Review

Cell PRESS

Transposable elements as genetic regulatory substrates in early development

Wesley D. Gifford[1], Samuel L. Pfaff[1], and Todd S. Macfarlan[2]

[1] Gene Expression Laboratory and the Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, CA 92037, USA
[2] Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Building 6B, Room 2B-211, Bethesda, MD 20892, USA

The abundance and ancient origins of transposable elements (TEs) in eukaryotic genomes has spawned research into the potential symbiotic relationship between these elements and their hosts. In this review, we introduce the diversity of TEs, discuss how distinct classes are uniquely regulated in development, and describe how they appear to have been coopted for the purposes of gene regulation and the orchestration of a number of processes during early embryonic development. Although young, active TEs play an important role in somatic tissues and evolution, we focus mostly on the contributions of the older, fixed elements in mammalian genomes. We also discuss major challenges inherent in the study of TEs and contemplate future experimental approaches to further investigate how they coordinate developmental processes.

Introduction

TEs comprise a substantial fraction of eukaryotic genomes; they account for over forty times the nucleotide content as protein-coding exons [1], equating to nearly half of the human genome [2]. Recent data suggest that the percentage may even be closer to two-thirds [3]. This fact re-raises a long-standing question: do TEs serve any functional role for their host? Although there is a considerable literature on this topic [4–7], there remain few direct data demonstrating a requirement for TEs during embryo development. This review discusses research focusing on recent evidence that TEs function in numerous early embryonic developmental processes. We start with a brief review of the various classes of eukaryotic TEs and the general types of processes that have lead to the genomes we see today.

Snapshot of mammalian genomic TEs

TEs are broadly characterized as either Class I retrotransposons or Class II DNA transposons (Figure 1a). Class II DNA transposons (Tc1/*mariner*) encode a transposase enzyme that excises the parental sequence and mediates reintegration in another location (Figure 1b). Hence, they use a 'cut and paste' mechanism and do not readily accumulate in copy number. We refer the reader to an insightful review on DNA transposons for further information [8]. By contrast, Class I retrotransposons are transcribed into an RNA intermediate that may then be reverse transcribed into DNA and reintegrated as an additional copy elsewhere in the genome. This 'copy and paste' mechanism explains the abundance of these elements.

Mammalian class I retrotransposons are further divided into two subdivisions defined by the presence or absence of flanking long terminal repeats (LTRs) (Figure 1b). Non-LTR retrotransposons include long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Both of these elements have transpositionally active subfamilies in human (LINE-1, *Alu*) and mouse (LINE-1, SINE-B1, SINE-B2), although SINEs require protein encoded by LINEs for retrotransposition [9–12]. LTR retrotransposons [including endogenous retroviruses (ERVs)] derive from infectious retroviruses that integrated in the germline [13,14]. They are thus present clonally in all cells of their host. Most ERVs do not encode a functional envelope protein and are incapable of horizontal transmission. ERVs are further categorized into three classes – I, II, or III – depending on the exogenous retrovirus genus they most closely resemble [15]. Numerous subtypes of ERVs are active transposons in mice, including intracisternal A-particles (IAPs) [16] and MuERV-L [17] (discussed below). At least one human subtype, HERV-K, is capable of producing intact viral particles in humans [18]. ERVs, however, are not the only endogenous viral elements (EVEs) embedded in animal genomes [19]. At least ten non-retroviral families have endogenous counterparts, some of which are transcribed and harbor open reading frames [20]. With this in mind, we expect to see future studies implicating non-retroviral EVEs in an array of important biological phenomena for their host.

The balance of selective forces acting on TEs is complex. Selection for active transposition has kept many subfamilies active for millions of years after their initial infection, whereas deleterious transposition events are rapidly purged from the genome [21]. Suspected TE-induced tumorigenic events have been reported [22], providing further evidence that new integration events are sometimes harmful to the host. One mechanism by which an

Corresponding author: Macfarlan, T.S. (todd.macfarlan@nih.gov, tmacfarl@hotmail.com).
Keywords: transposon; endogenous retrovirus; evolution; development; placentation; gene regulation.

0962-8924/$ – see front matter. Published by Elsevier Ltd. http://dx.doi.org/10.1016/j.tcb.2013.01.001 Trends in Cell Biology, May 2013, Vol. 23, No. 5

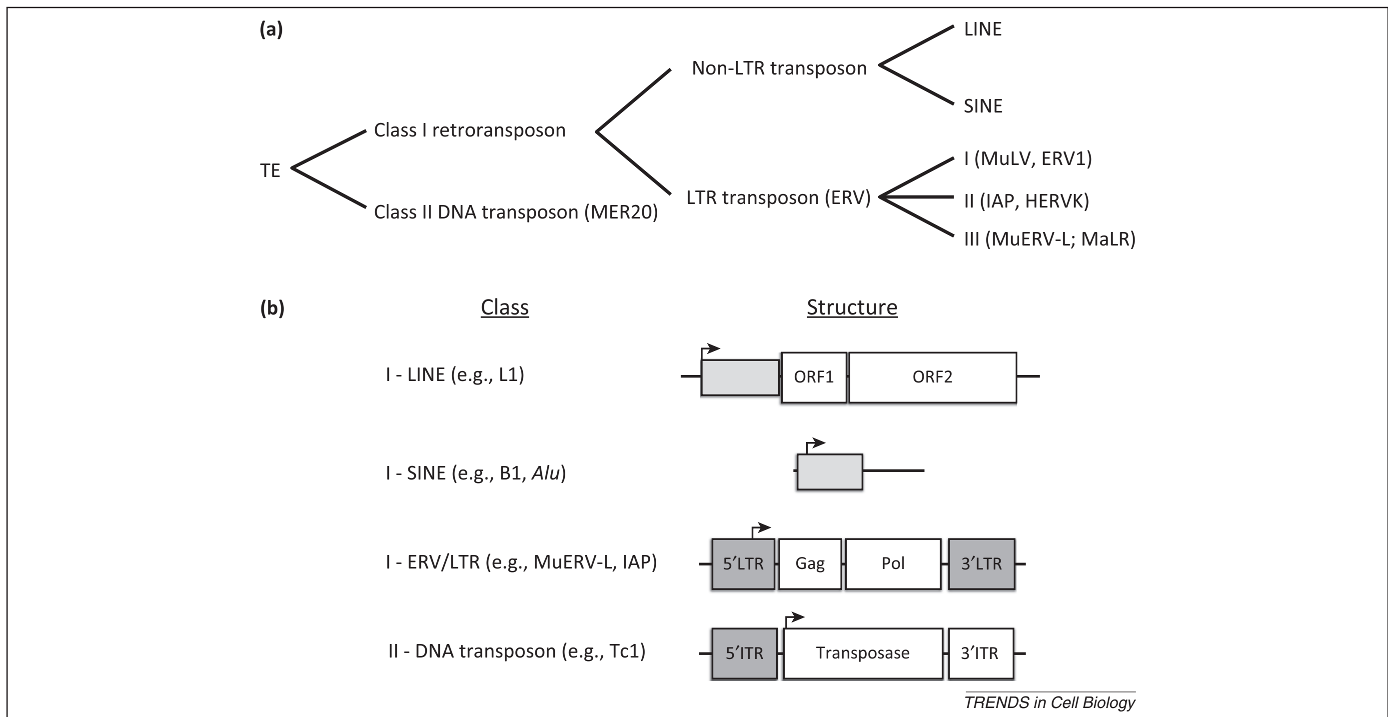

Figure 1. Classification of mammalian transposable elements (TEs). (**a**) TE are broadly categorized as DNA transposons or retrotransposons. Retrotransposons are further defined based on the presence of a long terminal repeat (LTR) that contains the element's regulatory information. Non-LTR retrotransposons comprise long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), whereas LTR retrotransposons (endogenous retroviruses [ERVs]) are further divided among three subclasses: I, II, and III. Examples of each subclass follow in parentheses. (**b**) The generic structures of major classes of TE are shown. LINE elements harbor a promoter sequence that is retained in the final transcript, shown in light grey, and two ORFs that mediate reverse transcription and transposition. SINE elements do not encode protein and rely on machinery encoded by LINE elements for their transposition. Most endogenous viral elements (EVEs) in mammalian genomes are endogenous retroviruses, some of which encode functional Gag and Pol proteins. More rarely, envelope proteins may also be present (not shown). These sequences are flanked by a direct LTR. Hence, these are named LTR-class retrotransposons. Most DNA transposons encode a transposase that cleaves and reinserts the parental element in a new location. The transposase is flanked by inverted terminal repeats (ITR).

ERV can be largely eliminated is through non-allelic homologous recombination between its two adjacent LTRs in *cis*, thereby excising the intervening sequence and leaving a single 'solo' LTR in its place. Solo LTRs, which outnumber full-length ERVs within mammalian genomes [23], can also be generated when two non-allelic LTRs recombine in *trans*. This process simultaneously leads to segmental duplications on the alternate chromosome. Some elements remain subject to neutral forces and slowly drift from their parental sequence and eventually lose the ability to transpose [21], whereas others remain intact for longer periods but are epigenetically silenced by various mechanisms. One such mechanism is DNA methylation, which accelerates the process of TE mutation through deamination of 5-methylcytosine residues to thymine [24–26]. The cumulative effects of these factors explain the genomes we see today: a heterogeneous mixture of elements – full length and truncated, active and broken, modern and ancient – distributed throughout the genome. Many of the resultant TE sequences are under strong purifying selection, despite being transcriptionally or transpositionally inactive, suggesting they retain function [27]. With this backdrop in mind, we will first highlight the mechanisms utilized by the host to keep TEs transcriptionally silent and then turn to the functional role that TEs play in their hosts.

Regulation of ERVs in the early mouse embryo

One area under active investigation is how host cells control the activity of TEs to prevent widespread retrotransposition. Recent evidence demonstrates that mouse cells utilize histone modification machinery to keep ERV subfamilies silenced in early embryos and embryonic stem (ES) cells, and that this silencing is independent of the DNA methylation machinery that is required for their silencing in somatic cells [15,28]. One such subfamily is the class III ERV MuERV-L, which belongs to an ancient endogenous retrovirus family termed ERV-L [29]. These ERV-L elements are likely to predate the placental mammal radiation (70 Mya) and are present in at least ten approximately full-length copies in all placentals examined to date, but generally do not encode intact viral proteins and, like most ERVs, do not harbor an *env* gene [17,29]. The exception to this includes ERV-L elements from mouse (MuERV-L), all extant simians (including humans [HERV-L]), and uranotherians (including elephants), which retain coding (mouse) or near-coding capacity (humans and uranotherians) for *gag* and *pol* genes [17,29–31]. The increased coding capacity is linked with the expansion of these elements in copy number in each of these lineages. In the case of the mouse, this expansion occurred after the mouse–rat divergence, because rats have ancestral levels of ERV-L [29] whereas mice contain roughly 350 fully coding elements per haploid genome [32]. The recent amplification and activity of these elements in the mouse lineage has necessitated tight control over transcription to prevent mutagenic events.

Many factors participate in the silencing of MuERV-L elements in ES cells and blastocyst-stage embryos,

Download English Version:

https://daneshyari.com/en/article/2204635

Download Persian Version:

https://daneshyari.com/article/2204635

Daneshyari.com