Mobile Genetic Elements 2:2, 106–109; March/April 2012; © 2012 Landes Bioscience

Positive selection of co-opted mobile genetic elements in a mammalian gene

If you can't beat them, join them

Lucía F. Franchini,[1] Flavio S.J. de Souza,[2] Malcolm J. Low[3] and Marcelo Rubinstein[4,*]

[1]Instituto de Investigaciones en Ingeniería Genética y Biología Molecular; Consejo Nacional de Investigaciones Científicas y Técnicas; Buenos Aires, Argentina; [2]Instituto de Investigaciones en Ingeniería Genética y Biología Molecular; Consejo Nacional de Investigaciones Científicas y Técnicas and Facultad de Ciencias Exactas y Naturales; Universidad de Buenos Aires; Buenos Aires, Argentina; [3]Department of Molecular and Integrative Physiology; University of Michigan; Ann Arbor, MI USA; [4]Instituto de Investigaciones en Ingeniería Genética y Biología Molecular; Consejo Nacional de Investigaciones Científicas y Técnicas and Facultad de Ciencias Exactas y Naturales; Universidad de Buenos Aires; Buenos Aires, Argentina

The *proopiomelanocortin* (*Pomc*) gene encodes a prepropeptide with essential functions in the response to stress and energy balance, which is expressed in the pituitary and hypothalamus of vertebrate animals. Neuronal expression of *Pomc* is controlled by two distal enhancers named nPE1 and nPE2. Using transgenic mice, we observed that both enhancers drive identical expression patterns in the mammalian hypothalamus, starting at embryonic day 10.5, when endogenous *Pomc* expression commences. This overlapping enhancer activity is maintained throughout hypothalamic development and into adulthood. We also found that nPE1 and nPE2 were exapted as neuronal enhancers into the *POMC locus* after the sequential insertion of two unrelated retroposons. Thus, nPE1 and nPE2 are functional analogs and represent an authentic first example of convergent molecular evolution of cell-specific transcriptional enhancers. In this Commentary we discuss the following questions that remain unanswered: (1) how does transcriptional control of POMC operate in hypothalamic neurons of non-mammalian vertebrates? (2) What evolutionary forces are maintaining two discrete neuronal POMC enhancers under purifying selection for the last ~100 million years in all placental mammals? (3) What is the contribution of MaLRs to genome evolution?

Keywords: exaptation, co-option, MaLR, CORE-SINE, convergent evolution, convergent molecular evolution, enhancer, proopiomelanocortin, pomc, in silico paleogenomics

Submitted: 01/23/12

Accepted: 04/04/12

http://dx.doi.org/10.4161/mge.20267

*Correspondence to: Marcelo Rubinstein; Email: mrubins@dna.uba.ar

Commentary to: Franchini LF, López-Leal R, Nasif S, Beati P, Gelman DM, Low MJ, et al. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. Proc Natl Acad Sci U S A 2011; 108:15270–5; PMID:21876128; http://dx.doi.org/10.1073/pnas.1104997108

The proopiomelanocortin gene (*POMC*) and its encoded peptides constitute a formidable theater of operations used by multiple evolutionary forces that shaped an adaptive polyfunctional toolkit that has increased the fitness of all vertebrate animals when exposed to environmental stress. Several unrelated peptide-coding sequences were recruited more than 400 million years ago (MYA) to assemble a modular prohormone, reminiscent of the polycystronic organization of prokaryotic genes. Targeted endopeptidase cleavage of the POMC prohormone gives rise to cell-specific combinations of the bioactive melanocortins α-, β- and γ-MSH, ACTH and the opioid peptide β-endorphin (**Fig. 1A**), which together coordinate the stress-response program.[1] *POMC* is mainly expressed in pituitary corticotrophs and melanotrophs and in a group of neurons present in the arcuate nucleus of the hypothalamus.

The tissue-specific control of *POMC* expression also follows a modular organization with an array of transcriptional enhancers, particularly identified in mammals, that are scattered along the *POMC* 5' flanking region. Pituitary specific *POMC* expression depends on the concerted action of proximal sequences located within 400 bp upstream of the transcription start site and a distal enhancer module,[2] whereas neuronal expression is independently controlled by cis-regulatory elements located further upstream.[3,4]

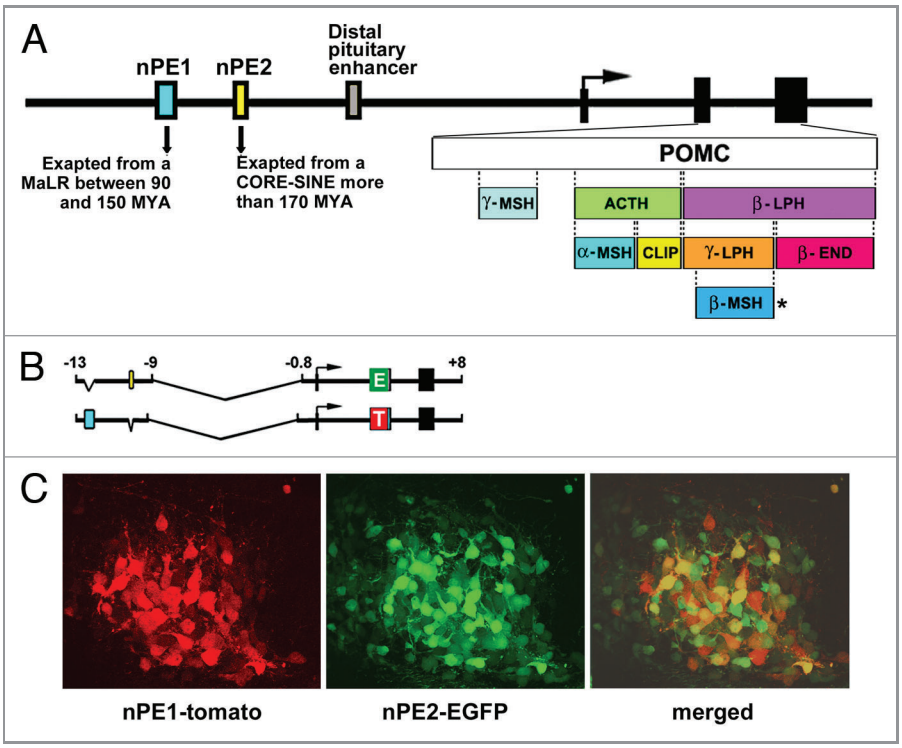

Figure 1. (A) Schematic of the mouse *POMC* gene. The exapted neuronal enhancers nPE1 and nPE2, and the distal pituitary enhancer are color-boxed. *Pomc* exons are in black boxes. Coding sequences for the POMC prohormone and the peptides obtained after endoproteolytic cleavage are indicated in colored boxes. The asterisk next to β-MSH denotes that this peptide is not released in mice. (B) Schematic of the transgenes nPE1Pomc–tomato and nPE2Pomc-EGFP transgenes described in Ref. 6. (C) Expression of tomato and EGFP in the developing hypothalamus of e13.5 compound transgenic mouse embryos showis extensive colocalization of both reportersin neurons.

A combination of genome comparison and expression studies in transgenic mice led us to the identification of two mammalian conserved **n**euronal *POMC* **e**nhancers, nPE1 and nPE2, that are able to independently drive reporter gene expression to POMC hypothalamic neurons when placed upstream of a minimal heterologous promoter.[3] We also demonstrated that only the concurrent removal of nPE1 and nPE2 from transgenic constructs leads to the loss of reporter gene expression in POMC hypothalamic neurons.[3] More recently, we found that nPE1 and nPE2 were exapted as neuronal enhancers into the *POMC locus* after the sequential nearby insertion of two unrelated retroposons. nPE2 originated from the earlier exaptation of a CORE-SINE retroposon[5] in the lineage leading to mammals sometime before the Prototheria/Metatheria split (~170 MYA) whereas nPE1 is a placental novelty that originated from a Mammalian-apparent LTR (MaLR) retroposon after the Metatheria/Eutheria split (~150 MYA) and before the wide mammalian radiation that occurred 90 MYA.[6] Thus, *POMC* represents one example of probably thousands of vertebrate genes that evolved following a "foreign legion" strategy through the recruitment of unrelated components that, once assembled into a functional gene, acquired a unified "esprit de corps" that contributes to improve fitness (**Fig. 1A**).

A special case is reserved for teleost fishes, because they have two paralog copies of *POMC* (α and β) that originated after a whole-genome duplication that occurred 320 MYA in the teleost lineage.[7] POMCα and POMCβ underwent a subfunctionalization process during teleost evolution through the partitioning of their expression domains and respective peptide repertoires which has maintained both paralogs under purifying selection.[7] POMCα retained its expression in the nucleus lateralis tuberis (homolog fish structure to the arcuate nucleus), but how the two paralogs regulate their individual expression patterns and coordinate their physiological functions remains to be studied.

To determine whether nPE1 and nPE2 control neuronal POMC expression in overlapping or complementary spatio-temporal domains within the developing and adult arcuate nucleus we generated transgenic mice expressing the fluorescent markers tomato and EGFP under the transcriptional control of nPE1 or nPE2, respectively (**Fig. 1B**). Analysis of compound transgenic mice demonstrated co-expression of both reporter genes in neurons along the entire arcuate nucleus from the earliest onset of POMC expression at e10.5.[6] Similar results were obtained at all later embryonic and postnatal stages examined (**Fig. 1C**). Altogether, these results demonstrated that nPE1 and nPE2 are two functional analogs and represent an authentic first example of convergent molecular evolution of cell-specific enhancers.[6]

Although our studies have provided several key insights about the *cis*-acting elements controlling mammalian *POMC* expression, several questions remain unanswered:

(1) How does transcriptional control of *POMC* operate in hypothalamic neurons of non-mammalian vertebrates? *POMC* is expressed in neurons of the mediobasal hypothalamus of all jawed vertebrates examined to date including mammalian, avian, reptile, amphibian and teleost fish species. However, the neuronal enhancers nPE1 and nPE2 have been identified only in mammals, consistent with the evolutionary origin and fate of the exapted CORE-SINE (nPE2) and MaLR (nPE1) retroposons. Our efforts to identify nPE1 or nPE2 orthologs in non-mammalian Classes have been unproductive. Although the existence of an ancestral neuronal enhancer(s) of *POMC* seems probable, phylogenetic footprinting analyses to identify alternative conserved non-coding regions in the vicinities of *POMC* have consistently failed. However, the proposition that each vertebrate Class independently acquired convergent molecular mechanisms to transactivate *POMC* in arcuate neurons does not seem parsimonious. Thus, based on the

existing sequence comparison data among and within vertebrate Classes together with what we have learned about the evolutionary origin of nPE1 and nPE2 we favor the hypothesis that an ancestral neuronal *POMC* enhancer existed in the last common forerunner to all extant vertebrates. This ancestral enhancer that may still exist in a particular lineage has been dynamically reshuffled following lineage-specific mechanisms, which could have involved, as in mammals, the exaptation of mobile genetic elements of distinct origin. *POMC* is one of many genes in which protein-coding sequences have been under greater selective constraint during longer evolutionary periods than transcriptional enhancers which remain conserved only within specific branches. We believe that this differential mechanism between coding and non-coding sequences is key to illuminate the genetic basis of developmental repatterning and animal diversity.

(2) What evolutionary forces are maintaining two discrete neuronal POMC enhancers under purifying selection for the last ~100 million years in all placental mammals? The adaptive value of having two functionally overlapping enhancers, instead of just one, has been proposed for developmental genes[8] in light of the canalization theory originally elaborated by Waddington[9] and Schmalhausen[10] who put forward the idea that the precise and stable spatio-temporal phenotypes normally observed during embryogenesis are assured by duplicated pathways that reduce variability by buffering environmental perturbations. Recently, two groups independently demonstrated for the *D. melanogaster* genes *snail*[11] and *shavenbaby*[12] that overlapping (shadow) enhancers provide adaptive robustness to overcome suboptimal environmental conditions during fly development.[11,12] Although *Pomc* expression starts at embryonic day 10.5 in the mouse, it cannot be considered a classical developmental brain gene because all known functions of POMC-derived peptides are exerted after birth and POMC knockout mice or humans do not exhibit overt central developmental defects.[13,14]

Thus, nPE1 and nPE2 probably constitute the first example of overlapping (shadow) enhancers acting in a gene primarily involved in postnatal development and adult physiology. For a gene like *Pomc*, the adaptive value of having two apparently redundant enhancers may be related to the probability of increasing transcriptional rate and/or minimizing its variance. These two, not mutually exclusive, hypotheses may be illustrated by a double scull boat that not only has the capacity to propel faster but also to keep on going in the event that one rower wears out (acquisition of a deleterious mutation). Overlapping enhancers might have occurred as a compensatory evolutionary mechanism during the unicellular to multicellular transition to circumvent the increased vulnerability to inactivating mutations imposed by the emergence of numerous cell-specific enhancers controlling expression of every gene in different cell types and developmental trajectories. Our expression studies in transgenic mice showing that nPE1 and nPE2 drive reporter gene expression to an identical group of neurons along the entire space-time dimension of the mouse arcuate nucleus do not support the possibility that the two enhancers underwent a subfunctionalization fission process, as observed for enhancer paralogs of duplicated genes.[7,15,16] To the contrary, the independent recruitment of two evolutionary unrelated enhancers to jointly subserve neuronal Pomc expression may be viewed as a case of superfunctionalization, a term recently coined for the duplication of genes that control circadian rhythms in prokaryotes.[17,18] Different from these reported cases, nPE1 and nPE2 illustrate superfunctionalization of transcriptional mammalian enhancers. Mutant mice carrying targeted deletions of nPE1 and/or nPE2 will constitute a valuable tool to experimentally test the functional importance of each enhancer in the control of neuronal *Pomc* expression and the adaptive value of enhancer superfunctionalization in adult mice.

(3) What is the contribution of MaLRs to genome evolution? Non-autonomous retroposons are intracellular parasites capable of colonizing their host genomes in association with autonomous retrotransposon partners that provide the necessary "copy and paste" enzymatic machinery.[19] Host genomes, in turn, developed insightful strategies throughout evolution to prevent massive transposon insertions that would otherwise impair their genetic information flow. After a long-lasting evolutionary arms race, mobile elements and genomes have reached a tolerable cease-fire.[20] However, beneath the apparently calm surface of a Cold War in which negotiations and betrayals are routine, some transposable elements silently defect to play active roles in the host genome camp. This mechanism called exaptation has been observed in a handful of retroposons that, after acquiring selectable mutations during the course of evolution, became transcriptional enhancers of mammalian genes. The neuronal POMC enhancers nPE1 and nPE2 are two of the five known and functionally verified cases of earlier retroposons that were exapted into transcriptional enhancers.[5,6] The incipient list also includes an LF-SINE exapted into a developing brain enhancer of *Isl1*,[21] an AmnSINE1 that evolved into a brain enhancer of *Fgf8*[22] and the more recent discovery of another AmnSINE1 exapted as an enhancer of the transcription factor *Satb2* which participates in cortical brain development.[23] Although these examples are sparse, we believe they represent the tip of an emerging iceberg that will buttress the theory developed in the past two decades that mobile genetic elements contributed extensively to gene and genome evolution as proposed many years ago.[24-28]

Because the mechanisms developed by each phylogenetic branch to inactivate and/or eliminate retroposons have been unequally effective on different types of transposable elements, particular retroposon families have been able to resist genomic surveillance, retain transpositional activity and, therefore, colonize their host genomes in large copy numbers. For example, members of the CORE-SINE MAR1 family are very abundant, and even still active, in marsupial genomes.[6,29] Similarly, MaLR THE1B retroposons are present in high copy number in primate genomes suggesting that they were particularly active until recently. The species-specific or order-specific abundance of particular families of retroposons in combination with the development of a bioinformatic algorithm

that we named "in silico paleogenomics" has helped us to identify the evolutionary origin of nPE1[6] and nPE2.[5] This method, based on the Blast algorithm, incorporates an evolutionary model to the search strategy that allows the identification of lower identity homolog sequences such as mobile elements that were co-opted by the host genome after suffering numerous mutations and becoming DNA relics from the once active transposable elements.[5,6] We believe that the use of in silico paleogenomics, or other similar strategies, will allow the identification of the origin of many transcriptional enhancers that were exapted throughout evolution to provide adaptive complexity and richness to the regulation of gene expression.

Acknowledgments

This work was supported by an NIH grant DK068400 (M.J.L. and M.R.), International Research Scholar Grant of the Howard Hughes Medical Institute (M.R.), Agencia Nacional de Promoción Científica y Tecnológica, Argentina (M.R. and L.F.F.), CONICET (L.F.F.) and Universidad de Buenos Aires (M.R.), Argentina.

References

1. Raffin-Sanson ML, de Keyzer Y, Bertagna X. Proopiomelanocortin, a polypeptide precursor with multiple functions: from physiology to pathological conditions. Eur J Endocrinol 2003; 149:79-90; PMID: 12887283; http://dx.doi.org/10.1530/eje.0.1490079
2. Langlais D, Couture C, Sylvain-Drolet G, Drouin J. A pituitary-specific enhancer of the POMC gene with preferential activity in corticotrope cells. Mol Endocrinol 2011; 25:348-59; PMID:21193556; http://dx.doi.org/10.1210/me.2010-0422
3. de Souza FS, Santangelo AM, Bumaschny V, Avale ME, Smart JL, Low MJ, et al. Identification of neuronal enhancers of the proopiomelanocortin gene by transgenic mouse analysis and phylogenetic footprinting. Mol Cell Biol 2005; 25:3076-86; PMID: 15798195; http://dx.doi.org/10.1128/MCB.25.8.3076-3086.2005
4. Young JI, Otero V, Cerdán MG, Falzone TL, Chan EC, Low MJ, et al. Authentic cell-specific and developmentally regulated expression of pro-opiomelanocortin genomic fragments in hypothalamic and hindbrain neurons of transgenic mice. J Neurosci 1998; 18:6631-40; PMID:9712635
5. Santangelo AM, de Souza FS, Franchini LF, Bumaschny VF, Low MJ, Rubinstein M. Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet 2007; 3:1813-26; PMID:17922573; http://dx.doi.org/10.1371/journal.pgen.0030166
6. Franchini LF, López-Leal R, Nasif S, Beati P, Gelman DM, Low MJ, et al. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. Proc Natl Acad Sci U S A 2011; 108:15270-5; PMID:21876128; http://dx.doi.org/10.1073/pnas.1104997108
7. de Souza FS, Bumaschny VF, Low MJ, Rubinstein M. Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. Mol Biol Evol 2005; 22:2417-27; PMID:16093565; http://dx.doi.org/10.1093/molbev/msi236
8. Hobert O. Gene regulation: enhancers stepping out of the shadow. Curr Biol 2010; 20:R697-9; PMID: 20833307; http://dx.doi.org/10.1016/j.cub.2010.07.035
9. Waddington CH. Canalization of development and the inheritance of acquired characters. Nature 1942; 150: 563-5; http://dx.doi.org/10.1038/150563a0
10. Schmalhausen. Factors of Evolution: The Theory of Stabilizing Selection. Chicago: University of Chicago, 1949.
11. Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol 2010; 20:1562-7; PMID: 20797865; http://dx.doi.org/10.1016/j.cub.2010.07.043
12. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 2010; 466:490-3; PMID:20512118; http://dx.doi.org/10.1038/nature09158
13. Challis BG, Coll AP, Yeo GS, Pinnock SB, Dickson SL, Thresher RR, et al. Mice lacking pro-opiomelanocortin are sensitive to high-fat feeding but respond normally to the acute anorectic effects of peptide-YY(3-36). Proc Natl Acad Sci U S A 2004; 101:4695-700; PMID:15070780; http://dx.doi.org/10.1073/pnas.0306931101
14. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet 1998; 19:155-7; PMID:9620771; http://dx.doi.org/10.1038/509
15. Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff JN, et al. Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics 2002; 161:259-67; PMID:12019239
16. McClintock JM, Kheirbek MA, Prince VE. Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. Development 2002; 129:2339-54; PMID:11973267
17. Dvornyk V, Vinogradova O, Nevo E. Long-term microclimatic stress causes rapid adaptive radiation of kaiABC clock gene family in a cyanobacterium, Nostoc linckia, from "Evolution Canyons" I and II, Israel. Proc Natl Acad Sci U S A 2002; 99:2082-7; PMID: 11842226; http://dx.doi.org/10.1073/pnas.261699498
18. Dvornyk V, Vinogradova O, Nevo E. Origin and evolution of circadian clock genes in prokaryotes. Proc Natl Acad Sci U S A 2003; 100:2495-500; PMID: 12604787; http://dx.doi.org/10.1073/pnas.0130099100
19. Goodier JL, Kazazian HH, Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 2008; 135:23-35; PMID:18854152; http://dx.doi.org/10.1016/j.cell.2008.09.022
20. Lisch D, Slotkin RK. Strategies for silencing and escape: the ancient struggle between transposable elements and their hosts. Int Rev Cell Mol Biol 2011; 292:119-52; PMID:22078960; http://dx.doi.org/10.1016/B978-0-12-386033-0.00003-7
21. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, et al. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 2006; 441:87-90; PMID:16625209; http://dx.doi.org/10.1038/nature04696
22. Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, et al. Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci U S A 2008; 105:4220-5; PMID: 18334644; http://dx.doi.org/10.1073/pnas.0709398105
23. Tashiro K, Teissier A, Kobayashi N, Nakanishi A, Sasaki T, Yan K, et al. A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born callosal projection neurons. PLoS One 2011; 6:e28497; PMID:22174821; http://dx.doi.org/10.1371/journal.pone.0028497
24. Britten RJ. Mobile elements inserted in the distant past have taken on important functions. Gene 1997; 205:177-82; PMID:9461392; http://dx.doi.org/10.1016/S0378-1119(97)00399-5
25. Brosius J. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica 1999; 107:209-38; PMID:10952214; http://dx.doi.org/10.1023/A:1004018519722
26. Brosius J. The contribution of RNAs and retroposition to evolutionary novelties. Genetica 2003; 118: 99-116; PMID:12868601; http://dx.doi.org/10.1023/A:1024141306559
27. McDonald JF. Evolution and consequences of transposable elements. Curr Opin Genet Dev 1993; 3:855-64; PMID:8118210; http://dx.doi.org/10.1016/0959-437X(93)90005-A
28. McDonald JF. Transposable elements: possible catalysts of organismic evolution. Trends Ecol Evol 1995; 10:123-6; PMID:21236980; http://dx.doi.org/10.1016/S0169-5347(00)89012-6
29. Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD, et al. Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. Genome Res 2007; 17:992-1004; PMID:17495012; http://dx.doi.org/10.1101/gr.6070707