Blood from ‘junk’: the LTR chimeric transcript Pu.2 promotes erythropoiesis

Kyle R Upton1 and Geoffrey J Faulkner1,2*

Abstract
Transposable elements (TEs) are a prominent feature of most eukaryotic genomes. Despite rapidly accumulating evidence for the role of TE-driven insertional mutagenesis and structural variation in genome evolution, few clear examples of individual TEs impacting biology via perturbed gene regulation are available. A recent report describes the discovery of an alternative promoter for the murine erythroid transcription factor Pu.1. This promoter is located in an ORR1A0 long terminal repeat (LTR) retrotransposon intronic to Pu.1 and is regulated by the Krüppel-like factors KLF1 and KLF3. Expression of the resultant chimeric transcript, called Pu.2, spontaneously induces erythroid differentiation in vitro. These experiments illustrate how transcription factor binding sites spread by retrotransposition have the potential to impact networks encoding key biological processes in the host genome.

Keywords: LTR retrotransposon, Transposable element, ORR1A0, KLF1, KLF3, Erythroid, Pu.1, Pu.2

Background
Transposable elements (TEs) have in the past been maligned as ‘junk’ [1], ‘selfish,’ and ‘parasitic’ [2,3]. These descriptions are likely apt for many if not most sequences derived from TEs. However, a more complete view is that the majority of DNA generated by TE activity evolves neutrally under selection, while a small yet important minority of TE-derived sequences continues to drive genome evolution and innovation. The potential importance of TEs to the host is suggested by their near ubiquitous presence in eukaryotes, often accounting for half or more of genome sequence content [4-6] and, more convincingly, their provision of regulatory or otherwise functional genetic elements [7]. In primates, for example, most order-specific regulatory sequences are derived from TEs [8]. Many of these regions are dynamically regulated during development [9,10] and incorporate internal binding sites for suppressor and activator complexes [11].

As a TE proliferates in its host genome, the number of loci subject to regulation by DNA binding proteins specific to that TE, such as transcription factors, also increases. During evolution, TEs have often contributed transcription factor binding sites to promoter, enhancer, and boundary elements [9-14]. In some cases, entire developmental pathways have been rewired as a result of TE mobilization [15-17]. Novel TE insertions can result in alternative splicing [18], exonization [19,20], altered mRNA translational efficiency [21,22], as well as the provision of distal enhancers [23]. Thus, an abundance of TE-derived sequences, including ready-made promoters, enhancers, and other regulatory units, points to a major role for TEs in shaping the regulatory landscape of the eukaryotic genome (see [24] for a recent review).

Discovery and characterization of Pu.2, an LTR-driven chimeric mRNA
Despite extensive evidence for gene regulatory and structural innovation produced by TEs, examples of phenotypic change due to this variation are comparatively limited in mammals. TE-derived alternative promoters, which generate a chimeric mRNA with an adjacent gene, are arguably one of the more straightforward scenarios to link a TE with a functional product, particularly when that gene encodes a protein of known function. In recent work, Mak et al. [25] report the discovery and functional characterization of a long terminal repeat (LTR) promoted chimeric mRNA of Pu.1, regulated in turn by Krüppel-like factors 1 (KLF1) and 3 (KLF3). Pu.1, KLF1, and KLF3 are transcription factors active during hematopoiesis, where Pu.1 favors...
myeloid differentiation and KLF1 and KLF3 drive erythroid maturation [26,27]. In wild-type mice, KLF1 and KLF3 recognize similar sequence motifs in an antagonistic manner. KLF1 generally acts as a transcriptional activator [27], while KLF3 recruits a repressive complex including epigenetic modifiers [28].

Analyzing a microarray screen of Klf3−/− knockout mice, Mak et al. first identified major de-repression of Pu.1 that, oddly, excluded the initial two exons of the gene. To resolve this discrepancy, the authors performed 5′ RACE primed from the third exon of Pu.1. The results of this approach indicated an alternative promoter in an ORR1A0 LTR [29] located in the second intron of Pu.1 (Figure 1). Subsequent qRT-PCR assays confirmed that the ORR1A0-Pu.1 chimeric transcript, named Pu.2, was upregulated in Klf3−/− fetal liver tissue, while KLF1 and KLF3 were found to bind sequence motifs contained in ORR1A0, based on electrophoretic mobility shift assay (EMSA) and luciferase reporter experiments. Taken together, these data suggest opposing roles for KLF1 and KLF3 in regulating Pu.2 expression.

Intriguingly, Pu.2 was predicted to encode an N-terminal truncated isoform of PU.1 that retained a DNA binding ETS domain (Figure 1). By EMSA, Mak et al. demonstrated that PU.2 could bind to a predicted Pu.1 target site. Moreover, PU.2 was found to counter the activity of Pu.1 in a dose-dependent manner, and promoted spontaneous erythroid differentiation in human K562 cells. These assays clearly demonstrated, in vitro, an erythropoietic function for Pu.2. Extending their results to an in vivo setting, the authors detected Pu.2 mRNA and protein in wild-type fetal liver, with overexpression observed in Klf3−/− animals, and also found in vitro KLF1 overexpression activated PU.2. These experiments provide evidence for PU.2 function in murine erythroid differentiation, albeit heavily repressed by KLF3 in vivo.

Finally, via RNA-seq and additional qRT-PCR, Mak et al. identified several other ORR1A0 alternative promoters differentially regulated by KLF3 and producing chimeric transcripts with adjacent protein-coding genes. This important, though preliminary, observation suggested that ORR1A0 may play a broader role in regulating erythroid differentiation beyond the highlighted example of Pu.2, and provides insight into the co-evolution of TE subfamilies, transcription factors, and core biological processes, as discussed elsewhere [30]. One reasonable conclusion from this work is that the amplification of the ORR1A0 LTR family in rodents generated a ready-made network of genetic material subject to control by KLF1 and KLF3, and capable of changing how erythroid maturation was regulated during development. Notably, the ORR1A0 LTR family is rodent-specific and almost certainly incapable of further mobilization [29]. As such, Pu.2 is not found in human cells, though it remains to be determined whether other TEs present in the human Pu.1 locus generate mRNAs functionally analogous to mouse Pu.2. Future experiments involving genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) to elucidate KLF bound sites in vivo may reveal human TE families dynamically regulated hematopoiesis, in the Pu.1 locus and elsewhere.

Conclusions
Among a host of alternative promoters derived from mammalian TEs and driving protein-coding and non-coding gene expression [9,31-33], we consider three reports as

---

**Figure 1** A potential role for Pu.2 in erythroid differentiation, as described by Mak et al. PU.1 is a master regulator of myeloid differentiation (top). KLF1 and KLF3, respectively, activate and repress an alternative promoter included in an ORR1A0 LTR located in the second intron of Pu.1. KLF3-mediated suppression of this LTR is the norm in wild-type fetal liver cells, permitting canonical PU.1 expression and myeloid differentiation. In the absence of KLF3, Mak et al. found the ORR1A0 LTR produced a chimeric mRNA with Pu.1 that, in turn, encoded a truncated protein isoform called Pu.2 (bottom) lacking the N-terminal activation domain of Pu.1. Pu.2 retained its DNA binding capacity, but did not interact with other transcription factors, leading to a putative role as a dominant negative isoform of Pu.1 promoting erythroid maturation.
landmark examples of TE insertions having a clear functional impact upon biology: an epigenetically regulated LTR upstream of the \textit{Agouti} gene in rodents [34], an LTR alternative promoter for the colony stimulating factor 1 receptor (CSF1R) proto-oncogene in lymphoma [35], and the convergent evolution of multiple LTRs to act as promoters for the neuronal apoptosis inhibitory protein (NAIP) gene [36]. In each case, an LTR produces a chimeric mRNA with the adjacent protein-coding gene. The discovery by Mak et al. that the LTR-initiated \textit{Pu.2} transcript can promote erythroid maturation in the absence of KLF3 is a valuable addition to this literature, and will likely increase future attention to the role of TEs in regulating various developmental processes, including hematopoiesis.

Abbreviations

\textit{Bra2}: Breast cancer 2, early onset; ChIP-seq: Chromatin immunoprecipitation sequencing; CSF1R: Colony stimulating factor 1 receptor; EMSA: Electrophoretic mobility shift assay; Kif: Krüppel-like factor; LTR: Long terminal repeat; NAIP: Neuronal apoptosis inhibitory protein; qRT-PCR: quantitative real time PCR; RACE: Rapid amplification of cDNA ends; RNA-seq: RNA sequencing; CSF1R: Colony stimulating factor 1 receptor; Klf: Krüppel-like factor; LTR: Long terminal repeat; NAIP: Neuronal apoptosis inhibitory protein; qRT-PCR: quantitative real time PCR; RACE: Rapid amplification of cDNA ends; RNA-seq: RNA sequencing; TE: Transposable element.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GJF wrote the manuscript. KRU provided edits and prepared Figure 1. Both authors read and approved the final manuscript.

Acknowledgements

We thank Sandra Richardson and Adam Ewing for helpful editorial comments. GJF acknowledges the support of an Australian NHMRC Career Development Fellowship (GNT1045237), NHMRC Project Grants GNT1042449 and GNT1068789 and the European Union’s Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 259743 underpinning the MODHEP consortium. KRU and GJF are additionally supported by NHMRC Project Grants GNT1045991 and GNT1067983. Funding bodies played no role in the preparation of, or decision to submit, the manuscript.

Received: 2 April 2014 Accepted: 2 May 2014

Published: 9 May 2014

References

1. Ohno S: So much “junk” DNA in our genome. Brookhaven Symp Biol 1972, 23:366–370.

2. Doolittle WF, Sapirza C: Selfish genes, the phenotype paradigm and genome evolution. Nature 1980, 284:601–605.

3. Orgel LE, Crick FH: Selfish DNA: the ultimate parasite. Nature 1980, 284:604–607.

4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Etalland R, MacVean G, McCombie WR, Mardis ER, Marth TG, McMichael FF, Miller CA, Murray JF, Nickerson DA, O’Roak BJ, Reich M, Yen IF, et al: Initial sequencing and analysis of the human genome. Nature 2001, 409:860–921.

5. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Alkinyo R, Alexandersson M, An P, Antorcanakis SE, Atwood T, Baetser R, Bailey J, Barlow K, Beck S, Berry E, Birken B, Bloom T, Bork P, Botcherby M, Bray N, Brent M, Brown GD, Brown S, Burt C, Burton J, Butler J, Campbell RD, et al: Initial sequencing and analysis of the human genome. Nature 2001, 409:860–921.

6. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pustenak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Cournot L, Kuchowski SS, Tomlinson C, Strong C, Deleashbury K, Friconic C, Cournot B, Rock SM, Belter E, Du F, Kim K, Abbotti RM, Cotton M, Levy A, Marchetti P, Ochoa K, Jackson SM, Gillam B, et al: The B73 maize genome: complexity, diversity, and dynamics. Science 2009, 326:1112–1115.

7. Britten RJ, Davidson EH: Gene regulation for higher cells: a theory. Science 1969, 165:349–357.

8. Jacques PE, Jeyakani J, Bourque G: The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet 2013, 9:e1003504.

9. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schröder K, Cloonan N, Stepole AL, Lassmann T, Waki K, Horning N, Arakawa T, Takashashi H, Kawai J, Forrest AR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, Grimmond SM, Caminci P: The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 2009, 41:563–571.

10. Lu X, Sachs F, Ransay L, Jacques PE, Goke J, Bourque G, Ng HH: The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol 2014, 21:423–425.

11. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET: Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 2008, 18:1752–1762.

12. Lunyak V, Prefontaine GG, Nunez E, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, Garcia-Diaz A, Zhu X, Yung Y, Montillo L, Glass CK, Rosenfeld MG: Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 2007, 317:248–251.

13. Schmidt D, Schwilke PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT: Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 2012, 148:335–348.

14. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D: Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci U S A 2010, 107:18613–18618.

15. Kunarso G, Chia NY, Jeyakani J, Bourque G: Contributions of retrotransposons by a sequence-specific transcription factor. Nat Genet 2006, 38:859–864.

16. Markus E, Rimmer AN, Bafna V, Brudno M: Alu-containing exons are alternatively spliced. Genome Res 2002, 12:1063–1067.

17. Nekrutenko A, Li WH: Transposable elements are found in a large number of human protein-coding genes. Trends Genet 2001, 17:519–521.

18. Carrieri C, Cimatti L, Biagioli M, Brugnera A, Zuccelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Caminci P, Biffi S, Stupka E, Gustinich S: Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 2012, 491:454–457.

19. Smallheiser NR, Torvik VI: Alu elements within human mRNAs are probable microRNA targets. Trends Genet 2006, 22:532–536.

20. Sasaki T, Nishihara H, Haikuma K, Tanaka M, Kobuk K, Kimura-Yoshida C, Matsuo I, Suyayama K, Satou N, Shimogori T, Okada N: Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci U S A 2008, 105:4220–4225.

21. Rebollo R, Romish-Miller MT, Mager D: Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 2012, 46:241–262.

22. Mak KS, Burdach J, Norton LJ, Pearson RC, Crossley M: Funnel APW: Repression of chimeric transgenes emanating from endogenous retrotransposons by a sequence-specific transcription factor. Genome Biol 2014, In Press.

23. Pearson R, Fleetwood J, Eaton S, Crossley M, Bao S: Kruppel-like transcription factors: a functional family. Int J Biochem Cell Biol 2008, 40:1996–2001.

24. Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Coonan N, Grimmond SM, Bailey TL, Perkins AC: A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome Res 2010, 20:1052–1063.
28. Turner J, Crossley M: Cloning and characterization of mCtBP2, a co-repressor that associates with basic Kruppel-like factor and other mammalian transcriptional regulators. EMBO J 1998, 17:5129–5140.

29. Smit AF: Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res 1993, 21:1863–1872.

30. Thomas JH, Schneider S: Coevolution of retroelements and tandem zinc finger genes. Genome Res 2011, 21:1800–1812.

31. Huda A, Bowen NJ, Conley AB, Jordan IK: Epigenetic regulation of transposable element derived human gene promoters. Gene 2011, 475:39–48.

32. Kapusta A, Kionenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C: Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 2013, 9:e1003470.

33. Speek M: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 2001, 21:1973–1985.

34. Morgan HD, Sutherland HG, Martin D, Whitelaw E: Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999, 23:314–318.

35. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenzé D, Kochert K, Bouhlet MA, Richter J, Soler E, Stadhouders R, Johrens K, Wurster KD, Callen DF, Harte MF, Gießing M, Barlow R, Stein H, Anagnostopoulos L, Ianz M, Cockerill PN, Siebert R, Dorken B, Bonifer C, Mathas S: Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med 2010, 16:571–579. 1p following 579.

36. Romanish MT, Lock WM, van de Lagemaat LN, Dunn CA, Mager DL: Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet 2007, 3:e10.

doi:10.1186/1759-8753-5-15

Cite this article as: Upton and Faulkner: Blood from ‘junk’: the LTR chimeric transcript Pu.2 promotes erythropoiesis. Mobile DNA 2014 5:15.