Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Transposable element activity, genome regulation and human health

Lu Wang[1,2] and I King Jordan[1,2]

A convergence of novel genome analysis technologies is enabling population genomic studies of human transposable elements (TEs). Population surveys of human genome sequences have uncovered thousands of individual TE insertions that segregate as common genetic variants, i.e. TE polymorphisms. These recent TE insertions provide an important source of naturally occurring human genetic variation. Investigators are beginning to leverage population genomic data sets to execute genome-scale association studies for assessing the phenotypic impact of human TE polymorphisms. For example, the expression quantitative trait loci (eQTL) analytical paradigm has recently been used to uncover hundreds of associations between human TE insertion variants and gene expression levels. These include population-specific gene regulatory effects as well as coordinated changes to gene regulatory networks. In addition, analyses of linkage disequilibrium patterns with previously characterized genome-wide association study (GWAS) trait variants have uncovered TE insertion polymorphisms that are likely causal variants for a variety of common complex diseases. Gene regulatory mechanisms that underlie specific disease phenotypes have been proposed for a number of these trait associated TE polymorphisms. These new population genomic approaches hold great promise for understanding how ongoing TE activity contributes to functionally relevant genetic variation within and between human populations.

Addresses
[1] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[2] PanAmerican Bioinformatics Institute, Cali, Colombia

Corresponding author: Jordan, I. King (king.jordan@biology.gatech.edu)

Current Opinion in Genetics & Development 2018, **49**:25–33

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Kathleen Burns** and **Damon Lisch**

https://doi.org/10.1016/j.gde.2018.02.006

Introduction

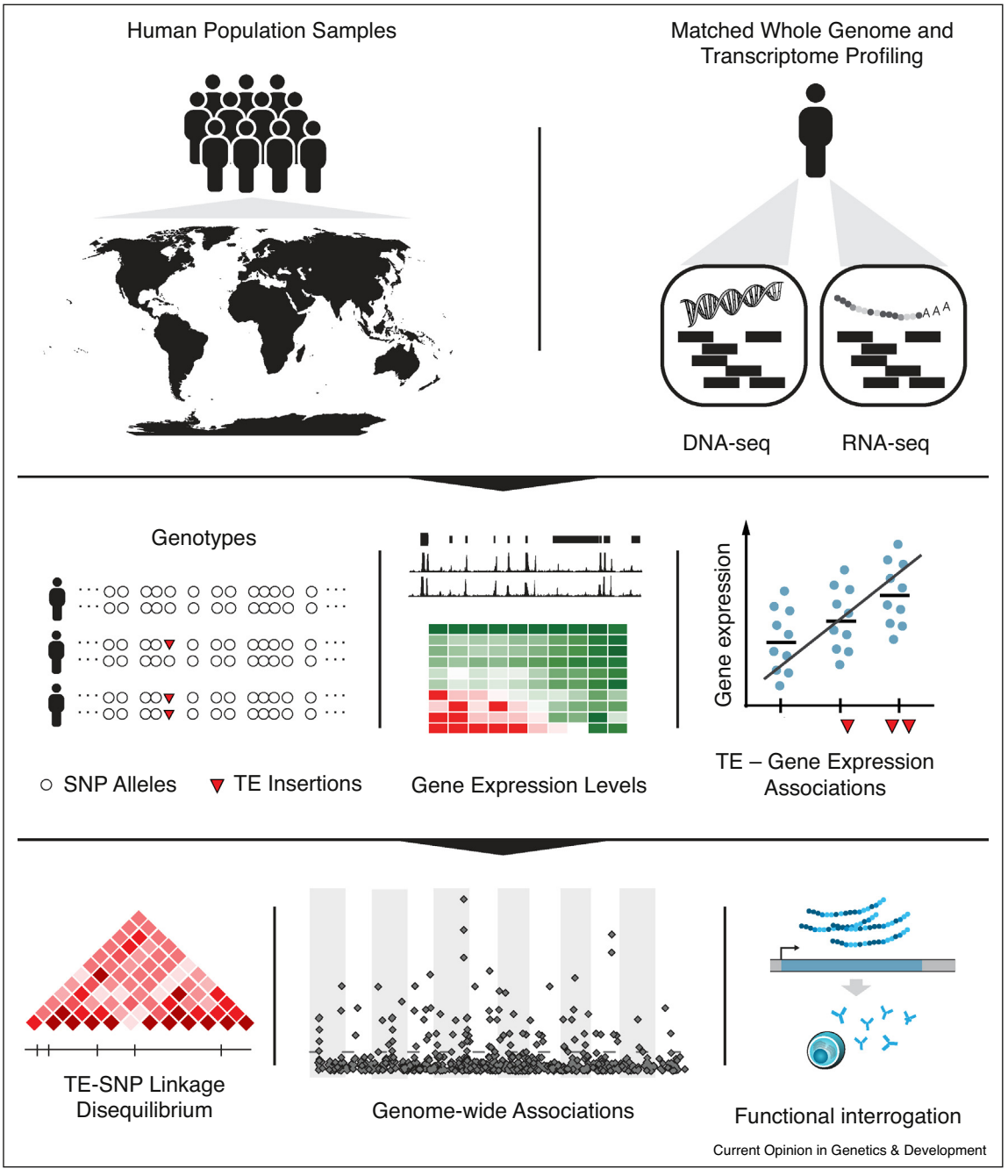
Transposable elements (TEs) are distinguished by their ability to move, i.e. transpose, among genomic locations, often making copies of themselves as they go. TEs can replicate to extremely high copy numbers over time; at least 50% of the human genome sequence is thought to be derived from TE insertions [1,2]. The abundance of TE sequences, along with their ability to colonize a seemingly endless variety of host genomes, begs an explanation for their evolutionary success. The selfish DNA theory holds that TEs are genomic parasites, which play no functional role for their hosts and exist simply by virtue of their ability to out-replicate the genomes in which they reside [3,4]. The selfish DNA theory is still widely considered to represent the null hypothesis that best explains the presence of TEs from an evolutionary standpoint. Nevertheless, numerous studies have revealed instances of exaptation [5], also referred to as molecular domestication [6], whereby formerly selfish TE sequences have been co-opted to provide some functional utility for their host genomes. The most widely observed route of molecular domestication entails the conversion of TE sequences into host genome regulatory elements [7–9].

TE-derived sequences provide a wide variety of regulatory elements to the human genome, including promoters [10–12], enhancers [13,14•,15–17], transcription terminators [18] and several classes of small RNAs [19–21]. Human TE-derived sequences can also exert higher order influences on gene regulation by shaping chromatin structure across the genome [22–26]. It is important to note that, until this time, nearly all studies on human TE regulatory elements have focused on TE-derived sequences that are remnants of relatively ancient insertion events and no longer capable of transposition. Accordingly, known human TE regulatory sequences largely correspond to so-called 'fixed' TE insertions, which are found at the same genomic insertion site locations within the genomes of all human individuals. This distinction is critical, since fixed TE insertions are not expected to contribute to regulatory variation among individual humans. In other words, fixed TE regulatory elements, while functionally important, do not provide a source of human population genetic variation.

Over the last several years, a convergence of genome-enabled technologies has begun to power studies that are focused squarely on structural variations generated by the ongoing activity of human TEs. There are several families of human TEs that retain the ability to transpose, primarily Alu [27,28], L1 [29,30], and SVA [31,32]. Alu and SVA elements are non-autonomous SINEs (Short Interspersed Nuclear Elements), which are mobilized *in*

trans by the transposition machinery encoded from autonomous LINEs (Long Interspersed Nuclear Elements) of the L1 family. Smaller numbers of HERV-K endogenous retroviruses also remain active in the human genome [33]. When members of these TE families transpose within the human genome, they generate inter-individual variations that segregate within and between populations in the form of TE insertion site polymorphisms. Given the known regulatory properties of human TEs, it is not unreasonable to expect that segregating TE polymorphisms could have significant regulatory consequences. In particular, some human TE polymorphisms may lead

Figure 1



The population genomic approach for the study of TE phenotypic effects. Individuals sampled from human populations are characterized using genome (DNA-seq) and transcriptome (RNA-seq) profiling techniques. Genome-wide TE insertion genotypes are compared to tissue-specific gene expression levels to uncover TE variants implicated in gene regulation. The linkage disequilibrium patterns (LD) among TE polymorphisms and SNPs are evaluated to identify TE insertions linked to genome-wide association study (GWAS) loci. Interrogation of functional information is used to hone in on likely TE causal variants.