Seminars in Cell & Developmental Biology 57 (2016) 51–56

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

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

# Massive contribution of transposable elements to mammalian regulatory sequences

Nirmala Arul Rayan[a,1], Ricardo C.H. del Rosario[b,1], Shyam Prabhakar[a,*]

[a] Computational and Systems Biology, Genome Institute of Singapore, #02-01 Genome, 60 Biopolis Street, 138672, Singapore
[b] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, 75 Ames St., Cambridge, MA 02142, USA

ARTICLE INFO

*Article history:*
Received 20 January 2016
Received in revised form 6 May 2016
Accepted 6 May 2016
Available online 10 May 2016

*Keywords:*
Transposons
Regulatory elements
Primate-specific
Transposable element functionalization

ABSTRACT

Barbara McClintock discovered the existence of transposable elements (TEs) in the late 1940s and initially proposed that they contributed to the gene regulatory program of higher organisms. This controversial idea gained acceptance only much later in the 1990s, when the first examples of TE-derived promoter sequences were uncovered. It is now known that half of the human genome is recognizably derived from TEs. It is thus important to understand the scope and nature of their contribution to gene regulation. Here, we provide a timeline of major discoveries in this area and discuss how transposons have revolutionized our understanding of mammalian genomes, with a special emphasis on the massive contribution of TEs to primate evolution. Our analysis of primate-specific functional elements supports a simple model for the rate at which new functional elements arise in unique and TE-derived DNA. Finally, we discuss some of the challenges and unresolved questions in the field, which need to be addressed in order to fully characterize the impact of TEs on gene regulation, evolution and disease processes.

© 2016 Elsevier Ltd. All rights reserved.

**Contents**



## 1. Introduction

Retrotransposons are repetitive DNA elements that propagate in genomes via a copy-paste duplication mechanism – they are transcribed into RNA and then reverse transcribed into a DNA copy that is inserted at another genomic location. DNA transposons, on the other hand, generally move by cut-and-paste (translocation). Recognizable copies of transposable elements constitute a massive 48% of the human genome [1], and even this value is an underestimate, since the vast majority of repeats older than ~150 million years have accumulated such a large number of mutations that their origins are now unrecognizable.

Barbara McClintock discovered the existence of transposable elements (TEs) while studying the response of maize genomes to DNA damage in the 1940s. Since the elements she identified had modulated the activity of genes by transposing into flanking regions, she proposed that transposons were gene regulatory elements. Moreover, since the observed gene regulatory changes occurred at a specific stage during the development of a maize kernel, she speculated that orderly waves of transposition in somatic cells could provide a mechanism for developmental changes in gene expression [2,3]. Subsequently, Britten and Davidson extended this idea to evolutionary changes in gene regulation. They hypothesized that the existence of nearly identical repetitive elements near

* Corresponding author.
*E-mail address:* prabhakars@gis.a-star.edu.sg (S. Prabhakar).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.semcdb.2016.05.004
1084-9521/© 2016 Elsevier Ltd. All rights reserved.

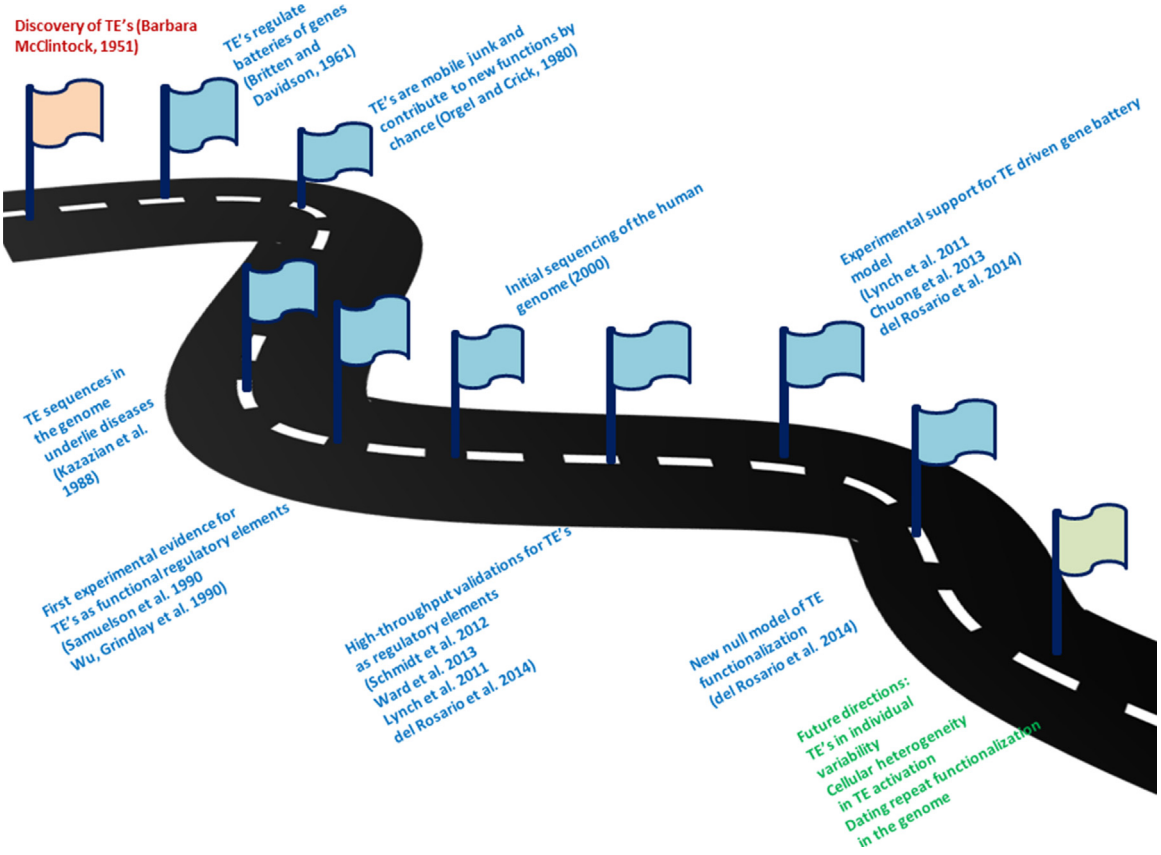


**Fig. 1.** Roadmap of transposon studies that have revolutionized our understanding of genome structure and function [63,64].

multiple functionally related genes was essential for their coordinated expression [4,5]. This "gene battery" hypothesis has been examined in numerous subsequent studies.

These early speculations on the functional role of transposons were rooted in the belief that more or less all DNA was functional, repetitive or not. However, this belief was soon displaced by the idea that the non-coding regions of genomes were mostly "junk DNA", i.e., DNA that served no specific organismal function. Along the same lines, TEs were seen as "selfish" or parasitic DNA elements, and it was thought that most of them existed for no other purpose than self-propagation [6–8]. Nevertheless, it was noted [7] that TEs could not be dismissed as entirely useless to the host: "Using the analogy of parasitism, slightly harmful infestation may ultimately be transformed into a symbiosis." Thus, the selfish-TE theory explicitly acknowledged the possibility of TEs becoming functional, though this was considered the exception rather than the norm. Of course, the same could be said of non-TE DNA. After all, the "birth rate" of new functional exons and *cis*-regulatory elements is unlikely to be high in any portion of the genome. This brings up a fundamental question: how quickly do new functions arise in TE-derived DNA, relative to non-TE DNA? If these two evolutionary birth rates are comparable, we should expect a massive number of TE-derived functional elements in the genomes of vertebrates.

In the discussion above, we have presented Orgel and Crick's conclusions in 1980 [7] as being entirely consistent with our current understanding of TE functionality. However, in the years immediately following the sequencing of human genome in 2000, it was actually common to discard TEs as being, by their very nature, non-functional. It is only over the last decade that the pendulum of scientific opinion has swung back to the recognition that they could indeed be functional, and that they do in fact contribute a massive number of functional elements to the human genome (Fig. 1). In this review, we discuss recent results demonstrating how TEs have contributed as a rich source of *cis*-regulatory elements in mammals, with special emphasis on primates.

## 2. TEs have a wide variety of functions in mammalian genomes

The early mammalian sequencing projects revealed a clear map of the abundance and distribution of TEs in the human and mouse genomes [9,10]. These and other sequencing efforts have led researchers to identify certain TE sequences that have been under functional constraint over the evolution of different mammalian species. One of the first studies looked at human and mouse orthologous constrained sequences and suggested that most repeat insertions occurred prior to the eutherian radiation [11]. Another pioneering study in 2003 showed that 25% of the experimentally characterized promoters were derived from ancient repeats [12]. Interest very soon picked up, and many ancient repeat elements in the human genome were shown to have evolved under strong purifying selection [13,14]. This provided strong evidence that TE-derived sequences could be functional and important for fitness. Confirming this, a SINE (short interspersed repetitive element)-derived developmental enhancer was identified [15] and shown to recapitulate the *in vivo* expression pattern of its target gene. Following this, there were other studies demonstrating SINEs as regulatory elements that altered gene regulation in the mammalian brain [16,17]. Studies that reported functional ancient repeats based on cross-species conservation increased in number as more mammalian species were sequenced. Analysis of the opossum genome highlighted that a substantial proportion of the eutherian-specific non-coding elements was TE-derived [18,19]. An even larger comparative analysis of 29 mammalian genomes revealed

Download English Version:

https://daneshyari.com/en/article/2202456

Download Persian Version:

https://daneshyari.com/article/2202456

Daneshyari.com