Table 1. Patient Characteristics

	Non-BCAA group	BCAA group	*P* value
n	182	85	
Age (*y*)	67.0 (35–89)	68.0 (31–86)	.817
Sex (female/male)	101/81	44/41	.569
Daily alcohol intake (none/<60 g/≥60 g) (%)	81.0/11.3/7.7	80.0/8.0/12.0	.450
BMI (kg/m^2)	22.9 (14.1–32.9)	23.9 (16.2–36.7)	.137
Platelet count ($\times 10^3/mm^3$)	92.5 (24.0–430.0)	85.0 (29.0–357.0)	.931
AST (*IU/L*)	41 (17–195)	48 (18–198)	.160
ALT (*IU/L*)	35 (7–173)	32 (9–100)	.762
Albumin (*g/dL*)	3.75 (2.31–4.95)	3.30 (1.99–4.20)	<.001
Prothrombin time (international normalized ratio)	1.08 (0.47–1.85)	1.14 (0.55–1.99)	.109
Total bilirubin (*mg/dL*)	1.00 (0.28–2.90)	1.10 (0.30–2.90)	.365
Total cholesterol (*mg/dL*)	157 (77–280)	142 (81–258)	.012
Triglycerides (*mg/dL*)	82 (30–367)	78 (25–233)	.449
Fasting blood glucose (*mg/dL*)	106 (55–280)	107 (73–256)	.360
HbA1c (%)	5.2 (4.0–9.9)	5.2 (4.6–9.5)	.808
Fasting insulin ($\mu U/mL$)	12.0 (2.0–167)	13.3 (3.6–106.3)	.072
HOMA-IR	3.3 (0.4–20.1)	4.0 (0.7–36.9)	.812
Ammonia ($\mu g/dL$)	38 (4–245)	54 (6–326)	<.001
BCAA-to-tyrosine ratio	3.71 (1.36–12.6)	2.87 (1.08–8.26)	<.001
Child–Pugh score	6 (5–10)	7 (5–11)	<.001
BUN (*mg/dL*)	15.0 (3–140)	16.0 (0.6–41)	.886
Creatinine (*mg/dL*)	0.70 (0.37–7.9)	0.72 (0.41–1.73)	.111
Sodium (*mEq/L*)	141 (124–148)	141 (131–146)	.930
Potassium (*mEq/L*)	4.1 (2.5–5.6)	4.2 (3.4–5.1)	.477
Zinc ($\mu g/dL$)	64 (38–116)	61 (21–116)	.210
Iron ($\mu g/dL$)	115 (22–332)	103 (17–274)	.107
TIBC ($\mu g/dL$)	344 (107–563)	328 (174–538)	.608
Ferritin (*ng/mL*)	90.4 (4.3–1624.7)	58.2 (4.0–478.4)	.023
AFP (*ng/mL*)	6.4 (0.6–191.0)	6.4 (1.0–188.6)	.351

NOTE. Data are expressed as number or median (ranges). Differences between the 2 groups were analyzed by using the Wilcoxon rank-sum test. *P* values <.05 were considered significant.

Competing Risk Analysis for the Onset of Hepatocellular Carcinoma and Death

In a multiple Cox regression analysis, serum AFP and TIBC levels were positive risk factors significantly associated with the onset of HCC. An increase in serum BCAA-to-tyrosine ratio was a negative risk factor significantly associated with the onset of HCC. These risk factors were also significantly associated with the onset of HCC in the Fine and Gray analysis (Table 3). Moreover, intake of BCAA supplementation was a negative risk factor significantly associated with the onset of HCC in both the multiple Cox and Fine and Gray analyses ($P = .026$ and $P = .019$, Table 3) after adjusting for other covariates.

Table 2. Cumulative Incidence of Events

	n	All subjects (n = 267)	Non-BCAA group (n = 182)	BCAA group (n = 85)
HCC	52	19.5% (52/267)	22.5% (41/182)	12.9% (11/85)
Death	18	6.7% (18/267)	8.8% (16/182)	2.4% (2/85)

In the multiple Cox regression analysis, the following 6 factors were significantly associated with death: Child–Pugh score, serum BUN level, platelet count, male sex, serum iron level, and HOMA-IR value. However, serum iron level and HOMA-IR value were not significantly associated with death in the Fine and Gray analysis (Table 3). In both the multiple Cox and Fine and Gray analyses, intake of BCAA supplementation was a negative risk factor significantly associated with death ($P = .007$ and $P = .015$, Table 3) after adjusting for other covariates.

Cumulative Incidence of Hepatocellular Carcinoma and Death Between the Branched-Chain Amino Acids and Non–Branched-Chain Amino Acids Groups

Cumulative incidence of the onset of HCC was significantly lower in the BCAA group compared with that in the non-BCAA group (relative risk, 0.45; 95% confidence interval, 0.24–0.88; $P = .019$) (red lines in Figure 1). Cumulative incidence of death was also significantly lower in the BCAA group compared with that in the non-BCAA group (relative risk, 0.009; 95% confidence interval, 0.0002–0.365; $P = .015$) (black lines in Figure 1).

Table 3. Competing Risk Analysis for Onset of HCC and Death

		Cox analysis			Fine and Gray analysis		
Event	Factors	Relative risk	95% confidence interval	*P* value	Relative risk	95% confidence interval	*P* value
HCC	BCAA supplementation	0.43[a]	0.21–0.90	.026	0.45[a]	0.24–0.88	.019
	AFP	1.01	1.00–1.02	.003	1.01	1.01–1.02	<.001
	TIBC	1.01	1.00–1.01	.006	1.01	1.00–1.01	.037
	BCAA-to-tyrosine ratio	0.74	0.56–0.99	.040	0.74	0.57–0.97	.029
Death	BCAA supplementation	0.002[a]	$3.09E^{-05}$–0.11	.007	0.009[a]	0.0002–0.365	.015
	Child–Pugh score	3.81	1.72–8.44	.003	2.78	1.45–5.32	.003
	BUN	1.04	1.02–1.06	<.001	1.03	1.02–1.04	<.001
	Platelet count	0.64	0.46–0.90	.014	0.69	0.53–0.89	.011
	Male sex	0.08	0.01–0.56	.014	0.22	0.05–0.90	.042
	Iron	0.96	0.93–0.98	.002	0.98	0.95–1.01	.246
	HOMA-IR	1.15	1.02–1.30	.029	1.05	0.90–1.23	.569

[a]Relative risk is adjusted by other covariates.

Discussion

There has been an increasing body of evidence to support the potential benefits of BCAA supplementation in cirrhotic patients in terms of reducing the risk of specific complications of liver cirrhosis. However, a BCAA supplementation–derived reduction in risk of overall mortality and/or development of HCC has not been demonstrated. Our study with competing risk analysis yielded several important findings in this regard. Most notably, BCAA supplementation was a significant negative risk factor for mortality. In addition, the present study found that BCAA supplementation significantly reduced the risk of hepatocarcinogenesis. We also made the novel observation that the serum BCAA-to-tyrosine ratio is independently associated with the onset of HCC in cirrhotic patients.

We showed that BCAA supplementation significantly reduced the risk of hepatocarcinogenesis in cirrhotic patients. Our findings are in contrast to those of 2 previous studies in which a preventive effect of BCAA supplementation on hepatocarcinogenesis was not observed.[10,11] Although the reason for this discrepancy is unclear, the following are possible explanations. In an Italian study performed in 1996, the study period was 1–15.5 months,

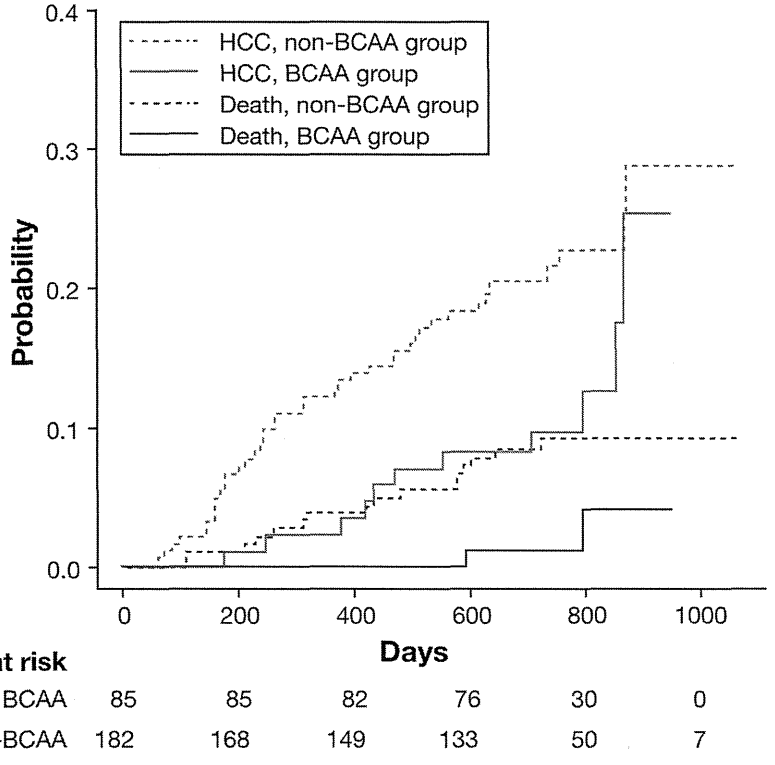


Figure 1. Cumulative incidences of HCC (*red lines*) and death (*black lines*) between BCAA and non-BCAA groups. *Solid lines* indicate the BCAA group, and *dashed lines* indicate the non-BCAA group.

and HCC developed in 5 patients.[10] The low rate of HCC onset would make it nearly impossible to observe a benefit in the treatment arm without a much larger number of events. In a Japanese study conducted in 1997, the obesity rate (BMI >25 kg/m^2) was 28.3%,[11] whereas the obesity rate in our study was 35.0%. BCAA supplementation is known to improve hepatic steatosis and reduce hepatic expression of insulin-like growth factors and the insulin-like growth factor-1 receptor, leading to suppression of obesity-related hepatocarcinogenesis in mice.[21] A previous study showed the preventive effect of BCAA supplementation on hepatocarcinogenesis in cirrhotic patients with obesity.[22] Although the event number is too small to perform a multivariate analysis stratified by BMI in the present study, the higher obesity rate could be a reason for the observed preventive effect of BCAA supplementation on hepatocarcinogenesis.

Another novel aspect of our study is the demonstration of a survival benefit for BCAA supplementation in cirrhotic patients. Chronic hepatic failure is one of the major causes of death in cirrhotic patients, with previous studies observing that hepatic failure is significantly prevented by BCAA supplementation.[10,11] In accordance with these previous reports, our results also suggest that BCAA supplementation prevents chronic hepatic failure and subsequent death. BCAAs are reported to promote the production of hepatocyte growth factor in hepatic stellate cells, asialoscintigraphic removal, and liver regeneration.[12,13] Thus, there are several possible mechanisms by which BCAA supplementation may suppress the progression of liver failure.

Bacterial infection increases mortality 4-fold in cirrhotic patients[23] and is an important issue in the management of cirrhosis.[16] In this study, bacterial infection accounted for 25.0% (4 of 16) of all causes of death and was a major cause of death along with hepatic failure in the non-BCAA group. However, no subject died of bacterial infection in the BCAA group, suggesting that BCAA supplementation prevented bacterial infection and consequent death. The preventive effects of BCAA supplementation on bacterial infection have been reported in liver transplant recipients.[24] In cirrhotic patients, BCAA supplementation increases lymphocyte counts and their natural killer activity[25] and improves the phagocytic function of neutrophils.[1,14] In addition, BCAAs reverse functional impairment and improve the maturation of dendritic cells, leading to an increase in interleukin-12 production.[15] Thus, BCAAs may exert a survival benefit through prevention of bacterial infection by up-regulation of immune functioning.

The final potentially important observation in this study was the predictivity of the BCAA-to-tyrosine ratio for the development of HCC. Decreased serum albumin levels have been reported to be a significant risk factor associated with HCC.[26] Although the serum albumin level was not a risk factor for HCC in this study, the BCAA-to-tyrosine ratio is a significant predictor of a decrease in serum albumin levels in patients with chronic liver disease.[27] In addition, a low BCAA-to-tyrosine ratio is important in the pathogenesis of insulin resistance, which is a known risk factor for HCC.[1] Moreover, our study showed that BCAA supplementation, which increases the BCAA-to-tyrosine ratio, was a negative risk factor for HCC. Taken together, these findings suggest that the BCAA-to-tyrosine ratio may be a useful predictor of the occurrence of HCC in cirrhotic patients.

Because BCAA supplementation has a bitter taste, compliance with BCAA supplementation is generally poor. In this study, the compliance rate was 73.9%, which was about 12% lower than the previous Japanese study (86%).[11] Although the reason for the difference is unclear, a possible reason could be that BCAA granules (12 g/day) were used in the previous Japanese study, whereas the present study used either BCAA granules or BCAA-enriched nutrients. The volume of BCAA-enriched nutrients is 200–600 mL/day, which could impact compliance with BCAA supplementation.

A limitation of this study is that this study is not a randomized controlled trial. Furthermore, the effects of BCAA supplementation on each cause of death remain unclear because of the small number of events that occurred during the study period. Therefore, our results are preliminary and would need to be confirmed by a randomized controlled trial with a much larger number of events.

In summary, this multicenter study showed that the BCAA-to-tyrosine ratio was independently associated with the onset of HCC in cirrhotic patients. In addition, this study demonstrated that BCAA supplementation prevented the onset of HCC and death in cirrhotic patients.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at http://dx.doi.org/10.1016/j.cgh.2013.08.050.

References

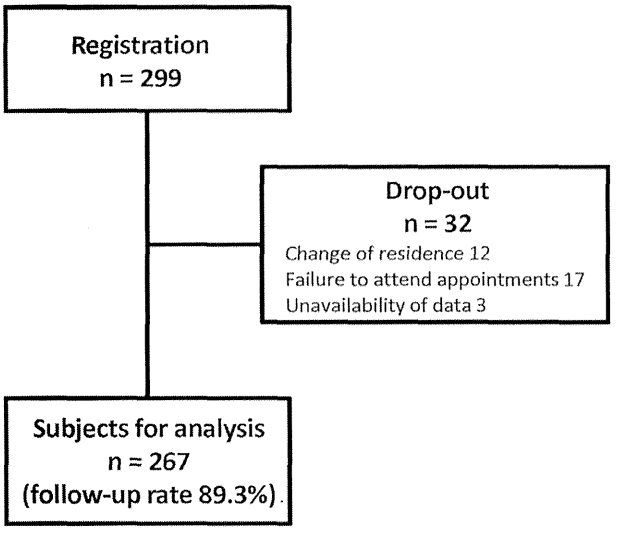
1. Kawaguchi T, Izumi N, Charlton MR, et al. Branched-chain amino acids as pharmacological nutrients in chronic liver disease. Hepatology 2011;54:1063–1070.
2. Charlton M. Branched-chain amino acid enriched supplements as therapy for liver disease. J Nutr 2006;136:295S–298S.
3. Holecek M. Three targets of branched-chain amino acid supplementation in the treatment of liver disease. Nutrition 2010; 26:482–490.
4. Moriwaki H, Miwa Y, Tajika M, et al. Branched-chain amino acids as a protein- and energy-source in liver cirrhosis. Biochem Biophys Res Commun 2004;313:405–409.
5. Hagiwara A, Nishiyama M, Ishizaki S. Branched-chain amino acids prevent insulin-induced hepatic tumor cell proliferation by inducing apoptosis through mTORC1 and mTORC2-dependent mechanisms. J Cell Physiol 2012;227:2097–2105.
6. Ichikawa K, Okabayashi T, Shima Y, et al. Branched-chain amino acid-enriched nutrients stimulate antioxidant DNA repair

in a rat model of liver injury induced by carbon tetrachloride. Mol Biol Rep 2012;39:10803–10810.
7. Shimizu M, Kubota M, Tanaka T, et al. Nutraceutical approach for preventing obesity-related colorectal and liver carcinogenesis. Int J Mol Sci 2012;13:579–595.
8. Nishikawa H, Osaki Y, Iguchi E, et al. The effect of long-term supplementation with branched-chain amino acid granules in patients with hepatitis C virus-related hepatocellular carcinoma after radiofrequency thermal ablation. J Clin Gastroenterol 2013; 47:359–366.
9. Ichikawa K, Okabayashi T, Maeda H, et al. Oral supplementation of branched-chain amino acids reduces early recurrence after hepatic resection in patients with hepatocellular carcinoma: a prospective study. Surg Today 2013;43:720–726.
10. Marchesini G, Bianchi G, Merli M, et al. Nutritional supplementation with branched-chain amino acids in advanced cirrhosis: a double-blind, randomized trial. Gastroenterology 2003;124: 1792–1801.
11. Muto Y, Sato S, Watanabe A, et al. Effects of oral branched-chain amino acid granules on event-free survival in patients with liver cirrhosis. Clin Gastroenterol Hepatol 2005;3: 705–713.
12. Tomiya T, Omata M, Fujiwara K. Branched-chain amino acids, hepatocyte growth factor and protein production in the liver. Hepatol Res 2004;30S:14–18.
13. Koreeda C, Seki T, Okazaki K, et al. Effects of late evening snack including branched-chain amino acid on the function of hepatic parenchymal cells in patients with liver cirrhosis. Hepatol Res 2011;41:417–422.
14. Nakamura I, Ochiai K, Imai Y, et al. Restoration of innate host defense responses by oral supplementation of branched-chain amino acids in decompensated cirrhotic patients. Hepatol Res 2007;37:1062–1067.
15. Kakazu E, Ueno Y, Kondo Y, et al. Branched chain amino acids enhance the maturation and function of myeloid dendritic cells ex vivo in patients with advanced cirrhosis. Hepatology 2009; 50:1936–1945.
16. Fernandez J, Gustot T. Management of bacterial infections in cirrhosis. J Hepatol 2012;56(Suppl 1):S1–S12.
17. Charlton M. Branched-chain amino-acid granules: can they improve survival in patients with liver cirrhosis? Nat Clin Pract Gastroenterol Hepatol 2006;3:72–73.
18. Wai CT, Greenson JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 2003;38:518–526.
19. Faller B, Beuscart JB, Frimat L. Competing-risk analysis of death and dialysis initiation among elderly (≥80 years) newly referred to nephrologists: a French prospective study. BMC Nephrol 2013;14:103.
20. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant 2007; 40:381–387.
21. Iwasa J, Shimizu M, Shiraki M, et al. Dietary supplementation with branched-chain amino acids suppresses diethylnitrosamine-induced liver tumorigenesis in obese and diabetic C57BL/KsJ-db/db mice. Cancer Sci 2010;101:460–467.
22. Muto Y, Sato S, Watanabe A, et al. Overweight and obesity increase the risk for liver cancer in patients with liver cirrhosis and long-term oral supplementation with branched-chain amino acid granules inhibits liver carcinogenesis in heavier patients with liver cirrhosis. Hepatol Res 2006;35:204–214.
23. Arvaniti V, D'Amico G, Fede G, et al. Infections in patients with cirrhosis increase mortality four-fold and should be used in determining prognosis. Gastroenterology 2010;139:1246–1256.
24. Shirabe K, Yoshimatsu M, Motomura T, et al. Beneficial effects of supplementation with branched-chain amino acids on postoperative bacteremia in living donor liver transplant recipients. Liver Transpl 2011;17:1073–1080.
25. Nakamura I, Ochiai K, Imawari M. Phagocytic function of neutrophils of patients with decompensated liver cirrhosis is restored by oral supplementation of branched-chain amino acids. Hepatol Res 2004;29:207–211.
26. Yamada S, Kawaguchi A, Kawaguchi T, et al. Serum albumin level is a notable profiling factor for non-B, non-C hepatitis virus-related hepatocellular carcinoma: a data-mining analysis. Hepatol Res 2013 Jul 2. Epub ahead of print.
27. Suzuki K, Koizumi K, Ichimura H, et al. Measurement of serum branched-chain amino acids to tyrosine ratio level is useful in a prediction of a change of serum albumin level in chronic liver disease. Hepatol Res 2008;38:267–272.

Reprint requests
Address requests for reprints to: Kazuyuki Suzuki, MD, PhD, Division of Gastroenterology and Hepatology, Department of Internal Medicine, Iwate Medical University, Morioka, Japan. e-mail: kasuzuki@morioka-u.ac.jp; fax: +81-652-6664.

Acknowledgments
The authors thank Drs Eitaro Taniguchi, Minoru Itou, Masahiro Sakata, and Kei Sawara for collection of data.

Conflicts of interest
These authors disclose the following: Takumi Kawaguchi has affiliation with a donation-funded department by MSD K.K. Yutaka Kohgo has received research grants/contracts from Mitsubishi Tanabe Pharma Corporation, Chugai Pharmaceutical Co, Ltd, and Novartis Pharma K.K. The remaining authors disclose no conflicts.



Supplementary Figure 1. Study design. A cohort of 299 cirrhotic patients without HCC were enrolled in 2009 and followed up until 2011. In the course of the study, 32 patients dropped out, and the remaining 267 patients were analyzed (follow-up rate, 89.3%).

Supplementary Table 1. Cause of Death and Incidence of Death in Each Group

Cause of death	All subjects (n = 18)	Non-BCAA group (n = 16)	BCAA group (n = 2)
Chronic hepatic failure (%)	27.8 (5/18)	25.0 (4/16)	50.0 (1/2)
Bacterial infection (%)	22.2 (4/18)	25.0 (4/16)	0 (0/2)
Rupture of varices (%)	11.1 (2/18)	12.5 (2/16)	0 (0/2)
Renal failure (%)	11.1 (2/18)	12.5 (2/16)	0 (0/2)
Pancreatic cancer (%)	5.6 (1/18)	6.3 (1/16)	0 (0/2)
Cardiovascular disease (%)	5.6 (1/18)	0 (0/16)	50.0 (1/2)
Trauma (%)	5.6 (1/18)	6.3 (1/16)	0 (0/2)
Unknown (%)	11.1 (2/18)	12.5 (2/16)	0 (0/2)

OPEN ACCESS Freely available online

PLOS | ONE

Comparison of Hepatocellular Carcinoma miRNA Expression Profiling as Evaluated by Next Generation Sequencing and Microarray

Yoshiki Murakami[1]*, Toshihito Tanahashi[2,3], Rina Okada[3], Hidenori Toyoda[4], Takashi Kumada[4], Masaru Enomoto[1], Akihiro Tamori[1], Norifumi Kawada[1], Y-h Taguchi[5], Takeshi Azuma[3]

1 Department of Hepatology, Osaka City University Graduate School of Medicine, Osaka, Japan, **2** Department of Medical Pharmaceutics, Kobe Pharmaceutical University, Kobe, Japan, **3** Division of Gastroenterology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kobe, Japan, **4** Department of Gastroenterology, Ogaki Municipal Hospital, Ogaki, Japan, **5** Department of Physics, Chuo University, Tokyo, Japan

Abstract

MicroRNA (miRNA) expression profiling has proven useful in diagnosing and understanding the development and progression of several diseases. Microarray is the standard method for analyzing miRNA expression profiles; however, it has several disadvantages, including its limited detection of miRNAs. In recent years, advances in genome sequencing have led to the development of next-generation sequencing (NGS) technologies, which significantly advance genome sequencing speed and discovery. In this study, we compared the expression profiles obtained by next generation sequencing (NGS) with the profiles created using microarray to assess if NGS could produce a more accurate and complete miRNA profile. Total RNA from 14 hepatocellular carcinoma tumors (HCC) and 6 matched non-tumor control tissues were sequenced with Illumina MiSeq 50-bp single-end reads. Micro RNA expression profiles were estimated using miRDeep2 software. As a comparison, miRNA expression profiles for 11 out of 14 HCCs were also established by microarray (Agilent human microRNA microarray). The average total sequencing exceeded 2.2 million reads per sample and of those reads, approximately 57% mapped to the human genome. The average correlation for miRNA expression between microarray and NGS and subtraction were 0.613 and 0.587, respectively, while miRNA expression between technical replicates was 0.976. The diagnostic accuracy of HCC, p-value, and AUC were 90.0%, 7.22×10^{-4}, and 0.92, respectively. In summary, NGS created an miRNA expression profile that was reproducible and comparable to that produced by microarray. Moreover, NGS discovered novel miRNAs that were otherwise undetectable by microarray. We believe that miRNA expression profiling by NGS can be a useful diagnostic tool applicable to multiple fields of medicine.

Citation: Murakami Y, Tanahashi T, Okada R, Toyoda H, Kumada T, et al. (2014) Comparison of Hepatocellular Carcinoma miRNA Expression Profiling as Evaluated by Next Generation Sequencing and Microarray. PLoS ONE 9(9): e106314. doi:10.1371/journal.pone.0106314

Editor: Max Costa, New York University School of Medicine, United States of America

Received March 10, 2014; **Accepted** July 29, 2014; **Published** September 12, 2014

Funding: The authors have no support or funding to report.

Competing Interests: The authors declare that they have no competing interests.

* Email: m2079633@med.osaka-cu.ac.jp

Introduction

MicroRNAs (miRNAs) are an abundant class of small (19–24 nt) and highly conserved, non-coding RNA. They act as post-transcriptional regulators of gene expression, altering mRNA transcription and translation by hybridizing to the untranslated regions (UTRs) of certain subsets of mRNAs [1] [2]. Since their initial discovery in *Caenorhabditis elegans* in 1993 [3], researchers have gained much insight into the prevalence of miRNAs in other species. The latest miRBase database (release 20) contains 1827 precursor miRNAs and 2578 mature miRNA products in *Homo sapiens* (http://www.mirbase.org/index.shtml).

Hepatocellular carcinoma (HCC) is a common cause of cancer-related deaths worldwide. There are more than 250,000 new HCC cases and an estimated 500,000–600,000 HCC deaths annually [4] [5]. The most frequent etiology of HCC is chronic hepatitis B and C (CHB, CHC), or alcoholic liver disease. Although recent advances in functional genomics provide a deeper understanding of viral associated hepatocarcinogenesis (review in [6]), the molecular pathogenesis of HCC remains unclear.

Altered miRNA expression has been observed in a large variety of HCC and a correlation has been found between miRNA expression and histological differentiation [7] [8]. For example, the expression level of miR-26 has been associated with hepatocarcinogenesis and response to interferon therapy [9]. Moreover recently, miR-122 expression was associated with hepatocarcinogenesis, liver homeostasis, and essential liver metabolism [10] [11]. miR-18 has also been highly associated with the occurrence and progression of different types of cancer [12] [13]. In other research, miRNA expression profiles were associated with vascular invasion, the levels of alpha-fetoprotein, and large tumor size [14].

To date, studies exploring the role of miRNAs in hepatocarcinogenesis have relied on microarrays to assay miRNA expression. Deep sequencing, a set of technologies that produce large amounts of sequence data from nucleic acid specimens, is rapidly replacing microarrays as the technology of choice for quantifying and annotating miRNAs [15] [16]. Deep sequencing has the superior ability to capture the scale and complexity of whole transcriptomes [17]. In particular, short read deep sequencing

(e.g., the Illumina MiSeq platform) is appropriate for miRNAs because a complete miRNA can be sequenced with a single read. While array design relies on prior knowledge of the miRNAs being investigated, deep sequencing allows for the discovery of novel miRNAs. Furthermore, microarray methods lack the dynamic range to detect and quantify low abundance transcripts, but deep sequencing can identify miRNAs that are expressed at levels that fall below microarray's detectable threshold. In addition, deep sequencing eliminates background problems that result from cross-hybridization in microarrays, thus facilitating interpretation of the signal and obviating the non-linear data manipulation steps required by microarrays. Therefore, the application of deep sequencing to miRNA profiling has the potential to uncover novel miRNAs and to detect expression of rare but functionally significant miRNAs. Recently, deep sequencing was used to analyze non-coding RNAs in HCC, by which miRNAs, PIWI-interacting RNA, and small nucleolar RNAs were identified [18].

In this study, we created miRNA expression profiles for HCC and non-tumorous tissue using NGS. We then compared the miRNA expression profiles obtained by NGA and microarray. Unlike previous studies, we sequenced un-pooled miRNA libraries to a previously unprecedented sequencing depth from multiple replicates and controls across multiple time-points, allowing us to explore the statistically significant temporal changes in miRNA expression in hepatocarcinogenesis.

Materials and Methods

Sample preparation

We isolated total RNA from 14 surgically resected HCC tumors and 6 matched adjacent non-tumor control tissues (Table 1). We confirmed that the 14 samples were accurately diagnosed as HCC by image diagnosis by CT and USTG and pathological findings. All patients or their guardians provided written informed consent, and Osaka City University and Ogaki Municipal Hospital approved all aspects of this study in accordance with the Helsinki Declaration.

RNA preparation, and miRNA deep sequencing and microarray

Total RNA from surgical resection was prepared using mirVana miRNA Isolation kit (Invitrogen), according to the manufacturer's instruction.

Total RNA, containing the small RNA fraction, was reverse transcribed into a cDNA library using the TruSeq Small RNA Sample Prep Kit (Illumina). Briefly, total RNA (1 μg per sample) was ligated overnight with adapters, reverse transcribed, RNase-treated, and PCR-amplified with unique barcode-labeled amplification primers. Then, size-selection was conducted on 6% native polyacrylamide gels. cDNA fragments between 145 and 160 bp corresponding to the miRNA populations were excised from the gel, then eluted and precipitated. The final cDNA pellet was air dried and resuspended in 10 μl of nuclease-free water. The quantity of cDNA in each final miRNA libraries was verified using Qubit fluorometer (Invitrogen). Equimoloar amounts for each final library were pooled at a final concentration of 2 nM cDNA. Barcoded templates were sequenced on a single flowcell of the Illumina MiSeq with 50-bp single-end reads. Eleven of 14 total HCC RNA samples were also assayed by microarray (Agilent human microRNA microarray release 14.0) (Table 1). Hybridization signals were detected with a DNA microarray scanner G2505B (Agilent Technologies) and the scanned images were analyzed using Agilent feature extraction software (v9.5.3.1). Raw data (gProcessedSignal) was normalized so that each expression had a mean of zero and a sample variance of one. The above processes were conducted with various packages and functions implemented in R {http://www.r-project.org}.

The sequence reads obtained in this study have been deposited in the DNA Data Bank of Japan Sequence Read Archive (http://www.ddbj.nig.ac.jp/index-e.html) under accession number DRA001067. All microarray data were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE31164.

Bioinformatics

In order to extract the adaptor sequence from each short read obtained by NGS, fastx_clipper from the fastx toolkit {http://hannonlab.cshl.edu/fastx_toolkit/} was used. The adaptor trimmed short reads were then mapped to the human reference genome sequence hg19 by mapper.pl script included in miRDeep2 [19] {http://genomewiki.ucsc.edu/index.php/Hg19_Genome_size_statistics}. miRNA mature and hairpin sequences were obtained from miRBase release 18 {http://www.mirbase.org}. Finally, the resulting fastq files were processed by miRDeep2.pl script. miRNA read counts were extracted from the "read_count" column of the file named"miRNAs_expressed_all_samples_sample_id.csv" file; while sample_id was given automatically by scripts (for sample scripts see Text S1 and for extracted read counts for each miRNA see Table S1). When drawing boxplots, the read count in each sample was normalized such that it had a zero mean and a variance of one. For microarray processing, gProcessedSignal values were extracted from raw data files. gProcessedSignal values were also normalized so that they had a zero mean and a variance of one. Average gProcessedSignal values over probes assigned to common mature miRNA were used to compute the correlation coefficient for NGS and microarray results. The correlation coefficients of logarithmic expression and subtracted expression were computed using only miRNAs with non-negative signals in both NGS and average microarray expression. P-values associated with boxplots were computed using Wilcoxon rank sum test. All bioinformatics computation was performed using functions implemented in R.

To discriminate HCC from non-tumorous tissue when miRNA expression was quantified by NGS, we combined principal components analysis (PCA)-based feature extraction and PCA-based linear discriminant analysis (LDA) [20] [21]. When PCA-based feature extraction was applied, each miRNA was embedded into a two-dimensional space by PCA and M miRNAs located far from the origin (outliers) were selected. Using these selected miRNA, each sample was embedded into a low dimensional space with dimension (M–M'). Samples were then discriminated by LDA using the M' dimensional PC scores. For more details, see Text 2. Novel miRNA candidates were selected from the total set if they satisfied the following criteria: 1) among the novel miRNAs identified by miRDeep2, those with a >80% probability of being a true positive, and 2) the miRNA was reproducibly detected in more than three samples.

Results

Analysis of miRNA sequence reads and reproducibility of NGS analysis

The average number of sequencing reads per sample exceeded 2.2 million, of which approximately 57% mapped to the human reference genome (for more details, see Table S2). A scatter plot of logarithmic miRNA expression measured by NGS and microarray using the first technical replicate of K-177 (K-177_1) is shown in Fig 1.

Table 1. Clinical information of 14 HCC patients analyzed in this study.

code No.	histological differentiation	sex	age	HBsAg	anti-HCV	AFP	PIVKA-II	no tumorous tissue	NGS analysis	microarray analysis
CU-070	moderately	M	53	(−)	(+)	3510	NI	LC	only HCC	only HCC
CU-083	moderately	M	49	(+)	(−)	143830	0.9	LC	both	only HCC
CU-085	moderately	M	55	(−)	(+)	113	0.06	LC	only HCC	only HCC
CU-087	well	M	62	(−)	(+)	NI	NI	NI	both	only HCC
CU-089	moderately	F	61	(−)	(+)	400	0.06	CH	both	only HCC
CU-091	moderately	M	64	(−)	(+)	5	0.06	CH	only HCC	only HCC
K-023	moderately	M	47	(+)	(−)	3500	2.2	LC	only HCC	only HCC
K-147	poorly	M	65	(−)	(+)	NI	NI	NI	only HCC	neither
K-175	moderately	M	67	(−)	(+)	3.4	32	CH	both	neither
K-177	moderately	M	68	(−)	(+)	13	29	NI	both	only HCC
K-181	well	M	LC	(−)	(−)	5.1	19	NI	both	neither
O-086	moderately	F	64	(−)	(+)	218	32	LC	only HCC	only HCC
O-088	moderately	M	74	(+)	(−)	13.4	7991	LC	only HCC	only HCC
O-089	moderately	M	68	(−)	(+)	8	25	LC	only HCC	only HCC

Abbreviations, moderately; moderately differentiated HCC, well; well differentiated HCC, poorly; poorly differentiated HCC, NI; no information, CH; chronic hepatitis, LC; liver cirrhosis, only HCC; only HCC was analyzed by NGS, both; both tumor and non-tumorous tissue were analyzed by NGS, neither; neither HCC nor non-tumorous tissue was not analyzed.

doi:10.1371/journal.pone.0106314.t001

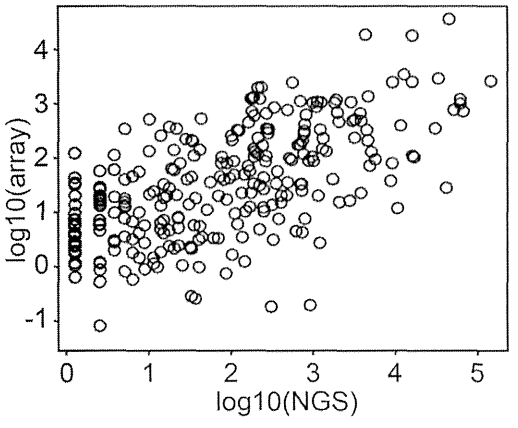


Figure 1. Comparison between logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis (K-177_1 means the first technical replicates of code No. K-177). One black circle showed one miRNA. Pearson's correlation coefficient is 0.6059.
doi:10.1371/journal.pone.0106314.g001

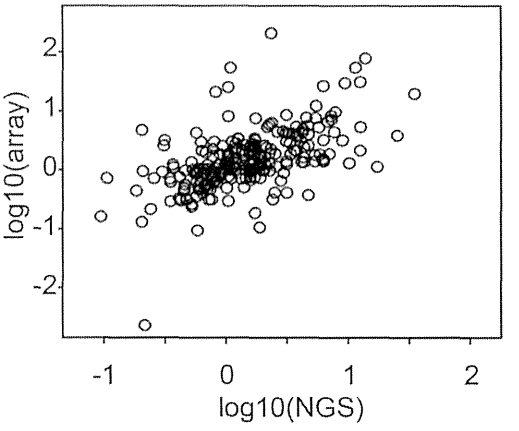


Figure 2. Comparison between differential (K-177 vs. CU-087) logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis. Pearson's correlation coefficient is 0.5555.
doi:10.1371/journal.pone.0106314.g002

Although the microarray and NGS expression levels are not perfectly correlated, they are approximately proportional to each other with a positive proportional coefficient. The corresponding scatter plots for the remaining 9 HCC samples are shown in Fig. S1. The similarity in the NGS and microarray miRNA profiling results was relatively independent of the samples considered. The correlation coefficient of logarithmic miRNA expression from 11 HCC miRNA expression profiles as measured by both NGS and microarray was 0.613. This demonstrates that NGS and microarray measurements give similar results.

We next validated the reproducibility of miRNA differential expression across the 11 samples. In general, miRNA expression profiles are not individually evaluated; instead, profiles for distinct samples are analyzed in pairs e.g., compared between tumor and adjacent non-tumorous tissue. Thus, reproducibility between NGS and microarray is more important in differential expression than in any single expression measurement by itself. Fig. 2 shows a scatter plot of differential (K-177 vs. CU-087) logarithmic miRNA expression obtained by NGS and microarray (Fig. S2 shows the full set of scatter plots). Again, we observed that the coincidence between NGS and microarray miRNA profiles was relatively strong irrespective of the sample pair that was considered. The correlation coefficient of differential logarithmic miRNA expression averaged over all pairs of 11 HCC is 0.587. This demonstrates a reasonable level of congruency between miRNA profiling results from NGS and microarray when considering differential logarithmic miRNA expression.

miRNA expression measured by NGS can be applied to diagnosing HCC

In order to discriminate between HCC and non-tumorous tissue, 11 miRNAs quantified by NGS (miR-10a-5p, miR-122-5p, miR-146b-5p, miR-148a-3p, miR-192-5p, miR-22-3p, miR-26a-5p, and miR-27b-3p, miR-10b-5p, miR-143-3p, and miR-21-5p) were chosen by PCA-based feature extraction. The miRNA expression levels in HCC and non-tumorous tissue is shown in Fig 3. The expression level of miR-10a-5p ($p<2.56\times10^{-2}$), miR-122-5p ($p < 1.55\times10^{-3}$), and miR-22-3 ($p<4.64\times10^{-3}$) among the 11 miRNAs differed significantly in the HCC and non-tumorous tissue. miRNA profiling allowed the accurate prediction of HCC with an overall cross-validation accuracy of 90.0% (18/20) by PCA-based feature extraction (Table 2). The p-value and AUC value for diagnostic ability were $<7.22\times10^{-4}$ and 0.92 respectively.

Reproducibility of NGS measurement among technical replicates

We have demonstrated that quantifying miRNA using NGS gives results similar to those obtained by microarray, and that miRNA expression measured by NGS can discriminate HCC from non-tumorous tissue. However, we have found that miRNA profiling using NGS is more accurate in cases where NGS measurement does not vary within multiple technical replicates. For those HCC samples in this study that had more than one technical replicate we validated the reproducibility of NGS miRNA expression between technical replicates. Fig. 4 shows examples of technical replicates (other scatter plots are available in Fig. S3). Among three technical replicates, the correlation coefficients of logarithmic miRNA expression are greater than 0.98. Additionally, the dynamic range is almost 5 digits. This means that technical replicates obtained from NGS measurements are highly reproducible.

Discovery of novel miRNAs in our analysis

NGS detected several miRNA candidates that are not registered in the present miRBase (Rel. 18) (see Materials and Methods for detection criteria, and see Supporting Information for more details about detected miRNAs). We speculate that four precursor miRNAs, hsa-mir-9985, hsa-mir-1843, hsa-mir-548bc, and hsa-mir-9986 and the corresponding four mature miRNAs, hsa-miR-9985, hsa-miR-1843, hsa-miR-548bc, and hsa-miR-9986, were not previously reported because they are not among the miRNAs found at the corresponding genomic coordinates. Fig. 5 shows the sequence of the new miRNA candidates, their alignment with their closest homologous miRNA, and the hairpin structure predicted by RNA-fold with default parameter settings (short reads mapped to these candidate miRNAs are available in Table S3).

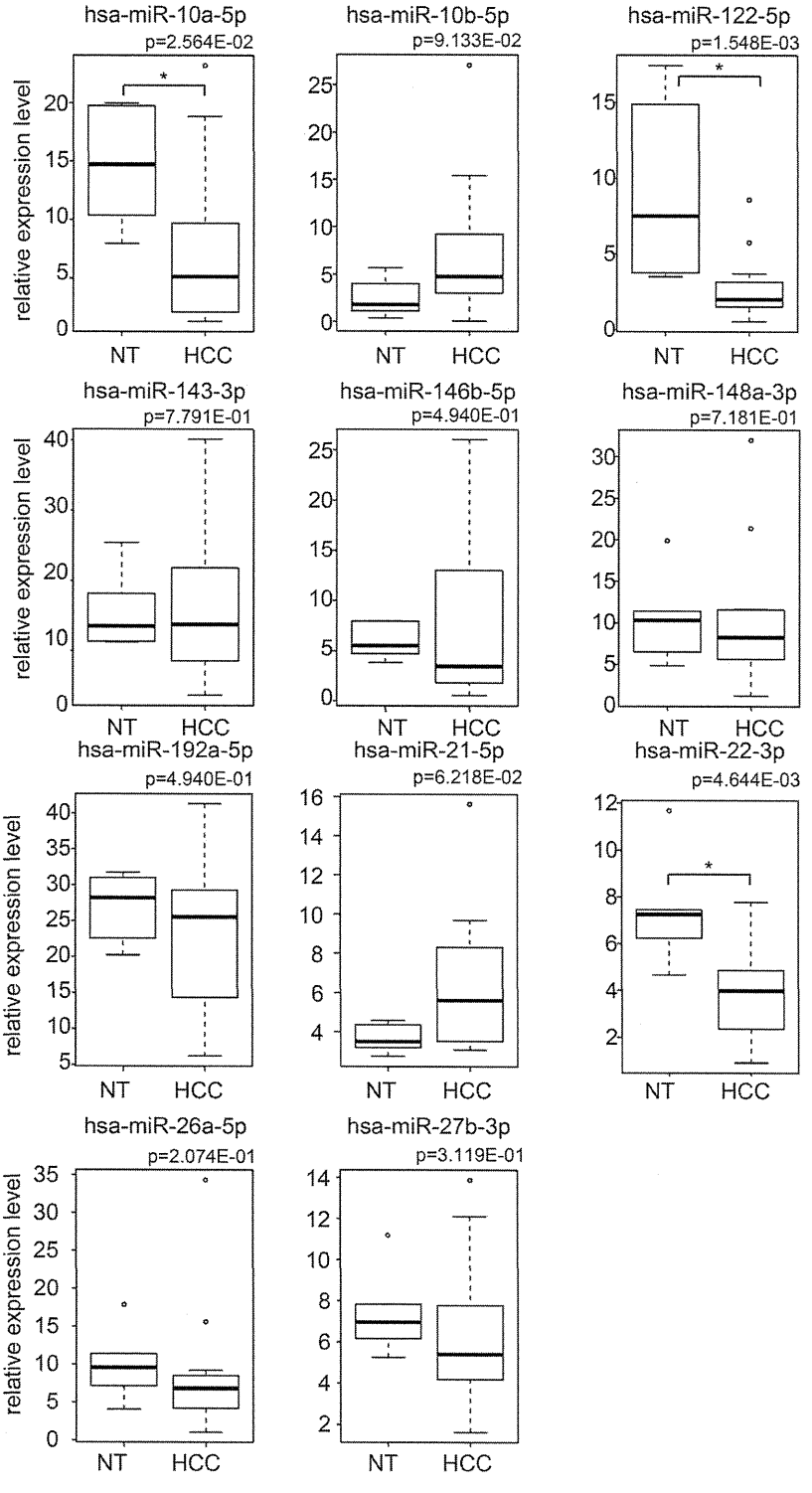


Figure 3. Boxplots of the expression of 11 miRNAs in HCC and non-tumorous tissue obtained by NGS, which were used for the differential analysis. P-values were computed using two-sided Wilcoxon Rank Sum test. Asterisk indicates a significant difference of $p<0.05$ (*).
doi:10.1371/journal.pone.0106314.g003

Table 2. Performance of discrimination between 14 HCC samples and 6 normal tissue samples using miRNA expression obtained by NGS analysis.

		Result	
		Control	Tumor
Prediction	Control	12	0
	Tumor	2	6

doi:10.1371/journal.pone.0106314.t002

Discussion

The clinical application of miRNA expression profiling, such as its use as a disease biomarker, has been extensively developed in recent years. This analysis demonstrates that miRNA profiling by NGS has the potential to diagnose HCC with high accuracy. Previous comprehensive analyses of miRNA expression have been performed by microarray; however, the miRBase is currently underdeveloped. An updated miRBase is required to create accurate microarray profiles, and especially because microarray experiments performed using previous version of miRBase are incompatible with microarrays based on the current version. The results from our miRNA expression analysis in HCC suggest that NGS may allow us to overcome this problem. Law et al. has previously reported small RNA transcriptome analysis in HCC by using NGS [18]. However, pre-miRNAs are of a similar length as other more numerous classes of ncRNA, including tRNA and snoRNA, making deep profiling of pre-miRNA sequences difficult [22]. Therefore, this study focused on the analysis on whole miRNA instead of ncRNA.

In order to determine if miRNA expression as measured by NGS technology can discriminate HCC from non-tumorous tissue, we adopted a recently proposed combination of PCA-based feature extraction and PCA-based LDA [20] [21] (for details, see Materials and Methods). Previously, we showed that miRNA expression profiles detected by microarray can accurately discriminate HCC from non-tumorous tissues [7] [23] [24] [25]. Therefore, because NGS produced results similar to those by microarray, and was capable of differentiating HCC from non-tumorous tissues, it is evident that miRNA expression quantified by NGS is as informative as is miRNA expression assessed by microarray. NGS's ability to quantify miRNA expression suggests

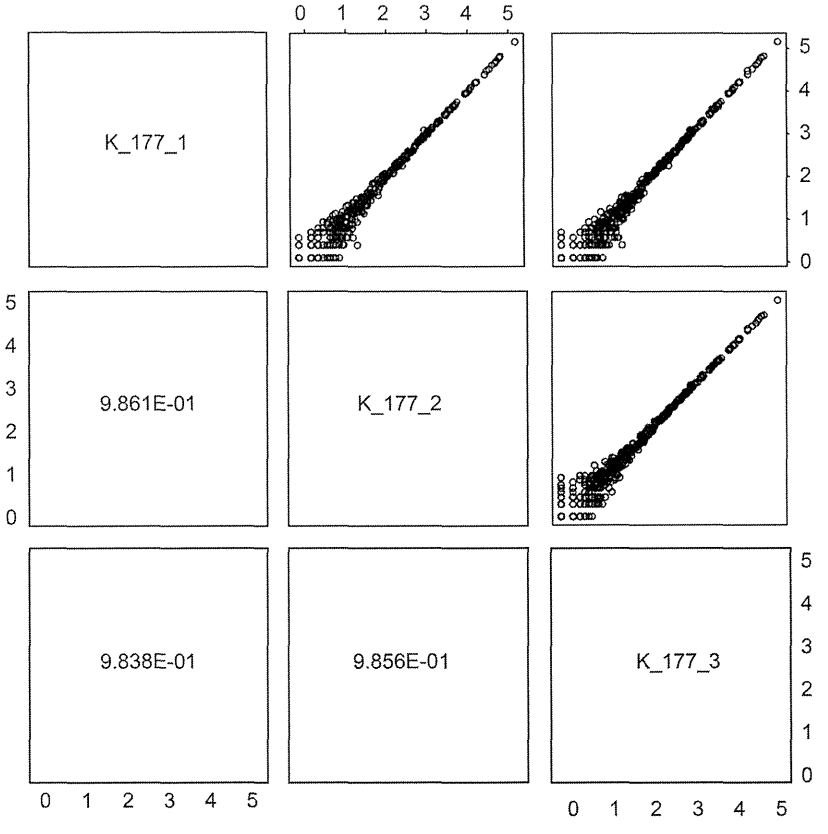


Figure 4. Comparison between logarithmic miRNA expression in HCC and NGS technical replicates (K-177_1 means the first technical replicates of code No. K-177). Pearson's correlation coefficients are greater than 0.9861.
doi:10.1371/journal.pone.0106314.g004

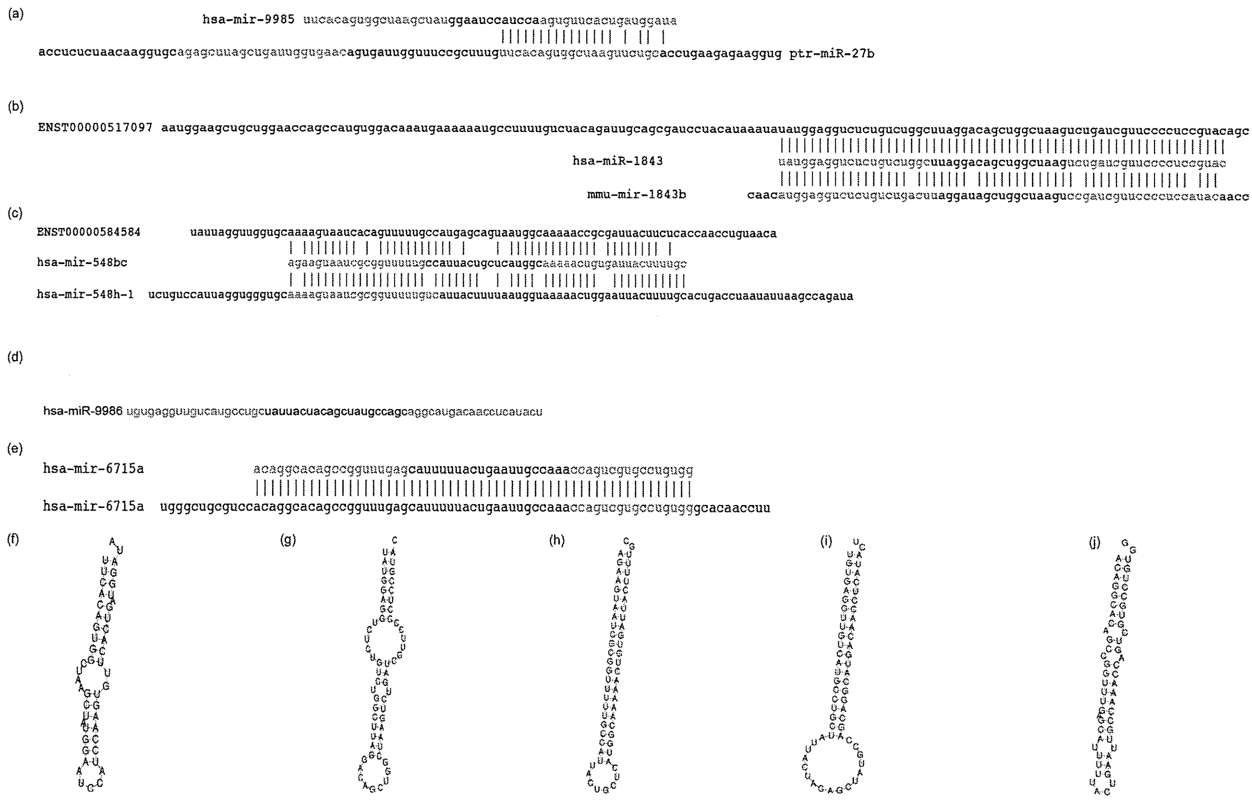


Figure 5. Comparison of novel miRNA candidates ((a) hsa-mir-9985, (b) hsa-mir-1843, (c) hsa-mir-548bc and (d) hsa-mir-9986) with known miRNAs and transcripts. (e) hsa-mir-6715a novel miRNA was not included in miRBase release 18, but was present in later releases. Nucleotides highlighted in red show hairpin constructs. F, g, h, i, and j are constructs of each miRNA and respectively correspond to a, b, c, d, and e.
doi:10.1371/journal.pone.0106314.g005

that the miRNA expression profile measured by NGS can be clinically applicable as a diagnostic tool.

HCC was discriminated from non-tumorous tissues using a linear combination of 11 miRNA expression profiles; however, it may be useful to investigate the expression of individual miRNA in order to understand their biological significance. Fig. 3 shows boxplots of the 11 miRNAs that were used to discriminate between HCC and non-tumorous tissue. Contrary to our expectations, only three (miR-10a-5p, 122-5p and 22-3p) of the 11 miRNAs showed significant differential expression ($p<0.05$) between HCC and non-tumorous tissue. Among these three miRNAs, miR-122-5p and miR-22-3p are well-known to be downregulated in HCC [10] [10] [26] [27]. Among the three miRNAs with significant differential expression between HCC and non-tumorous tissue, miR-10a-5p showed the least difference, with a p-value of 0.03. Recently, it has been reported that miR-10a-5p is involved in HCC metastasis [28]. Thus, the selection of these three miRNAs as potential biomarkers is biologically reasonable. Although the remaining eight miRNAs did not have significant differential expression between HCC and non-tumorous tissues, their inclusion did not reduce the discriminatory performance. This suggests that these eight miRNAs may contribute to diagnosing hepatocarcinogenesis only when combined with other miRNAs. Since controlling the expression of multiple miRNAs simultaneously is experimentally difficult, we were unable to confirm if these miRNAs work together. This may be worth further study in the future.

While investigating novel miRNA candidates, we found that mature miRNAs from hsa-mir-9985 and hsa-miR-9985-5p are almost identical to ptr-miR-27a-3p (Fig 5(a, f)). This may suggest that hsa-mir-9985 has been transposed during its evolution from chimpanzee to human. hsa-miR-1843 was homologous to miR-1843a, which was previously detected in mouse, Chinese hamsters, and brown rats. Near the genomic coordinate where hsa-mir-1843 was detected, Ender reported finding snoRNA (ENSR00000517097), therefore hsa-mir-1843 is likely to be a small nuclear RNA-derived microRNA [29] (Fig, 5(b, g)). It is possible that this explains why researchers have overlooked hsa-mir-1843 despite reports of homologous miRNAs in other animals. While hsa-mir-548bc appears homologous to hsa-mir-548h-1 (Fig. 5(c, h)), hsa-mir-548bc is detected at a distinct genomic coordinate from the original location of hsa-548h-1. Thus, hsa-miR-548bc seems to be a distinct microRNA from hsa-mir-548h-1.

It is interesting to note that the novel miRNA (EST00000584584) not included in the miRBase was reported to be located near the genomic coordinate where hsa-mir-9986 was detected. To our knowledge, this is the first experimental evidence that these predicted novel miRNAs exist. Finally, no homologous miRNA has been reported for hsa-mir-9986 (Fig. 5(d, i)), thus it is possible that hsa-mir-9986 is completely new. In addition to these four new miRNA candidates, hsa-mir-6715a was also detected in our analysis (Fig. 5(e, j)). hsa-miR-6715a was not included in the miRBase release 18, but was included in later releases, which lends support to the reliability and accuracy of our strategy to identify

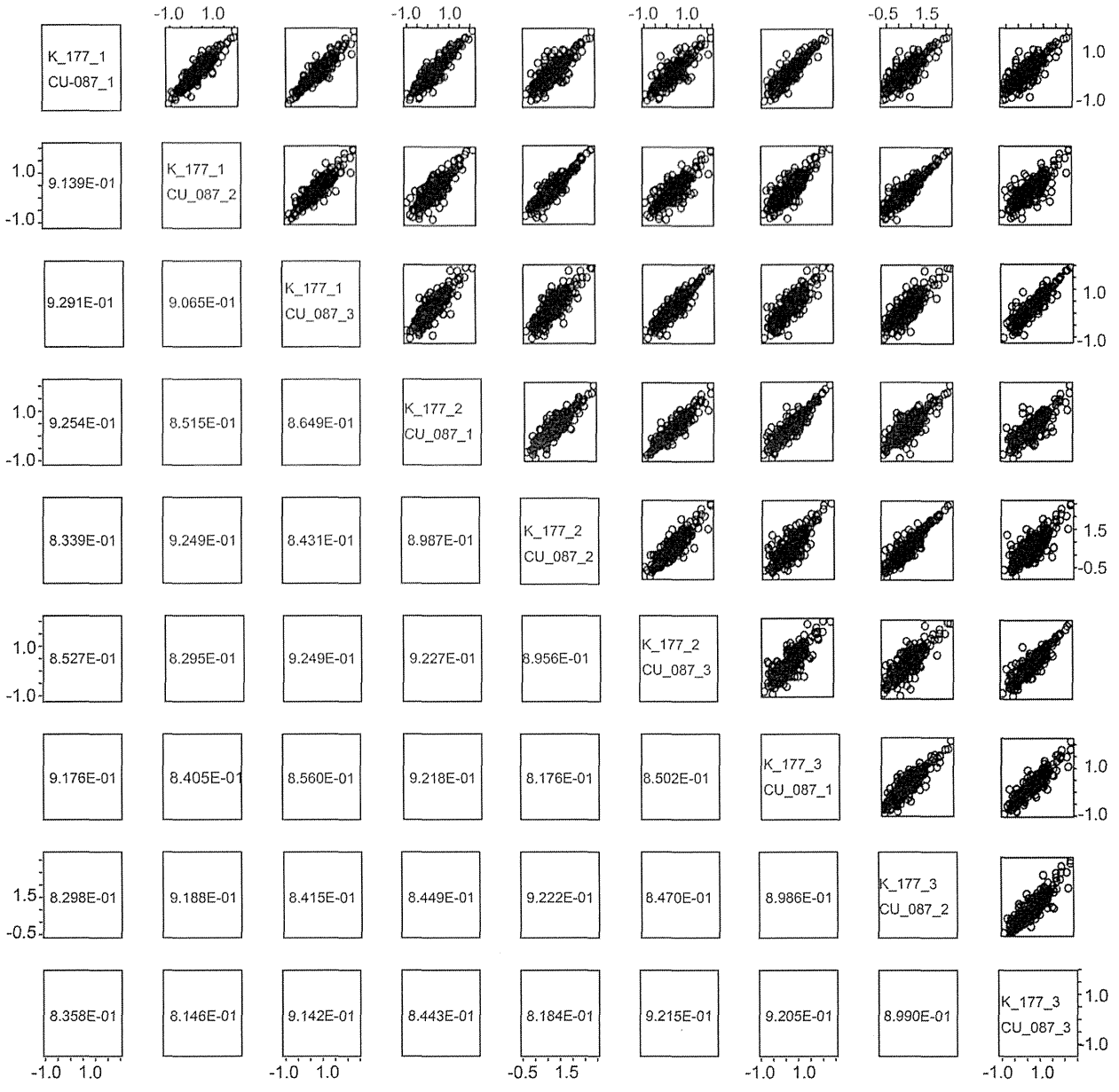


Figure 6. Comparison of differential (K-177 vs. CU-087) logarithmic miRNA expression in HCC among NGS technical replicates. Pearson's correlation coefficients range from 0.80 to 0.93.
doi:10.1371/journal.pone.0106314.g006

novel miRNAs. It is interesting that although hsa-miR-6715a-5p was not reported in the miRBase, it was detected in our analysis although only 10 reads were assigned to the miRNA. The biological significance of these candidate miRNAs requires future examination.

Finally, we also investigated the reproducibility of technical replicates for differential miRNA expression; a full list of scatter plots and correlation coefficients is available in Fig. S4. Fig. 6 shows examples of technical replicates with good correlation. We were unable to identify the specific conditions necessary to achieve highly reproducible technical replicates. However, several of our cross-replicate comparisons were of acceptable quality, leading us to believe that our method has the potential to adequately reproduce differential miRNA expression between technical replicates.

Conclusions

We have shown in this study that miRNA expression profiles obtained from NGS analysis are reproducible and are concordant with that obtained by the standard microarray procedure. Moreover, we have demonstrated that NGS can identify novel miRNAs that are otherwise undetectable by microarray analysis. HCC was distinguished from non-tumorous tissue with high

diagnostic accuracy, supporting the clinical application of NGS-based miRNA expression profiling.

Supporting Information

Figure S1 Full set of scatter plots of logarithmic miRNA expression in HCC samples by NGS and microarray analysis. Comparison between logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis. One black circle showed one miRNA.
(PDF)

Figure S2 Full set of scatter plots of differential logarithmic miRNA expression in HCC for NGS and microarray analysis. Comparison between differential logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis. One black circle showed one miRNA.
(PDF)

Figure S3 Comparison of logarithmic miRNA expression in HCC for NGS technical replicates not included in Fig. 3. Comparison between differential logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis. One black circle showed one miRNA.
(PDF)

Figure S4 Comparison of differential logarithmic miRNA expression in HCC for NGS technical replicates not included in Fig. 4. Comparison between differential logarithmic HCC miRNA expression in NGS (horizontal axis) and microarray (vertical) analysis. One black circle showed one miRNA.
(PDF)

Table S1 Extracted read counts of each miRNA obtained by NGS analysis.
(PDF)

Table S2 Detailed NGS analysis of HCC and non-tumorous tissue samples.
(PDF)

Table S3 Detailed mapping of short reads for novel miRNA candidates and hsa-mir-6715a.
(PDF)

Text S1 Sample script that processed short reads using fastx_clipper and miRDeep2.
(PDF)

Text S2 Detailed description of feature extraction and discriminant procedures.
(PDF)

Author Contributions

Conceived and designed the experiments: YM. Performed the experiments: YM TT RO HT. Analyzed the data: TT YT. Contributed reagents/materials/analysis tools: TT RO. Wrote the paper: YM TK ME AT NK YT TA. Read and approved the manuscript: YM TT RO HT TK ME AT NK YT TA.

References

1. Ambros V (2004) The functions of animal microRNAs. Nature 431: 350–355.
2. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297.
3. 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: 843–854.
4. El-Serag HB, Rudolph KL (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 132: 2557–2576.
5. Venook AP, Papandreou C, Furuse J, de Guevara LL (2010) The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. Oncologist 15 Suppl 4: 5–13.
6. Arzumanyan A, Reis HM, Feitelson MA (2013) Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. Nat Rev Cancer 13: 123–135.
7. Murakami Y, Yasuda T, Saigo K, Urashima T, Toyoda H, et al. (2006) Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. Oncogene 25: 2537–2545.
8. Braconi C, Patel T (2008) MicroRNA expression profiling: a molecular tool for defining the phenotype of hepatocellular tumors. Hepatology 47: 1807–1809.
9. Ji J, Shi J, Budhu A, Yu Z, Forgues M, et al. (2009) MicroRNA expression, survival, and response to interferon in liver cancer. N Engl J Med 361: 1437–1447.
10. Tsai WC, Hsu SD, Hsu CS, Lai TC, Chen SJ, et al. (2012) MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. J Clin Invest 122: 2884–2897.
11. Hsu SH, Wang B, Kota J, Yu J, Costinean S, et al. (2012) Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. J Clin Invest 122: 2871–2883.
12. Bjork JK, Sandqvist A, Elsing AN, Kotaja N, Sistonen L (2010) miR-18, a member of Oncomir-1, targets heat shock transcription factor 2 in spermatogenesis. Development 137: 3177–3184.
13. Murakami Y, Tamori A, Itami S, Tanahashi T, Toyoda H, et al. (2013) The expression level of miR-18b in hepatocellular carcinoma is associated with the grade of malignancy and prognosis. BMC Cancer 13: 99.
14. Toffanin S, Hoshida Y, Lachenmayer A, Villanueva A, Cabellos L, et al. (2011) MicroRNA-based classification of hepatocellular carcinoma and oncogenic role of miR-517a. Gastroenterology 140: 1618–1628 e1616.
15. McCormick KP, Willmann MR, Meyers BC (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. Silence 2: 2.
16. Wittmann J, Jack HM (2010) New surprises from the deep–the family of small regulatory RNAs increases. ScientificWorldJournal 10: 1239–1243.
17. Costa V, Angelini C, De Feis I, Ciccodicola A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol 2010: 853916.
18. Law PT, Qin H, Ching AK, Lai KP, Co NN, et al. (2013) Deep sequencing of small RNA transcriptome reveals novel non-coding RNAs in hepatocellular carcinoma. J Hepatol 58: 1165–1173.
19. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 40: 37–52.
20. Taguchi YH, Murakami Y (2013) Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers. PLoS One 8: e66714.
21. Murakami Y, Toyoda H, Tanahashi T, Tanaka J, Kumada T, et al. (2012) Comprehensive miRNA expression analysis in peripheral blood can diagnose liver disease. PLoS One 7: e48366.
22. Burroughs AM, Kawano M, Ando Y, Daub CO, Hayashizaki Y (2012) pre-miRNA profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5'/3' arm variation including concomitant cleavage and polyuridylation patterns. Nucleic Acids Res 40: 1424–1437.
23. Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, et al. (2007) MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. Gastroenterology 133: 647–658.
24. Gramantieri L, Ferracin M, Fornari F, Veronese A, Sabbioni S, et al. (2007) Cyclin G1 is a target of miR-122a, a microRNA frequently down-regulated in human hepatocellular carcinoma. Cancer Res 67: 6092–6099.
25. Thorgeirsson SS, Lee JS, Grisham JW (2006) Functional genomics of hepatocellular carcinoma. Hepatology 43: S145–150.
26. Xu J, Zhu X, Wu L, Yang R, Yang Z, et al. (2012) MicroRNA-122 suppresses cell proliferation and induces cell apoptosis in hepatocellular carcinoma by directly targeting Wnt/beta-catenin pathway. Liver Int 32: 752–760.
27. Zhang J, Yang Y, Yang T, Liu Y, Li A, et al. (2010) microRNA-22, downregulated in hepatocellular carcinoma and correlated with prognosis, suppresses cell proliferation and tumourigenicity. Br J Cancer 103: 1215–1220.
28. Yan Y, Luo YC, Wan HY, Wang J, Zhang PP, et al. (2013) MicroRNA-10a is involved in the metastatic process by regulating Eph tyrosine kinase receptor A4-mediated epithelial-mesenchymal transition and adhesion in hepatoma cells. Hepatology 57: 667–677.
29. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, et al. (2008) A human snoRNA with microRNA-like functions. Mol Cell 32: 519–528.

Taguchi and Murakami *BMC Research Notes* 2014, **7**:581
http://www.biomedcentral.com/1756-0500/7/581

BMC
Research Notes

RESEARCH ARTICLE Open Access

Universal disease biomarker: can a fixed set of blood microRNAs diagnose multiple diseases?

Y-h Taguchi[1*] and Yoshiki Murakami[2]

Abstract

Background: The selection of disease biomarkers is often difficult because of their unstable identification, i.e., the selection of biomarkers is heavily dependent upon the set of samples analyzed and the use of independent sets of samples often results in a completely different set of biomarkers being identified. However, if a fixed set of disease biomarkers could be identified for the diagnosis of multiple diseases, the difficulties of biomarker selection could be reduced.

Results: In this study, the previously identified universal disease biomarker (UDB) consisting of blood miRNAs that could discriminate between patients with multiple diseases and healthy controls was extended to the recently reported independent measurements of blood microRNAs (miRNAs). The performance achieved by UDB in an independent set of samples was competitive with performances achieved with biomarkers selected using lasso, a standard, heavily sample-dependent procedure. Furthermore, the development of stable feature extraction was suggested to be a key factor in constructing more efficient and stable (i.e., sample- and disease-independent) UDBs.

Conclusions: The previously proposed UDB was successfully extended to an additional seven diseases and is expected to be useful for the diagnosis of other diseases.

Keywords: Disease biomarker, Universality, Blood microRNA

Background

Identification of biomarkers is important for the diagnosis of disease. By using biomarkers with high specificity for certain diseases, patients can be identified without diagnosis by doctors. After diagnosis using biomarkers, it is hoped that fewer patients will require diagnosis by a doctor. This enables doctors to diagnose a limited number of screened patients in more detail. Blood is a useful source of biomarkers. Numerous compounds/proteins in blood have been identified as effective biomarkers that allow the early diagnosis of several diseases (e.g., [1-3]). One disadvantage of this system is that distinct compounds/proteins are required to diagnose individual diseases, because diagnoses are usually based on the observation of unexpected values of compounds/proteins. When following this strategy, new compounds/proteins that increase or decrease in specific diseases should be identified. This system of biomarker identification incurs high costs because of the measurements of each biomarker. Thus, it is difficult to test for many diseases simultaneously because the number of diseases tested is proportional to the cost. The identification of a universal disease biomarker (UDB) that can diagnose multiple diseases simultaneously would be useful and economically beneficial. However, identifying a UDB using the traditional strategy of one compound/protein for one disease is unlikely.

Despite this difficulty, several studies have attempted to identify UDBs. For example, interleukin-8 (IL-8) was thought to be a UDB [4] as it was reported to be a useful biomarker for multiple diseases including urinary bladder cancer, prostatitis, acute pyelonephritis, vesicoureteral reflux, pulmonary infections, osteomyelitis, inflammatory bowel disease, chorioamnionitis, nosocomial bacterial infections, and non-Hodgkin's lymphoma. Despite the apparent usefulness of IL-8 as a UDB, it has a strong tendency to increase non-specifically in individuals because most inflammatory conditions induce its production, therefore it might be considered together with other biomarkers. Another UDB is pHLIP and acidity,

*Correspondence: tag@granular.com
[1]Department of Physics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, 112-8551 Tokyo, Japan
Full list of author information is available at the end of the article

which although limited to cancer diagnosis was proposed to be a UDB for cancers [5]. Fendos and Engelman successfully and noninvasively labeled tumor tissues using a pH-sensitive biosensor. pHLIP also labeled tumors independent of the type of cancer. Another example of a UDB is FibroTest [6], which was used to diagnose several liver diseases including alcoholic liver disease, Hepatitis B virus, Hepatitis C virus, and Nonalcoholic fatty liver disease. FibroTest consists of a six-parameter blood test, α2-macroglobulin, Haptoglobin, Apolipoprotein A1, γ-glutamyl transpeptidase, Total bilirubin, and Alanine transaminase, combined with the age and gender of the patient. However, these biomarkers lacked either specificity (IL-8 is used in combination with other biomarkers for accurate diagnoses) or universality (pHLIP is used only for cancer diagnosis while FibroTest is only used to diagnose liver diseases). An ideal disease UDB should have the ability to diagnose multiple diseases compared with normal healthy controls. One method to achieve this is by the combination of multiple biomarkers, as used for the FibroTest. Although FibroTest has fixed coefficients to construct a UDB, if varying coupling constants allows the diagnosis of distinct multiple diseases, biomarkers that consist of multiple individual biomarkers have the potential to be UDBs.

Recently, blood microRNAs (miRNA) have been identified as promising disease biomarkers [7]; combinations of mir-498 clusters are potential biomarkers for pregnancy, although pregnancy is not a disease. Blood miRNAs were also identified as anti-doping biomarkers [8], biomarkers of peripheral arterial disease [9], acute myocardial infarction and underlying coronary artery stenosis [10], and acute graft-versus-host disease [11]. They are also stable biomarkers [12]. Furthermore, although combinatorial circulating biomarkers are considered potential effective biomarkers for various diseases [13-20], combinations for the diagnosis of individual diseases often fluctuate between studies. For example, two recent distinct studies that tried to construct combinatorial blood miRNA biomarkers for the diagnosis of Alzheimer's disease had no common miRNAs [21,22]. Even for the diagnosis of an individual disease, there is often no unique combination of blood miRNAs. This suggests that a UDB is unlikely to be constructed from multiple blood miRNAs.

In contrast to these studies, we recently identified a potential UDB consisting of blood miRNAs [23]. Ten to 12 common blood miRNAs could be used to diagnose 13 various diseases from normal controls. Although this demonstrated the potential of blood miRNAs to be used as a UDB, the study used samples taken from only one study with shared normal controls. Thus, further studies are required to provide convincing data. In the current study, we cross-validated the previously proposed UDB [23] of 12 fixed miRNAs by investigating whether miRNAs could diagnose an additional seven distinct diseases using blood miRNAs that were recently reported and were not available when the previous study [23] was performed. The discriminatory ability of a UDB composed of 12 fixed blood miRNAs was competitive compared with that using a conventional method and miRNAs selected by a recently proposed principal component analysis (PCA)-based unsupervised feature selection method [23].

Results and discussion

Universality of UDB

To determine whether previously identified UDBs consisting of blood miRNAs [23] were universal, we evaluated their performance using seven independent data sets targeting seven diseases (see Methods). Although 10 miRNAs were selected for each disease from a total of 13 diseases in the previous study, 12 combined blood miRNAs (hsa-miR-425, hsa-miR-15b, hsa-miR-185, hsa-miR-92a, hsa-miR-140-3p, hsa-miR-320a, hsa-miR-486-5p, hsa-miR-16, hsa-miR-191, hsa-miR-106b, hsa-miR-19b, and hsa-miR-30d) were used to form the UDB in this study. Missing miRNAs in the data sets were excluded from the discrimination.

In Figure 1, the accuracy achieved by PCA-based liner discriminant analysis (LDA, red crosses) and support vector machine (SVM, red x-marks) using UDB is shown (also red boxes in Figure 1(b)). Mean accuracies were 0.791 and 0.815, respectively, and they were coincident with the mean accuracy (0.784) estimated using PCA-based LDA with UDB in a previous study [23] (see Table 1). Values of accuracy together with sensitivity and specificity values are also listed in Table 1. It was observed that performances were independent of the methods and samples, demonstrating the usefulness of the UDB. More detailed performances and their evaluations, i.e., true and false positives and negatives in a 2×2 tables together with P-values computed by Fisher's exact test, odds ratio and area under the receiver operating characteristic (ROC) area under the curve (AUC), are shown in Additional file 1: Table S2.

Comparison of performances between UDB and lasso

Although Table 1 shows the usefulness of a UDB consisting of blood miRNAs, it is important to determine how effective the UDB is when compared with conventional methods (i.e., non-universal, sample-dependent sets). We performed lasso-based discrimination (see Methods) between healthy controls and patients of each disease. Lasso-based discrimination was used so that performances of feature extraction (FE) between unsupervised FE and lasso could be compared. In addition, there are generally limited numbers of individual miRNAs that exhibit significant differences between normal controls

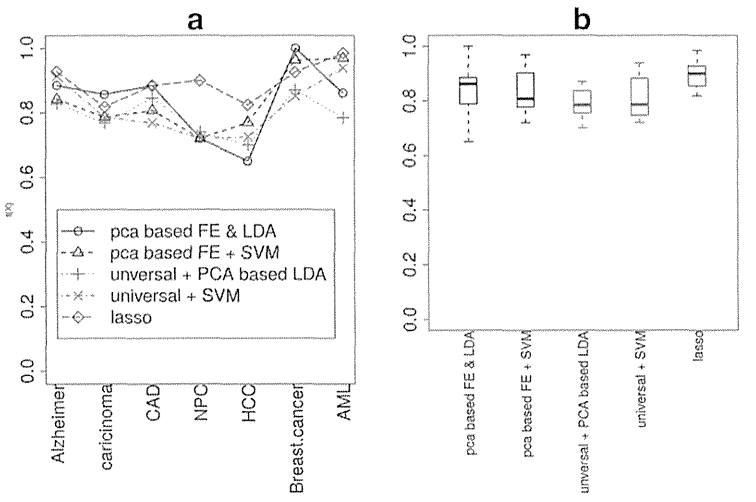


Figure 1 Accuracies achieved by various discrimination methods and FEs. (a) Dependence upon diseases and methods. **(b)** Boxplot of accuracies.

and patients (see below), thus selection based on significant differences between patients and healthy controls as usual was difficult. The results are shown in Table 2 and Figure 1 (blue diamonds and a blue box in Boxplot). More detailed performances and their evaluations, i.e., true and false positives and negatives in a 2 × 2 tables of lasso-based discrimination together with *P*-values computed by Fisher's exact test, odds ratios and AUC, are shown in Additional file 1: Table S3. Although performances achieved by lasso-based discrimination were better than by PCA-based LDA with UDB (Table 1), those achieved by SVM with UDB were not significantly lower than the lasso-based discrimination (although three tests were performed, *t*-test, Wilcoxon rank sum test and Kolmogorov-Smirnov test, no *P*-values lower than 0.05 were detected). Since the lack of significance was because of large fluctuations in performances achieved by SVM with UDB, this suggested UDB might not be as effective as lasso-based discrimination. However, the possibility that UDB is as effective as standard discrimination using sample-dependent (not universal) features is indicated.

Table 1 Performance of UDB with PCA-based LDA and SVM

Diseases	Accuracy	Sensitivity	Specificity
	PCA-based LDA		
AD	0.829	0.833	0.818
Carcinoma	0.768	0.730	0.800
CAD	0.846	0.846	0.846
NPC	0.740	0.806	0.632
HCC	0.700	0.700	0.700
BC	0.870	0.813	0.955
AML	0.784	0.769	0.846
Mean	0.791	0.785	0.800
Mean of previous study [23]	0.784	0.750	0.800
	SVM		
AD	0.914	0.917	0.909
Carcinoma	0.786	0.867	0.692
CAD	0.769	0.769	0.769
NPC	0.720	0.806	0.579
HCC	0.725	0.550	0.900
BC	0.852	0.813	0.909
AML	0.938	0.981	0.769
Mean	0.815	0.815	0.800

AD, Alzheimer's disease; CAD, coronary artery disease; NPC, nasopharyngeal carcinoma; HCC, hepatocellular carcinoma; BC, breast cancer; AML, acute myeloid leukemia; UDB, universal disease biomarker; SVM, support vector machine; LDA, linear discriminant analysis; PCA, principal component analysis. Data from previous study [23] are also shown for comparison.

Stability of FE: the condition to get UDB

To understand why we could successfully identified a UDB in the previous study that could never be indeitified by

Table 2 Performance of lasso-based discrimination

Diseases	Accuracy	Sensitivity	Specificity	Optimal *s*
AD	0.928	0.979	0.818	0.09
Carcinoma	0.818	0.867	0.760	0.9
CAD	0.884	0.769	1.000	0.24
NPC	0.900	0.935	0.842	1
HCC	0.825	0.650	1.000	0.03
BC	0.925	0.906	0.955	0.46
AML	0.985	1.000	0.923	0.64
Mean	0.895	0.872	0.900	

AD, Alzheimer's disease; CAD, coronary artery disease; NPC, nasopharyngeal carcinoma; HCC, hepatocellular carcinoma; BC, breast cancer; AML, acute myeloid leukemia. *s* (fraction) is used for the predict.lars function (see Methods).

anyone, the stabilities of FE were compared between lasso and PCA-based unsupervised FE. PCA-based unsupervised FE was used for the previous UDB discovery [23]. The importance of stability was previously demonstrated by Wehrens *et al.* [24], who showed that a stable FE improved the performance.

Figure 2 shows the stabilities S (see Methods) of lasso-based discrimination (blue diamonds). Generally, the stabilities were very low and each miRNA was selected as a biomarker at most for half the trials. Thus, lasso does not have the ability to provide UDBs, because it could not select stable (sample-independent) biomarkers for each disease. One may suppose that the stabilities will improve if miRNAs that exhibit significant differences between healthy controls and patients are identified and selected. However, this is not currently a realistic strategy, since there are insufficient numbers of miRNAs (often < 10) that exhibit significant differences between healthy controls and patients (Table 3). For coronary artery disease (CAD) and hepatocellular carcinoma (HCC), no miRNAs have been identified that exhibit significant differences between normal controls and patients in the present data sets.

However, PCA-based unsupervised FE (black circles in Figure 2) showed significantly larger S values than lasso. In addition, performances were comparative with those achieved by lasso (Table 4, black circles and triangles in Figure 1(a) and black box in Figure 1(b)). More detailed performances and their evaluations, i.e., true and false positives and negatives in a 2×2 tables together with P-values by Fisher' exact test, odds ratio and AUC, are shown in Additional file 1: Table S4.

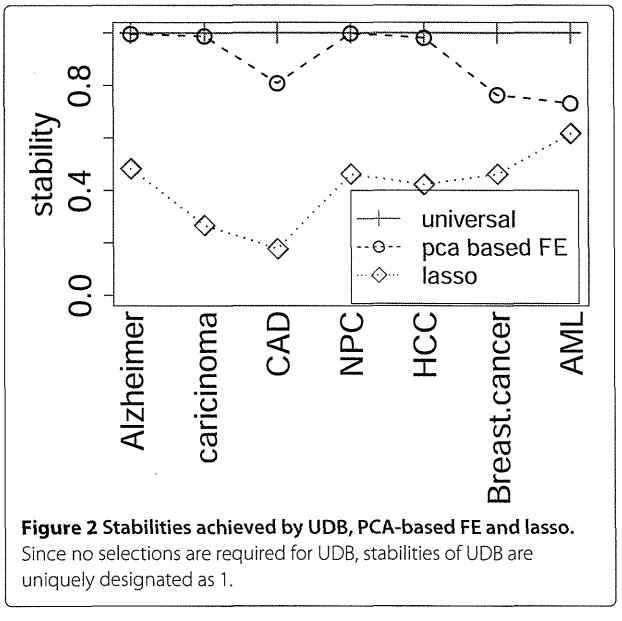


Figure 2 Stabilities achieved by UDB, PCA-based FE and lasso. Since no selections are required for UDB, stabilities of UDB are uniquely designated as 1.

Table 3 The number of miRNAs that exhibit significant differences between normal controls and patients for each disease

Diseases	Significant	Not significant
AD	4	498
Carcinoma	7	558
CAD	0	746
NPC	264	622
HCC	0	255
BC	86	188
AML	6	122

AD, Alzheimer's disease; CAD, coronary artery disease; NPC, nasopharyngeal carcinoma; HCC, hepatocellular carcinoma; BC, breast cancer; AML, acute myeloid leukemia. For more details, see Methods.

Why selected biomarkers are frequently varied between samples was attributed to the difference of data normalization. However, the results shown here indicate this might be caused by using incorrect and unstable FE methods. To obtain UDB, stable FE methods should be used [23].

The study by Wehrens *et al.* [24] used PCA-based LDA to maximize the stability of FE, whereas the current study did not require better stability, as this is automatically obtained when using PCA-based unsupervised FE. Thus, stability achieved by PCA-based unsupervised FE is expected to be more robust than feature selections by stability maximization using PCA-based LDA. Moreover, to rank features based on stability, Wehrens *et al.* [24] performed time-consuming iterative cross-validations that were not required by the PCA-based unsupervised FE. Thus, PCA-based unsupervised FE methodology is less computationally challenging than feature selections by stability maximization using PCA-based LDA.

The successful identification of UDBs [23] was possibly because of stable FE methods, which we suggest are important for developing UDBs, although the stability of FE is often overlooked. To determine more efficient UDBs, searching with efficient and stable FEs is required.

The number of features selected by FE

Previously [23], the number of features selected by PCA-based unsupervised FE was fixed at 10, because data sets analyzed previously were taken from a single study. Previous studies used the same microarray to measure miRNA expression in multiple diseases. In contrast, data sets used in the current study were heterogeneous. They were collected from multiple studies performed by independent research groups. Measurements were not performed by a single microarray but by various methods including qPCR. The sources of samples were also heterogeneous, ranging from whole blood to serum or plasma.

Table 4 Performance of miRNAs selected by PCA-based FE with PCA-based LDA and SVM

Diseases	Accuracy	Sens.	Spec.	Number of miRNAs*	Number of PCs+	Δ#
		PCA-based LDA				
AD	0.886	0.917	0.818	22	16	2.5
Carcinoma	0.857	0.846	0.867	36	2	7
CAD	0.885	0.923	0.846	16	14	9
NPC	0.720	0.806	0.579	28	18	5
HCC	0.650	0.600	0.700	8	1	7
BC	1.000	1.000	1.000	18	13	6
AML	0.862	0.846	0.923	11	8	7
Mean	0.837	0.848	0.819			
Mean of previous study [23]	0.784	0.750	0.800			
		SVM				
AD	0.843	0.833	0.864	22		
Carcinoma	0.786	0.807	0.767	36		
CAD	0.807	0.615	1.000	16		
NPC	0.720	0.774	0.632	28		
HCC	0.770	0.550	0.850	8		
BC	0.963	1.000	0.938	18		
AML	0.969	1.000	0.846	11		
Mean	0.837	0.797	0.842			

*number of miRNAs selected by PCA-based FE, +optimal number of PCs estimated by LOOCV, #threshold value of PCA-based FE. Data from previous study [23] are also shown for comparison. AD, Alzheimer's disease; CAD, coronary artery disease; NPC, nasopharyngeal carcinoma; HCC, hepatocellular carcinoma; BC, breast cancer; AML, acute myeloid leukemia; UDB, universal disease biomarker; SVM, support vector machine; LDA, linear discriminant analysis; PCA, principal component analysis.

Thus, we varied the number of features selected by PCA-based unsupervised FE between diseases (Additional file 2: Figure S1 for two-dimensional embeddings of miRNAs used for FE).

Interestingly, the optimal number of selected features was common between lasso and PCA-based unsupervised FE (Additional file 2: Figure S2). This suggests that the number of miRNAs required to discriminate healthy controls from patients is not dependent on the methods used but on the samples. This is not surprising because many sets of miRNAs discriminate between patients and normal controls if miRNAs are not independent of each other. In addition, the stability of FE is important, otherwise selected features will vary between trials.

This study did not identify a UDB from a data set we used, but rather validated the usefulness of UDBs identified in a previous study. To identify UDBs, sample preparation and measurements must be standardized to minimize the variance between samples. This should be possible because the target is uniquely independent of blood in disease.

Toward a mechanism-based biomarker

The UDB in this study was clearly decided by meta-analysis, and thus was not mechanism-based. However, if it also functions as a mechanism-based biomarker, this would be more plausible. To determine the possibility of using a UDB as a mechanism-based biomarker, we employed DIANA-mirpath [25]. Table 5 lists the 27 significant KEGG pathways reported by DIANA-mirpath (see Methods). Among 27 KEGG pathways, nine were cancer pathways (bold font). There were also five pathways (bold ilatic) that were disease pathways other than cancers. In addition, three pathways (italicized) were cancer-related pathways and four pathways (asterisked) were parts of "Pathways in cancer" (Figure 3). Thus, there were only five pathways that were not directly related to diseases. Therefore, miRNAs included in the UDB in this study were not only extensively included disease pathways, but also contributed to various disease pathways. Further experimental investigations of the expression of miRNA target genes will be required to demonstrate how UDB is involved in disease mechanisms.

Heterogeneity of blood sources

In contrast to previous research [23] where only serum samples were used, the blood sources in this study were heterogenous, ranging from whole blood [21] to serum [26] or plasma [27] (full list of sources is shown in Additional file 1: Table S1). One may wonder why UDB works well despite this heterogeneity of sources. However, in a previous study [23], we tried to select 12 miRNAs included in UDB, not based on inference accuracy but rather by stability. That study only checked

Table 5 KEGG pathway analysis of 12 miRNAs included in the UDB using DIANA-mirpath [25]

	KEGG.pathway	p.value	# of genes	# of miRNAs
1	***Prion diseases***	0.00e+00	6	2
2	*Pathways in cancer*	3.00e-13	39	6
3	PI3K-Akt signaling pathway*	1.07e-11	43	4
4	TGF-beta signaling pathway*	5.98e-10	14	4
5	*Viral carcinogenesis*	1.56e-09	27	5
6	Ribosome	6.04e-09	22	1
7	**Small cell lung cancer**	6.33e-09	17	5
8	**Colorectal cancer**	1.02e-08	9	6
9	Ribosome biogenesis in eukaryotes	2.02e-08	20	1
10	p53 signaling pathway*	6.75e-08	16	5
11	RNA transport	1.19e-07	28	1
12	Cell cycle*	1.56e-07	22	3
13	**Pancreatic cancer**	2.75e-07	11	5
14	***Hepatitis B***	1.17e-06	12	5
15	**Prostate cancer**	4.64e-06	16	5
16	**Bladder cancer**	6.94e-06	7	4
17	**Chronic myeloid leukemia**	2.01e-05	8	4
18	***Measles***	1.28e-04	19	5
19	Protein export	3.29e-04	9	1
20	**Non-small cell lung cancer**	3.69e-04	6	5
21	***HTLV-I infection***	1.43e-03	11	3
22	**Glioma**	1.46e-03	6	3
23	**Melanoma**	1.71e-03	7	5
24	*Transcriptional misregulation in cancer*	1.22e-02	9	3
25	Oocyte meiosis	1.36e-02	14	1
26	Focal adhesion*	1.50e-02	8	3
27	***Epstein-Barr virus infection***	2.45e-02	19	3

Bold faces: tumors/cancers, *Bold italic*: other diseases, *Italic*: tumors/cancers related, *parts of "Pathways in cancer", and surrounded by blue rectangular in Figure 3.

sample independency, but it is likely that sample independency is also related to source independency, since it is often as large as source dependency. miRNA expression is dependent upon both the source and patients' age, gender, and body mass index. In addition, UDB was independent of measurement methods, i.e., NGS, microarray or qPCR (a full list of measurement methods is shown in Additional file 1: Table S1). If UDB is independent of patient properties and measurement methods, it is not surprising that UDB is also independent of sources, since all sources were taken from blood. Source independency of UDB should be investigated in more detail in the future.

Usefulness of UDB as practical clinical tools

One may wonder if the expected accuracy (0.8) of UDB is useful or not. However, UDB can diagnose multiple diseases simultaneously. Therefore, by measuring 12 miRNAs in blood, over 20 diseases (14 diseases in the previous study [23] and seven diseases in this study) can be diagnosed. Thus, UDB can be used for pre-screening. For example, patients are diagnosed by UDB for the 20 diseases. Then, if patients are positive for one disease, further diagnosis using more precise biomarkers can confirm the diagnosis. This will be more effective and non-invasive than performing 20 independent diagnoses using disease-specific biomarkers.

Conclusion

In this study, we demonstrated that a predefined UDB [23] could discriminate seven diseases from healthy controls. Since the diseases and samples were not included in our previous study [23] that defined UDBs, this study suggests the robustness of UDB for disease diagnosis. The performance achieved by UDB was comparative with that of lasso, the standard sample-dependent FE. Because PCA-based unsupervised FE, used for UDB identification in a previous study, outperformed lasso in terms of stability, the use of stable FE will be a key factor for discovering UDBs.

Methods

Blood miRNA expression profiles

Seven blood miRNA expressions used in this study were from the Gene Expression Omnibus (GEO): Alzheimer's disease (AD) (GSE46579) [21], carcinoma (GSE37472) [26], CAD (GSE49823), nasopharyngeal carcinoma (NPC) (GSE43329), HCC (GSE50013) [27], breast cancer (BC) (GSE41922) [28] and acute myeloid leukemia (AML) (GSE49665) [29]. Detailed information is shown in Additional file 1: Table S1.

Principal component analysis-based unsupervised feature extraction

To select blood miRNAs for the diagnosis of seven diseases, blood miRNAs were selected using the recently proposed PCA-based unsupervised FE as previously described [23,30]. Briefly, suppose $\mathbf{X}$ is the matrix such that x_{ij} represents the amount of the ith miRNA expression in the jth sample. PCA is regarded as the eigenvalue problem

$$\frac{1}{N}\mathbf{X}^T\mathbf{X}\boldsymbol{u}_k = \lambda_k \boldsymbol{u}_k, (k = 1, \ldots, M)$$

where N and M are the total number of miRNAs and samples, respectively. Here M is assumed to be less than N