Identification of a novel circularized transcript of the AML1 gene

Ai-ning Xu, Xiu-hua Chen, Yan-hong Tan, Xi-ling Qi, Zhi-fang Xu, Lin-lin Zhang, Fang-gang Ren, Si-cheng Bian, Yi Chen & Hong-wei Wang

Department of Hematology, The Second Hospital of Shanxi Medical University, Taiyuan 030001, China

The AML1 gene is an essential transcription factor regulating the differentiation of hematopoietic stem cells into mature blood cells. Though at least 12 different alternatively spliced AML1 mRNAs are generated, three splice variants (AML1a, AML1b and AML1c) have been characterized. Here, using the reverse transcription-polymerase chain reaction with outward-facing primers, we identified a novel non-polyadenylated transcript from the AML1 gene, with exons 5 and 6 scrambled. The novel transcript resisted RNase R digestion, indicating it is a circular RNA structure that may originate from products of mRNA alternative splicing. The expression of the novel transcript in different cells or cell lines of human and a number of other species matched those of the canonical transcript in different cells or cell lines of human and a number of other species. Also, through several experiments, we investigated its origin and structure. Finally, the expression of this transcript in a number of human cell types, and several other species was studied.

INTRODUCTION

The Runt-related transcription factor 1 gene (RUNX-1 alias AML1), located on chromosome 21q22, consists of 12 exons and spans 261 kb of genomic DNA. It plays an important role in establishing hematopoiesis and the generation of hematopoietic stem cells in the embryo (1). Transcription of AML1 is initiated at two distinct 5' regions, a distal region-promoter-1 (P1) and a proximal region-promoter-2 (P2). The two promoters generate at least 12 alternatively spliced mRNAs, which differ within the 5' and 3' untranslated regions (UTR) and coding regions. Among these variants, the predominant transcripts AML1a and AML1b code for a 453-amino acid protein and a 480-amino acid protein, respectively, while the minor AML1c codes for a protein, which has a higher affinity for DNA binding and suppresses the transcriptional activation of AML1b. To date, 55 partner chromosome bands and 21 translocations have been described at the molecular level, which are considered as contributors to leukemia, especially acute myeloid leukemia (AML), using different pathways (1).

In our previous study, we used a pair of outward-facing primers to identify a novel partner gene of AML1 in an acute myeloid leukemia patient. We randomly selected 3 healthy individuals as negative control, and to our surprise, three unexpected bands of the same size could be amplified in all samples. By sequencing, we accidentally discovered the sequence of the exons was arranged in a different order compared to the genomic DNA, which consisted of a 5-6-5 exon alignment. These findings raise several fundamental questions: Where does it originate from? What does its whole structure look like?

In this study, we have presented evidence from RT-PCR using outward-facing primers that further examine its presence. Also, through several experiments, we investigated its structure. Finally, the expression of this transcript in a number of human cell types, and several other species was studied.

RESULTS

AML1 produces the abnormal transcript with the 5-6-5 exon alignment

To examine the abnormal transcript consisting of the 5-6-5 exon alignment, we performed PCR with total RNA from bone marrow or peripheral blood of 10 AML patients and 10 healthy individuals using a pair of outward-facing primers (oF/oR) from exon 5 (as in GenBank D43969). Fragments were observed in all samples at approximately 305 bp, and absent in the negative and blank controls. It was noteworthy that the 305 bp fragment contained an abnormal exon 5 following the original exons 5 and 6 (Fig. 1A).

To further examine whether the product with the abnormal exon order was an artifact, we used additional pairs of primers (6F/5R2 and 5F2/5R1). Also, primers F65 and R65 were designed to cover the boundary of the exons 6-5 and these were used in further PCRs with the primer pairs F65/5R3 and 5F1/R65. We confirmed that these amplified products contained the expected sequences (Fig. 1B).
Fig. 1. Analysis of RT-PCR results. (A) Detection of the novel $AML1$ transcript with the exon 5-6-5 alignment using RT-PCR and sequence analysis. Lanes 1 and 4 show the amplification products of the novel transcript in an AML patient and a healthy individual respectively, Lanes 2 and 5 are negative control, and Lanes 3 and 6 are blank control. The band above 305 bp in Lane 4 is nonspecific amplification product. The positions and directions of outward-facing primers (oF and oR) in the cDNA structure are indicated by arrows. The sizes of the expected fragment and the junction sequence of the 3’ end of exon 6 and 5’ end of exon 5 are also shown. (B) Detection of exon 5-6-5 alignment using different pairs of primers. Positions and directions of different pairs in the cDNA structure are indicated by arrows. The amplification products of primers 6F/5R2, 5F2/5R1, F65/5R3 and 5F1/R65 are shown. By sequencing, the bands of 118 bp and 284 bp in Lanes 1 and 4 and the bands of 116 bp and 315 bp in Lanes 7 and 10 respectively are the products, Lanes 2, 5, 8 and 11 are negative controls, Lanes 3, 6, 9 and 12 are blank controls. (C) Analysis of RT-PCR results and schematic diagram of the positions and directions of primers used to seek the neighboring exons of the exon 5-6-5 alignment. Lanes 1 to 4 are products of RT-PCR using 3F/7aR, 4F/7aR, F65/7aR and F65/7bR, respectively. Lane 5 shows the amplification product of F65/6R, Lane 10 is negative control.

Identification of the neighboring sequences upstream and downstream of the 5-6-5 exon alignment

To explore whether the novel transcript possesses the same neighboring sequences upstream and downstream as the canonical sequences, PCR was performed using a further set of primers. The fragment could be amplified using primers F65/6R (Fig. 1C), while the expected band was not present using the primers 3F/R65, 4F/R65, F65/7aR and F65/7bR. The canonical transcripts $AML1_a$ and $AML1_b$ were successfully amplified using 3F/7aR, 3F/7bR, 4F/7aR and 4F/7bR (Fig. 1C). Hence, we concluded that the 5-6-5 alignment was followed with an exon 6, but it did not contain exons 3 or 4 upstream or exons 7a or 7b downstream.

Poly(A) tail is absent in the exons 5 and 6 scrambled $AML1$ molecule

We performed Rapid Amplification of cDNA Ends (RACE) to obtain either the 3’ or 5’ end of the novel transcript, but were unable to form any fragments (data not shown). To investigate whether a polyadenylated tail was present in this scrambled exon alignment, total RNA was reverse-transcribed with either random primers or oligo(dT) primers. PCR was then performed using the primers shown in Table 1. Only the random primers, and not the oligo(dT) primers, produced the scrambled transcript (Fig. 2A). Also, poly(A)+ and poly(A)- RNA was prepared from total RNA, and the quantity of poly(A)+ and poly(A)- RNA were equal when reverse transcribed using random primers. Canonical transcripts were amplified to testify the purity of the three samples using the primers 4F/5R1. As a
result of amplification using the different pairs of primers shown in Table 1, the expected bands were present in the total RNA and poly(A)- RNA Lanes while absent in the poly(A)+ Lane (Fig. 2B).

The novel AML1 transcript is probably circular

As the transcript lacked the poly(A) tail, we used the One Step PrimeScript™ miRNA cDNA Synthesis Kit (Takara, Catalog No. D350A) to add a segment of polyadenylate to it when it was reverse transcribed, allowing its 3’ end to be amplified using our special forward primer and the Uni-miR qPCR Primer which is reverse to the adaptor sequence supplied in the kit. We were unable to add a tail to the novel transcript, while the reference gene U6 formed the expected mono peak after real-time PCR and the expected band on an agarose gel after conventional PCR (data not shown).

To further testify the confirmation of the non-canonical transcript, we determined its sensitivity to the RNA exonuclease R (RNase R), which specifically digests linear RNA, both structured and non-structured, but does not digest circle or lariat RNA (2, 3). We performed real-time PCR to detect the change of its expression after RNase R digestion using cDNAs of five healthy individuals. The results showed a significant increase in the circular transcript. The result of calculation using the equation $2^{-\Delta \Delta Ct}$ showed there was on average a 282-fold increase (Table 2), confirming its circular nature.

The expression of the novel transcript in different cells or cell lines of human and several other species matched those of the canonical transcripts

To determine the ubiquity of the novel transcript, we employed RT-PCR using cDNAs from different cells or cell lines of human and several other species (Fig. 2C). It turned out that the novel circular transcript of AML1 existed not in lower organisms or plant, but in higher animals we selected.

**Table 1. Nucleotide sequence of primers**

| Primer | Sequence |
|--------|----------|
| 3F     | 5'-TGCCCATCGCTTTCAAGGTG-3' |
| 4F     | 5'-ACCGCAGCCATGAAGAACCAGGTTG-3' |
| 5F1    | 5'-GGAAAAGCTTCACTCTGACCATC-3' |
| 5F2    | 5'-CGCCACCTACCACAGAGCCATCA-3' |
| oF     | 5'-GTCTTCACAAACCCACCG-3' |
| oR     | 5'-GTGAAGACAGTGATGGTCAG-3' |
| 5R     | 5'-GATGGCTCTGTGGTAGGTGGCGACT-3' |
| 6F     | 5'-AACCACTCCAACGGCGTCAATTAC-3' |
| 6R1    | 5'-CAGTGGAGTGTTGCTAGGAGGAG-3' |
| 6R2    | 5'-CTGCTATGACCTGTCCGGCTG-3' |
| F6S    | 5'-CTCACAGTCAAGTCAGAAGGGGA-3' |
| R6S    | 5'-CTCACAGTGAACTGTCTTTCCCCT-3' |
| 7aR    | 5'-ATCCTCCCACTCGTTGTGTCG-3' |
| 7bR    | 5'-CAGTATTTGGTACGACTGTAGTGTAAG-3' |

*Primers oF and oR are designed outward facing.

**DISCUSSION**

In this manuscript, we have successfully identified a novel AML1 transcript containing the donor site of exon 6 linked to the acceptor site of exon 5. Our investigations have gained an insight into this novel transcript, characterizing its sequence and conjecturing on its structure. The observation of the novel transcript prompted experimental analysis of several possible interpretations of this phenomenon. Originally, this transcript was thought to arise from either the expression of a duplicate exon 5 to exon 6 genomic segment or an intermolecular splicing (trans-splicing) of two AML1 pre-mRNAs. When reverse
A novel circularized transcript of AML1
Ai-ning Xu, et al.

Table 2. The enrichment of the novel transcript after RNase R digestion

| Sample | CT value | R+ ± Mean ± standard deviation |
|--------|----------|--------------------------------|
|        | Gene ABL1 | R- ± 2^\(\Delta C\) |
| 1      | 19.111 | 18.1087 | 27.3767 | 17.1736 | 588.4190 |
| 2      | 29.328 | 20.1606 | 35.2505 | 19.0501 | 130.9253 |
| 3      | 20.336 | 19.6205 | 26.3639 | 18.8059 | 114.7619 |
| 4      | 24.9551 | 21.1963 | 32.4336 | 20.4433 | 300.5581 |
| 5      | 20.9352 | 19.1885 | 28.8849 | 19.0151 | 278.8026 |

\(R^+\) and \(R^-\) represent the groups with or without RNase R digestion respectively.
The non-coding characteristic of the circular RNA species are reminiscent of the group of non-coding RNAs, which are also products of the splicing process and challenge the central dogma. Recent studies indicate that members of this non-coding RNA family are far more abundant and important than considered initially (22). What is the relationship between circular RNAs and non-coding RNAs? Could we suppose that these circular RNA transcripts belong to the non-coding RNA species? Since the discovery of microRNA and the reporting of their meaningful function, more attention has been paid to these non-canonical mRNAs.

Whether the circular transcripts encode protein or not, study on their biological function has been going through a bottleneck period. To date, due to the limited approach and technology, there is no way to figure out what physiological or pathological functions circular RNA may have. Therefore, it still requires us to seek new approach or develop new technology to further disclose the mysterious mask of these circular transcripts in the future.

In summary, we identified a novel transcript from the AML1 gene, which is likely to be a circular RNA molecule. We concluded that the novel circular transcript of AML1 is just consisted of exon5 and exon6. In other words, in the novel transcript exon5 and exon6 were joint head to tail. Therefore, we inferred its full length is length of exon5 plus that of exon6, that is 297 bp. The novel transcript was identified in different organisms. The circular RNA transcripts may suggest a novel RNA species, which may help to integrate the genomic context. But this still leads to the debate about their unknown biological role. They should attract our attention, just like the non-coding RNA. We believe there is likely to be circular RNA transcripts of other genes, though there is yet no dogma. Recent studies indicate that members of this non-coding RNA family are far more abundant and important than considered initially.

**MATERIALS AND METHODS**

**Sample preparation**

Bone marrow samples were from healthy individuals and AML patients, who were diagnosed using the MICM classification standard. Oral mucosa and nucleus pulposus cells of an intervertebral disc were also obtained from healthy individuals. All samples were collected with informed consent from the subjects. The patients’ IRB number is NO. 2012057 from Shanxi Medical University Medical Ethics Committee.

U266, K562, SMMC-7721, AGS and NRK52E cells were deposited in our laboratory, and cDNAs of tetrahymena, Euplotes and C. elegans were provided by Dr. Shen Quan of our laboratory, and cDNAs of tetrahymena, Euplotes and C. elegans were provided by Dr. Shen Quan. U266, K562, SMMC-7721, AGS and NRK52E cells were deposited in our laboratory, and cDNAs of tetrahymena, Euplotes elegans and Tartary buckwheat were given by Dr. Shen Quan.

**Preparation of total RNA, poly(A)+ and poly(A)-RNA**

Total RNA was extracted using RNeasy Plus (Takara, Catalog# D9108A) according to the instruction of manufacturer. The mRNA was isolated from total RNA using PolyATtract™ mRNA Isolation Systems III (Promega, Cat#Z5300), and the poly(A)-RNA was recycled from the supernatant by ethanol precipitation. Purity and concentration were detected by NanoDrop 2000 (Thermo, USA) before being stored at −80°C.

**Reverse transcription-polymerase chain reaction (RT-PCR)**

RNA (1 μg) was in vitro reverse-transcribed into cDNA with reverse transcriptase M-MLV (RNase H-) (Takara, Tokyo, Japan, Cat#D2640A) using random primers (Promega, Madison, WI, USA) and an oligo(dT) primer (Takara, Cat#D311), respectively. As a negative control, cDNA prepared without reverse transcriptase was employed.

The PCR was performed with a set of outward-facing primers designed to bind in exon 5, 3′ and 5′. Amplification conditions were as follows: denaturation at 94°C for 3 min; 35 cycles of denaturation at 94°C for 30 s, annealing at 53°C for 30 s, and extension at 72°C for 30 s; plus an final extension step at 72°C for 10 min. PCR was also employed with other sets of primers, using 60°C as the annealing temperature. The primers were designed by Primer 3.0 (Table 1). Each 25 μl PCR reaction contained 1 μl cDNA template, 2.5 μl of 10× Taq buffer, 1.5 μl of 25 mM MgCl2, 0.5 μl of 10 mM dNTPs, 1.5 μl of Taq DNA polymerase (Fermentas, Cat#EP0282) and 1 μmol/L of forward and reverse primers. Aliquots (10 μl) of the amplified PCR products were loaded onto 2% agarose gels for electrophoresis and visualization with ethidium bromide. The band was cut from the gel, purified using the Gel Extraction Kit (Omega, Cat#D2500-01), and sequenced by the Beijing Auqc Biotechnology Company.

**RNase R digestion**

RNA (20 μg) was incubated with or without 40 U of RNase R (Epicentre Biotechnologies, Cat# RNR07250) for 2 h at 37°C. The resulting RNA was purified by ethanol precipitation and quantified. Equal amounts of RNA with or without RNase R treatment were subjected to reverse transcription using random primers. Expression of the novel transcript was detected using SYBR®Green-based real-time PCR, which was performed using the 7300 Real-Time PCR System (Applied Biosystems, USA) with SYBR® Green I detection and 1 μmol/L of both forward and reverse primers. Aliquots (10 μl) of the amplified PCR products were loaded onto 2% agarose gels for electrophoresis and visualization with ethidium bromide. The band was cut from the gel, purified using the Gel Extraction Kit (Omega, Cat#D2500-01), and sequenced by the Beijing Auqc Biotechnology Company.

**Acknowledgements**

We would like to thank Jiayan Ran, Zhenhua Qiao, and Linhua Yang, the Second Hospital of Shanxi Medical University, who...
recruited the patients in the study. This work was supported by grant from the National Natural Science Fund (30772064) of China.

REFERENCES

1. De Braekeleer, E., Douet-Guilbert, N., Morel, F., Bris, M., Ferenc, C. and De Braekeleer, M. (2011) RUNX1 translocations and fusion genes in malignant hemopathies. Future Oncology. 7, 77-91.

2. Burd, C. E., Jeck, W. R., Liu, Y., Sanoff, H. K., Wang, Z. and Sharpless, N. E. (2010) Expression of linear and novel circular forms of an INK4/ARF-Associated Non-Coding RNA correlates with atherosclerosis risk. PLoS Genet. 6, e1001233.

3. Suzuki, H., Zuo, Y., Wang, J., Zhang, M. Q., Malhotra, A. and Mayeda, A. (2006) Characterization of RNase R-digestible cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res. 34, e63.

4. Miyoshi, H., Ohira, M., Shimizu, K., Mitanil, K., Hirai, H., and Imai, T. (1995) Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. Nucleic Acids Res. 23, 2762-2769.

5. Nigro, J., Cho, K., Fearon, E., Kern, S., Ruppert, J. and Oliner, J. (1991) Scrambled exons. Cell 64, 607-613.

6. Cocquerelle, C., Daubersies, P., Majerus, M. A., Kerckaert, J. P. and Bailleul, B. (1992) Splicing with inverted order of exons occurs proximal to large introns. EMBO J. 11, 1095-1098.

7. Cocquerelle, C., Mascrez, B., Hetuin, D. and Bailleul, B. (1993) Miss-splicing yields circular RNA molecules. FASEB J. 7, 155-160.

8. Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M. and Koopman, P. (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73, 1019-1030.

9. Zaphiropoulos, P. G. (1996) Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. Proc. Natl. Acad. Sci. 93, 6536-6541.

10. Zaphiropoulos, P. G. (1997) Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. Mol. Cell. Biol. 17, 2985-2993.

11. Surono, A., Takeshima, Y., Wibawa, T., Ikezawa, M., Nonaka, I. and Matsuo, M. (1999) Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. Hum. Mol. Genet. 8, 493-500.

12. Gualandi, F., Trabaneli, C., Rimessi, P., Calzolari, E., Toffolatti, L., Patamello, T., Kunz, G., Muntoni, F. and Ferlini, A. (2003) Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. J. Med. Genet. 40, e100.

13. Li, X. F. and Lytton, J. (1999) A circularized sodium-calcium exchanger exon 2 transcript. J. Biol. Chem. 274, 8153-8160.

14. Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. and Kleinschmidt, A. K. (1976) Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proc. Natl. Acad. Sci. 73, 3852-3856.

15. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. and Brown, P. O. (2012) Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS ONE 7, e30733.

16. Takahara, T., Kanazu, S., Yanagisawa, S. and Akanuma, H. (2000) Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing. J. Biol. Chem. 275, 38067-38072.

17. Flouriot, G., Brand, H., Seraphin, B. and Gannon, F. (2002) Natural trans-spliced mRNAs are generated from the human estrogen receptor-α (hERα) gene. J. Biol. Chem. 275, 38067-38072.

18. Axtell, M. J., Westholm, J. O. and Lai, E. C. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12, 221.

19. Chen, C. Y. and Sarnow, P. (1995) Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. Science 268, 413-417.

20. Kozak, M. (1989) The scanning model for translation: an update. J. Cell Biol. 108, 229-241.

21. Sachs, A. B. and Davis, R. W. (1989) The poly(A) binding protein is required for poly(A) shortening and 60s ribosomal subunit-dependent translation initiation. Cell 58, 857-867.

22. Storzm, G. (2002) An expanding universe of noncoding RNAs. Science 296, 1260-1263.