

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. Pre-analytical 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); 1381–7. ©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 compatibility 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

¹Department of Research, Cancer Registry of Norway, Oslo, Norway.

²Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Current address for T.B. Rounge: Department of Research, Cancer Registry of Norway, Oslo N-0304, Norway.

Corresponding Author: Trine B. Rounge, Cancer Registry of Norway, P.O. Box 5313 Majorstuen, Oslo, Norway. Phone: 47-99-60-43-04; Fax: 47-22-45-13-70; E-mail: trine.rounge@krefregisteret.no

doi: 10.1158/1055-9965.EPI-15-0289

©2015 American Association for Cancer Research.

extraction methods and diagnosis, and mean yield typically range from about 10 to at least 4000 ng DNA/mL serum (23, 24). The quantitative differences in circulating DNA in serum between samples from cancer patients and controls from healthy individuals have been replicated in numerous studies reviewed by Schwarzenbach and colleagues (25).

Archived serum from biobanks is potentially an invaluable source for miRNA and DNA biomarker research. However, characterization of sample quality, quantity, serum processing differences, and long-term storage effects are needed for optimal use. The purpose of the present study is to quantify DNA and miRNA in samples stored for up to 40 years in the Janus Serumbank repository. Furthermore, we will investigate the effects of pre-analytical sample handling and input volumes, and demonstrate that the quality of miRNA and DNA is sufficient for next-generation sequencing.

Materials and Methods

Serum samples grouped by collection time point and preanalytical treatment

The Janus Serumbank is a population-based cancer research biobank containing prediagnostic biospecimens from 318,628 Norwegians, of whom more than 69,000 have developed cancer (<http://www.kreftregisteret.no/en/Research/Janus-Serum-Bank/>). Samples were collected prospectively in the period 1972–2004 from health examinations in Norway (HE) and from Red Cross Blood donors (RCBD) in Oslo and surrounding areas. The samples were stored in freezers at -25°C from 1972. Most of the samples, 86.3% have never been thawed, while 11.1% have been thawed once, 2.2% have been thawed twice, and 0.4% have been thawed three times before included in the project.

miRNA was extracted from 319 samples and DNA was extracted from 177 samples divided by sample collection period and preanalytical treatments in eight groups (Table 1). HE samples from group 1 were collected in 10-mL tubes containing 5 mg sodium iodoacetate, group 2 were collected without additives, and group 3 were collected in gel vials. After coagulation at room temperature and centrifugation, the samples were shipped cold ($+1^{\circ}\text{C}$ – 10°C) to a central facility and frozen within days, depending on transportation route and analyses performed. The RCBD samples have followed the same protocol during the collection period. The tubes carried no additives and were stored overnight before being processed and frozen. Group 4B samples were lyophilized and have been rehydrated before use. Of note, 50 mL whole blood was kept at 4 to 8 hours at room temperature to coagulate. Serum was distributed in vials of 5 mL each and freeze

dried. Rehydration assumes 7% dried material and addition of 93% water (26). Clotting time varied between 14 hours and 28 hours for RCBD samples, whereas the HE samples, and the fresh ones had 1 hour clotting time. Unfortunately, detailed information about clotting time per samples is lacking.

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-PCR control were used to evaluate the yield of miRNA in 96 phenol-free extractions. RNA sequencing was prepared using the ScriptMiner 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

Table 1. Sample groups from the Janus Serumbank repository

Group	Sample collection period	Sample source	Serum processing	Analyzed for miRNA yield (N)	Analyzed for DNA yield (N)
1	1972–1978	HE	Iodoacetate added	44	25
2	1979–1986	HE	No additives	44	25
3	1987–2004	HE	Separating gel tubes	49	25
4A	1973–1979	RCBD	No additives	4	4
4B	1973–1979	RCBD	Lyophilization	42	24
5	1980–1990	RCBD	No additives	48	25
6	1997–2004	RCBD	No additives	45	25
7	2013–2014	Fresh	No additives	43	24
Total				319	177

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 two last columns show the number of samples with measured yield of miRNA and DNA, respectively.

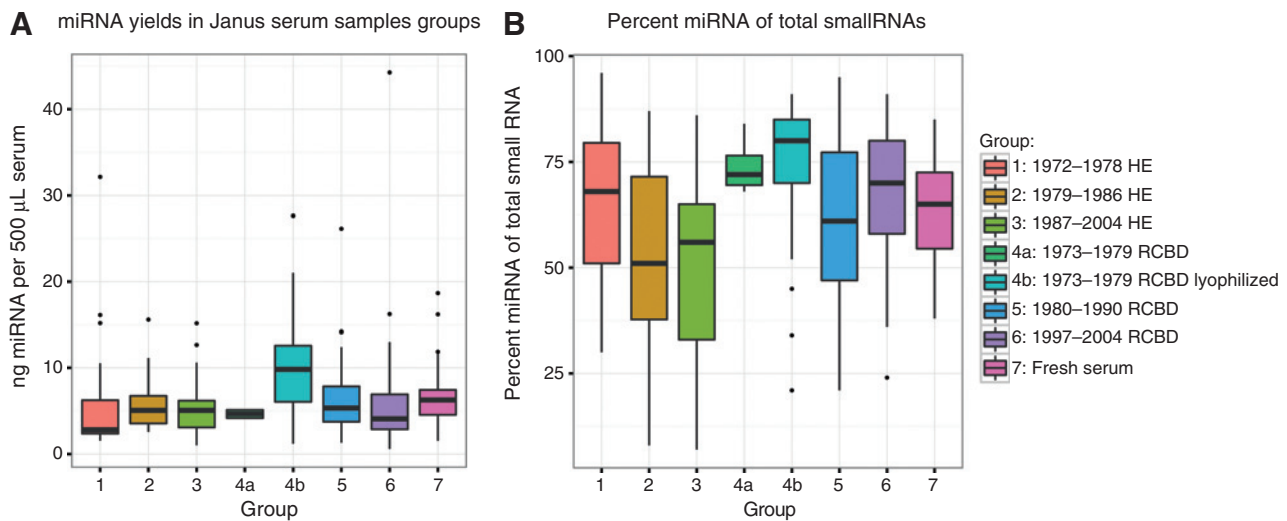


Figure 1. Distribution of miRNA yields from 500 µL serum, divided into groups by time of collection and preanalytical handling. A, the boxplot shows the mean and distribution of miRNA yields assessed by Bioanalyzer. B, distribution of the percentage of miRNA yield relative to yield of small RNA (0–150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

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 Nsoni-clip (<http://www.vicbioinformatics.com/software/nsoni.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 $F(6, 6288) = 17.68, P \leq 0,001$]. 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

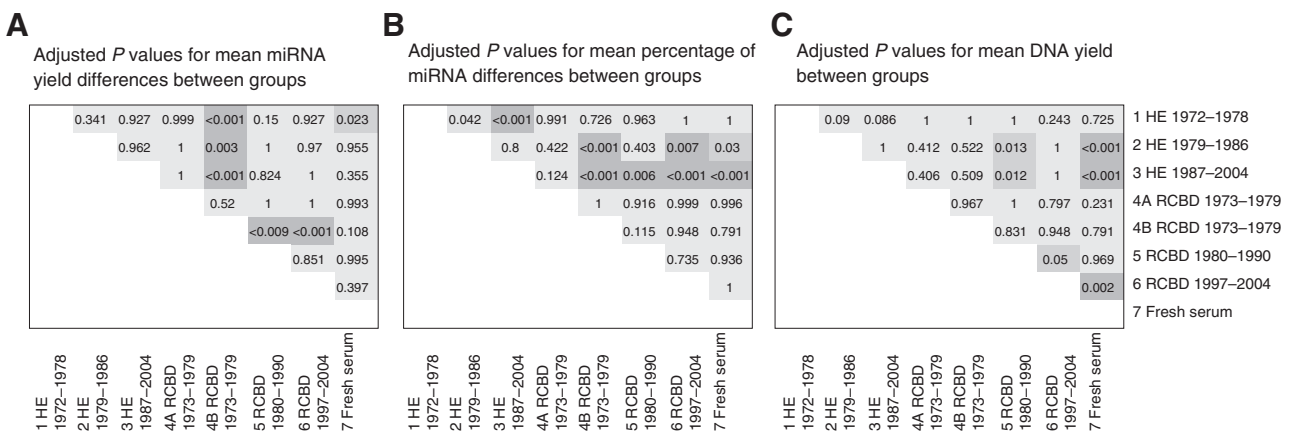


Figure 2. Heatmap of adjusted *P* values from *post hoc* analysis of the Tukey HSD test identifying differences in means between the groups. *P* values lower than 0.01 are colored in dark gray, *P* values between 0.01 and 0.05 are colored gray, and *P* values above 0.05 are colored light gray. A, the 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.

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

Sample id	Sample	Group	Input (μL)	miRNA yield (ng)	Reads mapping to miRNA	miRNAs (>0 reads)
412	SM3	4B	1,000	10.97	1023741	748
412	SM4	4B	500	4.38	1003809	724
038	SM1	5	1,000	40.89	710476	602
038	SM2	5	500	26.15	801742	617
059	SM6	6	1,000	18.7	378678	599
059	SM7	6	500	10.05	100023	439
214	SM5	3	500	12.65	1104383	763
Pool	SM8	QA sample	500	9.62	849473	698
Mean				16.68	746541	649

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 (hg19) and sequences mapped to regions of miRbase miRNA were counted. The total counts of unique miRNA are shown.

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 (η^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 nt) of the total small RNA (0–150 nt), illustrated on the bioanalyzer trace (Supplementary Fig. S1), was also significantly different between the groups [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.

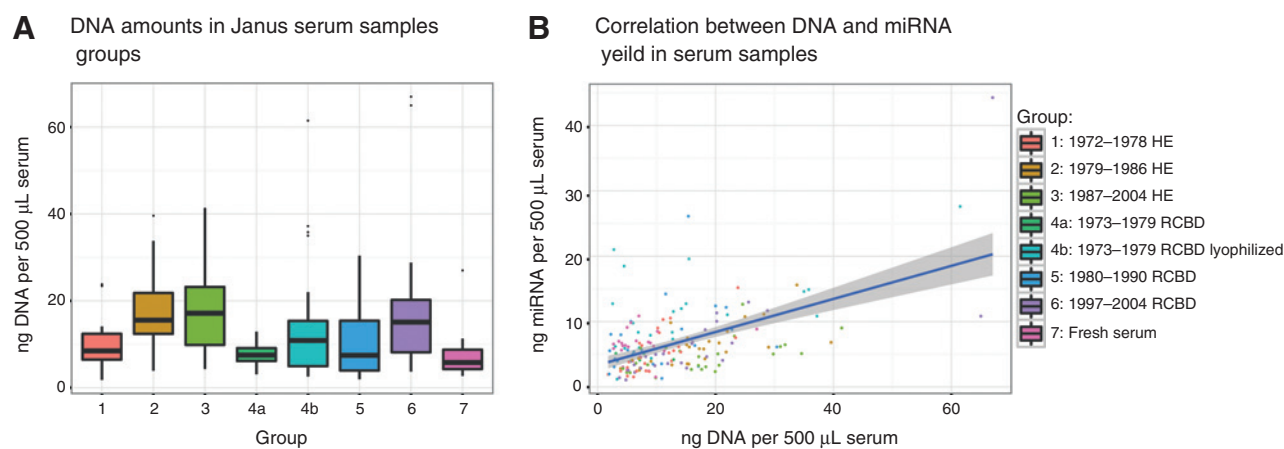
Successful miRNA sequencing using archived serum samples

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 ng) resulted in identification of 724 unique miR-

NAs. 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 (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 (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 (EGA), which is hosted by the EBI, 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 ng/500 μL serum; $N = 177$), excluding three extreme values (>3 SD). The boxplot (Fig. 3A) shows differences between groups confirmed by the ANOVA analyses ($F(5.835) = 19.365$, $P \leq 0.01$, 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

**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

ID	Group	Year of collection	DNA yield per 500 μ L serum (ng)	Number of reads (millions)	Theoretical read depth	Mean read depth	Number of SNPs	% SNPs in dbSNP
234	1	1972	13.75	1.33	0.03	0.03		
376	1	1973	187.50	39.32	0.98	0.54	34,242	94.6
341	2	1986	18.90	37.19	0.93	0.52	32,771	93.6
419	2	1985	33.80	32.85	0.82	0.48	24,528	93.7
214	3	1991	24.60	34.30	0.86	0.49	24,728	93.6
066	3	1987	41.40	34.28	0.86	0.50	27,099	93.7
055	4A	1976	7.85	33.30	0.83	0.47	30,867	92.8
304	4B	1974	35.80	36.97	0.92	0.52	29,535	93.8
445	4B	1973	146.00	37.26	0.93	0.52	29,286	94.5
431	5	1981	16.50	31.16	0.78	0.47	21,529	93.3
464	6	2003	16.10	36.47	0.91	0.51	32,742	92.6
369	6	2001	67.00	39.55	0.99	0.54	34,032	94.0

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

comparable with fresh serum and RCBD samples, although not significantly different to the other groups. The fresh serum (group 7) was significantly different than HE–no additive (group 2 and 3; $P < 0.001$) and the RCBD group 6 ($P = 0.002$). By only taking treatment into consideration regardless of time of storage, there are significant differences between HE samples with and without iodoacetate added, HE–no additive and RCBD and fresh serum, and RCBD and fresh serum (Supplementary Table S1 and Supplementary Fig. S4). The correlation between DNA and miRNA yield is statistically significant ($R^2 = 0.287$, $P \leq 0.001$) shown in Fig. 3B. Most samples showed some apoptotic DNA fragmentation on the Bioanalyzer trace (Supplementary Fig. S5).

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 ($\sim 0.5\times$). 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 μ L 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, give 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 (19–24 nt). The wide range attributed to miRNAs will not affect the relative difference in miRNA yields between groups. The Bioanalyzer method is 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 miRNA 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 μ L 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 μ L 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_t -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 μ L 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 \times 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

- Turchinovich A, Weiz L, Burwinkel B. Extracellular miRNAs: the mystery of their origin and function. *Trends Biochem Sci* 2012;37:460–5.
- Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev Clin Oncol* 2014;11:145–56.
- Ambros V. The functions of animal microRNAs. *Nature* 2004;431:350–5.
- Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001;294:862–4.
- Bartel DP, Lee R, Feinbaum R. MicroRNAs: genomics, biogenesis, mechanism, and function genomics: the miRNA genes. *Cell* 2004;116:281–97.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001;294:853–8.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;36:D154–8.
- Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. *Clin Chem* 2010;56:1733–41.
- Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol* 2008;141:672–5.

10. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* 2008;18:997–1006.
11. Barbarotto E, Schmittgen TD, Calin G a. MicroRNAs and cancer: profile, profile, profile. *Int J Cancer* 2008;122:969–77.
12. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435:834–8.
13. Keller A, Leidinger P, Gislefoss R, Haugen A, Langseth H, Staehler P, et al. Stable serum miRNA profiles as potential tool for non-invasive lung cancer diagnosis. *RNA Biol* 2011;8:506–16.
14. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 2008;105:10513–8.
15. Brase JC, Johannes M, Schlomm T, Haese A, Steuber T, Beissbarth T, et al. Circulating miRNAs are correlated with tumor progression in prostate cancer. *Int J Cancer* 2011;128:608–16.
16. Shen Y, Wang T, Yang T, Hu Q, Wan C, Chen L, et al. Diagnostic value of circulating microRNAs for lung cancer: a meta-analysis. *Genet Test Mol Biomarkers* 2013;17:359–66.
17. Ren A, Dong Y, Tsoi H, Yu J. Detection of miRNA as non-invasive biomarkers of colorectal cancer. *Int J Mol Sci* 2015;16:2810–23.
18. Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Front Genet* 2013;4:116.
19. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* 2011;108:5003–8.
20. Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res* 2011;39:7223–33.
21. Grasedieck S, Schöler N, Bommer M, Niess JH, Tumani H, Rouhi A, et al. Impact of serum storage conditions on microRNA stability. *Leukemia* 2012;26:2414–6.
22. Zhu W, Qin W, Atasoy U, Sauter ER. Circulating microRNAs in breast cancer and healthy subjects. *BMC Res Notes* 2009;2:89.
23. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* 1977;37:646–50.
24. Blömeke B, Bennett WP, Harris CC, Shields PG. Serum, plasma and paraffin-embedded tissues as sources of DNA for studying cancer susceptibility genes. *Carcinogenesis* 1997;18:1271–5.
25. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011;11:426–37.
26. Ørjasæter H. The "Janus" serum research programme: a multipurpose project for retrospective as well as prospective investigation of cancer. *Acta Pathol Microbiol Scand Sect* 1974;248:175–9.
27. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2008.
28. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
30. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map (SAM) format and SAMtools 1000 genome project data processing subgroup. *Bioinformatics* 2009;25:1–2.
32. Pallant J. SPSS Survival Manual. Allen & Unwin Book Publishers; 2010.
33. Blondal T, Jensby Nielsen S, Baker A, Andreassen D, Mouritzen P, Wrang Teilm M, et al. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* 2013;59:S1–6.
34. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet* 2012;13:358–69.
35. McDonald JS, Milosevic D, Reddi HV, Grebe SK, Algeciras-Schimnich A. Analysis of circulating microRNA: preanalytical and analytical challenges. *Clin Chem* 2011;57:833–40.
36. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, et al. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res* 2012;5:492–7.
37. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet* 2001;2:353–9.
38. Rasmussen KD, Simmini S, Abreu-Goodger C, Bartonicek N, Di Giacomo M, Bilbao-Cortes D, et al. The miR-144/451 locus is required for erythroid homeostasis. *J Exp Med* 2010;207:1351–8.
39. Huggins C, Jensen EV. Thermalcoagulation of serum proteins: The effect of iodoacetate and iodoacetamine, and thiol compounds on coagulations. *Clin Chem* 1949;179:645–54.
40. Gislefoss RE, Grimsrud TK, Mørkrid L. Long-term stability of serum components in the Janus Serum Bank. *Scand J Clin Lab Invest* 2008;68:402–9.
41. Zhang DJ, Elswick RK, Miller WG, Bailey JL. Effect of serum-clot contact time on clinical chemistry laboratory results. *Clin Chem* 1998;44:1325–33.
42. Lui YYN, Chik KW, Chiu RWK, Ho CY, Lam CWK, Lo YMD. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem* 2002;48:421–7.
43. Becker C, Hammerle-Fickinger A, Riedmaier I, Pfaffl MW. mRNA and microRNA quality control for RT-qPCR analysis. *Methods* 2010;50:237–43.
44. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 2007;129:1401–14.
45. Metpally RPR, Nasser S, Malenica I, Courtright A, Carlson E, Ghaffari L, et al. Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. *Front Genet* 2013;4:20.
46. Bucacas KL, Pandya GA, Pradhan S, Fleischmann RD, Peterson SN, Belmont JW. Assessing the utility of whole-genome amplified serum DNA for array-based high throughput genotyping. *BMC Genet* 2009;10:85.
47. Croft DT, Jordan RM, Patney HL, Shriver CD, Vernalis MN, Orchard TJ, et al. Performance of whole-genome amplified DNA isolated from serum and plasma on high-density single nucleotide polymorphism arrays. *J Mol Diagn* 2008;10:249–57.
48. Scheinin I, Olshen AB, Van Thuijl HF, Wesseling P, Pinkel D, Albertson DG. DNA copy number analysis of fresh and formalin-fixed specimens by whole-genome sequencing: improved correction of systematic biases and exclusion of problematic regions. *Genome Res* 2014;1–24.