17 of these patients had either idiopathic interstitial pneumonia or interstitial lung disease associated with collagen vascular disease. In addition, the median KL-6 level in the PCP-negative patients was 543 U/mL, which was higher than the normal limit (< 500 U/mL). This suggests that the highly elevated KL-6 in these patients could be associated with the underlying disease rather than being a false-positive finding. There was no significant difference in serum CRP levels between the PCP-positive and PCP-negative patients ($p = 0.70$; Fig 1, *bottom right, D*).

ROC curves constructed for each serum marker are presented in Figure 2. The areas under the curve values were 0.935 for β-D-glucan, 0.730 for KL-6, and 0.704 for LDH. The area under the ROC curve for β-D-glucan was significantly larger than that for the other two markers ($p < 0.001$), suggesting that β-D-glucan should be the most reliable for the diagnosis of PCP among the serum markers evaluated. Based on the ROC curves, the cut-off levels of β-D-glucan and LDH were estimated as 31.1 pg/mL and 268 IU/L, respectively. With this cut-off point, the sensitivity and specificity of β-D-glucan were 92.3% and 86.1%, respectively. The positive predictive value (PPV) and negative predictive value (NPV) of β-D-glucan were 0.610 and 0.980, respectively. The sensitivity of LDH was 86.0%, and the specificity was 45.3%. The PPV and NPV of LDH were 0.237 and 0.942, respectively.

There were 14 PCP-negative patients with serum levels of β-D-glucan > 31.1 pg/mL, the cut-off level; because of elevated serum Aspergillus galactomannan antigen levels, deep-seated aspergillosis was diagnosed in 7 of them. Although it was not clear whether the rest of the PCP-negative patients with elevated β-D-glucan had deep-seated mycosis, cytomegalovirus pneumonia was diagnosed in three patients and alveolar hemorrhage in two patients. We concluded that latent deep-seated mycosis, especially aspergillosis, could be the main cause of the false-positive finding.

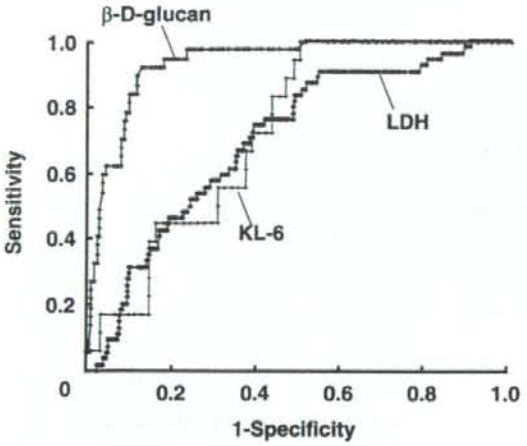

FIGURE 2. ROC curves for LDH, β-D-glucan, and KL-6. The areas under the curve values are 0.935 for β-D-glucan, 0.730 for KL-6, and 0.704 for LDH. There were significant differences in the area under the curve value between for β-D-glucan and for the other two markers ($p < 0.01$).

We evaluated the correlation between the markers in sera that were sampled from the patients with PCP. There was a significant correlation between β-D-glucan and LDH levels in serum ($p = 0.014$, $R^2 = 0.180$; Fig 3, *top left, A*). No significant correlation was observed between the serum levels of LDH and KL-6 ($p = 0.11$). There was no significant correlation between the levels of KL-6 and β-D-glucan ($p = 0.13$).

In PCP-positive patients, we also examined the correlation between the oxygenation index and the levels of serum markers. Figure 3, *top right, B* shows a significant inverse correlation between the oxygenation index and serum LDH level ($p = 0.001$, $R^2 = 0.244$). Neither KL-6 nor β-D-glucan were correlated with the oxygenation index (data not shown). Serum LDH might reflect lung tissue injury, which resulted in a statistical correlation between serum LDH and oxygenation impairment during PCP.

The correlation between serum markers and differential cell counts in BAL fluid was evaluated as well. The proportion of neutrophils in BAL fluid did not differ between the patients with and without PCP; however, in the PCP-positive patients, the serum LDH level was significantly correlated with the neutrophil proportion in BAL fluid ($p = 0.004$, $R^2 = 0.175$; Fig 3, *center left, C*). In the patients with PCP, the proportion of neutrophils in BAL fluid was also correlated with the serum level of β-D-glucan ($p = 0.008$, $R^2 = 0.216$; Fig 3, *center right, D*). No correlation was observed between the neutrophil proportion in BAL fluid and other serum markers examined (data not shown). The neutrophil proportion in BAL fluid was significantly correlated with the oxygenation index ($p = 0.001$, $R^2 = 0.235$; Fig 3, *bottom, E*). Neither lymphocyte nor eosinophil proportions were correlated with any of the serum markers examined (data not shown).

DISCUSSION

In the present study, we evaluated the roles of the serum markers LDH, β-D-glucan, KL-6, and CRP in the diagnosis of PCP. Although there have been several reports[1,3,5,6,9,10,13] describing the levels of

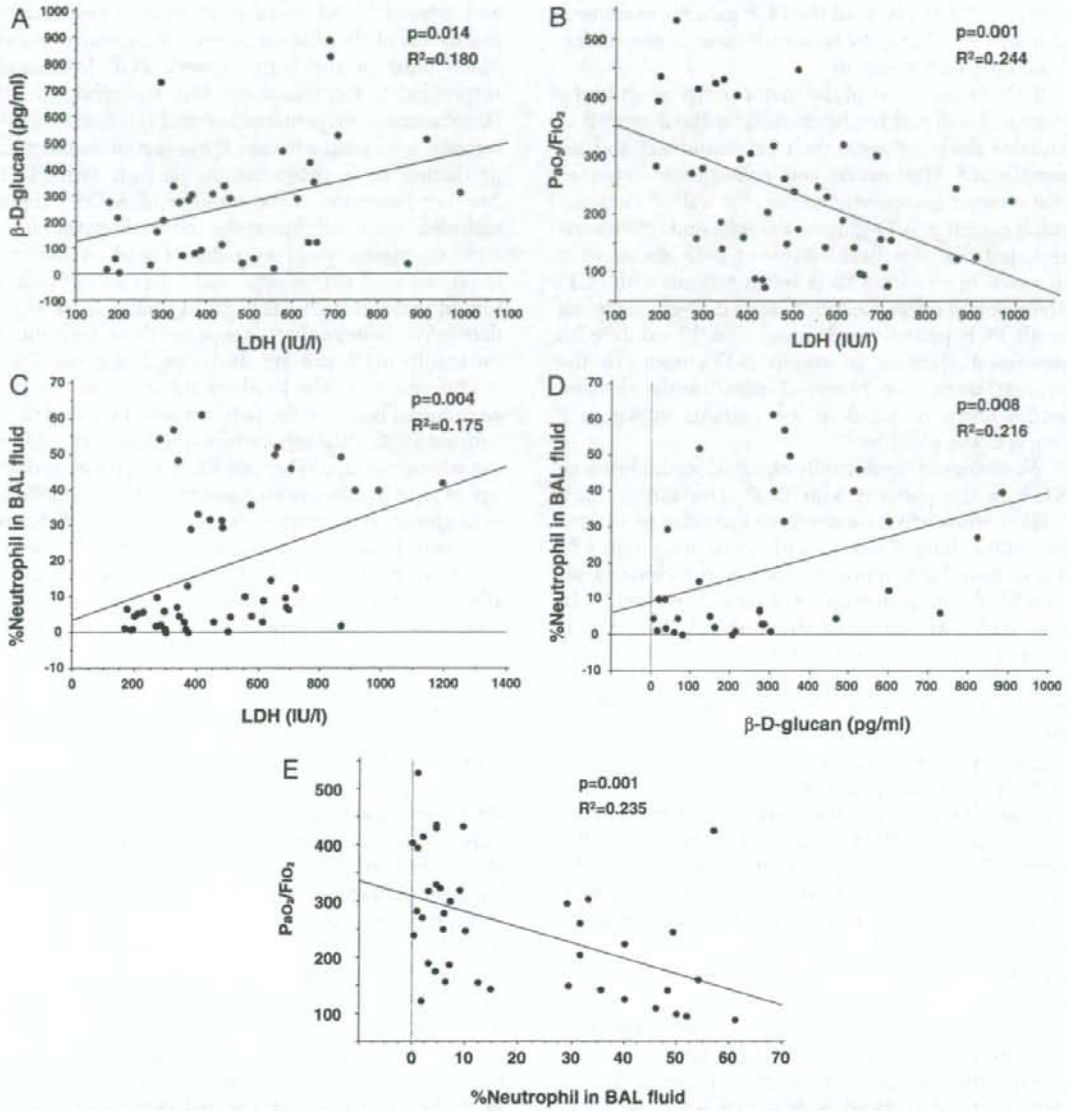

FIGURE 3. Relationships between the parameters evaluated. *Top left, A*: Serum levels of LDH and β-D-glucan are positively correlated ($p < 0.02$). *Top right, B*: Serum LDH level is inversely correlated with oxygenation index ($p < 0.01$). *Center left, C*: Serum LDH level is significantly correlated with neutrophil percentage in BAL fluid ($p < 0.01$). *Center right, D*: Serum β-D-glucan level is significantly correlated with neutrophil percentage in BAL fluid ($p < 0.01$). *Bottom, E*: The proportion of neutrophils in BAL fluid is inversely correlated with oxygenation index ($p < 0.01$).

serum markers in PCP patients, the diagnostic significance of serum markers, including the newly identified β-D-glucan and KL-6, remains to be evaluated in a larger cohort of patients with a variety of underlying diseases.

The results of this study revealed that β-D-glucan is the most reliable indicator for detecting *P jirovecii* infection, whereas only the LDH level was correlated with oxygenation impairment. Although these two serum markers correlate to each other, the findings suggested that LDH might be a good marker for organ damage rather than a specific marker for the disease. This result is compatible with the report by Quist and Hill[3] that described elevated

serum LDH levels in all the PCP patients examined, although it is likely to be a reflection of the underlying lung inflammation.

β-D-glucan is one of the major components of the yeast cell wall and has been used for the diagnosis of invasive deep mycosis, such as candidiasis and aspergillosis.[4] Matsumoto and colleagues[14] reported that a major component of the cyst wall of *Pneumocystis carinii* is β-D-glucan. Yasuoka and coworkers[5] reported that significant levels of β-D-glucan were detected in sera from six of seven patients with PCP. Shimizu and colleagues[6] evaluated β-D-glucan levels in 15 PCP patients and found that 13 exhibited a significant increase in serum β-D-glucan. In the present study, we observed significantly elevated serum levels of KL-6 in the patients in whom *P jirovecii* was identified.

We observed significantly elevated serum levels of KL-6 in the patients with PCP. The serum KL-6 level is known to be a sensitive indicator of various interstitial lung diseases and acute lung injury.[7,8] There have been reports[9,10] describing elevated serum KL-6 in a small number of patients with PCP. In this study, we observed that the KL-6 level was increased in the patients with PCP. KL-6 may not be as reliable as β-D-glucan for the diagnosis of PCP, possibly because KL-6 is too sensitive to underlying interstitial lung disease. Thirty-two of the PCP-negative patients had interstitial lung disease as an underlying disease, and 40 patients had collagen vascular disease. Since the median KL-6 level in the PCP-negative patients was also higher than the normal limit, we speculate that, in most cases, the increased KL-6 levels in the patients without PCP may have been due to an underlying lung disease, including interstitial lung disease, that was either idiopathic or associated with collagen vascular disease.

In the patients with PCP, whereas the proportion of neutrophils in BAL fluid was not increased compared with that in PCP-negative patients, it was significantly correlated with serum levels of LDH and β-D-glucan. It has been reported that *P carinii* enhances tumor necrosis factor-α release from alveolar macrophages through a β-glucan–mediated mechanism.[15] In this study, the neutrophil proportion in BAL fluid was also correlated with the oxygenation index, which is compatible with the findings of a previous report.[16] It is conceivable that β-D-glucan released from pneumocystis might cause subsequent neutrophil accumulation in the lung, leading to lung tissue damage and oxygenation impairment.

The limitation of this study is that we were unable to follow the time course of the serum markers after the diagnosis, partly because the chest radiographs and arterial blood gases were usually referred for evaluation of the disease course or treatment effect. Since most of the patients with PCP fortunately responded to the treatment with trimethoprim/sulfamethoxazole or pentamidine and recovered, it also remains uncertain whether these serum markers are predictive of a prognosis in patients with PCP. Another limitation is the number of AIDS patients included was small. Since the induced sputum has a high diagnostic yield in patients with AIDS, the importance of β-D-glucan and other serum indicators in patients with AIDS still remains to be elucidated. We believe that, except for these limitations, the results of the present study could be generalized.

Differences in the levels of three serum markers were found between the patients with PCP and those without PCP, although each marker reflects a different phenomenon. Whereas KL-6 responds to damage of type 2 pneumocytes and bronchial epithelium, β-D-glucan is a marker for a deep-seated fungal infection. In addition, an elevated serum LDH level, which we observed in PCP-positive patients, generally indicates damage in various organs. Although serum β-D-glucan appears to be the most reliable marker for *P jirovecii* infection, combined measurement of these three parameters may enable the rapid and noninvasive diagnosis of PCP, considering that most PCP patients have a variety of underlying diseases and simultaneous infections.

In the present study, GMS and Calcofluor white stain were used for *P jirovecii* detection, whereas polymerase chain reaction (PCR) assays have been shown to have greater sensitivity and specificity. It has been reported that PCR of BAL fluid can detect asymptomatic colonization of *P jiroveci* particularly in patients receiving corticosteroid therapy or immunocompetent patients with lung disease.[17,18] Although PCR assay may have an advantage in diagnostic sensitivity, we were concerned that positive PCR findings may indicate asymptomatic colonization of *P jirovecii* rather than clinically significant PCP, because most of the patients we evaluated were immunocompromised or had lung disease. Both GMS and Calcofluor white stain have been reported to have PPV and NPV > 90%.[19] Direct fluorescence monoclonal antibody stain was another sensitive method, although it has been shown that Calcofluor white stain has equal sensitivity in the detection of *P jirovecii* especially in BAL specimens.[11] We concluded that the microscopic method used to detect *P jirovecii* should have sufficient sensitivity and specificity.

In conclusion, after analyzing the diagnostic significance of serum markers in a larger group of patients, we found that β-D-glucan is a reliable marker for the diagnosis of PCP. Serum LDH, which

may not be as dependable as β-D-glucan in terms of the diagnosis of PCP, reflects lung tissue injury, resulting in its statistical correlation with oxygenation impairment during PCP. Since the BAL procedure is invasive, especially for patients with severe respiratory failure, measuring β-D-glucan should be considered as a primary modality for a noninvasive and accurate diagnosis of PCP.

ACKNOWLEDGMENT: We thank Dr. Gregory A. Plotnikoff for review of the manuscript. We also thank Dr. Satoru Fukinbara for contributing to the statistical analysis.

REFERENCES

1 Thomas CF Jr, Limper AH. Pneumocystis pneumonia. N Engl J Med 2004; 350:2487–2498
2 Klech H, Hutter C. Side-effects and safety of BAL. Eur Respir J 1990; 3:939–940, 961–969
3 Quist J, Hill AR. Serum lactate dehydrogenase in *Pneumocystis carinii* pneumonia, tuberculosis, and bacterial pneumonia. Chest 1995; 108:415–418
4 Obayashi T, Yoshida M, Mori T, et al. Plasma (1–>3)-β-D-glucan measurement in diagnosis of invasive deep mycosis and fungal febrile episodes. Lancet 1995; 345:17–20
5 Yasuoka A, Tachikawa N, Shimada K, et al. (1→3) β-D-glucan as a quantitative serological marker for *Pneumocystis carinii* pneumonia. Clin Diagn Lab Immunol 1996; 3:197–199
6 Shimizu A, Oka H, Matsuda T, et al. (1→3) β-D glucan is a diagnostic and negative prognostic marker for *Pneumocystis carinii* pneumonia in patients with connective tissue disease. Clin Exp Rheumatol 2005; 23:678–680
7 Kohno N, Kyoizumi S, Awaya Y, et al. New serum indicator of interstitial pneumonitis activity: sialylated carbohydrate antigen KL-6. Chest 1989; 96:68–73
8 Ishizaka A, Matsuda T, Albertine KH, et al. Elevation of KL-6, a lung epithelial cell marker, in plasma and epithelial lining fluid in the acute respiratory distress syndrome. Am J Physiol Lung Cell Mol Physiol 2004; 286:L1088–L1094
9 Hamada H, Kohno N, Yokoyama A, et al. KL-6 as a serologic indicator of *Pneumocystis carinii* pneumonia in immunocompromised hosts. Intern Med 1998; 37:307–310
10 Tanaka M, Tanaka K, Fukahori S, et al. Elevation of serum KL-6 levels in patients with hematological malignancies associated with cytomegalovirus or *Pneumocystis carinii* pneumonia. Hematology 2002; 7:105–108
11 Aslanzadeh J, Stelmach PS. Detection of *Pneumocystis carinii* with direct fluorescence antibody and Calcofluor white stain. Infection 1996; 24:248–250
12 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837–845
13 Skelly M, Hoffman J, Fabbri M, et al. S-adenosylmethionine concentrations in diagnosis of *Pneumocystis carinii* pneumonia. Lancet 2003; 361:1267–1268
14 Matsumoto Y, Matsuda S, Tegoshi T. Yeast glucan in the cyst wall of *Pneumocystis carinii*. J Protozool 1989; 36:21S–22S
15 Hoffman OA, Standing JE, Limper AH. *Pneumocystis carinii* stimulates tumor necrosis factor-α release from alveolar macrophages through a β-glucan–mediated mechanism. J Immunol 1993; 150:3932–3940
16 Smith RL, El-Sadr WM, Lewis ML. Correlation of bronchoalveolar lavage cell populations with clinical severity of *Pneumocystis carinii* pneumonia. Chest 1988; 93:60–64
17 Maskell NA, Waine DJ, Lindley A, et al. Asymptomatic carriage of *Pneumocystis jiroveci* in subjects undergoing bronchoscopy: a prospective study. Thorax 2003; 58:594–597
18 Sing A, Roggenkamp A, Autenrieth IB, et al. *Pneumocystis carinii* carriage in immunocompetent patients with primary pulmonary disorders as detected by single or nested PCR. J Clin Microbiol 1999; 37:3409–3410
19 Procop GW, Haddad S, Quinn J, et al. Detection of *Pneumocystis jiroveci* in respiratory specimens by four staining methods. J Clin Microbiol 2004; 42:3333–3335

Journal of Zoo and Wildlife Medicine 38(3): 471–474, 2007
Copyright 2007 by American Association of Zoo Veterinarians

PROFILES OF A PATHOGENIC *ENTAMOEBA HISTOLYTICA*–LIKE VARIANT WITH VARIATIONS IN THE NUCLEOTIDE SEQUENCE OF THE SMALL SUBUNIT RIBOSOMAL RNA ISOLATED FROM A PRIMATE (DE BRAZZA'S GUENON)

Jun Suzuki, Seiki Kobayashi, Ph.D., Rie Murata, Yoshitoki Yanagawa, D.V.M., Ph.D., and Tsutomu Takeuchi, M.D., Ph.D.

Abstract: A pathogenic *Entamoeba histolytica*–like variant (JSK2004) with genetic variations and a novel isoenzyme pattern was isolated from a De Brazza's guenon. A homology of 99.1% was found between the clones of *E. histolytica* (HM-1:IMSS) and JSK2004 in the 1,893 nucleotide bases of the small subunit rRNA (SSU-rRNA) gene. The DNA of the pathogenic amoeba species was also extracted from two sterile liver abscesses during the autopsies of an Abyssinian colobus and a Geoffroy's spider monkey occurring in the same institution in which JSK2004 was isolated, and the homology of the nucleotide sequences in the SSU-rRNA gene of the DNAs was identical to that of JSK2004.

Key words: *Entamoeba histolytica*–like variant, pathogenic isolate, primate, SSU-rRNA gene, De Brazza's guenon.

BRIEF COMMUNICATION

Entamoeba histolytica (pathogenic) and *Entamoeba dispar* (nonpathogenic) are parasitic amoebic species in humans and nonhuman primates, and they show significant genetic similarity.[2,12] Since it is difficult to morphologically differentiate the latter from the former, the detection of species-specific hexokinase (HK) and phosphoglucomutase (PGM) isoenzyme bands[9]; the detection of proteins by monoclonal antibodies[6]; and the detection of DNAs by a polymerase chain reaction (PCR)[4,10] have been used for identification. Previously, a pathogenic *E. histolytica* variant was detected with isoenzyme bands characteristic of both *E. histolytica* PGM and *E. dispar* HK. This exceptional isoenzyme pattern [zymodeme XIII (Z-XIII)] was detected in human cases in South Africa and Tanzania; however, the genetic profile of this variant has not been described.[9] In this study, we isolated a pathogenic *E. histolytica*–like strain (JSK2004) from a De Brazza's guenon (*Cercopithecus neglectus*). This strain did not satisfy the criteria for zymodeme classification; however, the isoenzyme bands were characteristic of both *E. histolytica* HK and *E. dispar* PGM that were inverse patterns of Z-XIII.

Prior to the isolation of JSK2004, the DNA of the pathogenic amoeba species was extracted from two sterile liver abscesses during the autopsies of an Abyssinian colobus (*Colobus guereza*) and a Geoffroy's spider monkey (*Ateles geoffroyi*). DNA was also extracted from the feces, including cysts, of a De Brazza's guenon without distinct symptoms and was identified as that of *E. histolytica* by using the *E. histolytica II* kit (TechLab, Blacksburg, Virginia 24060, USA); JSK2004 was isolated from the same individual. The above-mentioned three primate species were born and bred for several generations in Japan. The infection source could not be definitively identified. The specimens were subjected to PCR and multiplex PCR by using two primer sets targeting the 30-kDa proteins[10] and the small subunit rRNA (SSU-rRNA) genes[4] of *E. histolytica* and *E. dispar*, respectively. The expected 101–base pair (bp) fragments of the *E. histolytica* gene were produced by PCR. However, no fragment was produced by the multiplex PCR (data not shown).

A JSK2004 axenic culture was established in TYI-S-33 medium[1]; subsequently, four clones (JSK2004 cl1 to cl4) were obtained by the classical methods of Diamond.[1] Each clone was confirmed to possess the same genetic polymorphism[5] and zymodeme[9] profiles. We sequenced 1,893 bases of the SSU-rRNA gene of one clone (JSK2004 cl2). The gene was PCR-amplified using the *Entamoeba* species-specific primer set [Entam1 (forward: 5′-GTT GAT CCT GCC AGT ATT ATA TG-3′) and Entam2 (reverse: 5′-CAC TAT TGG AGC TGG AAT TAC-3′)][11] and two primer sets [Ent2F (forward: 5′-GTA ATT CCA GCT CCA ATA GTG-3′) and Ent2R (reverse: 5′-ACA CCA CTT ACT ATC CTT AAT-3′), Ent3F (forward: 5′-GTT ATC TAA

From the Division of Clinical Microbiology, Department of Microbiology, Tokyo Metropolitan Institute of Public Health, 3-24-1, Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan (Suzuki, Murata, Yanagawa); and the Department of Tropical Medicine and Parasitology, School of Medicine, Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (Kobayashi, Takeuchi). Correspondence should be directed to Dr. Seiki Kobayashi (skobaya@sc.itc.keio.ac.jp).

TTT CGG TTA GAC-3′), and Ent3R (reverse: 5′-ATC CTT CCG CAG GTT CAC CTA-3′)] that were designed from the SSU-rRNA gene sequence of an *E. histolytica* reference strain [HM-1:IMSS cl6 (HM-1)] (GenBank accession no. X64142). The amplification was performed in a reaction mixture (50 μl) containing 100 ng DNA, 1.0 U LA*Taq*DNA polymerase (Takara Bio, Inc., Shiga, Japan), 0.4 μM of each primer, and 0.25 mM dNTPs. The following cycling parameters were used: 1) *Taq* activation at 94°C for 3 min; 2) 35 cycles of denaturation at 94°C for 40 sec, annealing at 51°C for 40 sec, and extension at 72°C for 1 min; and 3) extension at 72°C for 5 min. The PCR products were sequenced using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, California 94404, USA) on an ABI PRISM 310 Genetic Analyzer.

A comparison of the SSU-rRNA sequences of JSK2004 cl2, HM-1, and SAW1734R clAR (SAW1734; an *E. dispar* reference strain; GenBank accession no. Z49256) indicated that 14 nucleotide bases of the SSU-rRNA sequence of JSK2004 cl2 differed from those of HM-1, and of these 14 bases, seven were similar to those of SAW1734 (Fig. 1). Similarly, the SSU-rRNA gene sequences of JSK2004 cl1, cl3, and cl4 corresponded to that of JSK2004 cl2. The SSU-rRNA sequences of the three other *E. histolytica* axenic strains with different zymodemes [Z-IIα-(SAW1627), Z-XIV (SAW755R clB), and Z-XIX (KU-2)][5] and two *E. dispar* axenic strains (CYNO 09:TPC and AS16IR) (Z-I)[8] corresponded to those of HM-1 (Z-II) and SAW1734 (Z-I), respectively.

The total DNA from the liver abscess of the Abyssinian colobus and Geoffroy's spider monkey were PCR-amplified using two primer sets, (Ent2F and Ent2R) and [Ent4F (forward: 5′-ATG GCC AAT TTA TTT AAA TGA-3′) and Ent4R (reverse: 5′-CAG ATT AAG AAA CAA TGC TTC-3′), designed from the SSU-rRNA gene sequence of JSK2004 cl2], for identifying the *E. histolytica*–like variant (JSK2004). The PCR was performed as follows: 5 min at 94°C; 35 cycles of 40 sec at 94°C, 40 sec at 55°C, 1 min at 72°C; and a final step of 5 min at 72°C. The PCR products from the total DNA from the liver abscesses of an Abyssinian colobus and Geoffroy's spider monkey were obtained
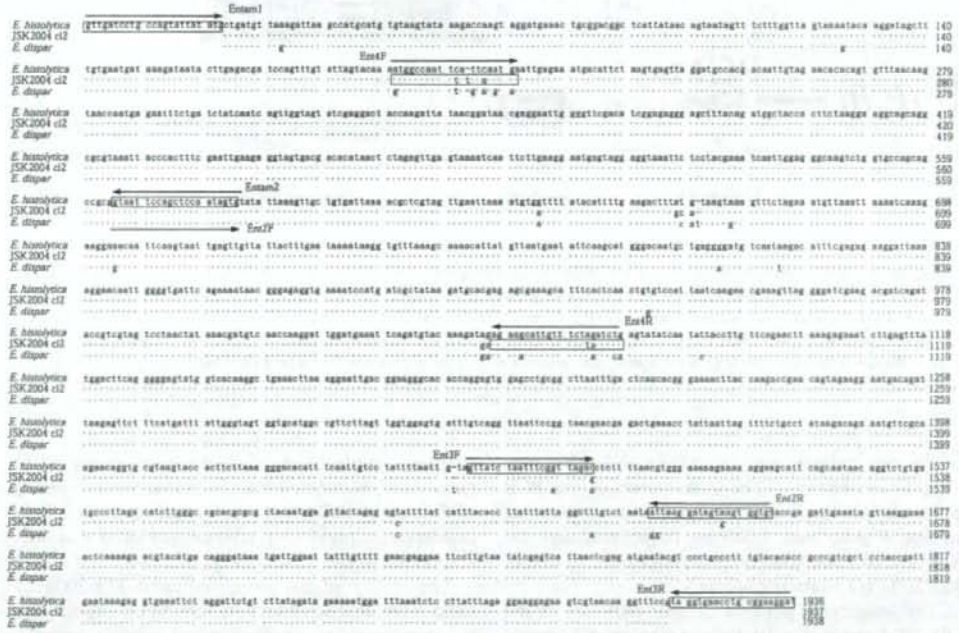
Figure 1. Profile of the nucleotide sequence of the small subunit rRNA of the *Entamoeba histolytica*–like variant (JSK2004 cl2) from a De Brazza's guenon. *E. histolytica*: *Entamoeba histolytica* (HM-1:IMSS cl6); *E. dispar*: *Entamoeba dispar* (SAW1734R clAR).

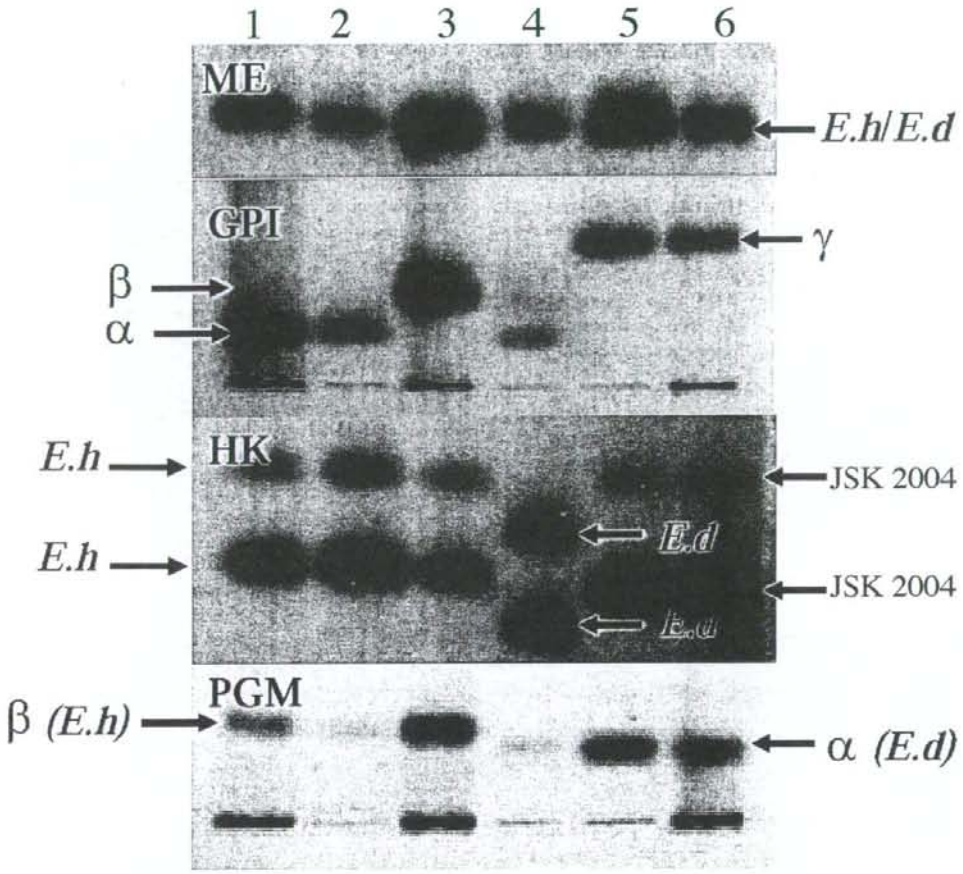

Figure 2. Isoenzyme patterns of four enzymes in the freeze-thawed lysate of the *Entamoeba histolytica*–like variant (JSK2004) where ME: malic enzyme; GPI: glucose phosphate isomerase; HK: hexokinase; and PGM: phosphoglucomutase. Lane 1: a clinical isolate (xenic) (Z-II); Lane 2: HM-1:IMSS cl6 (axenic) (Z-II); Lane 3: SAW1453 (xenic) (Z-XIV); Lane 4: SAW1734R clAR (axenic) (Z-I); Lane 5: JSK2004 (axenic); Lane 6: JSK2004 (xenic); Lanes 1, 2, and 3: *Entamoeba histolytica* (*E.h*); Lane 4: *Entamoeba dispar* (*E.d*); Lanes 5 and 6: JSK2004; α, β, and γ: α, β, and γ bands.

using two primer sets [(Ent2F and Ent2R) and (Ent4F and Ent4R)]. The authenticity of the PCR products was confirmed by nucleotide sequencing of the 878-bp and 1,088-bp fragments, and the SSU-rRNA gene sequences corresponded to that of JSK2004 cl2 (data not shown).

The isoenzyme patterns of four enzymes of the JSK2004 strain and its four clones were identical and novel and showed the following bands: 1) malic enzyme (ME), band with the same mobility as those of *E. histolytica* and *E. dispar*; 2) HK, fast-running double bands with the same mobility as those of *E. histolytica* HK; 3) PGM, band (α band) of *E. dispar*–PGM type; and 4) glucose phosphate isomerase, a previously unidentified band that corresponded to the γ band (Fig. 2).

The ability of the axenic-cultured JSK2004 to form liver abscesses and its infectious capacity in the large intestine were examined by inoculating 1 × 10^6 amoebae/head into the left hepatic lobes of female Syrian hamsters (3–4 wk old)[3] and the cecums of female C3H/HeJ mice (5–6 wk old).[7] Liver

abscess formation was confirmed in all of the three examined hamsters; JSK2004 established persistent infection in three mice for more than 6 mo. The homologies of 1,893 nucleotide bases of the SSU-RNA gene between JSK2004 cl2 and *E. histolytica* and *E. dispar* were 99.10% and 98.47%, respectively. Hence, the genetic homology between *E. histolytica* and *E. dispar* was 98.10%; JSK2004 cl2 is located at a position that is intermediate between the two species. The pathogenic *E. histolytica*–like variant (JSK2004) that is located at an intermediate position between *E. histolytica* and *E. dispar* and contains an *E. dispar*–PGM type isoenzyme and variations in the SSU-rRNA gene is expected to be a useful reference for phylogenetic studies.

Acknowledgments: We are grateful to veterinarians H. Tajima and F. Hashizaki for providing the samples. A part of this work was supported by a Health Sciences Research Grant-in-Aid for Emerging and Reemerging Infectious Diseases.

LITERATURE CITED

1. Diamond, L. S. 1983. Lumen-dwelling protozoa: *Entamoeba*, trichomonads and giardia. *In:* Jensen, J. B. (ed.). In Vitro Cultivation of Protozoan Parasites. CRC Press, Boca Raton, Florida. Pp. 65–109.

2. Diamond, L. S., and C. G. Clark. 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. J. Eukaryot. Microbiol. 40: 340–344.

3. Diamond, L. S., B. P. Phillips, and I. L. Bartgis. 1974. A comparison of the virulence of nine strains of axenically cultivated *E. histolytica* in hamster liver. Arch. Investig. Med. 5: 423–426.

4. Evangelopoulos, A., G. Spanakos, E. Patsoula, N. Vakalis, and N. Legakis. 2000. A nested, multiplex, PCR assay for the simultaneous detection and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* in faeces. Ann. Trop. Med. Parasitol. 94: 233–240.

5. Haghighi, A., S. Kobayashi, T. Takeuchi, G. Masuda, and T. Nozaki. 2002. Remarkable genetic polymorphism among *Entamoeba histolytica* isolates from a limited geographic area. J. Clin. Microbiol. 40: 4081–4090.

6. Haque, R., I. K. Ali, S. Akther, and W. A. Petri, Jr. 1998. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J. Clin. Microbiol. 36: 449–452.

7. Houpt, E. R., D. J. Glembocki, T. G. Obrig, C. A. Moskaluk, L. A. Lockhart, R. L. Wright, R. M. Seaner, T. R. Keepers, T. D. Wilkins, and W. A. Petri, Jr. 2002. The mouse model of amebic colitis reveals mouse strain susceptibility to infection and exacerbation of disease by CD4+ T cells. J. Immunol. 169: 4496–4503.

8. Kobayashi, S., E. Imai, A. Haghighi, S. A. Khalifa, H. Tachibana, and T. Takeuchi. 2005. Axenic cultivation of *Entamoeba dispar* in newly designed yeast extract-iron-gluconic acid-dihydroxyacetone-serum medium. J. Parasitol. 91: 1–4.

9. Sargeaunt, P. G. 1988. Zymodemes of *Entamoeba histolytica*. *In:* Radvin, J. I. (ed.). Amebiasis: Human Infection by *Entamoeba histolytica*. John Wiley and Sons, Inc., New York, New York. Pp. 370–387.

10. Tachibana, H., S. Kobayashi, M. Takekoshi, and S. Ihara. 1991. Distinguishing pathogenic isolates of *Entamoeba histolytica* by polymerase chain reaction. J. Infect. Dis. 164: 825–826.

11. Verweij, J. J., D. Laeijendecker, E. A. Brienen, L. van Lieshout, and A. M. Polderman. 2003. Detection and identification of *Entamoeba* species in stool samples by a reverse line hybridization assay. J. Clin. Microbiol. 41: 5041–5045.

12. Verweij, J. J., J. Vermeer, E. A. Brienen, C. Blotkamp, D. Laeijendecker, L. van Lieshout, and A. M. Polderman. 2003. *Entamoeba histolytica* infections in captive primates. Parasitol. Res. 90: 100–103.

Received for publication 2 January 2007

An Evolutionarily Conserved Mechanism for MicroRNA-223 Expression Revealed by MicroRNA Gene Profiling

Taro Fukao,[1,*] Yoko Fukuda,[2] Kotaro Kiga,[1] Jafar Sharif,[1] Kimihiro Hino,[1] Yutaka Enomoto,[1] Aya Kawamura,[1] Kaito Nakamura,[1] Tsutomu Takeuchi,[3] and Masanobu Tanabe[3]
[1] Department of Chemistry and Biotechnology, Graduate School of Engineering
[2] Department of Neurology, Graduate School of Medicine
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[3] Department of Tropical Medicine and Parasitology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
*Correspondence: tarofukao@2002.jukuin.keio.ac.jp
DOI 10.1016/j.cell.2007.02.048

SUMMARY

Many microRNAs (miRNAs) are evolutionarily conserved and have intriguing expression patterns. Tissue and/or time-specific expressions of some miRNAs are presumably controlled by unique *cis*-acting regulatory elements that coevolved with the miRNA sequences. Exploiting bioinformatics, we identified several miRNAs whose primary transcripts could be regulated by conserved genomic elements proximal to their transcription start sites. Such miRNAs include microRNA-223 (miR-223), which is reportedly controlled by a unique regulatory mechanism during granulopoiesis. Here, we define a mechanism distinct from that previously proposed to regulate miR-223 expression. We find that the *mir-223* gene resembles a "myeloid gene" and might be driven by the myeloid transcription factors, PU.1 and C/EBPs. This mechanism is specified by the conserved proximal *cis*-regulatory element and might be common among different species. Hence, it needs to be considered that two distinct mechanisms that would play critical roles in myeloid functions and differentiation are actually concerned with the regulation of miR-223.

INTRODUCTION

The miRNAs are a growing class of tiny noncoding RNAs (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001) that regulate the expression of genes by hybridizing the target sites with complementary sequences that are followed by translational repression, mRNA cleavage, or destabilization (Bartel, 2004; Yekta et al., 2004). Hundreds of miRNAs have been found in various organisms, and many miRNAs are evolutionarily conserved, suggesting their potential roles in essential physiological events (Bartel, 2004). Some miRNAs have limited expression patterns with strict tissue, cell, and temporal specificities, while others demonstrate ubiquitous or constitutive expression (Pasquinelli et al., 2005). Therefore, studies of regulatory mechanisms governing expression of specific miRNAs would contribute to better understanding of their physiological roles.

Despite growing knowledge on miRNA biology, little is known about the transcriptional regulation of miRNAs. However, cumulative evidence suggests that many miRNAs are transcribed by RNA polymerase II (Lee et al., 2004). This is consistent with spatiotemporally limited expression of miRNAs and indicates that numerous lineage-specific transcription factors would take part in transcriptional regulation of miRNAs. It is now widely accepted that most miRNAs are transcribed from their own promoters, even though a few primary transcripts have been fully defined to date. These primary miRNA transcripts, called pri-miRNAs, are generally thought to be much longer than the conserved stem loops currently used to define miRNA genes, as suggested by several observations. For example, clustered miRNAs are generated from a single primary transcript (Lagos-Quintana et al. 2001; Lau et al. 2001). Furthermore, some RT-PCR experiments amplified large fragments of the pri-miRNAs (Lee et al. 2002). Identification of full-length primary transcripts with regulatory genomic elements would therefore be a critical step for studying the transcriptional mechanisms for every single miRNA.

The hematopoietic system is a paradigm for the differentiation of distinct cell lineages from multipotent progenitors (Orkin and Zon, 2002). Although lineage commitment decisions in hematopoiesis involve a variety of modulators that are influenced by each other in an unusually complex manner, the central role of transcription factors in this bioprocess has been highlighted to date (Sieweke and Graf,

1998; Zhu and Emerson, 2002). The current concept favors the emerging hematopoietic lineages being controlled by a unique combination of transcription factors, each of which may appear singly in different lineages (Sieweke and Graf, 1998), rather than the lineages being regulated by single master transcription factors. Thus, diverse combinations of transcription factors specify the fate of distinct cell types by selection and maintenance of a lineage-specific gene-expression program.

Previous studies demonstrated that some miRNAs, such as miR-142, miR-181, and miR-223, are preferentially expressed in the hematopoietic system (Chen et al., 2004; Chen and Lodish, 2005). For example, miR-181 is highly expressed in lymphoid tissues and cells (Chen et al., 2004). Ectopic expression of miR-181 in murine hematopoietic progenitors resulted in significant increase of B lineage fraction, underscoring possible miRNA involvement in lymphopoiesis (Chen et al., 2004). Furthermore, a very recent study demonstrated the mechanism that specifies myeloid expression of miR-223 and proposed a unique "minicircuitry" comprised of miR-223 and transcription factors, NFI-A and CAAT enhancer-binding protein α (C/EBPα; Fazi et al., 2005). As shown previously, miR-223 is highly conserved, and its myeloid-specific expression is also well characterized in both human and mouse (Chen et al., 2004; Fazi et al., 2005), suggesting that the regulatory system is expected to be phylogenetically conserved.

Here, we demonstrate a comprehensive strategy for understanding miRNA gene expression and make use of transcriptome data, an integrated database of genome sequences or transcription factors, and bioinformatics; in addition, we follow experimental validation. With this approach, we identified dozens of cDNAs encoding miRNA precursor sequences. These cDNAs are actually considered as pri-miRNAs, and some of them are characterized with genomic information that implicates for their regulatory mechanisms.

As a result of this screening, we defined a regulatory mechanism for myeloid expression of miR-223, which is distinct from the previously proposed machinery but would be common among different species. We thus found that miR-223 originates in a relatively long noncoding RNA (pri-miR-223) driven by the conserved proximal genomic element in which possible binding sites for myeloid transcription factors, PU.1 and C/EBP, are present. Our analyses revealed an essential role of PU.1 in transcription of pri-miR-223. Moreover, C/EBP was shown to strongly enhance the promoter activity, especially in combination with PU.1. Therefore, the conserved proximal *cis*-acting regulatory element of pri-miR-223 shows general myeloid promoter characteristics, consistent with preferential expression of miR-223 in the myeloid lineage. This evolutionarily conserved promoter is distinct from the previously characterized promoter (Fazi et al., 2005). Importantly, our reassessment revealed that this conserved promoter is probably active in the induction of miR-223 during All-*trans* retinoic acid (ATRA)-induced differentiation of the APL cell line, NB4 cells, which is the main experimental system adopted in the previous study (Fazi et al., 2005). As a result of our study on miR-223 regulation and our reassessment of some parts in the previous report, it needs to be considered how two distinct mechanisms regulate the expression of miR-223.

RESULTS

Genome-Wide Screening and Characterization of Primary miRNA Transcripts

It is now widely accepted that many miRNAs are transcribed by RNA polymerase II (Lee et al., 2004; Kim, 2005). Therefore, we hypothesized that primary transcripts of some miRNAs could appear as mRNA-like RNAs. Recent fruits of FANTOM3, the project of mouse transcriptome analyses, published approximately 103,000 nonredundant cDNA sequences from various tissues (Carninci et al., 2005). Based upon the above hypothesis, we made use of this library for genome-wide screening of mRNA-like primary miRNAs. As an initial approach, we searched for cDNAs encoding full-length miRNA precursor sequences and reasoned that such cDNAs represent primary transcripts of "exonic" miRNAs (Kim and Nam, 2006).

Based on the homology, 102,802 cDNAs of the FANTOM3 database were exhaustively aligned with the 337 registered murine miRNA precursor sequences by BLASTN program (Figure 1A). We identified 42 pairings of cDNA-pre-miRNAs with a perfect match (Table 1). The majority of these miRNA precursors is present on paired cDNAs with +/+ direction, while a few miRNA precursors are on paired cDNAs with opposite direction (+/−; Table 1). Although these antisense transcripts could be involved in important gene-regulatory events as reported recently (Katayama et al., 2005), we excluded them from the candidates for this round of screening.

As is well known, 5′ genomic regions proximal to transcription start sites often act as critical *cis*-regulatory elements for gene transcription (Lipman, 1997; Venkatesh and Yap, 2004). Moreover, such proximal *cis*-acting regulatory elements are often highly conserved among different species. To identify 5′ proximal *cis*-regulatory elements for each pri-miRNA candidate, we then extracted upstream genomic regions of 38 cDNAs paired with pre-miRNAs with a +/+ direction (Figure 1A). A BLAT search of a 250 bp upstream sequence of each cDNA screened out 12 pri-miRNA candidates whose 5′ proximal genomic regions are highly conserved among mouse, rat, dog, and human (Table 2). Putative binding motifs for transcription factors of these 12 regions were then explored by conducting the TESS program (Table 2).

Our screening strategy here demonstrated that available cDNA libraries might be sources of information on mRNA-like primary transcripts of miRNAs. Furthermore, subsequent genome-based profiling of pri-miRNA indicated that some miRNA sequences might have coevolved with genomic elements residing close to their transcription start sites. Such genomic elements contain several consensus

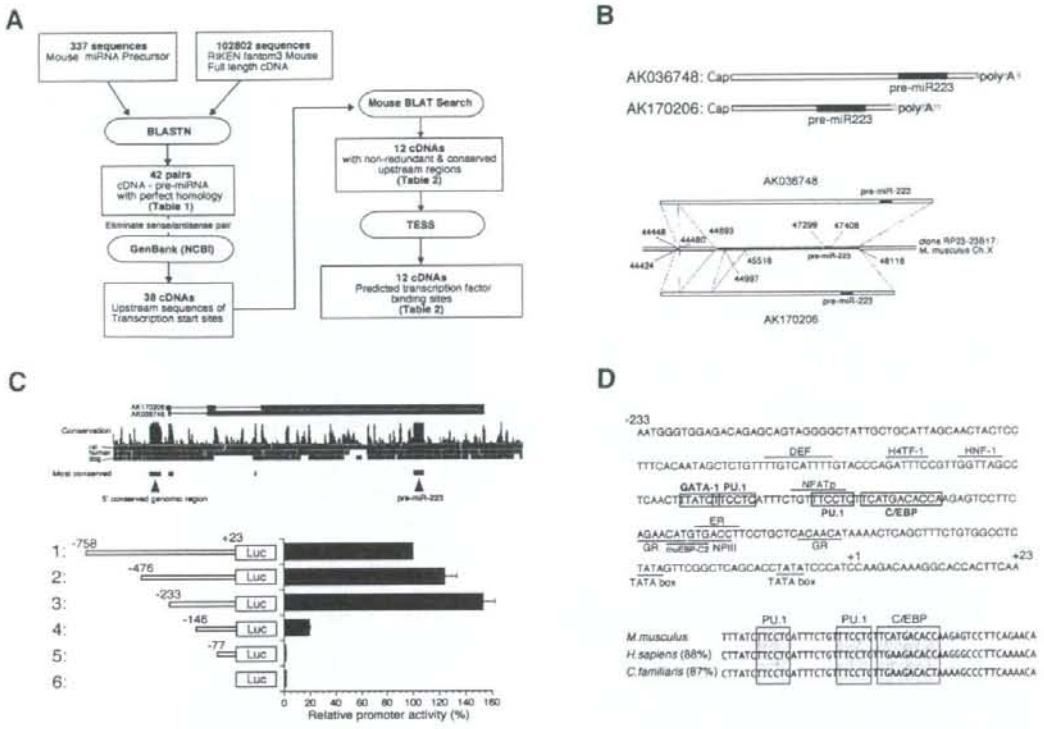

Figure 1. Mouse Pri-miR-223 Identified by Genome-Wide Screening of Pri-miRNAs

(A) Schematic for identification of cDNAs encoding known pre-miRNA sequences with conserved upstream genomic regions in FANTOM3.

(B) Two cDNAs, AK036748 and AK170206, encoding murine miR-223 precursor sequences (upper panel) and their gene structures (lower). The numbers denote the corresponding position at the *Mus musculus* genomic clone, RP23-23B17.

(C) Conservation of pri-miR-223 5′ proximal genomic region (upper). A luciferase reporter assay was performed in RAW264.7 cells with indicated reporters (lower). 1 indicates ppri-mmiR-223^{-758}-Luc, 2 indicates ppri-mmiR-223^{-476}-Luc, 3 indicates ppri-mmiR-223^{-233}-Luc, 4 indicates ppri-mmiR-223^{-146}-Luc, 5 indicates ppri-mmiR-223^{-77}-Luc, and 6 indicates pGL3 basic.

(D) Upper panel: The 5′ proximal promoter region of mouse pri-miR-223 with putative binding motifs for several transcription factors. The transcription start site of AK036748 is fixed as +1. Lower panel: Highly conserved genomic site that contains PU.1- and C/EBP-binding motifs.

motifs for intriguing transcription factors (Table 2), suggesting that these conserved genomic regions would act as *cis*-acting regulatory elements for pri-miRNAs and that miRNA expression might be regulated by such transcription factors.

Identification of Pri-miR-223 Candidates in FANTOM3

Of particular interest, AK170206 screened out as pri-miR-223 is derived from the cDNA library of CD11c+ dendritic cells (Table 1), consistent with previous studies reporting myeloid-specific expression of miR-223 (Chen et al., 2004). Furthermore, AK170206 could be regulated by myeloid transcription factors, PU.1 and C/EBP, as the conserved upstream genomic region contains binding motifs for these transcription factors (Table 2). Thus, this case may represent consistency between the tissue specificity and potential regulatory transcription factors.

Although another cDNA clone, AK036748, is also a candidate of pri-miR-223 (Table 1), we found that this clone is practically transcribed from the same gene locus for AK170206 (Figure 1B). Of note, the AK036748 clone is actually expressed in vivo with strict myeloid specificity, and ectopic expression of its splicing variant resulted in the generation of functional mature miR-223, indicating that AK036748 and AK170206 are probably mouse pri-miR-223 in vivo (Figures S1 and S2).

Since the conserved genomic region found by our genome profiling was supposed to be included in the proximal promoter of pri-miR-223, the 5′ flanking genomic region was amplified from mouse genome, which was followed by progressive deletions, and a set of reporter vectors was generated as shown in Figure 1C. Relative reporter activity in miR-223-expressing monocytic cell line, RAW264.7, was examined for each reporter vector using a luciferase reporter assay (Figure 1C). Transcriptional

Table 1. Murine cDNA Clones Encoding Known miRNA Precursors

miRNAs	Length[a]	Strand[b]	Clone	Location[c]	Tissue Origin
mmu-let-7d	103	+/+	AK160312	1006–1108	Wolffian duct includes surrounding region
mmu-let-7i	85	+/+	AK052706	109–193	kidney
mmu-mir-100	80	+/+	AK045205	1522–1601	parthenogenote
mmu-mir-101a	83	+/+	AK021368	207–289	eyeball
mmu-mir-124a-1	85	+/+	AK044422	736–820	retina
mmu-mir-130a	64	+/+	AK033350	467–530	testis
mmu-mir-133b	119	+/+	AK132542	3750–3868	head (brain, eyeball, and pituitary gland)
mmu-mir-144	66	+/+	AK158085	357–422	inner ear
mmu-mir-191	74	+/+	AK076537	788–861	head (brain, eyeball, and pituitary gland)
mmu-mir-196b	85	+/+	AK054058	368–452	oviduct
mmu-mir-202	72	+/+	AK144366	157–228	testis
mmu-mir-205	68	+/+	AK014513	1438–1505	skin
mmu-mir-21	92	+/+	AK043091	2481–2572	cerebellum
mmu-mir-214	110	+/+	AK051872	1754–1863	eyeball
mmu-mir-22	95	+/+	AK008813	133–227	stomach
mmu-mir-22	95	+/+	AK018628	242–336	cecum
mmu-mir-22	95	+/+	AK021103	151–245	corpus striatum
mmu-mir-223	110	+/+	AK036748	2437–2546	bone (os femoris)
mmu-mir-223	110	+/+	AK170206	1937–2046	NOD-derived CD11c +ve dendritic cells
mmu-mir-331	96	+/+	AK031192	1073–1168	forelimb
mmu-mir-339	96	+/+	AK019591	2509–2604	testis
mmu-mir-377	68	+/+	AK077315	913–980	pituitary gland
mmu-mir-423	109	+/+	AK038056	266–374	thymus
mmu-mir-425	85	+/+	AK076537	1246–1330	head (brain, eyeball, and pituitary gland)
mmu-mir-433	124	+/+	AK018276	194–317	olfactory brain
mmu-mir-451	72	+/+	AK158085	522–593	inner ear
mmu-mir-546	121	+/+	AK049271	3088–3208	ES cells
mmu-mir-678	84	+/+	AK077456	1829–1912	whole body
mmu-mir-678	84	+/+	AK144412	1827–1910	spinal cord
mmu-mir-678	84	+/+	AK145829	1829–1912	liver
mmu-mir-678	84	+/+	AK146754	1890–1973	kidney
mmu-mir-678	84	+/+	AK146966	1926–2009	heart
mmu-mir-678	84	+/+	AK159186	1835–1918	osteoclast-like cell
mmu-mir-686	109	+/+	AK133839	260–368	whole body
mmu-mir-698	109	+/+	AK163187	431–539	cerebellum
mmu-mir-705	82	+/+	AK078938	53–82	cecum
mmu-mir-705	82	+/+	AK154922	4698–4779	NOD-derived CD11c +ve dendritic cells
mmu-mir-707	73	+/+	AK030571	360–432	pituitary gland
mmu-mir-29c	88	+/–	AK081202	1364–1277	corpus striatum

Table 1. *Continued*

miRNAs	Length[a]	Strand[b]	Clone	Location[c]	Tissue Origin
mmu-mir-29b-2	81	+/−	AK081202	1871–1791	corpus striatum
mmu-mir-135a-1	90	+/−	AK052709	2901–2812	kidney
mmu-mir-181b-2	89	+/−	AK082091	1112–1024	cerebellum

The method used for extraction of these pre-miRNA/cDNA pairings is described in Experimental Procedures.
[a] Length of each predicted miRNA precursor in miRBase is shown.
[b] Coding direction of miRNA: cDNA pairing is shown as miRNA strand/cDNA strand.
[c] Position of each pre-miRNA sequence is shown as corresponding nucleotide numbers of pairing cDNA sequence.

activity was readily detected in the presence of regions that had been deleted for up to −146 bp of mouse pri-miR-223, although significant reduction was observed when the region from −233 bp to −146 bp was removed (Figure 1C). Strikingly, the promoter activity was completely abolished to the basal level when the region from −146 bp to −77 bp was deleted (Figure 1C), indicating that this region contains the essential promoter domain for mouse pri-miR-223 transcription. Notably, this region is actually the highly conserved region identified in the initial genome profiling and contains two binding motifs for PU.1: one nonconventional binding site for C/EBP (Grove and Plumb, 1993) and one site for GATA-1 (Figure 1D).

Together, our data suggest that the highly conserved genomic region that contains consensus motifs for hematopoietic transcription factors is probably the core promoter of mouse pri-miR-223.

Myeloid Transcription Factors Act on the Conserved Promoter of Pri-miR-223

To examine whether the myeloid transcription factors physically interact with the conserved promoter region in vivo, we performed chromatin immunoprecipitation (ChIP). DNA fragments containing mouse pri-miR-223 promoter as well as mouse Cathepsin C promoter were immunoprecipitated from genomic DNA of WEHI-3 by α-PU.1 and α-C/EBPβ antibodies, suggesting that these myeloid transcription factors are actually interacting with the corresponding promoter sites (Figure 2A).

The action of those transcription factors on the promoter was then examined. Single expression of PU.1 weakly activated mouse pri-miR-223 promoter in miR-223-negative NIH3T3 (Figure 2B). C/EBPβ alone stimulated mouse pri-miR-223 promoter activity, but such activation was weakened in a dose-dependent manner (Figure 2B). However, coexpression of PU.1 and C/EBPβ induced potent promoter activity, and C/EBPβ enhanced it in a dose-dependent manner (Figure 2C), indicating that the combinatorial action of PU.1 and C/EBP confers stable and high-level activity of mouse pri-miR-223 promoter.

We also examined action of GATA-1 on mouse pri-miR-223 promoter and observed a repression of the endogenous activity of pri-miR-223 promoter by this transcription factor (Figure 2D). Notably, however, such repression was observed even in the absence of the consensus-binding motif (data not shown), implying that GATA-1-mediated suppression of mouse pri-miR-223 promoter activity would be due to mechanisms without its DNA binding, as previously reported elsewhere (Nerlov et al., 2000).

Together these data suggest that the activity of mouse pri-miR-223 promoter is highly induced in the presence of myeloid transcription factors, PU.1 and C/EBP, whereas the erythroid transcription factor GATA-1 represses it, which is consistent with myeloid-specific expression of miR-223.

Critical Roles of Myeloid Transcription Factors in Mouse Pri-miR-223 Transcription

To test whether the putative binding sites for myeloid transcription factors in pri-miR-223 promoter are essential for its transcriptional activity, we introduced various mutations into those binding motifs (Figure 3A) and examined effects of those mutations on the transcriptional activity using a luciferase reporter assay in the RAW264.7 cell line. Although the promoter activity declined significantly with every single mutation (mutations 1, 2, and 3; Figure 3A), the mutation in the downstream PU.1-binding motif (mutation 2) resulted in a drastic decrease of transcriptional activity (Figure 3A), suggesting a dominant role of the corresponding site in transcriptional control of mouse pri-miR-223. Nonetheless, the most striking effect was observed when mutations were introduced in both of the two binding sites for PU.1 (mutation 4) and the transcriptional activity was completely abolished by this combination of mutations (Figure 3A), indicating that these two PU.1-binding sites are essential for induction of mouse pri-miR-223 promoter activity. Mutations in all three sites (mutation 7) logically resulted in the complete extinction of promoter activity, too (Figure 3A). Of note, physical binding of PU.1 to those putative binding sites was also confirmed (Figures S3A and S3B).

These results demonstrate that transcription of mouse pri-miR-223 is optimally induced in the presence of both of two PU.1-binding sites in the conserved promoter, therefore implying a critical role of PU.1 in pri-miR-223 transcription.

To further confirm the role of PU.1 in pri-miR-223 transcription, we investigated the effect of dominant-negative PU.1 (DN-PU.1) expression on pri-miR-223 promoter

Table 2. miRNA-Containing cDNAs with Conserved Upstream Genomic Elements

cDNAs (miRNAs)	Chromosome, Strand, Position[a]	Transcription Factors and Consensus Sequences	
AK021368 (mmu-mir-101A)	4(–) 100854121–100854208	ZF5 (1[b]) [GGCGCG]	IL-6 RE-BP (1) [CTGGAA]
		C/EBPβ (1) [CTGGRAA]	Sp1 (1) [GGCGGG]
AK144366 (mmu-mir-202)	7(–) 139809438–139809493	GR (1) [TGTGCC]	RAP1 (1) [CANCCNNNCA]
		TDEF (1) [CACGTG]	CAC-binding protein (1) [GGGGTG]
		C-Myb (1) [GGTGAG]	
		NF-1/L (1) [TTGGCA]	NF-1 (-like proteins) (1) [CTTGGC]
AK018628 AK008813 (mmu-mir-22)	11(+) 75277745–75277835	Sp1 (1) [GTGGGTGTGGc (GTGGGcGTGGT) (GGGCGG)]	
		TEF-2 (2) [GGGTGTGG]	AML1 (1) [TGTGGT]
		SpRunT-1(1) [TGTGGTC]	Core-binding factor (1)[TGTGGTCA]
		C/EBPAlpha (1) [CGTTGCNNNGCAAcG]	
		AP-2 (1) [GCCCAGCA]	CP1 (1) [GATTGG]
		CACCC-binding factor (1) [GGGTGG]	AP-1 (1) [TGCGTCA]
		XPF-1 (1) [CAGCTG]	E4F1 (1) [CGTNACG]
AK170206 AK036748 (mmu-mir-223)	X(+) 92442511–92442611	DEF (1) [TTGTCAgTTT]	H4TF-1 (1) [GATTTC]
		HNF-1 (1) [GGTTAG]	BBF1 (1) [ACTTTA]
		GATA-1 (1) [TTATCT]	PU.1 (2) [TTCCTC]
		NF-ATp (1) [TGTTTCCT (TTTCCTC)]	C/EBPα (1) [TCATGACACCAt]
AK044422 (mmu-mir-124a-1)	14(+) 63542041–63542225	Sp1 (3) [GGGCAG (GGGGGAGGGG) (GGGGC{C/G}GGGC)]	
		bZIP911 (1) [gGTGACGTGTGC]	E4F1 (1) [GTGACGT]
		ATF (1) [TGACGT]	bZIP910 (1) [TGACGTG]
		bZIP911 (1) [TGACGTGT]	HiNF-C (1) [GAGGGcGGGG]
		BGP1 (1) [GGGGGGG(7-16)]	Pur-1 (1) [AGGGGGGGGGTG]
		NF-S (1) [YGTCAGC]	E47 (1) [CAGATG]
		PU.1 (1) [CTTCTC]	C-MyC (1) [TCTCTTA]
		TII (1) [AATAATCC]	Kr (1) [tTAATCCGTT]
		C-ETs-2 (1) [AAGGAA]	N-OCT-3 (1) [MATNNWAAT]
		MEF-2 (1) [TTATTTTTAA(YTWWAAAtAR)]	
		AnTp (1) [TAATTaTTAAA]	SEF4 (1) [RTTTTTR]
		Dof2 (1) [AAAAAGGAgC]	
AK018276 (mmu-mir-433)	12(+) 110039132–110039250	PU.1 (4) [TTCCTC]	GR (6) [TCTTCT]
AK049271 (mmu-mir-546)	10(+) 126381790–126381867	HiNF-C (1) [GAGgGCGGGG]	CAC-binding protein (1) [GGGGTG]
		CACCC-binding factor (1) [GGGTGG]	PuF (1) [GGGTGGG]
		Sp1 (2) [GGTGGG (gGGGGAGGGGa)]	
		TTF-1 (1) [GCNCTNNAG]	Adf-1 (1) [GCYGYyGYCGY]
		HFH-8 (1) [TGTTTATNYR]	ANF (1) [cTTTATCTGG]
		GATA-1, -2, -3 (1) [TTATCT]	MAZ (1) [GGGAGGG]
		NF-Atp (1) [GGAGCC]	IL-6 RE-BP (1) [CTGGAA]
		C/EBPβ (1) [CTGGRAA]	
AK133839 (mmu-mir-686)	14(–) 53572279–53572320	Sp1 (1) [GAGGgGTGGT]	AML1 (1) [TGTGGT]

Table 2. *Continued*

cDNAs (miRNAs)	Chromosome, Strand, Position[a]	Transcription Factors and Consensus Sequences	
AK163187 (mmu-mir-698)	4(+) 124244095–124244140	IL-6 RE-BP (1) [CTGGGA]	Zmhox1a (1) [GCTGcCCGTC]
AK030571 (mmu-mir-707)	7(+) 44717300–44717355	N-Oct-3 (1) [MATNNWAAT]	GCF (2) [NNGCGGGGCN]

Predicted transcription factors that bind to highly conserved, upstream regions of cDNA sequences.
[a] The position of conserved genomic regions.
[b] Denotes number of binding sites.

activity. Two kinds of DN-PU.1 constructs with deletions of distinct regions in the transactivation domain were prepared (Figure S3C). We then examined actions of DN-PU.1 on mouse pri-miR-223 promoter activity using a luciferase reporter assay and observed that the expression of PU.1$^{\Delta AD}$, but not PU.1$^{\Delta GD}$, significantly repressed the promoter activity, demonstrating that PU.1 plays a significant role in mouse pri-miR-223 transcription and that the acidic domain of PU.1 is critical for transcriptional control of mouse pri-miR-223 (Figure 3B).

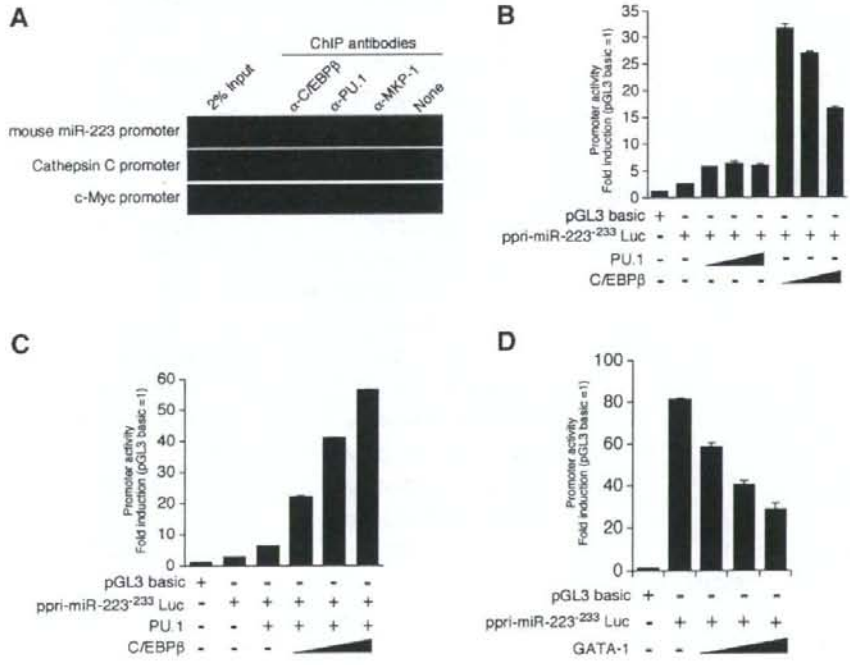

Figure 2. PU.1 and C/EBPβ Activate Mouse Pri-miR-223 Promoter

(A) Interactions of PU.1 and C/EBPβ with the core promoter region of mouse pri-miR-223 were examined in WEHI-3 cells using ChIP assay as described in Experimental Procedures. As controls, Cathepsin C promoter (positive control) and c-Myc promoter (negative control) were assayed. Two percent of input DNA was used as a positive control of PCR. The antibody against MKP-1 was used as negative control for ChIP.

(B) A graded amount (1[0.5 μg], 1/3, and 1/9) of each expression vector together with reporter vectors (Firefly Luc (FL): 0.2 μg; Renilla Luc (RL): 0.05 μg) was transfected into NIH3T3 cells and assayed for a luciferase reporter assay.

(C) A graded dose (1[0.3μg], 1/3, and 1/9) of C/EBPβ expression vector together with PU.1 expression plasmid (0.2 μg) and reporter vectors (FL: 0.2 μg; RL: 0.05 μg) were transfected in NIH3T3 and examined for the pri-miR-223 promoter activity by a luciferase reporter assay.

(D) A graded amount (1[0.5μg], 1/3, and 1/9) of mouse GATA-1 expression vector together with reporter vectors (FL: 0.2 μg; RL: 0.05 μg) was transfected into RAW264.7 cells and examined for pri-miR-223 promoter activity by a luciferase reporter assay. All data are represented as mean ± standard deviation (SD) from triplicate assays.

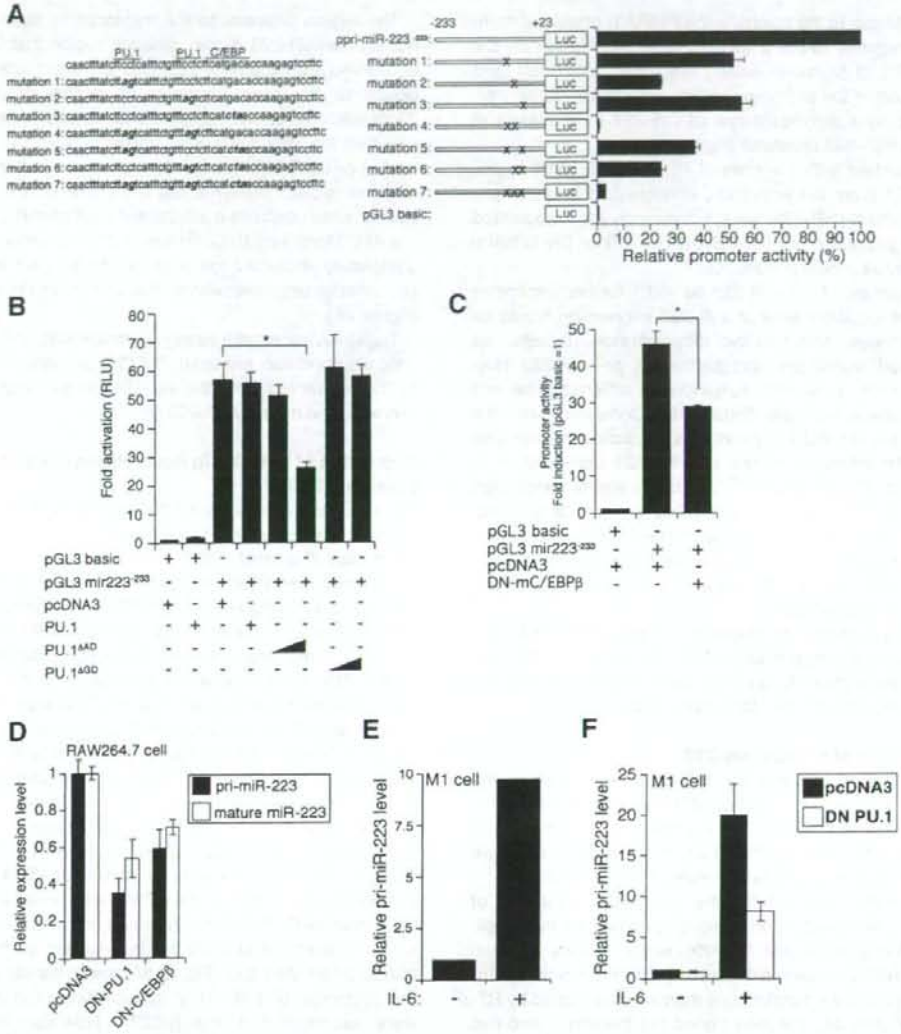

Figure 3. PU.1 as Well as C/EBP as a Regulator of Pri-miR-223 Transcription

(A) The mutant sequences of mouse pri-miR-223 promoter reporter vector are shown (left). The effects of mutations on endogenous pri-miR-223 promoter activity were examined by a luciferase reporter assay. Data are shown as promoter activity relative to the basal endogenous pri-miR-223 promoter activity.

(B and C) Repression of the endogenous mouse pri-miR-223 promoter activity by overexpression of dominant-negative PU.1 (DN-PU.1; B) or C/EBPβ (C). Together with reporter vectors (FL: 0.2 μg, RL; 0.05 μg), indicated expression vectors (0.5 μg) were transfected into RAW264.7 cells. The activity of mouse pri-miR-223 promoter was then examined by a luciferase reporter assay. For DN-PU.1 expression, each expression vector was used with two doses (1[0.5μg] or 1/10). Data are represented as mean ± SD from triplicate assays. * indicates $p < 0.05$ in a Student's t test.

(D) The levels of pri-miR-223 and mature miR-223 in RAW264.7 cells transfected with either pcDNA3, DN-PU.1 ($PU.1^{\Delta AD}$), or DN-C/EBPβ were examined by quantitative real-time RT-PCR as detailed in the Experimental Procedures. Transfectants were enriched as described in the Supplemental Data. Data are shown as mean ± SD from triplicate measurements.

(E) The level of pri-miR-223 in undifferentiated and differentiated M1 cells was examined by quantitative real-time RT-PCR. Data are shown as mean of triplicate assays.

(F) Stable transfectants of M1 cells with either control (pcDNA3) or DN-PU.1 ($PU.1^{\Delta AD}$) expression vector were prepared as described in the Supplemental Data and were cultured in the presence of mrIL-6 for 4 days to induce differentiation. The level of pri-miR-223 in each transfectant was then examined by quantitative RT-PCR. Data are represented as mean ± SD from triplicate tests.

In addition to the assays with DN-PU.1, effects of dominant-negative C/EBPβ (DN-C/EBPβ) expression on the pri-miR-223 promoter activity were also examined, and reduction of the promoter activity was observed as well, confirming a significant role of C/EBPs in activation of the pri-miR-223 promoter (Figure 3C).

Consistent with the roles of PU.1 and C/EBPβ in pri-miR-223 promoter activation, endogenous levels of pri- and mature miR-223 were significantly downregulated by the expression of DN-PU.1 ($PU.1^{\Delta AD}$) or DN-C/EBPβ in RAW264.7 cells (Figure 3D).

Regulation of pri-miR-223 by PU.1 further implicates that the inducible level of miR-223 expression would be synchronous with myeloid differentiation. Indeed, we observed significant upregulation of pri-miR-223 level in M1 cells along with IL-6-induced differentiation into macrophage-like cells (Figure 3E). Consistent with the above results, PU.1 appeared as a critical regulator also for differentiation-induced pri-miR-223 upregulation in M1 cells. The M1-$PU.1^{\Delta AD}$ line that is stably transfected with $PU.1^{\Delta AD}$ vector demonstrated a significantly lower level of pri-miR-223 than that in the M1-mock line after IL-6-induced differentiation (Figure 3F); this also indicates a significant role of PU.1 in pri-miR-223 regulation during myeloid differentiation.

Taken together, our observations indicated that hematopoietic transcription factor PU.1 as well as C/EBPβ are key transcription factors for pri-miR-223 transcription and, therefore, for miR-223 expression.

Regulation of Human miR-223

Given the mechanism for miR-223 expression in the mouse system, we realized a difference between our results and the previous work reporting the mechanism of human miR-223 regulation (Fazi et al., 2005). Hence, we sought to re-examine human miR-223 regulation.

We first searched for the candidate transcript of human pri-miR-223 and found one registered transcript, LOC389865 RNA (XM_374329), which encodes human pre-miR-223 sequence (Figure 4A). This transcript is actually expressed in human bone marrow as detected by RT-PCR (Figure 4A). We then cloned the transcript, and two alternative splicing forms were obtained (Figure 4B). A genome plot of these transcripts demonstrated that this gene has a conventional exon-intron structure as observed in the mouse *miR-223* gene (Figure 4B).

As LOC389865 may not be full-length transcript, we next performed rapid amplification of cDNA ends (RACE) to obtain the information from its 5′ end. This assay would also be informative in discovering whether other alternative transcripts are present or not. We obtained two distinct 5′ RACE products and estimated the consistent lengths for the scales from two alternative splicing forms of human pri-miR-223 (Figure 4C). We then sequenced a total of 36 clones of the RACE products, and 34 clones indicated the exact same transcription start site; also, the other two clones pointed to two sites that are near the major transcription start site (Figure 4D, lower panel).

The region proximal to the transcription start sites of human pri-miR-223 is that genomic region that is highly conserved among different species and is actually homologous to the core promoter of mouse pri-miR-223 (Figure 4D). A reporter assay similar to one in Figure 1C revealed that the 5′-flanking genomic region of cloned human pri-miR-223 has a potent promoter activity in human monocytic leukemia cell lines, THP-1 and U-937, both of which express a substantial level of miR-223 (Figure 4E). Moreover, the deletion of the conserved region completely abolished the promoter activity, showing the promoter structure similar to that of mouse pri-miR-223 (Figure 4E).

Together, our results strongly indicate that, under physiological condition, human miR-223 is probably regulated by the conserved promoter identified in our study as observed in the mouse *miR-223* gene.

Regulation of miR-223 in Acute Promyelocytic Leukemia Cells

We then sought to reassess the regulation of miR-223 in the exactly same experimental system adopted in the previous work (Fazi et al., 2005).

We first determined the transcription start sites of pri-miR-223 in both nontreated and ATRA-treated NB4 cells. Two relatively abundant RACE products (RP1 and RP2) and one minor product (RP3) were obtained in ATRA-treated NB4 cells (Figure 5A). In addition to these products, another minor RACE product (RP4) was detected in nontreated NB4 cells (Figure 5A). Cloning and sequencing of the RACE products revealed that transcripts represented by RP1, RP2, and RP4 are all generated from the exact same major transcription start site identified in the bone marrow sample (Figure 5A, lower panel). The RP4 is the other splicing variant of human pri-miR-223, which was undetectable in either bone marrow or ATRA-treated NB4 cells. The minor product, RP3, was undetectable in bone marrow RNA, and this revealed one alternative transcription start site at 2039 bp downstream of the major transcription start site (Figure 5A, lower panel). Despite the presence of this minor transcription start site, we were convinced that pri-miR-223 in NB4 cells is mainly transcribed from the exact same major transcription start site identified in human bone marrow. Although it is reported that some pri-miRNAs are hardly detectable due to Drosha activity (Lee et al., 2004), this is not the case for pri-miR-223 since the repertoire of pri-miR-223 transcripts in NB4 cells is not altered even by siRNA-mediated Drosha knockdown (Figure S4A).

Using a specific primer set detecting only the major pri-miR-223 transcripts (Figure 5B, middle), we found that the major variants were upregulated in NB4 cells upon ATRA treatment and that this increase of pri-miR-223 was correlated with the upregulation of mature miR-223 (Figure 5B, lower panel). Strikingly, the promoter containing the conserved genomic region was highly activated in NB4 cells during ATRA-induced differentiation (Figure 5C), consistent with the upregulation of pri-miR-223

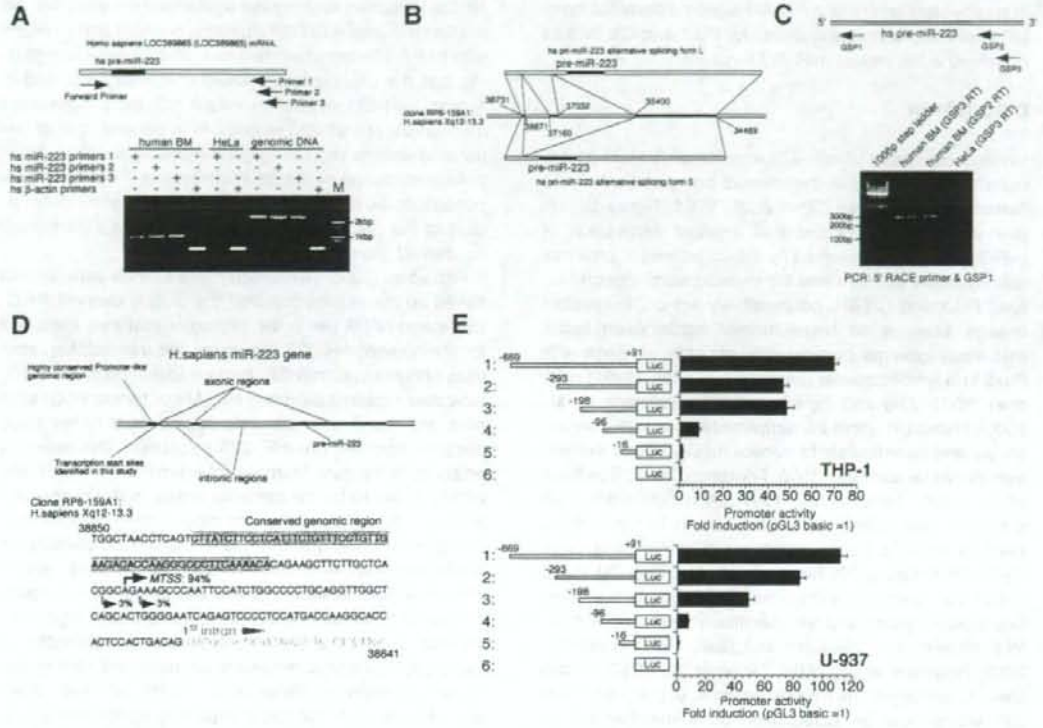

Figure 4. Cloning and Characterization of Human Pri-miR-223 in Bone Marrow

(A) Detection of the putative human pri-miR-223, LOC389865 mRNA, in bone marrow cDNA by RT-PCR with indicated primers.

(B) Cloning of human pri-miR-223. Two alternative splicing forms (Long, L, and Short, S) were cloned as human pri-miR-223, and their genome plots are shown. Numbers denote the position at *H. sapiens* genomic clone, RP6-159A1 from Xq12-13.3.

(C) Cloning of full-length 5′ ends of human pri-miR-223. Full-length transcripts were purified from bone marrow total RNA or HeLa cells and were reverse transcribed with GSP3 to obtain full-length human pri-miR-223 5′ ends.

(D) Schematic for positions of the conserved genomic region, identified transcription start sites, and pre-miR-223 sequence (upper panel). Details of sequences around the transcription start sites of human pri-miR-223 are shown (lower panel). Shaded region indicates the conserved site as shown in Figure 1D and contains the core promoter region. Faint letters denote the first intronic sequence of human pri-miR-223. MTSS indicates *major* transcription start site. Thin arrows indicate the positions of minor transcription start sites.

(E) Luciferase reporter assays were performed in either THP-1 or U937 cells using indicated reporters. 1 indicates ppri-hmiR-223^{-669}-Luc, 2 indicates ppri-hmiR-223^{-293}-Luc, 3 indicates ppri-hmiR-223^{-198}-Luc, 4 indicates ppri-hmiR-223^{-96}-Luc, 5 indicates ppri-hmiR-223^{-16}-Luc, and 6 indicates pGL3 basic.

and the resultant mature miR-223 level in NB4 cells by ATRA treatment.

The roles of PU.1 and C/EBPα in the ATRA-induced pri-miR-223 promoter activity were also examined using RNAi against these molecules. Knockdown of PU.1 resulted in a significant reduction of basal and ATRA-induced pri-miR-223 promoter activity (Figure 5D). The ATRA-inducible pri-miR-223 promoter activity was slightly reduced by a siRNA against C/EBPα (Figure 5D). Consistently, the endogenous level of pri-miR-223 in ATRA-treated NB4 cells was significantly repressed by knockdown of PU.1 using shRNA expression-vector-based RNAi (Figures 5E and 5F). Again, RNAi against C/EBPα resulted in only a slight reduction of ATRA-induced pri-miR-223 upregulation (Figures 5E and 5F). The levels of mature miR-223 were also assayed at the same time and found to be consistent with the levels of pri-miR-223 (Figure S4B). Weak effects of C/EBPα knockdown in the above assays are probably due to redundant activities of other inducible C/EBP members such as C/EBPβ (K.K. and T.F., unpublished data; Duprez et al. 2003) during ATRA-induced differentiation of NB4 cells. Indeed, we observed more effective reduction of the promoter activity and endogenous level of pri-miR-223 in ATRA-treated NB4 cells transfected with DN-C/EBPβ that would suppress actions of multiple C/EBPs (Figures S4C and S4D).

With these results taken together, we concluded that the conserved genomic region acts as pri-miR-223 promoter

in steady state and during ATRA-induced differentiation of NB4 cells and is probably driven by PU.1 and C/EBPs as observed in the mouse miR-223 regulation.

DISCUSSION

Lineage specificity of miR-223 expression is strict as it is detected exclusively in the myeloid compartment of the hematopoietic system (Chen et al., 2004; Figure S2). In this study, we concluded that myeloid expression of miR-223 might be specified by the conserved 5′ proximal *cis*-regulatory element that the myeloid transcription factors, PU.1 and C/EBP, cooperatively act on. In myeloid lineage, there is no single master transcription factor that alone governs myeloid differentiation as seen with Pax5 in B lymphopoiesis (Sieweke and Graf, 1998; Friedman, 2002; Zhu and Emerson, 2002; Rosmarin et al., 2005). Instead, multiple transcription factors work cooperatively and coordinately to control myeloid gene expression (Sieweke and Graf, 1998; Friedman, 2002; Rosmarin et al., 2005; Tenen et al., 1997). Although more than a dozen transcription factors are known to play critical roles in myeloid gene expression (Sieweke and Graf, 1998; Friedman, 2002; Rosmarin et al., 2005), PU.1 and C/EBPs are particularly important as they regulate numerous myeloid genes and are specifically required for myeloid differentiation (Sieweke and Graf, 1998; Friedman, 2002; Rosmarin et al., 2005; Tenen et al., 1997). In this view, together with its myeloid-specific expression, *miR-223* would now be categorized as a member of the so-called "myeloid genes." Neither PU.1 nor C/EBP is truly myeloid restricted since PU.1 is also found in B cells, and C/EBPs are expressed in multiple organs such as liver and adipocytes (Rosmarin et al., 2005). Therefore, myeloid-specific expression of miR-223 probably requires both of these transcription factors. In the presence of a substantial amount of PU.1, C/EBPβ could activate the core promoter in a dose-dependent manner (Figure 2C), indicating that both PU.1 and C/EBP are required for the proper activation of the pri-miR-223 promoter. Together, we speculate that high-level PU.1 confers stable inducibility of the pri-miR-223 promoter activity, and the C/EBP level determines magnitude of the promoter activation in myeloid cells. This idea is consistent with the notion that miR-223 is upregulated during the myeloid differentiation that is associated with increase of PU.1 and C/EBP levels. In addition, we observed repression of endogenous activity of pri-miR-223 promoter by overexpression of erythroid transcription factor, GATA-1 (Figure 2D). This indicates a critical regulatory role of the conserved element in determining pri-miR-223 expression at the turning point of lineage choice between myeloid and erythroid fates, showing an analogy to the regulation of other myeloid genes (Sieweke and Graf, 1998).

Through careful molecular analyses, our study contradicts some aspects of a previous study (Fazi et al., 2005). The miR-223 is evolutionarily conserved, and its expression pattern during myeloid differentiation is also similar in human and mouse systems. However, we did not find the unique C/EBPα-binding element that overlaps with NFI-A-binding site in mouse *miR-223* gene, suggesting that the previously proposed machinery reported in human miR-223 regulation might not be a conserved mechanism of miR-223 regulation. In general, critical genomic elements regulating gene transcription have been conserved during evolution. Furthermore, it has been proposed that the essential genomic elements of miRNAs, including the precursor sequences, might be evolutionary conserved (Berezikov et al., 2005).

Fazi et al. (2005) performed most of their experiments based on the assumption that the unique element for C/EBPα and NFI-A lies in the promoter that they identified for the human miR-223. However, the transcription start sites of human pri-miR-223 that we identified in this study indicated a distinct promoter site. Major transcription start sites are found in a genomic region at a rather large distance from the pre-miR-223 sequence. This region is adjacent to the core promoter of pri-miR-223, which was initially found to be the genomic region highly conserved among different species, including human, mouse, rat, and dog. According to our study, the unique element for C/EBPα and NFI-A competition actually lies in the second intronic region. Therefore, it was still thought to be possible that the element would act as an intronic regulatory element such as an intronic enhancer or repressor. In many cases, intronic regulatory elements are also evolutionarily conserved (Himes et al., 2001). However, there are a few exceptional cases reporting nonconserved intronic enhancers such as in the human adiponectin gene (Qiao et al., 2005). However, even this possibility is unlikely since we could not observe any effects of the intronic sequence on the activity of actual promoter we defined (Figure S4E). Therefore, the actual role of the intronic domain should be revisited in future investigations.

We believe that our study clearly demonstrated the importance of gaining information on pri-miRNAs and neighboring genomic elements for studying miRNA expression. In this view, our approach of finding cDNAs encoding miRNA precursors and following genomic profiling was successful and demonstrated that previously established databases are rich sources of primary miRNA transcripts. Because the databases, like FANTOM3, can provide some information on the tissue origin of each cDNA collection (Yamanaka et al., 2001; Okazaki et al., 2002) in association with EST databases, for example, we can estimate tissue-expression patterns of desired miRNAs by tracing distribution atlases of pairing cDNAs. Indeed, in this study, we could confirm this idea and found an overlapping tissue-expression pattern of miR-223 and its primary transcript, AK036748 (Table 1; Figure S2). In the FANTOM3 database, cDNAs were cloned using the Cap-trapper method that accumulates mRNAs with Cap structure, generating a trustworthy full-length cDNA library (Carninci et al., 2000; Carninci and Hayashizaki, 1999). Determining the exact transcription start sites is a necessary process for analyzing transcriptional regulation of pri-miRNAs,

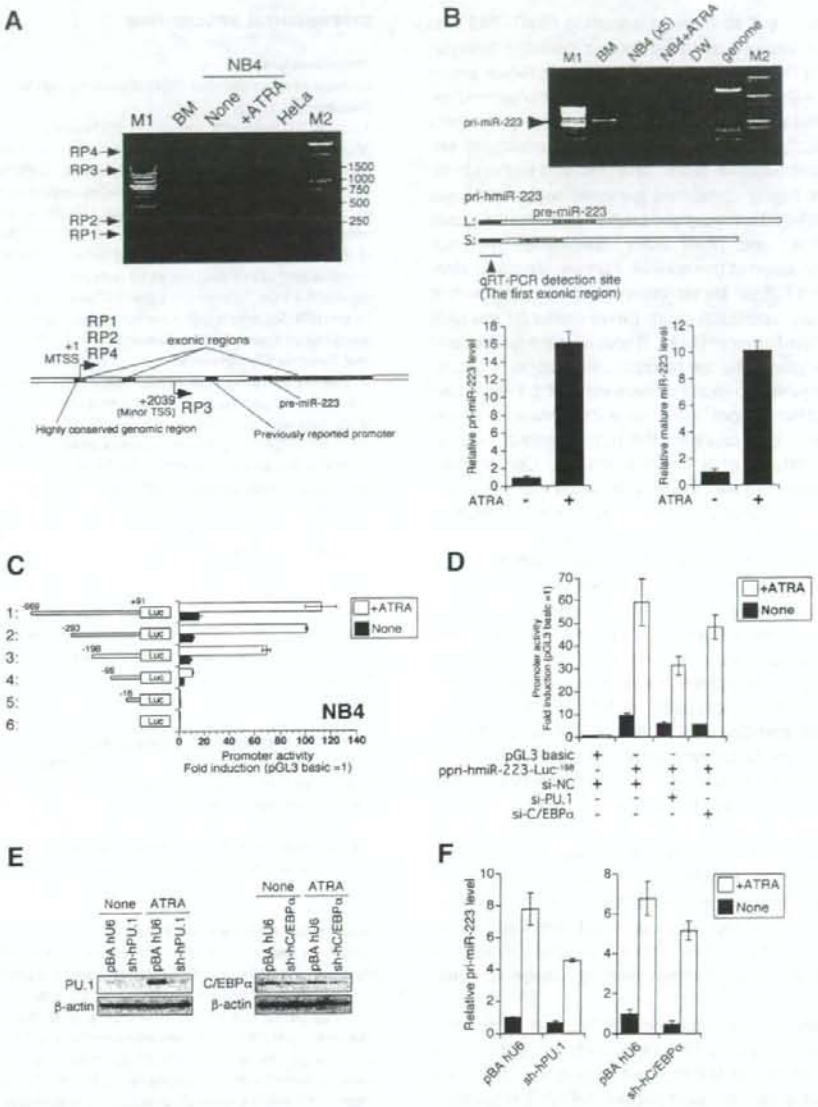

Figure 5. miR-223 Regulation during ATRA-Induced Differentiation of NB4 Cells

(A) Transcription start sites of pri-miR-223 in NB4 cells were determined by 5′ RACE. The 5′ RACE products (RPs) from total RNA samples of bone marrow (BM), NB4 cells (None), ATRA-treated NB4 cells, and HeLa cells are shown (upper panel). M1 indicates 100 bp step ladder; M2 indicates 1 kb DNA ladder (G5711, Promega). The transcription start sites indicated by the 5′ RACE products are shown (lower panel).

(B) Detection of the human pri-miR-223, LOC389865 mRNA, in ATRA-treated and non-treated NB4 cells using PCR with primer set 3 as indicated in Figure 4A (upper panel). BM indicates bone marrow; ×5 indicates five times more RNA template; and DW indicates distilled water. M1 indicates 100 bp step ladder; M2 indicates 1 kb DNA ladder (G5711, Promega). (Middle panel) The target site of human pri-miR-223 for a specific primer set for real-time PCR is shown. (Lower panel) Upregulation of pri-miR-223 (left) and mature miR-223 (right) during ATRA-induced differentiation of NB4 was observed by quantitative real-time PCR.

(C) Potent activation of the conserved pri-hmiR-223 promoter in NB4 cells during ATRA-induced differentiation. Luciferase reporter assay was performed as described in Experimental Procedures. 1 indicates ppri-hmiR-223^{-669}-Luc, 2 indicates ppri-hmiR-223^{-293}-Luc, 3 indicates ppri-hmiR-223^{-198}-Luc, 4 indicates ppri-hmiR-223^{-96}-Luc, 5 indicates ppri-hmiR-223^{-16}-Luc, and 6 indicates pGL3 basic. Data are shown as mean ± SD of triplicate measurements.