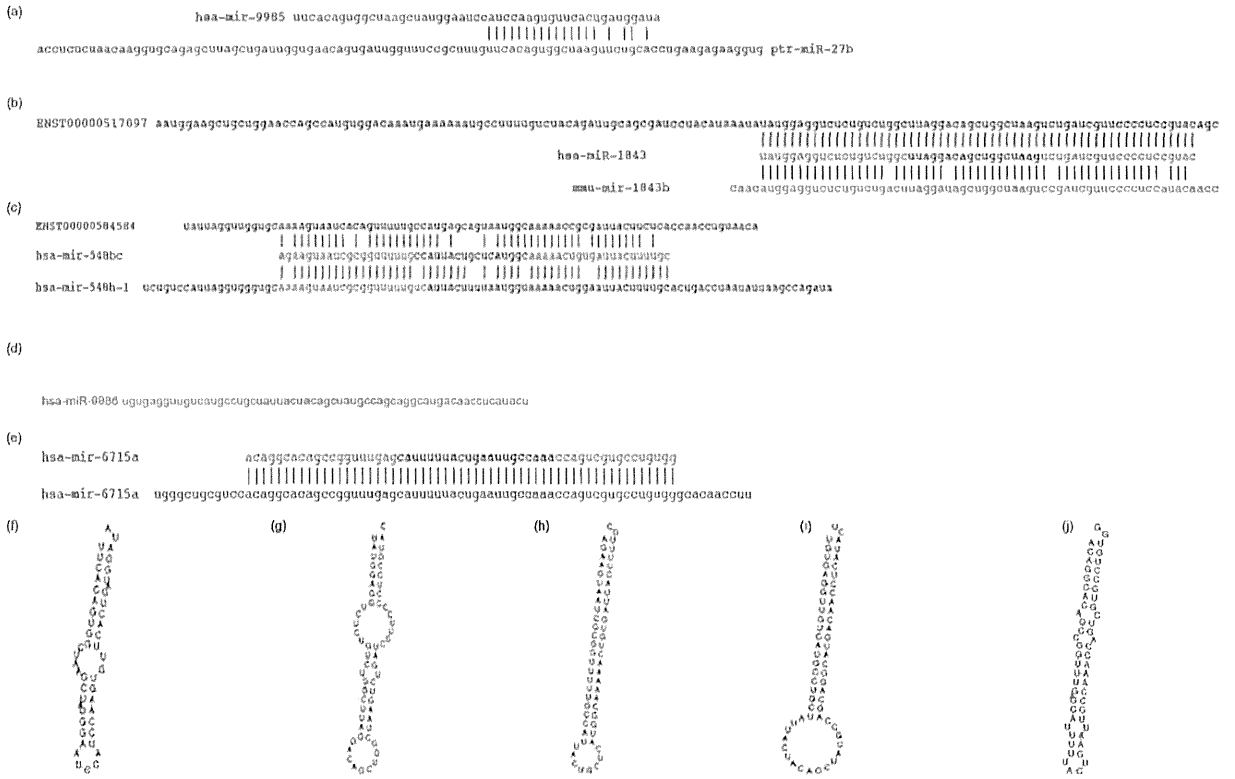


**Figure 5. Comparison of novel miRNA candidates ((a) hsa-mir-9985, (b) hsa-mir-1843, (c) hsa-mir-548bc and (d) hsa-mir-9986) with known miRNAs and transcripts.** (e) hsa-mir-6715a novel miRNA was not included in miRBase release 18, but was present in later releases. Nucleotides highlighted in red show 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

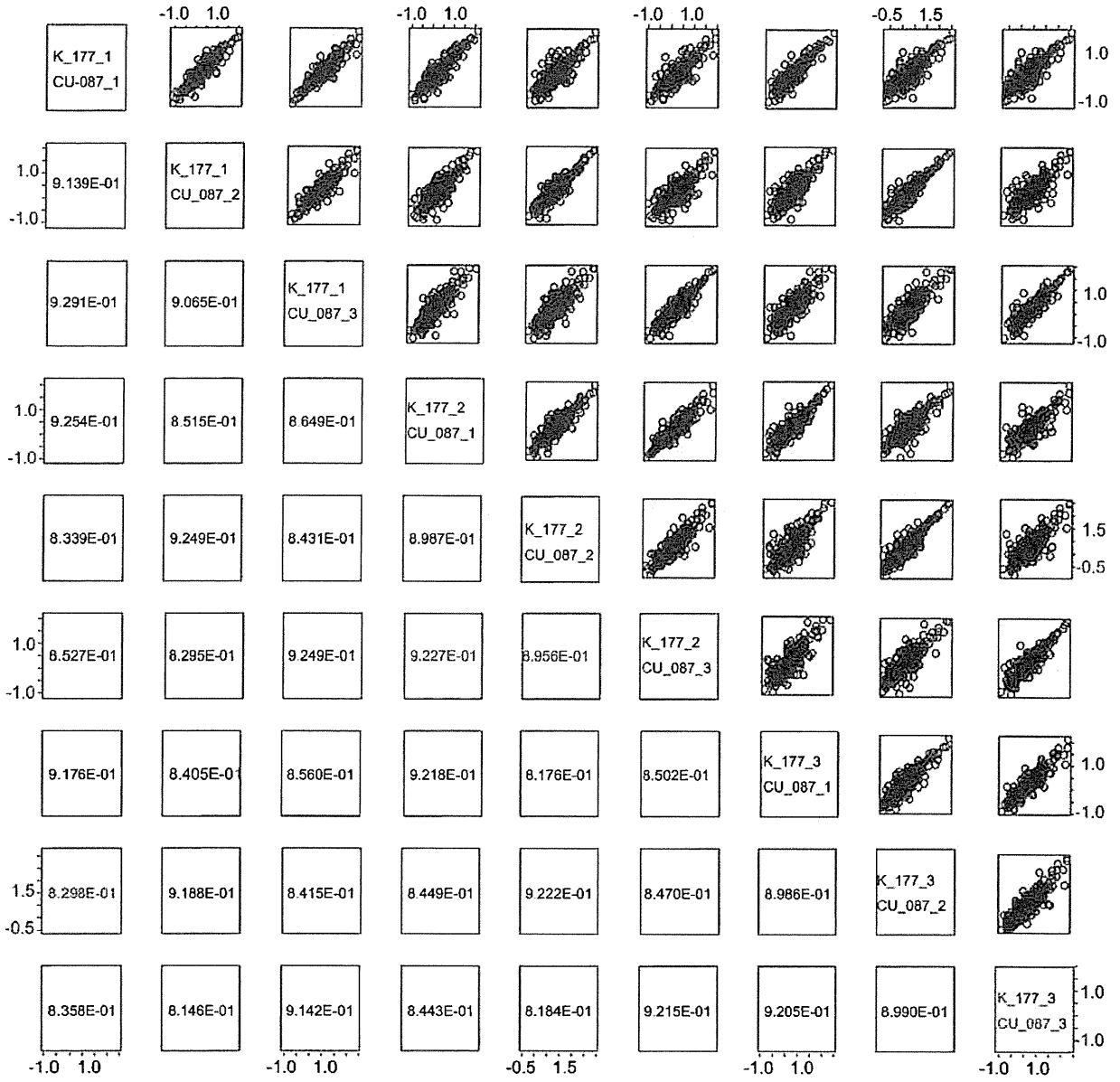


**Figure 6. Comparison of differential (K-177 vs. CU-087) logarithmic miRNA expression in HCC among NGS technical replicates.** Pearson's correlation coefficients range from 0.80 to 0.93.
doi:10.1371/journal.pone.0106314.g006

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, $\alpha$2-macroglobulin, Haptoglobin, Apolipoprotein A1, $\gamma$-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 $2 \times 2$ 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

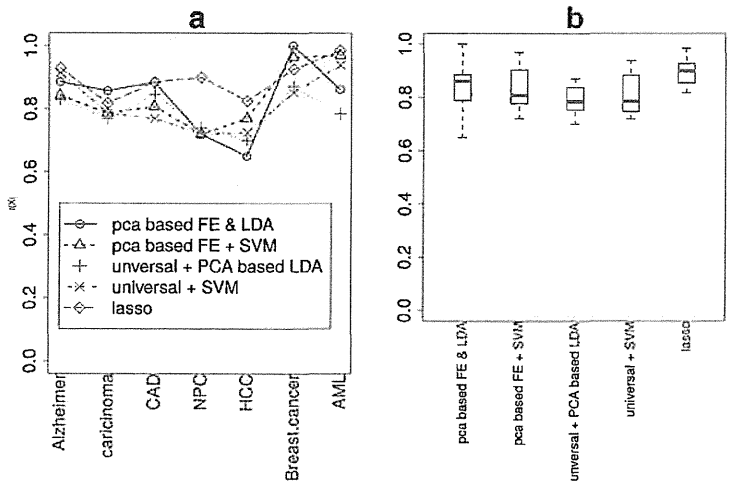


**Figure 1 Accuracies achieved by various discrimination methods and FEs. (a)** Dependence upon diseases and methods. **(b)** Boxplot of accuracies.

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 $2 \times 2$ tables together with $P$-values by Fisher' exact test, odds ratio and AUC, are shown in Additional file 1: Table S4.

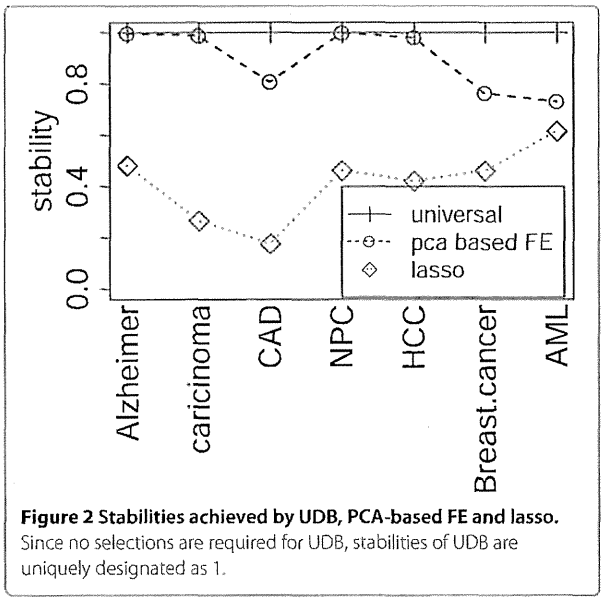


**Figure 2 Stabilities achieved by UDB, PCA-based FE and lasso.** Since no selections are required for UDB, stabilities of UDB are uniquely designated as 1.

**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 $i$th miRNA expression in the $j$th 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 $N$

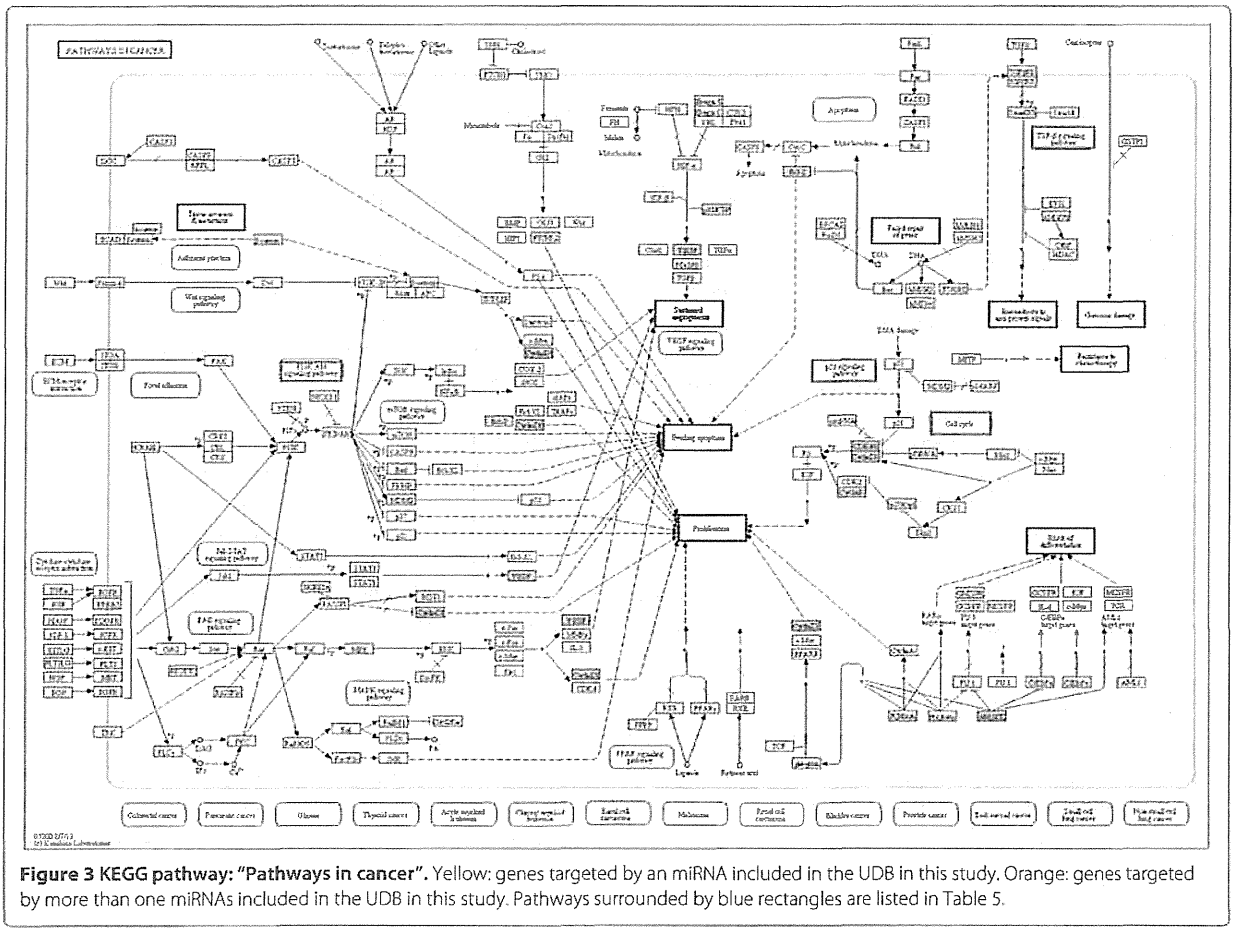

**Figure 3 KEGG pathway: "Pathways in cancer".** Yellow: genes targeted by an miRNA included in the UDB in this study. Orange: genes targeted by more than one miRNAs included in the UDB in this study. Pathways surrounded by blue rectangles are listed in Table 5.

as is usual. $\lambda_i$ and $\boldsymbol{u}_i$ represent the eigenvalue and vector, respectively.

$$x_{ik} \equiv \sum_j u_{kj} x_{ij}$$

gives the principal component score (PCS) of $i$th miRNA. Using the obtained $x_{ik}, k = 1, \ldots, D(< M)$, miRNAs were determined to be embedded into low $D$ dimensional space.

Multiplying $\mathbf{X}$ on both sides, the following is obtained:

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

where $\boldsymbol{v}_k = \mathbf{X}\boldsymbol{u}_k$ can be regarded as an eigenvector. Then,

$$x_{kj} \equiv \sum_i v_{ki} x_{ij}$$

gives the PCS of the $j$th sample. Using the obtained $x_{kj}, k = 1, \ldots, D(< M)$, samples were regarded to be embedded into low $D$ dimensional space.

PCA-based unsupervised FE selects outlier miRNAs in low $K(< M)$ dimensional embedding space,

$$r_{Ki} > \Delta$$

where

$$r_{Ki}^2 \equiv \sum_{k=1}^{K} x_{ik}^2$$

Typically $K$ is taken to be two. Since these outliers could have a major contribution to $\boldsymbol{u}_k$'s by definition, if there are a limited number of well-defined outliers, the exclusion of miRNAs other than outliers does not alter $\boldsymbol{u}_k$'s. Since $\boldsymbol{v}_k$ is a linear transformation of $\boldsymbol{u}_k$ as shown above, the exclusion of miRNAs other than outliers does not alter $\boldsymbol{v}_k$. Thus, retaining only outlier miRNAs may also preserve lower dimensional embeddings of samples that are important for disease diagnosis, e.g., discrimination between patients and healthy controls. Although this is only hypothetical, it explains why PCA-based unsupervised FE is expected to function well. Currently, there are no well-defined criteria for the selection of $\Delta$. Although

$\Delta$ was decided to include sufficient numbers (majority) of outliers, these were selected by the visual inspection of two-dimensional embedding of miRNAs. Singular decomposition-based interpretation is also available as Additional file 3: Text S1.

### Discriminatory analyses between patients and healthy controls with cross-validations

Three discriminant analyses were performed in this study as follows. The first, a PCA-based LDA, a discriminant counterpart of the partial least square (PLS), is defined as discrimination using the first $k$ PCSs (i.e., from the first to the $k$th PCSs). First, PCA was applied to all samples. Then, PCA-based LDA was performed using only PCSs in the training set. Since the learning process includes unlabeled information of the test set, it is semi-supervised learning. Samples in the test set were predicted using trained PCA-based LDA. LDA was performed using `lda` functions in R [31] and the prediction of samples in the test set was performed by `predict.lda` functions in R. Optimal $k$ was determined using cross-validations. The second analysis used an SVM trained with training set samples using `svm` function included in the `e1071` R package with default settings (e.g., with the usage of Gaussian kernel), other than `class.weight` argument that was set to attribute equal weights to sets of normal controls and patients when the number of samples in normal controls differed from that of patients. Then, samples in the test set were predicted using `predict.svm` function in R. Third, lasso was used for a discrimination study. Lasso was performed using the `lars` function included in `lars` R package, attributing 1 and 2 to healthy controls and patients, respectively, and using the setting `type='lasso'`. Then, samples in the test set were predicted using `predict.lars` function in R for $s = n/100, n = 0, \ldots, 100$ with `mode='fraction'`. Samples with predicted values larger (less) than 1.5 were regarded to be patients (healthy controls). Optimal $s$ was selected by cross-validation. For all cases, leave one out cross-validation (LOOCV) was employed.

### Data normalization

Since this study is a meta-analysis using data sets collected from various independent studies employing distinct measuring methods, we normalized data sets individually by distinct methods (Table 6). Data from multiple studies were treated identically and compared. In addition, some miRNAs with abnormally large values were excluded from the analysis. Excluded miRNAs were hsa-miR-486-5p (AD), hsa-miR-223 and hsa-miR-338 (CAD), and hsa-miR-451 (NPC).

### Stability test

On LOOCV FE, selected features (miRNAs) are listed. For lasso, miRNAs with non-zero $\beta$s were listed by setting `type='coefficients'` for `predict.lars` function with estimated optimal $s$. Because of LOOCV, FE was performed by $M$(=the number of samples) times. Then stability was defined as

$$S \equiv \frac{1}{\hat{N}} \sum_{i \in \{i | F_i \neq 0\}} \frac{F_i}{M}$$

where $F_i$ is the number of times that $i$th miRNA was selected within $M$ times FE. Summation was performed for miRNAs that were non-zero $F_i$ (i.e., selected at least once in FEs) and $\hat{N}$ is the number of miRNAs included in the summation. Larger $S$, $\left(\frac{1}{M} \leq S \leq 1\right)$ indicates more stable FEs.

### *P*-values computation for significant difference between healthy controls and patients

*P*-values computed for significant differences between healthy controls and patients of each disease were determined using $t$-test for each miRNA. Computed *P*-values were adjusted by BH-criterion [32] and miRNAs with

**Table 6 Details of data normalization**

| GEO ID | Disease | Data set names/ Data retrieval methods | Data normalization timing | Data normalization methods |
|---|---|---|---|---|
| GSE46579 | AD | GSE46579_AD_ngs_data_summarized.xls.gz | before FE | zero mean/variance is one |
| GSE37472 | carcinoma | getGEO | before FE | zero mean/variance is one |
| GSE49823 | CAD | getGEO | after FE | zero mean/variance is one* |
| GSE43329 | NPC | getGEO | before FE | zero mean/variance is one+ |
| GSE50013 | HCC | getGEO | before FE# | zero mean/variance is one* |
| GSE41922 | BC | GSE41922_series_matrix.txt.gz | after FE | zero mean/variance is one* |
| GSE49665 | AML | getGEO | after FE | zero mean/variance is one* |

*no normalization for SVM/lasso, +no normalization for SVM with PCA-based FE, #after FE for PCA-based LDA with universal features. All the sample normalizations were sample-based; i.e., each sample was normalized to have both zero mean and unit variance. AD, Alzheimer disease; CAD, coronary artery disease; NPC, nasopharyngeal carcinoma; HCC, hepatocellular carcinoma; BC, breast cancer; AML, acute myeloid leukemia. Data retrieval methods/data set names were used to name files and for analysis. getGEO indicates that individual sample profiles whose files names started with "GEO" were downloaded by the getGEO command in R.

adjusted $P$-values less than 0.05 were regarded to have significantly different expression between normal controls and patients.

### KEGG pathway analysis of UDB using DIANA-mirpath

DIANA-mirpath [25] was employed to investigate KEGG pathways enriched by miRNA target genes. Twelve genes were uploaded to DIANA-mirpath with the following settings: "Species" was "Human", "FDR" correction was "yes", "P-value threshold" was 0.05, and "Select the way to merge results" was "pathway union" (direct link to DIANA-mirpath and full list of KEGG pathways are shown in Additional file 3: Text S2 and Additional file 1: Table S5).

## Additional files

**Additional file 1: Supporting Tables.**
**Additional file 2: Supporting Figures.**
**Additional file 3: Supporting Texts.**

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

YHT and YM planned all the projects. YHT performed analyses and wrote the paper. All authors read and approved the final manuscript.

#### Acknowledgments

This research was supported by KAKENHI, 23300357 and 26120528.

#### Author details

[1]Department of Physics, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, 112-8551 Tokyo, Japan. [2]Department of Hepatology, Osaka City University, Graduate School of Medicine, 1-4-3 Asahimachi, Abeno-ku, 545-8585 Osaka, Japan.

**Received: 5 December 2013 Accepted: 14 August 2014**
**Published: 30 August 2014**

#### References

1. Hanash SM, Baik CS, Kallioniemi O: **Emerging molecular biomarkers–blood-based strategies to detect and monitor cancer.** *Nat Rev Clin Oncol* 2011, **8**(3):142–150.
2. Jellinger KA, Janetzky B, Attems J, Kienzl E: **Biomarkers for early diagnosis of Alzheimer disease 'ALZheimer ASsociated gene'–a new blood biomarker?** *J Cell Mol Med* 2008, **12**(4):1094–1117.
3. Petricoin EF, Belluco C, Araujo RP, Liotta LA: **The blood peptidome: a higher dimension of information content for cancer biomarker discovery.** *Nat Rev Cancer* 2006, **6**(12):961–967.
4. Shahzad A, Knapp M, Lang I, Kohler G: **Interleukin 8 (IL-8) - a universal biomarker?** *Int Arch Med* 2010, **3**:11.
5. Fendos J, Engelman D: **pHLIP and acidity as a universal biomarker for cancer.** *Yale J Biol Med* 2012, **85**(1):29–35.
6. Morra R, Munteanu M, Imbert-Bismut F, Messous D, Ratziu V, Poynard T: **FibroMAX: towards a new universal biomarker of liver disease?** *Expert Rev Mol Diag* 2007, **7**(5):481–490.
7. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T: **Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations.** *Proc Natl Acad Sci USA* 2013, **110**(11):4255–4260.
8. Leuenberger N, Robinson N, Saugy M: **Circulating miRNAs: a new generation of anti-doping biomarkers.** *Anal Bioanal Chem* 2013, **405**(309):617–623.
9. Sluijter JP, Doevendans PA: **Circulating microRNA profiles for detection of peripheral arterial disease: small new biomarkers for cardiovascular disease.** *Circ Cardiovasc Genet* 2013, **6**(5):441–443.
10. Wang F, Long G, Zhao C, Li H, Chaugai S, Wang Y, Chen C, Wang D. W: **Plasma microRNA-133a is a new marker for both acute myocardial infarction and underlying coronary artery stenosis.** *J Transl Med* 2013, **11**(1):222.
11. Xiao B, Wang Y, Li W, Baker M, Guo J, Corbet K, Tsalik EL, Li QJ, Palmer SM, Woods CW, Li Z, Chao NJ, He YW: **Plasma microRNA signature as a non-invasive biomarker for acute graft-versus-host disease.** *Blood* 2013, **122**(19):3365–3375.
12. Koberle V, Pleli T, Schmithals C, Augusto Alonso E, Haupenthal J, Bonig H, Peveling-Oberhag J, Biondi RM, Zeuzem S, Kronenberger B, Waidmann O, Piiper A: **Differential stability of cell-free circulating microRNAs implications for their utilization as biomarkers.** *PLoS ONE* 2013, **8**(9):75184.
13. Sheinerman KS, Umansky SR: **Circulating cell-free microRNA as biomarkers for screening, diagnosis and monitoring of neurodegenerative diseases and other neurologic pathologies.** *Front Cell Neurosci* 2013, **7**:150.
14. Recchioni R, Marcheselli F, Olivieri F, Ricci S, Procopio AD, Antonicelli R: **Conventional and novel diagnostic biomarkers of acute myocardial infarction a promising role for circulating microRNAs.** *Biomarkers* 2013, **18**(7):547–558.
15. Dorval V, Nelson PT, Hebert SS: **Circulating microRNAs in Alzheimer's disease: the search for novel biomarkers.** *Front Mol Neurosci* 2013, **6**:24.
16. Deddens JC, Colijn JM, Oerlemans MI, Pasterkamp G, Chamuleau SA, Doevendans PA, Sluijter JP: **Circulating microRNAs as novel biomarkers for the early diagnosis of acute coronary syndrome.** *J Cardiovasc Transl Res* 2013, **6**(6):884–898.
17. Ramshankar V, Krishnamurthy A: **Lung cancer detection by screening - presenting circulating miRNAs as a promising next generation biomarker breakthrough.** *Asian Pac J Cancer Prev* 2013, **14**(4):2167–2172.
18. Sita-Lumsden A, Dart DA, Waxman J, Bevan CL: **Circulating microRNAs as potential new biomarkers for prostate cancer.** *Br J Cancer* 2013, **108**(10):1925–1930.
19. Grasedieck S, Sorrentino A, Langer C, Buske C, Dohner H, Mertens D, Kuchenbauer F: **Circulating microRNAs in hematological diseases: principles, challenges, and perspectives.** *Blood* 2013, **121**(25):4977–4984.
20. Redova M, Sana J, Slaby O: **Circulating miRNAs as new blood-based biomarkers for solid cancers.** *Future Oncol* 2013, **9**(3):387–402.
21. Leidinger P, Backes C, Deutscher S, Schmitt K, Mueller SC, Frese K, Haas J, Ruprecht K, Paul F, Stahler C, Lang CJ, Meder B, Bartfai T, Meese E, Keller A: **A blood based 12-miRNA signature of Alzheimer disease patients.** *Genome Biol* 2013, **14**(7):78.
22. Kumar P, Dezso Z, MacKenzie C, Oestreicher J, Agoulnik S, Byrne M, Bernier F, Yanagimachi M, Aoshima K, Oda Y: **Circulating miRNA biomarkers for Alzheimer's disease.** *PLoS ONE* 2013, **8**(7):69807.
23. Taguchi YH, Murakami Y: **Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers.** *PLoS ONE* 2013, **8**(6):66714.
24. Wehrens R, Franceschi P, Vrhovsek U, Mattivi F: **Stability-based biomarker selection.** *Anal Chim Acta* 2011, **705**(1-2):15–23.
25. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, Paraskevopoulou MD, Prionidis K, Dalamagas T, Hatzigeorgiou AG: **DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways.** *Nucleic Acids Res* 2012, **40**(Web Server issue):498–504.
26. Maclellan SA, Lawson J, Baik J, Guillaud M, Poh CF, Garnis C: **Differential expression of miRNAs in the serum of patients with high-risk oral lesions.** *Cancer Med* 2012, **1**(2):268–274.
27. Shen J, Wang A, Wang Q, Gurvich I, Siegel AB, Remotti H, Santella RM: **Exploration of genome-wide circulating microRNA in hepatocellular carcinoma (HCC): MiR-483-5p as a potential biomarker.** *Cancer Epidemiol Biomarkers Prev* 2013, **22**(12):2364–2373.
28. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, Tan PH, Ho GH, Lee AS: **Identification of circulating microRNA signatures for breast cancer detection.** *Clin Cancer Res* 2013, **19**(16):4477–4487.
29. Rommer A, Steinleitner K, Hackl H, Schneckenleithner C, Engelmann M, Scheideler M, Vlatkovic I, Kralovics R, Cerny-Reiterer S, Valent P, Sill H, Wieser R: **Overexpression of primary microRNA 221/222 in acute myeloid leukemia.** *BMC Cancer* 2013, **13**:364.

30. Murakami Y, Toyoda H, Tanahashi T, Tanaka J, Kumada T, Yoshioka Y, Kosaka N, Ochiya T, Taguchi YH: **Comprehensive miRNA expression analysis in peripheral blood can diagnose liver disease.** *PLoS ONE* 2012, **7**(10):48366.
31. R Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013. http://www.R-project.org/.
32. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc Series B (Methodological)* 1995, **57**(1):289–300.

doi:10.1186/1756-0500-7-581
**Cite this article as:** Taguchi and Murakami: Universal disease biomarker: can a fixed set of blood microRNAs diagnose multiple diseases? *BMC Research Notes* 2014 **7**:581.

Journal of Gastroenterology and Hepatology

JGHF

doi:10.1111/jgh.12376

HEPATOLOGY

# Relationship between inosine triphosphate genotype and outcome of extended therapy in hepatitis C virus patients with a late viral response to pegylated-interferon and ribavirin

Hoang Hai,* Akihiro Tamori,* Masaru Enomoto,* Hiroyasu Morikawa,* Sawako Uchida-Kobayashi,* Hideki Fujii,* Atsushi Hagihara,* Etsushi Kawamura,* Le Thi Thanh Thuy,* Yasuhito Tanaka† and Norifumi Kawada*

*Department of Hepatology, Osaka City University Graduate School of Medicine, Osaka, and †Department of Virology and Liver Unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

**Key words**
extended therapy, HCV, *ITPA* genotype, treatment outcome.

Accepted for publication 12 August 2013.

**Correspondence**
Dr Akihiro Tamori, Department of Hepatology, Osaka City University Graduate School of Medicine, 1-4-3, Asahi-machi, Abenoku, Osaka 545-8585, Japan. Email: atamori@med.osaka-cu.ac.jp

Conflict of interest: In the past year, Dr Akihiro Tamori has received research funding from MSD K.K. Dr Yasuhito Tanaka has consulted for MSD K.K and Chugai Pharmaceutical Co., Ltd., and has received research funding from MSD K.K and Chugai Pharmaceutical Co., Ltd. Dr Norifumi Kawada has consulted for MSD K.K and Chugai Pharmaceutical Co., Ltd., and has received research funding from MSD K.K and Chugai Pharmaceutical Co., Ltd. Other authors report no conflicts of interest.

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see http://www.textcheck.com/certificate/mXLRyV

## Abstract

**Background and Aim:** It is not yet clear which factors are associated with the outcome of 72-week treatment with pegylated-interferon and ribavirin (RBV) in patients with chronic hepatitis C virus (HCV) infection.

**Methods:** In 66 patients with HCV genotype 1 who had a late viral response (LVR) to 72-week treatment of pegylated-interferon and RBV, we examined the factors that determined the outcome, including single nucleotide polymorphisms of interleukin-28B and inosine triphosphatase (*ITPA*) genes.

**Results:** Thirty seven of 66 (56%) patients with LVR achieved a sustained viral response (SVR). The mean age of these 37 SVR patients was 55, compared with 61 in 29 relapsed patients ($P = 0.009$). Twenty six of 54 (48%) patients with the CC genotype and 11 of 12 (92%) with the CA/AA genotype of *ITPA* rs1127354 achieved SVR ($P = 0.006$). The SVR rates were 79%, 40%, 60%, and 33% in patients with undetectable HCV RNA on weeks 16, 20, 24, and 28 or later, respectively ($P = 0.014$). Finally, serum RBV concentration at week 44 of treatment was significantly higher in the SVR group (2651 ng/mL) than in the relapse group (1989 ng/mL, $P = 0.002$). In contrast, the rate of the interleukin-28B genotype was not different between the groups. Multiple regression analysis showed that age < 60 years, *ITPA* CA/AA genotype, and serum RBV concentration were significant independent predictive factors for SVR.

**Conclusions:** Our findings elucidated the association of four factors, including *ITPA* genotype, with the outcome of 72-week treatment in LVR patients.

## Introduction

Hepatitis C virus (HCV) infection continues to be a major cause of liver cirrhosis and hepatocellular carcinoma.[1] An estimated 120–130 million people worldwide are infected with HCV.[2] Sustained viral response (SVR), defined as undetectable serum HCV RNA levels 24 weeks after cessation of therapy, is the aim of treatment. Although the current treatment regimen of pegylated-interferon (PEG-IFN) combined with ribavirin (RBV) greatly improved SVR in patients with HCV genotypes 2 and 3, the outcomes in patients with HCV genotype 1 and high viral load (> $10^5$ IU/mL) remain unsatisfactory, and SVR is attained in approximately 50% of cases.[3–8]

For HCV genotype 1, patients with rapid viral response, defined as undetectable serum HCV on week 4, achieve high rates of SVR up to 91% with combination therapy. Patients with early viral

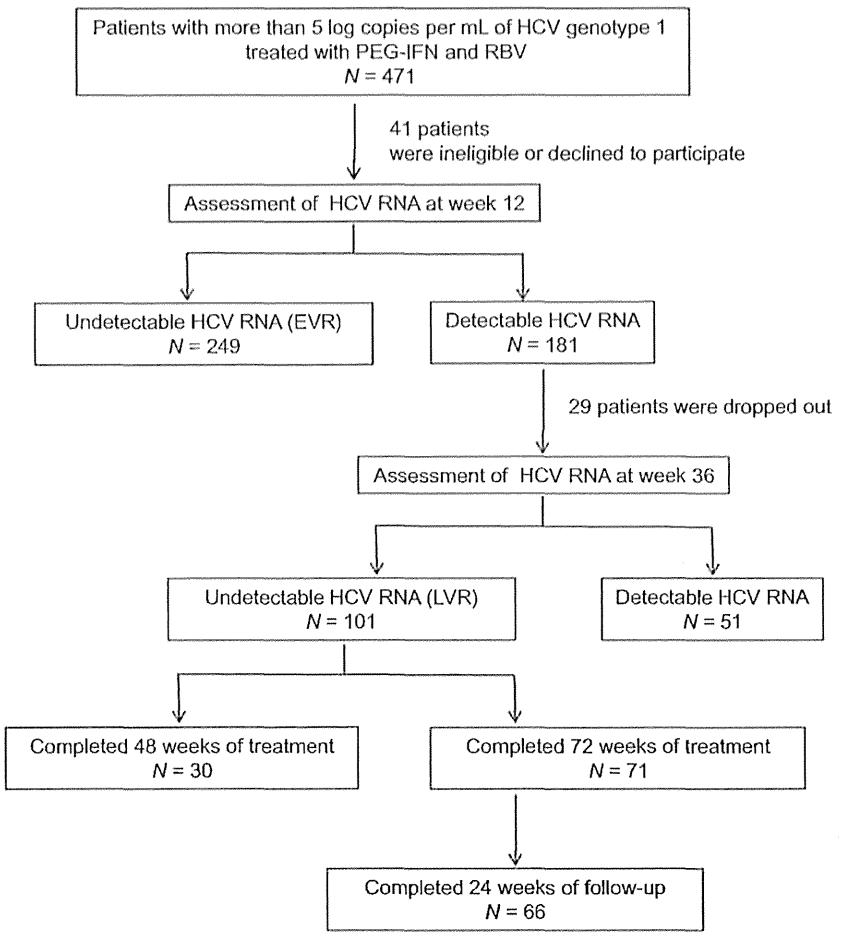


**Figure 1** Flow of participants throughout the study. EVR, early viral response; HCV, hepatitis C virus; LVR, late viral response; PEG-IFN, pegylated-interferon; RBV, ribavirin.

response, defined as undetectable serum HCV on week 12, achieved SVR rates of 65–81%. However, patients with a late viral response (LVR), who remained positive for HCV RNA on week 12 after the start of treatment but became negative for HCV RNA during weeks 13–36 of treatment, showed a lower SVR rate of 14–44%.[4,9–19] Although extending therapy to 72 weeks has been reported to decrease relapse in such patients,[12–17,20,21] it remains unclear which patient with LVR can benefit from extended treatment.

Inosine triphosphatase (*ITPA*) single nucleotide polymorphism (SNP) rs1127354, causing ITPase deficiency, was found to be associated with protection from RBV-induced anemia and to decrease the need for RBV dose reduction, but not to be associated with clinical outcome.[22–25] The present study was performed to identify that factors, including interleukin-28B (*IL28B*) and *ITPA* genotype, associated with the outcome of extended 72-week treatment in patients with HCV genotype 1 who had LVR to PEG-IFN and RBV.

## Methods

***Patients.*** A total of 471 patients were recruited at Osaka City University Hospital between December 2004 and June 2012. The flow of patients through the trial is presented in Figure 1. Sixty-six patients with HCV genotype 1 who were treated with PEG-IFN alpha 2a (Pegasys; Chugai Pharmaceutical Co., Ltd, Tokyo, Japan) or 2b (Pegintron; MSD, Osaka, Japan) and RBV (Rebetol, MSD) combination therapy were enrolled in this study. All patients had a viral load of $> 10^5$ IU/mL according to COBAS Amplicor HCV Monitor test, version 2.0 (Roche Diagnostics, Branchburg, NJ, USA), or a viral load of > 5 log copies/mL as determined by COBAS TaqMan HCV test (Roche Diagnostics). HCV RNA levels were investigated before and every 4 weeks after the start of treatment. All patients gave written informed consent to participate in this study, in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and according to the process approved by the ethical committee of Osaka City University, Graduate School of Medicine. Only the patients who completed 72-week combination therapy without discontinuation and in whom HCV RNA was detected on week 12 but not on weeks 13–36 were enrolled in this study.

Exclusion criteria included a history or evidence of a serious chronic or poorly controlled medical or psychiatric condition, infection with human immunodeficiency virus or hepatitis B virus, and receipt of systemic immunomodulatory or antineoplastic therapy within the previous 6 months. Pregnant or breastfeeding women and partners of pregnant women were also excluded.

The following factors were analyzed to determine whether they were related to the efficacy of combination therapy: patient age, gender, pretreatment biochemical parameters, such as neutrophil and platelet counts, hemoglobin concentration, levels of alanine transaminase, creatinine, HCV viral load, histopathological evaluation of hepatitis activity and hepatic fibrosis according to the METAVIR scoring system, total doses of PEG-IFN and RBV, and serum RBV concentration at week 44.

***Treatment protocol.*** The initial dose of PEG-IFN alpha 2a was 180 μg per week, and that of PEG-IFN alpha 2b was 1.5 μg per kg body weight per week. The initial dose of RBV was 400, 600, 800, or 1000 mg/day for patients weighing < 40 kg, 40–60 kg, 60–80 kg, or > 80 kg, respectively.

***RBV concentration.*** Serum RBV concentration was measured using an assay consisting of phenylboronic acid solid phase extraction, followed by HPLC at a commercial laboratory (SRL Inc., Osaka, Japan).[26] Briefly, the RBV concentrations in 200-μL samples were measured by validated HPLC with column switching. Serum samples deproteinized with perchloric acid were injected into the column, and RBV was detected by monitoring absorption of ultraviolet at 215 nm. The calibration curve was linear in the range of 50–20 000 ng/mL. A set of calibration standards at 0, 5, 10, 25, 50, 100, 250, 500, 1000, 2000, and 5000 mg/L RBV was prepared, extracted and analyzed with each series, together with internal quality controls at three levels.

***SNP genotyping.*** We examined genetic polymorphisms of the *IL28B* and *ITPA* genes in patients who consented to genome analysis. Whole blood was collected from all patients and centrifuged to separate the buffy coat. Genomic DNA was extracted from the buffy coat using a QIAamp DNA Blood Midi Kit (Qiagen Sciences Inc, Germantown, MD, USA). Genetic polymorphisms of *IL28B* rs8099917 and rs12979860 and *ITPA* rs1127354 were genotyped by TaqMan SNP Genotyping Assay on the 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). All samples were also genotyped by direct sequencing to confirm the genotype. Exon 2 of the *ITPA* gene and flanking intronic regions were amplified by polymerase chain reaction (PCR) using the following primers: forward, 5′-CTTTAGG AGATGGGCAGCAG-3′; reverse, 5′-CACAGAAAGTCAGGTC ACAGG-3′.[27] PCR was carried out in a total volume of 15 μL with 1× Premix Ex Tag (Applied Biosystems), 300 nM of each primer, and 100 ng of genomic DNA. The PCR profile consisted of 94°C for 10 min, followed by 35 cycles of 94°C for 30 s, 63°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 7 min. PCR products were sequenced bidirectionally using a BigDye Terminator v3.1 Cycle Sequencing Kit and ABI 3130XL Genetic Analyzer (Applied Biosystems). Genotyping analysis was permitted by the ethical committee of our university (approval number 1871).

***Statistical analysis.*** All data analyses were conducted using the JMP program, version 9.0 (SAS Institute, Cary, NC, USA). Individual characteristics between groups were evaluated by Wilcoxon's two-sample test for numerical variables, or Fisher's exact test for categorical variables. Variables exhibiting values of $P < 0.1$ on univariate analysis were subjected to stepwise multivariate logistic regression analysis. In the two-tailed test, $P < 0.05$ was taken to indicate statistical significance.

**Table 1** Characteristics of HCV genotype 1 patients with late virologic response†

| | Total ($n$ = 66) |
|---|---|
| Age (years) | 60 (21–77) |
| Gender (male/female) | 43/23 |
| HCV viral load (log copies/mL) | 6.5 (5.1–7.7) |
| Body mass index (kg/m$^2$) | 22.7 (16.5–28.0) |
| WBC (/μL) | 4600 (2000–8600) |
| Hb (g/dL) | 13.6 (10.9–17.3) |
| Platelet (× 10$^4$/μL) | 17.1 (8.5–45.0) |
| ALT (IU/L) | 54 (17–194) |
| Creatinine (mg/dL) | 0.66 (0.36–1.16) |
| *IL28B* rs8099917 (TT/TG+GG) | 45/21 |
| *IL28B* rs12979860 (CC/CT+TT) | 44/22 |
| *ITPA* rs1127354 (CC/CA+AA) | 54/12 |
| Liver biopsy | |
| Activity grading (0/1/2/3/ND) | 7/40/13/1/5 |
| Fibrosis staging (1/2/3/4/ND) | 38/15/8/0/5 |
| PEG-IFNα2a/PEG-IFNα2b | 12/54 |
| Week when HCV RNA was undetectable (16 weeks/20 weeks/24 weeks/28 weeks or delayed) | 24/20/10/12 |
| RBV concentration at week 44 (ng/mL) | 2467 (1006–4283) |
| Total dose of administered RBV (g/kg body weight) | 5.07 (1.56–7.21) |

†Continuous variables are medians (min-max).
ALT, alanine transaminase; Hb, hemoglobin; HCV, hepatitis C virus; *IL28B*, interleukin 28B gene; *ITPA*, inosine triphosphatase gene; ND, not done; PEG-IFN, pegylated-interferon; RBV, ribavirin; WBC, white blood cell.

## Results

***Patient profile and response rate.*** The characteristics of the overall 66 LVR patients, consisting of 43 men and 23 women, are shown in Table 1. The mean age of this cohort was 60 years. All of the patients who were infected with HCV genotype 1 with viral load > 5 log copies/mL, were treated with PEG-IFN/RBV for 72 weeks. HCV RNA was tested 24 weeks after completion of treatment when SVR and relapse were defined if HCV RNA was negative and positive, respectively. After 72 weeks of combination therapy, 37 (56%) patients achieved SVR, while the remaining 29 (44%) relapsed.

Direct sequencing and TaqMan SNP genotyping assay were used to genotype SNP *ITPA* rs1127354, and 100% of SNP results were concordant between both methods. Among the 66 LVR patients, 54 (82%) had the major CC genotype (wild-type), 10 (15%) were heterozygous for the CA genotype, and the remaining 2 (3%) had the minor AA genotype.

***Association between clinical factors and SVR rate.*** Among the 17 factors screened by univariate analysis, four factors were associated with treatment response, that is, patient

**Table 2** Comparison of the clinical characteristics of patients with SVR and those with relapse†

| | SVR ($n = 37$) | Relapse ($n = 29$) | *P*-value |
|---|---|---|---|
| Age (years) | 55 ± 7 | 61 ± 7 | 0.009* |
| Gender (male/female) | 24/13 | 19/10 | 0.956 |
| HCV viral load (log copies/mL) | 6.6 ± 0.4 | 6.4 ± 0.5 | 0.195 |
| Body mass index (kg/m$^2$) | 22.6 ± 2.9 | 22.1 ± 2.5 | 0.438 |
| WBC (/μL) | 4748 ± 1235 | 4693 ± 1281 | 0.861 |
| Hb (g/dL) | 13.9 ± 1.3 | 13.9 ± 1.6 | 0.892 |
| Platelets (× 10$^4$/μL) | 18.2 ± 7.8 | 16.6 ± 4 | 0.698 |
| ALT (IU/L) | 63.8 ± 40.6 | 62.6 ± 37.9 | 0.861 |
| Creatinine (mg/dL) | 0.71 ± 0.19 | 0.67 ± 0.12 | 0.473 |
| *IL28B* rs8099917 (TT/TG+GG) | 25/12 | 20/9 | 0.904 |
| *IL28B* rs12979860 (CC/CT+TT) | 24/13 | 20/9 | 0.930 |
| *ITPA* rs1127354 (CC/CA+AA) | 26/11 | 28/1 | 0.006* |
| Liver biopsy | | | |
| Activity grading (0/1/2/3/ND) | 3/24/6/1/3 | 4/16/7/0/2 | 0.563 |
| Fibrosis staging (1/2/3/4/ND) | 20/11/3/0/3 | 18/4/5/0/2 | 0.211 |
| PEG-IFNα2a/PEG-IFNα2b | 6/31 | 6/23 | 0.64 |
| Week when HCV RNA was undetectable (16 weeks/20 weeks/24 weeks/28 weeks or delayed) | 19/8/6/4 | 5/12/4/8 | 0.014* |
| RBV concentration at week 44 (ng/mL) | 2651 ± 675 | 1989 ± 525 | 0.002* |
| Total dose of RBV administered (g/kg body weight) | 5.08 ± 1.3 | 4.59 ± 1.11 | 0.059 |

*$P < 0.05$.

†Continuous variables are medians (min-max).

ALT, alanine transaminase; Hb, hemoglobin; HCV, hepatitis C virus; *IL28B*, interleukin 28B gene; *ITPA*, inosine triphosphatase gene; ND, not done; PEG-IFN, pegylated interferon; RBV, ribavirin; WBC, white blood cell.

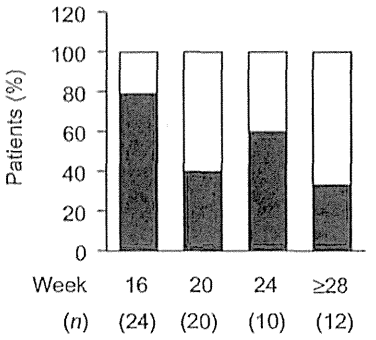


**Figure 2** Effects of combination therapy in patients with genotype 1 according to the time at which HCV was undetectable (week). Earlier HCV RNA disappearance was significantly associated with treatment outcome ($P = 0.014$). SVR rates were 79% (19/24) in patients with undetectable HCV RNA at week 16, 40% (8/20) at week 20, 60% (6/10) at week 24, and 33% (4/12) at week 28 or delayed. HCV, hepatitis C virus; SVR, sustained viral response. (□) Relapse, (■) SVR.

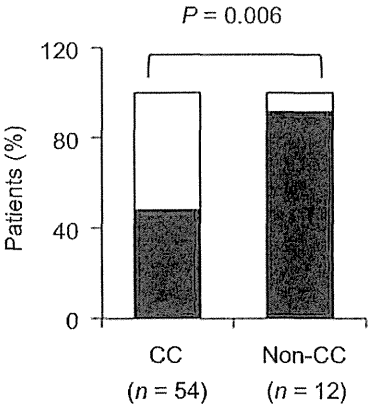


**Figure 3** Effects of combination therapy in patients with genotype 1 according to *ITPA* SNP rs1127354 genotype. SVR (black bar) was achieved in 48% (26/54) of patients with the rs1127354 CC genotype, and in 92% (11/12) of those with a non-CC genotype at rs1127354. Patients with the rs1127354 CA/AA genotype were significantly more likely to be associated with SVR ($P = 0.006$). In the relapse group (white bar), the major CC allele occurred in 28/54 patients but the minor CC allele in only 1/12 patients. *ITPA*, inosine triphosphatase gene; SNP, single nucleotide polymorphism; SVR, sustained viral response. (□) Relapse, (■) SVR.

age, *ITPA* SNP rs1127354, time of undetectable HCV RNA, and RBV concentration (Table 2). The mean age of patients with SVR was significantly younger than that of patients with relapse (55 *vs* 61 years, respectively, $P = 0.009$). Eleven of 37 (30%) patients with SVR and 1 of 29 (3%) patients with relapse had the CA/AA genotype of *ITPA*, indicating a significant association between the CA/AA genotype and SVR ($P = 0.006$). In contrast, the proportion of the *IL28B* genotype was not different between patients with SVR and relapse. Earlier HCV RNA disappearance was significantly associated with treatment outcome ($P = 0.014$); SVR rate was 79% (19/24) in patients with undetectable HCV RNA on week 16, 40% (8/20) on week 20, 60% (6/10) on week 24, and 33% (4/12) on or after week 28 (Fig. 2). Finally, when RBV concentration in the peripheral blood was examined on week 44 of treatment, it was significantly higher in the SVR group (2651 ng/mL) than the relapse group (1989 ng/mL, $P = 0.002$).

***Association between SNP ITPA rs1127354 and clinical factors.*** Twenty six of 54 (48%) patients with the CC genotype and 11 of 12 (92%) with the CA/AA genotype achieved SVR (Fig. 3), indicating a significant association between the CA/AA genotype and SVR ($P = 0.006$). The decline in hemoglobin concentration on week 12 from the baseline was 3.56 g/dL in patients with the CC genotype, compared with 2.16 g/dL in CA/AA patients ($P = 0.0004$, Fig. 4a).

Evaluation of the association between SNP rs1127354 and RBV concentration or total dose of administered RBV showed no significance ($P = 0.27$ and 0.65, respectively) (Fig. 4b,c).

***Independent predictive factors of combination therapy for SVR.*** Factors exhibiting values of $P < 0.1$ on univariate analysis were age, *ITPA* genotype, week at which HCV

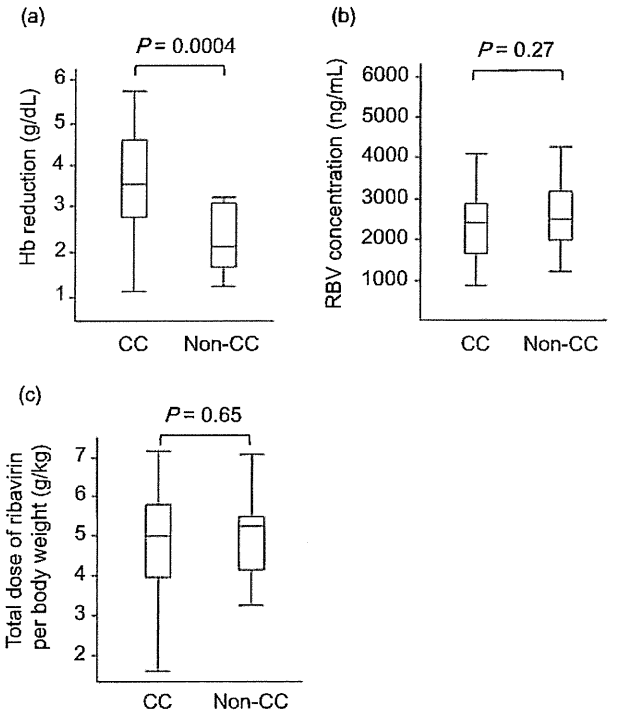


**Figure 4** Association between *ITPA* polymorphism and clinical factors: hemoglobin reduction at week 12 (a), ribavirin concentration at week 44 (b), and total dose of administered ribavirin (mg/kg of body weight (c). Hb reduction in wild-type (CC) was significantly higher than those with heterozygous (CA) or homozygous (AA) rs1127354 (3.56 *vs* 2.16 g/dL, respectively, $P = 0.0004$). There were no significant associations between *ITPA* SNP rs1127354 and ribavirin concentration and total ribavirin dose administered ($P = 0.27$ and 0.65, respectively). *ITPA*, inosine triphosphatase gene; SNP, single nucleotide polymorphism.

RNA was undetectable, RBV concentration, and total dose of RBV administered. These factors were categorized below: (i) younger or older than 60 years, (ii) CC or non-CC genotype of *ITPA* SNP rs1127354, (iii) HCV RNA undetectable at < 24 weeks or ≥ 24 weeks, (iv) RBV concentration < 2500 ng/mL or ≥ 2500 ng/mL, and (v) total RBV dose of < 4.9 g/kg or ≥ 4.9 g/kg. Multiple regression analysis indicated that age, *ITPA* rs1127354, and RBV concentration were significant independent predictive factors for SVR ($P = 0.002$, 0.006, and 0.045, respectively Table 3).

## Discussion

Previous studies have shown that extended 72-week combination therapy with PEG-IFN/RBV improves SVR rate,[14,15] while extended treatment is recommended only for HCV genotype 1 infection with LVR but not for general HCV patients.[13] However, Buti *et al.* showed that SVR rates were similar among LVR patients who received a standard dose of PEG-IFN alpha-2b and weight-based RBV for 48 or 72 weeks.[17] Although the overall SVR rate has been shown to improve in patients with LVR, it is necessary to determine which group of patients can benefit from extended therapy. The present study showed that age, timing of

**Table 3** Multiple regression analysis

| | Odds ratio (95% CI) | *P*-value |
|---|---|---|
| Age (≥ 60 years/< 60 years) | 9.7 (1.8–82.6) | 0.005* |
| *ITPA* rs1127354 (CA/AA *vs* CC) | 15.8 (1.7–415) | 0.012* |
| At week of undetectable HCV RNA (> 24 weeks/≤ 24 weeks) | 1.1 (0.2–6.4) | 0.897 |
| Ribavirin concentration on week 44 (≥ 2500 ng/mL/< 2500 ng/mL) | 12 (2.2–105.4) | 0.003* |
| Total dose of ribavirin administered (≥ 4.9 g/kg/< 4.9 g/kg) | 2 (0.4–9.7) | 0.361 |

*$P < 0.05$.
HCV, hepatitis C virus; *ITPA*, inosine triphosphatase gene.

HCV RNA disappearance, serum RBV concentration, and *ITPA* SNP rs1127354 were related to the outcome of 72-week PEG-IFN/RBV therapy for patients with LVR.

However, *IL28B* SNPs were not associated with the outcome of 72-week treatment in patients with LVR. *IL28B* SNP was originally reported as a host marker to predict null responders to 48-week treatment.[28] The patients enrolled in our study were late viral responders, but not null responders. Including only patients with a specific on-treatment viral response may reduce the influence of *IL28B* SNP on the outcome. Our results are consistent with those of Mangia *et al.*,[29] showing that *IL28B* genotyping had limited clinical utility in the arrangement of response-guided therapy for patients with genotype 1.

In contrast, 11 (92%) of 12 patients with CA or AA at *ITPA* SNP rs1127354 achieved SVR among 66 patients with LVR. Polymorphic variation in the *ITPA* gene causing ITPase deficiency leads to an elevated concentration of inosine triphosphate (ITP) in erythrocytes. Similarly, RBV-induced anemia is triggered by the accumulation of RBV active forms of triphosphate (RBV-TP) in erythrocytes. ITP competes with RBV-TP, thus protecting cells from the lytic effects of RBV-TP. Patients with the rs1127354 CA/AA genotype have a lower risk for a hemoglobin decline of > 3 g/dL.[22,30] In fact, we found that hemoglobin was significantly lower in patients with the CC genotype than in those with the CA/AA genotype during the initial 12 weeks of treatment (Fig. 4a). It has been reported that a cumulative reduction in RBV is more frequent in patients with the CC genotype than in patients who are non-CC. Additionally, *ITPA* SNP rs1127354 is one of the predictive factors for SVR.[31] However, other studies have shown that *ITPA* SNP is associated with RBV-induced anemia but not with treatment outcome in patients who undergo standard therapy.[22–25] In the 165 patients who underwent 48 weeks of therapy in our hospital, *ITPA* genotype was not related to outcomes of patients who underwent standard therapy (data not shown). In the present study, LVR patients were the subjects. We speculate that LVR patients have different clinical backgrounds, including genotype, related to outcome of PEG-IFN and RBV combination therapy. In a subset of patients with the favorable TT genotype of *IL28B* SNP rs8099917, rs1127354 SNP of *ITPA* seemed to be associated with the outcome of combination therapy.[32] This is the first study to demonstrate an association between *ITPA* SNP and SVR rate in LVR patients who underwent extended treatment.

It is unclear why the *ITPA* genotype was associated with outcomes of LVR patients who underwent extended treatment. It has been reported that expression of several genes before combination therapy is related to *ITPA* genotype.[30] One of these might play an important role in the response to elongated therapy. In the present study, seven of 10 patients with *ITPA* non-CC type showed > 2500 ng/mL RBV at week 44. In contrast, 19 of 41 patients with the *ITPA* CC type had > 2500 ng/mL RBV at week 44. Many patients with *ITPA* non-CC type had > 2500 ng/mL RBV at week 44.

We did not detect associations between *ITPA* variants and RBV concentrations at week 44 (Fig. 4b). RBV concentration is affected by both the dose administered and its clearance; the latter is regulated by renal function.[33] Serum creatinine level was within the normal range in the patients included in the present study, indicating that their renal function is sufficient to receive RBV adjusted by body weight. The RBV dose administered is dependent on body weight and is correlated with RBV-related adverse events, particularly anemia. Recently, it was reported that both *SLC28A2* rs11854484 genotype and *ITPA* genotype were related to RBV-related anemia. However, the factor associated with RBV concentration at weeks 4 and 8 was the *SLC28A2* rs11854484 genotype, but not the *ITPA* genotype.[34] In patients with LVR, RBV concentration and *ITPA* genotype were independently associated with the outcome of extended treatment (Table 3).

Our data suggest that serum RBV concentration at week 44 was significantly higher in patients with SVR than in those with relapse ($P = 0.002$). On the other hand, total dosage of RBV was not related to the outcome of extended therapy. In previously published data regarding 48-week therapy, both the RBV dose administered and the RBV concentration in peripheral blood were associated with the outcome of combination therapy with PEG-IFN and RBV.[35,36] Furusyo *et al.* reported that in both groups with < 60% and ≥ 60% of RBV assigned total dosage, the mean RBV concentration at 48 weeks in patients with SVR was > 1500 ng/mL and was significantly higher than in those with relapse, suggesting that RBV concentration was unaffected by the assigned total dosage.[37] In the present study, no association between RBV concentration on week 44 and the total dose of RBV administered was identified (data not shown).

Many novel interferon-free antiviral regimens for HCV are now under clinical investigation. Some of these include RBV in combination with one or two direct-acting antiviral agents.[38,39] RBV will remain a key drug for treatment of chronic HCV infection in the forthcoming era of oral combination antiviral therapy. Further studies are required to evaluate the significance of *ITPA* SNP as predictors of not only RBV-induced anemia but also of treatment outcome.

In conclusion, age, RBV concentration, timing of HCV RNA disappearance, and *ITPA* SNP rs1127354 were associated with a higher SVR rate in LVR patients given 72-week treatment. These predictive factors may allow more efficient extended treatment with PEG-IFN and RBV for patients with LVR.

## Acknowledgments

We thank Ms. Yoko Yasuhara and Ms. Sanae Deguchi for their assistance in data/sample collection.

## References

1 Shepard CW, Finelli L. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* 2005; **5**: 558–67.
2 Global Burden of Hepatitis C Working Group. Global burden of disease (GBD) for hepatitis C. *J. Clin. Pharmacol.* 2004; **44**: 20–9.
3 Manns MP, McHutchison JG, Gordon SC *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomized trial. *Lancet* 2001; **358**: 958–65.
4 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.* 2002; **347**: 975–82.
5 Hadziyannis S, Sette H Jr, Morgan TR *et al.* Peginterferon-alfa-2a plus ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.* 2004; **40**: 346–55.
6 Zeuzem S, Hultcrantz R, Bourliere M *et al.* Peginterferon alfa-2b plus ribavirin for treatment of chronic hepatitis C in previously untreated patients infected with HCV genotypes 2 or 3. *J. Hepatol.* 2004; **40**: 993–9.
7 Yamaguchi Y, Tamori A, Tanaka Y *et al.* Response-guided therapy for patients with chronic hepatitis who have high viral loads of hepatitis C virus genotype 2. *Hepatol. Res.* 2012; **42**: 549–57.
8 Zeuzem S. Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well? *Ann. Intern. Med.* 2004; **140**: 370–81.
9 Ferenci P, Fried MW, Shiffman ML *et al.* Predicting sustained virologic responses in chronic hepatitis C patients treated with peginterferon alfa-2a (40 KD)/ribavirin. *J. Hepatol.* 2005; **43**: 425–33.
10 Jensen DM, Morgan TR, Marcellin P *et al.* Early identification of HCV genotype 1 patients responding to 24 weeks peginterferon alpha-2a (40 kd)/ribavirin therapy. *Hepatology* 2006; **43**: 954–60.
11 Dienstag JL, McHutchison JG. American Gastroenterological Association medical position statement on the management of hepatitis C. *Gastroenterology* 2006; **130**: 225–30.
12 Ferenci P, Laferl H, Scherzer TM *et al.* Peginterferon alfa-2a/ribavirin for 48 or 72 weeks in hepatitis C genotypes 1 and 4 patients with slow virologic response. *Gastroenterology* 2010; **138**: 503–12.
13 Berg T, von Wagner M, Nasser S *et al.* Extended treatment duration for hepatitis C virus type 1: comparing 48 versus 72 weeks of peginterferon-alfa-2a plus ribavirin. *Gastroenterology* 2006; **130**: 1086–97.
14 Sanchez-Tapias JM, Diago M, Escartin P *et al.* Peginterferon-alfa2a plus ribavirin for 48 versus 72 weeks in patients with detectable hepatitis C virus RNA at week 4 of treatment. *Gastroenterology* 2006; **131**: 451–60.
15 Pearlman BL, Ehleben C, Saifee S. Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis C genotype 1-infected slow responders. *Hepatology* 2007; **46**: 1688–94.
16 Mangia A, Minerva N, Bacca D *et al.* Individualized treatment duration for hepatitis C genotype 1 patients: a randomized controlled trial. *Hepatology* 2008; **47**: 43–50.
17 Buti M, Lurie Y, Zakharova N *et al.* Randomized trial of peginterferon alfa-2b and ribavirin for 48 or 72 weeks in patients with HCV genotype 1 and slow virologic response. *Hepatology* 2010; **52**: 1201–7.
18 McHutchison JG, Lawitz EJ, Shiffman ML *et al.* Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N. Engl. J. Med.* 2009; **361**: 580–93.
19 Oze T, Hiramatsu N, Yakushijin T *et al.* Pegylated interferon alpha-2b (PEG-IFN a-2b) affects early virologic response

dose-dependently in patients with chronic hepatitis C genotype 1 during treatment with PEG-IFN a-2b plus ribavirin. *J. Viral Hepat.* 2009; **16**: 578–85.
20 Oze T, Hiramatsu N, Yakushijin T *et al.* The efficacy of extended treatment with pegylated interferon plus ribavirin in patients with HCV genotype 1 and slow virologic response in Japan. *J. Gastroenterol.* 2011; **46**: 944–52.
21 Di Martino V. Response-guided peg-interferon plus ribavirin treatment duration in chronic hepatitis C: meta-analyses of randomized, controlled trials and implications for the future. *Hepatology* 2011; **54**: 789–800.
22 Fellay J, Thompson AJ, Ge D *et al.* ITPA gene variants protect against anemia in patients treated for chronic hepatitis C. *Nature* 2010; **464**: 405–8.
23 Ochi H, Maekawa T, Abe H *et al.* ITPA polymorphism affects ribavirin-induced anemia and outcomes of therapy-A genome-wide study of Japanese HCV virus patients. *Gastroenterology* 2010; **139**: 1190–7.
24 Thompson AJ, Fellay J, Patel K *et al.* Variants in the ITPA gene protect against ribavirin-induced hemolytic anemia and decrease the need for ribavirin dose reduction. *Gastroenterology* 2010; **139**: 1181–9.
25 D'Avolio A, Ciancio A, Siccardi M *et al.* Inosine triphosphatase polymorphisms and ribavirin pharmacokinetics as determinants of ribavirin-associate anemia in patients receiving standard anti-HCV treatment. *Ther. Drug Monit.* 2012; **34**: 165–70.
26 Tsubota A, Horose Y, Izumi N *et al.* Pharmacokinetics of ribavirin in combined interferon-alpha 2b and ribavirin therapy for chronic hepatitis C virus infection. *Br. J. Clin. Pharmacol.* 2003; **55**: 360–7.
27 Bakker JA, Lindhout M, Habets DD, van den Wijngaard A, Paulussen AD, Bierau J. The effect of ITPA polymorphisms on the enzyme kinetic properties of human erythrocyte inosine triphosphatase toward its substrates ITP and 6-Thio-ITP. *Nucleosides Nucleotides Nucleic Acids* 2011; **30**: 839–49.
28 Tanaka Y, Nishida N, Sugiyama M *et al.* Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* 2009; **41**: 1105–9.
29 Mangia A, Thompson AJ, Santoro R *et al.* Limited use of interleukin 28B in the setting of response-guided treatment with detailed on-treatment virological monitoring. *Hepatology* 2011; **54**: 772–80.
30 Birerdinc A, Estep M, Afendy A *et al.* Gene expression profiles associated with anemia and ITPA genotypes in patients with chronic hepatitis C (CH-C). *J. Viral Hepat.* 2012; **19**: 414–22.
31 Azakami T, Hayes CN, Sezaki H *et al.* Common genetic polymorphism of ITPA gene affects ribavirin-induced anemia and effect of peg-interferon plus ribavirin therapy. *J. Med. Virol.* 2011; **83**: 1048–57.
32 Kurosaki M, Tanaka Y, Tanaka K *et al.* Relationship between polymorphisms of the inosine triphosphatase gene and anemia or outcome after treatment with pegylated interferon and ribavirin. *Antivir. Ther.* 2011; **16**: 685–94.
33 Morello J, Rodríguez-Novoa S, Jiménez-Nácher I, Soriano V. Usefulness of monitoring ribavirin plasma concentrations to improve treatment response in patients with chronic hepatitis C. *J. Antimicrob. Chemother.* 2008; **62**: 1174–80.
34 Rau M, Stickel F, Russmann S *et al.* Impact of genetic SLC28 transporter and ITPA variants on ribavirin serum level, hemoglobin drop and therapeutic response in patients with HCV infection. *J. Hepatol.* 2013; **58**: 669–75.
35 Ogawa E, Furusyo N, Kajiwara E *et al.* An inadequate dose of ribavirin is related to virological relapse by chronic hepatitis C patients treated with pegylated interferon alpha-2b and ribavirin. *J. Infect. Chemother.* 2012; **18**: 689–97.
36 Kurosaki M, Hiramatsu N, Sakamoto M *et al.* Age and total ribavirin dose are independent predictors of relapse after interferon therapy in chronic hepatitis C revealed by data mining analysis. *Antivir. Ther.* 2012; **17**: 35–43.
37 Furusyo N, Murata M, Ogawa E *et al.* Ribavirin concentration in the later stages of 48 week pegylated interferon-alpha2b plus ribavirin therapy for chronic hepatitis C is useful for predicting virological response. *J. Antimicrob. Chemother.* 2011; **66**: 1127–39.
38 Gane EJ, Stedman CA, Hyland RH *et al.* Nucleotide polymerase inhibitor sofosbuvir plus ribavirin for hepatitis C. *N. Engl. J. Med.* 2013; **368**: 34–44.
39 Poordad F, Lawitz E, Kowdley KV *et al.* Exploratory study of oral combination antiviral therapy for hepatitis C. *N. Engl. J. Med.* 2013; **368**: 45–53.