Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

The role of gene regulatory factors in the evolutionary history of humans

Alvaro Perdomo-Sabogal[1,2], Sabina Kanton[1,2,a], Maria Beatriz C Walter[1,2,a] and Katja Nowick[1,2]

Deciphering the molecular basis of how modern human phenotypes have evolved is one of the most fascinating challenges in biology. Here, we will focus on the roles of gene regulatory factors (GRFs), in particular transcription factors (TFs) and long non-coding RNAs (lncRNAs) during human evolution. We will present examples of TFs and lncRNAs that have changed or show signs of positive selection in humans compared to chimpanzees, in modern humans compared to archaic humans, or within modern human populations. On the basis of current knowledge about the functions of these GRF genes, we speculate that they have been involved in speciation as well as in shaping phenotypes such as brain functions, skeletal morphology, and metabolic processes.

Addresses
[1] TFome Research Group, Bioinformatics Group, Interdisciplinary Center of Bioinformatics, Department of Computer Science, University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany
[2] Paul-Flechsig-Institute for Brain Research, University of Leipzig, Jahnallee 59, D-04109 Leipzig, Germany

Corresponding author: Nowick, Katja (nowick@bioinf.uni-leipzig.de)

Current Opinion in Genetics & Development 2014, **29**:60–67

This review comes from a themed issue on **Genetics of human evolution**

Edited by **Aida M Andrés** and **Katja Nowick**

For a complete overview see the Issue and the Editorial

Available online 15th September 2014

http://dx.doi.org/10.1016/j.gde.2014.08.007

Introduction

The question of how relatively few genomic changes can result in comparatively large phenotypic differences has motivated much research. With the seminal paper by King and Wilson [1], it became increasingly acknowledged that gene expression differences are likely to drive many phenotypic distinctions. Accordingly, sequence changes in gene regulatory regions between and within species have frequently been found and linked to phenotypic alterations (Siepel and Arbiza, 2014, this issue). While it is expected that each single change of this type mainly affects the expression of only the respective gene, changes in the gene regulatory factors (GRFs) could potentially modify the phenotype considerably. This is because a GRF typically regulates the expression of several to many genes (Box 1). Whereas these pleiotropic characteristics of GRFs suggest that many GRFs should be conserved, non-deleterious evolutionary changes in GRFs are prime candidates for driving phenotypic diversity. Such changes can encompass differences in gene copy number, sequence, or expression.

In this review we will focus on two classes of GRFs, transcription factors (TFs) and long non-coding RNAs (lncRNAs). Both groups of genes primarily function at transcriptional level to regulate gene expression by directly or indirectly interacting with DNA. It is important to highlight that other types of non-coding RNA molecules also regulate gene expression, but they mainly operate at post-transcriptional or translational level, for instance, microRNAs, tRNAs, or snoRNAs [2]. Those readers interested in the evolution of e.g. microRNAs can refer to some work on miRNAs in modern human populations [3,4], miRNA changes in anatomically modern humans [5], and miRNA comparisons between humans and other primates [6–10]. We will present here examples of TF and lncRNA genes that have changed during human evolution and discuss how they might be related to human specific phenotypes or have medical consequences. To complement this review, we explored GRF genes within candidate regions that have been found to be under positive selection in recent genome-wide studies [11–13,14•].

Evolutionary changes after the split from chimpanzees

The appearance of new genes can have a strong impact on the way phenotypes evolve. Twenty-three (0.7%) of the 3315 human TFs are human specific (Perdomo-Sabogal *et al.*, in preparation) with no ortholog in other species including chimpanzees [30,31•]. For instance, some members of the family of *Krüppel*-type zinc finger (ZNF) genes and of the *FOXD* subfamily duplicated recently within primate lineages, creating human specific genes. Among them are seven ZNF genes [31•], forkhead box D4-like 5 (*FOXD4L5*), and synovial sarcoma X breakpoint 1 (*SSX1*). Genetic disease associations and medical

[a] They both equally contributed to this work.

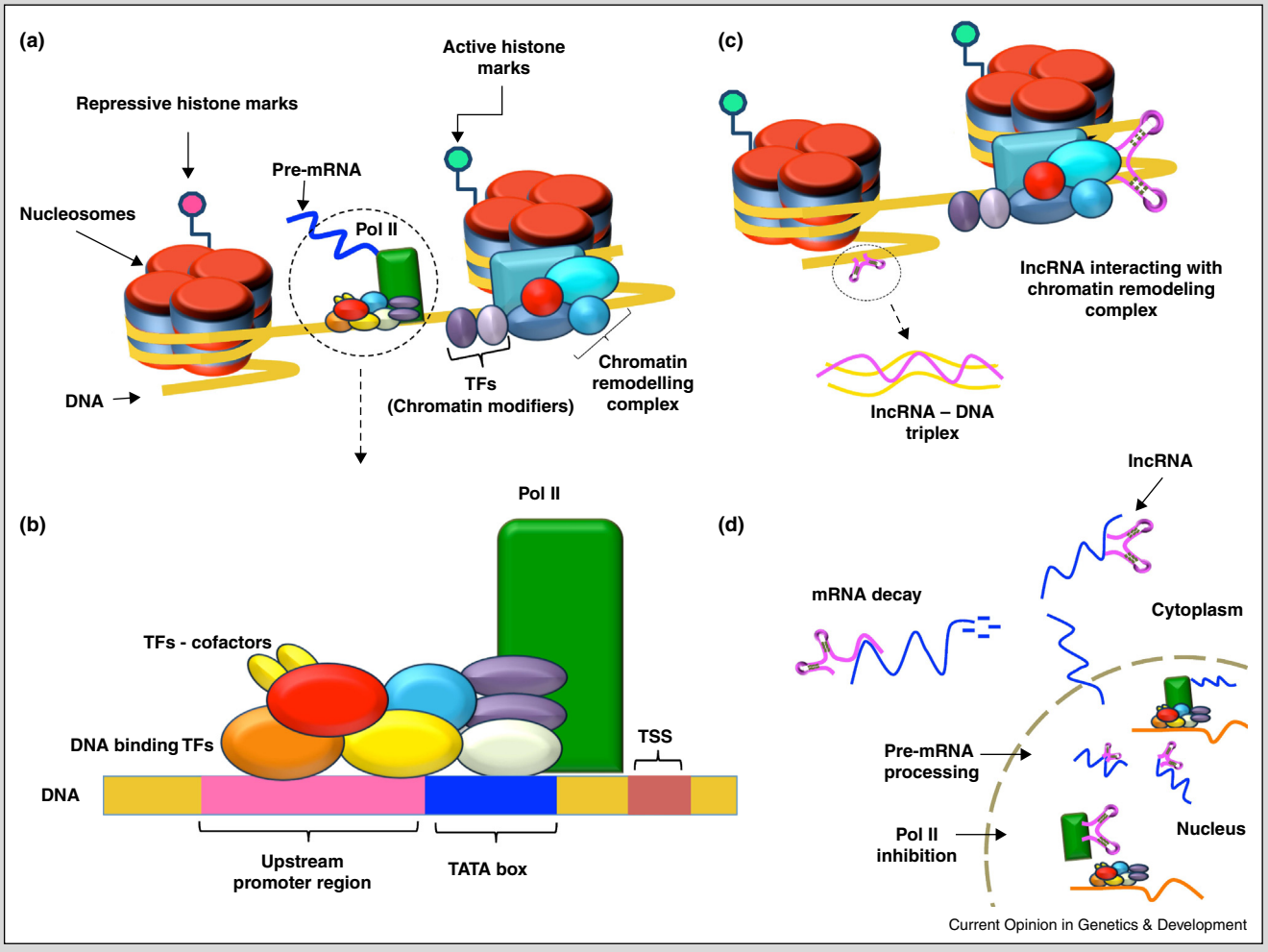
Box 1 The molecular biology of GRFs.

Throughout this paper we will refer to GRFs as molecules that directly or indirectly bind to DNA in order to regulate the expression of target genes. These molecules comprise two major classes, transcription factors (TFs) and long non-coding RNAs (lncRNAs). Different classification schemes for TFs [15,16,17•] and long non-coding RNAs have been developed [18,19••,20,21], but we focus here mainly on their functional classification.

TFs: TFs are proteins that activate or repress gene expression, either in a tissue or time point dependent manner, co-factors that interact with other TFs to fine-tune the expression, and chromatin modifiers that can activate or silence large parts of the genome (Figure B1a,b). The human genome harbors genes for about 3300 TFs (Perdomo-Sabogal *et al.*, in preparation).

LncRNAs: LncRNAs are longer than 200 nucleotides and are functionally similar to TFs in the sense that they are involved in transcriptional regulation [22]. Although doubts about their functional roles have been raised, multiple lines of evidence show that a significant portion of them are not only functional, but are key molecules in gene regulatory pathways [23,24]. LncRNAs are involved in several fundamental processes, such as recruitment of epigenetic modifier proteins, sequestration of TF proteins, post-transcriptional processing, and mRNA decay (Figure B1c,d) (for reviews, see [25,26]). The number of lncRNAs in humans is still a matter of debate. Recent studies reported between 14 000 [27••,28•] or 53 000 [29••] but this number is likely to change as more tissues are sequenced to higher depth.

Figure B1



(a) Schematic representation of chromatin structure and chromatin-mediated gene regulation. TFs acting as chromatin modifiers and chromatin remodeling complexes dynamically modify chromatin architecture and allow the access of the TF machinery to the DNA, thus regulating gene expression. **(b)** TFs bind DNA promoter regions and recruit other TF co-factors to regulate transcription. Upon TF binding, chromatin changes conformation for the RNA polymerase to initiate transcription. Similarly, TFs can bind and interact with other TFs to repress gene expression. **(c)** LncRNAs can sequester TFs or participate in chromatin remodeling. Similarly to TFs, they can activate or repress transcription. In addition, lncRNAs can form lncRNA-DNA triplex structures and regulate the assembling of the pre-initiation complex. **(d)** By binding to polymerase II, lncRNAs can also repress transcription. TSS: transcription start site, Pol II: polymerase II.