Leukemia Research 82 (2019) 7–10

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

Correspondence

A predicted novel protein isoform of *HOXA9*

1. Introduction

The *Hox* genes are master regulators of developmental processes [1,2]. In mouse and human, *HoxA9* is responsible for the development and differentiation of hematopoietic stem cells into myeloid and lymphoid progenitors [3–5]. *HoxA9* has also been implicated in oncogenesis of prostate cancer, breast cancer, ovarian cancer, bladder cancer, colorectal carcinomas and glioblastoma [6]. *HoxA9* acts as a tumour suppressor gene in carcinomas but as an oncogene in haematological malignancies [7]. Most importantly, *HoxA9* expression is critical in acute myeloid leukemia (AML); overexpression of *HoxA9* has been found in over 50% of AML [8]. *HoxA9* overexpression is strongly associated with poor AML prognosis and has been suggested as a potential therapeutic target [8].

As monomers, the HOX proteins weakly bind to DNA, facilitated by the homeodomain, a highly conserved region near the C-terminus [9,10]. Interactions between HOX proteins and other transcription factors significantly change binding sequences and affinities [11]. Of particular interest is that the interaction of the HOXA9 canonical protein with MEIS1 and PBX3 is critical for leukemogenesis [8].

Two main transcripts are associated with the *HoxA9* gene: the canonical transcript (henceforth referred to as the *HoxA9* transcript) arises from 2 exons (Fig. 1A), and an alternate transcript (*HoxA9T*) which arises when an intron is excised out of the first exon resulting in a frameshift (Fig. 1B and C) [12,13]. This splice event is conserved in chicken, mouse and human [12]. *HoxA9T* is thought to play a major role in leukemia [14] though the precise nature of its role has been elusive [14,15].

The two *HoxA9* transcripts potentially give rise to three proteins. It is currently assumed that the *HoxA9* and *HoxA9T* transcripts each encode only one isoform; with translation initiating at the canonical start codon giving the canonical protein (Fig. 1A) and a truncated protein (Fig. 1B) respectively [13]. In this study we describe a third alternate protein isoform that has been overlooked. We find that *HoxA9T* can potentially give rise to another isoform with translation initiating at an alternate ATG in both human and mouse resulting in a protein isoform with a novel amino terminus sequence, but retaining the full homeobox domain (Fig. 1C).

2. Methods

The alternate protein isoform in mouse was previously predicted [16]. For verification of other species, sequence data were downloaded from ENSEMBL and NCBI (Gg, NM_001305224; Mm, NM_010456, NM_001277238; Hs, NM_152739; Xl, NM_001097795.1; Rn, CV103558.1, CK480478.1).

PCR was performed using Taq polymerase (NEB) from mouse kidney cDNA using primers 5′-CTGCGCGATCCCTTTGCATAA (forward) and 5′-TCTCGGCATTGTTTTCGGAGAAG (reverse); annealing temperature 55 °C. Amplicons were separated on a TBE agarose gel and sequenced via Sanger sequencing (ACRF, ANU).

3. Results

The mouse alternate protein arising from the *HoxA9T* transcript was predicted alongside 2000 mammalian novel isoforms from our previous bioinformatics study [16]. The alternative human HOXA9 protein was also predicted in our bioinformatics study [16], however as human Unigene classifies *HoxA9* and *HoxA10* into a single gene cluster the protein was mis-labelled as arising from *HoxA10*. Fig. 1 compares these novel predicted proteins to the currently assumed HOXA9 isoforms. For ease of discussion we will refer to the protein isoforms according to Fig. 1: A is the canonical full-length HOXA9, B is the previously considered truncated isoform and C is our predicted novel HOXA9.

The alternate start codon in mouse was upstream from the annotated transcript on ENSEMBL and thus we confirmed its presence experimentally in cDNA prepared from mouse kidney lysate by PCR amplification. Sequencing confirmed that the cDNA of *HoxA9T* extends upstream to include the alternate ATG. Our novel mouse HOXA9 protein isoform is 295 amino acids long with a predicted molecular weight of 33.49 kDa.

Examination of human *HoxA9T* revealed a series of stop codons in the alternate reading frame that are present immediately upstream of the canonical start codon (red lines in Fig. 1C). This prevents a viable protein from being translated from an upstream alternate start codon. Translation initiating at the first alternate ATG downstream of the canonical ATG would produce a novel protein of 185 amino acids (of which 76 precede the cryptic intron and are read in an alternate frame) with a molecular weight of 21.89 kDa.

A diagrammatic representation of the canonical human HOXA9 protein is shown in Supplementary Fig. 1. Our alternate HOXA9 protein sequences in both mouse and human would maintain the PBX binding hexapeptide domain and the homeobox domain; but have a completely different upstream sequence to the canonical HOXA9 protein (blue in Fig. 1C). Alignments of the mouse and human HOXA9 proteins are shown in Fig. 2. It can be seen that the human and mouse alternate sequences show a high degree of homology, possibly implying selective pressure and functional significance.

Dintilhac et al. [12] experimentally verified the existence of the *HoxA9T* transcript in mouse, human and chicken and predicted its existence in rat and xenopus. We have identified a supporting EST for rat (CK480478.1). Currently ENSEMBL only annotates *HoxA9T* in mouse and human [12]. *HoxA9T* RNA sequences for chicken and xenopus were reconstructed using data from Dintilhac et al. [12] and ENSEMBL. Again, stop codons upstream of the canonical start codon prevent viable proteins being translated from upstream start codons. An alternate protein is unlikely for chicken as the first downstream ATG would result in a protein 44 amino acids in length and missing much of the

https://doi.org/10.1016/j.leukres.2019.05.002
Received 22 February 2019; Received in revised form 15 April 2019; Accepted 3 May 2019
Available online 04 May 2019
0145-2126/ © 2019 Elsevier Ltd. All rights reserved.

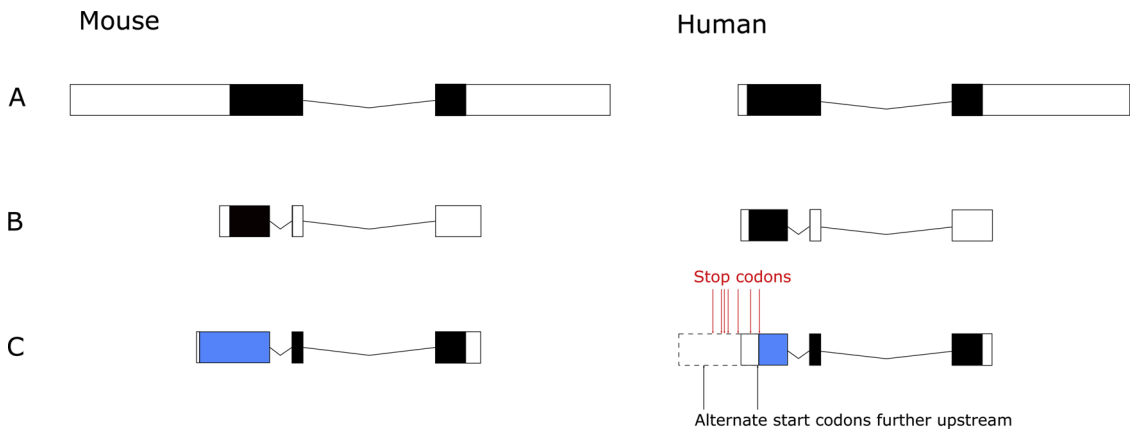

Fig. 1. Schematic representation of alternate splicing and protein isoforms in murine and human *HoxA9*. Boxes represent exons and lines represent introns. Blank boxes represent untranslated sequence; coloured boxes represent translated regions of exons; boxes translated in the same frame are given the same colour. Gene diagrams in row A are representations of the canonical *HoxA9* transcripts and the resulting proteins in mouse and human. The diagrams in row B represent the alternate *HoxA9T* transcripts in mouse and human and the previously predicted isoforms that are translated from the canonical start codon. The diagrams in row C also represent the alternate *Hoxa9T* transcripts but show how translation from an alternate ATG gives rise to the novel isoforms. The box with dotted lines in human represents sequence upstream of the transcription start site of human *HoxA9T* that has not been verified. Note that in the murine alternate transcript the alternate start is upstream of the canonical start, while in the human alternate transcript it is downstream. Start and stop codons in the alternate reading frame that are upstream of the canonical ATG are labelled in human *HoxA9*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

```
Mouse nov.   MAFAIKIMTAKHRAINSVRSDLRVIVLPGGHVTRVANGGAGAGNLLGDCTSPPPGATKLLHEICSFIISAGRAGPARRGLLQWPPPGP
Human nov
Mouse can.                          MATTGALGNYYVDSFLLGADAADELGAGRYAPGTLGQPPRQAAALAEHPDFSPCSFQSKAAVFGA
Human can.                          MATTGALGNYYVDSFLLGADAADELSVGRYAPGTLGQPPRQAATLAEHPDFSPCSFQSKATVFGA

Mouse nov.   WATTMWTPSCWAPTLLMSWVRDATLQGPWVNPQGRRQLWPNTPTSVLAASSPRRRCLVPRGTQCTRRAPM    RCLLQCIITTTTPT
Human nov.                          MRRGPWASLPGRRRRWPSTPTSARAASSPRRRCLAPRGTQCTRRAPTLYPLRCTTT  I TTTPT
Mouse can.   SWNPVHAAGANAVPAAVYHHHHH PYVHPQAPVAAAAPDGR(*)YMRSWLEPTPGALSFAGLPSSRPYGIKPEPLSARRGDCPTLDTH
Human can.   SWNPVHAAGANAVPAAVYHHHHHHPYVHPQAPVAAAAPDGR(*)YMRSWLEPTPGALSFAGLPSSRPYGIKPEPLSARRGDCPTLDTH

Mouse nov.   CIPRRPWRRRRRTAVDREKQPSEGAFSENNAENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRY
Human nov.   CTPRRPWRRRRRTAVDREKQPSEGAFSENNAENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRY
Mouse can.   TLSLTDYACGSPP VDREKQPSEGAFSENNAENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRY
Human can.   TLSLTDYACGSPP VDREKQPSEGAFSENNAENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRY

Mouse nov.   EVARLLNLTERQVKIWFQNRRMKMKKINKDRAKDE*
Human nov.   EVARLLNLTERQVKIWFQNRRMKMKKINKDRAKDE*
Mouse can.   EVARLLNLTERQVKIWFQNRRMKMKKINKDRAKDE*
Human can.   EVARLLNLTERQVKIWFQNRRMKMKKINKDRAKDE*
```

Fig. 2. Alignment of HOXA9 protein sequences in mouse and human. can. = canonical sequence; nov. = novel sequence. Black font is for sequence with little alignment. Red and gray font is for conserved sequence between novel isoforms and canonical isoforms respectively. Blue font is for the PBX-binding domain (hexapeptide sequence), the homeodomain region is boxed and the position of the MEIS1-interacting region underlined (these are explained in more detail in Supplementary Fig. 1). For the canonical sequences, the position of the premature stop codon arising from the alternate transcript is given by (*). Note that the PBX-interacting and the homeodomain regions are conserved across the isoforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

homeodomain. The predicted alternate proteins for xenopus and rat both include the homeodomains and, in the case of rat, also the PBX-interacting hexapeptide region (Supplementary Fig. 2). Thus, we predict alternate HOXA9 proteins are likely to exist in mouse, human, rat and possibly xenopus.

4. Discussion

We have identified an overlooked potential novel protein isoform (isoform C) arising from *HoxA9T* in human, mouse and rat. The isoform would have a novel amino terminal sequence yet would retain the PBX-interacting hexapeptide sequence and the DNA-binding homeobox domain. Currently it has been assumed that the only isoform arising from *HoxA9T* is a truncated form of the HOXA9 canonical (isoform B) which retains the amino terminus - a region that interacts with HOXA9 collaborating proteins and cofactors but not the C-terminus which contains the DNA-interacting homeodomain and the PBX-interacting hexapeptide sequence [13]. Previous studies have implicated *HoxA9T* as being important in leukemogenesis [14], but as isoform B cannot bind DNA or PBX3 (a critical cofactor in cytogenetically abnormal AML) the potential for isoform C should be investigated.

While the canonical HOXA9 transcripts both have optimal kozak sequences at the translation initiation site, the alternate transcripts have no kozak sequence [16]. However, it is worth noting that weak or missing kozak sequences do not preclude translation: Kozak herself showed that translation can occur in the absence of kozak sequences [17]. Moreover, a recent review of over 1500 transcripts in humans [18] found that a significant proportion of AUG start codons have no kozak sequence. More recently, sub-optimal kozak sequences were shown to be translated more efficiently than those with optimal kozak sequences in times of stress [19]. Together, these data suggest that kozak sequences are another mechanism to regulate gene expression. Hence, our novel HOXA9 isoform, though not necessarily the predominant form, is certainly worth consideration in tumorigenesis.

Our novel HOXA9 isoform is predicted to contain the PBX hexapeptide-interacting domain and the homeodomain, but lacks the MEIS1-interacting region. The functional implications of the lack of the MEIS1-interacting region are hard to predict: on the one hand MEIS1 has been postulated to accelerate (rather than induce) leukemic transformation [20]; on the other the HOXA9 protein fusions to NUP98 lack this region and still lead to AML [21]. Therefore, even though lacking the MEIS1-interacting site, the novel HOXA9 isoform could still be oncogenic.

Two publications have studied the leukemogenicity of the *HoxA9T* splice form [14,15]. He et al. [15] retrovirally transduced two constructs, full length human *HOXA9* and human *HOXA9T* into hematopoietic progenitor cells. Their *HOXA9* construct would encode isoform A, but also potentially isoforms B and C (as it could undergo further splicing as shown by Stadler et al. [14]). Their *HOXA9T* construct could potentially encode isoforms B and C. Without confirmation of which isoforms are expressed these findings do not shed light on the possibility of our hypothetical novel isoform [15].

Stadler et al. [14] retrovirally transduced three constructs (human *Hoxa9WT*, human *Hoxa9FLim*, mouse *Hoxa9T*) into bone marrow progenitor cells to determine the leukemogenicity conferred by alternate splicing. Human *Hoxa9WT* would encode isoform A, but also potentially isoforms B and C; human *Hoxa9FLim* could only encode human isoform A (as it contains point mutations to prevent splicing) and mouse *Hoxa9T* could only encode mouse isoform B.

Stadler et al. used three different assays to measure the leukemogenic potential of these constructs. Of these, results from serially replating transduced colony forming units suggests a role for our proposed novel protein. Here, cells co-transduced with *Hoxa9FLim/Hoxa9T* (producing both isoforms A + B) had substantially lower colony forming potential than *Hoxa9WT* cells (A + B + C) (Supplementary Fig. 3 of their paper). This would be consistent with a role for isoform C in leukemogenesis.

Despite its potential significance, it is perhaps unsurprising that this novel HOXA9 isoform has been overlooked, as proteins arising from alternative translation initiation sites are generally under-appreciated [16,22]. There are two main reasons for this: (1) empirical validation relies on generating novel antibodies and (2) theoretical validation requires protein mass spectral databases (which in turn need pre-existing annotated Uniprot and Genbank sequences) [22].

However, there is increasing evidence of such isoforms having biological significance. Two well studied examples of an alternative out-of-frame start codon giving rise to biologically significant alternative proteins are RNase K and Otubain 1. The alternative start codon in both cases is upstream of the canonical, producing isoforms with distinct N-terminal sequences but the same C-terminal sequences. In RNase K, the canonical and alternate proteins are co-expressed, with the ratio varying under different conditions [23] suggesting both isoforms have significant but differing functional roles. Otubain 1 offers an intriguing example where the alternate isoform has the complete opposite effect of the canonical isoform: expression of the canonical isoform enhances immune function, whereas expression of the alternate isoform dampens immune function [24]. These examples illustrate the important insight gained from investigating isoforms with distinct N-termini but conserved C-termini, such as our novel HOXA9 isoform.

In summary, we propose a novel, hitherto overlooked, protein encoded by *HoxA9T* which should be investigated. Though *HoxA9T* has an important influence on leukemogenesis it has been difficult to understand how, as the currently assumed isoform is missing key leukemogenic domains. In contrast, our isoform retains these key regions and thus warrants further investigation to provide new insight into the precise nature of *HoxA9T* in leukemogenesis.

Conflict of interest

All authors declare no conflicting interests.

Acknowledgements

ERA was supported by an Australian Government Research Training Program Stipend Scholarship. JDW by a National University Scholarship (ANU).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.leukres.2019.05.002.

References

[1] F.R. Goodman, Limb malformations and the human HOX genes, Am. J. Med. Genet. 112 (3) (2002) 256–265.
[2] R. Krumlauf, Hox genes in vertebrate development, Cell 78 (2) (1994) 191–201.
[3] H.J. Lawrence, G. Sauvageau, R.K. Humphries, C. Largman, The role of HOX homeobox genes in normal and leukemic hematopoiesis, Stem Cells 14 (3) (1996) 281–291.
[4] M.C. Magli, C. Largman, H.J. Lawrence, Effects of HOX homeobox genes in blood cell differentiation, J. Cell. Physiol. 173 (2) (1997) 168–177.
[5] V. Ramos-Mejía, O. Navarro-Montero, V. Ayllón, C. Bueno, T. Romero, P.J. Real, et al., HOXA9 promotes hematopoietic commitment of human embryonic stem cells, Blood 124 (20) (2014) 3065–3075.
[6] S. Bhatlekar, J.Z. Fields, B.M. Boman, HOX genes and their role in the development of human cancers, J. Mol. Med. 92 (8) (2014) 811–823.
[7] L. Zhou, Y. Wang, M. Zhou, Y. Zhang, P. Wang, X. Li, et al., HOXA9 inhibits HIF-1α-mediated glycolysis through interacting with CRIP2 to repress cutaneous squamous cell carcinoma development, Nat. Commun. 9 (1) (2018) 1480.
[8] C.T. Collins, J.L. Hess, Role of HOXA9 in leukemia: dysregulation, cofactors and essential targets, Oncogene 35 (9) (2016) 1090–1098.
[9] A. Laughon, DNA binding specificity of homeodomains, Biochemistry 30 (48) (1991) 11357–11367.
[10] I. Pellerin, C. Schnabel, K.M. Catron, C. Abate, Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster, Mol. Cell. Biol. 14 (7) (1994) 4532–4545.
[11] M. Slattery, T. Riley, P. Liu, N. Abe, P. Gomez-Alcala, I. Dror, et al., Cofactor binding evokes latent differences in DNA binding specificity between hox proteins, Cell 147 (6) (2011) 1270–1282.
[12] A. Dintilhac, R. Bihan, D. Guerrier, S. Deschamps, I. Pellerin, A conserved non-homeodomain Hoxa9 isoform interacting with CBP is co-expressed with the 'typical' Hoxa9 protein during embryogenesis, Gene Expression Patterns 4 (2) (2004) 215–222.
[13] S. Fujimoto, K. Araki, O. Chisaka, M. Araki, K. Takagi, K.-i. Yamamura, Analysis of the murine hoxa-9 cDNA: an alternatively spliced transcript encodes a truncated protein lacking the homeodomain, Gene 209 (1–2) (1998) 77–85.
[14] C.R. Stadler, N. Vegi, M.A. Mulaw, K.E. Edmaier, V.P.S. Rawat, A. Dolnik, et al., The leukemogenicity of Hoxa9 depends on alternative splicing, Leukemia 28 (9) (2014) 1838–1843.
[15] M. He, P. Chen, S. Arnovitz, Y. Li, H. Huang, M.B. Neilly, et al., Two isoforms of HOXA9 function differently but work synergistically in human MLL-rearranged leukemia, Blood Cells Mol. Dis. 49 (2) (2012) 102–106.
[16] L.O.W. Wilson, A. Spriggs, J.M. Taylor, A.M. Fahrer, A novel splicing outcome reveals more than 2000 new mammalian protein isoforms, Bioinformatics 30 (2) (2014) 151–156.
[17] M. Kozak, Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems, Mol. Cell. Biol. 9 (11) (1989) 5073–5080.
[18] S. Peri, A. Pandey, A reassessment of the translation initiation codon in vertebrates, Trends Genet 17 (12) (2001) 685–687.
[19] J.M. Acevedo, B. Hoermann, T. Schlimbach, A.A. Teleman, Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence, Sci. Rep. 8 (1) (2018) 4018.
[20] U. Thorsteinsdottir, E. Kroon, L. Jerome, F. Blasi, G. Sauvageau, Defining roles for HOX and MEIS1: genes in induction of acute myeloid leukemia, Mol. Cell. Biol. 21 (1) (2001) 224–234.
[21] S.M. Gough, C.I. Slape, P.D. Aplan, NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights, Blood 118 (24) (2011) 6247–6257.
[22] J.E. Kroll, S.J. de Souza, G.A. de Souza, Identification of rare alternative splicing events in MS/MS data reveals a significant fraction of alternative translation initiation sites, PeerJ 2 (2014).

[23] E.D. Karousis, D.C. Sideris, A subtle alternative splicing event gives rise to a widely expressed human RNase k isoform, PLoS One 9 (5) (2014) e96557.
[24] L. Soares, C. Seroogy, H. Skrenta, N. Anandasabapathy, P. Lovelace, C.D. Chung, et al., Two isoforms of otubain 1 regulate T cell anergy via GRAIL, Nat. Immunol. 5 (2003) 45.

Jo-Wai D. Wang[a,b], Gabrielle L. Gross[a,b], Erin R. Andrew[a,b]
[a] *Research School of Biology, The Australian National University, Acton, ACT 2601, Australia*
[b] *Medical School, The Australian National University, Garran, ACT 2605, Australia*

Amanda Padovan[a,b]
[a] *Research School of Biology, The Australian National University, Acton, ACT 2601, Australia*
[b] *CSIRO, Black Mountain Science and Innovation Park, Canberra, ACT 2601, Australia*

Aude M. Fahrer*
Research School of Biology, The Australian National University, Acton, ACT 2601, Australia
E-mail address: aude.fahrer@anu.edu.au.

* Corresponding author at: Building 134, The Australian National University, Acton, ACT 2601, Australia.