 (Supplementary Fig. S6A), with no evident bias due to GC content (Supplementary Fig. S6B). The coverage was also similar for all samples (Supplementary Fig. S6C and S6D). However, one sample (234) had considerably lower amount of sequences, for unknown reasons. To assess the ability to call variants from this data, we looked at regions with a depth of ≥5 and could identify on average 29,200 SNPs (Table 3). Of these SNPs, 93.8% are present in dbSNP, showing that the data is of sufficient quality for variant calling.

Discussion

The use of archived serum greatly reduces the time for sample collection in biomarker discovery studies, and long follow-up time may be achieved. However, there are controversies and challenges using archived material, specifically in quantitative studies such as miRNA expression (33–35). Varying preanalytical and analytical conditions both within and between biobanks, hemolysis, effect of storage and low yields from limited material are the main challenges (36). Detailed testing and quality assessment of materials and methods are necessary before and during large studies using serum samples in prediagnostic or prognostic biomarkers discovery. Small input volume protocols and maximized outputs in terms of yields and data produced are preferable when limited volumes of serum are available.

In this study, we show that 500 μL archived serum yield DNA and miRNA of sufficient quality and quantity to produce high-quality DNA and miRNA profiles using next-generation sequencing. The comprehensive cross sectional design, embracing different storage time and preanalytical treatments, gives statistical strength to analyze differences in miRNA and DNA yield between groups. There are no indications of substantial miRNA and DNA degradation during the 40 years of storage, and to our knowledge, this is the first time comprehensive data about effect of storage on miRNA and DNA in serum beyond 10 years are available. The high stability of miRNA and DNA is consistent with earlier observations (14, 19, 37) and also show that molecular analyses using samples archived for a substantial amount of time at ~25°C are possible. In comparison, small RNAs longer than 40 bp are not stable over time, and observed as close to absence of longer RNA in our Bioanalyzer measurements.

We attribute the differences in yield between groups to differences in preprocessing of the serum samples. The lyophilized samples (group 4B) showed increased miRNA yields. This is potentially due to increased cell disruption. The comparison of group 4B with the other groups assumes sufficient accuracy of the rehydration process. The ratio between mir-23a-5p and mir-451a might indicate erythrocyte contamination (38) and may be used as a quality control, as suggested by Blondal and colleagues (33), when full miRNA profiles are obtained. The sequenced lyophilized samples did not differ in mir-23a-5p and mir-451a ratio (data not shown). The addition of iodoacetate in sample preprocessing may affect DNA and miRNA yields. Serum preserved with iodoacetate for glucose analysis has shown systematic biases in analyses of sodium, potassium, chloride, and lactate dehydrogenase (LD; ref. 39). Only a nonsignificant reduction of DNA and miRNA yields in group 1 was observed. Serum separation using gel tubes is comparable with the no-additive protocol. The clotting times in serum preprocessing differed with collection groups and have been shown to affect other serum biomarkers (40, 41). It may also explain some of the variability in yields in DNA and miRNA; however, new experiments are needed to evaluate the effect of clotting time on miRNA and DNA yields in serum. Contamination of intracellular miRNA due to hemolysis can markedly influence the expression profile (2, 36). However, most
miRNAs are non-erythrocyte specific (33). High DNA yield in healthy donors may indicate a cell disruption releasing nucleic acids. In our study, we do find a correlation of DNA and miRNA yield, indicating release of miRNA from cells. Circulating DNA has been shown to be predominantly hematopoietic in origin (42), and may suggest hemolysis. However, the correlation between DNA and miRNA yield is weak and is not a suitable indicator of potential intercellular miRNA.

The Bioanalyzer measurements provide RNA in the range of 10 to 40 nt to miRNA yields, a wider range than considered miRNA volume. The results of the phenol-free and phenol extraction are not influenced by total RNA integrity and ongoing RNA degradation may lead to an overestimation of miRNA amount (43). However, longer unstable RNAs have degraded completely a long time ago in archived samples, thus any overestimation due to on-going degradation will be negligible.

The mRNA sequencing identified a large number of unique miRNAs compared with other studies (10, 44) from all samples sequenced, showing that high-quality miRNA profiles can be achieved from low volumes of serum. However, the number of miRNAs detected in each sample strongly depends on sequencing depth (45). Of note, 500 and 1,000 μL serum start volume did not notably change the diversity of miRNA detected (Supplementary Fig. S2). Small RNA sequencing library preparation of miRNA phenol-free extracted from 200 μL serum was not successful in our hands. This is potentially due to too low serum input or an inefficiency of the phenol-free extraction yielding low amounts of miRNA, supported by the high C_{r} values in the miR-16-1 qPCR assay. The results of the phenol-free and phenol extraction are not directly comparable because the methods differed in input volume.

The trace amounts of DNA from 500 μL serum were sufficient for genome sequencing. There was no bias in coverage across the genome or any bias due to GC content, across the 12 samples sequenced indicating that long-term stored samples could be used for whole-genome sequencing. In addition, even though we sequenced less than 1× coverage, high-quality SNP calls could be made, as shown by the fact that 93.8% of the SNPs we identified were present in dbSNP. The identification SNPs from serum DNA may be difficult. Whole-genome amplification (WGA) may compensate for low yields of DNA. Although WGA is well established, subsequent SNP calling has been shown to be difficult (46). Degradation of DNA as a result of freeze-thaw cycles may cause some of the problems (47). Shotgun sequencing of 12 samples, showing varying degree of apoptotic degradation on the Bioanalyzer trace, was successful and indicated that sequencing is a robust method for genome wide SNP calling of low amounts of DNA. The evenness in sequence distribution (Supplementary Fig. S7) also suggests that these results could be used to identify CNVs using methods developed for low coverage sequencing (48). A limitation of the study is that the sample size in the sequencing experiments is not sufficient to identify the variation in sequencing yield or the "failure rate", if any, should we sequence all samples in the SerumBank. However, the samples sequenced are representative of all preanalytical treatment groups and the result suggests a high success rate for a large-scale sequencing project using Janus SerumBank samples.

The robustness, decreased input requirements and reduced prices of sequencing of today's next-generation sequencing protocols, allows for new use of long-term stored serum samples. Biospecimen cohorts like the Janus Serumbank, including archived samples for up to 40 years, will in the future be favorable sources for biomarker discoveries; however, preanalytical conditions should be taken into account in project design and analyses.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: T.B. Rounge, H. Langseth, E. Enerly, R.E. Gislefoss
Development of methodology: T.B. Rounge, R. Lyle, R.E. Gislefoss
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T.B. Rounge, H. Langseth, R. Lyle
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): T.B. Rounge, M. Lauritzen, R. Lyle, R.E. Gislefoss
Writing, review, and/or revision of the manuscript: T.B. Rounge, M. Lauritzen, H. Langseth, E. Enerly, R. Lyle, R.E. Gislefoss

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T.B. Rounge, M. Lauritzen, H. Langseth
Study supervision: T.B. Rounge, R.E. Gislefoss

Acknowledgments
The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities.

Grant Support
This work was supported by the Norwegian Research Council grant number 229621/H10 to H. Langseth and Cancer Registry of Norway funds to R.E. Gislefoss.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 20, 2015; revised June 8, 2015; accepted June 15, 2015; published OnlineFirst June 24, 2015.

References


microRNA Biomarker Discovery and High-Throughput DNA Sequencing Are Possible Using Long-term Archived Serum Samples

Trine B. Rounge, Marianne Lauritzen, Hilde Langseth, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst June 24, 2015.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-15-0289

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2015/06/25/1055-9965.EPI-15-0289.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.