FEBS Letters xxx (2015) xxx–xxx

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Architectural hallmarks of the pluripotent genome

Britta A.M. Bouwman, Wouter de Laat *

Hubrecht Institute – KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 15 March 2015
Revised 24 April 2015
Accepted 27 April 2015
Available online xxxx

Edited by Wilhelm Just

Keywords:
Pluripotency
Genome architecture
Transcription
Enhancer clusters

ABSTRACT

Pluripotent stem cells (PSCs) have the ability to self-renew and are capable of generating all embryonic germ layers (Evans and Kaufman, 1981; Thomson et al., 1998). PSCs can be isolated from early embryos or may be induced via overexpression of pluripotency transcription factors in differentiated cells (Takahashi and Yamanaka, 2006). As PSCs hold great promise for regenerative medicine, the mechanisms underlying pluripotency and induction thereof are studied intensively. Pluripotency is characterized by a unique transcriptional program that is in part controlled by an exceptionally plastic regulatory chromatin landscape. In recent years, 3D genome configuration has emerged as an important regulator of transcriptional control and cellular identity (Taddei et al., 2004 [4]; Lanctot et al., 2007 [5]; Gibcus and Dekker, 2013; Misteli, 2009 [7]). Here we provide an overview of recent findings on the 3D genome organization in PSCs and discuss its putative functional role in regulation of the pluripotent state.

1. Introduction

Gene expression programs guide developmental decisions and underlie cell identity during all stages of development. Transcriptional activity is controlled by various factors including *trans*-acting chromatin and transcription factors, (distal) regulatory DNA elements, epigenetic decorations, and 3D chromatin organization [4–7]. As discussed in more detail below, the eukaryotic genome is stored in a compacted hierarchical fashion in the interphase cell nucleus (Fig. 1). Chromosomes occupy distinct nuclear sub-volumes that are called chromosome territories (CTs) [8]. Within a CT, along the linear chromosome axis, one can discern self-aggregating structural domains called topologically associated domains (TADs) [9–11]. These structural units serve as templates to accommodate physical contacts between genes and the cognate regulatory DNA elements that they encompass. At all levels of organization, genome architecture appears to be the result of a plethora of tissue-invariant and tissue-specific factors that compete for access to DNA to compact it or, oppositely, to expose sequences for reading, repairing and copying of the genetic code. Below, we first review current insight into the mechanisms that shape the genome and evaluate the functional implications of architecture at each topological level, starting at the sub-TAD level and gradually zooming out to higher-order genome structures. We next discuss the architectural specifics of the pluripotent 3D genome and elaborate on its significance for maintenance of the pluripotent state.

1.1. The dynamics and significance of enhancer–promoter contacts

Enhancers have emerged as important regulators of cell-specific gene expression patterns. Enhancers and other regulatory elements act on potentially distant target promoters via 3D chromatin contacts (Fig. 1), which can in some cases bridge distances of a megabase or more [12–15]. A forced enhancer–promoter loop was shown to be sufficient to induce recruitment of RNA Polymerase II and initiate transcription, even from a developmentally silenced gene although transcription elongation did not proceed at optimal rates [16,17]. The observation that loops persist when transcription is blocked [16,18,19] indicates that the process of transcription is not essential for maintenance of contacts, and that a different process is required to break up loops [20]. Taken together, current evidence suggests that enhancer–promoter loops form prior to and are required for efficient initiation of transcription. This is in contrast with elongation, the traversing of an RNA polymerase along the linear chromosome axis, which is likely not controlled at the 3D genome level.

The chromatin fiber behaves essentially like a polymer with certain flexibility when the effect of associated proteins is ignored. Chromatin loops are therefore likely to rely on random collisions between two sites and the further apart two sites sit on the linear chromosome, the less likely they are to autonomously contact each

* Corresponding author.
E-mail address: w.delaat@hubrecht.eu (W. de Laat).

http://dx.doi.org/10.1016/j.febslet.2015.04.055
0014-5793/

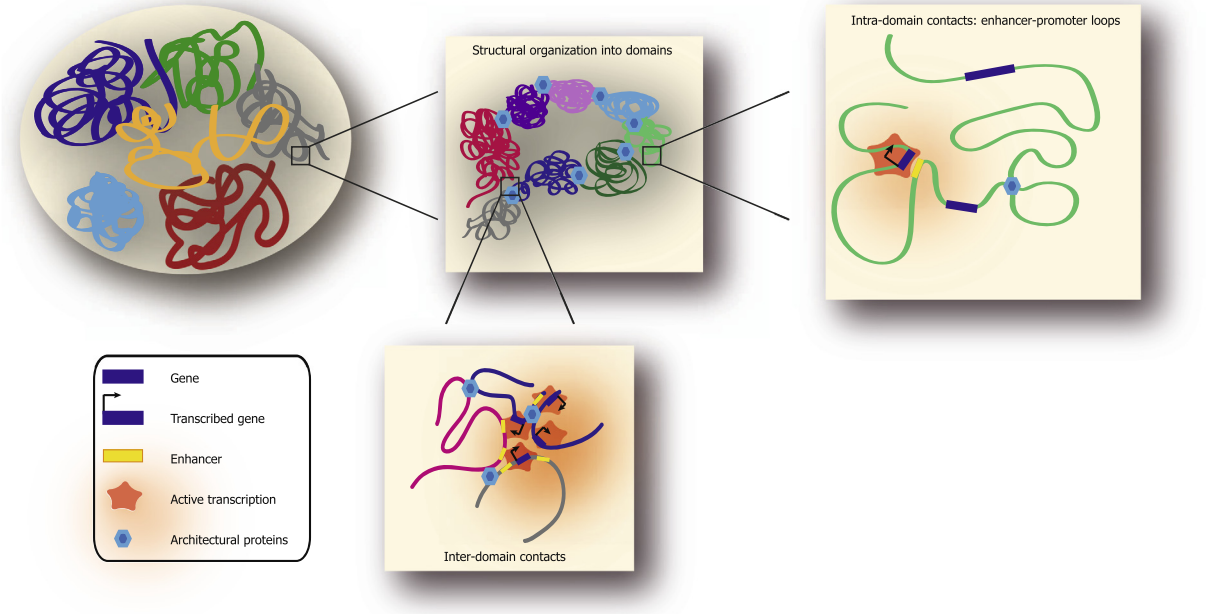

Fig. 1. Hierarchical levels of 3D genome organization. Interphase nuclei occupy distinct CTs (upper left). Within a CT, the chromosome is structurally organized into distinct TADs that are mostly demarcated by CTCF-associated TAD boundaries (upper middle). On a sub-TAD level regulatory elements such as enhancers sample the chromosome for compatible target genes that can be transcribed upon successful establishment of enhancer–promoter loops (upper right). Chromatin contacts between TADs that colocalize spatially have the ability to affect each other's transcription state (bottom).

other [21–26]. 3D contacts can stabilize when genome-associated proteins engage in protein-protein interactions [27]. Architectural proteins such as CTCF and cohesin [28–30] and general transcriptional co-activators such as Mediator and P300 [31,32], as well as more cell type-specific transcription factors [27,33,34] have been reported to be involved in loop formation and shaping of the genome. Enhancer–promoter contacts are frequently anchored by Mediator, cohesin, co-factor Nipbl, and (lineage-specific) transcription factors. These loop structures are proposed to be relatively dynamic during development, and are therefore considered important for regulation of key developmental genes [35–39].

Enhancer–promoter contacts correlate with but are not always sufficient to induce transcriptional activity; in some instances, they are believed to provide a spatial configuration that is poised for activation. Two distinct types of loops have been reported: pre-formed and *de novo* established loops, also referred to as permissive and instructive configurations, respectively [22]. The functional relevance of these differences in timing of loop formation is largely unknown and it is currently unclear whether these two configurations distinguish different categories of genes. Although the mechanisms that establish pre-formed loops require further investigation, they are speculated to facilitate rapid transcriptional activation: a permissive topology may optimally prime mammalian cells for a timely response to developmental stimuli [40–43]. Furthermore, pre-formed loops have been proposed to prevent bystander activation, via which unrelated neighboring genes can benefit from spurious contacts with unrelated regulatory DNA elements [11]. This contrasts with loops that are established de novo, presumably through the action of tissue- or lineage-specific transcription factors. These loops generally arise in a more tissue-specific manner at cell identity genes, which suggests a role in fate establishment [22,44].

Chromatin form generally precedes function and the 3D wiring of regulatory elements is assumed to coordinate cell type-specific expression patterns [16,45], appointing chromatin architecture as an integral feature of identity programming. Identity may be structurally safeguarded by the progressive formation of regulatory contacts required for later stages of lineage commitment, while 3D configurations required for earlier developmental stages are disrupted [46]. Recent work on the *Drosophila* genome revealed that only an estimated 6% of all identified enhancer–promoter interactions changes significantly during development. For the remaining 94% of loops, no dynamics in behavior were observed over time or between tissues, regardless of developmental transitions [47]. Based on these observations, it was proposed that enhancer-bound transcription factors assemble loops with target promoters, after which polymerase is recruited and maintained in a paused state. An additional cue, for example provided by recruited co-factors or looping of additional enhancers, may then trigger dispense of the paused state, allowing initiation and elongation of transcription [47]. High resolution Hi–C across a panel of human cell lines confirmed that many chromatin loops are conserved between cell types as well as during evolution, as evidenced from a comparison to Hi–C data generated in a mouse cell line. However, hundreds of tissue-specific loops between genes and enhancers were uncovered that corresponded almost exclusively with a highly increased transcriptional output of the contacted gene [44]. Thus, permissive and instructive configurations seem to co-exist in the genome to coordinate the faithful execution of cell-type specific transcriptional programs.

1.2. The functional importance of structural domains

The linear genome segregates into unit-like structural domains (Fig. 1) that are fairly conserved during differentiation and between mammalian species [9,10,44]. Initial Hi–C experiments revealed TADs with an estimated size of 1 Mb [9,11]. In the aforementioned more recent study, Hi–C experiments with increased sequencing depth and improved resolution allowed the appreciation of domain sizes ranging from 40 kb to 3 Mb, with a median of 185 kb [44]. TAD boundaries are enriched for CTCF-binding sites, housekeeping genes, short interspersed repeat elements, and tRNA

Please cite this article in press as: Bouwman, B.A.M. and de Laat, W. Architectural hallmarks of the pluripotent genome. FEBS Lett. (2015), http://dx.doi.org/10.1016/j.febslet.2015.04.055